

ORIGINAL PAPER

Different gene expression profiles between microsatellite instability-high and microsatellite stable colorectal carcinomasHyunki Kim^{1,2,8}, Suk Woo Nam^{3,6,8}, Hwanseok Rhee², Long Shan Li^{1,2}, Hyun Ju Kang^{1,2}, Kwi Hye Koh^{1,2}, Nam Kyu Kim⁴, Jaehwi Song^{3,7}, Edison Tak-Bun Liu⁵ and Hoguen Kim^{*1,3}

¹Department of Pathology, Yonsei University College of Medicine, Seoul 120-752, Korea; ²Brain Korea 21 Projects for Medical Sciences, Yonsei University College of Medicine, Seoul 120-752, Korea; ³Cancer Metastasis Research Center, Yonsei University College of Medicine, Seoul 120-752, Korea; ⁴Department of Surgery, Yonsei University College of Medicine, Seoul 120-752, Korea; ⁵Genome Institute of Singapore, 138672, Singapore; ⁶Department of Pathology, College of Medicine, The Catholic University of Korea, Seoul 137-710, Korea; ⁷Department of Genetic Engineering, College of Life Science and Natural Resources, Sungkyunkwan University, Suwon 440-746, Korea

Recent molecular genetic studies have revealed that two major types of genomic instabilities, chromosomal instability (CIN) and microsatellite instability (MSI), exist in colorectal carcinomas. In order to clarify the molecular signature related to the CIN and MSI in colorectal carcinomas, we performed transcriptomic expression analysis on eight microsatellite instability-high (MSI-H) colorectal carcinomas and compared the results obtained with that of nine microsatellite stable (MSS) colorectal carcinomas using oligonucleotide microarrays containing 17 334 known genes and 1331 unknown genes or expression sequence tags (ESTs). Unsupervised two-way hierarchical clustering with 5724 genes successfully classified tumors from normal mucosa, and displayed a distinctive MSI-H carcinomas subgroup. Based on intensive filtering, 57 known genes and eight ESTs were found to be highly relevant to the differentiation of MSI-H and MSS colorectal carcinomas. These genes successfully distinguish the new test set of six MSI-H and five MSS colorectal carcinomas. Many up- and downregulated genes in MSI-H colorectal carcinomas were related to the previously reported phenotypic characteristics; increased mucin production and intense peritumoral immune response in MSI-H carcinomas. Some of these differences were confirmed by semiquantitative reverse transcription-PCR and immunohistochemical analysis. Our findings indicate that there are many different genetic and transcriptomic characteristics between MSI-H and MSS colorectal carcinomas, and some of these differently expressed genes can be used as diagnostic or prognostic markers.

Oncogene advance online publication, 21 June 2004; doi:10.1038/sj.onc.1207853

Keywords: colorectal carcinomas; microsatellite instability; oligonucleotide microarray; gene expression profile; molecular classification

Introduction

The molecular genetics of colorectal carcinomas are among the best understood of common human cancers. Recent molecular genetic studies have revealed that two major types of genomic instabilities, chromosomal instability (CIN) and microsatellite instability (MSI), exist in colorectal carcinomas (Lengauer *et al.*, 1998). The majority of colorectal carcinomas are assigned to the CIN pathway, which is characterized by a high frequency of allelic losses, deletions and/or mutations of tumor suppressor genes, and an abnormal tumor DNA content (Kinzler and Vogelstein, 1996). The mechanism of tumorigenesis in CIN tumors involves the activation of oncogenes and the inactivation of tumor suppressor genes. The loss of one allele and the inactivation of the other allele by mutation or promoter methylation are accepted as a general mechanism of tumor suppressor gene inactivation.

The other pathway, the MSI pathway, begins with the inactivation of one of a group of genes responsible for DNA nucleotide mismatch repair, which leads to extensive mutations in both repetitive and nonrepetitive DNA sequences with low frequencies of allelic losses and rare alterations of tumor DNA content (Ionov *et al.*, 1993; Thibodeau *et al.*, 1993). The mechanism of tumorigenesis in high microsatellite instability (MSI-H) tumors is thought to involve frameshift mutations of microsatellite repeats within coding regions of the affected target genes, and the inactivation of these target genes is believed to directly contribute to tumor development and progression (Duval and Hamelin, 2002).

In addition to different genetic changes during the course of tumor development, different tumor progression patterns have been reported in CIN and MSI-H colorectal carcinomas. CIN colorectal carcinomas usually follow the classical adenoma–carcinoma sequence model and belong to the microsatellite stable (MSS) carcinomas. In contrast, adenoma formation in MSI-H colorectal carcinomas is infrequent, and the carcinomas that follow the MSI-H pathway reveal

*Correspondence: H Kim, Department of Pathology, Yonsei University College of Medicine, CPO Box 8044, Seoul, Korea; E-mail: hkyonsei@yumc.yonsei.ac.kr

⁸Contributed equally to this study

Received 2 February 2004; revised 9 April 2004; accepted 28 April 2004

distinct clinicopathologic patterns (Kim *et al.*, 1994; Alexander *et al.*, 2001). This is a complex and dynamic process, which is expected to involve genomic changes in many genes and altered gene expression profiles. The identification of genes and of gene expression profiles that contribute to these two different genetic pathways should significantly improve both tumor classification and therapy. To address this issue, we undertook a comprehensive genomic approach in eight MSI-H and nine MSS colorectal carcinomas by using oligonucleotide microarray. The molecular dissection of genes differently expressed in colorectal carcinomas was provoked by the knowledge that MSI-H colorectal carcinomas are distinguished from MSS carcinomas. Here, we demonstrate a molecular signature responsible for MSI in colorectal carcinoma.

Results

Hierarchical clustering analysis identifies two tumor subsets characterizing MSI in colorectal carcinoma

We analysed the gene expression of colorectal carcinomas and matched normal mucosae by using oligonucleotide microarray. We selected eight MSI-H (cases 1–8) and 10 MSS colorectal carcinomas (cases 9–18). The relative expression of each gene in tumor and normal tissue was measured by comparing its expression ratio to that of Universal Human Reference RNA (Stratagene). Adequate expression data could not be obtained from one MSS carcinoma (case 18) and three matched normal mucosae (cases 8, 12, and 17) due to artefacts in the hybridized arrays, or poor quality RNA. Thus, these were excluded from the analysis. We initially tried molecular pattern analysis to determine whether our spotted-oligoarray system was able to identify normal mucosae and colorectal carcinoma by molecular profil-

ing. With a relevant set of prefiltered 5724 genes (see ‘Materials and methods’), we conducted a complete-linkage hierarchical clustering analysis of 32 arrays (17 carcinomas and 15 normal mucosae). The results obtained from these 5724 genes in terms of the greatest expressional differences between the 17 colorectal carcinomas and the 15 normal mucosae samples are displayed in Figure 1. This result showed that normal mucosae and tumors were grouped together on a cluster dendrogram, implying that thousands of genes either contribute or are affected by colonic tumorigenesis. We next tried to identify a robust set of tumor-related genes by supervised rank-sum analysis by using the Mann–Whitney rank-sum test. Genes that were differently regulated in colorectal carcinoma were identified using stringent selection criteria (See ‘Material and methods’). Using these selection criteria, 520 genes were selected as being upregulated (285 genes) or downregulated (235 genes) compared to noncancerous mucosae. Supplementary Table 1 lists genes differentially expressed in normal mucosae and carcinomas.

Furthermore, we found that the carcinomas were clustered into two unique subgroups and that the MSI-H carcinomas were differentiated from the other carcinomas (Figure 1), thus indicating that the MSI-H carcinomas can be distinguished by their intrinsic molecular signatures. It is interesting that two distinctive subgroups in cluster analysis of tumors were correlated with MSI-H and MSS. Therefore, we next explored the molecular signature contributing to MSI-H in colorectal carcinoma.

Identification of differentially expressed genes in MSI-H and MSS carcinomas

We compared the relative expressions of each gene in the tumors in two ways. First, we compared the relative

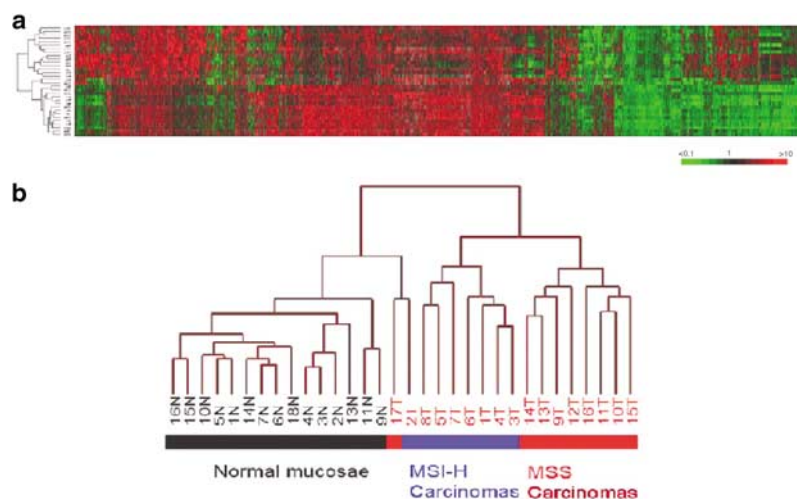


Figure 1 Unsupervised hierarchical clustering analysis of 17 colorectal carcinomas and 15 normal mucosae according to the gene expression. (a) Genes that passed the filtering criteria were used (gene expression values present in more than 75% in all arrays were taken, and genes having standard deviations of less than 0.35 were discarded). A total of 5724 genes were selected and applied to complete linkage hierarchical clustering analysis using the uncentered correlation similarity metric method. Red and green indicate transcript level above and below the median sample: Universal Human Reference RNA expression ratio for all each gene across all sample, respectively. (b) Separation of normal mucosal tissues and tumors, and separation of MSI-H and MSS tumors are evident

Table 1 List of upregulated genes in MSI-H carcinomas

UGcluster	Symbol	Gene name	MSS-H/MSS ^a	MSI-H/N ^b	MSS/N ^c	P-value ^d
Hs.76422	PLA2G2A	Phospholipase A2, group IIA (platelets, synovial fluid)	7.28	1.82	0.25	0.003
Hs.105806	GPLY	Granulysin	6.16	7.10	1.15	0.001
Hs.89603	MUC1	mucin 1, transmembrane	3.35	2.10	0.63	0.009
Hs.380933	MUC2	mRNA; cDNA DKFZp586O1224 (from clone DKFZp586O1224), mRNA sequence	3.25	5.48	1.69	0.001
Hs.91011	AGR2	Anterior gradient 2 homolog (<i>Xenopus laevis</i>)	3.14	2.85	0.91	0.009
Hs.103707	MUC5AC	Mucin 5, subtypes A and C, tracheobronchial/gastric	2.96	2.74	0.93	0.007
Hs.119140	EIF5A	Eucaryotic translation initiation factor 5A	2.92	4.34	1.49	0.001
Hs.15114	ARHD	ras homolog gene family, member D	2.85	1.70	0.60	0.005
Hs.18760	KIAA1389	KIAA1389 protein (<i>Homo sapiens</i>), mRNA sequence	2.73	1.86	0.68	0.003
Hs.343628	SIAT4B	Sialyltransferase 4B (beta-galactoside alpha-2,3-sialyltransferase)	2.71	1.39	0.51	0.003
Hs.118786	MT2A	Metallothionein 2A	2.70	0.49	0.18	0.001
Hs.59413	CTSL	Cathepsin L	2.65	2.13	0.81	0.009
Hs.14623	IFI30	Interferon, gamma-inducible protein 30	2.62	3.45	1.32	0.003
Hs.173043	MTA1L1	Metastasis-associated 1-like 1	2.62	2.91	1.11	0.001
Hs.8986	CIQB	cDNA FLJ12124 fis, clone MAMMA1000139	2.58	2.23	0.87	0.005
Hs.110796	SARI	Complement component 1, q subcomponent, beta polypeptide	2.56	1.88	0.73	0.001
Hs.270737	TNFSF13B	SARI protein	2.56	2.61	1.02	0.002
Hs.337461	TRIM15	Tumor necrosis factor (ligand) superfamily, member 13b	2.46	1.93	0.78	0.005
Hs.169401	APOE	Tripartite motif-containing 15	2.45	1.09	0.45	0.006
Hs.433300	FCER1G	Apolipoprotein E	2.39	2.26	0.94	0.003
Hs.2730	HNRPL	Fc fragment of IgE, high-affinity I, receptor for gamma polypeptide	2.35	2.64	1.12	0.003
Hs.151734	SECTM1	Heterogeneous nuclear ribonucleoprotein L	2.34	2.96	1.26	0.001
Hs.95655	NUTF2	Secreted and transmembrane 1	2.33	0.63	0.27	0.001
Hs.184390	LOC57168	Nuclear transport factor 2	2.33	0.27	0.11	0.007
Hs.13015	DNAJC1	Similar to aspartate beta hydroxylase (ASPH)	2.30	2.05	0.89	0.009
Hs.104741	TOPK	DnaJ (Hsp40) homolog, subfamily C, member 1	2.30	0.95	0.41	0.002
Hs.380778	MT1L	T-LAK cell-originated protein kinase	2.28	3.48	1.52	0.001
Hs.75367	SLA	Metallothionein 1L	2.23	0.53	0.24	0.004
Hs.325978	IL18BP	Sre-like adaptor	2.19	1.25	0.57	0.002
Hs.55968	GALNT5	Interleukin 18-binding protein	2.12	2.00	0.94	0.004
Hs.3164	NUCB2	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 5 (GalNAc-T5)	2.09	2.03	0.97	0.005
Hs.198248	B4GALT1	Nucleobindin 2	2.08	1.78	0.86	0.003
Hs.91448	DUSP14	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 1	2.07	1.00	0.48	0.001
Hs.1695	MMP12	Dual-specificity phosphatase 14	2.06	1.81	0.88	0.005
Hs.237856	PHT2	Matrix metalloproteinase 12 (macrophage elastase)	2.05	1.94	0.95	0.007
Hs.73010	IFNW1	Peptide transporter 3	2.05	2.10	1.02	0.005
Hs.75703	CCL4	Interferon, omega 1	2.04	2.75	1.35	0.001
Hs.75627	CD14	Chemokine (C-C motif) ligand 4	2.03	1.72	0.85	0.001
		CD14 antigen	2.01	1.32	0.66	0.006

^aMSS-H/MSS, average ratio of intensity of eight MSI-H carcinomas/average ratio of nine MSS carcinomas. ^bMSI-H/N, average ratio of intensity of eight MSI-H carcinomas/average ratio of intensity of 15 normal mucosae. ^cMSS/N, average ratio of intensity of nine MSS carcinomas/average ratio of intensity of 15 normal mucosae. ^dP-value, between eight MSI-H carcinomas and nine MSS carcinomas from Mann-Whitney rank-sum test

gene expressions of carcinomas with respect to Universal Human Reference RNA, and second we compared the gene expressions of each tumor by analysing the differences in the relative gene expression of each gene with respect to the average expressions in 15 matched normal mucosal tissues. Both methods produced the same result. Based on the level of gene expression, colorectal carcinomas were separated into two major groups. The seven MSI-H and eight MSS carcinomas showed strong tendency to form separate clusters on two distinct dendrogram branches, and one MSI-H and one MSS carcinomas formed a minor branch (Figure 1). Comparison with the cluster dendrogram showed several groups of genes common or specific to the genetic instability type. Supervised analysis identified 39 genes showing at least twofold higher mRNA expressions (Table 1), and 26 genes showing at least 0.5-fold lower expressions in MSI-H carcinomas compared with MSS carcinomas (Table 2) (Figure 2a).

Validation test of 65 genes as molecular classifiers for MSI-H colorectal carcinomas

In order to determine whether these 65 outlier genes can be used as unique molecular signatures discriminating the MSI-H colorectal carcinomas, we tested

for expression profiling of another set of 12 colorectal carcinomas including six MSI-H and six MSS carcinomas using new batch of spotted-oligonucleotide microarrays. Note that these arrays were manufactured at the DNA microarray core facility of the Department of Pathology, College of medicine, The Catholic University of Korea, but containing the same genetic elements as the previous ones. In this test, we noted that two genes out of a total of 65 outlier genes were missing during primary data extraction probably due to spotting error or stringent selection criteria. We then performed hierarchical cluster analysis of the 11 colorectal carcinomas by using selected 63 outlier genes (one array was eliminated due to poor quality of hybridization result). As expected, the classification of 63 outlier genes resulted in two distinct subgroups on cluster dendrogram, exhibiting a clear separation between all six MSI-H and five MSS carcinomas. This implies that, at least, our 63 outlier genes could be feasible genetic elements for the molecular classification of MSS and MSI colorectal carcinomas (Supplementary Figure 1).

Validation of differentially expressed genes

In order to examine the reliability of microarray data, we selected two upregulated genes, *mucin 1 (MUC1)* and

Table 2 List of downregulated genes in MSI-H carcinomas

UGCluster	Symbol	Gene name	MSI-H/MSS ^a	MSI-H/N ^b	MSS/N ^c	P value ^d
Hs.24395	CXCL14	Chemokine (C-X-C motif) ligand 14	0.23	0.29	1.23	0.004
Hs.87296		cDNA FLJ20269 fis, clone HEP01293, mRNA sequence	0.26	1.15	4.37	0.005
Hs.9029	HAIK1	Type I intermediate filament cytokeratin	0.29	1.08	3.80	0.002
Hs.22785	GABRE	Gamma-aminobutyric acid (GABA) A receptor, epsilon	0.32	0.73	2.24	0.001
Hs.282975	CES2	Carboxylesterase 2 (intestine, liver)	0.32	0.16	0.51	0.004
Hs.127337	AXIN2	Axin 2 (conductin, axil)	0.34	1.25	3.65	0.007
Hs.179704	MEP1A	Meprin A, alpha (PABA peptide hydrolase)	0.36	0.10	0.29	0.005
Hs.166705	GPR49	G protein-coupled receptor 49	0.39	1.19	3.03	0.006
Hs.275775	SEPP1	Selenoprotein P, plasma, 1	0.44	0.12	0.27	0.009
Hs.101850	RBP1	Retinol-binding protein 1, cellular	0.44	0.82	1.87	0.005
Hs.38738	CLDN15	Claudin 15	0.44	0.75	1.72	0.004
Hs.278997	CESR	Carboxylesterase-related protein	0.44	0.55	1.26	0.001
Hs.272245		cDNA FLJ11170 fis, clone PLACE1007301, mRNA sequence	0.45	0.29	0.65	0.010
Hs.248019	POU4F3	POU domain, class 4, transcription factor 3	0.45	0.67	1.51	0.007
Hs.49476		cDNA FLJ12815 fis, clone NT2RP2002546, mRNA sequence	0.46	0.54	1.15	0.009
Hs.1298	MME	Membrane metallo-endopeptidase (neutral endopeptidase, enkephalinase, CALLA, CD10)	0.46	0.84	1.82	0.002
Hs.86327	HOXB9	Homeo box B9	0.46	0.76	1.66	0.004
Hs.18457	FLJ20315	Hypothetical protein FLJ20315	0.47	1.28	2.71	0.002
Hs.151469	CASK	Calcium/calmodulin-dependent serine protein kinase (MAGUK family)	0.47	0.96	2.06	0.002
Hs.130714		HSPC323 (<i>Homo sapiens</i>), mRNA sequence	0.47	0.62	1.32	0.007
Hs.192927	PPP1R14D	Protein phosphatase 1, regulatory (inhibitor) subunit 14D	0.47	0.31	0.66	0.004
Hs.36927	HSP105B	Heat shock 105 kDa	0.48	2.50	5.18	0.005
Hs.274351	ZDHHC9	Zinc-finger, DHH domain containing 9	0.48	1.00	2.08	0.001
Hs.151301	CADPS	Ca ²⁺ -dependent activator protein for secretion	0.48	1.07	2.22	0.002
Hs.380831	FOXO3A	Forkhead box O3A	0.48	0.54	1.13	0.001
Hs.85835		cDNA: FLJ22841 fis, clone KAIA4844, mRNA sequence	0.49	0.31	0.63	0.001

^aMSI-H/MSS, average ratio of intensity of eight MSI-H carcinomas/average ratio of intensity of nine MSS carcinomas. ^bMSI-H/N, average ratio of intensity of eight MSI-H carcinomas/average ratio of intensity of 15 normal mucosae. ^cMSS/N, average ratio of intensity of nine MSS carcinomas/average ratio of intensity of 15 normal mucosae. ^dP-value, between eight MSI-H carcinomas and nine MSS carcinomas from Mann-Whitney rank-sum test

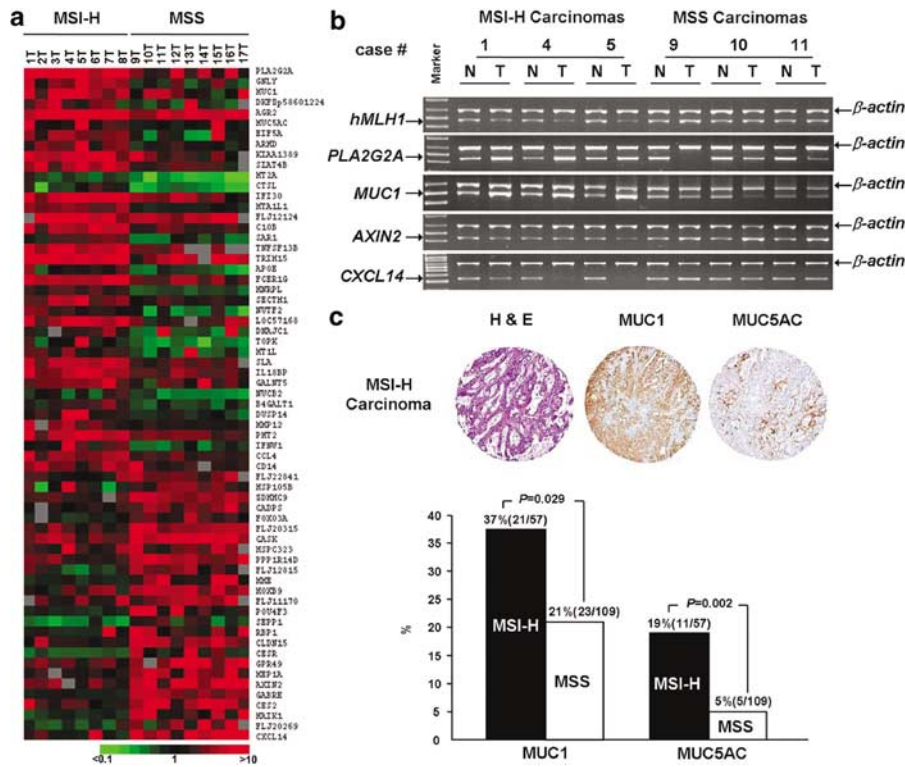


Figure 2 Expression analysis of differentially expressed genes between MSI-H and MSS carcinomas. **(a)** Transcriptional expression profile of 65 differently expressed genes between MSI-H and MSS carcinomas. Red and green indicate transcript level above and below the median tumor: Universal Human Reference RNA expression ratio for 65 genes, respectively. **(b)** Expression analysis of *hMLH1*, *PLA2G2A*, *MUC1*, *AXIN2*, and *CXCL14* by semiquantitative RT-PCR. Downregulation of *hMLH1* in MSI-H carcinomas is evident. Upregulations of *PLA2G2A* and *MUC1*, and downregulations of *CXCL14* are found in MSI-H carcinomas. The downregulation of *PLA2G2A* and the upregulation of *AXIN2* in MSS carcinomas are also shown. **(c)** Tumor histology and immunohistochemical analysis of MUC1 and MUC5AC in one MSI-H colorectal carcinoma. Increased expressions of MUC1 and MUC5AC are evident. Summary of MUC1 and MUC5AC expressions in the 57 MSI-H and 109 MSS colorectal carcinomas. In the MSI-H carcinomas, frequent expressions of MUC1 and MUC5AC are evident

phospholipase A2 group IIA (PLA2G2A) and one downregulated gene, *chemokine (C-X-C motif) ligand 14 (CXCL14)* in MSI-H carcinomas, and one upregulated gene *AXIN2* in MSS carcinomas. The expressions of these genes were analysed by semiquantitative reverse transcription (RT)-PCR using the RNA samples used for the microarray analysis. The results of semiquantitative RT-PCR were consistent for all the four genes (Figure 2b). We then performed an immunohistochemical analysis of MUC1 and mucin 5, subtypes A and C (MUC5AC) by using tissue array containing 57 MSI-H and 109 MSS carcinomas and their matched normal mucosae, then compared the results to that of microarray data. Both genes showed similar levels of mRNA and protein expression. The difference in immunohistochemical expression patterns between MSI-H and MSS colorectal carcinomas were consistently demonstrated in the 166 colorectal carcinomas. Expressions of MUC1 and MUC5AC were increased in MSI-H colorectal carcinomas (Figure 2c). MUC1 expression was found in 21 of 57 MSI-H and 23 of 109 MSS carcinomas ($P < 0.029$), and MUC5AC expression was found in 11 of 57 MSI-H and five of 109 MSS carcinomas ($P < 0.002$).

Discussion

In this study, we identified distinct gene expression profiles in MSI-H and MSS colorectal carcinomas by oligonucleotide microarray analysis. We identified a large number of genes that were differentially expressed in normal colonic mucosae and colorectal carcinomas. We also found two unique subgroups within carcinomas, namely, the MSI-H and MSS carcinomas, which showed different hierarchical cluster patterns. These data provide an insight into the extent of gene expression differences underlying different genetic instabilities in colorectal carcinomas (MSI-H vs MSS).

We identified 520 (2.8% of those detected) genes or expression sequence tag (ESTs) that appear to be differentially expressed (more than twofold and less than 0.5-fold) in carcinomas vs normal appearing mucosa. Many of these dysregulated genes at the transcript level were in accordance with previous reports (Alon *et al.*, 1999; Kitahara *et al.*, 2001; Notterman *et al.*, 2001; Birkenkamp-Demtroder *et al.*, 2002; Bertucci *et al.*, 2004). For example, *CA4*, *CNN1*, and

FCGBP were downregulated, whereas *CCNB2*, *CSE1L*, *NME1*, *RHEB2*, and *UBE2C* were upregulated in carcinomas. The functional categories of dysregulated genes in carcinomas revealed that the genes involved in nucleic acid metabolism and cell proliferation are overexpressed in carcinomas. The 285 upregulated genes in carcinomas belonged to the functional categories (<http://www.geneontology.org>) of nucleic acid metabolism (10%), cell proliferation (10%), signal transduction (7%), and biosynthesis (6%). The 235 downregulated genes in carcinomas belonged to transport (11%), signal transduction (8%), response to biotic stimulus (6%), and lipid metabolism (4%).

We successfully identified differential gene expression in two distinct subgroups of colorectal carcinomas, MSI-H and MSS carcinomas, by genome wide gene expression analysis. MSI-H carcinomas arise with the inactivation of one of a group of genes responsible for DNA nucleotide mismatch repair, which leads to extensive mutations in both repetitive and nonrepetitive DNA sequences with low frequencies of allelic losses and rare alterations of tumor DNA content (Ionov *et al.*, 1993; Thibodeau *et al.*, 1993). In contrast, the MSS carcinomas are characterized by a high frequency of allelic losses, deletions and/or mutations of tumor suppressor genes, and abnormal tumor DNA content (Kinzler and Vogelstein, 1996). Several clinicopathologic findings are also reported to be different in MSI-H and MSS carcinomas. MSI-H colorectal carcinomas show a preponderance for the right side, poor differentiation, mucin formation, and peritumoral lymphocytic infiltration (Kim *et al.*, 1994; Alexander *et al.*, 2001). However, these clinicopathologic features did not account for the separation of MSI-H and MSS colorectal carcinomas (Alexander *et al.*, 2001; Ward *et al.*, 2001). Based on these distinct molecular and clinical natures, MSI-H and MSS tumors are expected to be distinguished by an expression profiling. Also, it is possible to identify novel differentially expressed genes that may be related to the MSI-H or CIN molecular pathway. Currently, two studies are available for the gene expression differences between MSI-H and MSS tumors. One study used colorectal cancer cell lines (Dunican *et al.*, 2002) and the other study used tumor tissues (Mori *et al.*, 2003). Both studies reported a number of differentially expressed genes between MSI-H and MSS tumors. However, the number of reported genes was small and the usefulness of these genes as molecular classifiers is not defined statistically. In this study, we found that MSI-H and MSS carcinomas can be divided by unsupervised hierarchical clustering analysis, and identified 65 genes that completely differentiate MSI-H and MSS carcinomas by supervised analysis. These genes successfully differentiated the other test set of 11 colorectal carcinomas as MSI-H and MSS carcinomas. These findings indicate that MSI-H and MSS carcinomas are distinct types of colorectal carcinomas, and the 65 genes can be used as molecular classifiers.

Many of the dysregulated genes in MSI-H and MSS carcinomas reported herein make a good deal of sense even though some of them are unsuspected in colorectal carcinogenesis. For example, *MUC5AC* and *MUC1* are involved in mucin production, thus their increased expression is in good agreement with the high frequency of mucin production in MSI-H carcinomas (Kim *et al.*, 1994). Moreover, *MUC1* expression in colorectal carcinoma has been reported to be implicated in tumor progression and invasion in colorectal carcinoma (Hiraga *et al.*, 1998). An association between *MUC5AC* overexpression and MSI-H colorectal carcinomas had been reported (Biemer-Huttman *et al.*, 2000). These findings suggest that the proportions of the mucin component are different between MSI-H and MSS carcinomas. Increased granulysin, a protein present in the cytotoxic granules of cytotoxic T lymphocytes and natural killer cells (Stenger *et al.*, 1999), in MSI-H tumors might represent infiltrating intratumoral T cell in our MSI-H carcinomas. Intratumoral and peritumoral lymphocytic infiltration occurs frequently in MSI-H carcinomas, and a large proportion of cytotoxic T cells among infiltrating lymphocytes had been reported (Guidoboni *et al.*, 2001). These findings indicate that the molecular clusters that are involved in the differentiation toward MSI-H carcinomas are closely related with the clinicopathological characteristics of MSI-H carcinomas. The large numbers of molecular types related to the phenotypic alterations of MSI-H carcinomas are believed to enable MSI-H and MSS carcinoma differentiation at the gene level, while the morphologic classification of these tumors using clinicopathologic parameters is difficult (Alexander *et al.*, 2001).

Several genes were found that were related to the APC/ β -catenin pathway. *PLA2G2A* was downregulated and *AXIN2* upregulated in MSS carcinomas. These altered expressions are interesting because these two molecules may be directly related with the abnormal APC/ β -catenin pathway in MSS carcinomas. In APC^{Min/+} mouse, a mouse model of familial adenomatous polyposis, *PLA2G2A* was proposed to be a major modifier, which suppress the severity of intestinal neoplasia (Cormier *et al.*, 1997; MacPhee *et al.*, 1995). This finding suggests that the abnormal APC/ β -catenin pathway can be enhanced by the reduced expression of *PLA2G2A* in MSS colorectal carcinomas. It has been demonstrated that AXIN1 and/or AXIN2 form a multiprotein complex with APC, glycogen synthase kinase 3 β , and β -catenin (Ikeda *et al.*, 1998). Moreover, *AXIN2* induction by β -catenin activation has been reported (Leung *et al.*, 2002). Therefore, the specific overexpression of *AXIN2* in our MSS carcinomas reflects the selected activation of the APC/ β -catenin pathway in MSS carcinomas.

In summary, we identified several candidate colorectal carcinoma markers and molecular clusters capable of differentiating MSI-H and MSS colorectal carcinomas. Moreover, genes like these, which help to define specific genomic instability, may have potential values in differential diagnosis and prognosis evaluation.

Materials and methods

Case selection

Eight cases (cases 1–8), confirmed as MSI-H colorectal carcinoma, and 10 cases (cases 9–18) of MSS colorectal carcinoma were included in this study. In each case, grossly normal mucosa remote from the tumor was included as a control. All cases were selected from consecutively identified cases at the Gastrointestinal Tumor Working Group Tissue Bank at Yonsei University Medical Center (Seoul, Korea) between December 1996 and November 1999. We also selected six MSI-H and six MSS carcinomas for the test set. These cases were also selected from the Gastrointestinal Tumor Working Group Tissue Bank at Yonsei University Medical Center between December 2000 and November 2003. RNAs, proteins, and DNAs were extracted from fresh frozen tissues. Tumor specimens were microdissected on a cryostat and fractionated to enrich the tumor cell population. Genomic DNA preparation and the determination of the MSI status of eight MSI-H carcinomas have been previously reported (Kim *et al.*, 2001; Kim *et al.*, 2002).

Microarray formulation

High-density spotted-oligonucleotide microarrays were manufactured at the array core facility at the Genome Institute of Singapore. The human Oligolibrary™ was purchased from Compugen/Sigma-Genosys. It consisted of 18 861 oligonucleotides, which represent 18 664 LEADS™ clusters plus 197 controls *glyceraldehydes-3-phosphate dehydrogenase*. A total of 60 mers of synthesized oligos were robotically printed and processed. For analysis of the test set, oligonucleotide arrays containing the same genetic elements were manufactured at the core facility of Department of Pathology, College of Medicine, The Catholic University of Korea.

RNA preparation and hybridization

Total RNA was extracted from 100–200 mg of microdissected frozen tissues using a RNeasy Mini kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. Total RNA (20 µg) was used as input for cDNA targets synthesis as previously described (DeRisi *et al.*, 1997). The targets and Universal Human Reference RNA (Stratagene, La Jolla, CA, USA) were hybridized to an oligonucleotide microarray containing 18 664 probe sets representing 18 664 unique (LEADS™) genes, and the array was scanned using GenePix scanners. Expression values for each gene were calculated by using GenePix Pro 4.0 analysis software. Some of the hybridizations were carried out in duplicate with flouoreversal to compensate for the different chemical properties of the fluorescent dye molecules and for potential biases associated with normalization.

Of 18 colorectal carcinomas from 18 patients samples processed, 17 colorectal carcinomas and 15 corresponding normal mucosae were used for subsequent data analysis. The remaining one tumor and three normal samples either failed quality control restrictions concerning the amount and quality of RNA, as assessed by agarose gel electrophoresis, or yielded poor-quality scans.

Classification of colorectal carcinoma by molecular pattern analysis

Unsupervised hierarchical clustering analysis was used to analyse the classifications of 17 colorectal carcinomas and of 15 noncancerous mucosae according to gene expression. We

used a data set of genes that satisfied the filtering criteria (genes having more than 75% of log-transformed ratio values presenting in across all arrays were chosen and genes having less than 0.35 standard deviations of log-transformed ratio were discarded). The selected gene data set was then applied to complete-linkage hierarchical clustering analysis using the uncentered correlation similarity metric method in Cluster version 2.20, and the resulting expression map was visualized with Treeview version 1.60 (<http://rana.lbl.gov/EisenSoftware.htm>).

Identification of genes responsible for MSI in colorectal carcinoma

To determine whether colorectal carcinomas can be classified as MIS-H or MSS according to their transcriptomic patterns, 17 arrays (eight arrays of MSI-H and nine of MSS) of colorectal carcinomas were assessed using unsupervised hierarchical clustering analysis. To detect differentially expressed genes in a given subclass, we ranked the genes using the Mann–Whitney rank-sum test. Outlier genes responsible for MSI-H and MSS classifications were selected by $P < 0.01$. In addition, significant outlier subset genes were further narrowed by filtering genes showing greater than ± 2 -fold expression changes in colorectal carcinoma compared to the average values from normal mucosae.

Semiquantitative RT-PCR

First-strand cDNA was synthesized from 1 µg of total RNA using random hexamer primers (Qiagen) and M-MLV Reverse Transcriptase (Invitrogen, San Diego, CA, USA) according to the manufacturer's instructions. cDNA (20 ng) from each sample was used in each reaction. All RT-PCR primers were designed to contain an exon–exon junction. Reaction was performed with primer for the specific genes and for the β -actin in duplex reaction. The range of linear amplification for each gene and β -actin was examined with serial PCR cycles, and optimal PCR cycles were determined. For each gene, Gene Bank accession number, the sequences of the forward primer and reverse primer, respectively, are as follows: *human MutL homolog 1 (hMLH1)*, NM_000249, 5'-AGATCACGGTG GAGGACCTT-3', 5'-CCAATCTGTTGGGACTGATT-3'; *PLA2G2A*, NM_000300, 5'-CCTGGGGATACAACCTCTG GA-3', 5'-TTGCACAGGTGATTCTGCTC-3'; *MUC1*, NM_002456, 5'-AGTTCAGGCCAGGATCTGTG-3', 5'-CCCCA ACAAGTTGGCAGAAG-3'; *Axin2*, NM_004655, 5'-GTGT GAGGTCCACGGAAACT-3', 5'-TTCATCCTCTCGGATC TGCT-3'; *CXCL14*, NM_004887, 5'-TGTGGACGGGTC CAAATG-3', 5'-CTGCGCTTCTCGTTCCAG-3'; β -actin, NM_001101, 5'-TGCTATCCCTGTACGCCTCT-3', 5'-GTA CTTGCGCTCAGGAGGAG-3'. β -Actin was used as an internal control. After RT-PCR, 5 µl aliquots of the products were subjected to 2% agarose gel electrophoresis and stained with ethidium bromide.

Immunohistochemical analysis

Colorectal carcinoma tissue arrays containing 57 MSI-H carcinomas and 109 MSS carcinomas were constructed from formalin-fixed and paraffin-embedded tissues by **Petagen Inc. (Seoul, Korea)** and these arrayed slides were used for the immunostaining of MUC1 and MUC5AC. All eight MSI-H carcinomas and nine MSS carcinomas used in the oligonucleotide microarray analysis were included in the tissue array. Deparaffinization and rehydration were performed using xylene and alcohol. The sections were incubated for 1 h at room temperature with antibodies against MUC1

(Novocastra Laboratories Ltd, Newcastle, UK) and MUC5AC (Novocastra Laboratories Ltd). Avidin–biotin complex methodology was employed. The chromogen was diaminobenzidine and counterstaining was carried out with hematoxylin. The expression of the two gene products was categorized as expressed and negative. In the evaluation of MUC5AC, cases with definite cytoplasmic staining in more than 10% of the tumor cells were categorized as expressed, and cases with definite cytoplasmic staining in less than 10% of the tumor cells were categorized as negative. In the evaluation of MUC1, cases with definite membrane and/or cytoplasmic staining in

more than 10% of the tumor cells were categorized as expressed, and less than 10% of the tumor cells or complete absence of staining were as negative.

References

- Alexander J, Watanabe T, Wu TT, Rashid A, Li S and Hamilton SR. (2001). *Am. J. Pathol.*, **158**, 527–535.
- Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D and Levine AJ. (1999). *Proc. Natl. Acad. Sci. USA*, **96**, 6745–6750.
- Bertucci F, Salas S, Eysteris S, Nasser V, Finetti P, Ginestier C, Charafe-Jauffret E, Lloriod 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. (2004). *Oncogene*, **23**, 1377–1391.
- Biemer-Huttmann AE, Walsh MD, McGuckin MA, Simms LA, Young J Leggett BA and Jass JR. (2000). *Clin. Cancer Res.*, **6**, 1909–1916.
- Birkenkamp-Demtroder K, Christensen LL, Olesen SH, Frederiksen CM, Laiho P, Aaltonen LA, Laurberg S, Sorensen FB, Hagemann R and Orntoft TF. (2002). *Cancer Res.*, **62**, 4352–4363.
- Cormier RT, Hong KH, Halberg RB, Hawkins TL, Richardson P, Mulherkar R, Dove WF and Lander E. (1997). *Nat. Genet.*, **17**, 88–91.
- DeRisi JL, Iyer VR and Brown PO. (1997). *Science*, **278**, 680–686.
- Dunican DS, McWilliam P, Tighe O, Parle-McDermott A and Croke DT. (2002). *Oncogene*, **21**, 3253–3257.
- Duval A and Hamelin R. (2002). *Cancer Res.*, **62**, 2447–2454.
- Guidoboni M, Gafa R, Viel A, Doglioni C, Russo A, Santini A, Del Tin L, Macri E, Lanza G, Boiocchi M and Dolcetti R. (2001). *Am. J. Pathol.*, **159**, 297–304.
- Hiraga Y, Tanaka S, Haruma K, Yoshihara M, Sumii K, Kajiyama G, Shimamoto F and Kohno N. (1998). *Oncology*, **55**, 307–319.
- Ikedo S, Kishida S, Yamamoto H, Murai H, Koyama S and Kikuchi A. (1998). *EMBO J.*, **17**, 1371–1384.
- Ionov Y, Peinado MA, Malkhosyan S, Shibata D and Perucho M. (1993). *Nature*, **363**, 558–561.
- Kim H, Jen J, Vogelstein B and Hamilton SR. (1994). *Am. J. Pathol.*, **145**, 148–156.
- Kim N-G., Choi YR, Baek MJ, Kim YH, Kang H, Kim NK, Min JS and Kim H. (2001). *Cancer Res.*, **61**, 36–38.
- Kim N-G, Rhee H, Li LS, Kim H, Lee JS, Kim JH, Kim NK and Kim H. (2002). *Oncogene*, **21**, 5081–5087.
- Kinzler KW and Vogelstein B. (1996). *Cell*, **87**, 159–170.
- Kitahara O, Furukawa Y, Tanaka T, Kihara C, Ono K, Yanagawa R, Nita M E, Takagi T, Nakamura Y and Tsunoda T. (2001). *Cancer Res.*, **61**, 3544–3549.
- Lengauer C, Kinzler KW and Vogelstein B. (1998). *Nature*, **396**, 643–649.
- Leung JY, Kolligs FT, Wu R, Zhai Y, Kuick R, Hanash S, Cho KR and Fearon ER. (2002). *J. Biol. Chem.*, **277**, 21657–21665.
- MacPhee M, Chepenik KP, Liddell RA, Nelson KK, Siracusa LD and Buchberg AM. (1995). *Cell*, **81**, 957–966.
- Mori Y, Selaru FM, Sato F, Yin J, Simms LA, Xu Y, Olaru A, Deacu E, Wang S, Taylor JM, Young J, Leggett B, Jass JR, Abraham JM, Shibata D and Meltzer SJ. (2003). *Cancer Res.*, **63**, 4577–4582.
- Notterman DA, Alon U, Sierk AJ and Levine AJ. (2001). *Cancer Res.*, **61**, 3124–3130.
- Stenger S, Rosat JP, Bloom BR, Krensky AM and Modlin RL. (1999). *Immunol. Today*, **20**, 390–394.
- Thibodeau SN, Bren G and Schaid D. (1993). *Science*, **260**, 816–819.
- Ward R, Meagher A, Tomlinson I, O'Connor T, Norrie M, Wu R and Hawkins N. (2001). *Gut*, **48**, 821–829.

Supplementary Information accompanies the paper on Oncogene website (<http://www.nature.com/onc>).