Table 1. Association between Reg IV immunostaining and clinicopathological variables

	Reg IV		
	Positive	Negative	*P*-value
pT stage			
≤2b	6 (22%)	21	0.2001
≥2c	8 (11%)	63	
Gleason score			
≤7	8 (11%)	66	0.1001
≥8	6 (25%)	18	
Preoperative PSA concentration			
<20	12 (13%)	77	0.6122
≥20	2 (22%)	7	
MUC2			
Positive	14 (19%)	58	0.0182
Negative	0 (0%)	26	
Chromogranin A			
Positive	10 (32%)	21	0.0012
Negative	4 (6%)	63	
EGFR			
Positive	8 (32%)	17	0.0067
Negative	6 (8%)	67	
Phospho-EGFR (Tyr^{1068}) (n = 25)			
Positive	6 (67%)	3	0.0099
Negative	2 (13%)	14	

EGFR, epidermal growth factor receptor; PSA, prostate-specific antigen.

Table 2. Multivariate analysis of factors influencing relapse-free survival

	Hazard ratio	(95% CI)	χ^2-test	*P*-value
Reg IV staining				
Negative	1	(Reference)	4.640	0.0312
Positive	2.848	(1.099–7.381)		
Chromogranin A staining				
Negative	1	(Reference)	0.005	0.9431
Positive	1.034	(0.414–2.582)		
Gleason score				
≤7	1	(Reference)	10.235	0.0014
≥8	3.747	(1.668–8.416)		
Preoperative PSA concentration				
<20	1	(Reference)	4.410	0.0357
≥20	2.938	(1.074–8.036)		

CI, confidence interval; PSA, prostate-specific antigen.

characteristics. Reg IV staining was not correlated with pT stage, Gleason score or preoperative PSA concentration (Table 1).

We also performed an immunohistochemical analysis of MUC2 expression because Reg IV is associated with intestinal differentiation in gastric cancer.[5] In PCa tissues, MUC2 staining was observed in goblet cell-like vesicles (Fig. 2b) and perinuclear regions (Fig. 2b) of tumor cells. Of the 98 PCa cases, MUC2 staining was observed in 72 (73%). MUC2-positive PCa cells comprised 1–30% of tumor cells. MUC2 staining was considered positive if any tumor cells were stained. We analyzed the relation between MUC2 staining and clinicopathological characteristics. MUC2 positivity was found more frequently in PCa showing a pT stage of 2c or more (59/71, 83%) than in PCa showing a pT stage of 2b or less (13/27, 48%, $P = 0.0182$, Fisher's exact test). MUC2 staining was not correlated with Gleason score or preoperative PSA concentration. Association between Reg IV and MUC2 staining was also analyzed. Reg IV positivity was found more frequently in MUC2-positive cases (14/72, 19%) than in MUC2-negative cases (0/26, 0%, $P = 0.0182$, Fisher's exact test) (Table 1). We confirmed that almost all tumor cells showing mucin-like staining of Reg IV were positive for MUC2 by double-immunofluorescence staining (Fig. 2c). Some PCa cells showing perinuclear Reg IV staining also showed MUC2 staining (Fig. 2d). These results indicated that Reg IV is associated with intestinal differentiation of PCa.

Immunostaining of chromogranin A was also performed. Representative results of chromogranin A immunostaining in PCa are shown in Fig. 2(e). Of the 98 PCa cases, chromogranin A staining was observed in 31 (32%). Chromogranin A positivity was observed in 1–40% of PCa cells. Chromogranin A staining was considered positive if any tumor cells were stained. We analyzed the relation between chromogranin A staining and clinicopathological characteristics. Chromogranin A positivity were found more frequently in PCa showing a Gleason score of 8 or more (12/24, 50%) than in PCa showing a Gleason score of 7 or less (19/74, 26%, $P = 0.0418$, Fisher's exact test). Chromogranin A staining was not correlated with pT stage or preoperative PSA concentration. Association between Reg IV and chromogranin A staining was also analyzed. Reg IV positivity was found more frequently in chromogranin A-positive cases (10/31, 32%) than in chromogranin A-negative cases (4/67, 6%, $P = 0.0012$, Fisher's exact test) (Table 1). These results indicated that Reg IV is associated with neuroendocrine differentiation. However, double-immunofluorescence staining revealed that Reg IV-positive cells did not show chromogranin A staining (Fig. 2f). Therefore, a paracrine effect of secreted Reg IV may be involved in neuroendocrine differentiation in PCa.

Relation between Reg IV immunostaining and relapse-free survival of PCa patients. We next examined the relation between Reg IV immunostaining and relapse-free survival in PCa. Univariate analysis revealed that Reg IV staining ($P = 0.0004$, log–rank test), chromogranin A staining ($P = 0.0494$), Gleason score ($P < 0.0001$) and preoperative PSA concentration ($P = 0.0167$) (Fig. 3a–d) were significant prognostic factors for relapse-free survival in patients with PCa, whereas MUC2 staining and pT stage did not correlate with relapse-free survival. We then used the Cox proportional hazards multivariate model to examine the association of clinicopathological factors and Reg IV and chromogranin A staining with relapse-free survival. Multivariate analysis indicated that Reg IV staining, Gleason score and preoperative PSA concentration were independent predictors of relapse-free survival in patients with PCa (Table 2). These results suggested that Reg IV expression directly contributes to the malignant potential of PCa.

Reg IV activates EGFR in LNCaP cells. Statistical analysis revealed that Reg IV is an independent predictor of relapse-free survival in patients with clinically localized PCa. However, the underlying mechanism remains unclear. Therefore, Reg IV-CM was prepared and the function of Reg IV was analyzed. It has been reported that recombinant human Reg IV induces rapid phosphorylation of the EGFR at Tyr^{992} and Tyr^{1068} and of Akt at Thr^{308} and Ser^{473} resulting in increased AP-1 transcription factor activity.[12] Western blot analysis of the EGFR showed that colo320 did not express EGFR protein (data not shown). Therefore, we assumed that Reg IV overexpression has little effect in colo320 cells. Thus, Reg IV-CM was prepared with the colo320 cell line. With a specific antibody, we verified the expression of Reg IV protein in Reg IV-CM prepared from colo320 cells (Fig. 4a). We also examined whether EGF and TGF-α could be detected by ELISA in Reg IV-CM and control medium. EGF or TGF-α was not detected in Reg IV-CM or control medium (data not shown). We tested the specificity of the antiphospho-EGFR (Tyr^{1068}) and antiphospho-EGFR (Tyr^{992})

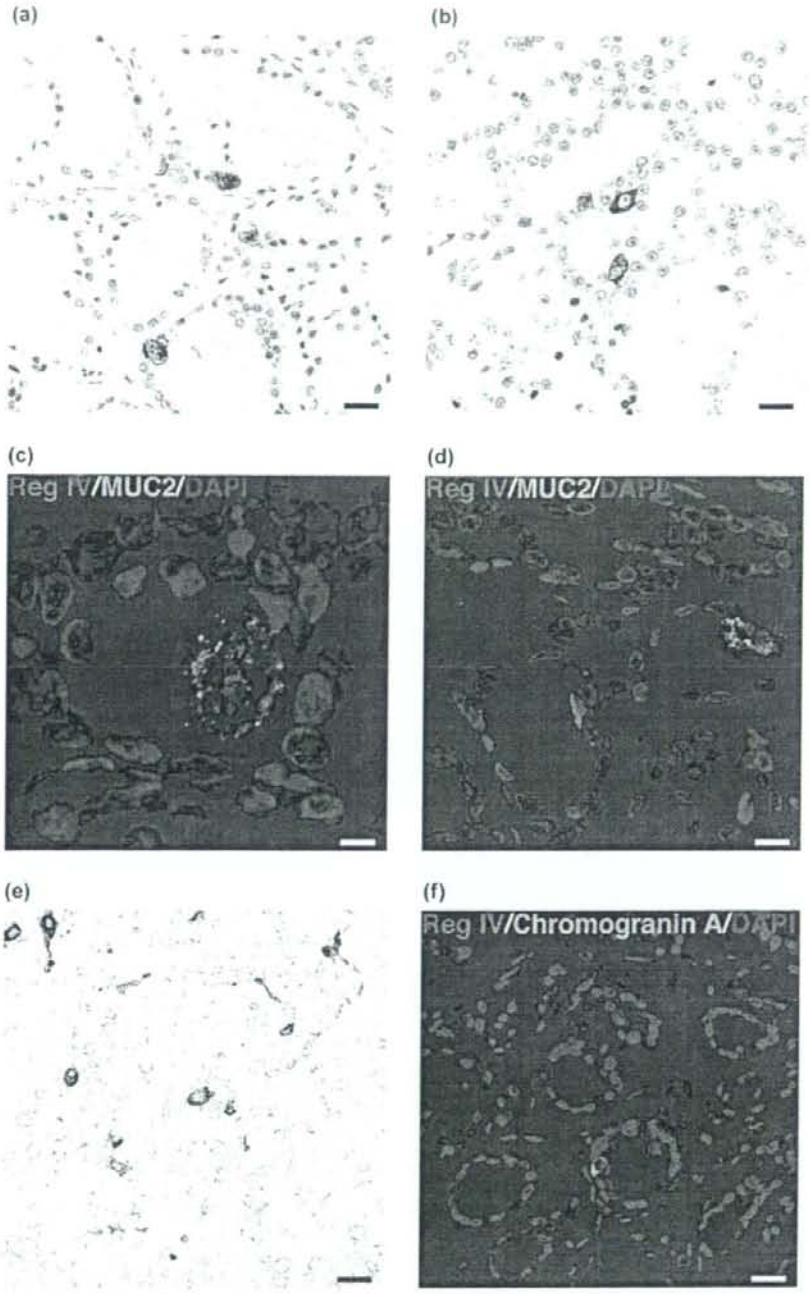

Fig. 2. Immunohistochemical analysis of MUC2 and chromogranin A expression in clinically localized prostate cancer (PCa). (a) Immunostaining of MUC2 in PCa. Mucin-like staining of MUC2 is present in goblet cell-like vesicles of a tumor cell. Scale line, 25 μm. (b) Immunostaining of MUC2 in PCa. Perinuclear staining of MUC2 is present in a tumor cell. Scale line, 25 μm. (c) Double-immunostaining of Reg IV (red) and MUC2 (green). Nuclei are stained with 4′,6-diamidino-2-phenylindole (DAPI; blue). Reg IV staining is present with MUC2 in goblet cell-like vesicles of a tumor cell. Scale line, 13 μm. (d) Double-immunostaining of Reg IV (red) and MUC2 (green). Nuclei are stained with DAPI (blue). A tumor cell showing perinuclear staining of Reg IV also shows MUC2 staining. Scale line, 25 μm. (e) Immunostaining of chromogranin A in PCa. Scale line, 25 μm. (f) Double-immunostaining of Reg IV (red) and chromogranin A (green). Nuclei are stained with DAPI (blue). Scale line, 25 μm.

doi: 10.1111/j.1349-7006.2008.00846.x

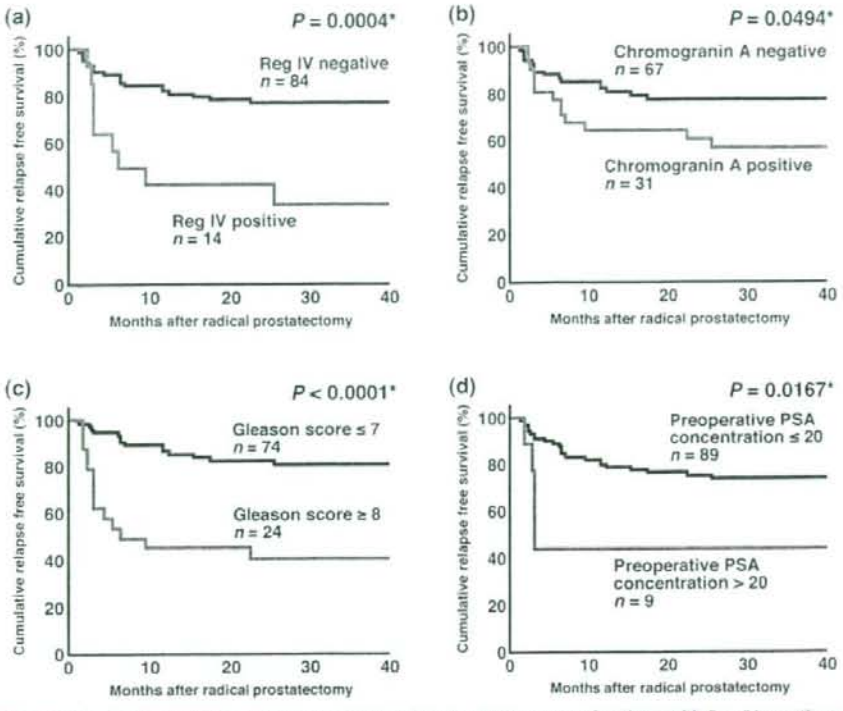

Fig. 3. Relapse-free survival of patients with prostate cancer (PCa). (a) Kaplan–Meier curves of patients with Reg IV-negative or Reg IV-positive PCa. (b) Kaplan–Meier curves of patients with chromogranin A-negative or chromogranin A-positive PCa. (c) Kaplan–Meier curves of patients with low Gleason score (≤7) or high Gleason score (≥8) PCa. (d) Kaplan–Meier curves of patients with low preoperative prostate specific antigen (PSA) concentration (≤20) or high preoperative PSA concentration (>20) PCa. *Log-rank test.

antibodies. Western blotting of lysates of LNCaP cell line was performed. The antiphospho-EGFR (Tyr^{1068}) antibody detected a single band of 170 kDa on Western blots of LNCaP cells treated with TGF-α or EGF (Fig. 4b). The antiphospho-EGFR (Tyr^{992}) antibody also detected a single band of 170 kDa on Western blots of LNCaP cells treated with TGF-α or EGF (data not shown). These data show that these antiphospho-EGFR (Tyr^{1068}) and antiphospho-EGFR (Tyr^{992}) antibodies specifically recognize phospho-EGFR protein. The effect of Reg IV-CM on EGFR phosphorylation was then investigated in LNCaP cells. EGFR was phosphorylated at Tyr^{1068} but not Tyr^{992} in LNCaP cells treated with Reg IV-CM (Fig. 4c). These findings indicated that Reg IV is involved in EGFR phosphorylation in an EGF- and TGF-α-independent manner in LNCaP cells. Because it has been reported that growth factors enhance *REG4* mRNA expression in colon cancer cells,[15] the effect of EGF and TGF-α on *REG4* expression was also investigated in LNCaP cells. Neither EGF or TGF-α had any significant effect on Reg IV protein expression in LNCaP cells (Fig. 4d). Similar results were also obtained from the DU145 cell line (data not shown).

We next examined whether expression of Reg IV activates phosphorylation of the EGFR at Tyr^{1068} in PCa tissue samples. Staining of the EGFR was found in 25 (26%) of the 98 PCa cases. All 25 PCa cases were considered EGFR-positive. Of 25 PCa cases positive for EGFR, eight (32%) were positive for Reg IV. In eight PCa cases positive for both Reg IV and EGFR, Reg IV-positive tumor cells were found near EGFR-positive tumor cells (Fig. 4e). Because Reg IV-CM phosphorylated the EGFR at Tyr^{1068} in LNCaP cells, we further investigated phosphorylation of the EGFR at Tyr^{1068} in 25 PCa cases positive for EGFR staining. Phosphorylation of the EGFR at Tyr^{1068} was found in nine (36%) PCa cases (Fig. 4e, inset). Reg IV positivity was found more frequently in Tyr^{1068}-positive EGFR cases than in Tyr^{1068}-negative EGFR cases ($P = 0.0099$, Fisher's exact test) (Table 1). Because it is possible that EGFR phosphorylation is due to the stimulation by some growth factors rather than Reg IV protein in the human PCa tissues, we performed immunohistochemical analysis of TGF-α. As reported previously,[16] in most cases, TGF-α expression was present in the stroma; however, co-expression of EGFR and TGF-α in tumor cells was observed in several PCa cases. Tumor cell staining of the TGF-α was found in 17 (17%) of the 98 PCa cases (Fig. 4f). We regarded these cases with TGF-α-positive tumor cells as TGF-α-positive. Of 25 PCa cases positive for EGFR, nine (36%) were positive for TGF-α. TGF-α positivity was found more frequently in Tyr^{1068}-positive EGFR cases (7/9, 78%) than in Tyr^{1068}-negative EGFR cases (2/16, 13%, $P = 0.0022$, Fisher's exact test). Because it has been reported that growth factors enhance *REG4* mRNA expression in colon cancer cells,[15] association between Reg IV and TGF-α staining was analyzed. Reg IV positivity was found more frequently in TGF-α-positive cases (8/17, 47%) than in TGF-α-negative cases (6/81, 7%, $P = 0.0003$, Fisher's exact test).

Discussion

Overexpression of *REG4* mRNA has been reported in PCa by *in situ* hybridization.[11] A majority of metastatic PCa tumors express high levels of *REG4* mRNA. In addition, *REG4*

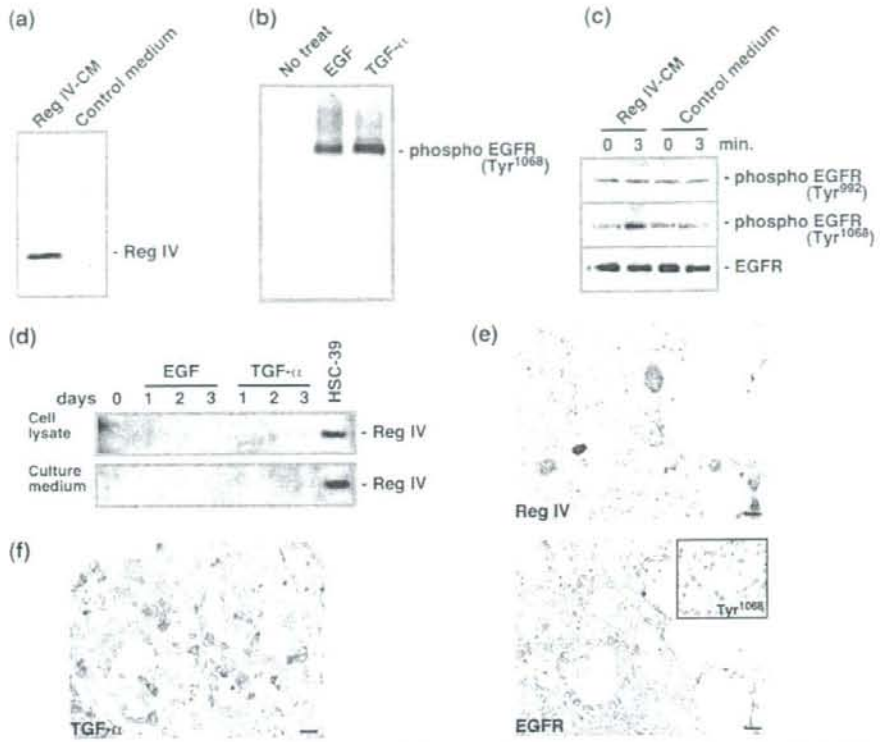

Fig. 4. Effect of Reg IV on epidermal growth factor receptor (EGFR) phosphorylation. (a) 10 μL of Reg IV-conditioned medium and control medium were analyzed by Western blot with a rabbit polyclonal antibody against Reg IV. (b) LNCaP cells were cultured with either EGF (100 nM) or transforming growth factor (TGF)-α (10 nM) for 3 min. Whole-cell lysates were prepared and analyzed by Western blot with antiphospho-EGFR (Tyr^{1068}) antibody. (c) LNCaP cells were cultured with either Reg IV-CM or control medium for 3 min. Whole-cell lysates were prepared and analyzed by Western blot with antiphospho-EGFR (Tyr^{992}) or antiphospho-EGFR (Tyr^{1068}) antibody. The samples were also probed with anti-EGFR antibody to verify equal loading. (d) LNCaP cells were cultured with either EGF (100 nM) or TGF-α (10 nM) for 1, 2 or 3 days. Whole-cell lysates and culture medium were analyzed by Western blot with anti-Reg IV antibody. (e) Expression of Reg IV and EGFR was examined by immunohistochemistry in serial sections of prostate cancer. Inset, phospho-EGFR (Tyr^{1068}) staining of a serial section. Scale line, 25 μm. (f) Expression of TGF-α was examined by immunohistochemistry. Scale line, 25 μm.

expression is significantly more intense in high-grade PCa (Gleason score, 7–10) than in low-grade PCa (Gleason score, 5–6). In the present study of clinically localized PCa, Reg IV-positive cases showed unfavorable prognosis with respect to relapse-free survival. Several autocrine and paracrine signaling pathways involving the EGFR and its ligands, EGF and TGF-α, are postulated to stimulate tumor cell proliferation independent of androgen activity.[17] These pathways become much more essential to PCa growth once androgen insensitivity occurs. EGFR signaling pathways are also involved in PCa invasion and angiogenesis, both of which are crucial for progression and metastasis.[18,19] In the present study, we showed that Reg IV-CM phosphorylated the EGFR at Tyr^{1068}. These findings suggest that Reg IV activates phosphorylation of the EGFR in human PCa tissues. Although the precise mechanism of EGFR phosphorylation by Reg IV remains unclear, it is not likely that secretion of the EGFR ligands, such as EGF and TGF-α, is involved because no EGF or TGF-α was detected in Reg IV-CM by ELISA. Taken together, these results suggest that Reg IV plays important roles in tumor progression and unfavorable prognosis. However, TGF-α positivity was also found more frequently in Tyr^{1068}-positive EGFR cases than in Tyr^{1068}-negative EGFR cases. Because frequency of Reg IV positivity in Tyr^{1068}-positive EGFR cases (67%) was lower than that of TGF-α positivity in Tyr^{1068}-positive EGFR cases (78%), EGFR phosphorylation may be due to the stimulation by TGF-α rather than Reg IV protein in PCa tissues.

Androgen-deprivation therapy has been used for decades in the treatment of prostate cancer.[20] Although this treatment is initially very effective in hormone-dependent cancers, they invariably become hormone-refractory and metastasize, resulting in death.[21] Novel therapies that target androgen-independent proliferation, as well as invasion and angiogenesis, may have enormous potential for improving the care of patients with advanced PCa. Several possible mechanisms by which PCa can escape the effects of androgen-deprivation therapy have been reported.[21] Among them, EGFR expression has been reported to be associated with hormone-refractory status.[22] Therefore, phosphorylation of the EGFR by Reg IV may participate in the acquisition of hormone-refractory status. In the present study, few tumor cells (1–10%) in clinically localized PCa showed Reg IV staining, although residual hormone-refractory PCa has been reported to express high levels of *REG4* mRNA.[11] These results suggest that Reg IV-positive tumor cells may escape the effects

doi: 10.1111/j.1349-7006.2008.00846.x

of androgen-deprivation therapy, resulting in an increased number of Reg IV-positive tumor cells in residual hormone-refractory PCa.

It has been reported that growth factors enhance *REG4* mRNA expression in colon cancer cells.[15] In fact, in PCa tissues, Reg IV positivity was found more frequently in TGF-α-positive cases than in TGF-α-negative cases. In contrast, neither EGF or TGF-α had any significant effect on Reg IV protein expression in LNCaP and DU145 cell lines, suggesting that, in addition to TGF-α, other factors are needed to express Reg IV in PCa cells.

In the present study, Reg IV-positive cases were frequently found in association with MUC2 positivity. It has been suggested that the acquisition by tumor cells of a mucinous phenotype, such as MUC2 expression, is involved in hormonal escape in PCa.[23] Reg IV-positive cases were frequently found in association with chromogranin A positivity. Neuroendocrine differentiation has been also reported to be correlated with tumor aggressiveness, short survival and poor response to androgen-deprivation therapy.[24,25] Therefore, Reg IV may be a key factor mediating the hormone-refractory phenotype in MUC2-positive or chromogranin A-positive PCa. We observed two Reg IV staining patterns, mucin-like staining and perinuclear staining. In gastric cancer, mucin-like Reg IV staining is associated with MUC2 positivity.[5] Perinuclear Reg IV staining is detected in cells with neuroendocrine differentiation and that show chromogranin A positivity. In PCa tissue, although almost all tumor cells showing mucin-like staining of Reg IV were positive for MUC2, some tumor cells showing perinuclear Reg IV staining also showed MUC2 staining. The significance of the difference between mucin-like Reg IV staining and perinuclear Reg IV staining is unclear; however, there were several PCa cases in which both staining patterns were observed and we presume that these staining patterns are not independent.

In conclusion, we showed that Reg IV immunostaining is an independent predictor of relapse-free survival in patients with clinically localized PCa. To clarify whether Reg IV immunostaining is useful for the identification of patients most likely to benefit from adjuvant treatment, the association between Reg IV staining and response to adjuvant therapies should be investigated. We also showed that Reg IV-CM induces EGFR phosphorylation. Because Reg IV expression is narrowly restricted in non-cancerous tissues, Reg IV may be a good therapeutic target for PCa.

Acknowledgments

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 from the Ministry of Health, Labor and Welfare of Japan. We thank Ms Emiko Hisamoto for 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 thank the Analysis Center of Life Science, Hiroshima University, for the use of their facilities.

References

1 Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. *CA Cancer J Clin* 2007; **57**: 43–66.
2 D'Amico AV, Whittington R, Malkowicz SB *et al.* Clinical utility of the percentage of positive prostate biopsies in defining biochemical outcome after radical prostatectomy for patients with clinically localized prostate cancer. *J Clin Oncol* 2000; **18**: 1164–72.
3 Oue N, Hamai Y, Mitani Y *et al.* Gene expression profile of gastric carcinoma: identification of genes and tags potentially involved in invasion, metastasis, and carcinogenesis by serial analysis of gene expression. *Cancer Res* 2004; **64**: 2397–405.
4 Aung PP, Oue N, Mitani Y *et al.* Systematic search for gastric cancer-specific genes based on SAGE data: melanoma inhibitory activity and matrix metalloproteinase-10 are novel prognostic factors in patients with gastric cancer. *Oncogene* 2006; **25**: 2546–57.
5 Oue N, Mitani Y, Aung PP *et al.* Expression and localization of Reg IV in human neoplastic and non-neoplastic tissues: Reg IV expression is associated with intestinal and neuroendocrine differentiation in gastric adenocarcinoma. *J Pathol* 2005; **207**: 185–98.
6 Mitani Y, Oue N, Matsumura S *et al.* Reg IV is a serum biomarker for gastric cancer patients and predicts response to 5-fluorouracil-based chemotherapy. *Oncogene* 2007; **26**: 4383–93.
7 Isaacs W, De Marzo A, Nelson WG. Focus on prostate cancer. *Cancer Cell* 2002; **2**: 113–16.
8 Hasegawa Y, Matsubara A, Teishima J *et al.* DNA methylation of the RIZ1 gene is associated with nuclear accumulation of p53 in prostate cancer. *Cancer Sci* 2007; **98**: 32–6.
9 Oue N, Kuniyasu H, Noguchi T *et al.* Serum concentration of Reg IV in patients with colorectal cancer: overexpression and high serum levels of Reg IV are associated with liver metastasis. *Oncology*, 2007; **72**: 371–80.
10 Takehara A, Eguchi H, Ohigashi H *et al.* Novel tumor marker REG4 detected in serum of patients with resectable pancreatic cancer and feasibility for antibody therapy targeting REG4. *Cancer Sci* 2006; **97**: 1191–7.
11 Gu Z, Rubin MA, Yang Y *et al.* Reg IV: a promising marker of hormone refractory metastatic prostate cancer. *Clin Cancer Res* 2005; **11**: 2237–43.
12 Bishnupuri KS, Luo Q, Murmu N, Houchen CW, Anant S, Dieckgraefe BK. Reg IV activates the epidermal growth factor receptor/Akt/AP-1 signaling pathway in colon adenocarcinomas. *Gastroenterology* 2006; **130**: 137–49.
13 Yasui W, Ayhan A, Kitadai Y *et al.* Increased expression of p34cdc2 and its kinase activity in human gastric and colonic carcinomas. *Int J Cancer* 1993; **53**: 36–41.
14 Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* 1966; **50**: 163–70.
15 Nanakin A, Fukui H, Fujii S *et al.* Expression of the REG IV gene in ulcerative colitis. *Lab Invest* 2007; **87**: 304–14.
16 Cohen DW, Simak R, Fair WR, Melamed J, Scher HI, Cordon-Cardo C. Expression of transforming growth factor-alpha and the epidermal growth factor receptor in human prostate tissues. *J Urol* 1994; **152**: 2120–4.
17 Culig Z, Hobisch A, Cronauer MV *et al.* Regulation of prostatic growth and function by peptide growth factors. *Prostate* 1996; **28**: 392–405.
18 Connolly JM, Rose DP. Angiogenesis in two human prostate cancer cell lines with differing metastatic potential when growing as solid tumors in nude mice. *J Urol* 1998; **160**: 932–6.
19 Bonaccorsi L, Carloni V, Muratori M *et al.* EGF receptor (EGFR) signaling promoting invasion is disrupted in androgen-sensitive prostate cancer cells by an interaction between EGFR and androgen receptor (AR). *Int J Cancer* 2004; **112**: 78–86.
20 Huggins C. Endocrine-induced regression of cancers. *Science* 1967; **156**: 1050–4.
21 Feldman BJ, Feldman D. The development of androgen-independent prostate cancer. *Nat Rev Cancer* 2001; **1**: 34–45.
22 Shah RB, Ghosh D, Elder JT. Epidermal growth factor receptor (ErbB1) expression in prostate cancer progression: correlation with androgen independence. *Prostate* 2006; **66**: 1437–44.
23 Legrier ME, de Pinieux G, Boye K *et al.* Mucinous differentiation features associated with hormonal escape in a human prostate cancer xenograft. *Br J Cancer* 2004; **90**: 720–7.
24 McWilliam LJ, Manson C, George NJ. Neuroendocrine differentiation and prognosis in prostatic adenocarcinoma. *Br J Urol* 1997; **80**: 287–90.
25 di Sant'Agnese PA, Cockett AT. Neuroendocrine differentiation in prostatic malignancy. *Cancer* 1996; **78**: 357–61.

Protection of telomeres 1 protein levels are associated with telomere length in gastric cancer

KIYOMU FUJII[1], TOMONORI SASAHIRA[1], YUKIKO MORIWAKA[1], NAOHIDE OUE[2], WATARU YASUI[2] and HIROKI KUNIYASU[1]

[1]Department of Molecular Pathology, Nara Medical University School of Medicine, Kashihara, Nara; [2]Department of Molecular Pathology, Hiroshima University Graduate School, Hiroshima, Japan

Received December 3, 2007; Accepted January 25, 2008

Abstract. Protection of telomeres 1 (Pot1) is a telomere-associated protein, which binds to the single-stranded DNA extensions of telomeres and regulates telomere length. Pot1 production was examined and compared with telomere length in gastric cancer. Pot1 production and telomere lengths were assessed in 5 human gastric cancer cell lines by immunoblotting and Southern blotting, respectively. Pot1 intracellular localization was examined with protein fractionation. Pot1 index and telomere volume were examined in human gastric mucosa and cancer by immunohistochemistry and *in situ* hybridization. Pot1 protein levels, which were lower than those in the lymphocytes of healthy persons, were significantly correlated with telomere length in gastric cancer cells (P=0.0167). Pot1 protein was mainly detected in the nuclear fraction and increased by G2/M blocking with nocodazole in MKN28 cells. Pot1 indexes were correlated with telomere volumes in gastric cancers (P<0.0001). Pot1 index was decreased in gastric epithelia distant from cancer (84±14%), in peritumoral epithelia (72±24%), and in stage I-II (39±14%) and stage III-IV (23±14%) gastric cancers (P<0.0001). Pot1 index was lower in stage III-IV than in stage I-II gastric cancers (P<0.05). Pot1-low cases showed advanced cancer invasion (P<0.05). Thus, Pot1 production was closely associated with telomere length in gastric mucosa and cancers. Pot1 might be a good *in situ* marker for the examination of cell-specific telomere length.

Correspondence to: Dr Hiroki Kuniyasu, Department of Molecular Pathology, Nara Medical University School of Medicine, 840 Shijo-cho, Kashihara, Nara 634-8521, Japan
E-mail: cooninh@zb4.so-net.ne.jp

Abbreviations: Pot1, protection of telomeres 1; *H. pylori*, *Helicobactor pylori*; hTERT, human telomerase reverse transcriptase; hTR, human telomerase RNA

Key words: telomere, telomere length, gastric cancer

Introduction

The telomere is a repetitive sequence of TTAGGG, located at the chromosomal ends. Telomere reduction is closely associated with aging, inflammatory, regenerative, and carcinogenic processes in various cells and tissues (1-4). In our previous study, the telomeres were reduced at various levels in the epithelial cells in the gastric mucosa in comparison with mucosa-infiltrating lymphocytes, smooth muscle cells, and endothelial cells (5). In particular, telomeres were notably reduced in the intestinal metaplasia epithelium with *Helicobactor pylori* (*H. pylori*) infection, in which human telomerase reverse transcriptase (hTERT) protein expression was found in association with a marked reduction of telomeres (5). Human telomerase RNA (hTR) is also expressed in relation to *H. pylori* infection in intestinal metaplasia (6). Telomere shortening, expression of hTR and hTERT, and telomerase activation are commonly found in gastric cancer (7). From these findings, we can hypothesize that the earliest stage of gastric cancer development might be associated with telomere reduction.

Protection of telomeres 1 (Pot1) is a telomere-associated protein, which is isolated from ciliated protozoa (8). Pot1 binds to the single-stranded G-rich DNA extensions of the telomere with its N-terminal DNA-binding domain, and protects the chromosomal ends from chromosomal instability (8,9). Pot1 localizes to telomeres in the interphase nuclei of human cells (10) or during periods of the cell cycle when t-loops are thought to be present (11). Pot1 regulates the telomere length: TRF1 regulates Pot1-binding with single-strand telomere DNA and Pot1 controls telomerase-mediated telomere elongation by *cis*-inhibition of telomerase (12-14). We previously reported that 3' telomeric overhang signals decreased in accordance with decreases in Pot1 expression levels and telomere shortening (15). In gastric cancer, the mRNA expression of *Pot1* is associated with telomere length and cancer progression (15,16).

In the present study, we demonstrated that Pot1 protein level were closely associated with telomere length in gastric cancer cells and were decreased in gastric mucosal epithelia and gastric cancer.

Materials and methods

Cell culture. Human gastric cancer cell lines, MKN28, MKN45, TMK1, HSC39 were routinely maintained in RPMI-1640 (Sigma Chemical Co., St. Louis, MO) containing 10% fetal bovine serum (FBS, Sigma Chemical Co.) under the conditions of 5% CO_2 in air at 37˚C. Lymphocytes used as the control for Pot1 expression and telomere length were obtained from the peripheral blood of 3 healthy volunteers, whose mean age was matched to gastric cancer patients (2 men and 1 woman, 59, 61 and 65 years of age, mean: 61.7 years). Lymphocytes ($1x10^7$) were collected from each volunteer and mixed for examination.

Patients and tumor specimens. Twenty-four gastric cancer patients (12 stage I-II and 12 stage III-IV cases; 15 men and 9 women, 52-78 years of age, mean: 63.6 years) were randomly selected from the cases analyzed in our previous study (5). A formalin-fixed, paraffin-embedded surgical specimen containing the deepest invasion site was chosen from the tissue specimens of each patient. Tumor staging and histopathological grading were classified according to the UICC TNM classification system (17) and Lauren's classification (18). Their medical records and prognostic follow-up data were obtained from the patient database maintained by the hospital.

Immunoblot analysis. Whole-cell lysates were prepared as described previously (19). Fifty-microgram lysates were subjected to immunoblot analysis in 12.5% sodium dodecyl sulfate-polyacrylamide gels followed by electrotransfer onto nitrocellulose filters. The filters were incubated with primary antibodies and then with peroxidase-conjugated IgG antibodies (Medical and Biological Laboratories, Nagoya, Japan). An α-tubulin antibody was used to assess the levels of protein loaded per lane (Oncogene Research Products, Cambridge, MA). The immune complex was visualized by a CSA system (Dako, Carpinteria, CA). Anti-protection of telomeres 1 (Pot1) antibodies (Santa Cruz Biotechnology, Santa Cruz, CA) were used as the primary antibodies.

Southern blot analysis. High molecular weight genomic DNA was extracted with a DNA Extraction Kit (Stratagene Cloning System, La Jolla, CA). DNA was digested with *Hin*fI (Takara Biomedicals, Tokyo, Japan), electrophoresed on 0.8% agarose-TAE gels, and blotted onto nitrocellulose filters. The filters were hybridized with $(TTAGGG)_4$ probe labeled with biotin at the 5' end (Sigma Genosys, Ishikari, Japan), which was detected with peroxidase-conjugated avidine and visualized with an ECL system (both from Dako). We estimated the telomere length as the peak signal using Kodak Digital Science 1D software (Eastman Kodak Company, New Haven, CT). DNA ladders (2.5-kb and 1-kb) (Takara Biomedicals) were used for the measurement of peak telomere length.

Intracellular localization. MKN28 cells grown in culture dishes were treated with or without nocodazole (Alexis Biochemicals, San Diego, CA; 100 ng/ml for 24 h). The cells were re-suspended in 500 μl of STKM buffer [50 mM Tris HCl pH 7.5, 25 mM KCl, 5 mM $MgCl_2$, 0.25 M sucrose, 10 μg/ml leupeptin, 50 μg/ml phenylmethylsulfonyl fluoride (PMSF)] and stroked 150 times in dounce pestles. After a 1,000 x g, 5-min centrifugation, the supernatant was centrifugated at 40,000 x g for 1 h at 4˚C to separate the membrane fraction (pellet) and the cytosol fraction (supernatant). The pellet was re-suspended by 1000 x g-spin into hypertonic buffer (25 mM Tris HCl pH 7.8, 10 mM KCl, 5 mM $MgCl_2$, 10 μg/ml leupeptin, 50 μg/ml PMSF) and incubated for 5 min on ice followed by the addition of the same amount of 2X STKM buffer. After centrifugation at 40,000 x g for 1 h at 4˚C, the pellet was re-suspended with nuclear extraction buffer (20 mM Hepes pH 7.9, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT, 25% glycerol) as the nuclear fraction. Loading amounts were monitored by Coomassie blue staining of the same amount of the proteins dot-blotted onto nitrocellulose membrane.

Immunohistochemistry. Immunohistochemistry was performed as previously described (20). Sections (4 μm thick) of each specimen were mounted on ProbeOn slides for ISH (Fisher Scientific, Pittsburgh, PA). An immunoperoxidase technique was used following antigen retrieval with microwave (1,000 W) treatment for 10 min three times in citrate buffer (pH 6.0). After blocking endogenous peroxidase activity with 3% H_2O_2-methanol for 15 min, the specimens were rinsed with phosphate-buffered saline (PBS). Anti-Pot1 antibody (Santa Cruz Biotechnology) diluted to 0.5 μg/ml was used as the primary antibody. After a 2-h incubation at room temperature, the slides were rinsed with PBS and incubated at room temperature for 1 h with a secondary antibody conjugated to peroxidase (1:200) (anti-goat IgG antibody, Medical and Biological Laboratories Co. Ltd). After being rinsed with PBS, all specimens were color-developed with diaminobenzidine (DAB) solution (Dako). Immunostaining of all specimens was performed simultaneously to ensure the same antibody reaction and DAB exposure conditions. Nuclear immunoreactivity was judged as positive. Pot1 positivity was examined as follows: 1,000 nuclei were observed in Pot1 immunostained slides under x200 magnification microscopy. The Pot1 index was designated as a percentage. Data of telomere volume in gastric cancer cases were derived from our previous study (5).

Telomere volume. Telomere volume was determined by fluorescent *in situ* hybridization (FISH). Sections (10 μm thick) of each specimen were mounted on ProbeOn slides (Fisher Scientific, Pittsburgh, PA) for ISH. The telomere repeat probe $(TTAGGG)_4$ was labeled with fluorescein isothiocyanate (FITC) on the 3'-tail (EspecOligo Service, Tsukuba, Japan) (5). The probe was diluted to 20 μg/ml by probe hybridization solution [50% formamide (Sigma Chemical Co.), 0.5 M NaCl, 5% polyethylene glycol 8000 [Sigma Chemical Co.)]. The specimens were dewaxed and dehydrated with xylene and 100% ethanol. They were then hydrated in Tris-buffered saline (Sigma Chemical Co.) and digested with 0.2% pepsin-2 M HCl (Dako) for 1 h at 37˚C, and then subjected to RNase A (10 μg/μl, Takara Biomedicals) treatment at 37˚C for 10 min. Specimens were then heated with probe solution at 100˚C for 5 min, cooled to 4˚C for

Table I. Pot1 protein levels and telomere length in gastric carcinoma cell lines.

Cell line	Pot1 level (%)[a]	Telomere length (kb)
MKN28	47	4.7
MKN45	42	3.8
TMK1	31	3.1
HSC39	25	2.8
Lymphocytes	100	9.9

[a]Signal density of immunoblotting was standardized with that in lymphocytes, which was set to 100%.

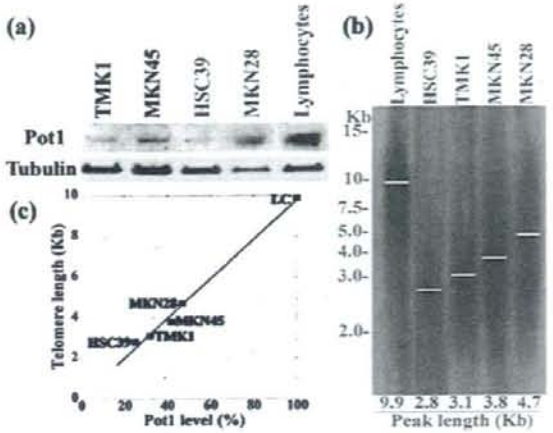

Figure 1. Pot1 production and telomere length in human gastric cancer cell lines. (a) Pot1 production was detected by immunoblotting in gastric cancer cells. Lymphocytes were examined as the normal control. Tubulin was detected as the loading control. (b) Telomere length was examined by Southern blot analysis in gastric cancer cells. Peak length is designated by a white line in each lane. (c) Correlation between Pot1 production and telomere length was examined. Pot1 production was designated as a percentage of that in lymphocytes (LC).

15 min, and then maintained at 37˚C for 2 h. The specimens were rinsed five times with 1X standard sodium chloride/sodium citrate at 45˚C.

Specimens hybridized with the $(TTAGGG)_4$ probe were examined with 520-nm light for FITC. Specimen images were stored on a computer and processed (Fig. 1a). Briefly, 100 nuclei were identified from the images, and the hybridization signals were scanned as inverted gray-scale images by means of NIH Image software (National Institute for Health, Bethesda, MD). The mean signal density and mean nuclear area of the 100 nuclei were calculated. The mean density was divided by the mean nuclear area (μm^2) to adjust for differences in the DNA amounts of the identified nuclei. The resulting value was considered to be representative of the telomere density of the tissue and was termed telomere volume to distinguish it from the telomere length determined by Southern blot analysis (5).

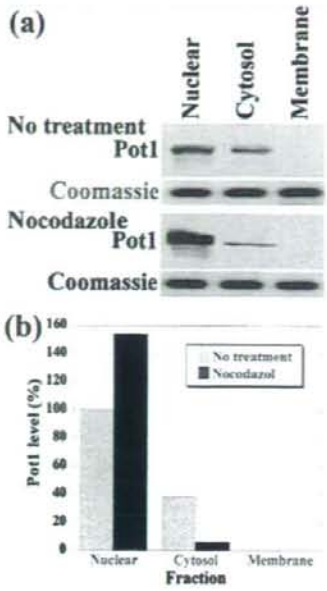

Figure 2. Intracellular localization of Pot1 was examined using fractionated protein extraction of MKN28 gastric cancer cells. (a) Pot1 was detected in each protein fraction by immunoblotting. Loading amounts were monitored by Coomassie blue staining of the same amount of protein dot-blotted onto nitrocellulose membrane. To increase G2/M phase cells, MKN28 cells were treated with nocodazole (100 ng/ml, 24 hrs). (b) Relative Pot1 protein levels. Pot1 level in the nuclear fraction of untreated cells was set to 100%.

Statistical analysis. Statistical significance was examined by the Spearman's Rank test, the two-tailed, unpaired Mann-Whitney U test, and Fisher's exact test using InStat software (Graphpad Software, Los Angeles, CA). Statistical significance was defined as a two-sided P-value of <0.05.

Results

Pot1 production and telomere length in gastric cancer cells (Table I). We first examined Pot1 protein production in gastric carcinoma cells (Fig. 1a). Pot1 expression in gastric cancer cells was decreased in comparison with human lymphocytes. The telomere lengths of these cells were next examined (Fig. 1b). The gastric cancer cell lines showed reduced telomere length in comparison with the lymphocytes of healthy persons. We compared Pot1 expression with telomere length in the cells (Fig. 1c). Pot1 production was significantly correlated with telomere length (Spearman R=0.9998, P=0.0167).

Intracellular localization of Pot1 in gastric cancer cells. Next, the intracellular localization of Pot1 was examined in MKN28 cells (Fig. 2). In untreated MKN28 cells, Pot1 protein was detected in the nuclear and cytosol fraction. Pot1 was not detected in the membrane fractions. When MKN28 cells were treated with nocodazole to stop the cell cycle at G2/M phase, Pot1 was increased in the nuclear fraction, whereas it was decreased in the cytosol fraction.

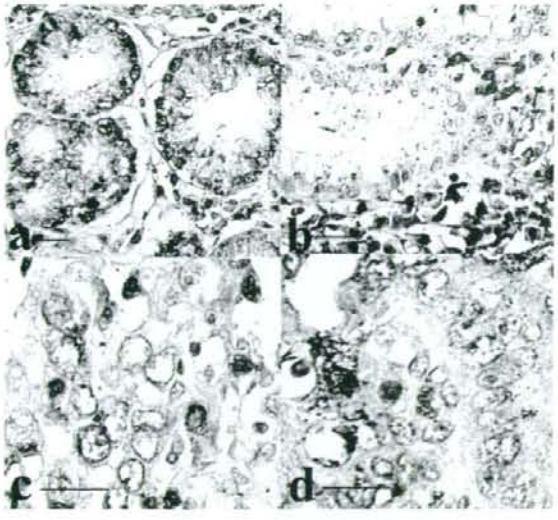

Figure 3. Immunohistochemistry of Pot1 in gastric mucosa and cancer. (a) Gastric mucosa distant from the tumor. (b) Peritumoral mucosa. (c) Early-stage gastric cancer. (d) Advanced gastric cancer. Nuclear immunoreactivity was judged as positive. Bar, 50 μm.

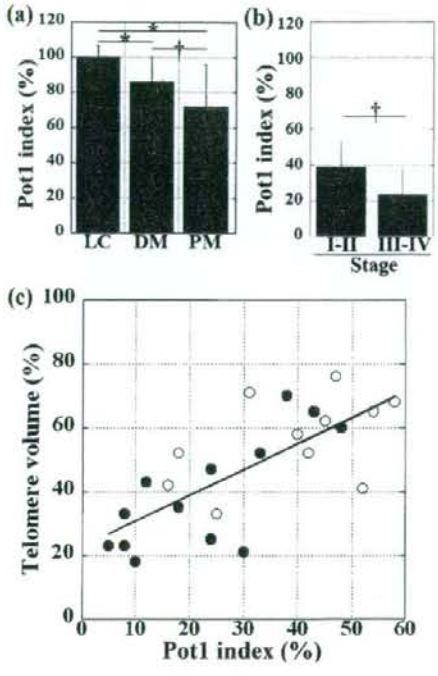

Figure 4. Pot1 positivity and telomere volume in human gastric mucosa and cancer. (a) Pot1 positivity in: lymphocytes (LC), mucosa distant from the tumor (DM), peritumoral mucosa (PM). (b) Pot1 positivity in stage I-II and stage III-IV cases. *P<0.0005, †P<0.05; error bar, SD. (c) Correlation between Pot1 positivity and telomere volume in gastric carcinomas. Telomere volume was designated as a percentage of that in lymphocytes. ○, stage I-II; ●, stage III-IV. Spearman R=0.7137, P<0.0001.

Table II. Pot1 positivity and telomere volume in human gastric cancer.

Case no.	Pathological stage	Pot1 positivity (%)	Telomere volume (%)
1	I	58	68
2	I	54	65
3	I	52	41
4	I	47	76
5	I	45	62
6	I	42	52
7	I	40	58
8	I	31	71
9	II	25	33
10	II	18	52
11	II	16	42
12	III	48	60
13	III	43	65
14	III	38	70
15	III	33	52
16	III	30	21
17	III	24	25
18	III	24	47
19	III	18	35
20	IV	12	43
21	IV	10	18
22	III	8	33
23	IV	8	23
24	IV	5	23

Immunohistochemistry of Pot1 in gastric mucosa and cancer. The Pot1 index was examined in non-cancerous gastric mucosa by immunohistochemistry (Fig. 3a and b). In mucosal epithelial cells distant from the tumors, the Pot1 index was reduced to 84±14% of that in the infiltrating lymphocytes (P=0.0003, Fig. 3a and Fig. 4a). In contrast, the Pot1 index was more reduced in the epithelial cells in the peritumoral mucosa (within 5 mm from the tumoral periphery), to 72±24% of that in the infiltrating lymphocytes (P=0.0001, Fig. 3b and Fig. 4a). The Pot1 index was significantly lower in the epithelial cells in the peritumoral mucosa than those in the mucosa distant from the tumors (P<0.05).

Next, the Pot1 index was examined in cancer tissues by immunohistochemistry (Fig. 3c and d, Table II). In Stage I-II gastric cancer, the Pot1 index was reduced to 39±14% of that in the infiltrating lymphocytes (P<0.0001, Fig. 4b). It was more reduced in Stage III-IV gastric cancer, to 23±14% of that in the infiltrating lymphocytes (P<0.0001, Fig. 4b). Pot1 positivity was significantly lower in stage III-IV cancers than in stage I-II cancers (P<0.05). We compared Pot1 positivity and telomere volume in the cases (Fig. 4c). The two parameters were significantly correlated in a linear regression manner (Spearman R=0.7137, P<0.0001).

Finally, we compared the Pot1 index and clinicopathological parameters (Table III). The 24 gastric cancer cases were divided into two groups by Pot1 index: Pot1-low (12 cases with lower Pot1 index) and Pot1-high (12 cases with higher

Table III. Comparison of Pot1 index with clinicopathological parameters in 24 gastric cancer cases.

	Pot1 high (n=12)	Pot1 low (n=12)	P
Pot1 index (%)	31-58 (median 44) 44.3±8.2	5-30 (median 17) 16.5±8.0	<0.0001
Telomere volume (%)[a]	41-76 (median 63.5) 61.7±9.8	18-52 (median 33) 32.9±11.2	0.0001
Age (years)	48-81 (median 64) 64.0±9.9	43-78 (median 66.5) 63.8±11.3	NS
Gender (M:F)	7:5	7:5	NS
Pathological stage[b] (I-II vs III-IV)	7:4	4:8	NS
T-factor[b] (T1-T2:T3)	11:1	5:7	<0.0500
N-factor[b] (N0:N1-N2)	6:6	4:8	NS
M-factor[b] (M0:M1)	12:0	9:3	NS
Histology[c] (intestinal vs diffuse type)	8:4	5:7	NS

[a]Telomere volume was examined by fluorescent *in situ* hybridization using $(TTAGGG)_4$ probe. [b]Pathological staging, T-factor, N-factor, and M-factor (all positive cases showed peritoneal dissemination in this study) were defined according to the TNM classification system (17). [c]Histological classification was done according to Lauren's classification (18). NS, not significant.

Pot1 index). The Pot1 indexes of Pot1-low and Pot1-high groups were 16.5±8.0% and 44.3±8.2%, respectively (P<0.0001). The telomere volumes in the Pot1-high group were higher (61.7±9.8%) than those in the Pot1-low group (32.9±11.2%) (P=0.0001).There was no difference between the two groups in age, gender, pathological stage, N-factor (nodal metastasis), M-factor (peritoneal dissemination), and histological types; however, T-factor (invasive depth) was advanced in Pot1-low. The number of T3 cases (invading the serosa) was 7 out of 12 (58%) in the Pot1-low group, and, in contrast, 1 out of 12 (8%) in the Pot1-high group (P<0.05).

Discussion

We confirmed that Pot1 expression was closely correlated with telomere length in gastric cancer cells and human lymphocytes. Pot1 was localized to the t-loops of telomeres in the interphase nuclei of human cells (10,11). We confirmed Pot1 intracellular localization by the fractionated cellular proteins. Pot1 was found in the nuclear and cytosol fractions in MKN28 cells cultured in regular conditions, whereas it was mainly found in the nuclear fraction in nocodazole-treated MKN28 cells, whose cell cycle was stopped at G2/M. This observation suggests that Pot1 might be recruited into the nuclei in the G2/M phase to bind to telomeres. The Pot1 localization detected by immunohistochemistry was found in whole nuclei or the nuclear periphery. This is similar to that of the telomeres in our previous observations (5), which suggests the co-localization of Pot1 and telomeres.

Immunohistochemical examination of the Pot1 index in gastric mucosa and cancer showed that it was reduced in gastric epithelia and gastric cancer in comparison with infiltrated lymphocytes. Cancer cells showed a lower Pot1 index than gastric epithelia. To confirm the relationship of Pot1 positivity with telomere reduction, we compared Pot1 positivity and our previous data for telomere volume in the same cases (5). The results showed significant correlation between the two parameters and suggest that Pot1 expression corresponds to telomere length in the cells or tissues.

Notably, the Pot1 index was significantly lower in the gastric epithelium adjacent to the tumoral periphery than in epithelia distant from the tumor. Two possibilities are proposed. One is that it was affected by the tumor: the mucosa adjacent to colon cancer showed hyperplastic, proliferative, and angiogenic properties, which were responses to the growth factors and cytokines produced by cancer cells (21). Another is that it is a precancerous change leading to gastric cancer. Telomere reduction is a strong stimulus for the reactivation of telomerase in somatic cells and is associated with carcinogenic processes in many cancers (3,22). Pot1 decrease might be associated with telomerase activation and hence with the transformation of gastric epithelia. In further study, we will endeavour to detect early-stage gastric cancer by examining Pot1 expression as a probe.

Acknowledgments

This work was supported in part by a Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science, Japan.

References

1. Harley CB: Human aging and telomeres. Ciba Found Symp 211: 129-139, 1997.

2. Greider CW and Blackburn EH: Telomeres, telomerase and cancer. Sci Am 274: 92-97, 1996.
3. Holt SE, Wright WE and Shay JW: Multiple pathways for the regulation of telomerase activity. Eur J Cancer 33: 761-766, 1997.
4. Weng NP, Palmer LD, Levine BL, Lane HC, June CH and Hodes RJ: Tales of tails: regulation of telomere length and telomerase activity during lymphocyte development, differentiation, activation, and aging. Immunol Rev 160: 43-54, 1997.
5. Kuniyasu H, Kitadai Y, Mieno H and Yasui W: *Helicobactor pylori* infection is closely associated with telomere reduction in gastric mucosa. Oncology 65: 275-282, 2003.
6. Kuniyasu H, Domen T, Hamamoto T, Yokozaki H, Yasui W, Tahara H and Tahara E: Expression of human telomerase RNA is an early event of stomach carcinogenesis. Jpn J Cancer Res 88: 103-107, 1997.
7. Tahara H, Kuniyasu H, Yokozaki H, Yasui W, Shay JW, Ide T and Tahara E: Telomerase activity in preneoplastic and neoplastic gastric and colorectal lesions. Clin Cancer Res 1: 1245-1251, 1995.
8. Baumann P and Cech TR: Pot1, the putative telomere end-binding protein in fission yeast and humans. Science 292: 1171-1175, 2001.
9. Lei M, Baumann P and Cech TR: Cooperative binding of single-stranded telomeric DNA by the Pot1 protein of *Schizosaccharomyces pombe*. Biochemistry 41: 14560-14568, 2002.
10. Baumann P, Podell E and Cech TR: Human Pot1 (protection of telomeres) protein: cytolocalization, gene structure, and alternative splicing. Mol Cell Biol 22: 8079-8087, 2002.
11. Wei C and Price CM: Cell cycle localization, dimerization, and binding domain architecture of the telomere protein cPot1. Mol Cell Biol 24: 2091-2102, 2004.
12. Loayza D and De Lange T: POT1 as a terminal transducer of TRF1 telomere length control. Nature 423: 1013-1018, 2003.
13. Colgin LM, Baran K, Baumann P, Cech TR and Reddel RR: Human POT1 facilitates telomere elongation by telomerase. Curr Biol 13: 942-946, 2003.
14. Ye JZ, Hockemeyer D, Krutchinsky AN, Loayza D, Hooper SM, Chait BT and De Lange T: POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. Genes Dev 18: 1649-1654, 2004.
15. Kondo T, Oue N, Yoshida K, Mitani Y, Naka K, Nakayama H and Yasui W: Expression of POT1 is associated with tumor stage and telomere length in gastric carcinoma. Cancer Res 64: 523-529, 2004.
16. Lin X, Gu J, Lu C, Spitz MR and Wu X: Expression of telomere-associated genes as prognostic markers for overall survival in patients with non-small cell lung cancer. Clin Cancer Res 12: 5720-5725, 2006.
17. Sobin LH and Wittekind C (eds): UICC TNM Classification of Malignant Tumours. 6th edition, John Wiley & Sons Inc., New York, 2003.
18. 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 64: 31-49, 1965.
19. Kuniyasu H, Yasui W, Kitahara K, Naka K, Yokozaki H, Akama Y, Hamamoto T, Tahara H and Tahara E: Growth inhibitory effect of interferon-beta is associated with the induction of cyclin-dependent kinase inhibitor p27Kip1 in a human gastric carcinoma cell line. Cell Growth Differ 8: 47-52, 1997.
20. Kuniyasu H, Oue N, Wakikawa A, Shigeishi H, Matsutani N, Kuraoka K, Ito R, Yokozaki H and Yasui W: Expression of receptors for advanced glycation end-products (RAGE) is closely associated with the invasive and metastatic activity of gastric cancer. J Pathol 196: 163-170, 2002.
21. Kuniyasu H, Yasui W, Shinohara H, Yano S, Ellis LM, Wilson MR, Bucana CD, Rikita T, Tahara E and Fidler IJ: Induction of angiogenesis by hyperplastic colonic mucosa adjacent to colon cancer. Am J Pathol 157: 1523-1535, 2000.
22. Greider CW: Telomere length regulation. Annu Rev Biochem 65: 337-365, 1996.

nature genetics

Genetic variation in *PSCA* is associated with susceptibility to diffuse-type gastric cancer

The Study Group of Millennium Genome Project for Cancer*

Gastric cancer is classified into intestinal and diffuse types, the latter including a highly malignant form, linitis plastica. A two-stage genome-wide association study (stage 1: 85,576 SNPs on 188 cases and 752 references; stage 2: 2,753 SNPs on 749 cases and 750 controls) in Japan identified a significant association between an intronic SNP (rs2976392) in *PSCA* (prostate stem cell antigen) and diffuse-type gastric cancer (allele-specific odds ratio (OR) = 1.62, 95% CI = 1.38–1.89, $P = 1.11 \times 10^{-9}$). The association was far less significant in intestinal-type gastric cancer. We found that *PSCA* is expressed in differentiating gastric epithelial cells, has a cell-proliferation inhibition activity *in vitro* and is frequently silenced in gastric cancer. Substitution of the C allele with the risk allele T at a SNP in the first exon (rs2294008, which has $r^2 = 0.995$, $D' = 0.999$ with rs2976392) reduces transcriptional activity of an upstream fragment of the gene. The same risk allele was also significantly associated with diffuse-type gastric cancer in 457 cases and 390 controls in Korea (allele-specific OR = 1.90, 95% CI = 1.56–2.33, $P = 8.01 \times 10^{-11}$). The polymorphism of the *PSCA* gene, which is possibly involved in regulating gastric epithelial-cell proliferation, influences susceptibility to diffuse-type gastric cancer.

Gastric cancer, the fourth most common cancer and the second leading cause of cancer death in the world, has an incidence and mortality particularly high in Japan and Korea[1,2]. Histopathological research has long suggested that gastric cancer is not a single disease and recognizes two major categories: intestinal and diffuse[1,3,4]. (**Supplementary Table 1** online). Besides their morphological differences, the two types may also be distinct in their pathogenesis.

The intestinal type predominates in high-risk geographic areas such as East Asia, showing a correlation with the prevalence in the region of *Helicobacter pylori* infection among the elderly. Typically, intestinal-type gastric cancer arises through a sequence of events: *H. pylori*–induced persistent inflammation, hypochlorhydria, atrophic gastritis, intestinal metaplasia and intestinal-type adenocarcinoma.

The diffuse type, however, is more uniformly distributed geographically, is apparently unrelated to *H. pylori* prevalence[1] and typically develops from *H. pylori*–free, morphologically normal gastric mucosa without atrophic gastritis or intestinal metaplasia. Unlike the decreasing incidence of the intestinal type, the prevalence of the diffuse type is reportedly increasing worldwide[1].

The intestinal and diffuse types show histological characteristics of well- and poorly differentiated adenocarcinomas, respectively, yet both can coexist in the same gastric cancer tissue specimen, which suggests that a degradation of a well-differentiated glandular architecture to a poorly differentiated morphology may occur in some cases during cancer development[5]. Such divergence may be mediated by somatic loss of *CDH1* function[6] and illustrates another example of the complexity in understanding the initial carcinogenesis mechanisms through the morphological features of the advanced stage of the cancer. Nevertheless, a *de novo* diffuse-type gastric cancer is believed to exist and develop from stem cells or precursors for gastric epithelial cells in the background of relatively normal gastric mucosa[4,7].

Notably, advanced diffuse-type gastric cancer includes a distinct form of gastric cancer with an extremely poor prognosis: linitis plastica[8] (**Supplementary Fig. 1** online). This cancer accounts for about 10% of all gastric cancer, and its 5-year survival rate is around 10–20% in Japan[9]. Unlike the intestinal types, the diffuse types, including linitis plastica, occur almost equally among males and females, and linitis plastica shows a relative predominance in younger individuals[8,10,11].

Close to 100 pedigrees with an autosomal-dominant hereditary form of diffuse-type gastric cancer showed germline mutations of the *CDH1* gene. However, the pathogenic *CDH1* mutation seems to be rare among Japanese individuals with familial gastric cancer[12]. A large-scale twin study in northern Europe, where *H. pylori* prevalence is relatively low, estimated that heritability accounts for 28% of the variation in susceptibility to gastric cancer[13], but data specific to diffuse- or intestinal-type cancers were not available. Despite suggestions that intestinal- and diffuse-type gastric cancers develop through different carcinogenic pathways and that genetic background is more important in the latter[14], information on the genetic susceptibility factors specific to the diffuse type is scarce[15]. This study aims to identify genetic factors influencing predisposition to the sporadic form of diffuse-type gastric cancer, which may constitute a distinct disease entity and should be analyzed separately from intestinal-type gastric cancer.

*A full list of the authors appears at the end of this paper.

Received 6 December 2007; accepted 26 March 2008; published online 18 May 2008; doi:10.1038/ng.152

Table 1 Ten SNPs with the smallest *P* values for allele model odds ratio in the stage 2 screening

SNP ID				Stage 2 screening				Stage 1 screening		
IMS-JST	rs number	Chr.	Gene	OR	95% CI	Fisher	Permutation	OR	95% CI	Fisher
089945	rs2976392	8q24.3	*PSCA*	1.62	1.38–1.89	1.1×10^{-9a}	2.3×10^{-5}	1.89	1.45–2.46	1.2×10^{-6}
003203	rs2075570	1q22	*MTX1*	1.65	1.34–2.02	9.2×10^{-7a}	0.0020	1.70	1.20–2.40	0.0025
005799	rs2070803	1q22	*TRIM46*	1.62	1.33–1.98	1.2×10^{-6a}	0.0026	1.68	1.20–2.35	0.0022
163807	rs3804775	3q12.3	*PCNP*	1.32	1.14–1.54	2.4×10^{-4}	0.31	0.74	0.59–0.93	0.010
136603	rs301451	9p24.2	*C9orf68*	0.67	0.54–0.84	2.5×10^{-4}	0.31	1.47	1.03–2.10	0.032
089943	rs2976391	8q24.3	*PSCA*	1.42	1.17–1.72	3.2×10^{-4}	0.37	1.90	1.34–2.70	1.7×10^{-4}
047367	rs2244163	8q24.3	*LY6K*	1.41	1.16–1.72	4.3×10^{-4}	0.45	1.90	1.33–2.73	2.9×10^{-4}
060142	rs2303474	3q12.3	*ZBTB11*	1.30	1.12–1.51	5.4×10^{-4}	0.52	0.70	0.55–0.89	0.0032
089950	rs2585174	8q24.3	*LY6K*	1.31	1.12–1.53	6.5×10^{-4}	0.58	1.62	1.25–2.09	2.2×10^{-4}
044453	rs2291905	3q13.2	–	0.72	0.59–0.89	0.0017	0.86	1.53	1.07–2.19	0.020

Listed are the gene nearest to the SNP, the allelic odds ratio (major allele/minor allele), the 95% confidence interval and *P* values obtained by Fisher's exact test or by permutation test. Please note that SNPs were selected in the first screening not only by allele model, but also by considering dominant and recessive models. Twenty-six pairs and a trio in the total 940 samples showed high concordance in the genotype data. See Methods for details. [a]Statistically significant after Bonferroni adjustment ($P < 1.8 \times 10^{-5}$).

A genome-wide association study was conducted on the basis of the JSNP database (see URLs section in Methods), which catalogs the gene-centric SNPs ascertained from Japanese individuals[16,17]. This study is a part of a five-disease collaborative national project in Japan, in which two-stage genome-wide association studies on 100,000 JSNPs were simultaneously done for Alzheimer's disease, gastric cancer, type 2 diabetes, hypertension and asthma[18,19].

The genome scan identified a SNP in *PSCA* to be associated with the diffuse type of gastric cancer even after Bonferroni correction of multiple testing. PSCA was first identified as a prostate-specific antigen overexpressed in prostate cancers[20,21], but it was also expressed in the bladder, esophagus and stomach[22].

To fortify the observed statistical association, we carried out several analyses to assess the functional significance of the gene. Taken together, our findings suggest that *PSCA* has a role in proliferation of differentiating gastric epithelial cells, and that variation in the regulatory region of *PSCA* influences its expression, which might be translated to a predisposition to the diffuse type of gastric cancer.

RESULTS

SNPs in *PSCA* highly related to diffuse-type gastric cancer

The design, positive predictive value and sensitivity of our JSNP genome scan have been described previously[23]. In the first stage of the scan, genotype data at 85,576 SNPs on 188 cases and 752 controls passed a quality check. The 752 control samples were a mixture of Japanese individuals ascertained for one of four diseases—Alzheimer's disease, type 2 diabetes, hypertension and asthma—in the joint JSNP genome scan project[18] (see Methods). For the second stage of screening, 2,880 SNPs were selected by criteria in which several factors such as odds ratios, *P* values of Fisher's exact test, allele frequency and linkage disequilibrium were taken into consideration (see Methods). The 2,880 SNPs were genotyped on separate stage 2 samples (749 cases and 750 healthy volunteer controls), and valid genotype data were obtained for 2,753 SNPs. **Table 1** lists ten SNPs showing the least *P* values for allelic contingency tables in the second stage. Of the ten SNPs, four are located at 8q24.3, where *PSCA* and *LY6K* map (**Table 1** and **Fig. 1a**).

Linkage disequilibrium (LD) analysis of the first-screening data showed that the *PSCA* and *LY6K* genes are in the same LD block. However, our separate dataset on 109,365 SNPs genotyped by Illumina Human-1 BeadChip on 379 Japanese controls offered denser SNP

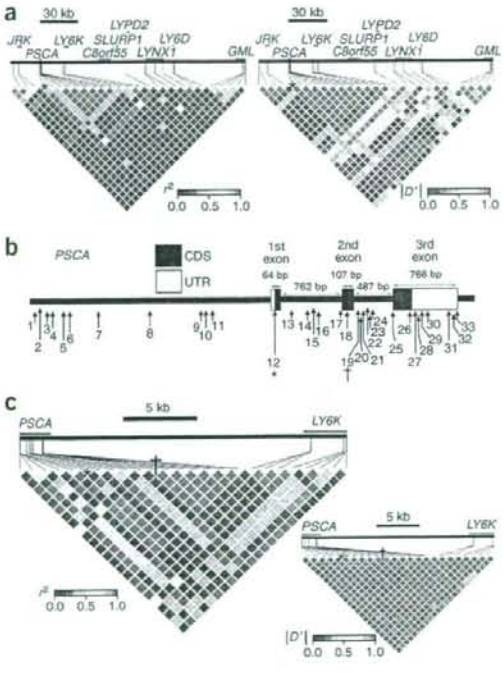

Figure 1 LD analyses of the SNPs in the genomic region around *PSCA*. (**a**) LD maps are shown by two parameters, r^2 and *D'* for 379 Japanese control individuals. The asterisk (*) marks SNP rs2294008. The other SNP, rs2976392, was not on the Illumina Human-1 BeadChip used for drawing the LD map. (**b**) SNPs (arrows) of *PSCA* identified by resequencing. The numbers correspond to those shown in **Table 2**. rs2294008 can modulate the transcriptional activity of the *PSCA* promoter *in vitro*. rs2976392 (†) was identified first by the second screening of this genome scan and is in strong LD with rs2294008. CDS, coding sequence. (**c**) Fine LD maps constructed by typing 750 Japanese control individuals for the 17 common SNPs (numbers 1, 4, 6, 10, 11, 12, 15–19, 22, 26, 27, 29, 30 and 32 in **Table 2**).

Table 2 SNPs in *PSCA* identified by resequencing and their association with diffuse-type gastric cancer in the stage 2 case and control subjects

				Resequencing		Genotyping on 749 cases and 750 controls				
SNP No.	Nucleotide position	rs number	SNP function	A/a	MAF	OR	95% CI	*P* (Fisher)	Risk allele	MAF
1	143755946	–	–	C/T	0.11	1.14	0.92–1.42	0.22	–	0.14
2	143756023	–	–	C/T	0.01	–	–	–	–	–
3	143756101	–	–	G/T	0.01	–	–	–	–	–
4	143756139	rs2978981	5′ flnk	T/C	0.29	1.59	1.36–1.86	4.0×10^{-9}	T	0.38
5	143756303	–	–	G/T	0.02	–	–	–	–	–
6	143756366	rs2976387	5′ flnk	A/G	0.42	1.61	1.39–1.86	1.2×10^{-10}	A	0.48
7	143756700	–	–	C/T	0.02	–	–	–	–	–
8	143757423	rs2976389[a]	5′ flnk	C/T	0.22	–	–	–	–	–
9	143758036	–	–	A/C	0.01	–	–	–	–	–
10	143758105	rs6471587	5′ flnk	C/G	0.35	1.14	0.92–1.41	0.24	–	0.14
11	143758174	rs13262164	5′ flnk	C/T	0.16	0.73	0.62–0.86	1.5×10^{-4}	T	0.26
12	143758933	rs2294008	Met/(Thr)[b]	T/C	0.33	1.58	1.35–1.85	6.3×10^{-9}	T	0.38
13	143759137	rs2920279[c]	Int	C/A	0.34	–	–	–	–	–
14	143759334	–	–	C/G	0.01	–	–	–	–	–
15	143759392	rs2294009	Int	G/A	0.01	1.02	0.61–1.73	1.0	–	0.03
16	143759432	rs2294010	Int	G/A	0.37	1.58	1.35–1.85	5.1×10^{-9}	G	0.38
17	143759726	rs2976391	Int	C/A	0.43	1.42	1.17–1.72	3.2×10^{-4}	C	0.20
18	143759809	rs3736001	Glu/Lys	G/A	0.09	1.22	0.95–1.56	0.11	–	0.11
19	143759934	rs2976392	Int	A/G	0.33	1.62	1.38–1.89	1.1×10^{-9}	A	0.38
20	143759969	rs3736003	Int	C/T	0.10	–	–	–	–	–
21	143759996	–	–	C/G	0.01	–	–	–	–	–
22	143760045	rs2920298	Int	G/A	0.21	1.58	1.35–1.85	5.0×10^{-9}	G	0.38
23	143760085	rs2920297[c]	Int	G/A	0.32	–	–	–	–	–
24	143760111	rs2920296[c]	Int	G/A	0.34	–	–	–	–	–
25	143760322	–	–	G/A	0.01	–	–	–	–	–
26	143760549	rs1045531	Leu/Leu	A/C	0.30	1.58	1.35–1.85	4.8×10^{-9}	A	0.38
27	143760624	rs2976394	3′ UTR	T/C	0.33	1.59	1.36–1.86	3.5×10^{-9}	T	0.38
28	143760654	–	–	A/T	0.03	–	–	–	–	–
29	143760692	rs10216533	3′ UTR	A/G	0.33	1.59	1.36–1.86	3.6×10^{-9}	A	0.38
30	143760752	rs2976395	3′UTR	A/G	0.25	1.59	1.36–1.86	3.9×10^{-9}	A	0.38
31	143760960	rs1045574[a]	Int	A/G	0.32	–	–	–	–	–
32	143761003	rs2976396	3′ UTR	A/G	0.34	1.59	1.35–1.86	4.9×10^{-9}	A	0.38
33	143761011	–	–	C/A	0.01	–	–	–	–	–

Position based on NCBI Build 36. SNP information obtained from dbSNP database. 5′ flnk and 3′ UTR for those SNPs located in the 5′ flanking region and 3′ untranslated region of the *PSCA* gene, respectively, and Int for intronic SNPs. SNPs in the coding region are shown for their effects on amino acid sequence (major allele/minor allele). Major/minor alleles and minor allele frequency were obtained by resequencing of 48 Japanese individuals. Risk allele and control minor allele frequency were obtained by an association study on the stage 2 subjects (749 cases and 750 controls). Dash in an empty cell means that the information is either unavailable or irrelevant.
[a]SNP not genotyped because TaqMan Assay-by-Design development was not successful. [b]When this codon codes for methionine (T allele), it is considered the translation starting site. For C allele, the protein may be translated from the next downstream methionine (**Supplementary Fig. 10** online). [c]SNP not genotyped because of their intronic location and closeness to other SNPs to be typed.

typing data and showed separation of the two genes into different blocks. Moreover, the association of the *LY6K* SNP with diffuse-type gastric cancer was neither significant after Bonferroni correction nor supported by permutation test (**Table 1**). Because no other known gene exists in the LD blocks (**Fig. 1a**), we assumed that the *PSCA* polymorphisms are responsible for the observed association with diffuse-type gastric cancer, and that the relatively low *P* values of the *LY6K* SNPs, albeit not significant after Bonferroni correction, are due to the LD between *LY6K* and *PSCA*.

We then resequenced *PSCA* and its 5′ upstream region on 48 Japanese controls and identified 33 SNPs (**Fig. 1b** and **Table 2**). After excluding some of the rare SNPs (minor allele frequency (MAF) <0.11) and several SNPs incompatible with our typing platform, we genotyped 17 SNPs on 749 cases and 750 controls (**Table 2**). Saturation genotyping on the *PSCA* SNPs revealed the presence of several other statistically significant SNPs, including a missense SNP located at the presumed translation-initiating codon, rs2294008 (OR = 1.58, 95% CI = 1.35–1.85, $P = 6.3 \times 10^{-9}$; **Table 2**).

Next, we inferred the haplotype structure of *PSCA* using the SNPs identified in the resequencing. We focused on both the 5′ upstream and the exonic regions of the gene and found four haplotypes in the upstream region determined by five SNPs, and three haplotypes in the exons determined by seven SNPs (**Table 3**). Odds ratios of each haplotype against the other haplotypes were calculated on the genotyping data (**Table 2**) on the 749 cases with diffuse-type gastric cancer and 750 controls. The odds ratio and its *P* value were not substantially improved by the haplotype analysis as compared to the single SNP–based analysis. Moreover, both upstream and exonic haplotypes contain rs2294008, the only statistically significant missense SNP, and the odds ratio and *P* value of the haplotype seem to depend simply on the fraction of the rs2294008 risk allele, T, in the case and control subjects (**Table 3**). Together with a reporter assay described

Table 3 Haplotypes for functional analyses in the exonic and upstream regions of *PSCA* and their association with diffuse-type gastric cancer

Haplotype ID	Frequency		OR	95% CI	*P* (Wald)	SNP No. list in the haplotype
	Case	Control				
Exonic region						12-18-26-27-29-30-32
PSCA-ExH1	0.63	0.51	1.6	1.4–1.9	5.9×10^{-11}	T- G- A- T- A- A- A
PSCA-ExH2	0.09	0.11	0.82	0.65–1.0	0.11	T- A- A- T- A- A- A
PSCA-ExH3	0.28	0.38	0.63	0.54–0.74	5.0×10^{-9}	C- G- C- C- G- G- G
Upstream region						4- 6-10-11-12
PSCA-UpH1	0.32	0.26	1.4	1.2–1.6	1.6×10^{-4}	T- A- C- T- T
PSCA-UpH2	0.28	0.22	1.3	1.1–1.6	6.1×10^{-4}	T- A- C- C- T
PSCA-UpH3	0.13	0.14	0.88	0.71–1.1	0.23	T- G- G- C- T
PSCA-UpH4	0.28	0.38	0.63	0.54–0.74	4.9×10^{-9}	C- G- C- C- C

Haplotype frequencies in the 749 cases and 750 controls analyzed for the stage 2 screening of the genome scan. Odds ratio and 95% CI were calculated from a contingency table based on the estimated haplotype frequencies. *P* values were calculated by Wald statistics assuming a multiplicative model. SNP number is defined in **Table 2** and **Figure 1b**; in the upstream region, SNPs 4–11 are at −2.8, −2.5, −0.8 and −0.76 kb upstream of the transcription starting site, respectively.

later, the resequencing-based high-density genotyping analysis suggests that rs2294008 is most likely responsible for the association, although other possibilities could not be excluded completely, such as polymorphisms undetectable by resequencing, accumulation of rare variants on a certain haplotype or simply other intronic or synonymous SNPs in this region with strong LD (**Fig. 1c**).

PSCA SNP effect larger in diffuse than in intestinal type

We examined the association of the SNPs in *PSCA* with intestinal-type adenocarcinoma by typing 11 SNPs on 599 cases and 648 controls. **Table 4** summarizes the results on two representative SNPs, which showed significant associations with diffuse- but not intestinal-type adenocarcinoma. When all the subjects genotyped on these SNPs in this study are combined (that is, those from the first and second screening, additional fine mapping and control typing for intestinal-type gastric cancer), our case-control panel consists of 932 diffuse-type gastric cancer cases and 1,398 controls for association analysis. Comparison of the dominant and recessive models suggested that the former fits better statistically to explain the effect of the SNPs, although detailed molecular mechanistic discussion should await further biological analyses. The confidence intervals for the odds ratio of the rs2976392 SNP for the intestinal and diffuse types did not overlap in the dominant model. Moreover, the allele and genotype frequencies of the rs2976392 SNP were significantly different between the two types of cancers ($P = 6.0 \times 10^{-4}$), suggesting that the effect of the *PSCA* polymorphisms is different between the two and is not significant in the intestinal type.

PSCA expressed in differentiating gastric epithelial cells

PSCA was originally identified as a prostate-specific stem cell antigen[20,21], and its expression was also reported in the bladder, esophagus and stomach[22]. However, the precise region of its expression in the gastric epithelium has been unknown. On the surface of the gastric mucosa, epithelial cells form tubular units that consist of four regions from the surface to the bottom: the pit, isthmus, neck and base regions

Table 4 Association of the SNPs in *PSCA* with two major gastric cancer types in Japan

Type	Case genotype			Allele model			Dominant model			Recessive model		
	AA	Aa	aa	OR	95% CI	*P* (Fisher)	OR	95% CI	*P* (logistic)	OR	95% CI	*P* (logistic)
rs2976392 (SNP No.19): risk allele (A) frequency in 1,397 control subjects = 0.616												
Diffuse 926 cases	469	419	38	1.71	1.50–1.94	1.5×10^{-16}	4.24	2.92–6.29	6.4×10^{-18}	1.66	1.39–1.99	1.5×10^{-8}
Intestinal 599 cases	268	272	59	1.29	1.12–1.49	5.0×10^{-4}	1.55	1.13–2.16	0.0059	1.24	1.02–1.52	0.035
1,397 controls	536	650	211	–	–	–	–	–	–	–	–	–
Intestinal vs. diffuse	–	–	–	1.32	1.13–1.56	6.0×10^{-4}	2.73	1.67–4.56	3.3×10^{-5}	1.35	1.07–1.71	0.012
rs2294008 (SNP No. 12): risk allele (T) frequency in 1,396 control subjects = 0.617												
Diffuse 925 cases	461	426	38	1.67	1.47–1.90	2.2×10^{-15}	4.18	2.88–6.21	1.5×10^{-17}	1.62	1.35–1.93	9.4×10^{-8}
Intestinal 599 cases	267	274	58	1.29	1.11–1.49	5.1×10^{-4}	1.59	1.15–2.21	0.0041	1.24	1.01–1.52	0.040
1,396 controls	536	650	210	–	–	–	–	–	–	–	–	–
Intestinal vs. diffuse	–	–	–	1.30	1.10–1.52	0.0015	2.70	1.64–4.50	4.7×10^{-5}	1.35	1.06–1.71	0.013

Genotypes are shown as AA for risk allele homozygotes, Aa for heterozygotes and aa for the nonrisk allele homozygotes. For dominant and recessive models, gender- and age-adjusted OR, its 95% CI and *P* values calculated by exact logistic regression are shown. Allele and genotype frequency differences between the intestinal- and diffuse-type gastric cancers were tested for statistical significance.

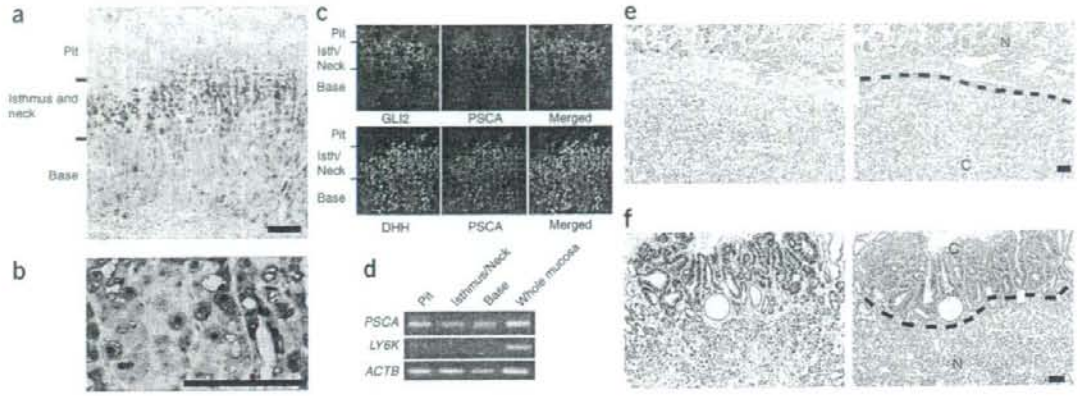

Figure 2 Expression of *PSCA* in the gastric epithelium. (**a**) Immunohistochemical double staining showed the presence of PSCA (blue staining) and PCNA (brown nuclear staining). Although PSCA protein was detected mainly in the cells of the isthmus and neck region, it was also expressed in some cells in the base region. Scale bar, 100 μm. (**b**) A higher magnification of **a**, showing colocalization of the PSCA protein (blue) with some, but not all, of the PCNA-positive cells (brown nuclear staining). Scale bar, 100 μm. (**c**) Immunohistochemical double staining of human gastric epithelium for GLI2 and PSCA and for DHH and PSCA. GLI2, a pit-cell lineage marker, and DHH, a parietal-cell lineage marker, are expressed mainly in the pit and base regions, respectively, although both were also observed in the isthmus/neck region. Merged figures show colocalization of PSCA with GLI2 and DHH. (**d**) RT-PCR analysis on the laser-captured microdissection samples showing *PSCA* transcripts in the pit, isthmus/neck and base regions. *LY6K*, which is located next to *PSCA* on human chromosome 8q24.2 (**Fig. 1**), was not expressed in the gastric epithelial cells. (**e**,**f**) Double staining for PSCA (blue) and PCNA (brown) proteins (left panels) showed silencing of PSCA in the diffuse-type (**e**) and intestinal-type (**f**) gastric cancer cells. Right panels show hematoxylin-eosin staining of the adjacent sections, indicating noncancerous (N) and cancerous (C) portions. Blue staining of the normal gastric glands is scored as 'positive' (**e**,**f**), whereas those of the diffuse- and intestinal-type gastric cancers are 'negative' (**e**) and 'weakly positive' (**f**), respectively. Scale bar, 100 μm.

(**Fig. 2a**). The isthmus is considered to contain stem cells and precursors of three cell lineages that move and differentiate to mature mucus-secreting pit cells (pit-cell lineage), hydrochloric acid–producing parietal cells (parietal-cell lineage) and mucus-secreting neck cells and pepsinogen-producing zymogenic cells (zymogenic-cell lineage)[24,25]. Diffuse-type gastric cancer may originate from the stem cells and/or the three precursors in the isthmus[7].

We raised a monoclonal antibody to PSCA and confirmed its specificity by flow cytometry and immunohistochemical analyses (**Supplementary Figs. 2** and **3** online). The PSCA protein was detected mainly in the middle portion of the gastric epithelium or isthmus and overlapped with the region where a proliferating-cell marker, proliferating cell nuclear antigen (PCNA), was also expressed (**Fig. 2a**). In the isthmus, PSCA protein was detected in cells with a variety of shapes, sizes and amounts of PCNA (**Fig. 2b**), suggesting that it is expressed both in precursor cells, such as pre-pit and pre-neck cells, and in differentiated cells, such as parietal cells. Double staining showed that PSCA and a pit-cell lineage marker, GLI2 (ref. 26), were coexpressed in some cells in the isthmus and neck regions. Similarly, PSCA and a parietal-cell lineage marker, DHH[26], were found to be colocalized in some cells in the region (**Fig. 2c**). The coexpression with the lineage markers in the isthmus region suggests that PSCA is expressed both in the pit- and parietal-cell lineages. In addition, several cells with weak staining were also observed in the pit and base regions (**Fig. 2a** and **Supplementary Fig. 4** online). We carried out RT-PCR analysis on the microdissected human gastric epithelium and detected *PSCA* transcripts in the pit, isthmus, neck and base regions (**Fig. 2d**). In sum, the results of the expression studies suggested that the PSCA protein is mainly expressed in the isthmus, probably in a variety of cell lineages and differentiation stages, with various proliferation activities. PSCA protein was undetectable in the normal intestinal epithelium and was downregulated in the gastric tissue with intestinal metaplasia (**Supplementary Figs. 3** and **4**).

Among cancerous tissues, quantitative RT-PCR (**Supplementary Fig. 5** online) and immunohistochemical (**Fig. 2e**,**f** and **Supplementary Fig. 6** online) analyses showed frequent suppression of PSCA expression in both the intestinal and diffuse types. However, when the immunohistochemical staining was scored in three grades—positive, weakly positive and negative—all 19 diffuse gastric cancer tissues analyzed were negative, whereas 20 out of 21 intestinal cancer tissues showed weakly positive staining, with the remaining one being negative. In all, the PSCA downregulation seemed more pronounced in the diffuse type by immunohistochemistry.

Cell-proliferation inhibition activity of *PSCA*

It has been suggested that PSCA is involved in the cell-proliferation inhibition of prostate epithelial cell lines[27,28]. Among the 12 gastric cancer cell lines we examined, HSC57 and some others did not express *PSCA* (**Supplementary Fig. 5**). We stably transfected *PSCA* cDNA to HSC57 and found that cells with *PSCA* showed fewer G418-resistant colonies than those without *PSCA*, indicating that PSCA has a cell-proliferation inhibition and/or cell-death induction activity (**Fig. 3a**). We isolated four clones with different amounts of *PSCA* expression (**Supplementary Fig. 7** online) and found that those with *PSCA* expression, HSC57P1, HSC57P2 and HSC57P3, showed a slower growth rate than HSC57P0, which did not have *PSCA* expression (**Fig. 3b**). We examined whether the growth suppression of the *PSCA*-transfected cells was due to a significant difference in the number of dead cells. The results suggested that *PSCA* does not induce significant cell death (**Supplementary Fig. 8** online and data not shown).

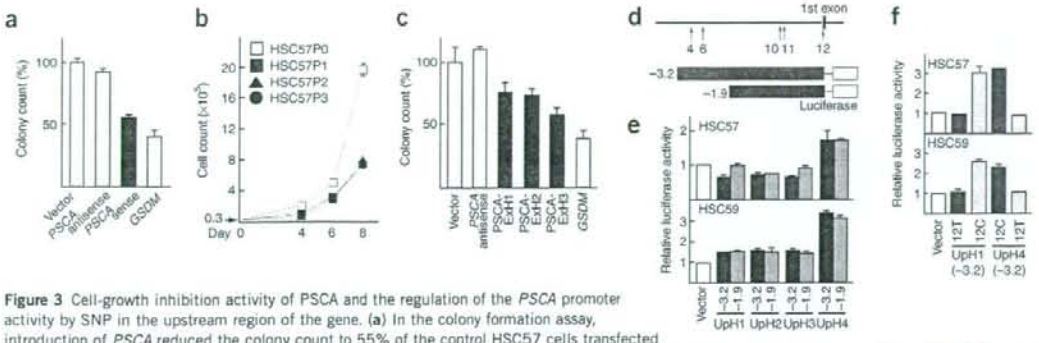

Figure 3 Cell-growth inhibition activity of PSCA and the regulation of the *PSCA* promoter activity by SNP in the upstream region of the gene. **(a)** In the colony formation assay, introduction of *PSCA* reduced the colony count to 55% of the control HSC57 cells transfected with a negative control empty vector. The *GSDM* expression vector was used as a positive control for its cell-proliferation inhibition activity[49]. Error bars, s.d. **(b)** Cell growth assay revealed that the three clones of HSC57 cells, HSC57P1, HSC57P2 and HSC57P3, stably transfected with *PSCA* cDNA showed slower growth rate than the parental HSC57P0, which does not express *PSCA*. **(c)** Colony formation assay showing no significant difference in the cell proliferation inhibition activity among the *PSCA* expression vectors representing the three major haplotypes, ExH1, ExH2 and ExH3, defined by the exonic SNPs (**Table 3**). Repeated experiments are shown in **Supplementary Figure 9** online. **(d)** The 3.2-kb and 1.9-kb upstream fragments cloned in luciferase reporter vectors were prepared by base-substitution to represent each haplotype (UpH1-UpH4) described in **Table 3**. **(e)** Reporter assay revealed a difference in the promoter activity among the 4 haplotypes of the *PSCA* upstream region. **(f)** Substitution of C allele of SNP No.12 (rs2294008) in the PSCA-UpH4(-3.2) for T allele reduced the reporter activity (UpH4(-3.2)-12T), and substitution of T allele of this SNP in the PSCA-UpH1(-3.2) for C increased the reporter activity (UpH1(-3.2)-12C).

The above experiments used *PSCA* cDNA corresponding to the exonic SNP–based haplotype ExH2 (**Table 3**). We next extended the colony formation assay to other haplotypes in order to evaluate their functional differences. *PSCA* cDNA expression constructs were made to represent each of the three major haplotypes covering the exonic regions of the gene (**Table 3** and **Fig. 1b**) by base substitution, and these were introduced to the HSC57 cells. Repeated experiments showed no detectable difference in the suppressive effect on colony formation (**Fig. 3c** and **Supplementary Fig. 9** online). PSCA has an N-terminal signal sequence corresponding to the first 20 residues[20]. *PSCA*-ExH3 has a C allele at rs2294008, which changes the presumed initiation ATG codon to ACG and which may result in a 9-amino-acid truncation of the signal peptide. An *in vitro* translation showed that the product of each haplotype has almost the same size, which is compatible with a difference of only 9 amino acids among the haplotypes (**Supplementary Fig. 10** online). However, a study using EYFP-fusion proteins detected no difference in the subcellular distribution between PSCA-ExH1-EYFP and PSCA-ExH3-EYFP.

Upstream SNP-based haplotypes affect *PSCA* expression

Haplotype analysis deduced four upstream SNP-based haplotypes (**Table 3**), with UpH1 giving the highest odds ratio and UpH4 the least in a one-to-the-others contingency table. To assess the functional significance of the upstream SNPs, we isolated a 3.2-kb genomic fragment including the upstream region and 5′ portion of the first exon of *PSCA*, corresponding to nucleotide positions –3,236 to +28 in relation to the transcription starting site and covering the region from SNPs rs2978981 to rs2294008. Four reporter plasmids constructed by base substitution represent the four haplotypes, designated as PSCA-UpH1(–3.2), PSCA-UpH2(–3.2), PSCA-UpH3(–3.2) and PSCA-UpH4(–3.2) (**Table 3** and **Fig. 3d**). A reporter assay using the HSC59 cells demonstrated that PSCA-UpH4(–3.2) showed the highest luciferase activity, which was 2.5-fold of the activity of the others, and this highest activity was also demonstrated in the assay using the HSC57 cells (**Fig. 3e**). When the C allele of rs2294008 SNP in PSCA-UpH4(–3.2) was replaced by T, the risk allele, the reporter assay showed a reduction of the reporter activity (**Fig. 3f**). In contrast, replacement of the alleles of rs2978981, rs2976387, or rs13262164 by the risk allele did not change the reporter activity (data not shown). The same transcription modulating effect of the haplotypes was also observed for the shorter 1.9-kb version of the *PSCA* upstream region (**Fig. 3e**). The results of the expression and functional analyses suggest that PSCA has an inhibitory effect on cell proliferation and that the rs2294008 SNP can modulate *PSCA* promoter activity.

Replication in Korea

Finally, the association of rs2294008 and rs2976392 was examined at the National Cancer Center (NCC) in Korea on 457 diffuse-type and 418 intestinal-type gastric cancers. Healthy volunteers who visited NCC in Korea for cancer screening provided 390 controls. The two SNPs are also strongly associated in Korea with diffuse-type gastric cancer, but less significantly with intestinal-type cancer (**Table 5**).

DISCUSSION

Through this two-stage genome-wide association study, we identified the *PSCA* gene as a previously unknown susceptibility gene for diffuse-type gastric cancer in Japan. Of the two main categories, the odds ratio was lower and the *P* value higher for the intestinal type, suggesting that *PSCA* is specifically involved in diffuse-type gastric adenocarcinomas. This association pattern was replicated in diffuse- and intestinal-type gastric cancer cases and controls in Korea (**Table 5**). It is notable that the risk alleles exist as major alleles in Japan, whereas the same alleles represent minor alleles in the other HapMap ethnic groups ($P = 5.6 \times 10^{-4}$ and 7.3×10^{-8} against CEU and HCB, respectively, for rs2294008). In Korea, the risk alleles matched those in Japan, but they are also minor alleles in the Korean population (**Supplementary Fig. 11** online).

The odds ratio of *PSCA* SNP rs2294008 in a dominant model was 4.18 (95% CI: 2.88–6.21; **Table 4**), possibly larger than the odds ratio of the –160C/A SNP (rs16260) of *CDH1* (refs. 29–31), the gene implicated in familial and sporadic diffuse-type gastric cancer. Because the –160C/A SNP was not included in the first-stage screening

Table 5 Association of SNPs in *PSCA* with two major gastric cancer types in Korea

	Case genotype			Allele model			Dominant model			Recessive model		
Type	AA	Aa	aa	OR	95% CI	*P* (Fisher)	OR	95% CI	*P* (logistic)	OR	95% CI	*P* (logistic)
rs2976392 (SNP No.19): risk allele (A) frequency in 390 control subjects = 0.463												
Diffuse 449 cases	159	240	50	1.90	1.56–2.33	8.0×10^{-11}	3.47	2.32–5.27	1.1×10^{-10}	1.64	1.17–2.30	0.0036
Intestinal 416 cases	119	213	84	1.37	1.12–1.68	0.0017	1.86	1.27–2.72	0.0010	1.24	0.86–1.80	0.26
390 controls	93	175	122	–	–	–	–	–	–	–	–	–
Intestinal vs. diffuse	–	–	–	1.39	1.14–1.69	9.0×10^{-4}	1.75	1.13–2.74	0.010	1.41	1.01–1.97	0.041
rs2294008 (SNP No. 12): risk allele (T) frequency in 390 control subjects = 0.462												
Diffuse 454 cases	159	246	49	1.91	1.57–2.33	6.3×10^{-11}	3.61	2.41–5.51	3.2×10^{-11}	1.61	1.15–2.26	0.0051
Intestinal 417 cases	118	215	84	1.37	1.12–1.68	0.0017	1.85	1.27–2.71	0.0011	1.22	0.84–1.77	0.31
390 controls	92	176	122	–	–	–	–	–	–	–	–	–
Intestinal vs. diffuse	–	–	–	1.39	1.14–1.69	7.9×10^{-4}	1.81	1.17–2.83	0.0066	1.39	1.00–1.94	0.050

Genotypes are shown as AA for risk allele homozygotes, Aa for heterozygotes and aa for the nonrisk allele homozygotes. For dominant and recessive models, gender- and age-adjusted OR, its 95% CI and *P* values calculated by exact logistic regression are shown. Allele and genotype frequency differences between the intestinal- and diffuse-type gastric cancers were tested for statistical significance.

markers in our genome scan, we genotyped it on 744 diffuse-type gastric cancer cases analyzed in the second-stage genome scan and on 1,397 controls described in **Table 4**. The –160C/A SNP showed a marginally significant protective effect in our diffuse-type gastric cancer cases only in a dominant model ($P = 0.038$, OR = 0.81, 95% CI = 0.66–0.99; **Supplementary Table 2** online).

Although certain transitions and overlaps exist for the two main pathological classifications of gastric cancer, as described earlier, we could still detect a difference between intestinal- and diffuse-type gastric adenocarcinomas with respect to genetic susceptibility. This result is congruous to the proposed notion that two distinct pathways of gastric carcinogenesis exist[32]: one arising in the atrophic gastritis with or without intestinal metaplasia, which develops to the intestinal-type cancer at least initially, and the other arising *de novo* from the gastric epithelial precursor cells, typically leading to diffuse-type cancer. Moreover, our immunohistochemical analysis localized the primary site of the PSCA protein expression at the isthmus of the gastric gland, which is considered to harbor proliferating precursor cells, the presumed origin of diffuse-type gastric adenocarcinoma[2,4]. Thus, the expression pattern may be compatible with the idea that *PSCA* is involved in the susceptibility to diffuse-type gastric cancer.

In this study, information about *H. pylori* infection status was unavailable for most of the subjects, although it is a significant risk factor for both intestinal and diffuse types in Japan, which has a high prevalence of *H. pylori* infection[33]. We do not, however, consider this a major flaw of this study. First, our explanatory variable is an SNP, which is not influenced by *H. pylori* infection. Thus, *H. pylori* infection cannot be a classical confounder, although it could be an intermediate variable between SNP and disease, the former of which may exert its risk effect by modulating the susceptibility of an individual to *H. pylori* infection. In that case, however, we would not want to adjust for *H. pylori* infection, in order not to miss such SNPs in our genome scan. The only important information possibly missed as a result of the lack of *H. pylori* infection data would be the interaction term between it and the SNP. In fact, numerous publications have examined interleukin gene polymorphisms that may increase cancer risk through modulating host response to virulent *H. pylori* strains[34]. However, in Japan, assessing a possible *H. pylori–PSCA* interaction would be difficult, as 99% of Japanese individuals with gastric cancer and 90% of the Japanese adult population are seropositive for *H. pylori*[33].

PSCA encodes a 123-amino-acid glycoprotein with 30% similarity to stem cell antigen 2 (Sca-2), a cell surface marker of immature thymic lymphocytes, and it belongs to the LY-6 family, which shares a glycosylphosphatidylinositol (GPI) anchor domain[20]. Although the function of PSCA is yet to be elucidated, it could have a role in signal transduction, as it is a GPI-anchored membrane protein, and several reports suggest its involvement in cell growth regulation in various systems[27,28,35,36]. *In vitro*, growth of HSC57 cells stably expressing *PSCA* was slower than HSC57 cells expressing no *PSCA* (**Fig. 3b**). Although *PSCA* was overexpressed in prostate cancer cells, it was suppressed in the cancers of the bladder, esophagus, skin and stomach (refs. 22 and 37 and this study). Those findings suggest that *PSCA* is actually expressed in epithelial cells of several tissues other than the prostate and downregulated by malignant transformation, suggesting that PSCA has a tumor suppressor–like character in certain types of cancers, although its roles in prostate cancer may differ.

Because the rs2294008 SNP may alter the first methionine, a polymorphic variation is possible in the length of the N-terminal signal peptide, which in turn can lead to a difference in protein folding, intracellular processing or subcellular localization. However, our *in vitro* analyses including the colony formation assays could detect no difference among the major haplotypes of the *PSCA* exonic regions. On the other hand, a reporter assay showed that rs2294008 can modulate transcriptional activity of the upstream region of *PSCA*. Because of strong LD between rs2294008 and rs2976392 ($r^2 = 0.995$, $D' = 0.999$), the *PSCA* gene was first identified by association with the latter SNP; however, we propose that the functional SNP is the former, rs2294008. Although the actual effect of the SNP *in vivo* is unknown and we could not exclude its possible effect on function of the N-terminal signal sequence, the risk allele, T, is associated with lower transcriptional activity of the supposedly tumor-suppressive *PSCA* gene, at least in the context of the –1.9 kb and –3.2 kb upstream regions of the gene.

Validating the allele-dependent difference in the transcriptional activity *in vivo* proved difficult in our preliminary experiments. We examined a small sample set of noncancerous stomach portions surgically dissected for gastric cancer. However, a quantitative RT-PCR analysis demonstrated no clear relationship between *PSCA* expression and the haplotypes (data not shown). Laser microdissection is at least necessary, but we may still need to normalize the value

of *PSCA* expression by the value of another reference gene whose expression is unique to *PSCA*-expressing cells. Such a reference gene is not yet established.

Finally, **Table 1** shows another candidate genomic region on chromosome 1 for association with diffuse-type gastric cancer that is statistically significant by permutation test and even by Bonferroni correction for multiple testing. The region contains multiple genes in the extended LD block and awaits future investigation. Once most of the common but low-penetrance polymorphisms are identified for this disease, the typing of these variants would have significant diagnostic power to promote a personalized preventive medicine[38]. Until such a collection is established, this line of research may contribute mostly to the understanding of carcinogenesis mechanisms.

METHODS

Classification of gastric adenocarcinoma and case selection. In Japan, pathological diagnosis of gastric cancer includes both macroscopic and microscopic (histological) typing of the gastric resection specimen according to the Japanese classification system of gastric carcinoma[39]. Histologically, the common type of gastric cancer is classified into seven 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, Lauren's classification of gastric cancer into two major categories, intestinal and diffuse types[3], is also widely used, especially for research purposes. In Korea, the pathohistological classification is based on Lauren's system and WHO classification[40]. The relationship of the three classification systems is shown in **Supplementary Table 1** online. In addition to these histological types, the Japanese system identifies types 0 to 5 by macroscopic examination: type 0 corresponds to early gastric cancer in most of the cases and is further subdivided into five subtypes, including a superficial depressed subtype, type 0 IIc. Linitis plastica, a highly malignant diffusely infiltrating form of gastric carcinoma, is type 4, which is almost always the diffuse type by histological classification[8,40].

The original purpose of this study was to identify genetic predisposition to linitis plastica, and we hypothesized that its typical precursor lesions should look like those of type 0 IIc. Therefore, we set linitis plastica as the case selection criterion for the stage 1 screening of the genome scan in Japan. For the stage 2 screening, we expanded inclusion criteria to include type 0 IIc with por2 or sig type histology. For the analysis of intestinal-type gastric cancers, we included pap, tub1 and tub2 by histological diagnosis and excluded type 4 by macroscopic examination.

In the replication study in Korea, the inclusion criterion for diffuse-type gastric cancer was linitis plastica or Lauren's diffuse type with poorly differentiated or signet-ring cell type histology by WHO classification. The inclusion criterion for intestinal-type gastric cancer was Lauren's intestinal type with papillary or well-differentiated or moderately differentiated type histology in WHO classification, and the exclusion criterion was linitis plastica.

Study design and subjects. In the first-stage screening of the genome scan, each SNP was genotyped on 188 cases and 752 controls. Because some DNA samples were used up and had to be replaced with other samples in the middle of the SNP-by-SNP typing in this genome scan, a total of 194 DNA samples (107 males and 87 females, mean age of 57 years) were obtained from individuals with the linitis plastica type of gastric cancer. The germline DNA was extracted either from peripheral blood leukocytes, from methanol-fixed paraffin-embedded tissues of noncancerous gastric mucosa or lymph nodes[41], or from fresh noncancerous gastric mucosa dissected by operation. The control or reference samples of the first screening were a combined group of the individuals with four other diseases in the joint five-disease genome scan project (Millennium Genome Project in Japan)[18]; the reference genotype frequency data were obtained on 752 subjects (188 individuals for each of the four diseases) per SNP from the total of 190 individuals with Alzheimer's disease, 213 with type 2 diabetes, 189 with hypertension, and 194 with asthma. The reference DNA samples were extracted from peripheral blood leukocytes.

In the following second-stage screening and a subsequent high-density SNP typing of the *PSCA* gene, 763 DNA samples from individuals with gastric cancer (male 403, mean age 56; female 360, mean age 55) were extracted either from methanol-fixed paraffin-embedded tissues of noncancerous gastric mucosa or lymph nodes or from peripheral blood of individuals with the linitis plastica type of gastric cancer (165 cases) or early-stage cancer diagnosed as macroscopic type 0 IIc with histological type of por2 and/or sig[39] (598 cases). The gastric cancer samples were collected at six institutions as follows: 365 paraffin-embedded tissues and 167 blood samples by the National Cancer Center Hospital in Tokyo, 128 blood samples by Nippon Medical School Hospital in Tokyo, 57 blood samples by Aichi Cancer Center in Aichi, 23 blood samples by Shikoku Cancer Center in Ehime, 19 blood samples by Hiroshima University Graduate School of Biomedical Sciences in Hiroshima and 4 blood samples by Hamamatsu University School of Medicine in Shizuoka. The control DNA samples for the second screening and for the high-density SNP typing on *PSCA* were from peripheral blood leukocytes of 751 volunteer individuals (male 387, mean age 55; female 364, mean age 54) with no known malignancies who offered blood at the occasion of a health check examination at two institutions: 204 and 335 individuals at Keio University campuses in Kanagawa and Tokyo, respectively, and 212 at Iwata City Hospital in Shizuoka. As in the first-stage screening, samples for 14 gastric cancer cases and a control were exhausted and replaced with another 15 samples; the final valid typing data were obtained on 749 cases and 750 controls for the second screening and also for the high-density typing of *PSCA*.

To construct an LD map around the *PSCA* gene, we obtained genotype data from a control population of 379 Japanese volunteers from the Iwata City Hospital (108 males), National Cancer Center (60 females) and Keio University (211 females).

Resequencing of *PSCA* was also done on DNA samples from 48 healthy Japanese volunteers at Keio University campuses, 47 of whom were not included in the control subjects genotyped in the second screening of the genome scan.

We obtained genotype data on the 11 *PSCA* SNPs for 599 intestinal-type gastric cancer cases and 648 controls. The DNA samples were extracted from 433 methanol-fixed paraffin-embedded tissue archives of noncancerous gastric mucosa or lymph nodes in the National Cancer Center Hospital and from 166 peripheral blood leukocytes from individuals at Nippon Medical School Hospital. The 648 control samples were derived from peripheral blood leukocytes of volunteers at Iwata City Hospital, who were different individuals from the 212 control Iwata subjects genotyped in the second screening by Invader assay. Therefore, data on the 11 *PSCA* SNPs were available for a total of 1,399 (751 + 648) control subjects.

In the Korean study, peripheral blood samples were donated from individuals with gastric cancer who were diagnosed or treated in the National Cancer Center in Seoul, Korea. The control subjects were volunteers who participated in the National Cancer Screening Program at the National Cancer Center and who were confirmed not to have gastric cancer by an endoscopy. The sample size of at least 350 subjects in each group was designed to provide the study with 90% power to detect an association of a SNP with minor allele frequency of 0.38 and allelic odds ratio of 1.45 in a two-sided test at a significance level of 5%.

Age and gender distributions of the subjects are shown in **Supplementary Figure 12** online. 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 component of the gastric cancer 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.

Depending on the samples, some restrictions may apply for sharing human materials with readers because of limited availability or conditions of the original informed consent.

Genotyping and resequencing. Genotyping for stages 1 and 2 of the genome scan was done using the Invader (Third Wave Technologies) assay combined with multiplex PCR[42], with some modifications such as the use of only 10 ng of genomic DNA in 10 μl total reaction volume. Fine typing of the *PSCA* region

was done on the SNPs selected by the resequencing data using either the Invader or TaqMan SNP Genotyping (Applied Biosystems) assays. We genotyped the intestinal-type gastric cancer cases and controls using Illumina GoldenGate assay (Illumina) developed for 14 SNPs of *PSCA*, with success on 11 SNPs. Part of genotype data to calculate linkage disequilibrium of the *PSCA*-containing genomic region was obtained by Sentrix Human-1 Genotyping BeadChip (Illumina). All the Korean samples were genotyped in the National Cancer Center, Korea by TaqMan SNP Genotyping Assays.

The Invader assay is a SNP-by-SNP genotyping system, and we excluded SNP markers that showed significant deviation from the Hardy-Weinberg equilibrium (χ^2 P value <0.001) in any of the reference disease groups. A final genotype call was made by visual quality-check inspection of SNP cluster plots of normalized fluorescence intensities. In the stage 1 screening, we genotyped 100,000 SNPs identified by the JSNP project and obtained 85,925 SNPs that passed our strict quality control process. After removing some SNPs that did not have JSNP ID assignment at that time, we subjected 85,576 SNPs to statistical analysis. In the stage 2 screening, 2,880 SNPs were genotyped, and 2,753 passed quality control.

Resequencing of *PSCA* was done on 48 control individuals using BigDye Terminator v1.1 Cycle Sequencing Kit and ABI PRISM 3700 DNA Analyzer (Applied Biosystems). The sequencing covered nucleotide positions 143,755,899 to 143,761,136 (NCBI Build 36), except for about 300 bp of sequence (for example, 143,757,522–143,757,748, 143,759,074–143,759,094, 143,760,414–143,760,439 and 143,760,801–143,760,837), because of difficulties in primer design.

Statistical analyses. The statistical power of this two-stage genome-wide scan study has been described elsewhere[18,23]. The statistical significance of the association was evaluated for each SNP by Fisher's exact tests. Crude odds ratios for allele, dominant and recessive inheritance models were calculated in the first-stage screening, in which two sets of control or reference data were used: one was the allele and genotype frequency data of the combined group of individuals with four other diseases (see above), and the other was the allele frequency reference data of the general Japanese population (752 individuals) available from the JSNP database (JSNP reference data). The SNPs subjected to the second stage of the genome scan were selected by the following step-wise procedure: (i) the SNPs were excluded if their minor allele frequencies were less than 0.1 in the JSNP reference data, (ii) the SNPs were excluded if they were not mapped on autosomal chromosomes, (iii) the odds ratios and P values against the JSNP reference data were ignored if the allele frequencies of the JSNP reference data and those of the reference disease group were significantly different (that is, P value $<10^{-4}$), (iv) the odds ratios and P values for the reference disease group were ignored if the reference disease group consisted of only one disease as a result of quality control of typing data, (v) genotype odds ratios and P values for the reference disease group were ignored when the P value for Hardy-Weinberg equilibrium was less than 0.01 in the reference disease group, (vi) the SNPs were excluded if their genotype odds ratios were less than 1.5 and allelic odds ratios were less than 1.3, (vii) the SNPs were excluded if the risk allele or genotypes against the reference disease group were opposite to those against the JSNP control data, (viii) the top 4,000 SNPs were selected without redundancy in the order of increasing P values by Fisher's exact tests, and (ix) pair-wise LD parameters were estimated among the 4,000 SNPs, and the SNP with a lower P value was selected from each pair of the SNPs showing a high LD parameter ($r^2 > 0.9$). Finally, we selected 2,880 SNPs for the stage 2 typing, which generated valid data on 2,753 SNPs after a quality check process.

The significance of association in the second screening was evaluated at the significance level of 0.05 by multiple testing using Bonferroni correction and permutation test[43]. Assuming 2,753 independent hypothesis tests, the conservative Bonferroni correction requires a P value of 1.8×10^{-5} before correction. Odds ratios adjusted for three age categories (≤ 39, 40–59 and ≥ 60) and gender were also calculated by exact logistic regression for dominant and recessive models for the stage 2 screening data and in the Korean replication study (**Supplementary Fig. 12**).

The haplotype-based association was tested on the basis of Wald statistics, which were obtained by CHAPLIN[44] case-control haplotype inference software, assuming a multiplicative model. Other statistical analyses were carried out using the Statistical Analysis System software version 9.1 (SAS Institute) and the R suite.

To evaluate the effect of population stratification in this study, we examined the population structure by the STRUCTURE software[45], the Genomic Control[46] and mixture model[47] methods using 1,025 and 501 loci from the stage 1 and 2 samples, respectively. None of the three analyses detected a significant subpopulation in the stage 1 or 2 samples after removing 28 samples from the 26 pairs and 1 trio that showed more than a 65% genotype concordance rate within the total of 980 samples analyzed in the first-stage genotyping. These samples may represent duplication, identical twins or first-degree relatives, depending on the degree of the concordance (**Supplementary Note** online). Even if we revise the first-stage screening data by removing 5 gastric cancer cases and 23 other-disease reference samples from the groups showing the high genotype concordance rate, 9 of the 10 SNPs listed in **Table 1** are nevertheless selected for the stage 2 screening, but the fourth SNP, rs3804775, does not pass the selection criteria because of its low odds ratio.

The genotype frequency data for the first and second screening are available in the GeMDBJ database (see URLs section below). Because the first-screening data may be used as an independent reference dataset by database users, the revised, 'cleaned-up' data after removing the high concordance samples are presented in the public database to minimize a spurious stratification.

Linkage disequilibrium analysis. LD analysis of the *PSCA-LY6K* region was done using genotype data of 379 control individuals obtained by Illumina Human-1 BeadChip, which contains 28 SNPs that reside in and around the *PSCA* gene: rs4587307, rs6988393, rs750529, rs2976399, rs2294008, rs2976391, rs2920298, rs2920297, rs2976394, rs10216533, rs2976395, rs1045547, rs1045605, rs1435453, rs2585174, rs2585153, rs4736323, rs7822193, rs2585187, rs13362, rs2585126, rs2239703, rs2572925, rs3750246, rs3764795, rs3802226, rs3819496 and rs3819497. The pattern of LD was analyzed using two parameters, r^2 and D' (ref. 48).

Functional analyses. Materials and methods for preparation of antibody to PSCA, immunohistochemistry, laser-captured microdissection and RT-PCR, cDNA isolation, colony formation assay, cell growth assay and reporter assay are described in the **Supplementary Methods** and **Supplementary Table 3** online.

URLs. JSNP database, http://snp.ims.u-tokyo.ac.jp/; GeMDBJ database, http://gemdbj.nibio.go.jp/.

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

ACKNOWLEDGMENTS

This work was supported in Japan by the program for promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NiBio) and by Health and Labour Sciences Research Grants by Ministry of Health, Labour and Welfare. The Korean part of the study was supported by grant 0710340 from the National Cancer Center, Korea.
We thank the following people (listed in alphabetical order) for discussion and technical and statistical assistance: M. Asako, T. Chujo, C. Hamada, T. Hayashida, C. Hirama, F. Igarashi, T. Imai, E. Inoue, S. Kamakami, A. Katoh, O. Kawaguchi, C. Kina, N. Kurata, Y. Liu, G. Maeno, S. Mimaki, N. Mitsuhashi, N. Miyahara, A. Miyaoka, R. Nakajima, J. Nakata, Y. Odaka, T. Ogiwara, N. Ohsawa, E. Ohshima, M. Okada, M. Okuyama, Y. Sakashita, M. Sato, M. Seishi, T. Sobue, H. Suganami, E. Takemoto, T. Taniguchi, S. Uchida, T. Urushidate, M. Ushiama, S. Yabe, N. Yamaguchi, S. Yamamoto and I. Yoshimura.

AUTHOR CONTRIBUTIONS

Principal investigators: H. Sakamoto, K. Yoshimura and N.S. **Gastric cancer project planning and design:** H. Sakamoto, K. Yoshimura, T.Y., H.K. and T.S. **Ascertainment of subjects for genetic analyses:** H.K., T.S., Y.M., D.S., H. Sugimura, F.T., S.K., N. Matsukura, N. Matsuda, T. Nakamura, I. Hyodo, T. Nishina, W.Y., H.H., M.H., E.T. and S.H. **Genetic analyses:** H. Sakamoto, Sumiko Ohnami, A.S. and Y.N. **Statistical analyses and database:** K. Yoshimura, Y.S., H.T., M.A., R.T., Y.T., M.O., K. Aoki, I. Honmyo and S.C. **Functional analyses:** N.S., K. Aoyagi, H. Sasaki, Sumiko Ohnami, Shunpei Ohnami, K. Yanagihara and H. Sakamoto. **National Cancer Center, Korea:** K.-A.Y., M.-C.K., Y.-S.L., S.R.P., C.G.K. and I.J.C. **Millennium Genome Project Cancer Subteam Leader:** T.Y. **Millennium Genome Project Human Genome Variation Team Leader:** Y.N. **Millennium Genome Project Disease Gene Team Leader:** S.H.