

- adiponectin and adiponectin receptors in the peripheral tissues and CNS. *FEBS Lett* 2008;582:74-80.
28. Cook JR, Semple RK. Hypoadiponectinemia--cause or consequence of human "insulin resistance"? *J Clin Endocrinol Metab* 2010;95:1544-1554.
29. Schloegl H, Percik R, Horstmann A, Villringer A, Stumvoll M. Peptide hormones regulating appetite--focus on neuroimaging studies in humans. *Diabetes Metab Res Rev* 2011;27:104-112.
30. Stofkova A. Leptin and adiponectin: from energy and metabolic dysbalance to inflammation and autoimmunity. *Endocr Regul* 2009;43:157-168.
31. Minokoshi Y, Shiuchi T, Lee S, Suzuki A, Okamoto S. Role of hypothalamic AMP-kinase in food intake regulation. *Nutrition* 2008;24:786-790.
32. Takahashi M, Arita Y, Yamagata K *et al.* Genomic structure and mutations in adipose-specific gene, adiponectin. *Int J Obes Relat Metab Disord* 2000;24:861-868.
33. Siitonen N, Pulkkinen L, Lindstrom J *et al.* Association of ADIPOQ gene variants with body weight, type 2 diabetes and serum adiponectin concentrations: the Finnish Diabetes Prevention Study. *BMC Med Genet* 2011;12:5.
34. Ling H, Waterworth DM, Stirnadel HA *et al.* Genome-wide linkage and association analyses to identify genes influencing adiponectin levels: the GEMS Study. *Obesity (Silver Spring)* 2009;17:737-744.
35. Heid IM, Henneman P, Hicks A *et al.* Clear detection of ADIPOQ locus as the major gene for plasma adiponectin: results of genome-wide association analyses including 4659 European individuals. *Atherosclerosis* 2010;208:412-420.
36. Wu Y, Li Y, Lange EM *et al.* Genome-wide association study for adiponectin levels in Filipino women identifies CDH13 and a novel uncommon haplotype at KNG1-ADIPOQ. *Hum Mol Genet* 2010;19:4955-4964.

37. Hivert MF, Manning AK, McAteer JB *et al.* Common variants in the adiponectin gene (*ADIPOQ*) associated with plasma adiponectin levels, type 2 diabetes, and diabetes-related quantitative traits: the Framingham Offspring Study. *Diabetes* 2008;57:3353-3359.
38. Choquette AC, Lemieux S, Tremblay A *et al.* Evidence of a quantitative trait locus for energy and macronutrient intakes on chromosome 3q27.3: the Québec Family Study. *Am J Clin Nutr* 2008;88:1142-1148.
39. Ioannidis JP. Non-replication and inconsistency in the genome-wide association setting. *Hum Hered* 2007;64:203-213.

Figure legends

Figure 1 Genome-wide $-\log_{10} P$ value plot for stage 1 discovery analysis of confectionery-intake score. Blue line indicates the criteria for stage 2 replication phase ($P = 1 \times 10^{-5}$).

Figure 2 Q-Q plot for stage 1 discovery analysis of confectionery intake score. $\lambda = 1.003$.

Table 1 Background characteristics of participants

Characteristics	Stage 1	Stage 2	Stage 1 + 2	<i>P</i> (stage 1 vs. 2)
N	939	4,491	5,430	
Women (%)	48.9	53.0	52.3	0.022
Age (years)	47.9 ± 16.3	55.8 ± 8.9	54.4 ± 10.9	< 0.001
Current drinkers (%)	56.0	56.1	56.1	0.95
Ex-drinkers (%)	4.5	1.9	2.4	< 0.001
Current smokers (%)	28.6	17.4	19.3	< 0.001
Ex-smokers (%)	20.5	22.7	22.3	0.14
Body mass index (kg/m ²)	22.5 ± 3.1	23.2 ± 3.2	23.0 ± 3.2	< 0.001
Confectionery intake score	0.30 ± 0.32	0.28 ± 0.27	0.28 ± 0.28	0.085

Values are means ± standard deviation (SD).

Table 2 SNPs identified in GWAS analysis for confectionery intake score (n = 939 for stage 1 and 4,491 for stage 2)

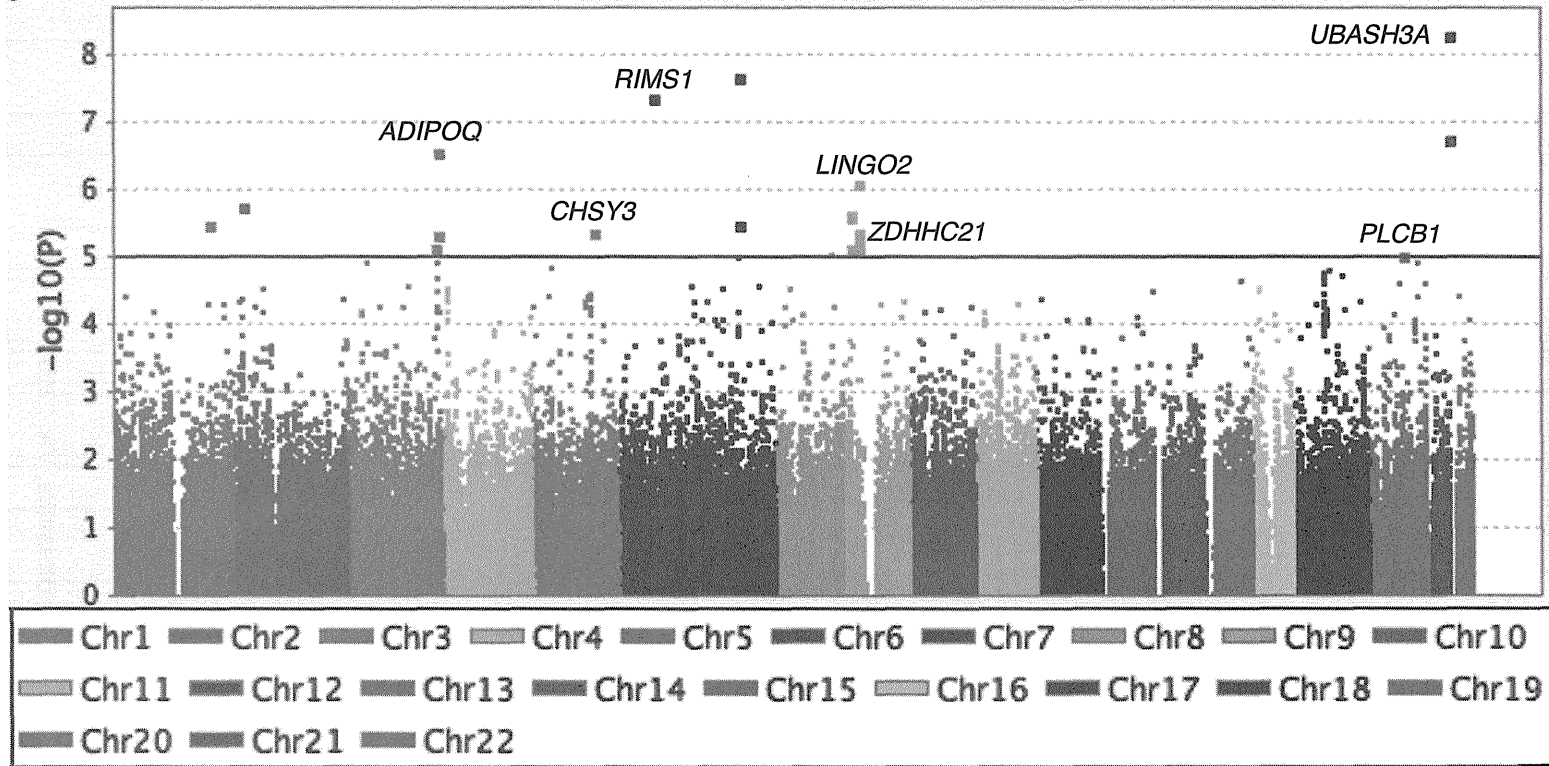
SNP ID	Chromo- some	Position (NCBI Build 36.3)	Major allele ^a	Minor allele ^a	Nearby gene	Minor allele frequency			β for additive model			P for additive model		
						Stage 1	Stage 2	Stage 1+2	Stage 1	Stage 2	Stage 1+2	Stage 1	Stage 2	Stage 1+2
rs17042603	2	21651923	A	G	No gene	0.052	0.050	0.050	0.1568	0.0155	0.0410	1.8×10 ⁻⁶	0.23	7.0×10 ⁻⁴
rs822396	3	188049571	A	G	<i>ADIPOQ</i>	0.055	0.060	0.060	0.1643	0.0234	0.0461	2.8×10 ⁻⁷	0.049	4.2×10 ⁻⁵
rs13356198	5	129284991	A	G	<i>CHSY3</i>	0.055	0.064	0.062	0.1432	0.0172	0.0362	4.7×10 ⁻⁶	0.14	9.1×10 ⁻⁴
rs1147522	6	72640321	C	T	<i>RIMS1</i>	0.017	0.017	0.017	0.3017	0.0043	0.0580	4.3×10 ⁻⁸	0.84	4.5×10 ⁻³
rs3897749	7	85972133	G	T	No gene	0.065	0.060	0.061	0.1379	-0.0104	0.0168	3.4×10 ⁻⁶	0.38	0.13
rs6474850	9	14652459	G	A	<i>ZDHHC21</i>	0.182	0.167	0.170	0.0911	-0.0044	0.0129	2.4×10 ⁻⁶	0.56	0.069
rs2890992	9	14662267	T	C	<i>ZDHHC21</i>	0.141	0.122	0.126	0.1001	-0.0006	0.0177	2.7×10 ⁻⁶	0.95	0.026
rs10810211	9	14666870	A	G	<i>ZDHHC21</i>	0.128	0.118	0.120	0.0976	-0.0010	0.0168	7.8×10 ⁻⁶	0.90	0.037
rs12351510	9	28641898	T	G	<i>LINGO2</i>	0.014	0.015	0.015	0.2834	-0.0039	0.0494	8.5×10 ⁻⁷	0.86	0.022
rs6039211	20	8564588	A	G	<i>PLCB1</i>	0.479	0.535	0.525	0.0652	-0.0035	0.0072	9.97×10 ⁻⁶	0.53	0.18
rs2839519	21	42740913	G	A	<i>UBASH3A</i>	0.177	0.170	0.171	0.0998	-0.0118	0.0077	1.9×10 ⁻⁷	0.11	0.27
rs2839525	21	42750006	T	G	<i>UBASH3A</i>	0.181	0.174	0.175	0.1103	-0.0084	0.0123	5.5×10 ⁻⁹	0.25	0.075

^a Alleles are indexed to the forward strand of Center for Biotechnology Information (NCBI) Build 36.3.

Table 3 Mean confectionery intake score by rs822396 genotype and background characteristics of participants in the pooled dataset (stage 1 and 2 studies)

Characteristics		Major homozygotes			Heterozygotes + minor homozygotes			<i>P</i>
		n	Mean (SE) ^a		n	Mean (SE) ^a		
All participants		4,800	0.278	(0.004)	628	0.326	(0.011)	4.6×10 ⁻⁵
Gender	Male	2,288	0.221	(0.005)	302	0.275	(0.014)	3.4×10 ⁻⁴
	Female	2,512	0.331	(0.006)	326	0.373	(0.017)	0.018
Age (years)	<55	2,085	0.271	(0.006)	254	0.354	(0.017)	5.1×10 ⁻⁶
	≥55	2,715	0.284	(0.005)	374	0.308	(0.014)	0.12
Smoking	Never smokers	2,784	0.318	(0.005)	380	0.362	(0.015)	4.7×10 ⁻³
	Current or ex-smokers	2,009	0.222	(0.006)	248	0.273	(0.017)	3.6×10 ⁻³
Alcohol drinking	Never drinkers	1,985	0.325	(0.007)	266	0.366	(0.018)	0.034
	Current or ex-drinkers	2,810	0.245	(0.005)	361	0.296	(0.014)	5.9×10 ⁻⁴
Body mass index (kg/m ²)	<25	3,638	0.278	(0.004)	494	0.330	(0.012)	5.8×10 ⁻⁵
	≥25	1,118	0.279	(0.009)	128	0.310	(0.026)	0.26

^a Adjusted for gender and age.



189

Accepted

Figure 1

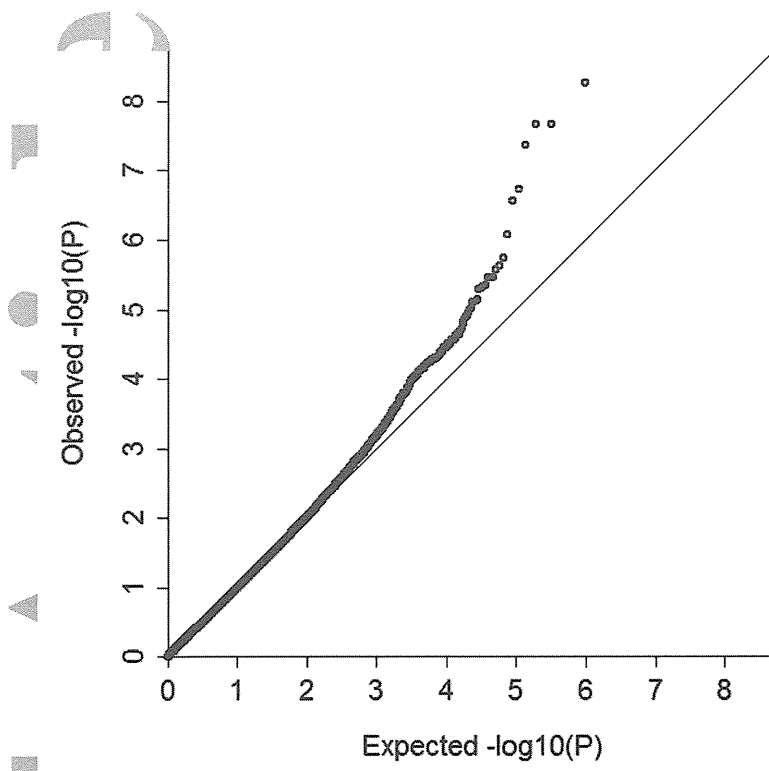


Figure 2

Accepted

Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population

Yukinori Okada^{1-3,40}, Chikashi Terao^{4,5,40}, Katsunori Ikari^{6,40}, Yuta Kochi^{1,2,40}, Koichiro Ohmura⁵, Akari Suzuki¹, Takahisa Kawaguchi⁴, Eli A Stahl^{7,8}, Fina A S Kurreeman⁷⁻⁹, Nao Nishida¹⁰, Hiroko Ohmiya³, Keiko Myouzen¹, Meiko Takahashi⁴, Tetsuji Sawada¹¹, Yuichi Nishioka¹², Masao Yukioka¹³, Tsukasa Matsubara¹⁴, Shigeyuki Wakitani¹⁵, Ryota Teshima¹⁶, Shigeto Tohma¹⁷, Kiyoshi Takasugi¹⁸, Kota Shimada¹⁷, Akira Murasawa¹⁹, Shigeru Honjo²⁰, Keitaro Matsuo²¹, Hideo Tanaka²¹, Kazuo Tajima²², Taku Suzuki^{6,23}, Takuji Iwamoto^{6,23}, Yoshiya Kawamura²⁴, Hisashi Tani²⁵, Yuji Okazaki²⁶, Tsukasa Sasaki²⁷, Peter K Gregersen²⁸, Leonid Padyukov²⁹, Jane Worthington³⁰, Katherine A Siminovitch³¹, Mark Lathrop^{32,33}, Atsuo Taniguchi⁶, Atsushi Takahashi³, Katsushi Tokunaga¹⁰, Michiaki Kubo³⁴, Yusuke Nakamura³⁵, Naoyuki Kamatani³⁶, Tsuneyo Mimori⁵, Robert M Plenge^{7,8}, Hisashi Yamanaka⁶, Shigeki Momohara^{6,41}, Ryo Yamada^{37,41}, Fumihiko Matsuda^{4,38,39,41} & Kazuhiko Yamamoto^{1,2,41}

Rheumatoid arthritis is a common autoimmune disease characterized by chronic inflammation. We report a meta-analysis of genome-wide association studies (GWAS) in a Japanese population including 4,074 individuals with rheumatoid arthritis (cases) and 16,891 controls, followed by a replication in 5,277 rheumatoid arthritis cases and 21,684 controls. Our study identified nine loci newly associated with rheumatoid arthritis at a threshold of $P < 5.0 \times 10^{-8}$, including *B3GNT2*, *ANXA3*, *CSF2*, *CD83*, *NFKBIE*, *ARID5B*, *PDE2A-ARAP1*, *PLD4* and *PTPN2*. *ANXA3* was also associated with susceptibility to systemic lupus erythematosus ($P = 0.0040$), and *B3GNT2* and *ARID5B* were associated with Graves' disease ($P = 3.5 \times 10^{-4}$ and 2.9×10^{-4} , respectively). We conducted a multi-ancestry comparative analysis with a previous meta-analysis in individuals of European descent (5,539 rheumatoid arthritis cases and 20,169 controls). This provided evidence of shared genetic risks of rheumatoid arthritis between the populations.

Rheumatoid arthritis is a complex autoimmune disease characterized by inflammation and the destruction of synovial joints and affects up to 1% of the population worldwide. To date, more than 35 rheumatoid arthritis susceptibility loci, including *HLA-DRB1*, *PTPN22*, *PADI4*, *STAT4*, *TNFAIP3* and *CCR6*, among others, have been identified by GWAS in multiple populations¹⁻¹² and by several meta-analyses of the original GWAS¹³⁻¹⁶. In particular, each meta-analysis of these GWAS uncovered a number of loci that were not identified in the single GWAS, leading to recognition of the enormous power of the meta-analysis approach for detecting causal genes in disease. However, these previous meta-analyses have been performed solely in European populations¹³⁻¹⁶ and not in

Asian ones. As multi-ancestry studies on validated rheumatoid arthritis susceptibility loci showed the existence of both population-specific and shared genetic components of rheumatoid arthritis^{10,17}, additional studies in Asian populations might provide useful insight into the underlying genetic architecture of rheumatoid arthritis, which would otherwise be difficult to capture using the studies in a single population. Here, we report a meta-analysis of GWAS and a replication study for rheumatoid arthritis in a Japanese population that was conducted by the Genetics and Allied research in Rheumatic diseases NETworking (GARNET) consortium^{10,12}. We subsequently performed a multi-ancestry comparative analysis that incorporated results from a previously conducted meta-analysis of individuals of European ancestry¹⁵.

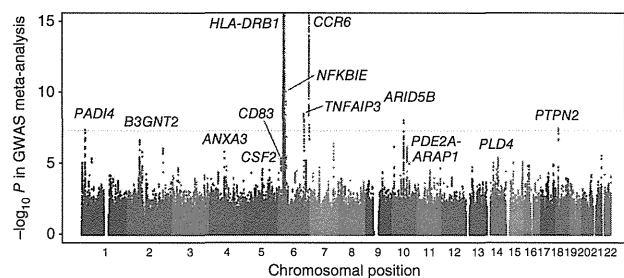


Figure 1 Manhattan plots of the GWAS meta-analysis for rheumatoid arthritis in the Japanese population. The genetic loci that satisfied the genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ (gray line) in the meta-analysis or in the combined study of the meta-analysis and the replication study are presented. The y axis shows the $-\log_{10} P$ values of the SNPs in the meta-analysis. The SNPs for which the P values were smaller than 1.0×10^{-15} are indicated at the upper limit of the plot.

A full list of author affiliations appears at the end of the paper.

Received 24 October 2011; accepted 1 March 2012; published online 25 March 2012; doi:10.1038/ng.2231

Table 1 Results of the GWAS meta-analysis and the replication studies for rheumatoid arthritis

rsID ^a	Chr.	Position (bp)	Cytoband	Gene(s)	Associations in Japanese										Associations in Europeans ^c							
					GWAS meta-analysis					Replication study					Combined study				GWAS meta-analysis			
					Allele 1/2	RA	Control	OR (95% CI) ^b	P	OR (95% CI) ^b	P	OR (95% CI) ^b	P	OR (95% CI) ^b	P	Allele 1 Freq.	RA	Control	OR (95% CI) ^b	P		
SNPs with significant associations ($P < 5.0 \times 10^{-8}$ in the combined study)																						
rs11900673	2	62306165	2p15	B3GN72	T/C	0.31	0.28	1.15 (1.08–1.21)	3.5×10^{-6}	1.09 (1.04–1.14)	6.0×10^{-4}	1.11 (1.07–1.15)	1.1×10^{-8}	0.13	0.13	1.05 (0.98–1.13)	0.17					
rs2867461	4	79732239	4q21	ANXA3	A/G	0.46	0.44	1.13 (1.08–1.19)	4.7×10^{-6}	1.12 (1.08–1.17)	1.2×10^{-7}	1.13 (1.09–1.17)	1.2×10^{-12}	0.37	0.37	0.98 (0.92–1.04)	0.52					
rs657075	5	131458017	5q31	CSF2	A/G	0.38	0.36	1.12 (1.06–1.18)	3.2×10^{-5}	1.11 (1.06–1.16)	3.8×10^{-6}	1.12 (1.08–1.15)	2.8×10^{-10}	0.10	0.10	1.04 (0.95–1.13)	0.37					
rs12529514	6	14204637	6p23	CD83	C/T	0.16	0.14	1.19 (1.10–1.27)	6.8×10^{-6}	1.11 (1.05–1.18)	6.0×10^{-4}	1.14 (1.09–1.19)	2.0×10^{-8}	0.055	0.053	1.11 (0.99–1.24)	0.074					
rs2233434	6	44340898	6p21.1	NFKBIE	G/A	0.24	0.21	1.23 (1.16–1.31)	9.2×10^{-11}	1.17 (1.11–1.23)	2.2×10^{-9}	1.19 (1.15–1.24)	5.8×10^{-19}	0.059	0.040	1.57 (1.11–2.21)	0.0099					
rs10821944	10	63455095	10q21	ARID5B	G/T	0.39	0.36	1.17 (1.11–1.23)	1.0×10^{-8}	1.15 (1.10–1.20)	3.0×10^{-10}	1.16 (1.12–1.20)	5.5×10^{-18}	0.29	0.26	1.11 (1.05–1.17)	1.9×10^{-4}					
rs3781913	11	72051144	11q13	PDE2A-ARAF1	T/G	0.71	0.69	1.11 (1.05–1.17)	3.2×10^{-4}	1.13 (1.08–1.18)	6.7×10^{-7}	1.12 (1.08–1.16)	5.8×10^{-10}	0.45	0.43	1.04 (0.99–1.09)	0.13					
rs2841277	14	104462050	14q32	PLD4	T/C	0.72	0.69	1.11 (1.05–1.18)	2.8×10^{-4}	1.18 (1.13–1.24)	7.0×10^{-12}	1.15 (1.11–1.19)	1.9×10^{-14}	0.47	0.46	1.02 (0.96–1.09)	0.54					
rs2847297	18	12787694	18p11	PTPN2	G/A	0.37	0.33	1.16 (1.11–1.23)	3.5×10^{-8}	1.06 (1.01–1.11)	0.013	1.10 (1.07–1.14)	2.2×10^{-8}	0.36	0.34	1.10 (1.05–1.15)	9.2×10^{-5}					
SNPs with suggestive associations ($5.0 \times 10^{-9} \leq P < 5.0 \times 10^{-6}$ in the combined study)																						
rs4937362	11	127997949	11q24	ETS1-FLI1	T/C	0.71	0.68	1.13 (1.07–1.19)	2.0×10^{-5}	1.07 (1.02–1.12)	0.0061	1.09 (1.06–1.13)	7.5×10^{-7}	0.46	0.44	1.06 (1.01–1.11)	0.015					
rs3783637	14	54417868	14q22	GCH1	C/T	0.76	0.74	1.13 (1.07–1.20)	6.5×10^{-5}	1.07 (1.02–1.13)	0.0062	1.10 (1.06–1.14)	2.0×10^{-6}	0.88	0.88	0.99 (0.88–1.11)	0.87					
rs1957895	14	60978085	14q23	PRKCH	G/T	0.40	0.39	1.12 (1.06–1.18)	4.1×10^{-5}	1.07 (1.02–1.12)	0.0022	1.09 (1.05–1.13)	3.6×10^{-7}	0.093	0.089	1.01 (0.95–1.07)	0.73					
rs6496667	15	88694672	15q26	ZNF774	A/C	0.38	0.35	1.13 (1.07–1.19)	4.7×10^{-5}	1.07 (1.02–1.11)	0.0050	1.09 (1.05–1.13)	1.4×10^{-6}	0.21	0.20	1.07 (1.01–1.13)	0.031					
rs7404928	16	23796341	16p12	PRKCB1	T/C	0.65	0.62	1.13 (1.07–1.19)	1.5×10^{-5}	1.05 (1.01–1.10)	0.026	1.08 (1.05–1.12)	4.0×10^{-6}	0.75	0.75	1.01 (0.94–1.09)	0.79					
rs2280381	16	84576134	16q24	IRF8	T/C	0.86	0.84	1.16 (1.08–1.25)	1.0×10^{-4}	1.09 (1.03–1.15)	0.0049	1.12 (1.07–1.17)	2.4×10^{-6}	0.62	0.60	1.05 (0.99–1.11)	0.081					
SNPs in previously reported rheumatoid arthritis susceptibility loci ($P < 5.0 \times 10^{-8}$ in the GWAS)																						
rs766449	1	17547439	1p36	PADI4	T/C	0.44	0.40	1.17 (1.11–1.24)	4.6×10^{-8}	-	-	-	-	0.38	0.37	1.09 (1.03–1.05)	0.0022					
rs2157337	6	32609122	6p21.3	HLA-DRB1	C/T	0.59	0.44	1.99 (1.88–2.11)	2.6×10^{-118}	-	-	-	-	0.69	0.46	2.50 (2.39–2.62)	$< 1.0 \times 10^{-300}$					
rs6932056	6	138284130	6q23	TNFAIP3	C/T	0.092	0.073	1.35 (1.23–1.49)	3.2×10^{-9}	-	-	-	-	0.044	0.034	1.41 (1.24–1.60)	1.3×10^{-7}					
rs1571878	6	167460832	6q27	CCR6	C/T	0.54	0.48	1.31 (1.24–1.39)	3.2×10^{-19}	-	-	-	-	0.47	0.43	1.13 (1.08–1.19)	5.9×10^{-7}					

Chr., chromosome; Freq., frequency; RA, rheumatoid arthritis; OR, odds ratio; CI, confidence interval.

^aSNPs with $P < 5.0 \times 10^{-6}$ in the combined study of the GWAS meta-analysis and the replication study or SNPs with $P < 5.0 \times 10^{-8}$ in the GWAS meta-analysis are annotated according to forward strand and NCBI Build 36.3. Full results of the replication study are provided in Supplementary Table 3. ^bOdds ratio of allele 1. ^cAssociations in the previous meta-analysis in European populations.⁵

The meta-analysis included 4,074 rheumatoid arthritis cases (with 81.4% and 80.4% of the subjects being positive for antibody to cyclic citrullinated peptide (anti-CCP) and rheumatoid factor, respectively) and 16,891 controls from three GWAS of Japanese subjects (from the BioBank Japan Project^{10,18}, Kyoto University¹² and the Institute of Rheumatology Rheumatoid Arthritis (IORRA)¹⁹; Supplementary Table 1). After the application of stringent quality control criteria, including principal-component analysis (PCA; Supplementary Fig. 1) for each GWAS, the meta-analysis was conducted by evaluating ~2.0 million autosomal SNPs with minor allele frequencies (MAFs) ≥ 0.01 , which were obtained through whole-genome imputation of genotypes on the basis of the HapMap Phase 2 East Asian panels (Japanese in Tokyo (JPT) and Han Chinese in Beijing (CHB)). The inflation factor of the test statistics in the meta-analysis λ_{GC} was as low as 1.036, suggesting no substantial effects of population structure (Supplementary Table 2). The quantile-quantile plot of P values showed a marked discrepancy in the values in its tail from those anticipated under the null hypothesis that there is no association—even after removal of the SNPs located in the human leukocyte antigen (HLA) region, the major rheumatoid arthritis susceptibility locus—thereby showing the presence of significant associations in the meta-analysis (Supplementary Fig. 2).

We identified seven loci in the current meta-analysis that satisfied the genome-wide significance threshold of $P < 5.0 \times 10^{-8}$. These included previously known rheumatoid arthritis susceptibility loci, such as *PADI4* at 1p36, *HLA-DRB1* at 6p21.3, *TNFAIP3* at 6q23 and *CCR6* at 6q27 (refs. 1,3,6,10,15) (the smallest $P = 2.6 \times 10^{-118}$ was found at the *HLA-DRB1* locus; Fig. 1 and Table 1). To our knowledge, the other three loci identified, *NFKBIE* at 6p21.1, *ARID5B* at 10q21 and *PTPN2* at 18p11, are newly associated ($P = 9.2 \times 10^{-11}$, 1.0×10^{-8} and 3.5×10^{-8} , respectively).

To validate the associations identified in the meta-analysis, we conducted a replication study of two independent Japanese rheumatoid arthritis case-control cohorts (cohort 1: 3,830 rheumatoid arthritis cases and 17,920 controls, cohort 2: 1,447 rheumatoid arthritis cases and 3,764 controls; Supplementary Table 1). To increase the number of subjects and enhance statistical power, genotype data obtained from other GWAS projects conducted for non-autoimmune diseases in Japanese using Illumina platforms were used for the replication control panels. For each of the 46 loci that exhibited $P < 5.0 \times 10^{-4}$ in



Figure 2 Regional plots of the loci newly associated with rheumatoid arthritis at the genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ in the combined study of the meta-analysis and the replication study. (a–i) Regional plots are shown at *B3GNT2* (a), *ANXA3* (b), *CSF2* (c), *CD83* (d), *NFKBIE* (e), *ARID5B* (f), *PDE2A-ARAP1* (g), *PLD4* (h) and *PTPN2* (i). Diamonds represent the $-\log_{10} P$ values of the SNPs, and the red diamonds represent the $-\log_{10} P$ values of the SNPs in the meta-analysis. Red color for the smaller circles represents the r^2 value with the most significantly associated SNP (larger red circle). The purple circle represents the P value in the combined study. The blue line shows the recombination rates given by the HapMap Phase 2 east Asian populations (release 22). RefSeq genes at the loci are indicated below. Genes nearest to the marker SNPs at the loci are colored blue (**Supplementary Note**), and genes implicated in eQTL analysis are colored red (**Supplementary Table 4**). At 11q13, two genes (*PDE2A* and *ARAP1*) that are nearest to the SNP selected for the replication study and the most significant SNP in the meta-analysis are highlighted. The plots were drawn using SNP Annotation and Proxy Search (SNAP) version 2.2.

the meta-analysis and had not been reported as rheumatoid arthritis susceptibility loci^{1–16}, we selected a marker SNP for the replication study (Online Methods and **Supplementary Table 3**).

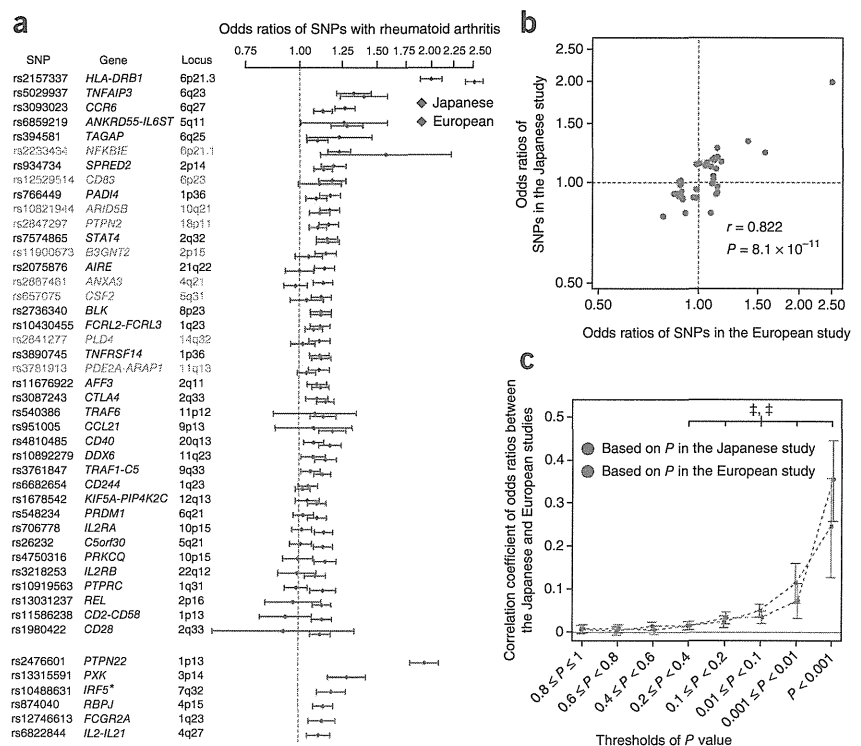
In the combined analyses of the meta-analysis and the replication study, including a total of 9,351 rheumatoid arthritis cases and 38,575 controls, we identified six newly associated loci, in addition to the *NFKBIE*, *ARID5B* and *PTPN2* loci, that satisfied the significance threshold of $P < 5.0 \times 10^{-8}$, including *B3GNT2* at 2p15, *ANXA3* at 4q21, *CSF2* at 5q31, *CD83* at 6p23, *PDE2A-ARAP1* at 11q13 and *PLD4* at 14q32 (**Figs. 1 and 2** and **Table 1**). Of these loci, *NFKBIE* had the smallest P value (5.8×10^{-19}). Although association with rheumatoid arthritis has been described for the *CSF2* and *PTPN2* loci^{11,15,16,20,21}, ours is the first report to our knowledge validating these associations with a threshold of $P < 5.0 \times 10^{-8}$. Suggestive associations were also observed in *ETS1-FLI1* at 11q24, *GCH1* at 14q22, *PRKCH* at 14q23, *ZNF774* at 15q26, *PRKCB1* at 16p12 and *IRF8* at 16q24 ($5.0 \times 10^{-8} \leq P < 5.0 \times 10^{-6}$). A summary of the genes in the newly associated loci and the results of *cis* expression quantitative trait locus (*cis* eQTL) analysis of the marker SNPs are provided (**Supplementary Table 4** and **Supplementary Note**).

Previous studies have reported associations of rheumatoid arthritis susceptibility loci with other autoimmune diseases^{4,10,15,16}. Therefore, we assessed the association of these newly identified susceptibility loci with systemic lupus erythematosus (SLE) by examining the results of an SLE GWAS in the Japanese population (891 cases and 3,384 controls)²² and in Graves' disease by genotyping 1,783 cases¹⁰ (the controls from the SLE analysis were used for testing for Graves'

disease). We observed significant associations of the *ANXA3* locus with SLE and of the *B3GNT2* and *ARID5B* loci with Graves' disease, which showed the same directional effects of the alleles as in rheumatoid arthritis ($P < 0.05/9 = 0.0056$, Bonferroni correction of the number of loci; **Supplementary Table 5**). It should be noted that relatively small sample sizes in the SLE and Graves' disease cohorts might yield limited statistical power, and further evaluations enrolling larger numbers of subjects would be desirable.

To highlight genetic backgrounds of rheumatoid arthritis that are common and divergent in different ancestry groups, we conducted a multi-ancestry comparative analysis of the present study in Japanese and a previous GWAS meta-analysis in Europeans that included 5,539 rheumatoid arthritis cases and 20,169 controls¹⁵ (**Fig. 3a–c**). First, we compared associations in the reported^{1–16} or newly identified rheumatoid arthritis susceptibility loci (**Fig. 3a** and **Supplementary Table 6**). Of the 46 rheumatoid arthritis risk variants evaluated, 6 were monomorphic in Japanese, and all were polymorphic in Europeans. We observed significant associations at 22 loci in Japanese and at 36 loci in Europeans (false discovery rate (FDR) < 0.05 , $P < 0.0030$), with 14 loci being shared between the populations. Of the newly associated rheumatoid arthritis susceptibility loci identified in our Japanese meta-analysis, significant associations were also observed in the European meta-analysis at the *ARID5B* and *PTPN2* loci ($P = 1.9 \times 10^{-4}$ and 9.2×10^{-5} , respectively; **Table 1**). Significant positive correlation of odds ratios was observed between the studies ($r = 0.822$, $P = 8.1 \times 10^{-11}$; **Fig. 3b**), suggesting that a substantial proportion of genetic factors are shared between

Figure 3 Overlap of the associations with rheumatoid arthritis between Japanese and European populations. (a) Forest plots of SNPs in the rheumatoid arthritis susceptibility loci (**Supplementary Table 6**). We selected the genetic loci that have been validated to be associated with rheumatoid arthritis susceptibility by showing associations in the reports of multiple cohorts or satisfying the genome-wide significant threshold ($P < 5.0 \times 10^{-8}$) in previous studies, including in the meta-analysis and replication phases^{1–16}. For each of the loci, the most significant SNP among those reported in the previous or present study were selected^{1–16}. SNPs in the newly identified rheumatoid arthritis susceptibility loci are colored green. Odds ratios and 95% confidence interval (CI) values are based on rheumatoid arthritis risk alleles, and the SNPs are ordered according to the odds ratios in the Japanese study. Several SNPs were monomorphic in the Japanese population. The odds ratios of these SNPs in the European study are presented below. The asterisk indicates that an association of another variant at the *IRF5* locus was reported in the Japanese population²⁴. (b) Correlation of the odds ratios of the SNPs in the validated rheumatoid arthritis susceptibility loci between the two populations. SNPs that were polymorphic in both populations were used; odds ratios were based on the minor allele in the Japanese population. (c) Correlation of the odds ratios of the genome-wide SNPs, excluding the rheumatoid arthritis susceptibility loci. Correlations were evaluated for sets of SNPs stratified by the thresholds based on the meta-analysis P values in each population after pruning of the SNPs by LD ($r^2 < 0.3$). Correlation coefficient and 95% CI are indicated on the y axis. Significant correlation of the odds ratios was observed (\ddagger , $P < 0.005$), even for the SNPs that showed moderate associations with rheumatoid arthritis (meta-analysis $P < 0.4$ in each population).



the two ancestry groups¹⁷. When the rheumatoid arthritis cases of the Japanese GWAS meta-analysis were stratified into anti-CCP-positive or rheumatoid factor-positive cases ($n = 3,209$) and controls ($n = 16,891$), similar results were observed (data not shown). Nevertheless, most of the SNPs assessed here are not necessarily causal variants, and further fine mapping of the loci is warranted to precisely evaluate the shared genetic predisposition between the populations.

Next, we compared regional associations within each of the loci and identified unique patterns in the *ARID5B* locus at 10q21 (**Supplementary Fig. 3**). In Japanese, three peaks of association were observed ($P = 1.0 \times 10^{-8}$ at rs10821944, $P = 5.7 \times 10^{-8}$ at rs10740069 and $P = 8.5 \times 10^{-6}$ at rs224311). These three variants were in weak linkage disequilibrium (LD) in Japanese ($r^2 < 0.10$), indicating independent associations with each of the other SNPs that satisfied a region-wide significance threshold of $P < 3.5 \times 10^{-5}$ (conditional $P = 4.3 \times 10^{-6}$, 1.7×10^{-5} and 1.8×10^{-5} , respectively) (**Supplementary Fig. 3**). In contrast, there was only one peak of association in Europeans ($P = 1.2 \times 10^{-6}$ at rs12764378; $r^2 = 0.59$ with rs10821944 in Europeans), and no additional association was observed in conditional analysis with rs12764378 (the smallest conditional $P = 2.2 \times 10^{-4}$), suggesting that the number of independent associations may be different at this locus in the two populations.

Finally, we conducted polygenic assessment for common variants showing modest associations to rheumatoid arthritis (those not meeting the genome-wide association threshold). This approach has been recognized to be a means to explain a substantial proportion of genetic risk²³. For the SNPs that were shared between the two meta-analyses but not included in the validated rheumatoid arthritis

susceptibility loci, we adopted LD pruning of the SNPs ($r^2 < 0.3$). We then evaluated the correlation of odds ratios of the SNPs between the two meta-analyses and observed a significant positive correlation ($r = 0.023$, $P < 1.0 \times 10^{-300}$). When the SNPs were stratified according to the P values in each meta-analysis, significant positive correlations of odds ratios were observed for the SNPs, even for those showing modest association ($P < 0.4$ in the meta-analysis of Japanese or Europeans; $r = 0.014$ – 0.36 for each P value range, $P < 0.005$ for each correlation test) (**Fig. 3c**). Correlations (r) of odds ratios observed herein suggest substantial overlap of the genetic risk of rheumatoid arthritis between the two populations, not only in the validated rheumatoid arthritis susceptibility loci but also at the loci showing nonsignificant associations. This suggests the usefulness of a meta-analysis approach involving multiple ancestry groups in identifying additional susceptibility loci.

In summary, we identified multiple new loci associated with rheumatoid arthritis through a large-scale meta-analysis of GWAS in Japanese. Multi-ancestry comparative analysis provided evidence of significant overlap in the genetic risks of rheumatoid arthritis between Japanese and Europeans. Thus, findings from the present study should contribute to the further understanding of the etiology of rheumatoid arthritis.

URLs. GARNET consortium, <http://www.twmu.ac.jp/IOR/garnet/home.html>; The BioBank Japan Project (in Japanese), <http://biobank.jp.org/>; International HapMap Project, <http://www.hapmap.org/>; PLINK, <http://pngu.mgh.harvard.edu/~purcell/plink/>; EIGENSTRAT, <http://genepath.med.harvard.edu/~reich/Software.htm>; MACH and mach2dat, <http://www.sph.umich.edu/csg/abecasis/MACH/index>.

html; R statistical software, <http://cran.r-project.org/>; SNAP, <http://www.broadinstitute.org/mpg/snap/index.php>; NCBI GEO database, <http://www.ncbi.nlm.nih.gov/geo/>.

METHODS

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

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

ACKNOWLEDGMENTS

The authors acknowledge the essential role of the GARNET consortium in developing the study. In this study, the following GARNET members are included: CGM of RIKEN, University of Tokyo, the BioBank Japan Project, Kyoto University and IORRA. We would like to thank all the doctors and staff who participated in sample collection for the RIKEN cohort and the BioBank Japan Project. We thank K. Kobayashi and M. Kitazato for their technical assistance. We thank T. Raj for calculation of composite of multiple signals (CMS). We thank M. Kokubo for DNA extraction, GWAS genotyping and secretarial assistance. We would also like to thank H. Yoshifuji, N. Yukawa, D. Kawabata, T. Nojima, T. Usui and T. Fujii for collecting DNA samples. We thank Y. Katagiri for her technical efforts. We also appreciate the contribution of E. Inoue and other members of the Institute of Rheumatology, Tokyo Women's Medical University, for their efforts on the IORRA cohort. This study was supported in part by grants-in-aid from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan, the Ministry of Health, Labour and Welfare (MHLW) in Japan, the Japan Society for the Promotion of Science (JSPS), Core Research for Evolutional Science and Technology (CREST), Solution-Oriented Research for Science and Technology (SORST), INSERM and the Okawa Foundation for Information and Telecommunications.

AUTHOR CONTRIBUTIONS

Y. Okada, C.T., K.I., Y. Kochi and K.O. designed the study and drafted the manuscript. Y. Okada, C.T., K.I., T.K., H.O., N.N., M.T., M.L., K. Tokunaga and M.K. managed genotyping and manipulation of GWAS data. Y. Okada, Y. Kochi, C.T. and K.I. managed genotyping of replication cohorts. Y. Okada, T.K., H.O., E.A.S., A. Takahashi and R.Y. performed statistical analysis. Y. Kochi, A.S., K. Myouzen, T. Sawada, Y. Nishoka, M.Y., T. Matsubara, S.W., R.T. and S.T. collected samples and managed phenotype data for the rheumatoid arthritis cohorts from the BioBank Japan Project and CGM, RIKEN. C.T., K.O., T.K., M.T., K. Takasugi, K.S., A.M., S.H., K. Matsuo, H. Tanaka, K. Tajima and M.L. collected samples and managed phenotype data for the rheumatoid arthritis cohorts from Kyoto University. K.I., T. Suzuki, T.I., Y. Kawamura, H. Tani, Y. Okazaki and T. Sakaki collected samples and managed phenotype data for the rheumatoid arthritis cohorts from IORRA. Y. Kochi managed the data for the SLE and Graves' disease cohorts. A.S., C.T. and K.I. analyzed the sera of subjects with rheumatoid arthritis. E.A.S., F.A.S.K., P.K.G., J.W., K.A.S., L.P. and R.M.P. managed the data for the rheumatoid arthritis cohorts in European populations. A. Taniguchi, A. Takahashi, K. Tokunaga, M.K., Y. Nakamura, N.K., T. Minori, R.M.P., H.Y., S.M., R.Y., F.M. and K.Y. supervised the overall study.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/naturegenetics/>.

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

- Suzuki, A. *et al.* Functional haplotypes of *PADI4*, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat. Genet.* **34**, 395–402 (2003).
- Kochi, Y. *et al.* A functional variant in *FCRL3*, encoding Fc receptor-like 3, is associated with rheumatoid arthritis and several autoimmunities. *Nat. Genet.* **37**, 478–485 (2005).
- The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
- Remmers, E.F. *et al.* *STAT4* and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N. Engl. J. Med.* **357**, 977–986 (2007).
- Plenge, R.M. *et al.* *TRAF1-C5* as a risk locus for rheumatoid arthritis—a genome-wide study. *N. Engl. J. Med.* **357**, 1199–1209 (2007).
- Plenge, R.M. *et al.* Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat. Genet.* **39**, 1477–1482 (2007).
- Barton, A. *et al.* Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. *Nat. Genet.* **40**, 1156–1159 (2008).
- Suzuki, A. *et al.* Functional SNPs in *CD244* increase the risk of rheumatoid arthritis in a Japanese population. *Nat. Genet.* **40**, 1224–1229 (2008).
- Gregersen, P.K. *et al.* *REL*, encoding a member of the NF- κ B family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat. Genet.* **41**, 820–823 (2009).
- Kochi, Y. *et al.* A regulatory variant in *CCR6* is associated with rheumatoid arthritis susceptibility. *Nat. Genet.* **42**, 515–519 (2010).
- Freudenberg, J. *et al.* Genome-wide association study of rheumatoid arthritis in Koreans: population-specific loci as well as overlap with European susceptibility loci. *Arthritis Rheum.* **63**, 884–893 (2011).
- Terao, C. *et al.* The human *AIRE* gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. *Hum. Mol. Genet.* **20**, 2680–2685 (2011).
- Raychaudhuri, S. *et al.* Common variants at *CD40* and other loci confer risk of rheumatoid arthritis. *Nat. Genet.* **40**, 1216–1223 (2008).
- Raychaudhuri, S. *et al.* Genetic variants at *CD28*, *PRDM1* and *CD2/CD58* are associated with rheumatoid arthritis risk. *Nat. Genet.* **41**, 1313–1318 (2009).
- Stahl, E.A. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.* **42**, 508–514 (2010).
- Zhernakova, A. *et al.* Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet.* **7**, e1002004 (2011).
- Kurreeman, F. *et al.* Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. *Am. J. Hum. Genet.* **88**, 57–69 (2011).
- Nakamura, Y. The BioBank Japan Project. *Clin. Adv. Hematol. Oncol.* **5**, 696–697 (2007).
- Yamanaka, H. *et al.* Influence of methotrexate dose on its efficacy and safety in rheumatoid arthritis patients: evidence based on the variety of prescribing approaches among practicing Japanese rheumatologists in a single institute-based large observational cohort (IORRA). *Mod. Rheumatol.* **17**, 98–105 (2007).
- Yamada, R. *et al.* Association between a single-nucleotide polymorphism in the promoter of the human interleukin-3 gene and rheumatoid arthritis in Japanese patients, and maximum-likelihood estimation of combinatorial effect that two genetic loci have on susceptibility to the disease. *Am. J. Hum. Genet.* **68**, 674–685 (2001).
- Tokuhiro, S. *et al.* An intronic SNP in a *RUNX1* binding site of *SLC22A4*, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nat. Genet.* **35**, 341–348 (2003).
- Okada, Y. *et al.* A genome-wide association study identified *AFF1* as a susceptibility locus for systemic lupus erythematosus in Japanese. *PLoS Genet.* **8**, e1002455 (2012).
- Stranger, B.E., Stahl, E.A. & Raj, T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* **187**, 367–383 (2011).
- Shimane, K. *et al.* A single nucleotide polymorphism in the *IRF5* promoter region is associated with susceptibility to rheumatoid arthritis in the Japanese patients. *Ann. Rheum. Dis.* **68**, 377–383 (2009).

¹Laboratory for Autoimmune Diseases, Center for Genomic Medicine (CGM), RIKEN, Yokohama, Japan. ²Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. ³Laboratory for Statistical Analysis, CGM, RIKEN, Yokohama, Japan. ⁴Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. ⁵Department of Rheumatology and Clinical Immunology, Graduate School of Medicine, Kyoto University, Kyoto, Japan. ⁶Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan. ⁷Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. ⁸Broad Institute, Cambridge, Massachusetts, USA. ⁹Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands. ¹⁰Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. ¹¹Department of Rheumatology, Tokyo Medical University Hospital, Tokyo, Japan. ¹²Yamanashi Prefectural Central Hospital, Yamanashi, Japan. ¹³Department of Orthopaedic Surgery, Yukioka Hospital, Osaka, Japan. ¹⁴Matsubara Mayflower Hospital, Hyogo, Japan. ¹⁵Osaka Minami National Hospital, Osaka, Japan. ¹⁶Department of Orthopedic Surgery, Tottori University, Tottori, Japan. ¹⁷Department of Rheumatology, National Hospital Organization, Sagami Hospital, Kanagawa, Japan. ¹⁸Center for Rheumatic Diseases, Dohgo Spa Hospital, Ehime, Japan. ¹⁹Department of Rheumatology, Niigata Rheumatic Center, Niigata, Japan. ²⁰Saiseikai Takaoka Hospital, Toyama, Japan. ²¹Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Aichi, Japan. ²²Aichi Cancer Center Hospital and Research Institute, Aichi, Japan. ²³Department of Orthopaedic Surgery, Keio University, Tokyo, Japan. ²⁴Yokohama Clinic, Warakukai Medical Corporation, Yokohama, Japan. ²⁵Department of Psychiatry, Mie University School of Medicine, Mie, Japan. ²⁶Metropolitan Matsuzawa Hospital, Tokyo, Japan. ²⁷Graduate School of Education, University of Tokyo, Tokyo, Japan. ²⁸The Feinstein Institute for Medical Research, North Shore–Long Island Jewish Health System, Manhasset, New York, USA. ²⁹Rheumatology Unit,

LETTERS

Department of Medicine in Solna, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden. ³⁰Arthritis Research Campaign–Epidemiology Unit, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK. ³¹Division of Medicine, University of Toronto, Mount Sinai Hospital and University Health Network, Toronto, Ontario, Canada. ³²Commissariat à l’Energie Atomique (CEA), Institut Genomique, Centre National de Genotypage, Evry, France. ³³Fondation Jean Dausset, Centre d’Etude du Polymorphisme Humain, Paris, France. ³⁴Laboratory for Genotyping Development, CGM, RIKEN, Yokohama, Japan. ³⁵Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan. ³⁶Laboratory for International Alliance, CGM, RIKEN, Yokohama, Japan. ³⁷Unit of Statistical Genetics, Center for Genomic Medicine Graduate School of Medicine Kyoto University, Kyoto, Japan. ³⁸Core Research for Evolutional Science and Technology (CREST) Program, Japan Science and Technology Agency, Kawaguchi, Japan. ³⁹Institut National de la Santé et de la Recherche Médicale (INSERM), Unité U852, Kyoto University Graduate School of Medicine, Kyoto, Japan. ⁴⁰These authors contributed equally to this work. ⁴¹These authors jointly directed this work. Correspondence should be addressed to Y.K. (ykochi@src.riken.jp) or K.O. (ohmurako@kuhp.kyoto-u.ac.jp).



ONLINE METHODS

Subjects. The Japanese participants in the meta-analysis (4,074 rheumatoid arthritis cases and 16,891 controls) and the replication study (5,277 rheumatoid arthritis cases and 21,684 controls) were obtained through the collaborations of the GARNET consortium (Supplementary Table 1)^{10,12}. The meta-analysis was conducted on three independent GWAS (from the BioBank Japan Project¹⁸ with 2,414 rheumatoid arthritis cases and 14,245 controls¹⁰, Kyoto University with 1,237 rheumatoid arthritis cases and 2,087 controls¹² and IORRA¹⁹ with 423 rheumatoid arthritis cases and 559 controls). The replication study consisted of two independent cohorts (cohort 1 included 3,830 rheumatoid arthritis cases and 17,920 controls, and cohort 2 included 1,447 rheumatoid arthritis cases and 3,764 controls). We employed a case-control cohort of SLE (891 cases and 3,384 controls)²² and 1,783 cases with Graves' disease¹⁰. Details of 5,539 rheumatoid arthritis cases and 20,169 controls included in the meta-analysis in European populations were described elsewhere¹⁵. All participants provided written informed consent for participation in the study, as approved by the ethical committees of the institutional review boards. Detailed descriptions of the participating subjects are provided (Supplementary Note).

Genotyping and quality control in the GWAS. Genotyping platforms and quality control criteria for the GWAS, including cutoff values for sample call rates, SNP call rates, MAF and Hardy-Weinberg P values, are given (Supplementary Table 2). For the subjects enrolled in each of three GWAS, we excluded closely related subjects with first- or second-degree kinship, which was estimated using PLINK version 1.06 (see URLs). We also excluded the subjects determined to be ancestry outliers from East Asian populations using PCA performed by EIGENSTRAT version 2.0 (see URLs) along with HapMap Phase 2 panels (release 24; Supplementary Fig. 1). Genotype imputation was performed on the basis of the HapMap Phase 2 East Asian populations, using MACH version 1.0.16 (see URLs) in a two-step procedure as described elsewhere²⁵. We excluded imputed SNPs with MAF < 0.01 or R_{sq} < 0.5 from each of the GWAS. Associations of the SNPs with rheumatoid arthritis were assessed by logistic regression models assuming additive effects of the allele dosages of the SNPs using mach2dat software (see URLs).

Meta-analysis. We included 1,948,139 autosomal SNPs that satisfied quality control criteria in all three GWAS (Supplementary Table 2). SNP information was based on a forward strand of the NCBI build 36.3 reference sequence. The meta-analysis was performed using an inverse variance method assuming a fixed-effects model from the study-specific effect sizes (logarithm of odds ratio) and the standard errors of the coded alleles of the SNPs determined with the Java source code implemented by the authors²⁵. Genomic control corrections²⁶ were carried out on test statistics of the GWAS using the study-specific inflation factor (λ_{GC}) and was applied or reapplied to the results of our current meta-analysis (Supplementary Fig. 2).

Replication study. We selected a SNP for the replication study from each of the loci that exhibited $P < 5.0 \times 10^{-4}$ in the meta-analysis that had not previously been reported as rheumatoid arthritis susceptibility loci¹⁻¹⁶ (Supplementary Table 3). For control subjects, we used genotype data obtained from additional GWAS for non-autoimmune diseases or healthy controls, genotyped using Illumina HumanHap550 BeadChips or HumanHap610-Quad BeadChips, and

the cases for rheumatoid arthritis and Graves' disease were genotyped with the TaqMan genotyping system (Applied Biosystems; Supplementary Table 1). Selection of the SNP was conducted according to the following criteria: if the SNP with the most significant association in the locus was genotyped in the replication control panel, then that SNP was selected; otherwise, a tag SNP in the replication control panel with the strongest LD was selected (mean $r^2 = 0.89$). For the three SNPs that yielded low call rates (<90%), we alternatively selected proxy SNPs with the second strongest LD. As a result, average genotyping call rates of the SNPs were 99.9% and 99.0% for the controls and cases, respectively. We then evaluated concordance rates between the assayed genotypes by applying these two different methods to samples from 376 subjects who were randomly selected. This procedure yielded high concordance rates of $\geq 99.9\%$. Associations of the SNPs were evaluated using logistic regression assuming an additive-effects model of genotypes in R statistical software version 2.11.0 (see URLs). The combined study of the meta-analysis and replication study was performed using an inverse variance method assuming a fixed-effects model²⁵.

Cis eQTL analysis. For each marker SNP of the newly identified rheumatoid arthritis susceptibility locus, correlations between SNP genotypes and expression levels of genes located 300 kb upstream or downstream of the SNP measured in B-lymphoblastoid cell lines (GSE6536) were evaluated using data from the HapMap Phase 2 east Asian populations²⁷.

Multi-ancestry analysis of the meta-analyses in Japanese and Europeans. We evaluated the associations of the variants in the validated rheumatoid arthritis susceptibility loci by comparing the results from the current meta-analysis in Japanese with those from a previous meta-analysis in Europeans¹⁵. We assessed two variants in the *IRF5* locus, where different causal variants were identified in the two populations²⁴. For the conditional analysis of the regional associations in the *ARID5B* locus (Supplementary Fig. 3), we repeated the meta-analysis at that locus by incorporating genotypes of the referenced SNP(s) as additional covariate(s). For comparison of the odds ratios of the SNPs, we first selected SNPs that were shared between the meta-analyses in Japanese and Europeans. Next, we removed the SNPs located more than 1 Mb away from each of the marker SNPs in the validated rheumatoid arthritis susceptibility loci, except for in the HLA region, where we removed the SNPs located between 24,000,000 bp to 36,000,000 bp on chromosome 6 because of the existence of long-range haplotypes with rheumatoid arthritis susceptibility in this region²⁸. LD pruning of the SNPs was conducted for the SNP pairs that were in LD ($r^2 \geq 0.3$) in both HapMap Phase 2 East Asian and Utah residents of Northern and Western European ancestry (CEU) populations (release 24). Correlations of the odds ratios were evaluated using R statistical software version 2.11.0.

25. Okada, Y. *et al.* Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. *PLoS Genet.* **7**, e1002067 (2011).
26. de Bakker, P.I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17**, R122-R128 (2008).
27. Stranger, B.E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39**, 1217-1224 (2007).
28. Okada, Y. *et al.* Contribution of a haplotype in the HLA region to anti-cyclic citrullinated peptide antibody positivity in rheumatoid arthritis, independently of HLA-DRB1. *Arthritis Rheum.* **60**, 3582-3590 (2009).



Genetic Polymorphisms of the Human PNPLA3 Gene Are Strongly Associated with Severity of Non-Alcoholic Fatty Liver Disease in Japanese

Takahisa Kawaguchi^{1,2}, Yoshio Sumida³, Atsushi Umemura⁴, Keitaro Matsuo⁵, Meiko Takahashi¹, Toshinari Takamura⁶, Kohichiroh Yasui⁷, Toshiji Saibara⁸, Etsuko Hashimoto⁹, Miwa Kawanaka¹⁰, Sumio Watanabe¹¹, Sumio Kawata¹², Yasuharu Imai¹³, Miki Kokubo¹, Toshihide Shima⁴, Hyohun Park⁴, Hideo Tanaka⁵, Kazuo Tajima⁵, Ryo Yamada¹, Fumihiko Matsuda^{1,2*}, Takeshi Okanoue⁴ for the Japan Study Group of Nonalcoholic Fatty Liver Disease (JSG-NAFLD)

1 Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, **2** Institut National de la Sante et de la Recherche Medicale (INSERM) Unite U852, Kyoto University Graduate School of Medicine, Kyoto, Japan, **3** Center for Digestive and Liver Diseases, Nara City Hospital, Nara, Japan, **4** Center of Gastroenterology and Hepatology, Saiseikai Suita Hospital, Suita, Japan, **5** Division of Epidemiology and Prevention, Aichi Cancer Center, Nagoya, Japan, **6** Department of Disease Control and Homeostasis, Kanazawa University, Graduate School of Medical Science, Kanazawa, Japan, **7** Department of Molecular Gastroenterology and Hepatology, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan, **8** Department of Gastroenterology and Hepatology, Kochi Medical School, Kochi, Japan, **9** Department of Internal Medicine and Gastroenterology, Tokyo Women's Medical University, Tokyo, Japan, **10** Center of Liver Diseases, Kawasaki Hospital, Kawasaki Medical School, Okayama, Japan, **11** Department of Gastroenterology, Juntendo University School of Medicine, Tokyo, Japan, **12** Department of Gastroenterology, Yamagata University School of Medicine, Yamagata, Japan, **13** Department of Internal Medicine, Ikeda Municipal Hospital, Ikeda, Japan

Abstract

Background: Nonalcoholic fatty liver disease (NAFLD) includes a broad range of liver pathologies from simple steatosis to cirrhosis and fibrosis, in which a subtype accompanying hepatocyte degeneration and fibrosis is classified as nonalcoholic steatohepatitis (NASH). NASH accounts for approximately 10–30% of NAFLD and causes a higher frequency of liver-related death, and its progression of NASH has been considered to be complex involving multiple genetic factors interacting with the environment and lifestyle.

Principal Findings: To identify genetic factors related to NAFLD in the Japanese, we performed a genome-wide association study recruiting 529 histologically diagnosed NAFLD patients and 932 population controls. A significant association was observed for a cluster of SNPs in *PNPLA3* on chromosome 22q13 with the strongest *p*-value of 1.4×10^{-10} (OR = 1.66, 95%CI: 1.43–1.94) for rs738409. Rs738409 also showed the strongest association ($p = 3.6 \times 10^{-6}$) with the histological classifications proposed by Matteoni and colleagues based on the degree of inflammation, ballooning degeneration, fibrosis and Mallory-Denk body. In addition, there were marked differences in rs738409 genotype distributions between type4 subgroup corresponding to NASH and the other three subgroups ($p = 4.8 \times 10^{-6}$, OR = 1.96, 95%CI: 1.47–2.62). Moreover, a subgroup analysis of NAFLD patients against controls showed a significant association of rs738409 with type4 ($p = 1.7 \times 10^{-16}$, OR = 2.18, 95%CI: 1.81–2.63) whereas no association was obtained for type1 to type3 ($p = 0.41$). Rs738409 also showed strong associations with three clinical traits related to the prognosis of NAFLD, namely, levels of hyaluronic acid ($p = 4.6 \times 10^{-4}$), HbA1c ($p = 0.0011$) and iron deposition in the liver ($p = 5.6 \times 10^{-4}$).

Conclusions: With these results we clearly demonstrated that Matteoni type4 NAFLD is both a genetically and clinically different subset from the other spectrums of the disease and that the *PNPLA3* gene is strongly associated with the progression of NASH in Japanese population.

Citation: Kawaguchi T, Sumida Y, Umemura A, Matsuo K, Takahashi M, et al. (2012) Genetic Polymorphisms of the Human PNPLA3 Gene Are Strongly Associated with Severity of Non-Alcoholic Fatty Liver Disease in Japanese. PLoS ONE 7(6): e38322. doi:10.1371/journal.pone.0038322

Editor: Takeshi Okanoue, Wageningen University, The Netherlands

Received: March 8, 2012; **Accepted:** May 3, 2012; **Published:** June 14, 2012

Copyright: © 2012 Kawaguchi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the grant from Ministry of Labor and Welfare Japan [T.O., H20-Hepatitis-general-008], Core Research of Evolutional Science & Technology (CREST). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: fumi@genome.med.kyoto-u.ac.jp

Introduction

Nonalcoholic fatty liver disease (NAFLD) includes a broad range of pathologies from fatty liver (simple steatosis), steatonecrosis, and steatohepatitis to cirrhosis [1–3]. NAFLD often accompanies other lifestyle-related pathologies of metabolic

syndrome such as diabetes mellitus, hypertension and dyslipidemia, and the number of NAFLD patients is increasing worldwide along with the escalation in the incidence of metabolic syndrome [4]. Prevalence of NAFLD is considered as approximately 8% in Japanese and 6–35% in Europeans [4,5]. The majority of NAFLD

shows simple steatosis with a good prognosis, but approximately 10–30% of NAFLD histologically diagnosed as nonalcoholic steatohepatitis (NASH) shows hepatocyte degeneration (ballooning hepatocyte), necrosis, inflammation and fibrosis, with a higher frequency of liver-related death both in Japanese and European populations [6,7]. Insulin resistance and oxidative stress are considered to be key players in the progression of NASH [8,9]. However, the progression of NASH has been considered to be complex involving multiple genetic factors interacting with the environment and lifestyle, because only a portion of NAFLD patients develops NASH.

The first Genome-wide association (GWA) study searching for such genetic factors identified the *PNPLA3* gene as a major genetic determinant for the predisposition to NAFLD in Hispanic, African American and European American populations according to liver fat contents [10], which was subsequently confirmed in Europeans and Asians according to liver biopsy. Association of *PNPLA3* with not only fatty liver and TG content, but also inflammation and fibrosis were shown in the subsequent studies, so *PNPLA3* may be widely associated with the development of NAFLD [11–13]. More recently, another GWA study reported the association of four additional genes with NAFLD in Europeans [14]. Also, a candidate gene-based approach revealed the association between NAFLD and the apolipoprotein C3 gene in Indians [15]. However, the precise role of such genes in the development of NASH still remains to be elucidated. In addition, no GWA study has been reported for Asian populations to date although the genetic components and their relative contribution may be different between ethnicities.

The Japan NASH Study Group was founded in 2008 aiming at the identification of genetic determinants predisposing to NASH in the Japanese population. Here we report the first GWA study of NAFLD in the Japanese using DNA samples of patients with liver histology-based diagnoses recruited through this multi-institutional research network.

Results

Genome-wide Association Analysis of NAFLD in Japanese

We conducted a GWA study using DNA samples of 543 patients with NAFLD and 942 controls. After quality controls of genotyping results (see materials and methods for details), a total of 529 patients consisting of four NAFLD subgroups according to Matteoni's classification [2] (type1; 100, type2; 73, type3; 29, type4; 327) and 932 controls were subjected to statistical analyses (Table 1). This index pathologically classifies NAFLD according to the degree of inflammation, hepatocyte degeneration, and the existence of fibrosis and Mallory-Denk body in the liver. Genome scan results of 932 DNA samples collected for other genetic studies were used as general Japanese population controls [16]. After standard quality control procedure as described in materials and methods, genotype distributions of 484,751 autosomal SNP markers were compared between the NAFLD cases and control subjects by exact trend test. A slight inflation of p -values was observed by genomic control method ($\lambda = 1.04$) (Figure S1).

We identified six SNP markers located at chromosome 22q13 showing genome-wide significance ($p < 1.04 \times 10^{-7}$) (Figure 1). Among them, four SNPs, namely, rs2896019, rs926633, rs2076211 and rs1010023, located in the *PNPLA3* gene and in strong linkage disequilibrium (LD) ($r^2 > 0.93$), returned p -values smaller than 1×10^{-9} ($p = 1.5 \times 10^{-10}$, 7.5×10^{-10} , 1.4×10^{-9} and 1.5×10^{-9} , respectively) (Table 2). Rs738407 and rs3810662 also located in *PNPLA3* showed significant but weaker associations

($p = 1.0 \times 10^{-7}$ and 1.0×10^{-7} , respectively) than the above four SNP markers. Rs738491, rs2073082, rs3761472, rs2235776, rs2143571 and rs6006473 were in the neighboring *SAMM50* gene which is outside of the linkage disequilibrium (LD) block where the top SNP markers were distributed (Figure 2). These markers were in moderate LD with each other ($r^2 > 0.42$) and showed p -values between 3.9×10^{-6} and 6.4×10^{-7} but did not reach genome-wide significance (Table S1). Rs738409, the SNP which showed the strongest association with NAFLD in the first GWA study [10], was not included in the SNP array used in our study. This SNP was therefore genotyped using Taqman technology in the same case and control samples that were used for genome scan. Rs738409 showed the strongest association with the disease ($p = 1.4 \times 10^{-10}$, OR = 1.66, 95%CI: 1.43–1.94) among all the SNP markers examined in this study. The association remained after the correction for population stratification with EIGENSTRAT [17] ($p = 2.3 \times 10^{-11}$). Although a peak consisting of a cluster of SNPs was observed at the *HLA* locus on chromosome 6 (minimal p -value of 4.10×10^{-7} for rs9262639 located at the 3' of *C6orf15* gene), the association disappeared when EIGENSTRAT was applied ($p > 1.6 \times 10^{-3}$). We consider this as a result of population stratification between the cases and controls.

Impact of *PNPLA3* Polymorphisms to the Pathogenicity of NAFLD

We next examined whether or not the seven SNPs in the *PNPLA3* gene were associated with the pathogenic status of NAFLD. The genotype distributions of these SNPs were compared by Jonckheere-Terpstra test among the four subgroups of NAFLD patients categorized by Matteoni's classification (type1 to type4). There was a significant increase in the frequency of the risk allele from Matteoni type1 to type4 for all of the seven SNPs (p -values ranging from 3.6×10^{-6} to 0.0017) (Table 2). Among them, rs738409 again showed the strongest association ($p = 3.6 \times 10^{-6}$) as seen in the simple case/control analysis. On the other hand, there was no significant association between control and Matteoni type1 ($p = 0.76$).

In order to clarify how rs738409 influences the pathogenicity of NAFLD, we performed pairwise comparisons of genotype distributions in the four subgroups of NAFLD patients. There were marked differences in genotype distributions between type4 subgroup and the other three subgroups by multivariable logistic regression adjusted for age, sex and body mass index (BMI) ($p = 2.0 \times 10^{-5}$, OR = 2.18, 95%CI: 1.52–3.18 between type1 and type4; $p = 1.4 \times 10^{-3}$, OR = 1.81, 95%CI: 1.26–2.62 between type2 and type4; $p = 0.027$, OR = 1.85, 95%CI: 1.07–3.19 between type3 and type4) (Figure 3). On the other hand, no significant associations were obtained for type1 to type3 in any combinations. When we performed the same analysis between type4 and the pooled genotypes of type1 to type3, we again obtained a significant difference ($p = 4.8 \times 10^{-6}$, OR = 1.96, 95%CI: 1.47–2.62).

We further examined the specific association of rs738409 with type4 subgroup by using the case/control association results of the initial genome scan. 529 NAFLD patients were divided into 202 patients with type1 to type3 and 327 patients with type4, and genotype distributions of rs738409 in each subgroup were compared with those of 932 control subjects. Exact trend test returned an extremely strong association of rs738409 with type4 subgroup ($p = 1.7 \times 10^{-16}$, OR = 2.18, 95%CI: 1.81–2.63) whereas no association was obtained for type1 to type3 subgroups ($p = 0.41$).

Table 1. Clinical characteristics according to the histological classification.

Phenotype	Matteoni classification of NAFLD				Control	p-value
	Type 1	Type 2	Type 3	Type 4		
Number of samples	100	73	29	327	932	
Sex (Male/Female)	59/41	47/26	13/16	130/197	471/461	0.0023‡
Age (year)	49.7±15.3	51.5±15.3	49.4±14.0	57.6±14.8	48.8±16.3	<0.001
Physical measurement						
BMI	26.2±4.3	27.7±4.8	27.6±3.5	27.7±5.2	–	0.054
Amount of visceral fat (cm ²)	146.8±65.3	154.3±47.7	136.8±53.8	151.7±57.4	–	0.46
Abdominal circumscript (cm)	90.9±9.9	94.1±10.0	88.5±10.2	94.1±11.8	–	0.10
Biochemical trait						
AST (IU/L)	31.1±14.6	36.4±18.5	52.4±35.1	57.7±48.4	–	<0.001
ALT (IU/L)	48.6±30.8	62.8±47.6	81.5±46.9	74.9±48.4	–	<0.001
GGT (IU/L)	71.0±62.5	67.1±66.9	96.1±91.3	76.6±73.9	–	0.25
Albumin (g/dL)	4.5±0.4	4.4±0.3	4.5±0.3	4.3±0.4	–	<0.001
Total bilirubin (mg/dL)	0.9±0.5	0.9±0.5	0.9±0.6	0.8±0.4	–	0.063
Cholinesterase (unit)	389.1±97.0	354.3±97.2	371.1±109.9	348.9±93.2	–	<0.001
Type IV collagen 7S (ng/dL)	3.8±0.7	3.9±0.9	3.9±0.8	5.1±1.7	–	<0.001
Hyaluronic acid (ng/dL)	25.6±22.5	33.6±29.5	31.5±24.0	80.9±84.3	–	<0.001
Triglycerides (mg/dL)	151.9±73.8	154.0±92.1	166.1±86.5	161.2±85.7	–	0.23
Total cholesterol (mg/dL)	209.1±32.8	194.0±38.0	203.0±39.9	200.3±39.0	–	0.093
HbA1c (%)	6.1±1.1	5.9±1.2	6.5±1.8	6.2±1.3	–	0.13
IRI (μg/dL)	9.1±5.4	11.4±9.0	10.4±6.3	14.9±9.9	–	<0.001
FPG (mg/dL)	112.9±33.7	107.3±27.4	109.9±27.7	114.8±33.8	–	0.14
HOMA-IR	2.4±1.5	2.9±2.4	3.0±2.1	4.2±3.0	–	<0.001
hs-CRP (mg/dL)	1078.9±1407	1048.3±1185.0	865.8±658.4	1579.2±2377.9	–	0.027
Adiponectin (μg/mL)	7.4±4.4	8.5±6.6	6.6±2.6	6.9±4.3	–	0.24
Leptin (ng/mL)	9.9±7.4	9.1±6.2	11.3±9.4	12.4±7.9	–	<0.001
Ferritin (ng/mL)	145.8±101.1	176.5±134.0	271.2±307.0	208.3±180.3	–	0.027
Uric acid (mg/dL)	5.9±1.5	5.7±1.2	5.4±1.9	5.7±1.6	–	0.77
PLT (×10 ⁶ /μL)	23.0±5.9	22.9±4.9	21.9±6.7	20.2±6.4	–	<0.001
ANA (0/1/2/3/4)	42/17/4/0/0	31/8/4/1/2	15/6/2/0/0	147/76/31/8/12	–	0.015
Clinical history						
Diabetes (NGT/IGT/DM)	36/11/34	24/7/27	12/8/7	103/35/119	–	0.45*
Hyperlipidemia (+/–)	31/68	31/42	9/20	120/206	–	0.60‡
Hypertension (+/–)	64/35	33/40	19/10	155/172	–	0.013‡
Liver biopsy feature						
Brunt grade (1/2/3)	–	–	19/3/2	149/133/44	–	<0.001‡
Brunt stage (1/2/3/4)	–	–	–	123/74/105/24	–	–
Fat droplet (1/2/3/4)	38/32/19/11	14/29/18/7	7/3/10/4	51/99/104/52	–	<0.001
Iron deposition (0/1/2/3/4)	30/14/21/10/1	24/9/12/2/1	10/5/2/2/0	132/56/29/29/11	–	0.16

Measurements are shown as mean ± standard deviation. Categorical values are shown by the count number. P-values are calculated by Jonckheere-Terpstra test unless otherwise stated;

‡Chochran-Armitage trend test,

*Kruskal-Wallis test. Abbreviations used for each trait are summarized in materials and methods.

doi:10.1371/journal.pone.0038322.t001

Association of rs738409 Genotypes with Clinical Traits

The quantitative effects of rs738409 genotypes to clinical traits were examined by multivariable regression adjusted for age, sex and BMI (statistical calculation 1, Table 3). Five categorical ordinals, namely, anti-nuclear antibody (ANA), Brunt grade, Brunt stage, fat deposition and iron deposition, were also tested by an ordinal logistic regression analysis. Potential associations

($p < 0.05$) were obtained for 11 traits, namely, aspartate transaminase (AST), alanine aminotransferase (ALT), type IV collagen 7S, hyaluronic acid, hemoglobin A1c (HbA1c), fasting immunoreactive insulin (IRI), fasting plasma glucose (FPG), platelet count (PLT), Brunt grade, fat deposition and iron deposition (Table 3). When the results were further adjusted for Matteoni type (statistical calculation 2), AST, hyaluronic acid, HbA1c, FPG,

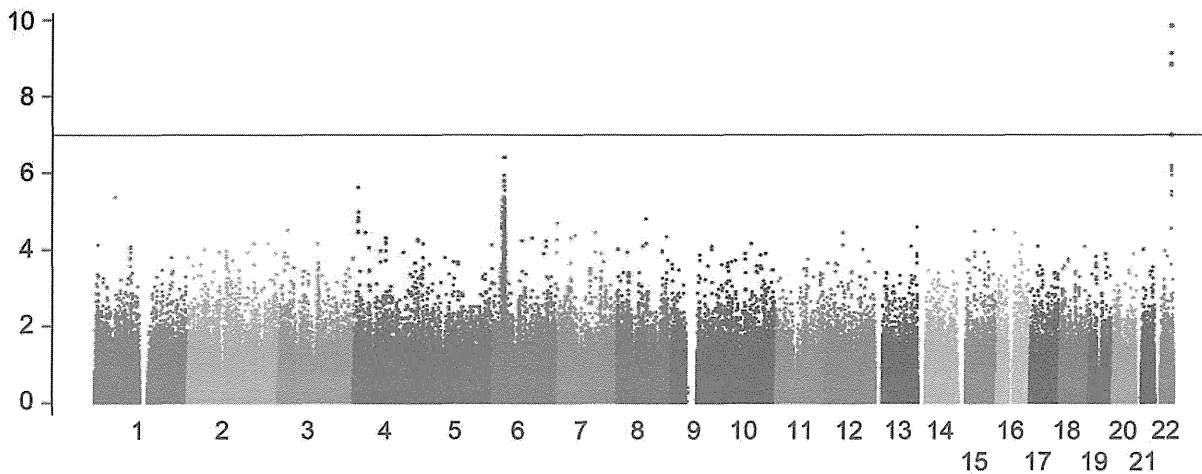


Figure 1. Manhattan plot of the GWA study. Association *p*-values are calculated by exact trend test and plotted along the chromosome in $-\log_{10}$ scale. The horizontal line indicates Bonferroni-adjusted significance threshold ($p = 1.03 \times 10^{-7}$).
doi:10.1371/journal.pone.0038322.g001

PLT, Brunt grade and iron deposition showed *p*-values smaller than 0.05. The level of serum triglyceride was not significant in the initial analysis, but became significant after being adjusted for Matteoni's type ($p = 0.013$). Among them, only three traits, namely, hyaluronic acid, HbA1c and iron deposition, remained significant ($p < 0.0021$) after Bonferroni's correction for multiple testing (Table 3).

Associations of Previously Reported SNPs with NAFLD

Previous genetic studies identified four chromosomal loci, namely, *LYPLAL1* at 1q41, *GCKR* at 2p23, *NCAN* at 19p12 and *PPPIR3B* at 8p23.1, associated with NAFLD in populations of

European descent [14]. We examined whether or not the associations were reproduced in the Japanese population by extracting genotype information of SNP markers corresponding to these four loci. As shown in Table 4, the association of rs780094 in *GCKR* with NAFLD was at the border of significance ($p = 0.011$, OR = 0.82, 95%CI: 0.70–0.91) in the case/control analysis. However, the association was lost when examined between rs780094 genotypes and Matteoni types. There were no associations of rs2228603 in *NCAN* and rs12137855 in *LYPLAL1* with either NAFLD or Matteoni types. Rs4240624 in *PPPIR3B* was not in the SNP array used for this study, and this marker was not polymorphic or at a very low frequency in the Japanese (0 in 90

Table 2. List of the SNP markers in the *PNPLA3* locus at chromosome 22q showing genome wide significance.

dbSNPID	Genotyping Result and Allele Frequency of A2							Statistics		
	A1/A2	NAFLD					NAFLD vs. Control		Matteoni	
		Control	Total	Type 1	Type 2	Type 3	Type 4	<i>p</i> -value†		OR (95%CI)
rs738407	T/C	124/447/361	46/200/283	12/51/37	10/28/35	4/14/11	20/107/200	1.0×10^{-7}	1.56(1.32–1.83)	3.4×10^{-5}
		(0.627)	(0.724)	(0.625)	(0.671)	(0.621)	(0.775)			
rs738409	C/G*	247/468/217	88/236/203	20/59/21	21/30/22	8/11/9	39/136/151	1.4×10^{-10}	1.66(1.43–1.94)	3.6×10^{-6}
		(0.484)	(0.609)	(0.505)	(0.507)	(0.518)	(0.672)			
rs2076211	C/T*	248/473/211	92/242/195	21/58/21	21/30/22	8/11/10	42/143/142	1.4×10^{-9}	1.61(1.38–1.87)	3.2×10^{-5}
		(0.480)	(0.597)	(0.500)	(0.507)	(0.534)	(0.653)			
rs2896019	T/G*	246/473/213	91/234/204	20/57/23	22/29/22	7/12/10	42/136/149	1.5×10^{-10}	1.66(1.42–1.93)	2.6×10^{-5}
		(0.482)	(0.607)	(0.515)	(0.500)	(0.552)	(0.664)			
rs1010023	T/C*	249/473/210	94/239/196	21/57/22	22/29/22	7/12/10	44/141/142	1.5×10^{-9}	1.61(1.38–1.87)	6.5×10^{-5}
		(0.479)	(0.596)	(0.505)	(0.500)	(0.552)	(0.650)			
rs926633	G/A*	247/474/211	93/237/199	21/56/23	22/29/22	7/12/10	43/140/144	7.5×10^{-10}	1.62(1.39–1.89)	5.8×10^{-5}
		(0.481)	(0.600)	(0.510)	(0.500)	(0.552)	(0.654)			
rs3810622	T*/C	330/445/157	263/208/58	40/48/12	28/29/16	14/12/3	181/119/27	1.0×10^{-7}	0.64(0.55–0.75)	0.0017
		(0.407)	(0.306)	(0.360)	(0.418)	(0.310)	(0.265)			

Reference (A1) and non-reference (A2) alleles refer to NCBI Reference Sequence Build 36.3 with the effective allele marked by an asterisk. Genotyping results are shown by genotype count of A1A1/A1A2/A2A2 with allele frequency of A2 in parenthesis.

†*P*-values are calculated by exact trend test with odds ratios (OR) calculated for A2 with 95% confidence interval (CI).

‡*P*-values are calculated by Jonckheere-Terpstra test in NAFLD patients for Matteoni type and additive model of genotype. SNPs are ordered by chromosomal location.
doi:10.1371/journal.pone.0038322.t002

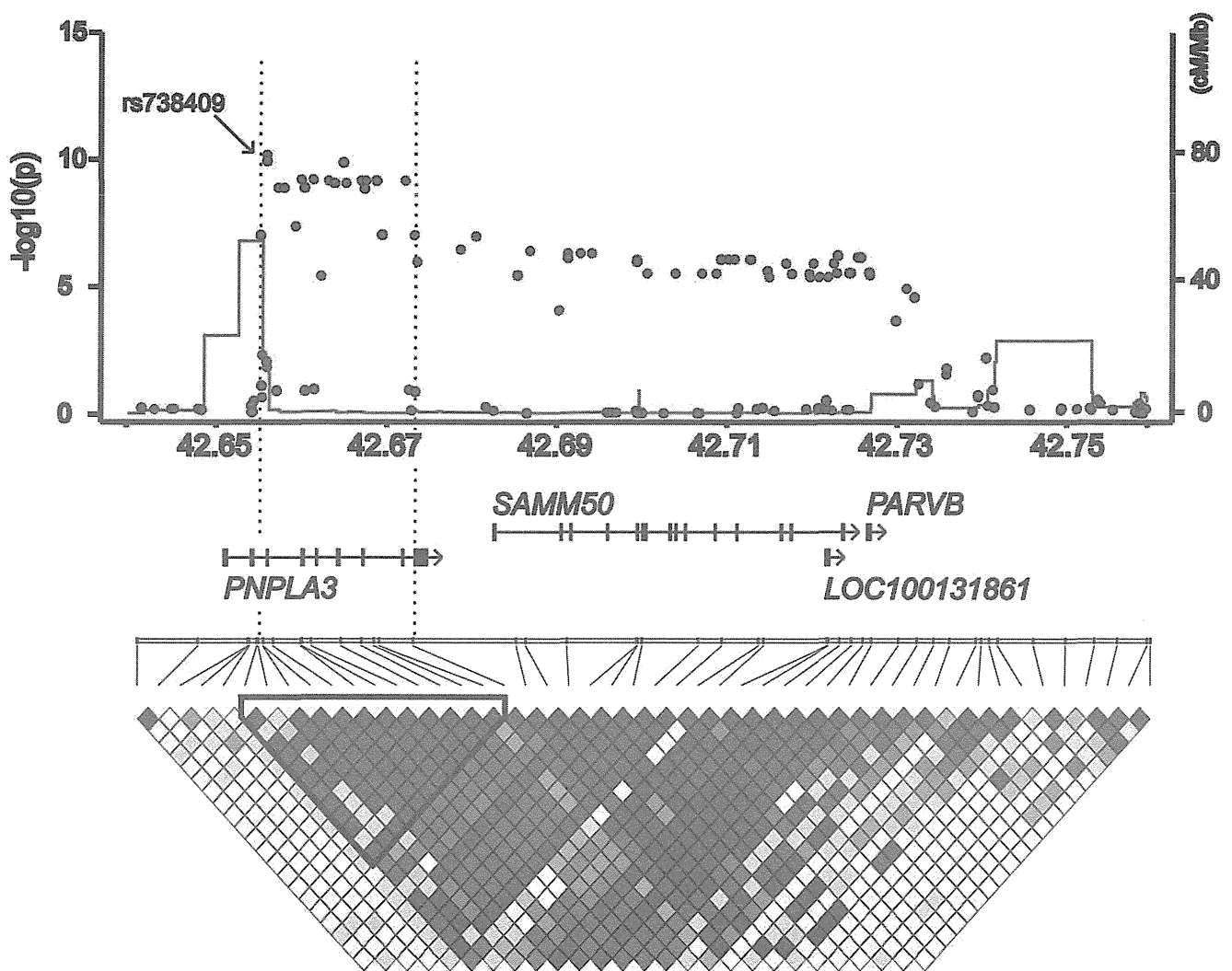


Figure 2. A schematic organization of the human *PNPLA3* locus at 22q13.31 with the genome scan results. P -values calculated by the exact trend test were plotted in $-\log_{10}$ scale. Red and blue dots indicate the p -values of genotyped and imputed SNPs, respectively. Local recombination rate obtained from HAPMAP release 22 is indicated by a red line plotted in cM/Mb scale. The structure and orientation of four genes in the region are shown below the plots with their transcriptional orientations according to NCBI Reference Sequence Build 36.3. LD blocks were generated according to pairwise LD estimates of the SNPs located within the region using the genome scan results. The LD block showing the strongest association is highlighted with the triangle, and the corresponding chromosomal region is represented by the dotted lines. doi:10.1371/journal.pone.0038322.g002

chromosomes in the Japanese result of the International HapMap Project).

Discussion

NASH is a type of hepatic steatosis in NAFLD with poor prognosis accompanying liver fibrosis, and subsequent liver cirrhosis and hepatocellular carcinoma [18]. Despite the extensive biochemical and histological investigation of NAFLD, whether or not NASH forms a distinct disease entity in NAFLD still remains unclear. The principle aim of this study was to identify the genetic factors related to the pathogenic status of NAFLD by collecting DNA samples of Japanese NAFLD patients with critically diagnosed disease status by liver biopsy. To our knowledge, this is the first GWA study of NAFLD using patients with known histology-based Matteoni type. In the initial association study using pooled genotyping results of all the cases, we found a significant association of the *PNPLA3* gene at chromosome

22q13.31 with NAFLD in the Japanese. Rs738409 which showed the strongest association with NAFLD in the GWA study of Caucasians was also genotyped and its strongest association with NAFLD was confirmed. These results were in agreement with the former GWA analyses in populations of European descent and in Hispanics, giving strong evidence of the involvement of *PNPLA3* in NAFLD beyond ethnicities. Rs738409 is located in exon3 of the *PNPLA3* gene which is expressed in the liver and adipose tissue. This SNP introduces an amino acid substitution from isoleucine to methionine (I148M), and biological studies demonstrated that its risk allele (G) abolishes the triglyceride hydrolysis activity of *PNPLA3* [19]. These observations strongly suggest rs738409 to be a causative genetic variation for NAFLD. However, future genomic analyses by fine mapping or extensive sequencing may identify additional genetic determinants within the *PNPLA3* locus.

In the current study we did not find other genetic loci showing genome-wide significance ($p < 1.0 \times 10^{-7}$). However, two additional chromosomal loci with p -values being smaller than 1×10^{-5} were