

## 基于生物信息学、机器学习与分子模拟研究槟榔碱对心肌肥厚的作用机制

丁志强<sup>1,2</sup>, 何星延<sup>1,2</sup>, 孙治粤<sup>1,2</sup>, 黄云龙<sup>1,2</sup>, 杨少敏<sup>1</sup>, 方红城<sup>1</sup>, 李志芳<sup>1\*</sup>

1. 深圳市中西医结合医院, 广东 深圳 518104

2. 广州中医药大学 深圳中西医结合临床医学院, 广东 深圳 518104

**摘要:** **目的** 基于生物信息学、机器学习与分子模拟研究槟榔碱对心肌肥厚的作用机制。**方法** 基于 GEO 数据库中与我心肌肥厚相关的转录组数据集 GSE249925 进行差异表达分析。采用 3 种机器学习算法 [最小绝对收缩与选择算子 (LASSO) 回归、随机森林 (RF)、支持向量机-递归特征消除 (SVM-RFE)] 筛选关键候选基因。通过基因本体论 (GO) 和京都基因和基因组百科全书 (KEGG) 富集分析探讨其潜在的生物学功能和信号通路。采用分子对接和分子动力学模拟评估槟榔碱与靶蛋白的结合特性和结构稳定性。**结果** 共识别出 28 个可能参与槟榔碱诱导心肌肥厚的候选基因, 经过 3 种机器学习算法联合筛选, 最终确定 4 个核心靶点 (*STAT3*、*SFRP4*、*S100A12*、*MT1A*)。这些基因在心肌肥厚组织中表达显著异常, 并具有良好的诊断效能。富集分析显示其涉及多个关键通路, 包括 Janus 激酶信号转导/转录激活因子 (JAK/STAT) 信号通路、缺氧诱导因子 1 (HIF-1) 信号通路、晚期糖基化终产物-其受体 (AGE-RAGE) 信号通路、以及磷脂酰肌醇 3-激酶/蛋白激酶 B (PI3K/Akt)。分子对接及动力学模拟结果表明, 槟榔碱可与 *STAT3* 和 *S100A12* 稳定结合。**结论** 阐明了槟榔碱诱导心肌肥厚的潜在分子机制, 识别出多个关键调控靶点及信号通路, 为进一步揭示槟榔毒性作用机制以及开发靶向干预策略提供了理论依据与计算支撑。

**关键词:** 槟榔碱; 心肌肥厚; 生物信息学; 机器学习; 最小绝对收缩与选择算子; 随机森林; 支持向量机-递归特征消除; Janus 激酶信号转导/转录激活因子信号通路

中图分类号: R996.2 文献标志码: A 文章编号: 1674-5515(2026)03-0596-10

DOI: 10.7501/j.issn.1674-5515.2026.03.006

## Mechanistic insights into arecoline in cardiac hypertrophy based on bioinformatics, machine learning, and molecular simulations

DING Zhiqiang<sup>1,2</sup>, HE Xingyan<sup>1,2</sup>, SUN Zhiyue<sup>1,2</sup>, HUANG Yunlong<sup>1,2</sup>, YANG Shaomin<sup>1</sup>, FANG Hongcheng<sup>1</sup>, LI Zhifang<sup>1</sup>

1. Shenzhen Hospital of Integrated Traditional Chinese and Western Medicine, Shenzhen 518104, China

2. Shenzhen Clinical College of Integrated Chinese and Western Medicine, Guangzhou University of Chinese Medicine, Shenzhen 518104, China

**Abstract: Objective** To study the mechanism of arecoline effect on myocardial hypertrophy based on bioinformatics, machine learning, and molecular simulation. **Methods** To conduct differential expression analysis based on the transcriptome dataset GSE249925 related to myocardial hypertrophy in the GEO database. Three machine learning algorithms (LASSO, RF, SVM-RFE) were used to screen key candidate genes. GO and KEGG enrichment analysis, their potential biological functions and signaling pathways were explored. Molecular docking and molecular dynamics simulations were employed to evaluate the binding characteristics and structural stability of arecoline with the target protein. **Results** A total of 28 candidate genes that might be involved in the induction of cardiac hypertrophy by palmatine were identified. After being jointly screened by three machine learning algorithms, four core targets (*STAT3*, *SFRP4*, *S100A12*, and *MT1A*) were finally determined. These genes showed significant abnormal expression in

收稿日期: 2025-12-26

基金项目: 广东省中医药局科研项目 (20231297); 深圳市科技计划项目 (JCYJ20230807141402005); 深圳市“医疗卫生三名工程”资助项目 (SZZYSM202106006); 深圳市宝安区医疗卫生科研项目 (2022JD260); 深圳市宝安区公立医院高质量发展研究项目 (BAGZL2024068)

作者简介: 丁志强, 男, 硕士, 研究方向为中西医结合防治心血管疾病。E-mail: d2361266657@163.com

\*通信作者: 李志芳, 女, 博士, 硕士研究生导师, 研究方向为中西医结合防治心血管疾病。E-mail: 18825286176@163.com

the cardiac hypertrophy tissues and had good diagnostic efficacy. Enrichment analysis revealed that they were involved in multiple key pathways, including JAK/STAT pathway, HIF-1 pathway, AGE-RAGE pathway, and PI3K/Akt pathway. Molecular docking and kinetic simulation results indicated that arecoline could stably bind to STAT3 and S100A12. **Conclusion** This study systematically elucidates the potential molecular mechanisms by which arecoline induces cardiac hypertrophy. Several key regulatory targets and signaling pathways were identified, providing novel insights into the toxicological impact of arecoline and its possible biomarker roles. These findings offer a theoretical and computational basis for future mechanistic research and targeted intervention strategies.

**Key words:** arecoline; cardiac hypertrophy; bioinformatics; machine learning; LASSO; RF; SVM-RFE; JAK/STAT pathway

槟榔是一种广泛分布于南亚和东南亚地区的热带植物,其种子因具有提神醒脑、增强注意力等作用而被广泛咀嚼食用<sup>[1]</sup>。近年来,槟榔对心血管系统的潜在毒性作用引起了学界关注。研究表明,槟榔摄入显著关联高血压、动脉粥样硬化、心律失常及心功能损伤等心血管疾病,提示其可能在心血管疾病的发生发展中具有促病理作用<sup>[2]</sup>。槟榔中主要生物碱具有拟副交感神经和拟毒蕈碱样活性<sup>[3]</sup>。槟榔碱被认为是槟榔主要的药理与毒理活性物质,具有显著的拟胆碱作用,可通过激活乙酰胆碱受体调控多种生理及病理过程<sup>[4]</sup>。已有研究表明,槟榔碱可诱导氧化应激、炎症反应及心肌细胞凋亡,破坏心肌结构完整性及代谢稳态<sup>[5-6]</sup>。动物实验证实,长期或高剂量暴露于槟榔碱可能导致心肌肥厚、心室重构,甚至心力衰竭,其机制涉及多个关键信号通路<sup>[7-8]</sup>。本研究利用 GEO 数据库中心肌肥厚组织的转录组数据,结合生物信息学分析与机器学习算法,筛选槟榔碱相关的关键靶基因,旨在为深入解析其致病机制及毒理学风险评估提供理论依据和研究框架。

## 1 资料与方法

### 1.1 槟榔碱在心肌肥厚中关键靶点的识别

为系统识别槟榔碱在心肌肥厚发生发展中的潜在分子靶点,首先从 PubChem 数据库获取其二维分子结构<sup>[9]</sup>,并利用 CTD、STITCH、SwissTarget Prediction、PharmMapper、GeneCards 数据库进行靶点预测,整合上述平台所得预测结果并去除重复项后,构建槟榔碱潜在作用靶点集合。选取了 GEO 数据库中的数据 GSE249925,该数据集包含来自肥厚型心肌病患者的 mRNA 转录组数据<sup>[10]</sup>。该数据集对应研究纳入 97 例有症状梗阻性肥厚型心肌病患者的室间隔切除术心肌组织及 23 例对照心脏组织,并进行了 RNA 测序。差异表达分析使用 R 语言中的 limma 包 (v 3.64.1) 完成。根据注释文件将样本分为心肌肥厚组与健康对照组,构建设计矩阵

与对比矩阵,随后进行线性模型拟合和经验贝叶斯统计检验。以  $|\log_2FC| > 2$  与校正后的  $P$  值  $< 0.05$  为阈值,筛选出显著差异表达基因 (DEGs),分别提取上调与下调基因用于后续分析。

将槟榔碱预测靶点集合与肥厚型心肌病 DEGs 进行交集分析,获得交叉重叠的关键基因。这些重叠基因被认为是可能介导槟榔碱诱导心肌肥厚的核心靶点,并被用于后续的功能富集分析及分子机制探讨。

### 1.2 交集靶点的基因本体 (GO) 注释与京都基因与基因组百科全书 (KEGG) 通路富集分析

为进一步探讨槟榔碱诱导心肌肥厚过程中关键靶点的潜在生物功能及其所涉及的信号通路,本研究采用综合功能注释平台 Metascape 对前期筛选得到的重叠基因进行功能富集分析。以“人类”作为参考物种。分析参数设置如下:最小重叠基因数设为 3,统计显著性阈值为  $P < 0.01$ ,最小富集因子设为 1.5。GO 注释从 3 个主要功能维度进行分类,即生物过程 (BP)、细胞组分 (CC)、分子功能 (MF)。此外,还进行了 KEGG 通路分析,系统评估这些核心基因可能参与的生物信号通路及调控网络。上述分析旨在为槟榔碱诱导的心脏重塑机制提供更全面的生物学解释。

### 1.3 多种机器学习算法识别心肌肥厚核心靶点

为系统识别在槟榔碱诱导心肌肥厚过程中具有潜在调控意义的核心靶基因,本研究对前期获得的 28 个重叠基因应用 3 种广泛使用的机器学习算法进行特征选择。这 3 种算法包括最小绝对收缩与选择算子 (LASSO) 回归、随机森林 (RF)、支持向量机-递归特征消除 (SVM-RFE)。为提高结果的稳健性与可靠性,引入交叉验证策略。将 3 种算法共同筛选出的交集基因定义为候选靶基因。

LASSO 逻辑回归模型通过 R 语言中的 glmnet 包 (v 4.1.10) 构建,并采用 10 折交叉验证以确定最优惩罚参数  $\lambda$ 。在该  $\lambda$  值下,具有非零回归系数

的基因被选为候选靶点。RF 方法使用 randomForest 包 (v4.7.1.2) 对基因表达矩阵进行训练, 构建包含 500 棵决策树的随机森林模型, 基于 Gini 不纯度的平均减少值对特征重要性进行排序, 选取信息量最大的前 10 个基因作为候选, 并利用 ggplot2 (v3.5.2)、ggpubr (v0.6.1) 进行可视化。SVM-RFE 算法以线性核 SVM 为基学习器, 先对表达特征进行 z-score 标准化, 再在 5 折交叉验证下递归消除权重最小特征, 比较不同特征数子集的准确率, 选择准确率最高对应的特征数作为最优子集, 并输出其基因列表。随后在外层 5 折交叉验证中报告曲线下面积 (AUC)、准确率、精确率、召回率、F1 分数、特异度。被 3 种算法共同识别的基因被视为在分类判别中具有一致性的重要特征, 即为最终核心靶点。

为进一步评估核心靶基因在心肌肥厚组与对照组中的表达差异及其潜在诊断价值, 采用受试者工作特征曲线 (ROC) 分析。利用 pROC 包 (v1.18.5) 对每个核心基因绘制 ROC 曲线, 并计算 AUC, 以评估其区分疾病状态与正常状态的能力。AUC 值接近 1 表示其具有优异的诊断性能, 提示其作为潜在生物标志物的应用前景。此外, 为增强区分能力并模拟多标志物联合诊断策略, 本研究整合核心基因表达水平构建逻辑回归模型, 生成综合 ROC 曲线并计算其 AUC 值, 以全面评估该模型对治疗组与对照组的判别能力, 从而验证多基因联合策略的预测性能及其潜在临床应用价值。

#### 1.4 分子对接与分子动力学模拟分析

为预测槟榔碱与候选核心靶蛋白之间的结合亲和力与相互作用稳定性, 首先进行分子对接分析。槟榔碱的三维结构从 PubChem 数据库获取, 靶蛋白的结构则来自 RCSB PDB 与 UniProt 数据库。采用 AutoDock Vina 进行刚性对接模拟, 并计算配体与靶标的结合能。对接构象通过 PyMOL 和 Discovery Studio 进行可视化和分析, 重点考察其结合模式与主要相互作用类型。结合稳定性评估依据结合能与对接构象特征进行, 结合能低于  $-5 \text{ kcal/mol}$  ( $1 \text{ cal}=4.2 \text{ J}$ ) 被视为具有良好结合潜力, 能量值越低表示结合相互作用越稳定。在分析氢键、疏水相互作用、 $\pi-\pi$  堆积等关键作用形式及药理学相关性的基础上, 选取结合模式最优的配体-受体复合物用于后续模拟。

基于 GROMACS 2022.3 软件进行了分子动力学模拟, 配体结构预处理通过 Amber Tools22 完成,

使用 GAFF (general AMBER force field) 进行参数化, 氢原子添加与电荷拟合采用 Gaussian 16W 结合限制静电势拟合完成。所得电荷信息被整合至系统拓扑文件中。模拟体系在恒温 (300 K) 与恒压 (1 bar) 条件下运行, 蛋白质使用 AMBER99SB-ILDN 力场, 并添加显式水分子与  $\text{Na}^+$  离子以维持体系电中性。在正式模拟前, 先通过最速下降法进行能量最小化, 随后在 NVT 与 NPT 系综下分别进行 100 ps 的平衡模拟, 耦合常数设定为 0.1 ps。执行主生产阶段的分子动力学模拟, 模拟步数为 5 000 000 步, 时间步长为 2 fs, 总时长为 100 ns。模拟结束后, 使用 GROMACS 内置分析工具对轨迹数据进行评估, 重点分析蛋白主链的均方根偏移 (RMSD)、均方根波动 (RMSF) 以及回转半径 ( $R_g$ ) 等结构稳定性指标。采用分子力学/广义波恩表面积 (MM/GBSA) 方法计算配体-靶点的结合自由能, 为理解其相互作用机制提供更深入的理论依据。

## 2 结果

### 2.1 槟榔碱诱导心肌肥厚的关键靶点识别

通过多种靶点预测数据库共获得与槟榔碱相关的潜在靶点 297 个。从心肌肥厚转录组数据集中共识别出 2 518 个差异表达基因, 这些基因在心肌肥厚组与健康对照组之间表现出明显分层, 聚类分析验证了样本间基因表达趋势的一致性, 表明数据集具有良好的稳定性与可信度, 见图 1A。将差异表达基因的分布情况通过火山图进行可视化, 其中包括 1 281 个上调基因与 1 237 个下调基因, 见图 1B。将槟榔碱相关的潜在靶点与心肌肥厚转录组数据集中差异表达基因进行交集分析, 共获得 28 个重叠基因, 被认为是槟榔碱诱导心肌肥厚的关键作用, 见图 1C。

### 2.2 槟榔碱相关靶基因的功能富集分析

针对前期筛选得到的 28 个与槟榔碱暴露及心肌肥厚共同相关的交集基因, 本研究采用 Metascape 平台进行功能富集分析。在 GO 分析 BP 层面, 共识别出 294 个显著富集的条目, 主要包括血液循环、细胞对含氮化合物的应答以及细胞迁移的正调控等类别。这提示槟榔碱可能通过影响体液稳态、炎症激活及组织重构等 BP 参与心肌肥厚的发生发展。在 CC 维度, 共识别到 14 个富集项, 主要包括囊泡腔、细胞外基质以及细胞外囊泡等, 表明这些关键靶点多定位于分泌囊泡系统或细胞外微环境, 可能通过调控细胞间通讯、基质稳态与免疫调节等

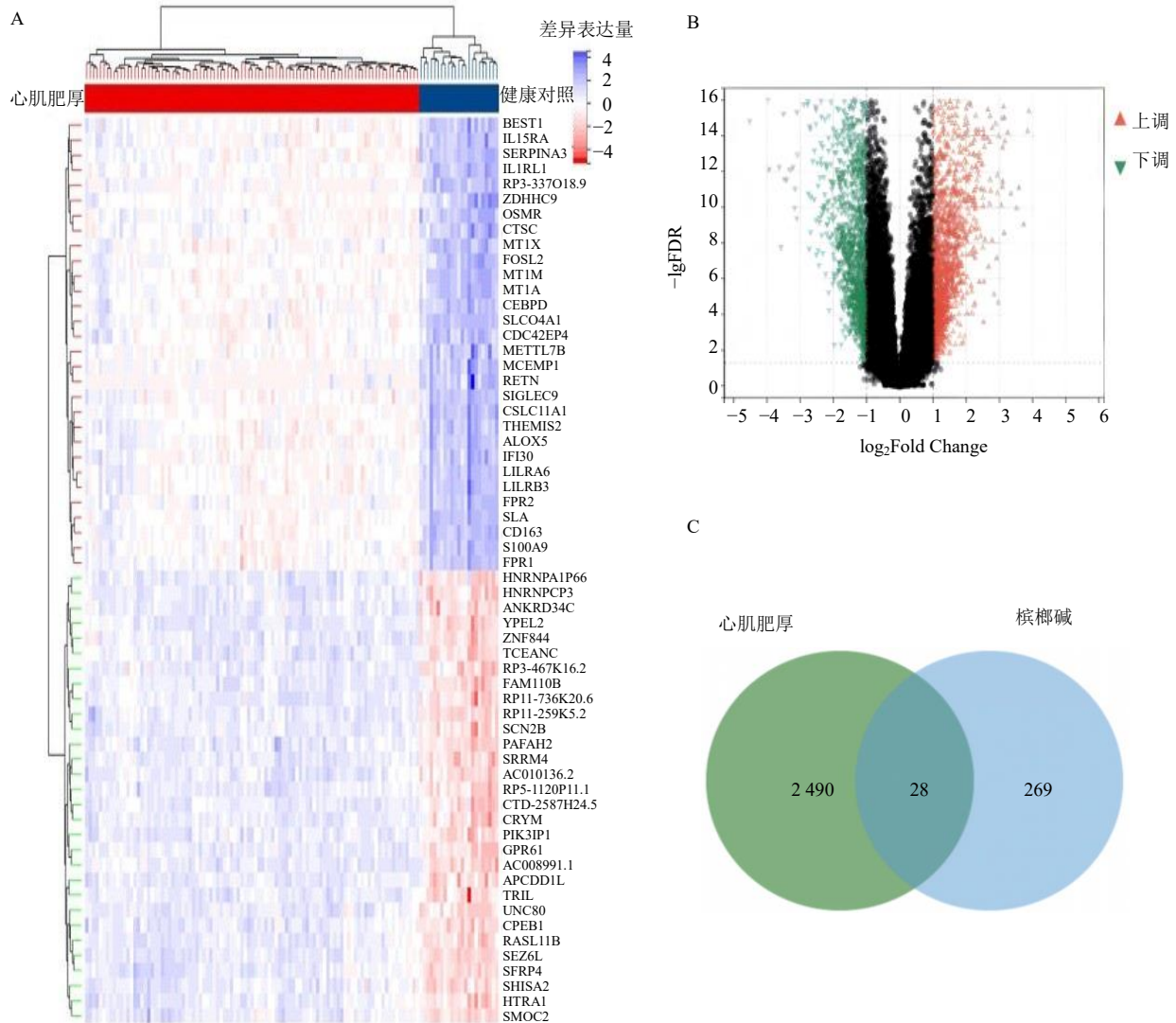


图 1 DEGs 表达模式及层次聚类分析 (A)、DEGs 火山图 (B)、槟榔碱与差异表达基因韦恩图 (C)

Fig. 1 Expression patterns of DEGs and hierarchical clustering analysis (A), volcano plot of DEGs (B), Venn diagram of arecoline and differentially expressed genes (C)

过程发挥作用。在 MF 方面, 共识别出 18 个显著术语, 其中主要包括细胞因子受体结合与一般配体结合, 进一步暗示这些靶点与信号转导、免疫应答调控及氧化还原反应密切相关, 见图 2A。

KEGG 通路富集分析出 72 条显著富集的信号通路, 涵盖多种与炎症反应、心血管毒性相关的机制, 包括 Janus 激酶信号转导/转录激活因子 (JAK/STAT)、缺氧诱导因子 1 (HIF-1)、晚期糖基化终产物-其受体 (AGE-RAGE) 以及磷脂酰肌醇 3-激酶/蛋白激酶 B (PI3K/Akt) 等信号通路。这些信号通路可能为构成槟榔碱诱导心脏重塑、细胞应激反应及炎症损伤的关键分子机制, 见图 2B。

### 2.3 核心靶基因的识别与验证

在 LASSO 回归模型中, 交叉验证结果显示,

$\lambda_{\min}=0.000\ 283\ 9$  (使二项偏差 Binomial deviance 最小),  $\lambda_{1se}=0.003\ 181\ 9$  (在最小偏差的 1 个标准误差范围内、对应更稀疏模型), 见图 3A。为获得更简洁且更稳健的特征集合、降低过拟合风险, 本研究选择  $\lambda_{1se}=0.003\ 181\ 9$  作为最终参数, 并筛选得到 10 个非零回归系数的特征基因 [IIA 组磷脂酶 A2 (*PLA2G2A*)、S100 钙结合蛋白 A12 (*S100A12*)、细胞色素 P450 家族 26 亚家族 B 成员 1 (*CYP26B1*)、白细胞介素-6 (*IL-6*)、分泌型卷曲相关蛋白 4 (*SFRP4*)、两面神激酶 2 (*JAK2*)、金属硫蛋白 1A (*MT1A*)、信号转导和转录激活因子 3 (*STAT3*)、基质金属蛋白酶 9 (*MMP9*)、血红素加氧酶 1 (*HMOX1*)], 见图 3B。在 RF 模型中, 基于候选差异基因的表达数据构建分类模型, 将特征重要性按 Gini 不纯度的

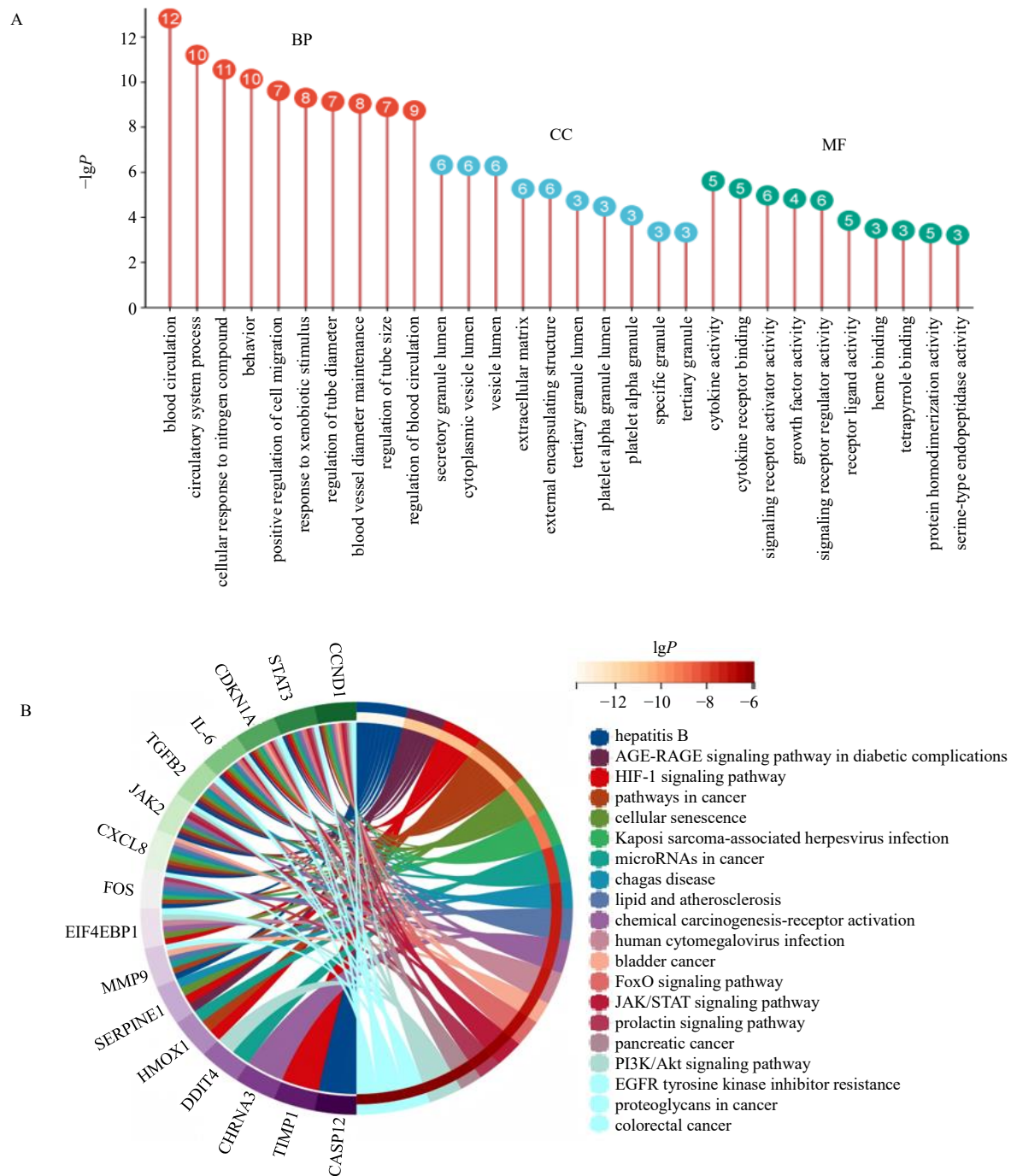


图 2 GO 功能分析 (A) 及 KEGG 通路富集分析 (B)  
Fig. 2 GO function analysis (A) and KEGG pathway enrichment analysis (B)

平均减少值从高到低排序后, 筛得重要性排名前 10 位的候选基因分别为 *SFRP4*、*MT1A*、*STAT3*、细胞周期素 D1 (*CCND1*)、*S100A12*、*PLA2G2A*、*IL-6*、*HMOX1*、*JAK2*、DNA 损伤诱导转录物 4 (*DDIT4*)，见图 3C。SVM-RFE 采用 5 折交叉验证评估不同特征数下的分类表现, 并以交叉验证准确度最高对应的特征数作为最优子集。结果显示最优特征数为 5, 最终筛得核心特征基因 *SFRP4*、*STAT3*、*CCND1*、

*MT1A*、*S100A12*, 见图 3D。对上述 3 种算法的结果进行交集分析, 最终确定 *SFRP4*、*MT1A*、*STAT3*、*S100A12* 4 个在所有模型中均被筛选出的核心靶基因, 见图 3E。

进一步的 ROC 曲线分析表明, 这 4 个核心基因均具有较强的判别能力, AUC 值均大于 0.94, 表现出优异的诊断性能。其中, *SFRP4* 的 AUC 最高 (0.998), 其次为 *STAT3* (0.986)、*MT1A* (0.971)、

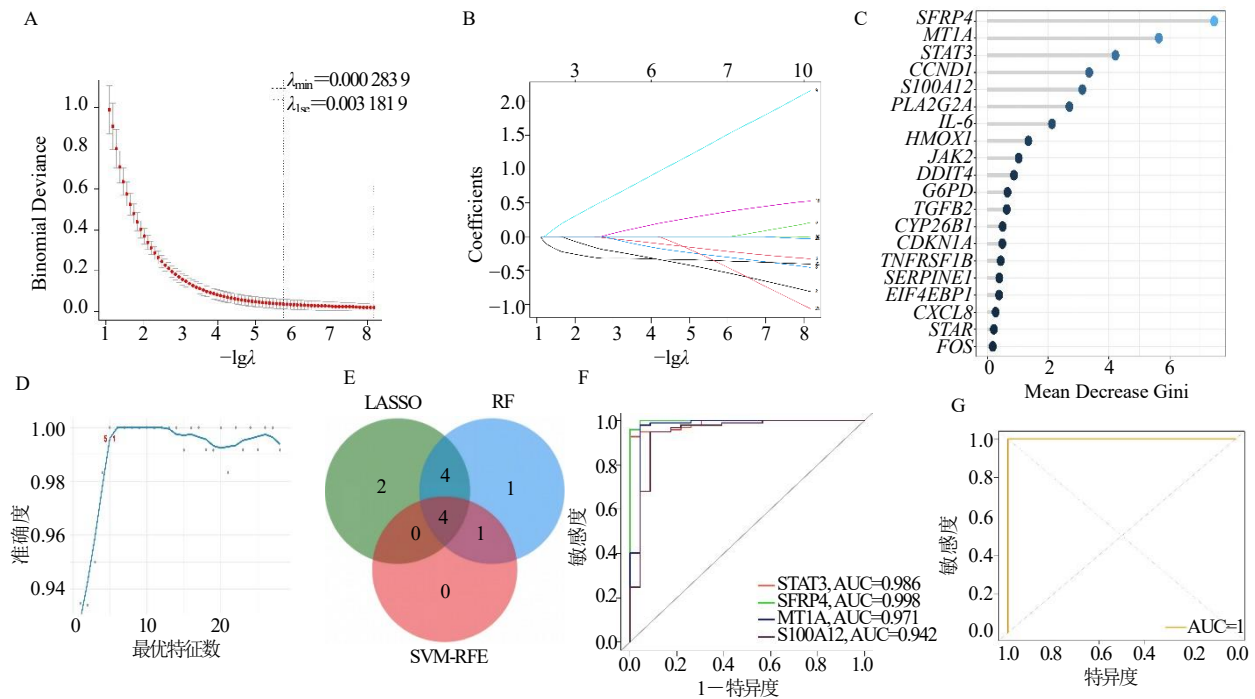


图 3 LASSO 回归的交叉验证曲线 (A)、LASSO 回归模型的回归系数路径图 (B)、RF 模型中基因特征 (C)、SVM-RFE 算法的交叉验证准确率曲线 (D)、3 种机器学习算法筛选韦恩图 (E)、4 个核心诊断基因的 ROC 曲线分析 (F)、核心基因构建的逻辑回归联合诊断模型的 ROC 曲线 (G)

Fig. 3 Cross-validation curve of LASSO regression (A), regression coefficient path diagram of LASSO regression model (B), gene features in RF model (C), cross-validation accuracy curve of SVM-RFE algorithm (D), Venn diagram of three machine learning algorithms (E), ROC curve analysis of four core diagnostic genes (F), ROC curve of the logistic regression combined diagnostic model constructed based on core genes (G)

*S100A12* (0.942), 均显示出良好的预测能力。值得注意的是, 将上述 4 个基因联合构建逻辑回归模型后, 综合 ROC 曲线分析显示 AUC 值达 1.00, 提示该联合基因特征在当前数据集中具有近乎完美的分类性能。这强调了其作为多标志物诊断工具在临床应用中的潜在价值, 见图 3F、G。

## 2.4 分子对接与分子动力学模拟分析

### 2.4.1 槟榔碱与核心靶点的分子对接分析

本研究利用 AutoDock Vina、PyMOL、Discovery Studio 软件, 对槟榔碱与 4 个筛选得到的核心靶点蛋白进行分子对接分析。结果显示, 所有蛋白-配体复合

物的结合能均优于  $-4.0$  kcal/mol, 提示槟榔碱对多种与心肌肥厚相关的靶点具有较强的结合潜力。槟榔碱-STAT3 复合物的结合能最低 ( $-5.4$  kcal/mol), 显示出最强的结合亲和力; 其次为槟榔碱-S100A12 复合物 ( $-5.1$  kcal/mol)、槟榔碱-SFRP4 ( $-4.9$  kcal/mol)、槟榔碱-MT1A ( $-4.2$  kcal/mol)。结果表明, 槟榔碱可能通过与这些关键靶点的稳定结合, 在心肌肥厚的病理过程中发挥潜在作用, 见图 4。

### 2.4.2 分子动力学模拟分析

为在有限计算资源下获得更具代表性的动态结合信息, 先对 4 个核心

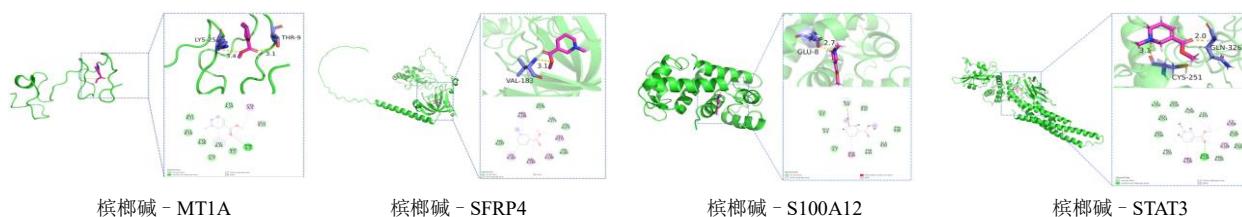


图 4 槟榔碱与核心靶点分子对接构象图

Fig. 4 Molecule docking conformation diagram of arecoline with core target

靶点 (STAT3、S100A12、MT1A、SFRP4) 与槟榔碱的对接构象进行比较, 综合对接结合能/得分排序、关键相互作用 (稳定氢键/疏水核心) 数量与口袋定位合理性筛选分子动力学对象。结果槟榔碱 - STAT3 与槟榔碱 - S100A12 在上述指标中整体表现更优 (结合能得分排名靠前, 且具有更稳定的关键作用位点与更紧密的相互作用网络), 而 MT1A 与 SFRP4 复合物相互作用较少、结合构象相对开放。基于“结合稳定性更高+潜在生物学相关性更强”的原则, 最终选取 STAT3 与 S100A12 复合物开展后续分子动力学模拟, 通过分析关键轨迹参数, 并结合 MM/GBSA 自由能分解, 深入探讨 2 种蛋白 - 配体复合物的构象行为与结合机制。

对于槟榔碱 - S100A12 复合物, 模拟初期系统存在一定波动, 但在约 20 ns 后趋于稳定。定量分析显示, 复合物的平均 RMSD 为  $(0.41 \pm 0.05)$  nm, 蛋白主链的 RMSD 为  $(0.33 \pm 0.05)$  nm, 表明构象稳定性较好。RMSF 均值为  $(0.19 \pm 0.11)$  nm, 提示残基级别的柔性适中; Rg 为  $(1.71 \pm 0.02)$  nm, 反映结构较为紧凑; 平均平均溶剂可及表面积 (SASA) 为  $(100.31 \pm 4.76)$  nm<sup>2</sup>, 表面暴露程度无

显著变化。MM/GBSA 能量分解显示总结合自由能为  $(-13.46 \pm 0.58)$  kcal/mol, 主要由范德华力和疏水作用驱动  $(-20.08 \pm 0.56)$  kcal/mol; 而电荷相互作用相对较弱  $(2.84 \pm 0.11)$  kcal/mol, 溶剂化能不利于结合  $(3.77 \pm 0.08)$  kcal/mol。

槟榔碱 - STAT3 复合物在 10~15 ns 内迅速趋于平衡, 展现出更快的稳定性。该复合物的平均 RMSD 为  $(0.41 \pm 0.04)$  nm, 蛋白主链 RMSD 为  $(0.37 \pm 0.05)$  nm, 均显示出系统良好的稳定性; 其 RMSF 略高  $(0.20 \pm 0.12)$  nm, Rg 显著较大  $(3.57 \pm 0.02)$  nm, 与 STAT3 较为开放的三级结构特征一致。SASA 值达到  $(286.54 \pm 3.77)$  nm<sup>2</sup>, 反映出较大的蛋白表面积。能量方面, 该复合物总结合自由能为  $(-16.05 \pm 0.45)$  kcal/mol, 表现出更强的结合能力。范德华力为  $(-19.09 \pm 0.01)$  kcal/mol, 且电荷相互作用显著增强  $(-9.86 \pm 0.33)$  kcal/mol, 尽管溶剂化能不利  $(12.90 \pm 0.31)$  kcal/mol, 但强烈的非键合相互作用对此产生了补偿。综上, 两种蛋白 - 槟榔碱复合物可形成构象稳定的结合结构, 槟榔碱 - STAT3 复合物结合能更低、电荷与疏水协同作用更强、蛋白柔性调控能力更优, 见图 5、表 1。

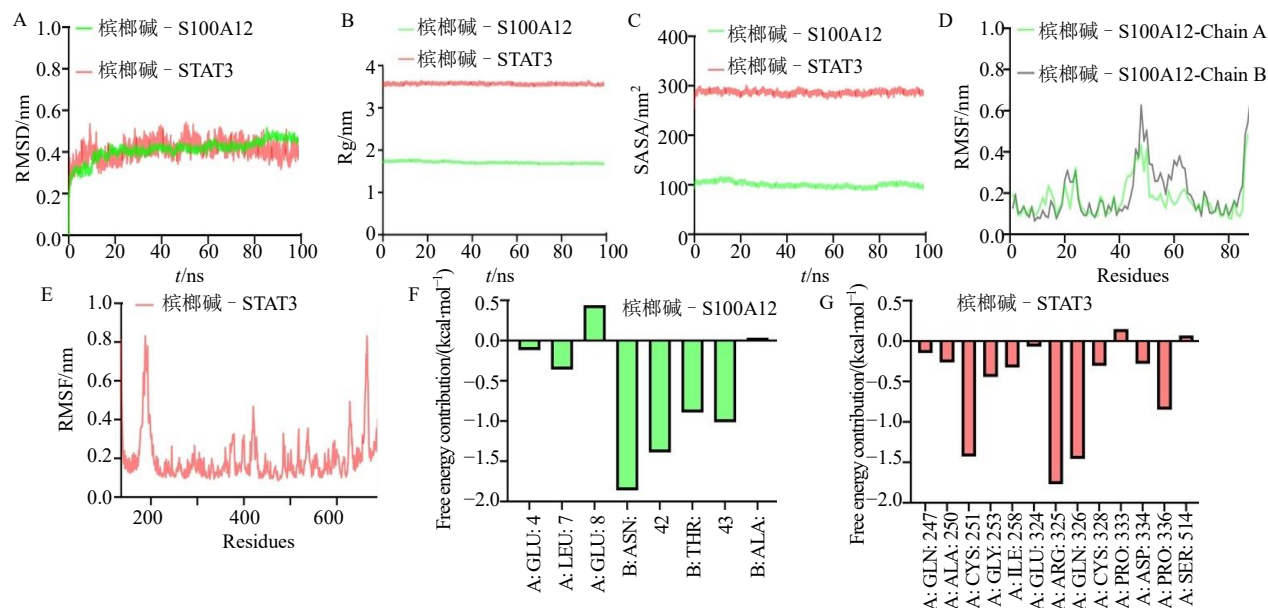


图 5 配体 - 蛋白复合物的 RMSD 轨迹图 (A)、蛋白质主链的 Rg (B)、SASA 变化图 (C)、S100A12 蛋白主链原子的 RMSF 分析 (D)、STAT3 蛋白主链原子的 RMSF 分析 (E)、槟榔碱 - S100A12 复合物 MM/GBSA 分解 (F)、槟榔碱 - STAT3 复合物 MM/GBSA 分解 (G)

Fig. 5 RMSD trajectory diagram of ligand-protein complex (A), Rg of protein main chain (B), SASA change diagram (C), RMSF analysis of S100A12 protein main chain atoms (D), RMSF analysis of STAT3 protein main chain atoms (E), MM/GBSA decomposition of arecoline -S100A12 complex (F), MM/GBSA decomposition of arecoline-STAT3 complex (G)

表 1 槟榔碱与 S100A12 及 STAT3 复合物的自由能分析  
Table 1 Free energy analysis of the complex formed by arecoline, S100A12 and STAT3

自由能	能量/(kcal·mol <sup>-1</sup> )	
	S100A12	STAT3
范德华力	-20.08±0.56	-19.09±0.01
电荷相互作用	2.84±0.11	-9.86±0.33
玻恩溶剂化能	6.35±0.08	15.65±0.31
表面能	-2.58±0.00	-2.74±0.02
气相自由能	-17.24±0.58	-28.95±0.33
吉布斯溶剂化自由能	3.77±0.08	12.90±0.31
总结合自由能	-13.46±0.58	-16.05±0.45

### 3 讨论

本研究系统性探讨了槟榔碱参与心肌肥厚发生发展的潜在分子机制。通过整合公共转录组数据与多平台靶点预测结果，揭示了槟榔碱诱导心脏病理改变的分子基础。在差异表达分析与交集筛选的基础上，结合 3 种主流机器学习算法，有效识别出关键分子特征。最终共筛选出 4 个核心靶基因 (*STAT3*、*SFRP4*、*MT1A*、*S100A12*)。这些基因在心肌肥厚组织中表达显著异常，且表现出良好的判别性能。功能富集分析提示，它们涉及多个关键通路，包括炎症反应、细胞凋亡、代谢调控与信号转导，表明其可能在槟榔碱诱导的心脏重塑过程中发挥重要调控作用。

*STAT3* 是 JAK/STAT 信号通路中的关键转录因子，在心肌中具有“双重作用”，在生理应激状态下发挥保护作用，而在过度激活时则参与病理性心肌肥厚的进程<sup>[11]</sup>。已有研究表明，*STAT3* 可被 IL-6、血管紧张素 II、JAK2 等上游信号激活，活化后的 *STAT3* 促进多种与心肌肥厚相关基因的表达，如脑钠肽 (BNP) 及促凋亡蛋白 (Bax)，从而诱导心肌细胞肥大<sup>[12]</sup>。此外，*STAT3* 还参与心肌细胞的代谢重编程，其激活可抑制葡萄糖氧化过程，导致能量代谢失衡，加速心肌肥厚的进展<sup>[11]</sup>。更重要的是，*STAT3* 在线粒体中的非经典功能近年来也被逐渐发现，包括调控氧化磷酸化与抗氧化能力<sup>[13]</sup>。然而，*STAT3* 的持续或过度激活亦可能诱发氧化应激与心肌纤维化，加重病理性心脏重构<sup>[14]</sup>。在本研究中发现，槟榔碱与 *STAT3* 之间具有稳定的结合能力，提示其可能通过增强 *STAT3* 介导的促炎信号通路并干扰心肌代谢程序，形成“慢性炎症 - 代谢失衡”的病理性正反馈回路。这一机制可能是槟榔碱诱导

心脏毒性效应的关键环节，并进一步确立了 *STAT3* 作为其核心分子介导因子的潜在地位。

*S100A12* 是一种由中性粒细胞分泌的钙结合蛋白，其主要通过与其受体 RAGE 的结合发挥促炎作用。该配体 - 受体相互作用可激活多条炎症相关信号通路，包括核因子- $\kappa$ B (NF- $\kappa$ B) 和丝裂原活化蛋白激酶 (MAPK) 级联反应。上述信号通路的激活可进一步促进心肌细胞中的炎症反应、细胞凋亡及纤维化过程，最终导致心脏重构与左心室肥厚的发生<sup>[15-16]</sup>。动物实验显示，在慢性肾病背景下过表达 *S100A12* 的小鼠会表现出明显的心肌肥厚与舒张功能障碍，并伴随成纤维细胞生长因子 23 水平显著升高<sup>[17]</sup>。在临床研究中，*S100A12* 血清水平升高已被证实与 2 型糖尿病患者心力衰竭风险增加显著相关，进一步支持其在心血管疾病发生发展中的病理作用<sup>[18]</sup>。

*SFRP4* 是 Wnt 信号通路的内源性抑制因子，在心脏病理过程中发挥关键调控作用。已有研究表明，*SFRP4* 可通过与整合素  $\alpha_1\beta_1$  以及 Notch1 信号通路的相互作用，干扰正常的心肌细胞分化，提示其可能参与异常的心肌细胞增殖过程<sup>[19]</sup>。在扩张型心肌病与心力衰竭患者中观察到 *SFRP4* 的表达水平升高，且与多种纤维化及凋亡相关基因呈正相关，支持其作为心脏重构潜在生物标志物的应用价值<sup>[20]</sup>。*SFRP4* 表达失调可能导致 Wnt 信号平衡破坏，从而促发心肌的结构性与功能性重构<sup>[21]</sup>。最新多组学和生物信息学研究进一步揭示，*SFRP4* 与心肌肥厚标志物如心钠素 (ANP) 与 BNP 呈显著正相关，提示其可能参与氧化应激反应、线粒体功能障碍及免疫 - 炎症调控<sup>[22-23]</sup>。因此，*SFRP4* 在心脏结构重塑与免疫调控中可能同时发挥作用，具有“双重调控潜力”，其具体分子机制及治疗价值仍需进一步深入研究与实验证实。

*MT1A* 是一种富含半胱氨酸、低相对分子质量的金属结合蛋白，在金属离子稳态维持及细胞抗氧化防御中发挥关键作用。多项转录组学分析与基于机器学习的生物信息学研究均证实 *MT1A* 在心肌肥厚患者中表达显著上调，提示其可能作为早期分子生物标志物具有潜在应用价值<sup>[24]</sup>。在相关动物实验中，在肥胖或高血压诱导的心肌肥厚与重构模型中，*MT1A* 缺失可显著加剧心脏病理变化，而其过表达则能有效缓解心肌氧化应激损伤与线粒体功能障碍<sup>[25]</sup>。此外，*MT1A* 基因的多态性与多种心

管疾病的易感性密切相关, 进一步强调其在临床诊断与疾病预测中的潜在价值<sup>[26]</sup>。

结合多维度研究结果, 本研究提出槟榔碱可能通过“炎症激活 - 免疫失衡 - 代谢紊乱”的三重机制诱导心肌肥厚。富集分析显示核心基因显著富集于 JAK/STAT、HIF-1、AGE-RAGE、PI3K/Akt 等炎症与应激相关通路, 提示槟榔碱可能通过持续激活炎症信号通路, 加剧心肌损伤。分子对接与分子动力学模拟结果证实, 槟榔碱可与 STAT3 和 S100A12 稳定结合, 提示其可能直接调控蛋白活性, 诱导氧化应激、能量代谢障碍及线粒体功能失衡。这一“炎症放大 - 免疫失衡 - 代谢损害”的整合模型为深入理解槟榔碱心脏毒性提供了新的理论视角。同时, 本研究也具有重要的临床与公共卫生意义。首先, STAT3、SFRP4、S100A12、MT1A 在心肌肥厚组织中均表现出卓越的诊断性能, 提示其作为早期分子标志物在槟榔碱相关心脏损伤的监测与风险评估中具有潜在价值。其次, 分子模拟结果表明, STAT3 与 S100A12 是槟榔碱的重要结合靶点, 为后续靶向干预药物开发提供了方向。最后, 考虑到槟榔在部分地区的广泛使用及其与心血管疾病的密切相关性, 本研究进一步强调了槟榔碱的潜在公共卫生风险, 提示临床医生在评估心血管病患者时应重视其槟榔使用史, 并加强对相关人群的健康宣教与预防干预。

本研究通过生物信息学分析、机器学习算法与分子模拟技术, 系统揭示了槟榔碱的潜在毒理靶点及其介导心肌肥厚的关键分子机制。同时, 本研究也验证了将生物信息学与结构药理学相结合在毒理机制研究中的实用价值与应用前景。研究结果不仅为进一步探索槟榔碱的毒理机制提供了理论依据, 也为相关疾病的靶向干预策略开发构建了分析框架与研究路径。

**利益冲突** 所有作者均声明不存在利益冲突

#### 参考文献

[1] 梁焕宴, 冀沛然, 弓宝, 等. 槟榔的药用历史及其应用现状 [J]. *中草药*, 2024, 55(2): 678-687.

[2] Liu H, Zheng H, Zhang J, *et al.* Review of the toxic effects and health functions of arecoline on multiple organ systems [J]. *Food Innov Adv*, 2024, 3(1): 31.

[3] Chen L H, Yuan F L, Chen S F, *et al.* Potential role of host microbiome in areca nut-associated carcinogenesis and addiction [J]. *Molecules*, 2022, 27(23): 8171.

[4] Liu Y J, Peng W, Hu M B, *et al.* The pharmacology, toxicology and potential applications of arecoline: A review [J]. *Pharm Biol*, 2016, 54(11): 2753-2760.

[5] Lin W Y, Tsai B C K, Day C H, *et al.* Arecoline induces heart injure via Fas/Fas ligand apoptotic pathway in heart of Sprague-Dawley rat [J]. *Environ Toxicol*, 2021, 36(8): 1567-1575.

[6] Ku C W, Day C H, Ou H C, *et al.* The molecular mechanisms underlying arecoline-induced cardiac fibrosis in rats [J]. *Open Life Sci*, 2021, 16(1): 1182-1192.

[7] Ho T J, Tsai B C K, Kuo C H, *et al.* Arecoline induces cardiotoxicity by upregulating and activating cardiac hypertrophy-related pathways in Sprague-Dawley rats [J]. *Chem Biol Interact*, 2022, 354: 109810.

[8] 聂安政, 高梅梅, 钞艳慧, 等. 槟榔药理毒理探讨与合理用药思考 [J]. *中草药*, 2020, 51(12): 3329-3336.

[9] Wang Y L, Xiao J W, Suzek T O, *et al.* PubChem's BioAssay database [J]. *Nucleic Acids Res*, 2012, 40 (Database issue): D400-412.

[10] Clough E, Barrett T, Wilhite S E, *et al.* NCBI GEO: Archive for gene expression and epigenomics data sets: 23-year update [J]. *Nucleic Acids Res*, 2024, 52(D1): D138-D144.

[11] Yang Y Q, Liu J, Wang L Y, *et al.* Oxytocin attenuates cardiac hypertrophy by improving cardiac glucose metabolism and regulating OXTR/JAK2/STAT3 axis [J]. *Peptides*, 2024, 182: 171323.

[12] Qi B, Xu R H, Jin Y Y, *et al.* A critical role for IL-21/IL-21 receptor signaling in isoproterenol-induced cardiac remodeling [J]. *Sci Rep*, 2025, 15(1): 18985.

[13] Lu X, Tong T T, Sun H L, *et al.* ECSIT-X4 is required for preventing pressure overload-induced cardiac hypertrophy via regulating mitochondrial STAT3 [J]. *Adv Sci*, 2025, 12(8): e2414358.

[14] Rostami A, Palomer X, Pizarro-Delgado J, *et al.* GADD45A suppression contributes to cardiac remodeling by promoting inflammation, fibrosis and hypertrophy [J]. *Cell Mol Life Sci*, 2025, 82(1): 189.

[15] Yan L, Bowman M A H. Chronic sustained inflammation links to left ventricular hypertrophy and aortic valve sclerosis: A new link between S100/RAGE and FGF23 [J]. *Inflamm Cell Signal*, 2014, 1(5): e279.

[16] Oesterle A, Bowman M A H. S100A12 and the S100/ Calgranulins: Emerging biomarkers for atherosclerosis and possibly therapeutic targets [J]. *Arterioscler Thromb Vasc Biol*, 2015, 35(12): 2496-2507.

[17] Yuan Y P, Shen Z Y, Teng T, *et al.* S100a8/9 (S100 Calcium Binding Protein a8/9) promotes Cardiac hypertrophy via

- upregulation of FGF23 (Fibroblast Growth Factor 23) in mice [J]. *J Am Heart Assoc*, 2024, 13(10): e028006.
- [18] Gellen B, Thorin-Trescases N, Thorin E, *et al.* Increased serum S100A12 levels are associated with higher risk of acute heart failure in patients with type 2 diabetes [J]. *ESC Heart Fail*, 2022, 9(6): 3909-3919.
- [19] Tian Y Y, Wang W P, Lu Q, *et al.* Cross-talk of SFRP4, integrin  $\alpha_1\beta_1$ , and Notch1 inhibits cardiac differentiation of P19CL6 cells [J]. *Cell Signal*, 2016, 28(11): 1806-1815.
- [20] Schumann H, Holtz J, Zerkowski H R, *et al.* Expression of secreted frizzled related proteins 3 and 4 in human ventricular myocardium correlates with apoptosis related gene expression [J]. *Cardiovasc Res*, 2000, 45(3): 720-728.
- [21] Bergmann M W. WNT signaling in adult cardiac hypertrophy and remodeling: Lessons learned from cardiac development [J]. *Circ Res*, 2010, 107(10): 1198-1208.
- [22] Ma Z T, Wang X Z, Lv Q B, *et al.* Identification of underlying Hub genes associated with hypertrophic cardiomyopathy by integrated bioinformatics analysis [J]. *Pharmacogenomics Pers Med*, 2021, 14: 823-837.
- [23] Ren M M, Ye X Q, Ouyang C, *et al.* JMJD2A mediates transcriptional activation of SFRP4 and regulates oxidative stress and mitochondrial dysfunction in heart failure [J]. *Pathol Int*, 2024, 74(4): 210-221.
- [24] Zhang F, Li C R, Zhang L L. Transcriptomic analysis and machine learning modeling identifies novel biomarkers and genetic characteristics of hypertrophic cardiomyopathy [J]. *Front Genet*, 2025, 16: 1596049.
- [25] Zhang H N, Zhou W Q, Wang X, *et al.* Exacerbation by knocking-out metallothionein gene of obesity-induced cardiac remodeling is associated with the activation of CARD9 signaling [J]. *Int J Biol Sci*, 2025, 21(3): 1032-1046.
- [26] Sarutipai boon I, Settasatian N, Komanasin N, *et al.* Association of genetic variations in NRF2, NQO1, HMOX1, and MT with severity of coronary artery disease and related risk factors [J]. *Cardiovasc Toxicol*, 2020, 20(2): 176-189.

[责任编辑 高源]