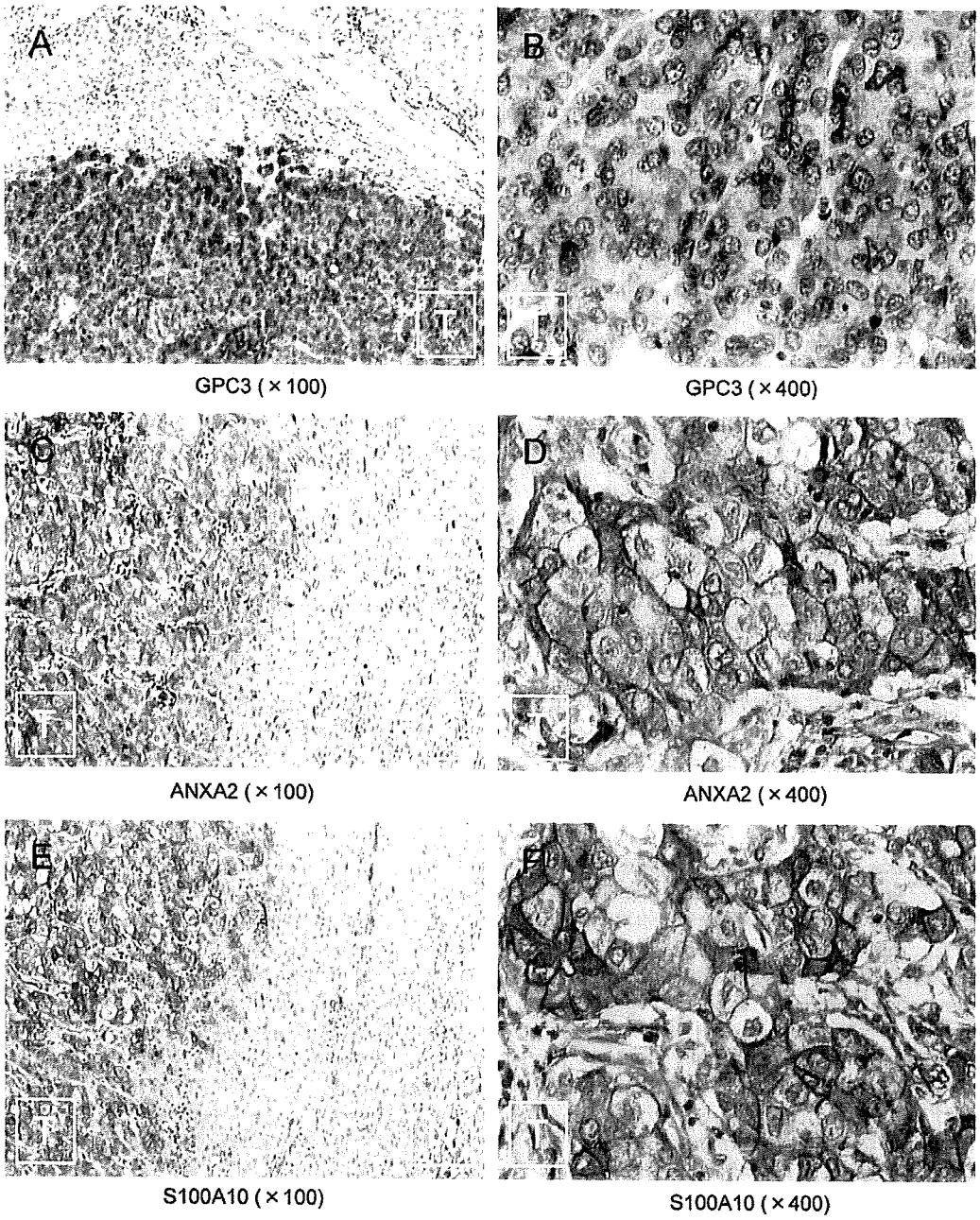

Fig. 3 – (A and B) GPC3 protein expression in human HCC tissue. Brown GPC3 immunostaining is evident in cancer cells. Note the diffuse non-granular staining pattern in the cytoplasm. (C and D) ANXA2 protein expression in human HCC tissue. Note ANXA2 staining in the cell membrane of cancer cells, with slight immunoreactivity in the cytoplasm. (E and F) S100A10 protein expression in human HCC tissues. Note S100A10 staining at the cell membrane of cancer cells. C, D, E and F are from the same tissue sample, and staining for S100A10 overlapped significant with ANXA2 staining. (A, C and E) ×100 magnification. (B, D and F) ×400 magnification. T; tumour region. N; normal liver.

and membrane-binding protein ANXA2 can form a hetero-tetrameric complex with S100A10 and this complex is thought to serve as a bridging or scaffolding function in the membrane underlying cytoskeleton.[14] Previous studies demonstrated both ANXA2 and S100A10 at the plasma membrane in hepatoblastoma HepG2 cell lines,[15] but not in human HCC tissues. Immunohistochemical staining of ANXA2 and S100A10 was stronger at the plasma membrane of the same samples than in the cytoplasm (Fig. 3C–F). The immunoreactivity for ANXA2 was heterogeneous, and cancerous tissues were immunopositive in 16 samples. Endothelial cells were immunopositive for ANXA2 in all samples tested. Similarly, staining of S100A10 was heterogeneous, and cancerous tissues were immunopositive in 17 samples.

Table 3 – Clinicopathological characteristics and results of immunohistochemical staining of ANXA2 and S100A10

Patient	Age	Gender	Hepatitis virus infection	Histological type	Vascular invasion	Annexin2 expression	S100A10 expression
Case 1	71	M	HCV	por	–	++	++
Case 2	66	F	HBV	mod	+	++	++
Case 3	38	M	HBV	mod	+	++	++
Case 4	65	M	HCV	por	+	++	++
Case 5	59	M	HCV	por	–	++	++
Case 6	72	F	HCV	mod	–	+	++
Case 7	71	M	–	por	–	++	+
Case 8	72	M	HBV, HCV	por	–	++	+
Case 9	61	M	–	por	–	++	+
Case 10	78	M	HCV	por	–	++	+
Case 11	49	F	HBV	por	+	++	+
Case 12	72	F	HBV	por	+	+	+
Case 13	68	M	HBV, HCV	mod	–	+	+
Case 14	78	F	HBV, HCV	mod	–	+	+
Case 15	74	M	HBV, HCV	well	–	+	+
Case 16	67	M	–	mod	–	–	+
Case 17	53	M	HBV	mod	+	–	–
Case 18	75	M	HCV	por	–	–	–
Case 19	66	M	HBV, HCV	mod	+	–	–
Case 20	60	M	HCV	por	–	–	–

++, strong immunopositive; +, partial immunopositive; –, immunonegative.

Colocalisation of ANXA2 and S100A10 was observed in 15 samples (Table 3).

3.5. *Correlation between gene expression signature of the two 'hotspot' and clinicopathological features*

Next, to better understand if any of the two 'hotspot' identified by this integrated approach correlates with clinicopathological features, a hierarchical clustering of all 100 HCC samples using the upregulated genes included in the 'hotspot' was performed. Fig. 4A shows the gene expression profiles using the 11 genes upregulated in the 'hotspot 1 (integrin signalling)'. Examination of this result allowed identification of three subgroups; 'relatively high-activated group ($n = 39$), defined as Group A1', 'intermediate-activated group ($n = 45$), defined as Group A2', and 'relatively low-activated group ($n = 16$), defined as Group A3'. Likewise, Fig. 4B shows the gene expression profiles using the 12 genes upregulated in the 'hotspot 2 (Akt/NF-κB signalling)'. Examination of this result also identified two subgroups; 'relatively high-activated group ($n = 17$), defined as Group B1' and 'relatively low-activated group ($n = 83$), defined as Group B2'. Having identified the two distinctive subgroups; 'relatively high-activated group' and 'relatively low-activated group' in each of the two 'hotspots', we examined the association between the activation of 'hotspot' and clinicopathological data (Table 4A and B). In the integrin signalling, the activated profile was significantly associated with intrahepatic metastasis ($p = 0.012$), tumour size ($p = 0.023$) and Edmonson grading ($p < 0.001$). On the other hand, in the Akt/NF-κB signalling, the activated profile was significantly associated with Edmonson grading ($p = 0.004$). Kaplan–Meier plot showed a significant difference in the probability of disease-free survival ($p = 0.037$) and overall survival ($p = 0.045$) between 'Group A1' and 'Group A3' in the integrin signalling (Fig. 4C and D). In the Akt/NF-κB signalling, a significant difference was observed in disease-free survival ($p = 0.045$, Fig. 4E and F).

3.6. *Overview of the distribution of differentially expressed genes on human chromosome*

To compare our microarray data with chromosomal aberrations in HCC, we also investigated the chromosomal region in which the differentially expressed genes were harboured. The public source for annotating the location of each gene was the National Center for Biotechnology Information (NCBI). This investigation revealed that the regions of high density of 100 upregulated genes tended to be at chromosomes 1q, 6p and 8q (Fig. 5A), whilst those of 100 downregulated genes were at chromosomes 4q and 16q (Fig. 5B). Furthermore, SPP1, GPC3, ANXA2 and S100A10, identified as key molecules, were separately located at chromosomes 4q, Xq, 15q and 1q.

4. Discussion

In the post-genomic period, DNA microarray technology is used to monitor disease progress and to individualise treatment regimens. However, extracting new biological insights from high-throughput genomic studies of cancer progression poses a challenge due to difficulties in recognising and evaluating relevant biological processes from vast quantities of experimental data. Although other high-throughput technologies in protein expression (proteomics) and low-molecular weight metabolite expression (metabolomics) have made remarkable progress, no comprehensive analytical techniques exist that can measure more than 500,000 protein forms and 100,000–1,000,000 metabolites quantitatively.[16] Generating biological networks from comprehensive gene expression profiles manually in a visual manner could be used to navigate

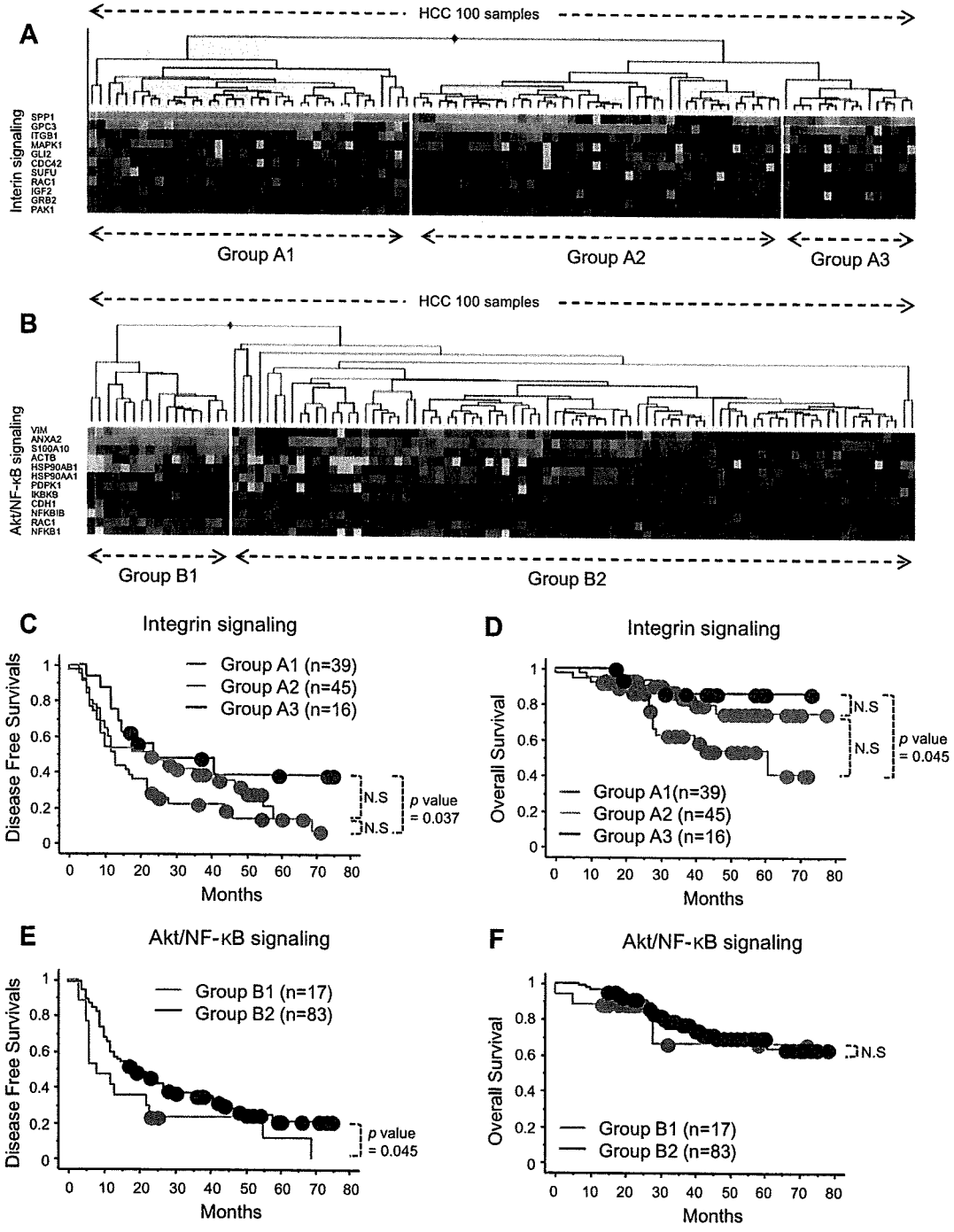

Fig. 4 – (A) Hierarchical clustering analysis of all 100 HCC samples using the 11 upregulated genes included in the 'hotspot 1 (integrin signalling)'. Red and green indicate relative high- and low-expression, respectively. Based on the similarities of their gene expression profiles, samples were grouped in 'relatively high-activated group ($n = 39$), defined as Group A1', 'intermediate-activated group ($n = 45$), defined as Group A2', and 'relatively low-activated group ($n = 16$), defined as Group A3'. (B) A hierarchical clustering analysis of all 100 HCC samples using the 12 upregulated genes included in the 'hotspot 2 (Akt/NF-κB signalling)'. Based on the similarities of their gene expression profiles, samples were grouped in 'relatively high-activated group ($n = 17$), defined as Group B1' and 'relatively low-activated group ($n = 83$), defined as Group B2'. (C and D) Disease-free survival and overall survival of each of the activated groups in the 'hotspot 1 (integrin signalling)' (Kaplan–Meier plot). The log-rank p value is shown. NS, not significant. (E and F) Disease-free survival and overall survival of each of the activated groups in the 'hotspot 2 (Akt/NF-κB signalling)' (Kaplan–Meier plot). The log-rank p value is shown. NS, not significant.

Table 4 – Clinical and pathological characteristics of the high-activated and low-activated groups in each of integrin signalling and AKT/NF-κB signalling

Characteristics	Integrin signalling ('Hotspot' 1)			p value
	Group A1 ($n = 39$)	Group A2 ($n = 45$)	Group A3 ($n = 16$)	
	No. of patients (%)	No. of patients (%)	No. of patients(%)	
A				
Intrahepatic metastasis	12(30.8)	(10(22.2))	0(0)	0.012
Tumour size (cm)	4.79 ± 3.12	(3.49 ± 1.75)	2.91 ± 1.22	0.023
Edmonson grading				
1–2	11(28.2)	22(48.9)	10(62.5)	<0.001
3–4	28(71.8)	23(51.1)	6(37.5)	
Pathological stage				
I	5(12.8)	15(33.3)	3(18.7)	0.642
II	22(56.4)	20(44.4)	10(62.5)	
III	9(23.1)	8(17.8)	3(18.7)	
IVA	3(7.7)	2(4.5)	0	

Characteristics	Akt/NF-κB signalling ('Hotspot' 2)		p value
	Group B1 ($n = 17$)	Group B2 ($n = 83$)	
	No. of patients (%)	No. of patients (%)	
B			
Intrahepatic metastasis	4(23.5)	18(21.7)	0.867
Tumour size (cm)	4.55 ± 2.58	3.88 ± 2.81	0.226
Edmonson grading			
1–2	2(11.7)	41(49.4)	0.004
3–4	15(88.3)	42(50.6)	
Pathological stage			
I	1(5.9)	22(26.5)	0.333
II	11(64.7)	41(49.4)	
III	4(23.5)	16(19.3)	
IVA	1(5.9)	4(4.8)	

p values of A are obtained by comparing Group A1 with Group A3.
p values of B are obtained by comparing Group B1 with Group B2.
p values for intrahepatic metastasis, Edmonson grading and pathological stage are obtained by χ^2 test.
p value for tumour size is obtained by t test.
Tumour size is mean ± SD.

through and unravel the complex networks involved in cancer progression. Here, we combined genome-wide expression analysis with a new bioinformatics method, Ingenuity Pathway Analysis, to clarify the relationship between the microarray datasets and the canonical pathways based on the published literature and identify functional networks, 'hotspot', responsible for the progression of HCC. Furthermore, we discovered several molecules commonly upregulated in HCC as potential key players in the neoplastic process.

This combined approach revealed that several distinct regions with upregulated genes were concentrated. These concentrations of activated genes included several genes involved in the WNT signalling pathway, which has been the subject of intense research in recent years. In addition, we highlighted integrin and Akt/NF-κB as two 'hotspot' signalling pathways, and propose that these signalling pathways are crucial for the core biological functions in HCC progression and potential intervention, such as cell proliferation, cell survival and apoptosis.[17]

In addition to studies of gene expression at the transcriptional level, protein analysis is vital for understanding the regulatory processes in living organisms, because emerging evidence suggests that mRNA expression patterns themselves are necessary but insufficient for quantitative description of biological systems. So far, comparative studies of mRNA and protein abundance indicate that only 20–28% of the total variation of protein abundance can be attributed to mRNA abundance alone.[18] The limiting factors were explained partly by translational processes (microRNAs repress the translation of mRNAs into proteins) and post-translational modification (such as phosphorylation, methylation, acetylation, glycosylation and ubiquitination). In fact, this was the basis for investigating the expression levels of proteins encoded by highly upregulated genes related to key signalling pathways in hepatocarcinogenesis.

Using our analysis protocol, integrin pathway-associated molecules, SPP1 and GPC3, were first identified as key for cell proliferation in HCC. HCC generally spreads throughout the

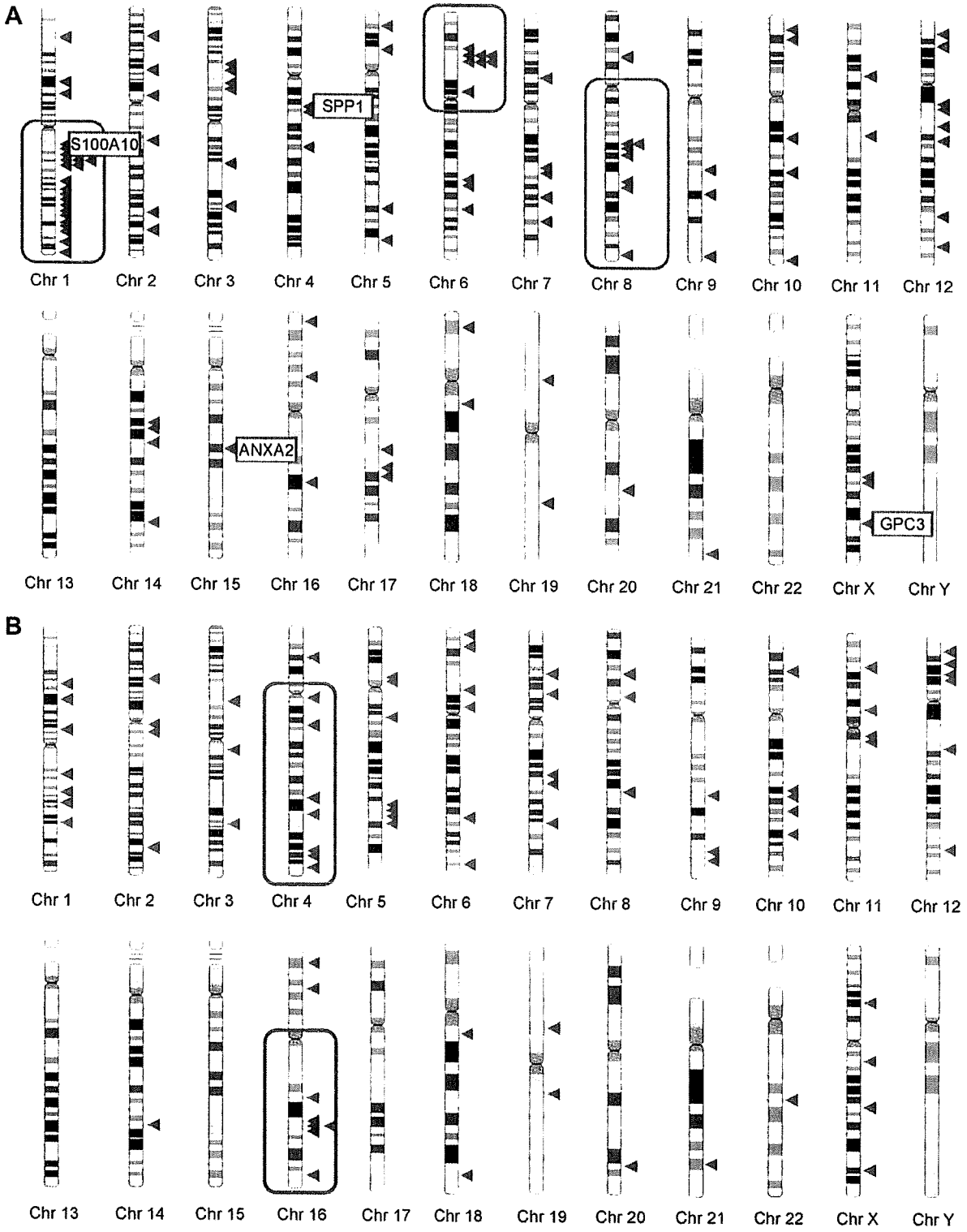

Fig. 5 – Location of 100 upregulated genes (A) and 100 downregulated genes (B) on human chromosomes. Upregulated genes are represented as red arrows, and downregulated genes are represented as blue arrows. Red-coloured circles (chromosomes 1q, 6p and 8q) represent regions with relative concentration of upregulated genes, whilst blue coloured circles (chromosomes 4q and 16q) represent regions with relative concentration of down-regulated genes. ANXA2, S100A10, SPP1 and GPC3 are located at chromosomes 15q21-22, 1q22, 4q21-25 and Xq26.1, respectively.

liver via the portal vein system even in advanced stages, and portal vein invasion is the most crucial histological feature associated with poor prognosis.[3] SPP1 protein expression is upregulated in primary HCC with accompanying metastasis, and SPP1 expression correlates with the invasiveness of HCC cells in tissue culture.[19] Based on network analysis, we speculated that the binding of integrins to SPP1[20] might be related to the progression and metastasis of HCC. GPC3, a heparan sulphate proteoglycan anchored to the plasma membrane, is also a good candidate marker of HCC. It is an oncofoetal protein overexpressed in HCC at both the mRNA and protein levels.[11] We also confirmed that GPC3 was overexpressed in HCC by the immunostaining of paraffin sections. In the activated integrin pathway, GPC3 interacts with IGF-2,[21] a protein that increases the phosphorylation of MAPK1.[22] GPC3 is also related to the zinc-finger transcription factors, GLI1 and GLI2, that are known players in WNT signalling and Sonic hedgehog signalling pathways.[23] Recently, WNT signalling was implicated in hepatocyte proliferation, which could be crucial in liver development, regeneration following partial hepatectomy, and pathogenesis of HCC.[24] In this context, SPP1 and GPC3 might participate in the activation of integrin signalling in HCC and based on the results of clustering analysis, might be implicated as mediators of intrahepatic metastasis, histopathological malignancy or poor prognosis.

Second, we focused on ANXA2, S100A10 and VIM, which were related to the Akt/NF-κB signalling pathway. ANXA2, also called calpactin I heavy chain, is a member of the annexin family of Ca^{2+}- and phospholipid-binding proteins and forms a heterotetrameric complex with S100A10, also called calpactin I light chain.[25] The ANXA2-S100A10 complex has been implicated in the structural organisation and dynamics of endosomal membranes, the organisation of cholesterol-rich membrane microdomains, and connecting lipid rafts with the actin cytoskeleton.[25] The ANXA2-S100A10 complex was also recently associated with recycling endosomes, and might be involved in the recycling of E-cadherin during the formation of the E-cadherin-based adherens junctions via the modulation of the actin cytoskeleton.[26] Moreover, ANXA2 was identified as a Rac binding partner and Rac activation is induced by the interactions of E-cadherin in the formation of adherens junctions.[27] In this way, cadherin–cadherin interactions initiate a cascade of signalling events that result in increased cadherin/Akt association, activation of Akt/NF-κB signalling, and increased cell survival and tumour growth.[27] Akt1 was found to associate structurally with VIM (a structural component of intermediate filaments),[28] which has been found in poorly differentiated HCC as well as hepatoblastomas.[29] Therefore, it is possible that binding of Akt1 and VIM activates downstream players (NF-κB signalling) as well as increasing the intrinsic activity of Akt1. This molecular understanding of HCC progression in Akt/NF-κB signalling was not so different from our result of correlations between clinicopathological features and gene expression profiles. It seems that the higher-activated group in Akt/NF-κB signalling has lower histopathological differentiation.

Currently, the array-based CGH approach is used to study chromosomal aberrations in human cancers. A previously reported meta-analysis[30] showed that the most common chromosomal arms containing gains were 1q, 6p and 8q, whereas the most common losses were found in chromosomes 4q, 8p and 16q. Comparing our expression data with the meta-analysis result of array-based CGH in HCC,[30] we found that our gene expression data surprisingly matched the chromosomal aberrations. The comprehensive analysis of 100 HCC samples using human 30 K DNA microarray revealed a potential association between the global copy number and expression. It is also noteworthy that our identified key molecules that operate synergistically in hepatocarcinogenesis are located at separate chromosomes, so chromosomal aberrations cannot prove a relationship of candidate genes such as ANXA2 and S100A10. Therefore, our integrative network approach can provide a significant clue to the discovery of novel genetic combinations that may be important for hepatocarcinogenesis.

Here, we highlighted the 'hotspot' canonical pathways in HCC and improved our molecular understanding of HCC progression. It is widely recognised that there are distinct molecular subtypes of HCC in the transcriptome space, and current interest of the community spread to include identification of subtype-specific aberration of genes/pathway. This functional genomics study could contribute towards the detection of several signalling pathways commonly activated in HCC. Moreover, we succeeded in detecting two potential disease markers, ANXA2 and S100A10, whose colocalisation in human HCC tissues has not been reported previously.

In conclusion, we reported an integrative approach of genome-wide microarray analysis and network analysis in HCC. This novel approach allows the extraction of deeper biological insight from microarray data and identifying potential key molecules in hepatocarcinogenesis.

Conflict of interest statement

None declared.

Acknowledgements

We thank EIJI MIYOSHI, Department of Biochemistry, Osaka University Medical School, and JORGE FILMUS, Division of Molecular and Cell Biology, Sunnybrook and Women's College Health Sciences Centre and Department of Medical Biophysics, University of Toronto, for providing a monoclonal antibody of GPC3.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejca.2008.02.019.

REFERENCES

1. Thomas MB, Abbruzzese JL. Opportunities for targeted therapies in hepatocellular carcinoma. *J Clin Oncol* 2005;23:8093–108.

2. Jemal A, Murray T, Ward E, Samuels A, Tiwari RC, Ghafoor A. Cancer statistics, 2005. *CA Cancer J Clin* 2005;**55**:10–30.
3. Pan HW, Ou YH, Peng SY, Liu SH, Lai PL, Lee PH. Overexpression of osteopontin is associated with intrahepatic metastasis, early recurrence, and poorer prognosis of surgically resected hepatocellular carcinoma. *Cancer* 2003;**98**:119–27.
4. Tsou AP, Wu KM, Tsen TY, Chi CW, Chiu JH, Lui WY. Parallel hybridization analysis of multiple protein kinase genes: identification of gene expression patterns characteristic of human hepatocellular carcinoma. *Genomics* 1998;**50**:331–40.
5. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V. Genetics of gene expression surveyed in maize, mouse and man. *Nature* 2003;**422**:297–302.
6. Brem RB, Yvert G, Clinton R, Kruglyak L. Genetic dissection of transcriptional regulation in budding yeast. *Science* 2002;**296**:752–5.
7. Thorgeirsson SS, Grisham JW. Molecular pathogenesis of human hepatocellular carcinoma. *Nat Genet* 2002;**31**:339–46.
8. Slaton JW, Perrotte P, Inoue K, Dinney CP, Fidler IJ. Interferon-alpha-mediated down-regulation of angiogenesis-related genes and therapy of bladder cancer are dependent on optimization of biological dose and schedule. *Clin Cancer Res* 1999;**5**:2726–34.
9. Yamamoto T, Nagano H, Sakon M, Wada H, Eguchi H, Kondo M. Partial contribution of tumour necrosis factor-related apoptosis-inducing ligand (TRAIL)/TRAIL receptor pathway to antitumor effects of interferon-alpha/5-fluorouracil against hepatocellular carcinoma. *Clin Cancer Res* 2004;**10**:7884–95.
10. Kondo M, Yamamoto H, Nagano H, Okami J, Ito Y, Shimizu J. Increased expression of COX-2 in nontumor liver tissue is associated with shorter disease-free survival in patients with hepatocellular carcinoma. *Clin Cancer Res* 1999;**5**:4005–12.
11. Capurro M, Wanless IR, Sherman M, Deboer G, Shi W, Miyoshi E. Glypican-3: a novel serum and histochemical marker for hepatocellular carcinoma. *Gastroenterology* 2003;**125**:89–97.
12. Zhang H, Ozaki I, Mizuta T, Yoshimura T, Matsuhashi S, Hisatomi A. Mechanism of beta 1-integrin-mediated hepatoma cell growth involves p27 and S-phase kinase-associated protein 2. *Hepatology* 2003;**38**:305–13.
13. Mottet D, Dumont V, Deccache Y, Demazy C, Ninane N, Raes M. Regulation of hypoxia-inducible factor-1alpha protein level during hypoxic conditions by the phosphatidylinositol 3-kinase/Akt/glycogen synthase kinase 3beta pathway in HepG2 cells. *J Biol Chem* 2003;**278**:31277–85.
14. Puisieux A, Ji J, Ozturk M. Annexin II up-regulates cellular levels of p11 protein by a post-translational mechanisms. *Biochem J* 1996;**313**:51–5.
15. Zobiack N, Gerke V, Rescher U. Complex formation and submembranous localization of annexin 2 and S100A10 in live HepG2 cells. *FEBS Lett* 2001;**500**:137–40.
16. Hollywood K, Brison DR, Goodacre R. Metabolomics: current technologies and future trends. *Proteomics* 2006;**6**:4716–23.
17. Cantley LC. The phosphoinositide 3-kinase pathway. *Science* 2002;**296**:1655–7.
18. Mootha VK, Bunkenborg J, Olsen JV, et al. Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* 2003;**115**:629–40.
19. Ye QH, Qin LX, Forgues M, He P, Kim JW, Peng AC. Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nat Med* 2003;**9**:416–23.
20. Hu DD, Lin EC, Kovach NL, Hoyer JR, Smith JW. A biochemical characterization of the binding of osteopontin to integrins alpha v beta 1 and alpha v beta 5. *J Biol Chem* 1995;**270**:26232–8.
21. Song HH, Shi W, Filmus J. OCI-5/rat glypican-3 binds to fibroblast growth factor-2 but not to insulin-like growth factor-2. *J Biol Chem* 1997;**272**:7574–7.
22. Moorehead RA, Sanchez OH, Baldwin RM, Khokha R. Transgenic overexpression of IGF-II induces spontaneous lung tumors: a model for human lung adenocarcinoma. *Oncogene* 2003;**22**:853–7.
23. Regl G, Kasper M, Schnidar H, Eichberger T, Neill GW, Ikram MS. The zinc-finger transcription factor GLI2 antagonizes contact inhibition and differentiation of human epidermal cells. *Oncogene* 2004;**23**:1263–74.
24. Apte U, Zeng G, Muller P, Tan X, Micsenyi A, Cieply B. Activation of Wnt/beta-catenin pathway during hepatocyte growth factor-induced hepatomegaly in mice. *Hepatology* 2006;**44**:992–1002.
25. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev* 2002;**82**:331–71.
26. Yamada A, Irie K, Hirota T, Ooshio T, Fukuhara A, Takai Y. Involvement of the annexin II-S100A10 complex in the formation of E-cadherin-based adherens junctions in Madin–Darby canine kidney cells. *J Biol Chem* 2005;**280**:6016–27.
27. Kovacs EM, Ali RG, McCormack AJ, Yap AS. E-cadherin homophilic ligation directly signals through Rac and phosphatidylinositol 3-kinase to regulate adhesive contacts. *J Biol Chem* 2002;**277**:6708–18.
28. Siu MK, Wong CH, Lee WM, Cheng CY. Sertoli-germ cell anchoring junction dynamics in the testis are regulated by an interplay of lipid and protein kinases. *J Biol Chem* 2005;**280**:25029–47.
29. Abenoza P, Manivel JC, Wick MR, Hagen K, Dehner LP. Hepatoblastoma: an immunohistochemical and ultrastructural study. *Hum Pathol* 1987;**18**:1025–35.
30. Moinzadeh P, Breuhahn K, Stutzer H, Schirmacher P. Chromosome alterations in human hepatocellular carcinomas correlate with aetiology and histological grade–results of an explorative CGH meta-analysis. *Br J Cancer* 2005;**92**:935–41.

Gene expression of colorectal cancer: Preoperative genetic diagnosis using endoscopic biopsies

TAKAMICHI KOMORI[1], ICHIRO TAKEMASA[1], MAKOTO YAMASAKI[1], MASAAKI MOTOORI[1], TAKESHI KATO[2], NOBUTERU KIKKAWA[2], NAOMASA KAWAGUCHI[3], MASATAKA IKEDA[1], HIROFUMI YAMAMOTO[1], MITSUGU SEKIMOTO[1], KENICHI MATSUBARA[4], NARIAKI MATSUURA[3] and MORITO MONDEN[1]

[1]Department of Surgery, Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871; [2]Department of Surgery, Minoh City Hospital, 5-7-1 Sugano, Minoh, Osaka 562-0014; [3]Department of Molecular Pathology, Graduate School of Medicine and Health Science, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871; [4]DNA Chip Research Inc., 1-1-43 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan

Received July 5, 2007; Accepted August 30, 2007

Abstract. In colorectal cancer, to predict the response to chemo- and/or radio-therapy or the existence of lymph node metastasis preoperatively, a more competent diagnostic system is required, in addition to conventional diagnosis based on morphology and pathology. The application of gene expression profiling to preoperative cancer diagnosis using endoscopic biopsies could enable the selection of a more appropriate therapy for patients. In this study, we evaluated the feasibility of gene expression profiling using preoperative biopsies of colorectal tumors in a clinical setting, by investigating the influence of intra-tumor heterogeneity on the profiles and testing the prediction ability of tumor malignancy. Under endoscopic examination, two biopsies were sampled from each of 10 colorectal cancers and 10 adenomas, and their gene expression data were obtained using cDNA microarrays. The intra- and inter-tumor heterogeneities of the profiles were compared with unsupervised clustering analysis. Molecular prediction of tumor malignancy using biopsies was performed with the supervised classification algorithm. In clustering analysis, almost all paired biopsies from the same tumors joined each other. Pearson's correlation coefficients of the profiles between biopsies from the same tumors (mean, 0.83) were significantly greater than those of the profiles between biopsies from other cancers (mean, 0.58) ($p<0.0001$). In the supervised classification method, malignancy was correctly predicted in 39 out of 40 biopsies with 8-71 informative genes. Gene expression profiling using endoscopic biopsies of colorectal tumors revealed that the intra-tumor heterogeneity was smaller than the inter-tumor heterogeneity and tumor malignancy was correctly predicted. Our findings suggest that the technique of gene expression profiling accurately represents the biological properties of colorectal cancer and could help the preoperative diagnosis of this disease.

Correspondence to: Dr Ichiro Takemasa, Department of Surgery, Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
E-mail: takemasa@surg2.med.osaka-u.ac.jp

Abbreviations: CRC, colorectal cancer; HCA, hierarchical clustering analysis; SNR, signal-to-noise ratio; IHC, immunohistochemical staining; pAb, polyclonal antibody; mAb, monoclonal antibody; FNAB, fine needle aspiration biopsy

Key words: colorectal cancer, gene expression profile, microarray, biopsy, tumor heterogeneity

Introduction

The incidence of colorectal cancer (CRC) is increasing and it is one of the leading causes of cancer death in Japan (1). Conventional diagnosis based on morphology and pathology, such as Dukes' classification and the tumor-node-metastasis (TNM) staging system, has played an important role in the clinical decision-making and evaluation of prognosis for CRC (2-4). However, it is difficult to differentiate the response to chemo- and/or radio-therapy or the existence of lymph node metastasis preoperatively by conventional diagnosis. To predict such individual heterogeneous cancers' characteristics preoperatively, a more competent diagnostic system is required.

Comprehensive gene expression assay using microarray technology has provided insights into cancer pathogenesis and is expected to help to fulfil the clinical demands for individualized medicine (5). The discovery of a set of new molecular markers, which can classify cancers according to their properties using surgical specimens, has been reported in various cancers (6). In CRC, this technology has been used to elucidate the mechanisms involved in carcinogenesis (7-16) as well as to predict various clinicopathological aspects, such as recurrence after surgery (17-22). However, in clinical practice, this fruitful molecular prediction using surgical specimens will be limited to the selection of postoperative medicine such as adjuvant therapy and follow-up schedules. On the other hand, analysis using preoperative endoscopic biopsies, instead of

Table I. Patients and tumor characteristics.

Case	Age (years)	Sex	Tumor location	Tumor size (cm)	Tumor type	TNM staging T	N	Stage
Ca1	67	M	D	2.0	Well differentiated adenocarcinoma	T1	N0	I
Ca2	55	F	R	1.5	Well differentiated adenocarcinoma	T1	N0	I
Ca3	69	F	R	5.6	Well differentiated adenocarcinoma	T2	N0	I
Ca4	59	F	R	3.5	Moderately differentiated adenocarcinoma	T2	N0	I
Ca5	66	M	R	5.5	Moderately differentiated adenocarcinoma	T2	N0	I
Ca6	48	M	S	2.8	Moderately differentiated adenocarcinoma	T2	N0	I
Ca7	73	F	S	5.7	Well differentiated adenocarcinoma	T3	N0	II
Ca8	55	M	R	3.8	Moderately differentiated adenocarcinoma	T3	N0	II
Ca9	67	F	R	3.4	Moderately differentiated adenocarcinoma	T3	N0	II
Ca10	60	F	R	3.0	Well differentiated adenocarcinoma	T2	N1	III
Ad1	60	M	S	4.0	Tubulovillous adenoma			
Ad2	74	M	S	2.5	Tubular adenoma			
Ad3	77	M	A	2.5	Tubulovillous adenoma			
Ad4	68	M	T	1.1	Tubular adenoma			
Ad5	62	F	A	1.0	Tubulovillous adenoma			
Ad6	60	M	D	2.0	Tubulovillous adenoma			
Ad7	72	M	R	0.8	Tubulovillous adenoma			
Ad8	61	M	R	0.8	Tubular adenoma			
Ad9	64	F	S	1.2	Tubular adenoma			
Ad10	78	F	S	1.8	Tubular adenoma			

M, male; F, female; A, ascending colon; T, transverse colon; D, descending colon; S, sigmoid colon; R, rectum.

surgically resected samples, would widen the utility of microarray technology.

In rectal cancer, preoperative rather than postoperative chemo- and/or radio-therapy reduces local recurrence after surgery (23). Using preoperative biopsies under colonoscopic examination, prediction of the response to preoperative chemo- and/or radio-therapy would be useful for the selection of patients who would most benefit from preoperative therapies aimed at a better prognosis or improved chances of sphincter preservation (24,25). The prediction of lymph node metastasis could contribute to the avoidance of unnecessary surgery for early invasive CRC. This is clinically important since lymph node metastasis is found in only approximately 10% of early invasive CRCs (26); the remaining 90% without metastasis are more suited to undergo local excision such as colonoscopic resection or transanal endoscopic microsurgery.

To determine the clinical importance of diagnosis based on gene expression profiles using preoperative endoscopic biopsies in CRC, we investigated the data quality from low volume samples and the influence of intra-tumor heterogeneity on the profiles. Wide differences in profiles within a tumor would hinder the adoption of this technique. There are no reports that compared intra- and inter-tumor heterogeneity in expression profiles in colorectal tumors. In the present study, we sampled two biopsy specimens from each of 10 cancers and 10 adenomas obtained under colonoscopic examinations, and determined their gene expression profiles using cDNA microarrays. By comparing their profiles, we assessed the intra-tumor heterogeneity in colorectal tumors. Moreover, as a first step for clinical applicability, by testing the molecular prediction ability of tumor malignancy, we investigated whether this technique, employing preoperative endoscopic biopsies, allows characterization of the biological properties of colorectal cancer.

Materials and methods

Patients and sample collection. From each of the 10 cancers and 10 adenomas, two biopsy specimens were obtained under colonoscopic examination. The clinicopathological data of patients and their tumors are summarized in Table I. Six bulky samples of the same cancers and 8 normal colorectal epithelium tissues were obtained from surgically resected specimens. None of the adenomas contained a cancerous component. Samples were stored in RNAlater (Qiagen, Valencia, CA) at -20°C after sampling until RNA extraction. This study was approved by the Institutional Review Board of Osaka University and Minoh City Hospital, and informed consent was obtained from all patients.

RNA extraction. Total RNA was extracted from biopsy samples using an RNeasy kit (Qiagen). The average volume of extracted total RNA from one biopsy was 21.1 μg in cancers and 22.6 μg in adenomas. From bulky samples, total RNA was extracted

using Trizol reagent (Invitrogen, Carlsbad, CA) and cleaned using the RNeasy kit. The quality of extracted total RNA was checked by 0.8% agarose-gel electrophoresis and a LabChip kit (Agilent Technologies, Palo Alto, CA), and the quantity was determined with a spectrophotometer. The total RNA solution was stored at -80°C until use.

Hybridization to cDNA microarray. We used cDNA microarrays containing 4608 clones which were derived from 30,000 clones expressed in CRC tissues (10,16,22). The Pearson's correlation coefficient of the data from replicated samples using our microarrays was 0.95. Gene expression data were obtained using a previously described method (10), with some improvements. As a standard normal control reference, a mixture of total RNA extracted from 40 normal colorectal epithelia was used. Labeled cDNA targets for hybridization were synthesized by reverse transcription from standard- and sample-total RNA respectively, with the indirect labeling method.

For each reverse transcription, 12 μg of total RNA was mixed with 1 μg of oligo-dT primer (Invitrogen) in a total volume of 15.5 μl, heated to 70°C for 10 min and cooled for 5 min. The mixture consisted of 6 μl of 5X first strand buffer, 3 μl of 0.1 M DTT, 3 μl of nucleotide cocktail (5 mM each dATP, dCTP, and dGTP, 3 mM dTTP and 2 mM aminoallyl-dUTP) and 0.5 μl of 40 U/μl RNase inhibitor was added. After incubation for 2 min at 42°C, 2 μl of 200 U/μl Superscript II reverse transcriptase (Invitrogen) was added. After incubation for 1 h at 42°C, 3.3 μl of 0.5 M EDTA was added and the RNA strand was degraded with 3.3 μl of 2 N NaOH and incubation at 70°C for 20 min. The mixture was neutralized with 3.3 μl of 2 N HCL and 60 μl of distilled water. This cDNA was purified using a QIAquick kit (Qiagen) and dried cDNA was dissolved in 9 μl of 0.2 M $NaHCO_3$-Na_2CO_3 (pH 9.0). Cy-dye solution (1 μl) (Amersham, Piscataway, NJ), Cy3 for the standard target, and Cy5 for the tested target, were added and incubated under shade at room temperature for 1 h, respectively. After purification using Micro Bio-Spin Columns P-30 Tris (Bio-Rad, Hercules, CA), the separately synthesized Cy3- and Cy5-labeled targets were combined and purified using Microcon YM-30 (Millipore, Billerica, MA), concentrated to a volume of 16.5 μl. To the concentrated target, 3.5 μl of human COT-1 DNA (Invitrogen), 7 μl of 20X SSC, 7 μl of 20X Denhart's solution, and 1 μl of 10% SDS were added. The 35 μl of target mixture was denatured by heating for 2 min at 95°C and cooled on ice. After the incubation at 50°C, the target was placed on the array. The array was incubated at 50°C for 14 h in a humid chamber. After hybridization, the slides were washed in 2X SSC with 0.1% SDS for 10 min, 0.1X SSC with 0.1% SDS for 10 min, and 0.1X SSC for 5 min at 30°C.

Scan and data processing. The array was scanned with ScanArray Lite (Perkin-Elmer, Wellesley, MA). The images were analyzed with QuantArray software (GSI Lumonics, Billerica, MA), converting signal intensities of each spot into numerical data. Data was processed after background subtraction, as described previously (10,16,22). Cy5/Cy3 ratios were log-transformed, and the global normalization was performed. Genes with >15% missing values in each group of cancers and adenomas were excluded from further analysis.

Statistical analysis. First, gene expression profiles of 20 cancer biopsies and 8 normal epithelia were compared. Hierarchical clustering analysis (HCA) was performed with the software GeneMaths version 2.0 (Applied Maths, Inc., Austin, TX) using all 1966 genes after data processing. Pearson's correlation was used as the similarity coefficient and the unweighted pair group method using arithmetic average as the clustering algorithm. In HCA, similarity in the expression pattern of paired biopsies from the same cancer was compared with those from other cancers, to define the difference between intra- and inter-tumor heterogeneity of profiles in cancers. Statistical significance of the difference was determined using t-test. To verify the validity of the list of differentially expressed genes involved in colorectal carcinogenesis using biopsy samples, examined genes in cancer biopsies and normal epithelia were ranked according to the signal-to-noise ratio (SNR) (5) and they were compared with previous reports using tissue samples. The p-value of SNR was calculated by performing the random permutation test 1,000 times (5).

Next, using 20 cancer biopsies and 20 adenoma biopsies, the prediction of malignancy in colorectal tumors by their gene expression profiles was tested. Genes differentially expressed in cancer and adenoma biopsies were ranked according to SNR, and differential diagnosis was performed using the weighted-votes method (5). The prediction accuracy was determined with the leave-one-out cross validation (5). Positive prediction strength was judged to be a cancer and negative to be an adenoma. Using the gene set with the highest accuracy, HCA was performed to identify the difference between the intra- and the inter-tumor heterogeneity of the profiles in cancers and adenomas.

Immunohistochemical staining. Immunohistochemical staining (IHC) was performed to investigate the translation of mRNA of differentially expressed genes to each coding protein and to examine the intra-tumor heterogeneity of expression patterns in cancers at the protein level. Among differentially expressed genes between cancer biopsies and normal epithelia in our transcriptional analyses, 7 genes whose antibodies were commercially available were selected. Buffered formalin-fixed (10%), paraffin-embedded sections were prepared from 10 surgically resected cancers, whose biopsy specimens were included in the microarray analyses. The streptavidin-biotin immunoperoxidase complex method (27) was used for IHC. Primary antibodies used were as follows; polyclonal antibody (pAb) to human peroxiredoxin 1 (PRDX1, Alexis Biochemicals, Lausen, Switzerland), pAb to high-mobility group box 1 (HMGB1, Santa Cruz Biotechnology, Santa Cruz, CA), pAb to DEK oncogene (DEK, Santa Cruz Biotechnology), pAb to poly(A) binding protein, cytoplasmic 1 (PABPC1, Santa Cruz Biotechnology), monoclonal antibody (mAb) to heat shock 60-kDa protein 1 (HSPD1, Sigma Aldrich, St. Louis, MO), mAb to nucleolin (NCL, Santa Cruz Biotechnology) and mAb to carbonic anhydrase II (CA2, Rockland, Gilbertsville, PA). Sections for negative control were tested by using normal mouse serum instead of primary

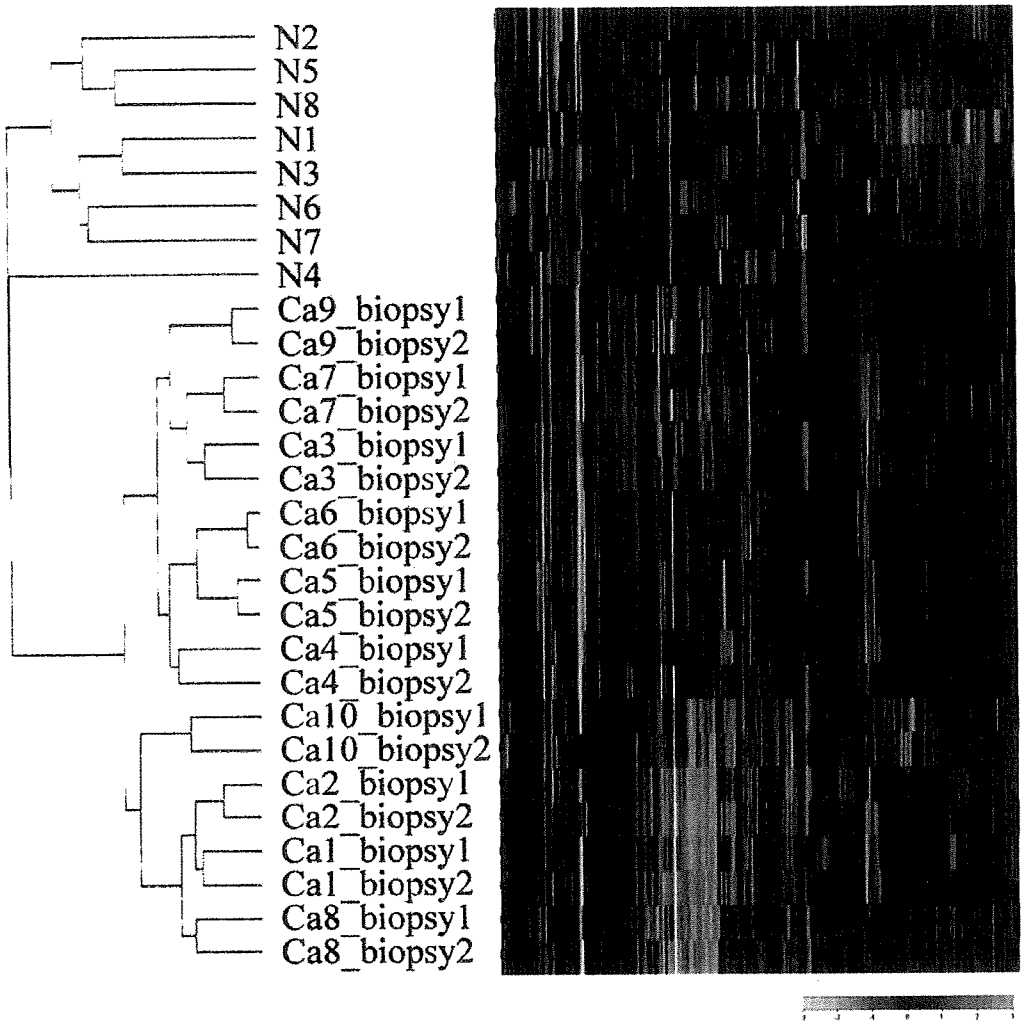

Figure 1 Hierarchical clustering analysis using full genes. Samples consisted of cancer biopsies (n=20; 10 pairs) and normal epithelia (n=8). Pearson's correlation was used as the similarity coefficient and the unweighted pair group method using the arithmetic average as the clustering algorithm. Red indicates overexpression, and green indicates underexpression. The respective paired biopsies from different areas of the same tumors are joined together.

antibody. Tissue sections of thyroid adenoma (for PRDX1), poorly differentiated gastric adenocarcinoma (for HMG1 and DEK), uterine cervical carcinoma (for HSPD1 and NCL), normal testis (for PABPC1) and normal colon (for CA2) were prepared as positive controls according to the recommendations of the manufacturer and previous publications on IHC. All slides were evaluated by a pathologist who was blinded to the microarray data. For each immunohistochemical analysis, the mean intensity in epithelial or tumor cells was evaluated in comparison with the positive controls as follows: weak, 1^+; moderate, 2^+; and strong, 3^+.

Results

Intra-tumor heterogeneity in colorectal cancers. Comparison of gene expression profiles between cancers and normal epithelia by HCA showed clear separation of the 20 cancer biopsies from the 8 normal tissues using all genes (Fig. 1). Moreover, the respective paired biopsies from different areas of the same cancers were joined together. Pearson's correlation coefficients of paired biopsies from the same cancers (mean, 0.83; SD, 0.08) were significantly greater than those of unpaired biopsies from other cancers (mean, 0.58; SD, 0.12; $p<0.0001$). Pearson's correlation coefficients between biopsies and their parent surgical bulky samples (mean, 0.60; SD, 0.15) were significantly greater than those between biopsies and surgical bulky samples from other cancers (mean, 0.37; SD, 0.16; $p<0.0001$).

Differentially expressed genes in cancer biopsies. Differentially expressed genes in cancer biopsies and normal epithelia were ranked according to SNR. Among them, 692 up-regulated and 219 down-regulated genes in cancer biopsies had SNR ($p<0.001$). The top 20 up-regulated and 20 down-regulated genes, excluding 4 down-regulated genes with no definition, are listed in Table II. PABPC1 and HSPD1 in the up-regulated genes and CA2, carboxylesterase 2 (CES2) and one EST in the down-regulated genes in cancer biopsies were also included in differentially expressed genes involved in colorectal carcinogenesis in our previous reports (10,16). Nine of the 40 differentially expressed genes were reported previously with respect to colorectal carcinogenesis with DNA microarray

Table II. Differentially expressed genes in colorectal cancer biopsies and normal colorectal epithelia.

Accession no.	Symbol	Gene definition	Previous report[a]	SNR
Up-regulated genes in cancer biopsies				
NM_181697	**PRDX1**	**peroxiredoxin 1, transcript variant 3**		**2.645**
NM_002128	**HMGB1**	**high-mobility group box 1**	**8,11,15**	**2.364**
NM_003472	**DEK**	**DEK oncogene (DNA binding)**	**11**	**2.341**
NM_001009	RPS5	ribosomal protein S5		2.052
NM_002568	**PABPC1**	**poly(A) binding protein, cytoplasmic 1**	**11,16**	**1.994**
NM_199440	**HSPD1**	**heat shock 60-kDa protein 1 (chaperonin), nuclear gene encoding mitochondrial protein, transcript variant 2**	**10,11,16,19**	**1.985**
NM_021130	**PPIA**	**peptidylprolyl isomerase A (cyclophilin A), transcript variant 1**		**1.957**
BE564899		EST		1.930
NM_001011	RPS7	ribosomal protein S7		1.894
NM_000978	RPL23	ribosomal protein L23		1.878
NM_002954	RPS27A	ribosomal protein S27a		1.860
AK090536		EST		1.826
NM_198829	**RAC1**	**ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1), transcript variant Rac1c**		**1.814**
NM_000971	RPL7	ribosomal protein L7		1.795
NM_003908	EIF2S2	eukaryotic translation initiation factor 2, subunit 2 ß, 38 kDa		1.772
NM_003756	**EIF3S3**	**eukaryotic translation initiation factor 3, subunit 3 γ, 40 kDa**		**1.763**
NM_000998	RPL37A	ribosomal protein L37a		1.732
NM_021034	**IFITM3**	**interferon induced transmembrane protein 3 (1-8U)**		**1.730**
NM_005381	**NCL**	**nucleolin**	**9**	**1.727**
NM_002106	H2AFZ	H2A histone family, member Z		1.721
Down-regulated genes in cancer biopsies				
NM_174977	SEC14L4	SEC14-like 4 (S. cerevisiae)		-2.518
NM_000067	**CA2**	**carbonic anhydrase II**	**8-11,15,16**	**-2.280**
NM_017958	PLEKHB2	pleckstrin homology domain containing, family B (evectins) member 2		-2.272
NM_005578	**LPP**	**LIM domain containing preferred translocation partner in lipoma**	**18**	**-2.236**
BF826364		EST		-1.926
BC051280		EST		-1.921
NM_003869	**CES2**	**carboxylesterase 2 (intestine, liver), transcript variant 1**	**10,14,16**	**-1.897**
BX647478		EST		-1.834
XM_375330		EST		-1.811
NM_002518	NPAS2	neuronal PAS domain protein 2		-1.789
CD652660		EST		-1.766
NM_020746	KIAA127	KIAA1271 protein		-1.742
NM_000355	TCN2	transcobalamin II; macrocytic anemia		-1.731
NM_005242	F2RL1	coagulation factor II (thrombin) receptor-like 1		-1.728
BM987276		EST		-1.698
XM_370781		**EST**	**16**	**-1.684**
AK129631		FLJ26120		-1.657
NM_002096	GTF2F1	general transcription factor IIF, polypeptide 1, 74 kDa		-1.623
NM_000014	A2M	α-2-macroglobulin		-1.617
NM_020675	Spc25	kinetochore protein Spc25		-1.606

Top 20 up-regulated and 20 down-regulated genes excluding 4 down-regulated genes with no definition, among 692 up-regulated and 219 down-regulated genes in cancer biopsies with SNR ($p<0.001$). Bold: genes reportedly involved in colorectal carcinogenesis based on microarray analysis or genes reportedly involved in carcinogenesis of other cancer types. [a]Previous report on colorectal carcinogenesis using microarray analysis (ref. no.). SNR, signal-to-noise ratio.

Table III. Immunohistochemical staining.

Gene name	Symbol		Expression grade[a]				Staining pattern[b]
			-	1+	2+	3+	
Peroxiredoxin 1	PRDX1	Normal	10				
		Cancer		6	4		Homogeneous
High-mobility group box 1	HMGB1	Normal		10			
		Cancer		1	7	2	Homogeneous
DEK oncogene	DEK	Normal			7	3	
		Cancer		4	6		Heterogeneous
Poly(A) binding protein, cytoplasmic 1	PABPC1	Normal		10			
		Cancer		3	7		Heterogeneous
Heat shock 60-kDa protein 1	HSPD1	Normal	10				
		Cancer			2	8	Homogeneous
Nucleolin	NCL	Normal	4	6			
		Cancer			5	5	Homogeneous
Carbonic anhydrase II	CA2	Normal			4	6	
		Cancer		8	2		Homogeneous

Microarray analyses revealed up-regulation of PRDX1, HMGB1, DEK, PABPC1, HSPD1 and NCL and down-regulation of CA2 in cancers. [a]The mean intensity of immunohistochemical staining in the epithelial or tumor cells evaluated relative to the positive controls as follows: weak, 1+; moderate, 2+; strong, 3+. [b]Intra-tumoral heterogeneity in immunohistochemical staining pattern.

assay. Among the 20 up-regulated genes, 6 genes of ribosomal proteins were included; this finding was in agreement with previous observations (7,9,11,18). In addition, relationships with carcinogenesis of other cancers have been reported regarding PRDX1 (28), peptidylprolyl isomerase A (PPIA) (29), ras-related C3 botulinum toxin substrate 1 (RAC1) (30), eukaryotic translation initiation factor 3, subunit 3 γ (EIF3S3) (31) and interferon induced transmembrane protein 3 (IFITM3) (32).

Immunohistochemical staining. The expression status of encoding proteins from 7 genes in IHC is summarized in Table III. The up- and down-regulation of PRDX1, HMGB1, PABPC1, HSPD1, NCL, and CA2 at the protein level was not in conflict with observations at the transcription level (Table III, Fig. 2). PRDX1, HMGB1, HSPD1, NCL and CA2 showed a homogeneous staining pattern in cancer tissues regardless of the region in the tumors, while the others showed a heterogeneous pattern.

Molecular prediction of tumor malignancy. In the differential diagnosis between 20 cancer biopsies and 20 adenoma biopsies by their gene expression profiles using the supervised classification method, the highest prediction accuracy was 97.5% when 8-71 genes were used (Fig. 3). Comparison between intra- and inter-tumor heterogeneity of the profiles using HCA showed that respective paired biopsies from the same tumors tended to join each other. When the selected 71-gene set with the highest accuracy was used, cancer biopsies and adenoma biopsies were clearly separated, and 18 of 20 paired biopsies were clustered side by side though a small number of genes was used (Fig. 4). Among the 71-gene set, COL1A1, COL1A2, and EIF2S2 were also reported in other studies as useful discriminating genes between cancers and adenomas (11).

Discussion

In the present study, we applied comprehensive expression analysis and found that intra-tumor heterogeneity of the gene expression profiles was smaller than inter-tumor heterogeneity, using preoperative endoscopic biopsies of colorectal tumors. We also showed that tumor malignancy could be accurately diagnosed with the profile in a single biopsy. Such accuracy is a promising first step in the clinical application of this technique for various settings such as prediction of the response to preoperative chemo- and/or radio-therapy or the existence of lymph node metastasis. Our findings suggest that this technique can be potentially used to define accurately the biological properties of colorectal tumors.

In CRC, changes in gene expression profiles that occurred during chemotherapy were detected using rectal cancer biopsies (33). The possible prediction of response to preoperative chemo- and/or radio-therapy for rectal cancers by using gene expression profiling of a single biopsy has also been reported (24,25). However, in the application of diagnosis in the clinical field based on gene expression profiling using preoperative endoscopic biopsies, it is not favorable that the profiles of biopsies from a tumor are widely different from each other. In this regard, our results added support to those

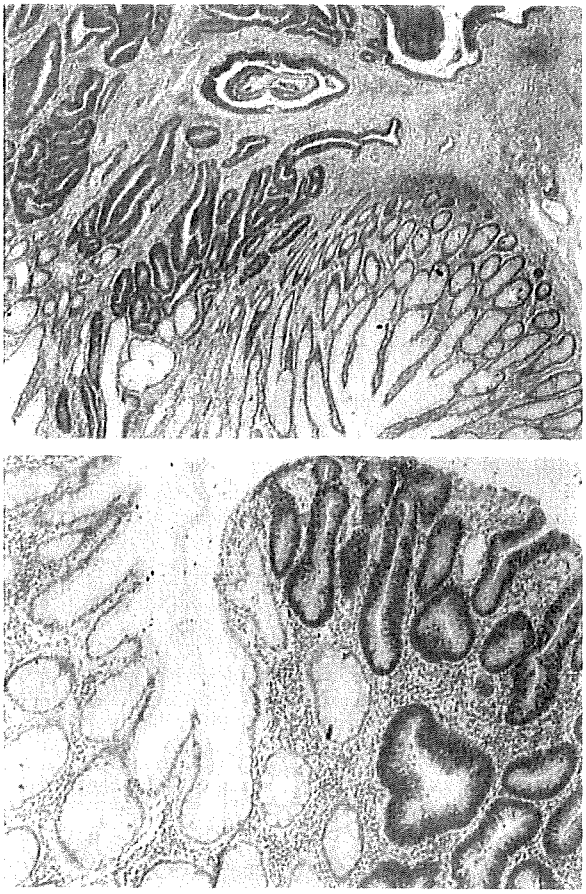

Figure 2. Immunohistochemical staining of HSPD1 (A) and NCL (B) in colorectal cancer tissues. (A) HSPD1 was expressed strongly in cancer cells while no HSPD1 expression was observed in adjacent normal epithelial cells (x4). (B) NCL was expressed strongly in cancer cells and weakly in normal epithelial cells in the lower portion of the colonic gland (x10).

of previous investigators (24,25), thus further promoting the technique.

Through using small-volume samples of a single biopsy of colorectal tumor, a gene expression profile was successfully obtained and a differentially expressed gene set related to colorectal carcinogenesis was detectable, as when using whole tissue samples. The feasibility of microarray-based study using fine-needle aspiration biopsy (FNAB) samples in some human tumors and endoscopically obtained tissues from precancerous lesions such as Barrett's esophagus was also reported (34,35). These results suggested that low-volume tissue samples such as endoscopic biopsy might give an accurate picture of gene expression in the whole tumor.

The histopathological features are not always homogeneous within a solid tumor. In CRC, the surface and invasive front of the tumor are sometimes histopathologically different. The influence of such morphological intra-tumor heterogeneity on gene expression profiles is not clear, although heterogeneity was detected in individual genes (36). Heterogeneity based on the superficial area of the tumor would have an unfavorable impact on molecular diagnosis when biopsy samples are used. In our study, the gene expression profiles of paired biopsies from the same tumors were identical in almost all tumors. At the protein level, IHC also showed a homogeneous staining

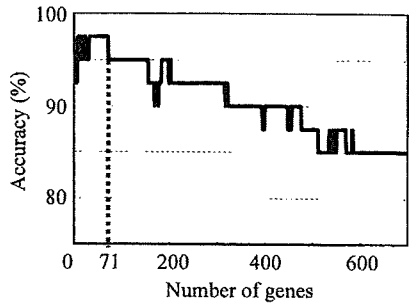

Figure 3. The accuracy curve of the differential diagnosis of 20 cancer biopsies and 20 adenoma biopsies using the supervised classification algorithm. Genes differentially expressed in cancer and adenoma biopsies were ranked according to the signal-to-noise ratio, and the accuracy was calculated with the weighted-votes method and the leave-one-out cross validation. The highest prediction accuracy was 97.5% when 8-71 genes were used.

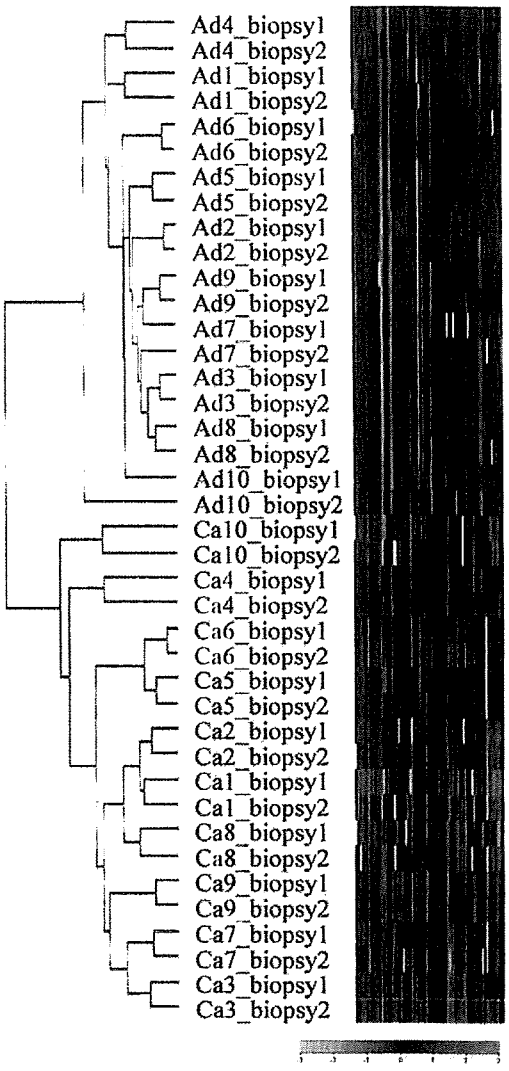

Figure 4. Hierarchical clustering analysis using 71 differentially expressed genes between cancer and adenoma biopsies. Samples consisted of cancer biopsies (n=20; 10 pairs) and adenoma biopsies (n=20; 10 pairs). Pearson's correlation was used as the similarity coefficient and the unweighted pair group method using arithmetic average as the clustering algorithm. Red indicates overexpression, and green indicates underexpression. Cancer biopsies and adenoma biopsies were clearly separated, and 18 of 20 paired biopsies were joined together.

pattern in cancer cells at the superficial position in 5 out of 7 genes. The heterogeneity in some varied genes at the protein level did not affect the intra-tumor homogeneity of the profiles at the transcriptional level. This may be because the use of tens to thousands of genes in profiling compensated for the difference due to heterogeneity in some varied genes or because the data at the transcriptional level do not always parallel those at the protein level (37).

Concerning the intra-tumor heterogeneity of the profiles in other cancers, differences in gene expression between cancer cells in the periphery and those in the center of breast cancers were detected using microdissected samples (38). However, only in a few candidate genes was the difference in the expression levels more than two-fold. In breast cancer, the profile of FNAB was reported to resemble closely that of the corresponding source tumor (39). In a comparison between FNABs and core biopsies, the difference in profiles within the same tumor was not greater than that between other tumors (40). Also in soft tissue sarcomas, the intra-tumor heterogeneity of the profiles was smaller than the inter-tumor heterogeneity (41).

It is not clear how many biopsies are sufficient to represent the biological properties of a tumor with gene expression profiles. The first part of our study indicated that the profile of one biopsy could distinguish cancers from adenomas, beyond the intra-tumor heterogeneity. However, our next pursuit is the differential diagnosis of more delicate and important differences, such as the potential of metastasis and sensitivity to chemo- or radio-therapy. Any unfavorable variation based on the sampling skill and the experiment may be a drawback on the actual clinical application of this method. Multiple biopsies from a tumor may be required to balance the differences within a tumor. The number of biopsies for microarray analysis was studied using biopsies from rectal epithelia and two biopsies per person were recommended, based on the equality of the expression data within a person (42), though similar analyses using biopsies from colorectal tumors have not reported.

Sampling techniques are important. Every biopsy was sampled by a specialist in the colonoscopic unit and very reliable sampling was carried out in this study. Inaccurate sampling would give confusing results showing a wide gap of profiles even between biopsies from one tumor. We took care to avoid contamination of the normal epithelium or adenomatous component on the periphery of the cancer tissue and confirmed the histology with pathological diagnosis of simultaneously obtained biopsies and surgical samples.

Our results suggest that gene expression profiling using endoscopic biopsies can accurately describe the biological properties of colorectal cancer. Further studies of gene expression profiling using preoperative endoscopic biopsies may allow the development of new diagnostic systems for the selection of neoadjuvant therapy or the method most appropriate for tumor resection. Ultimately, it may lead to individualized therapy for colorectal cancer.

References

1. Yiu HY, Whittemore AS and Shibata A: Increasing colorectal cancer incidence rates in Japan. Int J Cancer 109: 777-781, 2004.
2. Dukes CE: The classification of cancer of the rectum. J Pathol Bacteriol 35: 323-332, 1932.
3. Astler VB and Coller FA: The prognostic significance of direct extension of carcinoma of the colon and rectum. Ann Surg 139: 846-852, 1954.
4. Sobin LH and Witedkind CH (eds): TNM Classification of Malignant Tumors (5th edition). Wiley-Liss, New York, 1997.
5. Golub TR, Slonim DK, Tamayo P, *et al*: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286: 531-537, 1999.
6. Ramaswamy S and Golub TR: DNA microarrays in clinical oncology. J Clin Oncol 20: 1932-1941, 2002.
7. Alon U, Barkai N, Notterman DA, *et al*: Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci USA 96: 6745-6750, 1999.
8. Notterman DA, Alon U, Sierk AJ, *et al*: Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. Cancer Res 61: 3124-3130, 2001.
9. Kitahara O, Furukawa Y, Tanaka T, *et al*: Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. Cancer Res 61: 3544-3549, 2001.
10. Takemasa I, Higuchi H, Yamamoto H, *et al*: Construction of preferential cDNA microarray specialized for human colorectal carcinoma: molecular sketch of colorectal cancer. Biochem Biophys Res Commun 285: 1244-1249, 2001.
11. Lin YM, Furukawa Y, Tsunoda T, *et al*: Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. Oncogene 21: 4120-4128, 2002.
12. Zou TT, Selaru FM, Xu Y, *et al*: Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon. Oncogene 21: 4855-4862, 2002.
13. Ichikawa Y, Ishikawa T, Takahashi S, *et al*: Identification of genes regulating colorectal carcinogenesis by using the algorithm for diagnosing malignant state method. Biochem Biophys Res Commun 296: 497-506, 2002.
14. Birkenkamp-Demtroder K, Christensen LL, Olesen SH, *et al*: Gene expression in colorectal cancer Cancer Res 62: 4352-4363, 2002.
15. Williams NS, Gaynor RB, Scoggin S, *et al*: 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.
16. Komori T, Takemasa I, Higuchi H, *et al*: Identification of differentially expressed genes involved in colorectal carcinogenesis using a cDNA microarray. J Exp Clin Cancer Res 23: 521-527, 2004.
17. Wang Y, Jatkoe T, Zhang Y, *et al*: Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. J Clin Oncol 22: 1564-1571, 2004.
18. Bertucci F, Salas S, Eysteries S, *et al*: Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. Oncogene 23: 1377-1391, 2004.
19. Kwon HC, Kim SH, Roh MS, *et al*: Gene expression profiling in lymph node-positive and lymph node-negative colorectal cancer. Dis Colon Rectum 47: 141-152, 2004.
20. Li M, Lin YM, Hasegawa S, *et al*: Genes associated with liver metastasis of colon cancer, identified by genome-wide cDNA microarray. Int J Oncol 24: 305-312, 2004.
21. Arango D, Laiho P, Kokko A, *et al*: Gene-expression profiling predicts recurrence in Dukes' C colorectal cancer. Gastroenterology 129: 874-884, 2005.
22. Yamasaki M, Takemasa I, Komori T, *et al*: The gene expression profile represents the molecular nature of liver metastasis in colorectal cancer. Int J Oncol 30: 129-138, 2007.
23. Sauer R, Becker H, Hohenberger W, *et al*: Preoperative versus postoperative chemoradiotherapy for rectal cancer. N Engl J Med 351: 1731-1740, 2004.
24. Ghadimi BM, Grade M, Difilippantonio MJ, *et al*: Effectiveness of gene expression profiling for response prediction of rectal adenocarcinoma to preoperative chemoradiotherapy. J Clin Oncol 23: 1826-1838, 2005.
25. Watanabe T, Komuro Y, Kiyomatsu T, *et al*: Prediction of sensitivity of rectal cancer cells in response to preoperative radiotherapy by DNA microarray analysis of gene expression profiles. Cancer Res 66: 3370-3374, 2006.

26. Okabe S, Shia J, Nash G, *et al*: Lymph node metastasis in T1 adenocarcinoma of the colon and rectum. J Gastrointest Surg 8: 1032-1040, 2004.
27. Ito Y, Yoshida H, Matsuzuka F, *et al*: Expression of the components of the Cip/Kip family in malignant lymphoma of the thyroid. Pathobiology 71: 164-170, 2004.
28. Chang JW, Jeon HB, Lee JH, *et al*: Augmented expression of peroxiredoxin I in lung cancer. Biochem Biophys Res Commun 289: 507-512, 2001.
29. Campa MJ, Wang MZ, Howard B, *et al*: Protein expression profiling identifies macrophage migration inhibitory factor and cyclophilin a as potential molecular targets in non-small cell lung cancer. Cancer Res 63: 1652-1656, 2003.
30. Pan Y, Bi F, Liu N, *et al*: Expression of seven main Rho family members in gastric carcinoma. Biochem Biophys Res Commun 315: 686-691, 2004.
31. Saramaki O, Willi N, Bratt O, *et al*: Amplification of EIF3S3 gene is associated with advanced stage in prostate cancer. Am J Pathol 159: 2089-2094, 2001.
32. Hisamatsu T, Watanabe M, Ogata H, *et al*: Interferon-inducible gene family 1-8U expression in colitis-associated colon cancer and severely inflamed mucosa in ulcerative colitis. Cancer Res 59: 5927-5931, 1999.
33. Clarke PA, George ML, Easdale S, *et al*: Molecular pharmacology of cancer therapy in human colorectal cancer by gene expression profiling. Cancer Res 63: 6855-6863, 2003.
34. Centeno BA, Enkemann SA, Coppola D, *et al*: Classification of human tumors using gene expression profiles obtained after microarray analysis of fine-needle aspiration biopsy samples. Cancer 105: 101-109, 2005.
35. EL-Seraq HB, Nurqalieva Z, Souza RF, *et al*: Is genomic evaluation feasible in endoscopic studies of Barrett's esophagus? A pilot study. Gastrointest Endosc 64: 17-26, 2006.
36. Kaio E, Tanaka S, Kitadai Y, *et al*: Clinical significance of angiogenic factor expression at the deepest invasive site of advanced colorectal carcinoma. Oncology 64: 61-73, 2003.
37. Nishizaki S, Charboneau L, Young L, *et al*: Proteomic profiling of the NCI-60 cancer cell line using new high-density reverse-phase lysate microarrays. Proc Natl Acad Sci USA 100: 14229-14234, 2003.
38. Zhu G, Reynolds L, Crnogorac-Jurcevic T, *et al*: Combination of microdissection and microarray analysis to identify gene expression changes between differentially located tumour cells in breast cancer. Oncogene 22: 3742-3748, 2003.
39. Assersohn L, Gangi L, Zhao Y, *et al*: The feasibility of using fine needle aspiration from primary breast cancers for cDNA microarray analyses. Clin Cancer Res 8: 794-801, 2002.
40. Symmans WF, Ayers M, Clark EA, *et al*: Total RNA yield and microarray gene expression profiles from fine-needle aspiration biopsy and core-needle biopsy samples of breast carcinoma. Cancer 97: 2960-2971, 2003.
41. Francis P, Fernebro J, Eden P, *et al*: Intratumor versus intertumor heterogeneity in gene expression profiles of soft-tissue sarcomas. Gene Chromosome Cancer 43: 302-308, 2005.
42. Pellis L, Franssen-van Hal NL, Burema J, *et al*: The intraclass correlation coefficient applied for evaluation of data correction, labeling methods, and rectal biopsy sampling in DNA microarray experiments. Physiol Genomics 16: 99-106, 2003.

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

Received March 13, 2006; Accepted August 23, 2006

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.

Correspondence to: Dr Ichiro Takemasa, Department of Surgery and Clinical Oncology, Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita Osaka 565-0879, Japan
E-mail: alfa-t@sf6.so-net.ne.jp

Abbreviations: cDNA, complementary DNA; CEA, carcinoembryonic antigen; CRC, colorectal cancer; CT, computer tomography; *DCC*, deleted-in-colon carcinoma; ERK, extracellular-signal-regulated kinase; *GDI2*, *GDP dissociation inhibitor 2*; GO, gene ontology; HCA, hierarchical cluster analysis; MAPK, mitogen-activated protein kinase; RT-PCR, reverse transcription polymerase chain reaction; SAGE, serial analysis of a gene expression; TNM, tumor node metastasis; UPGMA, unweighted pair group method using arithmetic averages; VEGF, vascular endothelial growth factor

Key words: colorectal cancer, liver metastasis, gene expression profile, molecular analysis

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

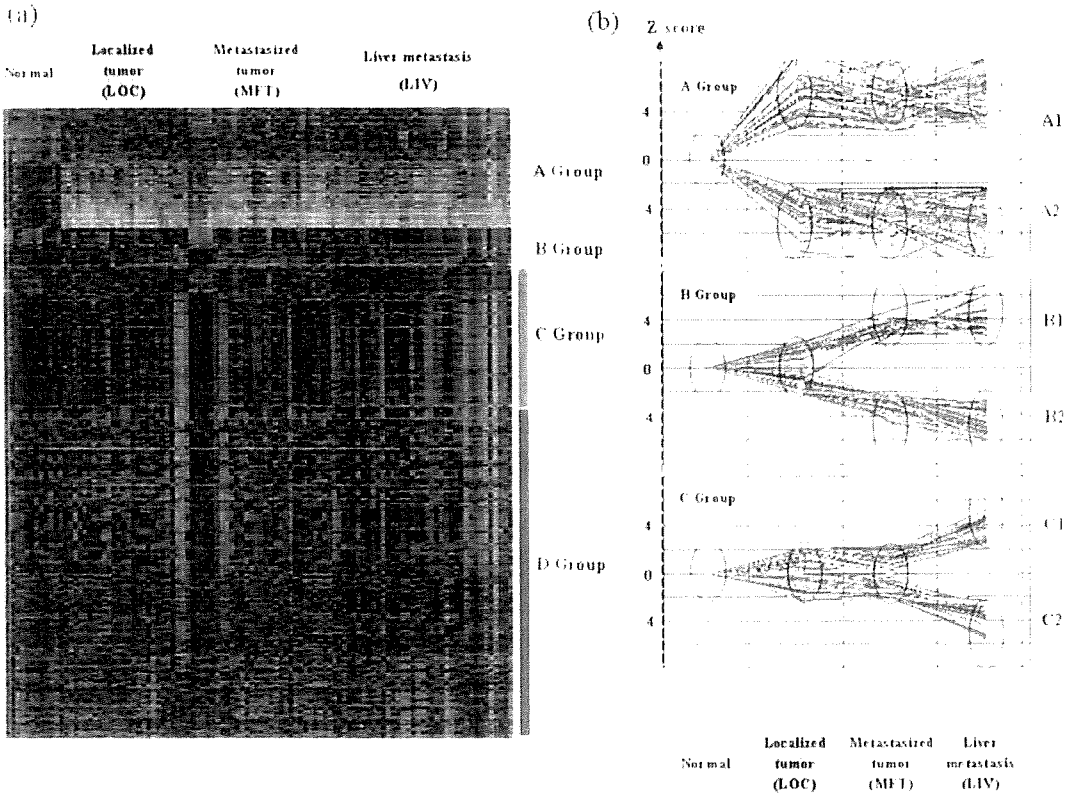

Figure 1. Gene expression patterns associated with the metastatic process. (a) The columns represent 12 normal colon samples and 92 colorectal tumor samples [32 localized tumors (LOC), 26 metastasized tumors (MET) and 34 liver metastatic tumors (LIV)]; the rows represent 1,953 genes. Differentially expressed 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.