Genome-wide analysis of DNA methylation in hypertension-discordant monozygotic twins

Zhentian Wu
Peking University
Wenjing Gao (pkuepigwj@126.com)
Peking University https://orcid.org/0000-0002-8242-0775
Weihua Cao
Peking University
Chunxiao Li
Peking University
Canqing Yu
Peking University
Jun Lv
Peking University
Zengchang Pang
Qingdao Center for Diease Control and Prevention
Liming Cong
Zhejiang Center for Diease Control and Prevention
Hua Wang
Jiangsu Center for Diease Control and Prevention
Xianping Wu
Sichuan Center for Diease Control and Prevention
Liming Li
Peking University

Research

Keywords: twin studies; DNA methylation; hypertension; blood pressure

Posted Date: March 11th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-16695/v1

Abstract

Background: DNA methylation has great potential for identifying the aetiology of hypertension. The aim of this study was to explore the correlation between hypertension and DNA methylation using twins discordant for hypertension in China. Methods: In this study, 43 pairs of monozygotic twins discordant for hypertension (age 31.9-72.3 years; 67.4% male) from the Chinese National Twin Registry were recruited. Genome-wide DNA methylation was measured using the Illumina Human methylation EPIC Beadchip in whole-blood-derived DNA. Standardized questionnaires were used to collect twin data on the following variables: age, gender, socioeconomic level, lifestyle factors (including smoking, alcohol drinking, vegetable intake, and physical activity). Blood pressure, height, weight, and other anthropometric indicators were obtained by physical examination. Empirical Bayes paired moderated t-test was utilized to compare the methylation data within twin pairs. Results: Four suspected significant methylation sites, cg00950476, cg08041400, cg26733338, and cg08580087 were identified. All of these four sites locate on known loci, which were LINC01252, BDP1, SYT1, and ODZ4, respectively. The main function includes transcriptional regulation, learning and cognitive, neurodevelopment. The significant sites were further replicated among two different replication population, the first replication population contained 38 hypertension concordant monozygotic twin pairs and 38 non-hypertension concordant monozygotic twin pairs matched in age, sex, region, and birth order, and the second replication group included 21 MZ twin pairs discordant for hypertension . None of them, however, were significant. The methylation variation in the above sites may influence blood pressure, independent of genetic and early-life environmental factors. Conclusions: This study found four suspected methylation sites correlated with hypertension. However, all four sites failed the replication analysis. More hypertension-discordant monozygotic twin pairs are needed to replicate these findings in the future to explore the stability of the results.

Introduction

Study Importance Questions

- What is already known about this subject?

- Numerous studies had focused on the correlation between hypertension and DNA methylation and tens of methylation sites on specific gene loci correlated with hypertension had been found.
- Limited DNA methylation studies were based on twin population worldwide. Moreover, there was a gap to be filled in this field in China.

- What does our study add?

- Although the four suspected methylation sites (cg00950476, cg08041400, cg26733338, and cg08580087) correlated with hypertension failed the replication analysis in this study, which might be related to imperfect replication design, it was worth to be further replicated.

- Genetic factors may influence DNA methylation. Disease-discordant monozygotic twin pairs are the ideal subjects when exploring potential contributions of DNA methylation to the aetiology of hypertension with minimum confounding from genetic heterogeneity.
- To our knowledge, this study was the first one to explore the correlation between hypertension and DNA methylation using twins discordant for hypertension in China and had the largest sample size in genome-wide methylation study of hypertension in twins.

Hypertension, the most common cardiovascular disease and the main risk factor for other severe cardiovascular diseases, has seriously consumed medical and social resources worldwide. Hypertension is currently considered to be a complex polygenic disease caused by genetic and environmental factors and their interactions[1]. According to a recent meta-analysis based on 17 studies globally, the heritability of systolic and diastolic blood pressure is 0.54 (95% CI: 0.48-0.60) and 0.49 (95% CI: 0.42-0.56), respectively[2]. Nevertheless, the genetic variants identified by genome-wide association studies (GWAS) can only explain less than 1% of blood pressure variation among individuals[3], which was called “missing heritability”.

DNA methylation plays a key role in the regulation of gene expression and may link between genes, environment, and hypertension, such as *NKCC1*[4,5], *AGTR1/AGTR2*[6,7]. Conversely, genetic factors may affect DNA methylation. Disease-discordant monozygotic (MZ) twin pairs, which share highly matched genetic structure and early environmental factors, provide an ideal model for exploring potential contributions of DNA methylation to the aetiology of hypertension with minimum confounding from genetic heterogeneity[8]. Analysis of discordant MZ twins has been successfully used to study epigenetic mechanisms in hypertension[9]. However, the sample size was mostly only dozens and the evidence from Asian twins was lacking.

In this study, we examine genome-wide DNA methylation using the Illumina Human methylation EPIC Beadchip in whole blood from 43 pairs of MZ twins discordant for hypertension. Then we attempted to further replicate significant cytosine-phosphate-guanine sites (CpG sites, where DNA methylation occurs) among two different replication population (detailed descripted in Methods section).

Results

According to the criteria in this study, a total of 43 MZ twin pairs discordant for hypertension were analysed in the discovery population. The age of the twins ranged from 32 to 72 years, with a mean of 53 years. There were 29 (67.4%) pairs of male twins. The first replication group included 76 cases and 76 controls, and the second replication group included 21 MZ twin pairs discordant for hypertension. No significant difference in socioeconomic level was found among discovery population, replication population 1 and 2. There were 18 (29.8%) twins had taken antihypertensive drugs within one month in the discovery population and 41 (27.0%) twins in replication population 1, 5 (11.9%) twins in replication population 2.

Table 1 shows the significant differences among discovery population, replication population 1 and 2. Notably, the difference in blood pressure between the case-controls was larger than that in the discovery population. The average blood pressure difference (SBP difference add DBP difference) between case-controls was 64mm Hg (ranged 11-151mm Hg). Only four pairs (5.3%) of case-controls had blood pressure difference less than 20mm Hg.

DNA methylation correlation analysis

Multivariate model 1 adjusted smoking, alcohol consumption, and SVA agent variables; model 2 additionally adjusted BMI and recalculated SVA agent variables.

Figure 1a displayed the volcano plot of paired model 1, which depicted the genome-wide DNA methylation difference at each CpG site between hypertension twins and non-hypertension co-twins. Only one CpG site showed *P* value < 10^{-6}, however, after correction for multiple testing, no CpG site remained significant under the threshold of FDR < 0.05. Figure 1b displayed the volcano plot of paired model 2, which showed similar results.

Then we filtered twin pairs among which blood pressure difference (SBP difference add DBP difference) ≥10mm Hg, ≥20mm Hg, ≥30mm Hg, ≥40mm Hg, respectively for empirical Bayes paired moderated t-test expecting to enhance the power to discover DNA methylation sites correlated with hypertension.

Three CpG sites with FDR < 0.05 were found in the twin paired model 1 among those with blood pressure difference ≥20 mm Hg. The results of the paired model for each blood pressure difference group were shown in Table 2. A CpG site with FDR=6.74E-02 was also found in the twin paired model 2 with blood pressure difference ≥20 mm Hg, which could be regarded as a suspected significant site to be replicated as well. The information of statistically significant sites in model 1 and model 2 was shown in Table 3 and the volcano plot was displayed in Figure 2.

Four significant CpG sites were hypomethylated in hypertensive twins. The significant CpG site (cg08580087) in model 2 was covered by 450K Beadchip, while the other three sites (cg00950476, cg08041400, cg26733338) were not.

Sensitivity analysis

Considering the influences from correlated diseases or medication treatment, we conducted the following sensitivity analyses in model 1 and model 2 by (1) Excluding twins discordant for diabetes mellitus and kidney disease; (2) Additionally adjusting self-reported medication use in the past month. All four significant sites maintained low *P* values, but failed to achieve a significant level of FDR < 0.05 (Supplement Table 1).

Replication analysis

Due to the small number of replication CpG sites, paired t-test was used in the replication analysis firstly. The significance level was $P < 0.05$ (two-sided test). Then empirical Bayes paired moderated t-test was used to test which site reached the Bonforroni test level ($P < 0.05$/ the number of significant sites found in the discovery stage, i.e. $P < 1.25E-2$). Since only four replication sites were not suitable for generating SVA, paired model 1 only adjusted smoking and alcohol consumption; model 2 additionally adjusted BMI. The replication results of four CpG sites were shown in Table 4.

In addition to the suspected significant CpG sites mentioned above, there were some other sites near the above CpG sites detected by MALDI-TOF-MS. Empirical Bayes paired moderated t-test was also performed on these sites, for which the significance level was FDR < 0.05. The result was shown in Supplement Table 2.

Cg08580087 was analysed by paired t-test and empirical Bayes paired moderated t-test in the replication population 2, the significance levels were all $P < 0.05$. Similarly, only smoking and drinking were adjusted in paired model 1, and BMI was further adjusted in model 2. The results were $P = 9.99E-01$ in paired t-test, $P = 4.66E-02$ in paired model 1 and $P = 1.15E-01$ in paired model 2. This CpG site could still not reach the significance level. However, these were not the main replication result considering insufficiency power led by the small sample size.

From the above results, we could conclude that four suspected significant CpG sites (cg00950476, cg08041400, cg26733338, cg08580087) failed in the replication phase, suggesting that these four sites found in the discovery stage were suspected false-positive sites.

In addition, we contrasted the P valule of methylation sites correlated to hypertension found in previous researches with in our models. One CpG site (cg17061862) correlated to SBP, which had a P value of 6.9E-05 in previous research[10], reached a P value of 8.2E-05 in paired model 1 with blood pressure difference ≥20 mm Hg and 5.1E-05 in paired model 2 with blood pressure difference ≥10 mm Hg, which replicated this site to a certain extent. Replication results in our models of the sites found in previous researches were displayed in Supplement Table3.

Enrichment analysis

The enrichment analysis of four sites was performed by comparing the GO pathway of four above genes with known genes covered by EPIC Beadchip. Table 5 showed the biological pathways with Fisher's exact test $P < 10^{-3}$. While FDRs of these pathways were all higher than 0.05, it could not be confirmed whether these genes were enriched in these pathways.

Discussion

In a two-stage design of discovery and replication analyses comprising 102 MZ Chinese twin pairs, four suspected significant CpG sites (cg00950476, cg08041400, cg26733338, and cg08580087) were found in twin pairs whose blood pressure difference ≥20 mm Hg. Although the four suspected CpG sites

correlated with hypertension failed the replication analysis in this study, which might be related to imperfect replication design, it was worth to be further validated.

All CpG sites failed to achieve a significant level of FDR < 0.05 in the sensitivity analyses. The disease status and medication treatment, however, were based on self-report rather than medical record or auxiliary diagnosis. Eliminating twins with inconsistent diabetes or kidney diseases inevitably leaded to a reduction in sample size. Therefore, we did not consider the results of sensitivity analyses as the main results.

The four suspected CpG sites are all on known genes (*LINC01252*, *BDP1*, *SYT1*, *ODZ4*).

LINC01252, long intergenic non-protein coding RNA 1252, has no clear functional report yet. Non-coding RNA is related to epigenetics, most of which are involved in heterochromatin formation, histone modification, targeted DNA methylation, and gene silencing. Long chain non-coding RNA (>200 nts) functions mainly for chromatin remodelling, transcriptional regulation, post-transcriptional regulation, and precursors of short interfering RNAs (siRNA). Many of long chain non-coding RNAs can enhance the catalytic effect of chromatin-modified proteins on specific sites, thereby altering chromatin status and affecting gene expression.

BDP1, B double prime 1, is a subunit of RNA polymerase III transcription initiation factor IIIB, which helps RNA polymerase III to target on the binding promoter to start transcription. Mutation of *BDP1* is related to hereditary deafness[11].

SYT1, synaptotagmin-1, is a membrane intrinsic protein on synaptic vesicles. It acts as a Ca^{2+} receptor in the process of vesicle transport and exocytosis. Ca^{2+} combines with synaptotagmin-1 to participate in the release of neurotransmitters to synapses. Current studies suggest that synaptotagmin-1 is related to learning, memory and cognitive function, and it has been found that *SYT1* gene is associated with attention deficit hyperactivity disorder (ADHD) and *SYT1* mutation leads to recurrent neurodevelopmental disorder[12].

ODZ4, also known as *TENM4* (teneurin transmembrane protein 4), plays an important role in the establishment of neuronal connectivity during development, and the defect of *ODZ4* is related to primary tremor[13,14].

Researchers had found that tens of methylation sites were correlated to the occurrence or development of hypertension. Candidate gene strategies found that the CpG sites on *NKCC1*, *SCNN1B*, *IL6*, *ADD1*, *ACE*, *AGTR1/AGTR2*, *ECE1*, *11betaHSD2*, *NET*, *TLR2*, *IFNG* were correlated with hypertension. The significant CpG sites found in genome-wide methylation studies (mostly 450K Beadchip) included CpGs on *ABCG4*, *SULF1*, *IGFBP3*, *KCNK3*, *PDE3A*, *PRDM6*, *ARHGAP24*, *OSR1*, *SLC227*, *TBX2*, *SYT7*, *ABLIM3*, *HDAC9*, *AMH*, *OVGP1* and cg00875989, cg09134341. All the above methylation sites were covered by the EPIC beadchip detection range except for the CpG site on *NKCC1*. In this study, the methylation level difference among these sites was consistent with previous studies, although they did not reach the significance

level of FDR < 0.05. Notably, all previous genome-wide methylation studies used Illumina 27 or 450K Beadchip. Since three of four suspected significant sites (cg00950476, cg08041400, and cg26733338) found in this study were all new CpG sites covered by the new EPIC Beadchip, there was no relevant study result regarding them.

So far, we could not find any other studies revealing these four significant CpG sites or their located genes related to human hypertension. It was worth noting that, however, *SYT1*, where cg26733338 was located, had been studied in renovascular hypertensive rats (RHR) with hypertensive stroke induced by artificial cold wave exposure. This study found that SYT1 protein was upregulated in the cerebral tissue of RHR[12]. A previous meta-analysis of GWAS found a SNP (rs751984) located on *SYT7* (synaptotagmin-7) was associated with mean arterial pressure (MAP)[15]. The protein encoded by *SYT7* regulated calcium-dependent membrane transport in synaptic transmission. *SYT7* was correlated with *SYT1*, suggesting that *SYT1* might be correlated with hypertension.

All four suspected significant CpG sites in this study failed replication. We consider it might due to the following issues: (1) Since there was no established methodology for calculating the power of genome-wide methylation study in twins, the sample size was still a problem. Reviews of twins for epigenetics suggested that “a relatively small number (15–25) of phenotypically discordant twin pairs had sufficient (>80%) power to detect epigenetic changes of 1.2-fold, where an effect size of 1.2-fold change was significantly greater than the null experimental variance threshold for the assay (1.15-fold change)”[16,17]. Usually, the sample size of previous genome-wide methylation MZ twin study targeted on hypertension was about ten pairs. Our study was the largest genome-wide methylation study based on twins discordant for hypertension so far; however, the sample size might still be insufficient. Moreover, about 2% of sites were missing in the replication analysis due to unsuccessful detection. (2) Most of the significant sites correlated with hypertension were found previously in the general population. Although potential confounding from age or gender was controlled by matching or covariable adjustment, the influence from the genetic background was not considered in the general population. In this study, MZ discordant for hypertension were used to match the genetic factors. Therefore, it was reasonable to presume that the CpG sites associated with hypertension found in the previous general population might be affected by genetic factors. The DNA methylation of parts of CpG sites might only be an intermediate factor between genetic factors to hypertension, rather than an intermediate factor between environmental factors to hypertension or a process of gene-environment interaction. (3) The replication design was imperfect. The first replication population included MZ twin pairs, which, however, were regarded as individual since they were hypertension or non-hypertension concordant pairs. Furthermore, the first replication population showed a higher difference in blood pressure between hypertension twins and non-hypertension co-twins than discovery population. These differences might originate from different treatments and proportion of controlled hypertension patients among discovery population and replication population. Although medication had been considered when performing sensitivity analyses, this might be a major limitation of this study, hampering replication of findings. Considering significant sites were discovered in the MZ with blood pressure difference ≥20mmHg, perhaps replications need to

be further investigated in case-controls with blood pressure difference ≥20mmHg. (4) BMI was significantly higher in hypertension twins compared to non-hypertension co-twins in discovery population and replication population. Although model 2 adjusted for BMI, the difference might impact the results of paired analysis. (5) Multiple testing correction was the preferable way to avoid over-inflating the results in the high throughput study[18,19]. In this study, we found that a large number of sites with original *P* values less than 10^{-5}; however, most of them were no longer significant after multiple testing correction. (6) Methodology in methylation data processing needed to be further improved. So far, there was no consistent method for data standardization[20]. In this study, we used DASEN with watermelon package for data standardization/normalization and SVA for potential confounding factors adjustment (mainly test plates and batches effect). These methods were commonly used in methylation studies. Different researchers might use different methods, followed by different findings. (7) Pyrosequencing was still the gold standard for methylation detection. In this study, MALDI-TOF-MS was used to detect methylation sites in candidate regions, which might not be perfectly consistent with pyrosequencing. In addition, the DNA extraction and methylation detection for discovery population and replication population were conducted in two different labs; and the DNA extraction kits used were from different manufacturers. Although all testing was conducted under strict quality control, the variance from different operators, reagents, and batches of experiments could not be excluded. (8) The disease status and medication treatment were based on self-report. For instance, if a participant not diagnosed with hypertension was taking thiazide diuretic drugs for some kidney disease without knowing that it is antihypertensive drug, and so answered no to question 2, this participant would end up in the non-hypertension group. Hence we attempted to exclude twins discordant for kidney disease in the sensitivity analysis.

The strength of our study included a twin design, a new generation of methylation detection array, and careful adjustment for established and potential risk factors for hypertension. However, this study had some limitations. The sample size might be insufficient even though it might be the largest in hypertension-discordant twin pairs. The first replication population included 76 MZ twin pairs, while they were analysed individually. However, we additionally included 21 pairs of hypertension-discordant MZ twin pairs in the second replication population, although only one CpG site covered by 450K beadchip could be replicated. DNA methylation was tested using peripheral blood sample instead of hypertension-related tissues. Considering the difficulty for human tissues such as kidney or myocardium, blood was a common fitting surrogate marker for reflecting DNA methylation in tissues[21]. Another disadvantage was that these CpG sites were identified through cross-sectional analyses, and we were not able to characterize the direction of the association. Despite these limitations, we discovered four CpG sites associated with BP among 43 twin pairs, although they could not reach the significance level in the replication stage, partly due to the imperfect replication population design.

Conclusion

This study used twins to analyse the correlation between hypertension and DNA methylation and found four suspected significant sites, cg00950476, cg08041400, cg26733338, and cg08580087 located on

LINC01252, *BDP1*, *SYT1,* and *ODZ4*, in discovery population; however, none of them survived the replication. In the future, we need a bigger sample size and different tissues for further replication. Longitudinal data will be needed for causality inference between these CpG sites and hypertension.

Methods

Participants

Twin pairs discordant for hypertension were identified from the Chinese National Twin Registry (CNTR). In 2013, the CNTR recruited 1281 twins from Shandong, Zhejiang, Jiangsu, and Sichuan provinces. Hypertension status was assessed through a self-reported questionnaire. Twins discordant for hypertension were selected using the following inclusion criteria: (1) One diagnosed with hypertension by county-level hospital or above, while the other one not; (2) MZ twins; (3) Twins aged 30 years or older; (4) Both twins finished the questionnaires and physical examination; (5) Blood sample available. Exclusion criteria included: (1) DNA extraction failed; (2) Pregnant women; (3) Twin pairs in which any one of the twins excluded. Ultimately, a total of 43 twin pairs discordant for hypertension satisfied the above criteria and were analysed as the discovery population. We had two groups of replication population: The first group included 38 hypertension concordant MZ twin pairs and 38 non-hypertension concordant MZ twin pairs, pairwise matched for age (±5 years), sex, region, and birth order. The second group included 21 MZ twin pairs discordant for hypertension. The study was approved by the Peking University Biomedical Ethics Committee (IRB00001052-16056), and informed consent was obtained from all participants.

Measurements

Standardized questionnaires were used to collect twin data on the following variables: age, gender, socioeconomic level, lifestyle factors (including smoking, alcohol drinking, vegetable intake, and physical activity). Blood pressure, height, weight, and other anthropometric indicators were obtained by physical examination. Two questions in questionnaires focused on hypertension: (1) Have you ever been diagnosed with hypertension by county-level or above? (2) Have you ever taken antihypertensive drugs within one month before the investigation? The subjects who met with any of the following criterias were defined as hypertension patients: (1) Yes to question 1; or (2) Yes to question 2; or (3) Systolic blood pressure (SBP) ≥140 mm Hg and/or diastolic blood pressure (DBP) ≥90 mm Hg. Blood pressure was measured twice by trained observers using the OMRON HEM-7200 sphygmomanometer on the right arm of seated participants after 5-min rest. The average SBP and DBP were calculated. If the difference between the two measurements was more than 10 mm Hg, the average of the nearest two measurements were calculated after the third measurement. Height was measured by height meter, waist circumference and hip circumference were measured by tape, and weight and percent body fat (PBF) were measured by TANITA professional body composition scales.

Sample collection and processing

Blood samples were collected, transported and stored according to a standardized CNTR sample handling and storage protocol for all participants in four provinces. Totally 15 ml fasting blood sample was collected from each participant using the vacutainer system in the morning and temporarily stored in 4°C portable refrigerator. The samples were transported to the local center and preserved in -40°C refrigerator the same day after sample aliquots. Within one month, the samples were dispatched by a commercial courier to the central laboratory at the School of Public Health Peking University in temperature-controlled shipping boxes.

DNA Extraction and Methylation analyses

DNA extraction and methylation detection of the discovery population was conducted in a commercial laboratory authorized by Illumina. DNA was extracted from 2ml blood samples in EDTA•K2•H20 anti-coagulant-containing tube. After natural thawing, the whole blood DNA was extracted using the E.Z.N.A. blood DNA kit (Omega Bio-tek, Inc, Norcross, GA). Then the genome-wide methylation profiles were detected using Illumina Human methylation EPIC Beadchip (Illumina, Inc, San Diego, CA). This array covers >850,000 CpG sites containing 91.1% of the Infinium Human Methylation450 BeadChip (450K Beadchip) sites plus over 330,000 CpGs that are located on enhancer regions[22]. Samples were randomly assigned to the patches and plates. The β-values and 0-100% methylation were used to report the result.

DNA extraction, methylation detection and zygosity determination of the first replication population were conducted in a third-party lab. Similarly, DNA was extracted after natural thawing using Bioteke whole blood DNA extraction kit (BioTeke, Co. Ltd, Beijing, China). Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) was used to validate the methylation level of candidate sites by EpiTYPER Reagent Kit (Agena Bioscience, Inc, San Diego, CA), and the MassARRAY Epityper (Agena Bioscience, Inc) software was used to obtain raw data and dot diagrams (visualizing the methylation level of each methylation site).

DNA methylation for the second replication group was detected using 450K Beadchip at Huazhong University of Science and Technology and has been described previously[23]. These methylation data were available before designing this study.

Zygosity analysis

For the discovery population, zygosity was confirmed by genotyping 59 SNPs (single nucleotide polymorphism) on EPIC Beadchip. The SNP correlation coefficient for MZ twin pairs was extremely close to 1, while it was generally less than 0.8 for the dizygotic (DZ) twin pairs. In the replication population, we used 21 short tandem repeats (STRs) detection to identify the zygosity. The accuracy of STRs comparison in zygosity determination could exceed 99%, which was validated in Chinese twins[23,24].

Quality control of the DNA methylation data

For the DNA methylation data of the discovery population, R minfi package was used to carry out the data quality control[25].

Sample quality control included: (1) Quality control report chart recommended by Illumina (Supplement Figure 1). (2) Multiple dimension-scale (MDS) figure to test outliers (Supplement Figure 2). (3) Two samples having 1% of sites with a detection p-value greater than 0.01 were removed.

Site quality control included: (1) 3,734 sites having a detection p-value greater than 0.01 in more than 1% of samples. (2) 1,960 probes with beadcount <3 in 5 % of samples. (3) 43,877 sites contained a locus with minor allele frequency (MAF)>0.05 in the Asian population or crossing with other loci were removed[26]. (4) 59 SNPs were removed.

Finally, there were 43 eligible twin pairs and 816,602 probes were left for the statistical analyses. The quality control in the replication population was similar.

Statistical analysis

In MZ discordant for hypertension, empirical Bayes paired moderated t-test was used to compare the methylation level between hypertension and non-hypertension co-twins and identify hypertension-related differentially methylated CpG sites (DMCs). The empirical Bayes paired moderated t-test has two following advantages for the application of MZ[27]: Firstly it helps to minimize the error between samples brought by the probe and it is suitable for studies with small sample size. Ebayes function in the R limma package was used[28].

Considering the same age, sex, and region between MZ twins, we adjusted for smoking, drinking, BMI, and surrogate variable analysis (SVA) agent variables[29-31] in multivariate models. False discovery rate (FDR) <0.05 was used for multiple testing correction.

To examine the robustness of our findings, we also conducted sensitivity analyses by excluding discordant twin pairs for self-reported diabetes mellitus or kidney disorders.

We applied Bonforroni correction for multiple testing ($P < 0.05$/ the number of CpGs significant in the discovery stage) for CpGs selection in the replication stage. If a significant CpG site found in the discovery stage was covered by the 450K Beadchip, it could be replicated both in the first and second replication groups.

Top differentially methylated CpG sites were used for enrichment analysis. By linking these sites with the gene symbol using the annotation file provided by Illumina, the gene set of all significant sites could be obtained, as well as a corresponding gene set for the remaining CpG sites. The gene ontology (GO) functions and processes of the genes of all significant CpG sites were investigated using the gene ontology enrichment analysis and visualization web-based tool (Gorilla, http://cbl-gorilla.cs.technion.ac.il/)[32,33]. P-values were corrected for multiple testing using FDR.

Declarations

Ethics approval and consent to participate

The study was approved by the Peking University Biomedical Ethics Committee (IRB00001052-16056), and informed consent was obtained from all participants.

Consent for publication

Not applicable.

Availability of data and material

The datasets generated and/or analysed during the current study are not publicly available due to confidentiality agreement but are available from the corresponding author on reasonable request.

Competing interests

The authors declared no conflict of interest.

Funding

This study was supported by the Special fund for Health scientific research in the public welfare (201502006, 201002007) and National Natural Science Foundation of China (81573223).

Authors' contributions

Gao Wenjing and Li Liming designed this research. Wu Zhentian analysed the data and wrote the initial manuscript and had primary responsibility for final content. Cao Weihua, Li Chunxiao, Yu Canqing, and Lv Jun provided critical input on drafts of the manuscript. Pang Zengchang, Cong Liming, Wang Hua, and Wu Xianping provided local twin data. All authors read and approved the final manuscript.

Acknowledgements

We appreciate the help from investigators regarding data collection in CDCs of Qingdao, Zhejiang, Jiangsu, and Sichuan.

Authors' information

Wu Zhentian[1], Gao Wenjing[1*], Cao Weihua[1*], Li Chunxiao[1], Yu Canqing[1], Lv Jun[1], Pang Zengchang[2], Cong Liming[3], Wang Hua[4], Wu Xianping[5], Li Liming[1].

[1]Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China. Correspondence: Gao Wenjing (pkuepigwj@126.com), Cao Weihua (caoweihua60@163.com)

[2]Qingdao Center for Disease Control and Prevention, Qingdao 266033, China

[3]Zhejiang Center for Disease Control and Prevention, Hangzhou 310051, China

[4]Jiangsu Center for Disease Control and Prevention, Nanjing 210009, China

[5]Sichuan Center for Disease Control and Prevention, Chengdu 610041, China.

References

1 Svetkey, L. P. *et al.* Modulation of the BP response to diet by genes in the renin-angiotensin system and the adrenergic nervous system. *Am J Hypertens* **24**, 209-217, doi:10.1038/ajh.2010.223 (2011).

2 Wang, B. *et al.* Genetic contribution to the variance of blood pressure and heart rate: a systematic review and meta-regression of twin studies. *Twin Res Hum Genet* **18**, 158-170, doi:10.1017/thg.2015.8 (2015).

3 International Consortium for Blood Pressure Genome-Wide Association, S. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-109, doi:10.1038/nature10405 (2011).

4 Cho, H. M., Lee, H. A., Kim, H. Y., Han, H. S. & Kim, I. K. Expression of Na+-K+ -2Cl- cotransporter 1 is epigenetically regulated during postnatal development of hypertension. *Am J Hypertens* **24**, 1286-1293, doi:10.1038/ajh.2011.136 (2011).

5 Lee, H. A. *et al.* Promoter hypomethylation upregulates Na+-K+-2Cl- cotransporter 1 in spontaneously hypertensive rats. *Biochem Biophys Res Commun* **396**, 252-257, doi:10.1016/j.bbrc.2010.04.074 (2010).

6 Purkait, P. *et al.* Association of angiotensinogen gene SNPs and haplotypes with risk of hypertension in eastern Indian population. *Clin Hypertens* **23**, 12, doi:10.1186/s40885-017-0069-x (2017).

7 Bogdarina, I., Welham, S., King, P. J., Burns, S. P. & Clark, A. J. Epigenetic modification of the renin-angiotensin system in the fetal programming of hypertension. *Circ Res* **100**, 520-526, doi:10.1161/01.RES.0000258855.60637.58 (2007).

8 Tan, Q., Christiansen, L., von Bornemann Hjelmborg, J. & Christensen, K. Twin methodology in epigenetic studies. *J Exp Biol* **218**, 134-139, doi:10.1242/jeb.107151 (2015).

9 Zhao, J., Goldberg, J., Bremner, J. D. & Vaccarino, V. Association between promoter methylation of serotonin transporter gene and depressive symptoms: a monozygotic twin study. *Psychosom Med* **75**, 523-529, doi:10.1097/PSY.0b013e3182924cf4 (2013).

10 Richard, M. A. *et al.* DNA Methylation Analysis Identifies Loci for Blood Pressure Regulation. *American journal of human genetics* **101**, 888-902, doi:10.1016/j.ajhg.2017.09.028 (2017).

11 Girotto, G. *et al.* Linkage study and exome sequencing identify a BDP1 mutation associated with hereditary hearing loss. *PLoS One* **8**, e80323, doi:10.1371/journal.pone.0080323 (2013).

12 Li, X. P. *et al.* Identification of potential biomarkers for artificial cold exposure-induced hypertensive stroke by proteomic analysis. *J Stroke Cerebrovasc Dis* **23**, 2671-2680, doi:10.1016/j.jstrokecerebrovasdis.2014.06.014 (2014).

13 Chao, Y. X. *et al.* Essential tremor linked TENM4 mutation found in healthy Chinese individuals. *Parkinsonism Relat Disord* **31**, 139-140, doi:10.1016/j.parkreldis.2016.05.003 (2016).

14 Hor, H. *et al.* Missense mutations in TENM4, a regulator of axon guidance and central myelination, cause essential tremor. *Hum Mol Genet* **24**, 5677-5686, doi:10.1093/hmg/ddv281 (2015).

15 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**, 1282-1293, doi:10.1038/ng.3405 (2015).

16 Kaminsky, Z. *et al.* Epigenetics of personality traits: an illustrative study of identical twins discordant for risk-taking behavior. *Twin Res Hum Genet* **11**, 1-11, doi:10.1375/twin.11.1.1 (2008).

17 Bell, J. T. & Spector, T. D. A twin approach to unraveling epigenetics. *Trends Genet* **27**, 116-125, doi:10.1016/j.tig.2010.12.005 (2011).

18 Bock, C. Analysing and interpreting DNA methylation data. *Nat Rev Genet* **13**, 705-719, doi:10.1038/nrg3273 (2012).

19 Noble, W. S. How does multiple testing correction work? *Nat Biotechnol* **27**, 1135-1137, doi:10.1038/nbt1209-1135 (2009).

20 Wang, T. *et al.* A systematic study of normalization methods for Infinium 450K methylation data using whole-genome bisulfite sequencing data. *Epigenetics* **10**, 662-669, doi:10.1080/15592294.2015.1057384 (2015).

21 Talens, R. P. *et al.* Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology. *FASEB J* **24**, 3135-3144, doi:10.1096/fj.09-150490 (2010).

22 Zhang, X. X., He, Z., Feng, B. & Shao, H. An epigenome-wide DNA methylation study of workers with an occupational exposure to lead. *J Appl Toxicol*, doi:10.1002/jat.3816 (2019).

23 Wang, B. *et al.* Determination of Zygosity in Adult Chinese Twins Using the 450K Methylation Array versus Questionnaire Data. *Plos One* **10**, e0123992 (2015).

24 Becker, A. *et al.* Twin zygosity. Automated determination with microsatellites. *J Reprod Med* **42**, 260-266 (1997).

25 Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-1369, doi:10.1093/bioinformatics/btu049 (2014).

26 McCartney, D. L. *et al.* Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genom Data* **9**, 22-24, doi:10.1016/j.gdata.2016.05.012 (2016).

27 Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, 1-25, doi:10.2202/1544-6115.1027 (2004).

28 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).

29 Barfield, R. T. *et al.* Accounting for population stratification in DNA methylation studies. *Genet Epidemiol* **38**, 231-241, doi:10.1002/gepi.21789 (2014).

30 Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* **3**, 1724-1735, doi:10.1371/journal.pgen.0030161 (2007).

31 Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883, doi:10.1093/bioinformatics/bts034 (2012).

32 Bell, J. T. *et al.* Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* **8**, e1002629, doi:10.1371/journal.pgen.1002629 (2012).

33 Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48, doi:10.1186/1471-2105-10-48 (2009).

Tables

Table 1 Blood pressure difference test among discovery population and replication population

Blood pressure	Hypertension twins	Non-hypertension co-twins	Mean of difference	t value	*P* value
discovery population	43 twins	43 twins			
Adjusted SBP (mm Hg)	144.9	126.0	18.9	9.77	1.12E-12
Adjusted DBP (mm Hg)	88.7	77.2	11.5	8.37	8.51E-11
Adjusted PP (mm Hg)	56.2	48.8	7.4	4.29	3.76E-05
Adjusted MAP (mm Hg)	107.4	94.5	13.9	10.21	2.94E-13
BMI (kg/m^2)	25.2	24.3	0.9	3.20	2.63E-03
replication population 1	76 twins	76 twins			
Adjusted SBP (mm Hg)	159.0	120.2	38.8	14.07	<2.20E-16
Adjusted DBP (mm Hg)	95.6	71.9	23.8	16.34	<2.20E-16
Adjusted PP (mm Hg)	63.4	48.4	15.0	6.86	8.46E-10
Adjusted MAP (mm Hg)	116.8	88.0	28.8	17.0	<2.20E-16
BMI (kg/m^2)	24.8	22.6	2.2	4.30	5.12E-05
replication population 2	21 twins	21 twins			
Adjusted SBP (mm Hg)	148.0	125.6	22.4	5.65	7.78E-06
Adjusted DBP (mm Hg)	87.8	75.7	12.0	4.87	4.60E-05
Adjusted PP (mm Hg)	107.9	92.3	15.5	5.71	6.83E-06
Adjusted MAP (mm Hg)	60.2	49.9	10.4	3.50	1.11E-03
BMI (kg/m^2)	25.9	24.4	2.5	4.30	5.12E-05

MZ, monozygotic; SBP, Systolic blood pressure; DBP, diastolic blood pressure; MAP, mean arterial pressure; PP, pulse pressure; BMI, body mass index; the test was one-sided for blood pressure, while two-sided for BMI.

Table 2 Paired model for hypertension-discordant MZ twins

Model	Model 1					Model 2				
Blood pressure difference (mm Hg)	-	≥10	≥20	≥30	≥40	-	≥10	≥20	≥30	≥40
Pairs of MZ	43	36	25	16	10	43	36	25	16	10
Most significant *P* value	3.9E-07	1.9E-06	3.7E-08	1.8E-07	1.3E-06	9.6E-07	8.7E-07	1.0E-07	2.4E-06	1.2E-05
Most significant FDR	3.2E-01	6.7E-01	1.7E-02	9.8E-02	4.6E-01	7.5E-01	4.7E-01	6.7E-02	6.9E-01	8.2E-01
Significant sites	0	0	3	0	0	0	0	1	0	0

MZ, monozygotic; blood pressure difference, SBP difference add DBP difference; FDR, false discovery rate.

Model 1 adjusted smoking, alcohol consumption, and SVA agent variables; model 2 additionally adjusted BMI and recalculated SVA agent variables.

Table 3 Statistically significant sites in the paired model for hypertension-discordant MZ twins

Significant site	Target gene	Model 1 (blood pressure difference ≥20mm Hg)		Model 2 (blood pressure difference ≥20mm Hg)	
		P value	FDR	*P* value	FDR
cg00950476	*LINC01252*	3.71E-08	1.74E-02	2.63E-06	5.38E-01
cg08041400	*BDP1*	4.27E-08	1.74E-02	1.65E-07	6.74E-02
cg26733338	*SYT1*	9.07E-08	2.47E-02	1.37E-05	9.72E-01
cg08580087	*ODZ4*	7.22E-05	1.76E-01	1.02E-07	6.74E-02

MZ, monozygotic; blood pressure difference, SBP difference add DBP difference; FDR, false discovery rate.

Model 1 adjusted smoking, alcohol consumption, and SVA agent variables; model 2 additionally adjusted BMI and recalculated SVA agent variables.

Table 4 Replication analysis of suspected significant sites in replication population 1

CpG	Target gene	Paired t-test	Model 1	Model 2
cg00950476	*LINC01252*	9.51E-01	9.78E-01	7.69E-01
cg08041400	*BDP1*	4.58E-01	3.73E-02	2.22E-01
cg26733338	*SYT1*	5.26E-01	7.00E-01	4.82E-01
cg08580087	*ODZ4*	8.82E-01	7.79E-01	8.15E-01

Model 1 adjusted smoking, alcohol consumption; model 2 additionally adjusted BMI.

Table 5 Enrichment analysis of suspected significant sites

GO term	Description	*P* value	FDR	N	B	n	b	Genes
GO:0098746	fast, calcium ion-dependent exocytosis of neurotransmitter	1.56E-4	1	19189	1	3	1	*SYT1*
GO:0060912	cardiac cell fate specification	1.56E-4	1	19189	1	3	1	*ODZ4*
GO:0032289	central nervous system myelin formation	4.69E-4	1	19189	3	3	1	*ODZ4*
GO:0099502	calcium-dependent activation of synaptic vesicle fusion	4.69E-4	1	19189	3	3	1	*SYT1*
GO:0031632	positive regulation of synaptic vesicle fusion to presynaptic active zone membrane	4.69E-4	1	19189	3	3	1	*SYT1*

GO, gene ontology; *P* value, computed according to Fisher exact test; FDR, false discovery rate; N, the total number of genes; B, the total number of genes associated with a specific GO term; n, the number of target genes; b, the number of genes in the intersection.

Figures

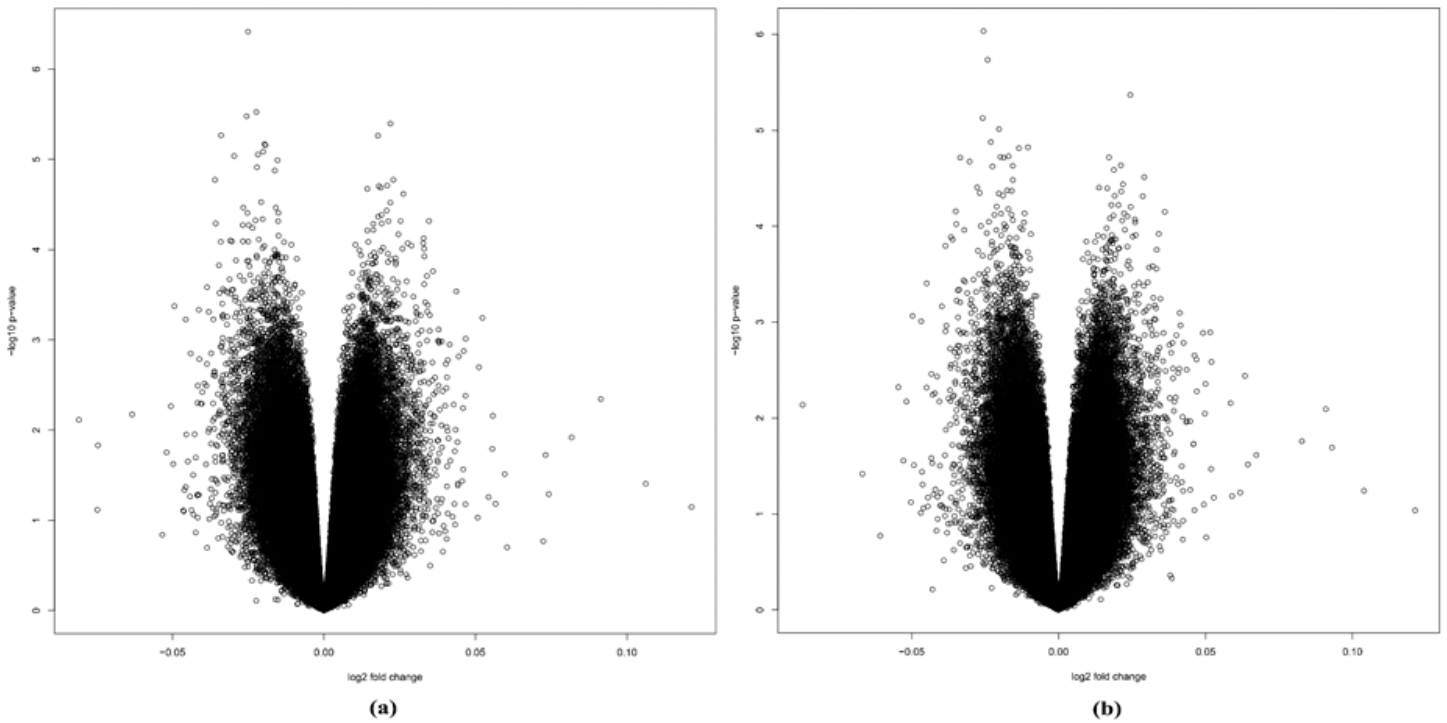

Figure 1

Volcano plot of paired model (a), volcano plot of paired model 1; (b), volcano plot of paired model 2. MZ, monozygotic; the horizontal axis represents the value of log2(methylation level in the hypertension twins/ methylation level in the non-hypertension co-twins) on each CpG sites, and the vertical axis represents the −log10(P value of paired model).

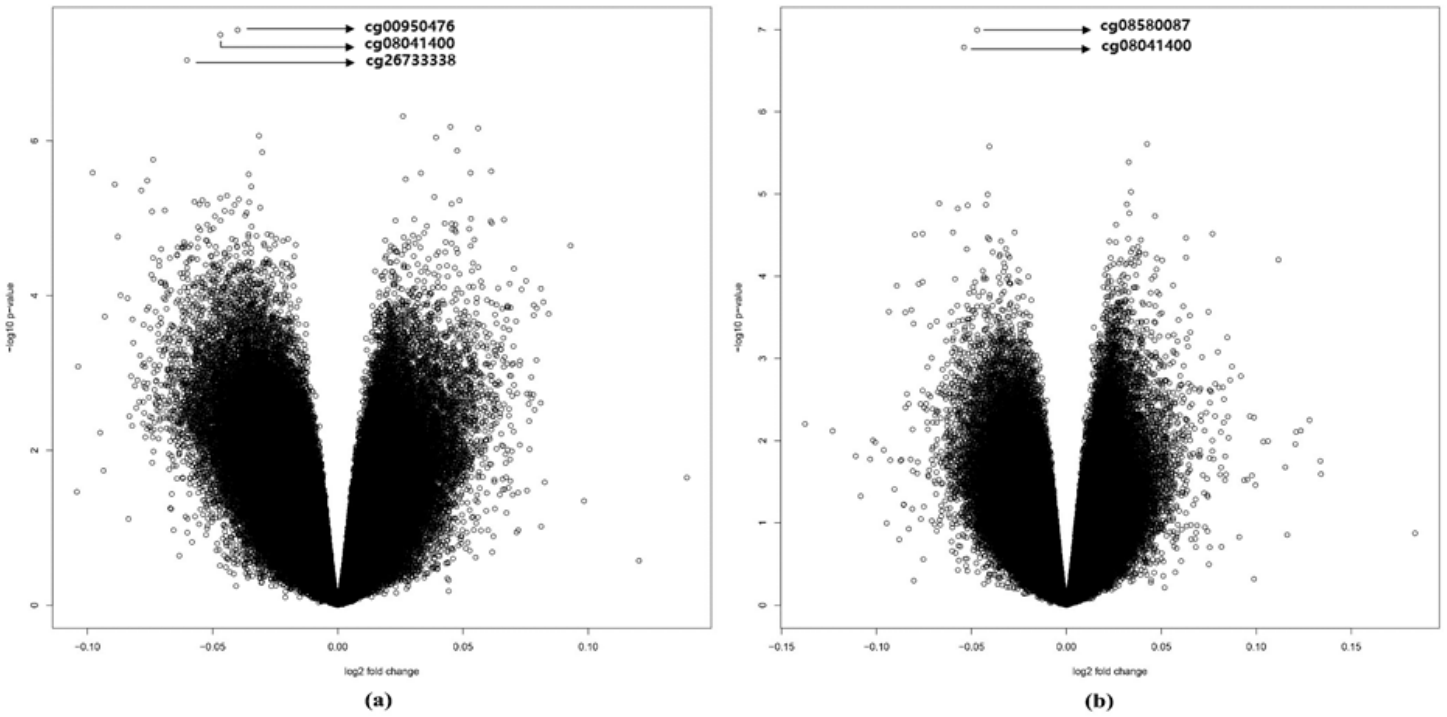

Figure 2

Volcano plot of the paired model with blood pressure difference ≥20 mm Hg (a), volcano plot of paired model 1; (b), volcano plot of paired model 2. MZ, monozygotic; the horizontal axis represents the value of log2(methylation level in the hypertension twins/ methylation level in the non-hypertension co-twins) on each CpG sites, and the vertical axis represents the −log10(P value of paired model).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf