Identification of Novel T1D Risk Loci and Their Association With Age and Islet Function at Diagnosis in Autoantibody-Positive T1D Individuals: Based on a Two-Stage Genome-Wide Association Study

Diabetes Care 2019;42:1414–1421 | https://doi.org/10.2337/dc18-2023

Meng Zhu,[1,2,3] Kuanfeng Xu,[1] Yang Chen,[1] Yong Gu,[1] Mei Zhang,[1] Feihong Luo,[4] Yu Liu,[5] Wei Gu,[6] Ji Hu,[7] Haixia Xu,[8] Zhiguo Xie,[9,10,11] Chengjun Sun,[4] Yuxiu Li,[12] Min Sun,[1] Xinyu Xu,[1] Hsiang-Ting Hsu,[1] Heng Chen,[1] Qi Fu,[1] Yun Shi,[1] Jingjing Xu,[1] Li Ji,[1] Jin Liu,[1] Lingling Bian,[1] Jing Zhu,[1] Shuang Chen,[1] Lei Xiao,[1] Xin Li,[1] Hemin Jiang,[1] Min Shen,[1] Qianwen Huang,[8] Chen Fang,[7] Xia Li,[9,10,11] Gan Huang,[9,10,11] Jingyi Fan,[2] Zhu Jiang,[2] Yue Jiang,[2] Juncheng Dai,[2] Hongxia Ma,[2] Shuai Zheng,[1] Yun Cai,[1] Hao Dai,[1] Xuqin Zheng,[1] Hongwen Zhou,[1] Shining Ni,[6] Guangfu Jin,[2,3] Jin-Xiong She,[13] Liping Yu,[14] Constantin Polychronakos,[15] Zhibin Hu,[2,3] Zhiguang Zhou,[9,10,11] Jianping Weng,[8] Hongbing Shen,[2,3] and Tao Yang[1,16]

OBJECTIVE

Type 1 diabetes (T1D) is a highly heritable disease with much lower incidence but more adult-onset cases in the Chinese population. Although genome-wide association studies (GWAS) have identified >60 T1D loci in Caucasians, less is known in Asians.

RESEARCH DESIGN AND METHODS

We performed the first two-stage GWAS of T1D using 2,596 autoantibody-positive T1D case subjects and 5,082 control subjects in a Chinese Han population and evaluated the associations between the identified T1D risk loci and age and fasting C-peptide levels at T1D diagnosis.

RESULTS

We observed a high genetic correlation between children/adolescents and adult T1D case subjects ($r_g = 0.87$), as well as subgroups of autoantibody status ($r_g \geq 0.90$). We identified four T1D risk loci reaching genome-wide significance in the Chinese Han population, including two novel loci, rs4320356 near *BTN3A1* (odds ratio [OR] 1.26, $P = 2.70 \times 10^{-8}$) and rs3802604 in *GATA3* (OR 1.24, $P = 2.06 \times 10^{-8}$), and two previously reported loci, rs1770 in MHC (OR 4.28, $P = 2.25 \times 10^{-232}$) and rs705699 in *SUOX* (OR 1.46, $P = 7.48 \times 10^{-20}$). Further fine mapping in the MHC region revealed five independent variants, including another novel locus, HLA-C position 275 (omnibus $P = 9.78 \times 10^{-12}$), specific to the Chinese population. Based on the identified eight variants, we achieved an area under the curve value of 0.86 (95% CI 0.85–0.88). By building a genetic risk score (GRS) with these variants, we observed that the higher GRS were associated with an earlier age of T1D diagnosis ($P = 9.08 \times 10^{-11}$) and lower fasting C-peptide levels ($P = 7.19 \times 10^{-3}$) in individuals newly diagnosed with T1D.

CONCLUSIONS

Our results extend current knowledge on genetic contributions to T1D risk. Further investigations in different populations are needed for genetic heterogeneity and subsequent precision medicine.

[1]Department of Endocrinology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China
[2]State Key Laboratory of Reproductive Medicine, Center for Global Health, Nanjing Medical University, Nanjing, China
[3]Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, China
[4]Department of Pediatric Endocrinology and Inherited Metabolic Diseases, Children's Hospital of Fudan University, Shanghai, China
[5]Department of Endocrinology and Metabolism, Sir Run Run Hospital, Nanjing Medical University, Nanjing, China
[6]Department of Endocrinology, Children's Hospital of Nanjing Medical University, Nanjing, China
[7]Department of Endocrinology and Metabolism, The Second Affiliated Hospital of Soochow University, Suzhou, China
[8]Department of Endocrinology and Metabolism, Guangdong Provincial Key Laboratory of Diabetology, Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China
[9]Department of Metabolism and Endocrinology, The Second Xiangya Hospital, Central South University, Changsha, China
[10]National Clinical Research Center for Metabolic Diseases, Changsha, China
[11]Key Laboratory of Diabetes Immunology, Central South University, Ministry of Education, Changsha, China
[12]Department of Endocrinology, Key Laboratory of Endocrinology, Ministry of Health, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China

Type 1 diabetes (T1D) is a chronic autoimmune disease characterized by T cell–mediated destruction of pancreatic β-cells. The incidence of T1D varies remarkably among ethnicities, ranging from <5 per 100,000 children in Asian countries (such as China, Japan, and Korea) to >25 per 100,000 children in European and North American countries (such as Finland and Canada) (1). Although the cause of this variation is not completely understood, both genetic susceptibility and environmental triggers likely contribute to disease incidence (2,3).

Polymorphisms in the genomic region 6p21, which encodes HLA class II genes (primarily *HLA-DRB1-DQA1-DQB1*), have long been recognized as a major T1D risk factor (4,5). Further fine mapping in this region also identified several amino acid positions in HLA-DQ and HLA-DR that increase T1D risk independently (6). To date, genome-wide association studies (GWAS) of T1D from Caucasian populations have identified ~60 T1D non-MHC susceptibility loci (2,7–13). Although these results provide an overview of the genetic basis of T1D, extrapolation to other populations must take into account the differences in risk allele frequency, linkage disequilibrium (LD) patterns, and phenotypic prevalence as well as effect size in different ethnic groups. Genetic studies in non-Caucasian populations are necessary to fully uncover T1D risk loci and generate medical advances addressing T1D in all populations.

Insulin dependence and early diagnosis in children or adolescents are two common criteria that have been used to define T1D cases in previous GWAS (7,8). However, a large proportion of new-onset T1D in China has occurred among adults (14). The correlations of the genetic basis among children, adolescents, and adults with T1D in Chinese individuals are still unclear. As pancreatic islet autoantibodies have been widely used to clinically distinguish T1D from type 2 diabetes, here we performed a GWAS of T1D in a Chinese Han population with autoantibody-positive T1D case subjects of all ages of diagnosis.

RESEARCH DESIGN AND METHODS

Participants

We performed a two-stage case-control analysis (Supplementary Fig. 1) of T1D in the Chinese Han population. The GWAS scan included 1,045 T1D case and 1,308 control subjects recruited from The First Affiliated Hospital of Nanjing Medical University and the Children's Hospital of Nanjing Medical University between January 2008 and December 2014. The validation stage included 1,553 case and 3,774 control subjects (Supplementary Table 1), who were recruited from southeast, central, and south China between January 2015 and December 2016.

Only patients with diabetes with insulin dependence within 6 months after diagnosis and the presence of at least one positive autoantibody (15) were included as T1D case subjects (regardless of age at diagnosis). The control subjects were outpatients without diabetes and family history of diabetes from the same geographic areas, and some of the control subjects were used in our previous study (16). All samples were collected with appropriate informed consent from all participants or their guardians. The study was approved by the ethics committee of The First Affiliated Hospital of Nanjing Medical University and conducted according to the principles of the Declaration of Helsinki.

Genotyping and Quality Control

The GWAS scan was conducted using the Illumina Human OmniZhongHua-8 platform, which provides exceptional coverage of common, intermediate, and rare variations specific to Chinese populations. Systematic quality control was performed before the association analysis (Supplementary Fig. 1), including sample exclusions for ambiguous sex, call rate <95%, extreme heterozygosity rate (out of the mean ± 6 SD), any duplicate or related individuals (PI_HAT > 0.25), single nucleotide polymorphism (SNP) exclusions for monomorphic SNPs, SNPs with minor allele frequency (MAF) <0.01, SNPs with a missingness rate >5%, SNPs not mapped on autosomal chromosomes (those on chromosome X were retained), SNPs that deviated from Hardy-Weinberg equilibrium ($P < 1 \times 10^{-4}$), and SNPs with significantly different missing rates between case and control subjects ($P < 1 \times 10^{-5}$). After exclusion of unqualified samples and SNPs, a total of 787,224 qualified SNPs in 1,005 T1D case and 1,257 control subjects remained in the association analysis. Principal component analysis showed that the case and control subjects were genetically matched (Supplementary Fig. 2). A small genomic-control inflation factor (λ) of 1.05 indicated minimal population stratification among the subjects (Supplementary Fig. 3).

Genotyping in the validation stage (13 SNPs) was mainly performed using the iPLEX Sequenom MassARRAY platform; the other four SNPs, which failed due to primer design in the first round, were genotyped using TaqMan assays (Applied Biosystems).

Imputation and Association Analysis

Different imputation strategies were used for the non-MHC and MHC regions. Briefly, imputation of the non-MHC region was conducted using IMPUTE2 software (V.2.2.2), which can automatically find haplotypes from the best matching population of the entire population in the

[13]*Center for Biotechnology and Genomic Medicine, Medical College of Georgia, Augusta University, Augusta, GA*
[14]*Barbara Davis Center for Childhood Diabetes, University of Colorado Denver, Aurora, CO*
[15]*The Endocrine Genetics Laboratory, Child Health and Human Development Program and Department of Pediatrics, McGill University Health Centre Research Institute, Montreal, Canada*
[16]*Key Laboratory of Human Functional Genomics of Jiangsu Province, Nanjing Medical University, Nanjing, China*

Corresponding author: Tao Yang, yangt@njmu.edu.cn, Hongbing Shen, hbshen@njmu.edu.cn, Jianping Weng, wjianp@mail.sysu.edu.cn, Zhiguang Zhou, zhouzhiguang@csu.edu.cn, or Zhibin Hu, zhibin_hu@njmu.edu.cn

Received 24 September 2018 and accepted 3 May 2019

This article contains Supplementary Data online at http://care.diabetesjournals.org/lookup/suppl/doi:10.2337/dc18-2023/-/DC1.

M. Zhu, K.X., Y.Ch., Y.G., M.Zha., F.L., Y.L., and W.G. contributed equally to this work and share co–first authorship.

1000 Genomes Project Phase 3. A prephasing strategy with SHAPEIT software (V.2) was adopted to improve imputation performance. Imputation of the MHC region was performed with SNP2HLA with default parameters, using a reference panel of Chinese Han provided by BGI (17), which included 10,689 healthy individuals of Chinese Han ancestry with classical typing for *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1*, and *HLA-DPB1*.

Variants that were poorly imputed (INFO score <0.3), had low frequency (MAF < 0.01), or deviated from Hardy-Weinberg equilibrium ($P < 1 \times 10^{-4}$) were excluded from the association analysis. The association of uncertain genotypes from IMPUTE2 was performed with SNPTEST (V2.5.2), and the association of the dosage file from SNP2HLA was tested with PLINK (version 1.9) as recommended with adjustments for significant eigenvectors and sex.

Genetic Correlation Analysis

Genetic correlations across different subgroups were calculated with LD score regression (LDSC) software. All genotyped SNPs that passed quality control were used for LDSC analysis with default parameters. The LD scores were computed using East Asians from 1000 Genomes. Due to the complexity of the LD structure in the MHC region, an R script named BUHMBOX was used to further validate the results in the current study (18). BUHMBOX evaluated the genetic similarity based on genetic risk scores (GRS), which were derived from independent top variants and less affected by LD structure. Variants with $P < 1.0 \times 10^{-4}$ were used in BUHMBOX with recommended parameters, and only the top variants were kept if multiple variants were in LD ($r^2 < 0.10$).

Fine Mapping Analysis of MHC

Stepwise conditional analysis was performed to explore independently effective loci in the MHC region in a forward stepwise manner. In each step, the most significant HLA variants were considered as covariates in the next logistic model until no variants satisfied the significance threshold of $P < 5.00 \times 10^{-8}$. To identify independent HLA associations outside of the *HLA-DRB1-DQA1-DQB1* locus, all four-digit classical alleles of *HLA-DRB1-DQA1-DQB1* were adjusted.

In Silico Bioinformatics Analysis

To define potentially causal variants for each locus, we calculated posterior probabilities (PPs) using the approximate Bayesian factor under the assumption of a single causal variant per locus, which has previously been described (19). Briefly, the association effect sizes were assumed to follow a distribution of N (0, V) under H0, with V being the squared SE. A distribution following N (0, V + W) was assumed under H1, where W is (ln[1.5]/1.96), reflecting the PP of observing an odds ratio (OR) of 1.5. For an observed effect size β, the approximate Bayesian factor was calculated as the ratio of P (β|H0)/P (β|HA). Following calculation of the PP, we created credible sets by sorting associations descending on the basis of their PP and including associations such that the sum of PP was >0.90. Conditional analyses in logistic regression were performed to determine the presence of multiple independent effects.

The most likely variants for each locus were then assessed for overlap with expression quantitative trait loci (eQTLs), DNAse-I hypersensitive sites, enhancers, promoters, and histone peaks. We used eQTLs found in GTEx (https://gtexportal.org/home/); a study assessing eQTLs in CD4 and CD8 T cells, B cells, natural killer cells, and monocytes in East Asians (20); a sequencing-based eQTL meta-analysis of 2,116 whole blood samples (21); and a study of splicing QTLs (22). Only the variant in high LD ($r^2 > 0.8$) with the top QTL for a given gene was considered as a QTL overlapping.

For comparison of the genetic basis of T1D between Caucasian and Chinese populations, all reported T1D loci from Caucasian GWAS were collected from the GWAS Catalog (https://www.ebi.ac.uk/gwas/). For variants without effect estimations in the GWAS Catalog or the original publications, the effects were obtained from the Wellcome Trust Case Control Consortium (WTCCC) and dbGaP (phs000180.v3.p2).

Utility and Effectiveness of GRS

To evaluate the effectiveness of our findings, we derived a GRS in three schemes: *1*) including the five independent variants of the MHC region (model A), *2*) including all the eight identified variants in the Chinese population (model B), and *3*) including both the 8 variants identified in the Chinese population and the other 14 risk loci reported in Caucasian populations that were consistently significant in our GWAS (model C).

In each scheme, the GRS was generated by multiplying the number of risk alleles for each variant by its respective weight (lnOR) and then summing all the variants together. Effect sizes for all variants were derived from the association with T1D of Chinese descent in the current study, and all ORs were calculated with the nonrisk allele as a reference.

Laboratory Measurements for Autoantibodies and C-Peptide in T1D Patients

Pancreatic islet autoantibodies were measured separately by protein A radiobinding assays (23) using [35s]methionine-labeled in vitro–transcribed/translated recombinant human GAD65, insulinoma-2–associated autoantibodies, and the carboxy-terminal portion of ZnT8 for the two major variants at amino acid 325. In addition, 300 patients newly diagnosed with T1D were assayed for fasting C-peptide levels by an automatic electrochemical luminescence instrument (Roche, Basel, Switzerland).

Statistical Analyses

The significant eigenvectors of ancestry between case and control subjects (principal components 1, 3, 4, and 6) and sex were included as covariates in the logistic regression model when estimating the ORs and 95% CIs. The results from different stages were combined with a fixed effects model in the meta-analysis. Subgroup analysis was performed according to age at diagnosis and autoantibody positivity. Assessment of the risk prediction model was performed with the R package "PredictABEL." The associations between GRS and age at T1D diagnosis were evaluated using Cox regression, and the associations between GRS and fasting C-peptide levels were evaluated with linear regression with adjustments for sex and significant eigenvectors. Heterogeneity of the associations was evaluated using Cochran's Q test, taking the log-ORs and SE estimated from the different subgroups. Cox regression was also used to evaluate the associations between genome-wide variants and age at T1D diagnosis directly in T1D patients

Table 1—Different T1D subgroups show high correlations genetically

			LDSC		BUHMBOX			
	Subgroup A	Subgroup B	r_g[a]	P[a]	Number of SNPs	GRS, P value[b]	GRS, β (95% CI)[b]	BUHMBOX, P value[b]
Age at diagnosis (years)[c]	≤17	>17	0.87	1.54×10^{-26}	106	2.70×10^{-130}	0.31 (0.28–0.34)	3.80×10^{-9}
Number of autoantibodies	Single	Multiple	0.90	1.01×10^{-20}	121	1.40×10^{-130}	0.22 (0.20–0.24)	7.29×10^{-48}
GADA	Negative	Positive	0.91	2.72×10^{-29}	134	1.21×10^{-56}	0.17 (0.15–0.19)	1.18×10^{-35}
IA-2A	Negative	Positive	0.98	3.29×10^{-42}	117	2.87×10^{-91}	0.17 (0.15–0.19)	4.89×10^{-60}
ZnT8A	Negative	Positive	0.92	1.71×10^{-4}	108	1.89×10^{-109}	0.21 (0.19–0.24)	1.12×10^{-24}

IA-2A, insulinoma-2–associated autoantibodies. [a]Estimated genetic correlation based on all genotyped SNPs with LDSC. [b]Estimated sharing of the genetic basis of different subgroups with Mendelian randomization and BUHMBOX using the independent significant SNPs of subgroup B ($P < 1 \times 10^{-4}$). A significant GRS P value indicates evidence of shared genetic structure; significant BUHMBOX P value indicates evidence of subgroup heterogeneity. [c]T1D patients were divided into early onset (≤17 years) and late onset (>17 years) subgroups based on median of age at diagnosis in case subjects.

with adjustments for eigenvectors and sex.

RESULTS

Different T1D Subgroups Share a Similar Genetic Basis in the Chinese Population

We first calculated the genetic associations according to age at diagnosis and autoantibody status (Supplementary Figs. 4 and 5) and then evaluated the genetic correlations across different subgroups. LD score regression analysis showed a highly correlated genetic basis in patients with early- and late-onset T1D ($r_g = 0.87$), even among children, adolescents, and adult T1D patients (Supplementary Table 2). Subgroups of different autoantibody status or the number of positive antibodies also showed similar correlations ($r_g \geq 0.9$) (Table 1 and Supplementary Fig. 6). The results from BUHMBOX further validated the genetic correlations among these different subgroups (Table 1).

Two Novel Risk Loci, rs4320356 near *BTN3A1* and rs3802604 in *GATA3*, Are Associated With T1D Risk

The high genetic correlation of different T1D subgroups indicated that these cases can be combined for analysis. Association analysis revealed 17 distinct genomic regions with P values $<1.00 \times 10^{-5}$ in

Table 2—Identification of four loci associated with T1D risk in the Chinese population

			Genotype distribution[b]		MAF					
Locus	Gene neighborhood	Stage	Case subjects	Control subjects	Case subjects	Control subjects	OR_{add} (95% CI)[c]	P[c]	P_{het}	I^2
6p22.2 rs4320356	*BTN3A1*	Discovery	112/458/435	81/512/664	0.34	0.27	1.40 (1.22–1.60)	1.72×10^{-6}		
C/T[a]		Replication[d]	146/668/729	272/1,569/1,899	0.31	0.28	1.19 (1.08–1.31)	6.74×10^{-4}	0.358	0.03
		Combined					**1.26 (1.16–1.36)**	$\mathbf{2.70 \times 10^{-8}}$	**0.131**	**0.47**
10p14 rs3802604	*GATA3*	Discovery	182/475/348	157/574/524	0.42	0.35	1.35 (1.18–1.53)	4.78×10^{-6}		
T/C[a]		Replication[d]	266/739/522	545/1,733/1,457	0.42	0.38	1.18 (1.08–1.30)	2.75×10^{-4}	0.080	0.60
		Combined					**1.24 (1.15–1.33)**	$\mathbf{2.06 \times 10^{-8}}$	**0.070**	**0.57**
6p21.3 rs1770	*HLA-DQB1*	Discovery	667/278/58	290/594/373	0.80	0.47	4.35 (3.85–5.00)	8.95×10^{-83}		
A/G[a]		Replication[d]	990/424/108	819/1,752/1,147	0.79	0.46	4.22 (3.79–4.72)	1.42×10^{-147}	0.066	0.63
		Combined					**4.28 (3.92–4.68)**	$\mathbf{2.25 \times 10^{-232}}$	**0.127**	**0.47**
12q13.2 rs705699	*SUOX*	Discovery	113/441/451	67/478/712	0.33	0.24	1.56 (1.36–1.80)	2.88×10^{-10}		
C/T[a]		Replication[d]	160/664/696	244/1,435/2,032	0.32	0.26	1.41 (1.27–1.55)	2.02×10^{-11}	0.304	0.16
		Combined					**1.46 (1.34–1.58)**	$\mathbf{7.48 \times 10^{-20}}$	**0.082**	**0.55**

Combined, derived from discovery stage and replication stage using meta-analysis under the assumption of a fixed model (combined results appear in boldface type); P_{het}, P value of heterogeneity test. [a]Major/minor alleles. [b]Individuals homozygous for the minor allele/heterozygous/homozygous for the major allele in control group. [c]OR_{add} (95% CI) and P_{add} values were derived from logistic regression analysis with adjustment for sex and the significant eigenvectors (for the discovery and replication stage) under the assumption of an additive genetic model. [d]Derived from three centers in China using meta-analysis under the assumption of a fixed model.

the discovery stage (Supplementary Fig. 7). Promising variants from each region were then genotyped with an additional 1,553 autoantibody-positive T1D case and 3,774 nondiabetic control subjects of Chinese Han ethnicity. Four of the tested variants showed significant associations in the same direction as that observed in the GWAS (Supplementary Table 3). After combining the results from different stages using meta-analysis, we found that these four variants were significantly associated with T1D risk at genome-wide significance ($P < 5.00 \times 10^{-8}$), including rs4320356 at 6p22.2 (OR 1.26, $P = 2.70 \times 10^{-8}$), rs1770 at MHC (OR 4.28, $P = 2.25 \times 10^{-232}$), rs3802604 at 10p14 (OR 1.24, $P = 2.06 \times 10^{-8}$), and rs705699 at 12q13.2 (OR 1.46, $P = 7.48 \times 10^{-20}$) (Table 2 and Supplementary Fig. 8).

Bayesian Fine Mapping Reveals Candidate Causal Variants

We constructed credible sets that were 90% likely to contain the potentially causal SNP based on PP (Supplementary Table 4). Of the identified four loci, rs4320356 at 6p22.2 (PP = 0.14), rs9273471 at MHC (PP = 0.43), rs10905277 at 10p14 (PP = 0.21), and rs773125 at 12q13.2 (PP = 0.27) were the probable causal variants. Further annotation revealed that rs4320356 and rs10905277 were involved in the regulation of T cells and associated with the expression of *BTN3A1* and *GATA3* in whole blood, respectively (Supplementary Tables 5 and 6), while rs9273471 ($r^2 = 0.99$ with rs1770) was probably related to primary B cells and natural killer cells and associated with the expression of *HLA-DQA2* in monocytes (Supplementary Tables 5 and 6). However, rs773125 at 12q13.2 was mainly overlapped with functional elements in the digestive system and associated with the expression of *SUOX* in these tissues, including the pancreas (Supplementary Table 7).

The Genetic Basis of T1D Risk Loci Has Both Similarities and Differences Between Caucasian and Chinese Populations

Of the four loci reaching genome-wide significance of our study, MHC and 12q13.2 have been linked to T1D risk previously reported in Caucasian populations (6,7); additionally, 6p22.2 and 10p14 were also significantly associated with increased risks of T1D in Caucasians according to the WTCCC ($P = 6.43 \times 10^{-6}$ and 5.05×10^{-4}, respectively) (Supplementary Table 8). Furthermore, we also observed 14 other reported T1D risk loci of Caucasians that were consistently associated with T1D risk in the same direction at $P < 0.05$ in our GWAS (Fig. 1 and Supplementary Table 9).

Even with shared susceptibility loci, the effect sizes were stronger for variants at 6p22.2, 10p14, and 12q13.2 in the Chinese population, and distinct differences in allele frequency were also observed, such as *INS* at 11p15.5 (MAF 0.04 in Chinese vs. 0.28 in Caucasian) (Fig. 1). Moreover, 13 of the 61 reported loci were nonpolymorphic or had very low MAF (<0.01) in the Chinese population, including *PTPN22* at 1p13.2. However, the other 32 reported T1D risk loci from Caucasian GWAS were not replicated in our study, which was probably due to

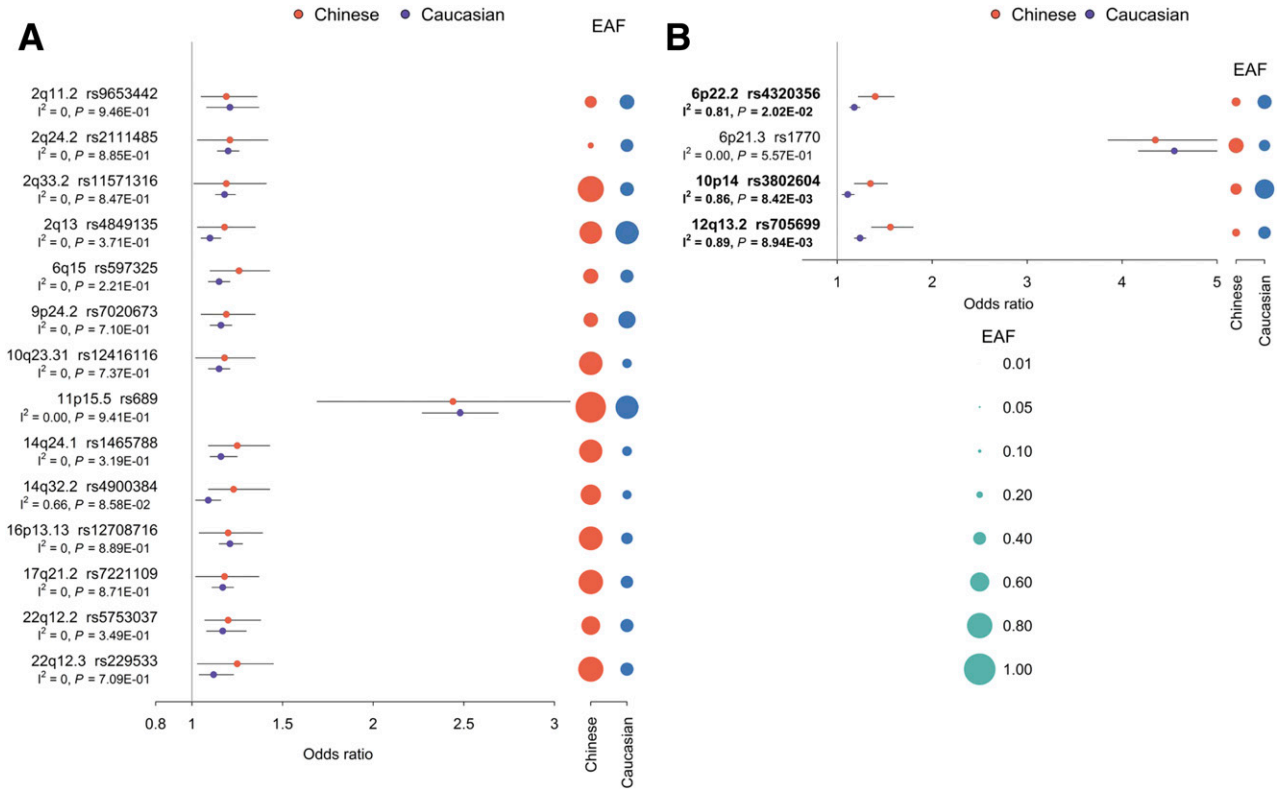

Figure 1—Frequencies and effect sizes of reported variants showed both similarities and differences between Caucasian and Chinese Han populations. *A*: Comparison of frequencies and effect sizes for the 14 validated variants between Caucasians and Chinese Han. *B*: Comparison for the four identified variants in the current study between Caucasians and Chinese Han. The effect sizes of Caucasians were collected from the GWAS Catalog (https://www.ebi.ac.uk/gwas/). For variants without CIs in GWAS Catalog and the original publications, the effects were obtained from the WTCCC. All the effect sizes and frequencies were calculated with the nonrisk allele as reference.

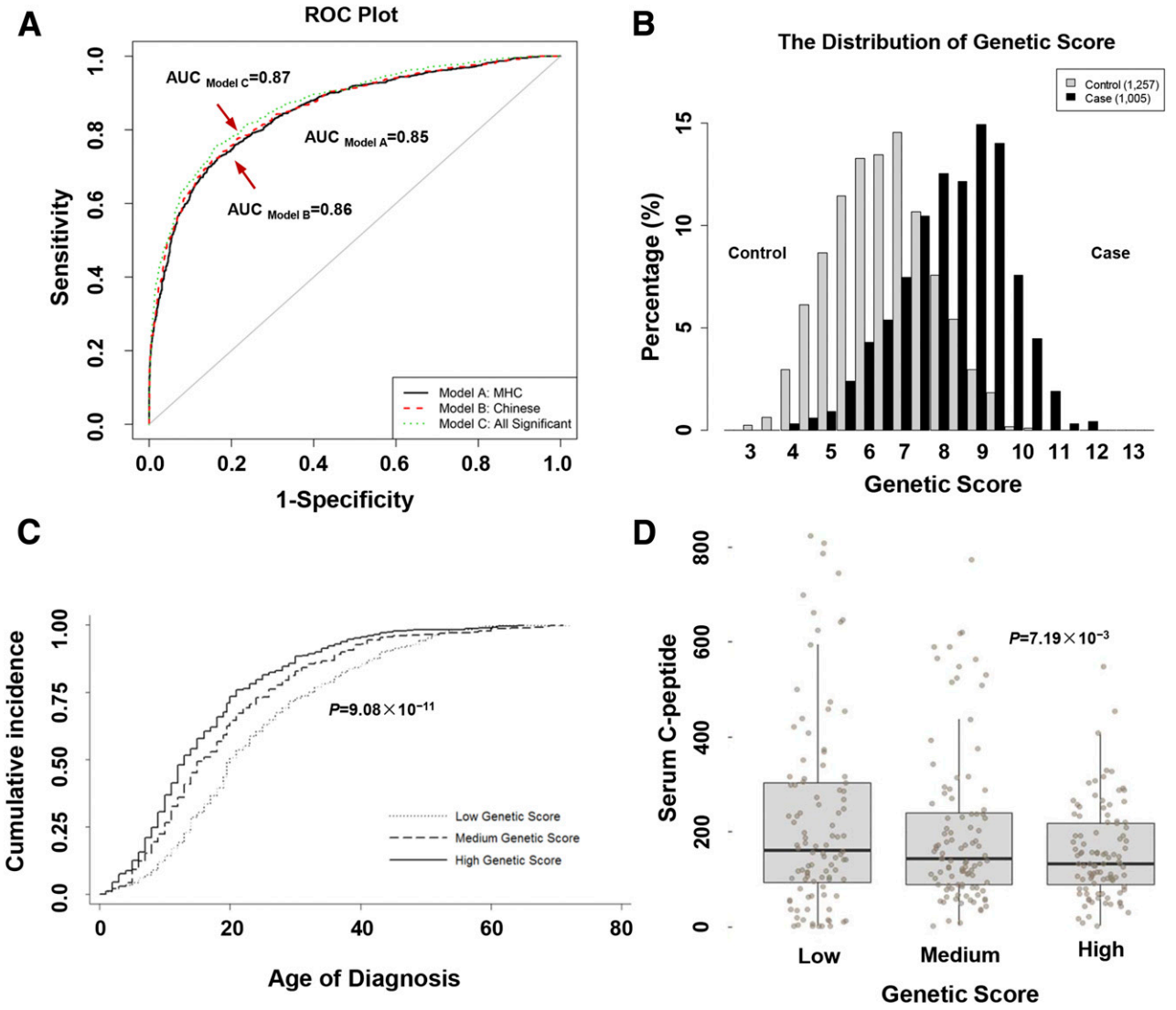

Figure 2—The identified variants showed a high prediction value of T1D risk and were associated with clinical features. *A*: The five independent variants of the MHC region (model A), the identified eight variants in the Chinese population (model B), and additional validated variants from the Caucasian population (model C). *B*: The distribution of GRS in T1D case subjects and control subjects without diabetes based on all the identified eight variants in the Chinese population (model C). *C*: The GRS were significantly associated with age of onset in T1D case subjects, and individuals carrying low GRS were older at onset (23.82 ± 13.60 years) than those carrying medium GRS (19.43 ± 13.25 years) and high GRS (16.54 ± 11.83 years). *D*: Individuals with low GRS showed significantly higher (162.30 [94.05–304.25]) fasting C-peptide levels at diagnosis than those with medium GRS (144.20 [89.43–240.85]) and high GRS (133.50 [89.78–218.20]). ROC, receiver operating characteristic.

the limited sample size of our study (Supplementary Table 9).

Fine Mapping of the MHC Region Identifies HLA-C 275 as a Novel Locus Specific to the Chinese Population

Inconsistent with prior results, rs1770 was the leading risk variant in our study (OR 4.35, $P = 8.95 \times 10^{-83}$) rather than alanine at HLA-DQβ1 position 57 (OR 2.44, $P = 2.46 \times 10^{-34}$) in Caucasian populations (Supplementary Fig. 9). Although rs1770 does not directly change the amino acid sequence of HLA molecules, it might modulate HLA class II expression pattern in humans (Supplementary Fig. 10).

Other than rs1770, HLA-DRβ1 position 74 (omnibus $P = 6.38 \times 10^{-24}$) and HLA-DRβ1 position 11 (omnibus $P = 8.47 \times 10^{-12}$) were also independently associated with T1D risk in the *HLA-DRB1-DQA1-DQB1* locus (Supplementary Table 10 and Supplementary Fig. 11). After adjustment for the four classical alleles in the *HLA-DRB1-DQA1-DQB1* locus, HLA-A position 9 (omnibus $P = 2.47 \times 10^{-11}$) and HLA-C position 275 (omnibus $P = 9.78 \times 10^{-12}$) were also independently linked to T1D risk in our study (Supplementary Table 11 and Supplementary Fig. 11).

Further analysis showed that the amino acid residues in HLA-DRβ1 and HLA-A identified in our study were in medium-high LD with those from Caucasian populations (Supplementary Table 11). Interestingly, the HLA-C position 275 was independent of any reported loci in Caucasian populations (6).

The Identified Variants Show a High Prediction Value of T1D Risk and Are Significantly Associated With Clinical Features

Building a multivariable logistic regression model that included the eigenvectors, sex, and the five independent MHC variants yielded an area under the curve (AUC) of 0.85 (0.84–0.87) (Fig. 2*A*).

Higher discriminations were achieved when additional non-MHC variants were added to the risk prediction model, with a predicted AUC of 0.86 (0.85–0.88). If we also included the other 14 validated variants, which were reported in Caucasian populations, the AUC improved to 0.87 (0.86–0.89). Based on the five independent MHC and three non-MHC loci identified in our study, an evident shift was observed for the GRS in T1D case subjects (Fig. 2*B*). Furthermore, higher GRS were associated with an earlier age at T1D diagnosis ($P = 9.08 \times 10^{-11}$) (Fig. 2*C*) and lower fasting C-peptide levels at T1D diagnosis ($P = 7.19 \times 10^{-3}$) (Fig. 2*D*).

MHC Showed Stronger Effect Sizes in Patients With Early-Onset T1D

We also found that the prediction value was more significant in early-onset T1D case subjects than in late-onset case subjects (Supplementary Fig. 12). Although the associations in MHC were consistently significant in both early-onset and late-onset subgroups, significant heterogeneity of effects was observed for rs1770, HLA-DRB1*74, and HLA-DRB1*11 across different age-groups ($I^2 \geq 75\%$ and $P \leq 0.05$), with almost double effects in the early-onset subgroup (Supplementary Table 12). When we tested for the association with age at T1D diagnosis directly with Cox regression, the MHC region was also the most significant locus, with rs1770 showing the most significant variant (hazard ratio 1.42 [95% CI 1.27–1.58], $P = 1.53 \times 10^{-10}$) (Supplementary Fig. 13).

CONCLUSIONS

The proportion of T1D patients diagnosed at ≥20 years of age was 65.3% in the Chinese population according to the National Registration System of China (14). Therefore, we performed the first GWAS of T1D in a Chinese population with autoantibody-positive T1D case subjects of all ages of diagnosis. We observed high genetic correlations across different T1D subgroups and identified two novel loci at 6p22.2 and 10p14, as well as the novel variant HLA-C 275 in MHC. Even though there were many shared loci between Chinese and Caucasian populations, heterogeneity of allele frequency and effects was observed across these two populations. Importantly, the GWAS-derived GRS significantly predicted T1D risk and was associated with age and fasting C-peptide levels at T1D diagnosis.

Of the two novel loci, rs4320356 is located 2 Mb away from the classical MHC region and marks an 83 kb haplotype block that contains *BTN3A1*. This marker is also a probable causal variant with an eQTL effect for the majority of the BTN family genes, including *BTN3A1*, in different immune cell subsets. Studies have indicated that BTN3A1 is involved in the stimulation of human γδ T cells, which are essential effectors of T1D (24–26). Second, rs3802604 is located in an 18 kb haplotype containing a single protein coding gene, *GATA3*, which has conventionally been regarded as a transcription factor that drives the differentiation of T helper cells (27–29). In addition, the haplotype has also been linked to rheumatoid arthritis in both Caucasians and Asians (30,31).

Interestingly, other than many similarities, substantial differences were also observed in comparison of susceptibility loci across Chinese and Caucasian populations. The most salient feature is that approximately one-fifth of the risk loci reported in Caucasians were nonpolymorphic or had a very low frequency, which might lead to a lower genetic risk load in the Chinese population and might be the reason for the lower T1D incidence in China (1). In addition, higher effects were observed for 6p22.2 and 10p14 in the Chinese population, and the HLA-C position 275 was identified only in the Chinese population, most likely because all T1D case subjects in our study were autoantibody positive and these loci are related to T-cell function (24,25,27,28,32). However, the heterogeneity of the population may also be another possible explanation.

T1D GRS combining MHC and non-MHC alleles have been developed and validated to improve T1D prediction in various settings from Caucasian populations (33,34). We achieved, for the first time, a similar prediction value in a Chinese population with five independent MHC and three non-MHC loci. Our study provided immediate cues for which variants should be included and the exact effect sizes for the T1D prediction model of the Chinese population. In addition, we also found that higher GRS was also an indicator of earlier age and reduced islet function at diagnosis in autoantibody-positive T1D patients. Considering that the majority of variants involved in the construction of GRS were related to the immune response (24, 26–29), a possible explanation is that the derived GRS was related to the severity of the autoimmunity of T1D in patients at diagnosis. However, the exact mechanisms remain to be explored in future studies.

In conclusion, the risk loci identified in the current study provided additional insights into the genetic basis of T1D risk, which would further advance our understanding of T1D susceptibility in the Chinese population. More importantly, we validated in the current study that GRS could be effectively used for T1D risk prediction in the Chinese population, potentially leading to practically feasible clinical screening for individualized intervention in the future.

Acknowledgments. The authors gratefully acknowledge the WTCCC group for generously allowing the use of their genotype data. The authors thank all the study participants, research staff, and students who participated in this work.

Funding. This work was supported in part by the National Key Project of Research and Development Plan (2016YFC1000204, 2016YFC1305302, 2016YFC1305000, and 2016YFC1305001), the National Key Technologies R&D Program of China (2015BAI12B13), the Innovation-Driven Program of Central South University (2015CX009), the Cheung Kong Scholars Program of China, the Key Program of National Natural Science Foundation of China (81830023, 81530026, and 81530025), the National Natural Science Foundation of China (81390543, 81270897, 81670715, 81670756, 81370939, 81300668, 81400808, 81400813, and 81803306), the Jiangsu Specially Appointed Professor project, the Science and Technology Innovation Team of Jiangsu Province Qinglan Project, the Priority Academic Program for the Development of Jiangsu Higher Education Institutions (Public Health and Preventive Medicine), the Top-Notch Academic Programs Project of Jiangsu Higher Education Institutions (PPZY2015A067), the Jiangsu Province Key Science and Technology Development Project (BE2017753), the Jiangsu Province Science Foundation for Youth (BK20171082 and BK20180675), the Three Big Constructions funds of Sun Yat-sen University (82000-31133402), the Science and Technology Commission of Shanghai Municipality (STCSM 15411961700), and the Beijing Science and Technology project (Z151100003915077).

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. M. Zhu performed statistical analyses, drafted the initial manuscript, and revised the manuscript. K.X. participated in study design, performed overall project management, and revised the manuscript. Y.Ch. and Y.G. performed overall project management and sample processing. M.Zha., M.Su., X.X., S.Z., Y.Ca., H.D., and X.Z. were responsible for control recruitment and sample preparation in the

validation stage. F.L., J.H., C.S., and C.F. contributed to the sample recruitment from southeast China. Y. Liu, Z.X., Y. Li, Xia Li, and G.H. were responsible for the preparation of samples from central China. W.G., H.C., Q.F., Y.S., J.X., L.J., J.L., L.B., J.Z., S.C., L.X., Xin Li, H.J., M.Sh., and S.N. were responsible for subject recruitment and sample preparation in the GWAS stage. H.X. and Q.H. contributed to the sample recruitment and sample preparation from south China. H.T.H., Y.J., J.D., H.M., and G.J. contributed to discussion and helped edit the manuscript. J.F. and Z.J. carried out genotyping. H.Z., J.-X.S., L.Y., and C.P. reviewed and commented on various versions of the manuscript and suggested revisions. Z.H., Z.Z., J.W., H.S., and T.Y. directed the study, obtained financial support, and were responsible for study design. M. Zhu, K.X., Y.Ch., Y.G., M.Zha., F.L., Y. Liu, W.G., J.H., H.X., Z.X., C.S., Y. Li, M.Su., X.X., H.T.H., H.C., Q.F., Y.S., J.X., L.J., J.L., L.B., J.Z., S.C., L.X., Xin Li, H.J., M.Sh., G.H., C.F., Xia Li, G.H., J.F., Z.J., Y.J., J.D., H.M., S.Z., Y.Ca., H.D., X.Z., H.Z., S.N., G.J., J.-X.S., L.Y., C.P., Z.H., Z.Z., J.W., H.S., and T.Y. approved the final manuscript. T.Y. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Prior Presentation. Parts of this study were presented in abstract form at the 78th Scientific Sessions of the American Diabetes Association, Orlando, FL, 22–26 June 2018.

References

1. Katsarou A, Gudbjörnsdottir S, Rawshani A, et al. Type 1 diabetes mellitus. Nat Rev Dis Primers 2017;3:17016
2. Pociot F, Lernmark Å. Genetic risk factors for type 1 diabetes. Lancet 2016;387:2331–2339
3. Rewers M, Ludvigsson J. Environmental risk factors for type 1 diabetes. Lancet 2016;387:2340–2348
4. Nerup J, Platz P, Andersen OO, et al. HL-A antigens and diabetes mellitus. Lancet 1974;2:864–866
5. Noble JA, Valdes AM, Cook M, Klitz W, Thomson G, Erlich HA. The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. Am J Hum Genet 1996;59:1134–1148
6. Hu X, Deutsch AJ, Lenz TL, et al. 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
7. Todd JA, Walker NM, Cooper JD, et al.; Genetics of Type 1 Diabetes in Finland; Wellcome Trust Case Control Consortium. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 2007;39:857–864
8. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–678
9. Barrett JC, Clayton DG, Concannon P, et al.; Type 1 Diabetes Genetics Consortium. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 2009;41:703–707
10. Bradfield JP, Qu HQ, Wang K, et al. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. PLoS Genet 2011;7:e1002293
11. Evangelou M, Smyth DJ, Fortune MD, et al. A method for gene-based pathway analysis using genomewide association study summary statistics reveals nine new type 1 diabetes associations. Genet Epidemiol 2014;38:661–670
12. Onengut-Gumuscu S, Chen WM, Burren O, et al.; Type 1 Diabetes Genetics Consortium. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat Genet 2015;47:381–386
13. Fortune MD, Guo H, Burren O, et al. Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls. Nat Genet 2015;47:839–846
14. Weng J, Zhou Z, Guo L, et al.; T1D China Study Group. Incidence of type 1 diabetes in China, 2010-13: population based study. BMJ 2018;360:j5295
15. Park Y, Wintergerst KA, Zhou Z. Clinical heterogeneity of type 1 diabetes (T1D) found in Asia. Diabetes Metab Res Rev 2017;33:e2907
16. Hu Z, Shi Y, Mo X, et al. A genome-wide association study identifies two risk loci for congenital heart malformations in Han Chinese populations. Nat Genet 2013;45:818–821
17. Zhou F, Cao H, Zuo X, et al. Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease. Nat Genet 2016;48:740–746
18. Han B, Pouget JG, Slowikowski K, et al.; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. Nat Genet 2016;48:803–810
19. Westra HJ, Martínez-Bonet M, Onengut-Gumuscu S, et al. Fine-mapping and functional studies highlight potential causal variants for rheumatoid arthritis and type 1 diabetes. Nat Genet 2018;50:1366–1374
20. Ishigaki K, Kochi Y, Suzuki A, et al. Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis. Nat Genet 2017;49:1120–1125
21. Zhernakova DV, Deelen P, Vermaat M, et al. Identification of context-dependent expression quantitative trait loci in whole blood. Nat Genet 2017;49:139–145
22. Zhang X, Joehanes R, Chen BH, et al. Identification of common genetic variants controlling transcript isoform variation in human whole blood. Nat Genet 2015;47:345–352
23. Orban T, Sosenko JM, Cuthbertson D, et al.; Diabetes Prevention Trial-Type 1 Study Group. Pancreatic islet autoantibodies as predictors of type 1 diabetes in the Diabetes Prevention Trial-Type 1. Diabetes Care 2009;32:2269–2274
24. Vavassori S, Kumar A, Wan GS, et al. Butyrophilin 3A1 binds phosphorylated antigens and stimulates human γδ T cells. Nat Immunol 2013;14:908–916
25. Markle JG, Mortin-Toth S, Wong AS, Geng L, Hayday A, Danska JS. γδ T cells are essential effectors of type 1 diabetes in the nonobese diabetic mouse model. J Immunol 2013;190:5392–5401
26. Aydintug MK, Zhang L, Wang C, et al. γδ T cells recognize the insulin B:9-23 peptide antigen when it is dimerized through thiol oxidation. Mol Immunol 2014;60:116–128
27. Ho IC, Tai TS, Pai SY. GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. Nat Rev Immunol 2009;9:125–135
28. Wan YY. GATA3: a master of many trades in immune regulation. Trends Immunol 2014;35:233–242
29. Gomez-Tourino I, Arif S, Eichmann M, Peakman M. T cells in type 1 diabetes: instructors, regulators and effectors: a comprehensive review. J Autoimmun 2016;66:7–16
30. Eyre S, Bowes J, Diogo D, et al.; Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate; Wellcome Trust Case Control Consortium. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet 2012;44:1336–1340
31. Okada Y, Wu D, Trynka G, et al.; RACI consortium; GARNET consortium. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2014;506:376–381
32. Blais ME, Dong T, Rowland-Jones S. HLA-C as a mediator of natural killer and T-cell activation: spectator or key player? Immunology 2011;133:1–7
33. Clayton DG. Prediction and interaction in complex disease genetics: experience in type 1 diabetes. PLoS Genet 2009;5:e1000540
34. Winkler C, Krumsiek J, Buettner F, et al. Feature ranking of type 1 diabetes susceptibility genes improves prediction of type 1 diabetes. Diabetologia 2014;57:2521–2529