Title	Gene-Based Meta-Analysis of Genome-Wide Association Study Data Identifies Independent Single-Nucleotide Polymorphisms in ANXA6 as Being Associated With Systemic Lupus Erythematosus in Asian Populations
Author(s)	Zhang, J; Zhang, L; Zhang, Y; Yang, J; Guo, M; Sun, L; Pan, HF; Hirankarn, N; Ying, D; Zeng, S; Lee, TL; Lau, WCS; Chan, DTM; Leung, AMH; Mok, CC; Wong, SN; Lee, KW; Ho, MHK; Lee, PPW; Chung, BHY; Chong, CC; Wong, RWS; Mok, TMY; Wong, WHS; Tong, KL; Tse, NKC; Li, XP; Avihingsanon, Y; Rianthavorn, P; Deekajorndej, T; Suphapeetiporn, K; Shotelersuk, V; Ying, SKY; Fung, SKS; Lai, WM; Garcia-Barceló, MM; Cherny, SS; Sham, PC; Cui, Y; Yang, S; Ye, DQ; Zhang, XJ; Lau, YL; Yang, W
Citation	Arthritis & Rheumatology, 2015, v. 67 n. 11, p. 2966-2977
Issued Date	2015
URL	http://hdl.handle.net/10722/217167
Rights	Arthritis & Rheumatology. Copyright © John Wiley & Sons, Inc.; This is the peer reviewed version of the following article: Arthritis & Rheumatology, 2015, v. 67 n. 11, p. 2966-2977, which has been published in final form at DOI: 10.1002/art.39275. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Arthritis & Rheumatology
DOI 10.1002/art.39275

Gene-based meta-analysis of GWAS data identifies independent SNPs in *ANXA6* as associated with SLE in Asian populations

Running title: *ANXA6* in association with SLE in Asians

[1,17]Jing Zhang, [1] Lu Zhang, [1]Yan Zhang, [1]Jing Yang, [1] Mengbiao Guo, [2]Liangdan Sun, [3]Hai-Feng Pan, [4]Nattiya Hirankarn, [1]Dingge Ying, [1]Shuai Zeng, [1]Tsz Leung Lee, [5]Chak Sing Lau, [5]Tak Mao Chan, [6]Alexander Moon Ho Leung, [7]Chi Chiu Mok, [8]Sik Nin Wong, [9]Ka Wing Lee, [1]Marco Hok Kung Ho, [1]Pamela Pui Wah Lee, [1]Brian Hon-Yin Chung, [1]Chun Yin Chong, [5]Raymond Woon Sing Wong, [5]Mo Yin Mok, [1]Wilfred Hing Sang Wong, [10]Kwok Lung Tong, [11]Niko Kei Chiu Tse, [12]Xiang-Pei Li, [13]Yingyos Avihingsanon, [14]Pornpimol Rianthavorn,[14]Thavatchai Deekajorndej, [14]Kanya Suphapeetiporn, [14]Vorasuk Shotelersuk, [10]Shirley King Yee Ying, [10] Samuel Ka Shun Fung, [11]Wai Ming Lai, [15]Maria-Mercè Garcia-Barceló, [16]Stacey S. Cherny**,** [16,19]Pak Chung Sham, [2]Yong Cui, [2]Sen Yang, [3]Dong Qing Ye, [2]Xue-Jun Zhang, [1,18 *]Yu Lung Lau, [1,19*]Wanling Yang

[1]Department of Paediatrics and Adolescent Medicine, Queen Mary Hospital, LKS Faculty of Medicine, The University of Hong Kong, 21 Sassoon Road, Hong Kong

Received: Jul 01, 2014; Revised: May 31, 2015; Accepted: Jul 02, 2015

[2]State Key Laboratory Incubation Base of Dermatology, Key Laboratory of Dermatology, Anhui Medical University, Ministry of Education, China, Hefei, Anhui 230032, China

[3]Department of Epidemiology and Biostatistics, School of Public Health, Anhui Medical University, Hefei, China

[4] Center of Excellence in Immunology and Immune-mediated Diseases, Department of Microbiology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

[5]Department of Medicine, Queen Mary Hospital, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong

[6]Department of Medicine, Queen Elizabeth Hospital, Hong Kong

[7]Department of Medicine, Tuen Mun Hospital, New Territory, Hong Kong

[8]Department of Paediatrics and Adolescent Medicine, Tuen Mun Hospital, Hong Kong

[9]Department of Medicine, Pamela Youde Nethersole Eastern Hospital, Hong Kong

[10]Department of Medicine, Princess Margaret Hospital, Hong Kong

[11]Department of Paediatrics and Adolescent Medicine, Princess Margaret Hospital, Hong Kong

[12]Department of Rheumatology, Anhui Provincial Hospital, Hefei, China

[13]Department of Medicine, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

[14]Center of Excellence for Medical Genetics, Department of Pediatrics, Faculty of

Medicine, King Chulalongkorn Memorial Hospital, Chulalongkorn University, Bangkok, Thailand

[15]Department of Surgery, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong

[16]Department of Psychiatry, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong

[17]Research Center, Eye & ENT Hospital of Fudan University, Shanghai, China

[18]Department of Pediatrics, The University of Hong Kong-Shenzhen Hospital, Shenzhen, China

[19]Centre for Genomic Sciences, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong

Correspondence should be addressed to

Yu Lung Lau, M. D.,

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) 2255 4481

Fax: (852) 2855 1523

Email: lauylung@ hku.hk

Or:

Dr. Wanling Yang

Department of Paediatrics and Adolescent Medicine

The University of Hong Kong

Tel: (852) 283-15077

Fax: (852) 281-98142

Email: yangwl@ hku.hk

Acknowledgement:

This study was partially supported by the generous donation from Shun Tak District Min Yuen Tong of Hong Kong. We thank Winnie Lau and her team for collection of samples and clinical records for Hong Kong patients. WY and YLL thank support from Research Grant Council of the Hong Kong Government (GRF 17125114, HKU783813M, HKU781709M, HKU 784611M, and HKU 770411M). We also thank support from S K Yee Medical Foundation general award (to BHYC, YLL, and WY). XJZ thanks for grant support from MOE of China (IRT-1046). SY 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). LDS thanks for 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). DQY thanks grant support from the key program

of National Natural Science Foundation of China (No. 30830089). NH was supported by grant for International Research Integration: Chula Research Scholar, Ratchadaphiseksomphot Endowment Fund. KS and VS thank the Ratchadapiseksomphot Endowment Fund of Chulalongkorn University (RES560530177-HR) and the Thailand Research Fund (RTA5680003). We thank PK and AC for their contributions to Thai pediatric patient care and sample collections.

Conflicts of interest

The authors confirm that there are no conflicts of interest.

Abstract

Background: Previous genome-wide association studies (GWAS), which were mainly based on single variant analysis, have identified many susceptibility loci for systemic lupus erythematosus (SLE). However, the genetic architecture of this complex disease is far from understood. A gene-based analysis may help identify novel loci by considering global evidence of association from a gene or a genomic region rather than focusing on evidence from individual variants. **Methods:** Based on meta-analysis results of two SLE GWASs, we performed gene- and region-based

analysis followed by replication with a total of 4,626 cases and 7,466 controls of Asian ancestry. Allelic differential expression was measured by pyrosequencing. **Results:** Over half of the reported SLE susceptibility loci showed evidence of independent effects, which is important for understanding the mechanisms of association and explaining disease heritability. *ANXA6* was detected as a novel susceptibility gene for SLE with several SNPs contributing independently to disease association. The risk allele of rs11960458 correlated significantly with increased expression of *ANXA6* in PBMCs of heterozygous healthy controls. Several other associated SNPs may also regulate *ANXA6* expression based on data from public databases. Higher expression of *ANXA6* in SLE cases was also reported previously. **Conclusion:** Our study demonstrated the merit of gene-based analysis in identifying novel susceptibility loci, especially those with independent effects, and also demonstrated a widespread presence of independent contributors in susceptibility genes for SLE.

Keywords

Gene-based analysis, GWAS, SLE susceptibility, *ANXA6*

Introduction

So far more than 50 loci were shown to have robust association with systemic lupus erythematosus (SLE) (1-23). However, its genetic basis is far from understood (24) . Several limitations may have contributed to the phenomenon of "missing heritability". Firstly, genetic variants may have different allele frequencies or linkage disequilibrium (LD) structures across major ethnicities, or even in different populations of the same ethnicity, so the pattern of disease association might be different between different populations (25), making it difficult to replicate previous findings and to predict risk. It is also possible that different variants in a gene having no LD to each other are associated with the underlying disease in different populations similar to what was shown for *IRF8* (15). Thus, replication efforts from different populations that focused on single reported SNPs may miss the association signal.

The second limitation may come from the insufficient exploration of independent effects within an individual locus. These independent contributors are often overlooked in the subsequent replication steps, and the validation of their effects may require much larger sample sizes than available in most studies. Additionally, the coverage of many genomic regions in the early stage SNP chips was also limited; hence reducing the chances of capturing independent contributors. However, there was strong support on the contribution from multiple independent effects from a

single locus to disease susceptibility. A recent study demonstrated multiple independent contributors in a number of risk loci for several complex diseases, and those independent contributors within a locus could explain more disease heritability than the single reported SNP (26). The same scenario might also be true for SLE. Two out of five loci showed evidence of independent effects in our recent study on SLE in Asian populations (15), and detection of independent contributors near an established locus was also reported recently (27).

Gene- or region-based association analysis is an approach that may improve the power of GWAS. Generally, this approach tests whether a set of SNPs in a given gene or region shows evidence of association with a trait of interest. Gene-based analysis has several attractive advantages(28). Firstly, by focusing on the association signal from the entire gene rather than the individual SNPs, it could detect associated genes with multiple independent effects of small effect size, which could have been missed by a single variant-based analysis. Besides, gene-based analysis can also alleviate the issue resulted from insufficient chip coverage of earlier GWASs (29). The effect from the untyped causal SNP might be captured more efficiently via detecting the joint effect of multiple markers that are in LD with it. In addition, this approach also reduces the multiple-testing burden of GWAS, since it only requires correction for ~20,000 genes rather than the hundreds of thousands of SNPs tested in a typical GWAS. Furthermore, the characteristics of a gene, such as its position, sequence, and function, are highly consistent across human populations. Therefore, gene-based

analysis is more likely to yield consistent results across major ethnicities, and it also provides an opportunity for understanding genetic findings at the functional level (25, 30).

Several gene-based association methods have been proposed, such as regression-based tests(25) and integration of the SNP-based test statistics or *P* values from a gene (31, 32). In the regression-based tests, all the SNPs in the genes are entered as predictor variables simultaneously, except for the redundant SNPs. This approach might suffer from low statistical power due to the high dimensionality. In addition, the requirements of raw genotype data and heavy computational demand also limited their application on a genome-wide scale. Recently, Li et al. (28) proposed GATES (Gene-Based Association Test using Extended Simes procedure) as an efficient approach for gene-based test via combination of SNP statistics. Using an extended Simes procedure, GATES integrates the *P* values of all the contributing SNPs within a gene and its up- and down-stream regions, obtaining an overall association *P* value for the gene. It could achieve comparable statistical power to other alternative gene-based analysis methods, and showed better performance when a gene has only one independent signal(28). This method has been successfully applied in the identification of variants conferring risk of schizophrenia in Han Chinese(33).

In this study, based on the meta-analysis of two SLE GWASs conducted on two Chinese populations (15), we performed an in-depth gene-based analysis by GATES and followed the findings by replication in a total of 4,626 cases and 7,466 controls in

several Asian cohorts. Our results confirmed most SLE susceptibility genes identified previously at the single SNP level, and showed widespread presence of independent effects among these genes. In addition, we identified *ANXA6* as a novel SLE susceptibility gene with multiple independently contributing variants, and the results provide new insight on the genetic architecture of this prototypic autoimmune disease.

Materials and Methods

Subjects

Samples used in this study were overlapped with our previous study (15). Briefly, samples were collected from Hong Kong (612 cases and 2193 controls in the GWAS stage, and 1027 cases and 1706 controls in the replication panel), Anhui Province, China (1044 cases and 1201 controls in the GWAS stage, and 1463 cases and 1398 controls in the replication panel) and Bangkok, Thailand (480 cases and 968 controls in the replication stage), 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. 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 performed in the University of Hong Kong, genotyped on the same

platform at the same time (GWAS stage). The samples for Anhui GWAS were obtained from several 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. SLE cases for Replication from Anhui were all self-reported Chinese ethnicity living in Anhui Province, attending Departments of Rheumatology at Anhui Provincial Hospital and the First Affiliated Hospital of Anhui Medical University in Hefei, Anhui Province, China. Controls were chosen from a pool of healthy blood donors recruited from Hefei, Anhui province, with an effort to match for the age and gender of the corresponding SLE patients. The Thai cases were SLE individuals visiting King Chulalongkorn Memorial Hospital, Bangkok. Thai controls were recruited from unrelated voluntary healthy donors from the same ethnic background and geographic area as the Thai SLE cases. All cases recruited 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

(15), both by Illumina 610-Quad Human Beadchip array. Further replication on selected SNPs was performed by TaqMan SNP genotyping method using assay-on-demand probes and primers (Applied Biosystems, Foster City, CA, USA), Catalogue nos. C__8723142_10 for rs1561535, C__29329619_10 for rs6707773, C__25934588_10 for rs3815700, C__11836916_10 for rs4958893, C__31986133_10 for rs11960458, C__29349750_10 for rs6896621, C__7513840_20 for rs868531, C__15758511_10 for rs2303028, C__3169127_20 for rs10036748. The accuracy of genotyping was confirmed by direct sequencing of PCR products for 96 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.

Association analysis

The association analysis and meta-analysis were performed as previously described (15). The quality control process included removing SNPs with low call rate (<90%), low minor allele frequencies (<1%), and violation of Hardy-Weinberg equilibrium in controls ($P \leq 10-4$), and removing samples with low call rate (<90%) and hidden relationship detected using PLINK (32). Then we applied two principle components (PCs) corrections to the whole genome association statistics on the two GWAS datasets. Two major PCs, PC1 and PC2 were added as covariates to adjust for

population stratification. PC1 and PC2 were calculated by EIGENSOFT 5.0 using the whole genome data, and SNPTEST was performed for association analysis. Then meta-analysis was done by METAL(34). Imputation was performed on Hong Kong and Anhui subjects from GWAS stage by IMPUTE2.1.1, using SNP genotypes of 286 Asians from the 1000 Genome project (June 2011 data release) as the reference. Imputed SNPs with an information score >0.9 were included for further analyses. The same quality control criteria were applied to process imputed SNPs.

Joint analysis of association, taking into consideration the differences between cohorts, was performed using Cochran-Mantel-Haenszel, and homogeneity of effect size between diverse cohorts and different stages of the study was examined by Breslow-Day test ($P_{\_het}$ in Table 2), both installed in PLINK. Test of independent effects towards disease association for SNPs in a single locus was done using logistic regression by PLINK, adjusted by the effect of the other SNPs in the same locus, also treating cohort as a covariate.

Gene-based and region-based analysis

GATES (28), which implemented in KGG (35) software package was used to perform gene-based analysis. SNPs were mapped onto genes according to gene coordinate information from NCBI GRCh37, and SNPs within 10 kilobase pairs (kb) both upstream and downstream of each gene were also included. A SNP that was located in the overlapping region of two genes was assigned into both genes. *The* *r*-square (r^2)

values from the HapMap CHB+JPT samples were used to adjust for marker dependency, and SNPs in high LD ($r^2 \geqslant 0.8$) were merged.

The construction of Gene-Based-Association *P* values was done as described previously(28). Briefly, let $P_{(1)}, \ldots, P_{(m)}$ be the ascending *p* values of *m* SNPs within a gene, the overall *P* value for the gene is calculated as follows:

$$P_G = Min\left(\frac{m_e p_{(j)}}{m_{e(j)}}\right)$$

Where m_e is the effective number of independent *P* values among the m SNPs and $m_{e(j)}$ is the effective number of independent *P* values among the top *j* SNPs.

The value of m_e is estimated to be $M - \sum_{i=1}^{M}[I(\lambda i > 1)(\lambda i - 1)]\lambda i > 0$, where *I*(x) is an indicator function and λi is the i^{th} eigenvalue of the *p* value correlation coefficient matrix $[\rho_{i,j}]$ of SNP-based statistic tests. When the SNPs were all independent, the eigenvalues were set at 1, so that m_e was equal to the number of SNPs. When all the SNPs were in absolute LD, the first eigenvalue was equal to the number of SNPs and the rest were equal to 0, so that m_e =1. For intermediate situations, the *P* value correlation coefficient ρ could be accurately approximated by a six-order polynomial function of the pair-wise allelic correlation coefficient *r*, and P_G will thus have an approximate uniform (0,1) distribution.

After correction for multiple testing, the significance level for Benjamini & Hochberg (1995) FDR test (36) to control the error rate at 0.05 on the whole genome was 2.0E-4 for the 23,411 genes tested. Briefly, the *P* values of each gene are ranked in ascending order and denoted by $P_{(1)} \ldots P_{(m)}$. Then, the Benjamini & Hochberg procedure controls the false discovery rate (α=0.05 here). The procedure works as follows: 1) For a given α, find the largest κ such that $P(\kappa) \leq \frac{\kappa}{m}\alpha$; 2) Then reject all $H_{(i)}$ for i=1,…, κ. Results

from genes with $P <$ 2.0E-4 were presented.

We designed a modification that would allow consideration of all the intergenic regions in GATES. We applied the sliding window algorithm with a window size of 200kb and 100kb overlap between two neighboring windows to let GATES treat all the intergenic regions as if they have real genes located within. This modification allowed us to analyze 43,729 autosomal regions in total (including both real genes and "artificial genes") using GATES for analysis of association.

Allele-Specific Transcription Quantification

Briefly, 43 healthy individuals heterozygous for rs11960458 were recruited for investigation of the relative *ANXA6* mRNA levels expressed from the two alleles, "T" and "C" using pyrosequencing. DNA detection was used as an internal control for normalization of the expression. Total RNA obtained from PBMCs from each individual were treated with DNAase first. Both cDNA and genomic DNA from the same individuals were amplified by PCR, and the PCR amplicons were then purified with the QIAquick PCR purification kit. The sequencing primers were designed with Pyrosequencing Assay Design Software v1.0. Reactions were then performed on a Biotage PSQ96MA machine, and allele quantification was analyzed by PSQMA 2.1 software. The ratio of T:C allelic detection was performed for both DNA and cDNA, with the ratio on DNA being used as a means of normalization on the detection

efficiency. A paired Student's *t* test was adopted to compare the expression level from the "T" and the "C" alleles for this gene.

Web resources

The URLs for data presented herein are as follows

PLINK version 1.07,http://pngu.mgh.harvard.edu/~purcell/plink/

SNAP, http://www.broadinstitute.org/mpg/snap/

Haploview, http://www.broad.mit.edu/mpg/haploview/

GHS Express, http://genecanvas.ecgene.net/uploads/ForReview/

NEXTBIO, https://www.nextbio.com/

BioGPS, http://biogps.org

SCAN, http://www.scandb.org/

RegulomeDB: http://www.regulomedb.org/

EIGENSOFT: https://github.com/DReichLab/EIG

SNPTEST: https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html

Results

Gene-based association analysis

As reflected by the results after adjusting for admixture by principle components

analysis (PCA), the two Chinese populations did not seem to have significant differences in terms of population structure. There was no substantial difference for the SNP-based *P* values before and after admixture adjustment, as majority (99.83%) of them differed within one order of magnitude (Figure S1). SNP-based *P* values after admixture adjustment were subjected to gene-based analysis using GATES. A total of 324,519 (60.44%) SNPs were mapped to one or more of the 23,411 genes. The inflation factor of gene-based association was 1.047. We examined the Quantile-Quantile plot (Q-Q plot) comparing expected and observed gene-based *P* values by removing genes in known susceptibility loci (4-6, 9, 10, 12-16, 18-20, 23, 27). After removal of these genes, we still observed an excess of association signal (Figure S2), suggesting additional associated genes for the disease. After correction for multiple testing, 83 genes reached gene-based genome-wide significance (P<2.0E-04), of which 41 genes were located in non-MHC regions. Twenty-five of the 41 genes within 18 loci were located in regions already established for SLE association in earlier studies based on single variant analysis (Table 1). Some of these genes are close to each other and located in the same LD block, thus were assigned into the same locus, such as *LRRC18* and *WDFY4*, *HIC2* and *UBE2L3*, and *BLK* and *FAM167A*.

It was shown that Ten of the eighteen known SLE susceptibility loci showed potential evidence of containing multiple independent contributors ($P_{\_adjusted}$<0.05 and $P_{\_unadjusted}$<0.001, Table 1, upper panel). The independent effects within nine of

them (except for *CLEC16A*) have already been confirmed via direct or indirect (by surrogate variants having high LD with them, r^2>0.8) tagging of at least one independent contributor by previous replication studies (4, 6, 12, 13, 15, 27, 37-40). Therefore, the results of gene-based analysis were proven consistent with what have been reported in the literature. For other potentially independent variants in the reported loci that did not reach genome-wide significance, their roles in disease association need confirmation through replication in additional samples.

The results also suggested 16 novel non-MHC genes within 13 loci might be associated with SLE. Three of them showed potential evidence of independent effects (Table S1). Thus, we performed replication on the best SNPs of these three loci, using 2,970 additional SLE cases and 4,072 controls collected from Hong Kong, Anhui and Thailand. A joint analysis of results from all three cohorts showed that an intronic variant in *ANXA6* (annexin Ⅵ) demonstrated significant SLE association (rs4958893, $P\_{combined}$=3.12E-07, Table S2), and this locus was followed up further. A flowchart describing the steps taken in the current study was shown in Figure 1.

Region-based association analysis

In order to detect association signals located in the intergenic regions, we designed a modification that would allow consideration of all the intergenic regions in a gene-based analysis (See Materials and Methods for details). After correction for multiple testing, we found additional 38 intergenic regions reaching genome-wide

significance of gene-based test. Seventeen of them were located in or near the MHC region, and the others were mostly in high LD with the know loci, or candidate loci identified by the analysis that only considered the real gene regions, such as those where *ETS1*, *TNFSF4*, *CDKN1B*, and *ANXA6* are located (Detailed results were shown in Table S3). Interestingly, with such modification, we further identified association signals from *miR146a*, *PRDM1* and *IKZF1* regions that were missed from analysis when the intergenic regions were not systematically considered.

ANXA6 locus in association with SLE

We examined the meta-analysis results in *ANXA6* region, and there were eleven SNPs that showed $P_{\_meta}$ <0.001 (Table S4 and Figure 2a), and LD patterns among them revealed that they could be grouped into five clusters (Figure S3). Imputation followed by meta-analysis was performed to obtain information on additional SNPs in this region. However, none of those imputed SNPs showed *P* values of higher significance than the interrogated SNPs. Additionally, there is no adequate evidence supporting their independent functional potentials on disease pathogenesis that would warrant additional replication. For example, none of them resulted in amino acid substitutions (Table S5, by SCAN annotation). RegulomeDB annotations were also investigated to identify SNPs in or near regulatory elements, and several such imputed SNPs were shown to have potential regulatory functions. However, these SNPs were all shown to have high LD with the genotyped SNPs (Table S5, RegulomeDB

annotation). Therefore, four additional SNPs, namely rs11960458, rs6896621, rs868531, and rs2303028, all interrogated in the GWAS stage, were selected for further replication to identify independent effects within *ANXA6*. In order to exclude the possibility that the association of *ANXA6* with SLE was dependent on the effect of the neighboring *TNIP1,* a gene that was previously reported for SLE association (13), SNP rs10036748 was also included for replication in additional samples.

Results in Table 2 showed that both rs11960458 and rs6896621 revealed suggestive significance for SLE association by joint analysis of data from all three cohorts ($P\_{combined}$ =5.66E-07 and 6.72E-06, respectively). The association of SNP rs10036748 in *TNIP1* was also replicated in the current study ($P\_{combined}$ =1.94E-10). For SNPs rs4958893, rs11960458, rs686921 and rs10036748, the same trends were observed in all the replication cohorts (Figure S4). Importantly, independence tests showed that SNPs rs4958893, rs11960458, and rs6896621 remained significant when the effect from rs10036748 in *TNIP1* was adjusted for (shown as conditional P^a). Furthermore, independent effects also existed among these four SNPs, since each SNP remained significant when the effects from the other three SNPs were all adjusted for (shown as conditional P^b). Consistent with the results of independence tests, these SNPs had low to moderate LD between each other ($r^2 < 0.4$, Figure 2b). In addition, the extended association down to *GPX3* observed from the data might be due to the *TNIP1* association, as indicated by the LD pattern between rs10036748 and the SNPs in this region (Figure S5). SNPs in *GPX3* were in moderate to high LD with

rs10036748, and their *P* values were no longer significant when the effect from rs10036748 was adjusted for by logistic regression (Shown in Table S6).

ANXA6 expression profile and eQTL analysis

In order to detect potential functions of the SNPs in transcription regulation, we queried the publicly available expression quantitative trait locus (eQTL) database GHS Express, which assessed the transcriptome of circulating monocytes in 1,490 unrelated individuals (41). We found that rs4958893 was in absolute LD (r^2=1) with rs868641, a cis eQTL for *ANXA6* (*P*=3.41E-11). Additionally, the regulatory effect of rs868641 on *ANXA6* expression in monocytes has also been confirmed by a recent study on cell type-specific eQTL analysis (42). SNP rs11960458 was also in high LD (r^2=0.94) with another cis eQTL for *ANXA6* (rs4958891, *P*=1.41E-10). Detailed results of the eQTL analyses were shown in Table S7-S8.

Interestingly, SNP rs11960458 is located in the 3'UTR of *ANXA6* and we speculated that it may affect the expression level of *ANXA6* or is in high LD with a variant that does. Therefore, we examined allelic differential expression of *ANXA6* from the two alleles of rs11960458 in peripheral blood mononuclear cells (PBMCs) from healthy individuals of Hong Kong Han Chinese population who are heterozygous on the SNP (n=43). The risk allele (T) was found to be significantly correlated with a higher level

of *ANXA6* expression (*P*=0.0017, Figure 3). The result suggested that rs11960458, or SNP(s) in high LD with it, might play a regulatory role and may predispose the risk allele carriers to SLE through increased expression of *ANXA6*.

Discussion

In this study, using gene-based test of meta-analysis results from two Asian GWASs, we confirmed multiple SLE susceptibility genes that were established before, and showed the widespread presence of independent effects among known susceptibility genes for SLE. Importantly, we identified *ANXA6* as a novel susceptibility gene for SLE, which contained several independent contributors. *ANXA6* did not contain a single SNP that could potentially reach genome-wide significance based on the current sample size, their effect size and allele frequency. Thus, the association of this gene would not have been recognized without application of an alternative approach such as the gene-based association analysis. We also identified rs11960458 as a regulatory SNP (rSNP) with its risk allele correlating with higher expression of *ANXA6*, providing a functional interpretation for the genetic association.

ANXA6 showed high expression level in immune-related cells, and increased expression of *ANXA6* was also observed in SLE cases (Figure S6-S7). In addition, ANXA6 protein was shown to interact with human complement receptor 2

(CR2/CD21) in an earlier study (43). CR2 is the receptor for complement C3Dd and Epstein-Barr virus (EBV) binding on B and T cells, and plays a central role in the immune responses to foreign antigens and the development of autoimmunity to nuclear antigens in SLE (44). EBV has been considered as an important environmental trigger for SLE (45, 46). Additionally, genetic variations in CR2 were found to be associated with SLE (47, 48). Thus, the potential functional link between ANXA6 and CR2 supported the genetic findings on *ANXA6* as a novel SLE susceptibility gene.

Furthermore, two variants in *ANXA6* (SNP rs4958893, and rs11960458) were in high LD with eQTLs that affecting *ANXA6* expression. SNP rs11960458 , which is located in the 3'UTR of *ANXA6*, showed evidence of allelic differential expression with the risk allele correlating with a higher expression of *ANXA6*. These data indicate that multiple mechanisms may be involved in the disease association of this locus. Interestingly, rs11960458 was also mapped onto a miRNA binding site of *ANXA6* in Sanger's miRBase (http://www.mirbase.org/). Destruction of the miRNA binding site (such as for miR-1234-3p and miR-4479) by this SNP might confer increased expression/stability although the detailed mechanisms still require further investigation. A summary of the current findings and the potential link between *ANXA6* with autoimmunity were shown in Figure 4.

The current study also suffers from several limitations and the exploration of SLE susceptibility genes is by no means complete. Genes like *ARID5B, ITGAM*, and

DRAM1, whose associations have been identified in our previous studies, did not show strong evidence of association when analyzed at the gene level. One of the underlying reasons is the definition of gene boundary. As reported by the ENCODE (The Encyclopedia of DNA Elements) project, the majority of the regulatory elements are located close to the gene although distant ones do exist (49), herein, the gene boundary was defined as 10kb up- and down-stream of each gene in the current study. The range of gene boundary, which determines the number of SNPs to be analyzed for the gene, may affect the results of gene-based associations. It is also important to recognize that there is a tradeoff for the gain in power for this method, since it may mask or dilute individual SNP effects when the SNPs in a gene are considered altogether, especially for large genes, such as *ARID5B*, *DRAM1*, and *ITGAM* as mentioned above. The lack of convincing signal in *ITGAM* might also be attributable to the low allele frequency of risk SNPs in this gene in Asian populations (MAF<5%, (11)).

In addition, a gene-based analysis can only cover SNPs within or near a gene, and some intergenic SNPs may also be associated with complex diseases and could be missed by this approach. In order to solve this issue, we designed an in-house algorithm to consider all the intergenic regions larger than 10kb in a gene-based association analysis. We found that most of the intergenic regions that showed significant association signals were those in high LD with the known association loci or candidate loci identified by the analysis only considered the real genes.

Interestingly, with such modification, we successfully identified the association signals from *miR146a*, *PRDM1* and *IKZF1*, which were not prominent by the gene-based method using a strict definition of genes. This region-based approach might find its usefulness in applying in the studies of other complex diseases. We have made this modification available from our laboratory website (paed.hku.hk/genome).

Recently, a re-sequencing-based study was conducted on 24,892 subjects with six autoimmune disease phenotypes and 17,019 controls (50). It concluded that many common variants with weak effect might have significant contribution to the missing heritability for complex diseases. These results support the merit of using gene-based approach for exploring independent effects in susceptibility loci for complex diseases. Additionally, with the support of many integrated bioinformatics databases, the application of gene-based approach could be further extended to explore novel biological pathways and protein-protein interaction networks that may play an important role in complex diseases.

Reference

1. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. The New England journal of medicine. 2007;357(10):977-86.
2. Cunninghame Graham DS, Graham RR, Manku H, Wong AK, Whittaker JC, Gaffney PM, et al. Polymorphism at the TNF superfamily gene TNFSF4 confers susceptibility to systemic lupus erythematosus. Nature genetics. 2008;40(1):83-9.
3. Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM, et al.

Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Nature genetics. 2008;40(9):1059-61.

4. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nature genetics. 2008;40(2):204-10.

5. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, et al. Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. The New England journal of medicine. 2008;358(9):900-9.

6. Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, Sanchez E, et al. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nature genetics. 2008;40(2):211-6.

7. Musone SL, Taylor KE, Lu TT, Nititham J, Ferreira RC, Ortmann W, et al. Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. Nature genetics. 2008;40(9):1062-4.

8. Nath SK, Han S, Kim-Howard X, Kelly JA, Viswanathan P, Gilkeson GS, et al. A nonsynonymous functional variant in integrin-alpha(M) (encoded by ITGAM) is associated with systemic lupus erythematosus. Nature genetics. 2008;40(2):152-4.

9. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nature genetics. 2009;41(11):1228-33.

10. Jacob CO, Zhu J, Armstrong DL, Yan M, Han J, Zhou XJ, et al. Identification of IRAK1 as a risk gene with critical role in the pathogenesis of systemic lupus erythematosus. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(15):6256-61.

11. Yang W, Zhao M, Hirankarn N, Lau CS, Mok CC, Chan TM, et al. ITGAM is associated with disease susceptibility and renal nephritis of systemic lupus erythematosus in Hong Kong Chinese and Thai. Human molecular genetics. 2009;18(11):2063-70.

12. Yang W, Shen N, Ye DQ, Liu Q, Zhang Y, Qian XX, et al. Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. PLoS genetics. 2010;6(2):e1000841.

13. Han JW, Zheng HF, Cui Y, Sun LD, Ye DQ, Hu Z, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nature genetics. 2009;41(11):1234-7.

14. Yang J, Yang W, Hirankarn N, Ye DQ, Zhang Y, Pan HF, et al. ELF1 is associated with systemic lupus erythematosus in Asian populations. Human molecular genetics. 2011;20(3):601-7.

15. Yang W, Tang H, Zhang Y, Tang X, Zhang J, Sun L, et al. Meta-analysis Followed by Replication Identifies Loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as Associated with Systemic Lupus Erythematosus in Asians. American journal of human genetics. 2013;92(1):41-51.

16. Okada Y, Shimane K, Kochi Y, Tahira T, Suzuki A, Higasa K, et al. A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus eyrthematosus in Japanese. PLoS genetics. 2012;8(1):e1002455.
17. Lee YH, Bae SC, Choi SJ, Ji JD, Song GG. Genome-wide pathway analysis of genome-wide association studies on systemic lupus erythematosus and rheumatoid arthritis. Molecular biology reports. 2012;39(12):10627-35.
18. Cunninghame Graham DS, Morris DL, Bhangale TR, Criswell LA, Syvanen AC, Ronnblom L, et al. Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. PLoS genetics. 2011;7(10):e1002341.
19. Lessard CJ, Adrianto I, Kelly JA, Kaufman KM, Grundahl KM, Adler A, et al. Identification of a systemic lupus erythematosus susceptibility locus at 11p13 between PDHX and CD44 in a multiethnic study. American journal of human genetics. 2011;88(1):83-91.
20. Lessard CJ, Adrianto I, Ice JA, Wiley GB, Kelly JA, Glenn SB, et al. Identification of IRF8, TMEM39A, and IKZF3-ZPBP2 as susceptibility loci for systemic lupus erythematosus in a large-scale multiracial replication study. American journal of human genetics. 2012;90(4):648-60.
21. Deng Y, Zhao J, Sakurai D, Kaufman KM, Edberg JC, Kimberly RP, et al. MicroRNA-3148 modulates allelic expression of toll-like receptor 7 variant associated with systemic lupus erythematosus. PLoS genetics. 2013;9(2):e1003336.
22. Kaufman KM, Zhao J, Kelly JA, Hughes T, Adler A, Sanchez E, et al. Fine mapping of Xq28: both MECP2 and IRAK1 contribute to risk for systemic lupus erythematosus in multiple ancestral groups. Annals of the rheumatic diseases. 2013;72(3):437-44.
23. Armstrong DL, Zidovetzki R, Alarcon-Riquelme ME, Tsao BP, Criswell LA, Kimberly RP, et al. GWAS identifies novel SLE susceptibility genes and explains the association of the HLA region. Genes and immunity. 2014;15(6):347-54.
24. Surakka I, Isaacs A, Karssen LC, Laurila PP, Middelberg RP, Tikkanen E, et al. A genome-wide screen for interactions reveals a new locus on 4p15 modifying the effect of waist-to-hip ratio on total cholesterol. PLoS genetics. 2011;7(10):e1002333.
25. Neale BM, Sham PC. The future of association studies: gene-based analysis and replication. American journal of human genetics. 2004;75(3):353-62.
26. Ke X. Presence of multiple independent effects in risk loci of common complex human diseases. American journal of human genetics. 2012;91(1):185-92.
27. Zhang J, Zhang Y, Yang J, Zhang L, Sun L, Pan HF, et al. Three SNPs in chromosome 11q23.3 are Independently Associated with Systemic Lupus Erythematosus in Asians. Human molecular genetics. 2013.
28. Li MX, Gui HS, Kwan JS, Sham PC. GATES: a rapid and powerful gene-based association test using extended Simes procedure. American journal of human genetics. 2011;88(3):283-93.
29. Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, et al. Powerful SNP-set analysis for case-control genome-wide association studies. American journal

of human genetics. 2010;86(6):929-42.
30. Cheung CL, Sham PC, Xiao SM, Bow CH, Kung AW. Meta-analysis of gene-based genome-wide association studies of bone mineral density in Chinese and European subjects. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2012;23(1):131-42.
31. Curtis D, Vine AE, Knight J. A simple method for assessing the strength of evidence for association at the level of the whole gene. Adv Appl Bioinform Chem. 2008;1:115-20.
32. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007;81(3):559-75.
33. Wong EH, So HC, Li M, Wang Q, Butler AW, Paul B, et al. Common Variants on Xq28 Conferring Risk of Schizophrenia in Han Chinese. Schizophrenia bulletin. 2013.
34. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.
35. Li MX, Sham PC, Cherny SS, Song YQ. A knowledge-based weighting framework to boost the power of genome-wide association studies. PloS one. 2010;5(12):e14480.
36. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological. 1995;57(1):289-300.
37. Yang W, Ng P, Zhao M, Hirankarn N, Lau CS, Mok CC, et al. Population differences in SLE susceptibility genes: STAT4 and BLK, but not PXK, are associated with systemic lupus erythematosus in Hong Kong Chinese. Genes and immunity. 2009;10(3):219-26.
38. Zhang Y, Yang W, Mok CC, Chan TM, Wong RW, Mok MY, et al. Two missense variants in UHRF1BP1 are independently associated with systemic lupus erythematosus in Hong Kong Chinese. Genes and immunity. 2011;12(3):231-4.
39. Zhao H, Yang W, Qiu R, Li J, Xin Q, Wang X, et al. An intronic variant associated with systemic lupus erythematosus changes the binding affinity of Yinyang1 to downregulate WDFY4. Genes and immunity. 2012;13(7):536-42.
40. Agik S, Franek BS, Kumar AA, Kumabe M, Utset TO, Mikolaitis RA, et al. The autoimmune disease risk allele of UBE2L3 in African American patients with systemic lupus erythematosus: a recessive effect upon subphenotypes. The Journal of rheumatology. 2012;39(1):73-8.
41. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, et al. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. PloS one. 2010;5(5):e10693.
42. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, et al. Genetics of gene expression in primary immune cells identifies cell type-specific

master regulators and roles of HLA alleles. Nature genetics. 2012;44(5):502-10.
43. Frade R, Gauffre A, Hermann J, Barel M. EBV/C3d receptor (CR2) interacts by its intracytoplasmic carboxy-terminal domain and two distinct binding sites with the p53 anti-oncoprotein and the p68 calcium-binding protein. Journal of immunology. 1992;149(10):3232-8.
44. Asokan R, Hua J, Young KA, Gould HJ, Hannan JP, Kraus DM, et al. Characterization of human complement receptor type 2 (CR2/CD21) as a receptor for IFN-alpha: a potential role in systemic lupus erythematosus. Journal of immunology. 2006;177(1):383-94.
45. Tsokos GC, Magrath IT, Balow JE. Epstein-Barr virus induces normal B cell responses but defective suppressor T cell responses in patients with systemic lupus erythematosus. Journal of immunology. 1983;131(4):1797-801.
46. Kang I, Quan T, Nolasco H, Park SH, Hong MS, Crouch J, et al. Defective control of latent Epstein-Barr virus infection in systemic lupus erythematosus. Journal of immunology. 2004;172(2):1287-94.
47. Wu H, Boackle SA, Hanvivadhanakul P, Ulgiati D, Grossman JM, Lee Y, et al. Association of a common complement receptor 2 haplotype with increased risk of systemic lupus erythematosus. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(10):3961-6.
48. Boackle SA, Holers VM, Chen X, Szakonyi G, Karp DR, Wakeland EK, et al. Cr2, a candidate gene in the murine Sle1c lupus susceptibility locus, encodes a dysfunctional protein. Immunity. 2001;15(5):775-85.
49. Consortium EP, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.
50. Hunt KA, Mistry V, Bockett NA, Ahmad T, Ban M, Barker JN, et al. Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. Nature. 2013.

Figure legends

Figure 1. Flowchart of the experimental process and SNP selection criteria.

Figure 2. Associated SNPs in *TNIP1-ANXA6* region and LD patterns among these SNPs in the current study. (a) SNPs are plotted by chromosomal position (NCBI genome build 36.3; *x* axis) and meta-analysis *P* values (-log10*P* value; *y* axis). SNP rs4958893, having the strongest signal of *ANXA6*, is labeled in the plot, and other SNPs are colored to reflect their LD with rs4958893 (based on pairwise r^2 values from the meta-analysis data). The reported SNP in *TNIP1* (rs10036748) is underlined. Estimated recombination rates (from HapMap CHB+JPT samples) are plotted to reflect the local LD structure. The figure was generated using SNAP. (b) Shown are r^2 values for CHB+JPT individuals in HapMap3 Release 2. Stars point to the reported SNP in *TNIP1* gene.

Figure 3. Allelic differential expression of *ANXA6* based on SNP rs11960458 in PBMCs of healthy individuals. PBMC cDNA and DNA prepared from 43 healthy individuals heterozygous on rs11960458 were used for allelic differential expression analysis of *ANXA6* by pyrosequencing. Shown are the ratio of T/C allelic detection for both DNA and cDNA. The median T/C ratio for DNA is 1.08 (95% CI: 1.07-1.11) and the median T/C ratio for cDNA expression is 1.18 (95% CI: 1.15-1.29), P=0.0017 by paired student's *t* test on the difference of the ratios.

Figure 4. Summary of association result of *ANXA6* with SLE and its potential functional link with autoimmunity.

Table 1. Gene-based association analysis confirms association of the known SLE susceptibility loci and reveals independent variants.

Known SLE susceptibility loci with evidence of independent effects									
Chr	Gene	$P_{\_Gene}$	Best reported SNP	$P_{\_Best}$	Candidate Independent SNPs	$P_{\_adjusted}$	$P_{\_unadjusted}$	Reported independent SNPs	Marker dependency*
1	*TNFSF4*	1.79E-08	rs2205960	6.97E-14	rs1418190	7.10E-05	1.30E-05	rs10798269 *(Ref.(4, 13))*	r^2=0.83
					rs1418191	1.90E-04	5.98E-05		
					rs1539259	2.08E-04	6.28E-05		
					rs1578624	2.52E-04	8.08E-05		
					rs704840	1.91E-02	9.77E-12	rs704840 *(Ref.(12))*	Direct tagging
2	*STAT4*	1.04E-19	rs7574865	6.22E-21	rs16833239	2.83E-05	5.02E-10		
					rs7594501	3.56E-05	6.96E-10		
					rs3771327	6.08E-05	2.95E-08		
					rs1400656	7.16E-05	5.08E-09		
					rs7601754	1.33E-04	2.43E-10	rs7601754 *(Ref.(13, 37))*	Direct tagging
					rs7572482	9.25E-04	0.004297		
					rs10931481	1.14E-03	1.07E-06		
					rs1517352	2.09E-02	6.59E-06		
					rs6740131	3.19E-02	6.19E-05		
2	*TET3*	2.25E-06	rs6705628	3.85E-07	rs4852324	8.71E-03	3.29E-08	rs4852324 *(Ref.(15)*)	Direct tagging
					rs940296	8.64E-03	2.10E-08	rs4852324 *(Ref.(15))*	r^2=0.92
4	*BANK1*	4.10E-05	rs4522865	9.26E-07	rs10516487	2.27E-02	2.76E-05	rs10516487 *(Ref.(6, 12))*	Direct tagging
6	*UHRF1BP1*	4.26E-12	rs13205210	5.51E-12	rs3734266	2.38E-02	7.17E-13	rs11755393 *(Ref.(38))*	Direct tagging
7	*IRF5*	2.88E-09	rs4728142	7.80E-09	rs729302	1.48E-06	8.84E-10	rs729302 *(Ref.(12, 13))*	Direct tagging
10	*LRRC18_WDFY4*	1.47E-06	rs1913517	2.52E-06	rs877819	2.03E-04	2.21E-08	rs877819 *(Ref.(13, 39))*	Direct tagging

					rs2943244 rs2663052	1.56E-02 3.83E-02	1.48E-06 1.27E-07	rs7097397 *(Ref.(12))*	r^2=0.81
12	*CDKN1B*	6.48E-06	rs10845606	6.05E-06	rs12822507 rs11055008	4.26E-03 1.17E-02	0.000203 6.16E-07	rs12822507 *(Ref.(15))*	Direct tagging
16	*CLEC16A*	4.29E-06	rs12599402	1.04E-07	rs17673553	1.91E-02	0.00065		
22	*HIC2_UBE2L3*	2.72E-07	rs463426	0.000289	rs5754217 rs4821112	7.51E-03 8.65E-03	2.99E-05 3.76E-05	rs5754217 *(Ref.(13, 40))*	Direct tagging

Known SLE susceptibility loci without evidence of independent effects

Chr	Gene	$P_{\_Gene}$	Best reported SNP	$P_{\_Best}$	Candidate Independent SNPs
3	*CD80*	1.33E-04	rs6804441	9.24E-06	N
5	*TNIP1*	6.75E-05	rs10036748	3.10E-06	N
6	*TNFAIP3*	8.81E-13	rs2230926	2.20E-13	N
8	*BLK_FAM167A*	1.81E-12	rs7812879	2.88E-13	N
11	*ETS1*	5.22E-05	rs6590330	1.37E-07	N
12	*SLC15A4*	3.55E-08	rs1385374	2.58E-09	N
13	*ELF1*	5.86E-05	rs7329174	6.21E-06	N
22	*YDJC*	2.61E-07	rs2298428	1.31E-07	N

$P_{\_Best}$ indicates the meta-analysis *P* value of the best SNP in each individual locus from the two SLE GWASs.

$P_{\_adjusted}$ stands for the *P* values for each candidate independent SNP, adjusted with the effect from the best reported SNP in each individual locus and potential cohort differences (using cohort as a covariate).

$P_{\_unadjusted}$ are the meta-analysis *P* values of the candidate independent SNPs based on the two SLE GWASs.

*indicates the LD between each candidate independent SNP and the reported independent SNP, shown as r^2 value in CHB+JPT populations.

$P_{\_adjusted}$ <0.05, and $P_{\_unadjusted}$<0.001 were set as the cutoff for screening candidate independent SNPs.

Table 2. Replication results on SNPs in *ANXA6* and *TNIP1*

SNP	Hong Kong *(1639/3899)*			An Hui *(2507/2599)*			Thailand *(480/968)*			$P_{\_combined}$	Combined OR(95% CI)	$P_{\_het}$	conditional P^a	conditional P^b
	Case_F	Ctrl_F	*P*_HK	Case_F	Ctrl_F	*P*_AH	Case_F	Ctrl_F	*P*_Thai					
rs10036748	0.22	0.26	1.03E-04	0.20	0.25	1.82E-07	0.31	0.33	3.50E-01	1.94E-10	0.81 (0.76-0.86)	0.2703	——	2.14E-07
rs4958893	0.27	0.31	2.12E-04	0.30	0.33	8.52E-04	0.33	0.36	1.89E-01	3.12E-07	0.86 (0.81-0.91)	0.8223	2.32E-64	0.009352
rs11960458	0.48	0.45	4.29E-04	0.48	0.45	1.16E-03	0.34	0.32	1.41E-01	5.66E-07	1.15 (1.09-1.21)	0.9518	8.51E-08	0.002212
rs6896621	0.33	0.35	4.70E-03	0.32	0.36	2.29E-04	0.40	0.41	6.32E-01	6.72E-06	0.88 (0.83-0.93)	0.4675	6.92E-15	8.15E-05
rs868531	0.46	0.42	1.62E-03	0.44	0.42	1.86E-02	0.29	0.33	5.56E-02	0.001939	1.09 (1.03-1.15)	0.00826	0.0646	
rs2303028	0.42	0.40	2.34E-02	0.38	0.37	2.08E-01	0.29	0.34	1.83E-02	0.1095	1.05 (0.99-1.11)	0.00748	0.3932	

Arabic numerals in brackets indicate number of cases and controls that used in this study

Case/Ctrl_F indicates minor allele frequencies of the SNP in cases or controls

[a] *P* values for each of the five replicated SNPs adjusting for the effect of rs10036748 from all the samples, while treating cohort as a covariate.

[b] For the four SNPs, rs10036748, rs4958893, rs11960458 and rs6896621, we show the *P* value of the SNP when adjusted by the effect of the other three SNPs.

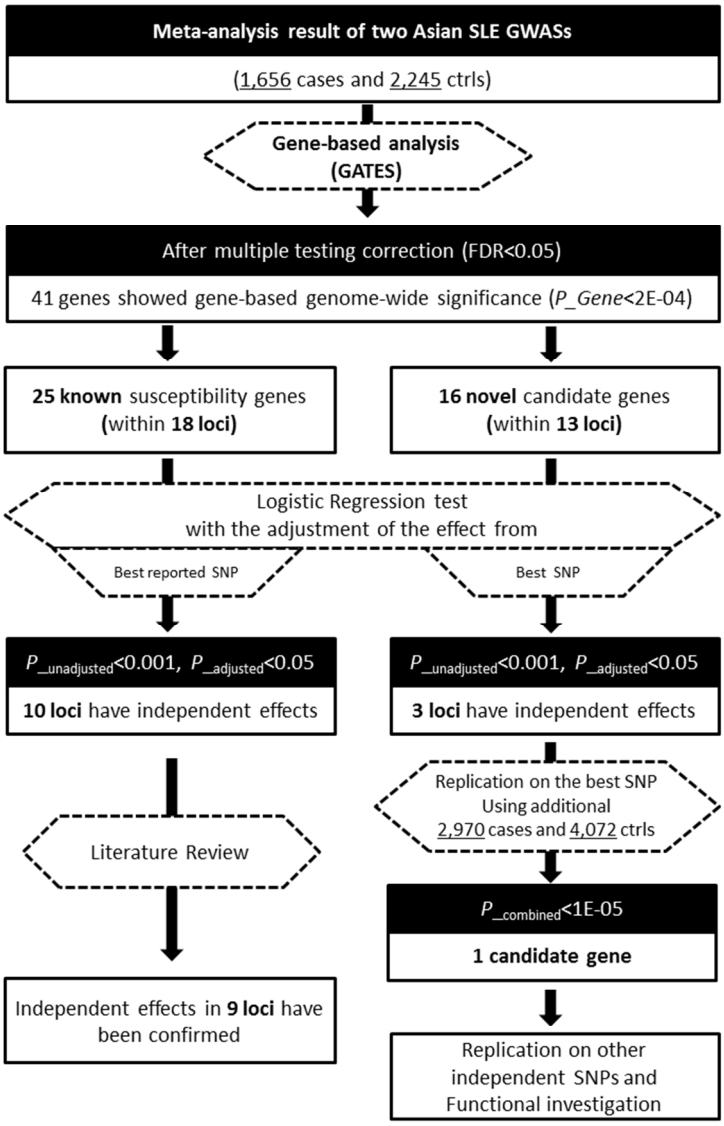

Figure 1. Flowchart of the experimental process and SNP selection criteria.
190x300mm (96 x 96 DPI)

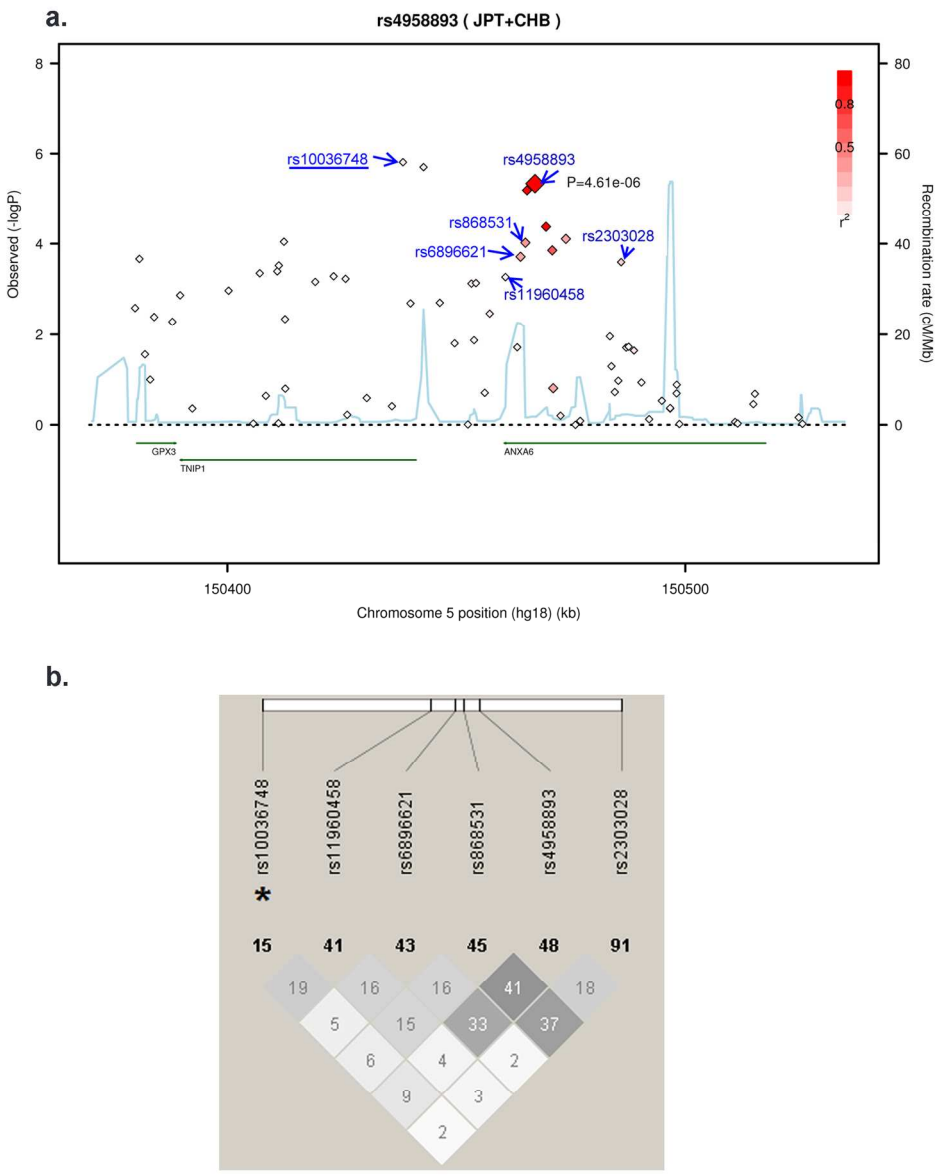

Figure 2. Associated SNPs in TNIP1-ANXA6 region and LD patterns among these SNPs in the current study. (a) SNPs are plotted by chromosomal position (NCBI genome build 36.3; x axis) and meta-analysis P values (-log10P value; y axis). SNP rs4958893, having the strongest signal of ANXA6, is labeled in the plot, and other SNPs are colored to reflect their LD with rs4958893 (based on pairwise r2 values from the meta-analysis data). The reported SNP in TNIP1 (rs10036748) is underlined. Estimated recombination rates (from HapMap CHB+JPT samples) are plotted to reflect the local LD structure. The figure was generated using SNAP. (b) Shown are r2 values for CHB+JPT individuals in HapMap3 Release 2. Stars point to the reported SNP in TNIP1 gene.
173x215mm (300 x 300 DPI)

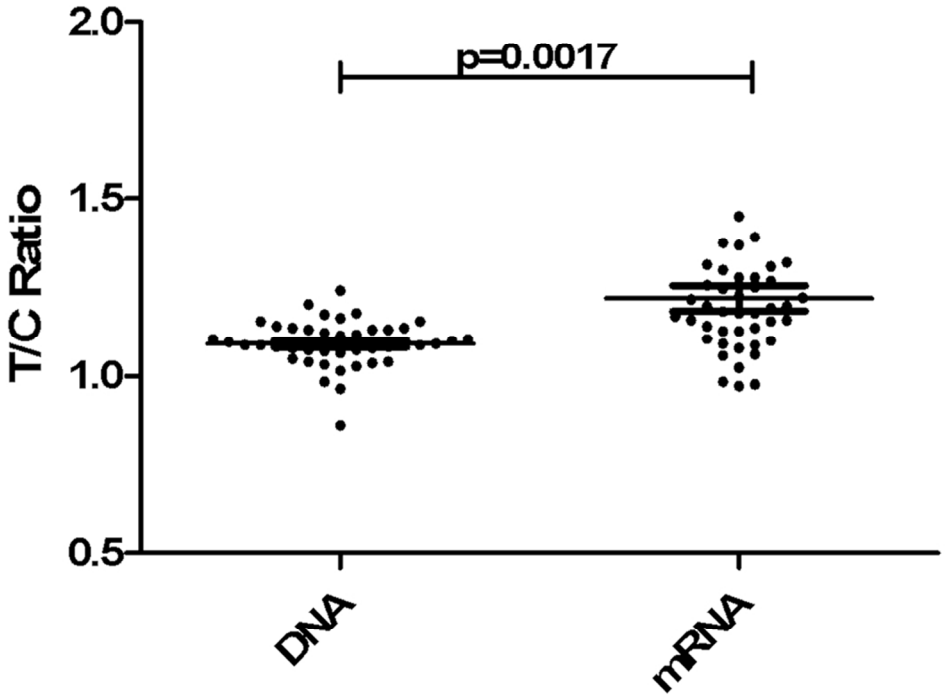

Figure 3. Allelic differential expression of ANXA6 based on SNP rs11960458 in PBMCs of healthy individuals. PBMC cDNA and DNA prepared from 43 healthy individuals heterozygous on rs11960458 were used for allelic differential expression analysis of ANXA6 by pyrosequencing. Shown are the ratio of T/C allelic detection for both DNA and cDNA. The median T/C ratio for DNA is 1.08 (95% CI: 1.07-1.11) and the median T/C ratio for cDNA expression is 1.18 (95% CI: 1.15-1.29), P=0.0017 by paired student's t test on the difference of the ratios.
95x74mm (300 x 300 DPI)

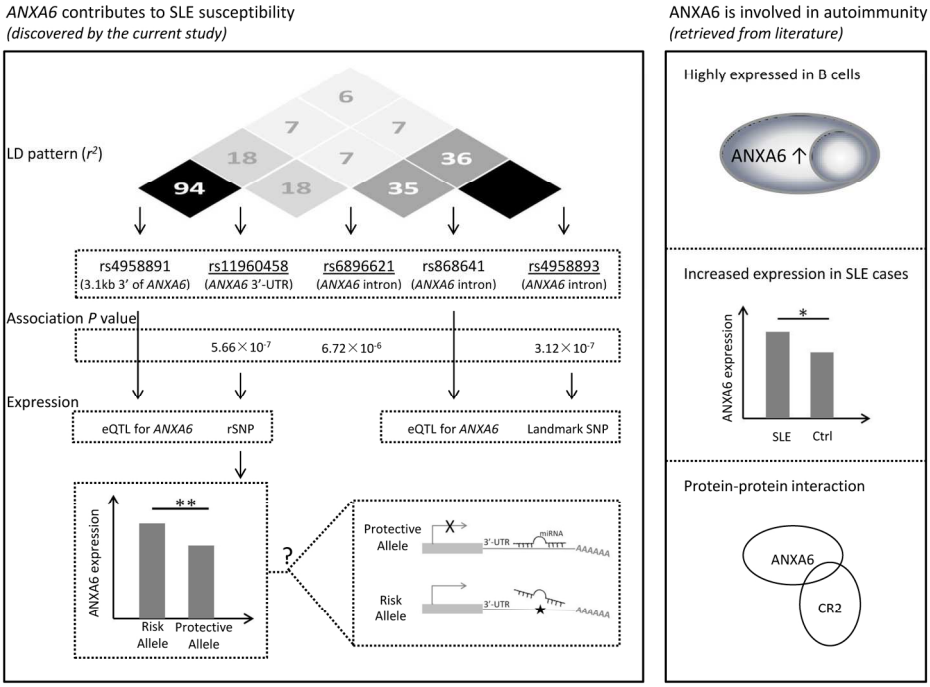

Figure 4. Summary of association result of ANXA6 with SLE and its potential functional link with autoimmunity.
170x127mm (300 x 300 DPI)