SPANDIDOS PUBLICATIONS

ONCOLOGY REPORTS 28: 1036-1042, 2012

Identification of candidate colon cancer biomarkers by applying a random forest approach on microarray data

ZHI YAN[1*], JIANGENG LI[2*], YIMIN XIONG[1], WEITIAN XU[1] and GUORONG ZHENG[1]

[1]Department of Digestive Diseases, Wuhan General Hospital of Guangzhou Command, Wuhan;
[2]Academy of Electronic Information and Control Engineering, Beijing University of Technology, Beijing, P.R. China

Received March 13, 2012; Accepted June 8, 2012

DOI: 10.3892/or.2012.1891

Abstract. Colon cancer is the third most common cancer and one of the leading causes of cancer-related death in the world. Therefore, identification of biomarkers with potential in recognizing the biological characteristics is a key problem for early diagnosis of colon cancer patients. In this study, we used a random forest approach to discover biomarkers based on a set of oligonucleotide microarray data of colon cancer. Real-time PCR was used to validate the related expression levels of biomarkers selected by our approach. Furthermore, ROC curves were used to analyze the sensitivity and specificity of each biomarker in both training and test sample sets. Finally, we analyzed the clinical significance of each biomarker based on their differential expression. A single classifier consisting of 4 genes (IL8, WDR77, MYL9 and VIP) was selected by random forests with an average sensitivity and specificity of 83.75 and 76.15%. The differential expression levels of each biomarker was validated by real-time PCR in 48 test colon cancer samples compared to the matched normal tissues. Patients with high expression of IL8 and WDR77, and low expression of MYL9 and VIP had a significantly reduced median survival rate compared to colon cancer patients. The results indicate that our approach can be employed for biomarker identification based on microarray data. These 4 genes identified by our approach have the potential to act as clinical biomarkers for the early diagnosis of colon cancer.

Introduction

Colon cancer is the third most common cancer, and one of the leading causes of morbidity and mortality in the world (1). According to the United States' statistics released in 2010 the incidence rate of colon has decreased (2). Over the last decade, many studies have proposed various kinds of statistical methods to analyze gene expression patterns and identify new biomarkers for prognostic and/or predictive information in relation to human diseases (3,4). However, most of the early studies applied unsupervised approaches to data-mining and identification of differential gene expressed profiling of certain diseases, such as hierarchical clustering for class discovering, taking an unbiased approach to searching for subgroups in the data (5). Along with the statistical methods extensively penetrated into the field of biomedicine, many supervised clustering analysis and machine learning approaches were adopted to deal with gene expression profiling data and sieved feature genes which contained more information to classify different kinds of diseases or subclasses of the same disease.

Various methods of statistics and machine learning, including clustering (6,7), Bayesian algorithms (8), and support vector machines (9), have been proposed to analyze microarray data generated through high-throughput experiments. Over the last few years, the technology of multiclassifier fusion developed substantially, and became very successful in improving the accuracy of certain classifiers. Random forests (RF) (10,11), a tree-based method of classification and regression, is one of the most important methods of multiclassifier fusion. Besides the outcome of classification, RF also returns several measures of variable importance according to which feature genes can be selected. Since RF is comparable with other methods and even better to a certain extent (12), it is used broadly especially for microarray data (13). Additionally, RF can be used as not only a supervised algorithm but also an unsupervised one (14), which depends on whether the gene expression data come from known classes or not.

In this study, we adopted an RF-based method for feature gene selection incorporating deductive reasoning to process the differential gene expressed profiling of colon cancer. We thus, selected 4 feature genes (IL8, WDR77, MYL9 and VIP) for colon cancer classification. Then, the differential expression level of each biomarker was validated by real-time PCR and in 48 test colon cancer samples compared to their matched normal tissues with high sensitivity and specificity. The results showed that our approach could filter out genes of great importance based on microarray data, and the genes selected by our approach were validated with high accuracy in classifying colon cancer and matched normal samples.

Correspondence to: Dr Guorong Zheng, Department of Digestive Diseases, Wuhan General Hospital of Guangzhou Command, Wuluo Road 627, Wuchang District, Wuhan 430070, P.R. China
E-mail: guorongzheng@sina.com

*Contributed equally

Key words: microarray, random forests, biomarker, colon cancer

Materials and methods

Micoarray data set. In 1999, Alon, *et al* (15) detected the whole genome of 40 colon tumor and 22 normal samples using an Affymetrix oligonuleotide array (Hum6000) and a two-way clustering approach to classify genes into functional groups. The microarray data was downloaded at: http://genomics-pubs.princeton.edu/oncology/affydata/index.html. To further study this group of microarray data and rediscover potential biomarkers not been mined completely, we used an RF-based machine learning method in our investigation.

RF algorithm. One of the most important supervised methods RF was used for data-mining in this study. The reliable measure is based on the decrease of classification accuracy when values of a variable in a node of a tree undergo random permutations (16). All training set observations were assigned to different terminal nodes in a tree and distinct split values were determined through several criterions such as the Gini index. The class of majority of training set observation in the terminal node was selected as the class of the node. We selected fewer genes with which the classifiers produced smallest out-of-bag (OOB) errors and highest classification scores.

For sample j, we defined mr_j as the difference between its accuracy rate and misclassifying rate. Additionally, we defined the mean decrease of accuracy rate of gene g as MDA (g). The calculating formulas of mr_j and MDA (g) are represented as follows:

$$mr_j = \left(\frac{\sum_{i=1}^{ntree} I(V_j(i) = Tclass) I(OOB_j(i) = T)}{\sum_{i=1}^{ntree} I(OOB_j(i) = T)}\right) - \left(1 - \frac{\sum_{i=1}^{ntree} I(V_j(i) = Tclass) I(OOB_j(i) = T)}{\sum_{i=1}^{ntree} I(OOB_j(i) = T)}\right) = 2\left(\frac{\sum_{i=1}^{ntree} I(V_j(i) = Tclass) I(OOB_j(i) = T)}{\sum_{i=1}^{ntree} I(OOB_j(i) = T)}\right) - 1 \quad \text{Eq. 1}$$

$$MDA(g) = \frac{1}{N}\sum_{j=1}^{N}(mr_j - mr_j(g)) = \frac{1}{N}\sum_{j=1}^{N}\frac{2}{\sum_{i=1}^{ntree} I(OOB_j(i) = T)} \times \sum_{i=1}^{ntree}(A - B) I(OOB_j(i) = T) \quad \text{Eq. 2}$$

$$A = I(V_j(i) = Tclass) \quad \text{Eq. 3}$$

$$B = I(V'_{j(g)}(i) = Tclass) \quad \text{Eq. 4}$$

I(g) denotes indicator function; ntree is the number of tree classifiers; N, total samples; $OOB_j(i) = T$, represents that sample j exists in OOB data set for tree i. If j is correctly classified by i, $V_j(i)$ = Tclass. Similarly, j is correctly classified after the value of gene g is randomly permuted $V_{j(g)}(i)$ = Tclass.

Table I. Top 20 genes with high classification score by the random forrest algorithm.

GenBank ID	Accession no.	Gene symbol	Score
M26383	**NM_000584**	**IL8**	**0.8776**
H08393	**NM_024102**	**WDR77**	**0.8520**
J02854	**BM473095**	**MYL9**	**0.8263**
M36634	**NM_003381**	**VIP**	**0.8124**
J05032	NM_001349	DARS	0.8108
T92451	CR590682	TPM2	0.8065
R36977	AK057993	GTF3A	0.8065
M22382	BC047350	HSPD1	0.8065
U25138	BC025707	KCNMB1	0.8065
D00860	NM_002764	PRPS1	0.8007
H43887	BQ712715	CFD	0.8007
X63629	NM_001793	CDH3	0.8007
T51571	BQ683841	S100A11	0.7963
Z50753	NM_007102	GUCA2B	0.7963
T96873	CR627338	CBWD1	0.7786
H64489	NM_005727	TSPAN1	0.7786
T60155	BX647362	ACTA2	0.7786
D14812	BC035249	MORF4L2	0.7786
T54303	CR607281	KRT8	0.7692
L41559	BM550965	PCBD1	0.7692

Bold indicates the genes selected as a classifier of colon cancer.

RNA isolation and real-time-PCR. A total of 48 colon cancer and matched normal tissues from Wuhan General Hospital of Guangzhou Command were used in this study for real-time-PCR experiment. Total RNA was extracted from the tissue samples according to a standard TRIzol protocol (Invitrogen, Carlsbad, CA, USA). Total RNA (5 μg) was reverse transcribed to cDNA with 200 U M-MLV reverse transcriptase (Promega, Madison) according to a standard manufacturer's protocol. The reverse transcription reaction conditions were: 37˚C for 60 min, 72˚C for 10 min. Real-time-PCR was performed in a total 20 μl reaction mixture with 2 μl of cDNA, 0.6 μl 20X EvaGreen (CapitalBio, Beijing, China), and 0.5 μl of each 10 μM forward and reverse primers, 0.5 μl of 2.5 mM dNTP, 1.5 U Cap Taq polymerase (CapitalBio), 10 μl of 2X PCR buffer for EvaGreen and 6.1 μl of ddH_2O. Quantification of differentially expressed genes was conducted with an RT-Cycle™ 2.0 system (CapitalBio). Real-time-PCR was carried out with programmed parameters, heating at 9˚C for 5 min followed by 40 cycles of a 3-stage temperature profile of 95˚C for 30 sec, 57˚C for 30 sec, 72˚C for 30 sec. All reactions were designed with 3 duplications and the final Ct values were determined by the average Ct value of the duplicated reaction. The melting curves for each PCR reaction were carefully analyzed to avoid non-specific amplifications in PCR products. The expression of each gene was transformed using the $2^{-\Delta\Delta Ct}$ formula and normalized with the β-actin expression (17).

Receiver operating curve (ROC) and statistical analysis. ROC curve analysis was conducted using the MedCalc software

Table II. Four genes selected by the random forest method based on microarray data.

Expression level	Gene symbol	GenBank ID	Fold-change	Q-value (%)
Upregulated	IL8	M26383	2.444	0.665
	WDR77	H08393	1.638	3.547
Downregulated	MYL9	J02854	0.011	0
	VIP	M36634	0.203	0

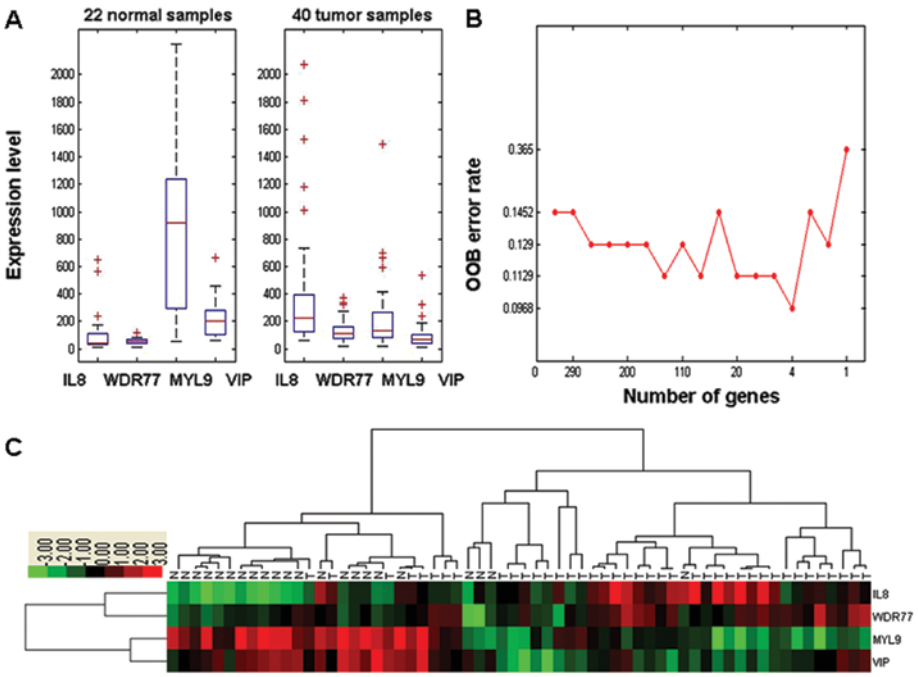

Figure 1. Identification of 4 genes by RF based on gene expression profiling. (A) The expression levels showed that IL8 and WDR77 were downregulated in 22 normal samples, and upregulated in tumor samples; MYL9 and VIP were downregulated in 40 tumor samples and upregulated in normal samples. (B) The smallest OOB error rate appeared when there were only 4 genes. The numbers of reserved genes were 320, 290, 260, 230, 200, 170, 140, 110, 80, 50, 20, 10, 5, 4, 3, 2 and 1. (C) Graphical overview of these 4 genes. Hierarchical clustering of the data matrix consists of 4 differential expressed genes by 40 colon cancers and 22 matched normal tissues. Columns represent samples and rows represent genes (black, green, and red correspond to no-change, downregulated and upregulated, respectively). T, tumor; N, normal.

packages (version 8.2.1.0; Mariakerke, Belgium). The AUC curves provided a measure of the overall performance of a diagnostic test. The ratio of gene signal intensities and the Ct value of each gene were used for ROC calculation in training and test samples, respectively. The clinical data were analyzed using the Chi-square test. The cumulative survival curve was compared by the log-rank test. For all analyses, a difference with $P<0.05$ was considered statistically significant.

Results

Biomarker rediscovery by the RF approach. We processed the microarray data of colon cancer using an RF-based algorithm. According to the OOB error rate, we identified 4 genes as a classifier to classify colon cancer and normal samples, composed of two upregulated genes IL8 and WDR77, and two downregulated genes MYL9 and VIP (Tables I and II). The classification accuracy of the 4-gene classifier was 91.94% (Fig. 1A). The average expression levels of each gene and the clustering graphical overview are shown in Fig. 1B and C.

Real-time PCR and IHC staining validation. cDNA from 48 colon cancer and matched normal tissues were used for real-time PCR experiment. The results showed that IL8 was upregulated in 37 of 48 cancer samples (77.1%) compared to the matched normal tissues with P-value of 0.032. Similarly, WDR77 was upregulated in 34 colon cancer samples (70.8%) with a P-value of 0.046. On the contrary, MYL9 was downregulated in 35 of 48 cancer samples (72.9%) with P-value of 0.028 and VIP was downregulated in 33 colon cancer samples (68.8%) with a P-value of 0.177 (Fig. 2).

ROC curve analysis. In order to analyze the classification sensitivity and specificity of the candidate biomarkers, we used ROC analysis both in training and test sample data. We observed a high sensitivity and specificity of the biomarkers and consistent results from both training and test samples. AUC-values of IL8, WDR77, MYL9 and VIP were 0.853, 0.875, 0.826 and 0.812 in the training group (Fig. 3A, Table III); 0.869, 0.867, 0.898 and 0.845 in the test group, respectively (Fig. 3B, Table III).

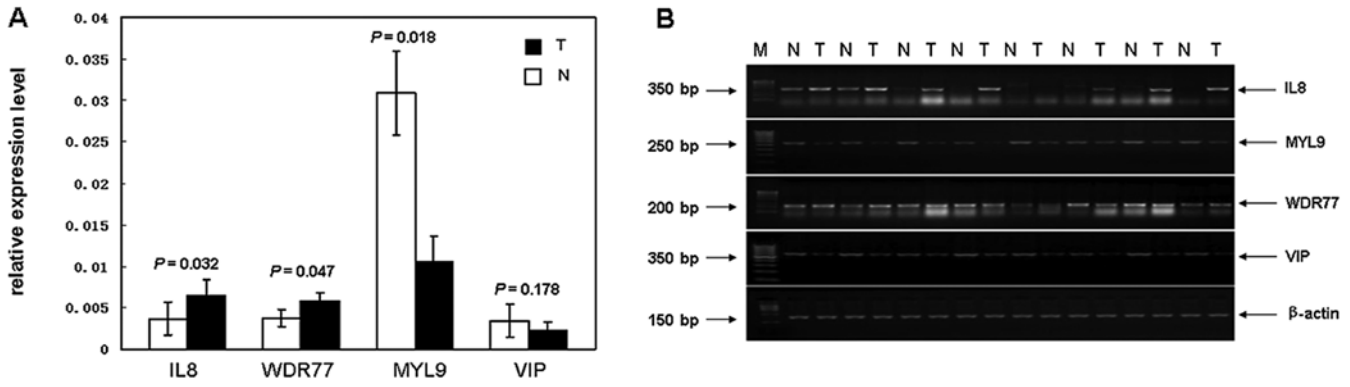

Figure 2. Relative expression levels of the candidate biomarkers validated by real-time PCR. (A) The $2^{-\Delta\Delta Ct}$ method was used to analysis the relative expression levels of the genes after real-time PCR. Quantitative real-time PCR results showed that IL8 and WDR77 were upregulated in colon cancer samples with P-values of 0.032 and 0.046; MYL9 and VIP were downregulated in colon cancer samples with the P-values of 0.028 and 0.177. (B) The same results were shown by semi-quantitative PCR.

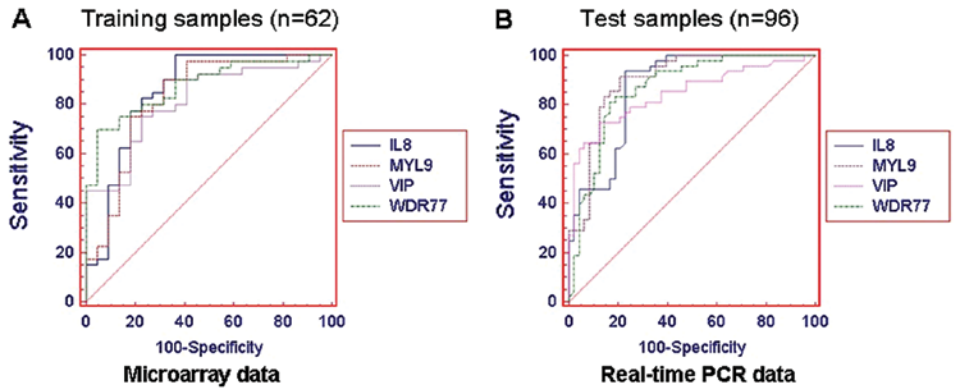

Figure 3. ROC curve analysis of candidate biomarkers. (A) ROC analysis based on microarray data. The AUC values of IL8, WDR77, MYL9 and VIP were 0.853, 0.875, 0.826 and 0.812 in the training group; (B) ROC analysis based on real-time PCR data (Ct value). The AUC values of IL8, WDR77, MYL9 and VIP were 0.869, 0.867, 0.898 and 0.845 in the test group, respectively.

Table III. ROC analyses of the sensitivity and specificity of candidate biomarkers.

Sample sets	Biomarkers	Sensitivity (%)	Specificity (%)	AUC	95% CI	SE	P-value
Training samples	IL8	100.0	63.6	0.853	0.740-0.930	0.0472	0.0001
	WDR77	70.0	95.5	0.875	0.766-0.945	0.0434	0.0001
	MYL9	90.0	68.2	0.826	0.709-0.910	0.0596	0.0001
	VIP	75.0	77.3	0.812	0.693-0.900	0.0614	0.0001
Test samples	IL8	93.8	77.1	0.869	0.785-0.929	0.0374	0.0001
	WDR77	81.2	83.3	0.867	0.782-0.928	0.0377	0.0001
	MYL9	91.7	79.2	0.898	0.820-0.951	0.0330	0.0001
	VIP	72.9	87.5	0.845	0.757-0.911	0.0405	0.0001

95% CI, 95% confidence interval; SE, standard error.

Clinical significance of the biomarkers. The expression levels of IL8, WDR77, MYL9 and VIP were used for comparing some of the clinical indicators in 48 colon cancer patients. A significant difference was observed in two groups which represent positive expression and negative expression of IL8 as follows: IL8(+) and IL8(-). Patients with IL8(+) had significantly reduced median survival compared to those with IL8(-) (P<0.001). Meanwhile, we observed that the positive expression of IL8 was associated with gender (P=0.029), clinical stage (P<0.001) and survival status (P<0.001) of colon cancer patients (Table IV). The expression levels of WDR77 were associated with the clinical stage (P=0.008),

Table IV. Statistical analyses of the biomarkers expression associated with the clinical significances of colon cancer patients.

	IL8			WDR77			MYL9			VIP		
Characteristics	(+) (n=37)	(-) (n=11)	P-value	(+) (n=34)	(-) (n=14)	P-value	(+) (n=13)	(-) (n=35)	P-value	(+) (n=15)	(-) (n=33)	P-value
Gender			**0.029**			0.338			0.234			0.133
Male	28	5		24	9		10	23		12	21	
Female	9	6		10	5		3	12		3	12	
Age (years)			0.082			0.063			0.101			0.348
Median	64	51		64	51		58	62		64	58	
Average	60	54		60	53		55	60		59	58	
Differentiation			0.404			0.146			0.462			**0.026**
Poor	22	7		22	7		8	21		6	23	
Moderate/well	15	4		12	7		5	14		9	10	
Lymph node resection			0.351			0.298			0.066			0.376
<12	8	3		7	4		1	10		3	8	
>12	29	8		27	10		12	25		12	25	
Clinical stage			**<0.001**			**0.008**			0.066			**0.037**
I/II	6	9		7	8		1	10		1	10	
III/IV	31	2		27	6		12	25		14	23	
Embolus			0.086			**0.035**			0.428			0.108
With	11	1		11	1		3	9		2	10	
Without	26	10		23	13		10	26		13	23	
Adjuvant chemotherapy			0.142			0.070			0.410			0.054
Performed	9	1		9	1		3	7		1	9	
Not performed	28	10		25	13		10	28		14	24	
Recurrence			0.479			0.097			0.120			**0.019**
Recurrence	7	2		8	1		1	8		0	9	
Non-recurrence	30	9		26	13		12	27		15	24	
Patients' status			**<0.001**			0.162			0.217			0.125
Survival	14	8		14	8		5	17		5	17	
Death	25	1		20	6		8	18		10	16	
Median survival time (month)	30.3	80.4	**<0.001**	35.3	41.8	**0.039**	31.8	41.8	**0.038**	30.3	44	**0.014**

Bold indicates P-values <0.05.

SPANDIDOS PUBLICATIONS
ONCOLOGY REPORTS 28: 1036-1042, 2012

numbers of the embolus (P=0.035) and the survival time of the patients. On the contrary, negative expression of MYL9 and VIP were associated with median survival time of colon cancer patients (Table IV). In addition, negative expression of VIP was associated with the differentiation status of cancer cell (P=0.026) and recurrence risk (P=0.019) of colon cancer patients (Table IV). The details of clinical significance for all the candidate biomarkers are shown in Table IV.

Discussion

Colon cancer is one of the most common diseases in the world, but only few tumor-specific gene products have been identified that could serve as targets to aid in the diagnosis of colon cancer. Its high prevalence and bad prognosis encourage researchers to find new biomarkers for the diagnosis and treatment of colon cancer. The microarray technique provides an effective method to identify a large scale of candidate biomarkers. Gene expression, methylation and microRNA profiling of colon cancer have been performed (18-20).

High-throughput microarray technologies have generated a large amount of data, where, various statistical and machine learning methods were adopted to analyze the data for finding gene or protein expression patterns and search for new biomarkers of human diseases. Microarray data analysis involves selecting the biomarkers which contain useful information necessary for molecular classification of human diseases and for establishing a gene expression profile. In this study, we present a concise investigative mode for feature gene selecting. We used a supervised machine learning algorithm RF to select gene a classifier based on differential gene expression profiling. A series of biological experiments were used to validate the results from high-throughput data.

RF is an effective algorithm with classifying quality comparable to other methods such as support vector machines (SVM) (10). It can also select featured genes which embody differentially expressed levels among different samples. We applied RF to deal with a colon cancer dataset and identified 4 genes which had great biological significance. The classifier composed of the 4 genes produced a high accuracy on both the training and the test samples. Bootstrap aggregating, a resample technique, is used when building the RF. This technique allows RF not to prune like other tree-based classification algorithm. Furthermore, RF can avoid over-fitting effectively although the mechanism is not currently clear. Besides dealing with the gene expression microarray data, RF has been extensively used in other aspects of biomedicine territory. In the latest years, RF was adopted extensively to analyze the single-nucleotide polymorphisms data (21) and the gene pathway building investigation (22).

In order to identify biomarkers with high sensitivity and specificity, verification in the laboratory and detection of new test clinical samples are important. Real-time PCR and tissuemicroarray-based IHC staining provided us convenient and precise approaches to detect the expression levels of candidate biomarkers. Our results also showed that real-time PCR was sensitive and specific for gene expression level validation. The PCR-based detection method therefore, appears to provide us with an easy way in early clinical diagnosis of human cancer.

The function and clinical significance of IL8 and VIP have been reported. There are 1,182 studies describing the gene function of IL8, including the biological mechanism in progress of most kinds of human cancer such as: glioblastoma, gastric carcinoma, small cell lung cancer, prostate cancer, esophageal squamous cell carcinoma, acute myelogenous leukemia, and colon cancer (23-29). It was confirmed that IL8 is differentially expressed in colon cancer, and is associated with proliferation, migration, angiogenesis and chemosensitivity in colon cancer cell line models (30). The VIP gene has also been the focus of investigation in many studies ralating to human cancer (31-36). WDR77, also known as p44, was reported to be related to the differentiation and proliferation in prostate epithelium (37). Its differential expression was observed in ovarian cancer (38). However, there is no report associating WDR77 with colon cancer. Thus, WDR77 is a novel potential biomarker of colon cancer. Similarly to WDR77, MYL9 has not been well-documented as being functionally associated with human cancer, including colon cancer. Therefore we reconfirmed its expression levels both at the RNA and protein levels by PCR and IHC methods, respectively.

In summary, we used an RF-based method to process a differential gene expression profile of colon cancer and selected 4 featured genes as candidate biomarkers of colon cancer. We validated these biomarkers in clinical colon cancer samples by a real-time PCR method. Our results showed that this approach filtered out genes of great importance, like IL8 and VIP based on microarray data, also including some new genes as WDR77 and MYL9 with the potential to act as cancer-related biomarkers.

Acknowledgements

This research was supported by the National Nature Science Foundation of China (no. 61075110) and the Scientific Plan of Beijing Municipal Commission of Education (JC002011200903). The authors wish to acknowledge Mr. Zhikun Gao (Beijing University of Technology) for providing assistance in processing the data with a machine learning algorithm.

References

1. West NP, Morris EJ, Rotimi O, Cairns A, Finan PJ and Quirke P: Pathology grading of colon cancer surgical resection and its association with survival: a retrospective observational study. Lancet Oncol 9: 857-865, 2008.
2. Jemal A, Siegel R, Xu J and Ward E: Cancer statistics, 2010. CA Cancer J Clin 60: 277-300, 2010.
3. Kosari F, Parker AS, Kube DM, Lohse CM, Leibovich BC, Blute ML, Cheville JC and Vasmatzis G: Clear cell renal cell carcinoma: gene expression analyses identify a potential signature for tumor aggressiveness. Clin Cancer Res 11: 5128-5139, 2005.
4. Zhang X, Yan Z, Zhang J, Gong L, Li W, Cui J, Liu Y, Gao Z, Li J, Shen L and Lu Y: Combination of hsa-miR-375 and hsa-miR-142-5p as a predictor for recurrence risk in gastric cancer patients following surgical resection. Ann Oncol 22: 2257-2266, 2011.
5. John Q: Microarray analysis and tumor classification. N Engl J Med 354: 2463-2472, 2006.
6. Wu GP, Chan KC and Wong AK: Unsupervised fuzzy pattern discovery in gene expression data. BMC Bioinformatics 12 (Suppl 5): S5, 2011.
7. Broom BM, Sulman EP, Do KA, Edgerton ME and Aldape KD: Bagged gene shaving for the robust clustering of high-throughput data. Int J Bioinform Res Appl 6: 326-343, 2010.

8. Pique-Regi R, Monso-Varona J, Ortega A, Seeger RC, Triche TJ and Asgharzadeh S: Sparse representation and Bayesian detection of genome copy number alterations from microarray data. Bioinformatics 24: 309-318, 2008.
9. Chen L, Xuan J, Riggins RB, Clarke R and Wang Y: Identifying cancer biomarkers by network-constrained support vector machines. BMC Syst Biol 5: 161, 2011.
10. Statnikov A, Wang L and Aliferis CF: A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. BMC Bioinformatics 9: 319, 2008.
11. Manilich EA, Özsoyoğlu ZM, Trubachev V and Radivoyevitch T: Classification of large microarray datasets using fast random forest construction. J Bioinform Comput Biol 9: 251-267, 2011.
12. Qi Y, Bar-Joseph and Klein-Seetharaman J: Evaluation of different biological data and computational classification methods for use in protein interaction prediction. Proteins 63: 490-500, 2006.
13. Ramon DU and Sara AA: Gene selection and classification of microarray data using random forest. BMC Bioinformatics 7: 3, 2006.
14. Cutler A and Stevens JR: Random forests for microarrays. Methods Enzymol 411: 422-432, 2006.
15. Alon U, Rarkai N, Notterman DA, Gish K, Ybarra S, Mack D and Levine AJ: Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci USA 96: 6745-6750, 1999.
16. Strobl C, Boulesteix AL, Zeileis A and Hothorn T: Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics 8: 25, 2007.
17. Kenneth JL and Thomas DS: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25: 402-408, 2001.
18. Alves PM, Lévy N, Stevenson BJ, Bouzourene H, Theiler G, Bricard G, Viatte S, Ayyoub M, Vuilleumier H, Givel JC, *et al*: Identification of tumor-associated antigens by large-scale analysis of genes expressed in human colorectal cancer. Cancer Immun 8: 11, 2008.
19. Schetter AJ, Leung SY, Sohn JJ, Zanetti KA, Bowman ED, Yanaihara N, Yuen ST, Chan TL, Kwong DL, Au GK, *et al*: MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. JAMA 299: 425-436, 2008.
20. Chung W, Kwabi-Addo B, Ittmann M, Jelinek J, Shen L, Yu Y and Issa JP: Identification of novel tumor markers in prostate, colon and breast cancer by unbiased methylation profiling. PLoS One 3: 2079, 2008.
21. Nicodemus KK, Wang W and Shugart YY: Stability of variable importance scores and rankings using statistical learning tools on single-nucleotide polymorphisms and risk factors involved in gene x gene and gene x environment interactions. BMC Proc 1 (Suppl 1): S58, 2007.
22. Pang H and Zhao H: Building pathway clusters from Random Forests classification using class votes. BMC Bioinformatics 9: 87, 2008.
23. de la Iglesia N, Konopka G, Lim KL, Nutt CL, Bromberg JF, Frank DA, Mischel PS, Louis DN and Bonni A: Deregulation of a STAT3-interleukin 8 signaling pathway promotes human glioblastoma cell proliferation and invasiveness. J Neurosci 28: 5870-5878, 2008.
24. Canedo P, Castanheira-Vale AJ, Lunet N, Pereira F, Figueiredo C, Gioia-Patricola L, Canzian F, Moreira H, Suriano G, Barros H, *et al*: The interleukin-8-251®T/®A polymorphism is not associated with risk for gastric carcinoma development in a Portuguese population. Eur J Cancer Prev 17: 28-32, 2008.
25. Yoshida C, Niiya K, Niiya M, Shibakura M, Asaumi N and Tanimoto M: Induction of urokinase-type plasminogen activator, interleukin-8 and early growth response-1 by STI571 through activating mitogen activated protein kinase in human small cell lung cancer cells. Blood Coagul Fibrinolysis 18: 425-433, 2007.
26. Araki S, Omori Y, Lyn D, Singh RK, Meinbach DM, Sandman Y, Lokeshwar VB and Lokeshwar BL: Interleukin-8 is a molecular determinant of androgen independence and progression in prostate cancer. Cancer Res 67: 6854-6862, 2007.
27. Savage SA, Abnet CC, Mark SD, Qiao YL, Dong ZW, Dawsey SM, Taylor PR and Chanock SJ: Variants of the IL8 and IL8RB genes and risk for gastric cardia adenocarcinoma and esophageal squamous cell carcinoma. Cancer Epidemiol Biomarkers Prev 13: 2251-2257, 2004.
28. Bruserud Ø, Ryningen A, Wergeland L, Glenjen NI and Gjertsen BT. Osteoblasts increase proliferation and release of pro-angiogenic interleukin 8 by native human acute myelogenous leukemia blasts. Haematologica 89: 391-402, 2004.
29. Landi S, Moreno V, Gioia-Patricola L, Guino E, Navarro M, de Oca J, Capella G, Canzian F; Bellvitge Colorectal Cancer Study Group: Association of common polymorphisms in inflammatory genes interleukin (IL)6, IL8, tumor necrosis factor alpha, NFKB1, and peroxisome proliferator-activated receptor gamma with colorectal cancer. Cancer Res 63: 3560-3566, 2003.
30. Ning Y, Manegold PC, Hong YK, Zhang W, Pohl A, Lurje G, Winder T, Yang D, LaBonte MJ, Wilson PM, *et al*: Interleukin-8 is associated with proliferation, migration, angiogenesis and chemosensitivity in vitro and in vivo in colon cancer cell line models. Int J Cancer 128: 2038-2049, 2011.
31. Ogasawara M, Murata J, Ayukawa K and Saiki I: Differential effect of intestinal neuropeptides on invasion and migration of colon carcinoma cells in vitro. Cancer Lett 119: 125-130, 1997.
32. Singh AT, Jaggi M, Prasad S, Dutt S, Singh G, Datta K, Rajendran P, Sanna VK, Mukherjee R and Burman AC: Modulation of key signal transduction molecules by a novel peptide combination effective for the treatment of gastrointestinal carcinomas. Invest New Drugs 26: 505-516, 2008.
33. Valdehita A, Carmena MJ, Collado B, Prieto JC and Bajo AM: Vasoactive intestinal peptide (VIP) increases vascular endothelial growth factor (VEGF) expression and secretion in human breast cancer cells. Regul Pept 144: 101-108, 2007.
34. Haberl I, Frei K, Ramsebner R, Doberer D, Petkov V, Albinni S, Lang I, Lucas T and Mosgoeller W: Vasoactive intestinal peptide gene alterations in patients with idiopathic pulmonary arterial hypertension. Eur J Hum Genet 15: 18-22, 2007.
35. Absood A, Hu B, Bassily N and Colletti L: VIP inhibits human HepG2 cell proliferation in vitro. Regul Pept 146: 285-292, 2008.
36. Collado B, Sánchez-Chapado M, Prieto JC and Carmena MJ: Hypoxia regulation of expression and angiogenic effects of vasoactive intestinal peptide (VIP) and VIP receptors in LNCaP prostate cancer cells. Mol Cell Endocrinol 249: 116-122, 2006.
37. Gao S, Wu H, Wang F and Wang Z: Altered differentiation and proliferation of prostate epithelium in mice lacking the androgen receptor cofactor p44/WDR77. Endocrinology 151: 3941-3953, 2010.
38. Ligr M, Patwa RR, Daniels G, Pan L, Wu X, Li Y, Tian L, Wang Z, Xu R, Wu J, *et al*: Expression and function of androgen receptor coactivator p44/Mep50/WDR77 in ovarian cancer. PLoS One 6: e26250, 2011.