

## ORIGINAL ARTICLE

# Identification of independent risk loci for Graves' disease within the MHC in the Japanese population

Kazuhiko Nakabayashi<sup>1,2,16</sup>, Atsushi Tajima<sup>3,16</sup>, Ken Yamamoto<sup>4</sup>, Atsushi Takahashi<sup>5</sup>, Kenichiro Hata<sup>2</sup>, Yasuo Takashima<sup>1,6</sup>, Midori Koyanagi<sup>1,6</sup>, Hirofumi Nakaoka<sup>7</sup>, Takashi Akamizu<sup>8,17</sup>, Naofumi Ishikawa<sup>9</sup>, Sumihisa Kubota<sup>10</sup>, Shiro Maeda<sup>11</sup>, Tatsuhiko Tsunoda<sup>12</sup>, Michiaki Kubo<sup>13</sup>, Naoyuki Kamatani<sup>5</sup>, Yusuke Nakamura<sup>14</sup>, Takehiko Sasazuki<sup>15</sup> and Senji Shirasawa<sup>1,6</sup>

To identify genetic variants that confer the risk of Graves' disease (GD) in the Japanese population, we conducted a two-stage genome-wide association study (GWAS) using 1119 Japanese individuals with GD and 2718 unrelated controls, and a subsequent replication study using independent 432 GD cases and 1157 controls. We identified 34 single nucleotide polymorphisms (SNPs) to be significantly associated with GD in the GWAS phase. Twenty-two out of 34 SNPs remained positive in the replication study. All 22 SNPs were located within the major histocompatibility complex (MHC) locus on chromosome 6p21. No strong long-range linkage disequilibrium (LD) was observed among the 22 SNPs, indicating independent involvement of multiple loci within the MHC with the risk of GD. Multivariate stepwise logistic regression analysis selected rs3893464, rs4313034, rs3132613, rs4248154, rs2273017, rs9394159 and rs4713693, as markers for independent risk loci for GD. The analysis of LD between these seven SNPs and tagging SNPs for GD-associated human leukocyte antigen (HLA) alleles in the Japanese population (HLA-DPB1\*0501 and HLA-A\*0206) demonstrated that all of and five of seven SNPs were not in strong LD with HLA-DPB1\*0501 and HLA-A\*0206, respectively. Although causal variants remain to be identified, our results demonstrate the existence of multiple GD susceptibility loci within the MHC region.

*Journal of Human Genetics* (2011) 56, 772–778; doi:10.1038/jhg.2011.99; published online 8 September 2011

**Keywords:** genome-wide association study; Graves' disease; major histocompatibility complex; single nucleotide polymorphism

## INTRODUCTION

Autoimmune thyroid disease (AITD), including Graves' disease (GD) and Hashimoto's thyroiditis (HT), is caused by immune response to self-thyroid antigens, and affects up to 2–5% of the general population.<sup>1</sup> GD is characterized by the production of thyroid-stimulating hormone receptor-stimulating antibodies, leading to hyperthyroidism, whereas HT is characterized by the apoptosis of the thyrocytes, resulting in hypothyroidism.<sup>2,3</sup> Although the molecular mechanisms underlying the initiation of AITD remain largely elusive, twin studies and familial aggregation have clearly demonstrated the involvement of genetic factors with AITD in addition to environmental factors.<sup>4,5</sup>

Human leukocyte antigen (HLA)/major histocompatibility complex (MHC) locus on the chromosome 6p21 was the first region identified to be associated with AITD 25 years ago, and has been consistently replicated. However, the primary causal variant in this region remains still unknown due to the long extended linkage disequilibrium (LD) feature of the HLA region.<sup>5,6</sup> Several AITD susceptibility genes from non-HLA loci have also been identified through candidate gene analyses and whole genome linkage studies, followed by locus-specific association analysis. Such non-HLA genes can be classified into at least two subgroups, thyroid-specific genes and immune-related genes.<sup>6</sup> The former group is comprised of the thyroglobulin gene (*TG*)<sup>6</sup> and

<sup>1</sup>Department of Cell Biology, Faculty of Medicine, Fukuoka University, Fukuoka, Japan; <sup>2</sup>Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development, Tokyo, Japan; <sup>3</sup>Department of Human Genetics and Public Health, Institute of Health Biosciences, The University of Tokushima Graduate School, Tokushima, Japan; <sup>4</sup>Department of Molecular Genetics, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan; <sup>5</sup>Laboratory for Statistical Analysis, Center for Genomic Medicine, RIKEN Yokohama Institute, Yokohama, Japan; <sup>6</sup>Center for Advanced Molecular Medicine, Fukuoka University, Fukuoka, Japan; <sup>7</sup>Division of Human Genetics, National Institute of Genetics, Mishima, Japan; <sup>8</sup>Translational Research Center, Kyoto University School of Medicine, Kyoto, Japan; <sup>9</sup>Ito Hospital, Tokyo, Japan; <sup>10</sup>Kuma Hospital, Kobe, Japan; <sup>11</sup>Laboratory for Endocrinology and Metabolism, Center for Genomic Medicine, RIKEN Yokohama Institute, Yokohama, Japan; <sup>12</sup>Laboratory for Medical Informatics, Center for Genomic Medicine, RIKEN Yokohama Institute, Yokohama, Japan; <sup>13</sup>Laboratory for Genotyping and Development, Center for Genomic Medicine, RIKEN Yokohama Institute, Yokohama, Japan; <sup>14</sup>Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan and <sup>15</sup>Institute for Advanced Studies, Kyushu University, Fukuoka, Japan

<sup>16</sup>These authors contributed equally to this paper.

<sup>17</sup>Current address: The First Department of Internal Medicine, Wakayama Medical University, Wakayama, Japan.

Correspondence: Dr K Nakabayashi, Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya, Tokyo 157-8535, Japan.

E-mail: knakabayashi@nch.go.jp

or Professor S Shirasawa, Department of Cell Biology, Faculty of Medicine, Fukuoka University, 7-45-1 Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan.

E-mail: sshirasa@fukuoka-u.ac.jp

Received 24 May 2011; revised 22 July 2011; accepted 28 July 2011; published online 8 September 2011

*TSHR*.<sup>7,8</sup> The latter group includes *CTLA4*,<sup>9</sup> *FCRL3*,<sup>10</sup> and *PTPN22*,<sup>6</sup> which are shown to be commonly involved in multiple autoimmune diseases. In addition to these genes, we have previously identified *ZFAT* (zinc-finger gene in AITD susceptibility region) on 8q24 as a susceptibility gene for both GD and HT,<sup>11</sup> and have revealed that *ZFAT* functions as a transcriptional regulator of immune-related genes.<sup>12</sup>

Recent genome-wide association studies (GWAS) have dramatically increased the numbers of susceptibility loci of autoimmune diseases such as Crohn's disease, systemic lupus erythematosus and type-1 diabetes.<sup>13</sup> In contrast, there has been only one GWAS for GD in the Western population.<sup>14</sup> We conducted a two-stage GWAS to elucidate genetic variations involved in GD susceptibility in the Japanese population.

## MATERIALS AND METHODS

### Study participants

For the GWAS, we selected case-control samples (case 1, control 1, case 2 and control 2) from the subjects enrolled in the Biobank Japan. We selected cases from individuals registered as having GD, and controls from individuals with diseases other than autoimmune diseases including GD. For the first stage of GWAS, 195 individuals with GD (case 1) and 1362 control individuals (control 1) were enrolled. The female ratio and the mean ( $\pm$  s.d.) age of the case and control groups are: 67.2% and 47.7 ( $\pm$  15.7) for case 1, and 49.8% and 59.8 ( $\pm$  12.2) for control 1. For the second stage of GWAS, 924 individuals with GD (case 2) and 1356 control individuals (control 2) were enrolled. The female ratio and the mean age of the groups are: 72.6% and 49.5 ( $\pm$  15.1) for case 2, and 39.0% and 61.2 ( $\pm$  18.6) for control 2. For the replication study, 432 individuals with GD (case 3) and 1157 control individuals (control 3) were enrolled. Diagnosis of GD was established as described previously.<sup>15</sup> Genomic DNA was isolated from peripheral blood cells using the QIAamp DNA blood Midi Kit (Qiagen, Tokyo, Japan). The ethical committees of all related institutes approved this study, and written informed consent was obtained from all subjects.

### Genotyping and selection of SNPs

In the first stage of GWAS, 268 068 autosomal single nucleotide polymorphisms (SNPs), selected as described previously,<sup>16</sup> were genotyped using high-density oligonucleotide arrays (Perlegen Sciences, Mount View, CA, USA). The differences between the case and control groups (case 1 and control 1) with regard to the genotype distribution or the allele frequency were assessed by Fisher's exact test using dominant, recessive and allelic models. SNPs were ranked by the minimum *P*-value among three models. Top 4000 SNPs were considered to be subjected to the second stage of GWAS. In the second stage, genotyping was conducted using the multiplex-PCR invader assay as described previously.<sup>17</sup> Invader assays were successfully designed for approximately 3900 out of the top 4000 SNPs. A cohort of GD cases (case 2, *n*=924) was genotyped by the invader assays for these 3900 SNPs. The obtained genotype data and those for a cohort of controls (control 2, *n*=1356) by Perlegen arrays were subjected to association tests by Fisher's exact tests. The control 2 cohort was genotyped by invader assays for 2720 SNPs whose minimum *P*-value among three models (dominant, recessive and allelic) surpassed the significance threshold of  $P < 10^{-4}$ . After excluding 12 SNPs whose genotype counts of minor allele homozygotes include 0 in either of case 2 or control 2 cohort data, the remaining 2708 SNPs were subjected to statistical analyses to select GD-associated SNPs in the GWAS stage. In the replication study, genotyping was conducted using the TaqMan assay as described previously.<sup>9</sup>

### Statistical methods

To select the GD-associated SNPs using the genome-wide screening data, a meta-analysis was performed for the integrated data set of the first and second GWAS stages, using the metafor package for R (<http://www.wvbauer.com/index.htm>). In the meta-analysis, the  $I^2$  metric<sup>18,19</sup> was used to consider the extent of heterogeneity between the first and the second stages of GWAS: *P*-value of fixed-effects (Mantel-Haenszel) meta-analysis was adopted when  $I^2$  metric is not larger than 50%, *P*-value of random-effects (DerSimonian-Laird)

meta-analysis was adopted when  $I^2$  metric is equal to or larger than 50%. In the replication study, disease associations were assessed by Cochran-Armitage trend test. Per allele odds ratios and their 95% confidence intervals were calculated using unconditional logistic regression. The thresholds for the Hardy-Weinberg equilibrium and minor allele frequency were set as  $P > 0.01$  and  $> 5\%$ , respectively. SNPAnalyze v7.0 software (DYNACOM, MObama, Japan) and Haploview version 4.0 software<sup>20</sup> were used for LD calculation. Average power to detect common SNPs (risk allele frequency range of 0.1–0.9) that confer a genotype relative risk of 1.5 at GWAS stage 1 (195 cases versus 1362 controls) and replication stage was calculated, assuming an additive model in log-odds scale and GD population prevalence of 0.5%. The average power in the GWAS stage 1 was 68%. The average power in the replication study exceeded 80%.

GD cases and controls with missing genotype data regarding the SNPs to be analyzed are excluded in the following analyses. The multivariate stepwise logistic regression analysis was performed in a stepwise (forward-backward) manner to select SNPs that represented the GD risk within the MHC region. The entry and removal criteria with *P*-values of 0.01 based on the likelihood ratio test were used. To evaluate the ability of the seven-SNP-logistic regression model for phenotype prediction, a receiver-operating characteristic (ROC) analysis was conducted. A ROC curve to discriminate GD cases from healthy controls was generated using the combined data set, and the area under the curve was calculated.

To assess the cumulative effects of risk alleles at seven SNPs that were included in the logistic regression model, the risk score for each of the all possible seven-locus genotypes was calculated as described previously.<sup>21</sup> Subsequently, the frequencies of each of the combinations of the genotypes of the seven SNPs (790 combinations observed) in the GD and control populations in the combined data set of GWA and replication studies were determined.

## RESULTS

To define genetic factors influencing the risk of GD in the Japanese population, we conducted a two-stage genome-wide association analysis in Japanese individuals with GD and unrelated controls. Characteristics of the cohort groups enrolled in each of the stages are described in the Materials and Methods section. In the first stage, we genotyped 268 042 autosomal SNPs, which covered approximately 56% of common SNPs in the Japanese,<sup>16</sup> in 195 individuals with GD (case 1) and 1362 unrelated controls (control 1) collected in the Biobank Japan. We subsequently assessed differences between the case and control groups with regard to genotype distribution or allele frequency by Fisher's exact test using dominant, recessive and allelic models. We ranked the genotyped SNPs by the minimum *P*-value among three models for each SNP and selected top 4000 SNPs showing the lowest *P*-values. As the second stage, we attempted to genotype these 4000 SNPs in 924 individuals with GD (case 2) and 1356 controls (control 2) by invader assays. After quality control processes as described in the Materials and Methods section, we successfully obtained genotyping data of 2708 SNPs to be subjected to association tests. Associations of these SNPs with GD were initially assessed by Cochran-Armitage trend test. The distribution of the test statistics for the SNPs in the second stage of GWAS was shown in Supplementary Figure S1. Considering the well-established association of HLA genes with GD<sup>5,6</sup> and the existence of multiple GD-associated SNPs in the MHC region revealed by the genome-wide analysis for the UK population,<sup>14</sup> we calculated the genomic inflation factor ( $\lambda_{GC}$ ) for the 2654 SNPs after excluding 54 SNPs located in the MHC region among the 2708 SNPs. The  $\lambda_{GC}$  for the 2654 SNPs was 1.088, suggesting that the data set was not heavily confounded by potential problems such as genotype misclassification or a population substructure.

Therefore, we tested for association between each SNP and GD by performing a meta-analysis of the results from the first and the second

stages, and identified 34 SNPs (Supplementary Table S1) that surpassed the multiple-testing corrected significance threshold  $P_c < 1.86 \times 10^{-7}$  (0.05/268 042). Aside from two SNPs, rs2240306 located at the 5'-upstream region of the *SENPI* (sentrin/SUMO-specific protease 1) gene on chromosome 12, and rs1149386 located at the 3' downstream region of the *SLAMF6* gene encoding an activating natural killer receptor precursor on chromosome 1, 32 out of 34 SNPs were located in an interval of 29.60–33.93 Mb within the MHC region on chromosome 6p21 (Figure 1). Replication of association of these 34 SNPs was tested in an independent set of cases and controls, comprising 432 Japanese individuals with GD (case 3) and 1157 controls (control 3), setting  $P_c < 0.00147$  (0.05/34) for significant replication. This replication analysis identified significant association of 22 SNPs, all of which are located in the MHC region, with GD (Table 1). All of these SNPs were in Hardy–Weinberg equilibrium in the controls ( $P > 0.019$ ). The genotype counts and the statistics of the 12 SNPs that did not surpass the significance threshold in the replication study are listed in Supplementary Table S2.

The MHC region is known to exhibit LD that extends across multiple HLA and non-HLA genes.<sup>22</sup> High-resolution LD mapping studies of the extended MHC region spanning about 7.6 Mb have revealed that LD exists in segments of over 2 Mb in some cases, but not all. Such studies have also demonstrated the existence of tight segment-to-segment LD observed between regions of strong LD that are located distantly from each other.<sup>22,23</sup> We calculated the pairwise LD of the 22 SNPs that were found to be associated with GD, using the genotyping data of enrolled control sets (control 1+2+3,  $n=3875$ ). Among the 22 GD-associated SNPs, strong pairwise LD ( $r^2 > 0.8$  and  $|D'| > 0.9$ ) was observed only within two pairs of closely located SNPs (SNPs #1/#2 and #6/#7, whose  $r^2$  values are highlighted in red in the schematic LD patterns in Figure 1). Among the pairs of SNPs that are located distantly ( $> 300$  kb) from each other, only limited pairs exhibited modest or high levels ( $r^2 = 0.21 \sim 0.51$  and  $|D'| = 0.50 \sim 0.90$ ) of LD. These LD patterns observed among the 22 SNPs in the MHC region suggest that the detected association of these SNPs with GD cannot be explained by a segment of extensive long-range LD, but that, at least some SNPs are likely to represent genomic segments that contain independent risk factors of GD. As a statistical means to assess the independence of these 22 SNPs in terms of association with GD, we conducted association tests for these SNPs conditioned on rs2273017, which showed the lowest  $P$ -value in the unconditional association analysis. Genotype data of all GD cases ( $n=1551$ ) and controls ( $n=3875$ ) were subjected to the analysis. Significant association with GD ( $P_c < 0.00238$  (0.05/21)) was detected for all SNPs (Supplementary Table S3), further indicating the independent involvement of multiple SNPs within the MHC region with GD. We subsequently conducted a multivariate stepwise logistic regression analysis to investigate the combinatorial effects of multiple SNPs with the risk of GD. Twenty out of 22 SNPs were subjected to the analysis as tagging SNPs (two SNPs, rs362540 and rs2254071 were excluded). GD cases and controls from GWAS and replication cohorts with genotypes for all 20 SNPs (GD,  $n=1152$ ; controls,  $n=2570$ ) were included in this analysis (as well as in the subsequent statistical analyses in this study). Consequently, we established a statistical model in which seven SNPs, rs3893464, rs4313034, rs3132613, rs4248154, rs2273017, rs9394159 and rs4713693, affected the risk of GD significantly and independently (Table 2). These seven SNPs are located discretely within a 4-Mb interval (Figure 1).

We further evaluated the seven-SNP-logistic regression model for its ability to predict the disease status of GD in a given individual, using the combined data of the GWA and replication studies. When we

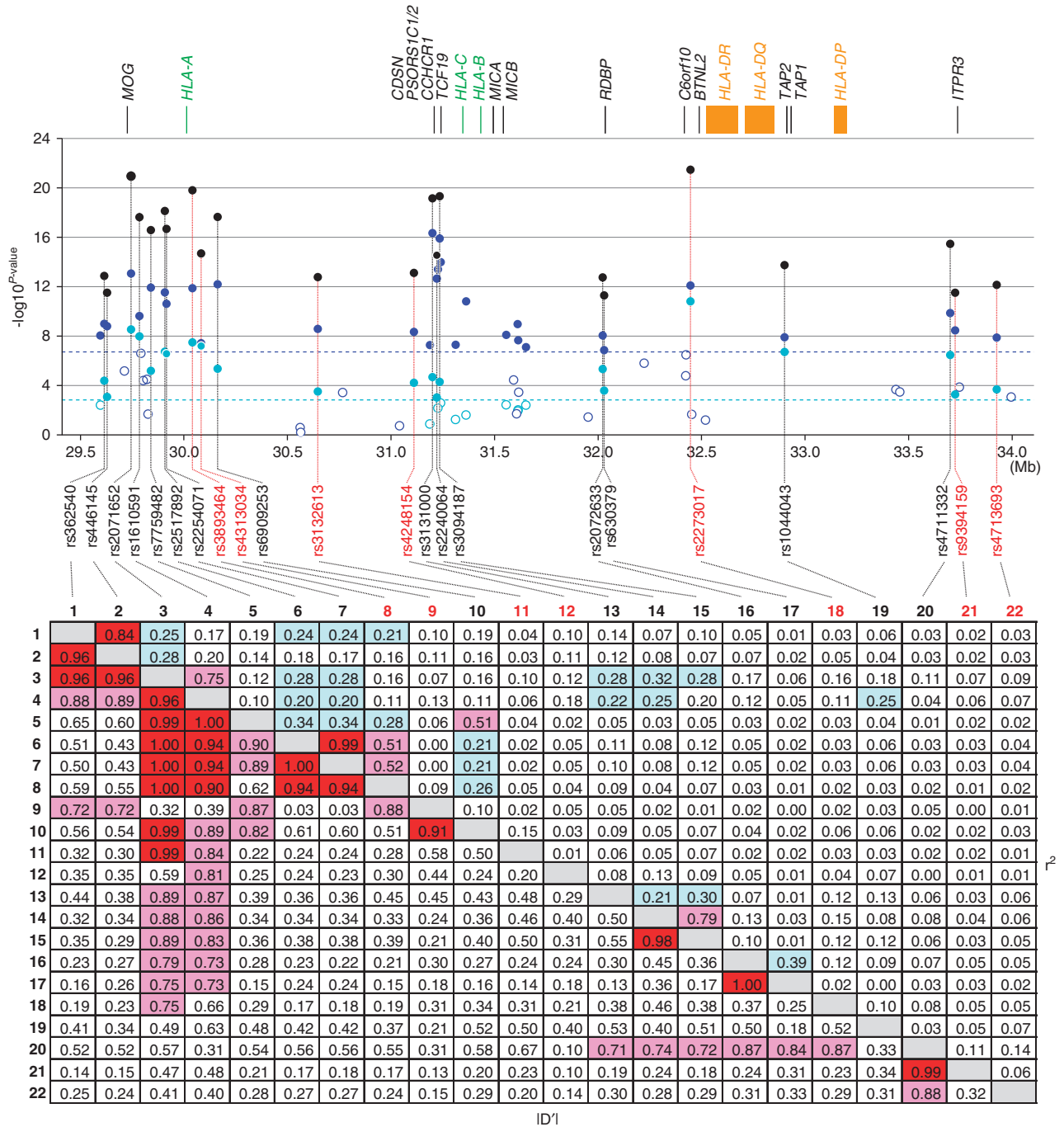
calculated the risk score for each subject by summing the logarithm of the logistic regression-based odds ratio (Table 2) for each risk allele that he/she harbored, the observed GD risk showed a significant linear relationship with risk score ( $P=2.1 \times 10^{-44}$ , Table 3), indicating cumulative effects of the risk alleles for the seven SNPs on the development of GD. Compared with the subjects having zero to three risk alleles, the odds ratios of GD exceeded over fivefold in those having more than seven risk alleles, and over tenfold in those having more than 10 risk alleles (Table 3 and Supplementary Figure S2). The ROC curve of the seven-SNP-logistic regression model was generated to further assess its sensitivity and specificity of predicting the development of GD (Figure 2). The area under the ROC curve was 0.645, indicating a modest diagnostic ability of the model.

## DISCUSSION

The MHC region represents one of the most extensively studied regions in the human genome because of the contribution of genetic variants in this region in autoimmune, infectious and inflammatory diseases. However, highly polymorphic nature and extensive LD of the MHC region has made it difficult to identify causal variants within this locus for the majority of diseases.<sup>23</sup> The Wellcome Trust Case Control Consortium (WTCCC) performed association scan of 14 436 non-synonymous SNPs and 897 MHC tag SNPs in four autoimmune diseases including GD.<sup>14</sup> In this WTCCC study, the strongest associations with GD were observed for SNPs in the MHC region. Although the maximum signal was observed around the MHC class I and class II genes, the strong association extended far beyond specific haplotypes. The International MHC Autoimmunity Genetics Network recently conducted an association analysis, using a panel of 1472 SNPs within the 3.44-Mb classic MHC region for a set of 10 576 DNA samples derived from patients with either of the seven immune-mediated diseases and appropriate control samples.<sup>24</sup> It should be noted that GD was not included in the seven diseases analyzed. This large-scale study has demonstrated that MHC associations with autoimmune diseases result from complex and multilocus effects that span the entire region.

We attempted to identify genetic variants involved in the risk of GD in the Japanese population by adopting a two-stage GWAS strategy. Whereas we detected strong association of multiple SNPs spanning several megabases within the MHC regions with GD, we failed to detect association of known and novel non-MHC loci with GD. We have previously reported that SNPs at *ZFAT*, *CTLA4*, *FCRL3* and *TSHR* loci are associated with GD in the Japanese population.<sup>7,9–11</sup> However, in this study, no SNP from these four loci, except for rs2210913 within the *FCRL3* gene, was included in the top 4000 SNPs that were selected in the first stage of GWAS. SNP rs2210913 did not surpass the significance threshold in the meta-analysis of the results from the first and the second GWAS stages. It should be mentioned that the average power to detect common SNPs that confer a genotype relative risk  $< 1.5$  did not exceed 80% in both of the current GWA and replication studies. Therefore, a lack of power to detect true association and an insufficient cover rate of the human genome in the first stage of GWAS in our study is likely to be mostly responsible for the discrepancy between the present results and past studies regarding the association of the genes located outside the MHC region with GD in the Japanese population. In the WTCCC study, although *TSHR* and *FCRL3* loci were confirmed to be associated with GD, no additional locus was identified as a novel GD susceptibility locus.<sup>14,25</sup>

Through the GWA and the subsequent replication studies, we identified 22 GD-associated SNPs within the MHC region. Among them, seven SNPs that were integrated into the multivariate logistic



**Figure 1** Association of SNPs in the MHC region (29.5–34.0 Mb of chromosome 6) with GD identified by GWA and replication studies. Representative genes in the MHC region are shown on the top. Classical MHC class I genes (HLA-A, -B and -C) and class II genes (HLA-DR, -DQ and -DP gene clusters) are shown in green and orange, respectively. Others are shown in black. Values of  $-\log_{10}(P\text{-value})$  are plotted against chromosomal positions based on NCBI Build 35 coordinates. Blue (open/closed), right blue (open/closed) and black (closed) circles represent the results of the GWAS (54 SNPs among the 2708 SNPs that were successfully genotyped, 1119 GD cases versus 2718 controls, meta-analysis  $P$ -values), the replication study (34 SNPs, 432 GD cases versus 1157 controls, Cochran–Armitage trend  $P$ -values) and the combined analysis of both studies (22 SNPs, 1551 GD cases versus 3875 controls, Cochran–Armitage trend  $P$ -values), respectively. The blue and light blue dashed-horizontal lines indicate the significance levels applied to the meta-analysis of the GWAS stages ( $\alpha=1.86 \times 10^{-7}$  (0.05/268042)) and applied to the association analysis in the replication study ( $\alpha=0.00147$  (0.05/34)), respectively. Closed and open circles (blue and light blue) indicate  $P$ -values smaller than and the ones not smaller than the significance level applied, respectively. The rs IDs of the 22 SNPs, whose associations with GD were replicated, are listed below the  $P$ -value plot. Each SNP ID and the corresponding circles representing  $-\log_{10}(P)$  values are connected by a vertical dashed line. Seven SNPs shown in red correspond to the ones suggested to represent independent genetic risk factors of GD in the multivariate step-wise logistic regression analysis. Below the SNP IDs, pairwise linkage disequilibrium (LD) patterns of the 22 SNPs are schematically represented as a block structure. Genotype data of the controls (control 1, 2 and 3,  $n=3875$ ) were used to determine the pairwise LD coefficients  $D'$  and  $r^2$ . In the schematic block for  $r^2$ , red highlight denotes pairwise LD of  $r^2 \geq 0.8$ , pink  $0.5 > r^2 \geq 0.5$ , right blue  $0.4 > r^2 \geq 0.2$  and no highlight  $0.2 > r^2$ . In the schematic block for  $D'$ , red highlight denotes pairwise LD of  $D' \geq 0.9$ , pink  $0.9 > D' \geq 0.7$  and no highlight  $0.7 > D'$ .

**Table 1** Summary of the results for the 22 GD-associated SNPs in the MHC region

SNP (risk allele)	Stage	GD			Control			Risk allele frequency		CA-trend	Per-allele	
		1/1	1/2	2/2	1/1	1/2	2/2	GD (%)	Control (%)	P-value	OR	95% CI
rs362540 (G)	GWAS	341	556	211	634	1344	737	55.9	48.1	8.18E-10	1.37	(1.24~1.51)
29618304	Replication	134	210	83	257	590	309	56.0	47.8	3.65E-05	1.40	(1.19~1.64)
	Combined	475	766	294	891	1934	1046	55.9	48.0	1.34E-13	1.37	(1.26~1.50)
rs446145 (C)	GWAS	377	545	192	711	1331	675	58.3	50.7	1.61E-09	1.36	(1.23~1.50)
29628766	Replication	142	208	79	286	597	271	57.3	50.6	7.12E-04	1.32	(1.12~1.55)
	Combined	519	753	271	997	1928	946	58.0	50.7	4.55E-12	1.35	(1.24~1.46)
rs2071652 (C)	GWAS	681	231	13	1653	915	140	86.1	77.9	4.77E-14	1.75	(1.51~2.03)
29743296	Replication	319	98	11	670	413	69	86.0	76.1	2.72E-09	1.90	(1.53~2.36)
	Combined	1000	329	24	2323	1328	209	86.1	77.4	1.19E-21	1.80	(1.59~2.03)
rs1610591 (G)	GWAS	848	250	14	1729	786	100	87.5	81.1	3.05E-11	1.62	(1.40~1.87)
29785814	Replication	336	87	8	726	373	56	88.1	79.0	1.06E-08	1.93	(1.53~2.42)
	Combined	1184	337	22	2455	1159	156	87.7	80.5	2.30E-18	1.71	(1.51~1.93)
rs7759482 (G)	GWAS	131	492	403	187	992	1261	36.7	28.0	4.15E-13	1.50	(1.35~1.68)
29839666	Replication	60	216	151	102	502	541	39.3	30.8	4.60E-06	1.48	(1.25~1.75)
	Combined	191	708	554	289	1494	1802	37.5	28.9	1.87E-17	1.49	(1.36~1.63)
rs2517892 (T)	GWAS	359	582	174	653	1375	690	58.3	49.3	4.94E-13	1.45	(1.31~1.61)
29909453	Replication	145	208	64	271	588	288	59.7	49.3	1.73E-07	1.55	(1.31~1.82)
	Combined	504	790	238	924	1963	978	58.7	49.3	6.01E-19	1.48	(1.35~1.61)
rs2254071 (T)	GWAS	353	582	179	654	1370	694	57.8	49.3	6.53E-12	1.42	(1.29~1.58)
29914041	Replication	151	200	72	269	578	292	59.3	49.0	2.79E-07	1.52	(1.30~1.79)
	Combined	504	782	251	923	1948	986	58.2	49.2	1.27E-17	1.45	(1.33~1.58)
rs3893464 (G)	GWAS	184	480	274	338	1221	1115	45.2	35.5	8.05E-14	1.51	(1.35~1.68)
30043229	Replication	103	183	122	143	548	447	47.7	36.6	3.86E-08	1.57	(1.34~1.85)
	Combined	287	663	396	481	1769	1562	46.0	35.8	1.88E-20	1.53	(1.39~1.67)
rs4313034 (T)	GWAS	871	213	24	1843	812	63	88.2	82.7	1.74E-09	1.57	(1.36~1.83)
30081904	Replication	351	76	4	780	336	34	90.3	82.4	6.09E-08	1.98	(1.54~2.55)
	Combined	1222	289	28	2623	1148	97	88.8	82.7	1.51E-15	1.67	(1.47~1.90)
rs6909253 (G)	GWAS	212	547	346	313	1258	1132	43.9	34.9	7.61E-14	1.48	(1.33~1.64)
30163622	Replication	93	205	126	140	575	438	46.1	37.1	2.84E-06	1.48	(1.25~1.74)
	Combined	305	752	472	453	1833	1570	44.5	35.5	1.72E-18	1.47	(1.35~1.61)
rs3132613 (C)	GWAS	111	491	515	121	764	1190	31.9	24.2	4.38E-11	1.47	(1.31~1.65)
30645585	Replication	43	193	195	73	453	626	32.4	26.0	3.24E-04	1.37	(1.15~1.63)
	Combined	154	684	710	194	1217	1816	32.0	24.9	1.43E-13	1.43	(1.30~1.57)
rs4248154 (C)	GWAS	437	515	167	808	1335	575	62.1	54.3	6.03E-10	1.37	(1.24~1.52)
31110595	Replication	158	197	75	299	595	261	59.7	51.6	5.56E-05	1.39	(1.18~1.63)
	Combined	595	712	242	1107	1930	836	61.4	53.5	1.06E-13	1.38	(1.27~1.50)
rs3131000 (C)	GWAS	460	515	133	820	1317	567	64.8	54.7	1.05E-15	1.52	(1.37~1.68)
31199252	Replication	175	185	62	360	546	249	63.4	54.8	2.67E-05	1.40	(1.20~1.65)
	Combined	635	700	195	1180	1863	816	64.4	54.7	1.76E-19	1.48	(1.36~1.62)
rs2240064 (A)	GWAS	476	446	103	963	1284	471	68.2	59.1	7.55E-13	1.48	(1.33~1.64)
31222552	Replication	182	181	59	380	572	196	64.6	58.0	9.04E-04	1.32	(1.12~1.55)
	Combined	658	627	162	1343	1856	667	67.1	58.7	4.80E-15	1.43	(1.31~1.56)
rs3094187 (T)	GWAS	473	507	138	822	1318	567	65.0	54.7	2.84E-16	1.53	(1.38~1.69)
31234923	Replication	169	188	68	330	581	242	61.9	53.8	5.65E-05	1.39	(1.18~1.63)
	Combined	642	695	206	1152	1899	809	64.1	54.4	8.75E-20	1.49	(1.37~1.62)
rs2072633 (A)	GWAS	287	468	172	619	1384	707	56.2	48.4	4.29E-09	1.38	(1.24~1.54)
32027557	Replication	134	210	81	252	580	320	56.2	47.0	4.58E-06	1.45	(1.24~1.70)
	Combined	421	678	253	871	1964	1027	56.2	48.0	1.18E-13	1.40	(1.28~1.53)
rs630379 (A)	GWAS	118	501	489	182	1082	1451	33.3	26.6	4.47E-09	1.38	(1.24~1.54)
32030233	Replication	45	201	186	83	458	608	33.7	27.2	2.82E-04	1.37	(1.16~1.63)
	Combined	163	702	675	265	1540	2059	33.4	26.8	5.57E-12	1.38	(1.26~1.51)
rs2273017 (A)	GWAS	419	493	182	679	1211	612	60.8	51.3	3.57E-13	1.45	(1.31~1.61)
32445608	Replication	162	195	60	282	551	314	62.2	48.6	4.06E-11	1.71	(1.46~2.02)
	Combined	581	688	242	961	1762	926	61.2	50.5	2.12E-22	1.53	(1.40~1.66)
rs1044043 (C)	GWAS	704	353	54	1451	1058	207	79.3	72.9	9.55E-09	1.41	(1.25~1.59)
32901959	Replication	278	142	11	602	457	96	81.0	71.9	2.21E-07	1.66	(1.37~2.02)
	Combined	982	495	65	2053	1515	303	79.7	72.6	2.48E-14	1.48	(1.34~1.63)

**Table 1 Continued**

SNP (risk allele)	Chr.6 position	Stage	GD			Control			Risk allele frequency		CA-trend	Per-allele	
			1/1	1/2	2/2	1/1	1/2	2/2	GD (%)	Control (%)	P-value	OR	95% CI
rs4711332 (G)		GWAS	1009	93	3	2260	431	21	95.5	91.3	2.12E-10	2.03	(1.63~2.54)
33703377		Replication	391	40	0	923	219	10	95.4	89.6	3.20E-07	2.42	(1.71~3.43)
		Combined	1400	133	3	3183	650	31	95.5	90.8	3.52E-16	2.15	(1.78~2.59)
rs9394159 (T)		GWAS	388	570	157	749	1381	587	60.4	53.0	2.15E-09	1.37	(1.23~1.51)
33726140		Replication	158	193	78	302	607	246	59.3	52.4	4.74E-04	1.33	(1.13~1.56)
		Combined	546	763	235	1051	1988	833	60.1	52.8	3.89E-12	1.36	(1.24~1.48)
rs4713693 (T)		GWAS	602	422	89	1178	1219	317	73.0	65.9	1.28E-09	1.40	(1.26~1.56)
33925907		Replication	205	188	37	443	550	157	69.5	62.4	1.81E-04	1.38	(1.17~1.64)
		Combined	807	610	126	1621	1769	474	72.1	64.8	6.60E-13	1.40	(1.28~1.53)

Abbreviations: CA-trend, Cochran-Armitage trend; Chr.6 position, chromosome 6 location; CI, confidence interval; GWAS, genome-wide association study; OR, odds ratio; SNPs, single nucleotide polymorphisms. Genotype counts, risk allele frequencies, CA-trend P-values and per-allele OR with CI are shown for GWAS, replication and combined data sets. The seven SNPs selected for the logistic regression model are underlined. NCBI Build 35 coordinates were used to indicate the chromosome 6 location of the SNPs. Risk allele is indexed to the forward strand of NCBI Build 35. Allele 1 in the genotype counts corresponds to the risk allele.

**Table 2 GD-associated SNPs selected by the step-wise logistic regression analysis**

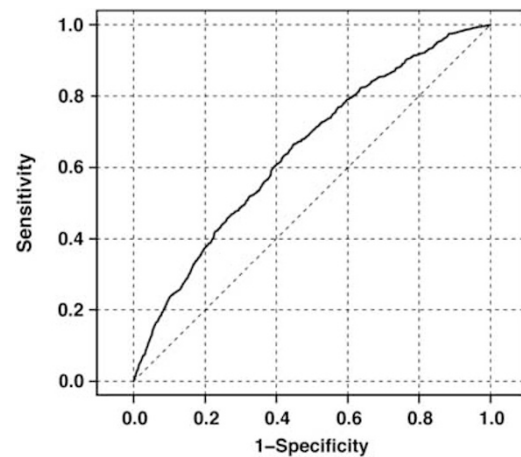
SNP	OR (95% CI)	Z-value	P-value	SNP location relative to the nearest gene
rs3893464	1.29 (1.15-1.44)	4.4	8.6E-06	21.3 kb downstream of the <i>HLA-A</i> gene
rs4313034	1.30 (1.01-1.52)	3.2	1.6E-03	In the first intron of the <i>HLA-J</i> gene
rs3132613	1.18 (1.05-1.33)	2.8	4.9E-03	1.8 kb upstream of the <i>ABCF1</i> gene
rs4248154	1.20 (1.08-1.33)	3.3	9.4E-04	19.1 kb upstream of the <i>HCG22</i> gene
rs2273017	1.30 (1.17-1.45)	5.0	7.0E-07	In the second intron of the <i>C6orf10</i> gene
rs9394159	1.28 (1.15-1.43)	4.5	6.7E-06	In the second intron of the <i>ITPR3</i> gene
rs4713693	1.20 (1.07-1.35)	3.2	1.4E-03	46.1 kb upstream of the <i>MLN</i> gene

Abbreviations: CI, confidence interval; GD, Graves' disease; OR, odds ratio; SNPs, single nucleotide polymorphisms.

**Table 3 Increased GD risk with increased risk score based on the genotypes of seven SNPs selected by the stepwise logistic regression analysis**

No. of risk alleles	Frequency (control/case)	Average risk score (min-max)	OR (95% CI)
0-3	0.07/0.02	0.57 (0.00-0.78)	1
4	0.07/0.03	0.91 (0.79-1.04)	2.17 (1.19-3.98)
5	0.10/0.07	1.14 (0.97-1.28)	3.01 (1.74-5.20)
6	0.14/0.09	1.36 (1.16-1.55)	2.98 (1.75-5.07)
7	0.15/0.13	1.59 (1.40-1.74)	4.06 (2.41-6.83)
8	0.15/0.16	1.80 (1.64-1.99)	5.20 (3.11-8.68)
9	0.14/0.16	2.02 (1.84-2.23)	5.31 (3.17-8.87)
10	0.10/0.15	2.25 (2.10-2.42)	7.37 (4.38-12.4)
11	0.05/0.12	2.45 (2.33-2.60)	11.7 (6.83-20.1)
12-14	0.03/0.07	2.74 (2.60-3.11)	11.8 (6.65-21.0)

Abbreviations: CI, confidence interval; GD, Graves' disease; OR, odds ratio; SNPs, single nucleotide polymorphisms. The P-value for the linear relationship of the risk score and the logarithm of OR (95% CI) are  $2.1 \times 10^{-44}$  and 2.72 (2.36-3.13), respectively. The plot of case/control frequencies is shown in Supplementary Figure S2.



**Figure 2** A receiver-operating characteristic (ROC) curve to discriminate GD cases from healthy controls.

regression model were considered to represent seven independent GD-risk loci in the Japanese population. We assessed the implications of these findings for identification of individuals with GD before morbid events by evaluating the cumulative effects of risk alleles and conducting a ROC analysis. Although the identified GD-associated SNPs turned out to exhibit only a modest level of ability to predict the

development of GD, the sensitivity/specificity could potentially be elevated by combining the information of GD-associated HLA alleles.

In the previous studies, *HLA-A\*0206* and *HLA-DPBI\*0501* have been reported to be associated with GD in the Japanese population.<sup>26,27</sup> Because we did not genotype HLA alleles in this study, to assess the extent of LD between these HLA alleles and the

seven GD-associated SNPs in the multivariate stepwise logistic regression model, we utilized the information of tagging SNPs in the GD-associated HLA alleles and the HapMap data for the Japanese population. Genotype data of the tagging SNPs for *HLA-DPBI\*0501* (rs10484569, rs2281388, rs6457713 and rs9380343)<sup>28</sup> and for *HLA-A\*0206* (rs892666–rs1632994 (rs2517699) haplotype),<sup>22</sup> and those of the seven GD-associated SNPs were retrieved from the HapMap2 (Release #24) or HapMap3 (Release #3) Public data set (<http://hapmap.ncbi.nlm.nih.gov/>) and analyzed together using the Haploview software.<sup>20</sup> No strong LD was detected between the tagging SNPs for *HLA-DPBI\*0501* and the seven GD-associated SNPs (Supplementary Figure S3A). However, the T-C haplotype among the rs892666–rs1632994 haplotypes, which is the tagging SNP haplotype for *HLA-A\*0206*, was in complete LD ( $|D'|=1$ ) with the risk allele (T allele) of rs4313034, and in high LD ( $|D'|=0.72$ ) with the risk allele (G allele) of rs3893464 (Supplementary Figure S3B). Collectively, these results indicate that, whereas the association of rs3893464 and rs4313034 with GD may be explained by their LD relationship with *HLA-A\*0206* (or vice versa), the other five SNPs represent GD-risk loci independent of *HLA-A\*0206* and *HLA-DPBI\*0501*.

Comprehensive and large-scale genotyping of HLA alleles and high-density SNPs for the entire MHC region is expected to provide us with genotype information of GD-associated HLA alleles and additional GD susceptibility variants that are independent of the GD-associated HLA alleles. As the MHC region encodes a large number of non-HLA genes involved in the immune system, such genes might be associated with the susceptibility to GD. It has been demonstrated that multiple uncommon or rare genetic variants across large regions of the genome can easily create synthetic genome-wide associations.<sup>29</sup> Therefore, considering the highly polymorphic and rapidly evolving nature of the extended MHC region, systematic identification of rare variants by targeted resequencing of this genomic interval seems to be inevitable to generate fundamental data to unequivocally identify causal variants of autoimmune diseases. Establishing a catalog of GD-associated genetic variants within the MHC region facilitates understanding the roles of the MHC in the genetic etiology of GD and improving the diagnostic ability to predict the GD phenotype of a given individual using genotype information.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We thank all participating doctors and staff from collaborating institutes for providing DNA samples. This work was supported by a grant from 'the Leading Project for Personalized Medicine' in the Ministry of Education, Culture, Sports, Science and Technology, Japan, and a Grant-in-Aid for Scientific Research (B) from the Japan Society for the Promotion of the Science.

- Jacobson, E. M. & Tomer, Y. The CD40, CTLA-4, thyroglobulin, TSH receptor, and PTPN22 gene quintet and its contribution to thyroid autoimmunity: back to the future. *J. Autoimmun.* **28**, 85–98 (2007).
- Fisfalen, M. E. & DeGroot, L. J. in (ed Weintraub, B.D.) *Grave's disease and autoimmune thyroiditis*, in *Molecular Endocrinology: Basic Concepts and Clinical Correlations* 319–370 (Raven Press: New York, 1995).

- Stassi, G. & Maria, R. D. Autoimmune thyroid disease: new models of cell death in autoimmunity. *Nat. Rev. Immunol.* **2**, 195–204 (2002).
- Tomer, Y. & Davies, T. F. Searching for the autoimmune thyroid disease susceptibility genes: from gene mapping to gene function. *Endocr. Rev.* **24**, 694–717 (2003).
- Simmonds, M. J. & Gough, S. C. Unravelling the genetic complexity of autoimmune thyroid disease: HLA, CTLA-4 and beyond. *Clin. Exp. Immunol.* **136**, 1–10 (2004).
- Zeitlin, A. A., Simmonds, M. J. & Gough, S. C. Genetic developments in autoimmune thyroid disease: an evolutionary process. *Clin. Endocrinol. (Oxf)* **68**, 671–682 (2008).
- Hiratani, H., Bowden, D. W., Ikegami, S., Shirasawa, S., Shimizu, A., Iwatani, Y. *et al.* Multiple SNPs in intron 7 of thyrotropin receptor are associated with Graves' disease. *J. Clin. Endocrinol. Metab.* **90**, 2898–2903 (2005).
- Brand, O. J., Barrett, J. C., Simmonds, M. J., Newby, P. R., McCabe, C. J., Bruce, C. K. *et al.* Association of the thyroid stimulating hormone receptor gene (TSHR) with Graves' disease. *Hum. Mol. Genet.* **18**, 1704–1713 (2009).
- Furugaki, K., Shirasawa, S., Ishikawa, N., Ito, K., Ito, K., Kubota, S. *et al.* Association of the T-cell regulatory gene CTLA4 with Graves' disease and autoimmune thyroid disease in the Japanese. *J. Hum. Genet.* **49**, 166–168 (2004).
- Kochi, Y., Yamada, R., Suzuki, A., Harley, J. B., Shirasawa, S., Sawada, T. *et al.* A functional variant in FcRH3, encoding Fc Receptor Homolog 3, is associated with rheumatoid arthritis and several autoimmunities. *Nat. Genet.* **37**, 478–485 (2005).
- Shirasawa, S., Harada, H., Furugaki, K., Akamizu, T., Ishikawa, N., Ito, K. *et al.* SNPs in the promoter of a B cell-specific antisense transcript, SAS-ZFAT, determine susceptibility to autoimmune thyroid disease. *Hum. Mol. Genet.* **13**, 2221–2231 (2004).
- Koyanagi, M., Nakabayashi, K., Fujimoto, T., Gu, N., Baba, I., Takashima, Y. *et al.* ZFAT expression in B and T lymphocytes and identification of ZFAT-regulated genes. *Genomics* **91**, 451–457 (2008).
- Lette, G. & Rioux, J. D. Autoimmune diseases, insights from genome-wide association studies. *Hum. Mol. Genet.* **17**, R116–R121 (2008).
- Wellcome Trust Case Control Consortium & Australo-Anglo-American Spondylitis Consortium. Association scan of 14 500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat. Genet.* **39**, 1329–1337 (2007).
- Sakai, K., Shirasawa, S., Ishikawa, N., Ito, K., Tamai, H., Kuma, K. *et al.* Identification of susceptibility loci for autoimmune thyroid disease to 5q31–q33 and Hashimoto's thyroiditis to 8q23–q24 by multipoint affected sib-pair linkage analysis in Japanese. *Hum. Mol. Genet.* **10**, 1379–1386 (2001).
- Unoki, H., Takahashi, A., Kawaguchi, T., Hara, K., Horikoshi, M., Andersen, G. *et al.* SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat. Genet.* **40**, 1098–1102 (2008).
- Ohnishi, Y., Tanaka, T., Ozaki, K., Yamada, R., Suzuki, H. & Nakamura, Y. A high-throughput SNP typing system for genome-wide association studies. *J. Hum. Genet.* **46**, 471–477 (2001).
- Ioannidis, J. P., Patsopoulos, N. A. & Evangelou, E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One* **2**, e841 (2007).
- Nakaoka, H. & Inoue, I. Meta-analysis of genetic association studies: methodologies, between-study heterogeneity and winner's curse. *J. Hum. Genet.* **54**, 615–623 (2009).
- Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
- Bilguvar, K., Yasuno, K., Niemelä, M., Ruigrok, Y. M., von Und Zu Fraunberg, M., van Duijn, C. M. *et al.* Susceptibility loci for intracranial aneurysm in European and Japanese populations. *Nat. Genet.* **40**, 1472–1477 (2008).
- de Bakker, P. I., McVean, G., Sabeti, P. C., Miretti, M. M., Green, T., Marchini, J. *et al.* A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.* **38**, 1166–1172 (2006).
- Fernando, M. M., Stevens, C. R., Walsh, E. C., De Jager, P. L., Goyette, P., Plenge, R. M. *et al.* Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet.* **4**, e1000024 (2008).
- Rioux, J. D., Goyette, P., Vyse, T. J., Hammarström, L., Fernando, M. M., Green, T. *et al.* Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. *Proc. Natl Acad. Sci. USA* **106**, 18680–18685 (2009).
- Newby, P. R., Pickles, O. J., Mazumdar, S., Brand, O. J., Carr-Smith, J. D., Pearce, S. H. *et al.* Follow-up of potential novel Graves' disease susceptibility loci, identified in the UK WTCCC genome-wide nonsynonymous SNP study. *Eur. J. Hum. Genet.* **18**, 1021–1026 (2010).
- Dong, R. P., Kimura, A., Okubo, R., Shinagawa, H., Tamai, H., Nishimura, Y. *et al.* HLA-A and DPB1 loci confer susceptibility to Graves' disease. *Hum. Immunol.* **35**, 165–172 (1992).
- Takahashi, M., Yasunami, M., Kubota, S., Tamai, H. & Kimura, A. HLA-DPBI\*0202 is associated with a predictor of good prognosis of Graves' disease in the Japanese. *Hum. Immunol.* **67**, 47–52 (2006).
- Kamatani, Y., Wattanapokayakit, S., Ochi, H., Kawaguchi, T., Takahashi, A., Hosono, N. *et al.* A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. *Nat. Genet.* **41**, 591–595 (2009).
- Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D. B. Rare variants create synthetic genome-wide associations. *PLoS Biol.* **8**, e1000294 (2010).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)