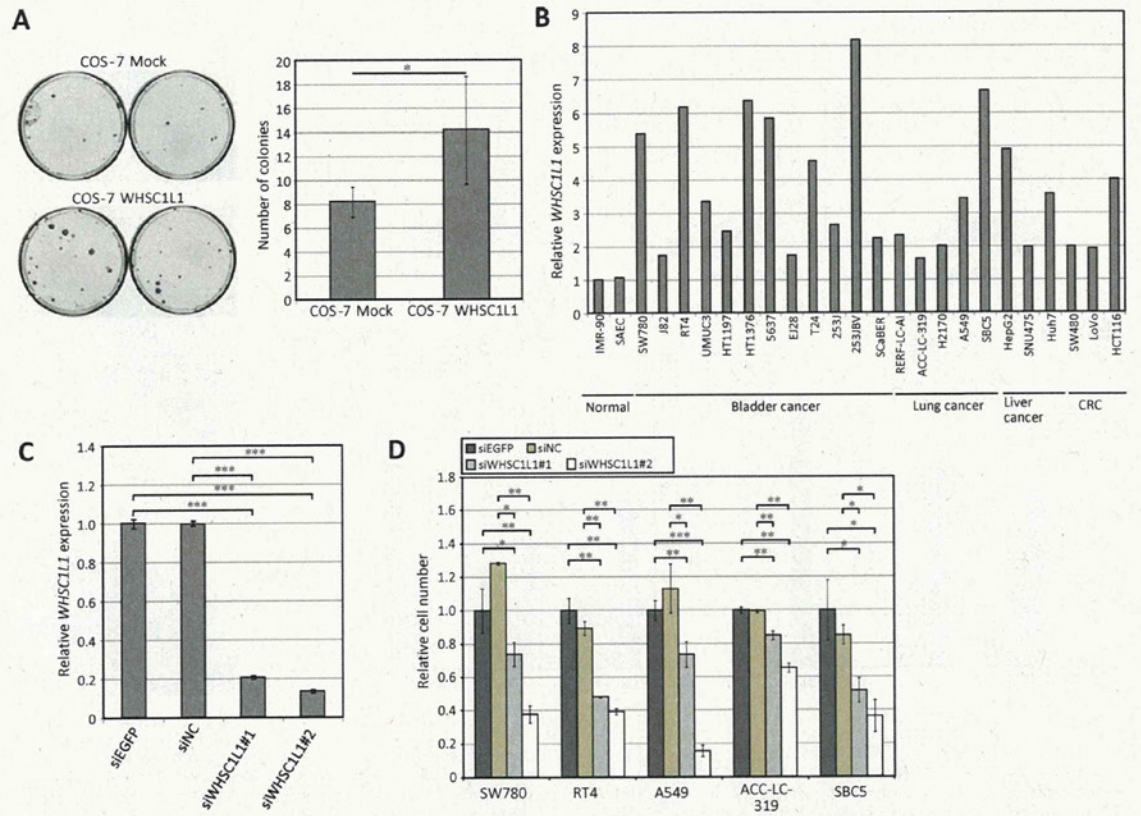

Figure 3. Involvement of WHSC1L1 in the growth of bladder and lung cancer cells. (A) Clonogenicity assays of COS-7 cells. Cells transfected with a 3xFLAG-mock vector and a 3xFLAG-WHSC1L1 vector were cultured in DMEM 10% FBS containing 0.4 (mg/mL) Geneticin/G-418 for 2 weeks, and colonies were stained with Giemsa. (B) Expression of *WHSC1L1* in 2 normal cell lines, 12 bladder cancer cell lines, 5 lung cancer cell lines, 3 liver cancer cell lines, and 3 colorectal cancer cell lines. mRNA expression levels were analyzed by quantitative real-time PCR and normalized by *GAPDH* and *SDH* expressions. Values are relative to IMR-90 (IMR-90 = 1). (C) Expression of *WHSC1L1* in A549 cells treated with two independent specific siRNAs targeting WHSC1L1 (siWHSC1L1#1, #2) was analyzed by quantitative real-time PCR. siRNAs targeting *EGFP* (siEGFP) and siNC were used as controls. mRNA expression levels were normalized by *GAPDH* and *SDH* expressions, and values are relative to siEGFP (siEGFP = 1). Results are the mean ± SD of three independent experiments. *P* values were calculated using Student's *t*-test (***$P < 0.001$). (D) Effects of WHSC1L1 siRNA knockdown on the viability of two bladder cancer cell line (SW780, RT4) and three lung cancer cell lines (A549, LC319, and SBC5). Relative cell number shows the value normalized to siEGFP-treated cells (siEGFP = 1). Results are the mean ± SD in three independent experiments. *P* values were calculated using Student's *t*-test (*$P < 0.05$; **$P < 0.01$; and ***$P < 0.001$). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

and siNC controls. We subsequently examined the effects of siRNAs on the growth of cancer cells using the cell counting kit-8 system (Fig. 3D) and found that two independent siWHSC1L1s significantly suppressed growth of two bladder cancer cell lines (SW780 and RT4) and three lung cancer cell lines (A549, ACC-LC-319 and SBC5), compared with siEGFP or siNC.

To further clarify the mechanism of growth suppression induced by siRNA, we performed flow cytometry assay (FACS) analysis (PI staining) and found that the proportion of cancer cells at G_2/M phase was significantly higher when they were treated with siWHSC1L1 than control siRNAs. On the other hand, the proportion of cancer cells at G_1 phase was markedly decreased in the cells treated with WHSC1L1 siRNA (Figs. 4A and 4B). This result was also confirmed by BrdU and 7-AAD double staining and FACS analysis (Fig. 4C and Supporting Information, Fig. S3). After the treatment with WHSC1L1 siRNA, the population of cancer cells with greater than 4N DNA content was significantly increased (Figs. 4A and 4C), implying that depletion of WHSC1L1 strongly induced the cell cycle arrest at the G_2/M phase and resulted in multinucleation of cancer cells. Taken together, these data indicate that WHSC1L1 plays a critical role in the growth through the regulation of cell-cycle progression, especially at G_2/M phase.

Candidate Downstream Genes Regulated by WHSC1L1 in Cancer Cells

To analyze the WHSC1L1-signal pathway, we performed microarray expression analysis using

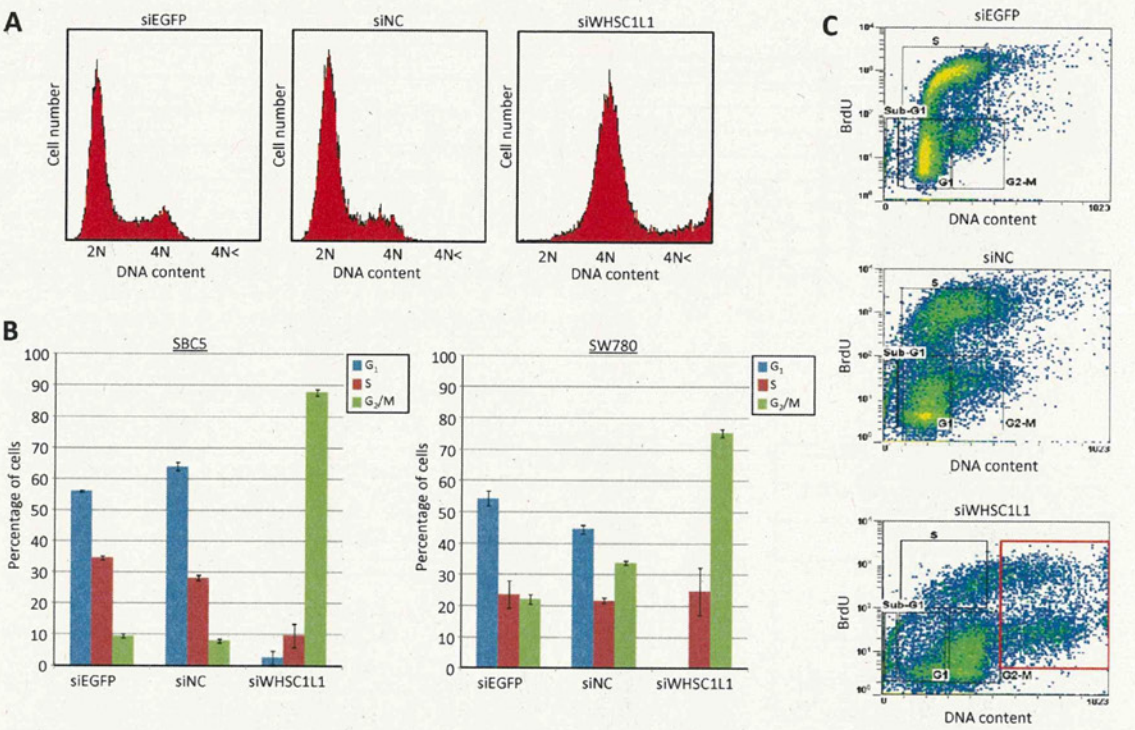

Figure 4. Effects of WHSC1L1 knockdown on cell cycle kinetics in cancer cells. (A,B) SBC5 and SW780 cells were treated with siRNAs and analyzed by FACS 72 hr after siRNA treatment. Representative histograms of this experiment (SBC5 cells) are shown (A). Numerical analysis of the FACS result, classifying cells by cell cycle status (B). Mean ± SD of three independent experiments. *P* values were calculated using Student's *t*-test. (C) Cell cycle distribution was analyzed by flow cytometry after coupled staining with FITC-conjugated anti-BrdU and 7-AAD. Mean ± SD of three independent experiments. *P* values were calculated using Student's *t*-test. Cells enclosed in the red frame are multinucleated.

Affymetrix's GeneChip® system. After knocking down of WHSC1L1 in SW780 and A549 cancer cells using siWHSC1L1#1, we isolated total RNA from SW780 and A549 24 hr after siRNA treatment, because WHSC1L1 protein expression was clearly knocked down at this time-point (Supporting Information, Fig. S2A). The expression profiles of these cells were compared to those treated with control siRNAs (siEGFP and siFFLuc) using HG-U133 Plus 2.0 Array. Expression levels of 42 genes significantly decreased to half or lower than the control cells, and those of four genes increased two times or higher than the control cells (-Table 2). We were able to validate the down-regulation of all randomly selected downstream gene candidates (Supporting Information, Fig. S4), so our microarray data are likely to be representative.

Signal pathway analysis of candidate downstream genes using the Gene Ontology database ("Materials and Methods" section and Table 2) indicated that alteration of WHSC1L1 expression level might affect expression of *Cyclin G1* (*CCNG1*) and *NEK7*, important regulators during mitosis (Seo et al., 2006; Salem et al., 2010). We statistically analyzed the expression changes for *CCNG1* and *NEK7* in A549, SBC5, and SW780 and validated that knockdown of WHSC1L1 by specific siRNA significantly attenuated the expression levels of *CCNG1* (Fig. 5A) and *NEK7* (Fig. 5B), implying that WHSC1L1 appears to activate the expression of *CCNG1* and *NEK7* in cancer cells.

DISCUSSION

In this study, we have shown that WHSC1L1 was overexpressed in bladder cancer, lung cancer, liver cancer, and CML and was likely to play a critical role in the growth of cancer cells through regulation of cell-cycle progression at G_2/M phase. Consistently, as WHSC1L1 is known to methylate lysine 36 of histone H3 (Morishita and di Luccio, 2011), we identified that H3K36 dimethylation was elevated in tumor cells overexpressing WHSC1L1 (Supporting Information, Fig. S5). In addition, this is the first report demonstrating the significant

TABLE 2. Affymetrix's Microarray Data Analysis of WHSC1L1

Probe ID	Gene symbol	Mean ratio	Up or down	Gene description	Gene ontology description
229450_at	IFIT3	4.227	Up	Interferon-induced protein with tetratricopeptide repeats 3	Cellular response to interferon-alpha; type I interferon-mediated signaling pathway
218943_s_at	DDX58	3.392	Up	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58	Positive regulation of transcription from RNA polymerase II promoter; positive regulation of sequence-specific DNA binding transcription factor activity
219863_at	HERC5	3.365	Up	Hect domain and RLD 5	Regulation of cyclin-dependent protein kinase activity; cytokine-mediated signaling pathway
204622_x_at	NR4A2	2.31	Up	Nuclear receptor subfamily 4, group A, member 2	Positive regulation of transcription from RNA polymerase II promoter; transcription, DNA-dependent
209209_s_at	FERMT2	0.498	Down	Fermitin family member 2	Actin cytoskeleton organization, cell junction assembly
200806_s_at	HSPD1	0.495	Down	Heat shock 60 kDa protein 1	Chaperone-mediated protein complex assembly; ATP catabolic process
203053_at	BCAS2	0.494	Down	Breast carcinoma amplified sequence 2	RNA splicing, via transesterification reactions
32042_at	ENOX2	0.492	Down	Ecto-NOX disulfide-thiol exchanger 2	Regulation of growth; oxidation-reduction process
232591_s_at	TMEM30A	0.49	Down	Transmembrane protein 30A	Positive regulation of protein exit from endoplasmic reticulum
212423_at	ZCCHC24	0.487	Down	Zinc finger, CCHC domain containing 24	Nucleic acid binding
211676_s_at	IFNGR1	0.486	Down	Interferon gamma receptor 1	Interferon-gamma-mediated signaling pathway
226994_at	DNAJA2	0.485	Down	Dnaj (Hsp40) homolog, subfamily A, member 2	Positive regulation of cell proliferation
229103_at	WNT3	0.482	Down	Wingless-type MMTV integration site family, member 3	Wnt receptor signaling pathway, calcium modulating pathway
230069_at	SFXN1	0.472	Down	Sideroflexin 1	Transmembrane transport
212573_at	ENDOD1	0.47	Down	Endonuclease domain containing 1	Nucleic acid binding
214831_at	ELK4	0.47	Down	ELK4, ETS-domain protein	Regulation of transcription, DNA-dependent
205371_s_at	DBT	0.465	Down	Dihydrolipoamide branched chain transacylase E2	Cellular nitrogen compound metabolic process; fatty-acyl-CoA biosynthetic process
221825_at	ANGEL2	0.462	Down	Angel homolog 2	None
225972_at	TMEM64	0.454	Down	Transmembrane protein 64	Integral to membrane
1555797_a_at	ARPC5	0.451	Down	Actin related protein 2/3 complex, subunit 5, 16kDa	Actin cytoskeleton organization
235204_at	ENTPD7	0.451	Down	Ectonucleoside triphosphate diphosphohydrolase 7	Hydrolase activity

(Continued)

TABLE 2. Affymetrix's Microarray Data Analysis of WHSC1L1 (Continued)

Probe ID	Gene symbol	Mean ratio	Up or down	Gene description	Gene ontology description
225489_at	TMEM18	0.449	Down	Transmembrane protein 18	Transcription, DNA-dependent
212530_at	NEK7[a]	0.449	Down	NIMA (never in mitosis gene a)-related kinase 7	Regulation of mitotic cell cycle; cytokinesis; spindle assembly
218236_s_at	PRKD3	0.443	Down	Protein kinase D3	Activation of protein kinase C activity by G-protein coupled receptor protein signaling pathway
214649_s_at	MTMR2	0.442	Down	Myotubularin related protein 2	Negative regulation of excitatory postsynaptic membrane potential head group
201151_s_at	MBNL1[b]	0.438	Down	Muscleblind-like	Regulation of RNA splicing
210959_s_at	SRD5A1	0.436	Down	Steroid-5-alpha-reductase, alpha polypeptide 1	Cellular response to cAMP
238756_at	GAS2L3	0.427	Down	Growth arrest-specific 2 like 3	Cell cycle arrest
209838_at	COPS2	0.424	Down	COP9 constitutive photomorphogenic homolog subunit 2	Transcription from RNA polymerase II promoter; cullin deneddylation
223208_at	KCTD10	0.424	Down	Potassium channel tetramerization domain containing 10	Proteasomal ubiquitin-dependent protein catabolic process
202020_s_at	LANCL1	0.412	Down	LanC lantibiotic synthetase component C-like 1	G-protein coupled receptor signaling pathway
208796_s_at	CCNG1[a,b]	0.407	Down	Cyclin G1	Mitotic cell cycle G2/M transition DNA damage checkpoint; cell growth; cell division
235931_at	FAM119A	0.407	Down	Methyltransferase like 21A	Methyltransferase activity
202438_x_at	IDS[b]	0.405	Down	Iduronate 2-sulfatase	Iduronate-2-sulfatase activity
218173_s_at	WHSC1L1[c]	0.392	Down	Wolf–Hirschhorn syndrome candidate 1-like 1	
217097_s_at	PHTF2	0.383	Down	Putative homeodomain transcription factor 2	Regulation of transcription, DNA-dependent
212919_at	DCP2	0.381	Down	DCP2 decapping enzyme homolog	Exonucleolytic nuclear-transcribed mRNA catabolic process involved in deadenylation-dependent decay
205122_at	TMEFF1[b]	0.371	Down	Transmembrane protein with EGF-like and two follistatin-like domains 1	Multicellular organismal development
222697_s_at	ABHD10	0.361	Down	Abhydrolase domain containing 10	Proteolysis
211023_at	PDHB	0.325	Down	Pyruvate dehydrogenase (lipoamide) beta	Regulation of acetyl-CoA biosynthetic process from pyruvate
221268_s_at	SGPP1	0.321	Down	Sphingosine-1-phosphate phosphatase 1	Sphingolipid biosynthetic process
212688_at	PIK3CB	0.311	Down	Phosphoinositide-3-kinase, catalytic, beta polypeptide	Phosphatidylinositol-3-phosphate biosynthetic process
214484_s_at	SIGMAR1	0.289	Down	Sigma nonopioid intracellular receptor 1	Ergosterol biosynthetic process
215017_s_at	FNBP1L	0.287	Down	Formin binding protein 1-like	Endocytosis
207543_s_at	P4HA1	0.222	Down	Prolyl 4-hydroxylase, alpha polypeptide 1	Peptidyl-proline hydroxylation to 4-hydroxy-L-proline
235016_at	REEP3	0.216	Down	Receptor accessory protein 3	Integral to membrane

[a]These genes are related to the regulation of the mitotic cell cycle.
[b]Expression change of these genes were also confirmed by real-time PCR.
[c]Confirmation of WHSC1L1 knockdown.

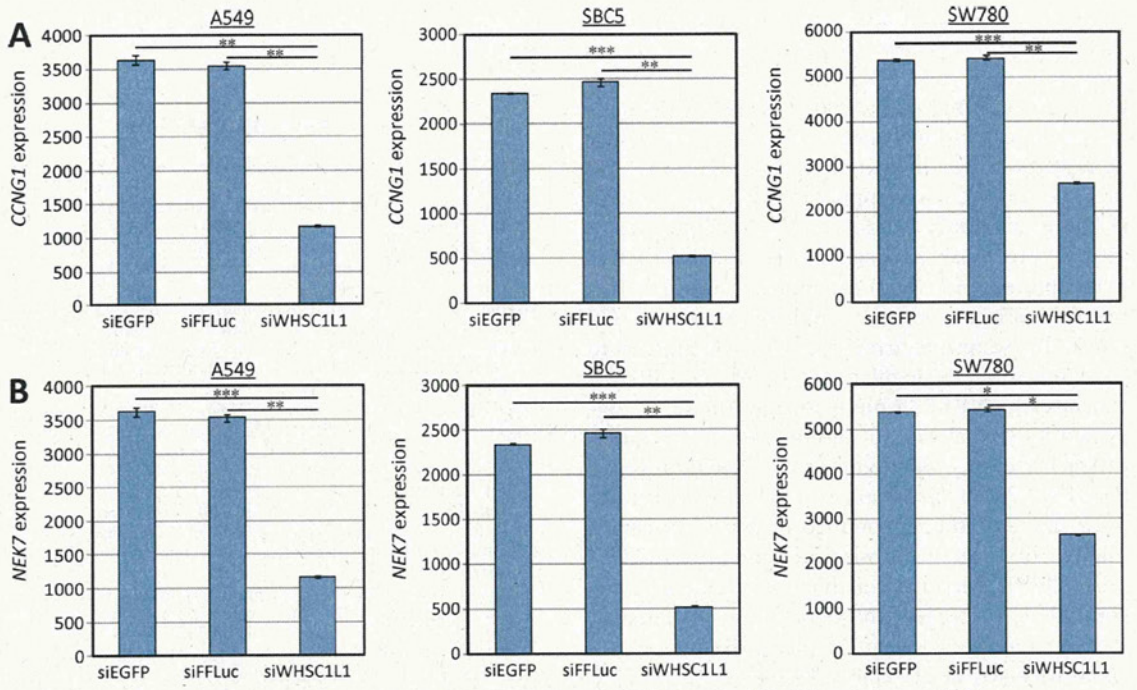

Figure 5. Expression profile data of *CCNG1* (A) and *NEK7* (B) analyzed by Affymetrix's GeneChip® system. A549, SBC5, and SW780 cells were used for the analysis, and expression levels of these genes were based on two independent experiments. *P* values were calculated using Student's *t*-test (*$P < 0.05$; **$P < 0.01$; and ***$P < 0.001$). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

increase of H3K36 methylation in tumor cells. As expression levels of *WHSC1L1* are significantly low in normal tissues, including heart, liver, lung, and kidney (Supporting Information, Fig. S6), WHSC1L1 is likely to be a promising target for cancer therapy. *WHSC1L1* is a member of a family of genes that currently includes two additional members *NSD1* and *WHSC1*, also known as *NSD2*. NSD1, WHSC1, and WHSC1L1 show strong sequence similarity, particularly in their C-terminal region, which includes the functional domains such as SET and PHD domains (Douglas et al., 2005). *NSD1* mutations and deletions are largely responsible for Sotos syndrome (Douglas et al., 2003; Nagai et al., 2003; Rio et al., 2003; Turkmen et al., 2003). The chromosomal translocation t(5;11)(q35;p15.5) detected in acute myeloid leukemia (AML) cells generates a fusion gene of *nucleoporin-98* (*NUP98*) and *NSD1* (Cerveira et al., 2003). It was recently reported that the NUP98-NSD1 fusion protein was implicated to sustain self-renewal of myeloid stem cells and induce expression of the *HOXA7*, *HOXA9*, *HOXA10*, and *MEIS1* proto-oncogenes. NUP98-NSD1 is known to bind genomic elements adjacent to *HOXA7* and *HOXA9*, maintains histone H3 Lys 36 (H3K36) methylation and histone acetylation, and prevents EZH2-mediated transcriptional repression of the *HOXA* locus during differentiation (Wang et al., 2007). Importantly, NSD1-dependent dysregulation of H3K36 methylation appears to be closely linked to tumorigenesis through the regulation of transcription of various downstream genes (Wang et al., 2007). As we also found that WHSC1 was overexpressed in many types of cancer (Toyokawa et al., 2011b), it seems that the WHSC1 family members are involved in human carcinogenesis commonly through their methyltransferase activity.

WHSC1L1 was first reported as a highly similar gene to *WHSC1* located in a WHS critical region (Stec et al., 1998, 2001) and was involved in chromosomal translocation t(4:14)(p18.3;q32.3) in multiple myelomas (Chesi et al., 1998; Stec et al., 1998). We here demonstrated that WHSC1L1 was likely to play a critical role in the G_2/M transition of cancer cells and may activate the expression of *CCNG1* and *NEK7*. CCNG1 is associated with cyclin-dependent kinase 5 (CDK5) and non-CDK-serine/threonine kinase, namely cyclin G-associated kinase (Kanaoka et al., 1997; Kimura et al., 1997). Furthermore, CCNG1 deregulation is associated with genomic instability (Tanaka and Diffley, 2002). In addition,

CCNG1 has been suggested to act as an oncogenic protein because of its overexpression in human tumor cells (Reimer et al., 1999; Baek et al., 2003; Perez et al., 2003). Some experimental evidences using cancer cell lines and tumor xenografts have indicated that suppression of CCNG1 resulted in suppression of tumor growth (Chen et al., 1997; Gordon et al., 2000). Seo et al. (2006) previously reported that CCNG1 was overexpressed in lung carcinoma, and CCNG1 overexpression overcame radiation-induced G2 arrest. These results reveal that CCNG1 appears to be a fundamental regulator of G_2/M transition of cancer cells. NEK7, a nimA-related kinase, is suggested its crucial role in mitotic spindle formation (Yissachar et al., 2006; Kim et al., 2007a; O'Regan and Fry, 2009). A recent study using knockout mice reported the importance of NEK7 during a number of steps of mitosis and cytokinesis (Salem et al., 2010). The kinase activity of NEK7, which is controlled by NEK9 phosphorylation, reached at the maximum level at an M phase (Belham et al., 2003; Richards et al., 2009). As expected, *Nek7*$^{-/-}$ mouse embryonic fibroblast (MEF) cells revealed the choromosome number increase after prolonged passages of the cells in tissue culture. In addition, genome-wide analysis revealed mutations in the *Nek7*-coding sequence in human lung and ovary cancers (Sjoblom et al., 2006) and in situ RNA analysis on tissue microarrays demonstrated *NEK7* overexpression in the breast, colon, and larynx cancers (Capra et al., 2006). These data imply that NEK7 also seems to be an important factor for the regulation of G_2/M transition in cancer cells. Our study has indicated that WHSC1L1 is likely to regulate the expression of both *CCNG1* and *NEK7*. Knockdown of WHSC1L1 induced the G_2/M arrest of cancer cells, and we assume that deregulation of *CCNG1* and *NEK7*expression might have played critical roles in this phenomenon. As our data are not sufficient enough to conclude whether WHSC1L1 directly or indirectly regulates the expression of these genes through its methyltransferase activity, additional functional analysis is essential to predict the detailed mechanism of WHSCL1-dependent transcriptional regulation of *CCNG1* and *NEK7*expression.

ACKNOWLEDGMENTS

The authors thank Professor Gillian Murphy and the members of her laboratory for substantial technical support. They also thank Dr. Masanori Yoshimatsu, Dr. Motoko Unoki, Mr. Kazuhiro Maejima, and Ms. Haruka Sawada for technical assistance.

REFERENCES

Baek WK, Kim D, Jung N, Yi YW, Kim JM, Cha SD, Bae I, Cho CH. 2003. Increased expression of cyclin G1 in leiomyoma compared with normal myometrium. Am J Obstet Gynecol 188:634–639.

Belham C, Roig J, Caldwell JA, Aoyama Y, Kemp BE, Comb M, Avruch J. 2003. A mitotic cascade of NIMA family kinases. Nercc1/Nek9 activates the Nek6 and Nek7 kinases. J Biol Chem 278:34897–34909.

Capra M, Nuciforo PG, Confalonieri S, Quarto M, Bianchi M, Nebuloni M, Boldorini R, Pallotti F, Viale G, Gishizky ML, Draetta GF, Di Fiore PP. 2006. Frequent alterations in the expression of serine/threonine kinases in human cancers. Cancer Res 66:8147–8154.

Cerveira N, Correia C, Doria S, Bizarro S, Rocha P, Gomes P, Torres L, Norton L, Borges BS, Castedo S, Teixeira,MR. 2003. Frequency of NUP98-NSD1 fusion transcript in childhood acute myeloid leukaemia. Leukemia 17:2244–2247.

Chen DS, Zhu NL, Hung G, Skotzko MJ, Hinton DR, Tolo V, Hall FL, Anderson WF, Gordon EM. 1997. Retroviral vector-mediated transfer of an antisense cyclin G1 construct inhibits osteosarcoma tumor growth in nude mice. Hum Gene Ther 8:1667–1674.

Chesi M, Nardini E, Lim RS, Smith KD, Kuehl WM, Bergsagel PL. 1998. The t(4;14) translocation in myeloma dysregulates both FGFR3 and a novel gene, MMSET, resulting in IgH/MMSET hybrid transcripts. Blood 92:3025–3034.

Cho HS, Kelly JD, Hayami S, Toyokawa G, Takawa M, Yoshimatsu M, Tsunoda T, Field HI, Neal DE, Ponder BA, Nakamura Y, Hamamoto R. 2011a. Enhanced expression of EHMT2 is involved in the proliferation of cancer cells through negative regulation of SIAH1. Neoplasia 13:676–684.

Cho HS, Suzuki T, Dohmae N, Hayami S, Unoki M, Yoshimatsu M, Toyokawa G, Takawa M, Chen T, Kurash JK, Field HI, Ponder BA, Nakamura Y, Hamamoto R. 2011b. Demethylation of RB regulator MYPT1 by histone demethylase LSD1 promotes cell cycle progression in cancer cells. Cancer Res 71:1–6.

Douglas J, Hanks S, Temple IK, Davies S, Murray A, Upadhyaya M, Tomkins S, Hughes HE, Cole TR, Rahman N. 2003. NSD1 mutations are the major cause of Sotos syndrome and occur in some cases of Weaver syndrome but are rare in other overgrowth phenotypes. Am J Hum Genet 72:132–143.

Douglas J, Coleman K, Tatton-Brown K, Hughes HE, Temple IK, Cole TR, Rahman N. 2005. Evaluation of NSD2 and NSD3 in overgrowth syndromes. Eur J Hum Genet 13:150–153.

Gordon EM, Liu PX, Chen ZH, Liu L, Whitley MD, Gee C, Groshen S, Hinton DR, Beart RW, Hall FL. 2000. Inhibition of metastatic tumor growth in nude mice by portal vein infusions of matrix-targeted retroviral vectors bearing a cytocidal cyclin G1 construct. Cancer Res 60:3343–3347.

Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, Yagyu R, Nakamura Y. 2004. SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. Nat Cell Biol 6:731–740.

Hamamoto R, Silva FP, Tsuge M, Nishidate T, Katagiri T, Nakamura Y, Furukawa Y. 2006. Enhanced SMYD3 expression is essential for the growth of breast cancer cells. Cancer Sci 97:113–118.

Hayami S, Yoshimatsu M, Veerakumarasivam A, Unoki M, Iwai Y, Tsunoda T, Field HI, Kelly JD, Neal DE, Yamaue H, Ponder BA, Nakamura Y, Hamamoto R. 2010. Overexpression of the JmjC histone demethylase KDM5B in human carcinogenesis: Involvement in the proliferation of cancer cells through the E2F/RB pathway. Mol Cancer 9:59.

Hayami S, Kelly JD, Cho HS, Yoshimatsu M, Unoki M, Tsunoda T, Field HI, Neal DE, Yamaue H, Ponder BA, Nakamura Y, Hamamoto R. 2011. Overexpression of LSD1 contributes to human carcinogenesis through chromatin regulation in various cancers. Int J Cancer 128:574–586.

Jenuwein T, Allis CD. 2001. Translating the histone code. Science 293:1074–1080.

Kanaoka Y, Kimura SH, Okazaki I, Ikeda M, Nojima H. 1997. GAK: A cyclin G associated kinase contains a tensin/auxilin-like domain. FEBS Lett 402:73–80.

Kaneta Y, Kagami Y, Tsunoda T, Ohno R, Nakamura Y, Katagiri T. 2003. Genome-wide analysis of gene-expression profiles in chronic myeloid leukemia cells using a cDNA microarray. Int J Oncol 23:681–691.

Kikuchi T, Daigo Y, Katagiri T, Tsunoda T, Okada K, Kakiuchi S, Zembutsu H, Furukawa Y, Kawamura M, Kobayashi K, Imai K, Nakamura Y. 2003. Expression profiles of non-small cell lung cancers on cDNA microarrays: Identification of genes for prediction of lymph-node metastasis and sensitivity to anti-cancer drugs. Oncogene 22:2192–2205.

Kim SM, Kee HJ, Eom GH, Choe NW, Kim JY, Kim YS, Kim SK, Kook H, Seo SB. 2006. Characterization of a novel WHSC1-associated SET domain protein with H3K4 and H3K27 methyltransferase activity. Biochem Biophys Res Commun 345:318–323.

Kim S, Lee K, Rhee K. 2007a. NEK7 is a centrosomal kinase critical for microtubule nucleation. Biochem Biophys Res Commun 360:56–62.

Kim SM, Kee HJ, Choe N, Kim JY, Kook H, Seo SB. 2007b. The histone methyltransferase activity of WHISTLE is important for the induction of apoptosis and HDAC1-mediated transcriptional repression. Exp Cell Res 313:975–983.

Kimura SH, Tsuruga H, Yabuta N, Endo Y, Nojima H. 1997. Structure, expression, and chromosomal localization of human GAK. Genomics 44:179–187.

Kitahara O, Furukawa Y, Tanaka T, Kihara C, Ono K, Yanagawa R, Nita ME, Takagi T, Nakamura Y, Tsunoda T. 2001. Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. Cancer Res 61:3544–3549.

Li Y, Trojer P, Xu CF, Cheung P, Kuo A, Drury WJ, III, Qiao Q, Neubert TA, Xu RM, Gozani O, Reinberg D. 2009. The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. J Biol Chem 284:34283–34295.

Margueron R, Trojer P, Reinberg D. 2005. The key to development: Interpreting the histone code?Curr Opin Genet Dev 15:163–176.

Morishita M, di Luccio E. 2011. Structural insights into the regulation and the recognition of histone marks by the SET domain of NSD1. Biochem Biophys Res Commun 412:214–219.

Nagai T, Matsumoto N, Kurotaki N, Harada N, Niikawa N, Ogata T, Imaizumi K, Kurosawa K, Kondoh T, Ohashi H, Tsukahara M, Makita Y, Sugimoto T, Sonoda T, Yokoyama T, Uetake K, Sakazume S, Fukushima Y, Naritomi K. 2003. Sotos syndrome and haploinsufficiency of NSD1: Clinical features of intragenic mutations and submicroscopic deletions. J Med Genet 40:285–289.

Nakamura T, Furukawa Y, Nakagawa H, Tsunoda T, Ohigashi H, Murata K, Ishikawa O, Ohgaki K, Kashimura N, Miyamoto M, Hirano S, Kondo S, Katoh H, Nakamura Y, Katagiri T. 2004. Genome-wide cDNA microarray analysis of gene expression profiles in pancreatic cancers using populations of tumor cells and normal ductal epithelial cells selected for purity by laser microdissection. Oncogene 23:2385–2400.

O'Regan L, Fry AM. 2009. The Nek6 and Nek7 protein kinases are required for robust mitotic spindle formation and cytokinesis. Mol Cell Biol 29:3975–3990.

Olsburgh J, Harnden P, Weeks R, Smith B, Joyce A, Hall G, Poulsom R, Selby P, Southgate J. 2003. Uroplakin gene expression in normal human tissues and locally advanced bladder cancer. J Pathol 199:41–49.

Perez R, Wu N, Klipfel AA, Beart RW, Jr. 2003. A better cell cycle target for gene therapy of colorectal cancer: Cyclin G. J Gastrointest Surg 7:884–889.

Rayasam GV, Wendling O, Angrand PO, Mark M, Niederreither K, Song L, Lerouge T, Hager GL, Chambon P, Losson R. 2003. NSD1 is essential for early post-implantation development and has a catalytically active SET domain. EMBO J 22:3153–3163.

Reimer CL, Borras AM, Kurdistani SK, Garreau JR, Chung M, Aaronson SA, Lee SW. 1999. Altered regulation of cyclin G in human breast cancer and its specific localization at replication foci in response to DNA damage in p53+/+ cells. J Biol Chem 274:11022–11029.

Richards MW, O'Regan L, Mas-Droux C, Blot JM, Cheung J, Hoelder S, Fry AM, Bayliss R. 2009. An autoinhibitory tyrosine motif in the cell-cycle-regulated Nek7 kinase is released through binding of Nek9. Mol Cell 36:560–570.

Rio M, Clech L, Amiel J, Faivre L, Lyonnet S, Le Merrer M, Odent S, Lacombe D, Edery P, Brauner R, Raoul O, Gosset P, Prieur M, Vekemans M, Munnich A, Colleaux L, Cormier-Daire V. 2003. Spectrum of NSD1 mutations in Sotos and Weaver syndromes. J Med Genet 40:436–440.

Salem H, Rachmin I, Yissachar N, Cohen S, Amiel A, Haffner R, Lavi L, Motro B. 2010. Nek7 kinase targeting leads to early mortality, cytokinesis disturbance and polyploidy. Oncogene 29:4046–4057.

Seo HR, Lee DH, Lee HJ, Baek M, Bae S, Soh JW, Lee SJ, Kim J, Lee YS. 2006. Cyclin G1 overcomes radiation-induced G2 arrest and increases cell death through transcriptional activation of cyclin B1. Cell Death Differ 13:1475–1484.

Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE. 2006. The consensus coding sequences of human breast and colorectal cancers. Science 314:268–274.

Stec I, Wright TJ, van Ommen GJ, de Boer PA, van Haeringen A, Moorman AF, Altherr MR, den Dunnen JT. 1998. WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphy gene maps in the Wolf–Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. Hum Mol Genet 7:1071–1082.

Stec I, van Ommen GJ, den Dunnen JT. 2001. WHSC1L1, on human chromosome 8p11.2, closely resembles WHSC1 and maps to a duplicated region shared with 4p16.3. Genomics 76:5–8.

Strahl BD, Allis CD. 2000. The language of covalent histone modifications. Nature 403:41–45.

Takata R, Katagiri T, Kanehira M, Tsunoda T, Shuin T, Miki T, Namiki M, Kohri K, Matsushita Y, Fujioka T, Nakamura Y. 2005. Predicting response to methotrexate, vinblastine, doxorubicin, and cisplatin neoadjuvant chemotherapy for bladder cancers through genome-wide gene expression profiling. Clin Cancer Res 11:2625–2636.

Takawa M, Masuda K, Kunizaki M, Daigo Y, Takagi K, Iwai Y, Cho HS, Toyokawa G, Yamane Y, Maejima K, Field HI, Kobayashi T, Akasu T, Sugiyama M, Tsuchiya E, Atomi Y, Ponder BA, Nakamura Y, Hamamoto R. 2011. Validation of the histone methyltransferase EZH2 as a therapeutic target for various types of human cancer and as a prognostic marker. Cancer Sci 102:1298–1305.

Tanaka S, Diffley JF. 2002. Deregulated G1-cyclin expression induces genomic instability by preventing efficient pre-RC formation. Genes Dev 16:2639–2649.

Toyokawa G, Cho HS, Iwai Y, Yoshimatsu M, Takawa M, Hayami S, Maejima K, Shimizu N, Tanaka H, Tsunoda T, Field HI, Kelly JD, Neal DE, Ponder BA, Maehara Y, Nakamura Y, Hamamoto R. 2011a. The histone demethylase JMJD2B plays an essential role in human carcinogenesis through positive regulation of cyclin-dependent kinase 6. Cancer Prev Res (Phila) 4:2051–2061.

Toyokawa G, Cho HS, Masuda K, Yamane Y, Yoshimatsu M, Hayami S, Takawa M, Iwai Y, Daigo Y, Tsuchiya E, Tsunoda T, Field HI, Kelly JD, Neal DE, Maehara Y, Ponder BA, Nakamura Y, Hamamoto R. 2011b. Histone lysine methyltransferase Wolf–Hirschhorn syndrome candidate 1 is involved in human carcinogenesis through regulation of the Wnt pathway. Neoplasia 13:887–898.

Toyokawa G, Masuda K, Daigo Y, Cho HS, Yoshimatsu M, Takawa M, Hayami S, Maejima K, Chino M, Field HI, Neal DE, Tsuchiya E, Ponder BA, Maehara Y, Nakamura Y, Hamamoto R. 2011c. Minichromosome maintenance Protein 7 is a potential therapeutic target in human cancer and a novel prognostic marker of non-small cell lung cancer. Mol Cancer 10:65.

Turkmen S, Gillessen-Kaesbach G, Meinecke P, Albrecht B, Neumann LM, Hesse V, Palanduz S, Balg S, Majewski F, Fuchs S, Zschieschang P, Greiwe M, Mennicke K, Kreuz FR, Dehmel HJ, Podeck B, Kunze J, Tinschert S, Mundlos S, Horn D. 2003. Mutations in NSD1 are responsible for Sotos syndrome, but are not a frequent finding in other overgrowth phenotypes. Eur J Hum Genet 11:858–865.

Turner BM. 2000. Histone acetylation and an epigenetic code. Bioessays 22:836–845.

Wagner EJ, Carpenter PB. 2012. Understanding the language of Lys36 methylation at histone H3. Nat Rev Mol Cell Biol 13:115–126.

Wallard MJ, Pennington CJ, Veerakumarasivam A, Burtt G, Mills IG, Warren A, Leung HY, Murphy G, Edwards DR, Neal DE, Kelly JD. 2006. Comprehensive profiling and localisation of the matrix metalloproteinases in urothelial carcinoma. Br J Cancer 94:569–577.

Wang GG, Cai L, Pasillas MP, Kamps MP. 2007. NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. Nat Cell Biol 9:804–812.

Yissachar N, Salem H, Tennenbaum T, Motro B. 2006. Nek7 kinase is enriched at the centrosome, and is required for proper spindle assembly and mitotic progression. FEBS Lett 580: 6489–6495.

Yoshimatsu M, Toyokawa G, Hayami S, Unoki M, Tsunoda T, Field HI, Kelly JD, Neal DE, Maehara Y, Ponder BA, Nakamura Y, Hamamoto R. 2011. Dysregulation of PRMT1 and PRMT6, Type I arginine methyltransferases, is involved in various types of human cancers. Int J Cancer 128:562–573.

OPEN ACCESS Freely available online

PLoS one

Association of Common Variants in *TNFRSF13B*, *TNFSF13*, and *ANXA3* with Serum Levels of Non-Albumin Protein and Immunoglobulin Isotypes in Japanese

Wael Osman[1], Yukinori Okada[2,3], Yoichiro Kamatani[4], Michiaki Kubo[5], Koichi Matsuda[1], Yusuke Nakamura[1]*

1 Laboratory of Molecular Medicine, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, **2** Laboratory for Statistical Analysis, Center for Genomic Medicine, Institute of Physical and Chemical Research (Center for Genomic Medicine, RIKEN), Kanagawa, Japan, **3** Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan, **4** Centre d'Etude du Polymorphisme Humain, Paris, France, **5** Laboratory for Genotyping Development, Center for Genomic Medicine, RIKEN, Kanagawa, Japan

Abstract

We performed a genome-wide association study (GWAS) on levels of serum total protein (TP), albumin (ALB), and non-albumin protein (NAP). We analyzed SNPs on autosomal chromosomes using data from 9,103 Japanese individuals, followed by a replication study of 1,600 additional individuals. We confirmed the previously- reported association of *GCKR* on chromosome 2p23.3 with serum ALB (rs1260326, $P_{meta} = 3.1 \times 10^{-9}$), and additionally identified the significant genome-wide association of rs4985726 in *TNFRSF13B* on 17p11.2 with both TP and NAP ($P_{meta} = 1.2 \times 10^{-14}$ and 7.1×10^{-24}, respectively). For NAP, rs3803800 and rs11552708 in *TNFSF13* on 17p13.1 ($P_{meta} = 7.2 \times 10^{-15}$ and 7.5×10^{-10}, respectively) as well as rs10007186 on 4q21.2 near *ANXA3* ($P_{meta} = 1.3 \times 10^{-9}$) also indicated significant associations. Interestingly, *TNFRSF13B* and *TNFSF13* encode a tumor necrosis factor (TNF) receptor and its ligand, which together constitute an important receptor-ligand axis for B-cell homeostasis and immunoglobulin production. Furthermore, three SNPs, rs4985726, rs3803800, and rs11552708 in *TNFRSF13B* and *TNFSF13*, were indicated to be associated with serum levels of IgG ($P < 2.3 \times 10^{-3}$) and IgM ($P < 0.018$), while rs3803800 and rs11552708 were associated with IgA ($P < 0.013$). Rs10007186 in 4q21.2 was associated with serum levels of IgA ($P = 0.036$), IgM ($P = 0.019$), and IgE ($P = 4.9 \times 10^{-4}$). Our results should add interesting knowledge about the regulation of major serum components.

Citation: Osman W, Okada Y, Kamatani Y, Kubo M, Matsuda K, et al. (2012) Association of Common Variants in *TNFRSF13B*, *TNFSF13*, and *ANXA3* with Serum Levels of Non-Albumin Protein and Immunoglobulin Isotypes in Japanese. PLoS ONE 7(4): e32683. doi:10.1371/journal.pone.0032683

Editor: Vladimir N. Uversky, University of South Florida College of Medicine, United States of America

Received December 2, 2011; **Accepted** January 29, 2012; **Published** April 27, 2012

Funding: This work was supported by Leading Project for Personalized Medicine in the Ministry of Education, Culture, Sports, Science and Technology, Japan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: yusuke@ims.u-tokyo.ac.jp

Introduction

Serum proteins possess various biological functions such as hormones, enzymes, antibodies, and clotting agents, and some serve as valuable biomarkers that reflect several disease conditions. Major components of serum proteins are ALB (approximately 60%), globulins (mainly as γ-globulins, approximately 30%), and fibrinogens. Total serum protein levels range from 6.5 to 8.5 g/dl and show significant inter-individual variation. These variations are found to be influenced by environmental factors. However, genetic factors are also known to affect their levels although the range of genetic effects varies by the reports from 20% to 77% [1]. Genome-wide association studies (GWAS) recently demonstrated that serum levels of several proteins can be strongly influenced by common genetic variants through either *cis* or *trans* effects [2–4].

We previously reported the GWAS results for hematological and biochemical traits, including TP and ALB, in the Japanese population [5]. An associated SNP for TP, rs4273077 (*P*-value $= 4.5 \times 10^{-10}$), is located in an intron of *TNFRSF13B* (Tumor Necrosis Factor Receptor Superfamily member 13B), which encodes TACI (transmembrane activator and calcium-modulator and cytophilin interactor), one of three TNF-receptor family members (BAFF-R, TACI, and BCMA) [6]. However, since rs4273077 showed no significant association with the serum ALB level ($P = 0.089$), we suspected that this SNP would have genetic effects primarily on the levels of the non-albumin fraction. TACI is expressed mainly in activated B cells and binds with a high affinity to two TNF ligands; APRIL (a Proliferation-Inducing Ligand, encoded by *TNFSF13*), and BAFF (B Cell-Activating Factor, encoded by *TNFSF13B*) [7]. TACI is implicated in B- cell homeostasis (including B- cell survival, activation, and differentiation), immunoglobulin production, and antibody class switching [8–10]. Hence, the association of variants in *TNFRSF13B* with TP is likely to reflect the immunoglobulin serum levels.

The aim of this study is to identify the genetic variations associated with serum levels of non-albumin proteins (NAP), particularly those of immunoglobulins by GWAS of Japanese subjects.

Results

GWAS of Total Protein (TP), Albumin (ALB), and Non-albumin Protein (NAP)

We conducted a GWAS using genotyping data and clinical information on 9,103 individuals who had been collected in the BioBank Japan Project [11] (Table 1, Table S1). Genotyping was performed using Illumina Human610-Quad BeadChip (Illumina, CA, USA). After applying stringent quality control (QC) filters for selection of individuals and SNPs (Materials and Methods), we additionally performed whole-genome imputation analysis using the data of HapMap Phase II East Asian populations, and we obtained the information of 2,178,644 SNPs on autosomal chromosomes with minor allele frequencies (MAF) of ≥0.01 and *Rsq* of ≥0.7. We then evaluated the association of the SNPs with the adjusted χ^2 scores of serum levels of total protein (TP), albumin (ALB), and non-albumin protein (NAP). A Quantile-quantile (Q-Q) plot for each trait indicated low possibility of population stratification (inflation factors (λ_{GC}) for TP, ALB and NAP were 1.04, 1.02 and 1.02, respectively) (Figure S2).

Several SNPs with strong linkage disequilibrium (LD) ($r^2>0.8$) in intronic regions of *TNFRSF13B* on chromosome 17p11.2 showed significant associations with both TP and NAP (rs4985726, $P=2.8\times10^{-12}$ and 2.4×10^{-22}, respectively) (Table 2, Table S2, Figure 1A and 1B, and Figure 2A and 2B). In addition, rs3803800 and rs11552708 in coding regions of *TNFSF13* on chromosome 17p13.1 demonstrated significant associations with NAP ($P=1.8\times10^{-12}$ and 7.0×10^{-9}, respectively) (Table 2, Figure 1B, and Figure 2C).

Since *TNFSF13* encodes APRIL, a ligand of TACI encoded by *TNFRSF13B*, this ligand- receptor interaction is likely to play a critical role in regulation of the serum NAP levels. However, we did not find any synergistic effects between SNPs in the receptor and ligand on NAP levels.

Rs10007186 located near *ANXA3* (annexin A3) on chromosome 4q21.2 also revealed significant association with NAP ($P=3.3\times10^{-9}$; Table 2, Figure 1B, and Figure 2D), and a cluster of highly linked SNPs near the 5′ flanking region of *AFF3* (AF4/FMR2 family, member 3) on 2q11.2 indicated suggestive associations with NAP (rs4851274, $P=9.95\times10^{-8}$) (Table S2). For serum ALB, SNPs rs1260326 (in exon) and rs3817588 (in intron) in *GCKR* (glucokinase regulator) on 2p23.3 revealed significant associations ($P=3.4\times10^{-8}$ and 4.1×10^{-8}, respectively) (Figure 2E, Table 2, and Table S2).

Conditional logistic regression analysis for the SNPs on 17p13.1 indicated that both rs3803800 and rs11552708 conferred independent associations with NAP levels when adjusted for each other ($P<0.023$). These two SNPs were in strong LD ($D'=0.99$, $r^2=0.30$) and the haplotype analysis of these two SNPs identified that a haplotype (rs3803800 [A] – rs11552708 [G]) revealed stronger association with NAP than individual SNP ($P=2.59\times10^{-13}$) (Table S3). Similarly, rs1260326 and rs3817588 in *GCKR* exhibited independent associations with ALB levels ($P<0.022$), and were in LD ($D'=0.95$, $r^2=0.50$). Moreover, the haplotype (rs1260326 [C]–rs3817588 [C]) indicated stronger association with serum ALB ($P=2.83\times10^{-9}$) (Table S4). For the 17p11.2 and 4q21.2 loci, no SNP remained significant after accounting for the effect of marker the SNPs rs4985726, and rs10007186, respectively.

When we examined the genetic contribution of these variations for the traits, the combinations of the SNPs indicated above could explain nearly 0.5%, 2.3%, and 0.3% of variations in serum TP, NAP, and ALB, respectively.

Replication Study

To validate the GWAS results, we performed a replication study using an independent set of ~1,600 subjects from BioBank Japan [11] (Table 1). For each trait, we selected marker SNPs for the replication analysis at each locus that indicated the genome-wide significant level of 5.0×10^{-8} (rs4985726 in *TNFRSF13B*, rs3803800 in *TNFSF13*, rs1260326 in *GCKR*, and rs10007186 on 4q21.2). In addition, the two SNPs that remained significant after accounting for the effect of each marker SNP at two loci (rs11552708 in *TNFSF13* and rs3817588 in *GCKR*) were also further investigated.

SNPs rs4985726 in the *TNFRSF13B* locus as well as rs3803800 and rs11552708 in the *TNFSF13* locus revealed significant associations with both TP and NAP (Table 2). The association of rs1260326 in *GCKR* with serum ALB was also replicated ($P=0.029$; Table 2). Meta-analyses combining the GWAS and the replication study yielded stronger associations of these SNPs than the GWAS alone (Table 2 and Figure 2A, B, C, and E).

Table 1. Characteristics of the examined proteins.

	TP		ALB		NAP		IgG *	IgA *	IgM *	IgE *
	GWAS	Replication	GWAS	Replication	GWAS	Replication				
No.	9,090	1,626	9,103	1,607	9,077	1,629	1,794	1,675	1,649	549
M±S.D [a]	7.10± 0.50	7.06± 0.73	4.25± 0.35	4.00± 0.51	2.85± 0.42	3.07± 0.57	1.44± 0.61	0.27± 0.15	0.11± 0.07	1306.54± 5598.06
Age [b]	69.52± 10.44	59.52± 15.43	69.52± 10.44	59.54± 15.39	69.51± 10.44	59.48± 15.52	59.70± 15.46	59.38± 15.73	59.42± 15.57	62.54± 18.61
Female %	37.45	45.08	37.41	45.12	37.46	45.12	55.30	54.57	54.88	63.93
BMI [b]	22.91± 3.45	23.31± 5.67	22.91± 3.45	23.34± 5.69	22.91± 3.45	23.29± 5.67	23.17± 5.00	23.20± 5.09	23.19± 5.07	22.73± 4.22
Smokers %	42.11	51.91	42.11	52.15	42.05	51.81	51.90	51.82	52.27	48.63
Drinkers %	29.37	51.97	29.37	52.08	29.40	51.81	51.00	50.81	50.82	41.35

[a]M±S.D: mean value±standard deviation of each protein is indicated in g/dl except for IgE, which is indicated as IU/ml.
[b]Age and body mass index (BMI) are indicated as mean values±standard deviation.
*Log-transformed values were applied in the analysis.
Abbreviations: GWAS: genome-wide association study, TP: total protein, ALB: albumin, NAP: non-albumin protein.
doi:10.1371/journal.pone.0032683.t001

Table 2. Summary results of the GWAS and the replication study of TP, ALB, and NAP.

Trait	SNP	Chr: Position	Nearest Gene	A1/A2[a]	MAF	GWAS		Replication		Meta analysis		% variance explained
						Effect[b] (s.e)	P[c]	Effect[b] (s.e)	P[c]	Effect[b] (s.e)	P[c]	
TP	rs4985726*	17:16804363	*TNFRSF13B*	C/G	0.375	0.108 (0.015)	2.8×10^{-12}	0.100 (0.030)	0.0010	0.107 (0.0138)	1.2×10^{-14}	0.53
ALB	rs1260326	2:27584444	*GCKR*	T/C	0.445	−0.082 (0.015)	3.4×10^{-8}	−0.070 (0.032)	0.029	−0.080 (0.014)	3.1×10^{-9}	0.32
NAP	rs4985726*	17:16804363	*TNFRSF13B*	C/G	0.375	0.148 (0.015)	2.4×10^{-22}	0.090 (0.028)	0.0013	0.135 (0.013)	7.1×10^{-24}	1.03
	rs3803800	17:7403693	*TNFSF13*	G/A	0.311	0.108 (0.015)	1.8×10^{-12}	0.090 (0.029)	0.0022	0.104 (0.013)	7.2×10^{-15}	0.53
	rs11552708	17:7403279	*TNFSF13*	G/A	0.401	−0.084 (0.015)	7.0×10^{-9}	−0.070 (0.027)	0.0091	−0.081 (0.013)	7.5×10^{-10}	0.36
	rs10007186*	4:79808069	*ANXA3*	T/C	0.307	0.095 (0.016)	3.3×10^{-9}	0.053 (0.029)	0.065	0.085 (0.014)	1.3×10^{-9}	0.38

[a]A1/A2: major/minor alleles.
[b]The effect of the minor allele on the normalized values based on an additive genetic model.
[c]For the GWAS and replication analysis, *P*-values were obtained by linear regression test model, for the Meta analysis by inverse-variance method.
*SNPs obtained by whole-genome imputation analysis.
Abbreviations: GWAS: genome-wide association study, MAF: minor allele frequency, TP: total protein, ALB: albumin, NAP: non-albumin protein, s.e: standard error.
doi:10.1371/journal.pone.0032683.t002

Rs10007186 near *ANXA3* revealed a suggestive association in the replication study ($P=0.065$), and meta-analyses indicated that the association was unlikely to be false positive ($P=1.3\times10^{-9}$) (Table 2 and Figure 2D).

Association of the SNPs Identified in the GWAS of NAP with Serum Immunoglobulin Isotypes

Immunoglobulin isotypes constitute the major components of NAP. Hence, we further examined the NAP-associated SNPs in the GWAS (*TNFRSF13B*, *TNFSF13*, and *ANXA3*) for the association with various serum immunoglobulins using the samples in BioBank Japan [11] (IgG: $n=1{,}794$, IgA: $n=1{,}675$, IgM: $n=1{,}649$, and IgE: $n=549$; Table 1).

We found significant associations of rs4985726 in *TNFRSF13B* as well as rs3803800 and rs11552708 in *TNFSF13* with serum levels of IgG ($P<0.0023$) and IgM($P<0.018$) (Table 3). For IgA, rs3803800 and rs11552708 in *TNFSF13* also revealed the significant association ($P<0.013$), while rs4985726 in *TNFRSF13B* revealed no significant association ($P=0.099$) (Table 3). Rs10007186 near *ANXA3* indicated significant association with IgA ($P=0.036$), IgM ($P=0.019$), and IgE ($P=4.9\times10^{-4}$). However, these associated SNPs explained only 1.4%, 0.9%, 1.3%, and 2.0% of the variances of log-transformed values of serum IgG, IgA, IgM, and IgE, respectively.

Discussion

On the basis of the information of 10,716 Japanese individuals, we identified one genetic locus (*TNFRSF13B*) on chromosome 17p11.2 associated with both TP and NAP, two loci (*TNFSF13* on 17p13.1 and a region near *ANXA3* on 4q21.2) associated with NAP, and one locus (*GCKR*) on 2p23.3 associated with ALB at the level of genome-wide significance.

The marker SNP rs4985726 shows association with TP and NAP is located in an intron of *TNFRSF13B* on chromosome 17p11.2. A possible mechanism for its association with these traits could be explained by its strong LD with rs34562254 ($D'=1$, $r^2=0.97$), which exhibits a missense variation (C>T, Pro251Leu) located in the intracellular domain of the receptor molecule. The *in silico* prediction of the amino acid substitution by rs34562254 in the PolyPhen-2 and SNPinfo database [12,13] suggested a "probably damaging" effect on the protein structure.

The SNPs in *TNFSF13* (encoding APRIL) that identified as being associated with NAP are missense variants; rs3803800 (A>G, Asn96Ser), and rs11552708 (G>A, Gly67Arg). APRIL was first described as having a promoter function for tumor-cell proliferation and survival [14]. APRIL is cleaved in the Golgi apparatus by furin at its 104Arg/105Ala site [15], and interestingly, rs3803800 is closely located to this cleavage site. Hence, this SNP might affect the cleavage affinity. Another possibility is the effect on splicing, because both SNPs are predicted to be located within binding sites of splicing regulatory elements [13]. However, further investigation should be required to address these possibilities.

The SNP rs4985726 in *TNFRSF13B* as well as rs3803800 and rs11552708, in *TNFSF13* also revealed significant associations with serum levels of IgG, IgA, and IgM. It is notable that the two genes encode a TNF-receptor and ligand axis that plays important roles for mediating antibody class switching and regulating immunoglobulin production [8,9]. Furthermore, knockout mice of either *TNFRSF13B* or *TNFSF13* presented a common phenotype of the IgA deficiency with impaired antibody response to T cell-independent antigens [16]. In addition, germ-line mutations in *TNFRSF13B* were reported in cases of common variable immunodeficiency (CVID; MIM # 607594) and selective IgA deficiency (IGAD; MIM # 137100) [17]. The combination of these significant statistical and biological evidences would suggest that the association of these two loci with NAP reflect at least their associations with regulation of serum immunoglobulin levels. It is also known that immunoglobulins are the major components of NAP, which provides compelling evidence for our results. The facts that both SNPs rs3803800 [A] and rs11552708 [G] in *TNFSF13* were reported to be associated with the susceptibility to the Systemic Lupus Erythematosus (SLE) in the Japanese population and that high serum APRIL was detected in the sera of individuals with the rs3803800 [A]–rs11552708 [G] haplotype [18] further support the significance of these SNPs in the regulation of immunoglobulin production. In this study, we observed that possession of two copies of SLE-risk alleles was associated with higher serum levels of NAP, IgG, IgA, and IgM (Figure S3), providing a good example of genetic loci that influence both quantitative traits and susceptibility to complex diseases.

Rs10007186, which was associated with NAP ($P_{meta}=1.3\times10^{-9}$) is located about 57.4 kb downstream of *ANXA3* encoding annexin A3, a member of annexin family of calcium-

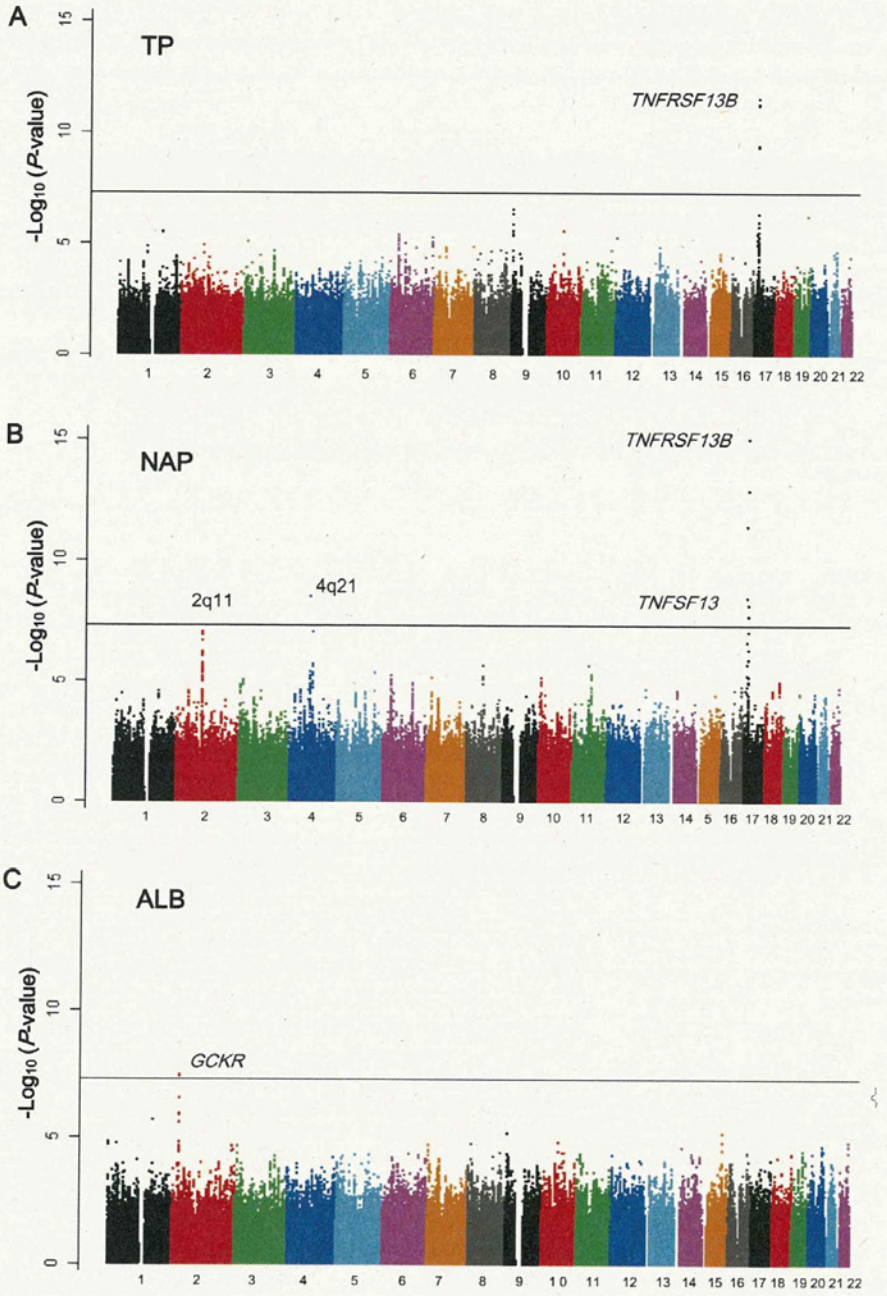

Figure 1. Manhattan plots for the GWAS of (A) TP, (B) NAP, and (C) ALB. SNPs were plotted based on their physical chromosomal positions (horizontal axis) together with their $-\log_{10}$ (*P*-values) in the GWAS (vertical axis). The black horizontal line shows the genome-wide significance threshold of $P = 5.0\times10^{-8}$. The SNPs for which *P*-values were smaller than 1.0×10^{-15} are indicated at the upper limit of the plots.
doi:10.1371/journal.pone.0032683.g001

dependent phospholipid-binding proteins [19]. Annexin A3 was found to be translocated into phagosomes in dendritic cells [20], which are antigen-presenting cells that serve as messengers between the innate and adaptive immune response, and play a key role in allergic, inflammatory, and autoimmune conditions. In addition, annexin A3 was also found to be associated with neutrophil granule membranes [21], where it can play a regulatory role in calcium-dependent granule secretions that contribute to acute inflammation and chronic tissue destruction. The association of rs10007186 with IgA, IgM, and IgE, would suggest additional biological roles of annexin A3 in the immune response.

We also confirmed the association of SNPs in *GCKR* with serum ALB levels (rs1260326, $P_{meta} = 3.1\times10^{-9}$). Rs1260326 is a missense variant (T>C, Leu446Pro) and predicted to cause a damaging effect on the protein structure. *GCKR* is a locus frequently associated with several metabolic traits [4,22–24] and rs1260326 has been reported to be associated with serum triglycerides [4].

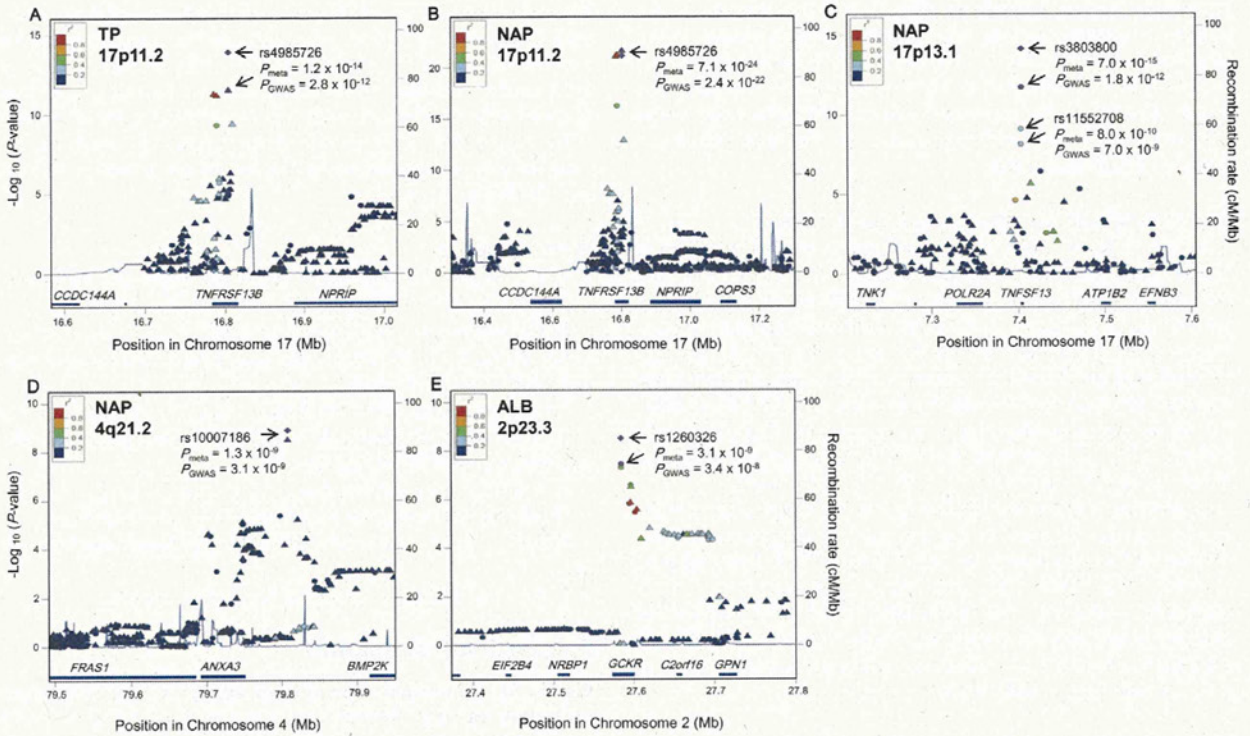

Figure 2. Regional plots for the associations of the SNPs in the GWAS stage of TP, ALB and NAP. SNPs plotted with their $-\log_{10}$ (*P*-values) in the GWAS based on their physical chromosomal positions. Genotyped SNPs are indicated as circles, while imputed SNPs are indicated as triangles. The color scheme indicated the linkage disequilibrium displayed as r^2 values between all SNPs and the top-ranked SNP in each plot. The tested trait, chromosomal locus, and the top-ranked SNPs (in purple color) in the GWAS and combined analyses together with their *P*-values are shown in each plot. The blue lines represent the recombination rates estimated based on HapMap Phase II database. The plots were drawn using Locus Zoom software.
doi:10.1371/journal.pone.0032683.g002

As a conclusion, the present study identified genetic loci that influence the inter-individual variation in serum levels of TP, ALB, and NAP. The loci associated with NAP encompass genes encoding a TNF-receptor and its ligand, which are implicated in biological roles in the immune system, and their associations with immunoglobulin isotypes were demonstrated here. Our results should add novel insight toward understanding the genetic background contributing to the regulation of the serum levels of NAP and its major components.

Materials and Methods

Study Cohorts

For the GWAS, 9,103 subjects derived from 10 disease cohorts (colorectal cancer, breast cancer, prostate cancer, lung cancer, gastric cancer, diabetes mellitus, peripheral artery disease, atrial fibrillation, ischemic stroke, and myocardial infarction) were selected, and for the replication study, we used data from >1600 independent individuals selected from the BioBank Japan Project [11] (Table 1 and Table S1). For immunoglobulin isotypes analyses, the data from ~1,600 additional individuals in BioBank

Table 3. Association of the SNPs in the GWAS of the NAP with immunoglobulin isotypes.

SNP	Gene	IgG			IgA			IgM			IgE		
		Effect [a] (s.e)	P [b]	%EV	Effect [a] (s.e)	P [b]	%EV	Effect [a] (s.e)	P [b]	%EV	Effect [a] (s.e)	P [b]	%EV
rs4985726	*TNFRSF13B*	0.071 (0.022)	1.4×10^{-3}	0.51	0.049 (0.030)	0.099	–	−0.090 (0.032)	5.9×10^{-3}	0.40	0.039 (0.064)	0.54	–
rs3803800	*TNFSF13*	−0.074 (0.024)	2.2×10^{-3}	0.47	−0.086 (0.031)	6.2×10^{-3}	0.39	−0.082 (0.034)	0.018	0.29	−0.117 (0.067)	0.080	–
rs11552708	*TNFSF13*	0.067 (0.022)	2.3×10^{-3}	0.46	0.072 (0.029)	0.013	0.31	0.078 (0.032)	0.014	0.31	0.059 (0.060)	0.33	–
rs10007186	*ANXA3*	−0.018 (0.022)	0.42	–	−0.063 (0.030)	0.036	0.20	−0.078 (0.033)	0.019	0.27	0.200 (0.057)	4.9×10^{-4}	2.02

[a]The effect of the minor alleles on the standardized values.
[b]*P*-values for the associations of SNPs with each normalized immunoglobulin isotype obtained by using a linear regression model.
Abbreviations: s.e: standard error, %EV: percentage of the explanatory variance.
doi:10.1371/journal.pone.0032683.t003

Japan [11] was used (Table 1). The clinical information for the samples is updated annually using a standard questionnaire in the 66 hospitals participating in the project. Written informed consent was obtained from all subjects. The research project was approved by the ethical committees in the Institute of Medical Science, the University of Tokyo, and the Center of Genomic Medicine, RIKEN, Yokohama, Japan.

Genotyping and Quality Control (Q.C) Filters

In the GWAS, SNPs were genotyped using the Illumina HumanHap610-Quad BeadChip (Illumina, CA, USA). After the exclusion of samples with call rates of <0.98, we excluded closely related individuals (in 1st or 2nd degree kinships) using identity-by-descent (IBD) evaluated by PLINK version 1.0.6 [25]. We also excluded individuals who were outliers in the cluster analysis using the principle component analysis performed by EIGENSTRAT 3.0 along with HapMap Phase II populations (Figure S1). In addition, SNPs with call rates of <0.99, MAF of <0.01 and Hardy Weinberg equilibrium of $P<1.0\times10^{-7}$ were excluded.

Genotyping data of the SNPs selected for replication analyses and for testing with immunoglobulin levels were generated using multiplex PCR- based Invader Assay (Third Wave Technologies, Madison, WI, USA) [26]. Genotypes were judged by visual inspection, following the application of QC measures of individuals' call rates of >98% and SNPs call rates of >99% of individuals. We could not obtain the genotype data of rs3817588 in *GCKR* using the Invader assay.

Whole-genome Imputation of Genotypes

We performed whole-genome imputation of the GWAS subjects in a two-step procedure, as described elsewhere [27]. HapMap phase II Japanese (JPT) and Han Chinese (CHB) individuals (release 24) were adopted as reference panels. We excluded the imputed SNPs with MAF of <0.01 or *Rsq* of <0.7. As a result, a total of 2,178,644 SNPs on autosomal chromosomes were used for the GWAS.

Statistical Analysis

We obtained the non-transformed values of TP, ALB and NAP (mg/dl) for the subjects from the clinical information stored in BioBank Japan [11], and adjusted them in linear regression models with age, gender, body mass index (BMI), smoking, drinking status, and affection status of the disease as covariates. The residuals were then normalized as Z scores and subjects with Z scores of <-4 or >4 were removed from each trait analysis. The associations of the SNPs with Z scores were evaluated in linear regression models assuming additive effects of allele dosages, using mach2qtl software. The same methods of data normalization and statistical models were applied for the replication analyses and for testing the association with common log-transformed values of immunoglobulin isotypes (IgG IgA, IgM, and IgE). Meta-analyses of the GWAS and the replication study were performed using the inverse-variance method assuming a fixed-effects model.

The significance level used was 5×10^{-8} in the GWAS stage. For the replication stage, we considered 0.05 as significant for the association of rs4985726 with TP and rs1260326 with ALB. For the association of SNPs rs4985726 in TNFRSF13B, rs3803800 and rs11552708 in TNFSF13 with NAP, 0.017 (0.05/3) was considered to be significant. These significance levels represent the Bonferroni correction for multiple statistical tests. In addition, we set a level of 0.05 to consider the association of the selected SNPs with immunoglobulin isotypes as significant.

The haplotype analyses were performed using the Haplo Stats package (version 1.4.0) implemented in *R* statistical software. Epistatic effects of the SNPs in *TNFRSF13B* and *TNFSF13* were evaluated using a linear regression model incorporating the product of the allele dosages of the SNPs in the loci as an independent variable. All statistical analyses including haplotype analyses were performed using the *R* statistical software version 2.9.1 except for genome-wide linear regression analyses. LD analyses were performed using Haploview 4.2 software, PLINK, and the SNAP database.

Web Resources

The URLs for the data presented in this paper are as follows:

The BioBank Japan Project, http://biobankjp.org/

PLIKN software, http://pngu.mgh.harvard.edu/purcell/plink/

EIGENSTRAT software, http://genepath.med.harvard.edu/reich/EIGENSTRAT.htm

The International HapMap Project, http://www.hapmap.org/

MACH and mach2qtl software, http://www.sph.umich.edu/csg/abecasis/MaCH/index.html

R statistical environment, http://www.r-project.org/

Haploview software, www.broad.mit.edu/mpg/haploview/

SNAP, http://www.broadinstitute.org/mpg/snap/ldsearch.php

Locus Zoom, http://csg.sph.umich.edu/locuszoom/

Supporting Information

Figure S1 Principal component analysis Plot of cohorts included in the GWAS. All individuals who were finally incorporated in the GWAS together with the four populations in the HapMap Phase II database (Japanese: JPT; Han Chinese: CHB; Africans: YRI, and European: CEU) were plotted based on the first two eigenvectors.
(PDF)

Figure S2 Quantile-Quantile (Q-Q) plots for the GWAS of (A) TP, (B) NAP, and (C) ALB. The inflation factor, λ_{GC}, for the analysis is shown in the legend of each plot. The SNPs for which *P*-values were smaller than 1.0×10^{-15} are indicated at the upper limit of the plots.
(PDF)

Figure S3 Relationship between the genotypes of SNPs identified in the study and the levels of tested proteins: (A) rs4985726, (B) rs3803800, (C) rs11552708, (D) rs10007186, and (E) rs1260326. For each box plot, the bold line indicates the median value which is the 50th quartile. The limits of each box are the 25th and 75th quartiles.
(PDF)

Table S1 Characteristics of the GWAS cohorts.
(DOC)

Table S2 SNPs showed suggestive associations with each examined trait ($P<1.0\times10^{-6}$).
(DOC)

Table S3 Haplotype analysis of rs3803800 and rs11552708 in *TNFSF13* in association with NAP.
(DOC)

Table S4 Haplotype analysis of rs1260326 and rs3817588 in *GCKR* in association with ALB.
(DOC)

Acknowledgments

We thank all subjects participating in this study for providing their clinical data and DNA samples. We acknowledge the staff in the BioBank Japan

Project [11] and the members of the Laboratory of the Genotyping Development at RIKEN Center for Genomic Medicine for their technical support.

Author Contributions

Conceived and designed the experiments: YN KM YO YK MK WO. Performed the experiments: WO MK. Analyzed the data: YO YK WO. Contributed reagents/materials/analysis tools: YN KM YO MK. Wrote the paper: WO YO YN. Summarized the whole results: WO.

References

1. Rahmioglu N, Andrew T, Cherkas L, Surdulescu G, Swaminathan R, et al. (2009) Epidemiology and genetic epidemiology of the liver function test proteins. PLoS One 4: e4435.
2. Melzer D, Perry J, Hernandez D, Corsi A, Stevens K, et al. (2008) A genome-wide association study identifies protein quantitative trait loci (pQTLs). PLoS Genet 4: e1000072.
3. Weidinger S, Gieger C, Rodriguez E, Baurecht H, Mempel M, et al. (2008) Genome-wide scan on total serum IgE levels identifies FCER1A as novel susceptibility locus. PLoS Genet 4: e1000166.
4. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, et al. (2009) Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet 41: 35–46.
5. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, et al. (2010) Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat Genet 42: 210–215.
6. Bossen C, Schneider P (2006) BAFF, APRIL and their receptors: structure, function and signaling. Semin Immunol 18: 263–275.
7. Mackay F, Kalled S (2002) TNF ligands and receptors in autoimmunity: an update. Curr Opin Immunol 14: 783–790.
8. Castigli E, Wilson S, Scott S, Dedeoglu F, Xu S, et al. (2005) TACI and BAFF-R mediate isotype switching in B cells. J Exp Med 201: 35–39.
9. Sakurai D, Hase H, Kanno Y, Kojima H, Okumura K, et al. (2007) TACI regulates IgA production by APRIL in collaboration with HSPG. Blood 109: 2961–2967.
10. Mackay F, Schneider P (2008) TACI, an enigmatic BAFF/APRIL receptor, with new unappreciated biochemical and biological properties. Cytokine Growth Factor Rev 19: 263–276.
11. Nakamura Y (2007) The BioBank Japan Project. Clin Adv Hematol Oncol 5: 696–697.
12. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. (2010) A method and server for predicting damaging missense mutations. Nat Methods 7: 248–249.
13. Xu Z, Taylor JA (2009) SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res 37: W600–605.
14. Hahne M, Kataoka T, Schröter M, Hofmann K, Irmler M, et al. (1998) APRIL, a new ligand of the tumor necrosis factor family, stimulates tumor cell growth. J Exp Med 188: 1185–1190.
15. López-Fraga M, Fernández R, Albar JP, Hahne M (2001) Biologically active APRIL is secreted following intracellular processing in the Golgi apparatus by furin convertase. EMBO Rep 2: 945–951.
16. Castigli E, Scott S, Dedeoglu F, Bryce P, Jabara H, et al. (2004) Impaired IgA class switching in APRIL-deficient mice. Proc Natl Acad Sci U S A 101: 3903–3908.
17. Castigli E, Wilson S, Garibyan L, Rachid R, Bonilla F, et al. (2005) TACI is mutant in common variable immunodeficiency and IgA deficiency. Nat Genet 37: 829–834.
18. Koyama T, Tsukamoto H, Masumoto K, Himeji D, Hayashi K, et al. (2003) A novel polymorphism of the human APRIL gene is associated with systemic lupus erythematosus. Rheumatology (Oxford) 42: 980–985.
19. Gerke V, Moss SE (2002) Annexins: from structure to function. Physiol Rev 82: 331–371.
20. Larsson M, Majeed M, Ernst JD, Magnusson KE, Stendahl O, et al. (1997) Role of annexins in endocytosis of antigens in immature human dendritic cells. Immunology 92: 501–511.
21. Rosales JL, Ernst JD (1997) Calcium-dependent neutrophil secretion: characterization and regulation by annexins. J Immunol 159: 6195–6202.
22. Ridker PM, Pare G, Parker A, Zee RY, Danik JS, et al. (2008) Loci related to metabolic-syndrome pathways including LEPR,HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study. Am J Hum Genet 82: 1185–1192.
23. Kolz M, Johnson T, Sanna S, Teumer A, Vitart V, et al. (2009) Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. PLoS Genet 5: e1000504.
24. Orho-Melander M, Melander O, Guiducci C, Perez-Martinez P, Corella D, et al. (2008) Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. Diabetes 57: 3112–3121.
25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.
26. Olivier M (2005) The Invader assay for SNP genotyping. Mutat Res 573: 103–110.
27. Okada Y, Hirota T, Kamatani Y, Takahashi A, Ohmiya H, et al. (2011) Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. PLoS Genet 7: e1002067.

nature genetics

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

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

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

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

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

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

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

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

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

Table 1 Characteristics of study populations

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

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

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

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

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

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

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

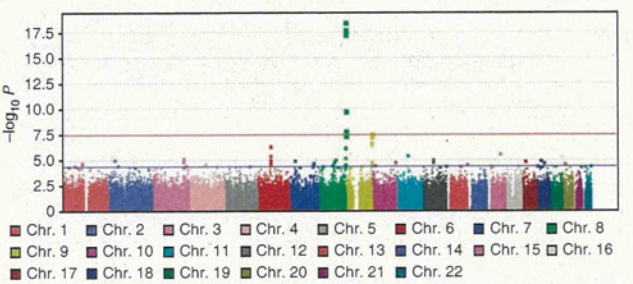

Figure 1 Manhattan plot showing genome-wide *P* values of association. The minimum *P* values under three genetic models (additive, recessive and dominant) were obtained by logistic regression analysis with adjustment for age and gender. The *y* axis shows the $-\log_{10} P$ values of 480,327 SNPs, and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

Table 2 Results of association analyses of duodenal ulcer

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

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

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

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

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

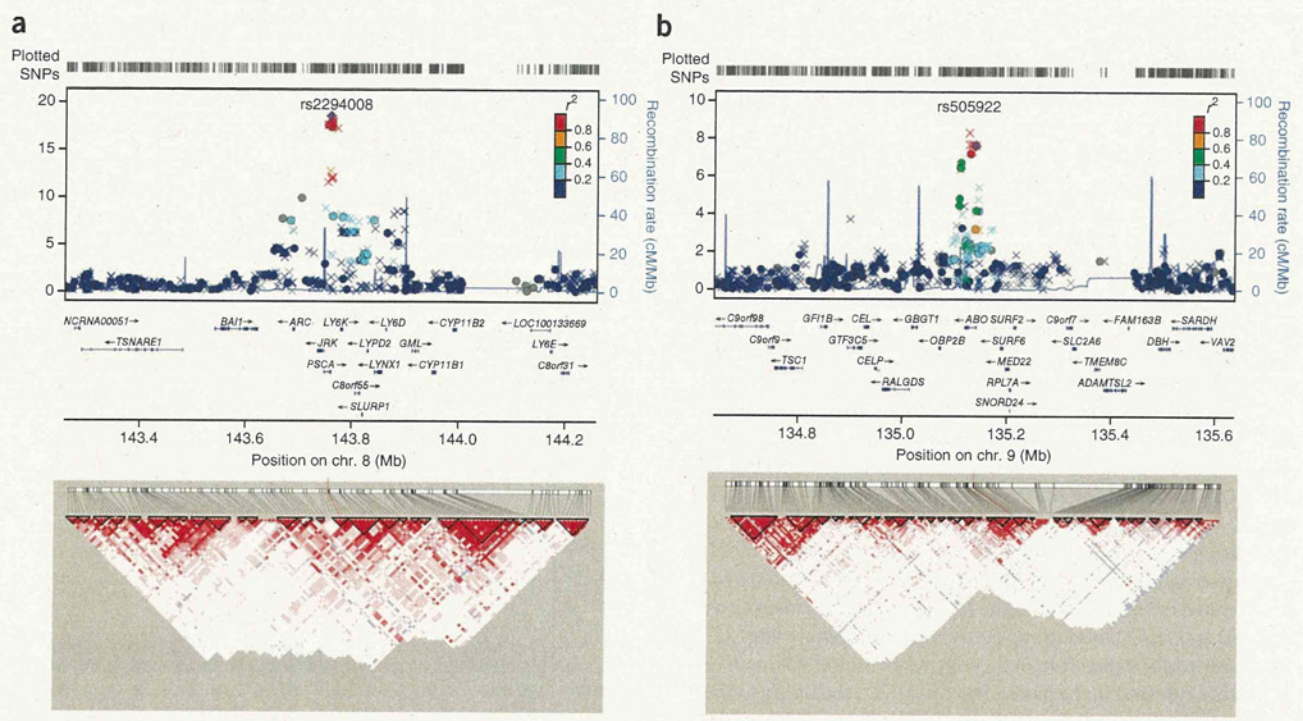

Figure 2 Regional association plots. (**a**,**b**) Data are shown for the associated regions on chromosome 8 including rs2294008 (in *PSCA*) (**a**) and on chromosome 9 including rs505922 (in *ABO*) (**b**). Upper, *P* values of genotyped SNPs (circle) and imputed SNPs (crosses) are plotted (as $-\log_{10}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

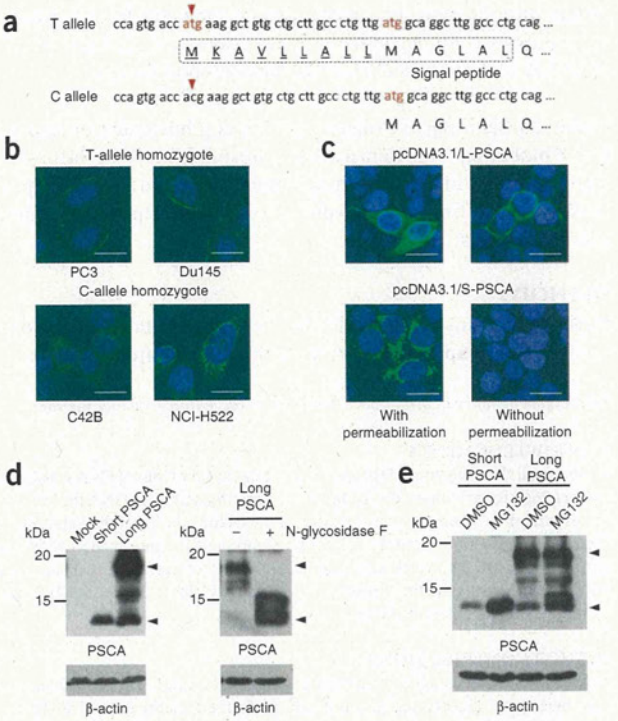

Figure 3 Effects of rs2294008 on subcellular localization and stability of PSCA protein. (**a**) The genomic structure around rs2294008 and corresponding amino-acid sequences. Arrowheads indicate the location of rs2294008. Potential start codons are shown in red. Underlined amino acids are unique to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

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

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

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

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

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

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

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

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

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

METHODS

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

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

ACKNOWLEDGMENTS

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

AUTHOR CONTRIBUTIONS

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

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

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