

利用加权基因共表达网络分析构建食管腺癌预后枢纽基因网络

陈超 童国俊 张建斌 沈亮 何焕钟

【摘要】 目的 利用加权基因共表达网络分析(WGCNA)构建食管腺癌预后枢纽基因网络,筛选与食管腺癌预后相关的枢纽基因。方法 提取癌症和肿瘤基因组计划(TCGA)数据库中食管腺癌的标本数据,利用 WGCNA 匹配基因表达数据与临床预后数据,构建食管腺癌的基因共表达网络模块,并筛选出与预后最相关的模块和枢纽基因,再根据共表达权重系数大小在 Cytoscape 软件中进行可视化。结果 共获得 9 例正常食管组织和 78 例食管腺癌的标本,其中男 67 例,女 11 例;年龄 28.0~86.6[68.4(58.0,77.1)]岁;TNM 分期:Ⅰ期 10 例,ⅡA 期 9 例,ⅡB 期 16 例,Ⅲ期 33 例,Ⅳ期 10 例。共筛选得到的多个模块中深蓝色模块与预后最相关,在深蓝色模块中识别出 19 个枢纽基因(PAQR8、MAMDC4、CEP44、SCRG1、UGT2B15、NUDT9、FOLH1、SFTPB、FBXO8、TENM1、CASR、NEB、SMIM19、SLC20A2、RNF170、SCN2A、GOLIM4、ICK、DNAH2),并构建了枢纽基因的共表达网络。其中基因间共表达权重系数最大的 3 对基因分别是 FOLH1 和 SCRG1、FOLH1 和 UGT2B15、FOLH1 和 SFTPB。结论 利用 WGCNA 可识别与食管腺癌预后相关的枢纽基因,为食管腺癌的治疗提供新靶点。

【关键词】 食管腺癌 加权基因共表达网络分析 预后

Construction of prognostic hub gene network of esophageal adenocarcinoma by weighted gene co-expression network analysis
CHEN Chao, TONG Guojun, ZHANG Jianbin, et al. Department of Anesthesiology, Huzhou Central Hospital, Huzhou 313000, China

【Abstract】 Objective To construct a prognostic hub gene network for esophageal adenocarcinoma with weighted gene co-expression analysis (WGCNA). Methods We extracted the specimen data of esophageal adenocarcinoma in the TCGA database, and used the WGCNA method matching the gene expression data with clinical prognostic data to construct a gene co-expression network module for esophageal adenocarcinoma. The modules and hub genes most relevant to prognosis were screened out, and then visualized according to the size of expression weights in Cytoscape software. Results Nice normal esophageal tissue specimens and 78 esophageal adenocarcinoma specimens were obtained from 67 males and 11 females patients aged 28.0~86.6 [68.4 (58.0, 77.1)] years, among whom there were 10 cases in stage I, 9 cases in stage II A, 16 cases in stage II B, 33 cases in stage III and 10 cases in stage IV. The midnight blue module in multiple modules screenings was the most relevant to prognosis. Nineteen hub genes were identified in the midnight blue module(PAQR8, MAMDC4, CEP44, SCRG1, UGT2B15, NUDT9, FOLH1, SFTPB, FBXO8, TENM1, CASR, NEB, SMIM19, SLC20A2, RNF170, SCN2A, GOLIM4, ICK, DNAH2), and a co-expression network of hub genes was constructed. The three pairs of genes with the largest co-expression weight coefficient were FOLH1 and SCRG1, FOLH1 and UGT2B15, FOLH1 and SFTPB. Conclusion WGCNA can be used to identify the hub genes associated with the prognosis of esophageal adenocarcinoma, providing a new target for the treatment of esophageal adenocarcinoma.

【Key words】 Esophageal adenocarcinoma WGCNA Prognosis

DOI: 10.12056/j.issn.1006-2785.2019.41.20.2019-1701

基金项目:湖州市公益性研究项目(2018GY10);浙江省实验动物科技计划项目(2018C37090)

作者单位:313000 湖州市中心医院麻醉科(陈超、何焕钟),精准医学临床研究中心(童国俊),胸外科(张建斌),神经外科(沈亮)

通信作者:何焕钟, E-mail: hhzhma@sina.com

食管癌是因癌症导致死亡的第六大原因,5 年生存率仅为 19%,其中晚期食管癌患者 5 年生存率仅 0.9%^[1]。近年来,食管腺癌发病率逐年上升;在西方国家,食管腺癌已成为发病率最高的食管恶性肿瘤^[2]。随着医疗水平的发展,食管腺癌患者的预后改善仍十分有限。在复杂的肿瘤微环境下,传统的单基因研究因不能揭示肿瘤基因复杂的信号传导网络而存在很大的局限性。随着基因

芯片、转录组测序(RNA-seq)技术的发展,应用生物信息分析实现多个基因表达及功能研究的方法,为疾病发生、发展的机制探索提供了新思路。本研究提取癌症和肿瘤基因图谱计划(TCGA)数据库中 78 例具有 RNA-seq 的食管腺癌标本信息,通过加权基因共表达网络分析(weighted gene co-expression network analysis, WGCNA)研究肿瘤的 RNA-seq 数据,筛选与预后相关的模块及枢纽基因,并根据表达谱信息构建多个枢纽基因的共表达网络关系。现将结果报道如下。

1 资料和方法

1.1 数据来源与预处理 9 例正常食管组织和 78 例食管腺癌组织的基因表达数据及临床预后数据均来自于免费、开源的 TCGA 数据库(<https://cancergenome.nih.gov/>)。将所有样本的基因名转化为标准基因名;数据的标准化处理是以正常食管组织为参照,对肿瘤标本数据进行归一化处理。

1.2 共表达网络的构建和模块的识别 利用 R-3.5.1 软件运行“WGCNA”包。为降低运算量,笔者筛选出基因表达量方差大于所有方差四分位数的基因用于共表达网络的构建。采用样本聚类树方法,根据聚类图剔除离群样本来保证构建稳定的共表达网络。构建无尺度网络使基因共表达网络符合无尺度现象,以无尺度网络指数(R^2)=0.9 作为满足无尺度条件的标准,同时根据平均连接度确定软阈值(β)。利用拓扑重叠(TOM)矩阵、相异度矩阵计算基因与基因间的关联程度^[3];对基因构建层次聚类树图形,采用动态剪枝法计算基因模块颜色。计算基因模块的特征值(ME),引入临床信息,对 ME 进行分层聚类并绘制树状图,设置高度值 0.25 为分割线,合并相似程度较高的基因模块,再用剪切后的模块绘制新的聚类树和模块图。

1.3 表观数据的纳入和枢纽基因的筛选 读取、清洗临床样本表观数据,主要包括性别、年龄、肿瘤级别、生存时间等,将样本与表达矩阵进行匹配重建样本聚类树。计算模块与表观数据相关性,绘制模块内基因表达热图,根据生存时间性状与模块特征向量基因的相关性及 P 值来挖掘与该性状相关的模块。计算相关模块内基因显著性(GS)以及基因在模块内的模块隶属度(MM),设置枢纽模块中候选枢纽基因 3 个筛选标准: $|GS| > 0.2$, $|MM| > 0.8$, $P.GS < 0.05$ 。同时计算加权基因共表达网络的权重值,筛选出权重值前 200 位的基因,与上述候选枢纽基因的交集为共同枢纽基因。

1.4 枢纽基因及共表达网络可视化 将共同枢纽基因

及基因共表达网络的权重信息导入 Cytoscape3.7.0 软件,根据互作网络关系绘制枢纽基因共表达网络图。

2 结果

2.1 TCGA 数据库中食管腺癌病例基本信息 本研究共纳入食管腺癌患者 78 例,其中男 67 例,女 11 例;年龄 28.0~86.6[68.4(58.0, 77.1)]岁;生存时间(634±513)d;随访时间(603±524)d;国际抗癌联盟食管癌 TNM 分期: I 期 10 例, II A 期 9 例, II B 期 16 例, III 期 33 例, IV 期 10 例。

2.2 共表达网络的构建与枢纽模块的选择 基因表达量方差大于所有方差四分位数的基因共 9 933 个,通过聚类树删除离群样本 14 例,剩余 64 例食管腺癌样本纳入下一步分析。根据 $R^2=0.9$, $\beta=5$ 作为标准,当 $\beta=5$ 时,无尺度拓扑网络的 $R^2=0.95$,见图 1。引入生存时间信息,采用动态剪切法将树剪切成不同的模块,合并相似度较高的模块,见图 2a(插页)。通过模块样本性状相关性热图,发现深蓝色模块与生存预后密切相关,相关系数(Cor)=0.3, $P=0.01$,即深蓝色模块为枢纽模块,见图 2b(插页)。枢纽模块中基因内部连接度与 MM 相关,深蓝色模块与基因显著性相关, $Cor=0.61$, $P < 0.01$,见图 3。

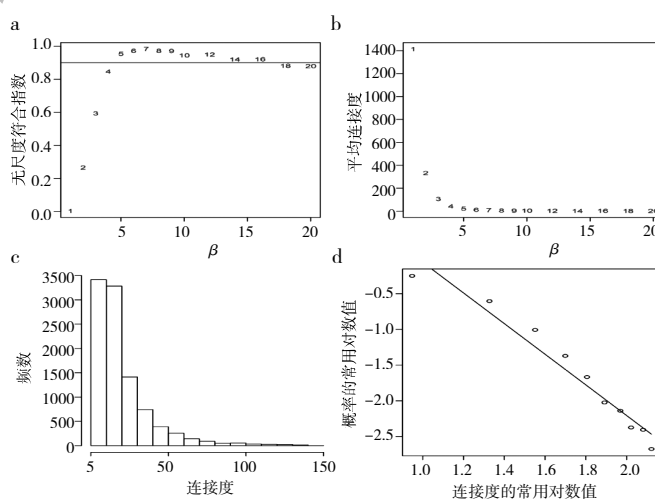


图 1 WGCNA 的 β 确定(a:不同 β 下计算的无尺度网络符合指数;b:不同 β 下计算的平均连接度;c: $\beta=5$ 时连接度分布直方图;d: $\beta=5$ 时无尺度网络拓扑检测)

2.3 枢纽基因确定及共表达网络可视化 在深蓝色枢纽模块中,根据 $|GS| > 0.2$, $|MM| > 0.8$, $P.GS < 0.05$ 的标准,筛选得到 20 个候选枢纽基因集 A,再根据基因权重共表达网络的权重大小筛选得到 32 个候选枢纽基因集 B,基因集 A 和 B 的交集基因有 19 个,见表 1。将 19

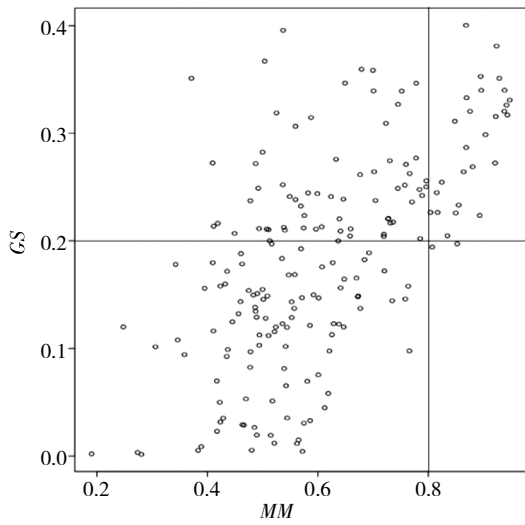


图 3 深蓝色模块中 GS 与 MM 的关系(右上角为 $|GS| > 0.2$ 、 $|MM| > 0.8$ 的基因分布)

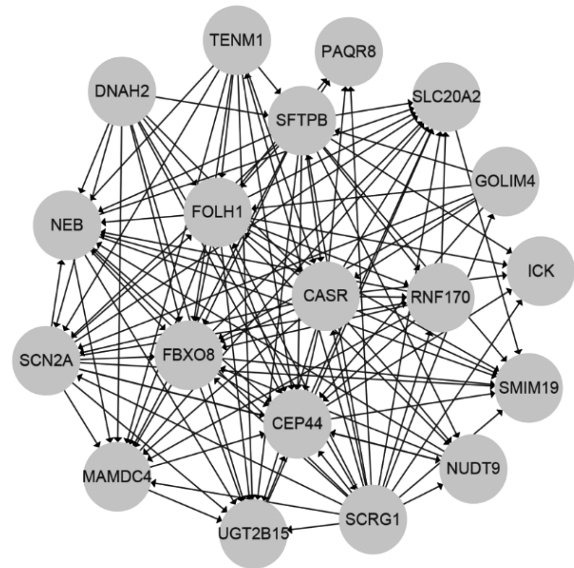


图 4 枢纽基因之间的共表达网络关系图

个基因间的权重信息导入 Cytoscape 软件后得到基因间互作网络图,其中基因间共表达权重系数最大的 3 对基因分别是 FOLH1 和 SCRG1、FOLH1 和 UGT2B15、FOLH1 和 SFTPB,见图 4。

表 1 19 个与临床预后相关的枢纽基因 GS、MM 和 P.GS

枢纽基因	GS	MM	P.GS
PAQR8	0.381067151	0.923868481	0.001892674
MAMDC4	0.352799115	0.895011207	0.004245409
CEP44	0.35109391	0.928840166	0.004447384
SCRG1	0.340001512	0.938365143	0.005981165
UGT2B15	0.339942209	0.896046781	0.005990481
NUDT9	0.332964343	0.869372822	0.007179661
FOLH1	0.330871749	0.947711621	0.007574409
SFTPB	0.326163229	0.941750114	0.008532686
FBXO8	0.320436089	0.937896834	0.009839428
TENM1	0.320355041	0.875854074	0.009859104
CASR	0.316968001	0.943456248	0.010712537
NEB	0.315549113	0.922265876	0.011088669
SMIM19	0.31115967	0.848180275	0.012325626
SLC20A2	0.298690164	0.903774899	0.01651039
RNF170	0.286755218	0.868841198	0.021601444
SCN2A	0.272391544	0.92118778	0.029440244
GOLIM4	0.268892387	0.880519545	0.031675456
ICK	0.264191204	0.864208979	0.034900044
DNAH2	0.25453708	0.824581356	0.042384281

3 讨论

胃食管反流、吸烟、肥胖、饮食习惯是食管腺癌的危险因素^[2]。而在危险因素影响下的表观遗传异常修饰导

致基因转录水平改变是肿瘤预后研究的基础。局限于某个基因或某几个分子的传统研究方法不能全面地阐述肿瘤的发展及预后。WGCNA 在 2005 年首先被提出,是指根据基因集的内连性和基因集与表型之间的关联鉴定候补生物标记基因或治疗靶点^[4]。相比于传统的单基因研究,WGCNA 利用数千或近万个变化最大的基因或全部基因的信息识别感兴趣的基因模块,并与表型进行显著性关联分析。该方法创新性强,研究结果可信度较高。近期亦有不少期刊发表了利用 WGCNA 的相关研究^[5-7]。

本研究利用 WGCNA 处理 78 例食管腺癌标本基因的 RNA-seq 数据及临床数据,通过深层次挖掘得到 19 个枢纽基因,并构建了枢纽基因的共表达网络。WGCNA 构建的预后相关的枢纽基因共表达网络图中,FOLH1 与 SCRG1、UGT2B15、SFTPB 等 3 个基因的共表达权重系数最高,属于共表达网络图中的核心基因。FOLH1 是一种 II 型跨膜糖蛋白,其分子量约为 100kDa,由 750 个氨基酸组成^[8]。FOLH1 主要在前列腺、中枢及外周神经系统、肾脏、小肠和肿瘤相关的新血管系统中表达^[9],可通过谷氨酸代谢来调节叶酸的吸收^[10]。有研究认为 FOLH1 表达水平能评估肿瘤患者的预后,尤其在已发生肿瘤转移的情况下;同时发现经普通筛查发现的恶性肿瘤分级与 FOLH1 表达亦密切相关^[11]。Chang 等^[12]称 FOLH1 在多种恶性肿瘤的新生血管系统中均有表达,可作为抗肿瘤新生血管生成治疗的有效靶点。目前关于 FOLH1 在前列腺肿瘤中表达的研究较多^[13-15],亦有 FOLH1 与乳腺癌^[16]、非小细胞肺癌^[17]等肿瘤预后有关的

报道。然而,关于 FOLH1 在食管恶性肿瘤中表达的研究尚未见报道。根据 WGCNA 构建的共表达网络分析以及关于 FOLH1 的现有研究结果,笔者认为 FOLH1 可能是食管腺癌中新的预后相关分子。

本研究通过 WGCNA 构建加权基因共表达网络,初步筛选得到与食管腺癌预后相关的 19 个枢纽基因及其共表达网络关系,为食管腺癌的治疗提供新靶点。

4 参考文献

- [1] Testa U, Castelli G, Pelosi E. Esophageal Cancer: Genomic and Molecular Characterization, Stem Cell Compartment and Clonal Evolution[J]. *Medicines(Basel)*, 2017, 4(3):67. DOI:10.3390/medicines4030067.
- [2] Smyth EC, Lagergren J, Fitzgerald RC, et al. Oesophageal cancer[J]. *Nat Rev Dis Primers*, 2017, 3:17048.
- [3] Li A, Horvath S. Network neighborhood analysis with the multi-node topological overlap measure[J]. *Bioinformatics*, 2007, 23(2):222-231. DOI:10.1093/bioinformatics/btl581.
- [4] Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis[J]. *Stat Appl Genet Mol Biol*, 2005, 4:17. DOI:10.2202/1544-6115.1128.
- [5] Schafer ST, Paquola ACM, Stern S, et al. Pathological priming causes developmental gene network heterochronicity in autistic subject-derived neurons[J]. *Nat Neurosci*, 2019, 22(2):245-255. DOI:10.1038/s41593-018-0295-x.
- [6] Tran SS, Jun HI, Bahn JH, et al. Widespread RNA editing dysregulation in brains from autistic individuals[J]. *Nat Neurosci*, 2019, 22(1):25-36. DOI:10.1038/s41593-018-0287-x.
- [7] Voineagu I, Wang X, Johnston P, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology[J]. *Nature*, 2011, 474(7351):380-384. DOI:10.1038/nature10110.
- [8] Sacha P, Zamecnik J, Barinka C, et al. Expression of glutamate carboxypeptidase II in human brain[J]. *Neuroscience*, 2007, 144(4):1361-1372. DOI:10.1016/j.neuroscience.2006.10.022.
- [9] Barinka C, Sacha P, Sklenar J, et al. Identification of the N-glycosylation sites on glutamate carboxypeptidase II necessary for proteolytic activity[J]. *Protein Sci*, 2004, 13(6):1627-1635. DOI:10.1110/ps.04622104.
- [10] Hiraoka M, Kagawa Y. Genetic polymorphisms and folate status [J]. *Congenit Anom (Kyoto)*, 2017, 57(5):142-149. DOI:10.1111/cga.12232.
- [11] Murphy GP, Barren RJ, Erickson SJ, et al. Evaluation and comparison of two new prostate carcinoma markers. Free-prostate specific antigen and prostate specific membrane antigen[J]. *Cancer*, 1996, 78(4):809-818. DOI:10.1002/(SICI)1097-0142(19960815)78:4<809::AID-CNCR18>3.0.CO;2-Z.
- [12] Chang SS, Reuter VE, Heston WD, et al. Five different anti-prostate-specific membrane antigen (PSMA) antibodies confirm PSMA expression in tumor associated neovasculature[J]. *Cancer Res*, 1999, 59(13):3192-3198.
- [13] Bakht MK, Derecichei I, Li Y, et al. Neuroendocrine differentiation of prostate cancer leads to PSMA suppression[J]. *Endocr Relat Cancer*, 2018, 26(2):131-146. DOI:10.1530/ERC-18-0226.
- [14] Owiti AO, Pal D, Mitra A. PSMA Antibody-Conjugated Pentablock Copolymer Nanomicellar Formulation for Targeted Delivery to Prostate Cancer[J]. *AAPS Pharm Sci Tech*, 2018, 19(8):3534-3549. DOI:10.1208/s12249-018-1126-9.
- [15] Zacho HD, Nielsen JB, Dettmann K, et al. 68Ga-PSMA PET/CT in Patients With Biochemical Recurrence of Prostate Cancer: A Prospective, 2-Center Study[J]. *Clin Nucl Med*, 2018, 43(8):579-585. DOI:10.1097/RLU.0000000000002169.
- [16] Kasoha M, Unger C, Solomayer EF, et al. Prostate-specific membrane antigen (PSMA) expression in breast cancer and its metastases[J]. *Clin Exp Metastasis*, 2017, 34(8):479-490. DOI:10.1007/s10585-018-9878-x.
- [17] Schmidt LH, Heitkotter B, Schulze AB, et al. Prostate specific membrane antigen (PSMA) expression in non-small cell lung cancer[J]. *PLoS One*, 2017, 12(10):e0186280. DOI:10.1371/journal.pone.0186280.

(收稿日期:2019-06-09)

(本文编辑:陈丹)

读者·作者·编者

《浙江医学》对医学论文中有关实验动物描述的要求

在医学论文的描述中,凡涉及实验动物者,在描述中应符合以下要求:(1)品种、品系描述清楚;(2)强调来源;(3)遗传背景;(4)微生物学质量;(5)明确等级及体重;(6)明确饲养环境和实验环境;(7)明确性别;(8)有无质量合格证;(9)有对饲养的描述(如饲料型、营养水平、照明方式、温度、湿度要求);(10)所有动物数量准确;(11)详细描述动物的健康状况;(12)对动物实验的处理方式有单独清楚的交代;(13)全部有对照,部分可采用双因素方差分析。

本刊编辑部