

分类号: R  
学号: 20222114193

密级: 公开  
单位代码: 10759

# 石河子大学

## 硕士学位论文



### 基于生物信息学的阿尔茨海默病线粒体基因筛选及 ceRNA 网络构建

学位申请人	李晓松
指导教师	王奎 副教授
申请学位类别	专业硕士
专业名称	公共卫生
研究领域	流行病学与卫生统计学
所在学院	医学院

中国·新疆·石河子  
2025年5月

分类号: R  
学号: 20222114193

密级: 公开  
单位代码: 10759

# 石河子大学

## 硕士学位论文



### 基于生物信息学的阿尔茨海默病线粒体基因筛选及 ceRNA 网络构建

学位申请人	李晓松
指导教师	王奎 副教授
申请学位类别	专业硕士
专业名称	公共卫生
研究领域	流行病学与卫生统计学
所在学院	医学院

中国·新疆·石河子

2025年5月

**Screening of mitochondrial genes and construction of ceRNA network  
for Alzheimer's disease based on bioinformatics**

A Dissertation Submitted to

**Shihezi University**

In Partial Fulfillment of the Requirements

for the Degree of

**Master of Public Health**

**By**

**Li Xiao-song**

**(Epidemiology and Health Statistics)**

Dissertation Supervisor: A/Prof. Wang Kui

May, 2025

# 石河子大学学位论文独创性声明及使用授权声明

## 学位论文独创性声明

本人所提交的学位论文是在我导师的指导下进行的研究工作及取得的研究成果。据我所知，除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确的说明并表示谢意。

研究生签名：李曉松

时间：2025年5月20日

## 使用授权声明

本人完全了解石河子大学有关保留、使用学位论文的规定，学校有权保留学位论文并向国家主管部门或指定机构送交论文的电子版和纸质版。有权将学位论文在学校图书馆保存并允许被查阅。有权自行或许可他人将学位论文编入有关数据库提供检索服务。有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

研究生签名：李曉松

时间：2025年5月20日

导师签名：李松

时间：2025年5月20日

## 摘要

**目的：**本研究旨在利用生物信息学方法挖掘阿尔茨海默病（AD）的线粒体相关生物标志物，并构建竞争性内源 RNA（ceRNA）调控网络，探讨线粒体功能障碍在 AD 中的作用机制。

**方法：**从 GEO 数据库获取 6 组 AD 相关芯片数据集（GSE109887、GSE132903、GSE118553、GSE63063、GSE120584 和 GSE182910），结合 MitoCarta3.0 线粒体基因集，通过差异表达分析、加权基因共表达网络分析筛选 AD 相关线粒体基因。利用 DAVID 数据库进行功能富集分析，STRING 数据库构建蛋白质互作网络，并通过 Cytoscape 软件筛选枢纽基因。进一步通过受试者工作特征曲线（ROC）筛选对 AD 具有一定鉴别能力的基因，并利用 PSPHunter 算法预测蛋白相分离能力。通过免疫浸润分析和孟德尔随机化（MR）分析，揭示免疫微环境与 AD 的关系，并基于 DsigDB 数据库筛选潜在靶向药物。针对 lncRNA、miRNA 和 mRNA 数据集进行差异表达分析，对线粒体相关差异表达 mRNA 进行富集分析，进一步利用 LASSO 回归筛选核心 mRNA，结合 ENCORI 数据库预测 miRNA 的 mRNA 与 lncRNA 互作靶标。最后，对预测结果进行整合分析，利用 Cytoscape 软件构建 ceRNA 网络并筛选关键分子对。

**结果：**筛选出 6 个 AD 线粒体相关基因，且各基因之间呈现出较强的相关性。经 ROC 曲线分析，VDAC1（AUC=0.655）、TOMM20（AUC=0.644）、CISD1（AUC=0.634）和 TSPO（AUC=0.640）在独立数据集验证中具有一定的鉴别能力，且 VDAC1（Prob=0.797）表现出较高的液-液相分离倾向；免疫浸润分析显示，枢纽基因与活化态树突状细胞和 M1 巨噬细胞具有一定的相关性，MR 分析揭示了 37 种免疫细胞表型与 AD 的潜在因果关联（ $P < 0.05$ ），尤其是多种树突状细胞表型（CD62L+ 树突状细胞、CD62L-浆细胞样树突状细胞、CD86+髓系树突状细胞）与 AD 存在因果关联（ $P < 0.05$ ）；DsigDB 数据库鉴定出 7 种与线粒体枢纽基因显著相关的化合物。在 ceRNA 方面，通过差异表达分析筛选出 1964 个 lncRNA、667 个 miRNA 和 59 个线粒体相关 mRNA，进一步利用 LASSO 回归筛选出 10 个核心 mRNA。最后，基于 ENCORI 数据库构建了包含 11 个 lncRNA、7 个线粒体相关 mRNA 与 16 个 miRNA 共 32 个 lncRNA-miRNA-mRNA 关系互作对，其中 XIST-has-miR-30d-5p-NDUFS5 和 XIST-has-miR-214-3p-NDUFS5 为核心调控对，可能通过 miRNA 调控 NDUFS5 的表达影响线粒体功能，进而影响 AD 的发生发展。

**结论：**研究最终鉴定出 4 个线粒体枢纽基因（VDAC1、TOMM20、CISD1、TSPO）可能作为潜在的生物标志物参与了 AD 的病理发生与疾病进展，VDAC1 相分离特性为 AD 病理机制提供了新视角，且枢纽基因可能通过免疫微环境影响 AD 的病理进程，ceRNA 网络关键分子对揭示了非编码 RNA 调控线粒体功能的潜在路径，为 AD 的分子病理机制及靶向干预策略提供了新型候选调控轴。

**关键词：**生物信息学；阿尔茨海默病；线粒体；枢纽基因；竞争性内源 RNA

## Abstract

**Objective:** This study aims to explore the mitochondrial related biomarkers of Alzheimer's disease (AD) using bioinformatics methods, and construct a competitive endogenous RNA (ceRNA) regulatory network to investigate the mechanism of mitochondrial dysfunction in AD.

**Method:** Six sets of AD related chip datasets (GSE109887, GSE132903, GSE118553, GSE63063, GSE120584, and GSE182910) were obtained from the GEO database. Combined with the MitoCarta 3.0 mitochondrial gene set, AD related mitochondrial genes were screened through differential expression analysis and weighted gene co expression network analysis. Perform functional enrichment analysis using DAVID database, construct protein-protein interaction network using STRING database, and screen hub genes using Cytoscape software. Further screening of genes with high discriminatory ability for AD through receiver operating characteristic (ROC) curves, and prediction of protein phase separation ability using PSPHunter algorithm. Through immune infiltration analysis and Mendelian randomization (MR) analysis, the relationship between immune microenvironment and AD was revealed, and potential targeted drugs were screened based on the DsigDB database. Perform differential expression analysis on lncRNA, miRNA, and mRNA datasets, enrich mitochondrial related differentially expressed mRNA, further screen core mRNA using LASSO regression, and predict miRNA mRNA lncRNA interaction targets in combination with ENCORI database. Finally, the predicted results were integrated and analyzed, and a ceRNA network was constructed using Cytoscape software to screen for key molecular pairs.

**Result:** Six AD mitochondrial related genes were screened, and there was a strong correlation between each gene. According to ROC curve analysis, VDAC1 (AUC=0.655), TOMM20 (AUC=0.644), CISD1 (AUC=0.634), and TSPO (AUC=0.640) have certain discriminative ability in independent dataset validation, and VDAC1 (Prob=0.797) exhibits a high tendency towards liquid-liquid phase separation; Immune infiltration analysis showed a certain correlation between hub genes and activated dendritic cells and M1 macrophages. MR analysis revealed a potential causal association between 37 immune cell phenotypes and AD ( $P<0.05$ ), especially multiple dendritic cell phenotypes (CD62L+dendritic cells, CD62L- plasma cell like dendritic cells, CD86+myeloid dendritic cells) with AD ( $P<0.05$ ); The DsigDB database identified 7 compounds significantly associated with mitochondrial hub genes. In terms of ceRNA, 1964 lncRNAs, 667 miRNAs, and 59 mitochondrial related mRNAs were screened through differential expression analysis, and 10 core mRNAs were further screened using LASSO regression. Finally, based on the ENCORI database, a total of 32 lncRNA-miRNA-mRNA interactions were constructed, including 11 lncRNAs, 7 mitochondrial related mRNAs, and 16 miRNAs.

XIST-has-miR-30d-5p-NDUFS5 and XIST-has-miR-214-3p-NDUFS5 were identified as the core regulatory pairs, which may affect mitochondrial function and subsequently impact the occurrence and development of AD.

**Conclusion:** The study ultimately identified four mitochondrial hub genes (VDAC1, TOMM20, CISD1, TSPO) as potential biomarkers involved in the pathological occurrence and disease progression of AD. The phase separation characteristics of VDAC1 provide a new perspective on the pathological mechanism of AD, and hub genes may affect the pathological process of AD through the immune microenvironment. Key molecules in the ceRNA network reveal potential pathways for non coding RNA regulation of mitochondrial function, providing novel candidate regulatory axes for the molecular pathological mechanism and targeted intervention strategies of AD.

**Key words:** Bioinformatics; Alzheimer's disease; Mitochondria; Hub genes; ceRNA

# 目录

摘要.....	I
Abstract.....	I
目录.....	IV
缩略词表.....	VI
第 1 章 前言.....	1
第 2 章 材料和方法.....	5
2.1 数据来源.....	5
2.1.1 基因芯片数据平台.....	5
2.1.2 KEGG 和 GO 数据库.....	5
2.1.3 STRING 数据库.....	6
2.1.4 DsigDB 数据库.....	6
2.1.5 ENCORI 数据库.....	6
2.1.6 数据下载.....	6
2.2 线粒体相关枢纽基因的筛选.....	7
2.2.1 数据预处理.....	7
2.2.2 差异表达基因的筛选.....	7
2.2.3 加权基因共表达网络分析.....	8
2.2.4 线粒体相关候选基因筛选.....	9
2.2.5 候选基因功能富集分析.....	9
2.2.6 PPI 网络分析及枢纽基因的筛选.....	9
2.2.7 独立数据集验证.....	9
2.2.8 蛋白相分离能力预测.....	10
2.2.9 免疫浸润分析.....	10
2.2.10 免疫细胞的 MR 分析.....	10
2.2.11 DsigDB 数据库药物筛选.....	11
2.3 ceRNA 调控网络构建.....	11
2.3.1 数据处理及差异表达分析.....	11
2.3.2 线粒体相关 mRNA 筛选.....	12
2.3.3 GO 和 KEGG 富集分析.....	12
2.3.4 LASSO 回归筛选 hub-DEmRNAs.....	12

2.3.5	制作 ceRNA 网络关系对文件 .....	12
2.3.6	ceRNA 网络可视化及关键分子对筛选 .....	13
2.4	技术路线图 .....	13
第 3 章	结果 .....	14
3.1	线粒体相关枢纽基因的筛选 .....	14
3.1.1	筛选 AD 的差异表达基因 .....	14
3.1.2	离群样本鉴定 .....	14
3.1.3	软阈值选择 .....	15
3.1.4	基因模块识别 .....	16
3.1.5	关键模块的识别与分析 .....	16
3.1.6	候选基因的筛选 .....	17
3.1.7	候选基因富集分析 .....	18
3.1.8	PPI 网络分析与枢纽基因的鉴定 .....	18
3.1.9	枢纽基因的验证 .....	20
3.1.10	蛋白相分离能力预测 .....	21
3.1.11	免疫细胞浸润及相关性分析 .....	21
3.1.12	免疫细胞与 AD 的 MR 分析 .....	23
3.1.13	DsigDB 数据库药物筛选 .....	30
3.2	ceRNA 调控网络构建 .....	30
3.2.1	差异表达分析 .....	30
3.2.2	线粒体相关 mRNA 筛选 .....	31
3.2.3	线粒体相关 mRNA 富集分析 .....	32
3.2.4	hub-DEmRNAs 的鉴定 .....	32
3.2.5	ceRNA 网络构建 .....	33
第 4 章	讨论 .....	37
4.1	线粒体相关基因和 AD .....	37
4.2	ceRNA 网络和 AD .....	40
4.3	优势与不足 .....	41
第 5 章	结论 .....	43
文献综述	.....	44
参考文献	.....	49
致谢	.....	57
作者简介	.....	58
石河子大学硕士研究生学位论文导师评阅表	.....	59

## 缩略词表

缩略词	英文全称	中文全称
AD	Alzheimer' s disease	阿尔茨海默病
A $\beta$	Amyloid $\beta$ -peptide	$\beta$ 淀粉样蛋白
BP	Biological Process	生物过程
CC	Cellular component	细胞组成
ceRNA	Competing endogenous RNA	竞争性内源 RNA
DElncRNAs	Differentially expressed lncRNAs	差异表达 lncRNAs
DEmiRNAs	Differentially expressed miRNAs	差异表达 miRNAs
DEmRNAs	Differentially expressed mRNAs	差异表达 mRNAs
GEO	Gene Expression Omnibus	基因表达综合数据库
GO	Gene ontology	基因本体
hub-DEmRNAs	Differentially expressed hub mRNAs	差异表达的枢纽 mRNAs
KEGG	Kyoto Encyclopedia of Genes and Genomes	京都基因组百科全书
LLPS	Liquid-liquid phase separation	液-液相分离
lncRNA	Long non-coding RNA	长链非编码 RNA
Log <sub>2</sub> FC	Log <sub>2</sub> fold-change	样本间差异表达倍数
MF	Molecular Function	分子功能
miRNA	MicroRNA	微小 RNA
MR	Mendelian randomization	孟德尔随机化
mRNA	Messenger RNA	信使 RNA
ncRNA	Non-coding RNA	非编码 RNA
NFTs	Neurofibrillary Tangles	神经原纤维缠结
P.adj	FDR-adjusted p-value	校正后的 P 值
PPI	Protein-Protein Interaction	蛋白质-蛋白质相互作用
ROS	Reactive Oxygen Species	活性氧
WGCNA	Weighted Gene Co-Expression Network Analysis	加权基因共表达网络分析

## 第 1 章 前言

阿尔茨海默病 (Alzheimer's disease, AD) 又称老年性认知障碍症, 是一种起病隐匿、病因未明的原发性、进行性发展的中枢神经系统退行性疾病。该疾病以多维度认知功能损伤为主要特征, 症状包括记忆系统衰退、语言功能异常 (如表达与理解障碍)、空间辨识能力缺失、执行功能受损以及人格特质改变等临床表现, 患者除呈现日常活动功能逐渐衰退外, 常并发神经精神系统症状及行为异常<sup>[1]</sup>。作为典型的年龄相关性疾患, 其病程呈不可逆性恶化趋势, 多数患者在确诊后 10-20 年间将逐步丧失自主生活能力, 最终因多系统并发症导致生命体征衰竭<sup>[2]</sup>。根据《World Alzheimer Report 2023》流行病学数据显示<sup>[3]</sup>, 全球痴呆症患者总数已突破 5500 万, 其中 AD 占比达 60-70% (约 3300 万-3850 万)。有研究指出, 随着人口老龄化进程加速及代谢综合征 (糖尿病、肥胖等) 风险因素剧增, 预测至 2050 年全球痴呆症患者规模将激增至 1.57 亿, 且 AD 仍为该疾病谱系的主体构成<sup>[4]</sup>。据统计, 我国作为人口老龄化速度最快的国家之一, 现存 AD 患者约 1000 万, 预计到 2050 年患者数将攀升至 4000 万以上<sup>[5]</sup>。同时, AD 造成的经济负担也呈指数级增长, 我国 2015 年度 AD 相关医疗支出已达 1677.4 亿美元, 占当年 GDP 的 1.47%, 显著高于全球平均水平 (1.09%)。单个 AD 患者年均治疗成本达 19144 美元, 这一数值不仅超出恶性肿瘤与心血管疾病的治疗支出, 更是高于全球平均水平 (15122-17483 美元)<sup>[6]</sup>。综上所述, AD 不仅给患者及家庭造成沉重的身心负担与经济压力, 更通过持续增长的医疗资源消耗深刻影响着社会经济运行。随着我国人口老龄化进程不断加快, AD 已逐渐发展为一项严峻的公共卫生议题, 迫切需要建立精准的早期诊断体系和探索有效的治疗方法。

AD 的发病过程具有渐进性特征, 其早期症状并不明显, 且具有个体特异性, 多数病例最初以近事记忆减退为主要表现, 随着病程进展, 患者将逐渐出现包括近期记忆功能衰退、执行功能障碍、异常行为模式及社交能力退化等复合型神经退行性症状<sup>[1]</sup>。然而, 当前医学领域尚未开发出能够有效逆转病程的特异性治疗方法, 这使得针对该疾病病理机制的深入研究, 以及建立精准的早期诊断体系和实施及时干预策略, 成为 AD 亟待突破的重要课题。

尽管众多学者对 AD 进行了深入研究, 但其发病机制仍未完全阐明。目前, AD 的典型病理特征包括海马区细胞外 $\beta$ 淀粉样蛋白 (Amyloid  $\beta$ -peptide, A $\beta$ ) 异常沉积引发的老年斑 (Senile Plaques, SP), 同时神经元内 tau 蛋白过度磷酸化形成的神经原纤维缠结 (Neurofibrillary Tangles, NFTs)<sup>[7]</sup>。基于上述病理标志, 目前形成两大核心理论框架——A $\beta$ 级联假说与 tau 蛋白异常磷酸化假说。在 A $\beta$ 假说中, 淀粉样前体蛋白

(Amyloid Precursor Protein, APP) 经异常酶切产生的 A $\beta$  多肽, 因代谢失衡导致其在中枢神经系统过度积累。这种异常沉积不仅形成具有特征性的淀粉样斑块, 其寡聚体形态还可通过激活胶质细胞、诱发氧化应激等机制产生神经毒性效应<sup>[8]</sup>。与之形成互补的 tau 蛋白学说则强调, 生理状态下通过结合微管维持轴突稳定性的 tau 蛋白, 在发生异常磷酸化、乙酰化等修饰后, 会丧失其生物学功能并聚集成 NFTs, 最终导致微管系统崩解和轴突运输障碍, 引发神经元退行性病变<sup>[9]</sup>。此外, 免疫细胞在 AD 的发病机制中也扮演着重要角色: 一方面通过 A $\beta$  和维持神经稳态发挥保护作用, 另一方面因慢性激活引发神经炎症和神经元损伤。研究指出, 中枢神经系统的小胶质细胞是核心免疫调控者, 其功能失调和促炎表型极化加剧 A $\beta$  斑块沉积并释放促炎因子, 驱动神经退行性变<sup>[10]</sup>。然而, 临床病理学研究显示, 部分认知功能正常个体的脑组织中也存在 A $\beta$  沉积现象, 且其浓度与痴呆严重程度无显著相关性; 同时, 针对 A $\beta$  清除和 tau 蛋白磷酸化抑制的临床试验均未取得预期疗效<sup>[11]</sup>。因此, 我们迫切需要从其他病理学角度探索治疗和改善 AD 的新方法。最新研究将关注重点转向线粒体功能障碍这一早期事件。证据表明, 在 AD 病理进程中, 线粒体结构异常和能量代谢障碍的出现早于 A $\beta$  沉积与 NFTs 的形成<sup>[12]</sup>。线粒体作为神经元能量代谢的核心细胞器, 其功能障碍可能通过氧化损伤、自噬异常等机制, 在 AD 发病过程中发挥始动作用<sup>[13]</sup>。这些发现为突破传统病理框架、开发新型治疗方法提供了重要研究方向。

线粒体作为真核细胞的能量代谢枢纽, 其功能障碍已被证实是神经退行性疾病尤其是 AD 的核心病理机制。早在 2004 年, Swerdlow 和 Khan 就提出了线粒体级联假说<sup>[14]</sup>, 该假说质疑了 A $\beta$  和 tau 蛋白是 AD 直接驱动力的观点, 而是强调了氧化应激诱导的线粒体功能障碍在 AD 中的关键作用。该理论假说指出, 由遗传易感性决定的基础线粒体效能, 在与环境诱因共同作用下, 通过级联反应推动 AD 发展, 当线粒体功能损伤突破生理代偿阈值时, 将导致 A $\beta$  异常累积、tau 蛋白过度磷酸化等特征性改变, 而这些病理产物本质上是能量代谢崩溃的继发性产物而非疾病始动因子。有研究显示, 在三重转基因 AD 小鼠模型的病理早期, 皮质梨状区与岛状区即出现葡萄糖代谢率下降和线粒体复合体活性减弱, 且这些改变早于 A $\beta$  斑块沉积和 NFTs 的出现, 有力佐证了线粒体功能衰竭在 AD 早期病理发展中的时序特征<sup>[15]</sup>。因此, 线粒体功能障碍可能是 AD 发生进程中最早出现的分子病理特征, 通过调控线粒体稳态有望为开发新型诊断标志物和治疗策略提供新方向。

随着高通量测序技术的革新, 基因表达谱研究已进入大数据驱动时代, 全球科研机构持续积累海量的多维度组学数据。为促进数据共享与知识挖掘, 国际学术界已构建多个权威开放数据库体系: 美国国立生物技术信息中心 (National Center for Biotechnology Information, NCBI) 的基因表达综合数据库 (Gene Expression Omnibus, GEO)<sup>[16]</sup> 与欧洲生物信息研究所的 ArrayExpress 构成了基因表达数据存储的国际核心平台<sup>[17]</sup>; 而美国国

家癌症研究所主导的癌症基因组图谱 (The Cancer Genome Atlas, TCGA)<sup>[18]</sup>则专注于整合恶性肿瘤多组学数据。这些开放共享的生物信息数据库通过整合海量研究数据并进行深度挖掘,不仅构建了生物信息学的理论方法体系,更驱动了生命科学研究的革新,为精准医学和系统生物学等前沿领域提供了关键数据支撑。

近年来,组学技术不断的迭代革新(如单细胞测序、空间转录组等),生物医学研究已步入大数据爆发时代。面对指数级增长的复杂组学数据,传统统计分析模型在解析系统生物学特征方面显现出明显局限。在此背景下,基于复杂网络理论的高通量数据分析方法迅速发展,其中加权基因共表达网络分析 (Weighted Gene Co-Expression Network Analysis, WGCNA) 因其独特的系统生物学视角成为关键工具<sup>[19]</sup>。该方法创新性地采用软阈值策略构建无尺度拓扑结构,通过优化模块划分算法捕获基因表达的动态关联特征——不仅识别具有生物学意义的共表达模块,更能通过模块特征向量与表型参数的深度关联分析,揭示从分子互作网络到宏观表型输出的调控路径。相较于传统硬阈值方法, WGCNA 更能反映基因调控网络的非线性特征,在疾病关键模块识别、驱动基因预测及治疗靶点发现等领域展现出显著优势,为转化医学研究提供了强有力的系统生物学分析框架。

内源竞争 RNA (competing endogenous RNA, ceRNA) 调控网络是基于共享 microRNA 应答元件 (microRNA response elements, MRE) 的分子互作体系,其核心机制在于不同 RNA 分子通过竞争性结合有限的微小 RNA (microRNA, miRNA) 池,动态调控靶基因表达水平<sup>[20]</sup>。该系统并非由特定 RNA 类别构成,而是涵盖信使 RNA (messenger RNA, mRNA)、长链非编码 RNA (Long non-coding RNA, lncRNA)、环状 RNA (circular RNA, circRNA) 等多种 RNA 分子的竞争性内源调控网络,通过 MRE 介导形成多层次调控轴。在病理层面, ceRNA 网络失调已被证实参与肿瘤增殖转移、心血管重构、代谢紊乱及神经退行性疾病等复杂病理过程<sup>[21]</sup>。近年来,越来越多的研究表明, ceRNA 调控网络在 AD 中发挥重要作用,其通过调控 A $\beta$ 异常沉积和 tau 蛋白磷酸化等关键病理进程,深入影响着 AD 的核心致病机制<sup>[22]</sup>。研究发现, lncRNA BDNF-AS 可通过吸附 miR-9-5p 解除其对 BACE1 mRNA 的抑制作用,进而促进 APP 异常切割;而干预该 ceRNA 轴能显著改善 AD 模型认知功能<sup>[23]</sup>。同时, circRNA FDMC 被证实通过构建 PINK1/Parkin 通路相关 ceRNA 网络,激活线粒体自噬途径加速 A $\beta$ 清除<sup>[24]</sup>。这些发现不仅阐明非编码 RNA (Non-coding RNA, ncRNA) 通过 ceRNA 机制调控 AD 核心病理的分子模式,更提示该网络具有时空特异性调控特征,特定脑区或病理阶段的 ceRNA 互作可能形成差异化调控节点。当前研究趋势正从单一 ceRNA 鉴定转向多组学整合的网络医学分析,通过构建动态 ceRNA-蛋白互作图谱,系统解析 AD 进程中 RNA 调控网络的层级演化规律。这种系统生物学研究范式为开发基于网络干预的精准治疗策略提供了新视角。

近年研究表明, 线粒体功能相关基因的表达受到 ceRNA 网络的动态调控, 而此类调控失衡可能通过加剧线粒体生物能衰竭、氧化应激及自噬障碍等机制, 推动 AD 病理进程<sup>[25]</sup>。因此, 线粒体功能障碍与 ceRNA 调控网络在 AD 病理进程中存在密切的交互作用。研究发现, circRNA FDMC 通过吸附 miR-188-3p, 解除其对 PINK1 基因的抑制作用, 激活 PINK1/Parkin 通路依赖的线粒体自噬<sup>[24]</sup>。在 AD 模型中, 该 ceRNA 轴的激活可促进受损线粒体的清除, 减少 A $\beta$  诱导的神经元毒性。这一机制揭示了 circRNA 通过调控线粒体质量控制参与了 AD 早期病理。Huang G 等<sup>[26]</sup>发现 MALAT1 通过吸附 miR-146a 促进 TLR4/NF- $\kappa$ B/MAPK 通路活化, 抑制线粒体自噬, 并加重心肌损伤, 敲低 MALAT1 可解除对 miR-146a 的抑制, 改善自噬和稳定线粒体膜电位并抑制线粒体活性氧 (Reactive Oxygen Species, ROS) 的产生。AD 患者脑组织中 MALAT1 表达显著下调, 而过表达的 MALAT1 通过吸附 miR-30b 上调 CNR1, 激活 PI3K/AKT 信号通路, 减少神经元损伤、炎症反应和细胞凋亡, 并促进神经恢复<sup>[27]</sup>。Zhou B 等发现, 在 A $\beta$ 1-42 诱导的 AD 模型中, ANRIL 通过吸附 miR-125a, 间接调控炎症通路和神经突生长相关蛋白, 敲除 ANRIL 可解除对 miR-125a 的抑制, 从而抑制凋亡、减轻炎症反应并促进神经元修复<sup>[28]</sup>。

随着 AD 病理机制研究的不断深入, 传统 A $\beta$  级联假说与 tau 蛋白异常磷酸化理论在解释疾病异质性和治疗耐药性方面面临挑战, 促使研究视角向多系统交互作用拓展。且 AD 的病理机制呈现多层次调控特征, 其核心在于由数千个基因构成的高度动态化互作系统。传统单基因研究范式在解析其分子机制时存在明显局限, 难以捕捉多靶点协同作用的生物学本质。在此背景下, 本研究旨在利用 WGCNA 筛选出 AD 中的线粒体相关基因, 结合多种生物信息学方法探讨相关基因在 AD 中的病理调控机制, 并构建 ceRNA 调控网络, 挖掘出线粒体相关 ceRNA 组合作为 AD 早期诊断及疗效评估的新型生物标志物, 为 AD 的精准医学研究提供新方向。

## 第 2 章 材料和方法

### 2.1 数据来源

#### 2.1.1 基因芯片数据平台

基因芯片作为高通量检测技术,通过将数千个单链 DNA 探针有序固定于载体表面,基于碱基互补配对原理实现核酸杂交分析,可同步检测 RNA 表达谱与 DNA 结构信息。其核心优势在于大规模并行检测能力,目前广泛应用于基因功能研究、疾病诊断、药物筛选及基因组学等领域。

GEO 数据库是由美国 NCBI 建立的全球权威公共基因数据存储库,自 2000 年上线以来已成为生命科学领域不可或缺的基础设施<sup>[16]</sup>。作为基因芯片数据存储的核心平台,GEO 数据库采用分层架构管理多维度生物信息:基因表达综合数据库样本(Gene Expression Omnibus Sample, GSM)描述单个样本的生物学特征与实验数据,基因表达综合数据库平台(Gene Expression Omnibus Platform, GPL)提供芯片探针的标准化注释,基因表达综合数据库系列(Gene Expression Omnibus Series, GSE)整合实验元数据及样本间的关联分析,而基因表达综合数据库数据集(Gene Expression Omnibus Dataset, GDS)则支持跨研究数据集的深度挖掘。研究人员可通过 GEO 直接获取原始数据、标准化矩阵及配套分析工具,显著提升基因表达谱研究的可重复性和跨平台整合效率,为生物医学研究提供关键数据支撑。

#### 2.1.2 KEGG 和 GO 数据库

京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)数据库<sup>[29]</sup>是由日本京都大学开发的全球权威生物通路资源库,致力于从系统生物学视角解析基因、代谢产物及其参与的生物通路网络。该数据库通过整合基因组学、蛋白质组学和代谢组学等多层次数据,构建了涵盖生物分子相互作用、代谢反应及信号转导路径的完整知识体系。不仅实现了对基因功能通路的可视化呈现,更为解析大规模基因集合在特定生物通路中的协同作用机制提供了强大的平台支持。

基因本体论(Gene ontology, GO)数据库<sup>[30]</sup>是由国际联盟维护的基因功能注释体系,采用层次化架构从三个层面解析基因功能:生物学过程(biological process, BP)揭示基因参与的时序性生命活动,分子功能(molecular function, MF)阐明基因产物的生化活性,细胞组分(cellular component, CC)定位基因产物的亚细胞分布。通过有效整

合这三个层面的功能注释信息,研究者不仅能精准定位基因产物在细胞内的空间分布特征,更能系统揭示基因群在复杂生物学网络中的功能模块及其动态调控关系。

### 2.1.3 STRING 数据库

STRING (Search Tool for the Retrieval of Interacting Genes) 数据库<sup>[31]</sup>作为蛋白质互作网络研究的核心平台,通过整合实验验证数据、共表达谱分析、文献文本挖掘、基因组共定位预测及系统进化关联等证据,构建了覆盖全物种的蛋白质功能互作图谱。该数据库采用机器学习算法对海量生物信息进行置信度评估,赋予蛋白质相互作用以定量化评分,支持研究者精准识别高可信度的核心功能模块。其可视化分析界面可动态展示蛋白质在代谢通路、信号转导及调控网络中的拓扑关系,同时提供与 KEGG、GO 等数据库的深度关联注释,是生物信息学研究不可或缺的工具。

### 2.1.4 DsigDB 数据库

药物靶点基因特征数据库 (Drug Signatures Database, DsigDB) 作为药物基因组学领域的特色数据库<sup>[32]</sup>,专注于药物干预下的基因表达特征与分子调控网络分析。该数据库整合了来自高通量筛选的实验、临床研究及药物组学数据,涵盖小分子化合物、天然产物及临床药物等数万方种生物活性物质的多维度特征信息。通过构建药物-靶点-通路的三维关联网络, DsigDB 提供基于模式匹配的化合物功能预测工具,支持药物重定位研究中的分子机制假说验证。其特色分析模块整合了富集分析、反向对接算法及网络药理学模型,可精准识别与特定疾病表型显著相关的候选药物和协同组合。

### 2.1.5 ENCORI 数据库

RNA 相互作用 (Encyclopedia of RNA Interactions, ENCORI) 数据库<sup>[33]</sup>是专注于 RNA 分子互作网络与调控机制解析的多维组学平台。该数据库系统整合了来自大规模 CLIP-seq、降解组测序、RNA-RNA 相互作用组及临床转录组等多维度数据,构建了覆盖 miRNA、lncRNA、circRNA、mRNA 及蛋白质间相互作用的全景式调控图谱。其核心算法融合了交叉关联实验验证数据与机器学习预测模型,可精准识别 miRNA-靶基因结合位点、RNA 结合蛋白调控元件及 ceRNA 网络拓扑结构。

### 2.1.6 数据下载

本研究数据来源于 NCBI 的 GEO 数据库 (<https://www.ncbi.nlm.nih.gov/geo>)<sup>[16]</sup>,通过关键词 “Alzheimer's disease” 系统检索并筛选符合标准的微阵列数据集。最终纳入分析的数据集包括 GSE109887、GSE132903、GSE118553、GSE63063、GSE120584 和