基于成分-基因双维度平行对比的中药寒热药性差异机器学习与生物信息学分析

杨　洋[1]，杨　淇[2]，李家诚[1*]

1. 南京中医药大学附属苏州市中医医院，江苏 苏州　215000
2. 广西中医药大学药学院，广西 南宁　530022

摘　要：目的　基于成分-基因双维度平行比较策略，结合机器学习与生物信息学分析中药寒热药性差异，为药性理论的现代阐释提供依据。**方法**　选取 16 味热性和 18 味寒性中药，在成分层面分析相对分子质量、辛醇-水分配系数、氢键供体数、氢键受体数及可旋转键数的全局分布与特异性成分；在基因层面构建中药-靶点二值化拓扑矩阵，采用基于同源药材隔离的组别留一交叉验证（leave-one-group-out cross-validation，Group-LOOCV）和置换检验评价分类模型，并结合 χ^2 特征筛选、Shapley 加性解释（Shapley additive explanations，SHAP）及功能富集分析解析关键靶点与相关通路。**结果**　寒热药成分的常规分子理化特征分布高度重叠。基因靶点分类模型中，脊回归分类器（ridge classifier，RidgeClass）的准确率为 0.82，受试者工作特征曲线下面积（area under the receiver operating characteristic curve，AUC）为 0.85，经检验 $P<0.01$。识别出再生家族成员 1α（regenerating family member 1 alpha，REG1A）、核受体亚家族 3 C 组成员 1（nuclear receptor subfamily 3 group C member 1，NR3C1）、分层蛋白（stratifin，SFN）及瞬时受体电位锚蛋白 1（transient receptor potential ankyrin 1，TRPA1）等 8 个核心特征基因。热性药相关基因主要关联心血管疾病与细胞应激，寒性药相关基因主要关联肿瘤及铂类耐药，二者下游信号通路存在交叠。**结论**　在数据集及分析框架下，基因靶点网络特征较常规分子理化特征更有助于区分中药寒热药性，提示生物网络调控可能是药性差异的重要基础，为后续机制验证与药性研究提供线索。

关键词：中药寒热药性；机器学习；生物信息；Shapley 加性解释；瞬时受体电位锚蛋白 1

中图分类号：R285；Q811.4　　**文献标志码：**A　　**文章编号：**0253 - 2670(2026)21 - 8606 - 11

DOI: 10.7501/j.issn.0253-2670.2026.21.018

Machine learning and bioinformatics analysis of differences between cold and heat properties of traditional Chinese medicine based on component-gene dual-dimensional parallel comparison strategy

YANG Yang[1], YANG Qi[2], LI Jiacheng[1]

1. Suzhou Traditional Chinese Medicine Hospital Affiliated to Nanjing University of Chinese Medicine, Suzhou 215000, China
2. School of Pharmacy, Guangxi University of Chinese Medicine, Nanning 530022, China

Abstract: Objective　To investigate differences between cold and hot properties of traditional Chinese medicine (TCM) using a component-gene dual-dimensional parallel comparison strategy integrating machine learning and bioinformatics. **Methods**　A total of 16 hot-property and 18 cold-property herbs were selected. At the component level, global distributions of molecular weight, octanol-water partition coefficient, hydrogen bond donor count, hydrogen bond acceptor count, and rotatable bond count, together with property-specific components, were analyzed. At the gene level, a binary herb-target topological matrix was constructed. Leave-one-group-out cross-validation (Group-LOOCV) with isolation of homologous herbs and permutation testing were used to evaluate classification models. Chi-square feature selection, Shapley additive explanations (SHAP), and functional enrichment analysis were combined to interpret key targets and associated pathways. **Results**　Conventional molecular physicochemical features of components from cold- and hot-property herbs showed substantial overlap. Among target-based classifiers, the ridge classifier (RidgeClass)

收稿日期：2026-04-26

基金项目：苏州市青年科技人才托举工程（QRC2507）；南京中医药大学 2025 年临床教学专项研究课题（NZYJY2025-L-45）；第三批吴门医派中医专家学术经验继承工作湿热治病专项课题（YSC2026024）；姑苏卫生人才计划（紧缺型人才）（SZWSRC2025078）

作者简介：杨　洋，副主任中医师，主要研究方向为中医药治疗肺病。E-mail: 42188714@qq.com

***通信作者：**李家诚，主治中医师，主要研究方向为中医药医疗大数据及人工智能研究。E-mail: lijiacheng@njucm.edu.cn

achieved an accuracy of 0.82 and an area under the receiver operating characteristic curve (AUC) of 0.85, with an empirical $P < 0.01$. A total of eight core feature genes were identified, including regenerating family member 1 alpha (REG1A), nuclear receptor subfamily 3 group C member 1 (NR3C1), stratifin (SFN), and transient receptor potential ankyrin 1 (TRPA1). Genes associated with hot-property herbs were mainly related to cardiovascular diseases and cellular stress, whereas those associated with cold-property herbs were mainly related to tumors and platinum drug resistance. Their downstream signaling pathways overlapped. **Conclusion** Within the present dataset and analytical framework, gene-target network features were more informative for distinguishing cold and hot medicinal properties than conventional molecular physicochemical features. These findings suggest that biological network regulation may contribute to medicinal property differences and provide leads for subsequent mechanistic validation and research.

Key words: cold-heat property of traditional Chinese medicine; machine learning; bioinformatics; Shapley additive explanations; transient receptor potential ankyrin 1

中药药性理论是中医药基础理论体系的核心组成部分，其中寒热药性作为“四性”理论的精髓，直接决定了药物对人体阴阳盛衰与寒热变化的作用倾向。自《神农本草经》以来，历代医家无不以“寒者热之，热者寒之”为组方立法之圭臬[1]。然而这一沿袭数千年的理论体系在当代临床与科研实践中正面临日益严峻的挑战。首先，传统药性判定主要依赖历代本草文献的经验总结与医家的个人体悟，其判别过程本质上是一个宏观疗效与主观感受相互参证的“黑箱”[2]，即药物为何表现为寒性或热性，其内在的化学与生物学逻辑始终未能阐明，导致药性理论与现代医学的靶点-通路-疾病认知体系之间存在难以逾越的鸿沟[3]。其次，寒热错杂证在临床中普遍存在，以及同一药物在不同剂量、不同配伍环境下可能呈现药性逆转的复杂现象，使得寒热属性的精确判定在临床实践中难以形成可量化的统一标准，对精准用药的指导价值亦因此受到限制。

近年来，围绕中药寒热药性的现代科学阐释学界已从多维度展开探索，形成了若干有代表性的研究范式[4]。在实验药理学层面，如有学者基于生物热动力学方法探究了附子与干姜配伍前后的寒热药性变化，从能量代谢角度初步阐释了“附子无干姜不热”的配伍规律[5]。王伽伯团队则通过单胺类神经递质代谢谱研究了姜黄与郁金的寒热药性差异，揭示了不同药性中药干预后儿茶酚胺类物质的差异性变化[6]。这些研究为寒热药性的物质基础与生物学效应提供了有价值的实证线索。然而，上述范式多围绕研究者预先设定的少数经典通路或候选靶标展开验证性分析，研究路径本身易陷入“先入为主、循环论证”的方法论困境，即以已知药性为起点设计实验，又以实验结果反推药性机制，而在此过程中可能损失大量未被预设的、隐藏在复杂生物网络中的关键信息[7]。

目前，已有学者将机器学习算法应用于寒热药性的判别与预测，如 Shapley 加性解释（Shapley additive ex planations，SHAP）。该算法基于博弈论中的 Shapley 值理论，能够将模型预测结果归因到每 1 个输入特征的边际贡献，从而实现对模型决策过程的透明化解构[8]。将 SHAP 解释框架与机器学习模型相结合已在中药产地鉴别、药效物质发现等领域展现出良好的应用前景，使“可解释的人工智能”在中医药研究中成为可能。但目前已有的机器学习分析策略主要为“分子特征”而非“基因特征”[9-10]，容易陷入理化参数的宏观混杂干扰。而若直接进行未去噪的成分-靶点拓扑富集又极易因公共数据库的“科研热度偏倚（annotation bias）”而引入高噪声风险[11]。如何跳出化学空间的直观属性局限，从全局生物网络层面提取高置信度的药性关键特征，符合“整体观念”与“辨证论治”的指导思想。

针对这一现状，本研究创新性地提出“成分-基因双维度平行比较”策略。首先，在成分层面表征寒热药化学成分的常规理化空间分布并捕获专属高频涌现特征；其次，在基因层面直接构建中药个体作用靶点的二值化拓扑矩阵，基于同源分组约束的留一交叉验证（leave-one-group-out cross-validation，Group-LOOCV）与置换检验的严谨约束，融合多维机器学习算法构建高严谨度约束下的药性判别模型，并借助 SHAP 算法定量解构决定分类走向的特征基因网络占有倾向。通过横向对比化学空间的高度重叠与基因拓扑空间的显著可分离性，预期揭示寒热药性的本质区分依据并非单纯的化学结构直观属性，而是深植于生物网络调控的深层逻辑，从而为中药药性理论的科学阐释提供一个兼顾可解释性与预测效能的双维度平行比较分析框架。

1 数据库的构建

1.1 寒热药性标签的构建与数据清洗

本研究通过 Symmap 数据库（http://www.symmap.org/）获得寒热药分类的中药，该数据库是一个通过临床症状与微观分子网络进行双向映射的大型集成化中医药系统生物学平台[12]。以此平台为基础，充分体现寒热的差异性，严格遵循“两极分化、去模糊化”的临床疗效导向策略，仅筛选具备绝对“寒”或“热”极性的中药（排除了药性相对缓和、边界模糊的温、凉及平性药）构建训练基准集。清洗条件：properties_Chinese（药性）列筛选“热”与“寒”选项，排除“温”“凉”和“平”性药，去重；中药名称以《中国药典》2025 年版为准，不被收录的中药则以行业/学术标准如《中药大辞典》或《中药学》教材第 2 版为准。最终筛选出巴豆等 16 个热性中药，设置为正标签“1”。在候选寒性中药池中，为消除类别不平衡对分类器构建的干扰，进一步以中药分子机制生物信息学分析工具（Bioinformatics Analysis Tool for Molecular mechANism of Traditional Chinese Medicine，BATMAN-TCM）数据库靶点注释完整度高和无数据缺失为约束条件，采取无放回抽样匹配出 18 味典型寒性中药作为负标签“0”。正负标签合计 34 个代表性样本，形成正负比例接近 1∶1 的训练分析集（表 1）。

1.2 寒热药成分数据库的构建

将带有寒热注释的中药导入 BATMAN-TCM 数据库（http://bionet.ncpsb.org.cn/batman-tcm/#/search）[13]，评分阈值（score cutoff）设置为 0.84［似然比（likelihood ratio，LR）＝80.88］，校正 *P* 值（adjusted *P* value）设置小于 0.05 获取所有寒热药的中药-成分-寒热注释信息构建数据库。

1.3 寒热药基因数据库的构建

将 34 个带寒热标签的中药数据集在 BATMAN-TCM 数据库中以单味中药为独立样本逐个进行基因靶点的检索，获得不同中药-靶点-寒热注释信息构建数据库。

2 寒热药差异性分析

2.1 分子结构差异分析

依据“1.2”项构建的寒热药分子数据集中的化合物标识符（compound identifier，CID），使用 PubChem 数据库中的批量检索功能[14]，以单味中药为单位分别检索相对分子质量（molecular weight，M_W）、辛醇-水分配系数（computed octanol-water partition coefficient，XLogP）、氢键供体数（hydrogen bond donors，HDs）、氢键受体数（hydrogen bond acceptors，HAs）和可旋转键数（rotatable bonds，RB）5 规则（rule of five，RO5）[15]相关的类药性分子特征。通过对 RO5 进行全局密度分布拟合与区间频数统计，直观考察寒热药化学成分在 5 维化学空间中的重叠程度与分布偏向，并利用箱线图辅

表 1 寒热药性中药训练分析数据集

Table 1 Training and analysis dataset for traditional Chinese medicine with cold-heat medicinal property

中药	拼音	药性	标签
巴豆	badou	辛，热，有大毒	1
斑蝥	banmao	辛，热，有大毒	1
荜茇	biba	辛，热	1
草乌	caowu	辛、苦，热，有大毒	1
川乌	chuanwu	辛、苦，热，有大毒	1
大风子	dafengzi	辛，热，有毒	1
干姜	ganjiang	辛，热	1
高良姜	gaoliangjiang	辛，热	1
胡椒	hujiao	辛，热	1
辣椒	lajiao	辛，热	1
两头尖	liangtoujian	辛，热，有毒	1
吴茱萸	wuzhuyu	辛、苦，热，有小毒	1
仙茅	xianmao	辛，热，有毒	1
炮姜	paojiang	辛，热	1
水蛭	shuizhi	辛，热，有大毒	1
制草乌	zhicaowu	辛、苦，热，有毒	1
乌敛莓	wulianmei	苦、酸，寒	0
鸭跖草	yazhicao	甘、淡，寒	0
牵牛子	qianniuzi	苦，寒，有毒	0
苦楝子	kulianzi	苦，寒，小毒	0
天竺黄	tianzhuhuang	甘，寒	0
甘遂	gansui	苦，寒，有毒	0
夏枯草	xiakucao	辛、苦，寒	0
青黛	qingdai	咸，寒	0
豨莶草	xixiancao	辛、苦，寒	0
禹州漏芦	yuzhou loulu	苦，寒	0
川射干	chuanshegan	苦，寒	0
黄柏	huangbo	苦，寒	0
车前子	cheqianzi	甘，寒	0
芦荟	luhui	苦，寒	0
地肤子	difuzi	辛、苦，寒	0
瓜蒂	guadi	苦，寒，有毒	0
瓜蒌皮	gualoupi	甘，寒	0
京大戟	jingdaji	苦，寒，有毒	0

以高频专属成分的特征涌现分析进行联合可视化呈现。

2.2 作用靶点差异分析

2.2.1 判别模型构建与 SHAP 可解释性分析 本研究引入中药单元整体观，以 34 味中药个体作为行，将其调控的去重基因靶点作为输入特征。采用二值化方法进行标准化编码，“1”表示该中药调控该靶点网络，“0”表示未调控，构建出 34 行的原始基因靶点拓扑矩阵。为定量评估样本间重叠度，首先计算了 34 味中药间的 Jaccard 拓扑网络相似度矩阵。鉴于特征空间维度远大于样本量，本研究首先切除极低频噪声特征，随后采用 χ^2 检验实施全局降维，精选出对寒热分类最具决定性贡献的 8 个核心 hub 特征基因。为彻底切断同源药材及炮制品（如草乌与制草乌、干姜与炮姜、瓜蒂与瓜蒌皮）因基因谱相似可能引发的“近邻识别”与数据泄露，本研究引入 Group-LOOCV 策略。将同源/炮制品独立分组，交叉验证迭代时强制隔离训练集与测试集中的同源关联样本。循环评估随机森林（random forest，RF）、支持向量机（support vector machine，SVM）、朴素贝叶斯（naive Bayes）及脊回归分类器（ridge classifier，RidgeClass）的判别性能，利用准确率（accuracy，Acc）、受试者工作特征（receiver operating characteristic，ROC）曲线下面积（area under the receiver operating characteristic curve，AUC）、F1 分数、精确率（precision，Pre）及混淆矩阵等量化评价指标锁定最优模型。并对最优模型执行 100 次标签随机置换检验以验证超小样本（N＝34）下的分类稳健性并计算经验 P 值。

在此基础上，采用 SHAP 算法量化核心特征靶点对模型分类结果的边际贡献，以进一步解析不同靶点对寒、热药性判别方向的影响。需要特别说明的是，本研究采用的特征矩阵为静态二值化矩阵，特征值仅表示相应靶点在药物作用网络中的占有状态（1）或非占有状态（0），因此 SHAP 分析反映的是靶点网络占有对模型分类输出的贡献方向，而不代表该基因在生物学意义上的表达上调或下调。在本研究的分类定义下，正 SHAP 值表示相应特征推动模型输出向热性药类别偏移，负 SHAP 值则表示其推动模型输出向寒性药类别偏移。在 SHAP summary plot 中，红色（high）表示该靶点处于网络占有状态，蓝色（low）表示该靶点处于非占有状态。通过比较不同靶点在占有状态下 SHAP 值的方向及幅度，评价其对寒、热药性分类的边际决策贡献及不同模型间的一致性。

2.2.2 全局生物信息富集分析 根据“2.2.1”项模型决策贡献与拓扑降维筛选出的关键药性判别基因，结合 34 味寒热中药在全量网络中分别下调的共有基因集进行点面整合分析。为了彻底滤除因公共数据库中部分热门基因高频注释而产生的科研热度偏倚，本研究未使用默认的全人类基因组作为富集背景，而是将 34 味中药以 10%频次初筛出的共有靶点基因作为专属定制背景集，引入超几何分布检验。

分别将寒热药调控的核心靶点和下调基因集导入 Appyters 数据库（https://appyters.maayanlab.cloud/#/）[16]进行在线人类孟德尔遗传（Online Mendelian Inheritance in Man，OMIM）数据库疾病条目、京都基因与基因组百科全书（Kyoto encyclopedia of genes and genomes，KEGG）信号通路及基因本体（gene ontology，GO）生物功能、分子功能的超几何分布富集，利用自组织映射与全局降维图谱观察寒热 2 极药性在全景上的空间分离特征与重叠趋势。

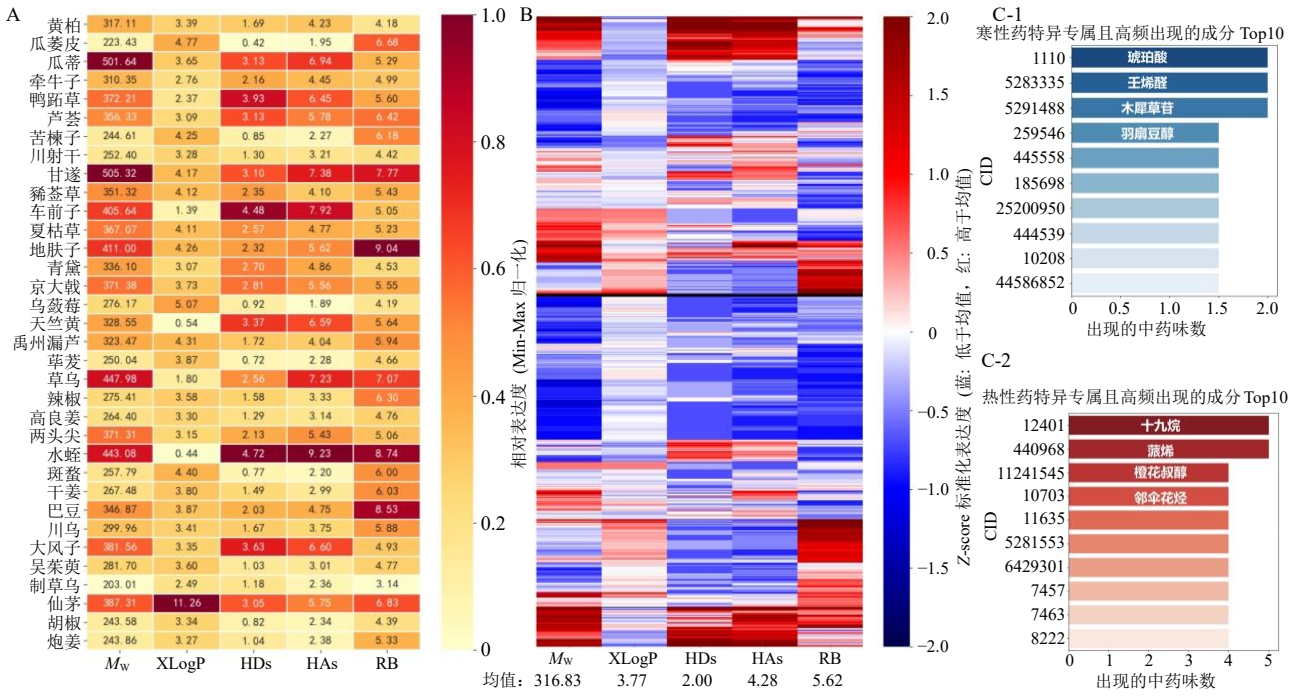
2.2.3 核心信号通路机制注释 为了在微观层面上打通核心特征基因到临床药效表达的现代医学因果解释桥梁，使用相互作用基因/蛋白质检索工具（Search Tool for the Retrieval of Interacting Genes/Proteins，STRING）数据库（https://string-db.org/）[17]对“2.2.1”项中 SHAP 贡献度最高且处于拓扑关键处的 8 个 hub 标志物进行深度交互网络重构与通路注释。物种限定为智人（*Homo sapiens*），重点解析并比较这 8 个决定药性决策的核心基因，在癌症通路（pathways in cancer）、p53 信号通路（p53 signaling pathway）、痛觉传导（pain transduction）、产热作用（thermogenesis）等特异性代表通路和信号节点的扰动作用。

本研究的原始数据与代码已上传至 github（https://github.com/GleasonCN/Han-Re-data.git）公开分享。

3 结果

3.1 寒热药成分的局部-全局差异观

由图 1-A、B 结果显示，寒性中药（如黄柏等）与热性中药（如干姜等）在 5 维常规化学理化特征的光谱带上表现出高度交织和重叠的态势。整体而言，全局成分的宏观均值分别收敛于 M_W（316.83）、



A-带寒热注释的单味中药 RO5 矩阵；B-寒热药的 RO5 全景光谱；C-1-寒性药特异性成分；C-2-热性药特异性成分。

A-RO5 matrix of single herbs with cold-heat property annotations; B-RO5 panoramic spectrum of cold and hot medicinal herbs; C-1-specific components of cold-natured herbs; C-2-specific components of hot-natured herbs.

图 1　寒热药-成分分子特征差异表现

Fig. 1　Differences in molecular characteristics of constituents between cold and hot property TCMs

XLogP（3.77）、HDs（2.00）、HAs（4.28）及 RB（5.62）。尽管部分热性中药（如仙茅等）在 XLogP 或 HDs 上存在极端离群分布，且部分寒性中药的 XLogP 分布相对内敛，但 2 组药物在微观理化参数的宏观投影上并未呈现出清晰的断层边界。这一全局光谱表征提示，单一或常规的化学结构直观参数在均值维度上无法有效区分中药的寒热药性。而图 1-C 表明，寒性药特异专属空间内高频涌现出琥珀酸（CID 1110）、壬烯醛（CID 5283335）等成分表现出明确的寒性药占有倾向。相反，热性药特异专属空间内则高频涌现出十九烷（CID 12401）、蒎烯（CID 440968）等挥发性单萜烃类及多碳长链烷烃成分，呈现出绝对的排他性现象。

3.2　寒热药-基因预测模型性能与特征重要性

本研究构建了 7 223 个热性中药调控基因和 7 404 个寒性中药调控基因的数据集。首先，图 2 展示了 34 个中药样本间的 Jaccard 拓扑相似度，其中可以看到炮制品、同源不同药用部位品的作用靶点差异较大，避免了“近邻识别”的发生。以中药作用的基因为特征开展寒热药性分类，各模型性能差异显著（图 3、表 2）。RidgeClass 在小样本约束下涌现出优异的分类稳健性，其 Acc 达到 0.82，马修斯相关系数（Matthews correlation coefficient，MCC）高达 0.70。混淆矩阵（图 3-E）与预测概率分布（图 3-D）清晰显示，RidgeClass 实现了对 16 味热性中药的无误判识别（敏感性为 1.00），而对 18 味寒性中药的特异性（specificity，Spe）为 0.67。与此同时，RF 展现出最强的全局特征空间流形捕获能力，其 AUC 达到了 0.88，Acc 为 0.79。Naive Bayes 模型的 AUC 同样达到了 0.88，但其整体偏向于 Class 0（寒性）概率池的离散分布，导致综合准确率仅为 0.68。SVM 的综合判别准确率最低（0.68），且概率自适应校准曲线（图 3-C）提示其盲测预测概率与真实标签概率存在较明显的非线性扭曲。

图 4 展示了 8 个 hub 基因在 RF 和 naive Bayes 模型中的 SHAP 分布特征及其对寒、热药性分类方向的边际贡献。总体来看，多数核心靶点在 2 个模型中表现出较为一致的 SHAP 极性，提示其对分类方向的贡献具有一定的跨模型稳定性。其中，当再生家族成员 1α（regenerating family member 1 alpha，REG1A）、核受体亚家族 3 C 组成员 1（nuclear receptor subfamily 3 group C member 1，NR3C1）

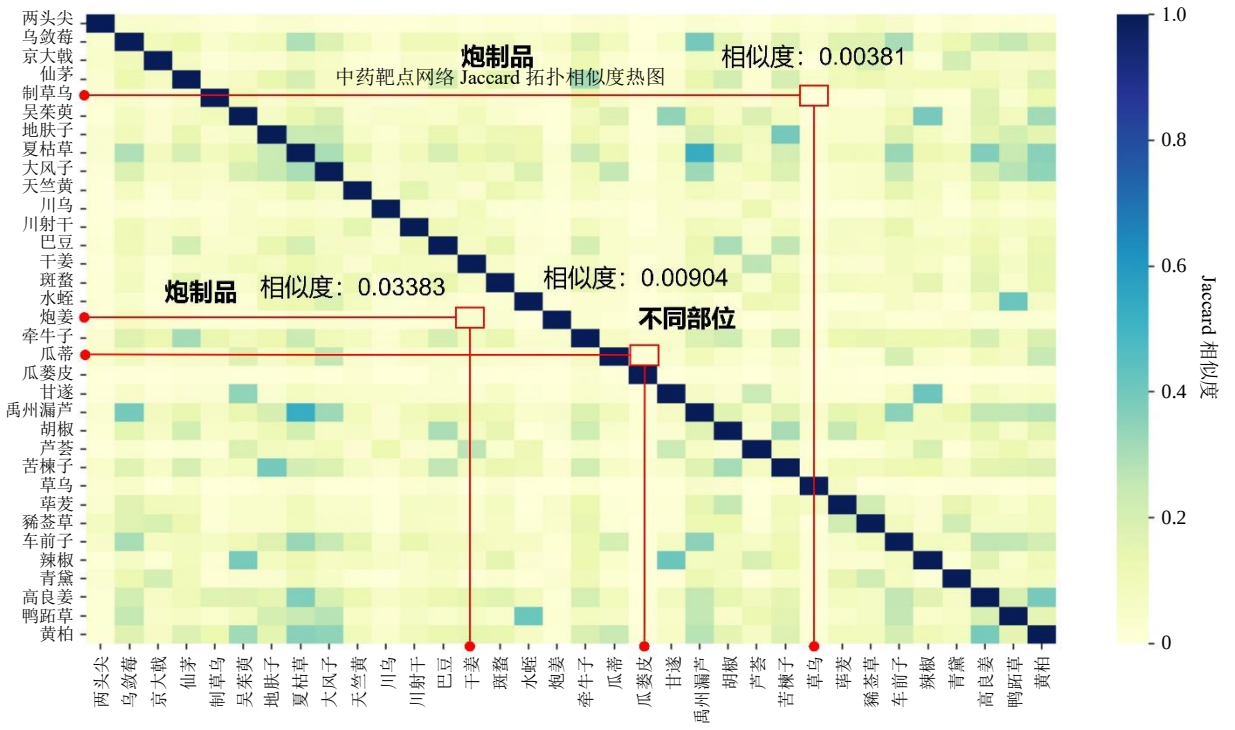


图 2　34 种中药靶点网络 Jaccard 拓扑相似度热图

Fig. 2　Heatmap of Jaccard topological similarity for target networks of 34 TCMs

表 2　寒热药-基因机器学习预测模型性能

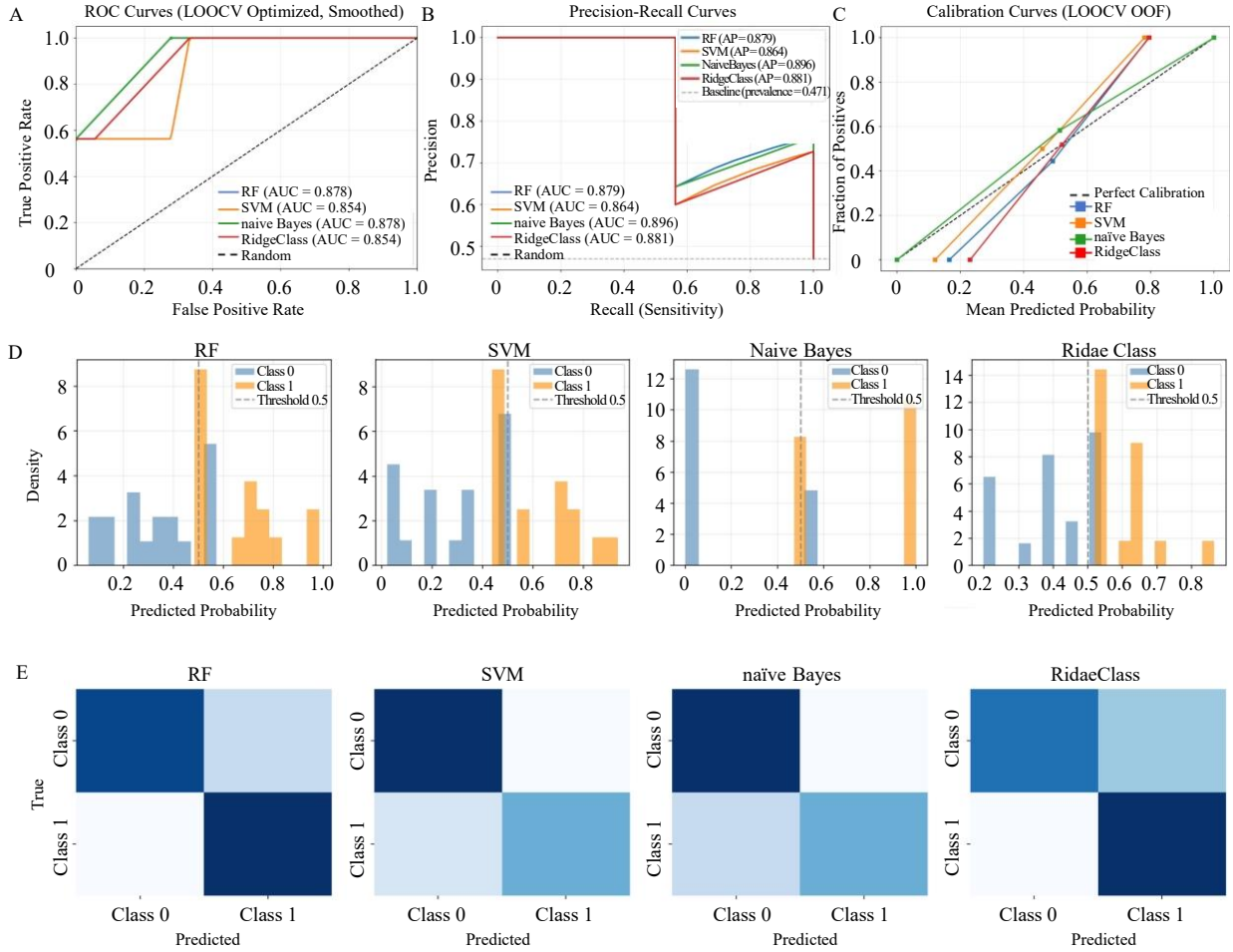
Table 2　Performance of machine learning prediction models for genes of cold and hot property TCMs

模型	Acc	AUC	F1	Pre	敏感性	Spe	MCC
RF	0.79	0.88	0.80	0.74	0.88	0.72	0.60
SVM	0.68	0.85	0.67	0.65	0.69	0.67	0.35
Naive Bayes	0.68	0.88	0.65	0.67	0.63	0.72	0.35
Ridge Class	0.82	0.85	0.84	0.73	1.00	0.67	0.70

和分层蛋白（stratifin，SFN）处于网络占有状态时，红色散点主要分布于负 SHAP 值区域，表明这些靶点的网络占有倾向于推动模型向寒性药类别判定。相反，瞬时受体电位锚蛋白 1（transient receptor potential ankyrin 1，TRPA1）、细胞色素 C 氧化酶亚基 5A（cytochrome C oxidase subunit 5A，COX5A）、DnaJ 热休克蛋白家族成员 B4[DnaJ heat shock protein family (Hsp40) member B4，DNAJB4] 和细胞周期蛋白依赖性激酶抑制因子 1B（cyclin dependent kinase inhibitor 1B，CDKN1B）在网络占有状态下主要表现为正 SHAP 值，提示其出现倾向于推动模型向热性药类别判定。Polo 样激酶 1（polo like kinase 1，PLK1）的 SHAP 分布则相对复杂，其贡献方向和幅度在不同样本及模型间存在一定差异。

3.3　寒热药-基因的全局扰动差异

如图 5 所示，寒性药与热性药的调控效应网络在生物信息景观中均有清晰的全局两极分离特征。然而，在宏观通路与功能底片的深层空间中，2 组药物也存在明显的共有重叠阴影区。结果表明，由 REG1A、NR3C1、SFN 和 PLK1 组成的寒性判定基因群（图 5-E-1），其效应集中卡位于细胞增殖与周期阻断的核心卡口。例如，通过在癌症通路中卡位 PLK1 控制有丝分裂以逆转分裂失控，在 p53 信号通路中卡位 SFN 响应 DNA 损伤以触发周期阻阻滞，从而共同服务于清热解毒、抗恶性增殖的寒性临床表达。相反，由 TRPA1、COX5A、DNAJB4 和 CDKN1B 组成的热性判定基因群（图 5-E-2），则高度特异性地聚焦于外源应激与能量代谢调节。其中，TRPA1 处于痛觉传导的核心关卡，直接响应外



A-ROC 曲线；B-精确率-召回率曲线；C-校正曲线；D-预测概率分布；E-混淆矩阵。

A-ROC curve; B-precision-recall curve; C-calibration curve; D-predicted probability distribution; E-confusion matrix.

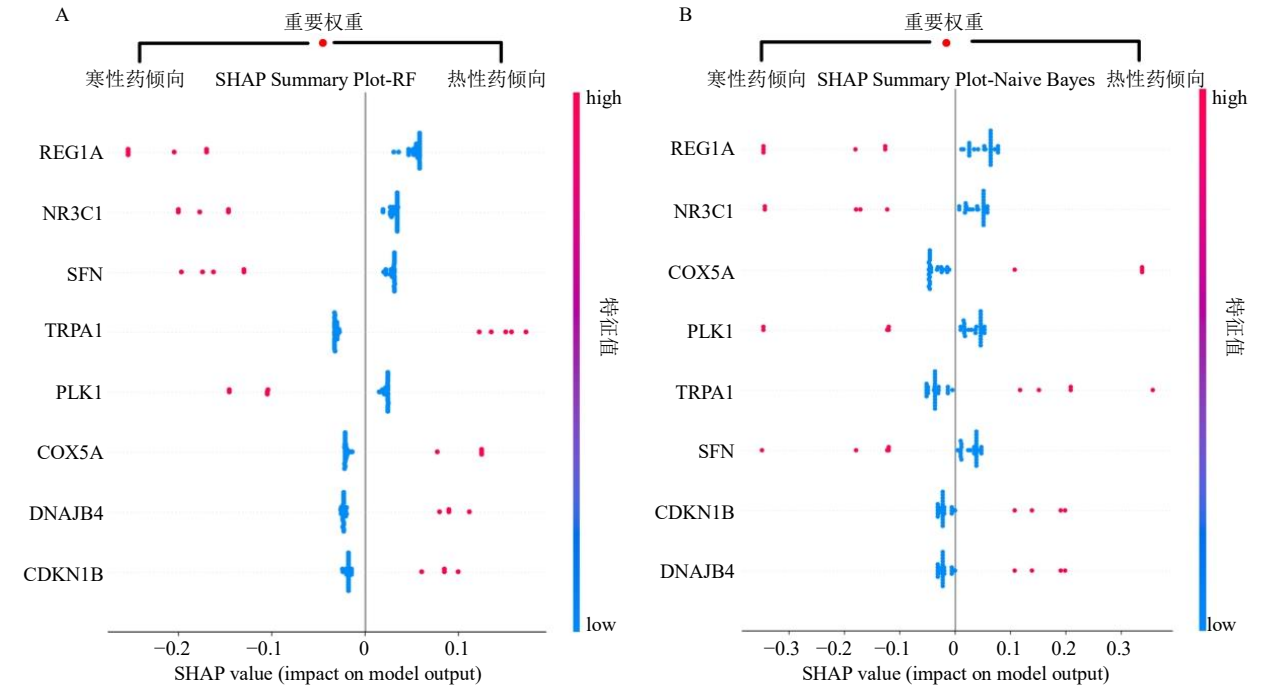
图 3　寒热药-基因机器学习模型性能

Fig. 3　Performance of machine learning prediction models for target genes of cold and hot property TCMs

伤刺激并介导疼痛传导；COX5A 则卡位于产热作用的关键限速步骤，在寒冷暴露下启动氧化产热。

4　讨论

中医药理论体系中，“寒、热、温、凉”四性是指引临床用药的核心法门，即“寒者热之，热者寒之”。然而，传统药性理论长期面临两大挑战：其一是宏观疗效与微观机制之间的不可解释性鸿沟；其二是现有寒热药性研究多局限于单一化学成分的含量测定或少数经典通路的验证，难以系统描绘寒热属性的全景式生物学差异。本研究创新性地提出了一种“成分-基因平行比较分析”策略，通过分类器效能横向互证，在排除样本规模与计算过拟合偏倚的 Group-LOOCV 隔离验证与置换检验数理强约束下，2 个维度的判别模型表现出了显著的效应差异。结果表明，仅依赖分子 RO5 特征构建的模型在复杂因果推断中易受混杂干扰，得到的结论可靠性有限。而在引入 χ^2 检验实施拓扑降维、构建中药单元基因效应二值化矩阵后，模型的判别性能发生根本性跨越，最佳模型盲测 AUC 稳健增长至 0.85（Acc 达到 0.82，经验 $P<0.01$）。这一显著的性能断层明确揭示了寒热药性的本质区分依据，并非单纯取决于化学结构空间的直观通用属性，而是深植于下游生物网络调控的深层极性逻辑之中。经 χ^2 特征筛选并通过 SHAP 解析的 8 个 hub 靶点（如寒性药占有倾向的 REG1A、SFN 与热性药占有倾向的 TRPA1、COX5A），在 2 类异质性分类器中体现出高度一致的决策贡献极性。该发现有效滤除了公共数据库高频注释产生的科研热度偏倚，为“药性源于机体应答”的中医原创思维提供了兼顾计算可解释性与现代医学靶向机制的生物信息学图谱。

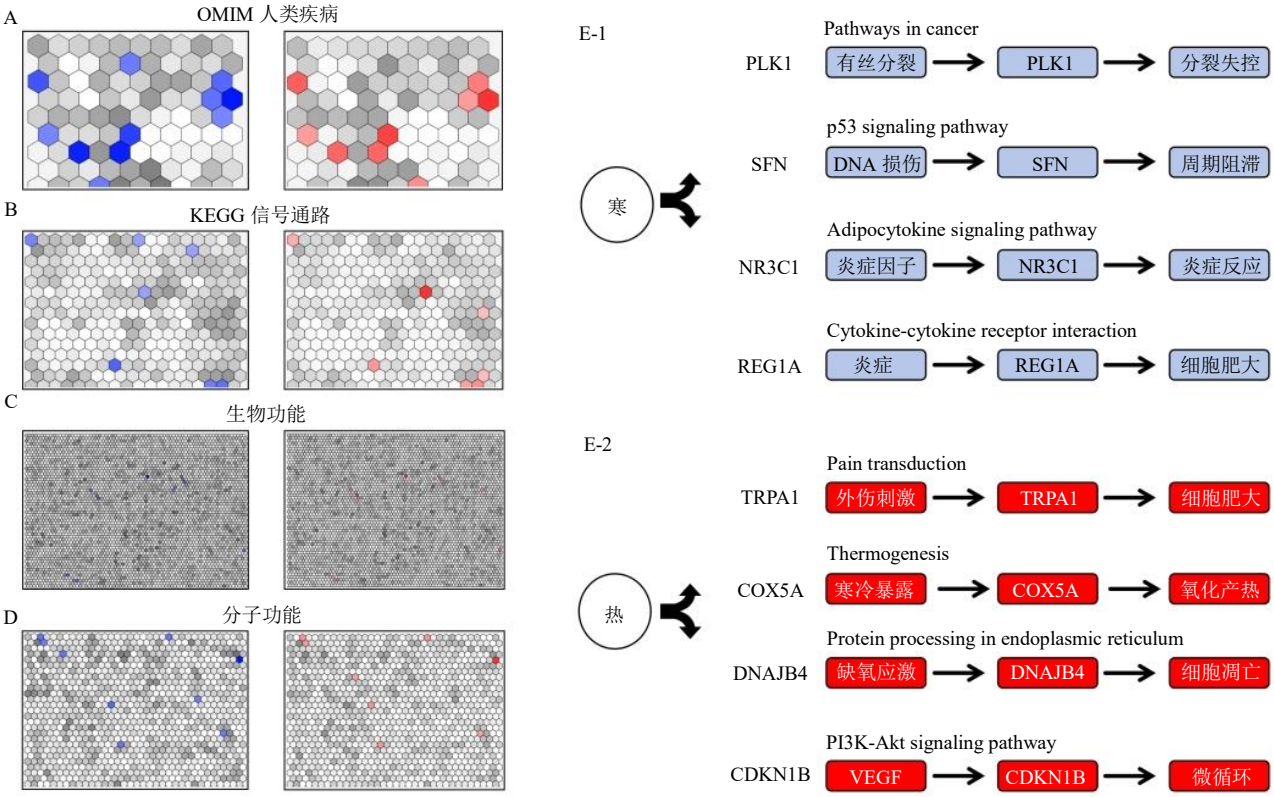


A-RF 模型 SHAP 特征重要性解释；B-NaiveBayes 模型 SHAP 特征重要性解释。

A-SHAP feature importance interpretation of RF model; B-SHAP feature importance interpretation of Naive Bayes model.

图 4　寒热药-基因机器学习预测 SHAP 分析（性能 Top 2）

Fig. 4　SHAP analysis of machine learning prediction for target genes of cold and hot property TCMs (Top 2 performing models)



A-疾病富集；B-信号通路富集；C-生物功能富集；D-分子功能富集；E-1-寒性药 hub 基因高关联通路；E-2-热性药 hub 基因高关联通路。

A-disease enrichment; B-signaling pathway enrichment; C-biological function enrichment; D-molecular function enrichment; E-1-pathways highly associated with hub genes of cold-natured medicines; E-2-pathways highly associated with hub genes of hot-natured medicines.

图 5　寒热药下调基因生物功能富集差异

Fig. 5　Differences in biological function enrichment of downregulated genes induced by cold and hot property TCMs

在寒性药方面，模型筛选出的核心特征基因及富集结果主要聚焦于肿瘤细胞增殖与周期调控。以寒性代表药夏枯草为例，本研究发现其作用网络显著抑制了与细胞周期、增殖相关的激酶活性及蛋白质磷酸化过程[18]。这与此前的药理学证据高度吻合，如研究表明夏枯草中的迷迭香酸、熊果酸等成分可通过下调细胞周期蛋白依赖性激酶 2（cyclin dependent kinase 2，CDK2）、抑制信号转导和转录激活因子 3（signal transducer and activator of transcription 3，STAT3）磷酸化来诱导肿瘤细胞凋亡[19]，临床上常用于辅助治疗甲状腺结节及淋巴瘤。富集结果提示的肝癌、内分泌腺癌及铂类耐药关联性也侧面印证了寒性清热解毒药在逆转肿瘤多药耐药中的潜在系统科学内涵。

在热性药方面，模型筛选出的 hub 基因及富集结果主要指向心血管疾病、缺血性疾病及神经系统调节。其中，TRPA1 作为本研究识别出的热性药判别相关核心特征之一，是一种多模态感觉离子通道，可感知温度变化、亲电性化学刺激及氧化应激，并参与伤害性感觉和疼痛传导[20]。既往研究表明，人源 TRPA1 具有内源性双向温度感受特征，可对冷、热刺激产生响应；同时，TRPA1 激活还可促进降钙素基因相关肽（calcitonin gene-related peptide，CGRP）释放并介导血管舒张，提示其在温度感知、神经感觉与心血管调节之间具有潜在桥接作用[21]。因此，本研究筛选得到的 TRPA1 及其相关功能富集结果为热性药温里止痛等传统功效与现代温度感知、疼痛及神经血管调控机制之间建立了较为合理的生物学联系。

本研究通过高维降维展现了寒热药性的拓扑分离，但在通路交汇的深层数据中，却精确映射出中医“阴阳互根、对立统一”的辨证智慧。全局网络拓扑显示，寒热药性在疾病功能条目上虽呈现高度的极性分离，但在信号通路富集层面存在显著的交叠区域，这构成了“同病异治”与“异病同治”在分子网络层面的数理表征。

中医理论强调“孤阴不生，独阳不长”[22]，寒与热并非绝对割裂的二元状态，自然辩证法中的矛盾对立统一揭示了这种规律[23]。在生物网络中，这种关系体现为寒性药对激酶活性的“下调抑制”与热性药对应激响应的“上调激活”，两者实质上作用于同一个核心网络模块（如磷脂酰肌醇-3-羟激酶（phosphatidylinositol-3-hydroxykinase，PI3K）-蛋白激酶 B（protein kinase B，Akt）或丝裂原活化蛋白激酶（mitogen-activated protein kinase，MAPK）级联反应）[24-25]。寒性药通过抑制异常增殖的病理网络以“祛邪存阴”，热性药通过促进微循环与能量代谢以“温阳消阴”，其靶点占有虽各具偏向，但最终效应均指向重建机体阴阳自和的稳态系统。本研究筛选出的 TRPA1 等关键靶点为连接中药寒热药性与现代感觉神经调控提供了重要分子线索。TRPA1 具有温度感知、化学刺激感知及疼痛传导等多模态功能，其网络占有在本研究中表现出明显的热性药判别倾向。结合寒热药在下游通路中的差异与交叠，提示 2 类药性可能通过不同的网络占有模式实现相反而又相互关联的生物调控。

尽管本研究提供了有价值的生物信息学见解，但仍存在若干局限性有待突破。第一，中药药性来源于中医典籍的记录，无法新增品种，热性中药的数量稀少问题不具有被解决特点。这导致了数据异质性限制，最终形成模型泛化性的固有屏障。第二，传统中药药性是一个多维复杂的调控系统，从药物本身而论，寒热药性往往与五味（如辛热、苦寒）、归经及药物毒性（如巴豆、斑蝥等热药多具大毒）存在强烈的数理交互与混杂效应。从药物外部而论，产地环境（道地性）[26]、采收季节[27]、煎煮条件[28]以及临床方剂配伍[29]等外源性非分子特征，也会直接干预效应成分的溶出与机体响应。受限于当前中药单元样本量的客观限制，本研究尚未能将五味、归经和毒性作为协变量引入分类器中开展高阶交互效应分析，在一定程度上限制了模型对药性全景特征的外延能力。第三，数据来源的静态性与滞后性，本研究依赖公共数据库，中药成分与靶点数据处于动态更新中。第四，本研究未纳入药物炮制（如生用与制用）对成分谱及靶点网络的颠覆性影响，如生地黄-熟地黄[30]、生天南星-胆南星[31]等。亦未涉及药动学（吸收、分布、代谢、排泄）过程，导致预测靶点与实际“入血成分”靶点可能存在偏差[32]。第五，生物学背景的简化，富集分析基于通用的人类基因注释，未考虑疾病状态下的细胞特异性及物种差异，异质性难以完全还原中药复方在复杂体内环境中的涌现效应[33]。展望未来，本研究仍具有进步的潜力：第一，引入血清药物化学与入血成分分析，剔除假阳性靶点，聚焦真正作用于靶器官的效应物质；第二，结合单细胞测序与空间转录组学，解析寒热药性在不同组织微环境及特定细胞

亚群中的精准调控差异；第三，纳入方剂配伍规律研究，探索“去性存用”配伍策略下寒热药物组合对生物网络的重编程作用[34]；第四，充分标准化中药生产与使用流程，减少药物从产生到发挥作用的中间信息交换给带来的噪声。通过上述多模态数据的整合与动态模拟，有望真正架起连接传统药性理论与现代精准医学的桥梁。

5 结论

本研究基于“成分-基因双维度平行比较”策略，系统解析了中药寒热药性的微观生物学内涵。分子结构差异分析表明，仅依赖常规分子理化特征（RO5）参数高度交织重叠，在宏观投影上无法有效表征和区分药性。而基于基因靶点二值化拓扑矩阵构建的判别模型在排除同源数据泄露与样本相似度干扰的 Group-LOOCV 严谨约束下展现出优异且稳健的盲测性能（最优模型准确率达 0.82，AUC 达 0.85，经验 $P<0.01$），明确表明寒热药性的本质差异并非单纯的静态化学直观属性，而是深植于下游生物网络的系统调控。全局扰动与生物信息富集分析显示，寒热药性在疾病功能层面呈现出高度的分离与通路交叠：寒性药下调基因聚焦于肿瘤增殖、铂类耐药与激酶活性调控，在微观上阐释了清热解毒药逆转病理过热的现代内涵；热性药下调基因则显著指向心血管、缺血性疾病与细胞应激响应，重在通过重塑微循环与能量代谢来恢复机体稳态。最终经特征筛选并结合 SHAP 解析识别出的 REG1A、SFN、NR3C1、TRPA1 等 8 个核心 hub 特征靶点，其网络占有倾向对模型判定具有高度一致的边际决策贡献（而非表达量上下调），在分子网络尺度上精确映射了中医“阴阳互根、对立统一”与“相反相成”的辨证智慧，为中药药性理论的现代科学阐释提供了全新的双维度平行比较与横向互证分析范式。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 林林, 李丽, 赵鹏程, 等. 中药四性研究进展及方法回顾 [J]. 中国实验方剂学杂志, 2023, 29(15): 267-272.

[2] 赵鼎, 翟长云, 乐根明, 等. 黑箱理论与中医学探析 [J]. 四川中医, 2019, 37(6): 13-15.

[3] 王正山, 张其成. 中医寒热概念的本质及相关问题辨析 [J]. 吉林中医药, 2014, 34(8): 757-760.

[4] 姜淼, 吕爱平. 基于药物生物效应的中药寒热属性分类研究策略 [J]. 中国中药杂志, 2014, 39(11): 2149-2152.

[5] 叶强, 郭一平, 彭成, 等. 附子无姜不热的成分研究 [J]. 中成药, 2013, 35(5): 1035-1039.

[6] 吴宏伟, 李洪梅, 唐力英, 等. 基于单胺类神经递质代谢谱研究姜黄和郁金的寒热药性差异 [J]. 高等学校化学学报, 2011, 32(4): 839-843.

[7] 韩森, 王佃勋, 魏佩煌, 等. 基于网络药理学方法的中药寒/热性药物特异性作用分子机制研究 [J]. 辽宁中医杂志, 2021, 48(8): 186-189.

[8] Harren T, Matter H, Hessler G, *et al*. Interpretation of structure-activity relationships in real-world drug design data sets using explainable artificial intelligence [J]. *J Chem Inf Model*, 2022, 62(3): 447-462.

[9] 张喜科, 赵文华, 付先军, 等. 基于 XGBoost 和 SHAP 的中药寒热药性识别及寒热特征标记可视化研究 [J]. 世界科学技术—中医药现代化, 2022, 24(12): 4984-4993.

[10] Ma S Y, Liu J L, Li W H, *et al*. Machine learning in TCM with natural products and molecules: Current status and future perspectives [J]. *Chin Med*, 2023, 18(1): 43.

[11] Haynes W A, Tomczak A, Khatri P. Gene annotation bias impedes biomedical research [J]. *Sci Rep*, 2018, 8(1): 1362.

[12] Wu Y, Zhang F L, Yang K, *et al*. SymMap: An integrative database of traditional Chinese medicine enhanced by symptom mapping [J]. *Nucleic Acids Res*, 2019, 47(D1): D1110-D1117.

[13] Kong X R, Liu C, Zhang Z Z, *et al*. BATMAN-TCM 2.0: An enhanced integrative database for known and predicted interactions between traditional Chinese medicine ingredients and target proteins [J]. *Nucleic Acids Res*, 2024, 52(D1): D1110-D1120.

[14] Wang Y, Xiao J, Suzek T O, *et al*. PubChem’s BioAssay database [J]. *Nucleic Acids Res*, 2012, 40(D1): D400-D412.

[15] Ding Y H, Qiang B, Chen Q X, *et al*. Exploring chemical reaction space with machine learning models: Representation and feature perspective [J]. *J Chem Inf Model*, 2024, 64(8): 2955-2970.

[16] Clarke D J B, Jeon M, Stein D J, *et al*. Appyters: Turning Jupyter Notebooks into data-driven web apps [J]. *Patterns*, 2021, 2(3): 100213.

[17] Szklarczyk D, Kirsch R, Koutrouli M, *et al*. The STRING database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest [J]. *Nucleic Acids Res*, 2023, 51(D1): D638-D646.

[18] 刘盛旺, 陈昳冰, 林静怡, 等. 夏枯草对乳腺疾病作用机制概述 [J]. 中国实验方剂学杂志, 2022, 28(5): 250-

255.

[19] 吴晓薇, 林锐珊. 熊果酸通过抑制 CDK2 活性诱导 HepG2 细胞凋亡 [J]. 中药材, 2014, 37(6): 1039-1042.

[20] Moparthi L, Kichko T I, Eberhardt M, *et al*. Human TRPA1 is a heat sensor displaying intrinsic U-shaped thermosensitivity [J]. *Sci Rep*, 2016, 6: 28763.

[21] Pozsgai G, Hajna Z, Bagoly T, *et al*. The role of transient receptor potential ankyrin 1 (TRPA1) receptor activation in hydrogen-sulphide-induced CGRP-release and vasodilation [J]. *Eur J Pharmacol*, 2012, 689(1/2/3): 56-64.

[22] 杜志峰. 孤阴不生, 独阳不长-系统生物学助推中医学发展的可行性分析及策略 [J]. 中西医结合心血管病电子杂志, 2019, 7(5): 25.

[23] 孙美堂, 崔馨丹. 基于复杂性思维的《矛盾论》再研究 [J]. 马克思主义哲学, 2026(1): 77-85.

[24] 陈亚容, 李纯, 李振南, 等. 基于 p38 MAPK/MK2 信号通路探讨桂林西瓜霜对口腔溃疡大鼠的治疗作用 [J]. 中国医院药学杂志, 2026, 46(12): 1381-1389.

[25] 徐明星,姜红梅. 石吊兰素调控 PI3K/AKT/HIF-1α 通路对胃癌 AGS 细胞增殖、自噬和糖代谢的影响 [J]. 中国老年学杂志, 2026, 46(7): 1301-1306.

[26] 黄姚姚, 申俊龙, 李洁. 逆境出质量：从种质与环境的相互作用探讨道地药材的质量形成机制 [J]. 中草药, 2025, 56(8): 2945-2953.

[27] 李聪, 黄诗雨, 陈丽华, 等. 药材部位、产地及采收期对中药挥发油成分的差异性分析 [J]. 中草药, 2020, 51(20): 5395-5404.

[28] 吴宏伟, 唐力英, 付梅红, 等. 不同煎煮条件下黄连解毒汤有效成分分析 [J]. 中国中医药信息杂志, 2010, 17(2): 42-44.

[29] 王晓娟, 季宇彬, 李志雄. 药对配伍对其成分药动学的影响 [J]. 中国医药导报, 2012, 9(5): 8-10.

[30] 阙晓慧, 桂蜀华, 钟清元, 等. 不同炮制程度地黄对小鼠卵巢的影响及其寒热药性相关成分分析 [J]. 中成药, 2021, 43(11): 3222-3226.

[31] 武旭, 王晶, 邹婷, 等. 天南星胆汁制前后药性变化的研究 [J]. 中药药理与临床, 2023, 39(6): 81-87.

[32] 官丽芳, 李赛玉, 董怡静, 等. 基于靶向胆汁酸代谢组学与分子对接技术的“人参-五味子”药对治疗肝细胞癌的药效物质与作用机制研究 [J]. 药学学报, 2026, 61(5): 1793-1803.

[33] Zhu F, Liu Y, Lei J H, *et al*. Single-cell eQTL mapping of human endogenous retroviruses reveals cell type-specific genetic regulation in autoimmune diseases [J]. *Nat Commun*, 2025, 16: 7534.

[34] 刘进娜, 谢鸣. “去性存用”配伍内涵的探讨 [J]. 北京中医药大学学报, 2011, 34(7): 444-447.

[责任编辑　潘明佳]