OPEN ACCESS

ORIGINAL ARTICLE

Fine mapping MHC associations in Graves' disease and its clinical subtypes in Han Chinese

Xun Chu,[1,2,3] Minjun Yang,[3] Zhen-Ju Song,[4] Yan Dong,[1,2] Chong Li,[3] Min Shen,[3] Yong-Qiang Zhu,[3] Huai-Dong Song,[5] Sai-Juan Chen,[5] Zhu Chen,[5] Wei Huang[3]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2017-105146).

[1]Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[2]Shanghai Key Laboratory of Pediatric Gastroenterology and Nutrition, Shanghai Institute for Pediatric Research, Shanghai, China
[3]Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center and Shanghai Academy of Science & Technology, Shanghai, China
[4]Department of Emergency Medicine, Zhongshan Hospital, Fudan University, Shanghai, China
[5]Shanghai Institute of Hematology, Ruijin Hospital Affiliated to Shanghai Jiaotong University (SJTU) School of Medicine, Shanghai, China, China

Correspondence to
Dr Xun Chu, Xinhua Hospital, Shanghai Institute for Pediatric Research, Shanghai Jiao Tong University School of Medicine, Shanghai 200092, China; chuxun@xinhuamed.com.cn

Received 6 November 2017
Revised 6 June 2018
Accepted 12 June 2018
Published Online First 9 July 2018

To cite: Chu X, Yang M, Song Z-J, *et al*. *J Med Genet* 2018;**55**:685–692.

ABSTRACT

Background The classical human leucocyte antigen (HLA) genes were the most important genetic determinant for Graves' disease (GD). The aim of the study was to fine map causal variants of the HLA genes.

Methods We applied imputation with a Pan-Asian HLA reference panel to thoroughly investigate themajor histocompatibility complex (MHC) associations with GD down to the amino acid level of classical *HLA* genes in 1468 patients with GD and 1490 controls of Han Chinese.

Results The strongest finding across the HLA genes was the association with HLA-DPβ1 position 205 ($P_{omnibus}=2.48\times10^{-33}$). HLA-DPA1*02:02 was the strongest association among the classical *HLA* alleles, which was in perfect linkage disequilibrium with HLA-DPα1 residue Met11 (OR=1.90, $P_{binary}=1.76\times10^{-31}$). Applying stepwise conditional analysis, we identified amino acid position 205 in HLA-DPβ1, position 66 and 99 in HLA-B and position 28 in HLA-DRβ1 explain majority of the MHC association to GD risk. We further evaluated risk of two clinical subtypes of GD, namely persistent thyroid stimulating hormone receptor antibody -positive (pTRAb+) group and 'non-persistent TRAb positive' (pTRAb−) group after antithyroid drug therapy. We found that HLA-B residues Lys66-Arg69-Val76 could drive pTRAb− GD risk alone, while HLA-DPβ1 position 205, HLA-B position 69 and 199 and HLA-DRβ1 position 28 drive pTRAb+ GD risk. The risk heterogeneity between pTRAb+ and pTRAb− GD might be driven by HLA-DPα1 Met11.

Conclusions Four amino acid positions could account for the associations of MHC with GD in Han Chinese. These distinct *HLA* association patterns indicated the two subtypes have distinct molecular mechanisms of pathogenesis.

INTRODUCTION

Graves' disease (GD) is a common autoimmune disease, and the major disease autoantigen, thyroid-stimulating hormone receptor (TSHR) autoantibody (TRAb), was found to underlie its pathogenesis. Conservative therapy with antithyroid drug (ATD) is proven to be effective in achieving euthyroidism in Graves' patients; however, 20%–50% of patients relapse after therapy withdrawal.[1 2] TRAb-positive patients are more prone to relapses than TRAb-negative patients at the time of ATD withdrawal, and it was found that these two subtypes have different genetic architecture. The genetic contribution of GD was estimated to be 79%, and it was found that the human major histocompatibility complex (MHC), human leucocyte antigen (HLA), was the most predominant genetic factor.[3–5] However, the high degree of linkage disequilibrium (LD) and genetic diversity challenges the determination of the causal functional variants in this region.

In Caucasian population, strong association with *HLA-DR3* (DRB1*03:01) has frequently been reported, which makes the DR3 haplotype as the predisposing factor to GD.[6–8] The 74th amino acids located in the peptide-binding grove of the HLA-β1 chain was found to determine the association of *HLA-DR3* with GD susceptibility in Caucasians.[7] In addition, the strongest association for GD in Caucasian has been reported with the HLA-class I allele, HLA-C*07.[9] Recently, our understanding of the landscape of MHC association was further extended by studies in Asian population.[5 10–12] Chen *et al*[12] performed direct and comprehensive genotyping of six classical *HLA* loci to four-digit resolution in Han Chinese population in Taiwan. Our previous genome-wide association study (GWAS) in Han Chinese population from China revealed three SNPs explained for most of association of MHC with GD.[5] More recently, Okada *et al*[11] applied HLA imputation to GWAS data for GD in Japanese and found that amino acid polymorphisms of multiple class I and class II *HLA* genes independently contribute to disease risk. The strongest impact was observed at *HLA-DPB1* in all of the above three Asian studies.[5 11 12]

Recent studies on *HLA* associations in GD and other diseases suggest that amino acid variants were the major driver conferring risk for diseases rather than the classical alleles.[11 13] In this study, we conducted a fine-mapping study assessing the GD associations with SNP, HLA amino acid variants, two-digit *HLA* alleles and four-digit *HLA* alleles simultaneously in Han Chinese populations by the imputation approach in our previous GWAS data.[5] Then, we explored the association patterns of *HLA* variants that distinguished the risks of different clinical subtypes.

METHODS

Study population and clinical characteristics

All samples were recruited from Han Chinese population through collaboration with multiple hospitals in China. All the enrolled subjects provided written informed consent for participation in the study approved by the local institutional review board. In

total, 1536 patients with GD and 1516 sex-matched controls were recruited for genotyping in GWAS.[5] Demographic information was shown in online supplementary table S1. Patients were divided into different subsets by levels of TRAb after ATD and with or without ophthalmopathy. Plasma levels of TRAb in all patients with GD who had been treated with ATD for ≥1 year were remeasured by quantitative ELISA (RSR Limited, UK). GD patients with TRAb levels ≥1.5 U/L were defined as 'persistent TRAb positive' (pTRAb+) and those with TRAb levels <1.5 U/L were defined as 'non-persistent TRAb positive' (pTRAb−). Of the 1536 patients with GD, 997 patients belong to the pTRAb+ group and 410 patients belong to the pTRAb− group. There were 602 patients diagnosed with GD orbitopathy and 866 patients without orbitopathy.

Genotyping was performed using Illumina Human660-Quad BeadChips.[5] The details of quality control (QC) steps have been described in previous study.[5] Briefly, after stringent QC, the genotypes of 486 049 SNPs in 1468 patients with GD and 1490 controls were kept for association analysis with an overall call rate of 98%, a minor allele frequency over 1% and the p values for Hardy-Weinberg equilibrium test in the controls over 10^{-6}. Principal component analysis (PCA) and multidimensional scaling analysis provided minimal evidence for population stratification in the current sample collection (λ_{GC}=1.02; online supplementary figure 1).

HLA imputation

We extracted 2676 SNP genotypes from ~25 to ~35 Mbp at chromosome 6 (Hg19, build 37) harbouring the extended MHC region from the GWAS data after QC. Imputation of two-digit and four-digit classical *HLA* alleles and amino acid polymorphisms of the HLA genes along with the SNPs that were not genotyped in the GWAS was performed using the SNP2HLA tool (https://www.broadinstitute.org/mpg/snp2hla/) and the Pan-Asian reference panel.[14–16] We applied postimputation QC criteria of minor allele frequency of ≥1% and PLINK INFO of ≥0.8 for the association analysis.

Association analysis

In the 5264 variants past quality control, a total of 84 amino acid sites were multiallelic amino acid polymorphisms; as a result, we considered the study-wide significance threshold to be $p=9.35\times10^{-6}$ (0.05/(5264+84)). Unless specified, all analyses were conducted using R 3.3.2 software.

We obtained dosage values of imputed markers and performed association analysis for both the genotyped SNPs and the imputed markers. Biallelic markers included two-allele SNPs, two-residue amino acid positions were encoded as allele 1 and allele 2. Multiresidue amino acid positions, two-digit and four-digit classical *HLA* alleles and haplotypes were encoded as presence and absence of each allele of multiallelic (≥1 alleles) markers. GD association of the imputed dosage of each marker was first examined by binary logistic regression analysis. To account for stratifications, we included sex and the top five principal components (PCs) as covariates. PCs were calculated by using SmartPCA.[17]

We also tested the possible influence of the positions with multiallelic residues using a logistic regression model by means of an omnibus test (see online supplementary methods).

Stepwise conditional analysis was performed to find additional markers with independent GD risk effect by adding the top associated markers as covariates in logistic regression. For conditional analysis on the specific HLA amino acid positions, we included the multiallelic variants of the amino acid residues as covariates.[18]

In order to account for LD within the region, the conditional haplotype method was also applied to a subset of amino acids accounting for MHC association identified from conditional logistic analysis.[19]

RESULTS

Summary of imputation results

The Pan-Asian reference panel including data of 530 individuals was used as reference in the imputation analysis.[14–16] Using the GWAS data of 1468 patients with GD and 1490 controls, we imputed 5267 markers (with minor allele frequency ≥1% and PLINK INFO ≥0.8) from ~29 Mb to ~35 Mb on chromosome 6p21.3 with SNP2HLA software.[14] The resulting data included 4299 SNPs, 173 two-digit and four-digit classical alleles and 792 amino acid residues for the eight classical *HLA* genes (*HLA-A*, *HLA-B*, *HLA-C*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1* and *HLA-DRB1*). Association analysis revealed that 1381 markers met the study-wide significance ($p\leq9.35\times10^{-6}$; online supplementary table S2).

Associations of classical *HLA* genes with GD

We first investigated the association-specific amino acid residues and classical alleles of each classical *HLA* gene with GD susceptibility (online supplementary table S2). The most significant amino acid residue at HLA-A was Arg151 (OR=0.58, $P_{binary}=1.71\times10^{-18}$; online supplementary figure 2A), and

Table 1 Associations of HLA amino acid residues with Graves' disease risk in Han Chinese

Genes	Effective residue	PLINK INFO	ERF (%)			OR (95% CI)	P values
			Cases (n=1468)	Controls (n=1490)	Europeans*		
HLA-A	Arg151	0.99	19.6	29.6	11.5	0.58 (0.51 to 0.65)	1.71×10^{-18}
HLA-B	Lys66-Arg69-Val76	0.95	14.1	6.5	–	2.38 (2.00 to 2.87)	5.81×10^{-21}
HLA-C	Tyr116	1.03	37.9	27.8	10.9	1.59 (1.41 to 1.76)	1.15×10^{-15}
HLA-DPA1	Met11	0.92	59.5	44.8	3.5	1.90 (1.70 to 2.12)	1.76×10^{-31}
HLA-DPB1	Leu35	0.93	50.9	36.6	9.4	1.80 (1.72 to 2.15)	8.73×10^{-29}
HLA-DQA1	Lys47- His52-Leu54	1.04	7.0	15.2	12.9	0.42 (0.36 to 0.51)	1.93×10^{-21}
HLA-DQB1	Ala-10	0.98	86.9	78.9	82.4	1.78 (1.55 to 2.04)	1.18×10^{-15}
HLA-DQB1	Asp57	0.96	72.7	63.6	54.7	1.53 (1.37 to 1.71)	1.79×10^{-13}
HLA-DRB1	Gly11-Tyr13-Lys14-Gln25-Leu30-Gln74	1.04	7.1	15.2	10.3	0.42 (0.36 to 0.51)	1.93×10^{-21}

*Amino acid residue frequencies of Europeans from the Hapmap CEU population.[14]
ERF, effective residue frequency; HLA, human leucocyte antigen.

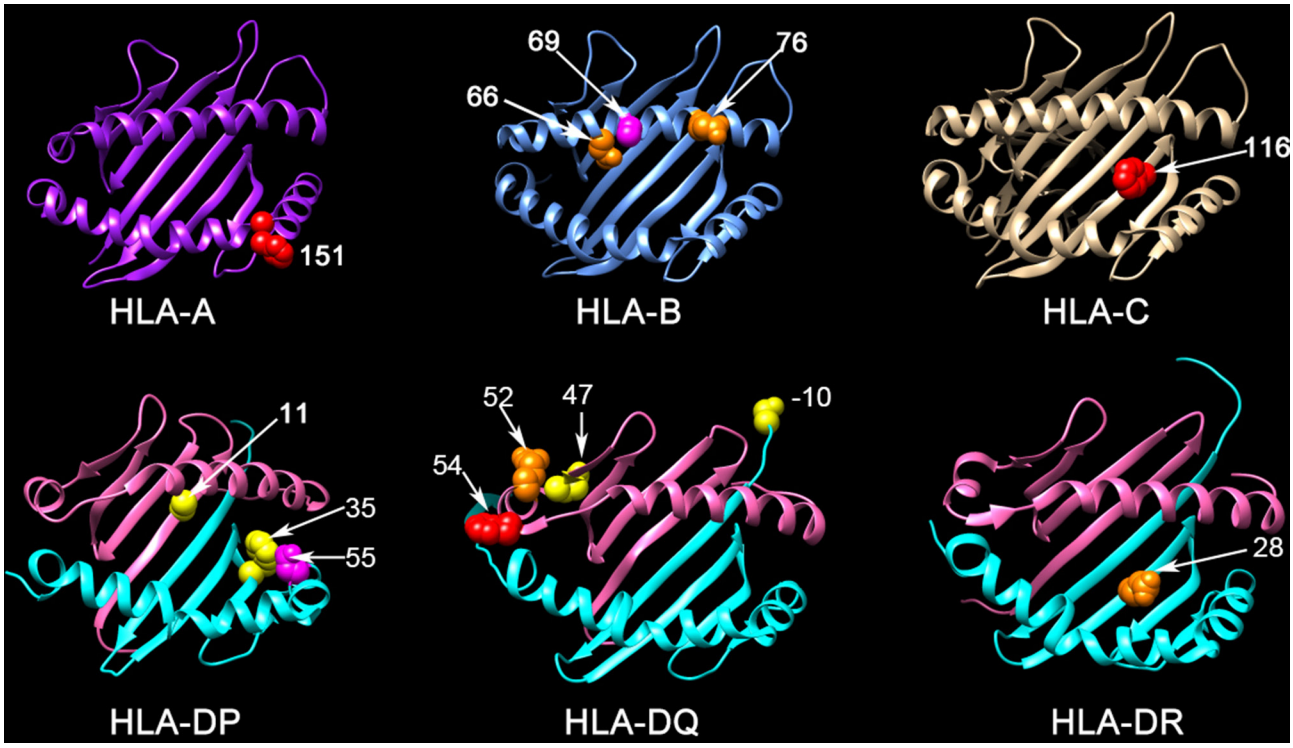

Figure 1 Three-dimensional ribbon models of HLA amino acid positions associated with GD risk. The protein structures of HLA-A, HLA-B, HLA-C, HLA-DP, HLA-DQ and HLA-DR are based on Protein Data Bank (PDB) entries 2XPG, 2BVP, 4NT6, 3LQZ, 4Z7W and 3PDO, respectively. GD associated amino acid positions identified by the association analysis are highlighted. This figure was prepared using UCSF Chimera.

HLA-A*02:07 was the top classical *HLA-A* allele tagged by Cys99 (OR=2.06, $P_{binary}=2.07\times10^{-12}$; table 1, figure 1). The most significantly associated residue of *HLA-B* gene is Arg69 (OR=2.38, $P_{binary}=5.81\times10^{-21}$), which was in perfect LD with Lys66 and Val76 ($r^2=1$; table 1, online supplementary figure 2B). The most associated *HLA-B* classical allele was HLA-B*46:01 (OR=2.38, $P_{binary}=8.78\times10^{-21}$; table 2), which could be tagged by residues Arg69, Lys66 and Val76. Positions 66, 69 and 76 of the HLA-B chain are all located on the α-helix wall of peptide binding pocket (figure 1). The top association signal at *HLA-C* was residue Tyr116 (OR=1.59, $P_{binary}=1.15\times10^{-15}$) and HLA-C*01:02 was the most associated *HLA-C* classical allele (OR=1.84, $P_{binary}=5.40\times10^{-15}$; table 2, online supplementary figure 2C).

HLA-DPα1 residue Met11 is the peak binary signal of *HLA-DPA1* as well as the most significant residues the MHC region ($P_{binary}=1.76\times10^{-31}$, OR=1.90; online supplementary figure 2D), which tagged HLA-DPA1*02:02 (tables 1 and 2). HLA-DPα1 position 11 is located within the β-sheet floor of peptide binding pocket (figure 1). The top residue association signal at HLA-DPB1 corresponded to Leu35 (OR=1.80, $P_{binary}=8.73\times10^{-29}$; online supplementary figure 2E). HLA-DPβ1 Glu55 was in strong LD with Leu35 ($r^2=0.96$) and had a similar association (OR=1.91, $P_{binary}=1.09\times10^{-28}$). HLA-DPβ1 Leu35 was the second significant residues across the MHC region, which exhibits moderate LD with HLA-DPα1 residue Met11 ($r^2=0.61$). The association of HLA-DPβ1 Leu35 was still significant, though not meet study-wide significance, when conditioning on HLA-DPα1 Met11 ($p=3.58\times10^{-5}$). Similarly, HLA-DPα1 Met11 was significantly associated with GD when conditioning on HLA-DPβ1 Leu35 ($p=2.13\times10^{-5}$). It is obvious these two amino acids represent independent associations. The most significant four-digit allele at *HLA-DPB1* was HLA-DPB1*05:01 (OR=1.72, $P_{binary}=1.73\times10^{-26}$), which was

Table 2 Associations of classical HLA alleles with Graves' disease susceptibility in Han Chinese

		Allele frequency (%)*				
HLA allele	PLINK INFO	Cases (n=1468)	Controls (n=1490)	Caucasian populations	OR (95% CI)	P values
HLA-A*02:07	0.90	9.7	4.9	0.007	2.10 (1.70 to 2.59)	2.07×10^{-12}
HLA-B*46:01	0.94	14.1	6.5	0.010	2.38 (1.99 to 2.86)	8.78×10^{-21}
HLA-C*01:02	1.02	18.4	10.9	3.432	1.83 (1.57 to 2.12)	5.40×10^{-15}
HLA-DPA1*02:02	0.92	59.5	44.8	6.324	1.90 (1.70 to 2.12)	1.76×10^{-31}
HLA-DPB1*05:01	0.92	44.0	31.3	1.927	1.90 (1.69 to 2.14)	1.73×10^{-26}
HLA-DQA1*02:01	1.04	7.0	15.2	12.658	0.43 (0.36 to 0.51)	1.93×10^{-21}
HLA-DQB1*02:01	1.00	10.9	17.8	15.437	0.57 (0.49 to 0.66)	2.31×10^{-13}
HLA-DRB1*07:01	1.04	7.1	15.3	13.337	0.43 (0.36 to 0.51)	2.49×10^{-21}

*The allele frequency of Caucasian populations was calculated from the average frequencies of all samples of Caucasiod ethnic origin from database: http://www.allelefrequencies.net.[23]
HLA, human leucocyte antigen.

found as the most associated *HLA* marker in our previous GWAS study and other studies.[5 11 12] HLA-DPB1*05:01 was in strong LD with Leu35 (r^2=0.76). The association of HLA-DPB1*05:01 was not significant when conditioning on HLA-DPβ1 Leu35 (p=0.09), whereas HLA-DPβ1 Leu35 was significantly associated with GD when conditioning on HLA-DPB1*05:01 ($p=7.12\times10^{-6}$).

The top residue association signal at *HLA-DQA1* was three residues in completely perfect LD (Lys47, His52 and Leu54; OR=0.42, $p=1.91\times10^{-21}$; figure 1, online supplementary figure 2F), which tagged the four-digit allele HLA-DQA1*02:01 (table 1). Ala-10 was the strongest associated residues of *HLA-DQB1* (OR=1.78, $P_{binary}=1.18\times10^{-15}$; online supplementary figure 2G) and the most significant four-digit allele at *HLA-DQB1* was HLA-DQB1*02:01 (OR=0.57, $p=2.31\times10^{-13}$). Interestingly, HLA-DQB1 Asp57 is the most protective allele for type 1 diabetes, and it is interesting that it was found to increase the risk of GD (OR=1.53, $p=1.79\times10^{-13}$).[20]

Six HLA-DRβ1 residues (Gly11, Tyr13, Lys14, Gln25, Leu30 and Gln74; OR=0.42, $p=1.93\times10^{-21}$) uniquely encoded by HLA-DRB1*07:01 (OR=0.43, $p=2.49\times10^{-21}$) had the strongest associations with GD among the *HLA-DRB1* alleles and amino acid residues (tables 1 and 2; online supplementary figure 2H).

Associations of HLA amino acid positions and dependence analysis

We also tested the influence of the polymorphic amino acid positions by means of an omnibus test (online supplementary table S2). The most associated association across all variants corresponded to the HLA-DPβ1 position 205 ($P_{omnibus}=2.48\times10^{-33}$), followed by HLA-DPα1 position 11 and HLA-DPβ1 position 35. There are three possible residues (Val205, Met205 and Deletion205) at the amino acid position 205 of HLA-DPβ1. Among them, Met205 showed risk effects for GD (OR=1.88, $P_{binary}=5.81\times10^{-26}$), while Val205 and Deletion205 were protective (OR=0.72, $P_{binary}=1.99\times10^{-9}$ for Val205 and OR=0.40, $P_{binary}=5.17\times10^{-15}$ for Deletion205).

We applied stepwise conditional regression analysis to find independent HLA amino acid positions confer independent risks on GD (table 3; figure 2). When conditioning on the top associated HLA-DPβ1 position 205, HLA-B position 66 showed the most significant independent evidence of association (conditional $P_{omnibus}=1.31\times10^{-16}$), with HLA-B position 69 demonstrating a similar level of evidence (conditional $P_{omnibus}=1.41\times10^{-16}$). Conditioning on positions HLA-DPβ1 position 205 and HLA-B position 66 demonstrated an independent association of HLA-B position 199 (conditional $P_{omnibus}=8.65\times10^{-10}$). When conditioning on HLA-DPβ1 position 205 and HLA-B position 66 and 199, we detected the most significant independent association at HLA-DRβ1 position 28 (conditional $P_{omnibus}=1.29\times10^{-8}$). No significant associations were observed after adjusting for the effects of HLA-DPβ1 position 205, HLA-B position 66 and 199 and HLA-DRβ1 position 28 (conditional $P_{omnibus}>9.35\times10^{-6}$), suggesting that the combination of these amino acid positions explain the majority of the *HLA* risk in Chinese.

To account the high LD of the HLA region, we further performed a conditional haplotype test combining HLA-DPβ1 205 and HLA-DRβ1 28 in the class II region as well as HLA-B 66 and 199 (online supplementary table S3). Although different haplotypes surpassed the study-wide significance threshold, none of them showed a significant improvement in the association observed for the associated amino acids of HLA-DPβ1 205, HLA-DRβ1 28 and HLA-B 66 and 199 independently. HLA-DRβ1 28 has independent haplotypic effect after we controlled for HLA-DPβ1 205 ($p=1.78\times10^{-8}$). HLA-B 199 has independent haplotypic effect after we controlled for HLA-B 66 ($p=3.27\times10^{-9}$). No evidence of epistatic interactions between known non-HLA risk loci[5] and any of the HLA alleles described here.

Associations of *HLA* genes with GD clinical subtypes

Evaluated levels of hyrotropin receptor antibodies (TRAb) after ATD therapy are the predictors of relapse. We further evaluated associations of *HLA* genes with risk for pTRAb+ and pTRAb− GD, respectively.

Table 3 Association of the HLA amino acid positions with Graves' disease risk in Han Chinese

	Frequency (%)*			
HLA variant	Case (n=1468)	Control (n=1490)	OR (95% CI) †	P values†
HLA-DPβ1 amino acid position 205				
Met	45.1	32.4	1.88 (1.67 to 2.12)	1.18×10^{-2}
Deletion	3.8	9.1	0.40 (0.32 to 0.5)	3.32×10^{-15}
Val	51.1	58.5	(reference)	
HLA-B amino acid position 66				
Lys	14.1	6.5	2.39 (2.00 to 2.87)	4.42×10^{-13}
Asn	5.0	6.5	0.76 (0.61 to 0.95)	6.17×10^{-2}
Ile	80.8	87.0	(reference)	
HLA-B amino acid position 199				
Val	2.2	5.3	0.40 (0.30 to 0.54)	1.80×10^{-7}
Ala	97.8	94.7	(reference)	
HLA-DRβ1 amino acid position 28				
Glu	18.9	29.5	0.57 (0.50 to 0.64)	1.55×10^{-7}
His	15.3	11.3	1.40 (1.20 to 1.62)	2.76×10^{-1}
Asp	64.3	58.7	(reference)	

*Amino acid residues with frequency ≥0.005 in the controls are shown.
†Obtained from the multivariate regression model that included position 205 in HLA-DPβ1, position 66 and 99 in HLA-B and position 28 in HLA-DRβ1 identified by stepwise regression analysis.

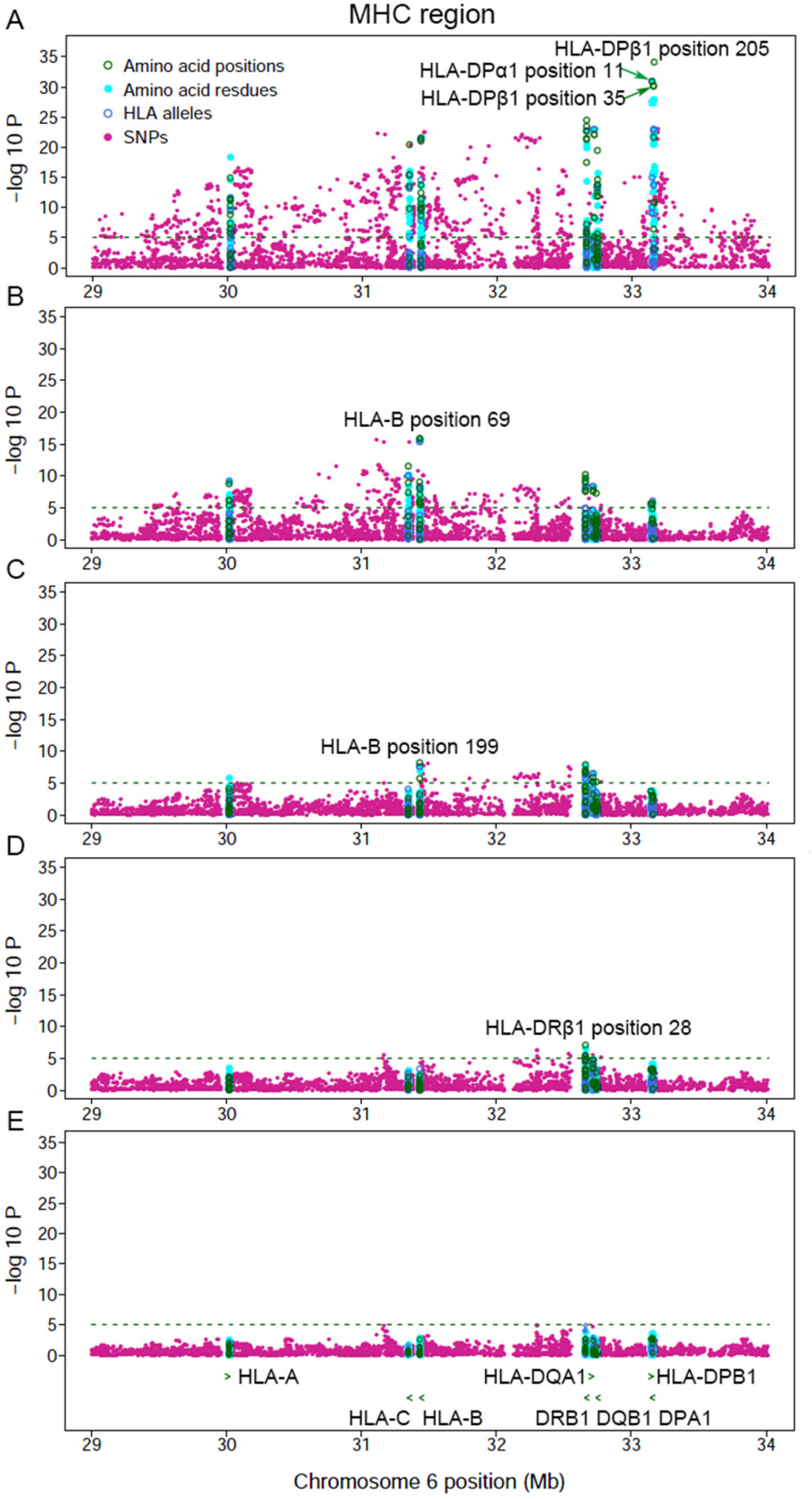

Figure 2 Regional association plots of *HLA* loci associated with GD in Han Chinese. Each dot represents the –log10 P of the variants, including SNPs, classical HLA alleles and amino acid polymorphisms encoded by the HLA genes. Physical positions are based on NCBI build 36 of the human genome. The green horizontal dashed line represents p=9.35×10−6. (A) The top signal was from HLA-DPβ1 position 205. (B) After conditioning on HLA-DPβ1 position 205, the top signal was from HLA-B position 66. (C) After conditioning on HLA-DPβ1 position 205 and HLA-B position 66, the top signal was from HLA-B position 199. (D) After conditioning on HLA-DPβ1 position 205 and HLA-B position 66 and 199, the most significant independent association was at HLA-DRβ1 position 28. (E) No study-wide significant associations were observed after adjusting for the effects of HLA-DPβ1 position 205, HLA-B position 66 and 199 and HLA-DRβ1 position 28 (conditional $P_{omnibus}$<9.35×10−6). GD, Graves' disease; MHC, major histocompatibility complex.

The most strongly associated variant with pTRAb+ GD risk was observed at HLA-DPβ1 positions 205 (997 pTRAb+ patients vs 1490 controls; $p=7.52\times10^{-35}$; online supplementary table S4). The effect sizes of HLA-DPβ1 position 205 tended to be greater in the pTRAb+ GD patients (997 pTRAb+ patients vs 1490 controls) than in the total of patients with GD (1468 patients vs 1490 controls). Then, we put HLA-DPβ1 position 205 in the regression model as the best locus and employed stepwise forward logistic regression to identify independent markers driving pTRAb+ GD risk. The analysis revealed that the HLA association with pTRAb+ GD could be explained by four amino acid positions, namely HLA-DPβ1 position 205, HLA-B position 66 and 199 and HLA-DRβ1 position 74 with conditional *P* ominibus values of 7.91×10^{-22}, 1.79×10^{-9}, 1.51×10^{-8} and 3.04×10^{-7}, respectively.

The strongest association with pTRAb− GD risk was at HLA-B residues Lys66-Arg69-Val76 (410 pTRAb− patients vs 1490 controls; OR=2.64, $P_{binary}=9.73\times10^{-13}$) uniquely encoded by HLA-B*46:01, which could explain the disease risk in pTRAb patients alone (online supplementary table S5).

When we directly assessed comparative risk between pTRAb+ and pTRAb− subjects, we found the lowest p value of the nominal association signal at HLA-DPα1 Met11 ($P_{binary}=1.12\times10^{-7}$ for 410 pTRAb− patients vs 997 pTRAb+ patients). After conditioning on HLA-DPα1 Met11, we observed no significant association in the MHC region (conditional $p>9.35\times10^{-6}$). HLA-DPα1 Met11 increased pTRAb+ susceptibility in comparison with pTRAb− susceptibility (OR=1.50, 95% CI 1.28 to 1.77). Examining the classical alleles, we noted that HLA-DPA1*02:02 demonstrated the lowest p value for pTRAb+ versus pTRAb− association ($P_{binary}=1.12\times10^{-7}$) similar significantly with HLA-DPα1 Met11.

We also investigated association of *HLA* with GD orbitopathy. The comparison of *HLA* variation between GD patients with orbitopathy and without orbitopathy showed no significant difference.

DISCUSSION

The HLA genes were responsible for the strongest association signals for GD susceptibility.[14 15] Numerous studies have tried to clarify the biological mechanism of classical HLA genes underlying disease susceptibility and found some molecular clues. In the current study, we found that combinations of amino acid polymorphisms in multiple class I and II *HLA* genes explained the majority of risk in the MHC region for GD in Han Chinese.

We observed the top association signal at HLA-DPβ1 position 205, which was among the most associated variants in Japanese.[11] In Asians, the *HLA-DPB1* locus was constantly not covered in association analysis of *HLA* genes with GD previously, and it was overlooked until the report in 1992.[21] Association of HLA-DPB1*05:01 with GD was first reported in Japanese[21] and later replicated in several independent sample set of Japanese and Chinese populations.[10 12 22] In our previous GWAS for GD, rs2281388 conferred the top risk with GD susceptibility in Chinese population. Since rs2281388 was a tagging SNP that could predict HLA-DPB1*05:01 ($r^2=0.90$), we suggested the association of HLA-DPB1*05:01 with GD susceptibility in previous report. The frequency of HLA-DPB1*05:01 is mostly less than 5% in Caucasians, and the association of HLA-DPB1*05:01 with GD has not been reported in Caucasians.[12 23] It could be speculated that *HLA-DPB1* contributed much less to autoimmune disease in Caucasians than in Asians.

In current analysis, HLA-DPA1*02:02 showed the strongest association among the *HLA* classical alleles with GD in Han Chinese population, and HLA-DPB*05:01 was the second significant *HLA* classical allele. Although HLA-DPA1*02:02 was not the most associated HLA alleles in recent Japanese study, it still belonged to the top association signals.[11] The frequency of HLA-DPA1*02:02 is 44.8% in the current control population, while it is about 3%–6% in Caucasians.[23] To our knowledge, no association of HLA-DPA1*02:02 and HLA-DPB1*05:01 with GD has been reported in Caucasians. Since only six loci are routinely typed by laboratories, that is, *HLA-A*, *HLA-B* and *HLA-C* for class I and *HLA-DRB1*, *HLA-DQB1*, *HLA-DPB1* for class II, the *HLA-DPA1* locus was constantly neglected in previous studies. Several studies that reported the association of HLA-DPB*05:01 with GD did not cover the *HLA-DPA1* locus in their study design including several recent comprehensive analysis of *HLA* association with GD in Chinese and Japanese.[10 12 22] Fortunately, the *HLA-DPA1* variation was included in the commercial genome-wide SNP arrays, and it showed association with chronic hepatitis B virus (HBV) infection and related diseases in Asians.[24 25] Special attention should be paid to the attribution of *HLA-DPA1* to susceptibility of autoimmune diseases in Asian populations in future studies.

We observed the peak residue signal at HLA-DPα1 Met11 in our sample set of Han Chinese, which conferred risk to GD susceptibility. The second leading signal is in HLA-DPβ1 residue Leu35, which exhibits moderate LD with HLA-DPα1 Met11 ($r^2=0.52$). Conversely, HLA-DPβ1 Leu35 as the strongest association with GD in Japanese, while HLA-DPα1 Met11 was the second significant *HLA* classical allele.[11] The association significances of HLA-DPα1 Met11 and HLA-DPβ1 Leu35 were similar, but conditional regression analysis showed that two amino acids conferred mutually independent risk to GD. The different significant level of HLA-DPα1 Met11 and HLA-DPβ1 Leu35 association with GD between Chinese and Japanese might be caused by genetic heterogeneity or stochastic nature of sample collection. Undoubtedly, both *HLA-DPA* and *HLA-DPB* contribute substantially to the genetic architecture of GD in Asians.

HLA-B harbours independent association signals for GD risk, and HLA-B positions 66, 69 and 76 showed the top associated signals among the *HLA-I* gene. HLA-B Arg69, Lys66 and Val76 is coded by HLA-B*46:01. In our data, HLA-B*46:01 was the most significantly associated classical allele of *HLA-I* genes, which had good replications in Asians.[26] HLA-B*46:01 is very rare in Caucasians, and the frequencies are mostly less than 1%. It should be noticed that the strongest association for Caucasian GD has been reported with HLA-C*07,[9] which showed a weak association with GD in our sample-set (p=0.009, online supplementary table S2).

The third independent association corresponded to HLA-*DRB1*. HLA-DRB1 position β74 was a well-established genetic and functional candidate locus for GD risk.[7 8 27] Arginine at position 74 of the HLA-DRβ1 chain conferred risk to GD susceptibility independently to *HLA-DR3* in Caucasians[7]; however, the susceptible molecule in the current Chinese population was not arginine (p=0.20 for Arg74) but glutamic acid and leucine ($p=1.16\times10^{-2}$ for Glu74; $p=7.66\times10^{-6}$ for Leu74; online supplementary table S2). More importantly, the most effective molecule at position 74 in Chinese population was glutamine, which has a strong protective effect against GD (OR=0.43, $p=3.14\times10^{-21}$). Although our data suggested HLA-DRβ1 position 28 as one of primary determinant of susceptibility to GD, we cannot rule out the possibility that this association is due to other variation at the *HLA-DRB1* locus.

ATD is the first choice of treatment for patients with GD in China, and the relapse rates after therapy withdrawal are estimated to be 30%–50%.[1] Evaluated levels of hyrotropin receptor antibodies (TRAb) after ATD therapy are predictors of relapse.[1] Since different genetic contributions associating with different phenotypes point to different etipathology, we illustrated that pTRAb+ GD and pTRAb− GD exhibit different genetic associations in our previous study.[5] TSHR gene variation was associated with disease risk in pTRAb+ patients but not in pTRAb− patients.[5] Our current data refined the *HLA* associations with these two clinical subtypes. Association of disease risk for pTRAb− patients with *HLA* genes could be explained by HLA-B Lys66-Arg69-Val76 alone. Association of disease risk for pTRAb+ patients with *HLA* genes was explained by combinations of amino acid polymorphisms in multiple class I and II *HLA* genes. These distinct *HLA* signatures underlying the two subtypes suggest that they are genetically heterogeneous and might have distinct molecular mechanisms of pathogenesis.

It is conceivable that viral or bacterial peptides and HLA class I complex caused autoimmune attack of the thyroid gland in both pTRAb+ and pTRAb− patients with GD.[9] After GD initiation, HLA class II molecules binding peptides derived from TSHR might influence the production of TRAb. Since the extracellular domain of human TSHR (TSHR-ECD) is shed into the circulation, TSHR-ECD is a preferentially immunogenic portion of TSHR.[28] The complex of HLA-class II and TSHR-ECD epitope is presented to $CD4^+$ T cells. The activated $CD4^+$ T cells help B cell differentiation into autoantibody producing plasma cells that produce TRAb.[28] The TSHR-ECD epitope encoded by high-risk allele might have high affinity to HLA-class II molecules high-risk alleles, and such complex probably stimulates persistent T cell response.[29] Conversely, certain HLA alleles may not present important epitopes that induce TSHR antibodies.[29] Therefore, low-risk and high-risk HLA and TSHR alleles could affect the production of TRAb. It could be hypothesised that pTRAb+ GD patients with the risk alleles of *HLA- II* and/or *TSHR* have a high efficiency in antigen presentation causing persistent TRAb production and therefore has poor clinical outcome after ATD therapy.

In conclusion, our study revealed four amino acids could account for most *HLA* association with GD in Han Chinese. Two clinical subphenotypes corresponding to different responses after ATD therapy correlated with the inheritance of different sets of disease-risk HLA amino acids variants. Our study extended the knowledge of *HLA-DPA1* association with GD association, and special attention should be paid to *HLA-DPA1* variation in further study of *HLA* contribution to disease susceptibility in Asians.

Acknowledgements The authors would like to thank all the clinical physicians for the sample collection in the collaborating hospitals and all participants in this research. We gratefully acknowledge Dr Zhibin Hu and Dr Meng Zhu for their valuable comments and kind help in improving the quality of the manuscript.

Contributors XC designed research and wrote the paper; XC, MY, Z-JS and CL analysed data; MS, YD, Y-Q Z, H-DS, S-JC, ZC and WH collected clinical samples and performed experiments.

Funding This work was supported by the National Natural Science Foundation of China (31671317, 31471190, 81471840 and 31271343).

Competing interests None declared.

Patient consent Obtained.

Ethics approval Ethical Committee of the Chinese National Human Genome Center at Shanghai.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The snp allele frequencies of the cases and controls could be shared with researchers by contacting corresponding author by email.

REFERENCES

1 Cappelli C, Gandossi E, Castellano M, Pizzocaro C, Agosti B, Delbarba A, Pirola I, De Martino E, Rosei EA. Prognostic value of thyrotropin receptor antibodies (TRAb) in Graves' disease: a 120 months prospective study. *Endocr J* 2007;54:713–20.
2 Konishi T, Okamoto Y, Ueda M, Fukuda Y, Harusato I, Tsukamoto Y, Hamada N. Drug discontinuation after treatment with minimum maintenance dose of an antithyroid drug in Graves' disease: a retrospective study on effects of treatment duration with minimum maintenance dose on lasting remission. *Endocr J* 2011;58:95–100.
3 Brix TH, Kyvik KO, Christensen K, Hegedüs L. Evidence for a major role of heredity in Graves' disease: a population-based study of two Danish twin cohorts. *J Clin Endocrinol Metab* 2001;86:930–4.
4 Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand WH, Samani NJ, Todd JA, Donnelly P, Barrett JC, Davison D, Easton D, Evans DM, Leung HT, Marchini JL, Morris AP, Spencer CC, Tobin MD, Attwood AP, Boorman JP, Cant B, Everson U, Hussey JM, Jolley JD, Knight AS, Koch K, Meech E, Nutland S, Prowse CV, Stevens HE, Taylor NC, Walters GR, Walker NM, Watkins NA, Winzer T, Jones RW, McArdle WL, Ring SM, Strachan DP, Pembrey M, Breen G, St Clair D, Caesar S, Gordon-Smith K, Jones L, Fraser C, Green EK, Grozeva D, Hamshere ML, Holmans PA, Jones IR, Kirov G, Moskivina V, Nikolov I, O'Donovan MC, Owen MJ, Collier DA, Elkin A, Farmer A, Williamson R, McGuffin P, Young AH, Ferrier IN, Ball SG, Balmforth AJ, Barrett JH, Bishop TD, Iles MM, Maqbool A, Yuldasheva N, Hall AS, Braund PS, Dixon RJ, Mangino M, Stevens S, Thompson JR, Bredin F, Tremelling M, Parkes M, Drummond H, Lees CW, Nimmo ER, Satsangi J, Fisher SA, Forbes A, Lewis CM, Onnie CM, Prescott NJ, Sanderson J, Matthew CG, Barbour J, Mohiuddin MK, Todhunter CE, Mansfield JC, Ahmad T, Cummings FR, Jewell DP, Webster J, Brown MJ, Lathrop MG, Connell J, Dominiczak A, Marcano CA, Burke B, Dobson R, Gungadoo J, Lee KL, Munroe PB, Newhouse SJ, Onipinla A, Wallace C, Xue M, Caulfield M, Farrall M, Barton A, Bruce IN, Donovan H, Eyre S, Gilbert PD, Hilder SL, Hinks AM, John SL, Potter C, Silman AJ, Symmons DP, Thomson W, Worthington J, Dunger DB, Widmer B, Frayling TM, Freathy RM, Lango H, Perry JR, Shields BM, Weedon MN, Hattersley AT, Hitman GA, Walker M, Elliott KS, Groves CJ, Lindgren CM, Rayner NW, Timpson NJ, Zeggini E, Newport M, Sirugo G, Lyons E, Vannberg F, Hill AV, Bradbury LA, Farrar C, Pointon JJ, Wordsworth P, Brown MA, Franklyn JA, Heward JM, Simmonds MJ, Gough SC, Seal S, Stratton MR, Rahman N, Ban M, Goris A, Sawcer SJ, Compston A, Conway D, Jallow M, Newport M, Sirugo G, Rockett KA, Bumpstead SJ, Chaney A, Downes K, Ghori MJ, Gwilliam R, Hunt SE, Inouye M, Keniry A, King E, McGinnis R, Potter S, Ravindrarajah R, Whittaker P, Widden C, Withers D, Cardin NJ, Davison D, Ferreira T, Pereira-Gale J, Hallgrimsdo'ttir IB, Howie BN, Su Z, Teo YY, Vukcevic D, Bentley D, Brown MA, Compston A, Farrall M, Hall AS, Hattersley AT, Hill AV, Parkes M, Pembrey M, Stratton MR, Mitchell SL, Newby PR, Brand OJ, Carr-Smith J, Pearce SH, McGinnis R, Keniry A, Deloukas P, Reveille JD, Zhou X, Sims AM, Dowling A, Taylor J, Doan T, Davis JC, Savage L, Ward MM, Learch TL, Weisman MH, Brown M. Wellcome Trust Case Control Consortium Australo-Anglo-American Spondylitis Consortium (TASC) Biologics in RA Genetics and Genomics Study Syndicate (BRAGGS) Steering Committee Breast Cancer Susceptibility Collaboration (UK). Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet* 2007;39:1329–37.
5 Chu X, Pan CM, Zhao SX, Liang J, Gao GQ, Zhang XM, Yuan GY, Li CG, Xue LQ, Shen M, Liu W, Xie F, Yang SY, Wang HF, Shi JY, Sun WW, Du WH, Zuo CL, Shi JX, Liu BL, Guo CC, Zhan M, Gu ZH, Zhang XN, Sun F, Wang ZQ, Song ZY, Zou CY, Sun WH, Guo T, Cao HM, Ma JH, Han B, Li P, Jiang H, Huang QH, Liang L, Liu LB, Chen G, Su Q, Peng YD, Zhao JJ, Ning G, Chen Z, Chen JL, Chen SJ, Huang W, Song HD. China Consortium for Genetics of Autoimmune Thyroid Disease. A genome-wide association study identifies two new risk loci for Graves' disease. *Nat Genet* 2011;43:897–901.
6 Zamani M, Spaepen M, Bex M, Bouillon R, Cassiman JJ. Primary role of the HLA class II DRB1*0301 allele in Graves disease. *Am J Med Genet* 2000;95:432–7.
7 Ban Y, Davies TF, Greenberg DA, Concepcion ES, Osman R, Oashi T, Tomer Y. Arginine at position 74 of the HLA-DR beta1 chain is associated with Graves' disease. *Genes Immun* 2004;5:203–8.
8 Simmonds MJ, Howson JM, Heward JM, Cordell HJ, Foxall H, Carr-Smith J, Gibson SM, Walker N, Tomer Y, Franklyn JA, Todd JA, Gough SC. Regression mapping of association between the human leukocyte antigen region and Graves disease. *Am J Hum Genet* 2005;76:157–63.
9 Simmonds MJ, Howson JM, Heward JM, Carr-Smith J, Franklyn JA, Todd JA, Gough SC. A novel and major association of HLA-C in Graves' disease that eclipses the classical HLA-DRB1 effect. *Hum Mol Genet* 2007;16:2149–53.
10 Nakabayashi K, Tajima A, Yamamoto K, Takahashi A, Hata K, Takashima Y, Koyanagi M, Nakaoka H, Akamizu T, Ishikawa N, Kubota S, Maeda S, Tsunoda T, Kubo M,

Kamatani N, Nakamura Y, Sasazuki T, Shirasawa S. Identification of independent risk loci for Graves' disease within the MHC in the Japanese population. *J Hum Genet* 2011;56:772–8.
11 Okada Y, Momozawa Y, Ashikawa K, Kanai M, Matsuda K, Kamatani Y, Takahashi A, Kubo M. Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. *Nat Genet* 2015;47:798–802.
12 Chen PL, Fann CS, Chu CC, Chang CC, Chang SW, Hsieh HY, Lin M, Yang WS, Chang TC. Comprehensive genotyping in two homogeneous Graves' disease samples reveals major and novel HLA association alleles. *PLoS One* 2011;6:e16635.
13 Kim K, Bang SY, Lee HS, Okada Y, Han B, Saw WY, Teo YY, Bae SC. The HLA-DRβ1 amino acid positions 11-13-26 explain the majority of SLE-MHC associations. *Nat Commun* 2014;5:5902.
14 Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, Raychaudhuri S, de Bakker PI. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One* 2013;8:e64683.
15 Pillai NE, Okada Y, Saw WY, Ong RT, Wang X, Tantoso E, Xu W, Peterson TA, Bielawny T, Ali M, Tay KY, Poh WT, Tan LW, Koo SH, Lim WY, Soong R, Wenk M, Raychaudhuri S, Little P, Plummer FA, Lee EJ, Chia KS, Luo M, De Bakker PI, Teo YY. Predicting HLA alleles from high-resolution SNP data in three Southeast Asian populations. *Hum Mol Genet* 2014;23:4443–51.
16 Okada Y, Kim K, Han B, Pillai NE, Ong RT, Saw WY, Luo M, Jiang L, Yin J, Bang SY, Lee HS, Brown MA, Bae SC, Xu H, Teo YY, de Bakker PI, Raychaudhuri S. Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. *Hum Mol Genet* 2014;23:6916–26.
17 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–9.
18 Zhu M, Dai J, Wang C, Wang Y, Qin N, Ma H, Song C, Zhai X, Yang Y, Liu J, Liu L, Li S, Liu J, Yang H, Zhu F, Shi Y, Shen H, Jin G, Zhou W, Hu Z. Fine mapping the MHC region identified four independent variants modifying susceptibility to chronic hepatitis B in Han Chinese. *Hum Mol Genet* 2016;25:1225–32.
19 Valdes AM, McWeeney S, Thomson G. HLA class II DR-DQ amino acids and insulin-dependent diabetes mellitus: application of the haplotype method. *Am J Hum Genet* 1997;60:717–28.
20 Hu X, Deutsch AJ, Lenz TL, Onengut-Gumuscu S, Han B, Chen WM, Howson JM, Todd JA, de Bakker PI, Rich SS, Raychaudhuri S. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat Genet* 2015;47:898–905.
21 Dong RP, Kimura A, Okubo R, Shinagawa H, Tamai H, Nishimura Y, Sasazuki T. HLA-A and DPB1 loci confer susceptibility to Graves' disease. *Hum Immunol* 1992;35:165–72.
22 Onuma H, Ota M, Sugenoya A, Inoko H. Association of HLA-DPB1*0501 with early-onset Graves' disease in Japanese. *Hum Immunol* 1994;39:195–201.
23 González-Galarza FF, Takeshita LY, Santos EJ, Kempson F, Maia MH, da Silva AL, Teles e Silva AL, Ghattaoraya GS, Alfirevic A, Jones AR, Middleton D. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Res* 2015;43:D784–8.
24 Kamatani Y, Wattanapokayakit S, Ochi H, Kawaguchi T, Takahashi A, Hosono N, Kubo M, Tsunoda T, Kamatani N, Kumada H, Puseenam A, Sura T, Daigo Y, Chayama K, Chantratita W, Nakamura Y, Matsuda K. A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. *Nat Genet* 2009;41:591–5.
25 Mbarek H, Ochi H, Urabe Y, Kumar V, Kubo M, Hosono N, Takahashi A, Kamatani Y, Miki D, Abe H, Tsunoda T, Kamatani N, Chayama K, Nakamura Y, Matsuda K. A genome-wide association study of chronic hepatitis B identified novel risk locus in a Japanese population. *Hum Mol Genet* 2011;20:3884–92.
26 Li Y, Yao Y, Yang M, Shi L, Li X, Yang Y, Zhang Y, Xiao C. Association between HLA-B*46 allele and Graves disease in Asian populations: a meta-analysis. *Int J Med Sci* 2013;10():164–70.
27 Doebele RC, Pashine A, Liu W, Zaller DM, Belmares M, Busch R, Mellins ED. Point mutations in or near the antigen-binding groove of HLA-DR3 implicate class II-associated invariant chain peptide affinity as a constraint on MHC class II polymorphism. *J Immunol* 2003;170:4683–92.
28 Inaba H, De Groot LJ, Akamizu T. Thyrotropin Receptor Epitope and Human Leukocyte Antigen in Graves' Disease. *Front Endocrinol* 2016;7.
29 Inaba H, Martin W, De Groot AS, Qin S, De Groot LJ. Thyrotropin receptor epitopes and their relation to histocompatibility leukocyte antigen-DR molecules in Graves' disease. *J Clin Endocrinol Metab* 2006;91:2286–94.