Gene 604 (2017) 33–40

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

A 15-gene signature for prediction of colon cancer recurrence and prognosis based on SVM

Guangru Xu [1], Minghui Zhang [1], Hongxing Zhu *, Jinhua Xu *

Department of Oncology, People's Hospital of Pudong District, Shanghai University of Medicine & Health Sciences, ShangHai 201299, China

ARTICLE INFO

Article history:
Received 26 May 2016
Received in revised form 7 November 2016
Accepted 14 December 2016
Available online 18 December 2016

Keywords:
Colon cancer
Support vector machine
Recurrence
Prognosis

ABSTRACT

Objective: To screen the gene signature for distinguishing patients with high risks from those with low-risks for colon cancer recurrence and predicting their prognosis.
Methods: Five microarray datasets of colon cancer samples were collected from Gene Expression Omnibus database and one was obtained from The Cancer Genome Atlas (TCGA). After preprocessing, data in GSE17537 were analyzed using the Linear Models for Microarray data (LIMMA) method to identify the differentially expressed genes (DEGs). The DEGs further underwent PPI network-based neighborhood scoring and support vector machine (SVM) analyses to screen the feature genes associated with recurrence and prognosis, which were then validated by four datasets GSE38832, GSE17538, GSE28814 and TCGA using SVM and Cox regression analyses.
Results: A total of 1207 genes were identified as DEGs between recurrence and no-recurrence samples, including 726 downregulated and 481 upregulated genes. Using SVM analysis and five gene expression profile data confirmation, a 15-gene signature (HES5, ZNF417, GLRA2, OR8D2, HOXA7, FABP6, MUSK, HTR6, GRIP2, KLRK1, VEGFA, AKAP12, RHEB, NCRNA00152 and PMEPA1) were identified as a predictor of recurrence risk and prognosis for colon cancer patients.
Conclusion: Our identified 15-gene signature may be useful to classify colon cancer patients with different prognosis and some genes in this signature may represent new therapeutic targets.

1. Introduction

Colon cancer is one of the most frequently diagnosed cancers and the third most common cause of death by cancer, with an estimated 93,090 newly diagnosed cases and 93,090 deaths annually in 2015 in the USA (Siegel et al., 2015). Surgical resection remains the mainstay of treatment for colon cancer. However, recurrence usually occurs in patients with distant metastasis, resulting in the treatment failure and poor prognosis (Lee et al., 2014). Therefore, it is a hot issue to investigate the prognostic factors able to early identify patients with a high recurrence risk in order to schedule individualized treatment strategies.

At present, the clinicopathologic staging is the gold standard for assessment of the recurrence risk of colon cancer. However, clinical outcomes are quite different, even among patients at the same tumor stage, suggesting the conventional staging may be unable to precisely characterize the cancer prognosis. Thus, there is a strong need for exploration of new tools (such as molecular markers) to more accurately identify patients with low and high risks of recurrence.

Recently, several studies have investigated the tumor markers associated with colon cancer recurrence. For example, Peng et al. demonstrated the carcinoembryonic antigen in draining venous blood (d-CEA) may be a prognostic factor for colon cancer patients, with the sensitivity and specificity of 90% and 40% in the prediction of metastasis or local recurrence (Peng et al., 2015). Belt et al. proved that low expressions of p21 and cyclin D1, and high expressions of p53 and aurora kinase A (AURKA) were significantly associated with recurrence. Further multivariate analysis confirmed p21 and AURKA to be independently predictive of disease recurrence (Eric et al., 2012). In addition, several gene signatures have also been identified by using the gene expression profile data generated by high-throughput platforms. For example, Wang et al. identified a 23-gene signature using Affymetrix U133a GeneChip data from 74 patients and demonstrated that this gene signature had a good performance for prediction of recurrence in Dukes' B patients with an accuracy of 78% (Wang et al., 2004). Bandrés et al.

Abbreviations: TCGA, The Cancer Genome Atlas; LIMMA, Linear Models for Microarray data; DEGs, differentially expressed genes; SVM, support vector machine; d-CEA, draining venous blood; GEO, Gene Expression Omnibus; RMA, Robust Multichip Average; RFE, recursive feature elimination; ROC, receiver operation characteristic; 2D, two-dimensional; HES5, Hairy and enhancer of split 5; RHEB, Ras homolog enriched in brain; mTORC1, mammalian target of rapamycin complex 1; SRY, sex-determining region Y.

* Corresponding authors at: Department of Oncology, People's Hospital of Pudong District, Shanghai University of Medicine & Health Sciences, No. 490 South Chuanhuan Road, ShangHai 201299, China.
E-mail address: wuwuuwgdg@aliyun.com (J. Xu).
[1] Co-first authors.

http://dx.doi.org/10.1016/j.gene.2016.12.016
0378-1119/

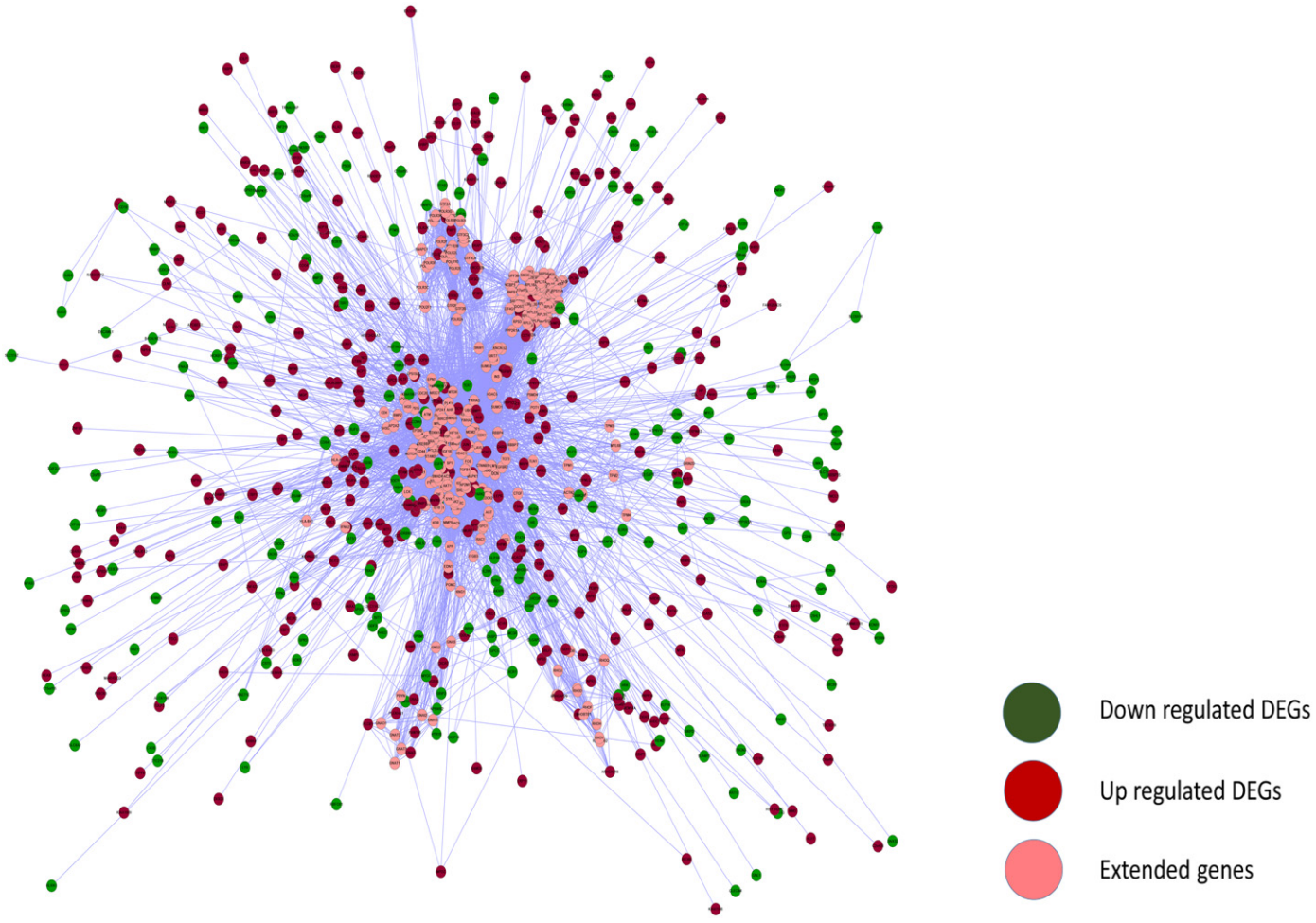

Fig. 1. PPI network construction.

retrospectively analyzed gene expression profiles in 16 tumor specimens from patients with Dukes' B colon cancer and screened a 8-gene signature to distinguish the risks of recurrence among patients (Bandrés et al., 2007). Smith et al. uncovered a 34-gene recurrence classifier in mouse models and confirmed them in colon cancer patients (Smith et al., 2010). However, the research on the genes that could predict the recurrence of colon cancer is still limited and needs further study.

The goal of this study was to further screen the gene signature for classification of high and low recurrence patients using a gene expression profile dataset and validated by four gene expression profile datasets based on the support vector machine (SVM) algorithm. The findings of our study may be more credible to be applied in clinic.

Table 1
The top 10 neighbor score.

Node	NS_score	LogFC	p-Value
EHBP1	0.96	1.015291	0.000397
EXOC6B	0.96	0.902543	0.00164
GRB10	0.92	0.948805	0.000933
AKAP12	0.91	0.97641	0.000658
SOX4	0.91	0.864663	0.002557
GLRA3	0.91	−0.83351	0.003639
GLRA2	0.91	−0.98549	0.000586
PPP1R3A	0.9	−1.04015	0.000285
FABP4	0.9	1.095293	0.000133
MED13L	0.9	0.710566	0.013178

NS, neighbor score; FC, fold change.

2. Materials and methods

2.1. Microarray data

To investigate the recurrence-related genes, the microarray data of colon cancer were collected from the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) database under the accession number GSE17537, in which 19 recurrence and 36 no-recurrence samples were included. In addition, other clinical characteristics, including age, sex, stage, death status, follow-up time, survival time, can also be available in this dataset to predict the prognosis value. Furthermore, to confirm the reliability of prognosis-related genes analyzed by GSE17537, four validation datasets were also obtained, consisting of three from GEO [GSE38832 (99 of 122 samples with survival time data), GSE17538 (68 recurrence and 109 no-recurrence samples), GSE28814 (126 samples with survival time data)] and one from The Cancer Genome Atlas (TCGA; http://tcga-data.nci.nih.gov/tcga/; 275 samples with survival time data).

2.2. Data normalization

The raw data (CEL files) downloaded were converted to expression values using the Robust Multichip Average (RMA) (Irizarry et al., 2003) implemented in the R statistical package (www.r-project.org). Preprocessed data were normalized by *Z*-score transformation, which adjusted the expression level of each gene according to the mean and standard deviation ($Z = (x - \mu) / \sigma$, where x is the raw expression value, μ is the mean and σ is the standard deviation). The final

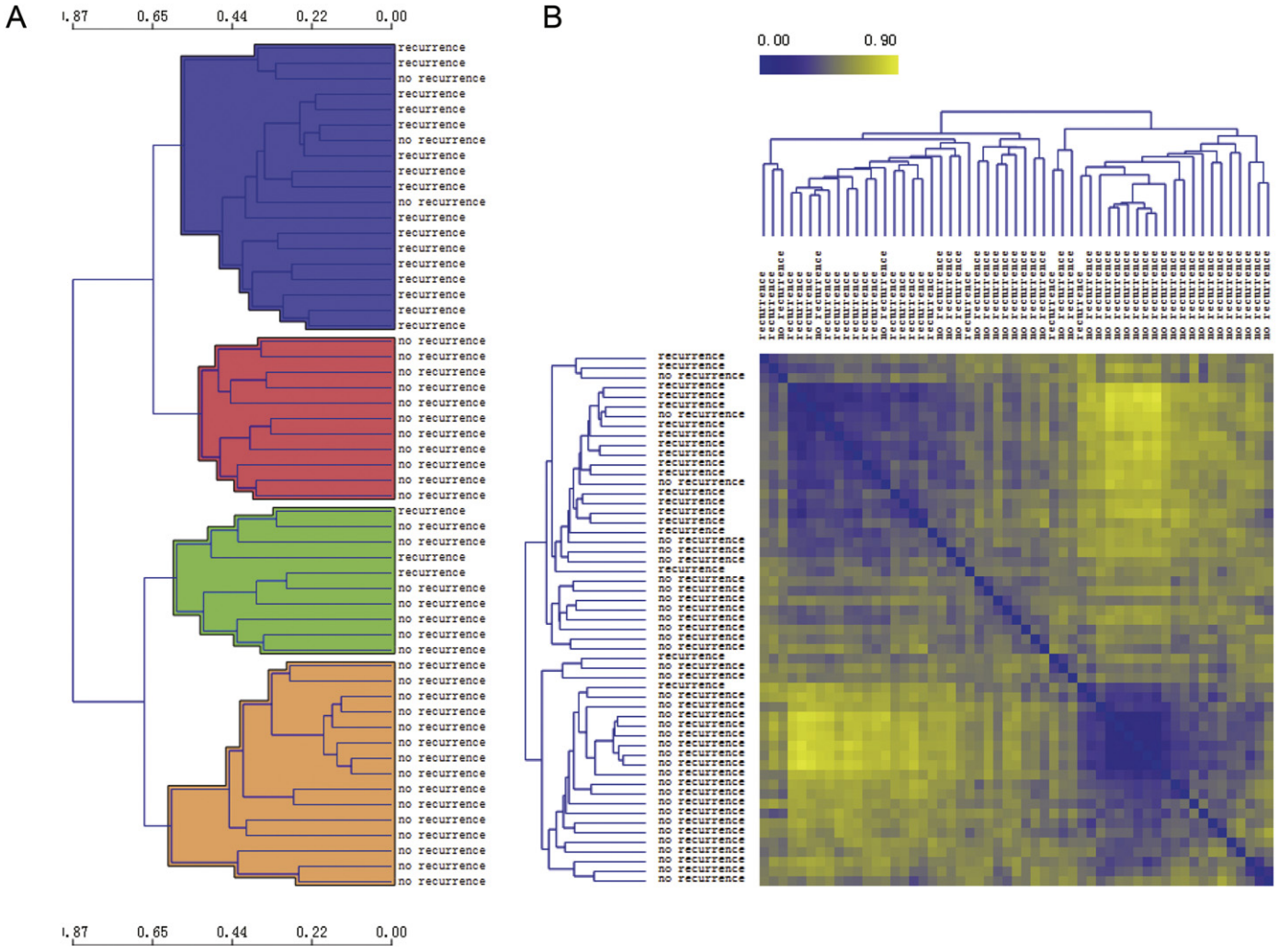

Fig. 2. Unsupervised hierarchical clustering analysis. A, clustering tree using the top 100 differentially expressed genes screened in GSE17537 dataset. The horizontal axis indicated the similarity and the vertical axis indicated the samples. Four samples clusters were obtained, displayed in yellow, green, red, and blue color. The yellow and red clusters could distinguish the non-recurrence sample with a 100% probability. The blue cluster could distinguish the recurrence sample with an 84.3% (16/19) probability; B, distance matrix. Both of the horizontal and vertical axes indicated samples. The similarity between samples was shown in heat map, with the higher similarity in blue and lower similarity in yellow.

expression level of genes was centered to mean zero and standard deviation of 1, with normal distribution.

2.3. Identification of differentially expressed genes (DEGs)

The analysis of DEGs between recurrence and no-recurrence samples was conducted using the Linear Models for Microarray data (LIMMA) method (Smyth, 2005) implemented in the R statistical package (www.r-project.org). The threshold for identification of DEGs was set as $p < 0.05$ and $|\log FC(\text{fold change})| > 0.7$.

2.4. Protein-protein interaction (PPI) network construction

To screen genes significantly correlated with recurrence, the DEGs were mapped into the PPI data downloaded from acknowledged HPRD (The Human Protein Reference Database; http://www.hprd.org) database (Peri et al., 2004). The protein that was not differentially expressed, was also included into the PPI network if it could interact with at least 20 DEGs encoded proteins. The PPI network was visualized using Cytoscape software (www.cytoscape.org/) (Kohl et al., 2011).

2.5. Network-based neighborhood scoring analysis

Neighborhood scoring is a local strategy for prioritizing candidates based on the expression of each differentially expressed gene (i) and its all direct neighboring genes [$N(i)$] in the network (Yang et al., 2014). Score 0 is assigned to genes that are neither differentially expressed or have any differentially expressed genes in their direct neighborhood. Genes were ranked by their scores that were calculated as follows:

$$\text{Score(i)} = \frac{1}{2} * FC(i) + \frac{1}{2} * \frac{\sum_{n \in N(i)} FC(n)}{N(i)}.$$

The top 100 genes were then selected as the feature genes according to the score.

2.6. Unsupervised hierarchical clustering analysis

An unsupervised hierarchical clustering analysis was used to investigate the effectiveness of feature genes in distinguishing recurrence and no-recurrence samples. Clustering was performed using Cluster software based on average linkage and Pearson correlation distance metric [6]. The clustering result was visualized by a heat map.

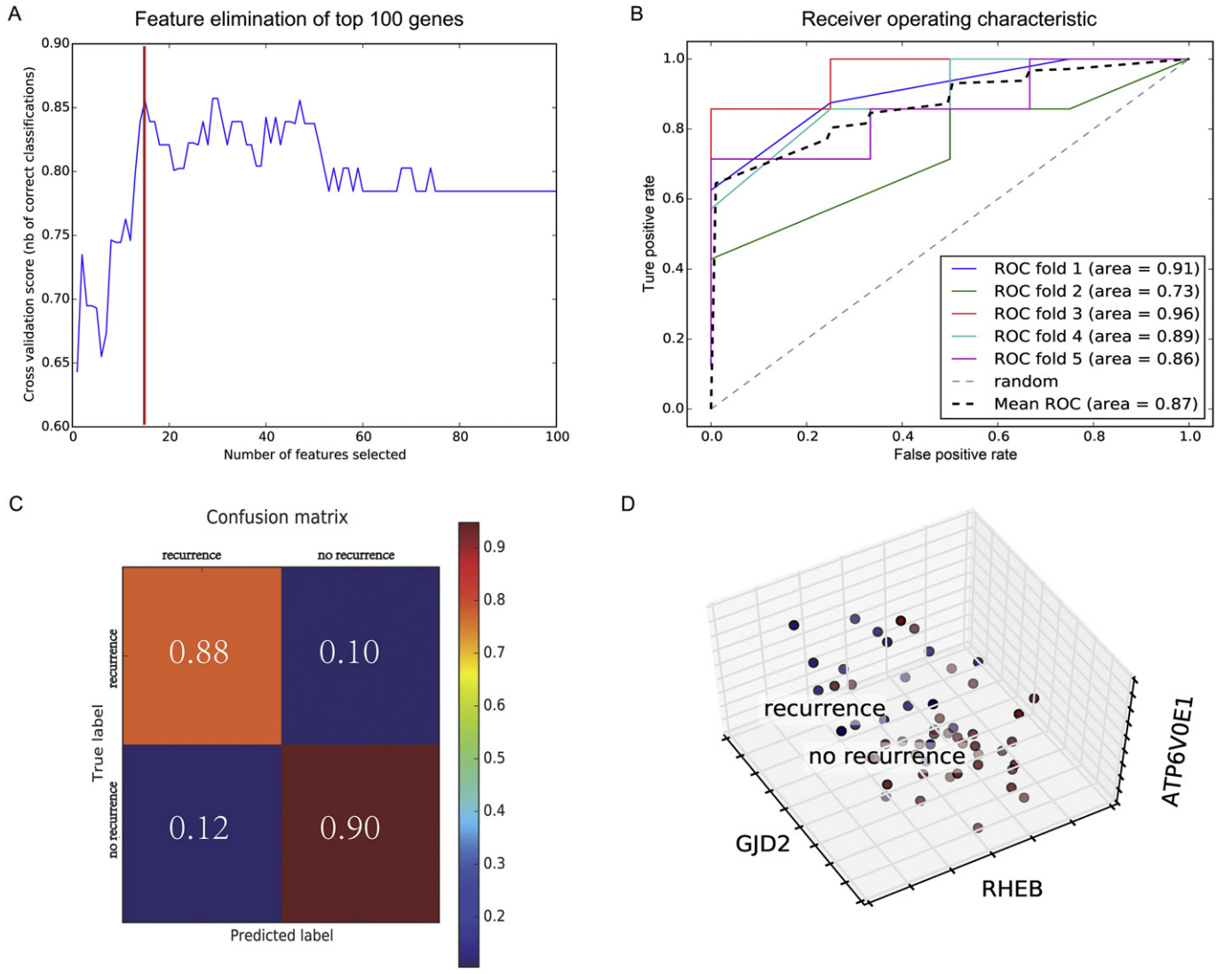

Fig. 3. Support vector machine analysis of top 100 differentially expressed genes. A, iterative feature elimination. The horizontal axis indicated the feature gene number and the vertical axis indicated the prediction accuracy; B, ROC curve using 15 genes as features. The 5-fold cross validation procedures were displayed with different colors, all which was finally fit into mean ROC. The horizontal axis indicated the false positive rate and the vertical axis indicated the true positive rate; C, confusion matrix using 15 genes. The horizontal axis indicated the predicted label and the vertical axis indicated the true label. The higher consistency between the predicted label and the true label was indicated to be in red, but lower consistency in blue.

2.7. Support vector machine (SVM) - recursive feature elimination analysis

To obtain the optimal feature genes to be applied in clinic for diagnosing recurrence, recursive feature elimination (RFE) algorithm, an iterative method based on the SVM technique, was performed (Guyon et al., 2002). The optimal gene subset was yielded by a leave-one-out cross-validation approach. To evaluate the prediction accuracy of the selected optimal feature genes combination for recurrence and no-recurrence samples, a supervised SVM classifier was constructed via SVM function in e1071 package of R (www.r-project.org) and trained with a 5-fold cross validation strategy followed by establishing receiver operation characteristic (ROC) curve. The area under the receiver operating curve (AUC) was estimated to indicate the prediction performance. To more detailedly confirm the prediction value of SVM classifier model, SVM with confusion matrix was constructed.

2.8. Validation analysis

To further confirm the classification reliability of the above selected optimal feature genes combination analyzed by GSE17537, SVM model was built using GSE17538 as a "testing" set and GSE17537 as a "training" set. Furthermore, cox regression analysis was performed based on GSE38832, GSE28814 and TCGA datasets via R survival package (Therneau, 2013) to evaluate the prognosis value of these feature genes.

3. Results

3.1. Identification of DEGs

Based on the threshold of $p < 0.05$ and $|\log FC| > 0.7$, 1207 genes were identified as DEGs between recurrence and no-recurrence samples of colon cancer, including 726 downregulated and 481 upregulated genes.

3.2. Network-based neighborhood scoring analysis

The 1207 DEGs were mapped into PPI data to construct a PPI network, consisting of 1085 nodes (protein) and 46,365 edges (interaction). As shown in Fig. 1, some DEGs exhibited an isolated characteristic, with only one or one more interaction relationships, but some DEGs were clustered, with several interaction relationships with other nodes. The DEGs that could influence multiple other nodes (that

Table 2
The selected 15 genes combination.

Gene	LogFC	p-Value
HES5	−0.89248	0.001848
ZNF417	−0.82599	0.003956
GLRA2	−0.98549	0.000586
OR8D2	−0.81352	0.004538
HOXA7	0.714953	0.012623
FABP6	0.923417	0.001275
MUSK	−0.79751	0.005399
HTR6	−0.76513	0.007601
GRIP2	−0.99734	0.000503
KLRK1	−0.9248	0.001254
VEGFA	0.84416	0.00323
AKAP12	0.97641	0.000658
RHEB	0.92876	0.001195
NCRNA00152	0.779584	0.017189
PMEPA1	0.780195	0.041686

FC, fold change; HES5, Hairy and enhancer of split 5; ZNF417, zinc finger protein 417; GLRA2, glycine receptor alpha 2; OR8D2, olfactory receptor family 8 subfamily D member 2; HOXA7, homeobox A7; FABP6, fatty acid binding protein 6; MUSK, muscle associated receptor tyrosine kinase; HTR6, 5-hydroxytryptamine receptor 6; GRIP2, glutamate receptor interacting protein 2; KLRK1, killer cell lectin like receptor K1; VEGFA, vascular endothelial growth factor A; AKAP12, A-kinase anchoring protein 12; RHEB, Ras homolog enriched in brain; NCRNA00152, CYTOR cytoskeleton regulator RNA; PMEPA1, prostate transmembrane protein, androgen induced 1.

is, high degree) may be more important to promote the recurrence of colon cancer. Thus, the top 100 DEGs (Table 1) were further selected by neighborhood scoring in consideration of the expression of each DEG and its all neighboring genes in the network.

To demonstrate the differential effectiveness of the 100 DEGs, an unsupervised hierarchical clustering analysis was performed. As shown in Fig. 2, the samples with a similar clinical manifestation tended to be clustered, indicating the excellent distinguishing capability of these 100 DEGs.

3.3. SVM analysis

To obtain the optimal feature genes for diagnosing recurrence, SVM-RFE algorithm was performed using genes as features and their expression levels as feature values. As a result, 85% prediction accuracy was found to be achieved using a 15-gene combination (Fig. 3A and Table 2). Subsequently, SVM analysis with a 5-fold cross validation procedure indicated an 87% prediction accuracy using this 15-gene combination (Fig. 3B). Furthermore, SVM with confusion matrix analysis also suggested an 88% prediction accuracy for recurrence samples and 90% prediction accuracy for no-recurrence samples (Fig. 3C). These findings all proved an outstanding prediction value of this 15-gene combination.

In addition to genes, the clinical features, including age, sex, stage, death status, follow-up time and survival time, may also exert important roles in prognosis. Thus, we also integrated the expression level of these 15 genes and the above clinical features as a feature set to investigate their prediction value. As anticipated, the prediction accuracy was improved to 92% from 87% (Fig. 4A), accompanied with an 90% prediction accuracy for recurrence samples and 94% prediction accuracy for no-recurrence samples (Fig. 4B).

3.4. Validation analysis

To further confirm the classification reliability of the above selected feature genes, another microarray dataset GSE17538 was collected and underwent the SVM model analysis. As expected, these 15 genes can provide a 77% prediction accuracy for recurrence samples and 84% prediction accuracy for no-recurrence samples in GSE17538 dataset (Table 3). In consideration of different intervention means (such as surgery or chemoradiotherapy) in 177 samples of GSE17538 dataset, the prediction accuracy may be lower than the expected value. To more comprehensively estimate the prognosis efficiency, we combined the survival analysis using three datasets (GSE38832, GSE28814 and TCGA). The results indicated that patients with different recurrent risks could exhibit a significantly different prognosis in two datasets (GSE38832, $p = 0.04$, Fig. 5A; GSE28814, $p = 0.0578$, Fig. 5B; and TCGA, $p = 0.0162$, Fig. 5C) using these 15 genes, demonstrating our classification model may be effective for clinical prognosis and treatment guidance.

4. Discussion

Our present study identifies a 15-gene signature (HES5, ZNF417, GLRA2, OR8D2, HOXA7, FABP6, MUSK, HTR6, GRIP2, KLRK1, VEGFA, AKAP12, RHEB, NCRNA00152 and PMEPA1) that can be used as a predictor of recurrence risk and prognosis for colon cancer patients. The genes in our signature seemed not to be the same with the findings of previous studies (Tang and Wei, 2004; Bandrés et al., 2007; Smith et al., 2010), but we believe our finding may be more credible to be applied in clinic because of two reasons: 1) this 15-gene signature was confirmed using five gene expression profile data in our study; 2) most of them (HES5, HOXA7, FABP6, KLRK1, AKAP12, VEGFA, RHEB and PMEPA1) have been experimentally demonstrated to be associated with cancer as following:

- HES5 (Hairy and enhancer of split 5) is a member of the basic helix–loop–helix (bHLH) superfamily of DNA-binding transcription factors. HES5 can be upregulated after the activation of the mammalian Notch pathway and then promotes proliferation by suppressing the transcription of its downstream gene Hash1, which ultimately induces the initiation and development of cervical carcinoma (Liu et al., 2010a). Zhu et al. also showed the expression of HES5 was significantly higher in hepatocellular carcinoma tissues and cell lines than that in normal control. The high expression of HES5 was significantly associated with histological grade, metastasis and poor prognosis. Knockdown of HES5 may inhibit cellular proliferation by increasing downstream Hash1 and repressing the activation of STAT3 (Zhu et al., 2015). These findings suggested the possible carcinogenic role of HES5 in colon cancer, however, rare studies were reported except Srinivasan et al. indicated that HES5 was upregulated in colon cancer initiating cells (Srinivasan et al., 2016). Unfortunately, an opposite result was obtained in our study ($LogFC = -0.89248$; $p = 0.001848$), indicating the dual function of HES5. Even, Massie et al. showed silencing of HES5 may be an early and recurrent change in prostate tumourigenesis (Massie et al., 2015).
- HOXA7 belongs to the HOX gene family that encodes homeodomain-containing transcription factors known to regulate many cellular processes, including cell proliferation, differentiation, and migration (Bhatlekar et al., 2014). Therefore, an aberrant expression of HOXA7 may be an important mechanism for the development of cancers, which have been demonstrated by several studies as following: Li et al. used siRNA and forced-expression to confirm the role of HOXA7 in hepatocellular carcinoma in vitro and in vivo. The results indicated overexpression of HOXA7 in hepatocellular carcinoma cells stimulated proliferation with the upregulation of cyclin E1 and CDK2 protein levels, but converse results were obtained after silencing of HOXA7 (Li et al., 2015). Zhang et al. also defined that knockdown of HOXA7 significantly decreased cell proliferation of breast cancer via inhibiting the expression of estrogen receptor-alpha (Zhang et al., 2013). Although previous studies have proved some of the HOX family genes exert crucial roles in colon cancer (Kanai et al., 2010; Alfredo, 2014), the studies of HOXA7 are fewly reported. In line with other cancers, we also demonstrated HOXA7 may be upregulated in colon cancer ($LogFC = 0.714953$; $p = 0.012623$).
- FABP6 (Fatty acid binding protein 6) is a protein that binds with bile acids in the ileal epithelium and participates in their metabolism to cholesterol. FABP6 expression is found to be upregulated by excess

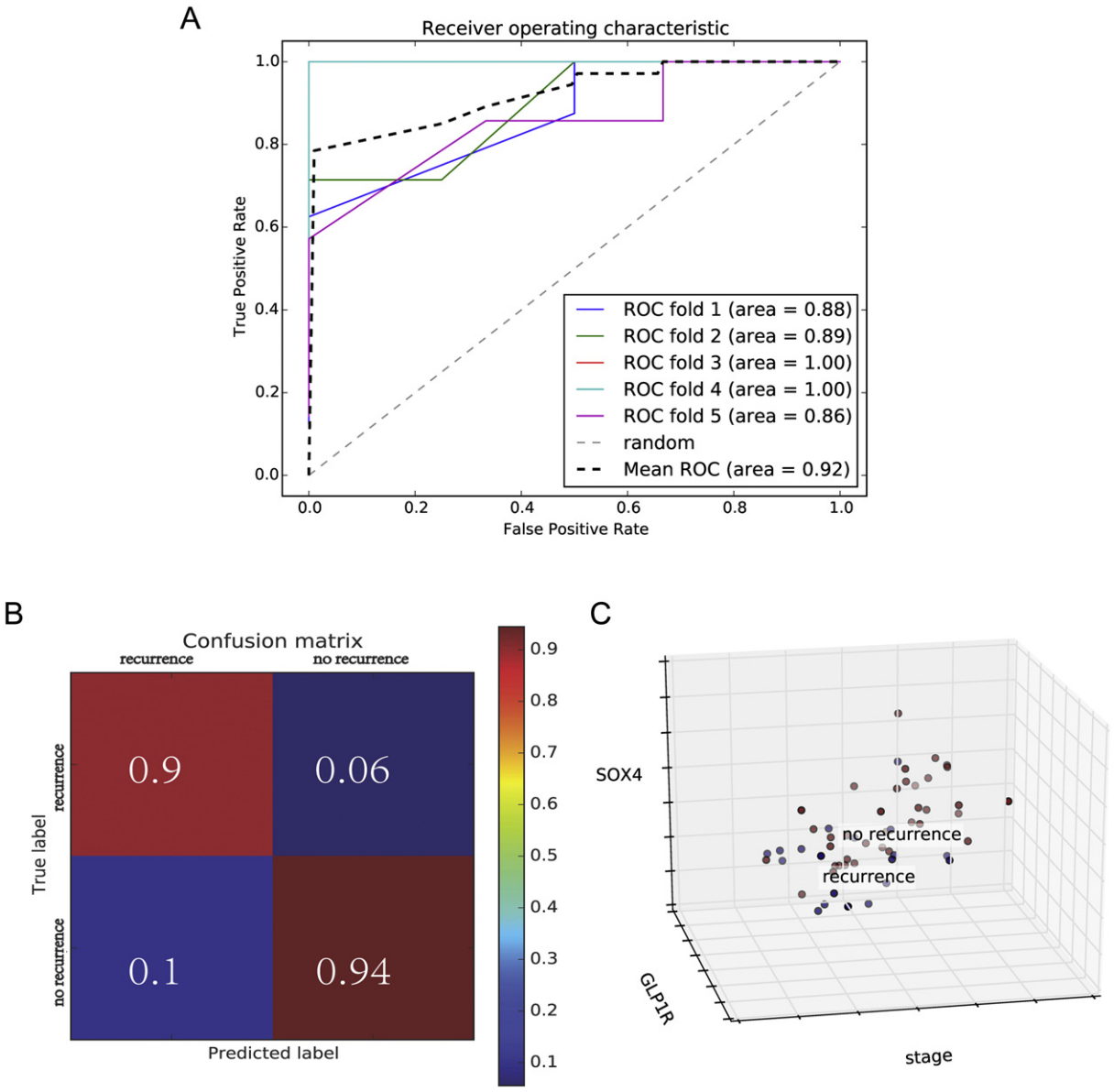

Fig. 4. Support vector machine analysis of 15 genes and clinical characteristics. A, ROC curve. The 5-fold cross validation procedures were displayed with different colors, all which was finally fit into mean ROC. The horizontal axis indicated the false positive rate and the vertical axis indicated the true positive rate; B, confusion matrix. The horizontal axis indicated the predicted label and the vertical axis indicated the true label. The higher consistency between the predicted label and the true label was indicated to be in red, but lower consistency in blue.

bile acids which infiltrate epithelial cells and cause DNA damage, ultimately inducing the development of colon cancer (Ohmachi et al., 2006). This conclusion was also preliminarily verified in our study (LogFC = 0.923417; p = 0.001275). However, highly expressed FABP6 was not demonstrated to be associated with metastasis and prognosis in colon cancer patients, which may be attributed to small sample size and further validation with multicenter samples is indispensable (Ohmachi et al., 2006).

- KLRK1 (killer cell lectin like receptor K1; also known as natural killer cell receptor G2D, NKG2D), is a stimulatory receptor expressed on natural killer (NK) and other cells active in the immune system (e.g. T cells and NKT cells), and thus plays important roles in induction of their immunosurveillance and cytotoxicity against tumors (López-Soto et al., 2015). The decrease in KLRK1 expression levels may lead to the suppression of activities of the above immune cells and initiates the development of colon cancer (Shen et al., 2011; Xu et al., 2011), which was also consentaneous in our study (LogFC = −0.9248; p = 0.001254). The lower expression of KLRK1 is shown to be significantly associated with lymph node metastasis, decreased disease-free survival and overall survival (Saito et al., 2011; Chen et al., 2015). Nevertheless, the relationships between KLRK1 expression and clinical characteristics in colon cancer remain under investigation, particular recurrence.
- VEGFA (vascular endothelial growth factor A) is an extensively studied gene that is involved in the development and progression of colon cancer by promoting angiogenesis (Mar et al., 2015). High expressed VEGF-A was associated with lymphatic metastases in colorectal cancer patients, and therefore may be a predictor for poor

Table 3
Validation analysis result of GSE17538 dataset.

	Precision	Recall	F1-score	Support
Recurrence	0.77	0.69	0.73	68
No recurrence	0.81	0.77	0.74	109
Total	0.79	0.74	0.74	177

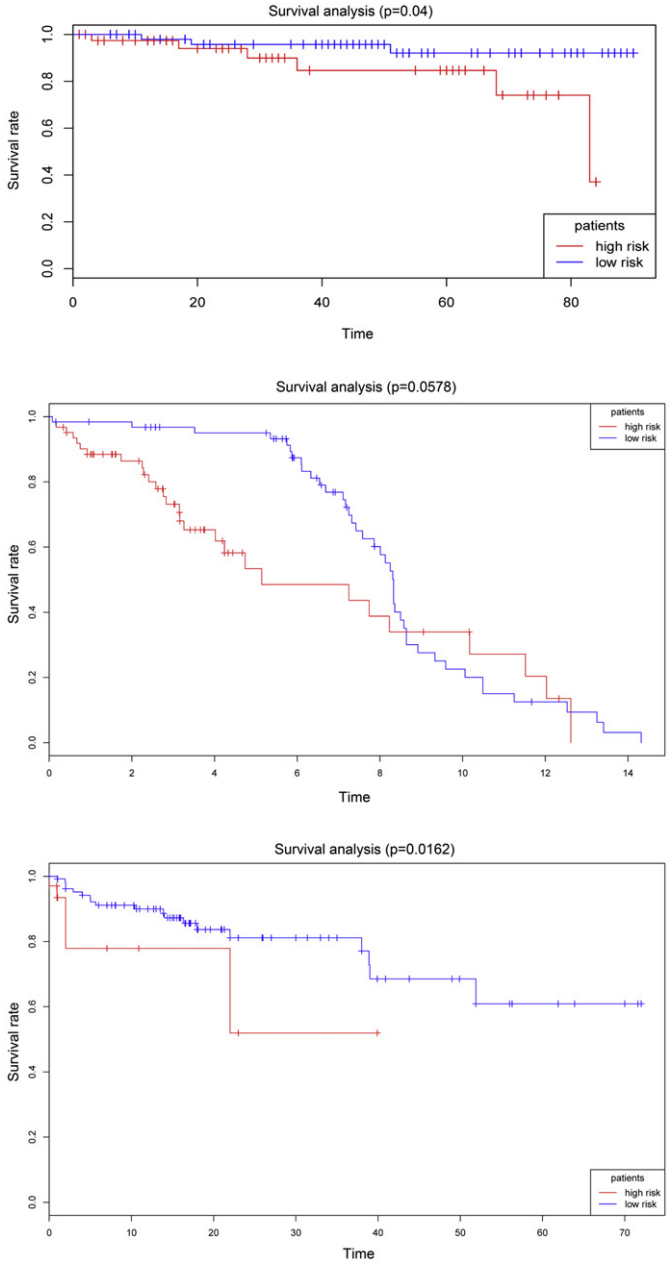

Fig. 5. Cox regression survival analysis using dataset GSE38832 (A), GSE28814 (B) and TCGA (C).

prognosis as reported in other cancers (George et al., 2001; Tanaka et al., 2010; Chen et al., 2014). As expected, VEGFA was also significantly upregulated in our study (LogFC = 0.84416; p = 0.00323).

- AKAP12 is a member of A-kinase-anchoring proteins (AKAPs) family that serves as a scaffold protein for the assembly of protein complexes (such as protein kinase A and C) to control oncogenic signaling pathways (Akakura and Gelman, 2011). Although AKAP12 is thought to be an important target for colon cancer, the expression of AKAP12 is not consistent in different literatures. Liu et al. found AKAP12 was downregulated in colon cancer, with the mechanism of promoter methylation (Liu et al., 2010b), but an elevated level was detected in the study of Yildirim et al. (2013). In line with the study of Yildirim et al. (2013), we also illustrated a significant upregulation of AKAP12 in our study (LogFC = 0.97641; p = 0.000658), suggesting a further confirmation is necessary.
- RHEB (Ras homolog enriched in brain) is a member of Ras family that functions as a direct upstream activator of mammalian target of rapamycin complex 1 (mTORC1), followed by stimulation of cell growth by promoting ribosomal biogenesis and protein synthesis through effectors p70S6K and eIF-4E-BP, finally leading to cancer progression (Kobayashi et al., 2010). Accumulated evidence suggested that RHEB was overexpressed in cancer and patients with high expression of RHEB possessed a reduced recurrence-free survival compared with patients with low expression of RHEB (Lu et al., 2010). Targeted inhibition of RHEB induced pronounced killing of cancer cells and improved the tumor cell sensitive to rapamycin (Schewe and Aguirre-Ghiso, 2008). Further study indicated the small GTPase RHEB may contribute to the development of colon cancer and/or resistance to mTOR inhibitors (such as rapamycin) by increasing autophagy in colon cancer cells. When blocking autophagy, the pro-survival effect of RHEB can be reversed (Campos et al., 2016). However, the mechanism of RHEB in colon cancer remains unclear and needs further investigation.
- PMEPA1 (prostate transmembrane protein, androgen induced 1) is originally identified as an androgen-induced gene and involved in the inhibition of prostate cancer (Xu et al., 2003). However, recent studies suggest PMEPA1 can act as a TGF-β signaling inducible gene and enhance tumorigenic activities in several cancers, including colon cancer (Nguyen et al., 2014; Koido et al., 2016; Nie, 2016). As anticipated, a significant upregulation of PMEPA1 was also observed in our study (LogFC = 0.780195; p = 0.041686).

However, further experiment studies using RT-PCR and/or western blot methods are needed to confirm the recurrence and prognosis classifier of our identified 15-gene signature based on the fact that: 1) a significantly differential prognosis was only proved in two of three datasets; 2) a 15-gene assay may be inconvenient and costly in clinical examination. Thus, a more simple gene combination may be essential to be screened by consideration of clinical results (Alberts et al., 2014; You et al., 2015); 3) as above descripted, the research in HES5, HOXA7, FABP6, KLRK1, AKAP12, VEGFA, RHEB and PMEPA1 genes in colon cancer and their relationships with the clinical characteristic of recurrence and prognosis remains rare. No studies have studies the role of ZNF417, GLRA2, OR8D2, MUSK, HTR6, GRIP2, and NCRNA00152 in cancer.

In conclusion, our findings suggest a 15-gene signature that may be useful as a predictor of recurrence risk and prognosis for colon cancer patients. Some of them that were not proved to be associated with colon cancer progression may represent new therapeutic targets.

References

Akakura, S., Gelman, I.H., 2011. Pivotal role of AKAP12 in the regulation of cellular adhesion dynamics: control of cytoskeletal architecture, cell migration, and mitogenic signaling. J. Signal Transduction 2012, 529179.

Alberts, S.R., Yu, T.M., Behrens, R.J., Renfro, L.A., Srivastava, G., Soori, G.S., Dakhil, S.R., Mowat, R.B., Kuebler, J.P., Kim, G.P., 2014. Comparative economics of a 12-gene assay for predicting risk of recurrence in stage II colon cancer. PharmacoEconomics 32, 1231–1243.

Alfredo, P., 2014. The paralogous group HOX 13 discriminates between normal colon tissue and colon cancer. J. Mol. Genet. Med. (2014).

Bandrés, E., Malumbres, R., Cubedo, E., Honorato, B., Zarate, R., Labarga, A., Gabisu, U., Sola, J.J., García-Foncillas, J., 2007. A gene signature of 8 genes could identify the risk of recurrence and progression in Dukes' B colon cancer patients. Oncol. Rep. 17, 1089–1094.

Bhatlekar, S., Fields, J.Z., Boman, B.M., 2014. HOX genes and their role in the development of human cancers. J. Mol. Med. 92, 811–823.

Campos, T., Ziehe, J., Palma, M., Escobar, D., Tapia, J.C., Pincheira, R., Castro, A.F., 2016. Rheb promotes cancer cell survival through p27Kip1-dependent activation of autophagy. Mol. Carcinog. 55, 220–229.

Chen, J., Xu, H., Zhu, X.X., 2015. Abnormal expression levels of sMICA and NKG2D are correlated with poor prognosis in pancreatic cancer. Ther. Clin. Risk Manag. 12, 11–18.

Chen, P., Zhu, J., Liu, D.Y., Li, H.Y., Xu, N., Hou, M., 2014. Over-expression of survivin and VEGF in small-cell lung cancer may predict the poorer prognosis. Med. Oncol. 31, 1–7.

Eric, J.T., Brosens, R.P., Delis-van Diemen, P.M., Bril, H., Tijssen, M., van Essen, D.F., Heymans, M.W., Beliën, J.A., Stockmann, H.B., Meijer, S., 2012. Cell cycle proteins predict recurrence in stage II and III colon cancer. Ann. Surg. Oncol. 19, 682–692.

George, M.L., Tutton, M.G., Janssen, F., Arnaout, A., Abulafi, A.M., Eccles, S.A., Swift, R.I., 2001. VEGF-A, VEGF-C, and VEGF-D in colorectal cancer progression. Neoplasia 3, 420–427.
Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46, 389–422.
Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., Speed, T.P., 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4, 249–264.
Kanai, M., Hamada, J.-I., Takada, M., Asano, T., Murakawa, K., Takahashi, Y., Murai, T., Tada, M., Miyamoto, M., Kondo, S., 2010. Aberrant expressions of HOX genes in colorectal and hepatocellular carcinomas. Oncol. Rep. 23, 843–851.
Kobayashi, T., Shimizu, Y., Terada, N., Yamasaki, T., Nakamura, E., Toda, Y., Nishiyama, H., Kamoto, T., Ogawa, O., Inoue, T., 2010. Regulation of androgen receptor transactivity and mTOR–S6 kinase pathway by Rheb in prostate cancer cell proliferation. Prostate 70, 866–874.
Kohl, M., Wiese, S., Warscheid, B., 2011. Cytoscape: software for visualization and analysis of biological networks. Data Mining in Proteomics: From Standards to Applications, pp. 291–303.
Koido, M., Sakurai, J., Tsukahara, S., Tani, Y., Tomida, A., 2016. PMEPA1, a TGF-β- and hypoxia-inducible gene that participates in hypoxic gene expression networks in solid tumors. Biochem. Biophys. Res. Commun.
López-Soto, A., Huergo-Zapico, L., Acebes-Huerta, A., Villa-Alvarez, M., Gonzalez, S., 2015. NKG2D signaling in cancer immunosurveillance. Int. J. Cancer 136, 1741–1750.
Lee, H., Choi, D.W., Cho, Y.B., Yun, S.H., Kim, H.C., Lee, W.Y., Heo, J.S., Choi, S.H., Jung, K.U., Chun, H.-K., 2014. Recurrence pattern depends on the location of colon cancer in the patients with synchronous colorectal liver metastasis. Ann. Surg. Oncol. 21, 1641–1646.
Li, Y., Yang, X.H., Fang, S.J., Qin, C.F., Sun, R.L., Liu, Z.Y., Jiang, B.Y., Wu, X., Li, G., 2015. HOXA7 stimulates human hepatocellular carcinoma proliferation through cyclin E1/CDK2. Oncol. Rep. 33, 990–996.
Liu, J., Lu, W.-G., Ye, F., Cheng, X.-d., Hong, D., Hu, Y., Chen, H.-z., Xie, X., 2010a. Hes1/Hes5 gene inhibits differentiation via down-regulating Hash1 and promotes proliferation in cervical carcinoma cells. Int. J. Gynecol. Cancer 20, 1109–1116.
Liu, W., Guan, M., Su, B., Ye, C., Li, J., Zhang, X., Liu, C., Li, M., Lin, Y., Yuan, L., 2010b. Quantitative assessment of AKAP12 promoter methylation in colorectal cancer using methylation-sensitive high resolution melting: correlation with Duke's stage. Cancer Biol. Ther. 9, 862–871.
Lu, Z.H., Shvartsman, M.B., Lee, A.Y., Shao, J.M., Murray, M.M., Kladney, R.D., Fan, D., Krajewski, S., Chiang, G.G., Mills, G.B., 2010. mTOR activator Rheb is frequently overexpressed in human carcinomas and is critical and sufficient for skin epithelial carcinogenesis. Cancer Res. 70, 3287.
Mar, A.C., Chu, C.H., Lee, H.J., Chien, C.W., Cheng, J.J., Yang, S.H., Jiang, J.K., Lee, T.C., 2015. Interleukin-1 receptor type 2 acts with c-Fos to enhance the expression of interleukin-6 and vascular endothelial growth factor a in colon cancer cells and induce angiogenesis. J. Biol. Chem. 290, 22212–22224.
Massie, C.E., Spiteri, I., Rossadams, H., Luxton, H., Kay, J., Whitaker, H.C., Dunning, M.J., Lamb, A.D., Ramosmontoya, A., Brewer, D.S., 2015. HES5 silencing is an early and recurrent change in prostate tumourigenesis. Endocrine related. Cancer 22, 131–144.
Nguyen, T.T.V., Watanabe, Y., Shiba, A., Noguchi, M., Itoh, S., Kato, M., 2014. TMEPAI/PMEPA1 enhances tumorigenic activities in lung cancer cells. Cancer Sci. 105, 334–341.
Nie, Z., 2016. Transforming growth factor-beta increases breast cancer stem cell population partially through upregulating PMEPA1 expression. Acta Biochim. Biophys. Sin. 48, 119–132.
Ohmachi, T., Inoue, H., Mimori, K., Tanaka, F., Sasaki, A., Kanda, T., Fujii, H., Yanaga, K., Mori, M., 2006. Fatty acid binding protein 6 is overexpressed in colorectal cancer. Clin. Cancer Res. 12, 5090–5095.
Peng, Y., Zhai, Z., Li, Z., Wang, L., Gu, J., 2015. Role of blood tumor markers in predicting metastasis and local recurrence after curative resection of colon cancer. Int. J. Clin. Exp. Med. 8, 982.
Peri, S., Navarro, J.D., Kristiansen, T.Z., Amanchy, R., Surendranath, V., Muthusamy, B., Gandhi, T., Chandrika, K., Deshpande, N., Suresh, S., 2004. Human protein reference database as a discovery resource for proteomics. Nucleic Acids Res. 32, D497–D501.
Saito, H., Osaki, T., Ikeguchi, M., 2011. Decreased NKG2D expression on NK cells correlates with impaired NK cell function in patients with gastric cancer. Gastric Cancer 15, 27–33.
Schewe, D.M., Aguirre-Ghiso, J.A., 2008. ATF6α-Rheb-mTOR signaling promotes survival of dormant tumor cells in vivo. Proc. Natl. Acad. Sci. 105, 10519–10524.
Shen, Y., Chao, L.U., Tian, W., Wang, L., Cui, B., Jiao, Y., Chunyan, M.A., Ying, J.U., Ling, Z., Shao, C., 2011. Possible association of decreased NKG2D expression levels and suppression of the activity of natural killer cells in patients with colorectal cancer. Int. J. Oncol. 40, 1285–1290.
Siegel, R.L., Miller, K.D., Jemal, A., 2015. Cancer statistics, 2015. CA Cancer J. Clin. 65, 5–29.
Smith, J.J., Deane, N.G., Fei, W.U., Merchant, N.B., Zhang, B., Jiang, A., Pengcheng, L.U., Johnson, J.C., Schmidt, C., Bailey, C.E., 2010. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. Gastroenterology 138, 958–968.
Smyth, G.K., 2005. Limma: Linear Models for Microarray Data, Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer, pp. 397–420.
Srinivasan, T., Walters, J., Bu, P., Than, E.B., Tung, K.L., Chen, K.Y., Panarelli, N., Milsom, J., Augenlicht, L., Lipkin, S.M., 2016. NOTCH signaling regulates asymmetric cell fate of fast- and slow-cycling colon cancer-initiating cells. Cancer Res. 76.
Tanaka, T., Ishiguro, H., Kuwabara, Y., Kimura, M., Mitsui, A., Katada, T., Shiozaki, M., Naganawa, Y., Fujii, Y., Takeyama, H., 2010. Vascular endothelial growth factor C (VEGF-C) in esophageal cancer correlates with lymph node metastasis and poor patient prognosis. J. Exp. Clin. Cancer Res. 29, 1–7.
Tang, C.H., Wei, Y., 2004. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 22, 1564–1571.
Therneau, T., 2013. A Package for Survival Analysis in S. R Package Version 2.37-4. URL. https://CRAN.R-project.org/package=survival. Box 980032: 23298-20032).
Wang, Y., Jatkoe, T., Zhang, Y., Mutch, M.G., Talantov, D., Jiang, J., McLeod, H.L., Atkins, D., 2004. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. J. Clin. Oncol. 22, 1564–1571.
Xu, L.L., Shi, Y., Petrovics, G., Sun, C., Makarem, M., Zhang, W., Sesterhenn, I.A., Mcleod, D.G., Sun, L., Moul, J.W., 2003. PMEPA1, an androgen-regulated NEDD4-binding protein, exhibits cell growth inhibitory function and decreased expression during prostate cancer progression. Cancer Res. 63, 4299–4304.
Xu, Y., Xu, Q., Ni, S., Liu, F., Cai, G., Wu, F., Ye, X., Meng, X., Mougin, B., Cai, S., 2011. Decrease in natural killer cell associated gene expression as a major characteristic of the immune status in the bloodstream of colorectal cancer patients. Cancer Biol. Ther. 11, 188–195.
Yang, J., Li, Z., Fan, X., Cheng, Y., 2014. A three step network based approach (TSNBA) to finding disease molecular signature and key regulators: a case study of IL-1 and TNF-alpha stimulated inflammation. PLoS One 9.
Yildirim, M., Suren, D., Yildiz, M., Sezgin, A.A., Eryilmaz, R., Goktas, S., Bulbuller, N., Sezer, C., 2013. AKAP12/Gravin gene expression in colorectal cancer: clinical importance and review of the literature. J. BUON 18, 635–640.
You, Y.N., Rustin, R.B., Sullivan, J.D., 2015. Oncotype DX colon cancer assay for prediction of recurrence risk in patients with stage II and III colon cancer: a review of the evidence. Surg. Oncol. 24, 61–66.
Zhang, Y., Cheng, J.-C., Huang, H.-F., Leung, P.C., 2013. Homeobox A7 stimulates breast cancer cell proliferation by up-regulating estrogen receptor-alpha. Biochem. Biophys. Res. Commun. 440, 652–657.
Zhu, G., Shi, W., Fan, H., Zhang, X., Xu, J., Chen, Y., Xu, Z., Tao, T., Cheng, C., 2015. HES5 promotes cell proliferation and invasion through activation of STAT3 and predicts poor survival in hepatocellular carcinoma. Exp. Mol. Pathol. 99, 474–484.