

利用蛋白互作网络分析挖掘肝癌分子机制

孙孝国,李力,王金华

(泰安市第一人民医院,山东 泰安 271000)

摘要: [目的] 从公共数据库收集大量肝癌候选癌基因,利用蛋白互作网络分析,挖掘肝癌分子机制。[方法] 收集 Liverome 数据库内肝癌候选癌基因,使用 String 数据构建肝癌候选癌基因的蛋白互作网络,进一步利用 MCODE 算法寻找定义网络中核心基因模块,然后通过功能富集分析探究模块相关的生物学功能与通路。[结果] 共收集 615 个肝癌候选癌基因,蛋白互作网络分析得到一个包含 486 个基因以及 2076 个交互关系的蛋白互作网络。同时,通过 MCODE 算法得到 6 个核心基因模块,这些模块分别富集到“细胞周期”、“药物代谢”、“DNA 复制与修复”等特定的生物学功能。[结论] 这些核心基因模块以及相关功能与肝癌发生发展显著相关。

关键词: 肝癌;基因互作网络;网络模块;分子机制

中图分类号: R735.7 **文献标识码:** A **文章编号:** 1004-0242(2015)10-0871-04

doi: 10.11735/j.issn.1004-0242.2015.10.A015

Explore the Molecular Mechanism of Liver Cancer by Protein-protein Interaction Network Analysis

SUN Xiao-guo, LI Li, WANG Jin-hua

(Tai'an First People's Hospital, Tai'an 271000, China)

Abstract: [Purpose] To collect abundance candidate liver cancer genes from common database, and to explore the molecular mechanism of liver cancer by protein-protein interaction network analysis. [Methods] The candidate cancer genes of liver cancer from Liverome database were collected, the protein-protein interaction network (PPIN) of candidate liver cancer were constructed by String database, then the hub gene network modules from the PPIN were analyzed by using MCODE software, then functional enrichment analysis were used to find the related functions of these hub gene network modules. [Results] A total of 615 candidate liver cancer genes were collected, and PPIN including 486 genes and 2076 interaction relationships were found by gene interaction network analysis. Moreover, 6 hub gene network modules were obtained by MCODE software, these modules were enriched in certain function clusters such as cell cycle, drug metabolism, DNA replication and repair. [Conclusion] Hub gene modules and corresponding functions are significant related to liver cancer.

Key words: liver cancer; protein-protein interaction network; network module; molecular mechanism

肝癌是我国常见恶性肿瘤之一,在我国每年约有 11 万人死于肝癌,占全世界肝癌死亡人数的 45%。研究表明原发性肝癌的预后比较差的主要原因是缺乏特异的早期诊断指标和治疗手段,而研究肝癌的发病机制有助于提高肝癌的早期诊断率,指导肝癌的正确治疗^[1,2]。目前研究认为,肝癌发生分子机制主要是由于多种癌基因变异的积累,继而进一步影响细胞内与肿瘤发展相关的信号通路,包括

大量的致癌基因、抑癌基因以及 DNA 修复基因的突变。这些基因的激活或者抑制直接或间接的影响细胞周期、增殖、凋亡等生物学功能^[3]。尽管基因芯片技术及新一代测序等技术的应用,已经揭示了很多与肝癌相关的分子以及调控信息。如何有效利用这些高通量技术积累的数据去探究肝癌以及用于临床实践仍需要探索。

肝癌发生发展过程中,肿瘤相关基因是一个相互调控相互关联的动态网络过程^[4]。随着分子生物学的进一步发展,积累了大量的蛋白质互作网络

收稿日期:2015-04-27;修回日期:2015-06-05

通讯作者:孙孝国, E-mail: xiaoguosun@163.com

(protein-protein interaction network, PPIN)数据,且这些 PPIN 数据也开始用于构建肿瘤等疾病的基因调控网络,取得一定的成功^[5,6]。据此,本研究收集 Liverome 数据库^[7]中积累的大量肝癌基因数据,尝试利用蛋白质互作网络分析寻找基因互作核心模块,并通过功能富集分析进一步确认这些网络模块的生物学功能,从全局上挖掘和探究肝癌的分子机制。

1 资料与方法

1.1 Liverome 数据库以及肝癌候选癌基因

Liverome 数据库^[7]主要收集目前已发表的高通量基因表达谱实验以及蛋白质谱实验的结果,通过重新分析和专家评定,筛选出与肝癌相关的基因或者蛋白质。目前该数据库已经收集 98 个高通量技术实验的数据,共涉及 6927 个基因。本研究取超过 5 个高通量实验报道的基因定义为肝癌候选癌基因,共收集基因 615 个。

1.2 基因互作网络分析

使用 String 9.1 数据库^[8]分析所有肝癌候选癌基因之间存在的蛋白互作关系。String 9.1 数据库是一个收集蛋白质互作关系、基因调控关系、以及文献挖掘分析和蛋白质共表达分析等计算得到物理互作和预测互作关系的数据库。且可通过算法给不同基因(蛋白质)之间的互作关系打分,并建立蛋白互作网络。本研究 615 个肝癌候选基因之间的蛋白互作关系打分阈值设定在 8.0。

1.3 基因(蛋白)互作网络核心模块以及模块功能富集分析

基因(蛋白)互作网络核心模块的定义:模块中蛋白间的互作关系显著多,而模块外蛋白与模块内蛋白的互作关系显著地少,采用 MCODE 软件^[9]计算网络中核心互作网络模块(Score>2)。同时,使用 DAVID 数据库^[10]将分析得到的核心模块进行基因功能富集分析。

2 结果

2.1 肝癌候选癌基因

我们收集了 615 个肝癌候选癌基因,此 615 个基因中有 169 个基因属于已知癌基因,已知癌基因

数据库来源于癌基因普查数据库^[11]。同时,我们使用 BiNGO 软件^[12]对这 615 基因进行功能富集分析(GOslim_generic)(Table 1),如 Table 1 所示,基因大量富集到“核酸和染色体绑定”、“细胞周期”、“抗氧化活性”、“细胞骨架”等与肿瘤发生发展有密切关系的功能中^[13-15]。

Table 1 The functional enrichment analysis of candidate liver cancer genes

Biological function	Gene	P value
Nucleotide binding	21	1.36E-05
Catalytic activity	35	1.61E-05
Chromatin binding	6	3.12E-05
Protein complex	22	1.46E-03
Cell cycle	13	2.04E-03
Carbohydrate metabolic process	9	2.05E-03
Response to stress	14	2.65E-03
Antioxidant activity	3	2.73E-03
Regulation of gene expression	5	3.10E-03
Actin binding	3	6.76E-03
Nuclear chromosome	7	9.47E-03
Cytoskeleton organization	6	1.32E-02

2.2 基因互作网络与核心网络模块分析

为进一步考察这些基因之间存在的相互作用关系,本研究使用 String 数据库对 615 个基因进行基因(蛋白质)互作网络分析,结果发现,这些基因形成了一个包含 486 个基因以及 2076 个基因互作关系的网络。这 486 个基因包含 143 个已知癌基因。

为了进一步挖掘网络中与肝癌相关的基因(蛋白质)互作网络模块。我们利用 MCODE 软件^[9]分析该互作网络,定义网络中的核心模块,分析结果得到 6 个基因模块(Figure 1, Table 2)。模块 1 包含 22 个基因涉及 213 个互作关系(MCODE score=12.11);模块 2 包含 14 个基因涉及 91 个互作关系(MCODE score=8.34);模块 3 包含 13 个基因涉及 75 个互作关系(MCODE score=7.61);模块 4 包含 8 个基因涉及 28 个互作关系(MCODE score=4.30);模块 5 包含 10 个基因涉及 29 个互作关系(MCODE score=3.83);模块 6 包含 7 个基因涉及 17 个互作关系(MCODE score=3.28)。

使用 DAVID 数据库针对 6 个模块分别进行功能富集发现,这 6 个模块分别可以富集到某些特定生物学功能类之中,例如模块 1 富集到“细胞周期”相关的功能;模块 2 富集到“炎症反应”相关的功能;

Table 2 The functional enrichment analysis of the PPI hub network modules

Module	Database	GO and pathway	Gene	P value
Module 1	KEGG	Cell cycle	7	3.79E-08
		Progesterone-mediated oocyte maturation	4	5.14E-04
	GO	Cell cycle process	16	5.63E-18
		Cell cycle	17	9.66E-18
		Nuclear division	12	1.09E-15
Module 2	KEGG	Complement and coagulation cascades	7	1.37E-10
	GO	Response to wounding	9	5.70E-09
		Acute inflammatory response	6	2.21E-08
		Inflammatory response	7	2.73E-07
Module 3	KEGG	Drug metabolism	12	4.15E-21
		Linoleic acid metabolism	6	2.68E-09
		Retinol metabolism	6	8.37E-08
	GO	Drug metabolic process	4	1.62E-07
		Aromatic compound catabolic process	2	0.014691
		Xenobiotic metabolic process	2	0.014691
Module 4	KEGG	Allograft rejection	2	0.028027
		Graft-versus-host disease	2	0.030336
		Autoimmune thyroid disease	2	0.03953
		Antigen processing and presentation	2	0.063727
	GO	Immune response	3	0.023436
		Antigen processing and presentation	2	0.030307
Module 5	KEGG	Arginine and proline metabolism	6	1.24E-08
		Alanine, aspartate and glutamate metabolism	5	1.39E-07
		Cysteine and methionine metabolism	3	0.001517
		Glutathione metabolism	3	0.003264
		Phenylalanine metabolism	2	0.038301
		Tyrosine metabolism	2	0.075292
	GO	Cellular amino acid biosynthetic process	6	7.76E-11
		Glutamate metabolic process	4	1.14E-07
		Cellular amino acid catabolic process	4	9.99E-06
		Carboxylic acid catabolic process	4	4.36E-05
Module 6	KEGG	DNA replication	5	2.12E-09
		Base excision repair	2	0.027257
	GO	DNA replication	5	5.53E-07
		DNA-dependent DNA replication	4	1.48E-06
		DNA replication initiation	3	1.96E-05
		DNA metabolic process	5	2.73E-05

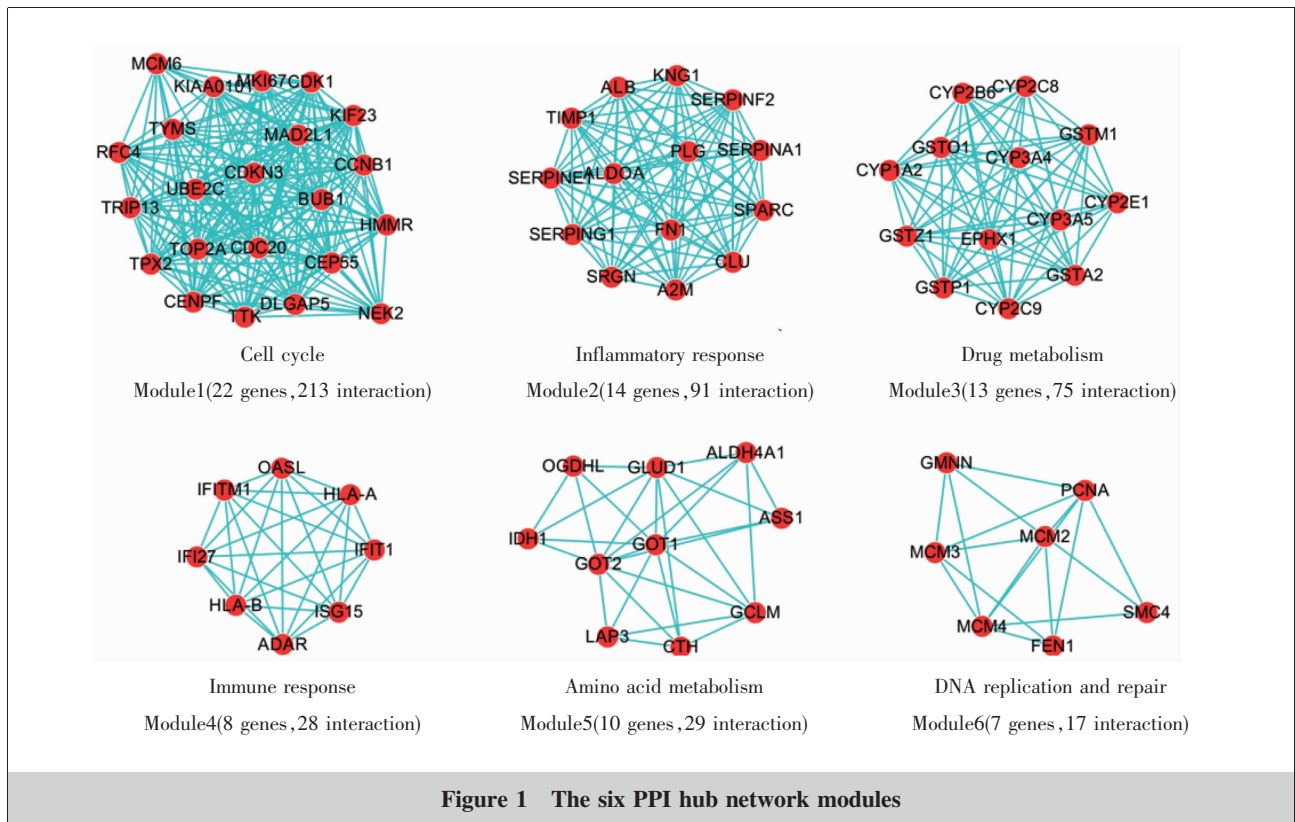
模块 3 富集到“药物以及化合代谢”相关功能;模块 4 富集到“免疫反应以及自身免疫疾病”相关功能;模块 5 富集到“各种氨基酸代谢”相关功能;模块 6 富集到“DNA 复制和修复”相关的功能(Figure 1)。

3 讨 论

本研究通过基因(蛋白)互作网络分析目前研究

积累的肝癌候选癌基因,尝试挖掘与肝癌发生显著性相关的基因互作网络核心模块,进一步探讨这些模块的生物学功能以及与肝癌的关系,结果显示 6 个基因互作网络模块分别与某特定的生物学功能或者通路相关,而这些功能与通路明显与肝癌甚至肿瘤的发生发展有着密切关系,如模块 1 包含 22 个基因,富集到“细胞周期”相关的功能。而众所周知,“细胞周期”是肿瘤发展过程中最关键的生物学功能之一。“细胞周期”的紊乱,进一步导致肿瘤细胞大量无限的分裂增殖^[16];模块 2 与模块 4 分别富集到“炎症反应”与“免疫反应”。而许多肿瘤特别是肝癌,经常是由于其组织的慢性炎症转化而成,同时“免疫逃逸”也是肿瘤细胞的基础特征之一^[17,18]。模块 3 与模块 5 分别富集到“药物代谢”与“氨基酸代谢”相关的功能。目前已有大量的文献报道了药物反应、氨基酸代谢等功能的紊乱与肿瘤细胞发展的关系密切^[18];而模块 6 富集到“DNA 复制与修复”相关功能。我们知道癌基因总体可分为致癌基因、抑癌基因以及 DNA 修复基因三个大类。DNA 复制与修复机制的紊乱往往导致细胞内基因组高频率的突变,最终导致细胞恶化^[11]。以上结果都提示,这些模块以及基因都参与特定的生物学功能而影响肝癌的发生发展。

近年来,随着高通量技术的发展,积累了大量与肝癌发生相关的数据以及相关的机制研究成果。但如何综合有效地利用这些海量的数据,仍然是一个值得探究的问题。本文通过收集大量的肝癌候选癌基因,通过基因(蛋白)互作网络分析的方式,从分子调控网络的角度出发去探究肝癌的分子机制,是系统生物学与肿瘤分子生物学研究结合的一个有效思路,其研究结果也值得进一步利用分子生物学方法去验证。当然,随着肝癌或其他肿



瘤高通量数据的进一步积累以及生物信息学方法与数据进一步发展,相信这种研究思路会发挥更大的作用。

参考文献:

- [1] Kimhofer T,Fye H,Taylor-Robinson S,et al. Proteomic and metabonomic biomarkers for hepatocellular carcinoma;a comprehensive review[J]. Br J Cancer,2015,112(7): 1141-1156.
- [2] Pinato DJ,Pirisi M,Maslen L,et al. Tissue biomarkers of prognostic significance in hepatocellular carcinoma [J]. Adv Anat Pathol,2014,21(4):270-284.
- [3] Liu M,Jiang L,Guan XY. The genetic and epigenetic alterations in human hepatocellular carcinoma;a recent update [J]. Protein Cell,2014,5(9):673-691.
- [4] Ma W,Yang D,Gu Y,et al. Finding disease-specific coordinated functions by multi-function genes:insight into the coordination mechanisms in diseases[J]. Genomics, 2009,94(2):94-100.
- [5] Du L,Qian X,Dai C,et al. Screening the molecular targets of ovarian cancer based on bioinformatics analysis [J]. Tumori,2015May 2.[Epub ahead of print]
- [6] Bi D,Ning H,Liu S,et al. Gene expression patterns combined with network analysis identify hub genes associated with bladder cancer [J]. Comput Biol Chem,2015,56:71-83.
- [7] Lee L,Wang K,Li G,et al. Liverome;a curated database of liver cancer-related gene signatures with self-contained context information [J]. BMC Genomics,2011,12 (Suppl 3):S3.

- [8] Franceschini A,Szklarczyk D,Frankild S,et al. STRING v9.1:protein-protein interaction networks,with increased coverage and integration[J]. Nucleic Acids Res,2013,41 (Database issue):D808-D815.
- [9] Bader GD,Hogue CW. An automated method for finding molecular complexes in large protein interaction networks [J]. BMC Bioinformatics,2003,4:2.
- [10] Huang daW,Sherman BT,Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources [J]. Nat Protoc,2009,4(1):44-57.
- [11] Futreal PA,Coin L,Marshall M,et al. A census of human cancer genes [J]. Nature Reviews,2004,4(3):177-183.
- [12] Maere S,Heymans K,Kuiper M. BiNGO:a cytoscape plug-in to assess overrepresentation of gene ontology categories in biological networks [J]. Bioinformatics,2005,21 (16): 3448-3449.
- [13] Zhong R,Liu L,Tian Y,et al. Genetic variant in SWI/SNF complexes influences hepatocellular carcinoma risk;a new clue for the contribution of chromatin remodeling in carcinogenesis [J]. Sci Rep,2014,4:4147.
- [14] Kongpetch S,Jusakul A,Ong CK,et al. Pathogenesis of cholangiocarcinoma;from genetics to signalling pathways [J]. Best Pract Res Clin Gastroenterol,2015,29(2):233-244.
- [15] Guo C,Liu S,Wang J,et al. ACTB in cancer[J]. Clin Chim Acta,2013,417:39-44.
- [16] Hanahan D,Weinberg RA. Hallmarks of cancer;the next generation [J]. Cell,2011,144(5):646-674.
- [17] Grivennikov SI,Greten FR,Karin M. Immunity,inflammation,and cancer [J]. Cell,2010,140(6):883-899.
- [18] Qiu F,Huang J,Sui M. Targeting arginine metabolism pathway to treat arginine-dependent cancers [J]. Cancer Lett,2015,364(1):1-7.