

# Identification of an intermediate signature that marks the initial phases of the colorectal adenoma-carcinoma transition

HUI TANG<sup>1</sup>, QIANG GUO<sup>2</sup>, CHAO ZHANG<sup>2</sup>, JUN ZHU<sup>3</sup>, HUI YANG<sup>4</sup>, YUN-LIAN ZOU<sup>1</sup>,  
YAN YAN<sup>5</sup>, DONG HONG<sup>1</sup>, TAO SOU<sup>1</sup> and XIN-MIN YAN<sup>1</sup>

<sup>1</sup>Institute of Basic Medicine, Key Laboratory of Yunnan Province; Departments of <sup>2</sup>Digestive System, <sup>3</sup>General Surgery II and <sup>4</sup>Pathology, The First People's Hospital of Yunnan Province, Kunming, Yunnan 650032; <sup>5</sup>College of Pharmacy, China Pharmaceutical University, Nanjing, Jiangsu 211198, P.R. China

Received June 7, 2010; Accepted August 24, 2010

DOI: 10.3892/ijmm\_00000508

**Abstract.** The colorectal adenoma-carcinoma sequence describes the stepwise progression from normal to dysplastic epithelium and then to carcinoma. Only a small proportion of colorectal adenomas (CRAs) progress to colorectal carcinomas (CRCs). Endoscopic intervention is currently being used on patients with high grade dysplasia CRAs, with diameters of >1 cm, or villous components of >25% who are at higher risk than other CRA sufferers. During the process, biopsy samples are taken for conventional histological diagnosis, but poor pathomorphological sensitivity and specificity greatly limit the diagnostic accuracy. Unfortunately, there are no reliable molecular criteria available that can predict the potential development of CRA to CRC. Gene expression profiles of normal colorectal mucosa (NOR), CRA and different Dukes' stages of CRC biopsy specimens, which represent the gradual progress of the CRA to CRC sequence, were determined by Affymetrix technology. Representative regulated genes were further analyzed by quantitative real-time PCR (qRT-PCR) and immunohistochemistry (IHC). Intersectional analyses of discriminative expression signatures of CRC vs. CRA and CRA vs. NOR allowed the identification of an intermediate signature of 463 probe sets (psets) that mark the NOR→CRA→CRC progression. This signature represents a reservoir of candidate markers for the early diagnosis of higher-risk CRA, thus allowing for timely therapeutic intervention and more selective treatment. A further 279 CRC-specific pssets pointing to the malignant transition from CRA to CRC were identified and these could represent potential therapeutic targets for CRC. The reliability of the results was further

confirmed by qRT-PCR and IHC analyses of the 4-gene sets randomly selected from the 463 pssets.

## Introduction

Colorectal carcinoma (CRC) is one of the most frequent cancers worldwide with very high mortality. According to WHO data, ~945,000 new CRC cases are diagnosed, and almost 492,000 CRC-related deaths occur worldwide every year (1). In Europe, CRC accounts for significant overall mortality (2,3). In China, CRC tops the list of the 5 most common causes of cancer death. Moreover, the incidence of CRC in China has been steadily increasing due to changes in lifestyle and nutritional habits (4). It is noticeable that the 5-year survival difference between early and late stages is still striking, dropping from 96% for stage I to only 5% for stage IV despite the recent improvements in surgical and therapeutic methods (3). Therefore, great attention has been directed at early detection in order to reduce mortality and morbidity due to CRC.

One of the most important fundamental concepts in CRC is the adenoma-carcinoma sequence, a term that describes the stepwise progression from normal to dysplastic epithelium and then to carcinoma (5). This concept has not been directly proven, and there are a few studies that have argued about its limitations pertaining to the colorectum (6,7). However, there is considerable indirect morphological (8), molecular (9), epidemiological and histopathological evidence that supports it (10). Most importantly, various clinical observations have clearly demonstrated that the removal of colorectal adenomas (CRAs) by endoscopic polypectomy is associated with a substantial reduction in the incidence of and mortality from CRC (11). Therefore, it has been widely accepted that CRAs represent the process by which most, if not all, CRCs arise (5). However, it is clear that only a small proportion of CRAs progress to malignancy (5). In fact, ~40% of the Western population will develop CRA during their lifetime, whereas <3% of these people will go on to suffer from CRC (12). Unfortunately, there are no reliable molecular criteria available that can predict higher-risk CRA patients during the early stages. Patients with high-grade dysplasia adenomas of with diameters of >1 cm or villous proportions of >25% are

---

*Correspondence to:* Dr Xin-Min Yan, Institute of Basic Medicine, Key Laboratory of Yunnan Province, The First People's Hospital of Yunnan Province, 157 Jinbi Road, Kunming, Yunnan 650032, P.R. China  
E-mail: yxmin08@163.com

**Key words:** colorectal carcinoma, colorectal adenoma, gene-expression profiling, microarray, signature

currently recognized as being at a much higher risk than other CRA sufferers. Therefore, endoscopic intervention is performed on such patients (13), and during the process, biopsy samples are taken for conventional histological diagnosis. Although a new technique based on computerized morphometry has been used to describe the growth pattern of colon carcinoma more objectively and quantitatively (14), unfortunately, the poor sensitivity and specificity of pathomorphology greatly limit the diagnostic accuracy. Therefore, if we could further identify patients with higher-risk CRAs with a great potential of developing into CRCs by detecting a diagnostic expression profile from the biopsy specimens taken during the endoscopic intervention, important questions pertaining to which CRA sufferer would be most likely to develop CRC and required follow-up protocols, could be answered. This could eventually lead to more accurate classification and more effective therapeutic regimens.

Gene expression analysis of colorectal biopsies using high density oligonucleotide microarrays could help detect such gene expression profiles that would establish the basis for new molecular diagnostic methods, and efforts have been made in this direction. In order to understand the gene expression background of CRC progression and metastasis development, many previous studies have primarily focused on the pair-wise comparisons of discriminative mRNA expression signatures between CRC and normal colorectal mucosa (NOR) samples, between CRA and NOR samples, as well as between the early and advanced stages of CRC (15-24). Results from various studies have demonstrated that CRC and/or CRA have widespread genetic alterations compared to NOR (25-27). However, much less is known in terms of direct comparisons between CRC and CRA, which could predict higher-risk CRA patients. To our knowledge, only a few research groups have reported the discriminative mRNA expression signatures between CRC vs. CRA and CRA vs. NOR (28,29), and the molecular mechanism of the CRA to CRC sequence remains to be elucidated.

In this study, NOR, CRA and different Dukes' stages of CRC biopsy specimens, representing the gradual progress of the CRA to CRC sequence, were obtained. mRNA expression analysis using Affymetrix Human Genome U133 Plus 2.0 Arrays was then performed to determine the difference in gene expression profiles between CRC and CRA. Our objective was to identify an intermediate gene expression signature that could be used to predict the potential conversion of CRA into CRC by detecting the gene expression trends of this particular signature. This would allow for more selective treatment strategies for higher-risk CRA patients.

## Materials and methods

**Ethics.** This study was carried out in accordance with the Declaration of Helsinki (2000) of the World Medical Association, and approved by the Ethics Committee of the First People's Hospital of Yunnan Province. All biopsy specimens were obtained after the patients gave their consent for the study.

**Patients and healthy control specimens.** For genome-wide expression profiling, 72 CRA, 99 CRC specimens and 60

carcinoma adjacent normal biopsies at least 8 cm away from the tumor, were obtained from patients who underwent either endoscopic intervention or colorectal resection at the First People's Hospital of Yunnan Province from October 2008 to June 2009. Six other NOR specimens from several locations throughout the colon were obtained from 6 Chinese individuals undergoing colonoscopy examinations. These healthy controls underwent colonoscopies for reasons ranging from abdominal pain and change in bowel habits to bleeding from hemorrhoids. These 6 healthy controls were found to have no polyps and no known family history or previous incidence of CRC. Altogether, 66 specimens were designated as the healthy controls. Of note, the 4 healthy control specimens that were used for the microarray analysis were selected from the 6 healthy controls. All the normal and cancerous portions of the specimens were snap-frozen in liquid nitrogen within 30 min of harvesting and stored thereafter at -80°C. Some of the samples were formalin-fixed and paraffin-embedded for subsequent immunohistochemistry (IHC) examination. Tumor staging was carried out using Dukes' staging system. Three CRA and 7 CRC samples used in the microarray analysis were specifically re-reviewed using paraffin-embedded tissues that were from areas adjacent or in close proximity to where the frozen samples were taken from. Clinical and pathological features of all the CRA, CRC and NOR specimens used for this study, are listed in Table I.

**RNA extraction and cDNA microarray.** The microarray experimental design is shown in Fig. 1A. Total RNA was extracted from each specimen using TRIzol RNA isolation reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's specifications. RNA integrity was checked using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), and concentrations were determined using NanoDrop technology (NanoDrop Technologies, Wilmington, DE). RNA samples isolated from several single individual tissues were pooled to reduce variations across the samples and to identify molecular markers that are expressed in the majority of individuals. The RNA was further cleaned using RNeasy mini kit columns (Qiagen, Izasa, Madrid, Spain). Next, biotinylated complementary RNA targets were prepared with 5 µg of total RNA using the IVT Labeling Kit (Affymetrix) according to the manufacturer's instructions. The targets were then fragmented at 94°C for 35 min and hybridized to 5 Human Genome U133 Plus 2.0 Arrays containing >54,000 probe sets (psets) for 16 h (Hybridization Oven 640, Affymetrix).

Following hybridization, the arrays were washed and stained with streptavidin/phycoerythrin conjugated and anti-streptavidin antibodies in the Affymetrix Fluidic Station 450, and scanned with the GeneChip Scanner 3000 (Affymetrix). Signals were then analyzed with the GeneChip Operating Software (GCOS 1.2, Affymetrix). The quality of the experiments was controlled by the visual inspection of the arrays and by the presence of the spike controls and housekeeping control genes. All data used for subsequent analyses passed the quality control criteria (GEO Accession: GSE22242).

**Statistical analysis.** Gene expression data, when classified as either flag-P (present) or flag-M (marginal) in >30% of all the samples, were loaded into the software. All expression

Table I. Characteristics of samples used in the present study.

	Group code	NOR	CRA	CRC		
				DA CRC	DB CRC	DC/D CRC
Microarray	Microarray code	H	A	B	C	D
	N	4	3	2	3	2
	Mean age (range, years)	59.75 (55-67)	68.00 (58-81)	59.50 (55-65)	61.33 (56-67)	62.50 (53-72)
	Gender (male/female)	2/2	1/2	1/1	2/1	1/1
qRT-PCR validation (including samples used in microarray)	N	18	22	7	8	8
	Mean age (range, years)	58.85 (50-73)	62.30 (52-72)	58.65 (51-73)	65.63 (51-73)	63.82 (51-73)
	Gender (male/female)	11/7	14/8	4/3	5/3	5/3
IHC validation	N	48	50	30	20	26
	Mean age (range, years)	64.64 (32-80)	53.94 (34-78)	61.47 (26-86)	59.65 (35-73)	60.68 (35-81)
	Gender (male/female)	29/19	26/24	11/19	10/10	15/11

DA CRC, Dukes' A CRC; DB CRC, Dukes' B CRC; DC/D CRC, Dukes' C/D CRC.

data on an array were normalized to the 50th percentile of all the values on that array followed by normalization to the median expression level of that gene across all the samples. Detection ('present', 'absent', or 'marginal'), change ('increased', 'decreased', or 'no change') and the fold change (FC) status (signal log ratio >0,  $FC=2^{\text{signal log ratio}}$ ; signal log ratio <0,  $FC=2^{-\text{signal log ratio}}$ ) of the signals generated from the CRA (microarray code A), CRC (microarray code B, C and D) and NOR samples (microarray code H) were analyzed with the GeneSpring software version 7.2 (Silicon Genetics, Redwood City, CA).

**Gene expression profile comparison of CRC vs. CRA and CRA vs. NOR and cluster analysis.** In order to find genes whose expression levels significantly differed between CRC and CRA, or CRA and NOR, we adopted a supervised method of analysis by using the GeneSpring 7.2 software. Mean values were calculated within the 2 classes for each pset and the FC ratios between 2 comparative groups, were derived. A 2-fold cut-off difference was applied to select the up- and down-regulated genes. Further statistical analysis was performed using Welch's approximate t-test and ANOVA. Two-dimensional hierarchical clustering was then applied to the log-transformed data with average-linkage clustering using a standard correlation as the similarity metric for discriminating genes that were identified as differentially expressed between CRC and CRA.

**Gene functional category analysis.** Gene Ontology (GO) categories were analyzed by the BioScript Library tool of GeneSpring 7.2. Genes were classified according to their annotated role in the biological processes, molecular functions,

as well as cellular components from the GO Consortium. A hypergeometric P-value was used to measure the statistical significance of the overlap,  $P<0.05$ .

**Quantitative real-time PCR (qRT-PCR) analysis.** The qRT-PCR assay was performed on 63 samples including the same 14 specimens used for the previous microarray analyses and 49 other specimens used for the validation (detailed clinical and pathological features are shown in Table I). Total RNA (1  $\mu$ g) was used for cDNA synthesis using random hexamer primers (RevertAid™ First Strand cDNA Synthesis Kit, Fermentas). Four differentially expressed genes, displaying an FC regulation ranging from -5.0 to 5.0, were selected and validated using an ABI PRISM 7000 sequence detector system (Applied Biosystems, Germany) in combination with a SYBR Premix Ex Taq II (Takara, Dalian, China) and the following sequence-specific primers (Invitrogen, Shanghai, China): DDIT4, forward 5'-GAG TCC CTG GAC AGC AGC AA-3' and reverse 5'-GGG TCA CTG AGC AGC TCG AA-3'; CXCL10, forward 5'-GGC CAT CAA GAA TTT ACT GAA AGC A-3' and reverse 5'-TCT GTG TGG TCC ATC CTT GGA A-3'; FOXQ1, forward 5'-GGC AAC GGG CTA CAG CTT TA-3' and reverse 5'-GGC ACC CCA CAT ACA TAA TCA A-3'; FOXM1, forward 5'-GCT TGC CAG AGT CCT TTT TGC-3' and reverse 5'-TTC TTG CAG GAA AGC TGA CTT G-3'; GAPDH, forward 5'-GCA CCG TCA AGG CTG AGA AC-3' and reverse 5'-TGG TGA AGA CGC CAG TGG A-3'. GAPDH was used as the internal control. Total RNA extractions from different specimens were used for the reverse transcription-PCR (RT-PCR), and 2  $\mu$ l of cDNA were then used for each 20- $\mu$ l qRT-PCR reaction. A total of 40 amplification cycles were performed after a denaturation



of 5 min at 95°C. Each cycle consisted of 30 sec at 94°C and 30 sec at 60°C. qRT-PCR analyses were performed in triplicate in 96-well plates. For each series, a threshold cycle value (Ct) was determined corresponding to the PCR cycle at which the fluorescent emission reached a threshold above the baseline emission, and then the same threshold was used for the comparison of the Ct values from the different experiments. The mean Ct value from each sample was normalized to the corresponding GAPDH Ct values, calculated as  $(Ct_{\text{experimental gene}} - Ct_{\text{GAPDH}})$ . The relative gene expression in a particular sample was then given by the following: Relative quantification or FC for each gene =  $2^{-\Delta\Delta CT}$  value, followed by the calculation of the relative FC of the 4-gene set in CRC and CRA compared to NOR. Dissociation curve analyses were also performed to confirm the specificity of the amplified PCR products.

**IHC analysis.** The expression patterns of 4 proteins corresponding to the selected 4-gene set were examined in an additional 174 samples, which contained 50 CRA, 76 CRC and 48 paired NOR tissues (detailed clinical and pathological features are shown in Table I), in order to further validate the consistency of the results.

Sections of paraffin-embedded specimens were deparaffinized in xylene and dehydrated through graded concentrations of ethanol, and antigen retrieval was performed by microwaving the samples in 0.01 mmol/l citric acid buffer (pH 6.0; Maixin-Bio, MVS-0100). Then, endogenous peroxidase was inhibited by immersing the sections in 3% hydrogen peroxide/methanol solution for 10 min. The slides were incubated in 100 ml/l non-immune goat serum (Maixin-Bio, SP KIT-B2) for 30 min to reduce background staining. The primary antibodies, rabbit polyclonal DDIT4 (Abcam, ab63059), rabbit polyclonal IP10/CXCL10 (Abcam, ab9807), rabbit polyclonal FOXQ1 (Abcam, ab51340) and mouse monoclonal FOXM1 (Abcam, ab5506), were each diluted 1:100 and incubated in a humidified chamber overnight at 4°C. After washing, the Ultra Sensitive S-P (Mouse/Rabbit) kit with biotin-labeled secondary antibody was used according to the manufacturer's instructions. The immunoreactivity was then revealed by DAB staining (DAB kit, Maixin-Bio, DAB-0031). The slides were then dipped in hematoxylin (Maixin-Bio, CTS-1099), dehydrated and mounted. Negative controls were included using negative antiserum as the first layer, with no first antibody.

The sections were photographed using an Olympus BX50 light microscope equipped with an HMIAS-2000 high distinct color medical imaging system (Qianping, Wuhan, China). The sum of the intensity and extension was then scored according to the method reported by Hao *et al.* (30): i) The intensity of staining was graded on a scale of 0-3 (0, negative; 1, weak; 2, moderate and 3, intense), and ii) the area of staining was evaluated as a percentage (0, 0-5; 1, 6-25; 2, 26-50; 3, 51-75 and 4, 76-100%). These 2 scores were multiplied to yield a final staining score. In addition, cellular localization (nuclear, cytoplasmic and cell surface) was assessed. All the sections were scored blindly by 2 independent investigators. The percentages of the samples with positive signals (IHC scores of 5-12) were obtained in the NOR, CRA and CRC tissues. Comparisons between the groups were performed by a 2-tailed

unpaired t-test with SPSS 11.0 software. A value of  $P < 0.05$  was considered statistically significant.

## Results

**Gene expression profiling: Class comparison of CRC vs. CRA and CRA vs. NOR.** A supervised analysis of discriminating genes between CRC and CRA identified a signature of 742 pssets ( $\geq 2$  FC, t-test with false discovery rate yielded a P-value of  $< 0.05$ ), which could probe into the mechanism of the CRA to CRC transition. At the same time, in order to identify which genes within the 742-psset list were already anomalously expressed during the NOR to CRA process, another independent analysis was performed to compare the gene expression profile of CRA with NOR. This allowed for the identification of the genes participating in the progression from NOR into CRA tissue, and we identified a list of 2950 pssets ( $\geq 2$  FC, P-value of  $< 0.05$  by t-test with false discovery rate; Fig. 1A and B).

**Evidence of progression from CRA to CRC.** We investigated the evidence for the NOR→CRA→CRC progression in order to identify an intermediate gene expression signature that could be used to predict the higher-risk CRA patients with the highest potential for developing CRC. We performed a statistical analysis following the analytical scheme depicted in Fig. 1B. Intersectional analyses of the 742 and 2950 pssets were performed and a signature of a 463-psset list was identified, which was coincidental in both lists. Within the 463-psset list, the distributions of up- and down-regulated events were 66 (305 pssets) and 34% (158 pssets), respectively. A list of these 305 up- and 158 down-regulated pssets is shown in Table II. The results of the two-dimensional hierarchical clustering analysis of this 463-psset list, are presented in Fig. 1C, indicating that this 463-psset list was able to accurately distinguish the CRC from the CRA class. More importantly, the results indicate that the onset and further tumor progression of CRC is a complex course involving numerous genes, which are already abnormally expressed at the pre-cancerous CRA stage. Thus, the malignant transformation from normal colorectal mucous membrane cells into CRC, already potentially exists in the pre-cancerous stage, and these genes could participate in the colorectal carcinogenesis processes. Of note, the CRC samples did not cluster according to their stages (tumor stages are reported in Fig. 1A and Table I). Therefore, the 463-psset signature marks the initial phases of the transition from NOR to CRA and then to CRC, and alterations in the expression of this intermediate gene signature could be used to predict the higher-risk CRA patients at the pre-cancerous stage.

As 2487 genes in the 2950 pssets (84.3%) did not overlap with 742 pssets, these 2487 pssets were specifically involved in the transition from NOR to CRA. These data suggest that most of the alterations in gene expression associated with the conversion from NOR to CRA are not retained in CRC, and that additional changes could occur later on, during tumor progression, differentiating CRA from CRC. This result was consistent with the epidemiological fact that CRAs have a much higher prevalence than CRCs, and that the majority of CRAs do not progress to become CRCs.

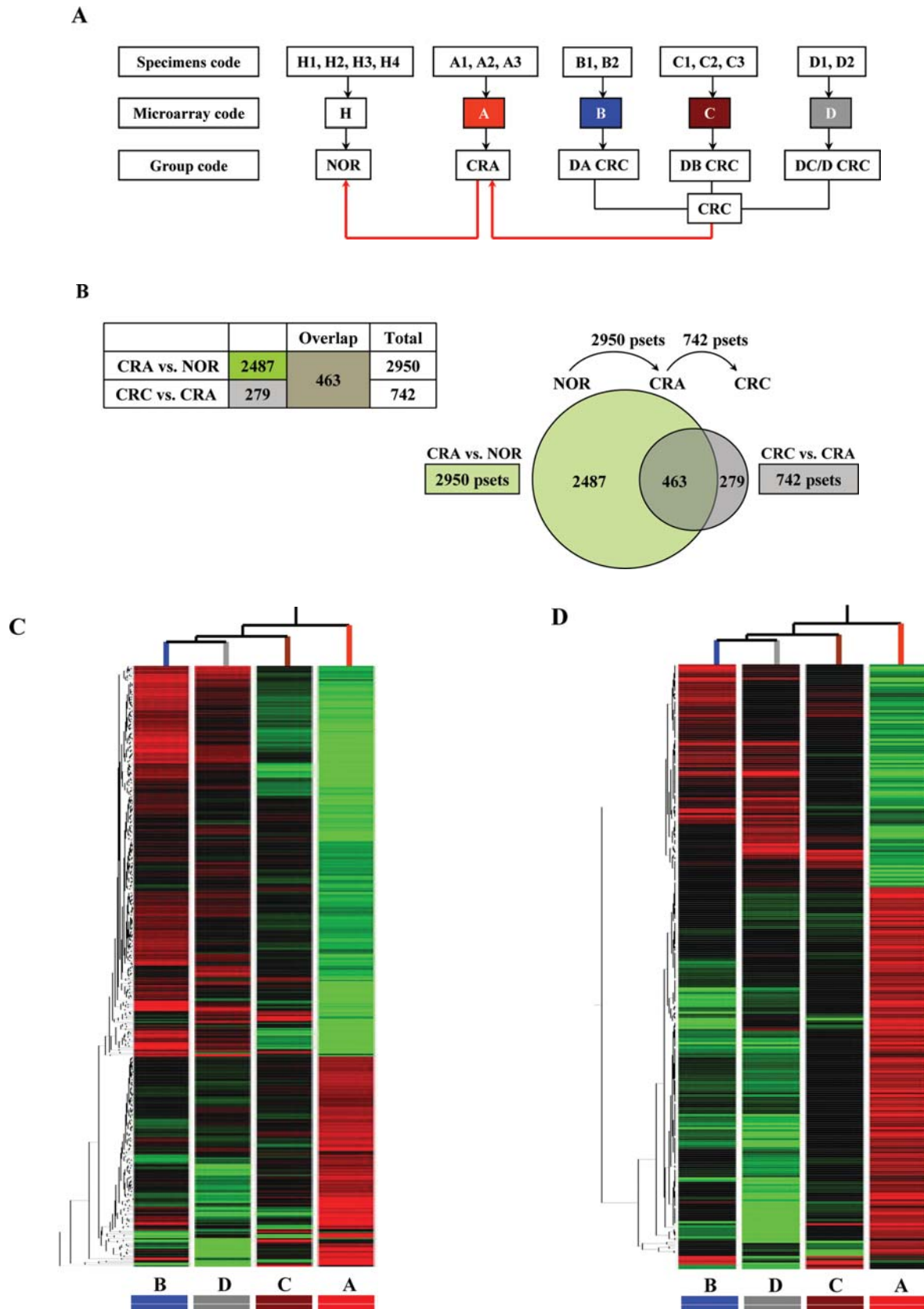


Figure 1. Microarray experimental design and corresponding statistical analysis of NOR, CRA and CRC samples. (A) Microarray experimental design. Specimen codes: H1-4, A1-3, B1-2, C1-3 and D1-2 corresponding to 14 specimens used in the study. Microarray code (group code): H (NOR), NOR; A (CRA), CRA; B (DA CRC), Dukes' A CRC; C (DB CRC), Dukes' B CRC; D (DC/D CRC), Dukes' C/D CRC. Group CRC consists of DA CRC, DB CRC and DC/D CRC. Red arrows show that gene expression comparisons were carried out between CRC and CRA, as well as between CRA and NOR. (B) Schematic representation and Venn diagram of the sequential statistical analysis performed on the NOR, CRA and CRC samples. Hierarchical clustering parameters of the 463 and 279 psets on 4 pooled samples. (C) Hierarchical clustering of the 463 psets on 4 samples. (D) Hierarchical clustering of the 279 psets on 4 samples. The 463- and 279-pset lists were able to distinguish CRA from the CRC class accurately. CRC samples did not cluster according to their stages. The rows show the differentially expressed genes. Red and green show relative high and low expression, respectively. Clusters, 1 cluster represents 1 microarray code. Microarray code (corresponding group code): Red, A (CRA); blue, B (DA CRC); brown, C (DB CRC); grey, D (DC/D CRC).

Table II. List of the up- and down-regulated genes of the 463-pset list in descending order of the mean fold change of the gene in the CRCs vs. CRAs.

#### Up-regulated psets

COL10A1, SPP1, SFRP2, TRPC1, OGN, MFAP5, LOC100128178, HNT, FOXP2, SFRP2, IFI44L, COL11A1, PRRX1, CXCL5, FABP4, H19, SFRP4, 233090\_at, THBS2, SERPINE1, GPM6B, RGS16, VIP, PLN, MAMDC2, CTHRC1, RGS1, COL11A1, EMP1, GAS1, GBP1, 226237\_at, CES1, ASPN, 226311\_at, 222877\_at, FN1, FN1, RGS1, PLN, IL8, CXCL13, BGN, ERAP2, FN1, ERAP2, FN1, DSEL, CYR61, COL10A1, FNDC1, TNFAIP6, 242397\_at, LOX, FN1, 227140\_at, HLA-DQA1, 229802\_at, BCL2A1, POSTN, IL1RN, GUCY1A3, AHNAK2, TPM2, POSTN, COL1A1, CTGF, COL1A1, LOC286167, ACTG2, CCDC80, BST2, ISG15, TNC, PCOLCE2, MEIS1, FLNA, TNS1, MYH11, CCL8, RSAD2, GPNMB, 243050\_at, COL1A2, PTPRR, COL5A2, ST6GALNAC5, GPC6, SULF1, SULF1, INHBA, LOC728190, TAGLN, VCAN, LOC727738, LY6E, NNMT, RAB31, LOC728264, SOCS3, 242881\_x\_at, 235821\_at, CYR61, FLNA, RAI14, TNFAIP6, CDKN2B, FBN1, APOE, VCAN, CXCL11, AKAP12, GREM1, ADAMTS1, SLC2A3, MAB21L2, IFIT1, COL5A2, BAG2, HBEGF, DPYSL3, 227462\_at, TIMP3, STC1, FCGR3A/FCGR3B, GPNMB, PCDH7, SULF1, CALD1, HSPA6, MYL9, COL1A2, CXCL10, CXCL11, MEIS1, SLC2A14/SLC2A3, EREG, VSIG4, LOC387763, IRS1, HBEGF, FSTL3, RAB31, MSR1, CCL3/CCL3L1/CCL3L3/LOC728830, TYMP, NNMT, TMEM45A, PLAUR, BGN, TIMP3, LOX, FBLN1, GPM6B, RGS16, IFI44, SCRNI, SPON1, VCAN, 228854\_at, FLNA, PLAUR, IL6, DUSP1, GEM, VCAN, FERMT2, SLC6A8, MFAP5, BCAT1, CPNE8, CALD1, LPHN2, PAPP, VCAN, DUSP5, 229530\_at, CCL21, AREG/LOC727738, 235419\_at, 235629\_at, DPT, H19, PTPRR, PMP22, TIMP3, MS4A4A, ROBO1, 228333\_at, FBLN1, TNFRSF12A, PTGS2, COL3A1, NR4A2, EMP1, VLDLR, CD14, RGS2, SGK1, FCGR1A, EIF2S3, CLIC4, CD109, CXCR4, PHLDA2, FCGR2A, S1PR3, COL3A1, MFGE8, 227235\_at, HIP1, SUCNR1, ARL4C, C5AR1, FOS, ERFFI1, C3, TDO2, MITF, 235392\_at, F13A1, DCN, CALD1, ROR1, ITGB2, EFEMP1, UPP1, DCN, ADH1B, FOXP2, PLAUR, UBE2E2, COL1A1, PRNP, SDS, SIRPA, G0S2, SRPX, RTN1, THBS1, DDIT4, PARVB, ATF3, IER3, SLC6A8, FCER1G, DCN, TSC22D3, KRT19, GBP1, 229549\_at, LDLR, 227827\_at, MXRA7, SLC2A3, IGFBP3, NR4A2, IFI16, NRP1, PLOD2, QSOX1, DCN, SLC20A1, RHOF, PLOD2, FCER1G, RHOB, LOC647190, SNF1LK, C2orf18, IFI16, TSPYL5, TPM4, PRNP, UBA6, SOX7, MS4A7, EPHA2, CLEC7A, PYGL, SLC20A1, IFI16, LATS2, FOSL2, LOC284454, C17orf91, 235735\_at, FPR3, ZFP36, SH2B3, NR4A2, NDN, CCL2, 239258\_at, TUBB6, CXCR4, FOXQ1, 217211\_at, RIN2, LAMA2, CXCL12, KLF2, SLC26A3, CXCL1, TACSTD2.

#### Down-regulated psets

DEFA6, DEFA5, PCSK1, C17orf92, RETNLB, RETNLB, KIAA1324, ENPP3, L1TD1, L1TD1, C17orf77, SH3GL2, 216050\_at, SLITRK6, TMED6, KIAA1324, LGR5, SPINK4, IGLV4-60, 1566698\_at, UGT1A6/UGT1A8/UGT1A9, C14orf53, FLJ30428, 239303\_at, ALDH3A1, 231250\_at, LGR5, ITFG2, 239624\_at, CFC1/CFC1B, ADH6, 233413\_at, SERPINA1, SCARNA2, PTC2, FZD10, C6orf117, FOXM1, ZG16, 1559538\_at, FLJ38028, QPCT, 235892\_at, 1556054\_at, SMAD9, IGFBP2, SERPINA1, 234158\_at, LOC283177, ENPP3, CHN2, ATOH8, CACNA2D2, 238632\_at, CTNNA2, 239723\_at, ACVR1C, CABP7, TBX3, PCLO, 229150\_at, SEMA5A, LOC606495, DUOX2, RORC, HOXB9, LOC285628, LOC730092, 1570005\_at, CRYM, REPS2, HOXA9, LOC643733, ABAT, 230119\_at, 217320\_at, 1560257\_at, AMY1A/AMY1B/AMY1C/AMY2A/AMY2B, CXorf10, ZDHHC8P, BEX2, LEFTY1, QPRT, SHANK2, ACVR1C, 1563331\_at, LOC100130097, 231979\_at, 230967\_s\_at, FGGY, 244383\_at, PITX2, EPHB2, RNF186, ZNF818, ENPP5, HOXB6, CYP39A1, CHDH, TMC5, SP5, PIWIL2, PLA2R1, SEMA5A, REPS2, MTSS1, C21orf57, 230088\_at, 238999\_at, PLCB4, SLC40A1, QPRT, C1orf161, YPEL1, HOXB5, EPHB2, 244878\_at, NPTX2, CPS1, DAPK2, MESPI, TLE2, MTL5, KIAA1244, DUSP18, COL9A2, PLCB4, FLJ44896, SLC40A1, TMC5, ETS2, THNSL1, LNX1, SEMA5A, ABAT, 216189\_at, 240045\_at, UNC5CL, UGT2B15, 1555846\_a\_at, TMC5, GCNT1, PKIB, FHDC1, MTERFD3, KIAA1984, 236963\_at, NSUN7, CLDN8, TBX3, LOC120376, 233377\_at, KIAA1147, CA1, BACE2, CEACAM7, 235274\_at, PAH.

In order to verify the consistency of the 463 psets in the majority of the individual samples, we also analyzed the expression trend of 4 differentially expressed genes by qRT-PCR: DDIT4, CXCL10, FOXQ1 and FOXM1 on a subset of 63 specimens. The average FC of the expression levels of the 4-gene set in CRC vs. CRA by qRT-PCR analysis compared to the microarray data, is shown in Fig. 3A. Together, gene profiling of qRT-PCR displayed excellent concordance with the data obtained by the microarray, suggesting that the gene expression signature of the 463 psets marked the progression of the CRA to CRC sequence, which made the discriminant differentiation of CRC from CRA reliable. In fact, qRT-PCR

was a more sensitive assay than microarray analysis, as the FC of the former was generally higher than the latter. DDIT4, CXCL10 and FOXQ1 displayed higher expression in CRC vs. NOR than in CRA vs. NOR, whereas FOXM1 displayed lower expression in CRC vs. NOR than in CRA vs. NOR (Fig. 3B). These data are not only in agreement with the overall profiles detected in the Affymetrix screening but also with the idea of a common origin of CRC and CRA, which could represent the progressive stages of the same tumor.

Furthermore, an independent assay using IHC in an additional 174 samples including CRC, CRA and NOR, was performed to further verify that the 4-gene set was indeed up-



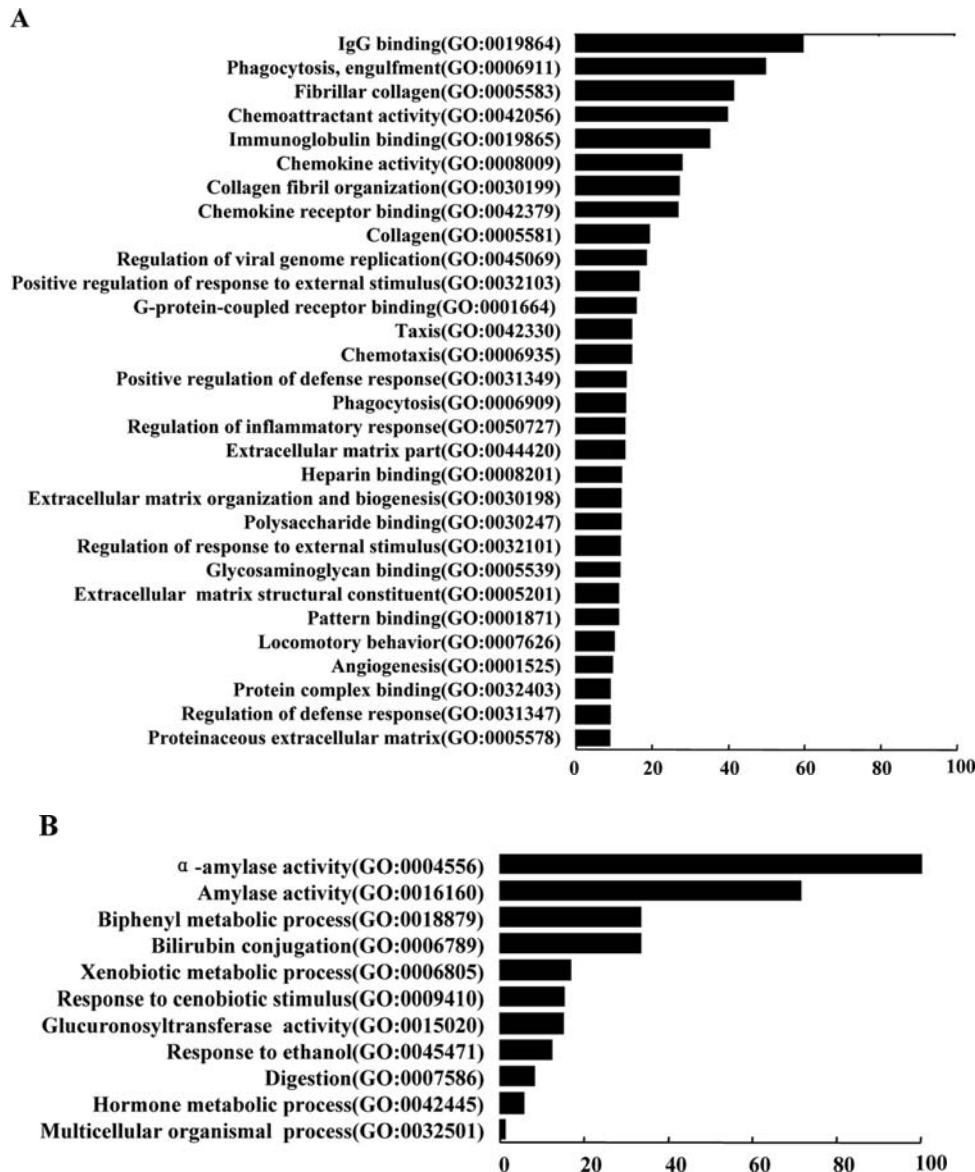


Figure 2. GO analysis of up- and down-regulated events of the 463 psets. (A and B) Top 30 of 108 and 11 GO categories involved in the up- and down-regulated events, respectively, that overlapped;  $P < 0.05$ . Left, GO categories; right, bar graph represents the psets proportion among the selected 305 and 158 psets, respectively.

or down-regulated at the protein level and was not an experimental artifact attributable to the microarray platform (Fig. 4). Of note, the microenvironment-associated genes, CXCL10 and FOXQ1, were significantly up-regulated in CRC compared to CRA ( $P < 0.05$ ) and NOR ( $P < 0.05$ ), and were only up-regulated in a low percentage at the CRA stage, indicating that their expressions could be specifically induced by the disease progress. Another member of the forkhead box (FOX) protein, FOXM1, was found to be significantly up-regulated in CRA compared to NOR ( $P < 0.05$ ), whereas it was significantly down-regulated in CRC compared to CRA ( $P < 0.05$ ). The change in expression of the DDIT4 protein by IHC was consistent with the microarray and qRT-PCR data, whereas the numbers of positive samples for the 3 comparison groups were too similar to obtain a meaningful analysis by IHC. Overall, different sensitivities of the 3 methods used in this study were considered. The IHC results indicated excellent

concordance with the data obtained by the microarray as well as by qRT-PCR, and further confirmed the reliability of the microarray data.

Together, our data suggest that the potential progression from CRA to CRC already exists at the pre-cancerous CRA stage. The mucosa of the patients was therefore primed for tumorigenesis at the pre-cancerous lesion stage, and additional changes could occur later on, during tumor progression, differentiating CRC from CRA, which are represented by a CRC-specific signature of the 279 psets (37.6% of 742 psets; Fig. 1B). Furthermore, two-dimensional hierarchical cluster analysis of the 279-pset list shown in Fig. 1D indicated that this list was also able to distinguish the CRA from the CRC class, whereas CRC samples did not cluster according to stage.

*Functional classification analysis of differentially expressed genes in the gene expression signature of 463 psets.* In order

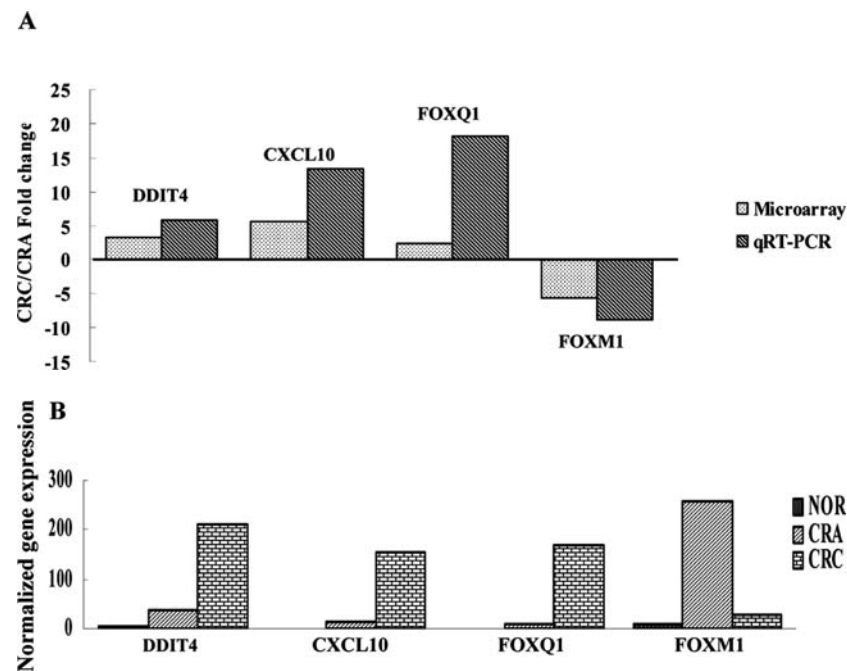


Figure 3. (A) Expression patterns of the 4-gene set in the CRC compared to the CRA specimens by using qRT-PCR, compared with microarray data. The average FC is the mean value of all the specimens used. (B) Relative FCs in the 4-gene set in the CRC and CRA compared to the NOR specimens by using qRT-PCR.

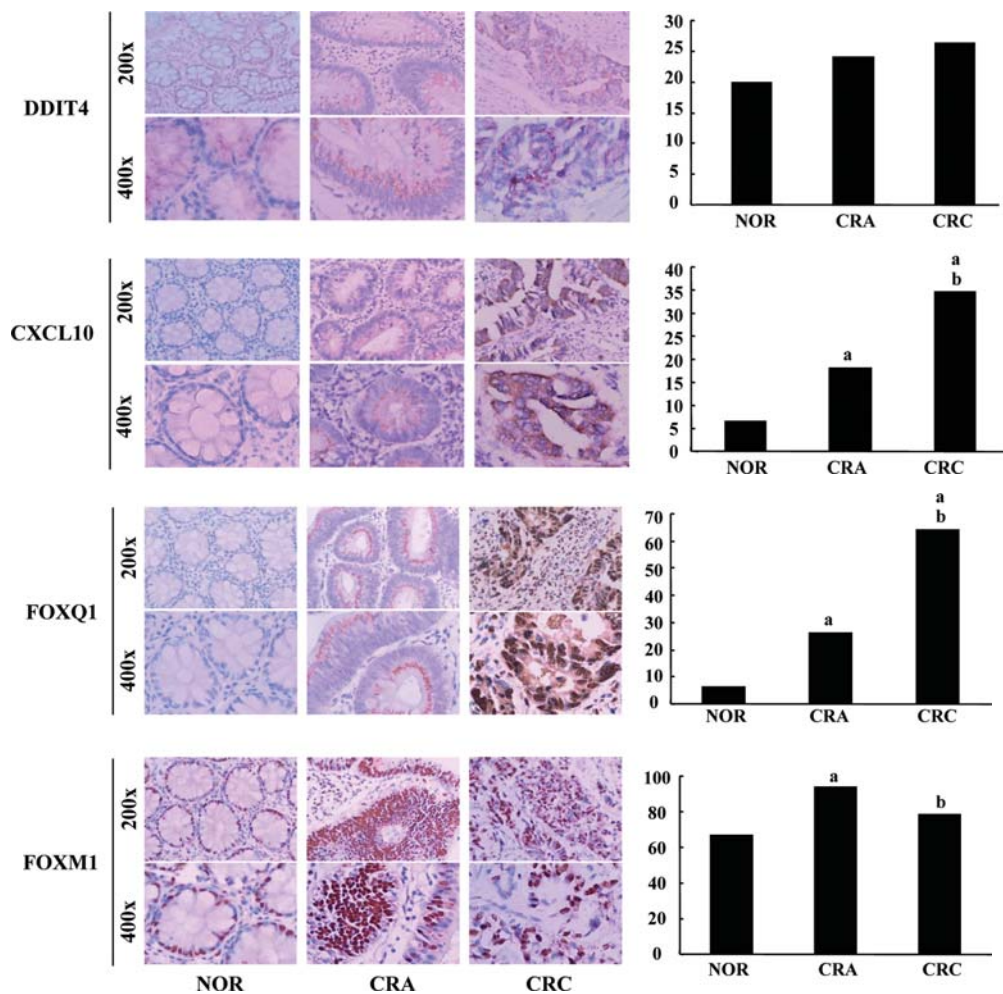


Figure 4. IHC analysis of the 4-gene set in NOR, CRA and CRC. Left, IHC images of the cytoplasm-positive (DDIT4 and CXCL10) and nuclei-positive (FOXQ1 and FOXM1) genes in the NOR, CRA and CRC samples at 200x and 400x magnifications. Right, bar graph representation of the percentage of samples in which a positive signal (IHC scores of 5-12) was observed in NOR, CRA and CRC. An ‘a’ indicates the statistical significance ( $P<0.05$ ) of differential expression in CRA vs. NOR, as well as in CRC vs. NOR, and a ‘b’ indicates the statistical significance ( $P<0.05$ ) of differential expression in CRC vs. CRA.



to further understand the biological functions of differentially expressed genes that make up the gene expression signature of the 463 psets, GO category analysis was performed on the up- and down-regulated events of the 463 psets, which corresponded to the 305 and 158 psets. The results indicated that 108 GO categories were involved in up-regulated events with a P-value of  $<0.05$  (the top 30 of these 108 GO categories are shown in Fig. 2A), which were enriched for genes involved in immunoglobulin and receptor binding, signal transduction, collagen fibril organization, regulation of defense response, response to wounds, regulation of cell motion and migration, angiogenesis, proteinaceous extracellular matrix, angiogenesis and chemokine activity. Conversely, the GO category enrichment of the down-regulated 158 psets showed that there were 11 GO categories involved in down-regulated events with a P-value of  $<0.05$  (Fig. 2B). These genes are involved in digestion, hormone, as well as biphenyl and xenobiotic metabolic processes. The activities of amylase and glucuronosyltransferase were down-regulated, pointing to a gradual loss of normal colorectal function with the disease progression.

## Discussion

Treating the pre-cancerous or early stage of CRC not only has a wider range of therapeutic options, but also a higher success rate of therapy compared to the advanced stages of CRC. At the same time, many long-term, follow-up clinical studies support the notion that only a small proportion of CRAs progress to CRCs. However, there are no reliable gene expression profile-based criteria available that can predict the proportion of higher-risk CRA patients before malignancy. The diagnostic standard currently available for higher-risk CRA detection is pathomorphism (13,14), but this has poor sensitivity and specificity, thus greatly limiting diagnostic accuracy. Here, we report that the signature of the 463 psets marks the initial phases of the NOR→CRA→CRC transition, and that the detection of the gene expression trends of this 463-pset signature could potentially be used for the screening of higher-risk CRA at the benign adenoma stage.

Often, the biopsy sample taken during the endoscopic intervention cannot be evaluated properly by conventional histology. Diagnostic expression profiling of the whole biopsy specimen could overcome this shortcoming.

Several long-term, follow-up clinical studies support the notion that human CRC is a multistep and multifactorial process (10,11). The progression of CRA to CRC has been clinically documented (5). The results of our expression profile analysis are in agreement with the clinical evidence and further show that, although CRA and CRC are molecularly distinct, they represent temporal variations in the natural history of the CRA to CRC progression. We initially identified a signature of 742 psets, characterizing the transition from CRA to CRC. Importantly, most of this signature (463 psets, 62.4% of 742 psets; Fig. 1B) was already anomalously expressed at the transformation stage of normal colorectal mucous membrane into CRA, and the 463 psets were present in the 2950 psets, representing the NOR to CRA progression (Fig. 1B). Therefore, this 463-pset signature points to the initial phases of the NOR→CRA→CRC progression. Addi-

tionally, this 463-pset signature was able to discriminate CRC from CRA by two-way hierarchical clustering (Fig. 1C). In summary, our data suggest that the potential malignant transformation from normal colorectal mucous membrane into CRC already exists at the pre-cancerous CRA stage. Therefore, mucosa of the patients is primed for tumorigenesis at this pre-cancerous stage, and additional changes could occur later on, during tumor progression, differentiating CRC from CRA, which are represented by the 279-pset signature (37.6% of 742 psets) identified (Fig. 1B).

Regarding the nature of the gene expression changes associated with the transition from normal to CRA to CRC, clues could be derived from a functional classification based on GO analysis. In the 463-pset signature, which represents this process, 66% (305 psets) were up-regulated and 34% (158 psets) were down-regulated. Up-regulated genes/GO categories were significantly, and not unexpectedly, enriched for those genes related to the stress/defense response, the response to wounding, cell motility, cytokine activity, angiogenesis and the extracellular matrix. This underscores the important contribution of the tumor microenvironment to the tumorigenesis and metastasis processes. It is well known that the tumor microenvironment is a mixture of extracellular matrix molecules, tumor cells, endothelial cells, fibroblasts and immune cells essential for tumor cell proliferation, angiogenesis, invasion and metastasis through the provision of survival signals, secretion of growth and pro-angiogenic factors, and direct adhesion molecule interactions (31). This notion was further reinforced by the IHC analysis of CXCL10 [chemokine (C-X-C motif) ligand 10], a well-known immune and defense response-related gene, which was found to gradually increase along with the NOR→CRA→CRC progression. This indicates that its expression can be specifically induced as the disease progresses. Previous studies have shown that CRC cells can produce CXCL10 (31) and that CXCL10 functions as a major chemoattractant for activated T cells and natural killer cells (33,34). Various studies have strongly suggested a role for CXCL10 in the regulation of angiogenesis activity in tumorigenesis (34), as well as in direct metastasis-promoting activities in CRC cells, and its influence in the tumor microenvironment (35). The importance of the cell cycle and cell proliferation are also underscored by the IHC results of FOXQ1 and FOXM1, 2 members of the FOX superfamily, which are known to be evolutionarily conserved transcriptional regulators and are involved in the control of a wide spectrum of biological processes. Furthermore, accumulating evidence indicates that FOXQ1 and FOXM1 are candidate oncogenes (36), although the precise mechanisms involved in CRC tumorigenesis remain to be elucidated. Furthermore, the undiscovered roles of FOXQ1 and FOXM1, as well as the regulation of FOXQ1 and FOXM1, also hold many remaining mysteries. Our IHC results for FOXQ1 were consistent with a previous study which found that FOXQ1 was up-regulated in CRC (37). Analogously, our results also show that the expression level of FOXM1 is increased both in CRC and CRA compared to NOR. The FOXM1 overexpression is essential for the development of various cancers (38,39) ranging from pancreatic (40), breast (41), non-small cell lung cancer (42) and basal cell carcinoma (43), to colorectal cancer (44). FOXM1

was found to be more intensively expressed in CRA than in CRC ( $P < 0.05$ ). Although the mechanism responsible for the FOXM1 gene function in tumorigenesis has not yet been fully clarified, increasing evidence indicates that the high expression levels of FOXM1 in tumor cells are more than a mere consequence of high proliferation rates, and that FOXM1 also actively participates in tumor development through its proliferation-stimulating activity (45). This could account for the higher FOXM1 levels in CRA compared to CRC, which could be related to the highly active participation of FOXM1 in tumor development via its proliferation-stimulating activity and disordered cell proliferation regulation during the pre-cancerous lesions.

A concern in GO analysis is the enrichment of extra-cellular matrix-related genes. In particular, the collagens, COL10A1 and COL11A1, were specifically identified in the 463-pset signature (COL11A1 was also found in the 279-pset signature). The importance of collagens was emphasized in a study identifying COL11A1 as a member of a metastasis-associated gene signature in both squamous cell carcinomas and adenocarcinomas of non-small cell lung cancer (46). The overexpression of COL11A1 was also reportedly correlated with lymph node metastasis and poor prognosis of patients with non-small cell lung cancer (47). The overexpression of COL10A1 was recently found in breast cancer (48), indicating the important participation of these particular collagens in tumorigenesis. In this study, COL10A1 and COL11A1 were significantly up-regulated and topped the list of the 20 most up-regulated genes in the comparison of CRC to CRA. This suggests that COL10A and COL11A could play significant roles in the progression of CRA to CRC.

Conversely, the down-regulated events in the NOR→CRA→CRC progression (158 psets, 34% of the 463-pset list) were affected in many cases. Genes/GO categories were implicated in specific functions of the colorectal mucosa (e.g. terminal digestion of the intermediate product of the hormone, biphenyl and xenobiotic metabolic processes and reduced activities of amylase and glucuronosyltransferase). This is consistent with the loss of normal colorectal function and differentiation as the disease progresses.

In conclusion, the salient finding from our study was the identification of the 463 psets as an intermediate signature that marks the initial phases of the transition from NOR→CRA→CRC. This represents a reservoir of candidate markers for higher-risk CRA. This signature could facilitate the early diagnosis of higher-risk CRA patients and enable timely and more selective therapeutic intervention. The length, duration and intensity of the patient follow-up could be determined by the predicted outcome of the gene expression trend of this 463-pset signature. Furthermore, some of the gene products of the CRC-specific signature (279 psets) could represent potential therapeutic targets for this malignancy. Although our study was limited by the small number of samples (14 patients), we used pooled RNA, which is thought to more accurately reflect the gene expression profile and yield more reproducible data compared to using individual samples (49,50). Therefore, before the 463-pset signature can be applied to a clinical setting, an independent set of samples from a larger number of patients is required to further confirm its validity. Finally, GO enrichment analysis of the 463-pset

signature, which represents the progression from NOR→CRA→CRC, could provide directions for further research into the molecular mechanism of the CRA to CRC sequence.

## Acknowledgements

This study was supported by grants from the Social Development of Yunnan Province, China (nos. 2006SG08 and 2009CA009), and the Science and Technology Department of Yunnan Province, China (no. 2008CD203).

## References

- O'Connell JB, Maggard MA and Ko CY: Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J Natl Cancer Inst* 96: 1420-1425, 2004.
- Boyle P and Ferlay J: Cancer incidence and mortality in Europe, 2004. *Ann Oncol* 16: 481-488, 2005.
- Sanjoaquin MA, Choodari-Oskoei B, Dolbear C, *et al*: Colorectal cancer incidence, mortality and survival in South-east England between 1972 and 2001. *Eur J Cancer Prev* 16: 10-16, 2007.
- Li M and Gu J: Changing patterns of colorectal cancer in China over a period of 20 years. *World J Gastroenterol* 11: 4685-4688, 2005.
- Leslie A, Carey FA, Pratt NR and Steele RJC: The colorectal adenoma-carcinoma sequence. *Br J Surg* 89: 845-860, 2002.
- Jeremy JR: Limitations of the adenoma-carcinoma sequence in colorectum. *Clin Cancer Res* 10: 5969-5970, 2004.
- Wada R: Proposal of a New hypothesis on the development of colorectal epithelial neoplasia: nonspecific inflammation-colorectal Paneth cell metaplasia-colorectal epithelial neoplasia. *Digestion* 79: 9-12, 2009.
- Snover DC, Jass JR, Fenoglio-Preiser C and Batts KP: Serrated polyps of the large intestine: a morphologic and molecular review of an evolving concept. *Am J Clin Pathol* 124: 380-391, 2005.
- Lee S, Bang S, Song K and Lee I: Differential expression in normal-adenoma-carcinoma sequence suggests complex molecular carcinogenesis in colon. *Oncol Rep* 16: 747-754, 2006.
- Song SY, Kim YH, Yu MK, Kim JH, Lee JM, Son HJ, Rhee PL, Kim JJ, Paik SW and Rhee JC: Comparison of malignant potential between serrated adenomas and traditional adenomas. *J Gastroenterol Hepatol* 22: 1786-1790, 2007.
- Kronborg O and Fenger C: Clinical evidence for the adenoma-carcinoma sequence. *Eur J Cancer Prev* 8 (Suppl 1): 73-86, 1999.
- Smith D, Ballal M, Hodder R, Selvachandran SN and Cade D: The adenoma carcinoma sequence: an indoctrinated model for tumorigenesis, but is it always a clinical reality? *Colorectal Dis* 8: 296-301, 2006.
- Qasim A, Muldoon C and McKiernan S: Colonic adenoma patients have higher incidence of hyperplastic polyps on surveillance colonoscopy. *Eur J Gastroenterol Hepatol* 21: 877-881, 2009.
- Franzén LE, Hahn-Strömberg V, Edvardsson H and Bodin L: Characterization of colon carcinoma growth pattern by computerized morphometry: Definition of a complexity index. *Int J Mol Med* 22: 465-472, 2008.
- Kwon HC, Kim SH, Roh MS, Kim JS, Lee HS, Choi HJ, Jeong JS, Kim HJ and Hwang TH: Gene expression profiling in lymph node-positive and lymph node-negative colorectal cancer. *Dis Colon Rectum* 47: 141-152, 2004.
- Agrawal D, Chen T, Irby R, Quackenbush J, Chambers AF, Szabo M, Cantor A, Coppola D and Yeatman TJ: Osteopontin identified as colon cancer tumor progression marker. *C R Biol* 326: 1041-1043, 2003.
- Bandres E, Catalan V, Sola I, Honorato B, Cubedo E, Cordeu L, Andion E, Escalada A, Zarate R, Salgado E, Zabalegui N, García F and Garcia-Foncillas J: Dysregulation of apoptosis is a major mechanism in the lymph node involvement in colorectal carcinoma. *Oncol Rep* 12: 287-292, 2004.
- Bertucci F, Salas S, Eysteries S, Nasser V, Finetti P, Ginestier C, Charafe-Jauffret E, Llorio B, Bachelart L, Montfort J, Victorero G, Viret F, Ollendorff V, Fert V, Giovannini M, Delpero JR, Nguyen C, Viens P, Monges G, Birnbaum D and Houlgate R: Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. *Oncogene* 23: 1377-1391, 2004.

19. Birkenkamp-Demtroder K, Christensen LL, Olesen SH, Frederiksen CM, Laiho P, Aaltonen LA, Laurberg S, Sørensen FB, Hagemann R and Ørntoft TF: Gene expression in colorectal cancer. *Cancer Res* 62: 4352-4363, 2002.
20. Williams NS, Gaynor RB, Scoggin S, Verma U, Gokaslan T, Simmang C, Fleming J, Tavana D, Frenkel E and Becerra C: Identification and validation of genes involved in the pathogenesis of colorectal cancer using cDNA microarrays and RNA interference. *Clin Cancer Res* 9: 931-946, 2003.
21. Li M, Lin YM, Hasegawa S, Shimokawa T, Murata K, Kamayama M, Ishikawa O, Katagiri T, Tsunoda T, Nakamura Y and Furukawa Y: Genes associated with liver metastasis of colon cancer, identified by genome-wide cDNA microarray. *Int J Oncol* 24: 305-312, 2004.
22. Lin YM, Furukawa Y, Tsunoda T, Yue CT, Yang KC and Nakamura Y: Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. *Oncogene* 21: 4120-4128, 2002.
23. Yanagawa R, Furukawa Y, Tsunoda T, Kitahara O, Kameyama M, Murata K, Ishikawa O and Nakamura Y: Genome-wide screening of genes showing altered expression in liver metastases of human colorectal cancers by cDNA microarray. *Neoplasia* 3: 395-401, 2001.
24. Frederiksen CM, Knudsen S, Laurberg S and Ørntoft TF: Classification of Dukes' B and C colorectal cancers using expression arrays. *J Cancer Res Clin Oncol* 129: 263-271, 2003.
25. Fan J, Peng Z, Zhou C, Qiu G, Tang H, Sun Y, Wang X, Li Q, Le X and Xie K: Gene-expression profiling in Chinese patients with colon cancer by coupling experimental and bioinformatic genomewide gene-expression analyses: identification and validation of IFITM3 as a biomarker of early colon carcinogenesis. *Cancer* 113: 266-275, 2008.
26. Bianchini M, Levy E, Zucchini C, Pinski V, Macagno C, De Sanctis P, Valvassori L, Carinci P and Mordoh J: Comparative study of gene expression by cDNA microarray in human colorectal cancer tissues and normal mucosa. *Int J Oncol* 29: 83-94, 2006.
27. Kleivi K, Lind GE, Diep CB, Meling GI, Brandal LT, Nesland JM, Myklebost O, Rognum TO, Giercksky KE, Skotheim RI and Lothe RA: Gene expression profiles of primary colorectal carcinomas, liver metastases, and carcinomatoses. *Mol Cancer* 6: 2, 2007.
28. Galamb O, Sipos F, Solymosi N, Spisák S, Krenács T, Tóth K, Tulassay Z and Molnár B: Diagnostic mRNA expression patterns of inflamed, benign, and malignant colorectal biopsy specimen and their correlation with peripheral blood results. *Cancer Epidemiol Biomarkers Prev* 17: 2835-2845, 2008.
29. Sabates-Bellver J, Van der Flier LG, de Palo M, Cattaneo E, Maake C, Rehauer H, Laczo E, Kurowski MA, Bujnicki JM, Menigatti M, Luz J, Ranalli TV, Gomes V, Pastorelli A, Faggiani R, Anti M, Jiricny J, Clevers H and Marra G: Transcriptome profile of human colorectal adenomas. *Mol Cancer Res* 5: 1263-1275, 2007.
30. Hao XP, Willis JE, Pretlow TG, Rao JS, MacLennan GT, Talbot IC and Pretlow TP: Loss of fragile histidine triad expression in colorectal carcinomas and premalignant lesions. *Cancer Res* 60: 18-21, 2000.
31. Nyberg P, Salo T and Kalluri R: Tumor microenvironment and angiogenesis. *Front Biosci* 13: 6537-6553, 2008.
32. Yang CK, Chen DD, Tian Y and Zhang JH: Chemokine IP-10 produced by colorectal carcinoma. *Shijie Huaren Xiaohua Zazhi* 11: 1703-1705, 2003.
33. Singh UP, Singh S, Iqbal N, Weaver CT, McGhee JR and Lillard JW Jr: IFN- $\gamma$ -inducible chemokines enhance adaptive immunity and colitis. *J Interferon Cytokine Res* 23: 591-600, 2003.
34. Lazzeri E and Romagnani P: CXCR3-binding chemokines: novel multifunctional therapeutic targets. *Curr Drug Targets Immune Endocr Metabol Disord* 5: 109-118, 2005.
35. Zipin-Roitman A, Meshel T, Sagi-Assif O, Shalmon B, Avivi C, Pfeffer RM, Witz IP and Ben-Baruch A: CXCL10 promotes invasion-related properties in human colorectal carcinoma cells. *Cancer Res* 67: 3396-3405, 2007.
36. Myatt SS and Lam EW: The emerging roles of forkhead box (Fox) proteins in cancer. *Nat Rev Cancer* 7: 847-859, 2007.
37. Kaneda H, Arao T, Tanaka K, Tamura D, Aomatsu K, Kudo K, Sakai K, De Velasco MA, Matsumoto K, Fujita Y, Yamada Y, Tsurutani J, Okamoto I, Nakagawa K and Nishio K: FOXQ1 is overexpressed in colorectal cancer and enhances tumorigenicity and tumor growth. *Cancer Res* 70: 2053-2063, 2010.
38. Pilarsky C, Wenzig M, Specht T, Saeger HD and Grützmann R: Identification and validation of commonly overexpressed genes in solid tumors by comparison of microarray data. *Neoplasia* 6: 744-750, 2004.
39. Van den Boom J, Wolter M, Kuick R, Misek DE, Youkilis AS, Wechsler DS, Sommer C, Reifemberger G and Hanash SM: Characterization of gene expression profiles associated with glioma progression using oligonucleotide-based microarray analysis and real-time reverse transcription-polymerase chain reaction. *Am J Pathol* 163: 1033-1043, 2003.
40. Wang Z, Banerjee S, Kong D, Li Y and Sarkar FH: Down-regulation of Forkhead Box M1 transcription factor leads to the inhibition of invasion and angiogenesis of pancreatic cancer cells. *Cancer Res* 67: 8293-8300, 2007.
41. Francis RE, Myatt SS, Krol J, Hartman J, Peck B, McGovern UB, Wang J, Guest SK, Filipovic A, Gojis O, Palmieri C, Peston D, Shousha S, Yu Q, Sicinski P, Coombes RC and Lam EW: FoxM1 is a downstream target and marker of HER2 overexpression in breast cancer. *Int J Oncol* 35: 57-68, 2009.
42. Wang IC, Meliton L, Tretiakova M, Costa RH, Kalinichenko VV and Kalin TV: Transgenic expression of the forkhead box M1 transcription factor induces formation of lung tumors. *Oncogene* 27: 4137-4149, 2008.
43. Teh MT, Wong ST, Neill GW, Ghali LR, Philpott MP and Quinn AG: FOXM1 is a downstream target of Gli1 in basal cell carcinomas. *Cancer Res* 62: 4773-4780, 2002.
44. Yoshida Y, Wang IC, Yoder HM, Davidson NO and Costa RH: The forkhead box M1 transcription factor contributes to the development and growth of mouse colorectal cancer. *Gastroenterology* 132: 1420-1431, 2007.
45. Laoukili J, Stahl M and Medema RH: FoxM1: At the crossroads of ageing and cancer. *Biochim Biophys Acta* 1775: 92-102, 2007.
46. Woenckhaus M, Klein-Hitpass L, Grepmeier U, Merk J, Pfeifer M, Wild P, Bettstetter M, Wuensch P, Blaszyk H, Hartmann A, Hofstaedter F and Dietmaier W: Smoking and cancer-related gene expression in bronchial epithelium and non-small-cell lung cancers. *J Pathol* 210: 192-204, 2006.
47. Chong IW, Chang MY, Chang HC, Yu YP, Sheu CC, Tsai JR, Hung JY, Chou SH, Tsai MS, Hwang JJ and Lin SR: Great potential of a panel of multiple hMTH1, SPD, ITGA11 and COL11A1 markers for diagnosis of patients with non-small cell lung cancer. *Oncol Rep* 16: 981-988, 2006.
48. Chang HJ, Yang MJ, Yang YH, Hou MF, Hsueh EJ and Lin SR: MMP13 is potentially a new tumor marker for breast cancer diagnosis. *Oncol Rep* 22: 1119-1127, 2009.
49. Helm J, Enkemann SA, Coppola D, Barthel JS, Kelley ST and Yeatman TJ: Dedifferentiation precedes invasion in the progression from Barrett's metaplasia to esophageal adenocarcinoma. *Clin Cancer Res* 11: 2478-2485, 2005.
50. Agrawal D, Chen T, Irby R, Quackenbush J, Chambers AF, Szabo M, Cantor A, Coppola D and Yeatman TJ: Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling. *J Natl Cancer Inst* 94: 513-521, 2002.