

doi:10.3969/j.issn.1672-5565.2014.03.07

癌症中 DNA 甲基化基因模块筛选

张淑梅,张 彬,刘军厚,刘洪波,苏建忠,王 芳,张 岩*

(哈尔滨医科大学生物信息科学与技术学院,黑龙江 哈尔滨 150081)

摘 要:肿瘤的发生受遗传学和表观遗传学修饰的共同影响。DNA 甲基化是一种重要的表观遗传修饰,在癌症的发生与发展中起着重大的作用。因此找到癌症的甲基化标记物在癌症的诊断和治疗中具有重大意义。本文利用权重基因共表达网络分析的方法(WGCNA)筛选出甲基化基因模块,并分析模块向量基因,进行功能注释,最后对基因模块进行功能分析,得到 DNA 甲基化与肿瘤间的关系。结果显示,这些甲基化异常的基因模块与癌症的发生有着显著的关联。同时还发现某些甲基化异常的基因模块与多种癌症的发生都有着显著的关联。

关键词:表观遗传修饰;肿瘤;甲基化;WGCNA;基因模块

中图分类号: Q81 **文献标志码:** A **文章编号:** 1672-5565(2014)-03-189-07

The screening of DNA methylation gene modules in cancers

ZHANG Shumei,ZHANG Bin,LIU Junhou,LIU Hongbo,SU Jianzhong,WANG Fang,ZHANG Yan*

(College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081,China)

Abstract: Tumorigenesis is affected by both genetic and epigenetic modifications. DNA methylation acts as an important epigenetic modification and plays a major role in the occurrence and development of cancers. Therefore, finding the methylation biomarkers in cancers is a great feat in the diagnosis and treatment of cancers. In this paper, we take the advantage of Weight Gene Co-expression Network Analysis method(WGCNA) to filter out methylation gene modules and analyze the module vector genes, afterwards, the functional annotation, then we conduct functional analysis of gene modules, finally the relationship between DNA methylation and cancers can be interpreted. The results showed that these aberrant methylated gene modules have significant associations with cancers. Interestingly, we also found that some abnormal methylation gene modules have great associations with the occurrences of various cancers.

Keywords: Cancer; Epigenetic modification; DNA methylation; WGCNA; Gene modules

目前,癌症是严重威胁人类健康的三大杀手之一,对于这种严重危害人类健康的顽疾,现在的医学界并不十分清楚它的发病机制。同时,人们对基因的本质也渐渐有了更深入地认识。很长一段时间里,人们认为癌症的形成只与基因突变有关^[1-4]。但是,越来越多的证据表明,表观遗传修饰对癌症的发生也起着十分重要的作用。基因序列不变,而基因的表型发生了可遗传的变化,称为表观遗传^[5]。这是由表观遗传修饰造成的。DNA 甲基化是一种重要的表观遗传修饰,它制约着基因的表达。

据报道,人类的基因只是果蝇的 2 倍多。很难想象 DNA 的遗传信息可以调控人类这样复杂的生命体发育和生存的全过程^[1-4]。维持细胞的功能,决定哪些基因表达、哪些基因不表达,是非常重要的,几个基因的错误表达便会诱发正常细胞发生癌变^[2]。

目前普遍认为,DNA 甲基化与癌症的发生有密切关系^[6]。癌症的甲基化异常表现为总体的甲基化水平降低与启动子区域的甲基化水平升高^[7]。例如,抑癌基因与修复基因的高甲基化会导致它们的失活,造成肿瘤抑制丧失与基因损伤增加。

收稿日期:2013-11-24;修回日期:2014-01-21.

资助项目:黑龙江省大学生创新创业训练计划(201210226050)资助。

作者简介:张淑梅,女,本科生,研究方向:计算表观遗传学;E-mail:shumei_1991@163.com.

* 通信作者:张岩,女,博士,教授,博士生导师,研究方向:生物信息学;E-mail:yanyou1225@gmail.com.

由于涉及基因的“开”与“关”,DNA 甲基化对肿瘤的产生起着重大的作用。同时研究表明,某些基因的异常甲基化与多种癌症的产生有着显著的关联^[8]。例如,基因 P15 的甲基化会使基因沉默,并使细胞过度激活与增殖,而这与白血病、淋巴瘤、鳞状细胞癌、肺癌的发生都有重要的联系^[9]。是否存在一组甲基化异常的基因,与多种癌症的发生有着重要的关联以及这些基因在不同的癌症中是否起着不同的作用,成为本文关心的问题。通过研究这些问题,会为癌症的预测提供必要的方法,同时也增强了人们对癌症与 DNA 甲基化关系更进一步的认识。

表观遗传标记可以在被割除的肿瘤和体液中检测到。例如,超甲基化的癌症基因可以在尿斑中检测到,这在膀胱癌的检测中很有意义^[10]。DNA 甲基化的生物标记物在疾病诊断和预后的领域正在兴起,并且需要在临床实践中广为应用和扩展。

本课题首先通过对不同癌症 DNA 甲基化数据进行预处理,利用权重基因共表达网络分析方法(WGCNA)筛选出甲基化基因模块,并分析模块向量基因,利用 DAVID(The Database for Annotation, Visualization and Integrated Discovery)进行功能注释,然后对基因模块进行功能分析,得到 DNA 甲基化与肿瘤间的关系。本课题有助于发现癌症中 DNA 甲基化的生物标记物,为肿瘤的诊断及治疗提供可能的靶点。

1 数据及方法

1.1 数据获取

本课题所用的癌症甲基化数据来自于 GEO 数据库,包括乳腺导管癌甲基化数据(编号为 GSE14865,平台为 GPL4126,6 个样本)^[11]、胃癌甲基化数据(编号为 GSE15291,平台为 GPL4126,7 个样本)、前列腺癌甲基化数据(编号为 GSE15298,平台为 GPL4126,20 个样本)^[12]、白血病甲基化数据(编号为 GSE18400,平台为 GPL4126,60 个样本,样本为婴儿期白血病数据和 1 个对照组)、食管鳞状细胞瘤甲基化数据(编号为 GSE21238,平台为 GPL4126,6 个样本,其中包括有淋巴结转移食道癌和无淋巴结转移食道癌样本以及转移淋巴结细胞样本)、肺鳞状细胞瘤甲基化数据(编号为 GSE9622,平台为 GPL4126,5 个样本)。

首先进行数据的预处理和标准化,标准化的原则是对同一基因的不同探针对应的数值取平均值,并且只选择对应于启动子的探针。最后获得包含 4 029 个基因的甲基化数据。

1.2 权重基因共表达网络分析(WGCNA)的简介

网络分析在生物信息学中得到越来越多的应用。WGCNA(Weight Gene Co-express Network Analysis)是一种描述各个样本的基因芯片相关的系统生物学方法。这种方法可以找到高相关的基因模块,可以使用模块特征基因(eigengene)或 hub 节点间的基因彼此间和外部采样特征来聚类^[13]。相关网络促进了基于基因筛选的方法的发展,可以用于识别候选生物标记物或治疗靶点。

1.3 甲基化基因模块筛选的原理

本文通过构建权重基因共表达网络来识别癌症中甲基化基因模块。

首先,利用公式 $a_{ij} = |\text{cor}(x_i, x_j)|^\beta$ 构建邻接矩阵,其中 x_i, x_j 分别代表两个基因的甲基化值、参数 β 根据模型适应指数,使得邻接矩阵大约是无尺度的。这个指数是线性模型的确定系数(R 的平方),线性模型是在 $\log(k)$ 上回归 $\log(p(k))$ 所得到的,其中 K 代表给定节点的度, $p(k)$ 代表度 K 在网络中的频率分布。

为了便于把显著差异的甲基化基因分类成模块,邻接矩阵被转换成拓扑重叠矩阵。拓扑重叠矩阵不仅可以捕捉到 x_i, x_j 的直接互作,也可以捕捉到间接互作。这样,定义了一个相似性测度:

$$TOM_{ij} = \frac{\sum_u \alpha_{iu} \alpha_{uj} + \alpha_{ij}}{\min(k_i, k_j) + 1 - \alpha_{ij}} \tag{1}$$

其中, $k_i = \sum_u \alpha_{iu}$ 代表点的连通性。 $1 - TOM_{ij}$ 是层次聚类的距离矩阵。

1.4 基因模块的功能分析

通过 WGCNA 筛选出甲基化基因模块并量化模块与表型的关系。分析与癌症表型显著相关的基因模块。挖掘出基因模块的向量基因,并利用 DAVID 生物信息学分析工具对基因模块进行 GO 功能注释与 KEGG 通路富集研究。

2 结 果

2.1 选择合适的阈值:网络拓扑结构分析

构建一个权重基因网络,选择一个合适的邻接矩阵的阈值 β ,得到的阈值满足网络接近于无尺度的标准。通过 WGCNA,选择一组候选的阈值,并返回被检测的网络参数(见图 1)。从图中可看出阈值选择为 5 最合适,它既保证了网络接近于无尺度网络(模型指数大于 0.9,完美无尺度网络的模型适应指数是 1),同时也是使曲线趋于平滑的最小阈值,并且它也使得网络的平均链接程度不会太小,这有利于网络包含足够的信息(例如,挖掘模块)。

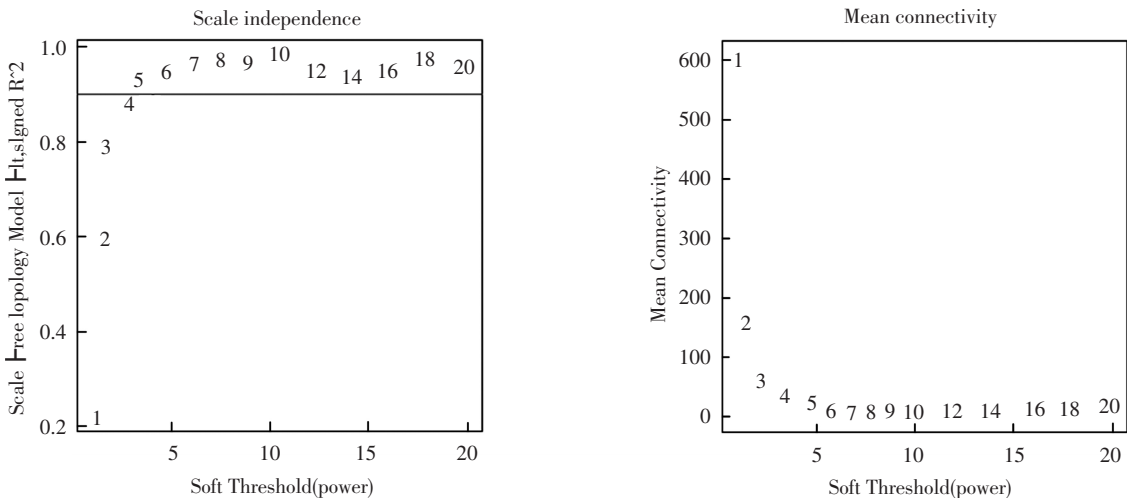


图 1 阈值分析

Fig.1 Threshold value analysis

注:左图横坐标为候选的阈值,纵坐标为对应的无尺度网络模型指数;右图代表了各阈值对应的网络的平均链接程度。

Notes:The horizontal ordinate in the left figure represent the candidate thresholds and Y-axis corresponds to the index of scale free network model; the right figure shows the mean connectivity of network corresponds to the thresholds.

2.2 使用 TOM(拓扑重叠矩阵) 聚类

为了减少噪声和伪关联的影响,将邻接矩阵转换为拓扑重叠矩阵(*TOM*)。通过 *TOM*,利用层次聚类产生一个基因的层次聚类树(见图 2)。

在层次聚类树中,通过各个分支的识别(即从树图“剪枝”)得到模块。使用 Dynamic Tree Cut 的方法^[14],期望获得较大较少的模块,所以设定参数

最小模块大小(*minModuleSize*)为 50,这样从树图中剪枝共得到 10 个模块,标签为 1 至 10,模块大小依次递减,从 806 至 65 个基因。模块 0 保存着所有模块外的基因。

树图中不同的深浅区域代表了不同的模块。找到匹配的模块,并返回各基因模块的宽度(见表 1)。

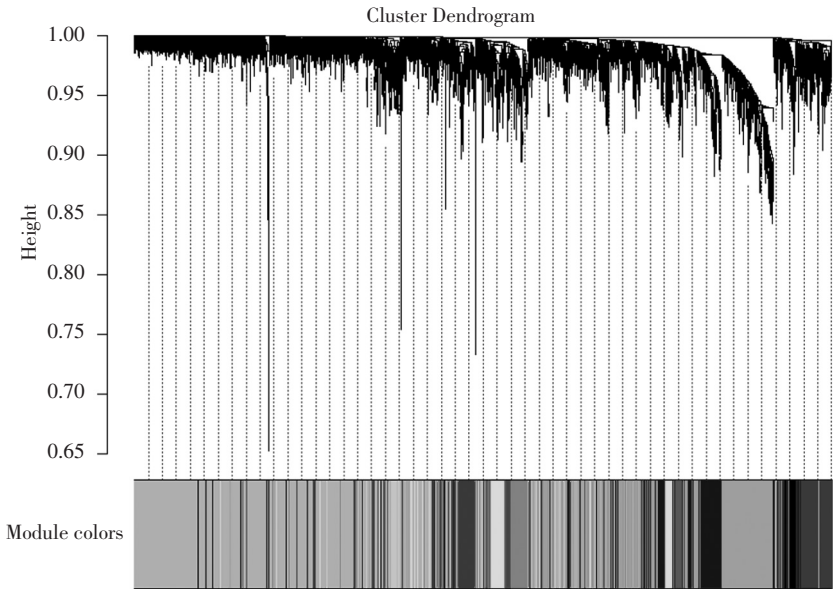


图 2 聚类树图

Fig.2 Cluster tree

注:树图中每个叶节点代表一个基因,其中密集连接的分支代表了甲基化数值接近的基因;图中不同的颜色代表不同的模块。

Notes:In the tree diagram, every leaf node represents a gene, and branches densely connected represent the genes which have the similar methylation values. The different colors represent different modules.

表 1 各基因模块中的向量基因
Table1 Vector genes in the modules

模块	0	1	2	3	4	5
基因	1 620	806	372	322	277	191
模块	6	7	8	9	10	
基因	160	81	68	67	65	

2.3 量化模块与表型的关联

分析模块与模型的显著关联。由于已有每个模块的特征基因 (eigengene), 使特征基因 (eigengene) 与表型相关联, 并找到最大相关性。由于已有模块与表型, 可以可视化这种关联, 用颜色标注相关性。

图 3 中可以清晰地看到模块与癌症表型的相关性。模块 0 的基因是树图中剔除的基因, 从图 3 中也

可看出它与各癌症表型的相关性较差, 因此不予考虑。

可以看到, 胃癌 (gastric cancer) 与模块 1、4、10 有较强的相关性 ($p \leq 0.05$); 前列腺癌 (prostate cancer) 与模块 2 有较强的相关性 ($p \leq 0.05$)。说明这两种癌症在上述模块中甲基化程度较高。

同时, 可以看到各种婴儿期白血病 (ALL), 如 MLL-AF4 白血病、MLL-ENL 白血病、未扩散白血病 (Untranslocated infant ALL) 在模块 7、9 都有较强的相关性 ($p \leq 0.05$), 而正常人体细胞在这两个模块中 p 值都大于 0.05, 没有显著的相关性。由于选用的基因来自于启动子区域, 可以得出结论: 在上述基因模块中, 白血病对应基因的甲基化程度要比正常细胞的甲基化程度高。

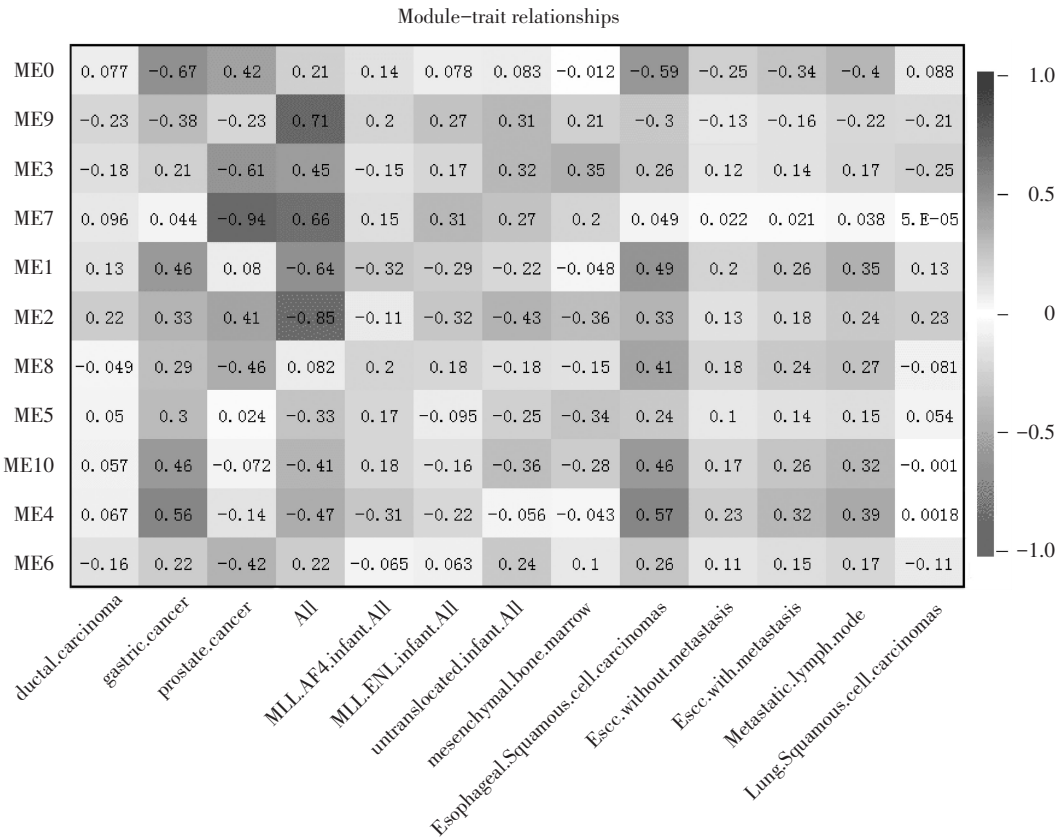


图 3 模块与癌症表型关系图
Fig.3 Module-trait relationship

注: 每一行对应一个模块, 每一列对应一个癌症表型, 其中每一单元对应相关值。
Notes: Each row corresponds to a module and each column corresponds to a phenotype, each cell represents the relative value between module and cancer.

食道鳞状细胞瘤 (Esophageal Squamous Cell Carcinomas) 在模块 1、4、10 都有显著的相关性。同时还发现, 有淋巴结转移食道鳞状细胞瘤 (ESCC with metastasis) 和转移淋巴结 (Metastatic lymph node) 比无淋巴结转移食道鳞状细胞瘤在上面的模块中具有更高的相关性。可以得出结论: 在上述模块中, 有淋巴结转移食道鳞状细胞瘤比无淋巴结转

移食道鳞状细胞瘤对应基因的甲基化程度高。同时, 也可看到胃癌与食管鳞状细胞瘤在模块 1、4、10 都有显著的相关性, 说明在这几个基因模块中, 两者甲基化程度较高。

2.4 基因模块的功能分析

为了进一步了解上面的基因模块与癌症发生与发展的关系, 挖掘上面得到的模块的向量基因。并

对这些向量基因进行基因本体功能分类及生物学通路分析。

首先,对与胃癌与食管鳞状细胞瘤显著相关的模块 1、4、10 的向量基因进行功能注释。这些模块中共得到 1 148 个基因。

通过 DAVID 分析,1 148 个基因有 647 个注释到了 189 类生物学过程,其余为未知功能基因。设定阈值为 $p \leq 0.05$,则基因注释到 96 类生物学过程。这些生物学过程主要包括:基因沉默,蛋白质降解过程,己糖降解,Wnt 受体信号通路,蛋白激酶活性负调节等(见表 2)。

表 2 与胃癌和食管鳞状细胞瘤显著相关的模块向量基因的功能富集聚簇

Table 2 Functional annotation for module vector genes significantly associated with gastric cancer and ESCC

GO term with $P \leq 0.05$	Count	Gene in Go Term
Proteasomal protein catabolic process	19	ANAPC1, RAD23B, TBL1XR1, EDEM1, SELS, OS9, CUL3, USP19, PSMA1, PSMC6, MAD2L1, PSME1, VCP, PSMD11, PSMC2, PSMA3, PPP2CB, BUB1B, PSMD3
gene silencing	12	EIF2C2, ZCCHC11, HDAC2, EIF2C1, DGCR8, MAEL, SIRT5, MBD3, KCNQ1, EIF2C3, HELLS, ADAR
negative regulation of catalytic activity	32	PALM, ADCY4, GNAI1, OPRK1, ADCY6, GABBR1, PYDC1, PDCD4, LATS2, GPX1, SPRY2, HRH3, HEXIM1, ARG2, PSMD3, ANAPC1, DUSP22, CDKN1C, PSMA1, PSMC6
proteolysis involved in cellular proteincatabolic process	68	TSG101, VPRBP, OS9, CUL3, USP19, MAP1LC3A, PSMD3, USP15, USP13, ANAPC1, TBL1XR1, FBXL21, USP1, GAN, UBE2I
Wnt receptor signaling pathway	18	CSNK1A1, DIXDC1, TBL1XR1, WNT10A, NKD1, CSNK1G2, WNT3A, CTNND1, TLE2, FRZB, FZD4, FZD6, WNT4, KREMEN2

注:表中为部分基因部分注释,count 代表注释到 GO term 中的基因个数。

Notes: The table illustrates part of the annotation in some genes, Count represents the number of genes annotated into the GO terms.

同理,对与前列腺癌显著相关的模块 2 的向量基因进行功能注释。注释的 372 个基因有 204 个注释到了 127 类生物学过程。设定阈值为 $p \leq 0.05$,则

基因注释到 79 类生物学过程。这些生物学过程主要包括:蛋白激酶活性负调节,细胞增殖调节,调控细胞死亡,参与细胞形态分化等(见表 3)。

表 3 与前列腺癌显著相关的模块向量基因的功能富集聚簇

Table 3 Functional annotation for module vector genes significantly associated with prostate cancer

GO term with $P \leq 0.05$	Count	Gene in Go Term
negative regulation of protein kinase activity	10	SH3BP5, TARBP2, YWHAG, PAK2, PSEN1, NF1, GADD45G, MLLT1, PKIB, GPS2
regulation of programmed cell death	36	HTATIP2, TRAIP, GCLM, SRC, MOAP1, BLOC1S2, BOK, NPM1, PLEKHG5, CHST11, TDGF1, TGM2, HSPA5, FGF4, EGFR, TP53BP2, NF1, MAL, SR2, YWHAE, TAX1BP1, SOD2, SMO, AMIGO2, TNFRSF10B, UACA
negative regulation of molecular function	18	CAST, NF1, PKIB, GPS2, SH3BP5, PSMF1, SMO, TARBP2, YWHAG, HTR1A, PSEN1, PAK2, GADD45G, PSMD1, MLLT1, ID3, HSPA5, UBA52
positive regulation of cell proliferation	24	EGFR, CDC7, UBE2A, NAP1L1, IRS1, HES1, SMO, HTR1A, HIF1A, BLOC1S2, NME1, NCK1, NPM1, CD81, TDGF1, TGM2, MDM2, LAMB1, CCNA2, TSHR, PGGT1B, RHOG, ALOX12, FGF4
negative regulation of transferase activity	10	SH3BP5, TARBP2, YWHAG, PAK2, PSEN1, NF1, GADD45G, MLLT1, PKIB, GPS2

注:表中为部分基因部分注释,count 代表注释到 GO term 中的基因个数。

Notes: The table illustrates part of the annotation in some genes, Count represents the number of genes annotated into the GO terms.

同理,对与白血病显著相关的模块 7、9 的向量基因进行功能注释。注释的 148 个基因有 141 个注释到了 27 类生物学过程。设定阈值为 $p \leq 0.05$,则基因注释到 14 类生物学过程。这些生物学过程主要包括:磷酸代谢过程;mRNA 代谢过程;转录调控;磷酸化蛋白质氨基酸等(见表 4)。

表 4 与白血病显著相关的模块向量基因的功能富集聚簇

Table 4 Functional annotation for module vector genes significantly associated with leukemia

GO term with $P \leq 0.05$	Count	Gene in Go Term
phosphorus metabolic process	19	ATP5E,PTPRK,PDP2,TRPM7,SSH3,MAP2K4,MKNK2,CDKL2,EPHB4,ATP5G3,MARK2,CDKL1,ERN1,LMTK2,SYNJ2,PRKAA1,BMP7,IGFBP3,MELK
mRNA metabolic process	9	SFRS6,DHX8,DDX46,MAGOH,ERN1,PTBP2,PRPF18,SNRPF,KIAA0427
transmembrane	17	MRPS17,ATP5E,SLC13A5,KCNAB3,MFSD3,TRPM7,MFSD1,ATP5G3,SLC2A8,KCNS3,NUP214,SLC26A5,SLC2A2,SLC37A2,KCNQ2,EIF5A2,SLC25A17
transport		
regulation of translation	6	TNRC6C,MAGOH,MKNK2,GCN1L1,EIF5A2,KIAA0427
posttranscriptional regulation of gene expression	7	TNRC6C,MAGOH,MKNK2,GCN1L1,EIF5A2,LIN28B,KIAA0427

注:表中为部分基因部分注释,count 代表注释到 GO term 中的基因个数。

Notes: The table illustrates part of the annotation in some genes, Count represents the number of genes annotated into the GO terms.

接着,对与胃癌与食管鳞状细胞瘤显著相关的模块 1、4、10 的向量基因进行 KEGG 通路分析。这些模块中共得到 1 148 个基因。

通过 DAVID 分析,对向量基因进行生物学通路富集分析。采用 Fisher 精确检验, $p < 0.05$ 表示一系列基因能代表与某些生物学通路相关的生物学功能发生了改变。本次分析中通路发生改变的主要有:产生癌症(Pathways in cancer),产生肾上皮细胞癌(Renal cell carcinoma)(见表 5)。

表 5 基因模块 1、4、10 生物学通路中富集情况

Table 5 Gene modules 1、4、10 biological pathway enrichment

Term	Count	P-value	Genes
Pathways in cancer	43	4.4E-3	FGF19,E2F2,FGFR1,PGF,GRB2,WNT3A,ARNT2,TGFB1,GLI1,WNT4,CDKN2B,SLC2A1,TGFA,WNT6,MYC,FGF3,TRAF4,FN1,AKT2,MSH6,WNT10A,RET,MSH3,MAP2K2,MET,BRCA2,PRKCG,FADD,CTNNA1,BIRC2,FZD4,RALGDS,DAPK1,RAD51,FZD6,CBLB,CDKN1B,HDAC2,LAMA5,NTRK1,PLCG2,TCEB2,PIAS1
Renal cell carcinoma	13	7.1E-3	MAP2K2,PGF,GRB2,MET,ARNT2,TGFB1,GAB1,SLC2A1,TCEB2,TGFA,RAP1A,RAP1B,AKT2

通过以上的富集情况,发现基因模块 1、4、10 的相关基因富集到了产生癌症的通路。由于基因的启动子区域甲基化程度较高,会产生抑制表达的作用。基因表达的缺失导致低氧诱导因子(缺氧诱导因子- α)的积累,从而产生多种生长因子,如血管内皮生长因子和血小板衍生生长因子,使细胞运动、细胞转化、防止细胞凋亡等生物学效应的调节功能缺失,造成了肿瘤的生成。

同时,模块 1、4、10 的基因也富集到了产生肾上皮细胞癌的通路,这也说明了相关基因启动子区域的甲基化程度较高,影响到多种癌症的发生。

再对与前列腺癌显著相关的模块 2 的向量基因进行 KEGG 通路分析。通过 DAVID 分析,采用 Fisher 精确检验,本次分析中通路发生改变的是:细胞分裂周期(见表 6)。

表 6 基因模块 2 生物学通路中富集情况

Table 6 Gene module 2 biological pathway enrichment

Term	Count	P-value	Genes
Cell cycle	10	4.1E-3	CDC7,MAD1L1,CCNB3,YWHAG,GADD45G,MDM2,ATR,CCNA2,YWHAH,CDC25A

通过上面的富集,基因模块 2 的相关基因富集到影响细胞分裂周期的通路上。有丝分裂是一个重

复序列的过程,细胞周期蛋白依赖性激酶(CDKs)是关键调节酶,它通过调控细胞基质来控制细胞进程。细胞周期蛋白依赖性激酶抑制剂(CKIs),如基因CDC7、MAD1L1、CCNB3等,参与CDKs的负调控,从而提供了一个通过该细胞周期负调控的通路。而它又反过来激活p53抑癌蛋白。基因CDC7、MAD1L1、CCNB3的高甲基化,抑制p53的表达,同时细胞不能进行正常分裂,从而造成肿瘤细胞的产生。

3 讨论

目前的研究认为DNA甲基化与肿瘤密切相关。肿瘤的DNA甲基化改变表现为总体的甲基化水平降低与启动子区域CpG岛的甲基化水平升高。所筛选的基因模块的向量基因的甲基化水平普遍较高,就是由于基因的启动子区域CpG岛的甲基化异常造成的。

通过对基因模块进行GO功能注释,发现了与各癌症显著相关的甲基化异常的基因模块内的相应基因注释到了诸如基因沉默,Wnt受体信号通路;蛋白激酶活性负调节;细胞增殖调节;调控细胞死亡;参与细胞形态分化等生物过程,而这些生物学过程又与癌症的发生有着显著的关联。说明这些甲基化异常的基因模块对肿瘤的发生与发展起着重大的作用。

同时,对与胃癌与食管鳞状细胞瘤显著相关的模块1、4、10的向量基因进行生物学通路富集分析,得到产生癌症的通路。说明甲基化异常的基因模块确实与肿瘤的生成有着重要的联系。而对于与胃癌与食管鳞状细胞瘤显著相关的模块1、4、10富集到产生肾上皮细胞癌的通路。也说明了甲基化异常的基因模块同时与多种癌症的发生有着千丝万缕的联系。

在本课题中,首先下载了乳腺导管癌、胃癌、前列腺癌、白血病、肺鳞状细胞瘤、食管鳞状细胞瘤等6种癌症及亚型的DNA甲基化数据,经过预处理后利用WGCNA筛选出了甲基化基因模块,通过量化模块与癌症表型的关系发现了与各癌症显著相关的6个基因模块。然后,挖掘这些基因模块的向量基因,对这些基因进行GO功能注释和KEGG通路富集分析。通过GO功能注释发现了基因模块内相应的基因与可能导致肿瘤产生的生物学过程有关;利用KEGG数据库对基因模块的向量基因进行功能聚类,发现模块内的基因富集到产生癌症的通路也说明甲基化异常的基因模块与癌症的发生有着显著的内在关联。同时,也发现某些甲基化异常的基因模块(模块1、4、10)与多种癌症的发生有着显著的关联。基于此,本文有助于发现癌症中的DNA甲基化的生物标记物,为肿瘤的诊断及治疗提供可能的靶点。

参考文献(References)

- [1] WU C, MORRIS J R. Genes, genetics, and epigenetics: A correspondence [J]. *Science*, 2001, 293(5532): 1103-5.
- [2] WOLFFE A P. Chromatin remodeling: Why it is important in cancer [J]. *Oncogene*, 2001, 20(24): 2988-90.
- [3] PENNISI E. Behind the scenes of gene expression [J]. *Science*, 2001, 293(5532): 1064-7.
- [4] VALLBOHMER D, BRABENDER J, YANG D, et al. DNA methyltransferases messenger RNA expression and aberrant methylation of CpG islands in non-small-cell lung cancer: association and prognostic value [J]. *Clinical Lung Cancer*, 2006, 8(1): 39-44.
- [5] ALVAREZ-VENEGAS R, AVRAMOVA Z. Methylation patterns of histone H3 Lys 4, Lys 9 and Lys 27 in transcriptionally active and inactive Arabidopsis genes and in atx1 mutants [J]. *Nucleic Acids Res*, 2005, 33(16): 5199-207.
- [6] AHMAD I, RAO, D. N. Chemistry and biology of DNA methyltransferases [J]. *Critical Reviews in Biochemistry and Molecular Biology*, 1996, 31(5-6): 361-380.
- [7] VERTINO P M, YEN R W, GAO J, et al. De novo methylation of CpG island sequences in human fibroblasts overexpressing DNA (cytosine-5-)-methyltransferase [J]. *Mol Cell Biol*, 1996, 16(8): 4555-65.
- [8] AHUJA N, LI Q, MOHAN A L, et al. Aging and DNA methylation in colorectal mucosa and cancer [J]. *Cancer Res*, 1998, 58(23): 5489-94.
- [9] WHEELER J M, BECK N E, KIM H C, et al. Mechanisms of inactivation of mismatch repair genes in human colorectal cancer cell lines: the predominant role of hMLH1 [J]. *Proc Natl Acad Sci U S A*, 1999, 96(18): 10296-301.
- [10] BAYLIN S B. Tying it all together: epigenetics, genetics, cell cycle, and cancer [J]. *Science*, 1997, 277(5334): 1948-9.
- [11] MURAKAMI J, ASAUMI J, MAKI Y, et al. Influence of CpG island methylation status in O6-methylguanine-DNA methyltransferase expression of oral cancer cell lines [J]. *Oncol Rep*, 2004, 12(2): 339-45.
- [12] NEPHEW K P, HUANG T H. Epigenetic gene silencing in cancer initiation and progression [J]. *Cancer Lett*, 2003, 190(2): 125-33.
- [13] ZHANG B, HORVATH S. A general framework for weighted gene co-expression network analysis [J]. *Stat Appl Genet Mol Biol*, 2005, 4(1): 1-43.
- [14] LANGFELDER P, ZHANG B, HORVATH S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R [J]. *Bioinformatics*, 2008, 24(5): 719-20.