EXTENDED REPORT

Genome-wide search followed by replication reveals genetic interaction of *CD80* and *ALOX5AP* associated with systemic lupus erythematosus in Asian populations

Yan Zhang,[1] Jing Yang,[1] Jing Zhang,[1] Liangdan Sun,[2] Nattiya Hirankarn,[3] Hai-Feng Pan,[4] Chak Sing Lau,[5] Tak Mao Chan,[5] Tsz Leung Lee,[1] Alexander Moon Ho Leung,[6] Chi Chiu Mok,[7] Lu Zhang,[1] Yongfei Wang,[1] Jiangshan Jane Shen,[1] Sik Nin Wong,[8] Ka Wing Lee,[9] Marco Hok Kung Ho,[1] Pamela Pui Wah Lee,[1] Brian Hon-Yin Chung,[1] Chun Yin Chong,[1] Raymond Woon Sing Wong,[5] Mo Yin Mok,[5] Wilfred Hing Sang Wong,[1] Kwok Lung Tong,[10] Niko Kei Chiu Tse,[11] Xiang-Pei Li,[12] Yingyos Avihingsanon,[13] Pornpimol Rianthavorn,[14] Thavatchai Deekajorndej,[14] Kanya Suphapeetiporn,[14] Vorasuk Shotelersuk,[14] Shirley King Yee Ying,[10] Samuel Ka Shun Fung,[10] Wai Ming Lai,[11] Chun-Ming Wong,[15] Irene Oi Lin Ng,[15] Maria-Merce Garcia-Barcelo,[16] Stacey S Cherny,[17] Yong Cui,[2] Pak Chung Sham,[17,18] Sen Yang,[2] Dong-Qing Ye,[3] Xue-Jun Zhang,[2] Yu Lung Lau,[1] Wanling Yang[1,18]

Handling editor Tore K Kvien

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/annrheumdis-2014-206367).

For numbered affiliations see end of article.

Correspondence to
Dr Wanling Yang, Department of Paediatrics and Adolescent Medicine, The University of Hong Kong, Room 58, 7/F, Laboratory Block, 21 Sassoon Rd., Pokfulam, Hong Kong, Hong Kong;
yangwl@hku.hk

Received 28 July 2014
Revised 4 March 2015
Accepted 22 March 2015

To cite: Zhang Y, Yang J, Zhang J, *et al*. *Ann Rheum Dis* Published Online First: [*please include* Day Month Year] doi:10.1136/annrheumdis-2014-206367

ABSTRACT

Objectives Genetic interaction has been considered as a hallmark of the genetic architecture of systemic lupus erythematosus (SLE). Based on two independent genome-wide association studies (GWAS) on Chinese populations, we performed a genome-wide search for genetic interactions contributing to SLE susceptibility.

Methods The study involved a total of 1 659 cases and 3 398 controls in the discovery stage and 2 612 cases and 3 441 controls in three cohorts for replication. Logistic regression and multifactor dimensionality reduction were used to search for genetic interaction.

Results Interaction of *CD80* (rs2222631) and *ALOX5AP* (rs12876893) was found to be significantly associated with SLE (OR_int=1.16, *P_int_all*=7.7E-04 at false discovery rate<0.05). Single nuclear polymorphism rs2222631 was found associated with SLE with genome-wide significance (*P_all*=4.5E-08, OR=0.86) and is independent of rs6804441 in *CD80,* whose association was reported previously. Significant correlation was observed between expression of these two genes in healthy controls and SLE cases, together with differential expression of these genes between cases and controls, observed from individuals from the Hong Kong cohort. Genetic interactions between *BLK* (rs13277113) and *DDX6* (rs4639966), and between *TNFSF4* (rs844648) and *PXK* (rs6445975) were also observed in both GWAS data sets.

Conclusions Our study represents the first genome-wide evaluation of epistasis interactions on SLE and the findings suggest interactions and independent variants may help partially explain missing heritability for complex diseases.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a prototypical autoimmune disease that mainly affects women of child-bearing age with multiorgan damage and production of antibodies against self. Genetic factors play important roles in SLE, supported by the fact that the concordance rate for SLE is much higher in monozygotic (25–70%) twins than in dizygotic twins (2–9%).[1 2] Genome-wide association studies (GWAS) have greatly improved our understanding of the genetic factors for the disease. Single locus analysis has been the major method used in most GWAS to date, despite the understanding that interplays between multiple genetic variants may be important in conferring disease susceptibility.[3] Thus, detecting genetic interactions between different loci may allow us to better understand the genetic factors involved in this disease.

Several methods have been developed to detect genetic interactions using genome-wide association data.[3] For one group of methods, interactions are tested by fitting a logistic regression model of effects from individual variants and effects from genetic interaction. However, logistic regression-based methods are often criticised for their inability to deal with non-linear models and with high-dimensional data that contain many potentially interacting predictor variables.[4 5] The second group of approaches seeks to identify combinations of loci that would together influence a disease outcome, possibly by interactions rather than by main individual effects. Multifactor dimensionality reduction (MDR)[6] seeks to identify evidence for higher-order genetic interactions in the absence of

statistically significant main effects to the disease. It has been used to identify potential interacting loci for several diseases such as type 2 diabetes[7] and rheumatoid arthritis.[8]

So far, most reported interactions on SLE were found through a candidate gene approach, for which prior biological knowledge and confirmed disease association were taken into consideration.[9–13] Moreover, unlike main effects for associated variants, very few interactions have been supported by further studies on additional independent populations or cohorts. In this study, we focused on testing pairwise genetic interactions for SLE using two GWAS conducted on Chinese populations. Potential interaction pairs were followed up by replication from three cohorts. We confirmed one interacting single nuclear polymorphism (SNP) pair as associated with SLE susceptibility, of which one SNP was located in a locus not previously known to be associated with SLE (*ALOX5AP*) and another one was an independent variant from a known associated gene (rs2222631 in *CD80*).

MATERIALS AND METHODS

Study participants

The samples included in the current study were collected from Hong Kong, Anhui, China, Bangkok and Thailand, as reported previously[14–16] (figure 1, see online supplementary table S1). All the cases have fulfilled the revised criteria of the American College of Rheumatology for diagnosis of SLE. Cases from Hong Kong were recruited from five hospitals (HK_GWAS and HK_REP). Controls from Hong Kong in the discovery stage were individuals from other GWAS studies conducted at the University of Hong Kong using the same platform (HK_GWAS) and blood donors from Hong Kong Red Cross in the replication stage (HK_REP). The cases for the Anhui GWAS (AH_GWAS) and replication stage (AH_REP) were recruited from several hospitals in central and southern China, and the controls were carefully selected from geographically matched unrelated healthy individuals from respective provinces (AH_GWAS and AH_REP). The cases from Thailand were patients visiting King Chulalongkorn Memorial Hospital and geographically matched healthy donors were used as controls (TH_REP).

Analysis of genetic interaction

Genetic interaction (case-control analysis) by logistic regression was done using PLINK.[17] A total of 33 704 and 33 168 interacting SNP pairs with *P*_int (interaction p values) <0.05 were retained for Anhui GWAS data and Hong Kong GWAS data, respectively. Overlapping SNP pairs in these two independent evaluations were used for further analysis. The false discovery rate (FDR) method described by Benjamini and Hochbergi was used to correct for multiple comparisons.[18]

MDR was used to detect and characterise high-order genetic interactions. The high-risk di-genotypes for SLE were defined when the percentage of patients with those di-genotypes was equal to or greater than that of controls, whereas the low-risk groups were defined when the ratio of percentages was lower than 1. The analysis was carried out using V.2.0β8.4 of the MDR software package (http://www.epistasis.org).

Genotyping

The GWAS data were generated using Illumina 610-Quad Beadchip for both data sets (HK_GWAS and AH_GWAS). Replication of the candidate SNPs was performed by the TaqMan genotyping method (Life Technologies) on the remaining samples from Hong Kong that were not included in the discovery stage (HK_REP), samples collected from Bangkok, Thailand (TH_REP), and samples from an independent Anhui cohort (AH_REP).

Analysis of association

For the GWAS data from Hong Kong and Anhui, the quality control processes were conducted separately using PLINK.[17] SNPs were removed if the genotype call rate was <90%, minor

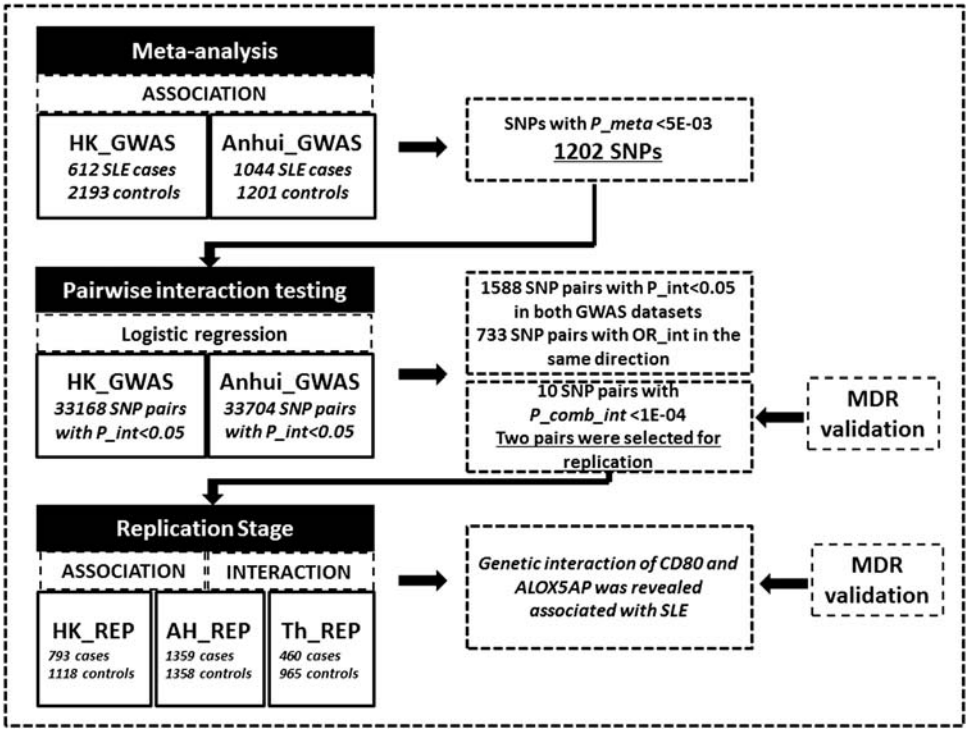

Figure 1 Flow chart on the experimental process and SNP selection criteria.

allele frequency was <1%, and there was violation of the Hardy-Weinberg equilibrium ($p \leq 10^{-4}$). The samples were removed from analysis if their SNP call rate was <90% or if a hidden relationship was detected. A total of 493 346 autosomal SNPs in 612 cases and 2193 controls from the Hong Kong cohort, and 493 645 autosomal SNPs in 1047 cases and 1205 controls from the Anhui cohort were analysed.[16]

We used the inverse variance method installed in METAL[19] for meta-analysis. Joint analysis of association was conducted using the Cochran–Mantel–Haenszel (CMH) test, and homogeneity of the effect size between different cohorts and different stages of the study was tested by the Breslow-Day test, both installed in PLINK. Conditional logistic regression was done by SNPTESTv2.4.1 to test the independent effect of an SNP, adjusting for the effect from other SNP(s) in the same locus and also treating the cohort as a covariate.

Real-time quantitative PCR on the expression of *CD80* and *ALOX5AP*

Total mRNA extracted from peripheral blood mononuclear cells (PBMCs) using the All Prep DNA/RNA mini kit (QIAGEN) was reverse-transcribed into cDNA (Life Technologies). Using real-time quantitative PCR (qRT-PCR), the expression levels of *CD80* (Life Technologies, Hs01045161_m1) and *ALOX5AP* (Hs00970920_m1), normalised by 18S rRNA (Hs99999901_s1), were calculated using the $2-\Delta\Delta Ct$ method. Log10 transformed *CD80* and *ALOX5AP* expression levels were compared between patients with SLE and healthy controls using Student's t test. Pearson correlations were computed between the expression of *CD80* and *ALOX5AP* in cases and controls using Graphpad Prism5. Logistic regression was used to examine the correlation of expression level of *CD80* and *ALOX5AP* and SNP genotypes.

RESULTS

Identification of genetic interactions from GWAS data

We tested pairwise interaction among the SNPs, using the two GWAS data sets collected from Hong Kong and Anhui, China, respectively. Considering the limitation of computational efficiency, the SNPs were preselected based on their association results from meta-analysis, and 1202 SNPs with meta-analysis p value (*P*_meta) <5E-03 were evaluated. Testing for pairwise interaction among these SNPs in both GWAS data sets was first performed by logistic regression using PLINK.[17] In total, 1588 SNP pairs remained nominally significant in both GWAS data sets (*P*_int <0.05), but 855 SNP pairs showed different directions in terms of the effect of interaction (OR_int<1 or OR_int>1) in the two data sets and were not further pursued.

For the remaining pairs, meta-analysis was performed on the interaction results from the two independent data sets using METAL.[19] Based on meta-analysis of interaction (*P*_comb_int), 15 SNP pairs (*P*_comb_int <1E-04) were further examined. When SNP pairs having high linkage disequilibrium (LD) with each other were collapsed, 10 pairs remained, and the results are shown in table 1. The FDR method was used to correct for multiple comparisons,[18] and all 10 candidate SNP pairs had an FDR <0.05. All 20 SNPs showed moderate association with SLE when they were analysed individually, based on the meta-analysis of the two GWAS data sets ($3.8E\text{-}04 \leq P\_meta \leq 2.2E\text{-}03$) (see online supplementary table S2).

Secondly, pairwise interaction tests were performed using MDR analysis on the 10 SNP pairs selected from logistic regression. MDR identified SNP pairs with maximum cross-validation consistency and balanced accuracy using the two-locus model (see online supplementary table S3). In agreement with the results from logistic regression, significant interactions ($p<0.01$) were observed in both GWAS data sets for all the 10 pairs. The directions of the interaction based on OR (OR_int) for this model were also examined for these pairs, and seven pairs showed consistency by the two methods. We selected pair rs2222631 (*CD80*) and rs12876893 (*ALOX5AP*), pair rs4073643 (*LHFPL3*) and rs7920305 (*Fas*) for further replication in additional cohorts based on significant combined interaction p value (*P*_comb_int) (table 1) and understanding of the biological functions of the genes involved. The selection process was further elaborated in the description of table 1.

Replication of the two candidate interaction pairs

Further replication for selected SNPs was performed by the TaqMan SNP genotyping method using assay-on-demand probes and primers (Life Technologies) on the remaining samples (figure 1, see online supplementary table S1). Association of each SNP with SLE was evaluated based on samples from discovery stage and replication stage (table 2). Among the four SNPs, rs2222631 (*CD80*) reached genome-

Table 1 Genetic interactions detected from two GWAS data sets by logistic regression

						AH_GWAS		HK_GWAS			
SNP1	CHR1	GENE1	SNP2	CHR2	GENE2	OR_int	*P_int*	OR_int	*P_int*	*P_comb_int*	HetPVal
rs1358357	7	*AUTS2*	rs7039222	9	*PXK*	0.71	1.0E-04	0.82	3.1E-02	1.5E-05	0.29
rs10160742	11	*CADM1*	rs2028416	14	*KIAA0284*	1.71	2.2E-04	1.31	4.6E-02	4.9E-05	0.30
rs4674516	2	NA	rs7494172	14	*IGH*	1.56	2.9E-04	1.26	3.2E-02	3.8E-05	0.38
rs10816735	9	*IKBKAP*	rs3812945	15	*SCAMP5*	3.4	3.5E-04	2.19	1.3E-02	1.7E-05	0.54
rs4888121	16	*PKD1L2*	rs243642	21	*LINC00113*	0.66	4.5E-04	0.75	1.4E-02	2.3E-05	0.55
rs2222631	**3**	***CD80***	**rs12876893**	**13**	***ALOX5AP***	**1.33**	**6.1E-04**	**1.22**	**3.5E-02**	**7.9E-05**	**0.44**
rs1319447	11	*ABCC8*	rs11226163	11	*PDGFD*	0.68	6.5E-04	0.79	4.1E-02	9.9E-05	0.42
rs1416819	1	*RNF186*	rs887564	4	NA	1.76	7.7E-04	1.43	4.1E-02	1.2E-04	0.43
rs4073643	**7**	***LHFPL3***	**rs7920305**	**10**	***FAS***	**1.37**	**9.0E-04**	**1.34**	**4.3E-03**	**1.2E-05**	**0.87**
rs17040235	2	*NRXN1*	rs3751120	11	*FLRT1*	1.34	9.3E-04	1.24	2.8E-02	8.7E-05	0.53

HetPVal is the heterogeneity p value. The two bold-faced SNP pairs were selected for further replication. The pair *CD80* (rs2222631) and *ALOX5AP* (rs12876893) was selected upon the known association of *CD80* to SLE, and the explainable biological function of ALOX5AP to leukotriene synthesis. The pair *LHFPL3* (rs4073643) and *FAS* (rs7920305) was selected as the most significant *P*_comb_int (1.2E-05) in the 10 pairs. Furthermore, the genes *CD80, ALOX5AP and FAS* were prioritised with obvious mouse knockout immune system phenotypes from public database, especially *Fas* showed a key role in the pathogenesis of various malignancies and immune system diseases.[43]
AHGWAS, Anhui GWAS data; GWAS, genome-wide association studies; HKGWAS, Hong Kong GWAS data; OR_int means OR for interaction, and a value of 1.0 indicates no effect; SLE, systemic lupus erythematosus.

Table 2 Association of interaction SNPs with SLE by joint analysis of results from three cohorts and combining data from GWAS and replication

SNP	Allele	AH_GWAS			HK_GWAS			AH_REP			HK_REP			TH_REP					
		F_A	F_U	p Value	F_A	F_U	p Value	F_A	F_U	p Value	F_A	F_U	p Value	CASE	CTRL	p Value	*P*_all	OR (95% CI)	*P*_het
rs2222631	G<A	0.43	0.47	1.90E-03	0.45	0.48	0.09	0.44	0.47	0.02	0.46	0.5	1.40E-03	0.46	0.5	0.12	**4.50E-08**	0.86 (0.81 to 0.91)	0.88
rs12876893	G<A	0.44	0.41	0.02	0.52	0.48	0.01	0.44	0.43	0.53	0.5	0.47	0.08	0.56	0.52	0.07	**6.20E-05**	1.12 (1.06 to 1.18)	0.57
rs4073643	G<A	0.37	0.32	4.90E-04	0.31	0.29	0.12	0.35	0.35	0.64	0.3	0.31	0.37	0.22	0.24	0.24	8.10E-02	1.06 (0.99 to 1.12)	0.01
rs7920305	G<A	0.4	0.36	0.01	0.52	0.48	9.00E-03	0.39	0.38	0.61	0.5	0.49	0.33	0.49	0.47	0.27	**7.40E-04**	1.10 (1.04 to 1.17)	0.4

The calculation of OR is also based on the minor allele of each SNP. Risk allele of each SNP is underlined.
The p values are shown in bold if the SNPs showed a consistent trend in the replication stage.
AHGWAS, Anhui GWAS data; AHREP, Anhui replication panel; GWAS, genome-wide association studies; HKGWAS, Hong Kong GWAS data; HKREP, Hong Kong replication panel; OR, OR from the combined analysis; *P*_all, p values from a combined analysis of the results from the two GWASs and the three replication panels with the use of the Cochran-Mantel-Haenszel test of disease association conditional on SNP-frequency differences among different panels; P_Het, p values of the homogeneity test of ORs from the five different panels with the use of the Breslow-Day test; SLE, systemic lupus erythematosus; TH, Thailand replication panel.

wide significance of disease association (*P*_all=4.5E-08, OR=0.86). Its interaction partner rs12876893 (*ALOX5AP*) showed suggestive association to SLE (*P*_all=6.2E-05, OR=1.12). For the other interaction pair, rs4073643 (*LHFPL3*) and rs7920305 (*Fas*), the former failed to show evidence of association based on joint analysis (*P*_all=8.1E-02, OR=1.06), while rs7920205 did not reach GWAS significance despite demonstrating the same trend in different cohorts in replication (*P*_all=7.4E-04, OR=1.1).

SNP rs2222631 is located in an intron of *CD80*, a susceptibility gene for SLE. There is moderate LD between rs2222631 and rs6804441, the reported SNP from our previous study[16] ($r^2<0.5$ in different cohorts, see online supplementary figure S1). Conditional logistic regression was performed to test independence between the newly identified SNP rs2222631 and rs6804441, both SNPs remained significant when the effect from the other SNP was adjusted for (see online supplementary table S4). The interaction between SNP rs6804441 and rs12876893 in *ALOX5AP* was also examined and no evidence of interaction was found (*P*_int=0.30, OR_int=0.96) (see online supplementary table S5).

Interaction between *CD80* (rs2222631) and *ALOX5AP* (rs12876893) and between *LHFPL3* (rs4073643) and *Fas* (rs7920305) by replication data

We tested whether the logistic model was significantly improved for the interaction between *CD80* (rs2222631) and *ALOX5AP* (rs12876893) and between *LHFPL3* (rs4073643) and *Fas* (rs7920305) using data from GWAS and the replication stage. As shown in figure 2 (see detailed information in supplementary table S6), significant joint interaction p values (*P*_int_all) were observed after adjusting for potential subpopulation differences among cohorts, although with more moderate effect sizes comparing with the results from the GWAS stage alone. For both pairs, the interactions in the Thailand cohort were not significant despite demonstrating the same trend, which might be due to the limited sample size in this cohort.

Pairwise MDR analysis was again used to test genetic interaction, incorporating samples from GWAS and the replication stage. As shown in online supplementary table S7, a significant effect for interaction ($p<0.01$) was observed in different cohorts for the *CD80* (rs2222631) and *ALOX5AP* (rs12876893) pair. Case-control comparison of different genotype combinations for the two SNPs in different cohorts was illustrated in figure 3. High-risk di-genotype (eg, AA/GG) was shown to be much higher in percentage in cases than in controls for the *CD80* (rs2222631-A) and *ALOX5AP* (rs12876893-G) pair (figure 3) across different cohorts, which is consistent with the results from logistic regression. For the *LHFPL3* (rs4073643) and *Fas* (rs7920305) pair, although the result was statistically significant across different populations (see online supplementary table S7), the case-control comparison of genotype combinations was not consistent among different cohorts (see online supplementary figure S2).

Expression of *CD80* and *ALOX5AP* in patients with SLE

Allelic association and genetic interaction pointed to the involvement of the two SNPs, rs2222631 and rs12876893 in SLE. To gain further biological insight, we examined the expression of *CD80* and *ALOX5AP* by qRT-PCR in PBMC from 96 SLE cases and 273 healthy controls, all from the Hong Kong cohort. It was showed that *ALOX5AP* expression was highly correlated with the expression of *CD80* in healthy controls, suggesting a potential functional link between these two genes

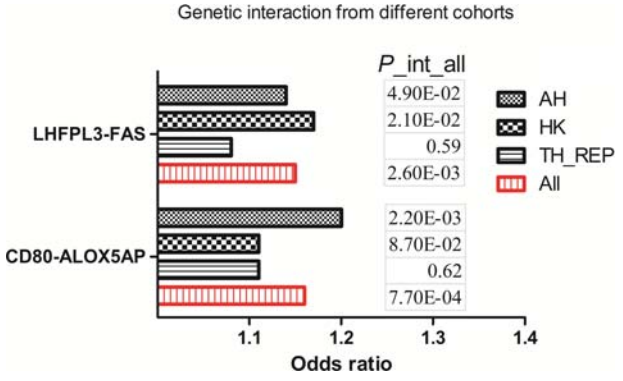

Figure 2 Genetic interaction by logistic regression from each cohort and meta-analysis. The analysis results combined the samples from genome-wide association studies (GWAS) and replication stages. The meta-analysis results were showed in red colour and the interaction patterns from different cohorts showed consistent direction (OR>1).

(Pearson's r=0.39, p<0.0001, figure 4). There is a similar but much weaker expression correlation in patients with SLE between these two genes (Pearson's r=0.16, p=0.16, figure 4).

In order to further examine whether the correlation of these two genes varies in different cell types, expression correlation of *CD80* and *ALOX5AP* was analysed across 26 different types of tissues and organs (http://www.genenetwork.org/) (see online supplementary figure S3), and good correlation was observed (Spearman's r=0.675, p=8.57E-05). Furthermore, both of these two genes were found to show high level expression in cells and tissues of the immune system (see online supplementary table S8).

Logistic regression was used to examine the *ALOX5AP* and *CD80* expression and different combination of genotypes of SNP rs12876893 (*ALOX5AP*) and rs2222631 (*CD80*) in controls and cases. There is evidence of correlation between the genotypes of rs12876893 (*ALOX5AP*) and the expression level of *CD80* in healthy controls (p=0.0056) (see online supplementary table S9). Meanwhile, the interaction between SNP rs12876893 and SNP rs2222631 might also be correlated with the expression level of *CD80* (p=0.03042) (see online supplementary table S9). However, no evidence of association was detected between the genotypes and expression of *ALOX5AP* in either cases or controls.

We also compared the gene expression level between SLE cases and controls, and significantly lower expression for *ALOX5AP* (p<0.0001) and higher expression for *CD80* were detected (p<0.0001) in SLE cases than in healthy controls (see online supplementary figure S4).

Interactions among reported SLE susceptibility variants

For the 42 SNPs from 37 known susceptibility loci for SLE (see online supplementary table S10), we tested pairwise interaction using logistic regression.[14 15 20–22] Six pairs showed meta-analysis interaction p value (P_comb_int) <0.01 and were listed in online supplementary table S11. Two pairs, rs13277113 (BLK) and rs4639966 (DDX6), and rs844648 (TNFSF4) and rs6445975 (PXK), showed consistently and statistically significant results (P_int<0.05) in both GWAS data sets. Further replication of these findings is necessary to examine the validity of these interactions, as the sample size is still small and the power is low in detecting genetic interactions.

DISCUSSION

In the present study, we have searched for pairwise genetic interactions using two GWAS data sets. To our knowledge, this is the first genome-wide analysis for genetic interaction for SLE. We identified an interaction pair between *CD80* (rs2222631) and *ALOX5AP* (rs12876893) with OR_int_all equal to 1.16 (*P_int_all*=7.70E-04) with consistent results from different cohorts.

SNP rs2222631 is located in *CD80*, which encodes a membrane receptor activated upon binding of CD28.[23] The percentage of CD80+ cells in the activated B cell subset in SLE was significantly higher than in controls.[24] Increased CD80 expression has been shown to play an important role in the development of lupus nephritis.[25] CD80 was also found to be associated with coeliac disease (rs11712165),[26] multiple sclerosis (rs1132200)[27] and primary biliary cirrhosis (rs2293370),[28] all of which have a strong autoimmune component. The SNPs reported from other diseases have low LD with the two SNPs detected for SLE (rs2222631 and rs6804441)[16] (see online supplementary figure S1).

Its interaction partner SNP rs12876893 is located in an intergenic region, downstream of *ALOX5AP* and *TEX26-AS1*. Functional annotation of the SNP and surrounding genes with

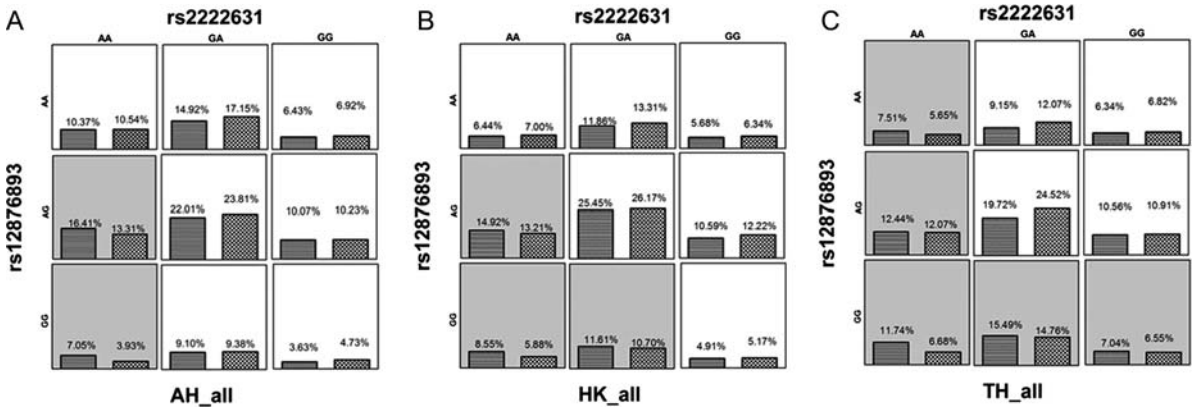

Figure 3 The optimal two-locus model as determined by multifactor dimensionality reduction analysis on variants in *CD80* and *ALOX5AP* in the three cohorts. The numbers within each small square represent the percentage of individuals with the specific genotype combinations (di-genotypes) in cases (left) and controls (right). The dark-shaded squares indicate risk for disease, whereas light shaded squares represent low risk for disease. Boxes were labelled as high risk if the ratio of percentage (labelled above each bar) of the patients to that of the controls ≥1. AH_all: combined the Anhui samples from the GWAS stage (AH_GWAS) and the replication stage (AH_REP); HK_all:combined the Hong Kong samples from the GWAS stage (HK_GWAS) and the replication stage (HK_REP). TH_all: all the Thailand samples in the replication stage (TH_REP).

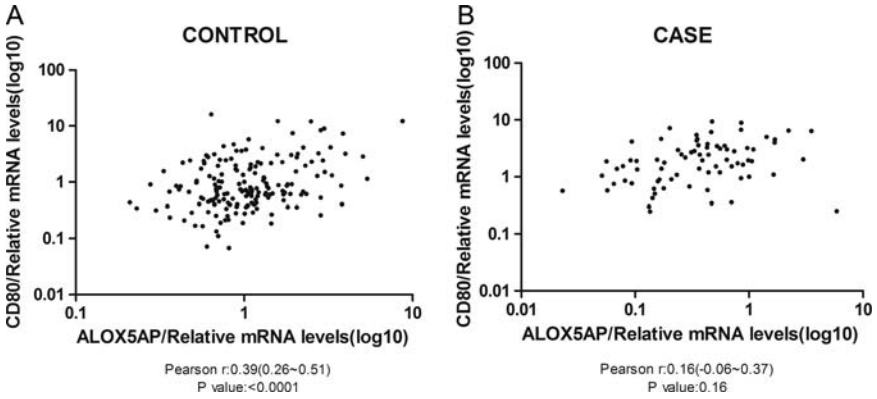

Figure 4 Expression correlation between CD80 and ALOX5AP in cases and controls. Expression correlation in controls (A) and cases (B) between *CD80* and *ALOX5AP*. Expression of the two genes was measured by TaqMan probe (Life Technologies) in peripheral blood mononuclear cells on 96 systemic lupus erythematosus cases (88 women and 8 men) and 273 healthy controls (206 women and 67 men), all from the Hong Kong cohort.

GREAT, a tool specifically designed to assign biological function to non-coding genomic regions,[29] indicated that the most probable association is with *ALOX5AP*, a gene that is involved in leukotriene synthesis. *ALOX5AP* is expressed in airway leucocytes in response to a number of stimuli including allergens and implicated in various types of inflammatory responses including asthma, arthritis and psoriasis.[30] Significantly larger effect (OR) of rs12876893 to the disease was observed in individuals carrying more risk alleles of rs2222631, and vice versa, supporting a role of interaction of these two genes in disease susceptibility (see online supplementary figure S5). However, the role of *ALOX5AP* in SLE remains speculative until functional characterisation clearly links this gene to SLE susceptibility and explains the genetic interaction observed in this study.

Abnormal expression of *CD80* and *ALOX5AP* were detected in patients with SLE compared with controls (see online supplementary figure S4). We also detected much stronger expression correlation between *CD80* and *ALOX5AP* in controls than in cases. The genotypes of interacting SNP rs12876893 (ALOX5AP) were found to correlate with the expression of *CD80*. Furthermore, the interaction of rs12876893 (*ALOX5AP*) and rs2222631 (*CD80*) was also found to correlate with the expression of *CD80* (see online supplementary table S9), consistent with the genetic interaction between these genes identified in this study. Freely accessible expression Quantitative Trait Loci (eQTL) databases were closely examined, including mRNA-by-SNP-Browser-V.1.0.1[31 32] and Genevar.[33] Consistent with our results, using the eQTL data acquired from Dimas *et al*[34] in the Genevar database, we found a marginal expression difference on *CD80* among individuals with different genotypes on SNP rs12876893 (*ALOX5AP)* (p=0.0573) in T cells (see online supplementary figure S6). Thereafter, eQTL data on circulating monocytes,[35] lymphoblastoid cell lines,[36] PBMCs,[37] sentinel dendritic cells,[38] T lymphocytes[39] and B lymphocytes[40] were further examined.[40] However, we did not find further supportive eQTL results from these analyses. This might be caused by the limited sample sizes on each individual study on eQTL. Further functional interpretations of the results from this work awaits additional studies on genetic replication and on functional characterisations of the genes and genetic variants involved, from wet laboratory experiments and genome biology data from public domains.

In addition, *LHFPL3* (rs4073643) and *FAS* (rs7920305) showed suggestive evidence for SLE association. The interaction detected by MDR showed suggestive significance. We did observe larger effect (OR) of rs4073643 to the disease in individuals carrying more risk alleles of rs7920305, and vice versa. However, the trend is not as clear as seen between *CD80* (rs2222631) and *ALOX5AP* (rs12876893) (see online supplementary figure S5). Based on the eQTL data from Dimas *et al*,[34] expression difference of Fas was observed among different genotypes of rs4073643 (p=0.0096) in lymphoblastoid cell lines but not in fibroblasts or T cells (see online supplementary figure S7).

Furthermore, we observed suggestive interactions between susceptibility SNPs rs13277113 (*BLK*) and rs4639966 (*DDX6*), and between rs844648 (*TNFSF4*) and rs6445975 (*PXK*) using logistic regression, with consistent results between the two GWAS data sets. Interestingly, there is no evidence of association for *PXK* with SLE in Asian populations.[41] The potential interaction between *TNFSF4* and *PXK* might suggest a role of this gene in SLE in Asians despite showing no or weaker main effect.

Several studies using candidate gene approach have identified genetic interactions for SLE. We thus examined evidence of these interactions in our GWAS data. As shown in online supplementary table S12, only the interaction between *STAT4* (rs7574865) and *IRF5* (rs2004640) was nominally significant in both GWAS data sets ($P\_int=0.015$). Population differences and detection power might be the explanations for these differences.

Instead of genetic variants between genes, interaction can also be found within a single gene leading to non-additive effects between two or more SNPs. As shown in our previous GWAS report,[14 42] two SNPs in moderate LD ($r^2=0.68$) located on *ETS1* showed significant synergistic interaction (p=0.0001). In the current study, though we did not limit the examination to interaction between different genes, no prominent interaction within a single gene was detected.

The current study might also suffer from several limitations. First, attribution of the genes involved based on SNP position and limited knowledge of gene function has its own flaws. As reported by the ENCODE project, although majority of the regulatory elements are located close to the genes, distant regulatory elements are not uncommon (49). It is possible that interactions among certain SNPs may indicate functional interplay between genes that are not necessarily closest to these SNPs. Using non-healthy individuals as controls in the Hong Kong GWAS in the discovery stage also deserves additional scrutiny. We have systematically analysed the controls used in this stage from different sources and removed the SNPs that showed significant differences in allele frequencies from any one of those sources. Furthermore, we have used two independent GWAS studies to cross-validate the pursued SNP pairs and replicated the interacting pairs in independent cohorts in the replication panel, with each having its own cases and matched healthy controls.

In summary, we used logistic regression and MDR to identify novel genetic interactions for SLE using meta-analysis of two

GWAS data on Han Chinese populations. The interaction between rs2222631 (*CD80*) and rs12876893 (*ALOX5AP*) was validated through replication in different cohorts. Interaction between rs7920305 (*Fas*) and rs4073643 (*LHFPL3*) was suggested but additional studies would be required for its validation. We also showed evidence on gene expression correlation between *CD80* and *ALOX5AP* but further work is also needed to elucidate the functional mechanism(s) of the genetic interaction.

Author affiliations
[1]Department of Paediatrics and Adolescent Medicine, LKS Faculty of Medicine, Queen Mary Hospital, The University of Hong Kong, Hong Kong, Hong Kong
[2]State Key Laboratory Incubation Base of Dermatology, Key Laboratory of Dermatology, Anhui Medical University, Ministry of Education, Hefei, Anhui, China
[3]Lupus Research Unit, Department of Microbiology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand
[4]Department of Epidemiology and Biostatistics, School of Public Health, Anhui Medical University, Hefei, China
[5]Department of Medicine, Queen Mary Hospital, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong
[6]Department of Medicine, Queen Elizabeth Hospital, Hong Kong, Hong Kong
[7]Department of Medicine, Tuen Mun Hospital, New Territory, Hong Kong, Hong Kong
[8]Department of Paediatrics and Adolescent Medicine, Tuen Mun Hospital, Hong Kong, Hong Kong
[9]Department of Medicine, Pamela Youde Nethersole Eastern Hospital, Hong Kong, Hong Kong
[10]Department of Medicine, Princess Margaret Hospital, Hong Kong, Hong Kong
[11]Department of Paediatrics and Adolescent Medicine, Princess Margaret Hospital, Hong Kong, Hong Kong
[12]Department of Rheumatology, Anhui Provincial Hospital, Hefei, China
[13]Department of Medicine, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand
[14]Department of Pediatrics, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand
[15]Department of Pathology, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong
[16]Department of Surgery, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong
[17]Department of Psychiatry, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong
[18]LKS Faculty of Medicine, Centre for Genomic Sciences, The University of Hong Kong, Hong Kong, Hong Kong

Acknowledgements The authors thank Winnie Lau and her team for collection of samples and clinical records for Hong Kong patients. WY and YLL thank the Research Grant Council of Hong Kong (GRF 17125114, HKU783813M, HKU781709M, HKU 784611M, and HKU 770411M) for support. The authors also thank the S K Yee Medical Foundation general award (to BH-YC, YLL and WY) for support. YZ was supported by Post-doctoral Fellow/Research Assistant Professor Scheme of the University of Hong Kong and Centre for Genomic Sciences of LKS Faculty of Medicine, HKU.

Contributors Conceived and designed the experiments: WY, YLL and YZ. Performed the experiments: YZ, JY and JZ. Analysed the data: YZ, LZ and YW. Contributed samples: DQ-Y, NH, H-FP, CCM, TMC, RWSW, KWL, MYM, SNW, AMHL, X-PL, YA, C-MW, TLL, MHKH, PPWL, WHSW, IOLN, CSL, PCS, WY and YLL. Wrote the paper: YZ and WY.

Funding This study was partially supported by the generous donation from Shun Tak District Min Yuen Tong of Hong Kong. WY and YLL received support from the Research Grant Council of Hong Kong (GRF 17125114, HKU783813M, HKU781709M, HKU 784611M and HKU 770411M). The authors also received support from a S K Yee Medical Foundation general award (to BH-YC, YLL and WY). YZ was supported by a Post-doctoral Fellow/Research Assistant Professor Scheme of the University of Hong Kong and Centre for Genomic Sciences of LKS Faculty of Medicine, HKU.

Competing interests None.

Patient consent Obtained.

Ethics approval Corresponding institutional review boards where the samples were collected.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Jarvinen P, Kaprio J, Makitalo R, *et al*. Systemic lupus erythematosus and related systemic diseases in a nationwide twin cohort: an increased prevalence of disease in MZ twins and concordance of disease features. *J Intern Med* 1992;231:67–72.
2 Grennan DM, Parfitt A, Manolios N, *et al*. Family and twin studies in systemic lupus erythematosus. *Dis Markers* 1997;13:93–8.
3 Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet* 2009;10:392–404.
4 McKinney BA, Reif DM, Ritchie MD, *et al*. Machine learning for detecting gene-gene interactions: a review. *Appl Bioinformatics* 2006;5:77–88.
5 Moore JH, Williams SM. New strategies for identifying gene-gene interactions in hypertension. *Ann Med* 2002;34:88–95.
6 Hahn LW, Ritchie MD, Moore JH. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics* 2003;19:376–82.
7 Cho YM, Ritchie MD, Moore JH, *et al*. Multifactor-dimensionality reduction shows a two-locus interaction associated with Type 2 diabetes mellitus. *Diabetologia* 2004;47:549–54.
8 Julia A, Moore J, Miquel L, *et al*. Identification of a two-loci epistatic interaction associated with susceptibility to rheumatoid arthritis through reverse engineering and multifactor dimensionality reduction. *Genomics* 2007;90:6–13.
9 Hughes T, Adler A, Kelly JA, *et al*. Evidence for gene-gene epistatic interactions among susceptibility loci for systemic lupus erythematosus. *Arthritis Rheum* 2012;64:485–92.
10 Leng RX, Wang W, Cen H, *et al*. Gene-gene and gene-sex epistatic interactions of MiR146a, IRF5, IKZF1, ETS1 and IL21 in systemic lupus erythematosus. *PLoS One* 2012;7:e51090.
11 Zhou XJ, Lu XL, Nath SK, *et al*. Gene-gene interaction of BLK, TNFSF4, TRAF1, TNFAIP3, and REL in systemic lupus erythematosus. *Arthritis Rheum* 2012;64:222–31.
12 Zhou XJ, Qi YY, Cheng FJ, *et al*. Genetic interactions between BANK1 and BLK in Chinese patients with systemic lupus erythematosus. *J Rheumatol* 2013;40:1772–3.
13 Dang J, Shan S, Li J, *et al*. Gene-gene Interactions of IRF5, STAT4, IKZF1 and ETS1 in systemic lupus erythematosus. *Tissue Antigens* 2014;83:401–8.
14 Yang W, Shen N, Ye DQ, *et al*. Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. *PLoS Genet* 2010;6:e1000841.
15 Han JW, Zheng HF, Cui Y, *et al*. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet* 2009;41:1234–7.
16 Yang W, Tang H, Zhang Y, *et al*. Meta-analysis followed by replication identifies loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as associated with systemic lupus erythematosus in Asians. *Am J Hum Genet* 2013;92:41–51.
17 Purcell S, Neale B, Todd-Brown K, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
18 Benjamini Y, Hochberg Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* 1995;57:289–300.
19 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;26:2190–1.
20 Yang J, Yang W, Hirankarn N, *et al*. ELF1 is associated with systemic lupus erythematosus in Asian populations. *Hum Mol Genet* 2011;20:601–7.
21 Harley JB, Alarcon-Riquelme ME, Criswell LA, *et al*. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat Genet* 2008;40:204–10.
22 Gateva V, Sandling JK, Hom G, *et al*. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nat Genet* 2009;41:1228–33.
23 Chen L, Flies DB. Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat Rev Immunol* 2013;13:227–42.
24 Folzenlogen D, Hofer MF, Leung DY, *et al*. Analysis of CD80 and CD86 expression on peripheral blood B lymphocytes reveals increased expression of CD86 in lupus patients. *Clin Immunol Immunopathol* 1997;83:199–204.
25 Sui MS, Zhou J, Jia XB, *et al*. [Expression and significance of CD80/CD86 in renal tissue of lupus nephritis]. *Zhonghua Nei Ke Za Zhi* 2010;49:691–5.
26 Dubois PC, Trynka G, Franke L, *et al*. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 2010;42:295–302.
27 Varade J, Comabella M, Ortiz MA, *et al*. Replication study of 10 genes showing evidence for association with multiple sclerosis: validation of TMEM39A, IL12B and CBLB [correction of CLBL] genes. *Mult Scler* 2012;18:959–65.
28 Nakamura M, Nishida N, Kawashima M, *et al*. Genome-wide association study identifies TNFSF15 and POU2AF1 as susceptibility loci for primary biliary cirrhosis in the Japanese population. *Am J Hum Genet* 2012;91:721–8.

29 McLean CY, Bristor D, Hiller M, *et al*. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010;28:495–501.
30 Strom JO, Strid T, Hammarstrom S. Disruption of the alox5ap gene ameliorates focal ischemic stroke: possible consequence of impaired leukotriene biosynthesis. *BMC Neurosci* 2012;13:146.
31 Dixon AL, Liang L, Moffatt MF, *et al*. A genome-wide association study of global gene expression. *Nat Genet* 2007;39:1202–7.
32 Moffatt MF, Kabesch M, Liang L, *et al*. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007;448:470–3.
33 Yang TP, Beazley C, Montgomery SB, *et al*. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* 2010;26:2474–6.
34 Dimas AS, Deutsch S, Stranger BE, *et al*. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 2009;325:1246–50.
35 Rotival M, Zeller T, Wild PS, *et al*. Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. *PLoS Genet* 2011;7:e1002367.
36 Lappalainen T, Sammeth M, Friedlander MR, *et al*. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 2013;501:506–11.
37 Westra HJ, Peters MJ, Esko T, *et al*. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013;45:1238–43.
38 Lee MN, Ye C, Villani AC, *et al*. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science* 2014;343:1246980.
39 Raj T, Rothamel K, Mostafavi S, *et al*. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* 2014;344:519–23.
40 Fairfax BP, Makino S, Radhakrishnan J, *et al*. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet* 2012;44:502–10.
41 Yang W, Ng P, Zhao M, *et al*. Population differences in SLE susceptibility genes: STAT4 and BLK, but not PXK, are associated with systemic lupus erythematosus in Hong Kong Chinese. *Genes Immun* 2009;10:219–26.
42 Zhang J, Zhang Y, Zhang L, *et al*. Epistatic interaction between genetic variants in susceptibility gene ETS1 correlates with IL-17 levels in SLE patients. *Ann Hum Genet* 2013;77:344–50.
43 Blake JA, Bult CJ, Eppig JT, *et al*. Mouse Genome Database Group. The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Res* 2014(Database issue);42:D810–817.