Journal of
Microbial & Biochemical Technology

Huang et al., J Microb Biochem Technol 2014, 6:2
DOI: 10.4172/1948-5948.1000129

Research Article Open Access

A Gene Expression Profile of Peripheral Blood in Colorectal Cancer

Chi-Shuan Huang[1], Harn-Jing Terng[2], Yu-Chin Chou[3], Sui-Lung Su[3], Yu-Tien Chang[3,4], Chin-Yu Chen[2], Woan-Jen Lee[2], Chung-Tay Yao[5], Hsiu-Ling Chou[6], Chia-Yi Lee[3,4], Chien-An Sun[7], Ching-Huang Lai[3], Lu Pai[3,8], Chi-Wen Chang[9]*, Kang-Hwa Chen[9], Thomas Wetter[10],Yun-Wen Shih[3,4] and Chi-Ming Chu[3,4]*

[1]Division of Colorectal Surgery, Cheng Hsin Rehabilitation Medical Center, Taiwan
[2]Advpharma, Inc., Taipei, Taiwan
[3]School of Public Health, National Defense Medical Center, Taiwan
[4]Division of Biomedical Statistics and informatics, School of Public Health, National Defense Medical Center, Taiwan
[5]Department of Surgery, Cathay General Hospital, Taipei, Taiwan.
[6]Department of Nursing, Far Eastern Memorial Hospital & Oriental Institute of Technology, New Taipei, Taiwan.
[7]Department of Public Health, Fu Jen Catholic University, Taiwan
[8]Taipei Medical University, Taiwan
[9]Department of Nursing, College of Medicine, Chang Gung University, Taiwan
[10]Institute of Medical Biometry and Informatics, Heidelberg University, Germany and Department of Biomedical Informatics and Medical Education, University of Washington Seattle, USA

Abstract

Background: Optimal molecular markers for detecting colorectal cancer (CRC) in a blood-based assay were evaluated. Microarray technology has shown a great potential in the colorectal cancer research. Genes significantly associated with cancer in microarray studies, were selected as candidate genes in the study. Pooling Internet public microarray data sets can overcome the limitation by the small number of samples in previous studies.

Objective: Using public microarray data sets verifies gene expression profiles for colorectal cancer.

Methods: Logistic regression analysis was performed, and odds ratios for each gene were determined between CRC and controls. Public microarray datasets of GSE 4107, 4183, 8671, 9348, 10961, 13067, 13294, 13471, 14333, 15960, 17538, and 18105 included 519 cases of adenocarcinoma and 88 controls of normal mucosa, which were used to verify the candidate genes from logistic models and estimated its external generality.

Results: A 7-gene model of CPEB4, EIF2S3, MGC20553, MAS4A1, ANXA3, TNFAIP6 and IL2RB was pairwise selected that showed the best results in logistic regression analysis (H-L p=1.000, R2=0.951, AUC=0.999, accuracy=0.968, specificity=0.966 and sensitivity=0.994).

Conclusions: A novel gene expression profile was associated with CRC and can potentially be applied to blood-based detection assays.

Keywords: Colorectal cancer; Gene expression; Microarray; Internet

Background

Colorectal cancer (CRC) is a common cancer worldwide [1]. An estimated 146,970 new cases of colon and rectal cancer and 49,920 deaths are expected to occur in 2009 in the United States [2]. CRC screening can possibly reduce the incidence of advanced disease and provide better overall, progression-free survival. Conventional CRC screening tests include fecal occult blood testing, flexible sigmoidoscopy, double-contrast barium enema X-ray, and colonoscopy [3]. Although they are commonly used, these tests have limitations, including highly variable sensitivity (i.e., 37% to 80%) and diet-test interactions [4].

The dissemination of malignant cells from a primary neoplasm is the pivotal event in cancer progression. In many clinical cases, tumor cells metastasize before the primary tumor is diagnosed. Individual circulating tumor cells may be the earliest detectable form of metastasis [5]. PCR-based analyses of mRNA from cytokeratins, the carcinoembryonic antigen (CEA), and epidermal growth factor receptor (EGFR) genes in peripheral blood samples from CRC patients have been reported [6]. However, the low sensitivities and specificities for these well-known genes are not considered acceptable for the detection of colorectal cancer. Recently, multiple biomarkers were reported for the detection of colorectal cancer that delivered a better sensitivity or specificity [7,8].

In the present study, expression levels of 28 cancer-associated candidate genes in the peripheral blood samples from 111 colorectal cancer patients and 227 non-cancer controls were analyzed using quantitative real time-PCR. Genes correlated with CRC were selected, and a discrimination model was constructed using multivariate logistic regression. Sensitivity, specificity, accuracy, positive and negative predictive values, and the AUC of the discrimination model are reported. Meanwhile, from this study (Model 1: 5 genes), Marshall et al. [7] (Model 2: 7 genes) and Han et al. [8] (Model 3: 5 genes), the 17 candidate genes were validated by pooling12 public microarray data sets as well as the external validation.

***Corresponding authors:** Prof. Dr. Chi-Ming Chu, Division of Bioinformatics and Statistics, Department of Epidemiology, School of Public Health, National Defense Medical Center, Taiwan, Tel: 886-2-87923100 X18438; Fax: 886-2-87923147; E-mail: chuchiming@web.de

Dr. Chi-Wen Chang, Department of Nursing, College of Medicine, Chang Gung University, Taiwan, E-mail: chuchiming@web.de

Received January 15, 2014; **Accepted** February 12, 2014; **Published** February 19, 2014

Citation: Huang CS, Terng HJ, Chang CW, Chou YC, Chang YT, et al. (2014) A Gene Expression Profile of Peripheral Blood in Colorectal Cancer. J Microb Biochem Technol 6: 102-109. doi:10.4172/1948-5948.1000129

Methods

Patients, controls, and blood samples

One hundred eleven patients with histologically confirmed colorectal cancer were enrolled (2006-2009) in a prospective investigational protocol, which was approved by the Institutional Review Board at Cheng Hsin Rehabilitation Medical Center (Taipei, Taiwan). CRC patients at different stages were classified according to the TNM system (Table 1). Peripheral blood samples (6-8 ml) were drawn from patients before any therapeutic treatment, including surgery, but after written informed consent was obtained. All blood samples were collected using BD vacutainer CPT™ tubes containing sodium citrate as an anti-coagulant (Becton Dickinson, NJ, USA) and were stored at 4°C.

The healthy controls were 227 volunteers who had come in for a routine health examination and had no evidence of any clinically detectable cancer disease. Each participant gave informed consent for the analysis. The same volume of peripheral blood was collected from controls as from patients. Samples were randomly divided into a training set (n=162) and a testing set (n=176). There were no significant differences in age, sex, cancer stage or tumor site between the two sets (Table 1).

RNA isolation and reverse transcription

The mononuclear cell (MNC) fraction was isolated within three hours after blood collection using BD vacutainer CPT™ tubes (Becton Dickinson, NJ, USA), according to the manufacturer's instructions. Total RNA was then extracted from the MNC fraction using the Super RNApure™ kit (Genesis, Taiwan) according to the manufacturer's instructions. The average yield of total RNA per milliliter of peripheral blood was 1.6 μg. The mRNA quality was assessed by the electrophoresis of total RNA, followed by staining with ethidium bromide, which showed two clear rRNA bands of 28S and 18S. Using a spectrophotometer, the ratio of the absorbance of each RNA at 260 and 280 nm (A260:A280) was confirmed to be greater than 1.7, which is an indicator of RNA purity [9]. One microgram of total RNA was used for cDNA synthesis with random hexamer primers (Amersham Bioscience, UK) and Superscript™ II reverse transcriptase (Invitrogen, USA).

Quantitative real-time PCR

Real-time PCR was performed using pre-designed, gene-specific amplification primer sets purchased from Advpharma, Inc. (Taiwan), nucleotide probes from Universal ProbeLibrary™ (Roche, Germany) and TaqMan® Master Mix (Roche) on a Roche LightCycler® 1.5 instrument. The hypoxanthine phosphoribosyltransferase 1 (HPRT1) gene was used as the internal control because its expression accurately reflects the mean expression of multiple commonly used normalization genes [10,11]. The cycle number for each candidate gene, Ct(test), was normalized against the cycle number of HPRT1, that is, C_t(HK). The calculation is performed as follows: ΔC_t(test)=Ct(HK) - Ct(test). The derived (normalized) value, ΔC_t(test), for each candidate gene is presented as the relative difference as compared to the mRNA expression level of the reference gene [12].

Preliminary screening of investigating genes

In the preliminary screening for CRC-associated genes, we selected candidate genes from the published microarray study [14], and tested for the relative range of expression levels using real-time PCR. There were totally 28 gene candidates for first run of screening, including 12 genes, which were reported as risk for cancer prognosis [14], 14 genes identified as correlated with the incidence of tumor tissues (unpublished data), and two genes with elevated expression in colon cancer patients, A3 adenosine receptor and CCSP-2 [15,16]. Since the measurement of a higher cycle number (i.e., Ct greater than 30) generally implies lower amplification efficiency [17,18], 15 genes were used for further analysis (Table 2) after eliminating genes with low amplification efficiency.

Statistical analysis

The chi-square test and *t*-test were performed to characterize sex and age distributions between cases and controls. The transcript levels of candidate genes were tested statistically for differences between the case and control samples, using the t-test. A logistic regression was performed, and odds ratios were determined in order to study the association of candidate genes with CRC. The power of the study was 100% for each candidate gene [13]. The statistical alpha level was 0.05.

Multivariate logistic regression was used to analyze the relationship of the cases and controls to the ΔC_t(test) values of candidate genes. The logistic probabilities were calculated using the modeling equations from logistic regression analysis. Diagnostic performances were further used to evaluate multivariate logistic models, including sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). We used the Hosmer-Lemeshow test to check goodness-of-fit. A receiver operating characteristic (ROC) curve analysis was

	Training set (n=162)			**Testing set (n=176)**			**P-value**	
	CRC (n=55)	Non-CRC (n=107)	**P-value**	**CRC (n=56)**	**Non-CRC (n=120)**	**P-value**	**Cases**	**Controls**
Age, yr (S.E.)	66.47 (1.50)	68.31 (1.12)	0.335	67.38 (1.83)	69.99 (1.03)	0.216	0.704	0.270
Gender, no. (%) Male Female	 32 (58.2) 23 (41.8)	 58(54.2) 49(45.8)	0.630	 28 (50.0) 28 (50.0)	 73 (60.8) 47 (39.2)	0.176	0.387	0.313
Stage, no. (%) I II III IV	 21 (38.2) 10 (18.2) 14 (25.5) 10 (18.2)	 - - - -	-	 15 (26.8) 9 (16.1) 21 (37.5) 11 (19.6)	 - - - -	-	0.447	-
Tumor site, no. (%) Colon Rectum Cecum Colon+Rectum	 28 (50.9) 22 (40.0) 4 (7.3) 1 (1.8)	-	-	 30 (53.6) 16 (28.6) 5 (8.9) 5 (8.9)	-	-	0.286	-

CRC: ColonRectal Cancer; *Data are given as means (SE) or as the number of cases (%); §P values were estimated using the t-test

Table 1: Characteristics of the training and testing sets*§.

			95% CI of OR		
	B	OR	Upper	Lower	P-value
Sex	0.577	1.780	7.582	0.418	0.435
Age	0.028	1.028	1.083	0.976	0.293
MCM4	0.142	1.152	4.504	0.295	0.838
ZNF264	1.450	4.265	18.208	0.999	0.050
RNF4	-0.550	0.577	5.146	0.065	0.622
GRB2	2.009	7.456	37.131	1.497	0.014
MDM2	1.359	3.892	15.166	0.999	0.050
STAT2	-1.178	0.308	1.466	0.065	0.139
WEE1	1.264	3.540	14.784	0.848	0.083
DUSP6	2.465	11.769	40.330	3.435	<0.001
CPEB4	2.045	7.725	27.695	2.155	0.002
MMD	-1.067	0.344	0.865	0.137	0.023
NF1	-1.417	0.243	1.517	0.039	0.130
IRF4	0.057	1.059	3.350	0.335	0.923
EIF2S3	-2.105	0.122	0.718	0.021	0.020
EXT2	-1.933	0.145	1.235	0.017	0.077
POLDIP2	-1.294	0.274	1.515	0.050	0.138

B: coefficient of logistic regression; OR: odds ratio; CI: confidence interval

Table 2: Multivariate analysis of colorectal cancer-related molecular markers and the discrimination model based on age, sex, and 15 genes using the logistic regression model on the training set.

performed to determine the cut-off logistic probabilities and the areas under the ROC curves (AUC), to identify the performance of each candidate gene and combinations of multiple genes. A sensitivity analysis demonstrated the influence on performance of different cut-off logistic probabilities [Logit(P)] in the logistic model.

Internet public microarray data sets

The microarray gene expression data are from searches using "colon cancer" AND "human [organism]" AND "expression profiling by array [dataset type]" as the key words in the GEO database of the National Center for Biotechnology Information (NCBI). The eligible criteria were 1) the examined samples were frozen tissue sections of normal human colorectal mucosa, primary colorectal cancer or hepatic metastases from colorectal cancer; 2) the microarray platform used was limited to single-color, whole genome gene chips from Affymetrix; and 3) the data were presented as gene expression level. The exclusion criteria were 1) data from cultured cell lines or other in vitro assays; 2) datasets without the original gene expression level data files; and 3) those with redundant sub-datasets. A total of 175 GEO series (GSE) datasets were finally excluded, leaving 12 public microarray dataset of GSE 4107, 4183, 8671, 9348, 10961, 13067, 13294, 13471, 14333, 15960, 17538, and 18105, which included 519 cases of adenocarcinoma and 88 controls of normal mucosa.

Furthermore, we validated the 17 CRC-associated genes from the studies (Model 1: 5 genes), Marshall et al. [7] (Model 2: 7 genes) and Han et al. [8] (Model 3: 5 genes) and performed the multivariate logistic regression analysis using the pooled 12 public microarray data sets as well as the external validation.

Results

Genes correlated with colorectal cancer

A multivariate analysis based on age, sex and 15 genes was used in a logistic regression model in the training set because the peripheral blood samples were drawn from patients before any therapeutic treatment. Although this full model seemed capable of discriminating between the CRC cases and controls, it may have resulted in over fitting(Table 2). The logistic regression analysis further resulted in the selection of five genes of significance (i.e., P-value<0.05), MDM2, DUSP6, CPEB4, MMD, and EIF2S3, with odds ratios of 2.978, 6.029, 3.776, 0.538, and 0.138, respectively. This model was reduced to a panel of five genes in a forward stepwise regression, which statistical powers of the five genes were 1.00 between case and control groups in training and testing sets.

Discrimination of colorectal cancer and non-cancer controls using five genes

Five genes, i.e., MDM2, DUSP6, CPEB4, MMD, and EIF2S3, were significantly associated with CRC. A five-gene logistic regression model provided good discriminative performance with 87.0% accuracy, 78.0% sensitivity, 92.0% specificity, 90.7% positive predictive value (PPV), and 80.7% negative predictive value (NPV) in the training set. The five-gene model performed with 94.9% accuracy, 97.1% sensitivity, 81.8% specificity, 96.9% PPV, 82.8% NPV, and an area under the ROC (receiver operating characteristic) curve of 0.978 (0.912-1) in the external validation. Discrimination models can be constructed with one of the five genes selected, based on forward multivariate logistic regression analysis using the training set. AUCs were used to compare the performance of discrimination models for single genes and combinations of two, three, four, or five marker genes. The DUSP6 model (Table 3) displayed the best discrimination ability, with an AUC of 0.804 (95% C.I.: 0.730-0.879), as compared to other one-gene models (AUC: 0.49-0.69). Distinct increases in the AUC of up to 0.905 (95% C.I.: 0.849-0.960) resulted from the combination of five genes (Table 3). The five-gene model fulfilled the criteria of good performance for diagnostic tests as well as accuracy (87%), sensitivity (78%), and specificity (92%); in addition, the Hosmer-Lemeshow test was non-significant (P-value=0.108).

The cut-off value of Logit(P) for the five-gene model could also be adjusted to achieve high sensitivity or specificity, i.e., 99%, 95% or 90% (Table 4). The five-gene model performed stably for the discrimination between CRC cases and controls in the training set, with accuracies

				95% CI	
Genes used for models	**AUC**	**S.E.**	**P-value**	**Lower**	**Upper**
DUSP6, CPEB4, EIF2S3, MDM2, MMD	0.905	0.028	<0.001	0.849	0.960
DUSP6, CPEB4, EIF2S3, MDM2	0.895	0.030	<0.001	0.838	0.953
DUSP6, CPEB4, EIF2S3	0.882	0.032	<0.001	0.820	0.945
DUSP6, CPEB4	0.855	0.032	<0.001	0.791	0.919
DUSP6	0.804	0.038	<0.001	0.730	0.879

ROC: receiver operating characteristic; AUC: area under the ROC curve; S.E.: Standard Error; CI: confidence interval. P-values for AUC were estimated using the Z test

Table 3: Discrimination power and ROC analysis of different combinations of CRC-associated genes in training set.

a: Logistic probabilities for the training set

Logit(P)	Sensitivity	Specificity	PPV	NPV	Accuracy
0.0198	99.0%	16.0%	2.3%	99.9%	44.2%
0.0511	95.0%	63.0%	12.1%	99.6%	73.9%
0.1783	90.0%	72.0%	41.1%	97.1%	78.1%
0.5	78.0%	92.0%	90.7%	80.7%	87.0%
0.4747	80.0%	90.0%	87.8%	83.3%	86.6%
0.6845	61.0%	95.0%	96.4%	52.9%	83.5%
0.9012	25.0%	99.0%	99.6%	12.6%	73.9%

Logit(P): Logistic Probabilities; PPV: Positive Predictive Value; NPV: Negative Predictive Value

b: Performance of the statistical model on the training and testing sets with Logit(P)=0.5

	Training set	Testing set	External validation
Non-Cancers	107	120	88
True negative	98	110	72
False positive	9	10	16
Colorectal Cancers	55	56	519
False negative	12	19	15
True positive	43	37	504
Total	162	176	519
Sensitivity	78.0%	66.0%	97.1%
Specificity	92.0%	92.0%	81.8%
PPV	90.7%	89.2%	96.9%
NPV	80.7%	73.0%	82.8%
Accuracy	87.0%	83.5%	94.9%

Logit(P): Logistic Probabilities; PPV: Positive Predictive Value; NPV: Negative Predictive Value

Table 4: Performance of the statistical model based on the five-gene profile.

ranging from 73.9% to 87.0%, with sensitivity of 95%, or with specificity of 95%. In addition, a well performance in the testing set was obtained using the discrimination model, with 84% accuracy, 66% sensitivity, 92% specificity, 89% PPV and 73% NPV (Table 4b).

Pooling 12 microarray studies to verify the 17 candidate genes and estimate its external generality

Furthermore, we performed the multivariate logistic regression analysis for pooled 12 public microarray data sets as well as the external validation to verify the CRC-associated genes from 3 studies (the present one of Chu et al., Marshall et al. and Han et al.) [7,8]. As the Table 5, we validated the 17 CRC-associated genes from this study (Model 1: 5 genes), Marshall et al. [7] (Model 2: 7 genes) and Han et al. [8] (Model 3: 5 genes) by pooling 12 public microarray dataset of GSE 4107, 4183, 8671, 9348, 10961, 13067, 13294, 13471, 14333, 15960, 17538, and 18105, which included 519 cases of adenocarcinoma and 88 controls of normal mucosa. The goodness-of-fit test of Hosmer-Lemeshow (H-L) showed statistical significance (p=0.044) for Model 2 of Marshall et al. [7], which observed event rates did not match expected event rates in subgroups of the model population. Models for which expected and observed event rates in subgroups are similar are called well calibrated (Model 1, 3 and 4). A 7-gene model (Model 4 with genes CPEB4, EIF2S3, MGC20553, MAS4A1, ANXA3, TNFAIP6 and IL2RB) was pairwise selected from genes of Model 1, 2 and 3 that showed the best results in logistic regression analysis (H-L p=1.000, R2=0.951, AUC=0.999, accuracy=0.968, specificity=0.966 and sensitivity=0.994).

Discussion

Common serum tumor markers used in primary care practice have not demonstrated a survival benefit in randomized controlled trials for screening in the general population. Most of them showed elevated levels only in some early-stage or late-stage cancer patients [19]. A recent review of real-time PCR-based assays with single molecular markers, such as CEA, CK19, and CK20, demonstrated low sensitivity, was ranging from 4% to 35.9%, 25.9% to 41.9%, and 5.1% to 28.3%, respectively [6]. One study, performed with a newly identified molecular marker known as ProtM [20], also attained unsatisfactory sensitivity.

Circulating tumor cells from any cancer type are capable of disseminating from solid tumor tissues, penetrating and invading blood vessels and circulating in the peripheral blood [21,22]. The number of circulating tumor cells has been used to predict the clinical outcome of

	Model 1			Model 2			Model 3			Model4		
	B	S.E.	p	B	S.E.	p	B	S.E.	p	B	S.E.	p
5 Candidate genes of this study;												
MDM2	6.069	1.461	<0.001									
DUSP6	1.360	0.235	<0.001									
CPEB4	-3.177	0.383	<0.001							-4.423	1.160	<0.001
MMD	0.335	0.442	0.448									
EIF2S3	1.462	0.244	<0.001							2.604	0.856	0.002
7 Candidate genes of Marshall et al. [7]												
ANXA3				0.559	0.212	0.008				1.566	0.485	0.001
CLEC4D				46.259	9.918	<0.001						
LMNB1				1.883	0.330	<0.001						
PRRG4				-1.284	0.371	0.001						
TNFAIP6				1.787	0.377	<0.001				2.031	0.572	<0.001
VNN1				0.207	0.159	0.194						
IL2RB				0.269	0.216	0.213				1.824	0.637	0.004
5 Candidate genes of Han et al. [8]												
CDA							-0.496	0.090	<0.001			
MGC20553							-1.386	0.197	<0.001	-1.751	0.619	0.005
BANK1							0.565	0.373	0.129			
BCNP1							-0.944	1.148	0.411			
MAS4A1							-1.483	0.457	0.001	-1.907	0.590	0.001
Constant	-32.758	6.001	<0.001	-124.678	25.437	<0.001	16.601	2.995	<0.001	-14.268	6.968	0.041
H-L		0.460			0.044			0.194			1.000	
R2		0.853			0.841			0.693			0.951	
AUC		0.978			0.985			0.957			0.999	
Accuracy		0.949			0.974			0.939			0.990	
Specificity		0.818			0.886			0.716			0.966	
Sensitivity		0.971			0.988			0.977			0.994	

Model 1: 5 candidate genes of this study; Model 2: 7 candidate genes of Marshall et al. ; Model 3: 5 candidate genes of Han et al. ; Model 4: stepwise 7 candidate genes from model 1, 2 and 3; B: logistic regression coifficient beta; S.E.: standard error of B; p: p value with statistical significance; H-L: Hosmer and Lemeshow test p value R2: Nagelkerke R Square; AUC: area under ROC

Table 5: The logistic regression models for pooled 12 microarray data sets as the external validation of CRC-associated genes from 3 studies.

cancer patients [23]. On the basis of the presence of circulating tumor cells, we identified five molecular markers, MDM2, DUSP6, CPEB4, MMD, and EIF2S3, which were differentially expressed between peripheral blood samples of CRC patients and healthy controls. The application of multivariate logistic regression analysis resulted in a five-gene discrimination model, which achieved good diagnostic performance and provided stable conditions with accuracies ranging from 73.9% to 87.0%, with sensitivity of 95%, or with specificity of 95%.

Both mRNAs and proteins in the peripheral blood have been tested for diagnostic use to detect circulating tumor cells of different solid tumors or to determine prognoses of various cancers. We confirmed, in our study, that the AUCs of the discrimination models greatly improved from 0.80 for the model based on a single gene (DUSP6) to 0.91 for the combined model with all five genes. More and more clinical studies show improvements in the sensitivity of cancer detection by assaying transcript levels of multiple genes in patient peripheral blood [7,8,24].

A higher sensitivity or specificity of the discriminatory performance of our five-gene model was achieved by adjusting the cut-off value of Logit(P) (Table 4a). This five-gene discrimination model with Logit(P)=0.0511 had a sensitivity of 95%, a specificity of 63%, and an accuracy of 74%, which is ideal for screening colorectal cancer. However, setting Logit(P) to 0.4747 resulted in specificity of 90%, sensitivity of 80% and an accuracy of 86%, which indicates that our five-gene model is robust and highly accurate for discriminating CRC from healthy or benign conditions. Similar accuracy rates (i.e., 80% to 86%) were achieved with Logit(P), ranging from 0.0511 to 0.4747. In the testing set, the five-gene model performed with satisfactory accuracy, sensitivity, and specificity.

Two reports [7,8] with similar screening approaches used different gene sets to detect CRC (Table 5). The two gene sets were obtained by direct selection from differentially expressed genes in peripheral blood samples using microarray techniques followed by real-time PCR. The biomarkers they selected may more or less reflect the static and dynamic changes of the immune system in response to cancer. The strategy of our study was to choose genes clinically confirmed to be cancer-associated in tumor tissues and to validate in peripheral blood samples. Five genes (MDM2, DUSP6, CPEB4, MMD and EIF2S3) identified here for discrimination between CRC patients and healthy controls showed strong association with CRC. MDM2 (Mouse double minute 2 homolog) gene, also known as HDM2 gene in human, is a negative regulator of the tumor suppressor protein p53 [25]. Overexpression of MDM2 gene was reported in several human tumor types, including osteosarcomas, melanoma, non-small cell lung cancer (NSCLC), esophageal cancer, leukemia, and non-Hodgkin's lymphoma [26-31]. Inhibition of MDM2 can restore p53 activity in cancers containing wild-type p53 and has recently become a strategy to develop anti-tumor drug [32-35].

DUSP6, the dual-specificity MAP kinase phosphatase 3 (also known as MKP3), inactivates ERK1/ERK2 [36,37]. Clinical studies

based on tumor tissues demonstrated that elevated DUSP6 transcript (mRNA) level was a risk factor for clinical outcome in non-small cell lung cancer (NSCLC) patients (hazard ratio=2.2) [14] and stronger protein level was identified in 31% of primary human NSCLC tumor using Immunohistochemistry [38]. Furthermore, overexpression of DUSP6 was associated with papillary and poorly differentiated thyroid carcinoma both at the mRNA and protein level [39,40] and with KRAS mutant colon cancer [41]. In addition, higher expression level of DUSP6 was found in the tamoxifen-resistant breast tumors group compared with the tamoxifen-sensitive tumor group [42] and tumor growth promotion in glioblastoma [43]. The DUSP6 function might vary in different cancer types. On the contrary, some other reports demonstrated that DUSP6 gene was a candidate tumor suppressor gene, for instance, in pancreatic cancer [44], esophageal squamous cell carcinoma [45], and lung cancer with 17.7% cases of study sample [46].

CPEB4, cytoplasmic polyadenylation element binding protein, targets mRNAs and promote translation by inducing cytoplasmic polyadenylation [47,48]. Overexpressed CPEB4 was identified in pancreatic ductal adenocarcinomas and glioblastomas compared with its corresponding normal tissue [49]. Increased CPEB4 mRNA was considered as a prognostic marker for poor clinical outcome in non-small cell lung cancer (NSCLC) patients (hazard ratio=1.8) [14]. In contrast, reduced or weaker CPEB4 expression was observed in most of hepatocellular carcinoma samples compared with normal tissues using IHC staining [50]. In addition, Xu and Liu [51] proposed that CPEB4 gene might be selectively overexpressed in metastasic cancers, such as metastatic prostate cancer, and potentially as a biomarker for chemotherapy resistance.

MMD is an integral membrane protein with seven putative transmembrane segments [52]. Its biological function is still unclear. EIF2S3 is the largest subunit (gamma) of eukaryotic translation initiation factor 2 (EIF2) [53] and might be indirectly involved in inhibition of prostate cancer metastasis through N-myc downstream regulated gene 1 [54].

Our study has firstly presented that four expressing genes in PBMC-derived fractions, including MDM2, CPEB4, EIF2S3 and MMD, have the direct association with CRC with significance. As many clinical studies have been reported, MDM2, DUSP6 and CPEB4 have been showed their association with other pathologies, especially different cancer types. These observations might provide the evidence that these biomarkers play central roles during carcinogenesis or malignance of tumor, but with different strength depending on cancer type. Indeed, it is important to have multiple biomarkers integrated in developed diagnostic or prognostic methods, while each candidate gene has its independent power and the efficacy to discriminate cancer and normal subject (Table 3).

There are several limitations of current study. Since the small number of different stages in the study CRC sample, we were not able to know whether individual gene expression or the five gene signature is stage-dependent. Secondly, the change of gene expression level in the BPMC fraction of CRC patient before and after treatment was not studied due to the restriction of single blood drawing of IRB-approved clinical protocol. Thirdly, the collection of survival status information is not completed and prognostic value of biomarkers could not be evaluated in this study due to many censored cases (over 50%; 5-year survival rate of CRC patients is around 50%).

Further investigation is warranted on the potential of gene signature for evaluation of clinical staging, metastatic probability and survival in CRC patients, when information for the disease progression and survival is completely collected. In addition, the application of currently identified gene signature for CRC detection is very important and it is also the goal for assay development. Discussion with physicians will be planning for integration the test of our CRC-specific gene signature into the national screening program for CRC. Especially, the diagnostic performance between this CRC-specific gene signature and current screening method, such fecal occult blood test (FOBT) or colonoscopy should be approached. As well as the potential of the individual gene expression or gene signature for evaluation of therapeutic response should be planned as future work.

Furthermore, we verified the CRC-associated genes by pooling 12 public microarray data sets that the four logistic models performed similar AUCs without statistically significant difference. In the future, the 7-gene logistic regression model (Model 4: CPEB4, EIF2S3, MGC20553, MAS4A1, ANXA3, TNFAIP6 and IL2RB) showed the best results that can be further verified for more samples. Meanwhile, the causal relations are needed to confirm among the selected genes and CRC. The expression signature of these CRC-associated genes should be evaluated for early detection of CRC, with more samples randomly screened from the population; in addition, subjects who eventually receive a diagnosis of CRC should be evaluated as well. Early CRC detection could provide inherent benefits to the patient and could also enable screening for post-operative residual tumor cells and occult metastases, an early indicator of tumor recurrence. Early detection could thus improve survival in patients before symptoms are detectable, during treatment, or during remission.

Conclusion

In conclusion, we found that the expression profile of 7 genes, CPEB4, EIF2S3, ANXA3, TNFAIP6, IL2RB, MGC20553 and MAS4A1, is highly associated with colorectal cancer. Detection of cancer cell-specific biomarkers in the peripheral blood can be an effective screening strategy for CRC.

Competing Interests

No other potential conflict of interest relevant to this article was reported. The authors have declared that no competing interests exist. Dr. Terng, Woan-Jen Lee, and Chin-Yu Chen report being employees of Advpharma. Authors of Advpharma involved in partial sample collection, laboratoryexperiment and manuscript.

Financial Disclosure

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Finance was supported by Taiwan's SBIR promoting program from the Department of Industrial Technology of Ministry of Economic Affairs and by research grants from National Defense Medical Center of Ministry of Defense.

Ethics Statement

Patients with histologically confirmed colorectal cancer were enrolled (2006-2009) in a prospective investigational protocol, which was approved by the Institutional Review Board at Cheng Hsin Rehabilitation Medical Center (Taipei, Taiwan). CRC patients at different stages were classified according to the TNM system (Table 1). Peripheral blood samples (6-8 ml) were drawn from patients before any therapeutic treatment, including surgery, but after written informed consent were obtained.

Authors' contributions

Conception and design: Chi-Ming Chu and Harn-Jing Terng. Administrative support: Chi-Ming Chu, Harn-Jing Terng, Woan-Jen Lee and Chin-Yu Chen. Provision of study materials or patients: Woan-Jen Lee and Chin-Yu Chen. Collection and assembly of data: Chi-Suan Huang, Woan-Jen Lee, Chin-Yu Chen, and Yun-Wen Shih. Data analysis and interpretation: Chi-Ming Chu and Yun-Wen

Shih. Manuscript writing: Chi-Ming Chu, Harn-Jing Terng, and Yun-Wen Shih. Final approval of manuscript: Chi-Ming Chu, Harn-Jing Terng, Chi-Suan Huang, Woan-Jen Lee, Chin-Yu Chen, and Mark L. Wahlqvist.

Acknowledgments

Research Funding: This work was supported by Taiwan's SBIR promoting program from the Department of Industrial Technology of the Ministry of Economic Affairs, Advpharma, Inc., and the National Defense Medical Center (NDMC), Bureau of Military Medicine, Ministry of Defense, Taiwan. Employment or Leadership Position: Chi-Ming Chu, NDMC; Harn-Jing Terng, Advpharma, Inc. Taiwan. Honoraria: Wu-ChienChien; Ching-Huang Lai, NDMC; Mark Sarno, Stason Pharmaceuticals, Inc., CA, USA.Expert Testimony: None. Other Remuneration: None.

References

1. Parkin DM, Bray F, Ferlay J, Pisani P (2005) Global cancer statistics, 2002. CA Cancer J Clin 55: 74-108.
2. Jemal A, Siegel R, Ward E, Hao Y, Xu J, et al. (2008) Cancer statistics, 2008. CA Cancer J Clin 58: 71-96.
3. Smith RA, Cokkinides V, Eyre HJ (2006) American Cancer Society guidelines for the early detection of cancer, 2006. CA Cancer J Clin 56: 11-25.
4. Levin B, Lieberman DA, McFarland B, Smith RA, Brooks D, et al. (2008) Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. CA Cancer J Clin 58: 130-160.
5. Fidler IJ (1990) Critical factors in the biology of human cancer metastasis: twenty-eighth G.H.A. Clowes memorial award lecture. Cancer Res 50: 6130-6138.
6. Sergeant G, Penninckx F, Topal B (2008) Quantitative RT-PCR detection of colorectal tumor cells in peripheral blood--a systematic review. J Surg Res 150: 144-152.
7. Marshall KW, Mohr S, Khettabi FE, Nossova N, Chao S, et al. (2010) A blood-based biomarker panel for stratifying current risk for colorectal cancer. Int J Cancer 126: 1177-1186.
8. Han M, Liew CT, Zhang HW, Chao S, Zheng R, et al. (2008) Novel blood-based, five-gene biomarker set for the detection of colorectal cancer. Clin Cancer Res 14: 455-460.
9. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd edn 1989, NY Cold Spring Harbor Laboratory Press.
10. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: RESEARCH0034.
11. de Kok JB, Roelofs RW, Giesendorf BA, Pennings JL, Waas ET, et al. (2005) Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes. Lab Invest 85: 154-159.
12. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402-408.
13. The Survey System: Sample Size Calculator.
14. Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, et al. (2007) A five-gene signature and clinical outcome in non-small-cell lung cancer. N Engl J Med 356: 11-20.
15. Xin B, Platzer P, Fink SP, Reese L, Nosrati A, et al. (2005) Colon cancer secreted protein-2 (CCSP-2), a novel candidate serological marker of colon neoplasia. Oncogene 24: 724-731.
16. Gessi S, Cattabriga E, Avitabile A, Gafa' R, Lanza G, et al. (2004) Elevated expression of A3 adenosine receptors in human colorectal cancer is reflected in peripheral blood cells. Clin Cancer Res 10: 5895-5901.
17. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29: e45.
18. Hunt M: Real Time Pcr. In Book Real Time Pcr (Editor Ed.^Eds.). City.
19. Perkins GL, Slater ED, Sanders GK, Prichard JG (2003) Serum tumor markers. Am Fam Physician 68: 1075-1082.
20. Schuster R, Max N, Mann B, Heufelder K, Thilo F, et al. (2004) Quantitative real-time RT-PCR for detection of disseminated tumor cells in peripheral blood of patients with colorectal cancer using different mRNA markers. Int J Cancer 108: 219-227.
21. Bogenrieder T, Herlyn M (2003) Axis of evil: molecular mechanisms of cancer metastasis. Oncogene 22: 6524-6536.
22. Carmeliet P, Jain RK (2000) Angiogenesis in cancer and other diseases. Nature 407: 249-257.
23. Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Matera J, et al. (2004) Circulating tumor cells, disease progression, and survival in metastatic breast cancer. N Engl J Med 351: 781-791.
24. Shen C, Hu L, Xia L, Li Y (2008) Quantitative real-time RT-PCR detection for survivin, CK20 and CEA in peripheral blood of colorectal cancer patients. Jpn J Clin Oncol 38: 770-776.
25. Reifenberger G, Liu L, Ichimura K, Schmidt EE, Collins VP (1993) Amplification and overexpression of the MDM2 gene in a subset of human malignant gliomas without p53 mutations. Cancer Res 53: 2736-2739.
26. Marchetti A, Buttitta F, Girlando S, Dalla Palma P, Pellegrini S, et al. (1995) mdm2 gene alterations and mdm2 protein expression in breast carcinomas. J Pathol 175: 31-38.
27. Bueso-Ramos CE, Yang Y, deLeon E, McCown P, Stass SA, et al. (1993) The human MDM-2 oncogene is overexpressed in leukemias. Blood 82: 2617-2623.
28. Ladanyi M, Cha C, Lewis R, Jhanwar SC, Huvos AG, et al. (1993) MDM2 gene amplification in metastatic osteosarcoma. Cancer Res 53: 16-18.
29. Momand J, Jung D, Wilczynski S, Niland J (1998) The MDM2 gene amplification database. Nucleic Acids Res 26: 3453-3459.
30. Onel K, Cordon-Cardo C (2004) MDM2 and prognosis. Mol Cancer Res 2: 1-8.
31. Heist RS, Zhou W, Chirieac LR, Cogan-Drew T, Liu G, et al. (2007) MDM2 polymorphism, survival, and histology in early-stage non-small-cell lung cancer. J Clin Oncol 25: 2243-2247.
32. Wasylyk C, Salvi R, Argentini M, Dureuil C, Delumeau I, et al. (1999) p53 mediated death of cells overexpressing MDM2 by an inhibitor of MDM2 interaction with p53. Oncogene 18: 1921-1934.
33. Chène P, Fuchs J, Bohn J, García-Echeverría C, Furet P, et al. (2000) A small synthetic peptide, which inhibits the p53-hdm2 interaction, stimulates the p53 pathway in tumour cell lines. J Mol Biol 299: 245-253.
34. Wang H, Nan L, Yu D, Lindsey JR, Agrawal S, et al. (2002) Anti-tumor efficacy of a novel antisense anti-MDM2 mixed-backbone oligonucleotide in human colon cancer models: p53-dependent and p53-independent mechanisms. Mol Med 8: 185-199.
35. Tortora G, Caputo R, Damiano V, Bianco R, Chen J, et al. (2000) A novel MDM2 anti-sense oligonucleotide has anti-tumor activity and potentiates cytotoxic drugs acting by different mechanisms in human colon cancer. Int J Cancer 88: 804-809.
36. Keyse SM (2008) Dual-specificity MAP kinase phosphatases (MKPs) and cancer. Cancer Metastasis Rev 27: 253-261.
37. Zhou B, Wu L, Shen K, Zhang J, Lawrence DS, et al. (2001) Multiple regions of MAP kinase phosphatase 3 are involved in its recognition and activation by ERK2. J Biol Chem 276: 6506-6515.
38. Zhang Z, Kobayashi S, Borczuk AC, Leidner RS, Laframboise T, et al. (2010) Dual specificity phosphatase 6 (DUSP6) is an ETS-regulated negative feedback mediator of oncogenic ERK signaling in lung cancer cells. Carcinogenesis 31: 577-586.
39. Degl'Innocenti D, Romeo P, Tarantino E, Sensi M, Cassinelli G, et al. (2013) DUSP6/MKP3 is overexpressed in papillary and poorly differentiated thyroid carcinoma and contributes to neoplastic properties of thyroid cancer cells. Endocr Relat Cancer 20: 23-27.
40. Lee JU, Huang S, Lee MH, Lee SE, Ryu MJ, et al. (2012) Dual specificity phosphatase 6 as a predictor of invasiveness in papillary thyroid cancer. Eur J Endocrinol 167: 93-101.
41. De Roock W, Janssens M, Biesmans B, Jacobs B, De Schutter J, et al. (2009) DUSPs as markers of MEK/Erk activation in primary colorectal cancer. J Clin Oncol 27: 4064.
42. Cui Y, Parra I, Zhang M, Hilsenbeck SG, Tsimelzon A, et al. (2006) Elevated

expression of mitogen-activated protein kinase phosphatase 3 in breast tumors: a mechanism of tamoxifen resistance. Cancer Res 66: 5950-5959.

43. Messina S, Frati L, Leonetti C, Zuchegna C, Di Zazzo E, et al. (2011) Dual-specificity phosphatase DUSP6 has tumor-promoting properties in human glioblastomas. Oncogene 30: 3813-3820.

44. Furukawa T, Yatsuoka T, Youssef EM, Abe T, Yokoyama T, et al. (1998) Genomic analysis of DUSP6, a dual specificity MAP kinase phosphatase, in pancreatic cancer. Cytogenet Cell Genet 82: 156-159.

45. Ma J, Yu X, Guo L, Lu SH (2013) DUSP6, a tumor suppressor, is involved in differentiation and apoptosis in esophageal squamous cell carcinoma. Oncol Lett 6: 1624-1630.

46. Okudela K, Yazawa T, Woo T, Sakaeda M, Ishii J, et al. (2009) Down-Regulation of DUSP6 Expression in Lung Cancer. Am J Pathol 175:867-881.

47. Huang YS, Kan MC, Lin CL, Richter JD (2006) CPEB3 and CPEB4 in neurons: analysis of RNA-binding specificity and translational control of AMPA receptor GluR2 mRNA. EMBO J 25: 4865-4876.

48. Hake LE, Richter JD (1994) CPEB is a specificity factor that mediates cytoplasmic polyadenylation during Xenopus oocyte maturation. Cell 79: 617-627.

49. Ortiz-Zapater E, Pineda D, Martínez-Bosch N, Fernández-Miranda G, Iglesias M, et al. (2011) Key contribution of CPEB4-mediated translational control to cancer progression. Nat Med 18: 83-90.

50. Tian Q, Liang L, Ding J, Zha R, Shi H, et al. (2012) MicroRNA-550a acts as a pro-metastatic gene and directly targets cytoplasmic polyadenylation element-binding protein 4 in hepatocellular carcinoma. PLoS One 7: e48958.

51. Xu H, Liu B (2013) CPEB4 is a candidate biomarker for defining metastatic cancers and directing personalized therapies. Med Hypotheses 81: 875-877.

52. Rehli M, Krause SW, Schwarzfischer L, Kreutz M, Andreesen R (1995) Molecular cloning of a novel macrophage maturation-associated transcript encoding a protein with several potential transmembrane domains. Biochem Biophys Res Commun 217: 661-667.

53. Gaspar NJ, Kinzy TG, Scherer BJ, Hümbelin M, Hershey JW, et al. (1994) Translation initiation factor eIF-2. Cloning and expression of the human cDNA encoding the gamma-subunit. J Biol Chem 269: 3415-3422.

54. Tu LC, Yan X, Hood L, Lin B (2007) Proteomics analysis of the interactome of N-myc downstream regulated gene 1 and its interactions with the androgen response program in prostate cancer cells. Mol Cell Proteomics 6: 575-588.