

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Meta-analysis and Multivariate GWAS Analyses in 77,850 Individuals of African Ancestry Identify Novel Variants Associated with Blood Pressure Traits

Brenda Udosen^{1,2,3,†}, Opeyemi Soremekun^{1,9,†}, Abram Kamiza^{1,4}, Tafadza Machipisa^{5,6}, Cisse Cheickna^{2,7}, Olaposi Omotuyi⁸, Mahmoud Soliman⁹, Mamadou Wélé^{2,7}, Oyekanmi Nashiru², Tinashe Chikowore^{10,11} and Segun Fatumo^{1,3,12,*}

¹ The African Computational Genomics (TACG) Research Group, MRC/UVRI, and LSHTM, Entebbe, Uganda

² The African Center of Excellence in Bioinformatics of Bamako (ACE-B), University of Sciences, Techniques and Technologies of Bamako, Bamako, Mali

³ H3Africa Bioinformatics Network (H3ABioNet) Node, Centre for Genomics Research and Innovation, NABDA/FMST, Abuja, Nigeria

⁴ Malawi Epidemiology and Intervention Research Unit, Lilongwe, Karonga, Malawi

⁵ Hatter Institute for Cardiovascular Diseases Research in Africa (HICRA), Department of Medicine, University of Cape Town, Cape Town, South Africa

⁶ Population Health Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton, Ontario, Canada

⁷ Department of Biological Sciences, Faculty of Sciences and Techniques, University of Sciences, Techniques and Technologies of Bamako, Bamako, Mali

⁸ Institute for Drug Research and Development, S.E. Bogoro Center, Afe Babalola University, Ado Ekiti, Nigeria

⁹ Molecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of Kwa-Zulu-Natal, Westville Campus, Durban, 4001, South Africa

¹⁰ Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

¹¹ MRC/Wits Developmental Pathways for Health Research Unit, Department of Pediatrics, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

¹² Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom

[†]Joint first authors

^{*}Corresponding author: Segun.fatumo@lshtm.ac.uk

Abstract: High blood pressure (BP) has been implicated as a major risk factor for cardiovascular diseases in several populations, including individuals of African ancestry. Despite the elevated burden of HBP-induced cardiovascular diseases in Africa and other populations of African descent, limited genetic studies have been carried out to explore the genetic mechanism driving this phenomenon.

We performed Genome-wide association univariate and multivariate analyses of both Systolic (SBP) and Diastolic Blood pressure (DBP) traits in 77,850 individuals of African ancestry. We used summary statistics data from six independent cohorts, including the African Partnership for Chronic Disease Research (APCDR), the UK Biobank, and the Million Veteran Program (MVP). FUMA was used to annotate, prioritize, visualize, and interpret our findings to gain a better understanding of the molecular mechanism(s) underlying the genetics of BP traits. Finally, we undertook a Bayesian fine-mapping analysis to identify potential causal variants.

Our meta-analysis identified 10 independent variants to be associated with both SBP and 9 with DBP traits. Whilst our multivariate GWAS method identified 21 independent signals, 18 of these SNPs have been previously identified. SBP was linked to gene sets involved in biological processes such as synapse assembly and cell-cell adhesion via plasma membrane adhesion. Of the 19 independent SNPs combinedly identified in the BP metaanalysis, only 11 variants had posterior probability (PP) of >50% including one novel variant rs562545 (*MOBP*, PP = 77%).

To facilitate further research and the fine-mapping of high-risk loci/variants in highly susceptible groups for cardiovascular disease and other related traits, large-scale genomic datasets are needed.



Our findings highlight the importance of including ancestrally diverse populations in large GWASs and the need for diversity in genetic research.

Keywords: systolic blood pressure; diastolic blood pressure; GWAS; high blood pressure; multivariate; univariate

1. Introduction

Blood pressure (BP) is a quantitative trait that is affected by both multifactorial genetic and environmental factors [1], [2], [3]. The heritability of high blood pressure is estimated to be 30-50% [4]. Elevated blood pressure otherwise called hypertension is the leading risk factor for many cardiovascular diseases like stroke and coronary artery diseases [5],[6]. The global prevalence of hypertension among adults aged 30-79 years increased significantly from 650 million in 1990 to 1.28 billion in 2019; with two-thirds of this burden coming from low-income countries (LMICs)[7]. When compared to other ethnic groups, African Americans and other African ancestry show a higher occurrence of high blood pressure [8], [9],[10],[11].

Despite the global rise in the disease burden among individuals of African ancestry, limited genome-wide association studies (GWASs) of blood pressure traits have been conducted or included individuals of African ancestry [12], [13]. For instance, the largest GWAS of blood pressure conducted to date in approximately a million individuals was predominantly consist of Europeans [15]. Additionally, only ~62% of all the genome-wide significant loci from this GWAS had the concordant direction of effects for individuals of African ancestry and moderate Pearson correlation coefficients with effect estimates in Europeans $r^2=0.37$ in Africans, compared to the strong $r^2=0.78$ for South Asians [15], [16], [17], [18]. Another example is that majority of blood pressure GWASs conducted in African ancestry populations have small sample sizes [19],[20],[21],[22],[23] and they mostly use single trait approach without giving due consideration to the phenotypic relatedness and the relationship between the two traits (SBP and DBP), which is a possible link between risk-related clinical measures and arterial properties [24],[25]. Thus, many novel insights into blood pressure traits in people of African ancestry remain to be discovered.

Furthermore, various GWAS reports have shown that the genetic determinants of blood pressure have small effect sizes and vary significantly between European and non-European populations [26]. Therefore, our study aim to extensively study the African population to better understand the genetic epidemiology underlying blood pressure traits in individuals of African descent. We also perform a multivariate GWAS in the hope that it will increase our study's statistical power over the univariate approach and consequently increase the overall number of novel loci observed in our study.

We conducted the largest GWAS of blood pressure in over 77,850 people from the African Partnership for Chronic Disease and Research (APCDR), African ancestry people from the United Kingdom (UK-Biobank), and the Million Veteran Program in this study (MVP). Figure 1 depicts the overall study design; we used fixed effects meta-analysis across the cohorts. We then performed a multivariate analysis, fine-mapping, pathway and tissue enrichment test analysis, and pathway and tissue enrichment test analysis to highlight relevant biological processes and investigate causal relationships with disease traits.

2. Materials and Methods

2.1. Study Population

The full description of the study population can be found in the supplementary 1 cohort description while the study design can be found in figure 1

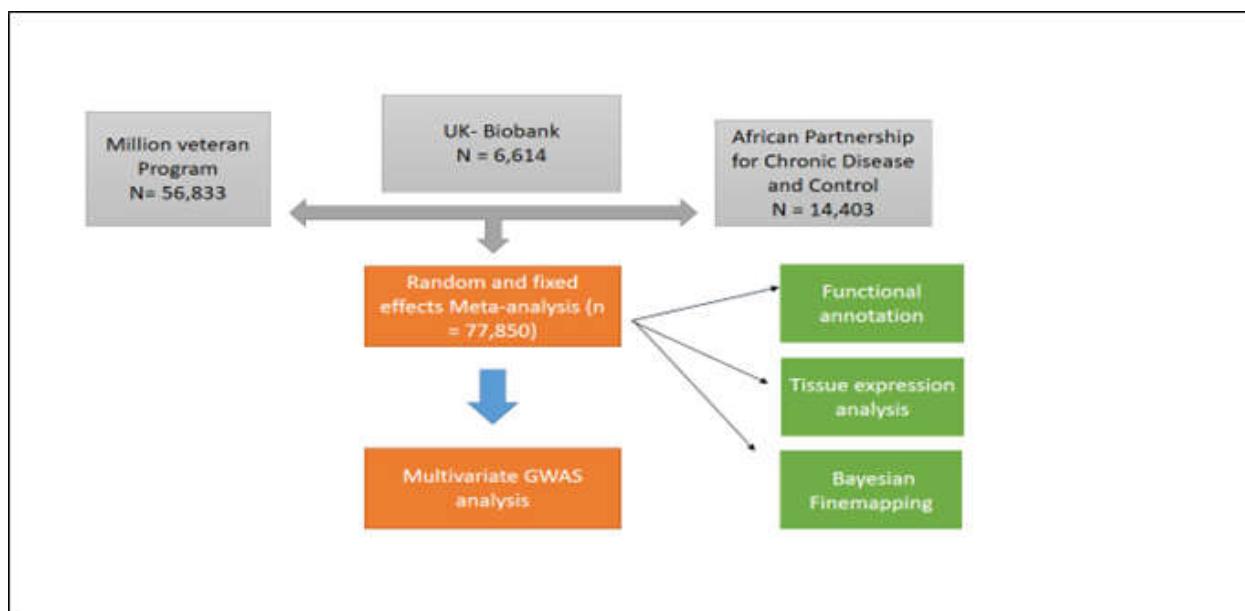


Figure 1. Study design schematic for discovery and validation of loci. APCDR; African Partnership for Control of Disease Research, UKB; United Kingdom Biobank, MVP; Million Veteran Program.

2.2. Meta-analysis of BP Summary Statistics in African Ancestry Individuals

We aggregated BP association summary statistics across the three cohorts (UK-BioBank, APCDR-UGR, DSS, DCC, and AADM, and the MVP) and performed an inverse-variance-weighted meta-analysis implemented in GWAMA [37]. We used a total of seventy-seven thousand eight hundred and fifty sample sizes across the studied cohorts (**Table 3**). The resulting output was used for subsequent downstream analyses, and we then plotted the resulting p-value in a Manhattan plot.

2.3. Tissue Expression Enrichment Pathway analysis

We performed a gene-based analysis with the MAGMA 1.6 software (Multi-marker Analysis of Genomic Annotation) [38], which is available in FUMA [39]. Magma gene-based analysis is useful for analyzing and detecting multiple genetic markers of individuals with a weak effect, which is common in polygenic traits. The 1000 Genomes were used as a reference dataset to account for LD between SNPs, and the confounding effects of gene density and gene size were used as covariates. The pathway and tissue expression analyses were performed using the default parameters in FUMA, using the results obtained from the meta-analysis.

2.4. Functional mapping and annotation analysis

We used an online functional mapping and annotation tool (FUMA) [39] to annotate SNPs from the GWAMA meta-analysis with their biological functionality and then mapped them to genes using positional mapping and QTL association (blood eQTL) [40]. The independent SNPs were classified based on their P-values as genome-wide significant ($P \leq 5.0 \times 10^{-8}$), independence from each other ($r^2 < 0.1$) and LD threshold within a 1 Mb window. Furthermore, the independent SNPs were annotated for functional consequences on gene function using ANNOVAR [41]. For the positional mapping, genes were mapped to SNPs if the physical distance between them was < 10 kb. The eQTL mapping used data from the blood cis-eQTL, and SNPs were mapped to genes on the premise that the SNPs had a significant effect on the expression of the gene. And also, SNPs were filtered using a CADD score > 12.37 , which is the threshold for deleterious scores (CADD scores are deleterious scores of genetic variants obtained by 63 functional annotations) [42]. Normalized gene expressions for 53 tissue types were obtained from GTEx. Other clumping parameters used were: reference panel to compute LD and MAF (minor allele

frequencies)= 1000 genome project was used (AFR) [43]; minor allele frequency filter > 0.01; maximum distance between LD blocks to merge into a single locus for genomic risk loci= 250 kb, lead SNPs were classified as SNPs that were in LD with each other at $r^2 < 0.1$.

2.5. Locus Definition

Lead SNPs from both univariate and multivariate analyses were defined based on positional mapping using 1Mb; SNPs that had reached the genome-wide significant threshold ($p < 5 \times 10^{-8}$) were considered to be associated with BP. Loci were defined by flanking distance mapping 500kb up and downstream of peak SNPs, and we retained SNPs with the lowest p-value from both the meta-analysis and multi-trait analysis.

2.6. Fine-mapping Analysis of Sentinel Variants

Following our result output from multi-trait analysis and meta-analysis, we performed Bayesian fine-mapping to identify possible causal variants for the locus $\pm 500\text{kb}$ of all the lead SNPs. We used a Bayesian approach [44] to fine-map the loci of the lead SNPs. The Z-scores for the SNPs were then used to compute the Bayes factor for each SNP denoted as BF_i , given by:

$$BF_i = e^{\frac{Z^*Z - \log(K)}{2}}$$

Where K is the number of studies). The posterior probability of driving the association for each SNP was computed by:

$$\text{Posterior probability} = \frac{BF_i}{\sum_j BF_j}$$

Where the summation in the denominator is over all SNPs at the locus.

Ninety-nine percent credible set sizes were calculated by sorting all SNPs at the locus according to their posterior probability from highest to lowest and then counting the number of SNPs required to achieve a cumulative posterior probability greater or equal to 0.99. High confidence was defined as index SNPs that account for more than 50% of the posterior probability of driving the BP association at a given signal.

2.7. Multivariate GWAS analysis

To further increase the statistical power for discovery, we employed a cross-phenotype approach implemented in the CPASSOC software [45]. The cross-phenotype association analysis accounts for the correlation of summary statistics data among traits and the participating cohorts and allows for both heterogeneity and homogeneity effects. The CPASSOC analysis generates two statistical tests: SHom and Shet, the latter of which is an extension of the former and improves statistical power when there is a difference in the genetic effect sizes across the traits. Meanwhile, the SHom test, which is similar to the fixed-effect meta-analysis approach, increases in power when the genetic effect sizes across the traits are the same.

3. Results

3.1. Results overview

We compiled GWAS summary statistics from three cohorts (Table 1), totaling 20,000,000 SNPs in 77,850 people of African descent. We used a univariate meta-analysis and a multivariate GWAS to find genetic variants linked to BP traits. At a genome-wide significant threshold of ($P < 5 \times 10^{-8}$) for both known and novel loci, the meta-analysis and multivariate approaches identified both known and novel loci. We used FUMA and fine mapping to gain more insight into the likely causal variants and molecular mechanism(s) that contribute to the genetics of BP traits.

Table 1. Description of study cohort used in this study.

Cohort	Continent	Country	Sample Size (N)	Phenotype	Type of Cohorts	Imputation Panel and Genome Build
APCDR-UGR (Gurdasani et.al.,2020)	Africa	Uganda	6,407	DBP SBP	Observational	Africa genome panel, hg19
APCDR-DCC (Gurdasani et.al.2020)	Africa	South-Africa	1,600	DBP SBP	Observational	Africa genome panel, hg19
APCDR-DDS (Gurdasani et.al.,2020)	Africa	South-Africa	1,165	DBP SBP	Case-control	Africa genome panel, hg19
APCDR-AADM (Gurdasani et.al.2020)	Africa	Nigeria Ghana Kenya	5,231	DBP SBP	Case-control	Africa genome panel, hg19
MVP – AFR	America	USA	56,833	DBP SBP	Observational	1000 Genome, hg19
UKB – AFR (Sudlow C,et.al.2015)	Europe	UK	6,614	DBP SBP	Observational	1000 Genome, hg19

3.2. Univariate GWAS meta-analysis

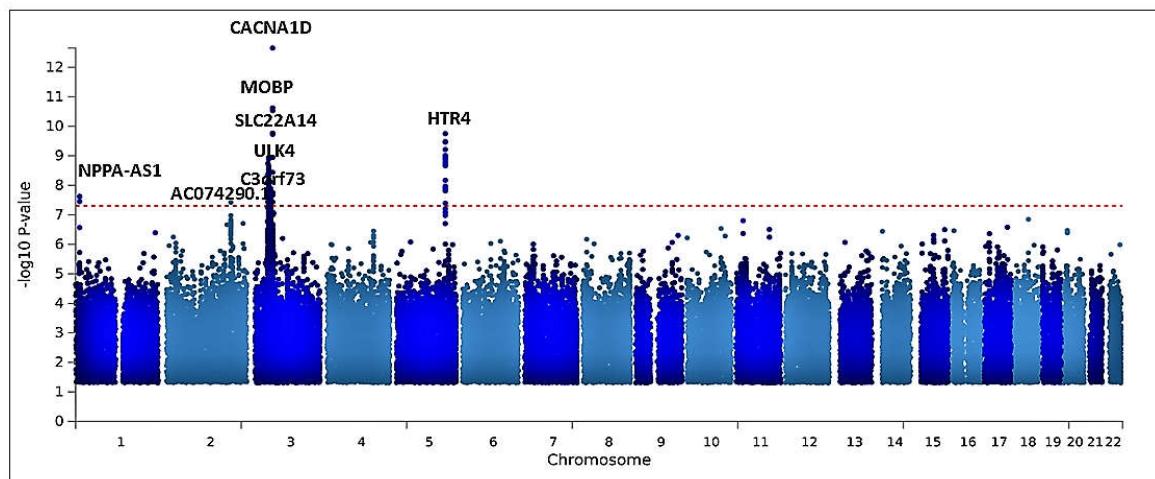
Meta-analysis of all six cohorts (n=77,850) identified 166 significant variants for SBP (**Supplementary Table 1**) and 184 genome-wide significance variants ($P<5\times10^{-8}$) for DBP (**Supplementary Table 2**). The significant SNPs for both blood pressure traits were clumped at $\pm 500\text{Kb}$ distance leaving 10 for SBP (**Supplementary Table 3**) and 9 lead SNPs for DBP (**Supplementary Table 4**). After clumping, out of the 19 SNPs identified across both traits, 2 were at least 1 Mbp away from any previously reported BP locus and therefore considered novel; an intergenic variant rs77534700 in *AC074290.1* ($p = 3.749\text{e}^{-08}$) and rs562545 an intronic variant at *MOBP* ($p = 1.823\text{e}^{-09}$) and are both associated with DBP trait (**Table 2, Figure 2A,2B**). Commonly known variants in *CACNA1D*, *HTR4*, *SLC22A14*, *NPPA-AS1*, *C3orf73*, *KCNK3*, *RPL35P4*, *CASZ1*, *NPPA-AS1*, *CTC-436K13.2*, *KCNN3*, *RSPO3*, *ATP2B1*, *FGF5*, *ULK4*, and *NPPA-AS1* were associated with SBP and DBP (**Supplementary table 3 and 4**).

The *CACNA1D* gene is an intron variant that has previously been identified in other populations, including African Americans, and is thought to regulate the renin-aldosterone-angiotensin system. The previously observed associations of the genetic variant in the meta-analyses were predominantly from the MVP cohorts, which might be driven by the fact that the largest proportion of our sample size came from the MVP's African American population. We plotted the resulting p-values from this association analysis on a Manhattan plot. (**Figure 2**).

Table 2. Novel distinct variant identified using a meta-analysis approach.

Nearest Gene	Lead SNPs	Chr	BP	Effect Allele	Other Allele	Trait	Beta	SE	MAF	P-value	Functional Consequence
AC074290.1	rs77534700	2	19467806 7	A	G	DBP	-0.0967	0.0176	0.0836	3.749e ⁻⁰⁸	Intergenic variant
<i>MOBP</i>	rs562545	3	39536524	A	G	DBP	0.0593	0.0099	0.8973	1.823e ⁻⁰⁹	Intron variant

(A)



(B)

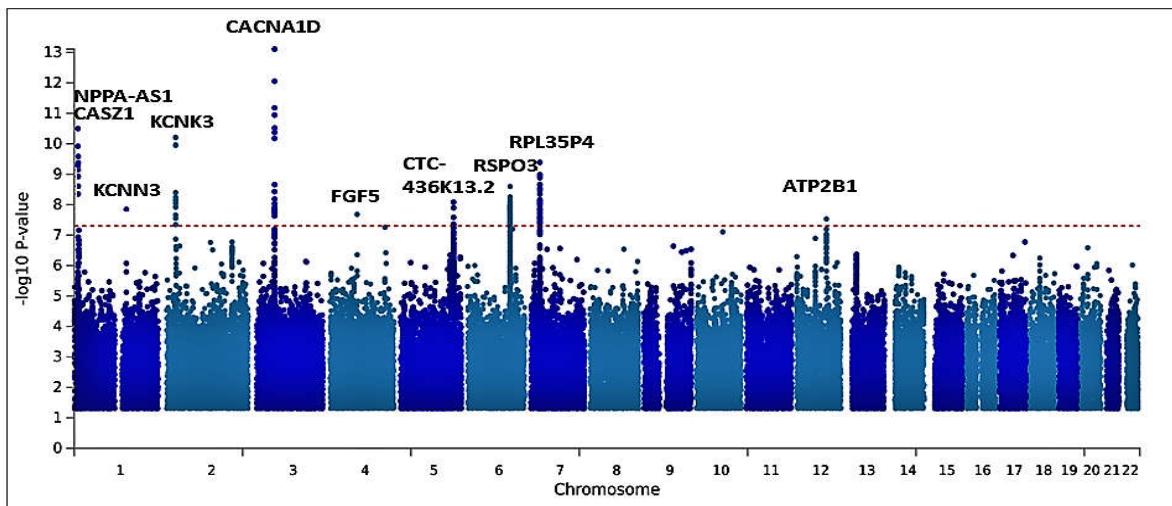


Figure 2. Manhattan plots showing the minimum P-value for the association across (A) DBP and (B) SBP blood pressure traits, computed using inverse-variance fixed-effect meta-analysis from 75,850 individuals. Each point on the Manhattan plots denotes a variant, with the X-axis representing the genomic position and the Y-axis representing the association level -log₁₀ (P-value). The horizontal red line shows the genome-wide significant threshold p-value = 5 X10⁻⁸.

3.3. Functional mapping and annotation analyses from FUMA from the meta-analysis

Using the default parameters on FUMA, we performed functional annotation on all SNPs in linkage disequilibrium (LD) to annotate and prioritize genes obtained from our meta-analysis. FUMA's SNP2GENE function revealed that the functional consequences of SNPs on genes in SBP included 12 lead SNPs (**Supplementary Table 5**), 20 independent SNPs (**Supplementary Table 6**), and 10 genomic risk loci (**Supplementary Table 7**). Following that, in DBP, 13 lead SNPs (**Supplementary Table 8**), 27 independent SNPs (**Supplementary Table 9**), and 8 genomic risk loci (**Supplementary Table 10**) were identified. The majority of the markers in SBP were intergenic, followed by those in the intronic region. In DBP, the most significant proportion of the markers were intronic SNPs, followed by intergenic SNPs (**Supplementary Figure 1**). 19 genes were identified through positional and/or eQTL mapping in SBP (**Supplementary Table 5**), and 34 genes for DBP SNPs (**Supplementary Table 6**).

The MAGMA gene set, tissue expression, and pathway analyses were carried out as part of the FUMA workflow. According to the MAGMA gene set study, after Bonferroni correction, no DBP gene sets were significant, but 10 SBP gene sets were (**Supplementary Table 7**). SBP was specifically linked to gene sets involved in biological processes such as synapse assembly (including presynaptic membrane assembly and organization, postsynaptic density, and specialization assembly), cell-cell adhesion via plasma membrane adhesion molecules (i.e., the connection of one cell to another cell through the use of adhesion molecules that are at least partially embedded in the plasma membrane), and the Takada gastric cancer copy number (i.e. candidate genes in the regions of copy number loss in gastric cancer cell lines).

Based on MAGMA tissue expression analysis, SBP was not associated with any gene property analysis for tissue specificity. However, DBP was associated with nine tissue specificities, significantly associated with brain tissues: particularly the hippocampus, brain substantia nigra, brain amygdala, brain putamen basal ganglia, hypothalamus, cortex, anterior cingulate cortex BA24, caudate basal ganglia, and nucleus accumbens basal ganglia (**Figure 3**). Notably, the strongest enrichment was observed for genes expressed in the hippocampus, followed by putamen basal ganglia.

As part of the FUMA pipeline, we used GENE2FUNC to test differentially expressed genes (DEGs); DBP found no association with our GTEx v8 54 tissue types, but SBP found two significantly up-regulated DEGs in the sigmoid and transverse colon (**Supplementary Figure 2**). Finally, we tested the enrichment of input gene sets (adjusted $p < 0.05$) and we found several gene sets previously associated with SBP, DBP, and correlated traits (**Supplementary Figure 2-8**).

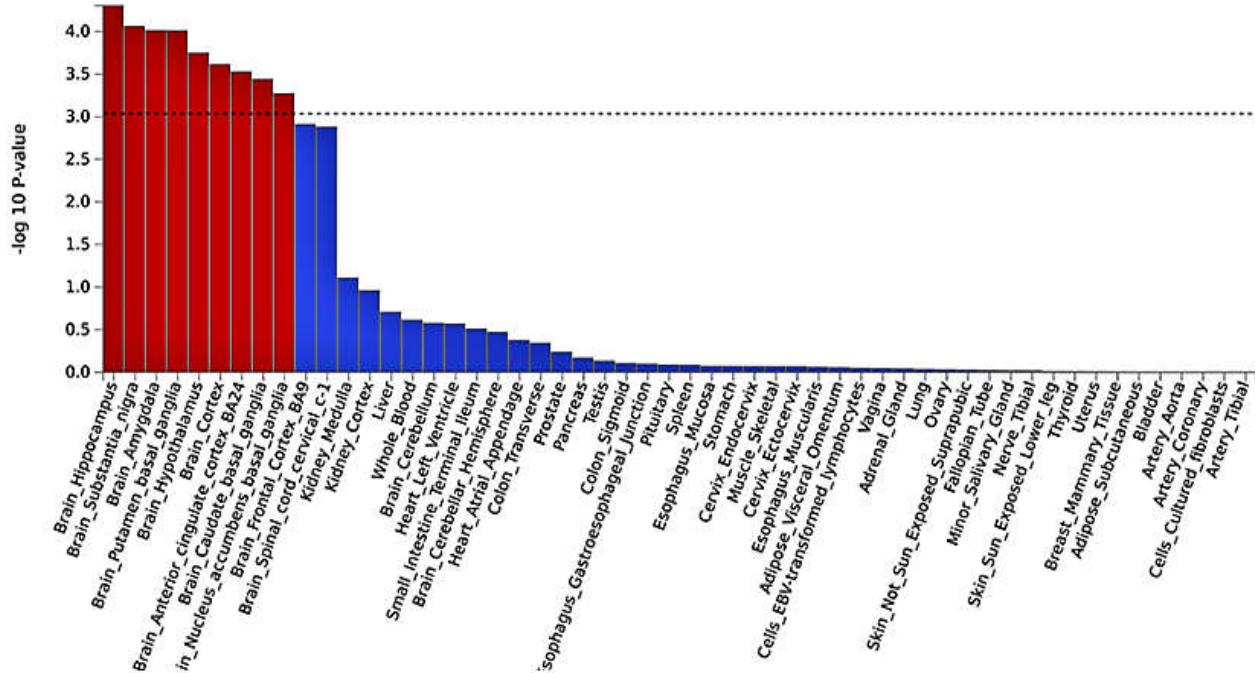


Figure 3. MAGMA tissue expression analysis using gene expression per tissue based on GTEx RNAseq data for 53 specific tissue types. Significant tissue is shown in red.

3.4. Fine-mapping of putatively causal variants

We performed Bayesian fine-mapping to pinpoint putative causal variants for distinct BP association signals using differences in the structure of LD between ancestry groups. Bayesian fine-mapping of the 19 distinct signals from the meta-analysis after clumping for DBP and SBP was undertaken in the region mapping 500kb up-and-downstream, which together accounted for 99% posterior probability and was based on association summary statistics from the meta-analysis GWAS. Only 11 variants from the lead

SNPs had >50% posterior probability (PP) of being causal, including the novel variant rs562545 (MOBP, PP = 77%) (**Supplementary Table 14, Figure 4**), and known variants; rs3821845 (CACNA1D, PP = 99%), rs12509595 (FGFR, PP = 99%), rs11129785 (SLC22A14, PP = 75%), rs12476527 (KCNK3, PP = 52%), rs5068 (NPPA-AS1, PP = 64%), rs73437338 (ATP2B1, PP = 52%), rs7720317 (CTC-436K13.2, PP = 59%), rs1984285 (KCNN3, PP = 99%). One of the lead SNPs, rs880315 (CASZ1, PP = 95%) was not the lead variant in the fine-mapping but was overlapped by other variants rs17035646 (PP = 99%) (**Supplementary Table 8**).

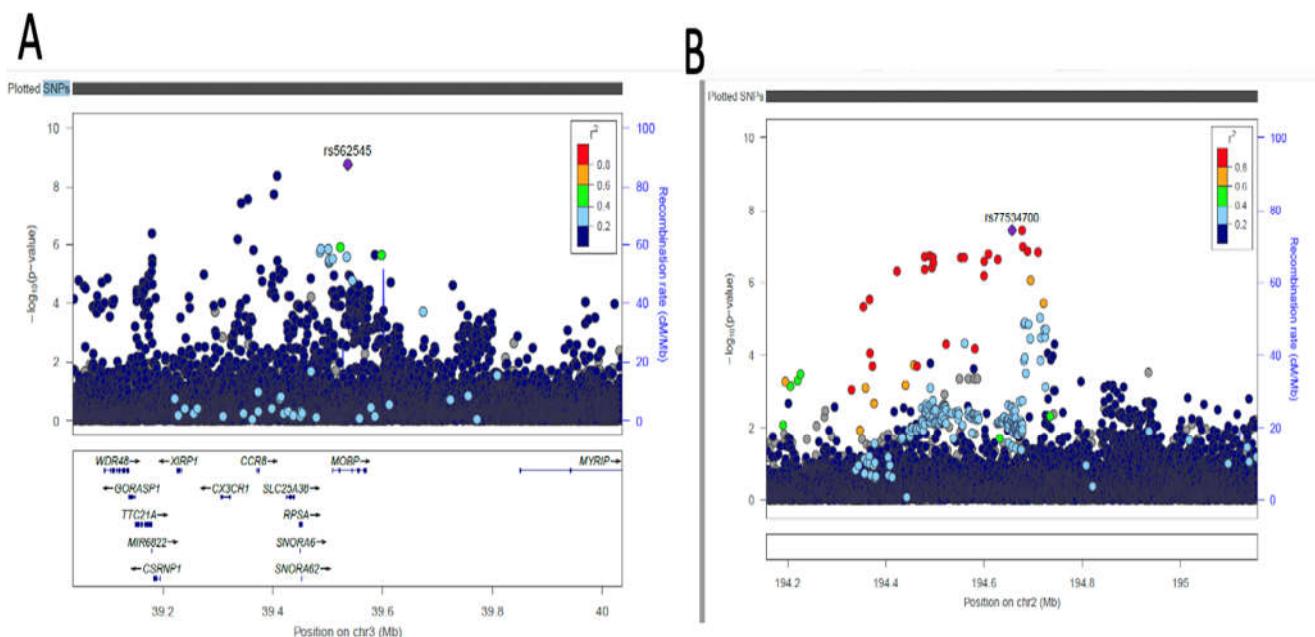


Figure 4. Regional visualization of the GWAS of $-\log_{10}$ of the P-value of genomic location MOBP (rs562545 in purple) (A), AC074290.1 (rs77534700 in purple) (B), with each dot representing SNP on the corresponding genes at the bottom. .

Multivariate GWAS analysis of blood pressure traits identifies additional novel loci

Using CPASSOC, we performed a multivariate analysis, this method identified 166 genome-wide significant loci associated with blood pressure (**Supplementary Table 10** ($p < 5 \times 10^{-8}$)). After clumping, we identified 21 independent significant SNPs, 3 novel SNPs, and 18 known SNPs (**Supplementary Table 16**). Interestingly, using the model assuming heterogeneity in CPASSOC, we identified 3 novel independent significant variants (**Table 3**), rs138493856 (*DNAJC17P1/GLULP6*, $p = 6.132e^{-9}$), rs139235642 (*RRM2*, $p = 2.798e^{-8}$) and rs72619992 (*LOC105377644*, $p = 1.134e^{-8}$). The resulting p-values were then plotted and visualized in a Manhattan plot (**figure 5**).

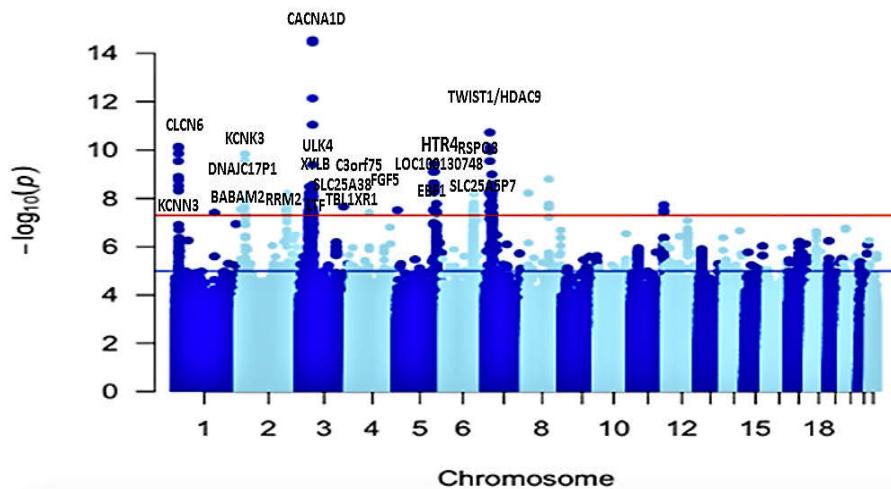


Figure 5. Manhattan plots showing P-value for the association computed from CPASSOC results output. Each point on the Manhattan plots denotes a variant, with the X-axis representing the genomic position and the Y-axis representing the association level $-\log_{10}(P\text{-value})$. The horizontal red line shows the genome-wide significant threshold $p\text{-value} = 5 \times 10^{-8}$.

Table 3. Novel variants of Blood pressure traits identified using multivariate methods.

Nearest Gene	Lead SNPs	Chr	BP	Effect Allele	Other Allele	HET_Pvalue	Functional Consequence
<i>DNAJC17P1/</i> <i>GLULP6</i>							
	rs138493856	2	194678067	A	G	6.1322e-09	Intergenic variant
<i>GLULP6</i> <i>GLULP6</i>							
<i>RRM2</i>	rs139235642	2	10278626	T	C	2.7981e-08	intron variant NMD transcript variant
<i>LOC105377644</i>	rs72619992	3	39407952	A	C	1.1339e-08	Intron variant

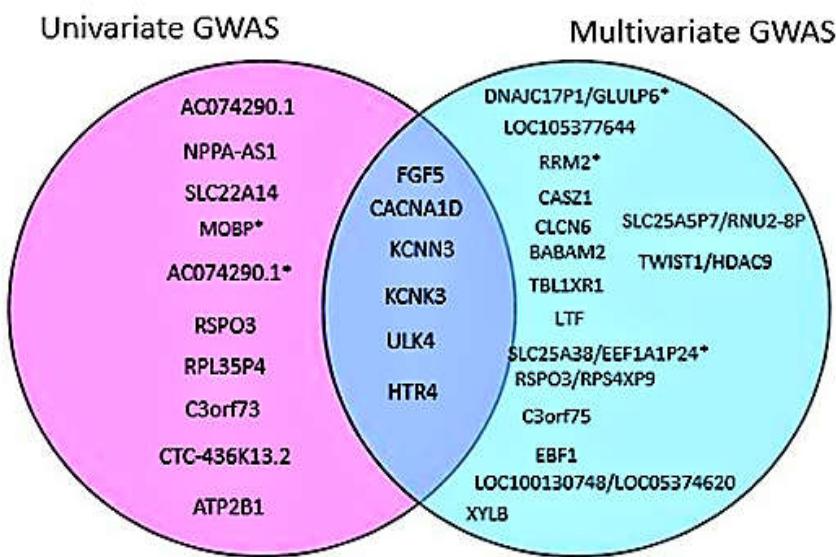


Figure 6. Venn diagram comparing the loci found by Meta-analysis and multivariate GWAS analysis. Both methods identified both known, the novel loci are indicated with an asterisk.

4. Discussion

This study describes the largest GWAS of blood pressure in African ancestry to date, involving a total of 77,850 individuals from the MVP, APCDR, and UK Biobank cohorts. The results of this analysis provide additional relevant information on the genetic and biological architecture of blood pressure traits in people of African ancestry.

According to our results, the multivariate GWAS approach had greater statistical power in identifying new variants than the univariate meta-analysis (Figure 6). Previous GWAS studies had shown the power of the multivariate approach, especially when dealing with traits that are highly correlated [27].

Five novel variants were discovered using both methods. The multivariate approach identified three variants: *DNAJC17P1/GLULP6* (rs138493856), *RRM2* (rs139235642), and *LOC105377644* (rs72619992), while the univariate approach identified two variants: *AC074290.1* (rs77534700) and *MOBP* (rs562545). The *DNAJC17P1/GLULP6* gene, which is located in the intergenic region, is known to be associated with susceptibility to infectious disease measurement [28] as well as educational attainment [29]. The *RRM2* is a protein-coding gene that encodes one of two non-identical subunits for ribonucleotide reductase and is highly expressed in the bone marrow (28.1) and lymph node (20.5), along with other tissues [30]. The high expression of this gene can lead to the abnormal proliferation of histiocytes and can also be used as a marker for malignant changes in ovarian endometriosis [31]. The rs72619992 variant in *LOC105377644* is an uncharacterized RNA gene that belongs to the ncRNA class and does not code for any protein. In *AC074290.1*, our univariate method identified an uncharacterized pseudogene. According to the GWAS catalog, the *MOBP* gene, which is a myelin-associated oligodendrocyte-associated protein, is linked to Alzheimer's disease [32], cognitive performance, and other brain-related disorders [33]. The *MOBP* gene is thought to be involved in both frontotemporal dementia and nervous system development. We used the largest BP summary statistics from European ancestry individuals to look up our lead SNPs, while some of the lead SNPs were found to be replicated at replicated at *P*-value 0.05. None of the SNPs identified as being novel replicated (Supplementary 13).

In the meta-analysis results, our in silico functional mapping and annotation analyses from FUMA revealed several biologically relevant signals. SBP gene sets, for example, were significantly associated with associated biological systems such as several synapse assembly components (such as components correlated to nervous system development/neurons and chemical or electrical synapses), candidate genes in regions of copy

number loss in gastric cancer cell lines, cell-cell adhesion via plasma membrane adhesion molecules (possibly, part of action potentials generated by the movement of ions through transmembranous channels), and cell-cell adhesion via plasma membrane adhesion.

In addition, the SBP meta-analysis tissue enrichment analysis was associated with significantly up-regulated DEGs in the sigmoid and transverse colon (**Supplementary figure 1**); which may suggest that gut microbiota may play a role in the regulation of gastro-renal axis and blood pressure [34]. Furthermore, the most interesting enrichment of input genes in gene sets significant in the Reactome was in the cardiac conduction and muscle contraction pathways for the SBP meta-analysis, which are the mechanisms and pathways that elicit rapid changes in the heart rate, blood pressure, and respond to changes in autonomic tone. On the other hand, our DBP MAGMA tissue expression analysis highlighted nine brain tissue types associated with DBP. For instance, the putamen, caudate, and nucleus accumbens basal ganglia are input nuclei as well as part of the corpus striatum, and the substantia nigra is a basal ganglia function-related nuclei, which are all involved in processing movement-related information. Dysfunction in this region is known to be associated with movement disorders like Huntington's, as correlated by GWAS catalog genes highlighted. In addition, the GWAS catalog genes included in gene sets included blood pressure traits and their interactions with alcohol and cigarette smoking, hence, these may be interesting environmental risk factors that should be investigated for their impact on BP traits in populations of African descent. Further investigation is needed to understand this, as different regions have different drinking and smoking habits.

Furthermore, our tissue expression analysis shows that DBP gene expression is enriched in the brain hippocampus (**Figure 3**), a brain region that is essential for learning and memory [35]. According to one study, hypertension is linked to decreased functional hippocampus connectivity and impaired memory [36]. As a result, more research is needed to understand our findings from *in silico* functional mapping and annotation analyses, as well as their mechanism.

Our current study has several strengths. First, our study is the largest SBP and DBP GWAS meta-analysis of an African population; thus, it has allowed us to find novel loci and replicate prior findings. Secondly, our functional mapping and annotation found several biologically relevant regions, that support our genetic findings, and these regions, tissues, and pathways are good candidates to explore further to elucidate the pathogenesis of blood pressure-related disorders like hypertension and prevent or treat them better. Finally, fine-mapping recommended target candidate loci to test *in vivo* and *in vitro* to improve our understanding of the regulators and genetic factors that affect blood pressure traits. The CPASSOC used for our multivariate GWAS increased statistical power and reflected the nature of the multivariate effect of traits on the genetic factor.

One of our limitations is that the "black" participants in our study are primarily from admixed regions with a variety of characteristics. Thus, our study used a small sample size from the continental African population, and this may be the reason why most of our variants were identified from the MVP dataset, as this data had the largest sample size (Table 3). Although our study is the largest study of SBP and DBP genetics, the overall sample size was small compared to contemporary GWASs for other traits. Thus, future studies will need to include more continental Africans to make sure our genetic risk factors can be used to make genetic risk scores that are inclusive of all or most African populations and their full range of diversity. Due to the diversity in African genome, latent sub-structuring could inflate the results, but this effect was minimized by adjusting for principal components in the GWAS model by the contributing cohorts. Second, the paucity of functional genomics information specific to African people makes it challenging to evaluate the functional relevance of the relationships found. Thirdly, regional environmental factors, including dietary variations, variances in the prevalence of TB and HIV, and other non-communicable disease factors, could potentially have an impact on BP outcomes; however, there isn't enough research on these aspects in our target group. Afrocentric GWAS data are grossly limited, hence we used Blood pressure GWAS data from individual of African ancestry available and accessible to the authors.

In conclusion, we have conducted the largest GWAS of blood pressure in African ancestry, which has significantly enabled an in-depth understanding of the genetic component. Our analysis emphasizes the relevance of applying fine-mapping and multivariate methods to correlated trait and their increase in statistical power toward the discovery of causal variants. These strategies offer a reliable approach to better understanding the genetic epidemiology of blood pressure disease in African ancestry and treatment development strategy. Lastly, to better understand the implication of these results, future studies could replicate the result on the European population.

Author Contributions: SF conceptualize the idea, BU and OS contributed to the preliminary analyses and manuscript drafting, OS contributed to the Meta-analysis and Multivariate analyses, TM contributed to the functional analysis, SF contributed to the supervision of the manuscript, TC and AK contributed to the overviewing of the manuscript.

Funding: OS is supported by the Africa Research Excellence Fund (AREF-325-SORE-F- C0904). SF is supported 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).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of respective data used in this study. The DDS was approved by the Biomedical Research Ethics Committee at the University of KwaZulu-Natal (reference: BF030/12) and the UK National Research Ethics Service (reference: 14/WM/1061). This UGR -GPC was approved by the Science and Ethics Committee of the UVRI, the Ugandan National Council for Science and Technology and the East of England-Cambridge South NHS Research Ethics Committee United Kingdom. The DCC was approved by the Biomedical Research Ethics Committee at the University of KwaZulu-Natal (reference: BF078/08) and the UK National Research Ethics Service (reference: 11/H0305/6). The AADM study obtained ethical approval from the Institutional Review Boards (IRB) of all participating institutions.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

- [1] M. Muñoz, R. Pong-Wong, O. Canela-Xandri, K. Rawlik, C. S. Haley, and A. Tenesa, "Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank," *Nat. Genet.*, vol. 48, no. 9, Art. no. 9, Sep. 2016, doi: 10.1038/ng.3618.
- [2] M. FEINLEIB *et al.*, "THE NHLBI TWIN STUDY OF CARDIOVASCULAR DISEASE RISK FACTORS: METHODOLOGY AND SUMMARY OF RESULTS," *Am. J. Epidemiol.*, vol. 106, no. 4, pp. 284–295, Oct. 1977, doi: 10.1093/oxfordjournals.aje.a112464.
- [3] N. R. Poulter, D. Prabhakaran, and M. Caulfield, "Hypertension," *The Lancet*, vol. 386, no. 9995, pp. 801–812, Aug. 2015, doi: 10.1016/S0140-6736(14)61468-9.
- [4] D. Levy *et al.*, "Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness," *BMC Med. Genet.*, vol. 8 Suppl 1, p. S3, Sep. 2007, doi: 10.1186/1471-2350-8-S1-S3.
- [5] C. L. Allen and U. Bayraktutan, "Risk factors for ischaemic stroke," *Int. J. Stroke Off. J. Int. Stroke Soc.*, vol. 3, no. 2, pp. 105–116, May 2008, doi: 10.1111/j.1747-4949.2008.00187.x.
- [6] European Stroke Organisation (ESO) Executive Committee and ESO Writing Committee, "Guidelines for management of ischaemic stroke and transient ischaemic attack 2008," *Cerebrovasc. Dis. Basel Switz.*, vol. 25, no. 5, pp. 457–507, 2008, doi: 10.1159/000131083.

[7] NCD Risk Factor Collaboration (NCD-RisC), "Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants," *Lancet Lond Engl.*, vol. 398, no. 10304, pp. 957–980, Sep. 2021, doi: 10.1016/S0140-6736(21)01330-1.

[8] A. P. Carson, G. Howard, G. L. Burke, S. Shea, E. B. Levitan, and P. Muntner, "Ethnic differences in hypertension incidence among middle-aged and older adults: the multi-ethnic study of atherosclerosis," *Hypertens. Dallas Tex* 1979, vol. 57, no. 6, pp. 1101–1107, Jun. 2011, doi: 10.1161/HYPERTENSIONAHA.110.168005.

[9] G. S. Berenson, W. A. Wattigney, and L. S. Webber, "Epidemiology of hypertension from childhood to young adulthood in black, white, and Hispanic population samples," *Public Health Rep. Wash. DC* 1974, vol. 111 Suppl 2, pp. 3–6, 1996.

[10] Writing Group Members *et al.*, "Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association," *Circulation*, vol. 133, no. 4, pp. e38–360, Jan. 2016, doi: 10.1161/CIR.0000000000000350.

[11] D. Chor *et al.*, "Prevalence, Awareness, Treatment and Influence of Socioeconomic Variables on Control of High Blood Pressure: Results of the ELSA-Brasil Study," *PLoS One*, vol. 10, no. 6, p. e0127382, 2015, doi: 10.1371/journal.pone.0127382.

[12] D. Levy *et al.*, "Genome-wide association study of blood pressure and hypertension," *Nat. Genet.*, vol. 41, no. 6, pp. 677–687, Jun. 2009, doi: 10.1038/ng.384.

[13] C. Newton-Cheh *et al.*, "Eight blood pressure loci identified by genome-wide association study of 34,433 people of European ancestry," *Nat. Genet.*, vol. 41, no. 6, pp. 666–676, Jun. 2009, doi: 10.1038/ng.361.

[14] "Genetic Variants in Novel Pathways Influence Blood Pressure and Cardiovascular Disease Risk," *Nature*, vol. 478, no. 7367, pp. 103–109, Sep. 2011, doi: 10.1038/nature10405.

[15] E. Evangelou *et al.*, "Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits," *Nat. Genet.*, vol. 50, no. 10, pp. 1412–1425, Oct. 2018, doi: 10.1038/s41588-018-0205-x.

[16] G. B. Ehret *et al.*, "Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk," *Nature*, vol. 478, no. 7367, pp. 103–109, Oct. 2011, doi: 10.1038/nature10405.

[17] N. Kato *et al.*, "Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians," *Nat. Genet.*, vol. 43, no. 6, pp. 531–538, Jun. 2011, doi: 10.1038/ng.834.

[18] H. R. Warren *et al.*, "Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk," *Nat. Genet.*, vol. 49, no. 3, pp. 403–415, Mar. 2017, doi: 10.1038/ng.3768.

[19] A. Adeyemo *et al.*, "A genome-wide association study of hypertension and blood pressure in African Americans," *PLoS Genet.*, vol. 5, no. 7, p. e1000564, Jul. 2009, doi: 10.1371/journal.pgen.1000564.

[20] N. Franceschini *et al.*, "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.*, vol. 93, no. 3, pp. 545–554, Sep. 2013, doi: 10.1016/j.ajhg.2013.07.010.

[21] A. Giri *et al.*, "Trans-ethnic association study of blood pressure determinants in over 750,000 individuals," *Nat. Genet.*, vol. 51, no. 1, pp. 51–62, Jan. 2019, doi: 10.1038/s41588-018-0303-9.

[22] J. Liang *et al.*, "Single-trait and multi-trait genome-wide association analyses identify novel loci for blood pressure in African-ancestry populations," *PLOS Genet.*, vol. 13, no. 5, p. e1006728, May 2017, doi: 10.1371/journal.pgen.1006728.

[23] E. R. Fox *et al.*, "Association of genetic variation with systolic and diastolic blood pressure among African Americans: the Candidate Gene Association Resource study," *Hum. Mol. Genet.*, vol. 20, no. 11, pp. 2273–2284, Jun. 2011, doi: 10.1093/hmg/ddr092.

[24] N. Franceschini *et al.*, "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.*, vol. 93, pp. 545–54, Sep. 2013, doi: 10.1016/j.ajhg.2013.07.010.

[25] L. M. Hendry, V. Sahibdeen, A. Choudhury, S. A. Norris, M. Ramsay, and Z. Lombard, "Insights into the genetics of blood pressure in black South African individuals: the Birth to Twenty cohort," *BMC Med. Genomics*, vol. 11, p. 2, Jan. 2018, doi: 10.1186/s12920-018-0321-6.

[26] J. He *et al.*, "Genome-wide association study identifies 8 novel loci associated with blood pressure responses to interventions in Han Chinese," *Circ. Cardiovasc. Genet.*, vol. 6, no. 6, pp. 598–607, Dec. 2013, doi: 10.1161/CIRCGENETICS.113.000307.

[27] S. Fatumo, T. Carstensen, O. Nashiru, D. Gurdasani, M. Sandhu, and P. Kaleebu, "Complimentary Methods for Multivariate Genome-Wide Association Study Identify New Susceptibility Genes for Blood Cell Traits," *Front. Genet.*, vol. 10, 2019, Accessed: Jun. 02, 2022. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fgene.2019.00334>

[28] R. Nudel *et al.*, "A large-scale genomic investigation of susceptibility to infection and its association with mental disorders in the Danish population," *Transl. Psychiatry*, vol. 9, no. 1, p. 283, Nov. 2019, doi: 10.1038/s41398-019-0622-3.

[29] A. Okbay *et al.*, "Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals," *Nat. Genet.*, vol. 54, no. 4, pp. 437–449, Apr. 2022, doi: 10.1038/s41588-022-01016-z.

[30] L. Fagerberg *et al.*, "Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics," *Mol. Cell. Proteomics MCP*, vol. 13, no. 2, pp. 397–406, Feb. 2014, doi: 10.1074/mcp.M113.035600.

[31] B. Yang, T. Wang, N. Li, W. Zhang, and Y. Hu, "The High Expression of RRM2 Can Predict the Malignant Transformation of Endometriosis," *Adv. Ther.*, vol. 38, no. 10, pp. 5178–5190, Oct. 2021, doi: 10.1007/s12325-021-01888-3.

[32] R. Sherva *et al.*, "Genome-wide association study of the rate of cognitive decline in Alzheimer's disease," *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 10, no. 1, pp. 45–52, Jan. 2014, doi: 10.1016/j.jalz.2013.01.008.

[33] A. C. Need *et al.*, "A genome-wide study of common SNPs and CNVs in cognitive performance in the CANTAB," *Hum. Mol. Genet.*, vol. 18, no. 23, pp. 4650–4661, Dec. 2009, doi: 10.1093/hmg/ddp413.

[34] J. Yang, P. A. Jose, and C. Zeng, "Gastrointestinal–Renal Axis: Role in the Regulation of Blood Pressure," *J. Am. Heart Assoc.*, vol. 6, no. 3, p. e005536, Mar. 2017, doi: 10.1161/JAHA.117.005536.

[35] K. S. Anand and V. Dhikav, "Hippocampus in health and disease: An overview," *Ann. Indian Acad. Neurol.*, vol. 15, no. 4, pp. 239–246, 2012, doi: 10.4103/0972-2327.104323.

[36] R. Feng, E. T. Rolls, W. Cheng, and J. Feng, "Hypertension is associated with reduced hippocampal connectivity and impaired memory," *EBioMedicine*, vol. 61, p. 103082, Nov. 2020, doi: 10.1016/j.ebiom.2020.103082.

[37] R. Mägi and A. Morris, "GWAMA: Software for genome-wide association meta-analysis," *BMC Bioinformatics*, vol. 11, p. 288, May 2010, doi: 10.1186/1471-2105-11-288.

[38] C. A. de Leeuw, J. M. Mooij, T. Heskes, and D. Posthuma, "MAGMA: Generalized Gene-Set Analysis of GWAS Data," *PLoS Comput. Biol.*, vol. 11, no. 4, p. e1004219, Apr. 2015, doi: 10.1371/journal.pcbi.1004219.

[39] K. Watanabe, E. Taskesen, A. Van Bochoven, and D. Posthuma, "Functional mapping and annotation of genetic associations with FUMA," *Nat. Commun.*, vol. 8, no. 1, pp. 1–11, Dec. 2017, doi: 10.1038/s41467-017-01261-5.

[40] H.-J. Westra *et al.*, "Systematic identification of trans eQTLs as putative drivers of known disease associations," *Nat. Genet.*, vol. 45, no. 10, pp. 1238–1243, Oct. 2013, doi: 10.1038/ng.2756.

[41] K. Wang, M. Li, and H. Hakonarson, "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data," *Nucleic Acids Res.*, vol. 38, no. 16, p. e164, Sep. 2010, doi: 10.1093/var/gkq603.

[42] K. M. W. Dm, J. P. O. Bj, C. Gm, and S. J., "A general framework for estimating the relative pathogenicity of human genetic variants," *Nat. Genet.*, vol. 46, no. 3, Mar. 2014, DOI: 10.1038/ng.2892.

[43] 1000 Genomes Project Consortium *et al.*, "A global reference for human genetic variation," *Nature*, vol. 526, no. 7571, pp. 68–74, Oct. 2015, DOI: 10.1038/nature15393.

[44] A. Hutchinson, H. Watson, and C. Wallace, "Improving the coverage of credible sets in Bayesian genetic fine-mapping," *PLOS Comput. Biol.*, vol. 16, no. 4, p. e1007829, Apr. 2020, DOI: 10.1371/journal.pcbi.1007829.

[45] X. Li and X. Zhu, "Cross-Phenotype Association Analysis Using Summary Statistics from GWAS," *Methods Mol. Biol. Clifton NJ*, vol. 1666, pp. 455–467, 2017, DOI: 10.1007/978-1-4939-7274-6_22.

