GENETICS OF COMPLEX TRAITS & DISEASES FROM UNDER-REPRESENTED POPULATIONS

EDITED BY: Segun Fatumo, Tinashe Chikowore and Karoline Kuchenbaecker

PUBLISHED IN: Frontiers in Genetics and Frontiers in Pediatrics

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88974-464-0
DOI 10.3389/978-2-88974-464-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

GENETICS OF COMPLEX TRAITS & DISEASES FROM UNDER-REPRESENTED POPULATIONS

Topic Editors:
Segun Fatumo, University of London, United Kingdom
Tinashe Chikowore, University of the Witwatersrand, South Africa
Karoline Kuchenbaecker, University College London, United Kingdom

Citation: Fatumo, S., Chikowore, T., Kuchenbaecker, K., eds. (2022). Genetics of Complex Traits & Diseases from Under-Represented Populations. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-464-0

Table of Contents

frontiers in Genetics

EDITORIAL
published: 17 January 2022
doi: 10.3389/fgene.2021.817683

Editorial: Genetics of Complex Traits and Diseases From Under-Represented Populations

Segun Fatumo[1,2]*, *Tinashe Chikowore*[3,4] *and Karoline Kuchenbaecker*[5,6]

[1]The African Computational Genomics (TACG) Research Group, MRC/UVRI and LSHTM, Entebbe, Uganda, [2]The Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom, [3]Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [4]MRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [5]Division of Psychiatry, University College of London, London, United Kingdom, [6]UCL Genetics Institute, University College London, London, United Kingdom

Keywords: genomic, underrepresentation, GWAS, genome-wide association study, genetic, complex disease

Editorial on the Research Topic

Genetics of Complex Traits and Diseases From Under-Represented Populations

The Inclusion of ethnically diverse populations in genetic research can provide deep insights into specific pathogenic variants that differ across and within populations. Such insights may be useful to guide diagnosis and provide relevant clinical interventions (Gurdasani et al., 2019; Chikowore et al., 2021; Fatumo et al., 2021a; Fatumo et al., 2021b). Unfortunately, the vast majority of the genetic studies have been performed in European ancestry populations and few in Africa (Kuchenbaecker et al., 2019; Chikowore et al., 2021).

We accepted only high-quality research work in this research topic, including review manuscripts describing the genetics of diseases in underrepresented populations such as African, Asian, Latin American, and other indigenous populations. Topics could include Genome-wide association studies, Gene-environment interactions, Population genetics, epigenetics, pharmacogenomics, Polygenic/Genetic Risk Scores. The final topic issue has 10 published articles out of 18 submitted manuscripts, covering various genetic studies in African, Asian, Arab, and other previously under-studied populations.

Most of the studies demonstrated that better representation of ethnically and ancestrally diverse populations is crucial for our ability to comprehensively characterise the genetic architecture of complex traits and human disease. For example, Font-Porterias et al. analysed genome-wide array data and exome sequences of 89 healthy Spanish Roma individuals to investigate chronic diseases and traits. The Roma represents the largest trans-national minority ethnic group in Europe, originating from South Asia and receiving extensive gene flow from West Eurasia. The authors found frequency differences between the Roma and other non-Roma populations for disease-causing variants and variants linked to drug response. The work demonstrated the value of merging population genetics and genetic epidemiology methodology.

Other papers demonstrated how such differences in clinically relevant genetic variation can impact health inequalities, if populations are not sufficiently represented in genetic research. Chande et al. analysed the effect of genetic ancestry and ethnicity on observed disease prevalence. They predicted disease risk in Colombia populations by mainly comparing the disease prevalence and socioeconomic indicators of Colombia's three major ethnic groups; from Mestizo, Afro-Colombian, and the Indigenous group. The study demonstrated a strong link between ethnicity, ancestry, and health outcomes. In their mini review Nagaraj and Toombs discussed "the history of pharmacogenomics and highlight the inequalities that must be addressed to ensure equal access

OPEN ACCESS

Edited and reviewed by:
Jordi Pérez-Tur,
Institute of Biomedicine of Valencia (CSIC), Spain

***Correspondence:**
Segun Fatumo
segun.fatumo@lshtm.ac.uk

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *18 November 2021*
Accepted: *21 December 2021*
Published: *17 January 2022*

Citation:
Fatumo S, Chikowore T and Kuchenbaecker K (2022) Editorial: Genetics of Complex Traits and Diseases From Under-Represented Populations. Front. Genet. 12:817683. doi: 10.3389/fgene.2021.817683

to pharmacogenomic-based healthcare and reviewed the efforts to conduct the pharmacogenomic profiling of chronic diseases among Australian Indigenous populations including the impact of the lack of drug safety-related information on potential ADRs among individuals in these Australian Indigenous communities".

Papers in this special issue actively addressed underrepresentation by investigating the genetics of complex traits in diverse populations. One approach used to address the imbalance is synthesising existing studies to empower more generalizable conclusions. Soremekun et al. meta-analysed summary statistics from genome-wide association studies of multiple African data, carried out a Bayesian fine-mapping, and assessed the causal relationship between White Blood Cell (WBC) subtypes and Asthma including a two-sample Mendelian Randomization (MR) analysis. This study identified five novel genes not previously reported associated with any WBC subtype. Their MR analysis showed that monocytes count and neutrophils counts are associated with an increased risk of asthma. Singh et al. conducted a systematic review to summarise studies on genetic association studies for blood pressure-related traits in populations with African ancestry. The review highlighted the limited number of available studies, especially in Africa. Nevertheless, across the five blood pressure-related traits, they found that 26 genome-wide significantly associated SNPs have been identified, including 12 associations that have not previously been described in non-African studies.

Similarly, Al-Homedi et al. leveraged several small-scale genome-wide association studies of metabolic syndrome (MetS) performed in Arab populations to investigate the genetic basis of MetS in the population. Their systematic review and meta-analysis identified 36 studies and shows that the most frequently studied genes were FTO and VDR. The study failed to find any unique genetic association between MetS and Arabian populations.

It may be unsurprising that other studies in this Special Issue focussed on cardiometabolic outcomes, given the immense public health burden of cardiovascular disease. Dlamini et al. investigated the associations between genetic variation in CYP17A1 and SERPINA6/A1 and circulating glucocorticoid concentrations in black South African adults. The authors found that a variant rs17090691 with G effect allele at *SERPINA6/A1* was associated with increased diastolic blood pressure in all adults at p-value of 9.47×10^{-6}. However, no association was seen between these genetic variants and glucocorticoids or between any variants in *CYP17A1* and metabolic outcomes after adjusting for multiple testing.

In Wang et al., the authors recruited over 1,400 Han Chinese patients with coronary artery disease. They conducted linear regression analyses on the genetic variants of mitochondrial DNA (mtDNA) and lipidomic traits in two independent groups where they found three associations. They also explored the associations between genetic variants in mitochondrial and lipids traits and reported two significant associations. The study provided "insights into the lipidomic context of mtDNA variations and highlighted the importance of studying mitochondrial genetic variants related to lipid species".

Two of the studies focused on mental health genetics in African populations. Genetic research could play an important role to decrease stigma and better understand the causes of this highly prevalent, yet poorly understood global health problem. Owalla et al. investigated the effects of stress and the genetic variation within the FK506 binding protein 5 (rs1360780) and glucocorticoid receptor (rs10482605) genes on internalizing disorders (ID) status in a CA-HIV cohort in Uganda. Using logistic regression, the authors assessed the relationships between IDs and recent stress, chronic stress, and the investigated genotypes. They reported no significant association between IDs and rs1360780 and rs10482605 polymorphisms within the FKBP5 and glucocorticoid receptor genes nor observed any gene–environment effect on vulnerability IDs in the Ugandan population, but shows that severe recent stress increases the vulnerability to IDs among CA-HIV. The findings further supported that polymorphisms at genetic loci only contributed minimal to the genetic vulnerability to IDs.

Similarly, Kalungi et al. assessed the relationships between acute stress and Internalizing mental disorders (IMDs), and moderation by chronic stress and SNPs in *SLC6A4* and *TPH2* in Ugandan populations. The authors observed a "statistically significant association between severe acute stress and IMDs. Acute stress was found to interact with *5-HTTLPR*-rs25531*S-A-S-A* haplotype to increase the risk for IMDs among Ugandan HIV^+ children and adolescents but no evidence for a combined interaction of acute stress, chronic stress, and *5-HTTLPR*-rs25531 on IMDs".

Overall, the studies in this Special Issue highlighted the importance of genetic research in diverse global populations. They synthesized existing data to enable conclusions for diverse populations, demonstrate population variation linked to clinically relevant genes, discover novel genetic associations, and investigate the implications for health inequalities.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

SF is an Intermediate international fellow funded by the Wellcome Trust grant (220740/Z/20/Z) at the MRC/UVRI and LSHTM. TC is an international training fellow supported by the Wellcome Trust grant (214205/Z/18/Z). KK is funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant agreement No. 948561).

ACKNOWLEDGMENTS

The guest editors wish to thank all the authors and reviewers for their valuable contributions to this Research Topic and we hope that this collection of articles will be of interest to the medical and genetics community.

REFERENCES

Chikowore, T., Kamiza, A. B., Oduaran, O. H., Machipisa, T., and Fatumo, S. (2021). Non-Communicable Diseases Pandemic and Precision Medicine: Is Africa Ready? *EBioMedicine* 65, 103260. doi:10.1016/j.ebiom.2021.103260

Fatumo, S., Chikowore, T., Kalyesubula, R., Nsubuga, R. N., Asiki, G., Nashiru, O., et al. (2021a). Discovery and Fine-Mapping of Kidney Function Loci in First Genome-Wide Association Study in Africans. *Hum. Mol. Genet.* 30, 1559–1568. doi:10.1093/hmg/ddab088

Fatumo, S., Karhunen, V., Chikowore, T., Sounkou, T., Udosen, B., Ezenwa, C., et al. (2021b). Metabolic Traits and Stroke Risk in Individuals of African Ancestry: Mendelian Randomization Analysis. *Stroke* 52, 2680–2684. doi:10.1161/STROKEAHA.121.034747

Gurdasani, D., Carstensen, T., Fatumo, S., Chen, G., Franklin, C. S., Prado-Martinez, J., et al. (2019). Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa. *Cell* 179 (4), 984–1002. e36. doi:10.1016/j.cell.2019.10.004

Kuchenbaecker, K., Telkar, N., Telkar, N., Reiker, T., Walters, R. G., Lin, K., et al. (2019). The Transferability of Lipid Loci across African, Asian and European Cohorts. *Nat. Commun.* 10 (1), 4330. doi:10.1038/s41467-019-12026-7

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 25 March 2021
doi: 10.3389/fgene.2021.630359

Associations of Mitochondrial Variants With Lipidomic Traits in a Chinese Cohort With Coronary Artery Disease

Zixian Wang[1,2,3,4†], Hui Chen[1,2,3†], Min Qin[1,2,3], Chen Liu[5], Qilin Ma[6], Xiaoping Chen[7], Ying Zhang[8], Weihua Lai[2], Xiaojuan Zhang[2] and Shilong Zhong[1,2,3,4]**

[1] Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, School of Medicine, South China University of Technology, Guangzhou, China, [2] Department of Pharmacy, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China, [3] Guangdong Provincial Key Laboratory of Coronary Heart Disease Prevention, Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China, [4] School of Biology and Biological Engineering, South China University of Technology, Guangzhou, China, [5] Department of Cardiology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [6] Department of Cardiology, Xiangya Hospital, Central South University, Changsha, China, [7] Department of Clinical Pharmacology, Xiangya Hospital, Central South University, Changsha, China, [8] Department of Cardiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China

OPEN ACCESS

Edited by:
Segun Fatumo,
University of London, United Kingdom

Reviewed by:
Corey Giles,
Curtin University, Australia
Yan Gong,
University of Florida, United States

***Correspondence:**
Xiaojuan Zhang
zhangxjtj@163.com
Shilong Zhong
gdph_zhongsl@gd.gov.cn;
zhongsl@hotmail.com

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: 17 November 2020
Accepted: 08 March 2021
Published: 25 March 2021

Citation:
Wang Z, Chen H, Qin M, Liu C, Ma Q, Chen X, Zhang Y, Lai W, Zhang X and Zhong S (2021) Associations of Mitochondrial Variants With Lipidomic Traits in a Chinese Cohort With Coronary Artery Disease. Front. Genet. 12:630359. doi: 10.3389/fgene.2021.630359

Plasma lipids have been at the center stage of the prediction and prevention strategies for cardiovascular diseases (CVDs), and novel lipidomic traits have been recognized as reliable biomarkers for CVD risk prediction. The mitochondria serve as energy supply sites for cells and can synthesize a variety of lipids autonomously. Therefore, investigating the relationships between mitochondrial single nucleotide polymorphism (SNPs) and plasma lipidomic traits is meaningful. Here, we enrolled a total of 1,409 Han Chinese patients with coronary artery disease from three centers and performed linear regression analyses on the SNPs of mitochondrial DNA (mtDNA) and lipidomic traits in two independent groups. Sex, age, aspartate aminotransferase, estimated glomerular filtration rate, antihypertensive drugs, hypertension, and diabetes were adjusted. We identified three associations, namely, D-loop$_{m.16089T>C}$ with TG(50:4) NL-16:0, D-loop$_{m.16145G>A}$ with TG(54:5) NL-18:0, and D-loop$_{m.16089T>C}$ with PC(16:0_16:1) at the statistically significant threshold of FDR < 0.05. Then, we explored the relationships between mitochondrial genetic variants and traditional lipids, including triglyceride, total cholesterol (TC), low-density lipoprotein cholesterol (LDLC), and high-density lipoprotein cholesterol. Two significant associations were found, namely *MT-ND6*$_{m.14178T>C}$ with TC and D-loop$_{m.215A>G}$ with LDLC. Furthermore, we performed linear regression analysis to determine on the SNPs of mtDNA and left ventricular ejection fraction (LVEF) and found that the SNP D-loop$_{m.16145G>A}$ was nominally significantly associated with LVEF ($P = 0.047$). Our findings provide insights into the lipidomic context of mtDNA variations and highlight the importance of studying mitochondrial genetic variants related to lipid species.

Keywords: mitochondrial DNA, polymorphisms, lipidomic, association analyses, coronary artery disease

INTRODUCTION

Coronary artery disease (CAD) is one of the greatest threats to human health (Virani et al., 2020). Lipid metabolism plays a central role in the development of cardiovascular diseases (CVDs) and is considered a therapeutic intervention target for CVDs (Ference et al., 2017; Ouchi et al., 2019; Echouffo-Tcheugui et al., 2020). Lipidomic traits can be used independently in predicting long-term CVD risk and adverse clinical outcome. The application of lipidomic traits for risk prediction is more comprehensive than traditional lipids, namely triglycerides (TRIG), total cholesterol (TC), low-density lipoprotein (LDLC), and high-density lipoprotein cholesterol (HDLC) (Razquin et al., 2018; Cadby et al., 2020; Poss et al., 2020).

Mitochondria serve as energy supply sites for cells and can synthesize a variety of lipids autonomously, such as phosphatidylglycerol, cardiolipin, and phosphatidic acid (Horvath and Daum, 2013). Mitochondrial dysfunction can cause various diseases, including CVDs (cardiomyopathy, heart failure, and arrhythmia) (Marín-García and Goldenthal, 2002; Yang et al., 2014; Wahbi et al., 2015). Human mitochondrial DNA (mtDNA) is in a circular double-stranded state and approximately 16.5 kb in length and encodes 37 genes (Sobenin et al., 2012). The replication of mtDNA starts with the displacement loop (D-loop) and contains a transcript promoter adjacent to the D-loop (Sharma and Sampath, 2019). Although the exact function of the D-loop is unclear, this region has a high sequence variability (Ingman et al., 2000). The sequences outside the D-loop region are all exons and encode genes, and thus mtDNA mutations are more likely to occur in the coding region than autosomal DNA mutations and directly affect the physiological functions of the genes encoded by them. Therefore, identifying the associations between mitochondrial genetic patterns and lipid species can provide a deeper understanding of lipid metabolism and its biological links to clinical features and disease progression.

Here, we performed large-scale association analyses on mtDNA variations and lipidomic traits in a Han Chinese population to obtain insights into the relationships between mitochondrial genetic variants and plasma lipidomic traits in CAD patients. We recruited 1,409 subjects with CAD from three centers and used multiple linear regression to explore the relationships between the 42 single nucleotide polymorphisms (SNPs) of mtDNA and 309 lipid species in two independent groups. Moreover, we investigated the relationships between mtDNA variations and the plasma profiling of traditional lipids, including TRIG, TC, LDLC, and HDLC to search more SNPs affecting these lipids. We then explored the effects of mitochondrial genetic variants on heart function by using linear regression to analyze the variations in mtDNA with left ventricular ejection fraction (LVEF). Our findings may provide novel insights into the lipidomic context of mtDNA variations and provide a potential target for pharmacological intervention and optimized treatment of patients with CAD.

Abbreviations: CAD, coronary artery disease; SNP, single nucleotide polymorphism; mtDNA, mitochondrial DNA; TRIG, triglyceride; TC, total cholesterol; LDLC, low-density lipoprotein cholesterol; HDLC, high-density lipoprotein cholesterol; LVEF, left ventricular ejection fraction; CVD, cardiovascular disease; D-loop, displacement loop; PCI, percutaneous coronary intervention; MG, monoglyceride; CE, cholesteryl esters; DG, diacylglycerol; TG, triacylglycerol; PA, phosphatidic acids; PC, phosphatidylcholines; PG, phosphatidylglycerol; PS, phosphatidylserines; PE, phosphatidylethanolamines; LPA, lysophosphatidic acids; LPC, lysophosphatidylcholine; LPE, lysophosphatidylethanolamine; LPS, hemolytic serine; Cer, ceramides; GSA, global screening array; AST, aspartate aminotransferase; eGFR, estimated glomerular filtration rate; FDR, false discovery rate.

MATERIALS AND METHODS

Study Population

Two groups were studied, and the Group I and Group II were mainly grouped according to the time of enrolment. Subjects of the Group I were sequentially enrolled from Guangdong Provincial People's Hospital between January 2010 and December 2013. The samples in Group I were obtained from patients who had complete baseline characteristics and genetic data in our previous study (Liu et al., 2019). Group II was a multicenter cohort including patients from three centers (Guangdong Provincial People's Hospital, Xiangya Hospital, Center South University, and the First Affiliated Hospital, Sun Yat-sen University) from September 2017 to October 2018 to illustrate the repeatability and representation of the results from the Group I. All patients enrolled in this study had taken statin lipid-lowering drugs, diagnosed with CAD through coronary angiography, and had undergone percutaneous coronary intervention (PCI). The exclusion criteria in Group II were the same as those used in the previous study, namely, (1) age $<$ 18 years or $>$80 years; (2) renal insufficiency [defined as serum creatinine concentration $>$ two times the upper limit of normal (230 μmol/L) and renal transplantation or dialysis]; (3) hepatic insufficiency [defined as serum transaminase concentration $>$ two times the upper limit of normal (80 U/L) or diagnosis of cirrhosis]; (4) pregnancy or lactating status; (5) advanced cancer or hemodialysis; (6) history of thyroid problems and the use of antithyroid drugs or thyroid hormone medication; and (7) incomplete information about cardiovascular events during follow-up.

Our study was approved by the Medical Ethical Review Committee of Guangdong General Hospital (Nos. GDREC2010137 and GDREC2017071H) and conducted in accordance with the Declaration of Helsinki. Informed consent (Nos. 20100910 and 20170211) was obtained each participant enrolled in the study. Baseline information, such as demographics, medical history, biochemical measurements, and medication, were all obtained from the hospital information database.

Lipidomic Data

The lipidomic data of the two groups were obtained through widely targeted lipidomic profiling. Details were described in our previous study (Qin et al., 2020), see **Supplementary Material**. In general, 667 lipid species were detected in the Group I,

which contained 14 categories, namely, monoglyceride (MG), cholesteryl esters (CE), diacylglycerol (DG), triacylglycerol (TG), phosphatidic acids (PA), phosphatidylcholines (PC), phosphatidylglycerol (PG), phosphatidylserines (PS), phosphatidylethanolamines (PE), lysophosphatidic acids (LPA), lysophosphatidylcholine (LPC), lysophosphatidylethanolamine (LPE), hemolytic serine (LPS), and ceramides (Cer). 687 lipid species were annotated in Group II, and a total of 309 lipid species were detected in both groups. For lipidomic analyses, raw signals with more than half of missing rate in the QC samples (those with zero ion intensity) were removed. Missing lipidomic data were imputed by replacing the missing value with a minimum value of the lipid species quantified. The lipidomic data after the batch effect correction by Quality Control–based Robust LOESS (Locally Estimated Scatterplot Smoothing) Signal Correction (QC-RLSC) algorithm was then scaled by pareto scaling, which utilized the square root of the standard deviation as the scaling factor (van den Berg et al., 2006).

Mitochondrial Genetic Data

For Group II, each eligible patient fasted for at least 8 h. This procedure was performed to minimize the influence of nutrition on lipid species levels. Blood samples were collected using ethylenediaminetetraacetic acid-coated tubes. Whole blood samples were separated into plasma and hemocytes within 2 h through centrifugation at 2,095 *g* for 10 min at 4°C and then stored at -80°C for further analysis. DNA was extracted from the hemocyte samples for genotyping with a TGuideM16 automatic nucleic acid extractor (Cat. No. OSE-M16) and genomic DNA extraction kit (Cat. No. OSR-M102; TIANGEN). Concentration quantification and electrophoresis were conducted for DNA QC, and qualified DNA was genotyped using an Illumina Infinium global screening array (GSA) bead chip. In brief, the 140 SNPs of mtDNA were genotyped for the two groups. We combined the mitochondrial genetic data of both groups, and the exclusion criteria were as follows: (1) maximum per-person missing rate > 0.05; (2) maximum per-SNP missing rate > 0.05; (3) Hardy–Weinberg disequilibrium *P* value < 1e-6; and (4) minor allele frequency < 0.01.

Association Analyses

Linear regression analyses were conducted at each qualified SNPs of mtDNA for the levels of 309 lipid species presented in Group I and Group II. Sex, age, aspartate aminotransferase (AST), estimated glomerular filtration rate (eGFR), antihypertensive drugs (including β-blockers, angiotensin converting enzyme inhibitors, calcium channel blockers, and proton pump inhibitors), hypertension, and diabetes mellitus were adjusted. All association analyses were conducted using PLINK v2.0 software (Chang et al., 2015).

Meta-Analysis

We used an inverse-variance weighted method meta-analysis, which was based on the effect sizes, and the standard errors adjusting for genomic control were used to combine the association results for the two groups. The heterogeneities in each association between the datasets were tested using the Cochran's Q test. The analyses above were conducted using METAL software (Willer et al., 2010). Only the associations that met the following conditions were presented: (1) *P*-value of both groups < 0.05 and (2) the *P*-value of heterogeneity of effects between the two groups > 0.001. The results of meta-analysis were adjusted to account for multiple testing by the Benjamini-Hochberg false discovery rate (FDR) method and associations were considered statistically

TABLE 1 | Baseline characteristics of the two groups.

Characteristics	Value *N* (%) or mean ± SD	
	Group I	**Group II**
Demographic data		
Size	914	495
Age	62.90 ± 10.06	61.98 ± 9.81
Sex (male)	733 (80.20)	367 (74.14)
BMI, kg/m^2	24.14 ± 3.10	24.13 ± 3.07
Comorbidities		
Arrhythmia	78 (8.53)	44 (8.89)
Diabetes	251 (27.46)	141 (28.48)
Heart failure	77 (8.42)	221 (44.65)
Hypertension	546 (59.74)	300 (60.61)
Hyperlipidemia	101 (11.05)	66 (13.33)
Baseline biochemical measurements		
ALT, U/L	27.54 ± 12.99	25.82 ± 18.71
AST, U/L	26.76 ± 10.77	26.13 ± 39.12
CK, U/L	114.09 ± 115.89	122.87 ± 352.43
eGFR, ml/min/1.73 m^2	95.85 ± 77.82	95.51 ± 126.21
CKMB, U/L	7.60 ± 6.08	12.58 ± 13.73
TC, mmol/L	4.25 ± 1.09	4.28 ± 1.78
LDLC, mmol/L	2.56 ± 0.90	2.70 ± 0.97
HDLC, mmol/L	0.96 ± 0.25	0.99 ± 0.25
TRIG, mmol/L	1.61 ± 1.13	1.84 ± 1.90
GLUC, mmol/L	6.66 ± 2.64	5.95 ± 2.12
Lpa, mg/L	306.68 ± 325.61	287.03 ± 332.98
APOA, g/L	1.04 ± 0.28	1.17 ± 0.24
BNP, pg/mL	782.52 ± 1658.15	1066.73 ± 3209.70
Medication		
β-blockers	810 (88.62)	427 (86.26)
ACEIs	580 (63.46)	239 (48.28)
CCBs	253 (27.68)	150 (30.30)
PPIs	445 (48.69)	346 (69.90)
SYNTAX	16.37 ± 10.71	15.97 ± 13.35

BMI, body mass index; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CK, creatine kinase; eGFR, estimated glomerular filtration rate; CKMB, creatine kinase MB; TC, total cholesterol; LDLC, low-density lipoprotein cholesterol; HDLC, high-density lipoprotein cholesterol; TRIG, triglyceride; GLUC, glucose; Lpa, lipoprotein (a); APOA, apolipoprotein a; BNP, B-type natriuretic peptide; ACEIs, angiotensin converting enzyme inhibitors; CCBs, calcium channel blockers; PPIs, proton pump inhibitors; SYNTAX, Synergy between PCI with TAXUS and Cardiac Surgery score.

significant at FDR < 0.05. All analyses were performed in R v3.6.1[1].

[1]https://www.r-project.org/

Associations of mtDNA With Traditional Lipids and Heart Function

To explore the influence of mitochondrial genetic variants with traditional lipids and heart function, multiple linear

TABLE 2 | Significant associations of mitochondrial genetic variants with lipid species.

Lipid species	mtSNP	Gene/ region	Test	EAF	Effect	SE	*P*-value of meta-analysis	FDR	*P*-value of group I	*P*-value of group II	heterogeneity
TG(50:4) NL-16:0	m.16089T > C	*D-loop*	T > C	0.98	0.65	0.18	0.00035	0.01462	0.01410	0.00929	0.78
TG(54:5) NL-18:0	m.16145G > A	*D-loop*	A > G	0.01	−0.54	0.17	0.00111	0.04670	0.02215	0.00403	0.08
PC(16:0_16:1)	m.16089T > C	*D-loop*	T > C	0.98	0.35	0.11	0.00113	0.04759	0.01040	0.03987	0.62
TG(58:9) NL-20:4	m.15924A > G	*tRNA Thr*	A > G	0.99	0.63	0.21	0.00256	0.10756	0.04615	0.01651	0.41
Cer(m18:1/26:0)	m.16217T > C	*D-loop*	T > C	0.87	0.29	0.10	0.00401	0.16859	0.04969	0.00794	0.10
TG(58:3) NL-18:2	m.16217T > C	*D-loop*	T > C	0.87	0.24	0.09	0.00567	0.23801	0.04579	0.02383	0.22
CE(18:1)	m.6680T > C	*MT-CO1*	T > C	0.91	0.08	0.03	0.00749	0.20076	0.04293	0.04623	0.33

EAF, effect allele frequency; TG, triacylglycerol; PC, phosphatidylcholines; Cer, ceramides; TG, triacylglycerol; CE, cholesteryl esters; SE, standard error; FDR, false discovery rate.

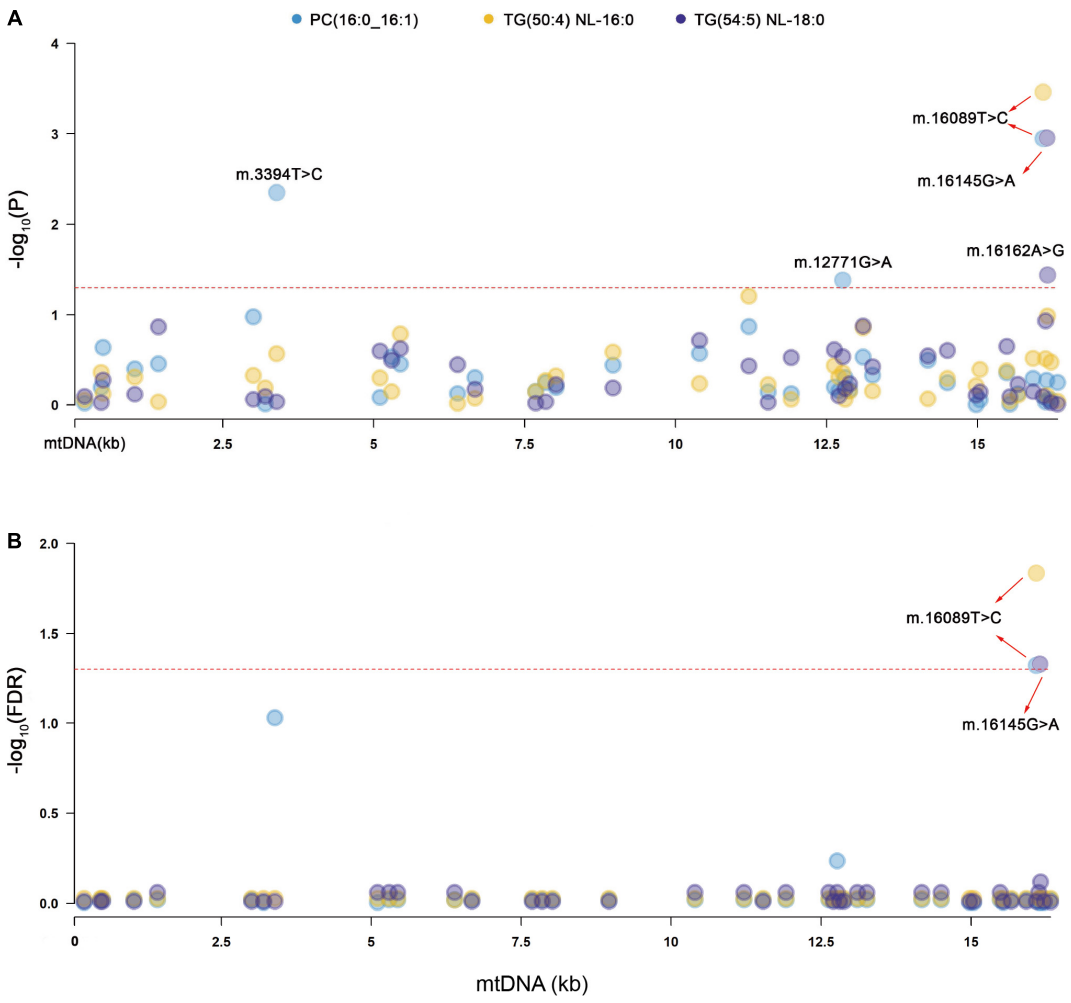

FIGURE 1 | Manhattan plots for the significant associations of mitochondrial genetic variants with lipid species. PC, phosphatidylcholines, TG, triacylglycerol. **(A)** The red dashed line indicates the threshold of nominal significance on lipid species ($P < 0.05$). **(B)** The red dashed line indicates the threshold of statistical significance on lipid species after corrected (FDR < 0.05).

TABLE 3 | Relationships of mtDNA variations with traditional lipids ($P < 0.05$).

Lipid	mtSNP	Gene/ region	Test	Effect	SE	*P*	FDR
TC	m.14178T > C	*MT-ND6*	T > C	1.45	0.34	0.000016	0.000691
	m.215A > G	*D-loop*	A > G	1.08	0.36	0.002684	0.056354
	m.6680T > C	*MT-CO1*	T > C	0.28	0.13	0.026426	0.369958
	m.12811T > C	*MT-ND5*	T > C	0.27	0.13	0.036252	0.380642
	m.3010G > A	*MT-RNR2*	G > A	−0.23	0.12	0.045509	0.382271
LDLC	m.215A > G	*D-loop*	A > G	0.80	0.24	0.001098	0.046129
	m.6680T > C	*MT-CO1*	T > C	0.25	0.09	0.002930	0.061529
	m.12811T > C	*MT-ND5*	T > C	0.24	0.09	0.006208	0.086907
	m.3010G > A	*MT-RNR2*	G > A	−0.20	0.08	0.012639	0.132706
	m.7684T > C	*MT-CO2*	T > C	0.19	0.08	0.015906	0.133611
HDLC	m.6680T > C	*MT-CO1*	T > C	0.06	0.02	0.009879	0.203592
	m.12811T > C	*MT-ND5*	T > C	0.06	0.02	0.010075	0.203592
	m.7684T > C	*MT-CO2*	T > C	0.05	0.02	0.014542	0.203592
	m.7853G > A	*MT-CO2*	G > A	0.05	0.02	0.025186	0.245955
	m.11914G > A	*MT-ND4*	G > A	−0.06	0.03	0.029280	0.245955

TC, total cholesterol; LDLC, low-density lipoprotein cholesterol; HDLC, high-density lipoprotein cholesterol; SE, standard error; FDR, false discovery rate.

regression of the qualified SNPs of mtDNA to the traditional lipids (TRIG, TC, LDLC, and HDLC), and the heart function (LVEF) were conducted with adjusting for confounding factors including sex, age, AST, eGFR, antihypertensive drugs, hypertension, and diabetes mellitus. Benjamini-Hochberg FDR was used in correcting the number of the SNPs of mtDNA for multiple hypothesis testing, and statistical analyses were performed using PLINK v2.0 and R v3.6.1.

RESULTS

Patients Characteristics

This study enrolled 1,409 Han Chinese patients with CAD who were undergoing PCI therapy. A total of 914 subjects were included in Group I (62.90 ± 10.06 years old), and 495 subjects from three centers (353 patients from Guangdong Provincial People's Hospital, 138 patients from Xiangya Hospital, Center South University, and 4 patients from the First Affiliated Hospital, Sun Yat-sen University) were included in Group II (61.98 ± 9.81 years old). All the subjects had complete mitochondrial genetic, lipidomic, and traditional lipids data. The average TRIG levels were 1.61 ± 1.13 and 1.84 ± 1.90 mmol/L for Group I and Group II, respectively. The average TC levels were 4.25 ± 1.09 and 4.28 ± 1.78 for Group I and Group II. The average LDLC levels were 2.56 ± 0.90 and 2.70 ± 0.97 mmol/L for Group I and Group II. The average HDLC levels were 0.96 ± 0.25 and 0.99 ± 0.25 mmol/L for Group I and Group II. The detailed demographic characteristics are presented in **Table 1**.

Mitochondrial Genetic Variants and Lipidomic Data

As indicated in the Materials and methods section, the 42 SNPs of mtDNA, including nine SNPs in the D-loop region, two SNPs on the 12S rRNA, two SNPs on the 16S rRNA, one SNP on the tRNA, and 28 SNPs in the exon region met the quality control criteria for merged genotype data. In the lipidomic data, 309 plasma lipid species, including 7 CE, 23 DG, 172 TG, 50 PC, 18 LPC, 19 LPE, 1 LPS, and 19 Cer were included for the association analyses conducted on each group.

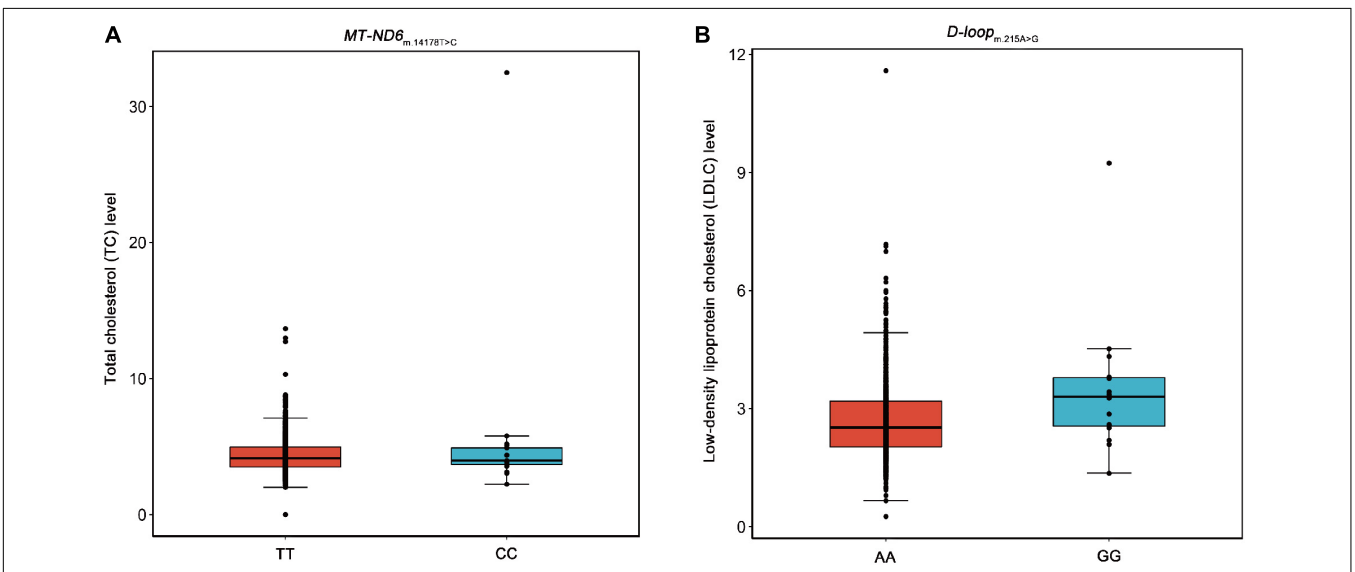

FIGURE 2 | Box plots for the SNPs m.14178T > C and m.215A > G against the levels of TC and LDLC. Five lines from top to bottom for each box were upper extreme, upper quartile, median, lower quartile, and lower extreme. Whiskers were drawn between upper extreme and upper quartile, lower quartile and lower extreme, and the dots above upper extreme or below lower extreme were outliers. **(A)** The SNP m.14178T > C (T > C) in *MT-ND6* against the level of TC. **(B)** The SNP m.215A > G (A > G) in D-loop against the level of LDLC.

TABLE 4 | Relationships of mtDNA variations with LVEF ($P < 0.05$).

mtSNP	Gene/region	Test	Effect	SE	*P*	FDR
m.1048C > T	*MT-RNR1*	C > T	−5.84	2.40	0.01509	0.63390
m.12705C > T	*MT-ND5*	T > C	−1.37	0.66	0.03899	0.65863
m.16145G > A	*D-loop*	G > A	−5.58	2.81	0.04705	0.65863

SE, standard error; FDR, false discovery rate.

Associations of mtDNA Variations With Lipid Species

Primary association analyses were conducted using the 42 SNPs of mtDNA with 309 lipid species in the two groups. The association results were merged through meta-analysis. *P*-values in the two groups are below 0.05, and no significant heterogeneity between them was observed (see section "Materials and Methods"; **Table 2**). In total, we identified seven associations of mitochondrial genetic variants with lipid species ($P < 0.05$). Three associations were statistically significant after FDR correction (**Table 2** and **Figure 1**). m.16089T > C (T > C) in D-loop was associated with the increased level of TG(50:4) NL-16:0 ($P = 0.00035$, FDR = 0.01462), m.16145G > A (A > G) in D-loop was associated with the decreased level of TG(54:5) NL-18:0 ($P = 0.00111$, FDR = 0.04670), and m.16089T > C (T > C) in D-loop was associated with the increased level of PC(16:0_16:1) ($P = 0.00113$, FDR = 0.04759). Four of them were in nominally significant associations ($P < 0.05$, but FDR > 0.05). Specifically, m.15924A > G (A > G) in *tRNA-Thr* was nominally significantly associated with the increased level of TG(58:9) NL-20:4 ($P = 0.00256$, FDR = 0.10756), m.16217T > C (T > C) in D-loop was nominally significantly associated with the increased levels of Cer(m18:1/26:0) ($P = 0.00401$, FDR = 0.16859), and TG(58:3) NL-18:2 ($P = 0.00567$, FDR = 0.23801), and m.6680T > C (T > C) in *MT-CO1* was nominally significantly associated with the increased level of CE(18:1) ($P = 0.00749$, FDR = 0.20076).

Relationships Between mtDNA Variations and Traditional Lipids

Moreover, we performed association analyses between the 42 SNPs of mtDNA and four traditional lipids to explore the relationships of mtDNA variants and traditional lipids (TRIG, TC, LDLC, and HDLC), using a merged genotype and lipids data ($N = 1409$). No SNP was statistically significantly associated with TRIG. The results of potential associations ($P < 0.05$) were presented in **Table 3**. After FDR corrections, two associations were considered statistically significant (FDR < 0.05). m.14178T > C (T > C) in *MT-ND6* was associated with the increased level of TC, and m.215A > G (A > G) in D-loop was associated with the increased level of LDLC (**Figure 2**). All 42 SNPs with the four lipids were presented in **Supplementary Tables 1–4**.

Relationships Between mtDNA Variations and LVEF

We performed further analyses to explore the relationships between the SNPs of mtDNA and heart function. LVEF is a continuous variable that can be used as an index for evaluating heart function, and a low LVEF value implies poor heart function. A total of 1,151 subjects with LVEF data were available in

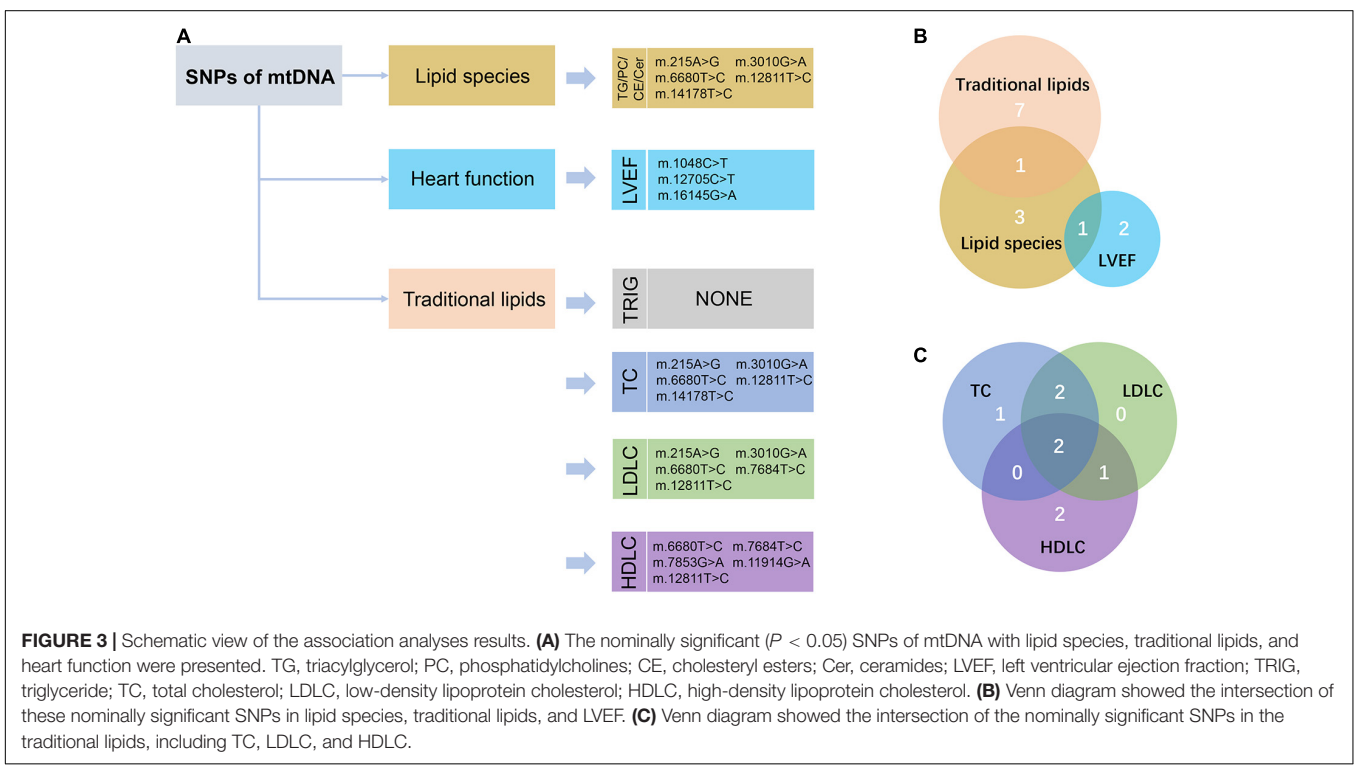

FIGURE 3 | Schematic view of the association analyses results. **(A)** The nominally significant ($P < 0.05$) SNPs of mtDNA with lipid species, traditional lipids, and heart function were presented. TG, triacylglycerol; PC, phosphatidylcholines; CE, cholesteryl esters; Cer, ceramides; LVEF, left ventricular ejection fraction; TRIG, triglyceride; TC, total cholesterol; LDLC, low-density lipoprotein cholesterol; HDLC, high-density lipoprotein cholesterol. **(B)** Venn diagram showed the intersection of these nominally significant SNPs in lipid species, traditional lipids, and LVEF. **(C)** Venn diagram showed the intersection of the nominally significant SNPs in the traditional lipids, including TC, LDLC, and HDLC.

linear regression analyses, whose LVEF(%) was 59.92 ± 11.47 (mean ± standard deviation). Three SNPs were nominally significantly associated with LVEF ($P < 0.05$ but FDR > 0.05). m.1048C > T (C > T) in *MT-RNR1* was nominally significantly associated with decreased LVEF, m.12705C > T (T > C) in *MT-ND5* was nominally significantly associated with decreased LVEF, and m.16145G > A (G > A) in D-loop was nominally significantly associated with decreased LVEF (**Table 4**). The SNP m.16145G > A showed a potential association with LVEF ($P < 0.05$) among the SNPs that were statistically significantly associated with lipid species (FDR < 0.05). All 42 associations were presented in **Supplementary Table 5**.

DISCUSSION

We performed large-scale association analyses between mitochondrial genetic variants and lipidomic traits in a Chinese cohort with CAD. We identified two SNPs (m.16089T > C and m.16145G > A) of mtDNA that were statistically significantly associated with three lipid species (FDR < 0.05) in two independent groups. Then, we performed linear regression to identify the relationships of the SNPs of mtDNA with traditional lipids, and heart function to explore the influence of mitochondrial genetic variants with clinical manifestations. We found that two SNPs (m.14178T > C and m.215A > G) were statistically significantly associated with TC and LDLC (FDR < 0.05), and three SNPs (m.1048C > T, m.12705C > T, and m.16145G > A) were nominally significantly associated with LVEF ($P < 0.05$; **Figure 3**). The results may provide a potential theoretical basis for further explaining the large difference in the levels of lipids and poor prognosis in patients with CAD and provide potential drug intervention targets for precision medicine in patients with CAD.

m.16145G > A (G > A) associated with the increased level of TG(54:5) NL-18:0 but with the decreased level of LVEF were

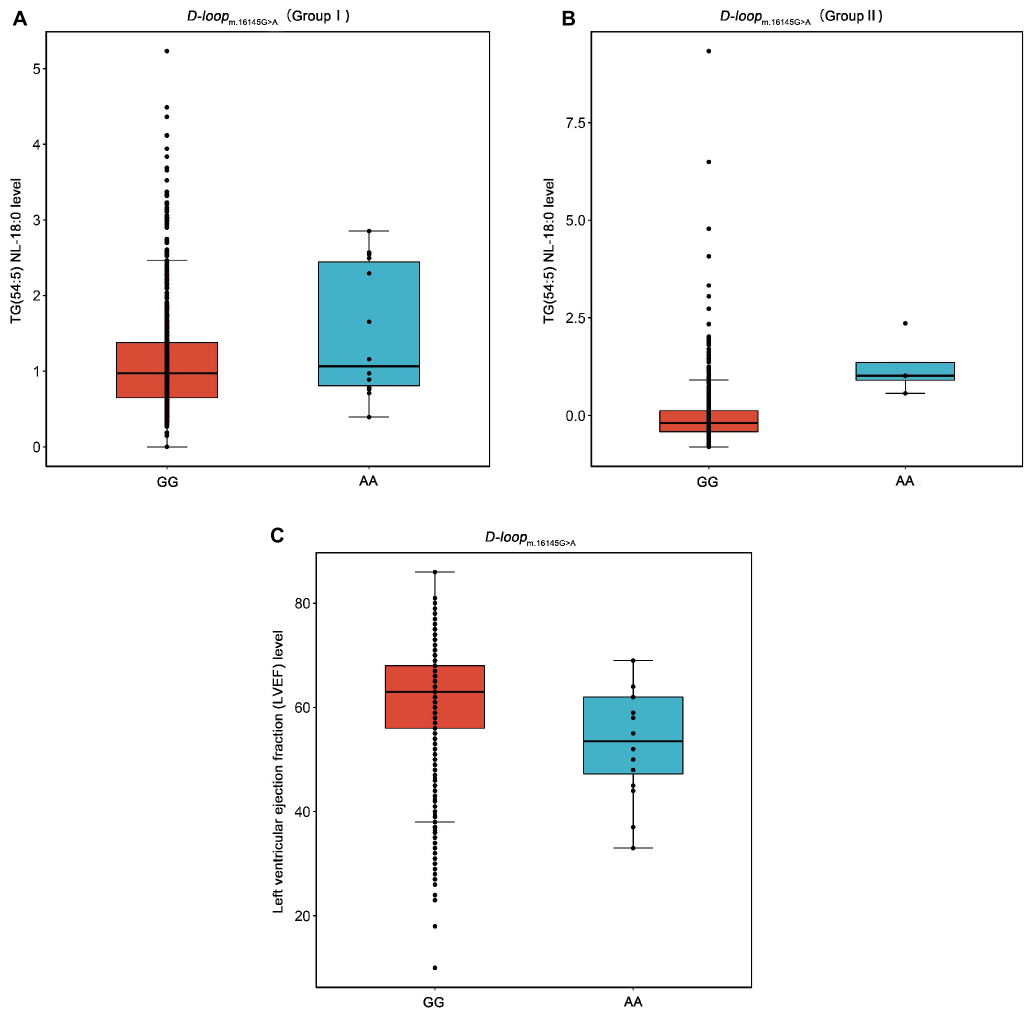

FIGURE 4 | Box plots for the SNP m.16145G > A against the levels of TG(54:5) NL-18:0 and LVEF. TG, triacylglycerol; LVEF, left ventricular ejection fraction. **(A)** The SNP m.16145G > A (G > A) in D-loop against the level of TG(54:5) NL-18:0 in the Group I. **(B)** The SNP m.16145G > A (G > A) in D-loop against the level of TG(54:5) NL-18:0 in the Group II. **(C)** The SNP m.16145G > A (G > A) in D-loop against the level of LVEF.

found in this study (**Figure 4**). This result suggests that the variation at m.16145G > A may bring adverse effect to patients with CAD. Interestingly, the SNPs of mtDNA that are statistically significantly related to the traditional lipids (FDR < 0.05) showed no significant correlation with LVEF (P > 0.05). The variation at the m.16145G > A was not statistically significantly related to the traditional lipids, suggesting the importance of studying mitochondrial genetic variants related to specific lipid species.

The SNPs of mtDNA were significantly associated with the level of lipid species (FDR < 0.05) in the D-loop. The D-loop region is a non-coding region but accumulates more mutations under increased oxidative stress than the other regions of mtDNA (Clayton, 2000). The initiation site for heavy chain replication and the promoter for heavy and light chain transcription are contained in this region (Chen et al., 2012). Therefore, the SNPs in the region of the D-loop may contain essential transcription and replication elements (Sharma et al., 2005). For example, the SNP m.16189T > C in D-loop is associated with vascular pathologies such as stroke and CAD (Liou et al., 2004; Mueller et al., 2011). We speculate that the variants in the D-loop region may affect the metabolism of lipid species, thereby affecting the cardiovascular prognosis.

By analyzing the correlation between mitochondrial genetic variation and plasma TC concentration, we found that m.14178T > C (T > C) in *MT-ND6* could significantly increase TC level (FDR < 0.001). The variation at m.14178T > C is a missense mutation that can change the encoded amino acid (T > C, isoleucine > valine), and *MT-ND6* is a mitochondrial gene that codes the NADH-ubiquinone oxidoreductase chain 6 protein (Shao et al., 2008). NADH is a reduced form of nicotinamide adenine dinucleotide, which is a cofactor in the metabolic center. NADH plays an important role in energy-producing functions of the mitochondria (Stein and Imai, 2012). Therefore, variations in *MT-ND6* may cause energy imbalance and metabolic diseases. A recent study reported that variants in *MT-ND6* can significantly increase the severity of cardiomyopathy (McManus et al., 2019). Our study found that the T to C base in *MT-ND6* m.14178T > C was associated with the increased level of plasma TC, thus possibly contributing to the occurrence and development of atherosclerosis. However, the causal relationship between this mutation and atherosclerosis is still unclear, and further research is needed.

Furthermore, we identified three SNPs that were nominally significantly associated with the level of LVEF (P < 0.05 but FDR > 0.05). The m.1048C > T in *MT-RNR1* was the most significant SNP of the results. *MT-RNR1* is an RNA gene that encodes 12S rRNA, which is mainly involved in the regulation of insulin sensitivity and metabolic balance (Lee et al., 2015). Previous studies reported that the SNP m.1555A > G in *MT-RNR1* was associated with atherosclerosis (Sobenin et al., 2012, 2013). We found that the mutation in this gene could affect heart function, suggesting that *MT-RNR1* may play an important role in the development of CAD. However, its relationship with LVEF level needs further analysis, and a large sample size is needed for the validation of this role.

Our present study has several limitations. Although we enrolled approximately 1,400 participants from three centers, a larger central cohort is needed for fully understanding the relationships between mitochondrial genetic variants and lipidomic profiling. Besides, the patients enrolled in this study were almost treated with statins for lipid-lowering medications, which may influence the plasma levels of cholesterol-derived lipids. Moreover, mitochondrial genetic data were obtained using GSA bead chips, and the number of SNPs detected was relatively small in this study. Increasing the number of SNPs and performing full mitochondrial sequencing could solve this problem in the future. Furthermore, our study can only provide the correlation, and the causality may need to be combined with causality analysis or other approaches, such as Mendelian randomization or molecular biology verification in the future.

CONCLUSION

We performed large-scale association analyses in a Han Chinese population with CAD to explore the relationships between mitochondrial genetic variants and lipidomic profiling. We identified three statistically significant associations (FDR < 0.05), namely, m.16089T > C with TG(50:4) NL-16:0, m.16145G > A with TG(54:5) NL-18:0, and m.16089T > C with PC(16:0_16:1). Then, we further investigated the relationships between mitochondrial genetic variants and traditional lipids. Two statistically significant associations were found (FDR < 0.05), namely, m.14178T > C with TC and m.215A > G with LDLC. Furthermore, we performed association analyses on the SNPs of mtDNA and LVEF. We found that the association between the SNP m.16145G > A and LVEF was nominally significant (P = 0.047). These findings highlight the importance of studying mitochondrial genetic variants related to lipid species and could provide a potential target for pharmacological intervention and optimized treatment of patients with CAD.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: EMBL-EBI (Project: PRJEB42554; Analyses: ERZ1714343).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethical Review Committee of Guangdong Provincial People's Hospital (Nos. GDREC2010137 and GDREC2017071H) and conducted according to the Declaration of Helsinki. Informed consent (Nos. 20100910 and 20170211) was obtained from all individual participants included in the study. The

patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ZW designed the study, performed the data analysis, and wrote the manuscript. HC participated in patient recruitment, performed the experiment, and revised the manuscript. MQ performed the experiment and the data analysis. CL, QM, and XC curated the data. YZ made the investigation. WL provided the resources. XZ and SZ designed the study, leaded the study, and provided the resources. All authors reviewed and approved the final manuscript.

FUNDING

This study was funded by the National Natural Science Foundation of China [No. 81872934 (Principle investigators: Shilong Zhong), 81673514 (Principle investigators: Shilong Zhong); http://www.nsfc.gov.cn/], the National Key Research and Development Program [No. 2017YFC0909301 (Principle investigators: Shilong Zhong); https://service.most.gov. cn/index/], and the Key Research and Development Program of Guangdong Province, China [2019B020229003 (Principle investigators: Shilong Zhong); http://pro.gdstc.gd.gov.cn/egrantweb/]. No funding bodies had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.630359/full#supplementary-material

REFERENCES

Cadby, G., Melton, P. E., McCarthy, N. S., Giles, C., Mellett, N. A., Huynh, K., et al. (2020). Heritability of 596 lipid species and genetic correlation with cardiovascular traits in the Busselton Family Heart Study. *J. Lipid Res.* 61, 537–545. doi: 10.1194/jlr.RA119000594

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7. doi: 10.1186/s13742-015-0047-8

Chen, J. B., Yang, Y. H., Lee, W. C., Liou, C. W., Lin, T. K., Chung, Y. H., et al. (2012). Sequence-based polymorphisms in the mitochondrial D-loop and potential SNP predictors for chronic dialysis. *PLoS One* 7:e41125. doi: 10.1371/journal.pone.0041125

Clayton, D. A. (2000). Transcription and replication of mitochondrial DNA. *Hum. Reprod.* 15(Suppl. 2), 11–17. doi: 10.1093/humrep/15.suppl_2.11

Echouffo-Tcheugui, J. B., Jain, M., and Cheng, S. (2020). Breaking through the surface: more to learn about lipids and cardiovascular disease. *J. Clin. Invest.* 130, 1084–1086. doi: 10.1172/jci134696

Ference, B. A., Ginsberg, H. N., Graham, I., Ray, K. K., Packard, C. J., Bruckert, E., et al. (2017). Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur. Heart J.* 38, 2459–2472. doi: 10.1093/eurheartj/ehx144

Horvath, S. E., and Daum, G. (2013). Lipids of mitochondria. *Prog. Lipid Res.* 52, 590–614. doi: 10.1016/j.plipres.2013.07.002

Ingman, M., Kaessmann, H., Pääbo, S., and Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* 408, 708–713. doi: 10.1038/35047064

Lee, C., Zeng, J., Drew, B. G., Sallam, T., Martin-Montalvo, A., Wan, J., et al. (2015). The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. *Cell Metab.* 21, 443–454. doi: 10.1016/j.cmet.2015.02.009

Liou, C. W., Lin, T. K., Huang, F. M., Chen, T. L., Lee, C. F., Chuang, Y. C., et al. (2004). Association of the mitochondrial DNA 16189 T to C variant with lacunar cerebral infarction: evidence from a hospital-based case-control study. *Ann. N. Y. Acad. Sci.* 1011, 317–324. doi: 10.1007/978-3-662-41088-2_31

Liu, Y., Ma, H., Zhu, Q., Zhang, B., Yan, H., Li, H., et al. (2019). A genome-wide association study on lipoprotein (a) levels and coronary artery disease severity in a Chinese population. *J. Lipid Res.* 60, 1440–1448. doi: 10.1194/jlr.P091009

Marín-García, J., and Goldenthal, M. J. (2002). Understanding the impact of mitochondrial defects in cardiovascular disease: a review. *J. Card. Fail.* 8, 347–361. doi: 10.1054/jcaf.2002.127774

McManus, M. J., Picard, M., Chen, H. W., de Haas, H. J., Potluri, P., Leipzig, J., et al. (2019). Mitochondrial DNA variation dictates expressivity and progression of nuclear DNA mutations causing cardiomyopathy. *Cell Metab.* 29, 78–90.e5. doi: 10.1016/j.cmet.2018.08.002

Mueller, E. E., Eder, W., Ebner, S., Schwaiger, E., Santic, D., Kreindl, T., et al. (2011). The mitochondrial T16189C polymorphism is associated with coronary artery disease in Middle European populations. *PLoS One* 6:e16455. doi: 10.1371/journal.pone.0016455

Ouchi, Y., Sasaki, J., Arai, H., Yokote, K., Harada, K., Katayama, Y., et al. (2019). Ezetimibe lipid-lowering trial on prevention of atherosclerotic cardiovascular disease in 75 or older (EWTOPIA 75): a randomized, controlled trial. *Circulation* 140, 992–1003. doi: 10.1161/circulationaha.118.039415

Poss, A. M., Maschek, J. A., Cox, J. E., Hauner, B. J., Hopkins, P. N., Hunt, S. C., et al. (2020). Machine learning reveals serum sphingolipids as cholesterol-independent biomarkers of coronary artery disease. *J. Clin. Invest.* 130, 1363–1376. doi: 10.1172/jci131838

Qin, M., Zhu, Q., Lai, W., Ma, Q., Liu, C., Chen, X., et al. (2020). Insights into the prognosis of lipidomic dysregulation for death risk in patients with coronary artery disease. *Clin. Transl. Med.* 10:e189. doi: 10.1002/ctm2.189

Razquin, C., Liang, L., Toledo, E., Clish, C. B., Ruiz-Canela, M., Zheng, Y., et al. (2018). Plasma lipidome patterns associated with cardiovascular risk in the PREDIMED trial: a case-cohort study. *Int. J. Cardiol.* 253, 126–132. doi: 10.1016/j.ijcard.2017.10.026

Shao, L., Martin, M. V., Watson, S. J., Schatzberg, A., Akil, H., Myers, R. M., et al. (2008). Mitochondrial involvement in psychiatric disorders. *Ann. Med.* 40, 281–295. doi: 10.1080/07853890801923753

Sharma, H., Singh, A., Sharma, C., Jain, S. K., and Singh, N. (2005). Mutations in the mitochondrial DNA D-loop region are frequent in cervical cancer. *Cancer Cell Int.* 5:34. doi: 10.1186/1475-2867-5-34

Sharma, P., and Sampath, H. (2019). Mitochondrial DNA integrity: role in health and disease. *Cells* 8:100. doi: 10.3390/cells8020100

Sobenin, I. A., Sazonova, M. A., Postnov, A. Y., Bobryshev, Y. V., and Orekhov, A. N. (2012). Mitochondrial mutations are associated with atherosclerotic lesions in the human aorta. *Clin. Dev. Immunol.* 2012:832464. doi: 10.1155/2012/832464

Sobenin, I. A., Sazonova, M. A., Postnov, A. Y., Bobryshev, Y. V., and Orekhov, A. N. (2013). Changes of mitochondria in atherosclerosis: possible determinant in the pathogenesis of the disease. *Atherosclerosis* 227, 283–288. doi: 10.1016/j.atherosclerosis.2013.01.006

Stein, L. R., and Imai, S. (2012). The dynamic regulation of NAD metabolism in mitochondria. *Trends Endocrinol. Metab. TEM* 23, 420–428. doi: 10.1016/j.tem.2012.06.005

van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., and van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 7:142. doi: 10.1186/1471-2164-7-142

Virani, S. S., Alonso, A., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., Carson, A. P., et al. (2020). Heart disease and stroke statistics-2020 update: a report from the American Heart Association. *Circulation* 141, e139–e596. doi: 10.1161/cir.0000000000000757

Wahbi, K., Bougouin, W., Béhin, A., Stojkovic, T., Bécane, H. M., Jardel, C., et al. (2015). Long-term cardiac prognosis and risk stratification in 260 adults presenting with mitochondrial diseases. *Eur. Heart J.* 36, 2886–2893. doi: 10.1093/eurheartj/ehv307

Willer, C. J., Li, Y., and Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190–2191. doi: 10.1093/bioinformatics/btq340

Yang, K. C., Bonini, M. G., and Dudley, S. C. Jr. (2014). Mitochondria and arrhythmias. *Free Radic. Biol. Med.* 71, 351–361. doi: 10.1016/j.freeradbiomed.2014.03.033

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 23 April 2021
doi: 10.3389/fgene.2021.649055

The *5-HTTLPR*-rs25531 *S-A-S-A* Haplotype and Chronic Stress Moderate the Association Between Acute Stress and Internalizing Mental Disorders Among HIV+ Children and Adolescents in Uganda

Allan Kalungi[1,2,3,4], Jacqueline S. Womersley[1,5], Eugene Kinyanda[2,3], Moses L. Joloba[4,6], Wilber Ssembajjwe[2,7], Rebecca N. Nsubuga[7], Soraya Seedat[1,5] and Sian M. J. Hemmings[1,5]*

[1]*Department of Psychiatry, Stellenbosch University, Cape Town, South Africa,* [2]*Mental Health Project, MRC/UVRI and LSHTM Uganda Research Unit, Entebbe, Uganda,* [3]*Department of Psychiatry, Makerere University, Kampala, Uganda,* [4]*Department of Medical Microbiology, Makerere University, Kampala, Uganda,* [5]*South African Medical Research Council/ Stellenbosch University Genomics of Brain Disorders Research Unit, Cape Town, South Africa,* [6]*School of Biomedical Sciences, College of Health Sciences, Makerere University, Kampala, Uganda,* [7]*MRC/UVRI and LSHTM Uganda Research Unit, Statistics and Data Science Section, Entebbe, Uganda*

OPEN ACCESS

Edited by:
Tinashe Chikowore, University of the Witwatersrand, South Africa

Reviewed by:
Aniefiok John Udoakang, University of Ghana, Ghana
Qingsen Ming, First Affiliated Hospital of Soochow University, China

***Correspondence:**
Allan Kalungi
allankalungi1@gmail.com

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *03 January 2021*
Accepted: *26 March 2021*
Published: *23 April 2021*

Citation:
Kalungi A, Womersley JS, Kinyanda E, Joloba ML, Ssembajjwe W, Nsubuga RN, Seedat S and Hemmings SMJ (2021) The 5-HTTLPR-rs25531 S-A-S-A Haplotype and Chronic Stress Moderate the Association Between Acute Stress and Internalizing Mental Disorders Among HIV+ Children and Adolescents in Uganda. Front. Genet. 12:649055. doi: 10.3389/fgene.2021.649055

Background: Internalizing mental disorders (IMDs) among HIV-positive (HIV+) children and adolescents are associated with poor disease outcomes, such as faster HIV disease progression. Although it has been suggested that the development of IMDs is moderated by interaction of stressful life events and vulnerability factors, the underlying etiology is largely unknown. Serotonin transporter gene [solute carrier family 6 member A4 (*SLC6A4*)] and human tryptophan hydroxylase 2 gene (*TPH2*) polymorphisms have been implicated in the development of IMDs. This study investigated the association between acute stress and IMDs, and moderation by chronic stress and genetic variants in *SLC6A4* and *TPH2*.

Hypothesis: Acute stress acts through genetic and environmental vulnerability factors to increase the risk of developing IMDs.

Methods: Polymorphisms in *SLC6A4* (*5-HTTLPR*, rs25531, *5-HTTLPR*-rs25531, and *STin2* VNTR) and *TPH2* (rs1843809, rs1386494, rs4570625, and rs34517220) were genotyped in 368 HIV+ children and adolescents (aged 5–17 years) with any internalizing mental disorder (depression, anxiety disorders, or posttraumatic stress disorder), and 368 age- and sex-matched controls, who were also HIV+. Chronic and acute stress categories were derived by hierarchical cluster analysis. Logistic regression analysis was used to assess the independent moderating effect of chronic stress and each selected polymorphism on the association between acute stress and IMDs.

Results: We observed a statistically significant association between severe acute stress and IMDs (p = 0.001). Children and adolescents who experienced severe acute stress were twice as likely to develop IMDs, compared to children and adolescents who experienced mild acute stress (p = 0.001). Chronic stress interacted with severe acute

stress to increase the risk of IMDs (p = 0.033). Acute stress was found to interact with *5-HTTLPR*-rs25531 *S-A-S-A* haplotype to increase the risk for IMDs among Ugandan HIV+ children and adolescents (p = 0.049). We found no evidence for a combined interaction of acute stress, chronic stress, and *5-HTTLPR*-rs25531 on IMDs.

Conclusion: The odds of having an internalizing mental disorder (IMD) were higher among HIV+ children and adolescents who experienced severe acute stress compared to HIV+ children and adolescents who experienced mild acute stress. Chronic stress and *5-HTTLPR*-rs25531 independently moderated the association between acute stress and IMDs.

Keywords: internalizing mental disorders, acute stress, serotonin transporter gene, *5-HTTLPR*-rs25531, chronic stress, HIV+ children and adolescents, Uganda

INTRODUCTION

HIV-positive (HIV+) children and adolescents suffer a considerable burden of internalizing mental disorders (IMDs; Mellins et al., 2012; Nachman et al., 2012; Kinyanda et al., 2019). IMDs are characterized by quiet, internal distress (Tandon et al., 2011) and include depressive and anxiety disorders, as well as posttraumatic stress disorder (PTSD; American Psychiatric Association, 2013). IMDs among people living with HIV/AIDS have generally been associated negative outcomes of more rapid HIV disease progression (Ironson et al., 2005; Chida and Vedhara, 2009), poor adherence to medication (Springer et al., 2012; Kinyanda et al., 2018), risky sexual behavior (Springer et al., 2012; Kinyanda et al., 2018), and poor linkages to care (Bhatia et al., 2011).

Internalizing mental disorders are complex disorders with gene-environment interactions contributing to their etiology (Musci et al., 2019), where a number of genes play a role with each gene contributing a small effect (Plomin and Davis, 2009; Assary et al., 2018). The role of psychosocial factors and their interaction with biological mechanisms in the etiology of IMDs is still poorly understood. HIV+ children and adolescents experience various chronic life stressors such as awareness of their HIV-status, increased levels of stigma and poorer parental mental health (Betancourt et al., 2014). As chronic stressors are reported to be risk factors for IMDs (Adelman et al., 2014; Robles et al., 2014; Revenson et al., 2016), chronic stressors likely contribute to the etiology of IMDs among HIV+ children and adolescents (Boyes and Cluver, 2015; Lwidiko et al., 2018).

Internalizing mental disorders are heritable (Smoller, 2016): twin studies have estimated a genetic heritability of 35% for depression (Otte et al., 2016) and 30-50% for PTSD (Smoller, 2016). In addition, a meta-analysis of family and twin studies estimated a genetic heritability of 31.6% for generalized anxiety disorder (GAD; Hettema et al., 2001). A more recent genome-wide association study (GWAS) of monozygotic and dizygotic female twins has estimated a genetic heritability of 42% for GAD (Davies et al., 2015). Despite being heritable, the underlying etiology of IMDs is largely unknown although dysregulation in serotonergic transmission has been implicated in depression among adults living with HIV (Hammoud et al., 2010).

Serotonergic pathways in the central nervous system is important in regulating mood and anxiety (Olivier, 2015). After an impulse is fired and serotonin (5-HT) is released into the synaptic cleft, 5-HT re-uptake reduces the activity of the serotonergic neurons, preparing the neuron for a new discharge (Artigas, 2013a,b; Olivier, 2015). Encoded by the serotonin transporter gene [solute carrier family 6 member 4 (*SLC6A4*)], the serotonin transporter (5-HTT) influences serotonergic transmission by regulating the duration of serotonin in the synaptic cleft (Rudnick, 2006; Kristensen et al., 2011). 5-HTT is a target for selective serotonin re-uptake inhibitors (SSRIs) that competitively block substrate binding and thereby prolong neurotransmitter action at the synapse (Kristensen et al., 2011; Cipriani et al., 2018). Tryptophan hydroxylase 2 (TPH2) catalyzes the rate-limiting step in 5-HT biosynthesis (Walther et al., 2003; Carkaci-Salli et al., 2006) and is expressed exclusively in the brainstem (Kennedy et al., 2012), an area which is the major locus of serotonin-producing neurons (Kennedy et al., 2012). Long-term treatment with the SSRI fluoxetine has been found to be associated with concurrent upregulation of *TPH2* messenger ribonucleic acid (mRNA) expression and alleviation of depressive symptoms (Shishkina et al., 2007), suggesting that low *TPH2* gene expression could represent a dysregulation corrected by fluoxetine.

SLC6A4 is located on chromosome 17 at position 17q11 (National Center for Biotechnology Information, 2021a). The serotonin transporter-linked polymorphic region (*5-HTTLPR*) within the promoter region of *SLC6A4* has been implicated in depression (Clarke et al., 2010; Wang et al., 2016), anxiety-related traits (Munafò et al., 2009b), and PTSD (Xie et al. 2009). *5-HTTLPR* variants comprise either 14 (short, *S*-allele) or 16 (long, *L*-allele) copies of a 22–23 base pair (bp) imperfect repeat (Heils et al., 1996). *In vitro* studies show that the *L*-allele has two to three times higher basal transcriptional activity compared to the *S*-allele (Lesch et al., 1996; Philibert et al., 2008). In proximity to *5-HTTLPR* is an *A* to *G* single nucleotide polymorphism (SNP), rs25531 which has been reported to alter expression of the *SLC6A4* by creating a functional AP2 transcription-factor binding site (Hu et al., 2006; Ehli et al., 2012). The rs25531 *G*-allele has been associated with reduced transporter gene expression (Hu et al., 2006; Ehli et al., 2012). This SNP, when analyzed in combination with *5-HTTLPR*, results in *L-A* and *L-G* haplotypes. The *L-G* haplotype has been associated with lower *SLC6A4* expression levels compared to the *L-A* haplotype (Hu et al., 2006; Lipsky et al., 2009).

However, some studies have found no effect of the *L-G* haplotype on *SLC6A4* expression (Martin et al., 2007; Philibert et al., 2008).

A variable number of tandem repeats (VNTR) polymorphism also occurs within the second intron (*STin2*) of *SLC6A4*. The polymorphism consists of multiple repeat copies of a 16–17 bp element (Battersby et al., 1996; Furlong et al., 1998). Three alleles have been reported (National Center for Biotechnology Information, 2021b), containing 9 (*STin2.9*), 10 (*STin2.10*), and 12 (*STin2.12*) copies of the repeat. The alleles have been associated with differential expression of *SLC6A4*, as *STin2.9* has been found to be associated with increased *SLC6A4* expression, and an increasing number of repeats associated with reduced reporter gene expression in rat neonate prefrontal cortical cultures (Ali et al., 2010). The *STin2* VNTR polymorphism has also been found to interact with *5-HTTLPR* to regulate expression of *SLC6A4*, with the combination of the *5-HTTLPR S*-allele and either *STin2.10* or *STin2.12* associated with increased expression, as compared to *STin2.9* whose combination with *5-HTTLPR S*-allele directed expression levels comparable to the *5-HTTLPR S*-allele alone (Ali et al., 2010; Haddley et al., 2012). The *STin2.9* allele has been associated with anxiety in patients with self-harming behaviors (Evans et al., 2013), *STin2.10* has been associated with both anxiety and PTSD (Evans, Li and Whipple, 2013; Xiao et al., 2019) and the *STin2.12* allele has been associated with depression, neuroticism, and suicide (Lopez de Lara et al., 2007; O'Gara et al., 2008; Kalungi et al., 2017).

The TPH2 gene is located on chromosome 12 at position 12q21.1. Located within intron 5 (rs1843809 and rs1386494), and within the promoter region (rs4570625), rs1843809, rs1386494, and rs4570625 have been linked to depression (Zill et al., 2004; Anttila et al., 2009; Gao et al., 2012). Located within the *TPH2* transcription factor binding site, rs34517220 SNP has been reported to modulate *TPH2* expression by altering binding sites for *foxa1* and *foxa2* transcription factors (Pristerà et al., 2015). Also, *foxa2* plays a role in establishing progenitor domains for serotonergic neuron precursors in the ventral hindbrain and in activating transcription factors required for the terminal differentiation of serotonergic neurons (Jacob et al., 2007). Therefore, rs34517220-driven variation in *foxa2* binding may have important effects on the development of neural serotonergic systems.

The diathesis-stress hypothesis of neuropsychiatric disorders postulates that a lower stress threshold is required for psychiatric disease to occur in individuals who harbor certain vulnerability factors, which may be genetic and/or acquired (Monroe and Simons, 1991; Silberg et al., 2001; Caspi et al., 2003; **Figure 1**). Stress is a common environmental risk factor for a number of mental disorders, including depression, anxiety, and PTSD (de Kloet et al., 2005; Popoli et al., 2011), and it is currently accepted that gene-environment interactions underlie the etiology of many, if not all, IMDs (Caspi and Moffitt, 2006; Willis and Brock, 2018; Jawahar et al., 2019).

In the context of the diathesis-stress hypothesis, with acute stress representing the exposure variable in a sample of HIV+ children and adolescents, we investigated polymorphisms in *SLC6A4* (*5-HTTLPR*, rs25531, and *STin2* VNTR) and *TPH2* (rs1843809, rs1386494, rs4570625, and rs34517220) as genetic vulnerability factors and chronic stress as an acquired vulnerability factor for IMDs. Our *a priori* selection of these polymorphisms was based on their associations with IMDs (*5-HTTLPR*, *STin2* VNTR, *TPH2* rs1843809, rs1386494, rs4570625, and rs34517220) and regulation of *SLC6A4* (*STin2* VNTR, rs25531).

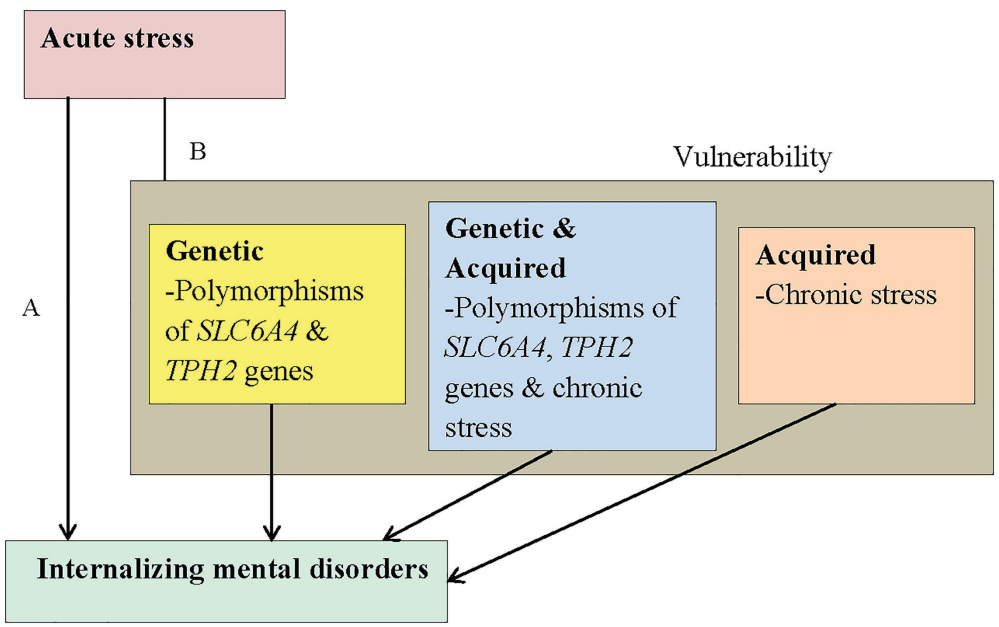

FIGURE 1 | The conceptual framework is adapted from the diathesis-stress model for depression (Monroe and Simons, 1991). The aim of the present study was to investigate the association between acute stress and IMDs **(A)**. We postulated that acute stress would interact with vulnerability factors [**B**; genetic (yellow block), acquired (pink block), or both (blue block)] to increase the risk for IMDs. SLC6A4, serotonin transporter gene; TPH2, tryptophan hydroxylase 2 gene.

MATERIALS AND METHODS

Study Design

This case-control study was nested within a study which investigated mental health among HIV infected children and adolescents in Kampala and Masaka, Uganda (CHAKA study). Characteristics of the study participants and the study design for the CHAKA study are detailed elsewhere (Mpango et al., 2017; Kinyanda et al., 2019).

All study participants were of Bagandan ethnicity, the largest ethnic group in Uganda, and were thus expected to be genetically similar since relatively modest genetic differentiation has been observed among populations representing the major sub-populations in sub-Saharan Africa (Gurdasani et al., 2015). All participants in the present study were HIV+ children and adolescents (aged 5–17 years) of black African (Ugandan) ancestry (Kinyanda et al., 2019). Cases (n = 368) were defined as participants who had any depressive disorder [depression or dysthymia (persistent depressive disorder)] or any anxiety disorder (GAD, separation anxiety disorder, social anxiety disorder, panic disorder, and agoraphobia) or PTSD. Controls (n = 368) were age-, site-, socio-economic status (SES)-, and sex-matched without any psychiatric disorder. Genomic DNA was extracted from an archived cell pellet sample for each participant.

Clinical Assessments

Children and assented adolescents, as well as their caregivers, were interviewed using a structured questionnaire. The questionnaire included, among others, socio-demographic characteristics (sex, study site, age, caregiver level of education, and SES) and modules on different psychiatric disorders from the DSM-5 referenced Children and Adolescent Symptom Inventory 5 (CASI-5; caregiver reported; Gadow and Sprafkin, 2013) and the Youth Inventory-4R (YI-4R; youth reported; Gadow and Sprafkin, 1999). The CASI-5 and YI-4R list the symptoms of a wide range of psychiatric disorders including major depressive disorder, GAD, PTSD, and attention-deficit/hyperactivity disorder, among others. Individual CASI-5 items are rated on a 4-point frequency of occurrence scale ranging from never (0) to very often (3). Though there are several CASI-5 scoring algorithms, in the present study, we used symptom count cut-off scores that reflect the prerequisite number of symptoms for a clinical diagnosis. About 4 ml of blood was withdrawn from each study participant by venipuncture into an EDTA vacutainer and was stored at −80°C pending DNA extraction for the genetics analyses.

Inclusion and Exclusion Criteria

Inclusion criteria were: (i) HIV-infected outpatients, registered with an HIV clinic at any of the study sites; (ii) aged between 5 and 17 years at the time of enrolment; (iii) conversant in English or Luganda, the language into which research assessment tools were translated; and (iv) able to provide written informed consent (caregiver) and assent (adolescents). Exclusion criteria were: (i) seriously ill and (ii) being unable to understand study procedures.

Ethical Considerations

The study was conducted in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). The CHAKA study obtained ethical and scientific clearance from the Uganda Virus Research Institute's Science and Ethical Committee (# GC/127/15/06/459) and the Uganda National Council of Science and Technology (# HS 1601). The study reported here obtained approval from the Higher Degrees Research and Ethics Committee of the School of Biomedical Sciences, Makerere University (# SBS 421) and the Health Research Ethics Committee of Stellenbosch University (# S17/09/179). All caregivers provided informed consent for their children/adolescents to participate in the study and for a blood specimen to be drawn from them (child/adolescent) for genetics analyses. Adolescents further provided informed assent to participate in the study. Study participants who were diagnosed with significant psychiatric problems were referred to mental health units at Entebbe and Masaka government hospitals.

Selection of Cases and Controls

The procedure for selection of cases and controls has been previously described (Kalungi et al., 2021). Briefly, all cases were stratified by site, sex, age category, and SES, resulting in a total of 36 strata. In each stratum, the number of cases was ascertained (e.g., for males in site 1 in the youngest age category and the lowest SES group there were nine cases). An equal number of controls were then randomly sampled from the stratum concerned, thus the controls were matched to the cases by site, sex, age, and SES.

DNA Extraction

DNA was extracted from whole blood using the QiAmp Mini DNA Extraction Kit according to manufacturer's instructions (Qiagen GmbH, Germany).

Selection of Variants Within the Serotonin Transporter and Tryptophan Hydroxylase 2 Genes

The selection of variants was based on their role in regulating gene expression or association with any of the IMDs. Also, because the Bantu group of people in East Africa have been reported to genetically cluster together as per principal components analysis (unpublished results from NeuroGAP data), a minor allele frequency cut-off of at least 0.1 was set for each selected SNP based on the Luhya, a population that belongs to same Bantu group as the population of the present study. The minor allele frequency data were obtained from an online genomic project (Clarke et al., 2017). **Supplementary Table S1** summarizes the biological data based on for selection for each of the *SLC6A4* and *TPH2* variants.

SLC6A4 Genotyping

All polymerase chain reactions (PCR) for *SLC6A4* polymorphisms were performed in a GeneAmp PCR System 9700 (Perkin Elmer Biosystems, Foster City, CA, United States). Amplification reactions were carried out in 25 μl reaction volumes containing: DNA template, 200 μM dNTP (Kapa Biosystems, Cape Town,

South Africa), 5 µl of 10X Taq DNA polymerase buffer (Kapa Biosystems), 1.0 mM magnesium chloride (Kapa Biosystems), 0.625 units (U) Taq DNA polymerase (Kapa Biosystems), and 0.5 µM of each primer (Integrated DNA Technologies, Coralville, IA, United States), with bi-distilled water.

The *5-HTTLPR*, *5-HTTLPR*-rs25531, and *STin2* VNTR polymorphisms were genotyped following a procedure described by Kalungi et al. (2017). Fragment sizes revealed by PCR were confirmed by restriction fragment length analysis on the ABI prism. Expected fragment sizes of the alleles at the *5-HTTLPR*-rs25531 locus were as follows: *S-A* = 281 bp, *L-A* = 325 bp and *S-G-L-G* = 151 bp, resulting into the following genotypes: *S-A-S-A* = 281 bp; *L-A-L-A* = 325 bp; *S-G-S-G*, *L-G-L-G*, *L-G-S-G* = 151 bp; *L-G-S-A* = 151 + 281 bp and *L-A-S-G*, *L-A-L-G* = 325 + 151 bp. Expected fragment sizes of the alleles at the STin2 VNTR locus were as follows: 9-repeat (*STin2.9*) = 250 bp, 10-repeat (*STin2.10*) = 265 bp and 12-repeat (*STin2.12*) = 300 bp, resulting in the following genotypes: 9/9 = 250 bp, 9/10 = 250 + 265 bp, 9/12 = 250 + 300 bp, 10/10 = 265 bp, 10/12 = 265 + 300 bp, and 12/12 = 300 bp.

Determining Genotypes for Selected Single Nucleotide Polymorphims in Tryptophan Hydroxylase 2 Gene

DNA samples were quantified and sent to LGC laboratory (LGC, Middlesex, United Kingdom) for automated SNP genotyping using the kompetitive allele-specific PCR (KASP) assay. Genotypes were determined for *TPH2* rs1843809, rs1386494, rs4570625, and rs34517220. However, none of these genotypes were validated using an alternate method.

Haplotype Analyses

We used Haploview version 4.2 (Barrett, 2009) to analyze linkage disequillibrium (LD) in *SLC6A4* and *TPH2*. *SLC6A4* analyses included 5-*HTTLPR*, rs25531, and *STin2* VNTR polymorphisms. *5-HTTLPR* and *STin2* VNTR alleles were coded as dummy variables, i.e., *S* = 1, *L* = 2, and *STin2.10* = 1 and *STin2.12* = 2. As very few participants carried the *STin2.9* allele (n = 6), they were excluded from the analysis.

TPH2 analyses included the three single nucleotide polymorphisms (SNPs) genotyped. Participants with more than 50% missing genotypes were excluded from the analysis, yielding data from 692 participants for each of the *SLC6A4* and *TPH2* analyses. None of the investigated *TPH2* polymorphisms were in LD (D' < 0.8), while *SLC6A4 5-HTTLPR* and rs25531 were in LD (D' > 0.8) (**Supplementary Figures S1**, **S2**).

Generation of Acute and Chronic Stress Class Categories

Social disadvantage variables were grouped into an index of acute and chronic stress. Caregiver mental state (assessed as psychological distress using the Self-report Questionnaire-20; Beusenberg et al., 1994), child-caregiver relationship [assessed as child-caregiver interactions, using data on how often the caregiver (i) beats, (ii) insults, (iii) spanks, or (iv) yells at the child/adolescent] and HIV symptoms were grouped together to constitute "acute stress" and orphanhood, study site (urban vs. rural), and caregiver level of education as variables constituting "chronic stress." Variables were scored on a disadvantage scale where, for example, double orphanhood carried a higher chronic stress score vs. single orphanhood or not orphaned; food availability: not enough food carried a higher chronic stress score vs. enough food; study site: urban carried a higher chronic stress score than rural; and caregiver level of education: no formal education carried a higher chronic stress scores than primary and primary a higher stress score than secondary, etc. Hierarchical cluster analysis using Statistica 13.5 software (TIBCO, CA, United States), Euclidian distance, as distance measure and Ward's method for clustering (Ward, 1963), was used to generate the different cut-off points for each acute and chronic stress class, respectively.

The acute stress index ranged from 0 to 2.46, with a normal distribution, while the chronic stress index ranged from 0 to 3.75, with a normal distribution as well. A total of three classes were generated for each type of stress by the hierarchical cluster analysis, i.e., mild, moderate, and severe. For acute stress, the mild class had an acute stress score of less than 0.362, the moderate class a score of 0.362–0.622, while the severe class a score of greater than 0.622. For chronic stress, the mild class had a chronic stress score of less than 1.375, the moderate class a score of 1.375–2.375, while the severe class a score of greater than 2.375.

Power of the Study

Using Stata 15 (StataCorp, TX, United States) software, we did a *post hoc* power calculation. Given a case-control ratio of 1:1, and assuming a zero correlation of exposure between cases and controls, a probability of 0.5 of exposure among controls at a 0.05 significance level, and an expected odds ratio of at least 1.8 for IMDs under severe acute stress, we needed 188 cases to achieve a power of 80% (Lachin, 1992). Our study comprising 368 cases was, therefore, adequately powered to address the main aim. Assuming similar assumptions, our interaction analyses achieved a power of 87.5%.

Statistical Methods

Statistical analyses were conducted using Stata 15 (StataCorp, TX, United States). Socio-demographic characteristics were compared between cases and controls. SES was generated from a scale of nine household items owned (car, motorcycle, refrigerator, electricity, bicycle, radio, telephone, cupboard, and flask) as previously described (Kalungi et al., 2019). A *t*-test was used to compare the distribution of cluster of differentiation 4 (CD4) counts between cases and controls. We computed 95% confidence intervals and statistical significance was set at a value of p less than or equal to 0.05.

Genotype distributions were compared between cases and controls. Likelihood-ratio tests were used to test genotypes for the Hardy-Weinberg equilibrium (HWE) in both cases and controls. Logistic regression was used to assess the relationship between *TPH2* rs1843809 and IMDs (equation for the model was $Y = \beta_0 + \beta_1 X_1$, where Y = IMDs, β_0 = constant,

X_1 = rs1843809, and β1 = coefficient for rs1843809). Logistic regression models were used to assess the relationship between acute stress and IMDs (equation for the model was $Y = \beta_0 + \beta_1 X_1$, where Y = IMDs, β_0 = constant, X_1 = acute stress, and β1 = coefficient for acute stress).

The moderating effect of chronic stress on the association between acute stress and IMDs was assessed by comparing logistic regression models of the association between acute stress and IMDs with and without chronic stress. Interactions between acute stress and chronic stress were tested using a likelihood ratio test (equation for the model was $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$, where Y = IMDs, β_0 = constant, X_1 = acute stress, X_2 = chronic stress and β_1–β_3 are the coefficients for acute stress, chronic stress, and the interaction term, respectively).

The moderating effect of each genotype on the association between acute stress and IMDs was assessed by comparing logistic regression models of the association between acute stress and IMDs with and without the polymorphism controlling for chronic stress. Interactions between acute stress and each of the polymorphism were tested using a likelihood ratio test, controlling for chronic stress (equation for the model was $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2$, where Y = IMDs, β_0 = constant, X_1 = acute stress, X_2 = polymorphism/SNP, X_3 = chronic stress and β_1–β_4 are the coefficients for acute stress, the polymorphism/SNP, chronic stress, and the interaction, respectively). These interaction models were performed on all the explanatory variables even without observing significant main effects, in order to rule out the possibility of cross-over interaction, where significant interactions may be observed for non-significant main effects (Widaman et al., 2012). Three-way interactions of acute stress, chronic stress, and any selected polymorphism on IMDs were not assessed as the two-way models were better than the three-way models, based on goodness of fit test, Arkaike informatics criteria, and the Bayesian information criterion. The 95% confidence intervals were calculated for all analyses.

RESULTS

Socio-demographic factors were similarly distributed between cases and controls as shown in **Table 1**.

The HWE *p*-values among cases and controls for each polymorphism are shown in **Table 2**. None of the genotype frequencies deviated significantly from HWE among controls except for rs1386494, which was thus omitted from further analysis. For *SLC6A4* polymorphisms, *5-HTTLPR* was in LD with rs25531 (D' > 0.8) (**Supplementary Figure S1**), while for *TPH2* SNPs, none of the selected polymorphisms were in LD (D' < 0.8) (**Supplementary Figure S2**).

The association between each investigated polymorphism and IMDs is shown in **Table 2**. There was a protective effect against IMDs in participants with *TPH2* rs1843809 *TG* and *TT* genotypes compared with participants with the *GG* genotype (**Table 3**).

TABLE 1 | Distribution of socio-demographic factors in cases and controls.

Variable (*n*, %)	Case *n* (%)	Control *n* (%)	*p*
Sex			0.111
Male (342, 46.5%)	160 (43.6)	182 (49.5)	
Female (393, 53.5%)	207 (56.4)	186 (50.5)	
Site			0.941
Urban (415, 56.4%)	208 (56.5)	207 (56.3)	
Rural (321, 43.6%)	160 (43.5)	161 (43.7)	
Age			0.374
7-11 years (389, 55.9%)	202 (57.6)	187 (54.2)	
12-17 years (307, 44.1%)	149 (42.4)	158 (45.8)	
Education level			0.371
No formal education (13, 1.8%)	9 (2.5)	4 (1.1)	
Primary (648, 88.4%)	323 (88.0)	325 (88.8)	
Secondary (72, 9.8%)	35 (9.5)	37 (10.1)	
Socioeconomic status			0.459
Low (332, 45.1%)	171 (46.5)	161 (43.8)	
High (404, 54.9%)	197 (53.5)	207 (56.2)	
Mean CD4 count at baseline	947.04	944.02	0.939

CD4, cluster of differentiation 4; primary, 0–7 years of formal education; Secondary, 8–14 years of formal education; Low socioeconomic status, 0–13; High socioeconomic status, >13. All numbers that do not add up to the total of 736 participants were due to missing data.

Table 4 shows the association between acute stress and IMDs. A significant association was observed between acute stress and being an IMD case. **Table 5** shows the independent effect of chronic stress and *SLC6A4 5-HTTLPR*-rs25531 haplotype on the association seen in **Table 4** (between acute stress and IMDs). The independent effect of all other polymorphisms on the association between acute stress and IMDs are shown in **Supplementary Table S2**.

Apart from the *SLC6A4 5-HTTLPR*-rs25531 haplotype (**Table 5**), we found no significant interactions between acute stress and any polymorphism on IMDs (**Supplementary Table S2**). A significant interaction between acute and chronic stress (p = 0.033) and acute stress and *5-HTTLPR*-rs25531 *S-A-S-A* haplotype (p = 0.049) was observed on IMDs (**Table 5**). The odds of having an IMD were 4.3 times higher among participants under both severe acute stress and severe chronic stress, compared to those under mild acute stress and mild chronic stress (**Table 5**). The odds for being a case of an IMD were 14.8 times higher in participants with the *S-A-S-A* haplotype who experienced acute stress as compared to participants with the *L-A-L-A* haplotype who experienced mild acute stress. Similarly, the odds of having an IMD were 11.95 times higher in *S-A-S-A* participants with severe acute stress compared to *L-A-L-A* participants with mild stress (**Table 5**).

DISCUSSION

This study investigated the association between acute stress and IMDs, and whether this association was moderated by chronic stress or selected genetic variants in *SLC6A4* and *TPH2*. To our knowledge, this is the first sub-Saharan African study to investigate these interactions among HIV+ children

TABLE 2 | Distribution of genotypes and Hardy-Weinberg equilibrium analysis between cases and controls for *SLC6A4* [*5-HTTLPR*, rs25531, *5-HTTLPR*-rs25531, and *STin2* VNTR] and *TPH2* [rs1843809, rs1386494, rs34517220, and rs4570625] polymorphisms.

Genotype	Cases (*n* = 368) *n* (%)	Controls (*n* = 368) *n* (%)	*p**	HWE Cases	HWE Controls
5-HTTLPR (*n* = 698)			0.286	0.323	0.1916
LL	223 (62.3)	231 (67.9)			
LS	115 (32.1)	94 (27.7)			
SS	20 (5.6)	15 (4.4)			
rs25531 (*n* = 685)			0.988	0.795	0.641
AA	227 (65.4)	221 (65.4)			
AG	108 (31.1)	106 (31.4)			
GG	12 (3.5)	11 (3.2)			
5-HTTLPR-rs25531 haplotype (*n* = 685)			0.688	N/A	N/A
L-A-L-A	130 (37.4)	136 (40.2)			
L-A-L-G	76 (21.9)	82 (24.3)			
L-A-S-A	79 (22.8)	71 (21.0)			
L-A-S-G	2 (0.6)	Absent			
L-G-L-G	12 (3.5)	11 (3.3)			
S-A-S-A	18 (5.2)	15 (4.4)			
S-A-L-G	30 (8.6)	23 (6.8)			
STin2 VNTR (*n* = 687)			0.203	0.090	0.440
10/10	26 (7.5)	28 (8.8)			
10/12	126 (36.3)	112 (35.3)			
12/12	189 (54.5)	177 (55.9)			
9/9	5 (1.4)	Absent			
9/12	1 (0.3)	Absent			
rs1843809 (*n* = 686)			0.003	0.009	0.537
TT	105 (30.1)	113 (33.5)			
TG	150 (43.0)	169 (50.2)			
GG	94 (26.9)	55 (16.3)			
rs1386494 (*n* = 685)			0.06	0.956	0.007
GG	210 (59.7)	202 (60.7)			
GA	124 (35.2)	125 (37.5)			
AA	18 (5.1)	6 (1.8)			
rs34517220 (*n* = 685)			0.945	0.557	0.913
AA	90 (25.6)	84 (25.1)			
AG	170 (48.4)	166 (49.7)			
GG	91 (25.9)	84 (25.2)			
rs4570625 (*n* = 683)			0.574	0.787	0.229
GG	88 (25.2)	90 (26.9)			
GT	177 (50.7)	156 (46.7)			
TT	84 (24.1)	88 (26.4)			

*5-HTTLPR, serotonin transporter linked polymorphic region; STin2 VNTR, serotonin transporter intron 2 variable number of tandem repeats; HWE, Hardy-Weinberg equilibrium; N/A, not applicable. *Value of p for association of each investigated polymorphism with IMDs.*

and adolescents. Results revealed a statistically significant association between acute stress and IMDs. This association was found to be moderated by *5-HTTLPR*-rs25531 in a haplotype-dependent manner. Specifically, in comparison to *L-A-L-A* haplotype carriers with experience of mild acute stress, individuals carrying the *S-A-S-A* haplotypes were more likely to have an IMD under conditions of moderate and severe acute stress. We found no evidence for a significant combined interaction of acute stress, chronic stress, and *5-HTTLPR*-rs25531 on IMDs.

TABLE 3 | Association between tryptophan hydroxylase 2 gene rs1843809 and internalizing mental disorders (IMDs).

IMDs	Odds ratio	*p* > |Z|	95% confidence interval
rs1843809 *GG*	Reference		
rs1843809 *TG*	0.52	0.001	0.349–0.774
rs1843809 *TT*	0.54	0.005	0.355–0.832

IMDs, internalizing mental disorders.

TABLE 4 | Association between acute stress and IMDs.

Acute stress class	OR (IMDs)	*p* > |Z|	95% CI	*p*
Mild	Reference			
Moderate	1.1	0.687	0.758–1.523	0.001
Severe	1.9	0.001	1.298–2.651	

OR, odds ratio; IMDs, internalizing mental disorders; CI, confidence interval.

Stress represents a prevalent environmental risk factor for many mental disorders, including depression and anxiety (de Kloet et al., 2005; Popoli et al., 2011). Acute stress has been associated with the likelihood of developing IMDs among samples of adolescent school children and adult disaster workers (Fullerton et al., 2004; O'Connor et al., 2010; Brown et al., 2016). In line with previous studies, where acute stress has been associated with IMDs such as depression, PTSD (Fullerton et al., 2004), and anxiety (Grillon et al., 2007), we found that risk for IMDs increased with increasing acute stress and was highest for severe acute stress. Chronic stress was found to significantly moderate the association between acute stress and IMDs. Previous studies have reported on chronic stress as a risk factor for IMDs (Charney and Manji, 2004; de Kloet et al., 2005; Evans et al., 2013; Adelman et al., 2014; Robles et al., 2014; Revenson et al., 2016). Vulnerability to IMDs by chronic stress (through an interaction with acute stress) could be due to alterations in the functioning of the hypothalamic-pituitary-adrenal axis that follow exposure to chronic stress (de Kloet et al., 2005; Fuchs and Flïugge, 2006). In addition, chronic stress has been reported to affect serotonergic signaling in the brain (van den Buuse and Hale, 2019). As serotonergic systems have been implicated in threat appraisal, psychophysiological measures of stress response (skin conductance and startle reactions), and attentional bias to negative stimuli, it is possible that alterations in 5-HT signaling due to chronic stress may at least partially contribute to some of the behavioral features of IMDs (Pergamin-Hight et al., 2012; Klumpp et al., 2014; Klumpers et al., 2015).

In contrast to previous findings of associations between *5-HTTLPR* and *STin2* VNTR and IMDs (Battersby et al., 1996; Evans et al., 2013; Gressier et al., 2013; Zhao et al., 2013; Smoller, 2016; Wang et al., 2016; Xiao et al., 2019), we found no direct association between any of the investigated *SLC6A4* polymorphisms and IMDs. This may be due to the complex nature of IMDs, where many genetic variants contribute to disease risk, each

TABLE 5 | Logistic regression analyses for the interaction of acute stress with *5-HTTLPR*-rs25531 and chronic stress, respectively, on IMDs.

Model	Variable	Odds ratio	$p > \|Z\|$	95% CI	*p*
Excluding any polymorphism or chronic stress	Mild acute stress	Reference			0.046[a]
	Moderate acute stress	1.22	0.282	0.848–1.761	
	Severe acute stress	1.92	0.001	1.327–2.785	
Including 5-*HTTLPR*-rs25531	**Acute stress*5-HTTLPR-rs25531**				
	Mild AS**L-A-S-G*	1			0.049[b]
	Moderate AS**L-A-L-G*	2.52	0.022	1.142–5.540	
	Moderate AS**L-A-S-A*	1.60	0.235	0.736–3.497	
	Moderate AS**L-A-S-G*	0.63	0.124	0.351–1.135	
	Moderate AS**L-G-L-G*	0.83	0.870	0.092–7.535	
	Moderate AS**S-A-L-G*	1.48	0.563	0.395–5.500	
	Moderate AS**S-A-S-A*	14.83	0.030	1.294–170.010	
	Severe AS**L-A-L-G*	3.72	0.002	1.630–8.474	
	Severe AS**L-A-S-A*	2.82	0.013	1.248–6.350	
	Severe AS**L-A-S-G*	1.16	0.629	0.641–2.091	
	Severe AS**L-G-L-G*	0.85	0.872	0.121–6.010	
	Severe AS**S-A-L-G*	1.58	0.493	0.426–5.881	
	Severe AS**S-A-S-A*	11.95	0.011	1.755–81.331	
Including chronic stress	**Acute stress*Chronic stress**				
	Moderate AS*moderate CS	1.04	0.896	0.607–1.768	0.033[c]
	Moderate AS*severe CS	1.80	0.035	1.041–3.106	
	Severe AS*moderate CS	0.95	0.904	0.406–2.217	
	Severe AS*severe CS	4.28	0.001	1.832–10.012	

*AS, acute stress; CS, chronic stress; Acute stress*5-HTTLPR-rs25531, interaction between acute stress and 5-HTTLPR-rs25531 genotypes on internalizing mental disorder (IMDs).*
[a]*Value of p for association between acute stress and IMDs (the model without chronic stress/SLC6A4).*
[b]*Value of p for the likelihood-ratio test of interaction between acute stress and 5-HTTLPR-rs25531 on IMDs.*
[c]*Value of p for the likelihood-ratio test of interaction between acute stress and chronic stress on IMDs.*

contributing a small effect (Plomin and Davis, 2009) and thus the effect of a few variants may not be powerful enough to show significant association. Indeed, studies examining the contributions of *5-HTTLPR* and *STin2* variants have yielded inconsistent results, with other published work similarly failing to find a significant association (Munafò et al., 2009a,b; Culverhouse et al., 2018; Xiao et al., 2019). In order to resolve these contradictory results, studies looking at epistatic interactions are required.

Based on the diathesis-stress hypothesis of neuropsychiatric disorders, we postulated that the association between acute stress and IMDs would be moderated by genetic polymorphisms, chronic stress, or a combination of genetic polymorphisms and chronic stress. Results revealed chronic stress and the *5-HTTLPR*-rs25531 *S-A-S-A* haplotype to each significantly moderate the association between acute stress and IMDs.

The interaction between the *5-HTTLPR* and stress has previously been investigated in gene-environment studies of IMDs (Karg et al., 2011; Sharpley et al., 2014). In contrast to previous studies that have reported the moderating role of *5-HTTLPR* on the association between stress and IMDs (Caspi et al., 2003; Karg et al., 2011; Conway et al., 2014; Sharpley et al., 2014), we found no significant moderating effect of *5-HTTLPR* on the association between acute stress and IMDs. However, our results revealed that the combined *5-HTTLPR*-rs25531 haplotype significantly moderates the association between acute stress and IMDs. These results revealed the *S-A-S-A* genotype as a vulnerability factor for IMDs. HIV+ children who experienced moderate and severe acute stress and possessed the *S-A-S-A* haplotype were more likely to have an IMD, compared to HIV+ children who experienced mild acute stress and possessed the *L-A-L-A* haplotype.

Previous studies have reported the *SS* genotype to be a risk for different IMD psychopathologies including depression (Zalsman et al., 2006; Cervilla et al., 2007; Karg et al., 2011; Sharpley et al., 2014; Haberstick et al., 2016), anxiety (Armbruster et al., 2009), and IMDs in general (Conway et al., 2014). However, we found no association between the *SS* genotype and IMDs in the present study. This observation could be due to insufficient power to detect the effect of *5-HTTLPR* on the association between acute stress and IMDs among our study participants. Our study sample size had a *post hoc* power of greater than 80% (87.5% for the interaction analyses) under assumption that there was zero correlation of exposure between cases and controls and that the probability of exposure among controls was 0.5. There is, however, a possibility that the probability of exposure could have been less than 0.5 among controls, thus reducing the power in the present study. The association between the *SS* genotype and IMDs has been suggested to be due to its lower transcriptional activity (Baca-García et al., 2002; Hu et al., 2006). Future functional studies that examine the effects of the *S-A-S-A* compared to the *L-A-L-A* haplotype are required for us to determine whether lower 5-HTT levels could be driving our results.

A significant association was observed between *TPH2* rs1843809 and IMDs. The functional role of *TPH2* rs1843809

is currently not known. Being an intronic variant, it is less likely that rs1843809 would have an effect on gene expression or protein structure, as introns are spliced during mRNA processing, although potential effects of intronic variants have been suggested (Kleinjan and van Heyningen, 2005). It is also possible that rs1843809 is in LD with a causal variant. For example, rs1843809 is in LD with a missense rs142055199 SNP among the Luhya, a population that speaks the same Niger-Congo language as the population of the present study (Baganda; Countries and their Cultures, 2021). East African populations that speak this class of language have been found to be genetically similar by principal components analysis on GWAS data from NeuroGAP pilot study (Unpublished results). The SNP rs142055199 is located in the zinc finger C3H1-type containing gene, which modulates interleukin-8 (IL-8) transcription (National Center for Biotechnology Information, 2021c). Modulation of IL-8 would be of interest since IMDs have been associated with inflammatory processes (Miller et al., 2009; Musinguzi et al., 2018; Hori and Kim, 2019; Wang et al., 2019). Further studies will be needed to determine the mechanisms underlying the influence of this SNP in IMDs.

None of the investigated polymorphisms in *TPH2* moderated the association between acute stress and IMDs. This suggests that although TPH2 is critical in 5-HT biosynthesis (Walther et al., 2003; Carkaci-Salli et al., 2006), genetic variants investigated in this study neither moderate an individual's response to stress nor are they causal polymorphisms that moderate an individual's response to stress.

Limitations and Recommendations

The following limitations should be noted. We defined IMDs as having any depressive disorder, anxiety disorder, or PTSD. As much as anxiety disorders are commonly comorbid with depression, the inclusion of PTSD (n = 60) is contentious as it has been excluded from the anxiety disorder category in the Diagnostic and Statistical Manual of Mental Disorders – 5th edition (DSM-5) and is not widely accepted as an IMD (American Psychiatric Association, 2013). The inclusion of PTSD samples allowed us to achieve the desired statistical power. However, the low correlation between genetic risk factors for PTSD and those of depression and anxiety would have reduced the power to detect true significant associations and interactions. Future studies should endeavor to stratify analyses by specific IMDs although this will require larger samples. In addition, both acute and chronic stress indices were measured using a number of context-specific indicators. This approach was adopted because there is no locally adapted tool for assessing these indicators in this setting and the variables used to generate these indices are known stressors in this population. However, this tool remains to be validated. Moreover, we did not control for population stratification at analysis. The study participants belong to the Bagandan population group, for which principal components analysis on a GWAS sample of over 4,700 individuals has shown less genetic variation among the Bagandan (unpublished data). There is, however, a possibility that some participants were not Bagandan, even though they could speak Luganda (the main language of Baganda), indicating that we may have had an admixed sampled. Future studies should, therefore, control for population stratification, in order to circumvent confounding of results due to possible ancestry associated differences in genetic architecture. Also, due to the exploratory nature of the study, multiple testing corrections were not done. Correcting for multiple testing would have rendered the interaction analysis non-significant. We, therefore, acknowledge the risk of false positive findings in this study which have potentially important clinical implications and need to be verified in future research.

CONCLUSION

Acute stress was found to be associated with an increased risk of IMDs, and this association was found to be moderated by the 5-*HTTLPR*-rs25531 *S-A-S-A* haplotype and chronic stress among Ugandan HIV+ children and adolescents. These results support the diathesis-stress model, though the mechanisms through which acute stress interacts with *5-HTTLPR*-rs25531 *S-A-S-A* haplotype to moderate the risk of IMDs need to be elucidated. This will allow for interventions to be targeted to at-risk individuals, an important consideration in resource constrained settings.

DATA AVAILABILITY STATEMENT

All information gathered about study subjects and their samples is confidential, with access limited to the research team. However, upon request, data from the MRC/UVRI and LSHTM Uganda Research Unit is currently accessed under a data sharing policy *via*: http://www.mrcuganda.org/sites/default/files/publications/MRC_UVRI_Data_sharing_policy_December2015.pdf.

ETHICS STATEMENT

This study was reviewed and approved by the Health Research Committee of Stellenbosch University (# S17/09/179) and the higher Degrees Research and Ethics Committee, School of Biomedical Sciences, College of Health Sciences, Makerere University (# SBS 421). The parent study (CHAKA) obtained ethics approval from the Uganda Virus Research Institute (UVRI) Science and Ethical Committee (# GC/127/15/06/459) and the Uganda National Council of Science and Technology (# HS 1601). All parents/caregivers provided written informed consent for their children/adolescents to participate in the study and for a blood specimen to be withdrawn from them for the genetics analyses. Adolescents further provided written informed assent to participate in the study. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AK, SH, EK, and SS: concept. AK, EK, SH, SS, and JW: data collection. WS, AK, RN, SH, JW, SS, and EK: data analysis. AK, SH, SS, EK, JW, MJ, WS, and RN: first draft and final revision. All authors contributed to the article and approved the submitted version.

FUNDING

The study was funded by the Alliance for Global Health and Science of the Center for Emerging and Neglected Diseases, grant number: 50288/N7145, the Medical Research Council/ Department for International Development – African Leadership Award (MRC/DfID/ARL) to EK, grant number: MR/L004623/1, the Africa Center of Excellence in Materials, Product Development and Nanotechnology (MAPRONANO ACE), Project ID/Number: P151847, IDA Number: 5797-UG, the South African Research Chairs Initiative in Post-traumatic Stress Disorder, funded by the Department of Science and Technology and the National Research Foundation of South Africa, and the South African Medical Research Council. AK was also supported by both the South African Research Chairs Initiative in Post-traumatic Stress Disorder and the mental health section of MRC/UVRI and LSHTM Uganda Research unit.

ACKNOWLEDGMENTS

The assistance of the following are gratefully acknowledged: study subjects, research assistants of the mental health section of MRC/UVRI and LSHTM Uganda Research Unit, Joint Clinical Research Centre, Nsambya Home Care, TASO – Masaka, Kitovu Mobile Clinic, Uganda Cares – Masaka, members of the Neuropsychiatric Genetics Laboratory at Stellenbosch University, Data and Statistics Section of the MRC/UVRI and LSHTM Uganda Research Unit, and the National Research Foundation of South Africa.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.649055/full#supplementary-material

REFERENCES

Adelman, R. D., Tmanova, L. L., Delgado, D., Dion, S., and Lachs, M. S. (2014). Caregiver burden: a clinical review. *JAMA* 311, 1052–1060. doi: 10.1001/jama.2014.304

Ali, F. R., Vasiliou, S. A., Haddley, K., Paredes, U. M., Roberts, J. C., Miyajima, F., et al. (2010). Combinatorial interaction between two human serotonin transporter gene variable number tandem repeats and their regulation by CTCF. *J. Neurochem.* 112, 296–306. doi: 10.1111/j.1471-4159.2009.06453.x

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*. 5th Edn.

Anttila, S., Viikki, M., Huuhka, K., Huuhka, M., Huhtala, H., Rontu, R., et al. (2009). TPH2 polymorphisms may modify clinical picture in treatment-resistant depression. *Neurosci. Lett.* 464, 43–46. doi: 10.1016/j.neulet.2009.08.018

Armbruster, D., Moser, D. A., Strobel, A., Hensch, T., Kirschbaum, C., Lesch, K. P., et al. (2009). Serotonin transporter gene variation and stressful life events impact processing of fear and anxiety. *Int. J. Neuropsychopharmacol.* 12, 393–401. doi: 10.1017/S1461145708009565

Artigas, F. (2013a). Serotonin receptors involved in antidepressant effects. *Pharmacol. Ther.* 137, 119–131. doi: 10.1016/j.pharmthera.2012.09.006

Artigas, F. (2013b). Future directions for serotonin and antidepressants. *ACS Chem. Neurosci.* 4, 5–8. doi: 10.1021/cn3001125

Assary, E., Vincent, J. P., Keers, R., and Pluess, M. (2018). Gene-environment interaction and psychiatric disorders: review and future directions. *Semin. Cell Dev. Biol.* 77, 133–143. doi: 10.1016/j.semcdb.2017.10.016

Baca-García, E., Vaquero, C., Diaz-Sastre, C., Saiz-Ruiz, J., Fernández-Piqueras, J., and de Leon, J. (2002). A gender-specific association between the serotonin transporter gene and suicide attempts. *Neuropsychopharmacology* 26, 692–695. doi: 10.1016/S0893-133X(01)00394-3

Barrett, J. C. (2009). Haploview: visualization and analysis of SNP genotype data. *Cold Spring Harb. Protoc.* 2009:pdb.ip71. doi: 10.1101/pdb.ip71

Battersby, S., Ogilvie, A. D., Smith, C. A., Blackwood, D. H., Muir, W. J., Quinn, J. P., et al. (1996). Structure of a variable number tandem repeat of the serotonin transporter gene and association with affective disorder. *Psychiatr. Genet.* 6, 177–181. doi: 10.1097/00041444-199624000-00001

Betancourt, T., Scorza, P., Kanyanganzi, F., Fawzi, M. C., Sezibera, V., Cyamatare, F., et al. (2014). HIV and child mental health: a case-control study in Rwanda. *Pediatrics* 134, e464–e472. doi: 10.1542/peds.2013-2734

Beusenberg, M., and Orley, J. H. (1994). *A User's Guide to The Self Reporting Questionnaire (SRQ)*. Geneva, Switzerland: World Health Organization.

Bhatia, R., Hartman, C., Kallen, M. A., Graham, J., and Giordano, T. P. (2011). Persons newly diagnosed with HIV infection are at high risk for depression and poor linkage to care: results from the steps study. *AIDS Behav.* 15, 1161–1170. doi: 10.1007/s10461-010-9778-9

Boyes, M. E., and Cluver, L. D. (2015). Relationships between familial HIV/AIDS and symptoms of anxiety and depression: the mediating effect of bullying victimization in a prospective sample of south African children and adolescents. *J. Youth Adolesc.* 44, 847–859. doi: 10.1007/s10964-014-0146-3

Brown, R. C., Nugent, N. R., Hawn, S. E., Koenen, K. C., Miller, A., Amstadter, A. B., et al. (2016). Predicting the transition from acute stress disorder to posttraumatic stress disorder in children with severe injuries. *J. Pediatr. Health Care* 30, 558–568. doi: 10.1016/j.pedhc.2015.11.015

Carkaci-Salli, N., Flanagan, J. M., Martz, M. K., Salli, U., Walther, D. J., Bader, M., et al. (2006). Functional domains of human tryptophan hydroxylase 2 (hTPH2). *J. Biol. Chem.* 281, 28105–28112. doi: 10.1074/jbc.M602817200

Caspi, A., and Moffitt, T. E. (2006). Gene-environment interactions in psychiatry: joining forces with neuroscience. *Nat. Rev. Neurosci.* 7, 583–590. doi: 10.1038/nrn1925

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., et al. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* 301, 386–389. doi: 10.1126/science.1083968

Cervilla, J. A., Molina, E., Rivera, M., Torres-González, F., Bellón, J. A., Moreno, B., et al. (2007). The risk for depression conferred by stressful life events is modified by variation at the serotonin transporter 5HTTLPR genotype: evidence from the Spanish PREDICT-gene cohort. *Mol. Psychiatry* 12, 748–755. doi: 10.1038/sj.mp.4001981

Charney, D. S., and Manji, H. K. (2004). Life stress, genes, and depression: multiple pathways lead to increased risk and new opportunities for intervention. *Sci. STKE* 2004, re5. doi: 10.1126/stke.2252004re5

Chida, Y., and Vedhara, K. (2009). Adverse psychosocial factors predict poorer prognosis in HIV disease: a meta-analytic review of prospective investigations. *Brain Behav. Immun.* 23, 434–445. doi: 10.1016/j.bbi.2009.01.013

Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., et al. (2018). Comparative efficacy and acceptability of 21

antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* 391, 1357–1366. doi: 10.1016/S0140-6736(17)32802-7

Clarke, H., Flint, J., Attwood, A. S., and Munafò, M. R. (2010). Association of the 5- HTTLPR genotype and unipolar depression: a meta-analysis. *Psychol Med.* 40, 1767–1778. doi: 10.1017/S0033291710000516

Clarke, L., Fairley, S., Zheng-Bradley, X., Streeter, I., Perry, E., Lowy, E., et al. (2017). The international genome sample resource (IGSR): a worldwide collection of genome variation incorporating the 1000 genomes project data. *Nucleic Acids Res.* 45, D854–D859. doi: 10.1093/nar/gkw829

Conway, C. C., Slavich, G. M., and Hammen, C. (2014). Daily stress reactivity and serotonin transporter gene (5-HTTLPR) variation: internalizing responses to everyday stress as a possible transdiagnostic phenotype. *Biol. Mood Anxiety Disord.* 4:2. doi: 10.1186/2045-5380-4-2

Countries and their Cultures (2021). Baganda. [Place unknown]. [Publisher unknown]. [Date unknown, cited 2021 Jan 02]. Available at: https://www.everyculture.com/wc/Tajikistan-to-Zimbabwe/Baganda.html (Accessed January 02, 2020).

Culverhouse, R. C., Saccone, N. L., Horton, A. C., Ma, Y., Anstey, K. J., Banaschewski, T., et al. (2018). Collaborative meta-analysis finds no evidence of a strong interaction between stress and 5-HTTLPR genotype contributing to the development of depression. *Mol. Psychiatry* 23, 133–142. doi: 10.1038/mp.2017.44

Davies, M. N., Verdi, S., Burri, A., Trzaskowski, M., Lee, M., Hettema, J. M., et al. (2015). Generalised anxiety disorder--a twin study of genetic architecture, genome-wide association and differential gene expression. *PLoS One* 10:e0134865. doi: 10.1371/journal.pone.0134865

de Kloet, E. R., Joëls, M., and Holsboer, F. (2005). Stress and the brain: from adaptation to disease. *Nat. Rev. Neurosci.* 6, 463–475. doi: 10.1038/nrn1683

Ehli, E. A., Hu, Y., Lengyel-Nelson, T., Hudziak, J. J., and Davies, G. E. (2012). Identification and functional characterization of three novel alleles for the serotonin transporter-linked polymorphic region. *Mol. Psychiatry* 17, 185–192. doi: 10.1038/mp.2010.130

Evans, G. W., Li, D., and Whipple, S. S. (2013). Cumulative risk and child development. *Psychol. Bull.* 139, 1342–1396. doi: 10.1037/a0031808

Fuchs, E., and Flïugge, G. (2006). Experimental animal models for the simulation of depression and anxiety. *Dialogues Clin. Neurosci.* 8, 323–333.

Fullerton, C. S., Ursano, R. J., and Wang, L. (2004). Acute stress disorder, posttraumatic stress disorder, and depression in disaster or rescue workers. *Am. J. Psychiatry* 161, 1370–1376. doi: 10.1176/appi.ajp.161.8.1370

Furlong, R. A., Ho, L., Walsh, C., Rubinsztein, J. S., Jain, S., Paykel, E. S., et al. (1998). Analysis and meta-analysis of two serotonin transporter gene polymorphisms in bipolar and unipolar affective disorders. *Am. J. Med. Genet.* 81, 58–63. doi: 10.1002/(SICI)1096-8628(19980207)81:1<58::AID-AJMG11>3.0.CO;2-V

Gadow, K. D., and Sprafkin, J. (1999). Youth's Inventory-4. Stony Brook, New York: Checkmate plus; [Revised 2016 Jan 23]. Available at: https://www.checkmateplus.com/product/yi-4.htm (Accessed April 07, 2021).

Gadow, K. D., and Sprafkin, J. (2013). Child and Adolescent Symptom Inventory-5. Stony Brook, New York: Checkmate plus; [Revised 2016 Jan 23]. Available at: https://www.checkmateplus.com/product/casi5.htm (Accessed April 07, 2021).

Gao, J., Pan, Z., Jiao, Z., Li, F., Zhao, G., Wei, Q., et al. (2012). TPH2 gene polymorphisms and major depression--a meta-analysis. *PLoS One* 7:e36721. doi: 10.1371/journal.pone.0036721

Gressier, F., Calati, R., Balestri, M., Marsano, A., Alberti, S., Antypa, N., et al. (2013). The 5-HTTLPR polymorphism and posttraumatic stress disorder: a meta-analysis. *J. Trauma. Stress.* 26, 645–653. doi: 10.1002/jts.21855

Grillon, C., Duncko, R., Covington, M. F., Kopperman, L., and Kling, M. A. (2007). Acute stress potentiates anxiety in humans. *Biol. Psychiatry* 62, 1183–1186. doi: 10.1016/j.biopsych.2007.06.007

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., et al. (2015). The African genome variation project shapes medical genetics in Africa. *Nature* 517, 327–332. doi: 10.1038/nature13997

Haberstick, B. C., Boardman, J. D., Wagner, B., Smolen, A., Hewitt, J. K., Killeya-Jones, L. A., et al. (2016). Depression, stressful life events, and the impact of variation in the serotonin transporter: findings from the National Longitudinal Study of adolescent to adult health (add health). *PLoS One* 11:e0148373. doi: 10.1371/journal.pone.0148373

Haddley, K., Bubb, V. J., Breen, G., Parades-Esquivel, U. M., and Quinn, J. P. (2012). Behavioural genetics of the serotonin transporter. *Curr. Top. Behav. Neurosci.* 12, 503–535. doi: 10.1007/7854_2011_186

Hammoud, D. A., Endres, C. J., Hammond, E., Uzuner, O., Brown, A., Nath, A., et al. (2010). Imaging serotonergic transmission with [11C]DASB-PET in depressed and non-depressed patients infected with HIV. *NeuroImage* 49, 2588–2595. doi: 10.1016/j.neuroimage.2009.10.037

Heils, A., Teufel, A., Petri, S., Stöber, G., Riederer, P., Bengel, D., et al. (1996). Allelic variation of human serotonin transporter gene expression. *J. Neurochem.* 66, 2621–2624. doi: 10.1046/j.1471-4159.1996.66062621.x

Hettema, J. M., Neale, M. C., and Kendler, K. S. (2001). A review and meta-analysis of the genetic epidemiology of anxiety disorders. *Am. J. Psychiatry* 158, 1568–1578. doi: 10.1176/appi.ajp.158.10.1568

Hori, H., and Kim, Y. (2019). Inflammation and post-traumatic stress disorder. *Psychiatry Clin. Neurosci.* 73, 143–153. doi: 10.1111/pcn.12820

Hu, X. Z., Lipsky, R. H., Zhu, G., Akhtar, L. A., Taubman, J., Greenberg, B. D., et al. (2006). Serotonin transporter promoter gain-of-function genotypes are linked to obsessive-compulsive disorder. *Am. J. Hum. Genet.* 78, 815–826. doi: 10.1086/503850

Ironson, G., O'Cleirigh, C., Fletcher, M. A., Laurenceau, J. P., Balbin, E., Klimas, N., et al. (2005). Psychosocial factors predict CD4 and viral load change in men and women with human immunodeficiency virus in the era of highly active antiretroviral treatment. *Psychosom. Med.* 67, 1013–1021. doi: 10.1097/01.psy.0000188569.58998.c8

Jacob, J., Ferri, A. L., Milton, C., Prin, F., Pla, P., Lin, W., et al. (2007). Transcriptional repression coordinates the temporal switch from motor to serotonergic neurogenesis. *Nat. Neurosci.* 10, 1433–1439. doi: 10.1038/nn1985

Jawahar, M. C., Toben, C. G., and Baune, B. T. (2019). "Gene-environment interactions and epigenetic mechanisms in depression" in *Neurobiology of Depression.* eds. J. Quevedo, A. F. Carvalho and C. A. Zarate (Academic Press), 17–25.

Kalungi, A., Kinyanda, E., Womersley, J. S., Joloba, M. L., Ssembajjwe, W., Nsubuga, R. N., et al. (2021). TERT rs2736100 and TERC rs16847897 genotypes moderate the association between internalizing mental disorders and accelerated telomere length attrition among HIV+ children and adolescents in Uganda. *BMC Med. Genet.* 14:15. doi: 10.1186/s12920-020-00857-z

Kalungi, A., Seedat, S., Hemmings, S. M. J., van der Merwe, L., Joloba, M. L., Nanteza, A., et al. (2017). Association between serotonin transporter gene polymorphisms and increased suicidal risk among HIV positive patients in Uganda. *BMC Genet.* 18:71. doi: 10.1186/s12863-017-0538-y

Kalungi, A., Womersley, J. S., Kinyanda, E., Joloba, M. L., Ssembajjwe, W., Nsubuga, R. N., et al. (2019). Internalizing mental disorders and accelerated cellular aging among perinatally HIV-infected youth in Uganda. *Front. Genet.* 10:705. doi: 10.3389/fgene.2019.00705

Karg, K., Burmeister, M., Shedden, K., and Sen, S. (2011). The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: evidence of genetic moderation. *Arch. Gen. Psychiatry* 68, 444–454. doi: 10.1001/archgenpsychiatry.2010.189

Kennedy, A. P., Binder, E. B., Bowman, D., Harenski, K., Ely, T., Cisler, J. M., et al. (2012). A common TPH2 haplotype regulates the neural processing of a cognitive control demand. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 159B, 829–840. doi: 10.1002/ajmg.b.32090

Kinyanda, E., Levin, J., Nakasujja, N., Birabwa, H., Nakku, J., Mpango, R., et al. (2018). Major depressive disorder: longitudinal analysis of impact on clinical and behavioral outcomes in Uganda. *J. Acquir. Immune Defic. Syndr.* 78, 136–143. doi: 10.1097/QAI.0000000000001647

Kinyanda, E., Salisbury, T. T., Levin, J., Nakasujja, N., Mpango, R. S., Abbo, C., et al. (2019). Rates, types and co-occurrence of emotional and behavioural disorders among perinatally HIV-infected youth in Uganda: the CHAKA study. *Soc. Psychiatry Psychiatr. Epidemiol.* 54, 415–425. doi: 10.1007/s00127-019-01675-0

Kleinjan, D. A., and van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* 76, 8–32. doi: 10.1086/426833

Klumpers, F., Kroes, M. C., Heitland, I., Everaerd, D., Akkermans, S. E., Oosting, R. S., et al. (2015). Dorsomedial prefrontal cortex mediates the impact of serotonin transporter linked polymorphic region genotype on anticipatory threat reactions. *Biol. Psychiatry* 78, 582–589. doi: 10.1016/j.biopsych.2014.07.034

Klumpp, H., Fitzgerald, D. A., Cook, E., Shankman, S. A., Angstadt, M., and Phan, K. L. (2014). Serotonin transporter gene alters insula activity to threat

in social anxiety disorder. *Neuroreport* 25, 926–931. doi: 10.1097/WNR.0000000000000210

Kristensen, A. S., Andersen, J., Jørgensen, T. N., Sørensen, L., Eriksen, J., Loland, C. J., et al. (2011). SLC6 neurotransmitter transporters: structure, function, and regulation. *Pharmacol. Rev.* 63, 585–640. doi: 10.1124/pr.108.000869

Lachin, J. M. (1992). Power and sample size evaluation for the McNemar test with application to matched case-control studies. *Stat. Med.* 11, 1239–1251. doi: 10.1002/sim.4780110909

Lesch, K. P., Bengel, D., Heils, A., Sabol, S. Z., Greenberg, B. D., Petri, S., et al. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science* 274, 1527–1531. doi: 10.1126/science.274.5292.1527

Lipsky, R. H., Hu, X. Z., and Goldman, D. (2009). Additional functional variation at the SLC6A4 gene. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 150B:153. doi: 10.1002/ajmg.b.30766

Lopez de Lara, C., Brezo, J., Rouleau, G., Lesage, A., Dumont, M., Alda, M., et al. (2007). Effect of tryptophan hydroxylase-2 gene variants on suicide risk in major depression. *Biol. Psychiatry* 62, 72–80. doi: 10.1016/j.biopsych.2006.09.008

Lwidiko, A., Kibusi, S. M., Nyundo, A., and Mpondo, B. C. T. (2018). Association between HIV status and depressive symptoms among children and adolescents in the southern highlands zone, Tanzania: a case-control study. *PLoS One* 13:e0193145. doi: 10.1371/journal.pone.0193145

Martin, J., Cleak, J., Willis-Owen, S. A., Flint, J., and Shifman, S. (2007). Mapping regulatory variants for the serotonin transporter gene based on allelic expression imbalance. *Mol. Psychiatry* 12, 421–422. doi: 10.1038/sj.mp.4001952

Mellins, C. A., Elkington, K. S., Leu, C. S., Santamaria, E. K., Dolezal, C., Wiznia, A., et al. (2012). Prevalence and change in psychiatric disorders among perinatally HIV-infected and HIV-exposed youth. *AIDS Care* 24, 953–962. doi: 10.1080/09540121.2012.668174

Miller, A. H., Maletic, V., and Raison, C. L. (2009). Inflammation and its discontents: the role of cytokines in the pathophysiology of major depression. *Biol. Psychiatry* 65, 732–741. doi: 10.1016/j.biopsych.2008.11.029

Monroe, S. M., and Simons, A. D. (1991). Diathesis-stress theories in the context of life stress research: implications for the depressive disorders. *Psychol. Bull.* 110, 406–425. doi: 10.1037/0033-2909.110.3.406

Mpango, R. S., Kinyanda, E., Rukundo, G. Z., Gadow, K. D., and Patel, V. (2017). Cross-cultural adaptation of the child and adolescent symptom Inventory-5 (CASI-5) for use in central and South-Western Uganda: the CHAKA project. *Trop. Dr.* 47, 347–354. doi: 10.1177/0049475517724688

Munafò, M. R., Durrant, C., Lewis, G., and Flint, J. (2009a). Gene X environment interactions at the serotonin transporter locus. *Biol. Psychiatry* 65, 211–219. doi: 10.1016/j.biopsych.2008.06.009

Munafò, M. R., Freimer, N. B., Ng, W., Ophoff, R., Veijola, J., Miettunen, J., et al. (2009b). 5-HTTLPR genotype and anxiety-related personality traits: a meta-analysis and new data. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 150B, 271–281. doi: 10.1002/ajmg.b.30808

Musci, R. J., Augustinavicius, J. L., and Volk, H. (2019). Gene-environment interactions in psychiatry: recent evidence and clinical implications. *Curr. Psychiatry Rep.* 21:81. doi: 10.1007/s11920-019-1065-5

Musinguzi, K., Obuku, A., Nakasujja, N., Birabwa, H., Nakku, J., Levin, J., et al. (2018). Association between major depressive disorder and pro-inflammatory cytokines and acute phase proteins among HIV-1 positive patients in Uganda. *BMC Immunol.* 19:1. doi: 10.1186/s12865-017-0239-3

Nachman, S., Chernoff, M., Williams, P., Hodge, J., Heston, J., and Gadow, K. D. (2012). Human immunodeficiency virus disease severity, psychiatric symptoms, and functional outcomes in perinatally infected youth. *Arch. Pediatr. Adolesc. Med.* 166, 528–535. doi: 10.1001/archpediatrics.2011.1785

National Center for Biotechnology Information (2021a). SLC6A4 solute carrier family 6 member 4 [*Homo sapiens* (human)]. Updated 2020 Dec 27. Available at: https://www.ncbi.nlm.nih.gov/gene/6532#location (Accessed December 29, 2020).

National Center for Biotechnology Information (2021b). STIN2-VNTR serotonin transporter intronic VNTR enhancer [*Homo sapiens* (human)]. Updated 2020 Nov 24. Available at: https://www.ncbi.nlm.nih.gov/gene/?term=serotonin+transporter+intronic+VNTR+enhancer+%5BHomo+sapiens+(human)%5D (Accessed December 29, 2020).

National Center for Biotechnology Information (2021c). ZFC3H1 zinc finger C3H1-type containing [*Homo sapiens* (human)]. Updated 2020 Dec 06. Available at: https://www.ncbi.nlm.nih.gov/gene/196441 (Accessed December 29, 2020).

O'Connor, R. C., Rasmussen, S., and Hawton, K. (2010). Predicting depression, anxiety and self-harm in adolescents: the role of perfectionism and acute life stress. *Behav. Res. Ther.* 48, 52–59. doi: 10.1016/j.brat.2009.09.008

O'Gara, C., Knight, J., Stapleton, J., Luty, J., Neale, B., Nash, M., et al. (2008). Association of the serotonin transporter gene, neuroticism and smoking behaviours. *J. Hum. Genet.* 53, 239–246. doi: 10.1007/s10038-007-0243-1

Olivier, B. (2015). Serotonin: a never-ending story. *Eur. J. Pharmacol.* 753, 2–18. doi: 10.1016/j.ejphar.2014.10.031

Otte, C., Gold, S. M., Penninx, B. W., Pariante, C. M., Etkin, A., Fava, M., et al. (2016). Major depressive disorder. *Nat. Rev. Dis. Primers.* 2:16065. doi: 10.1038/nrdp.2016.65

Pergamin-Hight, L., Bakermans-Kranenburg, M. J., van Ijzendoorn, M. H., and Bar-Haim, Y. (2012). Variations in the promoter region of the serotonin transporter gene and biased attention for emotional information: a meta-analysis. *Biol. Psychiatry* 71, 373–379. doi: 10.1016/j.biopsych.2011.10.030

Philibert, R. A., Sandhu, H., Hollenbeck, N., Gunter, T., Adams, W., and Madan, A. (2008). The relationship of 5HTT (SLC6A4) methylation and genotype on mRNA expression and liability to major depression and alcohol dependence in subjects from the Iowa adoption studies. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 147B, 543–549. doi: 10.1002/ajmg.b.30657

Plomin, R., and Davis, O. S. (2009). The future of genetics in psychology and psychiatry: microarrays, genome-wide association, and non-coding RNA. *J. Child Psychol. Psychiatry* 50, 63–71. doi: 10.1111/j.1469-7610.2008.01978.x

Popoli, M., Yan, Z., McEwen, B. S., and Sanacora, G. (2011). The stressed synapse: the impact of stress and glucocorticoids on glutamate transmission. *Nat. Rev. Neurosci.* 13, 22–37. doi: 10.1038/nrn3138

Pristerà, A., Lin, W., Kaufmann, A. K., Brimblecombe, K. R., Threlfell, S., Dodson, P. D., et al. (2015). Transcription factors FOXA1 and FOXA2 maintain dopaminergic neuronal properties and control feeding behavior in adult mice. *Proc. Natl. Acad. Sci. U. S. A.* 112, E4929–E4938. doi: 10.1073/pnas.1503911112

Revenson, T., Griva, K., Luszczynska, A., Morrison, V., Panagopoulou, E., Vilchinsky, N., et al. (2016). *Caregiving in the Illness Context*. London: Palgrave Macmillan.

Robles, T. F., Slatcher, R. B., Trombello, J. M., and McGinn, M. M. (2014). Marital quality and health: a meta-analytic review. *Psychol. Bull.* 140, 140–187. doi: 10.1037/a0031859

Rudnick, G. (2006). Serotonin transporters--structure and function. *J. Membr. Biol.* 213, 101–110. doi: 10.1007/s00232-006-0878-4

Sharpley, C. F., Palanisamy, S. K., Glyde, N. S., Dillingham, P. W., and Agnew, L. L. (2014). An update on the interaction between the serotonin transporter promoter variant (5-HTTLPR), stress and depression, plus an exploration of non-confirming findings. *Behav. Brain Res.* 273, 89–105. doi: 10.1016/j.bbr.2014.07.030

Shishkina, G. T., Kalinina, T. S., and Dygalo, N. N. (2007). Up-regulation of tryptophan hydroxylase-2 mRNA in the rat brain by chronic fluoxetine treatment correlates with its antidepressant effect. *Neuroscience* 150, 404–412. doi: 10.1016/j.neuroscience.2007.09.017

Silberg, J. L., Rutter, M., and Eaves, L. (2001). Genetic and environmental influences on the temporal association between earlier anxiety and later depression in girls. *Biol. Psychiatry* 49, 1040–1049. doi: 10.1016/S0006-3223(01)01161-1

Smoller, J. W. (2016). The genetics of stress-related disorders: PTSD, depression, and anxiety disorders. *Neuropsychopharmacology* 41, 297–319. doi: 10.1038/npp.2015.266

Springer, S. A., Dushaj, A., and Azar, M. M. (2012). The impact of DSM-IV mental disorders on adherence to combination antiretroviral therapy among adult persons living with HIV/AIDS: a systematic review. *AIDS Behav.* 16, 2119–2143. doi: 10.1007/s10461-012-0212-3

Tandon, M., Si, X., and Luby, J. (2011). Preschool onset attention-deficit/hyperactivity disorder: course and predictors of stability over 24 months. *J. Child Adolesc. Psychopharmacol.* 21, 321–330. doi: 10.1089/cap.2010.0045

van den Buuse, M., and Hale, M. W. (2019). "Serotonin in stress" in *Stress: Physiology, Biochemistry, and Pathology.* ed. G. Fink (Academic Press), 115–123.

Walther, D. J., Peter, J. U., Bashammakh, S., Hörtnagl, H., Voits, M., Fink, H., et al. (2003). Synthesis of serotonin by a second tryptophan hydroxylase isoform. *Science* 299:76. doi: 10.1126/science.1078197

Wang, Y., Sun, N., Liu, Z., Li, X., Yang, C., and Zhang, K. (2016). Psychosocial mechanisms of serotonin transporter's genetic polymorphism in susceptibility

to major depressive disorder: mediated by trait coping styles and interacted with life events. *Am. J. Transl. Res.* 8, 1281–1292.

Wang, W., Wang, L., Xu, H., Cao, C., Liu, P., Luo, S., et al. (2019). Characteristics of pro- and anti-inflammatory cytokines alteration in PTSD patients exposed to a deadly earthquake. *J. Affect. Disord.* 248, 52–58. doi: 10.1016/j.jad.2019.01.029

Ward, J. J. H. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58, 236–244.

Widaman, K. F., Helm, J. L., Castro-Schilo, L., Pluess, M., Stallings, M. C., and Belsky, J. (2012). Distinguishing ordinal and disordinal interactions. *Psychol. Methods* 17, 615–622. doi: 10.1037/a0030003

Willis, A. M., and Brock, M. S. (2018) Adverse Childhood Experience and Serotonin Transporters: A Gene Environmental Study of the Risk of PTSD in Soldiers (ACES). 59 MDW San Antonio United States.

Xiao, Y., Liu, D., Liu, K., Wu, C., Zhang, H., Niu, Y., et al. (2019). Association of DRD2, 5-HTTLPR, and 5-HTTVNTR gene polymorphisms with posttraumatic stress disorder in Tibetan adolescents: a case-control study. *Biol. Res. Nurs.* 21, 286–295. doi: 10.1177/1099800419838325

Xie, P., Kranzler, H. R., Poling, J., Stein, M. B., Anton, R. F., Brady, K., et al. (2009). Interactive effect of stressful life events and the serotonin transporter 5-HTTLPR genotype on posttraumatic stress disorder diagnosis in 2 independent populations. *Arch. Gen. Psychiatry* 66, 1201–1209. doi: 10.1001/archgenpsychiatry.2009.153

Zalsman, G., Huang, Y. Y., Oquendo, M. A., Burke, A. K., Hu, X. Z., Brent, D. A., et al. (2006). Association of a triallelic serotonin transporter gene promoter region (5-HTTLPR) polymorphism with stressful life events and severity of depression. *Am. J. Psychiatry* 163, 1588–1593. doi: 10.1176/ajp.2006.163.9.1588

Zhao, J., Miao, K., Wang, H., Ding, H., and Wang, D. W. (2013). Association between telomere length and type 2 diabetes mellitus: a meta-analysis. *PLoS One* 8:e79993. doi: 10.1371/journal.pone.0079993

Zill, P., Baghai, T. C., Zwanzger, P., Schüle, C., Eser, D., Rupprecht, R., et al. (2004). SNP and haplotype analysis of a novel tryptophan hydroxylase isoform (TPH2) gene provide evidence for association with major depression. *Mol. Psychiatry* 9, 1030–1036. doi: 10.1038/sj.mp.4001525

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

The handling editor declared a past co-authorship with the author RN.

frontiers in Genetics

ORIGINAL RESEARCH
published: 16 June 2021
doi: 10.3389/fgene.2021.683880

Admixture Has Shaped Romani Genetic Diversity in Clinically Relevant Variants

*Neus Font-Porterias[1], Aaron Giménez[2], Annabel Carballo-Mesa[3], Francesc Calafell[1] and David Comas[1]**

[1] Departament de Ciències Experimentals i de la Salut, Institut de Biologia Evolutiva (UPF-CSIC), Universitat Pompeu Fabra, Barcelona, Spain, [2] Facultat de Sociologia, Universitat Autònoma de Barcelona, Barcelona, Spain, [3] Facultat de Geografia i Història, Universitat de Barcelona, Barcelona, Spain

OPEN ACCESS

Edited by:
Karoline Kuchenbaecker, University College London, United Kingdom

Reviewed by:
Elizabeth Atkinson, Massachusetts General Hospital and Harvard Medical School, United States
Olivia J. Veatch, University of Kansas Medical Center, United States

***Correspondence:**
David Comas
david.comas@upf.edu

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *22 March 2021*
Accepted: *13 May 2021*
Published: *16 June 2021*

Citation:
Font-Porterias N, Giménez A, Carballo-Mesa A, Calafell F and Comas D (2021) Admixture Has Shaped Romani Genetic Diversity in Clinically Relevant Variants. Front. Genet. 12:683880. doi: 10.3389/fgene.2021.683880

Genetic patterns of inter-population variation are a result of different demographic and adaptive histories, which gradually shape the frequency distribution of the variants. However, the study of clinically relevant mutations has a Eurocentric bias. The Romani, the largest transnational minority ethnic group in Europe, originated in South Asia and received extensive gene flow from West Eurasia. Most medical genetic studies have only explored founder mutations related to Mendelian disorders in this population. Here we analyze exome sequences and genome-wide array data of 89 healthy Spanish Roma individuals to study complex traits and disease. We apply a different framework and focus on variants with both increased and decreased allele frequencies, taking into account their local ancestry. We report several OMIM traits enriched for genes with deleterious variants showing increased frequencies in Roma or in non-Roma (e.g., obesity is enriched in Roma, with an associated variant linked to South Asian ancestry; while non-insulin dependent diabetes is enriched in non-Roma Europeans). In addition, previously reported pathogenic variants also show differences among populations, where some variants segregating at low frequency in non-Roma are virtually absent in the Roma. Lastly, we describe frequency changes in drug-response variation, where many of the variants increased in Roma are clinically associated with metabolic and cardiovascular-related drugs. These results suggest that clinically relevant variation in Roma cannot only be characterized in terms of founder mutations. Instead, we observe frequency differences compared to non-Roma: some variants are absent, while other have drifted to higher frequencies. As a result of the admixture events, these clinically damaging variants can be traced back to both European and South Asian-related ancestries. This can be attributed to a different prevalence of some genetic disorders or to the fact that genetic susceptibility variants are mostly studied in populations of European descent, and can differ in individuals with different ancestries.

Keywords: Romani, whole-exome sequences, clinically relevant variants, drug-response variants, local ancestry inference, Eurocentric bias

INTRODUCTION

Human genetic diversity is a continuum, which means that there are no fixed, immutable or discrete boundaries between populations. Linguistic, geographic and social factors can lead to different demographic histories and, in turn, to patterns of inter-population variability (Lao et al., 2008; Barbujani et al., 2013; Batai et al., 2020; Heyer and Reynaud-Paligot, 2020). It has been previously reported that this stratification has consequences in the genetics of complex traits and diseases [reviewed in Bentley et al. (2017); Sirugo et al. (2019)]. On the one hand, disease-associated variants present allele frequency differences across populations. For example, 70% of the cystic fibrosis cases in Europeans are due to ΔF508 mutation in *CFTR* gene, while the most common causal variant in South Africans with African ancestry is 3120 + 1G→A, and different mutations have different therapeutic targets (Padoa et al., 1999; Stewart and Pepper, 2017; Sirugo et al., 2019). On the other hand, genomic variation across populations is also observed for treatment response differences, especially in genes related to absorption, distribution, metabolism, and excretion (ADME) of drugs (Dopazo et al., 2016; Škarić-Jurić et al., 2018; Sirugo et al., 2019). For example, the metabolism of the anticoagulant warfarin can differ due to several genetic polymorphisms; however, their frequencies are different in European and African descent groups, which challenges the correct dosage prescription (Bress et al., 2012; Johnson et al., 2017; Sirugo et al., 2019).

Thus, an accurate clinical assessment relies on the study of clinically relevant genetic variants with different allele frequencies across groups. Yet, there is an underrepresentation of human populations in the screening of these variants. Particularly, genetic studies show a strong and systematic Eurocentric bias (Need and Goldstein, 2009; Popejoy and Fullerton, 2016; Bentley et al., 2017; Martin et al., 2019; Sirugo et al., 2019). As a consequence, this bias prevents to fully understand the genetic architecture of human disease and leads to an incomplete genetic assessment of complex traits, and to an inaccurate disease diagnosis and treatment in under-represented groups (Martin et al., 2019; Sirugo et al., 2019).

The Roma population, also known by the misnomer of "Gypsies," has been under-represented in these genome-wide scans. They constitute the largest transnational minority ethnic group in Europe (Council of Europe, 2012). Linguistic and genetic evidence point to a South Asian origin and subsequent diaspora toward Europe, with extensive non-Roma gene flow and multiple founder effects shaping their demographic history (Hancock, 2002; Matras, 2002; Mendizabal et al., 2012; Moorjani et al., 2013; Morar et al., 2013; Font-Porterias et al., 2019). Most medical genetic studies on this population have been focused on targeting the genetic variants responsible for the increased prevalence of certain genetic diseases [reviewed in Kalaydjieva et al. (2001); Morar et al. (2013)]. In this sense, several founder mutations have been identified: e.g., the p.R299X mutation in the *LTBP2* gene, which is responsible for congenital glaucoma (Azmanov et al., 2011). However, the distribution of disease-associated variants in this population has not been fully characterized. In addition, drug response-related genome-wide variation has only been deeply examined in Croatian Roma, where variants in ADME genes were found to have increased allele frequencies (Škarić-Jurić et al., 2018).

To fill this gap, we examine whole exome sequences (WES) and genome-wide array data of 89 healthy Spanish Roma individuals and characterize the functionally relevant genomic variants (i.e., associated to disease or to drug response) with either increased or decreased allele frequencies in the Roma. Beyond frequency distribution differences, and taking into account that Roma is an admixed population, we describe the ancestral origin of multiple variants by leveraging on the estimated local ancestry of their background haplotypes.

MATERIALS AND METHODS

Data

We used WES (mean depth of 54X), and genome-wide autosomal SNP data (Affymetrix Axiom Genome-Wide Human Origins 1 array) for Spanish Roma individuals (89 and 62 samples, respectively) (Font-Porterias et al., 2021), deposited at EGA (EGAS00001004599). The Spanish Roma WES were merged with previously published non-Roma WES from 1000G (mean depth of 65.7X): Iberian Population in Spain (IBS), Toscani in Italia (TSI), Punjabi from Lahore (PJL), Indian Telugu from the United Kingdom (ITU), and Gujarati Indian from Houston (GIH) (Auton et al., 2015), resulting in a dataset with 512 individuals and 410,225 variants. The genome-wide SNP data was merged with IBS, TSI, PJL, ITU, and GIH from 1000G (1000 Genomes Project Consortium, 2012), with 474,632 genome-wide SNPs in 487 samples. Both datasets were then combined to increase the covered genomic variants, building a merged WES-array dataset with 487 individuals and 878,162 SNPs. Variant annotation was performed using the Variant Effect Predictor tool (VEP) from Ensembl (McLaren et al., 2016) focusing on three deleterious prediction scores: PolyPhen-2 (Adzhubei et al., 2010), GERP (Davydov et al., 2010) and CADD (Rentzsch et al., 2019), as previously explained (Font-Porterias et al., 2021). For each analysis, the corresponding dataset used is specified (WES dataset or merged WES-array dataset). In addition, we have included a glossary of terms that may have ambiguous meanings in genetic studies (**Supplementary Note 1**).

Local Ancestry Inference

The phasing of the merged WES-array dataset, with 405,814 variants with minor allele frequency (MAF) > 1%, was performed using SHAPEIT (O'Connell et al., 2014), using the population-averaged genetic map from the HapMap phase II (International HapMap Consortium, 2003) and the 1000G dataset as a reference panel (1000 Genomes Project Consortium, 2012). RFMix v1.5.4 (Maples et al., 2013) was run with one expectation-maximization (EM) iteration to infer the local ancestry of the phased haplotypes, using balanced reference panels representing European (IBS and TSI populations) and

South Asian (PJL, GIH, and ITU populations) ancestries. As previously explained (Font-Porterias et al., 2021), the Roma individuals included in the present study show, on average, 68.4% and 31.6% of European and South Asian global ancestry proportions, with a standard deviation of 7%. Ancestry was assigned when RFMix posterior probability was higher than 0.9, resulting in 96.3% of the variants with assigned ancestry. In order to match the local ancestry inference in heterozygous variants and obtain the ancestry background of the allele, we adjusted the RFMix rephasing as previously performed (Browning et al., 2018), since RFMix partially rephases the data when assigning local ancestry. However, when the variant was filtered out (MAF $< 1\%$) in the phasing, only genotype ancestries can be retrieved.

Genetic Portability

We computed the allele sharing ratio (proportion of variants at different frequency bins) from the Roma segregating in non-Roma populations from WES variants dataset. In addition, we compared the linkage disequilibrium (LD) decay patterns between Roma and non-Roma from the genome-wide array dataset using PopLDdecay (Zhang et al., 2018) with default parameters. We performed a two-sample Kolmogorov–Smirnov test to check whether the decay distributions of Roma and non-Roma groups were statistically different. For both analyses, we used the same number of individuals per population to avoid sample size biases (70 individuals per population for the allele sharing ratio and 62 individuals per population for the LD decay).

Gene Enrichment Analyses

Using the WES dataset, we performed two different gene enrichment analyses using WEB-based GEne SeT AnaLysis Toolkit (Liao et al., 2019) to identify categories (or classes) of genes that are over-represented in a particular set of genes, using a background gene set. First, we interrogated those genes with more deleterious mutations (i.e., deleterious $N_{alleles}$ or N_{hom} per individual per gene). Although we expect that the most constrained genes (i.e., lower values of deleterious $N_{alleles}$ or N_{hom} per individual per gene) will be shared across populations, we test whether there is a particular pathway enriched in the most mutated genes in the Roma samples and the non-Roma groups, independently. To do so, we normalized $N_{alleles}$ or N_{hom} by the number of variants in each gene and we then examined the correlation between each pair of Roma to non-Roma populations per each gene. The over-representation analysis was performed with default parameters (Liao et al., 2019), Gene Ontology (GO) was selected as the functional database, and all the genes included in our variants set was used as background gene set. This analysis does not take into account the frequency of the variants, since the calculation is performed per individual. We also performed a gene enrichment analysis to test whether the genes with deleterious variants showing allele frequency increases in Roma to non-Roma (or non-Roma to Roma) belong to specific genetic disease clusters. We included in the analysis those genes with variants with a fold increase in minor allele frequency (MAF) > 5 or fold increase in minor allele count (MAC) > 5 for monomorphic variants. By using this restrictive threshold, we decrease the number of false positive results in the enrichment analyses, although we may lose some pathogenic variants found at low frequencies (and for that reason, we then focus on ClinVar pathogenic variants with less conservative thresholds). The same number of individuals were considered for this analysis (70 individuals per population) to avoid biases in MAC and MAF calculations. The over-representation analysis was performed with default parameters (Liao et al., 2019), OMIM was selected as the functional disease database and all the genes included in our variants set was used as background gene set. Once the enriched pathways were identified, we computed a chi-squared test to check whether the associated genetic variants described in OMIM present in our dataset have statistically different genotype frequencies between Roma and non-Roma groups.

Screening of Known Disease-Associated Variants

We first identified previously reported Mendelian mutations in the Roma, annotated in Bianco et al. (2020) and checked the ancestry of their haplotypes to trace their putative origin. However, this approach does not allow us to examine if there is a different frequency spectrum of disease-associated variants comparing Roma and non-Roma. To that end, we then annotated the set of WES variants using ClinVar database (Landrum et al., 2014) and compared the frequency of clinically validated variants among populations. We kept only variants with a clinical significance of "pathogenic," which is the highest level of supported evidence. We selected those variants with a fold increase in risk allele frequency (RAF) > 1.5 between populations or 1.5% RAF for monomorphic variants. A chi-squared test was performed to test whether the genotype frequencies were significantly different across populations. The same number of individuals were considered for this analysis (70 individuals per population) to avoid biases in RAF calculations.

Screening Beyond Disease-Associated Variants

To examine pharmacogenetic variation in the WES dataset, we studied mutations that disrupt drug binding domains without being deleterious to the protein. These variants are based on function prediction: they might cause drug binding inhibition; however, not all variants have a reported association with drug response (Hopkins and Groom, 2002; Dopazo et al., 2016). We also examined variants found in 31 core ADME genes[1] (Škarić-Jurić et al., 2018). We selected those variants with a fold increase in MAF > 1.5 between populations or 1.5% MAF for monomorphic variants. A chi-squared test was performed to test whether the genotype frequency was significantly different across populations. The same number of individuals were considered for this analysis (70 individuals per population) to avoid biases in MAF calculations. The selected variants in both analyses were searched in PharmGKB (Whirl-Carrillo et al., 2012) and the

[1] www.pharmaadme.org

corresponding target drugs in DrugBank (Wishart et al., 2018) and PubChem (Kim et al., 2019).

RESULTS

Initial Assessment of Functional Variants as a First Evidence of Inter-Population Variability

In order to assess the genetic portability from non-Roma to Roma, we examined the allele sharing and linkage disequilibrium patterns between populations. We have previously shown that Roma exhibit a considerable amount of private variants; however, their proportion is lower than in other populations: 15,287 population-specific variants in Roma; 25,060 in IBS; 24,158 in TSI; 21,160 in PJL; 22,040 in GIH; and 24,070 in ITU (Font-Porterias et al., 2021). In addition, allele sharing is high for common variants (MAF > 5%) (over 86% of Roma variants are present in non-Roma) (**Figure 1**). However, for rare variants (MAF < 5%), the allele sharing is around 30–40% (**Figure 1**). Regarding linkage disequilibrium, decay patterns are not statistically different between Roma and non-Roma (**Supplementary Figure 1**, p-value > 0.95). The total number of deleterious alleles per individual is similar among Roma and non-Roma groups (Font-Porterias et al., 2021), and here we further show that the genomic distribution of accumulation of deleterious mutations (i.e., number of deleterious alleles per gene per individual) has comparable patterns between populations (**Supplementary Note 2**, **Supplementary Figures 2, 3**, and **Supplementary Tables 1, 2**). Thus, the overall allele sharing and linkage disequilibrium patterns are comparable among populations. However, rare and private variants can present some challenges in the genetic characterization of Roma population, especially for those variants with a frequency lower than 5%.

As shown in a previous study, Roma and non-Roma groups exhibit differences in the site frequency spectrum due to different demographic histories (Font-Porterias et al., 2021). An overrepresentation analysis including those genes with different allele frequency variants points to a differential OMIM trait enrichment in non-Roma and Roma (**Supplementary Note 3** and **Supplementary Tables 1, 3**). Regarding rare conditions, the genes related to non-Herlitz type junctional epidermolysis bullosa are enriched when comparing deleterious variants with higher frequency in Roma than non-Roma. The prevalence of this disease in Roma is higher than non-Roma in Spain (Martinez-Frias and Bermejo, 1992), although none of the pathogenic variants described in OMIM are present in our dataset and only benign ones show increased frequencies (**Supplementary Table 4**). In addition, the tetralogy of Fallot is enriched in non-Roma (**Table 1**), with two pathogenic variants in *GATA4*

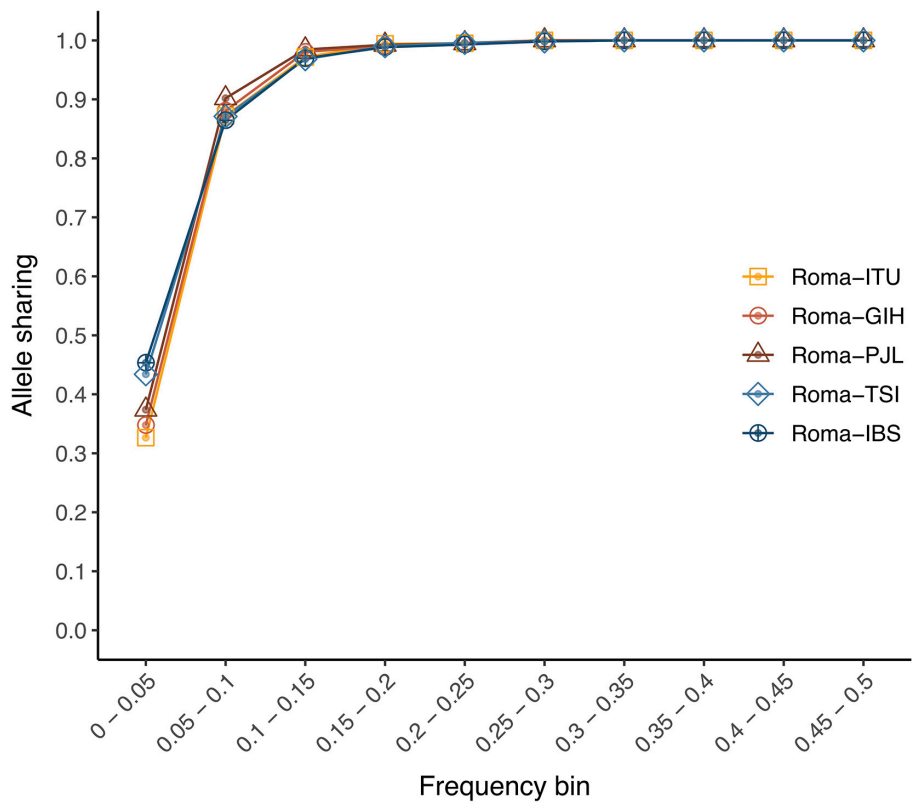

FIGURE 1 | Allele-sharing ratios among Roma and non-Roma groups. Proportion of Roma variants from each minor allele frequency bin (from >0% to 50%) which are also segregating at each non-Roma population.

TABLE 1 | Summary results of the overrepresentation analysis.

Type	Name	Comparison	Associated variants
Rare	Non-Herlitz Epidermolysis Bullosa junctional (OMIM 226650)	Roma > non-Roma	–
Rare	Tetralogy of Fallot (OMIM 187500)	Non-Roma > Roma	rs56208331, rs115099192 (pathogenic) (Tomita-Mitchell et al., 2007; Zhang et al., 2008)
Cardiovascular and metabolic	Obesity (OMIM 601665)	Roma > non-Roma	rs2282440 (associated) (Ha et al., 2006)
Cardiovascular and metabolic	Ischemic stroke (OMIM 601367)	Roma > non-Roma	rs6025 (risk factor) (Majerus, 1994; Casas et al., 2004)
Cardiovascular and metabolic	Insulin dependent diabetes (OMIM 125853)	Roma > non-Roma	–
Cardiovascular and metabolic	Non-insulin dependent diabetes (OMIM 222100)	Non-Roma > Roma	rs1800467 (likely benign) (Phani et al., 2014)
Other	Protection alcohol dependence (OMIM 103780)	Roma > non-Roma	rs1229984 (associated) (MacGregor et al., 2009)
Other	Breast cancer (OMIM 114480)	Non-Roma > Roma	–

Conditions (traits or diseases) are grouped in three categories: rare, cardiovascular/metabolic and other. Name and Phenotype MIM number from OMIM database are shown. Comparison shows the first population as the one with increased allele frequency variants. Associated variants with increased allele frequencies are shown with its clinical annotation and corresponding reference study. More details can be found in ***Supplementary Tables 3–6****.*

gene (rs56208331 and rs115099192) with increased frequencies in South Asian groups (**Supplementary Table 4**). This is consistent with the higher prevalence of this condition in Asia (Takkenberg and Roos-hesselink, 2011).

Multiple cardiovascular and metabolic disorders are also present in the gene overrepresentation analysis. Obesity is enriched comparing Roma with non-Roma, especially with Europeans. One variant in the *SDC3* gene (rs2282440), which is associated with obesity in Asians (**Table 1**) has a significantly increased frequency in Roma (19%) and it is virtually absent in other European groups (**Supplementary Tables 4, 5**). All risk alleles of this variant in the Roma have South Asian ancestry (**Supplementary Table 6**), suggesting that this allele in the Roma has a South Asian origin. In addition, genes related to ischemic stroke are overrepresented in Roma (**Table 1**). One variant in the *F5* gene, annotated as a risk factor in Europeans (rs6025), has significantly higher frequencies in Roma (8.6%) compared to all non-Roma (<1.5%) (**Supplementary Tables 4, 5**). This variant has a European ancestry background for 7 out of 8 risk alleles found (**Supplementary Table 6**), suggesting a European origin of this variant in the Roma. Lastly, both types of diabetes are enriched in our analysis: non-insulin dependent diabetes is enriched in non-Roma, while insulin dependent diabetes is overrepresented in Roma compared to non-Roma (**Table 1**). These results are consistent with previous literature (Mendizabal et al., 2013; Werissa et al., 2019), although our dataset does not include none of the pathogenic or risk-factor variants (more details in **Supplementary Note 3**).

Other conditions such as protection to alcohol dependence are enriched in Roma compared to non-Roma (**Table 1**). One variant in the *ADH1B* gene reported to be protective for alcohol dependence (rs1229984) shows a significantly higher allele frequency in Roma without a clear ancestry origin (**Supplementary Tables 5, 6**). The enrichment for breast cancer in non-Roma (**Table 1**) is due to variants with increased allele frequencies in *ATM*, *BRCA2*, and *NQO2* genes. Roma show higher prevalence of triple negative (TN) breast tumors (Reckova et al., 2017), and mutations in the enriched genes are mostly linked to non-TN types (Peshkin et al., 2010; Lin et al., 2016; Decker et al., 2017; Slavin et al., 2017; Sirisena et al., 2018).

These results show multiple OMIM traits enriched for genes with a 5-fold increased allele frequency deleterious variants both in Roma and in non-Roma. Although this is an exploratory approach, it is consistent with previous literature and reports the first evidence that Roma do not show a systematically increased genetic susceptibility to disease.

A Genome-Wide Screening Does Not Support an Increased Susceptibility for Genetic Disorders

The presence of mutations responsible for particular Mendelian disorders in the Roma has been reported in several studies (see (Kalaydjieva et al., 2001; Álvarez et al., 2005; Bouwer et al., 2007; Gamella et al., 2013; Sevilla et al., 2013; Rocha et al., 2014; Cabrera-Serrano et al., 2018) among others). From these previously reported mutations, we found seven variants in our dataset (**Table 2**). Only in two out of seven variants, the risk allele is present in non-Roma populations: rs1801968 (chr9:132580901) and rs1126809 (chr11:89017961) (**Table 2**). The risk allele of the latter has a higher frequency in Europeans than South Asians (**Table 2**) and, in the Roma, it has a European assigned ancestry in 17 out of 19 alleles (**Supplementary Table 7**), suggesting that this variant (responsible for Oculocutaneous albinism) in the Roma has a European origin, as previously identified (Bianco et al., 2020). In addition, two variants responsible for Charcot–Marie–Tooth disease (rs119483085; chr8:134270617 and rs80338934; chr5:148389835) (**Table 2**) both have one risk allele with European ancestry (**Supplementary Table 7**). On the contrary, rs77931234 (chr1:76226846; Acetyl-coA dehydrogenase deficiency) and rs104894396 (chr13:20763650; Deafness) appear to have a South Asian origin (**Supplementary Table 7**). In fact, rs104894396 risk allele is also present in PJL population

TABLE 2 | List of previously reported Mendelian mutations in Roma present in this study.

Variant – allele	Gene	Roma	IBS	TSI	PJL	ITU	GIH	Disease
rs77931234-G	*ACADM*	0.037	0.000	0.000	0.000	0.000	0.000	Acetyl-coA dehydrogenase deficiency (Rocha et al., 2014)
rs777176261-A	*BIN1*	0.012	0.000	0.000	0.000	0.000	0.000	Centronuclear myopathy (Cabrera-Serrano et al., 2018)
rs80338934-A	*SH3TC2*	0.006	0.000	0.000	0.000	0.000	0.000	Charcot–Marie–Tooth disease (Claramunt et al., 2007; Sevilla et al., 2013)
rs119483085-A	*NDRG1*	0.006	0.000	0.000	0.000	0.000	0.000	Charcot–Marie–Tooth disease (Claramunt et al., 2007; Sevilla et al., 2013)
rs1801968-G	*TOR1A*	0.019	0.1	0.141	0.157	0.165	0.211	Dystonia (Kalaydjieva et al., 2001)
rs1126809-A	*TYR*	0.154	0.305	0.269	0.079	0.006	0.103	Oculocutaneous albinism (Gamella et al., 2013)
rs104894396-T	*GJB2*	0.019	0.000	0.000	0.007	0.000	0.000	Deafness (Álvarez et al., 2005; Bouwer et al., 2007)

Variant rs ID and risk allele, risk allele frequency for each population and disease association are shown.

(**Table 2**), which is consistent with a South Asian origin. This variant, only present in Roma and Punjabi individuals in our dataset, is a non-synonymous mutation (W42X) responsible for autosomal recessive non-syndromic hearing loss. It has been previously reported in Spanish and Slovak Roma individuals and an Indian origin has been suggested (Minárik et al., 2003; Álvarez et al., 2005), which is congruent with the South Asian ancestry assignation of the risk allele.

We next investigated the 334 pathogenic variants described in the ClinVar database present in our dataset. In Roma, we found 60 out of 334 variants segregating at low RAF in the population. Only 27 variants have a RAF difference equal or higher than 1.5 comparing Roma and non-Roma (**Supplementary Table 8**). Although the RAF of these variants is low (below 5% in most cases), there are variants with increased frequency in Roma, but interestingly, we observe disease-associated variants with increased frequencies in European and South Asian non-Roma populations. For example, rs1799807 (chr 3:165548529) is a missense pathogenic variant causing the deficiency of butyrylcholine esterase and, consequently, postanesthetic apnea (McGuire et al., 1989; Jasiecki et al., 2019). This variant is only present in European populations and virtually absent in South Asia and Roma (**Supplementary Table 8**). On the contrary, rs137941190 (chr 11:126215441) is a missense pathogenic variant for Al-Raqad syndrome, described in Pakistani patients (Ahmed et al., 2014). The risk allele of this variant is absent in the European exomes, but it appears at low frequencies in the Roma and South Asian samples (**Supplementary Table 8**), although the genotype frequencies are not statistically different (**Supplementary Table 9**). Regarding the ancestry inference, European ancestry is assigned for the risk allele of most of these variants (**Supplementary Table 10**), except for rs104894396 (chr 13:20763650), as explained above.

The screening of known disease-associated variants together with the local ancestry inference has allowed to report both the presence and the absence of particular mutations in the Roma and to trace their most likely ancestral origin. However, most known clinically relevant variants have been discovered in European populations, which leads to an ascertainment bias that can weaken the results when studying the Roma population: out of the 334 pathogenic variants, less than 45 variants are segregating in South Asian populations (43 in PJL, 41 in GIH, and 31 in ITU), while there are 84 and 74 segregating in IBS and TSI, respectively.

Many Drug-Response Variants in Roma Are Related to Metabolic and Cardiovascular Disorders

Besides disease-associated variants, other functionally relevant mutations (e.g., pharmacogenomic variants) exhibit inter-population genetic variation in the human genome, as mentioned above. Regarding drug binding domains, we identified 101 variants in our dataset that disrupt the domains without being deleterious for the protein. This set is less biased toward European genetic variation, since it is based on impact prediction, rather than on previously discovered genetic associations (Dopazo et al., 2016). Only 26 variants were found to have a MAF fold increase $\geq$1.5 comparing Roma and non-Roma (**Supplementary Table 11**). Variants with known association drug phenotypes reported in PharmGKB (Whirl-Carrillo et al., 2012) have higher MAF in European populations, showing the European-centric bias in biomedical genetic studies. For example, the rs5918 variant, located in the *ITGB3* gene (chr17:45360730), reduces the efficacy of aspirin and clopidogrel (Dropinski et al., 2007; Motovska et al., 2009) (indicated for coronary artery disease and myocardial infarction) and it shows a higher allele frequency in IBS and TSI than in Roma (**Supplementary Table 11**). A variant found in the *GLP1R* gene (rs6923761; chr6:39034072) reduces the treatment efficacy for obesity and type II diabetes (i.e., sitagliptin, vildagliptin, and liraglutide) (Javorský et al., 2016) and its frequency is significantly lower in Roma than in the tested European populations (**Supplementary Table 11**).

On the contrary, there are 13 variants with increased MAF in Roma (**Supplementary Table 11**) with significantly different genotype frequencies compared to non-Roma groups (**Supplementary Table 12**). For example, there is a variant in the *CRYZ* gene (chr1:75175886) with significantly higher MAF in Roma and South Asians than in Europeans (**Supplementary Table 11**). Its protein contains a drug binding domain for dicumarol, indicated for deep vein thrombosis (**Supplementary Table 11**) and 19 out of 27 minor alleles in Roma have South Asian ancestry (**Supplementary Table 13**). In the *PTPRE* gene drug binding domain for alendronate (indicated for

osteoporosis) (**Supplementary Table 11**), we identify a variant (chr10:129868686) in Roma (3.7%), virtually absent in non-Roma, except GIH (0.5%) (**Supplementary Table 11**). Four Roma individuals were heterozygotes for this variant: one with both haplotypes assigned to South Asian ancestry and three with one European and one South Asian haplotype (**Supplementary Table 13**), which suggests that this variant in the Roma originated in South Asia. Regarding the ancestry inference and besides the mentioned examples, many of the variants have the minor allele with European ancestry: 72% of the minor alleles are assigned to a European-related ancestry; slightly above the mean genome-wide ancestry (68.4%), but within the first SD of the distribution. Although experimental evidence suggesting these variants affect the binding of these drugs is lacking, follow-up studies should be performed to validate the functional impact of these variants.

We next examined previously described variants in ADME genes. In our dataset, 14 out of 95 of them show increased MAF with a fold change equal or higher than 1.5 comparing Roma and non-Roma (**Table 3**). Some variants found in European groups are absent in the Roma exomes: e.g., rs34130495, which modifies the metabolism of tramadol (indicated for mild-to-moderate pain) (**Table 3**). However, many of the variants with increased frequencies in Spanish Roma are clinically associated with metabolic and cardiovascular-related drugs (**Table 3**) and some have significantly different genotype frequencies between Roma and non-Roma (**Supplementary Table 14**). For example, the rs4149056 variant shows a higher MAF in Roma than in IBS (18% and 11%, respectively) (**Table 3**) and it increases the risk of toxicity to simvastatin (indicated for hypercholesterolemia) (**Table 3**). A previous study reports a frequency of 17.2 and 18.9% of this variant in Roma and non-Roma groups from Hungary, respectively (Nagy et al., 2015), suggesting that this variant in the Spanish Roma was present before the arrival into the Iberian Peninsula. rs316019 variant also shows a significant MAF increase in Roma than in IBS and TSI populations (20%, 9%, and 11%, respectively) (**Table 3**) and it is reported to modify the metabolism of metformin, a drug used to treat type II diabetes (**Table 3**). Regarding ancestry inference, the minor alleles of these variants are almost exclusively of European ancestry (**Supplementary Table 15**). Lastly, three previously found variants with increased frequencies in Croatian Roma (Škarić-Jurić et al., 2018) do not show significantly higher MAFs in Spanish Roma (rs10509681, rs8192709, and rs34059508) (**Table 3**).

The screening beyond disease-associated variants in the Roma population reveals that most of them can change the response of drugs used for metabolic and cardiovascular disorders and that they might have a European origin. However, this analysis is based on the impact prediction and it is important to take into account that the phenotype is also influenced by the environment, non-coding variants and regulatory elements, among others.

DISCUSSION

The underrepresentation of human populations in genetic studies impairs the understanding of genome architecture and exacerbates health differences. In order to overcome this limitation, the Eurocentric bias in the discovery of functional variants has to be taken into account (Need and Goldstein, 2009; Popejoy and Fullerton, 2016; Bentley et al., 2017; Martin et al., 2019; Sirugo et al., 2019). In the case of the Roma population, we found that the genetic portability with European populations is overall high: allele sharing and LD decay patterns are comparable among groups. However, low frequency variants (MAF < 5%) can present some challenges in the genetic characterization of this population, since only half of these variants in the Roma are also segregating

TABLE 3 | List of variants in ADME genes found to have a fold increase in allele frequency equal or higher than 1.5 comparing Roma and non-Roma.

Variant – allele	Gene	Roma	IBS	TSI	PJL	ITU	GIH	Clinical annotation
rs12208357-T	*SLC22A1*	0.019 (0.022)	0.058	0.058	0.021	0.024	0.010	Metabolism metformin (Shu et al., 2008; Sundelin et al., 2017)
rs2282143-T	*SLC22A1*	0.037 (0.085)	0.005	0.013	0.079	0.059	0.082	Metabolism metformin (Yoon et al., 2013)
rs34130495-A	*SLC22A1*	0.000 (0.023)	0.032	0.019	0.000	0.000	0.000	Metabolism tramadol (Tzvetkov et al., 2011)
rs34059508-A	*SLC22A1*	0.012 (**0.048**)	0.021	0.006	0.000	0.000	0.000	Metabolism metformin (Shu et al., 2008)
rs316019-A	*SLC22A2*	0.204 (0.066)	0.089	0.109	0.107	0.165	0.134	Metabolism metformin; Toxicity cisplatin-anthracyclines (Visscher et al., 2012; Yoon et al., 2013)
rs1800460-T	*TPMT*	0.025 (0.002)	0.037	0.019	0.000	0.000	0.000	Toxicity azathioprine and mercaptopurine (Stocco et al., 2012; Steponaitiene et al., 2016)
rs717620-T	*ABCC2*	0.185 (0.253)	0.211	0.186	0.071	0.047	0.077	Efficacy and dosage atorvastatin; Toxicity fluorouracil (Cecchin et al., 2013; Prado et al., 2018)
rs4244285-A	*CYP2C19*	0.191 (0.155)	0.133	0.090	0.314	0.388	0.330	Efficacy and toxicity clopidogrel; Efficacy amitriptyline
rs10509681-C	*CYP2C8*	0.099 (**0.165**)	0.168	0.141	0.050	0.024	0.041	Metabolism rosiglitazone (Dawed et al., 2016)
rs1058930-C	*CYP2C8*	0.105 (0.056)	0.042	0.051	0.014	0.006	0.010	Metabolism diclofenac (Dorado et al., 2008)
rs4149056-C	*SLCO1B1*	0.179 (0.102)	0.111	0.224	0.050	0.076	0.021	Toxicity simvastatin (Shek et al., 2017)
rs1048943-C	*CYP1A1*	0.037 (0.033)	0.016	0.045	0.107	0.118	0.113	Efficacy capecitabine and docetaxel (Dong et al., 2012)
rs1799814-T	*CYP1A1*	0.043 (0.033)	0.089	0.013	0.014	0.018	0.005	Metabolism dacarbazine (Lewis et al., 2016)
rs8192709-T	*CYP2B6*	0.093 (**0.128**)	0.047	0.064	0.043	0.035	0.036	Toxicity efavirenz (Dhoro et al., 2014; Dickinson et al., 2016)

Variant rs ID and risk allele, gene, risk allele frequency for each population and PharmGKB main clinical annotations are shown. Croatian Roma frequencies from Škarić-Jurić et al., 2018 are also included within brackets in the Roma column (those variants with significantly higher frequencies are in bold).

in non-Roma populations. This is consistent with the fact that low frequency and rare variants are more population structured than common variants (Casals and Bertranpetit, 2012; Bomba et al., 2017).

The overrepresentation analysis shows enrichment of some gene sets with increased MAF deleterious variants for genetic disorders. Interestingly, this enrichment occurs both in Roma and in non-Roma, consistent with previous literature. For example, we identify an enrichment for non-triple negative breast cancer in non-Roma, in agreement with a lower incidence in Roma, which show more triple negative cases (Reckova et al., 2017). *Triple negative* refers to the overexpression of three common markers (i.e., estrogen receptor, progesterone receptor or *HER2* oncogene) (Foulkes et al., 2010). However, the overexpression of these markers is common in breast cancer patients of European descent, but not in other groups, such as African-descent (Brewster et al., 2014) or Roma (Reckova et al., 2017) patients.

In the present study, we also examine previously defined disease-associated variants. Besides confirming the ancestry origin of some Mendelian mutations in the Roma, we provide new evidence: the risk allele responsible for Acetyl-coA dehydrogenase deficiency (rs77931234) is traced to a South Asian-like ancestry, while the risk alleles for Charcot–Marie–Tooth disease variants (rs119483085, rs80338934) show European-related ancestry. In addition, we perform a comprehensive study of the pathogenic variation reported in ClinVar database (Landrum et al., 2014). The results show a different frequency spectrum of previously identified variants when comparing Roma and non-Roma groups. However, most of the variants in the Roma are traced to a European origin, evidencing the Eurocentric bias in public databases (Kessler et al., 2016).

Regarding pharmacogenetic variation, we identify variants with increased and decreased MAF in Roma. Many of the variants in drug-binding domains that we found with increased frequencies in non-Roma European populations have been previously associated with a drug response trait, while those with increased frequencies in Roma or South Asian groups are still not specifically characterized. This is particularly relevant, since many of them are related to metabolic and cardiovascular drugs and previous studies suggest that Roma show higher prevalence of these diseases (Vozarova De Courten et al., 2003; Živković et al., 2010). Following the expectations of European ancestry proportions, most drug-response variants in Roma are traced to a European-related origin, which is not a signal of bias, because the list of variants that disrupt drug binding domains is not based on previously known associations discovered in populations of European descent (Dopazo et al., 2016). Thus, these variants show inter-population variability in the Roma, since a differential European gene flow among Roma groups has been previously described (Mendizabal et al., 2012; Font-Porterias et al., 2019). In this sense, the variants reported with increased frequency in Spanish Roma do not completely overlap with a similar study in Croatian Roma (Škarić-Jurić et al., 2018). Moreover, the significantly increased frequency of a *CYP2C19* polymorphism (rs4244285) in Hungarian and Portuguese Roma groups (Sipeky et al., 2013; Teixeira et al., 2015) is not observed in Spanish Roma.

For complex diseases, polygenic risk scores (PRSs) are designed to predict the phenotype from genetic data, combining the effect sizes of multiple variants and their frequency (Wray et al., 2007). However, the pre-computed effect sizes usually derive from genome-wide association studies (GWAS), and given its Eurocentric bias, PRSs have a greater predictive accuracy in populations with European ancestry (Martin et al., 2017, 2019; De La Vega and Bustamante, 2018; Kim et al., 2018; Gurdasani et al., 2019). Particularly, they show a systematic bias when applied to other populations due to several factors: (i) GWAS are biased toward those variants segregating in the study population; (ii) when two populations have different LD patterns, tagSNPs and causal variants can differ; and (iii) environmental and genetic factors can be confounded when phenotypes are geographically stratified (Martin et al., 2019; Sirugo et al., 2019). Although new methods are emerging to overcome these limitations (Márquez-Luna et al., 2017; Marnetto et al., 2020), an in-depth analysis of PRS accuracy in non-European-descent populations is needed before implementing them to other under-represented populations (Sirugo et al., 2019).

Here, we provide new evidence of a different frequency spectrum of clinically relevant variants across populations: while some have increased allele frequencies in the Roma, others are virtually absent. This was possible due to the availability of a substantial number of whole-exome sequences at high coverage, which allows the study of clinically relevant genetic variation, enriched in low frequency variants (Bomba et al., 2017). However, this different frequency spectrum cannot be directly attributed to a lower prevalence of some diseases; instead, other unstudied variants might be responsible for disease susceptibility and drug response in this population. Although it is an exploratory approach without functional validation, this study aims to shift the traditional paradigm of focusing only on the increased genetic risk for some diseases in the so-called "isolated" populations. In fact, these results further confirm that Roma are not so genetically isolated. Gene flow with European groups accounts for 65% of their genetic ancestry (Font-Porterias et al., 2019); thus, clinically damaging variants are traced both to South Asian and European-related haplotypes. Lastly, we caution that these results are geographically limited to the Roma population from Spain, and further characterization should be performed in other groups with different demographic trajectories and with increasing sample sizes.

We would like to remark that this study does not aim to exacerbate the importance of inter-population variability, to justify health differences in minority ethnic groups, or to advocate for racialized medicine. In fact, genetic ancestry is not the only determinant of ethnicity or health, and social factors should be considered. Given that genetic diversity is a continuum, large scale genome-wide studies are needed to fully capture and represent human variation, without excluding any population while respecting their rights and interests and properly accounting for demographic differences. This would

prevent the current overgeneralization of the results obtained from genetic studies on populations with only European ancestry in the assessment of disease risk testing and treatment response (Martin et al., 2019; Batai et al., 2020; Hudson et al., 2020).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the CEIC-Parc de Salut Mar 2019/8900/I. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NF-P and DC contributed to the design and conception of the study. NF-P performed and implemented the data analysis. All authors contributed to the interpretation and discussion of the results and writing of the manuscript, and approved the submitted version.

FUNDING

This study was supported by the Spanish Ministry of Science, Innovation and Universities (MCIU), Agencia Estatal de Investigación (AEI) grant number PID2019-106485GB-I00/AEI/10.13039/501100011033, and "Unidad de Excelencia María de Maeztu" (AEI, CEX2018-000792-M). NF-P was supported by a FPU17/03501 fellowship.

ACKNOWLEDGMENTS

We would like to thank the invaluable collaboration of the Federacioì d'Associacions Gitanes de Catalunya (FAGiC). We would also like to thank the members of the Human Genome Diversity Group and Genomics of Individuality Lab (Institut de Biologia Evolutiva, UPF-CSIC) for the insightful discussions about the results of this study and the computational support from the Scientific IT Core Facility from Universitat Pompeu Fabra.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.683880/full#supplementary-material

REFERENCES

1000 Genomes Project Consortium, (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65. doi: 10.1038/nature11632

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. doi: 10.1038/nmeth0410-248

Ahmed, I., Buchert, R., Zhou, M., Jiao, X., Mittal, K., Sheikh, T. I., et al. (2014). Mutations in DCPS and EDC3 in autosomal recessive intellectual disability indicate a crucial role for mRNA decapping in neurodevelopment. *Hum. Mol. Genet.* 24, 3172–3180. doi: 10.1093/hmg/ddv069

Álvarez, A., Castillo, I., Villamar, M., Aguirre, L. A., González-Neira, A., López-Nevot, A., et al. (2005). High Prevalence of the W24X mutation in the gene encoding Connexin-26 (GJB2) in Spanish Romani (Gypsies) with autosomal recessive non-syndromic hearing loss. *Am. J. Med. Genet.* 137A, 255–258. doi: 10.1002/ajmg.a.30884

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393

Azmanov, D. N., Dimitrova, S., Florez, L., Cherninkova, S., Draganov, D., Morar, B., et al. (2011). LTBP2 and CYP1B1 mutations and associated ocular phenotypes in the Roma/Gypsy founder population. *Eur. J. Hum. Genet.* 19, 326–333. doi: 10.1038/ejhg.2010.181

Barbujani, G., Ghirotto, S., and Tassi, F. (2013). Nine things to remember about human genome diversity. *Tissue Antigens*82, 155–164. doi: 10.1111/tan.12165

Batai, K., Hooker, S., and Kittles, R. A. (2020). Leveraging genetic ancestry to study health disparities. *Am. J. Phys. Anthropol.* 175, 1–13.

Bentley, A. R., Callier, S., and Rotimi, C. N. (2017). Diversity and inclusion in genomic research: why the uneven progress? *J. Commun. Genet.* 8, 255–266. doi: 10.1007/s12687-017-0316-6

Bianco, E., Laval, G., Font-Porterias, N., García-Fernandez, C., Dobon, B., Sabido-Vera, R., et al. (2020). Recent common origin, reduced population size, and marked admixture have shaped European Roma genomes. *Mol. Biol. Evol.* 37, 3175–3187. doi: 10.1093/molbev/msaa156

Bomba, L., Walter, K., and Soranzo, N. (2017). The impact of rare and low-frequency genetic variants in common disease. *Genome Biol.* 18, 1–17.

Bouwer, S., Angelicheva, D., Chandler, D., Seeman, P., Tournev, I., and Kalaydjieva, L. (2007). Carrier rates of the ancestral Indian W24X mutation in GJB2 in the general Gypsy population and individual subisolates. *Genet. Test.* 11, 455–458. doi: 10.1089/gte.2007.0048

Bress, A., Patel, S. R., Perera, M. A., Campbell, R. T., Kittles, R. A., and Cavallari, L. H. (2012). Effect of NQO1 and CYP4F2 genotypes on warfarin dose requirements in Hispanic-Americans and African-Americans. *Pharmacogenomics* 13, 1925–1935. doi: 10.2217/pgs.12.164

Brewster, A. M., Chavez-MacGregor, M., and Brown, P. (2014). Epidemiology, biology, and treatment of triple-negative breast cancer in women of African ancestry. *Lancet Oncol.* 15, e625–e634.

Browning, S. R., Browning, B. L., Daviglus, M. L., Durazo, R. A., Schneiderman, N., Kaplan, R. C., et al. (2018). Ancestry-specific recent effective population size in the Americas. *PLoS Genet.* 14:e1007385. doi: 10.1371/journal.pgen.1007385

Cabrera-Serrano, M., Mavillard, F., Biancalana, V., Rivas, E., Morar, B., Hernández-Laín, A., et al. (2018). A Roma founder BIN1 mutation causes a novel phenotype of centronuclear myopathy with rigid spine. *Neurology* 91, e339–e348.

Casals, F., and Bertranpetit, J. (2012). Human genetic variation, shared and private. *Science* 336, 39–40. doi: 10.1126/science.1224528

Casas, J. P., Hingorani, A. D., Bautista, L. E., and Sharma, P. (2004). Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18 000 cases and 58 000 controls. *Arch. Neurol.* 61, 1652–1661. doi: 10.1001/archneur.61.11.1652

Cecchin, E., D'Andrea, M., Lonardi, S., Zanusso, C., Pella, N., Errante, D., et al. (2013). A prospective validation pharmacogenomic study in the adjuvant setting of colorectal cancer patients treated with the 5-fluorouracil/leucovorin/oxaliplatin (FOLFOX4) regimen. *Pharmacogenomics J.* 13, 403–409. doi: 10.1038/tpj.2012.31

Claramunt, R., Sevilla, T., Lupo, V., Cuesta, A., Millán, J. M., Vílchez, J. J., et al. (2007). The p.R1109X mutation in SH3TC2 gene is predominant in Spanish Gypsies with Charcot-Marie-Tooth disease type 4. *Clin. Genet.* 71, 343–349. doi: 10.1111/j.1399-0004.2007.00774.x

Council of Europe, (2012). *Human Rights of Roma and Travellers in Europe*. Strasbourg: Council of Europe.

Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* 6:e1001025. doi: 10.1371/journal.pcbi.1001025

Dawed, A. Y., Donnelly, L., Tavendale, R., Carr, F., Leese, G., Palmer, C. N. A., et al. (2016). CYP2C8 and SLCO1B1 variants and therapeutic response to thiazolidinediones in patients with Type 2 diabetes. *Diabetes Care* 39, 1902–1908. doi: 10.2337/dc15-2464

De La Vega, F. M., and Bustamante, C. D. (2018). Polygenic risk scores: a biased prediction? *Genome Med.* 10:100.

Decker, B., Allen, J., Luccarini, C., Pooley, K. A., Shah, M., Bolla, M. K., et al. (2017). Rare, protein-truncating variants in ATM, CHEK2 and PALB2, but not XRCC2, are associated with increased breast cancer risks. *J. Med. Genet.* 54, 732–741.

Dhoro, M., Ngara, B., Kadzirange, G., Nhachi, C., and Masimirembwa, C. (2014). Genetic variants of drug metabolizing enzymes and drug transporter (ABCB1) as possible biomarkers for adverse drug reactions in an HIV/AIDS Cohort in Zimbabwe. *Curr. HIV Res.* 11, 481–490. doi: 10.2174/1570162x113119990048

Dickinson, L., Amin, J., Else, L., Boffito, M., Egan, D., Owen, A., et al. (2016). Comprehensive pharmacokinetic, pharmacodynamic and pharmacogenetic evaluation of once-daily Efavirenz 400 and 600 mg in Treatment-Naïve HIV-Infected Patients at 96 weeks: results of the ENCORE1 study. *Clin. Pharmacokinet.* 55, 861–873. doi: 10.1007/s40262-015-0360-5

Dong, N., Yu, J., Wang, C., Zheng, X., Wang, Z., Di, L., et al. (2012). Pharmacogenetic assessment of clinical outcome in patients with metastatic breast cancer treated with docetaxel plus capecitabine. *J. Cancer Res. Clin. Oncol.* 138, 1197–1203. doi: 10.1007/s00432-012-1183-5

Dopazo, J., Amadoz, A., Bleda, M., Garcia-Alonso, L., Alemán, A., García-García, F., et al. (2016). 267 spanish exomes reveal population-specific differences in disease-related genetic variation. *Mol. Biol. Evol.* 33, 1205–1218. doi: 10.1093/molbev/msw005

Dorado, P., Cavaco, I., Cáceres, M. C., Piedade, R., Ribeiro, V., and LLerena, A. (2008). Relationship between CYP2C8 genotypes and diclofenac 5-hydroxylation in healthy Spanish volunteers. *Eur. J. Clin. Pharmacol.* 64, 967–970. doi: 10.1007/s00228-008-0508-4

Dropinski, J., Musial, J., Sanak, M., Wegrzyn, W., Nizankowski, R., and Szczeklik, A. (2007). Antithrombotic effects of aspirin based on PLA1/A2 glycoprotein IIIa polymorphism in patients with coronary artery disease. *Thromb. Res.* 119, 301–303. doi: 10.1016/j.thromres.2006.03.005

Font-Porterias, N., Caro-Consuegra, R., Lucas-Saìnchez, M., Lopez, M., Gimeìnez, A., Carballo, A., et al. (2021). The counteracting effects of demography on functional genomic variation: the Roma paradigm. *Mol. Biol. Evol.* msab070.

Font-Porterias, N., Arauna, L. R., Poveda, A., Bianco, E., Rebato, E., Prata, M. J., et al. (2019). European Roma groups show complex West Eurasian admixture footprints and a common South Asian genetic origin. *PLoS Genet.* 15:e1008417. doi: 10.1371/journal.pgen.1008417

Foulkes, W. D., Smith, I. E., and Reis-Filho, J. S. (2010). Triple-Negative Breast Cancer. *N. Engl. J. Med.* 363, 1938–1948.

Gamella, J. F., Carrasco-Muñoz, E. M., and Núñez Negrillo, A. M. (2013). Oculocutaneous albinism and consanguineous marriage among Spanish Gitanos or Calé-a study of 83 cases. *Coll Antropol.* 37, 723–734.

Gurdasani, D., Barroso, I., Zeggini, E., and Sandhu, M. S. (2019). Genomics of disease risk in globally diverse populations. *Nat. Rev. Genet.* 20, 520–535. doi: 10.1038/s41576-019-0144-0

Ha, E., Kim, M. J., Choi, B. K., Rho, J. J., Oh, D. J., Rho, T. H., et al. (2006). Positive association of obesity with single nucleotide polymorphisms of syndecan 3 in the Korean population. *J. Clin. Endocrinol. Metab.* 91, 5095–5099. doi: 10.1210/jc.2005-2086

Hancock, I. F. (2002). *We are the Romani People*. Hatfield: University of Hertfordshire Press.

Heyer, E., and Reynaud-Paligot, C. (2020). Us and them: from prejudice to racism. An original analysis of race and racism. *Am. J. Phys. Anthropol.* 175,1–9. doi: 10.4135/9781446218877.n1

Hopkins, A. L., and Groom, C. R. (2002). The druggable genome. *Nat. Rev. Drug Discov.* 1, 727–730.

Hudson, M., Garrison, N. A., Sterling, R., Caron, N. R., Fox, K., Yracheta, J., et al. (2020). Rights, interests and expectations: indigenous. *Nat. Rev. Genet.* 21, 377–384.

International HapMap Consortium, (2003). The international HapMap project. *Nature* 426, 789–796. doi: 10.1038/nature02168

Jasiecki, J., Żuk, M., Krawczyńska, N., Jońca, J., Szczoczarz, A., Lewandowski, K., et al. (2019). Haplotypes of butyrylcholinesterase K-variant and their influence on the enzyme activity. *Chem. Biol. Interact.* 307, 154–157. doi: 10.1016/j.cbi.2019.05.007

Javorský, M., Gotthardová, I., Klimčáková, L., Kvapil, M., Židzik, J., Schroner, Z., et al. (2016). A missense variant in GLP1R gene is associated with the glycaemic response to treatment with gliptins. *Diabetes Obes Metab.* 18, 941–944. doi: 10.1111/dom.12682

Johnson, J., Caudle, K., Gong, L., Whirl-Carrillo, M., Stein, C., Scott, S., et al. (2017). Clinical pharmacogenetics implementation consortium (CPIC) guideline for pharmacogenetics-guided warfarin dosing: 2017 update. *Clin. Pharmacol. Ther.* 102, 397–404. doi: 10.1002/cpt.668

Kalaydjieva, L., Gresham, D., and Calafell, F. (2001). Genetic studies of the Roma (Gypsies): a review. *BMC Med. Genet.* 2:5. doi: 10.1186/1471-2350-2-5

Kessler, M. D., Yerges-Armstrong, L., Taub, M. A., Shetty, A. C., Maloney, K., Jeng, L. J. B., et al. (2016). Challenges and disparities in the application of personalized genomic medicine to populations with African ancestry. *Nat. Commun.* 7, 1–8.

Kim, M. S., Patel, K. P., Teng, A. K., Berens, A. J., and Lachance, J. (2018). Genetic disease risks can be misestimated across global populations. *Genome Biol.* 19:179.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2019). PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* 47, D1102–D1109.

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., et al. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 42, 980–985.

Lao, O., Lu, T. T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., et al. (2008). Correlation between genetic and geographic structure in Europe. *Curr. Biol.* 18, 1241–1248.

Lewis, B. C., Korprasertthaworn, P., and Miners, J. O. (2016). Impaired dacarbazine activation and 7-ethoxyresorufin deethylation in vitro by polymorphic variants of CYP1A1 and CYP1A2: implications for cancer therapy. *Pharmacogenet. Genomics* 26, 453–461. doi: 10.1097/fpc.0000000000000236

Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* 47, W199–W205.

Lin, P. H., Kuo, W. H., Huang, A. C., Lu, Y. S., Lin, C. H., Kuo, S. H., et al. (2016). Multiple gene sequencing for risk assessment in patients with early-onset or familial breast cancer. *Oncotarget* 7, 8310–8320.

MacGregor, S., Lind, P. A., Bucholz, K. K., Hansell, N. K., Madden, P. A. F., Richter, M. M., et al. (2009). Associations of ADH and ALDH2 gene variation with self report alcohol reactions, consumption and dependence: an integrated analysis. *Hum. Mol. Genet.* 18, 580–593. doi: 10.1093/hmg/ddn372

Majerus, P. W. (1994). Bad blood by mutation. *Nature* 369, 14–15. doi: 10.1038/369014a0

Maples, B. K., Gravel, S., Kenny, E. E., and Bustamante, C. D. (2013). RFMix?: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* 93, 278–288. doi: 10.1016/j.ajhg.2013.06.020

Marnetto, D., Pärna, K., Läll, K., Molinaro, L., Montinaro, F., Haller, T., et al. (2020). Ancestry deconvolution and partial polygenic score can improve susceptibility predictions in recently admixed individuals. *Nat. Commun.* 11, 1–9.

Márquez-Luna, C., Loh, P. R., Price, A. L., Kooner, J. S., Saleheen, D., Sim, X., et al. (2017). Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet. Epidemiol.* 41, 811–823. doi: 10.1002/gepi.22083

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., et al. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* 100, 635–649. doi: 10.1016/j.ajhg.2017.03.004

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., and Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* 51, 584–591. doi: 10.1038/s41588-019-0379-x

Martinez-Frias, M. L., and Bermejo, E. (1992). Prevalence of congenital anomaly syndromes in a Spanish gypsy population. *J. Med. Genet.* 29, 483–486.

Matras, Y. (2002). *Romani: A Linguistic Introduction*. Cambridge, MA: Cambridge University Press.

McGuire, M. C., Nogueira, C. P., Bartels, C. F., Lightstone, H., Hajra, A., Van der Spek, A. F. L., et al. (1989). Identification of the structural mutation responsible for the dibucaine-resistant (atypical) variant form of human serum cholinesterase. *Proc. Natl. Acad. Sci. U.S.A.* 86, 953–957. doi: 10.1073/pnas.86.3.953

McLaren, W., Gil, L., Hunt, S.E., et al. (2016). The Ensembl Variant Effect Predictor. *Genome Biol.* 17, 122. doi: 10.1186/s13059-016-0974-4

Mendizabal, I., Lao, O., Marigorta, U. M., Kayser, M., and Comas, D. (2013). Implications of population history of European Romani on genetic susceptibility to disease. *Hum. Hered.* 76, 194–200. doi: 10.1159/000360762

Mendizabal, I., Lao, O., Marigorta, U. M., Wollstein, A., Gusmão, L., Ferak, V., et al. (2012). Reconstructing the population history of European Romani from genome-wide data. *Curr. Biol.* 22, 2342–2349. doi: 10.1016/j.cub.2012.10.039

Minárik, G., Ferák, V., Feráková, E., Ficek, A., Poláková, H., and Kádasi, L. (2003). High frequency of GJB2 Mutation W24X among Slovak Romany (Gypsy) patients with non-syndromic hearing loss (NSHL). *Gen. Physiol. Biophys.* 22, 549–556.

Moorjani, P., Patterson, N., Loh, P.-R., Lipson, M., Kisfali, P., Melegh, B. I., et al. (2013). Reconstructing Roma history from genome-wide data. *PLoS One* 8:e58633. doi: 10.1371/journal.pone.0058633

Morar, B., Azmanov, D. N., and Kalaydjieva, L. (2013). "Roma (Gypsies): genetic Studies," in *eLS* (Chichester: Wiley). Eds., John Wiley & Sons, Ltd

Motovska, Z., Widimsky, P., Kvasnicka, J., Petr, R., Bilkova, D., Hajkova, J., et al. (2009). High loading dose of clopidogrel is unable to satisfactorily inhibit platelet reactivity in patients with glycoprotein IMA gene polymorphism: a genetic substudy of PRAGUE-8 trial. *Blood Coagul. Fibrinolysis* 20, 257–262. doi: 10.1097/mbc.0b013e328325455b

Nagy, A., Sipeky, C., Szalai, R., Melegh, B. I., Matyas, P., Ganczer, A., et al. (2015). Marked differences in frequencies of statin therapy relevant SLCO1B1 variants and haplotypes between Roma and Hungarian populations. *BMC Genet.* 16:108. doi: 10.1186/s12863-015-0262-4

Need, A. C., and Goldstein, D. B. (2009). Next generation disparities in human genomics: concerns and remedies. *Trends Genet.* 25, 489–494. doi: 10.1016/j.tig.2009.09.012

O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., et al. (2014). A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet.* 10:e1004234. doi: 10.1371/journal.pgen.1004234

Padoa, C., Goldman, A., Jenkins, T., and Ramsay, M. (1999). Cystic fibrosis carrier frequencies in populations of African origin. *J. Med. Genet.* 36, 41–44.

Peshkin, B. N., Alabek, M. L., and Isaacs, C. (2010). BRCA1/2 mutations and triple negative breast cancers. *Breast Dis.* 32, 25–33. doi: 10.3233/bd-2010-0306

Phani, N. M., Guddattu, V., Bellampalli, R., Seenappa, V., Adhikari, P., Nagri, S. K., et al. (2014). Population specific impact of genetic variants in KCNJ11 gene to type 2 diabetes: a case-control and meta-analysis study. *PLoS One* 9:e107021. doi: 10.1371/journal.pone.0107021

Popejoy, A. B., and Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature* 538, 161–164. doi: 10.1038/538161a

Prado, Y., Arencibia, A., Zambrano, T., and Salazar, L. A. (2018). Gender-specific association between ABCC2 -24C>T SNP and reduction in triglycerides in chilean patients treated with atorvastatin. *Basic Clin. Pharmacol. Toxicol.* 122, 517–522. doi: 10.1111/bcpt.12943

Reckova, M., Mardiak, J., Plank, L., Vulevova, M., Cingelova, S., and Mego, M. (2017). Differences in incidence and biological characteristics of breast cancer between roma and non-roma patients in Slovakia. *Klin Onkol.* 30, 48–54. doi: 10.14735/amko201748

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 47, D886–D894.

Rocha, H., Castiñeiras, D., Delgado, C., Egea, J., Yahyaoui, R., González, Y., et al. (2014). Birth prevalence of fatty acid β-Oxidation disorders in iberia. *JIMD Rep.* 16, 89–94. doi: 10.1007/8904_2014_324

Sevilla, T., Martínez-Rubio, D., Márquez, C., Paradas, C., Colomer, J., Jaijo, T., et al. (2013). Genetics of the charcot-marie-tooth disease in the spanish gypsy population: the hereditary motor and sensory neuropathy-Russe in depth. *Clin. Genet.* 83, 565–570. doi: 10.1111/cge.12015

Shek, A. B., Kurbanov, R. D., Abdullaeva, G. J., Nagay, A. V., Hoshimov, S. U., Nizamov, U. I., et al. (2017). Simvastatin intolerance genetic determinants: some features in ethnic Uzbek patients with coronary artery disease. *Arch. Med. Sci. Atheroscler. Dis.* 2, 68–75. doi: 10.5114/amsad.2017.70597

Shu, Y., Brown, C., Castro, R. A., Shi, R. J., Lin, E. T., Owen, R. P., et al. (2008). Effect of genetic variation in the organic cation transporter 1, OCT1, on metformin pharmacokinetics. *Clin. Pharmacol. Ther.* 83, 273–280. doi: 10.1038/sj.clpt.6100275

Sipeky, C., Weber, A., Szabo, M., Melegh, B. I., Janicsek, I., Tarlos, G., et al. (2013). High prevalence of CYP2C19*2 allele in Roma samples: study on Roma and Hungarian population samples with review of the literature. *Mol. Biol. Rep.* 40, 4727–4735. doi: 10.1007/s11033-013-2569-4

Sirisena, N. D., Adeyemo, A., Kuruppu, A. I., Samaranayake, N., and Dissanayake, V. H. W. (2018). Genetic variants associated with clinicopathological profiles in sporadic breast cancer in Sri Lankan women. *J. Breast Cancer* 21, 165–172. doi: 10.4048/jbc.2018.21.2.165

Sirugo, G., Williams, S. M., and Tishkoff, S. A. (2019). The missing diversity in human genetic studies. *Cell* 177, 26–31. doi: 10.1016/j.cell.2019.02.048

Škarić-Jurić, T., Tomas, Ž, Petranović, M. Z., Božina, N., Narančić, N. S., Janićijević, B., et al. (2018). Characterization of ADME genes variation in Roma and 20 populations worldwide. *PLoS One* 13:e0207671. doi: 10.1371/journal.pone.0207671

Slavin, T. P., Maxwell, K. N., Lilyquist, J., Vijai, J., Neuhausen, S. L., Hart, S. N., et al. (2017). The contribution of pathogenic variants in breast cancer susceptibility genes to familial breast cancer risk. *npj Breast Cancer* 3:22.

Steponaitiene, R., Kupcinskas, J., Survilaite, S., Varkalaite, G., Jonaitis, L., Kiudelis, G., et al. (2016). TPMT and ITPA genetic variants in Lithuanian inflammatory bowel disease patients: prevalence and azathioprine-related side effects. *Adv. Med. Sci.* 61, 135–140. doi: 10.1016/j.advms.2015.09.008

Stewart, C., and Pepper, M. S. (2017). Cystic fibrosis in the African diaspora. *Ann. Am. Thorac. Soc.* 14, 1–7. doi: 10.1513/annalsats.201606-481fr

Stocco, G., Yang, W., Crews, K. R., Thierfelder, W. E., Decorti, G., Londero, M., et al. (2012). PACSIN2 polymorphism influences TPMT activity and mercaptopurine-related gastrointestinal toxicity. *Hum. Mol. Genet.* 21, 4793–4804. doi: 10.1093/hmg/dds302

Sundelin, E. I. O., Gormsen, L. C., Jensen, J. B., Vendelbo, M. H., Jakobsen, S., Munk, O. L., et al. (2017). Genetic polymorphisms in organic cation transporter 1 attenuates hepatic metformin exposure in humans. *Clin. Pharmacol. Ther.* 102, 841–848. doi: 10.1002/cpt.701

Takkenberg, J. J. M., and Roos-hesselink, J. W. (2011). Birth prevalence of congenital heart disease worldwide a systematic review and meta-analysis. *JAC* 58, 2241–2247.

Teixeira, J., Amorim, A., Prata, J. M., and Quental, S. (2015). Pharmacogenetic polymorphisms in a portuguese gypsy population. *Curr. Pharmacogenomics Person. Med.* 13, 36–40. doi: 10.2174/1875692113666150703180101

Tomita-Mitchell, A., Maslen, C. L., Morris, C. D., Garg, V., and Goldmuntz, E. (2007). GATA4 sequence variants in patients with congenital heart disease. *J. Med. Genet.* 44, 779–783. doi: 10.1136/jmg.2007.052183

Tzvetkov, M. V., Saadatmand, A. R., Lötsch, J., Tegeder, I., Stingl, J. C., and Brockmöller, J. (2011). Genetically polymorphic OCT1: another piece in the puzzle of the variable pharmacokinetics and pharmacodynamics of the opioidergic drug tramadol. *Clin. Pharmacol. Ther.* 90, 143–150. doi: 10.1038/clpt.2011.56

Visscher, H., Ross, C. J. D., Rassekh, S. R., Barhdadi, A., Dubé, M. P., Al-Saloos, H., et al. (2012). Pharmacogenomic prediction of anthracycline-induced cardiotoxicity in children. *J. Clin. Oncol.* 30, 1422–1428.

Vozarova De Courten, B., De Courten, M., Hanson, R. L., Zahorakova, A., Egyenes, H. P., Tataranni, P. A., et al. (2003). Higher prevalence of type 2 diabetes, metabolic syndrome and cardiovascular diseases in gypsies than in non-gypsies in Slovakia. *Diabetes Res. Clin. Pract.* 62, 95–103. doi: 10.1016/s0168-8227(03)00162-1

Werissa, N. A., Piko, P., Fiatal, S., Kosa, Z., Sandor, J., and Adany, R. (2019). SNP-based genetic risk score modeling suggests no increased genetic susceptibility of the roma population to Type 2 diabetes mellitus. *Genes* 10, 1–16.

Whirl-Carrillo, M., McDonagh, E. M., Hebert, J. M., Gong, L., Sangkuhl, K., Thorn, C. F., et al. (2012). Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* 92, 414–417.

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082.

Wray, N. R., Goddard, M. E., and Visscher, P. M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* 17, 1520–1528. doi: 10.1101/gr.6665407

Yoon, H., Cho, H. Y., Yoo, H. D., Kim, S. M., and Lee, Y. B. (2013). Influences of organic cation transporter polymorphisms on the population pharmacokinetics of metformin in healthy subjects. *AAPS J.* 15, 571–580. doi: 10.1208/s12248-013-9460-z

Zhang, C., Dong, S., Xu, J., He, W., and Yang, T. (2018). PopLDdecay?: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788. doi: 10.1093/bioinformatics/bty875

Zhang, W., Li, X., Shen, A., Jiao, W., Guan, X., and Li, Z. (2008). GATA4 mutations in 486 Chinese patients with congenital heart disease. *Eur. J. Med. Genet.* 51, 527–535. doi: 10.1016/j.ejmg.2008.06.005

Živković, T. B., Marjanović, M., Prgomelja, S., Soldatović, I., Koprivica, B., Acković, D., et al. (2010). Screening for diabetes among roma people living in Serbia. *Croat Med. J.* 51, 144–150. doi: 10.3325/cmj.2010.51.144

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 13 August 2021
doi: 10.3389/fgene.2021.687335

Associations Between *CYP17A1* and *SERPINA6/A1* Polymorphisms, and Cardiometabolic Risk Factors in Black South Africans

Siphiwe N. Dlamini[1,2], Ananyo Choudhury[3], Michèle Ramsay[3], Lisa K. Micklesfield[1], Shane A. Norris[1], Nigel J. Crowther[4], Andrew A. Crawford[5,6], Brian R. Walker[6,7], Zané Lombard[8] and Julia H. Goedecke[1,2]**

[1]South African Medical Research Council/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [2]Non-communicable Diseases Research Unit, South African Medical Research Council, Cape Town, South Africa, [3]Faculty of Health Sciences, Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa, [4]Department of Chemical Pathology, National Health Laboratory Service, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [5]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom, [6]BHF Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, United Kingdom, [7]Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, United Kingdom, [8]Division of Human Genetics, National Health Laboratory Service, Faculty of Health Sciences, School of Pathology, University of the Witwatersrand, Johannesburg, South Africa

OPEN ACCESS

Edited by:
Segun Fatumo,
University of London, United Kingdom

Reviewed by:
Enrique Lin Shiao,
University of California, Berkeley, United States
Xuanye Cao,
Baylor College of Medicine, United States

***Correspondence:**
Siphiwe N. Dlamini
siphiwe208@live.com
Julia H. Goedecke
julia.goedecke@mrc.ac.za

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: 08 April 2021
Accepted: 20 July 2021
Published: 13 August 2021

Citation:
Dlamini SN, Choudhury A, Ramsay M, Micklesfield LK, Norris SA, Crowther NJ, Crawford AA, Walker BR, Lombard Z and Goedecke JH (2021) Associations Between CYP17A1 and SERPINA6/A1 Polymorphisms, and Cardiometabolic Risk Factors in Black South Africans. Front. Genet. 12:687335. doi: 10.3389/fgene.2021.687335

Research in European and Asian populations has reported associations between single nucleotide polymorphisms (SNPs) in *CYP17A1* and *SERPINA6/A1* and circulating glucocorticoid concentrations, and some key cardiometabolic risk factors. This study aimed to investigate these associations in black South African adults, who are disproportionally affected by the metabolic syndrome and its related cardiometabolic risk factors. The dataset included black South African adults (n = 4,431; 56.7% women) from the AWI-Gen study, genotyped on the H3A genotyping array and imputed using the African reference panel at the Sanger imputation service. From the imputed data, 31 *CYP17A1* SNPs and 550 *SERPINA6/A1* SNPs were extracted. The metabolic syndrome and its components were defined using the 2009 harmonized guidelines. Serum glucocorticoid concentrations were measured in a subset of 304 men and 573 women, using a liquid chromatography-mass spectrometry method. Genetic associations were detected using PLINK. Bonferroni correction was used to control for multiple testing. A SNP at *SERPINA6/A1*, rs17090691 (effect allele G), was associated with higher diastolic blood pressure (BP) in all adults combined ($p = 9.47 \times 10^{-6}$). Sex-stratified analyses demonstrated an association between rs1051052 (effect allele G), another *SERPINA6/A1* SNP, and higher high-density lipoprotein (HDL) cholesterol concentrations in women ($p = 1.23 \times 10^{-5}$). No association was observed between these variants and glucocorticoids or between any of the *CYP17A1* SNPs and metabolic outcomes after adjusting for multiple testing. Furthermore, there were no associations between any of the SNPs tested and the

metabolic syndrome. This study reports novel genetic associations between two SNPs at *SERPINA6/A1* and key cardiometabolic risk factors in black South Africans. Future replication and functional studies in larger populations are required to confirm the role of the identified SNPs in the metabolic syndrome and assess if these associations are mediated by circulating glucocorticoids.

Keywords: *SERPINA6*, *SERPINA1*, *CYP17A1*, metabolic syndrome, blood pressure, lipids, insulin, cortisol

INTRODUCTION

Key cardiometabolic risk factors, including elevated waist circumference, blood pressure, and fasting blood glucose concentrations, and the presence of dyslipidemia, cluster to form the metabolic syndrome (Alberti et al., 2009). However, the prevalence of the metabolic syndrome generally differs by both sex and ethnicity, with urban-dwelling black South African women exhibiting a higher prevalence compared to their male counterparts, and women of other ethnic groups in similar urban settings (Goedecke et al., 2017; Gradidge and Crowther, 2017; Gerdts and Regitz-Zagrosek, 2019). The underlying mechanisms that explain these sex- and ethnic-based differences are unclear. Identifying biomarkers and common genetic variants that are associated with the metabolic syndrome and its related cardiometabolic risk factors, has the potential to improve the biological understanding of this condition and consequently the risk for common cardiometabolic diseases such as type 2 diabetes and cardiovascular disease.

Population-based cross-sectional studies have suggested that circulating cortisol, the primary glucocorticoid in humans, is associated with a higher risk of presenting with the metabolic syndrome and its related cardiometabolic risk factors (Walker et al., 2000; Reynolds et al., 2003; Ward et al., 2003). Likewise, single nucleotide polymorphisms (SNPs) within genes involved in glucocorticoid biology have been implicated in key components of the metabolic syndrome (Diver et al., 2016). SNPs in *CYP17A1* are reported to be associated with elevated blood pressure in Asians and Europeans (Levy et al., 2009; Kelly et al., 2013; Franceschini et al., 2016), and with adiposity in East Asians (Hotta et al., 2012). As the *CYP17A1* gene encodes an enzyme that converts precursors for corticosterone to precursors for cortisol (Yadav et al., 2017), the associations between *CYP17A1* SNPs and elevated blood pressure and adiposity are thought to be mediated by glucocorticoids (Diver et al., 2016). Considering that cardiometabolic risk factors that form the metabolic syndrome are interrelated (Alberti et al., 2009), the *CYP17A1* SNPs may also be associated with elevated fasting blood glucose and the presence of dyslipidemia, but this hypothesis is yet to be tested.

Further, in a genome-wide association study (GWAS) in Europeans, inter-individual variation in circulating cortisol concentrations was associated with SNPs at a locus that spans the *SERPINA6* and *SERPINA1* genes (Bolton et al., 2014). *SERPINA6* encodes corticosteroid binding globulin (CBG), the primary glucocorticoid-binding protein in circulation, while *SERPINA1* encodes alpha-1 antitrypsin, which is involved in the regulation of the binding and release of glucocorticoids from CBG (Henley and Lightman, 2011; Crawford et al., 2021). However, associations between SNPs at *SERPINA6/A1* and the metabolic syndrome and related cardiometabolic risk factors remain to be investigated.

The above-mentioned genetic studies were limited to non-Africans and only a few investigated sex-specific associations (Hotta et al., 2012; Bolton et al., 2014; Crawford et al., 2019). Genetic association studies in Africans have the potential to provide more effective localization of the disease-causing gene variants as a result of generally lower linkage disequilibrium in African genomes, compared to studies in populations of European ancestry (Gurdasani et al., 2015).

Hence, the primary aim of this study was to investigate the associations between *CYP171A* and *SERPINA6/A1* SNPs and the metabolic syndrome and its related cardiometabolic risk factors, in black South African men and women. The study also aimed to investigate the hypothesis that the observed genetic associations are mediated by circulating glucocorticoid concentrations.

MATERIALS AND METHODS

Study Population

This study included participants from the Africa Wits-INDEPTH partnership for Genomic Studies (AWI-Gen) project. Briefly, AWI-Gen is a population-based study comprising over 12,000 adult men and women of African ancestry from six research centers in the west, east, and southern parts of Africa (Ramsay et al., 2016). This study was restricted to the AWI-Gen research centers that are based in South Africa: the Dikgale Health and Demographic Surveillance System (HDSS) in the Limpopo Province, MRC/Wits Agincourt HDSS in the Mpumalanga Province near the Mozambican border, and MRC/Wits Developmental Pathways for Health Research Unit (DPHRU) in Soweto (Johannesburg). Most of the female participants from the Soweto DPHRU research center were the caregivers of the Birth to Twenty Plus cohort, which is an ongoing longitudinal study previously described (Richter et al., 2007). Serum samples for glucocorticoid determination were only available from a subset of 877 participants (304 men and 573 women) from the Soweto site. Overall, the present study included 5,268 black South Africans (2,226 men and 3,042 women).

The AWI-Gen data collection was approved by the University of Witwatersrand Human Research Ethics Committee (Medical;

Certificate numbers: M121029 and M170880). Collection of data and serum from the DPHRU sub-samples used for glucocorticoid determination were also approved by the same committee (Certificate numbers: M090620 and M160604). Likewise, the committee approved all secondary analyses described in the present study (Certificate number: M160225).

Participant Testing Procedures

Collection of the AWI-Gen phenotype and related data was described in detail previously (Ali et al., 2018), and only variables relevant to this study are briefly described below.

Anthropometry, Blood Pressure, and Lifestyle Factors

Weight was measured to the nearest 0.1 kg using a digital scale (Kendon Medical, South Africa), and height was measured to the nearest millimeter using a Harpenden digital stadiometer (Holtain, Wales). Weight and height were subsequently used to calculate body mass index (BMI = weight in kg/height in m^2) and obesity was classified as BMI $\geq$ 30 kg/m^2. Waist circumference was measured to the nearest millimeter using a stretch-resistant soft measuring tape (SECA, Hamburg, Germany), at the level of the narrowest part of the torso, halfway between the lowest rib and the iliac crest.

Blood pressure was measured on the right arm using a digital blood pressure reader (Omron M6, Kyoto, Japan) and appropriately sized cuffs. Each participant rested for at least 5 min prior to the blood pressure measurement in a seated position. The blood pressure was measured in triplicate at 2-min intervals, but only the average of the second and third readings were recorded and used in the analyses.

In terms of lifestyle factors, smoking and alcohol status were determined by asking the participants if they smoked any form of tobacco product (e.g., cigarettes, cigars, or pipes), or consumed any form of alcoholic beverage (e.g., beer, wine, spirits, fermented cider, or traditional beer). The participants were subsequently classified as current smokers/non-smokers and drinkers/non-drinkers. The participants were asked to bring all their medications (Meds) to the interview sessions for confirmation and recording of chronic Meds used.

Blood Sampling and Biochemical Analyses

Participants fasted overnight for 10–12 h prior to collection of fasting blood samples. A standard venepuncture technique was used to collect the blood samples for determination of serum biomarkers (fasting glucose, insulin, and lipids), and to obtain buffy coat for DNA extraction. Fasting glucose and lipid concentrations were measured using the Randox Plus chemistry analyzer (Crumlin, Northern Ireland) by colorimetric assays. The intra-assay coefficient of variation (CV) for glucose was 2.3%, while the intra-assay CV for lipids were all less than 1.5%. The concentration of low-density lipoprotein (LDL) cholesterol was subsequently estimated using the Friedewald Equation (Friedewald et al., 1972). Fasting insulin concentrations were determined by an enzyme-linked chemiluminescent immunometric assay on the Immulite 1000 immunoassay system (Siemens, Hamburg, Germany) and the intra-assay CV was less than 2.0%. Homeostatic Model Assessment of Insulin Resistance (HOMA-IR) was subsequently computed from the fasting glucose and insulin concentrations, using the HOMA2-IR calculator version 2.2.3.[1]

Serum corticosterone and cortisol concentrations were determined on the subset of 877 participants from the Soweto DPHRU cohort. Selection of these participants was primarily based on availability of serum samples for glucocorticoid determination. The time of fasting was recorded and ranged between 7:30 am and 11:59 am. A liquid-liquid extraction technique (Agilent Technologies, California, United States) was used to extract the glucocorticoids. The glucocorticoid concentrations were subsequently analyzed by ultra-high-pressure liquid chromatography-mass spectrometry (UPLC-MS; Waters Corp, Milford, United States). The inter and intra-assay CVs were 7.3 and 2.9% for corticosterone, and 13.6 and 9.6% for cortisol, respectively.

Definition of the Metabolic Syndrome

Presence of the metabolic syndrome was based on the 2009 harmonized criteria (Alberti et al., 2009). Briefly, the following cut off points were used to define components of the metabolic syndrome: (i) elevated waist circumference ($\geq$94 cm in men and $\geq$80 cm in women); (ii) elevated triglycerides ($\geq$1.7 mmol/L); (iii) low high-density lipoprotein (HDL) cholesterol (<1.0 mmol/L in men and <1.3 mmol/L in women); (iv) elevated blood pressure ($\geq$130 mmHg for systolic and/or $\geq$85 mmHg for diastolic and/or using blood pressure medication); and (v) elevated glucose ($\geq$5.6 mmol/L and/or using diabetes medication). Participants with three or more of the above components were classified as metabolic syndrome cases.

Genotyping, Imputation, and Data Quality Control

DNA was extracted from buffy coats using a modified salting-out method (Miller et al., 1988) or the QIAsymphony SP (QIAGEN GmbH, QIAGEN Strasse 1, 40724 Hilden, Germany). The Human Heredity and Health in Africa (H3Africa) genotyping array (Illumina Inc., California, United States) was used to perform genome-wide genotyping of over 2.3 million SNPs in all AWI-Gen participants (Illumina FastTrack service). The H3Africa genotyping array was specifically designed to account for the larger genetic diversity and smaller haplotype blocks observed in the genomes of populations of African ancestry. Pre-imputation quality control (QC) steps included removal of participants and SNPs with high data missingness (>5%), SNPs with low minor allele frequency (MAF; <1%) and those not in Hardy-Weinberg Equilibrium (HWE; $p < 1 \times 10^{-4}$). Mitochondrial, non-autosomal, and ambiguous SNPs that failed to match the *GRCh37* reference alleles and strands were also removed.

The Sanger Imputation Server (McCarthy et al., 2016) was used to perform imputation on the cleaned dataset,

[1]https://www.dtu.ox.ac.uk/homacalculator

which comprised of 1,729,661 SNPs and 10,903 participants from all AWI-Gen research centers. The African Genome Resources was used as the reference panel, EAGLE2 (Loh et al., 2016) for pre-phasing, and the default positional Burrows-Wheeler transform algorithm for the imputation (Rubinacci et al., 2020). Post-imputation QC steps included the removal of poorly imputed SNPs (info score < 0.6), SNPs with low MAF (<1%), and SNPs failing the HWE ($p < 1 \times 10^{-4}$). The final cleaned and imputed AWI-Gen dataset had 13.98 million SNPs.

From the imputed dataset, SNPs in the *CYP17A1* gene (all introns and exons) and in the 2 kb region upstream of the transcription start site of *CYP17A1* were included (*GRCh37* coordinates, 10:104590288–104599495), as SNPs in this region were previously associated with glucocorticoid concentrations and the metabolic syndrome in Europeans (Diver et al., 2016; MacKenzie et al., 2019). Secondly, all SNPs in the region from the start of transcription for the *SERPINA6* gene to the end of transcription for the *SERPINA1* gene including the region in between the two genes were also selected (*GRCh37* coordinates, 14:94769460–94857029). The intergenic region between *SERPINA6* and *SERPINA1* was included as it is known to harbor SNPs associated with circulating cortisol concentrations (Bolton et al., 2014). After removing the non-South African samples from the dataset, related participants were also removed (PI_HAT > 0.1875; $n = 758$ participants removed). Due to previously reported associations between glucocorticoid metabolism and critical illness (Boonen et al., 2013), participants previously diagnosed with cancer were also removed from the dataset ($n = 79$ participants removed). The final selected dataset used in the present analyses included 581 SNPs (31 from the *CYP17A1* and 550 from the *SERPINA6/A1* loci) in 4,431 black South Africans (1,918 men and 2,513 women).

Statistical Methods

The Shapiro-Wilk test and histogram plots were used to assess the distribution of continuous variables in STATA v.13.1 software (Stata Corp. LLC, Texas, United States). Non-normally distributed outcome variables were first mathematically transformed (using either log or square root transformations) to obtain normality prior to inclusion in the linear regression models. The Mann-Whitney U and Chi-square tests were used to test statistical differences between men and women in the continuous and categorical variables, respectively.

All genetic association tests were performed using PLINK v.1.9 software (Purcell et al., 2007). Associations between the selected SNPs and the metabolic syndrome and its related cardiometabolic risk factors were tested using logistic and linear regression models, respectively. Similarly, linear regression models were also used to examine the associations between all SNPs and serum glucocorticoid concentrations in the subset of Soweto participants. Age, sex, smoking, alcohol consumption, and BMI (except where BMI, waist circumference, or the metabolic syndrome were the outcome variables) were included as potential confounders in all regression models. Due to the diurnal nature of glucocorticoids, blood sampling time was also included as an additional covariate in models, where serum glucocorticoid concentrations were the outcome variables. To minimize confounding due to prescribed medication, participants taking prescribed diabetes, blood pressure, or lipid medications (see **Table 1** for numbers) were excluded from linear regression analyses involving glucose, insulin, blood pressure, or lipids, respectively, as outcomes.

Bonferroni correction was used to control for multiple testing in all genetic association tests. The number of independent SNPs were determined with the "indep-pairwise" method in PLINK using the following thresholds: window size = 50 kilobase pairs, variance inflation factor = 5, and $r^2 = 0.5$. The number of non-correlated outcomes (rho < 0.8) were determined using a Spearman's correlation test. Hence, the Bonferroni adjusted *p*-value threshold was determined as $p = 0.05/219$ (number of independent SNPs)/10 (number of non-correlated outcomes) = 2.28×10^{-5}. Sex interactions were also tested, and the sex interaction *p*-values were recorded as "Sex Int." All models showing evidence of sex interaction (Sex Int < 0.05) were subsequently stratified by sex.

Overall, the outcomes investigated in this study included the metabolic syndrome and its individual components (waist circumference, systolic and diastolic blood pressure, fasting glucose, triglycerides, and HDL), and other related cardiometabolic risk factors (BMI, LDL, and total cholesterol, insulin, HOMA-IR), and circulating glucocorticoids (corticosterone, cortisol, and the corticosterone/cortisol ratio). Only models that showed evidence of association after adjusting for multiple testing ($p < 2.28 \times 10^{-5}$) and models that showed evidence of sex interactions (Sex Int < 0.05) are presented in the results.

RESULTS

Study Sample Characteristics

The characteristics of the overall AWI-Gen black South African participants (main study sample) are presented and compared by sex in **Table 1**. The characteristics of the Soweto sub-sample on whom circulating glucocorticoids were measured, are compared to the main study sample in the **Supplementary Table S1**. Age, waist circumference, and alcohol consumption were not different; however, BMI and prevalence of smoking was higher in the glucocorticoid sub-sample compared to the main study sample. A comparison of the prevalence of obesity and the categorized individual components of the metabolic syndrome in the main study sample is presented in **Figure 1**.

There was no significant age difference between men and women in the main study sample. However, women had higher BMI and waist circumference compared to men ($p < 0.001$ for both). Accordingly, a greater proportion of women were obese and had elevated waist circumference compared to men ($p < 0.001$ for both). In contrast, men had higher diastolic blood pressure compared to women ($p = 0.027$), and this was accompanied by a greater proportion of men with elevated diastolic blood pressure compared to

TABLE 1 | Characteristics of the study sample of black South Africans and a comparison between sexes.

Study sample	Main genetic association study sample				
	N	All	Men	Women	*p*
Age (years)	4,431	52 (46–58)	52 (46–58)	52 (46–58)	0.185
BMI (kg/m^2)	4,147	27.2 (22.2–32.7)	23.2 (20.1–27.4)	30.5 (25.8–35.5)	<0.001
Waist circumference (cm)	4,148	95.8 (89.4–101.5)	88.0 (78.0–89.8)	98.0 (92.3–103.2)	<0.001
Systolic blood pressure (mmHg)	3,496	128.5 (116.0–144.0)	129.0 (116.5–144.0)	128.0 (115.5–144.0)	0.424
Diastolic blood pressure (mmHg)	3,499	83.0 (74.5–91.5)	84.0 (75.0–92.5)	82.5 (74.0–91.5)	0.027
Fasting glucose (mmol/L)	3,406	4.9 (4.6–5.4)	4.8 (4.4–5.3)	5.0 (4.6–5.5)	<0.001
Fasting insulin (uIU/ml)	3,417	7.2 (4.3–13.1)	7.0 (4.3–13.3)	7.4 (4.3–12.9)	0.327
HOMA-IR	2,985	1.6 (0.9–3.1)	1.6 (0.9–3.2)	1.7 (0.9–3.0)	0.768
Total cholesterol (mmol/L)	4,345	4.1 (3.5–4.9)	4.0 (3.3–4.7)	4.3 (3.6–5.0)	<0.001
LDL cholesterol (mmol/L)	4,309	2.5 (1.9–3.1)	2.3 (1.8–2.9)	2.6 (2.0–3.2)	<0.001
HDL cholesterol (mmol/L)	4,347	1.2 (1.0–1.4)	1.1 (0.9–1.4)	1.2 (1.0–1.4)	0.111
Triglycerides (mmol/L)	4,344	0.9 (0.7–1.3)	0.9 (0.6–1.3)	0.9 (0.7–1.3)	0.003
Blood pressure med [n/N (%)]	3,450	991/3,450 (28.7)	329/1,350 (24.3)	662/2,100 (31.5)	0.869
Diabetes med [n/N (%)]	3,450	366/3,450 (10.6)	134/1,362 (9.8)	232/2,088 (11.1)	0.259
Cholesterol med [n/N (%)]	2030	8/2,030 (0.4)	2/1,337 (0.1)	6/693 (0.9)	0.211
Smoking [n/N (%)]	3,108	597/3,108 (19.2)	538/1,078 (49.9)	59/2,030 (2.9)	<0.001
Alcohol [n/N (%)]	2,502	1,292/2,502 (51.6)	954/1,168 (81.7)	338/1,334 (25.3)	<0.001
*Corticosterone (nmol/L)	650	6.0 (3.5–12.2)	5.0 (3.4–8.8)	6.4 (3.6–14.6)	0.003
*Cortisol (nmol/L)	673	183.3 (96.6–322.6)	143.0 (74.1–295.7)	208.3 (84.2–343.8)	0.003
*Corticosterone/cortisol	548	0.05 (0.02–0.09)	0.04 (0.02–0.07)	0.05 (0.02–0.09)	0.050

Continuous data presented as median (IQR, interquartile range) and categorical data presented as number of prevalent/total number of observations (n/N) and percentage (%). Wilcoxon Rank Sum and Chi-square test were used to statistically compare the continuous and categorical variables, respectively, between the sex groups. p, value of p for the statistical difference between men and women; BMI, body mass index; HOMA-IR, homeostatic model assessment of insulin resistance; LDL, low density lipoprotein; HDL, high density lipoprotein; and Med, medication. All medication and lifestyle factors (smoking and alcohol) were self-reported during interview sessions.
**Measured in the Soweto sub-sample.*

women (p = 0.013). There were no differences in systolic blood pressure, the use of blood pressure medication and the prevalence of overall elevated blood pressure (which considered both systolic and diastolic blood pressures and medication) between the sexes.

Women had higher fasting glucose ($p < 0.001$) and a greater proportion of the women had elevated glucose ($p < 0.020$) compared to men. In contrast, there were no differences in fasting insulin, HOMA2-IR, or the use of diabetes medication, between men and women. For circulating lipids, women had higher total and LDL cholesterol ($p < 0.001$ for both), and triglycerides (p = 0.003), and a greater proportion of women had low serum HDL cholesterol ($p < 0.001$), than men. There were no sex differences in elevated serum triglycerides or the use of lipid-lowering medication.

Corresponding to the above observations, a greater proportion of women were classified as having the metabolic syndrome compared to men (58.4 vs. 37.6%; $p < 0.001$). Conversely, a greater proportion of men smoked and consumed alcohol ($p < 0.001$ for both). In the sub-sample, women had higher corticosterone and cortisol concentrations compared to men (p = 0.003 for both). There was also a tendency for higher corticosterone/cortisol ratio in women compared to men (p = 0.050).

Genetic Associations in Black South African Adults

Although, none of the SNPs tested were associated with odds of presenting with the metabolic syndrome (**Supplementary Table S2**), rs17090691-G in the *SERPINA6/A1* region, showed an association with higher diastolic blood pressure in the sample comprising both sexes (**Table 2**). **Figure 2A** shows the LocusZoom summarizing association signals and LD in the surrounding region. The box and whisker plot that compares diastolic blood pressure across the rs17090691 genotype groups is shown in **Figure 2B**.

In addition, there was evidence for sex interaction between five SNPs at *CYP17A1* and fasting insulin and HOMA-IR, as well as two *CYP17A1* SNPs and fasting cortisol. Likewise, there was evidence of a sex interaction for six SNPs at *SERPINA6/A1* and HDL cholesterol. Hence, the study sample was stratified by sex and the regression models that showed sex-interactions were repeated for men and women separately. These sex-stratified models are shown in **Figure 3**.

The associations between rs10883783-A, rs743575-G, rs4919687-A, rs10883784-T, and rs10786714-C at *CYP17A1*, and lower insulin and HOMA-IR were observed in men but not in women (**Figures 3A,B**). On the other hand, the associations between rs115978957-G and rs116032963-A at *CYP17A1*, and higher cortisol were observed in women but not in men (**Figure 3C**). Similarly, a set of SNPs, in *SERPINA6/A1*, rs60643124-G, rs12101216-T, rs74074941-T, rs74074947-T, and rs58460454-T showed association with lower HDL cholesterol in women only (**Figure 3D**). However, these signals did not remain significant after correction for multiple testing (all $p > 2.28 \times 10^{-5}$). Instead, the only association that remained significant even after adjusting for multiple testing was the women-specific association of rs1051052-G (*SERPINA6/A1* region) with higher HDL cholesterol ($p = 1.23 \times 10^{-5}$; **Figure 3D**). **Figure 3E** summarizes the LD architecture and association

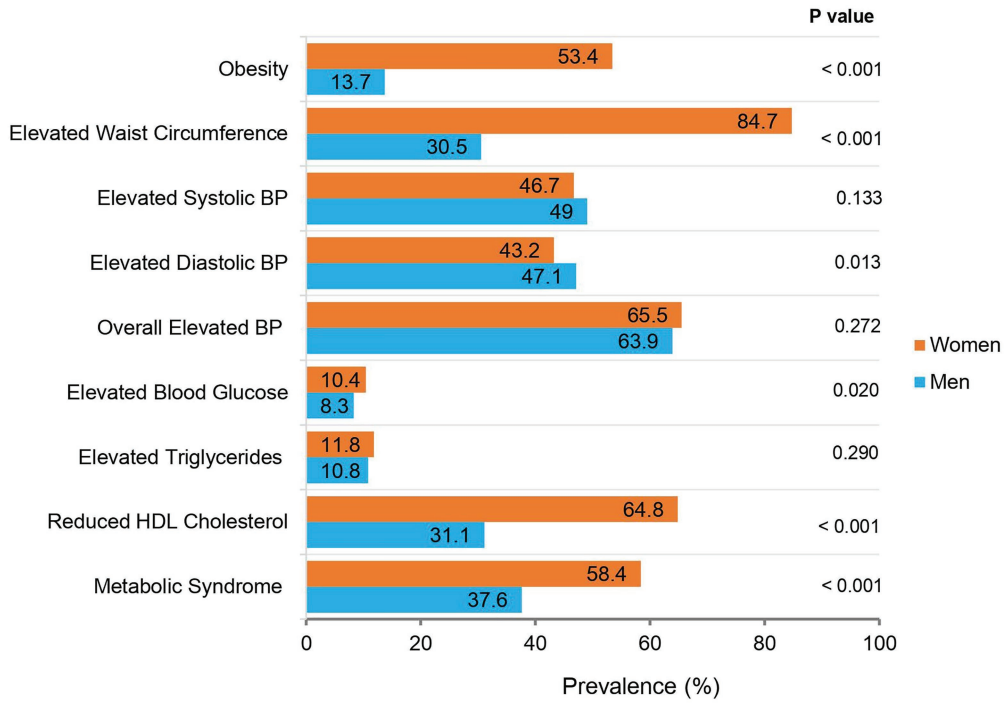

FIGURE 1 | The prevalence of obesity, the metabolic syndrome and its components in black South African men and women (n = 4,431). p: Value of p for the statistical difference between men and women, using a chi-square test. BP, blood pressure. Obesity: Body mass index (BMI) ≥ 30 kg/m^2; Elevated Waist Circumference: ≥94 cm in men and ≥80 cm in women; Elevated Systolic BP: ≥130 mmHg; Elevated Diastolic BP: ≥85 mmHg; Overall Elevated BP: Elevated Systolic BP and/or Diastolic BP and/or using BP medication; Elevated Blood Glucose: Elevated fasting glucose (≥5.6 mmol/L) and/or using diabetes medication; Elevated Triglycerides: Elevated fasting triglycerides (≥1.7 mmol/L); Reduced high-density lipoprotein (HDL) Cholesterol: Reduced fasting HDL-cholesterol (<1.0 mmol/L in men, <1.3 mmol/L in women); Metabolic Syndrome: 2009 Harmonized Criteria (Alberti et al., 2009).

scores in the genomic region. The box and whisker plot (**Figure 3F**) shows the variation of HDL cholesterol between the three rs1051052 genotype groups.

Frequencies of the Identified Effect Alleles

In summary, only the relationship between rs17090691-G allele and higher diastolic blood pressure in the sample comprising both men and women, as well as between rs1051052-G allele and serum HDL cholesterol in women only, provided sufficient evidence of association after adjusting for multiple testing. The allele frequencies for these two SNPs (rs17090691 and rs1051052) from the *SERPINA6/A1* region were compared to those of other global populations (Americans, Europeans, East Asians, and South Asians) in **Figure 4**. The frequencies for the two identified effect alleles, rs17090691-G and rs1051052-G were similar to other African populations (3 vs. 5% and 21 vs. 20%, respectively), but lower than non-African populations (10–32 and 42–66%, respectively).

DISCUSSION

The present study reports two novel genetic associations between SNPs at *SERPINA6/A1* and key cardiometabolic risk factors in black South Africans. Specifically, the rs17090691-G allele was associated with higher diastolic blood pressure in a combined sample of men and women, while the rs1051052-G allele was associated with higher HDL cholesterol in women only. There were potential sex-specific associations between minor alleles of *CYP17A1* SNPs and lower insulin and HOMA-IR in men only, as well as between other *CYP17A1* SNPs and higher cortisol concentrations in women only. However, the evidence of these associations was not sufficient after adjusting for multiple testing. Further, there was no evidence to show that any of the identified genetic associations with diastolic blood pressure and HDL cholesterol were mediated by circulating glucocorticoids in black South Africans (**Supplementary Table S2**).

The associations between *CYP17A1* SNPs and measures of elevated blood pressure and adiposity have been previously reported in non-Africans (Levy et al., 2009; Hotta et al., 2012; Kelly et al., 2013; Franceschini et al., 2016), but *SERPINA6/A1* SNPs have not been previously associated with any cardiometabolic risk factor in any population. Instead, a GWAS in European men and women demonstrated that a SNP at *SERPINA6/A1*, represented by the rs12589136-T allele, was associated with higher circulating cortisol concentrations, while the rs2749527-T and rs11621961-T alleles at the same locus were associated with lower circulating cortisol concentrations (Boonen et al., 2013). Accordingly, a recent European study suggested that these previously identified *SERPINA6/A1* SNPs are likely to influence circulating cortisol

TABLE 2 | Associations of *CYP17A1* and *SERPINA6/A1* SNPs in black South African men and women.

Phenotype	Locus	SNP	Minor (Effect) allele	Linear regression model adjusted for confounders		
				Beta (95% CI)	*p*	Sex Int
Fasting insulin	*CYP17A1*	rs10883783	A	−0.073 (−0.141, −0.005)	0.035	0.011
		rs743575	G	−0.072 (−0.140, −0.005)	0.036	0.009
		rs4919687	A	−0.072 (−0.139, −0.004)	0.037	0.010
		rs10883784	T	−0.067 (−0.134, 0.001)	0.053	0.006
		rs10786714	C	−0.067 (−0.135, 0.000)	0.051	0.005
HOMA-IR	*CYP17A1*	rs10883783	A	−0.085 (−0.157, −0.013)	0.021	0.033
		rs743575	G	−0.085 (−0.157, −0.014)	0.020	0.029
		rs4919687	A	−0.085 (−0.157, −0.013)	0.020	0.030
		rs10883784	T	−0.080 (−0.152, −0.008)	0.029	0.021
		rs10786714	C	−0.081 (−0.153, −0.009)	0.028	0.020
Fasting cortisol	*CYP17A1*	rs115978957	G	0.132 (0.032, 0.232)	0.010	0.047
		rs116032963	A	0.132 (0.032, 0.232)	0.010	0.047
Diastolic BP	*SERPINA6/A1*	rs17090691	G	0.065 (0.036, 0.094)	9.5×10^{-6}	0.813
HDL cholesterol	*SERPINA6/A1*	rs60643124	G	−0.021 (−0.036, −0.006)	0.008	0.047
		rs12101216	T	−0.018 (−0.032, −0.003)	0.016	0.026
		rs74074941	T	−0.016 (−0.032, −0.001)	0.040	0.013
		rs74074947	T	−0.016 (−0.032, −0.001)	0.040	0.010
		rs58460454	T	−0.017 (−0.032, −0.001)	0.033	0.013
		rs1051052	G	0.016 (0.004, 0.028)	0.008	0.006

Linear regression models adjusted for age, sex, smoking, alcohol, and BMI (except where waist circumference was the outcome); Blood sampling time was included as an additional covariate where cortisol was the outcome. SNP, single nucleotide polymorphism; M, major allele; m, minor allele; Beta, unstandardized beta coefficient for the linear regression model; 95% CI, 95% confident intervals; HOMA-IR, homeostatic model assessment of insulin resistance; HDL, high-density lipoprotein; BP, blood pressure; p, value of p for the unstandardized beta coefficient; Sex Int, value of p for sex interaction.

concentrations by altering hepatic CBG expression (Crawford et al., 2021). Additionally, many cross-sectional studies have suggested that higher circulating cortisol concentrations are associated with an increased risk of developing the metabolic syndrome and its related cardiometabolic risk factors in both African and non-African populations (Walker et al., 2000; Reynolds et al., 2003; Ward et al., 2003). Based on these previous findings, the present study investigated whether common genetic variants at both *CYP17A1* and *SERPINA6/A1* loci are associated with the metabolic syndrome, its related cardiometabolic risk factors, and circulating glucocorticoids, in black South Africans.

To our knowledge, the rs17090691 SNP, which was associated with diastolic blood pressure in the present study, has not been associated with any other human trait. In contrast, the rs1051052-G allele, which was associated with higher HDL cholesterol in women in the present study, has not been associated with any cardiometabolic risk factors, but rather, has been associated with respiratory disorders. Specifically, the rs1051052-G allele was associated with higher odds of having childhood asthma in European girls and boys, but not their African counterparts (Baye et al., 2011), as well as chronic obstructive pulmonary disease in East Asian men and women (Zhong et al., 2018). Although, these traits are not directly related to the metabolic syndrome, similar mechanisms involving circulating glucocorticoid concentrations may also be involved (Baye et al., 2011; Zhong et al., 2018).

Circulating glucocorticoids are known to increase blood pressure *via* several mechanisms, including impairment of nitric oxide-mediated renal vasodilation in the kidneys (De Matteo and May, 1997). Accordingly, higher concentrations of circulating cortisol, the primary glucocorticoid in humans, have been consistently shown to be associated with measures of elevated blood pressure in both men and women, regardless of ethnicity (Filipovský et al., 1996; Phillips et al., 1998; Walker et al., 2000; Reynolds et al., 2003; Ward et al., 2003; Weigensberg et al., 2008; Adam et al., 2010; Park et al., 2011; Tolmay et al., 2012; Guzzetti et al., 2014; Constantinopoulos et al., 2015; Schutte et al., 2016). As the rs17090691 SNP is intergenic between *SERPINA6* and *SERPINA1*, it is likely that this SNP is in LD with the causal variant. While common variants generally contribute a modest amount to complex phenotypes such as diastolic blood pressure, functional studies should still investigate whether having the identified effect allele is associated with expression of CBG and/or alpha-1 antitrypsin, to assess the mechanisms involved. Notably, the identified rs17090691-G allele (effect allele) has a higher frequency in populations of non-African ancestry (10–32 vs. 3–5%). Thus, a replication study that includes non-African populations is required to understand whether the association occurs in other ethnic groups.

The two *SERPINA6/A1* SNPs associated with cardiometabolic outcomes in the present study localize to a separate haplotype block from those previously associated with cortisol in Europeans (Bolton et al., 2014), and there are currently no reported functional studies for these SNPs. Furthermore, we did not find an association between the identified SNPs and circulating glucocorticoid concentrations in the sub-sample of men and women from Soweto. Notably, circulating glucocorticoid concentrations change dynamically and are influenced by several known and unknown confounders (Kelly et al., 2008). For example, circulating glucocorticoid concentrations exhibit

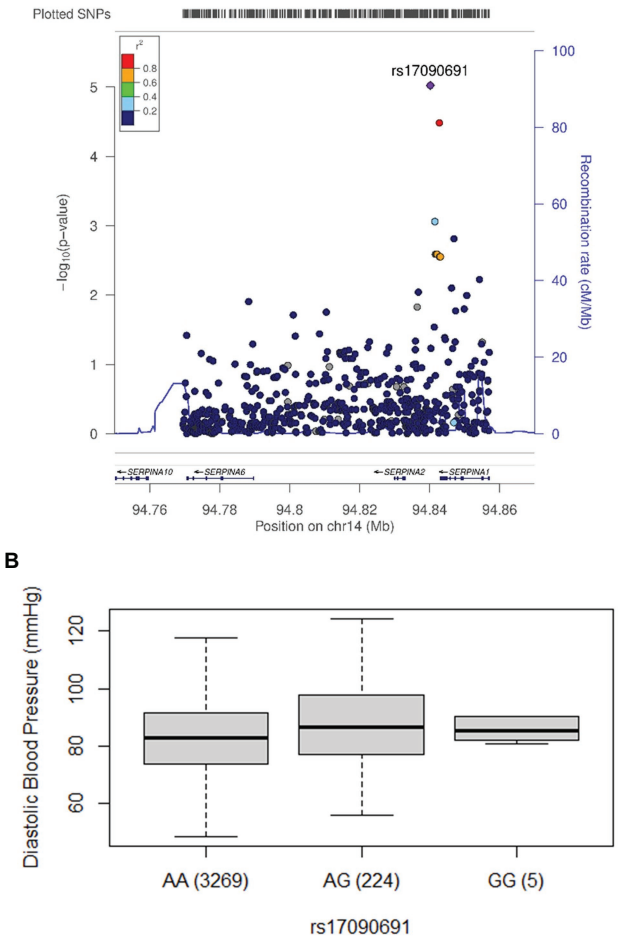

FIGURE 2 | The association between rs17090691 and diastolic blood pressure in black South African men and women. The LocusZoom **(A)** plot association data was plotted using the AFR LD backgrounds from the 1,000 Genomes Project phase 1 dataset. The rs17090691 single nucleotide polymorphism (SNP) is represented by a purple diamond. Each SNP is colored according to the r^2 value against rs17090691. SNPs with missing LD information are colored in gray. The box and whisker plot **(B)** shows medians and interquartile ranges (IQRs) of diastolic blood pressure across the rs17090691 genotype groups. The number of observations in each group is shown in brackets.

a diurnal nature, such that peaks are seen 30–45 min after awakening and the concentrations decline throughout the day (Clow et al., 2010). Morning glucocorticoid concentrations were measured in this study, but not all samples were taken at the same time. Although, blood sampling time was corrected for this may still have influenced the findings. Further, it is also possible that the identified *SERPINA6/A1* SNPs were associated with the variability in daily circulating glucocorticoid concentrations instead of the morning concentrations that were measured in the present study. Furthermore, our available sample size for glucocorticoid determination was relatively small and the lack of associations with glucocorticoids may have resulted from low statistical power. For these reasons, a larger and more detailed study in Africans is required to explore whether circulating glucocorticoids mediated the observed genetic associations.

The rs1051052-G allele, which was associated with higher HDL cholesterol in women in the present study, resides within the 3' untranslated region of *SERPINA1*. However, no functional studies have been undertaken to assess its influence on the genetic expression and function of its protein product, alpha-1 antitrypsin. Hence, the molecular consequences of rs1051052-G allele also remain to be explored. Nevertheless, this is the first study to show that this minor allele is associated with higher HDL cholesterol in women of African ancestry. These findings support the hypothesis that the often-observed sex differences in lipid and lipoprotein metabolism may be partly related to glucocorticoid signaling, which is often not considered as a potential confounder. However, the frequency of this allele is much lower in populations of African ancestry than in non-Africans (20–21 vs. 42–66%). Notably, black South African women typically present with lower HDL cholesterol compared to their European ancestry counterparts (Després et al., 2000; Punyadeera et al., 2001; Sumner and Cowie, 2008; Goedecke et al., 2010; Sliwa et al., 2012). However, results from the present study support previous findings that HDL cholesterol concentrations are similar between African men and women (Kodaman et al., 2016; van der Linden et al., 2019), an observation which is different to that reported in populations of European Ancestry (Balder et al., 2017). While sex and ethnic differences in serum lipid profiles are thought to be driven by differences in central adiposity, in particular differences in visceral adipose tissue (VAT; Keswell et al., 2016), lifestyle factors may also influence the differences in lipid profiles (Goedecke et al., 2010; Keswell et al., 2016; van der Linden et al., 2019). Indeed, HDL cholesterol concentrations are associated with alcohol intake, physical activity, and importantly, inflammation (Choi and Seeger, 2005; Keswell et al., 2016). Hence, exploring gene-environment interactions in African populations is required to understand sex and ethnic differences in serum lipid profiles.

Although, we detected some evidence for the associations between some of the *CYP17A1* SNPs and lower insulin and HOMA-IR in men only, and some of the *CYP17A1* SNPs with cortisol concentrations in women only, these signals were not conclusive as they did not remain significant after stringent Bonferroni corrections. Moreover, these sex-specific associations were possibly confounded by the vast differences in body fat between men and women in this cohort (obesity prevalence = 53.4% in women vs. 13.7% in men). As these sex differences are characteristic of the black South African population (Pillay et al., 2015), a larger sample size would be required to confirm the sex-specific associations between *CYP17A1* SNPs and insulin and HOMA-IR, as well as glucocorticoid concentrations in African men and women, respectively.

The lack of evidence of association between previously identified *CYP17A1* and *SERPINA6/A1* SNPs, and the tested cardiometabolic risk factors in the present study, was likely attributed to lower minor allele frequencies (**Supplementary Table S3**) in combination with the distinct

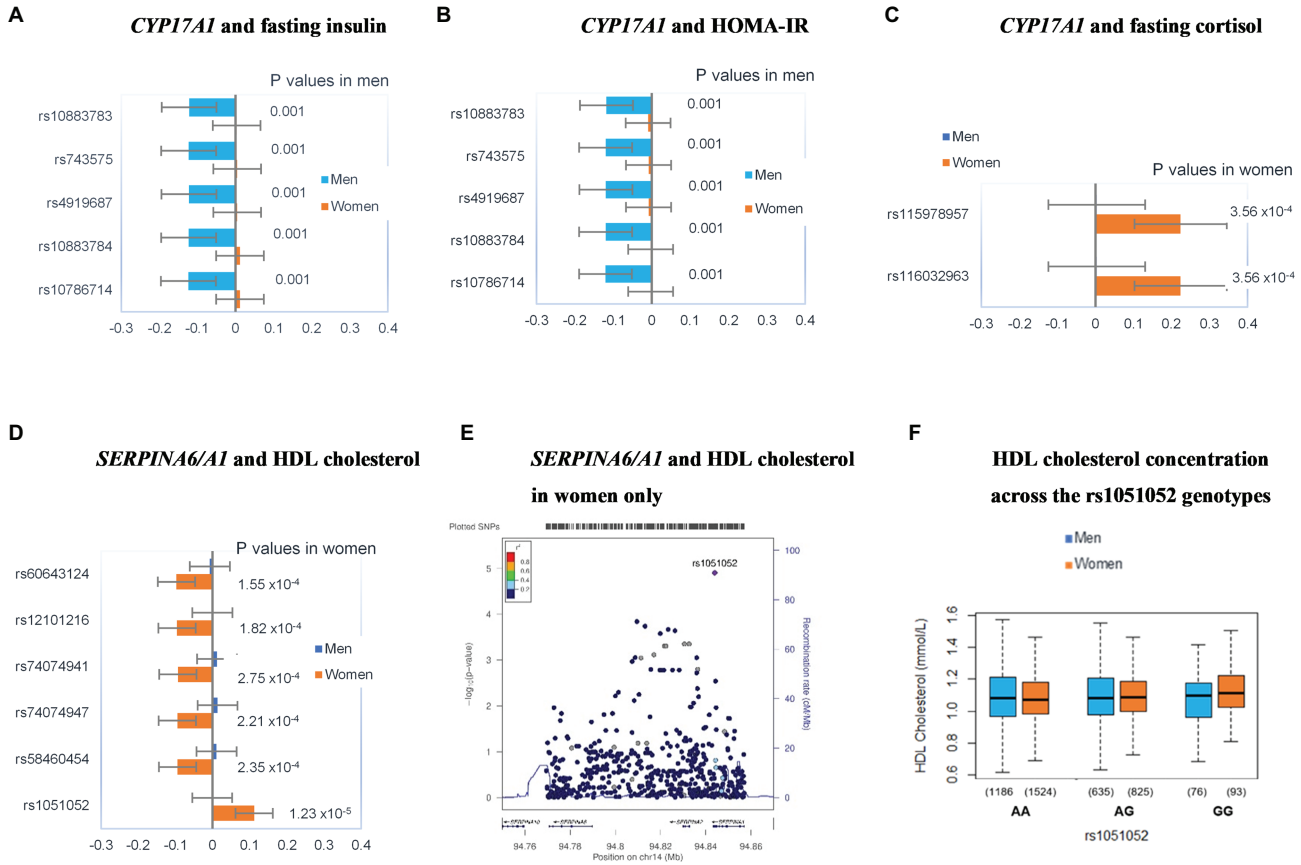

FIGURE 3 | Sex-specific genetic associations between *CYP17A1* **(A-C)** and *SERPINA6/A1* **(D–F)** variants and components of the metabolic syndrome. Plots **(A-D)** show standardized beta coefficients for the linear regression models after adjusting for age, smoking, alcohol, and BMI. Error bars in these plots represent 95% CIs for the standardized beta coefficients. Values of *p* are for the respective beta coefficients. The LocusZoom plot summarizes the association between rs1051052 and higher HDL cholesterol concentrations in women **(E)**. The LocusZoom plot association data was plotted against the AFR LD backgrounds from the 1,000 Genomes Project phase 1 dataset. The rs1051052 SNP is represented by a purple diamond. Each SNP is colored according to the r^2 value against rs1051052. SNPs with missing LD information are colored in gray. The box and whisker plot **(F)** shows the medians and IQRs of HDL cholesterol concentrations across the rs1051052 genotype groups in men and women. The number of observations in each group is shown in brackets.

metabolic profile of the studied population (e.g., almost half of the participants had the metabolic syndrome), compared to non-Africans. Due to these factors, a larger sample size may be required to replicate the previously identified associations. Nevertheless, investigating a black South African population of middle-aged men and women is a major strength of the present study. Considering the possibility that the observed associations may be African-specific, these novel genetic associations may have been previously missed, since most genetic association studies have been conducted in Europeans. A recent review has shown that the composition of participants in previously published GWAS is largely Eurocentric, with approximately 78% European, 10% Asian, and only 2% African (Sirugo et al., 2019). Therefore, large scale studies such as ours are necessary to test the transferability of European cohort-based signals to other ethnic groups and geographic regions. Moreover, the previous genetic association studies in Africans were limited by the use of genotyping arrays that are not designed to capture the diversity observed in the African genome (Pillay et al., 2015; Hendry et al., 2018). The use of an African-centric array and African enriched imputation panel enabled us to capture genetic diversity around these two loci more efficiently. The relatively larger sample size allowed for the inclusion of sufficient men and women participants in the present study for sex specific association testing. Notably, several explanations, including multifactorial models (interactions between biological and environmental factors; Reich et al., 1975), the sex-dependent liability threshold (Carter and Evans, 1969), and sex-disparities in the underlying genetic architecture (Khramtsova et al., 2019), have been proposed to explain sex-specific genetic associations in human phenotypes.

The present study also has some limitations. The statistical power of the study was limited by a moderate sample size at the beginning of the study and the criteria used to select the final participants. Hence, lack of genetic associations with some of the key cardiometabolic risk factors, and with the odds of having the metabolic syndrome, may be attributed to a lack of statistical power. Moreover, we did not adjust for possible

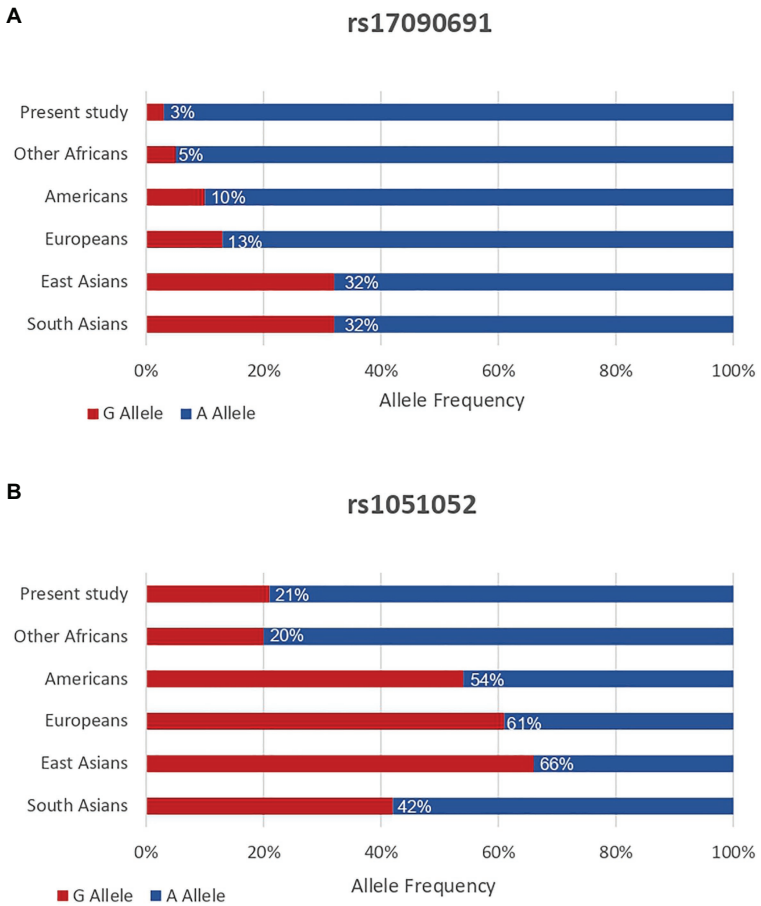

FIGURE 4 | Allele frequencies for *SERPINA6/A1* rs17090691-G **(A)** and rs1051052-G **(B)** variants in the present study sample (n = 4,431) and other global populations. Minor allele frequencies from the other populations were obtained from the 1,000 Genomes Project phase 3 dataset. Other Africans include Yoruba in Ibadan (Nigeria), Luhya in Webuye (Kenya), Gambian in Western Divisions (Gambia), Mende (Sierra Leone), Esan (Nigeria), Americans of African Ancestry in SW (United States), and African Caribbeans (Barbados). Europeans include Utah Residents (CEPH) with Northern and Western European Ancestry, Toscani (Italy), Finnish (Finland), British (England and Scotland), and Iberian Population (Spain). Americans include Mexican Ancestry from Los Angeles (United States), Puerto Ricans (Puerto Rico), Colombians from Medellin (Colombia), and Peruvians from Lima (Peru). East Asians include Han Chinese in Beijing (China), Japanese in Tokyo (Japan), Southern Han Chinese, Chinese Dai in Xishuangbanna (China), and Kinh in Ho Chi Minh City (Vietnam). South Asians include Gujarati Indian from Houston (Texas), Punjabi from Lahore (Pakistan), Bengali (Bangladesh), Sri Lankan Tamil (the United Kingdom), and Indian Telugu (the United Kingdom).

population substructure within the black South African population, which could have influenced some of the genetic association results (Sengupta et al., 2020). Furthermore, circulating glucocorticoid concentrations were only measured in a subset of the study population and the influence of the circadian rhythm could not be assessed. Future studies should also measure the diurnal curves to draw better conclusions regarding mediation by circulating glucocorticoids.

CONCLUSION

The present study reports novel genetic associations between two SNPs at *SERPINA6/A1* and key cardiometabolic risk factors in black South African men and women. These findings support the hypothesis that common variants in the *SERPINA6/A1* locus are associated with key cardiometabolic risk factors in humans. Future functional studies are also required to confirm the role of the identified SNPs in the metabolic syndrome and assess whether these associations are mediated by circulating glucocorticoids.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Witwatersrand Human Research

Ethics Committee (Medical). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SD, JG, ZL, AC, AAC, MR, and BW were responsible for the conception and planning of the study. MR, LM, JG, NC, and SN were involved in sample and data collection. SD was responsible for measuring serum glucocorticoid concentrations. SD, AC, JG, and ZL were involved in the data analyses. SD, AC, MR, LM, SN, NC, AAC, BW, ZL, and JG were involved in the interpretation of the results and the writing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The present study was funded by the Academy of Medical Sciences Newton Advanced Fellowship and the South African National Research Foundation (NRF). The South African Medical Research Council was also acknowledged for funds received from the South African National Department of Health, the UKMRC, the Newton Fund and GSK. The AWI-Gen Collaborative Centre was funded by the National Human Genome Research Institute (NHGRI), Office of the Director (OD), Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD), the National Institute of Environmental Health Sciences (NIEHS), the Office of AIDS research (OAR), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), and the National Institutes of Health (NIH) under award number U54HG006938 and its supplements, as part of the H3Africa Consortium. Additional funding came from the Department of Science and Technology, South Africa, award number DST/CON 0056/2014. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the research funders.

ACKNOWLEDGMENTS

The authors would like to thank all the participants who volunteered to be part of this study. The data collection and expertise contributed by members from the South African AWI-Gen research centers was also acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.687335/full#supplementary-material

REFERENCES

Adam, T. C., Hasson, R. E., Ventura, E. E., Toledo-Corral, C., Le, K. A., Mahurkar, S., et al. (2010). Cortisol is negatively associated with insulin sensitivity in overweight Latino youth. *J. Clin. Endocrinol. Metab.* 95, 4729–4735. doi: 10.1210/jc.2010-0322

Alberti, K. G. M. M., Eckel, R. H., Grundy, S. M., Zimmet, P. Z., Cleeman, J. I., Donato, K. A., et al. (2009). Harmonizing the metabolic syndrome: a joint interim statement of the international diabetes federation task force on epidemiology and prevention; national heart, lung, and blood institute; American heart association; world heart federation; international atherosclerosis society; and international association for the study of obesity. *Circulation* 120, 1640–1645. doi: 10.1161/CIRCULATIONAHA.109.192644

Ali, S. A., Soo, C., Agongo, G., Alberts, M., Amenga-Etego, L., Boua, R. P., et al. (2018). Genomic and environmental risk factors for cardiometabolic diseases in Africa: methods used for phase 1 of the AWI-gen population cross-sectional study. *Glob. Health Action* 11:1507133. doi: 10.1080/16549716.2018.1507133

Balder, J. W., de Vries, J. K., Nolte, I. M., Lansberg, P. J., Kuivenhoven, J. A., and Kamphuisen, P. W. (2017). Lipid and lipoprotein reference values from 133,450 Dutch lifelines participants: age- and gender-specific baseline lipid values and percentiles. *J. Clin. Lipidol.* 11, 1055–1064. doi: 10.1016/j.jacl.2017.05.007

Baye, T. M., Kovacic, M. B., Myers, J. M., Martin, L. J., Lindsey, M., Patterson, T. L., et al. (2011). Differences in candidate gene association between European ancestry and African American asthmatic children. *PLoS One* 6:e16522. doi: 10.1371/journal.pone.0016522

Bolton, J. L., Hayward, C., Direk, N., Lewis, J. G., Hammond, G. L., Hill, L. A., et al. (2014). Genome wide association identifies common variants at the *SERPINA6/SERPINA1* locus influencing plasma cortisol and corticosteroid binding globulin. *PLoS Genet.* 10:e1004474. doi: 10.1371/journal.pgen.1004474

Boonen, E., Vervenne, H., Meersseman, P., Andrew, R., Mortier, L., Declercq, P. E., et al. (2013). Reduced cortisol metabolism during critical illness. *N. Engl. J. Med.* 368, 1477–1488. doi: 10.1056/NEJMoa1214969

Carter, C. O., and Evans, K. A. (1969). Inheritance of congenital pyloric stenosis. *J. Med. Genet.* 6, 233–254. doi: 10.1136/jmg.6.3.233

Choi, H. K., and Seeger, J. D. (2005). Lipid profiles among US elderly with untreated rheumatoid arthritis—the third national health and nutrition examination survey. *J. Rheumatol.* 32, 2311–2316.

Clow, A., Hucklebridge, F., Stalder, T., Evans, P., and Thorn, L. (2010). The cortisol awakening response: more than a measure of HPA axis function. *Neurosci. Biobehav. Rev.* 35, 97–103. doi: 10.1016/j.neubiorev.2009.12.011

Constantinopoulos, P., Michalaki, M., Kottorou, A., Habeos, I., Psyrogiannis, A., Kalfarentzos, F., et al. (2015). Cortisol in tissue and systemic level as a contributing factor to the development of metabolic syndrome in severely obese patients. *Eur. J. Endocrinol.* 172, 69–78. doi: 10.1530/EJE-14-0626

Crawford, A. A., Bankier, S., Altmaier, E., Barnes, C. L. K., Clark, D. W., Ermel, R., et al. (2021). Variation in the *SERPINA6/SERPINA1* locus alters morning plasma cortisol, hepatic corticosteroid binding globulin expression, gene expression in peripheral tissues, and risk of cardiovascular disease. *J. Hum. Genet.* 66, 625–636. doi: 10.1038/s10038-020-00895-6

Crawford, A. A., Soderberg, S., Kirschbaum, C., Murphy, L., Eliasson, M., Ebrahim, S., et al. (2019). Morning plasma cortisol as a cardiovascular risk factor: findings from prospective cohort and Mendelian randomization studies. *Eur. J. Endocrinol.* 181, 429–438. doi: 10.1530/EJE-19-0161

De Matteo, R., and May, C. N. (1997). Glucocorticoid-induced renal vasodilatation is mediated by a direct renal action involving nitric oxide. *Am. J. Physiol.* 273, R1972–R1979. doi: 10.1152/ajpregu.1997.273.6.R1972

Després, J. P., Couillard, C., Gagnon, J., Bergeron, J., Leon, A. S., Rao, D. C., et al. (2000). Race, visceral adipose tissue, plasma lipids, and lipoprotein lipase activity in men and women: the health, risk factors, exercise training, and genetics (HERITAGE) family study. *Arterioscler. Thromb. Vasc. Biol.* 20, 1932–1938. doi: 10.1161/01.ATV.20.8.1932

Diver, L. A., MacKenzie, S. M., Fraser, R., McManus, F., Freel, E. M., Alvarez-Madrazo, S., et al. (2016). Common polymorphisms at the *CYP17A1* locus associate with steroid phenotype: support for blood pressure genome-wide association study signals at this locus. *Hypertension* 67, 724–732. doi: 10.1161/HYPERTENSIONAHA.115.06925

Filipovský, J., Ducimetière, P., Eschwège, E., Richard, J. L., Rosselin, G., and Claude, J. R. (1996). The relationship of blood pressure with glucose, insulin,

heart rate, free fatty acids and plasma cortisol levels according to degree of obesity in middle-aged men. *J. Hypertens.* 14, 229–235. doi: 10.1097/00004872-199602000-00012

Franceschini, N., Carty, C. L., Lu, Y., Tao, R., Sung, Y. J., Manichaikul, A., et al. (2016). Variant discovery and fine mapping of genetic loci associated with blood pressure traits in Hispanics and African Americans. *PLoS One* 11:e0164132. doi: 10.1371/journal.pone.0164132

Friedewald, W. T., Levy, R. I., and Fredrickson, D. S. (1972). Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* 18, 499–502. doi: 10.1093/clinchem/18.6.499

Gerdts, E., and Regitz-Zagrosek, V. (2019). Sex differences in cardiometabolic disorders. *Nat. Med.* 25, 1657–1666. doi: 10.1038/s41591-019-0643-8

Goedecke, J. H., Mtintsilana, A., Dlamini, S. N., and Kengne, A. P. (2017). Type 2 diabetes mellitus in African women. *Diabetes Res. Clin. Pract.* 123, 87–96. doi: 10.1016/j.diabres.2016.11.017

Goedecke, J. H., Utzschneider, K., Faulenbach, M. V., Rizzo, M., Berneis, K., Spinas, G. A., et al. (2010). Ethnic differences in serum lipoproteins and their determinants in south African women. *Metabolism* 59, 1341–1350. doi: 10.1016/j.metabol.2009.12.018

Gradidge, P. J. L., and Crowther, N. J. (2017). Review: metabolic syndrome in black south African women. *Ethn. Dis.* 27, 189–200. doi: 10.18865/ed.27.2.189

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., et al. (2015). The African genome variation project shapes medical genetics in Africa. *Nature* 517, 327–332. doi: 10.1038/nature13997

Guzzetti, C., Pilia, S., Ibba, A., and Loche, S. (2014). Correlation between cortisol and components of the metabolic syndrome in obese children and adolescents. *J. Endocrinol. Invest.* 37, 51–56. doi: 10.1007/s40618-013-0014-0

Hendry, L. M., Sahibdeen, V., Choudhury, A., Norris, S. A., Ramsay, M., and Lombard, Z. (2018). Insights into the genetics of blood pressure in black south African individuals: the birth to twenty cohort. *BMC Med. Genomics* 11:2. doi: 10.1186/s12920-018-0321-6

Henley, D. E., and Lightman, S. L. (2011). New insights into corticosteroid-binding globulin and glucocorticoid delivery. *Neuroscience* 180, 1–8. doi: 10.1016/j.neuroscience.2011.02.053

Hotta, K., Kitamoto, A., Kitamoto, T., Mizusawa, S., Teranishi, H., Matsuo, T., et al. (2012). Genetic variations in the *CYP17A1* and *NT5C2* genes are associated with a reduction in visceral and subcutaneous fat areas in Japanese women. *J. Hum. Genet.* 57, 46–51. doi: 10.1038/jhg.2011.127

Kelly, T. N., Takeuchi, F., Tabara, Y., Edwards, T. L., Kim, Y. J., Chen, P., et al. (2013). Genome-wide association study meta-analysis reveals transethnic replication of mean arterial and pulse pressure loci. *Hypertension* 62, 853–859. doi: 10.1161/HYPERTENSIONAHA.113.01148

Kelly, S. J., Young, R., Sweeting, H., Fischer, J. E., and West, P. (2008). Levels and confounders of morning cortisol collected from adolescents in a naturalistic (school) setting. *Psychoneuroendocrinology* 33, 1257–1268. doi: 10.1016/j.psyneuen.2008.06.010

Keswell, D., Tootla, M., and Goedecke, J. H. (2016). Associations between body fat distribution, insulin resistance and dyslipidaemia in black and white south African women. *Cardiovasc. J. Afr.* 27, 177–183. doi: 10.5830/CVJA-2015-088

Khramtsova, E. A., Davis, L. K., and Stranger, B. E. (2019). The role of sex in the genomics of human complex traits. *Nat. Rev. Genet.* 20, 173–190. doi: 10.1038/s41576-018-0083-1

Kodaman, N., Aldrich, M. C., Sobota, R., Asselbergs, F. W., Poku, K. A., Brown, N. J., et al. (2016). Cardiovascular disease risk factors in Ghana during the rural-to-urban transition: a cross-sectional study. *PLoS One* 11:e0162753. doi: 10.1371/journal.pone.0162753

Levy, D., Ehret, G. B., Rice, K., Verwoert, G. C., Launer, L. J., Dehghan, A., et al. (2009). Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* 41, 677–687. doi: 10.1038/ng.384

Loh, P. R., Danecek, P., Palamara, P. F., Fuchsberger, C., Reshef, Y. A., Finucane, H. K., et al. (2016). Reference-based phasing using the haplotype reference consortium panel. *Nat. Genet.* 48, 1443–1448. doi: 10.1038/ng.3679

MacKenzie, S. D., Crawford, A. A., Ackermann, D., Schraut, K. E., Hayward, C., Bolton, J. L., et al. (2019). Associations of 'relative corticosterone deficiency' with genetic variation in CYP17A1 and metabolic syndrome features. bioRxiv, 654269 [Preprint]. doi: 10.1101/654269

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* 48, 1279–1283. doi: 10.1038/ng.3643

Miller, S. A., Dykes, D. D., and Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16:1215. doi: 10.1093/nar/16.3.1215

Park, S. B., Blumenthal, J. A., Lee, S. Y., and Georgiades, A. (2011). Association of cortisol and the metabolic syndrome in Korean men and women. *J. Korean Med. Sci.* 26, 914–918. doi: 10.3346/jkms.2011.26.7.914

Phillips, D. I. W., Barker, D. J. P., Fall, C. H. D., Seckl, J. R., Whorwood, C. B., Wood, P. J., et al. (1998). Elevated plasma cortisol concentrations: a link between low birth weight and the insulin resistance syndrome? *J. Clin. Endocrinol. Metab.* 83, 757–760. doi: 10.1210/jcem.83.3.4634

Pillay, V., Crowther, N. J., Ramsay, M., Smith, G. D., Norris, S. A., and Lombard, Z. (2015). Exploring genetic markers of adult obesity risk in black adolescent south Africans—the birth to twenty cohort. *Nutr. Diabetes* 5, e157–e163. doi: 10.1038/nutd.2015.7

Punyadeera, C., van der Merwe, M. T., Crowther, N. J., Toman, M., Schlaphoff, G. P., and Gray, I. P. (2001). Ethnic differences in lipid metabolism in two groups of obese south African women. *J. Lipid Res.* 42, 760–767. doi: 10.1016/S0022-2275(20)31638-2

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Ramsay, M., Crowther, N., Tambo, E., Agongo, G., Baloyi, V., Dikotope, S., et al. (2016). H3Africa AWI-gen collaborative centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. *Glob. Health Epidemiol. Genom.* 1, 1–13. doi: 10.1017/gheg.2016.17

Reich, T., Cloninger, C. R., and Guze, S. B. (1975). The multifactorial model of disease transmission: description of the model and its use in psychiatry. *Br. J. Psychiatry* 127, 1–10. doi: 10.1192/bjp.127.1.1

Reynolds, R. M., Syddall, H. E., Walker, B. R., Wood, P. J., and Phillips, D. I. W. (2003). Predicting cardiovascular risk factors from plasma cortisol measured during oral glucose tolerance tests. *Metab. Clin. Exp.* 52, 524–527. doi: 10.1053/meta.2003.50090

Richter, L., Norris, S., Pettifor, J., Yach, D., and Cameron, N. (2007). Cohort profile: mandela's children: the 1990 birth to twenty study in South Africa. *Int. J. Epidemiol.* 36, 504–511. doi: 10.1093/ije/dym016

Rubinacci, S., Delaneau, O., and Marchini, J. (2020). Genotype imputation using the positional burrows wheeler transform. *PLoS Genet.* 16:e1009049. doi: 10.1371/journal.pgen.1009049

Schutte, C. E., Malan, L., Scheepers, J. D., Oosthuizen, W., Cockeran, M., and Malan, N. T. (2016). Cortisol: brain-derived neurotrophic factor ratio associated with silent ischaemia in a black male cohort: the SA BPA study. *Cardiovasc. J. Afr.* 27, 387–391. doi: 10.5830/CVJA-2016-065

Sengupta, D., Choudhury, A., Fortes-Lima, C., Aron, S., Gavin, W., Bostoen, K., et al. (2020). Genetic-substructure and complex demographic history of South African bantu speakers. bioRxiv [Preprint]. doi: 10.1101/2020.08.11.243840

Sirugo, G., Williams, S. M., and Tishkoff, S. A. (2019). The missing diversity in human genetic studies. *Cell* 177, 26–31. doi: 10.1016/j.cell.2019.02.048

Sliwa, K., Lyons, J. G., Carrington, M. J., Lecour, S., Marais, A. D., Raal, F. J., et al. (2012). Different lipid profiles according to ethnicity in the heart of Soweto study cohort of de novo presentations of heart disease. *Cardiovasc. J. Afr.* 23, 389–395. doi: 10.5830/CVJA-2012-036

Sumner, A. E., and Cowie, C. C. (2008). Ethnic differences in the ability of triglyceride levels to identify insulin resistance. *Atherosclerosis* 196, 696–703. doi: 10.1016/j.atherosclerosis.2006.12.018

Tolmay, C. M., Malan, L., and van Rooyen, J. M. (2012). The relationship between cortisol, C-reactive protein and hypertension in African and Causcasian women: the POWIRS study. *Cardiovasc. J. Afr.* 23, 78–84. doi: 10.5830/CVJA-2011-035

van der Linden, E., Meeks, K., Beune, E., Aikins, A. D. G., Addo, J., Owusu-Dabo, E., et al. (2019). Dyslipidaemia among Ghanaian migrants in three European countries and their compatriots in rural and urban Ghana: the RODAM study. *Atherosclerosis* 284, 83–91. doi: 10.1016/j.atherosclerosis.2019.02.030

Walker, B. R., Soderberg, S., Lindahl, B., and Olsson, T. (2000). Independent effects of obesity and cortisol in predicting cardiovascular risk factors in men and women. *J. Intern. Med.* 247, 198–204. doi: 10.1046/j.1365-2796.2000.00609.x

Ward, A. M. V., Fall, C. H. D., Stein, C. E., Kumaran, K., Veena, S. R., Wood, P. J., et al. (2003). Cortisol and the metabolic syndrome in south Asians. *Clin. Endocrinol.* 58, 500–505. doi: 10.1046/j.1365-2265.2003.01750.x

Weigensberg, M. J., Toledo-Corral, C. M., and Goran, M. I. (2008). Association between the metabolic syndrome and serum cortisol in overweight Latino youth. *J. Clin. Endocrinol. Metab.* 93, 1372–1378. doi: 10.1210/jc.2007-2309

Yadav, R., Petrunak, E. M., Estrada, D. F., and Scott, E. E. (2017). Structural insights into the function of steroidogenic cytochrome P450 17A1. *Mol. Cell. Endocrinol.* 441, 68–75. doi: 10.1016/j.mce.2016.08.035

Zhong, X. M., Li, L., Wang, H. Z., Zou, X. G., Zhang, P., Rexiati, M., et al. (2018). Associations of polymorphism of rs9944155, rs1051052, and rs1243166 locus allele in alpha-1-antitrypsin with chronic obstructive pulmonary disease in Uygur population of Kashgar region. *Chin. Med. J.* 131, 684–688. doi: 10.4103/0366-6999.226885

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 17 September 2021
doi: 10.3389/fgene.2021.690366

The Impact of Ethnicity and Genetic Ancestry on Disease Prevalence and Risk in Colombia

Aroon T. Chande[1,2,3], Shashwat Deepali Nagar[1,2,3], Lavanya Rishishwar[2,3], Leonardo Mariño-Ramírez[3,4], Miguel A. Medina-Rivas[5], Augusto E. Valderrama-Aguirre[6,7,8], I. King Jordan[1,2,3]† and Juan Esteban Gallo[9]*†*

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, United States, [2]IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, GA, United States, [3]PanAmerican Bioinformatics Institute, Cali, Colombia, [4]National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD, United States, [5]Centro de Investigación en Biodiversidad y Hábitat, Universidad Tecnológica del Chocó, Quibdó, Colombia, [6]Biomedical Research Institute (COL0082529), Cali, Colombia, [7]Department of Biomedical Sciences, Universidad Santiago de Cali, Cali, Colombia, [8]Department of Biological Sciences, Universidad de los Andes, Bogotá, Colombia, [9]GenomaCES, Universidad CES, Medellín, Colombia

Currently, the vast majority of genomic research cohorts are made up of participants with European ancestry. Genomic medicine will only reach its full potential when genomic studies become more broadly representative of global populations. We are working to support the establishment of genomic medicine in developing countries in Latin America *via* studies of ethnically and ancestrally diverse Colombian populations. The goal of this study was to analyze the effect of ethnicity and genetic ancestry on observed disease prevalence and predicted disease risk in Colombia. Population distributions of Colombia's three major ethnic groups – Mestizo, Afro-Colombian, and Indigenous – were compared to disease prevalence and socioeconomic indicators. Indigenous and Mestizo ethnicity show the highest correlations with disease prevalence, whereas the effect of Afro-Colombian ethnicity is substantially lower. Mestizo ethnicity is mostly negatively correlated with six high-impact health conditions and positively correlated with seven of eight common cancers; Indigenous ethnicity shows the opposite effect. Malaria prevalence in particular is strongly correlated with ethnicity. Disease prevalence co-varies across geographic regions, consistent with the regional distribution of ethnic groups. Ethnicity is also correlated with regional variation in human development, partially explaining the observed differences in disease prevalence. Patterns of genetic ancestry and admixture for a cohort of 624 individuals from Medellín were compared to disease risk inferred *via* polygenic risk scores (PRS). African genetic ancestry is most strongly correlated with predicted disease risk, whereas European and Native American ancestry show weaker effects. African ancestry is mostly positively correlated with disease risk, and European ancestry is mostly negatively correlated. The relationships between ethnicity and disease prevalence do not show an overall correspondence with the relationships between ancestry and disease risk. We discuss possible reasons for the divergent health effects of ethnicity and ancestry as well as the implication of our results for the development of precision medicine in Colombia.

Keywords: Colombia, Latin America, health disparities, precision medicine, ethnicity, genetic ancestry, cancer, malaria

OPEN ACCESS

Edited by:
Segun Fatumo, University of London, United Kingdom

Reviewed by:
Jovanny Zabaleta, Louisiana State University, United States
Pablo Duchen, University of Münster, Germany
Tesfaye B. Mersha, Cincinnati Children's Hospital Medical Center, United States
Victor Rivera, Baylor College of Medicine, United States

***Correspondence:**
I. King Jordan
king.jordan@biology.gatech.edu
Juan Esteban Gallo
jegallo@ces.edu.co

†These authors have contributed equally to this work and share senior authorship

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *02 April 2021*
Accepted: *11 August 2021*
Published: *17 September 2021*

Citation:
Chande AT, Nagar SD, Rishishwar L, Mariño-Ramírez L, Medina-Rivas MA, Valderrama-Aguirre AE, Jordan IK and Gallo JE (2021) The Impact of Ethnicity and Genetic Ancestry on Disease Prevalence and Risk in Colombia. Front. Genet. 12:690366. doi: 10.3389/fgene.2021.690366

INTRODUCTION

Genomic medicine is an emerging medical discipline that entails the use of genomic information about an individual as part of their clinical care – in support of better diagnostic, prognostic, and therapeutic decision-making (Collins and Varmus, 2015; Jameson and Longo, 2015). Genomic medicine promises to revolutionize healthcare, but the vast majority of genomics research cohorts are currently made up of individuals with European ancestry (Bustamante et al., 2011; Popejoy and Fullerton, 2016). Clinical insights based on the study of European ancestry genomes will not necessarily replicate across diverse populations (Martin et al., 2017). This genomics research gap limits the reach of genomic medicine and threatens to exacerbate existing health disparities (Petrovski and Goldstein, 2016; Martin et al., 2019). The promise of genomic medicine will not be fully realized until genomic studies become more broadly representative of global populations.

Colombia has a diverse, multi-ethnic population with major ancestry contributions from Europe, Africa, and the Americas (Bryc et al., 2010; Ruiz-Linares et al., 2014; Homburger et al., 2015). Colombian genomic diversity represents a rich and largely untapped resource that can be used to support the development of genomic medicine locally in Colombia and around the world. We have been working to build local capacity in precision medicine *via* population and clinical genomic studies of diverse Colombian populations over the last decade (Rishishwar et al., 2015a,b; Jordan, 2016; Medina-Rivas et al., 2016; Chande et al., 2017, 2020a,b; Conley et al., 2017; Norris et al., 2018, 2020; Nagar et al., 2019). These studies share the broad aims of (1) characterizing patterns of genetic ancestry and admixture within and between Colombian and other Latin American populations, and (2) exploring the relationship between ancestry and genetic determinants of health and disease in the region.

The goal of the current study was to analyze the effect of ethnicity and genetic ancestry on observed disease prevalence and predicted disease risk in Colombia. We focused on Colombia's three largest ethnic groups – Mestizo, Afro-Colombian, and Indigenous – along with corresponding genetic ancestry contributions from Europe, Africa, and the Americas. We studied health conditions and diseases that have been prioritized by the Colombian government as having an outsized impact on public health and the economy – six high impact non-cancer conditions along with eight of the most common cancers. Ethnicity and disease prevalence were analyzed at the level of administrative departments (states) and geographic regions, and genetic ancestry and disease risk were inferred using whole genome genotype (WGG) data for a diverse cohort of 624 individuals. We found that disease prevalence and risk are associated with ethnicity, geography, socioeconomics, and genetic ancestry in Colombia, and we discuss the implications of our findings with respect to the development of precision medicine in the country.

MATERIALS AND METHODS

Study Cohorts

A cohort of 624 individuals from Medellín, Colombia was recruited and genotyped by GenomaCES Biotechnologies of the Universidad CES (UniCES),[1] and a cohort of 99 individuals from Chocó, Colombia was recruited and genotyped as part of the ChocoGen research project.[2] All sample donors signed informed consent, and all participant recruiting, sampling, and genetic characterization was done following the Helsinki ethical principles for medical research involving human subjects. Human subject research in Colombia was conducted in accordance with article 11, resolution 8,430, 1993 of the Colombian Ministry of Health, which states that for every investigation in which a human being is the study subject, respect for their dignity and the protection for their rights should always be observed. The UniCES Medellín cohort was recruited and characterized with the approval of the Ethics and Research Committee of the Universidad CES. The ChocoGen project was conducted with the approval of the Ethics Committee of the Universidad Tecnológica del Chocó. Population genomic analysis was approved by the Institutional Review Board of the Georgia Institute of Technology.

Ethnicity, Disease Prevalence, and Socioeconomic Data

Data on Colombian ethnic groups was taken from the 2005 census as reported by the National Administrative Department of Statistics (DANE; Rojas Morales et al., 2007). Ethnic identity is self-reported in the Colombian census, and individuals choose from one of six ethnic groups: (1) *Indígena*, (2) *Rom*, (3) *Raizal del Archipiélago de San Andrés y Providencia*, (4) *Palanquero de San Basilio*, (5) *Negro(a)*, *mulato(a)*, *afrocolombiano(a) o afrodescendiente*, and (6) *Ninguna de las anteriores* (**Supplementary Figure 1**). We focused on the three largest ethnic groups in Colombia: Mestizo, Afro-Colombian, and Indigenous. The ethnic group labels, we use for this study are English translations of the officially used Spanish group names in Colombia, except for the Spanish word Mestizo, which we adopt here as it is widely used in both English and Spanish. Following the convention of DANE, Indigenous ethnic identity corresponds to question #1 *Indígena*, Afro-Colombian ethnic identity corresponds to questions #3, #4, and #5, all of which correspond to specific Afro-Colombian communities or identities, and Mestizo ethnic identity corresponds to question #6 *Ninguna de las anteriores* (none of the above). The Mestizo group, which may include individuals who identify as white or Mestizo, is also referred to as *sin pertenencia étnia* (no ethnicity), reflecting the fact that majority population individuals are not considered to belong to any of Colombia's officially recognized minority ethnic groups. The *Rom* census question refers the Roma community, which makes up less than 0.01% of the Colombian population and was therefore not considered here.

[1]https://www.genomaces.com/
[2]https://www.chocogen.com/

Population numbers and percentages for each of the three major ethnic groups are reported for the entire country and for each of 32 administrative departments plus the capital district of Bogotá. Regional population numbers and percentages for the three groups were calculated based on the administrative departments that make up each of the five geographic regions (Bustamente et al., 2012).

The Colombian *Instituto Nacional de Salud* and the non-governmental organization *Cuenta de Alto Costo* were used to identify health conditions and diseases that have a maximum impact on public health and the economy. We chose six non-cancer conditions and eight of the most common cancers for analysis. Prevalence data for these conditions and diseases were taken from two databases: (1) *Cuenta de Alto Costo*,[3] and (2) *Sistema Nacional de Vigilancia en Salud Pública – SIVIGILA*.[4] Disease prevalence values are calculated and expressed as age- and sex-adjusted prevalence per 100,000 population. Disease prevalence data are reported for the entire country and for each of 32 administrative departments plus the capital district of Bogotá.

The Human Development Index (HDI) is a composite index combining three dimensions of human development: long and healthy life, knowledge, and standard of living.[5,6] Each dimension is measured by specific indicators: life expectancy at birth, expected years of schooling and mean years of schooling, and gross national income *per capita*. Indicators are normalized, using fixed maximum and minimum values for each indicator, to yield I_{Health}, $I_{Education}$, and I_{Income} dimension index values that range from 0 to 1: dimension index (I) = (actual indicator value – minimum indicator value)/(maximum indicator value – minimum indicator value). The arithmetic mean of the two normalized education dimension index values is calculated for $I_{Education}$. The HDI is calculated as the geometric mean of the three dimension index values: $HDI = (I_{Health} \times I_{Education} \times I_{Income})^{1/3}$. HDI data for Colombian administrative departments was taken from the Global Data Lab,[7] and values range from 0.69 to 0.90.

Correlations between ethnic group percentages, HDI, and disease prevalence values for administrative departments were done using Pearson's correlation (*R*) using the cor.test function in base R v3.5.1 (R Core Team, 2020). Ethnic group percentages are the percentage of the population that each ethnic group makes up for each administrative department, values range from 0 to 98.27%. Disease prevalence values are taken for each administrative department, as described above, values range from 0 to 7,847 cases per 100,000 population.

Genetic Ancestry and Disease Risk Prediction

Universidad CES Medellín cohort participants' WGG were characterized using the Illumina Global Screening Array,[8] and ChocoGen cohort participants' WGG were characterized using the Illumina HumanOmniExpress-24 Array.[9] Full details on the sampling, DNA extraction, genotyping, and quality control procedures for these data, all of which were characterized for previous studies, have been reported elsewhere (Pato et al., 2013; Medina-Rivas et al., 2016; Conley et al., 2017; Bigdeli et al., 2020).

The UniCES Medellín cohort WGG data were merged and harmonized with whole genome sequence (WGS) data from global reference populations representing three continental population group ancestries – European, African, and Native American (**Supplementary Table 1**) – characterized as part of the 1,000 Genomes Project (1KGP), using the program PLINK v1.9 and bespoke scripts (1000 Genomes Project Consortium et al., 2015; Chang et al., 2015). Variant data from UniCES WGG and 1KGP WGS were merged to include variants that were present in both datasets with a missingness and minor allele frequency filters of 5 and 1%, respectively. Variant strand flips and identifier inconsistencies were corrected as needed. The merged and harmonized dataset contained 425,732 genome-wide variants. This combined dataset was then merged with WGG from Chocó, Colombia characterized as part of the ChocoGen project (see footnote 2), with a missingness threshold of 5%. The three-way merged UniCES-1KGP-ChocoGen dataset contained 77,575 variants. The three-way merged dataset was pruned for linkage disequilibrium (LD) using the "--indep" command in PLINK 1.9 with a window size of 50 kb, a step size of five variants, and a variant inflation factor (VIF) threshold of two to yield a final merged, harmonized, and LD pruned dataset of 63,852 variants, which was used for genetic ancestry characterization.

Principal component analysis (PCA) of the final variant dataset was performed using PLINK using the "--pca" option and the first two PCs for all samples were plotted using the ggplot2 package in R v3.5.1 (Wickham, 2016; R Core Team, 2020). The program ADMIXTURE v1.30 was used to characterize participants' genome-wide ancestry fractions for the three continental ancestry groups – European, African, and Native American (Alexander et al., 2009). ADMIXTURE was run in unsupervised mode with default settings and $K = 3$.

The NHGRI-EBI genome-wide association study (GWAS) Catalog was mined for trait-variant associations that correspond to the health conditions and diseases studied here.[10] Trait-associated variant sets corresponding to four of six non-cancer health conditions and seven of the eight common cancers were used to infer genetic disease risk *via* polygenic risk scores (PRS; **Supplementary Table 2**). GWAS were filtered by trait of interest, type of study, number of individuals and distinct ancestries in the discovery and replication cohorts, and finally by number of associations reported. We aimed to identify at least one large, multi-ethnic study for each trait, under the assumption that PRS derived from multi-ethnic cohorts are less likely to be biased. Only traits with at least 20 variant

[3]https://cuentadealtocosto.org/
[4]https://www.ins.gov.co/Direcciones/Vigilancia/Paginas/SIVIGILA.aspx
[5]http://hdr.undp.org/en/content/human-development-index-hdi
[6]http://hdr.undp.org/sites/default/files/hdr2020_technical_notes.pdf
[7]https://globaldatalab.org/shdi/shdi/
[8]https://www.illumina.com/products/by-type/microarray-kits/infinium-global-screening.html
[9]https://www.illumina.com/products/by-type/microarray-kits/infinium-omni-express.html
[10]https://www.ebi.ac.uk/gwas/

associations were retained for PRS calculation. Variant identifiers (rsid), effect alleles, effect sizes, and *p*-values were collected from for each selected trait-study combination. Curated variant lists for each trait were subjected to LD clumping ("--clump-r2 0.2") and used to calculate PRS for imputed genomic variant data for the UniCES Medellín cohort using PLINK (v1.90b6.16), as previously reported (Chande et al., 2018, 2020a). PRS for each condition or disease *i* were calculated as the sum of the effect alleles across all trait-associated SNPs as – $PRS_i = \sum_{j=1}^{n} EA_j / \sum_{j=1}^{n} A_j$ – where $E_j \in \{0,1,2\}$ corresponds to homozygous absent, heterozygous present, or homozygous present effect alleles at each variant, and $A_j \in \{0,1,2\}$ corresponds to the total number of alleles with base calls at each variant. WGG data were imputed using the 1KGP haplotype reference panel with the program IMPUTE2 version 2.3.2 (Howie et al., 2011, 2012). Imputed sites were retained for subsequent analysis if they had a 95% imputation rate across samples and an INFO score >0.4.

Correlations between individuals' genetic ancestry fractions and PRS were done using Pearson's correlation (*R*) using the cor.test function in base R v3.5.1 (R Core Team, 2020). Genetic ancestry fractions range from 0 to 0.925, and PRS values range from 0 to 1.

RESULTS

Ethnicity and Genetic Ancestry

The Colombian census reports the ethnic composition of the country as 85.9% Mestizo, 10.6% Afro-Colombian, and 3.4% Indigenous (Rojas Morales et al., 2007). There is high variation in the distribution of ethnic groups among Colombia's administrative departments and geographic regions (**Table 1**; **Supplementary Table 3**). The percent Mestizo population of individual departments ranges from 5.2% in the department of Chocó in the western Pacífico region to 98.4% in the capital district of Bogotá. Afro-Colombian population percentages range from 1.0% for the department of Guainía in the eastern Amazonía region to 82.1% in Chocó. Indigenous percentages range from 0.1% on the Caribbean Islands of San Andrés and Providencia to 66.7% for the Vaupés department in the southeastern Amazonía region. Broadly speaking, Mestizo population composition is highest in the central Andes region, Afro-Colombian populations are concentrated in the Pacífico and Caribe regions along the Pacific and Atlantic coasts, and Indigenous populations are highest in the Amazonía region (**Figure 1**).

TABLE 1 | Ethnic group population percentages for the five broad geographic regions and the entire country.

Region	Mestizo	Afro-Colombian	Indigenous
Amazonía	63.09	3.27	33.64
Andes	93.52	5.19	1.29
Caribe	70.33	19.13	10.53
Orinoquía	95.65	2.68	1.67
Pacífico	51.38	30.63	17.98
Total	85.94	10.62	3.43

Whole genome genotype data were characterized for a cohort of 624 individuals sampled from the city of Medellín in the department of Antioquia and 99 individuals sampled from Quibdó in the department of Chocó (**Figure 2A**). Comparison of individual WGG data from these samples with WGS data from European, African, and Native American ancestry reference populations shows evidence of substantial admixture for all three continental ancestry components. The Medellín sample characterized here is similar, but substantially more ancestrally diverse, compared to the previously characterized 1KGP population from Colombia, and falls between reference populations corresponding to European, African, and Native American ancestry (**Figure 2B**). The Chocó sample is closer to the African American and African Caribbean 1KGP populations.

The program ADMIXTURE was used to quantify ancestry components for the Colombian genomes analyzed here. ADMIXTURE was run for $K = 2$–10 ancestry components, with $K = 3$ ancestry components showing the best fit to the data. ADMIXTURE analysis confirms the ancestral diversity of the two Colombian cohorts: the Medellín cohort shows an average of 55.0% European ancestry, 32.4% Native American ancestry, and 12.6% African ancestry, and the Chocó cohort shows an average of 76.0% African ancestry, 12.5% Native American ancestry, and 11.5% European ancestry (**Figures 2C–E**). All of the Colombian genomes analyzed here show evidence of admixture with two or more ancestry components, with a minimum non-European ancestry component of 7.5%. European ancestry percentages for individual Colombian genomes studied here range from 2.5 to 92.5%, Native American ancestry from 1.3 to 73.6%, and African ancestry from 0.0 to 92.1%, underscoring the diversity of Colombian populations.

Ethnicity, Disease Prevalence, and Comorbidities

Department disease prevalence values, expressed as the age- and sex-adjusted number of cases per 100,000 individuals, were recorded for six health conditions and eight cancers prioritized as high impact by Colombian governmental and non-governmental organizations (**Supplementary Table 4**). Malaria shows the highest prevalence for the six prioritized non-cancer health conditions, with a per-department average of 618.5 cases per 100,000 individuals, followed by type 2 diabetes with 367.7 cases per 100,000. Breast cancer and prostate cancer were the most common cancer types, with prevalence values of 131.6 and 64.1 per 100,000 individuals, respectively. The highest combined prevalence for the six prioritized health conditions is seen for the departments of Guainía and Amazonas in the Amazonía region followed by Chocó in the Pacífico region. Disease prevalence in these departments is dominated by Malaria, but the combined disease prevalence remains high when Malaria is not considered. The Islands of San Andrés and Providencia are the exception to this trend with a high prevalence for non-Malaria conditions and little to no risk of Malaria. The highest combined prevalence across all eight cancers is seen for the departments of Risaralda, Antioquia,

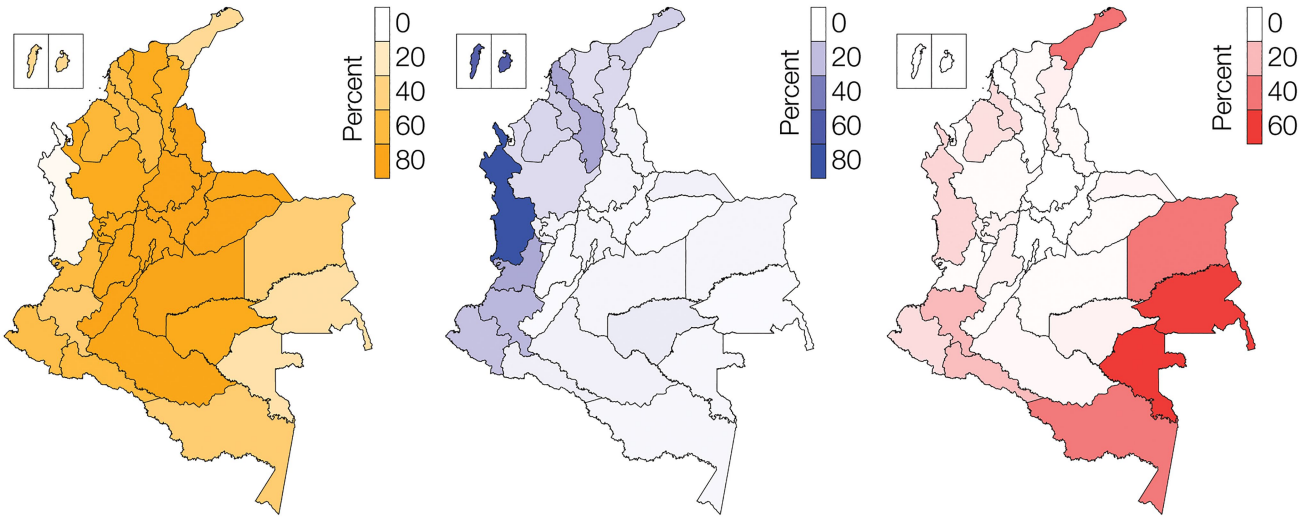

FIGURE 1 | Ethnic groups geographic distributions. Percentages of Mestizo (orange), Afro-Colombian (blue), and Indigenous (red) populations among Colombia's administrative departments.

and the capital district of Bogotá, all of which are located in the central Andes region.

Percentages of Mestizo, Afro-Colombian, and Indigenous ethnic populations across departments were compared to the disease prevalence data. The combined prevalence of the six prioritized non-cancer conditions is positively correlated with Indigenous population percentages and negatively correlated with Mestizo percentages (**Figure 3A**). Cancer shows the opposite pattern, a positive correlation with Mestizo population percentages and a negative correlation with Indigenous percentages. Afro-Colombian population percentages show comparatively smaller correlations with disease prevalence, only slightly negative for the non-cancer conditions and slightly positive for the cancers. Mestizo and Indigenous population patterns are mirror images of each other with respect to the prevalence of the individual non-cancer conditions and cancer (**Figure 3B**). Indigenous population percentages are positively correlated for six out of seven prioritized non-cancer conditions, whereas Mestizo and Afro-Colombian percentages are mostly negatively correlated with these same conditions. Malaria shows the highest overall correlations with ethnicity, with high positive correlations for both Afro-Colombian and Indigenous population percentages and the most extreme negative correlation with Mestizo percentages. Mestizo population percentages are positively correlated with seven out of eight cancer types and negatively correlated with Hodgkin's Lymphoma, whereas Indigenous percentages show the opposite patterns.

Non-cancer and cancer disease prevalence values were normalized within departments (**Supplementary Figure 2**), and the resulting relative disease prevalence values for all departments were correlated to calculate pairwise disease comorbidities among all departments. The resulting departmental pairwise comorbidity values were hierarchically clustered and compared to the departments' locations within Colombia's five geographic regions (**Figure 4**). Population comorbidity values are highest around the diagonal line and correspond well with individual departments' locations within broad geographic regions. The tightest clustering of comorbidity values, indicating similar relative disease prevalence values for departments within a given region, is seen for the Amazonía and Andes regions, corresponding to their high percentages of Indigenous and Mestizo populations, respectively. The Caribe region also shows high population comorbidity, whereas the comorbidity values for the Pacífico region are lower, reflecting greater demographic diversity for the region (**Figure 1**; **Table 1**; **Supplementary Table 3**).

Genetic Ancestry and Predicted Disease Risk

Genetic risk estimates for four of the six of the prioritized non-cancer conditions and seven of the eight most common cancers were inferred using PRS computed on genomic variant data from the UniCES Medellín cohort studied here. Individuals' genetic ancestry percentages – European, African, and Native American – were regressed against PRS for each trait to evaluate the relationship between ancestry and predicted disease risk.

European genetic ancestry is significantly positively correlated with the predicted risk of gastric carcinoma and negatively correlated with the risk of acute lymphoblastic leukemia and breast carcinoma (**Figure 5A**). African genetic ancestry shows statistically significant correlations with the predicted risk for five out of 11 diseases evaluated, compared to three and one significant correlation(s) for European ancestry and Native American, respectively. African genetic ancestry is positively correlated with the predicted risk of type 2 diabetes, colorectal carcinoma, and prostate cancer, and negatively correlated with gastric carcinoma and lung adenocarcinoma. The only significant

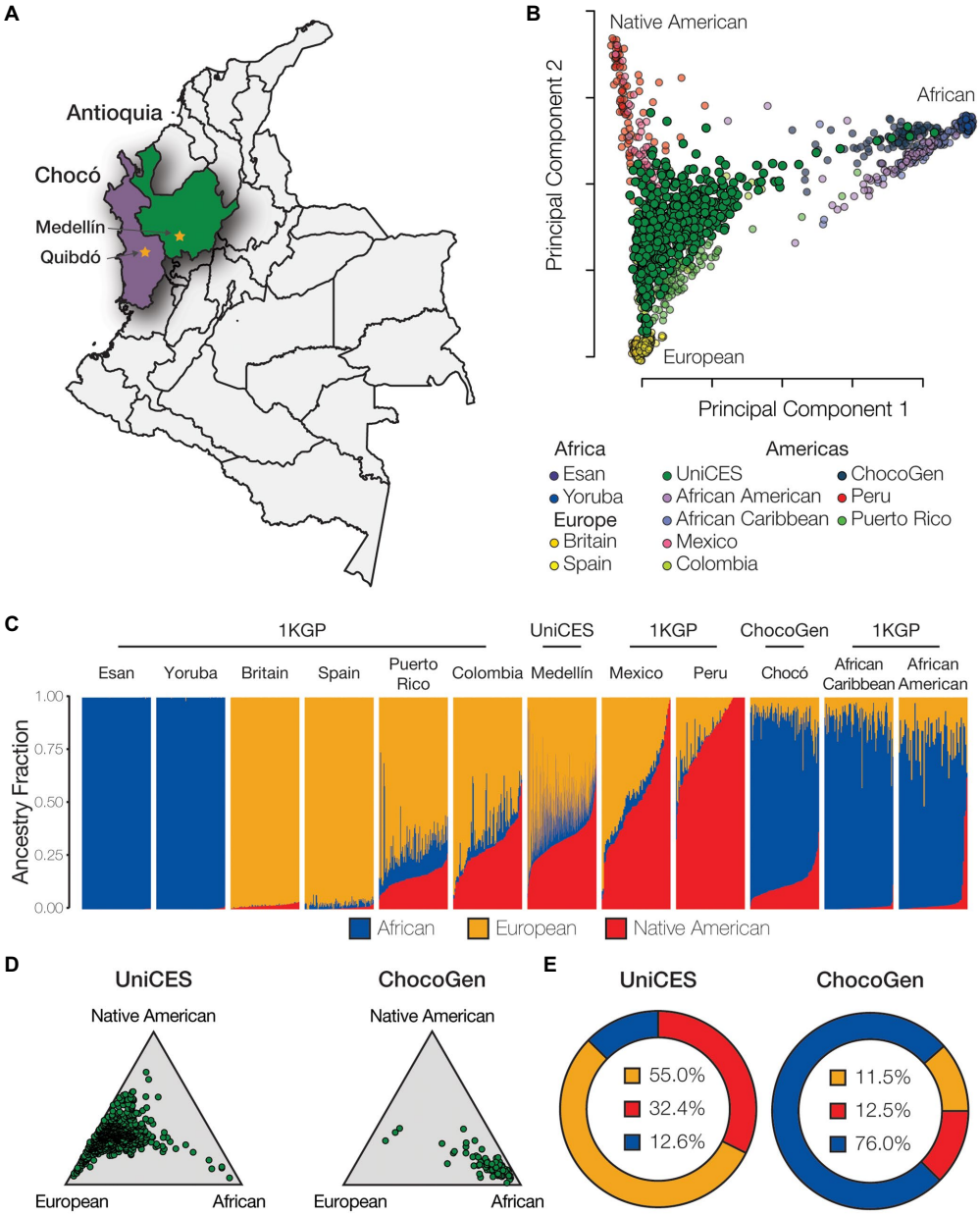

FIGURE 2 | Genetic ancestry and admixture. **(A)** Cohort sampling locations in Medellín, Antioquia and Quibdó, Chocó. **(B)** Principal component analysis (PCA) showing the genetic relationships among individuals from admixed American populations, including the Colombian populations studied here, together with European, African, and Native American reference populations. **(C)** ADMIXTURE plot showing the patterns of genetic ancestry fractions – European (orange), African (blue), Native American (red) – for individuals from admixed American and global reference populations. **(D)** Ternary plots showing the relative distributions of continental ancestry for the Medellín and Chocó cohorts studied here. **(E)** Population average ancestry fractions for the for the Medellín and Chocó cohorts. Population abbreviations: 1KGP – 1,000 Genomes Project, ChocoGen – Chocoano in Quibdó, Colombia, UniCES – Cosmopolitan Colombian in Medellín, Colombia.

correlation seen for Native American genetic ancestry is the negative correlation with the predicted risk for acute lymphoblastic leukemia. Overall, African genetic ancestry shows the highest effect on predicted disease risk, as measured by the sum of the absolute values of the ancestry-PRS correlations, followed by European ancestry and Native American ancestry, respectively (**Figure 5B**). African ancestry shows the highest overall positive correlation for predicted disease risk, as measured by the sum of the values of the ancestry-PRS correlations, whereas European ancestry shows the highest overall negative correlation with disease risk (**Figure 5C**). The overall correlation between Native American ancestry and predicted disease risk is negligible. Examples of regressions for individual ancestry-disease combinations are seen for European ancestry and breast carcinoma, African ancestry and type 2 diabetes, and Native American ancestry and acute lymphoblastic leukemia (**Figure 5D**).

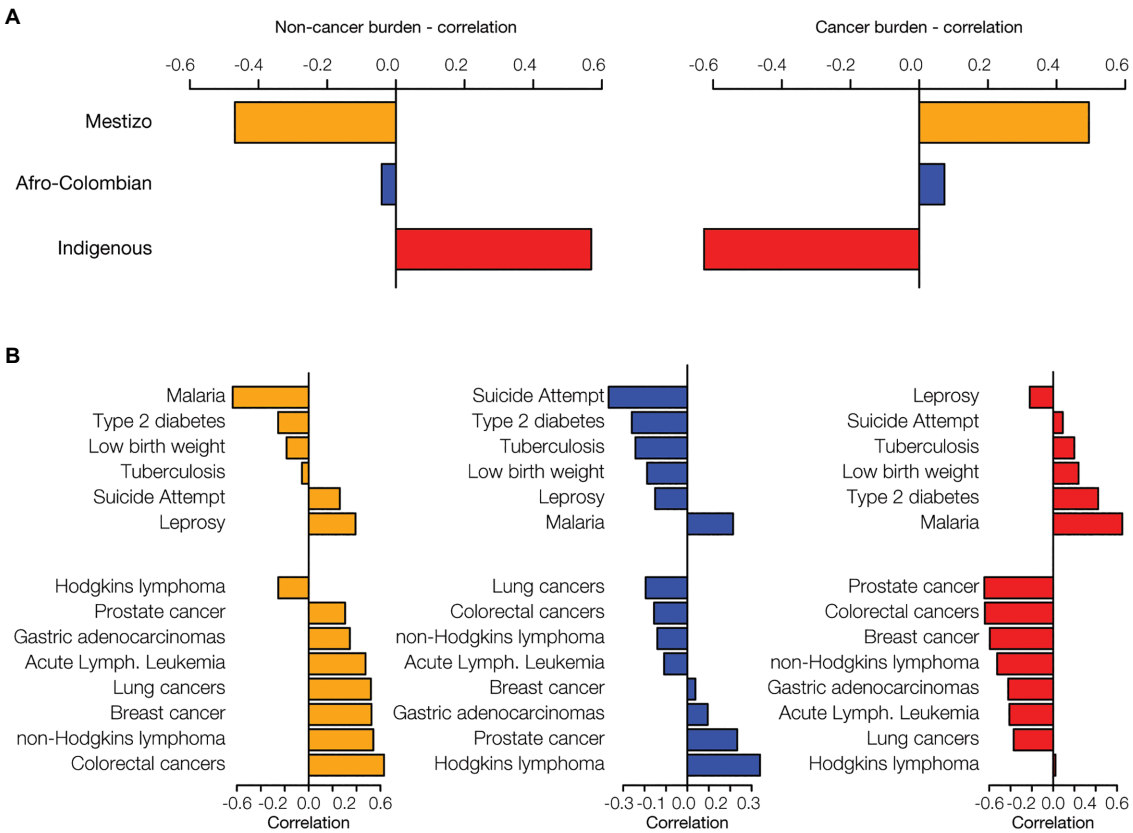

FIGURE 3 | Ethnicity and disease prevalence. Administrative department ethnic group percentages were correlated with disease prevalence data. Data are shown for the three largest ethnic groups in Colombia: Mestizo (orange), Afro-Colombian (blue), and Indigenous (red). **(A)** The overall correlation of ethnicity with disease prevalence was calculated by summing the disease-specific correlations for six non-cancer health conditions and eight cancers. **(B)** For each ethnic group, ethnicity-disease prevalence correlations are shown for six non-cancer conditions (top of each plot) and the eight cancers (bottom of each plot).

Differences Between Observed Disease Prevalence and Predicted Disease Risk

We analyzed the three major ethnic groups in Colombia – Mestizo, Afro-Colombian, and Indigenous – in comparison with three corresponding genetic ancestry groups – European, African, and Native American. Both ethnicity and genetic ancestry show a number of significant correlations with observed disease prevalence and predicted disease risk. However, the direction of the disease correlations for the corresponding ethnicity and ancestry groups often do not match. In other words, a positive (or negative) correlation between ethnicity and observed disease prevalence among departments does not necessarily entail a positive (or negative) correlation between genetic ancestry and predicted disease risk for the individuals in the ancestrally diverse cohort studied here, as may be expected given the correlation between ancestry and ethnicity in Colombia. For Afro-Colombian ethnicity and African genetic ancestry there are six concordant ethnicity/ancestry-disease correlation pairs and five discordant correlation pairs (**Figure 6**). For both Mestizo ethnicity-European genetic ancestry and Indigenous ethnicity-Native American genetic ancestry, there are four concordant correlation pairs and seven discordant pairs.

Human Development, Ethnicity, and Health

Health outcomes are strongly influenced by a variety of socioeconomic factors. The HDI is a composite index that is used to measure three key aspects of human development: life expectancy, education, and income. HDI in Colombia is related to demography with marked differences seen for the three ethnic groups studied here. The Mestizo population percentage is positively correlated with HDI, whereas the Afro-Colombian and Indigenous population percentages are negatively correlated with HDI (**Figure 7A**). Differences in the level of economic development and human capital captured by HDI may explain the differences in population health outcomes observed here. The prevalence of prioritized non-cancer conditions, Malaria in particular, tend to be negatively correlated with HDI, whereas cancer is uniformly positively correlated with HDI (**Figure 7B**). Clearly, human development and socioeconomic factors are strongly related with health outcomes in Colombia. Nevertheless, the patterns of genetic admixture observed for the Colombian population, which to some extent blur the genetic ancestry distinctions among ethnic groups, may obscure the relationship between socioeconomic factors, genetic ancestry, and disease risk in the country. This may explain the discordance seen between ethnicity and disease prevalence vs. ancestry and disease risk.

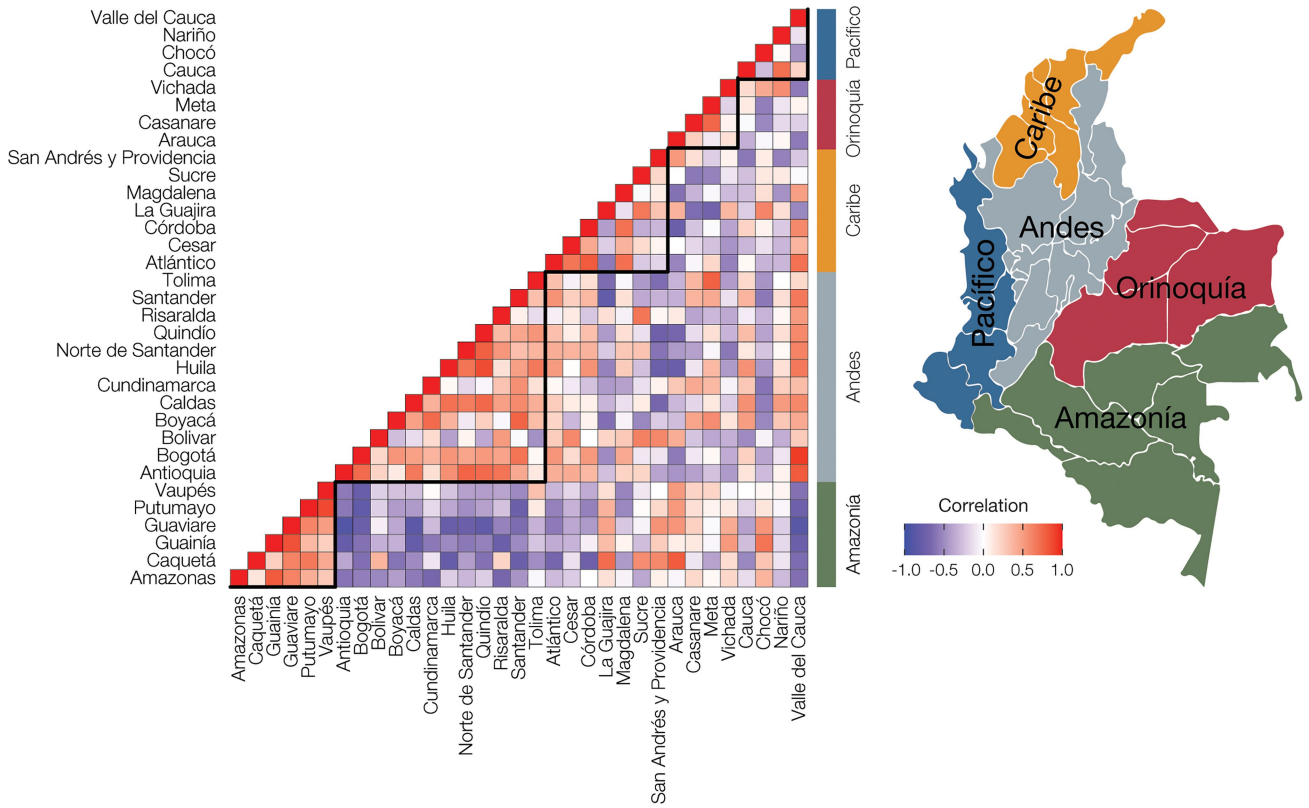

FIGURE 4 | Population comorbidity and geography. Disease prevalence values were normalized for each administrative department by dividing the observed prevalence value for each disease with the maximum disease prevalence value for the department to yield relative disease prevalence values. The relative disease prevalence values were correlated between all pairs of departments to infer overall comorbidity levels, i.e., similar relative disease prevalence values, among departments and geographic regions. Positive disease prevalence correlations between departments are shown in red, and negative correlations are shown in purple. Administrative department names are shown to the left of the plot, and geographic region names are shown to the right of the plot, with locations indicated on the adjacent map.

DISCUSSION

Ethnicity vs. Genetic Ancestry in Colombia

Ethnic designations in Colombia are based on self-identification. In the 2005 census, individuals could choose from one of six ethnic categories, including Indigenous, *Rom*, three categories of Afro-Colombian groups, or none of the above (**Supplementary Figure 1**; Rojas Morales et al., 2007). For this study, we focused on the three largest ethnic groups: Mestizo, which corresponds to the "none of the above" option from the census, and may also include individuals who self-identify as white, Afro-Colombian, and Indigenous. The *Rom* category from the census corresponds to the recently recognized Roma population, whose presence in the Americas dates to Columbus' earliest arrival but only make up 0.01% of the Colombian population. The small size of this ethnic group did not allow for sufficient resolution for the comparative analyses conducted here.

It is important to note that self-identified ethnicity is distinct from genetic ancestry as measured here. Colombians of all ethnicities show high levels of genetic admixture (Ruiz-Linares et al., 2014; Conley et al., 2017). Mestizos in particular have both substantial European and Native American genetic ancestry components and often show a small but not insubstantial African ancestry fraction as well. Even isolated Afro-Colombian and Indigenous Colombian communities show evidence of genetic admixture. Thus, while the ethnic categories used here are strongly correlated with genetic ancestry, they clearly do not capture the complexity of ancestry and admixture seen for the Colombian population. In addition to ancestry, self-identified ethnic categories also reflect individuals lived experiences, including many aspects of the environment and socioeconomic factors that can influence health outcomes. These distinctions between ethnicity and genetic ancestry may explain some of the differences observed between ethnic and genetic ancestry effects on disease prevalence and predicted risk observed here.

Implications for Colombia: Precision Public Health

The strong relationship between ethnicity, ancestry, and health outcomes in Colombia support the adoption of the precision public health paradigm as a way to bring genomic medicine to the country. Precision public health is an alternative to

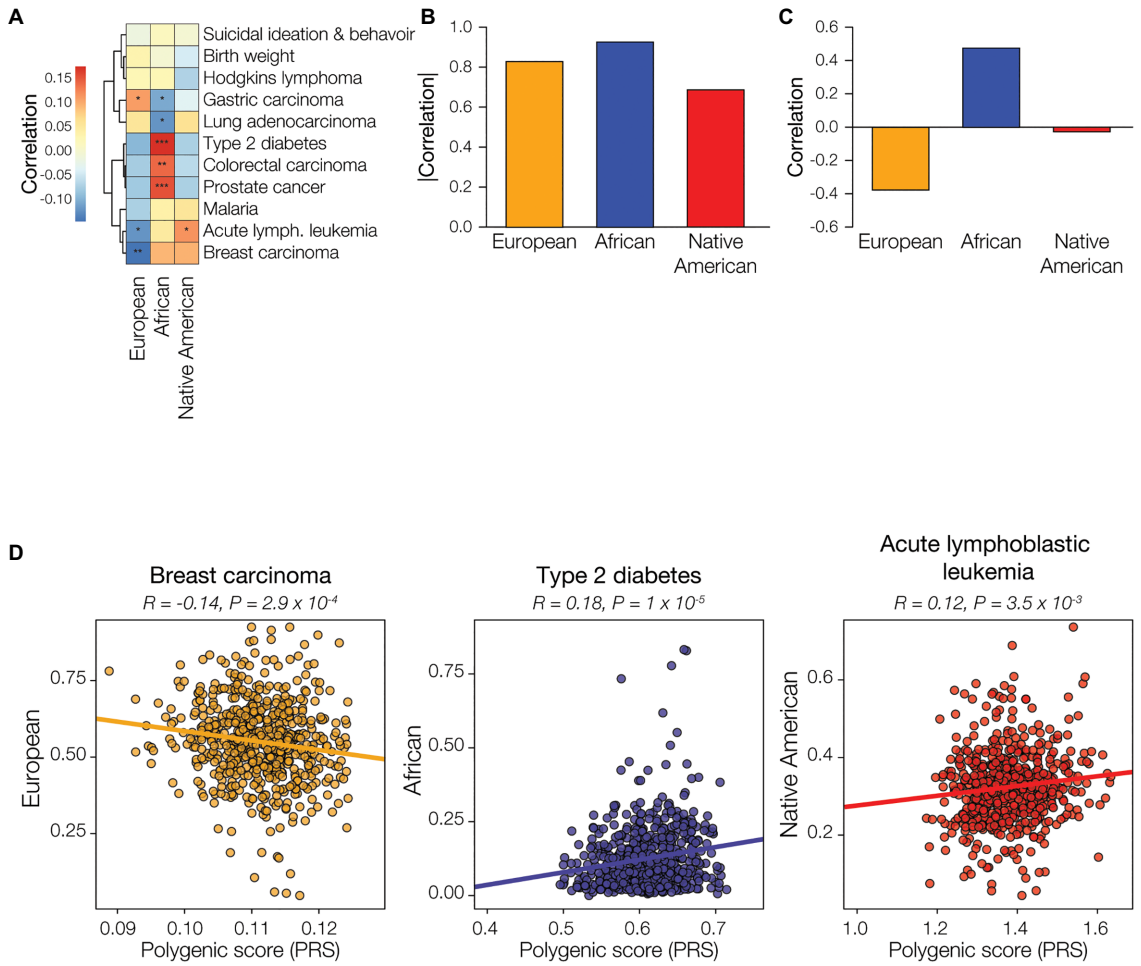

FIGURE 5 | Genetic ancestry and disease risk. Genetic ancestry percentages – European (orange), African (blue), and Native American (red) – and disease polygenic risk scores (PRS) were computed from participant genomic data and correlated. **(A)** Examples of correlations between genetic ancestry and PRS are shown for each of the three ancestries. Pearson correlation coefficient and *p*-values are shown for each. **(B)** Genetic ancestry-PRS correlations for all three ancestries and 11 health conditions or diseases are shown. Positive correlations are shown in red, and negative correlations are shown in blue. Correlation statistical significance levels are indicated ($0.01 < p < 0.075 = *$, $0.001 < p < 0.01 = **$, $p < 0.001 = ***$). The overall correlation of genetic ancestry with disease risk was calculated by summing the disease-specific correlations, absolute values **(C)** and raw correlations **(D)**, for the 11 conditions.

precision medicine, as it entails focus on populations instead of individuals (Khoury et al., 2016, 2018; Weeramanthri et al., 2018). Precision public health relies on detailed health, demographic, and population genomic profiles of developing countries, which allows for the allocation of resources and efforts where they will be most effective. The hope is that an emphasis on population-level interventions will enable the emerging genomic technologies that underlie precision medicine to be efficiently implemented in developing countries. We recently showed how population genomic profiles can be used to guide pharmacogenetic decision making in Colombia in a way that targets resources to the populations, where they will generate the maximum benefit to health and the best return on investment (Nagar et al., 2019). The results reported here, particularly as they relate to the distinct health characteristics of different Colombian populations and regions, underscore the need to focus health intervention efforts in a population-specific way. To the extent that genomic approaches to healthcare are implemented in the country, they should be targeted to groups and areas where they can be expected to yield the most benefit and the least harm. This approach will need to be attuned to the differences between self-identified ethnicity and genetically measured ancestry demonstrated here.

Limitations of the Study and Future Directions

One of the major limitations of this study is the lack of individual level health data for the cohort studied here. The ethnicity and disease prevalence data are reported at the level of administrative departments, which facilitates high-level comparisons but does not allow for detailed interrogation of the relationship between genetic ancestry and health. Future efforts to study the relationship between ethnicity, genetic

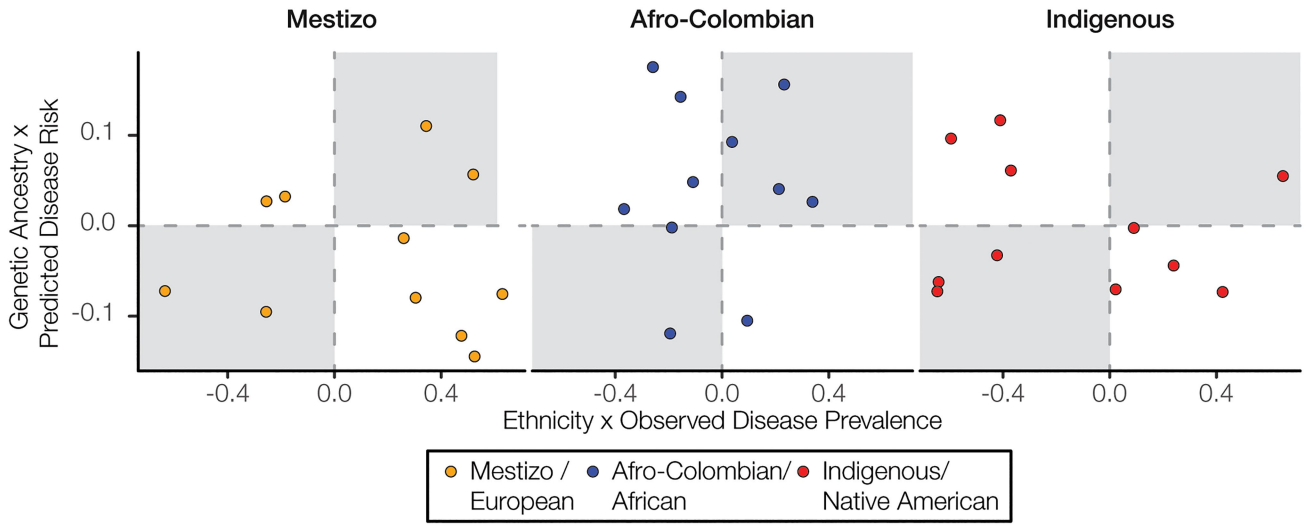

FIGURE 6 | Ethnicity and disease prevalence vs. genetic ancestry and disease risk. Correlation values for ethnicity and disease prevalence are plotted against correlation values for genetic ancestry and disease risk. Concordant correlations, both positive or both negative, are shown in the upper right and lower left quadrants, respectively.

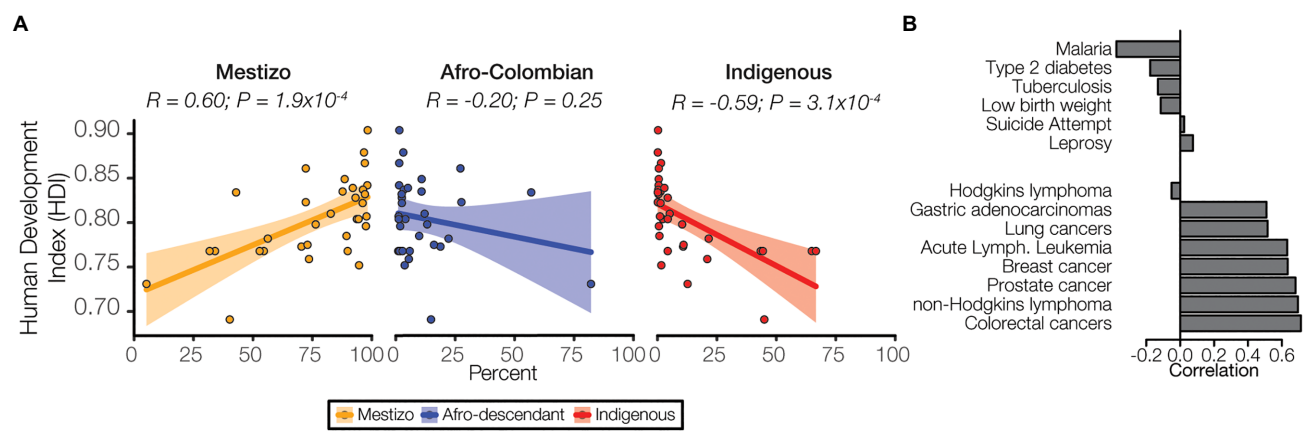

FIGURE 7 | Ethnicity, socioeconomic status, and disease. **(A)** Human develop index (HDI) values for each administrative department are correlated with ethnic group population percentages for each department: Mestizo (orange), Afro-Colombian (blue), Indigenous (red). Each point represents a single department, and the linear trend is indicated for each ancestry along with 95% CI (shaded area). Pearson correlation coefficient and *p*-values are shown for each. **(B)** Correlations between prevalence and HDI are shown for six non-cancer conditions (top) and the eight cancers (bottom).

ancestry, and health will be strengthened by the creation of population biobanks that include biological samples, deep phenotypic data, and electronic health records for individual participants. Biobanks also can and should include as many socioeconomic indicators as possible. Unfortunately, the goal of a country-level biobank, such as the United Kingdom Biobank or the United States All of Us project, remains out of reach for Colombia, at least for the time being. However, individual laboratories, including our own groups, are beginning to collaborate to marshal resources to create the kind of biobank data that will be needed to strengthen this line of research.

The use of different sources of data, both for disease prevalence values and for genome-wide genotyping, represents a potential source of bias in our study. However, this bias is likely mitigated by the choice of epidemiological databases and the methods for data harmonization used here. The disease prevalence databases were chosen as they are considered to be the most reliable epidemiological surveys of Colombia that are publicly available, with each database focused on a specific set of health conditions. For the two sources of genome-wide genotype data, we performed rigorous data quality control when harmonizing the data in an effort to avoid any potential bias. Visual inspection of the genomic relationships among individuals from each of the two genotype data sets does not point to any evidence of batch effects (**Figures 2B,C**).

Another potential limitation of this study relates to the predictive utility of PRS in the kind of diverse, admixed populations such as the Colombian cohort studied here. The vast majority of PRS have been developed and validated on cohorts with European genetic ancestry, and PRS may not transfer well across populations with distinct ancestry profiles. The PRS used here were mined from the NHGRI-EBI GWAS Catalog,[10] and we only used PRS that were developed using cohorts that included two or more diverse ancestries in an effort to mitigate problems with cross-population portability. Nevertheless, the development and validation of PRS on admixed American populations, and for Colombian cohorts in particular, will certainly allow for more robust genetic risk prediction in the future.

Finally, it should be noted that the results here apply to a single admixed American population within a single Latin American country. Latin American populations are highly diverse and characterized by distinct combinations of African, European, and Native American ancestry (Bryc et al., 2010; Ruiz-Linares et al., 2014; Homburger et al., 2015; Norris et al., 2018, 2020). In addition, as we have shown here, Colombia itself is an ethnically diverse country with distinct ancestry profiles for different regions in the country (Rojas Morales et al., 2007; Rishishwar et al., 2015b; Conley et al., 2017; Nagar et al., 2019; Chande et al., 2020a). Thus, the connections between ancestry and health outcomes reported here may or may not apply to other countries or even to different populations sampled within Colombia. Future studies of this kind in other countries, or other populations within Colombia, can shed light on the extent to which the trends observed here are unique or shared among populations.

DATA AVAILABILITY STATEMENT

Colombian ethnicity, disease prevalence, and HDI data are made freely available online as described in the Materials and Methods section, and are provided here as **Supplementary Material**. WGG data reported in this study are available for research purposes by request to the corresponding authors. WGS data are made freely available without use restriction through the 1KGP.[11]

[11]https://www.internationalgenome.org/

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics and Research Committee of the Universidad CES, Ethics Committee of the Universidad Tecnológica del Chocó, and Institutional Review Board of the Georgia Institute of Technology. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AV-A, IJ, and JG conceived and designed the study. AC, SN, and LR performed all data analysis. LM-R, MM-R, AV-A, IJ, and JG supervised and managed all aspects of the project in Colombia and the United States. MM-R and JG acquired study subject samples. AC, SN, LR, and IJ prepared the figures and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

AC, SN, LR, and IJ were supported by the IHRC-Georgia Tech Applied Bioinformatics Laboratory. LM-R was supported by the National Institutes of Health (NIH) Distinguished Scholars Program (DSP) and the Division of Intramural Research (DIR) of the National Institute on Minority Health and Health Disparities (NIMHD) at NIH (1ZIAMD000016 and 1ZIAMD000018). AV-A was supported by Fulbright Colombia. JG was supported by GenomaCES and Universidad CES.

ACKNOWLEDGMENTS

The authors would like to acknowledge the study participants from Medellín, Antioquia, Quibdó, and Chocó.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.690366/full#supplementary-material

REFERENCES

1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Bigdeli, T. B., Genovese, G., Georgakopoulos, P., Meyers, J. L., Peterson, R. E., Iyegbe, C. O., et al. (2020). Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. *Mol. Psychiatry* 25, 2455–2467. doi: 10.1038/s41380-019-0517-y

Bryc, K., Velez, C., Karafet, T., Moreno-Estrada, A., Reynolds, A., Auton, A., et al. (2010). Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc. Natl. Acad. Sci. U. S. A.* 107(Suppl. 2), 8954–8961. doi: 10.1073/pnas.0914618107

Bustamante, C. D., Burchard, E. G., and De la Vega, F. M. (2011). Genomics for the world. *Nature* 475, 163–165. doi: 10.1038/475163a

Bustamente, J., Jaramillo, C., Cárdenas, M. A., Acevedo, L., Freire, E., and Gutiérrez, C. (2012). *Atlas Estadistico.* Bogotá: Departamento Administrativo Nacional de Estadistica.

Chande, A. T., Rishishwar, L., Ban, D., Nagar, S. D., Conley, A. B., Rowell, J., et al. (2020a). The phenotypic consequences of genetic divergence between admixed latin american populations: Antioquia and Choco, Colombia. *Genome Biol. Evol.* 12, 1516–1527. doi: 10.1093/gbe/evaa154

Chande, A. T., Rishishwar, L., Conley, A. B., Valderrama-Aguirre, A., Medina-Rivas, M. A., and Jordan, I. K. (2020b). Ancestry effects on type 2 diabetes genetic risk inference in Hispanic/Latino populations. *BMC Med. Genet.* 21(Suppl 2):132. doi: 10.1186/s12881-020-01068-0

Chande, A. T., Rowell, J., Rishishwar, L., Conley, A. B., Norris, E. T., Valderrama-Aguirre, A., et al. (2017). Influence of genetic ancestry and socioeconomic status on type 2 diabetes in the diverse Colombian populations of Choco and Antioquia. *Sci. Rep.* 7:17127. doi: 10.1038/s41598-017-17380-4

Chande, A. T., Wang, L., Rishishwar, L., Conley, A. B., Norris, E. T., Valderrama-Aguirre, A., et al. (2018). GlobAl distribution of genetic traits (GADGET) web server: polygenic trait scores worldwide. *Nucleic Acids Res.* 46, W121–W126. doi: 10.1093/nar/gky415

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4:7. doi: 10.1186/s13742-015-0047-8

Collins, F. S., and Varmus, H. (2015). A new initiative on precision medicine. *N. Engl. J. Med.* 372, 793–795. doi: 10.1056/NEJMp1500523

Conley, A. B., Rishishwar, L., Norris, E. T., Valderrama-Aguirre, A., Marino-Ramirez, L., Medina-Rivas, M. A., et al. (2017). A comparative analysis of genetic ancestry and admixture in the Colombian populations of Choco and Medellin. *G3* 7, 3435–3447. doi: 10.1534/g3.117.1118

Homburger, J. R., Moreno-Estrada, A., Gignoux, C. R., Nelson, D., Sanchez, E., Ortiz-Tello, P., et al. (2015). Genomic insights into the ancestry and demographic history of South America. *PLoS Genet.* 11:e1005602. doi: 10.1371/journal.pgen.1005602

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* 44, 955–959. doi: 10.1038/ng.2354

Howie, B., Marchini, J., and Stephens, M. (2011). Genotype imputation with thousands of genomes. *G3* 1, 457–470. doi: 10.1534/g3.111.001198

Jameson, J. L., and Longo, D. L. (2015). Precision medicine--personalized, problematic, and promising. *N. Engl. J. Med.* 372, 2229–2234. doi: 10.1056/NEJMsb1503104

Jordan, I. K. (2016). The Columbian exchange as a source of adaptive introgression in human populations. *Biol. Direct* 11:17. doi: 10.1186/s13062-016-0121-x

Khoury, M. J., Bowen, M. S., Clyne, M., Dotson, W. D., Gwinn, M. L., Green, R. F., et al. (2018). From public health genomics to precision public health: a 20-year journey. *Genet. Med.* 20, 574–582. doi: 10.1038/gim.2017.211

Khoury, M. J., Iademarco, M. F., and Riley, W. T. (2016). Precision public health for the era of precision medicine. *Am. J. Prev. Med.* 50, 398–401. doi: 10.1016/j.amepre.2015.08.031

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., et al. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* 100, 635–649. doi: 10.1016/j.ajhg.2017.03.004

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., and Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* 51, 584–591. doi: 10.1038/s41588-019-0379-x

Medina-Rivas, M. A., Norris, E. T., Rishishwar, L., Conley, A. B., Medrano-Trochez, C., Valderrama-Aguirre, A., et al. (2016). Choco, Colombia: a hotspot of human biodiversity. *Rev. Biodivers. Neotrop.* 6, 45–54. doi: 10.18636/bioneotropical.v6i1.341

Nagar, S. D., Moreno, A. M., Norris, E. T., Rishishwar, L., Conley, A. B., O'Neal, K. L., et al. (2019). Population pharmacogenomics for precision public health in Colombia. *Front. Genet.* 10:241. doi: 10.3389/fgene.2019.00241

Norris, E. T., Rishishwar, L., Chande, A. T., Conley, A. B., Ye, K., Valderrama-Aguirre, A., et al. (2020). Admixture-enabled selection for rapid adaptive evolution in the Americas. *Genome Biol.* 21:29. doi: 10.1186/s13059-020-1946-2

Norris, E. T., Wang, L., Conley, A. B., Rishishwar, L., Marino-Ramirez, L., Valderrama-Aguirre, A., et al. (2018). Genetic ancestry, admixture and health determinants in Latin America. *BMC Genomics* 19(Suppl. 8):861. doi: 10.1186/s12864-018-5195-7

Pato, M. T., Sobell, J. L., Medeiros, H., Abbott, C., Sklar, B. M., Buckley, P. F., et al. (2013). The genomic psychiatry cohort: partners in discovery. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 162B, 306–312. doi: 10.1002/ajmg.b.32160

Petrovski, S., and Goldstein, D. B. (2016). Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. *Genome Biol.* 17:157. doi: 10.1186/s13059-016-1016-y

Popejoy, A. B., and Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature* 538, 161–164. doi: 10.1038/538161a

R Core Team (2020). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Rishishwar, L., Conley, A. B., Vidakovic, B., and Jordan, I. K. (2015a). A combined evidence Bayesian method for human ancestry inference applied to Afro-Colombians. *Gene* 574, 345–351. doi: 10.1016/j.gene.2015.08.015

Rishishwar, L., Conley, A. B., Wigington, C. H., Wang, L., Valderrama-Aguirre, A., and Jordan, I. K. (2015b). Ancestry, admixture and fitness in Colombian genomes. *Sci. Rep.* 5:12376. doi: 10.1038/srep12376

Rojas Morales, E., Fernández Ayala, P. J., Gallo Mejía, H., Serna Rios, C., and Poveda Gómez, M. (2007). *Colombia Una Nación Multicultural: Su Diversidad Étnia*. Bogotá: DANE.

Ruiz-Linares, A., Adhikari, K., Acuna-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., et al. (2014). Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet.* 10:e1004572. doi: 10.1371/journal.pgen.1004572

Weeramanthri, T. S., Dawkins, H. J. S., Baynam, G., Bellgard, M., Gudes, O., and Semmens, J. B. (2018). Editorial: precision public health. *Front. Public Health* 6:121. doi: 10.3389/fpubh.2018.00121

Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

MINI REVIEW
published: 20 September 2021
doi: 10.3389/fgene.2021.687116

The Gene-Drug Duality: Exploring the Pharmacogenomics of Indigenous Populations

Shivashankar H. Nagaraj[1]* *and Maree Toombs*[2]

[1] Centre for Genomics and Personalised Health, School of Biomedical Sciences, Faculty of Health, Queensland University of Technology, Brisbane, QLD, Australia, [2] School of Public Health, Faculty of Medicine, The University of Queensland, Herston, QLD, Australia

While pharmacogenomic studies have facilitated the rapid expansion of personalized medicine, the benefits of these findings have not been evenly distributed. Genomic datasets pertaining to Indigenous populations are sorely lacking, leaving members of these communities at a higher risk of adverse drug reactions (ADRs), and associated negative outcomes. Australia has one of the largest Indigenous populations in the world. Pharmacogenomic studies of these diverse Indigenous Australian populations have been hampered by a paucity of data. In this article, we discuss the history of pharmacogenomics and highlight the inequalities that must be addressed to ensure equal access to pharmacogenomic-based healthcare. We also review efforts to conduct the pharmacogenomic profiling of chronic diseases among Australian Indigenous populations and survey the impact of the lack of drug safety-related information on potential ADRs among individuals in these communities.

Keywords: pharmacogenomics, indigenous health, personalized medicine, chronic disease, adverse drug reaction

OPEN ACCESS

Edited by:
Segun Fatumo,
University of London, United Kingdom

Reviewed by:
Aniefiok John Udoakang,
University of Ghana, Ghana
Michelle Whirl-Carrillo,
Stanford University, United States
Iffat Elbarazi,
United Arab Emirates University,
United Arab Emirates

***Correspondence:**
Shivashankar H. Nagaraj
shiv.nagaraj@qut.edu.au

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *29 March 2021*
Accepted: *28 July 2021*
Published: *20 September 2021*

Citation:
Nagaraj SH and Toombs M (2021) The Gene-Drug Duality: Exploring the Pharmacogenomics of Indigenous Populations. Front. Genet. 12:687116. doi: 10.3389/fgene.2021.687116

INTRODUCTION

Pharmacogenomics, the "science of personalized medicine," is a branch of genetics research that focuses on predicting a given individual's responses to specific therapeutic drugs. The term "pharmacogenetics" was coined in 1959 by the German geneticist Friedrich Vogel to describe studies of individual genes and their impacts on drug activity (Kalow, 2006). Over time, pharmacogenetics evolved into the field of pharmacogenomics owing to the realization that most drug responses are multifactorial (Kalow, 2006).

The recent development of more potent and efficacious medications has highlighted the extreme variability in drug responses among individuals, leading to potentially life-threatening adverse drug reactions (ADRs), underscoring the need for a personalized approach to medicine. Pharmacogenomics studies seek to elucidate the genetic factors that influence a given individual's responses to drugs and susceptibility to ADRs. By better understanding these factors, clinicians can more safely deliver effective medications tailored to a given individual's genetic makeup (Kalow, 2006; Crews et al., 2012).

While there have been remarkable advancements in global genomic research during the last three decades, Indigenous populations have been severely underrepresented in most national and international pharmacogenomic studies (Hudson et al., 2020). Consequently, extant genomic data may not be relevant to these Indigenous individuals. Issues of trust, accountability, transparency,

equity, lack of ethical frameworks, community outreach, and proper research guidelines have historically undermined global Indigenous genetic research efforts (Anderson et al., 2016; Hudson et al., 2020).

Australia has one of the largest Indigenous populations in the world, with Indigenous individuals from culturally and ancestrally distinct communities and over 250 different language groups comprising approximately three percent of the total Australian population (Australian Bureau of Statistics, 2021). Pharmacogenomic studies of these diverse Indigenous Australian populations have been hampered by several issues, including ethical concerns and lack of transparency resulting in the introduction of moratoriums on Indigenous genetic research by academic institutions such as the Australian National University in the 1990s (Claw et al., 2018; Australian Human Rights Commission, 2021). To improve knowledge regarding the pharmacogenomic landscape of Indigenous Australian populations, an Australian Indigenous reference genome is being created. Such a reference will facilitate the detection of rare unmapped variants unique to this population that cannot be identified using standard non-Indigenous population-based reference genomes. Herein, we review efforts to conduct the pharmacogenomic profiling of chronic diseases among Australian Indigenous populations and survey the potential impact of a lack of drug safety-related information on ADR risk among these communities.

PHARMACOGENOMIC VARIANTS AND DRUG RESPONSES

Drug response variability may be attributable to genetic differences between individuals resulting in variations in drug pharmacokinetics and pharmacodynamics (Lauschke et al., 2019). Germline pharmacogenomic biomarkers are primarily associated with allelic variants linked to variations in drug distribution, absorption, metabolism, and elimination (Lauschke et al., 2019; Ahsan et al., 2020). Differences in drug responses can also arise as a consequence of human leukocyte antigen (HLA) and drug target gene variants (Ahsan et al., 2020). Information regarding these variants can guide efforts to modulate drug efficacy and control ADR incidence (Lauschke et al., 2019; Ahsan et al., 2020). To effectively catalog known allelic variants associated with specific drug responses, the widely utilized "star" (*) allele nomenclature system was developed. Under this system, gene names are appended with a number that denotes a particular allele or haplotype, with * being used to designate the variant allele (e.g., CYP2C19*2). This facilitates consistency among pharmacogenomic studies and clinical groups, ensuring the reliable classification of therapeutically relevant variants (Gaedigk et al., 2018). Currently, the ability to predict therapeutic outcomes and ADR incidence is of key importance in the oncology therapeutic space (Wheeler et al., 2013). To date, regulatory approval for the treatment of multiple cancers has been given to six small molecule inhibitors of epigenetic modifiers (epidrugs) with more under investigation (Yao et al., 2012; Wheeler et al., 2013). Significant variations are also observed in individual responses to psychiatric medications (Butler, 2018). A pharmacogenomic approach to the treatment of mental disorders would thus represent an efficacious means of reducing ADR incidence.

While early pharmacogenomic studies were reliant upon the targeted assessment of specific genotypic variations, recent advances in sequencing and bioinformatics have enabled comprehensive surveys of the pharmacogenomic landscape in specific populations. The Clinical Pharmacogenetics Implementation Consortium (CPIC) has published stringent guidelines for pharmacogenomics studies and for the high-quality annotation of datasets aimed at standardizing results from genome-wide association studies (GWAS; Manichaikul et al., 2010; Caudle et al., 2014). The CPIC currently lists 94 drug-dosage guidelines based upon pharmacogenomic findings, and the US Food and Drug Administration (FDA), 2021 has included pharmacogenomic label information on 372 drugs. As genome sequencing costs continue to decline and the value of personalized medicine becomes more fully realized, these tools will serve as the foundation for expanded pharmacogenomic medicine and guided prescription efforts (**Supplementary Table 1**).

POPULATION PHARMACOGENOMICS

Individual drug responses vary in relation to gene polymorphisms with different ethnic distributions (Ramamoorthy et al., 2015). Differing ethnicities may impact genetic polymorphisms, leading to population-specific differences in drug responses. For example, African Americans are poor responders to angiotensin-converting enzyme inhibitors and β-blockers (Palleria et al., 2013), while patients with Asian ancestry are prescribed lower statin drug doses due to a higher susceptibility to statin-associated myopathy (Brewster and Seedat, 2013). The impact of co-medication depends on drug-drug interactions, concentrations of inhibitors or inducers, and drug half-lives. Potential factors regulating this process may be intrinsic (genetics, metabolism, and elimination) or extrinsic (diet, environmental exposure, and sociocultural differences) (Ramamoorthy et al., 2015). A high frequency (∼4,000-fold higher than Caucasian populations) of homozygous silent mutations in the butyrylcholinesterase gene (BChE; responsible for apnea, respiratory failure, and prolonged paralysis after succinylcholine or mivacurium treatment) has been reported in the Arya Vysya community in India (Liao, 2007). Earlier studies of Eskimos and Persian Jews also found a high prevalence of the autosomal recessive BChE allele mutation (Scott and Wright, 1976; David et al., 2015), resulting in increased incidence of pseudocholinesterase deficiency and likely conferring cardiovascular protection to these high-fat-consuming populations (Kaback et al., 2010). The HLA-B*1502 allele, frequently found among members of the Han Chinese population, confers an elevated risk of Stevens-Johnson syndrome and associated severe toxic epidermal necrolysis when treated with carbamazepine (Pandit et al., 2011). However, despite such distinct pharmacogenomic variations, drug development programs rarely include individuals from these communities thus contributing to the paucity of data on

TABLE 1 | Chronic diseases in Australian indigenous populations.

Illness	Drugs with PGx information	PGx genes involved	Available guidelines	References
Chronic Kidney disease, Hypertension	ACE inhibitors, Losartan, Allopurinol	ACE, ABCB1, CYP2C9, CYP2D6, CYP3A4, CYP3A5, HLA-B	Allopurinol-HLA-B-CPIC	Saito et al., 2016
Coronary heart disease	Vitamin K antagonists, Platelet aggregation inhibitors, Selective Beta blockers, Antiarrhythmics, HMG CoA reductase inhibitors	ACE, ABCB1, CYP2C9, CYP2C19, CYP3A4, CYP3A5, CYP4F2, SLCO1B1, VKORC1	Vitamin K antagonists-CYP2C9, CYP4F2, VKORC1-CPIC, Platelet aggregation inhibitors-CYP2C19-CPIC, HMG CoA reductase inhibitors-SLCO1B1-CPIC	Scott et al., 2013; Ramsey et al., 2014; Johnson et al., 2017
Diabetes	Sulphonamides	CYP2C9	NA	–
Pain	NSAIDS, halogenated Anesthetics, Opioids	CACNA1S, CYP2C8, CYP2C9, CYP2D6, RYR1	NSAIDS-CYP2C9-CPIC, Anesthetics- CACNA1S, RYR1-CPIC, Codeine-CYP2D6-CPIC	Crews et al., 2014; Gonsalves et al., 2019; Theken et al., 2020
Mental disorders	Selective serotonin reuptake inhibitors (SSRI), Tricyclic anti-depressants (TCA), Haloperidol, Diazepines, Risperidone, Atomoxetine	CYP2C19, CYP2D6, CYP3A4	SSRI, TCA-CYP2C19, CYP2D6-CPIC,Atomoxetine-CYP2D6-CPIC	Hicks et al., 2015; Hicks et al., 2017; Brown et al., 2019
Cancer	Purine and Pyrimidine analog, platinum compounds, PEG-interferon-alpha	DPYD, IFNL3, NUDT15, TPMT, XPC	Purines-NUDT15, TPMT-CPIC, Pyrimidines-DPYD-CPIC, PEG-Iα-IFNL3-CPIC	Muir et al., 2014; Amstutz et al., 2018; Relling et al., 2019

Indigenous populations around the world in the context of genetic research.

PHARMACOGENOMICS OF AUSTRALIAN INDIGENOUS POPULATIONS

Australian Indigenous populations face a far greater burden of chronic diseases, substantially decreased life expectancy, and poorer overall general health relative to the non-Indigenous population, consistent with findings for Indigenous populations in other nations (**Table 1**; Thynne and Gabb, 2016; Beks et al., 2019). Most Indigenous Australians suffer from ≥1 chronic diseases (Tucci, 2011; Thynne and Gabb, 2016; Beks et al., 2019). Chronic diseases with a strong environmental and behavioral etiology, such as cardiovascular disease, hypertension, diabetes, obesity, chronic kidney disease, and depression, contribute to 80% of the mortality gap between Indigenous and non-Indigenous Australians under the age of 75 years (Tucci, 2011; Anderson et al., 2016). Kidneys from the Australian Indigenous individuals with no known kidney disease have been shown to have 30% fewer and significantly larger glomeruli than those from non-Indigenous individuals (Hoy et al., 2017; Thomson et al., 2019). Differences in the lipid profiles of Indigenous Australians relative to those of other Australians have also been observed, potentially influencing the efficacy of certain therapeutics (Hoy et al., 2017; Beks et al., 2019). Genetic predispositions have been implicated as factors influencing chronic mental health issues suffered by Indigenous Australians (Hoy et al., 2017; Das et al., 2018; Thomson et al., 2019). At least 20% of the 121 FDA-recognized pharmacogenetic markers are considered to be relevant in clinical practice involving psychiatric drugs (Currid and Mutsatsa, 2013; Butler, 2018; Nasir et al., 2018). Gene analytics and profiling programs have also shown associations between abnormal gene signaling networks and psychiatric illnesses like schizophrenia, bipolar disorder, and autism spectrum disorders (Butler, 2018). In Australia, little progress has been made in advancing the current understanding of these chronic diseases and mental disorders in Indigenous populations or improving the outcomes of affected individuals. More genetic epidemiological research must be conducted to properly address health outcomes among these Indigenous communities. A potential genetic predisposition to specific ADRs due to differences in allele frequencies of cytochrome P450 variants has been described in some earlier studies (Rheault et al., 2019). For example, the CYP2C19 and CYP2D6 genes in remote north western Australian Indigenous populations differ significantly from those observed in Australians of European ancestry, whereas these frequencies were similar to those observed in East Asian populations (Pandit et al., 2011; Tucci, 2011; Currid and Mutsatsa, 2013; Thynne and Gabb, 2016; Hoy et al., 2017; Das et al., 2018; Nasir et al., 2018; Beks et al., 2019; Rheault et al., 2019; Thomson et al., 2019). However, there remains a paucity of data linking ADRs to genetic predispositions in clinical settings, underscoring the urgent necessity of including the assessment and management of potential ADRs in any comprehensive healthcare program (Rheault et al., 2019).

INDIGENOUS PHARMACOGENOMIC RESEARCH AND THE ETHICAL DISSEMINATION OF INFORMATION

Multiple genomic consortiums have attempted to enlist Indigenous participants to map genomic diversity across

populations (**Supplementary Table 2**). However, most of these studies, including the Human Genome Diversity Project and the National Genographic Project, failed to adequately recruit and include Indigenous peoples due to a lack of a well-planned and inclusive ethical framework (Claw et al., 2018). The ongoing All of Us project—a research -project for gathering genetic, environmental, and lifestyle data from over one million US residents sponsored by the US National Institutes of Health, 2021—is currently attempting to engage in appropriate tribal consultations. However, despite initiatives to make genomic data more accessible, public trust across the diverse Indigenous communities is as yet lacking. As Bentley et al. (2017) stated in the context of genomic research in Africa, "public trust, oversight, and long-lasting relationships with communities who participate in genomic research are required to advance both data sharing, and diversity and inclusion—two major components of genomic research that must advance symbiotically for genomic research to benefit all."

Genomic data are more sensitive than other types of health data for Indigenous communities as they can influence traditional and cultural beliefs and also affect identity claims for rights to land and other resources. Targeted policies to promote responsible conduct of research have been developed by some Indigenous communities. Research guidelines (**Supplementary Table 3**) such as the Human Heredity and Health in Africa (H3Africa) guidelines for community engagement in Africa and the Te Mata Ira guidelines for Genomic Research with Maori in New Zealand, help Indigenous communities to develop an optimistic outlook regarding the outcomes of genomic research. The Navajo Nation in the United States has begun to develop a culturally informed genetic research policy replacing the 16-year moratorium on genetic research. Two similar independent initiatives–the "Silent Genomes" in Canada and the "Aotearoa Variome" in New Zealand have undertaken research to collect and curate genomic variation databases to better understand and undertake translational genetic research (Caron et al., 2020). These systematic efforts to engage Indigenous communities in healthcare and genomic research represent the first steps toward bridging the genomic knowledge gap regarding the Indigenous populations of the world, with years if not decades worth of research yet to be completed. Recognizing and understanding the inherent right of Indigenous people to develop socio-economically and culturally is the starting block for relationship building and must be reflected in the research frameworks (Claw et al., 2018; Garrison et al., 2019). One approach toward faciliating a paradigm shift toward equitable benefit sharing would be to ensure that Indigenous people have control over the data from Indigenous populations, including digital sequence information. Notably, Indigenous scholars and policy makers are leading initiatives to improve access to genomic research in health care throughout Canada, New Zealand, Australia, and the United States (Claw et al., 2018; Garrison et al., 2019; Caron et al., 2020). These countries have endorsed the United Nations Declaration on the Rights of Indigenous Peoples (Garrison et al., 2019). Genomics has the potential to revolutionize the way health care is delivered in Australia. Current ongoing research programs at the state and national level are attempting to introduce genomics into the Australian healthcare system. It has been identified that Australian Indigenous people are at risk of being "left out" of these developments due to the inability of researchers to engage with Indigenous communities and conduct genomics research in such settings. Recently, the Indigenous Genomics Health Literacy Project (IG-HeLP), 2021 has been implemented to educate and inform the Australian Aboriginal and Torres Strait Islander consumers and health workers on the topics of DNA, genes, genetic health, genetic testing, and precision medicine. Such efforts will facilitate cultural inclusivity in genomic research and equitable access to the healthcare benefits of clinical genomics across Australia, and thereby, has the potential to shape Australia's health future (GenetiQs Project, 2021; QIMR Berghofer, 2021).

CONCLUSION

Multiple changes in the scientific community are required to ensure that Indigenous populations benefit from genomic research. One method is the development of an Indigenous reference genome for cataloging genetic variants. There are over 500 different clan groups or "nations" in Australia, each with distinctive cultures, beliefs and languages. Ideally, a minimum of one reference genome per clan is required for effective high quality variant calling and to understand their underlying genomic architecture (Shumate et al., 2020). Efforts are ongoing to increase the number of Indigenous researchers working in genomic sciences and to develop novel culturally safe guidelines. Initiatives such as the AHRA, 2021 National Indigenous Research(er) Capacity Building Network (IRNet) are creating opportunities to improve Indigenous pharmacogenomic health literacy, increasing the number of Indigenous researchers working in genomic science, developing novel culturally safe guidelines, policies, and procedures for bridging the genomic divide, and respecting the rights, interests, and opinions of Indigenous participants in relation to their genomic data (Claw et al., 2018; Shumate et al., 2020; NHMRC, 2021). A proactive and comprehensive approach to address drug safety in Indigenous populations in the post-marketing space is also urgently required.

Pharmacogenomic variations remain understudied in Indigenous Australian populations and may have important implications for drug efficacy and ADR risk. This apparent failure, despite numerous funded Indigenous health programs functioning at a national level and under the "Closing the Gap" policy, has been attributed to issues of methodology relating to culturally appropriate acceptable research approaches accounting for the diversity of Indigenous populations, short government funding cycles, and no community ownership and consultation (Anderson et al., 2016; Rheault et al., 2019; Hudson et al., 2020). Initiatives like the National Center for Indigenous Genomics have been established to facilitate participatory, consensual, and ethical genomic research which may help to gradually

close this chronic disease gap affecting Indigenous Australian communities, thereby advancing pharmacogenomic research.

AUTHOR CONTRIBUTIONS

SN created Tables and Figures. Both authors conceived and prepared, read and approved the final manuscript.

FUNDING

Funding for this research was supported by the MRFF Genomics Health Futures Mission grant (76757).

ACKNOWLEDGMENTS

The authors would like to acknowledge Arvind Shankar for critically reading the manuscript and Debasri Mukherjee for editorial support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.687116/full#supplementary-material

REFERENCES

AHRA (2021). *National Indigenous Research(Er) Capacity Building Network*. Avaliable online at: https://ahra.org.au/our-work/national-indigenous-researcher-capacity-building-network-irnet/ (accessed July 16, 2021).

Ahsan, T., Urmi, N. J., and Sajib, A. A. (2020). Heterogeneity in the distribution of 159 drug-response related SNPs in world populations and their genetic relatedness. *PLoS One* 15:e0228000. doi: 10.1371/journal.pone.0228000

Amstutz, U., Henricks, L. M., Offer, S. M., Barbarino, J., Schellens, J. H. M., Swen, J. J., et al. (2018). Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing: 2017 update. *Clin. Pharmacol. Ther.* 103, 210–216. doi: 10.1002/cpt.911

Anderson, I., Robson, B., Connolly, M., Al-Yaman, F., Bjertness, E., King, A., et al. (2016). Indigenous and tribal peoples' health (The Lancet–Lowitja Institute Global Collaboration): a population study. *Lancet* 388, 131–157.

Australian Bureau of Statistics (2021). *Estimates of Aboriginal and Torres Strait Islander Australians, June*. Avaliable online at: https://www.abs.gov.au/statistics/people/ (accessed July 16, 2021).

Australian Human Rights Commission (2021). *The Protection of Genetic Information of Indigenous Peoples*. Avaliable online at: https://humanrights.gov.au/our-work/legal/protection-genetic-information-indigenous-peoples (accessed July 16, 2021).

Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) (2021). *Australian Institute of Aboriginal and Torres Strait Islander Studies*. Avaliable online at: https://aiatsis.gov.au/research/research-publications (Accessed July 16, 2021).

Beks, H., Binder, M. J., Kourbelis, C., Ewing, G., Charles, J., Paradies, Y., et al. (2019). Geographical analysis of evaluated chronic disease programs for aboriginal and torres Strait Islander people in the Australian primary health care setting: a systematic scoping review. *BMC Public Health* 19:1115. doi: 10.1186/s12889-019-7463-0

Bentley, A. R., Callier, S., and Rotimi, C. N. (2017). Diversity and inclusion in genomic research: why the uneven progress? *J. Community Genet*. 8, 255–266.

Brewster, L. M., and Seedat, Y. K. (2013). Why do hypertensive patients of African ancestry respond better to calcium blockers and diuretics than to ACE inhibitors and β-adrenergic blockers? A systematic review. *BMC Med.* 11:141. doi: 10.1186/1741-7015-11-141

Brown, J. T., Bishop, J. R., Sangkuhl, K., Nurmi, E. L., Mueller, D. J., Dinh, J. C., et al. (2019). Clinical pharmacogenetics implementation consortium guideline for cytochrome P450 (CYP)2D6 Genotype and Atomoxetine Therapy. *Clin. Pharmacol. Ther.* 106, 94–102. doi: 10.1002/cpt.1409

Butler, M. (2018). Pharmacogenetics and psychiatric care: a review and commentary. *J. Ment. Health Clin. Psychol.* 2, 17–24. doi: 10.29245/2578-2959/2018/2.1120

Caron, N. R., Chongo, M., Hudson, M., Arbour, L., Wasserman, W. W., Robertson, S., et al. (2020). Indigenous genomic databases: pragmatic considerations and cultural contexts. *Front. Public Health* 8:111. doi: 10.3389/fpubh.2020.00111

Caudle, K., Klein, T., Hoffman, J., Muller, D., Whirl-Carrillo, M., Gong, L., et al. (2014). Incorporation of pharmacogenomics into routine clinical practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline Development Process. *Curr. Drug Metab.* 15, 209–217. doi: 10.2174/1389200215666140130124910

Claw, K. G., Anderson, M. Z., Begay, R. L., Tsosie, K. S., Fox, K., Garrison, N. A., et al. (2018). A framework for enhancing ethical genomic research with Indigenous communities. *Nat. Commun.* 9:2957.

Crews, K. R., Gaedigk, A., Dunnenberger, H. M., Leeder, J. S., Klein, T. E., Caudle, K. E., et al. (2014). Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450 2D6 genotype and codeine therapy: 2014 update. *Clin. Pharmacol. Ther.* 95, 376–382. doi: 10.1038/clpt.2013.254

Crews, K. R., Hicks, J. K., Pui, C. H., Relling, M. V., and Evans, W. E. (2012). Pharmacogenomics and individualized medicine: Translating science into practice. *Clin. Pharmacol. Ther.* 92, 467–475.

Currid, T. J., and Mutsatsa, S. (2013). What effect will pharmacogenomics have on mental health nursing? *Neuropsychiatry*. 3, 309–320.

Das, M., Kini, R., Garg, G., and Parker, R. (2018). Australian aboriginal and torres strait islanders' mental health issues: a litany of social causation. *Indian J. Soc. Psychiatry* 34, 328–328.

David, S. M., Soundararajan, L., and Boopathy, R. (2015). The allele frequency of T920C mutation in butyrylcholinesterase gene is high in an Indian population. *Gene* 555, 409–413. doi: 10.1016/j.gene.2014.11.040

Gaedigk, A., Ingelman-Sundberg, M., Miller, N. A., Leeder, J. S., Whirl-Carrillo, M., Klein, T. E., et al. (2018). The Pharmacogene Variation (PharmVar) consortium: incorporation of the human cytochrome P450 (CYP) allele nomenclature database. *Clin. Pharmacol. Ther.* 103, 399–401. doi: 10.1002/cpt.910

Garrison, N. A., Hudson, M., Ballantyne, L. L., Garba, I., Martinez, A., Taualii, M., et al. (2019). genomic research through an indigenous lens: understanding the expectations. *Annu. Rev. Genomics Hum. Genet.* 20, 495–517. doi: 10.1146/annurev-genom-083118-015434

GenetiQs Project (2021). *GenetiQs Project*. Avaliable online at: https://www.qimrberghofer.edu.au/our-research/aboriginal-torres-strait-islander-health/genetiqs-project/ (accessed July16, 2021).

Gonsalves, S. G., Dirksen, R. T., Sangkuhl, K., Pulk, R., Alvarellos, M., Vo, T., et al. (2019). Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for the use of potent volatile anesthetic agents and succinylcholine in the context of RYR1 or CACNA1S genotypes. *Clin. Pharmacol. Ther.* 105, 1338–1344. doi: 10.1002/cpt.1319

Hicks, J. K., Bishop, J. R., Sangkuhl, K., Müller, D. J., Ji, Y., Leckband, S. G., et al. (2015). Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for CYP2D6 and CYP2C19 genotypes and dosing of selective serotonin reuptake inhibitors. *Clin. Pharmacol. Ther.* 98, 127–134. doi: 10.1002/cpt.147

Hicks, J. K., Sangkuhl, K., Swen, J. J., Ellingrod, V. L., Müller, D. J., Shimoda, K., et al. (2017). Clinical pharmacogenetics implementation consortium guideline (CPIC) for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants: 2016 update. *Clin. Pharmacol. Ther.* 102, 37–44. doi: 10.1002/cpt.597

Hoy, W. E., Kopp, J. B., Mott, S. A., and Winkler, C. A. (2017). Absence of APOL1 risk alleles in a remote living Australian Aboriginal group with high rates of CKD, hypertension, diabetes, and cardiovascular disease. *Kidney Int.* 91:990. doi: 10.1016/j.kint.2016.11.031

Hudson, M., Garrison, N. A., Sterling, R., Caron, N. R., Fox, K., Yracheta, J., et al. (2020). Rights, interests and expectations: Indigenous perspectives on unrestricted access to genomic data. *Nat. Res.* 21, 377–384. doi: 10.1038/s41576-020-0228-x

Indigenous Genomics Health Literacy Project (IG-HeLP) (2021). *Indigenous Genomics Health Literacy Project*. Avaliable online at: https://www.qimrberghofer.edu.au/our-research/aboriginal-torres-strait-islander-health/ig-help/ (accessed July 16, 2021).

Johnson, J. A., Caudle, K. E., Gong, L., Whirl-Carrillo, M., Stein, C. M., Scott, S. A., et al. (2017). Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for pharmacogenetics-guided warfarin dosing: 2017 update. *Clin. Pharmacol. Ther.* 102, 397–404. doi: 10.1002/cpt.668

Kaback, M., Lopatequi, J., Portuges, A. R., Quindipan, C., Pariani, M., Salimpour-Davidov, N., et al. (2010). Genetic screening in the persian jewish community: a pilot study. *Genetics Med.* 12, 628–633. doi: 10.1097/gim.0b013e3181edef5b

Kalow, W. (2006). Pharmacogenetics and pharmacogenomics: origin, status, and the hope for personalized medicine. *Pharmacogenomics J* 6, 162–165. doi: 10.1038/sj.tpj.6500361

Lauschke, V. M., Zhou, Y., and Ingelman-Sundberg, M. (2019). Novel genetic and epigenetic factors of importance for inter-individual differences in drug disposition, response and toxicity. *Pharmacol. Ther.* 197, 122–152. doi: 10.1016/j.pharmthera.2019.01.002

Liao, J. K. (2007). Safety and efficacy of statins in Asians. *Am. J. Cardiol.* 99, 410–414. doi: 10.1016/j.amjcard.2006.08.051

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., and Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867–2873. doi: 10.1093/bioinformatics/btq559

Muir, A. J., Gong, L., Johnson, S. G., Lee, M. T., Williams, M. S., Klein, T. E., et al. (2014). Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for IFNL3 (IL28B) genotype and PEG interferon-alpha-based regimens. *Clin. Pharmacol. Ther.* 95, 141–146. doi: 10.1038/clpt.2013.203

Nasir, B. F., Toombs, M. R., Kondalsamy-Chennakesavan, S., Kisely, S., Gill, N. S., Black, E., et al. (2018). Common mental disorders among Indigenous people living in regional, remote and metropolitan Australia: a cross-sectional study. *BMJ Open* 8:e020196. doi: 10.1136/bmjopen-2017-020196

National Institutes of Health (2021). *All of Us Tribal Engagement. National Institutes of Health All of Us Research Program*. Avaliable online at: https://allofus.nih.gov/about/diversity-and-inclusion/all-us-tribal-engagement (accessed July 16, 2021).

NHMRC (2021). *Ethical Conduct in Research With Aboriginal and Torres Strait Islander Peoples and Communities*. Avaliable online at: https://www.nhmrc.gov.au/about-us/resources/ethical-conduct-research-aboriginal-and-torres-strait-islander-peoples-and-communities (accessed July 16, 2021).

Palleria, C., Di Paolo, A., Giofrè, C., Caglioti, C., Leuzzi, G., Siniscalchi, A., et al. (2013). Pharmacokinetic drug-drug interaction and their implication in clinical management. *J. Res. Med. Sci.* 18, 601–610.

Pandit, J. J., Gopa, S., and Arora, J. (2011). A hypothesis to explain the high prevalence of pseudo-cholinesterase deficiency in specific population groups. *Eur. J. Anaesthesiol.* 28, 550–552. doi: 10.1097/eja.0b013e3283457cfb

QIMR Berghofer (2021). *Guidelines for Genomic Research With Aboriginal and Torres Strait Islander Peoples of Queensland*. Avaliable online at: https://www.qimrberghofer.edu.au/wp-content/uploads/2020/09/2019-Indigenous-Health-Genomics-Guide-v9-WEB.pdf (accessed July 16, 2021).

Ramamoorthy, A., Pacanowski, M. A., Bull, J., and Zhang, L. (2015). Racial/ethnic differences in drug disposition and response: review of recently approved drugs. *Nat. Publis. Group* 97, 263–273. doi: 10.1002/cpt.61

Ramsey, L. B., Johnson, S. G., Caudle, K. E., Haidar, C. E., Voora, D., Wilke, R. A., et al. (2014). The clinical pharmacogenetics implementation consortium guideline for SLCO1B1 and simvastatin-induced myopathy: 2014 update. *Clin. Pharmacol. Ther.* 96, 423–428. doi: 10.1038/clpt.2014.125

Relling, M. V., Schwab, M., Whirl-Carrillo, M., Suarez-Kurtz, G., Pui, C. H., Stein, C. M., et al. (2019). Clinical Pharmacogenetics Implementation Consortium Guideline for thiopurine dosing based on TPMT and NUDT15 Genotypes: 2018 update. *Clin. Pharmacol. Ther.* 105, 1095–1105. doi: 10.1002/cpt.1304

Rheault, H., Coyer, F., Jones, L., and Bonner, A. (2019). Health literacy in Indigenous people with chronic disease living in remote Australia. *BMC Health Services Res.* 19:523. doi: 10.1186/s12913-019-4335-3

Saito, Y., Stamp, L. K., Caudle, K. E., Hershfield, M. S., McDonagh, E. M., Callaghan, J. T., et al. (2016). Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for human leukocyte antigen B (HLA-B) genotype and allopurinol dosing: 2015 update. *Clin. Pharmacol. Ther.* 99, 36–37. doi: 10.1002/cpt.161

Scott, E. M., and Wright, R. C. (1976). A third type of serum cholinesterase deficiency in Eskimos. *Am. J. Hum. Genet.* 28, 253–256.

Scott, S. A., Sangkuhl, K., Stein, C. M., Hulot, J. S., Mega, J. L., Roden, D. M., et al. (2013). Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. *Clin. Pharmacol. Ther.* 94, 317–323. doi: 10.1038/clpt.2013.105

Shumate, A., Zimin, A. V., Sherman, R. M., Puiu, D., Wagner, J. M., Olson, N. D., et al. (2020). Assembly and annotation of an Ashkenazi human reference genome. *Genome Biol.* 21:129.

Theken, K. N., Lee, C. R., Gong, L., Caudle, K. E., Formea, C. M., Gaedigk, A., et al. (2020). Clinical Pharmacogenetics Implementation Consortium Guideline (CPIC) for CYP2C9 and nonsteroidal anti-inflammatory drugs. *Clin. Pharmacol. Ther.* 108, 191–200. doi: 10.1002/cpt.1830

Thomson, R. J., Mcmorran, B., Hoy, W., Jose, M., Whittock, L., Thornton, T., et al. (2019). New genetic loci associated with chronic kidney disease in an indigenous Australian population. *Front. Genet.* 10:330. doi: 10.3389/fgene.2019.00330

Thynne, T., and Gabb, G. M. (2016). Therapeutic drug safety for Indigenous Australians: how do we close the gap? *Med. J. Aust.* 204, 16–17. doi: 10.5694/mja15.00937

Tucci, J. D. (2011). Pharmacogenetic screening of indigenous australians. *Med. J. Aust.* 195, 386–387. doi: 10.5694/mja11.10678

US Food and Drug Administration (FDA) (2021). *Table of Pharmacogenomic Biomarkers in Drug Labeling*. Avaliable online at: https://www.fda.gov/drugs/science-and-research-drugs/table-pharmacogenomic-biomarkers-drug-labeling (accessed July 16, 2021).

Wheeler, H. E., Maitland, M. L., Dolan, M. E., Cox, N. J., and Ratain, M. J. (2013). Cancer pharmacogenomics: strategies and challenges. *Nat. Rev. Genet.* 14, 23–34. doi: 10.1038/nrg3352

Yao, S., Maghsoudlou, D., and Ambrosone, C. B. (2012). Breast cancer pharmacogenetics in the era of personalized medicine. *Curr. Breast Cancer Rep.* 4, 271–281. doi: 10.1007/s12609-012-0088-2

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

SYSTEMATIC REVIEW
published: 20 October 2021
doi: 10.3389/fgene.2021.699445

Systematic Review of Genomic Associations with Blood Pressure and Hypertension in Populations with African-Ancestry

S. Singh[1,2]*, *J-T. Brandenburg*[1], *A. Choudhury*[1], *F.X. Gómez-Olivé*[3] *and M. Ramsay*[1,2]*

[1]*Sydney Brenner Institute for Molecular Bioscience (SBIMB), Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa,* [2]*Division of Human Genetics, School of Pathology, National Health Laboratory Service and Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa,* [3]*MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, University of the Witwatersrand, Johannesburg, South Africa*

OPEN ACCESS

Edited by:
Karoline Kuchenbaecker, University College London, United Kingdom

Reviewed by:
Todd L Edwards, Vanderbilt University Medical Center, United States
Taobo Hu, Peking University People's Hospital, China

***Correspondence:**
S. Singh
surinsingh@hotmail.co.za
M. Ramsay
Michele.Ramsay@wits.ac.za

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *23 April 2021*
Accepted: *10 September 2021*
Published: *20 October 2021*

Citation:
Singh S, Brandenburg J-T, Choudhury A, Gómez-Olivé FX and Ramsay M (2021) Systematic Review of Genomic Associations with Blood Pressure and Hypertension in Populations with African-Ancestry. Front. Genet. 12:699445. doi: 10.3389/fgene.2021.699445

Background: Despite hypertension being highly prevalent in individuals with African-ancestry, they are under-represented in large genome-wide association studies. Inclusion of African participants is essential to better understand genetic associations with blood pressure-related traits in Africans. This systematic review critically evaluates existing studies with African-ancestry participants and identifies knowledge gaps.

Methods: We followed the PRISMA protocol, HuGE Review handbook to identify literature on original research, in English, on genetic association studies for blood pressure-related traits (systolic and diastolic blood pressure, pulse and mean-arterial pressure, and hypertension) in populations with African-ancestry (January 2007 to April 2020). A narrative synthesis of the evidence was conducted.

Results: Twelve studies with African-ancestry participants met the eligibility criteria, within which 10 studies met the additional genetic association data criteria (i.e., reporting only on African-ancestry participants). Across the five blood pressure-related traits, 26 genome-wide significantly associated SNPs were identified, with six SNPs linked to more than one trait, illustrating pleiotropic effects. Among the SNP associations, 12 had not previously been described in non-African studies.

Abbreviations: AA, African-American; AHM, antihypertensive medication; AWI-Gen Africa Wits-INDEPTH Partnership for Genomic Research; BMI, body mass index; BP, blood pressure; CARe, candidate gene association resource; COGENT, continental origins and genetic epidemiology network; DBP, diastolic blood pressure; pressure in the arteries when the heart rests between beats, measured as a continuous trait; GW, genome-wide; GWAS, genome-wide association study; HTN, hypertension; JNC7, guidelines were used i.e., HTN if SBP ≥140 mmHg and/or DBP>= 90 mmHg and/or treatment with AHM, measured as a binary trait; LD, linkage disequilibrium; LMM, linear mixed model; LoRM, logistic regression model; LRM, linear regression model; MAF, minor allele frequency; MAP, mean arterial pressure; sum of DBP and a third of the PP (i.e., DBP +1/3 PP), measured as a continuous trait; PC, principle component; PCA, principle components analysis; PP, pulse pressure; difference between the SBP and DBP (i.e., SBP-DBP), measured as a continuous trait; SBP, systolic blood pressure; pressure in the arteries when the heart beats, measured as continuous trait; SHet extension of SHom and it can increase the statistical power over SHom when a variant affects only a subset of traits; The distribution of SHet under the null hypothesis is obtained through an estimated beta distribution; SHom, Similar to fixed effect meta-analysis method but accounts for the correlation of summary statistics of the multi-traits and for overlapping or related samples among the cohorts; Uses the trait sample size as the weight, so that it is possible to combine traits with different measurement scales; SNP, single nucleotide polymorphism.

Discussion: The limited number of relevant studies highlights the dearth of genomic association studies on participants with African-ancestry, especially those located within Africa. Variations in study methodology, participant inclusion, adjustment for covariates (e.g., antihypertensive medication) and relatively small sample sizes make comparisons challenging, and have resulted in fewer significant associations, compared to large European studies. Regional variation in the prevalence and associated risk factors of hypertension across Africa makes a compelling argument to develop African cohorts to facilitate large genomic studies, using African-centric arrays. Data harmonisation and comparable study designs, such as described in the H3Africa CHAIR initiative, provide a good example toward achieving this goal.

Other relevant information: SS and J-TB were funded by the South African National Research Foundation. MR is a South African Research Chair in Genomics and Bioinformatics of African populations hosted by the University of the Witwatersrand, funded by the Department of Science and Innovation, and administered by the NRF. This review was registered at PROSPERO (registration number: CRD42020179221) and OSF (registration DOI: 10.17605/OSF.IO/QT2HA).

Keywords: African-ancestry, blood pressure, genetic association, genome-wide association study, hypertension, systematic review

INTRODUCTION

Hypertension (HTN) or high blood pressure is a risk factor for premature deaths and disability worldwide (Forouzanfar et al., 2016; Juma et al., 2016), accounting for 17.9 million deaths in 2018. It is present in approximately 22% of the global population (affecting approximately 1 in 4 men and 1 in 5 women), with the highest prevalence observed in Africa (27%) and lowest in America (average of 18%, where North America < 20% and South America 20–24.9%) (WHO, 2016; 2018).

Over the past decade, the genome-wide association study (GWAS) approach has become the preferred approach to identify genetic factors associated with complex diseases and traits, such as HTN and blood pressure (BP) (Tam et al., 2019). GWAS aims to identify genetic variants and genes associated with the regulation and control of BP. The first case-control studies for HTN and quantitative BP phenotypes were conducted in 2007 as part of the Wellcome Trust Case Control Consortium (Burton et al., 2007) and the Framingham Heart Study (Levy et al., 2007). The GWAS Catalog currently includes 5,152 genetic associations with BP based on 164 studies and 589 associations with HTN based on 76 studies (https://www.ebi.ac.uk/gwas, SNP (single nucleotide polymorphisms) trait p-value < 1E-5, accessed April 17, 2020).

The importance of increasing the representation of diverse populations in genomic research, including Africans, has been addressed in many studies (Ramsay et al., 2016; Martin et al., 2018; Popejoy et al., 2018; Bentley et al., 2020). Most GWAS studies have focused on the European ancestry populations (Azam and Azizan, 2018), and studies based on African-ancestry participants comprise mainly African Americans (AAs) (Manolio et al., 2009; Franceschini et al., 2013; Franceschini et al., 2016; Taylor et al., 2016; Liang et al., 2017). The first GWAS for HTN in AAs was conducted by Adeyemo et al. (2009). Larger studies are needed to understand the role of genetic variants and gene-environment interactions on BP and related phenotypes in African populations (Hendry et al., 2018).

This review critically evaluates the literature on GWASs for BP and HTN in participants with African-ancestry, using PubMed, PubMed Central (PMC), Scopus, Web of Science (WoS) and the GWAS Catalog databases.

METHODS

Protocol and Registration

The systematic review was conducted using: 1) The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) framework (Moher et al., 2009), 2) HuGENET™ (HuGE) Review handbook (Little et al., 2006), and 3) Systematic reviews of genetic association studies (Sagoo et al., 2009). The PRISMA checklist can be found in **Supplementary Table S1**.

To prevent any duplication of systematic reviews based on the topics addressed in this paper, the Cochrane Library (https://www.cochranelibrary.com), PROSPERO (the international prospective register of systematic review, https://www.crd.york.ac.uk/prospero), OSF (Open Science Framework, https://osf.io) and the Research Registry databases (https://www.researchregistry.com) were used to check for similar reviews. This systematic review was registered at PROSPERO database (registration number: CRD42020179221. Available at: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020179221). The systematic review was also registered with the OSF as recommended by PROSPERO

(registration DOI: 10.17605/OSF.IO/QT2HA, available at: https://osf.io/wg2ty), due to the large turnover time for PROSPERO (~ 4 months).

Eligibility Criteria

The eligibility criteria were based on study and report characteristics, outlined in Table S 2 (inclusion of articles were based on GWAS with BP traits and/or HTN with African-ancestries). Studies were excluded if they were: 1) not based on the inclusion criteria (Irrelevant articles, i.e., did not include GWAS, African, Human, BP or HTN); 2) published prior to 2007; 3) not based on whole genome arrays (Targeted or Non-genome-wide arrays); 4) not based on conducting a discovery GWAS within the study or focused on environmental interactions (Replication Study Only, Method Only, GxE Only); 4) focused on the effect of antihypertensive medication treatment (Clinical Study); or 5) not an experimental study (Reviews, Non-Journal Articles).

Additional eligibility criteria were included for genetic association data: 1) Meta-analyses across multiple studies (based on multi-ancestries) were included for only African-ancestry data (studies excluded if they were not separated by ethnicity). 2) Secondary studies (based on the same consortium) were included if the trait was different from that described in the primary study (studies excluded if the trait was previously reported by primary study).

Database Search Strategy

PubMed, Scopus and WoS were searched for relevant papers and additional papers were found using GWAS Catalog and PMC, published between January 2007 and April 2020 (search date: April 8, 2020). Publication references, from papers that met the eligibility criteria, were further searched to improve the search strategy.

Pubmed, PMC, Scopus and WoS: Search criteria are shown in **Supplementary Table S3**: Database search strategy. The query terms consisted of words related to BP and/or HTN, GWAS, genetic association and African-ancestry. MeSH (Medical Subject Headings) terms were searched and included for each keyword. Boolean operators such as AND/OR/NOT were used to string terms together. The initial PubMed search was without filter restrictions. Whereas the Scopus, WoS and PMC searches were further filtered by inclusion of keywords in the abstract, to get papers that were more relevant (**Supplementary Table S3**).

GWAS Catalog: Papers based on the regular expression "*blood pressure*" and "hypertens*" (trait codes: EFO_0004325, EFO_0000537) were searched in the GWAS database, in the "MAPPED_TRAIT" column. These traits included anything related to BP such as HTN, systolic BP (SBP), diastolic BP (DBP), pulse pressure (PP) and mean-arterial pressure (MAP). The study database was further filtered in "BROAD.ANCESTRAL.CATEGORY" column, using "African*" regular expression, to include only studies with African-ancestry participants. The inclusion criteria for papers within the GWAS catalog can be found at https://www.ebi.ac.uk/gwas/docs/methods/criteria.

Data Collection

Published studies were identified through database searching, against the selection criteria and screened by titles and abstracts. Full articles were obtained for studies that met the inclusion criteria and where inclusion was uncertain. SS recorded reasons for exclusion (full details in **Supplementary Table S4**: Inclusion/exclusion criteria of the screened records). Screening of the full-text articles, and selected studies in accordance with the inclusion criteria was conducted by SS and checked by RM, BJ, and CA, with final inclusion by consensus.

Data Extraction and Evaluation

Data was extracted from the included studies in the form of summary tables. All authors screened the extracted data and ambiguities were resolved through discussion. Data was evaluated by following the guidelines defined by Sagoo et al. (2009) i.e., assessment by: risk of bias in individual studies and across studies case/trait adjustment definition (such as HTN, anti-hypertensive medication (AHM) and covariant adjustment), population stratification, reporting of methods used (sample size of a study population, genotyping method and its reliability/accuracy, validation of results, statistical analyses). This reduced publication bias within this systematic review (avoiding risk of bias caused by selective reporting).

Summary Measures and Synthesis of Results

The summary of population characteristics and phenotype measures for BP traits for the eligible studies were derived from total population calculations for n and percentage average calculations for mean and SD (**Supplementary Table S5**). Population characteristics were reported only for African-ancestry. Phenotype measures were reported for BP traits i.e., SBP, DBP, PP and MAP (the effect of continuous variables was measured by beta values) and HTN (the effect of binary variables was measured by odds ratio). BP multi-trait analysis, for the SHet and SHom methods were also reported. The summary of genetic associations was reported for only genome-wide associations ($p < 5E-8$) (Suggestive associations as per study definition were included in supplementary tables, **Supplementary Table S6**).

Additional Analyses

The genetic associations were compared with findings in non-African-ancestry studies reported in the GWAS Catalog (under build 37) and further compared with Evangelou et al. (2018) summary statistics data (currently the largest published study with 757,601 European ancestry individuals), to determine which signals were uniquely identified in studies with African-ancestry populations.

RESULTS

Few studies were identified on genetic association of BP and HTN in African-ancestry populations, using the GWAS approach.

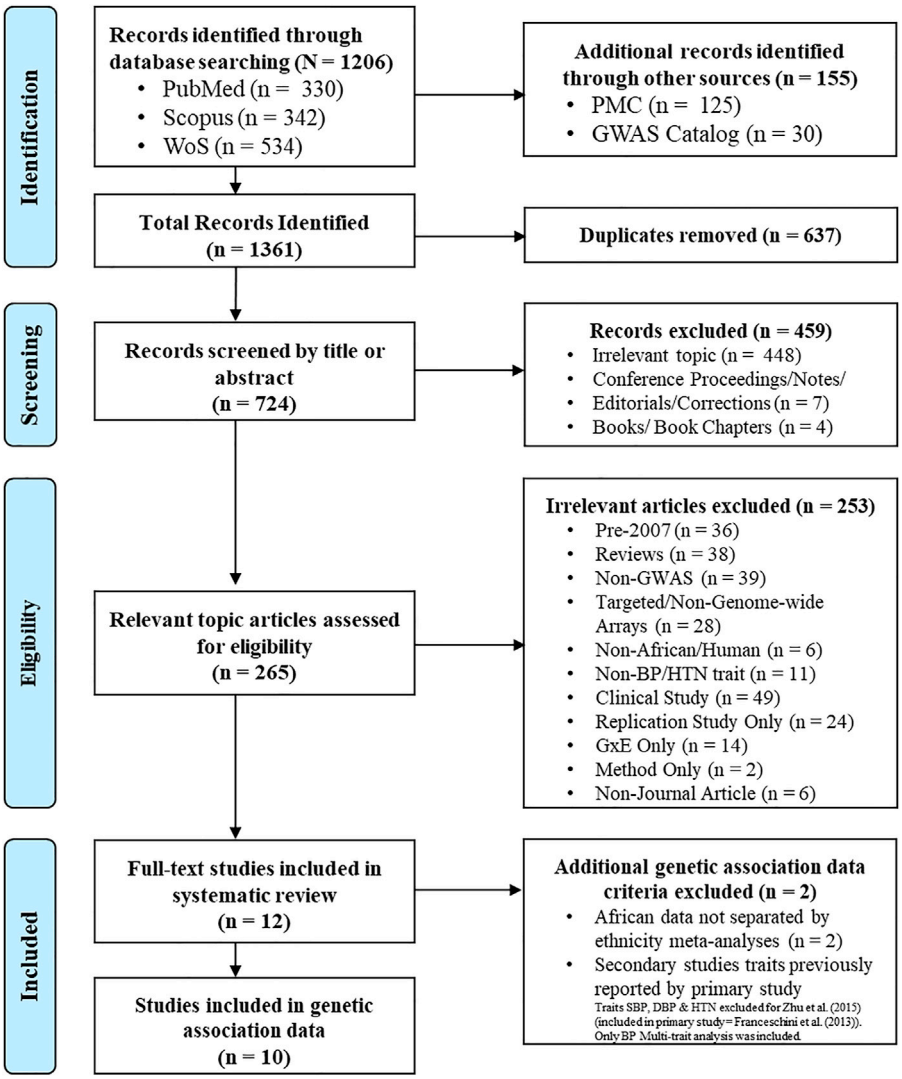

FIGURE 1 | Study schema. Study inclusion/exclusion criteria process, based on PRISMA 2009 flow diagram (Moher et al., 2009). Databases: PubMed (https://pubmed.ncbi.nlm.nih.gov), Scopus (http://www.scopus.com), WoS (http://thomsonreuters.com/thomson-reuters-web-of-science), PMC (https://www.ncbi.nlm.nih.gov/pmc) and GWAS Catalog (https://www.ebi.ac.uk/gwas).

Therefore, a narrative synthesis of the results was conducted, instead of a meta-analysis.

Selection of Publications for Inclusion

A total of 1,361 papers were identified (PubMed = 330, Scopus = 342, WoS = 534, PMC = 125 and GWAS Catalog = 30, publication references = 0). After removing duplicates, 724 papers remained eligible for screening. After screening by title, abstract and full-text, in cases where it was not clear if inclusion criteria had been met, only 12 papers were considered eligible and retained for the systematic review (**Figure 1**).

The first GWAS with BP and HTN in an African-ancestry population was conducted in 2009 (Adeyemo et al., 2009), with the most recent being conducted in 2020 lines (Willems et al., 2020). Over the years, the number of publications increased (**Supplementary Figure S1**).

Cohort Characteristics

The African-ancestry population size ranged between 815 (Lule et al., 2019) to 31968 (Liang et al., 2017) individuals (**Table 1**). Most studies were conducted on AAs (study setting = United States for nine studies) (**Table 1**). Four studies were from participants living in Africa i.e., Nigeria (Franceschini et al., 2013; Zhu et al., 2015) and Uganda (Gurdasani et al., 2019; Lule et al., 2019). All studies, where age was reported, included only adults (> 18 years), with the exception of Lule et al. (2019), which included 10–11 year old Ugandan adolescents. Meta-analyses were conducted for most studies (**Supplementary Table S5**), with exception of two (Adeyemo et al., 2009; Lule et al., 2019). Most studies included more women (**Supplementary Table S5**), with exception of Giri et al. (2019) and Lule et al. (2019) which had 8.5 and 48.7% women respectively.

TABLE 1 | Study characteristics of the 12 eligible publications.

Publication Year	Database pull	Age group (Years)	Study setting/region	Initial size (N)	Male composition (%)	Composition	African-ancestry size (n)	African-ancestry Composition	Ethnicity group	Africans ancestry only (%)	Meta-analysis	Consortium	References
2009	GWAS Catalog; PM; PMC; Scopus; WoS	NA	United States	1,017	41.2	509 AA cases, 508 AA controls	1,017	509 AA cases, 508 AA controls	AA	100	No	HUFS	Adeyemo et al. (2009)
2011	GWAS Catalog; PM; PMC; Scopus; WoS	>18	United States	7473	39.7	7,473 AA	7473	7,473 AA	AA	100	Yes	CARe (1a)	Fox et al. (2011)
2013	GWAS Catalog; PM; PMC; Scopus; WoS	≥ 20	United States; Nigera	29378	27.1	28,190 AA, 1,188 Nigerian	29378	28,190 AA, 1,188 Nigerian	AA; African-Ancestry (Nigerian)	100	Yes	COGENT (2a)	Franceschini et al. (2013)
2019	GWAS Catalog; PM	≥ 18	NR; United Kingdom	4,59,777	91.5	MVP (N = 318,891): whites (N = 5,265), blacks (N = 4,671), Asians (N = 3,936) and Hispanics (N = 1,838); United Kingdom Biobank: 140,886 empirically classified white	4671	4,671 Blacks	African-ancestry	1	Yes	#MVP and UKB	Giri et al. (2019)
2019	GWAS Catalog; WoS	>18	Uganda, Africa	14126	—	14,126 African-ancestry	14126	14,126 African-ancestry	African	100	Yes	UGR	Gurdasani et al. (2019)
2017	GWAS Catalog; Scopus; WoS	> 60.9	United States	99785	40.5	Non-Hispanic whites (81%; 80,792), Latinos (8%; 8,231), East Asians (7%; 7,243), AAs (3%; 3,058), and South Asians (1%; 461)	3058	3,058 AA	AA	3	Yes	*GERA	Hoffmann et al. (2017)
2011	GWAS Catalog; PM; PMC; Scopus; WoS	>18	United States	7917	39.7	4,069 AA cases, 3,848 AA controls; CARe; ARIC (N = 3,269), CARDIA (N = 1,209), CFS (N = 704), JHS (N = 2,200), and MESA (N = 1,737)	7917	4,069 AA cases, 3,848 AA controls	AA	100	Yes	CARe (1b)	Lettre et al. (2011)

(Continued on following page)

TABLE 1 | (*Continued*) Study characteristics of the 12 eligible publications.

Publication Year	Database pull	Age group (Years)	Study setting/region	Initial size (N)	Male composition (%)	Composition	African-ancestry size (n)	African-ancestry Composition	Ethnicity group	Africans ancestry only (%)	Meta-analysis	Consortium	References
2017	GWAS Catalog; PM; PMC; Scopus; WoS	NA	United States; NR	31968	30.3	31,968 of African-ancestry from 19 AA Cohorts and 2 African cohorts	31968	31,968 African-ancestry from 19 AA Cohorts and 2 African cohorts	African ancestry	100	Yes	21 Previous GWASs	Liang et al. (2017)
2019	GWAS Catalog; PM; PMC; Scopus; WoS	10 to 11	Uganda, Africa	815	51.1	815 African-ancestry early adolescents	815	815 African-ancestry early adolescents	African	100	No	EMaBS	Lule et al. (2019)
2020	GWAS Catalog	NA	United States	1,520	39.1	281 AA or Afro-Caribbean, 598 Hispanic or Latin American, 125 East Asian, 516 Europeanls	281	281 AA	AA	18	Yes	#GENNID	Willems et al. (2020)
2019	GWAS Catalog	NA	United States	49839	—	49,839 non-European; 11,863 AA cases, 8,663 Hispanic/Latino cases, 3,102 Asian ancestry cases, 2,645 Native Hawaiian ancestry cases, 342 Native American ancestry cases, 508 cases, 5,289 AA controls, 13,273 Hispanic/Latino controls, 1,561 Asian ancestry controls, 1,060 Native Hawaiian ancestry controls, 293 Native American ancestry controls, 542 controls	17152	11,863 AA cases,5,289AA controls	AA	34	Yes	*PAGE	Wojcik et al. (2019)
2015	PM; PMC; Scopus; WoS	≥ 20	United States; Nigera	29378	—	28,190 AA, 1,188 Nigerian 3000 total used	29378	28,190 AA, 1,188 Nigerian; 3000 total used	AA; African Ancestry (Nigerian)	100	Yes	COGENT (2b)	Zhu et al. (2015)

*Paper and population Characteristics of the 12 eligible publications. *Only African ancestry reported in this review (100% African ancestry), #Not 100% African ancestry. (a) Primary study (b) Secondary study. Red = Studies excluded when reporting genetic associations. Cohorts: AADM (Africa America Diabetes Mellitus Study), ARIC (Atherosclerosis Risk in Communities), CARDIA (Coronary Artery Risk Development in Young Adults), CARe (Candidate Gene Association Resource), CFS (Cleveland Family Study), CHS (Cardiovascular Health Study), COGENT (Continental Origins and Genetic Epidemiology Network), DCS (Diabetes Case-control Study), DDS (Durban Diabetes Study), EMaBS (Entebbe Mother and Baby Study), GENNID (GENetics of Noninsulin dependent Diabetes Mellitus), GERA (Genetic Epidemiology Research on Adult Health and Aging), GPC (General Population Cohort), HRC (Haplotype Reference Consortium), HUFS (Howard University Family Study), JHS (Jackson Heart Study).*

TABLE 2 | Phenotype characteristics (BP Traits and HTN) of the 12 eligible publications.

Instrument	Reading taken	Final reading	Patient position	Guidelines	DNA extracted from	BP traits studied	HTN cases (%)	AHM (%)	SBP (Mean)	SBP (SD)	DBP (Mean)	DBP (SD)	PP (Mean)	PP (SD)	References
Oscillometric device (Omron)	3X (every 10 min)	Average of 2nd and 3rd	Seated position	JNC7	NA	HTN SBP; DBP	NA	NA	131.3	16.5	81.35	10.95	—	—	Adeyemo et al. (2009)
Random-zero, Mercury and Autimated oscillometric Sphygmomanometer	ARIC and JHS = 3X; CARDIA = After 5 min rest period, 3X (every 1 min); CFS = 9X (3 every 18 h); MESA = 3X (every 1 min)	Average of 2nd and 3rd	Seated position	AHM adjusted = 10 and 5 mmHg added for SBP and DBP	NA	SBP; DBP	—	42.2	125.9	18.6	77.5	11.1	—	—	Fox et al. (2011)
NA	NA	NA	NA	AHM adjusted = 10 and 5 mmHg added for SBP and DBP; Outliers excluded = >4 SDs from mean	NA	HTN; SBP; DBP	26.2	44.9	130.6	20.5	78.1	11.8	—	—	Franceschini et al. (2013)
NA	used earliest recorded measured SBP and DBP	Median SBP and DBP	NA	AHM adjusted = 15 and 10 mmHg added for SBP and DBP; as per Newton-Cheh C et al.2009 and Taylor et al. 2016	Blood	SBP; DBP; PP	—	0.3	136.4	16.0	82.6	11.6	54.0	12.4	Giri et al. (2019)
Automated Omron M6-I	3X (every 3–5 min)	Average of 2nd and 3rd	Seated position	As described previously (Asiki et al., 2013); Fasting 30 min prior to the measurement	Blood	SBP; DBP	—	—	122.8	17	74	10.5	—	—	Gurdasani et al. (2019)
NA	NA	Average of repeated BP measures over a 5-years timespan	NA	AHM adjusted = 15 and 10 mmHg added for SBP and DBP; As per Tobin et al. 2005	NA	SBP; DBP; PP	—	—	127.8	14.8	76.2	9.6	52.1	11.7	Hoffmann et al. (2017)
NA	NA	NA	NA	NA	NA	HTN	31.9	—	—	—	—	—	—	—	Lettre et al. (2011)
NA	NA	NA	NA	AHM adjusted = 15 and 10 mmHg added for SBP and DBP; JNC	NA	HTN; SBP; DBP; BP (multi-trait analysis)	55.5	39.5	133.1	22.5	80.2	13.1	53.0	15.5	Liang et al. (2017)
Automated Omron (M6, HEM-700) machines	After 5 min rest period, BP measured 3X (every 5 min)	Average of 2nd and 3rd	Seated position	As decribed by Lule et al., 2019; "Pre-hypertension" was defined as systolic or diastolic BP ≥ 90th but <95th percentile	Red cell pellets	SBP; DBP	—	—	106	—	65.3	—	—	—	Lule et al. (2019)

(Continued on following page)

TABLE 2 | (*Continued*) Phenotype characteristics (BP Traits and HTN) of the 12 eligible publications.

Instrument	Reading taken	Final reading	Patient position	Guidelines	DNA extracted from	BP traits studied	HTN cases (%)	AHM (%)	SBP (Mean)	SBP (SD)	DBP (Mean)	DBP (SD)	PP (Mean)	PP (SD)	References
NA	NA	NA	NA	NA	Blood	SBP; DBP	–	–	–	–	–	–	–	–	Willems et al. (2020)
NA	NA	NA	NA	NA	NA	HTN; SBP; DBP	69.16	–	139.66	21.96	85.95	13.43	–	–	Wojcik et al. (2019)
NA	NA	NA	NA	NA	NA	HTN; SBP; DBP; BP (multi-trait analysis)	–	–	–	–	–	–	–	–	Zhu et al. (2015)

Eight studies were based on only participants of African-ancestry (**Table 1**). Of the 4 studies with other ethnicities (**Table 1**), Wojcik et al. (2019) and Hoffmann et al. (2017) reported ethnicity based GWASs and therefore only the African-ancestry data was reported for this systematic review (34 and 14% African-ancestry, respectively). Giri et al. (2019) and Willems et al. (2020) reported low percentages of African-ancestry within the conducted GWASs i.e., 14% and 18% respectively and the GWAS was based on multiple ethnicities, therefore, these studies were excluded when reporting genetic associations. There were two secondary studies. The Lettre et al. (2011) publication was the secondary study to that reported in Fox et al. (2011), based on the Candidate Gene Association Resource (CARe) consortium. The Zhu et al. (2015) publication was the secondary study of the Franceschini et al. (2013) study and was based on the Continental Origins and Genetic Epidemiology Network (COGENT) consortium.

Phenotype Measurement

From the studies that reported phenotype characteristics (**Table 2**), three used the automated Omron device and one used a sphygmomanometer to measure BP. From the reported readings, five studies used the average of the second and third recorded readings as the final reading, with the measurements taken with the patient in a seated position. All BP traits were derived from standard calculation, as reported by respective studies. For five of the studies, if a patient was taking AHM, then SBP and DBP was adjusted by adding 15 andand 10 mmHg (three studies) or adding 10 andand 5 mmHg (two studies), respectively. All publications examined SBP and DBP, with the exception of Lettre et al. (2011) who only looked at HTN as a trait (**Supplementary Figure S1**). SBP and DBP average measures were reported by most studies (**Table 2**).

Genotyping and Genome-Wide Association Study Parameters

DNA was extracted from blood samples (Giri et al., 2019; Gurdasani et al., 2019), red cell pellets (Lule et al., 2019) and Epstein-Barr virus (EBV) transformed blood lymphocyte cell lines (Willems et al., 2020) (**Table 2**).

Genotyping were performed using Illumina arrays (four studies), and Affymetrix arrays (eight studies) (**Table 3**). While the three initial studies were based on 600 K SNP arrays, studies that are more recent used the 1 M (Affymetrix array studies) and 2.5 M (Illumina studies) arrays. All the studies used standard GWAS quality control (QC) parameters. SNP imputation was done for all studies, with the exception of Adeyemo et al. (2009). A range of reference panels was used, with the most common being the 1,000 Genomes Project reference panel (six studies). Merged reference panels were used for six studies. For example, the African Genome Variation Project (AGVP) reference panel was used by both Ugandan studies (Gurdasani et al., 2019; Lule et al., 2019). The final number of SNPs ranged between 808,465 and 30,072,738 SNPs. Population structure was adjusted for in most studies.

TABLE 3 | Genotype and analyses characteristics of the 12 eligible publications.

Array type	Array name	QC	Population stucture/PCA	Imputed	References Panel	Final SNPs number	Association	Covariates/ Confounding variables adjusted for	Suggestive sig.	Replication	References
GW genotyping array	Affymetrix ® GW Human SNP Array 6.0	Yes	Structure = Non parametric clustering of genotypes using the AWClust algorithm; PC = using eigenstrat method	No	NA	8,08,465	LoRM under an additive model for HTN (PLINK V1.04); LRM for SBP and DBP (PLINK V1.04)	Age, Sex, BMI, First 2 PCs of the genotypes	p < 1E-4	Yes	Adeyemo et al. (2009)
GW genotyping array	Affymetrix GW Human SNP Array 6.0 array	Yes	Structure = MDS; PCs were calculated for all unrelated individuals and predicted forrelated individuals	Yes	Merged: YRI and CEU HapMap-phased haplotypes	2,500,000	LRM under an additive model	Age, Age2, Sex, BMI, First 10 PCs	p < 1E–6	Yes	Fox et al. (2011)
GW genotyping array	Affymetrix GeneChip SNP Array 6.0	Yes	PCA	Yes	Merged: HapMap Phase III and the Human Genome Diversity Project	~2,420,000	LRM/LMM for family data under additive model	Age2, Sex, BMI, First 10 PCs	p < 1E–5	Yes	Franceschini et al. (2013)
GW genotyping array	Affymetrix Axiom Biobank array designed specifically for the MVP	Yes	PCA using FlashPCA; Clusted by race using K-means approach K1-4	Yes	1000G phase 3, V5	NR	LRM with additive models (untransformed), using SNPTEST-v2.5.4-beta	Age, Age2, Sex, BMI, First 10 PCs	p < 1E–6	Yes	Giri et al. (2019)
GW genotyping array, GW sequencing	Illumina HumanOmni 2.5M BeadChip array(UGR); Illumina HumanOmni Multi-Ethnic GWAS/Exome Array (MEGA pre-commercial v1) using the Infinium Assay a.k.a MEGA Array (DDS and DCC and some of AADM); Affymetrix Axiom GW PanAFR Array Set (AADM)	Yes	PCA; fineSTRUCTURE analysis (Lawson et al., 2012)	Yes	Merged: Uganda Genome Resource, AGVP, 1000G phase 3	>24,423,923	LMM with two random effects	Age, Age2, Sex	None[a]	No	Gurdasani et al. (2019)
GW genotyping array	Affymetrix Axiom array	Yes	PCA; LMM using estimated kinship matrices with leave one chromosome out (LOCO) to account for population substructure and cryptic relatedness with Bolt-LMM	Yes	1000G phase I	>2,696,785	LMM using estimated kinship matrices in Bolt-LMM	Age, Age2, Sex, BMI, First 2 ancestry PCs	p ≤ 1E-2	Yes	Hoffmann et al. (2017)
GW genotyping array	Affymetrix GW Human SNP Array 6.0 platform	Yes	Family structure modeled using LME models; PCA using EIGENSTRAT	Yes	Merged: HapMap phase 2 CEU and YRI data	~2,740,000	LoRM under an additive genetic model In PLINK	First 10 PCs	p < 1E–6	Yes	Lettre et al. (2011)
GW genotyping array	Either Affymetrix or Illumina	Yes	PCA	Yes	1000G Phase 1	>30,072,738	LRM for unrelated data or by the generalized LMM for family data, under an additive genetic model	Age, Age2, Sex, BMI, up to first 10 PCs	p < 1E–6	Yes	Liang et al. (2017)
GW genotyping array	Illumina HumanOmni2.5M-8 ('octo') Beadchip arrays, version 1.1 (Illumina Inc., San Diego, United States)	Yes	Structure = LMM (GCTA V1.22)	Yes	Merged: 1000G, AGVP and UG2G	20,074,711	LMM regression (accounting for population substructure) (GCTA V1.22)	Age, BMI	p < 1E–6	Yes	Lule et al. (2019)

(Continued on following page)

TABLE 3 | (*Continued*) Genotype and analyses characteristics of the 12 eligible publications.

Array type	Array name	QC	Population stucture/PCA	Imputed	References Panel	Final SNPs number	Association	Covariates/ Confounding variables adjusted for	Suggestive sig.	Replication	References
GW genotyping array	Infinium Multi-Ethnic Global BeadChip (v1.0)	Yes	NA; TransMeta, follows a hierarchical framework that incorporates a kernel matrix K into the covariance structure of the effect size estimates	Yes	EA: HRC; AA, MA & JA: 1000G Phase 3	3,153,931	LMM using GCTA (Log & rank-based inverse normal transfomations)	Age, Sex & self-reported diabetes status (FEs)	p < 1E–6	No	Willems et al. (2020)
GW genotyping array	MEGA array	Yes	PCA, (SNPRelete in R)	Yes	1000G phase 3	28,263,875	LoRm and LMM, in both SUGEN and GENESIS program	AHM adjustment, (for BP traits only, not HTN), First 10 PCs	None[b]	Yes	Wojcik et al. (2019)
GW genotyping array	Either with Affymetrix or the Illumina whole-genome SNP genotyping arrays	Yes	NA	Yes	NA	~2,420,000	LRM, additive model; Plink	None	$p < 5 \times 10^{-7}$	No	Zhu et al. (2015)

GWAS sig. = $p < 5 \times 10^{-8}$ for all studies, except for [a]GWAS sig. = < 5E–9) [b]GWAS sig. = < 3E–9 for MAF <5%.

For genome-wide association the following tests were performed; 1) Linear regression models (LRMs)/linear mixed models (LMMs) regression, under an additive model, to measure continuous BP traits (i.e., SBP, DBP and PP) and 2) logistic regression models (LoRMs), under an additive model, to measure binary traits (i.e., HTN). All studies adjusted for covariates or confounding variables, with the exception of Zhu et al. (2015). The most common covariates adjusted for in the association testing were age, age2, sex, body mass index (BMI), and first 10 principal components (PCs) of the genotypes. The distributions of association were displayed by means of Manhattan and Q–Q plots for all studies, with most studies (seven) using locus/regional plots. Almost all studies used the standard cutoff for genome-wide significance (p < 5E-8), with exception to Gurdasani et al. (2019) who used p < 5E-9 and Wojcik et al. (2019) who used MAF (minor allele frequency) specific *p*-value thresholds (p < 5E-8 for MAF >5%; p < 3E-9 for MAF <5%). No suggestive significance was reported in these two studies. Replication of the genome-wide associations was conducted for most studies, with three exceptions (Zhu et al., 2015; Gurdasani et al., 2019; Willems et al., 2020).

Summary of Genetic Associations Detected

The analysis of genetic associations was based on the 10 studies with only African-ancestry participants (**Table 4**). Liang et al. (2017), based on 31968 African-ancestry individuals (largest), reported the most associations across all BP-related traits (SBP, DBP, HTN, PP and BP multi-trait analysis). The distribution of genome-wide signals across the 22 chromosomes, per trait, is shown in **Figure 2**. The 26 genome-wide significant SNPs found across BP-related traits, are summarized in **Table 5** (suggestive associations are shown in **Supplementary Table S6**). Six SNPs were associated with more than one BP phenotype from the same study (**Figure 3**), but none of associations were found in more than one study.

1) Genetic associations detected in each study

Genome-wide associations were reported for six studies, briefly summarised below.

Adeyemo et al. (2009) reported genome-wide significant associations with SBP for one intergenic SNP (rs16877320 in *AL365265.23*), and four intronic SNPs (rs5743185 in *PMS1*, rs17365948 in *YWHAZ*, rs12279202 in *IP O 7* and rs111600591 in *SLC24A4*). Replication was found in each gene for at least one SNP with low *p*-values for the Diabetes Genetics Initiative study (Saxena et al., 2007). A non-synonymous-coding SNP in *CACNA1H* was also discussed in study, however this SNP just missed the significance threshold (p = 6.71E-08) (**Supplementary Table S6**). Two of these genes (*CACNA1H* and *SLC24A4*), have annotations that suggest a role in BP regulation (Adeyemo et al., 2009).

Fox et al. (2011) identified two novel loci i.e., rs2258119 (genotyped SNP in *C21orf91*) and rs10474346 (imputed in *GPR98/ARRDC3*) for SBP and DBP respectively. These SNPs did not replicate in the AA population. The SNP rs10474346 is within the region of arrestin C, which is a peroxisome

proliferator-activated receptor gamma that may play a role in HTN through modification of inflammation and the innate immunity system in vascular cells. A genotyped SNP (rs1990151) in *IP O 7*, which showed suggestive significance with p = 7.39E-07 (Table S 6), is within the same importin beta protein gene (*IP O 7*) and may be a potential link between early-onset glucocorticoid exposure and HTN, through changes in gene expression and kidney function (Fox et al., 2011).

Franceschini et al. (2013) found only one genome-wide associated SNP for SBP i.e., rs11041530 which is 10 kb downstream of the cytochrome b5 reductase 2 gene (*CYB5R2 [MIM608342]*). In addition, an intergenic SNP rs17428471 (in *HOXA3*), which showed suggestive significance (p = 4E-07) (**Supplementary Table S6**), replicated for SBP in additional AA samples (another associated SNP, rs11564022, is also intergenic in *HOXA3*), and was found to be a *trans*-ethic SNP since it was found significant for both SBP and DBP in *trans*-ethnic meta analyses (Franceschini et al., 2013).

Lettre et al. (2011) found one novel genome-wide associated SNP for HTN (rs7801190, in *SLC12A9*; p = 2.5E-08). *SLC12A9* is a potassium/chloride transporter gene. However, this SNP was identified using imputed genotypes (imputation quality r2_hat = 0.70) and replication (also obtained by imputation) was not significant (p = 0.29), suggesting that this association may be due to chance (Lettre et al., 2011).

Liang et al. (2017) is to date the largest African-ancestry GWAS conducted (n = 31968). This study reported 14 independent genome-wide associated SNPs (**Table 5**). These were spread in 11 independent SNPs at 11 loci, including three novel loci (*TARID/TCF21, FRMD3,* and *LLPH/TMBIM4*). This study also pointed out that the African-specific SNPs were not well tagged by HAPMAP2 data (https://medicagohapmap2.org/) which could have limited the association detection in their previous study i.e., Franceschini et al. (2013). The SNPs identified in this study were found in immune, kidney, heart, and vascular system pathways and suggests that several associated genes may be involved in the kidneys renin-angiotensin pathways kidney during HTN (Liang et al., 2017).

Zhu et al. (2015) detected five genome-wide significant loci for the BP multi-trait analyses i.e., SHom detected the *HOXA-EVX1* locus, whereas SHet detected the loci *CHIC2, HOXA-EVX1, IGFBP1/IGFBP3*, and *CDH17* (**Table 5**). Three of the SHet associated loci (*CHIC2, HOXA-EVX1, and IGFBP1/IGFBP3*) were confirmed to be associated with HTN-related traits. It was suggested that SHet was more powerful compared to combined p-values when accounting for heterogeneity (Zhu et al., 2015).

2) Comparison of genetic associations detected

Six SNPs were pleiotropic (associated with more than one trait) (**Figure 3**). Liang et al. (2017) identified significant associations with five SNPs of the same polymorphisms for different traits, with most occurring with at least one BP Multi-trait analysis method. Zhu et al. (2015) reported that rs11564022 (HOXA-EVX1) was significant for both BP multi-trait analysis methods (SHet and SHom). This SNP was also found to be suggestively significant in Franceschini et al. (2013) for SBP (p = 1.83E-06), DBP (p = 7.66E-08) and HTN (p = 6.78E-08) (**Supplementary Table S6**).

The lookup of the 26 genome-wide associated lead SNPs detected in African populations (**Table 5**), in summary statistics of the largest GWAS to date for BP (Evangelou et al., 2018), found only 14 of 26 SNPs to be present in the study (for SBP, DBP and PP). Of these, 4 SNPs show at least a suggestive p-value (rs7651190b, rs11725861, rs11563582a,b, rs11977526b) in this European ancestry based GWAS (**Supplementarty Table S7**) and the other ten had p-values ranging from 1.5E-04 to 9.74E-01.

Furthermore, the look up of these 26 SNPs in other BP related GWASs (based on the GWAS Catalog, with a cut-off value p < 5E-6), detected three SNPs (also found in Evangelou et al. (2018), with p < 5E-6) that were associated with BP in four non-African-ancestry based GWASs. All three SNPs were associated in the African-ancestry population with multiple BP traits using the SHet method: rs7651190b (Liang et al., 2017), was found to be associated with PP in Japanese ancestry based GWAS (Takeuchi et al., 2018). rs11725861 (Zhu et al., 2015), was found to be associated with PP in European ancestry based GWAS (Hoffmann et al., 2017). rs11977526 (Zhu et al., 2015; Liang et al., 2017), was found to be associated with SBP in European ancestry based GWAS (Surendran et al., 2016), DBP in another European ancestry based GWAS (Hoffmann et al., 2017) and PP in both European (Hoffmann et al., 2017) and Japanese (Takeuchi et al., 2018) ancestry based GWAS.

After comparing the 26 genome-wide associated SNPs (**Table 5**) to the GWAS catalog (with a cut-off value p < 5E-6) and Evangelou et al. (2018) summary statistics, only 12 of the SNPs were observed in studies with African-ancestry participants (**Table 6**).

DISCUSSION

This systematic review of genetic association studies for BP, HTN and related phenotypes in African-ancestry populations, has highlighted the need for further studies from Africa, where HTN is most prevalent.

Despite the first GWAS with BP and HTN being conducted in 2007 and there being over 160 published studies, there are just a handful of studies on African-ancestry participants and even fewer on continental African populations. It is however encouraging that the number of studies is increasing and that sample sizes are also becoming larger in individual studies, although many remain small (e.g., Willems et al. (2020) had only 281 African-ancestry participants). Large sample sizes, especially in continental African populations are necessary to enhance our understanding of the genetic architecture of BP traits.

Meta-analysis of studies from diverse populations can lead to higher power for detecting associations across multiple populations and can begin to address the transferability of associations across populations (Willems et al., 2020).

Summary of Studies and Associations

The variability between associations found across studies may be partly attributable to differences in study protocols between

TABLE 4 | GWAS associated SNPs detected for BP-related traits in the 10 studies, with 100% African-ancestry participants.

BP traits	BP trait type	GW association number	Suggestive association number	Total associations	References
SBP	SBP	5	78	83	Adeyemo et al. (2009)
SBP	SBP	1	9	10	Fox et al. (2011)
SBP	SBP	1	18	19	Franceschini et al. (2013)
SBP	SBP (GERA AA)	0	10	10	Hoffmann et al. (2017)
SBP	SBP	4	6	10	Liang et al. (2017)
SBP	SBP	0	4	4	Lule et al. (2019)
SBP	SBP	1	0	1	Zhu et al. (2015)
DBP	DBP	0	82	82	Adeyemo et al. (2009)
DBP	DBP	1	3	4	Fox et al. (2011)
DBP	DBP	0	21	21	Franceschini et al. (2013)
DBP	DBP (GERA AA)	0	6	6	Hoffmann et al. (2017)
DBP	DBP	3	14	17	Liang et al. (2017)
DBP	DBP	0	4	4	Lule et al. (2019)
PP	PP (GERA AA)	0	10	10	Hoffmann et al. (2017)
PP	PP	3	16	19	Liang et al. (2017)
HTN	HTN	0	77	77	Adeyemo et al. (2009)
HTN	HTN	0	4	4	Franceschini et al. (2013)
HTN	HTN	1	0	1	Lettre et al. (2011)
HTN	HTN	1	6	7	Liang et al. (2017)
HTN	HTN	0	1	1	Zhu et al. (2015)
BP multi-trait	BP traits (SHet)	9	17	26	Liang et al. (2017)
BP multi-trait	BP traits (SHom)	2	15	17	Liang et al. (2017)
BP multi-trait	BP traits (SHet)	4	6	10	Zhu et al. (2015)
BP multi-trait	BP traits (SHom)	1	2	3	Zhu et al. (2015)

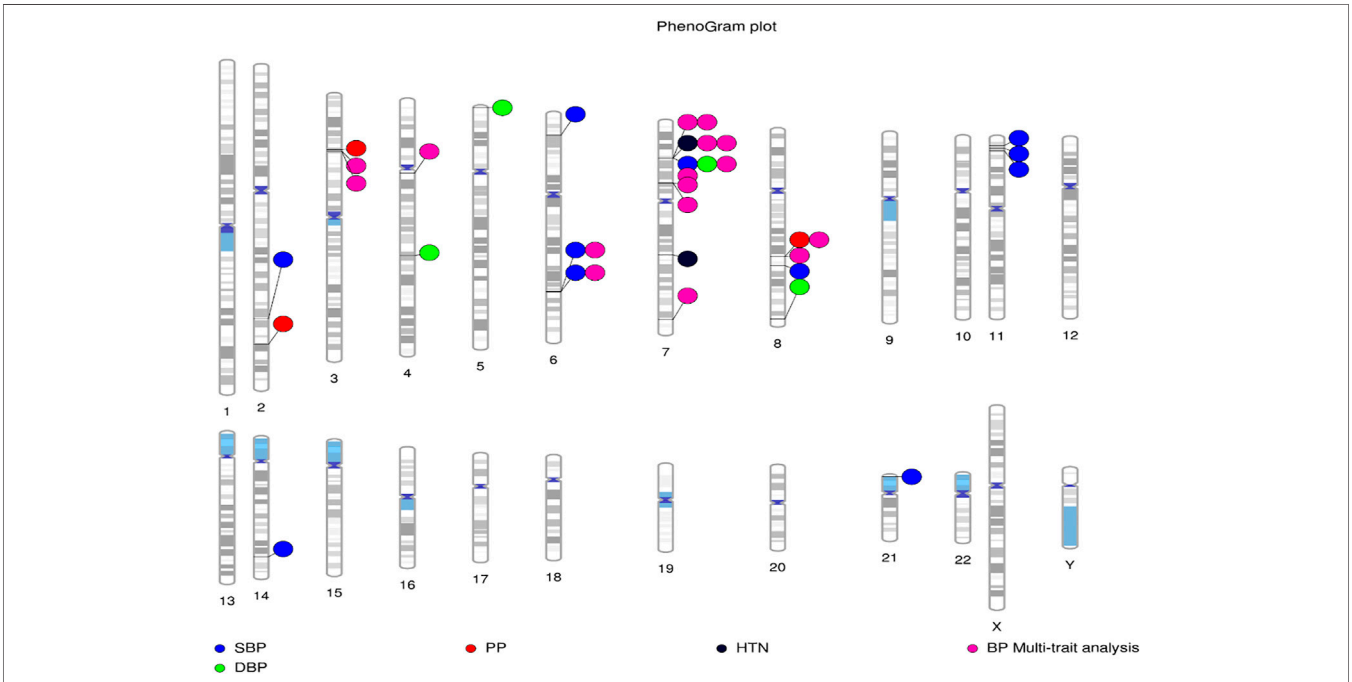

FIGURE 2 | Distribution of loci identified by GWAS for BP-associated traits in African-ancestry populations. Lines are plotted on the chromosomes corresponding to the base-pair location of each SNP, and the coloured circles represent the associated phenotype(s). Constructed using PhenoGram (Wolfe et al., 2013).

studies, but could also be due to major differences in sample size and sampling geography. Regional/geographical variability may also play a major role in the differences observed in genetic architecture. Most studies reported mainly on AA participants (**Table 1**), with African studies limited to Nigeria (Franceschini et al., 2013; Zhu et al., 2015) and Uganda (Gurdasani et al., 2019; Lule et al., 2019), and the origin of eight specified "African-ancestry" studies were not mentioned. African American populations are admixed and according to several studies and the recorded history of the

TABLE 5 | SNPs associations in the studies with only African-ancestry participants.

Trait	Chr	SNP	Position	Reported Gene	Distance to gene (kb)	Type	EA	RA	EAF	OR or BETA	SE or 95% CI	Heterogeneity	*p*-value	Top SNP rank	Analysis type	Reported Build	Ref
SBP	2	rs5743185	190446083	*PMS1*	0	Intronic	T	NA	0.14	-	-	—	2.09E-11	1	—	GRCh36	Adeyemo et al. (2009)
SBP	6	rs16877320	16031005	*AL365265.23*	-11987	Intergenic	G	T	0.13	-	-	—	3.42E-09	2	—	GRCh36	Adeyemo et al. (2009)
SBP	6	**rs76987554a,b**	134080855	—	—	—	t	c	0.1	−1.85	0.31	—	2.20E-09	1	—	GRCh37	Liang et al. (2017)
SBP	6	**rs79030490a,b**	134087689	—	—	—	a	c	0.1	−1.83	0.31	—	3.00E-09	2	—	GRCh37	Liang et al. (2017)
SBP	7	**rs11563582a,b**	27351650	—	—	—	a	g	0.13	1.61	0.28	—	7.10E-09	3	—	GRCh37	Liang et al. (2017)
SBP	8	rs17365948	102026053	*YWHAZ*	0	Intronic	A	C	0.11	—	—	—	1.59E-08	4	—	GRCh36	Adeyemo et al. (2009)
SBP	11	rs11041530	7658079	NA	—	Intergenic	c	g	0.11	−1.35	0.25	3.45E-01	4.04E-08	1	—	GRCh36	Franceschini et al. (2013)
SBP	11	rs12279202	9388666	*IPO7*	0	Intronic	A	C	0.12	—	—	—	4.80E-08	5	—	GRCh36	Adeyemo et al. (2009)
SBP	11	rs7941684a	5532222	—	—	—	t	g	1.0	−1.23	0.22	—	2.40E-08	4	—	GRCh37	Liang et al. (2017)
SBP	14	rs11160059	91877083	*SLC24A4*	0	Intronic	A	NA	0.28	—	—	—	1.54E-08	3	—	GRCh36	Adeyemo et al. (2009)
SBP	21	rs2258119	18 89 350	*C21orf91*	—	Genotyped	C	T	0.32	1.84	0.34	0.7	4.69E-08	1	—	GRCh36	Fox et al. (2011)
DBP	4	rs62312401a	116987529	—	—	—	a	g	0.94	1.31	0.24	-	3.50E-08	3	—	GRCh37	Liang et al. (2017)
DBP	5	rs10474346	9,05,99,895	*GPR98/ ARRDC3*		Imputed	C	T	0.33	1.1	0.2	0.11	3.56E-08	1	—	GRCh36	Fox et al. (2011)
DBP	7	**rs11563582a,b**	27351650	—	—	—	a	g	0.13	1.02	0.17	-	8.40E-10	1	—	GRCh37	Liang et al. (2017)
DBP	8	rs78192203a	142375073	—	—	—	a	t	0.2	−0.77	0.14	-	1.30E-08	2	—	GRCh37	Liang et al. (2017)
PP	2	rs556271823	210033781	—	—	—	a	at	0.1	−1.62	0.29	-	3.10E-08	3	—	GRCh37	Liang et al. (2017)
PP	3	rs114821199a	40965875	—	—	—	a	g	0.01	4.21	0.73	-	1.00E-08	2	—	GRCh37	Liang et al. (2017)
PP	8	**rs7006531a,b**	95110744	—	—	—	a	g	1.0	−1.16	0.17	-	5.00E-12	1	—	GRCh37	Liang et al. (2017)
HTN	7	**rs10279895a,b**	27328210	—	—	—	a	g	1.0	0.19	0.03	-	1.80E-08	1	—	GRCh37	Liang et al. (2017)
HTN	7	rs7801190	100296029	*SLC12A9*	—	—	-	C	0.73	1.35	1.22–1.50	-	2.50E-08	1	—	GRCh36	Lettre et al. (2011)
BP Multi-trait analysis	3	rs147428270b	41868721	—	—	—	t	c	0.63	—	—	—	2.49E-08	6	SHet	GRCh37	Liang et al. (2017)
BP Multi-trait analysis	3	rs7651190b	41765955	—	—	—	a	g	0.4	—	—	—	6.87E-09	4	SHet	GRCh37	Liang et al. (2017)
BP Multi-trait analysis	4	rs11725861	54497062	*CHIC2*	—	—	A	-	0.84	—	—	—	8.45E-09	2	SHet	GRCh36	Zhu et al. (2015)

(Continued on following page)

TABLE 5 | (*Continued*) SNPs associations in the studies with only African-ancestry participants.

Trait	Chr	SNP	Position	Reported Gene	Distance to gene (kb)	Type	EA	RA	EAF	OR or BETA	SE or 95% CI	Heterogeneity	*p*-value	Top SNP rank	Analysis type	Reported Build	Ref
BP Multi-trait analysis	6	**rs76987554a,b**	134080855	—	—	—	t	c	0.1	—	—	—	1.84E-08	5	SHet	GRCh37	Liang et al. (2017)
BP Multi-trait analysis	6	**rs79030490a,b**	134087689	—	—	—	a	c	0.1	—	—	—	2.50E-08	7	SHet	GRCh37	Liang et al. (2017)
BP Multi-trait analysis	7	**rs10279895a,b**	27328210	—	—	—	a	g	1.0	—	—	—	3.24E-08	8	SHet	GRCh37	Liang et al. (2017)
BP Multi-trait analysis	7	**rs10279895a,b**	27328210	—	—	—	a	g	1.0	—	—	—	2.16E-08	2	SHom	GRCh37	Liang et al. (2017)
BP Multi-trait analysis	7	rs115476423b	149199964	—	—	—	c	g	0.94	—	—	—	4.99E-08	9	SHet	GRCh37	Liang et al. (2017)
BP Multi-trait analysis	7	**rs11563582a,b**	27351650	—	—	—	a	g	0.13	—	—	—	1.08E-09	2	SHet	GRCh37	Liang et al. (2017)
BP Multi-trait analysis	7	**rs11563582a,b**	27351650	—	—	—	a	g	0.13	—	—	—	1.51E-10	1	SHom	GRCh37	Liang et al. (2017)
BP Multi-trait analysis	7	**rs11564022**	27303571	*HOXA-EVX1*	—	—	T	-	0.23	—	—	—	1.34E-08	3	SHet	GRCh36	Zhu et al. (2015)
BP Multi-trait analysis	7	**rs11564022**	27303571	*HOXA-EVX1*	—	—	T	-	0.23	—	—	—	2.35E-09	1	SHom	GRCh36	Zhu et al. (2015)
BP Multi-trait analysis	7	**rs11977526**	45974635	*IGFBP1,IGFBP3*	—	—	A	-	0.32	—	—	—	1.87E-08	4	SHet	GRCh36	Zhu et al. (2015)
BP Multi-trait analysis	7	**rs11977526b**	46008110	-	—	—	a	g	0.34	—	—	—	4.53E-09	3	SHet	GRCh37	Liang et al. (2017)
BP Multi-trait analysis	8	rs2446849	95172673	*CDH17*	—	—	T	-	1.0	—	—	—	7.01E-09	1	SHet	GRCh36	Zhu et al. (2015)
BP Multi-trait analysis	8	**rs7006531a,b**	95110744	-	—	—	a	g	1.0	—	—	—	7.56E-14	1	SHet	GRCh37	Liang et al. (2017)

Ranked by chromosome number for: (a) SBP. (b) DBP. (c) PP. (d) HTN. (e) BP Multi-trait analyses. Consisting of only GW associations (p < 5E-8). Duplicate SNPs in bold. Chr: Chromosome, EA: Effect allele, EAF: Effect allele frequency.

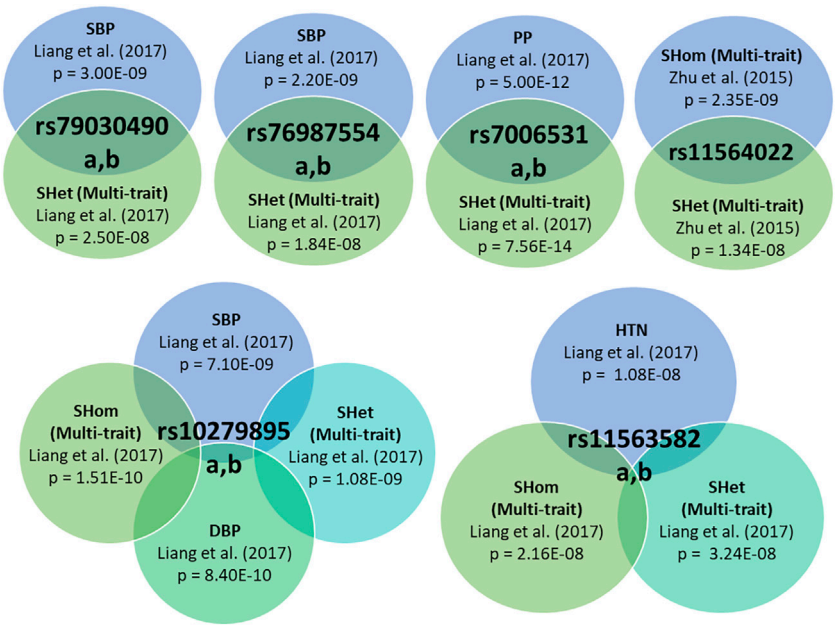

FIGURE 3 | Individual variants with associations to multiple BP-related traits, exhibiting pleiotropic effects.

trans-Atlantic slave trade, predominately represent a West African-ancestry, and is limited to specific countries or regions (Tishkoff et al., 2009; Patin et al., 2017). Therefore, genetic associations may differ within groups of AAs due to regional variability in the origins of their African ancestors, with a potential genetic stratification bias which would also be affected by the proportion of European ancestry in different participants and groups (Zakharia et al., 2009). Studies focusing on a selection of other countries within Africa are necessary to capture the ethnic diversity across sub-Saharan Africa and the role of genetic variation in BP related traits (Ramsay et al., 2016; Choudhury et al., 2020).

For studies that provided details on how phenotype data was collected, there was variability in the instrument used and how readings were taken to measure SBP and DBP, and the guidelines used to determine HTN and adjustments made for AHM (**Table 2**). Variability of BP-related phenotype traits in different populations could influence genetic associations, and may be partly attributed to environmental factors and comorbidities, as well as different age ranges in the participants.

The use of AHM was reported by four studies (Fox et al., 2011; Franceschini et al., 2013; Liang et al., 2017; Giri et al., 2019), however, the methods used for adjustments were not always mentioned. This may introduce bias when making comparisons across studies, as studies that make adjustments would not be comparable to studies that did not make adjustments. Two studies conducted BP multi-trait analyses (Zhu et al., 2015; Liang et al., 2017), using the SHet and SHom approaches. It was reported that these methods gave greater power to detect associations and the SHet method resulted in more associations compared to SHom for both studies. SHet was able to capture evidence for association, even when there was noise caused by heterogeneity among traits and cohorts (Zhu et al., 2015).

Data and statistical models were different in terms of final array sizes (depending on the genotyping array, imputation strategy and reference panel used), GWAS method (depending on adjustments for population structure/PCA (principal components analysis), association model, covariates adjusted for) and thresholds used for determining suggestive significance (**Table 3**). Adjusting for age, age2, sex and the first 10 PCs, is common practice for most GWASs, however, the reason BMI is used as an adjustment variable for most studies is not mentioned in the studies. This may have to do with the strong associations found between BMI and high BP/HTN (Landi et al., 2018).

Lack of Power to Detect Small Effects

The lack of significant associations found for BP traits from several of the African-ancestry studies is largely due to the low power to detect small effects of common SNPs, and the fact that genotyping arrays are usually Eurocentric and enriched for common European SNPs. Given that sample sizes tend to be smaller in African-ancestry studies, they only have power to detect large effect associations. Therefore, many true associations with small effects will not be significant and many suggestive associations may be spurious. The use of multiple cohorts, to maximize sample size and increase power, could bias findings as some associations may be due to heterogeneity in BP measurements across the different cohorts (Fox et al., 2011).

Many genetic associations found in European studies may not replicate because of the low power of the small studies to detect small effect variants. It is well recognised that BP traits are influenced by

TABLE 6 | Potential unique African-ancestry GW associations signals for 10 studies, based on 100% African-ancestry.

No	Chr	SNP	Position	Reported Gene	Distance to gene (kb)	Type	EA	RA	EAF	Associated Traits	Reported Build	EuA	African-ancestry	Higher freq
1	2	rs556271823	210033781	—	—	—	a	at	0.06	PP	GRCh37	0.1214	0.0483	EuA
2	2	rs5743185	190446083	*PMS1*	0	Intronic	T	NA	0.1418	SBP	GRCh36	0.0696	0.0182	EuA
3	3	rs114821199a	40965875	—	—	—	a	g	0.01	PP	GRCh37	0.0002	0.0145	African-ancestry
4	3	rs147428270b	41868721	—	—	—	t	c	0.63	BP Multi-trait analysis	GRCh37	0.0759	0.6448	African-ancestry
5	6	rs76987554a,b	134080855	—	—	—	t	c	0.09	SBP; BP Multi-trait analysis	GRCh37	0.0003	0.0910	African-ancestry
6	6	rs79030490a,b	134087689	—	—	—	a	c	0.09	SBP; BP Multi-trait analysis	GRCh37	0.0003	0.0931	African-ancestry
7	7	rs10279895a,b	27328210	—	—	—	a	g	0.90	HTN; BP Multi-trait analysis	GRCh37	0.0001	0.09283	African-ancestry
8	7	rs115476423b	149199964	—	—	—	c	g	0.94	BP Multi-trait analysis	GRCh37	6.485E-05	0.0544	African-ancestry
9	8	rs2446849	95172673	*CDH17*	—	—	T	-	0.80	BP Multi-trait analysis	GRCh36	0.9581	0.8059	EuA
10	8	rs70065314,b	95110744	—	—	—	a	g	0.85	PP; BP Multi-trait analysis	GRCh37	0.0006	0.1497	African-ancestry
11	8	rs78192203a	142375073	—	—	—	a	t	0.20	DBP	GRCh37	0.0003	0.1961	African-ancestry
12	14	rs11160059	91877083	*SLC24A4*	0	Intronic	A	NA	0.1782	SBP	GRCh36	1.0000	0.9288	EuA

[b]*Ranked by chromosome number for: (a) SBP. (b) DBP. (c) PP. (d) HTN. (e) BP Multi-trait analyses. Consisting of only GW associations ($p < 5E-8$). Chr: Chromosome, EA: Effect allele, EAF: Effect allele frequency, EuA: European allele frequency (Average of Finnish and non-Finnish population).*

many genetic variants, with each conferring a small change in BP, and modulated by environmental factors, behaviour and other physiological markers (Lule et al., 2019). This leads to failure to replicate most associated variants, since many GWAS are only powered to identify common variants of moderate to large effects (Adeyemo et al., 2009; Lettre et al., 2011). Since the African-ancestry samples are from different geographical regions and data is so limited, population admixture may result in different linkage disequilibrium (LD) patterns across African populations, thus limiting replication across different African populations (Fox et al., 2011). Failure to replicate SNPs in Lule et al. (2019) could be due to differences in MAF across the different populations or differences in LD patterns, as well the lack of knowledge of causative variants or due to spurious initial findings.

The Importance of Africa-Centric Arrays, Panels and Africa-Ancestry Studies

GWAS arrays are designed to capture common variation and rely on LD with causative variants to detect association. Most SNP arrays have generally been enriched for common European variants, suggesting that there may be many common African-ancestry SNPs that have not yet been identified as being associated with specific traits (Hoffmann et al., 2017). Interestingly, 14 of the 26 genome-wide SNPs associated in African-ancestry studies were also found in Evangelou et al. (2018) and 12 were potentially unique to African-ancestry populations (**Table 6**). This could, in part, be due to differences in linkage disequilibrium, genomic coverage of SNPs and allele frequencies that could lead to differences in associations between African and non-African-ancestry groups. African populations have over half a million more SNPs sites per genome in comparison to Europeans, therefore, GWASs performed on this population require more SNPs with a wide genomic coverage to ensure that most LD blocks are covered in the analyses (Ramsay, 2012; Choudhury et al., 2020). The development of Africa-centric arrays are important to detect signals that are unique to African-ancestries. One example of such an array is the H3Africa SNP array, consisting of ~2.3 million SNPs, including a large proportion of common African variants that effectively tag haplotype blocks in relevant populations and also included disease relevant variants from databases (such as the GWAS Catalog, ClinVar, and Cosmic).

African specific imputation panels are also important to better capture African specific variants. The AGVP reference panel, which includes 320 whole-genome sequences from 320 individuals across sub-Saharan Africa, has an efficient genotype array design capturing common genetic variation in Africa (Gurdasani et al., 2015) and hence this panel was used by both studies Ugandan studies. Moreover, the availability of larger panels, such as the TopMed with substantial African-ancestry representation, is expected to enhance the quality of imputation and thereby increasing the number of variants that could be included in future association studies (Bentley et al., 2020). This might also benefit the meta-analysis with some of the existing GWAS data.

This review emphasizes the value of including African-ancestry studies, as there are BP and HTN signals unique to African populations. Therefore, further studies in this under-represented region are needed to identify more associations and to validate signals that appear to be unique to African-ancestry

populations. Since most GWAS that include African-ancestry participants are AA, the results reported here are unlikely to be widely transferable to populations in Africa and emphasize the need for future studies based in continental Africa.

Limitations

There is a limited body of literature on GWAS in populations of African-ancestry, mainly represented by AA. This is well recognized in summaries of GWAS findings and the universal call for more representation of Africans and minority populations in such studies. Variations between population characteristics, phenotypic measures and genotypic analyses could cause bias when comparing different studies. It is therefore important for all studies to report more detail in the protocols to enable better replication and minimise bias between studies. Recent studies suggest that the genetic architecture of BP is different between longitudinal trajectories and cross sectional studies and therefore cohort analyses would be important in African settings. The relatively low overlap between signals detected in African and non-African-ancestry GWASs could be due to LD differences and the same locus being represented by different lead SNPs in the two groups. Although, a regional replication was not performed in the current review, which might have enhanced the overlap, we anticipate that many of the African-specific signals could indeed be continent-specific. Due to these limitations, the results are essentially a list of sentinel findings. A narrative synthesis of the results was conducted (limited to reporting lead SNPs), as a meta-analysis was not possible and additional analyses, such as regional mining or looking at credible sets, were challenging.

Future Directions

Genetic diversity is high within African populations, because of a deep evolutionary history, population admixture and genetic drift. It is therefore vital for future studies to capture contributions from African studies and to understand the functional and biological relevance of associated SNPs. Improved methods need to be developed to understand and compare heritability across populations and studying participants from other parts of the African continent, such as South and Central Africa are essential.

For enhanced discovery in GWAS in Africans, future studies should aim for increased sample size, multiple BP-related traits (including SBP, DBP, HTN, PP, MAP, BP multi-trait analysis and longitudinal data) to enhance discoverability across phenotypes, the use of Africa-centric genotyping arrays and larger imputation reference panels, which adequately represent African-genetic diversity (Hoffmann et al., 2017).

CONCLUSIONS

There is much need for more genetic-association studies based on BP and HTN in large-scale African-ancestry research, specifically in participants from the African continent. Further studies should also investigate rare variants, adjust appropriately for the use of anti-hypertensive medication, follow standardised procedures and provide justification for the different approaches used.

DATA AVAILABILITY STATEMENT

The original contribution presented in the study is included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

Screening of the full-text articles, and selected studies in accordance to the inclusion criteria was conducted by SS and checked by MR, J-TB, and AC. All authors screened the extracted data and disagreements were resolved through a discussion. SS drafted, formatted and edited the paper following feedback and edits by MR, J-TB and AC. Critical review and domain knowledge were added by FXG-O. All authors read and approved the submitted version.

FUNDING

The authors would like to acknowledge the South African Research Chairs Initiative (SARChI), NIH Africa Wits-INDEPTH Partnership for the Genomic Research (AWI-Gen) (Grant number U54HG006938) and NRF for financial support. SS and J-TB were supported by the South African National Research Foundation (NRF) through MR's South African Research Chair in Genomics and Bioinformatics of African populations hosted by the University of the Witwatersrand, funded by the Department of Science and Technology, and administered by the NRF. AC was supported by NIH grant U54HG006938. The funders played no role in the construction of this systematic review paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.699445/full#supplementary-material

REFERENCES

Adeyemo, A., Gerry, N., Chen, G., Herbert, A., Doumatey, A., Huang, H., et al. (2009). A Genome-wide Association Study of Hypertension and Blood Pressure in African Americans. *Plos Genet.* 5 (7), e1000564. doi:10.1371/journal.pgen.1000564

Azam, A. B., and Azizan, E. A. B. (2018). Brief Overview of a Decade of Genome-wide Association Studies on Primary Hypertension. *Int. J. Endocrinol.* 2018. doi:10.1155/2018/7259704

Bentley, A. R., Callier, S. L., and Rotimi, C. N. (2020). Evaluating the Promise of Inclusion of African Ancestry Populations in Genomics. *NPJ Genom. Med.* 5 (1), 5–9. doi:10.1038/s41525-019-0111-x

Burton, P. R., Clayton, D. G., Cardon, L. R., Craddock, N. J., Deloukas, P., Duncanson, A., et al. (2007). Genome-wide Association Study of 14,000 Cases of Seven Common Diseases and 3,000 Shared Controls. *Nature* 447 (7145), 661–678. doi:10.1038/nature05911

Choudhury, A., Aron, S., Botigué, L. R., Sengupta, D., Botha, G., Bensellak, T., et al. (2020). High-depth African Genomes Inform Human Migration and Health. *Nature* 586 (7831), 741–748. doi:10.1038/s41586-020-2859-7

Evangelou, E., Warren, H. R., Mosen-Ansorena, D., Mifsud, B., Pazoki, R., Gao, H. H., et al. (2018). Genetic Analysis of over 1 Million People Identifies 535 New Loci Associated with Blood Pressure Traits. *Nat. Genet.* 50 (10), 1412–1425. doi:10.1038/s41588-018-0205-x

Forouzanfar, M. H., Afshin, A., Alexander, L. T., Anderson, H. R., Bhutta, Z. A., Biryukov, S., et al. (2016). Global, Regional, and National Comparative Risk Assessment of 79 Behavioural, Environmental and Occupational, and Metabolic Risks or Clusters of Risks, 1990–2015: a Systematic Analysis for the Global Burden of Disease Study 2015. *The Lancet* 388 (10053), 1659–1724. doi:10.1016/S0140-6736(16)31679-8

Fox, E. R., Young, J. H., Li, Y., Dreisbach, A. W., Keating, B. J., Musani, S. K., et al. (2011). Association of Genetic Variation with Systolic and Diastolic Blood Pressure Among African Americans: the Candidate Gene Association Resource Study. *Hum. Mol. Genet.* 20 (11), 2273–2284. doi:10.1093/hmg/ddr092

Franceschini, N., Fox, E., Zhang, Z., Edwards, T. L., Nalls, M. A., Sung, Y. J., et al. (2013). Genome-wide Association Analysis of Blood-Pressure Traits in African-Ancestry Individuals Reveals Common Associated Genes in African and Non-african Populations. *Am. J. Hum. Genet.* 93 (3), 545–554. doi:10.1016/j.ajhg.2013.07.010

Franceschini, N., Carty, C. L., Lu, Y., Tao, R., Sung, Y. J., Manichaikul, A., et al. (2016). Variant Discovery and fine Mapping of Genetic Loci Associated with Blood Pressure Traits in Hispanics and African Americans. *PloS one* 11 (10), e0164132. doi:10.1371/journal.pone.0164132

Giri, A., Hellwege, J. N., Keaton, J. M., Park, J., Qiu, C., Warren, H. R., et al. (2019). Trans-ethnic Association Study of Blood Pressure Determinants in over 750,000 Individuals. *Nat. Genet.* 51 (1), 51–62. doi:10.1038/s41588-018-0303-9

Gurdasani, D., Carstensen, T., Fatumo, S., Chen, G., Franklin, C. S., Prado-Martinez, J., et al. (2019). Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa. *Cell* 179 (4), 984–1002. doi:10.1016/j.cell.2019.10.004

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., et al. (2015). The African Genome Variation Project Shapes Medical Genetics in Africa. *Nature* 517 (7534), 327–332. doi:10.1038/nature13997

Hendry, L. M., Sahibdeen, V., Sahibdeen, V., Choudhury, A., Norris, S. A., Ramsay, M., et al. (2018). Insights into the Genetics of Blood Pressure in Black South African Individuals: the Birth to Twenty Cohort. *BMC Med. Genomics* 11 (1), 2. doi:10.1186/s12920-018-0321-6

Hoffmann, T. J., Ehret, G. B., Nandakumar, P., Ranatunga, D., Schaefer, C., Kwok, P.-Y., et al. (2017). Genome-wide Association Analyses Using Electronic Health Records Identify New Loci Influencing Blood Pressure Variation. *Nat. Genet.* 49 (1), 54–64. doi:10.1038/ng.3715

Juma, P. A., Mohamed, S. F., Wisdom, J., Kyobutungi, C., and Oti, S. (2016). Analysis of Non-communicable Disease Prevention Policies in Five Sub-saharan African Countries: Study Protocol. *Arch. Public Health* 74 (1), 25. doi:10.1186/s13690-016-0137-9

Landi, F., Calvani, R., Picca, A., Tosato, M., Martone, A. M., Ortolani, E., et al. (2018). Body Mass index Is Strongly Associated with Hypertension: Results from the Longevity Check-Up 7+ Study. *Nutrients* 10 (12), 1976. doi:10.3390/nu10121976

Lettre, G., Palmer, C. D., Young, T., Ejebe, K. G., Allayee, H., Benjamin, E. J., et al. (2011). Genome-wide Association Study of Coronary Heart Disease and its Risk Factors in 8,090 African Americans: the NHLBI CARe Project. *Plos Genet.* 7 (2), e1001300. doi:10.1371/journal.pgen.1001300

Levy, D., Larson, M. G., Benjamin, E. J., Newton-Cheh, C., Wang, T. J., Hwang, S. J., et al. (2007). Framingham Heart Study 100K Project: Genome-wide Associations for Blood Pressure and Arterial Stiffness. *BMC Med. Genet.* 1(1), S3.doi:10.1186/1471-2350-8-S1-S3

Liang, J., Le, T. H., Edwards, D. R. V., Tayo, B. O., Gaulton, K. J., Smith, J. A., et al. (2017). Single-trait and Multi-Trait Genome-wide Association Analyses Identify Novel Loci for Blood Pressure in African-Ancestry Populations. *Plos Genet.* 13 (5), e1006728. doi:10.1371/journal.pgen.1006728

Little, J., Higgins, J., Bray, M., Ioannidis, J., Khoury, M., Manolio, T., et al. (2006). *The HuGENet™ HuGE Review Handbook*, Ottawa, Ontario, Canada: HuGENet Canada Coordinating Centre.

Lule, S. A., Mentzer, A. J., Namara, B., Muwenzi, A. G., Nassanga, B., Kizito, D., et al. (2019). A Genome-wide Association and Replication Study of Blood Pressure in Ugandan Early Adolescents. *Mol. Genet. Genomic Med.* 7 (10), e00950. doi:10.1002/mgg3.950

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., et al. (2009). Finding the Missing Heritability of Complex Diseases. *Nature* 461 (7265), 747–753. doi:10.1038/nature08494

Martin, A. R., Teferra, S., Möller, M., Hoal, E. G., and Daly, M. J. (2018). The Critical Needs and Challenges for Genetic Architecture Studies in Africa. *Curr. Opin. Genet. Develop.* 53, 113–120. doi:10.1016/j.gde.2018.08.005

Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *J. Clin. Epidemiol.* 62, 1006–1012. doi:10.1016/j.jclinepi.2009.06.005

Patin, E., Lopez, M., Grollemund, R., Verdu, P., Harmant, C., Quach, H., et al. (2017). Dispersals and Genetic Adaptation of Bantu-speaking Populations in Africa and North America. *Science* 356 (6337), 543–546. doi:10.1126/science.aal1988

Popejoy, A. B., Ritter, D. I., Crooks, K., Currey, E., Fullerton, S. M., Hindorff, L. A., et al. (2018). The Clinical Imperative for Inclusivity: Race, Ethnicity, and Ancestry (REA) in Genomics. *Hum. Mutat.* 39 (11), 1713–1720. doi:10.1002/humu.23644

Ramsay, M. (2012). Africa: Continent of Genome Contrasts with Implications for Biomedical Research and Health. *FEBS Lett.* 586 (586), 2813–2819. doi:10.1016/j.febslet.2012.07.061

Ramsay, M., Crowther, N., Tambo, E., Agongo, G., Baloyi, V., Dikotope, S., et al. (2016). H3Africa AWI-Gen Collaborative Centre: a Resource to Study the Interplay between Genomic and Environmental Risk Factors for Cardiometabolic Diseases in Four Sub-saharan African Countries. *Glob. Health Epidemiol. Genom.* 1, e20. doi:10.1017/gheg.2016.17

Sagoo, G. S., Little, J., and Higgins, J. P. (2009). Systematic Reviews of Genetic Association Studies. Human Genome Epidemiology Network. *Plos Med.* 6 (3), e28. doi:10.1371/journal.pmed.1000028

Saxena, R., Saxena, R., Voight, B. F., Lyssenko, V., Burtt, N. P., de Bakker, P. I. W., et al. (2007). Genome-wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels. *Science* 316 (5829), 1331–1336. doi:10.1126/science.1142358

Surendran, P., Drenos, F., Young, R., Warren, H., Cook, J. P., Manning, A. K., et al. (2016). Trans-ancestry Meta-Analyses Identify Rare and Common Variants Associated with Blood Pressure and Hypertension. *Nat. Genet.* 48 (10), 1151–1161. doi:10.1038/ng.3654

Takeuchi, F., Akiyama, M., Matoba, N., Katsuya, T., Nakatochi, M., Tabara, Y., et al. (2018). Interethnic Analyses of Blood Pressure Loci in Populations of East Asian and European Descent. *Nat. Commun.* 9 (1), 5052–5116. doi:10.1038/s41467-018-07345-0

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and Limitations of Genome-wide Association Studies. *Nat. Rev. Genet.* 20, 467–484. doi:10.1038/s41576-019-0127-1

Taylor, J. Y., Schwander, K., Kardia, S. L. R., Arnett, D., Liang, J., Hunt, S. C., et al. (2016). A Genome-wide Study of Blood Pressure in African Americans Accounting for Gene-Smoking Interaction. *Sci. Rep.* 6, 18812. doi:10.1038/srep18812

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., et al. (2009). The Genetic Structure and History of Africans and African Americans. *Science* 324 (5930), 1035–1044. doi:10.1126/science.1172257

Who (2016). *Global NCD Target: Reduce High Blood pressureNCD Management-Screening, Diagnosis and Treatment*. Geneva, World Health Organization.

Who (2018). *Noncommunicable Diseases Country Profiles 2018*. Geneva: World Health Organization.

Willems, E. L., Wan, J. Y., Norden-Krichmar, T. M., Edwards, K. L., and Santorico, S. A. (2020). Transethnic Meta-analysis of Metabolic Syndrome in a Multiethnic Study. *Genet. Epidemiol.* 44 (1), 16–25. doi:10.1002/gepi.22267

Wojcik, G. L., Graff, M., Nishimura, K. K., Tao, R., Haessler, J., Gignoux, C. R., et al. (2019). Genetic Analyses of Diverse Populations Improves Discovery for Complex Traits. *Nature* 570 (7762), 514–518. doi:10.1038/s41586-019-1310-4

Wolfe, D., Dudek, S., Ritchie, M. D., and Pendergrass, S. A. (2013). Visualizing Genomic Information across Chromosomes with PhenoGram. *BioData Min.* 6 (1), 18–12. doi:10.1186/1756-0381-6-18

Zakharia, F., Basu, A., Absher, D., Assimes, T. L., Go, A. S., Hlatky, M. A., et al. (2009). Characterizing the Admixed African Ancestry of African Americans. *Genome Biol.* 10 (12), R141. doi:10.1186/gb-2009-10-12-r141

Zhu, X., Feng, T., Tayo, B. O., Liang, J., Young, J. H., Franceschini, N., et al. (2015). Meta-analysis of Correlated Traits via Summary Statistics from GWASs with an Application in Hypertension. *Am. J. Hum. Genet.* 96 (1), 21–36. doi:10.1016/j.ajhg.2014.11.011

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Pediatrics

ORIGINAL RESEARCH
published: 26 October 2021
doi: 10.3389/fped.2021.666426

Association of Stress, Glucocorticoid Receptor, and FK506 Binding Protein Gene Polymorphisms With Internalizing Disorders Among HIV-Infected Children and Adolescents From Kampala and Masaka Districts—Uganda

Tonny Jimmy Owalla [1,2], Wilber Joseph Ssebajjwe [1], Dennis Muhanguzi [3], Jacqueline Samantha Womersley [4,5], Eugene Kinyanda [1,6] and Allan Kalungi [1,7]*

[1] Mental Health Unit, Medical Research Council/Uganda Virus Research Institute and London School of Hygiene and Tropical Medicine Uganda Research Unit, Entebbe, Uganda, [2] Med Biotech Laboratories, Kampala, Uganda, [3] Department of Bio-Molecular Resources and Bio-Laboratory Sciences, College of Veterinary Medicine, Animal Resources and Biosecurity, Makerere University, Kampala, Uganda, [4] Department of Psychiatry, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, [5] South African Medical Research Council, Stellenbosch University Genomics of Brain Disorders Research Unit, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, [6] Department of Psychiatry, College of Health Sciences, Makerere University, Kampala, Uganda, [7] Department of Immunology and Microbiology, College of Health Sciences, Makerere University, Kampala, Uganda

OPEN ACCESS

Edited by:
Karoline Kuchenbaecker, University College London, United Kingdom

Reviewed by:
Anjali Bhat, University College London, United Kingdom
Prasenjit Mitra, Post Graduate Institute of Medical Education & Research (PGIMER), India

***Correspondence:**
Tonny Jimmy Owalla
owallatonny@gmail.com

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Pediatrics

Received: *10 February 2021*
Accepted: *21 September 2021*
Published: *26 October 2021*

Citation:
Owalla TJ, Ssebajjwe WJ, Muhanguzi D, Womersley JS, Kinyanda E and Kalungi A (2021) Association of Stress, Glucocorticoid Receptor, and FK506 Binding Protein Gene Polymorphisms With Internalizing Disorders Among HIV-Infected Children and Adolescents From Kampala and Masaka Districts—Uganda. Front. Pediatr. 9:666426. doi: 10.3389/fped.2021.666426

Children and adolescents living with human immunodeficiency virus (CA-HIV) suffer a considerable burden of internalizing disorders (IDs; depressive and anxiety disorders). Environmental and genetic factors have been reported to influence the vulnerability to IDs in western settings; however, their role among African populations remains inadequately explored. We investigated the individual and interactive effects of stress and single-nucleotide polymorphisms within the FK506 binding protein 5 (rs1360780) and glucocorticoid receptor (rs10482605) genes on ID status in a cohort of CA-HIV in Uganda. We genotyped rs10482605 (309 cases and 315 controls) and rs1360780 (350 cases and 335 controls) among CA-HIV with and without IDs using Kompetitive Allele-Specific PCR. Socio-demographic variables, as well as allele and genotype distributions, were compared between cases and controls using chi-square tests. Genotypes were assessed for Hardy–Weinberg equilibrium. Composite indices of recent and chronic stress classes were also generated. A hierarchical cluster analysis was used to generate cutoff points within each of the indices of recent and chronic stress. Logistic regression was used to assess the association between IDs and each of recent stress, chronic stress, and the investigated genotypes. The interaction effect of chronic/recent stress on the association between each of the polymorphisms and IDs was determined using a likelihood ratio test. We observed no significant association between IDs and rs1360780 and rs10482605 polymorphisms within the FKBP5 and glucocorticoid receptor genes, respectively ($P > 0.050$). Severe recent stress increased the vulnerability

to IDs among CA-HIV ($P = 0.001$). We did not observe any gene–environment effect on vulnerability to IDs in this population. These findings support the currently held opinion that polymorphisms at single genetic loci only contribute a very small effect to the genetic vulnerability to IDs.

Keywords: stress, NR3C1, FKBP5, internalizing disorders, children, depression, anxiety

INTRODUCTION

Poor mental health has been recognized as a neglected global health burden, particularly for children and adolescents (1). Globally, 10–20% of children suffer from life-related disabilities arising from mental health disorders, including psychiatric disorders, general psychological distress, and emotional and behavioral problems (2). However, studies on mental health among children and adolescents have lagged considerably behind that of adults, especially in the African continent (3). Children and adolescents with human immunodeficiency virus (CA-HIV) represent one of the groups with a significant burden of mental health disorders (4, 5). A study conducted in the United States reported a higher prevalence of psychiatric cases in children with HIV compared to the general pediatric population (6.17 vs. 1.7 cases per 1,000 person-years) (6). Evidence for the adverse effects of HIV infection on mental health has also been found in sub-Saharan African populations (7), with a systematic review of 14 studies indicating a 25% prevalence for any psychiatric disorder and a 30–50% prevalence of psychological distress or emotional or behavioral difficulties in adolescents with HIV (8). One of the groups of mental illness that CA-HIV suffer from is internalizing disorders (IDs; depressive and anxiety disorders) (5, 9, 10).

Internalizing disorders are most often characterized by quiet, internal distress, which may be described as "intropunitive," rather than overt, socially negative, or disruptive behavior that characterizes externalizing disorders (11). The most prevalent and commonly diagnosed IDs include depression and anxiety, with studies from both Africa and the United States reporting ranges between 4–72 and 5–82%, respectively (5, 12). In Uganda, an epidemiological study, on which this analysis is based, reported the prevalence of IDs as 11.5% and of major depressive disorder as 3.9% and any anxiety disorder as 9.0% (5).

The risk factors for IDs in CA-HIV are multifactorial and include the direct and indirect effects of HIV neurotoxicity (13). Similarly, environmental risk factors such as bereavement, extreme poverty, food insecurity, and deprivation of basic needs have also been implicated in IDs (14, 15). These environmental factors predispose to acute/recent and/or chronically stressful life conditions, which shape the onset, temporal course, and consequences of depressive and anxiety disorders (16, 17). IDs have been associated with poor school performance, impaired social relationships and reduced drug adherence, risky sexual behavior as well as impaired quality of life (14, 15).

A familial history of psychiatric illness among children and adolescents has been demonstrated to be associated with IDs, lending credence to the role of genetic factors in determining internalizing disorder risk (18). Genetic polymorphisms within the neurotransmitter and neuropeptide signaling systems—for instance, brain-derived neurotropic factor (rs6265, rs10835210, and rs11030107) (19), dopamine transporter gene (20), monoamine oxidase A (21), and serotonin transporter (22) genes, have been associated with IDs in healthy children and adolescents in the western world. Albeit that the psychiatric–genomic data in African populations is limited, a study undertaken in Uganda (23) reported a significant association between polymorphisms of the serotonin transporter gene and suicidality among adults living with HIV/AIDS. Notably, the only study to date that has investigated genetic factors associated with psychiatric disorders among CA-HIV has reported a significant interaction between recent stress and the S_A haplotype of the serotonin-transporter-gene-linked polymorphic region, which was associated with higher odds of being an ID case (24).

On the other hand, genetic studies conducted in non-HIV-infected populations have reported associations between genes within the stress response-hypothalamic-pituitary-adrenal axis (HPA) and vulnerability to depression and anxiety. In particular, vulnerability to IDs has been associated with single-nucleotide polymorphisms (SNPs) within both the FK506 binding protein 51 (*FKBP5*) gene, which encodes a glucocorticoid receptor (GR) co-chaperone that is involved in the HPA axis negative feedback (25, 26), and the nuclear receptor subfamily-3 group-C member-1 (*NR3C1*) gene, which encodes the GR protein (27, 28). SNPs such as rs3800373, rs9296158, rs1360780, rs9470080, rs4713916, rs3000377, and rs47139611 in *FKBP5* (25, 26) and rs414232, rs6195, rs6189, rs10482605, and rs6190 in *NR3C1* (27, 28) are postulated to interact with early trauma or childhood adversity to predict negative outcomes later in life, such as vulnerability to IDs (29, 30). The minor *C* allele of the rs10482605 polymorphism, located within the promoter region of *NR3C1*, is associated with a reduced expression of the GR protein and has been inversely associated with the liability to develop IDs in non-HIV-infected subjects (27, 28). Similarly, rs1360780, a SNP located in an enhancer region 488 bp away from a glucocorticoid response element in intron 2 of *FKBP5*, moderates the GR-mediated expression of FKBP5 (25, 26). Several groups reported an association between *FKBP5* variants and IDs using American, German, and Polish non-HIV cohorts (31–33). However, other studies involving Spanish, Swedish, and Italian populations failed to replicate these findings (34–36). Whether common variants in *FKBP5* are associated with vulnerability to IDs thus remains inconclusive.

Additionally, these studies have been conducted among non-HIV-infected adults, primarily in populations of European ancestry. Therefore, the prevalence of these SNPs and their association with vulnerability to IDs (depression and anxiety) in

an African population of CA-HIV is unclear. We investigated GR and FK506 binding protein genotypes and stress as predictors of IDs among CA-HIV from Kampala and Masaka districts. Finally, the role of gene-environment interactions in predisposing to IDs was also explored.

MATERIALS AND METHODS

Study Design

This was a nested case-control study utilizing archived whole blood samples collected during a Medical Research Council/Department for International Development (MRC/DFID)-funded project entitled "Mental health among children and adolescents living with HIV/AIDS in Kampala and Masaka districts of Uganda" (CHAKA study). The aim of the CHAKA study was to investigate the epidemiology of psychiatric disorders among CA-HIV in Uganda and the implications of these for service provision (5). The present study investigated the association between stress on the one hand and selected stress hormone pathway gene polymorphisms on the other with IDs among CA-HIV.

Study Participants and Site

The CHAKA parent study enrolled a total of 1,339 children and adolescents [855 children (5–11 years) and 484 adolescents (12–17 years)] from two HIV clinics in Kampala (Joint Clinic Research Centre and Nsambya Home Care) and three HIV clinics in Masaka (The Aids Support Organization-Masaka, Kitovu Mobile Clinic, and Uganda Cares). The participants were recruited and assessed for psychiatric disorders (including IDs) and socio-demographic factors at three time points (baseline and at months 6 and 12 months) by trained psychiatric nurses and psychiatric clinical officers between December 2014 and August 2016 (5). All study participants were on antiretroviral treatment (ART). The current study is comprised of a total of 736 cases and matched control participants [389 children (5–11 years) and 307 adolescents (12–17 years)] who were recruited into the CHAKA study.

Inclusion and Exclusion Criteria

The following were the inclusion and exclusion criteria for the CHAKA study:

Inclusion criteria: (i) CA-HIV (with children defined as aged between 5 and 11 years and adolescents defined as aged between 12 and 17 years), and (ii) able to speak English or Luganda (the local language spoken in the study areas) and expected to remain in the study area for the subsequent 12 months.

Exclusion criteria: (i) concurrent enrollment in another study, (ii) sick and in need of immediate medical attention, and (iii) those unable to understand the study instruments for whatever reason.

The current study included all children (aged 5–11 years) and adolescents (aged 12–17 years) with IDs and their case-matched controls (children and adolescents without any psychiatric disorder). Details about the prevalence of psychiatric disorders diagnosed in the CHAKA study, including internalizing disorders considered in the current study, have been included in **Supplementary Table 1**.

Selection of Cases and Controls

The selection of cases and controls was as described elsewhere (37). Briefly, the cases were children/adolescents who had any IDs. All cases at baseline were ascertained, and the cases were then stratified by site (one of two sites), sex, age category (one of three categories), and socio-economic status (SES; one of three SES categories). This resulted in a total of 36 strata ($2 \times 2 \times 3 \times 3$). In each stratum, the number of cases was ascertained (e.g., for males in site 1 in the youngest age category and the lowest SES group, there were nine cases). An equal number of controls (children/adolescents without any psychiatric disorder) were then randomly sampled from the stratum concerned (so for males in site 1 in the youngest age category and the lowest SES group, nine controls were sampled). Thus, the controls were frequency matched to the cases on site, sex, age, and SES.

Power Calculation

We conducted a *post-hoc* power analysis for this study. The sample size calculation was done using an online calculator (http://osse.bii.a-star.edu.sg/calculation1.php). The default setting for a case–control study was used, with significance level of 5%, minor allele frequency in cases of 15%, minor allele frequency in controls of 7%, and ratio of cases to controls at 1:1. Given 309 cases and an equal number of controls with complete genotypic data, our study achieved a power of 89%.

Assessment of Psychiatric Disorders and Socio-Demographic Factors

The details of the procedures has been published elsewhere (37). Psychiatric disorders, including IDs, were assessed using the Children and Adolescent Symptoms Invetory-5 (CASI-5) (Gadow and Sprafkin, 2013). The CASI-5 has been locally adapted for use in Uganda (38). The CASI-5 was administered by trained psychiatric nurses and psychiatric clinical officers at two time points (baseline and 12 months). Individual CASI-5 items are rated on a four-point frequency of occurrence scale ranging from never (0) to very often (3). The socio-demographic characteristics for both the participants and their caregivers were collected using a socio-demographic proforma developed by the Mental Health Section of the MRC/Uganda Virus Research Institute and LSHTM Uganda Research Unit. The socio-economic status was generated from a scale of nine household items (car, motorcycle, refrigerator, electricity, bicycle, radio, telephone, cupboard, and flask). Each item was weighted in the respective order, with a car carrying a maximum weight of 9 and a flask a minimum weight of 1. A total score of the items was generated, whose median cutoff of 13 was used to classify low and high SES. A score <13 was classified as low SES, while that >13 was classified as high SES.

Assessment of Chronic and Recent Stress

The chronic and recent stress classes used in this study were generated by Kalungi et al. (37). Briefly, social disadvantage was used as a proxy of chronic stress, with a composite index

constructed based on the following variables: orphanhood, food availability, study site, and level of education of the caregiver. Hierarchical cluster analysis (HCA) was used to generate the different cutoff points for each of the chronic stress classes: mild with chronic stress scores of 0–1.375, moderate with scores ranging from 1.375 to 2.375, and severe with scores >2.375. A composite index for recent stress was constructed from the following variables: caregiver mental state, child–caregiver relationship, and HIV symptoms. Three recent stress classes were generated using HCA, i.e., mild with scores <0.362, moderate with scores between 0.362 and 0.622, and severe with scores >0.622.

DNA Extraction

Genomic DNA was extracted from a whole-blood specimen for each participant using the Qiamp mini DNA extraction kit (Qiagen GmbH, Germany), following the instructions of the manufacturer. Extracted DNA was quantified by ultraviolet spectrophotometry on NanoDrop 1000 spectrophotometer V3.7 (Thermo Fisher Scientific, Wilmington, MA). The DNA was subsequently diluted to 5 ng/μl, and 5 μl of each of the DNA samples was run on 1% agarose gel to assess the quality. The extracted DNA was then stored for ∼3 months at −20°C while awaiting further molecular analysis.

Kompetitive Allele-Specific PCR Genotyping

The Kompetitive Allele-Specific PCR (KASP) genotyping of rs10482605 (within *NR3C1*) and rs1360780 (within *FKBP5*) polymorphisms was conducted by LGC Genomics Ltd., United Kingdom. Briefly, the PCR reaction was carried out in 10 μl of reaction mixture containing 5 μl of 2X KASP master mix, KASP assay mix (primers) at 0.14 and 5 μl (5 ng/μl) of genomic DNA. The DNA sample was diluted to a final concentration of 5–10 ng per reaction. All reagents were vortex-mixed prior to use. The cycling conditions were as follows: one cycle of hot start activation at 94°C for 15 min, followed by 10 cycles of denaturation, annealing, and extension at 94°C for 20 s and 61–55°C for 60 s (dropping 0.6°C per cycle), respectively. A further 26 cycles were performed, consisting of denaturation at 94°C for 20 s and annealing and extension at 55°C for 60 s (39). Distilled water and PCR mixture without the DNA template were included on each genotyping plate as negative controls. A difference in fluorescent signal intensity between the presence and absence of template DNA allowed improved confidence in the validity of the genotyping results. Individual samples that failed to genotype for a particular SNP were removed from the analysis.

Statistical Analysis

Statistical analyses were conducted using Stata 15 (StataCorp, TX, USA). The socio-demographic characteristics (including SES) were dichotomized and described between cases and controls. The socio-economic status and chronic and recent stress were measured as described by Kalungi et al. (37). Hierarchical cluster analysis was used to generate the different cutoff points for each of the chronic and recent stress classes (mild, moderate, and severe). Chi-square tests were used to assess the differences between the socio-demographic characteristics according to ID status at baseline (cases vs. controls). The allele frequencies of the studied genes were tested for Hardy–Weinberg equilibrium (HWE), and the distribution of genotype in cases and controls was compared using chi-square tests. Any possible relationship between polymorphisms in each gene and ID was investigated using chi-square tests. Logistic regression models were used to investigate the association between recent stress, chronic stress, genotypes, and IDs and to calculate the odds ratio and 95% confidence intervals. In these models, recent and chronic stress were included as predictor variables, SNPs were entered as covariates, and ID case-control status was the binary outcome variable. The interaction effect of chronic stress/recent stress on the association between the polymorphisms (rs10482605 and rs1360780) and IDs was tested using a likelihood ratio test.

Ethical Considerations

The present study was nested in a parent study (37), which obtained ethical and scientific clearance from the Uganda Virus Research Institute Science and Ethical Committee (#GC/127/15/06/459) and the Uganda National Council of Science and Technology (#HS 1601). Written informed consent was obtained from all the caregivers for their children/adolescents to participate in the study and for a blood specimen to be drawn from their children/adolescents for genetic analyses. The adolescents provided further written informed assent to participate in the study. The study participants diagnosed with significant psychiatric problems were referred to the mental health units at Entebbe and Masaka government hospitals. Personal identifiers and data were kept under lock and key and were only available to the study team. Administrative clearance was obtained from the Uganda Virus Research Institute—MOH for the use of the archived samples.

RESULTS

Socio-Demographic Characteristics of the Study Participants

A total of 736 participants (389 children and 307 adolescents) were enrolled into this study from the Kampala and Masaka districts. Out of the 736, about 44% were from Kampala, and 56% were recruited from Masaka. The subjects were purposively selected and age- and sex-matched based on the availability of 368 presenting with at least one ID, hence giving both controls and cases equal representation of 50% (368/736 each). The rs10482605 polymorphism was assessed in 624 (309 cases and 315 controls) participants with complete data, while 685 participants (335 controls and 350 cases) had complete data for inclusion in the rs1360780 polymorphic analysis. The socio-demographic characteristics did not differ by disease status (**Table 1**). The socio-demographic characteristics were analyzed as described in the "Statistical Analysis" section.

TABLE 1 | Distribution of socio-demographic characteristics between cases and controls.

Characteristics	Controls, *n* (%)	Cases, *n* (%)	Total, *n* (%)	*P*
Gender				
Female	207 (56.40)	186 (50.54)	393 (53.47)	0.111
Male	160 (43.60)	182 (49.46)	342 (46.53)	
Age category				
Adolescents	149 (42.45)	158 (45.80)	307 (44.11)	0.374
Children	202 (57.55)	187 (54.20)	389 (55.89)	
Educational level				
No formal education	9 (2.45)	4 (1.09)	13 (1.77)	0.371
Pre-primary	323 (88.01)	325 (88.80)	648 (88.40)	
Secondary	35 (9.54)	37 (10.11)	72 (9.82)	
Socio-economic status				
High	197 (53.53)	207 (56.25)	404 (54.89)	0.459
Low	171 (46.47)	161 (43.75)	332 (45.11)	
Study site				
Kampala	165 (44.84)	156 (42.39)	321 (43.61)	0.504
Masaka	203 (55.16)	212 (57.61)	415 (56.39)	

n, number of participants; %, percentage; children, 5–11 years; adolescents, 12–17 years.

Distribution of the rs10482605 and rs1360780 Minor Allele Frequencies and Genotypes Between Cases and Controls

The comparison between minor allele frequencies of both rs10482605 (within *NR3C1*) and rs1360780 (within *FKBP5*) SNPs in the studied population and other ethnicities is shown in **Table 2**. The minor allele frequencies of rs10482605 and rs1360780 polymorphisms were 0.121 and 0.396, respectively, in this study population.

Frequencies and Genotype Distribution of NR3C1 rs10482605

The rs10482605 polymorphism was assessed in 624 (309 cases and 315 controls) participants with complete data. In particular, out of 315 participants without IDs, five carried the homozygous minor *GG* genotype for the rs10482605 polymorphism, while 60 and 250 participants carried the heterozygous *GA* and wild-type *AA* forms, respectively. Similarly, out of 309 subjects with IDs, four possessed the homozygous minor *GG* genotype of the rs10482605 polymorphism; 73 and 232 were heterozygous and homozygous wild type, respectively. In comparison, 44.44 and 55.56, 54.89 and 45.11, and 48.13 and 51.87% of cases and controls possessed the mutant, heterozygous, and wild type of the *NR3C1* genotypes, respectively (**Table 3**). According to the results of a chi-square test, there was no significant difference in the frequencies and genotype distribution of the rs10482605 polymorphism between cases and control populations ($P > 0.369$). The frequency of the minor *G* allele in the rs10482605 polymorphic site was 0.121 among the population studied (**Table 2**). The genotype distribution for the subjects did not significantly deviate from HWE ($P = 0.9595$).

Frequencies and Genotype Distribution of FKBP5 rs1360780

A total of 685 participants (335 controls and 350 cases) had complete data for inclusion in the rs1360780 polymorphic analysis, as shown in **Table 2**. The number of individuals carrying the homozygous minor *TT*, heterozygous *CT*, and wild-type *CC* genotypes were 60, 154, and 121, respectively, among the controls and 51, 166, and 133, respectively, among the cases (**Table 3**). The genotype frequencies were 54.05 and 45.95%, 48.13 and 51.88%, and 47.64 and 52.36% for homozygous minor *TT*, heterozygous *CT*, and wild-type *CC* among controls and cases, respectively (**Table 3**). There was no significant difference observed in the frequency and distribution of the different genotypes between the cases and controls for the rs1360780 polymorphism using the chi-square test ($P = 0.492$). In addition, the genotype distribution for subjects did not significantly deviate from the HWE as evidenced by the non-significant difference between the observed and expected alleles and genotype values of the rs1360780 polymorphisms ($P = 0.545$). The minor allele *T* frequency in this population was 0.396 (**Table 2**).

Genotypic and Environmental Predictors of IDs in CA-HIV From Kampala and Masaka Districts—Uganda

A total of 309 cases had complete data for inclusion in the analyses of the association between IDs and the potential risk factors of the investigated SNPs and environmental factors (acute and chronic stress). None of the allelic forms of the SNPs studied had any significant influence on the increased likelihood of developing IDs (**Table 4**). Using logistic regression analysis, the decreased risk of developing IDs in individuals with the *G* minor allele of the rs10142605 was not significant ($P = 0.143$) compared to individuals possessing the wild-type form. In individuals carrying the homozygous minor *T* allele of the rs1360780 variant, the odds of developing IDs was likewise non-significantly reduced ($P = 0.329$). Recent stress (specifically severe recent stress), as opposed to chronic stress, significantly increased the likelihood of developing IDs in CA-HIV as shown

TABLE 2 | Genome location and minor allele frequency of the rs10482605 and rs1360780 single nucleotide polymorphisms among different populations.

Genome location				Minor allele frequency					
Gene region	**Xsome**	**Nucleotide change**	**DBSNP Ref**	**1000G-Global**	**1000G-African**	**Est**	**Swe**	**Sib**	**Ug**
NR3C1-promoter	Chr5	A>G	rs10482605	0.10	0.08	0.12	0.13	0.50	0.12
FKBP5-intron 2	Chr6	T>C	rs1360780	0.33	0.42	0.26	0.28	0.24	0.39

Xsome, Chromosome; Est, Estonian; Swe, Sweden; Sib, Siberia; Ug, Uganda; DBSNP Ref, database single nucleotide polymorphism reference (allele frequency adopted from 1,000 genomes; https://www.ncbi.nlm.nih.gov/snp/).

TABLE 3 | Distribution of rs10482605 and rs1360780 genotypes between cases and controls.

Genotypes	**Cases, *n* (%)**	**Controls *n* (%)**	***P***
rs10482605	*N* = 309	*N* = 315	
Wild type (*AA*)	232 (48.13)	250 (51.87)	0.369
Heterozygous (*GA*)	73 (54.89)	60 (45.11)	
Mutant (*GG*)	4 (44.44)	5 (55.56)	
rs1360780	*n* = 350	*n* = 335	
Wild type (*CC*)	133 (52.36)	121 (47.64)	0.492
Heterozygous (*CT*)	166 (51.88)	154 (48.13)	
Mutant (*TT*)	51 (45.95)	60 (54.05)	

n, number of participants; %, percentage.

TABLE 4 | Association between chronic and recent stress and internalizing disorders (IDs) in children and adolescents living with human immunodeficiency virus from Kampala and Masaka Districts—Uganda.

Predictors of IDs	***n* = 309**	
	OR (95% CI)	***P***
Environmental factors		
Chronic stress		
Mild	1.0	Reference
Moderate	0.69 (0.49–0.95)	0.025[a]
Severe	0.92 (0.61–1.37)	0.673
Recent stress		
Mild	1.0	Reference
Moderate	1.07 (0.76–1.52)	0.687
Severe	1.85 (1.29–2.65)	0.001[a]
Genotype		
NR3C1 rs10482605		
Wild type (*AA*)	1.0	Reference
Heterozygous (*GA*)	1.34 (0.91–1.99)	0.140
Mutant (*GG*)	0.73 (0.19–2.81)	0.649
***FKBP5* rs1360780**		
Wild type (*CC*)	1.0	Reference
Heterozygous (*CT*)	1.08 (0.76–1.55)	0.636
Mutant (*TT*)	0.78 (0.48–1.27)	0.329

CI, confidence interval; OR, odds ratio.
[a] *Statistically significant.*

in **Table 4** ($P = 0.001$). Moderate chronic stress appears to protect against IDs ($P = 0.026$).

We interrogated the data to further assess whether individuals carrying the polymorphic risk variants (we predicted that these participants would be more vulnerable to the effect of environmental adversity) are more likely to develop psychopathology. All the environmental adversities assessed were summed up and dichotomized as chronic or recent stress as described. Using likelihood ratio analysis, there was no significant association between gene–environment interactions and susceptibility to IDs ($P > 0.050$), as shown in **Table 5**.

DISCUSSION

This study determined the allele and genotype frequencies of SNPs in *NR3C1* (rs10482605) and *FKBP5* (rs1360780) and examined the relationship between these SNPs, environmental stress, and vulnerability to IDs in CA-HIV in Kampala and Masaka districts in Uganda. The minor allele frequencies of rs10482605 and rs1360780 polymorphisms were 0.121 and 0.396, respectively, among this study population. The rs1360780 minor allele frequency of 0.396 is similar to the 0.39–0.45 documented in African Americans (40) but somewhat higher than those in Estonian, Swedish, and Siberian populations. Similarly, the rs10482605 minor allele frequency of 0.121 is higher than the global and African minor allele frequencies reported in the 1,000 Genomes, equal to those in the Estonians (0.12) and Swedish (0.13) but lower than those in the Siberian (0.50) populations. The inconsistencies in polymorphism frequencies might be due to genetic ancestry-related differences and reinforces the importance of studying genetically driven susceptibility to disorders in diverse populations. Our result show that *AA* was the most prevalent genotype of the rs10482605 SNP with a frequency of 77.2%, whereas the frequency of the minor *GG* genotype was 1.44%. This finding mirrors the frequencies of 93.1, 1.6, and 5.3% for *AA*, *AG*, and *GG*, respectively, reported among the Chinese population (41). However, this result conflicts with a previous study conducted in a German population where genotype frequencies of 67, 27.6, and 4.4% for *GG*, *AG*, and *AA*, respectively, were reported (42). These similarities and differences are probably due to differences in ancestry. With regards to the rs1360780 SNP in our sample, the following genotype frequencies of 37.0, 46.7, and 16.3% for *CC*, *CT*, and *TT*, respectively, were reported. Schneider et al. (42) also reported similar findings among the German population.

TABLE 5 | Gene-environment interaction.

Gene-environment interaction	OR (95% CI)	*P*	Overall *P*
rs10482605 × chronic stress (*N* = 624)			
AA—mild chronic stress	1.0	Reference	0.340
GA—moderate chronic stress	2.56 (105–6.24)	0.039	
GA—severe chronic stress	1.36 (0.49–3.76)	0.549	
GG—moderate chronic stress	0.765 (0.19–30.6)	0.887	
GG—severe chronic stress	1.09 (0.35–33.95)	0.959	
rs10482605 × recent stress (*N* = 624)			
AA—mild recent stress	1.0	Reference	0.553
GA—moderate recent stress	0.828 (0.33–2.07)	0.687	
GA—severe recent stress	1.75 (0.64–4.75)	0.270	
GG—moderate recent stress	0.939 (0.06–13.9)	0.964	
GG—severe recent stress	1.00	omitted	
rs1360780 × chronic stress (*N* = 685)			
CC—mild chronic stress	1.0	Reference	0.743
CT—moderate chronic stress	1.41 (0.673–2.962)	0.361	
CT—severe chronic stress	1.59 (0.629–4.039)	0.325	
TT—moderate chronic stress	0.83 (0.297–2.336)	0.728	
TT—severe chronic stress	1.2 (0.359–4.01)	0.767	
rs1360780 × recent stress (*N* = 685)			
CC—mild recent stress	1.0	Reference	0.2628
CT—moderate recent stress	2.44 (1.086–5.50)	0.031	
CT—severe recent stress	1.17 (0.52–2.63)	0.701	
TT—moderate recent stress	1.41 (0.487–4.101)	0.523	
TT—severe recent stress	0.894 (0.286–2.795)	0.848	

CI, confidence interval; OR, odds ratio.

The distribution as well as the role of the genotypes discussed above in vulnerability to IDs was investigated using a case–control study design. Overall, there was no significant difference in genotype distribution/frequencies between case and control populations for both rs10482605 and rs1360780. Remarkably, there was no association between the carriage of the rs10482605 polymorphic *G* risk allele in this population and vulnerability to IDs, consistent with previous studies among child psychiatric population (43) and the Chinese Han population (41). However, the rs10482605 polymorphism, which is in linkage disequilibrium with rs6198, is associated with increased GR beta mRNA stability, and previous authors have reported the indirect effects of this gene with the IDs of depression and anxiety (26, 42). Possession of the rs10482605 risk allele was associated with childhood mood disorders in a mixed Canadian population (44) and recurrent major depression in Belgian and Swedish populations (43). Recently, Womersley et al. in South Africa, reported that the rs10482605 risk allele was associated with increased sensitivity to anxiety among adolescents (45). In this study, the proportion of participants carrying the *TT* polymorphic risk allele of rs1360780, which is located in intron 2 of *FKBP5*, did not differ significantly between cases and controls. Schneider et al. in a study among pregnant women in Germany, observed no association between polymorphisms in the *FKBP5* gene and depressive scores (42). Other researchers, however, have reported conflicting results. Menke et al. reported a significant association between the *TT FKBP5* alleles and depression among an adult German population (46). Binder et al. in another German study, reported a significant association between polymorphisms in the *FKBP5* gene and the recurrence of depressive episodes and response to antidepressants (47). In a European study conducted among depression treatment-resistant adolescents, *FKBP5* rs1360780 was associated with suicide, a distal correlate of depression (48). It is plausible that the rs1360780 and rs10482605 polymorphisms may not be involved in the pathophysiology of IDs in CA-HIV in the Ugandan population.

There are a number of possible reasons for these discrepant results. Firstly, on a more general conceptual level, the IDs of depression and anxiety have a multi-factorial etiology that includes biological (including genetic) and environmental risk factors (5, 9, 49, 50). It is now believed that the predisposition to IDs, such as depression, is determined by the coordinated action of many genes and their interaction with each other and with diverse environmental factors (51). We did not observe any gene–environment effect on vulnerability to IDs in this population. It is also likely that each gene, by itself, makes a relatively small contribution to the pathogenesis of the disease (52). As a result of this and as shown in this study, it has been difficult to replicate genetic associations. Shadrina et al. (51), in a recent review article, highlighted the challenges of replicating genetic association studies despite the fact that a large number of

the associations between genes and different clinical depression variants and depressive sub-phenotypes have been published.

Secondly, another reason for the discrepant results could be methodological; the sample size for this study may have been inadequate to identify the small contribution that each of these SNPs is hypothesized to contribute to the risk of the IDs of depression and anxiety. Hence, this study was probably not adequately powered to test the association between these SNPs and IDs. This observation is supported by recent efforts to identify common genetic variants associated with depression in genome-wide association studies (GWAS), which have previously been unsuccessful largely because of inadequate samples. In a recent successful GWAS, in order to maximize the sample size, Howard et al. (49), in a meta-analysis, used data on 807,553 individuals (246,363 cases and 561,190 controls) and were able to identify 102 independent variants associated with depression. Lastly, another factor that may have contributed to the absence of an association between the investigated SNPs and IDs is the differences in participant ancestry between this and previous studies. While this study was undertaken among respondents of African ancestry, all the previous studies that reported positive results were of Caucasian ancestry (46–48). It is possible that it is not the rs1360780 and rs10482605 SNPs themselves that increase the risk for IDs but rather another variant or variants that are in close linkage disequilibrium (53). Ancestry-based differences in haplotype structure may thus account for discrepant findings across different study groups. Furthermore, ancestry may also affect the degree to which genetic variants play a role in determining local and more distal gene expression (54).

In this study, we found an association between IDs and both recent and chronic stress. The odds of IDs increased with increasing severe recent stress and decreased with moderate chronic stress. This finding is consistent with previous studies where recent stress has been associated with the IDs of depression (9, 55) and anxiety (56). Interestingly, the odds of developing IDs decreased with moderate chronic stress in CA-HIV. The seemingly protective effect of moderate chronic stress in this study could not be easily explained but could be attributed to physiological and biochemical cross-adaptations. Moderate chronic stress has also been shown to be associated with alleviation of depressive symptoms in open water swimmers (57). However, a recent study in animal models showed that, in young adult female mice, chronic stress resulted in decreased anxiety-like behavior and enhanced cognitive performance, whereas in old female mice it led to weight loss, dysregulated locomotion, and memory impairment, suggesting that the effect of chronic stress on behavior is age dependent (58). These results may have underlined the findings in our study as the study respondents were relatively very young (aged between 5 and 17 years). Further empirical evidence is warranted to understand the mechanisms underpinning the observed positive clinical effect. This could revolutionize the extant therapies and lead to potential non-drug pathways to recovery from IDs.

In conclusion, the results from this study seem to give credence to the stress-vulnerability diathesis of depression as postulated by Ising and Holsboer (59), where acute and chronic stress were both associated with IDs. However, in departure from the stress-vulnerability diathesis of Ising and Holsboer (59), moderate chronic stress among young respondents in this study seemed to be protective against IDs. These results, however, need to be replicated by further studies. Additionally, while the results from this study showed the contribution of psychosocial stressors on ID vulnerability, they did not show a contribution of genetic factors to vulnerability as originally hypothesized. We interrogated the data to further assess whether individuals carrying the polymorphic risk variants (we predicted that these participants would be more vulnerable to the effect of environmental adversity) were more prone to developing a psychopathology. Using logistic regression analysis, there was no significant association between gene–environment interaction and susceptibility to IDs.

The study has some limitations: (1) recall bias: the number and nature of adverse environmental episodes in the past month were obtained retrospectively from caregivers and participants (in the parent CHAKA study) by a psychiatric clinician during history taking. Studies have shown that the retrospective reporting of childhood trauma is valid (60) and is not influenced by psychiatric state at the time of the report (61). In this regard, it is noteworthy that all the participants were interviewed using standardized tools and by trained psychiatric clinicians; (2) study design: the study sample was small, given the relatively small contribution that each SNP makes to the risk of depression. Furthermore, our study was limited to two SNPs. Though these were selected based on previous publications, the more comprehensive, and expensive, genome- or transcriptome-wide approaches are better suited to investigating the role of multiple genetic variants in complex disorders such as IDs; and (3) control group composition: we acknowledge that factors such as ART could associate with ID status and influence the study findings. We also acknowledge that HIV positivity status constitutes a chronic stressor through the persistent disturbance of the immune, cardiovascular, metabolic, and behavioral homeostasis (62). However, any potential confounding effect of HIV diagnosis on the qualification of chronic stress in this study is buffered out by the fact that both cases and controls were HIV-positive. Nonetheless, this study would be preferred among a general population; but, since it was nested in a parent study which investigated psychiatric disorders among Ugandan HIV+ children and adolescents, the clinical data and sample available to us was in this cohort.

CONCLUSION

Neither of the SNPs investigated in this study was significantly associated with vulnerability to IDs. However, as previously reported by others, severe recent stress was a significant risk factor of IDs. Additionally, in this study, moderate chronic stress seemed to be a protective factor against IDs, an observation that has only been previously reported in animal models. We did not observe any gene–environment effect on vulnerability to IDs in this population.

RECOMMENDATION

There is a need to undertake a GWAS to illuminate the role of genetic factors in the risk of IDs among respondents of African ancestry. The observation of the possible protective role against IDs of moderate chronic stress among relatively young respondents observed in this study warrants further inquiry.

DATA AVAILABILITY STATEMENT

All information pertaining to participants and samples remain confidential with access limited to research team. However, upon request data from the MRC/UVRI and LSHTM Uganda research unit is currently accessed under a data sharing policy via http://www.mrcuganda.org/sites/default/files/publications/MRC_UVRI_Data_Sharing_Policy_December2015.pdf. Requests to access these datasets should be directed to http://www.mrcuganda.org/sites/default/files/publications/MRC_UVRI_Data_Sharing_Policy_December2015.pdf.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Uganda Virus Research Institute (UVRI) Science and Ethical Committee (# GC/127/15/06/459), the Uganda National Council of Science and Technology (# HS 1601). Written informed consent was obtained from all the caregivers for their children/adolescents to participate in the study and for a blood specimen to be drawn from their children/adolescents for genetic analyses. Adolescents provided further written informed assent to participate in the study. Personal identifiers and data were kept under lock and key and were only available to the study team. Administrative clearance was obtained from the UVRI-MOH for the use of the archived samples. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

TJO and AK conceived and conducted the study and drafted the first manuscript. WJS carried out the statistical analysis. DM, JSW, and EK supervised and edited the manuscript. All authors reviewed and approved the manuscript for publication.

FUNDING

This study was funded by a grant from the Alliance for Global Health Sciences, which was awarded to AK (Grant No. 50288/N7145). The parent study investigating mental health among children and adolescents living with HIV/AIDS in Kampala and Masaka districts of Uganda (CHAKA study) was funded by a Medical Research Council/Department for International Development African Research Leader award to EK (Grant No. MR/L004623/1).

ACKNOWLEDGMENTS

The authors extend their gratitude to the study subjects and the Data and Statistics Section of the MRC/UVRI and LSHTM Uganda Research Unit for all the support in statistical analyses.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fped.2021.666426/full#supplementary-material

REFERENCES

1. Marquez PV, Saxena S. Making mental health a global priority. *InCerebrum: The Dana Forum on Brain Science.* Dana Foundation (2016).
2. Kieling C, Baker-Henningham H, Belfer M, Conti G, Ertem I, Omigbodun O, et al. Child and adolescent mental health worldwide: evidence for action. *Lancet.* (2011) 378:1515–25. doi: 10.1016/S0140-6736(11)60827-1
3. Belfer ML. Critical review of world policies for mental healthcare for children and adolescents. *Curr Opin Psychiatry.* (2007) 20:349–52. doi: 10.1097/YCO.0b013e3281bc0cf4
4. Vreeman RC, McCoy BM, Lee S. Mental health challenges among adolescents living with HIV. *J Int AIDS Soc.* (2017) 20:21497. doi: 10.7448/IAS.20.4.21497
5. Kinyanda E, Salisbury TT, Levin J, Nakasujja N, Mpango RS, Abbo C, et al. Rates, types and co-occurrence of emotional and behavioural disorders among perinatally HIV-infected youth in Uganda: the CHAKA study. *Soc Psychiatry Psychiatr Epidemiol.* (2019) 54:415–25. doi: 10.1007/s00127-019-01675-0
6. Gaughan DM, Hughes MD, Oleske JM, Malee K, Gore CA, Nachman S. Psychiatric hospitalizations among children and youths with human immunodeficiency virus infection. *Pediatrics.* (2004) 113:e544–51. doi: 10.1542/peds.113.6.e544
7. Bankole KO, Bakare MO, Edet BE, Igwe MN, Ewa AU, Bankole IA, et al. Psychological complications associated with HIV/AIDS infection among children in South-South Nigeria, sub-Saharan Africa. *Cogent Med.* (2017) 4:1372869. doi: 10.1080/2331205X.2017.1372869
8. Dessauvagie AS, Jörns-Presentati A, Napp AK, Stein DJ, Jonker D, Breet E, et al. The prevalence of mental health problems in sub-Saharan adolescents living with HIV: a systematic review. *Glob Ment Health.* (2020) 7:e29. doi: 10.1017/gmh.2020.18
9. Kinyanda E, Hoskins S, Nakku J, Nawaz S, Patel V. Prevalence and risk factors of major depressive disorder in HIV/AIDS as seen in semi-urban Entebbe district, Uganda. *BMC Psychiatry.* (2011) 11:205. doi: 10.1186/1471-244X-11-205
10. Jörns-Presentati A, Napp AK, Dessauvagie AS, Stein DJ, Jonker D, Breet E, et al. The prevalence of mental health problems in sub-Saharan adolescents: a systematic review. *PLoS ONE.* (2021) 16:e0251689. doi: 10.1371/journal.pone.0251689
11. Forns M, Abad J, Kirchner, T. editors. Internalizing and Externalizing Problems. In: *Encyclopedia of Adolescence.* New York, NY: Springer (2011). p. 1464–9. doi: 10.1007/978-1-4419-1695-2_261
12. Achenbach TM, Ivanova MY, Rescorla LA, Turner LV, Althoff RR. Internalizing/externalizing problems: Review and recommendations for clinical and research applications. *J Am Acad Child Adolesc Psychiatry.* (2016) 55:647–56. doi: 10.1016/j.jaac.2016.05.012
13. Nachman S, Chernoff M, Williams P, Hodge J, Heston J, Gadow KD. Human immunodeficiency virus disease severity, psychiatric symptoms, and functional outcomes in perinatally infected youth. *Arch Pediatri Adolesc Med.* (2012) 166:528–35. doi: 10.1001/archpediatrics.2011.1785

14. Knizek BL, Mugisha J, Osafo J, Kinyanda E. Growing up HIV-positive in Uganda:"psychological immunodeficiency"? A qualitative study. *BMC Psychol.* (2017) 5:30. doi: 10.1186/s40359-017-0199-7
15. Mukangabire P, Katende G, Rugema J, Moreland P, Nkurunziza A. Factors associated with depression among adolescents living with HIV at health facilities in Rwanda. In: *Sigma's 30th International Nursing Research Congress*. Available online at: http://hdl.handle.net/10755/17685
16. Hammen C, Kim EY, Eberhart NK, Brennan PA. Chronic and acute stress and the prediction of major depression in women. *Depress Anxiety.* (2009) 26:718–23. doi: 10.1002/da.20571
17. Conway CC, Rutter LA, Brown TA. Chronic environmental stress and the temporal course of depression and panic disorder: a trait-state-occasion modeling approach. *J Abnorm Psychol.* (2016) 125:53–63. doi: 10.1037/abn0000122
18. Taporoski TP, Negrão AB, Horimoto AR, Duarte NE, Alvim RO, de Oliveira CM, et al. Shared genetic factors of anxiety and depression symptoms in a Brazilian family-based Cohort, the Baependi Heart Study. *PLoS ONE.* (2015) 10:e0144255. doi: 10.1371/journal.pone.0144255
19. McGregor NW, Dimatelis JJ, Van Zyl PJ, Hemmings SM, Kinnear C, Russell VA, et al. A translational approach to the genetics of anxiety disorders. *Behav Brain Res.* (2018) 341:91–7. doi: 10.1016/j.bbr.2017.12.030
20. Cimino S, Cerniglia L, De Carli P, Simonelli A. DNA methylation and allelic polymorphism at the dopamine transporter promoter affect internalizing and externalizing symptoms in preschoolers. *Child Psychiatry Hum Dev.* (2021) 52:281–90. doi: 10.1007/s10578-020-01009-1
21. Voltas N, Aparicio E, Arija V, Canals J. Association study of monoamine oxidase-A gene promoter polymorphism (MAOA-uVNTR) with self-reported anxiety and other psychopathological symptoms in a community sample of early adolescents. *J Anxiety Disord.* (2015) 31:65–72. doi: 10.1016/j.janxdis.2015.02.004
22. Chang CC, Chang HA, Fang WH, Chang TC, Huang SY. Gender-specific association between serotonin transporter polymorphisms (5-HTTLPR and rs25531) and neuroticism, anxiety and depression in well-defined healthy Han Chinese. *J Affect Disord.* (2017) 207:422–8. doi: 10.1016/j.jad.2016.08.055
23. Kalungi A, Seedat S, Hemmings SMJ, van der Merwe L, Joloba ML, Nanteza A, et al. Association between serotonin transporter gene polymorphisms and increased suicidal risk among HIV positive patients in Uganda. *BMC Genet.* (2017) 18:71. doi: 10.1186/s12863-017-0538-y
24. Kalungi A, Womersley JS, Kinyanda E, Joloba ML, Ssembajjwe W, Nsubuga RN, et al. The 5-HTTLPR-rs25531 S-A-S-A haplotype and chronic stress moderate the association between acute stress and internalizing mental disorders among HIV+ children and adolescents in Uganda. *Front Genet.* (2021) 12:649055. doi: 10.3389/fgene.2021.649055
25. Pérez-Pérez B, Cristóbal-Narváez P, Sheinbaum T, Kwapil TR, Ballespí S, Peña E, et al. Interaction between FKBP5 variability and recent life events in the anxiety spectrum: evidence for the differential susceptibility model. *PLoS ONE.* (2018) 13:e0193044. doi: 10.1371/journal.pone.0193044
26. Zimmermann P, Brückl T, Nocon A, Pfister H, Binder EB, Uhr M, et al. Interaction of FKBP5 gene variants and adverse life events in predicting depression onset: results from a 10-year prospective community study. *Am J Psychiatry.* (2011) 168:1107–16. doi: 10.1176/appi.ajp.2011.10111577
27. Gałecka E, Szemraj J, Bieńkiewicz M, Majsterek I, Przybyłowska-Sygut K, Gałecki P, et al. Single nucleotide polymorphisms of NR3C1 gene and recurrent depressive disorder in population of Poland. *Mol Boil Rep.* (2013) 40:1693–9. doi: 10.1007/s11033-012-2220-9
28. Kumsta R, Moser D, Streit F, Koper JW, Meyer J, Wüst S. Characterization of a glucocorticoid receptor gene (GR, NR3C1) promoter polymorphism reveals functionality and extends a haplotype with putative clinical relevance. *Am J Med Genet B Neuropsychiatr Genet.* (2009) 150:476–82. doi: 10.1002/ajmg.b.30837
29. Halldorsdottir T, Binder EB. Gene x environment interactions: from molecular mechanisms to behavior. *Ann Rev Psychol.* (2017) 68:215–41. doi: 10.1146/annurev-psych-010416-044053
30. Zannas AS, Wiechmann T, Gassen NC, Binder EB. Gene-Stress-Epigenetic regulation of FKBP5: clinical and translational implications. *Neuropsychopharmacology.* (2016) 41:261–74. doi: 10.1038/npp.2015.235
31. Lekman M, Laje G, Charney D, Rush AJ, Wilson AF, Sorant AJ, et al. The FKBP5-gene in depression and treatment response–an association study in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) cohort. *Biol Psychiatry.* (2008) 63:1103–10. doi: 10.1016/j.biopsych.2007.10.026
32. Zobel A, Schuhmacher A, Jessen F, Höfels S, Von Widdern O, Metten M, et al. DNA sequence variants of the FKBP5 gene are associated with unipolar depression. *Int J Neuropsychopharmacol.* (2010) 13:649–60. doi: 10.1017/S1461145709991155
33. Szczepankiewicz A, Leszczyńska-Rodziewicz A, Pawlak J, Narozna B, Rajewska-Rager A, Wilkosc M, et al. FKBP5 polymorphism is associated with major depression but not with bipolar disorder. *J Affect Disord.* (2014) 164:33–37. doi: 10.1016/j.jad.2014.04.002
34. Papiol S, Arias B, Gastó C, Gutiérrez B, Catalán R, Fañanás, et al. Genetic variability at HPA axis in major depression and clinical response to antidepressant treatment. *J Affect Disord.* (2007) 104:83–90. doi: 10.1016/j.jad.2007.02.017
35. Lavebratt C, Aberg E, Sjoholm LK, Forsell Y. Variations in FKBP5 and BDNF genes are suggestively associated with depression in a Swedish population-based cohort. *J Affect Disord.* (2010) 125:249–55. doi: 10.1016/j.jad.2010.02.113
36. Minelli A, Maffioletti E, Cloninger CR, Magri C, Sartori R, Bortolomasi M, et al. Role of allelic variants of FK506-binding protein 51 (FKBP5) gene in the development of anxiety disorders. *Depress Anxiety.* (2013) 30:1170–6. doi: 10.1002/da.22158
37. Kalungi A, Womersley JS, Kinyanda E, Joloba ML, Ssembajjwe W, Nsubuga RN, et al. Internalizing mental disorders and accelerated cellular aging among perinatally HIV-infected youth in Uganda. *Front Genet.* (2019) 10:705. doi: 10.3389/fgene.2019.00705
38. Mpango RS, Kinyanda E, Rukundo GZ, Osafo J, Gadow KD. Exploration of the understanding and etiology of ADHD in HIV/AIDS as observed by adolescents with HIV/AIDS, caregivers and health workers-using case vignettes. *Afr Health Sci.* (2018) 18:488–95. doi: 10.4314/ahs.v18i3.4
39. He J, Zhao X, Laroche A, Lu ZX, Liu H, Li Z. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front Plant Sci.* (2014) 5:484. doi: 10.3389/fpls.2014.00484
40. Roy A, Gorodetsky E, Yuan Q, Goldman D, Enoch MA. Interaction of FKBP5, a stress-related gene, with childhood trauma increases the risk for attempting suicide. *Neuropsychopharmacology.* (2010) 35:1674–83. doi: 10.1038/npp.2009.236
41. Lian Y, Xiao J, Wang Q, Ning L, Guan S, Ge H, et al. The relationship between glucocorticoid receptor polymorphisms, stressful life events, social support, and post-traumatic stress disorder. *BMC Psychiatry.* (2014)14:232. doi: 10.1186/s12888-014-0232-9
42. Schneider M, Engel A, Fasching PA, Häberle L, Binder EB, Voigt F, et al. Genetic variants in the genes of the stress hormone signalling pathway and depressive symptoms during and after pregnancy. *Biomed Res Int.* (2014) 2014:469278. doi: 10.1155/2014/469278
43. Van West D, Van Den Eede F, Del-Favero J, Souery D, Norrback KF, Van Duijn C, et al. Glucocorticoid receptor gene-based SNP analysis in patients with recurrent major depression. *Neuropsychopharmacology.* (2006) 31:620–7. doi: 10.1038/sj.npp.1300898
44. Mill J, Wigg K, Burcescu I, Vetro A, Kiss E, Kapornai K, et al. Mutation screen and association analysis of the glucocorticoid receptor gene (NR3C1) in childhood-onset mood disorders (COMD). *Am J Med Genet B Neuropsychiatr Genet.* (20090 150:866–73. doi: 10.1002/ajmg.b.30909
45. Womersley JS, Martin LI, van der Merwe L, Seedat S, Hemmings SM. Hypothalamic-pituitary-adrenal axis variants and childhood trauma influence anxiety sensitivity in South African adolescents. *Metab Brain Dis.* (2018) 33:601–13. doi: 10.1007/s11011-017-0138-6
46. Menke A, Klengel T, Rubel J, Brückl T, Pfister H, Lucae S, et al. Genetic variation in FKBP5 associated with the extent of stress hormone dysregulation in major depression. *Genes Brain Behav.* (2013) 12:289–96. doi: 10.1111/gbb.12026
47. Binder EB, Salyakina D, Lichtner P, Wochnik GM, Ising M, Pütz B, et al. Polymorphisms in FKBP5 are associated with increased recurrence of depressive episodes and rapid response to antidepressant treatment. *Nat genet.* (2004) 36:1319–25. doi: 10.1038/ng1479
48. Brent D, Melhem N, Turecki, G. Pharmacogenomics of suicidal events. *Pharmacogenomics.* (2010) 11:793–807. doi: 10.2217/pgs.10.64

49. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat neurosci.* (2019) 22:343–52. doi: 10.1038/s41593-018-0326-7
50. Musinguzi K, Obuku A, Nakasujja N, Birabwa H, Nakku J, Levin J, et al. Association between major depressive disorder and pro-inflammatory cytokines and acute phase proteins among HIV-1 positive patients in Uganda. *BMC Immunol.* (2018) 19:1. doi: 10.1186/s12865-017-0239-3
51. Shadrina M, Bondarenko EA Slominsky PA. Genetics factors in major depression disease. *Front Psychiatry.* (2018) 9:334. doi: 10.3389/fpsyt.2018.00334
52. Ebmeier KP, Donaghey C, Steele JD. Recent developments and current controversies in depression. *Lancet.* (2006) 367:153–67. doi: 10.1016/S0140-6736(06)67964-6
53. Polimanti R, Yang C, Zhao H, Gelernter, J. Dissecting ancestry genomic background in substance dependence genome-wide association studies. *Pharmacogenomics.* (2015) 16:1487–98. doi: 10.2217/pgs.15.91
54. Johnson EO, Hancock DB, Levy JL, Gaddis NC, Page GP, Glasheen C, et al. KAT2B polymorphism identified for drug abuse in African Americans with regulatory links to drug abuse pathways in human prefrontal cortex. *Addict Biol.* (2016) 21:1217–32. doi: 10.1111/adb.12286
55. Fullerton CS, Ursano RJ, Wang L. Acute stress disorder, posttraumatic stress disorder, and depression in disaster or rescue workers. *Am J Psychiatry.* (2004) 161:1370–6. doi: 10.1176/appi.ajp.161.8.1370
56. Grillon C, Duncko R, Covington MF, Kopperman L, Kling MA. Acute stress potentiates anxiety in humans. *Biol Psychiatry.* (2007) 62:1183–6. doi: 10.1016/j.biopsych.2007.06.007
57. Van Tulleken C, Tipton M, Massey H, Harper, C. M. Open water swimming as a treatment for major depressive disorder. *BMJ Case Rep*. (2018) 2018:bcr2018225007. doi: 10.1136/bcr-2018-225007
58. Lotan A, Lifschytz T, Wolf G, Keller S, Ben-Ari H, Tatarsky P, et al. Differential effects of chronic stress in young-adult and old female mice: cognitive-behavioral manifestations and neurobiological correlates. *Mol Psychiatry.* (2018) 23:1432–45. doi: 10.1038/mp.2017.237
59. Ising M, Holsboer F. Genetics of stress response and stress-related disorders. *Dialogues Clin Neurosci.* (2006) 8:433–44. doi: 10.31887/DCNS.2006.8.4/mising
60. Bifulco A, Brown GW, Lillie A, Jarvis J. Memories of childhood neglect and abuse: corroboration in a series of sisters. *J Child Psychol Psychiatry.* (1997) 38:365–74. doi: 10.1111/j.1469-7610.1997.tb01520.x
61. Fergusson DM, Woodward LJ, Horwood LJ. Risk factors and life processes associated with the onset of suicidal behavior during adolescence and early adulthood. *Psychol Med.* (2000) 30:23–39. doi: 10.1017/S00332917990 0135X
62. Chittiprol S, Kumar AM, Satishchandra P, Shetty KT, Rao RB, Subbakrishna DK, et al. Progressive dysregulation of autonomic and HPA axis functions in HIV-1 clade C infection in South India. *Psychoneuroendocrinology.* (2008) 33:30–40. doi: 10.1016/j.psyneuen.2007. 09.006

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

frontiers in Genetics

SYSTEMATIC REVIEW
published: 18 November 2021
doi: 10.3389/fgene.2021.733746

Genetic Studies of Metabolic Syndrome in Arab Populations: A Systematic Review and Meta-Analysis

*Zahrah Al-Homedi[1], Nariman Afify[1], Mashal Memon[1], Habiba Alsafar[2,3,4], Guan Tay[5,6], Herbert F. Jelinek[4,7], Mira Mousa[2,8], Nadia Abu-Samra[9] and Wael Osman[2,9]**

[1]College of Medicine and Health Sciences, Khalifa University, Abu Dhabi, United Arab Emirates, [2]Center for Biotechnology, Khalifa University, Abu Dhabi, United Arab Emirates, [3]Department of Genetics and Molecular Biology, College of Medicine and Health Sciences, Abu Dhabi, United Arab Emirates, [4]Department of Biomedical Engineering, College of Engineering, Khalifa University, Abu Dhabi, United Arab Emirates, [5]Department of Psychiatry, UWA Medical School, The University of Western Australia, Nedlands, WA, Australia, [6]School of Medical and Health Sciences, Edith Cowan University, Joondalup, WA, Australia, [7]Health Innovation Engineering Center, Khalifa University, Abu Dhabi, United Arab Emirates, [8]Nuffield Department of Women's and Reproduction Health, Oxford University, Oxford, United Kingdom, [9]Department of Biology, College of Arts and Sciences, Khalifa University, Abu Dhabi, United Arab Emirates

OPEN ACCESS

Edited by:
Karoline Kuchenbaecker, University College London, United Kingdom

Reviewed by:
Paul Lasko, McGill University, Canada
Vahidreza Ostadmohammadi, Kashan University of Medical Sciences, Iran

***Correspondence:**
Wael Osman
wael.osman@ku.ac.ae

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *30 June 2021*
Accepted: *20 October 2021*
Published: *18 November 2021*

Citation:
Al-Homedi Z, Afify N, Memon M, Alsafar H, Tay G, Jelinek HF, Mousa M, Abu-Samra N and Osman W (2021) Genetic Studies of Metabolic Syndrome in Arab Populations: A Systematic Review and Meta-Analysis. Front. Genet. 12:733746. doi: 10.3389/fgene.2021.733746

Background: The metabolic syndrome (MetS) is prevalent in Arabian populations. Several small-scale studies have been performed to investigate the genetic basis of MetS. This systematic review and meta-analysis aimed to examine whether candidate gene polymorphisms are associated with MetS susceptibility among ethnic groups of the Arabian world and to suggest possible directions for future research regarding genetic markers and MetS.

Methods: A search was conducted for peer-reviewed articles that examined the genetic association of MetS in Arabian populations in the following databases: Medline, Embase, Scopus, Direct Science, Web of Science, ProQuest, and Google Scholar until March 31, 2021. Articles were eligible if they were case-control studies, which investigated MetS as a dichotomous outcome (MetS vs no MetS). To assess the quality of the studies, the Q-Genie tool (Quality of Genetic Association Studies) was used. A non-central chi2 (random-effect) distribution was used to determine the heterogeneity (H) of Q and *I* (Galassi et al., The American journal of medicine, 2006, 119, 812–819) statistics.

Results: Our search strategy identified 36 studies that met our inclusion criteria. In most cases, studies were excluded due to a lack of statistical information such as odds ratios, confidence intervals, and *p*-values. According to the Q-Genie tool, 12 studies scored poorly (a score of≤35), 13 studies scored moderately (>35 and≤45), and 12 studies had good quality (>45 or higher). The most frequently studied genes were FTO and VDR (both included in four studies). Three SNPs indicated increased risk for MetS after calculating the pooled odds ratios: FTO-rs9939609 (odds ratio 1.49, 95% CI: 0.96–2.32); LEP-rs7799039 (odds ratio 1.85, 95% CI: 1.37–2.5); and SERPINA12-rs2236242 (odds ratio 1.65, 95% CI: 1.21–2.24). Meta-analysis studies showed no significant heterogeneity.

Conclusion: There were many sources of heterogeneity in the study settings. Most of the studies had low to moderate quality because of sample size and power issues, not

considering all potential sources of bias, and not providing details about genotyping methods and results. As most studies were small-scale, aimed to replicate findings from other populations, we did not find any unique genetic association between MetS and Arabian populations.

Keywords: metabolic syndrome, arab populations, genetic studies, systematic review, meta-analysis running title

INTRODUCTION

The metabolic syndrome (MetS), which is characterized by insulin resistance, hypertension, obesity, and dyslipidemia (Alberti et al., 2006), is a significant risk factor for atherosclerotic cardiovascular disease (ASCVD) and all-cause mortality (Alberti et al., 2006; Galassi et al., 2006). The prevalence of MetS has been reported as 27.3% in the Arabian Gulf Region (Shin and Jee, 2020). A number of factors can be attributed to this high prevalence, including lifestyle changes and rapid urbanization of the region in the last 40 years (Ramadan, 2015). In one study, Bayoumi and colleagues analyzed data from five large, extended, highly consanguineous Arabian families in Oman to determine the heritability (h2) of MetS and its different components (Bayoumi et al., 2007). The study examined 1,277 individuals in total. They reported that the prevalence of MetS was 23%, and the h2 value was 0.38. The figures are likely to be higher in larger samples due to the highly conserved gene pool in Arabian populations, the high rate of consanguinity in this region as well as the tribal structure of the society and the size of the families in this area (Zayed, 2016).

Several genetic loci are implicated in contributing to the different components of MetS (Barroso and McCarthy, 2019). However, not many genetic studies have focused on MetS as a binary phenotype. For example, searching the GWAS Catalogue (https://www.ebi.ac.uk/gwas/search?query=metabolic%20syndrome) revealed 18 publications on MetS but 193 publications when the search term diabetes was used. Using data from 291,107 individuals from the United Kingdom Biobank, Lind and his colleagues identified 93 independent loci associated with MetS, of which 80 were previously unidentified (Bayoumy et al., 2012). Furthermore, the study demonstrated that MetS is primarily determined by immunity and inflammation pathways and pathways that regulate lipids and lipoproteins. The purpose of this study was to review the current published data on MetS genetics in the Arabian population and suggest future directions.

METHODS

The flow chart in **Figure 1** illustrates the general set-up of the literature search, eligibility check, and extraction of reports, and inclusion of them into the final report.

Literature search: In reviewing the published data, these databases were searched: Medline (PubMed or OvidSP), Embase, Scopus, Direct Science, Web of Science, ProQuest, and Google Scholar, from database inception to March 31, 2021, using these search terms and filters:

- "Arab" OR "Arab ancestry" of "Arab ethnicity" OR one of the 22 Arab countries: "Algeria" OR "Bahrain" OR "Comoros" OR "Djibouti" OR "Egypt" OR "United Arab Emirate" OR "Iraq" OR "Jordan" OR "Kuwait" OR "Lebanon" OR "Libya" OR "Mauritania" OR "Morocco" OR "Oman" OR "Palestine" OR "Qatar" OR "Saudi" OR "Saudi Arabia" OR "Somalia" OR "Sudan" OR "Syria" OR "Tunisia" OR "Yemen. "
- AND "Metabolic syndrome" OR "MetS" OR "Syndrome X" OR "Dysmetabolic syndrome" OR "Insulin resistance syndrome. "
- AND "Genetic" OR "Genetics" OR "SNP" OR "Single nucleotide polymorphism" OR "Genetic polymorphism" OR "Genetic variant" OR "Genetic variation" OR "Genetic polymorphism" OR "Allele" OR "Genetic locus. "

A sample search strategy for Medline (via OvidSP) can be found in the **Supplementary Figure S1**.

Eligibility: The eligibility criteria comprised peer-reviewed articles on genetic association with MetS in Arabian populations (candidate-gene studies, genome-wide approaches, or sequencing

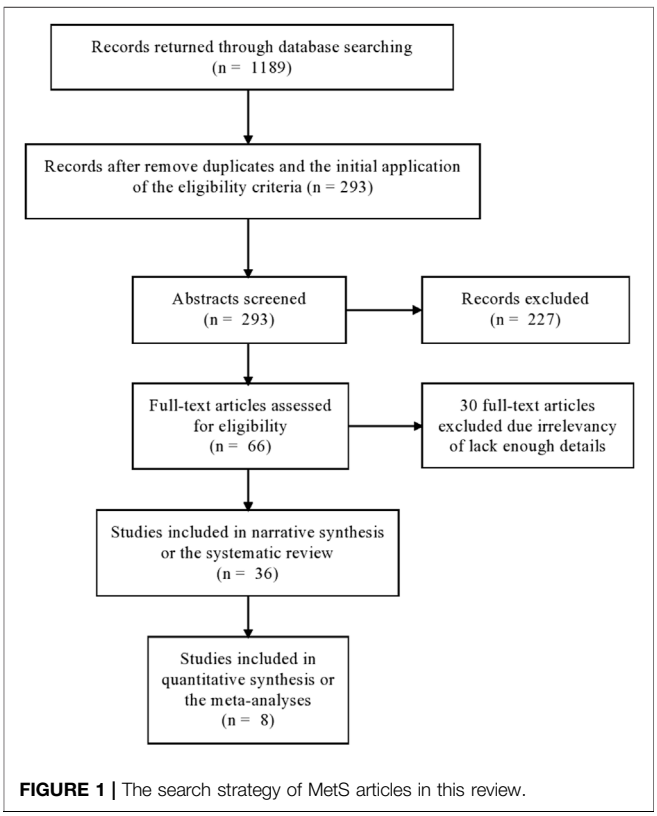

FIGURE 1 | The search strategy of MetS articles in this review.

approaches). In this context, we defined the Arabian population as those who are associated with the Arab League, as shown in the following table. Articles eligible for consideration included case-control studies examining MetS as a dichotomous outcome (MetS versus no MetS), and published in English. The exclusion criteria were studies that did not meet the inclusion criteria, grey literature, duplicate publications, and studies without detailed results, including odds ratios, confidence intervals, and *p* values. We also excluded case report review papers and family association studies. To reduce the risk of selection and methodological bias, studies that reported one of the standard definitions of MetS (IDF, AHA/NHLBI, or NCEP - ATP III) were included in the meta-analysis (**Supplementary Table S1**). Two authors identified and read an abstract to assess the article's eligibility and then confirmed by the primary investigator for meeting the criteria for inclusion and exclusion. When a paper did not contain an abstract or provide sufficient information, the full text or supplementary materials were reviewed for suitability. Abstracts were evaluated according to their eligibility as 'yes,' 'no' or 'unclear.' The unclear abstracts were kept until the full-text articles were reviewed before a decision was made regarding inclusion.

Data extraction: The full-text data was extracted by three authors (ZA, NA, and MM) after screening the titles and abstracts to determine their eligibility. Any disagreements were resolved by a fourth researcher (WO). Further, detailed data were extracted from standardized data extraction forms, including author affiliations, study designs and settings, study population characteristics, MetS classification criteria, number of cases and controls, obesity determination and measurements, genetic markers, tested alleles, and association analysis methods and outcomes.

Quality of studies: To assess the quality of the studies, the Q-Genie (Quality of Genetic Association Studies) tool was applied, which has a scoring system covering eleven different elements, as previously reported (Bayoumy et al., 2012). Studies were classified as poor quality (a score of ≤35), moderate: scores of >35 and ≤45; and good: scores of >45.

Statistical analysis: Statistical analysis was performed using R statistical software Version 4.0.3 and Stata software Version 14 (StataCorp, TX, United States). In Stata software, we used the "metan" command. The heterogeneity (H) was determined using the non-central chi (Galassi et al., 2006) (common-effect) distribution for Q and *I* (Galassi et al., 2006) statistics. According to initial analyses, most of the studies had significant among-study heterogeneity (*I* (Galassi et al., 2006) > 50%). Therefore, we conducted the meta-analysis by pooling aggregate data using the random-effects inverse-variance model, which included the DerSimonian-Laird model estimator for tau squared (between-study variance, also known as τ (Galassi et al., 2006)) in addition to *I* (Galassi et al., 2006). Furthermore, we generated Forest plots to illustrate the pooled prevalence estimates derived from the meta-analysis calculations. $p < 0.05$ was used as the threshold for statistical significance. The eyeball test (Funnel plot) and Egger's test were not considered because the number of studies used in the meta-analysis was small (<10 studies), and the test had a relatively low statistical power to determine whether the observed asymmetry was caused by chance.

RESULTS

Literature Search and Study Characteristics

Our search strategy (**Figure 1** and Methods section) identified 36 studies that met our inclusion criteria (Bayoumi et al., 2007; Bayoumy et al., 2012; Boumaiza et al., 2012a; Boumaiza et al., 2012b; Ghattas et al., 2013; Mohamed Youssef et al., 2013; Zadjali et al., 2013; Ajjemami et al., 2014; Al-Daghri et al., 2014; Mackawy and Badawi, 2014; Youssef et al., 2014; Ajjemami et al., 2015; Mehanna et al., 2015; Alnory et al., 2016; Mehanna et al., 2016; Elouej et al., 2016a; RaafatAbdRaboh et al., 2016; Elouej et al., 2016b; Elouej et al., 2016c; Khella et al., 2017; Hasan et al., 2017; Lakbakbi El Yaagoubi et al., 2017; Mackawy, 2017; Morjane et al., 2017; Ismail et al., 2019; Ibrahim and Bastawy, 2020; Salami and El Shamieh, 2019; ElGhareeb et al., 2019; Nizam et al., 2019; Boulenouar et al., 2019; Aljaibeji et al., 2020; Osman et al., 2020; Alenad et al., 2020; Hechmi et al., 2020; Kefi, 2021). In most cases, studies were excluded because they lacked detailed results, such as odds ratios, confidence intervals, and *p*-values. The study conducted by Bayoumi et al. did not study genetic markers. However, we included it because it reported the heritability of MetS and its components using data from Oman, which was found to be 0.38 [5]. Among the 36 association studies, the majority came from Egypt (13 studies) and Tunisia (10 studies) (**Table 1**). The majority of studies adopted the IDF definition of MetS over other definitions like the AHA/NHLBI (Grundy et al., 2005) and the NCEP — ATP III (Alexander et al., 2003) (**Supplementary Table S1**).

Genetics of Obesity in Arab Populations

All studies were gene-candidate studies (**Table 1**) that selected genes based on previous literature reports or their predicted functions in relation to MetS development. There was considerable variability in the number of genes selected, ranging from one to six. The most commonly studied genes were *FTO* and *VDR* (both four studies, **Supplementary Table S2**), whereas the most frequently tested markers were rs10735810 (FokI) and rs1544410 (BsmI); both in the VDR gene (**Supplementary Table S3**). Four genetic markers found eligible for the meta-analysis studies: rs9939609 (*FTO*), rs4420638 (*APOE*), rs7799039 (*LEP*), and rs2236242 (*SERPINA12*). As all of these indicated a significant heterogeneity [*I* (Galassi et al., 2006)> 50%] (**Figure 2**), we used the random-effect inverse-variance model for the meta-analysis studies (see the Methods section). Three of these markers indicated increased risk for MetS after calculations of the pooled odds ratios: *FTO*-rs9939609 (OR = 1.49, 95% CI: 0.96–2.32); *LEP*-rs7799039 (OR = 1.85, 95% CI: 1.37–2.5); and *SERPINA12*-rs2236242 (OR = 1.65, 95% CI: 1.21–2.24), **Figure 2**. These three genes are mainly implicated in obesity or lipid metabolism (**Supplementary Tables S4,5**).

Quality of the Studies

In order to assess the quality of the studies, the Q-Genie tool was utilized (Sohani et al., 2016). It is a validated tool of bias and quality for systematic reviews of genetic association studies. The tool employs 11 criteria, covering all association steps, including rationale, sampling, selection of cases and controls, methodology, analysis, and conclusions. The tool indicated that 12 studies had

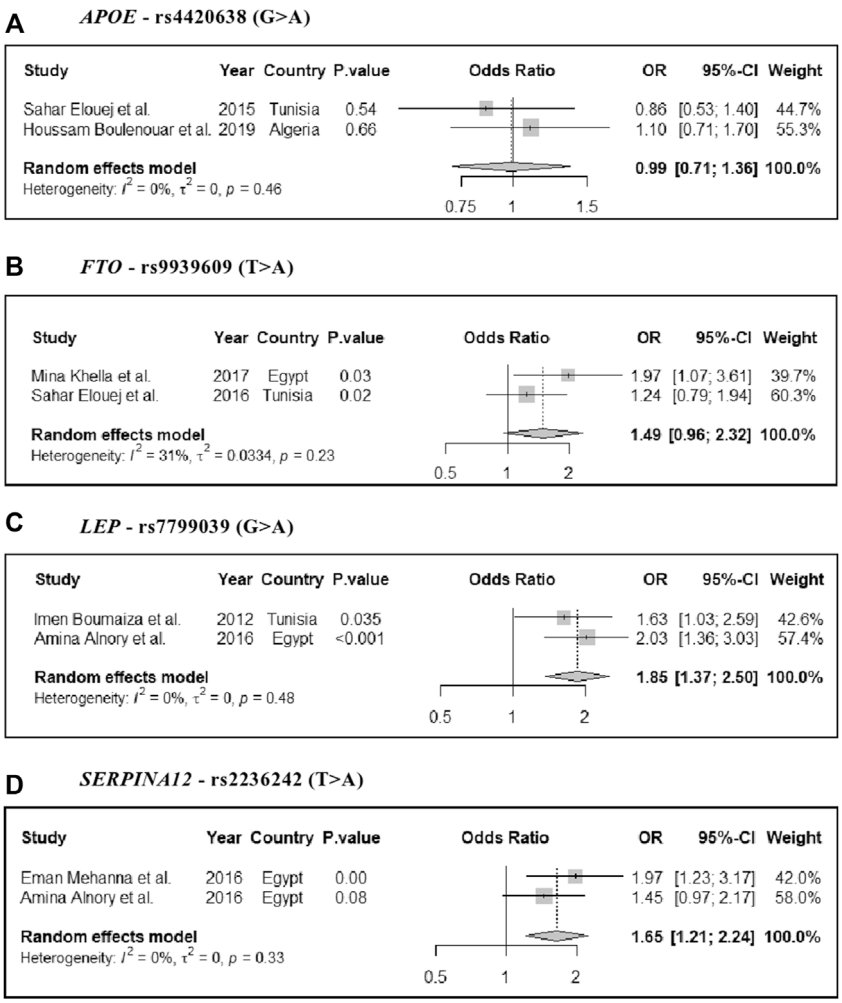

FIGURE 2 | Forest plot of the pooled odds ratios and 95% CI estimates of MetS for four genetic markers: **(A)** *APOE* - rs4420638 (G > A), **(B)** *FTO* - rs9939609 (T > A), **(C)** *LEP* - rs7799039 (G > A), and **(D)** *SERPINA12* - rs2236242 (T > A).

poor quality (a score of ≤35), 13 studies had moderate quality (scores >35 and ≤45); and only eight studies had good quality (scores >45). The detailed quality information for all studies is summarized in **Supplementary Table S6**. Specifically, three factors affected the study's quality: the non-technical classification of the genetic variants, the disclosure and discussion of biases, and the size and power calculations. (**Supplementary Table S6**).

DISCUSSION

Main Findings

Genetic studies should be conducted across populations due to differences in linkage disequilibrium (LD)/haplotype blocks and the type and frequency of monomorphic/polymorphic SNPs. Many studies have investigated the genetics of MetS, but few have examined MetS as a binary outcome. In light of the high prevalence of MetS in Arab populations and its components (Shin and Jee, 2020), the current study reviewed the studies that examined MetS in Arabic populations as a binary trait.

In the narrative synthesis, 36 studies were included after several were excluded due to insufficient methodology or statistical analysis information. All of the studies were candidate gene studies, and none of the studies utilized a genome-wide or sequencing approach. Moreover, there was considerable variation in the tested markers and genes, with most studies evaluating a single gene (Elouej et al., 2016b; Khella et al., 2017; Mackawy, 2017) to several genes in others (Elouej et al., 2016c; Osman et al., 2020). The study conducted by Bayoumi et al. reported the heritability of MetS using data from Oman, which was found to be 0.38. Based on their cultural similarity, we expected other Arab populations to have a similar prevalence of MetS.

Heterogeneity

There were several reasons for heterogeneity between these studies regarding MetS definitions, study populations, sample

sizes, research design, and selection of the genetic markers (**Table 1**). For example, while most of the studies focused on diverse populations, a few included only females in their research (Mehanna et al., 2015; Morjane et al., 2017), and one study examined MetS in children (Nizam et al., 2019). Similarly, the proportion of males and females, and the age range, varied in most studies; both of these factors are well-known sources of heterogeneity in meta-analyses (Pei et al., 2016). It is also important to note that the precise definition of the Arab population is challenging since there are different ethnic subgroups within the Arab world. For example, admixture studies revealed that the Qatari population is made up of three major sub-populations derived from Arabic, Persian, and African ancestries (Omberg et al., 2012). It indicates that several factors related to ethnic backgrounds, such as the LD patterns and allele frequencies, have not been considered by most reports. Moreover, heterogeneity in the study was assessed using the heterogeneity statistic I (Galassi et al., 2006), which estimates the proportion of the variance in a study due to heterogeneity. Nonetheless, I (Galassi et al., 2006) values can be imprecise and possess a substantial bias when the number of studies in a meta-analysis study is small, as in our study (von Hippel, 2015).

The Reported Genetic Markers

Most of the genetic markers were selected based on their functional predictability related to diabetes, insulin resistance, obesity, or previous reports of their association with metabolic syndrome. The functions and key information for these genes (*FTO*, *APOE*, *LEP*, and *SERPINA12*) are shown in **Supplementary Table S4**. The STRING database (Szklarczyk et al., 2019) has demonstrated that these genes are involved in several biological processes associated with the development of MetS, including steroid metabolism, lipid metabolism, apoptosis, glucose metabolism, cell signaling, and regulation of inflammatory responses response (**Supplementary Table S5**). Among the studies included in the meta-analysis study, three markers were found to be significant in the pooled associations with MetS: rs9939609 (odds ratio 1.49); rs7799039 (odds ratio 1.85); and rs2236242 (odds ratio 1.65). Despite several sources of heterogeneity, we observed no significant differences between the pooled and meta-analyzed studies (**Figure 2**). *FTO* gene encodes alpha-ketoglutarate-dependent dioxygenase, which regulates food intake and energy expenditure (Kalantari et al., 2016). Genetic variants in the *FTO* gene have been associated with many traits and diseases, such as obesity, lipid metabolism, type 2 diabetes, and MetS (Kucher, 2020). *APEO* encodes the glycoprotein apolipoprotein E, which is important in the metabolism of lipoproteins and lipids (Tudorache et al., 2017). In addition, *LEP* encodes leptin, a key hormone responsible for maintaining energy balance and regulating body weight (Pan and Myers, 2018). This indicates that the genes were mainly selected to be studied in association with MetS based upon their roles in obesity or lipid metabolism, not their roles in inflammatory pathways, which have been shown to be key contributors to MetS (Lind, 2019).

However, these associations with MetS are not limited to Arab populations (Teshigawara et al., 2012; Wang et al., 2012; Ghalandari et al., 2015; Albarracín and Torres, 2020; Hardy et al., 2020). For isntance, H Wang and colleagues (Wang et al., 2012) published a meta-analysis of 18 studies regarding the association of genetic variations in the FTO gene with MetS and reported significant pooled associations.

Quality of the Studies

Considering that the number of studies was small when divided by different genes and genetic variations, we could not determine the degree of publication bias. Consequently, it was challenging to use methods such as GRADE to evaluate the quality of studies. The Q-Genie tool was utilized as a quality assessment tool because genetic studies are unique compared to epidemiological studies, especially regarding bias sources, such as variations related to the collection, handling, and processing of DNA, genotype, and relatedness classification in the population under study (Boumaiza et al., 2012a). Twenty-five studies were rated as having a moderate to poor quality by the Q-Genie tool (67.6%) (**Supplementary Table S6**). The primary causes of low-quality studies were the sample size, insufficient power calculations, and not taking into account significant sources of bias when designing the study.

Furthermore, several studies have indicated insufficient information about the non-technical classification of genetic variations and how genotyping was performed. Additionally, many studies did not disclose other potential sources of bias (e.g., time-lag bias, attrition bias, recall bias) and how they were handled appropriately.

Limitations, Strengths, and Recommendations to Improve Future Studies

1) The protocol of this review has not been pre-registered for this (e.g., in PROSPERO), which may introduce potential bias per Cochrane guidelines.
2) There have been no comprehensive studies or genome-wide studies conducted on the association of MetS in any Arab population. Additionally, several studies have been conducted to examine some individual components of MetS rather than MetS as a whole. We recommend establishing a regional consortium that can address this problem and unify the definition of MetS, the study settings, and the analysis plans.
3) There was an underrepresentation of studies from the low- and middle-income regions of the Arab world, especially those that are subject to war and political instability (Iraq, Syria, Libya, Sudan, Yemen, and Palestine).
4) There were several studies with poor design or analysis of the results. In addition, several studies were excluded due to a lack of results detail, such as a lack of odds ratios or confidence intervals. Therefore, we recommend that MetS researchers in the region carefully design their studies, consider potential bias sources, and report results in detail.
5) In this systematic review, a significant concern was using different names for the same marker (e.g., rs1801282 in *PPARγ2* gene or rs2228570 in *VDR* gene), making it

difficult to conclude these studies. Therefore, we recommend using common gene or SNP markers nomenclature systems, such as the HUGO Gene Nomenclature for genes and the bSNP Reference SNP (rs or RefSNP) number for genetic markers.

6) Lastly, as reported in Beshyah and colleagues in 2018 (Beshyah et al., 2018), we observed that a significant number of papers are being published in predatory journals. A recent report revealed that the Middle East or Gulf Cooperation Council region is a significant target for predatory journals (Macháček and Srholec, 2021). Researchers are advised to increase their knowledge of predatory journals and become familiar with Beall's list of these journals and other resources available to them.

We summarized the current literature on the genetic associations of MetS in Arab populations in this review. We found that three markers, FTO-rs9939609, LEP-rs7799039, and SERPINA12-rs2236242, contributed significantly to the pooled risk for MetS development. As a result of considerable variations in study settings and numerous sources of heterogeneity, most of these studies are rated as of low to moderate quality.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

WO be designed the study protocol with the support of HA and GT; ZA-H, NA, and MM separately performed the search and summarized the search data, WO, ZA-H, NA, and MM wrote the manuscript, HFJ and MM assisted in the statistical analysis and summarizing the data, HFJ and NA-S helped in the data analysis and figure preparations and critically revised the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.733746/full#supplementary-material

REFERENCES

Ajjemami, M., Ouatou, S., Charoute, H., Fakiri, M., Rhaissi, H., Benrahma, H., et al. (2015). Haplotype Analysis of the Apolipoprotein A5 Gene in Moroccan Patients with the Metabolic Syndrome. *J. Diabetes Metab. Disord.* 14 (1), 1–8. doi:10.1186/s40200-015-0160-3

Ajjemami, M., Rhaissi, H, Houda, B, and Hicham, C (2014). Association between Polymorphism in APOC3, and Metabolic Syndrome in the Moroccan Population. *Int. J. Scientific Res. Publications* 12, 270.

Al-Daghri, N. M., Al-Attas, O. S., Alkharfy, K. M., Khan, N., Mohammed, A. K., Vinodson, B., et al. (2014). Association of VDR-Gene Variants with Factors Related to the Metabolic Syndrome, Type 2 Diabetes and Vitamin D Deficiency. *Gene* 542, 129–133. doi:10.1016/j.gene.2014.03.044

Albarracín, M. L. G., and Torres, A. Y. F. (2020). Adiponectin and Leptin Adipocytokines in Metabolic Syndrome: What Is its Importance. *Dubai Diabetes Endocrinol. J.*, 1–10.

Alberti, G., Zimmet, P., Shaw, J., and Grundy, S. M. (2006).The IDF Consensus Worldwide Definition of the Metabolic Syndrome. *Brussels: International Diabetes Federation* 23 (5), 469–80. doi:10.1111/j.1464-5491.2006.01858.x

Alenad, A., Alenezi, M. M., Alokail, M. S., Wani, K., Mohammed, A. K., Alnaami, A. M., et al. (2020). Association of ANGPTL8 (Betatrophin) Gene Variants with Components of Metabolic Syndrome in Arab Adults. *Sci. Rep.* 10, 6764–6810. doi:10.1038/s41598-020-63850-7

Alexander, C. M., Landsman, P. B., Teutsch, S. M., and Haffner, S. M. (2003). NCEP-defined Metabolic Syndrome, Diabetes, and Prevalence of Coronary Heart Disease Among NHANES III Participants Age 50 Years and Older. *Diabetes* 52, 1210–1214. doi:10.2337/diabetes.52.5.1210

Aljaibeji, H., Mohammed, A. K., Alkayyali, S., Hachim, M. Y., Hasswan, H., El-Huneidi, W., et al. (2020). Genetic Variants of the PLCXD3 Gene Are Associated with Risk of Metabolic Syndrome in the Emirati Population. *Genes* 11, 665. doi:10.3390/genes11060665

Alnory, A., Gad, H., Hegazy, G., and Shaker, O. (2016). The Association of Vaspin Rs2236242 and Leptin Rs7799039 Polymorphism with Metabolic Syndrome in Egyptian Women. *Turk J. Med. Sci.* 46, 1335–1340. doi:10.3906/sag-1502-138

Barroso, I., and McCarthy, M. I. (2019). The Genetic Basis of Metabolic Disease. *Cell* 177, 146–161. doi:10.1016/j.cell.2019.02.024

Bayoumi, R. A., Al-Yahyaee, S. A. S., Albarwani, S. A., Rizvi, S. G., Al-Hadabi, S., Al-Ubaidi, F. F., et al. (2007). Heritability of Determinants of the Metabolic Syndrome Among Healthy Arabs of the Oman Family Study*. *Obesity* 15, 551–556. doi:10.1038/oby.2007.555

Bayoumy, N. M., El-Shabrawi, M. M., and Hassan, H. H. (2012). Association of Cluster of Differentiation 36 Gene Variant Rs1761667 (G>A) with Metabolic Syndrome in Egyptian Adults. *Saudi Med. J.* 33, 489–494.

Beshyah, S., Hajjaji, I., and Elbarsha, A. (2018). Awareness of Predatory Journals Among Physicians from Africa and the Middle East: An Exploratory Survey. *Ibnosina J. Med. Biomed. Sci.* 10, 136. doi:10.4103/ijmbs.ijmbs_45_18

Boulenouar, H., Sounnia, M B, Hadjira, O D, Sarah, A L H, Leila, H, and Imane, H M (2019). Association Study of APOE Gene Polymorphisms with Diabetes and the Main Cardiometabolic Risk Factors, in the Algerian Population. *Egypt. J. Med. Hum. Genet.* 20, 1–8. doi:10.1186/s43042-019-0013-6

Boumaiza, I., Omezzine, A., Rejeb, J., Rebhi, L., Ben Rejeb, N., Nabli, N., et al. (2012). Association between FourResistinPolymorphisms, Obesity, and Metabolic Syndrome Parameters in Tunisian Volunteers. *Genet. Test. Mol. biomarkers* 16, 1356–1362. doi:10.1089/gtmb.2012.0156

Boumaiza, I., Omezzine, A., Rejeb, J., Rebhi, L., Ouedrani, A., Ben Rejeb, N., et al. (2012). Relationship between Leptin G2548A and Leptin Receptor Q223R Gene Polymorphisms and Obesity and Metabolic Syndrome Risk in Tunisian Volunteers. *Genet. Test. Mol. biomarkers* 16, 726–733. doi:10.1089/gtmb.2011.0324

ElGhareeb, M. I., Khater, M. H., Fakhr, A., and Khedr, H. A.-E. (2019). Risk and Severity of Psoriasis Vulgaris in Relation to Angiotensin II Type 1 Receptor Gene Polymorphism and Metabolic Syndrome. *Ccid* 12, 683–690. doi:10.2147/ccid.s212781

Elouej, S., Belfki-Benali, H., Nagara, M., Lasram, K., Attaoua, R., Sallem, O. K., et al. (2016). Association of Rs9939609 Polymorphism with Metabolic Parameters and FTO Risk Haplotype Among Tunisian Metabolic Syndrome. *Metab. Syndr. Relat. Disord.* 14, 121–128. doi:10.1089/met.2015.0090

Elouej, S., Nagara, M., Attaoua, R., Sallem, O. K., Rejeb, I., Hsouna, S., et al. (2016). Association of Genetic Variants in the FTO Gene with Metabolic Syndrome: A Case-Control Study in the Tunisian Population. *J. Diabetes its Complications* 30, 206–211. doi:10.1016/j.jdiacomp.2015.11.013

Elouej, S., Rejeb, I., Attaoua, R., Nagara, M., Sallem, O. K., Kamoun, I., et al. (2016). Gender-specific Associations of Genetic Variants with Metabolic Syndrome Components in the Tunisian Population. *Endocr. Res.* 41, 300–309. doi:10.3109/07435800.2016.1141945

Galassi, A., Reynolds, K., and He, J. (2006). Metabolic Syndrome and Risk of Cardiovascular Disease: a Meta-Analysis. *Am. J. Med.* 119, 812–819. doi:10.1016/j.amjmed.2006.02.031

Ghalandari, H., Hosseini-Esfahani, F., and Mirmiran, P. (2015). The Association of Polymorphisms in Leptin/leptin Receptor Genes and Ghrelin/ghrelin Receptor Genes with Overweight/obesity and the Related Metabolic Disturbances: a Review. *Int. J. Endocrinol. Metab.* 13, e19073. doi:10.5812/ijem.19073v2

Ghattas, M. H., Mehanna, E. T., Mesbah, N. M., and Abo-Elmatty, D. M. (2013). Association of Estrogen Receptor Alpha Gene Polymorphisms with Metabolic Syndrome in Egyptian Women. *Metabolism* 62, 1437–1442. doi:10.1016/j.metabol.2013.05.014

Grundy, S. M., Cleeman, J. I., Daniels, S. R., Donato, K. A., Eckel, R. H., Franklin, B. A., et al. (2005). Diagnosis and Management of the Metabolic Syndrome. *Circulation* 112, 2735–2752. doi:10.1161/circulationaha.105.169404

Hardy, D. S., Garvin, J. T., Mersha, T. B., and Racette, S. B. (2020). Ancestry Specific Associations of FTO Gene Variant and Metabolic Syndrome: A Longitudinal ARIC Study. *Medicine (Baltimore)* 99, e18820. doi:10.1097/MD.0000000000018820

Hasan, H. A., AbuOdeh, R. e. O., Muda, W. A. M. B. W., Mohamed, H. J. B. J., and Samsudin, A. R. (2017). Association of Vitamin D Receptor Gene Polymorphisms with Metabolic Syndrome and its Components Among Adult Arabs from the United Arab Emirates. *Diabetes Metab. Syndr. Clin. Res. Rev.* 11, S531–S537. doi:10.1016/j.dsx.2017.03.047

Hechmi, M., Dallali, H., Gharbi, M., Jmel, H., Fassatoui, M., Ben Halima, Y., et al. (2020). Association of Rs662799 Variant and APOA5 Gene Haplotypes with Metabolic Syndrome and its Components: a Meta-Analysis in North Africa. *Biosci. Rep.* 40. doi:10.1042/BSR20200706

Ibrahim, S. M., and Bastawy, A. A. (2020). The Relevance of Single-Nucleotide Polymorphism +62 G>A to the Expression of Resistin Gene Affecting Serum Resistin Levels in Metabolic Syndrome in the Egyptian Population. *Cpb* 21, 626–634. doi:10.2174/1389201021666191210122851

Ismail, N. A., Ragab, S., Abd, S M., and Dayem, E. (2019). Variant of TCF7L2 Gene and Obesity in Egyptian Children. *Research Journal Pharmaceutical Biological Chemical Sciences* 9, 1135–1142.

Kalantari, N., Doaei, S., Keshavarz-Mohammadi, N., Gholamalizadeh, M., and Pazan, N. (2016). Review of Studies on the Fat Mass and Obesity-Associated (FTO) Gene Interactions with Environmental Factors Affecting on Obesity and its Impact on Lifestyle Interventions. *ARYA Atheroscler.* 12, 281–290.

Kefi, R., Hechmi, M., Dallali, H., Elouej, S., Jmel, H., and Halima, Y. B. (2017). Association of Apolipoprotein A5 Gene Variants with Metabolic Syndrome in Tunisian Population. In *Annales D'endocrinologie* (Elsevier Masson) 78 (3), 146–155. doi:10.1016/j.ando.2017.01.005

Khella, M. S., Hamdy, N. M., Amin, A. I., and El-Mesallamy, H. O. (2017). The (FTO) Gene Polymorphism Is Associated with Metabolic Syndrome Risk in Egyptian Females: a Case- Control Study. *BMC Med. Genet.* 18, 101–108. doi:10.1186/s12881-017-0461-0

Kucher, A. N. (2020). The FTO Gene and Diseases: The Role of Genetic Polymorphism, Epigenetic Modifications, and Environmental Factors. *Russ. J. Genet.* 56, 1025–1043. doi:10.1134/s1022795420090136

Lakbakbi El Yaagoubi, F., Charoute, H., Morjane, I., Sefri, H., Rouba, H., Ainahi, A., et al. (2017). Association Analysis of Genetic Variants with Metabolic Syndrome Components in the Moroccan Population. *Curr. Res. translational Med.* 65, 121–125. doi:10.1016/j.retram.2017.08.001

Lind, L. (2019). Genome-wide Association Study of the Metabolic Syndrome in UK Biobank. *Metab. Syndr. Relat. Disord.* 17, 505–511. doi:10.1089/met.2019.0070

Macháček, V., and Srholec, M. (2021). Predatory Publishing in Scopus: Evidence on Cross-Country Differences. *Scientometrics* 126, 1897–1921.

Mackawy, A. M. H., and Badawi, M. E. H. (2014). Association of Vitamin D and Vitamin D Receptor Gene Polymorphisms with Chronic Inflammation, Insulin Resistance and Metabolic Syndrome Components in Type 2 Diabetic Egyptian Patients. *Meta gene* 2, 540–556. doi:10.1016/j.mgene.2014.07.002

Mackawy, A. M. (2017). Protein Tyrosine Phosphatase 1B (PTPN1) Gene Polymorphism (467T> C) and Metabolic Syndrome: A Pilot Study. *Int. J. Med. Res. Health Sci.* 6, 1–9.

Mehanna, E. T., Abo-Elmatty, D. M., Ghattas, M. H., Mesbah, N. M., and Saleh, S. M. (2015). Apelin Rs2235306 Polymorphism Is Not Related to Metabolic Syndrome in Egyptian Women. *Egypt. J. Med. Hum. Genet.* 16, 35–40. doi:10.1016/j.ejmhg.2014.11.004

Mehanna, E. T., Mesbah, N. M., Ghattas, M. H., Saleh, S. M., and Abo-Elmatty, D. M. (2016). Association of Chemerin Rs17173608 and Vaspin Rs2236242 Gene Polymorphisms with Metabolic Syndrome in Egyptian Women. *Endocr. Res.* 41, 43–48. doi:10.3109/07435800.2015.1066802

Mohamed Youssef, S., Mohamed, N., Afef, S., Khaldoun, B. H., Fadoua, N., Fadhel, N. M., et al. (2013). Interaction Effects of the Leu162Val PPAR α and Pro12Ala PPAR γ 2 Gene Variants with Renal Function in Metabolic Syndrome Population. *PPAR Res.* 2013, 329862. doi:10.1155/2013/329862

Morjane, I., Kefi, R., Charoute, H., Lakbakbi el Yaagoubi, F., Hechmi, M., Saile, R., et al. (2017). Association Study of HNF1A Polymorphisms with Metabolic Syndrome in the Moroccan Population. *Diabetes Metab. Syndr. Clin. Res. Rev.* 11, S853–S857. doi:10.1016/j.dsx.2017.07.005

Nizam, R., Al-Ozairi, E., Goodson, J. M., Melhem, M., Davidsson, L., Alkhandari, H., et al. (2019). Corrigendum: Caveolin-1 Variant Is Associated with the Metabolic Syndrome in Kuwaiti Children. *Front. Genetfrontiers Genetics* 10, 9689221. doi:10.3389/fgene.2019.00221

Omberg, L., Salit, J., Hackett, N., Fuller, J., Matthew, R., Chouchane, L., et al. (2012). Inferring Genome-wide Patterns of Admixture in Qataris Using Fifty-Five Ancestral Populations. *BMC Genet.* 13, 49–10. doi:10.1186/1471-2156-13-49

Osman, W. M., Khan, S. M., Jelinek, H. F., Almahmeed, W., Tay, G. K., and Alsafar, H. S. (2020). Clinical Correlations and Genetic Associations of Metabolic Syndrome in the United Arab Emirates. *Gene* 738, 144476. doi:10.1016/j.gene.2020.144476

Pan, W. W., and Myers, M. G., Jr (2018). Leptin and the Maintenance of Elevated Body Weight. *Nat. Rev. Neurosci.* 19, 95–105. doi:10.1038/nrn.2017.168

Pei, Y.-F., Tian, Q., Zhang, L., and Deng, H.-W. (2016). Exploring the Major Sources and Extent of Heterogeneity in a Genome-wide Association Meta-Analysis. *Ann. Hum. Genet.* 80, 113–122. doi:10.1111/ahg.12143

Raafat AbdRaboh, N., Louka, M. L., and Sabry, I. M. (2016). Leptin Gene Microsatellite Polymorphism: Relation to Metabolic Syndrome. *Gene Rep.* 4, 87–90. doi:10.1016/j.genrep.2016.04.002

Ramadan, E. (2015). Sustainable Urbanization in the Arabian Gulf Region: Problems and Challenges. *Arts Soc. Sci. J.* 6, 1–4.

Salami, A., and El Shamieh, S. (2019). Association between SNPs of Circulating Vascular Endothelial Growth Factor Levels, Hypercholesterolemia and Metabolic Syndrome. *Medicina* 55, 464. doi:10.3390/medicina55080464

Shin, S., and Jee, H. (2020). Prevalence of Metabolic Syndrome in the Gulf Cooperation Council Countries: Meta-Analysis of Cross-Sectional Studies. *J. Exerc. Rehabil.* 16, 27–35. doi:10.12965/jer.1938758.379

Sohani, Z. N., Sarma, S., Alyass, A., de Souza, R. J., Robiou-du-Pont, S., Li, A., et al. (2016). Empirical Evaluation of the Q-Genie Tool: a Protocol for Assessment of Effectiveness. *BMJ open* 6, e010403. doi:10.1136/bmjopen-2015-010403

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-wide Experimental Datasets. *Nucleic Acids Res.* 47, D607–D613. doi:10.1093/nar/gky1131

Teshigawara, S., Wada, J., Hida, K., Nakatsuka, A., Eguchi, J., Murakami, K., et al. (2012). Serum Vaspin Concentrations Are Closely Related to Insulin Resistance, and Rs77060950 atSERPINA12Genetically Defines Distinct Group with Higher Serum Levels in Japanese Population. *J. Clin. Endocrinol. Metab.* 97, E1202–E1207. doi:10.1210/jc.2011-3297

Tudorache, I. F., Trusca, V. G., and Gafencu, A. V. (2017). Apolipoprotein E - A Multifunctional Protein with Implications in Various Pathologies as a Result of its Structural Features. *Comput. Struct. Biotechnol. J.* 15, 359–365. doi:10.1016/j.csbj.2017.05.003

von Hippel, P. T. (2015). The Heterogeneity Statistic I(2) Can Be Biased in Small Meta-Analyses. *BMC Med. Res. Methodol.* 15, 35–38. doi:10.1186/s12874-015-0024-z

Wang, H., Dong, S., Xu, H., Qian, J., and Yang, J. (2012). Genetic Variants in FTO Associated with Metabolic Syndrome: a Meta- and Gene-Based Analysis. *Mol. Biol. Rep.* 39, 5691–5698. doi:10.1007/s11033-011-1377-y

Youssef, S. M., Mohamed, N., Afef, S., Khaldoun, B. H., Fadoua, N., Fadhel, N. M., et al. (2014). Combined Effects of the C161T and Pro12Ala PPARγ2 Gene Variants with Insulin Resistance on Metabolic Syndrome: A Case-Control Study of a Central Tunisian Population. *J. Mol. Neurosci.* 52, 487–492. doi:10.1007/s12031-013-0161-y

Zadjali, F., Al-Yahyaee, S., Hassan, M. O., Albarwani, S., and Bayoumi, R. A. (2013). Association of Adiponectin Promoter Variants with Traits and Clusters of Metabolic Syndrome in Arabs: Family-Based Study. *Gene* 527, 663–669. doi:10.1016/j.gene.2013.06.057

Zayed, H. (2016). The Arab Genome: Health and Wealth. *Gene* 592, 239–243. doi:10.1016/j.gene.2016.07.007

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, orclaim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 02 December 2021
doi: 10.3389/fgene.2021.749415

Genome-Wide Association and Mendelian Randomization Analysis Reveal the Causal Relationship Between White Blood Cell Subtypes and Asthma in Africans

*Opeyemi Soremekun[1], Chisom Soremekun[1,2], Tafadzwa Machipisa[3,4], Mahmoud Soliman[5], Oyekanmi Nashiru[2], Tinashe Chikowore[6,7] and Segun Fatumo[1,2,8]**

[1]The African Computational Genomics (TACG) Research Group, MRC/UVRI and LSHTM, Entebbe, Uganda, [2]H3Africa Bioinformatics Network (H3ABioNet) Node, Centre for Genomics Research and Innovation, NABDA/FMST, Abuja, Nigeria, [3]Department of Medicine, University of Cape Town, Groote Schuur Hospital, Cape Town, South Africa, [4]The Department of Pathology and Molecular Medicine, Population Health Research Institute (PHRI), Michael G. DeGroote School of Medicine, McMaster University, Hamilton, ON, Canada, [5]Molecular Bio-Computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban, South Africa, [6]Faculty of Health Sciences, Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa, [7]MRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [8]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom

OPEN ACCESS

Edited by:
Tommaso Pippucci, Policlinico Sant'Orsola-Malpighi, Italy

Reviewed by:
Diana Dunca, University College London, United Kingdom
Albert Henry, University College London, United Kingdom

***Correspondence:**
Segun Fatumo
segun.fatumo@lshtm.ac.uk

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *29 July 2021*
Accepted: *01 November 2021*
Published: *02 December 2021*

Citation:
Soremekun O, Soremekun C, Machipisa T, Soliman M, Nashiru O, Chikowore T and Fatumo S (2021) Genome-Wide Association and Mendelian Randomization Analysis Reveal the Causal Relationship Between White Blood Cell Subtypes and Asthma in Africans. Front. Genet. 12:749415. doi: 10.3389/fgene.2021.749415

Background: White blood cell (WBC) traits and their subtypes such as basophil count (Bas), eosinophil count (Eos), lymphocyte count (Lym), monocyte count (Mon), and neutrophil counts (Neu) are known to be associated with diseases such as stroke, peripheral arterial disease, and coronary heart disease.

Methods: We meta-analyze summary statistics from genome-wide association studies in 17,802 participants from the African Partnership for Chronic Disease Research (APCDR) and African ancestry individuals from the Blood Cell Consortium (BCX2) using GWAMA. We further carried out a Bayesian fine mapping to identify causal variants driving the association with WBC subtypes. To access the causal relationship between WBC subtypes and asthma, we conducted a two-sample Mendelian randomization (MR) analysis using summary statistics of the Consortium on Asthma among African Ancestry Populations (CAAPA: $n_{cases} = 7{,}009$, $n_{control} = 7{,}645$) as our outcome phenotype.

Results: Our metanalysis identified 269 loci at a genome-wide significant value of ($p = 5 \times 10^{-9}$) in a composite of the WBC subtypes while the Bayesian fine-mapping analysis identified genetic variants that are more causal than the sentinel single-nucleotide polymorphism (SNP). We found for the first time five novel genes (*LOC126987/MTCO3P14*, *LINC01525*, *GAPDHP32/HSD3BP3*, *FLG*-AS1/HMGN3P1, and *TRK-CTT13*-1/MGST3) not previously reported to be associated with any WBC subtype. Our MR analysis showed that Mon (IVW estimate = 0.38, CI: 0.221, 0.539, $p < 0.001$), Neu (IVW estimate = 0.189, CI: 0.133, 0.245, $p < 0.001$), and WBCc (IVW estimate = 0.185, CI: 0.108, 0.262, $p < 0.001$) are associated with increased risk of

asthma. However, there was no evidence of causal relationship between Lym and asthma risk.

Conclusion: This study provides insight into the relationship between some WBC subtypes and asthma and potential route in the treatment of asthma and may further inform a new therapeutic approach.

Keywords: white blood cell traits, asthma, metanalysis, Mendelian randomization, fine mapping

INTRODUCTION

Hematological cells play critical roles in protecting the host organism against immune assault (Li et al., 2013). Dysregulation or aberration within the hematopoietic system has been implicated in several diseases, and this could as well serve as prognostic markers. For example, an aberration in leukocytes could be an indicator of lymphoma, leukemia, heart failure, polycythemia, and hypertension (Buttari et al., 2015). White blood cells (WBCs) play a major role in both innate and adaptive immune systems, serving as the primary defense system against foreign assaults. Due to the role they play in defense and immunity, they are used as biomarkers for detecting inflammation (Gopal et al., 2012). High WBC count has been linked to the pathogenesis of different disease conditions such as cardiovascular disease and cancer (Keller et al., 2014). WBCs are categorized into five subtypes based on their functions and morphology: basophils (Baso), eosinophils (Eos), lymphocytes (Lym), monocytes (Mon), and neutrophils (Neu). WBC is an averagely heritable trait, with h^2 estimates between 0.14 and 0.40 across all the WBC types (Chinchilla-Vargas et al., 2020).

Asthma is a major health problem in the world, and scientific advances in the last two decades have improved our understanding and means of managing it effectively (Bateman et al., 2008). Studies have estimated the global prevalence of asthma to be 1%–18% (Masoli et al., 2004; Bateman et al., 2008; To et al., 2012). Despite its burden in Africa, asthma is classified as one of the neglected diseases with an estimated average of 12% (Asher et al., 2006; Adeloye et al., 2013; Hodonsky et al., 2020).

Several genome-wide association studies that have been carried out on WBC have reported more than 600 associated loci (Yang et al., 2007; Ferreira et al., 2009; Lo et al., 2011; Okada et al., 2011; Reiner et al., 2011; Keller et al., 2014; Astle et al., 2016). Despite the genetic diversity inherent within the African population, most WBC GWAS have been carried out in European- or East Asian-ancestry populations.

Mendelian randomization (MR) is a method that uses genetic proxies as instrumental variables and has been employed to investigate causal inference between an exposure and outcome phenotype (Fall et al., 2015).

Therefore, in this study, we performed an ancestry-specific metanalysis of GWAS of WBC and the five subtypes, Baso, Eos, Lym, Mon, and Neu in participants from two cohorts. The main aim of this study is to identify novel loci and signals associated with WBC traits and assess the causal relationships between these traits and asthma using SNPs as genetic instruments. Findings from this study may provide some biological and pathogenic insight into hematological disorders within the African population.

METHODS

Study Population

APCDR Cohort

The African Partnership for Chronic Disease Research (APCDR) is an organization set out to advance collaboration of epidemiological and genomic research of non-communicable diseases in sub-Saharan Africa. The APCDR cohort comprises four studies: the Ugandan Genome Resource (UGR), the Durban Diabetes Study (DDS), the Durban Case–Control Study (DCC), and the Africa America Diabetes Mellitus Study (AADM).

DDS is a population-based study carried out among non-pregnant black African individuals resident in eThekwini municipality in Durban South Africa from November 2013 to December 2014 (Hird et al., 2016). The survey, which had 1,165 individuals, combined different socioeconomic metrics with anthropometric measurements for infectious and non-communicable diseases. A detailed description of this study population and study design can be accessed in the paper (Hird et al., 2016).

The General Population Cohort (GPC) is a two-phase sample collection study composed of UGWAS (the first sample collected) and UG2G (second samples collected). GPC is a population-based study comprising approximately 22,000 residents of Kyamulibwa located in the southwestern part of Uganda. The goal of the study was to unravel the epidemiology and genetic drivers of non-communicable and communicable diseases using an Afrocentric population (Gurdasani et al., 2019).

The Diabetes Case–Control study is a study consisting of individuals of Zulu ancestry residing in KwaZulu-Natal, aged 40 and above that have diabetes, recruited in a tertiary health facility in Durban. A total of 1,600 individuals were recruited for this study.

The AADM study is a genetic epidemiological study of individuals with type 2 diabetes and associated diseases in Africans; this study has extensively been described in other studies (Rotimi et al., 2004; Adeyemo et al., 2015).

The Blood Cell Consortium Cohort

BCX2 is a consortium comprising of trans-ethnic data of blood cell traits of 746,667 individuals from five different ancestries. From the BCX2, we retrieved WBC traits from individuals with African ancestry composed of data pulled from the BioMe™

TABLE 1 | Total number of samples analyzed in APCDR and BCX2.

Traits	APCDR	BCX2	Total
WBC Count	2,741	15,061	17,802
Lymph count	2,681	13,477	16,158
Mono count	2,681	13,471	16,152
Eos count	2,671	11,615	14,286
Baso count	2,681	11,502	14,183
Neu count	2,671	13,476	16,147

**African Partnership for Chronic Disease Research (APCDR); Blood Cell Consortium Cohort (BCX2); Mon, Monocytes; Neu, Neutrophil count; WBC, White blood cell count (WBCC).*

BioBank Program, Cardiovascular Health Study, Genetic Epidemiology Research on Adult Health and Aging, Jackson Heart Study, The Multi-Ethnic Study of Atherosclerosis, and the UK Biobank African ancestry. The same units were used across the WBC traits in all the cohorts after inverse transformation [WBCc (10^9/L), Mon (10^9/L), Neu (10^9/L), Eos (10^9/L), Baso (10^9/L), and Lym (10^9/L)]. To obtain the pool effects of all the studies involved in the BCX2, an inverse variance-weighted fixed-effect meta-analyses was performed with the aid of GWAMA. For the study level association analysis, an additive genetic model of association was used to determine SNP association while a linear mixed-effect model was employed to factor in cryptic relatedness.

Hematological Phenotype

White blood and red blood indices of all participants in the DDS cohort were determined using a SYSMEX XT-2000i machine. For the UGR cohort, an 8.5-ml vacutainer was used to collect the venous blood while a 6-ml EDTA bottle was used to collect whole blood. The collected blood samples were stored at a temperature of 4–8°C. For hematological analysis, a swing bucket centrifuge was used to centrifuge the samples at 1,000–13,000 RCF (*g*) for 10 min.

Genotyping, Quality Control, and Imputation

Genotyping and quality control techniques have been described for each population previously (Gurdasani et al., 2019). Briefly, the DDS samples were genotyped on Illumina HumanOmni Multi-Ethnic GWAS/Exome Array employing the Infinium Assay. Illumina GenCall algorithm was used for genotype calling. The 5,000 GPC samples were genotyped on the Illumina HumanOmni 2.5M BeadChip array (**Table 1**). Quality control of the DDS cohort took into consideration the following criteria: exclusion of SNPs with heterozygosity > 4 SD from the mean, called proportion <97%, and sex check fails (*F* statistic >0.2 for women and <0.8 for men). Likewise, SNP QC ensured that called proportion was <−97%, relatedness (IBD >0.90), and Hardy–Weinberg disequilibrium ($p < 10^{-6}$). Imputation was done on pre-phased data with IMPUTE2 using a merged reference label of the whole genome sequence data from the African Genome Variation Project. Affymetrix Axiom PANAFR SNP array was used to genotype the AADM data as described previously (Adeyemo et al., 2015). Genotyping, imputation, and quality control of the African ancestry cohort of the BCX2 has been described somewhere else (Chen et al., 2020a).

Meta-Analysis

Prior to metanalysis, all summary statistics data were manually checked for integrity and accuracy (i.e., summary statistics downloaded have the required variables and are appropriately labeled). Some quality control measures were applied to the summary statistics, SNPs in each cohort having MAF >0.05 were selected. The SNP association *p* values from the summary statistics were meta-analyzed with the aid of GWAMA (Genome-Wide Association Meta-Analysis) (Morris, 2010). We further applied genomic control, and Manhattan plots and quantile–quantile plots were plotted for the meta-analyzed result.

Statistical Fine Mapping

Following our result output from meta-analysis, we used fine-mapping analysis to pick out possible causal SNPs for the locus ±250 kb of all the lead SNPs, using a Bayesian approach (Maller et al., 2012). The *Z* score was used to calculate the Bayes factor for each SNP denoted as $\boldsymbol{BF_i}$, given by

$$\boldsymbol{BF_i} = \boldsymbol{e}^{\left[\frac{Z*Z-\log(K)}{2}\right]}$$

Where *K* is the number of studies. The posterior probability of driving the association for each SNP was computed by

$$Posterior\ probability = \frac{BF_i}{\sum_j BF_j}$$

Where the summation in the denominator is over all SNPs at the locus.

Ninety-nine percent credible set sizes were derived by sorting all the SNPs according to their posterior probability $\boldsymbol{BF_i}$ at the locus from the highest to the lowest, and then counting the number of SNPs needed to attain a cumulative posterior probability that is greater than or equal to 0.99. Index SNPs accounting for more than 50% posterior probability of driving the WBC association at a given signal were defined as high confidence.

Identification of Potential Novel Variants and Locus Definition

Chen et al. (2020a) had previously performed a trans-ethnic and ancestry-specific GWAS of blood cell traits using 746,667 individuals (Chen et al., 2020b). To determine if the variants derived from our study are novel, and perhaps identify novel variants, we checked if any of our loci was reported or fall within ±250 kb window of those identified by Chen et al. A locus is further defined as ±250 kb around the significant SNPs ($p < 5 \times 10^{-09}$).

Two-Sample Mendelian Randomization

We performed a two-sample Mendelian randomization analysis using the R-based MR package Mendelian Randomization

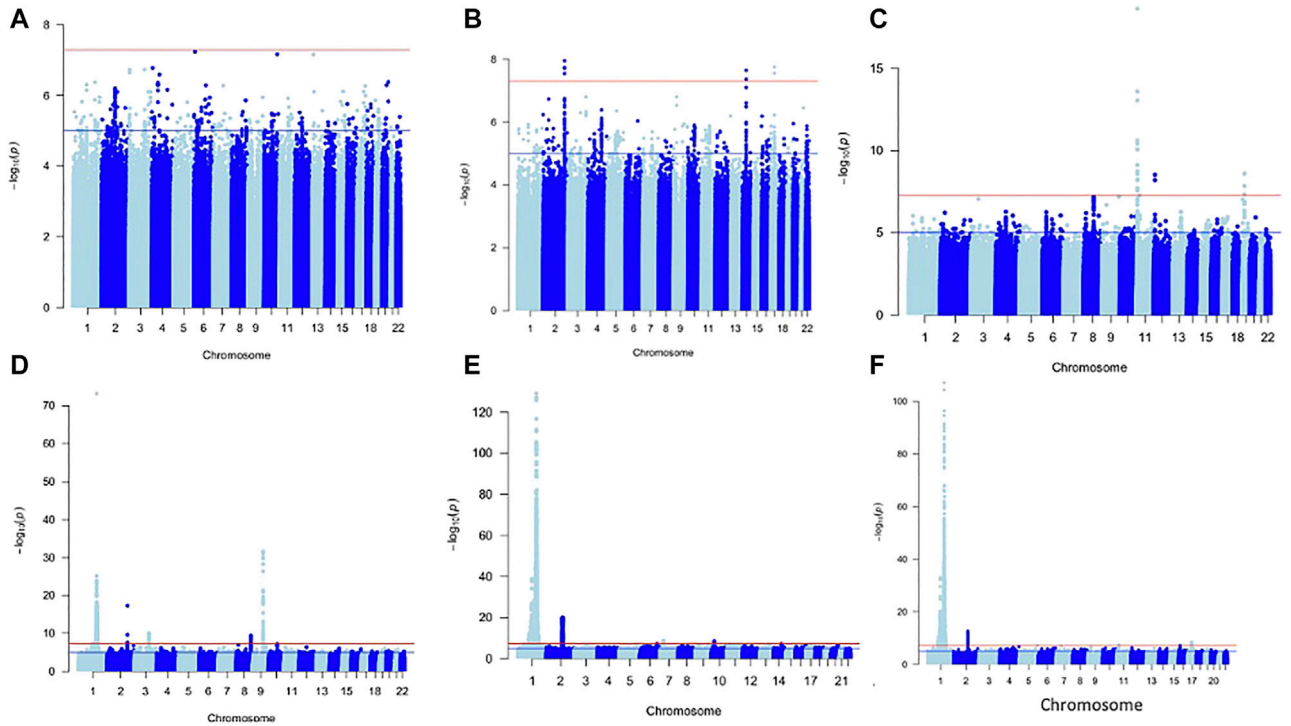

FIGURE 1 | Manhattan plot of metanalysis results of Baso **(A)**, Eos **(B)**, Lym **(C)**, Mon **(D)**, Neu **(E)**, and WBC **(F)**.

(Yavorska and Burgess, 2017). To identify independent genetic instruments, we used significant SNPs ($p < 5 \times 10^{-09}$) in the summary statistics of Lym, Mon, Neu, and WBCc derived from the metanalysis of APCDR and BCX2 summary statistics. We ensured that the instruments selected were not in LD with each other so that their impact on the exposure and outcome are uncorrelated. This was achieved by using $r^2 < 0.001$ and a 250-kb clumping upstream and downstream of the lead SNPs. We used asthma as our outcome phenotype. These data were selected from the Consortium on Asthma among African Ancestry Populations (CAAPA; 7,009 cases and 7,645 controls) (Daya et al., 2019). Causal estimates were calculated based on IVW and sensitivity analysis was carried out using MR-Egger and median-weighted methods. For each trait, we used *Q*-statistics to account for heterogeneity in the instruments and also excluded SNPs that may show pleiotropy. Proxy SNP for each missing SNP was obtained from LDProxy (Machiela and Chanock, 2015).

Functional Analysis

Functional analysis of SNPs identified by the metanalysis was carried out using FUMA (Watanabe et al., 2017). Independent SNPs in linkage disequilibrium or within the same genomic location with the sentinel SNPs were separated. A *p*-value cutoff of $p < 5 \times 10^{-09}$ and 1,000 G Phase3 AFR reference panel were used. We carried out other functional analysis such as eQTL tissue expression, pathway enrichment analysis, and biological process to get more insights into the functionality of the loci identified. eQTL mapping was performed by mapping SNPs to genes up to 1 Mb and using the Blood eQTL. We used GeneCards (Stelzer et al., 2016) to determine the functions of the gene. GWAS catalogue and Open Targets were used to identify any previously associated phenotypes of the lead SNPs. Annotation of all the genes identified in this study was done using NCBI's Genome data viewer. Pathway analysis of the identified loci was performed using Enrichr; Enrichr is an integrative web-based server that facilitates the visualization of the functional characteristics of a gene set. Enrichr is available online at http://amp.pharm.mssm.edu/Enrichr.

RESULTS

Description of Study

The total samples analyzed for each WBC trait are shown in **Table 1**. WBCc is the trait with the highest number of individuals and Baso has the lowest samples analyzed.

Metanalysis of WBC From APCDR and BCX2

For each of the WBC trait subtypes: Lym: 13 SNPs, Mon: 680 SNPs, Neu: 5,308 SNPs, and WBCc: 4,462 SNPs attained genome-wide significance ($p < 5 \times 10^{-09}$) (**Figure 1**). No SNPs in the Baso and Eos were significant at $p < 5 \times 10^{-09}$. Using a genomic distance of ±250 kb, we identified 4, 34, 124, and 108 lead SNPs within different loci in Lym, Mon, Neu, and WBCc, respectively, at $p < 5 \times 10^{-09}$ [**Supplementary Table S1 (ST1–ST4)**]. We examined if the sentinel SNPs were within ±250 kb of SNPs

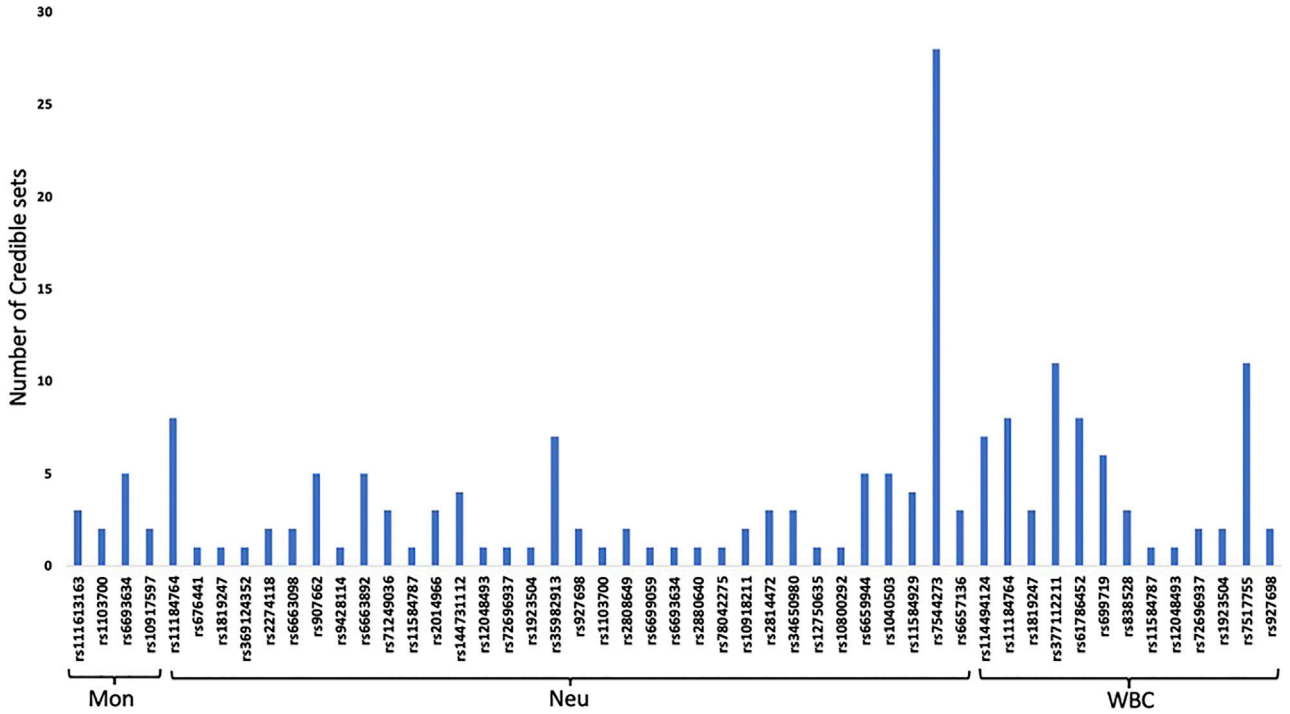

FIGURE 2 | 99% credible set of SNPs that jointly make 99% of the posterior probability of driving each WBC traits.

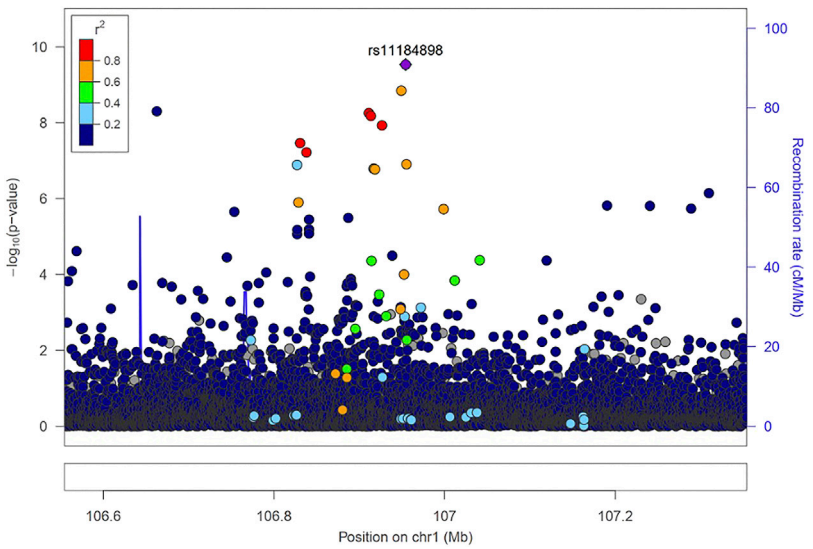

FIGURE 3 | Locus zoom regional association plot for rs1184898.

reported in the Chen et al. study. When compared to this, all the SNPs in Lym fall with ±250 kb of those previously reported, while only eight SNPs in Mon, 38 SNPs in Neu, and 13 SNPs in WBCc were unique. However, 4 SNPs out of the 8 SNPs unique in Mon, 3 out of 38 in Neu, and 1 out of 13 in WBCc have been previously identified to be associated with WBC traits when queried on the GWAS catalogue (Buniello et al., 2019) and Open Targets (Koscielny et al., 2017). Two SNPs (rs1103700 and rs6693634 mapped near *RPS10P8/CD1A* and *UHMK1/UQCRBP2*) are common in Mon and Neu.

Fine Mapping

Fine mapping seeks to analyze a trait-associated region to determine variants that are causal to a trait of interest. Fine

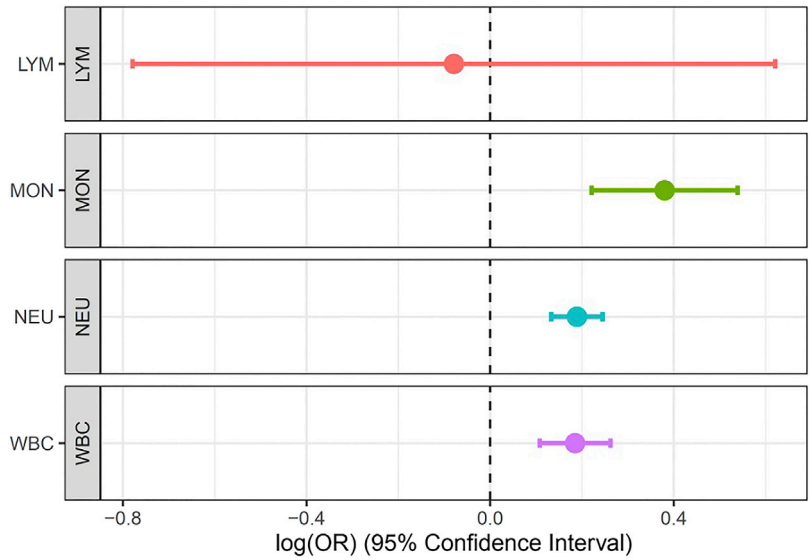

FIGURE 4 | Forest plot for the association between Lym, Mon, Neu, and WBC with asthma estimated using MR-IVW method.

mapping of loci was carried out within 250 kb upstream and downstream genomic distance of the sentinel variant identified *via* metanalysis. For each locus fine-mapped, we established the 99% credible set of SNPs that jointly make 99% of the posterior probability of driving the association (**Figure 2**). The fine-mapping analysis revealed that some of the lead variants such as rs369124352 (*MAGI3*), rs10918211 (*LRRC52*), and rs10800292 (*LINCO1363/POU2F1*) accounted for more than 50% of the posterior probability driving the association with these SNPs as the only variant within the 99% credible set. Several other variants were identified as the causal SNPs other than the sentinel SNP driving the association in WBC trait subtypes. Summary of the 99% credible set of variants driving the WBC trait can be found in **Supplementary Table S2**. We further went ahead and checked if these causal SNPs and their corresponding genes have been previously reported using the GWAS catalogue and Open Target; only five variants—rs11184898 (*LOC126987, MTCO3P14*) (**Figure 3**), rs907662 (*LINC01525*), rs7553527 (*GAPDHP32, HSD3BP3*), rs1923504 (*FLG-AS1, HMGN3P1*), and rs10733045 (*TRK-CTT13-1, MGST3*)—have not been reported to be associated with any WBC trait.

Causal Effects of WBC Traits on Asthma

We set out to evaluate the causal relationship between WBC traits (Lym, Mon, Neu, and WBCc) and asthma. Our MR analysis identified strong positive association of Mon (IVW estimate = 0.38, CI: 0.221, 0.539, $p < 0.001$), Neu (IVW estimate = 0.189, CI: 0.133, 0.245, $p < 0.001$), and WBCc (IVW estimate = 0.185, CI: 0.108, 0.262, $p < 0.001$) with increased risk of asthma. However, there were no evidence of causal relationship between Lym and asthma risk (IVW estimate = −0.079, CI: −0.779, 0.621, $p = 0.825$) (**Figure 4**, **Supplementary Table S3**), though there was an inverse relationship between Lym and asthma risk. The causal estimates of the weighted median and the MR-egger methods of Mon showed a similar effect to the IVW method (**Supplementary Table S3**).

Functional Analysis

MAGMA analysis using GTEx v8: 54 tissue types and GTEx v8: 54 general tissue types showed significant expression in whole blood (**Figure 5E**). The loci were also expressed in other tissues such as stomach and the colon but not at a significant level. Lipid and atherosclerosis, nitrogen metabolism, and FoxO signaling pathway are some of the pathways enriched by these loci [**Figure 5A**, **Supplementary Table S4 (ST1)**]. These loci were mapped at a significant level to thrombocytopenia-absent radius syndrome, autoimmune lymphoproliferative syndrome, lactose intolerance, central core myopathy etc. [**Figure 5B**, **Supplementary Table S4 (ST2)**]. They are also involved in the regulation of cellular response to transforming growth factor beta stimuli, regulation of transmembrane receptor protein serine/threonine kinase signaling pathway, regulation of cellular biosynthetic process, etc. (**Figure 5C**), **Supplementary Table S4 (ST3)**] and also in exogenous lipid antigen binding molecular functions (**Figure 5D**).

DISCUSSION

To the best of our knowledge, this is the first study to explore the causal relationship of WBC traits and asthma in an African population. Using metanalysis and Bayesian fine mapping, we identified 269 significant SNPs associated with WBC traits. Among these, five genes (*LOC126987*/*MTCO3P14*, *LINC01525*, *GAPDHP32*/*HSD3BP3*, *FLG*-AS1, and *TRK-CTT13*-1/MGST3) for the first time are reported to be associated with WBC traits. We also found causal relationship between Mon, Neu, and WBCc with asthma, while no causal relationship was seen between Lym and Asthma.

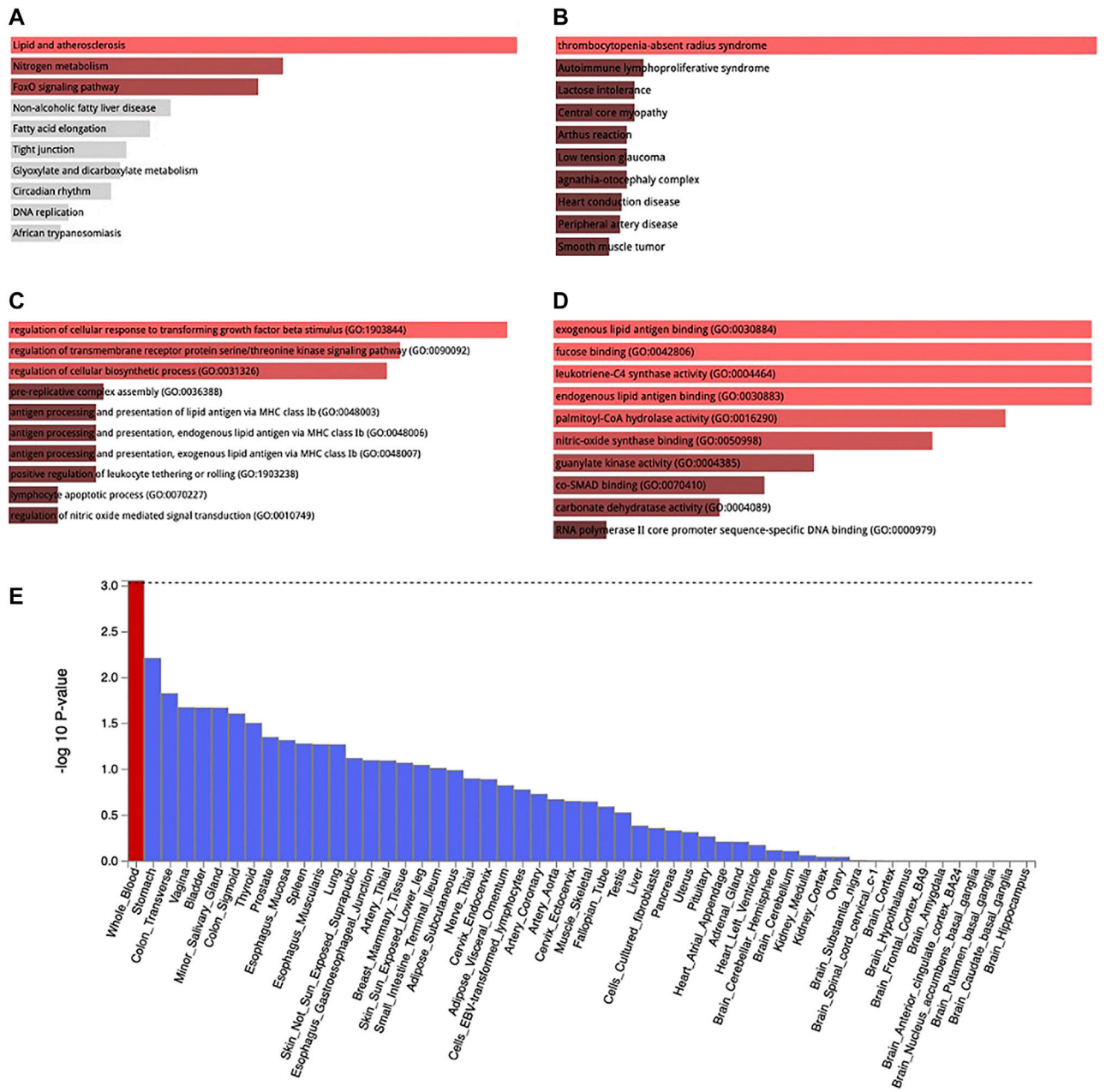

FIGURE 5 | KEGG pathway enrichment **(A)**, Jensen disease enrichment **(B)**, Biological process **(C)**, and Molecular function **(D)** of the genes associated with WBC subtypes identified by metanalysis. Tissue expression of enriched genes **(E)**. Red bar in 3 E connotes tissues with significant enrichment while blue bars indicate tissues with no significant enrichment.

FLG-AS1 (FLG Antisense RNA 1) is an RNA gene that is a member of the long non-coding RNA (lncRNA). FLG-AS1 is associated with asthma (Ferreira et al., 2019; Johansson et al., 2019; Pividori et al., 2019; Zhu et al., 2019; Olafsdottir et al., 2020), melanoma (Rashkin et al., 2020), eczema (Johansson et al., 2019; Kichaev et al., 2019), and acute myeloid leukemia (Lv et al., 2017). Microsomal glutathione S-transferase 3 encoded by *MGST3* has been shown to help in cellular defense of host organisms against lipid hydroperoxides, which may arise as a result of oxidative stress (Jakobsson et al., 1997; Stamova et al., 2013). There are evidences of oxidative stress in asthma (Misso and Thompson, 2005; Andrianjafimasy et al., 2017; Sahiner et al., 2018), this may explain the enrichment of this gene in the blood and causal association with asthma.

Interestingly, our result is consistent with the multivariable MR analysis of Astle et al. (2016), which showed a protective

effect of monocytes on Asthma. Furthermore, we also found a protective effect of neutrophils against asthma, consistent with Guyatt et al. (2020), which did not find substantial evidence for a harmful effect of neutrophils on asthma.

We found atherosclerosis to be significant in the KEGG pathway enrichment, and atherosclerosis is a major risk factor of coronary heart disease (CHD). Inflammation is an attribute of atherosclerosis; hence, several inflammatory cells such as Mon, Lym., Eos, and Neu have been implicated in CHD (Prentice et al., 1982; Madjid and Fatemi, 2013). Most importantly, several epidemiological studies have revealed that leukocyte count is an independent risk factor for CHD, and a risk factor for future cardiovascular events in individuals who do not have cardiovascular diseases (Weijenberg et al., 1996; Madjid et al., 2004; Chen et al., 2018). WBC count has also been suggested as a risk factor for atherosclerotic vascular diseases (Do LeeDo et al., 2001).

Significant enrichment of thrombocytopenia-absent radius syndrome (TAR), which is a rare congenital disorder (Greenhalgh et al., 2002), suggests the involvement of some WBC trait genes in the pathogenesis of this disease, as this disease is characterized by low levels of platelets in the blood (Greenhalgh et al., 2002).

Compared to other studies, one of the major strengths of this study is the use of African-ancestry data and the increased power from metanalysis. This enables the discovery of loci that have otherwise not been reported by previous studies. In addition, the CAAPA summary statistics is one of the latest cohorts of Asthma in Africa-admixed population.

We also applied different sensitivity analysis in our MR analysis, and we used strong instrumental variables by assessing the instrumental validity.

This study has its limitation, which is characteristic of two-sample MR analysis. Similar to any other non-experimental data that seek to make causal inference, where some experimentally unverifiable assumptions are made, our study is not an exception to this limitation.

Conclusively, we found evidence of causality between some WBC traits and asthma, and though some observational and MR data support these results, we believe more laboratorial experiments are needed to understand the biological mechanism of this causality.

KEY MESSAGES

- We carried out a well-powered meta-analysis genome-wide association study of white blood cell (WBC) traits in African populations.
- We identified not previously known genes driving WBC traits in an African population.
- Bayesian fine mapping identified more credible variants with high posterior probability associated with WBC subtypes.
- Mendelian Randomization Analysis found a causal relationship between monocyte count, neutrophil count, and white blood cell count with asthma in African populations.
- Findings from this study could provide more insight into the roles WBC traits play in the pathogenesis of asthma and could as well provide some directions in the treatment of asthma.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

SF conceptualized the study. OS led the main analyses. CS, TM, TC, and SF contributed to data analyses. OS wrote the first draft of the manuscript. TM, TC, and SF reviewed the first draft. SF and TC supervised the project. All the authors read and provided critical feedback on the paper.

FUNDING

SF is an international intermediate fellow funded by the Wellcome Trust grant (220740/Z/20/Z) at the MRC/UVRI and LSHTM. TC is an international training fellow supported by the Wellcome Trust grant (214205/Z/18/Z). MN acknowledges the support of Makerere University Non-Communicable Diseases (MakNCD). TM is an RHDGen PhD fellow supported by the Wellcome Trust, the University of Cape Town, and the inaugural Bongani Mayosi UCT-PHRI Scholarship (McMaster University). ON and SF are funded in part by the National Institutes of Health Common Fund to the H3ABioNet Project grant number 5U24HG006941-09.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.749415/full#supplementary-material

REFERENCES

Adeloye, D., Chan, K. Y., Rudan, I., and Campbell, H. (2013). An Estimate of Asthma Prevalence in Africa: a Systematic Analysis. *Croat. Med. J.* 54, 519–531. doi:10.3325/cmj.2013.54.519

Adeyemo, A. A., Tekola-Ayele, F., Doumatey, A. P., Bentley, A. R., Chen, G., Huang, H., et al. (2015). Evaluation of Genome Wide Association Study Associated Type 2 Diabetes Susceptibility Loci in Sub Saharan Africans. *Front. Genet.* 6, 335–339. doi:10.3389/fgene.2015.00335

Andrianjafimasy, M., Zerimech, F., Akiki, Z., Huyvaert, H., Le Moual, N., Siroux, V., et al. (2017). Oxidative Stress Biomarkers and Asthma Characteristics in

Adults of the EGEA Study. *Eur. Respir. J.* 50, 1701193. doi:10.1183/13993003.01193-2017

Asher, M. I., Montefort, S., Björkstén, B., Lai, C. K., Strachan, D. P., Weiland, S. K., et al. (2006). Worldwide Time Trends in the Prevalence of Symptoms of Asthma, Allergic Rhinoconjunctivitis, and Eczema in Childhood: ISAAC Phases One and Three Repeat Multicountry Cross-Sectional Surveys. *The Lancet* 368, 733–743. doi:10.1016/s0140-6736(06)69283-0

Astle, W. J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A. L., et al. (2016). The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* 167, 1415–1429. doi:10.1016/j.cell.2016.10.042

Bateman, E. D., Hurd, S. S., Barnes, P. J., Bousquet, J., Drazen, J. M., FitzGerald, M., et al. (2008). Global Strategy for Asthma Management and Prevention: GINA Executive Summary. *Eur. Respir. J.* 31, 143–178. doi:10.1183/09031936.00138707

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS Catalog of Published Genome-wide Association Studies, Targeted Arrays and Summary Statistics 2019. *Nucleic Acids Res.* 47, D1005–D1012. doi:10.1093/nar/gky1120

Buttari, B., Profumo, E., and Riganò, R. (2015). Crosstalk between Red Blood Cells and the Immune System and its Impact on Atherosclerosis. *Biomed. Res. Int.* 2015, 616834. doi:10.1155/2015/616834

Chen, H., Ding, X., and Li, J. (2018). *White Blood Cell Count: An Independent Predictor of Coronary Heart Disease Risk in Middle-Aged and Elderly Population with Hyperuricemia*. Medicine (Baltimore), 97. Available from: https://journals.lww.com/md-journal/Fulltext/2018/12210/White_blood_cell_count__an_independent_predictor.90.aspx.

Chen, M.-H., Raffield, L. M., Mousas, A., Sakaue, S., Huffman, J. E., Moscati, A., et al. (2020). Trans-ethnic and Ancestry-specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell* 182, 1198–1213. e14. doi:10.1016/j.cell.2020.06.045

Chen, M. H., Raffield, L. M., and Mousas, A. (2020). Trans-ethnic and Ancestry-specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell*.

Chinchilla-Vargas, J., Kramer, L. M., Tucker, J. D., Hubbell, D. S., Powell, J. G., Lester, T. D., et al. (2020). Genetic Basis of Blood-Based Traits and Their Relationship with Performance and Environment in Beef Cattle at Weaning. *Front. Genet.* 11, 717–813. doi:10.3389/fgene.2020.00717

Daya, M., Rafaels, N., Rafaels, N., Brunetti, T. M., Chavan, S., Levin, A. M., et al. (2019). Association Study in African-Admixed Populations across the Americas Recapitulates Asthma Risk Loci in Non-african Populations. *Nat. Commun.* 10, 880. doi:10.1038/s41467-019-08469-7

Do LeeDo, C., Folsom, A. R., and Nieto, F. J. (2001). White Blood Cell Count and Incidence of Coronary Heart Disease and Ischemic Stroke and Mortality from Cardiovascular Disease in African-American and White Men and Women: Atherosclerosis Risk in Communities Study. *Am. J. Epidemiol.* 154, 758–764. doi:10.1093/aje/154.8.758

Fall, T., Xie, W., Poon, W., Yaghootkar, H., Mägi, R., Knowles, J. W., et al. (2015). Using Genetic Variants to Assess the Relationship between Circulating Lipids and Type 2 Diabetes. *Diabetes* 64, 2676–2684. doi:10.2337/db14-1710

Ferreira, M. A. R., Hottenga, J.-J., Warrington, N. M., Medland, S. E., Willemsen, G., Lawrence, R. W., et al. (2009). Sequence Variants in Three Loci Influence Monocyte Counts and Erythrocyte Volume. *Am. J. Hum. Genet.* 85, 745–749. Available from. doi:10.1016/j.ajhg.2009.10.005

Ferreira, M. A. R., Mathur, R., Vonk, J. M., Szwajda, A., Brumpton, B., Granell, R., et al. (2019). Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct. *Am. J. Hum. Genet.* 104, 665–684. doi:10.1016/j.ajhg.2019.02.022

Gopal, S., Wood, W. A., Lee, S. J., Shea, T. C., Naresh, K. N., Kazembe, P. N., et al. (2012). Meeting the challenge of Hematologic Malignancies in Sub-saharan Africa. *Blood* 119, 5078–5087. doi:10.1182/blood-2012-02-387092

Greenhalgh, K. L., Howell, R. T., and Bottani, A. (2002). Thrombocytopenia-absent Radius Syndrome: a Clinical Genetic Study. *J. Med. Genet.* 39, 876–881881. doi:10.1136/jmg.39.12.876

Gurdasani, D., Carstensen, T., Fatumo, S., Chen, G., Franklin, C. S., Prado-Martinez, J., et al. (2019). Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa. *Cell* 179, 984–1002. doi:10.1016/j.cell.2019.10.004

Guyatt, A. L., John, C., and Williams, A. T. (2020). Mendelian Randomisation Analyses of Eosinophils and Other Blood Cell Types in Relation to Lung Function and Disease. *medRxiv*. doi:10.1101/2020.07.09.20148726

Hird, T. R., Young, E. H., Pirie, F. J., Riha, J., Esterhuizen, T. M., O'Leary, B., et al. (2016). Study Profile: The Durban Diabetes Study (DDS): A Platform for Chronic Disease Research. *Glob. Health Epidemiol. Genom* 1, e2. doi:10.1017/gheg.2015.3

Hodonsky, C. J., Baldassari, A. R., Bien, S. A., Raffield, L. M., Highland, H. M., Sitlani, C. M., et al. (2020). Ancestry-specific Associations Identified in Genome-wide Combined-Phenotype Study of Red Blood Cell Traits Emphasize Benefits of Diversity in Genomics. *BMC Genomics* 21, 228–314. doi:10.1186/s12864-020-6626-9

Jakobsson, P.-J., Mancini, J. A., Riendeau, D., and Ford-Hutchinson, A. W. (1997). Identification and Characterization of a Novel Microsomal Enzyme with Glutathione-dependent Transferase and Peroxidase Activities. *J. Biol. Chem.* 272, 22934–22939. doi:10.1074/jbc.272.36.22934

Johansson, Å., Rask-Andersen, M., Karlsson, T., and Ek, W. E. (2019). Genome-wide Association Analysis of 350 000 Caucasians from the UK Biobank Identifies Novel Loci for Asthma, hay Fever and Eczema. *Hum. .Mol .Genet* . 28, 4022–4041. doi:10.1093/hmg/ddz175

Keller, M. F., Reiner, A. P., Okada, Y., van Rooij, F. J. A., Johnson, A. D., Chen, M.-H., et al. (2014). Trans-ethnic Meta-Analysis of white Blood Cell Phenotypes. *Hum. Mol. Genet.* 23, 6944–6960. doi:10.1093/hmg/ddu401

Kichaev, G., Bhatia, G., Loh, P.-R., Gazal, S., Burch, K., Freund, M. K., et al. (2019). Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am. J. Hum. Genet.* 104, 65–75. doi:10.1016/j.ajhg.2018.11.008

Koscielny, G., An, P., and Carvalho-Silva, D. (2017). Open Targets: a Platform for Therapeutic Target Identification and Validation. *Nucleic Acids Res.*, D985–D994. doi:10.1093/nar/gkw1055

Li, J., Glessner, J. T., Zhang, H., Hou, C., Wei, Z., Bradfield, J. P., et al. (2013). GWAS of Blood Cell Traits Identifies Novel Associated Loci and Epistatic Interactions in Caucasian and African-American Children. *Hum. Mol. Genet.* 22, 1457–1464. doi:10.1093/hmg/dds534

Lo, K. S., Wilson, J. G., Lange, L. A., Folsom, A. R., Galarneau, G., Ganesh, S. K., et al. (2011). Genetic Association Analysis Highlights New Loci that Modulate Hematological Trait Variation in Caucasians and African Americans. *Hum. Genet.* 129, 307–317. doi:10.1007/s00439-010-0925-1

Lv, H., Zhang, M., Shang, Z., Li, J., Zhang, S., Lian, D., et al. (2017). Genome-wide Haplotype Association Study Identify the FGFR2 Gene as a Risk Gene for Acute Myeloid Leukemia. *Oncotarget* 8, 7891–7899. doi:10.18632/oncotarget.13631

Machiela, M. J., and Chanock, S. J. (2015). LDlink: a Web-Based Application for Exploring Population-specific Haplotype Structure and Linking Correlated Alleles of Possible Functional Variants: Fig. 1. *Bioinformatics* 31, 3555–3557. doi:10.1093/bioinformatics/btv402

Madjid, M., and Fatemi, O. (2013). Components of the Complete Blood Count as Risk Predictors for Coronary Heart Disease: In-Depth Review and Update. *Tex. Hear. Inst. J.* 40, 17–29. Available from: https://pubmed.ncbi.nlm.nih.gov/23467296.

Madjid, M., Awan, I., Willerson, J. T., and Casscells, S. W. (2004). Leukocyte Count and Coronary Heart Disease. *J. Am. Coll. Cardiol.* 44, 1945–1956. doi:10.1016/j.jacc.2004.07.056

The Wellcome Trust Case Control ConsortiumMaller, J. B., McVean, G., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., et al. (2012). Bayesian Refinement of Association Signals for 14 Loci in 3 Common Diseases. *Nat. Genet.* 44, 1294–1301. doi:10.1038/ng.2435

Masoli, M., Fabian, D., Holt, S., and Beasley, R. (2004). The Global burden of Asthma: Executive Summary of the GINA Dissemination Committee Report. *Allergy* 59, 469–478. doi:10.1111/j.1398-9995.2004.00526.x

Misso, N. L. A., and Thompson, P. J. (2005). Oxidative Stress and Antioxidant Deficiencies in Asthma: Potential Modification by Diet. *Redox Rep.* 10, 247–255. doi:10.1179/135100005x70233

Morris, P. A. (2010). GWAMA: Software for Genome-wide Association Meta-Analysis. *Bioinformatics* 11, 288.

Okada, Y., Hirota, T., Kamatani, Y., Takahashi, A., Ohmiya, H., Kumasaka, N., et al. (2011). Identification of Nine Novel Loci Associated with white Blood Cell Subtypes in a Japanese Population. *Plos Genet.* 7, e1002067–15. doi:10.1371/journal.pgen.1002067

Olafsdottir, T. A., Theodors, F., Bjarnadottir, K., Bjornsdottir, U. S., Agustsdottir, A. B., Stefansson, O. A., et al. (2020). Eighty-eight Variants Highlight the Role of T Cell Regulation and Airway Remodeling in Asthma Pathogenesis. *Nat. Commun.* 11, 393. doi:10.1038/s41467-019-14144-8

Pividori, M., Schoettler, N., Nicolae, D. L., Ober, C., and Im, H. K. (2019). Shared and Distinct Genetic Risk Factors for Childhood-Onset and Adult-Onset Asthma: Genome-wide and Transcriptome-wide Studies. *Lancet Respir. Med.* 7, 509–522. doi:10.1016/s2213-2600(19)30055-4

Prentice, R. L., Szatrowski, T. P., Fujikura, T., Kato, H., Mason, M. W., and Hamilton, H. H. (1982). Leukocyte Counts and Coronary Heart Disease in a Japanese Cohort. *Am. J. Epidemiol.* 116, 496–509. doi:10.1093/oxfordjournals.aje.a113434

Rashkin, S. R., Graff, R. E., Kachuri, L., Thai, K. K., Alexeeff, S. E., Blatchins, M. A., et al. (2020). Pan-cancer Study Detects Genetic Risk Variants and Shared Genetic Basis in Two Large Cohorts. *Nat. Commun.* 11, 4423. doi:10.1038/s41467-020-18246-6

Reiner, A. P., Lettre, G., Nalls, M. A., Ganesh, S. K., Mathias, R., Austin, M. A., et al. (2011). Genome-Wide Association Study of white Blood Cell Count in 16,388 African Americans: The continental Origins and Genetic Epidemiology Network (COGENT). *Plos Genet.* 7, e1002108. doi:10.1371/journal.pgen.1002108

Rotimi, C. N., Chen, G., Adeyemo, A. A., Furbert-Harris, P., Guass, D., Zhou, J., et al. (2004). A Genome-wide Search for Type 2 Diabetes Susceptibility Genes in West Africans: The Africa America Diabetes Mellitus (AADM) Study. *Diabetes* 53, 838–841. doi:10.2337/diabetes.53.3.838

Sahiner, U. M., Birben, E., Erzurum, S., Sackesen, C., and Kalayci, Ö. (2018). Oxidative Stress in Asthma: Part of the Puzzle. *Pediatr. Allergy Immunol.* 29, 789–800. doi:10.1111/pai.12965

Stamova, B. S., Tian, Y., Nordahl, C. W., Shen, M. D., Rogers, S., Amaral, D. G., et al. (2013). Evidence for Differential Alternative Splicing in Blood of Young Boys with Autism Spectrum Disorders. *Mol. Autism* 4, 30. doi:10.1186/2040-2392-4-30

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., et al. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr. Protoc. Bioinformatics* 54, 1–33. doi:10.30.110.1002/cpbi.5

To, T., Stanojevic, S., Moores, G., Gershon, A. S., Bateman, E. D., Cruz, A. A., et al. (2012). Global Asthma Prevalence in Adults: Findings from the Cross-Sectional World Health Survey. *BMC Public Health* 12, 204. doi:10.1186/1471-2458-12-204

Watanabe, K., Taskesen, E., Van Bochoven, A., and Posthuma, D. (2017). Functional Mapping and Annotation of Genetic Associations with FUMA. *Nat. Commun.* 8, 1–10. doi:10.1038/s41467-017-01261-5

Weijenberg, M. P., Feskens, E. J. M., and Kromhout, D. (1996). White Blood Cell Count and the Risk of Coronary Heart Disease and All-Cause Mortality in Elderly Men. *Atvb* 16, 499–503. doi:10.1161/01.atv.16.4.499

Yang, Q., Kathiresan, S., Lin, J. P., Tofler, G. H., and O'Donnell, C. J. (2007). Genome-wide Association and Linkage Analyses of Hemostatic Factors and Hematological Phenotypes in the framingham Heart Study. *BMC Med. Genet.* 8 Suppl 1, S12–S11. doi:10.1186/1471-2350-8-S1-S12

Yavorska, O. O., and Burgess, S. (2017). MendelianRandomization: an R Package for Performing Mendelian Randomization Analyses Using Summarized Data. *Int. J. Epidemiol.* 46, 1734–1739. doi:10.1093/ije/dyx034

Zhu, Z., Zhu, X., Liu, C. L., Shi, H., Shen, S., Yang, Y., et al. (2019). Shared Genetics of Asthma and Mental Health Disorders: a Large-Scale Genome-wide Cross-Trait Analysis. *Eur. Respir. J.* 54. doi:10.1183/13993003.01507-2019

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership