

基因富集及Meta分析对影响肝癌发生发展关键基因的筛选

曹 骥, 卢晓旭, 胡艳玲, 李 媛, 朱伶群, 杨 春, 欧 超, 唐艳萍

■背景资料

近年来随着人类和大鼠、小鼠等动物基因组测序工程的逐步完成, 人们开始意识到那些跨种属(cross-species)在不同物种共同表达的基因, 可能具有保守和更重要的功能。跨种属的肿瘤比较基因组学研究策略遂基于这一认识而产生。

曹骥, 卢晓旭, 李媛, 朱伶群, 杨春, 欧超, 唐艳萍, 广西壮族自治区肿瘤医院防治研究所实验研究部 广西壮族自治区南宁市 530021

胡艳玲, 广西医科大学医学科学实验中心 广西壮族自治区南宁市 530021

曹骥, 副研究员, 硕士生导师, 主要从事肝癌病因学、发病机制方面的研究。

国家自然科学基金资助项目, No. 30960428

作者贡献分布: 此课题由曹骥与李媛设计; 研究过程由曹骥、卢晓旭、胡艳玲、李媛、朱伶群、杨春、欧超及唐艳萍操作完成; 数据分析和论文写作由曹骥与卢晓旭完成。

通讯作者: 曹骥, 副研究员, 530021, 广西壮族自治区南宁市, 广西壮族自治区肿瘤医院防治研究所实验研究部. caojicn@yahoo.com.cn 电话: 0771-5310593

收稿日期: 2011-10-09 修回日期: 2012-01-30

接受日期: 2012-02-27 在线出版日期: 2012-03-28

Applying gene set enrichment analysis and meta-analysis to screen key genes controlling the development and progression of hepatic carcinoma

Ji Cao, Xiao-Xu Lu, Yan-Ling Hu, Yuan Li, Ling-Qun Zhu, Chun Yang, Chao Ou, Yan-Ping Tang

Ji Cao, Xiao-Xu Lu, Yuan Li, Ling-Qun Zhu, Chun Yang, Chao Ou, Yan-Ping Tang, Research Department of Guangxi Cancer Institute, Nanning 530021, Guangxi Zhuang Autonomous Region, China

Yan-Ling Hu, Medical Scientific Experimental Center of Guangxi Medical University, Nanning 530021, Guangxi Zhuang Autonomous Region, China

Supported by: National Science Foundation of China, No. 30960428

Correspondence to: Ji Cao, Associate Researcher, Research Department of Guangxi Cancer Institute, Nanning 530021, Guangxi Zhuang Autonomous Region, China. caojicn@yahoo.com.cn

Received: 2011-10-09 Revised: 2012-01-30

Accepted: 2012-02-27 Published online: 2012-03-28

Abstract

AIM: To analyze vast amounts of hepatic carcinoma-related microarray data and identify crucial genes that control the development and progression of hepatocellular carcinoma (HCC).

METHODS: Cross-species comparison could be used to explore the similarities between HCC-

related gene expression profiles of human beings and other species. In order to screen genes that are involved in hepatocarcinogenesis, gene set enrichment analysis (GSEA) and meta-analysis were performed to study five gene expression data sets of independent species.

RESULTS: Among the five gene expression data sets, three up-regulated and two down-regulated pathways were found to be consistent by gene set enrichment analysis. The up-regulated pathways are amino sugar and nucleotide sugar metabolism, cell cycle, and thyroid cancer, while the down-regulated pathways are linoleic acid metabolism and arachidonic acid metabolism. A total of 1 708 genes with a $P < 0.05$ were found in meta-analysis for five datasets, of which 720 could be assigned to functional pathways by DAVID and KEGG. These pathways include cell cycle, oocyte meiosis, and DNA replication. Cell cycle is the overlapping significant pathway between the two methods. Twenty-five genes with a $P < 0.05$ were identified in meta-analysis of cell cycle pathway. Five significant genes may be involved in the occurrence and progression of HCC.

CONCLUSION: Cell cycle may be the crucial pathway to affect signal transduction in hepatocarcinogenesis.

Key Words: Hepatocellular carcinoma; Gene set enrichment analysis; Meta-analysis; Cross-species; Key gene

Cao J, Lu XX, Hu YL, Li Y, Zhu LQ, Yang C, Ou C, Tang YP. Applying gene set enrichment analysis and meta-analysis to screen key genes controlling the development and progression of hepatic carcinoma. *Shijie Huaren Xiaohua Zazhi* 2012; 20(9): 754-758

摘要

目的: 筛选影响肝癌发生发展的关键基因。

方法: 运用跨种属肿瘤基因筛选策略比较不同种属的肝癌基因表达谱间的相似改变, 选择5套不同种属的肝癌基因表达芯片分别通过基

■同行评议者

王德盛, 副主任医师, 中国人民解放军第四军医大学西京医院肝胆外科

因组富集(gene set enrichment analysis, GSEA)以及对单套数据集单个基因元分析(meta-analysis, Meta)的分析方法, 筛选出在转录水平上影响肝癌的基因。

结果: 用GSEA方法分析, 5组数据中所得通路对比, 上调中皆有的通路为氨基糖核苷酸糖代谢、细胞周期、甲状腺癌; 下调中皆有的通路为亚油酸代谢、花生四烯酸代谢。对单套数据集单个基因进行Meta分析, 共筛出 $P<0.05$ 的基因1 708个。用DAVID和KEGG网站的分析工具发现这1 708个差异基因中有720个基因能够在KEGG库中筛出, 主要分布在细胞周期、卵母细胞减数分裂、DNA复制等通路。这两种分析方法得出的通路中, 重叠性较高的主要为细胞周期通路。在细胞周期通路中差异性有统计学意义($P<0.05$)的基因25个, 文献报道其中5个基因与肝癌有密切联系。

结论: 可能影响肝癌发生发展的信号传导通路是细胞周期通路, 后续我们将对细胞周期通路里的显著性基因进行验证。

关键词: 肝癌; 基因富集; Meta分析; 跨种属; 关键基因

曹骥, 卢晓旭, 胡艳玲, 李媛, 朱伶群, 杨春, 欧超, 唐艳萍. 基因富集及Meta分析对影响肝癌发生发展关键基因的筛选. 世界华人消化杂志 2012; 20(9): 754-758

<http://www.wjgnet.com/1009-3079/20/754.asp>

0 前言

肝细胞癌(hepatocellular carcinoma, HCC)是我国常见的恶性肿瘤, 预后较差, 如何寻找治疗HCC的药物靶点是当前攻克肝癌的重要任务。在肿瘤发生发展过程中, 有大量伴随基因参与癌基因扩增过程, 但是他们并不是我们所要找的关键基因^[1]。因此, 如何将那些影响肿瘤发生发展的关键性分子改变从伴随性改变中识别出来, 是目前肿瘤研究领域的重要挑战之一。近年来随着人类和大鼠、小鼠等动物基因组测序工程的逐步完成, 人们开始意识到那些跨种属(cross-species)在不同物种共同表达的基因, 可能具有保守和更重要的功能。跨种属的肿瘤比较基因组学研究策略遂基于这一认识而产生^[2]。Mattison等^[3]运用跨种属策略比较基因组杂交, 研究共同表达于人、小鼠、大鼠肝癌组织中的基因, 以期找到新颖的候选基因。我们考虑通过探索人和其他动物的肝癌形成过程中共同拥有的基

因改变, 有可能筛选出影响肝癌发生发展的关键基因。基因芯片已广泛用于肝癌的研究, 对于全基因组mRNA表达芯片的分析, 我们也得到可预测肝细胞癌预后、复发等HCC转录物, 但是如何挖掘出特殊的基因标记来作为治疗疾病的靶点, 还是一个巨大的挑战。为了解决这个问题, Mootha等^[4]提出基因组富集(gene set enrichment analysis, GSEA)分析, 该方法能在病例对照类型数据中, 基于基因组系统水平上来挖掘影响疾病的基因通路。通过分析一组处于2种生物学状态的基因表达谱杂交数据, 了解他们在特定功能基因集中的表达状况以及这种表达状况是否存在某种统计学显著性。另外, 因为实验平台、样本、标化方法、分析方法等问题的存在, 不同实验室的芯片数据有很多的差异, 元分析(Meta-analysis)是一种可行的解决方法, 可对同一个问题所发表相关研究报告的结果进行收集、统计上的整合, 以期获得更准确或更多的结果^[5]。因此, 本次研究我们以GSEA和Meta分析的方法对标准化以后的5套HCC全基因组表达芯片数据进行分析, 为HCC发病机制的阐明提供重要的理论基础。

1 材料和方法

1.1 材料 本文通过广泛的文献和数据库搜索, 找出迄今国内外所有已报道的任何病因引起的、无论是人还是任何其他动物的肝癌相关基因表达谱。有关数据从GEO数据库(<http://www.ncbi.nlm.gov/geo/>)中下载。搜索关键词为hepatocellular carcinoma, 限制研究类型为expression profiling by array, 结果有230篇文章提供基因芯片数据。符合以下标准的数据集将纳入我们的研究中: (1)数据集必须是全基因组的表达mRNA芯片数据; (2)这些数据是关于HCC和正常组织对照; (3)本次均考虑经标准化或者原始数据集; (4)数据集必须包括超过3个样本以上。最后, 只有5套样本纳入我们的研究中(表1)。本次研究我们以GSEA和Meta分析的方法对标准化的5套HCC全基因组表达芯片数据进行分析, 将得到的差异显著基因与我们以前的树鼯肝癌模型的mRNA芯片结果进行比较, 筛选出那些在2种以上种属(包括人)的肝癌出现特异改变的基因。

1.2 方法 通过Bioconductor^[11]的2.10.1版本来对数据进行标准化处理。用软件包affy中的RMA算法^[12,13]对affymetrix平台的原始数据进行背景校正、标准化和log2转换。对每一套数据每个探

■ 研发前沿

基因芯片已广泛用于肝癌的研究, 对于全基因组mRNA表达芯片的分析, 可得到预测肝细胞癌预后、复发等的HCC转录物, 但是如何挖掘出特殊的基因标记来作为治疗疾病的靶点, 还是一个巨大的挑战。

■ 相关报道

Mattison等运用跨种属策略比较基因组杂交, 研究共同表达于人、小鼠、大鼠肝癌组织中的基因, 以期找到新颖的候选基因。

■创新盘点

本研究在跨种属策略指导下采用GSEA和Meta分析方法对标准化以后的5套HCC全基因组表达芯片数据进行分析,发现可能影响肝癌发生发展的信号传导通路是细胞周期通路。

表 1 5套全基因组数据集的基本情况

类别	作者	芯片平台	实验设计	探针数	来源	样本(n)	对照(n)
GSE19665	Deng等 ^[6]	u133Plus2.0	paired, tissues	54 000	homo sapiens	10	10
GSE9809	Liao等 ^[7]	mouse430_2	unpaired+paired, tissues	45 000	mus musculus	3	7
GSE9012	Khetchoumian等 ^[8]	mouse430_2	unpaired, tissues	45 000	mus musculus	5	5
GSE19004	Viatour等 ^[9]	mouse430_2	unpaired, tissues	45 000	mus musculus	5	4
GSE2127	Sheth等 ^[10]	mouse430a	paired, tissues	22 000	mus musculus	9	6

Paried: 对hcc组的对照来自hcc本体; unpaired: 对hcc组的对照来自其他正常群体。

表 2 对5套数据集Meta分析差显基因分布

基因	P值	基因	P值	基因	P值	基因	P值
ABI3BP	3.23E-06	UBE2C	3.18E-05	FOXMI	6.58E-05	rrm2	2.74E-05
CCNB1	5.37E-06	mcm2	4.07E-05	BIRC5	7.18E-05	Ttc36	3.06E-05
NEK2	7.66E-06	ASPM	4.29E-05	STMN1	7.74E-05	Hist1h2ad	5.71E-05
MKI67	1.35E-05	NCAPH	5.23E-05	racgap1	7.84E-05	TOP2A	5.91E-05
cdc20	2.46E-05	TUBA1B	5.23E-05	Hist1h2ag	7.94E-05	Hist1h2ai	7.94E-05
angptl6	2.47E-05	CCNB2	5.45E-05	Hist1h2ah	7.94E-05	CDCA5	9.89E-05

针的检验采用成组 t 检验。我们只选取在KEGG库^[14]中存在的基因进行GSEA分析。剔除变异四分位距 <0.5 的基因。如果一个基因对应几个探针,我们只保留变异IQR最高的探针。GSEA通过Bioconductor的category包进行。只有超过10个基因的类保留,通过 t 检验对每一个通路中的基因进行检验。通过1 000次循环的排列组合(permutation)获得每个显著通路的 P 值。将得到的5套数据各自上调下调的通路进行总和比较,发现细胞周期(cell cycle)这个通路在5套数据中都表现为上调。我们将每套数据里这条通路的所含基因进行Meta分析。运用SAS9.13软件,通过 t 检验把每套数据里Cell cycle通路里的每个探针算出 P 值,再通过公式^[15]

$$LS = \frac{\sum_{i=1}^N (-\log(P_i))}{N} \quad (\text{自由度为数据集K的2倍})$$

算出每个基因的卡方值,最后保留 $P<0.05$ 的基因。对这些基因通路的分析通过DAVID(<http://david.abcc.ncifcrf.gov>)中的KEGG库进行分析。

2 结果

2.1 GSEA分析 通过GSEA方法对5套数据集进行功能基因富集,分别找出影响这几个数据集的主要上调通路和下调通路。GSE19665富集出27条上调通路,71条下调通路。GSE9809数据集富集出上调通路56条,下调通路4条。GSE9012数据集富集出上调通路66条,下调通路34条。GSE19004

数据集富集出上调通路73条,下调通路51条。GSE2127数据集富集出上调通路69条,下调通路50条。其中数据集GSE19665和数据集GSE19004通路重叠性比较高。5组数据中所得通路对比,下调中皆有的通路为亚油酸代谢(linoleic acid metabolism),花生四烯酸代谢(arachidonic acid metabolism);上调中皆有的通路为氨基糖核苷酸糖代谢(amino sugar and nucleotide sugar metabolism),细胞周期(cell cycle),甲状腺癌(thyroid cancer)。

2.2 Meta分析 用一般的成组 t 检验对5套数据集单独分析,其中GSE19665数据集中共筛出阳性基因220个,其余数据集筛出阳性基因均为213个。对5套数据单独进行 t 检验得出每个基因的 P 值后,通过软件SAS13.0运用选择的Meta公式进行整合分析,共筛出1 708个基因的 $P<0.05$ 。 $P<10^{-4}$ 有24个基因,具体每个基因名称及 P 值(表2)。

这1 708个基因通过DAVID的KEGG库进行通路富集,其中有720个基因能够在KEGG库中筛出,主要分布在cell cycle、卵母细胞减数分裂(oocyte meiosis)、DNA复制(DNA replication)、缬氨酸(valine)、亮氨酸(leucine)和异亮氨酸(isoleucine degradation)降解孕激素介导的卵母细胞成熟(progesterone-mediated oocyte maturation)、致病性大肠埃希氏菌感染(pathogenic escherichia coli infection)、甘氨酸(glycine)、丝氨酸(serine)和苏氨酸(threonine metabolism)

表 3 对5套数据集Meta分析存在于细胞周期通路里的差异基因分布

基因	χ^2 值	P值	基因	χ^2 值	P值	基因	χ^2 值	P值
CCNB2	46.94	9.67E-07	MCM4	46.94	0.0030	CDC25A	29.98	0.0009
mcm2	36.51	6.89E-05	cdc20	36.51	0.0065	Mad2l1	29.43	0.0011
YWHAB	35.63	9.74E-05	smc3	35.63	0.0067	MCM7	28.57	0.0015
CCNA2	33.83	0.0002	pcnA	33.83	0.0079	CCNE1	28.21	0.0017
CDKN2C	32.60	0.0003	RAD21	32.60	0.0090	TGFB1	30.15	0.0127
Cdk1	32.24	0.0004	CDKN1A	32.24	0.0096	YWHAZ	29.98	0.0206
MCM6	32.07	0.0004	CCND1	32.07	0.0108	YWHAG	29.43	0.0208
cdkn2b	30.15	0.0008	SMAD3	30.15	0.0116	YWHAH	28.57	0.0252
CCNB1	30.15	0.0008						

■名词解释

基因组富集分析 (Gene Set Enrichment Analysis, GSEA): 根据已有的对基因的定位、性质、功能、生物学意义等知识的基础上, 首先构建了一个分子标签数据库, 分析一组处于两个生物学状态的基因表达谱杂交数据, 他们在特定的功能基因集中的表达状况, 以及这种表达状况是否存在某种统计学显著性。

的代谢, 错配修复(mismatch repair)、丙酸代谢(propanoate metabolism)、原发性胆汁酸的生物合成(primary bile acid biosynthesis)、脂肪酸代谢(fatty acid metabolism)、前列腺癌(prostate cancer)、黏着(focal adhesion)、嘧啶代谢(pyrimidine metabolism)、p53信号通路(p53 signaling pathway)、神经胶质瘤(glioma)、核苷酸切除修复(nucleotide excision repair)、赖氨酸降解(lysine degradation)、小细胞肺癌(small cell lung cancer)、剪接(spliceosome)、癌症通路(pathways in cancer)、TGF- β 信号通路(TGF- β signaling pathway)、Fc γ R-介导的细胞的吞噬功能(Fc γ R-mediated phagocytosis)、同源重组(homologous recombination)、色氨酸代谢(tryptophan metabolism)、慢性粒细胞白血病(chronic myeloid leukemia)、泛酸和辅酶A合成(pantothenate and CoA biosynthesis)、泛素介导的蛋白水解(ubiquitin mediated proteolysis)、柠檬酸循环(TCA cycle)、叶酸碳库(one carbon pool by folate)、神经营养因子信号通路(neurotrophin signaling pathway)、butanoate代谢(butanoate metabolism)、甲状腺癌(thyroid cancer)、黑色素瘤(melanoma)、肌动蛋白骨架的调节(regulation of actin cytoskeleton)、胰腺癌(pancreatic cancer)等通路中。

2.3 两种方法所得结果分析 用GSEA和Meta两种方法, 我们得到重叠性较高的通路: 细胞周期。通过R命令语言, 得到5组数据集里cell cycle通路各自所含基因探针号。将探针号传至<http://david.abcc.ncifcrf.gov/conversion.jsp>网站上进行官方名称转换, 得到5组数据里该通路所含的基因名称。GSE19665里在细胞周期通路所含差异基因99个, GSE9809含96个, GSE9012含90个, GSE19004含106个, GSE2127含113个。通过上步

Meta运行结果可得细胞周期通路里差异性有统计学意义($P < 0.05$)的基因25个, 他们的名称、卡方值及P值(表3)。

3 讨论

目前, 对基因芯片数据的分析是研究基因芯片的重要部分。单独对某次实验结果进行分析, 而且只是单个基因分析, 由于样本问题, 可能会漏掉很多有用的信息; 而且对基因芯片单套的 t 检验有一定的局限性, 受到样本量的限制, 导致不可信的变异估计, 可产生较高的假阳性, 忽略了不同样本中表达水平的差异^[16]。基因富集方法通过分析一组处于2种不同生物状态(如正常和癌变)的芯片数据, 推断出现的基因是否有共同的表达趋势, 以此可以找出与疾病关联的基因或者通路^[17]。

本文结合GSEA和Meta两种方法对这5套数据进行分析, 对2种结果进行重叠对比, 希望能找到影响HCC相关的基因和通路。GSEA方法通过分析2组以上的样本之间差异表达基因, 对样本进行聚类以获得明显基因表达差异的样本分类。用R语言来处理数据, 进行统计分析, 得到5组数据共同改变的通路。Meta方法对单套数据集进行 t 检验, 将结果就行Meta分析, 得到差异表达的基因, 放入DAVID网站得到这些基因可能所在的通路。最终将重叠性较高的细胞周期通路进行分析, 得出差异性显著的基因。

在我们筛选出的细胞周期通路25个差异基因中, CCNB2(cyclinB2)、CCNB1(cyclinB1)、CCND1(cyclin D1)、CDC25A、Cdk1等5个基因与肝癌发生发展有密切联系。其中CDK是细胞周期调控网络的核心蛋白, 其表达活性的改变直接影响到细胞周期的长短, 决定着细胞的进程, 与机体细胞的生长、分化、运动、凋亡以

■同行评价

本文选题新颖, 具有很好的学术价值。

及肿瘤的发生、发展、转移关系密切^[18]. CDK在肝癌细胞增殖及凋亡的调控中起着重要的作用. cyclinB、cyclinD同属cyclin家族, 决定CDK何时何处将何种底物磷酸化. 研究表明^[19], cyclinD1在正常肝组织中正常表达, 而在肝细胞癌中过度表达, 与HCC的组织学分级相关, 提示参与了肝癌的演进发展, 在肝细胞癌发病机制中可能促进细胞分裂增殖, 形成肿瘤. 同家族中高表达的cyclinB1在肝细胞癌变过程中发挥重要作用. CDC25A为双重特异性磷酸酶, 能激活细胞周期素依赖性激酶(CDK), 推动细胞周期的进行. 目前国内未见CDC25A与肝癌关系的相关研究, 但是国外已有报道, 认为CDC25A在肝癌转移方面有重要作用^[20]. 后续我们将对这些差异显著基因进行验证.

4 参考文献

- Beroukhi R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, Cho YJ, Haery L, Greulich H, Reich M, Winckler W, Lawrence MS, Weir BA, Tanaka KE, Chiang DY, Bass AJ, Loo A, Hoffman C, Prensner J, Liefeld T, Gao Q, Yecies D, Signoretti S, Maher E, Kaye FJ, Sasaki H, Tepper JE, Fletcher JA, Tabernero J, Baselga J, Tsao MS, Demichelis F, Rubin MA, Janne PA, Daly MJ, Nucera C, Levine RL, Ebert BL, Gabriel S, Rustgi AK, Antonescu CR, Ladanyi M, Letai A, Garraway LA, Loda M, Beer DG, True LD, Okamoto A, Pomeroy SL, Singer S, Golub TR, Lander ES, Getz G, Sellers WR, Meyerson M. The landscape of somatic copy-number alteration across human cancers. *Nature* 2010; 463: 899-905
- 史俊林, 李媛. 应用跨种属肿瘤基因筛选策略探寻肿瘤关键分子. *世界华人消化杂志* 2010; 18: 1255-1260
- Mattison J, Kool J, Uren AG, de Ridder J, Wessels L, Jonkers J, Bignell GR, Butler A, Rust AG, Brosch M, Wilson CH, van der Weyden L, Largaespada DA, Stratton MR, Futreal PA, van Lohuizen M, Berns A, Collier LS, Hubbard T, Adams DJ. Novel candidate cancer genes identified by a large-scale cross-species comparative oncogenomics approach. *Cancer Res* 2010; 70: 883-895
- Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 2003; 34: 267-273
- Greenbaum D, Jansen R, Gerstein M. Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts. *Bioinformatics* 2002; 18: 585-596
- Deng YB, Nagae G, Midorikawa Y, Yagi K, Tsutsumi S, Yamamoto S, Hasegawa K, Kokudo N, Aburatani H, Kaneda A. Identification of genes preferentially methylated in hepatitis C virus-related hepatocellular carcinoma. *Cancer Sci* 2010; 101: 1501-1510
- Liao YJ, Liu SP, Lee CM, Yen CH, Chuang PC, Chen CY, Tsai TF, Huang SF, Lee YH, Chen YM. Characterization of a glycine N-methyltransferase gene knockout mouse model for hepatocellular carcinoma: Implications of the gender disparity in liver cancer susceptibility. *Int J Cancer* 2009; 124: 816-826
- Khetchoumian K, Teletin M, Tisserand J, Mark M, Herquel B, Ignat M, Zucman-Rossi J, Cammas F, Lerouge T, Thibault C, Metzger D, Chambon P, Losson R. Loss of Trim24 (Tif1alpha) gene function confers oncogenic activity to retinoic acid receptor alpha. *Nat Genet* 2007; 39: 1500-1506
- Viatour P, Sage J. Mouse HCC model. Available from: URL: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19004>
- Sheth SS, Bodnar JS, Ghazalpour A, Thippavong CK, Tsutsumi S, Tward AD, Demant P, Kodama T, Aburatani H, Lusis AJ. Hepatocellular carcinoma in Txnip-deficient mice. *Oncogene* 2006; 25: 3528-3536
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smyth G, Tierney L, Yang JY, Zhang J. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004; 5: R80. Epub 2004 Sep 15
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003; 4: 249-264
- Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004; 20: 307-315
- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000; 28: 27-30
- Daves MH, Hilsenbeck SG, Lau CC, Man TK. Meta-analysis of multiple microarray datasets reveals a common gene signature of metastasis in solid tumors. *BMC Med Genomics* 2011; 4: 56
- MacDonald JW, Ghosh D. COPA--cancer outlier profile analysis. *Bioinformatics* 2006; 22: 2950-2951
- Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics* 2007; 23: 3251-3253
- Liu L, Schwartz B, Tsubota Y, Raines E, Kiyokawa H, Yonekawa K, Harlan JM, Schnapp LM. Cyclin-dependent kinase inhibitors block leukocyte adhesion and migration. *J Immunol* 2008; 180: 1808-1817
- Greenbaum LE. Cell cycle regulation and hepatocarcinogenesis. *Cancer Biol Ther* 2004; 3: 1200-1207
- Wang XQ, Zhu YQ, Lui KS, Cai Q, Lu P, Poon RT. Aberrant Polo-like kinase 1-Cdc25A pathway in metastatic hepatocellular carcinoma. *Clin Cancer Res* 2008; 14: 6813-6820

编辑 张姗姗 电编 闫晋利