

文章编号:1000-5862(2020)05-0495-06

# 基于随机矩阵理论分析乳腺癌基因网络

李 蓉<sup>1</sup>, 郑 浪<sup>2</sup>, 任喜梅<sup>1</sup>, 刘超飞<sup>3</sup>

(1. 华东交通大学理工学院, 江西 南昌 330100; 2. 江西省教育考试院, 江西 南昌 330038; 3. 江西理工大学理学院物理电子系, 江西 赣州 341000)

**摘要:**利用随机矩阵理论分析乳腺癌基因微阵列数据,得到乳腺癌基因共表达网络,找出乳腺癌基因共表达网络中重要的增殖模块和免疫模块,并预测基因 PMSCL1 与乳腺癌细胞的增殖、侵袭及迁移有关,基因 CCAN2 与乳腺癌细胞的有丝分裂有关,基因 SCYA5 与乳腺癌细胞的免疫应答有关,基因 PRC1、RAB31、INHBA 可作为乳腺癌的靶向基因。

**关键词:**随机矩阵理论;乳腺癌;基因共表达网络;模块;靶向基因

**中图分类号:**O 411;Q 6-33;R 73 **文献标志码:**A **DOI:**10.16357/j.cnki.issn1000-5862.2020.05.08

## 0 引言

乳腺癌是女性最常发生的恶性肿瘤之一。近年来,中国乳腺癌发病率快速增长,成为女性发病率最高的癌症<sup>[1-2]</sup>。乳腺癌的发生是由抑癌基因失活或原癌基因激活为原因的多步骤过程,找出乳腺癌基因相互作用网络,对乳腺癌患者的治疗具有重要的意义。

关于癌细胞分子相互作用网络的实验和理论研究较多,一般认为癌细胞分子网络间的相互作用发生改变会激发肿瘤的产生。微阵列技术能够同时对成千上万个基因的表达式进行测量,其为生物网络的获得提供了可能性<sup>[3-5]</sup>。然而微阵列数据中的噪声会影响数据分析的结果。为了把生物真实网络模块从随机噪声中区分开来,近年来有科学家开始把随机矩阵理论(random matrix theory, RMT)应用到生物学领域中。人们把随机矩阵理论方法应用在酵母及拟南芥等生物体的代谢网络、蛋白质网络的研究中<sup>[6-8]</sup>,并成功地鉴别出代谢网络模块和蛋白质网络模块。RMT 现已被应用于人的肿瘤细胞网络的研究中<sup>[9-10]</sup>,笔者之前利用随机矩阵理论和层次聚类方法成功地研究了肝癌基因网络<sup>[11]</sup>,本文将采用随机矩阵理论结合拓扑结构分析法来研究乳腺癌基因网络,试图鉴别出乳腺癌基因共表达网络的模块,根据模块的功能特点预测与乳腺癌发生发展有关的基因。

## 1 随机矩阵理论

随机矩阵理论由维格纳、戴森在 20 世纪 60 年代提出,最初运用在复杂原子核光谱的研究中,是统计物理和材料科学中处理由噪声和无序导致的复杂系统相变非常有效的方法。到目前为止,它被广泛地应用于物理、数学、金融等研究领域<sup>[12-15]</sup>。目前,随机矩阵理论也逐步在生物学的各个领域开展研究工作。

随机矩阵理论的主要思想是实对称随机矩阵的特征值步长分布遵循 2 个规律,最近邻间隔分布(the nearest neighbor spacing distributions, NNSDs)的特征值(即 2 个最邻近的特征值之差的分布)之间具有相关性则系统遵循高斯正交系综(gauss orthogonal ensemble, GOE)统计,如果没有相关性,则系统遵循 Poisson 统计<sup>[16]</sup>。随机矩阵中的非零非对角元素表示了对角元素之间的相互作用,使得特征值之间具有较强的相关性,从而得到了 GOE 统计量。相反,矩阵特征值的间距分布是随机的只有对角或块对角部分非零值,则遵循 Poisson 统计量,因对角或块对角部分之间没有相互作用,该系统的特征值不相关。利用乳腺癌微阵列数据构建乳腺癌基因共表达相关矩阵(co-expression correlation matrices, CCMs)。由于基因共表达网络的模块化,在逐渐去除相关系数较低的值后,乳腺癌基因 CCMs 只对与基

收稿日期:2019-09-30

基金项目:国家自然科学基金面上课题(11875149),国家自然科学基金地区基金(61565007),江西省教育厅科学技术研究课题(GJJ181485)和江西省高等学校教学改革研究课题(JXJG-18-35-1)资助项目。

作者简介:李 蓉(1983-),女,湖南新宁人,讲师,主要从事生物信息网络研究。E-mail:14164121@qq.com

因模块对应的块对角部分具有非零元素. 由 RMT 可看出, 乳腺癌基因 CCMs 中的特征值间距分布经历了从 GOE 统计量到 Poisson 统计量的转变. 笔者研究发现乳腺癌基因网络的光谱波动可以用 GOE 统计量来描述. 此外, 虽然这些全局网络属于 GOE 统计, 但去除 CCMs 中相关系数的小值后会导致乳腺癌基因网络的光谱波动从 GOE 统计量到 Poisson 统计量的转换, 以及基因共表达网络从全局网络到孤立模块网络的分离. 这种急剧转变为全球生物网络中功能模块的识别提供了一种新的客观方法. 本文利用标准光谱展开技术研究乳腺癌基因 CCMs 的特征值间距分布. 在一般情况下, 矩阵的特征值密度是特征值  $E_i (i = 1, 2, \dots, N)$  的函数 ( $N$  是矩阵维数). 因此, 特征值间距分布是  $E_i$  的函数, 且与系统有关. 为了观测不同实对称矩阵的特征值波动函数, RMT 要求光谱展开具有恒定的特征值密度. 因此可用展开谱  $e_i$  代替特征值  $E_i$ , 其中  $e_i = Nav(E_i)$ ,  $Nav$  是通过拟合原始积分获得的特征值的平滑积分密度为 3 次样条或局部平均密度. 采用展开的特征值, 本文计算了乳腺癌基因共表达相关矩阵特征值最近邻间隔分布函数  $P(s)$ , 定义  $s = e_{i+1} - e_i$ . 由 RMT 可知, GOE 统计的  $P(s)$  近似服从 Wigner-Dyson 分布,  $P_{GOE}(s) \approx (\pi s \exp(-\pi s^2/4))/2$ . 计算结果如图 1 所示.

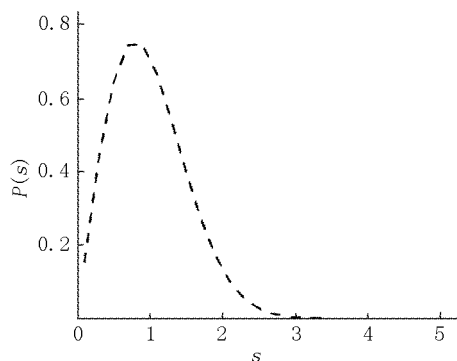


图 1 Wigner-Dyson 分布曲线

Poisson 统计的  $P(s)$  服从 Poisson 分布,  $P_{poisson} = \exp(-s)$ , 计算结果如图 2 所示.

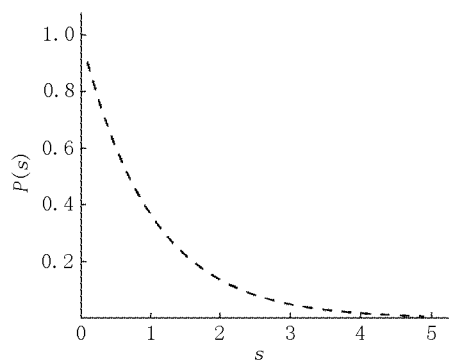


图 2 Poisson 分布曲线

Wigner-Dyson 分布和 Poisson 分布明显的差异是: 当  $s$  趋近于 0 时, Wigner-Dyson 分布中特征值的最近邻间隔分布趋近于 0, Poisson 分布中特征值的最近邻间隔分布趋近于 1.

## 2 构建乳腺癌基因共表达网络

研究的乳腺癌基因数据来自斯坦福大学对外发表的乳腺癌基因微阵列数据, 该数据由 3 044 个乳腺癌基因在 77 个乳腺癌样本条件下的表达量值组成. 下载的网址为 [http://genome-www.stanford.edu/breast\\_cancer/mopo\\_clinical/data.shtml](http://genome-www.stanford.edu/breast_cancer/mopo_clinical/data.shtml).

### 2.1 确定乳腺癌共表达网络模块分割阈值

由于基因在同一功能模块中的表达通常高度相关, 因此可以采用基因共表达网络方法来揭示基因功能模块. 由于背景噪声对微阵列数据的影响, 目前人们难以得到可信的基因网络结构. 为了从包含大量噪声的数据中提取更多的生物信息, 人们提出了  $K$  均值、层次聚类及自组织图等几种方法. 现有的分类方法不能自治地确定共表达网络临界阈值, 不能区分真实的模块网络和噪声引起的错误网络. 原则上, 由随机噪声引起的基因间相互作用弱于基因组模块间的相互作用, 可以通过增加阈值来消除随机噪声. 当消除噪声后, 基因表达网络从相互完全连接的网络变成了一个由具有独立功能模块的网络. 基于统计物理中对处理由噪声和无序导致的复杂系统相变非常有效的随机矩阵理论方法, 可以通过逐步消除基因间随机噪声, 网络系统的能级分布将发生从 GOE 统计到泊松统计的转变, 基因网络解体成为许多分立模块. 因此, 可以通过分析网络的能级统计分布来识别功能模块. 基因网络将完成从相互完全连接的网络到真正生物网络的转变, 从而阈值和基因共表达网络将自治地由随机矩阵理论给出.

利用乳腺癌微阵列数据构建乳腺癌基因共表达相关矩阵, 该相关矩阵通过计算乳腺癌基因间的皮尔逊相关系数而得到. 其中乳腺癌基因  $g_i$  与  $g_j$  间的相关系数由皮尔逊公式计算得到.

$$c(g_i, g_j) = \frac{1}{N} \sum_{k=1}^N ((g_{ik} - \bar{g}_i) / \sigma_{g_i}) ((g_{jk} - \bar{g}_j) / \sigma_{g_j}),$$

其中  $\sigma_{g_i}$  和  $\sigma_{g_j}$  代表基因  $g_i$  和  $g_j$  的标准误差;  $\bar{g}_i$ 、 $\bar{g}_j$  分别代表基因  $g_i$  和  $g_j$  的平均值;  $N$  为总基因个数.

设定去噪步长  $\Delta q = 0.01$ , 将去噪因子  $q$  由 0 逐步增加到 1, 计算每个去噪因子  $q$  值对应特征值的最近邻间隔分布. 发现随着  $q$  值的增加, 乳腺癌关联矩

阵特征值的最近邻间隔分布函数  $P(s)$  的分布曲线发生尖锐转变,由 Wigner-Dyson 分布锐变到 Poisson 分布,如图3所示.这说明通过增加  $q$  值,随机噪声引入的模块间弱相互作用逐渐被去除掉,体系保留下来的是基因间真实的强相互作用,代表乳腺癌基因网络从无规全局网络过渡到真实信息的生物网络.

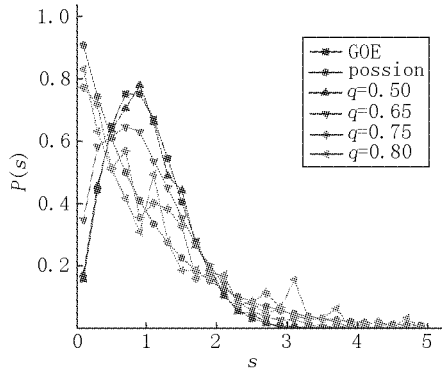


图3  $P(s)$  分布随  $q$  值变化曲线

通过计算乳腺癌关联矩阵特征值 NNDSs 对高斯分布的标准误差与对泊松分布的标准误差来确定乳腺癌基因共表达网络模块的分割阈值,标准误差的计算公式为

$$S_{D_{GOE}}(q) = \sqrt{\left( \sum_{i=1}^m (p(i) - P_{GOE}(i))^2 \right) / (m-1)},$$

$$S_{D_{Poisson}}(q) = \sqrt{\left( \sum_{i=1}^m (p(i) - P_{Poisson}(i))^2 \right) / (m-1)},$$

其中  $S_{D_{GOE}}(q)$  是乳腺癌关联矩阵特征值的 NNDSs 对高斯分布的标准误差;  $S_{D_{Poisson}}(q)$  为乳腺癌关联矩阵特征值的 NNDSs 对泊松分布的标准误差.  $P_{GOE}(i)$ 、 $P_{Poisson}(i)$  是第  $i$  个点对应高斯分布和泊松分布特征值的 NNDSs;  $p(i)$  是第  $i$  个点特征值

的 NNDSs.

通过作乳腺癌关联矩阵特征值 NNDSs 对高斯分布的标准误差与对泊松分布的标准误差的比值曲线图(见图4)可见,当  $q=0.78$  时标准误差的比值最大.此时系统完全过渡到泊松系统,基因模块间由噪声引入的弱相互作用被充分去掉,乳腺癌基因网络呈现基因间本质网络模块.因此,选取  $q=0.78$  为  $3\,044 \times 77$  的乳腺癌基因共表达网络模块的分割阈值.

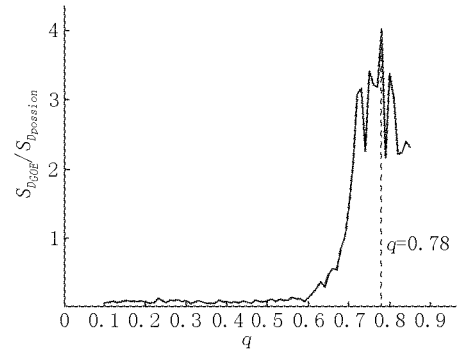


图4 NNDSs 的标准误差比值曲线

## 2.2 获取乳腺癌共表达网络模块

本文使用 BioLayout Express 3D 软件来可视化不同去噪因子对应的乳腺癌基因共表达网络结构的变化,结果如图5所示.模块中每个基因用一个节点来表示,节点间若用连线相连则代表基因间存在非零关联,零关联表示该模块中基因间的关联作用小于  $q$  值<sup>[17]</sup>.由图5可见,随着去噪因子  $q$  的增加,乳腺癌基因共表达网络从全局网络逐渐变化到孤立网络模块.当  $q=0.78$  时,独立的模块明显呈现,表示随着随机噪声的移除,乳腺癌基因真实的共表达网络呈现出来.本文研究的乳腺癌共表达网络正是阈值  $q=0.78$  的对应网络.

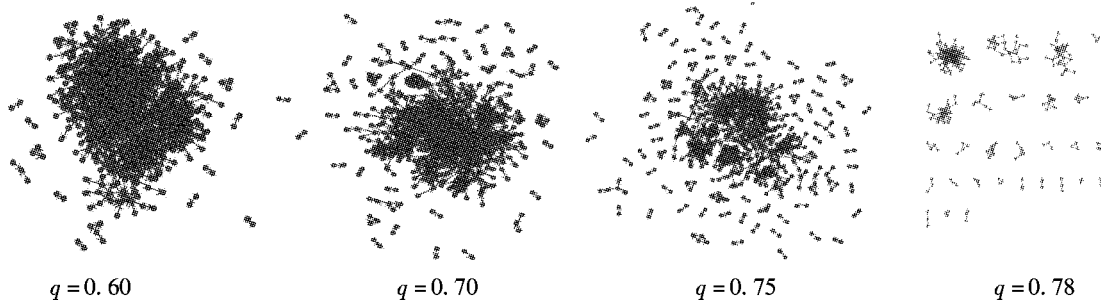


图5 不同去噪因子对应的乳腺癌基因共表达网络

在阈值为 0.78 对应的乳腺癌基因共表达网络中,得到了 19 个分立模块,模块中的基因数量为 3 ~ 86 个不等.

## 3 结果与分析

通过分析乳腺癌共表达网络,找到 2 个典型的

模块:乳腺癌增殖模块和免疫模块(见图6).

增殖模块由与细胞增殖相关的基因组成,这个模块中的基因绝大部分被报导在 G2/M 期异常表达.众所周知,肿瘤出现的重要原因是由细胞增殖和异常分化而导致的.在增殖模块中,位于模块最中心位置处的基因有 FEN1<sup>[18]</sup>、TRIP13<sup>[19]</sup>、E2-EPF<sup>[20]</sup>、CDC2<sup>[21]</sup>、KNSL5<sup>[22]</sup>、TEAD4<sup>[23]</sup>、BUB1<sup>[24]</sup>、FOXMI<sup>[25]</sup>、

RFC4<sup>[26]</sup>、ATRX<sup>[27]</sup>、TOPBP1<sup>[28]</sup>等,这些基因都已被发现与乳腺癌细胞的侵袭、增殖及迁移有关.免疫模块中的 TRA@、TRB@、TRD@、IL2RG、CD8A、CD3D、CD48 等基因与免疫功能的 T 细胞有关<sup>[29]</sup>;

HLA-DQB1、HLA-DRB5、HLA-DMA、HLA-DQA1、HLA-DPB1 等基因属于白细胞抗原 HLA 中的 II 类分子<sup>[30]</sup>,这类分子主要调控细胞的免疫应答.

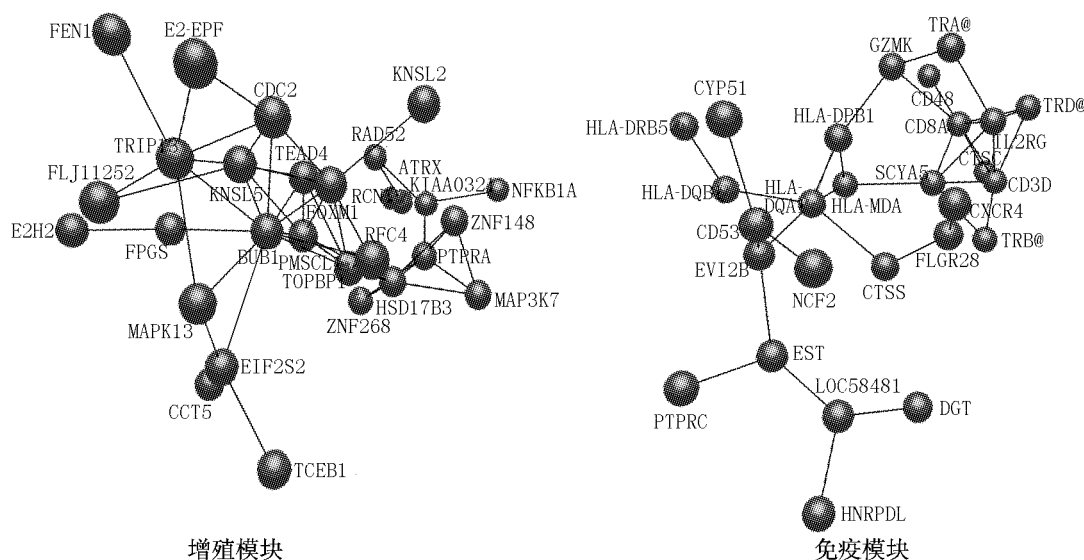


图6 增殖模块与免疫模块

通过分析乳腺癌共表达网络的增殖模块及免疫模块,发现 PMSCL1 基因及 SCYA5 基因在乳腺癌中的功能未知,将试图通过共表达网络模块中功能已知的基因来预测这 2 个功能未知的基因.从图 7(a)可看出, PMSCL1 基因与 TEAD4、BUB1、FOXM1、RFC4、TOPBP1、KNSL5 基因紧密相连,而基因 TEAD4、BUB1、FOXM1、RFC4、TOPBP1、KNSL5 均与乳腺癌细胞的侵袭、增殖及迁移有关.根据模块的特点:分布在同一个模块中的基因具有相同的功能性,

预测 PMSCL1 基因与乳腺癌细胞的侵袭、增殖及迁移有关.从图 7(b)可看出, SCYA5 基因与 TRB@、CD3D、IL2RG、CD8A、HLA-DMA 基因紧密相连,且 SCYA5 基因处于这些基因组成的共表达网络的中心点,研究发现 TRB@、CD3D、IL2RG、CD8A、HLA-DMA 基因与乳腺癌细胞的免疫功能及免疫应答相关,故根据分布在同一个模块中的基因具有相同的的功能性特点,预测 SCYA5 基因与乳腺癌细胞的免疫功能及应答有关.

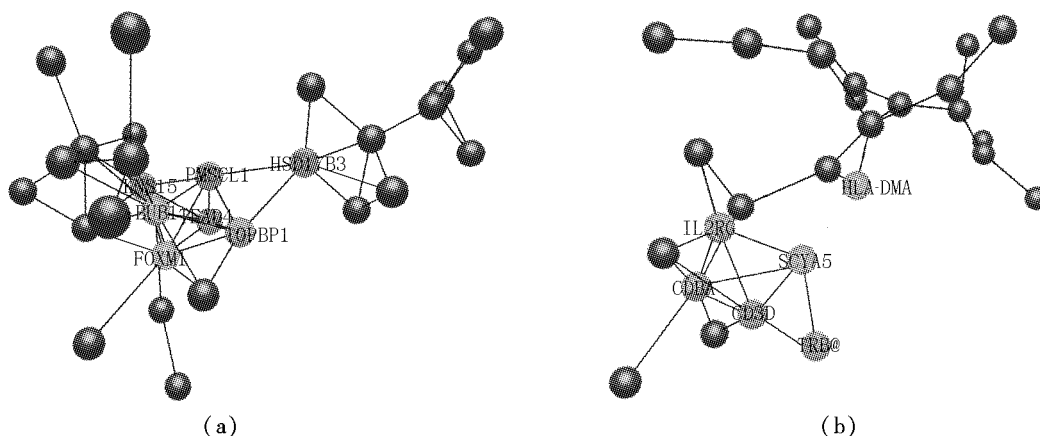


图7 PMSCL1 基因及 SCYA5 基因所在的模块

在乳腺癌基因共表达网络中,重点分析了 2 个基因个数较少但相关系数较强的模块(见图 8).

MAD2L1、TROAP、STK15、CCAN2、LRP8 这 5 个基因组成图 8 中的模块(a),其中 MAD2L1、TROAP、STK15、CCAN2 这 4 个基因相互紧密关联.研究发现:MAD2L1 基因主要参与有丝分裂,且 MAD2L1 基

因的高表达量与乳腺癌患者愈后复发呈正相关性<sup>[31-32]</sup>.TROAP 基因通常出现在细胞有丝分裂过程中,该基因的高表达可促使乳腺癌晚期患者的有丝分裂速度加快、加速乳腺癌细胞的增殖<sup>[33]</sup>.STK15 基因在多种恶性肿瘤中都被发现有高表达量,其也参与细胞的有丝分裂<sup>[34-35]</sup>.可见,MAD2L1、STK15、

TROAP 这3个基因均与肿瘤细胞的有丝分裂相关,而 CCAN2 基因与这3个基因紧密互联.根据模块的功能特点,预测 CCAN2 基因与乳腺癌细胞的有丝分裂有关.

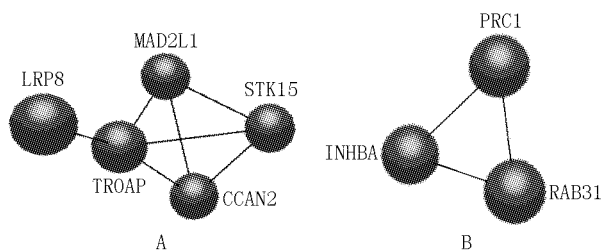


图8 2个分立的小模块

PRC1、RAB31、INHBA 这3个基因组成图8中的模块(b),从图8(b)可看出,3个基因为一个分立模块且构成一个环状连接.研究表明,PRC1 基因是一种蛋白基因,主要参与细胞质的分裂,常在有丝分裂的S和G2/M期出现,PRC1 基因能促进乳腺癌细胞的增殖速度<sup>[36]</sup>;Rab31 基因能激发G1/S细胞周期的转换的活性,是一种能够促使侵袭性细胞系增殖的基因<sup>[37]</sup>;INHBA 基因主要参与细胞的增殖及分化.研究结果表明:当INHBA 基因表达值上调时,会推进肿瘤细胞增长,且参与血管壁浸润及肿瘤细胞的迁移<sup>[38]</sup>.综上所述,PRC1、RAB31、INHBA 这3个基因均能促使乳腺癌细胞的增殖,3个基因构成环状连接,在乳腺癌的增殖过程中相互作用、相互依存,它们共同促使乳腺癌细胞的发展.若抑制其中任意1个基因的作用,则会导致另外2个基因相互促进作用减弱,从而抑制乳腺癌的增殖速度,故预测PRC1、RAB31、INHBA 这3个基因可用于乳腺癌的靶向治疗.

## 4 结论

本文利用随机矩阵理论处理了乳腺癌基因微阵列数据,研究表明:随机矩阵理论可以把乳腺癌基因网络中本质属性从随机噪声中分离出来,能客观真实地揭示乳腺癌基因共表达网络模块.通过乳腺癌共表达网络模块的分析,预测出与乳腺癌细胞发展相关的基因,为乳腺癌的临床研究提供一定的帮助.下一步的工作是将该方法运用到其他肿瘤的研究中,找出更多与肿瘤相关的基因,为肿瘤临床的诊断、治疗提供科学的理论依据.

## 5 参考文献

[1] Siegel R L, Miller K D, Jemal A. Cancer statistics, 2020 [J]. CA: A Cancer Journal for Clinicians, 2020, 70(1): 7-30.  
[2] Lei Fan, Kathrin S W, Jun Jieli, et al. Breast cancer in

China [J]. Oncology, 2014, 15(7): e279-e289.  
[3] 吴样平,郭飞,曾明华. 基于复杂网络的城市综合交通网络特征分析与优化研究 [J]. 江西师范大学学报:自然科学版, 2015, 39(3): 326-330.  
[4] Schena M, Shalon D, Davis R W, et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray [J]. Science, 1995, 270(5235): 467-470.  
[5] 于凤英,杨志豪,林鸿飞. 基于拓扑和生物特征的权重网络中络合物抽取 [J]. 江西师范大学学报:自然科学版, 2013, 37(3): 273-278.  
[6] Luo Feng, Zhong Jianxin, Yang Yunfeng, et al. Application of random matrix theory to microarray data for discovering functional gene modules [J]. Physical Review E, 2006, 73(3): 031924.  
[7] Luo Feng, Zhong Jianxin, Yang Yunfeng, et al. Application of random matrix theory to biological networks [J]. Physics Letter A, 2006, 357(6): 420-423.  
[8] Luo Feng, Yang Yunfeng, Zhong Jianxin, et al. Constructing gene co-expression networks and predicting functions of unknown genes by random matrix theory [J]. BMC Bioinformatics, 2007, 8: 299.  
[9] Ayumi K. Random matrix analysis for gene interaction networks in cancer cells [J]. Scientific Reports, 2018, 8(7): 10607.  
[10] Chen Jian, Zhang Kaiwan, Zhong Jianxin. Identifying functional modules of diffuse large B-cell lymphoma gene co-expression networks by hierarchical clustering method based on random matrix theory [J]. Nano Biomedicine and Engineering, 2011, 3(1): 57-64.  
[11] 李蓉,郑浪,任喜梅,等. 基于随机矩阵理论及层次聚类方法在肝癌基因网络中的研究 [J]. 湘潭大学学报:自然科学版, 2019, 41(2): 55-60.  
[12] Shuryak E V, Verbaarschot J J M. Random matrix theory and spectral sum rules for the Dirac operator in QCD [J]. Nuclear Physics A, 1993, 560(1): 306-320.  
[13] Zhong Jianxin, Geisel T. Level fluctuations in quantum systems with multifractal eigenstates [J]. Physical Review E, 1999, 59(4): 4071-4074.  
[14] 李蓉,颜平兰,陈建,等. 随机矩阵理论在肺癌基因网络识别中的应用 [J]. 物理学报, 2009, 58(10): 6703-6708.  
[15] Kwapien J, Drozd S, Oswiecimka P. The bulk of the stock market correlation matrix is not pure noise [J]. Physica A, 2006, 359(1): 589-606.  
[16] 谢赤,胡钰,王钢金. 基于随机矩阵理论的股市网络拓扑性质研究 [J]. 运筹与管理, 2018, 27(1): 144-152.  
[17] Theocharidis A, Dongen S V, Enright A J. Network visualization and analysis of gene expression data using biolayout [J]. Nature Protocol, 2009, 4(10): 1535-1550.  
[18] He Lingfeng, Yang Huan, Zhou Shiyang, et al. Synergistic antitumor effect of combined paclitaxel with FEN1 inhibitor in cervical cancer cells [J]. DNA Repair, 2018, 63(3): 1-9.

- [19] Kurita K, Maeda M, Mansour M, et al. TRIP13 is expressed in colorectal cancer and promotes cancer cell invasion [J]. *Oncology Letters*, 2016, 12(6): 5240-5246.
- [20] Zhang Xiaohui, Zhao fangbo, Zhang Shunjun, et al. Ectopic high expression of E2-EPF ubiquitin carrier protein indicates a more unfavorable prognosis in brain glioma [J]. *Genetic Testing and Molecular Biomarkers*, 2017, 21(4): 242-247.
- [21] Wang Jingnan, Zhang Zhirong, Che Yun, et al. Acetyl-macrocyclic B, an ent-kaurane diterpenoid, initiates apoptosis through the ROS-p38-caspase 9-dependent pathway and induces G2/M phase arrest via the Chk1/2-Cdc25C-Cdc2/cyclin B axis in non-small cell lung cancer [J]. *Cancer Biology and Therapy*, 2018, 19(7): 609-621.
- [22] Kress M, Kress C, Chauviere M. Disruption of the mitotic kinesin Eg5 gene (Kns11) results in early embryonic lethality [J]. *Biochemical and Biophysical Research Communications*, 2008, 372(4): 513-519.
- [23] 李臻, 李树安, 罗华友, 等. TEAD 对结肠直肠癌细胞增殖影响及机制探讨 [J]. *中华肿瘤防治杂志*, 2017, 24(15): 1035-1045.
- [24] 白晓斌, 霍龙伟, 谢万福, 等. BUB1 在胶质母细胞瘤中高表达并促进胶质母细胞瘤的增殖 [J]. *西安交通大学学报: 医学版*, 2019, 40(2): 278-283.
- [25] Gao Fangyuan, Feng Jing, Yao Hairong, et al. LncRNA SBF2-AS1 promotes the progression of cervical cancer by regulating miR-361-5p/FOXO1 axis [J]. *Artificial Cells Nanomedicine and Biotechnology*, 2019, 47(1): 776-782.
- [26] Hou Lingmi, Chen Maoshan, Wang Minghao, et al. Systematic analyses of key genes and pathways in the development of invasive breast cancer [J]. *Gene*, 2016, 593(1): 1-12.
- [27] Lui Minghong, Jiang Biaoabin, Bao Zhaoshi, et al. A mini-review on spatiotemporal evolution of glioma under treatment [J]. *Progress in Biochemistry and Biophysics*, 2019, 46(11): 1055-1062.
- [28] Leimbacher P A, Jones S E, Shorrock A K, et al. MDC1 interacts with TOPBP1 to maintain chromosomal stability during mitosis [J]. *Mol Cell*, 2019, 74(3): 571-583.
- [29] Johansson B, Karrman K, Laksosson M, et al. The insulin receptor substrate 4 gene (IRS4) is mutated in paediatric T-cell acute lymphoblastic leukaemia [J]. *British Journal of Haematology*, 2011, 155(4): 516-519.
- [30] Matei H V, Vică M L, Ciucă I, et al. Correlations among the HLA-DQB1 alleles and suicidal behavior [J]. *Journal of Forensic Science*, 2020, 65(1): 166-169.
- [31] Wang Zhanwei, Katsaros D, Shen Yi, et al. Biological and clinical significance of MAD2L1 and BUB1, genes frequently appearing in expression signatures for breast cancer prognosis [J]. *PLOS ONE*, 2015, 10(8): e0136246.
- [32] Shi Yuanxiang, Zhu Tao, Zou Ting, et al. Prognostic and predictive values of CDK1 and MAD2L1 in lung adenocarcinoma [J]. *Oncotarget*, 2016, 7(51): 85235-85243.
- [33] Jing Ke, Mao Qingsheng, Ma Peng. Decreased expression of TROAP suppresses cellular proliferation, migration and invasion in gastric cancer [J]. *Molecular Medicine Reports*, 2018, 18(3): 3020-3026.
- [34] 高峻, 杨勇, 陈霖, 等. STK15、MCM5 在胃癌前病变及癌变组织中的表达及相关性分析 [J]. *实用医学杂志*, 2014, 30(9): 1432-1434.
- [35] Green M R, Woolery J E, Mahadevan D. Update on aurora kinase targeted therapeutics in oncology [J]. *Expert Opinion on Drug Discovery*, 2011, 6(3): 291-307.
- [36] 施晓婷, 张斌, 邹晓平. PRC1 在细胞分裂及肿瘤发生中的作用 [J]. *现代肿瘤医学*, 2016, 24(23): 3860-3863.
- [37] Goldenring J R. A central role for vesicle trafficking in epithelial neoplasia Intracellular highways to carcinogenesis [J]. *Nature Reviews Cancer*, 2013, 13(11): 813-820.
- [38] 茅育蕾, 蔡宜玲. 乳腺癌中 INHBA 表达及与临床预后的关系 [J]. *临床与实验病理学杂志*, 2018, 34(3): 331-333.

## The Study of Breast Cancer Gene Network Based on Random Matrix Theory

LI Rong<sup>1</sup>, ZHENG Lang<sup>2</sup>, REN Ximei<sup>1</sup>, LIU Chao<sup>3</sup>

(1. Institute of Technology, East China Jiaotong University, Nanchang Jiangxi 330100, China;

2. Jiangxi Provincial Education Examination Authority, Nanchang Jiangxi 330038, China;

3. Jiangxi University of Science and Technology, Ganzhou Jiangxi 341000, China)

**Abstract:** By using random matrix theory, the breast cancer gene microarray data is analyzed. The proliferation and immune modules of breast cancer gene are found out from breast cancer gene express network. Meanwhile, it is shown that the gene PMSCL1 is related to breast cancer cell proliferation, invasion and migration, gene CCAN2 is related to the Mitosis of breast cancer cells, gene SCYA5 is related to immune response of breast cancer cells and gene PRC1, RAB31 and INHBA can be used as targeted genes for breast cancer treatment.

**Key words:** random matrix theory; breast cancer; gene co-expression network; module; targeted gene

(责任编辑: 刘显亮)