

## 网络毒理学结合机器学习与多维生物信息学解析全氟辛酸暴露致结直肠癌的分子毒理机制

石智尧<sup>1,2</sup>, 高宇<sup>2\*</sup>, 闫若菲<sup>2</sup>, 刘丽坤<sup>2</sup>, 王晞星<sup>1,2\*</sup>

1. 广州中医药大学 第一临床医学院, 广东 广州 510405

2. 山西省中医药研究院 肿瘤科, 山西 太原 030012

**摘要:** 目的 系统阐明全氟辛酸(PFOA)暴露致结直肠癌(CRC)的分子毒理机制, 筛选核心靶点并验证其表达特征、免疫相关性及临床价值, 构建PFOA诱导CRC的不良结局通路(AOP)假说框架。方法 整合网络毒理学、多机器学习算法(LASSO回归、SVM-RFE、随机森林)及多维生物信息学分析。从公共数据库收集PFOA与CRC相关靶点, 筛选交叉靶点并进行功能富集分析; 通过GEO和TCGA(UALCAN平台)数据集验证核心靶点的差异表达、免疫浸润及预后意义; 采用分子对接及100 ns分子动力学模拟评估PFOA与核心靶点的结合稳定性; 基于关键事件关联构建AOP框架。结果 共鉴定738个PFOA-CRC交叉靶点, 显著富集于凋亡通路、脂质与动脉粥样硬化通路及CRC通路。机器学习筛选出6个核心靶点[ATP结合盒亚家族G成员2(ABCG2)、紧密连接蛋白1(CLDN1)、ETS变异转录因子4(ETV4)、溶质载体家族7成员5(SLC7A5)、MMP7、CA9], 这些基因在不同队列CRC样本中表达持续失调( $P < 0.05$ ), 且与免疫细胞浸润密切相关: ABCG2增强抗肿瘤免疫应答, CLDN1、ETV4、SLC7A5、MMP7和CA9协同促进免疫抑制微环境形成。临床验证显示核心靶点表达与CRC患者预后仅呈关联趋势, 未达到统计学显著性( $P > 0.05$ )。分子对接证实PFOA与所有核心靶点稳定结合, 其中与CLDN1结合力最强( $-29.34 \text{ kJ} \cdot \text{mol}^{-1}$ ), 分子动力学模拟进一步验证了复合物的稳定性。结论 PFOA通过调控多核心靶点、重塑肿瘤免疫微环境促进CRC进展。本研究构建的整合计算框架及AOP假说, 为环境污染物健康风险评估及CRC靶向干预提供了理论基础和实验依据。

**关键词:** 全氟辛酸; 结直肠癌; 网络毒理学; 机器学习; 分子对接; 分子动力学模拟; AOP假说

中图分类号: R285.5 文献标志码: A 文章编号: 1674-6376(2026)10-3650-15

DOI: 10.7501/j.issn.1674-6376.2026.10.021

## Molecular toxicological mechanism of perfluorooctanoic acid exposure-induced colorectal cancer based on network toxicology combined with machine learning and multidimensional bioinformatics

SHI Zhiyao<sup>1,2</sup>, GAO Yu<sup>2</sup>, YAN Ruofei<sup>2</sup>, LIU Likun<sup>2</sup>, WANG Xixing<sup>1,2</sup>

1. The First Clinical Medical College, Guangzhou University of Chinese Medicine, Guangzhou 510405, China

2. Department of Oncology, Shanxi Academy of Traditional Chinese Medicine, Taiyuan 030012, China

**Abstract: Objective** To systematically clarify the molecular toxicological mechanisms of perfluorooctanoic acid (PFOA) exposure-induced colorectal cancer (CRC), identify core targets, verify their expression characteristics, immune relevance, and clinical value, and construct a hypothetical AOP framework for PFOA-induced CRC. **Methods** Network toxicology, multiple machine learning algorithms (LASSO regression, SVM-RFE, and Random Forest), and multidimensional bioinformatics analyses were integrated. PFOA-related and CRC-related targets were collected from public databases, and overlapping targets were screened and subjected to functional enrichment analysis. The differential expression, immune infiltration, and prognostic significance of core targets were

收稿日期: 2026-03-25

**基金项目:** 第四届国医大师传承工作室(中国中医药办人教函2022245号1123-04); 国家中医药传承创新中心的建设(202203); 山西省中医院管理局项目(zyytd2024001); 山西省中医院管理局项目(zyyys2024001); 山西省卫健委山西省医学科技创新团队项目(2020TD04); 山西省卫健委山西省医学重大科技攻关专项(2022XM10)

**作者简介:** 石智尧, 男, 博士研究生, 主要从事中西医结合防治肿瘤疾病的研究。E-mail: 18234069215y@163.com

**\*通信作者:** 高宇, 女, 副主任医师, 硕士生导师, 主要从事中西医结合防治肿瘤的临床疾病研究。E-mail: ggyy552716@163.com

王晞星, 男, 教授, 国医大师, 主任医师, 博士生导师, 主要从事中西医结合防治肿瘤的临床疾病研究。E-mail: wangxx315@163.com

validated using GEO and TCGA datasets (via the UALCAN platform). Molecular docking and 100-ns molecular dynamics (MD) simulations were performed to evaluate the binding stability between PFOA and core targets. An AOP framework was constructed based on the association of key events. **Results** A total of 738 overlapping PFOA-CRC targets were identified, which were significantly enriched in pathways such as “apoptosis”, “lipid and atherosclerosis”, and “colorectal cancer”. Six core targets (ABCG2, CLDN1, ETV4, SLC7A5, MMP7, and CA9) were screened by machine learning. These genes showed consistent dysregulation in CRC samples from different cohorts ( $P < 0.05$ ) and were closely associated with immune cell infiltration: ABCG2 enhanced anti-tumor immune responses, while CLDN1, ETV4, SLC7A5, MMP7, and CA9 synergistically promoted the formation of an immunosuppressive microenvironment. Clinical trials demonstrated that core target expression showed only an association trend with prognosis in CRC patients, failing to reach statistical significance ( $P > 0.05$ ). Molecular docking confirmed stable binding between PFOA and all core targets, with the strongest binding affinity observed for CLDN1 ( $-29.34 \text{ kJ} \cdot \text{mol}^{-1}$ ), which was further validated by molecular dynamics simulations confirming the stability of the complex. **Conclusion** PFOA promotes CRC progression by regulating multiple core targets and remodeling the tumor immune microenvironment. The integrated computational framework and AOP hypothesis constructed in this study provide a theoretical basis and experimental evidence for the health risk assessment of environmental pollutants and targeted interventions for CRC.

**Key words:** perfluorooctanoic acid; colorectal cancer; network toxicology; machine learning; molecular docking; molecular dynamics simulation; AOP hypothesis

结直肠癌 (CRC) 是全球第 3 大常见恶性肿瘤, 也是癌症相关死亡的第 2 大原因, 给全球公共卫生带来沉重负担<sup>[1]</sup>。其发病机制涉及遗传易感性、生活方式及环境暴露等多种因素的复杂相互作用<sup>[2]</sup>, 其中持久性有机污染物 (POPs) 已被证实是推动 CRC 发生发展的关键因素<sup>[3]</sup>。在各类环境污染物中, 全氟和多氟烷基物质 (PFAS) 凭借其持久性强、生物蓄积潜力高及人类暴露范围广等特点, 已成为全球性重要公共卫生问题<sup>[4]</sup>。

全氟辛酸 (PFOA) 作为典型的 PFAS, 因具备耐热、化学性质稳定及拒油等优良特性, 被广泛应用于防水涂料、纺织品处理等工业生产领域及各类消费品中<sup>[5]</sup>。人类主要通过受污染的饮用水和食物接触 PFOA, 生物监测数据显示, 超过 98% 的美国人口血清中可检测到 PFOA<sup>[6]</sup>。国际癌症研究机构已将 PFOA 归类为 1 类致癌物<sup>[7]</sup>, 相关研究也证实, PFOA 暴露与肾癌、睾丸癌等多种癌症的发病风险增加存在关联<sup>[8]</sup>。PFOA 具有多种潜在致癌生物学活性, 它可作为过氧化物酶体增殖物激活受体  $\alpha$  激动剂, 扰乱机体脂质代谢与能量稳态<sup>[9]</sup>; 同时还能诱导氧化应激、促进慢性炎症并调控免疫反应, 这些均是肿瘤促进微环境的典型特征<sup>[10-11]</sup>。流行病学调查发现, 化学厂周边居民的溃疡性结肠炎发病率与 PFOA 暴露水平呈显著正相关 ( $P < 0.001$ )<sup>[12]</sup>, 这提示 PFOA 暴露可能通过改变肠道菌群组成、损害肠道屏障功能, 为 CRC 的发生发展创造有利的局部环境<sup>[13]</sup>。然而, 目前关于 PFOA 致 CRC 的研究多

聚焦于单一通路或单个靶点, 缺乏对其致癌过程中全局分子调控网络及核心毒性靶点的系统性探索, 且针对 PFOA 暴露 CRC 进展轴的标准化不良结局通路 (AOP) 尚未建立, 这极大阻碍了对 PFOA 的全面环境风险评估及靶向干预策略的开发。

为填补上述研究空白, 本研究整合网络毒理学、机器学习与多维生物信息学构建计算分析框架, 从多维度开展系统性解析, 有效解决单一研究手段难以全面阐释污染物复杂致病机制的问题。具体而言, 网络毒理学通过构建化合物-靶点-疾病相互作用网络, 为解析环境污染物的复杂作用机制提供系统层面的研究支撑, 机器学习可从高维数据中稳健识别高置信度核心靶点, 提升关键因子筛选的准确性, 多维生物信息学分析则进一步阐明这些靶点的致病作用与临床相关性, 使分析结果更具实际意义。同时借助分子对接与分子动力学模拟探究 PFOA 与潜在蛋白靶点的物理相互作用, 综合各项研究结果提出新颖的 AOP 假说, 为 PFOA 暴露与 CRC 结局的关联提供标准化的机制解释框架。本研究依托 PFOA 毒理学靶点与 CRC 致病背景的综合分析, 全面揭示 PFOA 在结肠中的致癌潜力, 为环境健康风险评估提供可靠参考, 也为污染物相关 CRC 的预防及靶向干预策略制定提供科学依据。

## 1 材料与方法

### 1.1 PFOA 相关靶点的获取

从 PubChem 数据库 (<https://pubchem.ncbi.nlm>.)

nih.gov/) 中检索 PFOA (CID: 9554) 的化学结构及 SMILES 结构, 用于后续靶点预测和分子对接<sup>[14]</sup>。PFOA 相关靶点从生物活性分子数据库 (ChEMBL; <https://www.ebi.ac.uk/chembl/>)<sup>[15]</sup>、瑞士靶点预测数据库 (Swiss Target Prediction; <http://swisstargetprediction.ch/>)<sup>[16]</sup>、化合物-蛋白质相互作用网络数据库 (STITCH; <http://stitch.embl.de/>)<sup>[17]</sup> 及比较毒理基因组数据库 (CTD; <http://ctdbase.org/>)<sup>[18]</sup> 中检索, 且所有数据库检索过程均限定物种为 “Homo sapiens”, 确保靶点的人类物种相关性。之后利用 UniProt 数据库 (<https://www.uniprot.org/>)<sup>[19]</sup> 对筛选后靶点的基因名称进行标准化处理, 形成最终的 PFOA 相关靶点库。

## 1.2 CRC 相关靶点的获取

CRC 相关靶点从在线人类孟德尔遗传数据库 (OMIM, <https://omim.org/>)<sup>[20]</sup>、基因卡片数据库 (GeneCards, <https://www.genecards.org/>)<sup>[21]</sup>、治疗靶标数据库 (TTD, <https://db.idrblab.org/ttd/>)<sup>[22]</sup> 及 DrugBank 数据库 (<https://go.drugbank.com/>)<sup>[23]</sup> 中获取。为确保数据的全面性与相关性, 在 GeneCards 数据库中设置筛选条件为 “Relevance Score  $\geq 10$ ”。随后, 合并来自 4 个数据库的基因列表, 通过去除重复项并进行标准化, 最终构建 CRC 靶点库。

## 1.3 获取交集靶点与功能富集分析

通过 Venny 2.1 软件 (<https://bioinfogp.cnb.csic.es/tools/venny/>) 识别 PFOA 与 CRC 的交集靶点, 并通过 Venn 图进行可视化展示。随后, 通过 DAVID 在线分析平台<sup>[24]</sup>, 对交集靶点进行基因本体 (GO) 和京都基因与基因组百科全书 (KEGG) 富集分析。富集结果的统计学显著性以校正后  $P < 0.05$  作为判断标准。

## 1.4 机器学习筛选核心靶点

为筛选 PFOA 与 CRC 的核心交叉靶点, 本研究首先从 GEO 数据库 (<https://www.ncbi.nlm.nih.gov/geo/>)<sup>[25]</sup> 获取 3 个 CRC 转录组数据集 (GSE39582、GSE44076、GSE74602), 经去批次效应和标准化处理后 (见 “1.5” 项), 提取 “1.3” 项中获得的 738 个 PFOA-CRC 交叉靶点在这些数据集中的表达值, 构建 “738 个基因  $\times$  所有样本” 的表达矩阵, 作为机器学习的输入数据。随后采用 3 种机器学习算法对上述表达矩阵进行分析: 最小绝对收缩与选择算子 (LASSO) 回归<sup>[26]</sup>、支持向量机-递归特征消除 (SVM-RFE)<sup>[27]</sup> 及随机森林

(RandomForest)<sup>[28]</sup>。具体而言, 通过 R 语言 “glmnet” 包实现 LASSO 回归 (参数: 标准化=TRUE,  $\alpha=1$ , 分布=“binomial”, 三折交叉验证), 以交叉验证误差最小化的  $\lambda$  值确定最优特征子集; SVM-RFE 基于标准化表达数据, 利用 “mlbench” 和 “caret” 包迭代剔除低贡献特征; 随机森林通过 “randomForest” 包构建 500 棵决策树, 以基尼不纯度下降法评估特征重要性, 保留评分  $> 10$  的变量。最终取 3 种算法结果的交集确定核心靶点, 确保结果可靠性。

## 1.5 GEO 数据库中核心靶点表达与诊断效能的验证

为验证核心靶点, 本研究下载了 3 个 GEO 数据集: GSE39582 (566 个肿瘤样本和 19 个正常样本, 共计 585 例)、GSE44076 (98 个肿瘤样本和 148 个正常样本, 共计 246 例) 和 GSE74602 (30 个肿瘤样本和 30 个正常样本, 共计 60 例)。3 个数据集原始样本量之和为 891 例。采用 “sva” R 包中的 ComBat 算法去除数据集间的批次效应, 参数设置为默认值; 随后对合并后的表达矩阵进行分位数归一化处理, 以确保不同数据集间的数据可比性。合并后总样本量与原始样本量总和一致, 数据质量满足后续分析要求。将探针 ID 转换为基因符号, 按样本类型 (正常/肿瘤) 对样本进行分组。通过 limma 筛选差异表达基因 (DEGs) ( $|\log_2FC| > 1$ , 校正  $P < 0.05$ ), 并采用火山图/热图进行可视化。采用受试者工作特征 (ROC) 曲线及计算曲线下面积 (AUC) 值来评估核心生物标志物对 CRC 的诊断潜力。

## 1.6 GEO 数据库中核心靶点的免疫浸润分析

采用 GSVA (v1.46.0) R 软件包进行单样本基因集富集分析 (ssGSEA)<sup>[29]</sup>, 以检测 GEO 数据集中正常结直肠样本与 CRC 样本的免疫细胞浸润水平, 组间免疫细胞富集评分差异通过 Wilcoxon 秩和检验进行分析。此外, 通过相关性分析<sup>[30]</sup> 探究核心靶点表达与免疫细胞浸润之间的关联。

## 1.7 利用 UALCAN 数据库验证核心靶点

为进一步验证 CRC 中已识别核心靶点的临床相关性及预后价值, 本研究借助整合 TCGA 多组学与临床数据的 UALCAN 平台 (<http://ualcan.path.uab.edu/>)<sup>[31]</sup>, 对比 286 例 CRC 样本与 41 例正常样本中核心靶点的 mRNA 表达水平。采用  $t$  检验分析统计学显著性, 以箱线图可视化结果; 通过 “生存分析” 模块评估预后意义, 按中位表达水平将患者分为高表达组与中低表达组, 绘制 Kaplan-

Meier 生存曲线,并用对数秩检验比较总生存期(OS)差异。

### 1.8 分子对接及分子动力学模拟

核心靶点的三维结构数据来源于蛋白质数据库(PDB, <https://www.rcsb.org/>)<sup>[32]</sup>,PFOA 的分子结构则通过 PubChem 获取。使用 Open Babel 软件将 PFOA 结构转换为 Mol2 格式。靶蛋白经脱水、氢化处理后,通过 AutoDockTools-1.5.7 软件添加 Gasteiger 电荷。AutoDock Vina 生成的 pdbqt 文件对接结果导入 PyMOL 软件进行 pdb 格式转换。采用 PyMOL3.2 软件对结合能最高的重要蛋白质结构进行可视化展示<sup>[33]</sup>。

分子动力学模拟采用 Gromacs2022.3 软件<sup>[34-35]</sup>对 PFOA-CLDN1 复合物进行 100 ns 独立重复模拟 3 次。体系构建为十二面体周期性盒子,盒子边界距蛋白表面最小距离 1.0 nm,溶剂采用 TIP3P 水模型,并添加 Na<sup>+</sup>维持体系电中性。选用 Amber99sb-ildn 力场,能量最小化采用最速下降法。温度耦合采用 V-rescale 算法,耦合时间常数 0.1 ps,体系温度维持 300 K;压力耦合采用 Parrinello-Rahman 算法,耦合时间常数 2.0 ps,压力维持 1 bar。依次进行 10 ns NVT 平衡与 10 ns NPT 平衡,随后进行 100 ns 正式生产模拟,步长 2 fs。轨迹分析包括均方根偏差(RMSD)、回转半径(Rg)、溶剂可及表面积(SASA)、氢键,验证复合物的稳定性。

### 1.9 AOP 假说的构建

选取与 PFOA 暴露及 CRC 进展密切相关的表型作为潜在关键事件(KE)。基于靶点-表型关系及基因-表型网络,将这些表型对应的关联基因定义

为候选分子起始事件(MIE)。结合文献报道的证据等级与生物学合理性,构建假设性 AOP,以概念性串联 PFOA 暴露与 CRC 发生的关联机制。

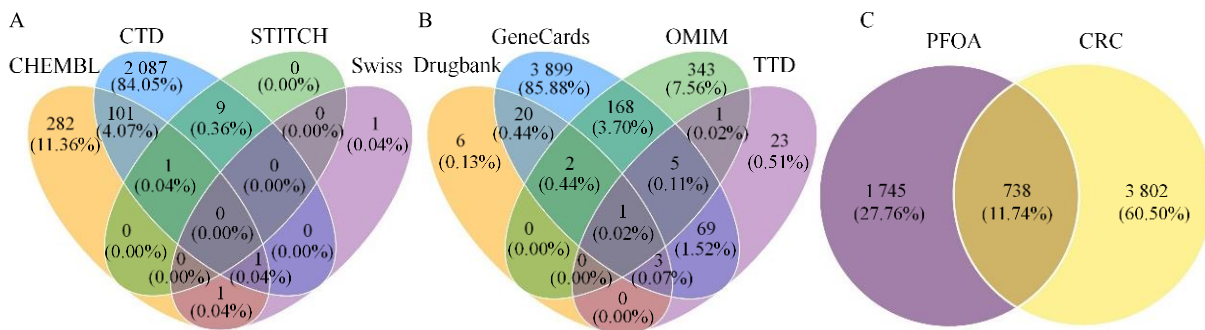
## 2 结果

### 2.1 PFOA 诱发 CRC 潜在靶点的获取

采用网络毒理学方法探究 PFOA 对 CRC 的潜在致病效应,系统收集并整合二者相关作用靶点:PFOA 相关靶点来源于 ChEMBL(386 个)、Swiss Target Prediction(3 个)、STITCH(10 个)及 CTD(2 199 个)数据库,经去重后获得 2 483 个靶点(图 1-A);CRC 相关靶点来源于 OMIM(520 个)、GeneCards(4 167 个)、TTD(102 个)及 DrugBank(32 个)数据库,经去重后获得 4 540 个靶点(图 1-B)。对 2 组靶点集进行交叉分析,共鉴定出 738 个交集靶点(图 1-C)。提示 PFOA 可能通过作用于这些与 CRC 致病相关的共同靶点,从而调控 CRC 的发生发展。上述结果提示 PFOA 与 CRC 存在大量共同作用靶点,二者可能通过共享的分子调控网络相互关联,为后续解析 PFOA 诱发 CRC 的潜在毒性靶点与分子机制奠定了靶点基础。

### 2.2 PFOA 诱导 CRC 交集靶点的功能富集分析

为阐明 PFOA 诱导 CRC 的潜在分子机制,本研究对 738 个交集靶点进行 GO 与 KEGG 通路富集分析(图 2)。选取各分类[生物学过程(BP)、细胞组分(CC)、分子功能(MF)]中错误发现率(FDR)最低的前 10 个术语进行可视化。在 BP 中,靶基因显著富集于“腺体发育”“对外源物质刺激的应答”和“凋亡信号通路的调控”,提示 PFOA 可能干扰内分泌稳态与组织重塑、激活细胞应激防御系统、并



A-不同数据库中 PFOA 相关靶点的交集分析; B-不同数据库中 CRC 相关靶点的交集分析; C-PFOA 相关靶点与 CRC 相关靶点的交集分析。

A-intersection analysis of PFOA-related targets obtained from different databases; B-intersection analysis of CRC-related targets obtained from different databases; C-intersection analysis of PFOA-related and CRC-related targets.

图 1 PFOA 与 CRC 相关靶点的筛选

Fig. 1 Identification of targets associated with PFOA and CRC

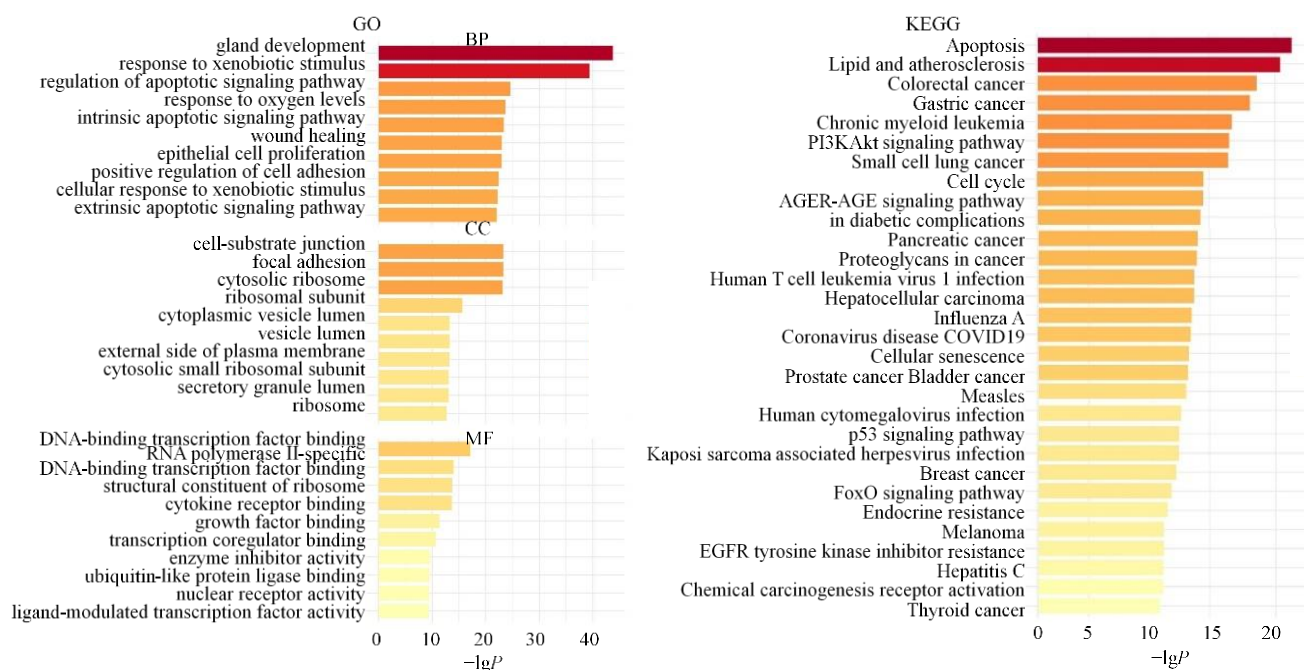


图 2 PFOA 诱导的 CRC 相关交集靶点的 GO 与 KEGG 通路富集分析

Fig. 2 GO and KEGG pathway enrichment analysis of CRC-related intersection targets induced by PFOA

破坏程序性死亡平衡，从而介导其多重毒性效应；在 CC 中，靶点高度集中于“细胞-基质连接”“黏着斑”和“胞质核糖体”，表明其作用可能与细胞黏附、迁移及蛋白质合成功能密切相关；在 MF 中，则主要富集于“DNA 结合转录因子结合”“RNA 聚合酶 II 特异性 DNA 结合转录因子结合”及“核糖体的结构成分”，突出了其在基因转录与蛋白质合成调控的潜在功能。KEGG 通路分析进一步揭示，这些靶基因显著富集于“凋亡”“脂质与动脉粥样硬化”及“结直肠癌”等通路。上述富集结果提示，PFOA 可能并非仅作用于单一通路，而是通过影响凋亡、脂质代谢及肿瘤相关通路的协同作用，从多方面推动肠上皮细胞恶性转化，进而参与 CRC 的发生发展。

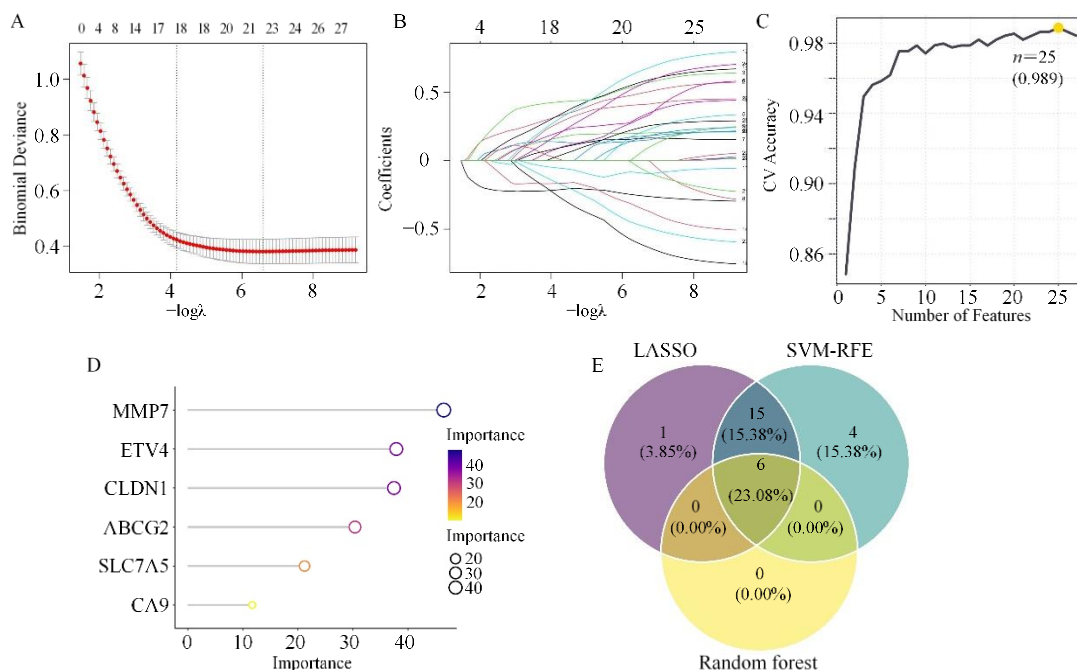
### 2.3 通过机器学习算法识别核心靶点

基于上述构建的“738 个基因×所有样本”表达矩阵，集成 3 种机器学习算法筛选核心基因。LASSO 回归（经交叉验证）鉴定出 22 个关键基因，SVM-RFE 算法鉴定出 25 个关键基因，随机森林算法则筛选出 6 个高重要性基因（图 3-A~D）。对 3 组结果取交集后，最终确定 6 个共同核心靶点：乳腺癌耐药蛋白(ABCG2)、紧密连接蛋白 1(CLDN1)、Ets 变异转录因子 4 (ETV4)、溶质载体家族 7 成员 5 (SLC7A5)、基质金属蛋白酶 7 (MMP7)、碳酸酐

酶 9 (CA9) (图 3-E)。该集成筛选策略有效提升了核心靶点的可靠性与稳定性，所筛选的 6 个核心靶点可能是 PFOA 诱发 CRC 的关键调控节点，可作为后续机制研究与风险评估的核心标志物。

### 2.4 基于 GEO 数据集的核心靶点差异表达分析与诊断评估

为验证核心靶点在独立队列中的表达模式与诊断价值，本研究整合了 GSE39582、GSE44076 和 GSE74602 3 个 CRC 转录组数据集。如图 4-A、B 所示，去批次效应和分位数标准化处理有效消除了数据集间的技术差异。差异表达分析共鉴定出 249 个 DEGs，并通过火山图及前 50 个 DEGs 的热图进行可视化（图 4-C、D）。随后，验证了 6 个核心靶标的表达及诊断潜力。与正常样本相比，CLDN1、ETV4、SLC7A5、MMP7 和 CA9 在 CRC 样本中的表达显著上调 ( $P < 0.01$ )，而 ABCG2 表达则明显下调 ( $P < 0.01$ ) (图 4-E)。诊断潜力评估通过绘制 ROC 曲线并计算 AUC 完成。6 个基因 (ABCG2、CLDN1、ETV4、SLC7A5、MMP7 和 CA9) 的 AUC 值分别为 0.887、0.882、0.864、0.866、0.826、0.707，均显示出良好的诊断区分能力（图 4-F）。以上结果表明，核心靶点在独立队列中表现出稳定的差异表达特征与良好的诊断效能，可作为区分 CRC 与正常组织的潜在生物标志物。



A-LASSO 回归的交叉验证曲线; B-LASSO 回归的特征系数路径; C-SVM-RFE 模型中特征数量与交叉验证准确率的关系; D-随机森林模型的特征重要性排序; E-LASSO、SVM-RFE 和随机森林 3 种算法筛选结果的交集分析。

A-cross-validation curve of LASSO regression model; B-feature coefficient profiles of LASSO regression model; C-relationship between number of selected features and cross-validation accuracy in SVM-RFE; D-feature importance ranking based on the random forest model; E-intersection of targets identified by LASSO, SVM-RFE, and random forest.

图 3 基于机器学习算法的 PFOA 诱导 CRC 核心靶点鉴定

Fig. 3 Identification of candidate core PFOA-CRC targets using machine learning algorithms

## 2.5 基于 GEO 数据集的核心靶点与免疫细胞浸润的相关性分析

为呈现肿瘤免疫微环境特征,通过免疫浸润分析及 22 种免疫细胞类型的关联研究,对比了正常组 ( $n=136$ ) 与 CRC 组 ( $n=499$ ) 的差异。结果显示两组间免疫组成存在显著差异: CRC 样本中  $CD4^+$  记忆活化 T 细胞、活化 NK 细胞、M0 型巨噬细胞、M1 型巨噬细胞、活化肥大细胞及中性粒细胞的浸润水平显著升高;而正常组中记忆 B 细胞、浆细胞、 $CD8^+$  T 细胞、 $CD4^+$  记忆静息 T 细胞、调节性 T 细胞 (Tregs)、 $\gamma\delta$  T 细胞、M2 型巨噬细胞、静息树突状细胞及静息肥大细胞的浸润水平显著高于 CRC 组 (图 5-A、B)。进一步探究 6 个核心靶点与免疫浸润模式的关系 (图 5-C): ABCG2 与浆细胞、 $CD8^+$  T 细胞、静息记忆  $CD4^+$  T 细胞、 $\gamma\delta$  T 细胞、M2 型巨噬细胞及静息肥大细胞等抗肿瘤免疫细胞呈正相关 ( $r=0.12\sim0.48$ ,  $P<0.05$ ), 而与初始  $CD4^+$  T 细胞、活化记忆  $CD4^+$  T 细胞、M0 型巨噬细胞、活化肥大细胞及中性粒细胞等免疫抑制相关细胞呈负相关 ( $r=-0.42\sim-0.18$ ,  $P<0.05$ )。CLDN1、

ETV4、SLC7A5、MMP7 及 CA9 则表现出一致的免疫调节倾向: 均与初始  $CD4^+$  T 细胞、活化记忆  $CD4^+$  T 细胞、M0 型巨噬细胞及活化肥大细胞呈正相关 ( $r=0.01\sim0.42$ ,  $P<0.05$ ), 且与浆细胞、 $CD8^+$  T 细胞、静息记忆  $CD4^+$  T 细胞、M2 型巨噬细胞及静息肥大细胞等抗肿瘤免疫细胞呈负相关 ( $r=-0.45\sim-0.09$ ,  $P<0.05$ ); 其中 CLDN1、ETV4、SLC7A5 与  $\gamma\delta$  T 细胞呈负相关, CA9 额外与静息树突状细胞呈负相关。结果提示, 核心靶点可通过双向调控抗肿瘤免疫与免疫抑制通路, 重塑 CRC 免疫微环境, 这可能是 PFOA 促进肿瘤免疫逃逸的重要免疫机制。

## 2.6 核心靶点独立临床验证

为进一步明确 6 个核心靶点的表达模式与临床关联, 利用 UALCAN 平台的 TCGA 转录组数据开展独立验证。结果与 GEO 3 个整合的数据集分析结论一致: 所有核心靶点在 CRC 组织中的表达均显著失调 ( $P<0.05$ , 图 6) 其中 CLDN1、ETV4、SLC7A5、MMP7 和 CA9 在 CRC 样本中显著上调, ABCG2 则明显下调 ( $P<0.001$ )。核心靶点在多队

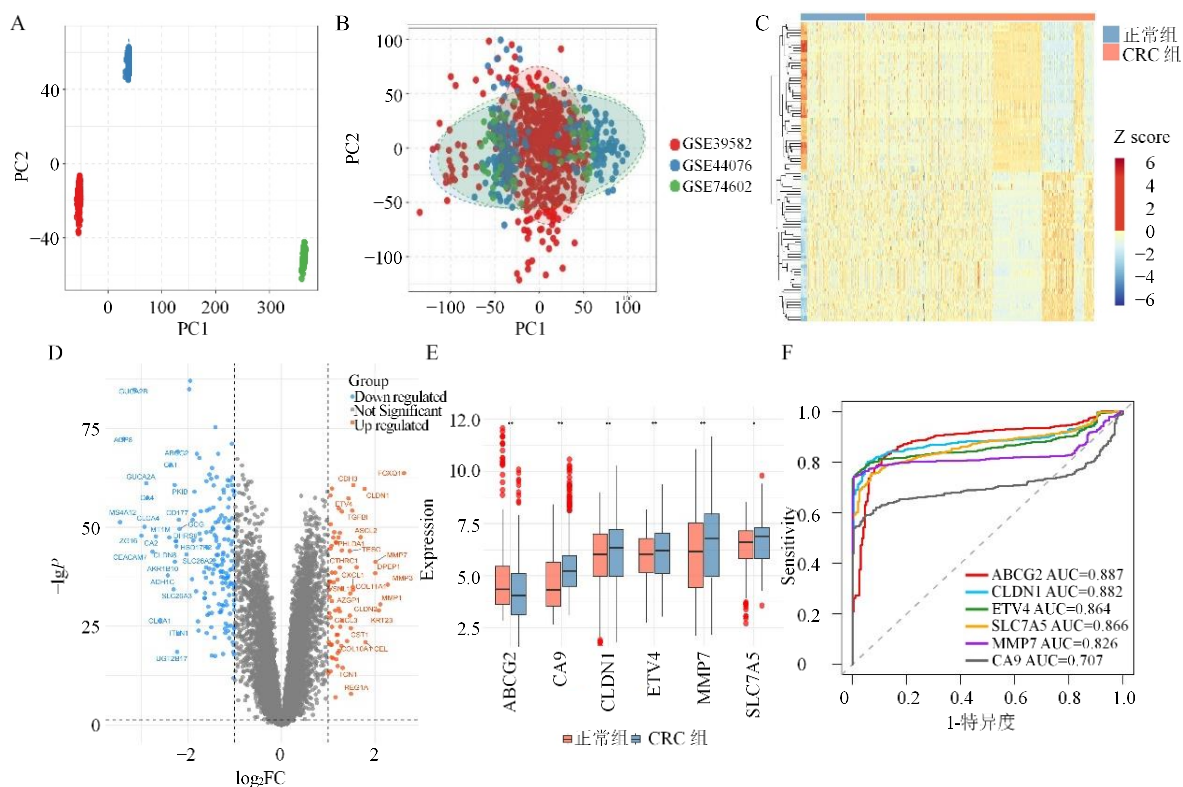


图 4 基于整合 GEO 数据集的核心靶点差异表达分析与验证  
Fig. 4 Expression and diagnostic performance of candidate core targets in the integrated GEO datasets  
A-批次效应校正前不同 GEO 数据集样本的主成分 (PCA) 分析; B-批次效应校正后不同 GEO 数据集样本的 PCA 分析; C-差异表达基因热图; D-差异表达基因火山图; E-6 个候选核心靶点在正常组织与 CRC 组织中的表达比较; F-6 个候选核心靶点区分 CRC 组织与正常组织的 ROC 曲线。  
A-principal component analysis (PCA) of samples from different GEO datasets before batch-effect correction; B-PCA of samples after batch-effect correction; C-heatmap of differentially expressed genes; D-volcano plot of differentially expressed genes; E-expression levels of the six candidate core targets in normal and CRC tissues; F-ROC curves of the six candidate core targets for discriminating CRC from normal tissues.

列中均表现出稳定的差异表达模式,进一步证实其与 CRC 发生发展的密切关联,可作为 PFOA 毒性评估的稳定标志物。

随后通过 Kaplan-Meier 生存分析探究核心靶点表达与 CRC 患者 OS 的关联,结果显示预后趋势存在差异 (图 7): *ABCG2*、*CLDN1*、*ETV4* 和 *CA9* 低表达且 *SLC7A5*、*MMP7* 高表达的患者随时间推移具有相对较高的生存概率。然而,核心靶点表达与患者 OS 仅呈现关联趋势,未达到统计学显著性 ( $P>0.05$ )。这说明核心靶点主要参与 CRC 的发生与进展过程,而非独立的预后标志物,其临床价值更体现在早期毒性预警与机制研究。

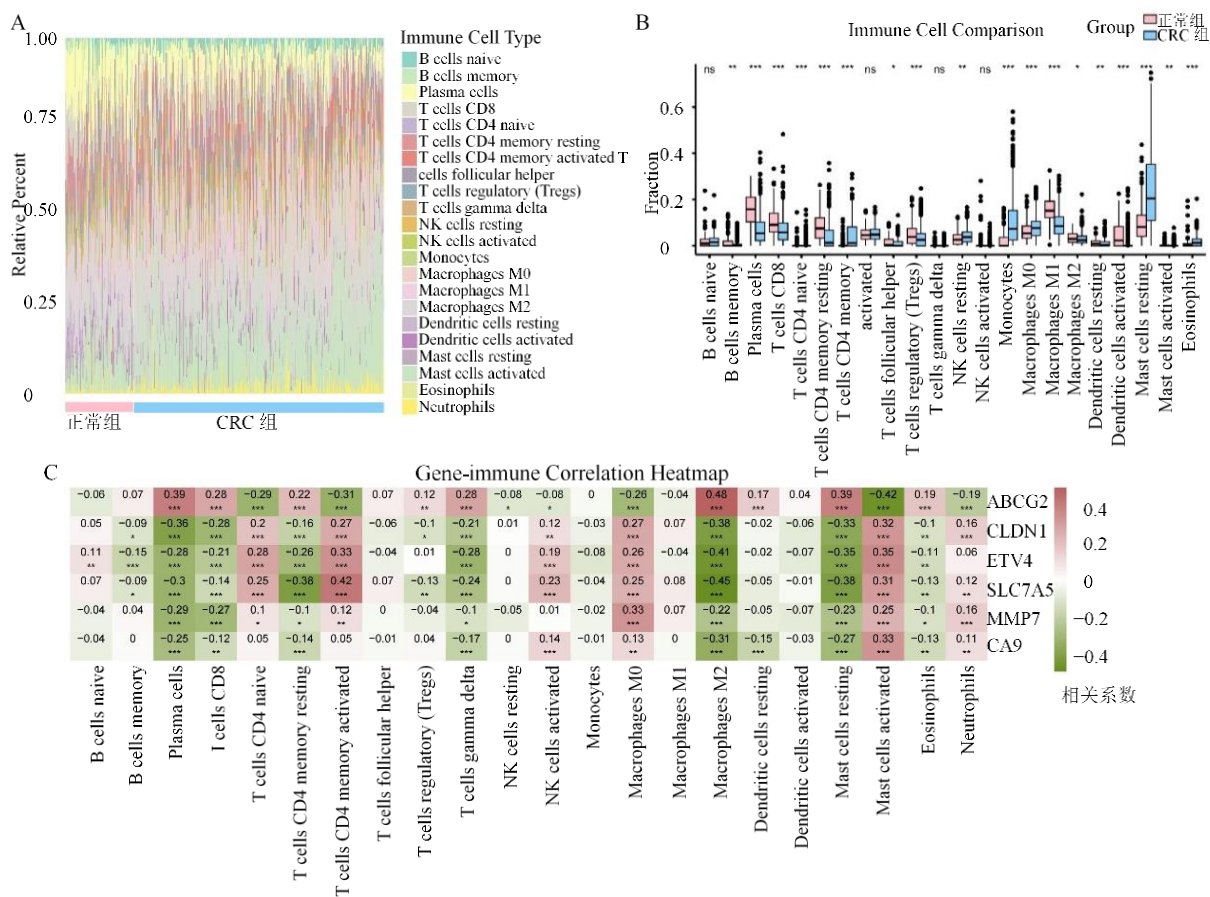
## 2.7 分子对接结果

作为网络毒理学框架的一部分,采用分子对接分析评估 PFOA 与 6 个核心靶点的相互作用。结果显示, PFOA 与 *ABCG2*、*CLDN1*、*ETV4*、*SLC7A5*、*MMP7* 和 *CA9* 的结合亲和力均低于  $-20.90 \text{ kJ} \cdot \text{mol}^{-1}$ ,

提示其具有良好的结合潜力<sup>[36]</sup>(图 8)。其中, PFOA 与 *CLDN1* 的结合亲和力最低 ( $-29.34 \text{ kJ} \cdot \text{mol}^{-1}$ ),预示二者之间存在强结合。PyMOL 可视化进一步揭示, PFOA 能够通过氢键与 *CLDN1* 的 CYS-157、LEU-185 位点结合,形成稳定的复合物,这可能是 PFOA 发挥结肠毒性的关键分子事件。

## 2.8 分子动力学模拟结果

本研究对 PFOA-*CLDN1* 复合物进行 3 次独立 100 ns 分子动力学重复模拟,结果趋势一致,表明模拟具有良好可重复性。沿 PC1 与 PC2 投影的吉布斯自由能景观图显示复合物形成稳定低能构象 (图 9-A)。蛋白质主链 RMSD 在 30 ns 后趋于收敛表明体系达到稳定平衡状态 (图 9-B)。回转半径 ( $R_g$ ) 与溶剂可及表面积 (SASA) 曲线平稳,证实复合物结构致密、疏水核心稳定 (图 9-C、D)。RMSF 分析显示,复合物在 *CLDN1* 的胞外环状区域呈现较高柔性波动,该区域恰好为 PFOA 结合口袋 (CYS-157、



A-正常组与 CRC 组样本中 22 种免疫细胞的相对浸润组成; B-正常组与 CRC 组 22 种免疫细胞浸润水平的比较; C-6 个候选核心靶点表达与 22 种免疫细胞浸润水平的相关性热图。

A-relative distribution of 22 immune cell types in normal and CRC samples; B-comparison of infiltration levels of 22 immune cell types between normal and CRC samples; C-correlation heatmap between the expression of six candidate core targets and the infiltration levels of 22 immune cell types.

图 5 基于整合 GEO 数据集的核心靶点与免疫细胞浸润的相关性

Fig. 5 Immune cell infiltration and its correlation with candidate core targets in integrated GEO datasets

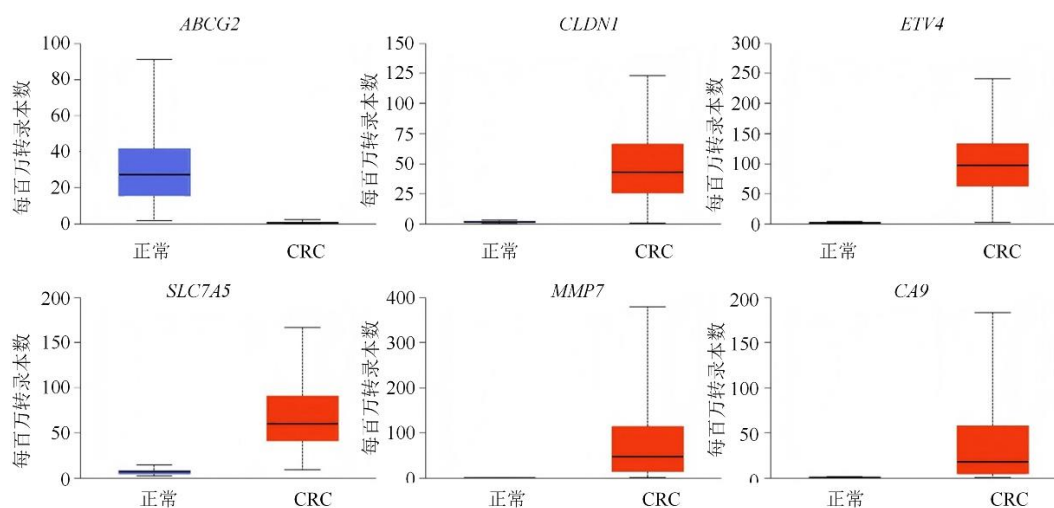


图 6 基于 UALCAN 数据库核心靶点差异表达分析

Fig. 6 Differential expression analysis of core targets based on UALCAN database

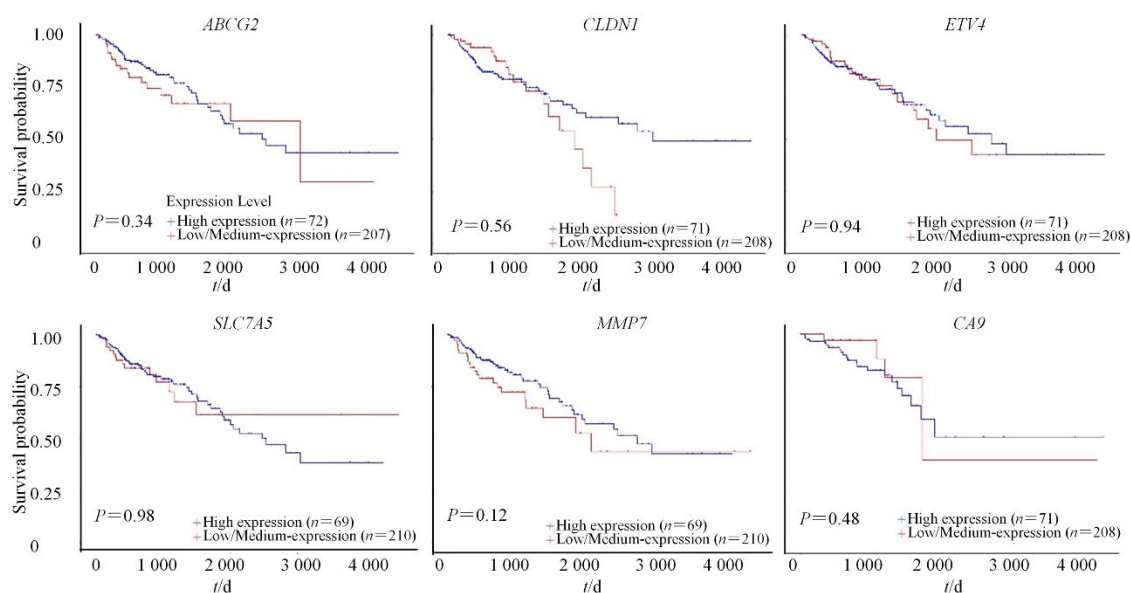


图 7 基于 UALCAN 数据库核心靶点的预后价值

Fig. 7 Prognostic value based on core targets from UALCAN database

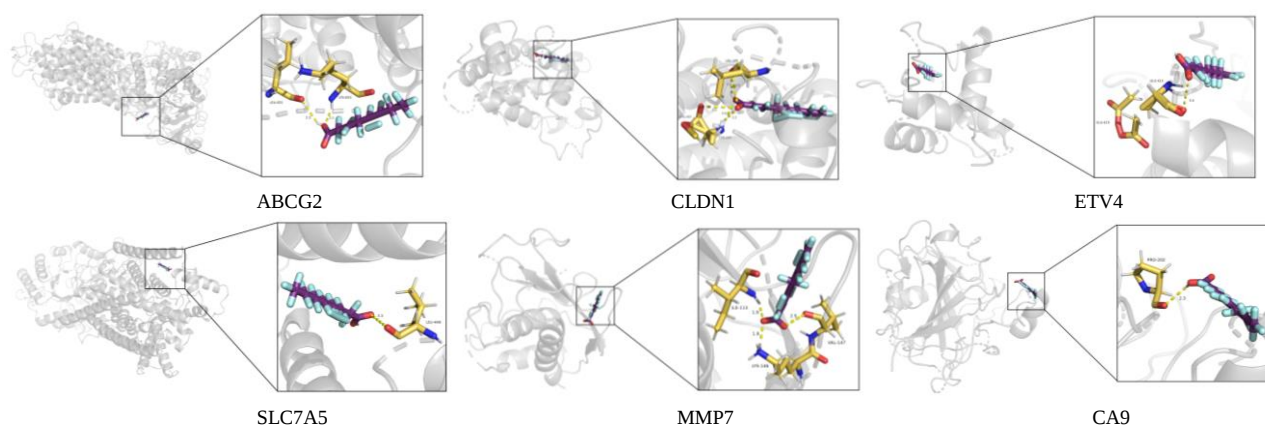


图 8 PFOA 与 6 个候选核心靶点的分子对接结果

Fig. 8 Molecular docking result of PFOA with six candidate core targets

LEU-185) 所在位置, 也是紧密连接功能与屏障调控的关键功能位点。高柔性特征有助于 PFOA 进入并稳定结合, 同时影响 CLDN1 的结构稳定性与屏障功能, 提示柔性波动区域直接对应 PFOA 作用位点与蛋白功能关键区 (图 9-E)。全程模拟中 PFOA 与 CLDN1 保持稳定结合, 氢键与疏水作用无明显断裂, 证实复合物构象稳定。

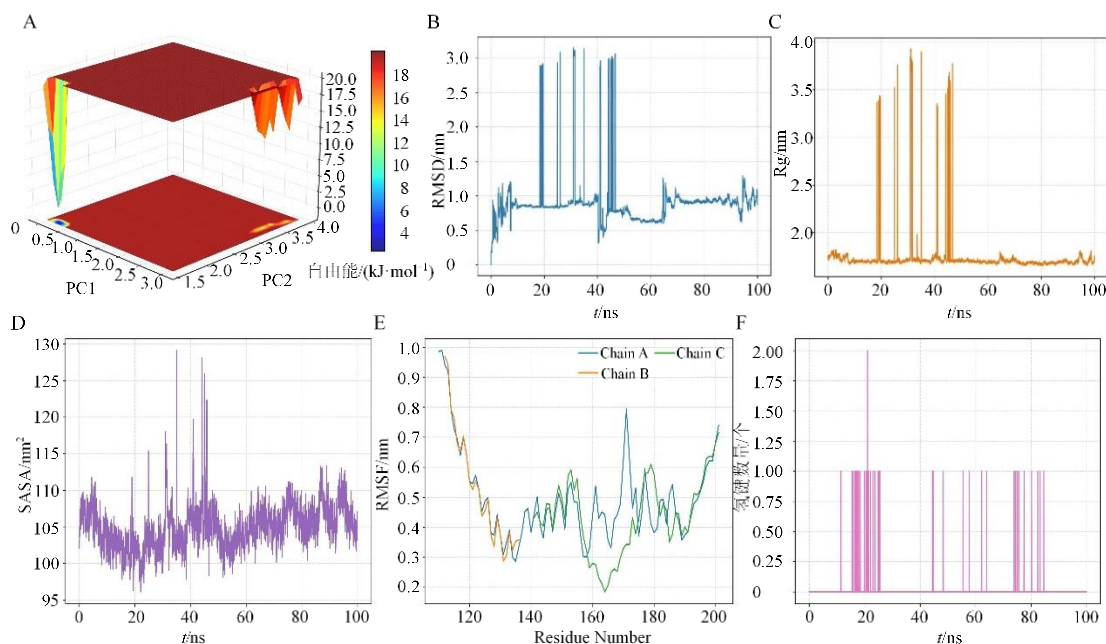
## 2.9 AOP 假说框架的构建

基于 Villeneuve 等<sup>[37]</sup>提出的方法论, 本研究开发了新型 AOP 框架 (图 10)。该框架以 PFOA 暴露为起点, 分子起始事件为 PFOA 诱导的 ABCG2、CLDN1、ETV4、SLC7A5、MMP7 和 CA9 异常表

达及功能改变, 这些核心靶点进而驱动 KE1 通路失调, 具体涉及“凋亡”“脂质与动脉粥样硬化”及“CRC”; 随后介导 KE2 结直肠免疫微环境失衡, 即通过核心靶点调控免疫细胞浸润实现; 与 KE3 结直肠上皮细胞基因组不稳定性及细胞周期紊乱, 最终导致 CRC 不良结局。该框架串联了 PFOA 暴露到 CRC 发生的主要环节, 可作为环境健康风险评估的参考框架。

## 3 讨论

PFOA 是一种普遍存在的环境污染物, 已证实可通过饮食摄入和饮用水污染在全球范围内导致人体暴露<sup>[38]</sup>。虽然先前的流行病学和毒理学研究已



A-复合物的吉布斯自由能景观; B-复合物蛋白质主链 RMSD 随时间的变化; C-复合物 Rg 随时间的变化; D-复合物 SASA 随时间的变化; E-CLDN1 残基 RMSF 分析; F-PFOA 与 CLDN1 之间氢键数量随时间的变化。

A-Gibbs free-energy landscape of complex; B-RMSD of protein backbone over time; C-Rg of complex over time; D-SASA of complex over time; E-RMSF of CLDN1 residues; F-number of hydrogen bonds between PFOA and CLDN1 over time.

图 9 PFOA-CLDN1 复合物的分子动力学模拟

Fig. 9 Molecular dynamics simulation analysis of the PFOA-CLDN1 complex

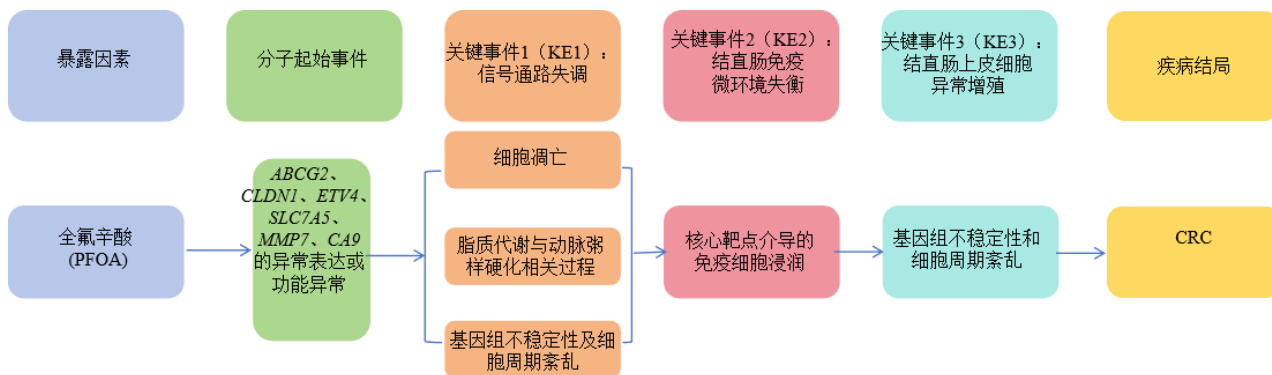


图 10 PFOA 诱导 CRC 的 AOP 假说

Fig. 10 AOP hypothesis for PFOA-induced CRC

确立 PFOA 的潜在致癌性及其与代谢紊乱的关联, 但对其在 CRC 发生发展中的具体分子机制仍缺乏系统阐明。本研究采用网络毒理学、机器学习与多维生物信息学相结合的整合计算方法, 系统揭示了 PFOA 暴露与 CRC 发病机制潜在的分子毒理学关联。研究发现 PFOA 可能通过 ABCG2、CLDN1、SLC7A5、ETV4、MMP7 和 CA9 6 个核心靶点破坏关键信号通路、重塑肿瘤免疫微环境, 最终驱动 CRC 的发生发展。

KEGG 通路富集分析揭示 3 条显著改变的通

路, 共同构成促进 CRC 发生的协同网络。凋亡通路的显著富集表明 PFOA 可能破坏程序性细胞死亡这一抵御癌变的关键防御机制, 已有研究证实环境污染物可通过失调 B 细胞淋巴瘤-2 (Bcl-2) 家族蛋白和半胱氨酸天冬氨酸特异性蛋白酶 (Caspase) 级联反应抑制细胞凋亡, 使受损细胞得以存活<sup>[39]</sup>。脂质与动脉粥样硬化通路的同步富集突显 PFOA 的代谢干扰作用, 作为已知的过氧化物酶体增殖物激活受体  $\alpha$  (PPAR $\alpha$ ) 激动剂, 该通路持续激活可重编程结肠上皮细胞脂质代谢, 创造支持肿瘤细胞增殖存

活的脂肪生成微环境<sup>[40]</sup>。CRC 通路的直接富集整合了上述效应,明确 PFOA 是结肠中已确立致癌级联反应的直接调节因子,可能涉及 Wnt/ $\beta$ -连环蛋白(Wnt/ $\beta$ -catenin)和磷脂酰肌醇 3-激酶/蛋白激酶 B(PI3K/Akt)等关键驱动通路<sup>[41]</sup>。细胞死亡受损、代谢异常和致癌信号激活这 3 大通路失调,为 PFOA 在 CRC 发生发展中的作用提供了连贯机制基础。

本研究采用 LASSO 回归、SVM-RFE、随机森林 3 种机器学习算法集成筛选核心靶点,三者具有良好互补性与稳定性优势:LASSO 回归可在高维数据中自动进行特征选择与参数收缩,有效排除多重共线性干扰;SVM-RFE 基于非线性核函数递归剔除低贡献特征,对复杂调控关系的挖掘更为灵敏;随机森林基于决策组合评估特征重要性,对噪声数据耐受性更强、不易过拟合。3 种算法原理不同、优势互补,取三者交集作为最终核心靶点,可显著降低单一模型偏倚,提升结果可靠性与稳定性。在此基础上,机器学习整合分析筛选出的 6 个核心靶点(ABCG2、CLDN1、ETV4、SLC7A5、MMP7、CA9),在 CRC BP 中功能多样且相互关联,其表达特征与作用机制如下:ABCG2 作为 ATP 结合盒转运蛋白及多药外排转运蛋白,在 CRC 中显著下调,其表达降低可能会减少内源性毒素和促癌代谢物的外排,增加基因组不稳定性<sup>[42]</sup>。CLDN1 是紧密连接关键组分,在 CRC 中显著上调。PFOA 可能靶向 CLDN1 破坏肠道屏障功能,为肠上皮细胞恶性转化创造条件,这与 PFOA 暴露损害溃疡性结肠炎患者肠道屏障的观察结果相符。其过表达与 CRC 上皮极性丧失、细胞运动性增强及转移相关,还会激活原癌基因酪氨酸蛋白激酶 Src (Src) /Akt 等致癌信号通路<sup>[43]</sup>。ETV4 为 ETS 家族转录因子,在 CRC 中上调且与抗肿瘤免疫细胞呈负相关,通过调节 MMP 和整合素促进上皮-间质转化及侵袭,契合其促肿瘤功能<sup>[44]</sup>。SLC7A5 (LAT1) 作为氨基酸转运蛋白,在 CRC 中上调且与抗肿瘤免疫细胞呈负相关,对癌细胞营养感知和哺乳动物雷帕霉素靶蛋白复合物 1 (mTORC1) 激活至关重要,可满足快速增殖 CRC 细胞的合成代谢需求<sup>[45]</sup>。MMP7 在 CRC 中上调表明其既能介导细胞外基质降解促进肿瘤侵袭转移,还能加工生长因子推动肿瘤血管生成,高表达常与肿瘤分化不良相关<sup>[46]</sup>。CA9 是缺氧诱导酶,在 CRC 中上调且与抗肿瘤免疫细胞呈负相关,与肿瘤微环境缺氧及免疫抑制密切相关,可调节细胞

内 pH 值以适应酸性微环境,且与不良预后相关<sup>[47]</sup>。因此,PFOA 暴露可能导致这些靶点协同失调,在激活促肿瘤进程的同时,抑制机体保护性机制。

免疫浸润分析表明 CRC 与正常肠黏膜组织存在显著的免疫组成差异,PFOA 可通过调控核心靶点重塑 CRC 免疫微环境,形成免疫监视缺陷与免疫抑制强化的双重病理特征。ABCG2 作为关键免疫调控分子,与 CD8<sup>+</sup>T 细胞、 $\gamma\delta$ T 细胞等抗肿瘤免疫细胞呈正相关,同时也与促肿瘤的 M2 型巨噬细胞相关。提示 ABCG2 并非单纯参与抗肿瘤免疫微环境构建,而是双向调控抗肿瘤与免疫抑制细胞浸润,参与复杂的肿瘤免疫调控网络<sup>[48]</sup>。此外,CLDN1、ETV4、SLC7A5、MMP7 及 CA9 呈现一致的免疫抑制促进模式,与 M0 型巨噬细胞、活化的肥大细胞及中性粒细胞呈正相关,而与 CD8<sup>+</sup>T 细胞及浆细胞呈负相关,提示这些分子可能通过协同机制推动肿瘤微环境向免疫耐受状态转变。这种协同作用体现在以下层面:第一,上游信号通路协同激活。缺氧作为实体瘤核心微环境特征,可同步上调 CA9、MMP7 与 CLDN1 表达:CA9 通过招募髓源性抑制细胞抑制 T 细胞功能<sup>[49]</sup>,MMP7 通过切割 Fas 配体削弱免疫介导的细胞毒性<sup>[50]</sup>。研究表明,PFOA 暴露可激活核因子- $\kappa$ B (NF- $\kappa$ B) 信号通路,进而驱动 MMP-2/MMP-9 等协同表达,加速免疫抑制微环境形成<sup>[51]</sup>。第二,分子间功能级联与正反馈放大。MMP7 不仅直接发挥免疫抑制作用,还可降解紧密连接蛋白,破坏肠道屏障功能,加剧局部炎症微环境,为免疫抑制细胞浸润创造条件<sup>[50]</sup>;CLDN1 异常表达可改变肠上皮通透性,进而通过 TLR 信号通路激活 NF- $\kappa$ B<sup>[43]</sup>;SLC7A5 可通过调控 mTOR 信号通路影响 T 细胞活化与功能<sup>[52]</sup>,其过表达可能与 MMP7 协同,共同抑制 CD8<sup>+</sup>T 细胞效应功能。第三,基质重塑与免疫细胞募集的时空协同。CA9 介导的微环境酸化可增强 MMP7 酶活性,抑制效应 T 细胞功能,同时促进巨噬细胞向 M2 型极化;MMP7 则通过基质降解招募免疫抑制细胞,二者形成酸化酶解招募级联反应,显著放大免疫抑制效应。第四,PFOA 与免疫抑制分子的协同调控网络。流行病学研究表明,PFAS 暴露与 CRC 患者转移性淋巴结数量增加呈正相关[PFOA 暴露者肠系膜淋巴结(MLNs)增加 27%,95%置信区间(CI):1%~60%],提示 PFAS 可能通过免疫抑制机制影响肿瘤进展<sup>[53]</sup>。体外研究证实,PFOA 可显著抑制

T 细胞依赖性抗体产生,降低免疫球蛋白 M (IgM) 水平,并减少 B 细胞向浆细胞分化<sup>[54]</sup>,这与本研究中浆细胞与免疫抑制分子负相关结果一致。同时, PFOA 可通过激活 NF- $\kappa$ B 通路、干扰 PPAR 信号的双重机制,协同上调 CLDN1、MMP7 等免疫抑制分子,共同构建免疫抑制微环境。综上, CLDN1、MMP7、CA9 等分子通过共同上游信号、功能级联及 PFOA 诱导的信号失衡,构成相互强化的免疫抑制网络。PFOA 通过下调 ABCG2 削弱抗肿瘤免疫轴,同时激活该网络,双重驱动肿瘤免疫逃逸。

TCGA 数据也验证了 6 个核心靶点在 CRC 组织中存在差异表达。虽然 Kaplan-Meier 分析未达到统计学显著性 ( $P>0.05$ ),但其趋势与已知生物学功能一致。例如,高 SLC7A5 和 MMP7 表达与较差的生存趋势一致,这与它们促进肿瘤生长和侵袭的作用相符。这些靶点的主要临床价值可能不在于作为独立的预后标志物,而在于作为多基因特征的补充组成部分,或作为患者分层的特定生物过程的指标。

分子对接与 100 ns 分子动力学模拟从原子层面揭示了 PFOA 与核心靶点的相互作用机制,3 次独立重复模拟均证实 PFOA-CLDN1 复合物构象稳定,结果可靠。分子对接显示 PFOA 与核心靶点结合亲和力均低于  $-20.90 \text{ kJ}\cdot\text{mol}^{-1}$ ,其中与 CLDN1 结合最强 ( $-29.34 \text{ kJ}\cdot\text{mol}^{-1}$ )。模拟轨迹分析显示, RMSF 高柔性区域集中于 CLDN1 胞外环,该区域恰好是 PFOA 结合口袋 (CYS-157、LEU-185) 与肠道屏障功能关键位点,提示环区柔性增强是 PFOA 干扰 CLDN1 紧密连接功能的重要结构基础。进一步分析表明, PFOA 通过氢键与疏水作用稳定占据 CLDN1 疏水口袋,干扰其屏障功能并引发下游代谢与免疫通路紊乱,推动结直肠上皮细胞发生代谢重编程,这一机制与既往研究中 CLDN1 功能异常参与结肠炎相关癌变早期进程的结论相符<sup>[55]</sup>。PFOA 与其余靶点的结合同样以疏水作用为主,提示 PFOA 作为多靶点干扰剂,影响转运蛋白 (ABCG2、SLC7A5)、转录因子 (ETV4) 及功能酶 (MMP7、CA9) 的正常生物学功能,通过协同扰乱细胞代谢、信号传导、免疫调控等多个生理过程,协同促进 CRC 发生发展。

本研究构建的 AOP 框架,以 PFOA 暴露作为起始事件,以 PFOA 与 ABCG2、CLDN1、ETV4、SLC7A5、MMP7、CA9 等靶点结合作为分子起始事件。从时间先后与生物学合理性来看, PFOA 作为

一种常见的全氟化合物,能够透过细胞膜进入细胞并与上述靶点结合,由此启动第一级关键事件 (KE1): 主要表现为细胞凋亡受到抑制、脂质代谢出现紊乱、CRC 相关信号通路异常激活,最终导致肠上皮细胞发生早期损伤。这一过程符合环境化学物先结合分子靶点、再引起细胞通路紊乱的基本规律,已有研究显示, PFOA 可在较短时间内进入细胞并引发早期损伤,为 KE1 优先发生提供了时间上的合理性。在此基础上,早期细胞损伤进一步通过损伤相关分子模式释放,在数天至数周内逐步引发免疫监视能力下降 (KE2),具体体现为免疫检查点调控失衡、抗肿瘤免疫细胞浸润减少。这一延迟过程符合细胞损伤 $\rightarrow$ 局部炎症 $\rightarrow$ 免疫状态改变的递进规律,与全氟化合物免疫毒性的发生时间过程基本一致。长期持续的免疫监视下降,再伴随基因组不稳定性逐步累积,最终在数月至数年的时间范围内推动肠上皮细胞恶性转化 (AO)。这一过程符合多数环境致癌物长期暴露 $\rightarrow$ 慢性损伤 $\rightarrow$ 逐步癌变的基本时间规律,也与人群流行病学中观察到的 PFOA 相关肿瘤潜伏期一致<sup>[56]</sup>。

值得注意的是,本研究基于计算毒理学与生物信息学预测的 PFOA 毒性效应,需结合真实环境暴露浓度与实验暴露浓度的差异客观解读<sup>[57]</sup>,普通人群血清 PFOA 浓度多在  $\text{ng}\cdot\text{mL}^{-1}$  级别,环境水体与饮用水中 PFOA 浓度多处于  $\text{ng}\cdot\text{L}^{-1}$  级别,仅高污染场地可达  $\mu\text{g}\cdot\text{L}^{-1}$  级别<sup>[58]</sup>,而多数体外与动物实验采用的 PFOA 浓度常为  $\mu\text{mol}\cdot\text{L}^{-1}$  至  $\text{mmol}\cdot\text{L}^{-1}$  级别,显著高于真实环境暴露水平<sup>[59]</sup>,尽管本研究为计算模拟且未涉及具体暴露剂量,但 PFOA 的生物富集性、长半衰期及低剂量长期累积效应已被证实可通过慢性持续作用调控信号通路及免疫微环境,推动 CRC 发生发展<sup>[60]</sup>,因此,本研究所提示的核心靶点与通路机制,更符合人群长期低剂量暴露的真实健康风险,可为流行病学研究与环境健康风险评估提供更贴合实际的机制依据。从公共卫生角度出发,本框架明确了潜在干预靶点,通过对核心靶点或关键通路进行干预,有望阻断 PFOA 的致病效应;同时结合 PFOA 在工业与消费品中的广泛使用及其环境持久性带来的 CRC 风险,可为更新其环境暴露限值提供理论支持。此外,计算毒理学与临床数据的结合,也进一步提升了该框架在环境风险评估中的应用价值<sup>[61]</sup>。

本研究存在一定局限性:其一,研究结论主要

基于公共数据库挖掘与计算生物学预测, PFOA 对核心靶点及通路的因果调控关系, 需通过细胞与动物实验进一步验证。其二, 真实环境中 PFOA 常与其他污染物共存形成混合暴露, 其潜在叠加或协同毒性未在单一化合物研究中体现, 联合效应有待深入探究<sup>[62]</sup>。其三, 预后分析显示核心靶点表达与患者 OS 未达到统计学显著性, 可能受样本量、临床分期异质性及治疗差异等影响, 需扩大样本量并开展多中心前瞻性研究进一步验证。其四, 未纳入个体遗传易感性、PFOA 相关肠道菌群紊乱等因素, 此类因素可调控机体对 PFOA 致 CRC 的易感程度<sup>[63]</sup>。此外, 研究所构建 AOP 框架中关键事件的时间序列与剂量-反应关系尚不明确, 各节点间的直接因果关联仍依赖推论, 缺乏实验证据支撑, 需通过纵向流行病学研究追踪高危人群暴露与发病轨迹以验证其科学性。未来应开展结合精确暴露评估、多组学整合分析的流行病学研究, 推动实验室机制发现向人群健康风险评估转化。

#### 4 结论

本研究整合网络毒理学、机器学习与多维生物信息学方法, 系统阐明了 PFOA 诱导 CRC 的分子毒理机制。通过机器学习筛选出 738 个候选靶点并鉴定出 ABCG2、CLDN1、ETV4、SLC7A5、MMP7 和 CA9 6 个核心靶点, 这些靶点在 CRC 样本中显著失调且与疾病相关免疫浸润密切相关。分子对接及 100 ns 分子动力学模拟证实 PFOA 与核心靶点稳定结合, 其中与 CLDN1 的结合稳定性尤为突出。基于上述研究发现, 本研究进一步构建了 PFOA 诱导 CRC 的 AOP 理论框架, 系统性梳理了从分子起始事件、关键生物学事件到最终不良结局的完整毒理学链条。这一整合分析思路弥补了既往单靶点研究的不足, 可为 PFOA 相关环境健康风险评估提供新的方法参考, 也为 CRC 的环境致病机制研究与潜在干预靶点探索提供一定理论依据。

**利益冲突** 所有作者均声明不存在利益冲突

#### 参考文献

- [1] Siegel R L, Miller K D, Wagle N S, et al. Cancer statistics, 2023 [J]. CA Cancer J Clin, 2023, 73(1): 17-48.
- [2] Keum N, Giovannucci E. Global burden of colorectal cancer: Emerging trends, risk factors and prevention strategies [J]. Nat Rev Gastroenterol Hepatol, 2019, 16(12): 713-732.

- [3] Arnold M, Sierra M S, Laversanne M, et al. Global patterns and trends in colorectal cancer incidence and mortality [J]. Gut, 2017, 66(4): 683-691.
- [4] Ohoro C R, Amaku J F, Conradie J, et al. Effect of physicochemical parameters on the occurrence of per- and polyfluoroalkyl substances (PFAS) in aquatic environment [J]. Mar Pollut Bull, 2024, 208: 117040.
- [5] Lau C, Anitole K, Hodes C, et al. Perfluoroalkyl acids: A review of monitoring and toxicological findings [J]. Toxicol Sci, 2007, 99(2): 366-394.
- [6] Calafat A M, Wong L Y, Kuklenyik Z, et al. Polyfluoroalkyl chemicals in the U.S. population: Data from the National Health and Nutrition Examination Survey (NHANES) 2003-2004 and comparisons with NHANES 1999-2000 [J]. Environ Health Perspect, 2007, 115(11): 1596-1602.
- [7] IARC Working Group on the Identification of Carcinogenic Hazards to Humans. *Perfluorooctanoic Acid (PFOA) and Perfluorooctanesulfonic Acid (PFOS)* [M]. Lyon (FR): International Agency for Research on Cancer, 2025.
- [8] SeyyedSalehi M S, Boffetta P. Per- and poly-fluoroalkyl substances (PFAS) exposure and risk of kidney, liver, and testicular cancers: A systematic review and Meta-analysis [J]. Med Lav, 2023, 114(5): e2023040.
- [9] Bjork J A, Wallace K B. Structure-activity relationships and human relevance for perfluoroalkyl acid-induced transcriptional activation of peroxisome proliferation in liver cell cultures [J]. Toxicol Sci, 2009, 111(1): 89-99.
- [10] Pierozan P, Jerneren F, Karlsson O. Perfluorooctanoic acid (PFOA) exposure promotes proliferation, migration and invasion potential in human breast epithelial cells [J]. Arch Toxicol, 2018, 92(5): 1729-1739.
- [11] Corsini E, Luebke R W, Germolec D R, et al. Perfluorinated compounds: Emerging POPs with potential immunotoxicity [J]. Toxicol Lett, 2014, 230(2): 263-270.
- [12] Steenland K, Zhao L P, Winkvist A, et al. Ulcerative colitis and perfluorooctanoic acid (PFOA) in a highly exposed population of community residents and workers in the mid-Ohio valley [J]. Environ Health Perspect, 2013, 121(8): 900-905.
- [13] Giménez-Bastida J A, Surma M, Zieliński H. *In vitro* evaluation of the cytotoxicity and modulation of mechanisms associated with inflammation induced by perfluorooctanesulfonate and perfluorooctanoic acid in human colon myofibroblasts CCD-18Co [J]. Toxicol In Vitro, 2015, 29(7): 1683-1691.
- [14] Wang Y L, Xiao J, Suzek T O, et al. PubChem's BioAssay

- database [J]. *Nucleic Acids Res*, 2012, 40(D1): D400-D412.
- [15] Gaulton A, Bellis L J, Bento A P, et al. ChEMBL: A large-scale bioactivity database for drug discovery [J]. *Nucleic Acids Res*, 2012, 40(D1): D1100-D1107.
- [16] 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.
- [17] Szklarczyk D, Gable A L, Lyon D, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets [J]. *Nucleic Acids Res*, 2019, 47(D1): D607-D613.
- [18] Davis A P, Murphy C G, Saraceni-Richards C A, et al. Comparative Toxicogenomics Database: A knowledgebase and discovery tool for chemical-gene-disease networks [J]. *Nucleic Acids Res*, 2009, 37(suppl\_1): D786-D792.
- [19] Consortium T U. UniProt: The universal protein knowledgebase [J]. *Nucleic Acids Res*, 2017, 45(D1): D158-D169.
- [20] Hamosh A, Amberger J S, Bocchini C, et al. Online Mendelian inheritance in man (OMIM®): Victor McKusick's magnum opus [J]. *Am J Med Genet A*, 2021, 185(11): 3259-3265.
- [21] Safran M, Solomon I, Shmueli O, et al. GeneCards 2002: Towards a complete, object-oriented, human gene compendium [J]. *Bioinformatics*, 2002, 18(11): 1542-1543.
- [22] Zhang Y T, Zhou Y, Xu H W, et al. Therapeutic target database 2026: Facilitating targeted therapies and precision medicine [J]. *Nucleic Acids Res*, 2026, 54(D1): D1692-D1701.
- [23] Wishart D S, Feunang Y D, Guo A C, et al. DrugBank 5.0: A major update to the DrugBank database for 2018 [J]. *Nucleic Acids Res*, 2018, 46(D1): D1074-D1082.
- [24] Dennis G Jr, Sherman B T, Hosack D A, et al. DAVID: Database for annotation, visualization, and integrated discovery [J]. *Genome Biol*, 2003, 4(5): P3.
- [25] Edgar R, Domrachev M, Lash A E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository [J]. *Nucleic Acids Res*, 2002, 30(1): 207-210.
- [26] Friedman J H, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent [J]. *J Stat Softw*, 2010, 033(i01): 1-22.
- [27] Sanz H, Valim C, Vegas E, et al. SVM-RFE: Selection and visualization of the most relevant features through non-linear kernels [J]. *BMC Bioinform*, 2018, 19(1): 432.
- [28] Ding J Y, Du J C, Wang H J, et al. A novel two-stage feature selection method based on random forest and improved genetic algorithm for enhancing classification in machine learning [J]. *Sci Rep*, 2025, 15: 16828.
- [29] Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data [J]. *BMC Bioinformatics*, 2013, 14: 7.
- [30] Akoglu H. User's guide to correlation coefficients [J]. *Turk J Emerg Med*, 2018, 18(3): 91-93.
- [31] Chandrashekar D S, Bashel B, Balasubramanya S A H, et al. UALCAN a portal for facilitating tumor subgroup gene expression and survival analyses [J]. *Neoplasia*, 2017, 19(8): 649-658.
- [32] Bernstein F C, Koetzle T F, Williams G J B, et al. The protein data bank: A computer-based archival file for macromolecular structures [J]. *J Mol Biol*, 1977, 112(3): 535-542.
- [33] Pinzi L, Rastelli G. Molecular docking: Shifting paradigms in drug discovery [J]. *Int J Mol Sci*, 2019, 20(18): 4331.
- [34] Van Der Spoel D, Lindahl E, Hess B, et al. GROMACS: Fast, flexible, and free [J]. *J Comput Chem*, 2005, 26(16): 1701-1718.
- [35] Hess B, Kutzner C, van der Spoel D, et al. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation [J]. *J Chem Theory Comput*, 2008, 4(3): 435-447.
- [36] Pagadala N S, Syed K, Tuszynski J. Software for molecular docking: A review [J]. *Biophys Rev*, 2017, 9(2): 91-102.
- [37] Villeneuve D L, Crump D, Garcia-Reyero N, et al. Adverse outcome pathway (AOP) development I: Strategies and principles [J]. *Toxicol Sci*, 2014, 142(2): 312-320.
- [38] Fenton S E, Ducatman A, Boobis A, et al. Per- and polyfluoroalkyl substance toxicity and human health review: Current state of knowledge and strategies for informing future research [J]. *Environ Toxicol Chem*, 2021, 40(3): 606-630.
- [39] Kuo M L, Shiah S G, Wang C J, et al. Suppression of apoptosis by Bcl-2 to enhance benzene metabolites-induced oxidative DNA damage and mutagenesis: A possible mechanism of carcinogenesis [J]. *Mol Pharmacol*, 1999, 55(5): 894-901.
- [40] Behr A C, Kwiatkowski A, Ståhlman M, et al. Impairment of bile acid metabolism by perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS) in human HepaRG hepatoma cells [J]. *Arch Toxicol*, 2020, 94(5): 1673-1686.

- [41] 王敦方, 冯雪, 张彩娟, 等. 基于代谢重编程及相关信号途径探讨中药干预结直肠癌的研究进展 [J]. 中草药, 2026, 57(6): 2326-2336.  
Wang D F, Feng X, Zhang C J, et al. Research progress on traditional Chinese medicine intervention in colorectal cancer based on metabolic reprogramming and related signal pathways [J]. Chin Tradit Herb Drugs, 2026, 57(6): 2326-2336.
- [42] Polgar O, Robey R W, Bates S E. ABCG2: Structure, function and role in drug response [J]. Expert Opin Drug Metab Toxicol, 2008, 4(1): 1-15.
- [43] Dhawan P, Singh A B, Deane N G, et al. Claudin-1 regulates cellular transformation and metastatic behavior in colon cancer [J]. J Clin Invest, 2005, 115(7): 1765-1776.
- [44] Oh S, Shin S, Janknecht R. ETV1, 4 and 5: An oncogenic subfamily of ETS transcription factors [J]. Biochim Biophys Acta BBA Rev Cancer, 2012, 1826(1): 1-12.
- [45] Cormerais Y, Massard P A, Vucetic M, et al. The glutamine transporter ASCT2 (SLC1A5) promotes tumor growth independently of the amino acid transporter LAT1 (SLC7A5) [J]. J Biol Chem, 2018, 293(8): 2877-2887.
- [46] Gobin E, Bagwell K, Wagner J, et al. A pan-cancer perspective of matrix metalloproteases (MMP) gene expression profile and their diagnostic/prognostic potential [J]. BMC Cancer, 2019, 19(1): 581.
- [47] Chiche J, Brahimi-Horn M C, Pouyssegur J. Tumour hypoxia induces a metabolic shift causing acidosis: A common feature in cancer [J]. J Cell Mol Med, 2010, 14(4): 771-794.
- [48] Robey R W, Pluchino K M, Hall M D, et al. Revisiting the role of ABC transporters in multidrug-resistant cancer [J]. Nat Rev Cancer, 2018, 18(7): 452-464.
- [49] Chafe S C, McDonald P C, Saberi S, et al. Targeting hypoxia-induced carbonic anhydrase IX enhances immune-checkpoint blockade locally and systemically [J]. Cancer Immunol Res, 2019, 7(7): 1064-1078.
- [50] Vargo-Gogola T, Fingleton B, Crawford H C, et al. Matrilysin (matrix metalloproteinase-7) selects for apoptosis-resistant mammary cells *in vivo* [J]. Cancer Res, 2002, 62(19): 5559-5563.
- [51] Miao C, Ma J, Zhang Y, et al. Perfluorooctanoic acid enhances colorectal cancer DLD-1 cells invasiveness through activating NF- $\kappa$ B mediated matrix metalloproteinase-2/-9 expression. Int J Clin Exp Pathol. 2015;8(9):10512-10522. Published 2015 Sep 1.
- [52] Huang X W, Sun T T, Wang J L, et al. Metformin reprograms tryptophan metabolism to stimulate CD8<sup>+</sup> T-cell function in colorectal cancer [J]. Cancer Res, 2023, 83(14): 2358-2371.
- [53] Cui J, Shi J X, Gao X W, et al. Associations of exposure to per- and polyfluoroalkyl substances mixture with the numbers of lymph nodes in colorectal cancer patients [J]. Environ Res, 2024, 240(Pt 1): 117529.
- [54] Iulini M, Bettinsoli V, Maddalon A, et al. *In vitro* approaches to investigate the effect of chemicals on antibody production: The case study of PFASs [J]. Arch Toxicol, 2025, 99(5): 2075-2086.[PubMed]
- [55] Weber C R, Nalle S C, Tretiakova M, et al. Claudin-1 and claudin-2 expression is elevated in inflammatory bowel disease and may contribute to early neoplastic transformation [J]. Lab Invest, 2008, 88(10): 1110-1120.
- [56] Wang Z Y, Cousins I T, Scheringer M, et al. Hazard assessment of fluorinated alternatives to long-chain perfluoroalkyl acids (PFAAs) and their precursors: Status quo, ongoing challenges and possible solutions [J]. Environ Int, 2015, 75: 172-179.
- [57] Sunderland E M, Hu X C, Dassuncao C, et al. A review of the pathways of human exposure to poly- and perfluoroalkyl substances (PFASs) and present understanding of health effects [J]. J Expo Sci Environ Epidemiol, 2019, 29(2): 131-147.
- [58] Lewis-Michl E L, Forand S P, Hsu W H, et al. Perfluorooctanoic acid serum concentrations and half-lives in a community exposed to contaminated drinking water in New York State [J]. J Expo Sci Environ Epidemiol, 2025, 35(3): 403-413.
- [59] Li K, Gao P, Xiang P, et al. Molecular mechanisms of PFOA-induced toxicity in animals and humans: Implications for health risks [J]. Environ Int, 2017, 99:43-54.
- [60] Stoffels C B A, Cambier S, Subirana M A, et al. When subcellular chemical imaging enlightens our understanding on intestinal absorption, intracellular fate and toxicity of PFOA *in vitro* [J]. J Hazard Mater, 2024, 480: 136205.
- [61] Ankley G T, Bennett R S, Erickson R J, et al. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment [J]. Environ Toxicol Chem, 2010, 29(3): 730-741.
- [62] Kortenkamp A, Faust M. Regulate to reduce chemical mixture risk [J]. Science, 2018, 361(6399): 224-226.
- [63] Wong S H, Yu J. Gut microbiota in colorectal cancer: Mechanisms of action and clinical applications [J]. Nat Rev Gastroenterol Hepatol, 2019, 16(11): 690-704.