

# 生物信息学技术在肺动脉高压研究中的应用

张弛 董浩如 徐 骞 厉秀纯 陈马云

**摘要** 生物信息学技术是随着生命科学及信息学的发展形成的一种新兴技术,被广泛用于疾病的研究。肺动脉高压的研究虽然在近几十年来取得了较大的进展,但其发病机制仍未完全被阐明。近年来,生物信息学技术在研究肺动脉高压的生物学标志物、竞争性内源性 RNA (competing endogenous RNA, ceRNA)、蛋白组学、表观遗传学以及网络药理学等方面得到了较大的应用,并取得了一定的进展,本文就生物信息学技术在肺动脉高压研究中的应用现状做一综述。

**关键词** 肺动脉高压 生物信息学 高通量组学 网络药理学

**中图分类号** R542 **文献标识码** A **DOI** 10.11969/j.issn.1673-548X.2020.11.007

肺动脉高压(pulmonary artery hypertension, PAH)是一种以肺血管收缩和重塑为特征的进展性疾病,病死率高<sup>[1]</sup>。目前 PAH 的发病机制尚未明确,同时,临床上治疗 PAH 的药物疗效不佳<sup>[2]</sup>。随着 2001 年人类基因组计划的完成、后基因组时代高通量技术的快速发展,生物信息学技术已经成为研究疾病必不可少的一样工具<sup>[3]</sup>。目前生物信息学技术在 PAH 研究中逐渐被应用,各种高通量技术、在线数据库以及相关软件是研究过程中必不可少的。其中 GEO(Gene Expression Omnibus, <http://www.ncbi.nlm.nih.gov/geo>)数据库是当今最大的、最全面的公共基因表达数据库之一,通过高通量技术以及公共数据库的挖掘,大量的研究成果被报道。本文就生物信息学技术在 PAH 研究中的应用现状做一综述,为今后对 PAH 发病机制及靶向治疗的研究提供新思路。

## 一、利用生物信息学技术寻找肺动脉高压的生物学标志物

当前,各种高通量组学和生物信息学技术已被广泛的用于寻找疾病相关基因。在 PAH 研究中,研究人员通过挖掘公共数据库,在 mRNA、lncRNA、miRNA 等水平鉴定出许多的生物学标志物。例如 Wang 等<sup>[4]</sup>基于 PAH 数据集(GSE703)进行了加权基因共表达网络分析(weighted gene co-expression network

analysis, WGCNA),筛选出与 PAH 相关性最强的 2 个模块中前 10 位 hub 基因,其中变化最大的为 YWHAB,其在肺动脉高压患者血清中高表达,并与患者的肺动脉压力呈正比,而后使用 RT-PCR 进行验证,证实了 YWHAB 可作为 PAH 的生物学标志物和治疗靶点。Sun 等<sup>[5]</sup>分析了 PAH 小鼠数据集(GSE49114),鉴定出了 77 个上调和 520 个下调的差异表达基因,而后进行了基因本体论和信号通路富集分析,最后发现 Smad9、BMP2、Eng 和 IL-4 参与了 PAH 的发展,但本研究鉴定出的 PAH 相关生物学标志物并未在人类样本中得到验证。

在 lncRNA 水平上,Gu 等<sup>[6]</sup>利用随机方差模型筛选出 PAH 患者肺组织内的差异表达基因后,通过构建共表达网络发现 NR-036693、NR-027783、NR-033766、NR-001284 发生了明显改变,从而推断这 4 个 lncRNA 在肺动脉高压的发生、发展过程中起到了重要的作用。Han 等<sup>[7]</sup>联合 mRNA 与 lncRNA 分析,得出 2.511 个差异表达的 lncRNA 和 1169 个差异表达的 mRNA。其中上调的 lncRNA 有 2004 个,下调的 lncRNA 有 507 个;上调的 mRNA 有 609 个,下调的 mRNA 有 560 个。进一步通过富集分析,推断下调的 lncRNA 可能参与 PAH 形成,此外差异表达的 lncRNA 可作为 PAH 诊断标志物。此外也有 miRNA 作为 PAH 生物学标志物的研究被报道:Zhu 等<sup>[8]</sup>通过公共数据挖掘,发现 PAH 大鼠模型中 miR-140-5p 下调,进一步实验证明上调的 miR-140-5p 可以通过靶向抑制 TNF- $\alpha$  来缓解肺动脉高压的进展,因此,miR-140-5p 可以作为 PAH 的诊断及治疗靶点。另外有研究通过对 miRNA 芯片数据进行差异分析及功能注释,发现 miR-1183 在风湿性心脏病并

基金项目:浙江省自然科学基金资助项目(LQ19H010003);浙江省医药卫生科技计划项目(2019RC047);温州医科大学本专科学科研基金资助项目(wyx2019101157)

作者单位:325000 温州医科大学第一临床医学院(张弛、董浩如、徐骞);325000 温州医科大学附属第一医院呼吸与危重症医学科、温州市呼吸循环重点实验室(厉秀纯、陈马云)

通讯作者:陈马云,电子邮箱:chenmayun@126.com

发肺动脉高压患者中过表达,并参与了肺动脉的重构<sup>[9]</sup>。以上研究结果表明生物信息学技术在寻找疾病标志物方面具有极大优势。

## 二、生物信息学分析揭示 ceRNA 在肺动脉高压中的作用

由 Salmena 等<sup>[10]</sup>提出的竞争性内源性 RNA 假说阐明了编码 RNA 和非编码 RNA 的相互作用在疾病的产生和发展过程中起到了重要作用,并且得到了大量的实验证明。然而,低通量的实验方法在构建与疾病相关的 ceRNA 网络时,具有成本高、效率低的缺点,而生物信息学技术恰好能弥补其不足。目前有许多通过生物信息学技术构建 PAH 相关 ceRNA 网络的研究被报道。Wang 等<sup>[11]</sup>使用高通量芯片在 PAH 小鼠肺组织中鉴定出 12 个差异表达的 circRNA,选择其中差异最大的两个 circRNA (mmu\_circRNA\_004592 和 mmu\_circRNA\_018351),利用了 TargetScan 与 miRanda 在内的多种生物信息学工具预测了靶向 miRNA 与 Mrna,使用了 Cytoscape 软件构建了 ceRNA 网络,结果显示这些差异 circRNA 可作为 PAH 的诊断和治疗靶点。

另有研究结合差异的 miRNA 与 circRNA 来构建 ceRNA 网络,如 Miao 等<sup>[12]</sup>应用高通量技术,在 PAH 患者外周血液中检测到了 212 个差异表达的 miRNA 和 61 个差异表达的 circRNA,而后构建了 miRNA - circRNA 调控网络,发现 hsa\_circ\_004615 可作为 miR - 1226 - 3p“海绵”来调控 ATP2A2 的表达,进而影响肺动脉高压的发生、发展。此外,也有研究通过建立 PAH 相关的 lncRNA - gene - miRNA 互作网络来阐明 PAH 发生、发展的潜在机制,研究结果显示基于 PDGFRB 和 HIF - 1 $\alpha$  的 ceRNA 网络 (miRNAs - PDGFRB - lncRNAs 和 miRNAs - HIF - 1 $\alpha$  - lncRNAs) 在 PAH 发展过程中起到了关键的作用<sup>[13]</sup>。Zhao 等<sup>[14]</sup>通过分析 587 例 PAH 患者和 736 例健康对照组外周血的单核苷酸多态性,发现 lncRNA MALAT1 中 rs619586A > G 单核苷酸多态性与 PAH 形成的相关性最高,进一步分析表明,变异的 MALAT1 可作为 miR - 214 的“海绵”进而影响 XBP1 的表达,对 PAH 的形成有保护作用。也有研究通过不同的算法模型来构建 ceRNA 网络,如 Feng 等<sup>[15]</sup>构建了一种新的算法模型——ce - Subpathway,在 PAH 数据集 (GSE33463) 中确定了 31 个由 ceRNA 介导的功能子通路,进一步的分析结果发现 miR - 30 家族在 EP300 和 JUN、CREBBP 和 TCF7L2、FBXW11 和 EP300 的

ceRNA 机制中有显著的调控作用。

综上所述,研究人员主要是通过生物信息学工具,预测 RNA 之间的靶向调控关系来构建 ceRNA 相关网络,而后通过富集分析、生存分析、分子生物学实验等验证其在疾病中的调控作用,生物信息学技术在其中发挥了重要的作用。

## 三、生物信息学技术用于肺动脉高压网络药理学研究

网络药理学是一门用于阐述疾病发生、发展,探究药物机体相互作用的新兴学科,在阐明疾病发生机制与中药药理学机制发挥了巨大的作用,现如今也有许多在线数据库被开发和应用,然而,目前网络药理学主要应用于癌症等领域的研究,在 PAH 领域只有少数研究被报道。如 Chen 等<sup>[16]</sup>研究显示,利用网络药理学方法,挖掘出染料木黄酮作用的靶点,通过构建蛋白互作网络和富集分析,发现该药物的抗 PAH 作用与凋亡信号通路和一氧化氮合成过程密切相关,接着通过分子对接模拟,发现染料木黄酮可与过氧化物酶体增殖物激活受体  $\gamma$  (PPAR $\gamma$ ) 直接作用,发挥抗肺动脉高压的作用。也有研究结合在线数据库 cMap (Connectivity Map) 与表达谱数据,进行差异分析与功能注释,发现活性氧的形成可能参与 PAH 的发生、发展,随后作者将差异基因分为上调与下调两组,通过在线数据库的检索,鉴定出了许多可抑制活性氧产生的药物<sup>[17]</sup>。但是,目前被报道的研究都缺乏体内实验的验证,因此通过网络药理学研究获得的药物需要通过进一步的细胞、动物实验以及大规模人群的随机对照实验来确定其临床疗效。

## 四、生物信息学技术用于肺动脉高压的蛋白质组学研究

蛋白质组学是对一个细胞或一种组织中全部蛋白质进行大规模分析的一门新兴学科,在多种疾病的研究中发挥了较大的作用。蛋白质组学在 PAH 领域研究已有多年,多种质谱 (MS) 分析方法被开发用于寻找 PAH 相关的生物学标志物,如在 2006 年便有文章报道:通过 SELDI - TOF MS 对 PAH 患者血液样本进行分析,鉴定出 234 个差异蛋白,进一步通过多因素回归分析发现质荷比为 8600 的离子是最有效的 PAH 候选生物学标志物<sup>[18]</sup>。

近年来蛋白质组学方法在揭示 PAH 潜在发病机制中也得到了广泛的应用,Meirick 等<sup>[19]</sup>通过 2D - DIGE/MS 分析 PAH 患者血清,得到了 9 个上调蛋白,7 个下调蛋白,进一步的分析发现其中一个蛋白

Grb2 参与 BMPR2 受体的信号转导,进而影响家族性肺动脉高压的发生、发展。此外,Xu 等<sup>[20]</sup>使用 LC-MS/MS,取 4 例 PAH 患者及 5 例健康对照的肺组织,进行了全局蛋白组学及磷酸蛋白组学分析,结果显示有 170 个蛋白和 240 个磷酸肽差异表达,其中 45 个蛋白和 18 个磷酸肽位于线粒体中,表明线粒体相关代谢途径的改变参与 PAH 的发生、发展。至于评价药物的治疗效果,Yao 等<sup>[21]</sup>通过检测加药动物模型相关蛋白表达,结合进一步的表型实验,证明了 osthole 具有治疗 PAH 的作用。YEAGER 等<sup>[22]</sup>研究了 8 例经扩血管治疗后预后良好患者和 7 例预后不良患者的血浆蛋白的差异水平,发现 SAA-4 在预后良好的患者中降低了 4 倍,在预后不良的患者中升高了 2 倍;paraoxonase/arylesterase-1 在预后良好的患者中升高了 2 倍;在预后良好的患者中,SAP 比治疗前降低了 1.3 倍;预后不良患者治疗后,结合珠蛋白和血凝蛋白分别降低了 1.45 和 1.80 倍。这些结果表明这些血浆蛋白可以作为评价 PAH 扩血管治疗预后的指标。总之,蛋白质组学在 PAH 研究中较早便得到了应用,近年来发表的研究较少,但其涉及了 PAH 相关生物学标志物、PAH 发病机制以及药物疗效评价等各个方面,应用较为广泛。

### 五、生物信息学技术用于肺动脉高压表观遗传学的研究

当前针对表观遗传学的研究包括了 DNA 甲基化、组蛋白修饰以及染色质重塑等,在已发表的研究中,利用生物信息学技术研究 PAH 的表观遗传学主要体现在 DNA 甲基化。如 Wang 等<sup>[23]</sup>利用高通量技术在肺动脉平滑肌细胞中检测到 6829 个 DNA 甲基化差异位点,其中高甲基化位点 4246 个,低甲基化位点 2583 个,将差异的甲基化位点进行基因功能与信号通路的富集分析,发现这些基因参与了细胞增殖、凋亡与迁移等生物学过程。筛选未被报道的 3 个基因(PIK3CA、HRAS 和 HIC1),使用焦磷酸测序来验证其上游启动子区甲基化水平,发现 HIC1 甲基化水平显著升高,而 PIK3CA 和 HRAS 甲基化水平显著降低,而后通过 RT-PCR 进一步验证了相对应的 mRNA 的表达。

有研究通过检测启动子区 CpG 岛,发现在不同病因 PAH 患者的肺动脉内皮细胞中部分基因的甲基化程度不同,通过主成分分析验证了不同病因 PAH 甲基化基因谱的差别,而后作者通过对编码转运蛋白的 46 个基因进行 Meta 分析与富集分析,发现 ABCA1

甲基化水平差异最明显,其参与调控了脂代谢,最后 qPCR 的结果证实了 PAH 患者肺组织中相应 mRNA 表达下调<sup>[24]</sup>。除此之外,有研究仅通过人类甲基化芯片,鉴定出风湿性心脏病并发肺动脉高压患者血液中共有 40 个低甲基化位点与 64 个高甲基化位点,但缺少了进一步实验的验证,而后此研究进行了富集分析,结果显示蛋白激酶/转移酶活性发生了变化,这些发现可以给相关基础研究人员提供新的思考方向<sup>[25]</sup>。总体而言,利用生物信息学对 PAH 表观遗传学研究主要集中于 DNA 甲基化,其他方面如乙酰化等,主要通过细胞动物实验直接探究其在 PAH 发生、发展中的作用,生物信息学技术应用较少。

### 六、展 望

随着各种高通量技术的快速发展,疾病研究已经进入了大数据整合分析的时代。通过利用各种芯片技术、测序技术以及各种在线数据库,研究人员完成了许多关于 PAH 发生、发展机制的研究,也鉴定出大量 PAH 相关的生物学标志物。然而,目前针对 PAH 的分析主要集中在某一特定组学以及单个数据,存在样本量少、可重复性差的问题,而整合多组学数据以及联合多数据集分析在癌症领域已经得到了广泛的应用,虽然已经开发出许多相关的算法、工具,但是多组学数据与多数据集的集成仍存在一定的误差。在未来,不断优化的算法以及不断更新的生物信息学技术可以帮助科研人员整合多维度的数据,寻找更加可靠的生物学标志物,推动相关分子机制的研究,最后可以更好地实现对 PAH 患者进行准确的诊断和个体化的靶向治疗。

### 参考文献

- 1 Simonneau G, Montani D, Celermajer DS, *et al.* Haemodynamic definitions and updated clinical classification of pulmonary hypertension [J]. *Eur Respir J*, 2019, 53(1): 1801913
- 2 Thenappan T, Ormiston ML, Ryan JJ, *et al.* Pulmonary arterial hypertension: pathogenesis and clinical management [J]. *BMJ*, 2018, 360: j5492
- 3 van Dijk EL, Jaszczyszyn Y, Naquin D, *et al.* The third revolution in sequencing technology [J]. *Trends Genet*, 2018, 34(9): 666-681
- 4 Wang T, Zheng X, Li R, *et al.* Integrated bioinformatic analysis reveals ywhab as a novel diagnostic biomarker for idiopathic pulmonary arterial hypertension [J]. *J Cell Physiol*, 2019, 234(5): 6449-6462
- 5 Sun Y, Lin X, Li L. Identification of biomarkers for schistosoma-associated pulmonary arterial hypertension based on rna-seq data of mouse whole lung tissues [J]. *Lung*, 2017, 195(3): 377-385
- 6 Gu S, Li G, Zhang X, *et al.* Aberrant expression of long noncoding rnas in chronic thromboembolic pulmonary hypertension [J]. *Mol*

Med Rep, 2015, 11(4): 2631 – 2643

- 7 Han B, Bu P, Meng X, *et al.* Microarray profiling of long non – coding rnas associated with idiopathic pulmonary arterial hypertension [J]. *Exp Ther Med*, 2017, 13(6): 2657 – 2666
- 8 Zhu TT, Zhang WF, Yin YL, *et al.* MicroRNA – 140 – 5p targeting tumor necrosis factor – alpha prevents pulmonary arterial hypertension [J]. *J Cell Physiol*, 2019, 234(6): 9535 – 9550
- 9 Li N, Lian J, Zhao S, *et al.* Detection of differentially expressed microRNAs in rheumatic heart disease: mir – 1183 and mir – 1299 as potential diagnostic biomarkers [J]. *Biomed Res Int*, 2015, 2015: 524519
- 10 Liu Y, Xue M, Du S, *et al.* Competitive endogenous rna is an intrinsic component of emt regulatory circuits and modulates emt [J]. *Nat Commun*, 2019, 10(1): 1637
- 11 Wang J, Zhu MC, Kalionis B, *et al.* Characteristics of circular rna expression in lung tissues from mice with hypoxia-induced pulmonary hypertension [J]. *Int J Mol Med*, 2018, 42(3): 1353 – 1366
- 12 Miao R, Gong J, Zhang C, *et al.* Hsa\_circ\_0046159 is involved in the development of chronic thromboembolic pulmonary hypertension [J]. *J Thromb Thrombolys*, 2019, 49(3): 386 – 394
- 13 Wang M, Gu S, Liu Y, *et al.* Mirna – pdgfrb/hif1a – lncrna ctephal network plays important roles in the mechanism of chronic thromboembolic pulmonary hypertension [J]. *Int Heart J*, 2019, 60(4): 924 – 937
- 14 Zhuo Y, Zeng Q, Zhang P, *et al.* Functional polymorphism of lncrna malat1 contributes to pulmonary arterial hypertension susceptibility in chinese people [J]. *Clin Chem Lab Med*, 2017, 55(1): 38 – 46
- 15 Feng C, Song C, Ning Z, *et al.* Ce – subpathway: identification of cerna – mediated subpathways via joint power of cernas and pathway topologies [J]. *J Cell Mol Med*, 2019, 23(2): 967 – 984
- 16 Chen Y, Chen D, Liu S, *et al.* Systematic elucidation of the mechanism of genistein against pulmonary hypertension via network pharmacology approach [J]. *Int J Mol Sci*, 2019, 20(22): 5569
- 17 Flynn C, Zheng S, Yan L, *et al.* Connectivity map analysis of non-sense – mediated decay – positive bmpr2 – related hereditary pulmonary arterial hypertension provides insights into disease penetrance [J]. *Am J Respir Cell Mol Biol*, 2012, 47(1): 20 – 27
- 18 Abdul – Salam VB, Paul GA, Ali JO, *et al.* Identification of plasma protein biomarkers associated with idiopathic pulmonary arterial hypertension [J]. *Proteomics*, 2006, 6(7): 2286 – 2294
- 19 Meyrick BO, Friedman DB, Billheimer DD, *et al.* Proteomics of transformed lymphocytes from a family with familial pulmonary arterial hypertension [J]. *Am J Respir Crit Care Med*, 2008, 177(1): 99 – 107
- 20 Xu W, Comhair SAA, Chen R, *et al.* Integrative proteomics and phosphoproteomics in pulmonary arterial hypertension [J]. *Sci Rep*, 2019, 9(1): 18623
- 21 Yao L, Yang Y, He G, *et al.* Global proteomics deciphered novel – function of osthole against pulmonary arterial hypertension [J]. *Sci Rep*, 2018, 8(1): 5556
- 22 Yeager ME, Colvin KL, Everett AD, *et al.* Plasma proteomics of differential outcome to long – term therapy in children with idiopathic pulmonary arterial hypertension [J]. *Proteomics Clin Appl*, 2012, 6(5 – 6): 257 – 267
- 23 Wang Y, Huang X, Leng D, *et al.* DNA methylation signatures of pulmonary arterial smooth muscle cells in chronic thromboembolic pulmonary hypertension [J]. *Physiol Genomics*, 2018, 50(5): 313 – 322
- 24 Hautefort A, Chesne J, Preussner J, *et al.* Pulmonary endothelial cell DNA methylation signature in pulmonary arterial hypertension [J]. *Oncotarget*, 2017, 8(32): 52995 – 53016
- 25 Zheng D, Chen X, Li N, *et al.* Differentially methylated regions in patients with rheumatic heart disease and secondary pulmonary arterial hypertension [J]. *Exp Ther Med*, 2017, 14(2): 1367 – 1372

(收稿日期: 2020 – 04 – 30)

(修回日期: 2020 – 05 – 12)

(上接第 28 页)

- 23 Fulton DJR, Li X, Bordan Z, *et al.* Galectin – 3: a harbinger of reactive oxygen species, fibrosis, and inflammation in pulmonary arterial hypertension [J]. *Antioxid Redox Sign*, 2019, 31(14): 1053 – 1069
- 24 Plecítá – Hlavatá L, Tauber J, Li M, *et al.* Constitutive reprogramming of fibroblast mitochondrial metabolism in pulmonary hypertension [J]. *Am J Respir Cell Mol Biol*, 2016, 55(1): 47 – 57
- 25 Rafikova O, Al Ghoulh I, Rafikov R. Focus on early events: pathogenesis of pulmonary arterial hypertension development [J]. *Antioxid Redox Signal*, 2019, 31(13): 933 – 953
- 26 Pullamsetti SS, Savai R, Seeger W, *et al.* Translational advances in the field of pulmonary hypertension. From cancer biology to new pulmonary arterial hypertension therapeutics. Targeting cell growth and proliferation signaling hubs [J]. *Am J Respir Crit Care Med*, 2017, 195(4): 425 – 437
- 27 D’Alessandro A, El Kasmi KC, Plecítá – Hlavatá L, *et al.* Hallmarks of pulmonary hypertension: mesenchymal and inflammatory cell metabolic reprogramming [J]. *Antioxid Redox Signal*, 2018, 28(3): 230 – 250
- 28 Li M, Riddle S, Zhang H, *et al.* Metabolic reprogramming regulates the proliferative and inflammatory phenotype of adventitial fibroblasts in pulmonary hypertension through the transcriptional corepressor C – terminal binding protein – 1 [J]. *Circulation*, 2016, 134(15): 1105 – 1121
- 29 Wakasugi T, Shimizu I, Yoshida Y, *et al.* Role of smooth muscle cell p53 in pulmonary arterial hypertension [J]. *PLoS One*, 2019, 14(2): e0212889
- 30 Zhang H, Wang D, Li M, *et al.* Metabolic and proliferative state of vascular adventitial fibroblasts in pulmonary hypertension is regulated through a MiR – 124/PTBP1/PKM axis [J]. *Circulation*, 2017, 136(25): 2468 – 2485

(收稿日期: 2020 – 05 – 15)

(修回日期: 2020 – 05 – 21)