

论著·临床研究 doi:10.3969/j.issn.1671-8348.2023.14.009

网络首发 [https://kns.cnki.net/kcms2/detail/50.1097.R.20230705.1308.004.html\(2023-07-05\)](https://kns.cnki.net/kcms2/detail/50.1097.R.20230705.1308.004.html(2023-07-05))

基于基因芯片技术的高原红细胞增多症 基因组 DNA 甲基化差异分析*

罗勇军¹, 陈 郁¹, 蒲 懿², 张 莉³, 董红梅^{3△}

(1. 陆军军医大学陆军卫勤训练基地军事医学地理学教研室, 重庆 400038; 2. 陆军军医大学基础医学院学员
五大队十五队, 重庆 400038; 3. 中国人民解放军联勤保障部队第 941 医院心肾内科, 西宁 810001)

[摘要] **目的** 利用全基因组 DNA 甲基化芯片技术, 发现与高原红细胞增多症(HAPC)相关的甲基化基因。**方法** 纳入 4 例男性 HAPC 患者和 5 例匹配的高原健康男性, 利用外周静脉血提取基因组 DNA, 利用 Illumina 850 k 甲基化芯片检测并比较两组样品的 DNA 甲基化位点差异分布及甲基化水平差异基因, 并对相关基因进行基因本体论(GO)功能富集分析和京都基因与基因组百科全书(KEGG)信号通路分析, 探讨甲基化差异基因的潜在功能。**结果** 共发现差异甲基化位点 96 360 个, 差异甲基化基因 24 054 个。GO 功能富集分析提示, 生物过程的正向调控、细胞质和解剖结构发展是差异甲基化基因的主要功能。KEGG 信号通路分析提示, 差异甲基化基因参与了代谢通路、癌症通路和丝裂原活化蛋白激酶(MAPK)信号通路。**结论** DNA 甲基化的异常变化可能参与了 HAPC 的发生, 甲基化位点可作为 HAPC 早期诊断和预警的标志物。

[关键词] 高原红细胞增多症; DNA 甲基化; 基因芯片; 生物信息学

[中图法分类号] R811.4 **[文献标识码]** A **[文章编号]** 1671-8348(2023)14-2132-06

Differential analysis of genomic DNA methylation in high altitude polycythemia based on gene chip technology*

LUO Yongjun¹, CHEN Yu¹, PU Yi², ZHANG Li³, DONG Hongmei^{3△}

(1. Teaching and Researching Section of Military Medical Geography, Army Health Service Training Base, Army Military Medical University, Chongqing 400038, China; 2. Fifteen Team, Fifth Battalion of Cadets, Basic Medical College, Army Medical University, Chongqing 400038, China; 3. Department of Cardiology and Nephrology, 941 Hospital of PLA Joint Logistic Support Force, Xining, Qinghai 810001, China)

[Abstract] **Objective** To find the high altitude polycythemia (HAPC) related methylated genes by using the genome wide DNA methylation chip technology. **Methods** Four male patients with HAPC and five matched plateau healthy males were enrolled. Genomic DNA was extracted from peripheral venous blood. The Illumina 850 k methylation chip was used to detect and compare the differential distribution of DNA methylation sites and differential genes of methylation levels between the two groups. In addition, the gene ontology (GO) function enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathway analysis were performed to explore the potential functions of methylated differentially expressed genes. **Results** A total of 96 360 differentially methylated sites and 24 054 differentially methylated genes were found. The GO function enrichment analysis suggested that positive regulation of biological processes, cytoplasm and anatomical structure development were the main functions of differentially methylated genes. The KEGG signal pathway analysis suggested that differentially methylated genes were involved in the metabolic pathways, cancer pathways and mitogen-activated protein kinase (MAPK) signaling pathways. **Conclusion** The abnormal change of DNA methylation may be involved in the occurrence of HAPC, and the methylation sites could be used as early diagnosis and early warning markers of HAPC.

[Key words] high altitude polycythemia; DNA methylation; gene chip; bioinformatics

* 基金项目: 陆军军医大学科技创新能力提升专项项目(2019XY09); 青海省科技厅基础 Research 计划项目(2018-ZJ-705)。 作者简介: 罗勇军(1976—), 教授, 博士, 主要从事时空大数据挖掘与战略管理、卫生应急管理、智能决策等交叉学科研究。 △ 通信作者, E-mail: hmdong@126.com。

当进入海拔超过 2 500 m 的高原地区后,人体会在低氧环境作用下出现明显的生理反应,因而医学界将海拔超过 2 500 m 定义为医学高原^[1]。为代偿低氧环境、改善低氧血症,红细胞会明显增生,这种增生在一定程度上能够提高运氧能力和血氧水平、减轻缺氧损伤,是人体习服高原环境的重要机制。然而,当红细胞增生超过一定范围时,就会演变为高原红细胞增多症(high altitude polycythemia, HAPC)^[2],其带来的危害已经抵消了原有的有益效应。过度增生的红细胞降低血流速度,而动脉血氧饱和度降低进一步加重低氧血症,可造成全身多器官多系统的缺氧损伤,并伴有相应的临床症状及体征,严重影响高原居民的身体健康,尤其在高原移居人群中更为普遍,造成的危害也更为严重。HAPC 是高原环境和人体遗传因素相互作用的结果,其发病机制较为复杂,也尚未完全阐明。因此,探讨 HAPC 的发病机制是降低 HAPC 危害的重要研究方向。

DNA 甲基化是调控基因表达的重要表观遗传学机制之一,能够显著影响疾病的发生与进展^[3]。HAPC 的遗传易感因素较为复杂,多个基因参与其中,单个基因的甲基化并不能解释 HAPC 的发病机制,在全基因组层面的甲基化研究也还处于初步阶段。本研究通过全基因组 DNA 甲基化芯片技术建立高原移居汉族 HAPC 特异性甲基化基因表达谱,对比 HAPC 患者与健康者基因组水平 DNA 甲基化差异,并借助生物信息学的方法和技术筛选相关差异基因,分析其潜在功能和分子作用网络,为探索 HAPC 的发病机制,以及可能的靶向治疗和预防方案提供理论依据。

1 资料与方法

1.1 一般资料

按照 HAPC 的诊断标准^[4],在世居中国青海省西宁市的汉族男性人群中筛选了 4 例 HAPC 患者(病例组),血红蛋白(hemoglobin, Hb)水平 216~253 g/L;另选取同期体检的 5 例高原汉族健康者作为对照组, Hb 水平 170~181 g/L,年龄、性别匹配。本研究经陆军军医大学伦理委员会批准(批准编号:2020 第 001-02 号),所有受试者均签署知情同意书。

1.2 方法

1.2.1 血样采集

抽取受试者晨起空腹外周静脉血 2 mL,经乙二胺四乙酸(EDTA)抗凝后,用于提取基因组 DNA。

1.2.2 基因组 DNA 提取与质检

采用美国 OMEGA 生物技术有限公司的全血基因组提取试剂盒(货号:D3392-02)提取血样中的基因组 DNA^[5]。提取完成后先用分光光度计定量,并将样品调到标准水平 50 ng/ μ L,取 20 μ L,然后用 0.8% 的琼脂糖凝胶电泳质检。样品电泳结果提示基因组 DNA 主带清晰,通常相对分子质量不小于 10×10^3 ,

没有明显降解,总量 5 μ g 以上,能够进行下游的甲基化芯片检测。

1.2.3 基因组 DNA 甲基化水平检测

采用美国 Illumina 公司的甲基化芯片 Infinium MethylationEPIC BeadChip(850 k 芯片)检测受试者基因组 DNA 甲基化水平^[6]。该芯片覆盖了 853 307 个 CpG 位点,并全面覆盖基因启动子区、基因编码区、CpG 岛、增强子区及 99% 的 RefSeq 基因。后续的亚硫酸盐转化,DNA 扩增,DNA 的片段化、沉淀和重悬,DNA 与芯片的杂交,芯片清洗、单碱基延伸、染色及芯片扫描和数据提取在北京博奥晶典生物技术有限公司完成。

1.3 统计学处理

将处理好的芯片放入扫描仪,利用激光激发芯片上的单碱基延伸产物的荧光基团,扫描仪获取由荧光基团发出的荧光,并生成原始数据,记录扫描结果存放的位置。由此所得的数据直接导入 GenomeStudio 软件进行分析,根据 Illumina 官方甲基化分析算法获得每个位点的原始信号强度值。然后经过对不同荧光、探针类型引起的偏差校正及归一化、位点过滤,得到高质量 CpG 位点的归一化后甲基化水平,即 β 值,用于质控。根据 β 值计算样品间 Pearson 相关系数(Pearson correlation coefficient, PCC),以评估受试样品芯片信号质量的重复率(β 值为 0 表示该位点无甲基化, β 值为 1 表示该位点完全甲基化)。质控后的数据,采用 R 软件包 IMA3. 1. 2 进行差异甲基化分析,其采用的方法为 Limma 中的经验贝叶斯统计^[7-8]。同时针对多重假设检验问题,计算错误发现率(false discovery rate, FDR)校正 P 值,以降低假阳性率。差异甲基化位点的选取标准为:校正 $P \leq 0.05$ (如果所有位点均校正 $P > 0.05$,或 ≤ 0.05 的位点很少,则以校正之前的 $P \leq 0.05$ 为标准)。针对差异甲基化位点,使用 R 脚本对样本进行聚类分析,并使用通路富集工具 KOBAS 注释这些差异位点映射到的基因,通过与京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG) GENES 数据库关联链接,挖掘出统计上明显的基因本体论(gene ontology, GO)功能富集分析和 KEGG 信号通路分析^[9],探讨差异甲基化基因的潜在功能及在 HAPC 发病机制中可能的作用。通过上述生物信息学分析,探讨差异甲基化基因及其作用网络在高原移居汉族 HAPC 发病机制中可能的作用。所有检验为双侧,以 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 甲基化芯片评价

将扫描得到的原始数据通过 GenomeStudio 软件获得每个位点的原始信号值和测量 P 值(Detection P 值),二者用于质控以评估数据质量和后续分析。测量 $P > 0.05$ 表明样品的质量较低,需要去除。原始

信号值提示所有样品的位点检测率均在 99.9% 以上,保证了检测范围。质控包括样本独立对照(sample-independent controls, SICs)、样本非独立对照(Sample-dependent controls, SDCs)、位点质控及个体质控。SICs 用于评估操作步骤,SDSs 用于评估样品质量。位点质控要求位点在 95% 以上的个体中测量 $P < 0.05$,同时去除位于 X、Y 染色体上的位点。个体质控要求个体在 95% 以上位点中的测量 $P < 0.05$ 。在经历严格的质控后,获取了病例组和对照组的 β 值。PCC 分析提示病例-对照 β 值的 PCC 为 0.997 9,表明芯片信号质量很好,所得数据可以用于后续分析,见图 1。

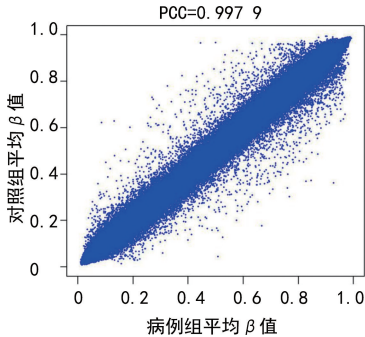


图 1 病例组与对照组甲基化水平 Pearson 相关性分析

2.2 差异甲基化分析

按照差异甲基化位点的筛选标准,共筛选得到 96 360 个差异甲基化位点。其中,根据位点所属区域差异,分为 CpG 岛(10 699,11.1%)、岛滩区(23 799,24.7%)、岛架区(7 370,7.6%)和其他区域(54 492,56.6%)。根据甲基化水平变化,病例组的高甲基化位点共 10 564 个(11.0%),低甲基化位点共 85 796 个(89.0%)。根据对照组与病例组对比结果显示:对照组中大部分 CpG 位点呈现高甲基化状态,而病例组大部分甲基化水平降低,与对照组相比差异甲基化基因 24 054 个。根据差异甲基化位点所映射的基因变化,与对照组相比,病例组的高甲基化基因共 5 981 个,低甲基化基因共 18 073 个。

2.3 差异甲基化基因谱

在前述筛选出的差异甲基化位点所处的基因中,按照筛选标准获取了 10 个最明显的高甲基化基因,分别是含黄素单氧化酶 3(flavin-containing monooxygenase 3,FMO3)、G 蛋白核仁 2(G protein nucleolar 2,GNL2)、钙调磷酸酶类似 EF-Hand 蛋白 2(cal-cineurin like EF-hand protein 2,CHP2)、酰基辅酶 A 硫酯酶 2(acyl-CoA thioesterase 2,ACOT2)、染色体 1 开放阅读框 25(chromosome 1 open reading frame 25,C1orf25)、OCA2 黑素体跨膜蛋白(OCA2 melano-somal transmembrane protein,OCA2)、肌动蛋白相关蛋白 2/3 复合亚基 4(actin related protein 2/3 complex subunit 4, ARPC4)、锚蛋白重复域 55 (ankyrin repeat domain 55,ANKRD55)、GRB10 交互作

用 GYF 蛋白 2(GRB10 interacting GYF protein 2,GI-GYF2)、溶酶体蛋白跨膜 4 β (lysosomal protein trans-membrane 4 beta,LAPTM4B),见表 1。筛选的 10 个最明显低甲基化基因分别是黏附蛋白偶联受体 B1(adhesion G protein-coupled receptor B1,ADGRB1)、Rab 亲和蛋白 3A 类似物(不含 C2 结构域)[rabphilin 3A Like (without C2 domains),RPH3AL]、酰辅酶 A 硫酯酶 1(acyl-CoA thioesterase 2,ACOT1)、ATP 结合盒亚家族 C 成员 13(ATP binding cassette subfamily C member 13,ABCC13)、钙粘蛋白 22(cadherin 22,CDH22)、染色体 1 开放阅读框 109(chromosome 1 open reading frame 109,C1orf109)、防御素 β 128(defensin beta 128,DEFB128)、锚蛋白重复域 23(ankyrin repeat do-main 23,ANKRD23)、酰基辅酶 A 合成酶家族成员 3(Acyl-CoA synthetase family member 3,ACSF3)、钙/钙调蛋白依赖性蛋白激酶 1D(calcium/calmodulin dependent protein kinase 1D,CAMK1D),见表 2。

表 1 筛选出的 10 个最明显高甲基化基因
(病例组 vs. 对照组)

基因名称	<i>P</i>	β 值差值	染色体	MAPINFO ^a
FMO3	4.84×10^{-9}	0.27	1	171 083 174
GNL2	8.30×10^{-7}	0.23	1	38 061 675
CHP2	6.35×10^{-5}	0.11	16	23 767 194
ACOT2	6.68×10^{-5}	0.17	14	74 034 893
C1orf25	7.39×10^{-5}	0.08	1	185 126 416
OCA2	8.92×10^{-5}	0.66	15	28 054 345
ARPC4	8.97×10^{-5}	0.12	3	9 834 004
ANKRD55	0.01×10^{-2}	0.22	5	55 444 662
GIGYF2	0.02×10^{-2}	0.01	2	233 671 240
LAPTM4B	0.02×10^{-2}	0.04	8	98 861 074

^a:甲基化位点在染色体上的位置(GRCh36)。

表 2 筛选出的 10 个最明显低甲基化基因
(病例组 vs. 对照组)

基因名称	<i>P</i>	β 值差值	染色体	MAPINFO ^a
ADGRB1	1.87×10^{-12}	-0.67	8	143 561 205
RPH3AL	4.63×10^{-10}	-0.38	17	133 110
ACOT1	9.42×10^{-9}	-0.54	14	74 003 199
ABCC13	2.12×10^{-7}	-0.18	21	15 646 114
CDH22	8.18×10^{-7}	-0.17	20	44 824 026
C1orf109	1.01×10^{-6}	-0.15	1	38 156 575
DEFB128	1.06×10^{-6}	-0.67	20	170 641
ANKRD23	1.07×10^{-6}	-0.24	2	97 505 464
ACSF3	1.66×10^{-6}	-0.09	16	89 187 848
CAMK1D	2.22×10^{-6}	-0.66	10	12 438 782

^a:甲基化位点在染色体上的位置(GRCh36)。

2.4 差异甲基化位点所在基因生物功能富集分析

2.4.1 GO 功能富集分析

按照筛选标准,病例组与对照组的差异甲基化位点所在基因相关的 GO 条目共 19 585 个,根据 *P* 值筛选了前 10 个最明显的 GO 条目。其中,主要富集于生物过程的物质定位、生物过程和细胞过程的正向调控及细胞发育等,以及细胞组分的细胞质及分子功能的结合,见表 3。

表 3 差异甲基化位点所在基因的 GO 分析

GO 编码	GO 结果	GO 分类	<i>P</i>	注释到该 Term 的 输入基因数目
0051179	定位(localization)	生物学过程	1.17×10^{-30}	4 677
0048518	生物过程的正向调控(positive regulation of biological process)	生物学过程	4.24×10^{-30}	4 625
0005737	细胞质(cytoplasm)	细胞组分	8.04×10^{-29}	8 499
0005488	结合(binding)	分子功能	1.12×10^{-27}	11 141
0048856	解剖结构发展(anatomical structure development)	生物学过程	6.86×10^{-26}	4 452
0009653	解剖结构形态发生(anatomical structure morphogenesis)	生物学过程	1.63×10^{-25}	6 510
0044444	细胞质部分(cytoplasmic part)	细胞组分	2.08×10^{-25}	3 954
0007275	多细胞有机体发展(multicellular organismal development)	生物学过程	4.15×10^{-25}	1 702
0048522	细胞过程的正向调控(positive regulation of cellular process)	生物学过程	1.04×10^{-24}	3 954
0048468	细胞发育(cell development)	生物学过程	2.50×10^{-24}	1 702

表 4 差异甲基化位点所在基因的 KEGG 通路分析

编号	通路名称	<i>P</i>
hsa01100	代谢通路(metabolic pathways)	0.04×10^{-3}
hsa05200	癌症通路(pathways in cancer)	0.08×10^{-2}
hsa04010	MAPK 信号通路(MAPK signaling pathway)	0.04×10^{-1}
hsa05166	人类嗜 T 淋巴细胞病毒 I 型感染(HTLV-I infection)	0.05×10^{-1}
hsa04015	Rap1 信号通路(Rap1 signaling pathway)	0.05×10^{-1}
hsa04510	黏着斑(focal adhesion)	0.06×10^{-1}
hsa04151	磷脂酰肌醇 3 激酶-蛋白激酶 B 信号通路(PI3K-akt signaling pathway)	0.07×10^{-1}
hsa04921	催产素信号通路(oxytocin signaling pathway)	0.09×10^{-1}
hsa04014	Ras 信号通路(Ras signaling pathway)	0.09×10^{-1}
hsa05205	癌症中的蛋白聚糖(proteoglycans in cancer)	0.01×10^{-1}

3 讨 论

HAPC 是高原地区的常见疾病,海拔 4 000 m 以上地区患病率高达 24.0%^[10],而在海拔超过 5 000 m 的喀喇昆仑山脉地区则更高,移居人群如汉族人群可达 80% 以上^[11],严重威胁着高原地区人群的健康。因此,探索 HAPC 的发病机制是维护高原人群生命安全的重要研究方向。HAPC 的发生是高原环境和人体相互作用的结果,涉及多个系统和多个环节,其分子机制不仅与核基因序列、线粒体 DNA(mtDNA)序列改变相关,还涵盖表观遗传学的变化,包括 DNA 甲基化、组蛋白修饰(如乙酰化、苏木化和磷酸化)及非编码 RNA 调控等多种机制。PENG 等^[12]发现了缺氧诱导因子(hypoxia inducible factor,HIF)途径的

2.4.2 KEGG 信号通路分析

病例组与对照组的差异甲基化位点所在基因相关的 KEGG 信号通路共 305 个,符合筛选条件的信号通路共 42 条(*P*<0.05),并根据 *P* 值筛选了前 10 个最明显的信号通路,见表 4。结果提示,差异甲基化位点所在基因涉及的信号通路主要包括代谢通路、癌症通路及丝裂原活化蛋白激酶(MAPK)信号通路。

内皮含 PAS 结构域蛋白 1(EPAS1)和脯氨酰羟化酶蛋白(EGLN1)基因可能是高海拔遗传适应的候选基因。FAN 等^[13]通过全外显子测序发现,在西藏世居人群中磷脂酰肌醇-3 激酶催化亚基 δ(PIK3CD)和Ⅳ型胶原蛋白 α3 链(COL4A3)基因上存在差异单核苷酸多态性(SNP)位点。ZHOU 等^[14]通过全基因组关联研究,发现红细胞生成调节基因小泛素相关修饰蛋白特异性蛋白酶 1(SENP1)和癌基因酸性富含亮氨酸的核磷蛋白 32 家族成员 D(ANP32D)在患有慢性高原疾病的患者中表达上调。陈辉等^[15]发现,白细胞介素-12 受体 B1(IL12RB1)基因的 SNPsrs393548、rs436857 和 rs845380 与西藏世居人群 HAPC 的发生有关。

目前,关于 HAPC 遗传易感机制的研究多为 DNA 水平,表观遗传机制研究较少。本研究借助高通量基因芯片技术,首次从全基因组层面分析了中国汉族 HAPC 患者的基因组 DNA 甲基化水平,发现 HAPC 患者基因组 DNA 大部分呈低甲基化状态,并筛选了以 FMO3 为代表的高甲基化基因和以 ADGRB1 为代表的低甲基化基因,借助生物信息学分析富集了相关的信号通路,提示基因组 DNA 的异常甲基化也是 HAPC 发生的重要机制。

除了基因本身的变化外,表观遗传学的改变也是人体适应高原的重要机制^[16]。DNA 甲基化作为表观遗传的主要机制之一,其核心是由 S-腺苷甲硫氨酸作为甲基供体,在 DNA 甲基化转移酶的催化下,提供的甲基被转移到相应碱基的一种 DNA 共价修饰方式,并不涉及 DNA 序列的改变^[17]。已有研究提示,作为高原适应的主要候选基因 EGLN1,其启动子区域的高甲基化与安第斯山印第安人 HAPC 有关^[16];而另一个高原适应候选基因 EPAS1,其甲基化水平与安第斯山印第安人的早期发育和终生高海拔暴露及高海拔适应性表型有关^[18-19],而血管紧张素 II 1 型受体相关蛋白的配体(Apelin)基因的高甲基化也在高原肺水肿的发生中发挥着重要作用^[20]。在本研究中,发现 HAPC 患者的基因组 DNA 多呈低甲基化状态,提示其基因表达较为活跃。在长期的低氧环境刺激下,作为低氧反应重要的调控因子 HIFs,在与特异的顺式作用元件相结合后,启动了一系列基因的转录,并以此形成信号转导的级联放大效应,建构复杂的分子作用网络,涵盖人体对缺氧的感知和反应,从造血干细胞分化成红系祖细胞再到成熟红细胞的形成,骨髓造血微环境的调节,促红细胞生成素的分泌与反馈,Hb 合成所需的铁、维生素 B₁₂ 等原料的吸收合成与利用。上述过程中不可避免的涉及多种表观遗传学变化,而本研究发现了 HAPC 患者和高原健康对照在多个基因的甲基化有着不同程度的变化,提示异常的甲基化改变在上述过程中发挥着重要的调节作用。

在大鼠的 HAPC 模型中,EGLN1 基因的甲基化水平与正常对照并无明显差异^[21],而 von Hippel-Lindau(VHL)基因启动子的高甲基化会降低 VHL 基因表达,进而促进 HIF-2 α 和促红细胞生成素(EPO)的表达,诱导 HAPC 的发生^[22]。本研究提示,HAPC 患者的 FMO3 呈现高甲基化状态,而抑制 FMO3 的表达能够降低全血中胶原依赖的血小板与基质的黏附性,对血栓形成有抑制作用^[23]。在 HAPC 患者中,由于红细胞过度增生易并发血栓,尤其是诱导微循环障碍,加重全身缺氧损伤。因此,推测 HAPC 患者的高甲基化 FMO3 状态可能是功能代偿,从血小板角度抑制血栓形成。在低甲基化基因方面,本研究提示 HAPC 患者的 ADGRB1 基因差异最为明显。ADGRB1 基因主要在脑部表达,在吞噬、炎

症、突触形成、抑制血管生成和成肌细胞融合等方面发挥着重要作用^[24]。研究发现,ADGRB1 基因的高表达能够抑制脑血管生成,与脑部肿瘤发生和肿瘤周围脑水肿呈负相关^[25]。此外,研究提示 HAPC 患者出现了不同程度的脑水肿^[26],而本研究中发现的低甲基化 ADGRB1 基因可能也是通过提高其表达水平以拮抗脑水肿的发生。

在本研究中,借助多种生物信息学分析方式探讨了差异甲基化基因的功能及其潜在的作用网络,探寻其在 HAPC 发生中的作用。综合分析两组差异甲基化位点,提示病例组和对照组基因组 DNA 甲基化水平差异明显,而且病例组绝大部分 CpG 差异甲基化位点数目低于对照组。基因组 DNA 的高甲基化位点意味着基因表达水平下降,而 HAPC 患者的高甲基化位点数目较健康者明显下降,表明在长期的低氧环境刺激下,人体内部分基因被激活,其形成的信号放大机制和分子相互作用网络在 HAPC 的发生、发展中发挥了重要作用。GO 分析提示,差异甲基化基因参与了生物过程的正向调控、细胞质、解剖结构发展和解剖结构形态发生等过程。上述富集的功能均可能涉及 HAPC 的发病机制,如红细胞生成与细胞发育和正向调控有关,细胞质的变化是红细胞在成熟过程中的必然反应,而结合既包含了转录因子和 DNA 的结合,也包括了蛋白和蛋白之间的结合与相互作用,在红细胞生成和成熟过程中均发挥了重要作用。KEGG 通路分析提示,包括代谢通路、癌症通路和 MAPK 信号通路等在 HAPC 对照组中呈现高富集。红细胞增生过程中涵盖了代谢通路的调整,也必然涉及能量代谢的改变,而 MAPK 信号通路在能量代谢过程中发挥着重要作用。此外,红细胞的过度增生,在生物学过程方面类似于癌细胞的过度增殖,故在信号通路富集过程中也涉及癌症通路的变化。上述结果进一步提示,HAPC 的发生是机体在应对低氧环境刺激时调控网络失衡的结果,低氧感知、红细胞发育分化与成熟、造血微环境如炎症状态、红细胞生成原料摄取等相关信号通路依然是探索 HAPC 发病机制的重要方向,如 FMO3 可影响血栓形成,ADGRB1 在炎症免疫、血管生成和肿瘤发生中均发挥着不同程度的效应,这也进一步提示甲基化在 HAPC 的发生中发挥了重要的调控作用。

综上所述,本研究借助高通量基因芯片技术,从基因组甲基化角度探索了 HAPC 可能的发生机制,并发现了甲基化异常与 HAPC 发生有关。虽然纳入的样本量较小,生物学重复也未完成,在结果解释方面有一定的局限性,但是目前 Infinium MethylationEPIC BeadChip(850 k 芯片)的应用越来越广,研究方法和技术稳定,研究结果仍有一定的借鉴意义,提示甲基化改变也在 HAPC 发生中发挥着重要作用,异常的甲基化位点可能是 HAPC 发生的诊断标

志物。本研究在基因组层面检测了甲基化水平,并未对某个具体部位或某个基因的具体位置再检测,针对差异甲基化基因,后期还需要借助甲基化特异性 PCR 进一步扩大样本验证结果。此外,虽然本研究只是甲基化改变在 HAPC 发生机制研究的开端,但是其中也提供了大量证据,提示 HAPC 的发病机制研究也要注意炎症免疫、血管生成和血栓形成等信号通路,其相关的表观遗传学调控机制是 HAPC 早期治疗和预警的客观依据。

参考文献

- [1] RONEN R,ZHOU D,BAFNA V,et al. The genetic basis of chronic mountain sickness[J]. *Physiology (Bethesda)*,2014,29(6):403-412.
- [2] CHEN Y,JIANG C,LUO Y,et al. An EPAS1 haplotype is associated with high altitude polycythemia in male Han Chinese at the Qinghai-Tibetan plateau[J]. *Wilderness Environ Med*,2014,25(4):392-400.
- [3] FUJII R,SATO S,TSUBOI Y,et al. DNA methylation as a mediator of associations between the environment and chronic diseases;a scoping review on application of mediation analysis[J]. *Epigenetics*,2022,17(7):759-785.
- [4] LEÓN-VELARDE F,MAGGIORINI M,REEVES J T,et al. Consensus statement on chronic and subacute high altitude diseases[J]. *High Alt Med Biol*,2005,6(2):147-157.
- [5] CHEN Y,JIANG C,LUO Y,et al. Interaction of CARD14, SENP1 and VEGFA polymorphisms on susceptibility to high altitude polycythemia in the Han Chinese population at the Qinghai-Tibetan Plateau[J]. *Blood Cells Mol Dis*,2016,57:13-22.
- [6] OH J Y,KO Y K,GIM J A. DNA methylation profiling for the diagnosis and prognosis of patients with nontuberculous mycobacterium lung disease[J]. *Curr Issues Mol Biol*,2021,43(2):501-512.
- [7] WANG D,YAN L,HU Q,et al. IMA:an R package for high-throughput analysis of Illumina's 450K Infinium methylation data[J]. *Bioinformatics*,2012,28(5):729-730.
- [8] TESCHENDORFF A E,MARABITA F,LECHNER M,et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data[J]. *Bioinformatics*,2013,29(2):189-196.
- [9] YANG W,HAN J,MA J,et al. Prediction of key gene function in spinal muscular atrophy using guilt by association method based on network and gene ontology[J]. *Exp Ther Med*,2019,17(4):2561-2566.
- [10] JIANG C,CHEN J,LIU F,et al. Chronic mountain sickness in Chinese Han males who migrated to the Qinghai-Tibetan plateau:application and evaluation of diagnostic criteria for chronic mountain sickness[J]. *BMC Public Health*,2014,14:701.
- [11] 张西洲,崔建华,王宏运,等. 驻喀喇昆仑山和西藏阿里某部队高原红细胞增多症患病率调查[J]. *高原医学杂志*,2007,17(1):58-60.
- [12] PENG Y,CUI C Y,HE Y X,et al. Down-regulation of EPAS1 transcription and genetic adaptation of Tibetans to high-altitude hypoxia[J]. *Mol Biol Evol*,2017,34(4):818-830.
- [13] FAN X W,MA L F,ZHANG Z Y,et al. Associations of high-altitude polycythemia with polymorphisms in PIK3CD and COL4A3 in Tibetan populations[J]. *Hum Genomics*,2018,12(1):37.
- [14] ZHOU D,UDPA N,RONEN R,et al. Whole-genome sequencing uncovers the genetic basis of chronic mountain sickness in Andean highlanders[J]. *Am J Hum Genet*,2013,93(3):452-462.
- [15] 陈辉,杨雪林,张致英,等. PDK1、IL12RB1 和 TCL1A 基因多态性与西藏世居人群高原红细胞增多症易感性分析[J]. *医学研究生学报*,2021,34(8):791-797.
- [16] JULIAN C G. Epigenomics and human adaptation to high altitude[J]. *J Appl Physiol (1985)*,2017,123(5):1362-1370.
- [17] CHILDEBAYEVA A,HARMAN T,WEINSTEIN J,et al. Genome-wide DNA methylation changes associated with high-altitude acclimatization during an everest base camp trek[J]. *Front Physiol*,2021,12:660906.
- [18] CHILDEBAYEVA A,JONES T R,GOODRICH J M,et al. LINE-1 and EPAS1 DNA methylation associations with high-altitude exposure[J]. *Epigenetics*,2019,14(1):1-15.
- [19] CHILDEBAYEVA A,GOODRICH J M,LEÓN-VELARDE F,et al. Genome-wide epigenetic signatures of adaptive developmental plasticity in the Andes[J]. *Genome Biol Evol*,2021,13(2):evaa239. (下转第 2142 页)