



OPEN

Genome-wide association study on blood pressure traits in the Iranian population suggests *ZBED9* as a new locus for hypertension

Goodarz Kolifarhood^{1,2}, Siamak Sabour¹, Mahdi Akbarzadeh², Bahareh Sedaghati-khayat², Kamran Guity², Saeid Rasekhi Dehkordi², Mahmoud Amiri Roudbar³, Farzad Hadaegh⁴, Fereidoun Azizi⁵ & Maryam S. Daneshpour²✉

High blood pressure is the heritable risk factor for cardiovascular and kidney diseases. Genome-wide association studies (GWAS) on blood pressure traits increase our understanding of its underlying genetic basis. However, a large proportion of GWAS was conducted in Europeans, and some roadblocks deprive other populations to benefit from their results. Iranian population with a high degree of genomic specificity has not been represented in international databases to date, so to fill the gap, we explored the effects of 652,919 genomic variants on Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), and Hypertension (HTN) in 7694 Iranian adults aged 18 and over from Tehran Cardiometabolic Genetic Study (TCGS). We identified consistent signals on *ZBED9* associated with HTN in the genome-wide borderline threshold after adjusting for different sets of environmental predictors. Moreover, strong signals on *ABHD17C* and suggestive signals on *FBN1* were detected for DBP and SBP, respectively, while these signals were not consistent in different GWA analysis. Our finding on *ZBED9* was confirmed for all BP traits by linkage analysis in an independent sample. We found significant associations with similar direction of effects and allele frequency of genetic variants on *ZBED9* with DBP (genome-wide threshold) and HTN (nominal threshold) in GWAS summary data of UK Biobank. Although there is no strong evidence to support the function of *ZBED9* in blood pressure regulation, it provides new insight into the pleiotropic effects of hypertension and other cardiovascular diseases.

High blood pressure as a heritable risk factor with direct and indirect effects on the incidence of cardiovascular diseases (CVD), stroke, and chronic renal failure is the leading cause of morbidity and mortality worldwide^{1,2}. In the analysis of world data, one in three adults is expected to be affected by high blood pressure up to 2025, while macro or microvascular complications are not limited to extreme ranges of systolic and diastolic blood pressures (SBP and DBP)³. Elevated BP is categorized into primary and secondary hypertension based on disease etiology. While the latter is secondary due to other diseases, some medications, and related side effects, the former is known as essential hypertension and is prevalent in all populations^{4,5}.

Genetic and environmental risk factors such as age, obesity, and lifestyle are attributed to the heterogeneity of essential hypertension risk in the general population^{6,7}. Emerging new genotyping technologies has facilitated studies with Genome-Wide Association (GWA) design in non-communicable and infectious diseases to detect

¹Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran. ²Cellular and Molecular Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. ³Department of Animal Science, Safiabad-Dezful Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Dezful, Iran. ⁴Prevention of Metabolic Disorders Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. ⁵Endocrine Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. ✉email: daneshpour@sbmu.ac.ir

risky or preventive genetic variants after adjusting for environmental risk factors. Promising results from these hypothesis-free studies addressing different pathophysiologic pathways and gene targets for future pharmacologic interventions in complex diseases⁸.

The majority of the genomic findings on BP traits show pleiotropic associations. Until recently, these findings were predominantly from Europe-centric studies, have not been translated into understanding of underpinning molecular pathways, new drug development and personalization of HTN treatment in diverse populations with different genetic backgrounds^{9,10}.

From the viewpoint of genomic architectures, different distribution of risk allele frequencies are contributed to the non-generality of findings between ethnicities¹¹. A recent review addressed evidence concerning the origin of non-generality of GWAS findings on BP traits in different ethnic groups. Accordingly, a large proportion of GWAS was conducted in individuals with European ancestry, and it was introduced as a major cause of low reproducibility of findings in other populations¹².

Iran has experienced the epidemiological transition from communicable to non-communicable diseases during the last four decades, and higher prevalence rates of pre-hypertension (47.3%) and hypertension (22.6%) have been reported in both adults and adolescents in recent years^{13,14}. High blood pressure is main risk factor for CVD events in the Iranian population and its prevalence among CVD cases aged over 18 years is between 17.3% to 20%^{15,16}. In comparing super populations in 1KG project including, Europeans, East Asians, South Asians, Africans and Hispanics, a large number of novel variants with a higher degree of genomic specificity were identified in the Iranian population¹⁷. Heritability estimations on BP traits provide evidence to support role of genetic variants are playing on BP regulation in this population (h^2_{SBP} : 25%, h^2_{DBP} : 26%)¹⁸. However, the Iranian population has not been represented in international genomic databases to date, so there is no evidence concerning the effect(s) of genetic variants in this population. To fill the gap, we conducted the first study using GWA design to investigate the replication or discovery of genetic variants on the Iranian population's BP traits.

Material and methods

Study subjects. The Tehran Cardiometabolic Genetic Study (TCGS) is a family-based genetic analysis of the Tehran Lipid and Glucose Study (TLGS) as the oldest Iranian cohort. In total, more than 20,000 participants are followed in 7 phases since 1999 and underwent a clinical examination on more than 230 metabolic-related traits in each phase of the study. The Medical Ethics Committee of Shahid Beheshti University of Medical Sciences approved this study and all methods were performed in accordance with the relevant guidelines and regulations. All participants gave written informed consent to participate in the original cohort and TCGS. In the case of younger participants, written formal consent was obtained from the parents or guardians. Principles for clinical investigations could be found in the original TCGS and TLGS papers^{19,20}.

In the present study, 17,462 participants from the five phases of TCGS (1999–2014) were recruited to investigate hypertension. More details on DNA sample collection, genotype quality control process, phenotype measurements, covariate imputation, case definition, and selection criteria are explained in the supplementary information 1.

Study design. Based on critical points of BP changes in the age trajectory, all study subjects aged 18 or above were included in the present study, and average values of SBP and DBP during follow up visits were considered in GWA analysis. For binary trait analysis, HTN incident cases and a random sample of healthy individuals with two or more follow-up records were included in the analysis (Supplementary file, Figure S1, S2). Age, sex, Body Mass Index (BMI), Waist Circumference (WC), insulin resistance and top 5 PCs were included in both GWA analysis on quantitative and binary traits after imputing their missing values, using the Expectation-Maximization method with Bootstrapping (EMB) approach by Amelia package in R²¹ (Supplementary file, Figure S3).

Quality control of genotypes. To maximize power against the removal of individuals and markers, quality control (QC) of genotyping data was implemented on a per-individual basis before per-marker using PLINK version 1.9 and R^{22,23}. Accordingly, a standard QC pipeline on 652,919 SNP, with an average mean distance of 4 kilobases, were performed in 7694 adults after excluding Individuals with missing phenotype, discordance of genetically inferred sex versus self-report, genotype calls $\leq 10\%$, and high heterozygosity ($F_{\text{statistics}} \pm 3$ standard deviation). Moreover, related subjects with Identity By Decent (IBD) $\geq 18.5\%$ were excluded from the study. In the genotype level, variants with minor allele frequency (MAF) $< 1\%$, missing genotype calls $> 5\%$, and Hardy Weinberg Equilibrium (HWE) P value $< 1e-6$ for quantitative traits, P value $< 1e-10$ in the HTN cases and P value $< 1e-6$ in controls were filtered out (Supplementary file, Figure S4).

Statistical analysis. After checking collinearity for all covariates ($r^2 > 0.8$), linear and logistic regression tests were performed. To explore any associations between BP traits and genetic variants, the covariates and top 5 Principal Components(PC) were adjusted and additive and overdominant inheritance models on autosomal chromosomes were checked for GWAS on the quantitative and binary traits in PLINK v1.9, respectively. A conventional genome-wide threshold of 5×10^{-8} considered for a significant P value. The genomic inflation factor was computed for each analysis, and observed versus expected P values were highlighted in the Q-Q plots to check for population stratification. Finally, four regression-based multivariate analyses evaluated the predictive accuracy of initial GWAS outputs for quantitative traits on discovery dataset and also an independent sample of TCGS ($N = 2799$)²⁴. Accordingly, Polygenic Risk Score (PRS) was calculated after adjusting for the covariates and top 5 PCs, to evaluate proportion of variance, which is explained by genomic variants (R^2). We used a standard pipeline to evaluate R^2 through calculating PRS on our GWAS summary data. Moreover, discrimination of best

Variable	Discovery		Confirmation
	SBP/DBP	HTN	SBP/DBP/HTN
	Adults (N = 4657)	Adults (N = 4214)	TCGS families (N = 210) Fam. Members (N = 1618)
Age (years)	40.8 ± 15.9	40.0 ± 13.5	35.6 ± 19.4
Female (%)	2523 (54%)	2143 (51%)	834 (52%)
SBP (mmHg)	116.29 ± 17.61	112.61 ± 12.29	117.61 ± 18.94
DBP (mmHg)	76.21 ± 9.72	74.70 ± 7.62	77.36 ± 10.49
BMI (kg/m ²)	27.07 ± 4.51	27.01 ± 4.28	24.45 ± 6.90
WC (cm)	90.82 ± 11.51	90.60 ± 10.72	86.06 ± 15.77
TG/HDL	3.81 ± 3.11	3.87 ± 2.08	–
HTN Case (%)	NA	970 (23%)	855 (53%)

Table 1. Descriptive characteristics of phenotypes and covariates in GWAS and confirmation datasets. Mean ± standard deviation, NA not applicable, *Fam. Members* family members.

fit PRS was assessed for each analysis according to sex. The PRS was computed in the following steps by PLINK and R:

- i. Pruning: A window size of 200 variants, sliding across the genome with step size of 50 variants at a time were considered, and any SNPs with $r^2 > 0.25$ were filter out,
- ii. Clumping: SNPs within 250 k of the index SNP were considered for clumping. Then, all SNPs are correlated with each other were removed ($r^2 > 0.2$) with P value ≤ 1 and only index SNP captured,
- iii. Seven P value thresholds (0.001, 0.05, 0.1, 0.2, 0.3, 0.4, and 0.5) for inclusion of SNPs in the PRS was extracted,
- iv. PRS generated in different ranges of P values for SBP and DBP,
- v. To find "best-fit" score for the quantitative traits, PRS calculated in different ranges of P values thresholds,
- vi. R^2 as an index of trait variance explained by genomic variants calculated in different ranges of P values thresholds for two sets of covariates, (including age, WC and Ins. resistance and top 5 PCs, and also age, BMI and Ins. resistance and top 5 PCs),
- vii. Calculated R^2 was plotted by different ranges of P values using "ggplot2",
- viii. The best-fit PRS was plotted against quantitative traits based on best result of R^2 by sex using "ggplot2".

Confirmation study. The confirmation study was conducted on 1618 participants in 210 selected TCGS families with an age range of 1 to 93 and familial aggregation of two or more affected (HTN) cases²⁵. Similar to initial GWAS, the QC processes were applied in the confirmation study. After removing Mendelian errors and pruning out SNPs in linkage disequilibrium (LD) with a $r^2 > 0.2$, the effects of significant independent SNPs in the initial GWAS were tested in the presence of the same covariate sets using two-level Haseman-Elston regression model by SAGE version 6.4²⁶. This model-free linkage analysis is mathematically equivalent to the likelihood-based score test in variance component linkage analysis, which compute a random effect for genetic variants with T score test.

Post GWAS. Four consecutive steps were followed to explain probable functions and pathophysiologic pathway(s) of discovered variants' effects on BP with P values less than 1×10^{-4} . In the first step, chromosomal coordinates, genes, transcripts, and variants on protein sequence were annotated in Ensemble Variant Effect Predictor²⁷. In the second step, GWAS catalog information was retrieved to identify the association of specified loci on BP traits²⁸. Moreover, we sought to map known BP loci by assessing these loci's functional consequences in Ensemble²⁹. Functional analysis was performed for those variants with consistent results and at least a P value less than 5×10^{-7} after adjusting for different covariate sets. In the case of a similar locus, llink browser by National Cancer Institute and LinDA (LINKage Disequilibrium-based Annotation) browser were checked for LD of detected and previously reported variant(s) in all populations using the website <http://analysistools.nci.nih.gov/LDlink/> and <http://linda.irgb.cnr.it/>, respectively. In the third step, a list of detailed information on all loci was retrieved separately in Open Targets POST GWAS, including disease associations, protein interactions, pathways, similar targets based on diseases in common, RNA, and protein baseline expression by the anatomical system and organ³⁰. In the final step, the overall association score was retrieved for each locus and other genes whose protein products interact with new loci through protein interaction networks^{31,32}. Open Target Platform provides the score from 20 data sources, is ranged from 0 to 1, that the former implies no evidence and later corresponds to the most reliable evidence supporting evidence based on frequency, severity, and significance of association³³.

Results

GWAS datasets. Table 1 describes the phenotypes and covariates characteristics in three datasets of discovery and confirmation studies. After applying DNA samples and genomic markers metrics, 4657 individuals with 616,263 SNPs on the quantitative traits and 4214 individuals with 616,308 SNPs on the HTN trait passed

Trait	Locus nearby	Position	rs ID	Effect/other allele	EAF	Discovery		Confirmation			Reported locus (Ref)
						Adults (+ 18)		TCGS families			
						St. β (95% CI)	P value	T value	P value	Trait	
SBP	<i>FBN1</i>	15:48420560	rs2303505	T/G	0.02	0.37 (0.23–0.50)	7.45E–08 [†]	0.44	0.24	–	Reported ^{34–37}
		15:48428455	rs363830	T/C	0.02	0.37 (0.24–0.51)	7.70 E–08 [†]	NA	NA		
		15:48420679	rs363838	G/T	0.02	0.37 (0.23–0.50)	7.67 E–08 [†]	NA	NA		
DBP	<i>ABHD17C</i>	15:80717774	rs1078107	C/T	0.01	0.10 (0.07–0.14)	1.55 E–08 ^{†*}	0.09	0.17	–	Reported ^{34–36,38–40}
		15:80689205	rs16972291	C/T	0.01	0.10 (0.06–0.13)	1.58 E–08 ^{†*}	NA	NA		
HTN	<i>ZBED9</i>	6:28574647	rs450630	A/G	0.43	0.64 (0.54–0.76) [‡]	3.58 E–07 [‡]	4.03	0.02	SBP,DBP,HTN	Reported ⁴¹
		6:28611694	rs9501180	T/C	0.43	0.64 (0.54–0.76) [‡]	4.13 E–07 [‡]	NA	NA		
		6:28602172	rs380914	T/C	0.43	0.65 (0.55–0.77) [‡]	9.41 E–07 [‡]	NA	NA		
		6:28554918	rs6456825	G/A	0.43	0.64 (0.54–0.76) [‡]	3.24 E–07 [‡]	NA	NA		
		6:28682576	rs9885928	A/G	0.43	0.64 (0.54–0.76) [‡]	3.76 E–07 [‡]	NA	NA		

Table 2. GWA findings on BP traits in discovery and confirmation studies. *EAF* effect allele frequency in TCGS, *St. β* standardized effect estimate, *CI* confidence interval. *NA* not applicable due to removing variant in LD pruning process ($r^2 > 0.2$). [†]Reach the genome-wide borderline *P* value threshold after adjustment for WC. ^{*}Reach genome-wide threshold after adjustment for age and sex. [‡]Reach the genome-wide borderline *P* value threshold after adjustment for BMI/WC. [#]Odds ratio.

genotype QC. The discovery sample had a larger proportion of normotensive individuals, whereas the proportion of affected and unaffected cases was balanced in the confirmation study. Further, there was similar distributions of covariates in discovery and confirmation datasets, while high collinearity between BMI and WC was highlighted; hence, these effects were adjusted in three different sets of covariates including, (1) age, BMI, Ins. resistance and top 5 PCs, (2) age, WC, Ins. resistance and top 5 PCs, and (3) age, sex and top 5 PCs (Supplementary file, Figures S5, S6).

GWAS result. Table 2 describes GWAS findings in the discovery and confirmation studies. In the genome-wide analysis of quantitative traits, two signals on *ABHD17C* detected for DBP corresponding to the genome-wide significance threshold. Moreover, suggestive signals within 6p22.1 (*ZBED9*) and *FBN1* with borderline *P* values detected with the HTN and SBP traits, respectively (Fig. 1). The detected variants on *ZBED9* and *ABHD17C* associated with HTN and DBP by two different sets of covariates, respectively. Moreover, *P* values for detected SNPs on *ZBED9* were slightly attenuated after adjusting for age, sex and top 5 PCs (Supplementary file, Table S1). The test statistic of genomic inflation was low for both analyses on quantitative and binary traits (λ : 1.001–1.01), and there was no population stratification (Fig. 2).

In the confirmation analysis for detected SNPs in initial GWAS, the association of rs450630 on *ZBED9* was replicated for SBP, DBP, and HTN traits after LD pruning on other variants and adjusting for three covariate sets (*P* value < 0.05). The associations of other genetic variants on *ZBED9* with QRS amplitude and interval in Europeans and also *ABHD17C* and *FBN1* with BP traits or cardiovascular diseases in Europeans and East Asians were previously reported in GWAS literature. However, there is no correlation (LD) between detected and previously reported genetic variants in the same locus (Supplementary file, Figure S7).

In an inquiry of GWAS summary data on blood pressure related phenotypes in UK Biobank and also Japanese population, we found out similar associations (direction of effect and allele frequency) of detected variants on *ZBED9* with DBP (*P* value $\leq 5 \times 10^{-8}$) and HTN (*P* value $\leq 1 \times 10^{-4}$) in UK Biobank, while the associations were not significant (even in nominal threshold of 5×10^{-2}) in Japanese population (Table 3). We checked out for LD of associated variants in 6p22.1 by LinDA. However, we could not find any reports in previous studies for association between correlated variants ($r^2 > 0.4$) and BP traits in the region. Except for Africans, allele frequencies of detected variants on *ZBED9* are similar in Europeans, South Asians, Middle East (Qatari) and Iranian population, while the allele frequencies of detected variants on *FBN1* and *ABHD17C* is less than 1% in Europeans (Supplementary file, Table S2).

Polygenic risk score estimation. Multivariate analysis of PRS on initial GWAS outputs in the discovery dataset showed evidence of correlations between calculated PRS with SBP and DBP in three out of four models. The genomic variants account for 4.5–12.7% of the total variance of quantitative traits. Moreover, a similar distribution of PRS-trait correlation was seen by sex for each statistical model in discrimination analysis (Fig. 3). However, the calculated R^2 in the independent sample of TCGS was much lower compared to discovery dataset ($\leq 1\%$) and there was low correlation between the PRS and BP traits by sex (Fig. 4).

Variant effects on protein-coding sequence. In the locus-based regulatory annotation, detected signals on *ABHD17C* and *ZBED9* were annotated as an intronic variant. All detected SNPs on the *FBN1* gene were in relatively high LD, while rs363830 and rs363838 were missense and splice variants. Also, two variants on *ABHD17C* were in moderate LD (r^2 : 0.4–0.6). In further investigation via sentinel variants, we found rs853684

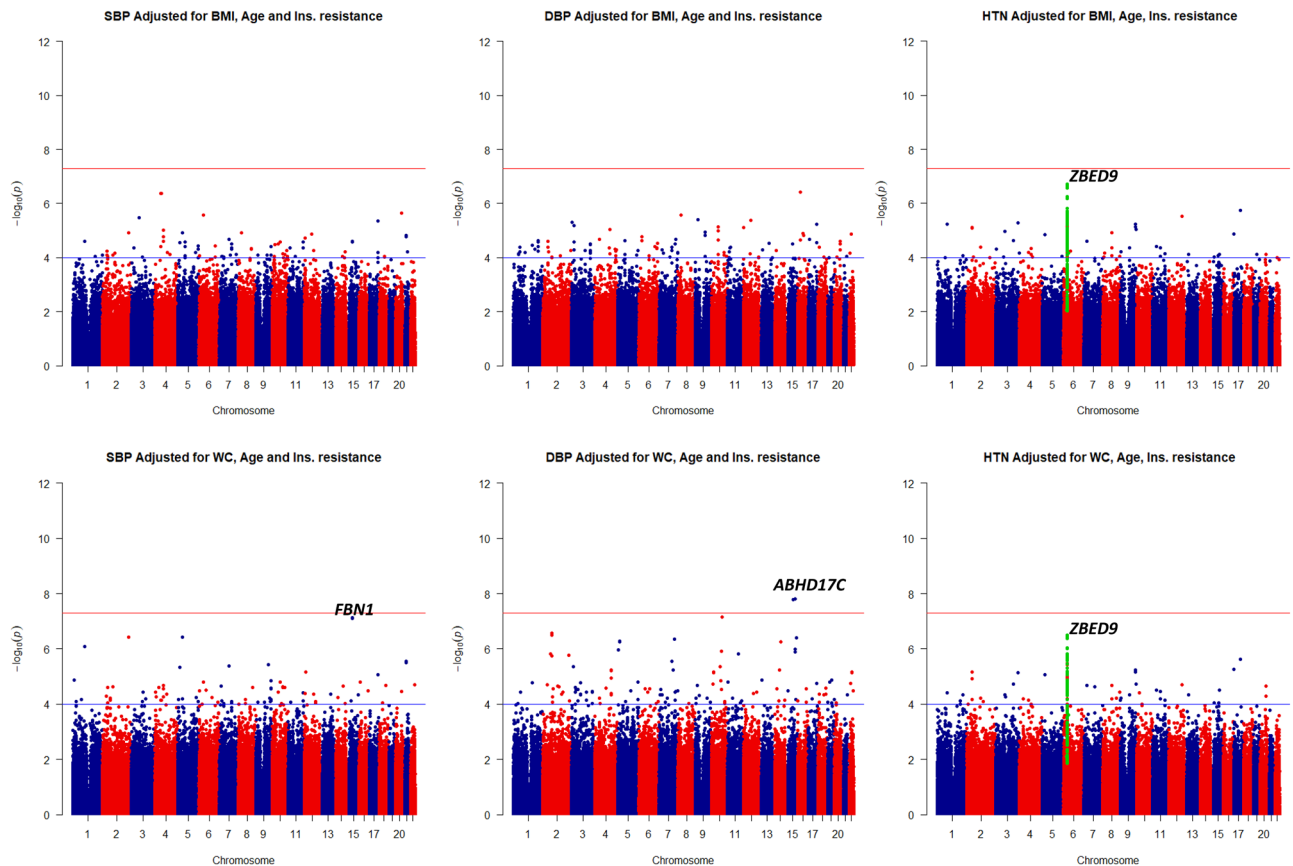


Figure 1. Manhattan plots of BP quantitative and binary traits, adjusted for two covariate sets of BMI and WC.

on *ZSCAN31* as a missense consequence in moderate LD ($r^2 > 0.60$) and five missense variants in perfect LD ($r^2 = 1.0$) with detected variants on *ZBED9* (Fig. 5).

Prioritization of targets by scoring target-disease association. The cumulative evidence for association with BP traits for three detected loci was retrieved in the Open Targets POST GWAS. The overall association scores were summarized in Fig. 6, with varying blue shades: the darker the blue, the stronger the association. Two previously reported loci of *FBN1* and *ABHD17C* to show evidence for a strong association with BP and cardiovascular disease traits, so they were prioritized as biological targets (overall association score = 1). Further, the protein interaction network suggests gene subnetworks for *ZBED9* with relatively low association with BP and cardiovascular diseases (overall association score ≤ 0.40), respectively.

Discussion

To our knowledge, this is the first GWAS to examine genetic associations with BP traits in the Iranian population. This study identified suggestive and strong signals on *ZBED9*, *FBN1*, and *ABHD17C*, associated with HTN, SBP, and DBP, respectively. We found significant associations with similar direction of effect and allele frequency of detected variants on *ZBED9* with DBP (genome-wide threshold) and HTN (nominal threshold) in UK Biobank. However, There is no LD between detected and previously reported genetic variants on three loci in Europeans and other populations including, South Asian, Hispanic or Latin American, East Asian, African American, and Afro-Caribbean populations. The detected signals were not consistent after adjusting for two different sets of covariates on *FBN1* and *ABHD17C*. On the other hand, the association of genetic variants on *ZBED9* and BP traits has not been previously reported in other populations, while their effects were consistent after controlling for environmental factors by two different GWA analysis. Further, the confirmed association of genetic variants on *ZBED9* in discovery study with SBP, DBP, and HTN in an independent sample of TCGS families substantially increase our knowledge of its effects on HTN in the Iranian population.

In an effort to explain the function of new loci by RNA and protein baseline expression, the rate of expression for all detected loci was in the range of low to moderate in the circulatory system, including blood, endothelial cells of the umbilical vein, aorta, coronary artery, left heart ventricle, tibial artery, atrium auricular region, and heart muscle, provided by Human Protein Atlas and Expression Atlas³⁰.

There is no evidence to support the functional effects of *ZBED9* on BP regulation in the literature. However, GeneHancer-gene association results indicate some known BP loci, including *EBF1*, *NR2F2*, *SOX5*, and *PRDM6*, are likely acting as transcription factor binding sites or gene targets for *ZBED9*^{34–37,42–46}. In this way, significant

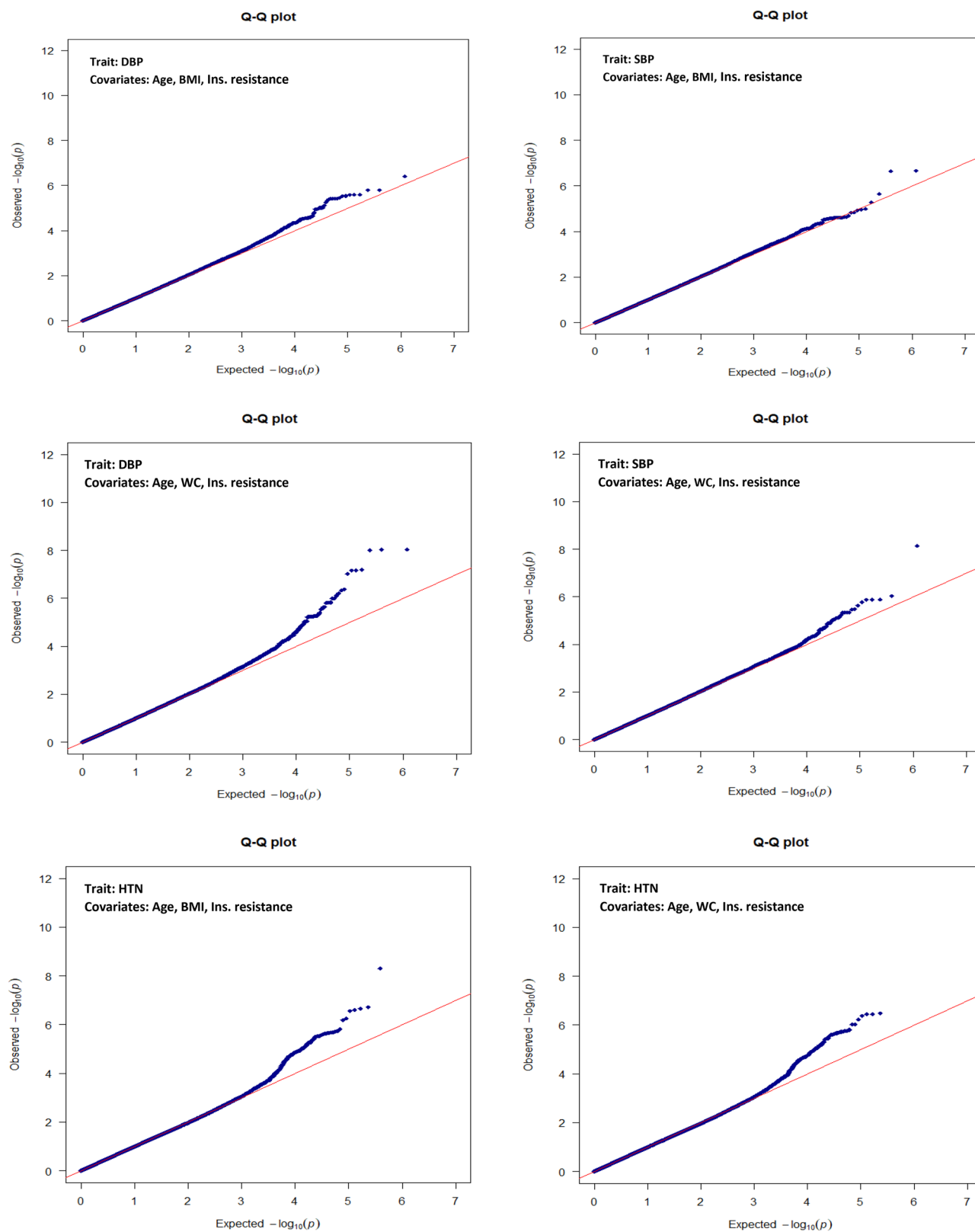


Figure 2. Q–Q plots of observed versus expected P values for BP quantitative and binary traits.

Locus nearby	Position	rs ID	Effect/ other allele	UKbiobank						Japanese pop					
				EAF	HTN		DBP		SBP		EAF	DBP		SBP	
					β (se)	P value	β (se)	P value	β (se)	P value		β (se)	P value	β (se)	P value
FBN1	15:48420560	rs2303505	T/G	0.008	7.2E-03 (5.5E-03)	1.9E-01	1.5E-02 (1.3E-02)	2.4E-01	2.3E-03 (1.2E-02)	8.4E-01	0.13	- 2.9E-03 (5.3E-03)	6.0E-01	- 6.3E-03 (5.6E-03)	2.6E-01
	15:48428455	rs363830	T/C	0.017	8.8E-03 (3.8E-03)	2.1E-02	2.3E-03 (9.2E-03)	7.9E-01	- 1.8E-04 (8.7E-03)	9.8E-01	0.13	3.0E-03 (5.6E-03)	5.9E-01	- 1.8E-04 (8.7E-03)	9.8E-01
	15:48420679	rs363838	G/T	0.02	7.0E-03 (5.5E-03)	2.0E-01	1.5E-02 (1.3E-02)	2.5E-01	2.6E-03 (1.2E-02)	8.3E-01	0.13	- 2.9E-03 (5.6E-03)	6.0E-01	- 6.3E-03 (5.6E-03)	2.6E-01
ABHD17C	15:80717774	rs1078107	C/T	0.003	- 1.3E-02 (8.4E-03)	1.0E-01	- 4.2E-02 (2.0E-02)	3.2E-02	- 2.6E-02 (1.9E-02)	1.6E-01	0.07	- 9.5E-02 (7.2E-03)	1.9E-01	- 4.7E-03 (7.2E-03)	5.1E-01
	15:80689205	rs16972291	C/T	0.003	- 1.2E-02 (8.7E-03)	1.6E-01	- 3.9E-02 (2.0E-02)	5.6E-02	- 2.5E-02 (1.9E-02)	2.0E-01	0.04	- 1.4E-02 (8.8E-03)	1.0E-01	3.3E-03 (8.8E-03)	7.0E-01
ZBED9	6:28574647	rs450630	A/G	0.46	- 3.6E-03 (1.0E-03)	2.5E-04	- 1.5E-02 (2.3E-03)	2.4E-10	6.2E-05 (2.2E-03)	9.7E-01	0.33	1.7E-03 (4.0E-03)	6.8E-01	5.0E-03 (4.0E-03)	2.1E-01
	6:28611694	rs9501180	T/C	0.47	3.9E-03 (1.0E-03)	9.5E-05	1.5E-02 (2.3E-03)	2.3E-10	- 1.3E-04 (2.2E-03)	9.5E-01	0.33	1.9E-03 (4.0E-03)	6.3E-01	5.0E-03 (4.0E-03)	2.1E-01
	6:28602172	rs380914	T/C	0.47	- 3.7E-03 (1.0E-03)	1.6E-04	- 1.4E-02 (2.3E-03)	3.4E-10	5.2E-04 (2.2E-03)	8.1E-01	0.33	1.9E-03 (4.0E-03)	6.4E-01	4.9E-03 (4.0E-03)	2.2E-01
	6:28554918	rs6456825	G/A	0.46	- 3.8E-03 (1.0E-03)	1.1E-04	- 1.5E-02 (2.3E-03)	1.3E-10	- 8.9E-05 (2.2E-03)	9.6E-01	0.33	1.8E-03 (4.0E-03)	6.5E-01	5.1E-03 (4.0E-03)	2.0E-01
	6:28682576	rs9885928	A/G	0.49	- 3.6E-03 (1.0E-03)	3.3E-04	- 1.3E-02 (2.3E-03)	2.3E-08	- 1.2E-03 (2.2E-03)	5.8E-01	0.43	2.1E-03 (4.0E-03)	6.0E-01	5.7E-03 (4.0E-03)	1.6E-01

Table 3. Association of GWA findings on BP traits in UKbiobank and Japanese population. EAF effect allele frequency, β effect estimate, se standard error.

association with QRS interval and amplitude as a surrogate of myocardial mass in European population⁴¹ and replication with SBP, DBP and also HTN using different criteria based on reference adjusted curves in the Iranian children and adolescents⁴⁷, who was included in TCGS confirmation study is a signal for further investigations on the functional role of ZBED9 in BP regulation pathway in other populations.

High blood pressure is directly associated with vascular mortality due to stroke, cardiomyopathy, aneurysm, cardiac hypertrophy, aorta stenosis, sudden cardiac arrest, and myocardial infarction⁴⁸. Accordingly, gene-set-based analysis by Open Target Post GWAS highlighted the probable pleiotropic effects of detected loci, which are likely acting as a common genetic etiology for BP and cardiovascular diseases. Accordingly, the lethal cardiovascular diseases may affect allele frequency of effective variants with advancing age in our study due to survival reduction, so overlook the effect of functional variants in older groups as a competing risk^{49,50}.

We acknowledge that some limitations are evident in our study. First, there is evidence for significant differences in the magnitude of genetic variants effects on high blood pressure between men and women⁵¹. However, there was inadequate statistical power to conduct GWAS by sex after removing related individuals in our study. Moreover, the calculated PRS for the quantitative traits showed similar patterns by sex. Second, despite the high predictive accuracy of PRS by genomic variants and confirm significant SNPs on ZBED9 with family-based regression analysis, it was a case for overfitting due to selecting individuals with similar genetic background from the TCGS cohort. Accordingly, our findings validated in an independent sample of the Iranian population. However, our results still needs to be cross-validated in a different sample of the Iranian population as well⁵². Third, using a family-based design for GWAS has the advantages of complete robustness against genetic heterogeneity. However, family-based designs are likely biased due to population substructures, and association tests yield at the price of inflated type I error and reduce statistical power^{53,54}. In this way, we included independent samples from a family bases cohort.

Additionally, there was no association between autozygosity and BP quantitative traits in an international meta-analysis, including TCGS families with a relatively high inbreeding rate⁵⁵. Finally, we found signals which were not in LD with previously reported genetic variants on BP traits. Further investigation is necessary to fine mapping of these regions when imputed variants on the Iranian genome become available⁵⁶.

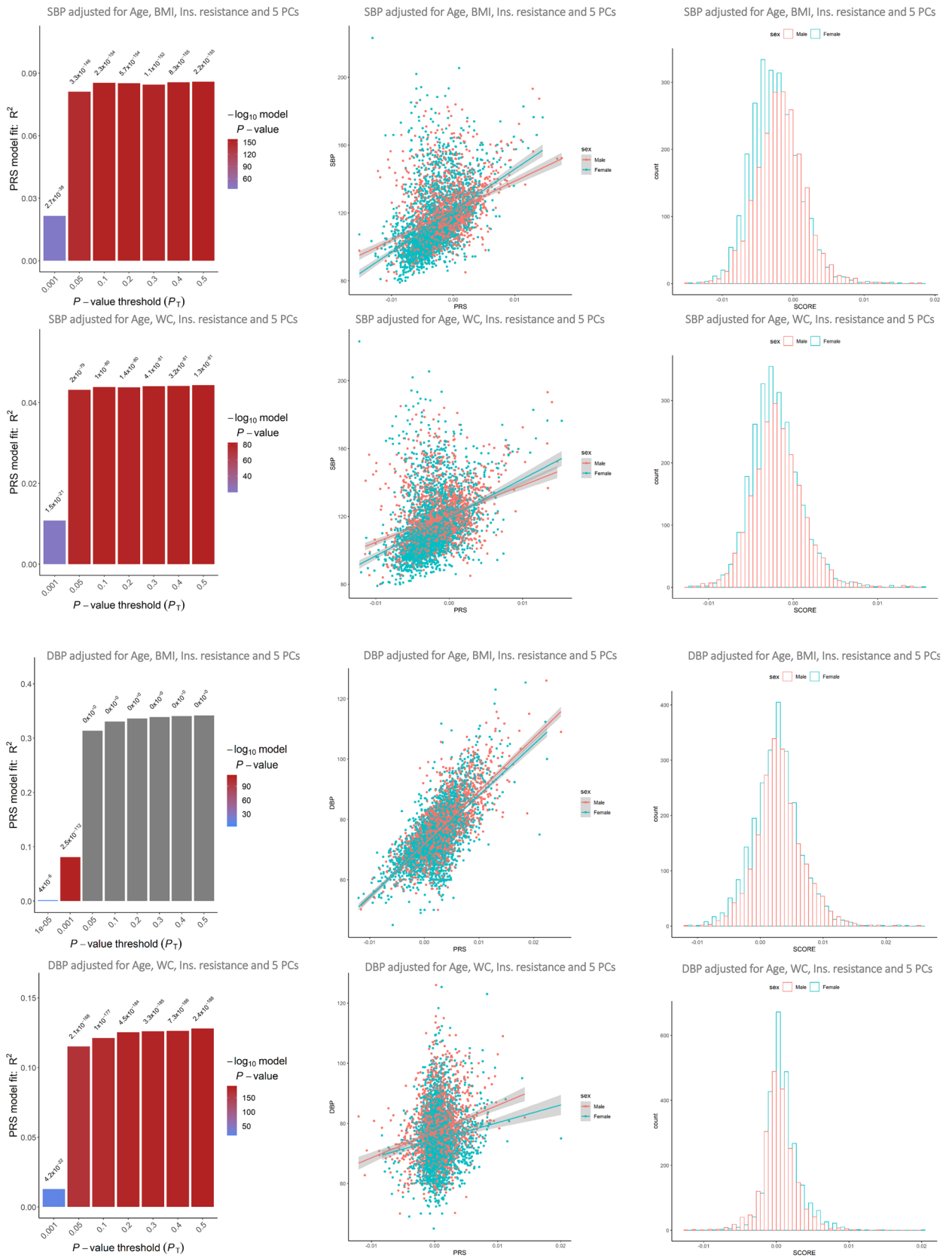


Figure 3. Predictive accuracy corresponding to a range of P value thresholds in regression models and the scatter plot of best fit PRS by sex in the discovery dataset.

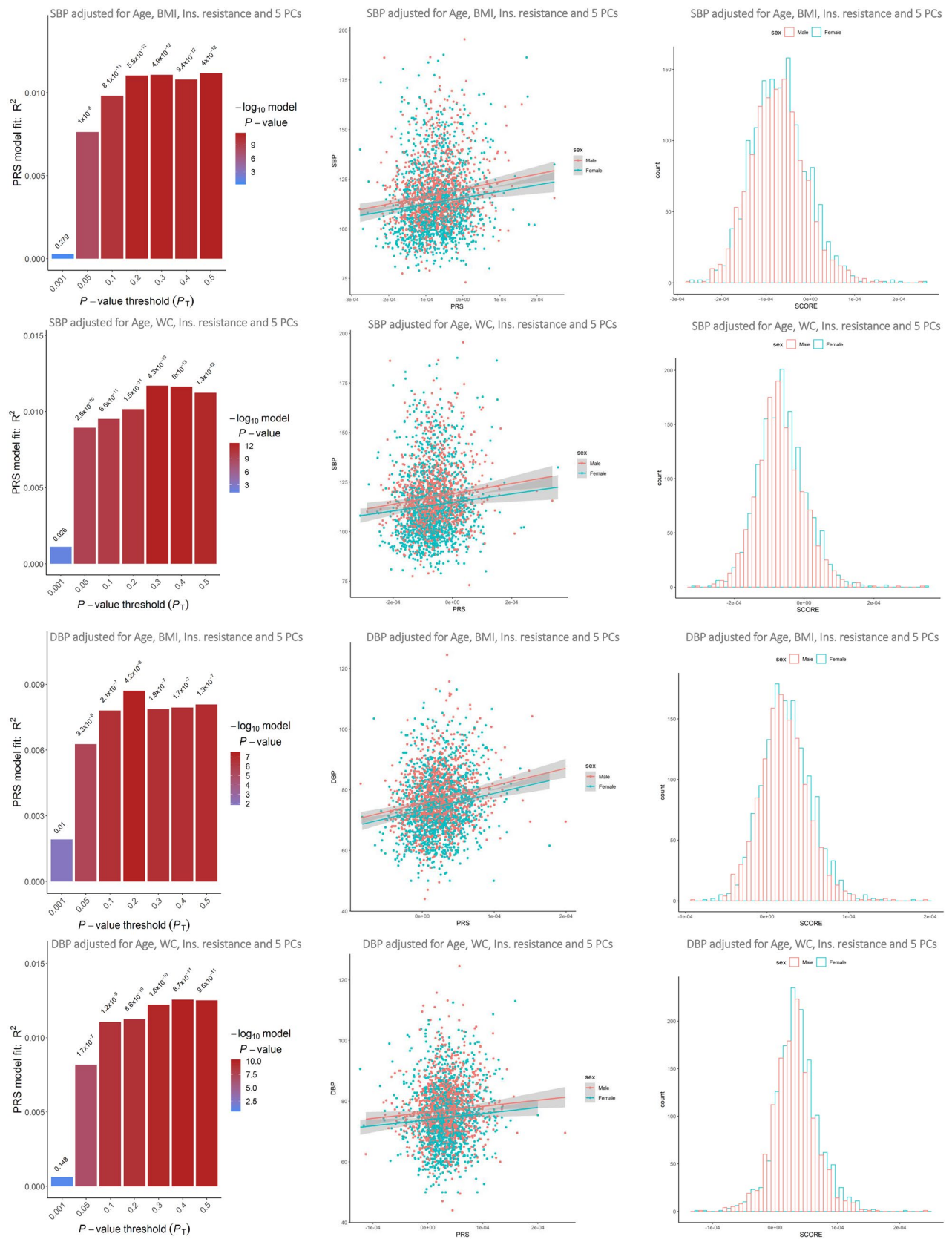


Figure 4. Predictive accuracy corresponding to a range of P value thresholds in regression models and the scatter plot of best fit PRS by sex in the independent sample of TCGs.

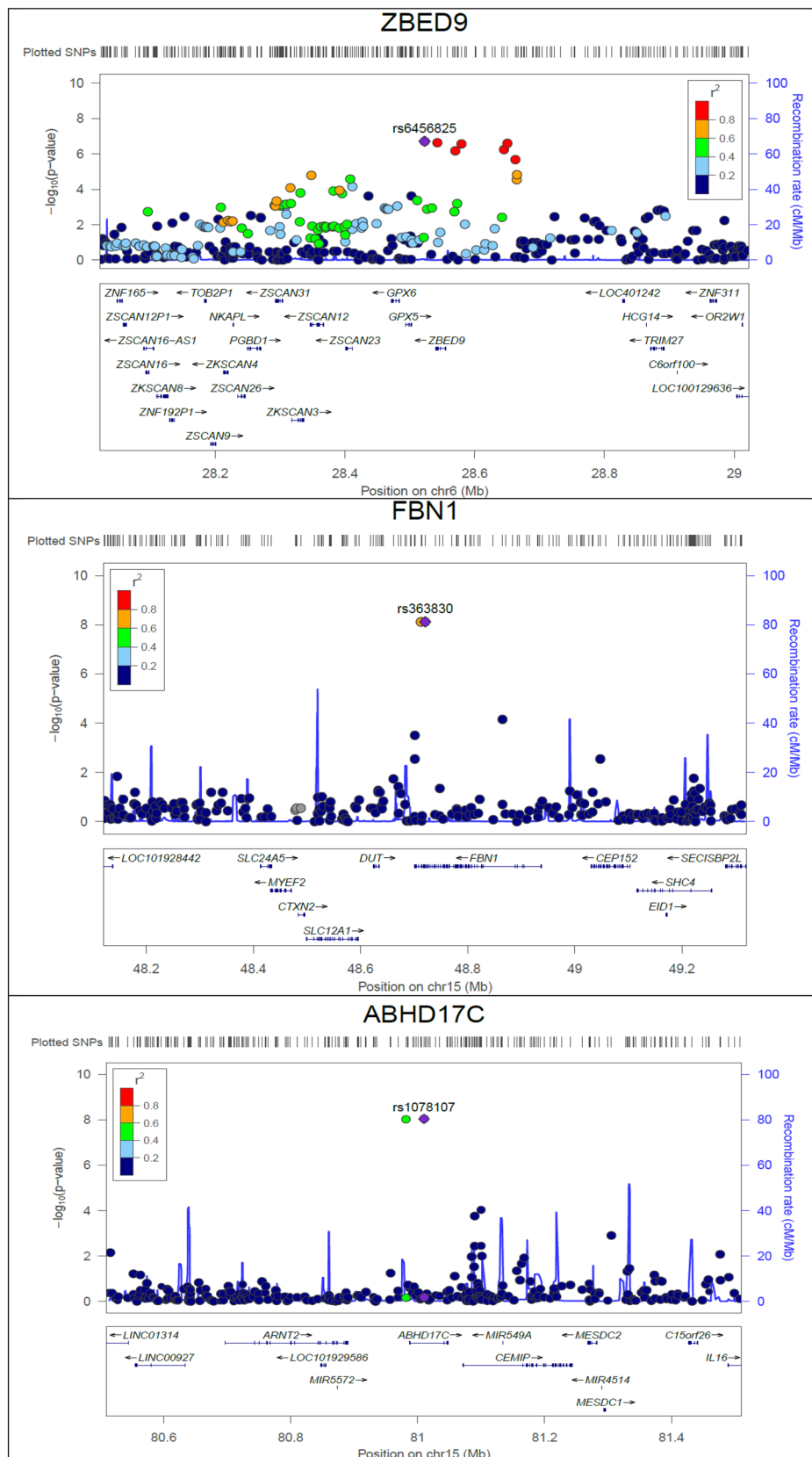


Figure 5. Locus zoom plot of GWAS findings in detected loci.

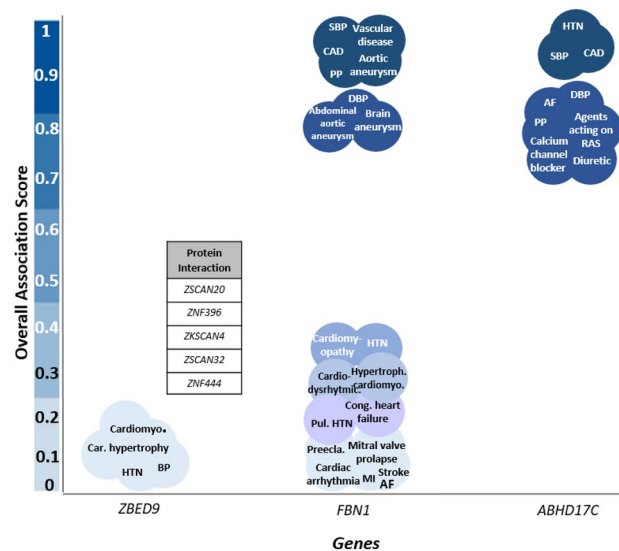


Figure 6. Prioritize targets by evidence of gene-disease overall association score in known loci and protein interaction networks of *ZBED9*.

Conclusion

We identified three loci, of which two loci were previously reported in individuals with European and Non-European ancestries. The association of genetic variants on *ZBED9* and BP traits has not been previously reported in other populations. Their effects were consistent after controlling environmental factors by two different GWA analyses and confirmed in a family-based linkage study. Although there is no strong evidence to support the function of *ZBED9* in blood pressure regulation by Open Target Post GWAS, its association with QRS interval and QRS amplitude as surrogates of myocardial mass may provide new insight into pleiotropic effects of hypertension and other cardiovascular diseases.

Data availability

Summary-level data for BP traits in UK Biobank and Japanese population are publicly available in (<http://www.nealelab.is/uk-biobank/>) and (<http://jenger.riken.jp/en/result>), respectively.

Received: 3 December 2020; Accepted: 11 May 2021

Published online: 03 June 2021

References

- NCD Risk Factor Collaboration (NCD-RisC). Contributions of mean and shape of blood pressure distribution to worldwide trends and variations in raised blood pressure: A pooled analysis of 1018 population-based measurement studies with 88.6 million participants. *Int. J. Epidemiol.* **47**(3), 872–883i. <https://doi.org/10.1093/ije/dyy016> (2018).
- Forouzanfar, M. H. *et al.* Global burden of hypertension and systolic blood pressure of at least 110 to 115 mm Hg, 1990–2015. *JAMA* **317**(2), 165–182 (2017).
- Kearney, P. M. *et al.* Global burden of hypertension: Analysis of worldwide data. *Lancet* **365**, 217–223 (2005).
- Mein, A., Caulfield, J., Dobson, J. & Munroe, B. Genetics of essential hypertension. *Hum. Mol. Genet.* **13**, R169–R175 (2004).
- Kolifarhood, G. *et al.* Heritability of blood pressure traits in diverse populations: A systematic review and meta-analysis. *J. Hum. Hypertens.* **24**, 1–1 (2019).
- Carretero, O. A. & Oparil, S. Essential hypertension: Part I: Definition and etiology. *Circulation* **101**(3), 329–335 (2000).
- Ehret, G. B. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
- Visscher, P. M. *et al.* 10 years of GWAS discovery: Biology, function, and translation. *Am. J. Hum. Genet.* **101**(1), 5–22 (2017).
- Lip, S. & Padmanabhan, S. Genomics of blood pressure and hypertension—extending the mosaic theory towards stratification. *Can. J. Cardiol.* **20**, 10 (2020).
- Rafiq, S., Anand, S. & Roberts, R. Genome-wide association studies of hypertension: Have they been fruitful?. *J. Cardiovasc. Trans. Res.* **3**, 189–196. <https://doi.org/10.1007/s12265-010-9183-9> (2010).
- Rosenberg, N. A. *et al.* Genome-wide association studies in diverse populations. *Nat. Rev. Genet.* **11**(5), 356–366 (2010).
- Kolifarhood, G. *et al.* Generality of genomic findings on blood pressure traits and its usefulness in precision medicine in diverse populations: A systematic review. *Clin. Genet.* **96**(1), 17–27 (2019).
- Tabrizi, J. S., Sadeghi-Bazargani, H., Farahbakhsh, M., Nikniaz, L. & Nikniaz, Z. Prevalence and associated factors of prehypertension and hypertension in Iranian population: The Lifestyle Promotion Project (LPP). *PLoS One* **11**(10), e0165264 (2016).
- Hadaegh, F. *et al.* Prehypertension tsunami: A decade follow-up of an Iranian adult population. *PLoS One* **10**(10), e0139412 (2015).
- Sarafzadegan, N. *et al.* The Isfahan cohort study: Rationale, methods and main findings. *J. Hum. Hypertens.* **25**(9), 545–553 (2011).
- Shirani, S. *et al.* Awareness, treatment and control of hypertension, dyslipidaemia and diabetes mellitus in an Iranian population: The IHHP study. *East Mediterr. Health J.* **15**(6), 1455–1463 (2009).
- Fattahi, Z. *et al.* Iranome: A catalog of genomic variations in the Iranian population. *Hum. Mutat.* **40**(11), 1–17. <https://doi.org/10.1002/humu.23880> (2019).

18. Zarkesh, M. *et al.* Heritability of the metabolic syndrome and its components in the Tehran Lipid and Glucose Study (TLGS). *Genet. Res.* **94**(6), 331–337 (2012).
19. Daneshpour, M. S. *et al.* Rationale and design of a Genetic Study on Cardiometabolic Risk Factors: Protocol for the Tehran Cardiometabolic Genetic Study (TCGS). *JMIR Res. Protoc.* **6**(2), e28 (2017).
20. Azizi, F., Zadeh-Vakili, A. & Takyar, M. Review of rationale, design, and initial findings: Tehran lipid and glucose study. *Int. J. Endocrinol. Metab.* **22**(16(4 Suppl)), e84777. <https://doi.org/10.5812/ijem.84777> (2018).
21. Honaker, J., King, G., Blackwell M. Amelia II: A program for missing data. *J. Stat. Softw.* **45**(7), 1–47. <http://www.jstatsoft.org/v45/i07/> (2011).
22. Anderson, C. A. *et al.* Data quality control in genetic case–control association studies. *Nat. Protoc.* **5**(9), 1564–1573 (2010).
23. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/> (2020).
24. Choi, S., Mak, T. & O'Reilly, P. Tutorial: A guide to performing polygenic risk score analyses. *Nat. Protoc.* **15**(9), 2759–2772. <https://doi.org/10.1038/s41596-020-0353-1> (2020).
25. Kolifarhood, G. *et al.* Familial genetic and environmental risk profile and high blood pressure event: A prospective cohort of cardio-metabolic and genetic study. *Blood Press.* **1**, 1–9. <https://doi.org/10.1080/08037051.2021.1903807> (2021).
26. Wang, T. & Elston, R. C. Regression-based multivariate linkage analysis with an application to blood pressure and body mass index. *Ann. Hum. Genet.* **71**(1), 96–106 (2007).
27. McLaren, W. *et al.* The ensembl variant effect predictor. *Genome Biol.* **17**(1), 122. <https://doi.org/10.1186/s13059-016-0974-4> (2016).
28. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**(Database issue), D1005–D1012 (2019).
29. Stelzer, G. *et al.* The GeneCards suite: From gene data mining to disease genome sequence analysis. *Curr. Protoc. Bioinform.* **54**, 1301–13033. <https://doi.org/10.1002/cpbi.5> (2016).
30. Peat, G. *et al.* The open targets post-GWAS analysis pipeline. *Bioinformatics* **36**(9), 2936–2937. <https://doi.org/10.1093/bioinformatics/btaa020> (2020).
31. Han, S. *et al.* Integrating GWASs and human protein interaction networks identifies a gene subnetwork underlying alcohol dependence. *Am. J. Hum. Genet.* **93**(6), 1027–1034. <https://doi.org/10.1016/j.ajhg.2013.10.021> (2013).
32. Wang, J., Zheng, J., Wang, Z., Li, H. & Deng, M. Inferring gene-disease association by an integrative analysis of eQTL genome-wide association study and protein–protein interaction Data. *Hum. Hered.* **83**(3), 117–129. <https://doi.org/10.1159/000489761> (2018).
33. Koscielny, G. *et al.* Open targets: A platform for therapeutic target identification and validation. *Nucleic Acids Res.* **45**(D1), D985–D994. <https://doi.org/10.1093/nar/gkw1055> (2017).
34. Wain, L. V. *et al.* Novel blood pressure locus and gene discovery using genome-wide association study and expression data sets from blood and the kidney. *Hypertension* **70**(3), e4–19 (2017).
35. Hoffmann, T. J. *et al.* Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat. Genet.* **49**(1), 54 (2017).
36. Giri, A. *et al.* Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat. Genet.* **51**(1), 51–62 (2019).
37. Feitosa, M. F. *et al.* Novel genetic associations for blood pressure identified via gene–alcohol interaction in up to 570K individuals across multiple ancestries. *PLoS One* **13**(6), e0198166 (2018).
38. Kichaev, G. *et al.* Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* **104**(1), 65–75. <https://doi.org/10.1016/j.ajhg.2018.11.008> (2019).
39. Warren, H. R. *et al.* Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat. Genet.* **49**(3), 403 (2017).
40. Wu, Y. *et al.* Genome-wide association study of medication-use and associated disease in the UK Biobank. *Nat. Commun.* **10**(1), 1 (2019).
41. Van der Harst, P. *et al.* 52 genetic loci influencing myocardial mass. *J. Am. Coll. Cardiol.* **68**(13), 1435–1448 (2016).
42. Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**(12), 1755. <https://doi.org/10.1038/s41588-018-0205-x> (2018).
43. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat. Genet.* **48**(10), 1151–1161 (2016).
44. Craddock, N. J., Jones, I. R. & Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**(7145), 661 (2007).
45. Takeuchi, F. *et al.* Interethnic analyses of blood pressure loci in populations of East Asian and European descent. *Nat. Commun.* **9**(1), 1–6 (2018).
46. Kato, N. *et al.* Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat. Genet.* **47**(11), 1282–1293 (2015).
47. Ataei, N. *et al.* Blood pressure percentiles by age and height for children and adolescents in Tehran, Iran. *J. Hum. Hypertens.* **30**(4), 268–277. <https://doi.org/10.1038/jhh.2015.52> (2016).
48. Prospective, L. S. & Collaboration, S. Age-specific relevance of usual blood pressure to vascular mortality: A meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* **360**(9349), 1903–1913 (2002).
49. Schooling, C. M. Biases in GWAS—the dog that did not bark. *bioRxiv* **1**, 709063 (2019).
50. Mostafavi, H. *et al.* Identifying genetic variants that affect viability in large cohorts. *PLoS Biol.* **15**(9), e2002458. <https://doi.org/10.1371/journal.pbio.2002458> (2017).
51. Faulkner, J. L. & Belin de Chantemèle, E. J. Sex differences in mechanisms of hypertension associated with obesity. *Hypertension* **71**(1), 15–21 (2018).
52. Akbarzadeh, M. *et al.* GWAS findings improved genomic prediction accuracy of lipid profile traits: Tehran Cardiometabolic Genetic Study. *Sci. Rep.* **11**(1), 1–9 (2021).
53. Lasky-Su, J. *et al.* On genome-wide association studies for family-based designs: An integrative analysis approach combining ascertained family samples with unselected controls. *Am. J. Hum. Genet.* **86**(4), 573–580. <https://doi.org/10.1016/j.ajhg.2010.02.019> (2010).
54. Fardo, D. W. *et al.* On family-based genome-wide association studies with large pedigrees: Observations and recommendations. *BMC Proc.* **8**(Suppl 1 Genetic Analysis Workshop 18Vanessa Olmo), S26. <https://doi.org/10.1186/1753-6561-8-S1-S26> (2014).
55. Clark, D. W. *et al.* Associations of autozygosity with a broad range of human phenotypes. *Nat. Commun.* **10**(1), 1–7 (2019).
56. Daneshpour, M. S. *et al.* Genetic identification for non-communicable disease: Findings from 20 years of the Tehran Lipid and Glucose Study. *Int. J. Endocrinol. Metab.* **16**(4), 20 (2018).

Acknowledgements

This study as part of a PhD project by Kolifarhood G has been funded by a Grant from the Deputy for Research Affairs of Shahid Beheshti University of Medical Sciences. We are grateful to all TLGS and Gemiran colleagues

that assisted with this study through the data collection process. Moreover, we would like to thank Dr. Hossein Mozafar Sadati for his comments on missing data imputation.

Author contributions

M.S.D. and G.K. had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis; G.K., M.S.D., S.S., M.A., and F.H. contributed to study concept, design, and supervision; M.S.D., B.S.K., K.G., M.A., and G.K. contributed in acquisition of data; G.K., M.S.D., S.S., M.A., F.H., and B.S.K. contributed in analysis and interpretation of data; G.K., M.S.D., M.A., S.S., and S.R. contributed in statistical analysis; F.A., M.S.D., and F.H. contributed in administrative, technical, or material support; G.K., M.S.D., B.S.K., M.A., S.S., F.H., K.G., M.A.R. and F.A.: drafting of the manuscript and critical revision of the manuscript for important intellectual content. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-90925-w>.

Correspondence and requests for materials should be addressed to M.S.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021