

基于蛋白质互作网络挖掘结直肠癌致病基因

吴慧慧¹ 唐旭清^{1,2}

(1. 江南大学理学院, 无锡, 214122; 2. 江南大学无锡市生物计算工程技术研究中心, 无锡, 214122)

摘要: 结直肠癌是消化系统常见的恶性肿瘤之一, 死亡率居发达国家恶性肿瘤死亡率的第3位。本文通过生物分析进行结直肠癌致病基因的识别。首先, 基于GEO中GSE9348基因表达数据集, 利用R语言的LIMMA包筛选出 $P < 0.05$, Fold change > 2 的结直肠癌差异基因339个; 其次, 基于OMIM数据库中已知结直肠癌的致病基因和STRING数据库, 获得差异表达基因与致病基因的蛋白质互作网络; 接着利用Cytoscape软件的ClusterONE插件进行蛋白质互作网络模块分析, 获得一个含有53个基因的子网络; 最后, 通过对子网络的拓扑分析, 获得了FOS、CCND1、CEBPB、EGR1和NOS3等5个新结直肠癌致病基因。同时, 通过功能富集分析和文献挖掘对新发现的致病基因进行验证。

关键词: 结直肠癌; 蛋白互作网络; 聚类分析; 网络拓扑分析; 功能富集分析

中图分类号: TP391; O29 **文献标志码:** A

Mining of Disease Genes for Colorectal Cancer Based on Protein Interaction Network

Wu Huihui¹, Tang Xuqing^{1,2}

(1. College of Science, Jiangnan University, Wuxi, 214122, China; 2. Wuxi Engineering Research Center for Biocomputing, Jiangnan University, Wuxi, 214122, China)

Abstract: Colorectal cancer is one of the common malignant tumors in digestive system, and its mortality rate is third of malignant tumor mortality in developed countries. The aim of this paper is to identify the pathogenic gene of colorectal cancer through biological analysis and data mining. Firstly, the expression spectrum dataset GSE9348 is downloaded from GEO database, and 339 differentially expressed genes are screened with $P < 0.05$ and Fold change > 2 in colorectal cancer by using LIMMA Package in R language. Secondly, based on the known disease genes of colorectal cancer, OMIM database and STRING database, the PPI network composed by the differentially expressed genes and known disease genes is obtained. Furthermore, the network module analysis is performed through ClusterONE plugin of Cytoscape software, and a subnetwork containing 53 genes is obtained. Finally, through network topology analysis, 5 candidate genes of colorectal cancer are considered to be candidate disease genes of colorectal cancer, including CCND1, EGR1, FOS, CEBPB and NOS3. Simultaneously, the newly discovered genes are verified by using the functional enrichment analysis and literature mining.

Key words: colorectal cancer; protein interaction network; clustering analysis; network topology analysis; functional enrichment analysis

引言

结直肠癌(Colorectal cancer, CRC)包括结肠癌和直肠癌,是目前全球范围内较为常见的消化系统恶性肿瘤之一^[1]。近几年来,随着经济的发展、人们生活水平的提高、饮食习惯的改变以及遗传因素的影响等,我国结直肠癌的发病率和死亡率逐年上升^[2],其病死率居世界范围内恶性肿瘤的第4位,发病率居第3位^[3],已严重威胁人类的健康。大多数研究者认为结直肠癌是一种复杂疾病,其发生与发展是在遗传因素、环境因素以及肿瘤发展过程的个体反应等背景下,多因素参与相互作用的结果。因此,预测并鉴定CRC的致病基因对深入理解其致病机制有着重要的意义。

由多个基因相互作用所产生的疾病通常被认为是复杂疾病,而挖掘与其相关的致病基因并阐释它们在疾病发展过程中所产生的作用已逐渐成为人们研究复杂疾病的主要目标之一。随着生物信息学分析的发展,基于网络的方法已经成为研究疾病机制的有力工具^[4-6],其主要包括共表达网络^[7]、蛋白质互作网络^[8]、蛋白质磷酸化网络^[9]和DNA甲基化网络^[10]等生物分子网络。而随着蛋白质相互作用(Protein-protein interaction, PPI)数据的丰富,基于蛋白质互作网络^[11]的方式逐渐成为挖掘复杂疾病候选基因的主要方法。其中,模块分析或聚类分析作为一种工具和算法已经广泛应用到蛋白质互作网络分析中,例如从局部扩展来探测社区的重叠社区发现算法(Local and wave-like extension algorithm of detecting overlapping community, LWS-OCD)算法^[12]能够有效地发现无向图中的重叠区域。而拓扑分析作为另外一种分析复杂网络的工具也被广泛应用,如其中的中心性分析中图中心的措施如度、介数和接近中心性对识别在网络中具有关键作用的节点十分有用。在蛋白质互作网络中,通常称度数高的节点为中心节点,介数高的节点为瓶颈节点,这些节点在网络中都发挥着举足轻重的作用。

疾病基因筛选面临的主要问题是数据高维且样本量少,因此其数据处理主要侧重于数据挖掘方法的多重筛选与验证。本研究是基于结直肠癌的公开表达谱数据,进行结直肠癌差异表达基因分析以获取疾病基因的初选;通过这些初选的疾病基因与结直肠癌已知致病基因之间的重要关系来组建互作网络,进行结直肠癌致病的候选基因筛选;最后,利用功能富集分析和子网络的拓扑分析进行结直肠癌疾病候选基因的筛选与验证。

1 实验材料与方法

1.1 材料来源

本文从美国生物技术中心GEO(<http://www.ncbi.nlm.nih.gov/geo/>)数据库下载结直肠癌基因表达谱数据,登录号为GSE9348,平台为Affymetrix Plus 2.0。下载其中的82例包含癌旁样本12例和癌样本70例作为样本数据,其中数据以DNA微阵列的表达谱数据形式给出,且维数是5万多。这一数据作为本研究的实验数据,记为 Ω ;从在线孟德尔人类遗传学数据库(Online Mendelian inheritance in man, OMIM)中下载135个结直肠癌的疾病基因,这些疾病基因是已被证实的结直肠癌致病基因,该数据在本文中当作验证数据,记为 Ω_0 。

1.2 方法

1.2.1 数据处理

在数据 Ω 的基础上,首先通过软件R对原始数据进行重复值合并、数据转换、标准化和缺失数据删除等处理。然后,利用T检验法进行基因筛选实现降维,挑选出样本间有显著性差异的基因,即筛选出结直肠癌的差异表达基因。这一过程是通过R中的LIMMA包(<http://www.bioconductor.org/packages/release/bioc/html/limma.html>)来实现,其中涉及参数 P 和Fold change。参数 P 为基因间差异的显著性水平,参数Fold change为实验组与对照组的差异表达的比值,且 $P < \alpha$ 和Fold change $>$

β 。一般地,参数 P 越小且 Fold change 越大,则筛选的基因就越多。在具体实验中,依据研究需要选取阈值 α 和 β ,一般分别取 0.05 和 2。通过数据处理获得的差异表达基因记为 Ω_1 。

1.2.2 蛋白质互作网络的构建

将基因集 Ω_0 和基因集 Ω_1 上传到 STRING(Search tool for the retrieval of interacting genes)数据库中,通过 SRTING 在线分析工具获得疾病基因与差异基因之间的相互作用^[13],然后选取可靠性指数 (Confidence scores) 大于 0.9 的基因对,以及它们之间的作用数据,得到数据集 Ω_2 。将 Ω_2 导入 Cytoscape 软件中,对结直肠癌的已知疾病与差异表达基因之间的相互作用进行可视化^[14],得到相互作用网络图。

1.2.3 子网络的识别

重叠领域扩展聚类(Clustering with overlap neighborhood expansion, ClusterONE)是一种从加权或非加权的蛋白质网络中挖掘重叠的密集连接区域的复合物识别算法^[15],已有很多研究者利用其识别能力在各种生物网络中检测有意义的局部结构^[16-17]。因此,本文可通过这种图聚类算法来挖掘蛋白质互作网络中节点高度连接的重叠区域,其算法的聚类凝聚力定义为

$$f(V) = \frac{W^{\text{in}}(V)}{W^{\text{in}}(V) + W^{\text{bound}}(V) + P|V|} \quad (1)$$

式中: $W^{\text{in}}(V)$ 表示一组顶点 V 内的边的总重量; $W^{\text{bound}}(V)$ 表示连接此组顶点到图的其余部分边的总重量; $P|V|$ 为惩罚项。

本文通过 Cytoscape 的 ClusterONE 插件来进行子网络的挖掘,其中,涉及到参数 minimum density、degree 和 P 的设定。通常情况下,参数 P 越大,一般被挖掘的聚类(模块)数量越少,minimum density 和 degree 越大,反之亦然。因此, P 值的选择应在一个合理范围内,一般认为不能太小。为便于分析,被挖掘的子网络模块的数量应在适宜的范围,因此参数 minimum density 和 degree 也有阈值,一般分别不小于 0.5 和 6。

1.2.4 子网络的拓扑分析与富集分析

在无标度网络中^[18]的中心节点通常由大量互作连线的节点来代表,而中心节点对应的蛋白质为核心蛋白质(基因),并且这些基因在生理调节的过程中扮演着重要的角色。因此,本文通过 Cytoscape 的 NetworkAnalyzer 插件来筛选子网络中的中心节点,NetworkAnalyzer 插件可对有向网络和无向网络进行拓扑分析,得到拓扑属性值如度、介数和接近中心等,定义如下:

(1) 度(Degree)

$$\text{deg}(v) = |N(v)| \quad (2)$$

式中: v 为节点, $N(v)$ 表示节点 v 邻节点的集合; $\text{deg}(v)$ 表示 v 与网络中其他节点的关联性,当 $\text{deg}(v)$ 越大时,说明 v 在网络中越重要。

(2) 接近中心性(Closeness centrality)

$$C_c(v) = \sum_{w \in V} \frac{1}{\text{dist}(v, w)} \quad (3)$$

式中: $\text{dist}(v, w)$ 表示节点 v, w 间的最短路径长度; $C_c(v)$ 表示 v 接近中心位置的程度,当 $C_c(v)$ 越小时,越说明 v 是网络的核心点,在网络中作用越重要。

(3) 介数(Betweenness centrality)

$$BC(v) = \sum_{s \neq t \neq v \in C(v)} \frac{\sigma_s(v)}{\sigma_s} \quad (4)$$

式中: $C(v)$ 为包含节点 v 的组成部分; σ_s 表示从节点 s 到节点 t 的最短路径数; $BC(v)$ 表示 v 对网络中其他节点之间通信连接的影响程度。当 $BC(v)$ 越大时,表明 v 的重要度越高。

CytoHubba 软件(<http://hub.iis.sinica.edu.tw/cytoHubba/index.html>)通过网络特征对网络中的节点进行排名,它实现了 11 个节点的排名方法来评估生物网络中节点的重要性,包括度、边缘渗出成

分、最大社区成分、最大邻域成分和最大团中心等,其中最大团中心(Maximal clique centrality, MCC)方式能够更准确地预测蛋白质互作网络中的重要基因^[19],其定义为

$$MCC(v) = \sum_{C \in S(v)} (|C| - 1)! \tag{5}$$

式中: $S(v)$ 是包含节点 v 的最大团簇, $(|C| - 1)!$ 是小于 $|C|$ 的所有正整数乘积。如果节点 v 的邻节点间没有边界,则节点 v 的最大团中心就是它的度。

子网络的拓扑分析就是利用网络节点度、介数、接近中心性等指标进行差异表达基因分析比较,筛选出子网络的中心节点,即结肠直肠癌的致病基因。子网络富集分析的目的是用来注释这些差异表达基因参与的生物学过程^[20],且通过选取基因本体论(Gene ontology, GO)条目来实现。在 GO 条目中涉及错误发现率(False discovery rate, FDR),且 $FDR < \gamma$ 。阈值 γ 越小表示 GO 条目样本变化判断越准确,试验中取 $\gamma = 0.05$ 。

综上,本文所采用方法的流程为

(1) 在 Ω_0 基础上,利用 R 软件和 LIMMA 包进行降维和筛选,获得差异表达基因,即基因集 Ω_1 ;

(2) 将基因集 Ω_0 和基因集 Ω_1 上传到 STRING 在线工具,获得可靠性大于 0.9 的基因对,并从中删除差异表达基因之间的基因对,得到数据集 Ω_2 ,利用 Cytoscape 得到 Ω_2 蛋白质互作网络;

(3) 在步骤(2)的基础上,通过 ClusterONE 聚类算法对蛋白质互作网络进行分析,获取其子网络,记子网络的基因集为 Ω_3 ;

(4) 在步骤(3)的基础上,进行子网络拓扑分析,列出节点度、介数、接近中心性排名前 10 的差异表达基因,并对这些差异表达基因进行分析比较,筛选出子网络中的中心节点,即结肠直肠癌的致病基因。

2 实验结果与分析

2.1 实验结果

在原始数据 Ω 的基础上,通过一系列的预处理和基因的筛选,共获得了 339 个差异基因 Ω_1 ,其中 $\alpha = 0.05, \beta = 2$ 。

采用 STRING 在线分析工具可获得 268 个可靠性指数大于 0.9 的基因对,即 $|\Omega_2| = 268$ 。利用 Cytoscape 对 Ω_2 可视化的结果如图 1 所示。在图 1 中,蛋白质相互作用网络由 132 个基因对组建。在此基

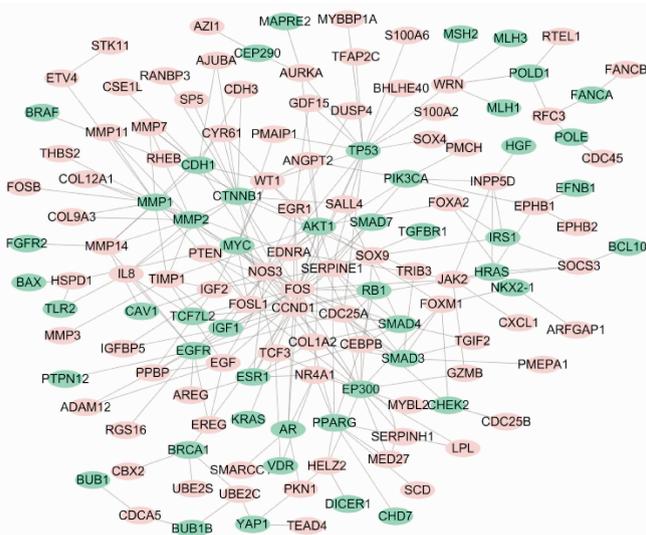


图 1 差异表达基因和已知疾病基因的蛋白质互作网络

Fig. 1 Protein-protein interaction network of identified differentially expressed genes and known disease genes

基础上,通过 ClusterONE 插件,取 $\text{minimum density} > 0.5$, $\text{degree} > 6$, $P < 0.001$,可挖掘出一个含有 53 个节点的子网络模块,如图 2 所示,其中图 1,2 中的粉色为识别差异表达基因,青色为结直肠癌的已知疾病基因。

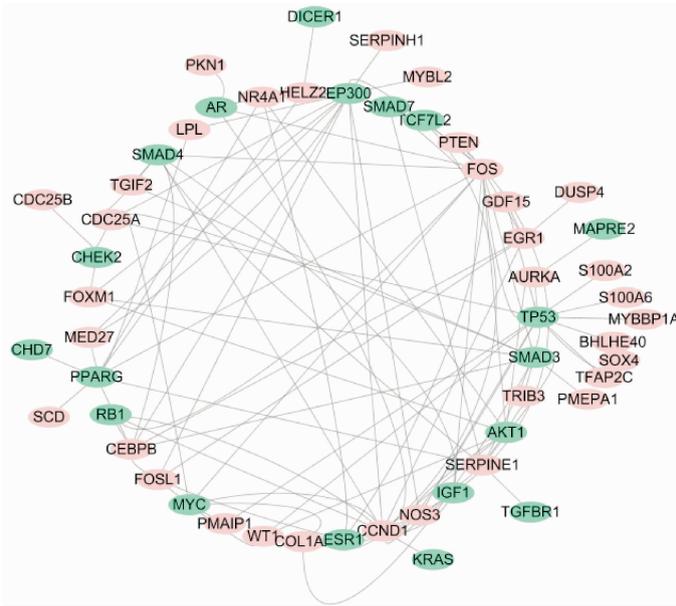


图 2 蛋白质互作网络中识别的子网络

Fig. 2 Sub-network identified from PPI network

为了识别子网络中潜在的枢纽基因,NetworkAnalyzer 插件统计网络中节点的度、介数和接近中心性,且节点度、介数和接近中心性最高的点被认为是中心节点,统计结果如表 1 所示,表中所列的是子网络中节点度、介数和接近中心性分别排名前 10 的差异表达基因。从表 1 可知,节点度、介数和接近中心性都排在前 10 的差异基因有 FOS,CCND1,CEBPB,EGR1 和 NOS3。因此,FOS,CCND1,CEBPB,EGR1 和 NOS3 在被识别的网络中具有重要作用,即为网络的中心节点。

此外,通过 Cytoscape 的 CytoHubba 插件获得了排名前 15 的差异表达基因,包括 TP53, FOS,

表 1 子网络中的中心基因

Tab. 1 Central genes in sub-network

Genes	Degree	Genes	Betweenness centrality	Genes	Closeness centrality
FOS	15	FOS	0.041 914 02	CCND1	1.000 00 000
CCND1	14	CDC25A	0.010 184 94	COL1A	1.000 000 00
SERPINE1	8	CCND1	0.008 495 83	CEBPB	0.727 272 73
CEBPB	7	FOXM1	0.005 674 06	EGR1	0.666 66 667
EGR1	6	CEBPB	0.005 124 70	FOS	0.655 172 41
CDC25A	6	NOS3	0.004 336 35	NR4A1	0.648 257 14
NR4A1	5	GDF15	0.003 519 36	HELZ2	0.583 333 34
NOS3	5	EGR1	0.002 211 02	NOS3	0.560 000 00
FOXM1	4	NR4A1	0.001 301 80	FOSL1	0.555 555 56
COL1A	4	WT1	0.001 057 90	LPL	0.545 454 55

EP300, CCND1, PPARG, AKT1, SMAD3, SERPINE1, CEBPB, MYC, SMAD4, CDC25A, EGR1 和 NOS3, 其中 FOS, CCND1, SERPINE1, CEBPB, CDC25A, EGR1 和 NOS3 为差异表达基因。结合筛选的中心节点, 可获得结直肠癌候选疾病基因 FOS, CCND1, CEBPB, EGR1, NOS3。

进一步, 对含有 53 个节点的子网络模块进行富集分析。依据 $FDR < 0.05$ 筛选出的子网络差异表达基因的 9 条 GO 条目测试, 结果见表 2, 其中 Count 记录的是子网络模块中参与相应生物学过程的基因数, FDR 是检验的错误发生率。由表 2 可知, 子网络模块中差异基因参与大分子代谢负调控过程、基因表达调控和细胞代谢负调控过程等 9 个生物学过程。

表 2 子网络中差异表达基因的基因本体论分析

Tab. 2 Gene ontology analysis of differentially expressed genes from sub-network

Term	Description	Count	FDR
GO:0010605	negative regulation of macromolecule metabolic process	13	0.019 898 13
GO:0010468	regulation of gene expression	17	0.019 900 88
GO:0006351	transcription, DNA-templated	15	0.020 214 02
GO:0031324	negative regulation of cellular metabolic process	13	0.027 334 26
GO:0048511	rhythmic process	6	0.043 049 34
GO:0006355	regulation of transcription, DNA-templated	15	0.043 238 84
GO:1903506	regulation of nucleic acid-templated transcription	15	0.046 263 93
GO:2001141	regulation of RNA biosynthetic process	15	0.048 218 21
GO:0009892	negative regulation of metabolic process	13	0.049 590 37

2.2 结果分析

对于实验所获结果, 细胞周期蛋白 D1 (Cyclin D1, CCND1) 为细胞周期蛋白的编码基因, 参与细胞周期 G1-S 期转换的调控, 目前已被公认为是一种原癌基因^[21]。它的过度表达能够使细胞 G1 期缩短, 促进 G1 期向 S 期过渡^[22], 从而导致 DNA 合成及增殖, 引起细胞增殖失控和癌变。此外, 已有研究表明, Cyclin D1 异常表达与包括结直肠癌在内的多种肿瘤发生密切相关^[23], 例如 Tan 等^[24]发现 CCND1 870A 能够增加结直肠癌的发病风险; Toncheva 等^[25]发现 Cycling D1 在结直肠癌中的蛋白表达阳性率明显高于正常结直肠组织。

c-fos 为早期反应因子, 能够对各种刺激做出迅速的反应。作为原癌基因的 c-fos 参与了细胞增殖、侵袭、凋亡的调控和血管生成、分化和凋亡等过程, 并在肿瘤细胞的运动、细胞外基质的降解、异常粘附及转移灶新生血管生长等多个环节中有着重要作用。此外, 有研究表明, c-fos 异常表达对肿瘤细胞的生长产生重要的影响^[26], 例如调节肿瘤细胞的侵袭性生长^[27]。

EGR1 为早期生长反应因子 1, 是一种对细胞生长起调节作用的转录因子。在细胞的增殖、分化、凋亡、信号传递和细胞生长调控等过程中产生了重要影响, 并且能够促进细胞增生和组织修复。近年来, 有研究表明外源性 Egr-1 基因有抑制肿瘤细胞的生长和恶性转化的作用, 其异常表达可引起细胞增殖与凋亡的失衡, 对乳腺癌、食管癌、胃癌、前列腺癌等癌细胞有着十分重要的影响。

NOS3 为人血管内皮型一氧化氮合酶 (endothelial Nitric oxide synthase, eNOS), 其释放的内皮型一氧化氮在血管生成过程中能够促进血管舒张, 调控血管的生成。事实上, 实体肿瘤的固有特征之一就是异常的血管增生, 而肿瘤细胞或激活的免疫细胞都能产生促血管因子。例如, 巨噬细胞、中性粒细胞、肥大细胞以及骨髓造血祖细胞等参与固有免疫的细胞在血管生成转换的过程中都能产生重要的作用。促进肿瘤细胞的增殖和迁移的 NO/NOS3 信号转导同时可以增强血管通透性、诱导细胞外基质降解以及肿瘤血管新生^[28]。目前, 已有多个研究发现 NOS 3 894G>T 增加了膀胱癌、结直肠癌和乳腺癌的发病

风险^[29]。

转录因子 CEBPB(CCAAT enhancer-binding protein, CCAAT 增强结合蛋白), 参与许多生物过程, 包括细胞分化、代谢平衡、增殖, 肿瘤发生、凋亡以及机体的免疫、应激反应, 能量代谢和血液生成^[30]。有研究表明 C/EBP 在肿瘤发生发展中通过与其他基因相互作用形成调控网络而影响肿瘤发生^[31], 也有研究显示 C/EBP 转录因子的 LIP 亚型可以诱导人乳腺癌细胞凋亡并诱发其自身的吞噬作用, 分析可能是一种诱发肿瘤自我吞噬的抑癌因袭。此外, PLAC1 基因在人类大部分恶性肿瘤中都是异常表达。例如, 在乳腺癌组织中, PLAC1 能够被 C/EBP 激活, 并在乳腺癌的恶性转化过程中产生重要的作用^[32]。另外, 通过检测得到 C/EBP 在肝癌、卵巢癌、淋巴细胞白血病、淋巴瘤、皮肤癌等多种肿瘤中均表达上调, 但其在结直肠癌中具体的表达机制尚不清楚, 需要进一步的深入研究。

综上, 本文获得的结直肠癌致病基因 FOS, CCND1, CEBPB, EGR1 和 NOS3 在结直肠癌的发展过程中起着重要的作用。

3 结束语

本研究通过生物学分析对结直肠癌致病基因进行了识别。首先, 基于 GEO 中 GSE9348 基因表达数据集, 利用 R 语言的 LIMMA 包筛选出 $P < 0.05$, Fold change > 2 的结直肠癌差异基因 339 个, 并从 OMIM 数据库中下载结直肠癌已知的致病基因 135 个; 其次, 将这 339 个差异基因和 135 个已知的致病基因上传到 STRING 数据库中, 获得了 228 个可靠性指数大于 0.9 的基因对, 并通过这些基因对构建了差异表达基因与致病基因的蛋白质互作网络; 进一步, 利用 Cytoscape 软件的 ClusterONE 插件进行了蛋白质互作网络模块分析, 获得了一个含有 53 个基因的子网络; 最后, 通过对子网络的拓扑分析, 获得了 FOS, CCND1, CEBPB, EGR1 和 NOS3 等 5 个新结直肠癌致病基因。同时, 通过功能富集分析和文献挖掘对新发现的致病基因进行了验证, 数据试验结果显示本文的研究方法是行之有效的。本文提供的方法具有通用性, 它对癌症发病机制的阐述以及分子靶向寻找等研究具有重要意义。

参考文献:

- [1] Siegel R L, Miller K D, Jemal A. Cancer statistics[J]. CA: A Cancer Journal for Clinicians, 2015, 65(1): 5-29.
- [2] Chen Q, Liu Z C, Cheng L P. An analysis of incidence and mortality of colorectal cancer in China, 2003—2007 [J]. China Cancer, 2012, 21(3): 179-182.
- [3] World Health Organization. Globocan 2012: Estimated cancer incidence, mortality, and prevalence worldwide in 2012[EB/OL]. Available at http://globocan.iarc.fr/Pages/fact_sheets_Population.aspx, 2014.
- [4] Bradley E W, Ruan M M, Vrable A, et al. Pathway crosstalk between Ras/Raf and PI3K in promotion of M-CSF-induced MEK/ERK-mediated osteoclast survival[J]. Journal of Cellular Biochemistry, 2008, 104(4): 1439-1451.
- [5] Li Y, Tang X Q, Bai Z, et al. Exploring the intrinsic differences among breast tumor subtypes defined using immunohistochemistry markers based on the decision tree[J]. Scientific Reports, 2016, 6: 35773.
- [6] Dai X, Li Y, Bai Z, et al. Molecular portraits revealing the heterogeneity of breast tumor subtypes defined using immunohistochemistry markers[J]. Scientific Reports, 2015, 5: 14499.
- [7] Gerits N, Kostenko S, Shiryaev A, et al. Relations between the mitogen-activated protein kinase and the cAMP-dependent protein kinase pathways: Comradeship and hostility [J]. Cellular Signalling, 2008, 20(9): 1592-1607.
- [8] Pham H, Chong B, Vincenti R, et al. Ang II and EGF synergistically induce COX-2 expression via CREB in intestinal epithelial cells[J]. Journal of Cellular Physiology, 2008, 214(1): 96-109.
- [9] Wang D, Xia D, DuBois R N. The crosstalk of PTGS2 and EGF signaling pathways in colorectal cancer [J]. Cancers, 2011, 3(4): 3894-3908.
- [10] Krysan K, Reckamp K L, Dalwadi H, et al. Prostaglandin E2 activates mitogen-activated protein kinase/Erk pathway signaling and cell proliferation in non-small cell lung cancer cells in an epidermal growth factor receptor-independent manner [J]. Cancer Research, 2005, 65(14): 6275-6281.
- [11] Han J D J, Bertin N, Hao T, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network [J]. Nature, 2004, 430(6995): 88-93.

- [12] 张海燕, 梁循, 周小平. 针对有向图的局部扩展的重叠社区发现算法[J]. 数据采集与处理, 2015, 30(3): 683-693.
Zhang Haiyan, Lang Xun, Zhou Xiaoping. Overlapping community detection from local extension in directed graphs[J]. Journal of Data Acquisition and Processing, 2015, 30(3): 683-693.
- [13] Huang K C, Yang K C, Lin H, et al. Analysis of schizophrenia and hepatocellular carcinoma genetic network with corresponding modularity and pathways: Novel insights to the immune system[J]. BMC Genomics, 2013, 14(5): S10.
- [14] Hindumathi V, Kranthi T, Rao S B, et al. The prediction of candidate genes for cervix related cancer through gene ontology and graph theoretical approach[J]. Molecular BioSystems, 2014, 10(6): 1450-1460.
- [15] Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks[J]. Nature Methods, 2012, 9(5): 471-472.
- [16] Srihari S, Leong H W. A survey of computational methods for protein complex prediction from protein interaction networks [J]. Journal of Bioinformatics and Computational Biology, 2013, 11(2): 1230002.
- [17] Van Landeghem S, De Bodt S, Drebert Z J, et al. The potential of text mining in data integration and network biology for plant research: A case study on Arabidopsis[J]. The Plant Cell, 2013, 25(3): 794-807.
- [18] Albert R. Scale-free networks in cell biology [J]. Journal of Cell Science, 2005, 118(21): 4947-4957.
- [19] Lu C, Hu X, Wang G, et al. Why do essential proteins tend to be clustered in the yeast interactome network[J]. Molecular Biosystems, 2010, 6(5): 871-877.
- [20] Consortium T G O. Gene ontology consortium: Going forward[J]. Nucleic Acids Research, 2015, 43 (Database Issue): 1049-1056.
- [21] Inaba T, Matsushime H, Valentine M, et al. Genomic organization, chromosomal localization, and independent expression of human cyclin D genes [J]. Genomics, 1992, 13(3): 565-574.
- [22] Solomon D A, Wang Y, Fox S R, et al. Cyclin D1 splice variants differential effects on localization, RB phosphorylation, and cellular transformation [J]. Journal of Biological Chemistry, 2003, 278(32): 30339-30347.
- [23] Knudsen K E, Diehl J A, Haiman C A, et al. Cyclin D1: Polymorphism, aberrant splicing and cancer risk [J]. Oncogene, 2006, 25(11): 1620-1628.
- [24] Tan X L, Nieters A, Kropp S, et al. The association of cyclin D1 G870A and E-cadherin C-160A polymorphisms with the risk of colorectal cancer in a case control study and meta-analysis [J]. International Journal of Cancer, 2008, 122(11): 2573-2580.
- [25] Wu M Y, Zhuang C X, Yang H X, et al. Expression of Egr-1, c-fos and cyclin D1 in esophageal cancer and its precursors: An immunohistochemical and in situ hybridization study [J]. World Journal of Gastroenterology, 2004, 10(4): 476-480.
- [26] Hein S, Mahner S, Kanowski C, et al. Expression of Jun and Fos proteins in ovarian tumors of different malignant potential and in ovarian cancer cell lines [J]. Oncology Reports, 2009, 22(1): 177.
- [27] Rupp B, Lorenz U, Schmidt J, et al. Discordant effects of activator protein-1 transcription factor on gene regulation, invasion, and metastasis in spontaneous, radiation-induced, and fos induced osteosarcomas [J]. Molecular Carcinogenesis, 1998, 23(2): 69-75.
- [28] Lim K H, Ancrile B B, Kashatus D F, et al. Tumour maintenance is mediated by eNOS[J]. Nature, 2008, 452(7187): 646-649.
- [29] Jang M J, Jeon Y J, Kim J W, et al. Association of eNOS polymorphisms (-786T> C, 4a4b, 894G> T) with colorectal cancer susceptibility in the Korean population[J]. Gene, 2013, 512(2): 275-281.
- [30] Ramji D P, Pelagia F. CCAAT/enhancer-binding proteins: Structure, function and regulation[J]. Biochemical Journal, 2002, 365(3): 561-575.
- [31] Abreu M M, Sealy L. The C/EBPbeta isoform, liver-inhibitory protein (LIP), induces autophagy in breast cancer cell lines [J]. Experimental Cell Research, 2010, 316(19): 3227-3238.
- [32] Koslowski M, Türeci Ö, Biesterfeld S, et al. Selective activation of trophoblast-specific PLAC1 in breast cancer by CCAAT/enhancer-binding protein β (C/EBP β) isoform 2[J]. Journal of Biological Chemistry, 2009, 284(42): 28607-28615.

作者简介:



吴慧慧(1993-),女,硕士研究生,研究方向:生物信息学, E-mail: 2825084383@qq.com.



唐旭清(1963-),男,通信作者,博士,教授,研究方向:智能计算、生物信息学和生态系统建模与仿真, E-mail: txq5139@jiangnan.edu.cn.