linitis plastica.[6] Identification of genetic predisposing factors and molecular pathways for the DGC development is one of the fundamentals for conceiving effective prevention, early diagnosis and therapeutic strategies.

Previously, we conducted a gene-centric genome-wide association study (GWAS) on DGC and identified 3 statistically significant single nucleotide polymorphisms (SNPs) at 2 loci after Bonferroni correction ($P < 1.8 \times 10^{-5}$) in the second stage of the screening.[7] Of the 3 SNPs, rs2976392 showed the lowest P value and tagged a linkage disequilibrium (LD) block at chromosome 8q24.3, in which we identified *prostate stem cell antigen* (*PSCA*) encoding prostate stem cell antigen as the novel DGC susceptibility gene. In the present study, we investigated the second genomic region of interest at chromosome 1q22, which harbors the remaining 2 SNPs, rs2075570 and rs2070803,[7] and identified *mucin 1* (*MUC1*) as the possible causal gene of the association of the region to DGC. The association between the gene and GC had been suggested also in previous reports.[8-11] However, unlike the previous candidate gene approach, we have reached the gene by performing a hypothesis-free GWAS followed by biologic studies in which a rationale of the association was obtained through the analyses of the function of a SNP rs4072037. Moreover, this study has a sufficient power as a systematic survey of genetic factors with the expected range of effect size and allele frequencies, generating a convincing level of statistical association ($P < 10^{-10}$ as compared with $P \sim 10^{-2}$ by the previous candidate gene approach[11]). The SNP rs4072037 is known to determine a splicing acceptor site in the second exon of *MUC1*.[12] In this study, we showed that the SNP is also related to major splicing variant selection in the stomach and has effect on the *MUC1* promoter activity, both of which may result in MUC1 functional difference between the individuals.

Materials and Methods

Samples

In Japan, the common type of GC is classified into 7 categories: papillary adenocarcinoma (pap), tubular adenocarcinoma (tub1 and tub2), poorly differentiated adenocarcinoma (por1 and por2), signet-ring cell carcinoma (sig), and mucinous adenocarcinoma (muc). However, a classification into 2 major categories by Lauren,[2] intestinal and diffuse types, is used worldwide especially for clinicoepidemiologic studies. A review of the classification systems is described elsewhere.[7] Basically, the DGC under the Lauren classification corresponds to por2 (nonsolid type) of poorly differentiated adenocarcinoma and sig by Japanese classification, although some investigators consider that por1 (solid type) is also included in DGC.[13]

Details of DNA samples used in the SNP typing and the association study of the chromosome 1q22 locus are as follows: In the Tokyo data set study, 610 DNA samples from patients with DGC (320 males; mean age, 55.4; 290 females; mean age, 54.0) were prepared either from methanol-fixed, paraffin-embedded tissues of noncancerous gastric mucosa or lymph nodes, or from peripheral blood, of patients with either the linitis plastica type of GC or early-stage cancer diagnosed as macroscopic type 0 IIc with histologic type of por2 and/or sig. The DGC samples in the Tokyo data set were collected at 4 institutions: the National Cancer Center Hospital in Tokyo: 360 paraffin-embedded tissues and 164 blood samples; Nippon Medical School Hospital in Tokyo: 76 blood samples; Aichi Cancer Center in Aichi: 1 blood sample; and Shikoku Cancer Center in Ehime: 9 blood samples. The control DNA samples were from peripheral blood leukocytes of 1266 volunteer individuals (male, 849; mean age, 67.2; female, 417; mean age, 59.8) with no known malignancies, who offered blood at a health check examination at Iwata City Hospital in Shizuoka and at Keio University campuses in Tokyo.

In the Aichi data set study, the DGC case samples were obtained from peripheral blood of 304 patients with histologic diagnosis of por1, por2, or sig (199 males; mean age, 57.3; 105 females; mean age, 56.4). Control blood samples were from 1467 volunteer individuals (1098 males; mean age, 59.8; 369 females; mean age, 57.3) with no known malignancies. Power calculations for the DGC analysis showed that the sample size of 304 cases and 1467 controls would provide the study with a power of over 98% for detecting an association of a SNP with a minor allele frequency of 0.2 or higher and per-allele odds ratio (OR) for risk allele of 1.63 or higher (estimates on rs2070803 obtained from Tokyo data set) in a 2-sided test at a significance level of .05.

In the Korea data set study, peripheral blood samples were donated from 455 patients with DGC who were diagnosed or treated at the National Cancer Center in Seoul, Korea (260 males; mean age, 52.4; 195 females; mean age, 48.5). The control subjects were 372 volunteers who participated in the National Cancer Screening Program at the National Cancer Center, Korea, and were confirmed by endoscopy not to have GC (191 males; mean age, 54.2; 181 females; mean age, 52.5). Power calculations showed that the sample size of the Korea data set would provide the study with a power of over 95% for detecting an association of rs2070803 at a significance level of .05 for the DGC study.

In the association studies (results shown in Figure 1, Table 1, and Supplementary Tables 1-4), 11 subjects (4 DGC and 2 controls from Tokyo data set, 2 controls from Aichi data set, 3 DGC cases from Korea data set) were excluded because of at least 1 missing covariate. Distributions of the covariates from subjects included in the studies are shown in Supplementary Figures 9-11.

Haplotype-based association study was performed on DNA samples from 380 DGC cases (200 males; mean age,

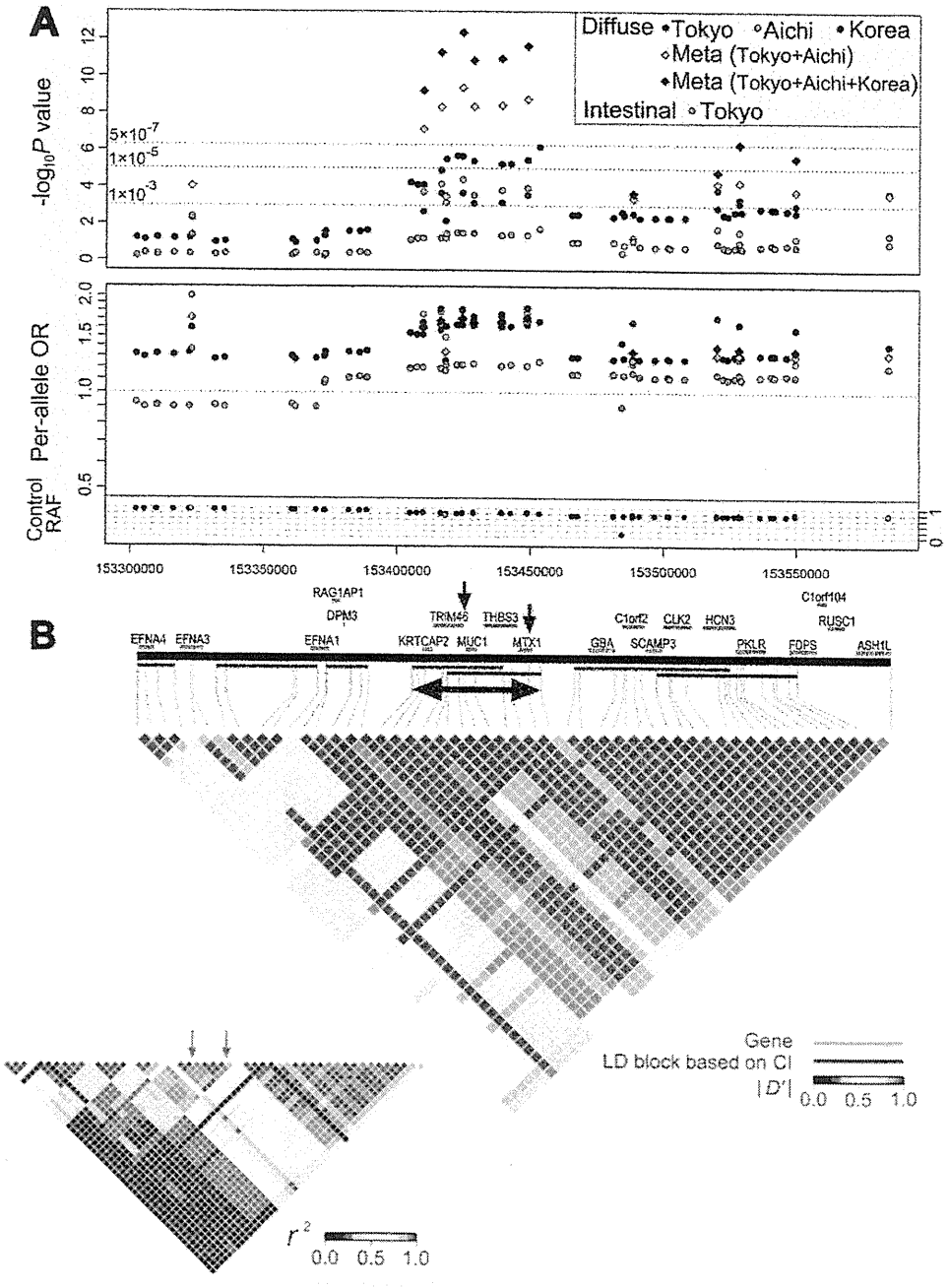


Figure 1. Association of the SNPs around rs2070803 and rs2075570 in chromosome 1q22 with GC and LD analyses of the SNPs. (*A*) The association study on DGC disclosed 8 SNPs with $P < 4 \times 10^{-5}$ within the LD block around rs2070803 and rs2075570 (*arrows* in *B*) in the Japanese population (Tokyo data set: *blue dots*). The association with DGC was replicated in another Japanese population (Aichi data set: *orange dots*) and also in the Korean population (Korea data set: *red dots*) for 4 selected SNPs: rs2070803, rs4072037, rs2066981, and rs2075570 ($P < 1 \times 10^{-3}$). Meta-analysis on these 3 data sets was also conducted (Tokyo and Aichi: *grey dots*; Tokyo, Aichi, and Korea: *black dots*). The study on IGC in the Japanese population (Tokyo data set: 599 cases, 1264 controls) showed no significant association of SNPs (*green dots*). *Upper panel* shows *P* value of each SNP in negative common logarithmic scale; *lower panel* shows OR and frequency of the risk allele (control RAF) of the SNPs. The position of the *dots* representing each SNP corresponds vertically to that in the physical map in *B*. (*B*) An LD analysis based on |*D'*| showed a strong LD around the 2 SNPs, rs2070803 and rs2075570 (*red arrows*), identified as DGC-associated SNPs in GWAS.[7] The strength of the LD is indicated by heat maps. *Blue bars* represent LD blocks defined by confidence intervals of |*D'*|.[15] Five genes—*KRTCAP2*, *TRIM46*, *MUC1*, *THBS3*, and *MTX1*—reside in the region of strong LD (*double-headed red arrow*) harboring rs2070803 and rs2075570. An LD map with r^2 is also shown in small scale. The analysis was performed with genotyping data of 1266 controls of the Tokyo data set.

Table 1. Results of Association Studies With 3 Independent Data Sets—Tokyo, Aichi, and Korea—for GC Susceptibility, on the 8 SNPs in the LD Block Containing rs2070803 and rs2075570 in Chromosome 1q22

SNPs, major/minor/risk[a]	Diffuse, Tokyo data set[b,c]	Diffuse, Aichi data set[c,d]	Diffuse, Korea data set[c,e]	Diffuse, Meta-analysis[f,g]	Diffuse, Meta-analysis[g,h]	Intestinal, Tokyo data set[c,i]
rs9426886 T/A/T	OR: 1.61 (1.32–1.97) *P*: 3.17×10^{-6} MAF (case): 0.133 MAF (control): 0.199					OR: 1.22 (1.01–1.46) *P*: 3.62×10^{-2} MAF (case): 0.169 MAF (control): 0.199
rs4971100 A/G/A	OR: 1.63 (1.33–1.99) *P*: 2.04×10^{-6} MAF (case): 0.133 MAF (control): 0.200					OR: 1.22 (1.02–1.47) *P*: 3.13×10^{-2} MAF (case): 0.169 MAF (control): 0.200
rs2070803[j] G/A/G	OR: 1.63 (1.33–1.99) *P*: 2.20×10^{-6} MAF (case): 0.133 MAF (control): 0.200	OR: 1.81 (1.36–2.40) *P*: 3.93×10^{-5} MAF (case): 0.104 MAF (control): 0.178	OR: 1.82 (1.32–2.49) *P*: 2.19×10^{-4} MAF (case): 0.103 MAF (control): 0.178	OR: 1.69 (1.43–1.99) *P*: 4.25×10^{-10}	OR: 1.71 (1.48–1.98) *P*: 4.33×10^{-13}	OR: 1.22 (1.02–1.47) *P*: 3.34×10^{-2} MAF (case): 0.168 MAF (control): 0.200
rs4072037 A/G/A	OR: 1.62 (1.32–1.99) *P*: 4.04×10^{-6} MAF (case): 0.126 MAF (control): 0.187	OR: 1.69 (1.27–2.25) *P*: 2.82×10^{-4} MAF (case): 0.099 MAF (control): 0.164	OR: 1.74 (1.26–2.39) *P*: 7.82×10^{-4} MAF (case): 0.093 MAF (control): 0.163	OR: 1.64 (1.39–1.94) *P*: 4.46×10^{-9}	OR: 1.66 (1.44–1.93) *P*: 1.43×10^{-11}	OR: 1.23 (1.02–1.48) *P*: 3.36×10^{-2} MAF (case): 0.157 MAF (control): 0.187
rs2066981 T/C/T	OR: 1.61 (1.31–1.98) *P*: 5.91×10^{-6} MAF (case): 0.125 MAF (control): 0.186	OR: 1.74 (1.31–2.32) *P*: 1.50×10^{-4} MAF (case): 0.099 MAF (control): 0.167	OR: 1.76 (1.27–2.43) *P*: 6.98×10^{-4} MAF (case): 0.092 MAF (control): 0.159	OR: 1.65 (1.40–1.95) *P*: 3.77×10^{-9}	OR: 1.67 (1.44–1.94) *P*: 1.11×10^{-11}	OR: 1.21 (1.01–1.46) *P*: 4.41×10^{-2} MAF (case): 0.158 MAF (control): 0.186
rs914615 G/A/G	OR: 1.61 (1.31–1.98) *P*: 5.83×10^{-6} MAF (case): 0.126 MAF (control): 0.187					OR: 1.22 (1.01–1.47) *P*: 3.94×10^{-2} MAF (case): 0.158 MAF (control): 0.187
rs2075570[j] A/G/A	OR: 1.63 (1.32–2.00) *P*: 3.45×10^{-6} MAF (case): 0.127 MAF (control): 0.189	OR: 1.77 (1.33–2.37) *P*: 1.12×10^{-4} MAF (case): 0.100 MAF (control): 0.170	OR: 1.84 (1.32–2.55) *P*: 2.73×10^{-4} MAF (case): 0.091 MAF (control): 0.163	OR: 1.67 (1.42–1.98) *P*: 1.73×10^{-9}	OR: 1.71 (1.47–1.98) *P*: 2.26×10^{-12}	OR: 1.21 (1.01–1.46) *P*: 4.25×10^{-2} MAF (case): 0.159 MAF (control): 0.189
rs1057941 T/C/T	OR: 1.67 (1.36–2.04) *P*: 6.59×10^{-7} MAF (case): 0.130 MAF (control): 0.200					OR: 1.25 (1.04–1.50) *P*: 1.95×10^{-2} MAF (case): 0.168 MAF (control): 0.200

NOTE. OR: per-allele odds ratio for risk allele with 95% confidence intervals in parentheses.
MAF, minor allele frequency; SNP, single nucleotide polymorphism.
[a]Major, minor, and risk alleles are common in Tokyo, Aichi, and Korea data sets.
[b]Association study on diffuse-type GC in the Japanese population (Tokyo data set, 606 cases, 1264 controls), performed with fine-mapping data.
[c]*P* values and ORs with 95% CI were calculated under an additive model using logistic regression adjusted for age, gender, and number of risk alleles of rs2294008 of the *PSCA* gene, which was associated with DGC in our previous GWAS (Study Group of Millennium Genome Project for Cancer[7]).
[d]Association study on diffuse-type GC in the Japanese population (Aichi data set: 304 cases, 1465 controls), for replication of the study on Tokyo data set.
[e]Association study on diffuse-type GC in the Korean population (Korea data set: 452 cases, 372 controls), for replication of the study on Tokyo data set.
[f]Meta-analysis on the data of the Tokyo and Aichi data sets.
[g]*P* values and ORs with 95% CI were calculated using a random effects model.
[h]Meta-analysis on the data of the Tokyo, Aichi, and Korea data sets.
[i]Association study on intestinal-type GC in the Japanese population (Tokyo data set: 599 cases, 1264 controls).
[j]SNPs identified in the GWAS (Study Group of Millennium Genome Project for Cancer[7]).

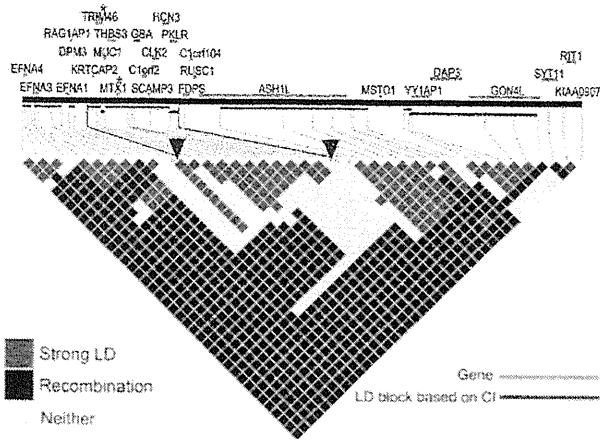


Figure 2. LD analyses on the SNPs in chromosome 1q22 using Gabriel's criteria[15] based on genotyping data of 680 Japanese controls. LD blocks with the criteria are in *blue horizontal lines*. *Red asterisks* indicate positions of rs2075570 and rs2070803. The LD block (both ends indicated by *red triangles*), in which the 2 SNPs reside, contains 12 SNPs and 11 genes. CI, confidence interval.

56.1; 178 females; mean age, 54.6; 2 missing gender information; 372 of which were the same as DGC cases in the Tokyo data set) and 309 controls (151 males; mean age, 49.1; 158 females; mean age, 46.1; 306 of which were included in the Tokyo data set). DGC samples were collected at 2 institutions as follows: 318 paraffin-embedded tissues at the National Cancer Center Hospital, and 62 blood samples at Nippon Medical School Hospital. Control DNA samples were from Keio University campuses.

The Japanese part of the study was approved by the ethics committees of the participating institutions in accordance with the Ethics Guidelines For Human Genome/Gene Analysis Research in Japan. The Korean side of the GC case control study was approved by the Ethics Committee of the National Cancer Center, Korea. Informed consent was obtained from all living subjects, including opt-out consent for the paraffin block archival samples.

LD Analysis

The LD map of chromosome 1q22 (Figure 2) was constructed based on the genotype data of 41 SNPs (Supplementary Table 10) obtained from 680 Japanese controls (436 males; mean age, 43.7; 242 females; mean age, 43.7; 2 missing gender information; 371 of which were from controls in the Tokyo data set) genotyped by Illumina Human610-Quad BeadChip (Illumina, San Diego, CA). The LD map shown in Figure 1 was constructed based on the fine mapping data of the 52 SNPs (Supplementary Table 10) from the Single Nucleotide Polymorphism database (http://www.ncbi.nlm.nih.gov/projects/SNP/) on the 1266 Japanese controls (the same as controls in the Tokyo data set). The pattern of LD was analyzed using 2 parameters, r^2 and $|D'|$,[14] and the confidence interval of the $|D'|$ was also utilized.[15]

Statistical Analyses

Statistical significance of the association was evaluated for each SNP by logistic regression. *P* values under an additive model adjusted for 3 age categories (≤39, 40–59, and ≥60 years), gender, and the risk alleles at rs2294008 in *PSCA* (Supplementary Figures 9–11). The significance level was set to .05 by Bonferroni correction for multiple testing, meaning $P = 8.9 \times 10^{-4}$ before correction for the Tokyo data set. Meta-analyses of the Tokyo data set and the Aichi and Korea replication data sets were performed using a random effects model.[16] Haplotype-based association was tested by Fisher exact test. Haplotype phases in each individual were inferred by fastPHASE software.[17] Other statistical analyses were carried out using the R suite (http://www.r-project.org/) and the StatXact 8 (Cytel Inc, Cambridge, MA). Population stratification of the Tokyo data set was examined previously by the STRUCTURE software,[18] the Genomic Control, and mixture model methods,[19,20] and no significant subpopulation was detected.[7]

In the association studies using 2-locus genotype data of rs4072037 and rs2294008, the biologic effect of the SNPs' risk allele was assumed to be recessive (rs4072037) or dominant (rs2294008), ie, the risk genotype for rs4072037 is AA and, for rs2294008, TT and TC (Figure 3). Risk factor variables consist of 4 categories based on the genotypes of rs4072037 and rs2294008. *P* value and OR and its 95% confidence interval (CI) for each category was obtained by logistic regression adjusted for age and gender.

Other Analyses

The materials and methods used in IGC association studies, genotyping, resequencing, and functional studies are described in Supplementary Materials and Methods and Supplementary Tables 9–11.

Results

Identification of the Susceptibility Region in Chromosome 1q22

Initially, we analyzed LD ($|D'|$) around the 2 marker SNPs based on the genotyping data on 680 control subjects. The criteria based on a confidence interval of $|D'|$[15] was applied to find an LD block containing 12 SNPs (including rs2075570 and rs2070803) and 11 genes (Figure 2). The second analysis of high-density genotyping around this block was performed on 610 cases of DGC and 1266 controls (Tokyo data set) for 52 SNPs selected from the Single Nucleotide Polymorphism database. A solid 49kb-LD block was identified spanning 13 SNPs including rs2075570 and rs2070803. Eight SNPs in the block showed strong associations ($P < 1.0 \times 10^{-5}$)

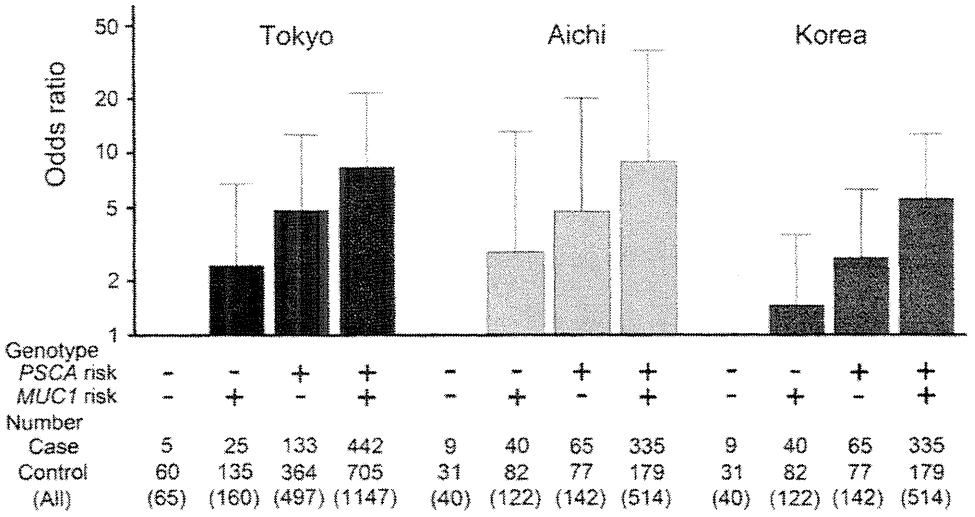


Figure 3. Association studies for DGC using 2-locus genotype data of rs4072037 in *MUC1* and rs2294008 in *PSCA*. The association studies were performed with a distinct model for each risk allele's effect, recessive for rs4072037 and dominant for rs2294008. *Bar*, upper bound of 95% confidence interval.

with DGC (Figure 1 and Table 1). Of the 8 SNPs, 4 were selected and genotyped on independent case-control sets in Japan (Aichi data set: 304 cases, 1467 controls) and in Korea (Korea data set: 455 cases, 372 controls), and the association was replicated in both data sets (Figure 1 and Table 1). A meta-analysis of the 3 case-control studies also showed significant correlation of the region: $P = 2.26 \times 10^{-12}$; OR, 1.71 for rs2075570 and $P = 4.33 \times 10^{-13}$; OR, 1.71 for rs2070803. Because this 1q22 region was originally identified by the GWAS on DGC,[7] an association with IGC was examined on the 3 populations: 601, 274, and 415 cases from Tokyo, Aichi, and Korea, respectively, using the same control subjects analyzed for DGC. The 1q22 region was not significantly associated with IGC (Figure 1, Supplementary Figure 2). Full data of the association studies are shown in Supplementary Tables 1–4.

Polymorphisms in MUC1 *Gene and Haplotype-Based Association Study*

The 49-kilobase (kb) block contained 5 genes encoding keratinocyte associated protein 2 (*KRTCAP2*), tripartile motif protein 46 (*TRIM46*), mucin 1 (*MUC1*), thrombospondin 3 (*THBS3*), and metaxin 1 (*MTX1*). Based on their expression patterns and gene annotations, we prioritized *MUC1* for further analyses because *MUC1* is expressed in pit cells in the pit region, mucous neck cells in the neck region, chief (zymogenic) cells in the base region, and parietal cells in the neck and base regions of the gastric epithelium (Figure 4*A*, Supplementary Figure 1).[21] Moreover, previous studies based on a candidate-gene approach reported an association between its polymorphisms and GC.[8–11]

The resequencing of the *MUC1* gene identified a total of 7 polymorphisms in 48 Japanese individuals: 4 SNPs without rs numbers (numbers 1, 2, 4, and 7), 1 indel (No. 6, rs66597679), rs12411216 (No. 3), and rs4072037 (No. 5) (Figure 4*B* and Supplementary Table 5). The 7 were genotyped on 380 Japanese cases and 309 controls (Table 2), and, in a subsequent haplotype analysis, SNPs numbers 1 and 7 were removed from analysis because they were monomorphic in the 689 Japanese individuals. The remaining 5 SNPs were used for a haplotype-based association study, which revealed 3 major haplotypes, numbers 1–3, with ORs of 1.32, 0.90, and 0.65, respectively, and 1 minor haplotype, No. 4, with minor allele frequency of 0.0105 in cases and 0.0032 in controls (Table 3).

Functional Analyses of MUC1 *SNPs*

Seven transcriptional variants are registered as *MUC1* messenger RNA in the National Center for Biotechnology Information database (http://www.ncbi.nlm.nih.gov/) (Supplementary Figure 3), and the rs4072037 SNP ($P = 1.43 \times 10^{-11}$ and OR of 1.66 by meta-analysis of the 2 Japanese and 1 Korea data sets, Table 1) located in exon 2 of *MUC1* had been found to be related to the splicing site selection in the exon.[12] To identify the variants expressed in the stomach, we conducted RNA ligase-mediated rapid amplification of the 5′ complementary DNA end procedure.

Our results showed that the major transcripts in the stomach are variants 2 and 3 (Figure 4*B* and Supplementary Figure 4) and that all the examined clones of variant 2 possessed G allele at rs4072037, in contrast to those of the variant 3 possessing the A allele, as reported previously (Supplementary Figure 5).[12] This suggests that rs4072037 is significantly involved in the splicing regulation of the second exon. In other words, it is likely that the SNP directly determines the relative dominance of the 2 major *MUC1* splicing variants, the variants 2 and 3, in the gastric epithelium.

As reported previously on the Caucasian population,[12] no polymorphisms other than rs4072037 were found in the region spanning from exon 1 to 2, which might affect the splicing of the second exon, by our resequencing of the

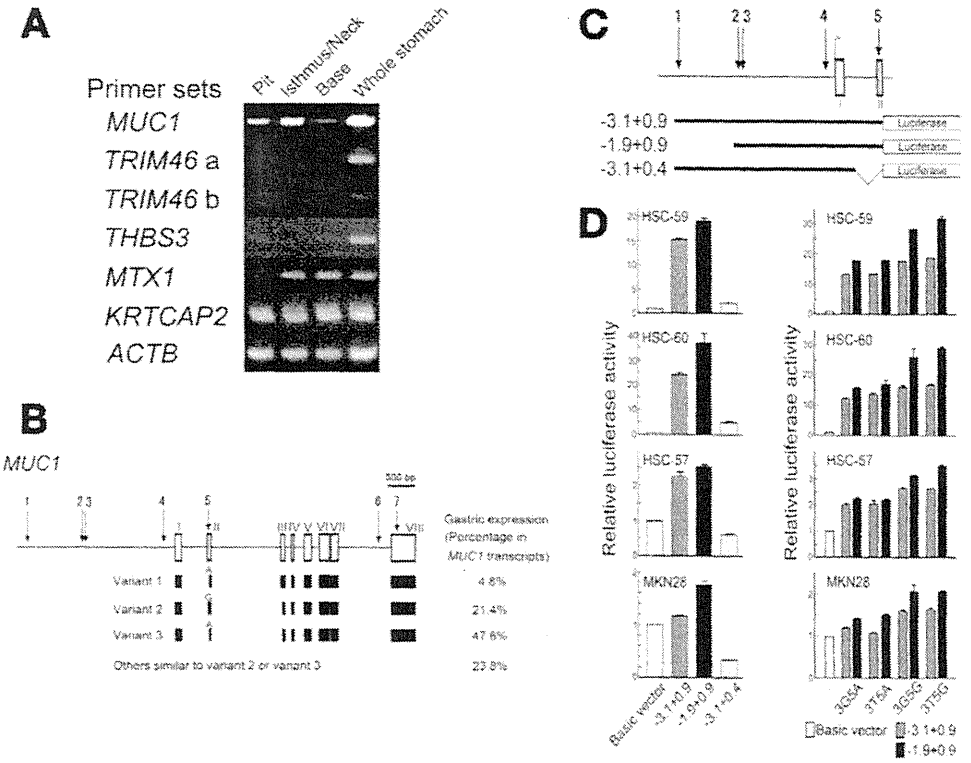


Figure 4. Functional analyses of MUC1 and its SNPs. (*A*) Expression analysis on the 5 genes in the LD block associated with DGC using microdissected gastric samples (reverse-transcription polymerase chain reaction). (*B*) Seven polymorphisms in the *MUC1* gene identified by resequencing of 48 Japanese controls and *MUC1* transcriptional variants detected in RNAs from the gastric mucosa by RNA ligase-mediated rapid amplification of the 5′ complementary DNA (cDNA) end procedure (5′RACE). SNP positions are indicated by *arrows* with numbers corresponding to Table 2 and Supplementary Table 5. The 5′RACE was conducted on a pooled gastric RNA sample from 21 Caucasian individuals. All the variant 2 clones contained exclusively the G allele of SNP No. 5 and all the variants 1 and 3 the A allele, without exception. Complete result of the 5′RACE is presented in Supplementary Figure 4. (*C*) *MUC1* genomic fragments examined in reporter assays. (*D*) Effect of SNP No.5 (rs4072037) on the transcriptional activity of the *MUC1* promoter (reporter assay). The transcriptional activity of the −3.1 to +0.9 region was significantly reduced by truncating the +0.4 to +0.9 region (*−3.1+0.4*) in gastric cancer cell lines HSC-59, HSC-60, HSC-57, and MKN28. The genomic fragments with G allele in the SNP (*3G5G* and *3T5G*) showed higher reporter activity than that with A allele (*3G5A* and *3T5A*).

MUC1 gene on 48 Japanese individuals (Figure 4*B* and Supplementary Table 5). As regards the splicing variation, the risk allele A at rs4072037, which is found in variants 1 and 3 (Supplementary Figure 5), causes a 9-amino acid deletion in the second exon and consequently modifies both the signal peptide and N-terminal amino acid of the mature protein by changing the signal-peptide cleavage site.[12] This may change the intracellular trafficking and glycosylation and folding of the protein, leading to alteration in the function of the mature protein.

To examine the difference in the MUC1 function among the variants, we evaluated in vitro functions of the

Table 2. Seven SNPs and Their Association With DGC Based on the Genotype Data of 380 Japanese Cases and 309 Controls

SNP no.	rs number	Major allele	Minor allele	Risk allele	MAF (case)	MAF (control)	OR[a]	95% CI	*P* value[b]
1		T	C		0.0000	0.0000			
2		C	T	C	0.1882	0.2039	1.11	0.85–1.44	.4945
3	12411216	G	T	G	0.1289	0.1851	1.53	1.14–2.06	.004344
4		G	A	A	0.0105	0.0032	3.28	0.69–15.49	.1998
5	4072037	A	G	A	0.1289	0.1857	1.54	1.15–2.07	.004275
6	66597679	AC	—	AC	0.1308	0.1869	1.53	1.14–2.05	.005225
7		C	T		0.0000	0.0000			

CI, confidence interval; MAF, minor allele frequency.
[a]Odds ratio for risk allele.
[b]*P* values obtained by Fisher exact test.

Table 3. Four Major Haplotypes Inferred and Their Association With DGC Based on the Genotype Data of 380 Japanese Cases and 309 Controls

	SNP no.									*P* values	
	2	3	4	5	6	Case[a]	Control[b]	OR	95% CI	Fisher[c]	Permutation[d]
Haplotype no.											
1	C	G	G	A	AC	0.6733	0.6084	1.32	1.06–1.65	.01501	.0483
2	T	G	G	A	AC	0.1884	0.2039	0.90	0.69–1.18	.4945	.9027
3	C	T	G	G	—	0.1278	0.1845	0.65	0.48–0.87	.004200	.0127
4	C	G	A	A	AC	0.0105	0.0032	3.28	0.69–15.49	.1998	.4058

CI, confidence interval.
[a]Frequency of case.
[b]Frequency of control.
[c]*P* values obtained by Fisher exact test.
[d]*P* values obtained by permutation test (100,000 permutations performed).

2 major variants expressed in the gastric epithelium: variants 2 and 3. Because the full-length product of *MUC1* is well-known for its cell growth-promoting activity in cancer cells,[22,23] we individually transfected a cytomegalovirus promoter-driven complementary DNA of *MUC1* variant 2 or 3 to the MKN28 cells, which express *MUC1* at an undetectable level (data not shown). Examination of their cell growth by both cell counting and colorimetric methods suggested that variant 2 is more potent in growth-promoting activity than variant 3 (Supplementary Figures 6 and 7). Although the observed difference seems to be small, this level of normal range of individual variation is generally expected for a common genetic variant influencing a common disease susceptibility and is probably because of the difference in the signal peptide or the N-terminal structure of the mature protein encoded by each variant because the other portion of the amino acid sequence is common between the 2 variants.

Next, we investigated the function of the *MUC1* SNPs in the context of the haplotypes. We selected haplotype No. 1 as the major risk haplotype and haplotype No. 3 as the most protective haplotype and analyzed the functions of SNPs numbers 3 and 5, excluding SNP No. 6 from our functional analyses because its location in the intron 7 made it unlikely to be involved in the transcriptional regulation and/or alternative splicing of the gene. SNP numbers 2 and 4 were also excluded because they were found on both the risk and the protective haplotypes. Because the remaining 2 SNPs, numbers 3 and 5, do not change amino acid, we first examined, by a reporter assay, their effect on the transcriptional regulation of *MUC1*; the region spanning −1.9 to 0.9 kb relative to the *MUC1* transcription start site had a transcriptional activity (Figure 4*C* and *D*). The reporter assay on base-substituted constructs showed that, in all the gastric carcinoma cell lines examined, the fragments containing the G allele at rs4072037 (SNP No. 5), which is present only in the protective haplotype, has a higher transcriptional activity than that with an A allele present in the risk haplotype (Figure 4*D*, *right panel*). The assay on truncated constructs showed that a removal of a +0.4 to +0.9-kb region, which contains rs4072037, significantly diminishes the transcriptional activity (Figure 4*D*, *left panel*). We also confirmed in the reporter assay that the T allele of SNP No. 2, which is unique to haplotype No. 2 showing OR of 0.9 but no significant *P* value, has no effect (Supplementary Figure 8).

In sum, the results in this study and from previous reports by other investigators suggest that rs4072037 SNP has at least 2 functions: (1) regulation of the alternative splicing at the second exon and (2) modification of the transcriptional activity of the promoter. The association study in the context of LD and the functional study strongly implicate rs4072037 as a functional cause of the association between the 1q22 region and DGC susceptibility.

Association Studies for DGC Using 2-Locus Genotype Data of rs4072037 in MUC1 *and rs2294008 in* PSCA

Finally, we examined the effect of 2 DGC susceptibility SNPs identified by our GWAS, rs4072037 in *MUC1* and rs2294008 in *PSCA*, both of which are functional, using Tokyo, Aichi, and Korea data sets. When a genetic model is tentatively selected for each locus by simply comparing *P* value, a recessive and dominant model was applied for rs4072037 and for rs2294008, respectively; the individuals possessing the risk genotype of both SNPs showed significant risk for developing DGC (eg, OR, 8.38 in Tokyo data set, Figure 3). Notably, individuals with protective alleles of both SNPs were observed only in controls (Supplementary Table 6).

Discussion

At chromosome 1q22, we focused on the region with strong LD around rs2075570 and rs2070803 using Gabriel et al's criteria.[15] The region contains 5 genes: *TRIM46*, *THBS3*, *MTX1*, *KRTCAP2*, and *MUC1* (Figure 1). We prioritized the genes for the subject of

further studies by the first criterion (whether the gene is expressed in the gastric epithelial cells) and then by the second one (whether annotated function suggests its involvement in carcinogenesis). We observed transcripts of 4 of the 5 genes in microdissected samples of the gastric epithelium by reverse-transcription polymerase chain reaction, but no transcript of *TRIM46* was detectable there (Figure 4*A*). *THBS3* encoding a multifunctional extracellular matrix glycoprotein is expressed in multiple human tissues including the stomach,[24] and no evidence of a causal relation to carcinogenesis has been obtained. *MTX1* encodes a component of a preprotein import complex in the outer membrane of the mammalian mitochondrion.[25] If it is involved in carcinogenesis, the effect of its SNP would be reflected in many types of cancer, yet no such involvement has surfaced. *KRTCAP2* encodes a protein possessing transmembrane domain, showing multitissue expression.[26] Its function is unknown, and its relation to carcinogenesis has not been demonstrated.

In this study, we considered *MUC1* as a strong candidate for the gene responsible for the association of 1q22 with DGC because, in addition to several previous candidate gene analyses showing an association with *MUC1* polymorphisms and GC,[8–11] MUC1 has been considered to possess an oncogenic property as described below.

Mucin family members are classified into 2 types, secreted or membranous, based on their localization, and MUC1 is a transmembrane mucin.[27] MUC1 is a multifunctional protein involved in mucosal lubrication, protection from pathogens, signal transduction, and cell-cell interaction.[27] MUC1 was over-expressed in breast, ovarian, lung, pancreatic, and prostate cancers and was a marker of poor prognosis in gastric cancer.[21,28] Several in vivo studies have provided evidence supporting its function in carcinogenesis. MUC1 has a role in cell growth, anchorage independence, cell migration, antiapoptotic property, and drug resistance of cancer cells,[22,29–33] all of which are accomplished through interaction with several signaling pathways,[34] although these lines of biologic evidence were obtained on the standard molecule containing tandem repeats (TR). Because the MUC1 expressed in the gastric epithelium has no TR, it is possible that its function in gastric epithelial cells is different from that of the TR-containing product in other epithelial cells.[35–37]

Recently, however, it is supposed that MUC1 has a protective function against environmental insults and acts against tumorigenesis in normal epithelial cells, which keep maintaining their cell polarity. In contrast, once the cells lose cell polarity in consequence of prolonged inflammation, MUC1 promotes cell growth and acts for tumorigenesis.[38] It was also reported that MUC1 functions as a growth factor receptor in human embryonic stem cells.[39] It is presumable that MUC1 is involved in growth regulation of gastric stem cells and progenitors, which are considered to be the origin of DGC. Like the function of PSCA, which is up-regulated in prostate and urinary bladder cancers but suppressed in gastric cancer,[7] the MUC1 function may differ between cell types, normal or malignant, and among different tissues. In the same manner as PSCA, MUC1 is down-regulated in intestinal metaplasia of the gastric epithelium from which IGC arises.[40] If MUC1 has some protective function in carcinogenesis, this down-regulation makes the stem and progenitor cells more susceptible to carcinogenic events. In any case, further research is needed to explore the pleiotropic functions of MUC1.

In this study, we demonstrated that rs4072037 has a role in transcriptional regulation and also in splicing site selection leading to the dominant variant determination of *MUC1* transcripts in gastric epithelial cells. If variant 3 is less functional in protection against DGC than variant 2, the possession of the A allele in the genome confers both quantitatively and qualitatively unfavorable consequences to MUC1 function, which may result in additive risk for DGC susceptibility.

In our GWAS on DGC susceptibility, the 2 loci showing the highest statistical significance directed us to the 2 functional SNPs: rs4072037 in *MUC1* and rs2294008 in *PSCA*.[7] It is noteworthy that both SNPs in the 2 genes appear to have dual functions: transcriptional regulation and signal-peptide modification.[7] Further investigation is required to validate the role of *MUC1* in DGC susceptibility and details of the mechanism that links the risk haplotype tagged by rs4072037 to DGC development.

The risk allele A of rs4072037 is in strong LD with the small allele of the variable numbers of tandem repeats in the second intron of the *MUC1* both in Europeans and Japanese; more than 90% of chromosomes have a nonrecombinant haplotype in both populations (Ng et al[12] and Supplementary Table 7), and the small allele was associated with GC in the European population.[8,9,12] However, the variable numbers of tandem repeats is unlikely to be the causal polymorphism for DGC susceptibility because the TRs are translated neither in normal nor malignant gastric epithelial cells (Supplementary Figure 4 and Supplementary Table 8).

This study has not only replicated the association of the *MUC1* SNP with GC in the Japanese and Korean populations, in addition to the previous reports on the Chinese and Caucasian population GCs,[8–11] but it has also disclosed that the association appears specific to DGC. Following discovery of the DGC-specific association of the *PSCA* polymorphism, this study has offered another piece of evidence to support distinct mechanisms for DGC and IGC development.

Although there was no significant interaction between the *MUC1* and *PSCA* SNPs for the DGC risk in our study (Figure 3, Supplementary Table 6, and data not shown), it is estimated that individuals with the double risk

genotype are the majority in Japanese (56%) and Korean (49%) populations with a significant OR, 8.38, in Japanese, as compared with the lowest risk category. GWAS and other emerging genome analysis tools may unveil a number of polymorphisms showing a significant statistical association, but it is important to identify functional SNPs potentially related to carcinogenesis. The accumulation of information on the functional SNPs, environmental factors, and their interactions, all of which are truly related to DGC susceptibility, will make the genotyping a more practical tool for evaluating the individual risk for DGC and offer effective prevention strategies in the future.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at doi: 10.1053/j.gastro.2010.10.058.

References

1. Brenner H, Rothenbacher D, Arndt V. Epidemiology of stomach cancer. In: Verma M, ed. Methods of molecular biology: cancer epidemiology. Volume 472. New Jersey: Humana, 2009:467–477.
2. Lauren P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. Acta Pathol Microbiol Scand 1965; 64:31–49.
3. Chiba T, Marusawa H, Seno H, et al. Mechanism for gastric cancer development by *Helicobacter pylori* infection. J Gastroenterol Hepatol 2008;23:1175–1181.
4. Schier S, Wright NA. Stem cell relationships and the origin of gastrointestinal cancer. Oncology 2005;69(Suppl 1):9–13.
5. Henson DE, Dittus C, Younes M, et al. Differential trend in the intestinal and diffuse types of gastric carcinoma in the United States, 1973–2000—increase in the signet ring cell type. Arch Pathol Lab Med 2004;128:765–770.
6. Rosai J. Rosai and Ackerman's surgical pathology. Edinburgh: Mosby, 2004.
7. Study Group of Millennium Genome Project for Cancer. Genetic variation in *PSCA* is associated with susceptibility to diffuse-type gastric cancer. Nat Genet 2008;40:730–740.
8. Carvalho F, Seruca R, David L, et al. *MUC1* gene polymorphism and gastric cancer—an epidemiological study. Glycoconj J 1997; 14:107–111.
9. Silva F, Carvalho F, Peixoto A, et al. *MUC1* gene polymorphism in the gastric carcinogenesis pathway. Eur J Hum Genet 2001;9: 548–552.
10. Xu Q, Yuan Y, Sun LP, et al. Risk of gastric cancer is associated with the *MUC1* 568 A/G polymorphism. Int J Oncol 2009;35: 1313–1320.
11. Jia Y, Persson C, Hou L, et al. A comprehensive analysis of common genetic variation in *MUC1, MUC5AC, MUC6* genes and risk of stomach cancer. Cancer Causes Control 2010;21:313–321.
12. Ng W, Loh AX, Teixeira AS, et al. Genetic regulation of *MUC1* alternative splicing in human tissues. Br J Cancer 2008;99:978–985.
13. Japanese Gastric Cancer Association. Japanese classification of gastric carcinoma. 2nd English edition. Gastric Cancer 1998;1: 10–24.
14. Nei M. Molecular evolutionary genetics. New York: Columbia University Press, 1987.
15. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in human genome. Science 2002;296:2225–2229.
16. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177–188.
17. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 2006; 78:629–644.
18. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 2003;164:1567–1587.
19. Bacanu SA, Devlin B, Roeder K. The power of genomic control. Am J Hum Genet 2000;66:1933–1944.
20. Zhu X, Zhang S, Zhao H, et al. Association mapping, using a mixture model for complex traits. Genet Epidemiol 2002;23: 181–196.
21. Senapati S, Sharma P, Bafna S, et al. The MUC gene family: their role in the diagnosis and prognosis of gastric cancer. Histol Histopathol 2008;23:1541–1552.
22. Li Y, Liu D, Chen D, et al. Human DF3/MUC1 carcinoma-associated protein functions as an oncogene. Oncogene 2003;4: 6107–6110.
23. Tsutsumida H, Swanson BJ, Singh PK, et al. RNA interference suppression of MUC1 reduces the growth rate and metastatic phenotype of human pancreatic cancer cells. Clin Cancer Res 2006;15:2976–2987.
24. Adolph KW. Relative abundance of Thrombospondin 2 and Thrombospondin 3 mRNAs in human tissues. Biochem Biophys Res Commun 1999;258:792–796.
25. Armstrong LC, Saenz AJ, Bornstein P. Metaxin 1 interacts with metaxin 2, a novel related protein associated with the mammalian mitochondrial outer membrane. J Cell Biochem 1999;74: 11–22.
26. Bonkobara M, Das A, Takao J, et al. Identification of novel genes for secreted and membrane-anchored proteins in human keratinocytes. Br J Dermatol 2003;148:654–664.
27. Gendler SJ. MUC1, the renaissance molecule. J Mammary Gland Biol Neoplasia 2001;6:339–353.
28. Taylor-Papadimitriou J, Burchell J, Miles DW, et al. MUC1 and cancer. Biochim Biophys Acta 1999;1455:301–313.
29. Suwa T, Hinoda Y, Makiguchi Y, et al. Increased invasiveness of *MUC1* cDNA-transfected human gastric cancer MKN74 cells. Int J Cancer 1998;76:377–382.
30. Raina D, Kharbanda S, Kufe D. The MUC1 oncoprotein activates the anti-apoptotic phosphoinositide 3-kinase/Akt and Bcl-xL pathways in rat 3Y1 fibroblasts. J Biol Chemi 2004;279:20607–20612.
31. Ren J, Agata N, Chen D, et al. Human MUC1 carcinoma-associated protein confers resistance to genotoxic anticancer agents. Cancer Cell 2004;5:163–175.
32. Huang L, Chen D, Liu D, et al. MUC1 oncoprotein blocks glycogen synthase kinase 3β-mediated phosphorylation and degradation of β-catenin. Cancer Res 2005;65:10413–10422.
33. Wei X, Xu H, Kufe D. Human mucin 1 oncoprotein represses transcription of the *p53* tumor suppressor gene. Cancer Res 2007;67:1853–1858.
34. Singh PK, Hollingsworth MA. Cell surface-associated mucins in signal transduction. Trends Cell Biol 2006;16:467–476.
35. Gendler SJ, Lancaster CA, Taylor-Papadimitriou J, et al. Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin. J Biol Chem 1990;265:15286–15293.

36. Zrihan-Licht S, Vos HL, Baruch A, et al. Characterization and molecular cloning of a novel MUC1 protein, devoid of tandem repeats, expressed in human breast cancer tissue. Eur J Biochem 1994;224:787–795.
37. Baruch A, Hartmann M, Zrihan-Licht S, et al. Preferential expression of novel MUC1 tumor antigen isoforms in human epithelial tumors and their tumor-potentiating function. Int J Cancer 1997; 71:741–749.
38. Kufe DW. Mucins in cancer: function, prognosis and therapy. Nat Rev Cancer 2009;9:874–885.
39. Hikita ST, Kosik KS, Clegg DO, et al. MUC1* mediates the growth of human pluripotent stem cells. PLoS One 2008;3:e3312.
40. Silva E, Teixeira A, David L, et al. Mucins as key molecules for the classification of intestinal metaplasia of the stomach. Virchows Arch 2002;440:311–317.

Received May 8, 2010. Accepted October 26, 2010.

Reprint requests

Address requests for reprints to: Teruhiko Yoshida, MD, Genetics Division, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. e-mail: tyoshida@ncc.go.jp; fax: (81) 3-3541-2685.

Acknowledgments

The authors thank Sachiyo Mimaki, Mineko Ushiama, Chie Naito, Yoko Odaka, Misuzu Okuyama, Miki Watanabe, and Hidemi Ito for their technical contributions and Seiji Ito for the clinical data collection of the Aichi data set.

Conflicts of interest

The authors disclose no conflicts.

Funding

Supported in Japan by the program for promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NiBio); in Aichi, supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, Culture and Technology of Japan and by a Grant-in-Aid for the Third Term Comprehensive 10-Year Strategy for Cancer Control from the Ministry of Health, Labour and Welfare of Japan; and, in Korea, part of the study was supported by grant 0710340 from the National Cancer Center, Korea.

Original Paper

Pathobiology

Pathobiology 2012;79:154–161
DOI: 10.1159/000335694

Received: November 7, 2011
Accepted after revision: December 12, 2011
Published online: January 28, 2012

Cytokeratin Expression Profiling in Gastric Carcinoma: Clinicopathologic Significance and Comparison with Tumor-Associated Molecules

Hokuto Takami Kazuhiro Sentani Miho Matsuda Naohide Oue
Naoya Sakamoto Wataru Yasui

Department of Molecular Pathology, Hiroshima University Graduate School of Biomedical Sciences, Hiroshima, Japan

Key Words
Gastric cancer · Cytokeratin · Mucin phenotype · Tumor-associated molecules

Abstract
Objective: The expressions of cytokeratin (CK) 7 and 20 have been studied in various primary and metastatic carcinomas, and their determination may help distinguish the site of origin of metastatic carcinomas. However, little is known about the molecular basis that determines variations in CK patterns in gastric cancers (GCs). The aim of the present study was to analyze the CK expression patterns in a large number of GCs and to investigate how the CK patterns correlate with clinicopathologic parameters, histology, mucin phenotype or several tumor-related molecules. ***Methods and Results:*** We immunohistochemically examined the CK7/CK20 patterns, mucin expression profiles (MUC5AC, MUC6, MUC2 and CD10), and the cancer-related molecules (CDX2, p53, EGFR and β-catenin), using a tissue microarray with 870 GCs. The GCs were divided into four patterns; 17% of CK7+/CK20+, 57% of CK7+/CK20–, 9% of CK7–/CK20+ and 17% of CK7–/CK20. GCs with the CK7–/CK20– pattern demonstrated a close relation to undifferentiated adenocarcinoma. CK7 expression was significantly correlated with the expression of MUC5AC and MUC6, while CK20 expression was correlated with MUC2 and CDX2. There were statistically significant associations between CK expression patterns and mucin phenotypes. ***Conclusion:*** These results indicate that the CK7/CK20 expression patterns in GCs demonstrated different clinicopathologic features and molecular signatures.

Introduction

Gastric cancer (GC) is one of the most common malignancies worldwide and often metastasizes to other organs, including the liver, lung and ovary [1]. In metastatic carcinomas of unknown primary site, identification of the origin as the stomach or other primary site is very difficult because of the heterogeneous histology of GC. This heterogeneity may be partly due to the fact that GC is caused by exogenous (nitrosamines and *Helicobacter pylori*) and endogenous (E-cadherin mutation) factors [2]. To identify these phenotypical differences at a morphologic level in a comparable and reproducible manner, it is

KARGER
Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1015–2008/12/0793–0154$38.00/0

Accessible online at:
www.karger.com/pat

Wataru Yasui, MD, PhD
Department of Molecular Pathology
Hiroshima University Graduate School of Biomedical Sciences
1-2-3 Kasumi, Minami-ku, Hiroshima 734-8551 (Japan)
Tel. +81 82 257 5145, E-Mail wyasui@hiroshima-u.ac.jp

necessary to have a novel classification system that recognizes all types of GCs adequately. In addition to classification by histology (the Lauren classification, the Japanese Classification of Gastric Carcinoma, etc.), GCs may also be classified into four phenotypes by the mucin expression profile: G type (gastric phenotype), I type (intestinal phenotype), GI type (gastric and intestinal mixed phenotype) and N type (neither gastric nor intestinal phenotype) [2, 3]. The G type is considered to behave more aggressively than the I type. Mutations of *p53* and loss of heterozygosity of the adenomatous polyposis coli gene occur more frequently in the I type than in the G type, while microsatellite instability and alterations in the *p73* gene are more common in the G type than in the I type. Microsatellite instability in the G type is usually associated with inactivation of hMLH1 following the promoter hypermethylation. The caudal-related homeobox gene 2 (CDX2) acts as an intestine-specific transcription factor and is expressed in I type GCs at high levels. CDX2 upregulates the expression of goblet-specific MUC2 [4].

Cytokeratin (CK), an intermediate filament observed mainly in epithelial cells, is an important cytoskeletal component involved in fixation of the nucleus and maintenance of cell morphology. CK consists of 20 subtypes, whose expression depends primarily on the epithelial cell type and the degree of differentiation [5]. The expression of CK8, CK18 and CK19 is observed at all levels of gastric mucosa, but CK20 expression is limited to mature superficial foveolar epithelium [6–8]. CK7 expression is absent in normal gastric mucosa but is observed in chronic mucosal irritation conditions such as *H. pylori* gastritis in the basal localization [9]. There are many reports of CK expression in various organs and the findings suggest that the CK expression profiles of metastatic cancers correspond to those of the primary sites [10–12]. In particular, in the gastrointestinal tract, colorectal carcinomas demonstrate a CK7–/CK20+ expression pattern, whereas adenocarcinomas of foregut origin demonstrate a CK7+/CK20– expression pattern [10–13]. Profiles for CK expression patterns may be helpful; however, the total number of GCs in each previous report about the expression of CK subtypes is one hundred cases at most [5, 6, 9, 12–21].

The aim of the present study was to analyze the CK expression patterns in a large number of GCs and to investigate how the CK expression patterns correlate with clinicopathologic parameters, histology, mucin phenotypes or several tumor-related molecules. Because the functional and biological properties of the GCs may reflect the tumors' ability to produce certain CKs, it would be of interest to determine which factors are best correlated with the CK immunophenotype.

Materials and Methods

Tissue Samples and Tissue Microarray Construction

The surgical pathology files of the Hiroshima University Hospital and its affiliated hospitals were used to randomly select 870 GCs from 51 adenocarcinomas of the esophagogastric junction (AEG) and 819 distal GCs (corpus or antrum). Surgically resected specimens were routinely fixed in 10% buffered formalin and examined macroscopically. Tumor staging was performed according to the Union Internationale Contre le Cancer (UICC) system [22]. There were 403 T1, 257 T2, 184 T3 and 26 T4 in these 870 cases. Nodal metastasis was present in 366 patients (42%). Tumor staging revealed 500 stage I, 152 stage II, 154 stage III and 64 stage IV. GCs were histologically classified as 507 intestinal type and 363 diffuse type according to the Lauren classification system. The 507 intestinal type GCs included 63 papillary, 173 well-differentiated tubular and 271 moderately differentiated tubular adenocarcinomas. The 363 diffuse type GCs consisted of 297 poorly differentiated adenocarcinomas, 52 signet-ring cell carcinomas and 14 mucinous adenocarcinomas according to the WHO histological classification. In addition, the GCs were classified as 63 papillary (pap), 173 well differentiated tubular (tub1), 271 moderately differentiated tubular (tub2), 75 solid type poorly differentiated (por1), 222 non-solid type poorly differentiated (por2) adenocarcinomas and 52 signet-ring cell carcinoma (sig) and 14 mucinous adenocarcinomas (muc) according to the Japanese Classification of Gastric Carcinoma [23]. In accordance with the Ethical Guidelines for Human Genome/Gene Research enacted by the Japanese Government, tissue specimens were collected and used after approval from the Ethical Review Committee of the Hiroshima University School of Medicine and from the ethical review committees of collaborating organizations.

The two most representative portions to be sampled for the tissue microarrays (TMAs) were carefully selected from different intratumoral areas in each case and marked on the HE-stained slide. Two superficial areas in mucosal GCs, and one superficial area and one deep area in GCs that had invaded beyond the submucosa were selected. A 2-mm-diameter tissue core of each donor block was punched out and transferred to a recipient block with a maximum of 48 cores using a tissue microarrayer (AZUMAYA KIN-1, Tokyo, Japan). 5-μm-thick sections were cut from the recipient block and transferred to glass slides. HE staining was performed on TMA for confirmation of the tumor tissue. Each tissue-array block contained 21 cases of GC and four cases of non-neoplastic stomach samples.

Immunohistochemistry

A Dako Envision Kit (Dako, Carpinteria, Calif., USA) was used for immunohistochemical analysis of all markers. In brief, sections were pretreated by microwaving (500 W) in a citrate buffer (pH 6.0) for 15 min to retrieve antigenicity. After endogenous peroxidase activity was blocked with 3% H_2O_2-methanol for 10 min, sections were incubated with normal goat serum (Dako) for

Table 1. Antibodies used in the current study and their positive rates in 870 GCs

Antibody	Clone	Dilution	Source	Positive cases n (%)
CK7	OV-TL 12/30	1:50	DAKO, Carpinteria, Calif., USA	648 (74)
CK20	IT-Ks 20.8	1:50	DAKO, Carpinteria, Calif., USA	232 (27)
MUC5AC	CLH2	1:50	Novocastra, Newcastle, UK	519 (60)
MUC6	CLH5	1:50	Novocastra, Newcastle, UK	233 (27)
MUC2	Ccp58	1:50	Novocastra, Newcastle, UK	199 (23)
CD10	56C6	1:50	Novocastra, Newcastle, UK	71 (8)
p53	DO-7	1:50	Novocastra, Newcastle, UK	293 (34)
EGFR	EGFR.113	1:50	Novocastra, Newcastle, UK	109 (13)
CDX2	AMT28	1:20	BioGenex, San Ramon, Calif., USA	183 (21)
β-Catenin	14/β-catenin	1:50	BD Biosciences, San Jose, Calif., USA	140 (16)

The cutoff point for antibody reactivity necessary to define a result as positive was staining of at least 10% of cancer cells in the TMAs.

20 min to block nonspecific antibody binding sites. Sections were then incubated with the following primary antibodies: anti-CK7, anti-CK20, anti-MUC5AC, anti-MUC6, anti-MUC2, anti-CD10, anti-p53, anti-EGFR, anti-CDX2 and anti-β-catenin. Suppliers and working dilutions are noted in table 1. Sections were incubated with a primary antibody for 1 h at 25°C, followed by incubations with peroxidase-labeled anti-rabbit or mouse IgG for 60 min. Staining was completed with a 10-min incubation with the substrate-chromogen solution. The sections were counterstained with 0.1% hematoxylin. Appropriate positive and negative control samples were used.

Evaluation of Positive Cases and Cutoff-Point Thresholds

Immunostaining results were evaluated independently by 3 investigators (H.T., K.S. and M.M.) and when the evaluations differed, a decision was made by consensus while the investigators reviewed the specimen with a multihead microscope. Neoplastic tissue was evaluated semiquantitatively at magnifications of ×100 and ×400. Cytoplasmic immunoreactivity for CK7, CK20, MUC5AC, MUC6 and MUC2, membranous reactivity for CD10 and EGFR, and nuclear reactivity for p53, CDX2 and β-catenin were assessed (fig. 1). The cutoff point for antibody reactivity necessary to define a result as positive was staining of more than 10% tumor cells in the TMAs.

CK Expression Profiles and Mucin Phenotypes

The 870 GCs were evaluated according to the CK7 and CK20 staining pattern and classified into four main groups: (1) coexpression of CK7 and CK20 (CK7+/CK20+), (2) no expression of CK7 and CK20 (CK7–/CK20–), (3) only CK7 expression (CK7+/CK20–) and (4) only CK20 expression (CK7–/CK20+). The criteria [24] for classification of G type and I type were as follows: GCs in which more than 10% of the cells displayed the gastric (MUC5AC and/or MUC6) or intestinal epithelial cell phenotype (MUC2 and/or CD10) were G type or I type, respectively. Those sections that showed both G and I types were classified as GI type, and those that lacked both G and I types were classified as N type.

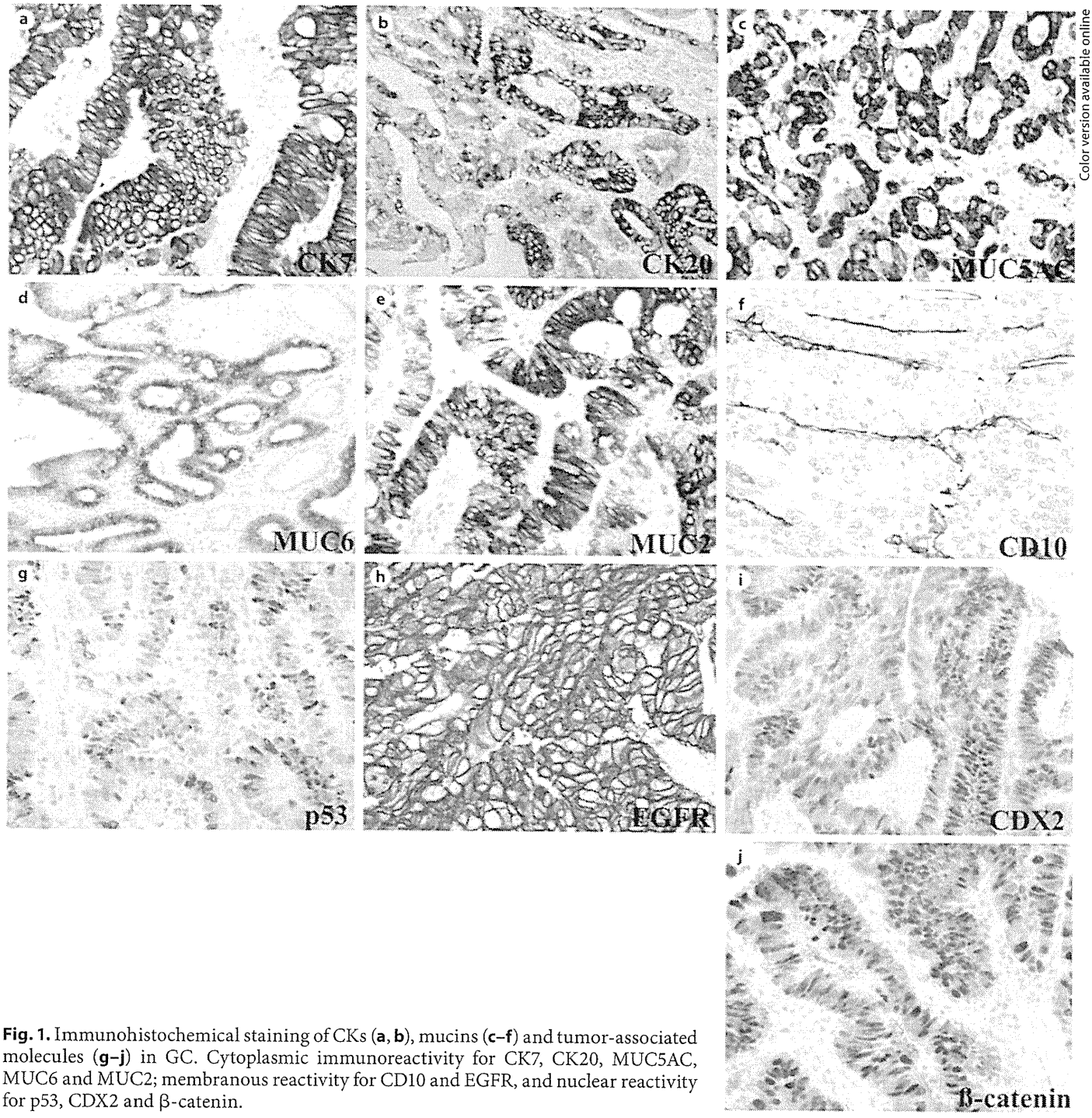
Statistical Methods

Associations between CK expression profiling and clinicopathologic variables, or immunostaining for various markers were analyzed by the χ^2 test. $p < 0.001$ was considered statistically significant.

Results

CK Expression Patterns in GCs and Their Correlation with Clinicopathologic Parameters

Immunohistochemical results in the current study are shown in table 1. The 870 GCs included 648 (74%) cases with CK7 expression and 232 (27%) cases with CK20 expression, and were classified into 156 (17%) cases with CK7+/CK20+ pattern, 492 (57%) cases with CK7+/CK20– pattern, 76 (9%) cases with CK7–/CK20+ pattern and 146 (17%) cases with CK7–/CK20– pattern. We investigated the relation between CK expression patterns and the clinicopathologic parameters including age, sex, tumor location, T grade, N grade, M grade, staging and histological type according to the Lauren classification. As shown in table 2, the CK7–/CK20– pattern was observed more frequently in the diffuse type of GC than in the intestinal type ($p = 0.0003$). In contrast, no differential trend was found between other CK expression patterns and clinicopathologic parameters. Regarding the Japanese Classification of Gastric Carcinomas, the CK7–/CK20– pattern was observed more frequently in the undifferentiated type (por1, por2, sig and muc) than in the differentiated type GC (pap, tub1 and tub2) (fig. 2; $p = 0.0003$). Furthermore, the analysis according to the WHO classification also yielded a similar result.



Color version available online

Fig. 1. Immunohistochemical staining of CKs (**a**, **b**), mucins (**c–f**) and tumor-associated molecules (**g–j**) in GC. Cytoplasmic immunoreactivity for CK7, CK20, MUC5AC, MUC6 and MUC2; membranous reactivity for CD10 and EGFR, and nuclear reactivity for p53, CDX2 and β-catenin.

Distribution of Mucin Phenotypes and Histological Types of GC

Next, we analyzed the relationships between the histological type and the mucin phenotype in the GCs. The 870 GCs included 519 (60%) cases with MUC5AC expression, 233 (27%) cases with MUC6 expression, 199 (23%) cases with MUC2 expression and 71 (8%) cases with CD10 expression. They were classified into 313 (36%) G type, 196 (23%) GI type, 149 (17%) I type and 212 (24%) N type. The distribution of each mucin phenotype and

Table 2. Relationships between CK expression patterns and clinicopathological findings in 870 GCs

CK pattern		CK7+/CK20+ (n = 156)	CK7+/CK20– (n = 492)	CK7–/CK20+ (n = 76)	CK7–/CK20– (n = 146)	p value
Age	>65 years	74 (47)	258 (52)	35 (46)	79 (54)	NS
	≤65 years	82 (53)	234 (48)	41 (54)	67 (46)	
Sex	Male	93 (60)	322 (65)	41 (54)	97 (66)	NS
	Female	63 (40)	170 (35)	35 (46)	49 (34)	
Tumor location	AEG	10 (6)	28 (6)	6 (8)	7 (5)	NS
	Distal GC	146 (94)	464 (94)	70 (92)	139 (95)	
T grade	T1	91 (58)	214 (43)	29 (38)	69 (47)	NS
	T2/T3/T4	65 (42)	278 (57)	47 (62)	77 (53)	
N grade	N0	107 (69)	279 (57)	36 (47)	82 (56)	NS
	N1	49 (31)	213 (43)	40 (53)	64 (44)	
M grade	M0	155 (99)	484 (98)	76 (100)	146 (100)	NS
	M1	1 (1)	8 (2)	0	0	
Staging	Stage I	106 (68)	277 (56)	36 (47)	81 (55)	NS
	Stage II/III/IV	50 (32)	215 (44)	40 (53)	65 (45)	
Histology	Intestinal type	105 (67)	285 (58)	52 (68)	61 (42)	0.0003
	Diffuse type	51 (33)	207 (42)	24 (32)	85 (58)	

Values in parentheses are percentages. AEG = Adenocarcinoma of the esophagogastric junction; NS = not significant. A $p < 0.001$ was considered statistically significant by χ^2 test. Tumor staging was classified according to the criteria of the International Union Against Cancer TNM classification of malignant tumors. Histology was according to the Lauren classification system.

histological type according to the Japanese Classification of Gastric Carcinomas is shown in figure 2. The N type was observed more frequently in the undifferentiated type GC.

Association of Expression between CK Patterns and Various Molecules

We then investigated the association between CK expression patterns and various molecules in the GCs. Of the 870 GCs examined, each tumor-associated molecule was detected in 293 (34%) cases for p53, 109 (13%) cases for EGFR, 183 (21%) cases for CDX2 and 140 (16%) cases for β-catenin (table 1). There are statistically significant associations between CK7 and MUC5AC expression, CK7 and MUC6 expression, CK20 and MUC2 expression, and CK20 and CDX2 ($p < 0.0001$) (fig. 3).

Association of CK Expression Patterns with Mucin Phenotypes

The relationship between each CK expression pattern and mucin phenotype in the 870 GCs was analyzed. As shown in figure 4, there are statistically significant associations between CK7 expression without CK20 expression and G type, CK7 expression and GI type, CK20 expression without CK7 expression and I type, and neither of them and N type, respectively ($p < 0.0001$).

Discussion

Much interest has focused on CK immunoprofiles in the classification of carcinomas, in particular the CK7 and CK20 profiles. There are many reports on the expression of CK7 and CK20 in relatively small numbers of GC cases [5, 6, 9, 12–21]. CK7 immunoreactivity is reported to range from 10 to 75%, while CK20 expression is frequently reported to range from 30 to 50%. Our results are also consistent with the findings of previous reports. The differences of positive rates in these previous reports are possibly due to evaluation scales and case groups. In this study, we used the TMA method to examine each molecule expression in the GCs. Although minute TMAs cannot ensure representative areas of donor specimen, we used 2-mm-diameter needles, which are large enough to evaluate the morphological appearance if representative regions are carefully selected with HE slides [25, 26]. In

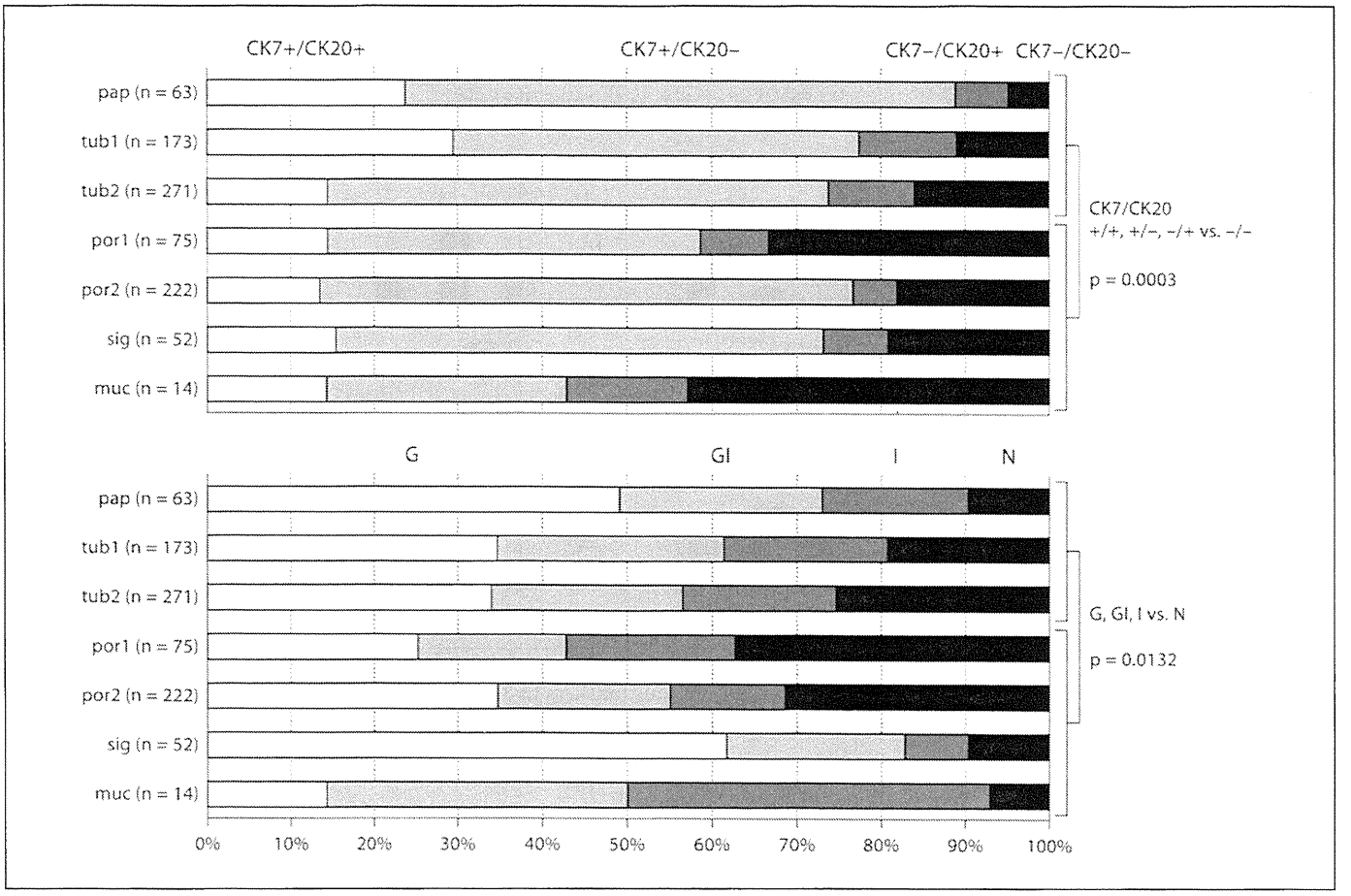


Fig. 2. The relationships between detailed histological type and CK expression patterns and mucin phenotypes in 870 GCs. The 870 GCs were histologically classified as 507 of the differentiated type (papillary adenocarcinoma or tubular adenocarcinoma) and 363 of the undifferentiated type (poorly differentiated adenocarcinoma, signet-ring cell carcinoma or mucinous adenocarcinoma). There are statistically significant associations between the undifferentiated type and neither CK7 nor CK20 expression, and the undifferentiated type and N mucin phenotype.

terms of the possible diversity of histological components or molecular abnormality in the GCs, several previous reports have shown an excellent concordance between the results obtained from TMAs and those from full sections [27, 28]. Furthermore, the effects of intratumoral heterogeneity can be averaged out in such a large-scale analysis as the present study. It is unlikely that the use of TMA biased the outcome.

To clarify the significance of the CK expression patterns, we analyzed the relation between CK expression patterns and the clinicopathologic parameters, histology, mucin phenotype or several tumor-related molecules. Histologically, expression of CK7 and/or CK20 showed a tendency toward a high positive rate in differentiated type GC and a low positive rate in undifferentiated type GC. This may reflect a loss of the ability to produce the CKs along with a decrease in histological differentiation in neoplastic cells. However, CK7 and/or CK20 expression was not associated with any other clinicopathologic features, consistent with the previous report [17, 29]. GCs have been classified into four mucin phenotypes. Previous reports provided evidence that mucin expression is closely associated with the differentiation of GCs [30, 31]. In the present study, G type was correlated with CK7 expression, especially in the absence of CK20 expression, whereas I type was correlated with CK20 expression, especially in the absence of CK7 expression. This result demonstrated statistical significance, but neither CK7 nor CK20 was sufficient for the discrimination of the mucin phenotype due to low sensitivity and specificity. In the present study,

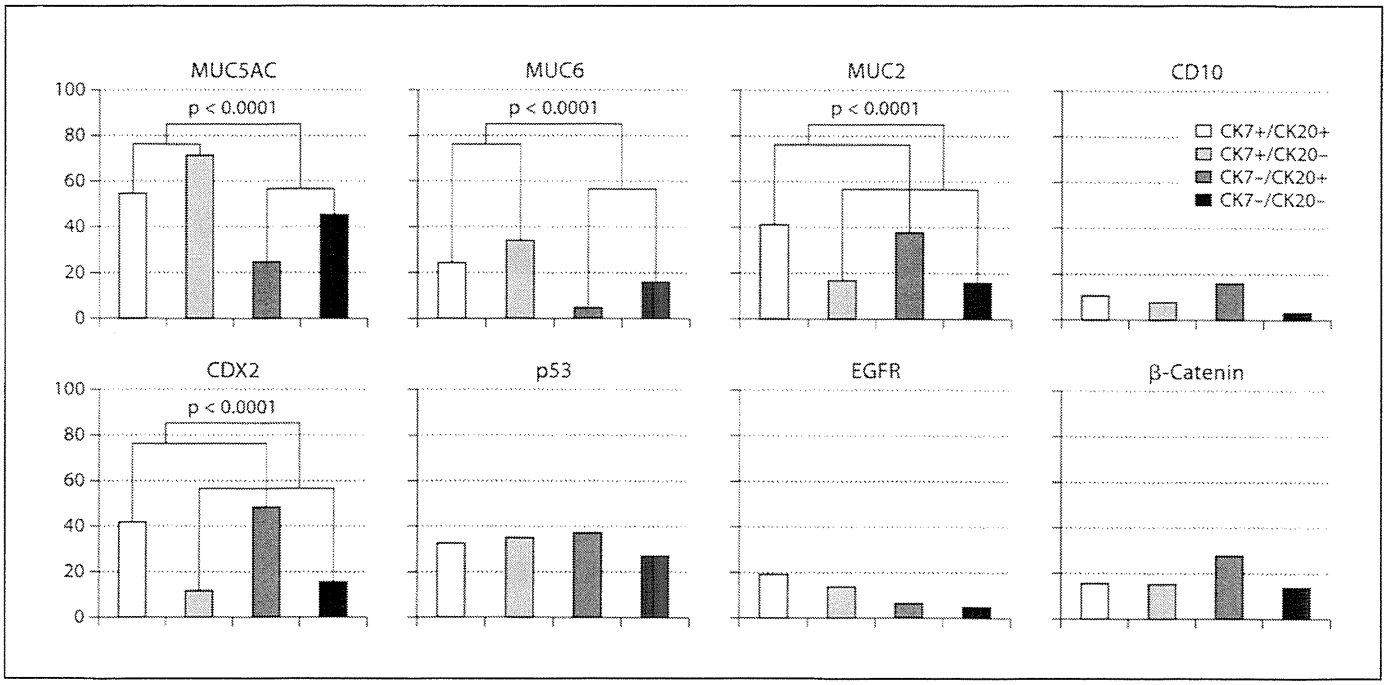


Fig. 3. The relationships between CK expression patterns and various markers in 870 GCs. There are statistically significant associations between CK7 and MUC5AC expression, CK7 and MUC6 expression, CK20 and MUC2 expression, and CK20 and CDX2 ($p < 0.0001$).

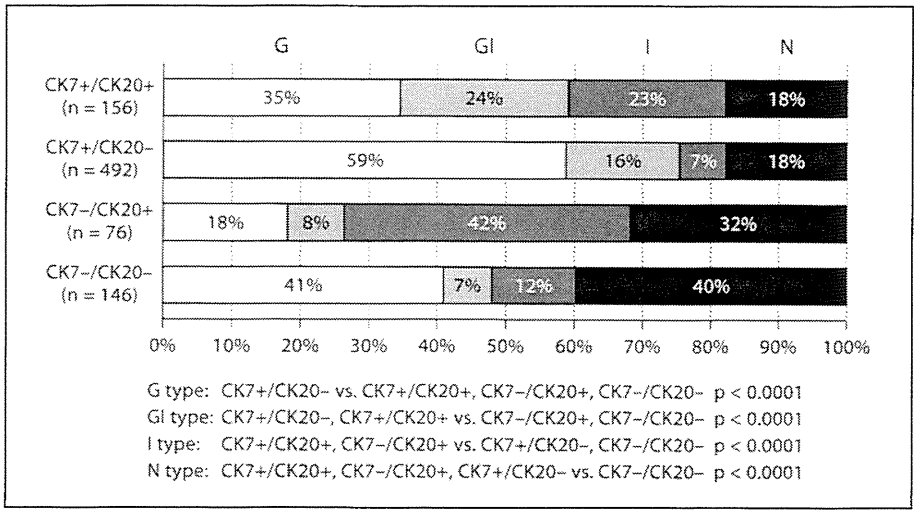


Fig. 4. The relationships between each CK expression pattern and mucin phenotypes in 870 GCs. There are statistically significant associations between CK7 expression without CK20 expression and G mucin phenotype, CK7 expression and GI mucin phenotype, CK20 expression without CK7 expression and I mucin phenotype, and neither of them and N mucin phenotype ($p < 0.0001$).

the positive expression of CK20 was frequently observed in GCs with the I type and showed significant correlation with the positive expression of Cdx2. There is no previous report showing a direct association between CK20 and Cdx2. Chan et al. [32] reported that CK20 is directly regulated by Cdx1. Therefore, there may also be a close correlation between CK20 and CDX2.

In summary, GCs showed heterogenous CK expression representing their histological features. Therefore, a single CK and its combination does not always provide diagnostic value in differentiating GCs. Advances in our understanding of the genetic and molecular bases of GC according to each CK expression, however, may lead to new therapy.

Acknowledgements

We thank Ms. Emiko Hisamoto and Mr. Shinichi Norimura for their excellent technical assistance and advice. This work was carried out with the kind cooperation of the Research Center for Molecular Medicine, Faculty of Medicine, Hiroshima University. We also thank the Analysis Center of Life Science, Hiroshima University, for the use of its facilities. This work was supported in part by Grants-in-Aid for Cancer Research from the Ministry of Education, Culture, Science, Sports and Technology of Japan and in part by a Grant-in-Aid for the Third Comprehensive 10-Year Strategy for Cancer Control and for Cancer Research from the Ministry of Health, Labour and Welfare of Japan.

References

1 Carneiro F, Sobrinho-Simoes M: Metastatic pattern of gastric carcinoma. Hum Pathol 1996;27:213–214.
2 Yasui W, Oue N, Kitadai Y, Nakayama H: Recent advances in molecular pathobiology of gastric carcinoma; in Kaminishi M, Takubo K, Mafune K (eds): The Diversity of Gastric Carcinoma: Pathogenesis, Diagnosis and Therapy. Tokyo, Springer, 2005, pp 51–71.
3 Tatematsu M, Tsukamoto T, Inada K: Stem cells and gastric cancer: role of gastric and intestinal mixed intestinal metaplasia. Cancer Sci 2003;94:135–141.
4 Yasui W, Sentani K, Motoshita J, Nakayama H: Molecular pathobiology of gastric cancer. Scand J Surg 2006;95:225–231.
5 Chu PG, Weiss LM: Keratin expression in human tissues and neoplasms. Histopathology 2002;40:403–439.
6 Chu P, Wu E, Weiss LM: Cytokeratin 7 and cytokeratin 20 expression in epithelial neoplasms: a survey of 435 cases. Mod Pathol 2000;13:962–972.
7 Moll R: Cytokeratins in the histological diagnosis of malignant tumors. Int J Biol Markers 1994;9:63–69.
8 Moll R, Lowe A, Laufer J, Franke WW: Cytokeratin 20 in human carcinomas. A new histodiagnostic marker detected by monoclonal antibodies. Am J Pathol 1992;140:427–447.
9 Schwerer MJ, Baczako K: Expression of cytokeratins typical for ductal and squamous differentiation in the human stomach: an immunohistochemical study of normal foveolar epithelium, *Helicobacter pylori* gastritis and intestinal metaplasia. Histopathology 1996;29:131–137.
10 Ramaekers F, van Niekerk C, Poels L, Schaafsma E, Huijsmans A, Robben H, Schaart G, Vooijs P: Use of monoclonal antibodies to keratin 7 in the differential diagnosis of adenocarcinomas. Am J Pathol 1990; 136:641–655.
11 Tot T: Cytokeratins 20 and 7 as biomarkers: usefulness in discriminating primary from metastatic adenocarcinoma. Eur J Cancer 2002;38:758–763.
12 Park SY, Kim HS, Hong EK, Kim WH: Expression of cytokeratins 7 and 20 in primary carcinomas of the stomach and colorectum and their value in the differential diagnosis of metastatic carcinomas to the ovary. Hum Pathol 2002;33:1078–1085.
13 Moll R: Cytokeratins as markers of differentiation in the diagnosis of epithelial tumors. Subcell Biochem 1998;31:205–262.
14 Wauters CC, Smedts F, Gerrits LG, Bosman FT, Ramaekers FC: Keratins 7 and 20 as diagnostic markers of carcinomas metastatic to the ovary. Hum Pathol 1995;26:852–855.
15 Kirchner T, Muller S, Hattori T, Mukaisyo K, Papadopoulos T, Brabletz T, Jung A: Metaplasia, intraepithelial neoplasia and early cancer of the stomach are related to dedifferentiated epithelial cells defined by cytokeratin-7 expression in gastritis. Virchows Arch 2001;439:512–522.
16 Ormsby AH, Goldblum JR, Rice TW, Richter JE, Gramlich TL: The utility of cytokeratin subsets in distinguishing Barrett's-related oesophageal adenocarcinoma from gastric adenocarcinoma. Histopathology 2001;38: 307–311.
17 Kim MA, Lee HS, Yang HK, Kim WH: Cytokeratin expression profile in gastric carcinomas. Hum Pathol 2004;35:576–581.
18 Shen B, Ormsby AH, Shen C, Dumot JA, Shao YW, Bevins CL, Gramlich TL: Cytokeratin expression patterns in noncardia, intestinal metaplasia-associated gastric adenocarcinoma: implication for the evaluation of intestinal metaplasia and tumors at the esophagogastric junction. Cancer 2002;94: 820–831.
19 Gurbuz Y, Kose N: Cytokeratin expression patterns of gastric carcinomas according to Lauren and Goseki classification. Appl Immunohistochem Mol Morphol 2006;14:303–308.
20 Lee MJ, Lee HS, Kim WH, Choi Y, Yang M: Expression of mucins and cytokeratins in primary carcinomas of the digestive system. Mod Pathol 2003;16:403–410.
21 Xue L, Zhang X, Li Y, Yang H, Li X, Mi J, Wang H, Wang J, Yan X: Differences of immunophenotypic markers and signaling molecules between adenocarcinomas of gastric cardia and distal stomach. Hum Pathol 2011;42:594–601.
22 Sobin LH, Wittekind CH (eds): TNM Classification Of Malignant Tumors, ed 6. New York, Wiley, 2002, pp 65–68.
23 Japanese Gastric Cancer Association: Japanese classification of gastric carcinoma, 3rd English edition. Gastric Cancer 2011;14:101–112.
24 Mizoshita T, Tsukamoto T, Nakanishi H, Inada K, Ogasawara N, Joh T, Itoh M, Yamamura Y, Tatematsu M: Expression of Cdx2 and the phenotype of advanced gastric cancers: relationship with prognosis. J Cancer Res Clin Oncol 2003;129:727–734.
25 Sauter G, Mirlacher M: Tissue microarrays for predictive molecular pathology. J Clin Pathol 2002;55:575–576.
26 Sentani K, Oue N, Tashiro T, Sakamoto N, Nishisaka T, Fukuhara T, Taniyama K, Matsuura H, Arihiro K, Ochiai A, Yasui W: Immunohistochemical staining of reg IV and claudin-18 is useful in the diagnosis of gastrointestinal signet ring cell carcinoma. Am J Surg Pathol 2008;32:1182–1189.
27 Hoos A, Urist MJ, Stojadinovic A, Mastorides S, Dudas ME, Leung DH, Kuo D, Brennan MF, Lewis JJ, Cordon-Cardo C: Validation of tissue microarrays for immunohistochemical profiling of cancer specimens using the example of human fibroblastic tumors. Am J Pathol 2001;158:1245–1151.
28 Lee HS, Cho SB, Lee HE, Kim MA, Kim JH, Park do J, Kim JH, Yang HK, Lee BL, Kim WH: Protein expression profiling and molecular classification of gastric cancer by the tissue array method. Clin Cancer Res 2007; 13:4154–4163.
29 Gulmann C, Counihan I, Grace A, Patchett S, Leen E, Leader M, Kay E: Cytokeratin 7/20 and mucin expression patterns in oesophageal, cardia and distal gastric adenocarcinomas. Histopathology 2003;43:453–461.
30 Pinto-de-Sousa J, David L, Reis CA, Gomes R, Silva L, Pimenta A: Mucins MUC1, MUC2, MUC5AC and MUC6 expression in the evaluation of differentiation and clinicobiological behaviour of gastric carcinoma. Virchows Arch 2002;440:304–310.
31 Pinto-de-Sousa J, Reis CA, David L, Pimenta A, Cardoso-de-Oliveira M: MUC5B expression in gastric carcinoma: relationship with clinico-pathological parameters and with expression of mucins MUC1, MUC2, MUC5AC and MUC6. Virchows Arch 2004;444: 224–230.
32 Chan CW, Wong NA, Liu Y, Bicknell D, Turley H, Hollins L, Miller CJ, Wilding JL, Bodmer WF: Gastrointestinal differentiation marker cytokeratin 20 is regulated by homeobox gene CDX1. Proc Natl Acad Sci USA 2009;106:1936–1941.

Pathology International

Pathology International 2012; **62**: 112–119 doi:10.1111/j.1440-1827.2011.02760.x

Original Article

Expression of cancer stem cell markers ALDH1, CD44 and CD133 in primary tumor and lymph node metastasis of gastric cancer

Yuta Wakamatsu, Naoya Sakamoto, Htoo Zarni Oo, Yutaka Naito, Naohiro Uraoka, Katsuhiro Anami, Kazuhiro Sentani, Naohide Oue and Wataru Yasui

Department of Molecular Pathology, Hiroshima University Graduate School of Biomedical Sciences, Hiroshima, Japan

Gastric cancer (GC) is one of the most common malignancies worldwide. Recently, cancer stem cells (CSCs) in tumors were found to possess the ability to sustain tumor self-renewal, initiate tumor progression, and possibly also contribute to cancer metastasis. We immunohistochemically examined expression and distribution of representative CSC markers ALDH1, CD44, and CD133 in primary tumors and lymph node metastasis of GC. Among 190 GC primary tumors, 104 (55%) were positive for ALDH1, 117 (62%) were positive for CD44, and 18 (9%) were positive for CD133. Expression of these three CSC markers was significantly associated with advanced clinicopathologic factors. Patients with CD44- and CD133-positive GC had a poorer survival rate than patients with CD44- and CD133-negative GC (CD44: $P < 0.001$, CD133: $P = 0.006$). Univariate and multivariate Cox proportional hazards analysis revealed tumor node metastasis stage, CD44 expression, and CD133 expression to be independent predictors of survival in patients with GC. Comparison of CSC markers in primary and metastatic sites showed ALDH1 positivity to be significantly higher in diffuse-type lymph node metastasis than in the primary tumor ($P < 0.001$). These results indicate that these CSC markers are important in tumor invasion and metastasis and may be good markers indicating long-term survival in patients with GC.

Key words: cancer stem cell, gastric cancer, metastasis, stem cell marker

Gastric cancer (GC) is one of the most common human cancers. According to the World Health Organization, GC is the fourth most common malignancy worldwide, with approximately 870 000 new cases occurring yearly.[1] Cancer develops as a result of multiple genetic and epigenetic alterations. Cancer at an advanced stage is a systemic disease and metastatic status significantly affects patient outcomes. In GC, lymph node metastasis is one of the most valuable prognostic factors. It is an important issue to examine the biological characteristics of cancer cells not only in primary tumors but also in metastatic tumors in the lymph node. However, the molecular pathological features of metastatic sites have not been sufficiently analyzed.

Recently, the cancer stem cell model suggests that in many cancers, tumor initiation and propagation is driven by a population of self-renewing tumor cells known as cancer stem cells (CSCs).[2] CSCs also promote tumor cell heterogeneity, metastasis, and therapeutic resistance, and are potentially driven by known oncogenic signaling pathways.[3,4] The study of CSCs would be greatly enhanced by the availability of specific markers to identify and isolate these cells. Through examinations using putative stem cell markers or side population (SP), unique subsets of cancer cells from different types of tumors have been detected. These markers include CD133, CD44, CD24, and CD166. Among them, both CD133 and CD44 are widely used for isolating CSCs from solid tumors. CD133 is a cell surface transmembrane glycoprotein, which exists in the cholesterol-rich domain of lipid rafts, and was identified in subpopulations of cells in brain and colon tumors.[5,6] Only one hundred CD133-positive cells implanted in a non-obese diabetic severe combined immunodeficient (NOD-SCID) mouse are sufficient to initiate a tumor, and isolated stem cells from this tumor can be serially passed to other NOD-SCID mice.[5,7] Expression of CD133 has been reported to be associated with poor prognosis in

Correspondence: Wataru Yasui, MD, PhD, Department of Molecular Pathology, Hiroshima University Graduate School of Biomedical Sciences, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8551, Japan. E-mail: wyasui@hiroshima-u.ac.jp

Received 25 July 2011. Accepted for publication 6 October 2011.

GC.[8,9] CD44 is also a transmembrane glycoprotein which participates in many cellular processes, including growth, survival, differentiation, and mortality,[10,11] and plays important roles in malignant behaviors of several human cancers including GC.[12–14] The CD44 marker can be used to isolate CSC populations of prostate,[15] pancreas,[16] and colorectal tumors.[17] Recent reports indicate that CD44-positive fractions of GC can generate spheroid colonies under non-adherent conditions and that small numbers of these cells can generate tumors in SCID mice.[18]

Beside these markers, a promising new marker for CSC is aldehyde dehydrogenase 1 (ALDH1).[19,20] Aldehyde dehydrogenase enzymes are a family of intracellular enzymes that participate in cellular detoxification, differentiation, and drug resistance through the oxidation of cellular aldehydes.[21] ALDH1 positive cells monitored by immunohistochemistry and flow cytometry occupy a considerably smaller subpopulation that is about one-seventh of the size of the CD44- and CD133-positive populations. Nevertheless, ALDH1-positive cell populations are capable of generating tumor xenografts.[20] Therefore, ALDH1 may be able to label a cell population closely related to stem cells. Moreover, high percentages of ALDH1-positive cells in most types of epithelial tumors are associated with poorer clinical outcomes for these patients.[19,22,23] However, expression of ALDH1 in GC has not been analyzed.

In this present study, we examined the expression and distribution of the representative CSC markers ALDH1, CD44 and CD133 in GC by immunohistochemistry, and studied their relationship with clinicopathologic features. Furthermore, the expression of CSC markers was compared between GC primary tumors and metastatic lymph nodes.

MATERIALS AND METHODS

Tissue samples

In total, 190 primary tumor samples were collected from patients diagnosed with GC. Patients were treated at the Hiroshima University Hospital or affiliated hospitals. Of the 190 GC samples, associated lymph node metastasis samples were available for 104 cases. Information on patient prognosis was available for 96 of the 190 GC cases. Because written informed consent was not obtained, for strict privacy protection, identifying information for all samples was removed before analysis. This procedure was in accordance with the Ethical Guidelines for Human Genome/Gene Research of the Japanese Government.

For immunohistochemical analysis, we used formalin-fixed, paraffin-embedded tissues from 190 patients who had undergone surgical excision for GC. One or two representative tumor blocks were examined from each patient by immunohistochemistry. Tumor staging was made according to the TNM classification system. Histological classification of GC was carried out according to the Lauren classification system; GC cases were classified into intestinal-, diffuse-, and mixed-type.

Immunohistochemical examination

Immunohistochemical staining was carried out according to the procedure previously reported.[24] Deparaffinized sections were deparaffinized in xylene, dehydrated in a graded ethanol series, and immersed in a 0.3% hydrogen peroxide solution in methanol for 10 min to inhibit endogenous peroxidase activity. The sections were placed in ethylenediaminetetraacetic acid buffer at pH 8.0 for CD133, or citrate buffer (pH 6.0) for ALDH1 and CD44 stainings. For antigen retrieval, the slides were heated at 95°C for 20 min in a microwave oven for CD133, or 95°C for 30 min in a microwave oven for ALDH1 and CD44 staining, and allowed to cool for 5 min at room temperature. Sections were incubated with the following antibody dilutions; primary anti-CD133 antibody (AC133; Miltenyi Biotec, Auburn, CA, USA) 1:100, anti-ALDH1 antibody (BD Biosciences; San Diego, CA, USA) 1:200, and anti-CD44 antibody (Novocastra; Newcastle, UK). Sections were incubated with primary antibody for 1 h at room temperature. The slides were washed three times with PBS, followed by incubations with Envision+ anti-mouse peroxidase for 1 h. For color reactions, sections were incubated with the DAB Substrate-Chromogen Solution (Dako Cytomation; Carpinteria, CA, USA) for 10 min. Sections were then counterstained with 0.1% hematoxylin. Previous reports demonstrated that significant correlation between poor clinical outcome and immunostaining with CD133 in colorectal cancer,[25] CD44 in squamous cell carcinoma of lung,[26] and ALDH1 in non-small cell carcinoma of lung.[27] According to the definition of positive staining in these previous reports, the result was considered positive if at least 10% of the cells were stained. When fewer than 10% of cancer cells were stained, the immunostaining was considered negative.

Statistical analysis

ALDH1, CD44, and CD133 expression and clinicopathologic features were tested for association by the χ^2 test. For each molecule, Kaplan–Meier survival curves were constructed to compare positive and negative patients. Differences between survival curves were tested for statistical significance by log-rank test. Univariate and multivariate Cox regression was used to evaluate the associations between clinical covariates and cancer-specific mortality in SPSS (SPSS Inc., Chicago, IL, USA). Hazard ratio (HR) and 95% confidence interval (CI)