

华中科技大学

本科毕业论文

多环芳烃代谢物相关的血浆中 microRNAs 表达谱研究

院 系 公共卫生学院

专业班级 预防医学 201803 班

姓 名 李若樨

学 号 U201812958

指导教师 邬堂春

2023 年 4 月 20 日

## 摘 要

**目的:** 多环芳烃 (polycyclic aromatic hydrocarbons, PAHs) 是一类由两个或更多的苯环结合而成的芳香烃类有机物, 具有遗传毒性、致癌性、免疫毒性和致畸性。现有研究表明, 多环芳烃暴露可以改变微小核糖核酸 (microRNA, miRNA) 的表达, 其发挥基因调控的作用, 影响机体表型, 甚至引起疾病的发生发展。然而当前的流行病学研究大多针对高暴露的职业人群, 对社区人群的研究较少, 且针对社区人群的研究多聚焦于某一特定人群或仅研究与某些特定 miRNA 的关联, 缺乏非预设的 miRNA 表达谱研究。为此, 本研究针对相对健康的社区人群, 旨在识别 PAHs 暴露相关的 miRNA, 探索 miRNA 的候选靶基因及其富集的生物学功能, 探讨暴露于 PAHs 潜在的健康损害和生物学机制。

**方法:** 本次研究在十堰社区选取相对健康的 160 人作为研究对象, 分别用气相色谱-质谱联用仪和酶联免疫吸附试验测定研究对象的 10 种尿羟基化多环芳烃 (Monohydroxy polycyclic aromatic hydrocarbons, OH-PAHs) 和血浆苯并[a]芘-7,8-二氢二醇-9,10-环氧化物 (Benzo[a]pyrene-diol-epoxide, BPDE)-白蛋白加合物作为 PAHs 内暴露标志物, 并采用 RNA 测序的方法测定血浆 miRNA 水平。使用多元线性回归评估 OH-PAHs 和 BPDE-白蛋白加合物浓度与 miRNA 水平之间的关联, 并针对年龄、性别、BMI、当前吸烟状况、当前饮酒状况和被动吸烟展开分层分析和交互作用分析。在 TissueAtlas 数据库中探索 miRNA 的组织特异性及其潜在组织来源。在 Targetscan、miRDB、DIANA-microT-CDS 和 miRTarBase 数据库寻找 miRNA 的候选靶基因, 并进一步通过 DAVID 数据库对候选靶基因进行富集分析。

**结果:** 本研究识别出 42 个与 1-羟基菲 (1-hydroxyphenanthrene, 1-OH-Phe)、9-羟基菲 (9-hydroxyphenanthrene, 9-OH-Phe) 和 2-羟基茚 (2-hydroxyfluorene, 2-OH-Flu) 显著相关的 miRNA ( $FDR < 0.10$ ), 其中有 4 个 miRNA 达到  $FDR < 0.05$  的筛选条件, 分别为与 2-OH-Flu 显著正相关的 miR-505-5p、miR-125b-2-3p 和 miR-483-5p 以及显著负相关的 miR-548e-3p, 本研究针对这 4 个最显著相关的 miRNA 进行进一步功能探索。分层分析显示在年龄  $< 45.0$  岁的人群中, 2-OH-Flu 与 miR-125b-2-3p、miR-483-5p、miR-548e-3p 之间的关联更加显著; 在否认存在被动吸烟的人群中, 2-OH-Flu 与 miR-483-5p、miR-548e-3p 之间的关联更加显著; 在 BMI  $< 25.0 \text{ kg/m}^2$  的人群中, 2-OH-Flu 与 miR-483-5p 间的关联更加显著。且 BMI

和 2-OH-Flu 对 miR-505-5p 和 miR-483-5p 表达水平存在交互作用。在 TissueAtlas 数据库中未发现这 4 个 miRNA 具有明显组织特异性。分别预测到 379 个 miR-505-5p 的候选靶基因, 534 个 miR-125b-2-3p 的候选靶基因, 168 个 miR-483-5p 的候选靶基因和 279 个 miR-548e-3p 的候选靶基因。富集分析发现 4 个 miRNA 的靶基因功能富集在心血管疾病、氧化应激、遗传损伤、免疫调控、炎症反应、癌症、丙型肝炎、单纯疱疹病毒 1 型感染、钙稳态、葡萄糖稳态、神经系统发育、胚胎发育、多个器官发育、记忆等通路与生物学过程上。

**结论:** 在相对健康的社区人群中, 我们识别出 42 个 PAHs 暴露相关的 miRNA, 其中有 4 个与 2-OH-Flu 相关的 miRNA 最为显著。综合其他已经发表的研究结果, 我们推测, 这 4 个 miRNA 可能介导了机体对 PAHs 的应答反应, 可能是 PAHs 导致畸形、癌症、糖尿病、心血管疾病、免疫失调、炎症反应、氧化应激和遗传损伤的重要介导机制。

**关键词:** 多环芳烃, 多环芳烃代谢物, 血浆 microRNA

## Abstract

**Backgrounds:** Polycyclic aromatic hydrocarbons (PAHs) are a class of aromatic hydrocarbons linked by two or more benzene rings, which are genotoxic, carcinogenic, immunotoxic, and teratogenic. Current studies have shown that exposure to PAHs can alter microRNA (miRNA) expression, which could further affect gene expression and physiological and pathological processes. Most of the current epidemiological studies focused on occupational groups with high PAHs exposure, and few studies were conducted among community populations; moreover, most of the studies on community populations focused on a specific group of people or only examined the association between PAHs exposure and certain specific miRNAs, showing a lack of non-predefined miRNA expression profiles. In this study, we focused on a generally healthy community population, searched for miRNAs associated with PAHs exposure and explored the candidate target genes of miRNAs and their enriched biological functions, in order to investigate the potential health damage of PAHs exposure and potential biological mechanisms.

**Methods:** In this study, 160 generally healthy participants in Shiyan community were selected as the study population. We measured ten urinary monohydroxy polycyclic aromatic hydrocarbons (OH-PAHs) by gas chromatography-mass spectrometry, plasma benzo[a]pyrene-diol-epoxide (BPDE)-albumin adducts by enzyme-linked immunosorbent assay and used them as internal exposure biomarkers of PAHs. We then measured plasma miRNA levels by RNA sequencing technique. Associations of the OH-PAHs concentrations and BPDE-albumin adduct concentrations with miRNAs levels were assessed by multivariable linear regression. Stratified analyses and interaction analyses were performed for age, sex, BMI, current smoking status, current drinking status and passive smoking. We explored the tissue specificity of miRNAs and their potential tissue sources in the TissueAtlas database. Candidate target genes of miRNAs were then searched in Targetscan, miRDB, DIANA-microT-CDS and miRTarBase databases, and further enrichment analysis of candidate target genes was

performed using the DAVID database.

**Results:** In this study, 1-hydroxyphenanthrene (1-OH-Phe), 9-hydroxyphenanthrene (9-OH-Phe) and 2-hydroxyfluorene (2-OH-Flu) were identified to be significantly associated with 42 miRNAs ( $FDR < 0.10$ ). Among these miRNAs, four miRNAs reached a more rigorous criteria of  $FDR < 0.05$ , including 2-OH-Flu positively associated miR-505-5p, miR-125b-2-3p, and miR-483-5p and negatively associated miR-548e-3p. Further functional exploration was performed within four most significantly associated miRNAs. Stratified analysis showed that 2-OH-Flu was more significantly associated with miR-125b-2-3p, miR-483-5p, and miR-548e-3p in those aged  $< 45.0$  years; 2-OH-Flu was more significantly associated with miR-483-5p, miR-548e-3p in those without passive smoking; 2-OH-Flu was more significantly associated with miR-483-5p in those with BMI  $< 25.0$  kg/m<sup>2</sup>. And there was interaction between 2-OH-Flu and BMI regarding the expression levels of miR-505-5p and miR-483-5p. There was no significant tissue specificity for miRNAs. There were 379 candidate target genes identified for miR-505-5p, 534 for miR-125b-2-3p, 168 for miR-483-5p, and 279 for miR-548e-3p, respectively. Enrichment analysis revealed that the functions of candidate target genes of four miRNAs were enriched in cardiovascular diseases, oxidative stress, genetic damage, immune regulation, inflammatory response, cancer, hepatitis C, herpes simplex virus type 1 infection, calcium homeostasis, glucose homeostasis, neurological development, embryonic development, organ development, memory, and other pathways and biological processes.

**Conclusions:** In a generally healthy community population, we identified 42 miRNAs that were significantly associated with PAHs exposure. Four of the miRNAs associated with 2-OH-Flu reached a more rigorous criteria and were used for further functional exploration. Taken together with the results of other published studies, we hypothesize that miRNAs may mediate the body's response to PAHs and may be important mediators in PAHs leading to malformation, cancer, diabetes, cardiovascular diseases, immune dysregulation, inflammatory response, oxidative stress and genetic damage.

**Key Words:** polycyclic aromatic hydrocarbons; polycyclic aromatic hydrocarbons  
metabolites; plasma microRNA

# 目 录

摘要.....	I
Abstract.....	III
1 绪论.....	1
2 方法.....	4
2.1 研究人群.....	4
2.2 尿 OH-PAHs 和血浆 BPDE-白蛋白加合物浓度检测.....	4
2.3 血浆 miRNA 定量分析.....	5
2.4 协变量的评估.....	5
2.5 探索 miRNA 的组织特异性.....	5
2.6 寻找 miRNA 的候选靶基因.....	6
2.7 探讨候选靶基因的生物学功能.....	7
2.8 统计分析.....	7
3 结果.....	9
3.1 基线描述.....	9
3.2 OH-PAHs 和 BPDE-白蛋白加合物之间的关联性.....	10
3.3 识别与 OH-PAHs 和 BPDE-白蛋白加合物相关的 miRNA.....	11
3.4 针对混杂因素的分层分析和交互作用分析.....	13
3.5 探索 miRNA 的关联性、组织特异性和候选靶基因.....	15
3.6 miRNA 候选靶基因的 KEGG 通路富集分析.....	15
3.7 miRNA 候选靶基因的 GO 富集分析.....	18
3.7.1 miR-505-5p 候选靶基因的 GO 富集分析结果.....	18
3.7.2 miR-125b-2-3p 候选靶基因的 GO 富集分析结果.....	22
3.7.3 miR-483-5p 候选靶基因的 GO 富集分析结果.....	26
3.7.4 miR-548c-3p 候选靶基因的 GO 富集分析结果.....	30
4 讨论.....	34
5 结论.....	38
致谢.....	39
参考文献.....	40

# 1 绪论

多环芳烃 (Polycyclic aromatic hydrocarbons, PAHs) 是由两个或更多的苯环以线性、簇状或  $\alpha$  角状排列方式结合而成且不含杂原子或取代基的芳香烃, 具有各种各样的结构构型<sup>[1, 2]</sup>。PAHs 在环境中无处不在。具有两至四个环的 PAHs 在水中的溶解度较低且挥发性较高, 主要以气态形式存在。具有五个或更多环的 PAHs 的溶解度较低且挥发性低, 主要以固体形式存在, 多与土壤、空气或沉积物中的微粒结合<sup>[3]</sup>。自然或人为的不完全燃烧, 已被确定为环境中 PAHs 的最大来源<sup>[4]</sup>。PAHs 在自然界中具有持久性, 人类在职业和非职业环境中都可能因为吸入、摄取和皮肤接触而暴露于 PAHs。职业性接触主要在职业人群, 包括采矿、金属加工或炼油的工人、机械工、街头小贩、机动车司机等。社区人群接触 PAHs 的途径则主要是呼吸室外和室内空气、食用含有 PAHs 的食物和吸烟<sup>[5-8]</sup>。由于亲脂性, PAHs 进入机体后生物利用度很高<sup>[9, 10]</sup>, 在体内通过以混合功能氧化酶为主的多种酶进行代谢激活, 形成多种 PAHs 代谢物, 主要以代谢物的形式经肾脏随尿排出。

接触 PAHs 造成的各种各样的不良健康影响是一个重要的公共健康问题<sup>[11]</sup>。职业性地接触含有高浓度 PAHs 的物质, 短期会导致眼睛刺激、恶心、呕吐、腹泻、皮肤刺激、精神错乱等症状<sup>[12, 13]</sup>。长期或慢性接触 PAHs 对健康的影响包括免疫功能下降、白内障、肾脏和肝脏损害 (如黄疸)、哮喘样症状和肺功能异常。同时, PAHs 反复接触皮肤可能会诱发红肿和皮肤炎症<sup>[14-16]</sup>。一些研究显示, PAHs 暴露与呼吸系统疾病如哮喘、慢性阻塞性肺病、肺喘息, 以及心血管疾病如胸闷和心率变化、血压升高和缺血性心脏病分别有关<sup>[17]</sup>。PAHs 在酶解过程中, 生成多种致癌代谢产物和大量活性氧自由基, 造成癌症、氧化性 DNA 损伤和氧化应激。PAHs 已被证明有遗传毒性、致癌性、免疫毒性和致畸性<sup>[18]</sup>。PAHs 对人类健康的影响主要取决于接触的时间和途径、接触的 PAHs 种类和浓度, 以及 PAHs 的相对毒性。其他因素如机体的健康状况和年龄, 也会造成不同的影响<sup>[9]</sup>。

微小核糖核酸 (microRNA, miRNA) 是一种内源性的短单链非编码 RNA, 通过与目标 mRNA 在 3' 非编码区域 (3 untranslated region, 3' UTR) 等区域结合抑制或降解 mRNA 的翻译, 发挥基因调控的作用<sup>[11]</sup>, 调节细胞功能以及机体生理和病理过程, 包括细胞的生长和死亡、发育和分化、内分泌平衡和致癌<sup>[19-21]</sup>。虽然大部分 miRNA 驻留在细胞内, 但也有部分通过微囊释放到细胞外空间, 通过体液



转运到其靶器官组织细胞,在各种生理和病理过程中发挥作用<sup>[22]</sup>。研究表明,这些细胞外循环的 miRNA 可以在无细胞环境中持续相当长的时间,因此它们可以作为多种疾病的诊断和预后生物标志物,如癌症、糖尿病、自身免疫和心血管疾病<sup>[11]</sup>。例如血清 miR-221 被认为是甲状腺癌与黑色素瘤的诊断与预后标志物<sup>[23]</sup>。此外,miRNA 的表达水平易受到环境因素的影响<sup>[11]</sup>。越来越多的体内与体外研究结果显示,PAHs 暴露可以改变 miRNA 的表达水平,从而影响基因表达与机体的表型<sup>[11, 18]</sup>。多项研究表明,PAHs 中致癌性最高的苯并[a]芘(benzo[a]pyrene, B[a]P)可代谢形成苯并[a]芘-7,8-二氢二醇-9,10-环氧化物(Benzo[a]pyrene-diol-epoxide, BPDE),可在恶性转化的人类支气管上皮细胞中诱导 miR-506 表达下调和 miR-22 表达上调<sup>[24-26]</sup>。流行病学研究也提示 PAHs 暴露与 miRNA 的表达水平有关。一项研究显示,职业性 PAHs 暴露与血浆 let-7i-3p 表达下调和 miR-885-5p 与 miR-20a-5p 表达上调相关<sup>[27]</sup>。然而当前的流行病学研究大多针对高暴露的职业人群(如焦炉工人、煤矿工人),对社区人群的研究较少,且针对社区人群的研究多聚焦于某一特定人群或仅研究 PAHs 与某些特定 miRNA 的关联,缺乏非预设的 miRNA 表达谱研究。

PAHs 在环境中无处不在,评估 PAHs 的外暴露难以反映机体内的真实内暴露。由于尿液的无创和易获取性,可以将尿液 PAHs 代谢物作为反映 PAHs 内暴露的标志物,以实施有效便捷的健康监护<sup>[28]</sup>。随着技术发展,气相色谱-质谱联用法(Gas chromatography-mass spectrometry, GS-MS)被用于分析尿中多种多环芳烃羟基代谢产物——羟基化多环芳烃(Monohydroxy polycyclic aromatic hydrocarbons, OH-PAHs),从而为相对全面的评价 PAHs 暴露提供了可能<sup>[29, 30]</sup>。但尿 OH-PAHs 仅能反映机体 24 小时以内的暴露情况,因此有必要联合使用能反映长期暴露情况的内暴露指标。B[a]P 经过代谢活化生成的 BPDE 是一种亲核代谢产物,其与生物大分子(如 DNA、蛋白质)结合形成的加合物具有较长的半衰期。人血中白蛋白的含量远超 DNA 的含量,因此 BPDE-白蛋白加合物更适合作为评价指标<sup>[31]</sup>。BPDE-白蛋白加合物半衰期约为 28 天,能够反映较为长期的 PAHs 暴露<sup>[32]</sup>。因此可以联合尿 OH-PAHs 和血浆 BPDE-白蛋白加合物综合反映人群的 PAHs 内暴露水平。而血浆中存在的游离 miRNA 具有性质稳定、抗 RNA 酶降解、容易提取、检测方法成熟简便等优点,是广为运用的生物标志物<sup>[33, 34]</sup>。

基于以上研究背景,本次研究采用人群的 10 种尿 OH-PAHs 和血浆 BPDE-白蛋白加合物共同作为 PAHs 内暴露标志物,来综合反映研究人群的 PAHs 暴露。10

种 OH-PAHs 包括 1-羟基萘 (1-hydroxynaphthalene, 1-OH-Nap), 2-羟基萘 (2-hydroxynaphthalene, 2-OH-Nap), 2-羟基芴 (2-hydroxyfluorene, 2-OH-Flu), 9-羟基芴 (9-hydroxyfluorene, 9-OH-Flu), 1-羟基菲 (1-hydroxyphenanthrene, 1-OH-Phe), 2-羟基菲 (2-hydroxyphenanthrene, 2-OH-Phe), 3-羟基菲 (3-hydroxyphenanthrene, 3-OH-Phe), 4-羟基菲 (4-hydroxyphenanthrene, 4-OH-Phe), 9-羟基菲 (9-hydroxyphenanthrene, 9-OH-Phe) 和 1-羟基芘 (1-hydroxypyrene, 1-OH-Pyr)。通过识别 OH-PAHs 和 BPDE-白蛋白加合物相关的血浆 miRNA, 探索 PAHs 暴露相关的 miRNA 表达谱特征。进一步寻找 miRNA 的候选靶基因并分析其生物学功能, 从而探讨 PAHs 暴露潜在的健康损害和生物学机制。

## 2 方法

### 2.1 研究人群

本研究是基于 160 名招募于中国十堰当地社区健康成年人的横断面研究。参与者经一夜禁食后到达招募现场, 签署书面知情同意书, 进行身体检查、回答问题并提供血液和尿液样本。经过统一培训的工作人员使用标准的结构化问卷收集人群的社会人口学信息, 如出生地、出生年月、性别等; 环境暴露史, 如烹饪习惯、居住环境和被动吸烟暴露等; 患病史, 如高血压、心肌梗塞、糖尿病、肿瘤等疾病的患病情况; 近两周用药情况, 如口服降糖药、胰岛素、降压药、降血脂药等药物的使用情况以及生活方式, 如当前吸烟状况、当前饮酒状况、饮食摄入(全谷物、肉类、鱼类或海鲜、蔬菜或水果的饮食摄入频率)等。

所有参与者均符合以下纳入标准: (1) 20-70 岁; (2) 未确诊肿瘤或心血管疾病; (3) 近两周内无急性感染, 未服用消炎药; (4) 提供足够的血液和尿液样本用于检测。该研究项目获得了华中科技大学同济医学院伦理委员会的批准。

### 2.2 尿 OH-PAHs 和血浆 BPDE-白蛋白加合物浓度检测

根据之前建立的方法<sup>[35]</sup>, 采用使用气相色谱-质谱联用仪 (Gas chromatography-mass spectrometry, Agilent 公司, 美国) 测定尿中 10 种 OH-PAHs 浓度。使用 QQQ 定量分析软件分析数据。使用全自动生化分析仪 BS-200 (深圳迈瑞生物医疗电子股份有限公司, 中国) 和肌酐测定试剂盒 (深圳迈瑞生物医疗电子股份有限公司, 中国) 通过肌氨酸氧化酶法测定尿样中的肌酐浓度。为获得最终纳入分析的 OH-PAHs 浓度, 首先通过标准曲线计算得到 OH-PAHs 浓度 (单位为  $\mu\text{g/L}$ ), 然后除以相应物质的分子量 (萘类分子量为 144.17, 茚类分子量为 182.22, 菲类分子量为 194.23, 芘类分子量为 218.25), 再除以对应样本的尿肌酐浓度 (单位为  $\text{mmol/L}$ ), 最终的 OH-PAHs 浓度单位为  $\mu\text{mmol/mmol}$  肌酐。

采用血浆 BPDE-白蛋白加合物检测试剂盒以酶联免疫吸附试验 (Enzyme-linked immunosorbent assay, ELISA) 方法检测血浆 BPDE-白蛋白加合物的含量<sup>[36]</sup>, 使用 Biotek Synergy2 多功能酶标仪 (Gene 公司, 美国), 选择 450nm 作为波长, 读取每孔吸光度。根据标准品拟合标准曲线, 计算样本 BPDE-白蛋

白加合物的浓度, 单位为 ng/mL。

## 2.3 血浆 miRNA 定量分析

从血浆提取总 RNA, 进行 RNA 清理, 并检查 RNA 洗脱液的质量。分别使用 miRNeasy Serum/Plasma 试剂盒(Qiagen 公司, 德国)和 QIAseq miRNA Library Kit 试剂盒(Qiagen 公司, 德国)对样品进行 miRNA 提取和文库制备。使用备用样本重建未通过质控的文库。构建的 miRNA 文库在 NextSeq500 (Illumina 公司, 美国)上进行测序。使用 CLC Genomics Workbench (v21.0.5, Qiagen 公司, 德国)按照默认设置的标准方案处理测序原始数据。最终共 1022 个血浆 miRNA 可在 80%以上的样本中检测到 (计数 $\geq 1$ ), 被用于统计分析。

## 2.4 协变量的评估

训练有素的访调员进行标准问卷调查, 收集有关年龄、性别(男性/女性)、当前吸烟状况(是/否)、当前饮酒状况(是/否)、被动吸烟(是/否)、身高、体重和全谷物、肉类、鱼类或海鲜、水果或蔬菜的饮食摄入的信息。身体质量指数 (Body mass index, BMI) 的计算方法是体重除以身高的平方, 单位为  $\text{kg}/\text{m}^2$ 。当前吸烟状况的定义是在过去 6 个月中每天至少吸一支烟。当前饮酒状况的定义是在过去 6 个月里每周至少喝一次酒精饮料。被动吸烟的定义是自我报告有室内环境烟草燃烧后产生的烟雾暴露史。对于每种食物类型, 膳食摄入频率分为 $\geq 5$ 次/周或没有。

## 2.5 探索 miRNA 的组织特异性

运用软件 TissueAtlas 数据库 (<https://ccb-web.cs.uni-saarland.de/tissueatlas2>) 探索 miRNA 的组织特异性。该软件使用来自六个人的 188 个样本来确定 21 个器官的表达数据, 从而得到人类表达图谱。软件计算了每个 miRNA 的组织特异性指数 (Tissue specificity index, TSI)。TSI 是一个定量的、分级的标度, 用于衡量一个 miRNA 在不同器官的表达特异性。数值范围从 0 到 1, 若  $\text{TSI} < 0.15$ , 则认为该 miRNA 是管家 miRNA; 若  $\text{TSI} > 0.85$ , 则认为该 miRNA 具有明显的

组织特异性<sup>[37]</sup>。

## 2.6 寻找 miRNA 的候选靶基因

现有较多预测 miRNA 靶基因的软件，多运用高通量交联和免疫沉淀测序（crosslinking and immunoprecipitation, CLIP）数据直接鉴定和定位靶基因 mRNA 分子上的 miRNA 识别元件（miRNA recognition elements, MRE），预测 miRNA 的靶基因。其中被运用最多的软件是 TargetScan 8.0（<http://www.targetscan.org/>），它的算法为在靶基因的 3' UTR 内搜索与 miRNA 的 5' 端 2-8 个核苷酸完全互补的序列，并评价结合位点的热力学稳定性。TargetScan 软件在 miRNA 靶基因预测方面展示出较低的假阳性率<sup>[38]</sup>。尽管大多数 MRE 在 3' UTR 中被发现，但也有个别研究报告 MRE 位于靶基因的蛋白质编码序列（coding sequence, CDS）中。另一个常用的预测软件 DIANA-microT-CDS（[https://dianalab.e-ce.uth.gr/html/dianauniverse/index.php?r=microT\\_CDS](https://dianalab.e-ce.uth.gr/html/dianauniverse/index.php?r=microT_CDS)）则运用 microT-CDS 算法，以识别 3' UTR 区域和 CDS 区域中的 miRNA 靶点<sup>[39, 40]</sup>。软件 miRDB（<https://mirdb.org/>）也是目前较为常用的 miRNA 靶基因预测软件之一，它利用了独立开发的 miRNA 靶点预测算法 MirTarget v4.0 开展靶基因预测。这一算法结合了当下流行的靶点分析策略，即利用 CLIP 数据进行靶基因预测，和另一种流行的策略，即鉴定 miRNA 过表达引起的下调转录物，来进行 miRNA 靶点预测。比较分析表明，MirTarget 在应用于独立实验数据时，比现有其他算法性能更高<sup>[41, 42]</sup>。此外，2021 年 9 月更新的 miRTarBase 9.0 数据库（[https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTarBase\\_2022/php/index.php](https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTarBase_2022/php/index.php)）汇总了已经被生物学实验验证过的 miRNA 的靶基因。到目前为止，miRTarBase 9.0 包含了超过 13389 篇提供了支持 miRNA-靶基因相互作用的实验证据的文章，这些文章的结果涉及来自 37 个物种的 27172 个靶基因<sup>[43]</sup>。

为了降低靶基因预测假阳性，此次研究选择综合三种软件 miRDB、DIANA-microT-CDS 和 TargetScan 8.0 的靶基因预测结果。当某基因均被这三种预测软件预测为靶基因时，才认定其为 miRNA 的预测靶基因。将预测靶基因与 miRTarBase 中来自实验验证的靶基因取并集，全部结果被视作该 miRNA 的候选靶基因。

## 2.7 探讨候选靶基因的生物学功能

使用在线生物信息学分析软件注释、可视化和集成发现数据库(The Database for Annotation, Visualization and Integrated Discovery, DAVID) (<http://david.abcc.ncifcrf.gov>), 开展京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)通路富集分析和基因本体论(Gene Ontology, GO)富集分析, 进而探讨靶基因的生物学功能<sup>[44, 45]</sup>。

KEGG 是一个从生物学信息中发掘生物系统高级功能和效用的数据库。KEGG 通路是一组人为描述的通路, 代表了 KEGG 对细胞过程、遗传信息处理、生物系统、新陈代谢、环境信息处理、人类疾病和药物开发中的分子相互作用、反应和关系网络的理解。在生物信息学分析过程中, KEGG 通路富集分析常常应用于差异表达基因的功能注释和对相关功能与作用通路的探索<sup>[46, 47]</sup>。 $P$  值越接近零, 与该组基因相关的 KEGG 通路越显著。本研究进行 KEGG 通路富集分析时, 对显著通路的筛选标准为  $P$  值  $< 0.05$ 。

GO 从三个方面描述了目前对生物领域的认知: 生物过程(Biological process, BP)、分子功能(Molecular function, MF)和细胞组分(Cell component, CC)。MF 描述了基因产物进行的分子水平的活动, 如“催化”或“运输”。CC 描述了基因产物发挥功能的相对于细胞结构的位置, 细胞区室(如线粒体)或它们所组成的稳定大分子复合物(如核糖体)。BP 描述了由多种分子活动完成的更大的过程, 如 DNA 修复或信号转导或葡萄糖跨膜转运。请注意, BP 并不等同于通路。给定一组在某些条件下表达改变的基因, GO 富集分析将使用该组基因的注释来发现哪些术语被表达过度或不足。 $P$  值越接近零, 与该组基因相关的特定 GO 术语就越显著<sup>[48, 49]</sup>。本研究进行 GO 富集分析时, 对显著富集 GO 术语的筛选标准为  $P$  值  $< 0.05$ 。本次分别展示候选靶基因最显著富集的前 20 个生物过程、分子功能、细胞组分, 若不足 20 个, 则展示全部结果。

## 2.8 统计分析

研究参与者的基本特征以平均值  $\pm$  标准差(SD)和计数(百分比)表示, OH-PAHs 和 BPDE-白蛋白加合物的浓度用中位数(P25, P75)表示。采用 Spearman

秩相关分析分别探索 10 种 OH-PAHs 和 BPDE-白蛋白加合物浓度之间的关联性以及 miRNA 水平之间的关联性。若  $P$  值  $< 0.05$ ，则认为有显著关联。 $r$  值介于 -1 至 1 之间，为正表示正相关，为负表示负相关，绝对值越接近 1，关联性越强。

以错误发现率 (False discovery rate,  $FDR$ )  $< 0.10$  为条件，识别与 OH-PAHs 和 BPDE-白蛋白加合物相关的单个 miRNA。本次研究使用多元线性回归评估尿 OH-PAHs 和血浆 BPDE-白蛋白加合物浓度与血浆 miRNA 水平之间的关联，并校正混杂因素，包括性别 (男性/女性)、年龄、BMI、当前吸烟状况 (是/否)、当前饮酒状况 (是/否)、被动吸烟 (是/否)，以及全谷物、肉类、鱼类或海鲜、水果或蔬菜的饮食摄入 ( $\geq 5$  次/周或没有)。为维持正态分布，对 OH-PAHs 和 BPDE-白蛋白加合物的含量进行自然对数转换并处理离群值，对 miRNA 的浓度进行  $\log_2$  转换。对转换后的 OH-PAHs、BPDE-白蛋白加合物浓度和 miRNA 水平进一步进行了标准分数转换，以获得标准的  $\beta$  系数。以  $FDR < 0.05$  为条件筛选 miRNA 进行后续分析。

针对混杂因素年龄、性别、BMI、当前吸烟状况、当前饮酒状况、被动吸烟进行分层，开展分层分析和交互作用分析。使用多元线性回归分析，校正饮食摄入，并在适当时校正年龄、性别、BMI、当前吸烟状况、当前饮酒状况、被动吸烟。WHO 规定  $BMI \geq 25.0 \text{ kg/m}^2$  为超重，因此以  $BMI < 25.0 \text{ kg/m}^2$  和  $\geq 25.0 \text{ kg/m}^2$  为标准将研究人群分为两组。

除非另有说明，所有的统计分析都使用 R (v 4.2.2) 进行。

## 3 结果

### 3.1 基线描述

本研究共纳入 160 名成年人, 平均年龄为  $41.98 \pm 9.23$  岁, 性别分布均匀 (49.38%为男性), 分别有 23 人 (14.38%)、50 人 (31.25%) 和 74 人 (46.25%) 存在当前吸烟状况、当前饮酒状况和被动吸烟。平均 BMI 为  $23.25 \text{ kg/m}^2$ 。95.00% 的参与者每周食用水果或蔬菜  $\geq 5$  次, 51.25% 的参与者每周食用肉类  $\geq 5$  次, 而只有 10.00% 和 13.75% 的参与者每周食用鱼类或海鲜和全谷物  $\geq 5$  次。

OH-PAHs 和 BPDE-白蛋白加合物的浓度用中位数(P25, P75)表示, 1-OH-Nap 浓度的中位数为  $0.84 \text{ nmol/mmol}$  肌酐, 2-OH-Nap 浓度的中位数为  $1.10 \text{ nmol/mmol}$  肌酐, 2-OH-Flu 浓度的中位数为  $0.17 \text{ nmol/mmol}$  肌酐, 9-OH-Flu 浓度的中位数为  $0.22 \text{ nmol/mmol}$  肌酐, 1-OH-Phe 浓度的中位数为  $0.12 \text{ nmol/mmol}$  肌酐, 2-OH-Phe 浓度的中位数为  $0.10 \text{ nmol/mmol}$  肌酐, 3-OH-Phe 浓度的中位数为  $0.10 \text{ nmol/mmol}$  肌酐, 4-OH-Phe 浓度的中位数为  $0.04 \text{ nmol/mmol}$  肌酐, 9-OH-Phe 浓度的中位数为  $0.05 \text{ nmol/mmol}$  肌酐, 1-OH-Pyr 浓度的中位数为  $0.06 \text{ nmol/mmol}$  肌酐, BPDE-白蛋白加合物浓度的中位数为  $122.53 \text{ ng/mL}$ 。具体数值见表 3-1。



表 3-1 人群的基本情况与 OH-PAHs 和 BPDE-白蛋白加合物浓度 (n=160)

变量	
男性, %	79 (49.38)
年龄, 岁	41.98 ± 9.23
当前吸烟状况 (% , 是)	23 (14.38)
当前饮酒状况 (% , 是)	50 (31.25)
被动吸烟 (% , 是)	74 (46.25)
身体质量指数, kg/m <sup>2</sup>	23.25 ± 2.54
<b>饮食摄入类别(% , ≥每周 5 次)</b>	
全谷物	22 (13.75)
肉类	82 (51.25)
鱼类或海鲜	16 (10.00)
水果或蔬菜	152 (95.00)
<b>OH-PAHs, μmol/mmol 肌酐</b>	
1-OH-Nap (×10 <sup>-3</sup> )	0.84 (0.56, 1.34)
2-OH-Nap (×10 <sup>-3</sup> )	1.10 (0.58, 2.46)
2-OH-Flu (×10 <sup>-4</sup> )	1.66 (1.19, 2.53)
9-OH-Flu (×10 <sup>-4</sup> )	2.19 (1.59, 2.94)
1-OH-Phe (×10 <sup>-4</sup> )	1.20 (0.85, 1.66)
2-OH-Phe (×10 <sup>-4</sup> )	0.96 (0.61, 1.45)
3-OH-Phe (×10 <sup>-4</sup> )	1.01 (0.71, 1.53)
4-OH-Phe (×10 <sup>-4</sup> )	0.36 (0.23, 0.52)
9-OH-Phe (×10 <sup>-4</sup> )	0.54 (0.31, 0.95)
1-OH-Pyr (×10 <sup>-4</sup> )	0.62 (0.35, 1.14)
BPDE-白蛋白加合物, ng/mL	122.53 (109.96, 134.57)

各变量由均数±标准差、中位数 (P25, P75) 或人数表示。

### 3.2 OH-PAHs 和 BPDE-白蛋白加合物之间的关联性

BPDE-白蛋白加合物与 OH-PAHs 之间的关联性不显著。除 2-OH-Phe 与 2-OH-Nap 关联性不显著之外, 10 种 OH-PAHs 之间都表现出显著的关联性 ( $P$  值  $< 0.05$ ), 相关系数范围为 0.14 至 0.83, 其中极强相关的有 2-OH-Phe 与 4-OH-Phe ( $r=0.83$ ), 2-OH-Phe 与 3-OH-Phe ( $r=0.82$ ), 3-OH-Phe 与 4-OH-Phe ( $r=0.81$ ), 强相关的有 4-OH-Phe 与 9-OH-Phe ( $r=0.73$ ), 2-OH-Flu 与 3-OH-Phe ( $r=0.68$ ), 1-OH-Phe 与 4-OH-Phe ( $r=0.65$ ), 1-OH-Phe 与 3-OH-Phe ( $r=0.65$ ),

2-OH-Flu 与 9-OH-Flu ( $r=0.61$ ), 1-OH-Phe 与 2-OH-Phe ( $r=0.60$ ), 3-OH-Phe 与 1-OH-Pyr ( $r=0.60$ )。详细结果见图 3-1。

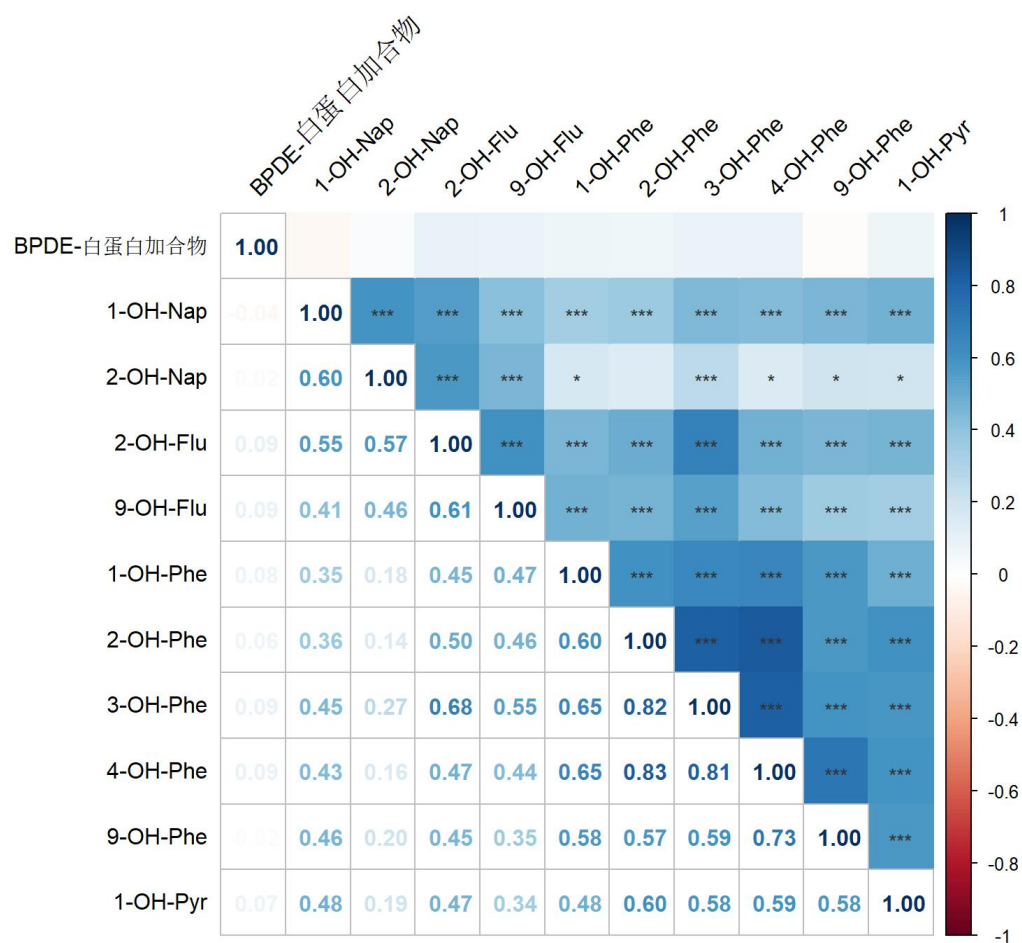


图 3-1 OH-PAHs 和 BPDE-白蛋白加合物关联性热图

注：颜色刻度代表 Spearman 秩相关分析的  $r$  值。右上角的\*代表  $P$  值 $<0.05$ ，\*\*代表  $P$  值 $<0.01$ ，\*\*\*代表  $P$  值 $<0.001$ ；左下角的数值为  $r$  值。

### 3.3 识别与 OH-PAHs 和 BPDE-白蛋白加合物相关的 miRNA

以  $FDR<0.10$  为条件, 识别出 42 个与 PAHs 内暴露水平显著相关的 miRNA, 其中 23 个仅与 2-OH-Flu 相关, 13 个仅与 9-OH-Phe 相关, 5 个与上述两者都相关, 1 个与 1-OH-Phe 相关。以  $FDR<0.05$  为条件筛选 4 个 miRNA 开展后续分析, 分别为与 2-OH-Flu 显著正相关的 miR-125b-2-3p、miR-505-5p 和 miR-483-5p 以及显著负相关的 miR-548e-3p。

表 3-2 与 OH-PAHs 相关的 miRNA ( $FDR < 0.10$ )

OH-PAHs <sup>a</sup>	miRNA <sup>b</sup>	$\beta_{\text{std}}$ (95% CI)	$P$ 值	$FDR$ <sup>c</sup>
2-OH-Flu	miR-10399-5p	-0.29 (-0.48, -0.11)	$2.60 \times 10^{-3}$	0.10
	miR-103a-3p	-0.30 (-0.49, -0.12)	$2.00 \times 10^{-3}$	0.09
	miR-122-5p	0.32 (0.13, 0.51)	$1.03 \times 10^{-3}$	0.06
	<b>miR-125b-2-3p <sup>d</sup></b>	<b>0.38 (0.19, 0.56)</b>	<b><math>8.08 \times 10^{-5}</math></b>	<b>0.04</b>
	miR-17-3p	-0.32 (-0.51, -0.14)	$7.77 \times 10^{-4}$	0.06
	miR-190b-5p	0.35 (0.16, 0.54)	$3.66 \times 10^{-4}$	0.06
	miR-193a-5p	0.35 (0.16, 0.53)	$3.20 \times 10^{-4}$	0.06
	miR-21-3p	-0.30 (-0.49, -0.12)	$1.66 \times 10^{-3}$	0.08
	miR-21-5p	-0.33 (-0.51, -0.14)	$8.30 \times 10^{-4}$	0.06
	miR-219a-1-3p	0.29 (0.11, 0.48)	$2.36 \times 10^{-3}$	0.10
	miR-29b-1-5p	0.33 (0.14, 0.51)	$7.41 \times 10^{-4}$	0.06
	miR-3180	0.30 (0.12, 0.49)	$1.74 \times 10^{-3}$	0.08
	miR-320c	0.34 (0.15, 0.52)	$5.26 \times 10^{-4}$	0.06
	miR-320d	0.34 (0.15, 0.52)	$6.26 \times 10^{-4}$	0.06
	miR-340-5p	-0.29 (-0.48, -0.11)	$2.44 \times 10^{-3}$	0.10
	miR-374a-3p	-0.30 (-0.49, -0.12)	$1.52 \times 10^{-3}$	0.08
	miR-375-3p	0.30 (0.12, 0.49)	$1.56 \times 10^{-3}$	0.08
	miR-4516	0.34 (0.15, 0.52)	$5.73 \times 10^{-4}$	0.06
	miR-4667-3p	0.29 (0.10, 0.47)	$2.68 \times 10^{-3}$	0.10
	miR-4732-5p	0.33 (0.15, 0.51)	$4.91 \times 10^{-4}$	0.06
	<b>miR-483-5p <sup>d</sup></b>	<b>0.34 (0.18, 0.54)</b>	<b><math>1.44 \times 10^{-4}</math></b>	<b>0.05</b>
	<b>miR-505-5p <sup>d</sup></b>	<b>0.40 (0.22, 0.57)</b>	<b><math>2.72 \times 10^{-5}</math></b>	<b>0.03</b>
	miR-548bc	-0.31 (-0.49, -0.13)	$9.72 \times 10^{-4}$	0.06
	<b>miR-548e-3p <sup>d</sup></b>	<b>-0.35 (-0.53, -0.17)</b>	<b><math>1.95 \times 10^{-4}</math></b>	<b>0.05</b>
	miR-651-5p	-0.32 (-0.51, -0.13)	$1.06 \times 10^{-3}$	0.06
	miR-766-5p	0.31 (0.12, 0.50)	$1.55 \times 10^{-3}$	0.08
	miR-9-5p	-0.31 (-0.50, -0.12)	$1.91 \times 10^{-3}$	0.09
	let-7d-5p	0.34 (0.16, 0.53)	$4.02 \times 10^{-4}$	0.06
1-OH-Phe				
	miR-380-3p	0.33 (0.17, 0.49)	$6.64 \times 10^{-5}$	0.07
9-OH-Phe				
	miR-10a-3p	0.27 (0.11, 0.44)	$1.47 \times 10^{-3}$	0.08
	miR-125b-2-3p	0.30 (0.14, 0.46)	$4.18 \times 10^{-4}$	0.06
	miR-153-3p	-0.29 (-0.46, -0.13)	$6.20 \times 10^{-4}$	0.08
	miR-15a-3p	-0.28 (-0.44, -0.12)	$9.45 \times 10^{-4}$	0.08
	miR-16-1-3p	-0.29 (-0.45, -0.12)	$7.64 \times 10^{-4}$	0.08
	miR-1914-3p	0.31 (0.14, 0.47)	$3.54 \times 10^{-4}$	0.06
	miR-193a-5p	0.28 (0.12, 0.45)	$9.77 \times 10^{-4}$	0.08
	miR-19a-3p	-0.27 (-0.44, -0.11)	$1.34 \times 10^{-3}$	0.08
	miR-320b	0.33 (0.16, 0.50)	$1.41 \times 10^{-4}$	0.05

续表 3-2 与 OH-PAHs 相关的 miRNA ( $FDR < 0.10$ )

OH-PAHs	miRNA <sup>a</sup>	$\beta_{\text{std}}$ (95% CI)	$P$ 值	$FDR$ <sup>b</sup>
	miR-320c	0.33 (0.16, 0.49)	$1.52 \times 10^{-4}$	0.05
	miR-320d	0.32 (0.15, 0.49)	$2.47 \times 10^{-4}$	0.05
	miR-33a-5p	-0.29 (-0.45, -0.12)	$8.28 \times 10^{-4}$	0.08
	miR-3613-5p	-0.27 (-0.43, -0.11)	$1.25 \times 10^{-3}$	0.08
	miR-483-5p	0.31 (0.15, 0.47)	$2.02 \times 10^{-4}$	0.05
	miR-503-3p	0.32 (0.15, 0.48)	$2.29 \times 10^{-4}$	0.05
	miR-545-3p	-0.28 (-0.44, -0.11)	$1.16 \times 10^{-3}$	0.08
	miR-590-5p	-0.28 (-0.44, -0.11)	$1.32 \times 10^{-3}$	0.08
	miR-6804-3p	-0.27 (-0.44, -0.11)	$1.18 \times 10^{-3}$	0.08

<sup>a</sup> 经自然对数转换。

<sup>b</sup> 经  $\log_2$  转换。

<sup>c</sup> 多元线性回归，校正了年龄、性别、BMI、当前吸烟状况、当前饮酒状况、被动吸烟、饮食摄入；利用 Benjamini & Hochberg 法控制错误发现率。

<sup>d</sup> 以  $FDR < 0.05$  为条件筛选的相关 miRNA。

### 3.4 针对混杂因素的分层分析和交互作用分析

在年龄  $< 45.0$  岁的人群中，尿 2-OH-Flu 与 miR-125b-2-3p ( $\beta_{\text{std}}=0.50$ ,  $P=2.20 \times 10^{-4}$ )、miR-483-5p ( $\beta_{\text{std}}=0.46$ ,  $P=7.14 \times 10^{-4}$ )、miR-548e-3p ( $\beta_{\text{std}}=-0.51$ ,  $P=3.39 \times 10^{-4}$ ) 之间的关联更加显著；在否认存在被动吸烟的人群中，尿 2-OH-Flu 与 miR-548e-3p ( $\beta_{\text{std}}=-0.51$ ,  $P=2.94 \times 10^{-4}$ ) 和 miR-483-5p ( $\beta_{\text{std}}=0.42$ ,  $P=1.79 \times 10^{-3}$ ) 之间的关联更加显著；在 BMI  $< 25.0 \text{ kg/m}^2$  的人群中，尿 2-OH-Flu 与 miR-483-5p ( $\beta_{\text{std}}=0.52$ ,  $P=3.70 \times 10^{-5}$ ) 之间的关联更加显著。

交互作用分析的结果显示，BMI 和尿 2-OH-Flu 与 miR-505-5p 和 miR-483-5p 之间存在交互作用 ( $P_{\text{交互}}$  分别为 0.03 和 0.01) (表 3-3)。

表 3-3 尿 2-OH-Flu 相关 miRNA 与混杂因素的亚组分析和交互作用分析

	miR-505-5p <sup>a</sup>			miR-125b-2-3p <sup>a</sup>			miR-483-5p <sup>a</sup>			miR-548e-3p <sup>a</sup>		
	$\beta_{\text{std}}$ (95% CI)	<i>P</i> 值	<i>P</i> 交互	$\beta_{\text{std}}$ (95% CI)	<i>P</i> 值	<i>P</i> 交互	$\beta_{\text{std}}$ (95% CI)	<i>P</i> 值	<i>P</i> 交互	$\beta_{\text{std}}$ (95% CI)	<i>P</i> 值	<i>P</i> 交互
性别			0.95			0.98			0.46			0.93
男性 (n=79)	0.45(0.21, 0.70)	5.70×10 <sup>-4</sup>		0.38(0.12, 0.63)	4.68×10 <sup>-3</sup>		0.32(0.07, 0.56)	1.37×10 <sup>-2</sup>		-0.40(-0.65, -0.17)	1.52×10 <sup>-3</sup>	
女性 (n=81)	0.44(0.17, 0.71)	2.30×10 <sup>-3</sup>		0.35(0.06, 0.64)	1.96×10 <sup>-2</sup>		0.44(0.15, 0.73)	3.78×10 <sup>-3</sup>		-0.36(-0.65, -0.08)	1.57×10 <sup>-2</sup>	
年龄, 岁			0.76			0.75			0.39			0.54
<45.0 (n=81)	0.46(0.22, 0.69)	3.31×10 <sup>-4</sup>		0.50(0.25, 0.76)	2.20×10 <sup>-4</sup>		0.46(0.20, 0.71)	7.14×10 <sup>-4</sup>		-0.51(-0.77, -0.24)	3.39×10 <sup>-4</sup>	
≥45.0 (n=72)	0.37(0.08, 0.66)	1.53×10 <sup>-2</sup>		0.16(-0.10, 0.43)	2.27×10 <sup>-1</sup>		0.20(-0.08, 0.47)	1.73×10 <sup>-1</sup>		-0.13(-0.38, 0.12)	3.27×10 <sup>-1</sup>	
BMI, kg/m <sup>2</sup>			0.03			0.21			0.01			0.57
<25.0 (n=120)	0.44(0.21, 0.67)	3.25×10 <sup>-4</sup>		0.43(0.18, 0.68)	9.56×10 <sup>-4</sup>		0.52(0.28, 0.75)	3.70×10 <sup>-5</sup>		-0.39(-0.63, -0.16)	1.28×10 <sup>-3</sup>	
≥25.0 (n=40)	0.44(0.13, 0.75)	9.60×10 <sup>-3</sup>		0.38(0.10, 0.65)	1.16×10 <sup>-2</sup>		0.18(-0.12, 0.47)	2.44×10 <sup>-1</sup>		-0.49(-0.81, -0.16)	6.90×10 <sup>-3</sup>	
当前吸烟状况			0.23			0.80			0.79			0.89
是 (n=23)	0.64(0.25, 1.02)	5.89×10 <sup>-3</sup>		0.30(0.05, 0.56)	3.57×10 <sup>-2</sup>		0.48(0.11, 0.85)	2.47×10 <sup>-2</sup>		-0.43(-0.75, -0.10)	2.13×10 <sup>-2</sup>	
否 (n=127)	0.35(0.15, 0.56)	9.87×10 <sup>-4</sup>		0.36(0.14, 0.58)	1.53×10 <sup>-3</sup>		0.35(0.14, 0.56)	1.19×10 <sup>-3</sup>		-0.36(-0.58, -0.15)	9.40×10 <sup>-4</sup>	
当前饮酒状况			0.05			0.80			0.44			0.79
是 (n=50)	0.62(0.31, 0.94)	3.53×10 <sup>-4</sup>		0.45(0.15, 0.74)	4.98×10 <sup>-3</sup>		0.40(0.09, 0.71)	1.58×10 <sup>-2</sup>		-0.30(-0.56, -0.04)	2.77×10 <sup>-2</sup>	
否 (n=110)	0.31(0.08, 0.53)	1.02×10 <sup>-2</sup>		0.32(0.07, 0.56)	1.26×10 <sup>-2</sup>		0.36(0.13, 0.60)	3.06×10 <sup>-3</sup>		-0.38(-0.63, -0.14)	2.50×10 <sup>-3</sup>	
被动吸烟			0.58			0.58			0.55			0.20
是 (n=74)	0.41(0.12, 0.70)	6.76×10 <sup>-3</sup>		0.33(0.04, 0.63)	2.86×10 <sup>-2</sup>		0.29(0.00, 0.58)	5.62×10 <sup>-2</sup>		-0.19(-0.46, -0.07)	1.54×10 <sup>-1</sup>	
否 (n=86)	0.39(0.15, 0.64)	2.08×10 <sup>-3</sup>		0.41(0.16, 0.66)	1.65×10 <sup>-3</sup>		0.42(0.16, 0.67)	1.79×10 <sup>-3</sup>		-0.51(-0.77, -0.25)	2.94×10 <sup>-4</sup>	

<sup>a</sup> 经 log2 转换。

多元线性回归分析, 校正饮食摄入, 在适当时校正年龄、性别、BMI、当前吸烟状况、当前饮酒状况、被动吸烟。

### 3.5 探索 miRNA 的关联性、组织特异性和候选靶基因

4 个 miRNA 表达水平均表现出显著的关联 ( $P$  值  $< 0.05$ )。相关系数范围为 -0.70 至 0.65，其中 miR-125b-2-3p 与 miR-483-5p ( $r = 0.65$ )，miR-505-5p 与 miR-548e-3p ( $r = -0.70$ ) 表现出强相关 (图 3-2)。

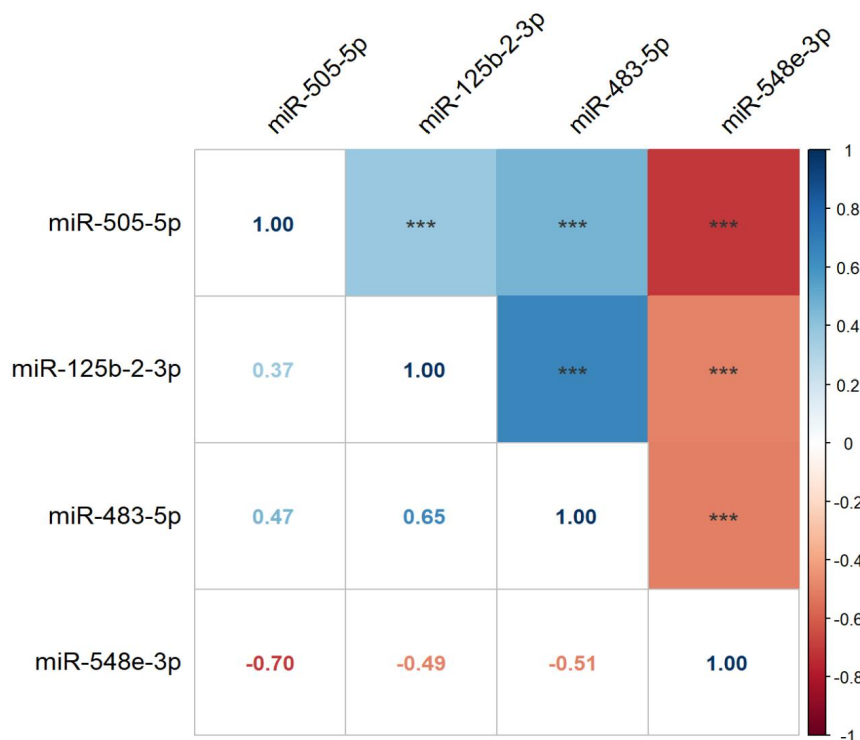


图 3-2 相关血浆 miRNA 表达水平关联性热图

注：颜色刻度代表 Spearman 秩相关分析的  $r$  值。右上角的\*代表  $P$  值  $< 0.05$ ，\*\*代表  $P$  值  $< 0.01$ ，\*\*\*代表  $P$  值  $< 0.001$ ；左下角的数值为  $r$  值。

miR-505-5p、miR125b-2-3p、miR483-5p 和 miR-548e-3p 的 TSI 分别为 0.33、0.30、0.37 和 0.59，提示这些 miRNA 没有明显组织特异性。

通过 Targetscan、miRDB、DIANA-microT-CDS 和 miRTarBase 软件，共得到 379 个 miR-505-5p 的候选靶基因，534 个 miR-125b-2-3p 的候选靶基因，168 个 miR-483-5p 的候选靶基因，279 个 miR-548e-3p 的候选靶基因。

### 3.6 miRNA 候选靶基因的 KEGG 通路富集分析

通过 DAVID 数据库对候选靶基因进行 KEGG 通路富集分析，筛选标准为  $P$

值 $<0.05$ 。miR-505-5p 的候选靶基因表达富集在 3 条通路上, 分别为 HIF-1 信号传导通路、卵母细胞减数分裂和 Wnt 信号传导通路; miR-125b-2-3p 的候选靶基因表达没有富集通路; miR-483-5p 的候选靶基因表达富集在 7 条通路上, 分别为 MAPK 信号通路、丙型肝炎、单纯疱疹病毒 1 型感染、癌症的中心碳代谢、癌症中的胆碱代谢、氨基酸的生物合成和甘氨酸、丝氨酸和苏氨酸代谢; miR-548e-3p 的候选靶基因表达富集在 3 条通路上, 分别为胰高血糖素信号传导通路、心肌细胞肾上腺素能信号传导、流体剪切应力和动脉粥样硬化。具体的通路信息见图 3-3, 表 3-4。

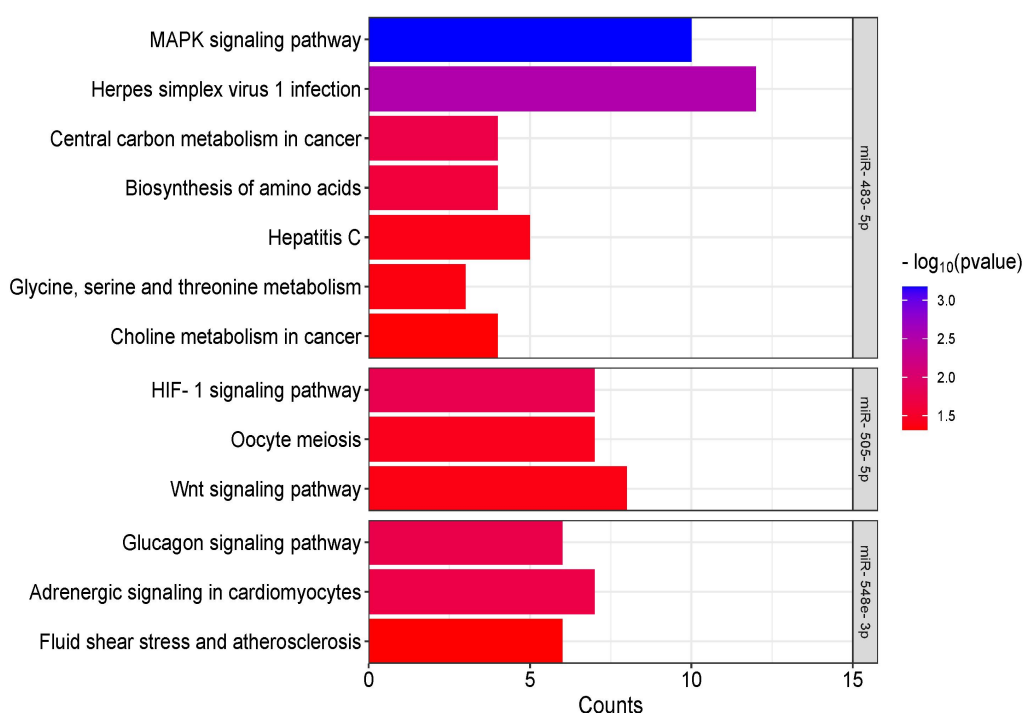


图 3-3 KEGG 通路富集分析直方图

注: 横轴为富集在该通路的输入基因的数量, 纵轴为通路名称。颜色刻度代表 $-\log_{10}(P)$ 值。

表 3-4 KEGG 通路富集分析结果 ( $P$  值  $< 0.05$ )

miRNA 种类	KEGG 通路英文名称	KEGG 通路中文名称	基因数 <sup>a</sup>	$P$ 值
miR-505-5p	HIF-1 signaling pathway	HIF-1 信号传导通路	7	$1.70 \times 10^{-2}$
miR-505-5p	Oocyte meiosis	卵母细胞减数分裂	7	$3.75 \times 10^{-2}$
miR-505-5p	Wnt signaling pathway	Wnt 信号传导通路	8	$4.19 \times 10^{-2}$
miR-483-5p	MAPK signaling pathway	MAPK 信号通路	10	$6.62 \times 10^{-4}$
miR-483-5p	Herpes simplex virus 1 infection	单纯疱疹病毒 1 型感染	12	$2.99 \times 10^{-3}$
miR-483-5p	Central carbon metabolism in cancer	癌症的中心碳代谢	4	$2.01 \times 10^{-2}$
miR-483-5p	Biosynthesis of amino acids	氨基酸的生物合成	4	$2.41 \times 10^{-2}$
miR-483-5p	Hepatitis C	丙型肝炎	5	$4.09 \times 10^{-2}$
miR-483-5p	Glycine, serine and threonine metabolism	甘氨酸、丝氨酸和苏氨酸代谢	3	$4.31 \times 10^{-2}$
miR-483-5p	Choline metabolism in cancer	癌症中的胆碱代谢	4	$4.74 \times 10^{-2}$
miR-548e-3p	Glucagon signaling pathway	胰高血糖素信号传导通路	6	$1.82 \times 10^{-2}$
miR-548e-3p	Adrenergic signaling in cardiomyocytes	心肌细胞肾上腺素能信号传导	7	$2.01 \times 10^{-2}$
miR-548e-3p	Fluid shear stress and atherosclerosis	流体剪切应力和动脉粥样硬化	6	$4.83 \times 10^{-2}$

<sup>a</sup> 富集在该通路的输入基因的数量。



### 3.7 miRNA 候选靶基因的 GO 富集分析

#### 3.7.1 miR-505-5p 候选靶基因的 GO 富集分析结果

对 miR-505-5p 的 379 个候选靶基因进行 GO 富集分析, 结果显示富集最显著的 5 个生物过程为突触组装、蛋白质结合的正向调控、蛋白酶体介导的泛素依赖性蛋白质分解代谢过程、轴突引导和 DNA 诱导的转录负调控, 其他显著富集的生物过程还包括血管生成、磷脂转运、中枢神经系统发育、成人行为、先天免疫反应的负调控等。最显著的 5 个分子功能为蛋白结合、血管内皮生长因子激活受体活动、DNA 结合、磷脂酰肌醇转运蛋白活动和 $\beta$ -1 连环蛋白结合。最显著的 5 个细胞组分为核质体、细胞质、细胞核、染色质和细胞内液。具体结果见表 3-5 至表 3-7。

表 3-5 miR-505-5p 候选靶基因最富集的 20 个生物过程

GO 术语编号	GO 术语英文名称	GO 术语中文名称	基因数 <sup>a</sup>	P 值
GO:0007416	synapse assembly	突触组装	8	$2.55 \times 10^{-4}$
GO:0032092	positive regulation of protein binding	蛋白质结合的正向调控	8	$4.21 \times 10^{-4}$
GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	蛋白酶体介导的泛素依赖性蛋白质分解代谢过程	12	$1.13 \times 10^{-3}$
GO:0007411	axon guidance	轴突引导	11	$2.00 \times 10^{-3}$
GO:0045892	negative regulation of transcription, DNA-templated	DNA 诱导的转录负调控	21	$3.94 \times 10^{-3}$
GO:0006357	regulation of transcription from RNA polymerase II promoter	RNA 聚合酶 II 启动子对转录的调控	46	$4.60 \times 10^{-3}$
GO:0015914	phospholipid transport	磷脂转运	5	$4.84 \times 10^{-3}$
GO:0007417	central nervous system development	中枢神经系统发育	9	$4.86 \times 10^{-3}$
GO:0045824	negative regulation of innate immune response	先天免疫反应的负调控	4	$7.45 \times 10^{-3}$
GO:0045747	positive regulation of Notch signaling pathway	Notch 信号通路的正调控	5	$8.07 \times 10^{-3}$
GO:0045893	positive regulation of transcription, DNA-templated	DNA 诱导的转录正调控	23	$9.54 \times 10^{-3}$
GO:0006635	fatty acid beta-oxidation	脂肪酸 $\beta$ -氧化	5	$1.16 \times 10^{-2}$
GO:0001525	angiogenesis	血管生成	11	$1.27 \times 10^{-2}$
GO:0046777	protein autophosphorylation	蛋白质自磷酸化	9	$1.44 \times 10^{-2}$
GO:0030534	adult behavior	成人行为	4	$1.70 \times 10^{-2}$
GO:0061024	membrane organization	膜的组织	4	$1.70 \times 10^{-2}$
GO:0036258	multivesicular body assembly	多囊体组装	4	$1.70 \times 10^{-2}$
GO:0010976	positive regulation of neuron projection development	神经元投射发育的正调控	7	$1.77 \times 10^{-2}$
GO:0033327	Leydig cell differentiation	Leydig 细胞分化	3	$1.83 \times 10^{-2}$
GO:0006486	protein glycosylation	蛋白质糖基化	7	$1.84 \times 10^{-2}$

<sup>a</sup> 富集在该术语的输入基因的数量。

表 3-6 miR-505-5p 候选靶基因富集的 19 个分子功能

GO 术语编号	GO 术语英文名称	GO 术语中文名称	基因数 <sup>a</sup>	P 值
GO:0005515	protein binding	蛋白结合	277	$1.95 \times 10^{-6}$
GO:0005021	vascular endothelial growth factor-activated receptor activity	血管内皮生长因子激活受体活动	4	$3.96 \times 10^{-4}$
GO:0003677	DNA binding	DNA 结合	43	$5.67 \times 10^{-4}$
GO:0008526	phosphatidylinositol transporter activity	磷脂酰肌醇转运蛋白活动	4	$9.59 \times 10^{-4}$
GO:0008013	beta-catenin binding	$\beta$ -1 连环蛋白结合	8	$1.70 \times 10^{-3}$
GO:0003682	chromatin binding	染色质结合	19	$3.22 \times 10^{-3}$
GO:0003723	RNA binding	RNA 结合	43	$3.83 \times 10^{-3}$
GO:0003714	transcription corepressor activity	转录辅阻遏物活动	11	$4.39 \times 10^{-3}$
GO:0003729	mRNA binding	mRNA 结合	12	$6.09 \times 10^{-3}$
GO:0008022	protein C-terminus binding	蛋白 C 端结合	10	$1.01 \times 10^{-2}$
GO:0004865	protein serine/threonine phosphatase inhibitor activity	蛋白丝氨酸/苏氨酸磷酸酶抑制剂活动	3	$2.38 \times 10^{-2}$
GO:0017154	semaphorin receptor activity	信号素受体活动	3	$2.74 \times 10^{-2}$
GO:0003713	transcription coactivator activity	转录辅助激活蛋白活动	11	$2.82 \times 10^{-2}$
GO:0048185	activin binding	激活素结合	3	$3.12 \times 10^{-2}$
GO:0008134	transcription factor binding	转录因子结合	5	$3.19 \times 10^{-2}$
GO:0005093	Rab GDP-dissociation inhibitor activity	Rab GDP 解离抑制因子活动	2	$3.71 \times 10^{-2}$
GO:0019904	protein domain specific binding	蛋白域特异性结合	10	$4.35 \times 10^{-2}$
GO:0008201	heparin binding	肝素结合	8	$4.69 \times 10^{-2}$
GO:0005524	ATP binding	ATP 结合	39	$4.84 \times 10^{-2}$

<sup>a</sup> 富集在该术语的输入基因的数量。

表 3-7 miR-505-5p 候选靶基因最富集的 20 个细胞组分

GO 术语编号	GO 术语英文名称	GO 术语中文名称	基因数 <sup>a</sup>	P 值
GO:0005654	nucleoplasm	核质体	116	$2.56 \times 10^{-11}$
GO:0005737	cytoplasm	细胞质	138	$4.82 \times 10^{-6}$
GO:0005634	nucleus	细胞核	119	$3.11 \times 10^{-5}$
GO:0000785	chromatin	染色质	109	$1.76 \times 10^{-4}$
GO:0005829	cytosol	细胞内液	89	$4.80 \times 10^{-4}$
GO:0042995	cell projection	细胞突起	52	$4.56 \times 10^{-3}$
GO:0098978	glutamatergic synapse	谷氨酸能突触	30	$5.60 \times 10^{-3}$
GO:0014069	postsynaptic density	突触后致密区	46	$5.80 \times 10^{-3}$
GO:0015630	microtubule cytoskeleton	微管细胞骨架	64	$6.17 \times 10^{-3}$
GO:0016328	lateral plasma membrane	侧质膜	16	$8.65 \times 10^{-3}$
GO:0032588	trans-Golgi network membrane	高尔基体反面网状膜	14	$9.42 \times 10^{-3}$
GO:0043229	intracellular organelle	细胞内细胞器	14	$1.01 \times 10^{-2}$
GO:0031594	neuromuscular junction	神经肌肉接头	13	$1.26 \times 10^{-2}$
GO:0030496	midbody	中间体	14	$1.54 \times 10^{-2}$
GO:0045171	intercellular bridge	细胞间桥	29	$2.27 \times 10^{-2}$
GO:0016020	membrane	膜	10	$2.97 \times 10^{-2}$
GO:0098793	presynapse	突触前	7	$3.03 \times 10^{-2}$
GO:0031527	filopodium membrane	丝状体膜	11	$3.80 \times 10^{-2}$
GO:0005819	spindle	纺锤体	6	$4.28 \times 10^{-2}$
GO:0000139	Golgi membrane	高尔基体膜	5	$4.36 \times 10^{-2}$

<sup>a</sup> 富集在该术语的输入基因的数量。

### 3.7.2 miR-125b-2-3p 候选靶基因的 GO 富集分析结果

对 miR-125b-2-3p 的 534 个候选靶基因进行 GO 富集分析, 结果显示富集最显著的 5 个生物过程为对 RNA 聚合酶 II 启动子转录的正调控、染色质编排、RNA 聚合酶 II 启动子转录的负调控、核转录 mRNA 分解代谢过程, 去烯基化依赖性衰变、蛋白质分泌的正调控和糖蛋白运输, 其他显著的生物过程还包括细胞对葡萄糖缺乏的反应、干细胞种群维护的调节、血管内皮生长因子受体信号传导途径调控、记忆等。最显著的 5 个分子功能为蛋白结合、金属离子结合、RNA 结合、视黄酸受体结合和转录辅助因子活动。最显著的 5 个细胞组分为核质体、细胞内液、轴突、细胞核和泛素连接酶复合物。具体结果见表 3-8 至表 3-10。

表 3-8 miR-125b-2-3p 候选靶基因最富集的 20 个生物过程

GO 术语编号	GO 术语英文名称	GO 术语中文名称	基因数 <sup>a</sup>	P 值
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	对 RNA 聚合酶 II 启动子转录的正调控	55	$1.76 \times 10^{-5}$
GO:0006325	chromatin organization	染色质编排	18	$5.86 \times 10^{-4}$
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	RNA 聚合酶 II 启动子转录的负调控	42	$8.13 \times 10^{-4}$
GO:0000288	nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay	核转录 mRNA 分解代谢过程, 去烯基化依赖性衰变	4	$1.14 \times 10^{-3}$
GO:0050714	positive regulation of protein secretion	蛋白质分泌的正调控	7	$1.42 \times 10^{-3}$
GO:0034436	glycoprotein transport	糖蛋白运输	3	$1.81 \times 10^{-3}$
GO:0042149	cellular response to glucose starvation	细胞对葡萄糖缺乏的反应	7	$3.07 \times 10^{-3}$
GO:0030947	regulation of vascular endothelial growth factor receptor signaling pathway	血管内皮生长因子受体信号传导途径调控	3	$3.56 \times 10^{-3}$
GO:0045056	transcytosis	胞吞作用	4	$3.60 \times 10^{-3}$
GO:0060213	positive regulation of nuclear-transcribed mRNA poly(A) tail shortening	核转录 mRNA poly(A) 尾缩短的正向调节	4	$3.60 \times 10^{-3}$
GO:0016477	cell migration	细胞迁移	16	$3.75 \times 10^{-3}$
GO:0045893	positive regulation of transcription, DNA-templated	DNA 诱导的转录正调控	31	$4.25 \times 10^{-3}$
GO:2000036	regulation of stem cell population maintenance	干细胞种群维护的调节	4	$4.50 \times 10^{-3}$
GO:0006511	ubiquitin-dependent protein catabolic process	泛素依赖的蛋白质分解过程	16	$4.88 \times 10^{-3}$
GO:0048009	insulin-like growth factor receptor signaling pathway	胰岛素样生长因子受体信号传导通路	5	$5.42 \times 10^{-3}$
GO:1900153	positive regulation of nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay	核转录 mRNA 分解代谢过程的正向调控, 去烯基化依赖性衰变	4	$5.52 \times 10^{-3}$
GO:0000381	regulation of alternative mRNA splicing, via spliceosome	通过剪接体调节替代性 mRNA 剪接	7	$5.83 \times 10^{-3}$
GO:0007613	memory	记忆	8	$5.86 \times 10^{-3}$
GO:0006357	regulation of transcription from RNA polymerase II promoter	调节 RNA 聚合酶 II 启动子的转录	60	$6.06 \times 10^{-3}$
GO:0080090	regulation of primary metabolic process	初级代谢过程的调节	4	$6.67 \times 10^{-3}$

<sup>a</sup> 富集在该术语的输入基因的数量。

表 3-9 miR-125b-2-3p 候选靶基因最富集的 20 个分子功能

GO 术语编号	GO 术语英文名称	GO 术语中文名称	基因数 <sup>a</sup>	P 值
GO:0005515	protein binding	蛋白结合	394	$2.13 \times 10^{-10}$
GO:0046872	metal ion binding	金属离子结合	95	$1.57 \times 10^{-3}$
GO:0003723	RNA binding	RNA 结合	58	$1.71 \times 10^{-3}$
GO:0042974	retinoic acid receptor binding	视黄酸受体结合	4	$4.19 \times 10^{-3}$
GO:0003712	transcription cofactor activity	转录辅助因子活动	11	$4.22 \times 10^{-3}$
GO:0005543	phospholipid binding	磷脂结合	10	$4.31 \times 10^{-3}$
GO:0003676	nucleic acid binding	核酸结合	18	$8.21 \times 10^{-3}$
GO:0000976	transcription regulatory region sequence-specific DNA binding	转录调控区序列特异性 DNA 结合	14	$8.84 \times 10^{-3}$
GO:0004407	histone deacetylase activity	组蛋白去乙酰化酶活动	4	$1.09 \times 10^{-2}$
GO:0004712	protein serine/threonine/tyrosine kinase activity	蛋白丝氨酸/苏氨酸/酪氨酸激酶活动	21	$1.22 \times 10^{-2}$
GO:0004843	thiol-dependent ubiquitin-specific protease activity	硫醇依赖性泛素特异性蛋白酶活动	9	$1.30 \times 10^{-2}$
GO:0061631	ubiquitin conjugating enzyme activity	泛素结合酶活动	5	$1.41 \times 10^{-2}$
GO:0003714	transcription corepressor activity	转录辅阻遏物活动	12	$1.58 \times 10^{-2}$
GO:0019208	phosphatase regulator activity	磷酸酶调控因子活动	3	$1.72 \times 10^{-2}$
GO:0003677	DNA binding	DNA 结合	48	$2.08 \times 10^{-2}$
GO:0003729	mRNA binding	mRNA 结合	13	$2.58 \times 10^{-2}$
GO:0000149	SNARE binding	SNARE 蛋白结合	6	$2.76 \times 10^{-2}$
GO:0044325	ion channel binding	离子通道结合	9	$2.78 \times 10^{-2}$
GO:0035925	mRNA 3'-UTR AU-rich region binding	mRNA 3'-UTR AU-rich 区域结合	4	$2.96 \times 10^{-2}$
GO:0019911	structural constituent of myelin sheath	髓鞘结构组成	3	$3.22 \times 10^{-2}$

<sup>a</sup> 富集在该术语的输入基因的数量。

表 3-10 miR-125b-2-3p 候选靶基因最富集的 20 个细胞组分

GO 术语编号	GO 术语英文名称	GO 术语中文名称	基因数 <sup>a</sup>	P 值
GO:0005654	nucleoplasm	核质体	153	$2.45 \times 10^{-9}$
GO:0005829	cytosol	细胞内液	186	$7.45 \times 10^{-7}$
GO:0030424	axon	轴突	25	$7.81 \times 10^{-6}$
GO:0005634	nucleus	细胞核	190	$3.17 \times 10^{-5}$
GO:0000151	ubiquitin ligase complex	泛素连接酶复合物	11	$4.44 \times 10^{-4}$
GO:0098978	glutamatergic synapse	谷氨酸突触	22	$1.37 \times 10^{-3}$
GO:0000118	histone deacetylase complex	组蛋白去乙酰化酶复合体	6	$4.34 \times 10^{-3}$
GO:0043005	neuron projection	投射神经元	19	$4.35 \times 10^{-3}$
GO:0043025	neuronal cell body	神经元细胞体	19	$4.49 \times 10^{-3}$
GO:0031092	platelet alpha granule membrane	血小板 $\alpha$ 颗粒膜	4	$7.76 \times 10^{-3}$
GO:0030672	synaptic vesicle membrane	突触囊泡膜	9	$1.16 \times 10^{-2}$
GO:0030425	dendrite	树突	20	$1.40 \times 10^{-2}$
GO:0005681	spliceosomal complex	剪接体复合体	9	$1.45 \times 10^{-2}$
GO:0005783	endoplasmic reticulum	内质网	41	$1.63 \times 10^{-2}$
GO:0098982	GABA-ergic synapse	GABA-能突触	7	$1.83 \times 10^{-2}$
GO:0036464	cytoplasmic ribonucleoprotein granule	细胞质核糖核酸蛋白颗粒	7	$1.93 \times 10^{-2}$
GO:0071013	catalytic step 2 spliceosome	催化步骤 2 剪接体	7	$2.03 \times 10^{-2}$
GO:0016020	membrane	膜	111	$2.14 \times 10^{-2}$
GO:0070382	exocytic vesicle	胞吐空泡	4	$2.79 \times 10^{-2}$
GO:0001650	fibrillar center	纤维中心	9	$3.22 \times 10^{-2}$

<sup>a</sup> 富集在该术语的输入基因的数量。



### 3.7.3 miR-483-5p 候选靶基因的 GO 富集分析结果

对 miR-483-5p 的 168 个候选靶基因进行 GO 富集分析, 结果显示富集最显著的 5 个生物过程为 RNA 聚合酶 II 启动子转录的调控、肺部形态发生、DNA 诱导的转录调控、RNA 聚合酶 II 启动子转录的正调控和 DNA 诱导的转录, 其他显著的生物过程还包括细胞对活性氧的反应、胸腺发育、面部发育、血管直径的调节、甲状腺发育、参与心脏发育的心脏神经嵴细胞发育、软骨内骨化、子宫内的胚胎发育等。最显著的 5 个分子功能为 RNA 聚合酶 II 调节区序列特异性 DNA 结合、金属离子结合、RNA 聚合酶 II 转录因子活动, 序列特异性 DNA 结合、离子结合和组蛋白结合。最显著的 5 个细胞组分为神经元细胞体、高尔基体、细胞核、过氧化物酶体膜和细胞内液。具体结果见表 3-11 至表 3-13。

表 3-11 miR-483-5p 候选靶基因最富集的 20 个生物过程

GO 术语编号	GO 术语英文名称	GO 术语中文名称	基因数 <sup>a</sup>	P 值
GO:0006357	regulation of transcription from RNA polymerase II promoter	RNA 聚合酶 II 启动子转录的调控	29	$6.27 \times 10^{-5}$
GO:0060425	lung morphogenesis	肺部形态发生	4	$1.11 \times 10^{-4}$
GO:0006355	regulation of transcription, DNA-templated	DNA 诱导的转录调控	20	$1.24 \times 10^{-4}$
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	RNA 聚合酶 II 启动子转录的正调控	21	$6.63 \times 10^{-4}$
GO:0006351	transcription, DNA-templated	DNA 诱导的转录	8	$1.30 \times 10^{-3}$
GO:0034614	cellular response to reactive oxygen species	细胞对活性氧的反应	4	$3.56 \times 10^{-3}$
GO:0048538	thymus development	胸腺发育	4	$5.25 \times 10^{-3}$
GO:1903351	cellular response to dopamine	细胞对多巴胺的反应	3	$6.24 \times 10^{-3}$
GO:0006801	superoxide metabolic process	超氧化物代谢过程	3	$8.77 \times 10^{-3}$
GO:0060324	face development	面部发育	3	$1.07 \times 10^{-2}$
GO:0006535	cysteine biosynthetic process from serine	丝氨酸生物合成半胱氨酸过程	2	$1.49 \times 10^{-2}$
GO:0051493	regulation of cytoskeleton organization	细胞骨架组织的调节	3	$1.49 \times 10^{-2}$
GO:1990000	amyloid fibril formation	淀粉样纤维的形成	3	$1.73 \times 10^{-2}$
GO:0097746	regulation of blood vessel diameter	血管直径的调节	3	$1.85 \times 10^{-2}$
GO:0030878	thyroid gland development	甲状腺发育	3	$1.98 \times 10^{-2}$
GO:0048009	insulin-like growth factor receptor signaling pathway	胰岛素样生长因子受体信号传导途径	3	$1.98 \times 10^{-2}$
GO:0019343	cysteine biosynthetic process via cystathionine	通过胱硫醚进行的半胱氨酸生物合成过程	2	$2.23 \times 10^{-2}$
GO:0061308	cardiac neural crest cell development involved in heart development	参与心脏发育的心脏神经嵴细胞发育	2	$2.23 \times 10^{-2}$
GO:0001958	endochondral ossification	软骨内骨化	3	$2.25 \times 10^{-2}$
GO:0001701	in utero embryonic development	子宫内的胚胎发育	6	$2.26 \times 10^{-2}$

<sup>a</sup> 富集在该术语的输入基因的数量。

表 3-12 miR-483-5p 候选靶基因最富集的 20 个分子功能

GO 术语编号	GO 术语英文名称	GO 术语中文名称	基因数 <sup>a</sup>	P 值
GO:0000977	RNA polymerase II regulatory region sequence-specific DNA binding	RNA 聚合酶 II 调节区序列特异性 DNA 结合	11	$5.92 \times 10^{-4}$
GO:0046872	metal ion binding	金属离子结合	37	$7.48 \times 10^{-4}$
GO:0000981	RNA polymerase II transcription factor activity, sequence-specific DNA binding	RNA 聚合酶 II 转录因子活动, 序列特异性 DNA 结合	22	$1.10 \times 10^{-3}$
GO:0043167	ion binding	离子结合	3	$3.24 \times 10^{-3}$
GO:0042393	histone binding	组蛋白结合	7	$4.09 \times 10^{-3}$
GO:0000978	RNA polymerase II core promoter proximal region sequence-specific DNA binding	RNA 聚合酶 II 核心启动子近端区域序列特异性 DNA 结合	19	$7.28 \times 10^{-3}$
GO:0070025	carbon monoxide binding	一氧化碳结合	2	$1.57 \times 10^{-2}$
GO:0004122	cystathionine beta-synthase activity	胱硫醚 $\beta$ 合成酶活动	2	$1.57 \times 10^{-2}$
GO:0004124	cysteine synthase activity	半胱氨酸合成酶的活动	2	$1.57 \times 10^{-2}$
GO:0004712	protein serine/threonine/tyrosine kinase activity	蛋白丝氨酸/苏氨酸/酪氨酸激酶活动	9	$2.34 \times 10^{-2}$
GO:0005515	protein binding	蛋白结合	112	$2.50 \times 10^{-2}$
GO:0070026	nitric oxide binding	一氧化氮结合	2	$3.11 \times 10^{-2}$
GO:0016301	kinase activity	激酶活动	6	$3.13 \times 10^{-2}$
GO:0031625	ubiquitin protein ligase binding	泛素蛋白连接酶结合	7	$3.44 \times 10^{-2}$
GO:0001228	transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific binding	转录激活因子活动, RNA 聚合酶 II 转录调节区序列特异性结合	9	$3.50 \times 10^{-2}$
GO:0000976	transcription regulatory region sequence-specific DNA binding	转录调控区序列特异性 DNA 结合	6	$3.80 \times 10^{-2}$
GO:0004674	protein serine/threonine kinase activity	蛋白丝氨酸/苏氨酸激酶活动	8	$3.80 \times 10^{-2}$
GO:0050421	nitrite reductase (NO-forming) activity	亚硝酸盐还原酶 (NO 形成) 活动	2	$3.88 \times 10^{-2}$

续表 3-12 miR-483-5p 候选靶基因最富集的 20 个分子功能

GO 术语编号	GO 术语英文名称	GO 术语中文名称	基因数 <sup>a</sup>	<i>P</i> 值
GO:0016167	glial cell-derived neurotrophic factor receptor activity	胶质细胞源性神经营养因子受体活动	2	$3.88 \times 10^{-2}$
GO:0004111	creatine kinase activity	肌酸激酶活动	2	$4.64 \times 10^{-2}$

<sup>a</sup> 富集在该术语的输入基因的数量。

表 3-13 miR-483-5p 候选靶基因富集的 8 个细胞组分

GO 术语编号	GO 术语英文名称	GO 术语中文名称	基因数 <sup>a</sup>	<i>P</i> 值
GO:0043025	neuronal cell body	神经元细胞体	10	$1.81 \times 10^{-3}$
GO:0005794	Golgi apparatus	高尔基体	18	$4.79 \times 10^{-3}$
GO:0005634	nucleus	细胞核	59	$1.33 \times 10^{-2}$
GO:0005778	peroxisomal membrane	过氧化物酶体膜	4	$1.44 \times 10^{-2}$
GO:0005829	cytosol	细胞内液	54	$2.24 \times 10^{-2}$
GO:0005770	late endosome	晚期内吞体	5	$2.81 \times 10^{-2}$
GO:0031226	intrinsic component of plasma membrane	质膜固有成分	3	$3.44 \times 10^{-2}$
GO:0000118	histone deacetylase complex	组蛋白脱乙酰酶复合物	3	$4.29 \times 10^{-2}$

<sup>a</sup> 富集在该术语的输入基因的数量。

### 3.7.3 miR-548e-3p 候选靶基因的 GO 富集分析结果

对 miR-548e-3p 的 279 个候选靶基因进行 GO 富集分析, 结果显示富集最显著的 5 个生物过程为蛋白质自磷酸化的正调控、mRNA 加工、脂质分解代谢过程、蛋白质在线粒体膜上定位的建立和翻译的负调控, 其他显著的生物过程包括有丝分裂细胞周期的 G2/M 期转换、葡萄糖稳态、钙离子跨膜转运蛋白活性的负调控、钙离子出细胞的负调控、ryanodine 敏感性钙释放通道活性的负调控、高电压门控钙通道活性的调控、心脏传导中电耦合对细胞通讯的调控等。最显著的 5 个分子功能为 mRNA 结合、RNA 结合、蛋白结合、蛋白质丝氨酸/苏氨酸激酶激活因子活动和 mRNA 5'-UTR 结合。最显著的 5 个细胞组分为核质体、细胞质、细胞内液、细胞质应激颗粒和细胞核。具体结果见表 3-14 至表 3-16。

表 3-14 miR-548e-3p 候选靶基因最富集的 20 个生物过程

GO 术语编号	GO 术语英文名称	GO 术语中文名称	基因数 <sup>a</sup>	P 值
GO:0031954	positive regulation of protein autophosphorylation	蛋白质自磷酸化的正调控	6	$1.73 \times 10^{-5}$
GO:0006397	mRNA processing	mRNA 加工	12	$1.87 \times 10^{-4}$
GO:0016042	lipid catabolic process	脂质分解代谢过程	8	$3.40 \times 10^{-4}$
GO:0090151	establishment of protein localization to mitochondrial membrane	蛋白质在线粒体膜上定位的建立	3	$5.26 \times 10^{-4}$
GO:0017148	negative regulation of translation	翻译的负调控	7	$9.02 \times 10^{-4}$
GO:0000086	G2/M transition of mitotic cell cycle	有丝分裂细胞周期的 G2/M 期转换	6	$9.63 \times 10^{-4}$
GO:0008380	RNA splicing	RNA 剪接	10	$1.49 \times 10^{-3}$
GO:0051343	positive regulation of cyclic-nucleotide phosphodiesterase activity	对环核苷酸磷酸二酯酶活性的正调控	3	$1.72 \times 10^{-3}$
GO:1905913	negative regulation of calcium ion export from cell	钙离子出细胞的负调控	3	$1.72 \times 10^{-3}$
GO:0010801	negative regulation of peptidyl-threonine phosphorylation	肽基苏氨酸磷酸化的负调控	4	$1.94 \times 10^{-3}$
GO:0060314	regulation of ryanodine-sensitive calcium-release channel activity	ryanodine 敏感性钙释放通道活性的负调控	4	$1.94 \times 10^{-3}$
GO:0006468	protein phosphorylation	蛋白质磷酸化	16	$2.45 \times 10^{-3}$
GO:1901020	negative regulation of calcium ion transmembrane transporter activity	钙离子跨膜转运蛋白活性的负调控	3	$3.56 \times 10^{-3}$
GO:0032092	positive regulation of protein binding	蛋白质结合的正向调节	6	$3.66 \times 10^{-3}$
GO:1901841	regulation of high voltage-gated calcium channel activity	高电压门控钙通道活性的调控	3	$4.70 \times 10^{-3}$
GO:0042593	glucose homeostasis	葡萄糖稳态	7	$4.83 \times 10^{-3}$
GO:0051028	mRNA transport	mRNA 运输	6	$5.58 \times 10^{-3}$
GO:0060316	positive regulation of ryanodine-sensitive calcium-release channel activity	ryanodine 敏感性钙释放通道活性的正调控	3	$5.99 \times 10^{-3}$
GO:1901844	regulation of cell communication by electrical coupling involved in cardiac conduction	心脏传导中电耦合对细胞通讯的调控	3	$5.99 \times 10^{-3}$
GO:0010800	positive regulation of peptidyl-threonine phosphorylation	肽基苏氨酸磷酸化的正调控	2	$6.62 \times 10^{-3}$

<sup>a</sup> 富集在该术语的输入基因的数量。

表 3-15 miR-548e-3p 候选靶基因最富集的 20 个分子功能

GO 术语编号	GO 术语英文名称	GO 术语中文名称	基因数 <sup>a</sup>	P 值
GO:0003729	mRNA binding	mRNA 结合	16	$1.45 \times 10^{-6}$
GO:0003723	RNA binding	RNA 结合	44	$1.96 \times 10^{-6}$
GO:0005515	protein binding	蛋白结合	202	$1.72 \times 10^{-4}$
GO:0043539	protein serine/threonine kinase activator activity	蛋白质丝氨酸/苏氨酸激酶激活因子活动	7	$2.50 \times 10^{-4}$
GO:0048027	mRNA 5'-UTR binding	mRNA 5'-UTR 结合	5	$4.17 \times 10^{-4}$
GO:0031997	N-terminal myristoylation domain binding	N 端肉豆蔻酰化结构域结合	3	$5.64 \times 10^{-4}$
GO:0019904	protein domain specific binding	蛋白质结构域特异性结合	12	$6.68 \times 10^{-4}$
GO:0031800	type 3 metabotropic glutamate receptor binding	3 型代谢性谷氨酸受体结合	3	$1.85 \times 10^{-3}$
GO:0010856	adenylate cyclase activator activity	腺苷酸环化酶激活因子活动	3	$2.74 \times 10^{-3}$
GO:0030235	nitric-oxide synthase regulator activity	硝酸合成酶调控因子活动	3	$5.03 \times 10^{-3}$
GO:0035925	mRNA 3'-UTR AU-rich region binding	mRNA 3'-UTR AU-rich 区域结合	4	$5.34 \times 10^{-3}$
GO:0046875	ephrin receptor binding	肾上腺素受体结合	4	$8.78 \times 10^{-3}$
GO:0008179	adenylate cyclase binding	腺苷酸环化酶结合	3	$1.14 \times 10^{-2}$
GO:0045296	cadherin binding	粘附蛋白结合	11	$1.27 \times 10^{-2}$
GO:0031432	titin binding	肌联蛋白结合	3	$1.34 \times 10^{-2}$
GO:0097718	disordered domain specific binding	无序结构域特异性结合	4	$1.43 \times 10^{-2}$
GO:0019855	calcium channel inhibitor activity	钙通道抑制因子活动	3	$1.55 \times 10^{-2}$
GO:0072542	protein phosphatase activator activity	蛋白磷酸酶激活因子活动	3	$1.55 \times 10^{-2}$
GO:0050998	nitric-oxide synthase binding	一氧化氮合成酶结合	3	$1.77 \times 10^{-2}$
GO:0019003	GDP binding	GDP 结合	5	$1.92 \times 10^{-2}$

<sup>a</sup> 富集在该术语的输入基因的数量。

表 3-16 miR-548e-3p 候选靶基因最富集的 20 个细胞组分

GO 术语编号	GO 术语英文名称	GO 术语中文名称	基因数 <sup>a</sup>	P 值
GO:0005654	nucleoplasm	核质体	84	$4.29 \times 10^{-6}$
GO:0005737	cytoplasm	细胞质	109	$5.07 \times 10^{-6}$
GO:0005829	cytosol	细胞内液	106	$1.11 \times 10^{-5}$
GO:0010494	cytoplasmic stress granule	细胞质应激颗粒	8	$2.29 \times 10^{-4}$
GO:0005634	nucleus	细胞核	107	$2.35 \times 10^{-4}$
GO:0030426	growth cone	生长锥	9	$4.94 \times 10^{-4}$
GO:1990904	ribonucleoprotein complex	核糖核蛋白复合物	10	$8.60 \times 10^{-4}$
GO:0032433	filopodium tip	丝状体尖端	4	$2.52 \times 10^{-3}$
GO:0030027	lamellipodium	板状伪足	9	$4.02 \times 10^{-3}$
GO:0031252	cell leading edge	细胞前缘	5	$5.92 \times 10^{-3}$
GO:1902494	catalytic complex	催化复合物	4	$6.40 \times 10^{-3}$
GO:0005813	centrosome	中心体	16	$7.39 \times 10^{-3}$
GO:0032991	macromolecular complex	大分子复合物	17	$1.92 \times 10^{-2}$
GO:0005876	spindle microtubule	纺锤体微管	4	$2.13 \times 10^{-2}$
GO:0008076	voltage-gated potassium channel complex	电压门控钾通道复合体	5	$2.22 \times 10^{-2}$
GO:0014069	postsynaptic density	突触后致密区	9	$2.22 \times 10^{-2}$
GO:0030992	intraciliary transport particle B	睫状体内运输粒子 B	3	$2.55 \times 10^{-2}$
GO:0043235	receptor complex	受体复合物	8	$2.63 \times 10^{-2}$
GO:0097225	sperm midpiece	精子中段	4	$2.81 \times 10^{-2}$
GO:0030054	cell junction	细胞连接	8	$3.64 \times 10^{-2}$

<sup>a</sup> 富集在该术语的输入基因的数量。



## 4 讨论

本研究选择相对健康的社区人群作为研究人群。在本研究人群中,我们发现除 2-OH-Phe 与 2-OH-Nap 关联性不显著之外,10 种 OH-PAHs 之间都相互表现出显著关联,而 BPDE-白蛋白加合物与 OH-PAHs 之间的关联不显著。有多种可能原因导致这一结果:如相似的代谢途径和相似的暴露来源等。

现有的研究提示,PAHs 暴露与 miRNA 的表达水平有关,且可导致 miRNA 表达改变,从而对细胞功能、机体生理和病理过程产生影响。细胞研究显示,PAHs 暴露可以导致 miR-126、mir-30c-1-3p 等 miRNA 表达下降和 miR-190a、miR-194-3p、miR-21 等 miRNA 表达上升,从而改变 *EGFL7*、*TP53INP1*、*PHLPP1* 等基因和炎症介质的表达,造成细胞活力、侵袭能力的变化等。流行病学研究则显示,PAHs 暴露与 miR-106a、miR-34c、miR-24-3p 等 miRNA 表达下降和 miR-126、miR-155、let-7b-5p 等 miRNA 表达上升有关,这些 miRNA 与精子活力、细胞微核率、心血管事件等有关,可调控相关靶基因参与 DNA 损伤、氧化应激等反应<sup>[11]</sup>。本次研究发现了 42 个与 PAHs 内暴露相关的 miRNA。其中 23 个与 2-OH-Flu 相关,13 个与 9-OH-Phe 相关,5 个与上述两者都相关,1 个与 1-OH-Phe 相关。在本研究发现的 42 个 miRNA 中,miR-21-3p、miR-320b 和 miR-33a-5p 曾被报道过在高 PAHs 暴露水平与对照组中差异表达<sup>[50]</sup>。

在本研究中发现的与 PAHs 暴露相关性最高的 miR-505-5p、miR-125b-2-3p、miR-483-5p 和 miR-548e-3p 中,未发现明显组织特异性,且彼此之间高度相关,提示 miRNA 可能具有相似的组织来源。未来需要更多研究探索人体不同组织中的 miRNA 特异性。

对 4 个 miRNA 筛选候选靶基因,开展 KEGG 通路富集分析和 GO 富集分析。发现 miR-505-5p 的靶基因显著富集于 KEGG 通路 HIF-1 信号传导通路、卵母细胞减数分裂和 Wnt 信号传导通路。HIF-1 信号通路是生物处于低氧状态下的应激信号通路,在缺氧条件下,该通路促进细胞厌氧代谢、提高 ATP 的合成、促使血管内皮细胞分裂增殖、诱导血管生成再生,维持心肌细胞功能。但过表达的 HIF-1 $\alpha$ 会过高促进细胞厌氧代谢,造成细胞能量代谢紊乱,进而破坏心肌细胞<sup>[51]</sup>。在发育过程中,Wnt 信号传导通路在细胞命运、增殖、迁移、极性和死亡过程中发挥作用,且是神经系统发育过程中的重要角色。在成人中,Wnt 信号传导通路

在维持体内稳态过程中发挥重要功能,其异常激活与多种癌症有关<sup>[52, 53]</sup>。GO 富集分析结果显示,miR-505-5p 显著富集于突触组装、轴突引导、血管生成、血管内皮生长因子激活受体活动、中枢神经系统发育、成人行为、先天免疫反应的负调控、Notch 信号通路的正调控等过程。Notch 信号通路是相邻细胞之间通讯进而调控细胞发育的重要通路,广泛参与了恶性肿瘤的发生发展<sup>[54]</sup>。简而言之,miR-505-5p 的靶基因功能与氧化应激、神经系统发育、癌症、胚胎发育、免疫调控等过程有关。

miR-125b-2-3p 的靶基因没有显著富集的 KEGG 通路。GO 富集分析结果显示靶基因富集于细胞对葡萄糖缺乏的反应、干细胞种群维护的调节、血管内皮生长因子受体信号传导途径调控、记忆、视黄酸受体结合等过程中。干细胞是人体的起源细胞,具有自我复制、高度增殖和多向分化潜能的特性,在胚胎发育、组织和器官的发育与修复中发挥重要作用<sup>[55]</sup>。视黄酸是体内维生素 A 的代谢中间产物,有促进角质溶解、上皮细胞增生和分化等代谢作用,且在胚胎发育阶段对血管发生的影响至关重要。此外,视黄酸可用于早幼粒白血病和表皮癌等癌症的防治<sup>[56, 57]</sup>。由此可见,miR-125b-2-3p 的靶基因功能与葡萄糖稳态、胚胎发育、记忆、癌症等过程有关。

miR-483-5p 的靶基因显著富集于 7 条 KEGG 通路,分别为 MAPK 信号通路、氨基酸的生物合成、甘氨酸、丝氨酸和苏氨酸代谢、丙型肝炎、单纯疱疹病毒 1 型感染、癌症的中心碳代谢、癌症中的胆碱代谢。MAPK 通路由一种三级激酶模式组成,调控细胞的生长、分化、环境应激、炎症反应等多种重要的细胞生理病理过程<sup>[58]</sup>。GO 富集分析结果显示靶基因富集于肺部形态发生、细胞对活性氧的反应、胸腺发育、血管直径的调节、甲状腺发育、参与心脏发育的心脏神经嵴细胞发育、软骨内骨化、子宫内的胚胎发育等过程。由此可见,miR-483-5p 的靶基因功能与蛋白质新陈代谢、多种器官发育、胚胎发育、氧化应激、遗传损伤、炎症反应、癌症、丙型肝炎、单纯疱疹病毒 1 型感染等过程有关。

miR-548e-3p 的靶基因显著富集于 KEGG 通路胰高血糖素信号传导通路、心肌细胞肾上腺素能信号传导、流体剪切应力和动脉粥样硬化。胰高血糖素具有促进糖原分解和糖异生、促进脂肪分解、促进胰岛素正性作用和胰岛生长抑素分泌的作用。心肌细胞中,肾上腺素作用于 $\beta 1$ 受体产生加强心肌收缩性,加速传导,

加速心率,提高心肌的兴奋性,提高心肌代谢。作用于 $\beta_3$ 受体则可产生负性肌力作用,可能参与了心衰的病理生理过程<sup>[59]</sup>。GO富集分析结果显示靶基因富集于脂质分解代谢过程、有丝分裂细胞周期的G2/M期转换、葡萄糖稳态、钙离子跨膜转运蛋白活性的负调控、钙离子出细胞的负调控、ryanodine敏感性钙释放通道活性的负调控、高电压门控钙通道活性的调控、心脏传导中电耦合对细胞通讯的调控等过程。结果可见,靶基因多方面影响钙稳态。结果提示miR-548e-3p的靶基因功能与细胞增殖、钙稳态、葡萄糖稳态、脂质分解、心肌细胞信号传导、动脉粥样硬化等过程有关。

总而言之,4个miRNA的靶基因可能参与了心血管疾病、氧化应激、遗传损伤、免疫调控、炎症反应、癌症、丙型肝炎、单纯疱疹病毒1型感染、钙稳态、葡萄糖稳态、神经系统发育、胚胎发育、多个器官发育、记忆等多个生物学过程。这与先前的研究结论相互印证。接触PAHs对健康的影响包括免疫功能下降、肺功能异常、精神错乱等。同时,PAHs反复接触皮肤可能会诱发红肿和皮肤炎症<sup>[14-16]</sup>。PAHs已被证明有遗传毒性、致癌性、免疫毒性和发育毒性。在酶解过程中,PAHs生成多种致癌代谢产物和大量活性氧自由基,造成癌症、氧化性DNA损伤和氧化应激<sup>[18]</sup>。母体血液中的PAHs可通过胎盘屏障进入胎儿体内,影响胎儿的生长发育,增加胎儿神经管缺陷患病风险<sup>[60]</sup>。此外,动物研究结果表明,长期暴露于环境多环芳烃会通过诱导促炎细胞因子产生增加2型糖尿病患病风险<sup>[61]</sup>。流行病学研究显示,低分子量PAHs水平与糖尿病患病风险呈正相关<sup>[62]</sup>,另一项包含6篇横断面研究的meta分析则发现OH-PAHs水平升高与2型糖尿病患病风险呈正相关<sup>[63]</sup>。一些研究则显示,接触PAHs与心血管疾病如动脉粥样硬化、胸闷和心率变化、血压升高和缺血性心脏病分别有关<sup>[17]</sup>。综合这些研究结果,我们推测,miRNA可能介导了机体对PAHs的应答反应,可能是PAHs导致免疫失调、炎症反应、氧化应激、遗传损伤、畸形、癌症、糖尿病和心血管疾病中的重要介导机制。

本研究具有如下优点。第一,本研究中的社区人群的性别分布平均、年龄分布广泛,得到的结果更具有广泛性和代表性;第二,对血浆中的miRNA进行测序,得到通过质控的1022个血浆miRNA,可探索式发现与PAHs暴露相关的miRNA。本研究也存在一些不足。第一,本研究为横断面研究,不能断定PAHs

暴露与 miRNA 之间的因果关联, 需要细胞、动物或者人群研究断定其之间的因果关联; 第二, 尽管已经校正了年龄、性别、BMI、当前吸烟状况、当前饮酒状况、被动吸烟和饮食摄入等主要混杂因素, PAHs 的来源较多, 未能排除其他混杂因素的影响; 但本研究已检测尿液 PAHs 浓度, 可以反应人体的内暴露水平。

综上所述, 本部分的研究结果表明, PAHs 暴露可能会影响 miRNA 表达情况。发现了 42 个 PAHs 暴露相关 miRNA, 其中有 4 个与 2-OH-Flu 相关的 miRNA 达到了更严格的显著水平, 分别为 2-OH-Flu 显著正相关的 miR-125b-2-3p、miR-505-5p 和 miR-483-5p 以及显著负相关的 miR-548e-3p。对候选靶基因富集分析后, 发现这些 miRNA 可能介导了机体对 PAHs 的应答反应, 可能是 PAHs 导致畸形、癌症、糖尿病、心血管疾病、免疫失调、炎症反应、氧化应激和遗传损伤中的重要介导机制。

## 5 结论

本研究在相对健康的社区人群中发现,除 2-OH-Phe 与 2-OH-Nap 之外,10 种 OH-PAHs 之间都相互表现出显著关联,而 BPDE-白蛋白加合物与 OH-PAHs 之间的关联不显著。本研究发现了 42 个与 PAHs 暴露显著相关的 miRNA。其中 23 个与 2-OH-Flu 相关,13 个与 9-OH-Phe 相关,5 个与上述两者都相关,1 个与 1-OH-Phe 相关。其中有 4 个与 2-OH-Flu 相关的 miRNA 达到了更严格的显著水平,分别为 2-OH-Flu 显著正相关的 miR-125b-2-3p、miR-505-5p 和 miR-483-5p 以及显著负相关的 miR-548e-3p,被用于后续分析。

在年龄 <45.0 岁的人群中,2-OH-Flu 与 miR-125b-2-3p、miR-483-5p、miR-548e-3p 关联更加显著;在否认存在被动吸烟的人群中,2-OH-Flu 与 miR-483-5p、miR-548e-3p 关联更加显著;在 BMI <25.0 kg/m<sup>2</sup> 的人群中,2-OH-Flu 与 miR-483-5p 关联更加显著。交互作用分析的结果显示,BMI 和 2-OH-Flu 对 miR-505-5p 和 miR-483-5p 表达水平存在交互作用。miR-505-5p、miR-125b-2-3p、miR-483-5p 和 miR-548e-3p 的 TSI 分别为 0.33、0.30、0.37 和 0.59,提示这些 miRNA 没有明显组织特异性。4 个相关 miRNA 之间相互表现出显著关联。

对于上述 4 个 miRNA,通过公共数据库分别预测到 379 个 miR-505-5p 的候选靶基因,534 个 miR-125b-2-3p 的候选靶基因,168 个 miR-483-5p 的候选靶基因,279 个 miR-548e-3p 的候选靶基因。开展 KEGG 通路富集分析和 GO 富集分析,发现 4 个 miRNA 的靶基因功能富集在心血管疾病、氧化应激、遗传损伤、免疫调控、炎症反应、癌症、丙型肝炎、单纯疱疹病毒 1 型感染、钙稳态、葡萄糖稳态、神经系统发育、胚胎发育、多个器官发育、记忆等通路与生物学过程上。综合其他已经发表的研究结果,我们推测,miRNA 可能介导了机体对 PAHs 的应答反应,可能是 PAHs 导致畸形、癌症、糖尿病、心血管疾病、免疫失调、炎症反应、氧化应激和遗传损伤中的重要介导机制。

## 致谢

五年的大学时光匆匆流逝，时光荏苒，回望本科五年的旅程，有许多人我想要感谢，感谢他们对我成长道路和科研学习的帮助、鼓励与陪伴，让我找到了属于我的方向。

毕业设计是本科阶段的最后一课，我首先要感谢邬堂春教授，感谢他对我毕业设计的建设性指导和意见，让我对科研有了更深刻的认识。我还要感谢王雨菲师姐在毕业设计过程中对我的帮助，从开题到结题，师姐不断地与我交流沟通，对试验方案提出意见并对论文做出修正。教授与师姐谦逊的科研态度和严谨的科研精神让我受益匪浅。

感谢本科辅导员赵金辉老师对我的帮助。在我陷入低谷的时候，他给予我的鼓励让我有了坚持的勇气。我还想感谢王超龙教授、郝兴杰教授与朱紫微师姐在本科大创项目期间对我的指导和教育，把我领进了科研的大门，让我坚定了就读研究生继续科研的想法。还要感谢所有教授和师长们的倾囊相授与谆谆教诲，以及学校提供的资源与平台。从小我就有着想要帮助他人的小小心愿，这一心愿逐渐成长为了少年时治病救人的理想。得益于学校与师生的帮助，我收获了太多知识和道理，更找到了生物统计这一今后前行的道路和目标。

我要感谢我的家人和朋友。他们的关心和陪伴让我的生活充满了温暖与爱；他们的支持和鼓励让我能够在自己所热爱的专业里坚持前行。最后，我也感谢过去五年的我自己。持之以恒的努力，总有一天会看到回报。五年时光很长，有过挫折、低谷、迷茫，但更多是幸福、坚定、满足；五年时间很短，一眨眼我就又站在了新旅程的起点。愿所有的经历都可以化作我成长的养料和前进的燃料，在未来的研究生阶段和更久远的以后，帮助我成长为更好的自己。长风破浪会有时，直挂云帆济沧海！