计算机科学
COMPUTER SCIENCE

基于基因优先级排序的活跃模块识别方法

张琦，潘可，朱凯

引用本文

张琦，潘可，朱凯. 基于基因优先级排序的活跃模块识别方法[J]. 计算机科学，2023，50(11A): 221200113-8.

ZHANG Qi, PAN Ke, ZHU Kai. Method for Identifying Active Module Based on Gene Prioritization[J]. Computer Science, 2023, 50(11A): 221200113-8.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.221200113

基于基因优先级排序的活跃模块识别方法

张 琦[1] 潘 可[2] 朱 凯[3]
1 沂蒙干部学院 山东 临沂 276000
2 南宁学院人工智能与软件学院 南宁 530000
3 武汉大学计算机学院 武汉 430000

摘 要 随着高通量测序技术的快速发展,海量的多组学数据有助于从分子水平上研究癌症的致病机理。近年来,活跃模块的识别问题已成为生物信息学领域的主要研究方向。然而,许多现有方法往往无法识别与癌症密切相关的强连通模块。通过整合蛋白质-蛋白质相互作用网(PPI)和基因表达两种组学数据,提出一种活跃模块识别方法 IdeMod。具体而言,IdeMod 提出了基于 p 步随机游走核回归模型的基因活跃评分函数,结合帕累托最优共识(POC)方法中的支配关系得到基因优先级排序列表。通过引入一种模拟退火算法 SA-PROX,寻找 PPI 网络中优先级高且连通性强的活跃模块。在真实的乳腺癌数据和宫颈癌数据下进行实验,与 SigMod,LEAN,RegMod 和 ModFinder 方法相比,IdeMod 可以识别一个连通性强且包含大量癌症相关基因的活跃模块。因此,IdeMod 方法可作为检测活跃模块的一种有效补充工具。

关键词:癌症;活跃模块;基因优先级排序;PPI 网络;基因表达;模拟退火算法
中图法分类号 TP301

Method for Identifying Active Module Based on Gene Prioritization

ZHANG Qi[1],PAN Ke[2] and ZHU Kai[3]
1 Yimeng Executive Leadership Academy,Linyi,Shandong 276000,China
2 College of Artificial Intelligence and Software,Nanning College,Nanning 530000,China
3 School of Computer Science,Wuhan University,Wuhan 430000,China

Abstract With the rapid development of high-throughput sequencing,a vast amount of multi-omics data has been contributed to investigating the pathogenesis of cancer at the molecular level. In recent years,the identification of active modules has become a major direction in bioinformatics. However,many existing approaches cannot identify a dense module that has strong association with cancer. A method called IdeMod is proposed by integrating protein-protein interaction network(PPI) and gene expression data. More concretely,a gene scoring function is devised by using the regression model with a p-step random walk kernel. By introducing the relationship of dominance in the POC method,a gene prioritization list is presented. A simulated annealing algorithm SA-PROX is introduced to find an active module with high gene prioritization and strong connectivity. Experiments are performed on real biological datasets,including breast cancer and cervical cancer. Compared with the previous methods SigMod,LEAN,RegMod and ModFinder,IdeMod can successfully identify a well-connected module that contains a large proportion of cancer-related genes. Therefore,the proposed approach may become a useful complementary tool for identifying active module.

Keywords Cancer,Active module,Gene prioritization,Protein-protein interaction network,Gene expression,Simulated annealing algorithm

1 引言

研究表明,癌症作为一种复杂的遗传疾病,其产生与发展是由模块\通路调控的[1-3]。近年来,随着高通量速测序技术的快速发展,癌症基因组图谱计划(TCGA)[4]、国际肿瘤基因组协作组(ICGC)[5]等大型癌症基因组测序项目收集了海量的癌症多组学数据,为实现精准医疗提供了新的思路。基因表达是基因经过转录、翻译,产生有生物活性的蛋白质的过程。基因表达谱可以检测在不同实验条件下特定的细胞或组织中基因的转录活动[6]。尽管差异表达分析可以识别与癌症相关的特定基因,但并没有考虑控制细胞功能的分子相互作用网络,从而无法从网络数据中获得重要信息[7]。蛋白质-蛋白质相互作用网络是单独蛋白彼此相互作用构成的,参与了新陈代谢、基因表达调节、细胞周期调控、细胞的增殖和凋亡等生命过程的各个环节。蛋白质-蛋白质相互作用网络为基因表达谱的研究补充了基因间的交互信息,具有与细胞过程或疾病状态相关的特异性信息[7]。越来越多的研究表明,活跃模块与癌症的产生和发展有关[8-11]。活跃模块/子网是一组相互关联且参与重要生物学功能的基因[6]。因此,识别活跃模块有助于了解癌症的致病机制,为临床抗癌药物研究提供重要依据。

活跃模块识别问题是一个 NP-难问题。目前大多数研究

通信作者:张琦(zhangqi5830@shandong.cn)

方法主要包括两个步骤：第一步，通过基因评分进行基因优先级排序；第二步，利用搜索算法提取生物网络中的活跃模块。2010 年，Qiu 等[8]提出了基于热扩散核回归模型的活跃模块识别方法 RegMod，用于提取 PPI 网络中上调的活跃模块和下调的活跃模块，但提取的模块在连通性上表现较差。2016 年，Gwinner 等[9]提出了 LEAN 方法，采用局部富集分析的方式识别网络中与疾病有关的模块。LEAN 方法的局限性在于仅考虑基因的直接邻居基因的影响。2017 年，Liu 等[10]提出了 SigMod 方法，用最小切割算法识别加权网络中的强连通模块。2018 年，Vaic 等[11]提出了 ModuleDiscover 方法来检测 PPI 网络中的活跃模块，该方法基于最大团枚举问题，通过从网络中的随机种子节点扩展生成的团的迭代枚举来近似表示 PPI 网络的群落，进而将显著差异表达的团确定为生物网络的活跃模块。然而，由于依赖于 STRING 数据库中的先验知识，所以该方法基本上不可能发现新的活跃模块。2020 年，Tian 等[12]提出了一种进化的多目标优化方法 EMODMI，该方法为每个疾病样本构建了一个特定的样本网络，然后利用多目标遗传算法从该网络中提取疾病模块。2022 年，Wu 等[13]提出了 ModFinder 方法，该方法通过确定基因的活跃评分和度的加权相对差距和，确定基因优先级排序，通过贪心算法 NSEA 识别活跃模块，但提出的基因优先级排序方法需要通过大量预实验测试选取实验参数。

本文提出了一种活跃模块识别方法 IdeMod。首先，基于 p 步随机游走核回归模型设计了计算基因活跃评分的评分函数。其次，结合帕累托最优共识(Pareto Optimality Consensus，POC)[14]确定每两个基因之间的支配关系，得到基因优先级排序。最后，引入模拟退火算法 SA-PROX 来识别一个优先级高且连通性强的活跃模块。实验结果表明，IdeMod 可识别出覆盖大量癌症基因的活跃模块，可能成为识别活跃模块的有力补充工具，为癌症研究工作发挥一定的辅助作用。

2 方法

假定存在一个 $n\times m$ 的基因表达矩阵 E，矩阵的行表示基因集 $G=\{g_1,g_2,\cdots,g_n\}$，其中 $g_i(1\leqslant i\leqslant n)$ 表示一个基因，n 为基因数；矩阵的列表示样本集 $S=\{s_1,s_2,\cdots,s_m\}$，其中 $s_i(1\leqslant j\leqslant m)$ 表示一个样本，m 为样本数。样本集 S 包含正常样本和癌症样本，令前 m_1 列为正常样本，后 m_2 列为癌症样本，且 $m_1+m_2=m$，矩阵中的元素 $e_{ij}(i=1,2,\cdots,n;j=1,2,\cdots,m)$ 表示基因 i 在样本 j 当中的表达值，通常情况下，$m\ll n$。

令无向图 $P=(V,A)$ 表示连通且加权的蛋白质-蛋白质相互作用网络(PPI)，其中节点 $v_i\in V$ 表示由基因 g_i 经过转录、翻译产生的蛋白质，无向边 $(v_i,v_k)\in A$ 表示基因 g_i 和基因 $g_k(i\neq k)$ 产生的蛋白质之间的相互作用。因此，基因 g_i 可以表示无向图 P 中的顶点 v_i。

2.1 构造加权 PPI 网络

给定 PPI 网络 $P=(V,A)$，根据基因表达矩阵 E 中的基因表达值和 PPI 网络中顶点的拓扑特性，分别定义了函数 $\dot{W}(\cdot)$ 和 $\ddot{W}(\cdot)$ 对 PPI 网络中的顶点(基因)和边(相互作用)进行赋权。由于共表达的基因之间更有可能具有相同的生物学功能，因此本文利用两个基因之间的共表达程度给 PPI 网络中边的赋权值。本文中，基因 g_i 和 g_k 之间边的权重 $\ddot{w}_{ik}$ 用基因表达矩阵 E 中行向量 e_{i-} 和 e_{k-} 的皮尔逊相关系数[15]表示。令 $\dot{w}_i=\dot{W}(v_i)$ 表示基因 g_i 对应顶点 $v_i\in V$ 的权重，其赋值过程包括计算观测活跃评分和预测潜在活跃评分两个步骤。因此，得到顶点加权和边加权的 PPI 网络 $P=(V,A,\dot{W},\ddot{W})$，其构造流程如图 1 所示。

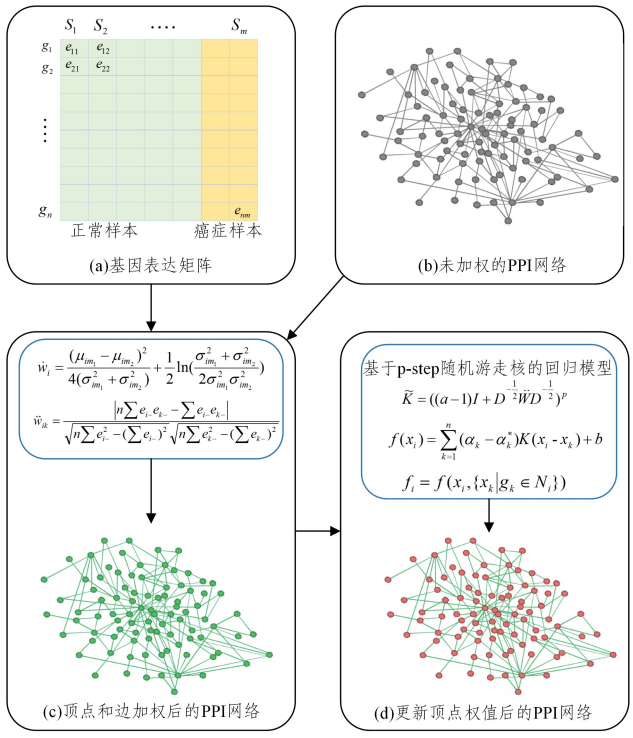

图 1 构造加权 PPI 网络流程图

Fig. 1 Flowchart of a weighted PPI network construction

2.1.1 计算观测活跃评分

给定网络 P 中的顶点 $v_i\in V$，其观测活跃评分利用贝叶斯距离公式[16]计算，以衡量基因 g_i 在癌症样本和正常样本中差异表达的显著性，如式(1)所示：

$$\dot{w}_i=\frac{(\mu_{im_1}-\mu_{im_2})^2}{4(\sigma_{im_1}^2+\sigma_{im_2}^2)}+\frac{1}{2}\ln\left(\frac{\sigma_{im_1}^2+\sigma_{im_2}^2}{2\sigma_{im_1}^2\sigma_{im_2}^2}\right) \tag{1}$$

其中，μ_{im_1}，μ_{im_2} 和 σ_{im_1}，σ_{im_2} 分别表示基因 g_i 在两类样本中表达量的均值和标准差。若基因表达量在癌症样本和正常样本中的分布均值和方差无明显差别，该基因与样本类别无关，将不会提供有用的生物学信息。然而，若基因表达量在两类样本中的分布均值和方差出现较大差异，那么从生物学的角度分析，该基因很可能与癌症紧密相关。

2.1.2 预测潜在活跃评分

由于高通量数据自身的局限性，在获得基因表达谱和数据转化过程中，不可避免地会引入系统噪声，由此计算出的基因观测活跃评分可能会存在许多异常值，并不能很好的反映相互作用的基因具有相似的活跃评分这一事实[7]。因此，利用基因观测活跃评分和基因在 PPI 网络中的拓扑相似性预测基因的潜在活跃评分，该评分更接近基因的真实活跃水平。

令 N_i 表示基因 g_i 及其邻居基因(直接/间接相连)的集合。基因 g_i 的潜在活跃评分 f_i 与集合 N_i 中基因的关系用潜在活跃评分函数 f 表示：

$$f_i=f(x_i,\{x_k\mid g_k\in N_i\}) \tag{2}$$

其中，x_i 表示每个基因对应的属性向量，记录基因的观测活跃评分。显然，定义的函数 f 中基因 g_i 的潜在活跃评分与其邻居基因有关，可以反映基因与其邻居节点相互作用模式。

基于上述定义，将活跃评分函数 f 拟合到基因的观测活跃评分上，构成了一个非线性回归问题。这个过程可以看作是一个平滑过程，消除网络中相邻基因之间活跃评分的剧烈变化，来预测基因的潜在活跃评分。本文中使用支持向量回归(SVR)模型[17]解决非线性回归问题。

给定训练数据集 $\{(x_i,\dot{w}_i)\mid g_i\in V\}$，将训练集数据 x_i 经过非线性变换 $\Phi(x)$ 映射到高维空间 F 中，将非线性回归问题转化为高维空间中的线性回归问题。回归决策函数如下所示：

$$f(x_i)=\sum_{k=1}^{n}(\alpha_k-\alpha_k^*)K(x_i,x_k)+b \tag{3}$$

其中，b 表示位移量；α_k 和 α_k^* 表示拉格朗日乘子；x_i 和 $\dot{w}_i$ 分别表示基因 g_i 在原始空间中的属性向量和目标值；$K(x_i,x_k)$ 是一个对称非负定函数，称为核函数。为捕获基因之间的相似性，同时考虑到整体的拓扑结构交互网络，SVR 模型的核函数取 p 步随机游走核[18]。

给定无向图 $P=(V,A)$ 的对称邻接矩阵 $\ddot{W}$，p 步随机游走核为：

$$\widetilde{K}=((a-1)I+D^{-\frac{1}{2}}\ddot{W}D^{-\frac{1}{2}})^p \tag{4}$$

其中，$\widetilde{K}$ 为核函数对应的格拉姆矩阵，其元素 $\widetilde{k}_{ij}$ 对应于核函数的值 $K(x_i,x_k)$；a 为大于 1 的常数，I 为单位矩阵，D 是对角元素为 $d_{ii}=\sum_{g_k\in N_i}\ddot{w}_{ik}$ 的对角矩阵；p 为步数，其大小决定集合 N_i 中基因 g_i 的邻居基因数目。换言之，通过设置 $p=2$，如果两个基因直接相连，则认为两个基因相似；通过设置 $p=3$，如果两个基因通过包含一个或两个顶点的路径连接，则认为两个基因相似。原则上也可以考虑更大的步数 p，但这可能会在基因之间引入过于遥远的相似性，从而在预测基因潜在活跃评分时产生大量的潜在噪音。在本文中，使用 LIBSVM[19]来解决 SVR 问题。

2.2　基因优先级排序方法

本节基于构造的加权 PPI 网络 P，提出了基因优先级排序方法。网络中顶点的度(degree)是指与顶点直接相连的边数。顶点的度是生物网络中顶点最基本的特征。度较高的顶点被称为“中心蛋白质(hubs)”，是生物分子网络中的关键参与者[20]。本文利用基因潜在活跃评分和在 PPI 网络中的度，结合帕累托最优共识(Pareto Optimality Consensus，POC)方法确定 PPI 网络中两两基因之间的支配关系，得到基因优先级排序列表。

POC 方法的基本原理描述如下：集合中的每个元素依据某种规则赋予两类得分 A 和 B，然后使用帕累托最优(Pareto Optimality)来达到两类得分之间的一致性，并确定集合中两两元素之间的支配关系[14]。在基因优先级排序方法中，若满足以下条件，则认为基因 g_i 支配基因 g_k($g_i\prec g_k$)：

$$f_i\geqslant f_k \text{ and } d_{ii}\geqslant d_{kk} \tag{5}$$

在基因集 $G=\{g_1,g_2,\cdots,g_n\}$ 中，若基因 g_i 支配基因 g_k($g_i\prec g_k$)，则基因 g_i 可以获得比基因 g_k 更高的排序。若基因 g_i 可以支配 $n-1$ 个基因，则基因 g_i 将获得最高排序；若基因 g_i 可以支配 $n-2$ 个基因，则基因 g_i 将处于排序的第二位，以此类推。基于上述每对基因之间的支配关系，根据如下定义：在基因集 G 中，若 $g_i\prec g_k$，则有 $c(g_i\prec g_k)=1$。基于 POC 方法的基因优先级排序函数 $pr(i)$ 的定义如下：

$$pr(i)=\sum_{g_k\in V}c(g_i\prec g_k) \tag{6}$$

基因 g_i 的 $pr(i)$ 值越高，表示该基因支配的其他基因就越多，基因优先级排序也就越高。

2.3　活跃模块识别方法

2.3.1　问题模型

假设有 $\hat{V}\subset V$ 表示一组选择的基因，即 PPI 网络中的一个模块。令 $pr(\hat{V})$ 表示模块 $\hat{V}$ 的优先级，定义为模块中所有基因的优先级的平均值，如下所示：

$$pr(\hat{V})=\frac{1}{|\hat{V}|}\sum_{g_i\in\hat{V}}pr(i) \tag{7}$$

根据以上定义，活跃模块识别问题模型可以描述为：给定加权 PPI 网络 $P=(V,A,\dot{W},\ddot{W})$、基因 $g_i(g_i\in V)$ 的优先级 $pr(i)$ 和基因潜在活跃评分 f_i、活跃模块规模大小 ms，试图识别一个模块规模为 ms 的活跃模块 $\hat{V}$，使得模块的优先级最大。

2.3.2　亲密度概念

为量化基因在 PPI 网络中的紧密程度，提出了基因与基因之间的亲密度概念。如式(8)所示，顶点 v_i 和顶点 v_k 的亲密度可以用函数 $prox(v_i,v_k)$ 表示：

$$prox(v_i,v_k)=\frac{1}{D_{ij}(v_i,v_k)} \tag{8}$$

其中，$D_{ij}(v_i,v_j)$ 表示顶点 v_i 和顶点 v_j 在 PPI 网络中的最短路径。特别地，若两个顶点不连通，则两个顶点的最短路径可以记为无限大，即亲密度近似于 0。类似地，顶点 v_k 与模块 $\hat{V}$($\hat{V}\subset V$)的亲密度可以用顶点 v_k 与模块 $\hat{V}\subset V$ 中所有顶点的亲密度之和表示，如式(9)所示：

$$prox(\hat{V},v_k)=\sum_{v_i\in\hat{V}}prox(v_i,v_k) \tag{9}$$

2.3.3　SA-PROX 算法

本节基于改进的基因优先级排序方法和问题模型，提出了求解活跃模块识别问题模型的模拟退火算法 SA-PROX。SA-PROX 算法的核心思想是通过寻找一个更优的种子顶点，不断扩充优质顶点，进而形成一个更优的活跃模块。算法步骤如算法 1 和算法 2 所示。

算法 1　SA-PROX 算法

输入：加权 PPI 网络 P，基因优先级排序列表 pr，基因活跃评分列表 f，参数 ms，迭代次数 L，起始温度 T_0，终止温度 T_{end}，降温系数 t

输出：规模大小为 ms 的活跃模块

```
1. 令当前温度 T 为 T0，随机生成种子顶点 v，v∈V
2. while T>Tend do
3.     for(l=0;l<L;l++) do
4.         sub_extend(P,f,ms,V̂)
5.         if(l=0) then
6.             best=V̂
7.         else if(pr(V̂)>pr(best)) then
8.             best=V̂
9.             else if(random(0,1)<exp(pr(best)−pr(V̂)/T) then
```

10. best=$\hat{V}$
11. end if
12. if l≤0.8L then
13. C={v_i | $v_i \in \hat{V}$, (f_i > $med_f(\hat{V})$) ∨ (d_i > $med_d(\hat{V})$)}
14. else
15. C={v_i | $v_i \in \hat{V}$, (f_i > $med_f(\hat{V})$) ∧ (d_i > $med_d(\hat{V})$)}
16. endif
17. endfor
18. T=tT
19. endwhile
20. $\hat{V}$=best
21. 输出模块 $\hat{V}$

算法 2 sub_extend

输入：加权 PPI 网络 P，基因优先级排序列表 pr，基因活跃评分列表 f，参数 ms

输出：规模大小为 ms 的活跃模块

1. repeat
2. U←{v_j | $v_i \in \hat{V}$, (v_i, v_j) ∈ A}
3. for(each $v_j \in U$) do
4. if({v_i | $v_i \in \hat{V}$, prox(v_j, v_i) < 1/3}) then
5. U=U−{v_j}
6. end if
7. end for
8. v^* = argmax(f_j + prox($\hat{V}$, v_j)), $v_j \in U$
9. $\hat{V}$ = $\hat{V}$ ∪ {v^*}
10. until U=Ø or $|\hat{V}|$ > ms

3 实验结果与分析

本节利用真实的乳腺癌数据和宫颈癌数据进行测试，比较 RegMod，LEAN，SigMod，ModFinder 和 IdeMod 这 5 种方法的识别性能。所有实验都是在一台 Huawei 工作站（AMD Ryzen 5 4600H with Radeon Graphics 3.00GHz，内存 16GB）上进行的，操作系统是 Windows 8；RegMod，ModFinder 和 IdeMod 方法的编译运行工具为 MATLAB 2018b，SigMod、LEAN 方法的编译运行工具为 R3.6.1。

3.1 实验数据和预处理

实验所采用的 PPI 数据从 HINT Database 2012[21] 下载，包含 9859 个顶点，40705 条边。基因表达数据从 TCGA 数据库[4] 中下载，其中乳腺癌（BRCA）数据包含 1091 例癌症样本和 113 例正常样本，宫颈癌（CC）数据包含 304 例癌症样本和 3 例正常样本。实验首先需要对基因表达数据进行预处理，删除缺失或低表达值的基因，并采用以 log 2 对数变换对数据进行标准化处理。后续实验仅考虑基因表达数据和 PPI 网络的共同基因，即 BRCA 数据集的 8601 个基因和 34449 条边，CC 数据集的 7847 个基因和 31111 条边。

3.2 评价指标

（1）癌症基因数量（NCG）

本文采用 NCG 指数[11] 评估基因优先级排序方法的识别性能。NCG 指数是指基因优先级排序列表的前 n 个基因中覆盖的癌症基因的个数。

（2）富集倍数（Fold enrichment）

富集倍数[22] 被广泛用来评估活跃模块中识别已知癌症基因的性能，其计算公式如下：

$$Fold\ enrichment = \frac{`Recovered' \times `All'}{`Selected' \times `Reference'} \tag{10}$$

其中，*All* 表示数据集中所有基因的数目，*Recovered* 表示参考基因中识别的癌症基因数目，*Selected* 表示模块的基因数目，*Reference* 表示参考基因的数目。

在线人类孟德尔遗传数据库（Online Mendelian Inheritance in Man，OMIM）[23] 是一个关于人类基因和遗传疾病信息的公共数据库。该数据库中收集了 1255 个被标记为癌症基因的关键基因。1255 个基因中有 996 个基因在 BRCA 数据集中覆盖。简而言之，在 BRCA 数据集中，*Reference*=996。宫颈癌基因数据库（Cervical Cancer gene DataBase，CCDB）[22] 是一个人工搜集整理的数据库，该数据库中包含经过实验证实的参与宫颈癌发展不同阶段的 537 个基因。537 个基因中有 362 个基因被宫颈癌数据集覆盖，即 *Reference*=362。

（3）连通强度（Connection Strength）

给定的活跃模块 $\hat{V}$ 中，λ 为模块 $\hat{V}$ 中边的数目。模块的连通强度[10] 的计算方法如式（11）所示：

$$\rho(\hat{V}) = \frac{2\lambda}{|\hat{V}|(|\hat{V}|-1)} \tag{11}$$

3.3 结果分析

基因优先级排序方法中设置参数 $p=2$。SA-PROX 算法的参数设置如下：迭代次数 $L=100$，起始温度 $T_0=97$，终止温度 $T_{end}=3$，降温系数 $t=0.95$。

（1）乳腺癌数据集（BRCA Dataset）

图 2 比较了 RegMod，LEAN，SigMod，ModFinder 和 IdeMod 方法在 $n=100\sim800$ 时的 NCG 指数。实验结果表明，IdeMod 方法可以识别出更多的被 OMIM 数据库证实的癌症基因，证明了提出的基因优先级排序方法的有效性，即排名越靠前的基因与癌症的相关性可能越大。表 1 列出了在 4 种参数 *ms* 设置下，IdeMod 方法所识别的活跃模块的富集倍数和连通强度。当 $ms=50$ 时，IdeMod 方法所识别的活跃模块有着最高的富集倍数和较高的连通强度。表 2 比较了 4 种不同方法所识别的活跃模块的富集倍数和连通强度。由于 LEAN 方法和 RegMod 识别了若干个活跃模块，因此无法对连通强度进行合理的比较。从表 2 中可以看出，在乳腺癌数据集中，IdeMod 方法识别的活跃模块比其他方法所识别的活跃模块具有更高的富集倍数和连通强度。表 3 给出在 IdeMod 方法识别规模 $ms=50$ 的活跃模块中，有 39 个基因已被 OMIM 数据库证实为与癌症有关的基因，其中 14 个用黑体标注的基因仅被 IdeMod 方法识别，其余 4 种方法并未识别。基因 TSC2 的多态性变异与乳腺癌的产生和发展有关[25]。基因 CDC6 是一个关键的复制许可因子，在调节 DNA 复制过程中起重要作用，与恶性肿瘤的早期发展有关[26]。基因 NPM1 是乳腺癌预后的不良因素[27]。基因 CEP55 与包括乳腺癌在内的多种恶性肿瘤有关[28]。基因 MCM7 通过抑制 AKT1/mTOR 信号通路促进肿瘤细胞增殖[29]。基因 MCM2 是口腔癌、胃癌、结肠癌和乳腺癌预后的良好增殖标志物[30]。基因 SKP2

在乳腺癌的发病机制中也起重要作用,是预防和治疗人类癌症的药物靶点[31]。基因 CDKN1B 的变异基因型可能参与乳腺癌的病因学[32]。基因 CDKN2A 在许多不同类型的肿瘤中经常发生突变,该基因的遗传变异与乳腺癌风险增加相关[33]。基因 CDK2 与多种癌症类型的肿瘤生长密切相关[34]。基因 CCNA2 为致癌基因,在调节癌细胞生长和凋亡中发挥重要作用[35]。基因 CDC7 参与细胞分裂和细胞周期等生命活动,并且与癌症进展有关[36]。MCM 基因家族的过表达与多种癌症进展之间的关联已在多项研究实为与癌症有关的基因[37]。

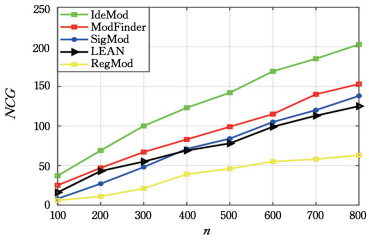

图 2　乳腺癌数据集中基因优先级排序方法结果比较

Fig. 2　Comparison of results of gene prioritization methods in BRCA dataset

表 1　乳腺癌数据集中不同 *ms* 规模下模块的富集倍数和连通强度

Table 1　Enrichment fold and connection strength with different *ms* in BRCA dataset

ms	Selected genes	Recovered edges	Recovered genes	Fold enrichment	Connection strength
30	30	140	22	6.33	0.321
50	50	376	39	6.73	0.306
70	70	464	51	6.29	0.192
90	90	584	56	5.37	0.145

表 2　乳腺癌数据集中识别的活跃模块性能比较

Table 2　Performance comparison of active modules in BRCA dataset

Methods	Selected genes	Recovered edges	Recovered genes	Fold enrichment	Connection strength
IdeMod	50	376	39	6.73	0.306
ModFinder	50	305	37	6.39	0.248
SigMod	299	1 227	94	2.71	0.027
LEAN	103	113	24	2.01	—
RegMod	298	8	23	0.66	—

表 3　乳腺癌数据集中 IdeMod 方法识别的致癌基因或抑癌基因(*ms*=50)

Table 3　Oncogenes or tumor suppressors identified by IdeMod in BRCA dataset (*ms*=50)

Identified oncogenes or tumor suppressors listed in OMIM							
TP53	BRCA1	RB1	**TSC2**	CTNNB1	PIN1	**CDC6**	**MCM5**
CDC7	HDAC1	SMAD2	PIK3R1	NPM1	GRB2	**CEP55**	**MCM7**
AR	**MCM2**	JUN	SKP2	ESR1	**CDKN1B**	RPA1	CDK4	PBK
CDK1	CREBBP	**MCM4**	**CDKN2A**	TP73	SMAD4	STAT3	SMAD3
CDK2	CCND1	**CCNA2**	CDKN1A	MYC	EP300		

为进一步验证 IdeMod 方法所识别的活跃模块,应用 DAVID 软件对 *ms*=50 的活跃模块(如图 3 所示)所包含的基因进行 KEGG 通路分析,得到的富集结果如表 4 所列。其中,KEGGID 表示 KEGG 通路编号,*P-value* 表示显著性结果 *P* 值,*Count* 表示富集基因个数,Term 表示生物过程描述。通过文献验证发现,p53 信号通路在抑制细胞生长、细胞迁移、细胞衰老或凋亡等方面发挥重要作用,与肿瘤的产生与发展有紧密联系[38]。FoxO 信号通路参与调节肿瘤细胞生存、增生、分化以及新陈代谢,对癌症的产生与发展至关重要[39]。PI3K-Akt 信号通路是乳腺癌的重要治疗靶点之一[40]。抑制 PI3K-Akt 信号通路可有效治疗多种类型的癌症[41]。JAK-STAT 信号通路参与细胞增殖、分化和凋亡,有助于肿瘤的形成[42]。Hippo 信号通路调控乳腺癌的发生和转移[43]。ErbB 信号通路的激活可以调节诱导上皮间质转化相关的正常和恶性乳腺上皮细胞的侵袭和迁移[44]。研究证明,ErbB 信号通路参与乳腺癌的细胞存活、细胞迁移和细胞增殖等活动[45]。TGF-beta 信号通路影响细胞生长、细胞分化、凋亡和细胞稳态等细胞功能,并通过抑制细胞增殖在乳腺癌中发挥肿瘤抑制作用[46]。HIF-1 信号通路在乳腺癌扩散与转移中起着重要作用[47]。AMPK 信号通路可以调节肿瘤细胞的侵袭和转移[48]。因此,活跃模块中的基因所调控的通路与乳腺癌的发展演变密切相关。

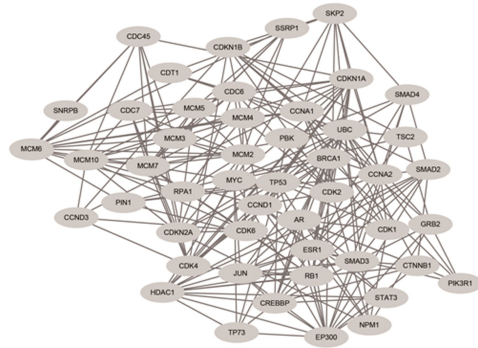

图 3　乳腺癌数据集中由 IdeMod 方法识别的 *ms*=50 的活跃模块

Fig. 3　Active module with *ms*=50 identified by IdeMod in BRCA dataset

表 4　乳腺癌数据集 KEGG 富集结果

Table 4　KEGG enrichment results in BRCA dataset

KEGGID	*P-value*	*Count*	Term
hsa04115	2.62×10^{-12}	11	p53 signaling pathway
hsa04068	4.42×10^{-11}	12	FoxO signaling pathway
hsa04151	1.67×10^{-7}	26	PI3K-Akt signaling pathway
hsa04630	1.62×10^{-6}	18	JAK-STAT signaling pathway
hsa04390	1.53×10^{-5}	16	Hippo signaling pathway
hsa04012	7.66×10^{-5}	12	ErbB signaling pathway
hsa04350	1.24×10^{-4}	12	TGF-beta signaling pathway
hsa04066	2.49×10^{-4}	12	HIF-1 signaling pathway
hsa04152	3.62×10^{-2}	10	AMPK signaling pathway

(2)宫颈癌数据集(CC Dataset)

图 4 给出了 RegMod,LEAN,SigMod,ModFinder 和 IdeMod 方法在 $n=100\sim800$ 时的 NCG 指数结果比较。同样,在宫颈癌数据集中,IdeMod 识别了更多的被 CCDB 数据证实的与宫颈癌有关的基因。

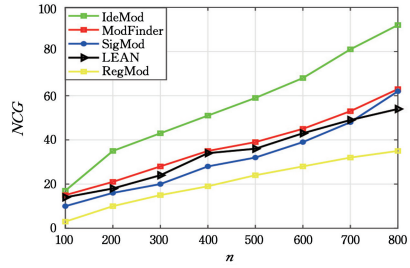

图 4　宫颈癌数据集中基因优先级排序方法结果比较

Fig. 4　Comparison of results of gene prioritization methods in CC dataset

表 5 列出了在宫颈癌数据集中不同规模活跃模块的富集倍数和连通强度。同样，当 $ms=50$ 时，IdeMod 识别的活跃模块(如图 5 所示)有着最高的富集倍数和较高的连通强度。表 6 对 5 种方法识别的活跃模块的性能进行了比较。结果显示，IdeMod 识别的活跃模块在富集倍数和连通强度上均优于其他方法，证明了 IdeMod 方法的识别性能。表 7 列出了 IdeMod 方法识别的活跃模块中所覆盖的与宫颈癌有关的致癌基因或抑癌基因，这些基因均被 CCDB 数据库所收录，对宫颈癌的发展过程中起着重要作用。其中，有 16 个基因是其他方法没有检测到的，在表中用粗体标注。

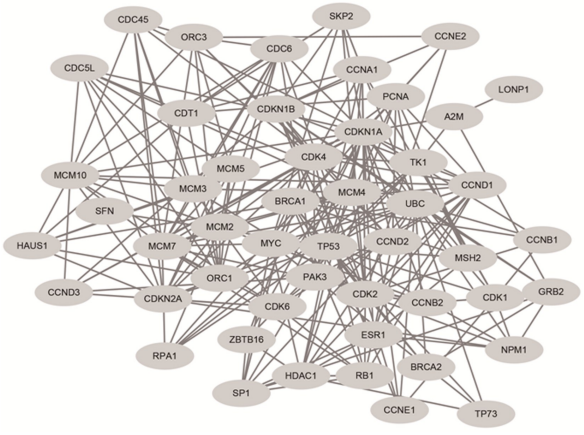

图 5 宫颈癌数据集中由 IdeMod 方法识别的 $ms=50$ 的活跃模块

Fig. 5 Active module with $ms=50$ identified by IdeMod in dataset CC

表 5 宫颈数据集中不同 *ms* 规模下模块的富集倍数和连通强度

Table 5 Enrichment fold and connection strength with different *ms* in CC dataset

ms	Selected genes	Recovered edges	Recovered genes	Flod enrichment	Connection strength
30	30	121	14	10.11	0.278
50	50	245	25	10.83	0.200
70	70	443	26	8.05	0.183
90	90	578	33	7.94	0.144

表 6 宫颈癌数据集中识别的活跃模块性能比较

Table 6 Performance comparison of active modules in CC dataset

Methods	Selected genes	Recovered edges	Recovered genes	Fold enrichment	connection strength
IdeMod	50	245	25	10.83	0.200
ModFinder	50	239	20	4.05	0.195
SigMod	107	138	20	4.05	0.024
LEAN	154	379	31	4.36	—
RegMod	194	12	9	1.00	—

表 7 宫颈癌数据集中 IdeMod 方法识别的致癌基因或抑癌基因($ms=50$)

Table 7 Oncogenes or tumor suppressors identified by IdeMod in CC dataset ($ms=50$)

Identified oncogenes or tumor suppressors listed in CCDB						
CCNE1	**BRCA1**	CDK2	**CDKN1A**	CDKN2A	**MCM2**	**SFN**
CDKN1B	**A2M**	**TP53**	**CDK6**	GRB2	**MYC**	ESR1
MSH2	**CCND2**	**MCM4**	**MCM3**	TP73	PCNA	CCND1
MCM5	**SKP22**	CCNB1	**CCNA1**			

通过查阅文献发现，基因 BRCA1 是一种抑癌基因，与遗传性乳腺癌的产生与发展有关[49]。基因 CDKN1A 与宫颈癌的产生、发展和患者预后有关[50]。基因 SFN 是癌症中的一种新型生物标志物，可以促进早期癌症的发展[51]。基因 CDKN1B 是肿瘤抑制基因，与多种癌症的产生和发展有关[52-53]。基因 A2M 可能与肿瘤发生有关[54]。肿瘤抑制基因 TP53 是人类癌症中最常发生突变的基因[55]。基因 CDK6 为细胞周期的重要调节因子，在宫颈癌的肿瘤发生中起关键作用[56]。基因 MYC 的高表达可能促进宫颈癌的侵袭和转移[57]。基因 MSH2 能促进细胞增殖、细胞周期进程和正向调节细胞侵袭性[58]。细胞周期异常是导致宫颈癌的重要早期事件。CCND2 的异常表达导致细胞周期失控，细胞无限增殖，丧失凋亡能力，从而促进了宫颈癌的形成[59]。基因 SKP2 表达增加是宫颈癌局部复发的独立预测因子[60]。基因 CCNA1 可能在 HPV 诱发的宫颈癌中发挥重要作用[61]。

通过对 $ms=50$ 的活跃模块包含的基因进行 KEGG 通路分析，得到的富集结果如表 8 所列。Pathways in cancer(癌症通路)是 KEGG pathway 数据库对各种癌症通路的整合，其中包括神经胶质瘤、子宫内膜癌、宫颈癌、肾细胞癌和结直肠癌等。癌症通路是与癌症有关的主要信号通路[62]。PI3K-Akt 信号通路促进细胞代谢、增殖、存活、生长，参与癌症的发生、化疗耐药和血管生成的调控[63]。FoxO 信号通路参与细胞分化、细胞凋亡、细胞增殖、DNA 损伤和修复等多种细胞功能，调控癌细胞的侵袭转移能力，被认为是治疗癌症的潜在靶点[64]。失调的 Jak-STAT 信号通路促进癌症产生和转移[65]。靶向 ErbB 信号通路是治疗宫颈癌的治疗策略之一[66]。通过 KEGG 富集分析，进一步证实了活跃模块所调控的通路对癌症具有重要影响。

表 8 宫颈癌数据集 KEGG 富集结果

Table 8 KEGG enrichment results in dataset CC

KEGGID	*P-value*	*Count*	Term
hsa04115	7.54×10^{-21}	16	p53 signaling pathway
hsa05200	2.83×10^{-15}	23	Pathways in cancer
hsa04151	2.36×10^{-8}	14	PI3K-Akt signaling pathway
hsa04068	3.87×10^{-7}	9	FoxO signaling pathway
hsa04012	1.08×10^{-3}	5	ErbB signaling pathway
hsa04630	1.68×10^{-3}	6	JAK-STAT signaling pathway

结束语 活跃模块识别问题是生物信息学的重要研究问题。本文通过融合 PPI 网络数据和基因表达数据，利用带 p 步随机游走核的回归模型设计了一个基因评分函数，提出了基于 POC 方法的基因优先级排序方法，避免了人为设置参数因素的干扰。在基因优先级排序方法的基础上，进一步提出了模拟退火算法 SA-PROX 用于识别 PPI 网络中优先级高且连通性强的活跃模块。

利用真实的乳腺癌数据集和宫颈癌数据集对 RegMod，LEAN，SigMod，ModFinder 和 IdeMod 方法进行实验测试。实验结果表明，IdeMod 方法可以成功识别一个包含大量的致癌基因或抑癌基因的强连通模块。综上所述，IdeMod 方法可以成为识别活跃模块的有效补充工具，在了解癌症的发病机制中发挥一定的辅助作用。

参 考 文 献

[1] SANTARIUS T, SHIPLEY J, BREWER D, et al. A census of amplified and overexpressed human cancer genes[J]. Nature Reviews Cancer, 2010, 10(1): 59-64.

[2] ALÓPEZ-CORTÉS, C PAZ-Y-MIÑO, GUERRERO S, et al. OncoOmics approaches to reveal essential genes in breast cancer: a

panoramic view from pathogenesis to precision medicine [J]. Scientific Reports,2020,10(1).

[3] LV K,YANG J,SUN J,et al. Identification of key candidate genes for pancreatic cancer by bioinformatics analysis[J]. Experimental and Therapeutic Medicine,2019,18(1).

[4] TOMCZAK K,CZERWINSKA P,WIZNEROWICZ M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge[J]. Contemporary Oncology,2015,19(1A):A68-77.

[5] HUDSON T J. International network of cancer genome projects [J]. Nature,2010,464(7291):993.

[6] HE H,LIN D,ZHANG J,et al. Comparison of statistical methods for subnetwork detection in the integration of gene expression and protein interaction network[J]. BMC Bioinformatics,2017,18(1):149.

[7] SHAH S D,BRAUN R. genesurrounder:network-based identification of disease genes in expression data[J]. BMC Bioinformatics,2019,229(20).

[8] QIU Y Q,ZHANG S,ZHANG X S,et al. Detecting disease associated modules and prioritizing active genes based on high throughput data[J]. BMC Bioinformatics,2010,11(1):1-12.

[9] FREDERIK G,B GWÉNOLA,CLAIRE V,et al. Network-based analysis of omics data:The LEAN method[J]. Bioinformatics,2017(5):701-709.

[10] LIU Y,BROSSARD M,D ROQUEIRO,et al. SigMod:an exact and efficient method to identify a strongly interconnected disease-associated module in a gene network[J]. Bioinformatics,2017,33(10):1536-1544.

[11] VLAIC S,CONRAD T,TOKARSKI S C,et al. ModuleDiscoverer:identification of regulatory modules in protein-protein interaction networks[J]. Scientific Reports,2018,8(1):1-11.

[12] TIAN Y,SU X,SU Y,et al. EMODMI:A multi-objective optimization based method to identify disease modules[J]. IEEE Transactions on Emerging Topics in Computational Intelligence,2020,5(4):570-582.

[13] WU J,ZHANG Q,LI G. Identification of cancer-related module in protein-protein interaction network based on gene prioritization [J]. Journal of Bioinformatics and Computational Biology,2022,20(1).

[14] LI Y,RATA I,CHIU S,et al. Improving predicted protein loop structure ranking using a Pareto-optimality consensus method [J]. BMC Structural Biology,2010,10(1):1-14.

[15] ZOU K H,KEMAL T,SILVERMAN S G. Correlation and simple linear regression[J]. Radiology,2003,227(3):617-622.

[16] BHATTACHARYYA A. On a measure of divergence between two statistical populations defined by their probability distributions[J]. Bull Calcutta Math Soc,1943,35:99-109.

[17] VAPNIK V,Statistical Learning Theory[M]. New York:John Wiley,1998.

[18] SMOLA A J,KONDOR R. Kernels and Regularization on Graphs[J]. Learning Theory and Kernel Machines,2003,(12):144-158.

[19] CHANG C C,LIN C J. LIBSVM:A library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology,2011,2(3):1-27.

[20] ANASTASIS O,GEORGE M,MARGARITA Z,et al. Systems Bioinformatics:increasing precision of computational diagnostics and therapeutics through network-based approaches[J]. Briefings in Bioinformatics,2019,20(3):806-824.

[21] PATIL A,NAKAMURA H. HINT:a database of annotated protein-protein interactions and their homologs[J]. Biophysics,2005,1(364):21-24.

[22] HAMOSH A. Online Mendelian Inheritance in Man(OMIM),a knowledgebase of human genes and genetic disorders[J]. Nucleic Acids Research,2002,30(1):52-55.

[23] SUBHASH,M,AGARWAL,et al. CCDB:a curated database of genes involved in cervix cancer[J]. Nucleic Acids Research,2011,39(suppl_1):975-979.

[24] AGARWAL S M,RAGHAV D,SINGH H,et al. CCDB:a curated database of genes involved in cervix cancer[J]. Nucleic Acids Research,2010,39(suppl_1):D975-D979.

[25] MEHTA M S,VAZQUEZ A,KULKARNI D A,et al. Polymorphic variants in TSC1 and TSC2 and their association with breast cancer phenotypes[J]. Breast Cancer Research and Treatment,2011,125(3):861-868.

[26] LIM N,TOWNSEND P A. Cdc6 as a novel target in cancer:Oncogenic potential,senescence and subcellular localisation[J]. International Journal of Cancer,2020,147(6):1528-1534.

[27] KARIMI D F,GHOLAMZADEH K S,AFSHAR S,et al. The potential role of nucleophosmin(NPM1) in the development of cancer[J]. Journal of Cellular Physiology,2021,236(11):7832-7852.

[28] SINHA D. Strategic inhibition of CEP55 in aggressive breast cancer leads to mitotic catastrophe[J]. Annals of Oncology,2019,30:v1.

[29] QIU Y T,WANG W J,ZHANG B,et al. MCM7 amplification and overexpression promote cell proliferation,colony formation and migration in esophageal squamous cell carcinoma by activating the AKT1/mTOR signaling pathway[J]. Oncology Reports,2017,37(6):3590-3596.

[30] DENG M,SUN J,XIE S,et al. Inhibition of MCM2 enhances the sensitivity of ovarian cancer cell to carboplatin[J]. Molecular Medicine Reports,2019,20(3):2258-2266.

[31] WANG Z,FUKUSHIMA H,INUZUKA H,et al. Skp2 is a promising therapeutic target in breast cancer[J]. Frontiers in Oncology,2012,1:57.

[32] MA H,JIN G,HU Z,et al. Variant genotypes of CDKN1A and CDKN1B are associated with an increased risk of breast cancer in Chinese women[J]. International Journal of Cancer,2006,119(9):2173-2178.

[33] SHAHIDSALES S,MEHRAMIZ M,GHASEMI F,et al. A genetic variant in CDKN2A/B gene is associated with the increased risk of breast cancer[J]. Journal of Clinical Laboratory Analysis,2018,32(1):e22190.

[34] TADESSE S,CALDON EC,TILLEY W,et al. Cyclin dependent kinase 2 inhibitors in cancer therapy:an update[J]. Journal of Medicinal Chemistry,2018,62(9):4233-4251.

[35] GAN Y,LI Y,LI T,et al. CCNA2 acts as a novel biomarker in regulating the growth and apoptosis of colorectal cancer[J]. Cancer Management and Research,2018,10:5113.

[36] MELLING N,MUTH J,SIMON R,et al. Cdc7 overexpression is an independent prognostic marker and a potential therapeutic target in colorectal cancer[J]. Diagnostic Pathology,2015,10(1):1-7.

[37] WANG D,LI Q,LI Y,et al. The role of MCM5 expression in cervical cancer:Correlation with progression and prognosis[J]. Biomedicine & Pharmacotherapy,2018,98:165-172.

[38] XIE B,NAGALINGAM A,KUPPUSAMY P,et al. Benzyl Isothiocyanate potentiates p53 signaling and antitumor effects

against breast cancer through activation of p53-LKB1 and p73-LKB1 axes[J]. Scientific Reports, 2017, 7(1): 1-14.
[39] FARHAN M, WANG H, GAUR U, et al. FOXO signaling pathways as therapeutic targets in cancer[J]. International Journal of Biological Sciences, 2017, 13(7): 815.
[40] MA C X. The PI3K pathway as a therapeutic target in breast cancer[J]. American Journal of Hematology/Oncology, 2015, 11(3).
[41] YANG J, NIE J, MA X, et al. Targeting PI3K in cancer: mechanisms and advances in clinical trials[J]. Molecular Cancer, 2019, 18(1): 1-28.
[42] THOMAS S J, SNOWDEN J A, ZEIDLER M P, et al. The role of JAK/STAT signalling in the pathogenesis, prognosis and treatment of solid tumours[J]. British Journal of Cancer, 2015, 113(3): 365-371.
[43] WU L, YANG X. Targeting the Hippo pathway for breast cancer therapy[J]. Cancers, 2018, 10(11): 422.
[44] HARDY K M, BOOTH B W, HENDRIX M J C, et al. ErbB/EGF signaling and EMT in mammary development and breast cancer[J]. Journal of Mammary Gland Biology and Neoplasia, 2010, 15(2): 191-199.
[45] PAZ-Y-MIÑO C, LÓPEZ-CORTÉS A, MUÑOZ M J, et al. Incidence of the L858R and G719S mutations of the epidermal growth factor receptor oncogene in an Ecuadorian population with lung cancer[J]. Cancer Genetics and Cytogenetics, 2010, 196(2): 201-203.
[46] ZUGMAIER G, ENNIS B W, DESCHAUER B, et al. Transforming growth factors type beta1 and beta2 are equipotent growth inhibitors of human breast cancer cell lines[J]. Journal of Cellular Physiology, 1989, 141(2): 353-361.
[47] LIU Z, SEMENZA G L, ZHANG H. Hypoxia-inducible factor 1 and breast cancer metastasis[J]. Journal of Zhejiang University-SCIENCE B, 2015, 16(1): 32-43.
[48] LI N, HUANG D, LU N, et al. Role of the LKB1/AMPK pathway in tumor invasion and metastasis of cancer cells[J]. Oncology Reports, 2015, 34(6): 2821-2826.
[49] RAKHA E A, EL-SHEIKH S E, KANDIL M A, et al. Expression of BRCA1 protein in breast cancer and its prognostic significance[J]. Human Pathology, 2008, 39(6): 857-865.
[50] ZHANG X, LI F, ZHU L. Clinical significance and functions of microRNA-93/CDKN1A axis in human cervical cancer[J]. Life Sciences, 2018, 209: 242-248.
[51] HU Y, ZENG Q, LI C, et al. Expression profile and prognostic value of SFN in human ovarian cancer[J]. Bioscience Reports, 2019, 39(5).
[52] LAUTER K B, ARNOLD A. Mutational analysis of CDKN1B, a candidate tumor-suppressor gene, in refractory secondary tertiary hyperparathyroidism[J]. Kidney International, 2008, 73(10): 1137-1140.
[53] CUSAN M, MUNGO G, DE MARCO ZOMPIT M, et al. Landscape of CDKN1B mutations in luminal breast cancer and other hormone-driven human tumors[J]. Frontiers in Endocrinology, 2018, 9: 393.
[54] ACUNER-OZBABACAN E S, ENGIN B H, GUVEN-MAIOROV E, et al. The structural network of Interleukin-10 and its implications in inflammation and cancer[J]. BMC Genomics, 2014, 15(4): 1-17.
[55] SZYMAŃSKA K, HAINAUT P. TP53 and mutations in human cancer[J]. Acta Biochimica Polonica, 2003, 50(1): 231-238.
[56] HU Q L, XU Z P, LAN Y F, et al. miR-636 represses cell survival by targeting CDK6/Bcl-2 in cervical cancer[J]. The Kaohsiung Journal of Medical Sciences, 2020, 36(5): 328-335.
[57] WU S H, ZENG X F, WANG P, et al. The Expression and Significance of c-myc and bcat1 in Cervical Cancer[J]. Sichuan da xue xue bao. Journal of Sichuan University(Medical Science Edition), 2018, 49(5): 725-730.
[58] SHEN K, YE Y J, JIANG K W, et al. MSH2 is required for cell proliferation, cell cycle control and cell invasiveness in colorectal cancer cells[J]. Chinese Science Bulletin, 2012, 57(20): 2580-2585.
[59] SHEN S N, WANG H, GONG B L. Expressions of miRNA29 target genes CCND2 and CDK6 in cervical cancer[J]. Disc Clin Cases, 2018, 5: 14-18.
[60] FU H C, YANG Y C, CHEN Y J, et al. Increased expression of SKP2 is an independent predictor of locoregional recurrence in cervical cancer via promoting DNA-damage response after irradiation[J]. Oncotarget, 2016, 7(28): 44047.
[61] ZUO Q, ZHENG W, ZHANG J, et al. Methylation in the promoters of HS3ST2 and CCNA1 genes is associated with cervical cancer in Uygur women in Xinjiang[J]. The International Journal of Biological Markers, 2014, 29(4): 354-362.
[62] GENG F G. A Functional Module Detection Method byCombining Cancer Multi-omics Data and PPINetwork [D]. Xi'an: Xidian University, 2019.
[63] FU K, ZHANG L, LIU R, et al. MiR-125 inhibited cervical cancer progression by regulating VEGF and PI3K/AKT signaling pathway [J]. World Journal of Surgical Oncology, 2020, 18: 1-10.
[64] FARHAN M, WANG H, GAUR U, et al. FOXO signaling pathways as therapeutic targets in cancer[J]. International Journal of Biological Sciences, 2017, 13(7): 815.
[65] GUTIÉRREZ-HOYA A, SOTO-CRUZ I. Role of the JAK / STAT Pathway in Cervical Cancer: Its Relationship with HPV E6/E7 Oncoproteins[J]. Cells, 2020, 9(10): 2297.
[66] HE C, MAO D, HUA G, et al. The Hippo/YAP pathway interacts with EGFR signaling and HPV oncoproteins to regulate cervical cancer progression[J]. EMBO Molecular Medicine, 2015, 7(11): 1426-1449.

ZHANG Qi, born in 1995, master. Her main research interests include bioinformatics and algorithm optimization.