

肝细胞癌术后生存率相关分子在复杂网络中的拓扑特性研究

单光宇, 卢一鸣, 屈武斌, 张成岗

[摘要] **目的** 从系统层面探索基因和微小 RNA (microRNA, miRNA) 的网络结构, 结合生存分析技术, 研究生存率相关分子的拓扑特性。**方法** 利用肝细胞癌的基因和 miRNA 表达谱数据构建调控网络, 并研究节点中患者的生存信息。**结果** 对基因共表达网络节点的度进行全局计算发现, 整个网络节点的度服从幂律分布, 在一个网络中, 度相对较高的节点称为“集线器 (Hub)”, 与普通节点相比, 这些 Hub 节点中与生存相关的基因更为富集。**结论** 基因调控网络中的 Hub 节点, 有望作为分子分型的潜在特征。

[关键词] 肝细胞癌; 外科手术; 复杂分子网络; 生存分析; 基因表达; miRNA

[中图分类号] R735.7

[文献标志码] A

[文章编号] 1674-9960(2015)--

DOI: 10.7644/j.issn.1674-9960.2015.00.000

Topological characteristics of postoperative survival rate-related molecules in hepatocellular carcinoma-associated complex networks

SHAN Guang-yu, LU Yi-ming*, QU Wu-bin, ZHANG Cheng-gang*

(Institute of Radiation Medicine, Academy of Military Medical Sciences, State Key Laboratory of Proteomics, Key Laboratory of Military Cognition and Mental Hygiene, Beijing 100850, China)

* Co-corresponding authors, LU Yi-ming, E-mail: luyim@outlook.com; ZHANG Cheng-gang, E-mail: zhangcg@bmi.ac.cn

[Abstract] **Objective** To explore the network structure of genes and microRNA (miRNA) at the system level and to study topological characteristics of survival rate related molecules by means of survival analysis. **Methods** Regulatory networks were established with genes of hepatocellular carcinoma (HCC) and with the expression profiling data of miRNA. Survival knowledge in nodes was studied. **Results** All the nodes in gene co-expression networks conformed to power-law distribution. In a network, high-degree nodes were called "Hub", which enriched more survival related genes than lower degree ones. **Conclusion** Hub nodes in gene regulatory networks promise to be potential features for molecular subtyping.

[Key words] hepatocellular carcinoma; surgical procedures, operative; complex molecular network; survival analysis; gene expression; miRNA

肝细胞癌是主要恶性肿瘤之一, 同时也是第三位致死性肿瘤, 全世界每年有 > 100 万人受其影响^[1]。肝细胞癌的致病机制异常复杂, 常常是很多基因、微小 RNA (miRNA)、外界环境、生活习惯等诸多因素共同作用的结果。早期诊断后通过切除、肝移植或射频消融等方法及时治疗可使 5 年生存率 > 50%。但是不同患者的预后却相差甚远, 短则数月, 长可达 10 年之久^[2]。深入理解其生物学机制对早期肝细胞癌的诊断、治疗和预防至关重要。

作为一类典型的复杂性疾病, 目前的形态学分类和病理学分型很难分析和理解这类差异性的原

因, 病症的发生发展通常是多个分子共同作用的结果^[3]。因此, 我们结合复杂分子网络技术, 从系统层面探究这些分子的拓扑特性与患者生存状态的关系。

微小 RNA (microRNA, miRNA) 是一类非编码小 RNA, 通常通过靶定信使 RNA (mRNA) 下调甚至封锁其表达^[4]。近年来, 大家已广泛认识到 miRNA 能通过调控基因表达干扰多类细胞过程与疾病发生^[5]。在癌症中常会发现多个基因的突变以及 miRNA 的失调, 肝癌也不例外^[6]。在另一些研究中还发现, miRNA 和基因对肿瘤的表型也具有调控作用, 并且认为这种表达谱能用来实现癌症分型^[7]。

然而, 目前并没有通过复杂网络的方法进行刻画分型特征的先例。为了研究利用网络拓扑结构数据进行特征刻画的可能性, 我们构建了 2 个分子网络, 并对节点中生存相关分子的富集程度进行计算, 发现其中集线器 (Hub) 节点具有显著富集的趋势, 表明的确有可能使用 Hub 节点作为分型的一个潜在特征。

[基金项目] 国家自然科学基金项目 (31401141, 31371345); 蛋白质组学国家重点实验室优秀青年课题 (SKLP-Y201303) 资助项目

[作者简介] 单光宇, 男, 硕士研究生, 研究方向: 生物信息学, E-mail: sci@shanguangyu.com。

[作者单位] 军事医学科学院放射与辐射医学研究所, 北京 100850

[通讯作者] 卢一鸣, E-mail: luyim@outlook.com; 张成岗, E-mail: zhangcg@bmi.ac.cn

1 材料与方法

1.1 数据集

使用公共数据库癌症基因组图谱(The Cancer Genome Atlas, TCGA, <http://cancergenome.nih.gov/>)作为数据源,下载肝细胞癌的 377 例样本的全部数据,包括临床生存数据、miRNA 表达谱数据、基因表达谱数据等。使用 Python 自编写程序提取临床生存数据、基因表达谱数据以及 miRNA 表达谱数据作为原始数据,共涉及 20 501 个基因、1046 个 miRNA。临床生存数据中包含性别、年龄、种族、生存状态、死亡时间、最后随访时间等信息。

1.2 方法

首先,构建基因共表达网络。与以往研究一致^[8,9],采用 $|\text{Spearman coefficient}| > 0.8$ 作为一对基因具有强相互作用关系的阈值,并考虑了基因互作的正相关性和负相关性两种可能性。对于 miRNA 与基因相互作用关系,先利用 TargetScan 数据库提取具有序列互补关系的 miRNA-基因关系对^[5],然后计算出 $P < 0.05$ 时的 $\text{Spearman coefficient} = -0.125$,提取出所有小于该系数具有负调控关系的基因和 miRNA 关系对作为 miRNA-基因相互作用网络源数据。然后利用 Cytoscape 软件构建出两个互作网络:基因共表达(co-expression)网络以及 miRNA-基因互作网络^[10];其次,以 $P < 0.05$ 为阈值对所有基因和 miRNA 进行生存分析,提取出患者术后生存时间显著相关基因和 miRNA。绘制生存时间显著相关基因度分布图及总体基因度分布图,并研究其分布特点;最后,计算 Hub 节点中生存相关基因的富集率,讨论后续以 Hub 节点作为潜在分型的特征的可能性。

本方法融合了肝细胞癌的网络信息和生存信息两类重要信息,不仅可探讨与生存密切相关的基因和 miRNA 的网络属性,更有助于从系统角度深化对肝细胞癌生物学机制的理解,为更有效进行癌症分型、药靶研发打下基础。

2 结果

利用基因表达谱数据进行 Spearman 相关性计算,提取相关基因并绘制基因共表达网络图(图 1)。该网络包括 2626 个基因的 67 303 个互作关系。

利用基因表达和 miRNA 表达谱数据进行 Spearman 相关性计算,提取负调控关系绘制 miRNA-基因互作网络图(图 2)。该网络包括 1601 个基因与 148 个 miRNA 的 3592 个相互作用关系。

显然,两个调控网络具有极其明显的“小世界”

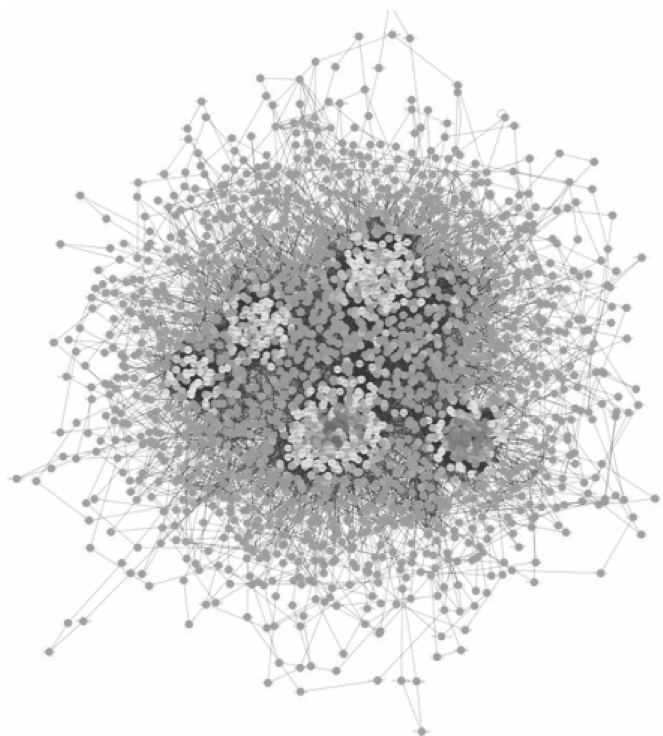


图 1 肝细胞癌相关的基因共表达网络

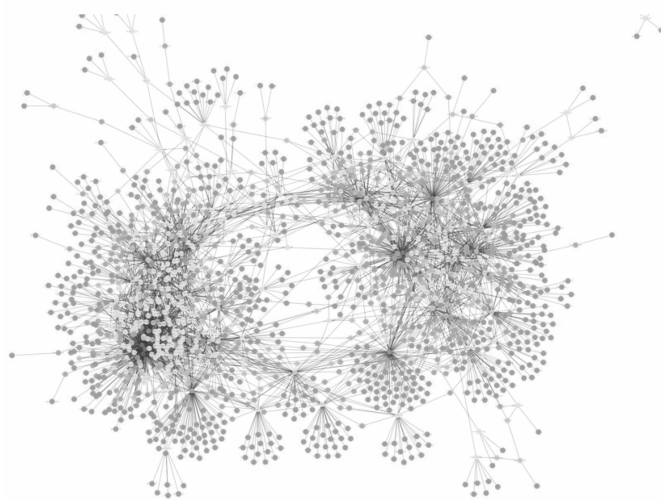


图 2 miRNA-基因互作网络

属性^[11]。基因共表达网络尤其明显,平均路径长度较短,聚类系数极高,大量节点都与少数几个 Hub 节点相连。

为了研究这些 Hub 节点与肝细胞癌生存数据的关系,我们挑选基因共表达网络,绘制了全局度分布图以及与生存紧密相关基因的度分布图(图 3)。

无论是全局度分布(图 3A),还是生存相关基因的度分布(图 3B),在绝对数量上,明显可见低度数基因完全占据主导地位。然而,有趣的是,在图 3B 中,生存相关基因在 Hub 节点中出现的数量似乎不减反增。为了验证是否存在这样的相对富集作用,我们构建了打分 Z-score, $Z\text{-score} = \log_2(R_{\text{survival}}/R_{\text{all}})$, R_{survival} 表示生存相关基因每个度区间基因个数占总体生存相关基因个数的比例, R_{all} 表示全局基因每个度区间基因个数占总体全局基因个数的比

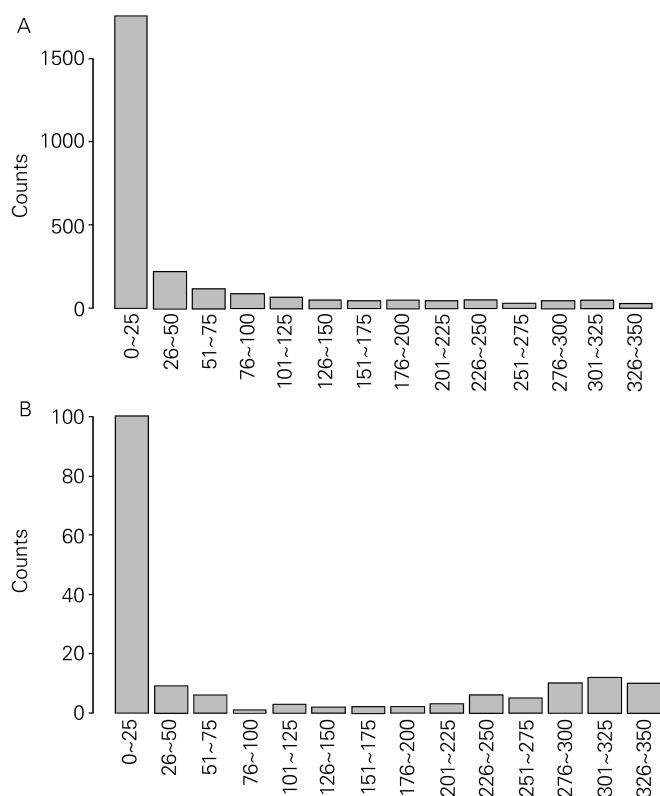


图3 全局度分布及生存相关基因度分布

例。该打分可用来描述生存相关基因在不同度数节点中的富集程度,富集分析结果见图4。在同一个度区间中, $Z\text{-score} > 0$ 表示生存相关基因的数量占比大于该区间中全局基因占比, $Z\text{-score} < 0$ 表示生存相关基因的数量占比小于该区间中全局基因占比, $Z\text{-score} = 0$ 表示生存相关基因的数量占比等于该区间中全局基因占比。

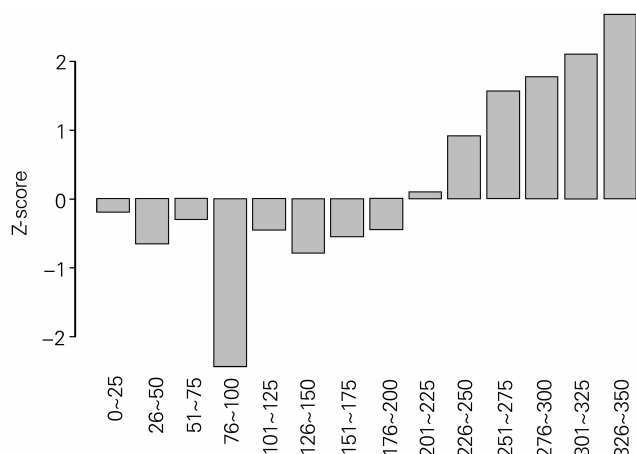


图4 对不同度数节点中生存相关基因的富集分析

可以看出,在 Hub 节点中,生存相关基因的数量大量富集,远高于低度数节点的富集率。这个现象为以后的肝癌分型提供了新的思路,使用 Hub 节点作为主要特征进行刻画很有可能成为提高分型准确率的新突破口。

3 讨论

肝细胞癌是一类典型的复杂性疾病,常用的连

锁分析、定位克隆以及关联基因等单基因病研究策略,在研究复杂疾病时存在一定的局限性。因此,我们从系统生物学的角度,借助 TCGA 公共数据库,探讨了与生存紧密相关的基因以及 miRNAs 的网络特性,为后续进行复杂分子网络的特征刻画、分子分型、预后诊断打下基础。

从逻辑上来说,基因和 miRNAs 在肝细胞癌的发生发展中总是处在一个不停的相互作用之中。为了从更加系统的层面理解肝细胞癌的发生发展,我们结合了复杂网络拓扑特性与患者生存信息进行组合研究,发现网络中的 Hub 节点很有可能作为一类新的潜在分型特征。

另外,本方法流程具有一定的普适性,对于研究其他癌症也有相当的参考价值。下一阶段我们会使用这种方法研究其他癌症中 Hub 节点生存相关基因的富集程度,并利用这个特性刻画出一个特征来对部分癌症进行分型,希望能够为癌症分型提供更科学、更符合逻辑且更加可靠的方法。

【参考文献】

- [1] Damania P, Sen B, Dar SB, *et al.* Hepatitis B virus induces cell proliferation via HBx-induced microRNA-21 in hepatocellular carcinoma by targeting programmed cell death protein4 (PDCD4) and phosphatase and tensin homologue (PTEN) [J]. *PLoS One*, 2014, 9(3): e91745.
- [2] Forner A, Llovet JM, Bruix J. Hepatocellular carcinoma [J]. *Lancet*, 2012, 379(9822): 1245 - 1255.
- [3] Huang S, He X. The role of microRNAs in liver cancer progression [J]. *Br J Cancer*, 2011, 104(2): 235 - 240.
- [4] Yang H, Kong W, He L, *et al.* MicroRNA expression profiling in human ovarian cancer; miR-214 induces cell survival and cisplatin resistance by targeting PTEN [J]. *Cancer Res*, 2008, 68(2): 425 - 433.
- [5] Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets [J]. *Cell*, 2005, 120(1): 15 - 20.
- [6] Bentwich I, Avniel A, Karov Y, *et al.* Identification of hundreds of conserved and nonconserved human microRNAs [J]. *Nature Genet*, 2005, 37(7): 766 - 770.
- [7] Kim HS, Lee KS, Bae HJ, *et al.* MicroRNA-31 Functions as a tumor suppressor by regulating cell cycle and epithelial-mesenchymal transition regulatory proteins in liver cancer [J]. *Oncotarget*, 2015, 6(10): 8089 - 8102.
- [8] Kauffman EC, Robinson BD, Downes MJ, *et al.* Role of androgen receptor and associated lysine-demethylase coregulators, LSD1 and JMJD2A, in localized and advanced human bladder cancer [J]. *Mol Carcinog*, 2011, 50(12): 931 - 944.
- [9] Balanescu S, Kopp P, Gaskill MB, *et al.* Correlation of plasma copeptin and vasopressin concentrations in hypo-, iso-, and hyperosmolar States [J]. *J Clin Endocrinol Metab*, 2011, 96(4): 1046 - 1052.
- [10] Smoot ME, Ono K, Ruscheinski J, *et al.* Cytoscape 2.8: new features for data integration and network visualization [J]. *Bioinformatics*, 2011, 27(3): 431 - 432.
- [11] Amaral LA, Scala A, Barthelemy M, *et al.* Classes of small-world networks [J]. *Proc Natl Acad Sci USA*, 2000, 97(21): 11149 - 11152.