A Functional Single Nucleotide Polymorphism in *Mucin 1*, at Chromosome 1q22, Determines Susceptibility to Diffuse-Type Gastric Cancer

NORIHISA SAEKI,[*] AKIRA SAITO,[*,‡] IL JU CHOI,[§] KEITARO MATSUO,[‖] SUMIKO OHNAMI,[*] HIROHIKO TOTSUKA,[*,¶] SUENORI CHIKU,[*,#] AYA KUCHIBA,[*] YEON–SU LEE,[§] KYONG–AH YOON,[§] MYEONG–CHERL KOOK,[§] SOOK RYUN PARK,[§] YOUNG–WOO KIM,[§] HIDEO TANAKA,[‖] KAZUO TAJIMA,[‖] HIROSHI HIROSE,[**] FUMIHIKO TANIOKA,[‡‡] YOSHIHIRO MATSUNO,[§§] HARUHIKO SUGIMURA,[‖‖] SHUNJI KATO,[¶¶] TSUNEYA NAKAMURA,[##] TOMOHIRO NISHINA,[***] WATARU YASUI,[‡‡‡] KAZUHIKO AOYAGI,[*] HIROKI SASAKI,[*] KAZUYOSHI YANAGIHARA,[§§§] HITOSHI KATAI,[‖‖‖] TADAKAZU SHIMODA,[¶¶¶] TERUHIKO YOSHIDA,[*] YUSUKE NAKAMURA,[###,****] SETSUO HIROHASHI,[****] and HIROMI SAKAMOTO[*]

[*]Genetics Division, National Cancer Center Research Institute, Tokyo; [‡]Statistical Genetics Analysis Division, StaGen Co, Ltd, Tokyo, Japan; [§]Research Institute and Hospital, National Cancer Center, Gyeonggi-do, Korea; [‖]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya; [¶]Bioinfomatics Group, Research and Development Center, Solution Division 4, Hitachi Government and Public Corporation System Engineering Ltd, Tokyo; [#]Science Solutions Division, Mizuho Information and Research Institute, Inc, Tokyo; [**]Department of Internal Medicine, Keio University School of Medicine, Tokyo; [‡‡]Iwata City Hospital, Shizuoka; [§§]Department of Surgical Pathology, Hokkaido University Hospital, Sapporo; [‖‖]First Department of Pathology, Hamamatsu University School of Medicine, Shizuoka; [¶¶]Department of Surgery, Nippon Medical School Hospital, Tokyo; [##]Department of Endoscopy, Aichi Cancer Center Hospital, Nagoya; [***]Department of Internal Medicine, Shikoku Cancer Center, Ehime; [‡‡‡]Department of Molecular Pathology, Hiroshima University Graduate School of Biomedical Sciences, Hiroshima; [§§§]Department of Life Sciences, Yasuda Women's University Faculty of Pharmacy, Hiroshima; [‖‖‖]Department of Surgical Oncology, National Cancer Center Hospital, Tokyo; [¶¶¶]Center for Cancer Control and Information Services, National Cancer Center, Tokyo; [###]Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo; and [****]National Cancer Center Research Institute, Tokyo, Japan

BACKGROUND & AIMS: Two major types of gastric cancer, intestinal and diffuse, develop through distinct mechanisms; the diffuse type is considered to be more influenced by genetic factors, although the mechanism is unknown. Our previous genome-wide association study associated 3 single nucleotide polymorphisms (SNPs) with diffuse-type gastric cancer (DGC); 1 was a functional SNP (rs2294008) in *prostate stem cell antigen* (*PSCA*), but the loci of the other 2 were not investigated. **METHODS:** We performed high-density mapping to explore a linkage disequilibrium status of the 2 SNPs at chromosome 1q22. A DGC case-control study was conducted using DNA from 606 cases and 1264 controls (all Japanese individuals) and validated using DNA from Japanese (304 cases, 1465 controls) and Korean (452 cases, 372 controls) individuals. The effects of SNPs on function were analyzed by reporter assays and analyses of splice variants. **RESULTS:** A region of a strong linkage disequilibrium with the 2 SNPs contained *mucin 1* (*MUC1*) and other 4 genes and SNPs significantly associated with DGC (rs2070803: $P = 4.33 \times 10^{-13}$; odds ratio [OR], 1.71 by meta-analysis of the studies on the 3 panels) but not with intestinal-type gastric cancer. Functional studies demonstrated that rs4072037 ($P = 1.43 \times 10^{-11}$; OR, 1.66 by meta-analysis) in *MUC1* affects promoter activity and determines the major splicing variants of *MUC1* in the gastric epithelium. Individuals that carry both SNPs rs2294008 in *PSCA* and rs4072037 in *MUC1* have a high risk for developing DGC (OR, 8.38). **CONCLUSIONS: *MUC1* is the second major DGC susceptibility gene identified. The SNPs rs2070803 and rs4072037 in *MUC1* might be used to identify individuals at risk for this type of gastric cancer.**

Keywords: Stomach Cancer; Risk Genotype; Cancer Prevention; Genome-Wide Association Study.

BASIC–ALIMENTARY TRACT

Gastric cancer (GC) is the fourth most common cancer and the second most common cause of cancer death in the world.[1] More than 90% of GC are adenocarcinomas, which are classified into diffuse-type GC (DGC) and intestinal-type GC (IGC).[2] Typically, IGC arises through a sequence of pathologic changes of the gastric epithelium: chronic gastritis mainly because of *Helicobacter pylori* infection, atrophic gastritis, intestinal metaplasia, dysplasia, and adenocarcinoma.[3] On the other hand, the origin of DGC has been considered to be gastric epithelial stem cells and/or precursors present in the isthmus region of the middle portion of the epithelium (Supplementary Figure 1). Genetic and epigenetic events acting on the stem/precursor cells may cause a deviation from their normal differentiation program and lead to a DGC development,[4] although details are yet to be revealed. In contrast to the steady decline of the incidence of IGC, mainly because of the decreasing prevalence of *H pylori* infection, DGC appears to be increasing.[5] Moreover, some DGC develops to a highly malignant form,

Abbreviations used in this paper: DGC, diffuse-type gastric cancer; GC, gastric cancer; GWAS, genome-wide association study; IGC, intestinal-type gastric cancer; kb, kilobase; LD, linkage disequilibrium; *MUC1*, mucin 1; OR, odds ratio; por, poorly differentiated adenocarcinoma; *PSCA*, *prostate stem cell antigen*; sig, signet-ring cell carcinoma; TR, tandem repeats.

0016-5085/$36.00
doi:10.1053/j.gastro.2010.10.058

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,