

利用基因芯片重注释筛选肝细胞肝癌术后复发相关 lncRNA

曾 玮, 陈 平

(第三军医大学第三附属医院, 重庆 400041)

摘 要: [目的] 利用基因芯片重注释寻找肝细胞肝癌(hepatocellular carcinoma, HCC)术后复发相关 lncRNA(long non-coding RNA)。[方法] 首先利用生物信息的方式重新注释 35/41 例复发/未复发 HCC 病例的基因表达谱数据, 筛查与 HCC 术后复发相关的差异表达 lncRNA。然后利用 ROC 曲线评价差异 lncRNA 的术后复发预测效能。最后利用基因共表达网络与功能富集分析探究差异 lncRNA 具体的生物学功能。[结果] 差异表达分析发现 3 个 lncRNA (LINC00342, LINC00607, LINC01553) 显著差异表达 [$\log(\text{fold change}) > 1, q \text{ value} < 0.05$]。ROC 曲线分析发现该 3 个差异 lncRNA 的 AUC 值为 0.693。基因共表达网络分析发现 LINC00342 相关共表达网络显著富集到“细胞周期”相关功能; LINC00607 相关的共表达网络中基因主要富集到“炎症反应”与“免疫反应”相关的生物学功能与通路; LINC01553 相关的共表达网络中基因主要富集到“DNA 复制与修复”相关的生物学功能与通路。[结论] 3 个 lncRNA 与 HCC 术后复发显著相关, 可作为一种潜在的诊断标志物组合, 值得进一步的挖掘与研究。

主题词: 肝细胞肝癌; lncRNA; 术后复发; 预后

中图分类号: R735.7 **文献标识码:** A **文章编号:** 1671-170X(2017)08-0658-05

doi: 10.11735/j.issn.1671-170X.2017.08.B002

Identification of lncRNAs Related to Postoperative Recurrence of Hepatocellular Carcinoma by Gene Microarray Re-annotation Strategy

ZENG Wei, CHEN Ping

(The Third Affiliated Hospital, Daping Hospital, Third Military Medical University, Chongqing 400041, China)

Abstract: [Objective] To identify the lncRNAs related to postoperative recurrence of hepatocellular carcinoma (HCC) by microarray re-annotation strategy. [Methods] The differentially expressed lncRNAs were identified in 35 cases of HCC with postoperative recurrence and 41 cases of HCC without recurrence by lncRNA re-annotation strategy. The classification power of the lncRNAs was analyzed by ROC curve. The biological functions and pathway related to these lncRNAs were investigated by gene co-expression network and gene functional enrichment analysis. [Results] Three differentially expressed lncRNAs (LINC00342, LINC00607, LINC01553) were found in the dataset [$\log(\text{fold change}) > 1, q \text{ value} < 0.05$]. The AUC score of the three lncRNAs was 0.693 by ROC curve analysis. Co-expression network analysis and gene functional enrichment analysis showed that the genes in LINC00342 related co-expression network were enriched in cell cycle related biological functions; genes in LINC00607 related co-expression network were enriched in inflammatory response and immune response related biological functions; genes in LINC01553 related co-expression network were enriched in DNA replication and repair related biological functions. [Conclusion] The lncRNAs identified in the study may be involved in postoperative recurrence of HCC, which might be a potential biomarker for HCC prognosis and worth of further investigation.

Subject words: hepatocellular carcinoma(HCC); lncRNA; postoperative recurrence; prognosis

肝细胞肝癌(hepatocellular carcinoma, HCC)是世

基金项目: 国家自然科学基金面上项目(30972895)

通讯作者: 陈 平, 主任医师, 教授, 博士生导师, 博士; 第三军医大学第三附属医院肝胆外科, 重庆市渝中区大坪长江支路 10 号(400041); E-mail: pingchen0089@163.com

收稿日期: 2016-12-06; **修回日期:** 2017-02-08

界第六大肿瘤, 发病率尤其以肝炎病毒流行的亚洲最高^[1]。在我国 90% 以上的肝癌为 HCC, 且大量患者是由慢性肝炎以及肝硬化发展而来的^[2]。HCC 恶性程度高, 且容易复发转移, 预后差。同时缺乏高效

的诊断手段,患者往往一经确诊即为中晚期,延误了最佳治疗时期。但基于传统临床病理特征的预测能力有限,且具有一定的滞后性^[3,4]。因此,目前前沿研究开始关注和寻找 HCC 早期诊断以及预后的分子标志物,综合这些已发现分子标志物的诊断效果来看,还存在准确性以及特异性不足的问题^[5]。基因芯片技术及新一代测序技术的应用,使得研究人员开始关注 RNA 分子对 HCC 的诊断作用,也发现了一些 mRNA 以及 miRNA 可能与 HCC 的早期发生以及预后相关^[6-8],但与临床应用还存在一定的距离。

长链非编码 RNA (long non-coding RNA, lncRNA) 是一类转录本长度超过 200nt、不编码蛋白的 RNA。研究表明, lncRNA 在细胞周期调控和细胞分化调控等众多生命活动中发挥重要作用,尤其在肿瘤发生发展过程中具有重要的意义,是遗传学研究热点^[9,10]。因此,本研究尝试利用生物信息的方式重新注释已有的 HCC 的肿瘤组织样本的全基因组表达谱芯片数据,去寻找 HCC 相关的潜在的术后复发 lncRNA,并探究其参与的生物学功能,评价其诊断价值。

1 材料与方法

1.1 基因表达谱数据以及数据处理

基因表达谱数据 (E-MEXP-84 与 E-TABM-292) 来自欧洲分子生物学实验数据库 (<http://www.ebi.ac.uk/arrayexpress/>)。该数据包括 76 个 HCC 患者的表达谱数据 (实验平台: Affymetrix GeneChip Human Genome HG-U133A 与 Affymetrix GeneChip Human Genome HG-U133B), 包括 35 例术后 2 年内复发的 HCC 患者样本数据, 41 例未复发的 HCC 患者样本数据^[6]。使用 gene expression console software (RMA 算法) 对表达谱原始数据进行预处理, 然后使用 SAM 分析算法筛选差异表达基因 (基于 EXCEL 的 SAM 软件), 本研究筛选差异表达基因的阈值为: $|\log(\text{fold change})| > 1$, $q \text{ value} < 0.05$ 。

1.2 基因芯片的 lncRNA 重注释

Affymetrix 公司的基因表达谱芯片中大量的探针不是仅仅针对所有的已知基因 (可表达蛋白的基因) 所设计的, 其还设计了大量的可检测人类基因组中大量的未知 EST 序列表达的探针, 而很多这些

EST 序列现在被证明是非编码 RNA (ncRNA), 包括 lncRNA^[11]。因此, 本文对这部分探针进行 lncRNA 重注释: (1) 使用 SeqMap 软件^[12] 将 Affymetrix GeneChip Human Genome HG-U133A 以及 Affymetrix GeneChip Human Genome HG-U133B 共有的探针序列在人全基因组序列 (GRCh38/hg38) 上进行重新比对定位。然后, 将可定位到基因组唯一位置的探针提取出来, 与 ENCODE 数据库 (release 21, GRCh38)^[13] 中的已知 lncRNA 位置进行匹配。最终得到 851 个探针可以重新注释到 649 个 lncRNA。针对多个探针对应到同一个 lncRNA 的情况, 本研究取所有探针其平均值作为该 lncRNA 的表达值。

1.3 统计分析

为了评价差异表达 lncRNA 组合对 HCC 术后复发的分型, 将所有差异表达 lncRNA 的几何平均值作为 lncRNA 组合的分型诊断值, 用于 HCC 术后复发的预测与诊断。同时, 使用 R 语言包 “survival ROC” 绘制 ROC 曲线, 并计算 AUC 值, 对 lncRNA 组合的分型效能进行评价。

lncRNA 共表达网络: 首先使用皮尔森相关计算 lncRNA 与所有差异表达基因间表达值向量 (某基因在所有样本中的表达值) 的相关系数, 将与 lncRNA 显著相关的基因定义为该 lncRNA 的共表达基因 ($|R| > 0.9$, $P < 0.05$)。而后将所有共表达的基因对使用 cytoscape 软件绘制形成基因共表达网络。GO 与 pathway 功能富集分析: 使用 DAVID 软件^[14] 对基因的功能进行富集分析 ($\text{adjust } P < 0.05$)。

2 结果

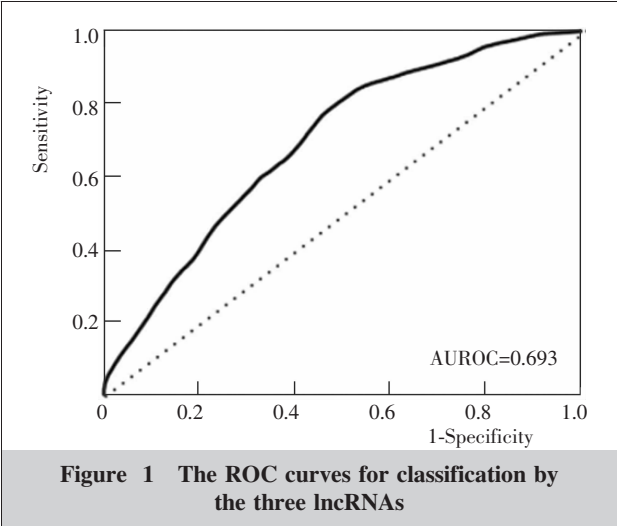
2.1 差异表达基因与 lncRNA

首先, 将 76 个样本的 HCC 表达谱芯片数据进行预处理, 得到标准化的表达谱数据。而后, 为了筛选可能与术后复发相关的基因与 lncRNA, 使用 SAM 分析寻找两组样本间差异表达基因与 lncRNA [$|\log(\text{fold change})| > 1$, $q \text{ value} < 0.05$]。结果共得到 324 个差异表达基因, 包括 208 个表达上调基因以及 116 个表达下调基因。同时, 还发现 3 个差异表达的 lncRNA (见 Table 1), 其中 2 个为表达上调 (LINC00342, LINC00607), 1 个为表达下调 (LINC01553)。这 3 个 lncRNA 均为基因间非编码

RNA(lincRNA)。

2.2 3 个 lincRNA 术后复发分型分析

为了考察 3 个 lincRNA 对 HCC 术后复发的预测能力，首先计算 3 个 lincRNA 在所有样本的分型诊断值(3 个 lincRNA 在某样本中表达值的几何均值)，然后绘制分型诊断值与术后复发的 ROC 曲线，结果如 Figure 1 所示，其 AUC 值为 0.693。证明这 3 个 lincRNA 联合诊断对 HCC 预测术后复发具有较好的效能。



2.3 lincRNA 与基因的共表达网络分析

为了进一步挖掘 3 个 lincRNA 可能参与的生物学功能，尝试利用皮尔森相关系数定义与 3 个 lincRNA 具有显著共表达的基因 ($|r|>0.9, P<0.05$)，结果如 Figure 2 所示，得到 3 个 lincRNA 相关的基因共表达网络，其中 LINC00342 相关的共表达网络中有 21 个基因与 LINC00342 显著共表达，涉及 213

Table 1 The information of three differential expressed lincRNAs

Ensembl id	Symbol	Chromosomal position	Ratio	Q value
ENSG00000232931	LINC00342	Chromosome 2: 95807118-95816215(-)	3.246	0.002
ENSG00000235770	LINC00607	Chromosome 2: 215611563-215843722(-)	5.123	0.003
ENSG00000235931	LINC01553	Chromosome 10: 59955430-59960913(-)	0.254	0.002

条共表达关系；LINC00607 相关的共表达网络中有 21 个基因与 LINC00607 显著共表达，涉及 128 条共表达关系；LINC01553 相关的共表达网络中有 16 个基因与 LINC01553 显著共表达，涉及 73 条共表达关系。

分别对三个网络中的基因进行 GO 与 pathway 功能富集分析，结果如 Table 2 所示，LINC00342 相关的共表达网络中基因主要富集到“细胞周期”相关的生物学功能与通路；LINC00607 相关的共表达网络中基因主要富集到“炎症反应”与“免疫反应”相关的生物学功能与通路；LINC01553 相关的共表达网络中基因主要富集到“DNA 复制与修复”相关的生物学功能与通路。

3 讨论

HCC 术后的高复发率是影响 HCC 治疗的关键因素^[15]，因此，建立一个有效的 HCC 术后评价体系对 HCC 外科治疗具有重要意义，而寻找和定义 HCC 术后的分子标志物是其中的关键。目前已有大量研究进行筛查对临床具有指导意义的分子标志物。而 lincRNA 作为一个近年才被发现和证明的遗传分子，已被发现参与了多种肿瘤的发生与发展过程，有研究开始研究其作为一些肿瘤诊断标志物的潜能^[16,17]。目前通过高通量技术，已得到大量对 HCC 具有诊断潜力的 lincRNA，如 ZNF674-AS1,DGCR5，

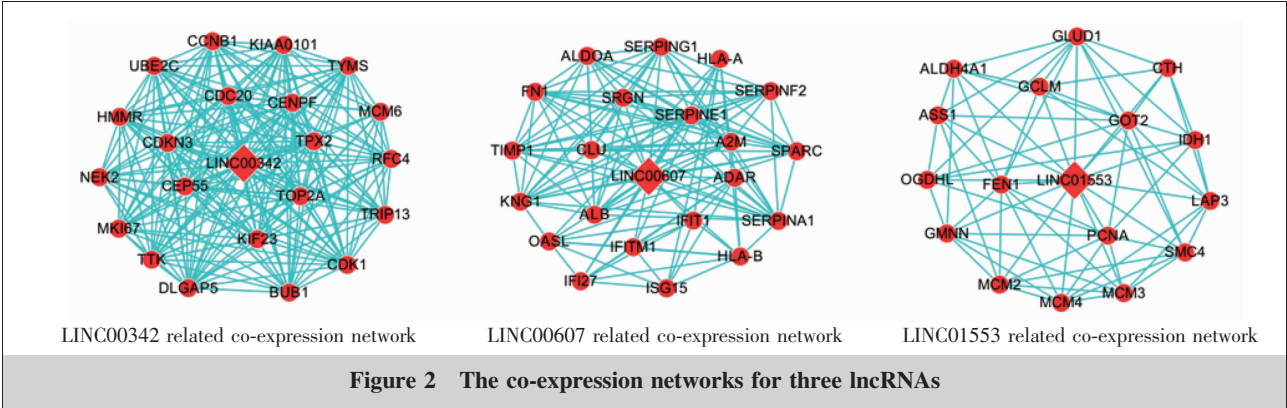


Table 2 The enrichment analysis of the genes in three lncRNA related co-expression networks

Gene	Database	Function	Gene number	Adjust-P
LINC00342 related co-expression network	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
LINC00607 related co-expression network	KEGG	Complement and coagulation cascades	7	1.37E-10
		Allograft rejection	2	0.028027
		Antigen processing and presentation	2	0.063727
	GO	Inflammatory response	7	2.73E-07
		Immune response	3	0.023436
LINC01553 related co-expression network	KEGG	Antigen processing and presentation	2	0.030307
		Arginine and proline metabolism	6	1.24E-08
		DNA replication	5	2.12E-09
	GO	Base excision repair	2	0.027257
		DNA replication	5	5.53E-07
		DNA-dependent DNA replication	4	1.48E-06
		DNA metabolic process	5	2.73E-05
		Cellular amino acid biosynthetic process	6	7.76E-11
		Glutamate metabolic process	4	1.14E-07

SRA1, SNHG1, GAS5, HOTAIR 等^[18-23]。其中 HOTAIR 被发现与 HCC 的术后复发相关^[23], DGCR5, SNHG1 以及 GAS5 的表达与 HCC 的不良预后相关^[19,21,22]。遗憾的是这些研究人员发现的相关 lncRNA 未在本研究的探针重注释列表中发现。因此,本研究的结果不能对这些 lncRNA 进行进一步的验证。不过,本研究尝试利用基因表达谱寻找 lncRNA 与 HCC 预后之间的关系,结果也发现 3 个与 HCC 术后复发相关的新的 lncRNA, ROC 曲线分析发现其联合诊断具有较高的灵敏度和特异性,值得进一步的挖掘与验证。

本研究对传统的全基因组表达谱芯片的所有探针在全基因组层面进行重注释,将目前新定义的大量 lncRNA 重新注释到芯片中。该注释方法是对传统全基因组芯片结果中所遗漏的大量有效信息的重新提炼。此外,该方法还可推广到包括 affymetrix 在内的各种芯片品牌之中,且不仅仅局限于表达谱芯片,还包括 SNP 芯片、CGH 芯片以及甲基化芯片等各种平台。对于目前公共数据平台中所积累的大量的芯片实验数据的重新利用具有非常好的借鉴意义。

本研究还进一步利用基因共表达网络分析探究了这 3 个 lncRNA (LINC00342, LINC00607 及 LINC01553)可能参与的生物学功能与通路。功能富集分析发现 LINC00342 相关的共表达网络中基因

主要富集到“细胞周期”相关的生物学功能与通路,提示 LINC00342 可能在肿瘤细胞的增殖过程中起重要作用,因“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^[24]。LINC00607 相关的共表达网络中基因主要富集到“炎症反应”与“免疫反应”相关的生物学功能与通路,提示 LINC00607 可能在肿瘤细胞的炎症与免疫中起重要作用。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^[24,25]。LINC01553 相关的共表达网络中基因主要富集到“DNA 复制与修复”相关的生物学功能与通路,提示 LINC01553 可能在肿瘤细胞的 DNA 复制与修复中起重要作用。已知 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^[26]。

总之,本研究尝试对一套 HCC 术后复发的基因表达谱数据的重新注释,将传统基因表达谱中大量遗漏的 lncRNA 信息重新提炼,得到 3 个与 HCC 术后复发相关的 lncRNA。且利用 ROC 曲线发现 3 个 lncRNA 对 HCC 术后复发的诊断具有较高的效能,同时还利用基因共表达网络分析发现 3 个 lncRNA 可能参与到多种与肿瘤发展相关的功能与通路。综合上述研究结果,提示 3 个 lncRNA 与 HCC 术后复发显著相关,可作为一种潜在的诊断标志物组合,值

得进一步的挖掘与研究。

参考文献:

- [1] Jemal A, Bray F, Center MM, et al. Global cancer statistics[J]. CA Cancer J Clin, 2011, 61(2): 69-90.
- [2] Bruix J, Gores GJ, Mazzaferro V. Hepatocellular carcinoma: clinical frontiers and perspectives [J]. Gut, 2014, 63(5): 844-855.
- [3] Altekruse SF, McGlynn KA, Reichman ME. Hepatocellular carcinoma incidence, mortality, and survival trends in the United States from 1875 to 2005 [J]. J Clin Oncol, 2009, 27(9): 1485-1491.
- [4] Liu G, Dong C, Liu L. Integrated multiple "-omics" data reveal subtypes of hepatocellular carcinoma[J]. PLoS One, 2016, 11(11): e0165457.
- [5] Terada T. Combined hepatocellular-cholangiocarcinoma with stem cell features, ductal plate malformation subtype: a case report and proposal of a new subtype[J]. Int J Clin Exp Pathol, 2013, 6(4): 737-748.
- [6] Wang SM, Ooi LL, Hui KM. Upregulation of Rac GTPase-activating protein 1 is significantly associated with the early recurrence of human hepatocellular carcinoma [J]. Clin Cancer Res, 2011, 17(18): 6040-6051.
- [7] Hao K, Lamb J, Zhang C, et al. Clinicopathologic and gene expression parameters predict liver cancer prognosis [J]. BMC Cancer, 2011, 11: 481.
- [8] Xiang ZL, Zhao XM, Zhang L, et al. MicroRNA-34a expression levels in serum and intratumoral tissue can predict bone metastasis in patients with hepatocellular carcinoma[J]. Oncotarget, 2016, 7(52): 87246-87256.
- [9] M MK, Goyal R. LncRNA as a therapeutic target for angiogenesis [J]. Curr Top Med Chem, 2016 Nov 16. [Epub ahead of print]
- [10] Previdi MC, Carotenuto P, Zito D, et al. Noncoding RNAs as novel biomarkers in pancreatic cancer: what do we know?[J]. Future Oncol, 2016, 13(5): 443-453.
- [11] Du Z, Fei T, Verhaak RG, et al. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer[J]. Nat Struct Mol Biol, 2013, 20(7): 908-913.
- [12] Jiang H, Wong WH. SeqMap: mapping massive amount of oligonucleotides to the genome[J]. Bioinformatics, 2008, 24(20): 2395-2396.
- [13] Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for The ENCODE Project[J]. Genome Res, 2012, 22(9): 1760-1774.
- [14] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources [J]. Nature Protoc, 2009, 4(1): 44-57.
- [15] Portolani N, Coniglio A, Ghidoni S, et al. Early and late recurrence after liver resection for hepatocellular carcinoma: prognostic and therapeutic implications [J]. Ann Surg, 2006, 243(2): 229-235.
- [16] Zhou M, Zhao H, Wang Z, et al. Identification and validation of potential prognostic lncRNA biomarkers for predicting survival in patients with multiple myeloma[J]. J Exp Clin Cancer Res, 2015, 34: 102.
- [17] Sun J, Chen X, Wang Z, et al. A potential prognostic long non-coding RNA signature to predict metastasis-free survival of breast cancer patients[J]. Sci Rep, 2015, 5: 16553.
- [18] Zhang L, He T, Yan Y, et al. Expression and clinical significance of the novel long noncoding RNA ZNF674-AS1 in human hepatocellular carcinoma [J]. Biomed Res Int, 2016, 2016: 3608914.
- [19] Huang R, Wang X, Zhang W, et al. Down-regulation of LncRNA DGCR5 correlates with poor prognosis in hepatocellular carcinoma[J]. Cell Physiol Biochem, 2016, 40(3-4): 707-715.
- [20] Luo P, Jing W, Zhu M, et al. Decreased expression of LncRNA SRA1 in hepatocellular carcinoma and its clinical significance[J]. Cancer Biomark, 2017 Jan 3. [Epub ahead of print]
- [21] Zhang M, Wang W, Li T, et al. Long noncoding RNA SNHG1 predicts a poor prognosis and promotes hepatocellular carcinoma tumorigenesis [J]. Biomed Pharmacother, 2016, 80: 73-79.
- [22] Tu ZQ, Li RJ, Mei JZ, et al. Down-regulation of long non-coding RNA GAS5 is associated with the prognosis of hepatocellular carcinoma[J]. Int J Clin Exp Pathol, 2014, 7(7): 4303-4309.
- [23] Yang Z, Zhou L, Wu LM, et al. Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation[J]. Ann Surg Oncol, 2011, 18(5): 1243-1250.
- [24] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation [J]. Cell, 2011, 144(5): 646-674.
- [25] Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer [J]. Cell, 2010, 140(6): 883-899.
- [26] Futreal PA, Coin L, Marshall M, et al. A census of human cancer genes [J]. Nat Rev Cancer, 2004, 4(3): 177-183.