Human Molecular Genetics, 2015, Vol. 24, No. 1 **274–284**
doi:10.1093/hmg/ddu429
Advance Access published on August 22, 2014

Meta-analysis of GWAS on two Chinese populations followed by replication identifies novel genetic variants on the X chromosome associated with systemic lupus erythematosus

Yan Zhang[1], Jing Zhang[1], Jing Yang[1], Yongfei Wang[1], Lu Zhang[1], Xianbo Zuo[7], Liangdan Sun[7], Hai-Feng Pan[8], Nattiya Hirankarn[9], Tingyou Wang[1], Ruoyan Chen[1], Dingge Ying[1], Shuai Zeng[1], Jiangshan Jane Shen[1], Tsz Leung Lee[1], Chak Sing Lau[2], Tak Mao Chan[2], Alexander Moon Ho Leung[12], Chi Chiu Mok[13], Sik Nin Wong[14], Ka Wing Lee[15], 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[16], Niko Kei Chiu Tse[17], Xiang-Pei Li[18], Yingyos Avihingsanon[10], Pornpimol Rianthavorn[11], Thavatchai Deekajorndej[11], Kanya Suphapeetiporn[11], Vorasuk Shotelersuk[11], Shirley King Yee Ying[16], Samuel Ka Shun Fung[16], Wai Ming Lai[17], Chun-Ming Wong[3], Irene Oi Lin Ng[3], Maria-Merce Garcia-Barcelo[4], Stacey S. Cherny[5], Paul Kwong-Hang Tam[4], Pak Chung Sham[5,6], Sen Yang[7], Dong Qing Ye[8], Yong Cui[7,*], Xue-Jun Zhang[7], Yu Lung Lau[1,*] and Wanling Yang[1,6,*]

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

Received March 24, 2014; Revised July 25, 2014; Accepted August 20, 2014

Systemic lupus erythematosus (SLE) is a prototypic autoimmune disease that affects mainly females. What role the X chromosome plays in the disease has always been an intriguing question. In this study, we examined the genetic variants on the X chromosome through meta-analysis of two genome-wide association studies (GWAS) on SLE on Chinese Han populations. Prominent association signals from the meta-analysis were replicated in 4 additional Asian cohorts, with a total of 5373 cases and 9166 matched controls. We identified a novel variant in *PRPS2* on Xp22.3 as associated with SLE with genome-wide significance (rs7062536, OR = 0.84, $P = 1.00E-08$). Association of the *L1CAM-MECP2* region with SLE was reported previously. In this study, we identified

*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. Tel: +852 28315077; Fax: +852 28198142; Email: yangwl@hkucc.hku.hk (W.Y.); Tel: +852 22554481; Fax +852 2855 1523; Email: lauylung@hkucc.hku.hk (Y.L.L.); State Key Laboratory Incubation Base of Dermatology, Key Laboratory of Dermatology, Anhui Medical University, Ministry of Education, China, Hefei, Anhui 230032, China. Email: wuhucuiyong@vip.163.com (Y.C.)

independent contributors in this region in *NAA10* (rs2071128, OR = 0.81, *P* = 2.19E−13) and *TMEM187* (rs17422, OR = 0.75, *P* = 1.47E−15), in addition to replicating the association from *IRAK1-MECP2* region (rs1059702, OR = 0.71, *P* = 2.40E−18) in Asian cohorts. The X-linked susceptibility variants showed higher effect size in males than that in females, similar to results from a genome-wide survey of associated SNPs on the autosomes. These results suggest that susceptibility genes identified on the X chromosome, while contributing to disease predisposition, might not contribute significantly to the female predominance of this prototype autoimmune disease.

INTRODUCTION

Systemic lupus erythematosus (SLE) is an autoimmune disease with an overwhelming female predominance. Women of childbearing age are most affected, though pre-menstrual and postmenopausal women also have higher prevalence than males of the same age group (1,2). The X chromosome certainly plays a role in SLE risk and prevalence, as demonstrated by many functional and susceptibility variants identified on the X chromosome (3–7). Interestingly, men with Klinefelter syndrome (XXY) have much higher disease prevalence than normal XY males (8), suggesting a dosage effect of the X chromosome genes in disease pathogenesis.

The mechanism through which X-linked genes contribute to female disease prevalence is poorly understood. In healthy women, one X chromosome is inactivated in a process known as X chromosome inactivation (XCI), to compensate for the dosage difference between the two genders. In this process, *Xist,* a long noncoding RNA, coats the chromosome to be inactivated and recruits other proteins to the X chromosome, leading to its silencing (9). It is now known that X chromosome is only partially inactivated under physiological conditions. About 10–15% of the genes may be expressed by both X chromosomes in female cells, thus leading to incomplete dosage compensation with a higher expression level than that in males (10). Female cells are mosaic for X-linked genes due to the random process of XCI. Skewed X inactivation refers to preferential expression of either maternal or paternal allele in majority of cells in a tissue and is thought to play a role in autoimmunity (11).

One particular question on the role of X chromosome genes in autoimmune diseases is that whether susceptibility genes on the X chromosome have more contribution toward the disease in females than in males, potentially through escaping dosage compensation with increased expression, or through skewed XCI with risk alleles expressed in most of female cells, leading to more females being affected by the risk allele (both homozygous and heterozygous genotypes) than males. Comparing the effect size of susceptibility variants in homozygous, heterozygous females and hemizygous males may shed light on the mechanism through which X-linked genes contribute to female disease prevalence. If the susceptibility genes are involved in escaping XCI or are involved in skewed X inactivation toward the risk allele, then the susceptibility variant may demonstrate a much larger effect size in females than in males or they may follow a near dominant rather than additive model of disease association.

Tukiainen *et al.* (12) studied association of >400000 X-chromosome SNPs to levels of twelve cardiometabolic and anthropometric traits on Finnish and Swedish individuals and identified three loci with association of genome-wide significance. Interestingly, the SNP in a locus near *ITM2A* associated with height showed a clear lack of dosage compensation, with estimated effect size twice as much in women as in men. For the other two loci, the proportion of variance explained was roughly twice the size in men as in women, as expected under the model of random XCI, different from the SNP in *ITM2A*, which showed a much higher variance explained in women than in men. Consistent with the genetic findings, they observed significantly higher expression of *ITM2A* in whole blood in women.

It is known that promoters on the X chromosome to be silenced are often heavily methylated (10). In SLE patients, however, this process might be compromised. For example, DNA methylation was found to be defective in CD4+ T cells in patients with lupus (13,14), and hypomethylation in these cells might lead to overexpression of genes that are important in SLE pathogenesis (15,16). Several X-linked genes including *CD40LG* were found to be overexpressed in female patients with SLE but not in male patients (17,18). *CXCR3*, *OGT* and microRNAs such as miR-98, miR188-3p, miR-421 and miR-503 were also found overexpressed in T cells in female but not male patients (16). This aberrant lack of methylation and gender-specific overexpression of genes might play important roles in female predominance of the disease.

A number of X-linked genes have been reported to be associated with SLE (3–6). An SNP on the *3′ UTR* of *TLR7* was shown to be associated with SLE in different ancestral groups, with higher effect size in males than in females (5). *IRAK1* and *MECP2* were reported to be associated with SLE by two independent studies using a candidate gene approach (3,4). These two genes are physically close to each other, and the strong linkage disequilibrium (LD) in this region led to a debate on which gene could be responsible for disease susceptibility (19). In a recent study (6), Kaufman *et al.* concluded that a nonsynonymous SNP rs1059702 (S196F) in *IRAK1* might be the causal variant in this region. The Xq28 region also has been suggested to be involved in several autoimmune diseases including rheumatoid arthritis (20) and systemic sclerosis (SSc) (21). Interestingly, though rs1059702 was reported in each study, in both cases, there are other SNPs in this region forming risk haplotypes with rs1059702, suggesting existence of independent contributors in this region.

In the current study, meta-analysis was first performed on data from two genome-wide association studies (GWAS) on Chinese populations on the X chromosome and potential association signals were followed up by replication in independent cohorts. We confirmed the association of the Xq28 region with SLE and identified independent variants centromeric to this region, as well as a novel susceptibility variant in *PRPS2* located on Xp22.2. The susceptibility variants on the X-chromosome showed a higher effect size in males than in females, with no evidence of lack of dosage compensation. Higher cumulative genetic risk in affected males was also observed in a genome-wide survey.

RESULTS

X chromosome imputation and meta-analysis

Our study comprised two sets of GWAS data with a total of 612 cases and 2193 controls from the Hong Kong GWAS and 1047 and 1205 controls from the GWAS conducted in Anhui, China. Based on the fact that 90% of the cases were females and the gender difference in X-chromosome dosage, we first focused only on female samples, analyzing a total of 559 cases and 1274 controls, and 984 cases and 532 controls from Hong Kong and Anhui, respectively. Meta-analysis of the two GWAS datasets was conducted using the inverse variance method installed in METAL (22). Q–Q plots (Fig. 1) showed a clear deviation at the tail from the diagonal line for both cohorts and the meta-analysis, suggesting existence of susceptibility variants on the X-chromosome and consistencies between the two cohorts.

Imputation of the X-chromosome SNPs was performed on the two GWAS data on both male and female samples, using data from 1000 Genomes project as references. Association analysis was carried out by SNPTEST (23), using sex as a covariate to correct for potential gender differences. Meta-analysis of imputed data showed that there were 284 SNPs reached P_meta of $<10^{-3}$. Two SNPs were removed due to inconsistent association direction between the two studies. Summary statistics of all the X-linked SNPs after imputation were made available to the public for further exploration in future studies (Supplementary Material, Table S1). The two GWAS datasets were subsequently assessed for population stratification using EIGENSTRAT (24). Statistical significance of the SNPs was re-assessed using SNPTEST, using the first two most prominent principal components (PC) to correct for potential population stratification (*P_adj* in Supplementary Material, Table S1).

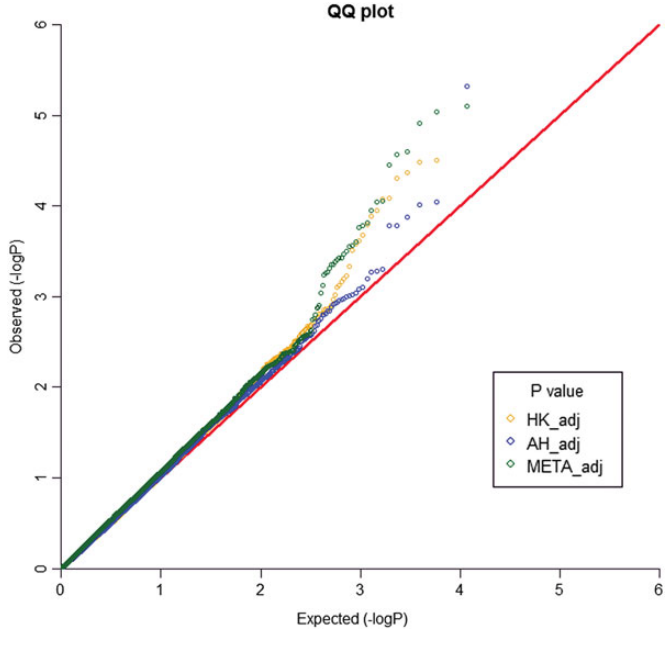

Figure 1. Quantile–quantile plot (Q–Q plot) of the observed versus expected *P*-values of X-chromosome SNPs from a female-only analysis. The green circles are based on the results of the meta-analysis. The yellow circles are from Hong Kong GWAS, and the blue ones are based on results from Anhui GWAS.

Comparing *P_meta* (without adjustment using PC as covariates) and *P_adj* values, it can be seen that there was no substantial difference before and after applying the correction using PCs. No imputed SNPs, however, showed better *P*-values or convincing predicted functional implications that would warrant further replication on top of the genotyped SNPs. Based on the *P*-value in each cohort and the LD information, 11 SNPs from 3 regions near or in *VCX2* (1 SNP), *PRPS2* (1 SNP) and *L1CAM1-MECP2* (9 SNPs) were selected for further replication in independent cohorts.

Replication and further analysis of selected SNPs

Further replication of the selected SNPs was performed by Sequenom or TaqMan on independent samples from Hong Kong, Anhui and Thailand. In the first stage of the replication, the 11 SNPs were analyzed on a total of 4271 cases and 6839 controls from Hong Kong Replication Panel (HK_REP), Anhui Replication Panel 1(AH_REP1) and Thailand Replication Panel (TH_REP). After analysis of the combined results from replication stage 1 and the GWAS, 10 of the 11 SNPs (except SNP rs5978830 in *VCX2* region) showed $P < 1E-05$ and were further studied using samples from Anhui Replication Panel 2, which involved 1102 cases and 2327 controls. The final combined *P*-values and ORs were calculated using logistic regression where gender and cohort information were used as covariates (Table 1).

Linkage disequilibrium (r^2) patterns of the nine SNPs in the *L1CAM-MECP2* region on Xq28 showed consistent LDs in both females and males among Anhui, Hong Kong and Thailand cohorts (Fig. 2 and Supplementary Material, Fig. S1). As shown in Figure 2 based on the Anhui population, SNPs rs4898457 and rs2070097 showed relatively low LD with other SNPs ($r^2 < 0.3$). Both of them did not reach genome-wide significance, whereas rs4898457 but not rs2070097 showed consistent trend in all the replications (P_combined $= 2.84E-06$, OR $=$ 0.87) and the role of rs4898457 in SLE requires further study. The seven SNPs that reached GWAS significance, from rs2071128 to rs2734647, can be roughly divided into three blocks according to their LD ($r^2 > 0.7$, Fig. 2). Of the three LD blocks, SNPs rs2071128 and rs1557501 have high LD with each other ($r^2 = 0.83$). SNPs rs17422, rs2266888 and rs6571303 have pairwise LD (r^2) bigger than 0.7. SNPs rs1059702 and rs2734647 around *IRAK1* and *MECP2* genes also have high LD with each other ($r^2 = 0.88$).

Independence test of the replicated SNPs in the *L1CAM-MECP2* region

The independent contribution of each SNP in this region to disease susceptibility was further examined. Pairwise independence tests of the nine SNPs were performed by logistic regression, also considering potential subpopulation and gender differences (Table 2). It is noted that the three haplotype blocks roughly defined by r^2 values shown in Figure 2 each had independent contributions toward disease susceptibility. Three representative SNPs of these blocks, rs2071128, rs17422 and rs1059702, showed consistent significance after adjusting for the effect of other individual SNPs in different LD blocks. A stepwise logistic regression analysis was also performed to test independence of these SNPs

Table 1. Replication results on SNPs on X chromosome selected through meta-analysis of GWAS data

Marker	Gene symbol	Allele	HK_GWAS Case/control *P*	AH_GWAS Case/control *P*	HK_REP Case/control *P*	AH_REP1 Case/control *P*	TH_REP Case/control *P*	AH_REP2 Case/control *P*	OR (95%)	*P*_combined
rs5978830	*near-VCX2*	G<A	0.49/0.43	0.49/0.48	0.46/NA	0.49/0.50	0.32/0.40	NA/NA	0.92 (0.86 ~0.98)	4.24E−02
			5.31E−03	1.41E−01	NA	2.54E−02	2.70E−01	NA		
rs7062536	*PRPS2*	A<G	0.27/0.32	0.28/0.33	0.29/0.32	0.30/0.30	0.28/0.34	0.28/0.32	0.84 (0.80 ~0.89)	**1.00E−08**
			4.61E−05	1.13E−02	3.15E−01	7.68E−01	7.07E−03	7.86E−03		
rs4898457	*L1CAM-upstream*	G<A	0.23/0.29	0.26/0.29	0.28/0.29	0.27/0.30	0.32/0.33	0.26/0.29	0.87 (0.82 ~0.92)	2.84E−06
			2.36E−05	5.79E−02	5.37E−01	2.14E−02	8.07E−01	3.51E−02		
rs2070097	*ARHGAP4*	T<C	0.24/0.29	0.25/0.27	0.27/0.28	0.26/0.29	0.26/0.29	0.26/0.21	0.95 (0.89 ~1.00)	5.03E−03
			6.64E−06	3.40E−01	2.93E−01	7.56E−02	2.71E−01	9.56E−05		
rs2071128	*NAA10*	G<A	0.22/0.31	0.31/0.34	0.26/0.28	0.31/0.37	0.22/0.27	0.29/0.34	0.81 (0.77 ~0.86)	**2.19E−13**
			7.67E−07	1.14E−01	2.14E−01	4.78E−04	5.82E−02	4.21E−05		
rs1557501	*NAA10*	C<T	0.19/0.27	0.27/0.31	0.23/0.25	0.28/0.32	0.21/0.23	0.26/0.30	0.83 (0.79 ~0.88)	**7.84E−10**
			1.68E−07	2.04E−02	4.32E−01	1.52E−02	5.27E−01	8.53E−04		
rs17422	*HCFC1*	T<C	0.17/0.23	0.20/0.25	0.18/0.22	0.20/0.24	0.17/0.21	0.18/0.23	0.75 (0.71 ~0.80)	**1.47E−15**
			3.06E−05	2.18E−03	4.32E−02	1.55E−03	9.96E−02	4.63E−08		
rs2266888	*TMEM187*	G<A	0.17/0.23	0.20/0.25	0.19/0.22	0.20/0.23	0.19/0.21	0.18/0.24	0.76 (0.72 ~0.81)	**8.20E−15**
			3.25E−05	3.56E−03	2.87E−02	1.27E−02	4.81E−01	3.05E−09		
rs6571303	*TMEM187*	C<T	0.20/0.26	0.25/0.28	0.22/0.25	0.25/0.29	0.21/0.23	0.23/0.29	0.80 (0.76 ~0.84)	**3.06E−13**
			3.61E−05	3.43E−02	7.46E−02	1.12E−02	5.12E−01	3.79E−08		
rs1059702	*IRAK1*	C<T	0.16/0.22	0.17/0.21	0.18/0.21	0.17/0.22	0.21/0.26	0.15/0.21	0.71 (0.67 ~0.76)	**2.40E−18**
			1.29E−05	3.74E−04	2.62E−01	2.18E−04	5.95E−02	5.27E−10		
rs2734647	*MECP2*	C<T	0.17/0.24	0.18/0.23	0.19/0.21	0.16/0.22	0.23/0.27	0.17/0.22	0.72 (0.67 ~0.76)	**5.22E−18**
			1.04E−05	4.75E−04	3.95E−01	1.83E−05	8.91E−02	2.60E−08		

Case/control (frequency) indicates minor allele frequency of the SNP in cases or controls; the calculation of OR was based on the minor allele of each SNP. Risk allele of each SNP is underlined. GWAS refers to the GWAS cohort whereas REP refers to replication cohort. SNPs with combined *P* value (*P*_combined) surpassing genome-wide significance (5E-8) were boldfaced.
HK, Hong Kong; AH, Anhui; TH, Thailand.

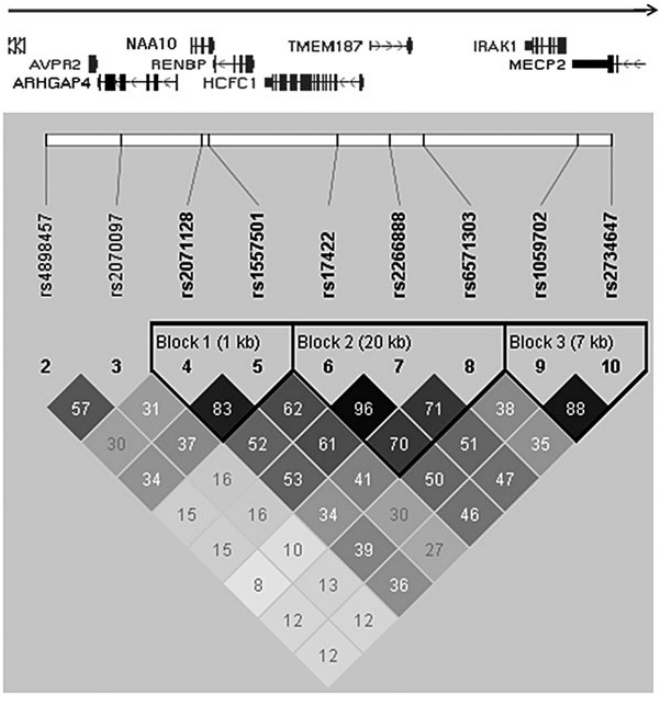

Figure 2. Linkage disequilibrium patterns (r^2) of the replicated SNPs in the region from *L1CAM* to *MECP2*. The analysis was based on Anhui samples (including AH_GWAS, AH_REP1 and AH_REP2), and similar patterns were observed in other cohorts.

(Table 3). This analysis showed that rs1059702 exhibited the strongest association with SLE ($P = 2.27E-08$). Subsequent addition of rs2071128 also significantly improved the model ($P = 1.49E-05$), providing further evidence that variants in this region are independently associated with the disease. Further addition of rs17422 showed marginal improvement to the model ($P = 0.043$). These data, together with analysis of results reported previously by Kaufman *et al.* (6) detailed in Discussion, support the notion that independent effects exist in this region for disease susceptibility. Due to the limited power in detecting independent variants in a region in association studies, the detailed relationships among these variants, however, require further replication in future studies.

Haplotype analysis was also performed using logistic regression on these three SNPs: rs2071128, rs17422 and rs1059702, adjusting for possible differences among cohorts. Analyses of males and females were done separately since treating males as homozygous and jointly analyzing samples would result in inflation of contributions from males. The result indicated that G-T-C haplotype formed by the three SNPs was the major protective haplotype with a combined OR of 0.76 and *P*-value of 9.78E−23, whereas the A-C-T haplotype formed by the major alleles of the three SNPs was the major risk haplotype with a combined OR of 1.19 and $P = 2.78E-27$ in female samples (Table 4). Similar results were observed in male samples with slightly larger effect size on the two major haplotypes (Supplementary Material, Table S2). Haplotype association in different cohorts showed consistent results (Fig. 3).

Table 2. Independence test by adjusting for the effects of other SNPs in the region and treating cohorts and gender as covariates

SNP	SNP whose effect was adjusted for*								
	rs4898457	rs2070097	rs2071128	rs1557501	rs17422	rs2266888	rs6571303	rs1059702	rs2734647
rs4898457	/	2.95E−04	0.25	0.19	0.06	0.04	2.52E−03	0.06	0.05
rs2070097	0.39	/	0.34	0.23	0.69	0.97	0.47	0.76	0.95
rs2071128	1.20E−08	2.30E−10	/	0.01	0.03	0.03	1.70E−04	0.04	0.02
rs1557501	1.77E−05	8.80E−09	0.70	/	0.61	0.56	0.02	0.08	0.03
rs17422	3.43E−11	1.12E−11	2.27E−05	3.56E−05	/	0.20	4.01E−04	0.04	1.54E−03
rs2266888	1.62E−10	8.63E−11	1.38E−04	5.70E−05	0.72	/	1.13E−04	0.06	3.22E−03
rs6571303	2.30E−09	1.85E−09	2.83E−04	6.31E−04	0.26	0.63	/	0.06	0.01
rs1059702	5.02E−14	3.00E−14	1.34E−08	6.40E−10	1.37E−06	8.88E−07	4.35E−09	/	3.56E−03
rs2734647	1.87E−13	3.71E−14	1.53E−08	1.31E−09	1.67E−05	1.33E−05	5.82E−08	0.56	/

*The data in each column represent the remaining effect of association (*P*-values) after adjusting for the effect of SNP(s) on the top row of each column.

Table 3. Significant results from stepwise logistic regression of the six SNPs in the L1CAM-MECP2 region and treating cohorts and gender as covariates

SNP added to the model	*P*-value	OR (95% CI)
rs1059702_C	2.27E−08	0.79 (0.73–0.86)
rs2071128_G	1.49E−05	0.93 (0.90–0.96)
rs17422_T	0.043	0.91 (0.83–0.99)

Differences in effect size for the X-linked susceptibility SNPs between females and males

Larger effect sizes, as measured by OR, were observed in males compared with females in all the susceptibility SNPs confirmed in this study (ORs are shown in Supplementary Material, Table S3). In order to further evaluate potential differences between females and males, heterozygous females were separately considered from homozygous females (Fig. 4, Supplementary Material, Table S3). It can be seen that the ORs (square) for the homozygous females are similar to those in males (triangle), which are much bigger (farther away from 1) than those when all female samples were considered (circle) (Fig. 4, Supplementary Material, Table S3). This is consistent among different cohorts. This result suggested that the X-linked susceptibility variants identified here are not involved in escaping dosage compensation.

In order to test whether the observed differences in effect size between the genders is a characteristic of the X-linked variants or is applicable to other susceptibility SNPs, we examined SNPs in the established loci that showed consistent signals in our meta-analysis results (Supplementary Material, Table S4) (25–30). Gender-specific aggregate genetic risk was examined by calculating the cumulative genetic risk for each gender (31). As shown in Figure 5, on average, male patients have a significantly higher genetic risk than female patients based on Hong Kong (unpaired *t* test, $P < 0.0001$) and Anhui (unpaired *t* test, $P < 0.0001$) GWAS data. Similar patterns were observed when eight SNPs reported in our recent study were examined (32), which were based on a much larger sample size including both GWAS samples and replication data (Supplementary Material, Fig. S2). The HLA region was not included in this estimation of genetic risk due to lack of clarity on independent signals, but preliminary data on selected SNPs in the HLA region showed similar differences between males and females.

DISCUSSION

Through meta-analysis of the two existing GWAS on Chinese Han populations, and further replication on three additional cohorts from Hong Kong, Anhui, and Thailand totaling 5373 cases and 9166 controls matched geographically and ethnically, we have identified multiple variants associated with SLE with evidence of independent effects.

Rs7062536 is located in an intron of *PRPS2*, is on the p arm of the chromosome and is clearly an independent contributor to SLE susceptibility. *PRPS2* was known to play a key role in the process of purine and pyrimidine synthesis (33), with no known function to autoimmunity or other immunological processes. Further work will be needed to understand the biological mechanism of this genetic association. The SNP (rs3853839) in the 3′ UTR of *TLR7* reported earlier, which is on the same arm of the chromosome, had poor imputation quality (impute score <0.6), and this region was also poorly covered by the Illumina 610 Quad Beadchip used in GWAS stage in both Hong Kong and Anhui cohorts.

The other signals identified are highly complex, covering a 134-kb region on Xq28, and overlapping with eight genes from upstream of *L1CAM* to the 3′ UTR of *MECP2*. Independence tests using logistic regression provided evidence of independent effects at least for SNPs rs2071128, rs17422 and rs1059702, although the detailed nature of their contribution toward disease susceptibility remains to be determined.

In the study of Kaufman *et al.* (6), other than rs1059702 (*IRAK1*), which was considered to be the most likely causal variant, six more SNPs reached genome-wide significance based on the meta-analysis results of four different populations. Furthermore, 27 SNPs showed suggestive association in one or more populations. There seem to be population differences in the association signals on the centromeric part of this large region, especially around *NAA10*, *RENBP* and *ARHGAP4* [Fig. 1 and Table 1 in Kaufman *et al.* (6)]. In European Americans (EAs) and to a lesser extent Asians (AA) and Hispanic Americans (HA), comparable association signals were observed in the region between *NAA10* and *ARHGAP4* to those found in the region where *MECP2* and *IRAK1* are located.

In order to further understand the association signals in this region, we tested the LD patterns in the *L1CAM-MECP2* region in different populations, including Asians, Europeans, Hispanics and Africans using data from 1000 Genome project

Table 4. Association of the haplotypes derived from rs2071128, rs17422 and rs1059702

rs2071128-rs17422-rs1059702	HK_GWAS		AH_GWAS		HK_REP		AH_REP1		TH_REP		AH_REP2		OR (CI95)	*P*
	Case	Control	Case	Control	Case	Control	Case	Control	Case	Control	Case	Control		
G-T-C	0.129	0.169	0.135	0.187	0.142	0.158	0.133	0.166	0.13	0.164	0.131	0.174	0.757 (0.753 – 0.761)	9.78E – 23
G-C-C	0.017	0.024	0.026	0.025	0.017	0.019	0.024	0.033	0.022	0.021	0.022	0.025	0.855 (0.806 – 0.907)	0.02
A-C-C	0.013	0.013	–[1]	–	0.013	0.023	–	–	0.063	0.08	–	–	0.610 (0.569 – 0.654)	1.92E – 04
G-T-T	0.038	0.04	0.062	0.052	0.039	0.047	0.059	0.063	0.03	0.044	0.051	0.066	0.982 (0.436 – 2.210)	0.83
G-C-T	0.049	0.063	0.089	0.079	0.063	0.061	0.099	0.102	0.045	0.044	0.091	0.079	1.120 (1.093 – 1.148)	2.76E – 03
A-C-T	0.754	0.691	0.688	0.657	0.716	0.682	0.684	0.636	0.696	0.637	0.706	0.657	1.190 (1.187 – 1.193)	2.78E – 27

Only female samples were analyzed and cohorts were treated as a covariate.
[1]No A-C-C haplotype exist in Anhui cohort, which were consistent among AHGWAS, AHREP1 and AHREP2 strata.

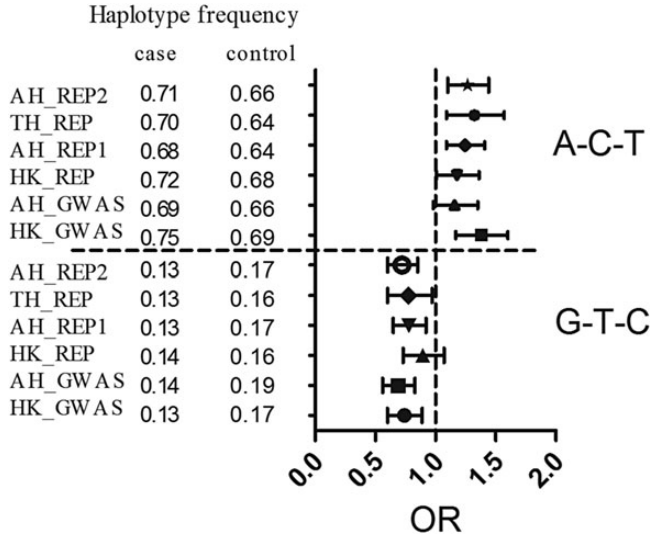

Figure 3. Haplotypic ORs for the A-C-T (Protective) and G-T-C (Risk) haplotypes defined by rs2071128(A), rs174322(C) and rs1059702(T). Shown are results from six independent replications. GWAS refers to the GWAS cohort whereas REP refers to replication cohort. HK, Hong Kong; AH, Anhui; TH, Thailand.

(Supplementary Material, Figs S3–S6). We included the three independent SNPs identified in this study and the six SNPs shown to be highly significant in the study of Kaufman *et al.* (6), in addition to SNP rs1059702 that was highly significant in both studies. In EA samples, SNPs in the region of *ARHGAP4-NAA10-RENBP* showed as strong association signals as those in the *IRAK1-MECP2* region. For example, rs2071129 in *NAA10*, rs13397 in *TMEM187* and rs1059702 in *IRAK1* had similar association *P*-values and ORs (Table 1 of Kaufman *et al.* (6)). There is only intermediate LD between these variants in Europeans, however (r^2 equals 0.73 and 0.62 between rs2071129 (*NAA10*) and rs4898374 (*TMEM187*) with rs1059702, respectively, Supplementary Material, Fig. S3), suggesting a possibility of independent association in the *ARHGAP4-NAA10-RENBP* region and the *TMEM187* region, or a haplotypic association effect from the three SNPs.

In Asians, SNPs in the region of *ARHGAP4-NAA10-RENBP* showed a clear drop in association signals from those in the *IRAK1-MECP2* region [comparing rs2071129 (*NAA10*) and rs1059702 (*IRAK1*) in Table 1 of Kaufman *et al.* (6)], and the LD is also much lower among these SNPs (r^2 between 0.4 and 0.5 between the two regions, Supplementary Material, Fig. S4), very similar to what we have observed in this study. However, rs4898374 in the *TMEM187* region showed similar strength of association compared with that of rs1059702 (*IRAK1*) (6). The LD between the two SNPs is moderate ($r^2 =$ 0.41, Supplementary Material, Fig. S4), far from explaining the comparable association signals observed and indicating independent contributions from SNPs in the *TMEM187* region. Hispanic Americans showed similar association signals as did Asians in the study of Kaufman *et al.* but relatively higher LD in this big region (Supplementary Material, Fig. S5). There is little evidence of association beyond rs13397 (*TMEM187*) in African Americans (AA), and there is also much lower LD for this population [Supplementary Material, Fig. S6 and Table 1 in Kaufman *et al.* (6)]. Overall, the data from Kaufman *et al.*

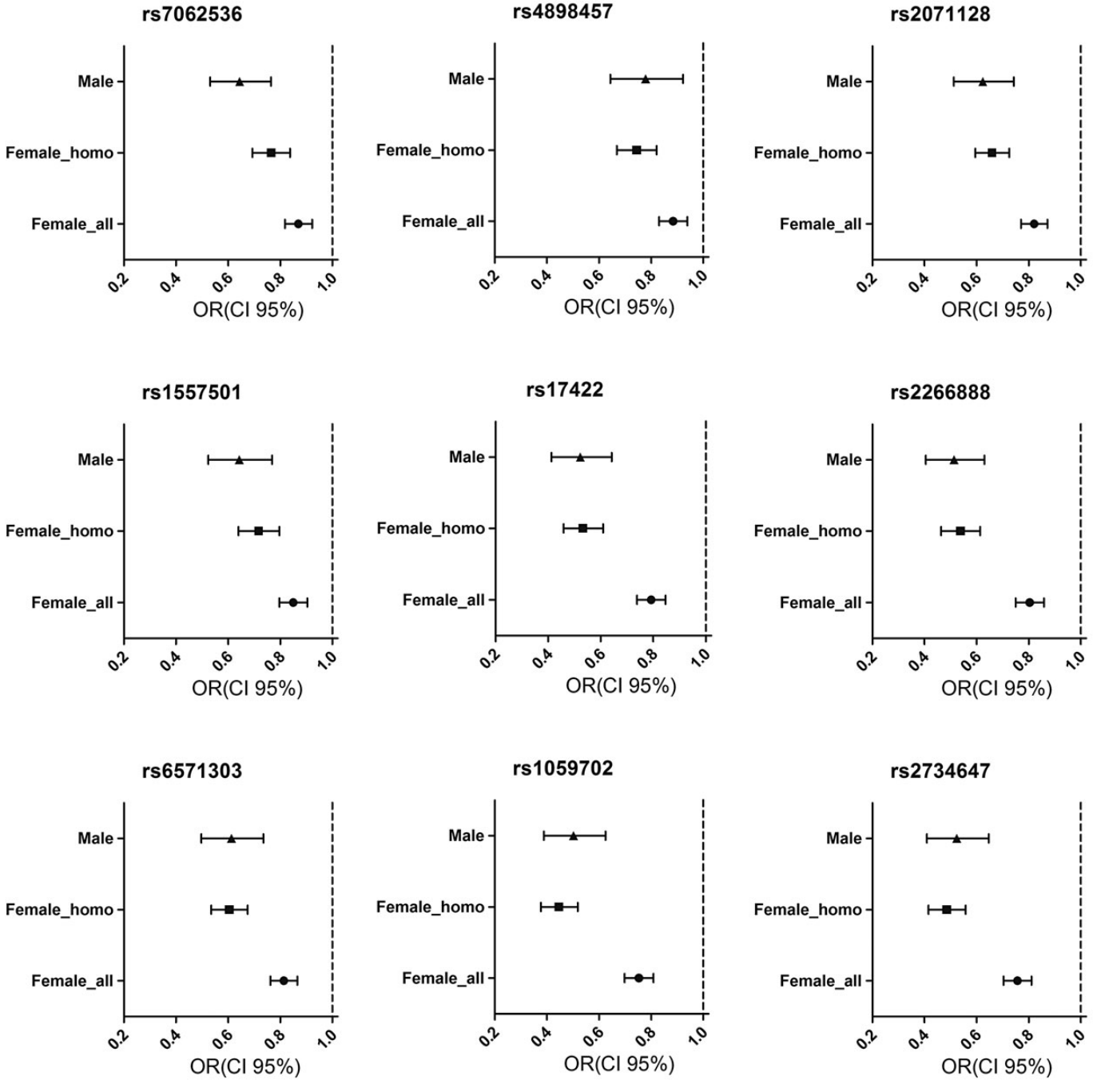

Figure 4. The OR comparison based on gender (Male and Female) and by separating females into homozygous risk allele genotypes (Female_homo) and all females (Female_all).

seem to support the notion that there are independent signals in the *L1CAM-NAA10* and *TMEM187* regions when adjusting for the effect from the *IRAK1-MECP2* region.

We examined the expression correlation of *NAA10*, *IRAK1* and *MECP2* with the replicated SNPs. Using isolated peripheral blood mononuclear cells from Hong Kong female healthy individuals, we have measured mRNA expression levels of *NAA10*, *IRAK1* and *MECP2* using real-time PCR. Females homozygous for the risk 'A' allele of SNP rs2071128 showed notably higher *NAA10* expression than those heterozygous or homozygous on the 'G' alleles ($P = 0.0183$, Supplementary Material, Fig. S7). *NAA10* encodes an N-terminal acetyltransferase that functions as the catalytic subunit of the major amino-terminal acetyltransferase A complex. Mutations in this gene are the cause of Ogden syndrome that is characterized by postnatal growth failure, global developmental delays and dysmorphic features (34). No other correlation between genotypes of the identified SNPs and these three genes was observed. We also did not observe

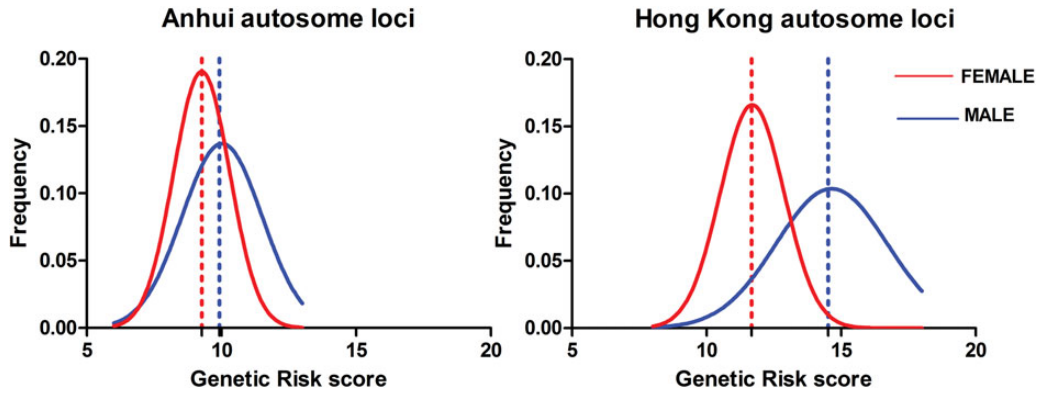

Figure 5. Distribution of cumulative genetic risk scores for SLE in male (blue) and female (red). A higher genetic risk in males than in females was observed

difference on expression levels of these three genes between males and females, consistent with the genetic findings showing no evidence of escaping dosage compensation for the susceptibility SNPs identified in this study. eQTL analysis was also performed using public databases such as mRNA by SNP Browser v 1.0.1 (35) and Genevar (36), and no statistically significant correlation was identified between these SNPs and the neighboring genes. Thus, the mechanisms for the disease association for these two regions found in this study await further investigation.

The association in males observed in this study had much larger effect sizes than those in females, which is consistent with a previous study on another X-linked susceptibility gene, *TLR7* (5). We thus evaluated the cumulative genetic risk scores genome-wide between males and females using known associated SNPs in the autosomes. Relatively higher genetic risk scores in males than those in females were observed in both Hong Kong and Anhui GWAS data (Fig. 5), which is consistent with the results from Hughes *et al.* (31). The results emphasized that a higher genetic predisposition might be required for males than for females to develop the disease. We also observed that the difference in genetic risk scores between males and females was much smaller for the Anhui cohort, as well as the total genetic risk explained by these genetic variants. Potential explanation of the differences between the two studies includes possible differences in diagnosis ascertainment and disease severity. Smaller sample size for male cases could also be a major reason for the differences observed. As shown in Supplementary Material, Table S5, the effective sample size for males was much smaller than that for females in both cohorts.

In conclusion, meta-analysis of GWAS data on Asian populations followed by replication identified association of variant in *PRPS2* and the *L1CAM-MECP2* region to SLE. We also showed evidence of association in the *L1CAM-NAA10-TMEM187* regions independent of the effect from variants in *IRAK1* and *MECP2*, a region reported in previous studies. There is no evidence of escaping XCI and dosage compensation for the SNPs identified in this study, as demonstrated by the higher effect size in males than that in females (Fig. 4). These findings indicate that X-chromosome susceptibility variants may have little contribution to the gender difference for this disease. However, it is possible that overexpression of other unknown susceptibility genes or X-linked genes without susceptibility variants could contribute to the higher female prevalence of the disease. With the increasing availability of expression data, it will be interesting to systematically compare the expression levels of X chromosome genes between males and females in relevant cells in both SLE patients and healthy individuals.

MATERIALS AND METHODS

Subjects

The samples included in the current study were collected from Hong Kong, Anhui, China, and Bangkok, Thailand (Supplementary Material, Table S4). 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 in Hong Kong: Queen Mary Hospital, Tuen Mun Hospital, Queen Elizabeth Hospital, Pamela Youde Nethersole Eastern Hospital and Princess Margaret Hospital (HK_GWAS and HK_REP). Clinical records were well documented with autoantibody profiles and subclinical symptoms. Controls from Hong Kong were individuals from other GWAS studies with disease without any overlapped manifestations with SLE in the discovery stage (HK_GWAS) and healthy blood donors from Hong Kong Red Cross in the replication stage (HK_REP). Cases from Anhui are all patients from the Department of Rheumatology at Anhui Provincial Hospital and the First Affiliated Hospital of Anhui Medical University in Hefei, Anhui province (AH_REP1), with corresponding controls from healthy blood donors in Anhui (AH_REP1). The cases for the Anhui GWAS (AH_GWAS) and Anhui Replication Panel 2 (AH_REP2) were recruited from several hospitals in central and southern China, and the controls were carefully selected with geographically matched and clinically unrelated individuals (AHGWAS and AH_REP2). The cases from Thailand were patients visiting King Chulalongkorn Memorial Hospital (TH_REP), and we took geographically matched healthy donors as controls (TH_REP). All the individuals involved this study gave informed consent. All studies were approved by the corresponding institutional review boards from all the institutions involved.

Genome-wide genotyping and quality control

The two GWAS from Hong Kong and Anhui were conducted by Illumina 610-Quad Human Beadchip array (25,30). Quality control for X chromosome was preceded as follows: the X-linked

SNP call rate for each individual was set at 90% as a cut-off. Any sample with ambiguous gender was removed. Any SNP with missing call rate of >10% was removed. Any SNP with allele frequency in healthy females smaller than 1% was removed. Any SNP that violated Hardy–Weinberg equilibrium ($P < 10^{-4}$) was discarded. The two GWAS datasets were treated following the same procedure, resulting in 11747 X-chromosome SNPs, 1047 cases and 1205 controls from Anhui cohort, together with 11 850 X-chromosome SNPs, 612 cases and 2193 controls from Hong Kong cohort being kept for further analysis.

Imputation

X-chromosome-specific imputation was done to obtain genotypes of additional SNPs using all the samples from 1000 Genome Project (released on September2013) as the reference for Hong Kong and Anhui data. IMPUTE2.1.1 was used to perform imputation. Only SNPs with an information score of >0.9 were considered for the further analysis.

Replication of candidate SNPs

There were two replication processes in this study. In the first stage of replication, 11 SNPs with a meta-analysis *P*_meta of $<1 \times 10^{-3}$ in were selected for replication using samples that are not included in GWAS discovery stage. Together with SNPs reported in a previous paper (32), SNPs were designed to be genotyped by MassARRAY iPLEX Gold system (Sequenom) on samples in HK_REP, AH_REP1 and TH_REP. SNPs that failed to be designed on Sequenom chip were genotyped by TaqMan assay (Life technologies). Ten SNPs with consistent association signals in the first stage of replication were selected for a second replication stage on samples from AHREP2. MassARRAY iPLEX GOLD system was adopted at this stage too.

Association analysis

We used inverse variance method for the meta-analysis. First of all, we aligned each allele from different cohorts to the same strand according to the Human Genome reference sequence in NCBI (Genome Build 37.5). Then, in order to estimate joint effect of strand-aligned reference allele from two different cohorts, inverse variances were used as weights to combine the effect sizes (β_i) of corresponding allele in independent studies. As shown in Equation (1), i indicates each independent study, w_i indicates weight where inverse variances were set as weights ($w_i = 1/se^2$) and β_i refers to the effect size. In this study, METAL (22) was used to calculate joint effect size and *P*-value, based on the inverse variance method.

$$Z = \frac{\left(\sum w_i \beta_i\right) / \sum w_i}{\sqrt{1/\sum w_i}}. \quad (1)$$

Joint analysis of association, eliminating the influence of gender and population stratification, was conducted using logistic regression among different cohorts and different stages of the replication studies. These two factors were utilized as discrete covariates in logistic regression to adjust for the real *P*-value as shown in Equation (2). Here, COV1 and COV2 indicate gender and population factors, adjusting for the true value of β and corresponding *P*-value. Samples from different cohorts were marked as discrete labels (which was approximately equivalent to running association tests in each stratum separately and then meta-analyzing results using an independent-effects meta-analysis).

$$\log \frac{p_i}{1 - p_i} = \beta.ADD + b_1.COV1 + b_2.COV2 + b_0. \quad (2)$$

Tests of independent contributions toward disease associations for SNPs in a single locus were done using logistic regression too, adjusting for the effect of a specific SNP (COV_{snp}) in the same locus, while also taking into account differences among cohorts and genders as shown in Equation (3).

$$\log \frac{p_i}{1 - p_i} = \beta.ADD + b_1.COV1 + b_2.COV2 + b_3.COV_{snp} + b_0. \quad (3)$$

SNPTEST v2.5β was used to perform the logistic regression tests in this study (23).

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_{10} Likelihood error for the included variables. After each variable was entered, each of the included variables was tested to see whether the model would be better off if the variable were excluded.

Linkage disequilibrium patterns and values were obtained using HaploView (37).

Cumulative genetic risk score calculation

We examined all the SNPs in the established loci showing consistency signals in our meta-analysis results (*P*_meta $< 5 \times 10^{-8}$ or replicated by our previous study) (25–30). For each of these loci, logistic regression was performed, adjusting for the effect from the best SNP in each individual locus to find potential independent variants. In total, thirty-two SNPs from twenty-three loci were considered.

The cumulative genetic risk scores were calculated following the equation from Hughes *et al.* (31). Cumulative risk was calculated in each patient by summing the risk scores from all the 32 independent SNPs on autosomes

$$\text{Cumulative genetic risk score} = \sum_{k=1}^{i} \ln(OR_i)N_i,$$

where OR_i is odds ratio of a given effect and N_i is number of risk alleles at a given SNP site.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Winnie Lau and her team for collection of samples and clinical records for Hong Kong patients. We thank P.K. and A.C.

for their contributions to Thailand pediatric patient care and sample collections.

Conflict of Interest statement. None declared.

FUNDING

This study was partially supported by the generous donation from Shun Tak District Min Yuen Tong of Hong Kong. W.Y. and Y.L.L. thank support from Research Grant Council of the Hong Kong Government (GRF HKU783813M, HKU781709M, HKU 784611M, 17125114 and HKU 770411M). We also thank support from S. K. Yee Medical Foundation general award (to B.H.Y.C., Y.L.L. and W.Y.). X.J.Z. thanks for grant support from MOE of China (IRT-1046). S.Y. thanks for grant from National Natural Science Foundation of China (No. 81171505 and 30972727) and Pre-project of State Key Basic Research Program 973 of China (No. 2011CB512103). L.D.S. thanks grant from Pre-project of State Key Basic Research Program 973 of China (No. 2012CB722404). X.J.Z. and Y.C. thank support from National Key Basic Research Program of China (2014CB541901), Program for New Century Excellent Talents in University (NCET-12-0600) and Key Project of Chinese Ministry of Education (213018A). D.Q.Y. thanks grant support from the key program of National Natural Science Foundation of China (No. 30830089). N.H. and Y.A. thank support from the National Research University Project of CHE and the Ratchadaphiseksomphot Endowment Fund (HR1163A) and the National Research Council of Thailand; K.S. and V.S. thank grant support from Thailand Research Fund. Y.Z. and W.Y. thank support from the Small Project/seed Funding from the University of Hong Kong (Nos. 201209176205, 201309176100, 201211159049). Y.Z. was supported by Post-doctoral Fellow/Research Assistant Professor Scheme of the University of Hong Kong and Centre for Genomic Sciences.

REFERENCES

1. Huang, J.L., Hung, J.J., Wu, K.C., Lee, W.I., Chan, C.K. and Ou, L.S. (2006) Septic arthritis in patients with systemic lupus erythematosus: salmonella and nonsalmonella infections compared. *Semin. Arthritis Rheum.*, **36**, 61–67.
2. Boddaert, J., Huong, D.L., Amoura, Z., Wechsler, B., Godeau, P. and Piette, J.C. (2004) Late-onset systemic lupus erythematosus: a personal series of 47 patients and pooled analysis of 714 cases in the literature. *Medicine*, **83**, 348–359.
3. 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.
4. Sawalha, A.H., Webb, R., Han, S., Kelly, J.A., Kaufman, K.M., Kimberly, R.P., Alarcon-Riquelme, M.E., James, J.A., Vyse, T.J., Gilkeson, G.S. *et al.* (2008) Common variants within MECP2 confer risk of systemic lupus erythematosus. *PLoS One*, **3**, e1727.
5. Shen, N., Fu, Q., Deng, Y., Qian, X., Zhao, J., Kaufman, K.M., Wu, Y.L., Yu, C.Y., Tang, Y., Chen, J.Y. *et al.* (2010) Sex-specific association of X-linked Toll-like receptor 7 (TLR7) with male systemic lupus erythematosus. *Proc. Natl. Acad. Sci. USA*, **107**, 15838–15843.
6. Kaufman, K.M., Zhao, J., Kelly, J.A., Hughes, T., Adler, A., Sanchez, E., Ojwang, J.O., Langefeld, C.D., Ziegler, J.T., Williams, A.H. *et al.* (2013) Fine mapping of Xq28: both MECP2 and IRAK1 contribute to risk for systemic lupus erythematosus in multiple ancestral groups. *Ann. Rheum. Dis.*, **72**, 437–444.
7. Webb, R., Wren, J.D., Jeffries, M., Kelly, J.A., Kaufman, K.M., Tang, Y., Frank, M.B., Merrill, J., Kimberly, R.P., Edberg, J.C. *et al.* (2009) Variants within MECP2, a key transcription regulator, are associated with increased susceptibility to lupus and differential gene expression in patients with systemic lupus erythematosus. *Arthritis Rheum.*, **60**, 1076–1084.
8. Scofield, R.H., Bruner, G.R., Namjou, B., Kimberly, R.P., Ramsey-Goldman, R., Petri, M., Reveille, J.D., Alarcon, G.S., Vila, L.M., Reid, J. *et al.* (2008) Klinefelter's syndrome (47,XXY) in male systemic lupus erythematosus patients: support for the notion of a gene-dose effect from the X chromosome. *Arthritis Rheum.*, **58**, 2511–2517.
9. Lee, J., Davidow, L.S. and Warshawsky, D. (1999) Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat. Genet.*, **21**, 400–404.
10. Carrel, L. and Willard, H.F. (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*, **434**, 400–404.
11. Ozcelik, T., Uz, E., Akyerli, C.B., Bagislar, S., Mustafa, C.A., Gursoy, A., Akarsu, N., Toruner, G., Kamel, N. and Gullu, S. (2006) Evidence from autoimmune thyroiditis of skewed X-chromosome inactivation in female predisposition to autoimmunity. *Eur. J. Hum. Genet.*, **14**, 791–797.
12. Tukiainen, T., Pirinen, M., Sarin, A.P., Ladenvall, C., Kettunen, J., Lehtimaki, T., Lokki, M.L., Perola, M., Sinisalo, J., Vlachopoulou, E. *et al.* (2014) Chromosome X-wide association study identifies Loci for fasting insulin and height and evidence for incomplete dosage compensation. *PLoS Genet.*, **10**, e1004127.
13. Deng, C., Kaplan, M.J., Yang, J., Ray, D., Zhang, Z., McCune, W.J., Hanash, S.M. and Richardson, B.C. (2001) Decreased Ras-mitogen-activated protein kinase signaling may cause DNA hypomethylation in T lymphocytes from lupus patients. *Arthritis Rheum.*, **44**, 397–407.
14. Richardson, B. (2003) DNA methylation and autoimmune disease. *Clin. Immunol.*, **109**, 72–79.
15. Ballestar, E., Esteller, M. and Richardson, B.C. (2006) The epigenetic face of systemic lupus erythematosus. *J. Immunol.*, **176**, 7143–7147.
16. Hewagama, A., Gorelik, G., Patel, D., Liyanarachchi, P., McCune, W.J., Somers, E., Gonzalez-Rivera, T., Michigan Lupus, C., Strickland, F. and Richardson, B. (2013) Overexpression of X-linked genes in T cells from women with lupus. *J. Autoimmun.*, **41**, 60–71.
17. Lu, Q., Wu, A., Tesmer, L., Ray, D., Yousif, N. and Richardson, B. (2007) Demethylation of CD40LG on the inactive X in T cells from women with lupus. *J. Immunol.*, **179**, 6352–6358.
18. Koshy, M., Berger, D. and Crow, M.K. (1996) Increased expression of CD40 ligand on systemic lupus erythematosus lymphocytes. *J Clin Invest*, **98**, 826–837.
19. Sawalha, A.H. (2009) Xq28 and lupus: IRAK1 or MECP2? *Proc. Natl. Acad. Sci. USA*, **106**, E62; author reply E63.
20. Han, T.U., Cho, S.K., Kim, T., Joo, Y.B., Bae, S.C. and Kang, C. (2013) Association of an activity-enhancing variant of IRAK1 and an MECP2-IRAK1 haplotype with increased susceptibility to rheumatoid arthritis. *Arthritis Rheum.*, **65**, 590–598.
21. Carmona, F.D., Cenit, M.C., Diaz-Gallo, L.M., Broen, J.C., Simeon, C.P., Carreira, P.E., Callejas-Rubio, J.L., Fonollosa, V., Lopez-Longo, F.J., Gonzalez-Gay, M.A. *et al.* (2013) New insight on the Xq28 association with systemic sclerosis. *Ann. Rheum. Dis.*, **72**, 2032–2038.
22. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
23. Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.
24. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
25. 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.
26. 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.
27. 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.

28. 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.
29. Zhang, Y., Yang, W., Mok, C.C., Chan, T.M., Wong, R.W., Mok, M.Y., Lee, K.W., Wong, S.N., Leung, A.M., Lee, T.L. *et al.* (2011) Two missense variants in UHRF1BP1 are independently associated with systemic lupus erythematosus in Hong Kong Chinese. *Genes Immun.*, **12**, 231–234.
30. 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.
31. Hughes, T., Adler, A., Merrill, J.T., Kelly, J.A., Kaufman, K.M., Williams, A., Langefeld, C.D., Gilkeson, G.S., Sanchez, E., Martin, J. *et al.* (2012) Analysis of autosomal genes reveals gene-sex interactions and higher total genetic risk in men with systemic lupus erythematosus. *Ann. Rheum. Dis.*, **71**, 694–699.
32. 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.
33. Ahmed, M., Taylor, W., Smith, P.R. and Becker, M.A. (1999) Accelerated transcription of PRPS1 in X-linked overactivity of normal human phosphoribosylpyrophosphate synthetase. *J.Biol. Chem.*, **274**, 7482–7488.
34. Rope, A.F., Wang, K., Evjenth, R., Xing, J., Johnston, J.J., Swensen, J.J., Johnson, W.E., Moore, B., Huff, C.D., Bird, L.M. *et al.* (2011) Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. *Am. J. Hum. Genet.*, **89**, 28–43.
35. Livengood, J.A., Schmit, V.L. and Gilmore, R.D. Jr (2008) Global transcriptome analysis of Borrelia burgdorferi during association with human neuroglial cells. *Infect. Immun.*, **76**, 298–307.
36. Yang, T.P., Beazley, C., Montgomery, S.B., Dimas, A.S., Gutierrez-Arcelus, M., Stranger, B.E., Deloukas, P. and Dermitzakis, E.T. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics*, **26**, 2474–2476.
37. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.