

中国中小学生对微量营养素缺乏风险关联 SNP 位点民族分布特征

张春红^{1,2}, 霍军生¹, 孙静¹, 黄建¹, 陈晨¹

1. 中国疾病预防控制中心营养与健康所, 北京 100050; 2. 中国人民解放军海军特色医学中心

【摘要】 目的 初步探索中国中小学生对微量营养素缺乏风险的关联 SNP 位点地域特征, 为评估微量元素缺乏风险提供基础依据。**方法** 搜集以往研究中报道的 143 个 SNP 位点, 研究对象来自于 2016 年中国贫困地区营养健康调查项目义务教育中一至九年级 1 130 名中小学生的凝固血细胞, 采用磁珠法提取 DNA, 竞争性等位基因法进行分型检测。对差异基因进行 GO 显著性富集, 采用 R 软件包进行 PCA、kinship 和 SNP 位点之间连锁不平衡分析, 分析候选 SNP 位点的特征。如有种群结构的存在, 采用 FaST-LMM 模型进行关联分析。**结果** 差异表达基因在生物过程分组中主要富集在催化活性、转运活性、能量代谢途径、类固醇激素、辅酶、维生素 A、维生素 D 和水溶性维生素的代谢等生物学过程中, 涉及转录、翻译以及能量代谢相关基因。143 个 SNP 位点分型结果在民族分布上差异有统计学意义, 3 号染色体上的基因多态性位点在民族间差异有统计学意义。主成分分析 PC1 显示, 位于 *RAB6B* 基因上的 rs2280673 解释效度在 25% 以下, 位于 *TF* 基因上的 rs1799852 解释效度为 25%~50%, 位于 *RBP2* 基因上的 rs2118981 位点和 *SRPRB* 基因上的 rs1830084 解释效度在 50%~75% 之间, 位于 *TF* 基因上的 rs1358024、rs1525892、rs1880669、rs3811647、rs3811658、rs6794945、rs7638018、rs8177248 等 8 个位点的解释效度为 75% 以上。**结论** 中国中小学生对微量营养素缺乏风险关联 SNP 位点存在民族分布差异。

【关键词】 微量营养系; 儿童营养障碍; 多态性, 单核苷酸; 基因; 血液化学分析; 学生

【中图分类号】 R 151 R 446 **【文献标识码】** A **【文章编号】** 1000-9817(2020)01-0025-04

Ethnic distribution characteristics of SNPs associated with micronutrient deficiency risk of Chinese primary and middle school students/ZHANG Chunhong, HUO Junsheng, SUN Jing, HUANG Jian, CHEN Chen. National Institute for Nutrition and Health, Chinese Center for Disease Control and Prevention, Beijing(100050), China

【Abstract】 Objective To explore ethnic distribution characteristics of SNPs associated with micronutrient deficiency risk of Chinese primary and middle school students, and to provide a basic reference for evaluating the risk of lack in micronutrient. **Methods** Totally 143 SNPs reported in previous studies were collected, and DNA was extracted by using magnetic beads in frozen blood cell samples from the 2016 nutrition health survey project of 1 130 primary and middle school students, competitive allele method was used to detect SNP genotyping. GO significant enrichment analysis R software package to PCA, kinship and linkage disequilibrium analysis were used for analysis of features of candidate SNPs. If there was a population structure, the FaST-LMM model was used for correlation analysis. **Results** The GO significant enrichment results showed that differentially expressed genes were mainly enriched in the biological process grouping, including catalytic activity, transport activity, energy metabolism pathway, steroid hormone, coenzyme, biological processes of vitamin A, D and metabolism of water-soluble vitamins, involving transcription, translation and energy metabolism related genes. The results of 143 SNPs showed statistically significant differences in ethnic distribution, and SNPs on chromosome 3 presented significant differences among ethnic groups. Principal component analysis 1 showed that rs1799852 on *TF* gene had 25%–50% explanatory validity, rs2118981 on *RBP2* gene and rs1830084 on *SRPRB* gene had 50%–75% explanatory validity, rs1358024, rs1525892, rs1880669, rs3811647, rs3811658, rs6794945, rs7638018 and rs8177248 on *TF* gene had more than 75% explanatory validity. **Conclusion** SNPs associated with micronutrient deficiency risk of Chinese primary and middle school students are characterized with ethnic distributions.

【Key words】 Micronutrients; Child nutrition disorders; Polymorphism, single nucleotide; Genes; Blood chemical analysis; Students

随着人类基因组计划的完成, 人类遗传学研究进入了全基因组时代, 其中单核苷酸的态性是人类可遗传变异中最常见的一种, 在人类基因组每 300~500 个碱基就存在 1 个 SNP, 整个人类基因组大约有 2 000 万个 SNP^[1]。

本课题前期应用荟萃分析得出母亲 *MTHFR* C677T、*MTRR* A66G 和 *MTR* A2756G 多态性的不同遗

【基金项目】 “农村学生营养改善计划”营养健康评估和膳食指导项目资助(2016)。

【作者简介】 张春红(1984-), 女, 山东德州人, 博士, 主要研究方向为营养与食品卫生学。

【通讯作者】 霍军生, E-mail: huoj@ninh.chinacdc.cn。

DOI: 10.16835/j.cnki.1000-9817.2020.01.007

传特征与叶酸、维生素 B6 和 B12 等在体内的吸收利用状况具有一定关联,母亲的不同位点分型特征影响到胎儿患神经管畸形的风险程度^[2-3]。Cummings 等^[4]研究显示,在相同的饮食条件下,CC 基因型个体相对于 TT 基因型个体的红细胞叶酸水平偏低。大量研究报道了维生素 A、维生素 D、维生素 E、维生素 K、维生素 B6、维生素 B12、叶酸、钙、铁、锌和硒等微量营养素缺乏风险相关联的 SNP 位点(MDR-SNPs)^[5-13]。人群研究显示,MDR-SNPs 可以作为生物标记物应用于多种微量营养素缺乏基因风险评估中^[14]。本研究中 143 个 MDR-SNPs 来源于已发表的全基因组关联分析(GWAS)、综述及荟萃分析等研究,从营养流行病学的角度分析了大量 SNP 位点与多种微量营养素缺乏风险之间的关联性^[5-13]。研究对象来自于 2016 年全国营养健康调查项目中九年制义务教育中一至九年级在校中小學生,涉及汉族、藏族、壮族、蒙古族和维吾尔族等多个民族,分析不同基因型以及等位基因在民族之间的分布差异,旨在探索其在我国的整体分布特征以及民族之间的差异。

1 资料来源与方法

1.1 MDR-SNPs 纳入原则及相关基因功能注释分析 采用竞争性等位基因 PCR 原理设计引物,纳入 143 个与维生素 A、维生素 D、维生素 E、维生素 B6、维生素 B12、叶酸、钙、铁、锌和硒微量营养素缺乏易感性关联较大的 SNP 位点。SNP 检索和纳入原则:在中国生物医学文献数据库、中国期刊全文数据库、万方数据库、重庆维普中文科技期刊全文数据库、PubMed 数据库和 Web of Science 检索从建库至 2017 年 6 月 25 日发表的相关文献,检索主题词分别为“single nucleotide polymorphism or SNP”和“vitamin A, D, E, B6, B12, FA, Ca, Iron, Zn, Se”和“单核苷酸多态性”和“维生素 A, D, E, B6, B12, 叶酸, 钙, 铁, 锌和硒”,同时手工检索文献,并辅以文献追踪法收集更多相关文献。

将以上 SNP 位点相关的基因基于数据库 NCBI (<http://www.ncbi.nlm.nih.gov/>)、GOC (<http://www.geneontology.org>)、Uniprot (<http://www.uniprot.org/>), 使用富集分析工具 FunRich Version 3.1.3 (Functional Enrichment analysis tool) 进行基因本体(gene ontology, GO)注释和京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)富集分析。使用富集分析工具探究表达基因富集的生物学功能物质转运的跨膜细胞组分。富集结果刻画多种营养素不同基因多态性位点之间的协同表达模式、生物学过程(biological process)、细胞组成(cellular component)、

分子功能(molecular function)和 KEGG 信号通路,得到基因参与的所有 GO,采用 Fisher 检验计算每个 GO 的统计学水平(*P* 值),从而筛选出差异基因富集的显著性 GO。其中 *P* 为 0.05 的用橘色标出,以期为探索营养缺乏风险的分子机制提供理论基础。

1.2 样本来源 来自我国 7 个省份的 1 130 名中小學生参与了该项研究(该项目是 2016 年农村义务教育学生营养改善计划——学生营养和健康状况深入监测与评价项目的一部分),采用多阶段分层抽样方法,根据城市、区、县、社区、学校开展中小學生抽样工作,每个县抽取 1~2 所农村寄宿制学校,每所学校选择一至九年级的男女學生,每个年级抽取 40 人左右,男、女基本各半。考虑到可用资金总额等一些限制因素,测量生化指标时在总样本中进行了再抽样。在监护人和参加者的充分知情同意下,采样地区涉及广西上林县、河北怀安县、内蒙古扎旗、宁夏盐池县、四川茂县、西藏巴宜区、新疆喀什地区泽普县。剔除个人基本信息或基因型或生化指标大量缺失数据后,27 份 CRP > 10 mg/L 的样本,126 份基因分型重复样本,13 份表型数据缺失样本,共 166 份样本被排除。获得 956 份样本数据纳入后续分析,其中广西上林县 199 份、河北怀安县 55 份、内蒙古扎旗 99 份、宁夏盐池县 4 份、四川茂县 198 份、西藏巴宜区 144 份、新疆喀什地区泽普县 257 份。参与者平均年龄为 (10.7±2.6) 岁;男性 493 名,女性 463 名,消除了年龄对分析结果的影响。该研究已获中国疾病预防控制中心机构评审委员会批准(2016-019)。

1.3 基因分型 采用自动化核酸提取工作站(百泰克,北京)和 DNA 提取试剂盒(百泰克,北京),提取带有凝胶的凝固血细胞中的 DNA, DNA 样本均冷冻于 -20 ℃,以便进一步使用。引物合成由生工生物工程(上海)股份有限公司完成,采用竞争等位基因特异性 PCR 法进行基因分型实验。

1.4 统计分析 基本信息采用 Excel 软件进行数据录入、核对和管理。采用 R 软件包进行 PCA、Kinship 和 SNP 位点之间连锁不平衡分析,分析候选 SNP 位点的特征^[15]。采用方差分析对主成分与种族之间的关系进行分析,检验水准 $\alpha=0.05$ 。

2 结果

2.1 基因功能注释及富集分析结果 利用生物学方法进行的基因功能分析从 BP、CC、MF 等 3 个层面进行 GO 注释,基于通路的分析进一步了解基因的生物学功能。将差异基因进行 GO 显著性富集,差异表达基因在生物过程分组中主要富集在催化活性、转

运活性、能量代谢途径、类固醇激素、辅酶、维生素 A、维生素 D 和水溶性维生素的代谢等生物学过程中,涉及转录、翻译以及能量代谢相关基因。

图 1 中每个方格中的数字为取 \log_2 (表达量) 后的结果,代表基因在样本中的表达量。结果显示,不同于其他基因,转铁蛋白基因呈现出显著高表达。



图 1 中小学生基因的生物功能分析

2.2 连锁不平衡和亲缘关系分析结果 SNP 位点颜色越深,2 个位点之间的联系就越有可能。连锁不平衡结果显示,该研究涉及到的 143 个位点之间的连锁不平衡程度是可以接受的,均可以纳入后续分析。本研究根据等位基因的频率特点分析亲缘关系,将 $kinship > 0.3$ 的样本剔除,不再纳入后续的关联分析。

2.3 主成分分析结果 主成分分析(PCA 分析)结果显示了存在群体遗传结构,见图 2。对主成分 1 和主成分 2 与种群间关系的方差分析结果显示,143 个 SNP 位点的民族差异均有统计学意义。

PCA 分析过程是从大量信息中提取关键的信息,

以便使用更少的变量(指标)实现对样本的有效区分。被提取出的信息按照效应从大到小排列,称为主成分 1(PC1)、主成分 2、主成分 3 等。采用函数 $snpGdsPCA$ 分析主成分中前 3 个成分与 SNP 位点之间的关系^[16],结果表明,与其他染色体上的多态性位点相比,3 号染色体上的基因多态性位点呈现显著的差异,主成分分析 PC1 中显示位于 *RAB6B* 基因上的 rs2280673 解释效率在 25% 以下,位于 *TF* 基因上的 rs1799852 解释效率为 25%~50%,位于 *RBP2* 基因上的 rs2118981 位点和 *SRPRB* 基因上的 rs1830084 位点解释效率在 50%~75% 之间,位于 *TF* 基因上的 rs1358024, rs1525892, rs1880669, rs3811647, rs3811658, rs6794945, rs7638018, rs8177248 等 8 个多态性位点的解释效率为 75% 以上。

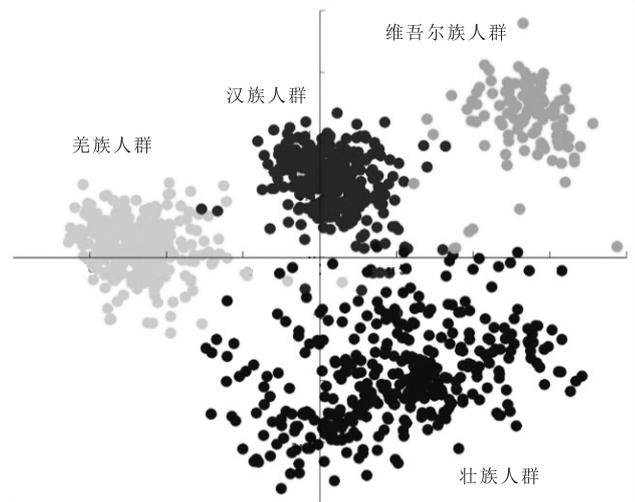


图 2 中小学生基因主成分分析

3 讨论

群体分层是全基因组关联研究(GWAS)的一个潜在问题,混淆结果并导致假阳性,因此,理解等位基因频率在不同地理区域或亚种群之间的变化很重要,也是分析基因多态性位点易感性的基础。本研究结果显示,群体遗传结构存在民族差异,壮族、羌族、汉族和维吾尔族各有独特的遗传特征。Chen 等^[17]对我国 10 个省份的 6 000 余名汉族人群进行了 GWAS 分析显示,我国南方和北方人群存在显著遗传结构差异。朱猛等^[18]研究发现,我国也存在东部和西部遗传结构差异,并通过分析第二主成分的影响因素发现哈尔滨人群的遗传特性与青岛地区极为相近,分析两者可能来源于相同亚群,并与我国曾有的“闯关东”历史相吻合。

主成分 1 与 SNP 位点之间的关系显示,3 号染色体上的基因多态性位点对于主成分 1 的解释效率具有

显著的独特性,尤其位于转铁蛋白(transferrin, TF)基因上的 8 个 SNP 位点的解释效度为 75% 以上。TF 是体内重要的铁调节蛋白,参与维持铁代谢平衡,涉及到铁的结合、转运及贮存。目前 TF 基因多态性与肿瘤、脑组织变性疾病等关联研究证实,TF 基因上多态性位点的变异可能干扰体内某些器官的铁代谢,正常状态下体内约 1/3 的 TF 被铁饱和。由于 TF G258S 位置 A 等位基因可以编码 TF 258 位置的丝氨酸,取代通常情况下的甘氨酸,氨基酸类型的改变可能使 TF 与铁结合的功能部位减少,对铁离子亲和力降低,导致结合和运送铁的能力下降,无法将足量铁离子从吸收和贮存的地方运送到红细胞以供合成血红蛋白,也无法随血液循环将足量的铁输送到骨髓等身体需铁或储铁部位,使得发育中的红细胞缺铁,血红蛋白合成不足,造成缺铁性贫血。《中国居民营养与慢性病状况报告(2015 年)》^[19]显示,我国贫血状况依然严峻,2013 年 6 岁以下儿童贫血患病率为 11.6% (城市 10.6%,农村 12.4%,其中贫困农村 16.6%),孕妇贫血率为 17.2%,城乡接近;乳母贫血率为 9.3% (城市 7.9%,农村 10.2%,其中贫困农村 14.4%)。与 2002 年相比,各组贫血率均下降,如 1 岁组下降 12.3 个百分点,2 岁下降 9.2 个百分点,孕妇下降 11.7 个百分点,乳母下降 21.4 个百分点^[19]。可见近 10 年来贫血患病率有所改善,但仍较高。近年来,人民生活水平得到了明显的提高,膳食血红素铁摄入量也有了增加,相对于农村,城市的生活条件虽然得到了更多的改善,但对贫血率的影响差异两者之间却并不显著,提示贫血状况不仅受到民族、地理环境、饮食习惯和气候因素的影响,也可能受到遗传因素的影响。

本研究显示,我国中小学生对微量营养素缺乏风险关联 SNP 位点存在民族分布差异。在接下来的研究中,将继续探索 SNP 位点与各种营养素相关的生化指标之间的关联性,以期建立合适的多基因风险评估模型,为评估微量营养素缺乏风险提供更多的科学依据。

4 参考文献

[1] 陈培战,王慧.精准医学时代下的精准营养[J].中华预防医学杂志,2016,50(12):1036-1042.

[2] 张春红,霍军生,孙静,等.中国人群母亲 5,10-亚甲基四氢叶酸还原酶基因 C677T 多态性与子代神经管畸形易感性关系的 Meta 分析[J].卫生研究,2018,47(2):312-317.

[3] 张春红,霍军生,孙静,等.母亲 MTRR 和 MTR 基因多态性与子代神经管畸形易感性 Meta 分析[J].中国预防医学杂志,2018,19(10):726-733.

[4] CUMMINGS D, DOWLING K F, SILVERSTEIN N J, et al. A cross-sectional study of dietary and genetic predictors of blood folate levels in healthy young adults[J]. *Nutrients*, 2017, 9(9):994-1011.

[5] JACOBSEN N. LNA-enhanced detection of single nucleotide polymorphisms in the apolipoprotein E [J]. *Nucleic Acids Res*, 2002, 30(19):100.

[6] MAJOR J M, YU K, CHUNG C C, et al. Genome-wide association study identifies three common variants associated with serologic response to vitamin E supplementation in men [J]. *J Nutr*, 2012, 142(5):866-871.

[7] O' SEAGHDHA C M, WU H, YANG Q, et al. Meta-analysis of genome-wide association studies identifies six new loci for serum calcium concentrations [J]. *PLoS Genet*, 2013, 9(9):e1003796.

[8] MCLAREN C E, MCLACHLAN S, GARNER C P, et al. Associations between single nucleotide polymorphisms in iron-related genes and iron status in multiethnic populations [J]. *PLoS One*, 2012, 7(6):e38339.

[9] FREDRIKSEN A, MEYER K, UELAND P M, et al. Large-scale population-based metabolic phenotyping of thirteen genetic polymorphisms related to one-carbon metabolism [J]. *Hum Mutat*, 2007, 28(9):856-865.

[10] LI W X, DAI S X, ZHENG J J, et al. Homocysteine metabolism gene polymorphisms (MTHFR C677T, MTHFR A1298C, MTR A2756G and MTRR A66G) jointly elevate the risk of folate deficiency [J]. *Nutrients*, 2015, 7(8):6670-6687.

[11] LIANG S, ZHOU Y, WANG H, et al. The effect of multiple single nucleotide polymorphisms in the folic acid pathway genes on homocysteine metabolism [J]. *Biomed Res Int*, 2014(2014):1-9.

[12] TANAKA T, SCHEET P, GIUSTI B, et al. Genome-wide association study of vitamin B6, vitamin B12, folate, and homocysteine blood concentrations [J]. *Am J Hum Genet*, 2009, 84(5):712.

[13] AHN J, YU K, STOLZENBERG-SOLOMON R, et al. Genome-wide association study of circulating vitamin D levels [J]. *Hum Mol Genet*, 2010, 19(13):2739-2745.

[14] STELUTI J, CARVALHO A M, CARIOCA A A F, et al. Genetic variants involved in one-carbon metabolism: polymorphism frequencies and differences in homocysteine concentrations in the folic acid fortification era [J]. *Nutrients*, 2017, 9(6):539-550.

[15] FUSI N, LIPPERT C, LAWRENCE N D, et al. Warped linear mixed models for the genetic analysis of transformed phenotypes [J]. *Nat Commun*, 2014, 5:4890-4897. DOI: 10.1038/ncomms5890.

[16] LIPPERT C, LISTGARTEN J, LIU Y C, et al. Fast linear mixed models for genome-wide association studies [J]. *Nat Methods*, 2011, 8(10):833-835.

[17] CHEN J, ZHENG H, BEI J X, et al. Genetic structure of the Han Chinese population revealed by genome-wide SNP variation [J]. *Am J Human Gen*, 2009, 85(6):775-785.

[18] 朱猛,吕筠,余灿清,等.中国不同地区群体遗传结构差异及调整策略研究[J].中华流行病学杂志,2019,40(1):20-25.

[19] 顾景范.《中国居民营养与慢性病状况报告(2015)》解读[J].营养学报,2016,38(6):525-529.