

基于复杂网络的胃癌关键基因筛选与分析

王晓曼 刘文奇

(昆明理工大学理学院, 昆明, 650093)

摘要: 运用复杂网络理论,对TCGA胃癌数据进行了筛选与降维,筛选出275个胃癌相关的基因,获得样本容量为40的胃癌ⅡB期样本组和样本容量为36的胃癌ⅢA期样本组。通过分析胃癌ⅡB期样本组与胃癌ⅢA期样本组的基因变化率,建立节点(基因)间的连边关系,从而构建了胃癌恶化过程的基因表达网络。引入综合中心性指标对网络进行分析,筛选出17个综合中心指数较高的基因。应用复杂网络的相关理论对胃癌基因网络进行社区划分,发现17个综合中心指数较高的基因全部落在一个规模较大的连通的子网络中,此拓扑结构与胃癌基因表达网络的关键节点一致。通过综合分析获取了胃癌恶化过程中的关键基因,提供了良好的胃癌恶化早期的预警信号。

关键词: 复杂网络;胃癌;基因;中心性

中图分类号: TP391 **文献标志码:** A

Screening and Analysis of Key Genes in Gastric Cancer Based on Complex Network

Wang Xiaoman, Liu Wenqi

(Faculty of Science, Kunming University of Science and Technology, Kunming, 650093, China)

Abstract: Through the complex network theory, screening and dimension reduction in accordance with the TCGA (The cancer genome atlas) gastric cancer data are dealt with. We selected 275 genes related to gastric cancer, 40 samples of patients with stage IIB of gastric cancer and 36 samples of patients with stage IIIA of gastric cancer. By analyzing the gene change rate of the gastric cancer IIB sample group and the gastric cancer IIIA sample group, the joint relationship between the nodes (genes) is established. Due to the above work, the gene expression network in the process of gastric cancer deterioration is constructed. The network is analyzed by making uses of comprehensive central indicators, and 17 genes with higher comprehensive index are screened out. Using the relevant theory of complex networks to divide the gastric cancer gene network into communities, it is found that all the genes with higher index of 17 comprehensive centers fall in a large connected sub-network. This topology is consistent with the key nodes of the gastric cancer gene expression network. Therefore we have verified our conclusions on the other hand. Through comprehensive analysis, the key genes in the process of gastric cancer deterioration are obtained, which provided an effective early warning signal for gastric cancer deterioration.

Key words: complex network; gastric cancer; gene; centrality

引言

近些年来,复杂性科学的实践研究迅速发展,分析方法不断更新,技术应用范围日益扩展。对于一个复杂系统,单独研究系统中的个体已经不能反映系统的性质。往往把系统中的个体视为节点,个体之间的关系抽象成边,这样就可以用复杂网络来研究系统的整体性质。目前,复杂网络理论应用的范围越来越广泛,不仅应用到军事^[1]、政治^[2]和交通^[3],且已被用来理解生物系统中的基因功能^[4]。

近些年来,癌症患者的数量逐步上升,中国每年就新增400多万癌症患者。胃癌作为如今最常见的恶性肿瘤之一,发病率仅次于肺癌、乳腺癌、结/直肠癌和前列腺癌。早期胃癌病人的生存率能达到90%,但是大部分胃癌患者确诊时就已经丧失治疗机会。杨维良等^[5]的研究表明胃癌为一种高发、高恶性度、预后不良的恶性肿瘤,长期以来包括手术切除及其他综合性治疗的应用,并没有带来满意的生存率。胃癌本身的生物学过程归根结底也是基因突变产生的,因此基因水平的治疗成为胃癌治疗学取得新突破的希望。如果能应用科学的数据分析方法,找到胃癌恶化过程中的关键基因,就能为癌症患者带来意想不到的收益。

王频等^[6]的研究已经通过RNA测序建立的癌症数据(The cancer genome atlas, TCGA)和Affymetrix芯片(Affymetrix genechip arrays, HG-U133 Plus 2.0)产生的GSE307207这两个数据库中正常胃组织和癌症组织之间的基因表达,筛选出存在显著差异的胃癌相关基因。这样就大量剔除了与胃癌无关的基因,为分析关键基因缩小了范围。

周晖杰^[7]与李星^[8]研究表明,基因与蛋白质很少单独起作用,而是通过网状的相互作用而影响生物表达,因此基因组的研究也从结构基因组转向了功能基因组的研究,通过对基因网络的分析来了解生物系统的功能。Laura等^[9]的研究表明,网络生物学是解释健康与疾病背景下基因组数据的一种系统性方法。其次王华等^[10]已经将复杂网络的分析方法应用到高血压特征基因的筛选中。由于高血压疾病在细胞、组织与生理水平上都表现出复杂性,因此应用网络分析方法来研究高血压疾病。根据基因表达数据的相关性构建网络模型,网络中的枢纽基因显示出与具有不同表达水平的大多数基因相同的变化趋势,并且具有相近的表达模式。所以枢纽基因比其他基因更具代表性,这些枢纽基因可成为病情分析的关键依据。

本文将复杂网络与胃癌测序数据相联系,将胃癌相关基因抽象为节点,依据胃癌ⅡB期样本与胃癌ⅢA期样本间的基因变化率构建胃癌基因表达复杂网络。通过分析该胃癌基因表达网络的相关拓朴性质,发现网络是稀疏的且具有小世界性。通过计算节点的度中心性、介数中心性和紧密度中心性,并引入综合中心性指标,筛选出中心性较高的基因并做出了验证,因此建议了胃癌恶化过程中的枢纽基因,为胃癌恶化提供了良好的预警信号。

1 数据的预处理

本文使用的原始数据来自TCGA数据库,包括7种胃癌分期共408个样本,每个样本包含60 484个基因测序。由于原始数据量非常庞大,应对数据进行初步筛选和降维。

首先使用两个可公开的国际肿瘤数据库:由RNA测序建立的癌症数据TCGA和Affymetrix芯片产生的GSE307207。比较这两个数据库中正常胃组织和癌症组织之间的基因表达,确定了存在显著差异的基因。TCGA和GSE30727分别有688个和3 239个基因达到了标准(2倍变化和校正后 p 值 < 0.05)。TCGA和GSE30727数据集之间存在275个重叠基因,将这275个基因保留,用于后续的分析^[6]。这样就筛选出与胃癌相关的基因,剔除了大量无关基因,为分析关键基因缩小了范围。

其次,原始数据包含胃癌不同分期不同分形共408个样本,样本数量较大。根据癌症TNM分期标准(依据肿瘤浸润的深度、是否存在淋巴结以及是否远处转移。 T 代表肿瘤浸润胃壁的深度, N 表示局

部淋巴结的转移情况, M 则代表有没有远处转移的情况, 可将胃癌分为 I, II, III, IV, 4 个临床病理分期, 每期又具有不同的分形, 本文筛选的样本应用第七版胃癌 TNM 分期标准如表 1 所示。根据大宗临床报道, 实行规范治疗的 1 期胃癌患者 5 年生存率为 82%~95%, 2 期为 55%, 3 期为 15%~30%, 4 期为 2%^[11]。在临床意义上讲, 当胃癌处于 II B 期时, 肿瘤未侵及胃的外部或者侵及胃的外部但是肿瘤未转移。处于此时的胃癌患者仍属于传统意义上的胃癌前期, 有很大的治愈希望。当胃癌处于 III A 期时, 肿瘤侵及腹腔脏层腹膜或侵及到胃的外部^[12]。处于这阶段的胃癌患者, 肿瘤已经侵及胃的外部并发生淋巴结转移, 在临床上已经发展为胃癌的中期, 治愈希望较低(表 2)。因此筛选处于胃癌 II B 期病人的基因序列(40 个样本)与处于胃癌 III A 期病人的基因序列(36 个样本)做为对照, 分析胃癌从 II B 期到 III A 期转变过程中的关键基因, 致力于为胃癌恶化提供良好的预警信号。

2 胃癌基因表达网络模型

本文选取了两个样本组的基因表达数据, 分别是处于 II B 期的胃癌患者, 共 40 个样本, 记为 H ; 以及处于 III A 期的胃癌患者, 共 36 个样本, 记为 S 。通过构建胃癌基因复杂网络模型, 来研究胃癌基因表达数据的变换趋势与胃癌表型之间的关系。

2.1 基胃癌计算胃癌基因表达数据变化率

选取第 2 部分中初步筛选出的 275 个胃癌相关基因, 将它们定义为网络节点, 网络规模 $N=275$ 。其中每 1 个节点都代表 1 个基因, 用基因名称进行标注。首先, 对于每个节点 i , 定义胃癌 II B 期样本到胃癌 III A 期样本的基因变化率为

$$R_i = \frac{\left| \frac{\sum S_i}{s} - \frac{\sum H_i}{h} \right|}{\frac{\sum H_i}{h}} \quad (1)$$

$$T = \frac{\sum_{i=1}^{275} R_i}{275} \quad (2)$$

式中: $i=1, 2, \dots, 275, S=36, h=40$ 。接下来计算出每个节点对应的基因变化率 R_i , 通过式(2)很容易得到基因变化率的平均值 $T=0.3617$ 。这个平均值可以为下一步的阈值选取作为参考。

2.2 基因复杂网络的连边规则

由于胃癌基因变化率的平均值较大, 应选择一个较小的阈值, 增大基因间的关联性, 降低节点间的扰动对网络的影响。通过多次实验, 选定 $0.008T$ 作为阈值, 此时网络结构较为稳定, 基因间的关联性较强, 节点间的扰动较小^[13]。建立连边的规则如下: 比较节点 i 与节点 j 的基因变化率差值, 当差值小于 $0.008T$ 时, 两个节点间的变化趋势相近, 则在这两个节点间建立一条边。式(3)为胃癌基因复杂网络连边规则的数学表达

$$A_{ij} = \begin{cases} 1 & |R_i - R_j| \leq 0.008T \\ 0 & |R_i - R_j| > 0.008T \end{cases} \quad (3)$$

应用上述思路, 构建了 1 个包含 275 个节点(基因)和 828 条边的胃癌基因表达网络。通过 Pajek 软件将该网络可视化。图 1 为胃癌基因表达网络示意图, 图中间是网络的密集部分, 有较高的连接数, 图 2

表 1 第 7 版胃癌 TNM 分期分型

Tab. 1 Seventh edition gastric cancer TNM staging classification

类别	N_0	N_1	N_2	N_3	任何 N, M_1
T_1	I A	I B	II A	II B	IV
T_2	I B	II A	II B	III A	IV
T_3	II A	II B	III A	III B	IV
T_{4a}	II B	III A	III B	III C	IV
T_{4b}	III B	III B	III C	III C	IV
任何 T, M_1	IV	IV	IV	IV	IV

表 2 胃癌不同分期的临床表现

Tab. 2 Clinical manifestations of different stages of gastric cancer

胃癌分期	临床表现
I A 期	肿瘤仅侵及粘膜层。
I B 期	肿瘤侵及肌层,或是肿瘤侵及粘膜层以及 1~2 个局部淋巴结。
II A 期	肿瘤侵及粘膜层及 3~6 个局部淋巴结;或是侵及肌层以及 1~2 个局部淋巴结;或是浸润至胃的外层(浆膜层),无临近组织或淋巴结侵犯。
II B 期	肿瘤侵及粘膜层及 7 个局部淋巴结;或是侵及肌层以及 3 至 6 个局部淋巴结;或是浸润至胃的外层(浆膜层)及 1~2 个淋巴结;或是侵及脏层腹膜,无临近组织或淋巴结侵犯。在临床意义上讲,肿瘤未侵及胃的外部或者侵及胃的外部但是肿瘤未转移。
III A 期	肿瘤侵及腹腔脏层腹膜以及 1~2 个局部淋巴结;或是侵及胃的外部以及 3~6 个局部淋巴结;或是侵及胃的肌层及 7 个以上淋巴结。
III B 期	肿瘤侵及邻近组织,及少于 2 个局部淋巴结;或是侵及腹腔脏层膜以及 3~6 个局部淋巴结;或是浸润至胃的外层(浆膜层)及大于 7 个局部淋巴结。
III C 期	肿瘤侵及邻近组织,及大于 3 个局部淋巴结;或是侵及腹腔脏层腹膜以及大于 7 个局部淋巴结。
IV 期	肿瘤远处转移,如转移到肝或肺。

是网络密集部分的放大图。

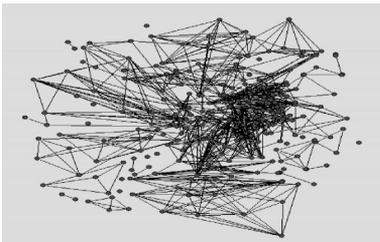


图 1 胃癌基因表达网络示意图

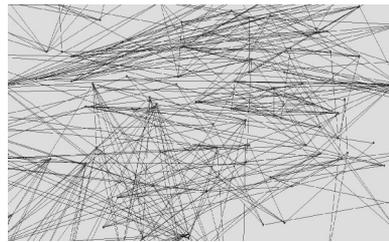


图 2 胃癌基因网络示意图密集部分放大图

Fig.1 Schematic diagram of gastric cancer gene network Fig.2 Enlarged view of the gastric cancer gene network

下面对该胃癌基因复杂网络的相关拓扑性质进行简单分析。

该胃癌基因表达网络包含 275 个节点,假设该网络是一个完全连通网络,则网络的最大连通度为

$$\frac{C_{275}^2}{275} = 137 \tag{4}$$

然而,该网络共有 828 条边,平均度为 6.02,其平均度远远小于该网络的最大连通度,故可以认为该胃癌基因表达网络是稀疏的。这表明,每个基因的表达趋势可能与少量的基因表达数据变化趋势一致,也就是说,每个基因受其他基因调控的平均数量较小。

平均路径长度是指网络中所有顶点对之间最短路径的平均值,可通过平均路径衡量网络中不同基因数据变化的差异^[14]。通过计算,该胃癌基因表达网络的平均路径长度约为 9.48,平均路径较小,有

$$C = \frac{1}{N} \sum_{i=1}^N C_i \tag{5}$$

通过式(5),可计算得到该胃癌基因表达网络的聚类系数 $C = 0.641$ 。该网络的聚类系数约是同规模随机网络的 28 倍。综上所述,胃癌基因表达网络具有较大的聚类系数与较小的平均路径长度,满足

小世界特性。

3 胃癌关键基因

构建出胃癌基因表达网络后,发现该网络具有稀疏性和小世界特性,这意味着平均每个基因可能受较少的基因调控,这些起到关键作用的基因可能具有较强的中心性。那如果能找到基因表达网络中的关键基因,就能为胃癌恶化提供良好的预警信号。下面将分析胃癌基因复杂网络中节点的3种中心性,包括介数中心性、紧密度中心性和度中心性,应用定量的方法对每个节点的中心度进行描述,进而确定胃癌基因表达网络中的关键基因。

$$\text{度中心性}^{[15]} \quad C_d = \frac{k_i}{N-1} \quad (6)$$

式中: k_i 为节点*i*的度, N 为网络节点总数。度中心性定义为节点的度与网络中其他节点总数的比值。通过度中心性可以迅速找到该网络中的关键节点。

$$\text{介数中心性}^{[15]} \quad C_b(i) = \frac{\sum_{j<k} \sum_k g_{jk}(i)}{g_{jk}} \quad (7)$$

式中: $g_{jk}(i)$ 为节点*j*与节点*k*的最短路径经过节点*i*的数量, g_{jk} 为节点*j*到节点*i*的最短路径数。介数中心性定义为节点*i*与节点*k*之间经过节点*i*的最短路径数与节点*i*与节点*k*之间的最短路径总数之比。介数中心性衡量的是经过该节点最短路径数的度量。通过介数中心性可以迅速找到该网络中的重要节点。

$$\text{紧密度中心性}^{[15]} \quad C_c(i) = (L_i)^{-1} = \frac{N-1}{\sum_{j=1}^N d_{ij}} \quad (8)$$

式中: L_i 节点*i*到网络中每个节点距离的平均值。紧密度中心性定义为节点*i*到网络中每个节点距离的平均值的倒数。一般观点认为,网络的拓扑中心是网络中其他节点到该节点距离最小的节点,因此,紧密度中心性可以作为复杂网络中的中心性度量。

分别应用式(1-3)对度中心指标、介数中心性指标和紧密度中心性指标进行计算,各中心指标结果靠前的基因分别列在了表3中。由于有多个基因的某个中心性指标相同,所以选定度中心性排名前32的基因、介数中心性排名前20的基因以及紧密度中心性排名前22的基因列在表3中。

从表3中可以得出,度中心性大的节点往往它们的紧密度中心性也较大,介数中心性大的节点其紧密度中心性也是相对较大的。但是介数中心性的值普遍较小,如果直接分析不能发挥介数中心性的重要性。

图3给出了基于度中心性、介数中心性、紧密度中心性的重要胃癌基因的分布示意图,可以发现,3种中心性指标均靠前的是MMP11基因。但是由于中心性指标间的差异性,并没有太多的重叠基因。

图4绘出了胃癌基因度中心性、介数中心性和紧密度中心性3种指标分布的三维图,从图4中可以看出同时具有较高的3种中心性指标的节点应分布于图的最上部,这些节点就有可能是网络的枢纽节点。

引用加权分析的方法将这3种指标定量化,来计算每个节点的综合中心性指标。根据式(9),将度中心性、介数中心性、紧密度中心性分别归一化为一个相对性中心性指标: $C_a(i)/C_{a,\max}$, $C_b(i)/C_{b,\max}$ 和

表3 胃癌度中心性、介数中心性、紧密度中心和综合中心性指标靠前的基因及数值
Tab. 3 Top genes and values for three central and comprehensive central indicators of gastric cancer

基因	度中心性	基因	介数中心性	基因	紧密度中心性	基因	综合中心性
IFITM1	0.051 095	TMEM63A	0.058 306 8	BCL2A1	0.055 775 803	MMP11	0.938 959 96
SLC6A6	0.051 095	MMP11	0.058 130 2	DDX18	0.055 775 803	TMEM63A	0.873 940 5
ALDH6A1	0.047 445	CKAP2	0.049 548 6	BGN	0.055 775 803	CKAP2	0.852 904 27
TIMP1	0.047 445	INTS8	0.048 608 3	ADH1C	0.055 715 958	IFITM1	0.832 668 51
RUNX1	0.043 796	AKR7A3	0.046 362 4	CYP4F12	0.055 715 958	NOP2	0.786 846 76
CKAP2	0.043 796	NOP2	0.046 322 3	PTGS1	0.055 715 958	ADAT1	0.786 846 76
PAK1IP1	0.043 796	ADAT1	0.046 322 3	PSMD3	0.055 715 958	ACTL6A	0.785 452 09
ECT2	0.043 796	ACTL6A	0.046 202	NOP2	0.055 477 855	FERMT1	0.785 452 09
MAD2L1	0.043 796	FERMT1	0.046 202	ADAT1	0.055 477 855	ECT2	0.774 811 94
MMP11	0.043 796	TNFRSF12A	0.043 887 7	ACTL6A	0.055 359 566	SLC6A6	0.764 381 82
SFRP4	0.043 796	ECT2	0.042 402 7	FERMT1	0.055 359 566	TNFRSF12A	0.751 470 49
CKB	0.040 146	CHORDC1	0.041 672	COL6A3	0.054 775 604	CHORDC1	0.750 915 62
ECHDC2	0.040 146	COL6A3	0.038 807 2	YEATS2	0.054 775 604	COL6A3	0.739 686 93
OSGIN1	0.040 146	YEATS2	0.038 807 2	TMEM63A	0.054 602 81	YEATS2	0.739 686 93
KIF11	0.040 146	IFITM1	0.032 951 8	ENTPD5	0.054 317 231	NCL	0.726 688 19
MCM2	0.040 146	AKR1B10	0.032 879 2	LY6E	0.054 260 473	INTS8	0.724 980 65
NOP56	0.040 146	ENTPD5	0.031 626 6	MMP11	0.053 699 351	SGSM3	0.705 809 87
OSMR	0.040 146	BCL2A1	0.030 480 5	SGSM3	0.053 368 215	LIMK1	0.704 915 42
XAF1	0.040 146	DDX18	0.030 480 5	NCL	0.053 313 422	ALDH6A1	0.699 946 03
HSP90AA1	0.040 146	BGN	0.030 480 5	COL8A1	0.053 204 173		
LIMK1	0.040 146			KAT2A	0.053 204 173		
LBR	0.040 146			DCBLD1	0.053 204 173		
RHBDF2	0.040 146						
GART	0.040 146						
UBFD1	0.040 146						
ZNF146	0.040 146						
GTPBP4	0.040 146						
HEATR1	0.040 146						
NCL	0.040 146						
CDC25B	0.040 146						
MEST	0.040 146						
SDS	0.040 146						
SNX10	0.040 146						

$C_c(i)/C_{c,max}$,将综合中心性指标定义为3个中心性指标的平均值^[10]。筛选出综合中心指标前20的基因,将其列在表3当中。

$$C_{integr}(i) = \frac{1}{3} [C_a(i)/C_{a,max} + C_b(i)/C_{b,max} + C_c(i)/C_{c,max}] \quad (9)$$

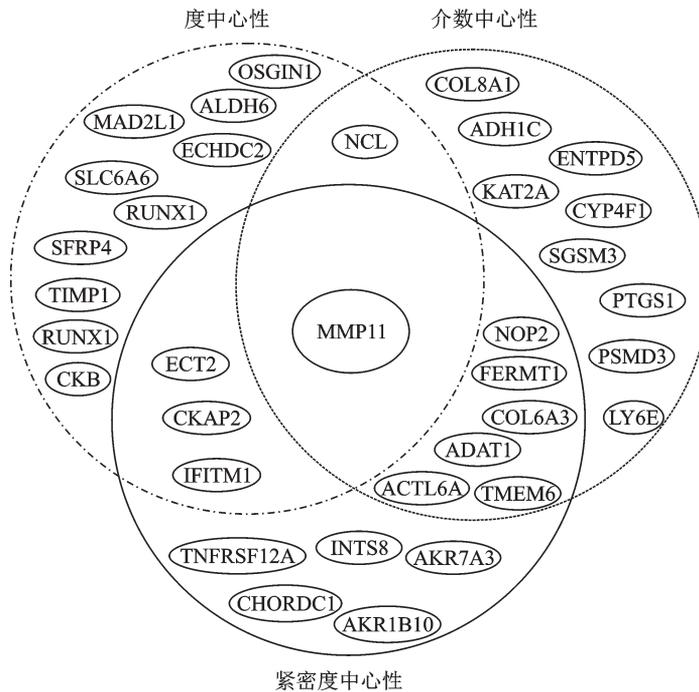


图3 基于3种中心性指标的重要胃癌基因分布图

Fig.3 Gene map of important gastric cancer based on three central indicators

通过对表3中的4种中心性指标进行分析,发现,综合性中心指标排名第1的基因MMP11,其3种中心性指标均是较高的。由于度中心性差值较小,其他综合性中心指数较高的基因其度中心性不一定排名靠前,但是相对拥有较高的介数中心性和紧密度中心性,这并不意味着其度中心性是较低的。通过对表3进行分析发现基因INTS8到基因SGSM3的综合性中心性指标下降过快,选定综合性指标 $C_{intgr} > 0.72$ 的16个基因作为关键基因,此外,由于ALDH6A1基因的度中心较高(排名第3位),在基因表达网络中拥有较高的度,连接着较多的节点,因此也应考虑其为关键基因。应说明的是,选择基因是否为关键基因的阈值可大可小,若想增大或减小关键基因的数量,应将阈值调小或调大。

将该胃癌基因复杂网络进行分区视图,如图5所示,文献[16]指出发现度中心性较小的节点往往分布更加广泛。而大多数基因都集合在一个完全连通的子网络中,在此子网络中,中心性较高的基因链接着更多的基因,在基因调控网络中处于关键地位。通过式(9)选取的中心度较高的17个基因全部落在该子网络中,这与网络中关键基因定量考虑的结果一致。

因此可以确定出胃癌基因表达网络中的17个重要的枢纽基因,分别是:MMP11, TMEM63A, CKAP2, IFITM1, NOP2, ADAT1, ACTL6A, FERMT1, ECT2, SLC6A6, TNFRSF12A, CHORDC1, COL6A3, YEATS2, NCL, INTS8和ALDH6A1。

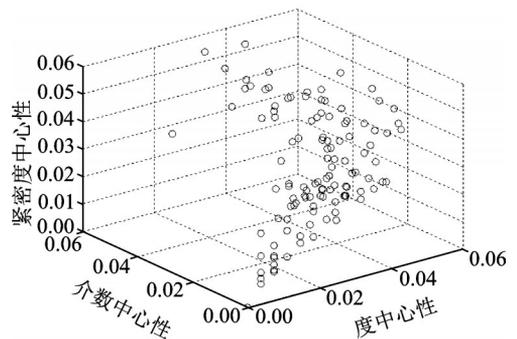


图4 度中心性、介数中心性和紧密度中心性分布的三维图

Fig.4 Three-dimensional map of three central distributions

依据胃癌基因在ⅡB与ⅢA样本间的表达变化率构建胃癌基因复杂网络,将每个基因抽象为节点,分析了该网络的拓扑特征。通过Pajek软件对基因的3种中心性指标进行计算,进一步引入综合中心性指标,进而筛选出上述17个关键的胃癌基因。

4 结束语

本文对TCGA胃癌数据进行初步的筛选和降维,筛选出与胃癌相关的275个胃癌基因,为分析关键基因缩小了范围。原始数据包包含胃癌不同分期不同分形共408个样本,样本数量较大。根据胃癌TNM分期,筛选胃癌ⅡB期(40个样本)与ⅢA期(36个样本)做为对照,分析胃癌从ⅡB期到ⅢA期转变的关键基因,致力于发现胃癌恶化过程中的特征基因。其次应用复杂网络的方法,将网络分析与胃癌基因表达数据结合起来。将胃癌基因抽象为节点,依据两组样本组间基因的变化率,选择合理的阈值来建立连边关系。然后分析了该网络的拓扑特征,发现胃癌基因表达网络是稀疏的,具有小世界特性。计算了该复杂网络节点的3种中心度指标,分别是度中心性、介数中心性和紧密度中心性,然后引入综合中心性指标,筛选出17个中心指标较高的基因。最后对该胃癌基因复杂网络进行社区划分,发现这17个中心度较高的基因都在一个规模较大的连通子网络中,进而验证了结论。因此建议了胃癌恶化过程中的关键基因,为胃癌恶化提供了良好的预警信号。

本文研究的问题是基于复杂网络理论和数据分析方法,筛选出胃癌恶化过程中17个的关键基因,但基因间的链路关系还需进一步的研究。另外,本文是基于病例数据做出的筛选与分析,缺乏动物实验模型验证,这需要继续研究学习。

参考文献:

- [1] 徐玉章,朱磊,张力.基于复杂网络理论的军事通信网信息传播模型[J].军事通信技术,2015,36(2):40-44.
Xu Yuzhang, Zhu Lei, Zhang Li. Information communication model of military communication network based on complex network theory[J]. Military Communications Technology, 2015, 36(2): 40-44.
- [2] 刘普.政治安全:网络时代的挑战与对策[D].北京:中国社会科学院研究生院,2012.
Liu Pu. Political security: Challenges and countermeasures in the network age [D]. Beijing:Graduate School of Chinese Academy of Social Sciences, 2012.
- [3] 吴建军,尹浩东,郭欣,等.基于复杂网络的城市轨道交通末班车衔接研究[J].电子科技大学学报(社科版),2018,20(5):48-53.
Wu Jianjun, Yin Haodong, Guo Xin, et al. Research on the connection of the last train of urban rail transit based on complex network[J]. Journal of University of Electronic Science and Technology of China(Social Sciences Edition), 2018, 20(5):48-53.
- [4] Diambra L. Coarse-grain reconstruction of genetic networks from expression levels[J]. Physica A: Statistical Mechanics and Its Applications, 2011, 390(11): 2198-2207.
- [5] 杨维良,张新晨.胃癌基因治疗的现状及展望[J].中华实验外科杂志,2004(5):128-129.
Yang Weiliang, Zhang Xinchun. Current status and prospects of gene therapy for gastric cancer[J]. Chinese Journal of Experimental Surgery, 2004(5): 128-129.
- [6] 王频.胃癌基因生物标志物的调控和功能及其临床相关研究[D].南京:南京医科大学,2017.
Wang Pin. The regulation and function of gastric cancer gene biomarkers and its clinical correlation research [D]. Nanjing:Nanjing Medical University, 2017.
- [7] 周晖杰.复杂网络理论在基因调控网络中的应用[J].重庆科技学院学报(自然科学版),2009,11(5):141-144.
Zhou Huijie. Application of complex network theory in gene regulatory networks[J]. Journal of Chongqing University of Science and Technology (Natural Science Edition), 2009, 11(5):141-144.
- [8] 李星.基于复杂网络的症状基因预测方法研究[D].北京:北京交通大学,2014.

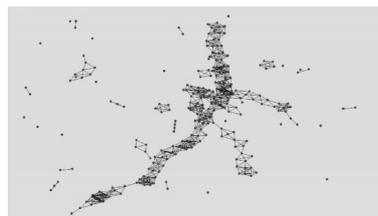


图5 胃癌基因复杂网络社区结构图
Fig.5 Gastric cancer gene complex network community structure

- Li Xing. Research on symptom gene prediction method based on complex network [D]. Beijing: Beijing Jiaotong University, 2014.
- [9] Laura I F. Human diseases through the lens of network biology[J]. *Trends in Genetics*, 2013, 29(3): 150-159.
- [10] 王欢. 基于复杂网络理论的高血压相关基因分析[D]. 昆明: 云南大学, 2013.
Wang Huan. Hypertension-related gene analysis based on complex network theory [D]. Kunming: Yunnan University, 2013.
- [11] 邹文斌. 中国胃癌发病率及死亡率研究进展[J]. *中国实用内科杂志*, 2014, 34 (4): 408-415.
Zou Wenbin. Research progress in gastric cancer incidence and mortality in China[J]. *Chinese Journal of Practical Internal Medicine*, 2014, 34 (4): 408-415.
- [12] 肖恒华, 宁文锋. 胃癌分子分型的研究进展[J]. *中国现代医生*, 2018, 56(6):164-168.
Xiao Henghua, Ning Wenfeng. Research progress on molecular typing of gastric cancer[J]. *Chinese modern doctor*, 2018, 56(6): 164-168.
- [13] 汪小帆, 李翔, 陈关荣. 复杂网络理论及其应用[M]. 北京: 清华大学出版社, 2006:129-133.
Wang Xiaofan, Li Xiang, Chen Guanrong. Complex network theory and its application [M]. Beijing: Tsinghua University Press, 2006: 129-133.
- [14] 何大初, 刘宗华, 汪秉宏. 复杂系统与复杂网络[M]. 北京: 高等教育出版社, 2009: 155-157.
He Dali, Liu Zonghua, Wang Binghong. Complex systems and complex networks [M]. Beijing: Higher Education Press, 2009: 155-157.
- [15] Newman M E J. 网络科学引论[M]. 北京: 电子工业出版社, 2014: 106-109, 115-117.
Newman M E J. Introduction to network science [M]. Beijing: Publishing House of Electronics Industry, 2014: 106-109, 115-117.
- [16] 史定华. 网络度分布理论[M]. 陈关荣主编. 网络科学与工程丛书1. 北京: 高等教育出版社, 2011:87-113.
Shi Dinghua. Network degree distribution theory [M]. Chen Guanrong editor. Network Science and Engineering Series 1. Beijing: Higher Education Press, 2011:87-113.

作者简介:



王晓曼(1995-),女,硕士研究生,研究方向:复杂网络、机器学习等, E-mail: 957216415@qq.com。



刘文奇(1965-),男,教授,博士生导师,研究方向:复杂大系统数据共享与分析,数据挖掘等。

(编辑:刘彦东)