GEO多芯片联合eQTL、机器学习预测抑郁症的生物标志物及免疫浸润分析

梁 科1*, 何乾超2#, 高玉广3, 赵 帅1, 唐海秀1

¹广西中医药大学研究生院,广西 南宁 ²广西中医药大学第一附属医院老年病科,广西 南宁 ³广西中医药大学第一附属医院急诊科,广西 南宁

收稿日期: 2025年5月1日; 录用日期: 2025年6月27日; 发布日期: 2025年7月8日

摘要

抑郁症是一种普遍的心理障碍,影响着全球数以百万计的人群。其疗效不稳定和高复发率。近年来,基因表达数量性状位点(expression quantitative trait loci, eQTL)研究为理解抑郁症的遗传基础提供了新的视角。本研究从eQTLGen联盟获取了顺式eQTL数据(ukb-d-F5_DEPRESSIO、ebi-a-GCST003769和ebi-a-GCST005902),并结合GEO数据库中的五个基因表达数据集(GSE98793、GSE19738、GSE44593、GSE54568和GSE54570),筛选出与抑郁症相关的差异表达基因。选择合适的工具变量进行MR分析,采用多种统计方法综合评估eQTL与抑郁症之间的因果关系。随后,利用受试者工作特征(Receiver Operating Characteristic, ROC)曲线分析突出诊断潜力的基因。基因本体(Gene Ontology, GO)和KEGG通路分析生物过程。LASSO回归分析进一步筛选核心基因。构建转录因子(TF)-miRNA-枢纽基因调控网络,识别miRNA和转录因子。最终,通过CIBERSORT算法分析抑郁症患者与健康对照组之间的免疫细胞浸润差异,并探讨核心基因与免疫细胞浸润之间的相关性。结果LASSO回归分析保留了4个关键基因(LTF、OLFM4、AKR1C3、WEE1),排除了EVI2A,因为其AUC值低于0.5。我们构建了一个转录因子(TF)-miRNA枢纽基因调控网络,识别出48个miRNA和56个转录因子。免疫细胞浸润分析显示抑郁症组与对照组之间存在显著差异,记忆性CD4+T细胞和巨噬细胞的比例发生了改变。这些发现强调了识别出的基因和免疫细胞作为治疗靶点的潜力。未来的研究应专注于这些发现的临床应用和机制探索,以改善抑郁症的早期诊断和治疗。

关键词

GEO, eQTL, 机器学习, 抑郁症, 免疫浸润

文章引用:梁科,何乾超,高玉广,赵帅,唐海秀.GEO 多芯片联合 eQTL、机器学习预测抑郁症的生物标志物及免疫 浸润分析[J]. 生物医学, 2025, 15(4): 688-701. DOI: 10.12677/hjbm.2025.154076

^{*}第一作者。

[#]通讯作者。

GEO Multi-Chip Integration of eQTL, Machine Learning Prediction of Biomarkers for Depression, and Immune Infiltration Analysis

Ke Liang^{1*}, Qianchao He^{2#}, Yuguang Gao³, Shuai Zhao¹, Haixiu Tang¹

¹Graduate School of Guangxi University of Chinese Medicine, Nanning Guangxi

²Department of Geriatrics, The First Affiliated Hospital of Guangxi University of Chinese Medicine, Nanning Guangxi

³The First Section of Department of Emergency, The First Affiliated Hospital of Guangxi University of Chinese Medicine, Nanning Guangxi

Received: May. 1st, 2025; accepted: Jun. 27th, 2025; published: Jul. 8th, 2025

Abstract

Depression is a common mental disorder that affects millions of people worldwide, characterized by unstable efficacy and a high relapse rate. In recent years, research on expression quantitative trait loci (eQTL) has provided new insights into the genetic basis of depression. This study obtained cis-eQTL data from the eQTLGen consortium (ukb-d-F5_DEPRESSIO, ebi-a-GCST003769, and ebi-a-GCST005902) and combined it with five gene expression datasets from the GEO database (GSE98793, GSE19738, GSE44593, GSE54568, and GSE54570) to screen for differentially expressed genes associated with depression. Appropriate instrumental variables were selected for MR analysis, and various statistical methods were used to comprehensively assess the causal relationship between eQTL and depression. Subsequently, receiver operating characteristic (ROC) curve analysis was utilized to highlight genes with diagnostic potential. Gene Ontology (GO) and KEGG pathway analyses were conducted to explore biological processes. LASSO regression analysis further filtered core genes. A transcription factor (TF)-miRNA-hub gene regulatory network was constructed to identify miRNAs and transcription factors. Finally, the CIBERSORT algorithm was used to analyze the differences in immune cell infiltration between depression patients and healthy controls, and to explore the correlation between core genes and immune cell infiltration. Results from the LASSO regression analysis retained four key genes (LTF, OLFM4, AKR1C3, WEE1) and excluded EVI2A due to its AUC value being below 0.5. We constructed a transcription factor (TF)-miRNA-hub gene regulatory network, identifying 48 miRNAs and 56 transcription factors. Immune cell infiltration analysis revealed significant differences between the depression group and the control group, with altered proportions of memory CD4+ T cells and macrophages. These findings emphasize the potential of the identified genes and immune cells as therapeutic targets. Future research should focus on the clinical applications and mechanistic exploration of these findings to improve early diagnosis and treatment of depression.

Keywords

GEO, eQTL, Machine Learning, Depression, Immune Infiltration

Copyright © 2025 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/

C O Open Access

1. 前言

抑郁症作为一种常见的精神障碍,对全球范围内的个体健康和社会功能造成了深远的影响。抑郁症 的全球负担在过去三十年中显著增加,尤其是在高社会发展指数(SDI)和高收入国家。女性和老年人的抑 郁症负担更为严重[1]。尽管抑郁症的发病率在某些地区有所下降,但整体负担仍在增加。这表明需要加 强对抑郁症的研究,以改善预防和治疗措施。尽管目前已有多种治疗方法,包括药物治疗、心理治疗和 物理治疗等,但仍有相当比例的患者对现有治疗反应不佳,复发率高。

抑郁症是一种复杂的心理疾病,其生理机制涉及多种生物学通路和神经生物学变化。近年来的研究 表明,抑郁症可能与神经递质的不平衡、炎症反应、内分泌失调以及神经可塑性变化等因素密切相关。 例如,5-羟色胺(5-HT)、去甲肾上腺素(NE)和多巴胺(DA)等神经递质的水平变化被认为是抑郁症的重要生 理基础[2]。此外,炎症反应在抑郁症的发病机制中也起着关键作用,炎症假说认为炎症细胞因子可以影 响抑郁症患者的神经递质代谢、神经内分泌功能和局部脑活动[3]。此外,内分泌系统的失调,特别是下 丘脑 - 垂体 - 肾上腺(HPA)轴的异常活动,也与抑郁症的发生密切相关。HPA 轴的过度激活会导致皮质 醇水平升高,从而影响情绪调节和认知功能[4]。这些生理机制的复杂交互作用,导致抑郁症的多样性和 个体差异,使得其治疗面临挑战。

尽管近年来在抑郁症生物标志物的研究中取得了一定进展,但现有生物标志物仍存在显著的局限性。 目前,许多生物标志物的研究主要集中在炎症因子、神经递质和激素水平等方面,但这些标志物的临床 应用仍然受到限制。例如,虽然如 BDNF(脑源性神经营养因子)和 IL-6(白细胞介素-6)等标志物在抑郁症 患者中显示出一定的相关性,但其特异性和敏感性不足,难以作为独立的诊断工具[5]。此外,现有的生 物标志物往往无法反映抑郁症的异质性和复杂性。抑郁症的病因和表现因个体差异而异,单一的生物标 志物难以全面评估疾病的严重程度或预测治疗反应。因此,亟需发展多重生物标志物组合,以提高抑郁 症的诊断准确性和治疗效果。

近年来,随着基因组学技术的飞速发展,表达数量性状位点(eQTL)研究为探索抑郁症的遗传基础提供了新的视角。eQTL 是指与基因表达水平相关的遗传变异位点,这些位点可以作为工具变量,用于孟德尔随机化(MR)分析,从而推断基因表达与疾病之间的因果关系[6],通过整合多个队列的 eQTL 数据,研究人员能够更全面地评估基因表达与抑郁症之间的潜在因果联系。此外,机器学习技术在生物医学领域的应用也为抑郁症生物标志物的发现提供了新的方法。机器学习算法能够处理大规模的基因表达数据,识别出与疾病相关的特征基因,并构建诊断模型,在神经系统方面的应用取得较好的典范[7]。

本研究旨在通过双样本孟德尔随机化(MR)分析方法,结合 GEO 数据库中的多芯片联合 eQTL 数据 和 IEU Open GWAS 平台的抑郁症相关结局数据,系统探讨 eQTL 与抑郁症之间的因果关系,并利用机器 学习技术筛选出具有诊断价值的生物标志物。构建转录因子(TF)-miRNA-枢纽基因调控网络,识别 miRNA 和转录因子。最终,通过 CIBERSORT 算法分析抑郁症患者的免疫细胞浸润特征,探讨免疫细胞在抑郁 症中的作用机制。通过这些综合分析,我们期望能够揭示抑郁症的潜在病理机制,为开发新的诊断工具 和治疗靶点提供理论依据。

2. 资料与方法

2.1. 科研设计

本研究旨在通过双样本孟德尔随机化(MR)分析方法,探讨 GEO 数据库中多芯片联合表达数量性状位点(eQTL)与抑郁症之间的因果关系及治疗靶点的预测。研究以 eQTL 作为工具变量,通过 SMR (基于 汇总统计的孟德尔随机化)格式化的顺式 eQTL 数据,评估其对抑郁症的潜在因果影响。研究设计遵循 MR 分析的基本原则,采用多种统计方法对因果关系进行综合评估。

2.2. 数据来源

eQTL 数据来源: eQTL 数据来源于 eQTLGen 联盟的顺式 eQTL 数据(<u>https://www.eqtlgen.org/cis-eqtls.html</u>),具体使用 SMR 格式化的 eQTL 结果。这些数据基于多个队列的联合分析,涵盖了大量样本的基因表达与 SNP 关联信息,为本研究提供了丰富的暴露变量数据。

结局数据来源:结局数据来源于 IEU Open GWAS 平台的三个独立队列,具体如下:

ukb-d-F5_DEPRESSIO.vcf.gz: 英国生物银行(UK Biobank)的抑郁症诊断数据(二分类变量),样本量为 361,194 例, SNP 数量为 9,518,776 个。ebi-a-GCST003769.vcf.gz: 重性抑郁障碍病例对照研究,样本量为 180,866 例, SNP 数量为 6,019,632 个。ebi-a-GCST005902.vcf.gz: 基于症状评分的抑郁严重程度数据(连 续变量),样本量为 322,580 例, SNP 数量为 7,624,934 个(如表 1)。

Table 1. EQTL data information 表 1. eQTL 数据信息

| _ | | | |
|---|--------------------|-----------|-------------|
| | 数据 ID | 样本量 | SNP 数量 |
| | ukb-d-F5_DEPRESSIO | 361,194 例 | 9,518,776 个 |
| | ebi-a-GCST003769 | 180,866 例 | 6,019,632 个 |
| | ebi-a-GCST005902 | 322,580 例 | 7,624,934 个 |
| | | | |

芯片数据来源:芯片数据来源于 GEO 数据库的多个基因表达数据集,具体如下:

GSE98793:包含 128 例抑郁症患者和 64 名健康对照的外周血基因表达谱芯片数据,是目前 GEO 数据库中样本量最大的抑郁症数据集。平台为 GPL570。GSE19738:包含 66 例抑郁症患者和 66 名健康对照的外周血基因表达谱芯片数据。平台为 GPL570。GSE44593:包含 14 例抑郁症患者和 14 名健康对照的外周血基因表达谱芯片数据。平台为 GPL570。GSE54568:包含 15 例抑郁症患者和 15 名健康对照的外周血基因表达谱芯片数据。平台为 GPL570。GSE54570:包含 13 例抑郁症患者和 13 名健康对照的外周血基因表达谱芯片数据。平台为 GPL570。GSE54570:包含 13 例抑郁症患者和 13 名健康对照的外周血基因表达谱芯片数据。平台为 GPL96。从 GEO 数据库中提取上述五个数据集的基因表达数据,利用 R 软件 limma 包对脑出血组和健康组数据进行差异表达分析,依据 llogFC > 0.2 和 P < 0.05 进行差异表达基因筛选(如表 2)。

| 表 2. GEO 数据信息 | | | | |
|----------------------|------------|----------|--------|--|
| 数据集 | 疾病组 | 健康组 | 平台 | |
| GSE98793 | 128 例抑郁症患者 | 64 名健康对照 | GPL570 | |
| GSE19738 | 66 例抑郁症患者 | 66 名健康对照 | GPL570 | |
| GSE44593 | 14 例抑郁症患者 | 14 名健康对照 | GPL570 | |
| GSE54568 | 15 例抑郁症患者 | 15 名健康对照 | GPL570 | |
| GSE54570 | 13 例抑郁症患者 | 13 名健康对照 | GPL96 | |

Table 2. GEO data information

2.3. 筛选工具变量

本研究采用以下步骤筛选合适的工具变量。

2.3.1. eQTL 数据筛选

eQTLGen 数据库中 SMR 格式化的顺式 eQTL 数据,选择与基因表达显著相关的 SNP 作为候选工具 变量。

2.3.2. 连锁不平衡(LD)筛选

设置连锁不平衡(LD)参数 r² 阈值为 0.001,遗传距离为 10,000 kb,以确保 SNP 之间相互独立,并排除可能的连锁不平衡影响。

2.3.3. 弱工具变量排除

剔除回文 SNP,并选取 F 统计量大于 10 的 SNP,以尽可能排除弱工具变量的影响。

2.4. MR 分析方法

本研究采用 R 软件中的 TwoSampleMR 包进行 MR 分析,具体方法如下。

2.4.1. 主要分析方法

逆方差加权法(IVW)作为主要分析方法,用于估计 eQTL 与抑郁症之间的因果效应。

2.4.2. 次要分析方法

MR-Egger 回归、加权中位数法、简单模型和加权模型作为次要分析方法,用于验证主要分析结果的 稳健性。

2.4.3. 异质性评估

采用 Cochran's Q 检验评估工具变量间的统计学异质性,若 P > 0.05,则认为工具变量间无显著异质性,采用固定效应模型进行因果关系估计;若 P ≤ 0.05,则采用随机效应模型。

2.4.4. 水平多效性评估

通过 MR-Egger 回归的截距项评估工具变量的水平多效性,若 P > 0.05,则认为不存在水平多效性。

2.4.5. 敏感性分析

采用留一法(leave-one-out)分析评估单个 SNP 对 MR 分析结果的影响,以验证结果的稳定性。

2.4.6. 偏倚评估

通过漏斗图分析评估 MR 分析结果是否受到潜在因素的影响。

2.5. 核心基因的筛选

将"1.2"筛选出的差异表达基因并与上述筛选出的工具变量取交集;表达上调的基因与 OR 值大于 1 的 SNP 取交集;表达下调的基因与 OR 值小于 1 的 SNP 取交集。

2.6. ROC 构建诊断模型

将 "1.5" 项获得的交集基因通过 cutpointr 程序包中的 multi_cutpointr 函数对差异表达基因进行批量 受试者工作特征(receiver operating characteristic, ROC)曲线分析,评估差异表达基因的诊断价值,AUC 曲线下面积越大的基因表达越显著,筛选 AUC 曲线下面积大于 > 0.5 的基因纳入诊断模型构建分析。

2.7. GO 富集与 KEGG 通路分析

将"1.6"项筛选出的基因进行 GO 和 KEGG 富集分析,以 P 值大小进行排序,将生物过程(Biological process, BP)、细胞组分(Cellular component, CC)、分子功能(Molecular function, MF)分别将 P 位于前十的

绘制气泡图; KEGG 分析则以 P 值大小进行排序, 将 P 位于前三十的绘制气泡图。

2.8. 通过机器学习进一步筛选关键基因

将"1.5"项获得的交集基因,采用 LASSO 回归对得到的差异泛素化基因进行进一步筛选,以得到 具有良好分类性能的 RIF 关键基因进行模型构建。在纳入每个特定基因的表达值后,构建每个患者的风险 评分公式,并在 LASSO 回归分析中用其估计的回归系数加权,根据风险评分公式计算每个患者的评分。

2.9. TF-miRNA-Hub 基因调控网络构建

通过 miRWalk (Version3.0, <u>http://mirwalk.umm.uni-heidelberg.de/</u>)预测 Hub 基因相应的 miRNA,并在 Targetscan 和 miRDBand 数据库进行验证。通过 TRRUST (Version3.0, <u>https://www.grnpedia.org/trrust/</u>)数据 库预测 Hub 基因相应的转录因子(Transcriptionfactor, TF)。通过 transmir (<u>https://www.cuilab.cn/</u>)数据库对 TF 和 miRNA 之间的相互作用进行验证。最后通过 cytoscape 对 TF、miRNA 和 hub 基因之间的关系进行 结果可视化。

2.10. 免疫细胞浸润分析

为了更好地了解抑郁症组与对照组之间免疫细胞的情况,采用 CIBERSORT 算法根据免疫细胞相关 基因的表达水平计算不同免疫细胞类型占比,它是表征复杂基因表达谱免疫细胞组成的常用工具。将 22 种浸润免疫细胞的输出结果进行整合,生成免疫细胞成分矩阵进行分析。利用 corrplot 包绘制相关性热 图,可视化 22 种免疫细胞浸润的相关性。利用 ggstatsplot (<u>https://github.com/IndrajeetPatil/ggstatsplot</u>)和 ggplot2 (<u>https://cran.r-project.org/web/packages/ggplot2/index.html</u>)包分析 AKR1C3 与免疫浸润细胞的 Spearman 相关性,并将结果可视化。

3. 结果

3.1. 与疾病相关的工具变量

根据 "2.4 MR 分析方法"项,ukb-d-F5_DEPRESSIO 数据得到 OR 值大于 1 的工具变量 115 个; OR 值小于 1 的工具变量 58 个。ebi-a-GCST003769 数据得到 OR 值大于 1 的工具变量 102 个; OR 值小于 1 的工具变量 103 个。ebi-a-GCST005902 数据得到 OR 值大于 1 的工具变量 160 个; OR 值小于 1 的工具 变量 108 个。

3.2. 差异基因鉴定结果

依据|logFC|>0.2 和 P<0.05 进行差异表达基因筛选,与健康对照组相比,从骨关节炎组中总共获得 1166 个差异表达基因,包括 1125 个上调的差异表达基因和 41 个下调的差异表达基因。

3.3. 筛选核心基因

ukb-d-F5_DEPRESSIO 数据与芯片数据交集得到 5 个表达上调的基因,分别是 GUCY1B2、WDR5B、CTSO、TTC23、TRPC6,与下调的基因无交集。ebi-a-GCST003769 数据与芯片数据交集得到 2 个表达上 调的基因,分别是 WEE1、GUCY1B2,得到 2 个表达下调的基因,分别是 AKR1C3、EVI2A。ebi-a-GCST005902 数据与芯片数据交集得到 6 个表达上调的基因,分别是 OLFM4、LTF、GP5、SERINC2、COL19A1、FEM1C,与下调的基因无交集。合并去除重复得到 14 的基因: GUCY1B2、WDR5B、CTSO、TTC23、TRPC6、WEE1、OLFM4、LTF、GP5、SERINC2、COL19A1、FEM1C、AKR1C3、EVI2A(见 图 1(a))。

3.4. ROC 曲线筛选结果

AUC 曲线下面积大于 >0.5 的基因, 依次是 LTF、OLFM4、GUCY1B2、WDR5B、SERINC2、GP5、AKR1C3、WEE1、TTC23、TRPC6、COL19A1、CTSO、FEM1C (如表 3)。

| Gene | AUC |
|---------|---------------------|
| LTF | 0.605 (0.549~0.661) |
| OLFM4 | 0.602 (0.546~0.654) |
| GUCY1B2 | 0.594 (0.539~0.646) |
| WDR5B | 0.591 (0.538~0.646) |
| SERINC2 | 0.585 (0.528~0.638) |
| GP5 | 0.581 (0.525~0.636) |
| AKR1C3 | 0.579 (0.526~0.635) |
| WEE1 | 0.573 (0.515~0.628) |
| TTC23 | 0.571 (0.512~0.628) |
| TRPC6 | 0.567 (0.511~0.622) |
| COL19A1 | 0.564 (0.510~0.622) |
| CTSO | 0.556 (0.500~0.610) |
| FEM1C | 0.547 (0.491~0.608) |
| EVI2A | 0.443 (0.389~0.500) |

Table 3. ROC diagnostic genes 表 3. ROC 诊断基因

3.5. GO 与 KEGG 结果

3.5.1. GO

BP主要包含(myeloid cell development)髓系细胞发育、(isoprenoid catabolic process)异戊二烯类化合物的分解代谢过程、(diterpenoid biosynthetic process)二萜类化合物的生物合成过程、(polyketide metabolic process)聚酮代谢过程、(aminoglycoside antibiotic metabolic process)氨基糖苷类抗生素代谢过程、(doxorubicin metabolic process)多柔比星代谢过程、(testosterone biosynthetic process)睾酮生物合成过程、(regulation of osteoclast development)对破骨细胞发育的调控、(positive regulation of platelet activation)对血小板激活的正向调控、(regulation of vitamin metabolic process)对维生素代谢过程的调控。CC主要包含(tertiary granule lumen)三级颗粒腔、(specific granule lumen)特异性颗粒腔、(specific granule)特异性颗粒、(tertiary granule lumen)三级颗粒粒、(secretory granule lumen)为泌颗粒腔、(cytoplasmic vesicle lumen)细胞质囊泡腔。MF主要包含(store-operated calcium channel activity)储存操作型钙通道活性、(bile acid binding)胆汁酸结合、(testosterone dehydrogenase [NAD(P)] activity)睾酮脱氢酶[NAD(P)]活性、(alditol: NADP+1-oxidoreductase activity)多元醇: NADP+1-氧化还原酶活性、(inositol 1,4,5 trisphosphate binding)肌醇 1,4,5-三磷酸结合、(NADP-retinol dehydrogenase

activity) NADP-视黄醇脱氢酶活性、(estradiol 17-beta-dehydrogenase [NAD(P)] activity)雌二醇 17-β-脱氢酶 [NAD(P)]活性、(NAD-retinol dehydrogenase activity) NAD-视黄醇脱氢酶活性、(alcohol dehydrogenase (NADP+) activity)醇脱氢酶(NADP+)活性、(aldo-keto reductase (NADP) activity)醛酮还原酶(NADP)活性(见图 1(b))。

3.5.2. KEGG

包含(Folate biosynthesis)叶酸生物合成、(Ovarian steroidogenesis)卵巢类固醇合成(见图 1(c))。



Figure 1. (a) Forest plot of MR-eQTL analysis; (b) Bar plot of GO functional enrichment analysis; (c) Bar plot of KEGG pathway enrichment analysis
图 1. (a) MR-eQTL 森林图; (b) GO 功能富集分析柱状图; (c) KEGG 通路富集分析柱状图

DOI: 10.12677/hjbm.2025.154076

3.6. LASSO 回归

构建 LASSO 回归分析结果显示保存了 5 个非零系数的特征基因 LTF、OLFM4、AKR1C3、WEE1、 EVI2A,其中 EVI2A 的 AUC 值小于 0.5,予剔除(见图 2(a),图 2(b))。

3.7. TF-miRNA-hubgenes 调控网络构建

将 4 个 Hub 基因输入 miRWalk 数据库预测相应的 miRNA, 再通过 mirTarBase、Targetscan、和 miRDB 数据库进行取交集筛选,获得 48 个 miRNA。通过 TRRUST 数据库筛选 Hub 基因相应的 TF,共获得 56 个 TF ((见图 2(c)), LncRNA 用蓝色表示, mRNA 用红色表示, miRNA 用绿色表示)。



Figure 2. (a) Path plot of LASSO regression coefficients; (b) Cross validation curve of LASSO regression; (c) TF-miRNAhub genes regulatory network diagram

图 2. (a) LASSO 回归系数路径图; (b) LASSO 回归交叉验证曲线; (c) TF-miRNA-hubgenes 调控网络图

3.8. 免疫细胞浸润分析

为了进一步探究抑郁症组和对照组的免疫细胞组成,我们利用 CIBERSORT 算法考察了训练集每个 样本中 22 种免疫细胞浸润的比例(图 3(a)),然后绘制了箱线图来比较两组免疫细胞浸润的差异。如图 3(b) 所示,抑郁症样本中激活状态的记忆 CD4+ T 细胞、促炎症的巨噬细胞、激活状态的肥大细胞的比例显 著低于对照组,而抑郁症组中记忆 B 细胞、未极化的巨噬细胞、抗炎症的巨噬细胞的比例显著高于对照 **组。**因此,这些类型的免疫细胞可能是潜在的核心与抑郁症相关的免疫细胞。此外,各免疫细胞之间的 相关性热图显示激活状态的自然杀伤细胞与激活状态的记忆 CD4+ T 细胞(Cor = 0.41)和激活状态的肥大 细胞(Cor = 0.52)呈正相关。激活状态的肥大细胞激活状态的记忆 CD4+ T 细胞(Cor = 0.41)呈正相关。相 反,中性粒细胞与激活状态的记忆 CD4+ T 细胞(Cor = -0.43)、激活状态的自然杀伤细胞(Cor = -0.49)和 T 细胞 CD8(Cor = -0.48) 呈负相关。记忆 B 细胞与激活状态的记忆 CD4+ T 细胞(Cor = -0.44)、未成熟 B 细胞(Cor = -0.58) **呈负相关**(图 3(c))。综合起来,上述发现揭示了抑郁症和对照样本之间免疫细胞浸润特 征的显著差异。然后,我们评估了四个核心基因与免疫细胞浸润的关系。通过相关性分析发现,AKR1C3 与单核细胞(Cor=0.35, p=1.8e-08)、激活状态的自然杀伤细胞(Cor=0.27, p=2.4e-05)、激活状态的记 忆 CD4+T 细胞(Cor=0.26, p=3.2e-05)、未成熟 B 细胞(Cor=0.26, p=3.4e-05)、静止状态的自然杀伤 细胞(Cor = 0.16, p = 0.010) **呈显著正相关。**与记忆 B 细胞(Cor = -0.37, p = 2.2e-09)、中性粒细胞(Cor = -0.18, p = 0.0048)、静止状态的记忆 CD4+ T 细胞(Cor = -0.18, p = 0.0052)、未极化的巨噬细胞(Cor = -0.16, p=0.013)呈显著负相关。我们的结果表明, AKR1C3 与免疫浸润细胞, 尤其是与抑郁症相关的潜 在核心免疫细胞具有统计学上显著的关系。此外,LTF 与静止状态的自然杀伤细胞呈显著正相关,与记 忆 B 细胞、静止状态的记忆 CD4+ T 细胞呈负相关(补充图 1(c))。OLFM4 与中性粒细胞呈显著正相关, 与 T 细胞 CD4 记忆激活呈负相关。WEE1 与巨噬细胞 M2 呈显著正相关,与巨噬细胞 M1、静止状态的 肥大细胞呈负相关。





Figure 3. (a) Stacked bar chart showing the proportions of 22 immune cell types. The height of the colored bars represents the proportion of immune cells; (b) Boxplot depicting differences in immune infiltration between the depression group and the control group. The x-axis indicates the names of the immune cells, while the y-axis indicates the proportion of immune cells. Green represents the control group, and red represents the depression group. A p-value of less than 0.05 indicates a significant difference; (c) Correlation matrix of the proportions of 22 immune cell types. The redder the color, the higher the positive correlation. The bluer the color, the higher the negative correlation; (d) Correlation analysis of AKR1C3 with 22 types of immune-infiltrating cells. The x-axis represents the correlation coefficient, and the y-axis represents the names of the immune cells. The size of the dots represents the absolute value of the correlation coefficient, and the numbers represent the p-values of the correlation tests. A p-value of less than 0.05 indicates a significant difference; (e)~(1) Correlation between AKR1C3 and monocytes, activated natural killer cells, activated memory CD4+ T cells, immature B cells, resting natural killer cells, memory B cells, neutrophils, resting memory, and un-polarized macrophages

图 3. (a) 22 种免疫细胞比例的堆积条形图。彩色柱的高度代表免疫细胞的比例;(b)箱线图描绘了抑郁症组和对照组 之间的免疫浸润差异。横轴表示免疫细胞的名称,纵轴表示免疫细胞的比例。绿色表示对照组,红色表示抑郁症组。 P<0.05 表示有显著差异;(c) 22 种免疫细胞比例的相关矩阵。颜色越红,正相关性越高。颜色越蓝,负相关性越高; (d) AKR1C3 与 22 种免疫浸润细胞的相关性分析。横坐标表示相关系数,纵坐标表示免疫细胞名称。点的大小代表 相关系数的绝对值,数字代表相关性检验的 p 值。P<0.05 表示有显著差异;(e)~(l) AKR1C3 与单核细胞、激活状态 的自然杀伤细胞、激活状态的记忆 CD4+ T 细胞、未成熟 B 细胞、静止状态的自然杀伤细胞、记忆 B 细胞、中性粒 细胞、静止状态的记忆、未极化的巨噬细胞之间的相关性。

4. 讨论

本研究通过 GEO 数据库中的五个基因表达数据集,结合 eOTLGen 联盟的 eOTL 数据,筛选出与抑 郁症相关的差异表达基因。结果显示,乳转铁蛋白(LTF)、OLFM4、AKR1C3 和 WEE1 等基因在抑郁症 患者中表现出显著的差异表达。乳转铁蛋白(Lactoferrin, LTF)是一种具有多种生物学功能的铁结合糖蛋白, 主要存在于乳汁、唾液、眼泪和其他体液中。有研究发现,LTF 缺乏会导致小鼠在成年后表现出抑郁样 行为,这与肠道屏障功能的损害和神经免疫功能的失调密切相关[8]。LTF 不仅能够调节肠道微生物组, 改善肠道屏障功能,还能通过抑制炎症反应和氧化应激来保护神经元[9]。研究表明,哺乳期摄入乳铁蛋 白通过调节小胶质细胞的激活来保护神经元,从而有效地减轻成年人的抑郁症状[10]。OLFM4 是一种含 有 olfactomedin 结构域,在脑组织中特异性表达的蛋白,与多种神经精神疾病相关。一项研究表明,OLFM4 与 LRFN5 的联合使用在诊断重度抑郁症(MDD)方面表现出优越的效果, AUC 值高达 0.974, 显示出其在 临床应用中的潜力[11]。另一项研究发现与健康对照组相比,重度抑郁症(MDD)患者的血清中 OLFM4 水 平显著升高[12]。一项随机临床试验的次要发现,补充益生菌治疗抑郁症,伴随 OLFM4 表达上调,激活 免疫反应[13]。AKR1C3 (Aldo-Keto Reductase Family 1 Member C3)是一种重要的酶,属于醛酮还原酶家 族,主要参与类固醇激素的代谢和药物的生物转化。一项研究通过机器学习算法构建了基于 AKR1C3 等 关键基因的抑郁症诊断模型,显示出良好的预测能力,这进一步支持了 AKR1C3 作为潜在生物标志物的 价值[14]。某些非甾体抗炎药(NSAIDs)如吲哚美辛己被发现能够选择性地抑制 AKR1C3 的活性[15], 而炎 症是诱发抑郁症的症状的主要原因之一。WEE1 基因是一种重要的细胞周期调节基因,主要负责在细胞 分裂过程中调控 G2/M 期的检查点。WEE1 未有与抑郁症相关的研究,其主要与癌症相关,一项研究表 明, 基因和化合物筛选确定了 PP2A 和 WEE1 的联合抑制在多种癌症模型中是协同的, 通过破坏 DNA 复 制和引发细胞死亡后的过早有丝分裂[16]。GO分析显示抑郁症相关基因主要富集在髓系细胞发育、三级 颗粒腔、特异性颗粒腔、储存操作型钙通道活性、胆汁酸结合等生物过程和分子功能。这些结果进一步 揭示了抑郁症的复杂生物学机制,为未来的研究提供了新的方向。一项实验研究表明,在脊髓损伤后, 诱导的脊髓背角和背根神经节髓样细胞活化,与未损伤组和假手术组相比,小鼠表现出伤害感受性和抑 郁样功能障碍[17]。通路富集分析显示,抑郁症相关基因主要富集在叶酸生物合成和卵巢类固醇合成通路。 这些通路为理解抑郁症的生物学基础提供了新的视角,提示叶酸生物合成和卵巢类固醇合成在情绪调节 和神经系统功能中的潜在作用。已知叶酸的衍生物如生物蝶呤和 S-腺苷甲硫氨酸(SAM)的合成或者与抑 郁症的改善有关,或者对抑郁症具有直接的治疗效果[18]。据文献表明补充叶酸可以提高传统抗抑郁药物 的疗效,临床医生可能希望考虑为患有抑郁症或可能有抑郁症状的患者补充叶酸[19]。一项荟萃分析显示, 孕期持续补充叶酸可能降低围产期抑郁症状的发生率[20]。绝经期妇女雌二醇等卵巢类固醇减少,相应对 HPA 轴的抑制性反馈较少,更容易引起抑郁症[21]。一项综述报道,卵巢类固醇和中枢神经递质的周期 性变化之间的复杂相互作用容易引起抑郁的前期症状[22]。本研究使用 LASSO 回归分析构建的预测模型 在抑郁症的分类中达到了 80%的准确率。这一高准确度的模型为临床早期诊断抑郁症提供了新的工具, 展示了机器学习方法在生物标志物发现中的优势。然而,模型的可解释性与临床应用的挑战仍需进一步 研究,以确保其在实际应用中的有效性和可靠性。并通过 CIBERSORT 算法分析抑郁症患者与健康对照 组之间的免疫细胞浸润差异,结果显示抑郁症患者的记忆 B 细胞、未极化的巨噬细胞 M0、抗炎症的巨 噬细胞 M2 密度显著增加,而激活状态的记忆 CD4+T 细胞、促炎症的巨噬细胞、激活状态的肥大细胞的 比例显著降低。这些发现表明免疫系统可能在抑郁症的发病机制中扮演重要角色,提示免疫治疗在抑郁 症管理中的潜在应用。

本研究通过多芯片联合 eQTL 分析和机器学习算法,系统地筛选了与抑郁症相关的生物标志物,并

分析了其与免疫浸润的关系。这些发现不仅为理解抑郁症的病理机制提供了新的视角,也为开发新的诊断工具和治疗靶点奠定了基础。然而,本研究仍存在一些局限性,例如样本量有限、数据来源单一等。 未来的研究应进一步扩大样本量,整合更多数据来源,以验证这些发现的普适性和可靠性。此外,未来的研究还应深入探讨这些生物标志物的生物学功能和作用机制,为抑郁症的精准治疗提供新的策略。

总之,本研究通过综合分析基因表达、eQTL、免疫浸润和机器学习等多种方法,揭示了抑郁症的潜 在病理机制,并筛选出多个具有诊断价值的生物标志物。这些发现为抑郁症的早期诊断和治疗提供了新 的理论依据,也为未来的研究提供了新的方向。

基金项目

广西医疗卫生适宜技术开发与推广应用项目(S2024099): 基于肠 - 靶轴探究盐酸度洛西汀对原发性 抑郁症作用的技术开发与推广应用研究,负责人:何乾超。

参考文献

- Liu, J., Liu, Y., Ma, W., Tong, Y. and Zheng, J. (2024) Temporal and Spatial Trend Analysis of All-Cause Depression Burden Based on Global Burden of Disease (GBD) 2019 Study. *Scientific Reports*, 14, Article No. 12346. https://doi.org/10.1038/s41598-024-62381-9
- [2] Engleman, E.A., Robertson, D.W., Thompson, D.C., Perry, K.W. and Wong, D.T. (1996) Antagonism of Serotonin 5-HT_{1A} Receptors Potentiates the Increases in Extracellular Monoamines Induced by Duloxetine in Rat Hypothalamus. *Journal of Neurochemistry*, 66, 599-603. <u>https://doi.org/10.1046/j.1471-4159.1996.66020599.x</u>
- [3] Zunszain, P.A., Hepgul, N. and Pariante, C.M. (2012) Inflammation and Depression. In: Cowen, P.J., Sharp, T. and Lau, J.Y.F., Eds., *Behavioral Neurobiology of Depression and Its Treatment*, Springer, 135-151. https://doi.org/10.1007/7854_2012_211
- [4] Bertollo, A.G., Santos, C.F., Bagatini, M.D. and Ignácio, Z.M. (2025) Hypothalamus-Pituitary-Adrenal and Gut-Brain Axes in Biological Interaction Pathway of the Depression. *Frontiers in Neuroscience*, 19, Article ID: 1541075. <u>https://doi.org/10.3389/fnins.2025.1541075</u>
- [5] Bürhan-Çavuşoğlu, P., İscan, E., Güneş, A., Atabey, N. and Alkın, T. (2021) Increased Telomerase Activity in Major Depressive Disorder with Melancholic Features: Possible Role of Pro-Inflammatory Cytokines and the Brain-Derived Neurotrophic Factor. *Brain, Behavior, & Immunity-Health*, 14, Article ID: 100259. https://doi.org/10.1016/j.bbih.2021.100259
- [6] Gleason, K.J., Yang, F. and Chen, L.S. (2021) A Robust Two-Sample Transcriptome-Wide Mendelian Randomization Method Integrating GWAS with Multi-Tissue eQTL Summary Statistics. *Genetic Epidemiology*, 45, 353-371.
- [7] Wang, F., Liang, Y. and Wang, Q. (2024) Interpretable Machine Learning-Driven Biomarker Identification and Validation for Alzheimer's Disease. *Scientific Reports*, 14, Article No. 30770. <u>https://doi.org/10.1038/s41598-024-80401-6</u>
- [8] Wang, W., Cheng, Z., Wang, X., et al. (2023) Lactoferrin Deficiency during Lactation Increases the Risk of Depressivelike Behavior in Adult Mice. BMC Biology, 21, Article No. 242.
- [9] Redel, A., Miry, F., Hellemons, M.E., Oswald, L.M.A. and Braunstahl, G.J. (2024) Effect of Lactoferrin Treatment on Symptoms and Physical Performance in Long COVID Patients: A Randomised, Double-Blind, Placebo-Controlled Trial. *ERJ Open Research*, 10, Article ID: 00031-2024. <u>https://doi.org/10.1183/23120541.00031-2024</u>
- [10] Wang, W., An, Q., Zou, Y., Dai, Y., Meng, Q. and Zhang, Y. (2025) Lactoferrin Alleviates the Adverse Effects of Early-Life Inflammation on Depression in Adults by Regulating the Activation of Microglia. *Molecular Medicine*, 31, Article No. 50. <u>https://doi.org/10.1186/s10020-025-01094-9</u>
- [11] Xu, K., Zheng, P., Zhao, S., et al. (2023) LRFN5 and OLFM4 as Novel Potential Biomarkers for Major Depressive Disorder: A Pilot Study. *Translational Psychiatry*, 13, Article No. 188.
- [12] Bouzid, A., Almidani, A., Zubrikhina, M., Kamzanova, A., Ilce, B.Y., Zholdassova, M., et al. (2023) Integrative Bioinformatics and Artificial Intelligence Analyses of Transcriptomics Data Identified Genes Associated with Major Depressive Disorders Including NRG1. Neurobiology of Stress, 26, Article ID: 100555. <u>https://doi.org/10.1016/j.ynstr.2023.100555</u>
- [13] Sempach, L., Doll, J.P.K., Limbach, V., Marzetta, F., Schaub, A., Schneider, E., et al. (2024) Examining Immune-Inflammatory Mechanisms of Probiotic Supplementation in Depression: Secondary Findings from a Randomized Clinical Trial. Translational Psychiatry, 14, Article No. 305. <u>https://doi.org/10.1038/s41398-024-03030-7</u>

- [14] Chen, B., Sun, X., Huang, H., Feng, C., Chen, W. and Wu, D. (2024) An Integrated Machine Learning Framework for Developing and Validating a Diagnostic Model of Major Depressive Disorder Based on Interstitial Cystitis-Related Genes. *Journal of Affective Disorders*, 359, 22-32.
- [15] Lolli, M.L., Carnovale, I.M., Pippione, A.C., et al. (2019) Bioisosteres of Indomethacin as Inhibitors of Aldo-Keto Reductase 1C3. ACS Medicinal Chemistry Letters, 10, 437-443.
- [16] Dias, M. H., Friskes, A., Wang, S., et al. (2024) Paradoxical Activation of Oncogenic Signaling as a Cancer Treatment Strategy. Cancer Discovery, 14, 1276-1301. <u>https://doi.org/10.1158/2159-8290.CD-23-1021</u>
- [17] Richards, J.H., Freeman, D.D. and Detloff, M.R. (2024) Myeloid Cell Association with Spinal Cord Injury-Induced Neuropathic Pain and Depressive-Like Behaviors in LysM-eGFP Mice. *The Journal of Pain*, 25, Article ID: 104433. https://doi.org/10.1016/j.jpain.2023.11.016
- [18] Paul, R.T.P., McDonnell, A.P. and Kelly, C.B. (2004) Folic Acid: Neurochemistry, Metabolism and Relationship to Depression. *Human Psychopharmacology: Clinical and Experimental*, **19**, 477-488. <u>https://doi.org/10.1002/hup.614</u>
- [19] Abdelmaksoud, A., Vojvodic, A., Ayhan, E., Dönmezdil, S., Jovicevic, T.V., Vojvodic, P., et al. (2019) Depression, Isotretinoin, and Folic Acid: A Practical Review. Dermatologic Therapy, 32, e13104. <u>https://doi.org/10.1111/dth.13104</u>
- [20] Jin, X., Cheng, Z., Yu, X., Tao, Q., Huang, R. and Wang, S. (2022) Continuous Supplementation of Folic Acid in Pregnancy and the Risk of Perinatal Depression—A Meta-Analysis. *Journal of Affective Disorders*, **302**, 258-272.
- [21] Young, E.A. and Altemus, M. (2004) Puberty, Ovarian Steroids, and Stress. Annals of the New York Academy of Sciences, 1021, 124-133. <u>https://doi.org/10.1196/annals.1308.013</u>
- [22] Haußmann, J., Goeckenjan, M., Haußmann, R. and Wimberger, P. (2024) Prämenstruelles Syndrom und prämenstruelle dysphorische Störung-Übersicht zu Pathophysiologie, Diagnostik und Therapie [Premenstrual Syndrome and Premenstrual Dysphoric Disorder-Overview on Pathophysiology, Diagnostics and Treatment]. Der Nervenarzt, 95, 268-274.