

# 基于机器学习的胃腺癌基因诱导药物再利用研究

师瑜英<sup>1,2\*</sup>

1. 山东大学齐鲁医学院公共卫生学院生物统计学系, 山东 济南 250002

2. 山东大学健康医疗大数据研究院, 山东 威海 264200

**摘要：**本研究通过药物再利用，开发了一个基于批量 RNA 转录组学胃腺癌药物的再利用计算流程。利用基因表达综合数据库（GEO）中的胃癌基因表达数据，我们鉴定出 1004 个差异表达基因，并将其作为潜在药物靶标的生成依据。此外，本项目从 DrugBank 中筛选了超过 10,000 种药物作为候选药物。为了深入探索药物再利用的潜力，我们构建了包括基因–基因相互作用和基因–靶标相互作用在内的多组学网络。这些网络有助于发现核心基因和再利用药物的目标，同时协助对药物进行有效排序。最终，结合药物–靶标相互作用的评分，我们对候选药物进行了综合排名。我们通过对排名靠前的药物在胃癌领域的临床试验证据和研究进行探讨，以验证这些药物的有效性和潜力。

**关键词：**药物再利用；机器学习；转录组学；胃癌

## Machine Learning-Based Study on Gene-Induced Drug Repurposing for Gastric Adenocarcinoma

Shi Yuying<sup>\*</sup>

1. Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Shandong Jinan 250002

2. Shandong University Health and Medical Big Data Research Institute, Shandong Weihai 264200

**Abstract :** In this study, we developed a computational pipeline for the repurposing of drugs for gastric adenocarcinoma, based on bulk RNA transcriptomic analysis. Using gene expression data from the Gene Expression Omnibus (GEO) database, we identified 1004 differentially expressed genes as potential drug targets. Additionally, over 10,000 drugs were screened from DrugBank as candidates. To further explore the potential of drug repurposing, we constructed multi-omics networks, including gene–gene and gene–target interactions. These networks aid in identifying core genes and targets for the repurposed drugs and assist in effective drug ranking. Ultimately, by combining drug–target interaction scores, a comprehensive ranking of the candidate drugs was conducted. We validated the efficacy and potential of the top-ranked drugs by exploring their clinical trial evidence and studies in the field of gastric cancer.

**Key words :** drug repurposing; machine learning; transcriptomics; gastric cancer

## 引言

胃癌是全球最常见的恶性肿瘤之一，也是全球主要的癌症死亡原因之一<sup>[1]</sup>。药物再利用策略旨在提出现有药物以抵抗癌细胞的生长，这种策略具有显著的优势，包括降低药物开发成本、节省时间，提高新治疗疾病的成功率<sup>[2]</sup>。基因组和药物数据可以帮助研究人员重新定位已研究药物的使用，并通过构建药物–疾病相互作用网络来研究它们的分子机制。在药物和疾病的比较中，通过比较药物治疗前后生物材料的基因表达情况，获取特定药物的转录组特性，然后将这些差异表达特性与通过差异表达分析获得的目标疾病和健康状况相关的表达情况进行比较，差异表达的基因为药物靶标提供了候选者<sup>[3]</sup>。在本研究中，我们基于转录组特性为胃癌开发了一种计算药物再利用方法，设计了一个端到端的计算流程（RepoGC）。我们利用 GEO 中的转录组数据，获得了 389 个胃癌样本中 1004 个差异基因，用于胃腺癌研究。我们通过 ShinyGO 进行基因富集分析，识别了作为药物靶标的中心关联和功能基因。我们还通过蛋白质–蛋白质相互作用网络推断出最大的相互作用靶标。最后，我们选择了具有目标的顶级候选药物进行进一步评估候选药物的可行性。

\* 作者简介：师瑜英 1995.10 女 汉 硕士胃癌的代谢组学及药物再利用 山西省晋中市平遥县

## 方法

### 数据集

我们从基因表达综合数据库中选择了四个与胃癌相关的转录组学档案，分别是：GSE13911、GSE54129、GSE19826、GSE11891。他们分别都包含不同数量的RNA测序数据及邻近正常或肿瘤配对的胃组织样本的基因表达数据。我们对这四个使用相同平台GPL570的微阵列数据集进行了整合。在对这4个数据集进行差异基因表达分析时，我们将胃肿瘤组织作为实验组，将非癌组织作为对照组进行比较。然后得出最为差异的基因。随后，通过在GEOexplorer上搜索相应的差异表达基因的ID，我们获得了与这四个基因表达数据集相关的1004个差异表达基因的信息。通过在Uniprot数据库上使用这些基因，映射了相应的药物靶标，再从DrugBank下载了超过10,000种具有结构的药物来进行筛选。

### 基因富集分析

基因富集分析用于将一组基因映射到多个功能途径上，旨在识别与差异表达基因相关的生物功能和途径。在本研究中，ShinyGO使用Bioconductor包通过API访问从KEGG网络服务器检索功能途径，构建一个包含丰富基因集的网络视图和层次聚类树。ShinyGO绘制了用户列表中所有基因的染色体位置，并对基因组特性进行了统计分析。

### 从DrugBank生成药物候选者

DrugBank（[www.drugbank.ca](http://www.drugbank.ca)）是一个在线数据库，包含有关药物、其机制、相互作用和靶标的分子信息。它可用于查看、排序和提取药物或药物-靶标数据，有助于深入分析药物作用和机制。在本研究中，提取了所有具有分子信息如结构、临床证据的药物。

### 药物和蛋白质靶标的特征化并进行相互作用预测

我们将药物结构从SMILE字符串转换为分子指纹。扩展连通性指纹（ECFPs）是一类结合了多种有用分子特征的特征化方法<sup>[4]</sup>。它们能将任意大小的分子转换成固定长度的向量，并使用相同的模型进行处理。指纹向量的每个元素表示特定分子特征的存在或缺失，这些特征由原子的某种局部排列定义。这些属性的每种独特组合都是一个特征，向量的相应元素被设置为1，以表示它们的存在。Seq2Vec用于将蛋白质序列表示为高维空间中的连续向量嵌入。这些嵌入捕捉了氨基酸及其基于大量蛋白质序列数据集中的共现模式的关系的语义信息。

DeepPurpose工具接受将化合物的SMILES和蛋白质氨基酸序列对作为输入，将学习到的蛋白质和化合物嵌入到MLP解码器中以生成预测，并返回预测分数，代表预测化合物和蛋白质之间结合的概率。

## 结果

### 差异基因表达分析和功能分析

我们通过ShinyGO鉴定了1004个差异表达基因并富集了989

个受干扰的功能途径。使用GSE13911和GSE54129数据对GC患者和健康对照组进行的差异基因表达分析，并对差异基因进行了功能富集分析。在胃癌发展中，KEGG中最显著改变的途径包括“蛋白质消化和吸收”、“胃酸分泌”和“Hippo信号途径”。这些途径已被确凿地证明与癌症有关。此外，GO分子功能中最显著改变的途径包括“苯并蒽9,10-单加氧酶活性”、“反式-1,2-二氢苯并二醇脱氢酶活性”、“靛醇脱氢酶活性”和“P型钾：质子转运蛋白活性”。最后，我们使用富集分析生成了一个网络图，这些途径之间的关联是基于重叠基因进行分析的。

### 通过DeepPurpose和iLINCS生成药物候选者

使用DrugBank，我们生成了11,912个药物候选。然后利用DeepPurpose测量药物-靶标相互作用，特别是药物分子与蛋白质靶标的结合。通过输入药物的SMILES和靶标的氨基酸序列，模型为每种药物生成一个结合分数。在选择了前10%的药物-靶标相互作用结构并对其进行排名后，使用L1000来预测潜在的候选药物。根据差异表达基因的映射，进一步汇总了药物-靶标相互作用分数，并获得了表1所示的排名最高的药物列表。

表1 顶级候选药物

药物名称	基因名称	靶向名称	分数
乌苏酸	MTHFD1L	Q6UB35	7.09
丙酸睾酮	PTPRS	G8JL96	6.87
盖尔德霉素	GPR155	Q7Z3F1	6.52
左炔诺孕酮	MTHFD1L	B7ZM99	6.52
雷地辛	MTHFD1L	Q6UB35	6.31
百里酚	PTPRS	G8JL96	6.29
塞鲁替尼	MTHFD1L	B7ZM99	6.16
提万替尼	KIF26B	B7WPD9	6.13
曲美替尼	PTPRS	G8JL96	6.08
沃诺他恰特	KIF26B	Q2KJY2	6.05

### 胃癌药物再利用

本研究确定了一些候选药物，包括乌苏酸、丙酸睾酮、盖尔德霉素、左炔诺孕酮、雷地辛、百里酚、塞鲁替尼、提万替尼、曲美替尼和沃诺他恰特。在这些药物中，乌苏酸有效抑制了体外胃癌细胞的生长，显著增加了体内外治疗的肿瘤细胞的凋亡率<sup>[5]</sup>。盖尔德霉素，抑制Hsp90功能，在人类胃癌细胞中诱导凋亡<sup>[6]</sup>。百里酚，在胃癌细胞中抑制细胞生长、增强凋亡，并使细胞对DPP治疗更加敏感<sup>[7]</sup>。塞鲁替尼，在具有MEK特征或RAS基因改变的胃癌患者中显示出有效的治疗效果和可容忍的安全性<sup>[8]</sup>。提万替尼，通过抑制c-MET或VEGFA扩增，从而抑制VEGF信号通路并在胃癌细胞中诱导凋亡<sup>[9]</sup>。曲美替尼，一种激酶抑制剂，通过靶向MAPK途径抑制胃肿瘤的发展和转移。<sup>[10]</sup>沃诺他恰特，在胃癌细胞系中触发细胞死亡和自噬，展现了作为胃癌

潜在治疗剂的可能性<sup>[11]</sup>。

## 结论

我们开发了一个基于转录组的胃腺癌计算药物再利用流程。我们从 GEO 数据库中选择了四个与癌症相关的基因表达档案，包括 258 个 RNA 测序数据集。通过在 GEOexplorer 上搜索相应的差异基因的 ID，我们获得了与这四个基因表达数据相关的 1004 个差

异基因的信息。随后，利用 Uniprot 数据库来识别相应的药物靶标。再从 DrugBank 下载了超过 10,000 种结构特征化的药物，利用深度神经网络，我们提高了药物 - 靶标相互作用的预测能力，并构建了多组学网络。我们进一步探索了排名最高的药物在癌症领域的临床试验证据。本研究展示了 RepoGC 在识别胃腺癌新药方面的潜力，并提供了一种有效的计算方法，以促进药物快速再利用。

## 参考文献：

- [1] Gao 等, 《A Pipeline to Call Multilevel Expression Changes between Cancer and Normal Tissues and Its Applications in Repurposing Drugs Effective for Gastric Cancer》.
- [2] Rudrapal, J. Khairnar 和 G. Jadhav, 《Drug Repurposing (DR)》.
- [3] Dudley, Deshpande 和 Butte, 《Exploiting Drug–Disease Relationships for Computational Drug Repositioning》.
- [4] Rogers 和 Hahn, 《Extended–Connectivity Fingerprints》.
- [5] Wang 等, 《Ursolic Acid Inhibits Proliferation and Induces Apoptosis of Cancer Cells in Vitro and in Vivo》.
- [6] Chen, Li 和 Pan, 《Geldanamycin Induces Apoptosis in Human Gastric Carcinomas by Affecting Multiple Oncogenic Kinases That Have Synergic Effects with TNF–Related Apoptosis–Inducing Ligand》.
- [7] Liu 等, 《Parthenolide Increases the Sensitivity of Gastric Cancer Cells to Chemotherapy》.
- [8] Lee 等, 《Selumetinib plus docetaxel as second-line chemotherapy in KRAS mutant, KRAS amplified or MEK signature gastric cancer patients: First arm of the umbrella trial in GC though the molecular screening, VIKTORY trial.》
- [9] Kim 等, 《Tivantinib Inhibits the VEGF Signaling Pathway and Induces Apoptosis in Gastric Cancer Cells with C-MET or VEGFA Amplification》.
- [10] Yamasaki 等, 《MEK Inhibition Suppresses Metastatic Progression of KRAS–Mutated Gastric Cancer》.
- [11] Claerhout 等, 《Gene Expression Signature Analysis Identifies Vorinostat as a Candidate Therapy for Gastric Cancer》.