The gene expression profile represents the molecular nature of liver metastasis in colorectal cancer

MAKOTO YAMASAKI[1,2], ICHIRO TAKEMASA[1], TAKAMICHI KOMORI[1,2], SHOUJI WATANABE[2], MITSUGU SEKIMOTO[1], YUICHIRO DOKI[1], KENICHI MATSUBARA[2] and MORITO MONDEN[1]

[1]Department of Surgery and Clinical Oncology, Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0879; [2]DNA Chip Research Inc., 1-1-43 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan

Abstract. The major cause of death in colorectal cancer is related to liver metastasis. Although the metastatic process has been well studied, many aspects of the molecular genetic basis of metastasis remain unclear. Elucidation of the molecular nature of liver metastasis is urgent to improve the outcome of colorectal cancer. We analyzed the chronological gene expression profiles of 104 colorectal samples corresponding to oncogenic development including normal mucosa, localized and metastasized primary tumors, and liver metastatic lesions as fundamental samples using a custom cDNA microarray. The gene expression patterns in 104 samples were classified into four groups closely associated with their metastatic status, and the genes of each group appropriately reflected the metastatic process. To investigate the existence of metastatic potential in primary tumors using metastasis-related genes detected by chronological analysis, we performed a hierarchical cluster and supervised classification analysis of 28 independent primary tumors. Hierarchical cluster analysis segregated the tumors according to their final metastatic status, rather than their clinical stages, and the profile of metastasized primary tumors resembled one of a metastatic lesion apart from a primary lesion rather than one of a non-metastasized primary tumor. Using the supervised classification approach, the expression profile of these genes allowed the classification of tumors diagnosed as localized cancer into two classes, the localized and the metastasized class, according to their final metastatic status. The disease-free survival and overall survival were significantly longer in the localized class than the metastasized class. Chronological analysis of the gene expression profile provides a better understanding of the metastatic process. Our results suggest that the metastatic potential is already encoded in the primary tumor and is detectable by a gene expression profile, which allows the prediction of liver metastasis in patients diagnosed with localized tumors and also the design of new strategies for the treatment and diagnosis of colorectal cancer.

Introduction

Colorectal cancer (CRC) is one of the most common cancers worldwide and considerable progress has been made in identifying the mechanisms of tumorigenesis based on molecular and biological research (1). The number of completely cured patients has increased because of improvements in screening technology and localized treatment; however, this has not resulted in major improvements in the prognosis of patients with advanced cancers. The major cause of death in patients with CRC is metastasis, particularly liver metastasis (2). Despite the development of various treatment modalities, once a metastatic lesion is discovered, the outcome for the patient is unfavorable. Therefore, elucidation of the global nature of liver metastasis in CRC is clinically important (3).

The phenomenon of cancer metastasis has been extensively studied morphologically and has been characterized as a complex, multistep process, and each step of this process is regulated by various changes on a genetic level (4-6). Thus, to understand cancer metastasis, it is important to comprehensively analyze the alterations of the genes involved in this process (7,8). DNA microarray technology is an efficient approach to a wide range of applications in cancer research (9). Gene expression profiling can identify biologically and clinically important subgroups of malignant tumors such as CRC (10-12). Indeed, it is possible to mark the histopathological differentiation at a molecular level, and the analysis could help to explain the complexity of the disease

---

process and can more accurately assess prognosis and distinct behavior.

Conventional diagnosis and surveillance have been performed by histopathological and clinical parameters such as the TNM stage and serum tumor markers such as the carcinoembryonic antigen (CEA). An elevated serum CEA level is regarded as a good indicator of metastatic disease and post-operative testing is recommended for the survey of recurrence, whereas ~30% of all CRC recurrences do not produce CEA (13). Currently, the prediction of metastasis or recurrence is also achieved by molecular parameters that focus on individual candidate genes such as the vascular endothelial growth factor (VEGF) and deleted-in-colon carcinoma (*DCC*) genes. However, these molecular diagnoses are not always accurate in predicting metastasis in the individual patient, and have not yet been validated for use as a diagnostic tool applicable to clinical practice (14). Thus, there has been a tremendous amount of effort to diagnose recurrent CRC in clinical practice.

Several investigators have reported that the gene expression profile makes it possible to predict the outcome including metastasis (15-17). These reports suggest that the gene expression profile has the potential to be applied clinically for the diagnosis of metastasis, and that a more precise diagnosis of metastasis than the current one can be achieved based on a comprehensive analysis of each distinct step of metastasis at a molecular level.

The present study was designed to analyze the chronological changes in the gene expression profiles of 104 patients as the fundamental set using cDNA microarray (Colonochip), which was developed for the analysis of CRC (18). We identified the gene expression profiles that reflected each step of cancer progression, including carcinogenesis, metastatic potential and metastatic colonization. To validate the predictability of metastasis by the gene expression profile in the bulk of primary tumors, we focused on the genes that correlated with the metastatic potential in the fundamental set. We then profiled the gene expression of 28 independent primary tumors that were diagnosed as localized or regional diseases at the time of surgery and investigated whether the expression profile in the primary tumors could determine their metastatic potential.

Materials and methods

Tissue samples. Tissues from 12 normal colonic tissues, 12 stage I and 20 stage II primary tumors without metastasis (stage I-LOC and stage II-LOC), 7 stage III and 19 stage IV primary tumors with liver metastasis (stage III-MET and stage IV-MET), and 34 liver metastatic specimens (LIV) were obtained from February 1989 to August 2000 and analyzed as the fundamental sample set. The clinicopathological backgrounds of the patients are summarized in Table I. All samples were immediately frozen in liquid nitrogen after surgical resection and stored at -80˚C until further use. Each sample was examined histopathologically and all normal tissues were confirmed to be free of cancer. Samples were collected from patients who had given informed consent and the study was approved by the Ethics Committee of Osaka University.

cDNA microarray, tissue preparation, hybridization and data processing. We used the cDNA microarray 'Colonochip' specifically designed for analyzing CRC. The reliability of Colonochip had been confirmed previously by comparison with the data obtained from two distinct assays: Reverse transcription-polymerase chain reaction (RT-PCR) and serial analysis of the gene expression (SAGE) database (18,19). Tissue preparation, hybridization, washing, scanning and image analysis were performed as described previously (18). In each sample, the values of the Cy3/Cy5 ratio of each spot were log-transformed and normalized so that the median Cy3/Cy5 ratio of whole genes was 1.0. Genes with <85% valid data points in the sample sets were excluded from further analysis, and hence, 1,953 genes were selected at this stage.

Statistical analysis. We calculated the signal-to-noise (S2N) statistics to identify differentially expressed genes with regard to each class distinction [(for example, normal mucosa vs primary cancer and normal mucosa vs metastatic cancer) $(\grave{I}_0-\grave{I}_1)/(\hat{U}_0+\hat{U}_1)$]. The symbol $\grave{I}$ represents the mean and $\hat{U}$ represents the standard deviation of expression for each class of each gene in the dataset. To compare these correlations to what would be expected by chance, permutation of the sample labels was performed and the S2N statistics in each permutation were calculated. One thousand global random permutations were used to build histograms. The assumption was made that these histograms were normally distributed. Based on these histograms, the z-score and significance levels were determined and then compared with the values obtained for the real dataset (20). We defined genes with a 99 percentile set as significantly different.

Gene ontology analysis. GoMiner (http://discover.nci.nih.gov/gominer/) was used to investigate the biological significance of a set of genes represented by the specific expression pattern during the progression of cancer metastasis (21). We calculated whether the gene ontology (GO) category was enriched or depleted for each group of genes represented by the specific expression pattern with respect to what would have been expected by chance alone, using the two-sided Fisher's exact test. We focused our analysis on a GO term that had >10 associated genes in order to evaluate the significance of a GO term.

Independent samples. An additional 28 CRC tumors were profiled as an independent sample set; 10 stage II and 8 stage III primary tumors without metastasis (stage II-LOC and stage III-LOC), and 4 stage II and 6 stage III with metachronous liver metastasis (stage II-MET and stage III-MET). Table I shows the clinicopathological backgrounds of these patients.

Hierarchical cluster analysis. To investigate whether the expression profile in the primary tumor reflected metastatic status, we analyzed the expression profiles of 92 tumors in the fundamental set, using 119 genes that were regarded as relevant to the metastatic potential by chronological analysis. Furthermore, the expression profiles of 120 tumors (adding the 92 tumors to the 28 independent tumors) were analyzed

Table I. Clinicopathological characteristics of the profiled cancer samples.

	Fundamental samples				Independent samples			
Sample group	Stage I-LOC	Stage II-LOC	Stage III-MET	Stage IV-MET	Stage II-LOC	Stage II-MET	Stage III-LOC	Stage III-MET
Number	12	20	7	19	10	4	8	6
Liver metastatic status	Negative	Negative	Negative	Positive	Negative	Positive	Negative	Positive
Sex								
Male	6	11	4	12	6	2	6	5
Female	6	9	3	7	4	2	2	1
Age (years)								
(mean ±SD)	59.3±8.7	59.3±10.2	63.3±8.9	59.4±10.2	68.3±10.2	64.0±2.7	58.3±8	60.8±6.7
Location								
Right (A,C,T)	2	6	3	4	3	2	0	1
Left (D,S)	2	8	2	10	2	1	2	2
Rectum	8	6	2	5	5	1	6	3
Depth								
Sm, mp	12	0	2	1	0	0	2	0
Ss, a1	0	13	2	9	7	2	3	3
Se, a2	0	7	3	8	3	2	3	3
si, ai	0	0	0	1	0	0	0	0
Lymph node metastatic status								
(-)	12	20	0	10	10	4	0	0
(+)	0	0	7	9	0	0	8	6
Follow-up (month)								
Median	84.5	70.8	51.2	34	77.9	27.2	82.6	46.6
(Range)	(24.8-102)	(25.1-101)	(13.4-76.0)	(4.4-80.5)	(63-115)	(12.7-80.1)	(47.1-102)	(22-56.2)
Month to metastasis								
Median	-	-	(6.2-39.8)	0	-	(6.7-20.1)	-	(6.8-18.5)
(Range)	-	-	17.2	0	-	15.4	-	17

LOC, localized primary tumor; MET, metastasized primary tumor.

using the same genes. Hierarchical cluster analysis (HCA) was performed with Pearson's correlation coefficient as a similarity coefficient and the unweighted pair group method using arithmetic averages (UPGMA) as the clustering algorithm.

Classification analysis with independent samples. We applied the weighted vote method as the class predictor (20). The prediction of a new sample is based on the 'weighted votes' of 119 informative genes. Each such gene g_i votes for either LOC or MET, depending on whether its expression level x_i in the sample is closer to $\grave{I}_{LOC}$ or $\grave{I}_{MET}$ (which denote, respectively, the mean expression levels of LOC and MET in the fundamental set). The magnitude of the vote is $w_i v_i$, where w_i is the S2N statistics between LOC and MET $[|(\grave{I}_{LOC}-\grave{I}_{MET})/(\hat{U}_{LOC}+\hat{U}_{MET})|]$ and $v_i=|x_i-(\grave{I}_{LOC}+\grave{I}_{MET})/2|$. The votes for each class are summed to obtain the total number of votes V_{LOC} and V_{MET}. The sample is assigned to the class (localized signature or metastasized signature) with the higher vote total.

Results

The 1,953 genes were classified into four distinct groups according to the following expression patterns. Group A (A1 and A2) consisted of 369 genes that were differentially expressed during all tumorgenic stages; group B (B1 and B2) consisted of 119 genes that were differentially expressed in synchronously or metachronously metastasized tumors and liver metastasis; group C (C1 and C2) consisted of 430 genes that were differentially expressed in liver metastasis; group D consisted of 1,035 genes that were not characterized by the expression pattern at all steps of cancer development (Fig. 1). The expression patterns of the genes of group A, B and C correlated with the phenomena of cancer progression in CRC; carcinogenesis, metastatic potential and metastatic
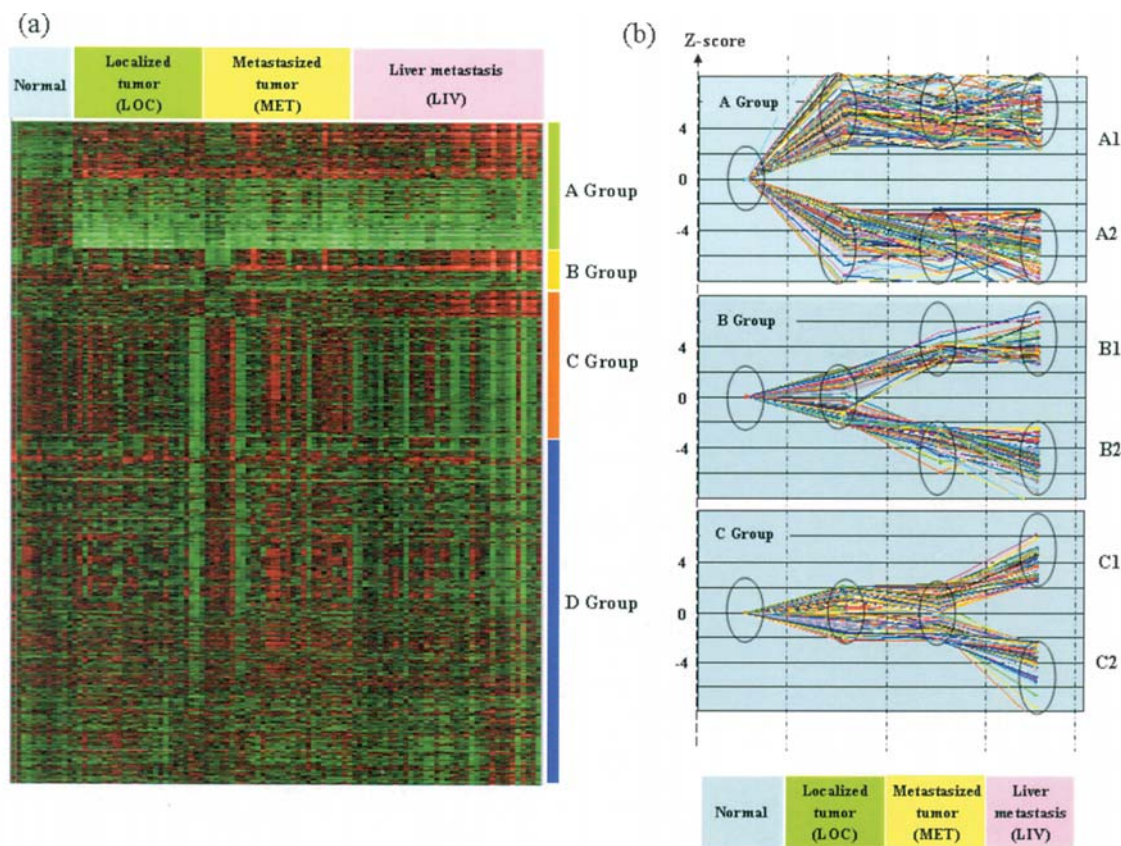
Figure 1. Gene expression patterns associated with the metastatic process. (a) The columns represent 12 normal colon samples and 92 colorectal tumor samples [32 localized tumors (LOC), 26 metastasized tumors (MET) and 34 liver metastatic tumors (LIV)]; the rows represent 1,953 genes. Differentially expressed genes during cancer development were identified by comparison of each sample group with 12 normal colon samples by permutation testing. (b) Line graphs represent the z-score calculated by permutation testing in each sample group at the stage of development. Group A (A1 and A2) consisted of 369 genes that were differentially expressed during the development of the carcinogenic process (A1, up-regulated genes; A2, down-regulated genes); group B (B1 and B2) consisted of 119 genes that were differentially expressed in localized tumors and metastasized tumors; group C (C1 and C2) consisted of 430 genes that were characterized by a differential expression pattern in liver metastasis.

colonization, respectively. The list of genes of group A, B or C are shown in Supplementary Table I, which is available on our website (http://www.dna-chip.co.jp/other/ other/ download_20060822.html).

Functional analysis based on GO was performed to translate the lists of differentially regulated genes in order to gain a better understanding of the underlying biological phenomenon. Table II shows the representative GO functional classes as upper-hierarchical terms in the final branch under the biological-process ontology with a significance of p<0.05 in each group, and all of the terms are shown in Supplementary Fig. 1, which is available on our website (http://www.dna-chip.co.jp/other/other/download_20060822.html), based on the directed acyclic graph structure of the GO and their hierarchy.

Group A1 included genes related to DNA metabolism, cell cycle, and response to endogenous stimuli such as response to DNA damage. Response to external stimuli such as defense and immune responses were represented in group A2. Genes related to apoptosis, cell motility and response to external stimuli were included in group B. Group C included genes involved in signal transduction such as the intracellular signaling cascade and cell-cell signaling, defense response and response to stress.

To investigate whether the expression profile in the bulk of primary tumors reflected metastatic status, we analyzed the gene expression profiles of 92 tumors in the fundamental set. HCA of 92 CRCs using the 119 genes in group B, which were detected by chronological analysis as genes closely associated with metastatic potential, was performed, and the tumors were segregated into two major categories, each consisting of 44 and 48 tumors, respectively (Fig. 2A). The left cluster (non-metastatic cluster) was comprised of 28 LOC, 10 MET and 6 LIV, whereas the right cluster (metastatic cluster) was comprised of 4 LOC, 16MET and 28 LIV. These two clusters correlated significantly with the metastatic status of the tumors.

Next, we performed HCA of 120 tumors (92 fundamental tumors and 18 independent ones) (Fig. 2B). The clustering pattern was robust against sample addition. For instance, the partitioning of the clusters and 92 fundamental tumors in the two sub-branches remained similar with little change.

As for the independent samples, 17 of the 18 stage II and III samples without metastasis belonged to the non-metastatic cluster despite of the clinical staging. Unexpectedly, all of the 10 stage II or III samples that metachronously developed liver metastasis although they had been diagnosed as localized or regional disease at the time of initial diagnosis belonged to the metastatic cluster. The expression profile of the independent samples was closely associated with the metastatic potential in the primary tumor rather than the histopathological staging.

Table II. Functional gene classes demonstrating significant differential expression in each group.

Function name	Total	P (X2)	P (X1)	P (X)
A group				
Regulation of transcription	189	0.998	0.008	0.581
Regulation of transcription, DNA-dependent	184	0.997	0.011	0.582
Response to external stimuli	135	0.008	0.932	0.14
Mitotic cell cycle	48	0.628	1E-04	0.003
Ion transport	46	0.057	0.214	0.026
Regulation of cell cycle	46	0.901	0.001	0.055
Lipid metabolism	36	0.013	0.844	0.082
Protein folding	22	0.926	0.038	0.261
Alcohol metabolism	20	1	0.026	0.365
DNA replication and chromosome cycle	20	0.668	0.001	0.009
Response to endogenous stimulus	20	1	0.026	0.365
DNA replication	18	0.609	0.003	0.015
Nuclear organization and biogenesis	18	1	0.016	0.278
DNA repair	17	1	0.013	0.236
Main pathways of carbohydrate metabolism	14	0.808	0.029	0.126
Nucleocytoplasmic transport	14	1	0.029	0.297
DNA packaging	13	1	0.004	0.096
Carbohydrate catabolism	11	1	0.012	0.157
B group				
Transport	170	0.041	0.617	0.07
Response to external stimuli	135	0.019	0.219	0.01
Electron transport	56	0.026	0.65	0.045
Apoptosis	44	0.109	0.187	0.043
Energy pathways	37	0.003	0.497	0.005
Cell motility	36	0.455	0.024	0.06
Induction of programmed cell death	18	0.037	0.04	0.003
C group				
Protein metabolism	340	0.002	0.895	0.009
Biosynthesis	190	0.041	0.99	0.181
Response to stress	96	0.045	0.076	0.01
Defense response	85	0.642	0.047	0.296
Intracellular signaling cascade	82	0.022	0.04	0.003
Cell-cell signaling	36	0.081	0.319	0.05
Vehicle-mediated transport	36	0.424	0.024	0.101
Ribosome biogenesis and assembly	33	0.046	1	0.114
Protein kinase cascade	24	0.335	0.038	0.077

[a]Total number of genes represented in each group corresponding to a specific gene ontology term, [b]p-value in the X group (X=A, B or C).

We assessed the utility of these 119 genes to identify patients who would develop liver recurrence after surgery. The expression of these 119 genes was used to build a weighted vote-based classifier that could predict liver metastasis at the time of surgery. We applied this classifier to the 28 independent samples. Eighteen and 10 patients were assigned to the localized signature and metastasized signature, respectively. Seventeen of the 18 (94.4%) patients assigned to the localized signature did not relapse at least 47 months after surgery, and 9 of the 10 (90%) patients assigned to the

Fundamental Set

▦	Stage I and II- LOC (n=32)
▦	Stage III -MET (n=7)
▦	StageⅣ -MET (n=19)
▦	LIV (n=34)

	Non-metastatic cluster	Metastatic cluster
Stage I and II-LOC	28	4
Stage III-MET	2	5
Stage IV-MET	8	11
LIV	6	28

Fundamental Set

▦	Stage I and II- LOC (n=32)
▦	Stage III -MET (n=7)
▦	StageⅣ -MET (n=19)
▦	LIV (n=34)

Independent Set

▦	Stage II-LOC (n=10)
▦	Stage III-LOC (n=8)
▦	Stage II-MET (n=4)
▦	Stage III-MET (n=6)

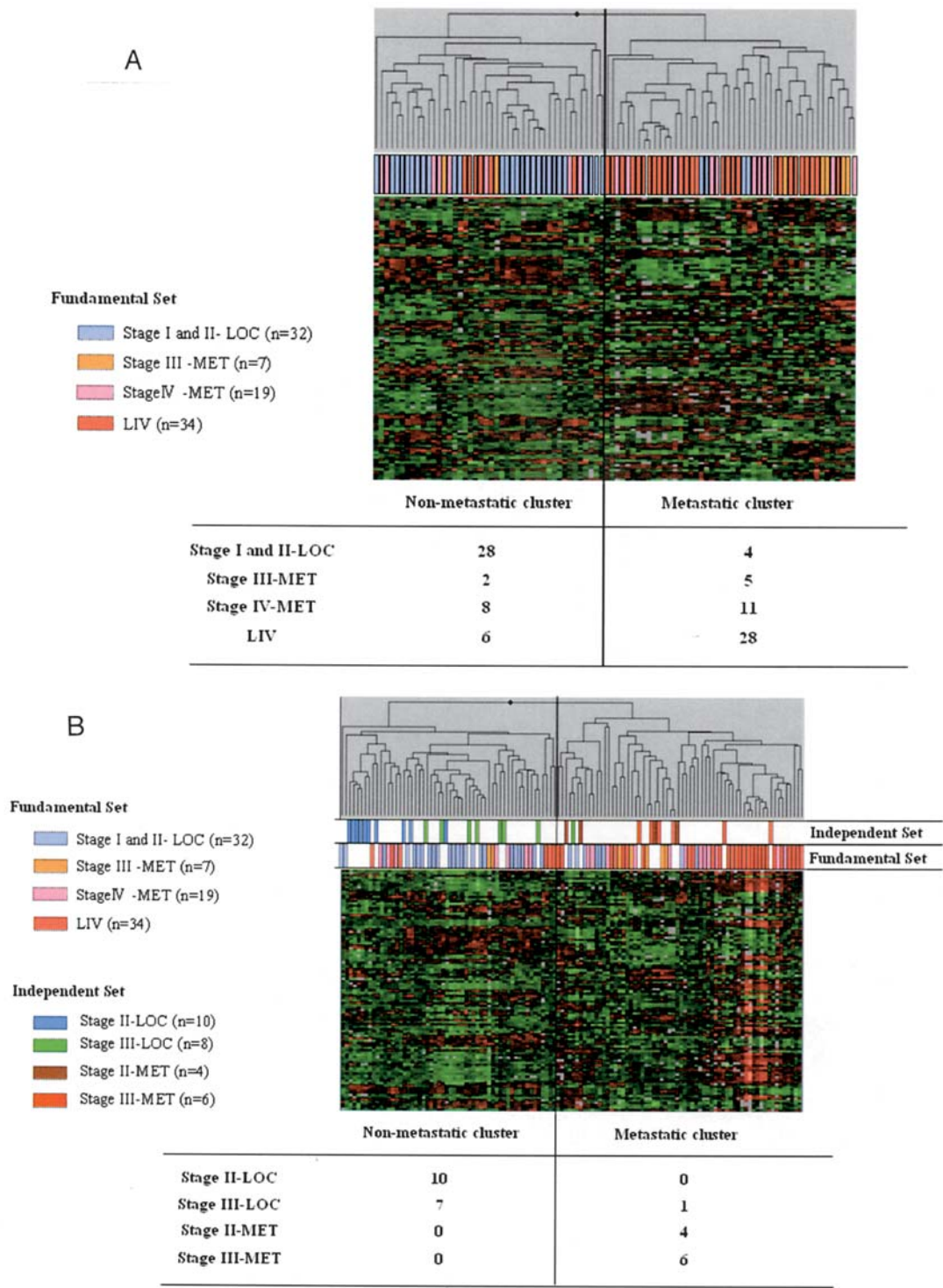|	Non-metastatic cluster	Metastatic cluster
Stage II-LOC	10	0
Stage III-LOC	7	1
Stage II-MET	0	4
Stage III-MET	0	6

Figure 2. Hierarchical cluster analysis of metastasized and non-metastasized tumor signatures. The genes in group B are hierarchically clustered with respect to the 92 tumors in the fundamental set (A) and the 120 tumors adding these 92 to the 18 tumors in the independent set (B). Each row represents a gene and each column represents a sample. For the samples, the color bars below the dendrogram represent sample annotation in clinicohistopathology. The aqua bars represent localized primary tumors (LOC), the orange and pink bars represent stage III- and stage IV- ones (stage III -MET and stage IV -MET) and the red bars represent liver metastatic lesions (LIV) in the fundamental set. The blue and green bars represent LOC with stage II and III (stage II-LOC and stage III-LOC), and the brown and red bars represent MET with stage II and III (stage II-MET and stage III-MET), respectively, in the independent set. (A) Two major branches of the hierarchical tree are observed. The left and right branches are mainly comprised of non-metastasized (non-metastatic cluster) and metastasized tumors (metastatic cluster), respectively. These two clusters significantly correlated with the metastatic status of the tumors. (B) The clustering pattern was so robust against sample addition that the partitioning of the 92 fundamental tumors remained the same and showed little change. Out of the 28 independent tumors, all of the 10 metastasized tumors are in the metastatic cluster, and 17 of the 18 non-metastasized tumors are in the non-metastatic cluster.

metastasized signature recurred to liver metastasis after surgery (Table III). The successful prediction of liver metastasis could be shown in the independent set of tumors.

To assess the association between class assigned by gene expression profile and survival, we plotted Kaplan-Meier survival curves of 28 patients in the independent set (Fig. 3A

Table III. The predictive diagnosis of liver metastasis in the independent samples.

		Clinical diagnosis	
		Non-relapse	Relapse
Profiling diagnosis	LOC	17	1
	MET	1	9

and B). A significant difference was observed in the disease-free survival (p<0.0001) and overall survival (p=0.0065) of patients between the localized and metastasized signature.

Discussion

The gene expression profile using DNA microarray is a promising approach to investigate the molecular complexity of cancer. Several studies have reported the usefulness of this technique such as the gene identification related to metastasis and possible prediction of metastasis in various cancers (16,17). However, these studies simply compared the gene expression profile of tumors between partial stages and did not directly address the metastatic process. To elucidate the phenomenon of metastasis more accurately, it is critical to analyze comprehensively the process of gradual change in cancer progression. Thus, we analyzed the gene expression profiles of 104 samples at four time points in cancer development; normal colon, localized and metastasized primary
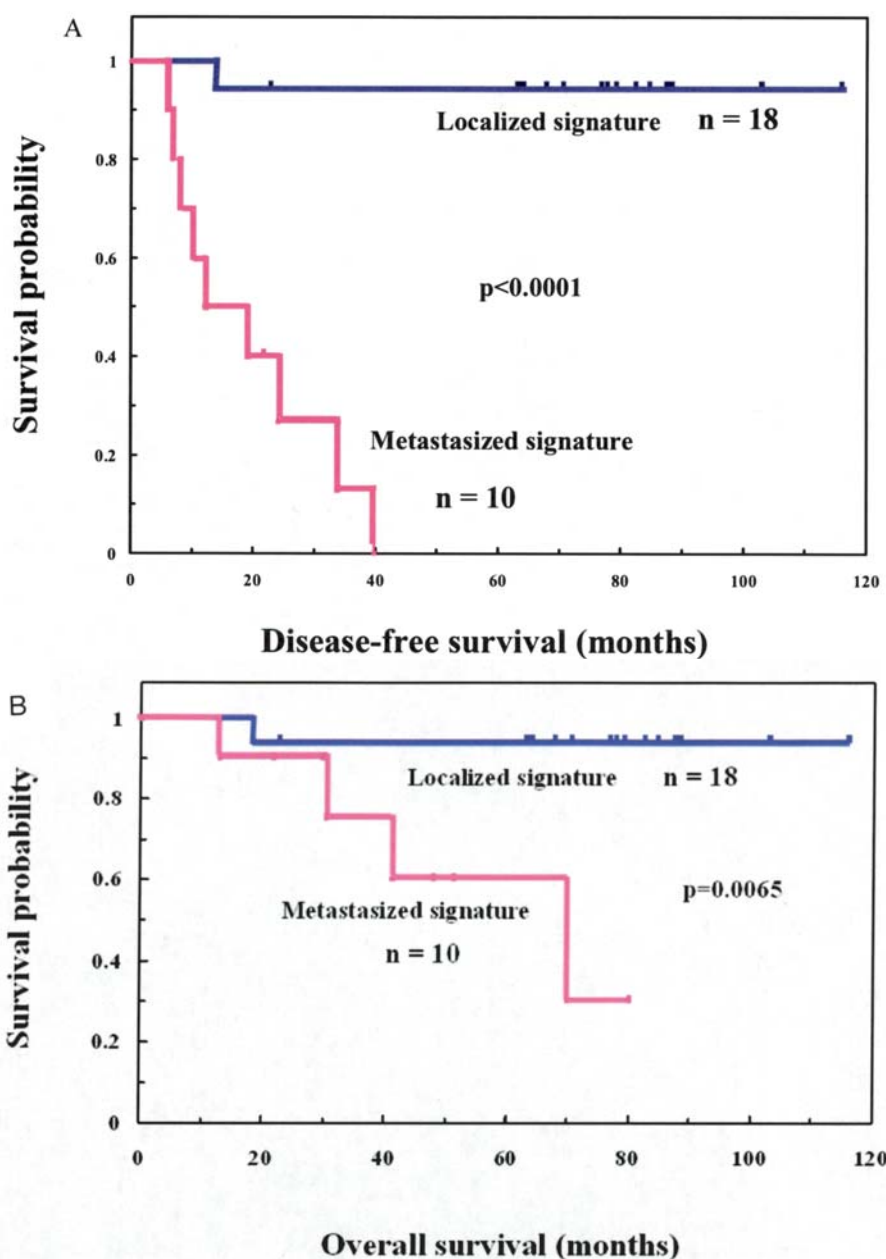

Figure 3. Kaplan-Meier plot of disease-free (A) and overall (B) survival of patients with the localized and metastasized signature. (A) Disease-free survival curves of patients with the localized signature were also higher than those of patients with the metastasized signature (log-rank test, p<0.0001). (B) Overall survival curves of patients with the localized signature were higher than those of patients with the metastasized signature (log-rank test, p=0.0065).

tumor and liver metastasis. These analyses indicate that each step of cancer progression can be characterized by a gene expression profile.

The expression pattern of group A is considered to be associated with the promotion of carcinogenesis in CRC. This group contained a significantly high proportion of genes with functions relevant to carcinogenesis such as cell proliferation and response to external and endogenous stimuli, and any imbalance in these functions initiates colorectal carcinogenesis. If these genes were expressed on the cell surface or secreted, this could be detected by easier methods such as immunohistochemical analysis and be applicable clinically as candidate diagnostic markers for the early detection of CRC (cellular-component, data not shown).

The expression of the genes represented in group B is significantly different in the primary tumor with or without metastasis, which directly reflects the event of metastasis itself. The loss or acquisition of this function in this group was closely relevant to the liver metastatic potential in colorectal cancer. This group contained a significantly high proportion of genes with functions relevant to the process of metastasis such as cell motility and apoptosis. The acquisition of cell motility is essential for the invasion of interstitial tissue and lymphatic or vascular vessel walls and the survival of cancer cells against immune and non-immune defenses while escape from apoptosis allows cancer cells to arrest a secondary organ apart from the primary lesion. Genes such as *DCC* and *GDP dissociation inhibitor 2* (*GDI2*), which are referred as relevant to metastasis, are included in group B. For example, *DCC* is a putative tumor suppressor gene, which has considerable homology to neural cell adhesion molecules and may be involved in the regulation of cell-to-cell or cell-to-substrate interactions with a potential functional role in the control of cell growth and the development of metastasis (22-24). Various studies reported that *DCC* is a predictor of distant metastasis and prognosis in colorectal cancer (24-26). Furthermore, *DCC* expression has been shown recently to predict clinical outcome to fluorouracil-based chemotherapy (27). Other studies have reported that the *DCC* gene product induced apoptosis through a caspase-dependent mechanism (28,29). In another study, *GDI2* was identified as a metastasis-suppressor gene (30). *GDI2* is an inhibitor of guanine nucleotide dissociation from RHO and RAC proteins, strongly inhibiting *GDP*/GTP exchange reaction. *GDI2*, RHO and RAC also modulate signal transduction such as the ERK, JNK and p38 MAPK pathways, three arms of which are involved in invasion, growth-factor-receptor signaling, the mitogen-activated protein kinase pathway, cell-cell communication and transcription, which are the main factors involved in the induction of liver metastasis (31,32).

The expression pattern of group C is closely associated with the colonization and maintenance of metastasis, which is the outgrowth of cancer cells after arrival at a secondary organ. Functional analysis indicated that metastatic colonization occurs as a result of the cancer cells' modification following interaction with the surrounding environment in the liver. Other investigators also reported similar results (16,33,34). Nishizuka *et al* (34) reported that cytokines produced by glial cells *in vivo* may contribute in a paracrine

manner to the development of brain metastases from breast cancer cells. Metastatic colonization differs from growth in the primary lesion by its requirement for angiogenesis, and changes in the local microenvironment in metastatic lesions and stressful conditions. The requirement for various growth factors and chemokines, the cell-cell or cell-extracellular matrix interaction and the activity of intracellular signaling pathways regulated by various stimuli and interactions could have considerable influence upon not only the initiation but also the maintenance of metastatic colonization (35,36). For example, *MAP3K4* and *ERK1*, which are differentially expressed in the metastatic tumor, stimulate the p38 and JNK MAPK pathway, and the MAPK-ERK pathway, respectively. Reduced JNK and p38 signals facilitate metastatic colonization by preventing apoptosis induced by the stress of a foreign environment (37). Numerous reports have linked the activation of *ERK1* with metastasis, motility and invasion, and angiogenesis (38,39). This result suggests that the cross-talk between signaling pathways, which were represented in the MAPK pathway, such as the ERK, p38 and JNK pathways, are linked closely within the cancer cells and interstitial tissue to adhesion, motility, proliferation, angiogenesis and apoptosis in metastatic lesions, which represents the final step in cancer development. These genes might be candidate markers of metastasis and could be suitable targets for the treatment of metastatic disease.

Chronological analysis allowed us to clarify each step of cancer progression, carcinogenesis to metastasis, at the molecular level. In particular, we were successful in separating metastasis into the acquisition of metastatic potential (group B) and the initiation and maintenance of metastatic colonization (group C).

Does the gene expression pattern correlate with the metastatic potential (such as that represented in group B) already present in the bulk of the primary tumor and detectable at the time of initial diagnosis? The conventional hypothesis about metastasis is that acquisition of the metastatic pheno-type may occur late in tumor progression. Primary CRC cells arise from the normal colorectal epithelium through onco-genic mutations. The occasional cancer cells possessing metastatic ability in the primary tumor population can accumulate additional genetic changes that mediate metastasis to the liver. Thus, it has been hypothesized that primary tumors have gene expression profiles in which only a fraction of the metastatic potential exists. However, some groups have challenged this hypothesis (16,17). Ramaswamy *et al* (16) reported that the expression profile of primary tumors that had metastasized resembled that of metastatic tumors, and that a predictive diagnosis for metastasis was possible based on the analysis of the primary tumor profile. Van't Veer *et al* (17) also reported that patients with early-stage breast cancer could be classified according to their prognosis based on the gene expression signature obtained from the primary tumors, and specific gene expression profiles were predictive for distant metastases. Thus, the new hypothesis indicates that the acquisition of metastatic ability does not begin in just a few cells in the primary tumor but instead most cells in such tumors are inherently capable of metastasis.

Our hierarchical clustering with the expression profile of the genes closely relevant to the metastatic potential shows

that the profile of the metastasized primary tumor resembled one of a metastatic lesion apart from a primary lesion rather than one of a non-metastasized primary tumor. This means that the tumors are segregated according to their final metastatic status rather than their clinical stages. This suggests that the metastatic potential of CRC is not acquired in proportion to cancer progression but has already been encoded in the primary tumor. Thus, our findings support the above new hypothesis rather than the conventional one.

We also succeeded in the verification of this hypothesis using independent primary tumors. Clustering analysis with these genes on CRC samples diagnosed as localized or regional disease allowed the robust classification of metastasized tumors vs non-metastasized tumors. Supervised classification analysis enabled us to predict the postoperative relapse risk of the patients who were diagnosed as localized or regional disease at the time of surgery. The classification appropriately reflected the outcome of CRC patients. Our microarray data provide important information on the metastatic potential in CRC and indicate that the expression profiles in the bulk of primary tumors reflect the character of individual tumors and are possible and valid enough to predict metastasis.

The clinical outcome of patients with CRC is governed to some extent, if not fully, by the inherent character of tumor cells. The programmed expression pattern is detectable in the bulk of primary tumors. The characterization of cancer based on gene expression profile promises to shift the paradigm of cancer diagnosis from general taxonomy by clinicohistopathological parameters to individualized risk assessment.

Currently, because it is clinically difficult to accurately predict metachronous liver metastasis at the time of initial diagnosis, a lot of time and effort is directed toward the detection of asymptomatic recurrences using postoperative monitoring such as CT and the CEA test. Therefore, a more accurate prediction of liver metastasis and clinical outcome should benefit the optimization of cancer treatment, including adjuvant therapy and postoperative surveillance, for individual patients, and at the same time avoid over- or unnecessary treatment and reduce overall cost. Chronological analysis of the gene expression profile should enable a better understanding of the metastatic process. The identified 'interesting' genes and profiles could help design new strategies for the treatment and diagnosis of metastasis.

References

1. Fearon ER, Hamilton SR and Vogelstein B: Clonal analysis of human colorectal tumors. Science 238: 193-197, 1987.
2. Hanahan D and Weinberg RA: The hallmarks of cancer. Cell 100: 57-70, 2000.
3. Fidler IJ: Critical factors in the biology of human cancer metastasis: twenty-eighth G.H.A. Clowes memorial award lecture. Cancer Res 50: 6130-6138, 1990.
4. Gutman M and Fidler IJ: Biology of human colon cancer metastasis. World J Surg 19: 226-234, 1995.
5. Woodhouse EC, Chuaqui RF and Liotta LA: General mechanisms of metastasis. Cancer 80: 1529-1537, 1997.
6. Chambers AF, Naumov GN, Varghese HJ, Nadkarni KV, MacDonald IC and Groom AC: Critical steps in hematogenous metastasis: an overview. Surg Oncol Clin N Am 10: 243-255, 2001.
7. Fidler IJ: Orthotopic implantation of human colon carcinomas into nude mice provides a valuable model for the biology and therapy of metastasis. Cancer Metastasis Rev 10: 229-243, 1991.
8. Welch DR: Technical considerations for studying cancer metastasis *in vivo*. Clin Exp Metastasis 15: 272-306, 1997.
9. Mohr S, Leikauf GD, Keith G and Rihn BH: Microarrays as cancer keys: an array of possibilities. J Clin Oncol 20: 3165-3175, 2002.
10. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO and Staudt LM: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403: 503-511, 2000.
11. Notterman DA, Alon U, Sierk AJ and Levine AJ: Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. Cancer Res 61: 3124-3130, 2001.
12. Bertucci F, Salas S, Eysteries S, Nasser V, Finetti P, Ginestier C, Charafe-Jauffret E, Loriod B, Bachelart L, Montfort J, Victorero G, Viret F, Ollendorff V, Fert V, Giovaninni M, Delpero JR, Nguyen C, Viens P, Monges G, Birnbaum D and Houlgatte R: Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. Oncogene 23: 1377-1391, 2004.
13. Safi F and Beyer HG: The value of follow-up after curative surgery of colorectal carcinoma. Cancer Detect Prev 17: 417-424, 1993.
14. Compton CC: Colorectal carcinoma: diagnostic, prognostic, and molecular features. Mod Pathol 16: 376-388, 2003.
15. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, Gaasenbeek M, Angelo M, Reich M, Pinkus GS, Ray TS, Koval MA, Last KW, Norton A, Lister TA, Mesirov J, Neuberg DS, Lander ES, Aster JC and Golub TR: Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med 8: 68-74, 2002.
16. Ramaswamy S, Ross KN, Lander ES and Golub TR: A molecular signature of metastasis in primary solid tumors. Nat Genet 33: 49-54, 2003.
17. Van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R and Friend SH: Gene expression profiling predicts clinical outcome of breast cancer. Nature 415: 530-536, 2002.
18. Takemasa I, Higuchi H, Yamamoto H, Sekimoto M, Tomita N, Nakamori S, Matoba R, Monden M and Matsubara K: Construction of preferential cDNA microarray specialized for human colorectal carcinoma: molecular sketch of colorectal cancer. Biochem Biophys Res Commun 285: 1244-1249, 2001.
19. Komori T, Takemasa I, Higuchi H, Yamasaki M, Ikeda M, Yamamoto H, Ohue M, Nakamori S, Sekimoto M, Matsubara K and Monden M: Identification of differentially expressed genes involved in colorectal carcinogenesis using a cDNA microarray. J Exp Clin Cancer Res 23: 521-527, 2004.
20. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286: 531-537, 1999.
21. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, Riss J, Barrett JC and Weinstein JN: GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol 4: R28, 2003.
22. Fearon ER, Cho KR, Nigro JM, Kern SE, Simons JW, Ruppert JM, Hamilton SR, Preisinger AC, Thomas G and Kinzler KW: Identification of a chromosome 18q gene that is altered in colorectal cancers. Science 247: 49-56, 1990.
23. Cho KR and Fearon ER: DCC: linking tumour suppressor genes and altered cell surface interactions in cancer? Eur J Cancer 31A: 1055-1060, 1995.
24. Shibata D, Reale MA, Lavin P, Silverman M, Fearon ER, Steele G Jr, Jessup JM, Loda M and Summerhayes IC: The DCC protein and prognosis in colorectal cancer. N Engl J Med 335: 1727-1732, 1996.
25. Reymond MA, Dworak O, Remke S, Hohenberger W, Kirchner T and Kockerling F: DCC protein as a predictor of distant metastases after curative surgery for rectal cancer. Dis Colon Rectum 41: 755-760, 1998.

26. Saito M, Yamaguchi A, Goi T, Tsuchiyama T, Nakagawara G, Urano T, Shiku H and Furukawa K: Expression of DCC protein in colorectal tumors and its relationship to tumor progression and metastasis. Oncology 56: 134-141, 1999.
27. Aschele C, Debernardis D, Lonardi S, Bandelloni R, Casazza S, Monfardini S and Gallo L: Deleted in colon cancer protein expression in colorectal cancer metastases: a major predictor of survival in patients with unresectable metastatic disease receiving palliative fluorouracil-based chemotherapy. J Clin Oncol 22: 3758-3765, 2004.
28. Chen YQ, Hsieh JT, Yao F, Fang B, Pong RC, Cipriano SC and Krepulat F: Induction of apoptosis and G2/M cell cycle arrest by DCC. Oncogene 18: 2747-2754, 1999.
29. Mehlen P, Rabizadeh S, Snipas SJ, Assa-Munt N, Salvesen GS and Bredesen DE: The DCC gene product induces apoptosis by a mechanism requiring receptor proteolysis. Nature 395: 801-804, 1998.
30. Gildea JJ, Seraj MJ, Oxford G, Harding MA, Hampton GM, Moskaluk CA, Frierson HF, Conaway MR and Theodorescu D: RhoGDI2 is an invasion and metastasis suppressor gene in human cancer. Cancer Res 62: 6418-6423, 2002.
31. Jaffe AB and Hall A: Rho GTPases in transformation and metastasis. Adv Cancer Res 84: 57-80, 2002.
32. Arozarena I, Matallanas D and Crespo P: Maintenance of CDC42 GDP-bound state by Rho-GDI inhibits MAP kinase activation by the exchange factor Ras-GRF. Evidence for Ras-GRF function being inhibited by Cdc42-GDP but unaffected by CDC42-GTP. J Biol Chem 276: 21878-21884, 2001.

33. LaTulippe E, Satagopan J, Smith A, Scher H, Scardino P, Reuter V and Gerald WL: Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. Cancer Res 62: 4499-4506, 2002.
34. Nishizuka I, Ishikawa T, Hamaguchi Y, Kamiyama M, Ichikawa Y, Kadota K, Miki R, Tomaru Y, Mizuno Y, Tominaga N, Yano R, Goto H, Nitanda H, Togo S, Okazaki Y, Hayashizaki Y and Shimada H: Analysis of gene expression involved in brain metastasis from breast cancer using cDNA microarray. Breast Cancer 9: 26-32, 2002.
35. Liotta LA and Kohn EC: The microenvironment of the tumour-host interface. Nature 411: 375-379, 2001.
36. Chambers AF, Groom AC and MacDonald IC: Dissemination and growth of cancer cells in metastatic sites. Nat Rev Cancer 2: 563-572, 2002.
37. Takekawa M, Posas F and Saito H: A human homolog of the yeast Ssk2/Ssk22 MAP kinase kinase kinases, MTK1, mediates stress-induced activation of the p38 and JNK pathways. EMBO J 16: 4973-4982, 1997.
38. Webb CP, Van Aelst L, Wigler MH and Woude GF: Signaling pathways in Ras-mediated tumorigenicity and metastasis. Proc Natl Acad Sci USA 95: 8773-8778, 1998.
39. Simon C, Hicks MJ, Nemechek AJ, Mehta R, O'Malley BW Jr, Goepfert H, Flaitz CM and Boyd D: PD 098059, an inhibitor of ERK1 activation, attenuates the *in vivo* invasiveness of head and neck squamous cell carcinoma. Br J Cancer 80: 1412-1419, 1999.