

• 论 著 •

胃癌转录调控网络构建及预后相关转录因子分析*

康争春¹, 李 勃², 蔡 慧¹, 马立业¹, 于恩达^{1△}

(1. 中国人民解放军海军军医大学长海医院普通外科, 上海 200433;

2. 中国人民解放军北京军区北戴河疗养院普通外科, 河北秦皇岛 066100)

摘要:目的 构建胃癌转录调控网络, 分析预后相关转录因子。方法 首先用基因芯片数据构建差异表达转录因子在癌组织中的共表达网络, 进而发现共表达网络中的核心亚网络, 然后用具有预后数据的合适表达谱芯片, 根据构建的网络筛选预后相关的转录因子, 采用样本量较大、有预后数据的表达谱芯片验证转录因子对胃癌患者预后是否有预测作用。结果 共发现 4 个核心网络、21 个核心转录因子。结合转录因子核心网络, 确定和预后相关的 4 个核心转录因子其中的 3 个和患者预后相关。结论 构建胃癌转录调控网络可为研究胃癌的潜在调控机制提供参考, 并可能为将来胃癌的诊断治疗提供分子标志物及靶点。

关键词:胃癌; 转录因子; 调控网络; 预后

DOI:10.3969/j.issn.1672-9455.2017.21.004 文献标志码:A 文章编号:1672-9455(2017)21-3142-03

Construction of transcriptional regulatory networks and analysis of prognostic transcription factors in gastric cancer*
KANG Zhengchun¹, LI Bo², CAI Hui¹, MA Liye¹, YU Enda^{1△}
(1. Department of General Surgery, Changhai Hospital, PLA Navy Medical University, Shanghai 200433, China;
2. Department of General Surgery, Beidaihe Sanatorium of Beijing Military Region of PLA, Qinhuangdao, Hebei 066100, China)

Abstract: **Objective** The purpose of this study is to construct the transcriptional regulatory network of gastric cancer, and analysis the prognostic transcription factors. **Methods** Firstly, using gene chip data to construct co-expression networks of differentially expressed transcription factors in cancer tissues, and then found the core subnetwork in the network, and then screen the core transcription factors based on the microarray with prognosis data. We finally used a large amount of data to verify whether these transcription factors can predict the prognosis of gastric cancer patients. **Results** There are 21 core transcription factors of 4 core subnetworks. We identified 4 key transcription factors of that 3 transcription factors had obvious function of prediction. **Conclusion** We have constructed the transcriptional regulatory network which provides references for the potential regulatory mechanism of gastric cancer. They may provide molecular markers for future diagnosis of gastric cancer, and possible targets for the treatments of gastric cancer.

Key words: gastric cancer; transcription factor; regulatory network; prognosis

胃癌是我国发病率、病死率最高的恶性肿瘤之一, 每年约有 50 万人死于胃癌, 其严重危害着人们的生命健康^[1]。目前, 科学界对胃癌的发病原因和机制都不甚清楚, 无法制订有效的防治措施。胃癌早期症状较少, 大多数患者通常在癌症进展到中晚期后才被确诊, 预后较差。胃癌是一种异质性疾病, 分子亚型不同的患者在疾病进展及预后上有较大差异^[2]。因此, 早期诊断和有效治疗是降低胃癌病死率的有效途径之一。建立转录调控网络, 发现与预后相关的转录因子将有助于理解胃癌潜在的调控机制, 为探索胃癌的治疗方案提供线索及证据。

1 资料与方法

1.1 芯片数据下载及预处理 从 NCBI 基因表达综合数据库 (GEO, <http://www.ncbi.nlm.nih.gov/gds>) 搜索并下载 3 项与胃癌发生及预后相关的表达谱芯片 (GSE66229、GSE26253、GSE15459) 原始研究数据。GSE66229 包含 300 个胃癌组织标本, 每个标本包含 54 675 个基因测试值; GSE15459 包含 200 个胃癌组织标本且包含预后信息, 每个标本包含 21 025 个基因测试值; GSE26253 包含 432 个胃癌组织标本且包含预后信息, 每个标本包含 17 418 个基因测试值。所有研究的检测平

台均为 Affymetrix Plus2.0。在 R 3.0.1 环境下, 原始数据 (.cel 格式) 首先用 fRMA 提取表达谱数据, 基于 barcode 文件剔除研究中均不表达的探针^[3], 再进一步基于 Jetset 文件选择标记为“Perfect”的探针^[4], 最终留存 9 384 个探针深入分析。

1.2 预后相关转录因子分析 GSE66229 和 GSE15459 用于建立核心网络及寻找与胃癌患者预后相关的转录因子, GSE26253 用于验证转录因子对胃癌患者预后是否有预测作用。采用上述预处理完的 GSE66229 数据, 以 GSVA 软件包进行差异转录因子的筛选及差异大小排序。采用上述预处理完的 GSE15459 数据, 以 survJmda 软件包进行 COX 单因素风险比例模型初筛分析^[5], 筛选出与胃癌患者预后相关的转录因子; 采用相同方法对上述预处理完的 GSE26253 数据进行预后相关转录因子初筛分析。

1.3 胃癌转录因子-靶基因调控网络的构建及可视化 采用 RTN 软件包网络分析方法, 比对目前已知的 1 727 个转录因子, $P=0.000\ 001$, $n\text{Permutations}=1\ 000$, 构建胃癌转录因子网络数据。选择 GSE66229 和 GSE15459 共表达分析有统计学意义 ($P<0.05$)、相关系数绝对值大于 0.6 的基因对, 在 Cy-

* 基金项目: 上海市卫生和计划生育委员会科研课题资助项目 (201640017)。
作者简介: 康争春, 男, 在读博士, 主要从事胃癌、结直肠癌方面研究。 △ 通信作者, E-mail: dryuenda@gmail.com。

toscape3.1.0 软件平台获得 2 组研究一致存在的共表达网络。

1.4 统计学处理 采用 SPSS20.0 进行数据处理。计数资料采用例数表示。 P 、 Q 以 $\alpha<0.05$ 为差异有统计学意义。

2 结 果

2.1 GSE66229 核心亚网络转录因子 通过 GSE66229,采用 R 语言编程、RTN 方法,共发现 4 个核心网络、21 个核心转录因子。见表 1。

表 1 GSE66229 核心亚网络转录因子			
转录因子	基因数(n)	转录因子	基因数(n)
TGFB11	127	LIN28B	23
NR1I2	115	CDX2	22
HLX	90	FOXA1	21
PRRX1	89	SOX11	17
BCL6	62	DUS3L	16
NAP1L3	62	SSH1	16
PLXND1	57	PDLIM4	15
ELK3	31	ZNF488	14
NOTCH3	30	PARP12	13
ZEB1	29	RORA	10
DZIP1	25		

2.2 GSE15459 筛选和预后相关转录因子 采用 GSE15459 结合预后信息发现共有 10 个转录因子和患者预后相关。见表 2。

表 2 GSE15459 筛选和预后相关转录因子			
转录因子	P	Q	Cor 偏差
BCL6	0.000 000 218	0.000 450 361	0.331 492 240
ZNF593	0.000 000 408	0.000 450 361	0.323 128 062
RORA	0.000 000 642	0.004 728 544	0.282 795 589
PLXND1	0.000 242 319	0.013 379 840	0.260 804 061
DSP	0.000 528 371	0.023 339 532	0.246 901 452
HIF1A	0.000 937 420	0.028 944 192	0.236 128 453
CHD7	0.001 023 298	0.028 944 192	0.234 436 532
NAP1L3	0.001 048 403	0.028 944 192	0.233 966 492
TRIB3	0.001 712 913	0.039 787 187	0.224 231 942
SNAPC1	0.001 801 441	0.039 787 187	—

注:—表示无数据

2.3 GSE26253 筛选和预后相关转录因子 结合转录因子核心网络,已确定和预后相关的 4 个核心转录因子,分别为 BCL6、PLXND1、NAP1L3、RORA,均位于 subnet1,因此确定对胃癌患者预后影响的主要调控网络为 subnet1。采用 GSE26253 验证,发现其中 3 个转录因子预测效果明显,BCL6、NAP1L3、PLXND1 均和患者预后相关($P<0.05$ 、 $Q<0.05$)。见表 3、4。

表 3 GSE26253 筛选和预后相关转录因子			
转录因子	P	Q	Cor 偏差
BCL6	0.001 317 249	0.017 505 230	0.153 854 726
DUS3L	0.014 326 674	0.048 764 434	−0.117 689 298
DZIP1	0.014 525 852	0.048 764 434	0.117 452 320
ELK3	0.000 583 229	0.010 334 232	0.164 541 252
FOXA1	0.001 859 802	0.019 716 526	0.149 116 617
LIN28B	0.013 938 359	0.048 764 434	0.118 159 746
NAP1L3	0.003 144 435	0.024 490 559	0.141 625 958

续表 3 GSE26253 筛选和预后相关转录因子			
转录因子	P	Q	Cor 偏差
NOTCH3	0.002 101 832	0.019 716 526	0.147 402 559
NR1I2	0.013 137 632	0.048 722 544	−0.119 167 374
PLXND1	0.000 213 155	0.004 798 336	0.176 896 732
PRRX1	0.000 000 646	0.001 030 812	0.214 395 984
SOX11	0.000 005 170	0.002 660 122	0.192 986 664
SSH1	0.000 240 713	0.004 798 336	0.175 448 865
ZEB1	0.001 684 958	0.019 716 526	0.150 486 838

表 4 validation 转录因子			
TFs	n	P	Q
BCL6	62	<0.05	0.017
NAP1L3	62	<0.05	0.024
PLXND1	57	<0.05	0.004

3 讨 论

相关报道中,学者发现不同分子亚型的患者在疾病进展及预后上有较大差异^[6],亦发现胃癌的大量致癌基因,并根据致癌基因尝试了靶向治疗的探索^[7-10]。在细胞信号通路中,人类的致癌基因远多于转录因子,且发现数量有限的转录因子在大多数人类肿瘤细胞中过度活跃。因此,学者认为转录因子是最直接、最有希望的癌症治疗靶点,转录因子成为了抗癌药物的发展目标。

人类基因组约编码 2 000 个转录因子。转录因子又称为反式作用因子,是控制基因转录活性的主要分子。这些 DNA 结合蛋白能够与真核基因启动子区域中顺式作用元件发生特异性相互作用,通过它们与其他相关蛋白间的相互作用,激活或抑制转录。通过确保正确表达特定基因,转录调控系统在众多生物学控制过程中起着核心作用,包括细胞周期的进展、细胞内代谢和生理平衡、细胞不同阶段的分化及发育。众多疾病的发生是由于调控系统出现异常,转录因子在致癌基因中过表达。约 1/3 的人类发育障碍是由转录因子异常导致,其中包括癌症和发育障碍综合征。

研究胃癌诊断及预后相关的转录因子须具有较大的样本量。然而,由于基因表达谱芯片价格昂贵,很少有单个实验室能够使用大量基因芯片进行胃癌差异转录因子的筛选。因此,开发新的数据分析和整合策略对组学数据挖掘、发现重要甚至全新的胃癌相关关键分子十分重要。本研究中,笔者首先采用既往样本量较大、基因测试值较多的基因表达谱芯片数据构建差异表达转录因子在癌组织中的共表达网络,进而发现共表达网络中的核心亚网络,然后采用具有预后数据的合适表达谱芯片,根据构建的网络筛选预后相关调控的转录因子,最后采用样本量较大、有预后数据的表达谱芯片验证转录因子是否对胃癌患者预后有预测作用。

本研究中,通过 GSE66229 基因数据集,采用 R 语言编程、RTN 方法,共发现 4 个核心网络、21 个核心转录因子。采用 GSE15459 结合预后信息发现,共有 10 个转录因子和患者预后相关。4 个核心转录因子均位于 subnet1,故确定对胃癌患者预后影响的主要调控网络为 subnet1。采用 GSE26253 验证,发现其中 3 个转录因子的预测效果明显,BCL6、NAP1L3、PLXND1 均和患者预后相关($P<0.05$ 、 $Q<0.05$)。上述分析不仅扩增了研究的样本量,也找到了不同实验室之间可重复的

研究结果,明显提高了胃癌预后相关转录因子的筛选效率。

BCL6 是一种对肿瘤起抑制作用的转录因子,本研究发现并验证了转录因子 BCL6 与胃癌患者的预后密切相关。目前普遍认为 BCL6 是一种在淋巴细胞分化过程中、对淋巴瘤有重要作用的转录抑制因子,但胃癌中关于 BCL6 的相关报道及其少见。有学者发现 BCL6 可以抑制胃癌细胞株中 Cyclin D2 的表达,且 BCL6 的表达与胃癌的组织学分型相关,提示 BCL6 可能在胃癌调控机制中起重要作用,表达下降可能促进胃癌的增殖、分化、侵袭。另有学者通过检测包括 BCL6 等 5 个基因签名表达情况,预测消化道肿瘤的炎症反应、癌前病变、癌的转变、进展状况,提示 BCL6 在胃癌的发生、进展过程中可能起重要调节作用,但具体调节机制有待于进一步深入研究。

PLXND1 是一种细胞表面受体,在细胞迁移、增殖和生存等方面起重要作用。目前研究发现,PLXND1 在心血管、神经和免疫系统对癌症的发生、发展均发挥重要作用^[11]。有学者发现 PLXND1 与胃癌的发生、发展、转移相关^[12],这与本研究结果一致,但是机制仍未明确,需要未来试验的探索和证明。

NAP1L3 是一种核小体组装蛋白,研究发现其与哺乳动物的神经系统发育有关^[13]。目前在胃癌中并未发现关于 NAP1L3 的相关研究,但有学者发现 NAP1L3 可明显促进乳腺癌的转移,NAP1L3 高表达的乳腺癌患者生存期明显缩短^[14]。本研究中,NAP1L3 升高与胃癌患者不良预后相关,笔者预测其可能在胃癌中起相同作用。

综上所述,本研究通过生物信息学方法,探索、发现并验证了与胃癌预后相关的转录因子,构建了转录调控网络,为研究胃癌的潜在调控机制提供了参考,并为将来胃癌的诊断提供可能的分子标志物,为胃癌的治疗提供可能的转录因子靶点。

参考文献

[1] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015[J]. CA Cancer J Clin, 2016, 66(2): 115-132.

[2] Deng N, Liu JW, Sun LP, et al. Expression of XPG protein in the development, progression and prognosis of gastric cancer[J]. PLoS One, 2014, 9(9): e108704.

[3] Mccall MN, Bolstad BM, Irizarry RA. Frozen robust multiarray analysis (fRMA) [J]. Biostatistics, 2010, 11(2): 242-253.

[4] Li QL, Birkbak NJ, gyorffy B, et al. Jetset: selecting the

optimal microarray probe set to represent a gene [J]. Bioinformatics, 2011, 12(1): 474-476.

[5] Yasrebi H. SurvJamda: an R package to predict patients' survival and risk assessment using joint analysis of microarray gene expression data [J]. Bioinformatics, 2011, 27(8): 1168-1169.

[6] Cristescu R, Lee J, Nebozhyn M, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes [J]. Nat Med, 2015, 21(5): 449-456.

[7] Ema A, Yamashita K, Ushiku H, et al. Immunohistochemical analysis of RTKs expression identified HER3 as a prognostic indicator of gastric cancer [J]. Cancer Sci, 2014, 105(12): 1591-1600.

[8] Shi HB, Ji M, Wu J, et al. Serum B7-H4 expression is a significant prognostic indicator for patients with gastric cancer [J]. World J Surg Oncol, 2014, 12(1): 1-3.

[9] Kwon CH, Park HJ, Lee JR, et al. Serpin peptidase inhibitor clade A member 1 is a biomarker of poor prognosis in gastric cancer [J]. Br J Cancer, 2014, 111(10): 1993-2002.

[10] Ivanova T, Zouridis H, Wu YH, et al. Integrated epigenomics identifies BMP4 as a modulator of cisplatin sensitivity in gastric cancer [J]. Gut, 2013, 62(1): 22-33.

[11] Gay CM, Zygmunt T, Torres-Vazquez J. Diverse functions for the semaphorin receptor PlexinD1 in development and disease [J]. Dev Biol, 2011, 349(1): 1-19.

[12] Maejima R, Tamai K, Shiroki T, et al. Enhanced expression of semaphorin 3E is involved in the gastric cancer development [J]. Int J Oncol, 2016, 49(3): 887-894.

[13] Attia M, Foerster A, Rachez C, et al. Interaction between Nucleosome Assembly Protein 1-like Family Members [J]. J Mol Biol, 2011, 407(5): 647-660.

[14] Ross JB, Huh D, Noble LB, et al. Identification of molecular determinants of primary and metastatic tumour re-initiation in breast cancer [J]. Nat Cell Biol, 2015, 17(5): 651-658.

(收稿日期: 2017-05-11 修回日期: 2017-07-18)

(上接第 3141 页)

attachment levels using a Florida probe versus traditional probes [J]. Med J Armed Forces India, 2015, 71(4): 352-358.

[3] Weidmann B, Sahrman P, Bindl A, et al. Depth determination of artificial periodontal pockets using cone-beam tomography and radio-opaque material: an in vitro feasibility study [J]. Swiss Dent J, 2014, 124(4): 406-415.

[4] Braun X, Ritter L, Jervoe-Storm PM, et al. Diagnostic accuracy of CBCT for periodontal lesions [J]. Clin Oral Investig, 2014, 18(4): 1229-1236.

[5] Radu C, Mihaela H, Pompei B, et al. The evaluation of 20

MHz ultrasonography, computed tomography scans as compared to direct microscopy for periodontal system assessment [J]. Med Ultrason, 2011, 13(2): 120-126.

[6] Zimbran A, Dudea S, Dudea D. Evaluation of periodontal tissues using 40 MHz ultrasonography. preliminary report [J]. Med Ultrason, 2013, 15(1): 6-9.

[7] Bartosz S, Andrew D, Wojciech B, et al. Assessment of gingival thickness using an ultrasonic dental system prototype: a comparison to traditional methods [J]. Ann Anat, 2015, 199(9): 98-103.

(收稿日期: 2017-05-09 修回日期: 2017-07-20)