HLA-DQβ1 amino acid position 87 and DQB1*0301 are associated with Chinese Han SLE

Jingying Sun¹ | Chao Yang¹ | Wenmin Fei¹ | Xuelei Zhang¹ | Yujun Sheng¹,² | Xiaodong Zheng¹ | Huayang Tang¹,² | Wanling Yang³ | Sen Yang¹,² | Xing Fan¹,² | Xuejun Zhang¹,²,⁴

¹Institute of Dermatology and Department of Dermatology at NO. 1 Hospital, Anhui Medical University, Hefei, China
²Key Laboratory of Dermatology, Anhui Medical University, Ministry of Education, Hefei, China
³LKS Faculty of Medicine, Department of Paediatrics and Adolescent Medicine, The University of Hong Kong, Pokfulam, Hong Kong
⁴Department of Dermatology, No. 2 Hospital, Anhui Medical University, Hefei, China

Correspondence
Xuejun Zhang and Xing Fan, Institute of Dermatology and Department of Dermatology at NO. 1 Hospital, Anhui Medical University, Hefei, China.

Emails: ayzxj@vip.sina.com and ms_fanxing@163.com

Funding information
the National Natural Science Foundation of China, Grant/Award Number: 81371722

Abstract
Background: Several susceptibility loci have been identified associated with Chinese Han systemic lupus erythematosus (SLE).

Methods: We carried out imputation of classical HLA alleles, amino acids and Single Nucleotide Polymorphisms (SNPs) across the MHC region in Chinese Han SLE genome-wide association study (GWAS) of mainland and Hong Kong populations for the first time using newly constructed Han-MHC reference panel followed by stepwise conditional analysis.

Results: We mapped the most significant independent association to HLA-DQβ1 at amino acid position (Phe87, \( p = 7.807 \times 10^{-9} \)) and an independent association at HLA-DQB1*0301 (\( P_{\text{conditional}} = 1.43 \times 10^{-7} \)).

Conclusion: Our study illustrates the value of population-specific HLA reference panel for fine-mapping causal variants in the MHC.

KEYWORDS
HAN, imputation, MHC, systemic lupus erythematosus

1 | BACKGROUND

The major histocompatibility complex (MHC) region (chr 6:29–34 Mb) harbors the human leukocyte antigen (HLA) genes of which many are associated with autoimmune diseases (Fernando et al., 2008). The risk of autoimmunity conferred by HLA polymorphisms is likely the result of variation in amino acid residues at specific positions, which may alter the structure and function of presented peptides (Astill, Ellis, Arif, Tree, & Peakman, 2003; Lee, Wucherpfennig, & Wiley, 2001; van Lummel et al., 2014; Scally et al., 2013). For certain disease, specific amino acid positions within HLA molecules may play an important functional role. SLE is a heterogeneous disease characterized by autoantibody production and damage to multiple organs due to immune complexes and inflammation (Chung et al., 2011). A genetic contribution of the human leukocyte antigen (HLA) region to SLE has been supported by epidemiological studies and several genetic studies (Deng & Tsao, 2010; Lee et al., 2014). Genes in class II...
HLA regions are dominantly represented as SLE susceptibility loci especially in T cell dependent antibody responses. Professors have shown that HLA-DRB1 was significantly associated with autoantibody subsets in SLE patients (Connolly & Hakonarson, 2012). Then studies fine-mapped the primary association within the MHC locus with SLE to HLA-DR, and further narrowed it down to specific amino-acid positions. For example, the role of amino acid position 11-13-26 in HLA-DRβ1 for Korean SLE susceptibility has been established (Kim et al., 2014). However, the effects of the genetic architecture of the MHC region on SLE risk have not yet to be fully elucidated due to the complexity of the region, extended regions of linkage disequilibrium (LD), and a lack of statistical power. Al-Motwee et al. found that HLA-DQB1*06 was associated with Saudis SLE patients (Al-Motwee et al., 2013). SNP rs2187668 at HLA-DR3 was significantly associated with anti-dsDNA and was stronger of association in anti-dsDNA positive SLE subjects compared with negative ones (Chung et al., 2011). However, those loci don’t fully explain the HLA mediated risk of SLE, and did not investigate the functional amino acids due to lack of a reference panel suitable for imputing their genetic variants. Data from GWAS also have had insufficient variant density to define the association signals within the MHC. In this study, we used recently established Han-MHC reference panel (Zhou et al., 2016) imputing Chinese SLE GWAS data to identify potential independent amino acid positions.

2 | METHODS

2.1 | Ethical compliance

The study was approved by the relevant local Institutional Ethical Committees and informed consent was obtained from patients and families.

2.2 | SLE GWAS data

We extracted genome-wide SNPs of 1047 cases and 1205 controls subjects (Chinese mainland SLE, data set #1) and SNP data of 612 cases and 2193 controls individuals (Hong Kong SLE, data set #2) in previously SLE GWAS, and filtered using standard quality control criteria (Han et al., 2009; Yang et al., 2010) (Table S1), including SNP and sample call rate, exclusion of closely related relative and outliers in terms of ancestry, and SNP minor allele frequency (MAF) and Hardy–Weinberg equilibrium cutoffs.

2.3 | HLA imputation

We imputed classical HLA alleles, HLA amino acid residues and untyped SNPs from each data set by SNP2HLA and the new Han-MHC reference panel. Imputed markers with minor allele frequency (≥1%) and imputation quality (PLINK R²) ≥0.3 were used in disease association tests. Basing on the known amino acid sequences of classical alleles of HLA-A, -B, -C, -DPA1, -DPB1, -DQA1, -DQB1 and -DRB1 in the IMGT/HLA database (database release 3.13.1) (Kallberg et al., 2007; Robinson et al., 2015) translated the amino acid residues of each HLA genes. All information about the SNPs, amino acid residues and two-digit and four-digit HLA alleles were encoded as binary variables and phased by Beagle 3.0.4 imputation program (Browning & Browning, 2009) powered by SNP2HLA (to extract SNP genotypes located in the MHC region) method (Jia et al., 2013).

All genotype data from the two data sets were merged as a single data set after excluding the SNPs that were not present in both data sets (n = 3,175 SNPs in mainland GWAS data; n = 2,885 SNPs in Hong Kong GWAS data; n = 6,060 SNPs in both panels). The study subject was highly homogenous in a principal component (PC) analysis (Gibbs et al., 2003).

2.4 | Association analysis

For each phenotype, we assessed variant risk with a logistic-regression model assuming additive effects of the allele dosages in the log-odds scale. We defined HLA variants to include two- and four-digit biallelic classical HLA alleles, biallelic HLA amino acid polymorphisms for respective residues, multiallelic HLA amino acid polymorphisms and biallelic SNPs across the entire MHC region. To account for potential population substructure, we included the top ten PCs and an indicator variable for each data set as covariates when examined SLE association of the imputed dosage of each marker with minor allele frequency of ≥1% and imputation quality (PLINK R²) of ≥0.3 by logistic regression. We also used a forward logistic regression model to find additional markers with independent SLE-risk effect by adding the identified markers as covariates for conditional analysis.

2.5 | Meta analysis

We used the inverse variance method for meta analysis, combining data from the two studies (Mainland and Hong Kong) for SNPs, alleles and amino acid with an imputation R² score of ≥0.3 in two studies.

3 | RESULTS

3.1 | HLA Imputation and association testing

After imputation of SLE GWAS data for Chinese Han subjects (1,659 cases versus 3,398 controls) we obtained
genotypes for 10 HLA two-digit alleles, 13 HLA four-digit alleles, 173 HLA amino acid positions and 231 SNPs encoded by HLA genes of the class I and class II ($p < 1.67 \times 10^{-6}$) from the genotyped MHC SNPs (29–34 Mb at chromosome 6). We set a study-wide significance threshold of $p = 1.67 \times 10^{-6}$ on the basis of the total number of imputed HLA variants (0.05/30,000).

3.2 | Risk of HLA-DQβ1 at amino acid position 87 associated with SLE subjects

When testing the imputed HLA variants in the MHC region for association with SLE risk, we demonstrated the principal association with HLA-DQβ1. We assessed the risk associated with both classical HLA alleles and HLA amino acid polymorphisms, and by a meta-analysis we found the most significant association was amino acid position 87 of HLA-DQβ1 ($P_{\text{meta}} = 7.81 \times 10^{-17}$, OR = 1.785; Figure 1a; Tables 1, S1).

3.3 | Independent HLA associations in HLA-DQB1*0301

We further investigated whether SLE risk was associated with other HLA genes independent of HLA-DQβ1 amino acid polymorphisms. When conditioning on the most significant amino acid in HLA-DQβ1 87, we detected significant independent associations at HLA-DQB1*0301 ($P_{\text{meta}} = 6.91 \times 10^{-15}$, OR = 0.5508) (Figure 1b). After conditioning on HLA-DQB1*0301 and HLA-DQβ1 amino acid position 87, no variants in the MHC region satisfied the study-wide significance threshold ($p > 1.67 \times 10^{-6}$; Figure 1c; Table S2). These results suggested that polymorphisms in class II HLA genes, particularly HLA-DQβ1, explain the majority of risk for SLE in Chinese Han population.

3.4 | HLA-DQB1 amino acid position 87 risks are shared between mainland and Hong Kong

After imputation in data set #2, we found that the effect size and direction of residue at HLA-DQβ1 amino acid position 87 and HLA-DQB1*0301 in Hong Kong subjects were highly consistent with the results in mainland ($p = 5.477 \times 10^{-4}$ and $3.79 \times 10^{-7}$).

4 | DISCUSSION

Systemic Lupus Erythematosus (SLE) (OMIM 152700) is a polygenic disorder characterized by chronic and systemic inflammation and affecting multiple organs due to a loss of immune tolerance against self-antigens. The genetic heritability of SLE ranged from 44% to 66% (Lawrence, Martins, & Drake, 1987; Wang et al., 2007). To date, more than
60 susceptibility loci have been identified in genome wide and candidate gene association studies (Boackle, 2013). Association analysis of HLA genes at amino acid sites have facilitated fine-mapping efforts in immune-related diseases (Pereyra et al., 2010; Raychaudhuri et al., 2012). To identify potentially causal variation within HLA genes associated with Chinese SLE patients, we carried out an imputation based on the Han-MHC reference panel for classical HLA alleles as well as amino acid polymorphisms in mainland and Hong Kong Han SLE patients. Our results support and refine previous findings of multiple signals in the MHC for population of Asia or European ancestry. By testing HLA alleles, amino acids, and SNPs at the same time, we were able to pinpoint the amino acid position 11 and 13 of HLA-DR\(_b\) as a significant signal in both mainland and Hong Kong Chinese Han populations (\(p = 2.91 \times 10^{-10}\) both), which was recognized as the major risk factor for SLE in Korean (Kim et al., 2014). This may suggest that genetic risk of the MHC region on SLE are generally shared within Asian populations to a certain extent (Kim et al., 2009, 2014).

We note that those sites explain part of the variation in MHC-mediated risk for both populations, and that the residues confer similar directions and relative magnitude of risk. We also observe that allele HLA-DRB1*15:01 (\(p = 8.08 \times 10^{-11}\)) and HLA-DQA1*0102 (\(p = 2.96 \times 10^{-10}\)) in European populations (Morris et al., 2012, 2014) is generally concordant with the results presented here for Chinese Han populations. HLA-DRB1 (MIM 142857) alleles are correlated with alleles of the class II loci HLA-DQA1 (MIM 146880) and HLA-DQB1 (MIM 604305) as a result of strong LD in the class II region (HLA-DR and HLA-DQ). However, amino acid in HLA-DQB1 have rarely been reported showing association with SLE. In our study amino acid position 87 of HLA-DQB1 has the most strongest association with SLE risk, which was firstly been reported till now. Furthermore, HLA-DQB1 amino acid position 87 and HLA-DQB1*0301 in Hong Kong individuals were highly consistent with mainland population. HLA genes encode cell surface proteins that display antigenic peptides to effector immune cells to regulate self-tolerance and downstream immune

### Table 1

Association results of the Top-20 associated markers after HLA imputation

| Marker ID\(^a\) | Position\(^b\) | Alleles\(^c\) | Frequency [%]\(^d\) | Nearest gene |
|----------------|----------------|--------------|----------------------|--------------|
| AA_DQB1_87_32632598_F | 32632598 | P/A | 0.3031 | 7.81E-17 | 1.785 | HLA-DQB1 |
| AA_DQB1_86_32632601_A | 32632601 | P/A | 0.4345 | 5.18E-16 | 1.652 | HLA-DQB1 |
| AA_DQB1_9_32632832_Y | 32632832 | P/A | 0.3523 | 3.05E-15 | 1.678 | HLA-DQB1 |
| AA_DQB1_140_32629890_A | 32629890 | P/A | 0.3967 | 6.30E-15 | 0.6247 | HLA-DQB1 |
| AA_DQB1_140_32629890_T | 32629890 | P/A | 0.3967 | 6.30E-15 | 0.6247 | HLA-DQB1 |
| AA_DQB1_182_32629764_N | 32629764 | P/A | 0.3967 | 6.30E-15 | 0.6247 | HLA-DQB1 |
| AA_DQB1_182_32629764_S | 32629764 | P/A | 0.3967 | 6.30E-15 | 0.6247 | HLA-DQB1 |
| AA_DQB1_55_32632694_P | 32632694 | P/A | 0.3356 | 6.91E-15 | 0.6188 | HLA-DQB1 |
| HLA_DQB1_03 | 32631061 | P/A | 0.3356 | 6.91E-15 | 0.6188 | HLA-DQB1 |
| AA_DQB1_125_32629935_A | 32629935 | P/A | 0.4646 | 8.81E-14 | 1.576 | HLA-DQB1 |
| AA_DQB1_220_32629141_H | 32629141 | P/A | 0.4646 | 8.81E-14 | 1.576 | HLA-DQB1 |
| AA_DQB1_220_32629141_R | 32629141 | P/A | 0.4646 | 8.81E-14 | 1.576 | HLA-DQB1 |
| AA_DQB1_221_32629138_H | 32629138 | P/A | 0.4646 | 8.81E-14 | 1.576 | HLA-DQB1 |
| AA_DQB1_221_32629138_Q | 32629138 | P/A | 0.4646 | 8.81E-14 | 1.576 | HLA-DQB1 |
| AA_DQB1_53_32632700 | 32632700 | Q/L | 0.4646 | 8.81E-14 | 1.576 | HLA-DQB1 |
| AA_DQB1_84_32632607 | 32632607 | E/Q | 0.4646 | 8.81E-14 | 1.576 | HLA-DQB1 |
| AA_DQB1_85_32632604 | 32632604 | V/L | 0.4646 | 8.81E-14 | 1.576 | HLA-DQB1 |
| AA_DQB1_86_32632601_E | 32632601 | P/A | 0.4646 | 8.81E-14 | 1.576 | HLA-DQB1 |
| AA_DQB1_87_32632598_L | 32632598 | P/A | 0.4646 | 8.81E-14 | 1.576 | HLA-DQB1 |
| AA_DQB1_89_32632592 | 32632592 | G/T | 0.4646 | 8.81E-14 | 1.576 | HLA-DQB1 |

\(^a\)Composite and low frequency markers (MAF < 1% in the entire sample set) are not shown.

\(^b\)Chromosome 6 positions for SNP markers according to genome build hg19.

\(^c\)Present/Absent for classical HLA alleles and in the case that a specific amino acid is given in the Marker ID (see first column).

\(^d\)Frequency and relative risk (OR) are given for the first allele denoted in the column ‘Alleles’.
responses. Variation in amino acid residues at specific positions within the antigen-bind grooves may alter the repertoire of presented peptides, and results in the risk of autoimmunity (van Lummel et al., 2014; Scally et al., 2013). A Phe in position 87 (Figure 1d) is located in peptide-binding grooves of HLA molecules according to our results, suggesting functional contributions to antigen-presentation ability or protein stability (Jin et al., 2014; Miyadera, Ohashi, Lernmark, Kitamura, & Tokunaga, 2015). These data suggest the involvement of peptide antigens bound to specific HLA molecules in controlling the development of SLE.

Yusmin et al. (Mohd-Yusuf, Phipps, Chow, & Yeap, 2011) has ever analyzed HLA alleles in 160 SLE patients (99 Chinese and 61 Malays) and 107 healthy control individuals. These data suggest the involvement of peptide antigens bound to specific HLA molecules in controlling the development of SLE.

In summary, by using imputation as an approach to identify disease risk variants within the MHC region, we determined the population-specific genetic variants of HLA polymorphisms and also reproduced known HLA alleles and amino acid polymorphisms effects from genotyped SNP data in SLE by using Han-MHC reference panel. Furthermore, given the prevalence of SLE, our study highlights the role of genetic variants in common diseases.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

ORCID
Jingying Sun http://orcid.org/0000-0002-2884-7225

REFERENCES
Al-Motwee, S., Jawdat, D., Jehani, G. S., Anazi, H., Shubaili, A., Sutton, P., . . . Hajeer, A. H. (2013). Association of HLA-DRB1*15 and HLADQB1*06 with SLE in Saudis. Annals of Saudi Medicine, 33(3), 229–234. https://doi.org/10.5144/0256-4947.2013.229
Astill, T. P., Ellis, R. J., Arif, S., Tree, T. I., & Peakman, M. (2003). Promiscuous binding of proinsulin peptides to Type 1 diabetes-permissive and -protective HLA class II molecules. Diabetologia, 46(4), 496–503. https://doi.org/10.1007/s00125-003-1070-3
Boackle, S. A. (2013). Advances in lupus genetics. Current Opinion in Rheumatology, 25(5), 561–568. https://doi.org/10.1097/BOR.0b013e32836eb4e
Browning, B. L., & Browning, S. R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. American Journal of Human Genetics, 84(2), 210–223. https://doi.org/10.1016/j.ajhg.2009.01.005
Chung, S. A., Taylor, K. E., Graham, R. R., Nithitham, J., Lee, A. T., Otmann, W. A., . . . Criswell, L. A. (2011). Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. PLoS Genetics, 7, e1001323. https://doi.org/10.1371/journal.pgen.1001323
Connolly, J. J., & Hakonarson, H. (2012). Role of cytokines in systemic lupus erythematosus: Recent progress from GWAS and sequencing. Journal of Biomedicine & Biotechnology, 2012, 798924. https://doi.org/10.1155/2012/798924
Deng, Y., & Tsao, B. P. (2010). Genetic susceptibility to systemic lupus erythematosus in the genomic era. Nature Reviews Rheumatology, 6, 683–692. https://doi.org/10.1038/nrrheum.2010.176
Fernando, M. M., Stevens, C. R., Walsh, E. C., De Jager, P. L., Goyette, P., Plenge, R. M., . . . Rioux, J. D. (2008). Defining the role of the MHC in autoimmunity: A review and pooled analysis. PLoS Genetics, 4(4), e1000024. https://doi.org/10.1371/journal.pgen.1000024
Gibbs, R. A., Belmont, J. W., Hardenbol, P., Willis, T. D., Yu, F. L., Yang, H. M., . . . Tanaka, T. (2003) The International HapMap Project Nature, 426(6968): 789–796. https://doi.org/10.1038/nature02168
Han, J. W., Zheng, H. F., Cui, Y., Sun, L. D., Ye, D. Q., Hu, Z., . . . Zhang, X. J. (2009). Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nature Genetics, 41(11), 1234–1237.
Jia, X., Han, B., Onengut-Gumuscu, S., Chen, W. M., Concannon, P. J., Rich, S. S., . . . de Bakker, P. I. (2013). Imputing amino acid polymorphisms in human leukocyte antigens. PLoS ONE, 8(6), e64683. https://doi.org/10.1371/journal.pone.0064683
Jin, H., Arase, N., Hirayasu, K., Kohyama, M., Suenaga, T., Saito, F., . . . Arase, H. (2014). Autoantibodies to IgG/HLA class II complexes are associated with rheumatoid arthritis susceptibility. Proceedings of the National Academy of Sciences of the United States of America, 111(10), 3787–3792. https://doi.org/10.1073/pnas.1401105111
Källberg, H., Padyukov, L., Plenge, R. M., Ronelid, J., Gregersen, P. K., van der Helm-van Mil, A. H., . . . Epidemiological Investigation of Rheumatoid Arthritis study group. (2007). Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. American Journal of Human Genetics, 80(5), 867–875. https://doi.org/10.1086/516736
Kim, K., Bang, S. Y., Lee, H. S., Okada, Y., Han, B., Saw, W. Y., . . . Bae, S. C. (2014). The HLA-DRB1*15 amino acid positions 11–13–26 explain the majority of SLE-MHC associations. Nature Communications, 5, 5902. https://doi.org/10.1038/ncomms6902
Kim, I., Kim, Y. J., Kim, K., Kang, C., Choi, C. B., Sung, Y. K., . . . Bae, S. C. (2009). Genetic studies of systemic lupus...
HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nature Genetics, 44(3), 291–296. https://doi.org/10.1038/ng.1076

Robinson, J., Halliwell, J. A., Hayhurst, J. D., Fliceck, P. P., Parham, P., & Marsh, S. G. (2015). The IPD and IMGT/HLA database: Allele variant databases. Nucleic Acids Research, 43 (Database issue), D423–D431. https://doi.org/10.1093/nat/gku1161

Scally, S. W., Petersen, J., Law, S. C., Dudek, N. L., Nel, H. J., Loh, K. L., ... Rossjohn, J. (2013). A molecular basis for the association of the HLA-DRB1 locus, citrullination, and rheumatoid arthritis. Journal of Experimental Medicine, 210(12), 2569–2582. https://doi.org/10.1084/jem.20131241

Wang, J., Yang, S., Chen, J. J., Zhou, S. M., He, S. M., Liang, Y. H., ... Zhang, X. J. (2007). Systemic lupus erythematosus: A genetic epidemiology study of 695 patients from China. Archives of Dermatological Research, 298(10), 485–491. https://doi.org/10.1007/s00105-007-0719-4

Yang, W., Shen, N., Ye, D. Q., Liu, Q., Zhang, Y., Qian, X. X., Hirankarn, N., ... ; Asian Lupus Genetics Consortium (2010). Genome-wide association study in Asian populations identifies variants in ETS1 and WDF4 associated with systemic lupus erythematosus. PLoS Genetics, 6(2), e1000841. https://doi.org/10.1371/journal.pgen.1000841

Zhou, F., Cao, H., Zuo, X., Zhang, T., Zhang, X. J. (2007). Systemic lupus erythematosus: A heritable trait with strong major gene effects. Nature Genetics, 39(5), 780–784. https://doi.org/10.1038/ng.2272

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article.