# 基于生物信息学分析原发性醛固酮增多症潜在 的核心基因

代黎明,曹彩霞\*,颜廷屹,孔 越,徐音飞

青岛大学附属医院老年医学科,山东 青岛

收稿日期: 2023年5月21日; 录用日期: 2023年6月14日; 发布日期: 2023年6月25日

# 摘要

目的: 运用生物信息学方法分析原发性醛固酮增多症的关键基因及涉及的信号通路。方法: 从公共基因 表达数据库(GEO)中下载原发性醛固酮增多症(primary aldosteronism, PA)的基因表达谱芯片 GSE60042,和DNA甲基化芯片GSE60043。通过R语言对PA的转录组进行差异表达分析,对PA DNA的 甲基化芯片进行差异甲基化基因筛选。通过对差异表达基因GO和KEGG富集分析筛选PA的核心基因。结 论:原发性醛固酮增多症的核心基因可能是ABCB4、AQP2、ARG2和CLRN1。

## 关键词

原发性醛固酮增多症,生物信息学分析,差异表达基因,甲基化

# **Analysis of Potential Core Genes in Primary Aldosteronism Based on Bioinformatics Analysis**

#### Liming Dai, Caixia Cao\*, Tingyi Yan, Yue Kong, Yinfei Xu

Department of Geriatrics Medicine, Affiliated Hospital of Qingdao University, Qingdao Shandong

Received: May 21<sup>st</sup>, 2023; accepted: Jun. 14<sup>th</sup>, 2023; published: Jun. 25<sup>th</sup>, 2023

#### Abstract

Objective: To analyze the key genes and signaling pathways of primary aldosteronism by bioinformatics analysis. Methods: GSE60042 gene expression microarray and GSE60043 DNA methyla-\*通讯作者。

文章引用: 代黎明, 曹彩霞, 颜廷屹, 孔越, 徐音飞. 基于生物信息学分析原发性醛固酮增多症潜在的核心基因[J]. 临床医学进展, 2023, 13(6): 9958-9968. DOI: 10.12677/acm.2023.1361392

tion microarray were used to profile primary aldosteronism from the public gene expression database (GEO). Differential expression analysis of the transcriptome of PA was performed by R language, and methylation microarrays of PA DNA were screened for differentially methylated genes. Core genes were screened by GO and KEGG enrichment analysis of differentially expressed genes. Conclusion: ABCB4, AQP2, ARG2 and CLRN1 may be the core genes of primary aldosteronism.

## **Keywords**

Primary Aldosteronism, Bioinformatics Analysis, Differentially Expressed Genes, Methylation

Copyright © 2023 by author(s) and Hans Publishers Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/

# 1. 引言

原发性醛固酮增多症(primary aldosteronism, PA)是一种常见的继发性高血压,占高血压患者 10%以上, 在难治性高血压中占 17%~23%。原醛症主要是肾上腺皮质球状带自主分泌过量醛固酮,过量的醛固酮作 用于肾脏远曲小管和集合管上的醛固酮受体,导致水钠潴留,钾排出量增多,血容量增多,肾素 - 血管 紧张素 - 醛固酮(RAAS)系统活性受制,临床表现主要为高血压、低血钾[1]。与原发性高血压(essential hypertension, EH)相比,更容易合并靶器官损害,有研究显示 PA 患者脑卒中、心肌梗死、心房颤动的发 生率分别为 EH 患者的 4.2,6.5 及 12.1 倍。还有学者发现, PA 患者与 EH 患者相比,靶器官损害分别为: 心肌梗死(4.4%和 1.7%)、心房颤动(3.9%和 1.1%)、冠心病(3.9%和 1.1%)、心力衰竭(4.1%和 1.2%) [2]。 其造成的心脏、肾脏等靶器官的损害严重影响患者的生活质量[3]。研究发现,DNA 的甲基化可以通过表 观遗传学修饰,调控基因转录过程,从而在基因未发生变化的情况下,基因功能发生了变化,最终导致 表观遗传变化[4]。有学者收集 15 例智力障碍或发育迟缓患儿,进行 SNP 微阵列检测后发现 2 例存在微 缺失,且均为新生变异;13 例 SNP 微阵列检测结果为阴性的患儿,经目标基因捕获高通量测序、基因变 异 -临床表型关联分析及遗传学分析,确诊 5 例患儿为单基因病, 2 例为疑似, 6 例阴性[5]。

近年来,随着微阵列分析和测序技术的发展,生物信息学技术已经成为筛选和识别致病基因,诊断和治疗疾病的有效技术手段[6]。因此,本研究旨在通过对 GEO 数据库中 PA 的基因表达谱及 DNA 甲基 化谱进行生物信息学分析,揭示与 PA 疾病发生相关的核心基因及信号通路,为进一步阐明 PA 的发病机 制提供有价值的信息,并为 PA 的诊断、治疗提供新的思路。

# 2. 研究对象

以"primary aldosteronism"作为关键词,在 GEO 数据库中筛选到两个 PA 的表达数据集: GSE60042 和 GSE60043。下载相应注释平台文件分别获得各数据集的基因表达矩阵。GSE60042 作为 PA 基因表达 数据集,GSE60043 作为 DNA 甲基化数据集[7]。

## 3. 研究方法

# 3.1. 微阵列芯片初始数据的获取

从 NCBI GEO 公共数据库下载 GSE60042 的 Series Matrix File 数据文件,注释平台为 GPL14550,共

14 组转录组数据, AAG 组 7 例(n = 7), APA 组 7 例(n = 7); 从 NCBI GEO 公共数据库下载 GSE60043 的 甲基化 450 k 矩阵, 其中 AAG 组 7 例(n = 7), APA 组 7 例(n = 7)。

#### 3.2. 差异基因的筛选

使用 R 包 "Limma" 对转录组数据集中的两组患者进行差异分析,差异基因筛选条件为|logFC| > 1 & adj.P.Val < 0.05。DNA 启动子甲基化通过转录沉默来调控其表达。本研究利用 R 包 "CHAMP"分析 GSE60043 数据集中的甲基化数据,分析 AAG 组与 APA 组中探针的甲基化水平,以及筛选差异甲基化 基因,差异甲基化的筛选条件为|logFC| > 0.2 & adj.P.Val < 0.05。通过韦恩图筛选高甲基化低表达和低甲 基化高表达的基因。

#### 3.3. GO 和 KEGG 功能注释

使用 R 包 "clusterProfiler"对基因进行功能注释,以全面探讨这些差异基因的功能相关性。基因本体论(GO)和京都基因与基因组百科全书(KEGG)被用来评估相关的功能类别。p 值和 q 值均小于 0.05 的 GO 和 KEGG 富集通路被认为是显著性类别。

#### 3.4. Lasso 回归和 SVM 算法的特征选择过程

我们使用 Lasso logistic 回归和 SVM 算法对醛固酮增多症的诊断标志物进行特征选择。Lasso 算法使用"glmnet"软件包。此外,SVM-RFE 是一种基于支持向量机的机器学习方法,通过删除 SVM 生成的特征向量来寻找最佳变量,并通过"e1071"软件包建立支持向量机模型,进一步识别这些生物标志物对疾病的诊断价值。

# 3.5. GSEA

基因集富集分析(GSEA)用于筛选数据集中可能与 PA 相关的 KEGG 通路。GSEA 采用 GSEA-4.0.1 (Java)版本进行,背景基因集为 c2.cp.kegg.v7.5.symbols.gmt,其中置换次数设置为 1000,置换类型设置为 phenotype。

#### 3.6. 统计分析

统计分析采用 R 语言(version 3.6)进行。所有统计检验均为双侧, p < 0.05 具有统计学意义。

#### 4. 结果

#### 4.1. 原发性醛固酮增多症与正常肾上腺组织的差异基因表达及分布特征

我们从 NCBI GEO 公共数据库下载 GSE60042 共 14 组转录组数据, 根据 GSE60042 的表达谱进行差 异分析, 差异基因筛选条件为: |Log2FC| > 1 & adj.P.Val < 0.05, 共筛选 500 个差异基因, 其中上调基因 192 个, 下调基因 308 个(图 1、图 2)。另一方面, 我们从 NCBI GEO 公共数据库下载 GSE60043 的数据 文件, 共 14 组甲基化 450 K 的数据, 用 ChAMP 包进行差异甲基化位点分析, 共筛选出 3015 个差异甲 基化探针, 其中下调探针 2975 个(|Log2FC| < 0.2 & adj.P.Val < 0.05), 上调探针 40 个(|Log2FC| > 0.2 & adj.P.Val < 0.05), 上调探针 40 个(|Log2FC| > 0.2 & adj.P.Val < 0.05), [图 3、图 4)。

### 4.2. 差异表达基因 GO 和 KEGG 富集分析结果和核心基因筛选结果

对高甲基化低表达以及低甲基化高表达的基因取交集,韦恩图结果显示共有 29 个交集的 mRNA(2 + 27),我们将这 29 个 mRNA 作为后续分析的候选基因集(图 5、图 6)。



**Figure 1.** Volcano map of differentially expressed genes. Note: Blue represents significantly down regulated genes, pink represents significantly up regulated genes, and gray represents genes with no statistically significant differences in expression

**图 1.** 差异表达基因的火山图。注: 蓝色为显著下调基因, 粉色为显著上调基因, 灰色为表达差异无统计学意义的基因



**Figure 2.** Heat map of differential gene expression profile. Note: Blue represents low expression, pink represents high expression, and the darker the color, the higher or lower the expression level of differential genes

**图** 2. 差异基因表达谱热图。注: 蓝色为低表达, 粉色为高表达, 颜色越深为差异基因的表达量越高或越低



Figure 3. Volcano map of differentially expressed genes. Note: Blue represents significantly down regulated genes, pink represents significantly up regulated genes, and gray represents genes with no statistically significant differences in expression

**图 3.** 差异表达基因的火山图。注: 蓝色为显著下调基因, 粉色为 显著上调基因, 灰色为表达差异无统计学意义的基因



#### **图 4.** 高表达前 10 位差异基因和低表达前 10 位差异表达基因热图。 注: 蓝色为低表达,红色为高表达,颜色越深为差异基因的表达 量越高或越低

**Figure 4.** Heat map of high expression top 10 differentially expressed genes and low expression top 10 differentially expressed genes. Note: Blue represents low expression, red represents high expression, and the darker the color, the higher or lower the expression level of differential genes



图 6.29 个交集的 mRNA (2 + 27)韦恩图

随后我们对这 29 个 mRNA 进行富集分析。GO 富集结果表明, 基因主要富集的通路是 ATPase activity、 polyol transport、regulation of interleukin-13 production、renal system process。KEGG 富集结果表明, 基因 主要富集的通路是 Pancreatic secretion、cAMP signaling pathway、Endocrine and other factor-regulated calcium reabsorption、Thyroid hormone signaling pathway、Aldosterone-regulated sodium reabsorption、 Aldosterone synthesis and secretion (图 7~9)。











**Figure 9.** KEGG analysis results. Note: Bubble size represents the number of enriched genes, and color represents the size of the P-value 图 9. KEGG 分析结果。注: 气泡大小表示富集的基因个数,颜色表示 P 值的大小

#### 4.3. 核心基因表达水平验证

为了进一步找出影响醛固酮增多症的特征基因,我们联合使用 lasso 回归和 SVM 特征选择算法对 29 个候选基因进行筛选。结果显示,Lasso 回归共识别 7 个基因作为醛固酮增多症的特征基因,图 10,另一方面,我们通过 SVM-RFE 算法评估醛固酮增多症中的特征基因。结果显示,通过筛选准确率最高以及错误率最低的前 10 个的特征基因,与 Lasso 回归算法筛选出来的特征基因取交集,共筛选出 4 个交集 基因。图 11 这 4 个基因将作为醛固酮增多症的核心基因,它们分别是:ABCB4、AQP2、ARG2 和 CLRN1



Figure 10. Lasso regression and SVM feature selection algorithm for screening 29 candidate genes 图 10. Lasso 回归和 SVM 特征选择算法对 29 个候选基因进行筛选



**Figure 11.** Screen the top 10 feature genes with the highest accuracy and lowest error rate, intersecting with the feature genes screened by the Lasso regression algorithm 图 11. 筛选准确率最高以及错误率最低的前 10 个的特征基因, 与 Lasso 回归算法筛选出来的特征基因取交集

# 5. 讨论

原发性醛固酮增多症是由于肾上腺皮质过度分泌醛固酮,使体内醛固酮水平升高,醛固酮的主要作 用是保钠、保水、排钾,从而使体内水钠潴留、血容量增加,引起血压升高,是临床上最常见的继发性 高血压[8]。PA 患者通常患有长期中、重度高血压,且对多种降压药疗效欠佳。PA 患病率在 1、2、3 级 高血压中分别为 1.99%、8.02%、13.2% [1]。长期的血压控制不佳使 PA 患者总体心血管事件发生率高于 EH 患者。已有研究表明, PA 中约有 50%的患者合并糖代谢紊乱以及 23%患者合并脂代谢异常[9]。目前 认为 PA 引起糖代谢异常的原因主要是醛固酮分泌过多影响胰岛 meta 细胞损害和靶器官对胰岛素敏感性降 低两个方面[10]。PA 导致脂代谢紊乱的机制目前尚不明确,有研究结果显示可能与高醛固酮影响脂肪细 胞的分化和功能有关[11]。PA 患者糖脂代谢的紊乱也会加重靶器官的损害。近年来根据经典遗传学方法 和基于二代测序的基因组测序、全外显子测序或转录组测序(RNA Sequencing, RNA-Seq)已报道过影响钾 通道的体细胞突变[12],影响 ATP 酶功能的体细胞突变[13],影响钙通道活性的 CACNA1D、CACNA1H 体细胞突变[14],影响氯通道的 CLCN2 体细胞突变,还有 CTNNB1 基因突变、PRKACA 基因突变等[15]。 生物信息学是将生物数据存储在计算机,并利用计算机作为工具对生物数据进行检索和分析的科学,涉 及生物学、医学统计学、应用数学以及计算机科学等学科知识。由于测序技术的不断提高和成本不断降 低,越来越多的领域应用到生物信息学分析。从疾病的筛查、诊断到治疗、预后,还有肿瘤个性化用药、 遗传病筛查等,通过从海量测序数据中挖掘有效的信息,为临床诊疗提供新思路[16]。基因表达综合数据 库(Gene Expression Omnibus, GEO)为全球研究学者提供高通量基因表达数据,通过 GEO 等数据库的公共 数据集进行生物信息学分析已广泛应用在基因组水平上筛选疾病的诊断、治疗和预后靶点[17],这些方法 帮助我们筛选原发性醛固酮增多症的差异表达基因(Differentially Expressed Genes, DEGs)。我们的研究发 现,原发性醛固酮增多症的核心基因可能是 ABCB4、AQP2、ARG2 和 CLRN1。ABCB4 基因最初发现 是与进行性家族性肝内胆汁淤积 3 型(PFIC3)的可遗传形式的胆汁淤积有关[18]。目前越来越多的研究表 明,该基因在各种肝胆疾病中都有参与。ABCB4 编码多药耐药蛋白 3 (MDR3),这是一种介导磷脂酰胆 碱(PC)流出胆汁的小管 ATP结合盒(ABC)蛋白。到目前为止,有证据表明双等位基因或单等位基因 ABCB4 缺陷会导致或诱发多种人类肝脏疾病(PFIC3、低磷脂相关胆石症综合征、妊娠肝内胆汁淤积症、药物性 肝损伤、短暂性新生儿淤积症)[19]。这种磷脂被认为可以保护肝细胞和胆管细胞的管腔膜免受游离胆汁

酸的去污作用[20]。AQP2 是广泛分布于哺乳动物组织中的水通道蛋白,在水平衡中起主要作用。AQP2 基因与胰岛素抵抗密切相关。血管紧张素Ⅱ和醛固酮可以通过改变细胞内 AOP2 靶向和/或 AOP2 丰度以 及加压素,在调节肾脏水的重吸收中发挥作用,引起血压升高[21]。目前有研究表明,AQP2蛋白在正常 肾上腺髓质组织和髓质肿瘤(包括 PCC、MPCC 和 NT)以及 EAPGL 中显著表达。AOP2 的表达可能提示 正常肾上腺组织的起源,其在肿瘤组织中的表达可能反映肿瘤发生过程中通过 AOP2 维持水代谢。AOP2 可作为鉴别诊断肾上腺肿瘤的一个有价值的指标[22]。精氨酸酶是尿素循环(UC)中普遍存在的酶,其将 L-精氨酸水解为尿素和L-鸟氨酸。两种哺乳动物精氨酸酶亚型,精氨酸酶1(ARG1)和精氨酸酶2(ARG2), 两者都是通过调节 L-精氨酸代谢、一氧化氮(NO)产生和炎症反应以及氧化应激,在β细胞功能、胰岛素 抵抗(IR)和血管并发症的调节中起重要作用[21]。研究证实 ARG2 基因与糖脂异常密切相关[23]。本研究 进一步分析了差核心基因涉及的具体信号通路,探讨核心基因影响醛固酮增多症进展的潜在分子机制。 低表达 ABCB4 主要富集的通路是 ALDOSTERONE-REGULATED-SODIUM-REABSORPTION、 GLUTATHIONE-METABOLISM、VEGF-SIGNALING-PATHWAY 等信号通路: 低表达的 AOP2 主要 富集的通路是 ALDOSTERONE-REGULATED-SODIUM-REABSORPTION 、 LEUKOCYTE-TRANSENDOTHELIAL-MIGRATION、KEGG-AUTOIMMUNE-THYROID-DISEASE 等信号通路: 高表 达 ARG2 主要富集的通路是 PROTEIN-EXPORT、RNA-DEGRADATION、AMINO-SUGAR-AND-NUCLEOTIDE-SUGAR-METABOLISM 等信号通路; 低表达 CLRN1 主要富集的通路是 TYROSINE-METABOLISM、ALDOSTERONE-REGULATED-SODIUM-REABSORPTION、AUTOIMMUNE-THYROID-DISEASE 等信号通路。提示核心基因通过这些信号通路影响醛固酮增多症的发生和发展。

### 6. 结论

上述研究帮助我们更好地理解基因改变如何参与 PA 的发生和发展,并确定应该研究哪些基因和通路。最重要的是,我们发现了 PA 的 4 个核心基因可能作用的信号通路,有助于进一步探讨核心基因影响醛固酮增多症进展的潜在分子机制。

# 参考文献

- [1] 医学会内分泌学分会. 原发性醛固酮增多症诊断治疗的专家共识(2020 版) [J]. 中华内分泌代谢杂志, 2020, 36(9): 727-736.
- [2] 黄素兰, 匡泽民, 陈国军, 等. 原发性醛固酮增多症靶器官损害的研究进展[J]. 心肺血管病杂志, 2016, 35(9): 765-767.
- [3] Reincke, M., et al. (2021) Diagnosis and Treatment of Primary Aldosteronism. The Lancet Diabetes & Endocrinology, 9, 876-892. <u>https://doi.org/10.1016/S2213-8587(21)00210-2</u>
- [4] Moore, L.D., Le, T. and Fan, G. (2013) DNA Methylation and Its Basic Function. *Neuropsychopharmacology*, 38, 23-38. <u>https://doi.org/10.1038/npp.2012.112</u>
- [5] 高志杰,姜茜,陈倩,等.第2代测序技术在甲基丙二酸尿症以及苯丙酮尿症诊断中的应用[J]. 医学研究杂志, 2015,44(3):111-114+168.
- [6] Shulaev, V. (2006) Metabolomics Technology and Bioinformatics. *Briefings in Bioinformatics*, **7**, 128-139. <u>https://doi.org/10.1093/bib/bbl012</u>
- [7] Gao, Y., et al. (2021) The Bioinformatics Analysis of Aldosterone-Producing Adenoma and Verification of Differentially Expressed Genes. International Journal of Endocrinology, 2021, Article ID: 4926323. <u>https://doi.org/10.21203/rs.3.rs-466199/v1</u>
- [8] 《原发性醛固酮增多症诊断治疗指南》解读[J]. 中国社区医师, 2011(6): 11.
- [9] Remde, H., Hanslik, G., Rayes, N. and Quinkler, M. (2015) Glucose Metabolism in Primary Aldosteronism. *Hormone and Metabolic Research*, 47, 987-993. <u>https://doi.org/10.1055/s-0035-1565208</u>
- [10] Chen, W., Li, F., He, C., Zhu, Y. and Tan, W. (2014) Elevated Prevalence of Abnormal Glucose Metabolism in Pa-

tients with Primary Aldosteronism: A Meta-Analysis. Irish Journal of Medical Science, **183**, 283-291. https://doi.org/10.1007/s11845-013-1007-x

- [11] Adolf, C., et al. (2016) Worsening of Lipid Metabolism after Successful Treatment of Primary Aldosteronism. Endocrine, 54, 198-205. <u>https://doi.org/10.1007/s12020-016-0983-9</u>
- [12] Yang, Y., et al. (2019) Primary Aldosteronism: KCNJ5 Mutations and Adrenocortical Cell Growth. Hypertension, 74, 809-816. <u>https://doi.org/10.1161/HYPERTENSIONAHA.119.13476</u>
- [13] Backman, S., et al. (2019) RNA Sequencing Provides Novel Insights into the Transcriptome of Aldosterone Producing Adenomas. Scientific Reports, 9, Article No. 6269. <u>https://doi.org/10.1038/s41598-019-41525-2</u>
- [14] Scholl, U.I. (2022) Genetics of Primary Aldosteronism. *Hypertension*, **79**, 887-897. https://doi.org/10.1161/HYPERTENSIONAHA.121.16498
- [15] Kamilaris, C.D.C., Hannah-Shmouni, F. and Stratakis, C.A. (2020) Adrenocortical Tumorigenesis: Lessons from Genetics. *Best Practice & Research Clinical Endocrinology & Metabolism*, 34, Article ID: 101428. https://doi.org/10.1016/j.beem.2020.101428
- [16] Oliver, G.R., Hart, S.N. and Klee, E.W. (2015) Bioinformatics for Clinical Next Generation Sequencing. *Clinical Chemistry*, 61, 124-135. <u>https://doi.org/10.1373/clinchem.2014.224360</u>
- [17] Clough, E. and Barrett, T. (2016) The Gene Expression Omnibus Database. In: Mathé, E. and Davis, S., Eds., Statistical Genomics: Methods and Protocols, Springer, Berlin, 93-110. <u>https://doi.org/10.1007/978-1-4939-3578-9\_5</u>
- [18] Olsen, J.A., Alam, A., Kowal, J., Stieger, B. and Locher, K.P. (2020) Structure of the Human Lipid Exporter ABCB4 in a Lipid Environment. *Nature Structural & Molecular Biology*, 27, 62-70. https://doi.org/10.1038/s41594-019-0354-3
- [19] Gordo-Gilart, R., et al. (2016) Heterozygous ABCB4 Mutations in Children with Cholestatic Liver Disease. Liver International, 36, 258-267. <u>https://doi.org/10.1111/liv.12910</u>
- [20] Davit-Spraul, A., Gonzales, E., Baussan, C. and Jacquemin, E. (2010) The Spectrum of Liver Diseases Related to ABCB4 Gene Mutations: Pathophysiology and Clinical Aspects. *Seminars in Liver Disease*, 30, 134-146. <u>https://doi.org/10.1055/s-0030-1253223</u>
- [21] Hasler, U., Leroy, V., Martin, P.-Y. and Féraille, E. (2009) Aquaporin-2 Abundance in the Renal Collecting Duct: New Insights from Cultured Cell Models. *American Journal of Physiology-Renal Physiology*, 297, F10-F18. https://doi.org/10.1152/ajprenal.00053.2009
- [22] Niu, D., Bai, Y.H., Yao, Q., Zhou, L.X., Huang, X.Z. and Zhao, C. (2020) AQP2 as a Diagnostic Immunohistochemical Marker for Pheochromocytoma and/or Paraganglioma. *Gland Surgery*, 2, 200-208. https://doi.org/10.21037/gs.2020.01.19
- [23] Ren, Y., et al. (2022) Arginase: Biological and Therapeutic Implications in Diabetes Mellitus and Its Complications. Oxidative Medicine and Cellular Longevity, 2022, Article ID: 2419412. <u>https://doi.org/10.1155/2022/2419412</u>