

黄曲霉毒素 B1 致癌毒性相关基因的生物信息学分析

李力, 曹以诚, 区镜深, 毛小帆

(华南理工大学生物科学与工程学院, 广东广州 510006)

摘要: 黄曲霉毒素是广泛存在于食品中的对人类危害最大的、最常见的霉菌毒素, 其中黄曲霉毒素 B1 是致癌毒性最强的一种。为了筛选与黄曲霉毒素 B1 致癌作用密切相关的基因, 利用生物信息学技术对从 GEO 数据库中下载的一组相关的人基因组芯片实验数据集作系统的聚类分析, 并对候选基因进行初步的数据库检索。首先选取数据集中表达比值相差 2 倍以上的基因做聚类分析, 然后对聚集结果中表达趋势最为接近的差异基因作生物信息学数据库的检索, 最后发现 KRT15、PCNA、MMP1 等多条基因与黄曲霉毒素 B1 关系紧密。通过对三组基因组芯片数据的生物信息学分析, 找出了多条可能是黄曲霉毒素 B1 致癌毒性相关的关键基因, 为进一步研究黄曲霉毒素诱导癌症发生、发展的具体机理提供了新的思路。

关键词: 黄曲霉毒素; 致癌物; 基因表达谱; 聚类分析

文章编号: 1673-9078(2013)8-1994-1998

Bioinformatic Analysis of Genes Relative to the Carcinogenic Toxicity of Aflatoxin B1

LI Li, CAO Yi-cheng, OU Jing-shen, MAO Xiao-fan

(School of Biological Science and Engineering, South China University of Technology, Guangzhou 510640, China)

Abstract: Aflatoxin is one of the most harmful mycotoxins existing extensively in daily food, among which carcinogenic toxicity of aflatoxin B1 (AFB1) showed the highest carcinogenic toxicity. To screen genes strongly associated with the carcinogenic effects of AFB1, three AFB1-related human genome microarray datasets were downloaded from GEO database and analyzed by clustering and bioinformatic retrieving in this study. A set of genes were selected for clustering, which expression differences were greater than 2 fold in three microarray datasets. Then the genes expressed in accordant trend in the cluster analysis were screened and the genomic information about those genes and the association with AFB1 were determined. Several genes such as KRT15, PCNA and MMP1 were identified as key AFB1-related genes through bioinformatics analysis on three relative microarray datasets. These data may be benefit for further investigations of the genesis and progression of the AFB1-induced cancer.

Key words: aflatoxins; carcinogens; gene expression profile; cluster analysis

黄曲霉毒素是由黄曲霉属中的黄曲霉菌株和寄生菌株产生的一类异环芳烃类化合物。其中黄曲霉毒素 B1 (aflatoxin B1, AFB1) 是在该类毒素中含量最多, 毒性最强的一种, 具有基因毒性和细胞毒性, 能造成多种类型的 DNA 损伤, 具有明显的免疫抑制作用, 从而引起动物免疫失败、癌症和其他感染性疾病的发生^[1]。由于 AFB1 分子结构中的二呋喃环使其具有亲肝特性以及与肝炎病毒的协同作用, 目前食物中 AFB1 污染已经成为我国肝癌高发区肝细胞肝癌

(Hepatocellular carcinoma, HCC) 重要的风险因素之一^[2]。此外 AFB1 的摄入与其他一些癌症包括: 食道癌、肺癌、结肠癌等也存在着重要联系。AFB1 具有的基因毒性和致癌性是主要通过损伤 DNA 而得以实现。AFB1 进入体内后和 DNA 共价结合, 由于分子内电子场的改变, 可自发形成其它 DNA 加合物, 导致 DNA 损伤, 损伤的 DNA 进一步发展导致碱基的替换、缺失、插入等, 从而影响 DNA 的转录和复制, 最终以染色体畸变的方式表现出来, 诱导基因突变和细胞癌变的发生。因此 DNA 损伤与修复在 AFB1 致癌毒性作用中占据重要地位, 众多研究也显示出多种 DNA 修复酶与 AFB1 毒性发挥与否密切相关, 但其内在作用机理仍然有待更多研究来阐明^[3]。

基因芯片特别是全基因组芯片作为一种高效、高通量获取基因表达数据的技术, 能够避免相关调控基

收稿日期: 2013-03-29

基金项目: 国家自然科学基金重大研究项目 (90412015); 教育部中国网络计划生物信息网格平台子项目 (B12137040130)

作者简介: 李力 (1981-), 男, 博士, 研究方向为生物信息学

通讯作者: 曹以诚 (1949-), 男, 教授, 博士生导师, 研究方向为生物医药开发和生物信息学

因的作用被忽略,反映不同实验条件下细胞中全部基因转录产物的丰度差异,从而为准确地了解与实验条件相关基因的作用规律和关系提供条件。本文从GEO (Gene Expression Omnibus) 数据库^[1]中搜索下载了三组与AFB1关联的人基因组芯片实验数据集,利用生物信息学方法作系统的聚类分析,对细胞中与AFB1致癌毒性相关的基因进行挖掘,并对其中相互作用关系比较集中的基因进行生物信息学分析,为进一步研究其功能提供依据。

1 材料与方法

1.1 材料和数据

从公共基因芯片数据库 GEO (<http://www.ncbi.nlm.nih.gov/geo/>) 中下载三组与 AFB1 相关的实验数据集: GSE40117、GSE25844 和 GSE19079, 其中从 GSE40117 中选取了与 AFB1 相关的 9 个样本数据和对应的 15 个对照样本数据, 从 GSE25844 选取全部 22 个样本数据包括 10 个样本和 12 个空白对照样本, 从 GSE19079 中选取了与 AFB1 有关的 5 个样本数据及另外五个样本用作对照。三组芯片共使用了两个不同的人基因组芯片平台: Affymetrix 公司的 Human Genome U133 Plus 2.0 Array 和 Agilent 公司的 014850 Whole Human Genome Microarray 4x44K G4112F 平台, 从芯片公司下载芯片平台数据, 将数据集中的表达值矩阵与平台探针基因名称对应起来。

1.2 聚类分析

首先删除芯片平台中的非特异性探针和无指定名称的探针数据。为了尽量减少原始实验数据的误差, 本文对原始数据采用 Lowess 方法^[1]对数据进行标准化, 选取表达值标准差大于 2 的数据进行分析, 并将筛选得到基因值进行对数转换, 排除掉比值差异在 2 倍以下的基因, 从三组实验数据集中分别得到 589、330、427 个差异表达基因。利用东京大学网站提供下载的 Gene Cluster 3.0 和 Alok Saldanha 的 Java TreeView 1.1.6 软件对以上三组数据做层次聚类^[6-7]。层次聚类的距离计算采用欧式距离算法, 聚类的计算方法采用平均距离法。从聚类结果中选择表达趋势接近的基因。

1.3 差异基因的生物信息学分析

GO (Gene Ontology) 分析是生物信息领域中一个极为重要的方法和工具, 它的目的是尽最大的努力去满足不同数据库中的基因持续信息描述的需求。它包

含了细胞组分、生物过程、分子功能三个分析模块, 一个基因产物可能会出现在不止一个细胞组分里面, 也可能在很多生物过程里面起作用, 并且在其中发挥不同的分子功能。所以 Ontology 的结构是一个有向无环图, 类似于分类树, 所不同的是 Ontology 的结构中一个树形分支可以有不止一个父结构。本文使用 Gene Ontology 工具对获得的共同差异表达基因进行注释分析。并进一步通过 STRING (Search Tool for the Retrieval of Interacting Genes/Proteins)、pSTING (Protein, Signalling, Transcriptional Interactions & Inflammation Networks Gateway)、GeneCards 等分析工具^[8-10]来了解各差异基因产物之间的相互作用和相关的信号通路情况, 建立基于文献的相关网络图谱。

2 结果与分析

2.1 聚类分析结果

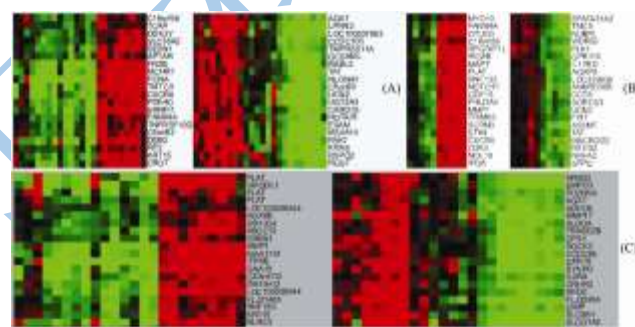


图1 三组芯片实验差异基因的聚类分析图谱

Fig.1 The clustergram of the differential genes screened from three microarray data.

注: (A)-GSE25844; (B)-GSE19079; (C)-GSE40117。

为了获得与AFB1相关的共同差异表达基因, 本文对三组相关的芯片实验数据做差异基因筛选, 然后使用层次聚类方法分析筛选出来的基因, 最后从三组聚类结果中获得表达趋势最为接近的上调和下调基因, 部分聚类图谱如图1所示, 在三个实验中筛选得到的共同差异表达的基因有107个, 其中表达上调51个, 表达下调56个。

2.2 差异基因的 GO 分析

将从三组数据中筛选出来的与AFB1有关的上、下调差异表达基因进行 GO 分析和 KEGG (Kyoto Encyclopedia of Genes and Genomes) 信号通路分析。结果发现上调的51个基因中, GO分类主要与碱基代谢、细胞表面受体信号通路、G蛋白偶联受体信号通路、初级代谢、细胞通讯和细胞的代谢及生理过程等有关, 这些说明了由于AFB1诱发产生DNA损伤, 激活了细胞

的DNA修复机制,导致核酸代谢相关基因的大量表达,另外,由于细胞表面受体基因的表达,为HBV(hepatitis B virus)及其卫星病毒对宿主细胞的感染起到了协同作用。对KEGG信号通路分析主要涉及神经刺激配体与受体之间的交互作用及多个氨基酸的生物合成。具体分析结果见表1。

表1 差异基因中表达上调基因的GO与KEGG通路分析

Table 1 The over-expressed genes analyzed by GO and KEGG

pathway			
Gene Ontology	#Genes	P Value	Bayes Factor
GO:0006139: nucleobase-containing compound metabolic process	6	<0.0001	8
GO:0007166: cell surface receptor signaling pathway	6	0.0002	5
GO:0044238: primary metabolism	7	0.0002	7
GO:0007186: G-protein coupled receptor signaling pathway	3	0.0002	5
GO:0008152: metabolism	8	0.0004	3
GO:0050875: cellular physiological process	9	0.002	4
GO:0007154: cell communication	3	0.003	5
GO:0044237: cellular metabolism	6	0.004	6
KEGG Pathway			
path:hsa04080: Neuroactive ligand-receptor interaction	4	0.004	6
path:hsa00400: Phenylalanine, tyrosine and tryptophan biosynthesis	2	0.009	3

对差异基因中表达下调的基因进行GO分析发现,主要分类为跨膜受体信号通路、酶联受体信号通路、酒精代谢、鸟氨酸循环、免疫激活等方面,一些主要发生在肝细胞内的代谢活动都受到了AFB1暴露的影响,细胞的免疫功能也受到较大的抑制。对其信号通路的分析发现他们主要涉及酪氨酸代谢、胆汁酸合成、甘油酯和脂肪酸的代谢。具体结果如表2所示。

2.3 差异基因编码蛋白的相互作用分析

STRING是一个由已知和预测的蛋白质相互作用数据组成的数据库,这些蛋白质间的相互作用包括直接和间接的相互作用,如今STRING已经包含了1133个物种的5214234种蛋白质信息。

将107个AFB1相关差异基因输入STRING工具,分析这些基因编码的蛋白质间的相互作用,结果如图2所示。从分析结果中得到,AFB1相关基因编码的蛋白间的相互作用主要集中在PCNA、PLAT、CXCR4、MMP1、KRT15、FRZB、TAT、ALDOA、AGXT、UCN2、CPS1

这11个蛋白,其他蛋白间的相互作用比较分散。

表2 差异基因中表达下调基因的GO与KEGG通路分析

Table 2 The under-expressed genes analyzed by GO and KEGG

pathway			
Gene Ontology	#Genes	P Value	Bayes Factor
GO:0007169: transmembrane receptor protein tyrosine kinase signaling	9	<0.0001	14
GO:0007167: enzyme linked receptor protein signaling pathway	9	<0.0001	10
GO:0006066: alcohol metabolism	10	<0.0001	10
GO:0006067: ethanol metabolism	3	<0.0001	9
GO:0006069: ethanol oxidation	3	<0.0001	9
GO:0000050: urea cycle	2	0.002	6
GO:0046649: lymphocyte activation	4	0.002	4
GO:0042110: T-cell activation	3	0.004	7
KEGG Pathway			
path:hsa00350: Tyrosine metabolism	3	0.0001	5
path:hsa00120: Bile acid biosynthesis	2	0.0007	3
path:hsa00561: Glycerolipid metabolism	2	0.001	7
path:hsa00071: Fatty acid metabolism	2	0.002	6

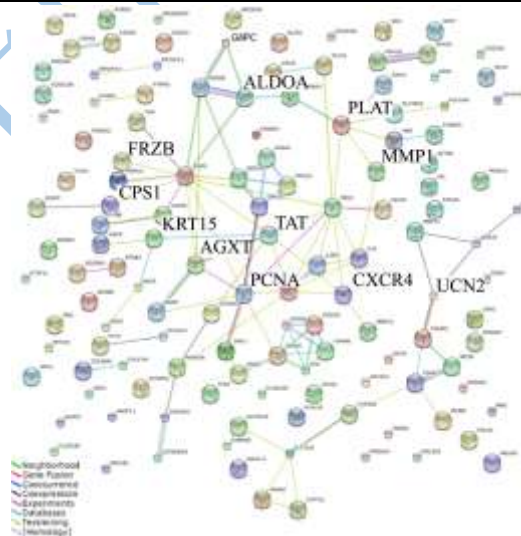


图2 差异基因编码蛋白的相互作用图

Fig.2 The Interaction network of the proteins coded by 107 differential genes

2.4 差异基因相关的通路分析

GeneCards是魏兹曼科学研究院建立的提供以基因为中心的信息自动整合及挖掘工具,现在的第三个版本使用了更复杂的检索引擎和新的底层设计,提供更为精准的基因信息。将前文得到的11个差异表达基因输入GeneCards,在通路和交互关系一栏里面得到各基因相关的通路信息。结果如表3所示,目前发现的许多信号通路都参与了AFB1对细胞毒性或致癌作用的过

程,如PLAU, MAPK, WNT通路等,从表中可以看出,差异表达的MMP1, FRZB基因分别参与了MAPK、WNT信号通路,说明这些基因很可能与AFB1的致癌毒性关系紧密。

2.5 差异基因相互调节关系网络

pSTIING是一个包含了蛋白-蛋白、蛋白-脂质体、蛋白-小分子、配体-受体相互作用信息,以及受体细胞类型信息和信号转导模型等信息的公用文献知识数据

库,它整合了文献、生化实验、功能分析与体内实验等生物学知识,为公众提供跨人、鼠、小鼠等6个物种的多重信息。主要侧重于慢性炎症、细胞运动及肿瘤的调控网络搜索蛋白结构域、基因本体注释及人类疾病间的网络关系。为了更好的了解前文得到的11个差异表达基因的相互作用网络,将这些基因输入pSTIING提供的Prot/Gene name(Batch)工具,显示参数设置为:Physical Interactions和Transcriptional Associations,保留复合网络。

表3 已知的与差异基因相关的信号通路

Table 3 The known pathways related to the 11 differential genes

基因名	信号通路	表达情况
PCNA	Gap-filling DNA repair synthesis and ligation in GG-NER/Base Excision Repair/Nucleotide Excision Repair Pathway	↑
PLAT	Cell adhesion_PLAU signaling/Selenium Pathway/Complement and coagulation cascades	↑
CXCR4	GPCR ligand binding/p38 Signaling/Cell adhesion_Integrin inside-out signaling	↑
MMP1	Degradation of the extracellular matrix/Immune response Oncostatin M signaling via MAPK in human cells/Degradation of collagen	↑
KRT15	Selected targets of p53	↑
FRZB	WNT Signaling/Adipogenesis/Endochondral Ossification	↑
TAT	Tyrosine metabolism/FOXA2 and FOXA3 transcription factor networks	↓
ALDOA	Glycolysis and gluconeogenesis/Glucose metabolism/Metabolic pathways	↓
AGXT	Peroxisome/Glycine, serine and threonine metabolism	↓
UCN2	Glucocorticoid Pathway	↓
CPS1	Urea cycle and metabolism of amino groups/superpathway of citrulline metabolism	↓

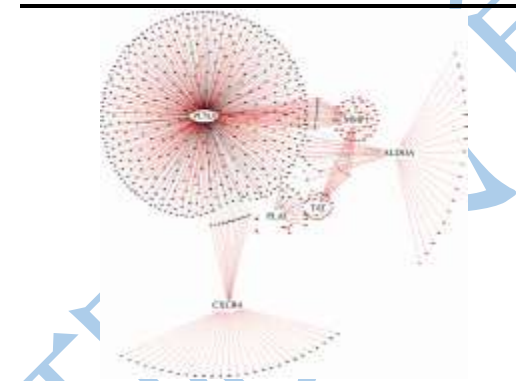


图3 差异基因的相互关系网络图谱

Fig.3 The Interaction network of the 11 differential genes

pSTIING 分析结果如上图3所示,网络图谱清楚的显示了差异基因的物理相关和转录影子相关关系,其中6个差异基因(PCNA、MMP1、TAT、ALDOA、PLAT、CXCR4)处在网络中相对关键的节点位置。

3 讨论

AFB1 在体内的基因毒性与致癌性是一个复杂的过程,是许多基因相互协同作用的结果。目前的研究认为: AFB1 在进入体内后,在细胞色素 P450 系统作

用下转变成 AFBO (AFB1-8, 9-epoxide), AFBO 能自发和核酸及蛋白质等生物大分子共价结合形成加合物,诱发 DNA 损伤并导致基因突变,最终表现为多种染色体畸变方式,诸如重复、互换、倒位和易位等。染色体畸变的形成直接影响结构基因在基因组内的正常排列,造成基因片段的丢失或重排,甚至可能改变基因的调控机理,成为 AFB1 诱导基因突变和细胞癌变的分子基础。研究还发现 AFB1 能诱导常见抑癌基因 p53 特定密码子产生特定碱基突变,即 249 号密码子的第三位碱基产生 G-T 错配^[11],而 p53 基因突变导致下游靶基因的异常表达,在三组芯片数据中均表现出表达上调的 KRT15 基因就是 p53 的靶基因,这也为 p53 基因突变频率与 AFB1 暴露程度之间的密切关系提供了佐证。

在真核细胞中,当遇到 AFB1 导致 DNA 损伤的情况时,它们会激活检查点的信号传导途径让细胞周期停下来并诱导基因的表达对 DNA 的损伤进行修复。低程度的 DNA 损伤可以由碱基切除修复机制来修复 1-13nt 内的碱基损伤,而在高水平的 AFB1 暴露的情况下,由于大量的 DNA 受到破坏,核苷酸切除修复

机制(NER)就发挥了重要的作用,对损伤的DNA链的两端进行切除^[12]。芯片实验中上调表达的增殖细胞核抗原(PCNA)基因就是核苷酸修复机制中的重要基因之一,在DNA缺失合成修复和核苷酸切除连接过程中发挥关键作用。

HBV和AFB1的协同致癌作用已经在大量的研究中得到了证明,但具体机理并不十分明确,部分研究者认为HBV病毒本身并不会引起肝细胞癌变,但HBV蛋白能影响宿主的DNA修复系统和药物代谢酶系统,抑制细胞对受损DNA的修复能力,因此一旦AFB1的代谢产物造成DNA损伤,受损的DNA由于得不到有效的修复而在体内大量累积,增加了机体对外来化合物的敏感性,最终引起肝癌发生率的增加。此外通常在HBV感染的慢性肝炎患者的血清中,基质金属蛋白酶(MMP)家族表达较健康人有着显著的升高,而该家族属于肿瘤细胞血管生成过程中负责分解细胞外基质功能的重要酶类,MMP1又是其中分解胶原等多因素的重要因子,其所表达的蛋白大量的分布于肿瘤细胞和内皮细胞胞质中。由此可见,不仅在AFB1的基因损伤和突变过程中,在后期肿瘤细胞的增值转移阶段,HBV病毒的感染也发挥了重要的协同作用。

FRZB (Frizzled motif associated with bone development) 基因是sFRP (secreted frizzled related proteins) 的家族成员,也是参与WNT(Wnt/beta-Catenin Signaling pathway)通路调节的相关基因。近来研究表明在不同种类的癌症中,Wnt信号常被异常激活^[13],而作为个体降低肿瘤细胞侵袭能力的相关基因之一,FRZB基因的表达均高于正常细胞,因此该基因也可以作为AFB1导致的肝细胞癌的早期观察指标进行研究。

Urocortin (UCN) 是促肾上腺皮质激素释放激素(CRH)肽类家族中的新成员,通过与神经肽CRF受体作用,下调肝癌组织的VEGF水平,从而抑制肿瘤生长和新生血管的生成^[14]。研究中UCN2的表达下调说明,在AFB1的暴露下,细胞的抑癌机制遭到破坏。而与细胞附着纤溶酶信号通路相关的PLAT (plasminogen activator, tissue)基因则在三组试验中表达上调,文献研究显示该基因可促进肿瘤细胞的转移和附着。这提示了在AFB1导致基因突变发生的同时还存在着与肿瘤细胞生长密切相关的多个基因的协调作用。

4 结论

本文从公共芯片数据库中选取三组与AFB1相关

的人基因组芯片实验数据,筛选其中差异表达的基因进行层次聚类,然后对得到的共同差异表达基因进行生物信息学分析。研究结果发现,11个相互作用比较集中的差异表达基因中,大部分基因参与了细胞的DNA修复机制、抑癌基因的表达和肿瘤细胞的发生、增值及转移等重要过程。揭示了AFB1的暴露在诱导癌症发生与发展过程中的部分机制,为进一步研究各相关基因的具体功能提供了新思路。

参考文献

- [1] 邱杨,刘建利,赵丽.食品中黄曲霉毒素B1的抽查检验与安全评价[J].现代食品科技,2008,24(10):1055-1057
Qiu Y, Liu J L, Zhao L. Determination and Safety Evaluation of Aflatoxin B1 in Foods [J]. Modern Food Science and Technology, 2008, 24(10): 1055-1057
- [2] 郭耀东,陈璐,袁亚宏,等.基于概率分析的玉米食品中黄曲霉毒素B1膳食暴露及风险评估研究[J].现代食品科技, 2013,5:1-6
Guo Y D, Chen L, Yuan Y H. et al. Determination and Safety Evaluation of Aflatoxin B1 in Foods [J]. Food Science, 2013, 5: 1-6
- [3] Shivender S S, Anupreet K. Aflatoxin B1: Toxicity, characteristics and analysis:Mini review[J]. Global Advanced Research Journal of Chemistry and Material Science, 2012, 1(4): 63-70
- [4] Barrett T, Edgar R. Gene Expression Omnibus (GEO): Microarray data storage, submission, retrieval, and analysis[J]. Methods in enzymology, 2006, 411: 352-369
- [5] Yang Y H, Dudoit S, Luu P. et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation [J]. Nucleic Acids Res, 2002, 30(4): e15
- [6] de Hoon M J L, Imoto S, Nolan J. et al. Open Source Clustering Software [J]. Bioinformatics, 2004, 20(9): 1453-1454
- [7] Saldanha A. Java TreeView [EB/OL]. [2009-04-19] <http://jtreeview.sourceforge.net/>
- [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, 2012, 41(D1):D808-815
- [11] Aylwin N, Boriss B, Qiong G, et al. STING: a 'systems' approach towards integrating signalling pathways, interaction and transcriptional regulatory networks in inflammation and cancer [J]. Nucleic Acids Res, 2006, 34: D527-D534

- [12] Belinky F, Bahir I, Stelzer G, et al. Non-redundant compendium of human ncRNA genes in GeneCards [J]. *Bioinformatics*, 2013, 29(2): 255-261
- [13] Roze L V, Hong S Y, Linz J E, et al. Aflatoxin Biosynthesis: Current Frontiers [J]. *Food Science and Technology*, 2013, 4: 293-311
- [14] Yves M, Brian D, Marc C, et al. DNA damage and repair in human oocytes and embryos: a review [J]. *Zygote*, 2010, 18(4): 357-365
- [15] 王震凯. Wnt 信号转导通路在肿瘤中的研究进展[J]. *医学研究生学报*, 2007, 20(12): 1294-1301
- Wang Z K. Update of the Wnt signaling path in tumor [J]. *Journal of Medical Postgraduates*, 2007, 20(12): 1294-1301
- [16] Lai J, Jie C, Shengnan L. Activation of Corticotropin-Releasing Factor Receptor 2 Inhibits the Growth of Human Small Cell Lung Carcinoma Cells [J]. *Cancer Investigation*, 2010, 28(2): 146-155

现代食品科技