in individuals of reproductive age and decreases in adolescence and old age.[16] While some European studies have reported differences in the age at onset between males and females,[17] a relatively large study involving 1790 cases from China did not detect a significant difference in the mean onset age between the sexes.[18] Thus, it is unclear whether there is no difference in the onset age of SLE between males and females or whether such differences are observed only in patients of European descent.

Signs and clinical markers of SLE

SLE produces a wide range of clinical signs, including physical signs and laboratory findings. Various reports have detected associations between the clinical features of SLE and age/sex, either at disease onset or throughout the clinical course of the condition. Thus, these reports suggested that age and/or sex can affect the signs of SLE. Efforts have been made to identify subgroups of SLE based on clinical manifestations.[19,20] However, limited power of previous reports made it difficult to draw conclusions. A detailed analysis of the clinical features of SLE in a large-scale study would increase our understanding of the clinical heterogeneity of SLE.

Here, we performed a nationwide surveillance study of patients with SLE in Japan to characterize the epidemiological and clinical features of SLE. As far as we know, this is the largest such study to have ever been conducted.

Patients and methods

SLE patient registration

In Japan, a total of 56 diseases are defined as "Nanbyo (intractable disease)" and patients are given a questionnaire about their clinical status and history, which is filled out by the clinician providing their care, during registration. The clinicians are not limited to specialists for the diseases. The registered information is used for making decisions by experts on the public financial support provided for their medical care. Each patient is enrolled as a new registrant in the first year after diagnosis, and his or her registration is renewed annually by different forms from the first ones (follow-up registry). SLE is one of these "Nanbyo." This registry-based financial support system is well known throughout the country, and Japanese public health departments and health care professionals believe that the vast majority of patients with the diseases that receive medical care are registered annually. Clinical information in the questionnaire for the SLE forms is listed in Supplementary Table 1.

We obtained text files electronically converted from nationwide registry data about SLE in Japan from 2003 to 2010.[14] Although the text files did not cover all the registrants, in total, 14,779 new registries were obtained from 2003 to 2010 and we adopted 2009 (44,249 patients), which covered the largest parts of the annual total registries (81.2%) as a year with representative follow-up data after we found that each year's follow-up registries displayed similar basic statistics. For new registries, we omitted suspected duplicate registries and identified 14,030 registrants as novel for the purposes of this study. We extracted 9374 registries for which information about disease onset was available and for which it could be confirmed that disease onset had occurred within the last year. Schematic images of quality control of the dataset were illustrated in Supplementary Figure 1. We evaluated two patient groups; the first group, which was collected from 2003 to 2010, consisted of patients who had been diagnosed with SLE within the last year, and the other group consisted of all patients in the representative year, 2009. We called these two groups the "novel SLE" and "all SLE" groups, respectively.

Clinical information

We extracted information about the patients' clinical features including the 11 major signs included in the American College of Rheumatology (ACR) classification,[5] age, sex, age at diagnosis, and complications (infection, bone necrosis, compression fracture of bone, gastric ulcers, myocardial infarction and cerebral infarction) from the registry for all registrants. Some items, including information about antinuclear antibody (ANA) positivity, anti-Smith (anti-Sm) antibody positivity, anti-double-stranded DNA (anti-dsDNA) antibody positivity, the occurrence of biological false-positives on the syphilis test, lupus anti-coagulant positivity and anticardiolipin antibody positivity, were available only for the novel group (Supplementary Table 1).

Sex ratio

The female:male ratio was estimated in the all SLE group.

Age distribution of SLE patients

Age at onset was compared between males and females in the novel SLE and all SLE groups.

Downloaded from lup.sagepub.com at Kyoto University on January 14, 2015

The significance of the difference was tested by logistic regression analysis.

Analysis of SLE signs and clinical markers in patients with SLE

The frequencies of SLE signs and clinical markers were analyzed in the novel and all SLE groups. The effects of age, sex and disease duration were assessed separately and in combination by multiple logistic regression analysis. Clustering of the major signs and patients was performed in 6637 patients in the novel SLE group for whom data regarding the 11 major signs and clinical markers were available and 10,000 randomly selected patients in the all SLE group for whom data regarding the 10 major signs and clinical markers other than ANA were available (Supplementary Figure 1). The associations between complications and the patients' basic information, SLE signs and clinical markers were also analyzed. We regarded autoantibody positivity at any point during the disease course as positivity.

Statistical analysis

Statistical analyses were performed using the R or SPSS (ver18) software.

Results

Female ratio of SLE

The female:male ratio was 8.14 in the all SLE group and was comparable to those described in previous reports (8.1–12.5).[21–23] A comparison of the age distributions of the male and female SLE patients in the all SLE group showed that the females were younger than the males ($p = 0.00031$, Figure 1(a)). The females were also younger at onset than the males ($p = 4.1 \times 10^{-62}$, Figure 1(b)).

Prevalence of clinical features and the effects of age and sex on them in the all SLE group

The prevalence of the 10 major signs of SLE (as outlined by the ACR, except for ANA positivity) varied (Figure 2(a), Supplementary Table 2). Cytopenia and arthritis were the two most common signs, and serositis was the least common sign. The frequencies of some of the 10 SLE signs differed markedly between males and females (Figure 2(b)). An analysis of the effects of age on the frequencies of these signs revealed four patterns: increases with age, decreases with age, a U-shaped age distribution (lowest in middle aged subjects), and an inverse-U shaped age distribution (highest in middle-aged subjects) (Figure 2(c)). An analysis of the effects of disease duration on the frequencies of these signs revealed that most of them were frequently observed in the short duration after onset. The signs' disease duration-based frequency patterns were similar to their age-based patterns. Photosensitivity was the only sign associated with a long disease duration (Figure 2(d)). Discoid eczema was the only sign that was not associated with disease duration. The detailed results are shown in Supplementary Figure 2 and Supplementary Table 3, and further analyses of the detailed signs of SLE are shown in the Supplementary notes and Supplementary Figure 3.

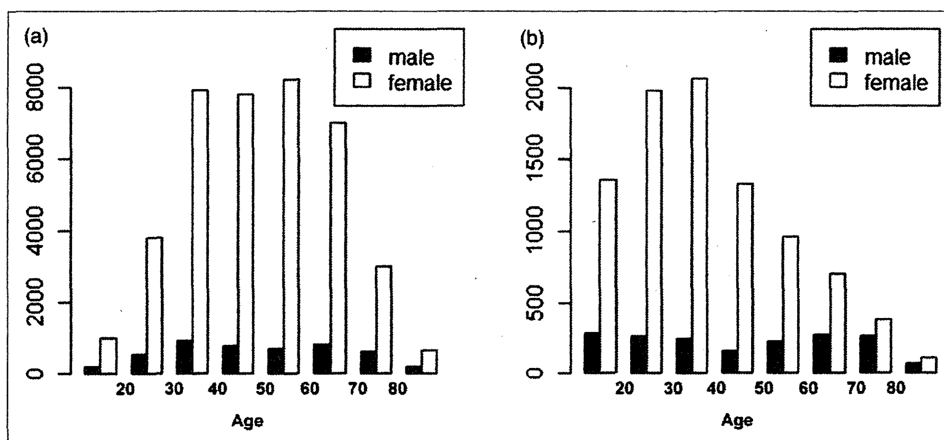

Figure 1 Distribution of patients who developed systemic lupus erythematosus (SLE). (a) Distribution of the current ages of the SLE patients. (b) Distribution of the age at onset of the SLE patients.

Downloaded from lup.sagepub.com at Kyoto University on January 14, 2015

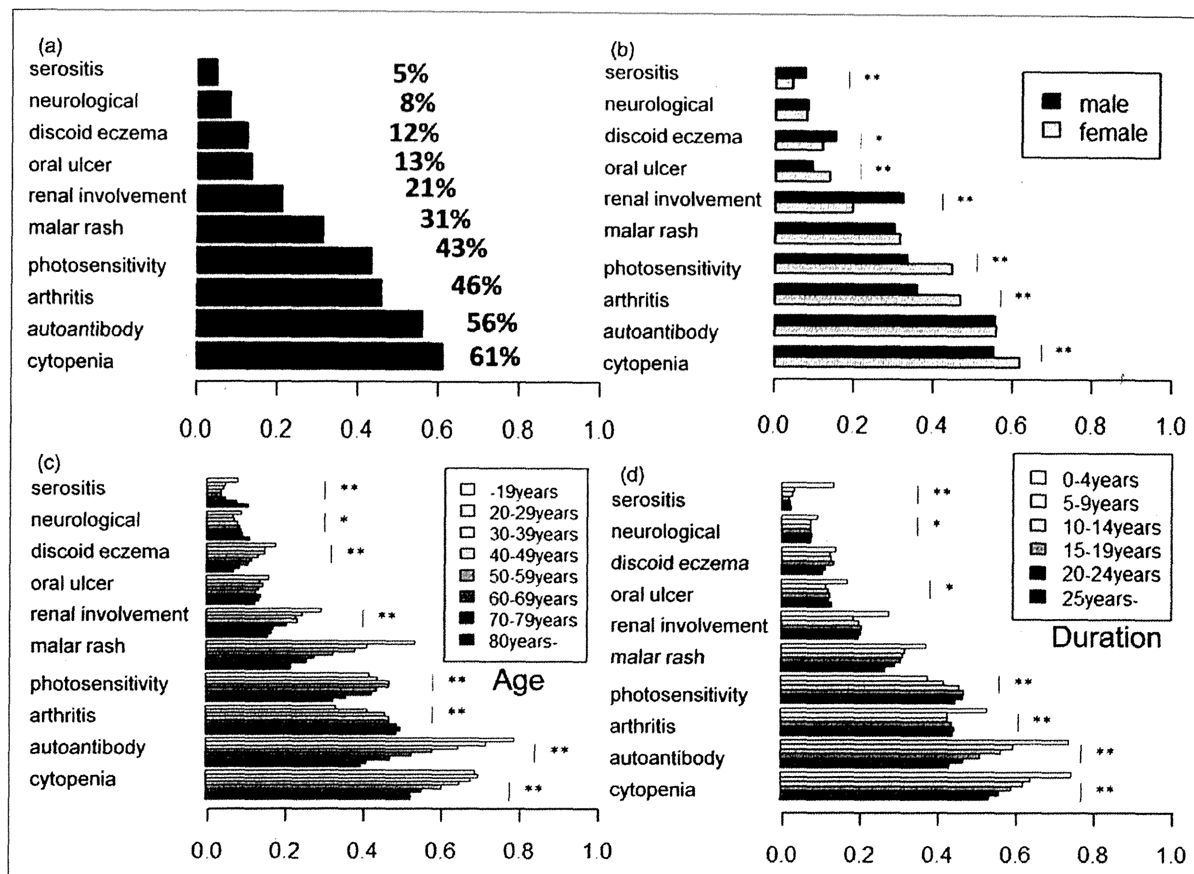

Figure 2 Distribution and clusters of systemic lupus erythematosus (SLE) signs and patients in the all SLE group. (a) Frequency of SLE signs in a year. Frequencies of SLE signs according to sex (b), age (c) and disease duration (d). *p value $<10^{-5}$, **p value $<10^{-10}$.

Prevalence of signs and clinical markers and the effects of age and sex on them in the novel SLE group

The prevalence of the major signs of SLE also varied in the novel SLE patients, and the order of the signs' frequencies (i.e. from highest to lowest) was different from that observed in the all SLE group (Figure 3(a) and Supplementary Table 2). Except for cytopenia, all of the SLE major signs were affected by sex in the same manner as was observed in the all SLE group according to multiple logistic regression analysis (Figures 2(b) and 3(b) and Supplementary Figure 4(a)). The associations between age and the SLE signs differed between the novel and all SLE groups for four of the 10 items (Figures 2(c) and 3(c) and Supplementary Figure 4). Two patterns of difference were observed. The first type involved a positive association with age being observed only in the novel SLE group. The other type involved a positive association with age not being observed in the novel SLE group. Oral ulcers exhibited the former pattern ($p = 3.9 \times 10^{-6}$), and renal involvement, cytopenia and arthritis displayed the latter pattern ($p > 0.019$). Sex-specific age associations showed a third pattern: opposite associations in the novel and all SLE groups. Namely, cytopenia was associated with old age in the males belonging to the novel SLE group, while it was associated with young age in the males in the all SLE group. In addition, three signs showed specific associations with age in the novel SLE group. The frequency of serositis increased age-dependently, whereas the frequencies of renal involvement and arthritis showed U and inverse-U patterns, respectively. The detailed results of the analyses and further analyses are shown in the Supplementary notes and Supplementary Table 4.

Clustering analysis of the coexistence of signs and clinical markers in the all SLE and novel SLE groups

Clustering analysis of the 11 signs in the patients in the novel SLE group revealed that they could be

Downloaded from lup.sagepub.com at Kyoto University on January 14, 2015

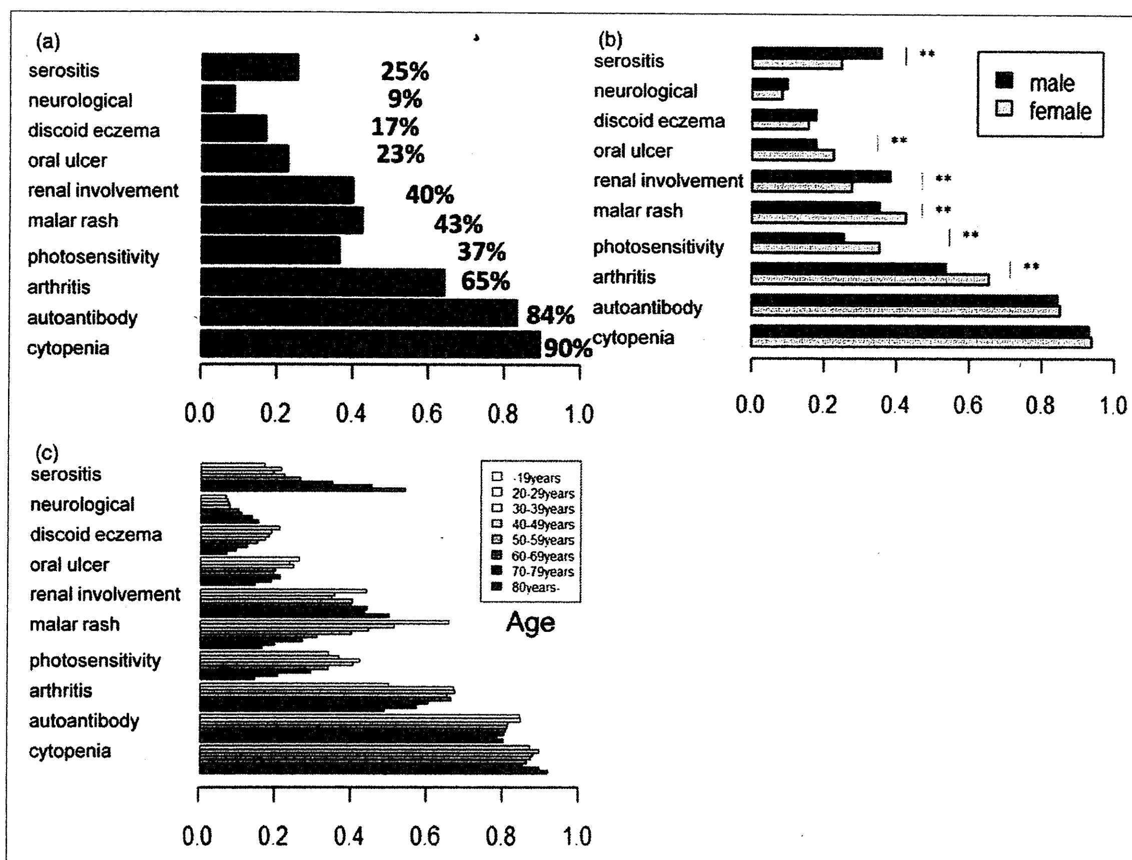

Figure 3 Distribution and clusters of systemic lupus erythematosus (SLE) signs and patients in the novel SLE group. (a) Frequencies of SLE signs during the first year after diagnosis. Frequency of SLE signs within a year of diagnosis according to sex (b) and age (c) based on multiple logistic linear regression analysis. *p value $<10^{-5}$, **p value $<10^{-10}$.

divided into two groups; namely, a group containing autoantibody positivity, ANA positivity, cytopenia and arthritis, and another group including the other seven signs and markers (Figure 4(a)). The novel SLE patients (6637) were also subjected to clustering analysis, which showed that they could be classified into 10 clusters according to their signs (Figure 4(b)). The sign frequencies and the numbers of SLE patients in each cluster are shown in Supplementary Table 5.

Cluster analysis of the 10 major SLE signs (not including ANA) in the all SLE group showed that they could be subgrouped into two clusters with the similar characteristics as those observed in the analysis of the novel SLE group although differences were observed among the finer cluster divisions (Figure 4(c)). Cluster analysis of 10,000 randomly selected SLE patients from the all SLE group produced eight clear clusters (Figure 4(d) and Supplementary Table 6). The patterns of clusters partly matched those observed in the novel SLE group.

Further analyses: complications of SLE and the distributions of specific autoantibodies

The complications of SLE were also assessed in the all SLE group, as were the effects of age, sex and disease duration. The associations of autoantibodies with complications were assessed according to age, sex and disease duration to assess their utility as predictive markers. The associations between complications and each SLE patient cluster were also analyzed. See the Supplementary notes for details.

Discussion

Although some small studies did not report a significant difference in age at onset between the sexes,[18] our large-scale study demonstrated that female patients developed SLE at a younger age than male patients. We evaluated the clinical features of two patient populations, "the novel

Downloaded from lup.sagepub.com at Kyoto University on January 14, 2015

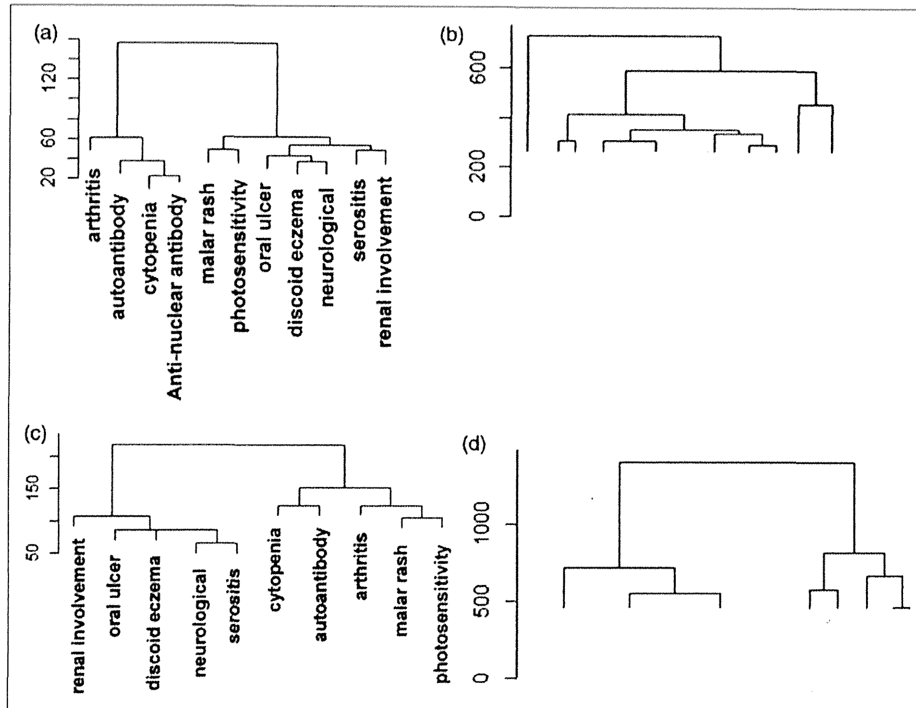

Figure 4 Clusters of systemic lupus erythematosus (SLE) signs and patients. (a) Clustering of 11 SLE major signs in patients who had been diagnosed with SLE within the last year. (b) Clustering of 6637 SLE patients who had been diagnosed with SLE within the last year. (c) Clustering of 10 major SLE signs in the all SLE group. (d) Clustering of 10,000 SLE patients in the all SLE group.

patients": i.e. patients who had been diagnosed with SLE within the last year, and "all patients": i.e. all patients regardless of their disease duration. As a result, we obtained evidence of associations between SLE signs and age, sex and disease duration. In our study, the frequencies of 11 major signs were similar to those obtained in previous reports from Asian and European countries both in the novel SLE and all SLE groups with the exception of serositis (25.3% in the novel SLE group, 4.6% in the all SLE group; 5%–22% at onset and 20%–40% prevalence in previous studies).[18,23–27] This difference might have been due to the relative difficulty of detecting serositis compared with other features.

We validated previous reports of higher frequencies of serositis,[28,29] renal involvement[29,30] and discoid eczema[29–31] in males and higher frequencies of photosensitivity[30] and oral ulcers[30] in females. Although neurological involvement was reported to be more common in males in two previous reports,[32,33] our study did not find any difference between the sexes. The difference between the sexes in the frequency of malar rash is disputed, and our study did not detect any sex difference. Our results indicate that any inter-sex difference in neurological involvement and malar rash is very small. The sex difference in the frequency of arthritis is also disputed, and we observed a significantly higher frequency of arthritis in females (47.0% in females and 36.0% in males with $p = 1.3 \times 10^{-44}$).

Only a few previous studies comprising more than 500 patients have examined the effects of age on the clinical manifestations of SLE.[18,28,34,35] Previous studies reported positive associations of younger age with malar rash, discoid eczema, autoantibody production and photosensitivity,[18,30] and we confirmed these associations. In addition, we demonstrated that serositis and neurological involvement were positively associated with older age. Renal involvement was associated with younger age only in the novel SLE group.

No studies have ever analyzed the detailed effects of disease duration on SLE signs. Most of the major signs and clinical markers of SLE, especially serositis, displayed higher prevalence in the patients with short disease durations. Only the prevalence of photosensitivity increased according to disease duration. Discoid eczema was not associated with disease duration.

Downloaded from lup.sagepub.com at Kyoto University on January 14, 2015

We performed similar analyses for more detailed signs of SLE (Supplementary notes).

The 11 SLE signs were classified into two groups according to their manifestation patterns in the novel SLE group: group 1 (ANA, autoantibody positivity (anti-Sm antibody and anti-dsDNA antibody), cytopenia, and arthritis) and group 2 (malar rash, discoid eczema, photosensitivity, oral ulcers, neurological involvement, serositis and renal involvement). The first group included hematoserological abnormalities such as cytopenia and arthritis was considered to be an inflammatory/autoimmunity-related reaction and so was classified with the hematoserological abnormalities because of its reduced organ specificity compared to the items in group 2. Therefore, we called group 1 the hematoserological group and group 2 the organ-specific group. In the all SLE group, such clear clustering was not very apparent, which might have been because individual patients tended to present with various features during their clinical courses.

The SLE patients in the novel SLE group were clustered into 10 groups according to the signs that they displayed. These groups were not associated with sex or age (analysis of variance (ANOVA), data not shown). At onset, the frequencies of the 10 groups ranged from 4.0% to 22.4%. The 10 groups were characterized as: represented by (1) neurological involvement (22.4%), (2) discoid eczema (10.6%), (3) a lack of autoantibodies other than ANA (12.7%), (4) oral ulcers (9.1%), (5) renal involvement (9.9%), (6) photosensitivity (5.7%), (7) a lack of arthritis (6.5%), (8) serositis (9.8%), (9) malar rash (4.0%) and (10) others (9.3%). It should be noted that each group was represented by one of the items in the organ-specific group or a lack of an item in the hematoserological group. These findings suggest that the items in the organ-specific group are the predominant determinants of a patient's condition. In the all SLE group, eight clusters, which displayed frequencies ranging from 3.9% to 31.5%, were determined. The clusters were characterized as follows: 1) no signs or markers (6.9%), 2) cytopenia alone (5.1%), 3) autoantibody positivity alone (3.9%), 4) cytopenia and autoantibody positivity only (5.9%), 5) arthritis (9.1%), 6) renal involvement (16.4%), 7) neurological signs and serositis (21.3%) and 8) others (31.5%). The novel SLE and all SLE groups shared two clusters with similar characteristics, i.e. the "neurological signs" and "renal involvement" clusters. The reduced frequencies of signs and clinical markers observed in the all SLE group led to clusters based on one or no signs. The lack of information about ANA during the chronic phase might also have reduced the number of clusters. We performed five rounds of resampling, each of which involved 10,000 patients, and the same clusters were maintained (data not shown). These results confirm that SLE patients and signs can be subgrouped into clear clusters. However, the 11 or 10 signs of SLE could not consistently explain the division of clusters among different stages of the disease. This raised the possibility that underlying factors related to the pathology of SLE other than the 11 signs exist. While we analyzed the associations between clusters and clinical signs or complications, we could not analyze the association between clusters and death because of a lack of information. Although the follow-up questionnaire included information about death causes (data not shown), this information was not filled out in most cases. This could be explained by the system of the nationwide study in which patients ask physicians to fill out the questionnaire. Associations between clusters in all SLE group and some complications (Supplementary notes) suggest the possibility that clusters are associated with severity and prognosis of SLE. Further follow-up studies would clarify the clinical characteristics of the abovementioned clusters.

Finally, we would like to comment on our data source. As the primary purpose of the national registry is to determine whether patients qualify for public financial aid, there could be a bias toward the over-rating of the signs. Despite our concern about such overestimation, the frequencies of individual signs in our study were similar to those described in previous reports from Asian countries,[36] indicating that any over-rating was not too problematic. Considering the number of subjects analyzed in the current study and the fact that the same tendencies were observed during each year (data not shown), our results regarding the patterns of signs and the associations between these signs and gender, age and disease duration in Japanese SLE patients should be regarded as conclusive.

In conclusion, we have obtained conclusive evidence about the distributions of the clinical features of SLE and their relationships with sex, age and age at onset.

Funding

This work was supported by grants-in-aid from the Ministry of Health, Labor, and Welfare of Japan.

Downloaded from lup.sagepub.com at Kyoto University on January 14, 2015

Conflict of interest statement

The authors have no conflicts of interest to declare.

References

1 Tsokos GC. Systemic lupus erythematosus. *N Engl J Med* 2011; 365: 2110–2121.
2 Bengtsson C, Ohman ML, Nived O, Rantapää Dahlqvist S. Cardiovascular event in systemic lupus erythematosus in northern Sweden: Incidence and predictors in a 7-year follow-up study. *Lupus* 2012; 21: 452–459.
3 Murray SG, Yazdany J, Kaiser R, *et al.* Cardiovascular disease and cognitive dysfunction in systemic lupus erythematosus. *Arthritis Care Res (Hoboken)* 2012; 64: 1328–1333.
4 Chiu CC, Huang CC, Chan WL, *et al.* Increased risk of ischemic stroke in patients with systemic lupus erythematosus: A nationwide population-based study. *Intern Med* 2012; 51: 17–21.
5 Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997; 40: 1725.
6 Petri M, Orbai AM, Alarcón GS, *et al.* Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. *Arthritis Rheum* 2012; 64: 2677–2686.
7 Ward MM. Prevalence of physician-diagnosed systemic lupus erythematosus in the United States: Results from the third national health and nutrition examination survey. *J Womens Health (Larchmt)* 2004; 13: 713–718.
8 Amor B, Bouchet H, Delrieu F. National survey on reactive arthritis by the French Society of Rheumatology [article in French]. *Rev Rhum Mal Osteoartic* 1983; 50: 733–743.
9 Gudmundsson S, Steinsson K. Systemic lupus erythematosus in Iceland 1975 through 1984. A nationwide epidemiological study in an unselected population. *J Rheumatol* 1990; 17: 1162–1167.
10 Somers EC, Thomas SL, Smeeth L, Schoonen WM, Hall AJ. Incidence of systemic lupus erythematosus in the United Kingdom, 1990–1999. *Arthritis Rheum* 2007; 57: 612–618.
11 Helve T. Prevalence and mortality rates of systemic lupus erythematosus and causes of death in SLE patients in Finland. *Scand J Rheumatol* 1985; 14: 43–46.
12 Hiraki LT, Feldman CH, Liu J, *et al.* Prevalence, incidence, and demographics of systemic lupus erythematosus and lupus nephritis from 2000 to 2004 among children in the US Medicaid beneficiary population. *Arthritis Rheum* 2012; 64: 2669–2676.
13 Chiu YM, Lai CH. Nationwide population-based epidemiologic study of systemic lupus erythematosus in Taiwan. *Lupus* 2010; 19: 1250–1255.
14 Ohta A, Nagai M, Nishina M, Tomimitsu H, Kohsaka H. Age at onset and gender distribution of systemic lupus erythematosus, polymyositis/dermatomyositis, and systemic sclerosis in Japan. *Mod Rheumatol* 2013; 23: 759–764.
15 Pons-Estel GJ, Alarcón GS, Scofield L, Reinlib L, Cooper GS. Understanding the epidemiology and progression of systemic lupus erythematosus. *Semin Arthritis Rheum* 2010; 39: 257–268.
16 Lahita RG. The role of sex hormones in systemic lupus erythematosus. *Curr Opin Rheumatol* 1999; 11: 352–356.
17 Lu LJ, Wallace DJ, Ishimori ML, Scofield RH, Weisman MH. Review: Male systemic lupus erythematosus: A review of sex disparities in this disease. *Lupus* 2010; 19: 119–129.
18 Feng JB, Ni JD, Yao X, *et al.* Gender and age influence on clinical and laboratory features in Chinese patients with systemic lupus erythematosus: 1,790 cases. *Rheumatol Int* 2010; 30: 1017–1023.
19 Levy DM, Peschken CA, Tucker LB, *et al.* Influence of ethnicity on childhood-onset systemic lupus erythematosus: Results from a multiethnic multicenter Canadian cohort. *Arthritis Care Res (Hoboken)* 2013; 65: 152–160.
20 Jacobsen S, Petersen J, Ullman S, *et al.* A multicentre study of 513 Danish patients with systemic lupus erythematosus. I. Disease manifestations and analyses of clinical subsets. *Clin Rheumatol* 1998; 17: 468–477.
21 Pons-Estel BA, Catoggio LJ, Cardiel MH, *et al.* The GLADEL multinational Latin American prospective inception cohort of 1,214 patients with systemic lupus erythematosus: Ethnic and disease heterogeneity among "Hispanics". *Medicine (Baltimore)* 2004; 83: 1–17.
22 Ballou SP, Khan MA, Kushner I. Clinical features of systemic lupus erythematosus: Differences related to race and age of onset. *Arthritis Rheum* 1982; 25: 55–60.
23 Wang F, Wang CL, Tan CT, Manivasagar M. Systemic lupus erythematosus in Malaysia: A study of 539 patients and comparison of prevalence and disease expression in different racial and gender groups. *Lupus* 1997; 6: 248–253.
24 Tan TC, Fang H, Magder LS, Petri MA. Differences between male and female systemic lupus erythematosus in a multiethnic population. *J Rheumatol* 2012; 39: 759–769.
25 Garcia MA, Marcos JC, Marcos AI, *et al.* Male systemic lupus erythematosus in a Latin-American inception cohort of 1214 patients. *Lupus* 2005; 14: 938–946.
26 Malaviya AN, Chandrasekaran AN, Kumar A, Shamar PN. Systemic lupus erythematosus in India. *Lupus* 1997; 6: 690–700.
27 Chahade WH, Sato EI, Moura JE Jr, Costallat LT, Andrade LE. Systemic lupus erythematosus in São Paulo/Brazil: A clinical and laboratory overview. *Lupus* 1995; 4: 100–103.
28 Cervera R, Khamashta MA, Font J, *et al.* Systemic lupus erythematosus: Clinical and immunologic patterns of disease expression in a cohort of 1,000 patients. The European Working Party on Systemic Lupus Erythematosus. *Medicine (Baltimore)* 1993; 72: 113–124.
29 Soto ME, Vallejo M, Guillén F, Simón JA, Arena E, Reyes PA. Gender impact in systemic lupus erythematosus. *Clin Exp Rheumatol* 2004; 22: 713–721.
30 Voulgari PV, Katsimbri P, Alamanos Y, Drosos AA. Gender and age differences in systemic lupus erythematosus. A study of 489 Greek patients with a review of the literature. *Lupus* 2002; 11: 722–729.
31 Font J, Cervera R, Navarro M, *et al.* Systemic lupus erythematosus in men: Clinical and immunological characteristics. *Ann Rheum Dis* 1992; 51: 1050–1052.
32 Stefanidou S, Benos A, Galanopoulou V, *et al.* Clinical expression and morbidity of systemic lupus erythematosus during a post-diagnostic 5-year follow-up: A male:female comparison. *Lupus* 2011; 20: 1090–1094.
33 Mok CC, To CH, Ho LY, Yu KL. Incidence and mortality of systemic lupus erythematosus in a southern Chinese population, 2000–2006. *J Rheumatol* 2008; 35: 1978–1982.
34 Ward MM, Polisson RP. A meta-analysis of the clinical manifestations of older-onset systemic lupus erythematosus. *Arthritis Rheum* 1989; 32: 1226–1232.
35 Lalani S, Pope J, de Leon F, Peschken C. Clinical features and prognosis of late-onset systemic lupus erythematosus: Results from the 1000 faces of lupus study. *J Rheumatol* 2010; 37: 38–44.
36 Jakes RW, Bae SC, Louthrenoo W, Mok CC, Navarra SV, Kwon N. Systematic review of the epidemiology of systemic lupus erythematosus in the Asia-Pacific region: Prevalence, incidence, clinical features, and mortality. *Arthritis Care Res (Hoboken)* 2012; 64: 159–168.

Downloaded from lup.sagepub.com at Kyoto University on January 14, 2015

ARTHRITIS & RHEUMATOLOGY
Vol. 66, No. 12, December 2014, pp 3395–3403
DOI 10.1002/art.38867
© 2014, American College of Rheumatology

Association Between Antinuclear Antibodies and the HLA Class II Locus and Heterogeneous Characteristics of Staining Patterns

The Nagahama Study

Chikashi Terao, Koichiro Ohmura, Ryo Yamada, Takahisa Kawaguchi, Masakazu Shimizu, Yasuharu Tabara, Meiko Takahashi, Kazuya Setoh, Takeo Nakayama, Shinji Kosugi, Akihiro Sekine, Fumihiko Matsuda, and Tsuneyo Mimori
on behalf of the Nagahama Study Group

Objective. **While antinuclear antibodies (ANAs) are observed in healthy populations as well as in patients with autoimmune diseases such as systemic lupus erythematosus (SLE), the detailed genetic background of ANAs has remained unclear. We undertook this study to identify the genetic determinants of ANAs in the general population in order to elucidate the underlying mechanisms of ANA production and to distinguish disease susceptibility genes from ANA production genes.**

Methods. **A total of 9,575 Japanese volunteers were registered, and their ANA levels were quantified using indirect immunofluorescence to analyze correlates of ANA positivity. Genetic studies were performed using 7,148 of the 9,575 subjects. We performed a genome-wide association study using 3,185 subjects genotyped for 303,506 single-nucleotide polymorphisms (SNPs), followed by a replication study of 3,963 subjects. HLA–DRB1 and HLA–DQB1 alleles were imputed, and associations between ANA positivity and the SNPs or the HLA alleles associated with SLE were analyzed.**

Results. **Female sex and old age were associated with ANA positivity, except for the nucleolar pattern. The T allele of rs2395185 in the HLA locus, which was in moderate linkage disequilibrium with HLA–DRB1*0405, was significantly associated with ANA positivity ($P = 1.3 \times 10^{-11}$). The T allele of rs2395185 displayed increasing effects on the frequency of speckled and homogeneous patterns ($P = 7.5 \times 10^{-12}$ and $P = 2.2 \times 10^{-11}$, respectively) but decreasing effects on the frequency of the nucleolar pattern ($P = 0.0045$). The 7 SNPs and 4 HLA–DRB1 alleles associated with SLE did not display strong associations with ANA positivity.**

Conclusion. **SNP rs2395185 linked with HLA–DRB1*0405 is a genetic determinant of ANA production in the Japanese population. Overlapping of loci for susceptibility to SLE and to ANA positivity was limited. The nucleolar pattern showed different associations from other staining patterns, both with correlates of ANA positivity and with the HLA locus.**

Antinuclear antibodies (ANAs) are autoantibodies that recognize various nuclear and cytoplasmic proteins, and they are frequently observed in patients with a broad range of diseases including systemic lupus erythematosus (SLE), hepatic disease, malignant disease, lung disease, and a variety of infections (1–6). The distribution patterns of fluorescent types of ANAs (such as speckled, homogeneous, nucleolar, or discrete speck-

Supported by the Ministry of Education, Culture, Sports, Science, and Technology of Japan (Grant-in-Aid for Scientific Research), Kyoto University (University Grant), the Japan Society for the Promotion of Science (Program for Enhancing Systematic Education in Graduate School Grant), and the Takeda Science Foundation (research grant).
Chikashi Terao, MD, PhD, Koichiro Ohmura, MD, PhD, Ryo Yamada, MD, PhD, Takahisa Kawaguchi, MSc, Masakazu Shimizu, PhD, Yasuharu Tabara, PhD, Meiko Takahashi, PhD, Kazuya Setoh, MSc, Takeo Nakayama, MD, PhD, Shinji Kosugi, MD, PhD, Akihiro Sekine, PhD, Fumihiko Matsuda, PhD, Tsuneyo Mimori, MD, PhD: Kyoto University, Kyoto, Japan.
Address correspondence to Chikashi Terao, MD, PhD, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Shogoin-Kawahara-cho 54, Sakyo-ku, Kyoto 606-8507, Japan. E-mail: a0001101@kuhp.kyoto-u.ac.jp.
Submitted for publication March 1, 2014; accepted in revised form August 28, 2014.

led patterns) also provide useful information for differential diagnosis (7–9). Previous studies have suggested that it is not unusual to find healthy individuals who are positive for ANAs (10). Since ANAs are included in the classification criteria for SLE as well as those for autoimmune hepatitis (11,12), analyzing the kinds of variables that affect the levels of ANAs would be helpful for avoiding excessive or deficient classification of these diseases as well as for gaining insight into their etiologies.

Although previous studies showed that ANA positivity was associated with female sex, old age, and being overweight (13,14), genetic components affecting ANA positivity in healthy individuals have never been addressed. Genome-wide association studies (GWAS) have detected many genes that confer susceptibility to connective tissue diseases, including SLE (15–18), and have elucidated the genetic background of biomarkers in general populations (19). Because almost all patients with SLE are positive for ANAs, it is important to confirm that SLE-related genes in the previous GWAS were not merely derived from their associations with ANA positivity.

At present, the number of large-scale studies addressing ANA levels in healthy subjects is quite limited. Detailed analyses of the correlates and genetic components of ANAs in healthy individuals would provide clues to the mechanisms responsible for the production of autoantibodies and the development of autoantibody-mediated autoimmune diseases (20,21). In the present study, we quantified circulating levels of ANAs in 9,575 Japanese volunteers for detailed analyses of the distributions and effects of correlates on ANA production. We also performed a GWAS in 7,148 of the 9,575 subjects to detect susceptibility loci that affect ANA production.

SUBJECTS AND METHODS

This study was approved by the Ethics Committee of Kyoto University Graduate School and Faculty of Medicine.

Study population. This study was performed as a part of the Nagahama Prospective Genome Cohort for Comprehensive Human Bioscience (the Nagahama Study), a community-based prospective multiomics cohort study conducted by the Center for Genomic Medicine at Kyoto University (22). A total of 9,809 volunteers ages 30–75 years in Nagahama City, Shiga Prefecture, Japan were recruited for this study. Written informed consent was obtained from each participant, and all were asked to complete a detailed questionnaire including present and past illnesses and lifestyle.

Exclusion criteria. We excluded volunteers from the association studies if they lacked necessary information or had ever been told that they have or had an autoimmune disease. We also excluded individuals whose answers to the questionnaire suggested that they might have an autoimmune disease. As a result, a total of 9,575 subjects remained for this study. A detailed flow chart of sample exclusion is shown in Supplementary Figure 1 (available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.38867/abstract).

Quantification of ANAs and C-reactive protein (CRP). ANAs and CRP in serum samples from volunteers were quantified (23) at SRL, one of the largest clinical laboratory testing companies in Japan. ANAs were quantified by serum dilution using indirect immunofluorescence with HEp-2 cells (TFB). Titers of ANAs with detailed staining patterns (speckled, homogeneous, nucleolar, cytoplasmic, and discrete speckled patterns) were also reported for these subjects. A cutoff level of 1:40 for positivity was applied according to the manufacturer's instructions.

Selection of potential correlates. Age, sex, body weight, smoking, alcohol use, and serum CRP level were selected as potential correlates based on a previous US study (14). CRP was quantified by highly sensitive methods using nephelometry, with a detection limit of 0.051 mg/liter, as previously reported (23).

Statistical analysis of nongenetic studies. The subjects were divided into 2 subgroups based on sex, 9 subgroups based on age (5-year intervals), and 18 subgroups based on sex and age. Associations between ANAs and age and/or sex were assessed by standardized logistic regression analysis. Odds ratios were also calculated with 95% confidence intervals. The associations between ANAs and potential correlates were analyzed by logistic regression analysis, with sex and age as covariates. Statistical analyses were performed using R statistical software (http://www.r-project.org) or SPSS version 18. We set significance levels in a conservative manner using Bonferroni correction for multiple testing.

GWAS. DNA samples from 3,710 of the 9,809 participants in the Nagahama Study were genome-scanned using Illumina HumanHap610, HumanHapOmni2.5-4, or Human HapOmni2.5-8 arrays. A total of 392,801 single-nucleotide polymorphisms (SNPs) that were common between the arrays were selected for the GWAS. We selected 3,185 subjects with call rates of >0.95 who did not show a high degree of kinship (PI_HAT <0.35) and who did not have connective tissue diseases. SNPs that showed P values less than 5×10^{-7} and in Hardy-Weinberg equilibrium ($P > 1 \times 10^{-7}$) with a success rate of >0.95 and a minor allele frequency of >0.05 were selected for a replication study using a TaqMan Assay (Applied Biosystems) with 3,963 of the participants. Population stratification was assessed with genomic control (24). Logistic regression analysis was performed to analyze the genetic influence on the production of ANAs for each SNP, corrected by age and sex. Logistic regression analysis was also used for the conditioning analysis. The associations of the 2 studies were combined using the inverse-variance method. The Jonckheere-Terpstra test was used to assess increasing effects of SNPs on ANA levels in subjects positive for ANAs.

HLA imputation. The HLA–DRB1 locus (the established HLA locus associated with SLE in previous reports) and the HLA–DQB1 locus were imputed using the GWAS data with HLA*IMP:02 (25). The imputation accuracy was evaluated by kappa coefficient with the use of imputation and genotyping data for 589 patients with rheumatoid arthritis and 932 healthy subjects for HLA–DRB1, as previously described

(23), and for 114 patients with thyroid diseases for HLA–DQB1 (Terao: unpublished observations). We analyzed whether each allele of HLA–DRB1 and HLA–DQB1 with imputation accuracy >70% was associated with ANA positivity by logistic regression analysis with additive or dominant models.

Evaluation of linkage disequilibrium (LD). LD between SNPs and HLA–DRB1 alleles was obtained from previous studies (17,26,27). For LD calculation between HLA–DRB1 and HLA–DQB1 alleles, we used genotyping data of 1,000 unrelated healthy Japanese subjects (Terao: unpublished observations).

Evaluation of effects of SLE-related SNPs. A total of 7 SNPs that displayed associations with SLE beyond levels significant in GWAS in a Japanese population (15) and the 5 SNPs in the HLA locus that displayed independent associations with SLE in Europeans (28) were selected to assess their effects on ANA positivity. The associations between these SNPs and ANA positivity were analyzed based on imputation by MaCH (29), using 192 samples in the Nagahama Study genotyped by HumanHapOmni2.5-8, HumanHapOmni2.5s, and HumanExome arrays or using East Asian panels in the 1000 Genomes Project as a reference when they were not directly genotyped.

Statistical analysis of genetic studies. Statistical calculations were performed using Plink software version 1.07 (30) and R statistical software. For all genetic analyses including the GWAS, we set significance levels using the Bonferroni correction for multiple testing.

RESULTS

A total of 9,575 subjects were analyzed for their ANA levels in the current study (Table 1). ANA titers in 45.2%, 12.5%, and 2.8% of the volunteers were ≥1:40, ≥1:80, and ≥1:160, respectively (see Supplementary Table 1, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.38867/abstract). When we analyzed potential correlates of ANA positivity, female sex and old age had higher correlations with ANA positivity, as shown in previous studies (13,14) (corrected P [P_{corr}] $< 1.0 \times 10^{-10}$) (see Supplementary Figure 2 and Supplementary Table 2, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.38867/abstract).

When we focused on each staining pattern, 43.7%, 25.3%, 4.7%, 0.9%, and 2.0% of subjects had ANAs with speckled, homogeneous, nucleolar, discrete speckled, and cytoplasmic patterns, respectively, at titers of ≥1:40 (Table 1). The multiple logistic regression analyses revealed that the nucleolar pattern was not associated with age or sex (see Supplementary Table 2 and Supplementary Figure 3, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.38867/abstract). Considering the higher frequency of the nucleolar pattern compared with that of the discrete speckled pattern, these results indicated that age and sex do not influence the positivity for each staining pattern in the same manner. Positivity for the speckled pattern was strongly correlated with positivity for all ANAs (see Supplementary Figure 4, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.38867/abstract). Associations between other potential correlates and ANAs are shown in Supplementary Table 3 (available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.38867/abstract). High CRP levels showed an association with ANA positivity (P_{corr} = 0.0029). We did not find a significant association between obesity and ANA positivity.

Table 1. Characteristics of the subjects in the current study*

	All subjects (n = 9,575)	GWAS (n = 3,185)†	Replication study (n = 3,963)†
Women	66.9	66.0	67.0
Age, mean ± SD years	53.3 ± 13.4	52.0 ± 14.1	53.7 ± 13.5
ANA titer ≥1:40			
All	45.2	48.4	42.5
Speckled	43.7	46.8	41.1
Homogeneous	25.3	29.0	21.3
Nucleolar	4.7	5.1	4.2
Discrete speckled	0.9	0.8	0.9
Cytoplasmic	2.0	1.6	2.3

* Except where indicated otherwise, values are the percent. ANA = antinuclear antibody.
† In the genome-wide association study (GWAS), DNA samples were genome-scanned using Illumina HumanHap610, HumanHapOmni2.5-4, or HumanHapOmni2.5-8 arrays. Genotyping in the replication study was performed using a TaqMan Assay.

Next, we performed a GWAS for ANA positivity. A total of 3,185 participants and 303,506 markers that had passed criteria of inclusion and quality control were used for logistic regression analysis, with age and sex as covariates. As a result, the Q–Q plot indicated an inflation factor of 1.02, suggesting that the current study was free from population stratification (Figure 1). A significant association of rs9405108 in the HLA locus was observed at a P value of 8.9×10^{-8}. Conditioning rs9405108 to detect further associated markers in this region did not result in any markers showing significant associations ($P > 1.0 \times 10^{-4}$) (data not shown). No SNPs in non-HLA regions displayed suggestive associations ($P > 1.0 \times 10^{-5}$). We performed a replication study for rs9405108 using 3,963 participants (Table 1). For technical reasons, SNP rs2395185, which is almost in complete LD with rs9405108 (D′ = 1 and r^2 = 0.999), was genotyped instead of rs9405108. As a result, the

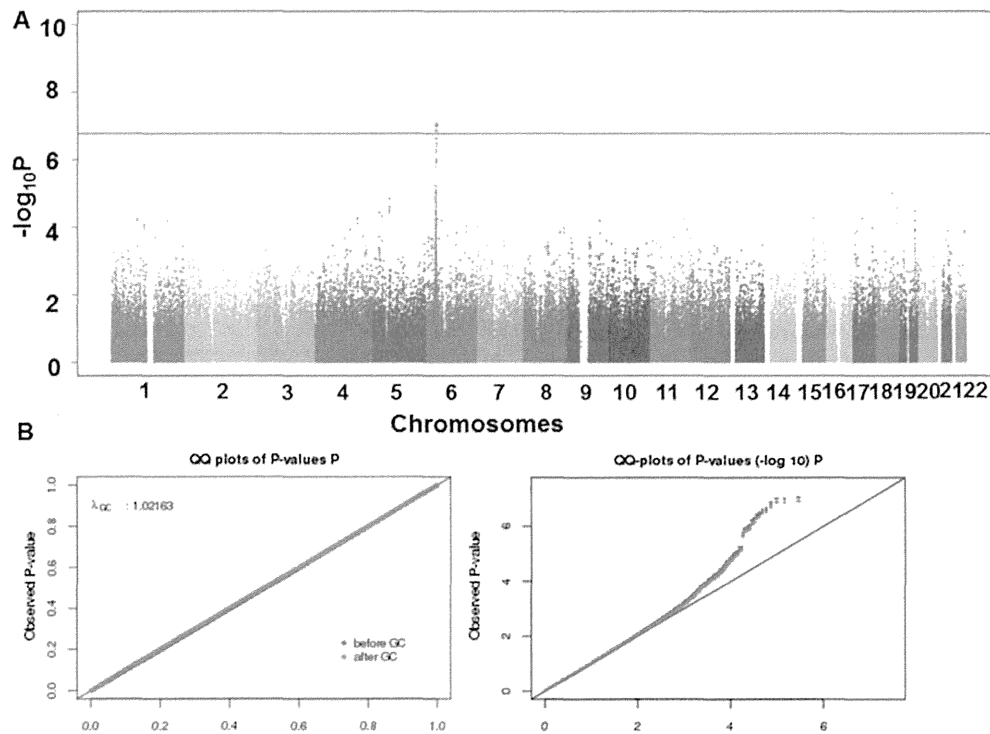

Figure 1. Genome-wide association study (GWAS) results for antinuclear antibody (ANA) production. **A,** Manhattan plot. The horizontal line indicates the significance level of the GWAS based on Bonferroni correction. **B,** Q–Q plots. λ_{gc} = genomic control inflation factor.

association of rs2395185 was replicated (overall $P = 1.3 \times 10^{-11}$) (Table 2).

SNP rs2395185 is located between the *HLA-DRA* and *HLA-DRB5* genes and is in moderate LD with HLA-DRB1*0405 ($r^2 = 0.42$). Considering that major histocompatibility complex proteins are respon-

Table 2. Associations of top SNPs with ANAs or their staining patterns*

SNP	Chr.	Position	ANA staining pattern	Nearest gene	Ref/var	Study	β	SE	OR (95% CI)	P
rs2395185	6	32541145	ANA (total)	*HLA-DRA*	G/T	GWAS	0.29	0.055	1.33 (1.20–1.48)	1.4×10^{-7}
						Replication	0.22	0.050	1.24 (1.12–1.37)	1.3×10^{-5}
						Overall	0.25	0.037	1.28 (1.19–1.38)	1.3×10^{-11}
rs2395185	6	32541145	Speckled	*HLA-DRA*	G/T	GWAS	0.29	0.055	1.33 (1.20–1.48)	1.4×10^{-7}
						Replication	0.22	0.050	1.25 (1.13–1.37)	8.3×10^{-6}
						Overall	0.25	0.037	1.29 (1.20–1.38)	7.5×10^{-12}
rs2395185	6	32541145	Homogeneous	*HLA-DRA*	G/T	GWAS	0.31	0.058	1.37 (1.22–1.54)	7.0×10^{-8}
						Replication	0.24	0.058	1.27 (1.13–1.42)	4.6×10^{-5}
						Overall	0.28	0.041	1.32 (1.22–1.43)	2.2×10^{-11}
rs6457300	6	31106721	Nucleolar	*C6orf205*	T/G	GWAS	−0.53	0.12	0.59 (0.46–0.74)	1.2×10^{-5}
						Replication	−0.13	0.11	0.88 (0.70–1.10)	0.26
						Overall	−0.32	0.083	0.73 (0.62–0.86)	0.00013
rs1611185	6	29876323	Discrete speckled	*HLA-G*	T/C	GWAS	1.28	0.29	3.61 (2.03–6.41)	1.2×10^{-5}
						Replication	0.19	0.25	1.21 (0.74–1.99)	0.44
						Overall	0.66	0.19	1.93 (1.32–2.80)	0.00060

* SNP = single-nucleotide polymorphism; ANAs = antinuclear antibodies; Chr. = chromosome; Ref/var = reference allele/variant allele; OR = odds ratio; 95% CI = 95% confidence interval; GWAS = genome-wide association study.

Table 3. Associations of SLE-related SNPs with ANA positivity and SLE susceptibility*

SNP	Chr.	Position	Gene	Ref/var	*P*	ANA OR (95% CI)†	SLE OR (95% CI)‡
Previous loci in Japanese population							
rs10168266	2	191644049	*STAT4*	T/C	0.20	1.08 (0.96–1.2)	1.59 (1.42–1.78)
rs340630	4	88177419	*AFF1*	A/G	0.13	1.08 (0.98–1.2)	1.21 (1.14–1.30)
rs9501626	6	32508322	*HLA*	A/C	0.62	1.04 (0.89–1.22)	1.86 (1.62–2.13)
rs2230926	6	138237759	*TNFAIP3*	G/T	0.15	1.16 (0.95–1.41)	1.75 (1.47–2.08)
rs6964720	7	75018280	*HIP1*	G/A	0.69	0.98 (0.86–1.1)	1.43 (1.27–1.63)
rs2254546	8	11381089	*BLK*	G/A	0.90	1.01 (0.9–1.13)	1.42 (1.25–1.61)
rs6590330	11	127816269	*ETS1*	A/G	0.015	1.14 (1.03–1.27)	1.44 (1.30–1.60)
Independent susceptibility SNPs of HLA locus in European population							
rs9265604	6	31407429	*HLA-B*	C/T	0.78	1.02 (0.92–1.13)	0.83 (0.78–0.89)
rs9378200	6	31680906	*BAT2*	C/T	0.17	0.92 (0.82–1.04)	0.59 (0.52–0.67)
rs9271731	6	32701590	*HLA-DRB1–HLA-DQA1*	G/A	0.41	1.06 (0.92–1.22)	1.34 (1.25–1.45)
rs9469220	6	32766288	*HLA-DQA1*	A/G	0.027	0.88 (0.78–0.98)	0.65 (0.61–0.68)

* SLE = systemic lupus erythematosus (see Table 2 for other definitions).
† For ANA positivity.
‡ For SLE susceptibility.

sible for self recognition and antigen presentation, the association between the polymorphisms in the HLA locus and ANAs seemed reasonable. HLA–DRB1*0405 is associated with a wide range of rheumatic and autoimmune diseases (26,31). This raised the possibility that autoimmune-related markers also had effects on ANA production. We selected SLE as being representative of autoimmune diseases with ANA production, and we analyzed the effects of a total of 7 markers that were reported to be associated with SLE in a previous Japa-

Table 4. Associations of ANA positivity with imputed HLA–DRB1 and HLA–DQB1 alleles*

HLA allele	Model	*P*	Corrected *P*†	OR (95% CI)	Accuracy
HLA–DRB1					
DRB1*0405	Dominant	3.0×10^{-5}	0.00081	1.43 (1.21–1.70)	0.902
DRB1*1302	Additive	3.6×10^{-5}	0.00097	0.69 (0.58–0.82)	0.997
DRB1*1201	Additive	0.00021	0.0057	0.58 (0.44–0.78)	0.704
DRB1*1401	Additive	0.069	1	0.80 (0.62–1.02)	0.746
DRB1*1101	Additive	0.095	1	0.77 (0.57–1.05)	0.827
DRB1*0901	Additive	0.11	1	1.13 (0.97–1.31)	1
DRB1*0701	Additive	0.23	1	0.37 (0.07–1.89)	1
DRB1*0803	Additive	0.33	1	1.10 (0.91–1.33)	0.987
DRB1*1502	Additive	0.52	1	0.95 (0.82–1.11)	0.998
DRB1*0401	Dominant	0.58	1	1.12 (0.74–1.71)	0.883
DRB1*1501	Additive	0.66	1	1.05 (0.86–1.28)	0.992
DRB1*1001	Additive	0.67	1	0.86 (0.42–1.74)	0.909
DRB1*1202	Additive	0.69	1	0.93 (0.64–1.34)	0.964
DRB1*0802	Additive	0.74	1	1.05 (0.77–1.45)	0.808
DRB1*0101	Dominant	0.90	1	1.01 (0.82–1.25)	0.992
HLA–DQB1					
DQB1*0301	Additive	3.5×10^{-5}	0.00095	0.71 (0.61–0.84)	0.888
DQB1*0604	Additive	0.00027	0.0073	0.71 (0.60–0.86)	1
DQB1*0401	Dominant	0.00031	0.0084	1.38 (1.16–1.65)	0.902
DQB1*0302	Dominant	0.0087	0.24	1.30 (1.07–1.59)	1
DQB1*0503	Additive	0.087	1	0.78 (0.58–1.04)	1
DQB1*0303	Additive	0.11	1	1.13 (0.97–1.31)	0.819
DQB1*0201	Dominant	0.15	1	3.46 (0.65–18.39)	1
DQB1*0402	Additive	0.20	1	1.18 (0.92–1.51)	0.907
DQB1*0602	Additive	0.49	1	1.08 (0.87–1.32)	1
DQB1*0601	Dominant	0.67	1	0.97 (0.83–1.12)	1
DQB1*0502	Dominant	0.75	1	0.95 (0.67–1.34)	1
DQB1*0501	Dominant	0.89	1	1.01 (0.83–1.24)	1

* See Table 2 for definitions.
† Corrected by Bonferroni adjustment.

nese study (15). The genotypes of these 7 markers were imputed using subjects in the Nagahama Study genotyped by denser arrays as a reference. All of the alleles showed good quality of imputation ($R^2 > 0.95$), but none of them displayed strong associations with ANA positivity ($P > 0.01$) (Table 3).

Since the HLA locus, especially HLA–DRB1, is the established locus for susceptibility to SLE with multiple independent associations shown beyond ethnicity (15,28,32), we analyzed detailed associations between the HLA locus and ANA positivity. A previous European study identified 5 independent SNPs that confer susceptibility to SLE (28). Because 1 of the 5 SNPs (rs1150703) is monomorphic in Japanese, the results for 4 SNPs are given in the current study (Table 3). None of them showed comparable associations in Europeans. We also performed imputation of HLA–DRB1 and HLA–DQB1 alleles (see Subjects and Methods). While the previous European study suggested the independent association of HLA–DQA1*0102 with SLE, we used HLA–DRB1*1501 and *1302 instead, which explained large parts of the association between HLA–DQA1*0102 and SLE (28). HLA–DRB1*0405, which was moderately tagged by rs2395185, showed a positive association with the smallest P value (P_{corr} = 0.00081) (Table 4). HLA–DQB1*0401 also showed a positive association, and HLA–DRB1*1302 and *1201 and HLA–DQB1*0301 and *0604 showed negative associations ($P_{corr} \leq 0.0084$) (Table 4). The associations of HLA–DQB1*0401, *0604, and *0301 seemed to be explained with HLA–DRB1*0405, HLA–DRB1*1302, and a combination of HLA–DRB1*1201 and HLA–DRB1*1101, respectively (r^2 values of 0.99, 0.92, and 0.59, respectively). HLA–DRB1*1501, the strongest susceptibility allele in Japanese (32), did not show a significant association (Table 4).

Considering the negative association of HLA–DRB1*1302 and the lack of association of HLA–DRB1*1501, HLA–DQA1*0102 was assumed to display a suggestive negative association. HLA–DRB1*0901, *0802, and *0401, which showed independent significant positive associations with SLE in Japanese (32), were not associated with ANA positivity.

Next, we addressed the similarities and differences of associations in the HLA locus among ANA staining patterns. Among the 2,820 SNPs in the HLA locus, rs9368726 and rs1964995, both of which were in strong LD with rs2395185 (r^2 values of 1.0 and 0.72, respectively), showed the strongest associations with speckled and homogeneous patterns, respectively ($P = 1.1 \times 10^{-7}$ and $P = 3.6 \times 10^{-8}$, respectively, in the GWAS) (Figure 2A). When we used the genotyping

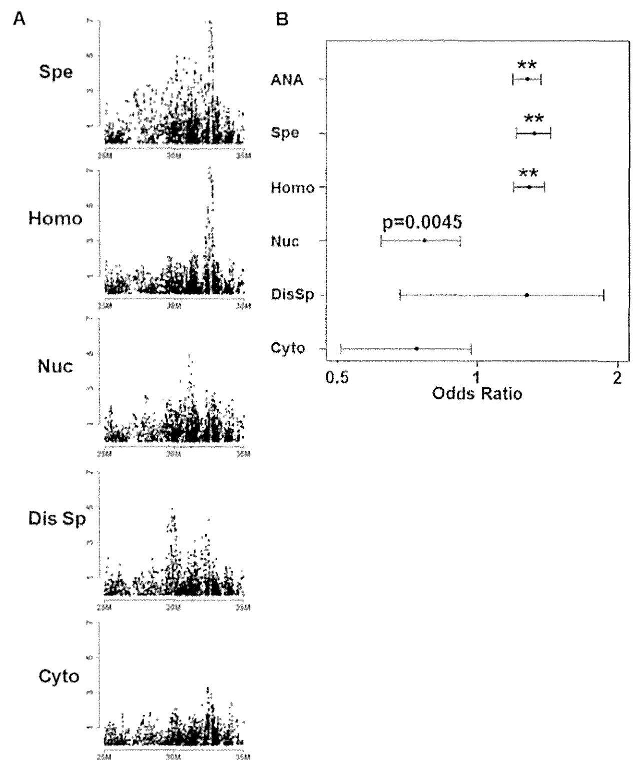

Figure 2. Heterogeneous association of the HLA locus among staining patterns of antinuclear antibodies (ANAs). **A,** Regional Manhattan plots for different staining patterns in the HLA region. **B,** Odds ratios and 95% confidence intervals of associations between rs2395185 and ANAs or their staining patterns. ** = $P < 1.0 \times 10^{-10}$. Spe = speckled; Homo = homogeneous; Nuc = nucleolar; Dis Sp = discrete speckled; Cyto = cytoplasmic.

results of rs2395185 instead of the 2 SNPs, the associations were also observed in the replication study (overall $P = 7.5 \times 10^{-12}$ and overall $P = 2.2 \times 10^{-11}$ for speckled and homogeneous patterns, respectively) (Table 2). The strongest associations with nucleolar and discrete speckled patterns in the HLA locus were observed for rs6457300 and rs1611185, respectively (both $P = 1.2 \times 10^{-5}$) (Table 2). Both SNPs are located >1.4 Mb from rs2395185. The cytoplasmic pattern showed the strongest association with rs9268347 ($P = 0.00052$), which is located 101 kb from rs2395185. We further genotyped rs6457300 and rs1611185 in the replication study, but the associations were not replicated (Table 2).

We focused on rs2395185 since it was the only SNP that demonstrated increasing effects on speckled and homogeneous patterns beyond levels significant in GWAS. Despite its increasing effects on the production of speckled and homogeneous patterns, the SNP displayed a significant decreasing effect on the nucleolar

pattern (P = 0.0045) (Figure 2B). Next, we analyzed whether rs2395185 had increasing effects on ANA levels in subjects positive for ANAs. When we examined subjects with ANA titers ≥1:40 and reviewed the staining patterns, the T allele of rs2395185 showed suggestive or significant increasing effects on levels of total, speckled, and homogeneous patterns (P = 0.12, P = 0.016, and P = 0.00030, respectively, by Jonckheere-Terpstra test).

DISCUSSION

The current study provided solid evidence of the distribution and correlates of ANAs in a Japanese adult population. This is the first study to perform GWAS of ANAs in healthy populations and detect a significant locus. The nucleolar pattern has characteristics that differ from those of other staining patterns. Autoantibodies such as anti–U3 RNP, anti-Th/To, or antiribosomal antibodies, associated with systemic sclerosis or SLE, are classified as having the nucleolar pattern of ANAs.

In our study, 12.5% of healthy participants had ANA titers of ≥1:80, which is comparable to previous results in the US (4,754 individuals, 13.8%) (14). The percentages were slightly higher than in previous studies for the cutoff level of 1:40 and comparable for the cutoff level of 1:160 (~26.8–31.7% and ~5.0–8.1%, respectively, in previous studies). Of the 201 subjects who were excluded due to the possibility of having autoimmune diseases, 141 had ANA titers of ≥1:40 (70.1%) (data not shown), suggesting the validity of the exclusion criteria. The increase in ANA positivity in women was confirmed, and this association could partly be explained by sex hormones (33–35). Considering the sex difference in onset of autoimmune diseases, the same undetermined mechanisms related to sex may underlie ANA production in healthy populations.

This study showed a strong effect of age on positivity for ANAs. We did not observe an increase in positivity for ANAs with aging in subjects 30–50 years old (P = 0.20) (data not shown); therefore, the elderly populations largely accounted for the association between aging and ANA positivity. The increase in ANAs after age 50 years matches the results in the US study. This association might be explained by dysregulation of immunologic tolerance in the elderly population. Considering the previous reports of high ANA levels in the adolescent population (13,36), the association between ANA positivity and aging in the general population seems to have a "U" pattern (lowest ANA levels at ages with most frequent reproduction). The effects of age and sex on ANAs seemed to differ among the staining patterns. The nucleolar pattern did not display significant associations with age and sex. As discrete speckled patterns showed positive associations, the lack of association of the nucleolar pattern with age and sex cannot be explained by its frequency.

Correlates of ANAs seemed to partly differ between different populations. The current study did not find a significant association between obesity and ANA positivity. However, obesity tended to be inversely related to ANA positivity as in the US study, and the limited number of obese individuals in the current study might explain this nonsignificant association. The association between increased CRP levels and ANA positivity was not found in the previous study. Chronic mild inflammation would lead to the production of ANAs. Since the distribution of CRP levels in subjects differs greatly between the 2 studies, further analysis would clarify the association.

The current study identified rs2395185 in the HLA class II locus as a marker of susceptibility to ANA positivity. It should be noted that a previous study showed an association between rs2395185 and ulcerative colitis (37), suggesting the involvement of rs2395185 with autoimmune processes. Because a previous study showed that the type I interferon (IFN) signature is up-regulated in healthy populations with high ANA titers (38), it will be interesting to analyze the functional roles of rs2395185 in the type I IFN pathway.

The T allele of rs2395185 showed increasing effects on levels of speckled and homogeneous patterns, but a decreasing effect on levels of the nucleolar pattern. This indicates that the nucleolar pattern also differs from the speckled and homogeneous patterns in terms of HLA association. The detailed plots in the HLA locus support the notion of different association patterns among ANA staining patterns. The opposing effect of rs2395185 on levels of the nucleolar pattern indicates that the lack of common association of rs2395185 over staining patterns of ANAs was not due to lower positivity for several staining patterns. As the HLA class II locus is strongly associated with presentation and recognition of antigen, the current results may suggest that ANA production is associated with binding affinity of antigens to the HLA molecule. Since antigens recognized by ANAs contain a wide variety of molecules, the common strong association of 1 polymorphism with speckled and homogeneous patterns suggests similarity or cross-reactivity of antigens that correspond to speckled or homogeneous patterns. The opposing effect also suggests that antigens corresponding to the nucleolar pattern are not presented by common HLA class II alleles with speckled and homogeneous patterns.

As HLA–DRB1*0405 is associated with susceptibility to immunologic disorders or autoantibody production in autoimmune diseases (27,39), the association between ANA production and rs2395185 in LD with HLA–DRB1*0405 might suggest a common mechanism between HLA–DRB1*0405–related autoimmune disease susceptibility and production of ANAs. At the same time, the association raises the possibility that genes conferring susceptibility to ANA positivity might be identified as genes conferring susceptibility to connective tissue diseases.

However, the current study did not detect significant associations between SLE-related SNPs or HLA alleles and ANA positivity. These results indicated that SNPs significantly associated with SLE in the previous study were associated with SLE itself and not with ANAs. Lack of association between ANA production in healthy subjects and rs9501626 or HLA–DRB1*1501, the most significant HLA SNP or HLA–DRB1 allele associated with SLE in the Japanese population, may suggest that autoantigens recognized by ANAs in SLE patients are different from those recognized by ANAs in healthy populations. In fact, a previous study showed that healthy subjects with high ANA titers exhibited an autoantibody profile distinct from that in SLE patients (38). These results may also suggest the involvement of immunologic molecular pathways in SLE development that are not related to ANA production in healthy populations. While we did not find associations of the 7 SNPs in Japanese and the 4 SNPs in Europeans, we observed that 9 of the 11 SNPs had a common direction of association between SLE susceptibility and ANA positivity. All the susceptibility DRB1 alleles in Japanese (HLA–DRB1*1501, *0901, *0802, and *0401) also showed a trend toward increasing ANAs. The common directionality between SLE susceptibility and ANA positivity may be meaningful.

It will be interesting to finely genotype the HLA locus to determine the polymorphisms and mechanisms responsible for causing the associations with ANAs or speckled and homogeneous patterns. None of the polymorphisms display significant associations with nucleolar, discrete speckled, or cytoplasmic patterns. However, considering the low positivity for these staining patterns and the strength of associations in the HLA locus in the current study, increasing the number of subjects would identify yet-to-be-determined polymorphisms associated with these staining patterns. We did not observe significant associations with ANA positivity outside the HLA locus. In addition, none of the polymorphisms outside the HLA locus showed suggestive associations with ANA staining patterns (data not shown). The significance and roles of ANAs in healthy populations have not yet been clarified. Because a previous study showed that the type I IFN signature is up-regulated in healthy populations with high ANA titers (38), it is possible that high ANA titers in healthy populations reflect a pre–autoimmune disease state. Further followup and analyses are necessary to address these points.

Taken together, the current study determined that the HLA class II locus is a locus for susceptibility to ANA production. Genetic overlap between SLE susceptibility and ANA production in healthy populations is limited. The current results indicate that ANAs are not homogeneous autoantibodies with similar characteristics. It is feasible to analyze whether the current results are observed in different populations, especially in Europeans.

ACKNOWLEDGMENTS

We are grateful to the Nagahama City Office and the nonprofit organization Zeroji Club for their help in performing the Nagahama Study.

AUTHOR CONTRIBUTIONS

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

Study conception and design. Terao, Ohmura, Yamada, Kawaguchi, Shimizu, Tabara, Takahashi, Setoh, Nakayama, Kosugi, Sekine, Matsuda, Mimori.

Acquisition of data. Terao, Ohmura, Yamada, Kawaguchi, Shimizu, Tabara, Takahashi, Setoh, Nakayama, Kosugi, Sekine, Matsuda, Mimori.

Analysis and interpretation of data. Terao, Kawaguchi.

REFERENCES

1. Czaja AJ, Nishioka M, Morshed SA, Hachiya T. Patterns of nuclear immunofluorescence and reactivities to recombinant nuclear antigens in autoimmune hepatitis. Gastroenterology 1994;107:200–7.
2. Solomon DH, Kavanaugh AJ, Schur PH, and the American College of Rheumatology Ad Hoc Committee on Immunologic Testing Guidelines. Evidence-based guidelines for the use of immunologic tests: antinuclear antibody testing. Arthritis Rheum 2002;47:434–44.
3. Petri M, Karlson EW, Cooper DS, Ladenson PW. Autoantibody tests in autoimmune thyroid disease: a case-control study. J Rheumatol 1991;18:1529–31.
4. Rich S, Kieras K, Hart K, Groves BM, Stobo JD, Brundage BH. Antinuclear antibodies in primary pulmonary hypertension. J Am Coll Cardiol 1986;8:1307–11.
5. Guyomard S, Salles G, Coudurier M, Rousset H, Coiffier B, Bienvenu J, et al. Prevalence and pattern of antinuclear autoantibodies in 347 patients with non-Hodgkin's lymphoma. Br J Haematol 2003;123:90–9.
6. Bonnet F, Pineau JJ, Taupin JL, Feyler A, Bonarek M, de Witte S, et al. Prevalence of cryoglobulinemia and serological markers of

autoimmunity in human immunodeficiency virus infected individuals: a cross-sectional study of 97 patients. J Rheumatol 2003;30: 2005–10.
7. Koenig M, Dieude M, Senecal JL. Predictive value of antinuclear autoantibodies: the lessons of the systemic sclerosis autoantibodies. Autoimmun Rev 2008;7:588–93.
8. Buyon JP, Ben-Chetrit E, Karp S, Roubey RA, Pompeo L, Reeves WH, et al. Acquired congenital heart block: pattern of maternal antibody response to biochemically defined antigens of the SSA/Ro-SSB/La system in neonatal lupus. J Clin Invest 1989;84:627–34.
9. Hahn BH. Antibodies to DNA. N Engl J Med 1998;338:1359–68.
10. Tan EM, Feltkamp TE, Smolen JS, Butcher B, Dawkins R, Fritzler MJ, et al. Range of antinuclear antibodies in "healthy" individuals. Arthritis Rheum 1997;40:1601–11.
11. Petri M, Orbai AM, Alarcon GS, Gordon C, Merrill JT, Fortin PR, et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64:2677–86.
12. Alvarez F, Berg PA, Bianchi FB, Bianchi L, Burroughs AK, Cancado EL, et al. International Autoimmune Hepatitis Group Report: review of criteria for diagnosis of autoimmune hepatitis. J Hepatol 1999;31:929–38.
13. Hayashi N, Koshiba M, Nishimura K, Sugiyama D, Nakamura T, Morinobu S, et al. Prevalence of disease-specific antinuclear antibodies in general population: estimates from annual physical examinations of residents of a small town over a 5-year period. Mod Rheumatol 2008;18:153–60.
14. Satoh M, Chan EK, Ho LA, Rose KM, Parks CG, Cohn RD, et al. Prevalence and sociodemographic correlates of antinuclear antibodies in the United States. Arthritis Rheum 2012;64:2319–27.
15. Okada Y, Shimane K, Kochi Y, Tahira T, Suzuki A, Higasa K, et al. A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus erythematosus in Japanese. PLoS Genet 2012;8:e1002455.
16. Terao C, Yamada R, Ohmura K, Takahashi M, Kawaguchi T, Kochi Y, et al. The human AIRE gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. Hum Mol Genet 2011;20:2680–5.
17. Terao C, Ohmura K, Katayama M, Takahashi M, Kokubo M, Diop G, et al. Myelin basic protein as a novel genetic risk factor in rheumatoid arthritis—a genome-wide study combined with immunological analyses. PLoS One 2011;6:e20457.
18. Terao C, Yoshifuji H, Kimura A, Matsumura T, Ohmura K, Takahashi M, et al. Two susceptibility loci to Takayasu arteritis reveal a synergistic role of the IL12B and HLA-B regions in a Japanese population. Am J Hum Genet 2013;93:289–97.
19. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat Genet 2010;42: 210–5.
20. Schur PH, Sandson J. Immunologic factors and clinical activity in systemic lupus erythematosus. N Engl J Med 1968;278:533–8.
21. Koffler D, Agnello V, Winchester R, Kunkel HG. The occurrence of single-stranded DNA in the serum of patients with systemic lupus erythematosus and other diseases. J Clin Invest 1973;52: 198–204.
22. Yoshimura K, Nakayama T, Sekine A, Matsuda F, Kosugi S, Sugino Y, et al, and the Nagahama Cohort Research Group. Prevalence of postmicturition urinary incontinence in Japanese men: comparison with other types of incontinence. Int J Urol 2013;20:911–6.
23. Terao C, Ohmura K, Ikari K, Kawaguchi T, Takahashi M, Setoh K, et al, on behalf of the Nagahama Study Group. The effects of smoking and shared epitope on the production of ACPA and RF in a Japanese adult population: the Nagahama Study. Arthritis Care Res (Hoboken) 2014. E-pub ahead of print.
24. Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999;55:997–1004.
25. Dilthey A, Leslie S, Moutsianas L, Shen J, Cox C, Nelson MR, et al. Multi-population classical HLA type imputation. PLoS Comput Biol 2013;9:e1002877.
26. Terao C, Ohmura K, Kochi Y, Ikari K, Maruya E, Katayama M, et al. A large-scale association study identified multiple HLA-DRB1 alleles associated with ACPA-negative rheumatoid arthritis in Japanese subjects. Ann Rheum Dis 2011;70:2134–9.
27. Terao C, Ikari K, Ohmura K, Suzuki T, Iwamoto T, Takasugi K, et al. 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. Ann Rheum Dis 2012;71:1095–7.
28. Morris DL, Taylor KE, Fernando MM, Nititham J, Alarcon-Riquelme ME, Barcellos LF, et al. Unraveling multiple MHC gene associations with systemic lupus erythematosus: model choice indicates a role for HLA alleles and non-HLA genes in Europeans. Am J Hum Genet 2012;91:778–93.
29. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 2010;34:816–34.
30. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559–75.
31. Yamashita H, Awata T, Kawasaki E, Ikegami H, Tanaka S, Maruyama T, et al, on behalf of the Japanese Study Group on Type 1 Diabetes Genetics. Analysis of the HLA and non-HLA susceptibility loci in Japanese type 1 diabetes. Diabetes Metab Res Rev 2011;27:844–8.
32. Shimane K, Kochi Y, Suzuki A, Okada Y, Ishii T, Horita T, et al. An association analysis of HLA-DRB1 with systemic lupus erythematosus and rheumatoid arthritis in a Japanese population: effects of *09:01 allele on disease phenotypes. Rheumatology (Oxford) 2013;52:1172–82.
33. Nashold FE, Spach KM, Spanier JA, Hayes CE. Estrogen controls vitamin D_3-mediated resistance to experimental autoimmune encephalomyelitis by controlling vitamin D_3 metabolism and receptor expression. J Immunol 2009;183:3672–81.
34. Wilder RL. Neuroendocrine-immune system interactions and autoimmunity. Annu Rev Immunol 1995;13:307–38.
35. Strickland FM, Hewagama A, Lu Q, Wu A, Hinderer R, Webb R, et al. Environmental exposure, estrogen and two X chromosomes are required for disease development in an epigenetic model of lupus. J Autoimmun 2012;38:J135–43.
36. Yoshida H, Imafuku Y, Morita EO, Nagai T, Kato Y, Motegi S. Detection of antinuclear antibodies and their significance as disease markers. Rinsho Byori 1994;42:455–9. In Japanese.
37. Silverberg MS, Cho JH, Rioux JD, McGovern DP, Wu J, Annese V, et al. Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. Nat Genet 2009; 41:216–20.
38. Li QZ, Karp DR, Quan J, Branch VK, Zhou J, Lian Y, et al. Risk factors for ANA positivity in healthy persons. Arthritis Res Ther 2011;13:R38.
39. Terao C, Ohmura K, Ikari K, Kochi Y, Maruya E, Katayama M, et al. ACPA-negative RA consists of two genetically distinct subsets based on RF positivity in Japanese. PLoS One 2012;7: e40067.

OPEN ACCESS Freely available online PLOS ONE

Large-Scale East-Asian eQTL Mapping Reveals Novel Candidate Genes for LD Mapping and the Genomic Landscape of Transcriptional Effects of Sequence Variants

Maiko Narahara[1], Koichiro Higasa[2], Seiji Nakamura[3], Yasuharu Tabara[2], Takahisa Kawaguchi[2], Miho Ishii[3], Kenichi Matsubara[3], Fumihiko Matsuda[2], Ryo Yamada[1]*

1 Statistical Genetics, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, 2 Human Disease Genomics, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, 3 DNA Chip Research Inc., Kanagawa, Japan

Abstract

Profiles of sequence variants that influence gene transcription are very important for understanding mechanisms that affect phenotypic variation and disease susceptibility. Using genotypes at 1.4 million SNPs and a comprehensive transcriptional profile of 15,454 coding genes and 6,113 lincRNA genes obtained from peripheral blood cells of 298 Japanese individuals, we mapped expression quantitative trait loci (eQTLs). We identified 3,804 *cis*-eQTLs (within 500 kb from target genes) and 165 *trans*-eQTLs (>500 kb away or on different chromosomes). *Cis*-eQTLs were often located in transcribed or adjacent regions of genes; among these regions, 5′ untranslated regions and 5′ flanking regions had the largest effects. Epigenetic evidence for regulatory potential accumulated in public databases explained the magnitude of the effects of our eQTLs. *Cis*-eQTLs were often located near the respective target genes, if not within genes. Large effect sizes were observed with eQTLs near target genes, and effect sizes were obviously attenuated as the eQTL distance from the gene increased. Using a very stringent significance threshold, we identified 165 large-effect *trans*-eQTLs. We used our eQTL map to assess 8,069 disease-associated SNPs identified in 1,436 genome-wide association studies (GWAS). We identified genes that might be truly causative, but GWAS might have failed to identify for 148 out of the GWAS-identified SNPs; for example, *TUFM* (P=3.3E-48) was identified for inflammatory bowel disease (early onset); *ZFP90* (P=4.4E-34) for ulcerative colitis; and *IDUA* (P=2.2E-11) for Parkinson's disease. We identified four genes (P<2.0E-14) that might be related to three diseases and two hematological traits; each expression is regulated by *trans*-eQTLs on a different chromosome than the gene.

Citation: Narahara M, Higasa K, Nakamura S, Tabara Y, Kawaguchi T, et al. (2014) Large-Scale East-Asian eQTL Mapping Reveals Novel Candidate Genes for LD Mapping and the Genomic Landscape of Transcriptional Effects of Sequence Variants. PLoS ONE 9(6): e100924. doi:10.1371/journal.pone.0100924

Editor: Amanda Ewart Toland, Ohio State University Medical Center, United States of America

Received January 22, 2014; **Accepted** June 2, 2014; **Published** June 23, 2014

Copyright: © 2014 Narahara et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by Ministry of Health, Labour and Welfare, Japan (2009-2013), Kyoto University (2006), and the Research Grant for Intractable Diseases from the Ministry of Health, Labour and Welfare of Japan (201238002A). Co-authors Seiji Nakamura, Miho Ishii and Kenichi Matsubara are employed by DNA Chip Research Inc. DNA Chip Research Inc. provided support in the form of salaries for authors Seiji Nakamura, Miho Ishii and Kenichi Matsubara, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the "author contributions" section.

Competing Interests: The authors have the following interests: Co-authors Seiji Nakamura, Miho Ishii and Kenichi Matsubara are employed by the DNA Chip Research Inc. The affiliation to this company does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

* Email: ryamada@genome.med.kyoto-u.ac.jp

Introduction

Variation in gene expression levels is one of the major factors causing phenotypic variation and disease susceptibility. Although gene expression levels are influenced by environmental factors, genetic variations also play an important role in transcriptional regulation; notably, about 30% of transcriptional phenotypes are heritable (h^2>30%) [1]. Additionally, many loci identified in genome-wide association studies (GWAS) are located in non-coding regions that have no known protein-coding genes, suggesting that these loci influence transcriptional regulation. Expression quantitative trait locus (eQTL) mapping is a common approach to locate genetic loci that regulate transcription, and recent development with genome-wide SNP typing arrays and gene expression microarrays has enhanced genome-wide eQTL mapping. Genome-wide eQTL maps can substantially improve our understanding of transcriptional regulation at the genetic level; they can also improve the interpretability of the results of GWAS. Moreover, comprehensive hypothesis-free scans of eQTLs can provide hypothesis-generating results; this approach may lead to the unexpected discovery of important biological phenomena. Consequently, eQTL mapping has been intensively studied in humans [1–9]. However, further eQTL mapping studies would be valuable because technical advances in high-throughput genome analysis are being made in terms of experiments, accumulation of knowledge, and computation. Moreover, non-coding RNAs are important regulators of gene expression, and these RNAs greatly influence many phenotypes [10,11]. Therefore, profiling eQTLs of non-coding RNAs should be very valuable for biomedical research; however, previous eQTL studies have focused almost exclusively on protein-coding genes. Here, our study included

6,113 lincRNA probes; we identified 278 unique eQTLs that affected 326 unique lincRNA probes.

Notably, expression levels of many individual genes vary among human populations [7,12–16], and this variation among populations is primarily explained by differences in genotype frequencies (R^2 of ~0.81) among populations; nevertheless, population-specific genotypic effects may also be an important source of this variation (R^2 of ~0.31) [13]. Additionally, low between-population replication rates of eQTLs indicate that population-specific eQTL effects exist; for example, only 37% of *cis*-eQTLs and 15% of *trans*-eQTLs identified in one population were also identified in a second population [7]. Therefore, ethnicity-specific eQTL maps may be very useful for basic and applied research. Here, we describe large-scale eQTL mapping in a Japanese population; the sample size ($n = 298$ unrelated individuals) was 3-fold larger than that in any preceding eQTL study of East Asian individuals [7,13,17]; moreover, updated genome and gene data were used to improve the coverage of tested transcripts over that in preceding studies. In this study, we report genome-wide, high-resolution eQTL association mapping for baseline gene expression levels in peripheral blood cells.

We identified 3,804 *cis*-eQTLs (defined as a SNP that affects expression of a gene located within 500 kb) that affected 16.9% of genes; among these *cis*-eQTLs, the mean fold difference in gene expression levels between two homozygous genotypes was 1.6-fold, and the mean proportion of transcriptional variance explained by genotype was 0.19. We also identified 165 *trans*-eQTLs (defined as a SNP that affects expression of any transcript more than 500 kb away or on a different chromosome); among these *trans*-eQTLs, the mean fold difference in gene expression levels between two homozygous genotypes was 2.1-fold, and the mean proportion of transcriptional variance explained by genotype was 0.27. *Cis*-eQTLs were more likely to be located in gene structure and the adjacent regions; specifically, 45.7% of *cis*-eQTLs were located within 1 kb of the respective differentially expressed gene (genic *cis*-eQTLs). The genic *cis*-eQTLs had a larger effect than other *cis*-eQTLs (mean $|\beta|$: 0.33 vs. 0.31, $P = 0.00093$; mean R^2 0.21 vs. 0.17, $P = 7.8\text{E-}11$). *Cis*-eQTLs with the largest effects (top 10%) were located predominantly in genic regions (58% in genic vs. 42% in the others). Among the genic regions, 5′ untranslated regions (UTR) and upstream regions within 1 kb of a transcription start site had relatively more *cis*-eQTLs than the other regions, and *cis*-eQTLs with larger effects also tended to be located in these two types of genic regions; the mean effect size of *cis*-eQTLs in these regions were 1.4-fold larger than those of others (mean $|\beta|$: 0.45 vs. 0.32, $P = 0.0033$). The density of *cis*-eQTLs decreased exponentially with distance from respective structural genes; the majority (70%) of *cis*-eQTLs were located within 17 kb-flanking or within a target protein coding gene; and effects of individual *cis*-eQTLs became small with distance from a target gene.

eQTL analyses have been used to reliably identify variant-gene pair(s) among potential combinations of SNPs identified by GWAS and the nearby genes [18]; nevertheless, a considerable fraction of GWAS have not included eQTLs evaluation. In many GWAS, the gene closest to the significant SNP is reported as a probable causative gene. However, there are two major caveats with this practice: 1) when multiple genes are in strong linkage disequilibrium (LD) in the detected region, the reported SNP may capture an effect of a faraway gene, and thus, GWAS cannot determine which gene in the LD region is truly causative; and 2) the reported SNP may capture a transcriptional regulatory site that is located far from the regulated and causative gene. Therefore, eQTL maps may improve interpretation of GWAS results and overcome these two caveats by identifying causative genes whose expression is actually altered. We used our eQTL map to reassess 8,069 trait/disease-associated SNPs identified in 1,436 published GWAS; our eQTL map suggested different causative genes from those reported in published GWAS for 148 of the GWAS-identified SNPs.

Our eQTL mapping project is part of the Human Genetic Variation Browser (http://www.genome.med.kyoto-u.ac.jp/SnpDB/), an open-access database; this project is intended to provide researchers with integrative genomic data–including our eQTL map, summary statistics for genotypes of all SNPs used in this study, and exome sequencing data–for biomedical studies.

Results

Gene expression profile

Our study population comprised 298 individuals (102 male and 196 female); the mean age was 55.1 years, and age ranged from 32 to 66 years (Table S1). We treated each probe as though it represented a unique transcript, and each Entrez Gene ID represented a distinct gene. With this definition of genes, our expression profile was comprised of 30,395 autosomal transcripts (17,598 genes): 19,818 *mRNA* transcripts representing 15,454 genes, 6,113 *lincRNA* transcripts (no gene ID was assigned for any of them), and 4,464 *other* transcripts representing 3,288 genes (see Methods for classification). The numbers of genes (15,454 and 3,288) do not add up to the total (17,598) because 1,144 gene IDs were found in both mRNA and others as different transcripts. Definitions of *cis*- and *trans*-eQTLs, and local and distant SNPs are described in Methods.

Cis-eQTL analysis

A distribution of P values for all local SNP-transcript pairs showed an excess of small P values (Figure S1A), suggesting that a substantial fraction of associations are truly positive. With the false discovery rate (FDR) <5%, we identified 3,804 *cis*-eQTLs transcript pairs (Figure 1, Table 1). 12.5%, or 16.9%, of all tested transcripts, or genes, were *cis*-regulated (Table 1). The complete list of the *cis*-eQTLs with annotation and statistics is provided in File S1.

We used two statistics as measures for magnitudes of effects of eQTLs; the coefficient of genotypes was designated β, or its absolute value $|\beta|$, and the proportion of transcriptional variance explained by genotypes was designated R^2 (see supplementary note in File S3 for more explanation). *Cis*-eQTLs with large effects were abundant (Figure 2A, 2B and Table 1): for example, the number of *cis*-eQTLs with $|\beta|$ values larger than 0.3, which corresponds to a 1.5-fold change between two homozygous genotypes, was 1,440 (4.7%) of all tested transcripts. The numbers of *cis*-eQTLs with R^2 values larger than 0.1 were 2,568 (8.4%) of all examined transcripts.

Gene-based functional categories and protein consequences. Next, we analyzed the *cis*-eQTLs in terms of gene-based functional categories of SNPs. Here, we analyzed the *cis*-eQTLs that affected mRNAs because the structures of the coding genes represented by these transcripts were the most clearly annotated. First, we compared SNPs in genic regions, those within genes and 1 kb upstream or downstream of genes, with SNPs in intergenic regions. We define *enrichment* as the fold change in proportion that each group constitutes among *cis*-eQTLs compared to among all local SNPs. The enrichment of genic SNPs was 7.04 (45.74% of *cis*-eQTLs vs. 6.50% of all local SNPs, Table 2). Moreover, *cis*-eQTLs had significantly stronger effects than did intergenic *cis*-eQTLs (mean $|\beta|$ values 0.33 vs. 0.31, $P = 0.00093$; mean R^2 values 0.21 vs. 0.17, $P = 7.8\text{E-}11$, Table 2, Figure 3A,

Table 1. Summary statistics and counts of *cis*- and *trans*-eQTLs at thresholds by *R2* or |β|.

		cis-eQTL (*n*=3,804 by FDR <5%)				*trans*-eQTL (*n*=165 by FWER <5%)			
		All	**mRNA**	**lincRNA**	**Other**	**All**	**mRNA**	**lincRNA**	**Other**
#eQTLs-transcript pairs		3,804	2,995	293	516	165	91	49	25
#unique eQTLs		3,385	2,779	244	440	105	65	34	21
#unique transcripts (%)		3,804 (12.5%)	2,995 (15.1%)	293 (4.8%)	516 (11.6%)	114 (0.4%)	60 (0.3%)	34 (0.6%)	20 (0.4%)
#unique genes (%)		2,973 (16.9%)	2,667 (17.3%)	0	357 (10.9%)	74 (0.4%)	57 (0.4%)	0	17 (0.5%)
#unique transcripts without gene ID		455	28	293	134	54	1	49	4
R^2 mean±SD		0.19±0.15	0.19±0.15	0.20±0.16	0.21±0.18	0.27±0.12	0.27±0.12	0.29±0.13	0.23±0.07
R^2 median±IQR		0.13±0.15	0.13±0.14	0.13±0.15	0.15±0.17	0.23±0.12	0.23±0.12	0.26±0.11	0.21±0.09
|β| mean±SD		0.33±0.33	0.32±0.32	0.38±0.33	0.38±0.34	0.53±0.35	0.50±0.38	0.59±0.31	0.51±0.33
|β| median±IQR		0.24±0.24	0.23±0.23	0.30±0.31	0.28±0.25	0.47±0.41	0.40±0.44	0.58±0.34	0.43±0.34
Stat.	**Cutoff**								
R^2	**0.1**	2,568	2,022	192	354	165	91	49	25
	0.3	665	500	56	109	45	25	15	5
	0.5	245	171	22	52	10	5	5	0
	0.7	67	43	8	16	2	2	0	0
	0.9	2	2	0	0	0	0	0	0
|β|	**0.3**	1,440	1,053	146	241	118	60	40	18
	0.9	155	122	11	22	23	13	6	4
	1.5	55	44	3	8	3	1	1	1
	2.1	26	19	2	5	1	1	0	0
	2.7	12	9	1	2	1	1	0	0

FDR: false discovery rate; FWER: family-wise error rate; SD: standard deviation; IQR: inter-quartile range; R^2: proportion of phenotypic variances explained by genotypes; |β|: absolute value of coefficient of genotypes.
The sum of #unique eQTLs counted within RNA types is not necessarily equal to #unique eQTLs counted for all transcripts because the same eQTLs may be counted in more than one RNA types. The number of genes for All and each type do not match for a similar reason.

doi:10.1371/journal.pone.0100924.t001

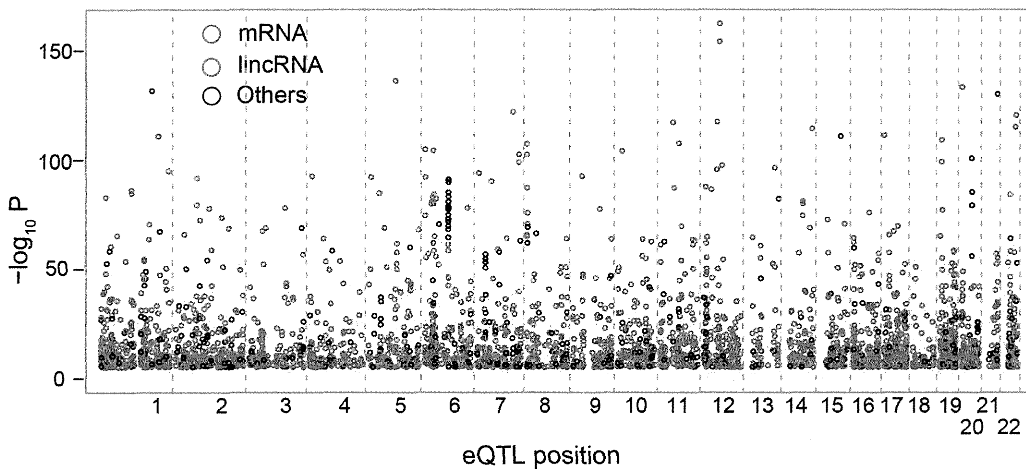

Figure 1. Cis-eQTL map. $-\log_{10}$ P values of cis-eQTLs are plotted against the respective chromosomal positions. eQTLs for mRNA transcripts are shown in red; lincRNA transcripts are shown in green; and other transcripts are shown in black. The vertical dashed lines separate chromosomes.
doi:10.1371/journal.pone.0100924.g001

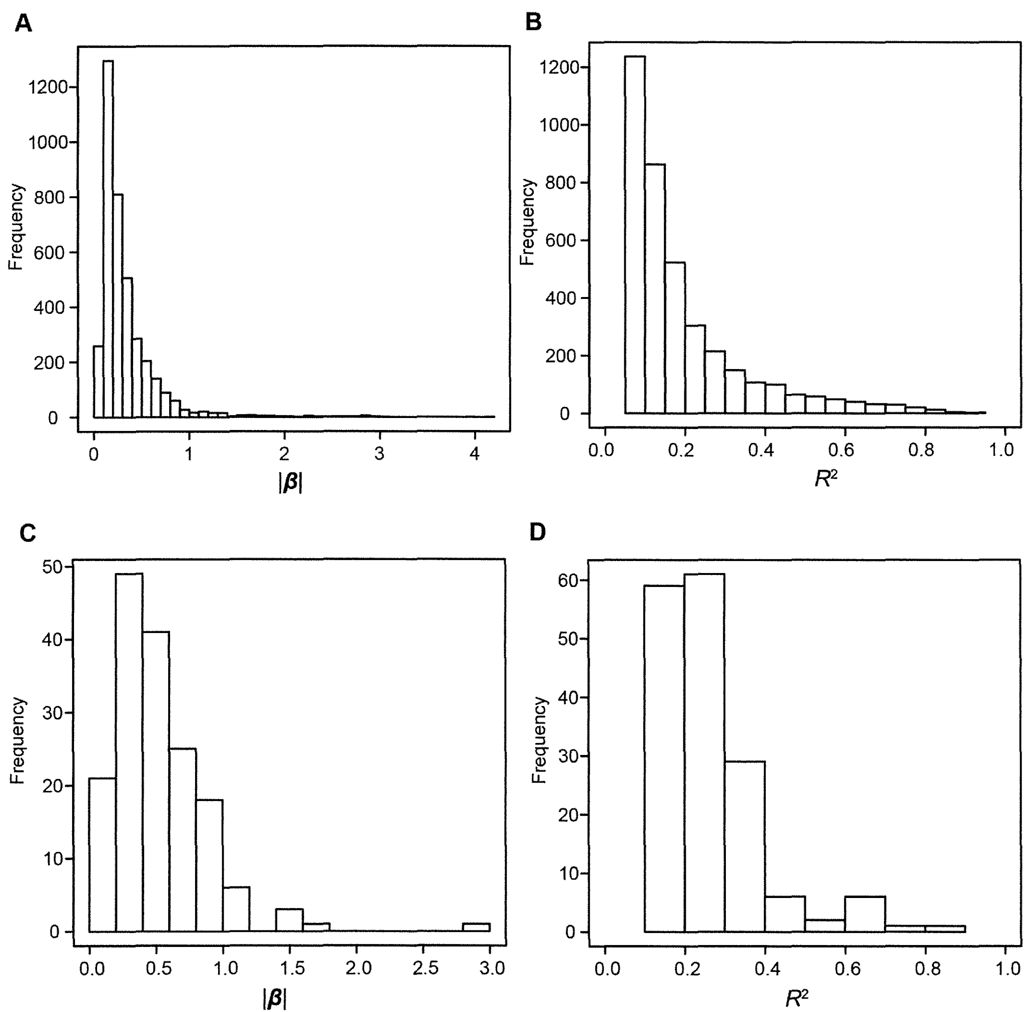

Figure 2. Histograms of effects of eQTLs. A, B) Histograms of |β| values (A) and of R^2 values (B) of *cis*-eQTLs are shown. C, D) Histograms of |β| values (C) and R^2 values (D) of *trans*-eQTLs are shown.
doi:10.1371/journal.pone.0100924.g002