TABLE 3. POPULATION MODEL EXAMPLES GIVEN AS HAPLOTYPE-ALLELE FREQUENCIES

	Frequency in population	
Haplotype-allele	(i) $\mathcal{M}_1$	(ii) $\mathcal{M}_2$
1111	0.40	0.20
1110	0.00	0.07
1101	0.20	0.08
1011	0.25	0.10
0011	0.00	0.05
0110	0.10	0.30
0101	0.00	0.05
1100	0.00	0.05
1000	0.05	0.10
Others	0.00	0.00

Against each of the 8,930 blocks, we performed Ward's clustering algorithm (see Section 2.3.1) based on the ASD and also did the same based on the HHD, and compared the CERs (see Section 2.3.2) of their results (Table 6). The difference of the results in relation to the number of htSNPs, i.e., h_B (see Section 3.3.2), is also shown.

The mean of CERs based on the HHD (i.e., 0.3557) is better than that for the ASD (i.e., 0.3611). The P-value of the t-test to see the difference between them is 0.004177, which means the CERs of the HHD is significantly better than that of the ASD. The number of data sets where the HHD (or the ASD) shows better performance than the ASD (or the HHD) are checked with the sign test. Among all the data sets, the HHD is superior to the ASD on 4366 data sets and inferior to the ASD in 3696 data sets. The results of two measures were the same in the other 868 data sets. The P-value of the sign test of all of these results is $8.98 \cdot 10^{-14}$, which means that the HHD is significantly superior to the ASD.

The CERs decrease with increasing h_B for both the ASD and the HHD, but the differences of CERs between the ASD and the HHD also increases as h_B increase (Fig. 4). We call the result HDD's success if the HHD's CER is lower than that of the ASD, and vice versa. The ratio of the HHD's success increases with increasing h_B. The ratio of ASD's success also increases with increasing h_B. The difference of ratios of success between the ASD and the HHD is getting larger as h_B increases. The ratio of the case when the ASD and the HHD have the same results are getting lower as h_B increases (Fig. 5).

The HHD is superior to ASD especially when $80 \leq h_B < 90$. It is a reasonable result as we should be able to better cluster individuals if we have more information (i.e., LE). The difference of ratios of success

TABLE 4. CONDITIONAL PROBABILITIES OF CANDIDATE HAPLOTYPE-DIPLOTYPES FOR INDIVIDUALS
IN FIGURE 2 BASED ON THE POPULATION MODELS IN TABLE 3

			Conditional probability	
Individual	Unphased-diplotype	Candidate haplotype-diplotype	(i) $\mathcal{M}_1$	(ii) $\mathcal{M}_2$
	{1,0}-{1,0}-{1,1}-{1,0}	{1011, 0110}	1.0000	0.8955
a		{1110, 0011}	0.0000	0.1045
		Others	0.0000	0.0000
	{1,0}-{1,1}-{1,0}-{1,0}	{1101, 0110}	1.0000	0.8727
b		{1110, 0101}	0.0000	0.1273
		Others	0.0000	0.0000
	{1,1}-{1,0}-{1,0}-{1,0}	{1111, 1000}	1.0000	0.8000
c		{1011, 1100}	0.0000	0.2000
		Others	0.0000	0.0000

TABLE 5. MEANS OF VARIANCES OF ASD/HHD MEASURES ON THE REGIONS WHERE THE SNPs ARE WEAKLY CORRELATED AND HIGHLY CORRELATED IN CHROMOSOME 1

	Mean of variances		
	LE	LD	P-value
ASD	0.00267	0.00546	$2.066 \cdot 10^{-16}$
HHD	0.00248	0.00539	$1.637 \cdot 10^{-17}$

The LE and LD columns show the means of variances on the LE regions (i.e., regions with many htSNPs) and those on the LD regions (i.e., regions with a few htSNPs), respectively. The difference of the variances between weakly and highly correlated regions are tested by t-test for each of the measures. The P-value column shows the P-value of the t-test.

between the ASD and the HHD also becomes largest when $80 < h_B < 90$. In this case, the HHD is superior on 13 data sets, while the ASD is superior only on six data sets among the remaining 18 data sets.

5. CONCLUSION

We proposed a new inter-diplotype similarity measure that we call the PMD. The PMD improves the previous ASD measure by utilizing a population model. As one of such population models, we propose to use the HMM population model used in the phasing algorithm HIT. We call the PMD based on the HIT's HMM the HHD. The HHD utilizes the predicted conditional probabilities of haplotype-diplotypes of unphased-diplotype emitted from the HIT's HMM. Based on comprehensive experiments over 8930 genome-wide data sets of HapMap, we showed that the HHD significantly outperforms the ASD. We also discussed the relationships between the clustering accuracies and the LD.

There are many future tasks to do related to this work. The HHD requires much larger computation time than the ASD, and one future task should be to improve the computation speed of the HHD. There are still data sets for which the HHD is not superior to the ASD. It would be very interesting if we can predict the regions where the HHD is inferior to the ASD, before computing these measures. Another future task is to improve the population model, as it should directly improve the performance of the PMD. From the biological viewpoint, it would also be very interesting if we can utilize our clustering algorithms to identify

TABLE 6. THE EXPERIMENTAL RESULTS AND THEIR RELATIONSHIPS TO THE H_B VALUES

h_B	♯blocks	Mean of CERs		Comparison of CERs			
		ASD	HHD	$CER_{ASD} < CER_{HHD}$	$CER_{HHD} < CER_{ASD}$	$CER_{ASD} = CER_{HHD}$	P-value of sign test
$0 \sim 10$	1	0.5630	0.5630	0 (0.0)	0 (0.0)	1 (1.0)	
$10 \sim 20$	44	0.4733	0.4678	9 (0.2045)	13 (0.2955)	22 (0.5)	0.5235
$20 \sim 30$	223	0.4363	0.4305	62 (0.2780)	82 (0.3677)	79 (0.3543)	0.1130
$30 \sim 40$	993	0.4240	0.4207	380 (0.3827)	418 (0.4209)	195 (0.1964)	0.1902
$40 \sim 50$	2364	0.3929	0.3877	975 (0.4124)	1131 (0.4784)	258 (0.1091)	$7.276 \cdot 10^{-4*}$
$50 \sim 60$	3063	0.3567	0.3514	1327 (0.4332)	1528 (0.4989)	208 (0.06793)	$1.808 \cdot 10^{-4*}$
$60 \sim 70$	1822	0.3052	0.2997	772 (0.4237)	970 (0.5324)	80 (0.04391)	$2.303 \cdot 10^{-6*}$
$70 \sim 80$	399	0.2584	0.2465	165 (0.4135)	211 (0.5288)	23 (0.05764)	$0.02018*$
$80 \sim 90$	21	0.2178	0.1944	6 (0.2857)	13 (0.6190)	2 (0.09524)	0.1671
$90 \sim 100$	0	—	—	—	—	—	—
Total	8930	0.3611	0.3557	3696 (0.4139)	4366 (0.4889)	868 (0.09720)	$8.98 \cdot 10^{-14*}$

The ♯blocks column shows the numbers of blocks with the specified h_B values. In the Comparison of CERs columns, the $CER_{ASD} < CER_{HHD}/CER_{ASD} > CER_{HHD}/CER_{ASD} = CER_{HHD}$ columns show the numbers (and the ratios) of data (with the specified h_B values) where the ASD performed better/the HHD performed better/the performance of the two measures are exactly the same, respectively. $x \sim y$ indicates that $x \le h_B < y$, and * means the result of the sign test is significant (i.e., ≤ 0.05).

FIG. 4. The plot of h_B values and the means of CERs for both the ASD and the HHD. $x \sim y$ indicates that $x \leq h_B < y$. The HHD is superior to the ASD in all the cases.
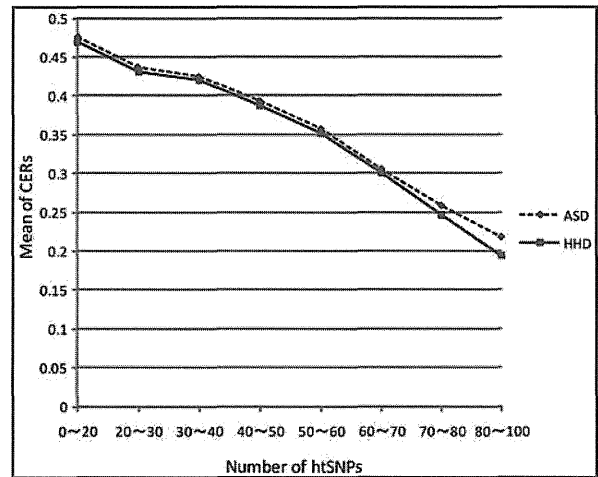
gene functions of the target genome regions, especially the regions that affect the disease prevalence and drug responses (Bamshad et al., 2004; Wiencke, 2004; Wilson et al., 2001).

6. APPENDIX

*Counting number of mutations under founder hypothesis*

Suppose that founder haplotype-alleles $f_1, \ldots, f_m$ has been evolved into the present-day haplotpye-alleles of individuals p and q, without any recombinations. Let p_1 and p_2 be the haplotype-alleles of p and q_1 and q_2 be the haplotype-alleles of q. We can consider that the number of mutations between p and q under the assumption of founders $f_1, \ldots, f_m$ as

$$S_{f_1, \ldots, f_m}(p, q) = \min \left\{ \sum_{i=1}^{2} \min_{j=1}^{m} \{dist(p_i, f_j) + dist(q_i, f_j)\}, \right.$$

$$\left. \sum_{i=1}^{2} \min_{j=1}^{m} \{dist(p_i, f_j) + dist(q_{2-i}, f_j)\} \right\}, \tag{7}$$

where $dist()$ denotes the ordinary number of mutations between the two sequences.

But we cannot know the appropriate set of founder haplotype-alleles. Instead, we can define the number of mutations between two individuals under the assumption that there are m founders as

FIG. 5. The plot of h_B values and the ratios of success for both the ASD and the HHD. The line ASD = HHD indicates the results in which the performance of the two measures are the exactly the same. $x \sim y$ indicates that $x \leq h_B < y$. The HHD is superior to the ASD in all the cases.
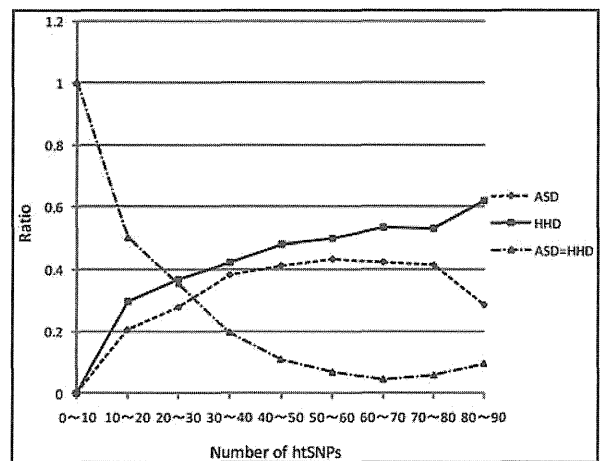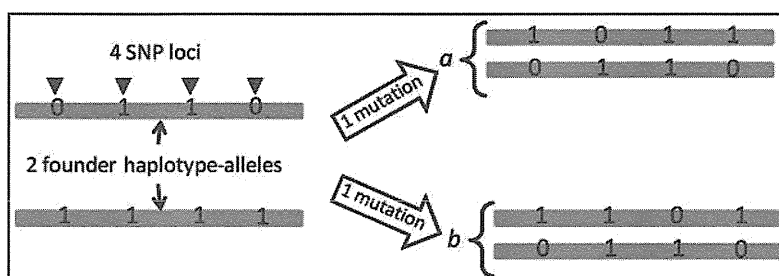
FIG. 6. The optimal founder haplotype-allele pair (when $m = 2$) for the individuals a and b in Figure 2.

$$S^*_m(p, q) = \min_{f_1, ..., f_m} S_{f_1, ..., f_m}(p, q).$$ (8)

Table 2 shows all the $S^*_2()$ values for all the pairs among individuals a, b, and c in Figure 2. Figure 6 shows the founder pair f_1, f_2 that minimizes the $S_{f_1, f_2}(a, b)$ value.

ACKNOWLEDGMENTS

DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES

Ahmad, T., Neville, M., Marshall, S.E., et al. 2002. Haplotype-specific linkage disequilibrium patterns define the genetic topography of the human MHC. *Hum. Mol. Genet.* 12, 647–656.

Bamshad, M., Wooding, S., Salisbury, B.A., et al. 2004. Deconstructing the relationship between genetics and race. *Nat. Rev. Genet.* 5, 598–609.

Beaty, T.H., Fallin, M.D., Hetmanski, J.B., et al. 2005. Haplotype diversity in 11 candidate genes across four populations. *Genetics* 171, 259–267.

Bhatia, G., Bansal, V., Harismendy, O., et al. 2010. A covering method for detecting genetic associations between rare variants and common phenotypes. *Plos Comput. Biol.* 6, 1–12.

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., et al. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368, 455–457.

Carlson, C.S., Eberle, M.A., Rieder, M.J., et al. 2004. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.* 74, 106–120.

Chung, P.Y.J., Beyens, G., Guanabens, N., et al. 2008. Founder effect in different European countries for the recurrent P392L SQSTM1 mutation in Paget's disease of bone. *Calcif. Tissue. Int.* 83, 34–42.

Cirulli, E.T., and Goldstein, D.B. 2010. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* 11, 415–425.

Conrad, D.F., Jakobsson, M., Coop, G., et al. 2006. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat. Genet.* 38, 1251–1260.

Cornuet, J.M., Sylvain, P., Luikart, G., et al. 1999. New methods employing multilocus genotypes to select or exlcude populations as origins of individuals. *Genetics* 153, 1989–2000.

Cover, T.M., and Thomas, J.A. 1991. *Elements of Information Theory*, John Wiley & Sons, New York.

de Bakker, P.I.W., Yelensky, R., Pe'er, I., et al. 2005. Efficiency and power in genetic association studies. *Nat. Genet.* 37, 1217–1223.

Durbin, R., Eddy, S., Krogh, A., et al. 1998. *Biological Sequence Analysis*. Cambridge Press, New York.

Ester, M., Kriegel, H.P., Sander, J., et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. *Proc. 2nd Int. Conf. Knowl. Discov. Data Mining* 226–231.

Fallin, D., Cohen, A., Essioux, L., et al. 2001. Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE locus variation and Alzheimer's disease. *Genome Res.* 11, 143–151.

Falush, D., Stephens, M., and Pritchard, J.K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587.

Gao, X., and Martin, E.R. 2009. Using allele sharing distance for detecting human population stratification. *Hum. Hered.* 68, 3.

Gao, X., and Starmer, J. 2007. Human population structure detection via multilocus genotype clustering. *BMC Genet.* 8, 34.

Gaudieri, S., Leelayuwat, C., Tay, G.K., et al. 1997. The Major Histocompatability Complex (MHC) contains conserved polymorphic genomic sequences that are shuffled by recombination to form ethnic-specific haplotypes. *J. Mol. Evol.* 45, 17–23.

Gonzalez, E., Bamshad, M., Sato, N., et al. 1999. Race-specific HIV-1 disease-modifying effects associated with CCR5 haplotypes. *Proc. Natl. Acad. Sci. USA* 96, 12004–12009.

Haiman, C.A., Stram, D.O., Pike, M.C., et al. 2003. A comprehensive haplotype analysis of CYP19 and breast cancer risk: The Multiethnic Cohort. *Hum. Mol. Genet.* 12, 2679–2692.

International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature* 437, 1299–1320. Available at www.hapmap.org. Accessed November 1, 2011.

Isaev, A. 2004. *Introduction to mathematical methods to bioinformatics.* Springer, New York.

Jakobsson, M., Scolz, S.W., Scheet, P., et al. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* 451, 998–1003.

Jin, L., Zhu, W., and Guo, J. 2010. Genome-wide association studies using haplotype clustering with a new haplotype similarity. *Genet. Epidemiol.* 34, 633–641.

Johnson, G.C.L., Esposito, L., Barratt, B.J., et al. 2001. Haplotype tagging for the identification of common disease genes. *Nat. Genet.* 29, 233–237.

Kaufman, L, and Rousseeuw, P. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis.* John Wiley and Sons, New York.

Ke, X., and Cardon, L.R. 2003. Efficient selective screening of haplotype tag SNPs. *Bioinformatics* 19, 287–288.

Kim, S., and Misra, A. 2007. SNP genotyping: technologies and biomedical applications. *Annu. Rev. Biomed. Eng.* 9, 289–320.

Lesk, A.M. 2005. Introduction to Bioinformatics, 2nd ed. *Oxford,* New York.

Li, J., and Jiang, T. 2005. Haplotype-based linkage disequilibrium mapping via direct data mining. *Bioinformatics* 21, 4384–4393.

Li, J., Zhou, Y., and Elston, R.C. 2006. Haplotype-based quantitative trait mapping using a clustering algorithm. *BMC Bioinform.* 7, 258.

Mao, X., Bigham, A.W., Mei, R., et al. 2007. A genomewide admixture mapping panel for Hispanic/Latino populations. *Am. J. Hum. Genet.* 80, 1171–1178.

Meng, Z., Zaykin, D.V., Xu, C., et al. 2003. Selection of genetic markers for association analyses, using linkage disequilibrium and haplotypes. *Am. J. Hum. Genet.* 73, 115–130.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.

Rabiner, L.R., and Juang, B.H. 1986.An introduction to hidden Markov models. *IEEE ASSP Mag.* 3, 4–16.

Rastas, P., Koivisto, P.M., Mannila, H., et al. 2005. A hidden Markov technique for haplotype reconstruction. *Lect. Notes Bioinform.* 3692, 140–151.

Rinaldo, A., Bacanu, S., Devlin, B., et al. 2005. Characterization of multilocus linkage disequilibrium. *Genet. Epidemiol.* 28, 193–206.

Saitou, N., and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.

Small, K.M., Mialet-Perez, J., Seman, C.A., et al. 2004. Polymorphisms of cardiac presynaptic α_{2C} adrenergic receptors: diverse intragenic variability with haplotype-specific functional effects. *Proc. Natl. Acad. Sci. USA* 101, 13020–13025.

Team RDC. 2007. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.*

Tzeng, J.Y., Devlin, B., Wasserman, L., et al. 2003. On the identification of disease mutations by the analysis of haplotype similarity and goodness of fit. *Am. J. Hum. Genet.* 72, 891–902.

Ward, J.H. 1963. Hierarchical grouping procedure to optimize an objective function. *J. Am. Stat. Assoc.* 58, 236–244.

Ward, J.H., and Hook, M.E. 1963. Application of an hierarchical grouping procedure to a problem of grouping profiles. *Educ. Psychol. Measure.* 23, 69–81.

Wiencke, J.K. 2004. Impact of race/ethnicity on molecular pathways in human cancer. *Nat. Rev. Cancer* 4, 79–84.

Wilson, F.W., Weale, M.E., Smith, A.C., et al. 2001. Population genetic structure of variable drug response. *Nat. Genet.* 29, 265–269.

Witherspoon, D.J., Wooding, S., Rogers, A.R., et al. 2007. Genetic similarities within and between human populations. *Genetics* 176, 351–359.

Yang, Y., and Tabus, I. 2007. Haplotype block partitioning using a normalized maximum likelihood model. *Proc. IEEE Genomic Signal Process. Stat.* 1–4.

Address correspondence to:
Ms. Ritsuko Onuki
Bioinformatics Center
Institute for Chemical Research
Kyoto University
Gokasho, Uji
Kyoto 611-0011, Japan

E-mail: onuki@hgc.jp

A Genome-Wide Association Study Identified *AFF1* as a Susceptibility Locus for Systemic Lupus Eyrthematosus in Japanese

Yukinori Okada[1,2,3☺], Kenichi Shimane[1,2☺], Yuta Kochi[1,2☺]*, Tomoko Tahira[4], Akari Suzuki[1], Koichiro Higasa[3], Atsushi Takahashi[3], Tetsuya Horita[5], Tatsuya Atsumi[5], Tomonori Ishii[6], Akiko Okamoto[2], Keishi Fujio[2], Michito Hirakata[7], Hirofumi Amano[8], Yuya Kondo[9], Satoshi Ito[9], Kazuki Takada[10], Akio Mimori[11], Kazuyoshi Saito[12], Makoto Kamachi[13], Yasushi Kawaguchi[14], Katsunori Ikari[14], Osman Wael Mohammed[15], Koichi Matsuda[15], Chikashi Terao[16,17], Koichiro Ohmura[16], Keiko Myouzen[1], Naoya Hosono[18], Tatsuhiko Tsunoda[19], Norihiro Nishimoto[20], Tsuneyo Mimori[16], Fumihiko Matsuda[17], Yoshiya Tanaka[12], Takayuki Sumida[9], Hisashi Yamanaka[14], Yoshinari Takasaki[8], Takao Koike[5], Takahiko Horiuchi[21], Kenshi Hayashi[4], Michiaki Kubo[18], Naoyuki Kamatani[3], Ryo Yamada[1,17], Yusuke Nakamura[15], Kazuhiko Yamamoto[1,2]

1 Laboratory for Autoimmune Diseases, Center for Genomic Medicine (CGM), RIKEN, Yokohama, Japan, 2 Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan, 3 Laboratory for Statistical Analysis, CGM, RIKEN, Yokohama, Japan, 4 Division of Genome Analysis, Research Center for Genetic Information, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan, 5 Department of Medicine II, Hokkaido University Graduate School of Medicine, Sapporo, Japan, 6 Department of Hematology and Rheumatology, Tohoku University Graduate School of Medicine, Sendai, Japan, 7 Division of Rheumatology, Department of Internal Medicine, Keio University School of Medicine, Tokyo, Japan, 8 Department of Internal Medicine and Rheumatology, Juntendo University School of Medicine, Tokyo, Japan, 9 Division of Clinical Immunology, Doctoral Program in Clinical Sciences, Graduate School of Comprehensive Human Science, University of Tsukuba, Tsukuba, Japan, 10 Departments of Medicine and Rheumatology, Graduate School, Tokyo Medical and Dental University, Tokyo, Japan, 11 Division of Rheumatic Diseases, National Center for Global Health and Medicine, Tokyo, Japan, 12 First Department of Internal Medicine, University of Occupational and Environmental Health, Kitakyushu, Japan, 13 Department of Immunology and Rheumatology, Unit of Translational Medicine, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan, 14 Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan, 15 Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan, 16 Department of Rheumatology and Clinical immunology, Graduate School of Medicine, Kyoto University, Kyoto, Japan, 17 Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, 18 Laboratory for Genotyping Development, CGM, RIKEN, Yokohama, Japan, 19 Laboratory for Medical Informatics, CGM, RIKEN, Yokohama, Japan, 20 Laboratory of Immune Regulation, Wakayama Medical University, Wakayama, Japan, 21 Department of Medicine and Biosystemic Science, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan

Abstract

Systemic lupus erythematosus (SLE) is an autoimmune disease that causes multiple organ damage. Although recent genome-wide association studies (GWAS) have contributed to discovery of SLE susceptibility genes, few studies has been performed in Asian populations. Here, we report a GWAS for SLE examining 891 SLE cases and 3,384 controls and multi-stage replication studies examining 1,387 SLE cases and 28,564 controls in Japanese subjects. Considering that expression quantitative trait loci (eQTLs) have been implicated in genetic risks for autoimmune diseases, we integrated an eQTL study into the results of the GWAS. We observed enrichments of cis-eQTL positive loci among the known SLE susceptibility loci (30.8%) compared to the genome-wide SNPs (6.9%). In addition, we identified a novel association of a variant in the AF4/FMR2 family, member 1 (*AFF1*) gene at 4q21 with SLE susceptibility (rs340630; $P = 8.3 \times 10^{-9}$, odds ratio = 1.21). The risk A allele of rs340630 demonstrated a cis-eQTL effect on the *AFF1* transcript with enhanced expression levels ($P < 0.05$). As *AFF1* transcripts were prominently expressed in $CD4^+$ and $CD19^+$ peripheral blood lymphocytes, up-regulation of *AFF1* may cause the abnormality in these lymphocytes, leading to disease onset.

Author Summary

Although recent genome-wide association study (GWAS) approaches have successfully contributed to disease gene discovery, many susceptibility loci are known to be still uncaptured due to strict significance threshold for multiple hypothesis testing. Therefore, prioritization of GWAS results by incorporating additional information is recommended. Systemic lupus erythematosus (SLE) is an autoimmune disease that causes multiple organ damage. Considering that abnormalities in B cell activity play essential roles in SLE, prioritization based on an expression quantitative trait loci (eQTLs) study for B cells would be a promising approach. In this study, we report a GWAS and multi-stage replication studies for SLE examining 2,278 SLE cases and 31,948 controls in Japanese subjects. We integrated eQTL study into the results of the GWAS and identified *AFF1* as a novel SLE susceptibility loci. We also confirmed cis-regulatory effect of the locus on the *AFF1* transcript. Our study would be one of the initial successes for detecting novel genetic locus using the eQTL study, and it should contribute to our understanding of the genetic loci being uncaptured by standard GWAS approaches.

Introduction

Systemic lupus erythematosus (SLE) is an autoimmune disease characterized by autoantibody production, complement activation, and multi-organ damage [1]. Familial aggregation demonstrates that both genetic and environmental factors play a role in pathogenesis of SLE [2]. Genetic studies using candidate gene-approaches, and recently, genome-wide association studies (GWAS), have uncovered more than 25 SLE susceptibility genes, including *HLA-DRB1*, *IRF5*, *STAT4*, *ITGAM*, *BLK*, *TNFAIP3*, and others [3–18]. However, most of these studies were conducted in European populations [3–13,15,17], and few studies have been conducted in Asian populations [14,16,18]. Since the epidemiology of SLE has demonstrated that the prevalence of disease substantially differs among populations, genetic backgrounds of SLE should be also heterogeneous across populations [19,20]. Therefore, additional studies in Asians might provide novel insights. It is of note that GWAS for SLE in Chinese populations identified novel loci that had not been detected in Europeans, such as *ETS1*, *IKZF1*, and *WDFY4* [14,16].

Another issue raised by the previous GWASs for complex diseases is that many susceptibility loci still remained uncaptured, owing to its strict significance threshold for multiple hypothesis testing [21]. In SLE, for example, the 26 risk loci identified by the previous GWAS explained only an estimated 8% of the total genetic susceptibility to the disease [15]. Therefore, it is still important to examine the sub-loci of GWAS, in order to reveal the entire picture of genetic etiology. To effectively explore these uncaptured loci, prioritization of GWAS results by incorporating additional information implicated in the disease pathophysiology is recommended [22,23]. Considering that abnormalities in B cell activity play essential roles in SLE [1] and that expression quantitative trait loci (eQTL) have been implicated to comprise approximately a half of genetic risks for autoimmune diseases [24], prioritization based on an eQTL study for B cells would be a promising approach for SLE [25]. Moreover, an eQTL itself assures the presence of functional variant(s) that regulate gene expression. Thus, eQTL increases the prior probability of the presence of disease-causal variant(s) in the locus more effectively

and unbiasedly, compared to other knowledge-based prioritizations such as gene pathway analysis [24].

Here, we report a GWAS and multi-stage replication studies for SLE examining 2,278 SLE cases and 31,948 controls in Japanese subjects. We integrated eQTL study into the results of the GWAS, which effectively enabled to detect a novel SLE susceptibility locus.

Results

GWAS for SLE

In the GWAS, 891 SLE cases and 3,384 controls in Japanese subjects were genotyped over 550,000 single nucleotide polymorphism (SNP) markers (Table S1, S2 and Figure 1). We applied stringent quality control (QC) criteria and evaluated associations of 430,797 autosomal SNPs, as previously described [26]. No substantial population stratification was demonstrated through principal component analysis (Figure S1) or a Quantile–Quantile plot of P-values (inflation factor, $\lambda_{GC} = 1.088$, Figure S2), suggesting homogenous ancestries of our study population [27].

We identified significant associations in six chromosomal loci that satisfied the genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ (Table 1 and Figure 2A). These loci have been reported to be associated with SLE susceptibility (*STAT4*, *TNFAIP3*, *HIP1*, *BLK*, *ETS1*, and the HLA region) [3–18]. We also observed significant replications in 17 of the previously reported SLE susceptibility loci [3–18] ($\alpha = 0.01$; Table 2). Of these, significant replications were enriched in the loci identified through the studies in Asian populations (80%; 8 of the 10 loci), including *RASGRP3*, *IKZF1*, *HIP1*, *WDFY4*, intergenic region at 11q23, *ETS1*, *SLC15A4*, *ELF1*, and *HIC2-UBE2L3* [14,16,18], compared to those in European populations (56.3%; 9 of the 16 loci) [3–13,15,17].

Incorporation of eQTL study into GWAS results

For the selection of SNPs incorporated in the replication studies of the potential association signals, we evaluated cis-eQTL effects of the SNPs using publically available gene expression data [28], and prioritized the results of the GWAS. After applying QC criteria, we evaluated the expression levels of 19,047 probes assayed in lymphoblastoid B cell lines from Phase II HapMap East-Asian individuals [29] using Illumina's human whole-genome expression array (WG-6 version 1) [28]. For each of the SNPs included in our GWAS, probes located within ±300 kbp regions were focused on as cis-eQTLs (average 4.93 probes per SNP). We denoted the SNPs which exhibited significant associations with expression levels of any of the corresponding cis-eQTLs as eQTL positive (false discovery rate (FDR) Q-values<0.2). We observed enrichments of eQTL positive loci among the SLE susceptibility loci (30.8%; 8 of the 26 evaluated loci) including a well-known eQTL gene of *BLK* [11,25] (Table 2), compared to the genome-wide SNPs (6.9%) and compared even to the SNPs in the vicinity of expressed loci (among the SNPs located within ±10 kbp of probes used for the expression analysis, 13.1% were eQTL positive; Table S3).

By prioritizing the results of the GWAS using the eQTL study, we selected 57 SNPs from 1,207 SNPs that satisfied $P < 1.0 \times 10^{-3}$ in the GWAS. We subsequently referred the associations of the selected SNPs using the results of the concurrent genome-wide scan for SLE in an independent Japanese population (Tahira T et al. Presented at the 59th Annual Meeting of the American Society of Human Genetics, October 21, 2009). In the scan, 447 SLE cases and 680 controls of Japanese origin were evaluated using a pooled DNA approach [30]. We selected SNPs if any association signals were observed in the neighboring SNPs of the
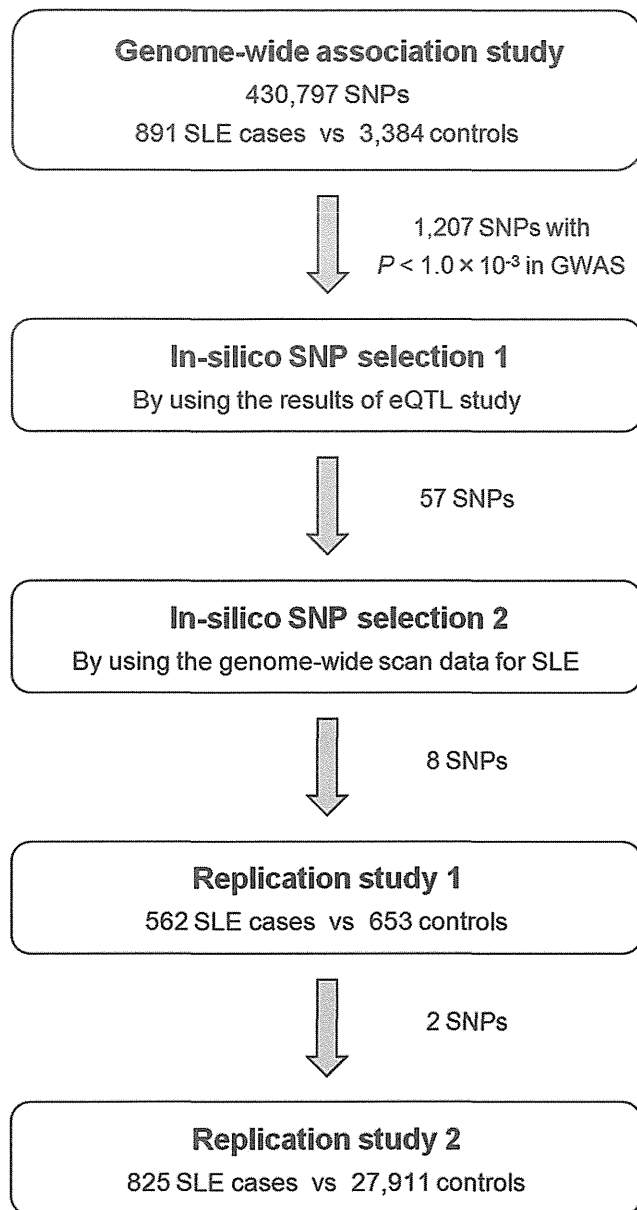
Figure 1. Design of the GWAS and multi-stage replication studies for SLE in Japanese subjects. A total of 2,278 SLE cases and 31,948 controls were enrolled. The clinical characteristics of the subjects are summarized in Table S1 and S2. Details of the genome-wide scan data for SLE referenced in the *in silico* SNP selection 2 are described elsewhere (Tahira T et al. Presented at the 59th Annual Meeting of the American Society of Human Genetics, October 21, 2009).
doi:10.1371/journal.pgen.1002455.g001

pooled analysis. As a result, 8 SNPs remained for further investigation (Table S4).

Replication studies and identification of *AFF1*

Then, we performed two-stage replication studies using independent SLE cohorts for Japanese subjects (cohort 1 with 562 SLE cases and 653 controls, and cohort 2 with 825 SLE cases and 27,911 controls). First, we evaluated the selected 8 SNPs in the replication study 1. In the replication study 2, 2 SNPs that satisfied $P < 1.0 \times 10^{-6}$ in the combined study of GWAS and replication

study 1 were further evaluated (Figure 1). Among the evaluated SNPs, we observed significant replications in the SNP located in the genomic region of the AF4/FMR2 family, member 1 gene (*AFF1*) at 4q21 (rs340630; $P = 4.6 \times 10^{-5}$ and $P = 0.0094$ in the two individual cohorts, respectively; Table 3, Table S5, and Figure 2B). The combined study for the GWAS ($P = 1.5 \times 10^{-4}$) and the replication studies demonstrated significant associations of rs340630 that satisfied the genome-wide significance threshold ($P = 8.3 \times 10^{-9}$, OR = 1.21, 95% CI 1.14–2.30).

Cis-eQTL effect of rs340630 on *AFF1* transcripts

Since the landmark SNP in the *AFF1* locus, rs340630, was prioritized through the eQTL study as an eQTL positive SNP (Table 3), we further validated its cis-eQTL effect using Epstein-Barr virus (EBV)-transfected B cell lines established from Japanese individuals (Pharma SNP Consortium (PSC) cells, $n = 62$). The correlation between rs340630 genotypes and the expression levels of *AFF1* was significant in the PSC cells stimulated with phorbol myristate acetate (PMA) ($R^2 = 0.074$, $P = 0.033$; Figure 3A). The expression levels increased with the number of SLE-risk (A) alleles. To further confirm this cis-regulatory effect, we performed allele-specific transcript quantification (ASTQ) of *AFF1*. The transcript levels of each allele were quantified by qPCR using an allele specific probe for a SNP in the 5′-untranslated region (rs340638), which was in absolute LD with rs340630 ($r^2 = 1.0$, $D' = 1.0$). We examined PSC-cells ($n = 17$) that were heterozygous for both rs340630 and rs340638. The mean ratio of each transcript (A over G allele; the A allele comprises a haplotype with the risk (A) allele of rs340630) were significantly increased to 1.07 compared to the ratio of the amount of DNA (1.00, $P = 0.012$) (Figure 3B). These results suggest that rs340630, or SNP(s) in LD with it, are a regulatory variant predisposing SLE susceptibility through increased expression levels of *AFF1*.

Expression of *AFF1* in CD4⁺ and CD19⁺ peripheral blood lymphocytes

AFF1 is known to be involved in cytogenetic translocations of acute lymphoblastic leukemia (ALL) [31]. Its fusion protein with the mixed-lineage leukemia gene (*MLL*) is implicated in the regulation of transcription and the cell cycle of lymphocytes [31]. To investigate the expression pattern of *AFF1* in normal tissues, we evaluated the transcript levels of *AFF1* in a panel of various tissues. We observed prominent expression of *AFF1* in CD4⁺ and CD19⁺ peripheral blood lymphocytes, implying an important role for *AFF1* in helper-T-cells and B-cells (Figure 3C).

Discussion

Through a GWAS and multi-staged replication studies consisting of 2,278 SLE cases and 31,948 controls in Japanese subjects, our study identified that the *AFF1* locus was significantly associated with SLE susceptibility.

As well as the identification of the novel SLE susceptibility locus, we observed significant replications of associations in the previously reported susceptibility loci. The replications were especially enriched in the loci identified through the studies in Asian populations, compared to those in European populations. Considering the ethnical heterogeneities in the epidemiology of SLE [19,20], these observations suggest the similarities in the genetic backgrounds of SLE shared within Asian populations, and also the existence of the both common and divergent genetic backgrounds encompassed between European and Asian populations.

129

Table 1. Results of a genome-wide association study for Japanese patients with SLE.

rsID[a]	Chr	Position (bp)	Cytoband	Gene	Allele[b] 1/2	No. subjects Case	No. subjects Control	Allele 1 freq. Case	Allele 1 freq. Control	OR (95%CI)	P
rs10168266	2	191,644,049	2q32	STAT4	T/C	891	3,384	0.37	0.27	1.59 (1.42–1.78)	2.7×10^{-16}
rs9501626	6	32,508,322	6p21	HLA region	A/C	891	3,381	0.20	0.12	1.86 (1.62–2.13)	1.0×10^{-18}
rs2230926	6	138,237,759	6q23	TNFAIP3	G/T	891	3,377	0.11	0.069	1.75 (1.47–2.08)	1.9×10^{-10}
rs6964720	7	75,018,280	7q11	HIP1	G/A	891	3,384	0.25	0.19	1.43 (1.27–1.63)	1.3×10^{-8}
rs2254546	8	11,381,089	8p23	BLK	G/A	891	3,384	0.78	0.72	1.42 (1.61–1.25)	4.1×10^{-8}
rs6590330	11	127,816,269	11q24	ETS1	A/G	891	3,368	0.48	0.39	1.44 (1.30–1.60)	1.3×10^{-11}

[a]SNPs that satisfied the threshold of $P < 5.0 \times 10^{-8}$ were indicated.
[b]Based on forward strand of NCBI Build 36.3.
SLE, systemic lupus erythematosus; OR, odds ratio.
doi:10.1371/journal.pgen.1002455.t001

To effectively detect the novel SLE susceptibility locus, we integrated cis-eQTL effects of the SNPs and prioritized the results of the GWAS. In addition to identifying a novel locus for SLE-susceptibility, our study demonstrated approximately 30% of confirmed SLE-susceptibility loci were comprised of cis-eQTLs. We also confirmed cis-regulatory effect of the landmark SNP in the AFF1 locus, rs340630, on AFF1 transcripts, which had been prioritized through the eQTL study. These results would suggest that accumulation of quantitative changes in gene expression would accelerate the disease onset of SLE. It would also demonstrate the validity of applying eQTL study in the search of the susceptible genes for SLE or other autoimmune diseases, as previously suggested in the study for celiac disease [24]. To our knowledge, this is one of the initial studies to successfully discover a new locus by prioritizing GWAS results using eQTLs, and should contribute to the approaches assessing genetic loci still being uncaptured by recent large-scaled GWASs due to stringent significance threshold for multiple hypothesis testing [21].

We observed prominent expression levels of AFF1 in CD4+ and CD19+ peripheral blood lymphocytes, which would imply an important role for AFF1 in helper-T-cells and B-cells. In fact, AFF1 is essential for normal lymphocyte development, as demonstrated in mice deficient for AFF1; severe reduction were observed in the thymic double positive CD4/CD8 population and the bone marrow pre-B and mature B-cell numbers [32]. The risk A allele of rs340630 demonstrated a cis-eQTL effect on the AFF1 transcript with enhanced expression levels. As the AFF1 locus was also demonstrated as an eQTL in primary liver cells [33], the cis-regulatory effect may hold in primary cells as well as lymphoblastoid cells used in the present study. However, because the mechanism of transcriptional regulation is substantially different among cell types [34], cell-type specific analyses including those for primary T-cells and B-cells are needed for understanding the precise role of AFF1 variant in primary lymphocytes. Although further functional investigation is necessary, our observation suggested that AFF1 is involved in the etiology of SLE through the regulation of development and activity of lymphocytes. It is of note that AFF3, which also belongs to the AF4/FMR2 family, is associated with susceptibility to autoimmune diseases [35].

One of our study's limitations is the selection of SNPs for the replication study using the results of the pooled DNA approach [30], which used a different genotyping platform from that of the present GWAS. Moreover, the association signals based on Silhouette scores in pooled analysis would be less reliable compared to those based on individual genotyping. Since direct comparisons of the association signals of the same single SNPs

between the studies would be difficult due to these issues, we adopted the complementary approach that referred the association signals of the multiple SNPs in the pooled analysis for each of the single SNPs in the GWAS, taking account of LD and physical distances between the SNPs. However, there would exist a possibility that the variant(s) truly associated with SLE was left not to be examined in the replication study. It should be noted that only 1 SNP among the 8 selected SNPs yielded the significant association with SLE, although further enrichments of the significant associations might be anticipated. To elucidate effectiveness and limitation of our approach, further assessments of the studies on the remaining loci would be desirable. It should also be noted that the control-case ratio of the subjects were relatively high in the replication study 2 (= 33.8), and this disproportionate ratio could have induced potential bias on the results of the association analysis of the SNPs. However, considering the homogeneous ancestries of the Japanese population [27] and that principal component analysis did not demonstrate significant population stratification in the control subjects of the replication study 2 (data not shown), the bias owing to population stratification might not be substantial.

In summary, through a GWAS and multi-staged replication studies in a Japanese population integrating eQTL study, our study identified AFF1 as a novel susceptibility locus for SLE.

Materials and Methods

Subjects

We enrolled 2,278 systemic lupus erythematosus (SLE) cases and 31,948 controls. SLE cases enrolled in the genome-wide association study (GWAS) (n = 891) or part of the 2nd replication study (n = 83) were collected from 12 medical institutes in Japan under the support of the autoimmune disease study group of Research in Intractable Diseases, Japanese Ministry of Health, Labor and Welfare: Hokkaido University Graduate School of Medicine, Tohoku University Graduate School of Medicine, the University of Tokyo, Keio University School of Medicine, Juntendo University School of Medicine, University of Occupational and Environmental Health, University of Tsukuba, Tokyo Medical and Dental University, National Center for Global Health and Medicine, Nagasaki University, Wakayama Medical University, and Jichi Medical University. SLE cases (n = 562) and controls (n = 653) enrolled in the 1st replication study were collected from Kyushu University. Some of the SLE cases (n = 742) and controls (n = 27,911) enrolled in the 2nd replication study were collected from Kyoto University, Tokyo Women's
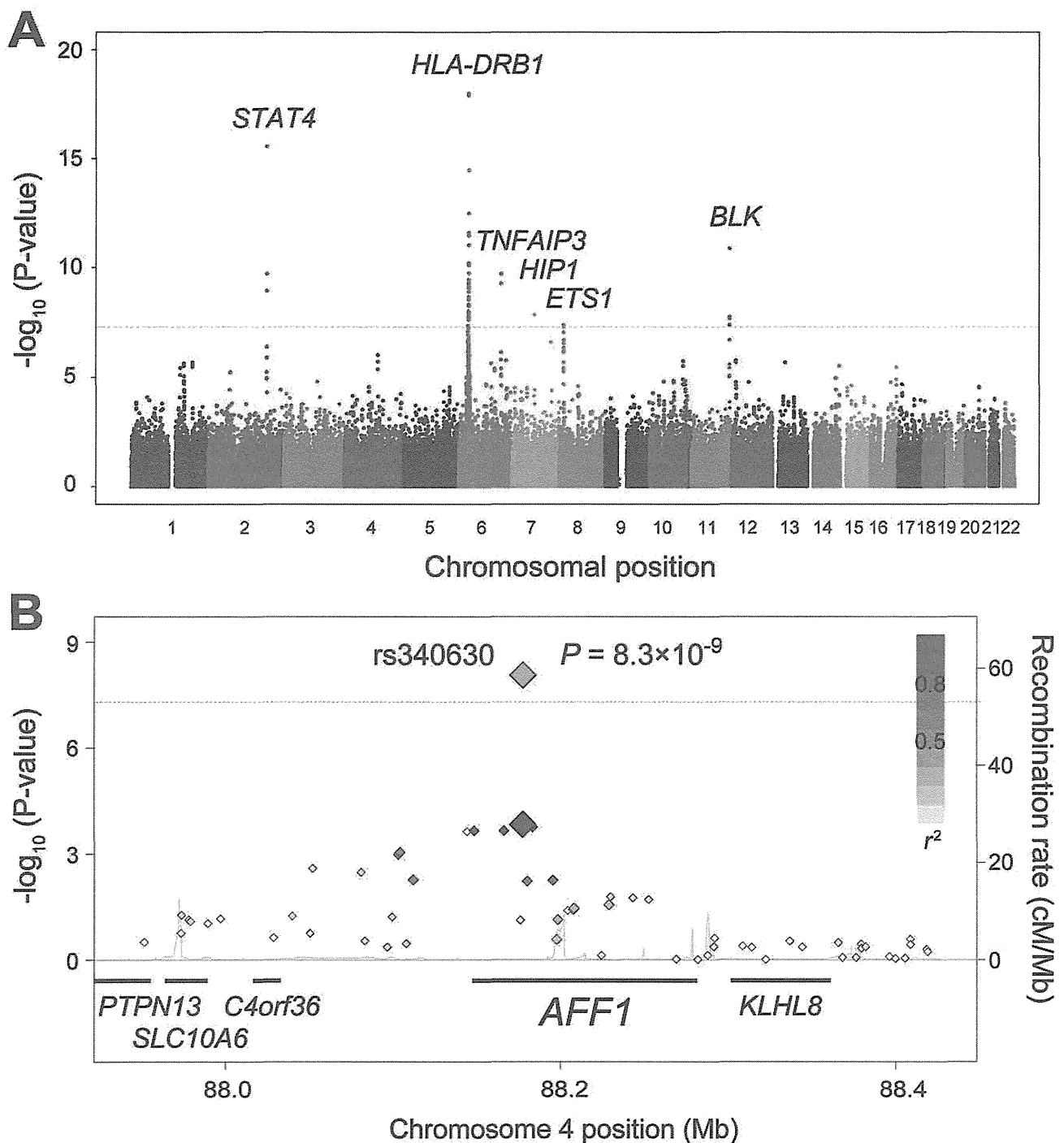
130

Figure 2. Associations of the *AFF1* locus with SLE. (A) A chromosomal plot of *P*-values in GWAS for SLE. (B) A regional plot in the *AFF1* locus. Diamond-shaped data points represent $-\log_{10}$ (*P*-values) of the SNPs. Large-sized points indicate the *P*-values of the landmark SNP, rs340630 (green for the combined study and red for the GWAS). Density of red color represents r^2 values with rs340630. Blue line represents recombination rates. Lower part indicates RefSeq genes. Gray dashed horizontal lines represent the threshold of $P=5.0\times10^{-8}$. The plots were drawn using SNAP, version 2.1 [47].
doi:10.1371/journal.pgen.1002455.g002

Medical University, the University of Tokyo, and the BioBank Japan Project [36]. All subjects were of Japanese origin and provided written informed consent. SLE cases met the revised American College of Rheumatology (ACR) criteria for SLE [37]. Control subjects were confirmed to be free of autoimmune disease. Some of the SLE cases were included in our previous studies [38–40]. Details of the subjects are summarized in Table S1 and S2. This research project was approved by the ethical committees of the University of Tokyo, RIKEN, and affiliated medical institutes.

Table 2. Associations among previously reported SLE-related loci.

rsID	Chr	Position (bp)	Cytoband	Gene	Allele[a] 1/2	Allele 1 freq. Case	Allele 1 freq. Control	OR (95%CI)	P	eQTL[b]	Identified by the studies in[c] Caucasians	Identified by the studies in[c] Asians
rs2205960	1	171,458,098	1q25	TNFSF4	T/G	0.23	0.18	1.35 (1.19–1.54)	3.0×10^{-6}		+	
rs3024505	1	205,006,527	1q32	IL10	A/G	0.019	0.014	1.34 (0.90–2.00)	0.15		+	
rs13385731	2	33,555,394	2p22	RASGRP3	C/T	0.90	0.87	1.37 (1.15–1.64)	6.0×10^{-4}	+		+
rs10168266	2	191,644,049	2q32	STAT4	T/C	0.37	0.27	1.59 (1.42–1.78)	2.7×10^{-16}		+	
rs6445975	3	58,345,217	3p14	PXK	G/T	0.25	0.23	1.09 (0.96–1.23)	0.18	+	+	
rs10516487	4	102,970,099	4q24	BANK1	G/A	0.91	0.89	1.28 (1.07–1.53)	0.0070		+	
rs10036748	5	150,438,339	5q33	TNIP1	T/C	0.75	0.72	1.16 (1.03–1.31)	0.014			+
rs9501626	6	32,508,322	6p21	HLA-DRB1	A/C	0.20	0.12	1.86 (1.62–2.13)	1.0×10^{-18}		+	
rs548234	6	106,674,727	6q21	PRDM1	C/T	0.40	0.34	1.30 (1.16–1.44)	2.3×10^{-6}	+	+	
rs2230926	6	138,237,759	6q23	TNFAIP3	G/T	0.11	0.069	1.75 (1.47–2.08)	1.9×10^{-10}	+	+	
rs849142	7	28,152,416	7p15	JAZF1	C/T	0.999	0.999	2.72 (0.25–29.8)	0.41		+	
rs4917014	7	50,276,409	7p12	IKZF1	T/G	0.58	0.53	1.24 (1.11–1.38)	8.1×10^{-5}			+
rs6964720	7	75,018,280	7q11	HIP1	G/A	0.25	0.19	1.43 (1.27–1.62)	1.3×10^{-8}			+
rs4728142	7	128,361,203	7q32	IRF5	A/G	0.16	0.11	1.48 (1.28–1.72)	2.4×10^{-7}	+	+	
rs2254546	8	11,381,089	8p23	BLK	G/A	0.78	0.72	1.42 (1.25–1.61)	4.1×10^{-8}	+	+	
rs1913517	10	49,789,060	10q11	WDFY4	A/G	0.32	0.28	1.20 (1.07–1.35)	0.0013			+
rs4963128	11	579,564	11p15	KIAA1542	T/C	0.98	0.97	1.58 (1.03–2.44)	0.038	+	+	
rs2732552	11	35,041,168	11p13	PDHX, CD44	T/C	0.75	0.73	1.13 (1.00–1.27)	0.056		+	
rs4639966	11	118,078,729	11q23	Intergenic	T/C	0.32	0.28	1.22 (1.09–1.36)	7.3×10^{-4}			+
rs6590330	11	127,816,269	11q24	ETS1	A/G	0.48	0.39	1.44 (1.30–1.60)	1.3×10^{-11}			+
rs1385374	12	127,866,647	12q24	SLC15A4	T/C	0.19	0.16	1.21 (1.06–1.38)	0.0057			+
rs7329174	13	40,456,110	13q14	ELF1	G/A	0.30	0.25	1.32 (1.18–1.49)	2.2×10^{-6}			+
rs7197475	16	30,550,368	16p11	Intergenic	T/C	0.12	0.10	1.20 (1.02–0.41)	0.031			+
rs11150610	16	31,241,737	16p11	ITGAM	C/A	0.20	0.19	1.07 (0.94–1.22)	0.32	+	+	
rs12949531	17	13,674,531	17p12	Intergenic	T/C	0.28	0.27	1.02 (0.91–1.15)	0.73		+	
rs463426	22	20,139,185	22q11	HIC2,UBE2L3	T/C	0.52	0.48	1.20 (1.08–1.33)	6.1×10^{-4}		+	

[a]Based on forward strand of NCBI Build 36.3.
[b]Defined using gene expression data measured in lymphoblastoid B cell lines [28].
[c]Based on the previously reported studies for SLE susceptibility loci [3–18].
SLE, systemic lupus erythematosus; OR, odds ratio; eQTL, expression quantitative trait locus; GWAS, genome-wide association study.
doi:10.1371/journal.pgen.1002455.t002

Genotyping and quality control

In GWAS, 946 SLE cases and 3,477 controls were genotyped using Illumina HumanHap610-Quad and Illumina Human-Hap550v3 Genotyping BeadChips (Illumina, CA, USA), respectively. After the exclusion of 47 SLE cases and 92 controls with call rates <0.98, SNPs with call rates <0.99 in SLE cases or controls,

Table 3. Results of combined study for Japanese patients with SLE.

rsID	Chr	Position (bp)	Cytoband	Gene	Allele 1/2	Stage	No. subjects Case	No. subjects Control	Allele 1 freq. Case	Allele 1 freq. Control	OR (95%CI)	P	eQTL[a]
rs340630	4	88,177,419	4q21	AFF1	A/G	GWAS	891	3,383	0.56	0.51	1.22 (1.10–1.36)	1.5×10^{-4}	+
						Replication study 1	550	646	0.57	0.49	1.40 (1.19–1.64)	4.6×10^{-5}	
						Replication study 2	820	27,911	0.56	0.53	1.14 (1.03–1.26)	0.0094	
						Combined study	2,261	31,940	0.56	0.52	1.21 (1.14–1.30)	8.3×10^{-9}	

[a]Defined using gene expression data measured in lymphoblastoid B cell lines [28].
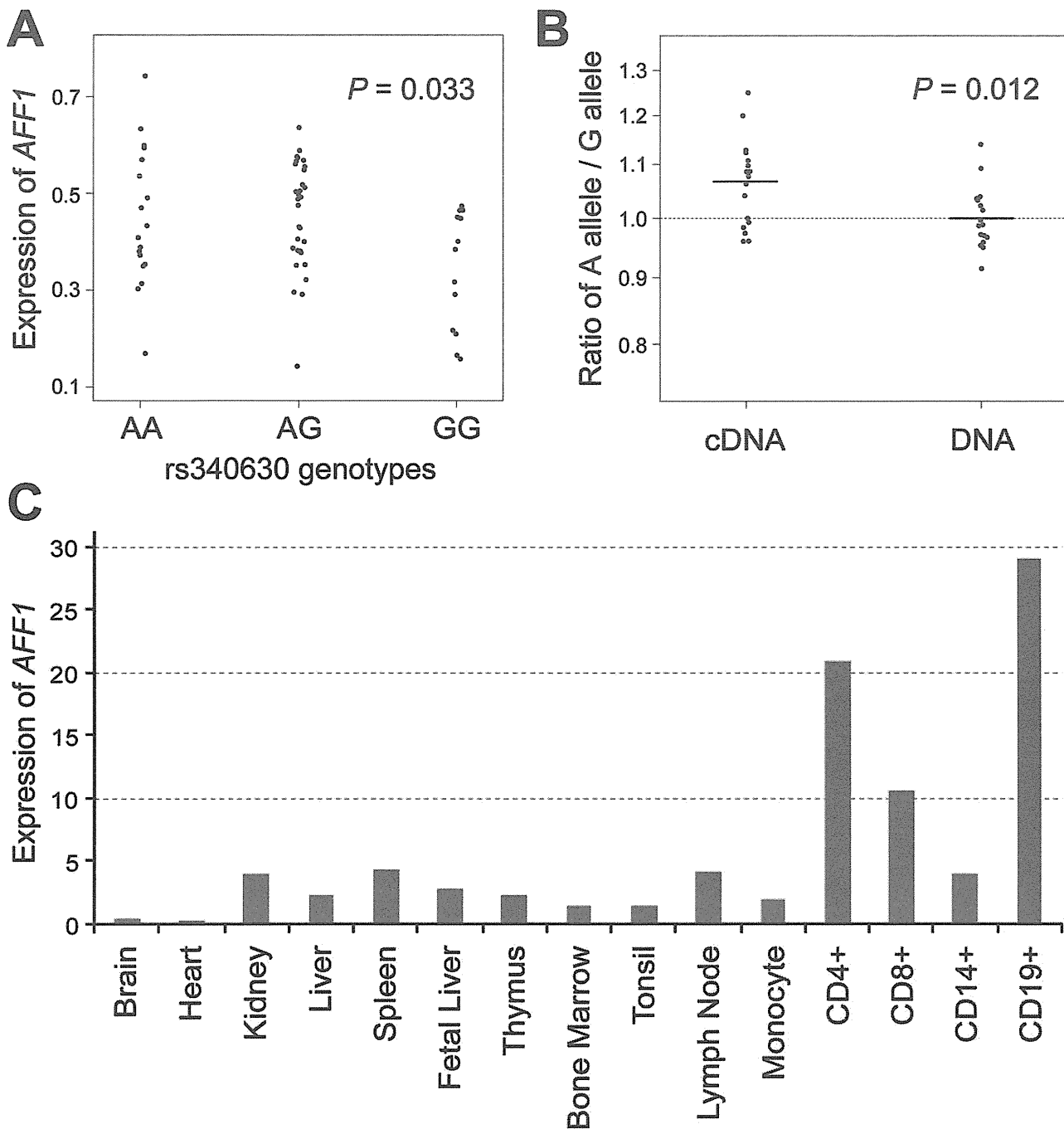doi:10.1371/journal.pgen.1002455.t003

132

Figure 3. Association of rs340630 with *AFF1* expression. (A) Correlation between rs340630 genotypes and transcript levels of *AFF1* (NM_001166693) in EBV-transfected cell lines (*n* = 62) stimulated with PMA. (B) Allele-specific quantification (ASTQ) of *AFF1* transcripts. Allele specific-probes for rs340638 were used for quantification by qPCR. The ratios of A allele over G allele for the amounts of both cDNAs and DNAs were plotted in log scale for each cell line. (C) *AFF1* expression in various tissues. Transcripts levels of *AFF1* were quantified by qPCR and were normalized by *GAPDH* levels.
doi:10.1371/journal.pgen.1002455.g003

non-autosomal SNPs, and SNPs not shared between SLE cases and controls, were excluded. We excluded 7 closely related SLE cases in a 1st or 2nd degree of kinship based on identity-by-descent estimated using PLINK version 1.06 [41]. We then excluded 1 SLE cases and 1 controls whose ancestries were estimated to be distinct from East-Asian populations using PCA performed along with the genotype data of Phase II HapMap populations (release 24) [29] using EIGENSTRAT version 2.0 [42]. Subsequently,

SNPs with minor allele frequencies <0.01 in SLE cases or controls, SNPs with exact *P*-values of Hardy-Weinberg equilibrium test <1.0×10^{-6} in controls, or SNPs with ambiguous cluster plots were excluded. Finally, 430,797 SNPs for 891 SLE cases and 3,384 controls were obtained. Genotyping of SNPs in replication studies was performed using TaqMan Assay or Illumina HumanHap610-Quad Genotyping BeadChip (Illumina, CA, USA).

133

Association analysis of the SNPs

Association of SNPs in GWAS and replication studies were tested with Cochran-Armitage's trend test. Combined analysis was performed with Mantel-Haenzel method. Associations of previously reported SLE susceptibility loci [3–18] were evaluated using the results of the GWAS. Genotype imputation was performed for non-genotyped SNPs using MACH version 1.0 [43] with Phase II HapMap East-Asian individuals as references [29], as previously described [44]. All imputed SNPs demonstrated imputation scores, *Rsq*, >0.70.

eQTL study

We analyzed gene expression data previously measured in lymphoblastoid B cell lines from Phase II HapMap East-Asian individuals using Illumina's human whole-genome expression array (WG-6 version 1) (accession number; GSE6536) [28]. Expression data were normalized across the individuals. We used BLAST to map 47,294 Illumina array probes onto human autosomal reference genome sequences (Build 36). We discarded probes mapped with expectation values smaller than 0.01 to multiple loci, or for which there was polymorphic HapMap SNP(s) inside the probe. Then, 19,047 probes with exact matches to a unique locus with 100% identity and with a mean signal intensity greater than background were obtained. Genotype data of HapMap individuals were obtained for SNPs included in the GWAS. Associations of SNP genotypes (coded as 0, 1, and 2) with expression levels of each of the cis-eQTL probes (located within ±300 kbp regions of the SNPs) were evaluated using linear regression assuming additive effects of the genotypes on the expression levels. Considering the significant overlap between eQTL and genetic loci responsible for autoimmune diseases [24], we applied relatively less stringent multiple testing threshold of FDR *Q*-values<0.2 for the definition of eQTL. SNPs that exhibited this threshold with any of the corresponding cis-eQTL probes were denoted as eQTL positive.

Selection of SNPs enrolled in the replication studies

In order to select SNPs for further replication studies, we firstly integrated the results of GWAS and eQTL study. SNPs that satisfied $P<1.0\times10^{-4}$ in GWAS, or the SNPs that satisfied $1.0\times10^{-4}\leq P<1.0\times10^{-3}$ in GWAS and denoted as eQTL positive, were selected. Among these, SNPs most significantly associated in each of the genomic loci and not included in the previously reported SLE susceptibility loci [3–18] were further evaluated.

Then, the results of the concurrently proceeding genome-wide scan for SLE in the Japanese subjects using a pooled DNA approach were referred (Tahira T et al. Presented at the 59th Annual Meeting of the American Society of Human Genetics, October 21, 2009). In the scan, DNA collected from 447 SLE cases and 680 controls of Japanese origin were pooled respectively, and genotyped using GeneChip Human Mapping 500K Array Set (Affymetrix, CA, USA). SNPs were ranked according to the Silhouette scores estimated based on relative allele scores (RAS) between SLE cases and controls, and rank-based *P*-values were assigned [30]. By referring to association signals in multiple neighboring SNPs in the pooled analysis, we selected SNPs for replication study 1. Namely, if the SNP of interest was in LD ($r^2>0.5$) or was located within ±100 kbp of SNPs showing association signals in the pooled analysis (rank-based $P<0.01$), it would be selected. SNPs that satisfied $P<1.0\times10^{-6}$ in the combined study of GWAS and replication study 1 were further evaluated in replication study 2 (Figure 1).

Quantification of *AFF1* expression

EBV-transformed lymphoblastoid cell lines ($n=62$) were established by Pharma SNP Consortium (Tokyo, Japan) using peripheral blood lymphocytes of Japanese healthy individuals. Cells were incubated for 2 h in medium alone (RPMI 1640 medium containing 10% FBS, 1% penicillin, and 1% streptomycin) or with 100 ng/ml PMA. Conditions for cell stimulation were optimized before the experiment as previously described [45]. Cells were then harvested and total RNA was isolated using an RNeasy Mini Kit (Qiagen) with DNase treatment. Total RNA (1 μg) was reverse transcribed using TaqMan Gold RT-PCR reagents with random hexamers (Applied Biosystems). Real-time quantitative PCR was performed in triplicate using an ABI PRISM 7900 and TaqMan gene expression assays (Applied Biosystems). Specific probes (Hs01089428_m1) for transcript of *AFF1* (NM_001166693) were used. Expression of *AFF1* in various tissues was also quantified using Premium Total RNA (Clontech). The data were normalized to *GAPDH* levels. *GUS* levels were also evaluated for internal control, and similar results were obtained. Correlation coefficient, R^2, between rs340630 genotypes and transcript levels of *AFF1* was evaluated.

Allele-specific transcript quantification (ASTQ)

ASTQ of *AFF1* in PSC cells was performed as previously described [46]. DNAs were extracted by using a DNeasy Kit (QIAGEN). RNA extraction and cDNA preparation were performed as described above. For PSC cells ($n=17$) that were heterozygous for both rs340630 (the landmark SNP of GWAS) and rs340638 (located in the 5′-untranslated region of *AFF1* and in absolute LD with rs340630), expression levels of *AFF1* were quantified by qPCR on an ABI Prism 7900 using a custom-made TaqMan MGB-probe set for rs340638. Primer sequences were 5′-CTAACTGTGGCCCGCGTTG-3′ and 5′-CCCGGCGCA-GTTTCTGAG-3′. The probe sequences were 5′-VIC-CGAA-GACCGCCAGCGCCCAAC-TAMRA-3′ and 5′-FAM-CGAA-GACCGCCGGCGCCCAA-TAMRA-3′. Ct values of VIC and FAM were obtained for genomic DNA and cDNA samples after 40 cycles of real-time PCR. We also prepared genomic DNA of samples homozygous for each allele and mixed them at different ratios (2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2) to create a standard curve by plotting Ct values of VIC/FAM against the allelic ratio of VIC/FAM for each mixture. Using the standard curve, we calculated the allelic ratios for each genomic DNA and cDNA samples. We measured each sample in quadruplicate in one assay; tests were independently repeated twice.

Web resources

The URLs for data presented herein are as follows.

NCBI GEO, http://www.ncbi.nlm.nih.gov/geo

BioBank Japan Project, http://biobankjp.org

PLINK software, http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml

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

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

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

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

Supporting Information

Figure S1 Principal component analysis (PCA) plot of the subjects. PCA plot of subjects enrolled in the GWAS for SLE. SLE cases and the controls enrolled in the GWAS are plotted based on

eigenvectors 1 and 2 obtained from the PCA using EIGEN-STRAT version 2.0 [42], along with European (CEU), African (YRI), Japanese (JPT), and Chinese (CHB) individuals obtained from the Phase II HapMap database (release 22) [29]. Subjects who were estimated to be outliers in terms of ancestry from East-Asian (JPT+CHB) clusters and excluded from the study are indicated by black arrows.
(TIF)

Figure S2 Quantile-Quantile plot (QQ-plot) of *P*-values in the GWAS for SLE. The horizontal axis indicates the expected $-\log_{10}$ (*P*-values). The vertical axis indicates the observed $-\log_{10}$ (*P*-values). The QQ-plot for the *P*-values of all SNPs that passed the quality control criteria is indicated in blue. The QQ-plot for the *P*-values after the removal of SNPs included in the previously reported SLE susceptibility loci is indicated in black. The gray line represents $y = x$. The SNPs for which the *P*-value was smaller than 1.0×10^{-15} are indicated at the upper limit of the plot.
(TIF)

Table S1 Basal characteristics of cohorts.
(DOC)

Table S2 Frequency of clinical characteristics of SLE in this GWAS.
(DOC)

Table S3 Distributions of eQTL positivity rates of the SNPs.
(DOC)

Table S4 Results of replication study 1 for Japanese patients with SLE.
(DOC)

Table S5 Results of replication studies 1 and 2 for Japanese patients with SLE.
(DOC)

Author Contributions

Conceived and designed the experiments: Y. Okada, K. Shimane, Y. Kochi, A. Suzuki, M. Kubo, N. Kamatani, R. Yamada, Y. Nakamura, K. Yamamoto. Performed the experiments: Y. Kochi, N. Hosono, M. Kubo, K. Myouzen, T. Horiuchi, T. Tahira, K. Hayashi. Analyzed the data: Y. Okada, A. Takahashi, T. Tsunoda, K. Higasa, R. Yamada. Contributed reagents/materials/analysis tools: T. Horita, T. Atsumi, T. Koike, T. Ishii, A. Okamoto, K. Fujio, M. Hirakata, H. Amano, Y. Takasaki, Y. Kondo, S. Ito, T. Sumida, K. Takada, A. Mimori, K. Saito, Y. Tanaka, M. Kamachi, N. Nishimoto, Y. Kawaguchi, C. Terao, K. Ohmura, T. Mimori, F. Matsuda, O.W. Mohammed, K. Matsuda, K. Shimane, K. Ikari, H. Yamanaka. Wrote the paper: Y. Okada, K. Shimane, Y. Kochi, A. Suzuki, R. Yamada, K. Yamamoto.

References

1. Lipsky PE (2001) Systemic lupus erythematosus: an autoimmune disease of B cell hyperactivity. Nat Immunol 2: 764–766.
2. Sestak AL, Shaver TS, Moser KL, Neas BR, Harley JB (1999) Familial aggregation of lupus and autoimmunity in an unusual multiplex pedigree. J Rheumatol 26: 1495–1499.
3. Sigurdsson S, Nordmark G, Goring HH, Lindroos K, Wiman AC, et al. (2005) Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. Am J Hum Genet 76: 528–537.
4. Graham RR, Kozyrev SV, Baechler EC, Reddy MV, Plenge RM, et al. (2006) A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. Nat Genet 38: 550–555.
5. Graham RR, Kyogoku C, Sigurdsson S, Vlasova IA, Davies LR, et al. (2007) Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. Proc Natl Acad Sci U S A 104: 6758–6763.
6. 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.
7. Cunninghame Graham DS, Graham RR, Manku H, Wong AK, Whittaker JC, et al. (2008) Polymorphism at the TNF superfamily gene TNFSF4 confers susceptibility to systemic lupus erythematosus. Nat Genet 40: 83–89.
8. Nath SK, Han S, Kim-Howard X, Kelly JA, Viswanathan P, et al. (2008) A nonsynonymous functional variant in integrin-alpha(M) (encoded by ITGAM) is associated with systemic lupus erythematosus. Nat Genet 40: 152–154.
9. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, et al. (2008) Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat Genet 40: 204–210.
10. Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, et al. (2008) Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nat Genet 40: 211–216.
11. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, et al. (2008) Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. N Engl J Med 358: 900–909.
12. Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, et al. (2008) Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Nat Genet 40: 1059–1061.
13. Musone SL, Taylor KE, Lu TT, Nititham J, Ferreira RC, et al. (2008) Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. Nat Genet 40: 1062–1064.
14. Han JW, Zheng HF, Cui Y, Sun LD, Ye DQ, et al. (2009) Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat Genet 41: 1234–1237.

15. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, et al. (2009) A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nat Genet 41: 1228–1233.
16. Yang W, Shen N, Ye DQ, Liu Q, Zhang Y, et al. (2010) Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. PLoS Genet 6: e1000841. doi:10.1371/journal.pgen.1000841.
17. Lessard CJ, Adrianto I, Kelly JA, Kaufman KM, Grundahl KM, et al. (2011) Identification of a systemic lupus erythematosus susceptibility locus at 11p13 between PDHX and CD44 in a multiethnic study. Am J Hum Genet 88: 83–91.
18. Yang J, Yang W, Hirankarn N, Ye DQ, Zhang Y, et al. (2011) ELF1 is associated with systemic lupus erythematosus in Asian populations. Hum Mol Genet 20: 601–607.
19. Hopkinson ND, Doherty M, Powell RJ (1994) Clinical features and race-specific incidence/prevalence rates of systemic lupus erythematosus in a geographically complete cohort of patients. Ann Rheum Dis 53: 675–680.
20. Danchenko N, Satia JA, Anthony MS (2006) Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. Lupus 15: 308–318.
21. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42: 565–569.
22. Raychaudhuri S, Plenge RM, Rossin EJ, Ng AC, Purcell SM, et al. (2009) Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet 5: e1000534. doi:10.1371/journal.pgen.1000534.
23. Cantor RM, Lange K, Sinsheimer JS (2010) Prioritizing GWAS results: A review of statistical methods and recommendations for their application. Am J Hum Genet 86: 6–22.
24. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, et al. (2010) Multiple common variants for celiac disease influencing immune gene expression. Nat Genet 42: 295–302.
25. Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M (2009) Mapping complex disease traits with global gene expression. Nat Rev Genet 10: 184–194.
26. 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.
27. Yamaguchi-Kabata Y, Nakazono K, Takahashi A, Saito S, Hosono N, et al. (2008) Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. Am J Hum Genet 83: 445–456.
28. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, et al. (2007) Population genomics of human gene expression. Nat Genet 39: 1217–1224.
29. The International HapMap Consortium (2003) The International HapMap Project. Nature 426: 789–796.

135

30. Pearson JV, Huentelman MJ, Halperin RF, Tembe WD, Melquist S, et al. (2007) Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide-polymorphism association studies. Am J Hum Genet 80: 126–139.

31. Xia ZB, Popovic R, Chen J, Theisler C, Stuart T, et al. (2005) The MLL fusion gene, MLL-AF4, regulates cyclin-dependent kinase inhibitor CDKN1B (p27kip1) expression. Proc Natl Acad Sci U S A 102: 14028–14033.

32. Isnard P, Core N, Naquet P, Djabali M (2000) Altered lymphoid development in mice deficient for the mAF4 proto-oncogene. Blood 96: 705–710.

33. Schadt EE, Molony C, Chudin E, Hao K, Yang X, et al. (2008) Mapping the genetic architecture of gene expression in human liver. PLoS Biol 6: e107. doi:10.1371/journal.pbio.0060107.

34. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, et al. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473: 43–49.

35. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, et al. (2010) Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet 42: 508–514.

36. Nakamura Y (2007) The BioBank Japan Project. Clin Adv Hematol Oncol 5: 696–697.

37. Hochberg MC (1997) Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 40: 1725.

38. 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.

39. Shimane K, Kochi Y, Horita T, Ikari K, Amano H, et al. (2010) The association of a nonsynonymous single-nucleotide polymorphism in TNFAIP3 with systemic lupus erythematosus and rheumatoid arthritis in the Japanese population. Arthritis Rheum 62: 574–579.

40. Myouzen K, Kochi Y, Shimane K, Fujio K, Okamura T, et al. (2010) Regulatory polymorphisms in EGR2 are associated with susceptibility to systemic lupus erythematosus. Hum Mol Genet 19: 2313–2320.

41. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.

42. 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.

43. Li Y, Willer C, Sanna S, Abecasis G (2009) Genotype imputation. Annu Rev Genomics Hum Genet 10: 387–406.

44. Okada Y, Takahashi A, Ohmiya H, Kumasaka N, Kamatani Y, et al. (2011) Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. Hum Mol Genet 20: 1224–1231.

45. Aikawa Y, Yamamoto M, Yamamoto T, Morimoto K, Tanaka K (2002) An anti-rheumatic agent T-614 inhibits NF-kappaB activation in LPS- and TNF-alpha-stimulated THP-1 cells without interfering with IkappaBalpha degradation. Inflamm Res 51: 188–194.

46. Akamatsu S, Takata R, Ashikawa K, Hosono N, Kamatani N, et al. (2010) A functional variant in NKX3.1 associated with prostate cancer susceptibility down-regulates NKX3.1 expression. Hum Mol Genet 19: 4265–4272.

47. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, et al. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24: 2938–2939.

136

International Journal of Cardiology xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

International Journal of Cardiology

journal homepage: www.elsevier.com/locate/ijcard
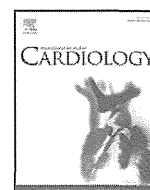
Increased aortic wave reflection and smaller pulse pressure amplification in smokers and passive smokers confirmed by urinary cotinine levels: The Nagahama Study ☆,☆☆

Yasuharu Tabara [a],*, Yoshimitsu Takahashi [b], Kazuya Setoh [a], Shigeo Muro [c], Takahisa Kawaguchi [a], Chikashi Terao [a], Shinji Kosugi [d], Akihiro Sekine [e], Ryo Yamada [a], Michiaki Mishima [c], Takeo Nakayama [b], Fumihiko Matsuda [a] on behalf of the Nagahama Study Group

[a] Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan
[b] Department of Health Informatics, Kyoto University School of Public Health, Kyoto, Japan
[c] Department of Respiratory Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan
[d] Department of Medical Ethics and Medical Genetics, Kyoto University School of Public Health, Kyoto, Japan
[e] EBM Research Center, Kyoto University Graduate School of Medicine, Kyoto, Japan

ARTICLE INFO

ABSTRACT

Background: Central blood pressure (cSBP) is suggested to be a better predictor of cardiovascular risk than brachial BP. Although brachial BP levels among smokers have been reported to be the same or somewhat lower than those in nonsmokers, it is suggested that smoking might have a substantial impact on cSBP.
Methods: We conducted a cross-sectional study to clarify the association of smoking habit with arterial tone and cSBP in a general population of 8557 participants using urinary cotinine levels as an objective marker of smoking intensity. Absolute pressure of the late systolic peak (SBP2) was obtained by calibrating the radial waveform with brachial systolic BP (bSBP) and considered to be the cSBP.
Results: Confounding factor-adjusted mean pulse pressure amplification (PPa = bSBP − cSBP) was significantly smaller in habitual smokers (current, 9.3 ± 0.15; past, 10.2 ± 0.13; never, 10.6 ± 0.10 mm Hg; $p < 0.001$). Further, among smokers, PPa was linearly decreased with increasing urinary cotinine quartile (Q1, 10.9 ± 0.38; Q2, 10.9 ± 0.39; Q3, 10.4 ± 0.39; Q4, 9.7 ± 0.41 mm Hg; $p = 0.020$). Multiple linear regression analysis identified both smoking habit ($p = 0.003$) and urinary cotinine levels ($p = 0.008$) as independent determinants of PPa. Urinary cotinine was also detected in a small fraction of never smokers (1.8%). These passive smokers showed a smaller PPa (passive smoker, 9.4 ± 0.4; never smoker, 10.4 ± 0.12 mm Hg, $p = 0.020$) but not bSBP (122.7 ± 0.6, 123.1 ± 0.2 mm Hg, $p = 0.474$).
Conclusions: Not only habitual smoking but also passive smoking had harmful effects on AIx and central BP. Our results strongly emphasize the importance of avoiding passive smoking to the prevention of cardiovascular risks of which the subject is likely unaware.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Hypertension is a major health burden, particularly in developed countries. Although accumulated clinical and epidemiological evidences for high blood pressure (BP) risks have been based on BP measured at brachial artery, recent epidemiological studies suggest that cardiovascular risk might more closely correlated with central aortic

systolic pressure (cSBP) [1–3]. Further, a large-scale clinical trial examining the impact of two different BP-lowering drugs, namely beta-blocker atenolol-based therapy and calcium channel blocker amlodipine therapy, clearly showed that differences in cSBP were more closely associated with cardiac outcome than differences in brachial systolic BP (bSBP), which in fact showed equivalent levels between the two treatment arms [4]. These results for the superiority of cSBP in estimating BP risks emphasize the importance of identifying factors that might affect the difference between bSBP and cSBP, i.e. pulse pressure amplification (PPa).

Smoking is another factor which strongly increases cardiovascular (CV) risk. Because BP levels among smokers were suggested to be the same as or somewhat lower than those in nonsmokers [5], the harmful effects of smoking on cardiovascular outcome were thought to be independent of BP risks. However, recent epidemiological study in a

general population reported that augmentation index (AIx) was increased in habitual smokers, a change which can result in a higher central BP [6], and higher AIx and cSBP, but not bSBP, have been observed in persons with a smoking habit [7]. Further, several experimental trials have shown that cigarette or cigar smoking [8,9], even passive smoking [10], acutely increased both brachial and central BP, as well as arterial tone as measured by aortic pulse wave velocity or AIx of the arterial waveform, and that the effects of smoking on arteries lasted for at least 2 h [9]. Given these previous findings, we hypothesized that the increased CV risk in smokers might be partially due to elevated aortic tone and consequent raised central BP, which could not be detected by a simple brachial BP monitoring.

Here, we conducted a cross-sectional study to clarify the association of smoking habit with arterial tone and PPa in our large-scale general population sample in which urinary cotinine levels are recorded as an objective marker for smoking intensity.

2. Methods

2.1. Study subjects

The study subjects consisted of 8557 of a total of 9804 participants recruited from 2008 to 2010 to the Nagahama Prospective Genome Cohort for the Comprehensive Human Bioscience (The Nagahama Study) who were free from any symptomatic cardiovascular diseases and whose fasting plasma and urine samples were available. The Nagahama Study cohort was recruited from the general population living in Nagahama City, a largely rural city of 125,000 inhabitants in Shiga Prefecture, located in the center of Japan. Smoking habit was investigated using a structure questionnaire, and study subjects were classified into three sub-groups according to their smoking habit; current smoker, now smoking cigarettes every day or some days; past smoker, not smoking at the time of the interview, but have an experience of continuous smoking; and never smoker, having no experience of smoking. Medical history was also investigated using a structured questionnaire. All study procedures were approved by the ethics committee of Kyoto University Graduate School of Medicine.

2.2. Measurement of BP and augmentation index (AIx)

Brachial BP and radial arterial waveform were measured simultaneously (HEM-9000AI: Omron Healthcare, Kyoto, Japan) after 5 min rest in the sitting position. Briefly, brachial BP was measured on the right upper arm using a cuff-oscillometric device, and the radial arterial waveform was simultaneously obtained from the left wrist using a multi-element tonometric sensor. Radial AIx was calculated from the waveform as the ratio of the height of the late systolic peak to that of the first peak [11–13]. Absolute pressure of the late systolic peak (SBP2) was obtained by calibrating the radial waveform with bSBP and considered to be the cSBP. The validity of SBP2 for use in estimating cSBP has been demonstrated by invasive simultaneous measurement of ascending aorta and radial artery pressure [14,15]. We also reported that radial SBP2 was closely related to cSBP, as calculated using the widely used generalized transfer function [16]. The difference between bSBP and cSBP was used as an index of brachial-to-central PPa (PPa = bSBP − cSBP). Heart rate (HR) was also measured simultaneously.

2.3. Measurement of urinary cotinine level

A urine sample was collected in the morning and immediately frozen until used for measurement. Cotinine levels were measured at a commercial laboratory (BML Inc., Tokyo, Japan) by gas chromatography–mass spectrometry (QP-5050, Shimadzu Corporation, Kyoto, Japan). The inter- and intra-assay coefficients of variation of the urinary cotinine assay were 1.97% and 2.70%, respectively.

2.4. Statistical analysis

Values are the mean ± standard deviation unless otherwise specified. Differences in numeric variables by smoking habit were assessed by analysis of variance, while the frequency of differences was assessed by the chi-squared test. Factors independently associated with AIx and ΔSBP2 were identified by multiple linear regression analysis. All statistical analysis was performed using a commercially available statistical package, JMP 9.0.2 (SAS Institute, Cary, NC), with p-values less than 0.05 considered to indicate statistical significance.

3. Results

Clinical characteristics of the study subjects are summarized in Table 1. Current smokers were significantly younger and had a higher body stature and body weight, and were more commonly male.

AIx of current smokers was slightly lower than that of non-smokers and past-smokers, contrary to previous reports (Table 1). In contrast, clinical features in this group, i.e. younger age (r = 0.407, p < 0.001), higher stature (r = −0.416, p < 0.001) and male sex (male 76 ± 14, female 83 ± 12%, p < 0.001), are well-known to show a strong inverse association with AIx. We accordingly made a separate analysis by sex, and found that the AIx of current or past male smokers was somewhat higher than that of never smokers (never 73 ± 15, past 77 ± 13, current 77 ± 14%, p < 0.001), whereas no clear trend was observed in female subjects (never 84 ± 12, past 79 ± 13, current 82 ± 14, p < 0.001). Multiple linear regression analysis adjusted for these confounding factors identified smoking habit as an independent determinant for both AIx and PPa (Table 2), as well as AIx adjusted HR at 75 beats/min (AIx75) (β = 0.051, p < 0.001). Fig. 1A shows adjusted mean AIx and PPa by smoking habit: habitual smokers showed significantly higher AIx and smaller PPa, whereas bSBP of current smokers was somewhat lower than that of never smokers (Table 3).

Distribution of urinary cotinine levels is illustrated in Fig. 2. Urinary cotinine was chiefly detected in current smokers (1.08 ± 0.79 μg/ml), and levels were broadly distributed among current smokers. Cotinine levels were significantly associated with AIx and PPa (Table 2), as well as AIx75 (β = 0.064, p < 0.001), independently of smoking habit. Although a simple correlation analysis within current smokers showed no direct relationship between cotinine level and AIx (r = 0.033, p = 0.247) or PPa (r = 0.040, p = 0.169), presumably for the same reason as that for the lower AIx and larger PPa in current smokers,

Table 1
Clinical characteristics of study subjects.

	Total (8557)	Smoking habit			
		Never (5591)	Past (1764)	Current (1202)	p
Age (years)	54 ± 13	55 ± 13	55 ± 14	50 ± 13	<0.001
Sex (male/female)	2925/5632	743/4848	1301/463	881/321	<0.001
Height (cm)	160 ± 9	157.1 ± 7.4	165.1 ± 7.3	166.3 ± 8.1	<0.001
Weight (kg)	57 ± 11	54.6 ± 9.6	62.8 ± 10.6	62.9 ± 12.3	<0.001
BMI (kg/m²)	22.3 ± 3.3	22.1 ± 3.3	23.0 ± 3.1	22.6 ± 3.4	<0.001
bSBP (mm Hg)	124 ± 18	123 ± 18	127 ± 18	125 ± 18	<0.001
cSBP (mm Hg)	114 ± 18	114 ± 18	116 ± 18	114 ± 18	<0.001
PP amplification (mm Hg)	9.5 ± 5.8	8.9 ± 5.4	10.9 ± 6.1	10.7 ± 6.6	<0.001
MBP (mm Hg)	92 ± 13	91 ± 13	95 ± 13	93 ± 13	<0.001
DBP (mm Hg)	76 ± 11	75 ± 11	79 ± 11	77 ± 12	<0.001
Antihypertensive medication (%)	17.6	17.3	23.1	11.3	<0.001
Heart rate (beats/min)	69 ± 10	70 ± 10	68 ± 11	69 ± 10	<0.001
Augmentation index (%)	81 ± 13	82 ± 13	78 ± 13	78 ± 14	<0.001

BMI, body mass index calculated by weight (kg)/height² (m); bSBP, brachial systolic blood pressure; cSBP, central aortic systolic pressure; DBP, diastolic BP; MBP, mean BP calculated by DBP + (SBP − DBP) / 3. Pulse pressure (PP) amplification was calculated by bSBP − cSBP. Differences in numeric variables by smoking habit were assessed by analysis of variance. Frequency differences were assessed by the chi-squared test.

Table 2
Multiple linear regression analysis for AIx and PP amplification.

	AIx		PP amplification	
	β (s.e.)	p	β (s.e.)	p
Sex (male)	−0.236 (0.179)	<0.001	0.215 (0.088)	<0.001
Age (years)	0.291 (0.010)	<0.001	−0.167 (0.005)	<0.001
Height (cm)	−0.192 (0.021)	<0.001	0.178 (0.010)	<0.001
Weight (kg)	−0.098 (0.013)	<0.001	0.112 (0.007)	<0.001
Antihypertensive medication	−0.058 (0.149)	<0.001	0.074 (0.073)	<0.001
Heart rate (beats/min)	−0.462 (0.010)	<0.001	0.417 (0.005)	<0.001
MBP (mm Hg)	0.271 (0.010)	<0.001	−0.138 (0.005)	<0.001
Current smoking	0.049 (0.236)	<0.001	−0.041 (0.116)	0.003
Urinary creatinine (mg/dl)	0.013 (0.002)	0.116	−0.022 (0.001)	0.017
Urinary cotinine (μg/ml)	0.057 (0.334)	<0.001	−0.037 (0.164)	0.008

Smoking status is included as a dichotic value, namely 1 = current smoker, 0 = past and never smoker.

Table 3
Adjusted mean SBP and smoking property.

		Adjusted bSBP (mm Hg)	p
Smoking habit (total subjects)	Never smoker	125.0 ± 0.2	p = 0.003
	Past smoker	124.3 ± 0.2	
	Current smoker	124.4 ± 0.1	
Cotinine quartile (current smokers)	Q1	124.2 ± 0.4	p = 0.004
	Q2	125.6 ± 0.4	
	Q3	125.3 ± 0.4	
	Q4	125.7 ± 0.4	
Smoking quantity (current smokers)	≤10 cigarettes/day	124.8 ± 0.4	p = 0.322
	11–20 cigarettes/day	125.3 ± 0.4	
	>21 cigarettes/day	125.5 ± 0.4	
Cotinine levels (never smokers)	N.D.	123.1 ± 0.2	p = 0.474
	>0.05	122.7 ± 0.6	

Values are mean ± standard error. The following factors were adjusted by a linear regression model: sex, age, height, weight, antihypertensive medication, HR, MBP, and urinary creatinine. N.D. indicates urinary cotinine levels below the detection threshold (0.005 μg/ml).

adjusted mean AIx and PPa showed stepwise association with urinary cotinine quartile (Fig. 1B), with mean cotinine levels of each quartile as follows: Q1, 0.25 ± 0.14; Q2, 0.72 ± 0.13; Q3, 1.19 ± 0.16; and Q4, 2.18 ± 0.63 μg/ml; $p < 0.001$. The same relationship was seen in the analysis of smoking quantity obtained by a self-reported questionnaire (Fig. 1C). Adjusted mean AIx and PPa were associated in a dose-dependent manner. Self-reported smoking quantity was also significantly
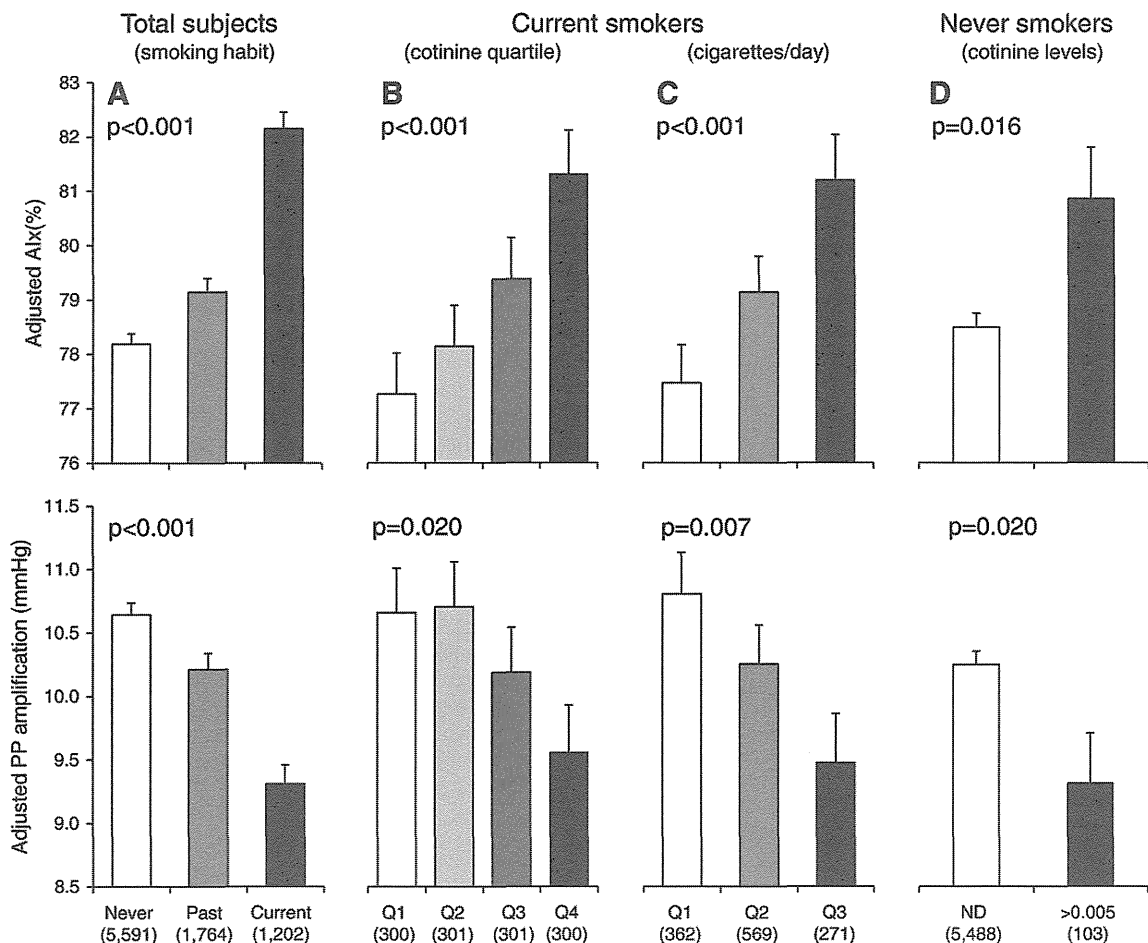

Fig. 1. Adjusted mean AIx and PP amplification by smoking status. Values are mean ± standard error. PP amplification was calculated by bSBP − cSBP. The following factors were adjusted by a linear regression model: sex, age, height, weight, antihypertensive medication, HR, MBP, and urinary creatinine. ND indicates urinary cotinine levels below the detection threshold (0.005 μg/ml).
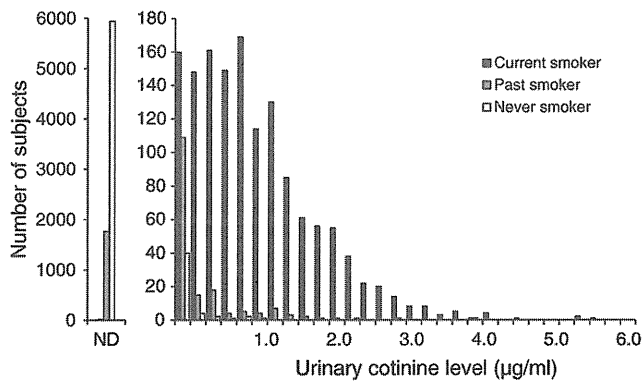
Fig. 2. Distribution of urinary cotinine levels. ND indicates urinary cotinine levels below the detection threshold (0.005 μg/ml).

associated with urinary cotinine levels (≤10 cigarettes/day, 0.66 ± 0.60; 11–20, 1.17 ± 0.75; ≥21, 1.46 ± 0.83 μg/ml, p < 0.001).

Urinary cotinine was also detected in a small fraction of never smokers (1.8%) (Fig. 2). This sub-population, regarded as passive smokers, showed higher adjusted AIx and smaller PPa (Fig. 1D) but not bSBP (Table 3). Mean AIx of passive smokers was almost equal to that of the highest quartile of current smokers.

4. Discussion

In this study, we found that habitual smoking was significantly associated with increased arterial tone as evaluated by AIx and smaller PPa in a general population sample. In smokers, AIx and PPa were linearly associated with smoking intensity. Further, passive smokers, as defined by urinary cotinine levels, also showed higher AIx and smaller PPa, with levels closely similar to those of current smokers. The major strength of this study is its large-scale sample with urinary cotinine levels, which has enabled us to detect the harmful impact of passive smoking on arterial pressure.

Our present and previous epidemiological studies have reported a weak or negative correlation between smoking habit and brachia BP. However, our results showed that the PPa of habitual smokers was substantially smaller than that of never or past smokers presumably due to increased arterial tone. Further, PPa was linearly decreased with smoking intensity as evaluated by urinary cotinine levels. Similar results were reported in an analysis of 443 normotensive Japanese men [7], namely that cSBP and AIx were significantly higher in current smokers than never smokers whereas no substantial differences were observed in brachial SBP. Given suggestions that cSBP is superior in predicting CV risks [1–3], relatively higher cSBP in smokers might partially explain the additive risk of habitual smoking on CV disease that cannot be attributed to BP.

Urinary cotinine was detected in a number of subjects without experience of smoking. These subjects, regarded as passive smokers, had increased levels of AIx and smaller PPa which were equivalent to those of current smokers. A recent prospective study of 13,443 participants living in England and Scotland [17] clearly showed a higher incidence of CV death and all-cause death during an 8-year follow-up period in passive smokers defined by salivary cotinine levels. Further, exposure to secondhand smoke was suggested to confer an increased risk of cognitive impairment [18]. Since increased cSBP may be associated with not only CV disease but also intracerebral small vessel disease [19], relatively higher cSBP may partially explain the excessive risk in passive smokers independently of the increased systemic inflammation [17].

Smoking cessation reduces AIx and pulse wave velocity in the aorta [20], which in turn decreases aortic BP. While the decrease in cardiovascular risk enjoyed by smokers who stop smoking is well known,

our results emphasize the importance of avoiding exposure to secondhand smoke by non-smokers to reduce a CV risk of which they are likely unaware.

We showed here clear relationships between smoking intensity as evaluated by urinary cotinine level and both AIx and cSBP. Smoking intensity has been usually evaluated by questionnaire, but the reliability of questionnaire-based measurements, such as the Brinkman index, has been questioned. Previous studies which investigated the relationship between self-reported smoking status and smoking intensity as confirmed by cotinine measurement in a large number of pregnant women [21], as well as a meta-analysis of 67 studies [22], have identified a trend toward underestimation in self-reported data. In the present study, analysis of self-reported smoking quantity showed that AIx and PPa tended to associate in a dose-dependent manner; however, distribution of urinary cotinine levels was largely overlapped among the subgroups. Although both urinary cotinine levels and self-reported smoking quantity were independently associated with higher AIx and smaller PPa in our cross-sectional investigation, these findings warrant further longitudinal study to clarify which of the two parameters provides a more sensitive indication risks for future CV events.

Chronic cigarette smoking has been shown to be associated with increased arterial stiffness [23,24]. Further, cigarette smoking acutely evokes sympathetic nerve activity [25] which in turn increases arterial tone. The harmful effects of smoking on AIx and PPa might result from both chronic and acute effects of smoking. Activation of the sympathetic nervous system concomitantly increases HR, and higher HR is associated with a lower AIx and, consequently, with a larger PPa. Higher AIx and smaller PPa in smokers are therefore independent of changes in HR. By contrast, smoking habit and smoking intensity had no substantial impact on brachial BP. Measurement of ambulatory BP revealed that daytime brachial BP was significantly higher in smokers [26,27]. The possible effects of smoking on bSBP levels might not be detectable by simple cross-sectional BP measurement.

5. Study limitations

Several study limitations also warrant mention. First, we investigated smoking habit by self-reported questionnaire. Thus, a degree of misclassification in smoking habit might have occurred; if so, however, any such misclassification might have been independent of smoking status and be non-differential. Second, our observational study design does not allow us to discriminate whether the higher AIx and smaller PPa in passive smokers were a long-lasting phenomenon due to continuous exposure to secondhand smoke, or a transient reaction to passive smoking occurring in the few days before measurements.

6. Conclusions

In summary, we have shown that not only habitual smoking but also passive smoking had harmful effects on AIx and central BP in a large-scale general population sample. Our results suggest that the changes in aortic pressure in smokers represent one reason for the elevated CV risks of smoking, and strongly emphasize the importance of avoiding passive smoking in preventing CV disease.