大肠癌转移相关分子标签的筛选*

齐 鲁[1] 袁 理[1] 吴 萍[1] 冶亚平[1] 丁彦青[1,2△]

(1 南方医科大学基础医学院病理学系 广东 广州 510515;2 南方医科大学南方医院病理科 广东 广州 510515)

摘要 目的:筛选与转移相关的大肠癌分子标签。**方法:**本文通过对大肠癌表达谱数据进行分析,按照表达谱中大肠癌转移情况,组织学分化程度以及患者生存时间进行显著性分析,将与大肠癌转移相关的具有显著性意义的基因群进行聚类,通过主成分分析以及自组织映射的方法计算出差异表达基因中,起着主体分类作用的基因群。**结果:**筛选出与肿瘤组织分化程度以及患者生存时间相关的差异表达基因,进行功能富集,并筛选出了一批与大肠癌转移密切相关的重要基因,这些基因对大肠癌早期诊断,及时治疗,预后评估有着重要意义。**结论:**细胞代谢,趋化因子信号通路和细胞因子受体等分子事件与大肠癌分化程度密切相关,是否发生转移与大肠癌的预后生存期密切相关。

关键词:大肠癌;转移;分子标签

中图分类号:R735.34 **文献标识码:**A **文章编号:**1673-6273(2012)27-5225-05

Screen Molecular Signatures Related to Colorectal Cancer Metastasis*

QI Lu [1], YUAN Li[1], WU Ping[1], YE Ya-ping[1], DING Yan-qing[1,2△]

(1Department of Pathology Southern Medical University, Guangzhou, Guangdong, 510515, China;

2 Department of Pathology Southern Hospital, Guangzhou, Guangdong, 510515, China)

ABSTRACT Objective: To screen the molecular signature that related to the colorectal cancer metastasis. Methods: In this study, gene expression data of colorectal cancer were analyzed. Significant analysis were based on the gene expression data of colorectal cancer metastasis, histological differentiating degree and patients' survival time. The significant meaning genes that related to the colorectal cancer metastasis were clustered. The subject classification genes in differently expressed genes were calculated by the way of principal component analysis and self-organizing mapping calculate. Results: Differently expressed genes that related to differentiation degree of tumor tissue and patients' survival time were selected, then its function was enriched. A collection of genes closely related to colorectal cancer metastasis was selected. These genes were great significance for the early diagnosis, prompt treatment and prognostic appraisal of colorectal cancer. Conclusion: Cell metabolism, signaling pathway of chemokine, cytokine receptor and other molecular event are closely related to histological differentiating degree of colorectal cancer. Metastasis or not is closely related to patients' survival time.

Key words: Colorectal cancer; Metastasis; Molecular signature

Chinese Library Classification(CLC): R735.34 Document code: A

Article ID:1673-6273(2012)27-5225-05

前言

大肠癌晚期可通过淋巴道或血道进行淋巴结或远隔器官转移[1],转移是大肠癌患者死亡的主要原因[2],而血道转移远比淋巴结转移危害大[3],目前主要治愈大肠癌的方法为手术切除[4,5],因此转移极大地影响了大肠癌的治疗和预后。所以及早发现大肠癌,在转移前进行手术切除,对于治愈大肠癌有着重要意义[6]。大肠癌转移相关分子标签,是能够在大肠癌转移中起着重要识别作用的基因,这些基因在转移与非转移肿瘤组织中表达差异明显[7]。但单个基因的表达情况往往难以区分转移与非转移的肿瘤,因此需要多个基因构成的共同表达模式才能够用于区分肿瘤的转移情况[8]。现有的研究在大肠癌转移相关分子标签方面研究较少,通常只研究单个分子标签[9,10]或者少数几个分子标签[10,11],其筛选出的分子标签并不能很好的区分大肠癌的转移情况,本文对大肠癌表达谱进行分析,筛选出的差异表达基因在聚类分析中能够明显区分早期大肠癌与转移后的大肠癌表达谱数据,并通过主成分分析筛选出具有相似表达模式的分子标签集,期望通过这些分子标签集,来尽早地预测大肠癌的转移情况。

1 材料和方法

1.1 对大肠癌分化程度和生存期相关差异基因的显著性分析

取来自 GEO 数据库中的芯片数据 GSE17538 系列中的大肠癌数据进行分析[12],此系列本身包含 238 例表达谱数据,去掉其中 6 例为小鼠数据,剩下的 232 例大肠癌表达谱数据通过显著性分析软件 SAM3.01 按照数据中的大肠癌组织分化程度以及确诊大肠癌后的生存时间进行显著性分析[13]。此芯片平台为 HG-U133 Plus 2.0[14],芯片平台中共有 54613 个探针,对差异

* 基金项目:国家自然科学基金资助项目(30670967)

作者简介:齐鲁(1986-),男,硕士研究生,主要研究方向:生物信息学,E-mail:kuuga888@qq.com

△通讯作者:丁彦青,E-mail:dyq@fimmu.com

(收稿日期:2012-02-25 接受日期:2012-03-21)

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

度不明显的基因进行剔除，经过筛减后用于后续分析的探针数为 26765。数据中分化程度分为三等，高分化，中分化，和低分化，有分化程度资料的标本例数有 206 例，有生存期资料的标本例数有 226 例，经过显著性分析，取假阳性率 FDR 值小于 1%的差异基因数，去除重复探针后，得到与大肠癌分化程度相关的差异基因共有 332 个，与生存期相关的差异基因有 235 个。用 FunNet 在线工具对上述差异基因进行通路分析，可以得到这些基因所参与的分子事件。

1.2 对大肠癌远隔器官转移相关差异表达基因的主成分分析

取来自 GEO 数据库中的芯片数据 GSE2109 系列中的大肠癌数据进行分析，此系列本身包含 2158 个各种不同肿瘤的数据并包含各种临床信息。在此主要研究大肠癌，因此在这 2158 例各种不同肿瘤表达谱数据中挑选大肠癌数据并包含 TNM 分期数据的表达谱数据进行分析，共挑选出 343 例大肠癌数据，从 343 例大肠癌表达谱数据中挑选出属于 T1，T2 和 M1 分期的大肠癌表达谱数据共 121 例进行分析，将 TNM 分期中属于 T1，T2 的归为一类，M1 的归为一类。通过显著性分析软件 SAM3.01 按照 T1，T2 这类与 M1 这类进行显著性分析，取假阳性率 FDR 值小于 1%的差异基因数，去除重复探针后，得到与大肠癌转移相关的上调基因数有 177 个，下调基因数有 382 个。对上调和下调的基因通过 matlab7.90 聚类算法进行聚类分析，为了更精细的筛选出与大肠癌转移密切相关的分子标签，并能够很好的将大肠癌是否发生转移的标本区分，因此对这些基因通过 matlab7.90 主成分分析及自组织映射算法进行主成分分析以及自组织映射，得到与大肠癌转移关系密切的重要基因集。

根据 Matlab7.90 中所提到的聚类和主成分分析及自组织映射算法如下：

```
load Colorectalcancer.mat
mask = genevarfilter(Colorectalcancer);
Colorectalcancer = Colorectalcancer(mask,:);
genes = genes(mask);
numel(genes)
[mask, Colorectalcancer, genes] = genelowvalfilter(Colorectalcancer,genes,'absval',log2(3));
[mask, Colorectalcancer, genes] = geneentropyfilter(Colorectalcancer,genes,'prctile',15);
clustergram (Colorectalcancer (:,1:end),'RowLabels',genes, 'ColumnLabels',type(1:end))
mapcaplot(Colorectalcancer,genes)
[pc, zscores, pcvars] = princomp(Colorectalcancer);
pcvars./sum(pcvars) * 100
cumsum(pcvars./sum(pcvars) * 100)
figure
scatter(zscores(:,1),zscores(:,2));
xlabel('First Principal Component');
ylabel('Second Principal Component');
title('Principal Component Scatter Plot');
P = zscores(:,1:2)';
net = newsom(P,[4 4]);
net = train(net,P);
figure
plot(P(1,:),P(2,:),'.g','markersize',20)
hold on
plotsom(net.iw{1,1},net.layers{1}.distances)
hold off
```

2 结果

2.1 大肠癌分化程度和生存期相关差异基因的显著性分析结果

用 FunNet 在线工具对上述 332 个大肠癌分化程度相关的差异基因以及 235 个与大肠癌生存期相关的差异基因进行通路分析[15]。可以发现，与大肠癌分化程度相关 332 个差异基因主要参与了细胞代谢，趋化因子信号通路和细胞因子受体等分子事件（图 1），说明分化越低的大肠癌细胞，代谢越旺盛，分泌的趋化因子以及合成细胞因子受体能够有效促进大肠癌变形，游走，阿米巴样运动，进而突破血管壁，容易随血流进行远隔器官转移。接着对与生存期相关的 235 个差异基因进行通路分析可以发现，相关的信号通路主要参与了细胞粘附，ECM 受体相互作用，阿米巴样运动等与大肠癌转移高度相关的分子事件，因此可以看出，是否发生大肠癌转移与患者生存期有着密切联系。与分化程度相关的显著性最高的前十个基因分别为：AQP3、DUSP4、ANXA10、SLAIN1、PCSK1、CFTR、GBP5、SMAD9、MUC5AC、HSPA2。与生存期相关的显著性最高的前十个基因分别为：AKAP12、ULBP2、AHNAK2、SPON1、DCBLD2、FABP4、SPOCK1、ETV1、PLK2、PI15。

2.2 大肠癌远隔器官转移相关差异表达基因的主成分分析结果

2.2.1 大肠癌转移相关表达上调基因的主成分分析结果 对大肠癌转移相关的表达上调的 177 个差异基因的数据进行聚类分析（图 2）。其中图下标记为 1 的部分为属于 T1T2 这类的样本，显示为绿色，标记为 2 的部分为属于 M1 这类样本，显示为红色，所以 M1 类样本这些基因表达上调，T1T2 类样本这些基因表达下调。并可以发现，高表达红色标记的位点和低表达绿色标记的位点分界较为明显，因此这些基因能够初步区分出 T1，T2 与 M1 分期大肠癌样本。对这些基因进行主成分分析，经计算，第一主成分和第二主成分能够代表原 121 例表达谱差异数据的 94%，因此能够很好的代替原数据进行分析，将第一主成分和第二主成分进行二维作图并进行分类（图 3）。将聚集度最高的顶点部分的基因提取出来，此部分基因为大肠癌转移的关键基因并高度聚集，说明功能最为相近，这些基因包括：SPP1、NRP2、UACA、CLEC11A、ITGB5、DOK5、DACT1、LOX、TMCC1、SLIT2、S1PR3、CTNNB1、UNC5B、NOX4、GLIS2、KLC1、MAML2、HMCN1、LOC100132999、SCUBE3、LOC400043、ATXN1、THBS3、LTBP3、MRC2、MAP4K4、PHF20L1、RUNX2、ZNF532、ETV5、SYNPO、OLR1、RUNX1、CD109、RGS4、WWTR1、ZFHX3、PLXDC2、FSTL3、ZNF449、AMACR、TNRC6A、RCN3、HNT、GFPT2、ANTXR1、SOCS5、PHLDB1、DBT、C14orf37、TREM2、PFTK1、PTK2、NOV、SLC2A3、PRKD1、KIF1B、VEGFC、

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

XYLT1、SLC2A14、DFNA5、PDGFRL、SLC16A4、YAP1、LRRC32、CLDN11、TRPC1、MGC24103、LOC100132891。对数据进一步进行自组织映射分析，构建自组织映射图（图 4），可以看出，功能相关的基因都聚集于一处，红色点聚集度很高，因此有着很高的特异性。

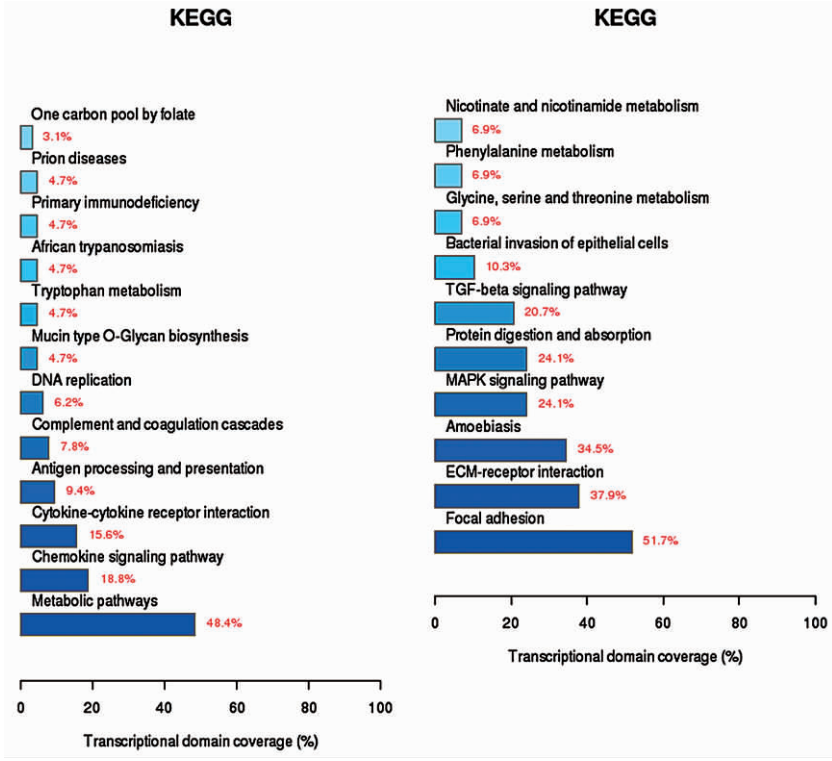

图 1 与生存期相关差异表达基因的功能富集

Fig.1 Enrich the function of genes related to patients' survival time

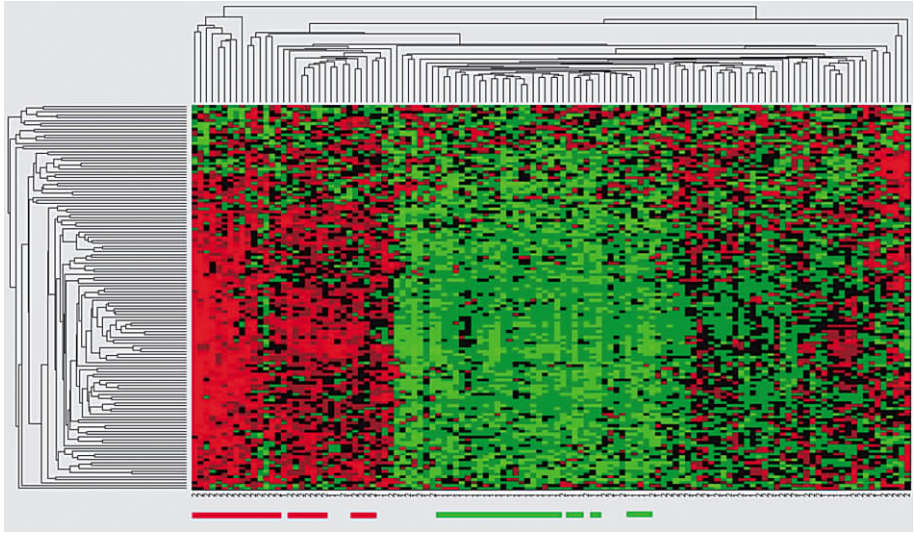

图 2 表达上调的差异表达基因聚类图

Fig.2 Dendrogram of differentially expressed genes that up-regulated expressed

2.2.2 大肠癌转移相关表达下调基因的主成分分析结果 对提取经过显著性分析得到的与大肠癌转移相关的表达下调的差异基因的数据进行聚类分析（图 5）。其中图下标记为 1 的部分为属于 T1T2 这类的样本，显示为红色，标记为 2 的部分为属于 M1 这类样本，显示为绿色，所以 T1T2 类样本这些基因表达上调，M1 类样本这些基因表达下调。高表达的红色标记的位点和低表达绿色标记的位点分界也较为明显。进行主成分分析，经计算，第一主成分和第二主成分能够代表原 121 例表达谱差异数据的 91.6%，因此能够很好的代替原数据进行分析，将第一主成分和第二主成分进行二维作图并进行分类（图 6）。将聚集度最高的顶点部分的基因提取出来，此部分基因为大肠癌转移的关键基因并高度聚集，说明功能最为相近，因此提取出这些基因，这些基因包括：MOBKL2B、COX11、CALCOCO2、CASP1、MUC4、SLC25A10、HELLS、IGHG1、CANT1、C4orf19、

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

C6orf136、ERAL1、PLK4、DNAJC11、TNFRSF11A、C9orf40、GGT6、FAM55A、CCNA2、ARF6、SULT1B1、COX10、BUB1、ATP2A3、GNA11、CDCA2、HHLA2、CNOT1、HOOK1、CORO2A、SLC25A11、WDR68、AOF1、ARHGDIA、PPP5C、CNNM4、TPMT、MLX、PLK1、LETM1、C12orf49、DLAT、BIRC5、CDC42SE2、PAG1、RBM14、GBA2、SELI、CHAC2、C9orf100、G3BP2、OIP5、NR3C2、CXCR6、SMC2、IRF4、ZWILCH、SPCS3、GPT、AMN、PARP3、NAT2、DGCR8、MTUS1、CDC6、MPHOSPH9、TIPIN、LRRC1、RNF157、CARD16、C6orf105、LOC153346、HINT1、IGHA1、SOLH、ACVRL1、MYO19、XPO7、C15orf23、KBTBD11、LARS2、CAPN5、ELL3、CCNYL1、SLAMF7、METTL2A、CENPM、ZFP91、LPCAT4、C12orf48、ACOT11、ERAP1、NCAPG2、MGC13057、FBF1、RAB27A、SCYL2、RAB11FIP4、LAMP3、TC2N、HAPLN1、BCL10、TYMS、MCM5、CDC45L、SSX2IP、CHAF1A、TPK1、DERL3、CENPH。对数据进一步进行自组织映射分析，构建自组织映射图（图 7），可以看出，功能相关的基因都聚集于一处。

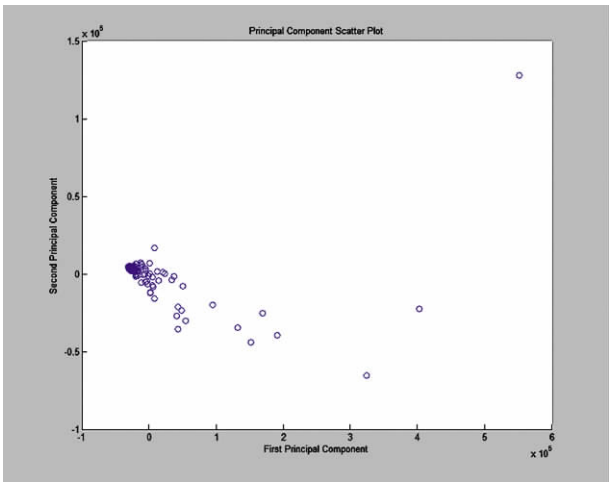

图 3 主成分分析图

Fig.3 The map of PCA

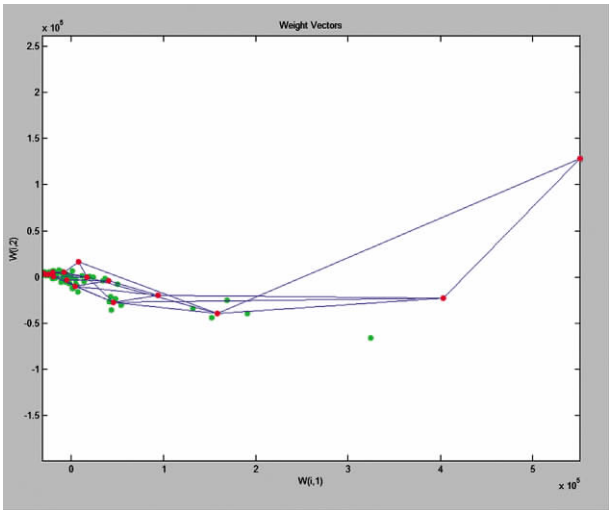

图 4 自组织映射图

Fig.4 The map of SOM

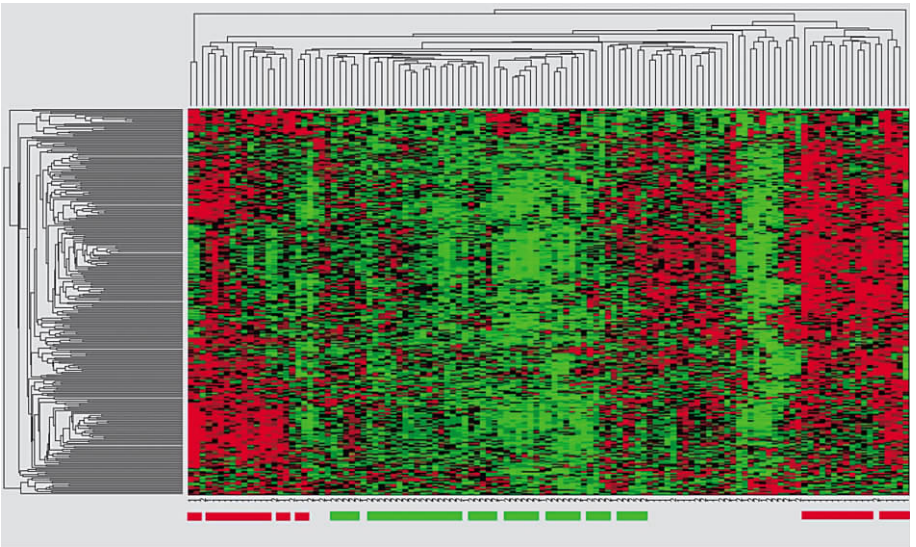

图 5 表达下调的差异表达基因聚类图

Fig.5 Dendrogram of differentially expressed genes that down-regulated expressed

3 讨论

大肠癌转移涉及到的通路很多[16]，决定大肠癌转移，组织分化程度以及患者生存时间的因素有很多[17]。对大肠癌转移来说，转移相关的分子标签集参与了多个信号通路的传递，研究发现在大肠癌 T1T2 期高表达的基因主要参与了细胞代谢与增殖相关的信号通路，而在 M1 期中高表达的基因主要参与与细胞粘附，细胞外基质受体信号通路[18]。因此判断大肠癌是否发生转移也许能够通过差异表达的基因集进行区分。本文对大肠癌转移差异表达基因进行显著性分析，并筛选出了一批起着协同作用的差异表达基因，这些差异表达基因从聚类分析和主成分分析都能得出其归为同一类，是大肠癌转移的关键基因集。进行筛选的表达谱数据来源于各个不同的大肠癌患者的组织样本，因此，这些基因集对大肠癌转移的判断有着重要指导意义。

本文还依据大肠癌组织学分化程度以及患者生存时间进

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

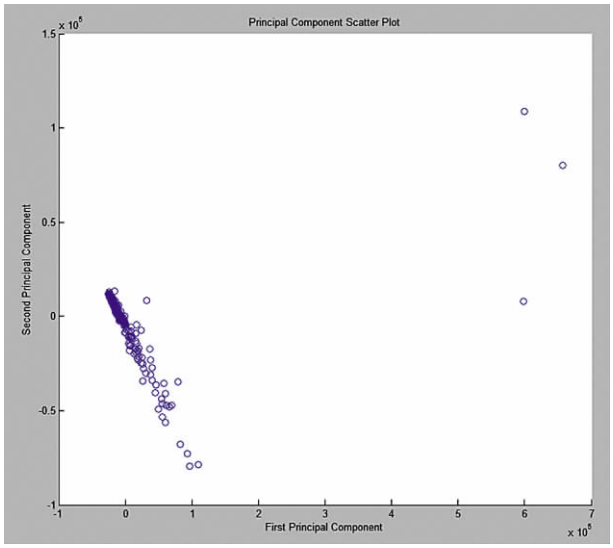

图 6 主成分分析图

Fig.6 The map of PCA

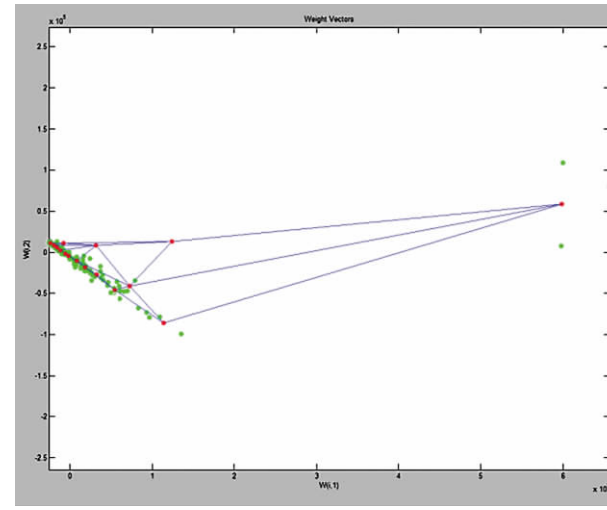

图 7 自组织映射图

Fig.7 The map of SOM

行分类，筛选出的差异基因可以看出分化程度低的大肠癌高表达着与合成趋化因子，细胞因子受体等相关的蛋白质，这些趋化因子和细胞因子受体能够促进大肠癌细胞的远隔器官转移，具有趋向性和游走性。而根据大肠癌患者生存时间的长短筛选出的高表达基因参与着细胞粘附，细胞外基质受体等分子功能，说明生存时间越短的病人，其肿瘤细胞越容易从原位脱落，变形游走，因此可以认为肿瘤细胞是否从原位脱落是大肠癌患者生存时间长短的重要因素。大肠癌的恶性程度高，容易通过淋巴道和血道转移，转移是影响大肠癌患者疗效，复发，死亡的重要原因，因此大肠癌的早期诊断和治疗对患者治愈或延长生存时间有着重要意义，本文所筛选的与大肠癌转移相关分子标签集，可对制作大肠癌转移检测基因芯片提供参考，对指导临床治疗以及分子诊断有着重要意义。

参 考 文 献(References)

[1] Yamauchi H, Togashi K, Kawamura Y J, et al. Pathological predictors for lymph node metastasis in T1 colorectal cancer [J]. Surg Today,2008,38(10):905-910

[2] Tsai M H, Wu C C, Peng P H, et al. Identification of secretory gelsolin as a plasma biomarker associated with distant organ metastasis of colorectal cancer[J]. J Mol Med (Berl),2012,90(2):187-200

[3] Tsutsumi S, Watanabe R, Tabe Y, et al. Extranodal metastasis predicts poor survival in advanced colorectal cancer[J]. Hepatogastroenterology,2012,59(114):372-374

[4] Makela J T, Kiviniemi H. Surgical treatment of colorectal cancer in patients aged over 80 years [J]. Int J Colorectal Dis,2012,27(8):1055-1060

[5] Chirica M, Leconte M, Oberlin O, et al. Surgical treatment of liver metastasis in patients with colorectal cancer [J]. Presse Med,2012,41(1):58-67

[6] Barbulescu M, Alecu L, Boeti P, et al. Port-site metastasis after laparoscopic surgery for colorectal cancer--still a real concern? Case report and review of the literature[J]. Chirurgia (Bucur),2012,107(1):103-107

[7] Lee J S, Leem S H, Lee S Y, et al. Expression signature of E2F1 and its associated genes predict superficial to invasive progression of blad- der tumors[J]. J Clin Oncol,2010,28(16):2660-2667

[8] Croner R S, Fortsch T, Bruckl W M, et al. Molecular signature for lymphatic metastasis in colorectal carcinomas[J]. Ann Surg,2008,247(5):803-810

[9] Joyce T, Cantarella D, Isella C, et al. A molecular signature for Epithelial to Mesenchymal transition in a human colon cancer cell system is revealed by large-scale microarray analysis [J]. Clin Exp Metastasis,2009,26(6):569-587

[10] Joyce T, Oikonomou E, Kosmidou V, et al. A molecular signature for oncogenic BRAF in human colon cancer cells is revealed by microarray analysis[J]. Curr Cancer Drug Targets,2012,Apr 19[Epub ahead of print]

[11] Carvalho L, Yu J, Schwartsmann G, et al. RNA expression of the molecular signature genes for metastasis in colorectal cancer [J]. Oncol Rep,2011,25(5):1321-1327

[12] Wilhite S E, Barrett T. Strategies to explore functional genomics data sets in NCBI's GEO database[J]. Methods Mol Biol,2012,802:41-53

[13] Lin D, Shkedy Z, Burzykowski T, et al. An investigation on performance of Significance Analysis of Microarray (SAM) for the comparisons of several treatments with one control in the presence of small-variance genes[J]. Biom J,2008,50(5):801-823

[14] Zhang L, Farrell J J, Zhou H, et al. Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer[J]. Gastroenterology,2010,138(3):949-957

[15] Prifti E, Zucker J D, Clement K, et al. FunNet: an integrative tool for exploring transcriptional interactions[J]. Bioinformatics,2008,24(22):2636-2638

[16] Simone C. Signal-dependent control of autophagy and cell death in colorectal cancer cell: the role of the p38 pathway[J]. Autophagy,2007,3(5):468-471

[17] Georges R B, Adwan H, Hamdi H, et al. The insulin-like growth factor binding proteins 3 and 7 are associated with colorectal cancer and liver metastasis[J]. Cancer Biol Ther,2011,12(1):69-79

[18] Takayama T, Miyanishi K, Hayashi T, et al. Colorectal cancer: genetics of development and metastasis [J]. J Gastroenterol,2006,41(3):185-192

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net