


## Article

# Whole-Genome Sequencing of 100 Genomes Identifies a Distinctive Genetic Susceptibility Profile of Qatari Patients with Hypertension

Alsamman M. Alsamman <sup>1</sup>, Hakeem Almbrazi <sup>2</sup> and Hatem Zayed <sup>3,\*</sup>

<sup>1</sup> Department of Genome Mapping, Agricultural Genetic Engineering Research Institute, Agricultural Research Center, Giza P.O. Box 12619, Egypt; smahmoud@ageri.sci.eg

<sup>2</sup> Research Branch, Sidra Medicine, Doha P.O. Box 26999, Qatar; hakeem.almbrazi@gmail.com

<sup>3</sup> Department of Biomedical Sciences, College of Health Sciences, QU Health, Qatar University, Doha P.O. Box 2713, Qatar

\* Correspondence: hatem.zayed@qu.edu.qa

**Abstract:** Essential hypertension (EH) is a leading risk condition for cardiovascular and renal complications. While multiple genes are associated with EH, little is known about its genetic etiology. Therefore, this study aimed to screen for variants that are associated with EH in 100 hypertensive/100 control patients comprising Qatari individuals using GWASs of whole-genome sequencing and compare these findings with genetic data obtained from more than 10,000 published peer-reviewed studies on EH. The GWAS analysis performed with 21,096 SNPs revealed 38 SNPs with a significant  $\geq 4$  log-*p* value association with EH. The two highest EH-associated SNPs (rs921932379 and rs113688672) revealed a significance score of  $\geq 5$  log-*p* value. These SNPs are located within the inter-genic region of *GMPS-SETP14* and *ISCA1P6-AC012451.1*, respectively. Text mining yielded 3748 genes and 3078 SNPs, where 51 genes and 24 SNPs were mentioned in more than 30 and 10 different articles, respectively. Comparing our GWAS results to previously published articles revealed 194 that are unique to our patient cohort; of these, 13 genes that have 26 SNPs are the most significant with  $\geq 4$  log-*p* value. Of these genes, *C2orf47-SPATS2L* contains nine EH-associated SNPs. Most of EH-associated genes are related to ion gate channel activity and cardiac conduction. The disease–gene analysis revealed that a large number of EH-associated genes are associated with a variety of cardiovascular disorders. The clustering analysis using EH-associated SNPs across different ethnic groups showed high frequency for the minor allele in different ethnic groups, including Africans, East Asians, and South Asians. The combination of GWAS and text mining helped in identifying the unique genetic susceptibility profile of Qatari patients with EH. To our knowledge, this is the first small study that searched for genetic factors associated with EH in Qatari patients.

**Keywords:** genome-wide association; GWAS; hypertension; genomics; whole-genome sequencing; genomic biomarkers; SNPs; precision medicine; Qatar Biobank; QBB; Qatar



**Citation:** Alsamman, A.M.; Almbrazi, H.; Zayed, H. Whole-Genome Sequencing of 100 Genomes Identifies a Distinctive Genetic Susceptibility Profile of Qatari Patients with Hypertension. *J. Pers. Med.* **2022**, *12*, 722. <https://doi.org/10.3390/jpm12050722>

Academic Editor: George P. Patrinos

Received: 3 March 2022

Accepted: 26 April 2022

Published: 29 April 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Cardiovascular disorders cause about 17 million deaths worldwide, with about one-third of these being due to hypertension complications [1]. Essential hypertension (EH) is a chronic and age-related disorder that frequently causes cardiovascular and renal risks. EH affects 25–35% of the adult population in both developed and developing countries, leading to stroke and cardiovascular disorders. Of these, up to 60–70% are in their mid-sixties [2,3]. Several factors, such as the large arteries, endocrine factors, central nervous system, and microcirculation, are involved. The correlation between these factors varies with age and reflects the heterogeneous pattern of hemodynamic changes [2]. A data survey of the global burden of disease showed that in 2015, 7.8 million deaths were related to a systolic blood pressure of  $\geq 140$  mmHg, which is the current clinical threshold for identifying hypertension [4].

Both genetic and environmental factors are involved in the development of EH symptoms [5,6]. Although several research efforts have been made, the genetic etiology of EH remains partly unknown. EH is a polygenic condition that results from the inheritance of a variety of risk genes, each with a small influence according to sex, age, race, or lifestyle [2,7]. Population-based studies have reported that hypertension runs in approximately 20% of families, which might be related to genetic factors and/or dietary, lifestyle, and activity levels [8]. In twin studies, this risk was up to 60%, where dizygotic twins showed a lower association than monozygotic twins [2,9].

The hereditary contribution to hypertension variations is estimated to be around 30%, but the genetic history of critical hypertension is ambiguous and poorly understood [10]. Several studies have associated dozens of genes with hypertension. However, genes linked to the function of the kidneys that are involved in fluid and electrolyte balance control are the most significant determinants of blood pressure [11]. These genes include aldosterone synthase genes *CYP11B1* and *CYP11B2* [12], *HSD11B2* [13] *eNOS* [14], and others. Genetic variants play a significant role in the malfunctioning of such genes, resulting in unpredictability in the primary biological pathways controlling blood pressure. Furthermore, recent research indicates that epigenetic mechanisms, such as DNA methylation, post-translational histone modifications, and microRNAs, play a significant role in the molecular dynamics underlying EH etiopathogenesis. Some of these modifications affect blood pressure and thus lead to a variety of cardiovascular disorders [12].

Single-nucleotide polymorphisms (SNPs) are one of the most common genetic variations in humans, and they are the most common and functionally important source of evolution [15]. Rapid advances in sequencing technologies permit affordable, reliable population SNP genotyping. These technologies offer the ability to screen thousands of SNPs that could be used as biomarkers to identify hypertension susceptibility loci through genome-wide association studies (GWAS) [7,16]. GWASs apply a statistical inference approach, applying estimation and hypothesis testing on the correlation between a trait and variants such as SNPs [17]. The rate of identification of validated SNPs in GWASs associated with blood pressure has recently increased exponentially [18,19]. GWASs enabled medical researchers to identify several loci that may contain causative variants or genes that lead to overall EH susceptibility at a single locus [20]. Furthermore, the ongoing global development of national biobanks has provided researchers with sufficient bio-specimens for well-established cohort studies to explore genome-wide significant associations for EH in specific ethnic groups [21].

Drawing inferences from what is already known is an important process in creating knowledge [22]. In this context, we can draw more insightful conclusions by analyzing the available information about EH. Text mining is a method of data mining in which a large volume of structured and unstructured text data are processed to generate potentially valuable information [23]. Due to the significant expansion of medical knowledge, text mining is being used to enhance medical analysis and provide more comprehensive techniques for data processing. It is a rich resource for the acquisition of knowledge from the current research literature. The only downside is that it requires a high level of data filtration and manipulation skills [24,25]. It enables researchers to classify various database resources, speed up the structuring of information derived from unstructured data, and gain deeper insights. A text mining approach was used to organize and evaluate published scientific literatures on cancer [26], diabetes [27], and schizophrenia [28]. Some reports used text mining to classify biomedical literatures in hypertension and to identify useful treatments using linguistic techniques [29].

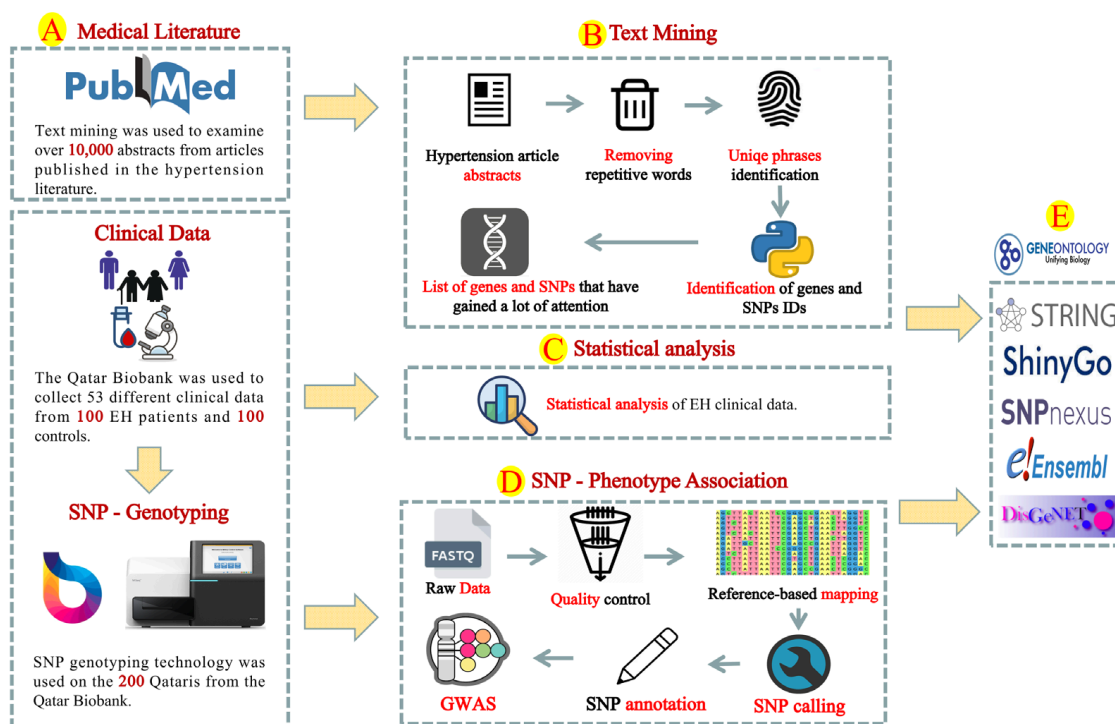
We conducted a GWAS on 100 Qatari patients with hypertension and 100 normal Qatar individuals. Additionally, we performed text-mining analysis to identify the unique variants that predispose Qataris to hypertension.

## 2. Methodology

### 2.1. SNPs Associated with EH in Qatari Population

#### 2.1.1. Study Design and Statistical Analysis

Figure 1 is a schematic representation of the methodology section. The data of the cohort studied were acquired from the Qatar Biobank (QBB), which comprised 100 patients with EH and 100 normal controls. All participants were Qatari [30] and were sequenced through the national Qatar genome project [31] (Figure 1A). The eligibility criteria for hypertensive subjects were: (1) age  $\geq 35$ –70; (2) systolic blood pressure (SBP)  $\geq 140$  mmHg and/or diastolic blood pressure (DBP)  $\geq 90$  mmHg; (3) absence of secondary causes of hypertension based on comprehensive biochemical and clinical studies; and (5) absence of pharmacological therapy for hypertension. Fifty-three different associated data of both the control and patients were retrieved from QBB and used for statistical analysis (Supplementary S1). These data include clinical parameters, such as age, blood pressure factors, body mass index, sodium, potassium, chloride, bicarbonate, urea, creatine kinase, creatinine, glucose, total protein, albumin, cholesterol, triglyceride, calcium, phosphorus, iron, magnesium, Fibrinogen, vitamin B12, insulin, and others. The statistical analyses, which included principal component analysis and correlation, were carried out in R, and missing values were imputed using the “mice” package with the predictive mean matching (PMM) method [32].



**Figure 1.** A flowchart representing the genetic resources and bioinformatics methods used in the present research. Thousands of hypertension-related articles were reviewed, as well as the SNP genotyping of 200 Qataris with high blood pressure (A) using text mining (B), statistical analysis (C), and GWAS (D). The text mining analysis included eliminating redundant words, recognizing unique phrases, detecting genes and SNP expressions, and categorizing these genes and SNPs based on their prevalence and redundancy (B). Fifty-three different clinical data of the studied patients were retrieved from QBB and used for statistical analysis (C). The SNP genotyping data were analyzed using a quality control method before being aligned/mapped to the human genome, where discovered SNPs were annotated and their association with high blood pressure was investigated using GWAS analysis (D). The details obtained from text mining and GWAS analysis was analyzed using gene ontology and enrichment analyses (E).

### 2.1.2. SNP Genotyping and Computational Analysis

DNA was isolated from 5 mL of blood samples using the Puregene DNA extraction kit (Gentra Systems, Minneapolis, MN, USA) according to the manufacturer's instructions. DNA quantification was performed using Fluorometer Qubit 2.0 (Invitrogen, Carlsbad, CA, USA). Whole-genome sequencing (WGS) was conducted by QBB on DNA samples from 200 participants. Illumina HiSeq X Ten sequencers were used for WGS analysis, where raw data were processed using bioinformatics pipelines (Figure 1C). FastQ data were converted to paired-end FASTQ format using bcl2fastq conversion tool. FastQC software was used to assess the raw data quality. The human genome GRCh37 version was used as a reference for data passing quality control using Burrow-Wheeler Aligner (BWA) aligner (v7.12). SNP calling was performed using HaplotypeCaller provided by Genome Analysis Toolkit (GATK v3.3) [33]. The SNP effect and annotation were categorized using SnpEff (v4.1) [34]. SNP was discarded from analysis if the genotyping call rate was <95%, the minor allele frequency (MAF) was <1%, the heritability error rate was >1%, or the Hardy-Weinberg equilibrium was  $p < 10^{-6}$ . Filtered SNPs were used for further analysis. Fisher's exact test [35] was conducted to determine the statistical significance of allele frequency differences between case and control groups (Figure 1C).

### 2.2. Exploring Published Hypertension Literature

The available hypertension reports were explored in order to obtain more context on the prior information of the genetic mechanism of hypertension (Figure 1A,B). PubMed-NCBI was used to retrieve all abstracts of the scientific articles that reported hypertension-associated genes and SNPs (Figure 1A,B). The query of "hypertension genes" was used to download all abstracts of medical articles published up to "Tue, 24 November 2020 05:35:29". The text analysis included "10,001" different articles. The data mining was conducted through Python programming language (Supplementary S2). Common English phrases and word redundancy were removed. A list of human gene terminology was prepared using human genome (GRCh38 version) obtained from the NCBI database. The SNP reference numbers (RSs) were extracted using regular expression and string-searching algorithms (Supplementary S2).

### 2.3. Gene and SNP Enrichment Analysis

The collected hypertension-associated genes from the GWAS and text mining analyses were used for gene annotation and enrichment analysis. Comprehensive computational analysis was conducted using several bioinformatics tools (Figure 1D). Gene enrichment analysis was conducted using ShinyGo [36] and uniprot database [37]. SNPnexus and Ensembl [38] platforms were used to collect more information about hypertension-associated SNPs retrieved from text mining and GWAS analyses, including allele frequencies in different ethnic populations, gene annotation, and pathway analysis. We analyzed protein-protein interactions (PPIs) using the STRING database and Cytoscape software, and gene names were used as queries to extract information from gene ontology databases in STRING platform [39] (Figure 1D). Text mining and GWAS results are represented using Circos software [40], R-ggplot2 [41], and GeneSyno [25] according to the human genomic data. Online Clustvis tool was used to perform statistical clustering [42]. The disease gene analysis was conducted using disease genomics of DisGeNET [43].

## 3. Results

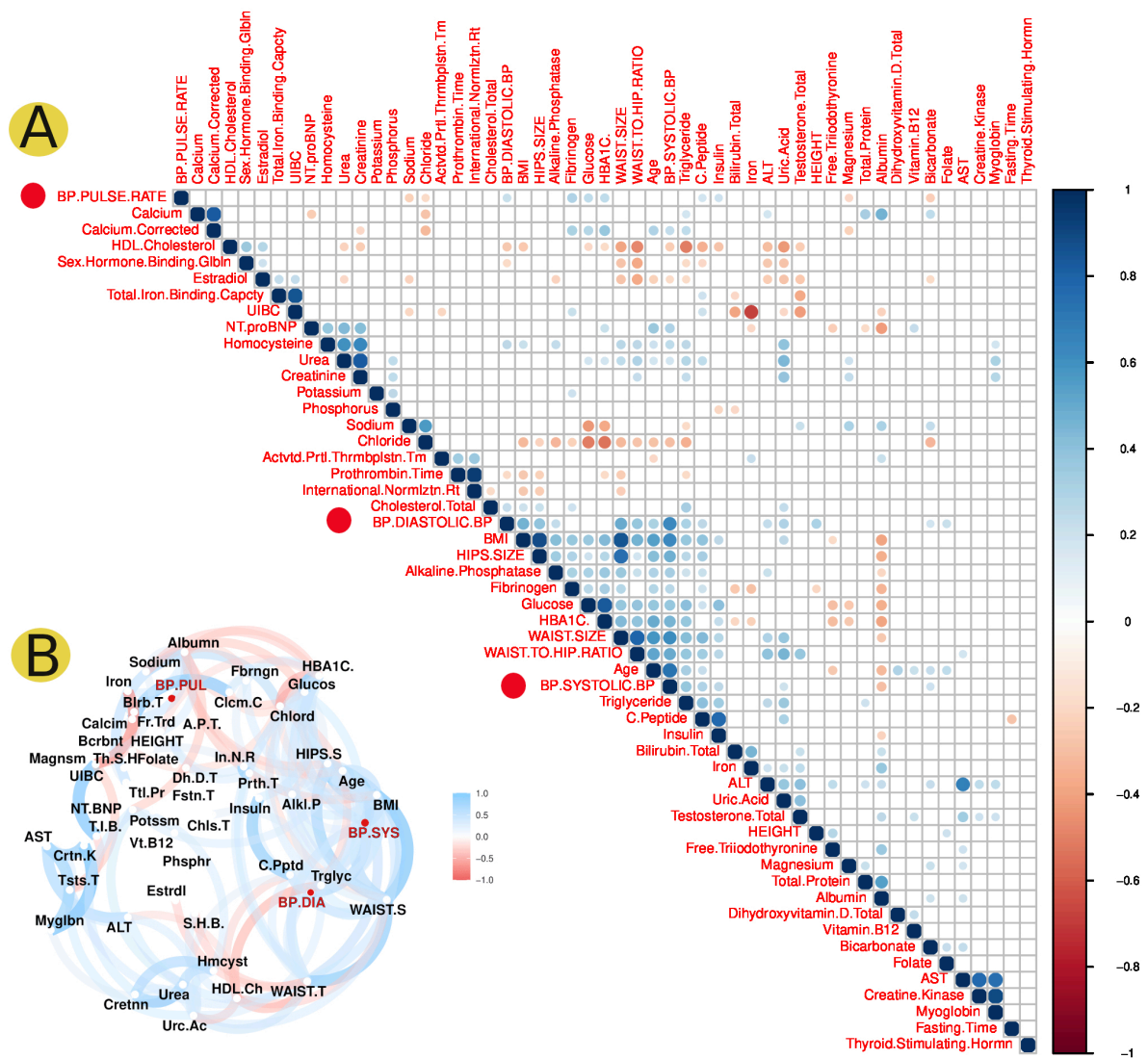
### 3.1. Statistical Analysis

Principle component analysis (PCA) is a common analysis used for sample classification and clustering. It reduces the dimensionality of the investigated dataset, improving interpretability while minimizing information loss. We used this type of analysis to group the studied individuals (100 case and 100 controls) based on several clinical data parameters that did not include the medical diagnosis. The PCA analysis showed that the clinical parameters obtained in this study were useful in highlighting the main characteristics that

distinguish hypertension. The PCA analysis revealed that the patients studied were clustered based on their hypertension diagnosis (Figure 2). The statistical correlation of clinical data revealed a significant correlation between blood pressure factors such as systolic blood pressure, diastolic blood pressure, blood pressure/pulse rate, and several clinical parameters (Figure 3A,B). For instance, systolic and diastolic blood pressure showed a significant positive correlation with body mass index, Alkaline phosphatase, and C peptide, while blood pressure/pulse rate was positively correlated with C peptide, insulin, Fibrinogen, and HbA1C (Figure 3A).



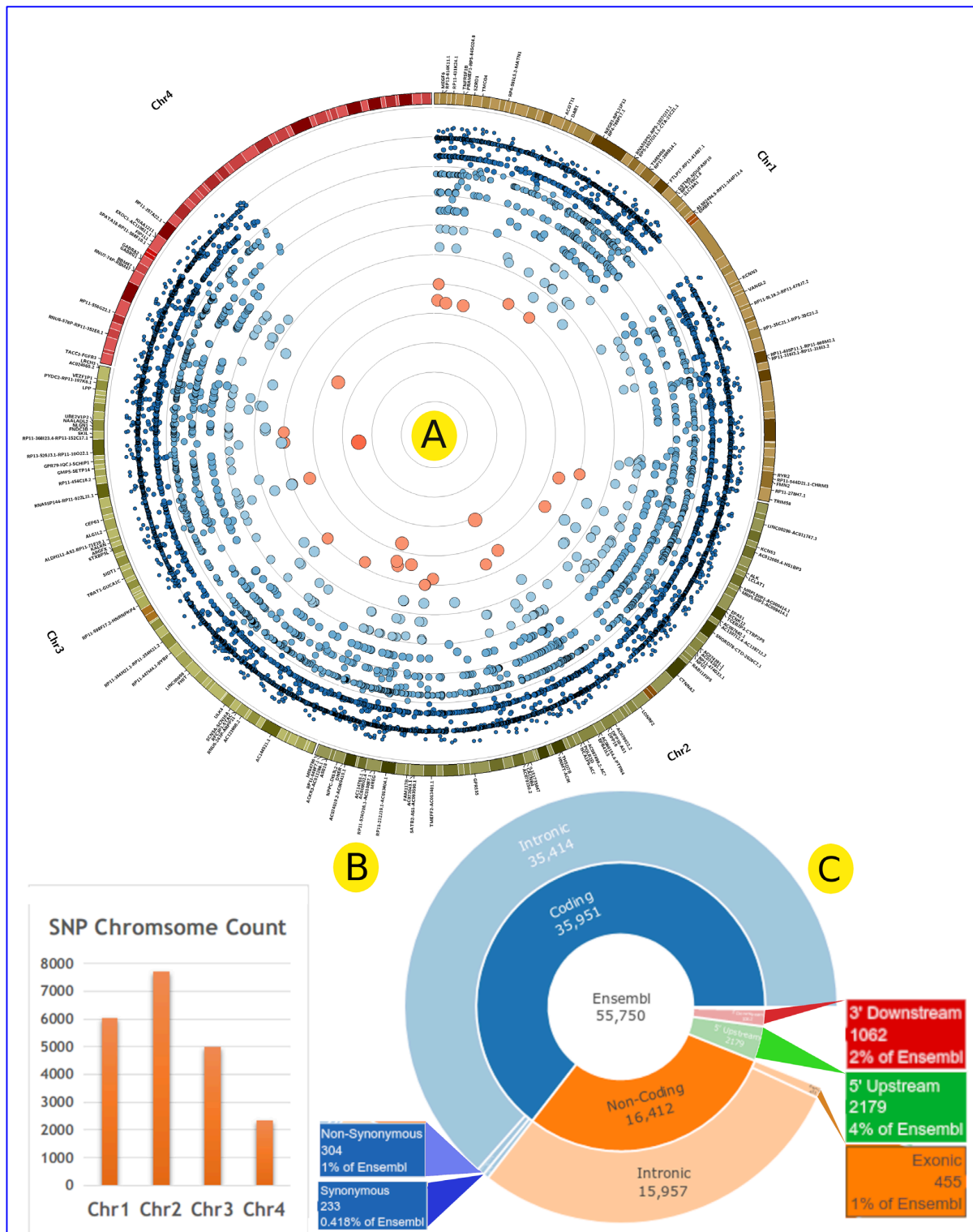
**Figure 2.** The principal component analysis of hypertension data. The PCA analysis of the 200 Qataris clustering based on 53 different clinical parameters, including blood pressure measurements.



**Figure 3.** The statistical correlation analysis of hypertension data. The correlation analysis of different clinical parameters from the 200 Qataris obtained from QBB, with blood pressure parameters highlighted in red circles. (A) Graph of a correlation matrix showing positive and negative correlations with significance indicated by  $p$  values less than 0.01. (B) A correlation network graph displaying the statistical interaction between various clinical parameters, including blood pressure factors in red.

### 3.2. SNPs Associated with EH in Qatari Population

In this study, we investigated the association between hypertension and 21,096 SNPs in 1503 genes across the human genome that are specific to our patient cohort (100 cases vs. 100 normal controls) (Figure 4A). Our analysis revealed hypertension-associated genes in four human chromosomes (Chr 1, 2, 3, and 4) that are specific to our 100 patients. Chr2 had the highest number of SNPs, followed by chromosome Chr1. The SNP density for 1 Mbp was 30, 36, 29, and 28 for Chr1, 2, 3, and 4, respectively (Figure 4A). The estimated functional consequences of the distribution indicate that the used SNPs are mainly located within coding regions (Figure 4C). Most of the SNPs are located in genes (12,913), while 8995 SNPs are upstream or downstream genes with a distance varying from 5 bp to 21 Mbp. Genes such as *DPP10* (119 SNPs), *AC007319.1* (105 SNPs), *RN7SKP61* (101 SNPs), and *RN7SL647P* (92 SNPs) acquired the highest number of SNPs (Supplementary S3).



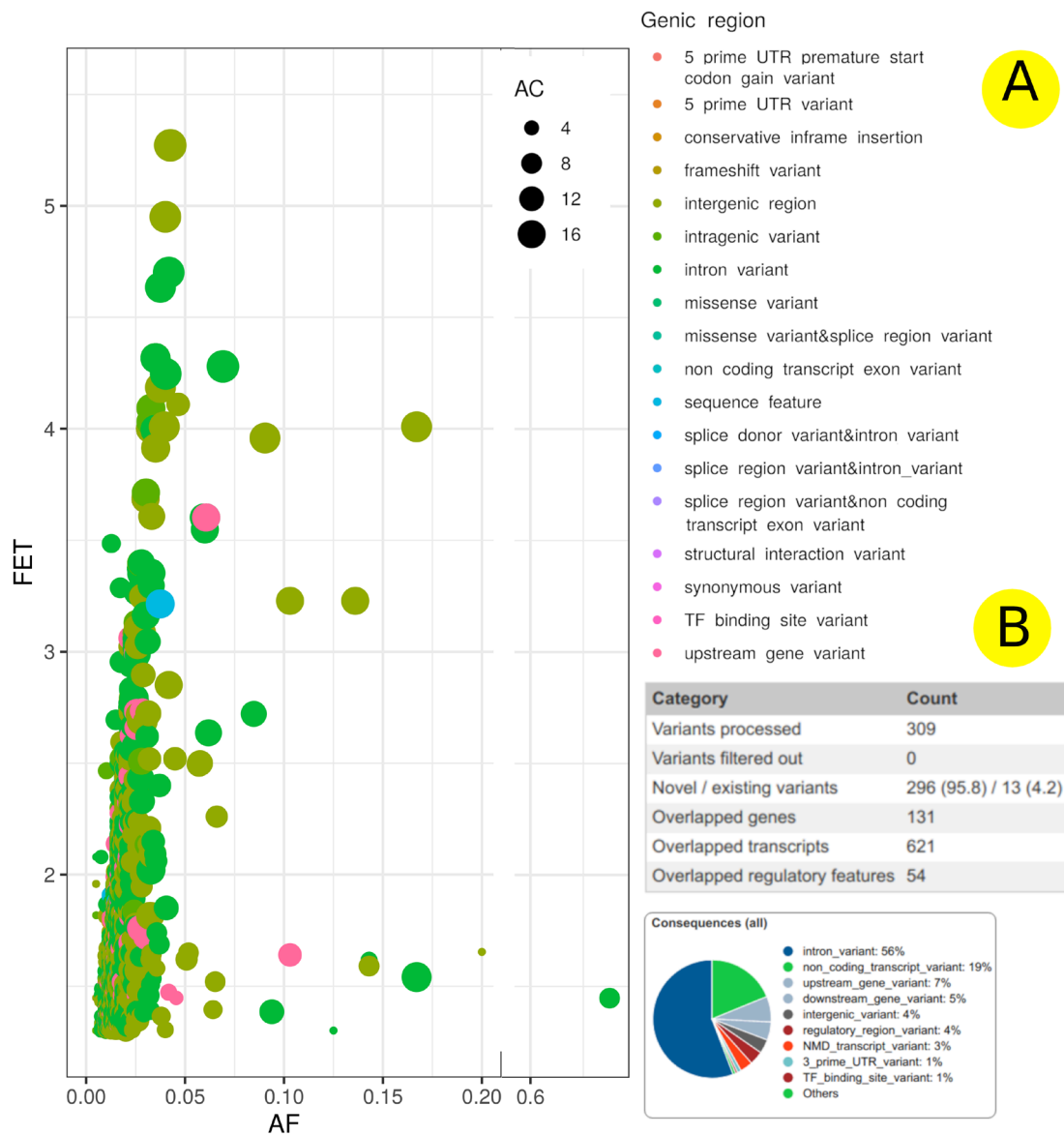
**Figure 4.** The chromosomal distribution of the identified SNPs and their association scores with EH. The colored small circles represent SNPs' location on the human genome and their hypertension association score ( $-\log_{10}$  of  $p$  value) according to the Fisher's exact test. Blue and small circles represent non-significant SNPs, while large and red circles represent significant SNPs (A). The number of SNPs according to human chromosome (B). Predicted functional consequences. SNPs are classified into several groups based on their location in and effect on gene structures, which include coding and non-coding regions according to different databases (C).

The correlations between allele count (AC), Fisher’s exact test (FET), genic region, and allele frequency are plotted in Figure 5. This correlation is important to reveal any potential misinterpretation of the genetic association with EH and visually demonstrate the potential impact of each genetic variation feature on SNP association with EH manifestation. Furthermore, examining the relationship between genetic variation components and EH may reveal previously unknown parameters that could be used in future EH genetic research to increase the number of significant SNPs. Mostly, smaller numbers of SNPs with higher FET have more AC and AF, which could indicate their importance in EH. SNPs with  $FET \geq 4$  are located in the intergenic region (20 SNPs), intron variant (10 SNPs), intragenic variant (5 SNPs), and upstream gene variant (3 SNPs) (Figure 5, Table 1 and Supplementary S4). The FET analysis revealed 336 SNPs with  $\geq 3$  log-*p* value, 38 SNPs with  $\geq 4$  log-*p* value, and 2 SNPs (rs921932379 and rs113688672) with  $\geq 5$  log-*p* value (located within regions of *GMPS-SETP14* and *ISCA1P6-AC012451.1* genes) (Figure 4, Table 1, and Supplementary S4). These SNPs are near/adjacent to 215 genes (Figures 5 and 6 and Supplementary S5), and the SNPs’ distribution across the gene structure can be seen in Figures 5 and 6.

**Table 1.** Top SNP markers associated with EH in studied Qatari population ( $p \leq 1 \times 10^{-5}$ ).

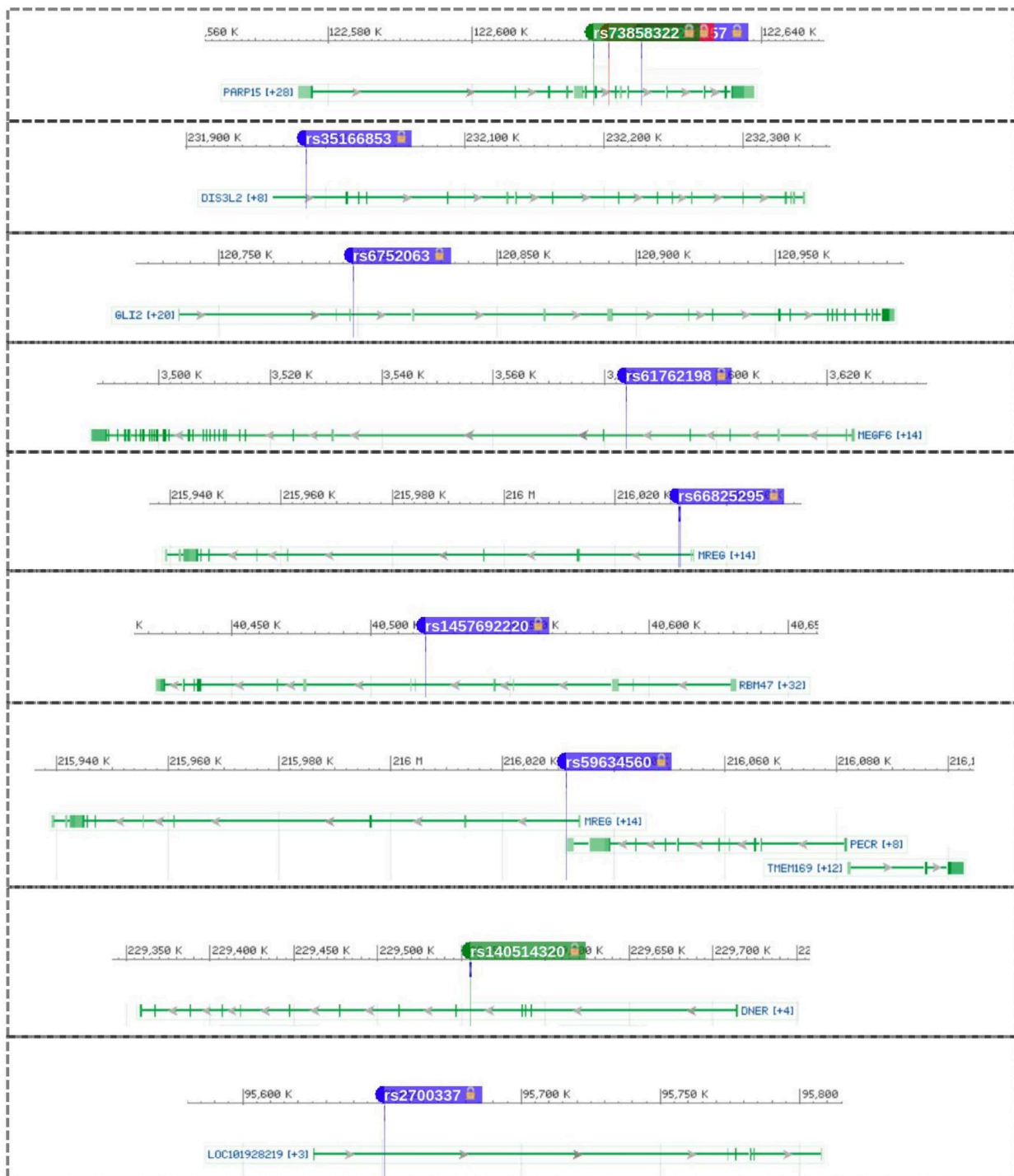
SNP ID	Chr	Position	Ref	Alt	Gene	AC	FET Logpval
rs921932379	3	155674348	T	TA	GMPS-SETP14	17	5.27
rs113688672	2	129289164	C	T	ISCA1P6-AC012451.1	16	4.95
rs_new-95	4	40521676	CTTTTTTTTTTTTT	C	RBM47	16	4.7
rs35166853	2	232850730	C	T	DIS3L2	15	4.63
rs11591086	1	30925979	G	A	RP4-591L5.2-MATN1	14	4.32
rs11893181	2	18313644	G	A	KCNS3	14	4.32
rs1491266756	2	52292323	TTA	T	AC007682.1	14	4.32
rs17199733	2	232824585	G	A	DIS3L2	14	4.32
rs34189801	2	232808692	G	A	NPPC-DIS3L2	14	4.32
rs34553499	2	232820162	C	T	NPPC-DIS3L2	14	4.32
rs60554757	3	122342303	G	A	PARP15	14	4.32
rs652625	1	12225351	T	A	TNFRSF1B	14	4.32
rs66825295	2	216896216	GAGAA	G	MREG	14	4.32
rs6752063	2	121556047	G	A	GLI2	14	4.32
rs73858324	3	122337678	G	A	PARP15	14	4.32
rs61762198	1	3500434	G	GCAGCCACCAGACAACGCA	MEGF6	17	4.28
rs1217727360	2	230419573	A	AAAGAG	DNER	16	4.25
rs369201055	2	138779802	A	AAC	HNMT-AC069394.1	15	4.19
rs_new-56	2	193355414	A	ACAGGGCTGCAGGAAAAAGGG AATGCCTATAGAC	TMEFF2-AC013401.1	8	4.11
rs1345206935	3	18848380	G	GTTTTTTTTTTTTTTTTTTTTTTTTTTT	AC144521.1	13	4.09
rs57679512	3	61057088	GA	G	FHIT	13	4.03
rs7646137	3	162408311	T	A	RP13-526J3.1-RP11-10022.1	15	4.01
rs770631619	1	72759661	C	CTTTTTTTTTTTTT	NEGR1-RPL31P12	15	4.01
rs10931882	2	200917613	T	C	C2orf47-SPATS2L	13	4
rs10931883	2	200917675	T	C	C2orf47-SPATS2L	13	4
rs11291368	3	172047831	AG	A	FNDC3B	13	4
rs11897782	2	200916495	G	A	C2orf47-SPATS2L	13	4
rs11903185	2	200916506	T	C	C2orf47-SPATS2L	13	4
rs12032899	1	3999592	G	C	RP13-614K11.1	13	4
rs12464998	2	200893820	T	C	C2orf47-SPATS2L	13	4
rs12470336	2	201000860	A	T	C2orf47-SPATS2L	13	4
rs2700337	1	96116213	A	G	RP11-286B14.1	13	4
rs4673652	2	200903578	C	T	C2orf47-SPATS2L	13	4
rs73141887	2	8590888	T	G	LINC00299-AC011747.3	13	4
rs73858322	3	122335609	T	A	PARP15	13	4
rs7608345	2	8590943	G	A	LINC00299-AC011747.3	13	4
rs77321003	2	201005171	C	A	C2orf47-SPATS2L	13	4
rs79953652	2	200913138	T	C	C2orf47-SPATS2L	13	4





**Figure 5.** The genetic structure of SNPs associated with EH. (A) The allele frequency vs. FET scores  $-\log_{10}$  of the studied SNP variants. Allele frequency of SNP variants ( $x$ -axis) vs. FET scores  $-\log_{10}$  ( $y$ -axis), the SNP genic region (color) and the maximum allele count in studied population (size). (B) The percentage of SNPs distribution across gene structure.

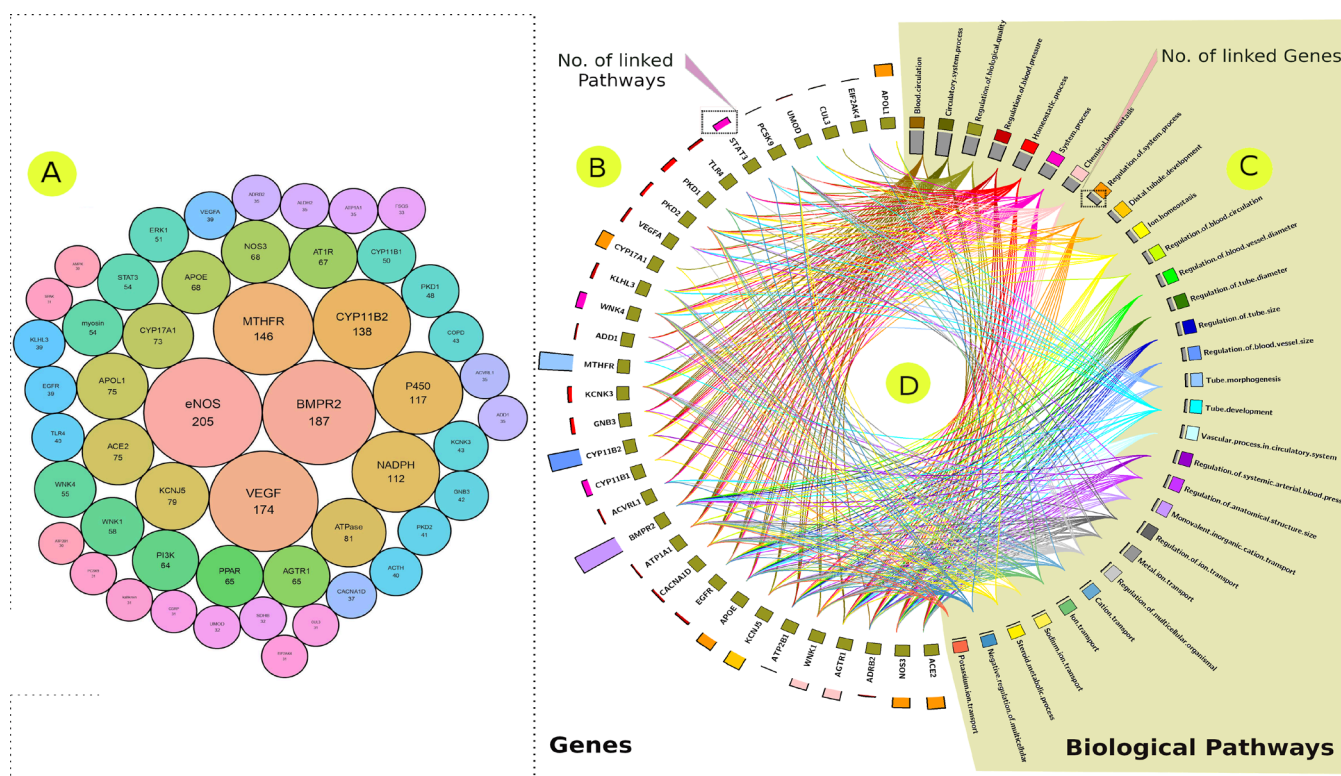
We identified three novel SNPs: rs\_new-95 (4.70  $\log-p$  value), rs\_new-58 (3.06  $\log-p$  value), and rs\_new-56 (4.11  $\log-p$  value). These SNPs are located in Chr4:40521676, Chr2:221398860, and Chr2:193355414, respectively (Table 1 and Supplementary S4). A total of seven genes have ten or more SNPs with  $\geq 3$   $\log-p$  value, including *MAIP1* (23 SNPs), *SPATS2L* (21 SNPs), *ULK4* (17 SNPs), *PKN2-AS1* (14 SNPs), *AC092966.1* (11 SNPs), *FHIT* (10 SNPs), and *RNA5SP52* (10 SNPs) genes (Figure 6 and Supplementary S3). The PPI analysis for the genes where significant SNPs are located demonstrated interaction activity across the hypertension-associated proteins. On the other hand, it shows high significance of functional analysis for genes related to ion-gated channel activity and cardiac conduction (Figure 7, Table 1).



**Figure 6.** Graphical representation of the location of some SNPs linked to hypertension. The physical location of several SNPs discovered to be associated with hypertension using GWAS analysis.







**Figure 9.** The most frequently mentioned genes in the hypertension literature (A,B). The size of the circle, as well as the numbers included, represent the number of articles that mentioned a genetic link between these genes and EH disorder (A). The most frequently mentioned genes in the hypertension literature and their corresponding biological processes according to the enrichment analysis (C). The histogram depicts the number of hypertension publications that listed these genes (A,B). Biological processes were allocated to hypertension-associated genes (C), with gray bars representing the number of genes found within these processes and colored connections representing genes that contribute to these processes. A colored link is shown between the biological process and the corresponding EH-linked genes, with each gene potentially linked to multiple pathways or vice versa (D).

### 3.4. GWAS and Text Mining Results Comparison

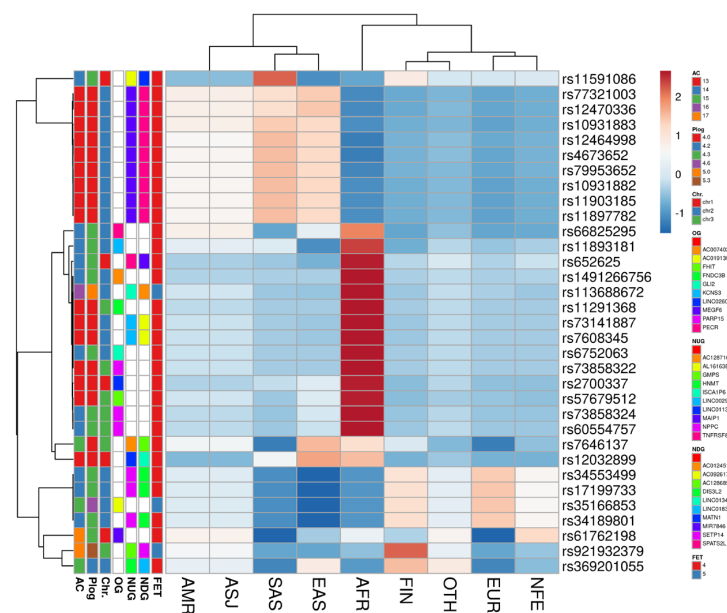
The text mining analysis yielded 3748 genes that are associated with hypertension in the published literature (Supplementary S7). We compared this list of genes with our GWAS hypertension-related genes with SNPs with  $\geq 3$  log-*p* value association scores. We identified 21 genes that are common between the GWASs and text mining that contain 50 SNPs significantly associated with hypertension in our studied cohort (Figure 10 and Supplementary S4 and S8). Genes of *ULK4* and *FHIT* have the highest number of hypertension-associated SNPs for genes mentioned in previous studies (Supplementary S5). *ULK4* and *FHIT* genes have 17 and 10 EH-associated SNPs, respectively, while *FNDC3B*, *FHIT*, *NPPC-DIS3L2*, *GLI2*, and *RBM47* include SNPs that have a  $\geq 4$  log-*p* value association with hypertension in the studied cohort. On the other hand, 194 are unique to our patient cohort; of these, 13 genes that have 26 SNPs are the most significant with  $\geq 4$  log-*p* value. Of these genes, *C2orf47-SPATS2L* contains nine EH-associated SNPs.



**Figure 10.** Comparison between the EH-associated genes identified in the GWAS and text mining analysis.

### 3.5. Allele Frequency of EH-Associated SNPs across Ethnic Groups

To study the frequency of the SNPs identified through our GWAS with  $\geq 3$  log-*p* value in different ethnic populations, we screened our associated SNPs in EUR, SAS, AFR, AMR, ASJ, EAS, FIN, NFE, and OTH populations (Figure 11 and Supplementary S9). The clustering analysis of SNPs' frequency across different ethnic groups demonstrated high frequency for the minor alleles, especially in AFR, EAS, and SAS populations. For instance, rs1004840 shows a high-frequency rate of 0.24 for the minor allele (T) in the AFR population compared to the major allele (C:0.76) (Supplementary S9). Additionally, the clustering analysis using EH-associated SNPs revealed some consistency for SNP clustering depending on the *p* values, AC, overlapped gene (OG), nearest upstream gene (NUG), and nearest downstream gene. Furthermore, a high number of EH-SNPs (rs11591086, rs77321003, rs12470336, rs10931883, rs12464998, rs4673652, rs79953652, rs10931882, rs11903185, and rs11897782) have separated the populations of AMR, ASJ, SAS, and EAS with high frequencies for the major allele (Figure 11). These SNPs are located near the *SPATS2L* (nine SNPs) and *MATN1* (one SNP).



**Figure 11.** The clustering analysis of hypertension-associated SNPs ( $\leq 4$  log-*p* value) across different ethnic groups. The SNP identity number (SNPID), maximum allele count (AC), EH association log-*p* value score, chromosome (Chr), overlapped gene (OG), nearest upstream gene (NUG), and nearest downstream gene (NDG).

## 4. Discussion

Hypertension is a complicated cardiovascular condition that is influenced by various genetic and environmental factors. More basic research and sophisticated tools are needed to understand the missing part of heritability of hypertension. The strength of this study is the combination of a GWAS of 100 hypertensive cases and 100 controls of whole-genome-sequenced individuals and text mining of past reports about the genetic association with hypertension in the past ten years.

We studied 53 different health parameters of both the cases and controls. PCA demonstrated that the clinical parameters obtained in this study were useful in highlighting the main characteristics that distinguish hypertension (Figure 2). PCA demonstrated a good clustering of EH patients and was useful in distinguishing the majority of patients from controls based on the clinical data. A low number of explanations were detected on both axes. On the other hand, this could be attributed to the high complexity of EH and the low precision of clinical parameters alone in detecting EH contributing factors. The correlation analysis of the clinical data revealed a strong association between high blood pressure factors and several clinical parameters (Figure 3). Additionally, clinical parameters related to alkaline phosphatase and C peptide, fibrinogen, HbA1C, iron, homocysteine, and serum ferritin showed a significant association with hypertension (Figure 3).

### 4.1. Genome-Wide Association

The outcome of the GWAS of 100 hypertensive patients and 100 normal controls yielded 21,096 SNPs within 1503 genes that associated with hypertension (Figures 4 and 6). Our study identified two SNPs, rs921932379 and rs113688672, associated with hypertension ( $\geq 5$  log-*p* value), which are located within the intergenic region of *GMPS-SETP14* and *ISCA1P6-AC012451.1* genes, respectively (Table 1, Figures 5 and 6). The association between these loci and hypertension risk has not been reported in the medical literature. *GMPS* (guanine monophosphate synthesis) is engaged in guanine nucleotide de novo synthesis providing GTP, which is involved in cellular processes that are critical for cell division [44]. GTP is the primary source of cyclic guanosine monophosphate (cGMP), which catalyzes a variety of cardio-protective functions [45]. cGMP signaling stimulation is a possible therapeutic strategy for heart failure [46]. The association between the *AC012451.1* gene (a long non-coding RNA/lincRNA) and hypertension may be due to their role in gene regulation [47]. On the other hand, *SETP14* and *ISCA1P6* are pseudogenes, which are non-functional segments of DNA that resemble functional genes. The link between certain pseudogene genes and heart disease has previously been reported in human and model animal studies and is currently an active area of medical research [48–50]. For example, the *HK2P1* pseudogene may contribute to preeclampsia by acting as a competing endogenous RNA for hexokinase 2 and impairing decidualization [48].

We identified three novel SNPs through our GWAS: rs\_new-95 (4.70 log-*p* value), rs\_new-56 (4.11 log-*p* value), and rs\_new-58 (3.06 log-*p* value) (Table 1, Supplementary S4). These are insertion and deletion SNPs found within *RBM47*, *TMEFF2-AC013401.1*, and *AC067956.1*, respectively. The *RBM47* (RNA Binding Motif Protein 47) gene is an RNA-binding protein that controls cell fate decisions and has been suggested to be a tumor suppressor [51]. Rare putative functional hypertension-associated variants in the *RBM47* gene were discovered [52]. These reported rare SNPs could promote hypertension under some environmental factors [53,54]. *TMEFF2* (Transmembrane Protein with EGF-Like and Two Follistatin-Like Domains 2) has two follistatin-like domains that interact directly with TGF- $\beta$  and thus regulate associated growth factor signaling and, as a result, blood pressure regulation. *TMEFF2* has been linked to hypertension and cardiac hypertrophy, according to previous research [55,56].

Seven genes have more than 10 SNPs with  $\geq 3$  log-*p* value (Supplementary S2). Among these genes, Spermatogenesis-Associated Serine-Rich 2 (*SPATS2L*) revealed 21 hypertension-associated SNPs, of which 9 SNPs were  $\geq 4$  log-*p* value (Table 1). *SPATS2L* has a well-known role in the manifestation of some human disorders such as asthma [57] and hepatocellular

carcinoma [58]. Recently, some reports have defined a clear link between *SPATS2L* and atrial fibrillation, which is an abnormal and sometimes rapid heart rate that occurs when the two upper chambers of the heart have unstable electrical signals [59]. Additionally, the Unc-51-Like Kinase 4 (*ULK4*) gene revealed 17 SNPs with  $\geq 3$  log-*p* value (Figures 4 and 6). Several reports have mentioned the association between the *ULK4* gene and heart disorders, including acute aortic dissections [60] and hypertension [61]. *ULK4* has a well-known role in neuronal growth and tyrosine kinase activity [62]. Among genes with a high number of hypertension-associated SNPs, the PKN2-AS1 lncRNA region revealed 14 SNPs. The Protein Kinase N2 (*PKN2*) gene plays a role in the regulation of cell cycle progression, cell migration, actin cytoskeleton assembly, and tumor cell invasion. Several blood-pressure-associated loci were reported in *PKN2* [63]. Additionally, the Fragile Histidine Triad Diadenosine Triphosphatase (*FHIT*) gene revealed 10 hypertension-associated SNPs with  $\geq 3$  log-*p* value, 1 of which was an SNP (rs57679512) with  $\leq 4$  log-*p* value. *FHIT* was a hot point for cancer research for decades [64]. Recently, *FHIT* has gained great interest as a key factor for pulmonary arterial hypertension, where its reductions were associated with endothelial and smooth muscle cell dysfunction [65].

The PPI analysis revealed interaction activity across the previously hypertension-associated genes; this could be due to their unknown interaction on the cellular level and possible unseen links. On the other hand, the PPI shows high significance of functional analysis for genes related to ion-gated channel activity and cardiac conduction (Figure 7). This is highly accepted as most of genes are correlated with signal transduction activity and muscle proliferation and regulation. The relationship between hypertension and cell voltage-gated ion channels, especially in arteries, is well-known where this cellular process is related to multifactorial mechanisms [66].

#### Disease–Gene Relationship

The disease–gene relationship analysis reveals that cardiovascular disorders, including hypertension, are more closely associated with genetic variants located in a specific several hub genes (Figure 8B,C). These genes are linked to a number of biological pathways that are important in regulating cellular processes, such as signal transduction and cell development. We found that 61 of the detected EH-associated genes are linked to 270 cardiovascular disorders (Figure 8 and Supplementary S6). *SCN5A* was linked to 121 cardiovascular disorders (Figure 5 and Supplementary S6). Several studies have examined the relationship between *SCN5A* and a variety of cardiac disorders [67]. Dilated cardiomyopathy, cardiac conduction disease, and sick sinus syndrome are all cardiac conditions linked to variants in *SCN5A* [68,69]. The alpha subunit of the main cardiac sodium channel Nav1.5 is encoded by *SCN5A*, which regulates cardiac electrophysiological function [69]. Our study revealed that rs74947646 (located in *SCN5A-SCN10A*) is associated with EH (log-*p* value  $\geq 3$ ) (Supplementary S4). Similarly, *RYR2*, which contains rs1391189881 (log-*p* value  $\geq 3$ ), has been linked to 60 different cardiovascular disorders (Supplementary S4 and S6). The *RYR2* gene codes for a protein known as ryanodine receptor 2, which forms channels within cells that transport positively charged calcium ions. These channels are important in the heart contraction process [70]. The identified EH-associated *SCN10A* is known for its influence on cardiac conduction by controlling the activity of sodium channels, which are required for electrical signal transmission in cells [71]. *SCN10A* has been linked to 51 cardiovascular disorders and shares an EH-associated SNP with *SCN5A* according to our GWAS analysis (rs1391189881 with log-*p* value  $\geq 3$ ) (Supplementary S4 and S6). Interestingly, some known *SCN10A* variants have been reported to modulate cardiac *SCN5A* expression, influencing cardiac physiology and factors that predispose individuals to arrhythmia [71]. Furthermore, *EPAS1* has a high number of cardio disorders and is linked to two different EH-SNPs (Supplementary S4). *EPAS1* induces adrenomedullin hormone expression and plays an important role in cardiac myocyte inflammation [72,73]. These results suggest that our identified SNPs associated with EH are likely to be correlated with EH. In addition, we



discovered that the majority of EH-associated genes are strongly linked to cardiovascular disorders via 14 different genes and 16 SNPs (Supplementary S4).

#### 4.2. Text Mining in the Hypertension Literature

The text mining analysis was used to explore the previously reported genes associated with hypertension. The hypertension articles published in the last decade revealed a small number of highly redundant hypertension-associated genes and SNPs. We identified 51 genes; each of them was mentioned at least in 30 articles. In addition, 24 SNPs were captured, each of which was mentioned in at least 10 research articles (Figure 9 and Supplementary S7). This could lend credence to the known complexities of the key factors that govern such a condition [74]. Most of these genes are correlated with blood circulation, while others are related to ion transportation, the metamorphosis of coronary arteries, and blood vessels (Figure 9C). Among these genes, eNOS/NOS3, *BMP2*, *VEGF*, *MTHFR*, *CYP11B2*, *P450*, and *NADPH* were highly redundant (Figure 9B). Because of their well-known and mostly direct link to hypertension molecular mechanisms, these genes are frequently mentioned in the hypertension-related studies.

The PPI network was constructed for the genes identified with text mining to visualize the EH-associated genes. Interaction activity revealed a collection of genes forming a hot spot of biological activity (Figure 7). These highly interactive genes are mostly related to the homeostatic process, nitric-oxide-mediated signal transduction, blood circulation, and regulation of blood pressure. Genes such as *MAPK3*, *NOS3*, *AGTR1*, *DEC1*, *VEGFA*, *GNB3*, *APOE*, *CYP11B2*, *EGFR*, *POMC*, and *STAT3* were highly interactive (Figure 7). Some of these genes are a part of the AGE/RAGE pathway (*NOS3*, *MAPK3*, *STAT3*, and *EGFR*). Several studies have shown a link between AGE–RAGE stress and the pathophysiology of a variety of cardiovascular disorders, including essential and pulmonary hypertension [75]. On the other hand, the *AGTR1* and *CYP11B2* genes are linked to the ACE inhibitor pathway and are used to treat cardiovascular disorders such as hypertension [76]. ACE inhibitors prevent ACE from transforming angiotensin I into angiotensin II, lowering blood pressure by suppressing smooth muscle constriction and decreasing aldosterone release [77]. Additionally, *GNB3*, *VEGFA*, and *APOE* genes function on the cell-size level, where they are mostly linked to artery morphogenesis [78,79].

#### 4.3. Differences and Similarities between the GWAS and Text Mining

Combining text mining and GWAS analysis yielded a set of genes that were shared by both methods of analysis (Figure 10 and Supplementary S8). Most of these genes are correlated with heart muscle activity and heart rate and are associated with cell proliferation and cell migration (*HDAC4*, *EPAS1*, *SCN5*, *FGFR3*, *AKT3*, and *ULK4*). For instance, histone deacetylases (HDACs) and histone acetyltransferases (HATs) regulate a broad range of biological processes by controlling transcription factors' accessibility to the gene promoter by histone acetylation or deacetylation [80]. *HDAC4* is one of these genes that has been connected to cardiovascular disorders, where in cultured cardiomyocytes, it regulates neointimal hyperplasia by inducing the proliferation and migration of vascular smooth muscle cells [81]. *CYP11B2* is one of the most important genes that regulate hypertension manifestation due to its role as a key enzyme in aldosterone biosynthesis and aldosterone synthase. because the renin-angiotensin system is critical in regulating intravascular volume and blood pressure. The text mining analysis highlighted the importance of this gene by reflecting its high frequency of EH articles (Figure 7) and high interaction activity in PPI analysis (Figure 7). The GWAS reported no SNPs near or close to this or similar genes; this could be due to the exclusion criteria used to recruit the Qatari cases for EH, which excluded cases with secondary causes of EH such as kidney disease and diabetes. We excluded these individuals to focus our search on just the genes that are associated with only the EH phenotype without interference from secondary causes.

#### 4.4. Allele Frequency of EH-Associated SNPs across Ethnic Groups

Additionally, we studied the allele frequency of EH-associated SNPs across a range of ethnic groups (Figure 11 and Supplementary S9). The clustering analysis of SNPs' frequency across different ethnic groups shows a high frequency for the minor allele, especially in AFR, EAS, and SAS populations, which are more likely to be risk alleles [82] (Figure 11 and Supplementary S9). The AFR population was distinguished from the other populations, whereas SAS and EAS were grouped together. According to the World Health Organization (WHO), hypertension affects approximately 46% of adults aged 25 and older in Africa, compared to 35 to 40% in other parts of the world [83]. This could explain the high divergence of the AFR population in the EH-associated SNP data compared to other ethnic groups (Supplementary S5). The majority of the EH-associated SNPs that differentiated between ethnic groups are located near the *SPATS2L* gene, which has previously been linked to blood pressure in some ethnic groups such as Koreans [84,85].

## 5. Conclusions

We combined a GWAS for 100 hypertensive patients and 100 normal controls and text mining of over 10,000 research articles relevant to EH-gene association to map unique genes/variants to our patients with hypertension; we mapped 194 unique to our Qatari patients that are significantly associated with hypertension. Of these, 26 SNPs in 13 genes have a  $\log\text{-}p$  value  $\geq 4$ . The most significant region was the intergenic region of *C2orf47-SPATS2L*, which contains nine EH-associated SNPs that are highly related to hypertension, which was not previously reported in the literature. Interestingly, we identified three novel SNPs, rs\_new-95 (4.70  $\log\text{-}p$  value), rs\_new-58, and rs\_new-56, located in the loci *RBM47*, *TMEFF2-AC013401.1*, and *AC067956.1*, respectively. The gene–disease association analysis identified 60 EH-associated genes that are linked to 270 different cardiovascular diseases. The *SCN5A* gene that showed a correlation between rs74947646 and EH with a  $\log\text{-}p$  value of 3 is linked to more than 121 cardiovascular disorders. The *RYR2* gene was linked to 60 different cardiovascular diseases, the majority of which affect the functionality of ventricular hypertrophy. Our research found that rs1391189881, which is located in *RYR2*, has a significant association with EH, with a  $\log\text{-}p$  value of 3. Furthermore, genes of *SCN10A*, *EPAS1*, *NPPC*, *ACKR3*, *CHRM3*, and *HDAC4* were linked to 51, 28, 27, 27, 23, and 23 cardio disorders, respectively. In addition, we identified 21 common genes that are shared between the GWAS and text mining analyses; of these, *FNDC3B*, *FHIT*, *NPPC-DIS3L2*, *GLI2*, and *RBM47* genes are the most highly associated with hypertension. Successfully, our analysis revealed two SNPs with a hypertension association of  $\geq 5 p$  value  $\log\text{-}p$  value (rs921932379 and rs113688672), which are located within the intergenic region of *GMPS-SETP14* and *ISCA1P6-AC012451.1* genes. Allele frequency of our GWAS indicates that our patients' susceptibility to hypertension are more related to Africans and Asians than other ethnic groups. Text mining analysis aided GWAS analysis by providing a comprehensive catalogue of known and previously undiscovered EH-associated genes. The limitation of our study is the low number of WGS samples; therefore, more patients with EH need to be included to increase the statistical power of association with hypertension in our population.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jpm12050722/s1>. Supplementary S1: A summary table comparing the values of control and patient clinical parameters. Supplementary S2: Computer programming scripts used for text mining; Supplementary S3: Genes studied according to their number of EH-associated SNPs; Supplementary S4: Fisher's exact test scores for SNPs' association with EH; Supplementary S5: Genes containing EH-associated SNPs; Supplementary S6: The cardiovascular disorders related to EH-associated SNPs detected in the current study; Supplementary S7: The highest frequently mentioned genes and SNPs in the hypertension literature; Supplementary S8: Text mining and GWAS comparison using Venn analysis; Supplementary S9: The allele frequency of EH-associated SNPs in different ethnic groups according to Ensemble database.

**Author Contributions:** Conceptualization, H.Z.; methodology, H.Z. and H.A.; writing—original draft preparation, H.Z. and A.M.A.; writing—review and editing, H.Z. and A.M.A.; supervision, H.Z.; project administration, H.Z.; formal analysis, H.Z., H.A. and A.M.A.; visualization, H.Z. and A.M.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Qatar university grants QUST-1-CHS-2019-11 and QUST-1-CHS-2019-12.

**Institutional Review Board Statement:** The study was approved through the Expedited category of the QBB Institutional Review Board (QBB IRB) E-2019-QF-QBB- RES-ACC-0159-0082.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data available within the article or its supplementary materials.

**Acknowledgments:** We thank Qatar Biobank and Qatar genome project for their efforts to share with us the data used in this manuscript.

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

## References

- Lim, S.S.; Vos, T.; Flaxman, A.D.; Danaei, G.; Shibuya, K.; Adair-Rohani, H.; AlMazroa, M.A.; Amann, M.; Anderson, H.R.; Andrews, K.G.; et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **2012**, *380*, 2224–2260. [[CrossRef](#)]
- Staessen, J.A.; Wang, J.; Bianchi, G.; Birkenhger, W.H. Essential hypertension. *Lancet* **2003**, *361*, 1629–1641. [[CrossRef](#)]
- Lawes, C.M.M.; Hoorn, S.V.; Rodgers, A.; International Society of Hypertension. Global burden of blood-pressure-related disease, 2001. *Lancet* **2008**, *371*, 1513–1518. [[CrossRef](#)]
- Forouzanfar, M.H.; Liu, P.; Roth, G.A.; Ng, M.; Biryukov, S.; Marczak, L.; Alexander, L.; Estep, K.; Abate, K.H.; Akinyemiju, T.F.; et al. Global burden of hypertension and systolic blood pressure of at least 110 to 115 mm hg, 1990–2015. *JAMA* **2017**, *317*, 165–182. [[CrossRef](#)]
- Yagil, Y.; Yagil, C. The search for the genetic basis of hypertension. *Curr. Opin. Nephrol. Hypertens.* **2005**, *14*, 141–147. [[CrossRef](#)]
- Chobanian, A.V.; Bakris, G.L.; Black, H.R.; Cushman, W.C.; Green, L.A.; Izzo, J.L., Jr.; Jones, D.W.; Materson, B.J.; Oparil, S.; Wright, J.T., Jr.; et al. Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. *Hypertension* **2003**, *42*, 1206–1252. [[CrossRef](#)]
- Singh, M.; Singh, A.K.; Pandey, P.; Chandra, S.; Singh, K.A.; Gambhir, I.S. Molecular genetics of essential hypertension. *Clin. Exp. Hypertens.* **2016**, *38*, 268–277. [[CrossRef](#)]
- Niiranen, T.J.; McCabe, E.L.; Larson, M.G.; Henglin, M.; Lakdawala, N.K.; Vasani, R.S.; Cheng, S. Risk for hypertension crosses generations in the community: A multi-generational cohort study. *Eur. Heart J.* **2017**, *38*, 2300–2308. [[CrossRef](#)]
- Luft, F.C. Twins in cardiovascular genetic research. *Hypertension* **2001**, *37*, 350–356. [[CrossRef](#)]
- Naber, C.K.; Siffert, W. Genetics of human arterial hypertension. *Minerva Med.* **2004**, *95*, 347–356.
- Cowley, A.W. The genetic dissection of essential hypertension. *Nat. Rev. Genet.* **2006**, *7*, 829–840. [[CrossRef](#)]
- Arif, M.; Sadayappan, S.; Becker, R.C.; Martin, L.J.; Urbina, E.M. Epigenetic modification: A regulatory mechanism in essential hypertension. *Hypertens. Res.* **2019**, *42*, 1099–1113. [[CrossRef](#)]
- Ueda, K.; Nishimoto, M.; Hirohama, D.; Ayuzawa, N.; Kawarazaki, W.; Watanabe, A.; Shimosawa, T.; Loffing, J.; Zhang, M.-Z.; Marumo, T.; et al. Renal dysfunction induced by kidney-specific gene deletion of *hsd11b2* as a primary cause of salt-dependent hypertension. *Hypertension* **2017**, *70*, 111–118. [[CrossRef](#)]
- Sogawa, Y.; Nagasu, H.; Itano, S.; Kidokoro, K.; Taniguchi, S.; Takahashi, M.; Kadoya, H.; Satoh, M.; Sasaki, T.; Kashihara, N. The eNOS-NO pathway attenuates kidney dysfunction via suppression of inflammasome activation in aldosterone-induced renal injury model mice. *PLoS ONE* **2018**, *13*, e0203823. [[CrossRef](#)]
- Woo, Y.H.; Li, W.-H. DNA replication timing and selection shape the landscape of nucleotide variation in cancer genomes. *Nat. Commun.* **2012**, *3*, 1004. [[CrossRef](#)]
- Doris, P.A. Hypertension genetics, single nucleotide polymorphisms, and the common disease: Common variant hypothesis. *Hypertension* **2002**, *39*, 323–331. [[CrossRef](#)]
- Bodily, P.M.; Fujimoto, M.S.; Page, J.T.; Clement, M.J.; Ebbert, M.T.W.; Ridge, P.G. A novel approach for multi-SNP GWAS and its application in Alzheimer’s disease. *BMC Bioinform.* **2016**, *17*, 455–463. [[CrossRef](#)]
- Padmanabhan, S.; Joe, B. Towards precision medicine for hypertension: A review of genomic, epigenomic, and microbiomic effects on blood pressure in experimental rat models and humans. *Physiol. Rev.* **2017**, *97*, 1469–1528. [[CrossRef](#)]
- Giri, A.; Hellwege, J.N.; Keaton, J.M.; Park, J.; Qiu, C.; Warren, H.R.; Torstenson, E.S.; Kovesdy, C.P.; Sun, Y.V.; Wilson, O.D.; et al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat. Genet.* **2019**, *51*, 51–62. [[CrossRef](#)]
- Flister, M.J.; Tsaih, S.-W.; O’Meara, C.C.; Endres, B.; Hoffman, M.J.; Geurts, A.M.; Dwinell, M.R.; Lazar, J.; Jacob, H.J.; Moreno, C. Identifying multiple causative genes at a single GWAS locus. *Genome Res.* **2013**, *23*, 1996–2002. [[CrossRef](#)]

21. Fowdar, J.Y.; Grealy, R.; Lu, Y.; Griffiths, L.R. A genome-wide association study of essential hypertension in an Australian population using a DNA pooling approach. *Mol. Genet. Genom.* **2017**, *292*, 307–324. [[CrossRef](#)]
22. Davies, R. The creation of new knowledge by information retrieval and classification. *J. Doc.* **1989**, *45*, 273–301. [[CrossRef](#)]
23. Bhardwaj, P.; Khosla, P. Review of text mining techniques. *IITM J. Manag. IT* **2017**, *8*, 27–31.
24. Luque, C.; Luna, J.M.; Luque, M.; Ventura, S. An advanced review on text mining in medicine. *WIREs Data Min. Knowl. Discov.* **2019**, *9*, e1302. [[CrossRef](#)]
25. Alsamman, A.M.; Habib, P.T. GeneSyno: Simple tool to extract gene sequence from the human genome despite synonymous gene terms. *Highlights Biosci.* **2019**, *2*. [[CrossRef](#)]
26. Baker, S.; Ali, I.; Silins, I.; Pyysalo, S.; Guo, Y.; Högberg, J.; Stenius, U.; Korhonen, A. Cancer Hallmarks Analytics Tool (CHAT): A text mining approach to organize and evaluate scientific literature on cancer. *Bioinformatics* **2017**, *33*, 3973–3981. [[CrossRef](#)]
27. Hansson, L.K.; Hansen, R.B.; Pletscher-Frankild, S.; Berzins, R.; Hansen, D.H.; Madsen, D.; Christensen, S.B.; Christiansen, M.R.; Boulund, U.; Wolf, X.A.; et al. Semantic text mining in early drug discovery for type 2 diabetes. *PLoS ONE* **2020**, *15*, e0233956. [[CrossRef](#)]
28. Nam, H.J.; Ryu, S. Text-Mining Analyses of News Articles on Schizophrenia. *Korean J. Schizophr. Res.* **2020**, *23*, 58–64. [[CrossRef](#)]
29. Rodzuan, N.A.S.; Kasim, S.; Sithambranathan, M.; Hassan, M.Z. Classification of biomedical literature in hypertension and diabetes. *Int. J. Data Sci.* **2020**, *1*, 114–119. [[CrossRef](#)]
30. Al Kuwari, H.; Al Thani, A.; Al Marri, A.; Al Kaabi, A.; Abderrahim, H.; Afifi, N.; Qafoud, F.; Chan, Q.; Tzoulaki, I.; Downey, P.; et al. The Qatar Biobank: Background and methods. *BMC Public Health* **2015**, *15*, 1–9. [[CrossRef](#)]
31. Zayed, H. The Arab genome: Health and wealth. *Gene* **2016**, *592*, 239–243. [[CrossRef](#)]
32. Zhang, Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Ann. Transl. Med.* **2016**, *4*, 30. [[CrossRef](#)]
33. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [[CrossRef](#)]
34. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w*<sup>1118</sup>; *iso-2*; *iso-3*. *Fly* **2012**, *6*, 80–92. [[CrossRef](#)]
35. Fisher, R.A. On the interpretation of  $\chi^2$  from contingency tables, and the calculation of P. *J. R. Stat. Soc.* **1922**, *85*, 87–94. [[CrossRef](#)]
36. Ge, S.X.; Jung, D.; Yao, R. ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **2020**, *36*, 2628–2629. [[CrossRef](#)]
37. UniProt Consortium. UniProt: A hub for protein information. *Nucleic Acids Res.* **2015**, *43*, D204–D212. [[CrossRef](#)]
38. Ullah, A.Z.D.; Oscanoa, J.; Wang, J.; Nagano, A.; Lemoine, N.R.; Chelala, C. SNPnexus: Assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic Acids Res.* **2018**, *46*, W109–W113. [[CrossRef](#)]
39. von Mering, C.; Huynen, M.; Jaeggi, D.; Schmidt, S.; Bork, P.; Snel, B. STRING: A database of predicted functional associations between proteins. *Nucleic Acids Res.* **2003**, *31*, 258–261. [[CrossRef](#)]
40. Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [[CrossRef](#)]
41. Wickham, H. ggplot2. *WIREs Comput. Stat.* **2011**, *3*, 180–185. [[CrossRef](#)]
42. Metsalu, T.; Vilo, J. ClustVis: A web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res.* **2015**, *43*, W566–W570. [[CrossRef](#)]
43. Piñero, J.; Ramírez-Anguita, J.M.; Saüch-Pitarch, J.; Ronzano, F.; Centeno, E.; Sanz, F.; Furlong, L.I. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* **2020**, *48*, D845–D855. [[CrossRef](#)]
44. Wang, J.; Wu, Y.; Li, Y.; Wang, Y.; Shen, F.; Zhou, J.; Chen, Y. Guanosine monophosphate synthase upregulation mediates cervical cancer progression by inhibiting the apoptosis of cervical cancer cells via the Stat3/P53 pathway. *Int. J. Oncol.* **2021**, *58*, 3. [[CrossRef](#)]
45. Ying, W.; Zhao, D.; Ouyang, P.; Subramanya, V.; Vaidya, D.; Ndumele, C.E.; Guallar, E.; Sharma, K.; Shah, S.J.; Kass, D.A.; et al. Associations between the cyclic guanosine monophosphate pathway and cardiovascular risk factors: MESA. *J. Am. Heart Assoc.* **2019**, *8*, e013149. [[CrossRef](#)]
46. Muraki, Y.; Naito, T.; Tohyama, K.; Shibata, S.; Kuniyeda, K.; Nio, Y.; Hazama, M.; Matsuo, T. Improvement of pulmonary arterial hypertension, inflammatory response, and epithelium injury by dual activation of cAMP/cGMP pathway in a rat model of monocrotaline-induced pulmonary hypertension. *Biosci. Biotechnol. Biochem.* **2019**, *83*, 1000–1010. [[CrossRef](#)]
47. Cheng, G.; He, L.; Zhang, Y. LincRNA-Cox2 promotes pulmonary arterial hypertension by regulating the let-7a-mediated STAT3 signaling pathway. *Mol. Cell. Biochem.* **2020**, *475*, 239–247. [[CrossRef](#)]
48. Lv, H.; Tong, J.; Yang, J.; Lv, S.; Li, W.-P.; Zhang, C.; Chen, Z.-J. Dysregulated pseudogene *HK2P1* may contribute to preeclampsia as a competing endogenous RNA for hexokinase 2 by impairing decidualization. *Hypertension* **2018**, *71*, 648–658. [[CrossRef](#)]
49. Nojima, H.; Sokabe, H. Genes and pseudogenes for calmodulin in the spontaneously hypertensive rat. *J. Hypertens.* **1988**, *6*, S231–S233. [[CrossRef](#)]
50. Qi, Y.; Wang, X.; Li, W.; Chen, D.; Meng, H.; An, S. Pseudogenes in cardiovascular disease. *Front. Mol. Biosci.* **2021**, *7*, 622540. [[CrossRef](#)]

51. Radine, C.; Peters, D.; Reese, A.; Neuwahl, J.; Budach, W.; Jänicke, R.U.; Sohn, D. The RNA-binding protein RBM47 is a novel regulator of cell fate decisions by transcriptionally controlling the p53-p21-axis. *Cell Death Differ.* **2020**, *27*, 1274–1285. [[CrossRef](#)]
52. Russo, A.; Di Gaetano, C.; Cugliari, G.; Matullo, G. Advances in the genetics of hypertension: The effect of rare variants. *Int. J. Mol. Sci.* **2018**, *19*, 688. [[CrossRef](#)]
53. Ji, L.-D.; Tang, N.L.S.; Xu, Z.-F.; Xu, J. Genes regulate blood pressure, but “environments” cause hypertension. *Front. Genet.* **2020**, *11*, 580443. [[CrossRef](#)]
54. Surendran, P.; Drenos, F.; Young, R.; Warren, H.; Cook, J.P.; Manning, A.K.; Grarup, N.; Sim, X.; Barnes, D.R.; Witkowska, K.; et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat. Genet.* **2016**, *48*, 1151–1161. [[CrossRef](#)]
55. Overcash, R.F.; Chappell, V.A.; Green, T.; Geyer, C.B.; Asch, A.S.; Ruiz-Echevarría, M.J. Androgen signaling promotes translation of *TMEFF2* in prostate cancer cells via phosphorylation of the  $\alpha$  subunit of the translation initiation factor 2. *PLoS ONE* **2013**, *8*, e55257. [[CrossRef](#)]
56. Kumarasamy, S.; Gopalakrishnan, K.; Toland, E.J.; Yerga-Woolwine, S.; Farms, P.; Morgan, E.E.; Joe, B. Refined mapping of blood pressure quantitative trait loci using congenic strains developed from two genetically hypertensive rat models. *Hypertens. Res.* **2011**, *34*, 1263–1270. [[CrossRef](#)]
57. Himes, B.E.; Jiang, X.; Hu, R.; Wu, A.C.; Lasky-Su, J.A.; Klanderma, B.J.; Ziniti, J.; Senter-Sylvia, J.; Lima, J.J.; Irvin, C.G.; et al. Genome-wide association analysis in asthma subjects identifies SPATS2L as a novel bronchodilator response gene. *PLoS Genet.* **2012**, *8*, e1002824. [[CrossRef](#)]
58. Min, P.; Li, W.; Zeng, D.; Ma, Y.; Xu, D.; Zheng, W.; Tang, F.; Chen, J.; Shi, J.; Hu, H.; et al. A single nucleotide variant in microRNA-1269a promotes the occurrence and process of hepatocellular carcinoma by targeting to oncogenes SPATS2L and LRP6. *Bull. Cancer* **2017**, *104*, 311–320. [[CrossRef](#)]
59. van Ouwerkerk, A.F.; Bosada, F.M.; van Duijvenboden, K.; Hill, M.C.; Montefiori, L.E.; Scholman, K.T.; Liu, J.; de Vries, A.A.F.; Boukens, B.J.; Ellinor, P.T.; et al. Identification of atrial fibrillation associated genes and functional non-coding variants. *Nat. Commun.* **2019**, *10*, 1–14. [[CrossRef](#)]
60. Guo, D.-C.; Grove, M.L.; Prakash, S.K.; Eriksson, P.; Hostetler, E.M.; LeMaire, S.A.; Body, S.C.; Shalhub, S.; Estrera, A.L.; Safi, H.J.; et al. Genetic variants in *LRP1* and *ULK4* are associated with acute aortic dissections. *Am. J. Hum. Genet.* **2016**, *99*, 762–769. [[CrossRef](#)]
61. Datta, A.S.; Zhang, Y.; Zhang, L.; Biswas, S. Association of rare haplotypes on *ULK4* and *MAP4* genes with hypertension. *BMC Proc.* **2016**, *10*, 363–369. [[CrossRef](#)]
62. Lang, B.; Pu, J.; Hunter, I.; Liu, M.; Martin-Granados, C.; Reilly, T.J.; Gao, G.-D.; Guan, Z.-L.; Li, W.-D.; Shi, Y.-Y.; et al. Recurrent deletions of *ULK4* in schizophrenia: A novel gene crucial for neuritogenesis and neuronal motility. *J. Cell Sci.* **2013**, *127*, 630–640. [[CrossRef](#)]
63. Kraja, A.T.; Cook, J.P.; Warren, H.R.; Surendran, P.; Liu, C.; Evangelou, E.; Manning, A.K.; Grarup, N.; Drenos, F.; Sim, X.; et al. New blood pressure—Associated loci identified in meta-analyses of 475,000 individuals. *Circ. Cardiovasc. Genet.* **2017**, *10*, e001778. [[CrossRef](#)]
64. Croce, C.M.; Sozzi, G.; Huebner, K. Role of *FHIT* in human cancer. *J. Clin. Oncol.* **1999**, *17*, 1618. [[CrossRef](#)]
65. Prosseda, S.D.; Tian, X.; Kuramoto, K.; Boehm, M.; Sudheendra, D.; Miyagawa, K.; Zhang, F.; Solow-Cordero, D.; Saldivar, J.C.; Austin, E.D.; et al. *FHIT*, a novel modifier gene in pulmonary arterial hypertension. *Am. J. Respir. Crit. Care Med.* **2019**, *199*, 83–98. [[CrossRef](#)]
66. Cox, R.H.; Rusch, N.J. New expression profiles of voltage-gated ion channels in arteries exposed to high blood pressure. *Microcirculation* **2002**, *9*, 243–257. [[CrossRef](#)]
67. Glazer, A.M.; Wada, Y.; Li, B.; Muhammad, A.; Kalash, O.R.; O'Neill, M.J.; Shields, T.; Hall, L.; Short, L.; Blair, M.A.; et al. High-throughput reclassification of *SCN5A* variants. *Am. J. Hum. Genet.* **2020**, *107*, 111–123. [[CrossRef](#)]
68. Wilde, A.A.M.; Amin, A.S. Clinical spectrum of *SCN5A* mutations: Long QT syndrome, brugada syndrome, and cardiomyopathy. *JACC Clin. Electrophysiol.* **2018**, *4*, 569–579. [[CrossRef](#)]
69. Li, W.; Yin, L.; Shen, C.; Hu, K.; Ge, J.; Sun, A. *SCN5A* variants: Association with cardiac disorders. *Front. Physiol.* **2018**, *9*, 1372. [[CrossRef](#)]
70. Hamilton, S.; Terentyeva, R.; Martin, B.; Perger, F.; Li, J.; Stepanov, A.; Bonilla, I.M.; Knollmann, B.C.; Radwański, P.B.; Györke, S.; et al. Increased RyR2 activity is exacerbated by calcium leak-induced mitochondrial ROS. *Basic Res. Cardiol.* **2020**, *115*, 38. [[CrossRef](#)]
71. Chambers, J.C.; Zhao, J.; Terracciano, C.M.N.; Bezzina, C.R.; Zhang, W.; Kaba, R.; Navaratnarajah, M.; Lotlikar, A.; Sehmi, J.S.; Kooner, M.K.; et al. Genetic variation in *SCN10A* influences cardiac conduction. *Nat. Genet.* **2010**, *42*, 149–152. [[CrossRef](#)]
72. Tanaka, T.; Akiyama, H.; Kanai, H.; Sato, M.; Takeda, S.; Sekiguchi, K.; Yokoyama, T.; Kurabayashi, M. Endothelial PAS domain protein 1 (EPAS1) induces adrenomedullin gene expression in cardiac myocytes: Role of EPAS1 in an inflammatory response in cardiac myocytes. *J. Mol. Cell. Cardiol.* **2002**, *34*, 739–748. [[CrossRef](#)]
73. Luque, Y.; Lenoir, O.; Bonnin, P.; Hardy, L.; Chipont, A.; Placier, S.; Vandermeersch, S.; Xu-Dubois, Y.-C.; Robin, B.; Lazareth, H.; et al. Endothelial *Epas1* deficiency is sufficient to promote parietal epithelial cell activation and FSGS in experimental hypertension. *J. Am. Soc. Nephrol.* **2017**, *28*, 3563–3578. [[CrossRef](#)]

74. Shih, P.B.; O'Connor, D.T. Hereditary determinants of human hypertension: Strategies in the setting of genetic complexity. *Hypertension* **2008**, *51*, 1456–1464. [[CrossRef](#)]
75. Prasad, K. Age–rage stress in the pathophysiology of pulmonary hypertension and its treatment. *Int. J. Angiol.* **2019**, *28*, 71–79. [[CrossRef](#)]
76. Wald, N.J.; Law, M.R. A strategy to reduce cardiovascular disease by more than 80%. *BMJ* **2003**, *326*, 1419. [[CrossRef](#)]
77. Flaten, H.K.; Monte, A.A. The pharmacogenomic and metabolomic predictors of ace inhibitor and angiotensin ii receptor blocker effectiveness and safety. *Cardiovasc. Drugs Ther.* **2017**, *31*, 471–482. [[CrossRef](#)]
78. Hasty, A.H.; Linton, M.F.; Brandt, S.J.; Babaev, V.R.; Gleaves, L.A.; Fazio, S. Retroviral gene therapy in ApoE-deficient mice: ApoE expression in the artery wall reduces early foam cell lesion formation. *Circulation* **1999**, *99*, 2571–2576. [[CrossRef](#)]
79. Holderfield, M.T.; Hughes, C.C.W. Crosstalk between vascular endothelial growth factor, notch, and transforming growth factor- $\beta$  in vascular morphogenesis. *Circ. Res.* **2008**, *102*, 637–652. [[CrossRef](#)]
80. de Ruijter, A.J.; van Gennip, A.H.; Caron, H.N.; Kemp, S.; van Kuilenburg, A.B. Histone deacetylases (HDACs): Characterization of the classical HDAC family. *Biochem. J.* **2003**, *370*, 737–749. [[CrossRef](#)]
81. Backs, J.; Song, K.; Bezprozvannaya, S.; Chang, S.; Olson, E.N. CaM kinase II selectively signals to histone deacetylase 4 during cardiomyocyte hypertrophy. *J. Clin. Investig.* **2006**, *116*, 1853–1864. [[CrossRef](#)]
82. Kido, T.; Sikora-Wohlfeld, W.; Kawashima, M.; Kikuchi, S.; Kamatani, N.; Patwardhan, A.; Chen, R.; Sirota, M.; Kodama, K.; Hadley, D.; et al. Are minor alleles more likely to be risk alleles? *BMC Med. Genom.* **2018**, *11*, 1–11. [[CrossRef](#)]
83. Dzudie, A.; Rayner, B.; Ojji, D.; Schutte, A.E.; Twagirumukiza, M.; Damasceno, A.; Ba, S.A.; Kane, A.; Kramoh, E.; Kacou, J.B.A.; et al. Roadmap to achieve 25% hypertension control in Africa by 2025. *Glob. Heart* **2018**, *13*, 45–59. [[CrossRef](#)]
84. Hong, K.-W.; Lim, J.-E.; Oh, B. A regulatory SNP in AKAP13 is associated with blood pressure in Koreans. *J. Hum. Genet.* **2011**, *56*, 205–210. [[CrossRef](#)]
85. de las Fuentes, L.; Sung, Y.J.; Schwander, K.L.; Kalathiveetil, S.; Hunt, S.C.; Arnett, D.K.; Rao, D.C. The role of SNP-loop diuretic interactions in hypertension across ethnic groups in HyperGEN. *Front. Genet.* **2013**, *4*, 304. [[CrossRef](#)]