

## 基于个性化网络标志物的药物推荐方法研究

刘文斌<sup>①②</sup> 吴倩<sup>①</sup> 杜玉改<sup>①</sup> 方刚<sup>②</sup> 石晓龙<sup>②</sup> 许鹏<sup>\*②③</sup>

<sup>①</sup>(温州大学计算机与人工智能学院 温州 325035)

<sup>②</sup>(广州大学计算科技研究院 广州 510006)

<sup>③</sup>(黔南民族师范学院计算机与信息学院 都匀 558000)

**摘要:** 基于个性化标志物的药物推荐研究, 有助于实现个性化用药及推动精准医疗的发展。该文利用基因表达谱数据及蛋白质网络信息, 基于基因2维高斯分布方法筛选出个性化网络标志物。进而综合考虑靶基因的重要性的和药物的副作用, 提出了一种计算药物对个性化标志物影响权重的方法。将该方法应用于肺腺癌、肾透明细胞癌和子宫内膜癌数据集, 通过启发式搜索方法, 得到每个疾病样本重要药物推荐列表。结果表明, 推荐的药物列表在同种癌症不同样本中既存在一致性, 也表现出很大的差异性, 如药物种类及药物排序差异, 这说明个性化药物在疾病治疗中的重要性及必要性。通过从药物数据库中搜索药物组合对疾病治疗的影响作用表明, 该文方法筛选得到的许多药物组合对具体疾病治疗具有积极影响, 这进一步证明该文基于个性化网络标志物的药物推荐方法的准确性。该文的研究将有效促进精准化医疗的发展。

**关键词:** 精准医疗; 个性化标志物; 网络标志物; 药物推荐

中图分类号: TP301

文献标识码: A

文章编号: 1009-5896(2020)06-1340-08

DOI: 10.11999/JEIT190837

## Drug Recommendation Based on Individual Specific Biomarkers

LIU Wenbin<sup>①②</sup> WU Qian<sup>①</sup> DU Yugai<sup>①</sup> FANG Gang<sup>②</sup>

SHI Xiaolong<sup>②</sup> XU Peng<sup>\*②③</sup>

<sup>①</sup>(College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou 325035, China)

<sup>②</sup>(Institute of Computing Science and Technology, Guangzhou University, Guangzhou 510006, China)

<sup>③</sup>(School of Computer Science of Information Technology, Qiannan Normal University for Nationalities, Duyun 558000, China)

**Abstract:** Drug recommendation research based on personalized markers can help to achieve personalized medicine and promote the development of precision medicine. In this paper, a method for calculating the weight of drugs on personalized markers is proposed, which first uses gene expression profile data and protein network information to filter out personalized network markers based on gene two-dimensional Gaussian distribution and then uses the importance degree of genes and the drugs side effect data to calculate the weight of drugs. This method is applied to lung adenocarcinoma, kidney renal clear cell carcinoma and uterine corpus endometrial carcinoma. Through the iterative process, a list of important drug recommendations for each disease sample is got. The results show that there are some differences in the recommended drug list and the ordering importance of drugs in different cases of the same kind of cancer, which indicates the importance and necessity of personalized drugs in the treatment of diseases. By querying the relationship between drugs from the drug database, many of the drug combinations screened by this method have a positive effect on the treatment of specific diseases, which further proves the accuracy of the drug recommendation methods based on personalized network markers. This study will effectively promote the development of precision medicine.

**Key words:** Precision medicine; Personalized biomarkers; Network biomarkers; Drug recommendation

收稿日期: 2019-10-29; 改回日期: 2020-01-20; 网络出版: 2020-02-27

\*通信作者: 许鹏 gdxupeng@gzhu.edu.cn

基金项目: 国家重点研发计划(2019YFA0706402), 国家自然科学基金(61572367, 61573017, 61972107, 61972109)

Foundation Items: The National Key R&D Program of China (2019YFA0706402), The National Natural Science Foundation of China (61572367, 61573017, 61972107, 61972109)

## 1 引言

癌症是一种具有高发病率和死亡率的恶性疾病，严重威胁人类的健康和生命<sup>[1,2]</sup>。随着科技的进步，癌症诊断和治疗的医疗技术越来越先进。但是对于许多癌症，依然没有非常有效的控制和治疗方法，需要更进一步地深入研究。精准医学的提出为癌症的诊断及治疗，提供了一种可能的解决思路。

识别疾病个性化标志物是实现精准医疗用药的首要解决问题。真正的个性化标志物应该是基于一系列个体正常状态、疾病产生发展及药物响应的完整表达谱数据获取到的，但目前无法有效获取这样的数据。于是，有研究学者提出基于多个正常样本与单个疾病样本进行个性化样本分析方法来识别疾病个性化标志物。其基本思想是正常群体的基因表达代表了一种“模式”，个体样本中偏离这种模式的分子就是个性化标志物。Wang等人<sup>[3]</sup>提出了基于稳定基因表达关系对的排名比较(RankComp)方法，来识别个体样本基因表达相对差异的个性化标志物。该方法简单直观，方便整合跨平台的表达数据。Liu等人<sup>[4]</sup>基于扰动理论，认为当加入一个样本后原有正常样本基因对的相关关系发生显著变化，则该基因对是疾病产生的原因，并据此提出了基于单样本的网络标志物方法(Single-Sample-specific Networks, SSN)。该方法发现了样本间差异关系的模式，可用于癌症分型和耐药性分析。文献<sup>[5]</sup>根据网络中基因对之间的变化，提出了基于单样本禽流感动态感染的预测方法，取得了良好的预测结果。之后，Yu等人<sup>[6]</sup>进一步将边构成的网络分析扩展到基于样本个体的边构成的网络分析，用于预测单个样本特异的癌症标志物。

由于仍有大量药物-靶点关系没有被发掘出来，近年来，研究学者提出了基于计算的方法来扩充已有的药物-靶点关系。Perlman等人<sup>[7]</sup>提出了基于药物-药物及基因-基因相似性的药物靶点预测方法，并应用逻辑回归对潜在的药物-靶点进行预测。基于“连坐”的思想，Wang等人<sup>[8]</sup>集成了已知的药物-靶点关系，及药物-药物和靶点-靶点相似网络，提高了预测新的药物-靶点关系性能。Zong等人<sup>[9]</sup>基于拓扑结构相似性，采用DeepWalk方法预测新的药物-靶点关系。

精准化用药的顺利实施涉及两个重要问题：如何识别疾病患者个性化标志物；如何基于个性化标志物给出适合的治疗药物及策略推荐。本文基于基因表达谱似然矩阵<sup>[10]</sup>以及蛋白质关系网络信息，筛选出癌症样本个性化网络标志物。进而结合已有的

药物-靶点信息，应用启发式搜索算法根据病人个性化特点推荐用药及治疗策略。本文分别在肺腺癌(LUng ADenocarcinoma, LUAD)、肾透明细胞癌(KIDney Renal Clear cell carcinoma, KIRC)和子宫内膜癌(Uterine Corpus Endometrial Carcinoma, UCEC)上展示该方法的推荐药物清单，并通过药物数据库中收录的药物组合对疾病治疗效果的信息，间接验证推荐药物的有效性。

## 2 研究方法

### 2.1 实验数据集及预处理

#### 2.1.1 TCGA样本数据

本文从癌症基因组图谱(The Cancer Genome Atlas, TCGA)数据库<sup>[11]</sup>下载了LUAD, KIRC和UCEC 3种癌症样本的基因表达数据。表1列出了3种癌症所包含的正常和癌症样本的数量。

表 1 3种癌症数据集统计信息

癌症类型	样本数量(正常/癌症)
LUAD	609(95/514)
KIRC	602(72/530)
UCEC	578(35/543)

为了消除部分基因在样本中表达缺失对结果的影响，首先对基因表达数据进行预处理，对于在超过1/2的样本中均未表达的基因进行删除；对于在少量样本中未表达的基因，用该基因其余样本表达量的均值进行填充。

#### 2.1.2 PPI网络数据

STRING<sup>[12]</sup>数据库是当前发展比较成熟的存储蛋白质相互作用(Protein-Protein Interaction, PPI)的数据库，该数据库包括经过实验验证的、从PubMed中挖掘的以及其它生物信息学方法预测的蛋白质互作信息。本文下载了10.5版本的PPI网络数据，为了保证所用数据的准确性，仅筛选出综合得分大于900的具有高置信度的功能关联网，过滤之后的PPI网络共包括10197个基因和190692条边<sup>[12]</sup>。

#### 2.1.3 药物-靶点数据

本文从DrugBank<sup>[13]</sup>和DGIdb<sup>[14]</sup>数据库整理出53810个药物-靶点关系，包含2325种药物。由于数据库中收录的药物-靶基因关系仍然比较稀疏，因此本文首先利用Wang等人<sup>[8]</sup>提出的异构图推理(Heterogeneous Graph Based Inference, HGBI)方法对已有的药物-靶基因关系进行扩展，从而获取比较完善的药物-靶点关系数据。在使用HGBI方法

过程中，需要用到3种数据：已知的药物-靶基因关系矩阵、基于DrugBank数据库获取的药物化学结构相似性计算得到的药物相似矩阵以及基于国家生物技术信息中心(National Center for Biotechnology Information, NCBI)获取到的基因对应蛋白质的氨基酸序列的相似性计算得到基因相似矩阵。取HGBI方法中的衰减系数 $\alpha$ 为0.4，通过不断迭代，药物-靶基因之间的关系达到稳定状态。在已有药物-靶基因关系上，为每个药物扩展10个新的权重最大的关系作为新的药物-靶点关系。最终，本文共获得55种LUAD相关药物，41种KIRC相关药物，31种UCEC相关药物。

2.2 方法

2.2.1 个性化网络标志物筛选方法

细胞癌变并非仅由单个基因发生变异所致，而是多个相关致癌基因共同作用的结果。假设PPI网络中的基因对在生物系统中服从2维高斯分布<sup>[15,16]</sup>，而在正常样本和癌症样本中分别服从两个不同的2维高斯分布。假设存在一对基因，表达量为 $x_1, x_2$ ，且某组样本中的均值为 $\mu_1, \mu_2$ ，标准差为 $\sigma_1, \sigma_2$ ，相关系数为 $\rho$ ，根据2维高斯分布，则该基因对的概率密度函数 $f(x_1, x_2)$ 为

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp \left[ -\frac{1}{2(1-\rho^2)} \left( \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right) \right] \quad (1)$$

同理，可以分别得到该基因对在正常样本和癌症样本中概率密度函数 $f^N(x_1, x_2)$ 和 $f^T(x_1, x_2)$ ，然后根据对数比计算每个基因对的表达量在两个高斯分布中的差异

$$\lambda(x_1, x_2) = \log_2(f^T(x_1, x_2)/f^N(x_1, x_2)) = \log_2(f^T(x_1, x_2)) - \log_2(f^N(x_1, x_2)) \quad (2)$$

而 $\lambda(x_1, x_2)$ 能反映出基因对更可能属于哪一个表型。 $\lambda(x_1, x_2) \geq 0$ 时，基因对的表达偏向于癌症状态， $\lambda(x_1, x_2) < 0$ 时，基因对的表达偏向于正常状态。

将基因表达数据和过滤的PPI网络数据相结合，提出基于基因2维高斯分布方法筛选出个性化网络标志物(2D Sample-Specific Networks, SSN-2D)的方法。流程如图1所示，首先，根据式(1)得到所有基因对在正常样本和癌症样本中的概率密度函数；

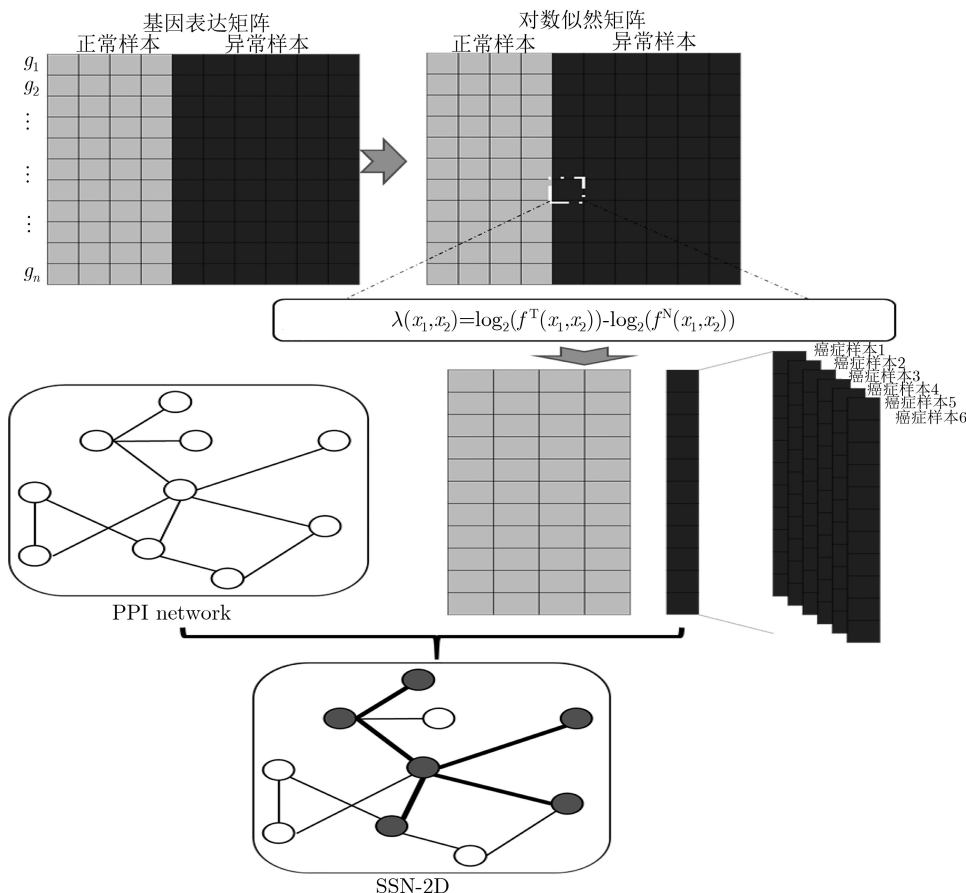


图1 癌症个性化网络标志物获取流程

然后, 根据式(2)构建基于基因对的对数似然矩阵; 接下来, 根据正常样本中各基因对值的分布情况, 对每个癌症样本的各基因对值进行假设检验, 保留显著性 $p < 0.05$ 的差异基因对作为癌症个性化标志物; 最后, 将上一步中得到的基因对投影到过滤后的PPI网络中, 得到癌症个性化标志物的PPI网络信息。

### 2.2.2 个性化药物推荐方法

由于每种癌症对应着多种治疗药物。在具体的临床用药过程中, 医生只能凭借经验制定用药策略。因此, 需要根据个性化差异特点, 使用计算的方法推荐不同的药物清单。一方面, 由于每种药物总能靶向多个基因, 而这些靶基因在PPI网络中所处位置不同, 具有不同的重要性, 且不同药物靶基因可能存在重叠。另一方面, 药物存在副作用。因此, 当选择较少的药物时, 本文倾向于选择副作用少但靶向重要基因多的药物。

给定一个疾病相关药物集合 $D = \{d_1, d_2, \dots, d_n\}$ 及个性化标志物集合 $T = \{t_1, t_2, \dots, t_m\}$ 本文的目标是寻找最佳药物组合, 具有靶向重要基因多且副作用小的特性。为了得到这样的药物组合, 定义一个药物权重函数 $S(d_i)$ , 如式(3)所示

$$S(d_i) = \frac{\sum_{j=1}^k d(t_j)}{\text{Num}(DS)} \quad (3)$$

其中,  $k$ 表示药物 $d_i$ 的靶基因数量,  $d(t_j)$ 表示药物第 $j$ 个靶点在蛋白质网络中的度,  $DS$ 表示药物的副作用,  $\text{Num}(DS)$ 表示药物副作用的个数。对于副作用信息缺失的药物, 本文将其副作用的数量定义为其他药物副作用个数的中值。

为了获得副作用少, 对个性化标志物具有多靶向性且趋向于靶向重要关键基因的最佳药物排序清单, 采用简单的启发式搜索方法。具体用一个迭代的过程实现, 伪代码如表2所示。

通过上述不断的迭代, 直至各个药物靶向的基因数量少于6时停止。此时所选中的药物组成的集合为最终的药物推荐列表。选择靶基因数量大于6的药物是因为当一个药物仅能靶向很少一部分相对不重要的基因时, 该药物在实际的应用中, 效果往往不理想。

## 3 实验结果及分析

### 3.1 药物副作用与药物靶点数量关系分析

基于药物数据库如DrugBank<sup>[13]</sup>和DGIdb<sup>[14]</sup>, 图2给出了3类癌症中, 药物靶基因数量与药物副作用数量之间的关系。总体上看, 药物靶基因越多,

表2 启发式搜索的迭代过程

个性化药物推荐算法
输入: 物集合 $D = \{d_1, d_2, \dots, d_n\}$ ; 个性化标志物集合 $T = \{t_1, t_2, \dots, t_m\}$ ;
输出: 个性化药物推荐列表 (Personalized Drug, PD);
(1) Initialization: Set $k = 1$ ;
(2) DO
(3) for $i = 1, 2, \dots, n$
(4) Compute $S(d_i)$ ;
(5) EndFor
(6) If $S(d_i)$ is the maximum among all drugs in $D$ then
(7) PD( $k$ ) = $d_i$ ;
(8) $k = k + 1$ ;
(9) EndIf
(10) Update $D$ : Delete $d_i$ from $D$ ;
(11) Update $T$ : Delete all targets of $d_i$ from $T$ ;
(12) WHILE Max(targets number of each drug in $D$ ) $\geq 6$

其副作用也越大。此外, 在靶基因数量相当时, 有些药物的副作用数量少, 有些药物的副作用数量就很大。从癌症治疗角度出发, 即希望药物靶点尽可能多, 又希望药物副作用尽可能少, 如何在二者之间达到平衡非常重要。

### 3.2 个性化药物推荐结果

本文结合药物副作用数量、药物靶基因数量以及靶基因重要程度, 通过药物权重计算, 选择最优的候选药物组合。为验证该方法的效果, 将考虑药物副作用信息所得到的药物集合, 与不考虑副作用信息所得到的药物集合进行比较。由于癌症样本数量非常可观, 利用以上两种方法将得出的药物罗列并逐一比较并不现实, 因此计算出每个药物在所有推荐列表中的平均排名, 通过平均排名, 比较两种方法的性能。图3为3种癌症中考虑副作用(灰色)和不考虑副作用(黑色)的情况下所得到的药物排名, 其中 $X$ 轴为药物,  $Y$ 轴表示各个药物的平均排名。

由图3可知, 考虑药物副作用信息, 得出副作用小的药物数量更多, 且这些药物的平均排名普遍更加靠前, 符合筛选副作用小但靶向更多更重要基因的药物的标准。为进一步分析两种方法的区别, 本文分别考察在两种方法中排名前5的药物。在LUAD中, 当考虑药物副作用的情况下, 排名前5的药物是crizotinib, carboplatin, metformin, ruxolitinib和vinblastine。而不考虑副作用的情况时, 排名前5的药物是bortezomib, paclitaxel, docetaxel, temozolomide和dasatinib。两种方法筛

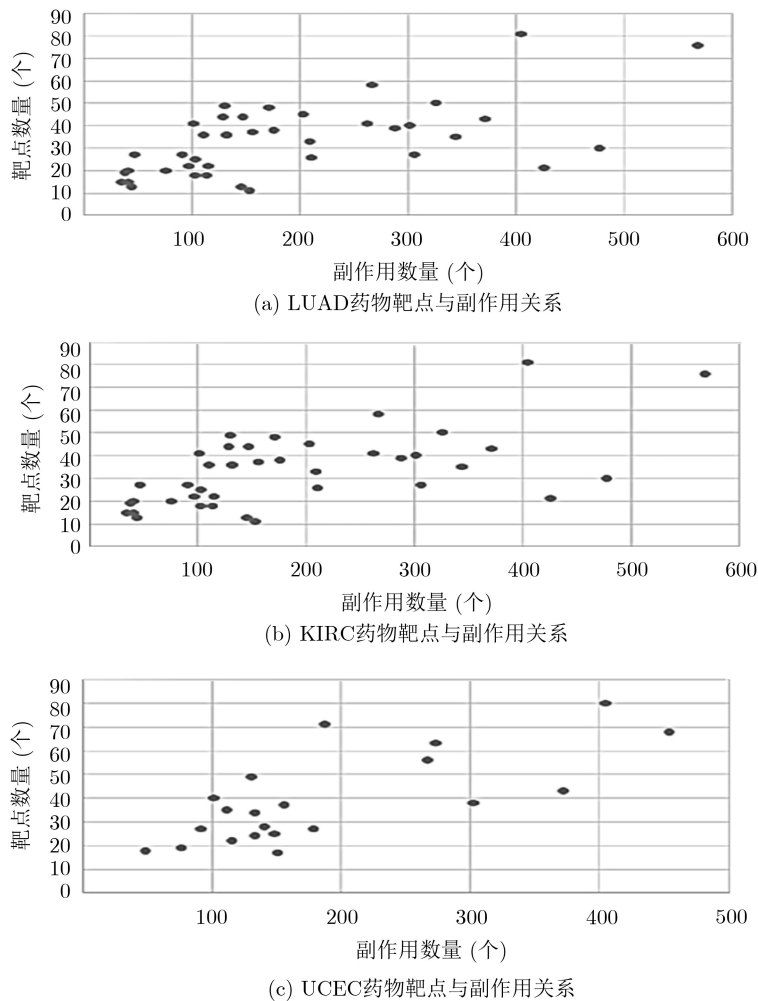


图2 3种癌症中药物靶基因数量与药物副作用数量之间的散点图

选得到的排名前5的药物并不一致, 本文发现, 在不考虑药物副作用的情况下, 筛选得到的药物副作用的数量都很大。例如, 使用bortezomib治疗疾病时, 患者会出现血小板减少, 中性粒细胞减少, 神经病和呼吸困难等症状<sup>[17-19]</sup>。使用paclitaxel<sup>[20,21]</sup>和docetaxel<sup>[22,23]</sup>也会出现相似的多种副作用的情况。同样, 在KIRC中, 当考虑药物副作用时, 排名前5的是vorinostat, erlotinib, carboplatin, etoposide和axitinib。而不考虑药物副作用时, 排名前5的是erlotinib, doxorubicin<sup>[24]</sup>, everolimus<sup>[25]</sup>, vincristine和etoposide。在UCEC中, 当考虑药物副作用的情况时, 排名前5的是vismodegib, afatinib, crizotinib, metformin和carboplatin。而不考虑药物副作用的情况时, 排名前5的是paclitaxel, everolimus, dasatinib, doxorubicin和dabrafenib。虽然这些药物都与癌症的发生有关系, 但在同时考虑药物靶基因信息和药物副作用数量的情况下, 本文能获得更多的副作用小的药物, 对临床应用具有更重要的意义。

在考虑药物副作用的情况下, 本文得到了候选药物对应的靶基因的重要程度, 分别计算各样本中候选药物靶基因的平均度和SSN-2D方法筛选出的样本特异的标志物的平均度, 在3类癌症中分别有85.77%, 74.15%和72.38%的样本, 候选药物的靶基因的平均度更大。由此可见, 利用本文所提的筛选候选药物的方法, 可以获取到靶向重要基因的药物。

为了更清楚地比较3类癌症得到的候选药物集合的分布, 本文对各类癌症中各个样本筛选得到的候选药物进行统计。结果如图4所示, 图4中行表示药物, 列表示癌症样本, 颜色的深浅表示某种药物在个性化样本推荐药物中排名, 颜色越深排名越靠前。由图4可知, 从样本角度出发, 在同种癌症中, 不同样本筛选得到的候选药物集合不尽相同, 这些候选药物的排序也有差异; 从药物角度出发, 有一些药物在大多数样本中都有, 说明这些药物在治疗这类癌症上比较重要, 而有一些药物仅出现在部分样本中, 这就说明个性化药物治疗在癌症治疗中起

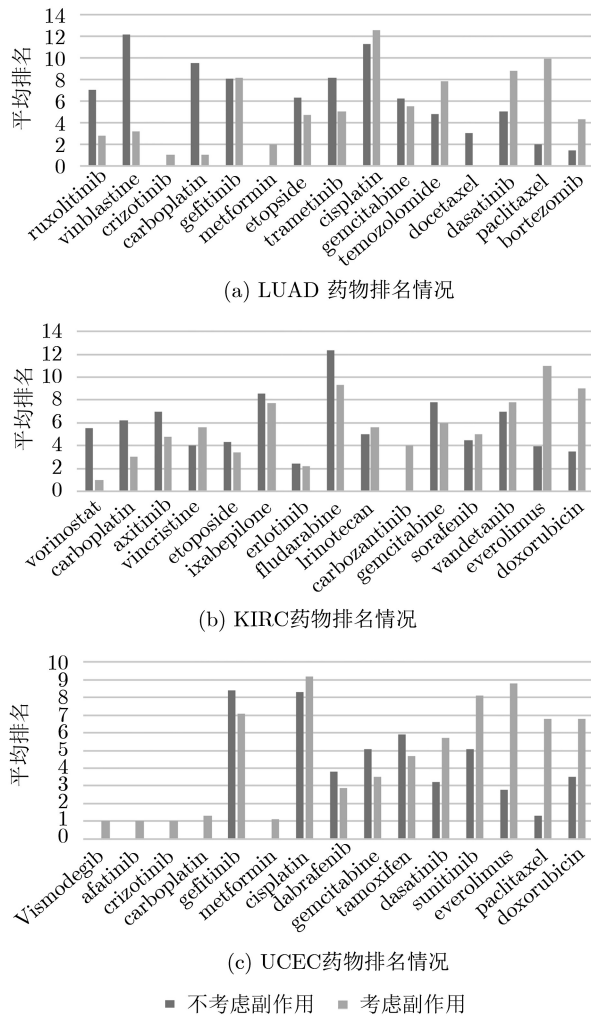


图 3 3种癌症中考虑药物副作用和不考虑药物副作用时药物的排名

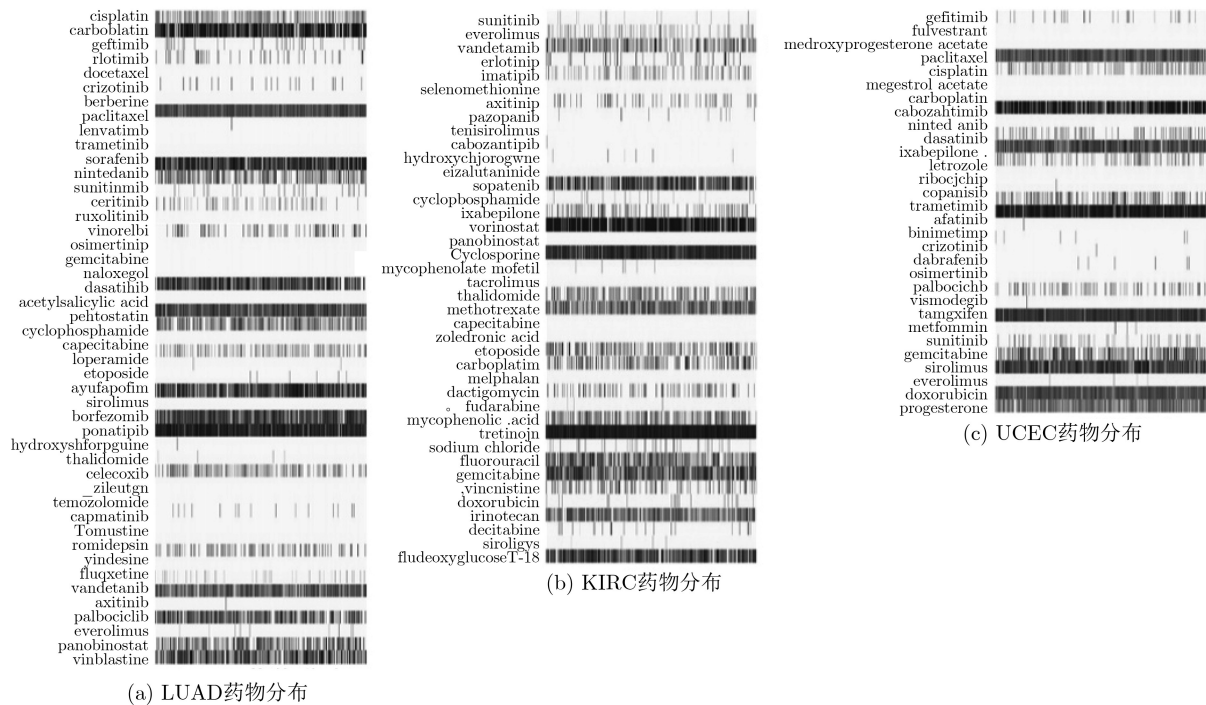


图 4 3类癌症得到的候选药物集合在各个样本中的具体分布

着至关重要的作用；从整体看，针对不同的癌症患者，需要制定不同的药物治疗方案，这也是个性化药物在癌症治疗中受到广泛关注的原因。

### 3.3 个性化组合药物分析

组合药物治疗能够促进药物疗效，因此组合药物已经被大范围采用。为了更深入地分析本文方法所得到的药物在疾病治疗中所起到的作用，本文从DrugBank数据库整理出这些药物之间的关系，根据药物之间的关系描述，将药物间的关系划分为两类，一类是将两种药物同时应用于疾病治疗时，具有促进疾病治疗的目的(协同作用)；一类是将两种药物同时应用于疾病治疗时，并不能促进疾病的治疗(拮抗作用)。

针对每一个癌症样本筛选得到的药物集合，分别分析这些药物组合在疾病治疗中的作用。图5展示了在3种癌症中，各个癌症样本中具有协同作用的药物关系分布。由图5可知，筛选出的药物集合中除KIRC得到的具有积极作用的药物关系比较少之外，LUAD和UCEC两种癌症样本获得的药物集合中，都有许多协同作用的药物关系，这也验证了通过药物副作用的数量，药物靶基因的数量以及靶基因的重要程度之间的关系筛选个性化药物是具有生物学意义的。由于基于个性化标记物得到的与KIRC相关的药物在DrugBank数据库中，这些药物之间协同作用的数量也较少，导致KIRC得到具有积极作用药物关系比较少。

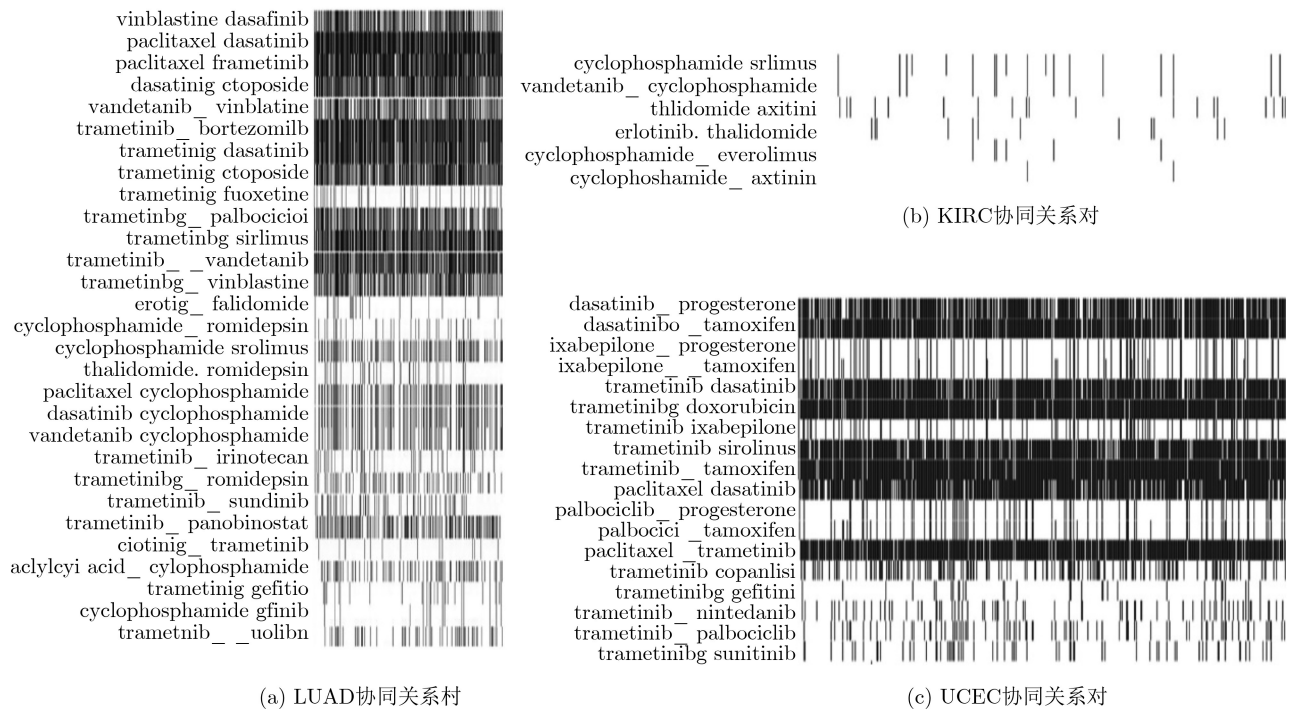


图 5 DrugBank数据库中具有协同作用的药物对在各个样本中的分布情况

## 4 结束语

随着高通量生物测序技术的发展,为人们认识疾病的发生、发展的分子机理提供了数据保障。精准化医疗是生命科学领域热点研究内容。如何根据病人的个性化特征设计相应的治疗策略,是实现精准化医疗需要解决的重要问题之一。本文利用基因表达谱数据及PPI网络信息,基于基因2维高斯分布方法筛选出样本个性化网络标志物。综合考虑基因的重要性和药物副作用,提出了一种计算药物对个性化标志物影响权重的方法。将本方法应用于3个癌症数据集,通过迭代过程,得到每个疾病样本重要药物推荐列表。结果表明,在同种癌症不同病例中,推荐的药物列表既有一致性也有差异性。即说明特定癌症在不同人群中存在着潜在的相同分子机理,可以采用一些广谱类抗癌药物;也说明了根据个性化标志物进行个性化用药的必要性。通过从药物数据库中查找药物组合对疾病治疗的联合作用效果,发现本文方法筛选得到的许多药物组合对具体疾病治疗具有积极的作用,进一步证明基于个性化网络标志物的药物推荐方法的准确性。

本文所提药物推荐方法,对于个性化医疗的发展具有一定的推动作用。同时,本研究还存在着一些不足。如所使用的数据集正常样本不足(UCEC正常样本仅有35例),以及药物-靶点信息的缺失等都会对本研究结果有一定的影响。随着数据样本的不断更新,本文将会致力于提出更加高效的个性化药物推荐方法。

## 参考文献

- [1] SIEGEL R L, MILLER K D, and JEMAL A. Cancer statistics, 2016[J]. *CA: A Cancer Journal for Clinicians*, 2016, 66(1): 7–30. doi: [10.3322/caac.21332](https://doi.org/10.3322/caac.21332).
- [2] TORRE L A, BRAY F, SIEGEL R L, *et al.* Global cancer statistics, 2012[J]. *CA: A Cancer Journal for Clinicians*, 2015, 65(2): 87–108. doi: [10.3322/caac.21262](https://doi.org/10.3322/caac.21262).
- [3] WANG Hongwei, SUN Qiang, ZHAO Wenyuan, *et al.* Individual-level analysis of differential expression of genes and pathways for personalized medicine[J]. *Bioinformatics*, 2015, 31(1): 62–68. doi: [10.1093/bioinformatics/btu522](https://doi.org/10.1093/bioinformatics/btu522).
- [4] LIU Xiaoping, WANG Yuetong, JI Hongbin, *et al.* Personalized characterization of diseases using sample-specific networks[J]. *Nucleic Acids Research*, 2016, 44(22): e164. doi: [10.1093/nar/gkw772](https://doi.org/10.1093/nar/gkw772).
- [5] ZHANG Wanwei, ZENG Tao, and CHEN Luonan. EdgeMarker: Identifying differentially correlated molecule pairs as edge-biomarkers[J]. *Journal of Theoretical Biology*, 2014, 362: 35–43. doi: [10.1016/j.jtbi.2014.05.041](https://doi.org/10.1016/j.jtbi.2014.05.041).
- [6] YU Xiangtian, ZHANG Jingsong, SUN Shaoyan, *et al.* Individual-specific edge-network analysis for disease prediction[J]. *Nucleic Acids Research*, 2017, 45(20): e170. doi: [10.1093/nar/gkx787](https://doi.org/10.1093/nar/gkx787).
- [7] PERLMAN L, GOTTLIEB A, ATIAS N, *et al.* Combining drug and gene similarity measures for drug-target elucidation[J]. *Journal of Computational Biology*, 2011, 18(2): 133–145. doi: [10.1089/cmb.2010.0213](https://doi.org/10.1089/cmb.2010.0213).
- [8] WANG Wenhui, YANG Sen, and LI Jing. Drug Target

- Predictions Based on Heterogeneous Graph Inference[M]. ALTMAN R B, DUNKER A K, HUNTER L, *et al.* Biocomputing 2013. Hawaii, USA: World Scientific, 2013: 53-64 doi: [10.1142/9789814447973\\_0006](https://doi.org/10.1142/9789814447973_0006).
- [9] ZONG Nansu, KIM H, NGO V, *et al.* Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations[J]. *Bioinformatics*, 2017, 33(15): 2337-2344. doi: [10.1093/bioinformatics/btx160](https://doi.org/10.1093/bioinformatics/btx160).
- [10] SU Junjie, YOON B J, and DOUGHERTY E R. Accurate and reliable cancer classification based on probabilistic inference of pathway activity[J]. *PLoS One*, 2009, 4(12): e8161. doi: [10.1371/journal.pone.0008161](https://doi.org/10.1371/journal.pone.0008161).
- [11] TOMCZAK K, CZERWIŃSKA P, and WIZNEROWICZ M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge[J]. *Contemporary Oncology (Poznan, Poland)*, 2015, 19(1A): A68-A77. doi: [10.5114/wo.2014.47136](https://doi.org/10.5114/wo.2014.47136).
- [12] SZKLARCZYK D, MORRIS J H, COOK H, *et al.* The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible[J]. *Nucleic Acids Research*, 2017, 45(D1): D362-D368. doi: [10.1093/nar/gkw937](https://doi.org/10.1093/nar/gkw937).
- [13] 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 Research*, 2018, 46(D1): D1074-D1082. doi: [10.1093/nar/gkx1037](https://doi.org/10.1093/nar/gkx1037).
- [14] GRIFFITH M, GRIFFITH O L, COFFMAN A C, *et al.* DGIdb: Mining the druggable genome[J]. *Nature Methods*, 2013, 10(12): 1209-1210. doi: [10.1038/nmeth.2689](https://doi.org/10.1038/nmeth.2689).
- [15] 李玉博, 陈邈. 几乎完备高斯整数序列构造法[J]. 电子与信息学报, 2018, 40(7): 1752-1758. doi: [10.11999/JEIT170844](https://doi.org/10.11999/JEIT170844).  
LI Yubo and CHEN Miao. Construction of nearly perfect gaussian integer sequences[J]. *Journal of Electronics & Information Technology*, 2018, 40(7): 1752-1758. doi: [10.11999/JEIT170844](https://doi.org/10.11999/JEIT170844).
- [16] 陈曦, 张坤. 一种基于树增强朴素贝叶斯的分类器学习方法[J]. 电子与信息学报, 2019, 41(8): 2001-2008. doi: [10.11999/JEIT180886](https://doi.org/10.11999/JEIT180886).  
CHEN Xi and ZHANG Kun. A classifier learning method based on tree-augmented naïve bayes[J]. *Journal of Electronics & Information Technology*, 2019, 41(8): 2001-2008. doi: [10.11999/JEIT180886](https://doi.org/10.11999/JEIT180886).
- [17] FANUCCHI M P, FOSSELLA F V, BELT R, *et al.* Randomized phase ii study of bortezomib alone and bortezomib in combination with docetaxel in previously treated advanced non-small-cell lung cancer[J]. *Journal of Clinical Oncology*, 2006, 24(31): 5025-5033. doi: [10.1200/JCO.2006.06.1853](https://doi.org/10.1200/JCO.2006.06.1853).
- [18] DAVIES A M, RUEL C, LARA P N, *et al.* The proteasome inhibitor bortezomib in combination with gemcitabine and carboplatin in advanced non-small cell lung cancer: A california cancer consortium phase I study[J]. *Journal of Thoracic Oncology*, 2008, 3(1): 68-74. doi: [10.1097/JTO.0b013e31815e8b88](https://doi.org/10.1097/JTO.0b013e31815e8b88).
- [19] DAVIES A M, CHANSKY K, LARA Jr P N, *et al.* Bortezomib plus gemcitabine/carboplatin as first-line treatment of advanced non-small cell lung cancer: A phase II southwest oncology group study (S0339)[J]. *Journal of Thoracic Oncology*, 2009, 4(1): 87-92. doi: [10.1097/JTO.0b013e3181915052](https://doi.org/10.1097/JTO.0b013e3181915052).
- [20] KELLY K, CROWLEY J, BUNN PA, *et al.* Randomized phase III trial of paclitaxel plus carboplatin versus vinorelbine plus cisplatin in the treatment of patients with advanced non-small-cell lung cancer: A southwest oncology group trial[J]. *Journal of Clinical Oncology*, 2001, 19(13): 3210-3218. doi: [10.1200/JCO.2001.19.13.3210](https://doi.org/10.1200/JCO.2001.19.13.3210).
- [21] SANDLER A, GRAY R, PERRY M C, *et al.* Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer[J]. *New England Journal of Medicine*, 2006, 355(24): 2542-2550. doi: [10.1056/NEJMoa061884](https://doi.org/10.1056/NEJMoa061884).
- [22] MCNAMARA M, SWEENEY C, ANTONARAKIS E S, *et al.* The evolving landscape of metastatic hormone-sensitive prostate cancer: A critical review of the evidence for adding docetaxel or abiraterone to androgen deprivation[J]. *Prostate Cancer and Prostatic Diseases*, 2018, 21(3): 306-318. doi: [10.1038/s41391-017-0014-9](https://doi.org/10.1038/s41391-017-0014-9).
- [23] ROBINET G, BARLESI F, FOURNEL P, *et al.* Second-line therapy with gefitinib in combination with docetaxel for advanced non-small cell lung cancer: A phase II randomized study[J]. *Targeted Oncology*, 2007, 2(2): 63-71. doi: [10.1007/s11523-007-0042-9](https://doi.org/10.1007/s11523-007-0042-9).
- [24] ANGELOPOULOU A, KOLOKITHAS-NTOUKAS A, FYTAS C, *et al.* Folic acid-functionalized, condensed magnetic nanoparticles for targeted delivery of doxorubicin to tumor cancer cells overexpressing the folate receptor[J]. *ACS Omega*, 2019, 4(26): 22214-22227. doi: [10.1021/acsomega.9b03594](https://doi.org/10.1021/acsomega.9b03594).
- [25] TEDESCO-SILVA H, PASCUAL J, VIKLICKY O, *et al.* Safety of everolimus with reduced calcineurin inhibitor exposure in de novo kidney transplants: An analysis from the randomized TRANSFORM study[J]. *Transplantation*, 2019, 103(9): 1953-1963. doi: [10.1097/TP.0000000000002626](https://doi.org/10.1097/TP.0000000000002626).
- 刘文斌: 男, 1969年生, 教授, 研究方向为生物信息学.  
吴倩: 女, 1994年生, 硕士, 研究方向为生物信息学.  
杜玉改: 女, 1993年生, 硕士, 研究方向为生物信息学.  
方刚: 男, 1969年生, 教授, 研究方向为生物信息学.  
石晓龙: 男, 1975年生, 教授, 研究方向为生物信息学.  
许鹏: 男, 1986年生, 博士后, 研究方向为生物信息学.