Gene Copy Number Change Events at Chromosome 20 and Their Association with Recurrence in Gastric Cancer Patients

Sang Hwa Yang, Min Young Seo, Ha Jin Jeong, et al.

Clin Cancer Res 2005;11:612-620.

Gene Copy Number Change Events at Chromosome 20 and Their Association with Recurrence in Gastric Cancer Patients

Sang Hwa Yang,[1] Min Young Seo,[1,2,3] Ha Jin Jeong,[1,2,3] Hei-Cheul Jeung,[1,2,3] Jihye Shin,[1,2,3] Sang Chul Kim,[1,2,3,4] Sung Hoon Noh,[1,3,4,6] Hyun Cheol Chung,[1,2,3,4,5] and Sun Young Rha[1,2,3,4]

[1]Cancer Metastasis Research Center, [2]Yonsei Cancer Center, [3]Yonsei Cancer Research Institute, [4]Brain Korea 21 Project for Medical Sciences, [5]Departments of Internal Medicine, and [6]Surgery, Yonsei University College of Medicine, Seoul, Korea

ABSTRACT

Purpose: **This study examined the gene copy number change events at chromosome 20 in gastric cancer, and their possible relationship with recurrence using cDNA microarray-based comparative genomic hybridization.**

Experimental Design: **Thirty pairs of gastric tumor and normal gastric tissues were used in the cDNA microarray-based comparative genomic hybridization. The cDNA microarrays containing 17,000 sequence-verified human gene probes were used in a direct comparison design, where genomic DNAs from the normal and tumor tissues were labeled with fluorescent dyes Cy3 and Cy5, respectively, and cohybridized. Genes with $\log_2$ (Cy5/Cy3) $\geq$ 0.58 in at least one case were selected as the amplified genes. In order to search for the association between gene copy number changes and the recurrence status, patients were grouped according to their recurrence status. Gene selection between the two groups was done, and each patient was given a score based on the sum of the selected genes' ratios. Logistic regression analysis was carried out in order to determine if the score of a group of patients was correlated with a recurrence.**

Results: **A group of genes including NCOA6, CYP24A1, PTPN1, and ZNF217 was amplified in gastric cancer. Another group of 39 genes, whose sum of copy number change levels was significantly associated with a poor prognosis for recurrence, was selected ($P < 0.05$).**

Conclusion: **Ninety-six amplified genes at chromosome 20 of gastric cancer are reported. A scoring system based on gene copy changes at chromosome 20 can provide an independent patient grouping system that can distinguish patient recurrence status and survival.**

INTRODUCTION

Even with the dramatic decline in the incidence of gastric cancer in developed countries, it still remains a leading cause of cancer death worldwide (1). Even with the rapid advancement of molecular biological techniques, there has been slow progress in the understanding of the molecular mechanism of gastric carcinogenesis. This is partly due to the multistep nature of the carcinogenesis and the need to probe the expression and interaction of a large number of genes and proteins simultaneously. Genome-wide monitoring of both of the gene expression and gene copy number changes has now become possible due to microarray technology. In gastric cancer genomic research, microarrays spotted with more than 20,000 human genes are routinely being used to monitor the expression pattern of those genes in a single experiment. Oligonucleotide or cDNA microarrays have been used for a comprehensive survey of the gene expression profiles of the gastric carcinoma cell lines (2), to identify the gastric adenocarcinoma metastasis-related genes in the cell lines derived from a primary tumor and its metastasis counterpart (3, 4), and to select the gene sets that can distinguish between the normal and tumor gastric tissues, or the subtypes of gastric cancer based on gene expression profiling (5–9).

Another important tool for understanding the various types of tumorigenesis has been comparative genomic hybridization (CGH), which traditionally employed whole chromosomes (10, 11). Even though these chromosome-scale CGHs have been informative, the low resolution of sometimes megabases-long inserts has led to the development of matrix-based CGH using P1, PAC clones, where the insert sizes range from 75 to 130 kb in length (12). A breakthrough in the field of CGH was made when Pollack et al. (13, 14) applied cDNA microarray technology to probe the gene copy changes in breast cancer.

Many groups have reported genome instabilities related with gastric cancer. A recent report by Weiss et al. (15) specifically determined an amplicon boundary at 20q13.2 region using BAC clone-based microarrays. The amplified region was shown to be within an ~800-kb-long region of 20q13.2. Two of the well-known oncogene candidates, ZNF217 and CYP24, were located within this region. However, many of these BAC array-based CGH studies have a low resolution and few have searched for an association between gene copy number change and any clinical information.

In order to delineate the individual genes that undergo copy number changes in gastric cancer, a cDNA microarray-based CGH with 30 pairs of normal and tumor gastric tissues using cDNA microarrays containing 17,000 human genes were done. This report highlights two separate but related aspects of the gene copy number change events in chromosome 20 of gastric

Received 5/7/04; revised 10/19/04; accepted 10/28/04.
Grant support: Korea Science and Engineering Fund through the Cancer Metastasis Research Center at Yonsei University College of Medicine.
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
Requests for reprints: Sun Young Rha, Cancer Metastasis Research Center, Yonsei University College of Medicine, 134 Shinchon-Dong, Seodaemun-Ku, Seoul 120-752, Korea. Phone: 82-2-361-7650; Fax: 82-2-362-5592; E-mail: rha7655@yumc.yonsei.ac.kr.

cancer. The first is a list of amplified genes in human chromosome 20 of gastric cancer compared with normal gastric tissues. We report that 96 genes including ZNF217, CYP24A, and PTPN1 show more than a 1.5-fold increase in the Cy5/Cy3 intensity ratio in the tumor tissues tested.

In many cases of microarray experiments, a correlation between the expression level of selected genes and the clinical parameters was obtained when the gene expression signature comprised of ratios or intensities of several genes. A group of selected genes, when combined, forms a molecular signature that provides a high correlation with the clinical information, even though they do not seem to provide meaningful information individually (16). Therefore, in the second approach, instead of selecting the genes that pass the 1.5-fold cutoff line, all of the genes that passed the microarray spot quality filtering and are located on chromosome 20 were examined using a *t* test to determine if any of the genes, whose copy changes are related with a recurrence, are selected. For this, the patients were divided into the recurrent and nonrecurrent group, and the genes that can discriminate between the two groups were searched. Using the *t* test, 5 genes ($P < 0.01$) and 39 genes ($P < 0.05$), whose combined score of copy change ratios could distinguish between the recurrent and nonrecurrent groups in a statistically significant manner, were selected.

Table 1 Clinical information of the patients

Category	Class	Cases	Total	Cox regression *P**
Age	< 65 y	15	30	0.823
	≥ 65 y	15		
Sex	Female	3	30	0.831
	Male	27		
Stage	I	12	30	0.005
	II			
	III	18		
	IV			
Differentiation	Poor	5	30	0.344
	Well	25		
Lymph node metastasis	N_{1-2}	20	27	0.003
	N_3	7		
Recurrence	Recurrent	13	27	Dependent on variables
	Nonrecurrent	14		
5-Gene grouping†	Low	13	27	0.000
	High	14		
39-Gene grouping‡	Low	8	27	0.017
	Middle	7		
	High	12		

*P values from the Cox regression analysis denote the statistical significance of survival between clinical classes in each category. Classes in each category are compared with its relation to survival. Age, sex, and differentiation do not hold statistical significance at $P < 0.05$ in survival.

†The group divided by five genes that have different copy number changes between recurrent group and nonrecurrent group.

‡The group divided by 39 genes that have different copy number changes between recurrent group and nonrecurrent group.

MATERIALS AND METHODS

Tissue Samples and DNA Isolation. All the experiments using the patient tissues were done after gaining the approval of the supervisory committee of the Yonsei University College of Medicine. Only tissue from the patients who underwent surgery at the Yonsei University College of Medicine (1997-1999), and were followed up for at least 5 years after surgery, were used. Tissues with an at least 70% tumor content were saved as the tumor tissues. Thirty pairs of normal and tumor gastric tissues used in these experiments (Table 1) were stored in liquid nitrogen immediately after surgery. The frozen tissues were ground to a powder, and the genomic DNA was prepared using the phenol/chloroform/isoamylalcohol method (17).

cDNA Microarray-Based CGH. The cDNA microarrays containing 17,000 human gene probes (Genomictree, Daejeon, Republic of Korea) were used for the microarray CGH. Six micrograms of the *Dpn*II-digested genomic DNA were purified using a QIAquick PCR purification kit (QIAgen, Düsseldorf, Germany), and was subjected to fluorescent labeling using a Bioprime labeling kit (Invitrogen, Carlsbad, CA). The reaction mixtures were incubated at 37°C for 2 hours in the dark and were quenched by adding 5 μL of 0.5 mol/L EDTA (pH 8.0). The microarray slide was incubated in 3.5× SSC, 0.1% SDS with 10 mg/mL bovine serum albumin for 1 hour at 42°C for blocking. The genomic DNAs from the normal and tumor tissues were labeled with Cy3-dCTP and Cy5-dCTP, respectively. A pair of Cy3- and Cy5-labeled DNAs from a patient was mixed with 30 μg human Cot-1 DNA, 20 μg poly (dA)-poly (dT) oligonucleotides, and 100 μg yeast tRNA. A Microcon-30 filter (Amicon, Bedford, MA) was used to purify and concentrate the hybridization mixture, which was then adjusted to the final concentration of 3.5× SSC and 0.3% SDS in 60 μL. Following denaturation at 100°C for 1.5 minutes and preannealing at 37°C for 30 minutes, the labeled DNA mixture was applied to the cDNA microarray and incubated at 65°C for 20 to 30 hours. The slide was then washed for 2 minutes each in 0.5× SSC/0.01% SDS, 0.06× SSC/0.01% SDS, and 0.05× SSC, consecutively, at room temperature and spun-dried. The microarray slides were scanned using a GenePix 4000B scanner (Axon Instruments, Foster City, CA).

Data Preprocessing. The fluorescence signal was calculated by subtracting the background intensity from the total intensity of a spot in GenePix Pro 4.1 software. Spots with poor signals (F532 ¡−1.5 × B532 < 0 or F635 ¡−1.5 × B635 < 0) were removed from further analysis (spot quality filtering). The discrepancies in the labeling efficiencies were adjusted by a "within print-tip Lowess normalization" (18). Information regarding the chromosomal locations of the genes were obtained from SOURCE http://genome-www5.stanford.edu/cgi-bin/source/sourceSearch).

Statistical Methods I: Selection of Amplified Genes. This paper classified the "copy number alteration" into two categories. The first was "gene amplification", meaning the conversion of an intact copy of a gene into two copies. For this, the "noise level" of the fluorescence intensity ratio within the normal tissues was first calculated. When the normal gastric tissues from the male (XY) and female (XX) genomic DNA were cohybridized in the control microarray CGH experiments, the genes on the autosomes were varied in $\log_2$ (Cy5/Cy3) ratios within ± 0.2 ranges. After adding 1 SD to this, it was decided that the genes on the normal gastric tissues were varied between the ± 0.3 range of the $\log_2$ (Cy5/Cy3) ratios. The average male to female ratio of the genes on the X chromosomes, XX/X, which could represent a single gene amplification into two genes, was calculated and a $\log_2$ value of 0.58 was obtained.

Table 2 Amplified genes in chromosome 20 of gastric cancer

No.	Genbank accession No.	Symbol	Cytoband	Average ratio	Frequency (%)	Reference
1	AI015986	RASSF2	20pter-p12.1	0.94	6.7	
2	AA450265	PCNA	20pter-p12	0.61	10	
3	AA455969	PRNP	20pter-p12	0.9	13.3	
4	R55796	PDYN	20pter-p12	0.66	10	
5	AI279333	C20orf155	20p13-p12.3	0.74	10	
6	H10761	RNF24	20p13-p12.1	0.74	10	
7	AA417279	PTPNS1	20p13	0.79	3.3	
8	AA448659	CDC25B	20p13	0.76	3.3	
9	AA461071	EST	20p13	0.74	10	
10	AA866160	SOX12	20p13	0.61	3.3	
11	AA974817	ATRN	20p13	0.95	3.3	
12	AA974946	SDCBP2	20p13	0.63	3.3	
13	AA974960	NSFL1C	20p13	0.77	3.3	
14	AI368570	CDS2	20p13	0.65	3.3	
15	AI399990	PTPRA	20p13	0.65	6.7	
16	AI538521	STK35	20p13	0.66	10	
17	AI928276	IDH3B	20p13	0.79	6.7	
18	AI017696	MKKS	20p12	1.21	3.3	
19	AI620336	PLCB1	20p12	0.98	16.7	
20	AI913732	HAO1	20p12	0.6	3.3	
21	AA035728	XRN2	20p11.2-p11.1	0.96	3.3	
22	AA922705	PYGB	20p11.2-p11.1	0.85	6.7	
23	AA424824	DSTN	20p11.23	0.63	6.7	
24	AA634360	SEC23B	20p11.23	1.19	3.3	
25	AW007634	HARS2	20p11.23	0.59	3.3	
26	AA599177	CST3	20p11.21	1.3	3.3	
27	AI018501	CRNKL1	20p11.2	0.6	6.7	
28	AI348549	C20orf147	20p11.1	1.03	30	
29	AA456022	FLRT3	20p11	0.85	10	
30	AA910881	EST	cen	0.67	3.3	
31	AI948413	CBFA2T2	20q11	0.86	3.3	
32	AA864524	NCOA6	20q11	0.67	3.3	(22)
33	AA934558	DEFB123	20q11.1	0.9	23.3	
34	AA922700	MAPRE1	20q11.1-11.23	1	66.7	
35	N39452	EST	20q11.2	0.65	3.3	
36	AA699926	SNTA1	20q11.2	1.08	46.7	
37	R36886	C20orf104	20q11.22-q11.23	0.7	6.7	
38	AA448286	ITCH	20q11.22-q11.23	0.73	16.7	
39	AI016688	C20orf52	20q11.23	0.81	40	
40	AA996156	SLA2	20q11.23	0.71	3.3	
41	AA485922	CPNE1	20q11.23	0.88	10	
42	T50121	MAFB	20q11.2-q13.1	0.86	13.3	
43	T95014	STK4	20q11.2-q13.2	0.7	13.3	
44	AA149097	HCK	20q11-q12	0.81	3.3	
45	AI302669	NCOA3	20q12	0.8	20	
46	AI279830	PPP1R16B	20q12	0.81	6.7	
47	AA683520	SLPI	20q12	0.82	6.7	
48	AA283091	KIAA1219	20q12	0.68	3.3	
49	AW082097	PI3	20q12-q13	0.71	6.7	
50	AI380629	SRC	20q12-q13	0.71	20	
51	H18950	SLC13A3	20q12-q13.1	0.89	3.3	
52	AI927795	SEMG2	20q12-q13.1	1.8	3.3	
53	AI301528	HNF4A	20q12-q13.1	0.95	13.3	
54	AA664210	PKIG	20q12-q13.1	0.71	10	
55	AA402874	PLTP	20q12-q13.1	0.64	3.3	
56	AA683578	ADA	20q12-q13.11	0.83	3.3	
57	H82992	PIGT	20q12-q13.12	0.83	3.3	
58	AI678253	SEMG1	20q12-q13.2	0.84	13.3	
59	W73473	BMP7	20q13	0.61	3.3	
60	AI674491	KCNG1	20q13	0.87	36.7	
61	AA877213	CYP24A1	20q13	0.72	6.7	(15, 26, 29)
62	H26184	CEBPB	20q13.1	0.83	6.7	(24)
63	AI302973	C20orf111	20q13.11	0.74	6.7	
64	AI304790	PTGIS	20q13.11-q13.13	0.64	6.7	
65	AA436260	ATP9A	20q13.11-q13.2	0.63	3.3	
66	R61289	L3MBTL	20q13.12	0.6	3.3	
67	AA918380	ZSWIM1	20q13.12	0.61	3.3	

(Continued on next page)

Table 2 Amplified genes in chromosome 20 of gastric cancer (cont'd)

No.	Genbank accession No.	Symbol	Cytoband	Average ratio	Frequency (%)	Reference
68	AA864812	SLC35C2	20q13.12	0.62	6.7	
69	AI961669	ARFGEF2	20q13.13	0.7	20	
70	AA897418	MOCS3	20q13.13	0.7	3.3	
71	AA886333	ADNP	20q13.13	0.9	3.3	
72	AA504825	ZNF313	20q13.13	1.06	6.7	
73	AA480906	PRKCBP1	20q13.13	0.74	6.7	
74	AA004759	DPM1	20q13.13	1.2	3.3	
75	R06605	PTPN1	20q13.1-q13.2	0.99	6.7	(22, 23)
76	AA682649	SPATA2	20q13.1-q13.2	0.61	10	
77	AI559473	ZNF217	20q13.2	0.86	10	(15, 23, 24, 26, 27)
78	AI302759	C20orf17	20q13.2	0.9	16.7	
79	AA069770	KCNB1	20q13.2	0.79	3.3	
80	T67004	EDN3	20q13.2-q13.3	0.68	3.3	
81	AI393442	SPO11	20q13.2-q13.3	0.79	6.7	(24)
82	AI365196	NFATC2	20q13.2-q13.3	0.8	6.7	(24)
83	AI268751	GNAS	20q13.2-q13.3	1.15	56.7	
84	AI700308	PPP1R3D	20q13.3	0.68	3.3	
85	AI681381	RPS21	20q13.3	0.94	16.7	
86	AA887585	PTK6	20q13.3	0.66	6.7	
87	AA700688	ATP5E	20q13.3	0.72	10	
88	AI339492	EST	20q13.31	0.7	10	
89	AA504128	RAE1	20q13.31	0.75	6.7	
90	AA676370	STX16	20q13.32	0.65	10	
91	AI986098	MYT1	20q13.33	1.28	6.7	
92	AA946732	GTPBP5	20q13.33	0.93	23.3	
93	AA910559	C20orf40	20q13.33	1.74	3.3	
94	AA863149	PSMA7	20q13.33	0.81	20	
95	AA412500	TCEA2	20q13.33	1.24	6.7	
96	AA872091	TCFL5	20q13.3-qter	0.63	3.3	

Therefore, the genes with $\log_2$ (Cy5/Cy3) $\geq$ 0.58 were determined to be those showing gene amplification. The term "frequency" in the text refers to the percentage of copy number alteration incidences for each gene from the 30 experiments. Therefore, after combining the intensity ratios and frequency, the ratio of $\log_2$ (Cy5/Cy3) $\geq$ 0.58 or $\leq$ ¡−0.58 was observed in at least 1 of the 30 cases tested (3.3% frequency) for the gene to be qualified as amplified.

Statistical Methods II: Selection of Genes That Distinguish Recurrent and Nonrecurrent Groups. The second aspect of the gene copy number change events in this paper is the effect of combined gene copy change ratios. Without consideration of $\log_2$ (Cy5/Cy3) having to be > 0.58 or < ¡−0.58, all the genes that passed quality filtering and that were located at chromosome 20 were considered. Patients were first divided into a recurrent and nonrecurrent groups. To evaluate the gene copy number change and their possible relation to the distinction of recurrence status of patients, genes that showed differential copy number changes between the recurrent and nonrecurrent groups were selected by a *t* test. The tissue samples without information for recurrence status were removed from the analysis. From the *t* test, 5 genes ($P < 0.01$), and 39 genes ($P < 0.05$) containing these 5 genes, were selected. A modified version of prognostic scoring system used by Inoue et al. (19) was devised, and each patient was given a score S_i:

$$S_i = 100\Sigma_k G_{ik}$$

where S_i is the score of patient i, and G_{ik} is the $\log_2$ (Cy5/Cy3) of gene k of patient i.

Logistic regression analysis was done to determine if the score of each patient was correlated with a recurrence. The patients were classified into two or three groups based on the fact that S_i was likely to reflect the genetic alteration level. According to the scores of the 5 or 39 genes, 27 patients were divided into two groups (Table 1; 5-gene grouping of patients) or three groups (39-gene grouping of patients). In the 5-gene grouping of the patients, the patients were divided into a high group with $S_i \geq 30$ (14 cases) and a low group with $S_i \leq 30$ (13 cases). In the 39-gene grouping, the patients were divided into the high group with $S_i \geq 292$ (12 cases), a middle group with $0 \leq S_i \leq 292$ (7cases), or a low group with $S_i \leq 0$ (8 cases).

The disease-free survival curves were constructed according to the Kaplan-Meier method and compared with the G-rho family of tests including the log-rank and Gehan-Wilcoxon test. Cox regression analysis was used for multivariate survival analysis with a backward stepwise selection procedure, and the following were entered as the categorical covariates: the tumor stage (I, II versus III, IV), the age at diagnosis (< 65 years, $\geq$ 65 years), gender, lymph node metastasis ($N_{1\text{-}2}$ versus N_3), and genetic group based on the selected genes (5-gene grouping, 39-gene grouping; Table 1). The association between the genetic groups and the stage groups were analyzed using Fisher's exact test.

RESULTS

Global View of Gene Copy Number Changes on Chromosome 20 in Gastric Cancers. From the cDNA microarray-based CGH of the 30 pairs of gastric samples, ~13,000 genes passed the spot quality filtering. Of these, 7,064 genes had

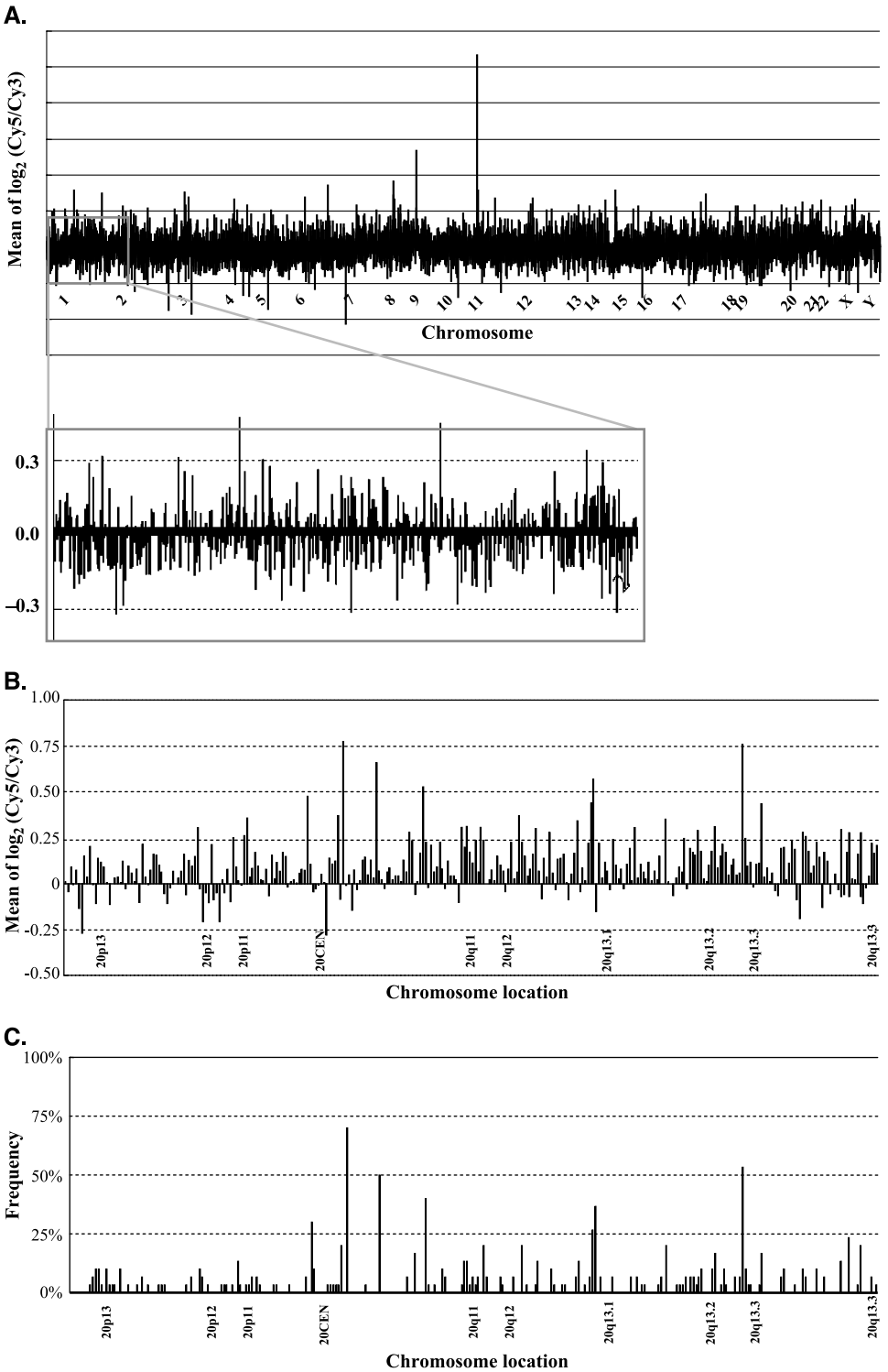

Fig. 1 Global view of gene copy number changes on chromosome 20 in 30 cases of gastric cancer analyzed by cDNA microarray-based CGH. *A,* 7,064 genes with known chromosomal location information were plotted for their mean of $\log_2$ (Cy5/Cy3) and by chromosome location. Amplified view of gene copy change pattern for chromosome 1 is shown in the box with $\log_2$ (Cy5/Cy3) range (*Y*-axis). Of the 7,064 genes located on chromosome 20, 296 are plotted by their relative location (*X*-axis; *A*), by the mean of $\log_2$ (Cy5/Cy3; *B*), or by frequency (*C*).

chromosomal location information according to SOURCE (Data Preprocessing in MATERIALS AND METHODS), and the mean of $\log_2$ (Cy5/Cy3) for 7,064 genes in 30 samples were plotted according to their chromosomal locations (Fig. 1*A*): 296 of the 7,064 genes were located on chromosome 20. Of these, 84 (~30%) were located on the p arm, 8 had map locations spanning the centromere, and the remaining 204 were at the q arm of chromosome 20. In order to search for the amplified genes in chromosome 20, 296 genes were screened for the $\log_2$ (Cy5/Cy3) ratios of either ≥ 0.58 in at least 1 (3.3%) out of the 30 samples tested. One hundred and twenty-three genes were selected and their mean $\log_2$ (Cy5/Cy3) values (Fig. 1*B*) and

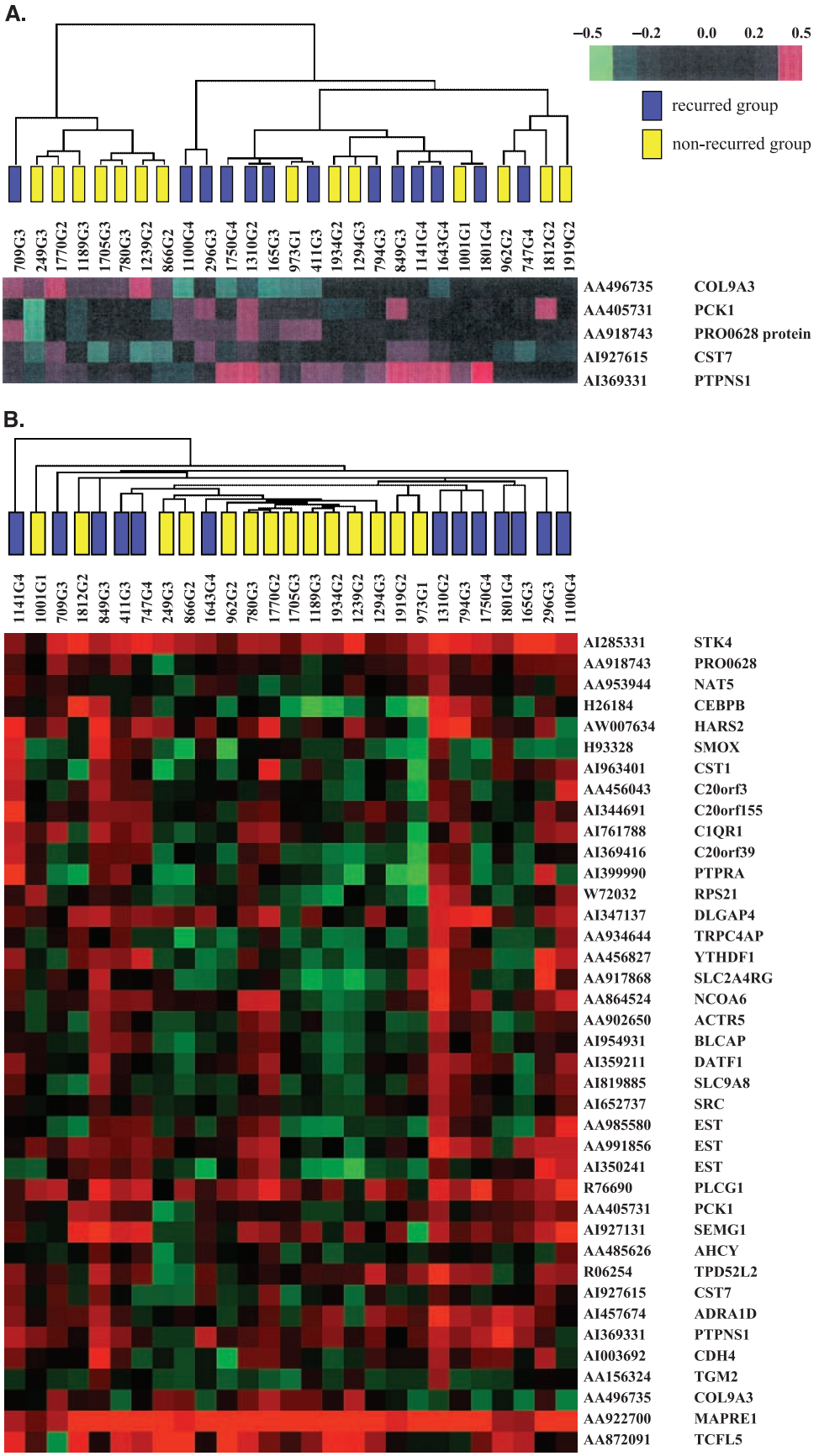

Fig. 2 A two-way hierarchical clustering of genes whose copy number changes are correlated with recurrence. Five genes selected at $P < 0.01$ (*A*) or 39 genes selected at $P < 0.05$ (*B*) were clustered and shown as TreeView dendrograms. Dendrogram of relative gene copy increase (*red*) or decrease (*green*) following the scale bar in log scale (*bottom right*). *Yellow bars* and *blue bars* under the dendrograms, nonrecurrent and recurrent groups, respectively. Patient labels are shown below the yellow and blue bars.

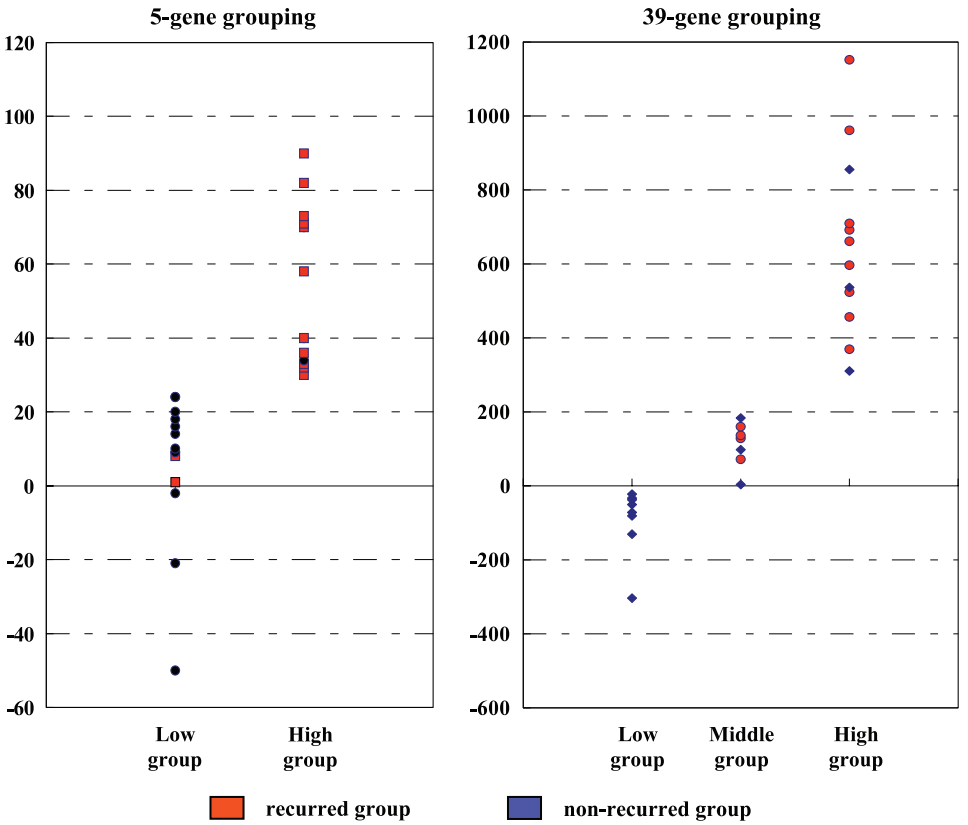

Fig. 3 Grouping of recurrent and nonrecurrent patients based on the combination of gene copy level. According to the S_i for each patient (see the text), 27 cases are divided into two groups of low and high in a "5-gene grouping" system or three groups of low, middle, and high group in a "39-gene grouping" system as summarized in Table 1. The low groups in a 5-gene grouping and a 39-gene grouping system show scores in the range of −50 to 24 and 304 to −22, respectively. The scores of the high group were from 30 to 90 in a 5-gene grouping system, and from 311 to 1,152 in the 39-gene grouping system. These scores significantly correlate with the disease-free survival of each patient (Fig. 4). In a 5-gene grouping system, the median survival was 30 months for 14 patients who belonged to the high group, and the median survival was also 30 months for 12 patients who belonged to the high group in a 39-gene grouping system.

incidences (Fig. 1*C*) were plotted according to the map locations. Approximately 30% of the copy number-altered genes (39 of 123 genes) were located on the p arm of chromosome 20, whereas the other 70% (84 of 123) had map locations at the q arm.

Amplified Genes at Chromosome 20 in Gastric Cancer. This study surveyed the gene amplification only (not the deletions) at human chromosome 20 [$\log_2$ (Cy5/Cy3) $\geq$ 0.58, 3.3%]. Of the 123 unique genes, 96 passed these criteria (Table 2); and of the 96 amplified genes, 29 genes (30%) were located at the p arm, 1 at the centromere region, and 66 (70%) at the q arm of chromosome 20. The higher frequency of the gene amplification events at the q arm of chromosome 20 compared with the p arm, does not necessarily guarantee the true biologically active gene copy change distribution at the q arm until proven otherwise, but represents the relative proportions of the copy changes among the genes spotted on the microarray at this point. ZNF217 and CYP24A, which are well-known oncogene candidates that have previously been shown to be within an amplicon boundary in 20q13.2 of gastric cancer (15), were among the genes selected.

Association Between the Gene Copy Number Changes at Chromosome 20 and the Clinical Outcome. Because no correlations could be found between the gene amplification and the clinical parameters when the criteria of $\geq$ 1.5-fold of Cy5/Cy3 and $\geq$ 3.3% frequency was used, we examined whether or not the molecular signature of the combined Cy5/Cy3 of multiple genes, albeit with low ratios, could provide any useful information. In order to select the discriminatory genes, we compared the genetic alteration pattern in the recurrent and the nonrecurrent group comprising 27 out of the 30 patients whose recurrence status was known (Table 1). Initially, *t* tests were used to select a subset of discriminatory genes from all the genes on chromosome 20. *t* tests with nominal *P* values of 0.01 and 0.05 produced 5 and 39 genes, respectively, of which the 5 genes, PTPNS1, CST7, PRO0628, PCK1, and COL9A3, with $P < 0.01$ were also in the 39-gene list with $P < 0.05$. Two-way hierarchical clustering showed that these 5 and 39 genes, which were visualized via TreeView dendrogram, were able to classify most of the 27 cases into either the recurrent group or nonrecurrent group based on their copy number change patterns (Fig. 2). The next stage was to identify any correlation between the selected groups of 5 or 39 genes and the recurrence status of the patients. In order to classify the patients according to the genetic alteration levels of the selected genes, each patient was given a score using a logistic regression analysis (MATERIALS AND METHODS). The 27 patients were divided into two or three groups based on the patients' scores reflecting the genetic alteration level (Table 1; Fig. 3), and logistic regression analysis was done to determine if the score-based grouping of the patients reflects the recurrence versus nonrecurrence status. The results of logistic regression analysis confirmed that the scores of each patient for both of the 5 and 39 genes were strongly correlated with the recurrence of cancer ($P < 0.05$ in both cases). Using Fisher's exact test, it was confirmed that the genetic subgroups and the gastric cancer stage were not significantly associated

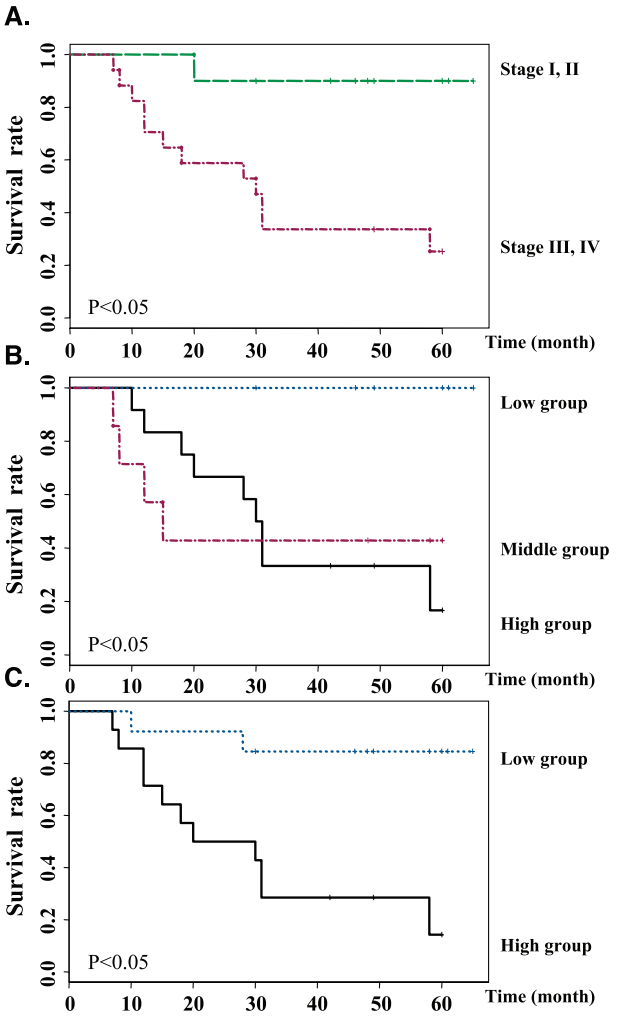

Fig. 4 Survival rate of 27 gastric cancer patients depending on the status of gene copy changes in chromosome 20. *A,* Kaplan-Meier survival curves of gastric cancer patients divided by tumor stages (stages I and II versus stages III and IV). Patients with an early stage have a higher disease-free survival rate than those patients with a late stage disease. *B,* survival curves of high, middle, and low groups divided by the genetic alteration levels of 39 genes that had different copy number changes between the recurrent group and nonrecurrent group. *C,* survival curves of high and low groups divided by the genetic alteration levels of 5 genes that had different copy number changes between the recurrent group and nonrecurrent group.

($P = 0.12$). Although the use of a traditional staging system for determining the prognosis of gastric cancer represents a statistically significant difference for a nonrecurrence, the Cox regression analysis result and the Kaplan-Meier curves showed that the patients grouped according to the alteration of the specific genes were also highly associated with nonrecurrence (Fig. 4). The group with a higher level of copy number changes in the recurrence-related genes on chromosome 20 showed a shorter survival rate (Fig. 4*B* and *C*). Although it was shown that patients with a late stage disease have a shorter survival rate than those patients with an early stage disease ($P < 0.05$; Fig. 4*A*), the recurrence difference based on the gene copy number changes were also statistically meaningful ($P < 0.05$).

DISCUSSION

Human chromosome 20 is of special interest due to the presence of a susceptibility locus for hereditary diseases, which include various types of cancers (20). Genomic instabilities at chromosome 20, particularly at the 20q13 region, are frequently examined for the presence of specific oncogenes and gene amplification events in various types of tumors and/or cancer cell lines. These genes include CYP24, ZNF217 (15), and BTAK (21) in gastric cancer; CTSZ, NCOA6, and PTPN1 in pancreatic cancer (22); E2F1, TGIF2, PTPN1 BTAK, and ZNF217 in ovarian cancer (23); and AIB4, ZNF217, CYP24, and aurora 2 kinase in breast (24, 25) and colorectal cancer (26). Almost all these gene amplification events have been examined using the BAC clone-based approaches. In many cases, these BAC clones are several hundred kilobases in length, and approaches of single gene probing, one at a time on a small scale, need to be taken in order to identify the copy-changed genes specifically. These laborious and time-consuming approaches often result in the reconfirmation of a previously known specific gene amplification event. On the other hand, cDNA microarray-based CGH can look at the gene copy change events for all the genes printed on the array in an unbiased manner. The chromosome 20q13.2 region contains differentially amplified genes in primary and metastatic colorectal cancer, and the amplification has been associated with metastasis (27). It has been suggested that there are amplified genes in the 20q13.2 region other than the well-known genes including ZNF217. Many reports repeatedly emphasized that chromosome 20 needs to be examined in more detail because not only are CYP24 and ZNF217 located in this specific region of 20q but other BAC array-based CGH experiments have continuously highlighted this region as being amplified in various types of tumors. The current study of cDNA microarray-based CGH analysis provides a list of candidate amplified genes in chromosome 20 (Table 2). Log_2 (Cy5/Cy3) $\geq$ 0.58 was used to specifically select genes that are amplified. This is a relatively high level of intensity ratio change for a gene to be qualified as copy number–changed. We specifically used this line to select amplified genes that can be confirmed by independent investigators beyond experimental differences. Several experiments are in the planning stage in order to confirm the gene copy number changes shown in this report. First, some of the well-known amplified genes in gastric cancer including BTAK are located in 20q. Unfortunately, the cDNA microarrays used in this study did not contain such probes. Independent confirmation of the result by fluorescence *in situ* hybridization or Southern blotting for some of the amplified genes in this report together with the genes that have been reported to be amplified but were not spotted in the cDNA microarrays, notably, BTAK, aurora kinase, and AIB4, will be highly meaningful. Second, we were looking for the gene amplification that could contribute to the linear amplification of mRNA yield by selecting the genes that are more likely to be amplified in this study. We have done RNA expression profiling experiments using the same tissue samples and cDNA microarrays. The genes that are highly correlated in the copy change and RNA expression are strong candidates for novel oncogenes. The initial analysis suggests that a comparison of the gene copy change and RNA expression level might require more than just a fold change–based gene selection in the two databases and

direct comparison. Instead, a comparison of Pearson correlation coefficients between gene copy number change profile and RNA expression profile, following the idea of correlation analysis between gene expression profiles and chemosensitivity profiles (28), would provide meaningful information.

This study showed that the minimum five genes of PTPNS1, CST7, PRO0628, PCK1, and COL9A3 comprise a gene copy change signature and that the copy numbers of these genes was higher in the patient group with a recurrent gastric cancer. Furthermore, the patient groups with copy increase for these genes are strongly correlated with a poor survival (Fig. 4). This is probably the first study where the gene copy number changes in chromosome 20 show a direct relationship with the patient recurrence status. Ongoing data analysis for the remaining 17,000 genes in the 30 cases of the microarray CGH will yield an almost complete view of the gene copy number change events that occur in gastric cancer.

ACKNOWLEDGMENTS

We thank G.Y. Lee, C.H. Park, and Y.J. Yang for technical support and helpful comments.

REFERENCES

1. Parkin DM, Pisani P, Ferlay J. Estimates of the worldwide incidence of eighteen major cancer in 1985. Implications for prevention and projections of future burden. Int J Cancer 1993;54:594–606.

2. Ji J, Chen X, Leung SY, et al. Comprehensive analysis of the gene expression profiles in human gastric cancer cell lines. Oncogene 2002; 21:6549–56.

3. Wang J, Chen S. Screening and identification of gastric adenocarcinoma metastasis-related genes using cDNA microarray coupled to FDD-PCR. J Cancer Res Clin Oncol 2002;128:547–53.

4. Wang J, Chen S. Screening and identification of gastric adenocarcinoma metastasis-related genes using cDNA mi4. Sakakura C, Hagiwara A, Nakanishi M, et al. Differential gene expression profiles of gastric cancer cells established from primary tumour and malignant ascites. Br J Cancer 2002;87:1153–61.

5. Meireles SI, Carvalho AF, Hirata R, et al. Differentially expressed genes in gastric tumors identified by cDNA array. Cancer Lett 2003;190:199–211.

6. Hippo Y, Taniguchi Y, Tsutsumi S, et al. Global gene expression analysis of gastric cancer by oligonucleotide microarrays. Cancer Res 2002;62:233–40.

7. Boussioutas A, Li H, Liu J, et al. Distinctive patterns of gene expression in premalignant gastric mucosa and gastric cancer. Cancer Res 2003;63:2569–77.

8. Tay ST, Leong SH, Yu K, et al. A combined comparative genomic hybridization and expression microarray analysis of gastric cancer reveals novel molecular subtypes. Cancer Res 2003;63:3309–16.

9. Chen X, Leung SY, Yuen ST, et al. Variation in gene expression patterns in human gastric cancers. Mol Biol Cell 2003;14:3208–15.

10. Weiss NM, Kuipers EJ, Postma C, et al. Genomic profiling of gastric cancer predicts lymph node status and survival. Oncogene 2003;22: 1872–9.

11. Kallioniemi A, Kallioniemi OP, Sudar D, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 1992;258:818–21.

12. Solinas-Toldo S, Lampel S, Stilgenbauer S, et al. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer 1997;20:399–407.

13. Pollack JR, Perou CM, Alizadeh AA, et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet 1999; 23:41–6.

14. Pollack J, Sørlie T, Perou C, et al. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci U S A 2002;99: 12963–8.

15. Weiss MM, Snijders AM, Kuipers EJ, et al. Determination of amplicon boundaries at 20q13.2 in tissue samples of human gastric adenocarcinomas by high-resolution microarray comparative genomic hybridization. J Pathol 2003;200:320–6.

16. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 2002;346:1937–47.

17. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning, a laboratory manual. New York: Cold Spring Harbor Press; 1989.

18. Yang YH, Dudoit S, Luu P, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 2002;30:E15.

19. Inoue H, Matsuyama A, Mimori K, et al. Prognostic score of gastric cancer determined by cDNA microarray. Clin Cancer Res 2002;8: 3475–9.

20. Deloukas P, Matthews LH, Ashurst J, et al. The DNA sequence and comparative analysis of human chromosome 20. Nature 2001;414: 865–71.

21. Sakakura C, Hagiwara A, Yasuoka R, et al. Tumour-amplified kinase BTAK is amplified and overexpressed in gastric cancers with possible involvement in aneuploid formation. Br J Cancer 2001;84:824–31.

22. Mahlamäki EH, Bärlund M, Tanner M, et al. Frequent amplification of 8q24, 11q, 17q, and 20q-specific genes in pancreatic cancer. Genes Chromosomes Cancer 2002;35:353–8.

23. Watanabe T, Imoto I, Katahira T, et al. Differentially regulated genes as putative targets of amplifications at 20q in ovarian cancers. Jpn J Cancer Res 2002;93:1114–22.

24. Hodgson JG, Chin K, Collins C, Gray JW. Genome amplification of chromosome 20 in breast cancer. Breast Cancer Res Treat 2003;78: 337–45.

25. Nonet GH, Stampfer MR, Chin K, Gray JW, Collins CC, Yaswen P. The ZNF217 gene amplified in breast cancers promotes immortalization of human mammary epithelial cells. Cancer Res 2001;61:1250–4.

26. Nakao K, Mehta KR, Fridlyand J, et al. High-resolution analysis of copy number alterations in colorectal cancer by array-based comparative genomic hybridization. Carcinogenesis 2004;25:1345–57.

27. Hidaka S, Yasutake T, Takeshita, H, et al. Differences in 20q13.2 copy number between colorectal cancers with and without liver metastasis. Clin Cancer Res 2000;6:2712–7.

28. Scherf U, Ross DT, Waltham M, et al. A gene expression database for the molecular pharmacology of cancer. Nat Genet 2000; 24:236–44.

29. Veltman JA, Fridlyand J, Pejavar S, et al. Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors. Cancer Res 2003;63:2872–80.