Three SNPs in chromosome 11q23.3 are independently associated with systemic lupus erythematosus in Asians

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

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

Systemic lupus erythematosus (SLE) has a complex etiology and is affected by both genetic and environmental factors. Although more than 40 loci have shown robust association with SLE, the details of these loci, such as the independent contributors and the genes involved, are still unclear. In this study, we performed meta-analysis of two existing genome-wide association studies (GWASs) on Chinese Han populations from Hong Kong and Anhui, China, and followed the findings by further replication on three additional Chinese and Thailand cohorts with a total of 4254 cases and 6262 controls matched geographically and ethnically. We discovered multiple susceptibility variants for SLE in the 11q23.3 region, including variants in/near *PHLDB1* (rs11603023, $P_{\text{combined}} = 1.25E-08$, OR = 1.20), *DDX6* (rs638893, $P_{\text{combined}} = 5.19E-07$, OR = 1.22) and *CXCR5* (rs10892301, $P_{\text{combined}} = 2.51E-08$, OR = 0.85). Genetic contributions from the newly identified variants were all independent of SNP rs4639966, whose association was reported from the previous GWAS. In addition, the three newly identified variants all showed independent association with the disease through modeling by

*To whom correspondence should be addressed at: Department of Paediatrics and Adolescent Medicine, Li Ka Shing Faculty of Medicine and Queen Mary Hospital, The University of Hong Kong, Hong Kong, China. Tel: +852 22554481; Fax: +852 28551523; Email: lauylung@hkucc.hku.hk (Y. L. L.); Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, Hong Kong, China. Tel: +852 283 15077; Fax: +852 281 98142; Email: yangwl@hkucc.hku.hk (W. Y.)

both stepwise and conditional logistic regression. The presence of multiple independent variants in this region emphasizes its role in SLE susceptibility, and also hints the possibility that distinct biological mechanisms might be involved in the disease involving this genomic region.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a heterogeneous disease characterized by various aberrations in the immune system, typically leading to the formation of autoantibodies against nucleic acid and associated proteins (1). Both cellular and hormonal regulatory mechanisms fail to terminate the process, while inflammation and tissue injury lead to amplification and perpetuation of end-organ damage. Due to tissue damage caused by immune complex deposition, SLE usually presents with a diverse spectrum of clinical symptoms, including skin rash, anti-double stranded DNA (dsDNA) autoantibody production, renal involvement and other manifestations (2).

SLE has a complex etiology, affected by both genetic and environmental factors. The disease concordance rate is between 20 and 59 for monozygotic twins, and the risk for siblings of affected individuals is 30 times higher than that for the general population, indicating important genetic contributions (3,4). The disease prevalence also varies with ethnicity, being higher in non-European populations (approximately 1:500 in populations with African ancestry versus. 1:2500 in Northern Europeans) (5). In Hong Kong, the prevalence and annual incidence of SLE are estimated to be 0.1% and 6.7 cases/100 000 individuals, respectively (6). In addition to genetic effects, environmental factors such as smoking and exposure to ultraviolet light have also been implicated in the disease (7). The possibility that viruses may trigger SLE has also been speculated in the past, with observations of faster seroconversion to Epstein-Barr virus (EBV) infection (8) and higher viral load (9) in individuals affected with SLE than in normal subjects. In addition, the molecular similarity between EBV nuclear antigen and the common lupus autoantigen Ro also suggested potential roles of virus infection in SLE pathogenesis (10).

Since the application of genome-wide association studies (GWASs) on complex traits, significant progress has been made in identifying novel susceptibility loci for SLE (11–27). However, the genetic basis of SLE still remained largely undetermined. So far more than 40 loci were demonstrated to have robust association with the disease, yet they only explain a small proportion of the disease heritability (28). Even among those established loci, there still exists much ambiguity. One of the possible reasons is that for most GWASs, only the most significant SNP from a locus is followed up and eventually established, while in reality there could be multiple variants independently associated with the underlying disease in a given region. The independent variants are often overlooked and this may result in underestimation of explained disease heritability. A recent study carried out by Ke (29) demonstrated the presence of those multiple independent effects in loci of several other complex diseases, and found that the variance explained by the multiple effects in a locus was much higher than the variance explained by the single SNP reported. In addition, a recent study by our own group reported five loci to be associated with SLE susceptibility, and for two of them, multiple independent variants were confirmed (30). Thus, further exploration of those

independent effects will not only enrich the list of SLE susceptibility genes, but also has a potential to advance our understanding on the etiology of this complicated disease.

The 11q23.3 genomic region has attracted much attention for its role in several complex diseases. Multiple genes reside in this region spanning 5.3 Mb, including *PHLDB1* (pleckstrin homology-like domain, family B, member 1), *DDX6* (DEAD (Asp-Glu-Ala-Asp) box helicase 6), *TREH* (trehalase, brush-border membrane glycoprotein) and *CXCR5* (chemokine (C-X-C motif) receptor 5). SNP rs498872, which was located in the upstream of *PHLDB1*, was found to be associated with glioma (31). SNP rs6421571, a variant ~10 kb upstream of *CXCR5,* showed close association with primary billiary cirrhosis (PBC) (32), and another SNP located in the promoter region of *CXCR5*, rs630923, was reported as associated with multiple sclerosis (MS) (33). SNP rs10892279, located in the promoter region of *DDX6* gene, showed significant association with both celiac disease (CD) and rheumatoid arthritis (RA) (34). Two studies on SLE also identified signals in this region. SNP rs503425, a variant located between *DDX6* and *CXCR5*, showed suggestive association with SLE in US-Swedish cohorts ($P_{combined} = 3.30E-06$) (19). Another SNP, rs4639966, located downstream of *DDX6*, was established as associated with SLE in our previous work on a Chinese Han population from Anhui, China ($P_{combined} = 1.30E-16$) (35). However, due to the high gene density in this region, many details still remain elusive. Moreover, it is not clear whether these association signals point to a single biological pathway or a single gene, or they have different underlying biological implications. A more detailed elucidation of genetic variants in this region may help us better understand these issues.

In this study, we focused on exploring independent susceptibility variants for SLE in the 11q23.3 region and the relationship of these variants with those identified for SLE in different populations and from studies of other diseases. Through meta-analysis of two existing GWASs on Chinese Han populations from Hong Kong and Anhui, and further replication on three additional Chinese and Thailand cohorts with a total of 4254 cases and 6262 controls matched geographically and ethnically, we identified multiple independent variants associated with SLE in the 11q23.3 region, and showed that they had little linkage disequilibrium (LD) with variants found in previous studies, suggesting complex biological mechanisms for the involvement of this region in disease pathogenesis.

RESULTS

Meta-analysis of two existing GWASs on Chinese Han population from Hong Kong and Anhui

First, meta-analysis on genotyped SNPs was performed by METAL (36), using effective sample size as weight from the two GWASs (940 and 1110 for the Hong Kong and Anhui studies, respectively) (35,37,38). The two Chinese populations did not seem to have significant difference in terms of population

structure as measured by principal component analysis (Supplementary Material, Fig. S1), and a modest genomic inflation factor of 1.04 was also observed from meta-analysis (30). We examined meta-analysis results in 11q23.3 and there were 1072 SNPs interrogated in this 5.3 Mb region, from 115905730 bp to 121181848bp (GRCh37/hg19). Based on meta-analysis *P*-values, 10 SNPs showed $P_{meta} < 0.001$, aggregating in a 250 Kb span (Table 1 and Fig. 1). Further investigation of LD patterns among those SNPs revealed that they can be grouped into four clusters (Supplementary Material, Fig. S2A and B). Therefore, three SNPs, SNP rs11603023, an intronic variant in *PHLDB1*, SNP rs638893, which is located ∼30 kb upstream of *DDX6*, and SNP rs10892301, a variant ∼20 kb upstream of *CXCR5,* together with rs4639966, the SNP previously reported in the Anhui GWAS (35), were selected for further replication, to identify independent variants associated with the disease.

In order to extensively explore this region, we also performed imputation on the two Chinese populations to obtain genotypes of additional SNPs in 11q23.3. Meta-analysis of the imputed results was also performed using METAL (Supplementary Material, Table S1, shown as P_{meta}). The results showed that a total of 169 imputed SNPs reached meta-analysis *P*-value <0.001. Although many imputed SNPs have comparable meta-analysis *P*-values as those interrogated in the two GWASs and eventually replicated SNPs (such as rs10892301), few showed consistent results between Hong Kong and Anhui datasets, and the ones having a nominal *P* < 0.05 in both cohorts (highlighted in Supplementary Material, Table S2), all have high LD with the interrogated SNPs ($r^2 > 0.96$, Supplementary Material, Table S2). Functionally, none of these SNPs result in amino acid substitutions (Supplementary Material, Table S1, SCAN annotation). RegulomeDB (40) annotations were also investigated to identify SNPs in or near regulatory elements, and several such imputed SNPs were found and they are in high LD with the genotyped SNPs ($r^2 \geq 0.8$) (Supplementary Material, Table S3). Although there is no evidence that these regulatory SNPs may be contributing to disease pathogenesis (they did not show better *P*-values and there are quite a few of them that are in high LD with each other), this analysis may provide information for future studies for understanding this region biologically.

Replication experiment

Further replication for selected SNPs were performed by a TaqMan SNP genotyping method using assay-on-demand probes and primers (Applied Biosystems) on the remaining samples from the Hong Kong cohort that were not included in the GWAS stage, samples collected from Bangkok, Thailand and samples from Anhui, China, that are independent of the samples from the Anhui GWAS cohort. As shown in Table 2, both rs11603023 and rs10892301 showed genome-wide significance in association with SLE by joint analysis of data from all these cohorts ($P_{combined} = 1.25E-08$ and $2.51E-08$, respectively), while SNP rs638893 had suggestive association with a $P_{combined}$ value of $5.19E-07$. The same trends were observed in all the cohorts for the three SNPs (Table 2), and testing of between-population heterogeneity of odds ratios (ORs) by the Breslow–Day method (41) did not show significant difference between cohorts ($P_{het} > 0.09$, Table 2).

Table 1. Meta-analysis results for genotyped SNPs in the 11q23.3 region with *P*-values <0.001

Marker	Gene symbol	Allele	P_{meta}	I^2	HK_GWAS Case (freq)	Ctrl (freq)	OR (95%CI)	P_{HKGWAS}	Case (freq)	Ctrl (freq)	AH_GWAS P_{AHGWAS}	OR (95% CI)
rs11603023	PHLDB1	T<C	6.81E−04	0	0.36	0.33	1.15 (1.01–1.31)	3.86E−02	0.26	0.22	6.67E−03	1.21 (1.05–1.39)
rs2077173	PHLDB1	T<C	9.09E−04	0	0.33	0.37	0.84 (0.73–0.96)	8.95E−03	0.42	0.45	3.55E−02	0.88 (0.78–0.99)
rs633683	PHLDB1	T<C	9.06E−04	0	0.36	0.33	1.15 (1.01–1.31)	4.22E−02	0.27	0.23	8.29E−03	1.20 (1.05–1.38)
rs494560	PHLDB1	A<G	3.51E−04	0	0.36	0.32	1.16 (1.01–1.32)	3.06E−02	0.26	0.22	4.12E−03	1.22 (1.07–1.40)
rs12798453	PHLDB1	G<T	9.84E−04	0	0.32	0.36	0.86 (0.75–0.98)	2.44E−02	0.41	0.45	1.61E−02	0.86 (0.77–0.97)
rs573905	DDX6	A<G	5.59E−05	48	0.45	0.42	1.12 (0.98–1.27)	8.81E−02	0.34	0.29	9.37E−05	1.29 (1.13–1.46)
rs4639966	DDX6	C<T	4.77E−05	57	0.45	0.42	1.11 (0.98–1.26)	1.01E−01	0.34	0.28	5.85E−05	1.30 (1.14–1.47)
rs638893	DDX6	C<T	5.27E−05	17	0.21	0.18	1.17 (1.00–1.37)	5.33E−02	0.14	0.10	2.02E−04	1.40 (1.17–1.68)
rs7117594	DDX6	T<C	1.54E−04	0	0.22	0.20	1.17 (1.00–1.36)	5.14E−02	0.16	0.12	8.09E−04	1.34 (1.13–1.58)
rs10892301	CXCR5	G<A	3.69E−05	0	0.39	0.44	0.82 (0.72–0.93)	2.40E−03	0.51	0.55	4.92E−03	0.84 (0.75–0.95)

Case/Ctrl (freq) indicates minor allele frequency of the SNP in cases or controls, the calculation of OR is also based on the minor allele of each SNP. Risk allele of each SNP is underlined.
I^2: heterogeneity index (range 0–100).
The nearby representative gene for each SNP was determined by LD block containing the corresponding SNP based on the data from HapMap3 Release 2 (CHB + JPT population).
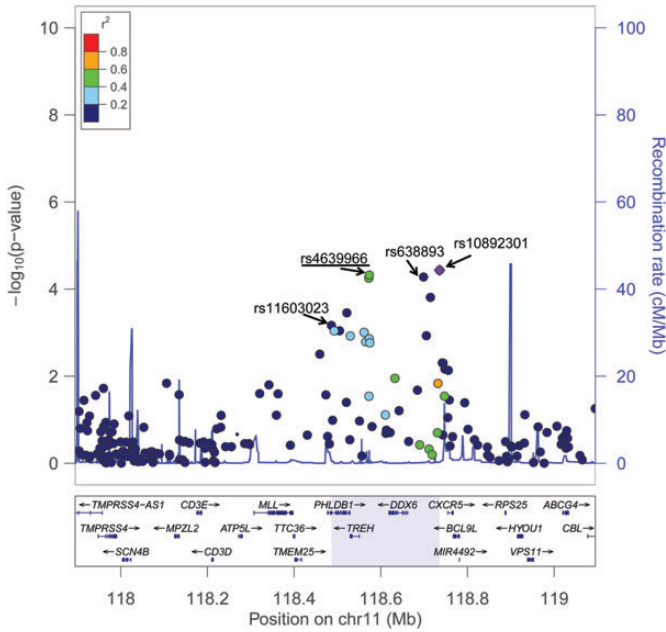
Figure 1. Regional association plot for SNPs in the 11q23.3 region. SNPs are plotted by the chromosomal position (GRCh37/hg19; *x* axis) and association with SLE from the GWAS ($-\log_{10}P$-value; *y*-axis). The strongest signal from SNP rs10892301 is labeled in the plot, and other SNPs are colored to reflect their LD with the top SNP (based on the pairwise r^2 values from the meta-analysis data). Estimated recombination rates (from hg19/1000 Genomes Mar 2012 ASN) are plotted to reflect the local LD structure. Genes are indicated in the lower panel of the plot. The four selected SNPs for replication are labeled by arrows, including rs4639966 reported in the previous GWAS. The figure was generated using LocusZoom (39).

Independence test and estimation of variance in liability explained (Vg)

A stepwise logistic regression analysis was performed to test independence of these SNPs. The method begins with an empty model, to which the variables are added one by one. This analysis showed that rs11603023 exhibited the strongest association with SLE. Subsequent additions of rs10892301 and rs638893 also significantly improved the model, providing further evidence to support the notion that multiple variants in the 11q23.3 region are independently associated with the disease (Table 3). Further addition of rs4639966 did not show significant improvement in this model ($P = 0.306$), which might be explained by its relatively small effect size in the current samples. The independent contributions of the identified SNPs were also supported by an increase in OR values in the analysis of each individual cohort (Supplementary Material, Table S4).

Conditional logistic regression analysis was also performed to investigate the independent effects among these replicated SNPs. The result showed that the newly identified SNPs (rs11603023, rs10892301 and rs638893) were all independent of the reported SNP rs4639966, since all of them remained significant while the effect from rs4639966 was adjusted (shown in Table 4, last column). In order to find residual association after each step of the analysis, we also showed the *P*-values of these SNPs after adjusting for the effect from (i) the most significant SNP (rs11603023); (ii) rs11603023 plus the variant with the highest residual association (rs10892301) and (iii) rs11603023,

Table 2. Association results of the identified SNPs from each cohort and joint analysis

SNP	Hong Kong (1611/3219) Case (freq)	Ctrl (freq)	P_{HK}	An Hui (2331/2333) Case (freq)	Ctrl (freq)	P_{AH}	Thailand (312/710) Case (freq)	Ctrl (freq)	P_{Thai}	$P_{combined}$	OR$_{combined}$ (95% CI)	P_{het}	Vg (%)
rs11603023	0.36	0.32	3.79E−05	0.25	0.23	3.46E−03	0.43	0.36	2.99E−03	1.25E−08	1.20 (1.12–1.27)	0.3826	0.12
rs638893	0.20	0.18	6.62E−03	0.14	0.11	2.11E−04	0.20	0.16	1.39E−02	5.19E−07	1.22 (1.13–1.32)	0.3775	0.09
rs10892301	0.41	0.45	1.13E−04	0.51	0.54	4.10E−03	0.36	0.44	3.55E−04	2.51E−08	0.85 (0.80–0.90)	0.08699	0.12
rs4639966	0.45	0.41	1.21E−03	0.33	0.31	2.49E−02	0.46	0.41	3.76E−02	1.48E−05	1.14 (1.07–1.20)	0.5868	0.07

Arabic numerals in brackets indicate the number of cases and controls that are used in this study.
Case/Ctrl (freq) indicates the minor allele frequency of the SNPs in cases or controls.
Vg value indicates variance in liability explained by individual genetic variant.

rs10892301 and the variant with the next highest residual association (rs638893). The results support independent contributions of the three newly identified SNPs to SLE susceptibility (Table 4). Consistent with the independence test results, these SNPs had low LD among each other ($r^2 < 0.2$, Supplementary Material, Fig. S2A and B).

With the presence of multiple independent effects in this region, the proportion of heritability explained by this region would probably increase. For the assessment of such increase, the contribution of individual variant to disease heritability was estimated with an algorithm developed by So *et al*. (28) under a disease-liability model. The results showed that the heritability explained by the three independent SNPs was much higher than that explained by the previously reported SNP rs4639966 (Table 2, shown as Vg).

In addition, haplotype analysis indicated that TCA haplotype formed by the three SNPs, rs11603023, rs4639966 and rs10892301, was the major risk haplotype, while the CTG haplotype was the major protective haplotype in all the three cohorts (Supplementary Material, Table S5). The same trend was observed when another SNP with suggestive significance, rs638893, was included in the haplotype analysis. However, probably due to the lower allele frequency of rs638893 (18, 11 and 16% in HK, AH and Thai controls, respectively), including this SNP did not increase significance in disease association (Supplementary Material, Table S5).

Sub-phenotype analysis

A case-only analysis was performed on the Hong Kong cohort to detect potential genetic association of the variants with specific subphenotypes. As presented in Table 5, SNP rs11603023 showed significant association with anti-dsDNA antibody production (OR = 1.32, $P_{corrected}$ = 0.0153), and SNP rs10892301 showed marginal significance on early-onset cases versus those with onset age $\geq$ 18 years (OR = 1.29, uncorrected P = 0.04935), although the P-values did not withstand correction for multiple testing by 10 000 permutations. For the other 22 sub-phenotypes tested, no significant association was observed and further confirmation from additional cohorts would be needed.

SNPs in this region associated with other complex diseases

For the purpose of finding potentially shared susceptible variants/pathways between SLE and other immune-related diseases, we examined our meta-analysis data on SNPs in this region which showed association with other immune-related diseases. Results suggested that none of them showed strong evidence of association with SLE in these Asian populations (Table 6 and Supplementary Material, Table S6), and these SNPs are also in low LD with the four replicated SNPs in the current study and previous GWASs (Supplementary Material, Fig. S2C and D) pointing to potential differences in the biological mechanisms behind these association signals. Our comparison also suggests differences in genetic variants in SLE association between different populations. Of note, based on imputation data, rs503425, an SNP with suggestive significance for SLE association in the US–Swedish cohorts (19) ($P_{combined}$ = 3.30E−06) did not show evidence of association in the two Chinese cohorts (P_{meta} = 0.3059), and this SNP have high allele frequencies in both populations (36% in Asians and 20% in Europeans). Thus, study power is unlikely to be the explanation for this population difference in association signals.

DISCUSSION

In this study, through meta-analysis of two existing GWASs on Chinese Han populations, and further replication on three additional Chinese and Thailand cohorts, with a total number of 4254 cases and 6262 controls matched geographically and ethnically, we have identified multiple independent variants associated with SLE in the 11q23.3 region. Among those variants, an intronic SNP in *PHLDB1*, rs11603023, demonstrated the most significant association with SLE ($P_{combined}$ = 1.25E−08, OR$_{combined}$ = 1.20). SNP rs10892301, which was located upstream of *CXCR5*, showed genome-wide significance as well

Table 3. Independent effects among the identified SNPs in the 11q23.3 region. (a) Stepwise logistic regression of three SNPs model

SNP added to the model	*P*-value	OR (95% CI)
rs11603023	1.55E−08	1.19 (1.12–1.27)
rs10892301	2.20E−04	0.89 (0.84–0.95)
rs638893	0.02997	0.91 (0.83–0.99)

Variable not in the equation: rs4639966, *P* = 0.306.

Table 4. Independent effects among the identified SNPs in the 11q23.3 region. (b) Conditional logistic regression analysis

| | Adjusted *P*-value and OR | SNP(s) whose effect is adjusted for | | | |
		rs11603023	rs11603023, rs10892301	rs11603023, rs10892301, rs638893	rs4639966
rs11603023	*P*				3.50E−03
	OR (95% CI)				1.04 (1.02–1.07)
rs10892301	*P*	9.05E−04			4.42E−03
	OR (95% CI)	0.96 (0.93–0.98)			0.96 (0.94–0.99)
rs638893	*P*	2.58E−03	3.48E−02		1.41E−04
	OR (95% CI)	0.95 (0.91–0.98)	0.96 (0.92–1.00)		0.93 (0.89–0.97)
rs4639966	*P*	0.5438	0.3891	0.5039	
	OR (95% CI)	0.99 (0.96–1.02)	1.01 (0.98–1.04)	1.01 (0.98–1.04)	

Independence test was performed using the combined dataset, which includes 1611 cases and 3219 controls from the HK cohort, 2331 cases and 2333 controls from the Anhui cohort, and 312 cases and 710 controls from the Thailand cohort. Stepwise logistic regression was done by SPSS 16.0, and conditional logistic analysis was performed by SNPTEST v2.4.1, both treating potential difference among cohorts as a covariate.

Table 5. Association of SLE risk loci analyzed by subphenotype stratification in HK SLE cases

| | Anti-dsDNA antibody | | | | Early onset age | | | |
	+ (freq)	− (freq)	Nominal P (P_corrected)	OR (95% CI)	≤18 years (freq)	>18 years (freq)	Nominal P (P_corrected)	OR (95% CI)
rs11603023	0.37	0.31	0.003254 (*0.0153)	1.32 (1.09−1.60)	0.35	0.35	0.8961 (1)	0.99 (0.75−1.30)
rs10892301	0.41	0.43	0.3009 (0.76)	0.91 (0.76−1.09)	0.36	0.42	0.0485 (0.1954)	0.76 (0.58−1.00)

822 HK SLE cases showed dsDNA-positive manifestation, while 345 HK SLE cases showed dsDNA-negative manifestation. 122 HK SLE cases developed the disease on or before 18 years. old, and 1009 HK SLE cases developed the disease after 18 years.
P_corrected is the value that is corrected for multiple comparisons by controlling the familywise type I error rate (FWER) using 10 000 permutations.

(P_combined $= 2.51E-08$, OR_combined $= 0.85$). In addition, another SNP, rs638893 in the upstream of *DDX6,* showed suggestive association with SLE susceptibility (P_combined $= 5.19E-07$, OR_combined $= 1.22$), the validity of this association may need further replication in additional cohorts in future studies.

The independent effects found for the three novel variants in this region were supported by strong evidence. Based on our results, all the three SNPs remained significant while the effect of rs4639966 was adjusted, and independent contributions also existed among the three novel variants (Tables 3 and 4). These three SNPs, together with the previously established SNP rs4639966, increased the proportion of heritability explained by this region. Our current study also highlighted the complexity of this region. Although many GWAS hits were aggregated in this region, none of them (or proxies of them based on LD information from Asian populations or European populations, $r^2 \geq$ 0.8) showed strong evidence of association with SLE by the current investigation, and they also have little LD between each other. Interestingly, SNP rs638893, which revealed suggestive association with SLE susceptibility in the current study, was recently demonstrated as associated with vitiligo, an autoimmune disease presented with progressive loss of skin pigmentation, in a Chinese Han population (P_combined $= 2.47E-09$, OR = 1.22) (42). The same susceptible variant for both SLE and vitiligo in this locus may imply a potentially shared biological pathway for the two autoimmune diseases. Interestingly, rs503425, a SNP showed association with SLE in a study of European populations, did not show evidence in our meta-analysis (P_meta = 0.3059), hinting potential differences between major ethnicities even the same region was found to be associated with the same disease, as demonstrated on several loci in our previous study (30).

Obviously, the exploration of independently contributing variants from this region is mainly based on the meta-analysis data and on replication of a limited number of variants and is by no means complete. For example, two SNPs in high LD with each other in *PHLDB1,* SNP rs2077173 and SNP rs12798453, showed suggestive association signals with SLE (P_meta = $9.10E-04$ and $9.84E-04$, respectively). These two SNPs do not have high LD with the newly identified SNP rs11603023 or the other three SNPs shown in Table 2 ($r^2 < 0.3$) and whether this association is valid remains to be determined. Therefore, it is possible that even more variants may be independently associated with SLE in this region of high interest and identification of their roles requires further investigation in the future. It is also realized that the power of detecting independent signals is much smaller than focusing on a top signal from a single locus, thus much larger sample size is needed. In addition,

there still exists certain possibility that currently identified susceptibility variants for SLE in the 11q23.3 region are just the tag-SNPs that have high LD with an unknown functional variant, although imputation results did not reveal consistent and clear signals that may better explain the association seen in this region (Supplementary Material, Tables S1−S3). The imputed SNPs either showed differences in association between Hong Kong and Anhui cohorts, or they have high LD with the replicated SNPs and comparable P-values. Functional annotations from RegulomeDB also did not reveal any clear hint worth pursuing at this moment.

To identify how those loci might influence SLE susceptibility, we further investigated the potential biological function of the three putative susceptibility genes in this region (*PHLDB1, DDX6* and *CXCR5*). *PHLDB1* has been shown to be an insulin-responsive protein that enhances AKT activation (43). AKT signaling pathway plays an important role in cellular proliferation and growth signaling. It was demonstrated that murine models with an activated PI3 K/AKT/mTOR signaling pathway in lymphocytes developed features of systemic autoimmunity, linking this pathway to autoimmune diseases (44). Abnormal activation of the AKT signaling pathway was also found in peripheral blood T cells from individuals with SLE (45,46). In addition, based on the results from the ENCODE project (http://genome.ucsc.edu/ENCODE/), the intronic susceptible variant, rs11603023, is suggested to locate in a binding site for transcription factor TAF1, which is known to be involved in the progression of G1 phase of the cell cycle. SNP rs11603023 is in high LD ($r^2 \geq 0.8$) with several imputed SNPs that may locate in regulatory elements based on RegulomeDB annotations (Supplementary Material, Table S3), although further study is needed to elucidate functional implications, if any, of all these SNPs in disease pathogenesis.

DDX6 encodes a member of the DEAD box protein family (47). The protein is an RNA helicase found in P-bodies and stress granules, which functions in translation suppression, mRNA degradation and microRNA-induced gene silencing (48). Of note, transfected miR-146a was demonstrated to produce a comparable increase in the number of P-bodies, and this was accompanied by a reduction in major cytokines/chemokines induced by lipopolysaccharide (49). Interestingly, a functional variant related to miR-146a expression was also identified to confer disease risk of SLE in our previous study (rs57095329, P_meta $= 2.74E-08$, OR = 1.29) (27), and these findings together may indicate a vital role of *DDX6* and microRNA in the etiology of SLE.

While the detailed functions of *PHLDB1* and *DDX6* are still unknown, another potential susceptibility gene suggested in this study, *CXCR5* is well studied and its involvement in

Table 6. Meta-analysis results for SNPs reported in previous GWASs on other immune-related diseases

SNP	Associated diseases	Gene	Allele	I^2	P_{meta}	HK_GWAS Case (freq)	Ctrl (freq)	P_{HKGWAS}	OR (95% CI)	AH_GWAS Case (freq)	Ctrl (freq)	P_{AHGWAS}	OR (95% CI)
rs498872	Glioma[31]	PHLDB1	T<C	0	0.2412	0.25	0.26	0.7667	0.98 (0.85–1.13)	0.27	0.28	0.1869	0.92 (0.80–1.04)
rs10892279	CD,RA[34]	DDX6	A<G	0	0.2049	0.35	0.33	0.2315	1.09 (0.95–1.24)	0.27	0.26	0.5341	1.04 (0.91–1.19)
rs503425(*)	SLE[19]	DDX6	T<C	0	0.3059	0.37	0.36	0.5911	1.04 (0.91–1.18)	0.34	0.32	0.3694	1.06 (0.93–1.20)
rs6421571	PBC[32]	CXCR5	T<C	0	0.00498	0.11	0.13	0.09811	0.84 (0.69–1.03)	0.12	0.14	0.02177	0.81 (0.68–0.97)
rs630923	MS[33]	CXCR5	A<C	10	0.08922	0.08	0.06	0.05387	1.27 (1.00–1.62)	0.07	0.07	0.5923	1.06 (0.85–1.34)

Case/Ctrl (freq) indicates minor allele frequency of the SNP in cases or controls, the calculation of OR is also based on the minor allele of each SNP. Risk allele of each SNP is underlined.
I^2: heterogeneity index. (range 0–100).
*The data of SNP rs503425 are based on the imputed result using SNP genotypes in Chromosome 11 of 286 Asians from the 1000 Genome Project (June 2011 data release) as the reference.
CD, celiac disease; RA, rheumatoid arthritis; PBC, primary biliary cirrhosis; MS, multiple sclerosis.

autoimmunity is well supported by numerous functional studies. *CXCR5* encodes a multi-pass membrane protein that belongs to the CXC chemokine receptor family. Being involved in B-cell migration, CXCR5 is enriched in mature B-cells and Burkitt's lymphoma (50,51). In addition, CXCR5 has also been demonstrated to be constitutively expressed by follicular helper T lymphocytes (T_{fh}), which are involved in autoantibody production and systemic autoimmunity (52,53).

In conclusion, focusing on the 11q23.3 region with proven association with SLE from previous work, we have identified three novel and independent susceptibility variants for SLE in this region via meta-analysis and further replication in independent cohorts. The presence of multiple independent-associated variants in this region significantly increased its total contribution to genetic variance explained. The putative susceptibility genes identified in this study suggest roles of AKT signaling pathway, RNA-induced silencing and chemokine/chemokine receptor axis in SLE pathogenesis. Our findings may enrich the list of SLE susceptibility genes as well as advance our knowledge on the etiology of this complicated disease. In addition, the discrepancy of risk variants in the 11q23.3 region in association with SLE and other diseases with immune-related components (shown in Table 6 and Supplementary Material, Table S6) indicates that certain regions in our genome may be involved in multiple diseases but through different and complicated biological mechanisms. Understanding the biology of these regions has a long way to go, but they may open a new door for us to better elucidate disease pathogenesis.

MATERIALS AND METHODS

Samples used in this study were collected from Hong Kong, Anhui province, China and Bangkok, Thailand, respectively. The Hong Kong cases were SLE patients visiting five hospitals in Hong Kong: Queen Mary Hospital, Tuen Mun Hospital, Queen Elizabeth Hospital, Pamela Youde Nethersole Eastern Hospital and Princess Margaret Hospital. They were all of self-reported Chinese ethnicity living in Hong Kong. Clinical and serological data and autoantibody profiles were both recorded at the time of diagnosis and subsequent doctor visits, which were also reviewed at the time of sample collection. The records include 22 kinds of information, including gender, age of onset, autoantibody profiles (anti-ANA, anti-dsDNA, IgG, IgM, anti-Sm, aPL, anti-La, anti-Ro, anti-nuclear factor and anti-RNP) and clinical manifestations including malar rash, discoid rash, photosensitivity, oral ulcer, arthritis, serositis, renal involvement, neurological involvement, haematological involvement and immunological involvement. Controls for the Hong Kong cohort were healthy blood donors from the Hong Kong Red Cross (for the Hong Kong Replication Panel) and individuals from other GWAS studies conducted in the University of Hong Kong, genotyped on the same platform at the same time (GWAS stage). SLE cases for Anhui Replication were all self-reported Chinese ethnicity living in Anhui province, central China, visiting Departments of Rheumatology at Anhui Provincial Hospital and the First Affiliated Hospital of Anhui Medical University in Hefei, Anhui Province. Controls were selected from a pool of healthy blood donors recruited from Hefei, Anhui province, with an effort to match for the age and sex of

the corresponding SLE patients. The samples for the Anhui GWAS were obtained from multiple hospitals in two geographic regions (central and southern China), and the corresponding controls were clinically assessed to be without SLE, other autoimmune disorders, systemic disorders or family history of autoimmune diseases. The Thai cases were SLE patients attending King Chulalongkorn Memorial Hospital, a tertiary referral center in Bangkok. Thai controls were recruited from unrelated voluntary healthy donors from the same ethnic background and geographic area as the Thai SLE patients. All cases involved in this study had medical records documenting fulfillment of the revised criteria of the American College of Rheumatology for diagnosis of SLE. The studies were approved by the respective Institutional Review Board of all the institutions listed above, and all subjects gave informed consent.

Genotyping

The two GWASs from Hong Kong and Anhui were conducted as previously reported ([35,37,38]), both by Illumina 610-Quad Human Beadchip array. Further replication for selected SNPs was performed by a TaqMan SNP genotyping method using assay-on-demand probes and primers (Applied Biosystems, Foster City, CA, USA; Catalog nos. C__31695728_10 for rs11603023, C__27999313_10 for rs4639966, C__1052268_10 for rs638893, C__31035641_10 for rs10892301) on the remaining samples from Hong Kong that were not included in the GWAS stage, samples collected from Bangkok, Thailand, and samples from an independent Anhui cohort. Genotyping accuracy was confirmed by direct sequencing of polymerase chain reaction products for some randomly chosen samples, which showed 100% concordance. Genotyping concordance between Illumina Human 610-Quad Beadchip and Taqman SNP genotyping method was also checked on randomly selected samples and the two methods showed complete concordance.

Imputation

To obtain genotypes of additional 11q23.3 SNPs, SNP genotypes in Chromosome 11 of 286 Asians from the 1000 Genome Project (June 2011 data release) were used as the reference in imputation for Hong Kong and Anhui subjects. Imputation was performed using IMPUTE2.1.1 ([54]), and imputed SNPs with an information score of >0.9 were included for further analyses.

Association analysis

The same quality control criteria were applied to process both genotyped and imputed SNPs. After a quality control process removing SNPs with low call rates (<90%), low minor allele frequencies (<1%) and violation of the Hardy–Weinberg equilibrium ($P \leq 10^{-4}$), and removing samples with low call rates (<90%) and hidden relationship detected using PLINK ([41]), 325 324 variants in Chr11, 1047 cases and 1205 controls from the Anhui cohort, and 324 795 variants in Chr11, 612 cases and 2193 controls from the Hong Kong cohort were analyzed.

A weighted z-score method was used for meta-analysis. To combine results across different cohorts, alleles were oriented to the forward strand of NCBI reference sequence of the human genome to avoid ambiguity associated with C/G and A/T SNPs. Association P-values for each cohort were converted

to z scores, taking into account the direction of association relative to an arbitrary reference allele. A weighted sum of z scores was calculated by weighing each z score by the square root of the quotient of effective sample size for each cohort and sum of the total cohort sample size. The meta-analysis z score was then converted to P-values from a chi-square distribution. The number of overlapped genotyped SNPs and imputed SNPs in 11q23.3 from Hong Kong and Anhui cohorts is 1072 and 14 980, respectively. The meta-analysis was carried out using METAL ([36]).

Joint analysis of association, taking into account the effect of SNP differences between cohorts, was conducted using Cochran–Mantel–Haenszel, and homogeneity of the effect size between different cohorts and different stages of the study was tested by the Breslow–Day test ($P_{\_het}$ in Table [2]), both installed in PLINK. Conditional and stepwise logistic regression methods were used to determine whether independent effects existed. Conditional logistic regression was done by SNPTESTv2.4.1, adjusted for the effect of other SNP(s) in the same locus, also treating potential differences among cohorts as a covariate.

Stepwise logistic regression was performed by SPSS 16.0. Briefly, variables were added to the logistic regression equation one at a time, using the statistical criterion of reducing the −2Log Likelihood error for the included variables. After each variable was entered, each of the included variables was tested to see if the model would be better off if the variable was excluded. The process of adding more variables stops when all of the available variables had been included or when it was not possible to make a statistically significant reduction in −2Log Likelihood using any of the variables not yet included. The order of entry the variables to the logistic regression equation can be used as a measure of relative importance.

Explained variance was estimated with the R algorithm reported by So *et al.* ([28]) by using the OR, allele frequency and disease prevalence. Briefly, a multifactorial liability threshold model was applied for assessing the contribution of individual variants to the total heritability. This model proposed a latent continuous liability, which was assumed to follow a normal distribution with mean 0 and variance 1. The model was theoretically justified by the central limit theorem, as the overall liability was assumed to be composed of many genetic variants and other risk factors with modest effects. The disease was assumed to be present in individuals whose liability exceeded a certain threshold (*T*), and was absent in other individuals. Variance in liability explained (Vg) by a variant can be directly interpreted as the locus-specific heritability.

URLS:

SCAN: http://www.scandb.org
RegulomeDB: http://regulome.stanford.edu/
SNPTESTv2.4.1:https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html
LocusZoom: http://csg.sph.umich.edu/locuszoom/
SNAP: http://www.broadinstitute.org/mpg/snap/

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Tsokos, G.C. (2011) Systemic lupus erythematosus. *N. Engl. J. Med.*, **365**, 2110–2121.
2. Tan, E.M., Cohen, A.S., Fries, J.F., Masi, A.T., McShane, D.J., Rothfield, N.F., Schaller, J.G., Talal, N. and Winchester, R.J. (1982) The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.*, **25**, 1271–1277.
3. Ramos-Niembro, F. and Alarcon-Segovia, D. (1978) Familial aspects of mixed connective tissue disease (MCTD). I. Occurrence of systemic lupus erythematosus in another member in two families and aggregation of MCTD in another family. *J. Rheumatol.*, **5**, 433–440.
4. Sestak, A.L., Shaver, T.S., Moser, K.L., Neas, B.R. and Harley, J.B. (1999) Familial aggregation of lupus and autoimmunity in an unusual multiplex pedigree. *J. Rheumatol.*, **26**, 1495–1499.
5. Johnson, A.E., Gordon, C., Palmer, R.G. and Bacon, P.A. (1995) The prevalence and incidence of systemic lupus erythematosus in Birmingham, England. Relationship to ethnicity and country of birth. *Arthritis Rheum.*, **38**, 551–558.
6. Mok, C.C. (2011) Epidemiology and survival of systemic lupus erythematosus in Hong Kong Chinese. *Lupus*, **20**, 767–771.
7. Simard, J.F., Costenbader, K.H., Liang, M.H., Karlson, E.W. and Mittleman, M.A. (2009) Exposure to maternal smoking and incident SLE in a prospective cohort study. *Lupus*, **18**, 431–435.
8. Tsokos, G.C., Magrath, I.T. and Balow, J.E. (1983) Epstein-Barr virus induces normal B cell responses but defective suppressor T cell responses in patients with systemic lupus erythematosus. *J. Immunol.*, **131**, 1797–1801.
9. Kang, I., Quan, T., Nolasco, H., Park, S.H., Hong, M.S., Crouch, J., Pamer, E.G., Howe, J.G. and Craft, J. (2004) Defective control of latent Epstein-Barr virus infection in systemic lupus erythematosus. *J. Immunol.*, **172**, 1287–1294.
10. Poole, B.D., Scofield, R.H., Harley, J.B. and James, J.A. (2006) Epstein-Barr virus and molecular mimicry in systemic lupus erythematosus. *Autoimmunity*, **39**, 63–70.
11. Remmers, E.F., Plenge, R.M., Lee, A.T., Graham, R.R., Hom, G., Behrens, T.W., de Bakker, P.I., Le, J.M., Lee, H.S., Batliwalla, F. *et al.* (2007) STAT4 And the risk of rheumatoid arthritis and systemic lupus erythematosus. *N. Engl. J. Med.*, **357**, 977–986.
12. Cunninghame Graham, D.S., Graham, R.R., Manku, H., Wong, A.K., Whittaker, J.C., Gaffney, P.M., Moser, K.L., Rioux, J.D., Altshuler, D., Behrens, T.W. *et al.* (2008) Polymorphism at the TNF superfamily gene TNFSF4 confers susceptibility to systemic lupus erythematosus. *Nat. Genet.*, **40**, 83–89.
13. Graham, R.R., Cotsapas, C., Davies, L., Hackett, R., Lessard, C.J., Leon, J.M., Burtt, N.P., Guiducci, C., Parkin, M., Gates, C. *et al.* (2008) Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat. Genet.*, **40**, 1059–1061.
14. Harley, J.B., Alarcon-Riquelme, M.E., Criswell, L.A., Jacob, C.O., Kimberly, R.P., Moser, K.L., Tsao, B.P., Vyse, T.J., Langefeld, C.D., Nath, S.K. *et al.* (2008) Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat. Genet.*, **40**, 204–210.
15. Hom, G., Graham, R.R., Modrek, B., Taylor, K.E., Ortmann, W., Garnier, S., Lee, A.T., Chung, S.A., Ferreira, R.C., Pant, P.V. *et al.* (2008) Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N. Engl. J. Med.*, **358**, 900–909.
16. Kozyrev, S.V., Abelson, A.K., Wojcik, J., Zaghlool, A., Linga Reddy, M.V., Sanchez, E., Gunnarsson, I., Svenungsson, E., Sturfelt, G., Jonsen, A. *et al.* (2008) Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nat. Genet.*, **40**, 211–216.
17. Musone, S.L., Taylor, K.E., Lu, T.T., Nititham, J., Ferreira, R.C., Ortmann, W., Shifrin, N., Petri, M.A., Kamboh, M.I., Manzi, S. *et al.* (2008) Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. *Nat. Genet.*, **40**, 1062–1064.
18. Nath, S.K., Han, S., Kim-Howard, X., Kelly, J.A., Viswanathan, P., Gilkeson, G.S., Chen, W., Zhu, C., McEver, R.P., Kimberly, R.P. *et al.* (2008) A nonsynonymous functional variant in integrin-alpha(M) (encoded by ITGAM) is associated with systemic lupus erythematosus. *Nat. Genet.*, **40**, 152–154.
19. Gateva, V., Sandling, J.K., Hom, G., Taylor, K.E., Chung, S.A., Sun, X., Ortmann, W., Kosoy, R., Ferreira, R.C., Nordmark, G. *et al.* (2009) A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nat. Genet.*, **41**, 1228–1233.
20. Jacob, C.O., Zhu, J., Armstrong, D.L., Yan, M., Han, J., Zhou, X.J., Thomas, J.A., Reiff, A., Myones, B.L., Ojwang, J.O. *et al.* (2009) Identification of IRAK1 as a risk gene with critical role in the pathogenesis of systemic lupus erythematosus. *Proc. Natl Acad. Sci. USA*, **106**, 6256–6261.
21. Webb, R., Merrill, J.T., Kelly, J.A., Sestak, A., Kaufman, K.M., Langefeld, C.D., Ziegler, J., Kimberly, R.P., Edberg, J.C., Ramsey-Goldman, R. *et al.* (2009) A polymorphism within IL21R confers risk for systemic lupus erythematosus. *Arthritis Rheum.*, **60**, 2402–2407.
22. Yang, W., Zhao, M., Hirankarn, N., Lau, C.S., Mok, C.C., Chan, T.M., Wong, R.W., Lee, K.W., Mok, M.Y., Wong, S.N. *et al.* (2009) ITGAM Is associated with disease susceptibility and renal nephritis of systemic lupus erythematosus in Hong Kong Chinese and Thai. *Hum. Mol. Genet.*, **18**, 2063–2070.
23. Chang, Y.K., Yang, W., Zhao, M., Mok, C.C., Chan, T.M., Wong, R.W., Lee, K.W., Mok, M.Y., Wong, S.N., Ng, I.O. *et al.* (2009) Association of BANK1 and TNFSF4 with systemic lupus erythematosus in Hong Kong Chinese. *Genes Immun.*, **10**, 414–420.
24. Yang, W., Ng, P., Zhao, M., Hirankarn, N., Lau, C.S., Mok, C.C., Chan, T.M., Wong, R.W., Lee, K.W., Mok, M.Y. *et al.* (2009) Population differences in SLE susceptibility genes: STAT4 and BLK, but not PXK, are associated with systemic lupus erythematosus in Hong Kong Chinese. *Genes Immun.*, **10**, 219–226.
25. Cunninghame Graham, D.S., Morris, D.L., Bhangale, T.R., Criswell, L.A., Syvanen, A.C., Ronnblom, L., Behrens, T.W., Graham, R.R. and Vyse, T.J. (2011) Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. *PLoS Genet.*, **7**, e1002341.
26. Zhao, J., Wu, H., Khosravi, M., Cui, H., Qian, X., Kelly, J.A., Kaufman, K.M., Langefeld, C.D., Williams, A.H., Comeau, M.E. *et al.* (2011) Association of genetic variants in complement factor H and factor H-related

genes with systemic lupus erythematosus susceptibility. *PLoS Genet*, **7**, e1002079.

27. Luo, X., Yang, W., Ye, D.Q., Cui, H., Zhang, Y., Hirankarn, N., Qian, X., Tang, Y., Lau, Y.L., de Vries, N. *et al.* (2011) A functional variant in microRNA-146a promoter modulates its expression and confers disease risk for systemic lupus erythematosus. *PLoS Genet.*, **7**, e1002128.

28. So, H.C., Gui, A.H., Cherny, S.S. and Sham, P.C. (2011) Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet. Epidemiol.*, **35**, 310–317.

29. Ke, X. (2012) Presence of multiple independent effects in risk loci of common complex human diseases. *Am. J. Hum. Genet.*, **91**, 185–192.

30. Yang, W., Tang, H., Zhang, Y., Tang, X., Zhang, J., Sun, L., Yang, J., Cui, Y., Zhang, L., Hirankarn, N. *et al.* (2013) 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.*, **92**, 41–51.

31. Shete, S., Hosking, F.J., Robertson, L.B., Dobbins, S.E., Sanson, M., Malmer, B., Simon, M., Marie, Y., Boisselier, B., Delattre, J.Y. *et al.* (2009) Genome-wide association study identifies five susceptibility loci for glioma. *Nat. Genet.*, **41**, 899–904.

32. Mells, G.F., Floyd, J.A., Morley, K.I., Cordell, H.J., Franklin, C.S., Shin, S.Y., Heneghan, M.A., Neuberger, J.M., Donaldson, P.T., Day, D.B. *et al.* (2011) Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis. *Nat. Genet.*, **43**, 329–332.

33. Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C., Hunt, S.E. *et al.* (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*, **476**, 214–219.

34. Zhernakova, A., Stahl, E.A., Trynka, G., Raychaudhuri, S., Festen, E.A., Franke, L., Westra, H.J., Fehrmann, R.S., Kurreeman, F.A., Thomson, B. *et al.* (2011) Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet.*, **7**, e1002004.

35. Han, J.W., Zheng, H.F., Cui, Y., Sun, L.D., Ye, D.Q., Hu, Z., Xu, J.H., Cai, Z.M., Huang, W., Zhao, G.P. *et al.* (2009) Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat. Genet.*, **41**, 1234–1237.

36. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.

37. Yang, J., Yang, W., Hirankarn, N., Ye, D.Q., Zhang, Y., Pan, H.F., Mok, C.C., Chan, T.M., Wong, R.W., Mok, M.Y. *et al.* (2011) ELF1 Is associated with systemic lupus erythematosus in Asian populations. *Hum. Mol. Genet.*, **20**, 601–607.

38. Yang, W., Shen, N., Ye, D.Q., Liu, Q., Zhang, Y., Qian, X.X., Hirankarn, N., Ying, D., Pan, H.F., Mok, C.C. *et al.* (2010) Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. *PLoS Genet.*, **6**, e1000841.

39. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) Locuszoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.

40. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S. *et al.* (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*, **22**, 1790–1797.

41. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, **81**, 559–575.

42. Tang, X.F., Zhang, Z., Hu, D.Y., Xu, A.E., Zhou, H.S., Sun, L.D., Gao, M., Gao, T.W., Gao, X.H., Chen, H.D. *et al.* (2013) Association analyses identify three susceptibility loci for vitiligo in the Chinese Han population. *J. Invest. Dermatol.*, **133**, 403–410.

43. Zhou, Q.L., Jiang, Z.Y., Mabardy, A.S., Del Campo, C.M., Lambright, D.G., Holik, J., Fogarty, K.E., Straubhaar, J., Nicoloro, S., Chawla, A. *et al.* (2010) A novel pleckstrin homology domain-containing protein enhances insulin-stimulated Akt phosphorylation and GLUT4 translocation in adipocytes. *J Biol Chem*, **285**, 27581–27589.

44. Wu, T., Qin, X., Kurepa, Z., Kumar, K.R., Liu, K., Kanta, H., Zhou, X.J., Satterthwaite, A.B., Davis, L.S. and Mohan, C. (2007) Shared signaling networks active in B cells isolated from genetically distinct mouse models of lupus. *J Clin Invest*, **117**, 2186–2196.

45. Tang, H., Tan, G., Guo, Q., Pang, R. and Zeng, F. (2009) Abnormal activation of the Akt-GSK3beta signaling pathway in peripheral blood T cells from patients with systemic lupus erythematosus. *Cell Cycle*, **8**, 2789–2793.

46. Besliu, A.N., Pistol, G., Marica, C.M., Banica, L.M., Chitonu, C., Ionescu, R., Tanaseanu, C., Tamsulea, I., Matache, C. and Stefanescu, M. (2009) PI3K/Akt Signaling in peripheral T lymphocytes from systemic lupus erythematosus patients. *Roum Arch Microbiol Immunol*, **68**, 69–79.

47. Lu, D. and Yunis, J.J. (1992) Cloning, expression and localization of an RNA helicase gene from a human lymphoid cell line with chromosomal breakpoint 11q23.3. *Nucleic Acids Res*, **20**, 1967–1972.

48. Akao, Y., Marukawa, O., Morikawa, H., Nakao, K., Kamei, M., Hachiya, T. and Tsujimoto, Y. (1995) The rck/p54 candidate proto-oncogene product is a 54-kilodalton D-E-A-D box protein differentially expressed in human and mouse tissues. *Cancer Res*, **55**, 3444–3449.

49. Pauley, K.M., Satoh, M., Pauley, B.A., Dominguez-Gutierrez, P.R., Wallet, S.M., Holliday, L.S., Cha, S., Reeves, W.H. and Chan, E.K. (2010) Formation of GW/P bodies as marker for microRNA-mediated regulation of innate immune signaling in THP-1 cells. *Immunol Cell Biol*, **88**, 205–212.

50. Burkle, A., Niedermeier, M., Schmitt-Graff, A., Wierda, W.G., Keating, M.J. and Burger, J.A. (2007) Overexpression of the CXCR5 chemokine receptor, and its ligand, CXCL13 in B-cell chronic lymphocytic leukemia. *Blood*, **110**, 3316–3325.

51. Gunn, M.D., Ngo, V.N., Ansel, K.M., Ekland, E.H., Cyster, J.G. and Williams, L.T. (1998) A B-cell-homing chemokine made in lymphoid follicles activates Burkitt's lymphoma receptor-1. *Nature*, **391**, 799–803.

52. Fazilleau, N., Mark, L., McHeyzer-Williams, L.J. and McHeyzer-Williams, M.G. (2009) Follicular helper T cells: lineage and location. *Immunity*, **30**, 324–335.

53. Chevalier, N., Jarrossay, D., Ho, E., Avery, D.T., Ma, C.S., Yu, D., Sallusto, F., Tangye, S.G. and Mackay, C.R. (2011) CXCR5 Expressing human central memory CD4T cells and their relevance for humoral immune responses. *J Immunol*, **186**, 5556–5568.

54. Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*, **5**, e1000529.