• 数据挖掘与循证医学 •

网络药理学参数与成分双扰动鲁棒性分析及其在草莓叶抗氧化活性成分研究中的应用

王宝军[1]，李树祥[1#]，赵志磊[1]，王宝刚[2*]，陈璐瑶[2*]

1. 河北大学质量技术监督学院，河北 保定 071002
2. 北京市农林科学院农产品加工与食品营养研究所，北京 100097

摘 要：目的 针对传统网络药理学固定阈值及输入成分构成易导致结果不稳定的问题，构建参数与成分双扰动鲁棒性框架，将单一条件下的靶点筛选转化为跨参数、跨成分的稳定性排序，并以筛选草莓叶抗氧化活性成分为例示范应用。**方法** 基于文献报道的液相色谱-串联质谱（liquid chromatography-tandem mass spectrometry，LC-MS/MS）注释结果，经结构规范化和去重获得 18 个化合物；采用 SwissTargetPrediction 预测靶点，整合 GeneCards、OMIM 和 TTD 构建氧化应激相关基因集合。围绕靶点预测概率、STRING 互作置信度和按节点度值排序选取前 *K* 个枢纽靶点（Hub Top K）开展 112 组全因子参数扰动，并逐一剔除 18 个化合物，完成 2 016 组成分扰动运行。依据跨参数出现频率、靶点平均归一化度值（degree）或功能条目平均富集强度筛选参数稳健结果；再根据成分剔除保留率确定基于预设操作性阈值的双重稳健候选结果；结合反向药效团匹配靶点预测平台（PharmMapper）和相似性集成方法（similarity ensemble approach，SEA）跨平台靶点预测、PubMed 共现、节点度匹配置换检验和重复分子对接进行评估。**结果** 共获得 684 个非重复预测靶点，其中 286 个预测概率大于 0 的靶点与氧化应激相关基因存在交集。候选靶点和功能富集结果随筛选阈值变化而波动。最终获得 50 个参数稳健靶点和 70 个参数稳健功能条目，其中 29 个靶点和 58 个功能条目依据预设操作性阈值，在成分扰动后被判定为具有相对稳健性。PharmMapper 和 SEA 分别为 13 个和 5 个基于预设操作性阈值判定的相对稳健候选靶点提供补充性计算支持，其中碳酸酐酶 2（carbonic anhydrase 2，CA2）和 CA12 同时获得 3 个靶点预测平台的支持。信号转导和转录激活因子 3（signal transducer and activator of transcription 3，STAT3）在全成分条件下的参数出现频率为 100%，且在全部成分剔除条件下均保持参数稳健状态；表皮生长因子受体（epidermal growth factor receptor，EGFR）、酪氨酸激酶 Src（proto-oncogene tyrosine-protein kinase Src，SRC）、肿瘤坏死因子（tumor necrosis factor，TNF）、热休克蛋白 90α 家族 A 类成员 1（heat shock protein 90 alpha family class A member 1，HSP90AA1）和 B 淋巴细胞瘤-2（B-cell lymphoma-2，BCL2）等亦具有较高稳定性。稳健功能条目主要涉及氧化应激响应、炎症调控、细胞凋亡和激酶信号转导等共享功能模块。**结论** 参数与成分双扰动框架可同时评估人为阈值和输入成分依赖性，从而提高网络药理学候选结果的稳定性与可审计性。

关键词：网络药理学；参数敏感性；鲁棒性分析；分子对接；靶点优先级；抗氧化；草莓叶

中图分类号：R285 **文献标志码：**A **文章编号：**0253 - 2670(2026)22 - 9013 - 22

DOI: 10.7501/j.issn.0253-2670.2026.22.017

Dual-perturbation robustness analysis for network pharmacology parameters and component and its application in antioxidant study of *Folia Fragariae*

WANG Baojun[1], LI Shuxiang[1], ZHAO Zhilei[1], WANG Baogang[2], CHEN Luyao[2]

1. College of Quality and Technology Supervision, Hebei University, Baoding 071002, China
2. Institute of Agri-Food Processing and Nutrition, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

Abstract: Objective To address the instability of conventional network pharmacology analyses arising from fixed screening thresholds and dependence on input compound composition, a dual-perturbation robustness framework was constructed to transform

收稿日期：2026-05-20

基金项目：北京市农林科学院科技创新能力建设专项（KJCX20261801，KJCX20261003）

作者简介：王宝军，副教授，研究方向为食品药品质量检测技术。E-mail: wbj@hbu.edu.cn

#共同第一作者：李树祥，硕士研究生，研究方向为数据科学。E-mail: lishuxiang09@outlook.com

***通信作者：**王宝刚，研究员，研究方向贮藏保鲜技术。E-mail: wangbaogang@baafs.net.cn

陈璐瑶，助理研究员，研究方向农产品贮藏与加工。E-mail: chenluyao@baafs.net.cn

the target screening under a single condition into stability ranking across parameters and components. The application was demonstrated by taking the screening of antioxidant active components from Caomeiye (*Folia Fragariae*) as an example. **Methods** Eighteen compounds were retained after structural standardization and deduplication of previously reported liquid chromatography-tandem mass spectrometry (LC-MS/MS) annotations. Potential targets were predicted using SwissTargetPrediction, and an oxidative stress-related gene set was compiled from GeneCards, OMIM, and TTD. Based on the predicted probability of the target, the confidence of STRING interaction, and the ranking by node degree value, the Top *K* Hub targets (Hub Top K) were selected to carry out 112 sets of full-factor parameter perturbations. Each of the 18 compounds was then excluded in turn, resulting in 2 016 compound-perturbation runs. The robust results of the parameters were screened based on the cross-parameter occurrence frequency, the average normalization degree value (degree) of the target, or the average enrichment intensity of the functional items. Dual-robust candidates were subsequently defined based on retention rates following compound exclusion under prespecified operational thresholds. Complementary evaluations included cross-platform target prediction using PharmMapper and similarity ensemble approach (SEA), PubMed co-occurrence analysis, degree-matched permutation testing, and repeated molecular docking. **Results** A total of 684 nonredundant predicted targets were obtained, 286 with positive prediction probabilities overlapped the oxidative stress-related gene set. Candidate targets and functional enrichment results varied across screening thresholds. A total of 50 parameter-robust targets and 70 parameter-robust functional terms were identified, of which 29 targets and 58 functional terms were classified as relatively robust based on prespecified operational thresholds after compound perturbation. PharmMapper and SEA provided complementary computational support for 13 and five relatively robust candidate targets identified based on prespecified operational thresholds, respectively, carbonic anhydrase 2 (CA2) and CA12 were identified by all three target-prediction platforms. Under the full-compound condition, the parameter occurrence frequency of signal transducer and activator of transcription 3 (STAT3) was 100%, and the parameter maintained robust status under every compound-exclusion condition. Epidermal growth factor receptor (EGFR), proto-oncogene tyrosine-protein kinase Src (SRC), tumor necrosis factor (TNF), heat shock protein 90 alpha family class A member 1 (HSP90AA1), and B-cell lymphoma-2 (BCL2) also showed high stability. The robust functional terms primarily represented shared modules related to oxidative stress responses, inflammatory regulation, apoptosis, and kinase signaling. **Conclusion** The parameter and component dual-perturbation framework can simultaneously evaluate artificial thresholds and input component dependencies, thereby improving the stability and auditability of candidate prioritization in network pharmacology.

Key words: network pharmacology; parameter sensitivity; robustness analysis; molecular docking; target prioritization; antioxidant activity; *Folia Fragariae*

网络药理学通过整合化合物、靶点、蛋白互作及功能富集信息，常用于解析天然产物和中药复方多成分、多靶点、多功能模块的潜在作用机制[1-4]。然而，其分析结果具有明显的流程依赖性。首先，化合物结构规范化、靶点预测算法及数据库覆盖范围均可能影响上游靶点集合，不同预测工具所得结果也可能存在差异，单一平台预测因而可能带来算法和数据库偏倚。其次，传统流程通常采用固定的靶点预测概率、蛋白质相互作用（protein-protein interaction，PPI）置信度及枢纽（Hub）基因数量阈值，而这些阈值缺乏统一标准；阈值变化可进一步影响候选靶点组成、网络拓扑排序及功能富集结果，使单一参数条件下得到的“核心靶点”仅代表参数空间的一个局部结果。再次，PPI 网络中的高连接度节点（靶点）可能同时受到网络拓扑特征、数据库收录程度及文献关注度影响；功能富集结果也受背景基因集、数据库版本和显著性阈值制约，部分以疾病命名的通路实际反映炎症、氧化应激、凋亡及激酶信号转导等共享功能模块，不能直接等同于特定疾病适应证或因果机制。此外，计算预测结果主要用于提出机制假设，仍需结合外部证据、统计检验和实验研究验证[5-8]。

除上述参数和数据库依赖性外，天然产物网络药理学还可能受到输入成分集合的影响。植物样本中不同化合物对靶点网络的贡献通常并不均衡，完整成分集合中排名靠前的靶点可能主要由单个或少数成分驱动。若缺少成分扰动分析，便难以区分代表多成分体系的稳定信号与高度依赖特定输入成分的敏感信号。因此，网络药理学研究不仅需要回答“筛选出哪些候选靶点”，还应进一步评估候选结果对参数设定、输入成分变化和网络拓扑偏倚的稳健性。现有研究较多关注网络药理学流程规范化、数据库质量和外部证据验证[9-10]。近年来，随着计算模型可重复性和不确定性评价理念的发展，针对分析流程参数选择、输入变量变化及结果稳定性的敏感性分析逐渐受到重视[11-12]。然而，针对网络

药理学中参数选择与成分构成 2 类不确定性进行联合量化的研究仍相对有限，这正是本研究拟解决的主要方法学问题。

草莓叶 *Folia Fragariae* 是草莓种植过程中的重要副产物，含有黄酮、多酚、有机酸和含氮小分子等多类活性成分。既往研究显示，草莓叶及其提取物具有较高的多酚含量，并具有抗氧化、抗糖基化及抑制还原型烟酰胺腺嘌呤二核苷酸磷酸氧化酶 2 活性相关潜力[13-16]。已有研究基于液相色谱-串联质谱（liquid chromatography-tandem mass spectrometry，LC-MS/MS）、网络药理学和分子对接初步分析了草莓叶抗氧化作用的候选靶点及潜在机制[17]，但这些候选结果是否受参数选择和输入成分构成影响尚缺乏系统评估。氧化应激不仅表现为氧化还原稳态失衡[18]，还可通过活性氧及蛋白氧化修饰参与细胞信号转导和药理调控[19]，因而适合作为网络药理学鲁棒性分析的应用场景。

基于此，本研究将传统网络药理学中的候选靶点筛选重新定义为稳定性优先级排序问题，并以草莓叶抗氧化活性成分研究为例构建参数与成分双扰动鲁棒性分析框架（图 1）。首先，围绕 SwissTargetPrediction 概率（probability）、STRING 综合得分（combined score）和按节点度值排序选取前 *K* 个枢纽靶点（Hub Top K）构建 112 组全因子参数组合，以评估候选靶点和功能条目对筛选参数变化的稳定性。其次，对 18 个结构规范化后的化合物开展逐一剔除化合物分析（leave-one-compound-out），形成 2 016 组成分扰动运行，以识别多成分体系中的稳定信号及成分依赖性。此外，采用反向药效团匹配靶点预测平台（PharmMapper）和相似性集成方法（similarity ensemble approach，SEA）开展补充性多平台靶点预测交叉验证，以评价主分析候选靶点在不同预测原理下的可重复性。最后，结合 PubMed 共现审计、节点度匹配置换检验和重复分子对接，对候选结果的文献背景、网络拓扑偏倚及结构结合可能性进行补充评估。

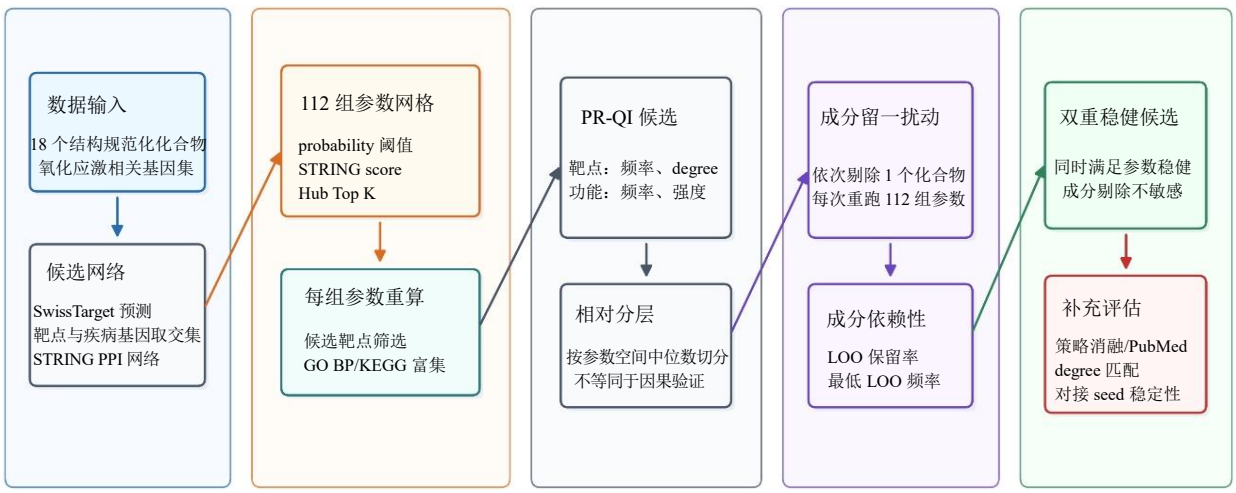


图 1　参数与成分双扰动鲁棒性分析框架

Fig. 1　Dual-perturbation robustness analysis framework for parameters and components

1　材料与方法

1.1　草莓叶活性成分整理

以文献报道的 LC-MS/MS 注释成分[17]为输入，原始记录共 19 条，其中海藻糖（trehalose）与 α,α-海藻糖（alpha,alpha-trehalose）被视为同一化学实体，经结构规范化去重后获得 18 个化合物。化合物名称统一后，通过 PubChem 获取 SMILES、CID 和结构信息[20]；对 PubChem 检索不稳定或无法可靠解析的化合物采用人工校准结构信息补充，以保证靶点预测输入一致。

1.2　靶点预测与疾病基因集合构建

将每个结构规范化后的化合物 SMILES 输入 SwissTargetPrediction，物种设定为 *Homo sapiens*[21]。对输出靶点进行基因符号规范化，并拆分包含多个基因符号的条目；同一基因在不同化合物中重复出现时，保留最高概率。在 2026 年 4 月 3～4 日，以“oxidative stress”为关键词分别检索 GeneCards[22]、OMIM[23]和 TTD[24]数据库，构建氧

化应激相关基因集合。GeneCards 未另设相关性评分（relevance score）截断值，纳入检索结果中基因符号（gene symbol）非空的全部条目，并保留基因符号、UniProt 数据库标识符（UniProt ID）、相关性评分和描述（description）等字段。OMIM 导出文件显示该检索共返回 2 739 个条目，受数据库下载数量限制，原始文件保留前 200 个检索结果；数据处理时仅纳入孟德尔人类遗传编号（Mendelian inheritance in man number，MIM number），以“*”开头且可从标题（title）字段解析出基因符号的基因条目，排除以“#”开头的疾病表型条目、无明确基因符号条目及其他非基因条目。TTD 纳入基因名称（gene names）字段非空且不为“N.A.”的靶点条目；对于包含多个基因名称的记录，按照逗号进行拆分，无明确基因名称的条目予以排除。随后清理各数据库中的基因符号并进行合并，按照基因符号精确匹配去重；同一基因同时来源于多个数据库时，按照 GeneCards、OMIM 和 TTD 的顺序优先保留相应记录，最终构建氧化应激相关基因集合。预测靶点与该基因集合取交集后，获得氧化应激相关候选靶点。

1.3　PPI 网络构建与 Hub 靶点筛选

疾病相关候选靶点被提交至 STRING 数据库构建 PPI 网络，物种设定为 *Homo sapiens*[25]。STRING 综合得分用于表示蛋白互作置信度。每组参数组合中，先按概率阈值筛选候选靶点，再按综合得分过滤互作边，构建无向加权 PPI 网络。随后按节点度值（degree）排序，选取 Hub Top K 进入富集分析。

1.4　参数扰动设计

本研究选择 SwissTargetPrediction 概率、STRING 综合得分和 Hub Top K 作为关键扰动参数。概率阈值设置为 0、0.03、0.05、0.08、0.10、0.15 和 0.20；STRING 综合得分阈值设置为 0.40、0.50、0.70 和 0.90；Hub Top K 设置为 10、20、30 和 50。富集分析校正后 *P* 阈值固定为 0.01。由此形成 7×4×4＝112 组参数组合。每组组合输出候选靶点数、PPI 节点数、PPI 边数、Top K 靶点、基因本体-生物过程（gene ontology-biological process，GO-BP）/基因组百科全书（Kyoto encyclopedia of genes and genomes，KEGG）显著功能富集条目及其功能分类结果。若筛选后靶点数、PPI 节点数或 Top K 基因数不足，则该组合被标记为跳过。

1.5　参数稳健第一象限候选（parameter-robust quadrant I candidate targets，PR-QI）靶点定义

对 112 组参数组合得到的 Top K 靶点计算每个基因的参数出现频率，表示其跨参数稳定性。同时，将每组 PPI 网络中节点度值除以该组合内最大度值，得到归一化度值，并计算每个基因的平均归一化度值。以参数出现频率和平均归一化度值构建二维鲁棒性空间。采取中位数分割方法，2 个指标均以当前参数网格中的中位数为分界，同时高于或等于中位数的靶点定义为 PR-QI 靶点。PR-QI 表示参数空间内的相对稳健优先级，不作为独立统计显著性检验。

1.6　GO-BP/KEGG 功能富集与功能条目鲁棒性定义

每组参数组合的 Top K 靶点用于 GO-BP[26]和 KEGG 富集分析[27]。鲁棒性统计仅纳入 GO-BP 和 KEGG 结果，关注参数扰动下抗氧化相关生物过程和信号网络解释是否稳定，后续将 GO-BP 和 KEGG 富集结果统称为功能富集条目，其中 KEGG 记录称为通路，GO-BP 记录称为生物过程；富集结果进一步按校正后 *P* 值、命中基因数和富集因子（enrichment factor）过滤[28]，并参考 Enrichr 交互式基因集富集分析工具的结果组织方式[29]：保留校正后 *P* 值＜0.01、命中基因数 $k \geqslant 2$ 且富集因子≥3.0 的条目，并在每个数据库内按校正后 *P* 值排序保留前 20 个条目。功能富集条目鲁棒性采用与靶点相同的二维思路定义。每个功能富集条目的跨参数出现频率表示稳定性，平均−lg 校正后 *P* 值表示富集强度。2 项指标均以中位数为分界，高频率且高富集强度的功能富集条目定义为 PR-QI 功能条目。

对筛选后的 GO-BP/KEGG 功能条目采用预设的互斥单标签规则进行分类，判定顺序如下：（1）直接氧化应激（direct oxidative stress），GO-BP 条目自身的 GO 编号或其在基因本体有向无环图（directed acyclic graph，DAG）中的至少一个祖先条目（即沿上位关系追溯得到的条目）命中预设氧化应激 GO 种子集合；（2）交叉调控（cross-talk），KEGG 条目根据 KEGG 功能层级分类体系（BRITE）归属于癌症、心血管疾病、神经退行性疾病、内分泌和代谢疾病、感染性疾病、免疫系统、内分泌系统、神经系统或细胞群落等类别，或精确命中预设的氧化应激交叉通路清单；（3）信号骨架（signaling backbone），KEGG 条目在 BRITE 中归属于信号转导（signal transduction）类别，或尚未归入前 2 类的

GO-BP/KEGG 条目名称命中丝裂原活化蛋白激酶（mitogen-activated protein kinase，MAPK）、磷脂酰肌醇-3-羟激酶（phosphatidylinositol-3-hydroxykinase，PI3K）-蛋白激酶 B（protein kinase B，Akt）、Janus 激酶（Janus kinase，JAK）/信号转导和转录激活因子（signal transducer and activator of transcription，STAT）、受体酪氨酸激酶、磷酸化或激酶级联等预设信号关键词；（4）其他，未满足上述规则的条目。若同一条目满足多个条件，按照直接氧化应激、KEGG BRITE 映射、交叉调控精确清单、信号骨架关键词和其他的固定优先级赋予唯一类别。分类由固定脚本自动批量完成，不使用条目的出现频率、富集显著性或留一法（leave-one-out，LOO）保留结果作为分类依据，也未进行逐条人工双盲判定；相同输入和规则可得到相同分类结果，完整种子集合、映射表、关键词及分类脚本可见复现代码。

1.7 成分 LOO 扰动分析

为评估输入成分集合对结果的影响，对 18 个结构规范化后的化合物逐一执行 leave-one-compound-out 分析，目的是每次剔除 1 个化合物后，重新构建靶点集合、疾病相关候选靶点集、PPI Top K 靶点和功能富集条目，也就是运行同一套 112 组参数组合，共形成 18×112＝2016 组成分扰动运行。

定义 LOO PR-QI 保留率＝（剔除某化合物后该靶点仍被识别为 PR-QI 的次数）/18。对每个剔除条件重新计算 PR-QI 靶点和 PR-QI 功能条目，并与全成分条件结果比较，若某一全成分 PR-QI 靶点的 LOO PR-QI 保留率不低于 0.80，且最低 LOO 参数频率大于 0，则定义为双重稳健候选靶点。功能条目采用相同标准定义双重稳健功能条目。

1.8 PharmMapper 和 SEA 靶点预测交叉验证

为评价 SwissTargetPrediction 主分析候选靶点在不同预测原理下的可重复性，本研究采用 PharmMapper 反向药效团匹配和 SEA 配体集合相似性方法开展补充性多平台靶点预测交叉验证[30-31]。该分析用于提供独立计算支持，不参与原 112 组参数扰动和 2 016 组成分扰动候选集合的重新定义，亦不等同于实验验证。后续实验验证优先级按验证目的分层：网络机制节点主要依据参数-成分双重稳健性排序，多平台一致性作为补充支持；直接结合或酶活验证则优先考虑同时具有双重稳健性和多平台预测支持的靶点。

PharmMapper 分析采用经 PubChem 核对的三维 SDF 结构，物种设为 Homo sapiens，固定保留每个化合物 best rank≤100 的靶点；SEA 分析采用经 PubChem 核对的 isomeric SMILES，保留返回结果中的 Homo sapiens 靶点，不另设 Top N 或评分截断。两平台输出均统一映射为标准基因符号，并分别按化合物-基因关系和基因符号去重，随后分别与 GeneCards、OMIM 和 TTD 合并得到的氧化应激相关基因集合取交集。

对两平台的氧化应激交集靶点分别构建 STRING PPI 网络，物种设为 Homo sapiens，最低互作置信度设为 0.700，不添加额外节点，并按 degree 排序筛选 Hub Top30。以 Hub Top30 靶点开展 GO BP 和 KEGG 富集分析，筛选标准与主分析一致，即 adjusted P-value＜0.01、命中靶点数≥2 且 enrichment factor≥3.0，各数据库按 adjusted P-value 排序保留前 20 个条目。最后，将两平台预测靶点及 Hub Top30 与 SwissTargetPrediction 主分析得到的 29 个双重稳健候选靶点进行比较，以评价跨平台计算支持情况。

1.9 关键判定参数的敏感性分析

以 LOO 保留率 0.80 和最低 LOO 参数频率＞0 作为基准判定条件。LOO 保留率 0.80 要求候选结果在 18 种成分剔除条件中的至少 15 种条件下保持 PR-QI 状态，既允许少数成分剔除对结果产生影响，又可排除高度依赖单一成分的候选结果。最低 LOO 参数频率＞0 要求候选结果在每种成分剔除条件下至少由 1 组参数组合支持。上述阈值均为操作性筛选条件，其对候选集合的影响通过敏感性分析进一步评价。

采用单因素变化法评估 PR-QI 分割阈值、LOO 保留率阈值及最低 LOO 参数频率阈值对分析结果的影响。Q_{40}、Q_{50} 和 Q_{60} 分别表示以 PR-QI 二维指标的第 40、第 50（中位数）和第 60 百分位数作为分割阈值；R_{70}、R_{80} 和 R_{90} 分别表示将 LOO 保留率阈值设为 0.70、0.80 和 0.90，对应至少在 13、15 和 17 种成分剔除条件下保持 PR-QI 状态；F_0、F_5 和 F_{10} 分别表示将最低 LOO 参数频率阈值设为＞0、＞0.05 和＞0.10，对应每种成分剔除条件下至少获得 1、6 和 12 组参数组合的支持。敏感性分析依次设置为固定 R_{80} 和 F_0，比较 Q_{40}、Q_{50} 与 Q_{60}；固定 Q_{50} 和 F_0，比较 R_{70}、R_{80} 与 R_{90}；固定 Q_{50} 和 R_{80}，比较 F_0、F_5 与 F_{10}。评价指标包括不同条件下双重稳健靶点和功能条目的数量、集合包含关系、两两 Jaccard 相

似系数及主要候选靶点的保留情况。主分析采用 Q_{50}、R_{80} 和 F_0 作为预设的操作性判定条件：Q_{50} 通过中位数进行相对分割，以避免依赖特定量纲的绝对界值，并兼顾候选集合的覆盖度与筛选严格度；R_{80} 要求候选条目至少在 15/18 种成分剔除条件下保持 PR-QI 状态，用于识别对多数单一成分缺失不敏感的结果；F_0 用于排除在任一成分剔除条件下参数出现频率降为 0 的候选条目。上述阈值仅用于相对稳健性排序，不具有统计学显著性含义，也不代表普遍适用的最优界值，其对结果的影响通过相邻阈值敏感性分析进行评价。

1.10 外部证据审计与消融分析

比较传统固定参数前 30 位、单一最佳参数前 30 位、PR-QI 稳健前 30 位的外部证据背景，包括 PubMed 共现。为避免靶点集合规模差异影响比例比较，策略横向比较统一采用 Top 30 口径；双重稳健候选靶点为成分扰动后保留的 29 个双重稳健靶点，作为压缩候选集合单独参照。PubMed 共现用于文献背景检查，不参与靶点筛选，也不作为鲁棒性策略优越性的主要判据；检索逻辑为基因符号与“oxidative stress”相关词共现，并设置≥10、≥50、≥100、≥200、≥500 和≥1 000 篇共现文献的阈值敏感性分析。采用超几何检验评估不同靶点集合相对于背景靶点集合的外部证据富集，并使用 Benjamini-Hochberg 方法校正错误发现率（false discovery rate，FDR）[32]。为进一步控制 PPI 度值偏差，进一步构建节点度匹配随机集合，执行 1 000 次置换检验，评估双重稳健候选靶点的文献背景是否超过相同度值分布的随机集合。此外，设置参数选择消融、富集质量比较和成分扰动消融。参数选择消融比较传统固定参数前 30 位、单一最佳参数前 30 位和 PR-QI 稳健前 30 位 3 种策略。富集质量从显著条目数、机制相关性、富集因子、−lg$P$ 和 Top 10 条目相关性进行评价。成分扰动消融计算每个化合物被剔除后的靶点损失、功能条目损失、PR-QI 靶点损失和 PR-QI 功能条目损失，并据此定义化合物影响评分。

1.11 分子对接分析

根据双重稳健候选靶点优先级及网络功能代表性，选择信号转导和转录激活因子 3（signal transducer and activator of transcription 3，STAT3）、表皮生长因子受体（epidermal growth factor receptor，EGFR）、酪氨酸激酶 Src（proto-oncogene tyrosine-protein kinase Src，SRC）、肿瘤坏死因子（tumor necrosis factor，TNF）、热休克蛋白 90α 家族 A 类成员 1（heat shock protein 90 alpha family class A member 1，HSP90AA1）和 B 淋巴细胞瘤-2（B-cell lymphoma-2，BCL2）与儿茶素（catechin）、槲皮素-3-*O*-葡萄糖醛酸苷（miquelianin）和原花青素 B1（procyanidin B1）进行分子对接。受体在蛋白质结构数据库（Protein Data Bank，PDB）中的结构编号分别为分别为 6NJS、1M17、2H8H、2AZ5、4LVT 和 3TUH。阳性对照采用各共晶结构的原配体 SD36（KQV）、厄洛替尼（erlotinib，AQ4）、H8H、307、纳维托克（navitoclax，1XJ）和甘特西匹（ganetespib，TUH），用于界定实验已解析口袋并进行原配体重对接。受体去除水分子和非目标杂原子，添加氢并赋予 Gasteiger 方法计算的原子部分电荷后转换为 PDBQT 格式。

配体以 PubChem 异构 SMILES 为起点并保留立体化学，采用 RDKit ETKDGv3（seed 20260716）生成三维构象，经 MMFF94 优化后，使用 Open Babel 按 pH 7.4 质子化并分配 Gasteiger 电荷。对接采用 AutoDock Vina 相关流程[33]，软件版本为 v1.2.0[34]，搜索详尽度（exhaustiveness）设为 16，每次输出 9 个构象。搜索盒以共晶原配体几何中心定义：STAT3（13.028, 55.637, 0.250；20 Å×30 Å×20 Å，1 Å＝0.1 nm）、EGFR（22.014, 0.253, 52.794；24 Å×20 Å×20 Å）、SRC（20.255, 20.924, 58.253；22 Å×20 Å×20 Å）、TNF（−19.410, 74.651, 33.850；20 Å×20 Å×20 Å）、BCL2（7.631, −2.985, −8.722；20 Å×34 Å×20 Å）和 HSP90AA1（22.126, −5.336, 8.222；20 Å×20 Å×20 Å）。每组使用 seed 1001、1002 和 1003 重复 3 次，以排名第 1 构象的 Vina 对接评分（Vina score）$\bar{x} \pm s$ 汇总。Vina score 仅为预测评分，不等同于实测结合自由能。

STAT3 口袋为 6NJS 中参与磷酸化酪氨酸识别和二聚化的 SH2 共晶位点；TNF 口袋为 2AZ5 中由 A/B 亚基形成、与小分子诱导三聚体结构扰动相关的界面共晶位点，因而均具有功能依据。原配体重对接以排名第 1 构象的重原子均方根偏差（root mean square deviation，RMSD）≤2.0 Å 为验证标准，并检查全部 9 个输出构象，以区分近天然构象的采样能力与评分排序能力。

2 结果

2.1 参数扰动下的靶点鲁棒性

参数扰动分析首先用于检验传统固定阈值流

程是否会产生稳定结果。18 个结构规范化后的草莓叶抗氧化活性化合物经 SwissTargetPrediction 预测后，共得到 684 个去重并集靶点；与 GeneCards、OMIM 和 TTD 整合得到的氧化应激疾病基因集合取交集后，获得 286 个正概率疾病相关候选靶点。112 组参数组合均成功完成。随着 SwissTargetPrediction 概率、STRING 综合得分和 Hub Top K 变化，不同参数组合得到的 Top K 靶点、显著功能富集条目数量和机制分类结果存在差异，表明候选结果对参数设定具有敏感性。

二维候选靶点鲁棒性分布如图 2 所示，部分靶点在多数参数组合中反复出现，并保持较高归一化度值。基于参数频率和平均归一化度值的二维筛选，本研究获得 50 个 PR-QI 靶点。PR-QI 靶点中，STAT3 的参数频率为 1.0，在全部 112 组参数组合中均进入 Top K。BCL2、基质金属蛋白酶 9（matrix metalloproteinase 9，MMP9）、前列腺素内过氧化物合酶 2（prostaglandin-endoperoxide synthase 2，PTGS2）、SRC、EGFR、半胱天冬酶 3（caspase 3，CASP3）、MMP2、HSP90AA1、TNF、MAPK1 和雌激素受体 1（estrogen receptor 1，ESR1）等靶点也具有较高参数频率和网络连接度，主要涉及氧化应

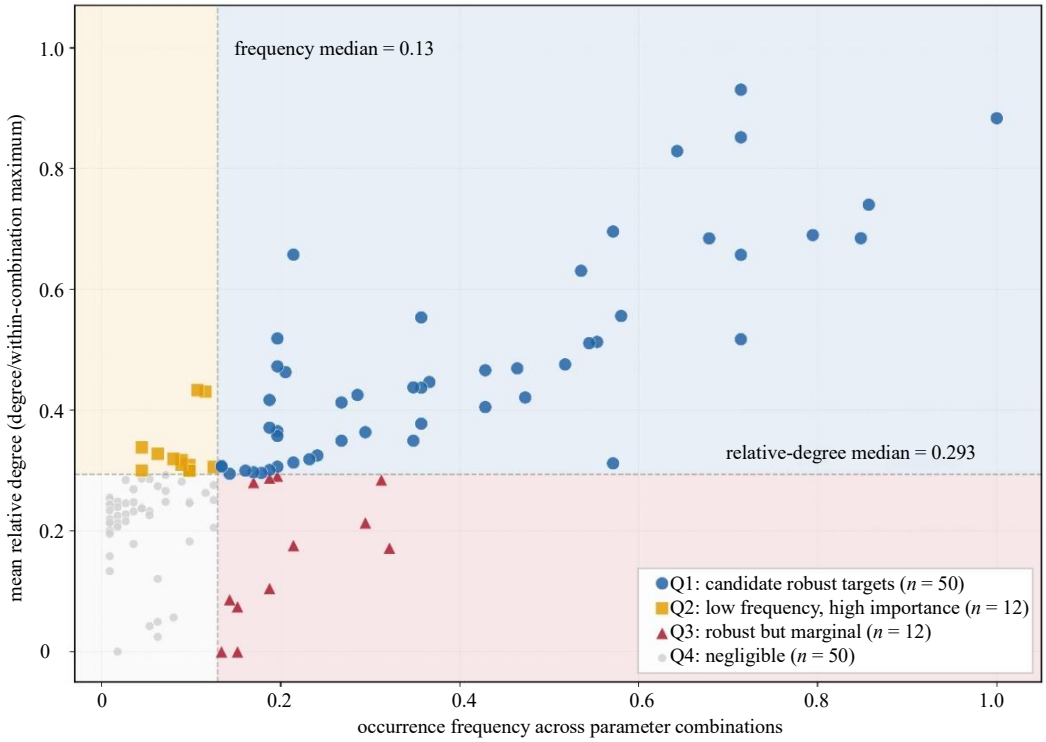


图 2　参数扰动下候选靶点鲁棒性分布

Fig. 2　Robustness distribution of candidate targets under parameter perturbation

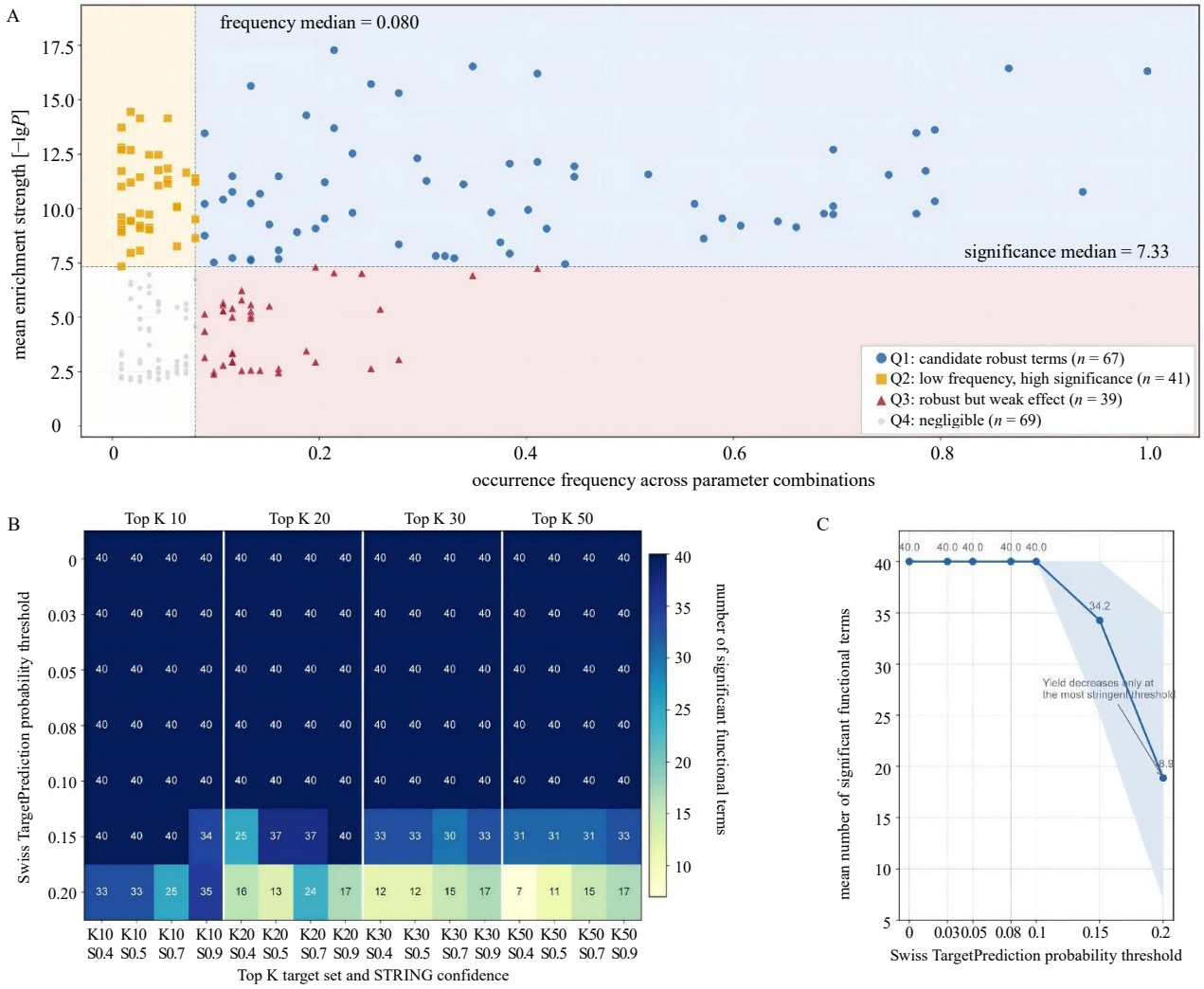
激、炎症反应、细胞凋亡、血管内皮功能、激酶信号和免疫调控。

与传统固定参数 Top 30 相比，PR-QI 靶点更强调跨参数复现性。固定参数 Top 30 中存在参数频率较低的靶点，最低参数频率仅为 0.036，提示部分候选可能是单一阈值下的偶然结果；PR-QI 稳健前 30 位的最低参数频率明显提高，说明鲁棒性筛选能够减少参数脆弱靶点。

2.2　参数扰动下的 GO-BP/KEGG 功能富集条目鲁棒性

在靶点层面确认参数敏感性后，研究进一步评估功能富集解释是否同样受参数扰动影响。功能富集条目包括 GO-BP 生物过程和 KEGG 通路；对每组参数组合得到的 Top K 靶点分别进行 GO-BP/KEGG 富集（图 3），结果显示草莓叶抗氧化活性成分候选靶点反复富集于氧化应激响应、炎症调控、凋亡调节及激酶信号转导等共享功能模块。为区分偶然出现的富集结果和跨参数稳定的功能解释，进一步统计每个 GO-BP/KEGG 功能富集条目的跨参数出现频率和平均–lg*P*。基于这 2 个指标的二维筛选，最终获得 70 个 PR-QI 功能条目。高频且高富集强度的条目被认为更能代表参数扰动下稳定保留的抗氧化相关功能模块。

癌症通路（pathways in cancer）、糖尿病并发症



A-跨参数组合的 GO-BP 功能条目稳健性分层散点图；B-不同 Top K 靶点集和 STRING 置信度下的 GO-BP 显著功能条目数量热图；C-SwissTargetPrediction 概率阈值对平均显著 GO-BP 功能条目数量的影响。

A-stratified scatter plot of GO-BP term robustness across parameter combinations; B-heatmap of the number of significant GO-BP terms across Top K target sets and STRING confidence thresholds; C-effect of SwissTargetPrediction probability threshold on the mean number of significant GO-BP terms.

图 3　参数扰动下 GO-BP/KEGG 功能富集条目的鲁棒性分布

Fig. 3　Robustness distribution of GO-BP/KEGG functional enrichment terms under parameter perturbation

中的晚期糖基化终末产物-受体信号通路（advanced glycation end products-receptor for advanced glycation end products signaling pathway in diabetic complications，AGE-RAGE signaling pathway in diabetic complications）、乙型肝炎（hepatitis B）、脂质与动脉粥样硬化（lipid and atherosclerosis）等条目具有较高出现频率和富集强度。上述条目名称来源于 KEGG 通路分类，其共同涉及氧化应激响应、炎症调控、凋亡调节及激酶信号转导等功能模块，不代表草莓叶活性成分对相应疾病具有治疗作用。

2.3　成分扰动下的双重稳健子集

参数扰动能够筛出对阈值变化较稳定的 PR-QI 靶点，但天然产物网络药理学还存在另一类不确定性，即结果可能依赖某一个或少数输入成分。为进一步判断参数稳健靶点是否也能抵抗成分集合变化，本研究对 18 个结构规范化后的化合物执行 leave-one-compound-out 分析。该分析的目的不是进行实验验证，而是从输入成分依赖性角度解析多成分体系中的稳定机制信号。每次剔除 1 个化合物后，重新运行完整的 112 组参数组合，共形成 2 016 组成分扰动运行，所有组合均成功完成。

结合全成分条件下的 PR-QI 靶点与各剔除条件下的 PR-QI 保留情况，筛选得到 29 个双重稳健候选靶点。该集合同时满足参数扰动稳定和成分扰动稳

定，因此比单一全成分条件下的 PR-QI 靶点更适合作为后续机制解释和实验验证的候选集合。同时功能条目分类结果显示，70 个 PR-QI 功能条目中，直接氧化应激、交叉调控、信号骨架和其他类分别为 1 个（1.4%）、20 个（28.6%）、19 个（27.1%）和 30 个（42.9%）。经成分 leave-one-compound-out 扰动分析后，其中 58 个同时满足 PR-QI 和预设 LOO 保留标准，被定义为双重稳健功能条目；其 4 类分布分别为 0 个、20 个（34.5%）、15 个（25.9%）和 23 个（39.7%）。前 10 位双重稳健候选靶点的鲁棒性指标见表 1。

表 1　前 10 位双重稳健候选靶点的鲁棒性指标

Table 1　Robustness metrics for top 10 dual-robust candidate targets

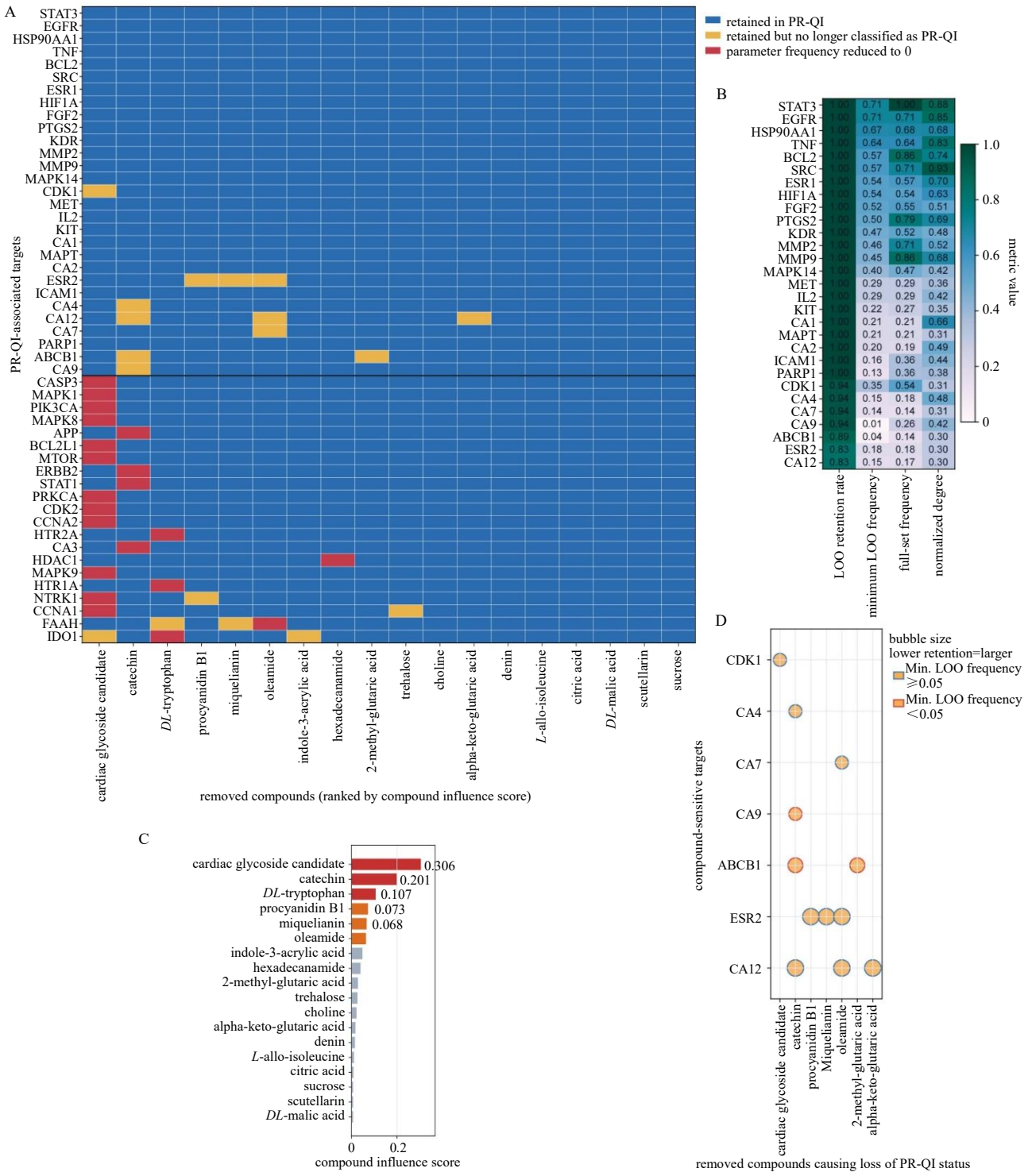
靶点	LOO 保留率	最低 LOO 频率	全成分频率	平均归一化度值	主要缺失化合物
STAT3	1.000	0.714	1.000	0.883	−
EGFR	1.000	0.714	0.714	0.852	−
HSP90AA1	1.000	0.670	0.679	0.684	−
TNF	1.000	0.643	0.643	0.829	−
BCL2	1.000	0.571	0.857	0.740	−
SRC	1.000	0.571	0.714	0.931	−
ESR1	1.000	0.536	0.571	0.696	−
HIF1A	1.000	0.536	0.536	0.631	−
FGF2	1.000	0.518	0.554	0.513	−
PTGS2	1.000	0.500	0.795	0.690	−

LOO 保留率＝删除单个化合物后仍保留为 PR-QI 靶点的次数/18；最低 LOO 频率＝在所有“删除一个化合物”的条件中，该靶点进入 PR-QI 靶点集合的参数频率最小值；STAT3 为 Tier 1 靶点，其参数频率为 100%，LOO 保留率为 100%；LOO 保留率＜1.0 表示该靶点在特定化合物剔除后退出 PR-QI 状态；“−”表示未发现使该靶点退出 PR-QI 状态的单一剔除化合物。

LOO retention rate was defined as the proportion of the 18 leave-one-compound-out conditions in which a target retained PR-QI status; Minimum LOO parameter frequency was the lowest parameter-selection frequency of a target across all LOO conditions. STAT3 was classified as a Tier 1 target because both its full-compound parameter-selection frequency and LOO retention rate were 100%; LOO retention rate below 1.0 indicates loss of PR-QI status after exclusion of at least one compound; “–” indicates that no single-compound exclusion caused the target to lose PR-QI status.

在双重稳健候选靶点中，STAT3 表现出最强稳定性：其在全参数空间中的出现频率为 1.0，并在所有成分剔除条件下均保持 PR-QI 状态，因此，按照预设的网络稳定性分级规则，将其定义为 Tier 1 候选靶点。该分级仅依据参数出现频率和成分剔除保留率，不包含分子对接结果，也不表示 STAT3 已经获得直接结合或结构可药性验证。EGFR、SRC、TNF、HSP90AA1、BCL2、PTGS2 和 MMP9 等靶点也具有较高 LOO PR-QI 保留率，提示这些节点并非由单一参数组合或单一输入成分驱动。结合其已知生物学功能，这些双重稳健靶点共同指向炎症反应、激酶信号、细胞凋亡和氧化应激调控等功能模块。这一结果为后续机制解释提供了层级化候选集合：STAT3 更像参数和成分扰动下均稳定保留的下游转录调控枢纽，EGFR 和 SRC 则更可能代表上游激酶信号节点。后续外部证据审计和分子对接分析将进一步评估这些核心节点是否同时具有文献背景和结构结合可行性。在获得双重稳健靶点后，为进一步评估不同化合物对整体网络稳定性的贡献，若剔除某一化合物后靶点集合、功能富集条目或 PR-QI 结果明显下降，则说明该化合物对维持当前网络解释具有较高影响。因此定义化合物影响评分指标用于衡量单个化合物缺失对靶点和功能条目稳定性的综合影响。

图 4-A 展示不同成分剔除条件下 PR-QI 相关靶点的状态变化，图 4-B 汇总相应的 LOO 稳定性指标，图 4-C 给出化合物影响评分排序。如图 4-D 所示，29 个双重稳健候选靶点中有 22 个在 18 种成分剔除条件下始终保持 PR-QI 状态，仅 7 个出现成分特异性状态变化。剔除儿茶素与 CA4、CA9、ABCB1 和 CA12 退出 PR-QI 相关，剔除油酰胺（oleamide）与 CA7、ESR2 和 CA12 退出 PR-QI 相关；ABCB1、ESR2 和 CA12 分别对 2、3 和 3 种成分剔除敏感，而 CDK1 仅在剔除该强心苷类候选化合物时退出 PR-QI。需要说明的是，“退出 PR-QI”表示该靶点未再同时达到相对频率和拓扑阈值，并不等同于其从预测靶点集合中完全消失。成分-靶点敏感性关联总体呈稀疏且不均匀分布，提示多数核心节点具有多成分冗余支持，而部分接近判定边界的节点更依赖特定输入成分。影响评分最高的强心苷类候选化



A-不同剔除化合物对 PR-QI 相关靶点状态影响的 LOO 热图；B-PR-QI 相关靶点的 LOO 稳定性指标热图；C-剔除化合物的化合物影响评分排序；D-导致靶点退出 PR-QI 的关键化合物及成分敏感靶点分布。

A-LOO heatmap showing the effects of individual compound exclusions on PR-QI-related target status; B-heatmap of LOO stability metrics for PR-QI-related targets; C-ranking of excluded compounds by compound influence score; D-distribution of compound-sensitive targets and the key compounds whose exclusion caused targets to lose PR-QI status.

图 4　成分 leave-one-compound-out 扰动下双重稳健候选靶点的保留与成分依赖性

Fig. 4　Retention and compound dependence of dual-robust candidate targets under leave-one-compound-out perturbation

合物在图 4-D 中仅与 CDK1 的状态变化相关，说明整体网络影响度与特定核心靶点依赖性并非同一概念。

如表 2 所示，剔除强心苷类候选化合物后的化

表 2 前 10 位化合物影响度评分

Table 2 Influence scores of top 10 compounds

化合物	化合物影响评分	平均靶点保留率	平均功能条目保留率	PR-QI 靶点保留率	PR-QI 功能条目保留率
强心苷类候选化合物（cardiac glycoside candidate）	0.306	0.757	0.655	0.720	0.643
儿茶素（catechin）	0.201	0.859	0.768	0.840	0.729
DL-色氨酸（*DL*-tryptophan）	0.107	0.944	0.892	0.920	0.814
原花青素 B1（procyanidin B1）	0.073	0.939	0.836	0.960	0.971
槲皮素-3-*O*-葡萄糖醛酸苷（miquelianin）	0.068	0.955	0.900	0.960	0.914
油酰胺（oleamide）	0.065	0.946	0.917	0.920	0.957
吲哚-3-丙烯酸（indole-3-acrylic acid）	0.050	0.969	0.952	0.980	0.900
十六酰胺（hexadecanamide）	0.041	0.964	0.948	0.980	0.943
2-甲基戊二酸（2-methylglutaric acid）	0.029	0.975	0.971	0.980	0.957
海藻糖（trehalose）	0.027	0.980	0.988	0.980	0.943

化合物影响评分＝1－平均保留率，其中平均保留率为靶点保留率、功能条目保留率、PR-QI 靶点保留率和 PR-QI 功能条目保留率 4 项指标的均值；得分越高表示该化合物缺失对网络稳定性的影响越大；强心苷类候选化合物为 (3β,9ξ)-3-(β-*D*-吡喃葡萄糖氧基)-14-羟基卡-20(22)-烯内酯（$C_{29}H_{44}O_9$，相对含量 2.84%[17]），其归属依据为高分辨 LC-MS/MS 数据库匹配；因未公开该条目的保留时间、前体 *m*/*z*、碎片离子及对照品验证信息，本研究按 Level 3 暂定候选结构处理。

Compound influence score = 1－mean retention rate, where mean retention rate is the average of the target, functional-term, PR-QI target, and PR-QI functional-term retention rates. Higher scores indicate greater network sensitivity to compound exclusion. The cardiac glycoside candidate was annotated as (3β,9ξ)-3-(β-*D*-glucopyranosyloxy)-14-hydroxycard-20(22)-enolide ($C_{29}H_{44}O_9$, relative abundance 2.84%[17]), the annotation was based on high-resolution LC-MS/MS database matching; because compound-specific retention time, precursor *m*/*z*, fragment ions, and reference-standard validation were not reported, it was classified as a Level 3 tentative candidate in the present study.

合物影响评分最高（0.306），表明网络结果对该输入条目的缺失较敏感。儿茶素、*DL*-色氨酸（*DL*-tryptophan）、原花青素 B1 和槲皮素-3-*O*-葡萄糖醛酸苷的影响度依次降低。该指标仅反映计算模型对输入成分缺失的敏感性，不代表相应成分的实际含量、药理活性或生物学贡献。成分扰动结果仅用于描述网络稳定性，并为代表性对接配体的选择提供计算依据。然后进一步考察剔除结果其平均靶点保留率、平均功能条目保留率、PR-QI 靶点保留率和 PR-QI 功能条目保留率分别为 0.757、0.655、0.720 和 0.643。在 29 个双重稳健候选靶点中，28 个在该剔除条件下仍保持 PR-QI 状态，仅 CDK1 转为稳定但外围的 Q3 类型。STAT3 仍保持 PR-QI 状态，EGFR、SRC、TNF、HSP90AA1、BCL2、PTGS2 和 MMP9 等主要候选靶点也均被保留。因此，排除该强心苷类暂定候选结构后，主要候选靶点的双重稳健性结论仍然成立。

2.4 关键判定参数的敏感性分析结果

PR-QI 分割阈值敏感性分析显示，Q_{40}、Q_{50} 和 Q_{60} 条件下分别获得 34、29 和 24 个双重稳健候选靶点，以及 69、58 和 41 个双重稳健功能条目。随着分割标准趋于严格，候选集合规模逐渐减小，且不同阈值下的候选集合呈 $Q_{60}\subseteq Q_{50}\subseteq Q_{40}$ 的嵌套关系。不同分割阈值下双重稳健靶点集合的两两 Jaccard 相似系数为 0.706～0.853，功能条目集合的两两 Jaccard 相似系数为 0.594～0.841。

当 LOO 保留率分别设为 0.70、0.80 和 0.90 时，双重稳健候选靶点数量分别为 29、29 和 26 个，双重稳健功能条目数量分别为 60、58 和 47 个。R_{70} 与 R_{80} 条件下的靶点集合完全一致，R_{80} 与 R_{90} 条件下靶点集合的 Jaccard 相似系数为 0.897。最低 LOO 参数频率分别设为＞0、＞0.05 和＞0.10 时，双重稳健候选靶点数量分别为 29、27 和 27 个，双重稳健功能条目数量分别为 58、50 和 39 个；不同最低 LOO 参数频率下靶点集合的两两 Jaccard 相似系数为 0.931～1.000。

总体而言，功能条目集合对判定阈值变化相对敏感，而靶点集合保持了较高一致性。STAT3、EGFR、SRC、TNF、HSP90AA1、BCL2、PTGS2 和 MMP9 在上述不同判定条件下均被保留，表明主要候选靶点的识别不依赖于单一分割方式或 LOO 判定阈值。关键判定参数的敏感性分析结果见表 3。

2.5 鲁棒性策略与传统策略的消融比较

在获得 PR-QI 和双重稳健候选靶点后，研究进一步通过消融比较检验鲁棒性筛选策略是否优于传统固定参数流程。首先控制靶点集合大小均为

表 3　关键判定参数的敏感性分析结果

Table 3　Sensitivity analysis results of key determination parameters

敏感性因素	判定条件	双重稳健靶点数	靶点 Jaccard	双重稳健功能条目数	功能条目 Jaccard
PR-QI 分割阈值	Q40	34	0.853	69	0.841
PR-QI 分割阈值	Q50	29	1.000	58	1.000
PR-QI 分割阈值	Q60	24	0.828	41	0.707
LOO 保留率	R70	29	1.000	60	0.967
LOO 保留率	R80	29	1.000	58	1.000
LOO 保留率	R90	26	0.897	47	0.810
最低 LOO 参数频率	F0	29	1.000	58	1.000
最低 LOO 参数频率	F05	27	0.931	50	0.862
最低 LOO 参数频率	F10	27	0.931	39	0.672

Jaccard 相似系数均以相应因素的主分析条件为参照，即 PR-QI 分割阈值以 Q_{50} 为参照、LOO 保留率以 R_{80} 为参照、最低 LOO 参数频率以 F_0 为参照；由于 18 种成分剔除条件和 112 组参数组合均为离散计数，名义阈值对应的最低保留数分别为 13/18、15/18 和 17/18，最低参数支持数分别为 1/112、6/112 和 12/112。

Jaccard similarity was calculated relative to the main-analysis setting for each criterion: Q_{50} for the PR-QI split threshold, R_{80} for the LOO retention threshold, and F_0 for the minimum LOO parameter-frequency threshold. Because the 18 LOO conditions and 112 parameter combinations yield discrete counts, the nominal thresholds correspond to minimum retention counts of 13/18, 15/18, and 17/18 and minimum parameter-support counts of 1/112, 6/112, and 12/112, respectively.

Top 30，同时比较 3 种 Top 30 策略、全部双重稳健候选靶点在完整参数空间中的连续型稳定性分布。比较对象包括传统固定参数 Top 30、单一最佳参数 Top 30 以及 PR-QI 靶点内综合频率、归一化度值、平均排名和介数中心性得到的 PR-QI 稳健前 30 位。双重稳健候选靶点的全部 29 个靶点可作为参照分析。

如表 4 所示，传统固定参数前 30 位虽可捕捉部分核心靶点，但包含较多低复现节点。单一最佳参数前 30 位的整体稳定性提高，但依赖 1 次“最佳参数组合”的选择。PR-QI 稳健前 30 位不包含参数频率低于 0.25 的靶点，其中 PR-QI 稳健前 30 位在保持较高参数频率的同时具有最高平均归一化度值，更适合作为参数稳健候选靶点的代表性 Top 30 集合，双重稳健候选靶点（29 个）效果略差于 PR-QI 稳健前 30 位。各策略的靶点网络、参数频率分布、平均归一化度值及低频率靶点数量分别见图 5-A～D。

靶点稳定性提高并不必然意味着功能解释更合理，所以进一步比较这几种策略的 GO-BP/KEGG 功能富集质量，如图 6 和表 5 所示。该比较用于判断鲁棒性筛选是否在提高参数复现性的同时，仍能保留抗氧化相关生物过程和信号网络解释。4 种策略均获得 40 个显著条目，但机制相关性和统计强度不同。传统固定参数前 30 位平均富集因子（enrichment factor，EF）较高，但 Top 10 条目中仅

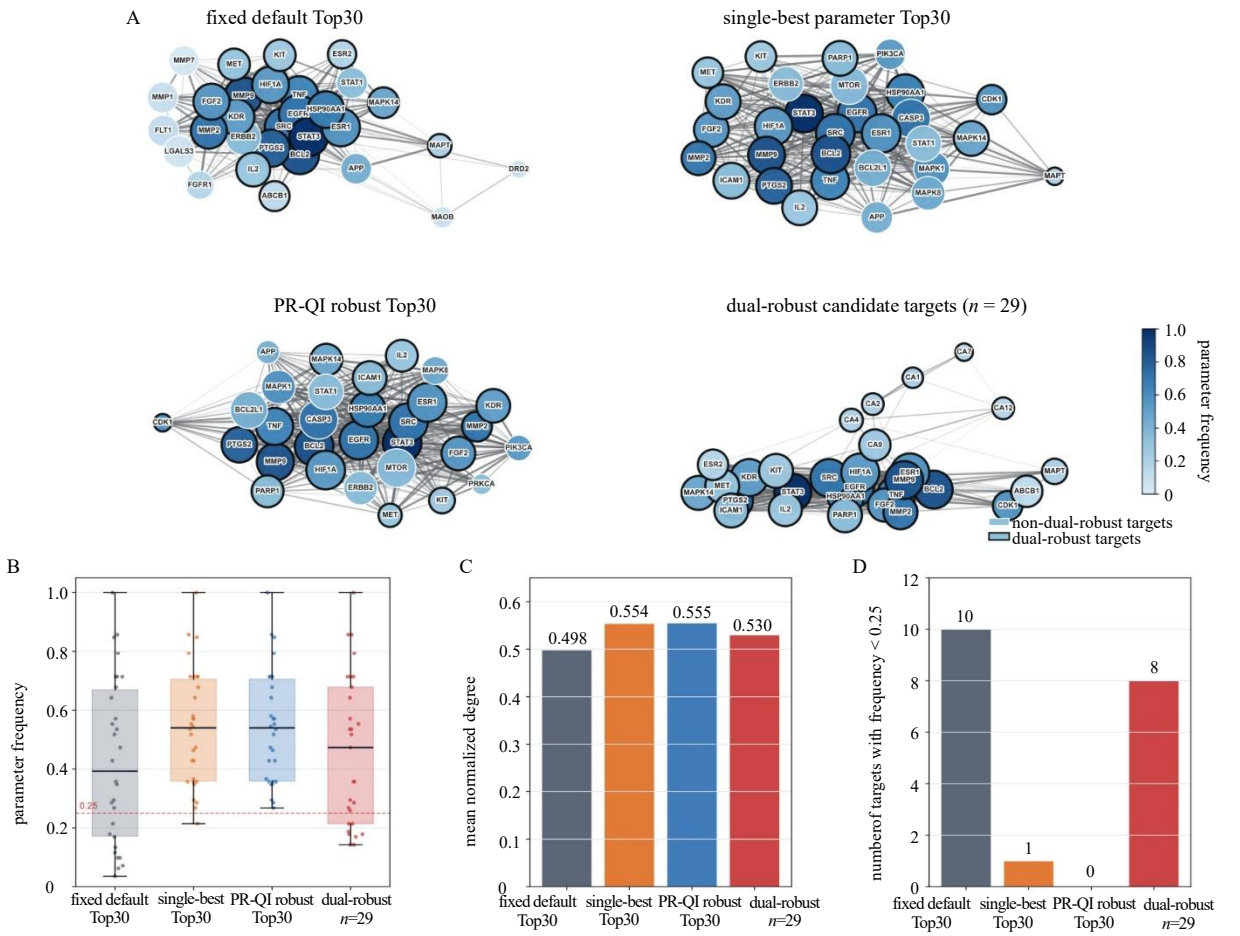
表 4　参数选择策略消融分析

Table 4　Ablation analysis of parameter-screening strategies

策略	平均参数频率	最低参数频率	频率＜0.25 靶点数	频率≥0.50 靶点数	平均归一化度值
传统固定参数前 30 位	0.426	0.036	10	13	0.498
单一最佳参数前 30 位	0.539	0.214	1	17	0.554
PR-QI 稳健前 30 位	0.543	0.268	0	17	0.555
双重稳健候选靶点（29 个）（参照）	0.462	0.143	8	14	0.530

传统固定参数采用概率＞0.05、STRING 综合得分≥0.40 和度值 Top 30 的传统阈值；单一最佳参数表示按氧化应激相关富集条目选择单一最佳参数组合后取度值 Top 30；PR-QI 稳健策略表示在 PR-QI 靶点内按参数频率、归一化度值、平均排名和介数中心性综合排序取前 30 位；双重稳健候选靶点为成分 leave-one-compound-out 扰动后保留的 29 个双重稳健靶点，作为最终压缩候选集合参照，不参与统一 Top 30 主比较。

Fixed default used conventional thresholds of prediction probability > 0.05, STRING combined score ≥ 0.40, and the top 30 targets by degree; single-best comprised the top 30 targets by degree under the parameter combination yielding the strongest oxidative stress-related enrichment; PR-QI robust comprised the top 30 PR-QI targets ranked by parameter-selection frequency, normalized degree, mean rank, and betweenness; 29 dual-robust candidate targets retained after leave-one-compound-out perturbation were included only as a compressed reference set and were not part of the size-matched Top 30 comparison.

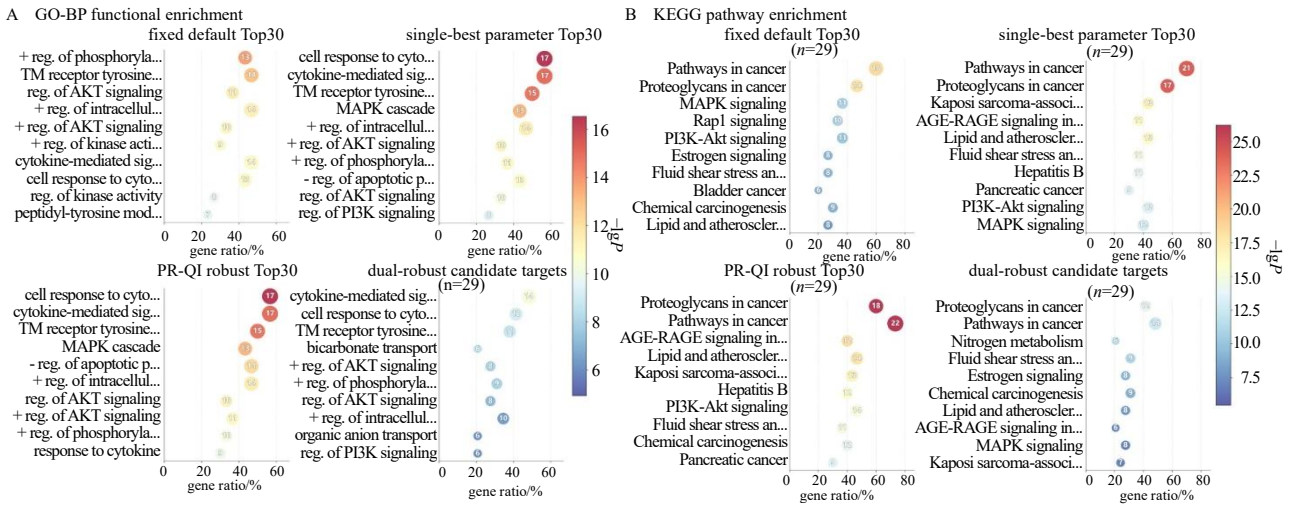


A-不同靶点筛选策略的 Top 靶点 PPI 网络图；B-不同靶点筛选策略的参数敏感性频率分布箱线图；C-不同靶点筛选策略的平均归一化度值比较；D-不同靶点筛选策略中低参数频率靶点数量比较。

A-PPI networks of top-ranked targets obtained using different target-screening strategies; B-box plots of parameter sensitivity frequency distribution for different target screening strategies; C-comparison of mean normalized degree of different target screening strategies; D-comparison of the numbers of low-parameter frequency targets in different target screening strategies.

图 5　不同靶点筛选策略的参数稳定性比较

Fig. 5　Comparison of parameter stability across target-screening strategies



A-不同靶点筛选策略的 GO-BP 功能富集气泡图；B-不同靶点筛选策略的 KEGG 通路富集气泡图。

A-GO-BP enrichment bubble plots across different target-screening strategies; B-KEGG pathway enrichment bubble plots across different target-screening strategies.

图 6　不同靶点筛选策略的 GO-BP/KEGG 功能富集比较

Fig. 6　Comparison of GO-BP/KEGG functional enrichment across different target-screening strategies

表 5 GO BP/KEGG 功能富集质量横向比较

Table 5 Horizontal comparison of GO-BP/KEGG functional enrichment metrics

靶点集合	相关条目数	功能分类 (ox/ct/bb)	Top 10 相关	平均 EF	平均−lgFDR
传统固定参数前 30 位	26	0/17/9	6	47.23	9.75
单一最佳参数前 30 位	30	2/15/13	10	39.59	12.30
PR-QI 稳健前 30 位	27	0/13/14	10	43.59	12.99
双重稳健候选靶点（29 个）	24	0/16/8	8	41.37	7.19

ox、ct 和 bb 分别表示直接氧化应激、交叉调控和信号骨架功能条目的数量；EF 表示富集因子，传统固定参数的平均 EF 较高，但机制指向相对分散；单一最佳参数的机制相关性较高，但可能存在参数选择偏倚；PR-QI 稳健策略在统计强度和拓扑稳健性之间较为均衡。

ox, ct, and bb denote the numbers of terms classified as direct oxidative stress, crosstalk, and signaling backbone, respectively; EF indicate enrichment factor, fixed default showed a higher mean EF but less coherent mechanistic relevance; single-best parameter showed stronger mechanistic relevance but may be affected by parameter-selection bias; PR-QI robust provided a more balanced profile of statistical strength and topological robustness.

6 个属于氧化应激相关信号网络，且无直接氧化应激条目。单一最佳参数前 30 位的氧化应激相关条目最多，并命中对活性氧的响应（response to reactive oxygen species）和细胞对氧化应激的响应（cellular response to oxidative stress），但该策略按富集表现选择参数组合，存在选择偏倚。PR-QI 稳健前 30 位的平均−lgFDR 最高，统计显著性较强。Top 10 条目均为氧化应激相关信号网络，且兼顾参数复现性和网络拓扑信息。

综合来看，PR-QI 稳健前 30 位在避免单一最佳参数选择偏倚的同时，保持较高 Top 10 机制集中度和统计强度，更适合用于挖掘作用机制。双重稳健候选靶点在部分统计指标上整体弱于 PR-QI 稳健前 30 位，这与本研究预期一致。PR-QI 稳健前 30 位是在全成分条件下根据参数频率、归一化度值、平均排名和介数中心性综合排序得到的代表性高稳定靶点集合，其筛选目标偏向保留参数空间中最强的网络中心节点；而双重稳健候选靶点是在参数稳健基础上进一步叠加成分 leave-one-compound-out 约束后形成的收缩集合，会排除一部分在全成分条件下表现较强但依赖特定化合物支撑的高分靶点，因此，双重稳健集合的平均参数频率、平均归一化度值或部分富集质量指标低于 PR-QI 稳健前 30 位。换言之，PR-QI 稳健前 30 位主要用于体现参数稳健筛选的代表性优势，双重稳健候选靶点则用于机制解释和后续实验验证候选的进一步收敛。

2.6 PharmMapper 和 SEA 靶点预测交叉验证结果

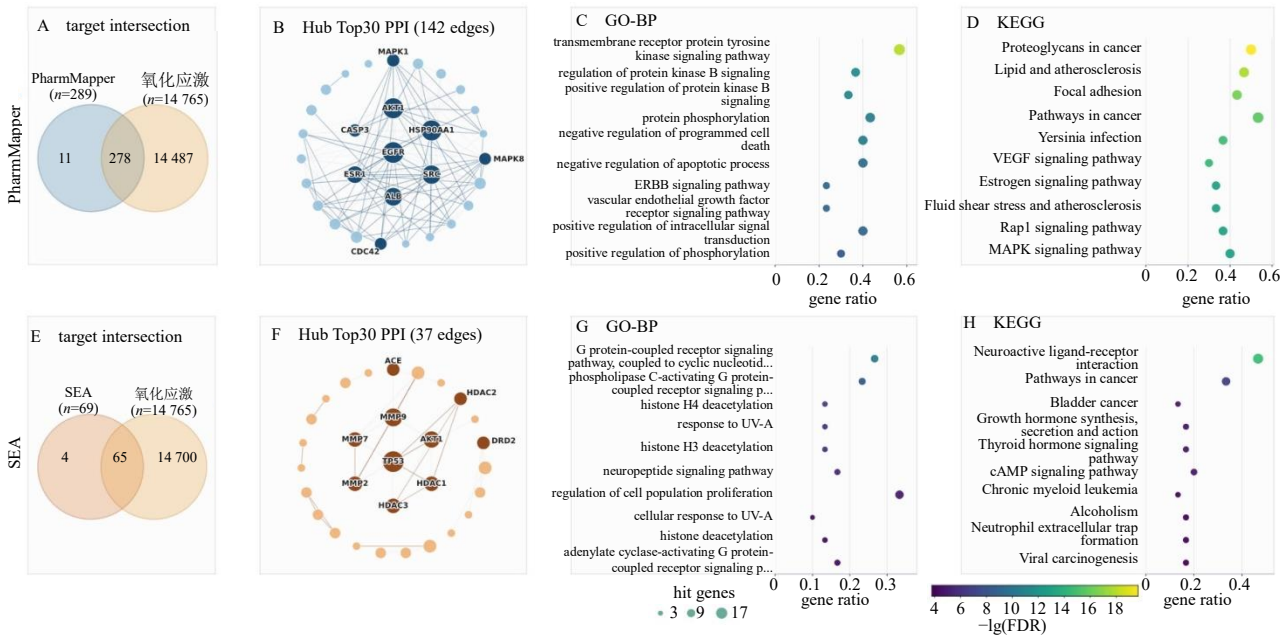
18 个化合物中 17 个获得 PharmMapper 预测结果，胆碱（choline）未获得可报告匹配。按每个化合物最佳排名（best rank）≤100 筛选并合并去重后，共得到 289 个非重复人源靶点，其中 278 个与氧化应激相关基因集合重叠（图 7-A）。在 STRING 综合得分≥0.700 条件下，242 个节点形成 751 条无向互作边；按度值筛选的 Hub Top30 子网络包含 142 条互作边。PharmMapper 预测集合直接支持 29 个双重稳健候选靶点中的 13 个，其中 EGFR、ESR1、HSP90AA1、SRC 等靶点同时进入 PharmMapper Hub Top 30 子网络（图 7-B）。

SEA 分析中，10 个化合物获得人源靶点，野黄芩苷（scutellarin）虽有返回结果但无人源靶点，另有 7 个化合物未获得预测结果。合并去重后得到 69 个非重复人源靶点，其中 65 个与氧化应激相关基因集合重叠（图 7-E）。在 STRING 综合得分≥0.700 条件下，36 个节点形成 42 条无向互作边，Hub Top30 子网络包含 37 条互作边。SEA 预测集合直接支持 CA2、CA9、CA12、MMP2 和 MMP9 共 5 个双重稳健候选靶点，其中 MMP2 和 MMP9 等靶点进入 SEA Hub Top 30（图 7-F）；CA2 和 CA12 同时获得 SwissTargetPrediction、PharmMapper 和 SEA 3 个平台支持。

功能富集结果（图 7-G、H）显示，PharmMapper Hub Top 30 主要涉及跨膜受体蛋白酪氨酸激酶信号、蛋白磷酸化、细胞凋亡调节以及 ERBB 受体家族（Erb-B receptor tyrosine kinase family，ERBB）、血管内皮生长因子（vascular endothelial growth factor，VEGF）和 MAPK 等信号模块；SEA Hub Top 30 主要涉及 G 蛋白偶联受体、环核苷酸第 2 信使、细胞增殖和环磷酸腺苷（cyclic adenosine monophosphate，cAMP）等信号模块。富集结果中的疾病命名通路主要反映氧化应激响应、炎症调控、凋亡调节及激酶信号转导等共享功能模块，不作为相应疾病适应证的直接证据。

2.7 外部证据审计与分子对接分析

在参数扰动和成分扰动筛选后，本研究进一步



A-PharmMapper 预测靶点与氧化应激相关靶点的交集；B-PharmMapper 交集靶点的 Hub Top30 蛋白质相互作用网络；C-PharmMapper 交集靶点的 GO 生物过程富集分析；D-PharmMapper 交集靶点的 KEGG 通路富集分析；E-SEA 预测靶点与氧化应激相关靶点的交集；F-SEA 交集靶点的 Hub Top30 蛋白质相互作用网络；G-SEA 交集靶点的 GO 生物过程富集分析；H-SEA 交集靶点的 KEGG 通路富集分析。

A-overlap between PharmMapper-predicted targets and oxidative stress-related targets; B-protein-protein interaction (PPI) network of the Top 30 hub targets in the PharmMapper overlap; C-gene ontology biological process enrichment of the PharmMapper overlapping targets; D-KEGG pathway enrichment of the PharmMapper overlapping targets; E-overlap between SEA-predicted targets and oxidative stress-related targets; F-PPI network of the Top 30 hub targets in the SEA overlap; G-gene ontology biological process enrichment of the SEA overlapping targets; H-KEGG pathway enrichment of the SEA overlapping targets.

图 7　PharmMapper 和 SEA 靶点预测交叉验证分析

Fig. 7　Cross-validation analysis of target predictions using PharmMapper and SEA

对双重稳健靶点进行外部证据审计。该审计不参与靶点筛选，也不作为双重稳健策略优越性的主要证据，而是用于检查候选靶点是否具有既有氧化应激相关文献背景，并评估这种文献背景是否可能受到 PPI 度值偏差影响。

外部证据比较显示，传统固定参数前 30 位、单一最佳参数前 30 位、PR-QI 稳健前 30 位和双重稳健候选靶点（n＝29）均具有一定文献或数据库支持，但这些集合代表的筛选逻辑不同。传统固定参数前 30 位能够捕捉部分文献和数据库中较活跃的核心靶点，但也包含参数敏感或成分敏感节点；单一最佳参数前 30 位的功能富集表现较强，但依赖某一后验最优参数组合；PR-QI 稳健前 30 位强调靶点在参数空间内稳定出现；双重稳健候选靶点则进一步要求候选靶点在成分 leave-one-compound-out 扰动下仍保持稳定。传统固定参数前 30 位与双重稳健候选靶点共有 20 个重叠靶点，占传统固定参数前 30 位的 66.7%，占双重稳健候选靶点的 69.0%。这说明传统固定参数方法能够捕捉部分核心靶点，但双重扰动框架进一步排除了一部分参数敏感或成分敏感候选。双重稳健候选靶点的平均 LOO PR-QI 保留率为 0.977，显示其对输入成分扰动具有较高稳定性。结果显示，在相同靶点规模下，PR-QI 稳健前 30 位在高共现阈值下的文献证据密度高于传统固定参数前 30 位，并接近单一最佳参数前 30 位。

如表 6 和图 8 所示，在 Top 30 比较中，PR-QI 稳健前 30 位在 PubMed 共现阈值（T）≥50、≥100、≥200、≥500 和≥1 000 下均等于或高于传统固定参数前 30 位，并与单一最佳参数前 30 位接近。这说明 PR-QI 稳健前 30 位并未因避免单一参数选择而损失高阈值文献背景。双重稳健候选靶点的宽松阈值覆盖率低于各 Top 30 策略，但在 T≥500 和≥1 000 下仍保留 8/29 和 7/29 个高共现靶点，提示双重稳健集合在压缩候选规模的同时仍保留部分高文献背景节点。整体来看在所有策略比较中，PR-QI 稳健前 30 位在高共现阈值下的文献证据

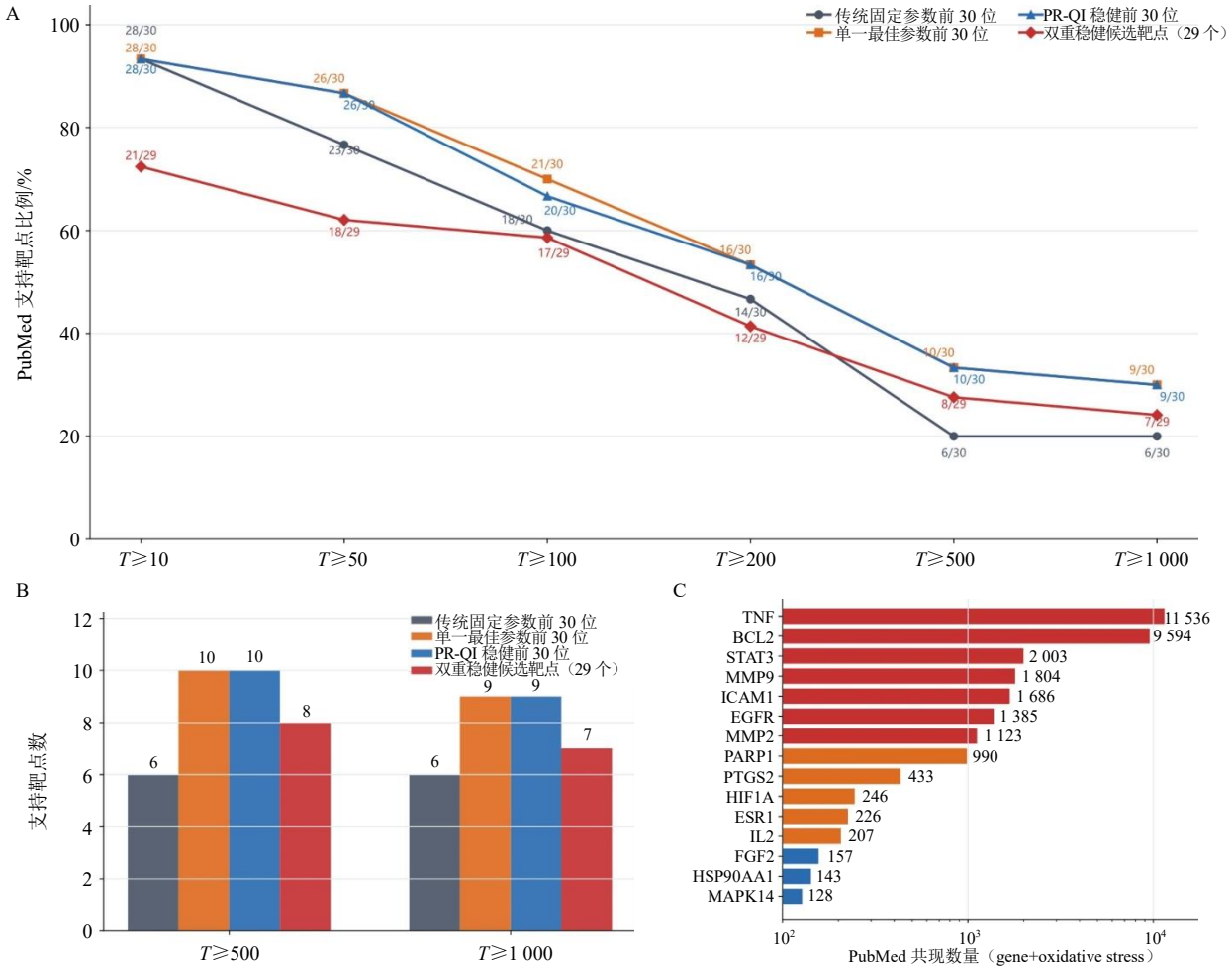
表 6　PubMed 共现证据密度比较

Table 6　Comparison of PubMed co-occurrence density

靶点集	$T \geqslant 10$	$T \geqslant 50$	$T \geqslant 100$	$T \geqslant 200$	$T \geqslant 500$	$T \geqslant 1\,000$
传统固定参数前 30 位	28/30 (93.3%)	23/30 (76.7%)	18/30 (60.0%)	14/30 (46.7%)	6/30 (20.0%)	6/30 (20.0%)
单一最佳参数前 30 位	28/30 (93.3%)	26/30 (86.7%)	21/30 (70.0%)	16/30 (53.3%)	10/30 (33.3%)	9/30 (30.0%)
PR-QI 稳健前 30 位	28/30 (93.3%)	26/30 (86.7%)	20/30 (66.7%)	16/30 (53.3%)	10/30 (33.3%)	9/30 (30.0%)
双重稳健候选靶点（29 个）	21/29 (72.4%)	18/29 (62.1%)	17/29 (58.6%)	12/29 (41.4%)	8/29 (27.6%)	7/29 (24.1%)

T 表示 PubMed 共现阈值，当靶点“基因名＋oxidative stress”相关文献数量≥*T* 时，认为该靶点具有文献支撑；同一比较族的超几何检验显示，各策略在表中相应阈值下均达到显著富集（BH-FDR≤0.006 3），其中双重稳健候选靶点在 *T*≥50 及更严格阈值下达到显著富集。

T denotes the PubMed co-occurrence threshold, when the number of literature related to the target “gene name +oxidative stress” is ≥ *T*, it is considered that the target has literature support; Within each comparison family, hypergeometric tests showed significant enrichment at the corresponding thresholds (BH-adjusted FDR ≤ 0.006 3), the dual-robust candidate targets showed significant enrichment at *T* ≥ 50 and all stricter thresholds.



A-不同靶点筛选策略在不同 PubMed 共现阈值下的文献支持比例变化；B-不同靶点筛选策略在高共现阈值下的文献支持靶点数量比较；C-代表性靶点与氧化应激相关的 PubMed 共现数量排序。

A-changes in proportion of literature support under different PubMed co-occurrence thresholds across different target-screening strategies; B-comparison of the number of targets supported by literature under high co-occurrence thresholds for different target screening strategies; C-Ranking of pubMed co-occurrence counts related to representative targets and oxidative stress.

图 8　不同靶点筛选策略的 PubMed 共现证据比较

Fig. 8　Comparison of PubMed co-occurrence profiles across different target-screening strategies

密度高于传统固定参数前 30 位，并接近单一最佳参数前 30 位。

考虑到网络中心性高的靶点更容易被数据库和文献记录，构建节点度匹配随机对照集。对双重稳健

集合中的每一个真实靶点均从背景靶点池中随机选择 1 个 PPI 度值相近的替代靶点，若某真实靶点的 PPI 度值较高，则只从度值同样较高或接近的候选靶点中抽取替代基因；若某真实靶点度值较低，则从度值较低的候选靶点中抽取替代基因，得到与真实双重稳健集合规模相同、PPI 度值分布相近的随机靶点集，统计其中靶点的 PubMed 共现次数≥100。重复该过程 1 000 次后，得到节点度匹配随机背景下的空分布。结果（表 7）显示，随机集合中 PubMed≥100 的靶点数平均为 8.21 个，95%区间为 5～12 个，低于真实的双重稳健集合。双重稳健靶点的外部文献背景不能完全由高度值偏差解释，但仍应视为补充性信息，而非独立机制验证。结合前述双重稳健靶点结果及网络功能代表性，本研究选取 STAT3、EGFR、SRC、TNF、HSP90AA1 和 BCL2 这 6 个候选靶点作为对接对象；从表 2 所列成分中选取儿茶素、槲皮素-3-*O*-葡萄糖醛酸苷和原花青素 B1 这 3 个成分作为代表性配体，形成 18 组受体-配体组合。该组合属于代表性对接对象的选取，不表示按化合物影响评分直接取前 3 位，也不构成已证实的活性成分结论。原配体重对接的 18 次运行中，12 次排名第 1 构象 RMSD≤2.0 Å（STAT3 0/3、EGFR 1/3、SRC 3/3、TNF 2/3、BCL2 3/3、HSP90AA1 3/3）。STAT3 的 3 次排名第 1 构象 RMSD 分别为 3.542、2.736、3.267 Å，均未通过预设标准；但检查全部输出构象后，2 次运行采样到 RMSD≤2.0 Å 的近天然构象，最小 RMSD 为 1.576 Å，表明其主要限制在于近天然构象未被 Vina 稳定排至首位。该现象可能与 KQV 构象自由度较高以及刚性受体对接对 SH2 位点构象适配、静电和溶剂化效应描述有限有关。

表 7 双重稳健靶点的节点度匹配置换检验结果

Table 7 Degree-matched permutation test results for dual-robust candidate targets

靶点集	指标	观察值	Null 均值	Null95%区间	*Z*-score	Empirical *P*
双重稳健	PubMed_ge_50_count	18	13.75	10～18	2.05	0.036
双重稳健	PubMed_ge_100_count	17	8.21	5～12	4.70	0.001
双重稳健	PubMed_ge_200_count	12	5.75	3～9	3.84	0.001
双重稳健	PubMed_count_sum	32 104	19 009	3 424～27 484	2.07	0.001
双重稳健	PubMed_count_median	128	44.6	17～88	4.95	0.001

基于 1 000 次节点度匹配置换检验；Empirical *P* 为单侧检验 *P* 值；Null 95%区间为随机集合指标的 2.5%～97.5%分位数。

Results are based on 1 000 node degree-matched permutations; Empirical *P* is the one-sided empirical *P* value, and the 95% Null interval corresponds to the 2.5%—97.5% percentiles of the random-set distribution.

分子对接仅为网络鲁棒性分析提供结构层面的补充信息。STAT3 的 Tier 1 分级仅依据其参数出现频率和 LOO 保留率，不依赖分子对接结果；鉴于其排名第 1 构象重对接 0/3 通过，STAT3 与候选成分的对接结果仅作为假设性结合模式参考。SRC、EGFR、TNF 和 HSP90AA1 的部分组合在相应共晶口袋中获得较低且随机种子间稳定的 Vina 预测评分，但不同受体和配体间的绝对评分不宜直接解释为实验亲和力差异。代表性对接构象及相互作用示意图见图 9。图 10-A、B 分别汇总 6 个候选靶点与 3 个代表性成分的平均 Vina 评分及 3 个随机种子的重复结果，图 10-C、D 分别展示各靶点中评分最低的成分及各成分跨靶点的评分分布。例如，SRC 与槲皮素-3-*O*-葡萄糖醛酸苷、EGFR 与原花青素 B1 组合的 Vina 评分分别为（−10.497 ± 0.023）和（−9.637±0.016）kcal/mol，仅用于描述相应口袋中的预测结合模式和计算重复性。

分子对接结果与网络鲁棒性分析形成互补信息：STAT3 在系统层面最稳健；SRC、EGFR、TNF 和 HSP90AA1 的部分组合在相应口袋中预测存在结合可能性，但仍需实验验证。

3 讨论

112 组参数扰动显示，SwissTargetPrediction 概率、STRING 综合得分和 Hub Top K 的变化均会影响 Top K 靶点及 GO-BP/KEGG 功能富集结果。依据跨参数出现频率及平均归一化度值或富集强度，最终识别出 50 个参数稳健靶点和 70 个参数稳健功能条目，表明参数扰动的主要价值在于区分偶然入选结果与跨阈值可重复的候选信号。

成分 LOO 扰动进一步表明，部分 PR-QI 靶点依赖特定输入成分，而 29 个靶点在参数与成分两类扰动下均保持稳健。该集合较单一全成分条件下

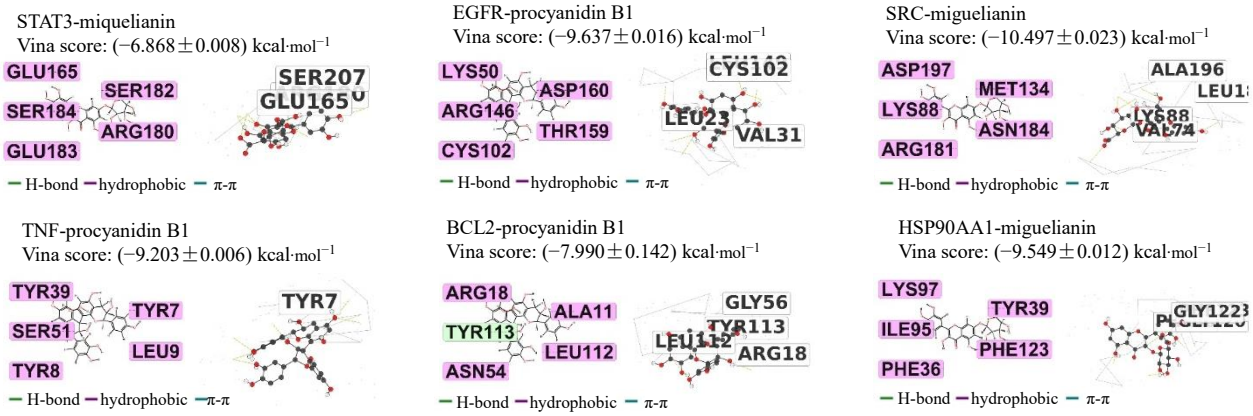
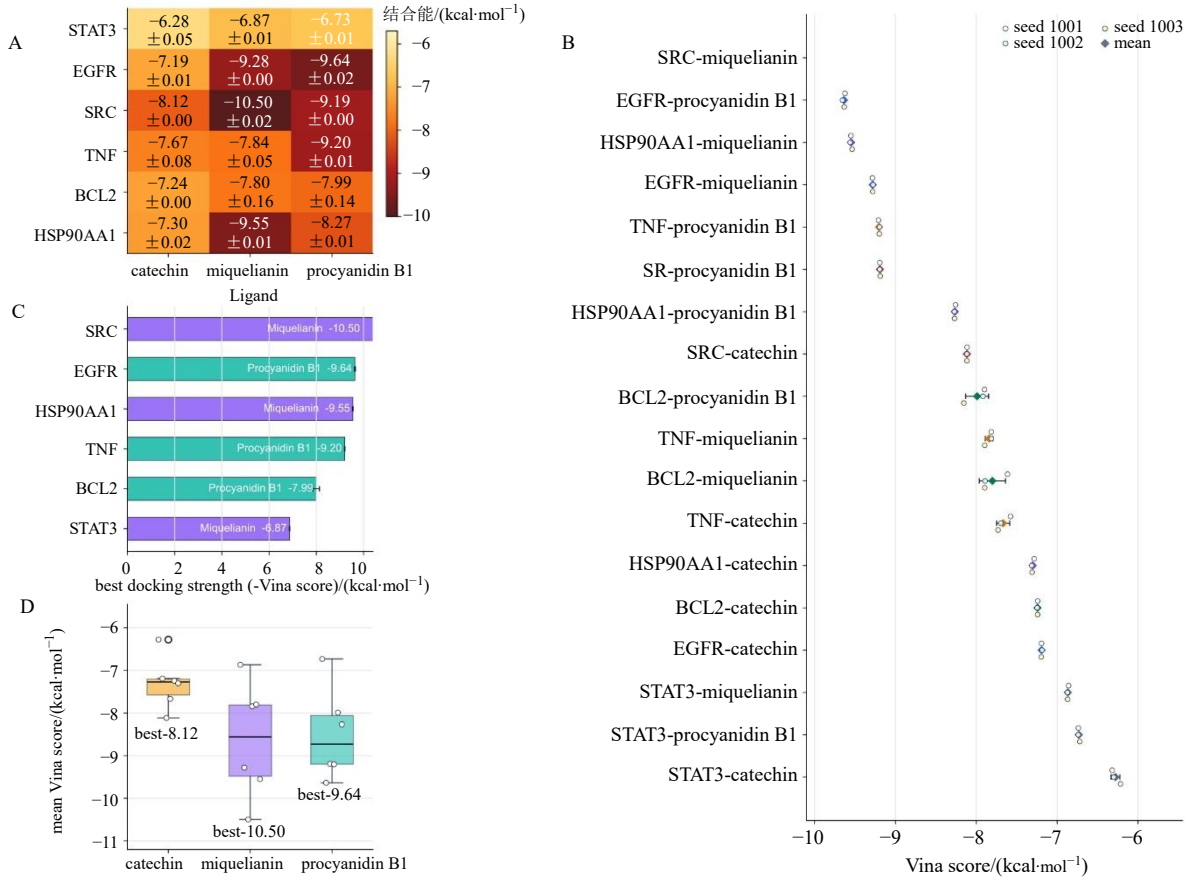


图 9　代表性靶点与候选成分的分子对接结果

Fig. 9　Molecular docking results for representative targets and candidate compounds



A-6 个候选靶点与 3 个候选成分的平均 Vina 评分热图；B-18 组 receptor-ligand 组合的随机种子重复结果及 $\bar{x} \pm s$；C-各候选靶点中平均 Vina 评分最低的候选成分；D-3 个候选成分在 6 个候选靶点中的 Vina 评分分布。

A-heatmap of mean Vina scores for six candidate targets and three candidate compounds; B-random-seed reproducible results and their $\bar{x} \pm s$ for 18 receptor–ligand pairs; C-candidate compound yielding the lowest mean Vina score for each target; D-distributions of Vina scores for the three candidate compounds across six candidate targets.

图 10　候选靶点与候选成分分子对接及随机种子稳定性分析

Fig. 10　Molecular docking and random-seed stability analysis for candidate targets and candidate compounds

的候选结果具有更低的输入成分依赖性，可作为后续机制研究的优先候选，但不代表其已获得实验验证。同时采用的中位数分割、LOO 保留率 0.80 及最低 LOO 参数频率＞0 均属于预先设定的操作性

判定标准，并非具有唯一性的统计学界值。敏感性分析显示，随着 PR-QI 分割阈值、LOO 保留率和最低 LOO 参数频率趋于严格，双重稳健靶点及功能条目的数量总体减少，其中功能条目对阈值变化更为敏感，而靶点集合保持了较高的一致性。尤其是 STAT3、EGFR、SRC、TNF、HSP90AA1、BCL2、PTGS2 和 MMP9 在不同判定条件下均被保留，说明主要候选靶点并非由某一特定分割方式或 LOO 阈值单独驱动。因此，本研究保留 Q_{50}、LOO 保留率 0.80 和最低 LOO 参数频率＞0 作为主分析条件，主要是为了在候选集合覆盖度与筛选严格度之间取得平衡，而非将其视为普遍适用的最优阈值。

功能条目对判定阈值变化更敏感，而主要靶点集合保持较高一致性，可能与二者所处的分析层级不同有关。靶点集合是预测靶点筛选和 PPI 拓扑排序的直接结果，STAT3、EGFR、SRC 等核心节点具有较高的跨参数出现频率，并受到多个输入成分及网络连接关系的共同支持，因此小幅调整判定阈值通常不会改变其优先级。功能富集则是靶点集合的下游映射，少量靶点的进入或退出即可同时改变多个相互重叠基因集的命中数、富集强度、校正 *P* 值和排序；每个数据库内保留前 20 个条目的规则及后续相对分位数分割还可能进一步放大接近边界条目的变化。因此，功能条目集合的波动并不必然表示核心靶点不稳定，而是反映从靶点层向功能解释层传播时存在非线性放大效应。该结果提示，网络药理学研究应优先报告跨参数和成分扰动稳定保留的核心靶点及反复出现的功能模块，避免仅依据单个富集条目或疾病命名通路提出确定性机制结论，并应同时公开靶点层和功能条目层的敏感性结果。

PharmMapper 和 SEA 交叉预测为部分双重稳健候选靶点提供了不同计算原理下的补充证据。SwissTargetPrediction 主要依据已知活性配体的二维和三维结构相似性，PharmMapper 采用反向药效团匹配，SEA 则依据配体集合相似性[21,30-31]，因此不同平台预测结果仅表现出部分重叠属于预期现象。上述原理及数据库覆盖范围的系统性差异可能导致一定的靶点类别偏向：SwissTargetPrediction 相对有利于识别具有结构相近已知活性配体的靶点，PharmMapper 更依赖具有可用蛋白结构和明确药效团特征的靶点，SEA 则倾向于识别已积累一定规模且化学特征相对一致的已知配体集合。同时 PharmMapper Hub Top30 主要涉及受体酪氨酸激酶及蛋白磷酸化信号，而 SEA Hub Top30 更多涉及 G 蛋白偶联受体和 cAMP 信号，提示不同预测原理和数据库覆盖可能造成平台特异性的靶点类别分布；但基于本研究的单一应用案例，尚不能将该差异完全归因于算法本身。

PharmMapper 直接支持 13 个双重稳健候选靶点，其中 EGFR、SRC、HSP90AA1、KDR、MAPK14、MET 和 PARP1 等同时具有较高 PPI 网络优先级；SEA 直接支持 CA2、CA9、CA12、MMP2 和 MMP9，其中 CA2 和 CA12 同时获得 3 个平台支持。上述结果表明，部分主分析候选靶点能够被不同计算方法重复识别，提高了其作为后续实验验证候选的优先级。需要指出的是，跨平台未重叠不能被解释为否定证据。PharmMapper 预测受到药效团及蛋白结构模型覆盖度的影响，SEA 则依赖已知配体集合及其化学空间覆盖情况，且本研究 SEA 结果表现出不同化合物靶点贡献不均衡的特点。因此，STAT3 等未获得 PharmMapper 或 SEA 直接预测支持的靶点不能据此被排除，其优先级主要来自参数扰动、成分扰动、PPI 网络稳定性及其他补充证据。3 个平台的预测均属于计算证据，不能替代细胞、体外或分子生物学实验验证。

从生物学角度看，STAT3、EGFR、SRC、TNF、HSP90AA1、BCL2、PTGS2 和 MMP9 等靶点共同指向炎症信号、激酶调控、细胞凋亡和氧化应激反应。STAT3 可参与炎症因子介导的转录调控，并与氧化应激状态下的细胞存活和免疫调节相关[35-36]。EGFR 和 SRC 是酪氨酸激酶信号网络中的关键节点[37-38]，蛋白激酶可受氧化还原状态调控[39]，并可通过核因子-κB（nuclear factor-κB，NF-κB）相关炎症信号影响细胞增殖、炎症和氧化还原稳态[40]。MAPK 通路也参与氧化应激响应及其下游细胞效应[41]。TNF、PTGS2 和 MMP9 则与炎症反应、前列腺素代谢、细胞外基质重塑和氧化应激损伤密切相关。这些靶点在参数扰动和成分扰动中保持稳定，提示草莓叶抗氧化作用可能是多层次信号网络调控，而非单一通路效应。功能富集层面也支持这一解释。癌症通路、乙型肝炎、晚期糖基化终末产物-受体信号通路和脂质与动脉粥样硬化等疾病命名通路共享氧化应激响应、炎症调控、凋亡调节及激酶信号转导等分子过程，其中 AGE-RAGE 通路与活性氧（reactive oxygen species，ROS）生成和炎症调控密切相关[42-43]。

因此，这些条目的富集仅提示候选靶点可能参与相关功能模块，不应解释为草莓叶对相应疾病具有治疗作用。成分扰动分析显示，强心苷类候选化合物和儿茶素的化合物影响评分较高，说明网络结果对其缺失较敏感。该差异可能与预测靶点集合及网络覆盖范围有关；鉴于前者仅为 Level 3 暂定候选结构，其较高影响评分应解释为模型对该候选注释。

在候选机制模型中，STAT3 更适合作为稳定保留的下游转录枢纽，EGFR 和 SRC 则可能代表上游激酶节点。STAT3 是唯一 Tier 1 候选靶点，在所有参数组合和成分剔除条件下均保持 PR-QI 状态；EGFR 和 SRC 的参数频率略低，但网络连接度较高。分子对接仅提示 SRC-槲皮素-3-*O*-葡萄糖醛酸苷、EGFR-原花青素 B1 等组合在相应口袋中预测存在结合可能性，Vina 评分不能证明直接作用或实验亲和力。既往研究显示，槲皮素等多酚可通过抑制 JAK2/STAT3 磷酸化、STAT3 二聚化或核转位影响该信号轴并降低 ROS[44-48]，而氧化应激本身也可激活 JAK2 等上游信号[49]。因此，STAT3 的机制优先级主要来自网络稳定性和生物学证据，而非其对接评分。EGFR/SRC-STAT3 轴可作为优先验证的候选机制链条，但其调控方向和因果关系仍需实验确认。

本研究仍存在以下局限：候选靶点主要来源于数据库预测和网络拓扑，SwissTargetPrediction、STRING 及疾病基因数据库的覆盖范围和收录偏倚可能影响结果，高度值靶点也更易获得文献支持；跨平台预测、文献共现、置换检验和分子对接均不能替代实验验证。后续应在细胞或体外体系中检测代表性候选成分对 ROS 水平及 EGFR、SRC 和 STAT3 磷酸化状态的影响，并结合靶点抑制或敲降实验明确 EGFR/SRC-STAT3 轴的调控方向和因果关系。

4 结论

本研究构建了参数与成分双扰动鲁棒性框架，将网络药理学候选筛选转化为跨阈值、跨输入成分的稳定性排序。通过 112 组参数组合和 2016 组成分扰动运行，从 684 个预测靶点中识别出 50 个参数稳健靶点和 29 个双重稳健靶点，并获得 70 个参数稳健功能条目和 58 个双重稳健功能条目；STAT3 为唯一 Tier 1 候选靶点，EGFR/SRC-STAT3 轴可作为优先验证的候选机制链条。该框架可提高网络药理学候选结果的稳定性与可审计性，为天然产物及其他多成分体系的候选靶点优先级识别和后续实验验证提供方法学参考；但上述结果仍属于计算预测，其调控方向和因果关系尚需实验确认。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 李泮霖, 苏薇薇. 网络药理学在中药研究中的最新应用进展 [J]. 中草药, 2016, 47(16): 2938-2942.

[2] 解静, 高杉, 李琳, 等. 网络药理学在中药领域中的研究进展与应用策略 [J]. 中草药, 2019, 50(10): 2257-2265.

[3] 牛明, 张斯琴, 张博, 等. 《网络药理学评价方法指南》解读 [J]. 中草药, 2021, 52(14): 4119-4129.

[4] 赵韦欣, 王晴, 王梦齐, 等. 基于网络药理学和分子对接探讨鸦胆子治疗结直肠癌的作用机制 [J]. 中草药, 2023, 54(6): 1850-1859.

[5] Guney E, Menche J, Vidal M, *et al*. Network-based in silico drug efficacy screening [J]. *Nat Commun*, 2016, 7: 10331.

[6] Cheng F X, Desai R J, Handy D E, *et al*. Network-based approach to prediction and population-based validation of in silico drug repurposing [J]. *Nat Commun*, 2018, 9(1): 2691.

[7] Morselli Gysi D, do Valle Í, Zitnik M, *et al*. Network medicine framework for identifying drug-repurposing opportunities for COVID-19 [J]. *Proc Natl Acad Sci U S A*, 2021, 118(19): e2025581118.

[8] Himmelstein D S, Lizee A, Hessler C, *et al*. Systematic integration of biomedical knowledge prioritizes drugs for repurposing [J]. *eLife*, 2017, 6: e26726.

[9] Thombre K R, Gabhane V, Gupta K R, *et al*. Network pharmacology in the multi-omics era: Uncovering novel therapeutic strategies [J]. *Discov Chem*, 2026, 3(1): 111.

[10] Wang Y Y, Liu M X, Jafari M, *et al*. A critical assessment of traditional Chinese medicine databases as a source for drug discovery [J]. *Front Pharmacol*, 2024, 15: 1303693.

[11] Wang R, Fan P B, Deng X, *et al*. Reproducible computational workflow for drug discovery to standardize network pharmacology and molecular docking analyses [J]. *J Vis Exp*, 2026(230): e70171.

[12] Najjar A, Hamadeh A, Krause S, *et al*. Global sensitivity analysis of Open Systems Pharmacology Suite physiologically based pharmacokinetic models [J]. *CPT Pharmacometrics Syst Pharmacol*, 2024, 13(12): 2052-2067.

[13] Fecka I, Bednarska K, Włodarczyk M. *Fragaria* × *Ananassa* cv. *Senga* Sengana leaf: An agricultural waste with antiglycation potential and high content of

ellagitannins, flavonols, and 2-pyrone-4, 6-dicarboxylic acid [J]. *Molecules*, 2022, 27(16): 5293.

[14] Salas-Arias K, Irías-Mata A, Sánchez-Calvo L, *et al*. Eliciting polyphenols in strawberry leaves: Preliminary experiments in *Fragaria* × *Ananassa* cv. *festival* [J]. *Molecules*, 2024, 29(11): 2467.

[15] Cvetković D J, Stanojević L P, Stanković M Z, *et al*. Antioxidant activity of strawberry (*Fragaria* × *ananassa* Duch.) leaves [J]. *Sep Sci Technol*, 2017, 52(6): 1039-1051.

[16] Ladika G, Christodoulou P, Kritsi E, *et al*. Exploring postharvest metabolic shifts and NOX2 inhibitory potential in strawberry fruits and leaves via untargeted LC-MS/MS and chemometric analysis [J]. *Metabolites*, 2025, 15(5): 321.

[17] 周友良, 路婉杉, 成志伟, 等. LC-MS/MS 联合网络药理学、分子对接探究草莓叶抗氧化作用靶点及机制 [J]. 日用化学工业 (中英文), 2025, 55(1): 63-74.

[18] Sies H. Oxidative stress: A concept in redox biology and medicine [J]. *Redox Biol*, 2015, 4: 180-183.

[19] Wani R, Nagata A, Murray B W. Protein redox chemistry: Post-translational cysteine modifications that regulate signal transduction and drug pharmacology [J]. *Front Pharmacol*, 2014, 5: 224.

[20] Kim S, Chen J, Cheng T J, *et al*. PubChem 2023 update [J]. *Nucleic Acids Res*, 2023, 51(D1): D1373-D1380.

[21] Daina A, Michielin O, Zoete V. SwissTargetPrediction: Updated data and new features for efficient prediction of protein targets of small molecules [J]. *Nucleic Acids Res*, 2019, 47(W1): W357-W364.

[22] Safran M, Dalah I, Alexander J, *et al*. GeneCards Version 3: The human gene integrator [J]. *Database*, 2010, 2010: baq020.

[23] Amberger J S, Bocchini C A, Schiettecatte F, *et al*. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders [J]. *Nucleic Acids Res*, 2015, 43(D1): D789-D798.

[24] Zhou Y, Zhang Y T, Lian X C, *et al*. Therapeutic target database update 2022: Facilitating drug discovery with enriched comparative data of targeted agents [J]. *Nucleic Acids Res*, 2022, 50(D1): D1398-D1407.

[25] 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.

[26] Consortium T G O, Aleksander S A, Balhoff J, *et al*. The gene ontology knowledgebase in 2023 [J]. *Genetics*, 2023, 224: iyad031.

[27] Kanehisa M, Furumichi M, Sato Y, *et al*. KEGG for taxonomy-based analysis of pathways and genomes [J]. *Nucleic Acids Res*, 2023, 51(D1): D587-D592.

[28] Kuleshov M V, Jones M R, Rouillard A D, *et al*. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update [J]. *Nucleic Acids Res*, 2016, 44(W1): W90-W97.

[29] Chen E Y, Tan C M, Kou Y, *et al*. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool [J]. *BMC Bioinform*, 2013, 14(1): 128.

[30] Wang X, Shen Y H, Wang S W, *et al*. PharmMapper 2017 update: A web server for potential drug target identification with a comprehensive target pharmacophore database [J]. *Nucleic Acids Res*, 2017, 45(W1): W356-W360.

[31] Keiser M J, Roth B L, Armbruster B N, *et al*. Relating protein pharmacology by ligand chemistry [J]. *Nat Biotechnol*, 2007, 25(2): 197-206.

[32] Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing [J]. *J R Stat Soc Ser B Stat Methodol*, 1995, 57(1): 289-300.

[33] Seeliger D, de Groot B L. Ligand docking and binding site analysis with PyMOL and Autodock/Vina [J]. *J Comput Aided Mol Des*, 2010, 24(5): 417-422.

[34] Eberhardt J, Santos-Martins D, Tillack A F, *et al*. AutoDock vina 1.2.0: New docking methods, expanded force field, and Python bindings [J]. *J Chem Inf Model*, 2021, 61(8): 3891-3898.

[35] 杜雨芯, 操娇娇, 张倩霞, 等. 银杏二萜内酯通过拮抗 PAFR 和调节 SIRT1/STAT3 抑制氧化应激诱导的 PC12 神经元衰老研究 [J]. 中草药, 2023, 54(9): 2793-2801.

[36] Kasembeli M M, Bharadwaj U, Robinson P, *et al*. Contribution of STAT3 to inflammatory and fibrotic diseases and prospects for its targeting for treatment [J]. *Int J Mol Sci*, 2018, 19(8): 2299.

[37] Heppner D E, van der Vliet A. Redox-dependent regulation of epidermal growth factor receptor signaling [J]. *Redox Biol*, 2016, 8: 24-27.

[38] Heppner D E. Structural insights into redox-active cysteine residues of the Src family kinases [J]. *Redox Biol*, 2021, 41: 101934.

[39] Truong T H, Carroll K S. Redox regulation of protein kinases [J]. *Crit Rev Biochem Mol Biol*, 2013, 48(4): 332-356.

[40] Liu T, Zhang L Y, Joo D, *et al*. NF-κB signaling in inflammation [J]. *Sig Transduct Target Ther*, 2017, 2: 17023.

[41] Runchel C, Matsuzawa A, Ichijo H. Mitogen-activated protein kinases in mammalian oxidative stress responses [J]. *Antioxid Redox Signal*, 2011, 15(1): 205-218.

[42] Yan S F, Ramasamy R, Schmidt A M. The receptor for advanced glycation endproducts (RAGE) and cardiovascular disease [J]. *Expert Rev Mol Med*, 2009, 11: e9.

[43] Sparvero L J, Asafu-Adjei D, Kang R, *et al*. RAGE (receptor for advanced glycation endproducts), RAGE ligands, and their role in cancer and inflammation [J]. *J Transl Med*, 2009, 7: 17.

[44] Shin E J, Lee J S, Hong S, *et al*. Quercetin directly targets JAK2 and PKCδ and prevents UV-induced photoaging in human skin [J]. *Int J Mol Sci*, 2019, 20(21): 5262.

[45] Wang D, Ali F, Liu H X, *et al*. Quercetin inhibits angiotensin II-induced vascular smooth muscle cell proliferation and activation of JAK2/STAT3 pathway: A target based networking pharmacology approach [J]. *Front Pharmacol*, 2022, 13: 1002363.

[46] Cao H H, Tse A K, Kwan H Y, *et al*. Quercetin exerts anti-melanoma activities and inhibits STAT3 signaling [J]. *Biochem Pharmacol*, 2014, 87(3): 424-434.

[47] Granato M, Gilardini Montani M S, Zompetta C, *et al*. Quercetin interrupts the positive feedback loop between STAT3 and IL-6, promotes autophagy, and reduces ROS, preventing EBV-driven B cell immortalization [J]. *Biomolecules*, 2019, 9(9): 482.

[48] Yin Q Q, Wang L Y, Yu H Y, *et al*. Pharmacological effects of polyphenol phytochemicals on the JAK-STAT signaling pathway [J]. *Front Pharmacol*, 2021, 12: 716672.

[49] Mazière C, Conte M A, Mazière J C. Activation of JAK2 by the oxidative stress generated with oxidized low-density lipoprotein [J]. *Free Radic Biol Med*, 2001, 31(11): 1334-1340.

[责任编辑　潘明佳]