A genome-wide association study identifies two susceptibility loci for duodenal ulcer in the Japanese population

Article *in* Nature Genetics · March 2012

DOI: 10.1038/ng.1109 · Source: PubMed

10 authors, including:

Hidemi Ito
Aichi Cancer Center
242 PUBLICATIONS **4,023** CITATIONS

SEE PROFILE

Yusuke Nakamura
Tokai University
1,622 PUBLICATIONS **76,599** CITATIONS

SEE PROFILE

Koichi Matsuda
The University of Tokyo
196 PUBLICATIONS **6,187** CITATIONS

SEE PROFILE

nature genetics

A genome-wide association study identifies two susceptibility loci for duodenal ulcer in the Japanese population

Chizu Tanikawa[1], Yuji Urabe[1], Keitaro Matsuo[2], Michiaki Kubo[3], Atsushi Takahashi[3], Hidemi Ito[2], Kazuo Tajima[2], Naoyuki Kamatani[3], Yusuke Nakamura[1] & Koichi Matsuda[1]

Through a genome-wide association analysis with a total of 7,035 individuals with duodenal ulcer and 25,323 controls from Japan, we identified two susceptibility loci at the *PSCA* gene (encoding prostate stem cell antigen) at 8q24 and at the *ABO* blood group locus at 9q34. The C allele of rs2294008 at *PSCA* was associated with increased risk of duodenal ulcer (odds ratio (OR) = 1.84; $P = 3.92 \times 10^{-33}$) in a recessive model but was associated with decreased risk of gastric cancer (OR = 0.79; $P = 6.79 \times 10^{-12}$), as reported previously[1]. The T allele of rs2294008 encodes a translation initiation codon upstream of the reported site and changes protein localization from the cytoplasm to the cell surface. rs505922 at *ABO* was also associated with duodenal ulcer in a recessive model (OR = 1.32; $P = 1.15 \times 10^{-10}$). Our findings demonstrate a role for genetic variants in the pathogenesis of duodenal ulcer.

Duodenal ulcer is one of the most common gastrointestinal disorders, with a lifetime prevalence of 4–15% (refs. 2,3). *Helicobacter pylori* infection is a major cause of duodenal ulcer and is observed in 70–90% of individuals with this condition[2] as well as in individuals with gastric ulcer and cancer[3]. Eradication of *H. pylori* by antibiotics can effectively cure duodenal ulcer[4], showing the causal role of *H. pylori* in disease pathogenesis. Because of the high prevalence of *H. pylori* infection in the Japanese population, the incidence of peptic ulcer and gastric cancer is much higher in Japanese individuals than in individuals of European descent[5,6]. Among Europeans and non-Japanese Asian populations, duodenal ulcer is more common than gastric ulcer[5,7], while gastric ulcer is more common among Japanese and Japanese-Americans[6,8]. In addition, individuals with duodenal ulcer are well known to have a lower risk for gastric cancer[9]. These heterogeneities in disease susceptibility are influenced by both bacterial and host factors[10]. Smoking and nonsteroidal anti-inflammatory drugs have been shown to increase the risk of peptic ulcer[11]. In addition, in a Finnish twin study in which the concordance rate between probands was 23.6% in monozygotic twins and 14.8% in dizygotic twins, 39% of the liability to peptic ulcer was explained by genetic factors[12]. To identify genetic susceptibility factors for duodenal ulcer, we performed a genome-wide association study (GWAS).

We genotyped DNA samples from 1,043 individuals with duodenal ulcer (cases) and 21,694 controls without peptic ulcer from BioBank Japan[13]. (The characteristics of each cohort and the study design are shown in **Table 1** and in **Supplementary Fig. 1**.) After performing a standard quality control procedure, we had genotyping results for 480,327 SNPs and performed logistic regression analysis by including age and gender as covariates. In the GWAS stage, we calculated the minimum *P* value under three genetic models (additive, recessive and dominant). The genomic inflation factor λ was calculated to be 1.014, indicating minimal evidence of population stratification (**Supplementary Fig. 2**). The association analyses identified two significantly associated loci on chromosomes 8q24 and 9q34 ($P = 2.84 \times 10^{-19}$ and 2.27×10^{-8}, respectively; **Fig. 1**, **Table 2** and **Supplementary Table 1**). Overall, a total of 101 SNPs from 42 distinct genomic regions showed suggestive evidence of association ($P < 5 \times 10^{-5}$).

In the replication stage, we selected the 42 SNPs that had the lowest *P* values in each genomic region and genotyped them in 5,992 independent duodenal ulcer cases and 3,629 controls. We performed logistic regression analysis, adjusting for age and gender, and observed significant associations at rs2294008 on 8q24 (OR = 1.73 and $P = 6.60 \times 10^{-16}$ in a recessive model) and at rs505922 on 9q34 (OR = 1.23 and $P = 1.78 \times 10^{-4}$ in a recessive model) after applying a Bonferroni correction ($P < 0.05/42 = 1.19 \times 10^{-3}$; **Table 2** and **Supplementary Table 2**). A meta-analysis of the GWAS and replication stages identified significant associations for rs2294008 and rs505922 without any heterogeneity between the two stages (OR = 1.84; $P = 3.92 \times 10^{-33}$ and OR = 1.32; $P = 1.15 \times 10^{-10}$, respectively; **Table 2**). Under an additive model, rs2294008 showed equivalent association with duodenal ulcer ($P = 1.79 \times 10^{-33}$) as in the recessive model. Logistic regression analysis of various risk factors showed that the two SNPs, age, gender and smoking are independent risk factors for duodenal ulcer (**Supplementary Table 3**).

The development of duodenal ulcer occurs as a multistep process, involving persistent *H. pylori* infection and subsequent inflammation and damage of the duodenal mucosa. To elucidate the physiological

[1]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [2]Divison of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Aichi, Japan. [3]Center for Genomic Medicine, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan. Correspondence should be addressed to K. Matsuda (koichima@ims.u-tokyo.ac.jp).

Received 12 September 2011; accepted 20 January 2012; published online 4 March 2012; doi:10.1038/ng.1109

Table 1 Characteristics of study populations

Stage	Sample type	*H. pylori* status	Source	Platform	Number of samples	Female (%)	Age (mean ± s.d.)
GWAS	Duodenal ulcer		BioBank Japan	HumanHap610K	1,043	247 (23.7%)	63.2 ± 10.6
	Controls		BioBank Japan	HumanHap610K	21,694	10,162 (46.8%)	62.6 ± 12.5
Replication	Duodenal ulcer		BioBank Japan	Invader assay	5,992	1,639 (27.4%)	63.0 ± 12.1
	Controls		BioBank Japan	HumanHap550K	2,722	2,273 (62.6%)	46.3 ± 15.7
			Healthy controls		907		
Gastric cancer	Gastric cancer		BioBank Japan	HumanHap610K	2,346	521 (22.2%)	64.9 ± 9.1
	Controls		BioBank Japan	HumanHap610K	16,882	8,663 (51.3%)	62.6 ± 13.0
H. pylori susceptibility[a]	Duodenal ulcer	+	Aichi Cancer Center	TaqMan assay	37	8 (21.6%)	53.4 ± 13.3
	Healthy controls	+	Aichi Cancer Center	TaqMan assay	284	144 (50.7%)	54.4 ± 14.3
		–	Aichi Cancer Center	TaqMan assay	509	268 (52.7%)	42.0 ± 15.6

[a]Positivity for *H. pylori* infection was defined by plasma levels of immunoglobulin G (IgG) to *H. pylori* that were greater than 10 U/ml.

roles of the variants, we genotyped the two associated SNPs in additional cohorts consisting of healthy controls, with or without *H. pylori* infection, as well as duodenal ulcer cases with *H. pylori* infection (**Table 1**). Neither SNP showed significant association with *H. pylori* infection in healthy controls (**Supplementary Table 4**). However, we found significant association of rs2294008 ($P = 0.021$; OR = 2.59) and marginal association of rs505922 ($P = 0.076$; OR = 1.90) with duodenal ulcer susceptibility among *H. pylori* carriers (**Supplementary Table 5**). In addition, both SNPs showed significant association with duodenal ulcer risk in individuals with gastric cancer ($P = 4.7 \times 10^{-4}$ and 0.041; OR = 2.88 and 1.62, respectively; **Supplementary Table 5**), of whom nearly 80% were infected with *H. pylori*. Taken together, the data indicate that these SNPs are likely to be associated with duodenal ulcer development after *H. pylori* infection and not with susceptibility to persistent *H. pylori* infection *per se*.

In both the GWAS and replication studies, we used individuals with various other diseases as controls. To exclude the possibility of confounding effects caused by the varied backgrounds of the control samples, we conducted an association analysis using only healthy volunteers as controls. In this analysis, we found strong association of rs2294008 ($P = 2.97 \times 10^{-15}$; OR = 1.75) and rs505922 ($P = 0.006$; OR = 1.17) with duodenal ulcer (**Supplementary Table 6**). In addition, the allelic frequencies of rs2294008 and rs505922 were not significantly different between controls with other disease and healthy controls ($P = 0.19$ and 0.46, respectively; **Supplementary Table 7**). Therefore, the use of controls with other diseases is not likely to have affected the association results in our study.

rs2294008 is located within a genomic region that encodes *ARC*, *JRK*, *PSCA*, *LY6K*, *C8orf55*, *SLURP1*, *LYPD2*, *LYNX1* and *LY6D*, while rs505922 is in a region encoding *OBP2B*, *ABO*, *SURF6*, *MED22*, *SNORD24*, *RFL7A*, *SNORD36B*, *SNORD36A*, *SNORD36C*, *SURF1* and *SURF2*. To further characterize these two loci, we performed imputation analysis (**Fig. 2a**,**b**). The regional association plots using genotyped and imputed data show that all strongly associated SNPs are confined to regions around the *PSCA* (encoding prostate stem cell antigen) and *ABO* blood group genes (**Supplementary Fig. 3**). Next, we examined the expression of these genes in multiple human tissues. *ABO* was expressed in the gastrointestinal tract, including in the duodenum, concordant with previous reports[14], whereas *PSCA* was highly expressed in stomach but not in normal duodenum (**Supplementary Fig. 4**). However, as metaplasia consisting of gastric-type mucous-secreting cells has been observed in duodenal ulcer lesions[15], *PSCA* is likely to be expressed in the duodenum in such cells.

PSCA encodes a glycosylphosphatidylinositol (GPI)-anchored membrane glycoprotein that is involved in cell renewal and proliferation[16]. PSCA was shown to be upregulated in various cancers, including bladder, pancreatic and kidney[17], and PSCA expression has been correlated with higher tumor grade and metastatic properties of prostate cancer[16]. However, PSCA downregulation and growth-suppressive effects in esophageal and gastric cancers have also been reported[17]. Hence, *PSCA* might function as an oncogene in some epithelial cells and as a tumor suppressor gene in others. Although previous studies identified an association between rs2294008 and risk for gastric and bladder cancers[1,18], the only known functional consequence of this variation is reduced transcriptional activity. The C allele of rs2294008 is common in European and African populations, whereas the T allele is dominant in Japanese. The T allele of rs2294008 encodes a translation initiation codon for the *PSCA* gene upstream of the known site, resulting in a PSCA protein with an additional nine amino acids at its N-terminus (long PSCA, 123 amino acids; **Fig. 3a**) relative to the reported PSCA protein (short PSCA, 114 amino acids). According to PSORT II (ref. 19), long PSCA contains an N-terminal signal peptide, which is predicted to be N-glycosylated, with the protein localizing to the plasma membrane, whereas short PSCA is predicted to not be glycosylated and to localize to the cytoplasm. As expected, immunocytochemical analysis showed membrane localization of the long PSCA protein in PC3 and Du145 cells (prostate cancer cell lines that have two T alleles at rs2294008) and cytosolic localization of the short PSCA protein in C42B and NCI-H522 cells (prostate and lung cancer cell lines that have two C alleles at rs2294008) (**Fig. 3b**).

We additionally constructed plasmids expressing short or long PSCA protein (pcDNA3.1/S-PSCA and pcDNA3.1/L-PSCA, respectively; **Supplementary Fig. 5**). HEK293T or MKN1 cells transfected

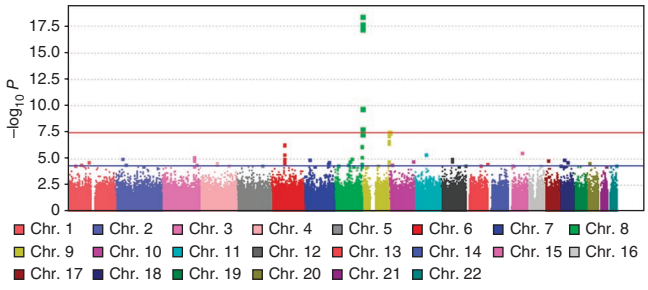

Figure 1 Manhattan plot showing genome-wide *P* values of association. The minimum *P* values under three genetic models (additive, recessive and dominant) were obtained by logistic regression analysis with adjustment for age and gender. The *y* axis shows the $-\log_{10} P$ values of 480,327 SNPs, and the *x* axis shows their chromosomal positions. Horizontal red and blue lines represent the thresholds of $P = 3.47 \times 10^{-8}$ for Bonferroni significance and $P = 5 \times 10^{-5}$ for selecting SNPs for replication, respectively.

Table 2 Results of association analyses of duodenal ulcer

SNP	Chr.	Stage	Allele 1/2[a]	Cases 11	Cases 12	Cases 22	Cases RAF	Controls 11	Controls 12	Controls 22	Controls RAF	Additive[b] P	Additive[b] OR	Dominant[b] P	Dominant[b] OR	Recessive[b] P	Recessive[b] OR	P^{c}_{het}
rs2294008	8q24	GWAS	C/T	252	473	318	0.468	2,984	9,947	8,758	0.367	2.92×10^{-18}	1.49	6.60×10^{-9}	1.50	2.84×10^{-19}	1.99	0.163
		Replication		1,387	2,735	1,866	0.460	503	1,665	1,456	0.369	2.11×10^{-17}	1.36	9.99×10^{-10}	1.38	6.60×10^{-16}	1.73	
		Combined[d]		1,639	3,208	2,184	0.461	3,487	11,612	10,214	0.367	1.79×10^{-33}	1.41	5.81×10^{-17}	1.42	3.92×10^{-33}	1.84	
rs505922	9q34	GWAS	T/C	389	451	203	0.589	6,425	10,758	4,504	0.544	1.60×10^{-5}	1.22	0.24	1.10	2.27×10^{-8}	1.45	0.053
		Replication		1,989	2,805	1,194	0.566	1,055	1,817	757	0.541	1.83×10^{-3}	1.12	0.23	1.08	1.78×10^{-4}	1.23	
		Combined[d]		2,378	3,256	1,397	0.570	7,480	12,575	5,261	0.544	3.43×10^{-7}	1.15	9.55×10^{-2}	1.09	1.15×10^{-10}	1.32	

We analyzed 7,035 duodenal ulcer cases (1,043 in the GWAS and 5,992 in replication) and 25,323 controls (21,694 in the GWAS and 3,629 in replication). Chr., chromosome; RAF, risk allele frequency.

[a]Allele 1, risk allele; allele 2, non-risk allele. [b]*P* values and ORs were calculated by logistic regression analysis, with age and gender as covariates. Non-risk alleles were considered as references in the three genetic models: additive, 1 versus 2; recessive, 11 versus 12 + 22; dominant, 11 + 12 versus 22. [c]Heterogeneity across the two stages was examined by Cochran *Q* test under a genetic model which provided the minimum *P* value in the screening stage. [d]ORs and *P* values were calculated using the Mantel-Haenszel fixed-effects model.

with either of these plasmids were stained with antibody to PSCA, with or without membrane permeabilization. We found that PSCA protein was localized to the plasma membrane only in cells transfected with pcDNA3.1/L-PSCA (**Fig. 3c** and **Supplementary Fig. 6a**). In addition, in protein blots, we observed that the band corresponding to long PSCA was actually 18 kDa rather than the predicted 13 kDa, whereas short PSCA was the predicted size. As the 18-kDa protein band for long PSCA was shifted to approximately 13 kDa by treatment with N-glycosidase (**Fig. 3d**), the modification accounting for its increased size was considered to be N-glycosylation. In addition, short PSCA was degraded through the ubiquitin proteasome pathway, which was inhibited by MG-132 treatment, whereas the long PSCA protein was relatively stable (**Fig. 3e** and **Supplementary Fig. 6b**,**c**).

These findings show that the genetic variation in *PSCA* could have a considerable effect on the biological function of the PSCA protein by altering its subcellular localization and stability.

Of note, when we analyzed rs2294008 and rs505922 in 2,346 individuals with gastric cancer and 16,882 controls (**Table 1**), we found that rs2294008 had opposing effects on gastric cancer and duodenal ulcer risk. Whereas the C allele of rs2294008 increased the risk of duodenal ulcer (OR = 1.84) in a recessive model, it showed a protective effect for gastric cancer, as reported previously[1] ($P = 6.79 \times 10^{-12}$ and OR = 0.79 in an additive model; **Supplementary Table 8**). We also estimated the population attributable risk (PAR) of rs2294008 to be as high as 23.0% for duodenal ulcer (C allele) and 39.2% for gastric cancer (T allele) in the Japanese population.

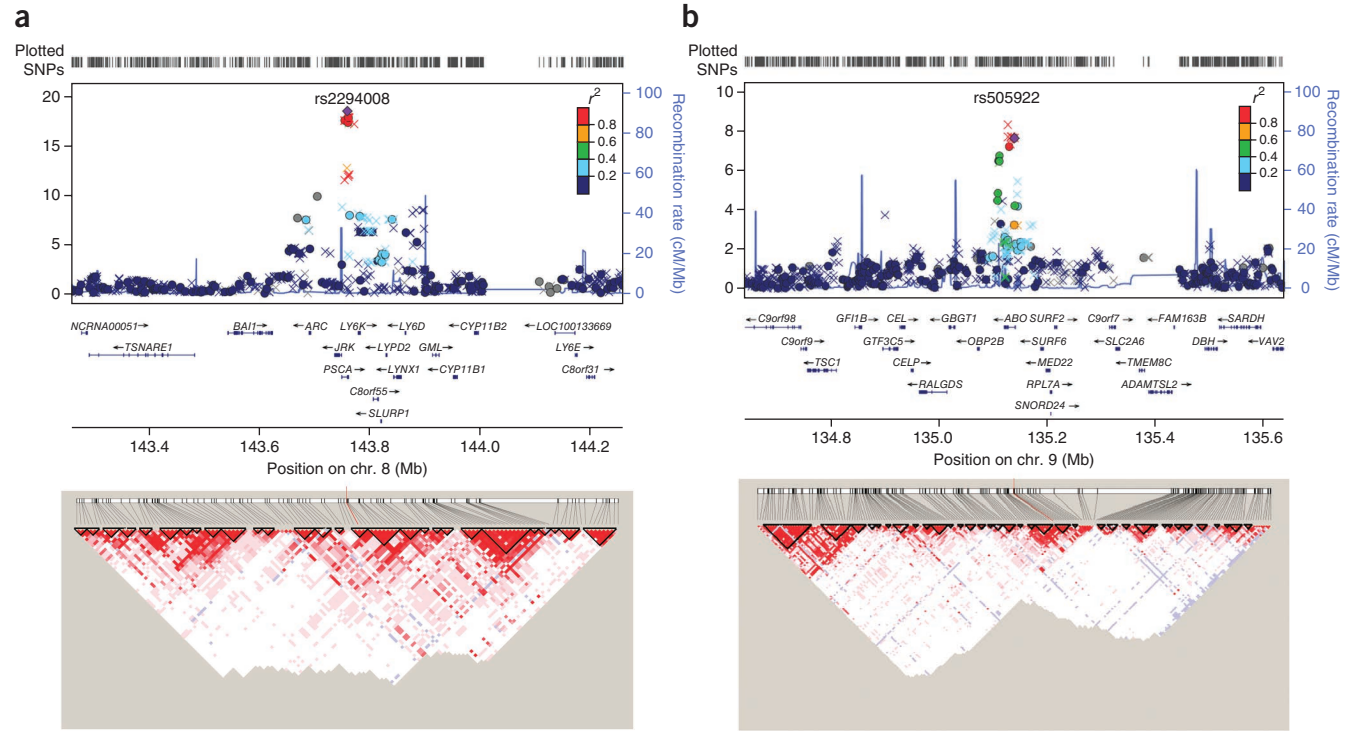

Figure 2 Regional association plots. (**a**,**b**) Data are shown for the associated regions on chromosome 8 including rs2294008 (in *PSCA*) (**a**) and on chromosome 9 including rs505922 (in *ABO*) (**b**). Upper, *P* values of genotyped SNPs (circle) and imputed SNPs (crosses) are plotted (as $-\log_{10} P$) against their physical position on the chromosomes (NCBI Build 36). Through imputation analyses, we obtained genotypes for 527 and 580 SNPs in addition to 155 and 219 genotyped SNPs within the 1-Mb genomic regions surrounding the marker SNPs on chromosome 8 and 9, respectively. Estimated recombination rates from the HapMap Japanese in Tokyo (JPT) population show the local LD structure. The color of each SNP indicates LD with rs2294008 or rs505922 based on pairwise r^2 values from HapMap JPT data. Middle, gene annotations from the UCSC genome browser. Lower, LD map based on *D'* (coefficient of linkage disequilibrium) in the associated regions using genotyping results from 907 healthy control samples from the replication study.

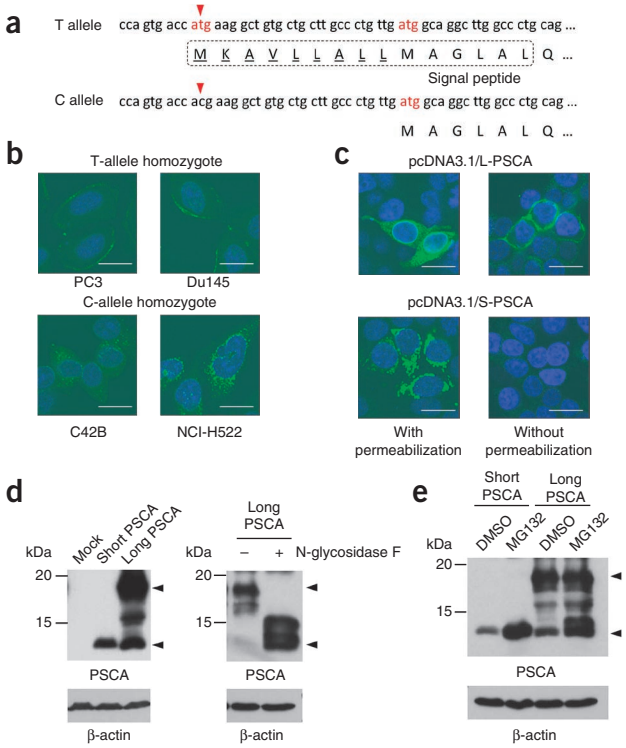

Figure 3 Effects of rs2294008 on subcellular localization and stability of PSCA protein. (**a**) The genomic structure around rs2294008 and corresponding amino-acid sequences. Arrowheads indicate the location of rs2294008. Potential start codons are shown in red. Underlined amino acids are unique to the T allele of rs2294008. (**b**) Representative images of cells stained with antibody to PSCA. PC3 and Du145 cells are homozygous for the T allele, whereas C42B and NCI-H522 cells are homozygous for the C allele of rs2294008. Scale bars, 20 μm. (**c**) HEK293T cells were transiently transfected with the indicated plasmid. Subcellular localization of PSCA protein was evaluated with antibody to PSCA, either with or without membrane permeabilization. Scale bars, 20 μm. (**d**) Left, expression of PSCA protein in HEK293T cells after transfection with the indicated expression plasmids encoding PSCA or with empty vector (mock). β-actin was used to normalize expression levels. Right, lysate from HEK293T cells transfected with pcDNA3.1/L-PSCA treated with N-glycosidase F. (**e**) After transfection with pcDNA3.1/L-PSCA or pcDNA3.1/S-PSCA vector, HEK293T cells were incubated with 10 μM of MG132 for 10 h before harvesting. DMSO was used as control. β-actin was used to normalize expression levels.

In contrast, rs505922 showed no association with gastric cancer in our analysis, which is discordant with previous reports[20,21].

On the basis of our findings, we propose that susceptibility to duodenal ulcer and gastric cancer is influenced by genetic variation in *PSCA* through a growth-promoting effect of the T allele and an effect on T-cell activation by the C allele (**Supplementary Fig. 7**). We hypothesize that, in response to damage in mucosal cells in the duodenum, the tissue repair system is switched on by the aggregation of platelets and the release of growth factors, which is followed by the proliferation and migration of epithelial cells. PSCA is highly expressed in various cancer tissues, and cells treated with small interfering RNA (siRNA) targeting *PSCA* or with antibody against PSCA exhibited a substantially suppressed growth[22,23], indicating a role for cell surface PSCA in cell proliferation. Our findings suggest that individuals with homozygous C alleles might have insufficient epithelial proliferation to counteract the damage because of a lack of functional cell surface PSCA, causing slow recovery from duodenal tissue damage.

A possible alternative mechanism involves antigen presentation. PSCA-derived peptides were reported to be a target of T-cell–based immunotherapy for advanced prostate cancer[24]. Our findings suggest that cytosolic short PSCA protein is likely to be more susceptible to proteasomal degradation than the long PSCA protein at the cell surface. Peptides presented by the human leukocyte antigen (HLA) molecules induce activation of $CD4^+$ and/or $CD8^+$ T cells, which were shown to be involved in peptic ulcer[25] and which also inhibit tumor formation. Therefore, individuals with a C allele might have higher risk for duodenal ulcer and lower risk for gastric cancer as a result of accelerated proteasomal degradation of PSCA protein and subsequent activation of immune responses. Thus, our findings could partially explain why individuals with duodenal ulcer have a low risk for gastric cancer.

We also found that individuals with homozygous T alleles of rs505922 have significantly higher risk for duodenal ulcer. This SNP is located within the *ABO* gene, which encodes a glycosyltransferase. The synthesis of blood group ABH antigens is determined by variations in the *ABO* gene[26]. An analysis of 94 subjects showed that the T allele of rs505922 was in strong linkage disequilibrium (LD; $r^2 = 0.97$) with the *ABO* gene encoding the O blood type, which produces nonfunctional protein due to a single-nucleotide deletion in codon 87 (rs8176719), concordant with a previous report[27]. By using the genotyping results of two tagging SNPs, through which *ABO* alleles can be inferred (rs505922 and rs8176746; **Supplementary Table 9**)[27], we could successfully determine the ABO blood type in 98.6% of samples[28] (**Supplementary Table 10**). Association analysis showed that individuals with blood type O exhibited significantly higher risk for duodenal ulcer than those with blood type A ($P = 2.04 \times 10^{-6}$; OR = 1.43; **Supplementary Table 10**). In contrast, blood type B was associated with a lower risk of intestinal-type gastric cancer than blood type A ($P = 0.019$; OR = 0.85; **Supplementary Table 11**). Even though the extent of association was slightly different between the GWAS (OR = 1.45) and replication (OR = 1.23) samples for rs505922, our findings are consistent with previous epidemiological studies showing an association between the O blood type and duodenal ulcer[29]. Taken together, our data suggest that ABO blood type could be a marker for duodenal ulcer susceptibility.

Recent GWAS have identified association of the *ABO* gene with various diseases, such as pancreatic cancer[30] and myocardial infarction[31]. In addition, the severity of infectious diseases caused by *Escherichia coli* O157 or *Vibrio cholera* has been linked to ABO blood type[32]. ABH antigens are highly expressed in gastrointestinal epithelium[14], and the South American *H. pylori* strain P466 was shown to bind to the H antigen but not to the A antigen[33]. However, the absence of correlation between *H. pylori* infection and ABO blood groups has also been reported[34]. Therefore, further analyses are necessary to fully elucidate the role of the *ABO* gene in the development of duodenal ulcer.

We also investigated the association between previously reported genes and duodenal ulcer using samples from the GWAS stage (**Supplementary Table 12**). Of the 27 SNPs analyzed, 4 at the *VEGFA*, *IL6* and *COX1* loci showed suggestive associations ($P < 0.05$), although these associations were not significant after Bonferroni correction ($P < 0.0019$).

Through the analysis of 7,035 duodenal ulcer cases and 25,323 controls, we have demonstrated a role for two genetic variants in the development of duodenal ulcer. Of note, genotype frequencies for the risk alleles of rs2294008 and rs505922 in the Japanese population are lowest among the 11 HapMap populations (14.2% and 31.8%, respectively; **Supplementary Table 13**). Taken together, our findings provide new

insight into the molecular mechanism responsible for the lower risk of gastric cancer among individuals with duodenal ulcer and the lower incidence of duodenal ulcer in the Japanese population.

URLs. BioBank Japan Project participating hospitals (in Japanese), http://biobankjp.org/plan/member_hospital.html; R, http://www.r-project.org/; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; Primer3, http://frodo.wi.mit.edu/; LocusZoom, http://csg.sph.umich.edu/locuszoom/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank all the subjects and the members of the Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan, who donated their DNA for this work. We also thank A. Matsui and H. Tagaya and the technical staff of the Laboratory for Genotyping Development at the Center for Genomic Medicine at RIKEN for their technical support. This work was conducted as a part of the BioBank Japan Project, which was supported by the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government.

AUTHOR CONTRIBUTIONS

C.T., Y.N. and K. Matsuda conceived and designed the study. Y.U., K. Matsuo and M.K. performed genotyping. A.T. and N.K. performed quality control analysis for the GWAS. Y.N., K. Matsuda and M.K. managed DNA samples belonging to BioBank Japan. H.I. and K.T. managed DNA samples from the Aichi Cancer Center. C.T. analyzed and summarized all the results. C.T., Y.N. and K. Matsuda wrote the manuscript. Y.N. obtained funding for the study.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Sakamoto, H. *et al.* Genetic variation in *PSCA* is associated with susceptibility to diffuse-type gastric cancer. *Nat. Genet.* **40**, 730–740 (2008).
2. Wyatt, J.I., Rathbone, B.J., Dixon, M.F. & Heatley, R.V. *Campylobacter pyloridis* and acid induced gastric metaplasia in the pathogenesis of duodenitis. *J. Clin. Pathol.* **40**, 841–848 (1987).
3. Helicobacter and Cancer Collaborative Group. Gastric cancer and *Helicobacter pylori*: a combined analysis of 12 case control studies nested within prospective cohorts. *Gut* **49**, 347–353 (2001).
4. Hopkins, R.J., Girardi, L.S. & Turney, E.A. Relationship between *Helicobacter pylori* eradication and reduced duodenal and gastric ulcer recurrence: a review. *Gastroenterology* **110**, 1244–1252 (1996).
5. Schlemper, R.J., van der Werf, S.D., Vandenbroucke, J.P., Biemond, I. & Lamers, C.B. Peptic ulcer, non-ulcer dyspepsia and irritable bowel syndrome in The Netherlands and Japan. *Scand. J. Gastroenterol. Suppl.* **200**, 33–41 (1993).
6. Araki, S. & Goto, Y. Peptic ulcer in male factory workers: a survey of prevalence, incidence, and aetiological factors. *J. Epidemiol. Community Health* **39**, 82–85 (1985).
7. Khuroo, M.S., Mahajan, R., Zargar, S.A., Javid, G. & Munshi, S. Prevalence of peptic ulcer in India: an endoscopic and epidemiological study in urban Kashmir. *Gut* **30**, 930–934 (1989).
8. Kurata, J., Nogawa, A., Watanabe, Y. & Kawai, K. Peptic ulcer disease mortality. Comparison of native Japanese, Japanese Americans, and Caucasian Americans. *J. Clin. Gastroenterol.* **18**, 145–154 (1994).
9. Hansson, L.E. *et al.* The risk of stomach cancer in patients with gastric or duodenal ulcer disease. *N. Engl. J. Med.* **335**, 242–249 (1996).
10. Kusters, J.G., van Vliet, A.H. & Kuipers, E.J. Pathogenesis of *Helicobacter pylori* infection. *Clin. Microbiol. Rev.* **19**, 449–490 (2006).
11. Ainley, C.C., Forgacs, I.C., Keeling, P.W. & Thompson, R.P. Outpatient endoscopic survey of smoking and peptic ulcer. *Gut* **27**, 648–651 (1986).
12. Räihä, I., Kemppainen, H., Kaprio, J., Koskenvuo, M. & Sourander, L. Lifestyle, stress, and genes in peptic ulcer disease: a nationwide twin cohort study. *Arch. Intern. Med.* **158**, 698–704 (1998).
13. Kamatani, Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat. Genet.* **42**, 210–215 (2010).
14. Clausen, H. & Hakomori, S. ABH and related histo-blood group antigens; immunochemical differences in carrier isotypes and their distribution. *Vox Sang.* **56**, 1–20 (1989).
15. Walker, M.M. & Dixon, M.F. Gastric metaplasia: its role in duodenal ulceration. *Aliment. Pharmacol. Ther.* **10** (suppl. 1), 119–128 (1996).
16. Gu, Z. *et al.* Prostate stem cell antigen (PSCA) expression increases with high gleason score, advanced stage and bone metastasis in prostate cancer. *Oncogene* **19**, 1288–1296 (2000).
17. Saeki, N., Gu, J., Yoshida, T. & Wu, X. Prostate stem cell antigen: a Jekyll and Hyde molecule? *Clin. Cancer Res.* **16**, 3533–3538 (2010).
18. Wu, X. *et al.* Genetic variation in the prostate stem cell antigen gene *PSCA* confers susceptibility to urinary bladder cancer. *Nat. Genet.* **41**, 991–995 (2009).
19. Nakai, K. & Kanehisa, M. A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* **14**, 897–911 (1992).
20. Edgren, G. *et al.* Risk of gastric cancer and peptic ulcers in relation to ABO blood type: a cohort study. *Am. J. Epidemiol.* **172**, 1280–1285 (2010).
21. Nakao, M. *et al.* ABO genotype and the risk of gastric cancer, atrophic gastritis, and *Helicobacter pylori* infection. *Cancer Epidemiol. Biomarkers Prev.* **20**, 1665–1672 (2011).
22. Marra, E. *et al.* Growth delay of human bladder cancer cells by Prostate Stem Cell Antigen downregulation is associated with activation of immune signaling pathways. *BMC Cancer* **10**, 129 (2010).
23. Gu, Z., Yamashiro, J., Kono, E. & Reiter, R.E. Anti-prostate stem cell antigen monoclonal antibody 1G8 induces cell death *in vitro* and inhibits tumor growth *in vivo* via a Fc-independent mechanism. *Cancer Res.* **65**, 9495–9500 (2005).
24. Raff, A.B., Gray, A. & Kast, W.M. Prostate stem cell antigen: a prospective therapeutic and diagnostic target. *Cancer Lett.* **277**, 126–132 (2009).
25. Ohara, T., Morishita, T., Suzuki, H., Masaoka, T. & Ishii, H. Perforin and granzyme B of cytotoxic T lymphocyte mediate apoptosis irrespective of *Helicobacter pylori* infection: possible act as a trigger of peptic ulcer formation. *Hepatogastroenterology* **50**, 1774–1779 (2003).
26. Yamamoto, F., Clausen, H., White, T., Marken, J. & Hakomori, S. Molecular genetic basis of the histo-blood group ABO system. *Nature* **345**, 229–233 (1990).
27. Nakao, M. *et al.* ABO blood group alleles and the risk of pancreatic cancer in a Japanese population. *Cancer Sci.* **102**, 1076–1080 (2011).
28. Fujita, Y., Tanaka, K. & Tanimura, M. The distribution of the Rh(D) blood types in Japan. *Jinrui Idengaku Zasshi* **23**, 197–209 (1978).
29. Clarke, C.A. *et al.* The relationship of the ABO blood groups to duodenal and gastric ulceration. *BMJ* **2**, 643–646 (1955).
30. Amundadottir, L. *et al.* Genome-wide association study identifies variants in the *ABO* locus associated with susceptibility to pancreatic cancer. *Nat. Genet.* **41**, 986–990 (2009).
31. Reilly, M.P. *et al.* Identification of *ADAMTS7* as a novel locus for coronary atherosclerosis and association of *ABO* with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet* **377**, 383–392 (2011).
32. Harris, J.B. *et al.* Blood group, immunity, and risk of infection with *Vibrio cholerae* in an area of endemicity. *Infect. Immun.* **73**, 7422–7427 (2005).
33. Aspholm-Hurtig, M. *et al.* Functional adaptation of BabA, the *H. pylori* ABO blood group antigen binding adhesin. *Science* **305**, 519–522 (2004).
34. Wu, T.C., Chen, L.K. & Hwang, S.J. Seroprevalence of *Helicobacter pylori* in school-aged Chinese in Taipei City and relationship between ABO blood groups. *World J. Gastroenterol.* **9**, 1752–1755 (2003).

ONLINE METHODS

Sample collection. We obtained DNA from 7,035 subjects with duodenal ulcer from the BioBank Japan Personalized Medicine Project of the Ministry of Education, Culture, Sports, Science and Technology of Japan[35]. In the BioBank Japan Project, DNA and serum of subjects were collected through a collaborating network of 66 hospitals throughout Japan (see **Supplementary Note**). Clinical information on subjects, including drug history, was obtained from clinical records of each participating hospital. More than 200,000 individuals with 47 diseases, irrespective of prior treatment, were enrolled in this project from 2003 through 2009. A list of participating hospitals is provided at the BioBank Japan website (see URLs). A total of 25,323 control samples without duodenal ulcer were obtained from BioBank Japan or from healthy volunteers from the Midosuji Rotary Club in Osaka, Japan. Controls for the GWAS included subjects with colon, breast, prostate, lung, pancreatic or liver cancer, diabetes, myocardial infarction, brain infarction, arteriosclerosis obliterans, atrial fibrillation, cholangiocarcinoma, drug eruption, liver cirrhosis and amyotrophic lateral sclerosis. Controls for the replication analysis included healthy volunteers and individuals with chronic hepatitis B or C, cervical, esophageal, hematological or ovarian cancer, pulmonary tuberculosis, keroidosis and endometriosis. Controls for gastric cancer included subjects with diabetes, myocardial infarction, brain infarction, arteriosclerosis obliterans, atrial fibrillation, drug eruption, liver cirrhosis and amyotrophic lateral sclerosis. A total of 37 individuals with *H. pylori*–positive duodenal ulcer or 793 healthy controls were randomly selected from non-cancer outpatient visitors to the Aichi Cancer Center Hospital in Nagoya, Japan between January 2001 and November 2005 (ref. 36). *H. pylori* infection status was confirmed by determining plasma levels of IgG to the bacterium with a commercially available direct ELISA kit (Eiken Kagaku). Positivity for *H. pylori* infection was defined by *H. pylori*–specific IgG levels of greater than 10 U/ml in serum. We excluded subjects with gastric cancer or gastric ulcer from all controls. We excluded individuals treated with nonsteroidal anti-inflammatory drugs or steroid hormones from both duodenal ulcer cases and controls. We excluded subjects with a history of any cancer or peptic ulcer from the controls for gastric cancer. All subjects were Japanese and provided written informed consent. The clinical and demographic details of the samples are summarized in **Table 1**. This research project was approved by the ethical committees of The University of Tokyo, RIKEN and the Aichi Cancer Center.

SNP genotyping and quality control. In the GWAS stage, 1,043 duodenal ulcer and 21,694 controls were genotyped using Illumina HumanHap610-Quad BeadChip arrays. From the 101 SNPs from 42 distinct genomic regions with $P < 5 \times 10^{-5}$ in the GWAS stage, we selected the most strongly associated SNPs from each genomic region for the replication stage. Because we failed to design probes for rs1051631 and rs16896391, we selected rs10516733 and rs7775478 instead. In the replication stage, we genotyped 5,992 independent duodenal ulcer cases and 3,629 controls without peptic ulcer using the multiplex PCR–based Invader assay (Third Wave Technologies) and Illumina HumanHap550v3 BeadChip arrays, respectively. Because we failed to obtain genotype data for rs2240458 with Illumina HumanHap550v3, the genotyping results for rs2240458 in the replication control samples were imputed using a hidden Markov model. *H. pylori*–positive duodenal ulcer cases and healthy controls from the Aichi Cancer Center were genotyped by TaqMan assay. We observed 100% concordance between results from direct sequencing and those from Invader assays, Illumina BeadChips and TaqMan assays when we analyzed 94 samples for the two significantly associated SNPs (rs505922 and rs2294008). We performed a standard quality control procedure to exclude SNPs with low call rate (<99%), *P* value for the Hardy-Weinberg equilibrium test of $<1.0 \times 10^{-7}$ for controls and minor allele frequency (MAF) of < 0.01. In total, we obtained 480,327 SNPs for analysis in the screening stage.

Statistical analysis. The association of SNPs with the phenotype in the GWAS, replication and combination analyses was tested by multivariate logistic regression analysis with adjustment for age at recruitment and gender, by assuming an additive, dominant or recessive model using PLINK[37]. At the GWAS stage, the genomic inflation factor λ was derived by applying *P* values from logistic regression in an additive model for all the tested SNPs. Quantile-quantile plots were drawn using the R program. Odds ratios were calculated by considering the non-risk allele as a reference, unless otherwise stated. Significance levels after Bonferroni correction for multiple testing were $P = 3.47 \times 10^{-8}$ (0.05/(480,327 × 3)) in the GWAS stage and $P = 1.19 \times 10^{-3}$ (0.05/42) in the replication stage, under a genetic model that provided the minimum *P* value in the screening stage. Heterogeneity across two stages was examined by the Cochran *Q* test[38]. We calculated the population attributable risk by $\mathrm{PAR} = (f_{\mathrm{homo}}(\mathrm{OR}_{\mathrm{homo}} - 1) + f_{\mathrm{hetero}}(\mathrm{OR}_{\mathrm{hetero}} - 1))/(1 + f_{\mathrm{homo}}(\mathrm{OR}_{\mathrm{homo}} - 1) + f_{\mathrm{hetero}}(\mathrm{OR}_{\mathrm{hetero}} - 1))$, where *f* is the frequency in the control subjects. PAR is defined as the reduction in incidence that would be achieved if the population had been entirely unexposed.

Imputation analysis. Sequences from the GWAS samples or replication controls that were not genotyped were imputed using a hidden Markov model in MACH[39] and HapMap JPT data (release 27). We applied the same quality control procedure as in the GWAS stage. For all SNPs that passed the quality control criteria, association was tested by logistic regression analysis with adjustment for age and gender.

Cell culture and transfections. Cell lines were purchased from the American Type Culture Collection, Lonza Biologics or the Japanese Collection of Research Bioresources. Cells were transfected with plasmids using FuGENE6 (Roche). For the analysis of proteasomal degradation of PSCA protein, cells were treated with 10 μM of proteasome inhibitor MG132 (Calbiochem) for 10 h before harvest.

Plasmid construction. cDNA fragments of *PSCA* were amplified and cloned into the pcDNA3.1 vector (Invitrogen). Plasmid structure is shown in **Supplementary Figure 5**. The primer sequences for cloning are given in **Supplementary Table 14**.

Quantitative RT-PCR. PolyA RNA or total RNA for normal tissues was purchased from Calbiochem and BioChain. Most of these samples were pooled RNA from multiple individuals; only duodenal tissue was derived from a single individual carrying homozygous C alleles at rs2294008 and homozygous C alleles at rs505922 (estimated from cDNA sequence at rs2294008 and rs8176719). cDNA was synthesized with the SuperScript Preamplification System (Invitrogen). Quantitative RT-PCR was conducted using the SYBR Green I Master on a LightCycler 480 (Roche). The primer sequences are given in **Supplementary Table 14**.

Protein blotting. To prepare whole-cell extracts, cells were collected and lysed in SDS sample buffer with sonication and centrifuged at 16,000*g* for 15 min. For treatment with N-glycosidase, cells were lysed in RIPA buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1% SDS, 0.5% DOC, 1% NP-40 and 1 mM PMSF) and incubated with 5 U of N-glycosidase F (Calbiochem) at 37 °C for 20 h. Samples were subjected to SDS-PAGE, and immunoblotting was performed using standard procedures.

Antibodies. Polyclonal antibodies to PSCA (NB100-91938 and ab64919) were purchased from Novus Biologicals and Abcam, respectively. Monoclonal antibody to β-actin (clone AC15) was purchased from Sigma.

Immunocytochemistry. To select cell lines for immunocytochemistry, we analyzed *PSCA* mRNA expression and genotype at rs2294008 in a total of 44 cell lines (40 lung cancer and 4 prostate cancer cell lines). From these, we selected seven (three lung cancer and four prostate cancer cell lines) that were homozygous (or hemizygous) for the T or C allele at rs2294008 and expressed high levels of *PSCA* mRNA. We detected PSCA protein in four cell lines (PC3, Du145, C42B and NCI-H522) by immunocytochemistry and have presented the result for these cell lines. Immunocytochemistry was performed as previously described[40]. For staining of endogenous PSCA protein, cells were fixed using methanol. For cell surface staining, cells were labeled under nonpermeabilizing conditions using buffer without Triton X-100.

ABO blood type. In this study, we examined three SNPs at the *ABO* gene (rs505922, rs8176719 and rs8176746). These SNPs have been shown

to be tagging SNPs for the ABO blood type in the Japanese population[21]. Single-nucleotide deletion in exon 6 (rs8176719) affecting at amino-acid position 87 results in the O blood type, and a C or A at position 796 in exon 7 (rs8176746) distinguishes the B from the A blood type, respectively. We also confirmed strong LD between rs505922 and rs8176719 ($r^2 = 0.97$) by genotyping both SNPs in 94 individuals. Thus, we can deduce ABO blood types based on the genotypes of rs505922 and rs8176746 (**Supplementary Table 9**).

Software. For general statistical analysis, we employed the R statistical environment version 2.6.1 or PLINK-1.06. Haploview software version 4.2 (ref. 41) was used to calculate LD and to draw Manhattan plots. The Primer3 v0.3.0 web tool was used to design primers. We employed LocusZoom to plot regional association plots.

35. Nakamura, Y. The BioBank Japan Project. *Clin. Adv. Hematol. Oncol.* **5**, 696–697 (2007).
36. Tajima, K. *et al.* A model of practical cancer prevention for out-patients visiting a hospital: the hospital-based epidemiologic research program at Aichi Cancer Center (HERPACC). *Asian Pac. J. Cancer Prev.* **1**, 35–47 (2000).
37. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
38. Breslow, N.E. & Day, N.E. Statistical methods in cancer research. Volume II–The design and analysis of cohort studies. *IARC Sci. Publ.* 1–406 (1987).
39. Scott, L.J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
40. Tanikawa, C. *et al.* Regulation of protein Citrullination through p53/PADI4 network in DNA damage response. *Cancer Res.* **69**, 8761–8769 (2009).
41. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).