

Colon cancer recurrence-associated genes revealed by WGCNA co-expression network analysis

XIAOFENG ZHAI^{1,2*}, QINGFENG XUE^{3*}, QUN LIU^{1,2}, YUYU GUO^{1,2} and ZHE CHEN^{1,2}

¹Department of Integrative Oncology, Changhai Hospital of Traditional Chinese Medicine, Second Military Medical University; ²Department of Integrative Oncology, Changhai Hospital, Shanghai 200433; ³Department of Anesthesiology, 264 Hospital of The People's Liberation Army, Taiyuan, Shanxi 030001, P.R. China

Received March 31, 2016; Accepted March 20, 2017

DOI: 10.3892/mmr.2017.7412

Abstract. The present study aimed to identify the recurrence-associated genes in colon cancer, which may provide theoretical evidence for the development of novel methods to prevent tumor recurrence. Colon cancer and matched normal samples microarray data (E-GEOD-39582) were downloaded from ArrayExpress. Genes with significant variation were identified, followed by the screening of differentially expressed genes (DEGs). Subsequently, the co-expression network of DEGs was constructed using the weighted correlation network analysis (WGCNA) method, which was verified using the validation dataset. The significant modules associated with recurrence in the network were subsequently screened and verified in another independent dataset E-GEOD-33113. Function and pathway enrichment analyses were also conducted to determine the roles of selected genes. Survival analysis was performed to identify the association between these genes and survival. A total of 434 DEGs were identified in the colon samples, and stress-associated endoplasmic reticulum protein family member 2 (SERP2) and long non-coding RNA-0219 (LINC0219) were determined to be the vital DEGs between all the three sub-type groups with different clinical features. The brown module was identified to be the most significant module in the co-expression network associated with the recurrence of colon cancer, which was verified in the E-GEOD-33113 dataset. Top 10 genes in the brown module, including EGF containing fibulin like extracellular matrix protein 2 (EFEMP2), fibrillin 1 (FBN1) and secreted protein acidic and cysteine rich (SPARC) were also associated with survival time of colon cancer patients. Further analysis

revealed that the function of cell adhesion, biological adhesion, extracellular matrix (ECM) organization, pathways of ECM-receptor interaction and focal adhesion were the significantly changed terms in colon cancer. In conclusion, SERP2, EFEMP2, FBN1, SPARC, and LINC0219 were revealed to be the recurrence-associated molecular and prognostic indicators in colon cancer by WGCNA co-expression network analysis.

Introduction

Colon cancer is one of the most common malignant tumors, with a high incident rate in the 40-50 age group. Colon cancer affects ~150,000 patients in the USA annually (1). Due to the changing of diets, colon cancer has become the 4th cause for malignant tumor mortality in China and there are ~140,000 diagnosed cases annually (2). Surgery is the primary therapy for colon cancer and patients exhibit 5-year survival rate of 50% following surgery (3).

However, 15-20% patients experience recurrence following treatment. Tumor recurrence following curative surgery is a major hindrance for the improvement of overall survival (4). Therefore, it is important to identify the molecular changes in patients and to determine the underlying reason for colon cancer recurrence. Biomarkers have been used as tools in the detection and management of the disease in patients with colon cancer (5). For instance, the CpG island methylator phenotype is independently associated with an unfavorable prognosis in patients with colon cancer (6). Epithelial cell adhesion molecule, cluster of differentiation (CD)26, musashi RNA binding protein 1, CD29, CD24, leucine rich repeat containing G protein-coupled receptor 5 and aldehyde dehydrogenase 1 family member A1 have been identified as potential putative markers for colon cancer (7,8). DNA methylation may also predict recurrence of resected stage III proximal colon cancer (9). MicroRNA-93 inhibited the early relapse of colon cancer by targeting cell cycle-associated genes (10). Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit a mutation in colorectal cancer may act as a predictive molecular biomarker for adjuvant aspirin therapy in colon cancer (11).

An improved understanding of the biology of recurrence may improve the development of novel recurrence prevention or treatment methods in colon cancer. In order to investigate

Correspondence to: Dr Zhe Chen, Department of Integrative Oncology, Changhai Hospital, 168 Changhai Road, Shanghai 200433, P.R. China
E-mail: chendhdhd@163.com

*Contributed equally

Key words: colon cancer, co-expression, differentially expressed genes, recurrence, survival

the recurrence-associated genes in colon cancer for future therapy, a co-expression network of differentially expressed genes (DEGs) in colon cancer was constructed in the present study and the most significant modules in the network were used to reveal the recurrence-associated genes. Subsequently, the functions of recurrence-associated genes were enriched to determine the importance of these genes in the relapse of patients with colon cancer.

Materials and methods

Microarray profiles. Two microarray profiles of colon cancer samples including recurrence information (E-GEOD-39,582 and E-GEOD-33,113) were downloaded from ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>). E-GEOD-39,582 included 566 samples, based on the platform of AFFY HG-U133_Plus_2, which were divided into the training dataset and the validation dataset for the weighted correlation network analysis (WGCNA) network construction. There were a total of 90 colon cancer samples in E-GEOD-33,113, with the recurrence status and clinical information of the samples, which was used as the validation dataset for mining vital module associated with recurrence.

Sample classification. Samples in the profiles were classified using Consensus Clustering in R (<http://www.bioconductor.org/packages/release/bioc/html/ConsensusClusterPlus.html>) to identify their sub-types. The parameters set in this analysis are presented in Table I. Following the classification of samples, the association between colon cancer samples and corresponding clinical performance [survival, sex, cghdata, chemotherapy, mismatch repair (Mmr) status, tumor node metastasis (TNM) stage, and Tumor location] were determined using the χ^2 test. $P < 0.05$ was considered to indicate a statistically significant difference.

Pretreatment of genes expression data. Data in the profiles were initially normalized using the robust multi-array analysis (RMA) method in Affy package (<http://www.bioconductor.org/packages/release/bioc/html/affy.html>) and were transformed using a \log_2 transformation. Probes were converted into gene symbols and the average expression value was used as the only value of the gene with multiple corresponding probes. Following that, the cv method in the genefilter package was applied to filter the genes with significant variation. Genes with coefficient of variation > 0.4 were recognized as the candidate genes.

Screening of DEGs. The limma package in R (<http://www.bioconductor.org/packages/release/bioc/html/limma.html>) was used to identify the DEGs from the candidate genes. $P < 0.01$ and log fold change (FC) > 1.5 were set as the cut-off criteria.

Construction of the WGCNA co-expression network. The WGCNA method (12) was used to construct the co-expression network of the genes in the test samples of colon cancer. The interaction coefficient between genes was calculated using the following formula:

$$a_{ij} = S_{ij}^{\beta}, S_{ij} = |\text{cor}(x_i, x_j)| \quad (1)$$

Where x_i and x_j are the expression vectors of gene i and j , respectively. The Pearson coefficient of these two vectors was cor , which was transitioned into the interaction coefficient using S_{ij} . This transition intended to give more weight to the strong connections and reduce the importance of the weak connections in the predicted co-expression network in order to improve the reliability of the co-expression network.

The connection coefficient will be transformed into a weighted coefficient W_{ij} using the following formula:

$$W_{ij} = \frac{l_{ij} + a_{ij}}{\min\{k_i, k_j\} + 1 - a_{ij}} \quad (2)$$

$$l_{ij} = \sum_u a_{iu} a_{uj}, k_i = \sum_u a_{iu}$$

Subsequently, the co-expression network would be constructed based on the W matrix, followed by module mining. The reliability of the mined modules was verified using the verify sample set E-GEOD-33113. The topological properties were also confirmed. Modules with module significance < 0.05 were identified to be recurrence-associated modules.

Verification of recurrence-associated modules. The average expression value of genes in the significant modules in each sample was calculated. Subsequently, the samples were ranked based on the expression level of modules. According to 1/4 and 3/4 value of expression level, the modules were divided into high expression level, median expression level, and low expression level. Finally, the Kaplan-Meier (KM) curves of recurrence-associated modules expression level and recurrence status were drawn. The significance of the different expression level modules was compared, and $P < 0.05$ was considered to indicate a statistically significant difference. The identified vital module was verified using the independent dataset E-GEOD-33113.

Function enrichment of recurrence associated modules. In order to determine the functions of the genes in the recurrence-associated modules, genes were subjected to Database for Annotation, Visualization and Integrated Discovery (13) for function and pathway enrichment. $P < 0.05$ was the threshold used for the significant terms.

Survival analysis of genes in recurrence-associated modules. Survival analysis was performed on the genes in the recurrence-associated modules. The degree (k) of a gene and the significance (P) between each gene and sample survival time were also calculated using a Cox regression. Subsequently, the interaction coefficient (coef) between k and $-\log_{10}(P)$ was computed to identify the hub genes associated with survival. Finally, selected genes were determined in the verified samples for an association with survival. Recurrence associated genes were also subject to sample classification.

Results

Data preprocessing and sample classification. From the 19,846 irredundant genes in E-GEOD-39582, 6,600 were

Table I. Parameters used in sample classification using consensus clustering.

Parameter	Value
Max cluster group	20
Subsample number	5,000
Proportion sample	0.9
Proportion feature	1
Distance method	Pearson
Clustering method	K-means dist

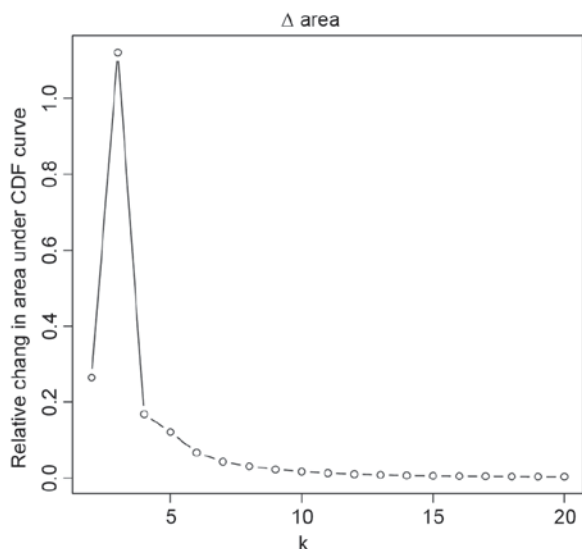


Figure 1. Correlation between the number of subgroups and CDF values. CDF, cumulative distribution function.

the variation genes. Consensus Clustering analysis revealed that cumulative distribution function (CDF) was at a high level when there were 3 sub-types, accompanied with a satisfied classification effect (Fig. 1). Therefore, the 556 samples in E-GEOD-39582 were divided into 3 subgroups: G1, G2, and G3 (Fig. 2). There were significant survival differences among these 3 types of samples, and as depicted in Fig. 3, G1 exhibited the highest survival status, whereas G3 exhibited the lowest survival status. Additionally, the χ^2 test determined that all clinical data, with the exception of sex, were significantly different among these 3 groups (Table II).

DEGs among the 3 sub-groups. From the 434 DEGs, 76 were the DEGs between G1 and G2, 390 were the DEGs between G1 and G3 and 63 were DEGs between G2 and G3. A total of 2 DEGs were identified in all 3 groups (Fig. 4). These two common DEGs were stress-associated endoplasmic reticulum protein family member 2 (SERP2) and long non-coding RNA-0219 (LINC0219).

Co-expression network construction and module mining. The co-expression network of the training dataset was divided into 4 modules (Fig. 5), which were verified with the validation dataset. By calculating the correlation coefficient, the connections between the genes in each module and survival status of

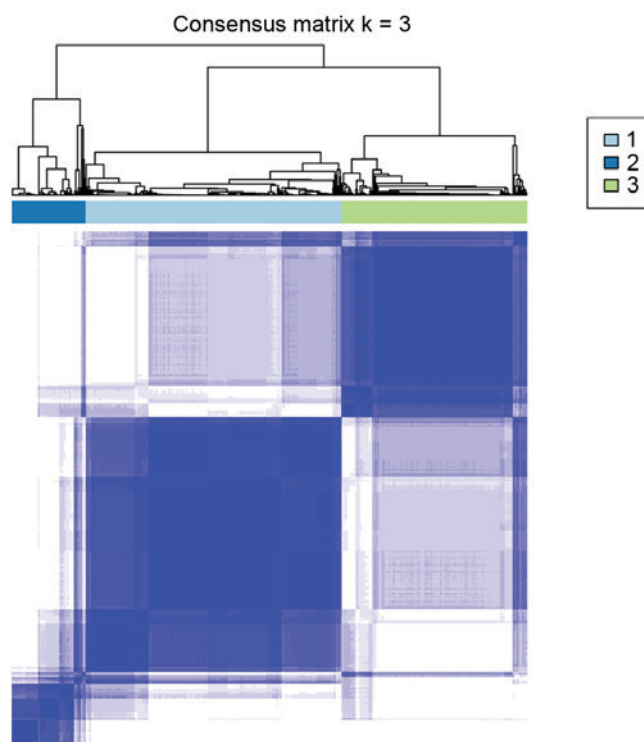


Figure 2. The three subgroups of the samples in E-GEOD-39582.

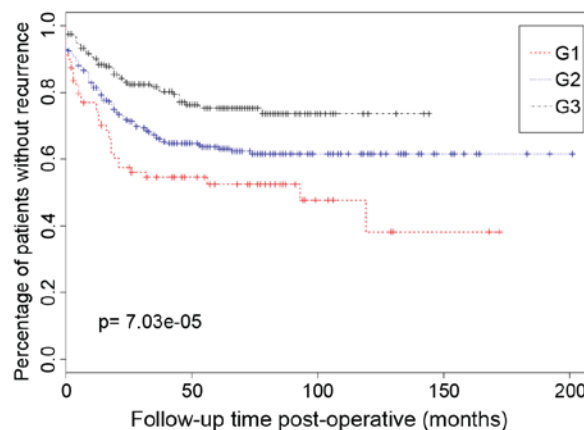


Figure 3. Recurrence analysis of the three subgroups.

the samples were identified. It was determined that the Brown module had the highest module significance and there were various survival-associated genes (hub genes) in this module (Fig. 6). These connections were also verified in the validation dataset.

Recurrence-associated module analysis. All 431 genes in the Brown module were subject to a clustering analysis and it was revealed that there were 3 types of samples (Fig. 7). As presented in Fig. 7, certain G2 type samples were also observed in with G1 and G2 type samples, suggesting it may have a medium role for the connection between the G1 and G3 status. Additionally, the recurrence interval of G3 samples was short, accompanied with downregulated genes.

The top 10 genes were collagen type VI α 3 chain (COL6A3), EGF containing fibulin like extracellular matrix

Table II. Association analysis on clinical data with sub-samples groups.

Characteristic	χ^2	df	P-value
Sex	1.1653	2	5.58×10^{-1}
CGH data	17.0540	2	1.98×10^{-4}
Chemotherapy	19.7370	4	5.63×10^{-4}
MMR status	23.4140	4	1.05×10^{-4}
TNM stage	26.2990	8	9.34×10^{-4}
Tumor location	14.7690	2	6.21×10^{-4}

MMR, mismatch repair; TNM, tumor node metastasis; CGH, comparative genomic hybridization; df, degrees of freedom.

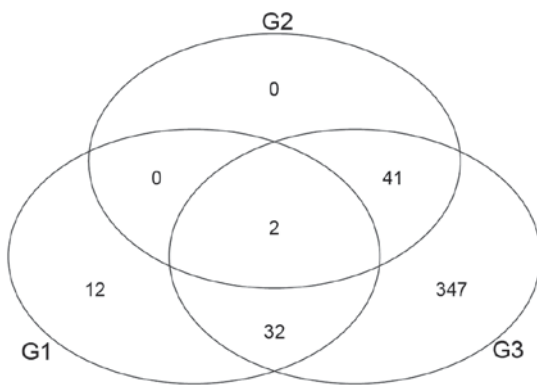


Figure 4. Differentially expressed genes in the 3 subgroups of G1, G2, and G3.

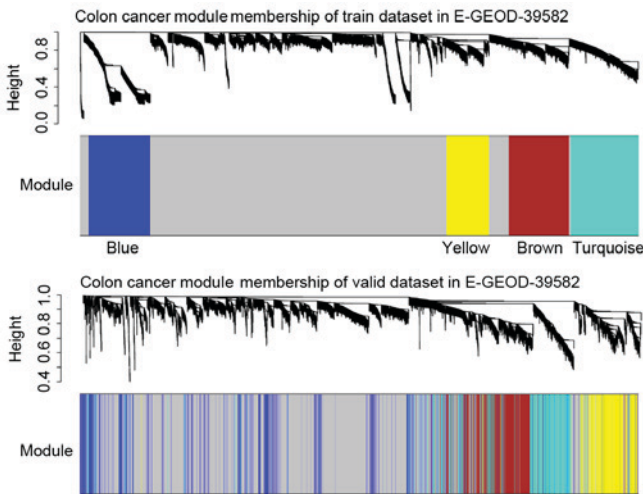


Figure 5. Clustering results of minded modules in train (upper) and valid (below) dataset. Genes in modules are marked with different colors (blue, yellow, brown and turquoise), with grey color representing no genes in any modules.

protein 2 (EFEMP2), fibrillin 1 (FBN1), follistatin-like 1 (FSTL1), glycosyltransferase 8 domain containing 2 (GLT8D2), heart development protein with EGF like domains 1 (HEG1), RAB31, member RAS oncogene family (RAB31), secreted protein acidic and cysteine rich (SPARC), SPARC/osteonectin, cwcv and kazal like domains proteoglycan 1 (SPOCK1) and

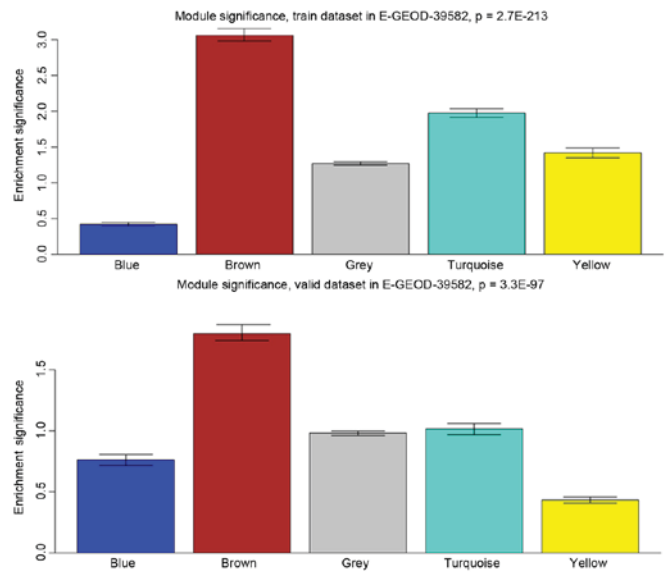


Figure 6. Distribution of survival related genes in all modules. Genes are presented in x-axis, and the enrichment significance is shown in y-axis.

TIMP metalloproteinase inhibitor 2 (TIMP2) in the Brown module with higher degrees were associated with survival (Table III). Genes in the Brown module were primarily enriched in tumor recurrence-associated functions and pathways, including cell adhesion, biological adhesion, ECM organization, the ECM-receptor interaction and the focal adhesion pathways (Table IV).

Validation of recurrence related modules. As presented in the KM curves (Fig. 8), there was a significant difference in terms of recurrence status ($P=1.5 \times 10^{-6}$) among the samples with high, median, and low gene expression level of the Brown module. High expression levels indicated higher incidence of recurrence, whereas low expression levels show lower recurrence incidence. This association between gene expression level and recurrence incidence was validated in the dataset of E-GEOD-33113 ($P=2.8 \times 10^{-2}$).

Discussion

Classification of the colon samples was downloaded from ArrayExpress revealed that there were significant differences of survival status and clinical data among the 3 subtype samples. From the 434 DEGs, SERP2 and LINC0129 were the common DEGs of the 3 subgroups, suggesting they may have an important role in the recurrence of colon cancer. The Brown module was the recurrence-associated module in the co-expression network of DEGs and the top 10 genes (COL6A3, EFEMP2, FBN1, FSTL1, GLT8D2, HEG1, RAB31, SPARC, SPOCK1, and TIMP2) in this module with higher degrees were demonstrated to be significantly associated with survival. Enrichment analysis revealed that genes in the Brown module were primarily enriched in tumor recurrence-associated functions and pathways, including cell and biological adhesion, ECM organization, the ECM-receptor interaction, and the focal adhesion pathways. Additionally, the association between the module and tumor recurrence were verified in another dataset E-GEOD-33113.

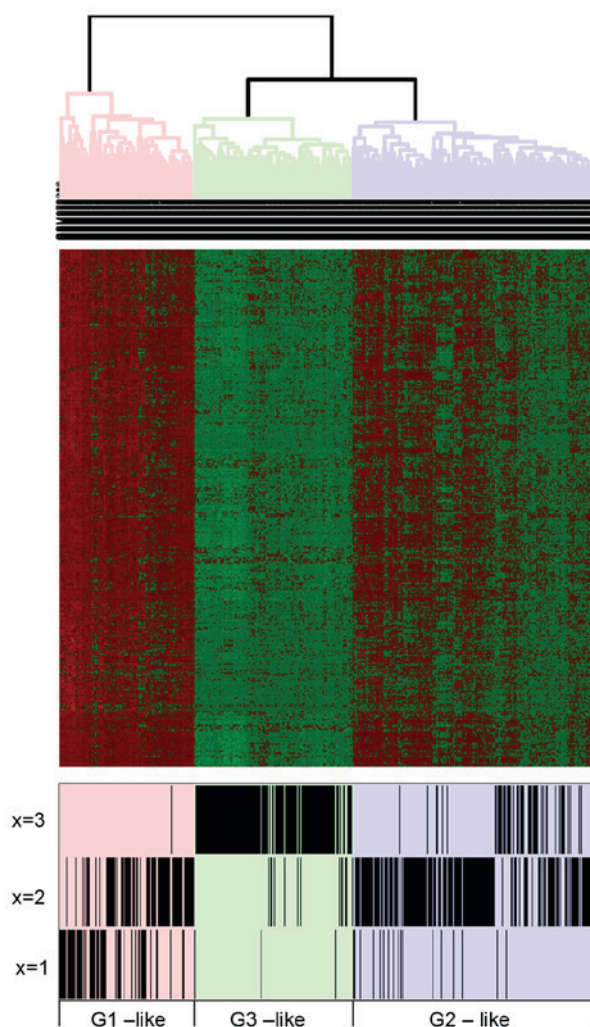


Figure 7. Clustering results of genes in the brown module. Pink, green, light green, and light blue represent G1-, G2-, and G3-type samples, respectively.

SERP2 belongs to the serine proteinase inhibitor family, which are key regulators for the biological pathways that initiate coagulation, inflammation, angiogenesis, apoptosis, complement activation response and ECM composition (14). SERP2 methylation was identified to be a marker for the detection and diagnosis of colon cancer (15). Expression of SERP2 is reported to be an early event in colon cancer, and is associated with carcinogenesis and its development (16). LINC0129 is a long non-coding RNA (lincRNA) and lincRNAs are RNAs >200 nt, which are not translated into proteins. Dysfunctions of lincRNAs have been associated with cancer. A previous study revealed that the downregulation of lincRNA BRAF-activated non-protein coding RNA may promote the proliferation of colorectal cancer cells (17). Another previous study reported that lincRNA HOX transcript antisense RNA expression may be a poor prognosis indicator in colon cancer (18). Additionally, overexpression of lincRNA prostate cancer associated transcript 1 was identified to be a novel biomarker of poor prognosis in patients with colon cancer (19). However, there are currently no direct findings that have determined the connection between LINC0129 and colon cancer, all aforementioned findings may have suggested that it may have an important role in colorectal cancer by contributing to the process of relapse.

Table III. Top 10 genes with high degrees in Brown module.

Gene name	Coefficient	P-value	k total	k within
<i>COL6A3</i>	2.17×10^{-4}	1.9×10^{-6}	136.82	91.86
<i>EFEMP2</i>	2.10×10^{-3}	9.2×10^{-7}	172.24	110.00
<i>FBN1</i>	1.42×10^{-3}	5.7×10^{-7}	137.03	92.14
<i>FSTL1</i>	1.50×10^{-3}	5.4×10^{-7}	143.42	91.08
<i>GLT8D2</i>	2.91×10^{-3}	7.5×10^{-7}	137.54	91.21
<i>HEG1</i>	2.17×10^{-3}	2.3×10^{-5}	168.71	94.84
<i>RAB31</i>	5.27×10^{-4}	5.9×10^{-5}	138.09	93.74
<i>SPARC</i>	3.06×10^{-4}	9.8×10^{-8}	132.03	95.57
<i>SPOCK1</i>	1.49×10^{-3}	9.7×10^{-7}	128.75	88.57
<i>TIMP2</i>	5.34×10^{-4}	4.2×10^{-7}	166.30	105.73

EFEMP2 is a serum biomarker for the early detection of colon cancer (20) and a superior biomarker compared with carcinoembryonic-antigen, which is the sole biomarker currently used for the diagnosis and treatment monitor in colon cancer (21). FBN1 is a component of the extracellular microfibril and the hypermethylation status of its promoter is a

Table IV. Significantly enriched functional terms of genes in Brown module.

Term ID	Function	Count	P-value
GO:0007155	Cell adhesion	84	7.29x10 ⁻³⁶
GO:0022610	Biological adhesion	84	8.12x10 ⁻³⁶
GO:0030198	Extracellular matrix organization	30	2.31x10 ⁻²³
GO:0043062	Extracellular structure organization	32	1.46x10 ⁻¹⁹
GO:0001501	Skeletal system development	40	2.50x10 ⁻¹⁷
GO:0001944	Vasculature development	35	1.35x10 ⁻¹⁶
GO:0001568	Blood vessel development	34	5.71x10 ⁻¹⁶
GO:0009611	Response to wounding	45	3.90x10 ⁻¹³
GO:0006928	Cell motion	40	1.36x10 ⁻¹¹
GO:0042060	Wound healing	25	2.72x10 ⁻¹¹
GO:0016477	Cell migration	28	5.08x10 ⁻¹⁰
GO:0048870	Cell motility	29	1.19x10 ⁻⁹
GO:0051674	Localization of cell	29	1.19x10 ⁻⁹
GO:0048514	Blood vessel morphogenesis	24	1.25x10 ⁻⁹
GO:0051270	Regulation of cell motion	22	6.93x10 ⁻⁹
GO:0035295	Tube development	23	1.44x10 ⁻⁸
GO:0030334	Regulation of cell migration	20	2.15x10 ⁻⁸
GO:0040012	Regulation of locomotion	21	3.35x10 ⁻⁸
GO:0042127	Regulation of cell proliferation	45	1.12x10 ⁻⁷
GO:0016337	Cell-cell adhesion	22	2.97x10 ⁻⁶
GO:0007167	Enzyme linked receptor protein signaling pathway	23	2.49x10 ⁻⁵
GO:0010033	Response to organic substance	36	5.15x10 ⁻⁵
GO:0008285	Negative regulation of cell proliferation	23	5.67x10 ⁻⁵
GO:0006954	Inflammatory response	21	1.07x10 ⁻⁴
GO:0000902	Cell morphogenesis	22	1.32x10 ⁻⁴
GO:0032989	Cellular component morphogenesis	23	2.24x10 ⁻⁴
GO:0008284	Positive regulation of cell proliferation	23	3.97x10 ⁻⁴
GO:0030182	Neuron differentiation	23	8.49x10 ⁻⁴
hsa04512	ECM-receptor interaction	26	1.80x10 ⁻²¹
hsa04510	Focal adhesion	35	1.85x10 ⁻²⁰
hsa05200	Pathways in cancer	22	3.45x10 ⁻⁵
hsa04060	Cytokine-cytokine receptor interaction	15	4.34x10 ⁻³
hsa04810	Regulation of actin cytoskeleton	13	5.83x10 ⁻³

GO, gene ontology.

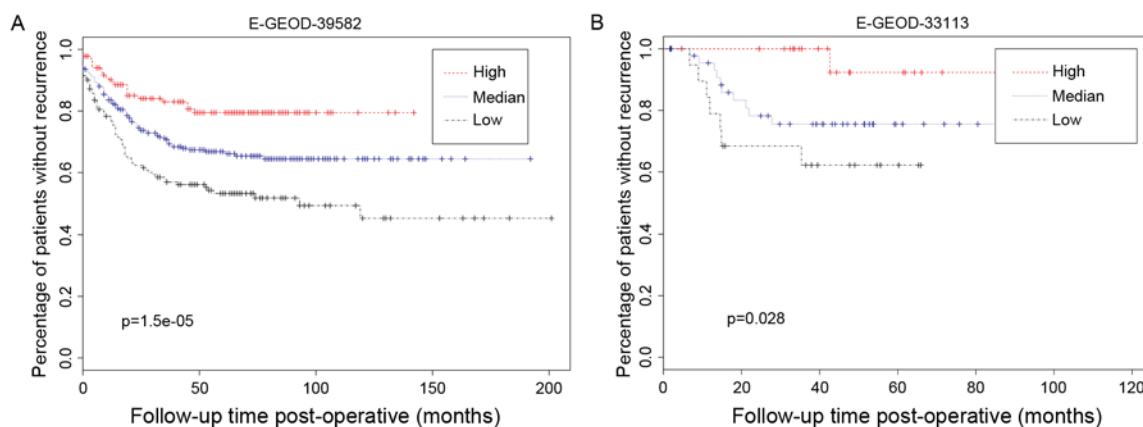


Figure 8. Correlation analysis between the genes and recurrence time. Kaplan-Meier curves for the brown module in (A) E-GEOD-39582 and (B) E-GEOD-33113 datasets.

specific and sensitive biomarker for colon cancer (22). SPARC is a matricellular protein involved in cell migration, angiogenesis and tissue remodeling. High SPARC expression may be associated with an improved clinical outcome in stage II colon cancer (23). A previous study determined that the absence of stromal SPARC is an independent prognostic predictor for poor prognosis of colon cancer (24). The high degrees of these genes in the recurrence-associated modules indicated that they have important roles in colon cancer relapse. Among the significantly enriched pathways, the ECM-receptor interaction and focal adhesion pathways were functionally clearly associated with the progression and prognosis of colon cancer (25).

By constructing the co-expression network of genes and identifying the recurrence related modules in the network, the present study identified several survival and recurrence-associated genes in colon cancer. These genes, including SERP2, EFEMP2, FBNI, SPARC, and LINC0219 were identified as recurrence-associated molecular and prognosis indicators in colon cancer.

Acknowledgements

The present study was supported by the Shanghai Municipal Commission of Health and Family Planning (grant no. ZYSNXD-CC-ZDYJ032), the Fund from Shanghai Science and Technology Committee (grant no. 12401907800) and the Shanghai Science and Technology Committee (grant no 15401931700).

References

- American Cancer Society: Colorectal cancer facts & figures, 2011-2013, Atlanta, GA: American Cancer Society, 2011.
- Liang Z, Baige L, Shen G, Lina S, Lingxia L and Chengyan H: Study on the Serum Periostin Concentrations of colon cancer patients. *Chin J Lab Dia* 4: 42, 2015.
- Yangming G, Chunxiao W, Minlu Z, Minlu Z, Peng P, Kai GU, Pingping B, Zhezhou H, Yongmei X and Ying Z: Colorectal cancer survival analysis in major areas in Shanghai. *China Oncol* 25: 497-504, 2015.
- Gerger A, Zhang W, Yang D, Bohanes P, Ning Y, Winder T, LaBonte MJ, Wilson PM, Benhaim L, Paez D, *et al*: Common cancer stem cell gene variants predict colon cancer recurrence. *Clin Cancer Res* 17: 6934-6943, 2011.
- Duffy M, Lamerz R, Haglund C, Nicolini A, Kalousova M, Holubec L and Sturgeon C: Tumor markers in colorectal cancer, gastric cancer and gastrointestinal stromal cancers: European group on tumor markers 2014 guidelines update. *Int J Cancer* 134: 2513-2522, 2014.
- Juo YY, Johnston FM, Zhang DY, Juo HH, Wang H, Pappou EP, Yu T, Easwaran H, Baylin S, van Engeland M and Ahuja N: Prognostic value of CpG island methylator phenotype among colorectal cancer patients: A systematic review and meta-analysis. *Ann Oncol* 25: 2314-2327, 2014.
- Sanders MA and Majumdar AP: Colon cancer stem cells: Implications in carcinogenesis. *Front Biosci (Landmark Ed)* 16: 1651-1662, 2011.
- Vermeulen L, Todaro M, de Sousa Mello F, Sprick MR, Kemper K, Perez Alea M, Richel DJ, Stassi G and Medema JP: Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity. *Proc Natl Acad Sci USA* 105: 13427-13432, 2008.
- Ahn JB, Chung WB, Maeda O, Shin SJ, Kim HS, Chung HC, Kim NK and Issa JP: DNA methylation predicts recurrence from resected stage III proximal colon cancer. *Cancer* 117: 1847-1854, 2011.
- Yang IP, Tsai HL, Hou MF, Chen KC, Tsai PC, Huang SW, Chou WW, Wang JY and Juo SH: MicroRNA-93 inhibits tumor growth and early relapse of human colorectal cancer by affecting genes involved in the cell cycle. *Carcinogenesis* 33: 1522-1530, 2012.
- Liao X, Lochhead P, Nishihara R, Morikawa T, Kuchiba A, Yamauchi M, Imamura Y, Qian ZR, Baba Y, Shima K, *et al*: Aspirin use, tumor PIK3CA mutation, and colorectal-cancer survival. *N Engl J Med* 367: 1596-1606, 2012.
- Langfelder P and Horvath S: WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9: 559, 2008.
- Huang da W, Sherman BT and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44-57, 2009.
- Richardson J, Viswanathan K and Lucas A: Serpins, the vasculature, and viral therapeutics. *Front Biosci* 11: 1042-1056, 2006.
- Zhang X, Song YF, Lu HN, Wang DP, Zhang XS, Huang SL, Sun BL and Huang ZG: Combined detection of plasma GATA5 and SFRP2 methylation is a valid noninvasive biomarker for colorectal cancer and adenomas. *World J Gastroenterol* 21: 2629-2637, 2015.
- Ning L, Haixing J, Zhenyu H, Shanyu Q and Xin L: Expression of SFRP2, β -catenin and their roles in colorectal carcinoma. *J Dig Oncol (Electronic Version)*, 2008.
- Shi Y, Liu Y, Wang J, Jie D, Yun T, Li W, Yan L, Wang K and Feng J: Downregulated long noncoding RNA BANCR promotes the proliferation of colorectal cancer cells via downregulation of p21 expression. *PLoS One* 10: e0122679, 2015.
- Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, Tanaka F, Shibata K, Suzuki A, Komune S, *et al*: Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res* 71: 6320-6326, 2011.
- Ge X, Chen Y, Liao X, Liu D, Li F, Ruan H and Jia W: Overexpression of long noncoding RNA PCAT-1 is a novel biomarker of poor prognosis in patients with colorectal cancer. *Med Oncol* 30: 588, 2013.
- Yao L, Lao W, Zhang Y, Tang X, Hu X, He C, Hu X and Xu LX: Identification of EFEMP2 as a serum biomarker for the early detection of colorectal cancer with lectin affinity capture assisted secretome analysis of cultured fresh tissues. *J Proteome Res* 11: 3281-3294, 2012.
- McPherson RA and Pincus MR: Henry's Clinical Diagnosis and Management by Laboratory Methods. 22nd edition. Elsevier Saunders, Philadelphia, PA, 2011.
- Guo Q, Song Y, Zhang H, Wu X, Xia P and Dang C: Detection of hypermethylated fibrillin-1 in the stool samples of colorectal cancer patients. *Med Oncol* 30: 695, 2013.
- Chew A, Salama P, Robbshaw A, Klopcec B, Zeps N, Platell C and Lawrance IC: SPARC, FOXP3, CD8 and CD45 correlation with disease recurrence and long-term disease-free survival in colorectal cancer. *PLoS One* 6: e22047, 2011.
- Liang JF, Wang HK, Xiao H, Li N, Cheng CX, Zhao YZ, Ma YB, Gao JZ, Bai RB and Zheng HX: Relationship and prognostic significance of SPARC and VEGF protein expression in colon cancer. *J Exp Clin Cancer Res* 29: 71, 2010.
- Lascorz J, Chen B, Hemminki K and Försti A: Consensus pathways implicated in prognosis of colorectal cancer identified through systematic enrichment analysis of gene expression profiling studies. *PLoS One* 6: e18867, 2011.