



Universiteit
Leiden
The Netherlands

Interactions between genome-wide genetic factors and smoking influencing risk of systemic lupus erythematosus

Cui, J.; Raychaudhuri, S.; Karlson, E.W.; Speyer, C.; Malspeis, S.; Guan, H.S.; ... ; Costenbader, K.H.

Citation

Cui, J., Raychaudhuri, S., Karlson, E. W., Speyer, C., Malspeis, S., Guan, H. S., ... Costenbader, K. H. (2020). Interactions between genome-wide genetic factors and smoking influencing risk of systemic lupus erythematosus. *Arthritis And Rheumatology*, 72(11), 1863-1871.
doi:10.1002/art.41414

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3184341>

Note: To cite this publication please use the final published version (if applicable).

Interactions Between Genome-Wide Genetic Factors and Smoking Influencing Risk of Systemic Lupus Erythematosus

Jing Cui,¹  Soumya Raychaudhuri,¹ Elizabeth W. Karlson,¹ Cameron Speyer,¹ Susan Malspeis,¹ Hongshu Guan,¹ Jeffrey A. Sparks,¹  Hongru Ni,¹ Xinyi Liu,¹ Emma Stevens,¹  Jessica N. Williams,¹  Emma E. Davenport,¹ Rachel Knevel,²  and Karen H. Costenbader¹ 

Objective. To identify interactions between genetic factors and current or recent smoking in relation to risk of developing systemic lupus erythematosus (SLE).

Methods. For the study, 673 patients with SLE (diagnosed according to the American College of Rheumatology 1997 updated classification criteria) were matched by age, sex, and race (first 3 genetic principal components) to 3,272 control subjects without a history of connective tissue disease. Smoking status was classified as current smoking/having recently quit smoking within 4 years before diagnosis (or matched index date for controls) versus distant past/never smoking. In total, 86 single-nucleotide polymorphisms and 10 classic *HLA* alleles previously associated with SLE were included in a weighted genetic risk score (wGRS), with scores dichotomized as either low or high based on the median value in control subjects (low wGRS being defined as less than or equal to the control median; high wGRS being defined as greater than the control median). Conditional logistic regression models were used to estimate both the risk of SLE and risk of anti-double-stranded DNA autoantibody-positive (dsDNA+) SLE. Additive interactions were assessed using the attributable proportion (AP) due to interaction, and multiplicative interactions were assessed using a chi-square test (with 1 degree of freedom) for the wGRS and for individual risk alleles. Separate repeated analyses were carried out among subjects of European ancestry only.

Results. The mean $\pm$ SD age of the SLE patients at the time of diagnosis was 36.4 ± 15.3 years. Among the 673 SLE patients included, 92.3% were female and 59.3% were dsDNA+. Ethnic distributions were as follows: 75.6% of European ancestry, 4.5% of Asian ancestry, 11.7% of African ancestry, and 8.2% classified as other ancestry. A high wGRS (odds ratio [OR] 2.0, $P = 1.0 \times 10^{-51}$ versus low wGRS) and a status of current/recent smoking (OR 1.5, $P = 0.0003$ versus distant past/never smoking) were strongly associated with SLE risk, with significant additive interaction (AP 0.33, $P = 0.0012$), and associations with the risk of anti-dsDNA+ SLE were even stronger. No significant multiplicative interactions with the total wGRS ($P = 0.58$) or with the *HLA*-only wGRS ($P = 0.06$) were found. Findings were similar in analyses restricted to only subjects of European ancestry.

Conclusion. The strong additive interaction between an updated SLE genetic risk score and current/recent smoking suggests that smoking may influence specific genes in the pathogenesis of SLE.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a complex and heterogeneous inflammatory autoimmune disease associated with

a wide range of symptoms and organ involvement. The disease primarily affects women of childbearing age and has an enormous social and financial impact (1). Complete understanding of the pathogenesis of SLE remains elusive, but both genetic and

Supported by the NIH (grants UM1-CA-186107, P01-CA-87969, R01-CA-49449, R01-HL-034594, R01-HL-088521, U01-CA-176726, R01-CA-67262, R01-AR-057327, R01-AR-057327-S1, K24-AR-066109, R01-AR-AR049880, and P30-AR-072577, and grants 1R01-AR-063759-01A1 and UM1-AI-109565 to Dr. Raychaudhuri). Dr. Knevel's work was supported by the Dutch Arthritis Foundation (grant 15-3-301). The Partners HealthCare Biobank is supported by Partners HealthCare System.

¹Jing Cui, PhD, Soumya Raychaudhuri, MD, PhD, Elizabeth W. Karlson, MD, Cameron Speyer, BS, Susan Malspeis, MS, Hongshu Guan, MS, Jeffrey A. Sparks, MD, Hongru Ni, Xinyi Liu, MS, Emma Stevens, BS, Jessica N. Williams, MD, Emma E. Davenport, PhD (current address: Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK), Karen H. Costenbader,

MD: Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; ²Rachel Knevel, MD: Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, and Leiden University Medical Center, Leiden, The Netherlands.

Dr. Sparks has received consulting fees from Bristol Myers Squibb, Gilead, Inova, Janssen, and Optum (less than \$10,000 each). No other disclosures relevant to this article were reported.

Address correspondence to Jing Cui, PhD, Brigham and Women's Hospital, Section of Clinical Sciences, Division of Rheumatology, 60 Fenwood Road, Sixth Floor, Boston, MA 02115. Email: jcui@bwh.harvard.edu.

Submitted for publication March 11, 2020; accepted in revised form June 18, 2020.

environmental factors are known to play a role (2). In a recent study using data from the Taiwan National Health Insurance Research Database, it was found that SLE heritability was estimated to be 44%, with ~26% of phenotypic variance attributed to shared environmental factors (2). The capacity to genotype numerous single-nucleotide polymorphisms (SNPs) across the genome in a robust, efficient, and cost-effective manner has facilitated comprehensive genome-wide association studies (GWAS) in SLE. SLE GWAS have, to date, identified more than 100 risk loci for SLE susceptibility across populations, adding to our understanding of its pathogenesis (3–8).

In a previous meta-analysis of epidemiologic studies, we reported that exposure to current smoking increased the risk of SLE by 50%, whereas past smoking was not associated with SLE (9). We also demonstrated that a status of current or recent smoking specifically elevated the risk of the subtype of SLE characterized by the presence of anti-double-stranded DNA (anti-dsDNA) autoantibodies (10). Interactions between genetic and environmental factors are involved in disease etiology, but in SLE, these interactions have yet to be examined across the genome. In statistical terms, a gene–environment interaction is present when the effect of genotype on disease risk depends on the level of exposure to an environmental factor, or vice versa (11). In the current study, we sought to identify gene–smoking interactions as a potential influence on SLE susceptibility. We constructed an updated weighted genetic risk score (wGRS) for SLE, utilizing data from recent large GWAS and meta-analyses, to assess for interactions between exposure to current smoking or having recently quit smoking and known genetic risk factors, in aggregate and individually, in relation to a potential risk of developing SLE.

SUBJECTS AND METHODS

Study population and collection of smoking exposure data. The Partners Healthcare Institutional Review Board approved all aspects of this study. We utilized genetic and epidemiologic data from 1,274 Nurses' Health Study (NHS) cohort participants and 2,675 Partners Healthcare Biobank (PHB) participants, including a total of 673 patients with SLE and their 3,272 matched control subjects. The NHS and NHSII are 2 prospective cohort studies of female subjects in which data on lifestyle, behavioral factors, and disease outcomes have been collected by questionnaire every 2 years (10). The NHS, which began in 1976, enrolled 121,700 registered female nurses ages 30–55 years from 11 US states. The NHSII, which began in 1989, enrolled 116,430 registered female nurses ages 25–42 years from 14 US states. The preponderance of participants in the NHS cohorts are of European ancestry, given the demographics of the nursing profession in the years of enrollment.

On biennial questionnaires, participants were asked to report whether they had received a new physician diagnosis of SLE. Those indicating having received a new SLE diagnosis were asked

to complete the Connective Tissue Disease Screening Questionnaire (12) and to consent to the release of their medical records. Released medical records of all nurses who indicated experiencing SLE symptoms on this questionnaire were independently reviewed by 3 board-certified rheumatologists (EWK, JAS, and KHC). Cases of SLE were identified based on the presence of at least 4 criteria from the American College of Rheumatology (ACR) 1997 updated criteria for the classification of SLE and also based on reviewers' consensus (13–15).

For each case in the NHS, 10 control subjects (individuals without any history of a connective tissue disease and with available GWAS data) were selected. Controls were matched to cases on the basis of age at the index date of SLE diagnosis (within 5 years), self-reported race, and genotyping platform (as detailed below). In the NHSII, 4 control subjects were similarly matched to each SLE case. Detailed smoking exposure data for all cases and controls, which included the age at the time of initiation of smoking and the updated numbers of cigarettes smoked per day at each 2-year time interval, were prospectively reported on every questionnaire. Given our past findings indicating that current smokers and those who had recently quit smoking (within 4 years before the diagnosis of SLE) had an elevated risk of SLE, whereas those who had quit smoking in the more distant past or who had never smoked did not show any elevation in disease risk, we defined a binary variable for smoking status (9,10).

The PHB includes data on self-reported smoking exposure collected at the time of enrollment, as well as blood samples, GWAS results, and linked electronic medical records for >40,000 patient volunteers from the Partners Healthcare System in the greater Boston area (16). All patients diagnosed as having SLE in the PHB and included in this study met at least 4 of the 11 ACR 1997 updated classification criteria for SLE (13–15). Cases were identified as those individuals previously included in the Brigham and Women's Hospital Lupus Registry or those with ≥ 3 International Classification of Diseases, Ninth Revision (ICD-9)/ICD-10 codes for SLE, each noted ≥ 30 days apart, followed by medical record review to identify the presence of any of the ACR 1997 criteria for SLE (13–15).

For each SLE case selected from the PHB, we matched 4 controls, each of whom had available GWAS results and PHB questionnaire data. Cases and controls were matched by age at the index date of SLE diagnosis (within 5 years), sex, and race (the first 3 genetic principle components). Patients with a history of any ICD-9/ICD-10 codes for SLE (or related diseases, including rheumatoid arthritis [RA], mixed connective tissue disease, scleroderma, or Sjögren's syndrome) were excluded from the potential control-matching pool. Each control was assigned an index date that matched the date of SLE diagnosis in the matched case. Moreover, since cases and controls were matched by age at the time of blood collection, they were of a similar age at the matched index date as well. Detailed smoking data for cases and controls, including amount and duration of smoking at different time points

in life, were collected on the PHB enrollment questionnaire. As in the NHS cohort, smoking status at or prior to the time of SLE diagnosis (or the matched index date in controls) was defined as either current smoking/recently having quit smoking within 4 years before the SLE diagnosis versus having quit smoking in the more distant past/never smoking. Missing data on smoking were collected by medical records review.

Genotyping. Since prior GWAS studies have been performed among subjects in the NHS and NHSII cohorts, a total of 6 different genotyping platforms were employed, including 5 that were used in prior studies and the Illumina MEGA chip, which was performed for the subjects with no available GWAS results. All PHB subjects were genotyped on Illumina arrays, including the Infinium MEGA array, Infinium Expanded Multi-Ethnic Genotyping Array (MEGA Ex) (both are prerelease forms of the Illumina Multi-Ethnic Global [MEG] Array), and Infinium MEG BeadChip. Standard quality control procedures were carried out on the genotype data, followed by imputation using the 1000 Genomes reference panel from the Michigan imputation.

Principal components analysis was performed and used to match the PHB cases to controls on the first 3 genetic principal components of age, sex, and race. Classic *HLA* alleles were imputed separately for the NHS/NHSII and PHB matched cases and controls, using the SNP2HLA tool (17). SNP genotyped/imputed data were extracted from the imputed data set using VCFtools (18).

Calculation of SLE wGRS. All SNPs that have been previously reported to be associated with SLE at a genome-wide significance level ($P \leq 5 \times 10^{-8}$) (see the GWAS Catalog at <https://www.ebi.ac.uk/gwas/>) were identified. SNPs in linkage disequilibrium (LD) were pruned in favor of the most significant SNP, and the remaining SNP set included 86 SNPs with a pairwise LD of $R^2 \leq 0.3$. If an imputed SNP was used, the imputation information score had to be ≥ 0.8 . As the literature supports the observation that SNPs for complex traits are often consistent across populations with similar effect directions of the alleles upon traits (19–21), and as most of our sample was of European or Asian ancestry, we included SNPs reportedly associated with SLE in populations of either European or Asian ancestry.

Our wGRS was weighted by the odds ratio (OR) of reported associations, and calculated as follows:

$$wGRS = \sum_{i=1}^{86} \ln(OR_i) \times SNP_i$$

where OR_i is the effect size reported for SNP_i , and SNP_i is the number of copies (0, 1, or 2) of the risk alleles or dosage (range 0–2) for imputed SNPs. Ten classic *HLA* alleles, previously described in a study by Langefeld et al (3), were included in the wGRS using SLE association results. The 86 SNPs and the ORs employed for

weighting our wGRS are summarized in Supplementary Table 1 (available on the *Arthritis & Rheumatology* website at <http://online.library.wiley.com/doi/10.1002/art.41414/abstract>). The 10 *HLA* alleles included from the literature and their effect sizes used for weighting are summarized in Supplementary Table 2 (available on the *Arthritis & Rheumatology* website at <http://online.library.wiley.com/doi/10.1002/art.41414/abstract>). With the use of receiver operating characteristic curve (ROC) analysis, we tested whether the wGRS based on European population-derived GWAS SLE risk SNPs enabled separation of SLE cases from controls among all our samples, and whether SNPs that were derived from GWAS in Asian populations could discriminate between cases and controls.

Main effect analyses of gene and smoking status in relation to SLE risk. We studied the associations of smoking status and the wGRS, both as a continuous variable and dichotomized as low or high (with low wGRS being defined as less than or equal to the median value in controls, and high wGRS as greater than the median value in controls), with the risk of developing SLE, utilizing conditional logistic regression in which the data were conditioned on the matched sets. In addition, we studied each SLE risk locus (86 SNPs and 10 classic *HLA* alleles) for their association with SLE risk as a secondary analysis, controlling for smoking status. Overall, smoking status and wGRS association effects in the NHS/NHSII and PHB data sets were combined using the inverse variance weighting meta-analysis method.

Gene-smoking interaction analyses. After pooling the NHS and PHB data, we first examined additive interactions between the SLE wGRS and smoking status, using the attributable proportion (AP) due to interaction (22) to assess biologic interaction. The AP tests whether subjects with both a high-risk wGRS and exposure to current/recent smoking have excess risk of SLE attributable to the interaction between the 2 factors. Using methods previously described (23–27), we calculated the AP for interaction between the dichotomous variable of low SLE wGRS versus high SLE wGRS and smoking status in relation to the risk of developing SLE. We also performed a sensitivity analysis that was restricted to only subjects of European ancestry, in whom we assessed the same variables for associations and interactions.

We next tested for multiplicative interactions between smoking status and the SLE wGRS, as well as individual SLE risk SNPs and *HLA* alleles, as they pertained to the influence on SLE risk. We assessed the multiplicative interaction term using a chi-square test (with 1 degree of freedom) in a conditional logistic regression model, adjusted for cohort. The SLE wGRS was examined as a continuous variable, as well as dichotomized into high and low wGRS, using the median value in controls as the cutoff. We also plotted the distribution of SLE wGRS scores for cases and controls separately, stratified by smoking status. We applied a false discovery rate (FDR) of <0.05 to correct for multiple comparisons. Analyses were conducted using SAS and R statistical software.

Table 1. Demographic and clinical characteristics of the SLE cases and their matched controls at the index date of SLE diagnosis*

	NHS combined cohort		PHB cohort	
	SLE cases (n = 138)	Controls (n = 1,136)	SLE cases (n = 535)	Controls (n = 2,136)
Age, mean $\pm$ SD years	51.8 $\pm$ 10.9	53.5 $\pm$ 10.3	32.4 $\pm$ 13.7	32.4 $\pm$ 13.8
Non-European ancestry	1.5	1.4	30.4	30.8
Female	100	100	90.3	90.3
Distant past/never smoker	82.6	85.1	78.1	84.2
Current/recent smoker	17.4	14.9	21.9	15.8
No. of ACR criteria met, mean $\pm$ SD†	4.7 $\pm$ 1.2	–	5.1 $\pm$ 1.6	–
ANA+	94.9	–	96.8	–
Arthritis	76.1	–	74.6	–
Hematologic involvement	53.6	–	67.3	–
Renal involvement	14.5	–	35.3	–
dsDNA+	39.1	–	64.5	–

* Except where indicated otherwise, values are the percentage of subjects. NHS combined = Nurses' Health Study I and II; PHB = Partners Healthcare Biobank; ANA+ = antinuclear antibody-positive; dsDNA+ = double-stranded DNA antibody-positive.

† According to the American College of Rheumatology (ACR) 1997 updated classification criteria for systemic lupus erythematosus (SLE) (13).

RESULTS

Characteristics of the subjects with SLE. We identified 138 SLE cases in the NHS/NHSII cohort and 535 SLE cases in the PHB cohort. GWAS results and smoking data from the time prior to SLE onset were available for all of these subjects. The characteristics of the subjects in each data set, including smoking exposures, are shown in Table 1. In the 2 NHS cohorts, 100% of subjects were female and 98% were of European ancestry. Among the subjects with SLE in the NHS/NHSII cohort, 17.4% reported being current smokers at or within 4 years of the SLE diagnosis compared to 14.9% of controls at the matched index date. Among the subjects with SLE and their matched controls in the PHB cohort, 90.3% were female, and 68.1% were of European ancestry, 4.9% were of Asian ancestry, 14.2% were of African ancestry, and 12.8% were classified as other ancestry. Among the subjects with SLE in the PHB cohort, 21.9% were current smokers or had quit within 4 years of the time of SLE diagnosis, compared to 15.8% of matched controls at a similar index date.

The clinical manifestations of SLE in each population are also summarized in Table 1. Among the subjects with SLE in both the

NHS/NHSII and PHB cohorts, all had similar numbers of criteria meeting the ACR 1997 classification criteria for SLE, while a higher proportion of SLE patients in the PHB cohort were anti-dsDNA positive and had renal disease. Furthermore, more of the SLE patients in the PHB cohort reported a status of current/recent smoking compared to those in the NHS/NHSII cohort (21.9% versus 17.9%) (Table 1).

Main effects. In our pooled case-control analyses of main effects, current/recent smoking was associated with a higher risk of SLE (OR 1.49, 95% confidence interval [95% CI] 1.20–1.85; $P = 0.0003$) (Table 2). The distribution of our updated wGRS scores in cases and controls is shown in Figure 1, which comprised 86 SNPs previously reported in genotyped or imputed SNP data sets, as well as imputed *HLA* classic alleles associated with SLE risk, as described above. Our updated wGRS, which was based on European population-derived GWAS SLE risk SNPs, enabled good differentiation of SLE cases from controls among all samples, with an ROC of 69.1% (see Supplementary Figure 1, available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41414/abstract>). The SNPs derived from GWAS

Table 2. Main effects of SLE wGRS scores and a status of current/recent smoking in relation to SLE risk*

	Beta	SE	Odds ratio	<i>P</i>
SLE wGRS				
Total wGRS	0.70	0.05	2.01	1.0×10^{-51}
Non- <i>HLA</i>	0.62	0.05	1.87	3.2×10^{-40}
<i>HLA</i> only	0.35	0.04	1.42	1.1×10^{-17}
Current/recent smoking vs. distant past/never smoking	0.40	0.11	1.49	0.00031

* The weighted genetic risk scores (wGRS) for association with risk of systemic lupus erythematosus (SLE) were assessed in total or dichotomized according to the absence or presence of *HLA* alleles. Smoking status was defined as either current smoking/having recently quit smoking versus having quit smoking in the distant past/never smoking.

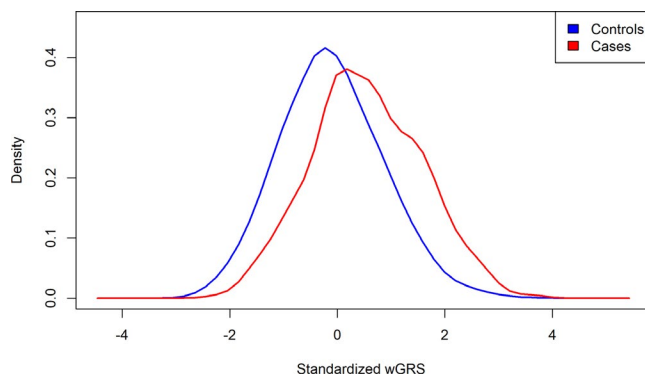


Figure 1. Distribution of standardized weighted genetic risk scores (wGRS) for the risk of systemic lupus erythematosus in pooled cases and controls.

in Asian populations also worked well to separate cases from controls, with an ROC of 62.7% (Supplementary Figure 1).

In our cohorts, the SLE wGRS was significantly associated with SLE risk, in that each standard deviation increase in the wGRS corresponded to an increase in the OR for SLE risk of 2.01 ($P = 1.0 \times 10^{-51}$) (Table 2). The association with SLE for each individual risk SNP/allele showed a similar effect size, as is reported in the GWAS Catalog. We did not observe smoking status to be a confounder in the evaluation of associations with any of the risk SNPs (data not shown). The alleles most significantly associated with SLE risk were *HLA* alleles *DRB1*0301*, *IRF5* on chromosome 7, *BLK* on chromosome 8, *STAT4* on chromosome 2, and *TNIP1* on chromosome 5. For example, *HLA-DRB1*0301* had an allele frequency of 0.18 in SLE cases compared to an allele frequency of 0.11 in controls ($P = 7.9 \times 10^{-16}$).

Additive interactions between genetic risk for SLE and current/recent smoking. SLE susceptibility was also strongly associated with a high wGRS versus a low wGRS, dichotomized using the median value among the controls. Compared to the reference group of nonsmokers in the low wGRS category, current/recently quit smokers with a low wGRS had an OR for SLE risk of 1.28 (95% CI 0.81–2.01), while nonsmokers

with a high wGRS had an OR of 3.52 (95% CI 2.78–4.44). Among those who were both current/recent smokers and in the high SLE wGRS category, the OR for SLE risk was 5.67 (95% CI 4.12–7.79).

We observed a significant additive interaction between both a high wGRS and a status of current/recent smoking, with an AP due to interaction of 33% (95% CI 13–53%; $P = 0.0012$) (Table 3). Among those who had dsDNA+ SLE, as compared to their matched controls, we observed a stronger AP due to interaction (smoking + wGRS) of 45% (95% CI 24–66%; $P = 0.00003$) (Table 3).

The strongest additive-scale interactions between individual SLE risk SNPs and current/recent smoking are reported in Table 4. Using FDRs to control for multiple comparisons, we found a significant additive interaction of the individual SLE risk SNPs *ETS1*, *FLI1*, *KIT*, *UHRF1BP1*, *BLK*, *GPR78*, *NR*, *UBE2L3*, *SMG7*, and *NCF2* with current/recent smoking.

Multiplicative interactions between genetic risk for SLE and current/recent smoking. We did not detect statistically significant multiplicative interactions between current/recently quit smoking and the continuous wGRS ($P = 0.58$) (see Supplementary Table 3, available on the *Arthritis & Rheumatology* website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41414/abstract>). However, we did observe a borderline multiplicative interaction of current/recently quit smoking with the *HLA* allele-only wGRS ($P = 0.06$), as well as with the non-*HLA* wGRS ($P = 0.1$). Similar results were found when we restricted analyses to only subjects of European ancestry ($P = 0.04$ for current/recently quit smoking with *HLA* allele-only wGRS, and $P = 0.12$ for current/recently quit smoking with non-*HLA* wGRS). Of note, *HLA-DRB1*0102*, a risk allele for SLE, showed a significant multiplicative interaction with current/recently quit smoking ($P = 0.002$). The SLE-associated SNP in *SPATA8*, rs8023715, also had a multiplicative interaction with current/recent smoking ($P = 0.004$). However, after FDR correction for multiple comparisons, these multiplicative interactions did not remain significant.

The distribution of the SLE wGRS scores among cases and controls separately, stratified by smoking status, is shown in Supplementary Figure 2 (available on the *Arthritis & Rheumatology*

Table 3. Gene–smoking interactions in relation to risk of SLE and risk of dsDNA+ SLE*

	No. of SLE cases	No. of controls	All SLE	dsDNA+ SLE
Additive interactions, wGRS + smoking				
OR (95% CI)				
Low wGRS–distant past/never smoking	126	1,371	1.0 (referent)	1.0 (referent)
Low wGRS–current/recent smoking	30	265	1.28 (0.81–2.01)	1.73 (0.93–3.21)
High wGRS–distant past/never smoking	406	1,394	3.52 (2.78–4.44)	4.64 (3.33–6.46)
High wGRS–current/recent smoking	111	242	5.67 (4.12–7.79)	9.74 (6.24–15.19)
AP due to interaction (95% CI)			0.33 (0.13–0.53)	0.45 (0.24–0.66)
<i>P</i> for interaction			0.0012	0.000027
Multiplicative interaction, wGRS × smoking, <i>P</i>			0.38	0.58

* The weighted genetic risk score (wGRS) for the risk of systemic lupus erythematosus (SLE) and risk of double-stranded DNA antibody-positive (dsDNA+) SLE includes 86 SLE single-nucleotide polymorphisms and 10 *HLA* alleles. Scores are dichotomized into low or high based on the median value among the 3,272 control subjects without a history of connective tissue disease. OR = odds ratio; 95% CI = 95% confidence interval; AP = attributable proportion.

Table 4. Strongest additive interactions between individual SLE-risk SNPs and smoking status in relation to risk of SLE*

SNP†	rs number	Gene	AP (95% CI)	P	FDR	OR for SLE risk		
						Genetic only	Smoking only	Both
11:128499000	rs7941765	<i>ETS1, FLI1</i>	0.46 (0.22–0.69)	0.00018	0.0097	0.90	1.17	1.96
4:55548475	rs2855772	<i>KIT</i>	0.38 (0.18–0.59)	0.00027	0.0097	4.63	1.34	8.07
6:35033854	rs820077	<i>UHRF1BP1</i>	0.36 (0.16–0.57)	0.00041	0.0097	3.05	1.12	4.99
8:11343973	rs2736340	<i>BLK</i>	0.41 (0.18–0.63)	0.00042	0.0097	1.40	1.21	2.72
4:8558266	rs13116227	<i>GPR78</i>	0.43 (0.18–0.67)	0.00066	0.011	1.04	1.13	2.03
11:118573519	rs4639966	<i>NR</i>	0.34 (0.14–0.53)	0.00071	0.011	4.15	1.18	6.54
22:21976934	rs7444	<i>UBE2L3</i>	0.32 (0.11–0.52)	0.0021	0.025	3.95	1.26	6.16
1:183542323	rs17849501	<i>SMG7, NCF2</i>	0.34 (0.1–0.58)	0.0048	0.045	2.54	1.36	4.40

* SLE = systemic lupus erythematosus; OR = odds ratio; AP = attributable proportion; 95% CI = 95% confidence interval; FDR = false discovery rate.

† Single-nucleotide polymorphisms (SNPs) are identified by chromosome:basepairs.

website at <http://onlinelibrary.wiley.com/doi/10.1002/art.41414/abstract>). It is to be expected that control subjects would have a lower wGRS than SLE cases, and that those with SLE who are smokers would have a lower wGRS than those with SLE who are nonsmokers, although the relatively small number of cases in the present study produced somewhat overlapping wGRS in the latter 2 groups.

DISCUSSION

In this gene–environment interaction study of SLE risk, the largest conducted to date and the first, to our knowledge, to employ GWAS results rather than to investigate individual candidate polymorphisms, we created an updated, weighted GRS based on existing GRS studies of SLE (28,29), and tested for interactions between SLE genetic factors and smoking status in relation to their potential influence on SLE risk. We found that having a high wGRS and being a current or recent smoker were each individually strongly associated with SLE risk: each standard deviation increase in the wGRS more than doubled the risk of SLE, while current smokers and those who had recently quit smoking had an increased SLE risk of ~50% compared to never and more distant past smokers. It is believed that testing for the presence of additive interaction may be more relevant for some scientific objectives (30). A number of researchers have shown that conceptual models for biologic interaction translate to the presence of interaction on the additive scale, and not necessarily on the multiplicative scale (22,27).

We discovered a significant additive interaction between a high wGRS versus low wGRS and current/recent smoking versus distant past/never smoking, in which 33% of the excess risk of SLE among smokers with a high wGRS was due to interaction. We also found a similar additive interaction between a high wGRS and current/recent smoking in relation to the risk of dsDNA+ SLE, in which 45% of the excess risk of SLE among smokers with a high wGRS was due to interaction. Moreover, we found significant additive interactions for several individual SLE-related SNPs and current/recent smoking, including *ETS1*, *FLI1*, *KIT*, *UHRF1BP1*, *BLK*, *GPR78*, *NR*, *UBE2L3*, *SMG7*, and *NCF2*.

Significant multiplicative interactions were not found using the continuous SLE wGRS, but were found for current/recent smoking and specific SLE risk genotypes, including *HLA-DRB1*0102* and *SPATA8*. However, these associations were not significant after adjusting for multiple comparisons. We suspect that the reason that no formal interaction between the wGRS and smoking in determining SLE risk was found is that we were testing for multiplicative interactions, rather than additive interactions. Multiplicative interaction analysis that tests for a significantly greater than multiplicative interaction between the 2 risk factors (on a logarithmic scale) in a logistic regression analysis is a higher bar for formal interaction testing, as compared to testing for a greater than additive interaction between the 2 risk factors (30).

Previous studies and meta-analyses demonstrated a strong association between current cigarette smoking and increased SLE risk (9,10). Among women in the NHS cohorts, we reported strong and specific associations of current smoking, particularly with the subtype of dsDNA+ SLE (hazard ratio 1.86, 95% CI 1.14–3.04), suggesting that ongoing smoking is involved in the pathogenesis of SLE and that effects diminish after cessation of smoking (10). Potential biologic mechanisms by which smoking may contribute to or accelerate SLE pathogenesis include damage to DNA and proteins, induction of oxidative stress (31), stimulation of proinflammatory cytokines such as tumor necrosis factor and IL-6, and increased CD95 expression on B cells and CD4 T cells, inducing apoptosis and inducing autoimmunity (32). Smoking also impairs natural killer cell functions and may influence inflammatory Th17 and Th22 cell functions via the aryl hydrocarbon receptor, activated by benzopyrenes in cigarette smoke. We have also found that current smoking was associated with increased levels of the SLE-specific cytokines interferon- α and B lymphocyte stimulator among women without SLE in the NHS cohorts (33).

In a related autoimmune disease, RA, there is a statistically significant multiplicative interaction as well as additive interaction between RA genetic risk and smoking, in particular with *HLA-DRB1* genes (26,34). This interaction, strongest for seropositive RA, has been replicated in several populations, including women in the US and in Swedish, Korean, and Malaysian

populations (26). The smoking–RA genetic risk interaction is thought to accelerate the citrullination of peptides at mucosal surfaces, such as in the lungs, and the development of anti-citrullinated peptide antibody–targeted autoimmunity in RA. It is possible, although not yet thoroughly investigated, that similar processes arise from smoking and genetic interactions in driving the production of SLE-specific cytokines and autoimmune reactions, accelerating pathogenesis. Our current findings of interactions between current/recent smoking and the SLE wGRS (with an AP due to interaction of 33% [95% CI 13–53%]; $P = 0.0012$) and stronger interaction for dsDNA+ SLE imply that smoking has a greater influence if these SLE genetic factors are present. Future studies should investigate the biologic mechanisms of gene–smoking interactions, such as immunologic mechanisms, anatomic site, autoantibody production, and timing prior to clinical onset, as they pertain to an influence on SLE risk.

To date, there have been only a few investigations of gene–environment interactions in SLE susceptibility, and these have been limited by small sample sizes (35,36). The present study is the first to investigate all previously identified SLE genetic risk polymorphisms and to combine them into a single wGRS. Small case–control studies in Japanese populations have reported additive interactions between candidate polymorphisms, including the SLE risk gene *TNFRSF1B* as well as the detoxification genes and genotypes *CYP1A1*, *GSTM1*, and *NAT2*, and ever/never smoking in elevating SLE risk (35–37). In a study of 152 SLE cases and 427 controls in Japan, 2 candidate genes were studied for potential interactions with smoking. An interaction between the *TNFRSF1B* risk genotypes GG or GT and smoking (ever/never) was found, with a significant AP due to interaction of 0.49 (indicating that 49% of the excess risk of SLE among smokers with a G allele was due to an interaction) (36). No interaction between *STAT4* SLE risk alleles and smoking was found. As *TNFRSF1B* has not been confirmed as an SLE risk gene in other populations and did not reach genome-wide significance, it was not included in our updated wGRS for SLE.

The same Japanese research group also studied interactions between smoking and the detoxification genes *CYP1A1*, *GSTM1*, and *NAT2* in SLE risk in the same population (37). An interaction between smoking (ever/never) and the *NAT2* slow acetylator genotype was found, with a significant AP due to interaction of 0.50 (35). Again, these genotypes were not included in our wGRS, as they have not reached genome-wide significance in past SLE studies.

The current study has several strengths. It leveraged rich, prospectively collected smoking exposure and genetic data from a large US cohort of women who have been followed up with detailed smoking exposure information from the time prior to SLE onset, as well as a large hospital-based SLE cohort with both self-reported questionnaire exposure data and linked electronic medical records. For both populations, nonaffected controls were closely matched by age, sex, and race. Importantly,

smoking exposure data were collected prospectively from the time prior to the SLE diagnosis or matched index date in controls and categorized based on smoking status and duration of cessation of smoking. All SLE cases were carefully validated by expert review of the medical records for presence of at least 4 of the ACR updated criteria for SLE and by expert reviewers' consensus. Using GWAS data, we developed an updated wGRS for SLE risk using the most comprehensive set of risk alleles currently available for SLE, derived from European and Asian populations. We investigated the interactions of current/recent smoking and genetic risk factors in determining the risk of dsDNA+ SLE (59% of our cases), given our past finding that current smoking was a stronger risk factor for this SLE subtype, and found that the additive interaction effects were stronger in that subgroup.

No SLE gene–environment studies using GWAS data have been performed to date, as the populations in which SLE GWAS have been conducted have not had detailed data on environmental exposures available. However, GWAS gene–environment interaction studies have been fruitful in many other complex diseases, in which they have helped to identify new pathologic mechanisms and synergies between genetic and environmental factors (38).

This study does have some limitations. We pooled data from 2 different SLE case–control populations. The NHS cohorts have prospectively collected data, and thus, the risk of potential misclassification of smoking status is lower than in the PHB cohort, for which past exposure data were collected from prevalent SLE cases. On the other hand, the NHS cohorts are all female and >98% of the participants are of European ancestry, potentially limiting the generalizability of the findings from that population. Thirty percent of the PHB population was of Asian, African, or Hispanic descent.

Our wGRS SNPs and *HLA* data were obtained from large European and Asian GWAS, as large SLE GWAS in populations of other ancestry have not been conducted to date (4,39–45). One recent SLE GWAS was performed in Latin Americans (predominantly those of European and Native American origin) (46). There is debate as to whether the risk alleles identified in European and Asian populations are similar to those found in populations of African or South American ancestry; studies point to many risk alleles having similar effects in different ancestry populations, whereas there are likely many more yet to be identified, given the lack of studies of diverse populations (3,47). We found strong associations with SLE risk using European population–derived SNPs in non-European populations and using Asian population–derived SNPs in European populations. When we limited our study to those of European descent and the wGRS to European-only genotypes, we observed similar main effects and interaction results.

In this large SLE case–control study of gene–environment interactions, we have found that individuals with both high SLE genetic risk (a high wGRS) and current or recent smoking had a greater than additive risk of developing SLE, compared to having only genetic risk or only smoking as a risk factor. We found a greater

than multiplicative interaction between the *HLA-DRB1*0102* SLE risk allele and current/recent smoking in influencing SLE risk ($P = 0.002$). These novel findings support the hypothesis that current smoking triggers SLE in presence of genetic factors, with a greater than additive interaction with overall genetic risk and potentially greater than multiplicative interaction with certain genotypes. Studies should now investigate the mechanisms of these interactions and whether smoking influences function of these genes in SLE pathogenesis.

ACKNOWLEDGMENTS

We thank the participants in the NHS and NHSII cohorts for their dedication and continued participation in these longitudinal studies, as well as the staff in the Channing Division of Network Medicine (Department of Medicine, Brigham and Women's Hospital) and Harvard Medical School for their assistance with this project.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Cui had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Cui, Raychaudhuri, Karlson, Speyer, Malspeis, Guan, Sparks, Ni, Liu, Stevens, Williams, Davenport, Knevel, Costenbader.

Acquisition of data. Cui, Raychaudhuri, Karlson, Speyer, Malspeis, Guan, Sparks, Ni, Liu, Stevens, Williams, Davenport, Knevel, Costenbader.

Analysis and interpretation of data. Cui, Speyer, Malspeis, Guan, Sparks, Ni, Liu, Stevens, Williams, Davenport, Knevel, Costenbader.

REFERENCES

- Lotstein DS, Ward MM, Bush TM, Lambert RE, van Vollenhoven R, Neuwelt CM. Socioeconomic status and health in women with systemic lupus erythematosus. *J Rheumatol* 1998;25:1720–9.
- Kuo CF, Grainge MJ, Valdes AM, See LC, Luo SF, Yu KH, et al. Familial aggregation of systemic lupus erythematosus and coaggregation of autoimmune diseases in affected families. *JAMA Intern Med* 2015;175:1518–26.
- Langefeld CD, Ainsworth HC, Graham DS, Kelly JA, Comeau ME, Marion MC, et al. Transancestral mapping and genetic load in systemic lupus erythematosus. *Nat Commun* 2017;8:16021.
- Bentham J, Morris DL, Graham DS, Pinder CL, Tomblinson P, Behrens TW, et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet* 2015;47:1457–64.
- 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. *Nat Genet* 2009;41:1228–33.
- Morris DL, Sheng Y, Zhang Y, Wang YF, Zhu Z, Tomblinson P, et al. Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. *Nat Genet* 2016;48:940–6.
- 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 erythematosus in Japanese. *PLoS Genet* 2012;8:e1002455.
- Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, et al, for the International Consortium for Systemic Lupus Erythematosus Genetics. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat Genet* 2008;40:204–10.
- Costenbader KH, Kim DJ, Peerzada J, Lockman S, Nobles-Knight D, Petri M, et al. Cigarette smoking and the risk of systemic lupus erythematosus: a meta-analysis. *Arthritis Rheum* 2004;50:849–57.
- Barbhaiya M, Tedeschi SK, Lu B, Malspeis S, Kreps D, Sparks JA, et al. Cigarette smoking and the risk of systemic lupus erythematosus, overall and by anti-double stranded DNA antibody subtype, in the Nurses' Health Study cohorts. *Ann Rheum Dis* 2018;77:196–202.
- Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet* 2001;358:1356–60.
- Karlson EW, Sanchez-Guerrero J, Wright EA, Lew RA, Daltroy LH, Katz JN, et al. A connective tissue disease screening questionnaire for population studies. *Ann Epidemiol* 1995;5:297–302.
- Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus [letter]. *Arthritis Rheum* 1997;40:1725.
- Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1982;25:1271–7.
- Merola JF, Bermas B, Lu B, Karlson EW, Massarotti E, Schur PH, et al. Clinical manifestations and survival among adults with systemic lupus erythematosus according to age at diagnosis. *Lupus* 2014;23:778–84.
- Karlson EW, Boutin NT, Hoffnagle AG, Allen NL. Building the Partners HealthCare Biobank at Partners Personalized Medicine: informed consent, return of research results, recruitment lessons and operational considerations. *J Pers Med* 2016;6:2.
- Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One* 2013;8:e64683.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics* 2011;27:2156–8.
- Ntzani EE, Liberopoulos G, Manolio TA, Ioannidis JP. Consistency of genome-wide associations across major ancestral groups. *Hum Genet* 2012;131:1057–71.
- Marigorta UM, Navarro A. High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS Genet* 2013;9:e1003566.
- Wang C, Ahlford A, Jarvinen TM, Nordmark G, Eloranta ML, Gunnarsson I, et al. Genes identified in Asian SLE GWASs are also associated with SLE in Caucasian populations. *Eur J Hum Genet* 2013;21:994–9.
- Andersson T, Alfredsson L, Kallberg H, Zdravkovic S, Ahlbom A. Calculating measures of biological interaction. *Eur J Epidemiol* 2005;20:575–9.
- Kim K, Jiang X, Cui J, Lu B, Costenbader KH, Sparks JA, et al. Interactions between amino acid-defined major histocompatibility complex class II variants and smoking in seropositive rheumatoid arthritis. *Arthritis Rheumatol* 2015;67:2611–23.
- Han SS, Rosenberg PS, Garcia-Closas M, Figueroa JD, Silverman D, Chanock SJ, et al. Likelihood ratio test for detecting gene (G)-environment (E) interactions under an additive risk model exploiting G-E independence for case-control data. *Am J Epidemiol* 2012;176:1060–7.

25. Kim S, Wang M, Tyrer JP, Jensen A, Wiensch A, Liu G, et al. A comprehensive gene-environment interaction analysis in ovarian cancer using genome-wide significant common variants. *Int J Cancer* 2019;144:2192–205.
26. Karlson EW, Chang SC, Cui J, Chibnik LB, Fraser PA, de Vivo I, et al. Gene-environment interaction between HLA-DRB1 shared epitope and heavy cigarette smoking in predicting incident rheumatoid arthritis. *Ann Rheum Dis* 2010;69:54–60.
27. Kallberg H, Ahlborn A, Alfredsson L. Calculating measures of biological interaction using R. *Eur J Epidemiol* 2006;21:571–3.
28. Chen L, Wang YF, Liu L, Bielowska A, Ahmed R, Zhang H, et al. Genome-wide assessment of genetic risk for systemic lupus erythematosus and disease severity. *Hum Mol Genet* 2020;29:1745–56.
29. Reid S, Alexsson A, Frodlund M, Morris D, Sandling JK, Bolin K, et al. High genetic risk score is associated with early disease onset, damage accrual and decreased survival in systemic lupus erythematosus. *Ann Rheum Dis* 2020;79:363–9.
30. Sparks JA, Costenbader KH. Genetics, environment, and gene-environment interactions in the development of systemic rheumatic diseases [review]. *Rheum Dis Clin North Am* 2014;40:637–7.
31. Pryor WA, Stone K. Oxidants in cigarette smoke: radicals, hydrogen peroxide, peroxyhydrate, and peroxyhydrate. *Ann N Y Acad Sci* 1993;686:12–27.
32. Walczak H, Krammer PH. The CD95 (APO-1/Fas) and the TRAIL (APO-2L) apoptosis systems. *Exp Cell Res* 2000;256:58–66.
33. Leatherwood C, Liu X, Malspeis S, Roberts A, Sparks JA, Karlson E, et al. Associations between current cigarette smoking and SLE-related cytokine and chemokine biomarkers among US female nurses without SLE [abstract]. *Arthritis Rheumatol* 2018;70 Suppl 10. URL: <https://acrabstracts.org/abstract/associations-between-current-cigarette-smoking-and-sle-related-cytokine-and-chemokine-biomarkers-among-u-s-female-nurses-without-sle/>.
34. Lundstrom E, Kallberg H, Alfredsson L, Klareskog L, Padyukov L. Gene-environment interaction between the DRB1 shared epitope and smoking in the risk of anti-citrullinated protein antibody-positive rheumatoid arthritis: all alleles are important. *Arthritis Rheum* 2009;60:1597–603.
35. Kiyohara C, Washio M, Horiuchi T, Tada Y, Asami T, Ide S, et al. Cigarette smoking, N-acetyltransferase 2 polymorphisms and systemic lupus erythematosus in a Japanese population. *Lupus* 2009;18:630–8.
36. Kiyohara C, Washio M, Horiuchi T, Tada Y, Asami T, Ide S, et al. Cigarette smoking, STAT4 and TNFRSF1B polymorphisms, and systemic lupus erythematosus in a Japanese population. *J Rheumatol* 2009;36:2195–203.
37. Kiyohara C, Washio M, Horiuchi T, Asami T, Ide S, Atsumi T, et al. Risk modification by CYP1A1 and GSTM1 polymorphisms in the association of cigarette smoking and systemic lupus erythematosus in a Japanese population. *Scand J Rheumatol* 2012;41:103–9.
38. Simon PH, Sylvestre MP, Tremblay J, Hamet P. Key considerations and methods in the study of gene-environment interactions. *Am J Hypertens* 2015;29:891–9.
39. 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, PTK, KIAA1542 and other loci. *Nat Genet* 2008;40:204–10.
40. 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. *N Engl J Med* 2008;358:900–9.
41. Nath SK, Han S, Kim-Howard X, Kelly JA, Viswanathan P, Gilkeson GS, et al. A nonsynonymous functional variant in integrin- α (M) (encoded by ITGAM) is associated with systemic lupus erythematosus. *Nat Genet* 2008;40:152–4.
42. Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Reddy MV, Sanchez E, et al. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nat Genet* 2008;40:211–6.
43. 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. *Nat Genet* 2009;41:1234–7.
44. 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. *Hum Mol Genet* 2009;18:2063–70.
45. 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 Genet* 2010;6:e1000841.
46. Alarcon-Riquelme ME, Ziegler JT, Moliner J, Howard TD, Moreno-Estrada A, Sánchez-Rodríguez E, et al. Genome-wide association study in an Amerindian ancestry population reveals novel systemic lupus erythematosus risk loci and the role of European admixture. *Arthritis Rheumatol* 2016;68:932–43.
47. Hanscombe KB, Morris DL, Noble JA, Dilthey AT, Tomblinson P, Kaufman KM, et al. Genetic fine mapping of systemic lupus erythematosus MHC associations in Europeans and African Americans. *Hum Molecul Genet* 2018;27:3813–24.