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).

Table 2. Association of HLA-DRB1 alleles with ACPA-negative RF-negative 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	69 (12.6%)	340 (11.3%)	0.37	1.13 (0.86–1.49)	57 (12.2%)	129 (12.9%)	0.70	0.94 (0.67–1.31)	126 (12.4%)	469 (11.7%)	0.52	1.07 (0.87–1.32)
*09:01	74 (13.5%)	432 (14.3%)	0.61	0.93 (0.72–1.22)	65 (13.9%)	154 (15.4%)	0.45	0.89 (0.65–1.21)	139 (13.7%)	586 (14.6%)	0.46	0.93 (0.76–1.13)
*12:01	28 (5.1%)	91 (3.0%)	0.012	1.73 (1.12–2.67)	27 (5.8%)	37 (3.7%)	0.070	1.59 (0.96–2.65)	55 (5.4%)	128 (3.2%)	0.00071	1.74 (1.26–2.40)
*13:02	28 (5.1%)	273 (9.1%)	0.0023	0.54 (0.36–0.81)	34 (7.3%)	52 (5.2%)	0.070	1.59 (0.96–2.65)	62 (6.1%)	325 (8.1%)	0.033	0.74 (0.56–0.98)
*14:03	12 (2.2%)	39 (1.3%)	0.10	1.71 (0.89–3.29)	10 (2.1%)	14 (1.4%)	0.30	1.54 (0.68–3.49)	22 (2.2%)	53 (1.3%)	0.047	1.65 (1.00–2.73)
*15:02	51 (9.3%)	369 (12.2%)	0.051	0.74 (0.54–1.00)	36 (7.7%)	113 (11.3%)	0.033	0.65 (0.44–0.97)	87 (8.6%)	482 (12.0%)	0.0020	0.69 (0.54–0.87)
SE	131 (23.9%)	677 (22.4%)	0.45	1.09 (0.88–1.34)	103 (22%)	233 (23.3%)	0.58	0.93 (0.71–1.21)	234 (23.0%)	910 (22.7%)	0.80	1.02 (0.87–1.2)
DR14	69 (12.6%)	253 (8.4%)	0.0016	1.57 (1.19–2.09)	51 (10.9%)	73 (7.3%)	0.021	1.55 (1.07–2.26)	120 (11.8%)	326 (8.1%)	0.00022	1.52 (1.21–1.89)
Diplotype												
DR8/DR8	12 (4.4%)	17 (1.1%)	9.1×10^{-5}	4.02 (1.90–8.51)	7 (3.0%)	8 (1.6%)	0.21	1.90 (0.68–5.29)	19 (3.7%)	25 (1.2%)	0.00013	3.08 (1.68–5.64)
*12:01/*09:01	4 (1.5%)	10 (0.66%)	0.25	2.22 (0.50–7.76)	4 (1.7%)	3 (0.60%)	0.22	2.88 (0.48–19.80)	8 (1.6%)	13 (0.6%)	0.040	2.46 (1.01–5.96)

doi:10.1371/journal.pone.0040067.t002

In our previous study [14], HLA-DRB1*09:01 was not significantly associated with ACPA-negative RA, in spite of the association it displayed in combination with HLA-DRB1*12:01. In the current study, we showed that HLA-DRB1*09:01 displayed a strong dose-dependent association with ACPA-negative RF-positive RA, but not with ACPA-negative RF-negative RA. These findings were confirmed by a direct comparison between the two subsets. A small study in Japan suggested that HLA-DRB1*09:01 is associated with ACPA-negative RA [15]. Our results suggest that their study mainly included ACPA-negative RF-positive RA patients. HLA-DRB1*09:01 was shown to reduce the ACPA titer in Japanese ACPA-positive RA patients [19–20]. Therefore, HLA-DRB1*09:01 might increase the titer of RF and decrease that of ACPA, although our study also showed that HLA-DRB1*09:01 is associated with ACPA-positive RF-negative RA.

HLA-DRB1*04:05, which is a major component of the SE in Asians [17], was shown to be significantly associated with ACPA-negative RA in our previous study. The current study showed that it is only associated with ACPA-negative RF-positive RA. This predisposition was also confirmed by direct comparison of the two subsets. As we could not detect a dosage effect of HLA-DRB1*04:05, its susceptibility effect might occur in a dominant manner. It is interesting that of the many SE alleles only HLA-DRB1*04:05 is associated with ACPA-negative RF-positive RA. This does not seem to be due to the relatively low frequencies of the other SE alleles (Table 1). Therefore, the common amino acid sequence that extends from the 70th to the 74th amino acid of the HLA-DRβ chain might not be important for the development of ACPA-negative RF-positive RA. As immunization of citrullinated peptide induced arthritis in HLA-DR4 transgenic mice [21] and citrullinated peptides were shown to have higher affinity to HLA-DR4 [22], high affinity of SE to citrullinated antigen is hypothesized to be the link between SE and RA development. Our findings may raise possibility of another mechanism of SE in developing arthritis.

It is quite interesting that HLA-DRB1*04:05 and *09:01, strongly associated alleles with ACPA-positive RA, are associated with ACPA-negative RF-positive RA. Although there are genetic similarities between ACPA-negative RF-positive RA and ACPA-positive RA, they should be considered to be different subsets as SE alleles other than HLA-DRB1*04:05 are not associated with ACPA-negative RF-positive RA and the HLA-DRB1*09:01 and *12:01 diplotype is strongly associated with ACPA-negative RF-positive RA.

When we analyzed the HLA-DR14 serotype, it showed a strong association with ACPA-negative RF-negative RA, largely due to HLA-DRB1*14:03 and *14:06. When we compared the frequency of DR14 in each ACPA-negative subset after stratifying the data according to the presence of HLA-DRB1*09:01 and *04:05, DR14 did not display a significant effect. In this sense, the specific association of DR14 with ACPA-negative RF-negative RA needs to be confirmed.

The HLA-DR8 homozygote displayed an association with ACPA-negative RA in our previous study [14]. The current study demonstrated that its association is specific to ACPA-negative RF-negative RA. As the number of HLA-DR8 homozygote is limited, further replication is necessary for this association. No association between the HLA-DR8 and 14 diplotype and susceptibility to ACPA-negative RF-negative RA was found (data not shown).

It is interesting that HLA-DR14 and HLA-DR8, associated serotype with ACPA-negative RF-negative RA, were reported association with psoriatic arthritis [23]. HLA-DR14 is often linked with HLA-Cw*06, susceptibility serotype to psoriasis arthritis in European [24]. HLA-Cw*06 is rare in Japanese (<1%) and the

Table 3. Direct comparison of HLA-DRB1 allele frequency between ACPA-negative RF-positive RA and ACPA-negative RF-negative RA.

HLA-DRB1	ACPA(−)RF(+)RA Number of allele (%)	ACPA(−)RF(−)RA Number of allele (%)	*p*	OR (95%CI)
*09:01	169 (18.9%)	139 (13.7%)	0.0018	1.47 (1.15–1.88)
*04:05	153 (17.2%)	126 (12.4%)	0.0034	1.46 (1.13–1.89)
*08:02	24 (2.7%)	52 (5.1%)	0.0068	0.51 (0.31–0.84)
*14:06	8 (0.9%)	21 (2.1%)	0.037	0.43 (0.19–0.97)
SE	256 (28.7%)	234 (23.0%)	0.0047	1.35 (1.09–1.65)
DR14	78 (8.7%)	120 (11.8%)	0.028	0.72 (0.53–0.97)
DR8/DR8	6 (1.3%)	19 (3.7%)	0.021	0.35 (0.14–0.89)

doi:10.1371/journal.pone.0040067.t003

strong association between HLA-Cw*06 and HLA-DR14 is not observed in Japan (<10%). While psoriatic arthritis is not reported to be associated with these serotypes in Japan, association between these serotypes and arthritis is interesting.

It could be argued that ACPA-negative RA includes some non-RA arthritic diseases such as psoriasis, seronegative spondyloarthropathy and other collagen vascular diseases. Thus, we analyzed the associations between the above-mentioned alleles and diplotypes with ACPA-negative RA displaying bone erosion to examine whether the same association patterns were present in this strictly defined cohort. The typical bone erosions of RA are rarely seen in other arthritic disorders. As a result, we found the same associations. Therefore, we are convinced that our findings were not caused by the contamination of our study population by patients with other diseases. Since RF sometimes normalizes after treatment, the RF-negative RA patients whose RF titers were not measured at multiple points might not have been RF-negative. So, we re-analyzed our data by excluding the RA patients for whom consecutive RF titers were not available. As a result, we found the same tendency of associations for each allele and diplotype in each subset (data not shown), indicating that these subsets are stable.

Analysis using ACPA-positive RF-positive RA and ACPA-positive RF-negative RA patients compared with healthy controls did not result in distinct differences in HLA-DRB1 association. The SE is associated with both ACPA-positive RF-positive and RF-negative RA. HLA-DRB1*09:01 was found to be associated with both subsets after stratifying the patients according to their SE alleles. We also did not detect an association between HLA-DR14 or the HLA-DR8 homozygote and either subset. While 154 ACPA-positive RF-negative RA patients in our study are too small in number to detect the difference in HLA-DRB1 alleles with weak effect size between the two ACPA-positive subsets, these results suggest that there are no big differences in the HLA usage of the two subsets in ACPA-positive RA. To confirm our results and to detect possible different frequency of other HLA-DRB1 alleles in the two ACPA-positive subsets, replication study is necessary.

Table 4. Logistic regression analysis of HLA-DRB1 alleles with ACPA-positive RF-positive RA and ACPA-positive RF-negative RA.

	ACPA(+)RF(+)RA		ACPA(+)RF(−)RA	
HLA-DRB1	*p**	OR (95%CI)*	*p**	OR (95%CI)*
SE	$<2\times10^{-16}$	3.21 (2.72–3.78)	$<2\times10^{-16}$	3.03 (2.33–3.94)
*09:01	2.4×10^{-9}	1.83 (1.5–2.25)	0.0035	1.67 (1.17–2.37)

*p-values and odds ratios in logistic regression analysis using SE and HLA-DRB1*09:01.

doi:10.1371/journal.pone.0040067.t004

In the current study, we performed multiple comparisons in each subset and between subsets. The associations should be evaluated in the combined analysis with significant level corrected by Bonferronii's method and independency of each association should be evaluated by logistic regression analysis or RPE method. In this sense, p-values around cut-off level in the combined analysis should be taken with caution and the associations should be confirmed by independent study.

We have shown that ACPA-negative RA includes two genetically distinct subsets in Japanese population: RF-positive and RF-negative RA. This is the first report in Asians to show that these subsets are genetically distinct. We have to clarify the clinical difference between these two subsets. We also have to clarify whether non-HLA genes display different associations with each subset. So far, many genome wide association studies (GWAS) of RA and ACPA-positive RA have been performed, and more than twenty genes or loci have been shown to be susceptibility loci [25–38]. However, no GWAS studies have detected susceptibility genes for ACPA-negative RA with genome-wide significance [39]. This is probably due to the relatively small number of patients studied, but it might be overcome by stratifying ACPA-negative RA patients into RF-positive and RF-negative subsets. Since RA susceptibility genes usually cross ethnic boundaries [40], global collaboration might result in a fruitful dissection of these minor subsets.

Materials and Methods

Ethics Statement

This study was approved by the local ethical committees at each institution, namely, Kyoto University Graduate School and Faculty of Medicine, Ethics Committee, Tokyo Women's Medical University Genome Ethics Committee, and the ethics committee of RIKEN, and written informed consent was obtained from all patients.

Study Subjects

DNA samples were collected from ACPA-negative RA patients at Kyoto University Hospital, Tokyo Women's Medical University [41], and RIKEN with the support of BioBank Japan. All patients were Japanese and had been diagnosed by rheumatologists

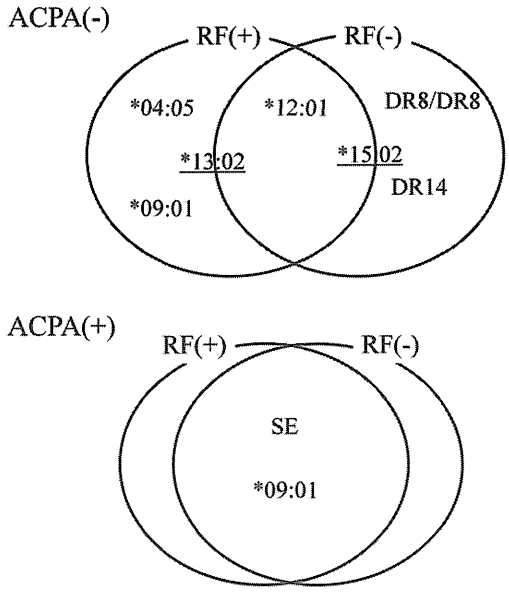

Figure 1. Summary of the HLA-DRB1 alleles associated with ACPA-negative RA and ACPA-positive RA. The relationships between the RF-positive and RF-negative subsets of ACPA-negative and ACPA-positive RA in terms of their associations with HLA-DRB1 alleles are illustrated. While the two subsets of ACPA-positive RA seem to share most associations with HLA-DRB1, the two ACPA-negative RA subsets possess specific alleles and HLA-DRB1 diplotypes. The underlined alleles are protective alleles.
doi:10.1371/journal.pone.0040067.g001

according to the 1987 American College of Rheumatology revised criteria for RA [42]. The control DNA samples were collected at Aichi Cancer Center Hospital, the DNA banks of the Pharma SNP Consortium [43], and HLA laboratory. A more detailed description of the collection procedure was given in a previous study [14]. We performed association studies using similar study design of the two collections to our previous study; namely, collection 1 for 456 ACPA-negative RA and 1508 healthy subjects, and collection 2 for 501 ACPA-negative RA and 500 healthy people. RF data were available for 453 out of 456 cases in collection 1 and all of 501 cases in collection 2. 179 patients were RF-positive in collection 1 and 267 patients were RF-positive in collection 2. We also collected DNA samples from 531 ACPA-positive RF-positive RA patients at Kyoto University Hospital and 154 ACPA-positive RF-negative RA patients at Kyoto University and Tokyo Women's Medical University.

ACPA Detection

The MESACUP CCP ELISA kit (Medical and Biological Laboratories Co., Ltd, Nagoya, Japan) was used to detect 2nd generation ACPA in each RA patient, according to the manufacturer's instructions. A cut-off value of 4.5 U/ml was used to define ACPA positivity.

RF Detection

The serum RF concentrations of samples in collection 1 were quantified using a latex agglutination turbidimetric immunoassay. An ELISA assay was used to determine the RF levels of samples in collection 2. When multiple values for RF had been obtained at different visits, we used the maximum RF value for each patient. The cut off values of each detection kit in each hospital were employed.

HLA-DRB1 Genotyping

The HLA-DRB1 typing methods were previously described [14]. Briefly, the WAKFlow system or the AlleleSEQR HLA-DRB1 typing kit (Abbott, Tokyo, Japan) was used for the HLA-DRB1 typing. The following HLA-DRB1 alleles were classified as belonging to the SE: 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.

Statistical Analysis

The frequency of each allele or diplotype was compared among the ACPA-negative RF-positive RA, ACPA-negative RF-negative RA patients, and the healthy controls in each set and combined set using the chi-square test or Fisher's exact test. The same analyses were performed in ACPA-positive RA patients classified according to their RF possession. Ninety-five percent confidence intervals (CI) for the OR were also calculated. Logistic regression analysis was used to evaluate the effects of each allele by adjusting for the influence of strongly-associated alleles. Single alleles were regarded as significant when they showed p-values of less than 0.0026 in a combined study, which is obtained by Bonferroni's correction. For diplotype analyses, we regarded 0.025 as the cut off level for significance because we performed just two tests. All statistical analyses were performed using the R statistic system (http://www.R-project.org) or SPSS (version 18).

Supporting Information

Figure S1 Dosage effects of HLA-DRB1*04:05 and *09:01 alleles on ACPA-negative RF-positive RA susceptibility. Each column represents the odds ratio for developing ACPA-negative RF-positive RA associated with possessing one (red column) or two (green column) alleles of HLA-DRB1*04:05 or *09:01.
(TIF)

Table S1 Logistic regression analysis of associated alleles with ACPA-negative RF-positive RA. *p-values and odds ratios in logistic regression analysis using the six alleles listed above.
(DOC)

Table S2 Association between HLA-DR14 and ACPA-negative RF-negative RA.
(DOC)

Table S3 Logistic regression analysis of associated alleles with ACPA-negative RF-negative RA. *p-values and odds ratios in logistic regression analysis using HLA-DR14 and three HLA-DRB1 alleles listed above.
(DOC)

Table S4 Association of HLA-DRB1 with ACPA-negative RA erosive subsets. [a)]Total allele number is 268. [b)]Total allele number is 212.
(DOC)

Table S5 Logistic regression analysis of assoicated alleles with ACPA-negative RF-positive RA, compared with ACPA-negative RF-negative RA. *p-values and odds ratios in logistic regression analysis using HLA-DRB1*09:01, *04:05, and HLA-DR14. [a)]HLA-DRB1 alleles which showed $p<0.05$ in Table 3 were used for analysis.
(DOC)

Table S6 Comparison between ACPA-positive RF-positive RA and ACPA-positive RF-negative RA. [a)] Alleles with frequency more than 1% in any groups are shown.
(DOC)

Acknowledgments

We would like to thank Dr. Naoichiro Yukawa, Dr. Hajime Yoshifuji, Dr. Daisuke Kawabata, Dr. Takaki Nojima, Dr. Takashi Usui, and Dr. Takao Fujii of Kyoto University for collecting the DNA samples. We would also like to thank Mr. Taishi Shigeki for developing the clinical database software used at Dohgo Spa hospital. We would like to thank Dr. Yasuo Miura and Dr. Taira Maekawa of Kyoto University for their support of HLA-DRB1 genotyping. Moreover, we wish to thank all of the doctors and medical staff who collected the samples. This study was performed with the support of the Genetics and Allied research in Rheumatic diseases Networking (GARNET) consortium.

Author Contributions

Conceived and designed the experiments: CT KO KI YK RY FM TM. Performed the experiments: CT KI YK EM K. Yurugi MK AS HS. Analyzed the data: CT. Contributed reagents/materials/analysis tools: KI EM KS AM SH K. Takasugi KM K. Tajima SM HY K. Yamamoto HS TM. Wrote the paper: CT KO.

References

1. Firestein GS (2003) Evolving concepts of rheumatoid arthritis. Nature 423: 356–361.
2. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, et al. (2000) Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum 43: 30–37.
3. Kallberg H, Ding B, Padyukov L, Bengtsson C, Ronnelid J, et al. (2011) Smoking is a major preventable risk factor for rheumatoid arthritis: estimations of risks after various exposures to cigarette smoke. Ann Rheum Dis 70: 508–511.
4. Deighton CM, Walker DJ, Griffiths ID, Roberts DF (1989) The contribution of HLA to rheumatoid arthritis. Clin Genet 36: 178–182.
5. Gorman JD, Lum RF, Chen JJ, Suarez-Almazor ME, Thomson G, et al. (2004) Impact of shared epitope genotype and ethnicity on erosive disease: a meta-analysis of 3,240 rheumatoid arthritis patients. Arthritis Rheum 50: 400–412.
6. Gregersen PK, Silver J, Winchester RJ (1987) The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum 30: 1205–1213.
7. Ohmura K, Terao C, Maruya E, Katayama M, Matoba K, et al. (2010) Anti-citrullinated peptide antibody-negative RA is a genetically distinct subset: a definitive study using only bone-erosive ACPA-negative rheumatoid arthritis. Rheumatology (Oxford) 49: 2298–2304.
8. van der Woude D, Lie BA, Lundstrom E, Balsa A, Feitsma AL, et al. (2010) Protection against anti-citrullinated protein antibody-positive rheumatoid arthritis is predominantly associated with HLA-DRB1*1301: a meta-analysis of HLA-DRB1 associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in four European populations. Arthritis Rheum 62: 1236–1245.
9. Schellekens GA, Visser H, de Jong BA, van den Hoogen FH, Hazes JM, et al. (2000) The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide. Arthritis Rheum 43: 155–163.
10. Morgan AW, Thomson W, Martin SG, Carter AM, Erlich HA, et al. (2009) Reevaluation of the interaction between HLA-DRB1 shared epitope alleles, PTPN22, and smoking in determining susceptibility to autoantibody-positive and autoantibody-negative rheumatoid arthritis in a large UK Caucasian population. Arthritis Rheum 60: 2565–2576.
11. Huizinga TW, Amos CI, van der Helm-van Mil AH, Chen W, van Gaalen FA, et al. (2005) Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins. Arthritis Rheum 52: 3433–3438.
12. Verpoort KN, van Gaalen FA, van der Helm-van Mil AH, Schreuder GM, Breedveld FC, et al. (2005) Association of HLA-DR3 with anti-cyclic citrullinated peptide antibody-negative rheumatoid arthritis. Arthritis Rheum 52: 3058–3062.
13. Lundstrom E, Kallberg H, Smolnikova M, Ding B, Ronnelid J, et al. (2009) Opposing effects of HLA-DRB1*13 alleles on the risk of developing anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. Arthritis Rheum 60: 924–930.
14. Terao C, Ohmura K, Kochi Y, Ikari K, Maruya E, et al. (2011) A large-scale association study identified multiple HLA-DRB1 alleles associated with ACPA-negative rheumatoid arthritis in Japanese subjects. Ann Rheum Dis 70(12): 2134–2139.
15. Furuya T, Hakoda M, Ichikawa N, Higami K, Nanke Y, et al. (2007) Differential association of HLA-DRB1 alleles in Japanese patients with early rheumatoid arthritis in relationship to autoantibodies to cyclic citrullinated peptide. Clin Exp Rheumatol 25: 219–224.
16. Mackie SL, Taylor JC, Martin SG, Wordsworth P, Steer S, et al. (2012) A spectrum of susceptibility to rheumatoid arthritis within HLA-DRB1: stratification by autoantibody status in a large UK population. Genes Immun 13: 120–128.
17. Lee HS, Lee KW, Song GG, Kim HA, Kim SY, et al. (2004) Increased susceptibility to rheumatoid arthritis in Koreans heterozygous for HLA-DRB1*0405 and *0901. Arthritis Rheum 50: 3468–3475.
18. Payami H, Joe S, Farid NR, Stenszky V, Chan SH, et al. (1989) Relative predispositional effects (RPEs) of marker alleles with disease: HLA-DR alleles and Graves disease. Am J Hum Genet 45: 541–546.
19. Okada Y, Suzuki A, Yamada R, Kochi Y, Shimane K, et al. (2010) HLA-DRB1*0901 lowers anti-cyclic citrullinated peptide antibody levels in Japanese patients with rheumatoid arthritis. Annals of the Rheumatic Diseases 69: 1569–1570.
20. Terao C, Ikari K, Ohmura K, Suzuki T, Iwamoto T, et al. (2012) Quantitative effect of HLA-DRB1 alleles to ACPA levels in Japanese rheumatoid arthritis: no strong genetic impact of shared epitope to ACPA levels after stratification of HLA-DRB1*09:01. Annals of the Rheumatic Diseases 71: 1095–1097.
21. Hill JA, Bell DA, Brintnell W, Yue D, Wehrli B, et al. (2008) Arthritis induced by posttranslationally modified (citrullinated) fibrinogen in DR4-IE transgenic mice. Journal of Experimental Medicine 205: 967–979.
22. Hill JA, Southwood S, Sette A, Jevnikar AM, Bell DA, et al. (2003) Cutting edge: the conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1*0401 MHC class II molecule. Journal of Immunology 171: 538–541.
23. Queiro-Silva R, Torre-Alonso JC, Tinture-Eguren T, Lopez-Lagunas I (2004) The effect of HLA-DR antigens on the susceptibility to, and clinical expression of psoriatic arthritis. Scand J Rheumatol 33: 318–322.
24. Ho PY, Barton A, Worthington J, Plant D, Griffiths CE, et al. (2008) Investigating the role of the HLA-Cw*06 and HLA-DRB1 genes in susceptibility to psoriatic arthritis: comparison with psoriasis and undifferentiated inflammatory arthritis. Ann Rheum Dis 67: 677–682.
25. Suzuki A, Yamada R, Chang X, Tokuhiro S, Sawada T, et al. (2003) Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. Nat Genet 34: 395–402.
26. Kochi Y, Yamada R, Suzuki A, Harley JB, Shirasawa S, et al. (2005) A functional variant in FCRL3, encoding Fc receptor-like 3, is associated with rheumatoid arthritis and several autoimmunities. Nat Genet 37: 478–485.
27. Suzuki A, Yamada R, Kochi Y, Sawada T, Okada Y, et al. (2008) Functional SNPs in CD244 increase the risk of rheumatoid arthritis in a Japanese population. Nat Genet 40: 1224–1229.
28. Kochi Y, Okada Y, Suzuki A, Ikari K, Terao C, et al. (2010) A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. Nat Genet 42: 515–519.
29. Terao C, Ohmura K, Katayama M, Takahashi M, Kokubo M, et al. (2011) Myelin basic protein as a novel genetic risk factor in rheumatoid arthritis-a genome-wide study combined with immunological analyses. PLoS One 6: e20457.
30. Terao C, Yamada R, Ohmura K, Takahashi M, Kawaguchi T, et al. (2011) 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.
31. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, et al. (2007) TRAF1-C5 as a risk locus for rheumatoid arthritis–a genomewide study. N Engl J Med 357: 1199–1209.
32. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, et al. (2007) STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 357: 977–986.
33. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, et al. (2007) Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 39: 1477–1482.
34. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678.
35. Thomson W, Barton A, Ke X, Eyre S, Hinks A, et al. (2007) Rheumatoid arthritis association at 6q23. Nat Genet 39: 1431–1433.
36. Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, et al. (2008) Common variants at CD40 and other loci confer risk of rheumatoid arthritis. Nat Genet 40: 1216–1223.
37. Gregersen PK, Amos CI, Lee AT, Lu Y, Remmers EF, et al. (2009) REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. Nat Genet 41: 820–823.
38. Raychaudhuri S, Thomson BP, Remmers EF, Eyre S, Hinks A, et al. (2009) Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. Nat Genet 41: 1313–1318.
39. Padyukov L, Seielstad M, Ong RT, Ding B, Ronnelid J, et al. (2011) A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis. Ann Rheum Dis 70: 259–265.
40. Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, et al. (2012) Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nature Genetics 44: 511–516.

41. Matsuda Y, Singh G, Yamanaka H, Tanaka E, Urano W, et al. (2003) Validation of a Japanese version of the Stanford Health Assessment Questionnaire in 3,763 patients with rheumatoid arthritis. Arthritis Rheum 49: 784–788.
42. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, et al. (1988) The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 31: 315–324.
43. Kamatani N, Sekine A, Kitamoto T, Iida A, Saito S, et al. (2004) Large-scale single-nucleotide polymorphism (SNP) and haplotype analyses, using dense SNP Maps, of 199 drug-related genes in 752 subjects: the analysis of the association between uncommon SNPs within haplotype blocks and the haplotypes constructed with haplotype-tagging SNPs. Am J Hum Genet 75: 190–203.

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 Sumida[3], Atsushi Umemura[4], Keitaro Matsuo[5], Meiko Takahashi[1], Toshinari Takamura[6], Kohichiroh Yasui[7], Toshiji Saibara[8], Etsuko Hashimoto[9], Miwa Kawanaka[10], Sumio Watanabe[11], Sumio Kawata[12], Yasuharu Imai[13], Miki Kokubo[1], Toshihide Shima[4], Hyohun Park[4], Hideo Tanaka[5], Kazuo Tajima[5], Ryo Yamada[1], Fumihiko Matsuda[1,2]*, Takeshi Okanoue[4] 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 ($p=3.6\times10^{-6}$) 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 ($p=4.8\times10^{-6}$, 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 ($p=1.7\times10^{-16}$, 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 ($p=4.6\times10^{-4}$), HbA1c ($p=0.0011$) and iron deposition in the liver ($p=5.6\times10^{-4}$).

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

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 ($p<1.04\times10^{-7}$) (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} ($p = 1.5\times10^{-10}$, 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 ($p = 1.0\times10^{-7}$ 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 ($p = 1.4\times10^{-10}$, 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] ($p = 2.3\times10^{-11}$). 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 ($p>1.6\times10^{-3}$). 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 ($p = 3.6\times10^{-6}$) 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) ($p = 2.0\times10^{-5}$, OR = 2.18, 95%CI: 1.52–3.18 between type1 and type4; $p = 1.4\times10^{-3}$, 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 ($p = 4.8\times10^{-6}$, 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 ($p = 1.7\times10^{-16}$, 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 ($\times 10^4/\mu L$)	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,

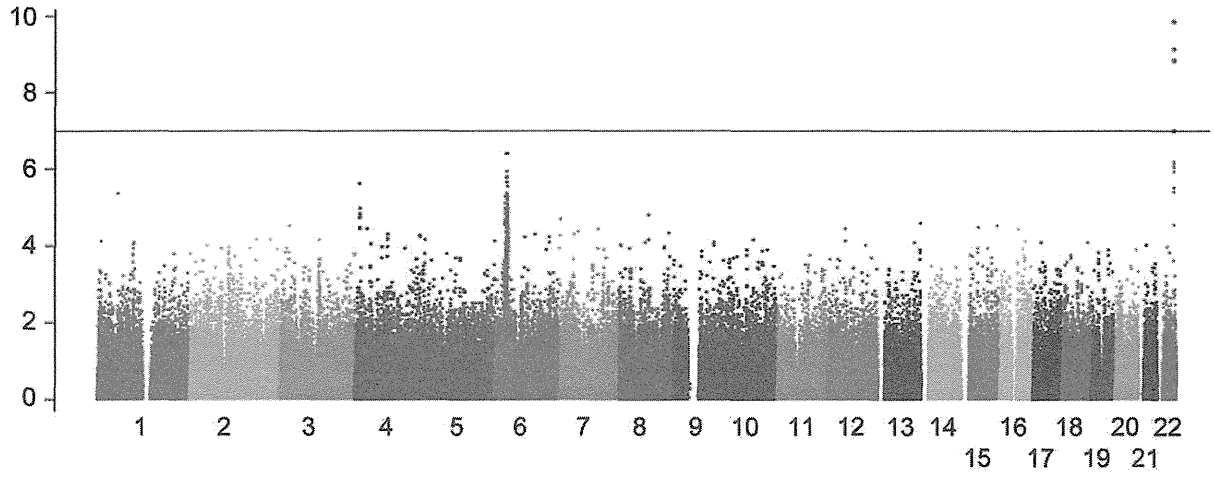

Figure 1. Manhattan plot of the GWA study. Association *p*-values are calculated by exact trend test and plotted along the chromosome in $-\log_{10}$ scale. The horizontal line indicates Bonferroni-adjusted significance threshold ($p = 1.03 \times 10^{-7}$).
doi:10.1371/journal.pone.0038322.g001

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 *PPP1R3B* 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 *PPP1R3B* 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.

		Genotyping Result and Allele Frequency of A2						Statistics		
			NAFLD					NAFLD vs. Control		Matteoni
dbSNPID	A1/A2	Control	Total	Type 1	Type 2	Type 3	Type 4	*p*-value†	OR (95%CI)	*p*-value‡
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

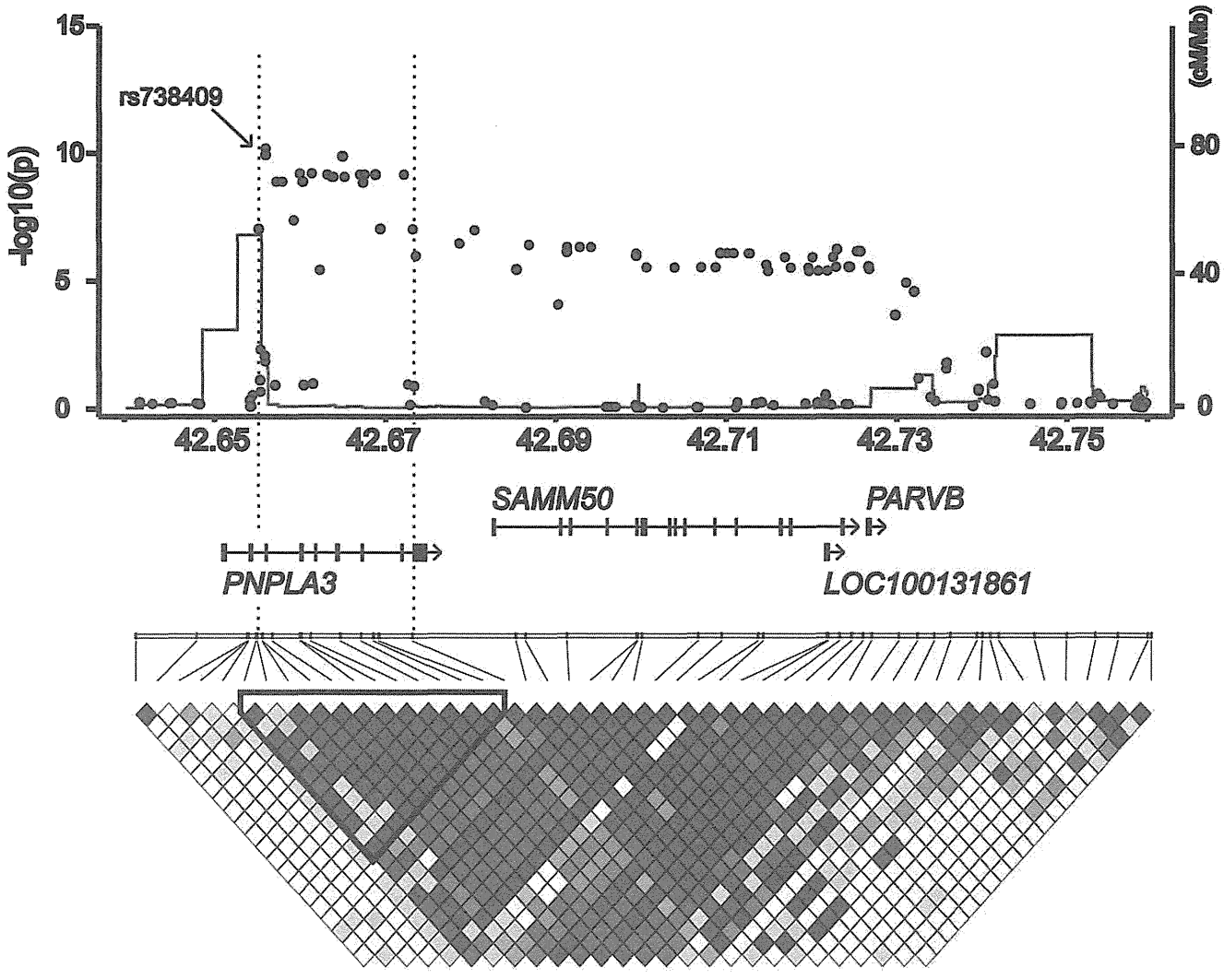

Figure 2. A schematic organization of the human *PNPLA3* locus at 22q13.31 with the genome scan results. *P*-values calculated by the exact trend test were plotted in $-\log_{10}$ scale. Red and blue dots indicate the *p*-values of genotyped and imputed SNPs, respectively. Local recombination rate obtained 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 ($p<1.0\times10^{-7}$). However, two additional chromosomal loci with *p*-values being smaller than 1×10^{-5} were

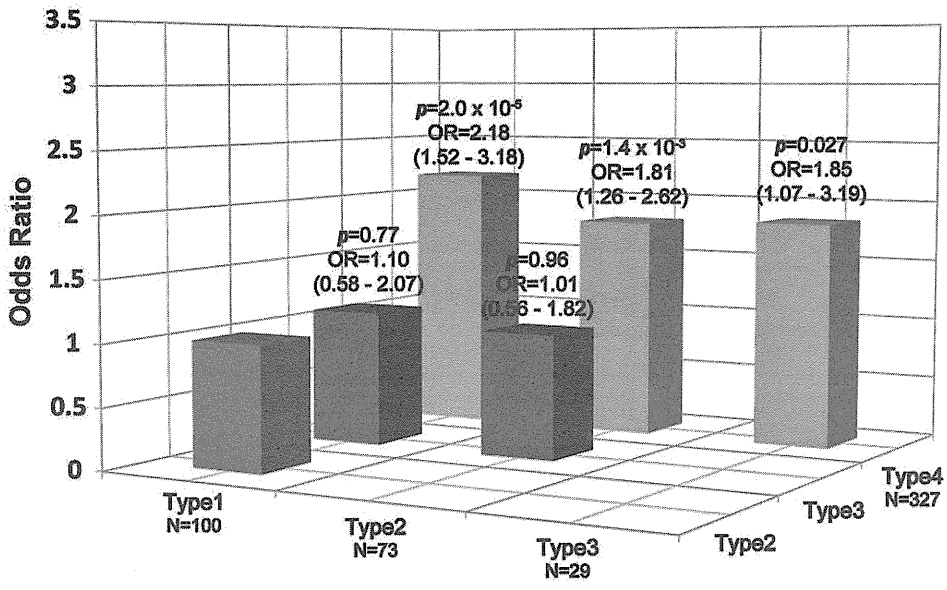

Figure 3. Histogram of odds ratios for genotype distribution of rs738409 between Matteoni types. Each box denotes the odds ratio (OR) comparing the corresponding Matteoni types on the horizontal axes. N represents the number of samples. Odds ratios and *p*-values are calculated for the higher Matteoni type per risk allele (G) on additive model by multivariable logistic regression adjusted for age, sex and BMI, and are shown with 95% CI above each box.
doi:10.1371/journal.pone.0038322.g003

identified on chromosome 1p (rs11206226) and chromosome 4p (rs1390096) neither of which has been reported as being associated with NAFLD in Caucasians (Table S1). Statistical calculation by taking their allele frequencies and effect sizes into account showed that approximately three times as many case and control samples are required to obtain sufficient statistical power (>0.8) for genome wide significance. Hence, further confirmation is required using a larger collection of patients and controls although they may be potential candidates of low-penetrance genes for susceptibility to NAFLD in Japanese.

Subsequent analyses through comparison of genotype distribution among four subgroups of NAFLD (type1 to type4) categorized by Matteoni's classification revealed that the seven NAFLD-associated SNPs in the *PNPLA3* gene were also significantly associated with the pathogenic status of NAFLD. There were also marked differences in genotype distribution of rs738409 between type4 subgroup and the other three groups ($p = 4.8 \times 10^{-6}$, OR = 1.96, 95%CI: 1.47–2.62 between type4 and pooled genotypes of type1 to type3). Moreover, a case/control analysis of rs738409 between Matteoni type4 and controls returned a surprisingly strong association ($p = 1.7 \times 10^{-16}$) which was much stronger than the initial analysis using all NAFLD cases ($p = 1.4 \times 10^{-10}$), whereas the analysis using Matteoni type1 to type3 as cases didn't show significance ($p = 0.41$). There were differences in the score of HOMA-IR and hs-CRP, indicators of insulin resistance and inflammation, respectively, between Matteoni type1 to type3 and type4 subgroups (Table 1). Our results provide compelling evidence that NASH corresponding to Matteoni type4 is both a clinically and genetically different disease subset from other spectrums of NAFLD. Previous studies showed association between *PNPLA3* and fatty liver, inflammation, fibrosis grade and NASH [13]. In our result, strong association between rs738409 and fatty liver was not observed by comparing control and Matteoni type1. In addition, strong association between rs738409 and lobular inflammation was not observed by comparing Matteoni Type1 and Type2. In contrast, a strong association between rs738409 and NASH was observed. Although we could not observe the strong association between rs738409 and fibrosis stage, strong association between rs738409 and Hyaluronic acid suggests that an association exists between *PNPLA3* and fibrosis.

We have also undertaken association analyses of rs738409 and clinical traits in the patients. The multivariable regression analysis adjusted for age, sex, BMI and Matteoni type followed by the correction for multiple testing revealed hyaluronic acid and HbA1c as being significantly associated with rs738409. Hyaluronic acid is one of the principle components of the extracellular matrix and its involvement in fibrosis has been previously suggested [20]. This may indicate another possible functional involvement of *PNPLA3* in the progression of liver fibrosis by influencing the circulating hyaluronic acid levels. A weak association of rs738409 and HbA1c levels was observed in our study population. However, there are no reports to date indicating such an association, and confirmation with different sample sets is needed for definitive conclusion. Also, the association between rs738409 and iron deposition was demonstrated by an ordinal logistic regression analysis. Since the association still remained after the results were adjusted with Matteoni type, rs738409 may play a functional role in the oxidative stress through iron absorption in the liver.

Recently, a genetic analysis of Japanese NAFLD patients was reported demonstrating a significant association in the increase of AST, ALT, ferritin levels and fibrosis stage (Brunt stage) and in the decrease of serum triglyceride with the risk allele (G) of rs738409 [12]. In our study, the association of rs738409 with AST ($p = 1.2 \times 10^{-4}$) and ALT ($p = 0.0016$) was reproduced and that of AST still remained after the results were adjusted for Matteoni type ($p = 0.038$). No association was observed for ferritin level. Brunt stage was available for Matteoni type4 patients only in our study. Although the odds ratio was slightly high (OR = 1.28, 95%CI: 0.95–1.72), it was not possible to examine the association. In addition, the inverse association of the risk allele of rs738409 with decrease of serum triglyceride was confirmed in our study ($p = 0.013$ after being adjusted for Matteoni type). For all of these

Table 3. Association of rs738409 with clinical traits.

Biochemical traits	Statistical calculation1		Statistical calculation 2	
Phenotype	**Coef. (S.E.)**	***p*-value**	**Coef. (S.E.)**	***p*-value**
Biological traits				
AST (IU/L)	0.22 (0.056)	**1.2×10^{-4}**	0.11 (0.052)	0.038
ALT (IU/L)	0.19 (0.058)	**0.0016**	0.093 (0.056)	0.098
GGT (IU/L)	−0.056 (0.061)	0.37	−0.088 (0.062)	0.16
Albumin (g/dL) *	0.015 (0.051)	0.77	−0.012 (0.052)	0.81
Total bilirubin (mg/dL)	−0.011 (0.063)	0.86	0.0059 (0.064)	0.93
Cholinesterase (unit) *	0.062 (0.040)	0.12	0.069 (0.041)	0.092
Type IV collagen 7S (ng/dL) *	−0.19 (0.064)	0.0025	−0.11 (0.062)	0.069
Hyaluronic acid (ng/dL)	0.30 (0.065)	**4.9×10^{-6}**	0.22 (0.063)	**4.6×10^{-4}**
Triglycerides (mg/dL)	−0.10 (0.058)	0.072	−0.15 (0.059)	0.013
Total cholesterol (mg/dL)	−0.066 (0.060)	0.27	−0.057 (0.061)	0.34
HbA1c (%)	−0.17 (0.053)	**0.0012**	−0.18 (0.054)	**0.0011**
IRI (μg/dL)	0.16 (0.063)	0.012	0.086 (0.061)	0.16
FPG (mg/dL)	−0.14 (0.049)	0.0047	−0.15 (0.05)	0.0035
HOMA-IR	0.084 (0.064)	0.19	0.0092 (0.062)	0.88
Hs-CRP (mg/dL)	−0.013 (0.048)	0.79	−0.031 (0.049)	0.52
Adiponectin (μg/mL)	0.048 (0.066)	0.47	0.12 (0.066)	0.072
Leptin (ng/mL)	0.11 (0.068)	0.11	0.10 (0.069)	0.15
Ferritin (ng/mL)	0.031 (0.047)	0.51	−0.0042 (0.048)	0.93
Uric acid (mg/dL)	−0.097 (0.061)	0.11	−0.11 (0.062)	0.067
PLT (x10^4/μL)	−0.056 (0.020)	0.0052	−0.045 (0.020)	0.028
Immunological/histological traits				
ANA (0/1/2/3/4)	0.92 (0.70–1.21)	0.56	0.86 (0.65–1.15)	0.31
Brunt grade (1/2/3)	1.42 (1.06–1.92)	0.021	1.38 (1.02–1.87)	0.036
Brunt stage (1/2/3/4)	1.28 (0.95–1.72)	0.11		
Fat deposition (1/2/3/4)	1.44 (1.15–1.81)	0.0019	1.24 (0.98–1.56)	0.76
Iron deposition (0/1/2/3/4)	0.61 (0.47–0.80)	**3.0×10^{-4}**	0.62 (0.47–0.81)	**5.6×10^{-4}**

Associations between distribution of rs738409 genotypes and clinical traits are calculated by multivariable regression. Statistical calculation1 is adjusted for age, sex and BMI, while the Matteoni types are additionally included as covariate in statistical calculation 2. Statistics are calculated by multivariable linear regression for biochemical traits and by multivariable ordinal logistic regression for immunological and histological traits.

Coefficients and odds ratios are calculated for the increase of each trait per risk allele (G). The *p*-values showing significance after Bonferroni's correction for multiple testing ($p = 0.0021$) was shown in bold.

*Reciprocal numbers are used for normalization and a negative coefficient implicates an increase in value according to the increase of the risk allele.

doi:10.1371/journal.pone.0038322.t003

biomarkers, however, the significance was lost after the correction for multiple testing.

A replication analysis of other genetic loci that had been reported for their association with NAFLD in East coast white Americans [14] was performed in our sample collection. We confirmed the association of rs780094 in *GCKR* with NAFLD in a case/control analysis but at a much weaker level ($p = 0.011$, OR = 0.82, 95%CI: 0.70–0.95) than that shown for the populations of European-descent. No associations were found for *LYPLAL1* and *NCAN* loci in our study. There are several reasons to explain such differences, such as the insufficient statistical power with a limited number of study subjects in our study due to the difficulty in the collection of a larger number of histologically diagnosed NAFLD patients. The difference in genetic background between the Japanese and Europeans is also conceivable. Indeed, the risk allele frequency of rs12137855 in *LYPLAL1* was 0.944 in our control subjects but approximately 0.79 in the European populations [14]. Similarly, there was a difference in the risk allele frequency of rs2228603 in *NCAN* (0.049 in Japanese and 0.08 in Europeans). Rs4240624 in *PPP1R3B* was not polymorphic in the Japanese while its risk allele frequency was 0.91 in Europeans.

Materials and Methods

Ethics Statement

In compliance with the Declaration of Helsinki, ethical approval for this study was given by the respective Institutional Review Board and subject written informed consent were obtained for all subjects (Ethical committee of Nara City Hospital; Ethical committee of Saiseikai Suita Hospital; Medical Ethics Committee of Kanazawa University; Ethics committee of Kyoto Prefectural University of Medicine; Ethical Committee of Aichi Cancer Center; Ethical Committee of Kochi Medical School, Kochi University; Ethics Committee of Tokyo Women's Medical University; Ethical Committee on Kawasaki Medical School and Kawasaki Medical School Hospital; Ethical Committee of

Juntendo University; Ethics Committee of Yamagata University School of Medicine; Ethical Committee of the Ikeda Municipal Hospital; Institutional Review Board and Ethics Committee of Kyoto University School of Medicine).

Study Population

A total of 543 patients histologically diagnosed for NAFLD in 2007–2009 were recruited through the Japan study of Nonalcoholic Fatty Liver Disease. Biopsy specimens were stained with H&E and Masson's trichrome for morphological review and assessment of fibrosis. Perl's Prussian blue was performed to evaluate iron load. Biopsy specimens were reviewed by a hepatopathologist (T.O). NAFLD patients were classified into four categories by liver histology according to the classification by Matteoni *et al* [2] as follows; type1: fatty liver alone, type2: fat accumulation and lobular inflammation, type3: fat accumulation and ballooning degeneration, type4: fat accumulation, ballooning degeneration, and either Mallory-Denk body or fibrosis. With these criteria, the 543 patients were classified as type1; 102, type2; 75, type3; 31 and type4; 335. The histological grade and fibrosis stage were also evaluated by the classification of Brunt *et al* [21] for advanced NAFLD cases (type3 and type4) as follows; grade 1: steatosis involving up to 66% of biopsy, occasional ballooned zone 3 hepatocytes and absence or mild portal chronic inflammation, grade2: steatosis, ballooning hepatocytes mild to moderate chronic inflammation, grade3: panacinar steatosis, ballooning and disarray obvious and mild or portal mild to moderate inflammation, stage1: perivenular and/or perisinusoidal fibrosis in zone3, stage2: combined pericellular portal fibrosis, stage3: septal/bridging fibrosis, stage4: cirrhosis. The degree of fat deposition was evaluated by amount of fat droplets as observed under the microscope as follows; 0: <5%, 1: 5–<10%, 2: 10–<34%, 3: 34–<67%, 4: >67%. The degree of iron deposition was categorized by the presence of granules of free iron observed under the microscope as follows; 0: absence by x400, 1: easily identifiable by x400 and rarely identifiable by x250, 2: identifiable by x100, 3: identifiable by x25, 4: identifiable at lower than x25.

Inclusion criteria for NAFLD patients were as follows; (i) no history of alcoholism, (ii) no history for HBV/HCV/HIV infection, (iii) diagnosed by liver biopsy, (iv) information regarding age and BMI available. The sex of two samples was unknown, and was imputed from the results of the genome scan. As general Japanese population controls, the genome scan results of 942 healthy Japanese volunteers from Aichi Cancer Center Hospital and Research Institute were used [22].

Anthropometric and Laboratory Evaluation

We employed conventional methods for the measurement of anthropometry (height, weight, amount of visceral fat and abdominal circumscript). BMI was calculated from the measurements. The following biochemical/hematological/immunological traits were also measured by conventional methods; aspartate aminotransferase (AST), alanine aminotransferase (ALT), γ-glutamyl transpeptidase (GGT), albumin, total bilirubin, cholinesterase, type IV collagen 7S, hyaluronic acid, triglyceride, total cholesterol, hemoglobin A1c (HbA1c), fasting immunoreactive insulin (IRI), fasting plasma glucose (FBS), high sensitive CRP (hs-CRP), adiponectin, leptin, ferritin, uric acid, and platelet (PLT) count. Anti nuclear antibody (ANA) was measured by ELISA and categorized by the detection limit in a serial dilution as follows; 0: <40x, 1: 40–80x, 2: 81–160x, 3: 160x, 4: >320x. Homeostasis model assessment-insulin resistance (HOMA-IR) was calculated from the measurements. Patients were assigned a diagnosis of diabetes mellitus (DM) when they had documented use of oral hypoglycemic medication, a random glucose level >200 mg/dl, or FPG >126 mg/dl. Hyperlipidemia was diagnosed with the cholesterol level being >200 mg/dl and/or triglyceride level being >160 mg/dl. Hypertension was diagnosed when the patient was taking antihypertensive medication and/or had a resting recumbent blood pressure ≥140/90 mmHg on at least two occasions.

DNA Preparation

Genomic DNA was extracted from peripheral blood mononuclear cells by standard phenol-chloroform extraction and resuspended in TE buffer. DNA concentration and purity were measured with Nanodrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). The samples were stored at −20°C until use.

Genome-wide Genotyping and Quality Control

Genome scan was conducted for 543 patients with NAFLD and 942 healthy subjects using Illumina Human 610-Quad Bead Chip on a Bead Station 500G Genotyping System (Illumina, Inc., San Diego, CA, USA) and subjected to the following quality controls. Initially, ten patients and six control subjects were removed due to low call rates (<0.99). Regarding the SNP markers, 85,472 SNPs with minor allele frequency of smaller than 0.01 in either case or control group, 6,479 SNPs with lower success rates (<0.98) and 35 SNPs with distorted Hardy-Weinberg equilibrium ($p<10^{-7}$) were removed, resulting in 484,751 SNP markers being used for analysis. Principal component analysis by EIGENSOFT [17] including phase II HapMap (http://hapmap.ncbi.nlm.nih.gov/) samples identified no samples that were deviated from the Japanese population. Subsequently, the degree of kinship between individuals was examined by pi-hat in PLINK 1.07 (http://pngu.mgh.harvard.edu/purcell/plink/) [23]. Of the eight pairs of samples (four case pairs and four control pairs) showing high degrees of kinship (PI-HAT>0.4), the sample with the lower call rate in each pair was removed. After these steps, 529 case and 932 controls were used for the analysis.

Statistical Analysis

A case/control association analysis was performed by exact trend test between NAFLD patients and control subjects [24]. The correction of obtained *p*-values for population stratification was performed using EIGENSTRAT [17]. In addition, an association between Matteoni classification (type1 to type4) and additive model of genotype for each SNP was examined using Jonckheere-Terpstra test for NAFLD patients. Assessment of population stratification of inflation of *p*-value was carried out by the genomic control method for asymptotic trend test [25]. Association between each quantitative trait and the genotype of significant SNPs in NAFLD patients were calculated by multivariable linear regression or multivariable ordinal regression adjusted for age, sex and BMI. Each quantitative trait was transformed as follows; natural log for ALT, AST, HOMA-IR, HbA1c, IRI, triglyceride, total bilirubin, adiponectin, hs-CRP, hyaluronic acid, leptin, reciprocal number for albumin, cholinesterase, type IV collagen 7S and square root for uric acid and ferritin. The values of FPG, PLT, total cholesterol, amount of visceral fat, and abdominal circumscript were not transformed. For each trait, values that were within only 4 S.D. were included for analysis. LD indices were calculated by default setting of Haploview [26] and the LD block was defined manually.

Table 4. Replication study of previously reported SNPs.

			Genotyping Result and Allele Frequency of A2					Statistics		
			NAFLD					NAFLD vs. Control		Matteoni
dbSNPID	A1/A2	Gene	Control	Type 1	Type 2	Type 3	Type 4	*p*-value†	OR (95%CI)	*p*-value‡
rs12137855	C*/T	*LYPLAL1*	828/102/2	90/10/0	67/6/0	24/5/0	294/33/0	0.55	0.89	0.98
			(0.056)	(0.050)	(0.041)	(0.086)	(0.050)		(0.64–1.25)	
rs780094	T*/C	*GCKR*	321/433/178	34/54/12	28/34/11	17/11/1	133/139/55	0.011	0.82	0.92
			(0.423)	(0.390)	(0.383)	(0.224)	(0.381)		(0.70–0.95)	
rs4240624	G/A	*PPP1R3B*	–	–	–	–	–	–	–	–
rs2228603	C/T*	*NCAN*	842/88/2	93/7/0	65/8/0	28/1/0	292/31/4	0.80	1.05	0.58
			(0.049)	(0.035)	(0.054)	(0.017)	(0.059)		(0.75–1.48)	

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.
doi:10.1371/journal.pone.0038322.t004

Supporting Information

Figure S1 QQ plot of the GWA study comparing distribution of the observed and expected *p*-values. Upper box is expressed in antilog scale and the lower box is expressed in $-\log_{10}$ scale. The X- and Y-axis correspond to expected and observed *p*-values. Blue and red dots denote before and after correction by genomic control method ($\lambda = 1.04$), respectively.
(DOC)

Table S1 List of the SNPs showing $p<1.0\times10^{-5}$ in the GWA study. 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.
(DOC)

Acknowledgments

The authors would like to thank Yutaka Kohgo, Hirofumi Uto and Tetsuo Takehara for sample collection and Hisako Imamura and Hiroyuki Uneme for data management.

Author Contributions

Conceived and designed the experiments: FM TO. Performed the experiments: MT M. Kokubo. Analyzed the data: TK RY FM. Contributed reagents/materials/analysis tools: TK YS AU KM MT TT KY T. Saibara EH M. Kokubo SW SK YI M. Kawanaka T. Shima HP HT KT RY. Wrote the paper: TK MT RY FM TO.

References

1. Ludwig J, Viggiano TR, McGill DB, Oh BJ (1980) Nonalcoholic steatohepatitis: Mayo Clinic experiences with a hitherto unnamed disease. Mayo Clin Proc 55: 434–438.
2. Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, et al. (1999) Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity. Gastroenterology 116: 1413–1419.
3. Cohen JC, Horton JD, Hobbs HH (2011) Human fatty liver disease: old questions and new insights. Science 332: 1519–1523. doi:10.1126/science.1204265.
4. Vernon G, Baranova A, Younossi ZM (2011) Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. Aliment Pharmacol Ther 34: 274–285. doi:10.1111/j.1365–2036.2011.04724.x.
5. Okanoue T, Umemura A, Yasui K, Itoh Y (2011) Nonalcoholic fatty liver disease and nonalcoholic steatohepatitis in Japan. J Gastroenterol Hepatol 26 Suppl 1: 153–162. doi:10.1111/j.1440–1746.2010.06547.x.
6. Williams CD, Stengel J, Asike MI, Torres DM, Shaw J, et al. (2011) Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study. Gastroenterology 140: 124–131. doi:10.1053/j.gastro.2010.09.038.
7. Okanoue T (2011) Recent progress in the research of NASH/NAFLD in Japan. Nihon Shokakibyo Gakkai Zasshi 108: 1161–1169.
8. Berson A, De Beco V, Lettéron P, Robin MA, Moreau C, et al. (1998) Steatohepatitis-inducing drugs cause mitochondrial dysfunction and lipid peroxidation in rat hepatocytes. Gastroenterology 114: 764–774.
9. Day CP (2006) From fat to inflammation. Gastroenterology 130: 207–210. doi:10.1053/j.gastro.2005.11.017.
10. Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, et al. (2008) Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 40: 1461–1465. doi:10.1038/ng.257.
11. Sookoian S, Castaño GO, Burgueño AL, Gianotti TF, Rosselli MS, et al. (2009) A nonsynonymous gene variant in the adiponutrin gene is associated with nonalcoholic fatty liver disease severity. J Lipid Res 50: 2111–2116. doi:10.1194/jlr.P900013-JLR200.
12. Hotta K, Yoneda M, Hyogo H, Ochi H, Mizusawa S, et al. (2010) Association of the rs738409 polymorphism in PNPLA3 with liver damage and the development of nonalcoholic fatty liver disease. BMC Med Genet 11: 172. doi:10.1186/1471–2350–11–172.
13. Sookoian S, Pirola CJ (2011) Meta-analysis of the influence of I148M variant of patatin-like phospholipase domain containing 3 gene (PNPLA3) on the susceptibility and histological severity of nonalcoholic fatty liver disease. Hepatology 53: 1883–1894. doi:10.1002/hep.24283.
14. Speliotes EK, Yerges-Armstrong LM, Wu J, Hernaez R, Kim LJ, et al. (2011) Genome-wide association analysis identifies variants associated with nonalcoholic Fatty liver disease that have distinct effects on metabolic traits. PLoS Genet 7: e1001324. doi:10.1371/journal.pgen.1001324.
15. Petersen KF, Dufour S, Hariri A, Nelson-Williams C, Foo JN, et al. (2010) Apolipoprotein C3 gene variants in nonalcoholic fatty liver disease. N Engl J Med 362: 1082–1089. doi:10.1056/NEJMoa0907295.
16. Terao C, Yamada R, Ohmura K, Takahashi M, Kawaguchi T, et al. (2011) The human AIRE gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. Human Molecular Genetics 20: 2680–2685. doi:10.1093/hmg/ddr161.
17. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904–909. doi:10.1038/ng1847.
18. Yasui K, Hashimoto E, Komorizono Y, Koike K, Arii S, et al. (2011) Characteristics of patients with nonalcoholic steatohepatitis who develop hepatocellular carcinoma. Clin Gastroenterol Hepatol 9: 428–433; quiz e50. doi:10.1016/j.cgh.2011.01.023.

19. He S, McPhaul C, Li JZ, Garuti R, Kinch L, et al. (2010) A Sequence Variation (I148M) in PNPLA3 Associated with Nonalcoholic Fatty Liver Disease Disrupts Triglyceride Hydrolysis. J Biol Chem 285: 6706–6715. doi:10.1074/jbc.M109.064501.
20. Ueno T, Inuzuka S, Torimura T, Tamaki S, Koh H, et al. (1993) Serum hyaluronate reflects hepatic sinusoidal capillarization. Gastroenterology 105: 475–481.
21. Brunt EM, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR (1999) Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. Am J Gastroenterol 94: 2467–2474. doi:10.1111/j.1572–0241.1999.01377.x.
22. Suzuki T, Matsuo K, Sawaki A, Mizuno N, Hiraki A, et al. (2008) Alcohol Drinking and One-Carbon Metabolism-Related Gene Polymorphisms on Pancreatic Cancer Risk. Cancer Epidemiology Biomarkers & Prevention 17: 2742–2747. doi:10.1158/1055–9965.EPI-08–0470.
23. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet 81: 559–575.
24. Yamada R, Okada Y (2009) An optimal dose-effect mode trend test for SNP genotype tables. Genet Epidemiol 33: 114–127. doi:10.1002/gepi.20362.
25. Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55: 997–1004.
26. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263–265. doi:10.1093/bioinformatics/bth457.

nature genetics

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 Ohmura[5], Akari Suzuki[1], Takahisa Kawaguchi[4], Eli A Stahl[7,8], Fina A S Kurreeman[7–9], Nao Nishida[10], Hiroko Ohmiya[3], Keiko Myouzen[1], Meiko Takahashi[4], Tetsuji Sawada[11], Yuichi Nishioka[12], Masao Yukioka[13], Tsukasa Matsubara[14], Shigeyuki Wakitani[15], Ryota Teshima[16], Shigeto Tohma[17], Kiyoshi Takasugi[18], Kota Shimada[17], Akira Murasawa[19], Shigeru Honjo[20], Keitaro Matsuo[21], Hideo Tanaka[21], Kazuo Tajima[22], Taku Suzuki[6,23], Takuji Iwamoto[6,23], Yoshiya Kawamura[24], Hisashi Tanii[25], Yuji Okazaki[26], Tsukasa Sasaki[27], Peter K Gregersen[28], Leonid Padyukov[29], Jane Worthington[30], Katherine A Siminovitch[31], Mark Lathrop[32,33], Atsuo Taniguchi[6], Atsushi Takahashi[3], Katsushi Tokunaga[10], Michiaki Kubo[34], Yusuke Nakamura[35], Naoyuki Kamatani[36], Tsuneyo Mimori[5], Robert M Plenge[7,8], Hisashi Yamanaka[6], 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 populations[1–12] and by several meta-analyses of the original GWAS[13–16]. 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 populations[13–16] 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 ancestry[15].

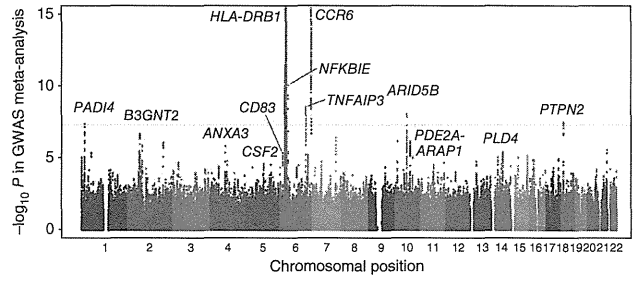

Figure 1 Manhattan plots of the GWAS meta-analysis for rheumatoid arthritis in the Japanese population. The genetic loci that satisfied the genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ (gray line) in the meta-analysis or in the combined study of the meta-analysis and the replication 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

					Associations in Japanese								Associations in Europeans[c]			
					GWAS meta-analysis			Replication study		Combined study			GWAS meta-analysis			
					Allele	Allele 1 freq.							Allele 1 Freq.			
rsID[a]	Chr.	Position (bp)	Cytoband	Gene(s)	1/2	RA	Control	OR (95% CI)[b]	*P*	OR (95% CI)[b]	*P*	OR (95% CI)[b]	*P*	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	*B3GNT2*	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-ARAP1*	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 ($5.0 \times 10^{-8} \leq P < 5.0 \times 10^{-6}$ 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.
[a]SNPs 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**. [b]Odds ratio of allele 1. [c]Associations in the previous meta-analysis in European populations[15].

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 University[12] 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

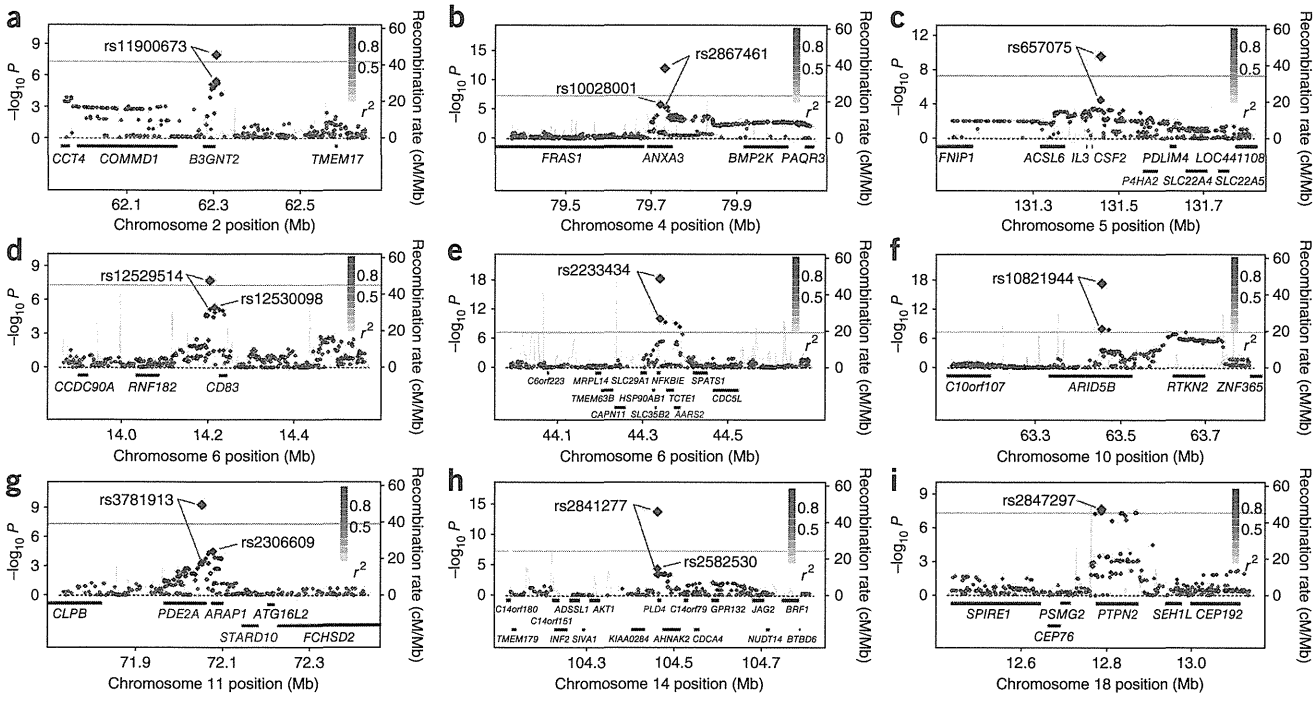

Figure 2 Regional plots of the loci newly associated with rheumatoid arthritis at the genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ in the combined study of the meta-analysis and the replication study. (**a–i**) Regional plots are shown at *B3GNT2* (**a**), *ANXA3* (**b**), *CSF2* (**c**), *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 cases[10] (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 controls[15] (**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