

成年双生子血压水平与 DNA 甲基化相关性研究

吴震天¹, 高文静^{1△}, 王碧琦¹, 曹卫华¹, 吕 筠¹, 余灿清¹, 逢增昌², 丛黎明³, 汪 华⁴, 吴先萍⁵, 李立明¹

(1. 北京大学公共卫生学院流行病学与卫生统计学系, 北京 100191; 2. 青岛市疾病预防控制中心, 山东青岛 266033; 3. 浙江省疾病预防控制中心, 杭州 310051; 4. 江苏省疾病预防控制中心, 南京 210009; 5. 四川省疾病预防控制中心, 成都 610041)

[摘要] **目的:**在成年双生子人群中探索与血压指标(收缩压、舒张压、平均动脉压、脉压)存在相关性的 DNA 甲基化位点。**方法:**研究人群来自中国双生子登记系统,共 476 名双生子,问卷调查包括一般人口学特征、生活方式及疾病状况等信息,体格检查包括血压、身高、体重等信息,使用 Infinium HumanMethylation450 BeadChip 芯片对外周全血进行全基因组 DNA 甲基化检测。在调整潜在混杂因素的基础上,通过构建混合效应模型在全基因组范围寻找与血压指标存在相关性的 DNA 甲基化位点,显著性水平为错误发现阳性位点率 < 0.05 。**结果:**经过数据质量控制最终纳入 465 名双生子(122 对同卵双生子,104 对异卵双生子,13 对双生子的其中之一),年龄 (44.8 ± 13.2) 岁,男性多于女性,同卵略多于异卵,目前吸烟和目前规律饮酒者所占比例均大于 30%。所有双生子个体均行全基因组 DNA 甲基化与血压指标相关分析,经多重校正后未发现显著的甲基化位点,但位于 10 号染色体的 cg07761116 在 3 个血压指标(收缩压、舒张压、平均动脉压)相关分析中的 P 值相对较小,提示其可能是一个与血压相关的位点。还有 7 个位点在两个血压指标相关分析中的 P 值较小,所在基因与神经发育、蛋白质稳态、炎症反应等功能相关。**结论:**没有明确证据支持与血压水平存在相关性的甲基化位点,可能由于样本量不足等原因,可以为后续开展类似的双生子研究提供参考,后续研究可以关注 10 号染色体上的 cg07761116 及其他 P 值较小的位点。

[关键词] 血压;DNA 甲基化;双生子研究

[中图分类号] R181.3 **[文献标志码]** A **[文章编号]** 1671-167X(2018)03-0387-08

doi:10.3969/j.issn.1671-167X.2018.03.001

Correlation between blood pressure and DNA methylation in adult twins

WU Zhen-tian¹, GAO Wen-jing^{1△}, WANG Bi-qi¹, CAO Wei-hua¹, LV Jun¹, YU Can-qing¹, PANG Zeng-chang², CONG Li-ming³, WANG Hua⁴, WU Xian-ping⁵, LI Li-ming¹

(1. Department of Epidemiology and Biostatistics, Peking University School of Public Health, Beijing 100191, China; 2. Qingdao Municipal Center for Disease Control and Prevention, Qingdao 266033, Shandong, 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)

ABSTRACT Objective: To explore the DNA methylation sites correlated with blood pressure (systolic blood pressure, diastolic blood pressure, mean arterial pressure, pulse pressure) in adult twin population. **Methods:** A total of 476 twins from the Chinese National Twin Registry were selected as the research population. Questionnaires were used to collect demographic characteristics, lifestyle, disease status and other information, and blood pressure, height, weight and other anthropometric indicators were measured. The genome-wide DNA methylation of whole blood samples was detected by using Infinium HumanMethylation450 BeadChip. The DNA methylation sites correlated with blood pressure were analyzed by constructing mixed effect model with adjusting potential confounding factors, and the significant level was false discovery rate < 0.05 . **Results:** After data quality control, 465 twins (122 pairs of monozygotic twins, 104 pairs of dizygotic twins, 13 individuals from 13 pairs of twins) aged (44.8 ± 13.2) years were finally enrolled. There were more males and more monozygotic twins, and the current smokers and current regular drinkers both accounted for more than 30%. No significant CpG site was found after multiple testing in the correlation study between genome-wide DNA methylation and blood pressure by using the collected twins. However, the cg07761116 located on chromosome 10 had low P value in the correlation analysis of 3 blood pressure indices (systolic blood pressure, diastolic blood pressure, mean arterial pressure), suggesting that this site might be correlated with blood pressure. The

基金项目:公益性行业科研专项(201502006、201002007)和国家自然科学基金(81573223)资助 Supported by the Special Fund for Health Scientific Research in the Public Welfare (201502006, 201002007) and the National Natural Science Foundation of China (81573223)

△ Corresponding author's e-mail, pkupegwj@126.com

网络出版时间:2018-5-8 11:19:26 网络出版地址: <http://www.cnki.net/kcms/detail/11.4691.R.20180508.1119.046.html>

other 7 sites had low P value in the correlation analysis of the two blood pressure indices, respectively, which pointed to genes involved in neurological development, protein homeostasis, inflammatory reaction and other pathways. **Conclusion:** There is no sufficient evidence to support any DNA methylation site correlated with blood pressure, which may be caused by insufficient sample size and other reasons. This study could provide a reference for subsequent similar twin studies, and subsequent studies can focus on the cg07761116 located on chromosome 10 and other sites with low P values.

KEY WORDS Blood pressure; DNA methylation; Twin study

血压指标一般包括收缩压 (systolic blood pressure, SBP)、舒张压 (diastolic blood pressure, DBP)、平均动脉压 (mean arterial pressure, MAP) 和脉压 (pulse pressure, PP)。一个心动周期中的最高压力为 SBP, 最低压力为 DBP, 每一心动周期中的动脉血压平均值称为 MAP, SBP 与 DBP 的差值称为 PP, $MAP = (SBP + 2 \times DBP) / 3$, $PP = SBP - DBP$, MAP 与 PP 分别与外周血管阻力及大动脉弹性功能密切相关, 具有重要的病理生理意义。目前世界卫生组织和我国推荐的高血压定义一致, 即在未使用降压药物的情况下, 非同日 3 次测量血压, 平均 SBP ≥ 140 mmHg (1 mmHg = 0.133 kPa) 和/或平均 DBP ≥ 90 mmHg^[1]。患者既往有高血压史, 目前正在使用降压药物, 血压虽然低于 140/90 mmHg 也诊断为高血压。

高血压作为一种复杂性疾病, 涉及多种遗传、环境因素, 并且基因-环境交互作用也影响高血压的发生。目前认为, 表观遗传学可能是基因-环境交互作用背后的一种机制, 在理解基因-环境-高血压表型的关系上建立了相关联系, 为高血压病因学的研究开拓了思路。DNA 甲基化 (DNA methylation) 是最常见、研究最多的表观遗传现象^[2]。近年来, 随着甲基化检测技术由部分位点向全基因组发展, 许多研究对大规模人群的 DNA 样本进行了全基因组甲基化分析, 已发现的甲基化位点主要作用于以下两条通路: 肾素-血管紧张素-醛固酮系统 (renin-angiotensin-aldosterone system, RAAS) 和肾钠潴留系统 (renal sodium retention system, RSRS)^[3-5]。双生子人群由于基因和环境高度匹配, 有助于提高统计效力、节约样本量, 是高血压 DNA 甲基化研究的宝贵资源^[6], 如果双生子之间血压存在差异, 由于基因结构及共享环境因素高度相似, 那么血压的差异很可能是表观遗传现象或非共享环境的差异导致的。我国目前缺少此类研究证据, 因此本研究旨在成年双生子人群中探索与血压指标相关的 DNA 甲基化位点。

1 资料与方法

1.1 研究对象

本研究对象来自于中国双生子登记系统 (Chinese national twin registry, CNTR), 分布于青岛、江苏、浙江和四川 4 省市, 根据下述纳入、排除标准共收集双生子 476 人, 数据收集于 2013 年, 双生子均签署知情同意书, 本研究经北京大学生物医学伦理委员会审查并通过 (批准号: IRB00001052-13022)。

纳入标准: (1) 年龄 18 岁及以上非多胞胎双生子, (2) 血压相关指标及 DNA 甲基化数据完整者。

排除标准: (1) 分开抚养的双生子, (2) 自报患有肿瘤、冠心病 (包括心绞痛、心肌梗死发作、急性冠状动脉综合征等) 和/或脑卒中疾病病史, (3) 妊娠期妇女。

1.2 研究内容

本研究的问卷信息包括: 一般信息 (性别、年龄、地区等), 社会经济状况 (受教育水平等), 疾病相关信息 (患病情况、用药史等), 行为信息 (吸烟、饮酒、体力活动、膳食、睡眠等) 及其他信息 (女性生育史等)。问卷信息由经过培训的调查员面对面访问获得, 其中高血压患病情况及用药史均为自报, 问题是“是否曾在区、县级及以上医院明确诊断患有高血压”, 以及“是否在近 1 个月服用任何降血压药物”。

体格检查信息包括: 收缩压、舒张压、身高、体重、腰围等指标, 由工作人员使用统一的仪器测量, 其中血压采用欧姆龙 HEM-7200 电子血压计测量, 要求间隔 2 ~ 3 min 测量两次血压, 若前后 2 次 SBP/DBP 相差大于 10 mmHg 则需要进行第 3 次测量, 取最接近的两次测量值的平均值用作数据分析。

双生子卵型信息利用 Human Omni ZhongHua-8 BeadChip 芯片全基因组基因型检测, 根据基因数据进行鉴定。

双生子的空腹外周全血使用 Infinium Human-Methylation450 BeadChip 芯片 (Illumina 公司, 简称 450K 芯片) 进行 DNA 甲基化检测, 该芯片覆盖 485 553 个 CpG (胞嘧啶-磷酸-鸟嘌呤) 位点。外周全血标本储存于 -80 °C 冰箱, 经水浴解冻后使用 BioTeke 全血 DNA 提取试剂盒 (BioTeke whole blood DNA extraction kit) 提取 DNA。经过电泳和紫外分光光度计测定, 提取后的 DNA 样本的纯度

($D_{280\text{ nm}/260\text{ nm}}$ 为 1.5 ~ 2.0) 和浓度 (50 ~ 200 mg/L) 均满足实验要求。DNA 样本经过亚硫酸氢盐转换、全基因组扩增、DNA 片段化、杂交、洗脱、延伸及成像过程形成甲基化数据。

1.3 统计学分析

本研究数据采用 R 3.4.1 统计软件进行统计学分析。利用 R 软件 minfi 程序包读取 CpG 位点的红绿信号 IDAT 文件,并转换成 CpG 位点的 β 值(β 值定义为该位点甲基化信号占甲基化和非甲基化信号之和的比例),再进行样本质控和位点质控^[7]。

在进行血压指标与甲基化的相关分析时,对于自报正在服用降压药的研究对象,根据既往文献需对其血压值进行如下调整^[8]: SBP + 15 mmHg, DBP + 10 mmHg, PP + 5 mmHg, MAP + 11.7 mmHg。进行 DNA 甲基化检测时,不同的检测平板、实验批次以及操作人员、操作技术和仪器、试剂等实验因素的差异,会使得到的 DNA 甲基化数据存在一定的误差,需要使用代理变量分析方法(surrogate variable analysis, SVA)调整潜在混杂因素^[9-10],使用 R 软件 sva 程序包可生成批次效应等潜在混杂因素的代理变量^[11]。

为控制双生子同胞之间的相关性(基因结构及共享环境因素),本研究采用混合效应模型(mixed effect model, ME)构建回归方程,以血压指标的 lg 值作为连续型自变量, DNA 甲基化水平作为因变量,调整潜在混杂因素包括年龄、性别、卵型、双生子出生顺序、体重指数(body mass index, BMI)、吸烟、饮酒、蔬菜、水果、红肉、高盐饮食、体力活动、睡眠时间、教育水平、调查地区以及 SVA 作为固定效应,双生子对编号作为随机效应,使用 R 软件 nlme 程序包中 lme 函数构建如下的回归方程^[12]: CpG 位点 β 值 = lg 血压指标(SBP/DBP/MAP/PP) + 年龄 + 性别 + 卵型 + 双生子出生顺序 + BMI + 吸烟 + 饮酒 + 蔬菜 + 水果 + 红肉 + 高盐饮食 + 体力活动 + 睡眠时间 + 教育水平 + 调查地区 + SVA(以上变量均为固定效应) + 双生子对编号(随机效应)。

因甲基化分析位点较多,需要多重检验校正,上述分析的显著性水平定为错误发现阳性位点率(false discovery rate, FDR) < 0.05, FDR 估计按照 Benjamini & Hochberg 法^[13]。

2 结果

2.1 数据质控

进行数据分析前需要对甲基化数据进行样本质控和位点质控。

样本质控包括:(1) Illumina 公司推荐形成的质控报告图(未发现异常数据);(2)根据 450K 芯片中 65 个单核苷酸多态性(single nucleotide polymorphisms, SNP)位点与调查对象已有全基因组基因数据比对,检查样本污染或加错样本情况,剔除样本 8 个;(3)利用 MDS(multiple dimension-scale)图检验离群值(未发现离群值);(4)样本检出率(与芯片内置空白对照相比,某个位点的检出 $P > 0.05$, 则该位点检出缺失,样本检出率即该样本中检出非缺失的位点数占所有位点数的比例) < 0.99, 剔除样本 3 个。

位点质控包括:(1)位点检出率(与芯片内置空白对照相比,某个位点的检出 $P > 0.05$, 则该位点检出缺失,位点检出率为某个位点在所有样本非缺失的比例) < 0.99, 剔除位点 3 081 个;(2)CpG 位点上含有基因位点,基因位点在亚洲人群中最小等位基因频率(minor allele frequency, MAF) > 0.05 或与其他基因位点交叉^[14], 剔除位点 41 269 个;(3)剔除 65 个 SNP 位点。

最终合格样本 465 个, CpG 位点 442 836 个,再利用 R 软件 wateRmelon 程序包进行 DASEN 标准化(控制 II 型探针的测量偏倚、背景噪音偏倚和芯片间偏倚)^[15], 形成待分析的 DNA 甲基化数据。

2.2 双生子基本分布

质量控制后合格双生子样本 465 人,包括 122 对同卵双生子、104 对异卵双生子以及 13 对双生子的其中之一(1 名同卵、7 名异卵、5 名未知卵型,这些双生子其中的一个仍然可以作为个体纳入模型中进行分析)。双生子年龄均值为 44.8 岁(范围 18.0 ~ 80.9 岁),男性多于女性,目前吸烟和目前规律饮酒者在 CNTR 中所占比例均大于 30%。另外,新鲜蔬菜、新鲜水果摄入量和体力活动量呈偏态分布,大部分研究对象集中在摄入量或体力活动量较低的水平,为减小偏态分布对结果的影响,将这 3 个变量进行五等分后再作为协变量分析,其他特征描述见表 1。

2.3 血压水平与全基因组 DNA 甲基化相关性研究

将 465 名双生子作为个体,以血压指标(SBP、DBP、MAP、PP)作为连续型自变量, DNA 甲基化水平作为因变量,纳入相关协变量用混合效应模型进行分析,找出与血压相关的 CpG 位点。

DNA 甲基化研究分析阶段需要利用 SVA 控制人群分层可能造成的影响,从本研究的 q-q 图(图 1)可以看出观测 P 值基本未偏离期望 P 值,基因控制因子(genomic control factor, λ) 均接近 1 ($\lambda_{\text{SBP}} =$

1.02, $\lambda_{DBP} = 0.99$, $\lambda_{MAP} = 1.00$, $\lambda_{PP} = 1.02$), 说明血压指标与甲基化相关分析的结果已经较好控制人群分层可能的影响。

表 1 双生子一般人口学特征、生活方式、血压指标及社会经济状况

Table 1 Demographic characteristics, lifestyle, blood pressure, and socioeconomic status in adult twins

Variable	$\bar{x} \pm s$ or n (%)
Age/years	44.8 ± 13.2
Men	308 (66.2)
Monozygotic twins	245 (52.7)
SBP/mmHg	133.2 ± 22.2
DBP/mmHg	79.4 ± 12.4
MAP/mmHg	97.3 ± 14.7
PP/mmHg	53.8 ± 15.2
BMI/(kg/m ²)	25.0 ± 3.8
Current smokers	148 (31.8)
Current regular drinkers	161 (34.6)
Intake of fresh vegetables/(g/week)	1 566.3 ± 951.7
Intake of fresh fruits/(g/week)	1 093.2 ± 1 104.1
Intake of red meat	
Less than once a week	72 (15.5)
1 - 3 times a week	117 (25.2)
4 - 6 times a week	52 (11.2)
Once a day	137 (29.5)
≥ 2 times a day	83 (17.8)
Intake of high salt food	
Less than once a week	256 (55.1)
1 - 3 times a week	106 (22.8)
4 - 6 times a week	31 (6.7)
Once a day	51 (11.0)
≥ 2 times a day	16 (3.4)
Amount of physical activity/(MET/week)	5 779.7 ± 5 212.3
Sleep time	
Less than 4 hours a day	4 (0.9)
4 - 6 hours a day	34 (7.3)
6 - 8 hours a day	230 (49.5)
More than 8 hours a day	196 (42.2)
Education	
Primary school not completed	51 (11.0)
Primary school	79 (17.0)
Junior school	225 (48.4)
High school	71 (15.3)
Junior college/Bachelor degree and above	39 (8.4)
Area	
Shandong	135 (29.0)
Zhejiang	129 (27.7)
Jiangsu	151 (32.5)
Sichuan	50 (10.8)
Currently diagnosed with hypertension	86 (18.5)
Take medicine in last month	167 (35.9)
Take antihypertensive drugs in last month	66 (14.2)

SBP, systolic blood pressure; DBP, diastolic blood pressure; MAP, mean arterial pressure; PP, pulse pressure; BMI, body mass index; MET, metabolic equivalent of tasks. 1 mmHg = 0.133 kPa.

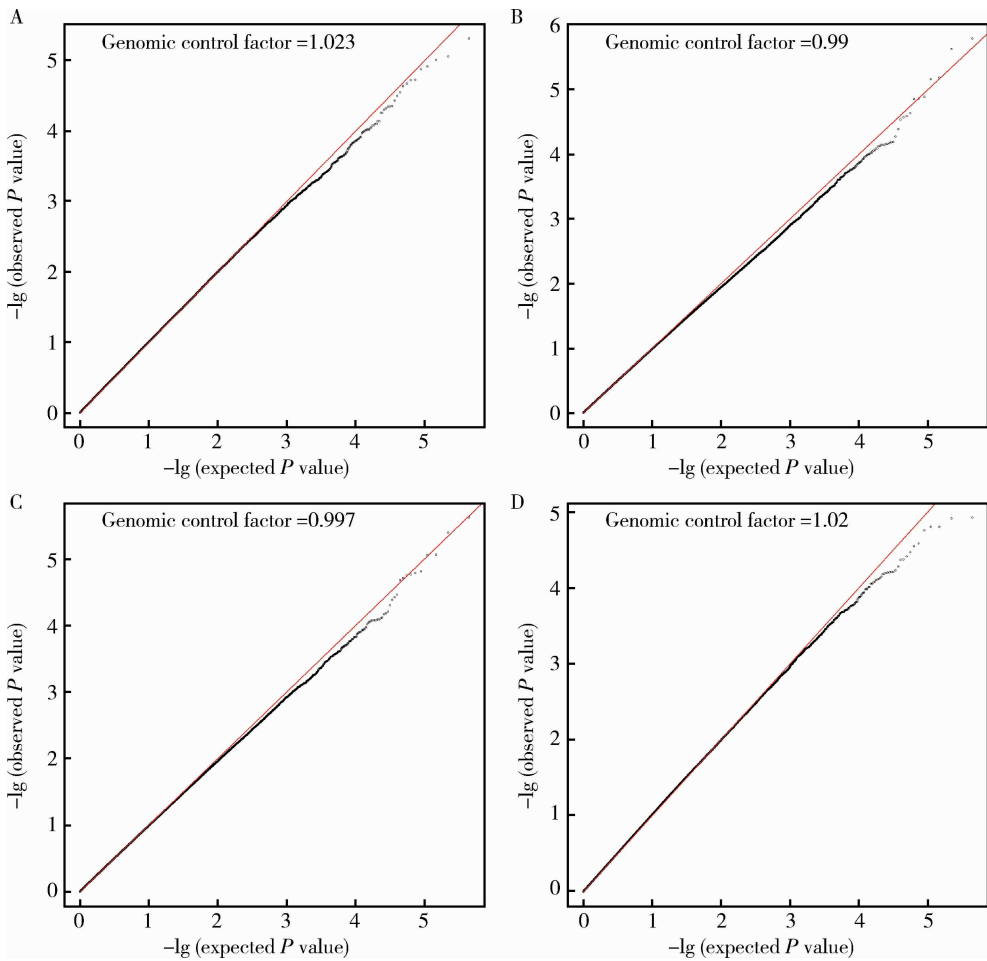
用曼哈顿图 (Manhattan plot) 描述全基因组范围上每个 CpG 位点与血压指标相关的显著性水平及其所在染色体上的位置 (图 2)。如果以 10^{-6} 为 P 值界限, 在 4 个血压指标中均未发现满足此显著性水平的 CpG 位点。

4 个血压指标的甲基化相关分析中 P 值最小位点分别为位于 11 号染色体 *NTM* 基因上的位点 cg17090716 ($P = 4.88 \times 10^{-6}$)、位于 14 号染色体 *KIF26A* 基因上的位点 cg00016406 ($P = 1.62 \times 10^{-6}$)、位于 17 号染色体 *PSMD11* 基因上的位点 cg00459043 ($P = 2.37 \times 10^{-6}$)、位于 2 号染色体 *FOXN2* 基因上的位点 cg19376844 ($P = 1.19 \times 10^{-5}$)。

4 个血压指标的甲基化相关分析 P 值最小的前 10 个位点及位置关系等信息见表 2。根据表 2 结果可见, 利用双生子构建 DNA 甲基化与血压指标混合效应模型, 按照 $FDR < 0.05$ 的显著性水平, 在 4 个血压指标 (SBP、DBP、MAP、PP) 均未发现相关阳性 CpG 位点, 但在 4 个血压指标相关分析中, P 值最小的前 10 个位点有 8 个重叠位点, 重叠的位点信息见表 3。

虽然本研究未发现 $FDR < 0.05$ 的甲基化位点, 但在 SBP、DBP 和 MAP 相关分析中, 位于 10 号染色体的位点 cg07761116 其相关分析 P 值均较小, 且随着 3 个血压指标的增加, cg07761116 甲基化水平均降低, 提示其可能是一个与血压相关的位点。

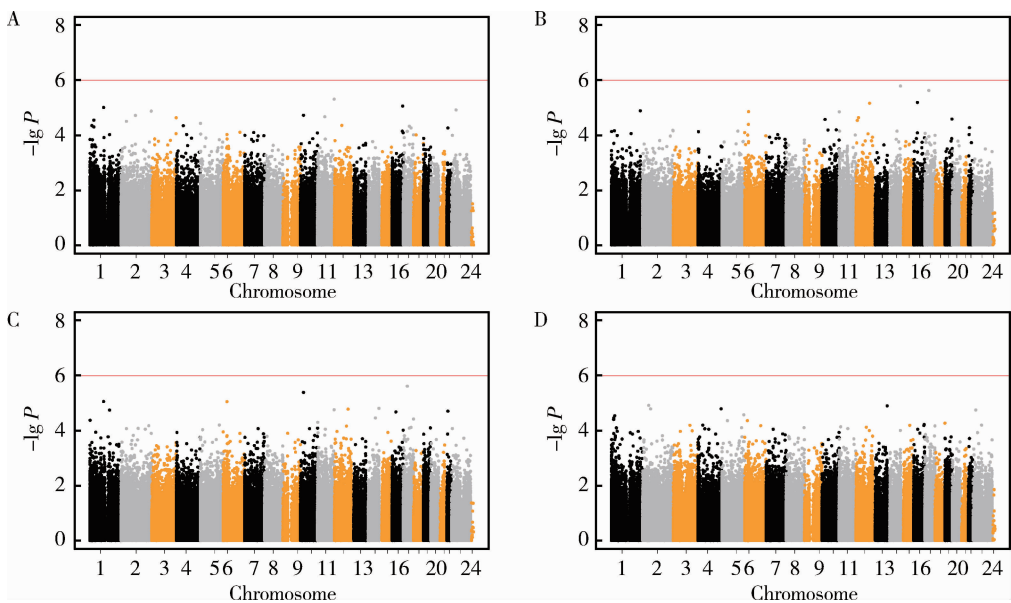
其他 7 个位点均在两个血压指标的相关分析中 P 值较小, 所在的基因及其功能简介如下: *NTM* (neurotrimin) 编码的蛋白质可以促进神经突触生长和黏附, *Clorf59* (HEN methyltransferase 1) 编码的产物是一种 RNA 甲基化转移酶, *KIF26A* (kinesin family member 26A) 编码一种在肠神经发育中起关键作用的非典型驱动蛋白, *PSMD11* (proteasome 26S subunit non-ATPase 11) 编码的是 26S 蛋白酶体的一种组分, 该蛋白酶体通过去除可能损害细胞功能的错误折叠或受损蛋白质以及不再需要功能的蛋白质而在维持蛋白质稳态中起关键作用, *MVP* (major vault protein) 编码穹窿体复合体的主要成分, *BTBD11* (BTB domain containing 11) 编码的是一种包含 BTB 结构域的蛋白, 可能与蛋白质的异二聚化活性 (protein heterodimerization activity) 有关, *HCG22* (HLA complex group 22) 编码的是一种与泛细支气管炎相关的黏蛋白样蛋白。



A, q-q plot of DNA methylation-wide correlation study with SBP; B, q-q plot of DNA methylation-wide correlation study with DBP; C, q-q plot of DNA methylation-wide correlation study with MAP; D, q-q plot of DNA methylation-wide correlation study with PP. Abbreviations as in Table 1.

图 1 全基因组 DNA 甲基化与血压指标相关分析 q-q 图

Figure 1 The q-q plot of DNA methylation-wide correlation study of blood pressure



A, Manhattan plot of DNA methylation-wide correlation study with SBP; B, Manhattan plot of DNA methylation-wide correlation study with DBP; C, Manhattan plot of DNA methylation-wide correlation study with MAP; D, Manhattan plot of DNA methylation-wide correlation study with PP. The horizontal axis represents the 24 chromosomes and their location, where 23 is the X chromosome and 24 is the Y chromosome. The ordinate axis represents the $-\lg P$. The red horizontal line represents the P value equals to 10^{-6} . Abbreviations as in Table 1.

图 2 全基因组 DNA 甲基化与血压相关的曼哈顿图

Figure 2 Manhattan plot of DNA methylation-wide correlation study of blood pressure

表 2 全基因组 DNA 甲基化与血压指标相关分析

Table 2 Correlation analysis of DNA methylation-wide and blood pressure

CpG site	Chromosome	Chromosomal coordinate	Target gene	Gene region	Coefficient	Standard error	P value	FDR
SBP								
cg17090716 *	11	131778850	<i>NTM</i>	Body	0.053	0.010 9	4.88×10^{-6}	0.99
cg00961177	16	84213833	<i>TAF1C</i>	Body	-0.019	0.004 0	8.70×10^{-6}	0.99
cg21689075 *	1	109192664	<i>Clorf59</i>	Body	0.029	0.006 2	9.79×10^{-6}	0.99
cg20448001	X	38424760	<i>TSPAN7</i>	Body	0.063	0.013 6	1.20×10^{-5}	0.99
cg00094744	2	239927431	-	-	-0.071	0.015 5	1.33×10^{-5}	0.99
cg07761116 *	10	23005837	-	-	-0.050	0.011 0	1.88×10^{-5}	0.99
cg01467417	2	113874068	<i>ILIRN</i>	TSS1500	-0.018	0.004 1	1.91×10^{-5}	0.99
cg11376147	11	57261198	<i>SLC43A1</i>	Body	-0.018	0.004 1	2.13×10^{-5}	0.99
cg10237854	3	195539463	<i>MUC4</i>	TSS1500	0.022	0.004 9	2.32×10^{-5}	0.99
cg08847173	1	31654792	<i>NKAIN1</i>	Body	0.150	0.034 5	2.83×10^{-5}	0.99
DBP								
cg00016406 *	14	104643443	<i>KIF26A</i>	Body	0.016	0.003 2	1.62×10^{-6}	0.52
cg00459043 *	17	30771460	<i>PSMD11</i>	TSS200	-0.016	0.003 2	2.36×10^{-6}	0.52
cg27575052 *	16	29855904	<i>MVP</i>	Body	0.026	0.005 5	6.46×10^{-6}	0.76
cg13492159 *	12	107714826	<i>BTBD11</i>	Body	0.012	0.002 4	6.88×10^{-6}	0.76
cg20703242	1	230279135	<i>GALNT2</i>	Body	0.044	0.009 6	1.29×10^{-5}	0.89
cg17038049 *	6	31021365	<i>HCG22</i>	TSS1500	-0.037	0.008 2	1.39×10^{-5}	0.89
cg22747142	11	795582	<i>SLC25A22</i>	5'UTR	-0.036	0.007 8	1.41×10^{-5}	0.89
cg11802843	12	19441930	<i>PLEKHA5</i>	Body	-0.043	0.009 7	2.29×10^{-5}	1.00
cg23956238	19	55672036	<i>Cl9orf51</i>	Body	0.016	0.003 7	2.59×10^{-5}	1.00
cg07761116 *	10	23005837	-	-	-0.046	0.010 5	2.68×10^{-5}	1.00
MAP								
cg00459043 *	17	30771460	<i>PSMD11</i>	TSS200	-0.018	0.003 5	2.37×10^{-6}	0.89
cg07761116 *	10	23005837	-	-	-0.056	0.011 4	4.01×10^{-6}	0.89
cg21689075 *	1	109192664	<i>Clorf59</i>	Body	0.031	0.006 6	8.59×10^{-6}	0.91
cg17038049 *	6	31021365	<i>HCG22</i>	TSS1500	-0.042	0.009 0	8.71×10^{-6}	0.91
cg00016406 *	14	104643443	<i>KIF26A</i>	Body	0.016	0.003 6	1.52×10^{-5}	0.91
cg13492159 *	12	107714826	<i>BTBD11</i>	Body	0.012	0.002 7	1.62×10^{-5}	0.91
cg17090716 *	11	131778850	<i>NTM</i>	Body	0.052	0.011 5	1.71×10^{-5}	0.91
cg07022769	1	157162190	-	-	-0.046	0.010 1	1.76×10^{-5}	0.91
cg02948259	22	24988020	<i>C22orf36</i>	Body	-0.035	0.007 7	1.93×10^{-5}	0.91
cg27575052 *	16	29855904	<i>MVP</i>	Body	0.027	0.006 1	2.06×10^{-5}	0.91
PP								
cg19376844	2	48543191	<i>FOXP2</i>	5'UTR	0.019	0.004 1	1.19×10^{-5}	0.97
cg05668422	13	114026346	-	-	0.041	0.008 8	1.23×10^{-5}	0.97
cg02552216	4	184426031	<i>ING2</i>	TSS200	0.015	0.003 2	1.58×10^{-5}	0.97
cg04342863	2	62560135	-	-	0.039	0.008 6	1.59×10^{-5}	0.97
cg00772918	X	24487341	<i>PDK3</i>	Body	0.037	0.008 2	1.74×10^{-5}	0.97
cg26615160	5	174151043	<i>MSX2</i>	TSS1500	0.009	0.001 9	2.62×10^{-5}	0.97
cg01973987	1	26394300	<i>TRIM63</i>	TSS200	0.029	0.006 6	2.83×10^{-5}	0.97
cg15564444	1	18589627	<i>IGSF21</i>	Body	-0.042	0.009 8	3.42×10^{-5}	0.97
cg24738993	1	18167815	-	-	0.040	0.009 3	3.88×10^{-5}	0.97
cg10558740	6	22148738	<i>FLJ22536</i>	Body	0.021	0.005 0	4.24×10^{-5}	0.97

FDR, false discovery rate; Other abbreviations as in Table 1. Body, gene body; TSS1500, 200 - 1 500 bases upstream of the transcriptional start site; TSS200, 0 - 200 bases upstream of the transcriptional start site; 5'UTR, 5' untranslated region; “-” represents the CpG site does not target on known genes; * represents the CpG site was also among the 10 smallest P value sites in the correlation analysis of other blood pressure.

表3 全基因组 DNA 甲基化与血压指标相关分析重叠位点

Table 3 Overlapped sites in correlation analysis of DNA methylation-wide and blood pressure

CpG site	Target gene	SBP		DBP		MAP	
		<i>P</i> value	FDR	<i>P</i> value	FDR	<i>P</i> value	FDR
cg07761116	-	1.88×10^{-5}	0.99	2.68×10^{-5}	1.00	4.01×10^{-6}	0.89
cg17090716	<i>NTM</i>	4.88×10^{-6}	0.99	-	-	1.71×10^{-5}	0.91
cg21689075	<i>C1orf59</i>	9.79×10^{-6}	0.99	-	-	8.59×10^{-6}	0.91
cg00016406	<i>KIF26A</i>	-	-	1.62×10^{-6}	0.52	1.52×10^{-5}	0.91
cg00459043	<i>PSMD11</i>	-	-	2.36×10^{-6}	0.52	2.37×10^{-6}	0.89
cg27575052	<i>MVP</i>	-	-	6.46×10^{-6}	0.76	2.06×10^{-5}	0.91
cg13492159	<i>BTBD11</i>	-	-	6.88×10^{-6}	0.76	1.62×10^{-5}	0.91
cg17038049	<i>HCG22</i>	-	-	1.39×10^{-5}	0.89	8.71×10^{-6}	0.91

Abbreviations as in Table 1 and 2. “-” represents the CpG site was not among the 10 smallest *P* value sites.

3 讨论

表观遗传学是高血压遗传研究的新方向,可解释血压水平变异不能被传统孟德尔遗传定律阐明的部分^[16],已有大量研究证明,DNA 甲基化是基因变异影响血压水平的重要调控通路,目前寻找与血压有关的甲基化位点仍是研究热点。DNA 甲基化研究不仅能够帮助揭示遗传和环境对于血压的交互影响作用,解释高血压发生的机制,而且相比于基因的不可改变性,未来有可能通过改变相关的甲基化位点的甲基化水平从而对血压水平进行调控,对于高血压的干预和治疗也有重要意义。

本研究利用收集到的 476 名双生子个体进行全基因组 DNA 甲基化与血压指标相关分析,经多重校正后未发现血压指标相关阳性甲基化位点,但本研究中位于 10 号染色体的位点 cg07761116 在 3 个血压指标的相关分析中 *P* 值相对较小,提示其可能是一个与高血压发生相关的位点。本研究其他相关分析中 *P* 值较小的重叠位点所在基因的功能主要与神经系统和炎症反应有关。目前有研究表明高血压与精神紧张有关^[17],机体免疫功能紊乱是高血压的发病机制之一^[18]。2013 年一项全基因组关联研究 (genome-wide correlation studies, GWAS) 的 meta 分析发现了数十个与高血压和血压相关指标之间关联具有显著性的基因位点^[19],其中部分基因的功能与神经系统和炎症反应有关,而既往的甲基化研究也有发现神经系统^[20]和免疫系统^[21]相关的基因上的 CpG 位点与高血压存在相关,炎症相关基因的甲基化可通过调控炎症通路导致血压的改变。

利用甲基化芯片对全基因组位点进行研究,因其高覆盖性以及相对经济,成为目前甲基化研究的

重要检测手段,但是因为研究一次性关注的 CpG 位点较多,可能会发现虚假相关,所以为了控制整体假阳性必须进行多重校正。既往也有较多研究应用 450K 芯片发现了许多甲基化位点在高血压组与对照组间存在差异,但经多重校正后,并没有满足显著性水平要求的阳性位点。本研究采用常用的 FDR 标准下的 Benjamini-Hochberg 校正方法,其前提假设是所有多重比较的位点都是阴性的(即没有差异),而且位点间的关系是独立的,这显然过于严格,但即使存在一些待解决的问题,多重检验校正仍是必须进行的一步,以防止对假阳性位点不必要的验证及探索。

本研究的主要优势在于:(1)针对中国双生子人群在全基因组范围内进行血压与 DNA 甲基化相关研究,虽然经过多重校正后未能发现阳性位点,但提示某些位点可能与血压存在相关;(2)样本量充足,共收集 476 名双生子个体,该样本量在全球双生子甲基化研究中属于前列;(3)既往甲基化研究多采用一般人群,本研究采用双生子这一天然匹配人群,能更好地控制遗传因素的影响;(4)制作了标准化的现场操作手册和现场模拟视频,调查人员经过了统一培训,确保收集到的信息标准、统一、内容真实;(5)研究较充分地考虑了可能影响相关研究的混杂因素,包括年龄、性别、体力活动、饮食及家庭经济状况等,并且使用了 SVA 方法调整潜在混杂因素。

本研究的局限性:(1)DNA 甲基化水平是具有组织特异性的表观遗传现象,而本研究仅采用外周血样本进行甲基化检测,无其他器官组织的 DNA 甲基化信息,未来需要进一步验证。但是,目前有研究表明外周血 DNA 甲基化水平与其他组织的甲基化

水平存在一定相关^[8]; (2) 本研究仅采用双生子作为研究人群, 可能存在双生子对内研究中过度匹配问题以及双生子人群特有的高血压相关甲基化位点的问题^[22-24], 但在模型中将双生子作为个体分析, 一定程度上减小了上述问题; (3) 本研究中并没有收集到完整的食物摄入情况, 因此无法完全控制膳食摄入的影响, 但已经控制了常见的与血压关系较为密切的膳食摄入信息 (如红肉、高盐饮食等摄入量); (4) 本研究为横断面研究, 只能探索血压与 DNA 甲基化的相关性, 研究设计上无法确定两者发生的时序, 因此需要后续研究中结合纵向数据继续研究血压与 DNA 甲基化的关联。

综上所述, 本研究利用双生子人群, 全基因组范围内探索与血压相关的 DNA 甲基化位点, 虽然没有明确证据支持与血压水平存在相关的 DNA 甲基化位点, 但本研究作为国内利用较大样本量双生子人群进行的血压与 DNA 甲基化相关分析, 能够为后续研究提供重要的参考。

参考文献

- [1] 刘力生. 中国高血压防治指南 2010 [J]. 中华高血压杂志, 2011, 19(8): 701-743.
- [2] Holliday R. Epigenetics: an overview [J]. Dev Genet, 1994, 15(6): 453-457.
- [3] 李莉, 杨昭庆. 原发性高血压相关基因 DNA 甲基化的研究进展 [J]. 现代生物医学进展, 2017, 17(26): 5197-5200.
- [4] 翟秀伟, 胡大春. DNA 甲基化在高血压疾病中的研究 [J]. 国际检验医学杂志, 2014, 35(16): 2206-2207.
- [5] 范瑞, 钟琦珑, 毛书奇, 等. DNA 甲基化修饰与原发高血压发病关联的研究 [J]. 生命科学, 2015, 27(5): 590-594.
- [6] 王碧琦, 周斌, 高文静, 等. 双生子研究在表观遗传学研究中的价值 [J]. 中华流行病学杂志, 2015, 36(4): 402-404.
- [7] Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays [J]. Bioinformatics, 2014, 30(10): 1363-1369.
- [8] Kato N, Loh M, Takeuchi F, et al. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation [J]. Nat Genet, 2015, 47(11): 1282-1293.
- [9] Barfield RT, Almlı LM, Kilaru V, et al. Accounting for population stratification in DNA methylation studies [J]. Genet Epidemiol, 2014, 38(3): 231-241.
- [10] Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis [J]. PLoS Genet, 2007, 3(9): 1724-1735.
- [11] Leek JT, Johnson WE, Parker HS, et al. The sva package for removing batch effects and other unwanted variation in high-throughput experiments [J]. Bioinformatics, 2012, 28(6): 882-883.
- [12] Carlin JB, Gurrin LC, Sterne JA, et al. Regression models for twin studies: a critical review [J]. Int J Epidemiol, 2005, 34(5): 1089-1099.
- [13] Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing [J]. J R Statist Soc B, 1995, 57(1): 289-300.
- [14] Chen YA, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray [J]. Epigenetics, 2013, 8(2): 203-209.
- [15] Pidsley R, Wong CCY, Volta M, et al. A data-driven approach to preprocessing Illumina 450K methylation array data [J]. BMC Genomics, 2013, 14(1): 293.
- [16] Raftopoulos L, Katsi V, Makris T, et al. Epigenetics, the missing link in hypertension [J]. Life Sci, 2015, 129: 22-26.
- [17] 黄文辉, 胡丙杰, 韩柯, 等. 职业紧张与高血压关系的研究进展 [J]. 广州医药, 2008, 39(4): 4-7.
- [18] Pashkow FJ. Oxidative stress and inflammation in heart disease: Do antioxidants have a role in treatment and/or prevention? [J]. Int J Inflam, 2011, 514623. doi: 10.4061/2011/514623.
- [19] Lind JM, Chiu CL. Genetic discoveries in hypertension: steps on the road to therapeutic translation [J]. Heart, 2013, 99(22): 1645-1651.
- [20] Esler M, Eikelis N, Schlaich M, et al. Human sympathetic nerve biology: parallel influences of stress and epigenetics in essential hypertension and panic disorder [J]. Ann N Y Acad Sci, 2008, 1148: 338-348.
- [21] Alexeeff SE, Baccarelli AA, Halonen J, et al. Association between blood pressure and DNA methylation of retrotransposons and pro-inflammatory genes [J]. Int J Epidemiol, 2013, 42(1): 270-280.
- [22] Amin V, Behrman JR, Kohler HP, et al. Causal inferences: Identical twins help and clarity about necessary assumptions is critical [J]. Soc Sci Med, 2015, 127: 201-202.
- [23] Boardman JD, Fletcher JM. To cause or not to cause? That is the question, but identical twins might not have all of the answers [J]. Soc Sci Med, 2015, 127: 198-200.
- [24] Goetz M, Shan A, Goldberg J, et al. Posttraumatic stress disorder, combat exposure, and carotid intima-media thickness in male twins [J]. Am J Epidemiol, 2014, 180(10): 989-996.

(2018-02-09 收稿)

(本文编辑:任英慧)