nature genetics

Genome-wide association analyses in east Asians identify new susceptibility loci for colorectal cancer

Wei-Hua Jia[1,16], Ben Zhang[2,16], Keitaro Matsuo[3], Aesun Shin[4], Yong-Bing Xiang[5], Sun Ha Jee[6], Dong-Hyun Kim[7], Zefang Ren[1], Qiuyin Cai[2], Jirong Long[2], Jiajun Shi[2], Wanqing Wen[2], Gong Yang[2], Ryan J Delahanty[2], Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO)[8], Colon Cancer Family Registry (CCFR)[8], Bu-Tian Ji[9], Zhi-Zhong Pan[1], Fumihiko Matsuda[10], Yu-Tang Gao[5], Jae Hwan Oh[11], Yoon-Ok Ahn[12], Eun Jung Park[6], Hong-Lan Li[5], Ji Won Park[11], Jaeseong Jo[6], Jin-Young Jeong[7], Satoyo Hosono[3], Graham Casey[13], Ulrike Peters[14,15], Xiao-Ou Shu[2], Yi-Xin Zeng[1,17] & Wei Zheng[2,17]

To identify new genetic factors for colorectal cancer (CRC), we conducted a genome-wide association study in east Asians. By analyzing genome-wide data in 2,098 cases and 5,749 controls, we selected 64 promising SNPs for replication in an independent set of samples, including up to 5,358 cases and 5,922 controls. We identified four SNPs with association *P* values of 8.58×10^{-7} to 3.77×10^{-10} in the combined analysis of all east Asian samples. Three of the four were replicated in a study conducted in 26,060 individuals of European descent, with combined *P* values of 1.22×10^{-10} for rs647161 (5q31.1), 6.64×10^{-9} for rs2423279 (20p12.3) and 3.06×10^{-8} for rs10774214 (12p13.32 near the *CCND2* gene), derived from meta-analysis of data from both east Asian and European-ancestry populations. This study identified three new CRC susceptibility loci and provides additional insight into the genetics and biology of CRC.

CRC is one of the most commonly diagnosed malignancies in east Asia and many other parts of the world[1]. Genetic factors have an important role in the etiology of both sporadic and familial CRC[2]. However, less than 6% of CRC cases can be explained by rare, high-penetrance variants in the CRC susceptibility genes identified to date, such as the *APC*, *SMAD4*, *AXIN2*, *BMPR1A*, *POLD1*, *STK11*, *MUTYH* and DNA mismatch repair genes[2]. Over the past two decades, many candidate gene studies have evaluated common genetic risk factors for CRC; only a few of these have been replicated in subsequent studies[3]. Recent genome-wide association studies (GWAS) have identified approximately 15 common genetic susceptibility loci for CRC[4–12]. However, these newly identified genetic factors, along with known high-penetrance variations in CRC susceptibility genes, explain less than 15% of the heritability for this common malignancy[10,11]. Furthermore, with the exception of a small study conducted in Japan[12], all other GWAS have been conducted in populations of European ancestry, which differ from other populations in certain features of genetic architecture. Many of the variants discovered in populations of European ancestry show only weak or no association with CRC in other ancestry groups[13]. Therefore, additional GWAS are needed, particularly in populations not of European ancestry, to fully uncover the genetic basis for CRC susceptibility.

In 2009, we initiated the Asia Colorectal Cancer Consortium (ACCC), a GWAS in east Asians, to search for previously unknown genetic risk factors for CRC. The discovery stage (stage 1) consisted of five GWAS conducted in China, Korea and Japan, including 2,293 CRC cases and 5,780 controls (**Supplementary Table 1**). Cases and controls were genotyped using several SNP arrays, including the Affymetrix Genome-Wide Human SNP Array 6.0 (906,602 SNPs), the Affymetrix Genome-Wide Human SNP Array 5.0 (443,104 SNPs), the Illumina Infinium HumanHap610 BeadChip (592,044 SNPs), the Illumina Human610-Quad BeadChip (620,901 SNPs) and the Illumina HumanOmniExpress BeadChip (729,462 SNPs) (**Supplementary Table 1**). After quality control exclusions as described previously[14–17], 2,098 cases and 5,749 controls remained for this study (**Supplementary Tables 1** and **2**). Also excluded from the analyses were SNPs with call rate of <95%, genotype concordance rate of <95%

[1]State Key Laboratory of Oncology in South China, Cancer Center, Sun Yat-sen University, Guangzhou, China. [2]Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. [3]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan. [4]Molecular Epidemiology Branch, National Cancer Center, Goyang-si, Korea. [5]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China. [6]Institute for Health Promotion, Department of Epidemiology and Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Korea. [7]Department of Social and Preventive Medicine, Hallym University College of Medicine, Okcheon-dong, Korea. [8]A complete list of members is provided in the Acknowledgements. [9]Division of Cancer Epidemiology & Genetics, National Cancer Institute, Bethesda, Maryland, USA. [10]Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [11]Center for Colorectal Cancer, National Cancer Center, Goyang-si, Korea. [12]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea. [13]Department of Preventive Medicine, University of Southern California, Los Angeles, California, USA. [14]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [15]Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington, USA. [16]These authors contributed equally to this work. [17]These authors jointly directed this work. Correspondence should be addressed to W.Z. (wei.zheng@vanderbilt.edu).

Received 29 March; accepted 29 November; published online 23 December 2012; doi:10.1038/ng.2505

Table 1 Association of CRC risk with the top four risk variants identified in east Asian samples

						Cases		Controls		Per-allele association		Heterogeneity	
SNP	Alleles[a]	Chr.	Gene[b]	Location (bp)[c]	Stage	Sample size	MAF	Sample size	MAF	OR (95% CI)[d]	P_{trend}	P^e	I^2
rs10774214	T/C	12p13.32	*CCND2*	4238613	GWAS	2,098	0.373	5,749	0.348	1.20 (1.09–1.32)	2.03×10^{-4}		
					Replication	5,197	0.381	5,797	0.355	1.16 (1.09–1.23)	5.80×10^{-7}		
					Overall	7,295	0.379	11,546	0.352	1.17 (1.11–1.23)	5.48×10^{-10}	0.615	0%
rs647161	A/C	5q31.1	*PITX1*	134526991	GWAS	2,098	0.353	5,749	0.308	1.22 (1.12–1.33)	3.29×10^{-6}		
					Replication	5,217	0.344	5,815	0.319	1.14 (1.07–1.21)	1.15×10^{-5}		
					Overall	7,315	0.347	11,564	0.313	1.17 (1.11–1.22)	3.77×10^{-10}	0.444	0%
rs2423279	C/T	20p12.3	*HAO1*	7760350	GWAS	2,098	0.339	5,749	0.307	1.16 (1.07–1.26)	4.96×10^{-4}		
					Replication	5,227	0.315	5,811	0.297	1.13 (1.06–1.19)	1.22×10^{-4}		
					Overall	7,325	0.322	11,560	0.302	1.14 (1.08–1.19)	2.29×10^{-7}	0.331	12%
rs1665650	T/C	10q26.12	*HSPA12A*	118477090	GWAS	2,098	0.346	5,749	0.310	1.20 (1.10–1.31)	3.88×10^{-5}		
					Replication	5,192	0.328	5,808	0.320	1.10 (1.04–1.17)	0.0018		
					Overall	7,290	0.333	11,557	0.315	1.13 (1.08–1.19)	8.58×10^{-7}	0.404	4%

Chr., chromosome; OR, odds ratio; CI, confidence interval.
[a]Minor/major allele for east Asians. OR was estimated for the minor allele. [b]Closest gene. [c]Location based on NCBI Human Genome Build 36.3. [d]Adjusted for age, sex, the first ten principal components (stage 1) and study site. [e]*P* for heterogeneity across studies in GWAS and replication was calculated using Cochran's *Q* test.

between positive control samples, minor allele frequency (MAF) of <5% or *P* value for Hardy-Weinberg equilibrium of $<1.0 \times 10^{-5}$ in controls for each study. Imputation was conducted for each study following the MaCH algorithm[18] using phased HapMap 2 Han Chinese in Beijing, China (CHB) and Japanese in Tokyo, Japan (JPT) samples as the reference. No apparent genetic admixture was detected, except for one sample from KCPS-II (**Supplementary Fig. 1**). Associations between CRC risk and each of the genotyped and imputed SNPs were evaluated using logistic regression within each study after adjusting for age, sex and the first ten principal components using mach2dat[18]. Meta-analyses were conducted under a fixed-effects model using the METAL program[19]. There was little evidence for inflation in the association test statistics for any of the five studies (genomic inflation factor (λ) range of 1.02 to 1.04) or for all studies combined ($\lambda = 1.01$) (**Supplementary Fig. 2** and **Supplementary Table 1**). The observed number of SNPs with small *P* values was slightly larger than that expected by chance (**Supplementary Fig. 2**).

Multiple genomic locations were found that were potentially related to CRC risk (**Supplementary Fig. 3**). Nine SNPs identified from published GWAS conducted in populations of European ancestry showed associations with CRC risk at $P < 0.05$ in stage 1 (data not shown). To improve the statistical power for evaluating these SNPs, we genotyped 6,476 additional samples to bring the total sample size to 5,252 cases and 9,071 controls. Except for the 2 SNPs that are monomorphic in east Asians (rs6691170 and rs16892766), all 16 of the other SNPs identified from published GWAS conducted in European-ancestry populations showed association with CRC risk in the same direction as reported previously (**Supplementary Table 3**). A significant association with CRC risk at $P < 0.05$ was found for 13 SNPs, including rs6687758, rs10936599, rs10505477, rs6983267, rs7014346, rs10795668, rs3802842, rs4444235, rs4779584, rs9929218, rs4939827, rs10411210 and rs961523. Except for two SNPs (rs6983267 and rs4779584), no statistically significant heterogeneity at $P < 0.05$ was observed between east Asian and European-ancestry populations (**Supplementary Table 3**).

To identify new genetic factors for CRC, we selected 64 SNPs for replication in an independent set of 5,358 cases and 5,922 controls recruited in 5 studies conducted in China, Korea and Japan (**Supplementary Table 2**). SNPs were selected from among those that (i) had MAF of >5%; (ii) showed no heterogeneity across studies ($P_{het} > 0.05$ and $I^2 < 25\%$); (iii) were not in linkage disequilibrium (LD; $r^2 < 0.2$) with any known CRC risk variant reported from previous GWAS; (iv) had high imputation quality in each of the five studies (RSQ > 0.5); and (v) were associated at $P < 0.01$ in the combined analysis of all five studies included in stage 1. These criteria were used to prioritize SNPs for replication.

Of the 64 SNPs evaluated in stage 2, 7 showed association with CRC risk at $P < 0.05$ with a direction of association consistent with that observed in stage 1 (**Table 1** and **Supplementary Table 4**). In the combined analysis of data from stages 1 and 2, *P* values for associations with two SNPs (rs647161 at 5q31.1, odds ratio (OR) = 1.17, $P = 3.77 \times 10^{-10}$, and rs10774214 at 12p13.32, OR = 1.17, $P = 5.48 \times 10^{-10}$) were lower than the conventional genome-wide significance level of 5.0×10^{-8}, providing convincing evidence for an association of these SNPs with CRC risk (**Table 1**). An additional SNP, rs2423279, showed a significant association in stage 2 after Bonferroni correction (corrected $P < 7.8 \times 10^{-4}$) but did not reach the conventional GWAS significance level for association with CRC risk in the combined analysis of all samples (OR = 1.14, $P = 2.29 \times 10^{-7}$). The association between CRC risk and each of these three SNPs was consistent across most studies (**Fig. 1**). Results for the other four SNPs that replicated in stage 2 at $P < 0.05$ (rs1665650, rs2850966, rs1580743 and rs4503064) are also presented (**Supplementary Table 4**), including one SNP (rs1665650) with an association *P* value of 8.58×10^{-7} in the combined analysis of all data from both stages (**Table 1**).

We next evaluated these top four SNPs (**Table 1**) using data from GWAS in the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) and the Colon Cancer Family Registry (CCFR), which together include 11,870 cases and 14,190 controls of European ancestry[4,20,21]. Three of the four SNPs were replicated in the GECCO and CCFR sample, although the strength of the associations was weaker than in east Asians (**Table 2**). These results provide independent support of our findings in the east Asian population. Meta-analyses of data from both east Asian and European-ancestry populations provided strong evidence for associations of CRC risk with three SNPs, with *P* values all below the genome-wide significance threshold of 5×10^{-8} (**Table 2**). The weaker associations observed in European-ancestry populations could be explained, in part, by differences in LD patterns at these loci for east

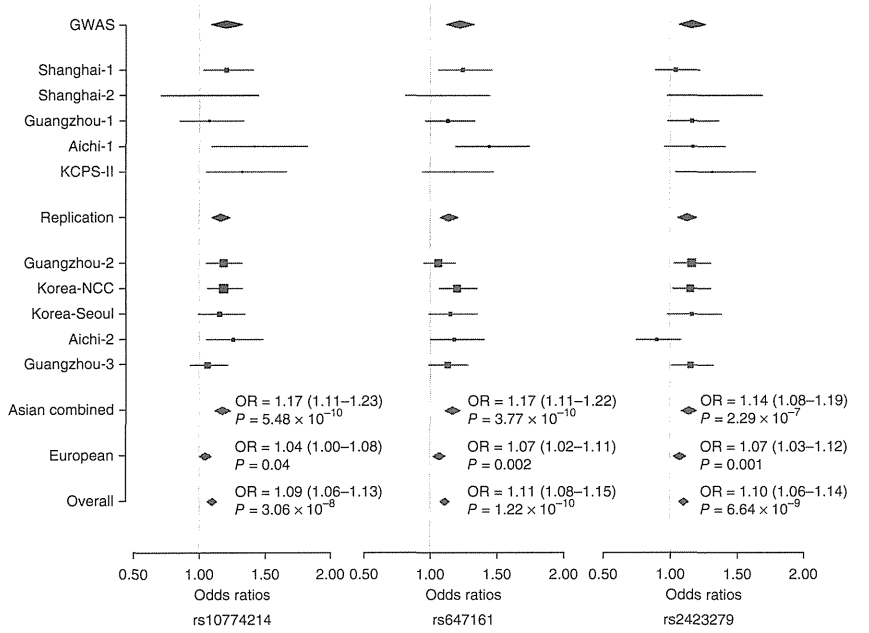

Figure 1 Forest plots for the three SNPs showing evidence of an association with CRC risk. Per-allele ORs are presented, with the area of each box proportional to the inverse variance weight of the estimate. Horizontal lines represent 95% confidence intervals.

Asians and Europeans (**Supplementary Fig. 4**). It is possible that causal variants in these regions are tagged by different SNPs in these two populations or that there is allelic heterogeneity, in which different underlying causal variants exist in populations of Asian and European ancestry. The difference in LD structure between Asian and European descendants and possible allelic heterogeneity in these two populations might explain, in part, why these loci were not discovered in previous studies conducted in individuals of European ancestry. The fourth SNP evaluated in the GECCO and CCFR sample, rs1665650, however, was not replicated in individuals with European ancestry (OR = 0.96, $P = 0.05$).

Stratification analyses showed that the association of CRC risk with each of the three replicated SNPs was generally consistent in Chinese, Korean and Japanese individuals ($P_{het} > 0.05$), although the association with rs2423279 was not statistically significant in Japanese, perhaps owing to a small sample size (**Supplementary Table 5**). Associations of these three SNPs with CRC risk were similar for men and women ($P_{het} > 0.05$) (**Supplementary Table 6**).

The rs10774214 SNP is located just 15 kb upstream of *CCND2*, the gene encoding cyclin D2 (**Fig. 2a**), a member of the D-type cyclin family, which also includes cyclins D1 and D3. These cyclins have a critical role in cell cycle control (from G1 to S phase) through activation of cyclin-dependent kinases (CDKs), primarily CDK4 and CDK6 (ref. 22). *CCND2* is closely related to *CCND1*, a well-established human oncogene[22,23]. Although *CCND2* has been less well studied than *CCND1*, several studies, including The Cancer Genome Atlas (TCGA), have shown that *CCND2* is overexpressed in a substantial proportion of human colorectal tumors[22–25]. Overexpression of this cyclin may be an independent predictor of survival in individuals with CRC[24]. Several other genes, including *PARP11*, *FGF23*, *FGF6*, *C12orf5* and *RAD51AP1*, are also in close proximity to the SNP identified in our study, of which both *C12orf5* (also known as *TIGAR*, encoding TP53-induced glycolysis and apoptosis regulator) and *RAD51AP1* were found to be overexpressed in CRC tissue included in TCGA[25]. rs10774214 is in strong LD with several SNPs that are located in potential transcription factor–binding sites, as determined using the TRANSFAC database[26]. Additional research may be warranted regarding possible mechanisms by which this SNP is related to CRC risk.

The rs647161 SNP is located on chromosome 5q31.1, where a cluster of SNPs were associated with CRC risk (**Fig. 2b**). Of the genes in this region (including *PITX1*, *CATSPER3*, *PCBD2*, *MIR4461* and *H2AFY*), *PITX1* is the closest to rs647161 (approximately 129 kb upstream). The *PITX1* gene (encoding paired-like homeodomain 1) has been described as a tumor suppressor gene and may be involved in the tumorigenesis of multiple human cancers[27–31], including CRC[27,32]. *PITX1* has been reported to suppress tumorigenicity by downregulating the RAS pathway, which is frequently altered in colorectal tumors[27]. Inhibition of *PITX1* induces the RAS pathway and tumorigenicity, and restoring *PITX1* in colon cancer cells inhibits tumorigenicity[27]. It also has been reported that *PITX1* may activate *TP53* (ref. 33) and regulate telomerase activity[34]. Consistent with its possible function as a tumor suppressor gene, *PITX1* has been found to be downregulated in human cancer tissue samples and cell lines[27–30,32]. CRC tissue expressing wild-type KRAS showed significantly lower expression of *PITX1* than tissue with mutant KRAS[32]. Most recently, low *PITX1* expression was found to be associated with poor survival in individuals with CRC[35]. In addition, rs6596201, which is in moderate LD with rs647161 ($r^2 = 0.25$), is an expression quantitative trait locus (eQTL) ($P = 2.42 \times 10^{-28}$) for the *PITX1* gene[36]. Several other genes at this locus, including *C5orf24*, *H2AFY* and *NEUROG1*, were also found to be highly expressed in colorectal tumors included in TCGA ($P < 0.001$)[25]. Additional studies are warranted to explore a possible role for these genes in the etiology of CRC.

Table 2 Association of CRC risk with the top three risk variants in European descendants and east Asian and European descendants combined

		MAF[b]		European-ancestry populations[c]				East Asian and European-ancestry populations combined[c]			
SNP	Alleles[a]	Cases	Controls	Cases	Controls	OR (95% CI)	P_{meta}	Cases	Controls	OR (95% CI)	P_{meta}
rs10774214	T/C	0.385	0.379	11,870	14,190	1.04 (1.00–1.09)	0.040	19,165	25,736	1.09 (1.06–1.13)	3.06×10^{-8}
rs647161	A/C	0.680	0.667	11,870	14,190	1.07 (1.02–1.11)	0.002	19,185	25,754	1.11 (1.08–1.15)	1.22×10^{-10}
rs2423279	C/T	0.263	0.252	11,870	14,190	1.07 (1.03–1.12)	0.001	19,195	25,750	1.10 (1.06–1.14)	6.64×10^{-9}

[a]Alelles (minor/major) for east Asians. [b]MAF in European-ancestry populations. [c]Summary statistics were generated using inverse variance–weighted fixed-effects meta-analysis.

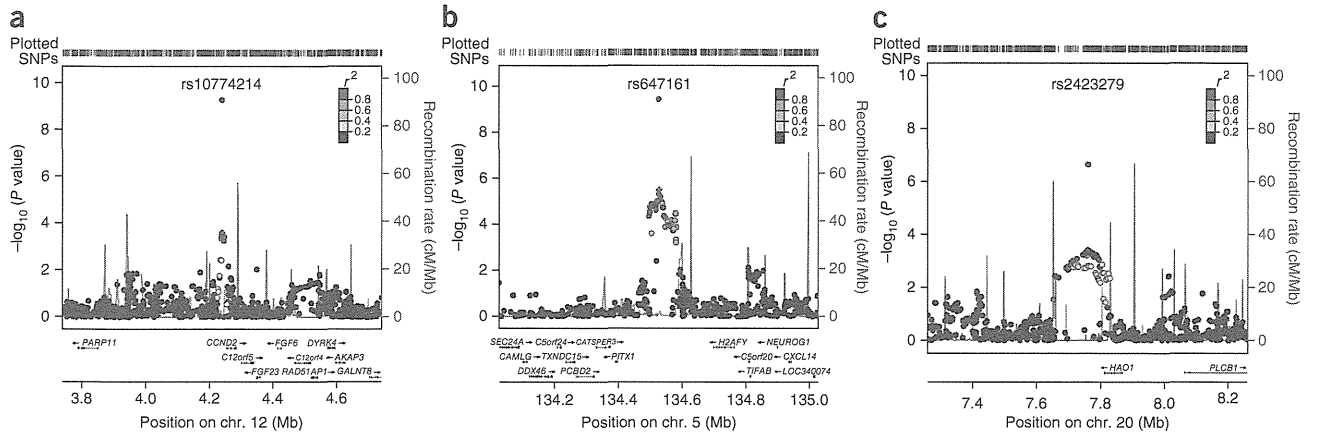

Figure 2 Regional plots of association results and recombination rates for the three SNPs showing evidence of association with CRC risk. Genotyped and imputed data from GWAS samples are plotted on the basis of their chromosomal position in NCBI Human Genome Build 36.3. For each region, the SNP selected for stage 2 replication is denoted with a diamond, and the *P* value from the combined analysis of stage 1 and 2 data is provided. (**a**–**c**) Data are shown for rs10774214 (**a**), rs647161 (**b**) and rs2423279 (**c**).

The rs2423279 SNP is located on chromosome 20p12.3, close to the *HAO1* and *PLCB1* genes (**Fig. 2c**). *HAO1* encodes hydroxyacid oxidase, which oxidizes 2-hydroxyacid. *PLCB1* encodes phospholipase C-β1, which has an important role in the intracellular transduction of many extracellular signals. Overexpression of the *PLCB1* gene has been observed in CRC tissue[25]. Possible mechanisms by which these genes are involved in CRC carcinogenesis are unknown. The rs2423279 SNP is 1,408,069 bp downstream of rs961253, a SNP previously identified in a European GWAS as being associated with CRC risk[10]. However, these two SNPs are not correlated in east Asians ($r^2 = 0$) or in Europeans ($r^2 = 0$). Adjustment for rs961253 did not change the results for rs2423279 (data not shown).

To our knowledge, this is the largest GWAS performed for CRC in east Asians, a population that differs from populations of European ancestry in CRC risk and certain aspects of genetic architecture. Results from our study, along with data from a large study conducted in a population of European ancestry, provide convincing evidence of associations with CRC risk for three new independent susceptibility loci at 5q31.1, 12p13.32 and 20p12.3. Results from this study provide new insights into the genetics and biology of CRC.

URLs. Cancer Genetic Markers of Susceptibility (CGEMS), http://cgems.cancer.gov/; Database of Genotypes and Phenotypes (dbGaP), http://www.ncbi.nlm.nih.gov/gap; EIGENSTRAT, http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm; eqtl.uchicago.edu, http://eqtl.uchicago.edu/Home.html; GTEx eQTL Browser, http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi; Haploview, http://www.broad.mit.edu/mpg/haploview/; HapMap Project, http://hapmap.ncbi.nlm.nih.gov/; IntOGen, http://www.intogen.org/home; LocusZoom, http://csg.sph.umich.edu/locuszoom/; MaCH 1.0, http://www.sph.umich.edu/csg/abecasis/MACH/; mach2dat, http://genome.sph.umich.edu/wiki/Mach2dat:_Association_with_MACH_output; METAL, http://www.sph.umich.edu/csg/abecasis/Metal/; PLINK version 1.07, http://pngu.mgh.harvard.edu/~purcell/plink/; R version 2.13.0, http://www.r-project.org/; SAS version 9.2, http://www.sas.com/; SNAP, http://www.broadinstitute.org/mpg/snap/; TRANSFAC, http://www.gene-regulation.com/pub/databases.html; UCSC Genome Browser, http://genome.ucsc.edu/; WHI investigators, https://cleo.whi.org/researchers/SitePages/Write%20a%20Paper.aspx.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. The authors wish to thank the study participants and research staff for their contributions and commitment to this project, R. Courtney for DNA preparation, J. He for data processing and analyses, and M.J. Daly for clerical support in manuscript preparation. This research was supported in part by US National Institutes of Health (NIH) grants R37CA070867, R01CA082729, R01CA124558, R01CA148667 and R01CA122364, as well as by Ingram Professorship and Research Reward funds from the Vanderbilt University School of Medicine. Participating studies (grant support) in the consortium are as follows: Shanghai Women's Health Study (US NIH, R37CA070867), Shanghai Men's Health Study (US NIH, R01CA082729), Shanghai Breast and Endometrial Cancer Studies (US NIH, R01CA064277 and R01CA092585; contributing only controls), Guangzhou Colorectal Cancer Study (National Key Scientific and Technological Project, 2011ZX09307-001-04, and the National Basic Research Program, 2011CB504303, contributing only controls; the Natural Science Foundation of China, 81072383, contributing only controls), Aichi Colorectal Cancer Study (Grant-in-Aid for Cancer Research, the Grant for the Third Term Comprehensive Control Research for Cancer and Grants-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology, 17015018 and 221S0001), Korea–National Cancer Center Colorectal Cancer Study (Basic Science Research Program through the National Research Foundation of Korea, 2010-0010276; National Cancer Center Korea, 0910220), Korea-Seoul Colorectal Cancer Study (none reported) and KCPS-II colorectal Cancer Study (National R&D Program for Cancer Control, 0920330; Seoul R&D Program, 10526).

We wish to thank all participants, staff and investigators of GECCO and CCFR for making it possible to present the results in individuals of European ancestry for new CRC susceptibility loci identified in east Asians. Investigators (institution and location) from GECCO and CCFR who provided support for this project include (in alphabetical order) Aaron K. Aragaki (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA), John A. Baron (Division of Gastroenterology and Hepatology, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA), Sonja I. Berndt (Division of Cancer Epidemiology and Genetics, National Cancer Institute, US NIH, Bethesda, Maryland, USA), Stéphane Bezieau (Service de Génétique Médicale, Centre Hospitalier Universitaire (CHU) Nantes, Nantes, France), Hermann Brenner, Katja Butterbach (Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany), Bette J. Caan (Division of Research, Kaiser Permanente Medical Care Program, Oakland, California, USA), Christopher S. Carlson (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA, and School of Public Health,

University of Washington, Seattle, Washington, USA), Graham Casey (Department of Preventive Medicine, University of Southern California, Los Angeles, California, USA), Andrew T. Chan (Division of Gastroenterology, Harvard Medical School and Massachusetts General Hospital, Boston, Massachusetts, USA, and Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA), Jenny Chang-Claude (Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany), Stephen J. Chanock (Division of Cancer Epidemiology and Genetics, National Cancer Institute, US NIH, Bethesda, Maryland, USA), Lin S. Chen (Department of Health Studies, University of Chicago, Chicago, Illinois, USA), Gerhard A. Coetzee (Keck School of Medicine, University of Southern California, Los Angeles, California, USA), Simon G. Coetzee (Keck School of Medicine, University of Southern California, Los Angeles, California, USA), David V. Conti (Department of Preventive Medicine, University of Southern California, Los Angeles, California, USA), Keith Curtis (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA), David Duggan (Translational Genomics Research Institute, Phoenix, Arizona, USA), Todd L. Edwards (Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, USA), Charles S. Fuchs (Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA, and Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA), Steven Gallinger (Department of Surgery, Mount Sinai Hospital, Toronto, Ontario, Canada, and Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada), Edward L. Giovannucci (Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA, and Departments of Epidemiology and Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA), Stephanie M. Gogarten (School of Public Health, University of Washington, Seattle, Washington, USA), Stephen B. Gruber (Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, California, USA), Robert W. Haile (Department of Preventive Medicine, University of Southern California, Los Angeles, California, USA), Tabitha A. Harrison (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA), Richard B. Hayes (Division of Epidemiology, Department of Environmental Medicine, New York University School of Medicine, New York, New York, USA), Michael Hoffmeister (Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany), John L. Hopper (Melbourne School of Population Health, The University of Melbourne, Melbourne, Victoria, Australia), Li Hsu (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA, and Department of Biostatistics, University of Washington, Seattle, Washington, USA), Thomas J. Hudson (Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada, and Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada), David J. Hunter (Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA), Carolyn M. Hutter (Division of Cancer Control and Population Sciences, National Cancer Institute, US NIH, Bethesda, Maryland, USA), Rebecca D. Jackson (Division of Endocrinology, Diabetes, and Metabolism, Ohio State University, Columbus, Ohio, USA), Mark A. Jenkins (Melbourne School of Population Health, The University of Melbourne, Melbourne, Victoria, Australia), Shuo Jiao (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA), Charles Kooperberg (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA), Sébastien Küry (Service de Génétique Médicale, CHU Nantes, Nantes, France), Andrea Z. LaCroix (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA), Cathy C. Laurie (Department of Biostatistics, University of Washington, Seattle, Washington, USA), Cecelia A. Laurie (Department of Biostatistics, University of Washington, Seattle, Washington, USA), Loic Le Marchand (Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA), Mathieu Lemire (Ontario Institute for Cancer Research, Toronto, Ontario, Canada), David Levine (School of Public Health, University of Washington, Seattle, Washington, USA), Noralane M. Lindor (Department of Health Sciences Research, Mayo Clinic, Scottsdale, Arizona, USA), Yan Liu (Stephens and Associates, Carrollton, Texas, USA), Jing Ma (Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA), Karen W. Makar (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA), Polly A. Newcomb (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA, and Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington, USA), Ulrike Peters (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA, and Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington, USA), John D. Potter (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA, Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington, USA, and Centre for Public Health Research, Massey University, Palmerston North, New Zealand), Ross L. Prentice (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA), Conghui Qu (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA), Thomas Rohan (Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Yeshiva University, Bronx, New York, USA), Robert E. Schoen (Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA), Fredrick R. Schumacher (Department of Preventive Medicine, University of Southern California, Los Angeles, California, USA), Daniela Seminara (Division of Cancer Control and Population Sciences, National Cancer Institute, US NIH, Bethesda, Maryland, USA), Martha L. Slattery (Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, Utah, USA), Darin Taverna (Translational Genomics Research Institute, Phoenix, Arizona, USA), Stephen N. Thibodeau (Department of Laboratory Medicine, Mayo Clinic, Rochester, Minnesota, USA, and Department of Pathology and Laboratory Genetics, Mayo Clinic, Rochester, Minnesota, USA), Cornelia M. Ulrich (Division of Preventive Oncology, German Cancer Research Center, Heidelberg, Germany, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA, and Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington, USA), Raakhee Vijayaraghavan (Genetic Basis of Human Disease Division, Translational Genomics Research Institute, Phoenix, Arizona, USA), Bruce Weir (Department of Biostatistics, University of Washington, Seattle, Washington, USA), Emily White (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA, and Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington, USA) and Brent W. Zanke (Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada).

We also thank B. Buecher of *ASTERISK*; U. Handte-Daub, M. Celik, R. Hettler-Jensen, U. Benscheid and U. Eilber of DACHS; P. Soule, H. Ranu, I. Devivo, D. Hunter, Q. Guo, L. Zhu and H. Zhang of HPFS, NHS and PHS; C. Berg and P. Prorok of PLCO; T. Riley of Information Management Services Inc.; B. O'Brien of Westat Inc; B. Kopp and W. Shao of SAIC-Frederick; investigators from the Women's Health Initiative (WHI; see URLs) and the GECCO Coordinating Center. Participating studies (grant support) in the GECCO and CCFR GWAS meta-analysis are as follows: GECCO (US NIH, U01 CA137088 and R01 CA059045), DALS (US NIH, R01 CA048998), Colo2&3 (US NIH, R01 CA060987), DACHS (German Federal Ministry of Education and Research, BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, CH 117/1-1, 01KH0404 and 01ER0814), HPFS (US NIH, P01 CA055075, UM1 CA167552, R01 137178 and P50 CA127003), MEC (US NIH, R37 CA054281, P01 CA033619 and R01 CA063464), NHS (US NIH, R01 137178, P50 CA127003 and P01 CA087969), OFCCR (US NIH, U01 CA074783), PMH (US NIH, R01 CA076366), PHS (US NIH, CA042182), VITAL (US NIH, K05 CA154337), WHI (US NIH, HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, HHSN271201100004C and 268200764316C) and PLCO (US NIH, Z01 CP 010200, U01 HG004446 and U01 HG 004438). CCFR is supported by the National Cancer Institute, US NIH, under RFA CA-95-011 and through cooperative agreements with members of the Colon Cancer Family Registry and principal investigators of the Australasian Colorectal Cancer Family Registry (U01 CA097735), the Familial Colorectal Neoplasia Collaborative Group (U01 CA074799; USC), the Mayo Clinic Cooperative Family Registry for Colon Cancer Studies (U01 CA074800), the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783), the Seattle Colorectal Cancer Family Registry (U01 CA074794) and the University of Hawaii Colorectal Cancer Family Registry (U01 CA074806). The GWAS work was supported by a National Cancer Institute grant (U01CA122839). OFCCR was supported by a GL2 grant from the Ontario Research Fund, the Canadian Institutes of Health Research and the Cancer Risk Evaluation (CaRE) Program grant from the Canadian Cancer Society Research Institute. B.Z. is a recipient of Senior Investigator Awards from the Ontario Institute for Cancer Research, through support from the Ontario Ministry of Economic Development and Innovation. ASTERISK was funded by a Regional Hospital Clinical Research Program (PHRC) and supported by the Regional Council of Pays de la Loire, Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), Association Anne de Bretagne Génétique and Ligue Régionale Contre le Cancer (LRCC). PLCO data sets were accessed with approval through dbGaP (Cancer Genetic Markers of Susceptibility (CGEMS) prostate cancer scan, phs000207.v1.p1 and GWAS of Lung Cancer and Smoking, phs000093.v2.p2).

AUTHOR CONTRIBUTIONS

W.Z. conceived and directed ACCC as well as the Shanghai-Vanderbilt Colorectal Cancer Genetics Project. W.-H.J., Y.-X.Z., K.M., A.S., Y.-B.X., S.H.J., D.-H.K., U.P.

and G.C. directed CRC projects at Guangzhou, Aichi, Korea-NCC, Shanghai, KCPS-II, Korea-Seoul, GECCO and CCFR, respectively. B.Z., Q.C. and W.W. coordinated the project. Q.C. directed laboratory operations. J.S. performed genotyping experiments. B.Z., J.L. and W.W. performed statistical analyses. W.Z. wrote the manuscript with substantial contributions from B.Z., Q.C., J.L., X.-O.S. and R.J.D. Z.R., G.Y., B.-T.J., Z.-Z.P., F.M., Y.-T.G., J.H.O., Y.-O.A., E.J.P., H.-L.L., J.W.P., J.J., J.-Y.J. and S.H. contributed to data and biological sample collection in the original studies included in ACCC and contributed to manuscript revision. Members of GECCO and CCFR contributed to data and biological sample collection in studies included in these consortia.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/ng.2505.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Jemal, A. *et al.* Global cancer statistics. *CA Cancer J. Clin.* **61**, 69–90 (2011).
2. de la Chapelle, A. Genetic predisposition to colorectal cancer. *Nat. Rev. Cancer* **4**, 769–780 (2004).
3. Dong, L.M. *et al.* Genetic susceptibility to cancer: the role of polymorphisms in candidate genes. *J. Am. Med. Assoc.* **299**, 2423–2436 (2008).
4. Zanke, B.W. *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* **39**, 989–994 (2007).
5. Tomlinson, I. *et al.* A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat. Genet.* **39**, 984–988 (2007).
6. Broderick, P. *et al.* A genome-wide association study shows that common alleles of *SMAD7* influence colorectal cancer risk. *Nat. Genet.* **39**, 1315–1317 (2007).
7. Jaeger, E. *et al.* Common genetic variants at the *CRAC1* (*HMPS*) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat. Genet.* **40**, 26–28 (2008).
8. Tenesa, A. *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat. Genet.* **40**, 631–637 (2008).
9. Tomlinson, I.P. *et al.* A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.* **40**, 623–630 (2008).
10. Houlston, R.S. *et al.* Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat. Genet.* **40**, 1426–1435 (2008).
11. Houlston, R.S. *et al.* Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat. Genet.* **42**, 973–977 (2010).
12. Cui, R. *et al.* Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. *Gut* **60**, 799–805 (2011).
13. He, J. *et al.* Generalizability and epidemiologic characterization of eleven colorectal cancer GWAS hits in multiple populations. *Cancer Epidemiol. Biomarkers Prev.* **20**, 70–81 (2011).
14. Zheng, W. *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat. Genet.* **41**, 324–328 (2009).
15. Bei, J.X. *et al.* A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. *Nat. Genet.* **42**, 599–603 (2010).
16. Jee, S.H. *et al.* Adiponectin concentrations: a genome-wide association study. *Am. J. Hum. Genet.* **87**, 545–552 (2010).
17. Nakata, I. *et al.* Association between the *SERPING1* gene and age-related macular degeneration and polypoidal choroidal vasculopathy in Japanese. *PLoS ONE* **6**, e19108 (2011).
18. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
19. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
20. Peters, U. *et al.* Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum. Genet.* **131**, 217–234 (2012).
21. Figueiredo, J.C. *et al.* Genotype-environment interactions in microsatellite stable/microsatellite instability–low colorectal cancer: results from a genome-wide association study. *Cancer Epidemiol. Biomarkers Prev.* **20**, 758–766 (2011).
22. Musgrove, E.A., Caldon, C.E., Barraclough, J., Stone, A. & Sutherland, R.L. Cyclin D as a therapeutic target in cancer. *Nat. Rev. Cancer* **11**, 558–572 (2011).
23. Mermelshtein, A. *et al.* Expression of D-type cyclins in colon cancer and in cell lines from colon carcinomas. *Br. J. Cancer* **93**, 338–345 (2005).
24. Sarkar, R. *et al.* Expression of cyclin D2 is an independent predictor of the development of hepatic metastasis in colorectal cancer. *Colorectal Dis.* **12**, 316–323 (2010).
25. Gundem, G. *et al.* IntOGen: integration and data mining of multidimensional oncogenomic data. *Nat. Methods* **7**, 92–93 (2010).
26. Matys, V. *et al.* TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34**, D108–D110 (2006).
27. Kolfschoten, I.G. *et al.* A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity. *Cell* **121**, 849–858 (2005).
28. Chen, Y. *et al.* Decreased *PITX1* homeobox gene expression in human lung cancer. *Lung Cancer* **55**, 287–294 (2007).
29. Chen, Y.N., Chen, H., Xu, Y., Zhang, X. & Luo, Y. Expression of pituitary homeobox 1 gene in human gastric carcinogenesis and its clinicopathological significance. *World J. Gastroenterol.* **14**, 292–297 (2008).
30. Lord, R.V. *et al.* Increased *CDX2* and decreased *PITX1* homeobox gene expression in Barrett's esophagus and Barrett's-associated adenocarcinoma. *Surgery* **138**, 924–931 (2005).
31. Nagel, S. *et al.* Activation of paired-homeobox gene *PITX1* by del(5)(q31) in T-cell acute lymphoblastic leukemia. *Leuk. Lymphoma* **52**, 1348–1359 (2011).
32. Watanabe, T. *et al.* Differential gene expression signatures between colorectal cancers with and without *KRAS* mutations: crosstalk between the KRAS pathway and other signalling pathways. *Eur. J. Cancer* **47**, 1946–1954 (2011).
33. Liu, D.X. & Lobie, P.E. Transcriptional activation of p53 by Pitx1. *Cell Death Differ.* **14**, 1893–1907 (2007).
34. Qi, D.L. *et al.* Identification of *PITX1* as a *TERT* suppressor gene located on human chromosome 5. *Mol. Cell Biol.* **31**, 1624–1636 (2011).
35. Knösel, T. *et al.* Loss of desmocollin 1-3 and homeobox genes *PITX1* and *CDX2* are associated with tumor progression and survival in colorectal carcinoma. *Int. J. Colorectal Dis.* **27**, 1391–1399 (2012).
36. Zeller, T. *et al.* Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* **5**, e10693 (2010).

ONLINE METHODS

Study populations. After quality control filtering, 7,456 cases and 11,671 controls from 10 studies were included in the consortium (**Supplementary Table 2**). Detailed descriptions of participating studies and demographic characteristics of study participants are provided in the **Supplementary Note**. Briefly, the consortium included 10,730 Chinese participants, 5,544 Korean participants and 2,853 Japanese participants. Chinese participants were from five studies: the Shanghai Study 1 (Shanghai-1, $n = 3,102$), the Shanghai Study 2 (Shanghai-2, $n = 485$), the Guangzhou Study 1 (Guangzhou-1, $n = 1,613$), the Guangzhou Study 2 (Guangzhou-2, $n = 2,892$) and the Guangzhou Study 3 (Guangzhou-3, $n = 2,638$). Korean participants were from three studies: the Korean Cancer Prevention Study-II (KCPS-II, $n = 1,301$), the Seoul Study ($n = 1,522$) and the Korea–National Cancer Center (Korea-NCC) Study ($n = 2,721$). Japanese participants were from two studies: the Aichi Study 1 (Aichi-1, $n = 1,346$) and the Aichi Study 2 (Aichi-2, $n = 1,507$). We also evaluated associations for the top 4 SNPs using data from 11,870 CRC cases and 14,190 controls of European ancestry included in GECCO and CCFR, which included 14 studies from the United States, Europe, Canada and Australia[4,20,21]. Approval was granted from the relevant institutional review boards at all study sites, and all included participants gave informed consent.

Genotyping and quality control procedures. Detailed descriptions of genotyping and quality control procedures as well as design of plates and control samples are given in the **Supplementary Note**. Briefly, in stage 1, 481 cases and 2,632 controls from Shanghai-1 were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0 as described previously[14]. The average concordance percentage of quality control samples was 99.7%, with a median value of 100% in Shanghai-1 (refs. 14,37,38). Stage 1 genotyping for 296 cases and 257 controls in Shanghai-2 was performed using Illumina HumanOmniExpress BeadChips. The same method was used to genotype cases from the Guangzhou-1 ($n = 694$) and Aichi-1 ($n = 497$) studies in stage 1. The positive quality control samples in these studies had an average concordance percentage of 99.41% and a median value of 99.97%. Cases and controls in KCPS-II were genotyped using the Affymetrix Genome-Wide Human SNP Array 5.0 (ref. 16). Controls for the Guangzhou-1 and Aichi-1 studies were genotyped previously using the Illumina Human610-Quad BeadChip[15] and Illumina Infinium HumanHap610 BeadChip[17] platforms, respectively. Details of quality control procedures for these samples have been described previously[15–17]. We excluded from the analysis samples that were genetically identical or duplicated, had a genotype-determined sex that was inconsistent with self-reported data, had unclear population structure, had close relatives with a PI-HAT estimate greater than 0.25 or had a call rate of <95%. Within each study, SNPs were excluded if (i) MAF was <5%, (ii) the call rate was <95%; (iii) the genotyping concordance percentage was <95% in quality control samples; (iv) the P value for Hardy-Weinberg equilibrium was $<1.0 \times 10^{-5}$ in controls; or (v) SNPs were not on the 22 autosomes. The final numbers of cases, controls and SNPs remaining for analysis in each participating study are presented in **Supplementary Table 1**.

Genotyping for stage 2 was completed using the iPLEX Sequenom MassARRAY platform as described previously[14,39]. With the exception of samples from the Guangzhou-3 study, which were genotyped at Fudan University (Shanghai), all other samples were genotyped at the Vanderbilt Molecular Epidemiology Laboratory. The average concordance percentage of the genotyping data for positive control samples was >99% with a median value of 100% for each of the five studies. SNPs were excluded from the analysis if (i) the call rate was <95%, (ii) the genotyping concordance percentage was <95% in control samples, (iii) there was an unclear genotype call or (iv) the P value for Hardy-Weinberg equilibrium was $<7.8 \times 10^{-4}$. The numbers of SNPs remaining for analysis in each participating study in stage 2 are presented in the **Supplementary Note**.

Genotyping for samples included in the GECCO and CCFR GWAS was conducted using Illumina BeadChip arrays, with the exception of the Ontario Familial Colorectal Cancer Registry study, for which Affymetrix arrays were used[4,20,21]. Details of the quality control procedures for these samples are presented in the **Supplementary Note**.

SNP selection for replication. SNPs were selected for stage 2 replication if (i) data were available in each of the five stage 1 studies; (ii) MAF was >5% in each stage 1 study; (iii) no heterogeneity was detected across the five studies included in stage 1 ($P_{het} > 0.05$ and $I^2 < 25\%$); (iv) there was no LD ($r^2 < 0.2$) with any known risk variant reported from previous GWAS; (v) there was no LD ($r^2 < 0.2$) with the other SNPs identified in this study; (vi) there was high imputation quality in each of the five studies (RSQ > 0.5); and (vii) $P < 0.01$ in combined analysis of all stage 1 studies.

Evaluation of population structure. We evaluated population structure in each of the five participating studies included in stage 1 by using principal-components analysis (PCA). Genotyping data for uncorrelated genome-wide SNPs were pooled with data from HapMap to generate the first ten principal components using EIGENSTRAT software[40] (see URLs). The first two principal components for each sample were plotted using R (see URLs). We identified and excluded one participant of KCPS-II who was more than 6 s.d. away from the means of principal components 1 and 2 (**Supplementary Fig. 1**). The remaining 7,847 samples showed clear east Asian origin, and these samples were included in the final genome-wide association analysis. Cases and controls in each of the five studies were in the same cluster as HapMap Asian samples. The estimated inflation factor λ ranged from 1.02 to 1.04 in these studies after adjusting for age, sex and the first ten principal components, with a λ of 1.01 for combined stage 1 data (**Supplementary Fig. 2** and **Supplementary Table 1**).

Imputation. We used the MaCH 1.0 program[18] (see URLs) to impute genotypes for autosomal SNPs that were present in HapMap Phase 2 release 22 separately for each of the five studies included in stage 1. Genotype data from the 90 Asian subjects from HapMap were used as the reference. For Guangzhou-1 and Aichi-1, cases and controls were genotyped using different platforms. To improve imputation quality[41], we identified SNPs for which data were available in both cases and controls (250,612 SNPs in Guangzhou-1 and 232,426 SNPs in Aichi-1) and used them to impute genotyping data. A total of 1,636,380 genotyped SNPs or imputed SNPs with high imputation quality (RSQ > 0.50) in all five studies were tested for association with CRC. To directly evaluate the imputation quality for the top four SNPs identified in our study, we genotyped them in approximately 2,500 samples included in stage 1. The agreement of genotype calls derived from direct genotyping and imputation was very high, with mean concordance rates of 98.05%, 95.61%, 99.84% and 97.90% for rs647161, rs10774214, rs2423279 and rs1665650, respectively (**Supplementary Table 7**).

Statistical analyses. Dosage data for genotyped and imputed SNPs for participants in each stage 1 study were analyzed using the program mach2dat[18] (see URLs). We coded 0, 1 or 2 copies of the effect allele as the dosage for genotyped SNPs, and, for imputed SNPs, we used the expected number of copies of the effect allele as the dosage score. This approach has been shown to give unbiased estimates in meta-analyses[42]. Associations between SNPs and CRC risk were assessed using ORs and 95% CIs derived from logistic regression models. ORs were estimated on the basis of the log-additive model and adjusted for age, sex and the first ten principal components. PLINK version 1.07 (see URLs) also was used to analyze genotype data[43] and yielded results virtually identical to those derived from dosage data using mach2dat[18]. Meta-analyses were performed using the inverse-variance method, assuming a fixed-effects model, and calculations were implemented in the METAL package[19] (see URLs).

Similar to stage 1, we used logistic regression models to derive ORs and 95% CIs for the 64 selected SNPs in stage 2, assuming a log-additive model with adjustment for age and sex. We performed joint analyses to generate summary results for combined samples from all studies, with additional adjustment for study site. We also conducted stratification analysis for the top four SNPs by population ancestry (Chinese, Korean or Japanese) and by sex. We used Cochran's Q statistic to test for heterogeneity[44] and the I^2 statistic to quantify heterogeneity[45] across studies as described elsewhere in detail[46]. Analyses for stage 2, as well as combined stage 1 and 2 data, were conducted using SAS, version 9.2 (see URLs), with the use of two-tailed tests. P values of $<5\times10^{-8}$ in the combined analysis was considered statistically significant.

We used Haploview version 4.2 (see URLs; ref. 47) to generate a genome-wide Manhattan plot for results from the stage 1 meta-analysis. Forest plots

and quantile-quantile plots were drawn using R. We drew regional association plots using the website-based tool LocusZoom, version 1.1 (see URLs; ref. 48). LD plots were generated using Haploview[47] and the UCSC Genome Browser (see URLs).

37. Long, J. *et al.* Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. *PLoS Genet.* **8**, e1002532 (2012).
38. Shu, X.O. *et al.* Identification of new genetic risk variants for type 2 diabetes. *PLoS Genet.* **6**, pii: e1001127 (2010).
39. Zheng, W. *et al.* Genetic and clinical predictors for breast cancer risk assessment and stratification among Chinese women. *J. Natl. Cancer Inst.* **102**, 972–981 (2010).
40. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
41. Sinnott, J.A. & Kraft, P. Artifact due to differential error when cases and controls are imputed from different platforms. *Hum. Genet.* **131**, 111–119 (2012).
42. Jiao, S., Hsu, L., Hutter, C.M. & Peters, U. The use of imputed values in the meta-analysis of genome-wide association studies. *Genet. Epidemiol.* **35**, 597–605 (2011).
43. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
44. Lau, J., Ioannidis, J.P. & Schmid, C.H. Quantitative synthesis in systematic reviews. *Ann. Intern. Med.* **127**, 820–826 (1997).
45. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
46. Zhang, B., Beeghly-Fadiel, A., Long, J. & Zheng, W. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Lancet Oncol.* **12**, 477–488 (2011).
47. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
48. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).

ARTHRITIS & RHEUMATISM
Vol. 65, No. 2, February 2013, pp 472–480
DOI 10.1002/art.37777

PLD4 as a Novel Susceptibility Gene for Systemic Sclerosis in a Japanese Population

Chikashi Terao,[1] Koichiro Ohmura,[1] Yasushi Kawaguchi,[2] Tetsuya Nishimoto,[3] Aya Kawasaki,[4] Kazuhiko Takehara,[5] Hiroshi Furukawa,[6] Yuta Kochi,[7] Yuko Ota,[2] Katsunori Ikari,[2] Shinichi Sato,[8] Shigeto Tohma,[6] Ryo Yamada,[1] Kazuhiko Yamamoto,[8] Michiaki Kubo,[7] Hisashi Yamanaka,[2] Masataka Kuwana,[3] Naoyuki Tsuchiya,[4] Fumihiko Matsuda,[1] and Tsuneyo Mimori[1]

Objective. **Systemic sclerosis (SSc) is an autoimmune disease for which multiple susceptibility genes have been reported. Genome-wide association studies have shown that large numbers of susceptibility genes are shared among autoimmune diseases. Recently, our group identified 9 novel susceptibility genes associated with rheumatoid arthritis (RA) in a Japanese population. The aim of this study was to elucidate whether the 18 genes that displayed associations or suggestive associations for RA in our previous study are associated with SSc in Japanese.**

Methods. **We performed an association study that included 415 patients with SSc and 16,891 control subjects, followed by a replication study that included 315 patients and 21,054 control subjects. The 18 markers reported to display association with RA were analyzed for their associations with SSc in the first study, and 5 markers were further analyzed in the replication study. The inverse variance method was used to evaluate the associations of these markers with SSc in a combined study.**

Results. **In the phospholipase D4 gene (*PLD4*), rs2841277 displayed a significant association with SSc in Japanese patients ($P = 0.00017$). We observed that rs2841280 in exon 2 of *PLD4* was in strong linkage disequilibrium with rs2841277 and introduced an amino acid alteration. We also observed associations between SSc and rs6932056 in *TNFAIP3* and rs2280381 in *IRF8* ($P = 0.0000095$ and $P = 0.0030$, respectively), both of which displayed associations with SSc in a European population.**

Conclusion. **We determined that *PLD4* is a novel susceptibility gene for SSc in Japanese, thus confirming the involvement of *PLD4* in autoimmunity. Associations between SSc and *TNFAIP3* or *IRF8* were also detected in our Japanese population. SSc and RA appear to share relatively large proportions of their genetic backgrounds.**

Systemic sclerosis (SSc) is a connective tissue disease that affects 7–489 individuals per million worldwide and is characterized by the excess production of extracellular matrix molecules and fibrosis (1). Patients with SSc display skin sclerosis, obliterative microvasculopathy such as Raynaud's phenomenon, and multiorgan involvement. Severe complications of SSc sometimes develop, including interstitial lung disease, pulmonary hypertension, and renal crisis. These severe symptoms

Supported by grants-in-aid from the Ministry of Health, Labor, and Welfare of Japan and from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, by research grants from the Japan Rheumatism Foundation, the Waksman Foundation, and the Mitsubishi Pharma Research Foundation, and by the Genetics and Allied Research in Rheumatic Diseases Networking consortium.

[1]Chikashi Terao, MD, PhD, Koichiro Ohmura, MD, PhD, Ryo Yamada, MD, PhD, Fumihiko Matsuda, PhD, Tsuneyo Mimori, MD, PhD: Kyoto University, Kyoto, Japan; [2]Yasushi Kawaguchi, MD, PhD, Yuko Ota, MD, Katsunori Ikari, MD, PhD, Hisashi Yamanaka, MD, PhD: Tokyo Women's Medical University, Tokyo, Japan; [3]Tetsuya Nishimoto, PhD, Masataka Kuwana, MD, PhD: Keio University, Tokyo, Japan; [4]Aya Kawasaki, PhD, Naoyuki Tsuchiya, MD, PhD: University of Tsukuba, Tsukuba, Japan; [5]Kazuhiko Takehara, MD, PhD: Kanazawa University, Kanazawa, Japan; [6]Hiroshi Furukawa, MD, PhD, Shigeto Tohma, MD: Sagamihara National Hospital, National Hospital Organization, Sagamihara, Japan; [7]Yuta Kochi, MD, PhD, Michiaki Kubo, MD, PhD: RIKEN, Yokohama, Japan; [8]Shinichi Sato, MD, PhD, Kazuhiko Yamamoto, MD, PhD: University of Tokyo, Tokyo, Japan.

Address correspondence to Chikashi Terao, MD, PhD, or Koichiro Ohmura, MD, PhD, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan. E-mail: a0001101@kuhp.kyoto-u.ac.jp or ohmurako@kuhp.kyoto-u.ac.jp.

Submitted for publication June 4, 2012; accepted in revised form October 23, 2012.

and complications of SSc result in a poor prognosis and a shortened lifespan (2,3). No effective method for preventing or curing SSc has been established (4).

It is well known that SSc has genetic components (5); for example, a US study revealed that the incidence of SSc was much higher among the families of patients with SSc compared with the general population (6). Recent technologic developments enabled the use of genome-wide association studies (GWAS) to identify novel susceptibility loci for autoimmune diseases (7). GWAS of European patients with SSc revealed that *CD247* (8), HLA (8), *TNIP1*, *PSORS1C1*, and *RHOB* (9) are susceptibility loci for SSc. In addition, another GWAS identified associations between *IRF8*, *GRB10*, and *SOX5* and limited cutaneous SSc (lcSSc) in a European population (10). Furthermore, studies adopting a candidate gene approach based on subjecting genes to functional inference analysis led to the identification of *STAT4* (11), *IRF5* (12), *TBX21* (13), *NLRP1* (14), *TNFSF4* (15), *CD226* (16), *BLK* (17), and *TNFAIP3* (18) as novel susceptibility genes for SSc in Europeans. SSc association studies in Japanese populations confirmed that *STAT4* (19), *IRF5* (20), and *BLK* (21) are associated with SSc and identified *UBE2L3* as a susceptibility gene for diffuse cutaneous SSc (dcSSc) (22). An association between HLA and SSc was also detected in Asians (23). These findings suggest a clear overlap in the genetic background of SSc between different populations.

It is well known that susceptibility genes are shared by various autoimmune diseases (24). In fact, HLA (25), *STAT4* (26), and *TNFAIP3* (27,28), which are susceptibility genes for SSc, have also been reported to be associated with rheumatoid arthritis (RA). In addition, *PTPN22*, which was shown to be strongly associated with RA in a European population (29), showed a suggestive association with SSc in Europeans (30). The sharing of these susceptibility genes between RA and SSc raises the possibility that newly identified susceptibility genes for RA could also be susceptibility genes for SSc. Recently, a large Japanese consortium, the Genetic and Allied research in Rheumatic diseases Networking consortium, identified 9 novel susceptibility genes and 6 candidate susceptibility genes for RA using a meta-analysis of GWAS and replication studies (31). Four other genes, namely, HLA, *PADI4*, *CCR6*, and *TNFAIP3*, were also confirmed to display associations with RA. Here, we performed a 2-stage association study of Japanese patients with SSc, in which we genotyped these genes as candidate susceptibility loci.

PATIENTS AND METHODS

Study subjects. DNA samples were obtained from 415 patients with SSc at Kyoto University Hospital and Tokyo Women's Medical University; these samples comprised the first set. Independent DNA samples were obtained from 315 patients with SSc at Keio University Hospital, Sagamihara National Hospital, and Kanazawa University Hospital; these samples were used as the replication set. All patients were Japanese, all had a diagnosis of SSc as determined by a rheumatologist, and all fulfilled the 1980 American College of Rheumatology classification criteria for SSc (32). The patients with SSc for whom clinical information was available were classified as having lcSSc or dcSSc, according to the definitions developed by LeRoy et al (33). The control samples were described in detail in our previous study (31). The current study was approved by the local ethics committees at each institution, and written informed consent was obtained from all subjects. The basic characteristics of the study subjects are shown in Table 1.

Genotyping. The 9 novel susceptibility markers, 6 potentially associated markers, and 4 confirmed markers of RA that were identified in our previous study in a Japanese population (31) were chosen as candidate susceptibility markers for SSc in Japanese. Eighteen of the 19 markers (HLA was excluded; see Results), none of which had previously been reported to be associated with SSc in Japanese individuals, were genotyped in the current study. The 5 candidate markers in the first set that showed associations with P values less than 0.1 were further genotyped in the replication study. Single-nucleotide polymorphisms (SNPs) rs2841280 and rs894037 were chosen as candidate causative variants in the phospholipase D4 gene (*PLD4*) region. Because rs894037 was shown to be monomorphic in Japanese, rs2841280 was genotyped in 334 control subjects, in addition to all patients, for imputation reference. The patients in the first and replication studies were genotyped at Kyoto University or Tokyo Women's Medical University and at Keio University or University of Tsukuba, respectively, using TaqMan assays (Applied Biosystems). The genotyping methods in control subjects were described in detail in our previous study (31).

Briefly, control genotypes in the first set were imputed based on the genome-scanning data, using mach2dat software with HapMap Phase II East Asian Populations as reference. The control genotypes for the replication study were extracted from genome-scanning data for the markers included on Illumina HumanHap610 Quad BeadChips. The genotypes for rs6932056 (which is not included in the array) were imputed based on the genome-scanning data, using mach2dat software with HapMap Phase II East Asian Populations as reference, and were used as control data for the replication set. The genotypes for rs2841280 (which is not included in the HapMap data or the array) were also imputed in control subjects, based on the genome-scanning data, using mach2dat software. Genotyping data for the 334 control subjects as determined by TaqMan assay in combination with genome-scanning data were used as reference.

Statistical analysis. The associations between the genotyped markers and SSc were analyzed using a Cochran-Armitage trend test in both the first and replication studies. Subanalyses were performed by comparing the genotypes of

Table 1. Characteristics of the study population*

	Patients	Controls
First set		
Institutions	Kyoto University, Tokyo Women's Medical University	Kyoto University, Tokyo Women's Medical University, BioBank Japan
Typing	TaqMan assay	Illumina HumanHap610 Quad BeadChip, Illumina HumanHap550 BeadChip, Affymetrix Genome-Wide Human SNP Array 6.0
Limited SSc/diffuse SSc, %	49.6/50.4	Not applicable
Anti–topo I/ACA, %	30.6/31.1	Not applicable
Interstitial lung disease, %	48.9	Not applicable
Age, mean ± SD years	50.9 ± 14.7	60.9 ± 12.5
Female, %	91.3	44.9
Replication set		
Institutions	Keio University, Sagamihara National Hospital, Kanazawa University	Kyoto University, BioBank Japan
Typing	TaqMan assay	Illumina HumanHap550 BeadChip, Illumina HumanHap610 Quad BeadChip
Limited SSc/diffuse SSc, %	63.8/34.6	Not applicable
Anti–topo I/ACA, %	29.5/35.2	Not applicable
Interstitial lung disease, %	43.2	Not applicable
Age, mean ± SD years	51.4 ± 14.1	59.3 ± 14.2
Female, %	87.3	48.4

* The first set included 415 patients with systemic sclerosis (SSc) and 16,891 control subjects. The replication set included 315 patients with SSc and 21,054 control subjects. Anti–topo I = anti–topoisomerase I; ACA = anticentromere antibody.

the control subjects with those of patients in the SSc subgroups based on the disease phenotypes. The subanalyses used the same control subjects as were used in the association studies. Intracase analyses based on phenotypes were also performed.

Odds ratios (ORs) and 95% confidence intervals were also calculated. The associations detected in the first and replication studies were then meta-analyzed using the inverse variance method. The resultant *P* values were corrected using the Benjamini-Hochberg false discovery rate (FDR) criterion, and corrected *P* values less than 0.05 were regarded as significant in both the combined study and the subanalyses. The efficiency of the current study was estimated by calculating the likelihood of detecting 3 significant markers (after correcting the *P* values using the FDR method) among 18 randomly selected markers. After the statistically significant markers were identified, the best-fit model for each association was analyzed using dominant, recessive, trend, and allelic chi-square tests or models. Statistical analyses were performed using R or SPSS (version 18) software.

RESULTS

Analyses of candidate genes for SSc in a Japanese population. The 415 patients with SSc and 16,891 control subjects in the first set were genotyped for the 18 markers that were shown to have associations or suspected associations with RA in our previous study. The HLA region was excluded from the genotyped markers, because this region has already been shown to be associated with SSc in Asians. The allele frequencies of the patients were compared with those of the control subjects, using a Cochran-Armitage trend test.

As a result, 3 markers that demonstrated associations with *P* values less than 0.01 in the first set (Table 2) were identified, namely, rs6932056 in the *TNFAIP3* region ($P = 0.0000038$, OR 1.69), rs10821944 in the *ARID5B* region ($P = 0.0025$, OR 1.25), and rs2841277 in the *PLD4* region ($P = 0.0054$, OR 1.25). Two loci that showed suggestive associations with *P* values less than 0.1 (Table 2) were also identified, namely, rs12529514 in the *CD83* region ($P = 0.083$, OR 1.18) and rs2280381 in the *IRF8* region ($P = 0.095$, OR 1.19). The *TNFAIP3* and *IRF8* regions were previously reported to display associations with SSc and lcSSc, respectively, in European populations (10,18). These 5 markers were selected as candidate susceptibility markers for SSc in Japanese and were subjected to validation.

Next, a replication study consisting of 315 patients with SSc and 21,054 control subjects was performed to validate the associations of the 5 markers with SSc. The patients were genotyped for the 5 markers. The genotypes of the control subjects for the 5 markers, except rs6932056, were extracted from the Illumina Infinium HumanHap610 Quad array, as reported previously (31). The genotypes for rs6932056 were imputed based on genome-scanning data using mach2dat soft-

Table 2. Association studies of Japanese patients with SSc*

				Allele 1 frequency								
				First set			Replication set			Combined study		
SNP	Chr	Gene	Allele 1/2	Controls	Patients	*P*	Controls†	Patients	*P*	*P*, patients vs. controls	OR (95% CI)	*P*, patients without overlapping RA vs. controls
rs766449	1	*PADI4*	T/C	0.40	0.37	0.12	–	–	–	–	–	–
rs11900673	2	*B3GNT2*	T/C	0.29	0.28	0.65	–	–	–	–	–	–
rs2867461	4	*ANXA3*	A/G	0.44	0.43	0.57	–	–	–	–	–	–
rs657075	5	*IL3-CSF2*	A/G	0.36	0.34	0.25	–	–	–	–	–	–
rs12529514	6	*CD83*	C/T	0.14	0.16	0.083	0.15	0.16	0.31	0.046	1.15 (1.00–1.33)	0.040
rs1571878	6	*CCR6*	C/T	0.49	0.47	0.28	–	–	–	–	–	–
rs6932056	6	*TNFAIP3*	C/T	0.069	0.11	3.8×10^{-6}	0.067	0.079	0.23	9.5×10^{-6}	1.50 (1.25–1.80)	5.4×10^{-6}
rs2233434	6	*NFKBIE*	G/A	0.21	0.21	0.93	–	–	–	–	–	–
rs10821944	10	*ARID5B*	G/T	0.36	0.41	0.0025	0.36	0.37	0.64	0.0073	1.16 (1.04–1.29)	0.010
rs3781913	11	*PDE2A-CENTD2*	T/G	0.69	0.69	0.91	–	–	–	–	–	–
rs4937362	11	*ETS1-FLI1*	T/C	0.68	0.68	0.88	–	–	–	–	–	–
rs2841277	14	*PLD4*	T/C	0.69	0.74	0.0054	0.69	0.73	0.012	0.00017	1.25 (1.11–1.41)	0.00052
rs3783637	14	*GCH1*	C/T	0.74	0.73	0.54	–	–	–	–	–	–
rs1957895	14	*PRKCH*	G/T	0.39	0.41	0.26	–	–	–	–	–	–
rs6496667	15	*ZNF774*	A/C	0.35	0.37	0.33	–	–	–	–	–	–
rs7404928	16	*PRKCB1*	T/C	0.62	0.63	0.51	–	–	–	–	–	–
rs2280381	16	*IRF8*	T/C	0.84	0.86	0.095	0.83	0.87	0.0099	0.0030	1.26 (1.08–1.47)	0.0021
rs2847297	18	*PTPN2*	G/A	0.34	0.34	0.85	–	–	–	–	–	–

* SSc = systemic sclerosis; SNP = single-nucleotide polymorphism; Chr = chromosome; OR = odds ratio; 95% CI = 95% confidence interval; RA = rheumatoid arthritis.
† The control rs6932056 genotypes used in the replication study were imputed using genome-scanning data obtained for 3,765 subjects.

ware, because rs6932056 was not included in the array. As a result, rs2841277 in the *PLD4* region and rs2280381 in the *IRF8* region showed relatively strong associations with SSc ($P = 0.012$, OR 1.25 and $P = 0.0099$, OR 1.37, respectively) (Table 2). Interestingly, we observed that all 5 of the markers that displayed associations in the first study also demonstrated the same association directions in the replication study.

The inverse variance method was used to combine the data for the first and replication studies. SNPs rs2841277 in the *PLD4* region, rs6932056 in the *TNFAIP3* region, and rs2280381 in the *IRF8* region showed significant associations with SSc even after correcting the associated *P* values using the FDR method for multiple testing (Table 2). Importantly, all 3 of these loci shared risk alleles with RA. Although rs6932056 in the *TNFAIP3* region did not show a strong association with SSc in the replication study, its association was significant in the combined study. The *PLD4* region was shown to be a novel susceptibility gene for SSc, and, for the first time, the *TNFAIP3* and *IRF8* regions were confirmed to be associated with SSc in Japanese.

The association between rs2841277 and SSc was then investigated in detail. When the 200-kbp region around rs2841277 was evaluated, 2 hypothetical genes and cell division cycle associated 4 gene (*CDCA4*) were located at the region, in addition to *PLD4*. *PLD4* was the only gene whose region showed moderate to strong linkage disequilibrium (LD) with rs2841277, indicating *PLD4* as a susceptibility gene (Figure 1A). We vigorously searched candidate markers in exons of *PLD4* that showed strong LD with rs2841277 and selected 2 markers registered in the 1000 Genomes Project (34) that displayed >5% frequency in genotyped subjects, namely, rs2841280 (Figure 1B) and rs894037 in exon 2. Genotyping of these polymorphisms revealed strong LD between rs2841280 (E27Q) and rs2841277 ($D' = 0.98$, $r^2 = 0.75$) and monomorphism of rs894037 in Japanese. An association study of rs2841280 using control genotypes obtained by imputation supported association of *PLD4* with SSc ($P = 6.3 \times 10^{-5}$) (see Supplementary Tables 1 and 2, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/doi/10.002/art.37777/abstract).

Because the 3 loci were associated with RA in a Japanese population, we analyzed whether the associations with SSc in the current study were contributed by patients with both RA and SSc. When 22 patients who had RA as well as SSc were excluded, significant associations for the 3 loci were still observed (Table 2). A

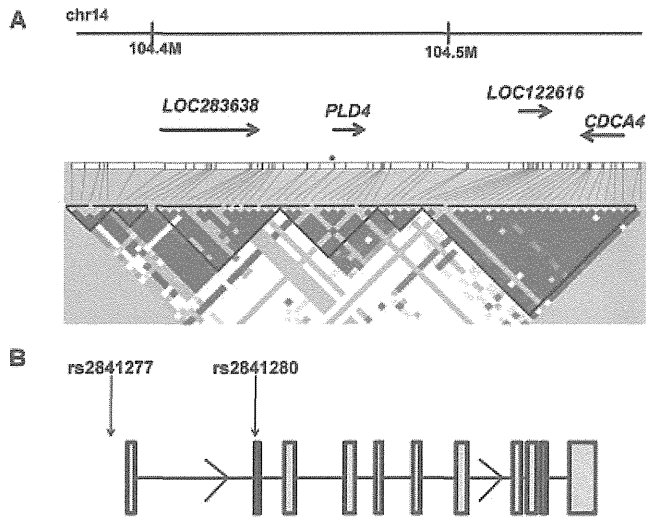

Figure 1. Linkage disequilibrium (LD) block around the *PLD4* region and the *PLD4* structure. **A,** LD block and genes around *PLD4*. The LD block is based on HapMap phase 3 data. **Asterisk** indicates rs2841277. **B,** Schematic view of *PLD4* structure. Rectangles represent exons of *PLD4*.

further stringent analysis excluding patients with other autoimmune diseases demonstrated significant associations of the 3 genes (see Supplementary Table 2). When we compared SSc patients with and those without other autoimmune diseases for the associated alleles, no differences were observed (data not shown).

Subanalysis of types of SSc. Previous studies have revealed that the genetic background of SSc varies between different types of SSc (11,18). Thus, subanalyses of the 5 regions examined in the combined study were performed, in which the allele frequencies of the control subjects were compared with those of the patients with lcSSc or dcSSc. The control subjects were the same as those used in the first study or the combined study. Although *PLD4* and *TNFAIP3* did not display a preference for either SSc phenotype, *IRF8* and *ARID5* showed suggestive preferences for lcSSc, and *CD83* showed a suggestive preference for dcSSc (Table 3).

We also investigated whether the susceptibility loci affect autoantibody status and severe complications. The association studies revealed an association of *TNFAIP3* with SSc patients who possess anticentromere antibodies (ACAs) (see Supplementary Table 3, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.37777/abstract), but intracase analyses did not demonstrate clear significance ($P = 0.043$). We did not observe other associations between the susceptibility loci and clinical phenotypes of SSc, in either case–control analyses or intracase analyses.

Efficacy of the current study. In the current study, a candidate gene analysis was performed based on a meta-analysis of RA GWAS, because many susceptibility genes for autoimmune disease have been reported

Table 3. Associations of the 2 SSc subtypes*

SNP	Chr	Gene	Allele 1/2	Controls, allele 1 frequency	Limited cutaneous SSc (n = 408)			Diffuse cutaneous SSc (n = 318)		
					Allele 1 frequency	*P*	OR (95% CI)	Allele 1 frequency	*P*	OR (95% CI)
rs766449	1	*PADI4*	T/C	0.40	0.39	0.52	0.94 (0.77–1.14)	0.36	0.11	0.85 (0.69–1.04)
rs11900673	2	*B3GNT2*	T/C	0.29	0.25	0.096	0.82 (0.66–1.03)	0.31	0.32	1.11 (0.9–1.38)
rs2867461	4	*ANXA3*	A/G	0.44	0.42	0.40	0.92 (0.75–1.12)	0.44	0.97	1.00 (0.82–1.22)
rs657075	5	*IL3-CSF2*	A/G	0.36	0.34	0.54	0.94 (0.76–1.15)	0.33	0.23	0.88 (0.72–1.08)
rs12529514	6	*CD83*	C/T	0.14	0.15	0.79	1.03 (0.85–1.25)	0.18	0.0075	1.32 (1.08–1.62)
rs1571878	6	*CCR6*	C/T	0.49	0.48	0.81	0.98 (0.80–1.19)	0.46	0.20	0.88 (0.72–1.07)
rs6932056	6	*TNFAIP3*	C/T	0.069	0.093	0.0062	1.40 (1.1–1.78)	0.10	0.00063	1.57 (1.21–2.04)
rs2233434	6	*NFKBIE*	G/A	0.21	0.20	0.60	0.94 (0.73–1.20)	0.22	0.70	1.05 (0.83–1.33)
rs10821944	10	*ARID5B*	G/T	0.36	0.40	0.0085	1.22 (1.05–1.41)	0.38	0.30	1.09 (0.93–1.29)
rs3781913	11	*PDE2A-CENTD2*	T/G	0.69	0.69	0.98	1.00 (0.81–1.24)	0.69	0.90	1.01 (0.82–1.25)
rs2841277	14	*PLD4*	T/C	0.69	0.73	0.0067	1.24 (1.06–1.45)	0.74	0.0049	1.29 (1.08–1.55)
rs2841280	14	*PLD4*	C/G	0.64	0.69	0.0011	1.30 (1.11–1.52)	0.69	0.0086	1.27 (1.06–1.51)
rs2847297	18	*PTPN2*	G/A	0.34	0.33	0.67	0.96 (0.78–1.18)	0.34	0.87	1.02 (0.83–1.25)
rs4937362	11	*ETS1-FLI1*	T/C	0.68	0.68	0.75	0.97 (0.78–1.19)	0.69	0.92	1.01 (0.82–1.25)
rs3783637	14	*GCH1*	C/T	0.74	0.73	0.69	0.96 (0.77–1.19)	0.73	0.65	0.95 (0.76–1.18)
rs1957895	14	*PRKCH*	G/T	0.39	0.40	0.84	1.02 (0.84–1.25)	0.42	0.16	1.15 (0.95–1.41)
rs6496667	15	*ZNF774*	A/C	0.35	0.39	0.088	1.19 (0.97–1.45)	0.34	0.75	0.97 (0.79–1.19)
rs7404928	16	*PRKCB1*	T/C	0.62	0.61	0.60	0.95 (0.78–1.16)	0.66	0.15	1.17 (0.95–1.44)
rs2280381	16	*IRF8*	T/C	0.84	0.88	0.0038	1.36 (1.11–1.68)	0.86	0.21	1.16 (0.92–1.45)

* SSc = systemic sclerosis; SNP = single-nucleotide polymorphism; Chr = chromosome; OR = odds ratio; 95% CI = 95% confidence interval.

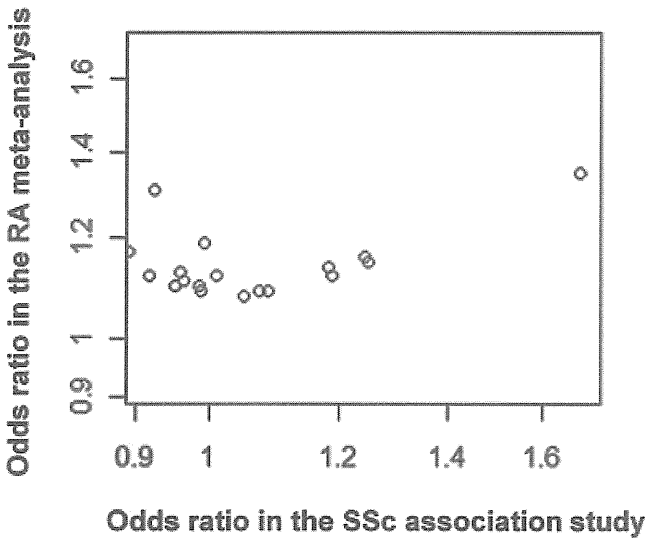

Figure 2. Comparison of associations for systemic sclerosis (SSc) and rheumatoid arthritis (RA). The odds ratios obtained for 18 genes in association studies of SSc and RA are plotted.

to be shared by a wide range of diseases. As a result, 3 susceptibility genes for SSc in Japanese were identified. Thus, we analyzed whether the candidate gene approach taken in the current study for detecting novel susceptibility genes for SSc was effective. When the likelihood of finding 3 susceptibility genes among 18 genes by chance was calculated, the likelihood was determined to be 2.5×10^{-8}. These results indicated that our approach to identifying novel susceptibility genes for systemic diseases is effective. It would be interesting to compare the risk direction of the genotyped markers between RA and SSc. Although the 3 susceptibility loci for SSc shared risk direction with RA, no correspondence of the risk directions of the markers between the 2 diseases was detected (Figure 2). This indicated that a large proportion of the 18 RA markers are not shared by SSc, and that the lack of association between the 13 markers and SSc was not attributable to the low power produced by the relatively small number of SSc patients included in this study.

DISCUSSION

Because SSc can lead to severe complications, poor quality of life, and shortened survival, clarifying the characteristics of SSc is important. Clarification of the disease would aid the search for novel therapeutic targets and the development of new therapeutic strategies. Detecting susceptibility genes using GWAS or a candidate gene approach would also help to uncover the pathophysiology underlying SSc.

Previous studies have revealed that more than 15 markers and loci are associated with SSc. However, the markers detected so far cannot fully explain the genetics of SSc, indicating that many susceptibility genes are yet to be identified. Because a relatively large proportion of RA susceptibility genes are shared by other autoimmune diseases (24), a candidate gene approach using novel markers observed in GWAS of RA is a fascinating way of identifying new SSc markers. In fact, some of the novel susceptibility markers for RA identified in the meta-analysis were shown to be susceptibility markers for systemic lupus erythematosus (SLE) and Graves' disease (31).

In the current study, we successfully identified 3 susceptibility genes for SSc in Japanese. No studies have identified *PLD4* as an SSc-associated locus. The current study is also the first to detect *TNFAIP3* and *IRF8* as susceptibility genes for SSc in a Japanese population. The best-fit models for each association are shown in Supplementary Table 4, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/doi/10.002/art.37777/abstract.

It is conceivable that these 3 associations might have been obtained due to the overlap of RA and SSc. Even after excluding the patients with both RA and SSc based on physicians' reports, the significant associations for the 3 loci were still observed (Table 3). Information regarding rheumatoid factor (RF) and anti–citrullinated protein antibody (ACPA) was available for 371 SSc patients without RA and 65 SSc patients without RA, respectively, of whom 21.6% and 10.8% were positive for RF and ACPA, respectively. These prevalences are compatible with those previously observed in SSc patients without RA (35,36). Moreover, we showed that the effect sizes and risk direction of the markers tested in this study were dissociated between SSc and RA. In addition, further stringent analysis comprising SSc patients without any autoimmune disease also showed the associations of the 3 loci. These results indicate that the associations of the 3 loci are not attributable to overlapping of RA or other diseases.

Although the associations of the *ARID5B* and *CD83* loci with SSc did not reach a stringently significant level in the combined study, the tendencies toward an association with SSc displayed by rs10821944 in the *ARID5B* locus and rs12529514 in the *CD83* region in the first study were maintained in the replication study. This indicates that these loci are potential susceptibility regions for SSc. Further replication studies are needed to

address the associations of these 2 loci with SSc in a Japanese population.

Because *TNFAIP3* was reported to be strongly associated with SSc in a European population (18), the significant associations detected in the combined study indicate that *TNFAIP3* displays general associations with SSc that go beyond ethnic boundaries. In addition, rs6932056, which displayed a strong association with SSc in a European population (18), is in strong LD with rs5029939 ($r^2 = 0.85$) in the Japanese population. SNP rs6932056 also displays strong LD with rs2230926, a missense mutation of *TNFAIP3* ($r^2 = 0.85$), in Japanese. The rs2230926 missense mutation leads to an amino acid alteration in the OTU (ovarian tumor) domain of the A20 protein, which is considered to result in decreased NF-κB signaling. Because we did not observe strong associations between rs6932056 and SSc in the replication study, it will be necessary to reexamine the association between *TNFAIP3* and SSc using independent sample sets of Japanese patients with SSc, in spite of the significant associations detected in this study.

PLD4 is a recently reported member of the phospholipase family without phospholipase D activity. *PLD4* is expressed in the spleen and early postnatal microglia in the white matter of mice (37). The phenotypes of *Pld4*-deficient mice have not been reported. In addition, little is known about the expression or distribution of *PLD4* in humans. Although the functions of *PLD4* are also poorly understood, it is known to be involved in the phagocytosis of microglia (38). The expression of *PLD4* around the marginal zone in the spleen might support the functional involvement of *PLD4* in immunologic systems. It is interesting that rs2841280, which alters an amino acid of PLD-4, is associated with SSc. Minor allele G of rs2841280 is associated in a protective manner. The impact of an amino acid alteration brought by rs2841280 on the effect of PLD-4 protein is not known.

When we analyzed the impact of the amino acid alteration using in silico analysis (SIFT software; http://sift.jcvi.org/), it was shown to result in a small effect. However, the association raises the possibility that this polymorphism leads functional modulation of PLD-4, and it is feasible to analyze the functional change of PLD-4 protein with rs2841280, using animal models of SSc. When we performed an in silico analysis of the effect of rs2841277 and rs2841280 on *PLD4* expression, we did not detect any clear associations between the 2 genotypes and *PLD4* transcription ($P > 0.05$) (39). Therefore, in spite of the association of these 2 mutations, it has not been confirmed whether one of these 2 polymorphisms is the causative mutation.

Although the detection of a P value less than 5×10^{-8} in a GWAS is stringent evidence of an association between a marker and a particular disease, the detection of suggestive associations between the *PLD4* region and SSc in European GWAS would indicate that associations exist between *PLD4* and SSc in other populations. However, when we examined the associations between the *PLD4* locus or nearby loci and SSc in GWAS involving a European population, we did not detect any strong associations ($P < 10^{-4}$) (8,9). According to the HapMap database, the European population displays a higher risk allele frequency for rs2841277 than the Japanese population. In addition, the HapMap database also indicates that the LD block spanning *PLD4*, which includes rs2841277, is similar in Europeans and Japanese. Nevertheless, a European population did not show a strong association between *PLD4* and SSc, suggesting that *PLD4* has a stronger effect on autoimmune diseases in Japanese than in Europeans. There is also a possibility that these 2 polymorphisms are only markers, and that a rare variant in LD with the 2 markers affects disease onset. A rare causative variant might explain a different association of *PLD4* with SSc between populations.

IRF8 was shown to be associated with SLE in a European population (40). Interferon regulatory factor 8 (IRF-8) protein is a transcription factor involved in the interferon pathway. The interferon pathway has been shown to be involved with a broad range of autoimmune diseases, including SSc (41). Thus, it is interesting that *IRF5* and *IRF8*, both of which belong to the IRF family, displayed associations with SSc. Although a European GWAS of SSc patients revealed suggestive associations between the *IRF4* locus and SSc, the results were not successfully replicated (8), indicating that the different functional roles of each IRF family molecule might influence the development of SSc. *IRF8* promotes B cell differentiation; however, the roles and importance of B cells in skin fibrosis in SSc patients have not been established (42–44). *IRF8* and its mutant variants are also known to be involved in the development of dendritic cells (45). Thus, the association between *IRF8* and SSc might indicate the involvement of B cells and dendritic cells in the development of SSc.

When the patients with SSc were classified as having either lcSSc or dcSSc and subanalyses were performed, *ARID5B*, *IRF8*, and *CD83* displayed stronger associations with one of the 2 phenotypes. However, the associations of these 3 markers with the phenotypes

were not strong enough to provide convincing evidence of a clear distinction between the genetic backgrounds of the 2 SSc phenotypes. When the associations of the SSc subtypes with the other 13 markers in the first set were analyzed, no strong association was detected ($P > 0.05$). Other subanalyses of the susceptibility loci in the combined set did not show significant results between disease phenotypes, due to lack of power. Because classification according to disease phenotypes resulted in limited numbers of subjects in each subset, we conducted this subanalysis only in the combined set. The association between *TNFAIP3* and ACAs should be confirmed in a large-scale association study.

Although GWAS are an extremely powerful way to detect novel susceptibility genes for diseases, GWAS of patients with SSc have been performed only in European populations. Our study detected strong evidence for the sharing of susceptibility genes between RA and SSc in a Japanese population. In addition, the current study indicated that a candidate gene approach based on the results of GWAS of other diseases that display pathologic signaling pathways or mechanisms similar to those associated with the disease being examined is an effective approach to identifying novel susceptibility genes.

It will be interesting to perform GWAS of Japanese patients with SSc and analyze the similarities and differences in the detected associations not only between Japanese and Europeans but also between Japanese patients with SSc and Japanese patients with RA.

ACKNOWLEDGMENT

We thank the staff of the BioBank Japan Project for collecting DNA samples from control subjects.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Terao had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Terao, Ohmura, Kawaguchi, Nishimoto, Kawasaki, Takehara, Furukawa, Kochi, Ota, Ikari, Sato, Tohma, Yamada, Yamamoto, Kubo, Yamanaka, Kuwana, Tsuchiya, Matsuda, Mimori.

Acquisition of data. Terao, Ohmura, Kawaguchi, Nishimoto, Kawasaki, Takehara, Furukawa, Kochi, Ota, Ikari, Sato, Tohma, Yamada, Yamamoto, Kubo, Yamanaka, Kuwana, Tsuchiya, Matsuda, Mimori.

Analysis and interpretation of data. Terao, Ohmura.

REFERENCES

1. Chifflot H, Fautrel B, Sordet C, Chatelus E, Sibilia J. Incidence and prevalence of systemic sclerosis: a systematic literature review. Semin Arthritis Rheum 2008;37:223–35.
2. Kawut SM, Taichman DB, Archer-Chicko CL, Palevsky HI, Kimmel SE. Hemodynamics and survival in patients with pulmonary arterial hypertension related to systemic sclerosis. Chest 2003;123:344–50.
3. Ioannidis JP, Vlachoyiannopoulos PG, Haidich AB, Medsger TA Jr, Lucas M, Michet CJ, et al. Mortality in systemic sclerosis: an international meta-analysis of individual patient data. Am J Med 2005;118:2–10.
4. Bhattacharyya S, Wei J, Varga J. Understanding fibrosis in systemic sclerosis: shifting paradigms, emerging opportunities. Nat Rev Rheumatol 2012;8:42–54.
5. Romano E, Manetti M, Guiducci S, Ceccarelli C, Allanore Y, Matucci-Cerinic M. The genetics of systemic sclerosis: an update. Clin Exp Rheumatol 2011;29:S75–86.
6. Arnett FC, Cho M, Chatterjee S, Aguilar MB, Reveille JD, Mayes MD. Familial occurrence frequencies and relative risks for systemic sclerosis (scleroderma) in three United States cohorts. Arthritis Rheum 2001;44:1359–62.
7. Terao C, Ohmura K, Katayama M, Takahashi M, Kokubo M, Diop G, et al. Myelin basic protein as a novel genetic risk factor in rheumatoid arthritis: a genome-wide study combined with immunological analyses. PLoS One 2011;6:e20457.
8. Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet 2010;42:426–9.
9. Allanore Y, Saad M, Dieude P, Avouac J, Distler JH, Amouyel P, et al. Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. PLoS Genet 2011;7:e1002091.
10. Gorlova O, Martin JE, Rueda B, Koeleman BP, Ying J, Teruel M, et al. Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. PLoS Genet 2011;7:e1002178.
11. Rueda B, Broen J, Simeon C, Hesselstrand R, Diaz B, Suarez H, et al. The STAT4 gene influences the genetic predisposition to systemic sclerosis phenotype. Hum Mol Genet 2009;18:2071–7.
12. Dieude P, Guedj M, Wipff J, Avouac J, Fajardy I, Diot E, et al. Association between the IRF5 rs2004640 functional polymorphism and systemic sclerosis: a new perspective for pulmonary fibrosis. Arthritis Rheum 2009;60:225–33.
13. Gourh P, Agarwal SK, Divecha D, Assassi S, Paz G, Arora-Singh RK, et al. Polymorphisms in TBX21 and STAT4 increase the risk of systemic sclerosis: evidence of possible gene–gene interaction and alterations in Th1/Th2 cytokines. Arthritis Rheum 2009;60:3794–806.
14. Dieude P, Guedj M, Wipff J, Ruiz B, Riemekasten G, Airo P, et al. NLRP1 influences the systemic sclerosis phenotype: a new clue for the contribution of innate immunity in systemic sclerosis-related fibrosing alveolitis pathogenesis. Ann Rheum Dis 2011;70:668–74.
15. Bossini-Castillo L, Broen JC, Simeon CP, Beretta L, Vonk MC, Ortego-Centeno N, et al. A replication study confirms the association of TNFSF4 (OX40L) polymorphisms with systemic sclerosis in a large European cohort. Ann Rheum Dis 2011;70:638–41.
16. Dieude P, Guedj M, Truchetet ME, Wipff J, Revillod L, Riemekasten G, et al. Association of the CD226 Ser^{307} variant with systemic sclerosis: evidence of a contribution of costimulation pathways in systemic sclerosis pathogenesis. Arthritis Rheum 2011;63:1097–105.
17. Gourh P, Agarwal SK, Martin E, Divecha D, Rueda B, Bunting H, et al. Association of the C8orf13-BLK region with systemic sclerosis in North-American and European populations. J Autoimmun 2010;34:155–62.
18. Dieude P, Guedj M, Wipff J, Ruiz B, Riemekasten G, Matucci-Cerinic M, et al. Association of the TNFAIP3 rs5029939 variant

with systemic sclerosis in the European Caucasian population. Ann Rheum Dis 2010;69:1958–64.
19. Tsuchiya N, Kawasaki A, Hasegawa M, Fujimoto M, Takehara K, Kawaguchi Y, et al. Association of STAT4 polymorphism with systemic sclerosis in a Japanese population. Ann Rheum Dis 2009;68:1375–6.
20. Ito I, Kawaguchi Y, Kawasaki A, Hasegawa M, Ohashi J, Hikami K, et al. Association of a functional polymorphism in the IRF5 region with systemic sclerosis in a Japanese population. Arthritis Rheum 2009;60:1845–50.
21. Ito I, Kawaguchi Y, Kawasaki A, Hasegawa M, Ohashi J, Kawamoto M, et al. Association of the FAM167A–BLK region with systemic sclerosis. Arthritis Rheum 2010;62:890–5.
22. Hasebe N, Kawasaki A, Ito I, Kawamoto M, Hasegawa M, Fujimoto M, et al. Association of UBE2L3 polymorphisms with diffuse cutaneous systemic sclerosis in a Japanese population. Ann Rheum Dis 2012;71:1259–60.
23. Zhou X, Lee JE, Arnett FC, Xiong M, Park MY, Yoo YK, et al. HLA–DPB1 and DPB2 are genetic loci for systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans. Arthritis Rheum 2009;60:3807–14.
24. Suzuki A, Kochi Y, Okada Y, Yamamoto K. Insight from genome-wide association studies in rheumatoid arthritis and multiple sclerosis. FEBS Lett 2011;585:3627–32.
25. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, et al. TRAF1-C5 as a risk locus for rheumatoid arthritis: a genomewide study. N Engl J Med 2007;357:1199–209.
26. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 2007;357:977–86.
27. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 2007;39:1477–82.
28. Thomson W, Barton A, Ke X, Eyre S, Hinks A, Bowes J, et al. Rheumatoid arthritis association at 6q23. Nat Genet 2007;39: 1431–3.
29. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, et al. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet 2004;75:330–7.
30. Diaz-Gallo LM, Gourh P, Broen J, Simeon C, Fonollosa V, Ortego-Centeno N, et al. Analysis of the influence of PTPN22 gene polymorphisms in systemic sclerosis. Ann Rheum Dis 2011; 70:454–62.
31. Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, et al. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nat Genet 2012;44:511–6.
32. Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Arthritis Rheum 1980;23:581–90.
33. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA Jr, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988;15:202–5.
34. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature 2010;467: 1061–73.
35. Mimura Y, Ihn H, Jinnin M, Asano Y, Yamane K, Yazawa N, et al. Rheumatoid factor isotypes and anti-agalactosyl IgG antibodies in systemic sclerosis. Br J Dermatol 2004;151:803–8.
36. Santiago M, Baron M, Miyachi K, Fritzler MJ, Abu-Hakima M, Leclercq S, et al. A comparison of the frequency of antibodies to cyclic citrullinated peptides using a third generation anti-CCP assay (CCP3) in systemic sclerosis, primary biliary cirrhosis and rheumatoid arthritis. Clin Rheumatol 2008;27:77–83.
37. Yoshikawa F, Banno Y, Otani Y, Yamaguchi Y, Nagakura-Takagi Y, Morita N, et al. Phospholipase D family member 4, a transmembrane glycoprotein with no phospholipase D activity, expression in spleen and early postnatal microglia. PLoS One 2010;5: e13932.
38. Otani Y, Yamaguchi Y, Sato Y, Furuichi T, Ikenaka K, Kitani H, et al. PLD4 is involved in phagocytosis of microglia: expression and localization changes of PLD4 are correlated with activation state of microglia. PLoS One 2011;6:e27544.
39. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 2007;315: 848–53.
40. Cunninghame Graham DS, Morris DL, Bhangale TR, Criswell LA, Syvanen AC, Ronnblom L, et al. Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. PLoS Genet 2011;7:e1002341.
41. Higgs BW, Liu Z, White B, Zhu W, White WI, Morehouse C, et al. Patients with systemic lupus erythematosus, myositis, rheumatoid arthritis and scleroderma share activation of a common type I interferon pathway. Ann Rheum Dis 2011;70:2029–36.
42. Whitfield ML, Finlay DR, Murray JI, Troyanskaya OG, Chi JT, Pergamenschikov A, et al. Systemic and cell type-specific gene expression patterns in scleroderma skin. Proc Natl Acad Sci U S A 2003;100:12319–24.
43. Lafyatis R, Kissin E, York M, Farina G, Viger K, Fritzler MJ, et al. B cell depletion with rituximab in patients with diffuse cutaneous systemic sclerosis. Arthritis Rheum 2009;60:578–83.
44. Smith V, Van Praet JT, Vandooren B, Van der Cruyssen B, Naeyaert JM, Decuman S, et al. Rituximab in diffuse cutaneous systemic sclerosis: an open-label clinical and histopathological study. Ann Rheum Dis 2010;69:193–7.
45. Hambleton S, Salem S, Bustamante J, Bigley V, Boisson-Dupuis S, Azevedo J, et al. IRF8 mutations and human dendritic-cell immunodeficiency. N Engl J Med 2011;365:127–38.

ACPA-Negative RA Consists of Two Genetically Distinct Subsets Based on RF Positivity in Japanese

Chikashi Terao[1,2]*, Koichiro Ohmura[1]*, Katsunori Ikari[3], Yuta Kochi[4], Etsuko Maruya[5], Masaki Katayama[1], Kimiko Yurugi[6], Kota Shimada[7], Akira Murasawa[8], Shigeru Honjo[9], Kiyoshi Takasugi[10], Keitaro Matsuo[11], Kazuo Tajima[11], Akari Suzuki[4], Kazuhiko Yamamoto[12], Shigeki Momohara[3], Hisashi Yamanaka[3], Ryo Yamada[2], Hiroo Saji[5], Fumihiko Matsuda[2,13,14], Tsuneyo Mimori[1]

1 Department of Rheumatology and Clinical Immunology, Kyoto University Graduate School of Medicine, Kyoto, Japan, **2** Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, **3** Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan, **4** Laboratory for Autoimmune Diseases, Center for Genomic Medicine, RIKEN, Yokohama, Japan, **5** HLA Laboratory, Kyoto, Japan, **6** Department of Transfusion Medicine and Cell Therapy, Kyoto University Hospital, Kyoto, Japan, **7** Department of Rheumatology, Sagamihara National Hospital, National Hospital Organization, Sagamihara, Japan, **8** Department of Rheumatology, Niigata Rheumatic Center, Niigata, Japan, **9** Rheumatoid Arthritis Center, Saiseikai Takaoka Hospital, Toyama, Japan, **10** Department of Internal Medicine, Center for Rheumatic Diseases, Dohgo Spa Hospital, Matsuyama, Japan, **11** Aichi Cancer Center Hospital and Research Institute, Nagoya, Japan, **12** Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan, **13** CREST program, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan, **14** Institut National de la Sante et de la Recherche Medicale (INSERM) Unite U852, Kyoto University Graduate School of Medicine, Kyoto, Japan

Abstract

HLA-DRB1, especially the shared epitope (SE), is strongly associated with rheumatoid arthritis (RA). However, recent studies have shown that SE is at most weakly associated with RA without anti-citrullinated peptide/protein antibody (ACPA). We have recently reported that ACPA-negative RA is associated with specific HLA-DRB1 alleles and diplotypes. Here, we attempted to detect genetically different subsets of ACPA-negative RA by classifying ACPA-negative RA patients into two groups based on their positivity for rheumatoid factor (RF). HLA-DRB1 genotyping data for totally 954 ACPA-negative RA patients and 2,008 healthy individuals in two independent sets were used. HLA-DRB1 allele and diplotype frequencies were compared among the ACPA-negative RF-positive RA patients, ACPA-negative RF-negative RA patients, and controls in each set. Combined results were also analyzed. A similar analysis was performed in 685 ACPA-positive RA patients classified according to their RF positivity. As a result, HLA-DRB1*04:05 and *09:01 showed strong associations with ACPA-negative RF-positive RA in the combined analysis ($p = 8.8 \times 10^{-6}$ and 0.0011, OR: 1.57 (1.28–1.91) and 1.37 (1.13–1.65), respectively). We also found that HLA-DR14 and the HLA-DR8 homozygote were associated with ACPA-negative RF-negative RA ($p = 0.00022$ and 0.00013, OR: 1.52 (1.21–1.89) and 3.08 (1.68–5.64), respectively). These association tendencies were found in each set. On the contrary, we could not detect any significant differences between ACPA-positive RA subsets. As a conclusion, ACPA-negative RA includes two genetically distinct subsets according to RF positivity in Japan, which display different associations with HLA-DRB1. ACPA-negative RF-positive RA is strongly associated with HLA-DRB1*04:05 and *09:01. ACPA-negative RF-negative RA is associated with DR14 and the HLA-DR8 homozygote.

Citation: Terao C, Ohmura K, Ikari K, Kochi Y, Maruya E, et al. (2012) ACPA-Negative RA Consists of Two Genetically Distinct Subsets Based on RF Positivity in Japanese. PLoS ONE 7(7): e40067. doi:10.1371/journal.pone.0040067

Editor: Pierre Bobé, Institut Jacques Monod, France

Received March 10, 2012; **Accepted** May 31, 2012; **Published** July 6, 2012

Funding: This study was supported by Grants-in-aid from the Ministry of Health, Labor, and Welfare of Japan and from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, as well as by research grants from the Japan Rheumatism Foundation, the Waksman Foundation, and the Mitsubishi Pharma Research Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No additional external funding received for this study.

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

* E-mail: a0001101@kuhp.kyoto-u.ac.jp (CT); ohmurako@kuhp.kyoto-u.ac.jp (KO)

Introduction

Rheumatoid arthritis (RA) is the most common cause of chronic arthritis worldwide and results in severe joint destruction [1]. Genetic and environmental factors have been shown to be associated with its onset [2–3]. Among the susceptibility genes to RA, HLA-DRB1 has been shown to be the strongest genetic determinant of RA susceptibility, and its association with RA susceptibility has been repeatedly shown to be independent of ethnicity [4–5]. A common amino acid sequence extending from the 70th to 74th in the HLA-DRβ chain, which is known as the "shared epitope (SE)", is considered to be the reason for the association between HLA-DRB1 and RA, and the association between the SE and RA has been reported to be ethnicity-independent [6–8]. However, recent studies have shown that the SE is strongly associated with RA patients who have anti-citrullinated peptide/protein antibodies (ACPA), which is a highly specific marker of RA [9], but that it is not or only weakly associated with RA without ACPA [7,10–11]. Among the various HLA-DRB1 alleles, HLA-DR3 [12] and HLA-DR13 [13] were reported to be associated with ACPA-negative RA in populations of European descent, but these results were not confirmed in a

meta-analysis of a large Caucasian cohort [8]. In Asian populations, we recently reported that DRB1*12:01 is a HLA-DRB1 susceptibility allele for ACPA-negative RA in Japanese populations and that DRB1*04:05, the most common SE allele in Japanese, and *14:03 showed moderate associations with ACPA-negative RA susceptibility [14]. We also reported that DRB1*15:02 and *13:02 displayed protective associations with ACPA-negative RA and that being homozygous for HLA-DR8 was associated with ACPA-negative RA susceptibility. While a very small Japanese study suggested that HLA-DRB1*09:01 is associated with ACPA-negative RA [15], our study did not detect a significant association between them. These findings suggest that ACPA-negative RA is genetically different from ACPA-positive RA in terms of its associations with HLA-DRB1 alleles. While some specific alleles and diplotypes seem to be associated with ACPA-negative RA, the genetic characteristics of ACPA-negative RA have not been fully elucidated. Recently, UK group reported that SE is associated with ACPA-negative RF-positive RA in UK population [16]. However, whether this is true to other population is uncertain. Moreover, the associations of other alleles than SE with subgroups of ACPA-negative RA have never been reported. Here, we show that when we classified ACPA-negative RA into two subsets based on rheumatoid factor (RF) positivity, we were able to clearly distinguish them from each other according to their associations with HLA-DRB1 alleles, not only with SE, but with other alleles. We also compared ACPA-positive RA patients based on their RF positivity to examine whether we can apply this classification to ACPA-positive RA.

Results

HLA-DRB1 Alleles Associated with ACPA-negative RF-positive RA

We compared 179 ACPA-negative RF-positive RA with 1508 controls in collection 1 for their frequency of HLA-DRB1 alleles, followed by comparison of 267 ACPA-negative RF-positive RA with 500 controls in collection 2. Significant association was evaluated in the combined analysis. Regarding HLA-DRB1 alleles that were previously shown to be associated with ACPA-negative RA, we found that all of the alleles, namely, HLA-DRB1*12:01, *04:05, *13:02, *14:03, and *15:02 showed association tendency with ACPA-negative RF-positive RA in the combined study (Table 1). Interestingly, HLA-DRB1*04:05 ($p = 8.8 \times 10^{-6}$, odds ratio (OR): 1.57) showed the strongest association, while its association with entire ACPA-negative RA was moderate in the previous study. When we analyzed the associations of the SE, we found that it displayed a significant association ($p = 0.00013$, OR: 1.37). HLA-DRB1*04:05 was responsible for most of the association of SE because none of the other SE alleles showed significant associations with ACPA-negative RF-positive RA. We also found that HLA-DRB1*09:01, which was not associated with ACPA-negative RA as a single allele, was found to be significantly associated with ACPA-negative RF-positive RA ($p = 0.0011$, OR: 1.37). Importantly, these association tendencies written above were observed in both collections (Table 1). Logistic regression analysis was carried out to examine whether the susceptibility associations were dependent on a lack of protective alleles or vice versa. As a result, it was demonstrated that HLA-DRB1*04:05, *09:01, and *12:01 showed significant associations ($p<0.0005$), while the associations of HLA-DRB1*14:03, *13:02, and *15:02 were moderate to suggestive (Table S1). Next, we analyzed the dosage effects of the alleles and found that the association between HLA-DRB1*09:01 and ACPA-negative RF-positive RA showed a clear dosage effect (Figure S1). HLA-DRB1*12:01 also showed a dosage effect (data not shown due to small number). HLA-DRB1*04:05 did not show a dosage effect, suggesting that the effect of HLA-DRB1*04:05 on the predisposition to ACPA-negative RF-positive RA is a dominant effect.

HLA-DRB1 Alleles Associated with ACPA-negative RF-negative RA

Next we compared 274 ACPA-negative RF-negative RA with 1,508 controls, followed by comparison between 234 ACPA-negative RF-negative RA and 500 controls. Interestingly, we did not observe association of HLA-DRB1*04:05 and *09:01 with ACPA-negative RF-negative RA, while HLA-DRB1*12:01, *13:02, *14:03, and *15:02 were moderately associated with ACPA-negative RF-negative RA (Table 2). The SE was not associated with ACPA-negative RF-negative RA. DR14 was found to be significantly associated with ACPA-negative RF-negative RA and HLA-DRB1*14:03 and *14:06 comprised the association of HLA-DR14 (Table S2). These association tendencies in ACPA-negative RF-negative RA were observed in both sets (Table 2). Logistic regression analysis confirmed that none of the associations were mutually dependent and that the association of DR14 remained significant ($p = 0.00069$, Table S3). DR14 could not be evaluated the dosage effect because neither the cases nor controls included DRB1*14:03 or *14:06 homozygotes or the DRB1*14:03 and *14:06 diplotype.

HLA Diplotype Analysis: DR8 Homozygote and *12:01/*09:01 Diplotype

As we previously showed that the DR8 homozygote was significantly associated with susceptibility to ACPA-negative RA, we analyzed its associations with ACPA-negative RF-positive RA and RF-negative RA. As a result, we found that the HLA-DR8 homozygote is exclusively associated with ACPA-negative RF-negative RA in the combined study ($p = 0.00013$, OR: 3.08 for ACPA-negative RF-negative RA, Table 2; $p = 0.86$, OR: 1.08 for ACPA-negative RF-positive RA, Table 1). The effect of DR8 on the susceptibility to ACPA-negative RF-negative RA was not dose-dependent (OR: 1.04 for HLA-DR8 heterozygote).

We also found that the combination of HLA-DRB1*12:01 and *09:01, the diplotype that was most strongly associated with susceptibility to ACPA-negative RA in the previous study, was especially strongly associated with ACPA-negative RF-positive RA ($p = 5.0 \times 10^{-6}$, OR: 4.97 for ACPA-negative RF-positive RA; $p = 0.040$, OR: 2.46 for ACPA-negative RF-negative RA).

We found that the similar associations were seen between the alleles/diplotypes and ACPA-negative RF-positive erosive RA and ACPA-negative RF-negative erosive RA (except for that between HLA-DRB1*12:01 and the ACPA-negative RF-negative subset), even though the number of patients was limited (Table S4).

Comparison between ACPA-negative RF-positive RA and ACPA-negative RF-negative RA

To compare the usage of HLA-DRB1 allele between ACPA-negative RF-positive RA and ACPA-negative RF-negative RA, we directly compared the allele and diplotype frequencies between the two groups (Table 3). As expected, HLA-DRB1*09:01 and *04:05 showed significant differences in their frequencies between the two subsets ($p = 0.0018$ and 0.0034, respectively). The SE was more common in the ACPA-negative RF-positive RA patients ($p = 0.0047$), whereas DR14 was more prevalent in the ACPA-negative RF-negative RA patients ($p = 0.028$). The DR8 homozygote was more frequently seen in the ACPA-negative RF-negative RA patients than in the ACPA-negative RF-positive RA patients

Table 1. Association of HLA-DRB1 alleles with ACPA-negative RF-positive RA.

	1st set				2nd set				combined analysis			
HLA-DRB1 allele	§ACPA (-)RF(+)RA	§control	p	OR	§ACPA (-)RF(+)RA	§control	p	OR	§ACPA (-)RF(+)RA	§control	p	OR
*04:05	65 (18.2%)	340 (11.3%)	0.00015	1.75 (1.30–2.34)	88 (16.5%)	129 (12.9%)	0.055	1.33 (0.99–1.79)	153 (17.2%)	469 (11.7%)	8.8×10^{-6}	1.57 (1.28–1.91)
*09:01	70 (19.6%)	432 (14.3%)	0.0086	1.45 (1.10–1.92)	99 (18.5%)	154 (15.4%)	0.11	1.25 (0.95–1.65)	169 (18.9%)	586 (14.6%)	0.0011	1.37 (1.13–1.65)
*12:01	13 (3.6%)	91 (3%)	0.53	1.21 (0.67–2.19)	35 (6.6%)	37 (3.7%)	0.012	1.83 (1.14–2.93)	48 (5.4%)	128 (3.2%)	0.0014	1.73 (1.23–2.43)
*13:02	21 (5.9%)	273 (9.1%)	0.043	0.63 (0.40–0.99)	18 (3.4%)	52 (5.2%)	0.10	0.64 (0.37–1.1)	39 (4.4%)	325 (8.1%)	0.00013	0.52 (0.37–0.73)
*14:03	7 (2.0%)	39 (1.3%)	0.31	1.52 (0.68–3.43)	13 (2.4%)	14 (1.4%)	0.14	1.76 (0.82–3.77)	20 (2.2%)	53 (1.3%)	0.040	1.71 (1.02–2.88)
*15:02	43 (12.0%)	369 (12.2%)	0.90	0.98 (0.70–1.37)	37 (6.9%)	113 (11.3%)	0.0060	0.58 (0.4–0.86)	80 (9.0%)	482 (12.0%)	0.010	0.72 (0.56–0.93)
SE	106 (29.6%)	677 (22.4%)	0.0024	1.45 (1.14–1.85)	150 (28.1%)	233 (23.3%)	0.039	1.29 (1.01–1.63)	256 (28.7%)	910 (22.7%)	0.00013	1.37 (1.17–1.62)
DR14	29 (8.1%)	253 (8.4%)	0.85	0.96 (0.64–1.44)	48 (9.0%)	73 (7.3%)	0.24	1.25 (0.86–1.83)	78 (8.7%)	326 (8.1%)	0.55	1.08 (0.83–1.40)
Diplotype												
DR8/DR8	3 (1.7%)	17 (1.1%)	0.46	1.49 (0.28–5.24)	3 (1.1%)	8 (1.6%)	0.76	0.70 (0.12–2.94)	6 (1.3%)	25 (1.2%)	0.86	1.08 (0.44–2.65)
*12:01/*09:01	5 (2.8%)	10 (0.66%)	0.0041	4.30 (1.45–12.74)	9 (3.3%)	3 (0.60%)	0.0051	5.76 (1.42–33.42)	14 (3.1%)	13 (0.6%)	5.0×10^{-6}	4.97 (2.32–10.66)

OR: odds ratio.
SE: shared epitope: HLA-DRB1*01:01, *01:02, *04:01, *04:04, *04:05, *04:08, *04:10, *04:13, *04:16, *10:01, *14:02, and *14:06.
doi:10.1371/journal.pone.0040067.t001

(p = 0.021). When we applied logistic regression analysis to the HLA-DRB1*09:01, *04:05, and HLA-DR14, their associations were revealed to be significant and do not depend on each other (p = 0.00067 and 0.00072, respectively, Table S5), except for that of DR14 (p = 0.30).

Comparison between ACPA-positive RF-positive RA and ACPA-positive RF-negative RA

Next, we analyzed whether these allele usage differences are also seen in ACPA-positive RA. We collected data about the HLA-DRB1 genotypes of 154 ACPA-positive RF-negative RA patients and 531 ACPA-positive RF-positive RA patients. As the SE and HLA-DRB1*09:01 were found to be associated with ACPA-positive RA, we analyzed the differences in the frequencies of these alleles [17]. In comparison with the healthy controls, SE and HLA-DRB1*09:01 were associated with a predisposition to ACPA-positive RF-positive RA as well as ACPA-positive RF-negative RA and displayed comparable odds ratios in logistic regression analysis (Table 4). No HLA-DRB1 alleles showed a strong specific association with a particular subset. When we directly compared the two subsets of ACPA-positive RA, no alleles displayed significant associations (Figure 1, Table S6). However, whether the two subsets of ACPA-positive RA share most of HLA-DRB1 susceptibility associations is inconclusive due to the small number of RF-negative subset.

Discussion

In this study, we demonstrated that classifying Japanese ACPA-negative RA patients based on their RF positivity successfully divided them into two genetically different subsets, which displayed different associations with HLA-DRB1. We showed that HLA-DRB1*09:01 and *04:05, strong susceptibility alleles to ACPA-positive RA, were also associated with ACPA-negative RF-positive subset, and that DR14 and the DR8 homozygote were associated only with the ACPA-negative RF-negative subset (Figure 1). Since the titer of RF fluctuates along with disease activity much more than that of ACPA, we were very careful to take the maximum RF titer when multiple titers were available for a particular patient, in order to prevent the RF positive subset from being contaminated with RF negative RA patients. The Recent UK population study reported the association of SE with ACPA-negative RF-positive RA [16]. Our study not only confirmed this association in Japanese RA, but also showed that the association of SE with ACPA-negative RF-positive RA is mainly due to the effect of HLA-DRB1*04:05 and that HLA-DRB1*09:01, HLA-DR14, and homozygote of HLA-DR8 are specifically associated with subsets of ACPA-negative RA.

These above-mentioned association tendencies were observed in the first set and successfully replicated in the second set, indicating that we can avoid population stratification or sampling bias. The effect sizes (odds ratio) of the alleles were comparable in each cohort (Tables 1 and 2) and the associations in the combined analysis reached significant level, although the p-values in each set did not reach the significance level due to the limited number of samples they contained. These data indicate that our results are reliable, at least in Japanese populations, although further replication studies including other populations are favorable. In the current study, we used logistic regression analysis to confirm independency of associated alleles in each comparison. When we used relative predispositional effects (RPE) method [18] to stratify associated alleles, we obtained the similar results to those we obtained by logistic regression analysis (data not shown).