**ORIGINAL ARTICLE**

Genetic risk and longitudinal disease activity in systemic lupus erythematosus using targeted maximum likelihood estimation

MA Gianfrancesco, L Balzer, KE Taylor, L Trupin, J Nititham, MF Seldin, AW Singer, LA Criswell, and LF Barcellos

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease associated with genetic and environmental risk factors. However, the extent to which genetic risk is causally associated with disease activity is unknown. We utilized longitudinal-targeted maximum likelihood estimation to estimate the causal association between a genetic risk score (GRS) comprising 41 established SLE variants and clinically important disease activity as measured by the validated Systemic Lupus Activity Questionnaire (SLAQ) in a multiethnic cohort of 942 individuals with SLE. We did not find evidence of a clinically important SLAQ score difference (> 4.0) for individuals with a high GRS compared with those with a low GRS across nine time points after controlling for sex, ancestry, renal status, dialysis, disease duration, treatment, depression, smoking and education, as well as time-dependent confounding of missing visits. Individual single-nucleotide polymorphism (SNP) analyses revealed that 12 of the 41 variants were significantly associated with clinically relevant changes in SLAQ scores across time points eight and nine after controlling for multiple testing. Results based on sophisticated causal modeling of longitudinal data in a large patient cohort suggest that individual SLE risk variants may influence disease activity over time. Our findings also emphasize a role for other biological or environmental factors.

*Genes and Immunity* (2016) 17, 358–362; doi:10.1038/gene.2016.33; published online 28 July 2016

**INTRODUCTION**

Systemic lupus erythematosus (SLE) is a complex, heterogeneous autoimmune disease caused by both genetic and environmental factors. A substantial genetic component to SLE is supported by data demonstrating high heritability of the disease, a higher concordance rate of SLE in monozygotic twins than dizygotic twins or siblings and a high sibling recurrence risk ratio (that is, greater likelihood of disease, given that one’s sibling is affected, compared with disease prevalence in the general population). The strongest genetic risk factor established for SLE resides within the major histocompatibility complex (MHC) on chromosome 6, where HLA-DRB1*03:01 and other MHC variants have been strongly implicated. Recent genome-wide association studies have also identified more than 40 independent loci related to SLE onset. A weighted genetic risk score (GRS) comprising these variants was significantly higher on average in SLE cases compared with healthy controls. However, measures of association between genetic variants and SLE are generally modest (with odds ratios ranging from 1.1 to 2.3), gene–environment interactions are poorly understood and there is little awareness of how genetic profiles have an impact on disease activity.

Autoimmune diseases such as SLE are commonly treated with immunosuppressive therapy, and large efforts are dedicated to the development of targeted therapies such as biologic agents. In order to accurately determine the effect of treatment, it is critical to understand the factors that influence variation in disease activity, including genetic susceptibility. Whereas genetic risk has been demonstrated to be associated with disease onset, the extent to which it is associated with disease activity over time has not been extensively studied. Previous studies have found significant associations between individual gene variants and more severe clinical SLE characteristics, such as production of antibodies against double-stranded DNA (ds-DNA), nephritis and early age at diagnosis. One study found that a weighted GRS comprising 22 SLE risk loci demonstrated significant associations with ds-DNA, immunologic disorder, renal disorder, hematologic disorder and early age at diagnosis. Further, understanding of how genetic variants influence disease activity may lead to personalized, targeted treatments for patients with SLE. This study aimed to examine clinically important marginal effects of a GRS composed of 41 established genetic risk loci on disease activity over a period of 9 years, accounting for time-dependent confounding of missing visits, in a multiethnic cohort of SLE patients. We were also interested in conducting individual single-nucleotide polymorphism (SNP) analyses in order to determine whether specific genetic variants were associated with disease activity.

**RESULTS**

Demographic and disease characteristics are shown in Table 1. Over 90% of study participants were female, which is consistent with the striking female predominance of the disease. The mean age of participants at enrollment was 56.2 (±13.6) years, with the mean disease duration of 9.1 (±8.3) years and average age at diagnosis of 33.5 (±13.0). The mean follow-up time for participants was 6.7 (±2.7) years. The average Systemic Lupus Activity Questionnaire (SLAQ) score at baseline was 12.5 (±7.9). Treatment of patients included plaquenil (56%), oral prednisone or other glucocorticoid (46%), other disease-modifying therapies (31%), cyclophosphamide or chlorambucil (3%), or biologics (1%).

1Division of Epidemiology, University of California, Berkeley, CA, USA; 2Division of Biostatistics, University of California, Berkeley, CA, USA; 3Department of Medicine, Rosalind Russell/Ephraim P. Engleman Rheumatology Research Center, University of California, San Francisco, CA, USA and 4Department of Biochemistry and Molecular Medicine, University of California, Davis, CA, USA. Correspondence: Dr LF Barcellos, Division of Epidemiology, University of California, 324 Stanley Hall, M C#3220, Berkeley, CA 94720-3220, USA.

E-mail: lbarcellos@genepi.berkeley.edu

2These authors contributed equally to this work.

Received 25 May 2016; accepted 9 June 2016; published online 28 July 2016
Table 1. Characteristics of SLE participants at baseline (n = 942)

| Characteristic                  | Mean ± s.d. or N (%) |
|---------------------------------|----------------------|
| **Age, years**                  | 56.2 ± 13.6          |
| **Sex**                         |                      |
| Females                         | 866 (91.9)           |
| Males                           | 76 (8.1)             |
| **Race/ethnicity**              |                      |
| African American                | 104 (11.0)           |
| Asian                           | 80 (8.5)             |
| Caucasian                       | 617 (65.5)           |
| Hispanic                        | 82 (8.7)             |
| Mixed ethnicity                 | 59 (6.3)             |
| **Disease duration, years**     | 9.1 ± 8.3            |
| **Age at diagnosis, years**     | 33.5 ± 13.0          |
| **Genetic risk score (binary, n = 903)** |                      |
| High (≥33)                      | 498 (55.0)           |
| Low (< 33)                      | 405 (45.0)           |
| **Genetic risk score (extremes, n = 493)** |                  |
| Upper 25% (≥36)                 | 257 (52.0)           |
| Lower 25% (< 30)                | 236 (48.0)           |
| **Systemic Lupus Activity Questionnaire Score** (range 0–47) | 12.5 ± 7.9          |
| **Treatment**                   |                      |
| Oral prednisone or other glucocorticoid | 429 (45.5)         |
| Plaquenil                       | 525 (55.7)           |
| Biologics                       | 8 (0.9)              |
| Cyclophosphamide or chlorambucil| 27 (2.9)             |
| Other disease-modifying therapies| 288 (30.6)          |
| Depression (range 0–59)         | 15.3 ± 12.7          |
| **Smoker**                      |                      |
| Ever                            | 519 (57.2)           |
| Never                           | 388 (42.8)           |
| **Renal transplant status**     |                      |
| Yes                             | 6 (0.6)              |
| No                              | 936 (99.4)           |
| **Dialysis**                    |                      |
| Yes                             | 39 (4.1)             |
| No                              | 903 (95.9)           |
| **Education**                   |                      |
| Less than high school graduate  | 25 (2.7)             |
| High school graduate            | 128 (13.6)           |
| Some college                    | 247 (26.2)           |
| Associate/trade degree          | 176 (18.7)           |
| Bachelor’s degree               | 215 (22.8)           |
| Master’s, Doctoral, Post-Baccalaureate Degree | 151 (16.0)  |

Abbreviations: N, number of observations; SLE, systemic lupus erythematos.

Table 2. Estimates of the marginal association of GRS and disease activity (SLAQ score) among individuals over study perioda

| L-TMLE | Expected difference (high versus low GRS; 95% CI) | P-value |
|--------|--------------------------------------------------|---------|
| T1     | −1.7 (−2.7, −0.7)                                | 0.001   |
| T2     | −1.5 (−2.7, −0.4)                                | 0.01    |
| T3     | −1.4 (−2.5, −0.3)                                | 0.01    |
| T4     | −1.2 (−2.3, −0.1)                                | 0.03    |
| T5     | −2.1 (−3.4, −0.8)                                | 0.001   |
| T6     | −0.5 (−2.2, 1.3)                                 | 0.61    |
| T7     | −1.0 (−2.4, 0.4)                                 | 0.17    |
| T8     | −1.3 (−4.3, 1.7)                                 | 0.39    |
| T9     | −2.6 (−5.4, 0.1)                                 | 0.06    |

Abbreviations: CI, confidence interval; GRS, genetic risk score; L-TMLE, longitudinal-targeted maximum likelihood estimation; N, number of observations without missing outcome at time t; SLAQ, Systemic Lupus Activity Questionnaire. aAdjusted for sex, ancestry, disease duration, renal status, renal involvement, dialysis, treatment, depression, smoking and education.

Figure 1. Genetic risk score (GRS) and longitudinal clinically significant disease activity.

first and fourth quartiles) indicated no clinically important difference of SLAQ score between groups throughout the follow-up period after controlling for covariates (Supplementary Table 1). Analyses examining the effect of GRS extremes without the HLA-DRB1*03:01 on SLAQ score were similar (data not shown).

Individual SNP analyses demonstrated that clinically meaningful SLAQ score differences were observed for two SNPs at time point 8 and ten SNPs at time point 9 after adjusting for covariates and correcting for multiple testing (Table 3). Eight of the twelve risk SNP associations were associated with an increased SLAQ score, indicating higher disease activity. We found no significant association between DRB1*03:01 and SLAQ score at any of the time points.

**DISCUSSION**

Our study describes the first longitudinal study examining how genetic factors influence disease activity in a large multiethnic cohort of individuals with SLE. Using a robust method of statistical analysis, our findings do not support a strong causal relationship between an overall GRS comprising established SLE SNPs and disease activity as measured by the validated self-reported SLAQ. Whereas results from analyses of the GRS and SLAQ score provided some evidence for an inverse relationship at years 1–5, the magnitude of these SLAQ score differences is not within the range established as clinically important.9 Results from individual SNP analyses provide important insight to the overall GRS
findings; specifically, evidence for significant associations between certain SNPs and SLAQ score at two time points during the longitudinal study was demonstrated. Some SLE risk alleles were associated with an increased SLAQ score, whereas others were associated with a decreased SLAQ score. Of the 41 SNPs tested across various time points, 12 were associated with clinically significant SLAQ score differences. Score differences varied, depending on risk SNP tested, and help to explain the overall null findings that were observed when all risk SNPs were considered as a single combined GRS. More work is needed to further clarify the relationship between established SLE risk variants and patterns of disease activity for clinical outcomes that the SLAQ may or may not capture.

Previous studies have found significant associations between specific gene variants and SLE disease characteristics, which capture severity, not specifically activity. Allele 2 of the 1LRN polymorphism was associated with SLE disease characteristics as defined by discoid rash and photosensitivity. In a Danish study, the MBL2 gene was associated with an increased risk of sustained disease activity and a tendency to acquire infection, but not identified as a susceptibility locus. Taylor et al. also concluded that a common polymorphism within STAT4 was associated with more severe SLE characteristics, including production of antibodies against ds-DNA, nephritis and age at diagnosis. Status for ds-DNA, which can fluctuate over time with disease activity or severity, was previously associated with the HLA-DRB1*0401 allele, as well as ITGAM, UBE2L3 and HLA-DRB1*1501. Although we did not find a significant association between HLA alleles or risk variants in ITGAM and SLAQ score across time points, we did observe a clinically important difference in disease activity for those carrying one to two copies of a risk allele in UBE2L3 (rs754217) at time point 9. An earlier investigation reported significant associations between a weighted GRS comprising 22 SLE risk loci and several SLE-related clinical manifestations including ds-DNA, immunologic disorder, renal disorder, hematologic disorder and early age at diagnosis in individuals of European ancestry. Whereas previous studies support the influence of genetics on certain SLE outcomes, our study is the first to examine the relationship between a 41-SNP GRS and SLAQ score as a measure of disease activity.

Some differences in findings between previous studies of disease severity measured by ds-DNA status and disease activity measured by SLAQ score in the current study may be explained by the self-reported nature of the SLAQ score, which could be influenced by a number of factors that were not measured or appropriately controlled for in our analysis. We found SLAQ score within individuals to be highly variable over time, indicating that it may be driven by characteristics not closely linked to genetics, such as psychosocial and other biological factors. Indeed, previous studies have demonstrated that specific factors such as age, sex, income, education, as well as presence of renal and lung diseases are associated with SLAQ scores. Our results indicate primarily a nonlinear and clinically nonsignificant relationship between the GRS and SLAQ after controlling for these other factors, suggesting that established genetic risk factors for developing SLE, when considered together, do not have a major role in disease activity, at least as measured with the SLAQ score. However, we identified specific genetic variants through individual SNP analysis causally associated with disease activity, even after correction for multiple testing.

Strengths of this study include the application of causal methods to an important research hypothesis focused on a comprehensive, validated clinical outcome instrument for SLE. Traditional longitudinal methods, such as simple random effects models, are likely to be subject to bias from time-dependent confounding: time-varying covariates that could influence variation in disease activity and the measurement process, and that are also affected by the baseline exposure. However, using our approach, we were able to much better control for these covariates using longitudinal-targeted maximum likelihood estimation (L-TMLE). Additional studies examining the extent to which the GRS is associated with other measures of disease activity using methods that account for time-varying covariates are warranted; our current study is a model of how future investigations can be approached. Our study also focused on a large number of established genetic variants for SLE risk with large effect sizes.

In addition, the current study included a large, multiethnic group of individuals, which is important for external validation of our findings. Although power was limited for analyzing each race/ethnicity group separately, overall results were consistent when the study sample was restricted to Caucasian participants only (controlling for intra-European principal components), the largest group in our patient cohort. Further, we do not present analyses stratified by ancestry because of the fact that the GRS was predominately derived from risk alleles associated with SLE in European populations and may under-represent relative genetic risk of non-Europeans. Future studies examining other populations using population-specific SLE risk alleles are needed.

Study limitations include examination of the GRS as a binary variable in the statistical model, which is unlikely to fully capture the relationship between the GRS values and shifts in disease activity. However, extreme quartiles were utilized for analyses and comparable finding were demonstrated. Further, use of a binary variable reduces the increased likelihood of positivity violations and extrapolation that would occur when using a continuous variable. Additional risk SNPs will no doubt be identified through larger genome-wide association studies in SLE, and pathway analysis may soon inform subscore development for future longitudinal studies of a GRS and disease activity in SLE. Additional limitations include reliance on prevalent cases to estimate effects over time, self-report of symptoms in the SLAQ score and imputation of missing covariates using the last value forward method. We also accounted for informative measurement (that is, selection bias) as well as possible given data; however, there may be additional unmeasured time-dependent covariates that influence whether or not a participant was interviewed at time t. In other words, the sequential randomization assumption may not hold for the measurement process. There was some attrition over the 9-year follow-up period. Whereas we controlled for the potentially informative measurement process with L-TMLE, the wider confidence intervals at later time points reflect, in part, the fewer number of individuals interviewed in

| SNP          | Risk allele | Time point 8 | Time point 9 |
|--------------|-------------|--------------|--------------|
| rs3184504    | A           | 4.82 (3.10, 6.55) | —            |
| rs17696736   | G           | 4.24 (2.53, 5.95) | —            |
| rs2618473    | A           | 4.30 (1.95, 6.65) | —            |
| rs11574914   | A           | 4.28 (1.71, 6.86) | —            |
| rs7197475    | T           | 5.72 (3.56, 7.89) | —            |
| rs3135388    | A           | 4.39 (2.38, 6.39) | —            |
| rs3024505    | T           | 7.50 (5.27, 9.73) | —            |
| rs5754217    | T           | 5.63 (3.69, 7.57) | —            |

Abbreviations: CI, confidence interval; SLAQ, Systemic Lupus Activity Questionnaire; SNP, single-nucleotide polymorphism. Estimates reflect differences in the expected SLAQ score for participants with 1–2 versus 0 risk alleles (95% CI). Adjusted for sex, ancestry, disease duration, renal status, renal involvement, dialysis, treatment, depression, smoking and education. All P-values are Bonferroni-corrected and ≤0.001.
the later years of the study, which may also contribute to a higher potential for positivity violations.

In summary, SLE is a multifactorial disease with both genetic and environmental risk factors. In line with previously demonstrated associations of genetic factors with disease severity, for example, ds-DNA status as examined in other studies,\(^5\)\(^-\)\(^7\)\(^-\)\(^9\)\(^-\) we observed evidence of clinically important differences in SLAQ scores for individual SNPs. We did not observe a clinically important association of the overall GRS with disease activity as measured by the SLAQ score, which highlights the strong potential for heterogeneous genetic effects influencing disease activity over time. Genetic susceptibility appears to have a role in the development of SLE, as well as some measures of disease activity and severity. However, other biological and unknown environmental factors may also contribute significantly to disease activity. Future work to identify specific patient subgroups based on genetic and environmental factors and the development of corresponding targeted treatment interventions is needed, and has the potential to improve SLE symptoms.

**MATERIALS AND METHODS**

**Participants**

Data were collected from participants of the Lupus Outcome Study at the University of California, San Francisco. This longitudinal cohort study was designed to prospectively investigate health and quality of life outcomes in a set of SLE patients who were also participants in a larger study of SLE genetic risk factors and outcomes.\(^6\)\(^-\)\(^7\) Participants were followed from January 2004 to December 2012. Study methodology and details have been reported previously.\(^6\)\(^-\)\(^7\) Participants had a diagnosis of SLE confirmed by medical record review by a rheumatologist or a nurse working under a rheumatologist's supervision. The cohort was composed of 942 individuals self-identified as Caucasian, African American, Hispanic, Asian or mixed race/ethnicity. The Committee on Human Research at the University of California, San Francisco approved this study, and all participants provided informed consent before enrollment.

**Data collection**

Participants completed up to nine annual structured telephone interviews comprising questions related to various events and exposures. Disease activity was measured by a validated comprehensive self-reported measure, the SLAQ. The SLAQ identifies clinical outcomes such as disease symptoms including weight loss, fatigue, fever, skin rash, vasculitis, arthralgia, lymphadenopathy, shortness of breath, chest pain with deep breath, Raynaud’s, abdominal pain, stroke syndrome, seizures, forgetfulness, depression, headache, muscle pain, muscle weakness, joint pain or stiffness and joint swelling.\(^8\)\(^-\)\(^9\)\(^-\)\(^10\) Items were weighted and aggregated into a scoring system, resulting in scores ranging from 0 to 47, with higher scores indicating greater disease activity over the prior 3 months. A conservative estimate of the minimal clinically important difference for SLAQ score in the current study was considered to be 4.0 points, as previously indicated.\(^9\)

Participants were genotyped using the Illumina Immunochip platform.\(^11\)\(^-\)\(^12\) Quality-control procedures included removal of subjects for low genotyping (< 5%), sex mismatch and relatedness (first degree as determined by identity-by-descent analysis). A GRS was calculated for each individual from 41 established SLE susceptibility loci (Supplementary Table 2).\(^13\)\(^-\)\(^14\)\(^-\)\(^15\) Proxies if available for SNPs not genotyped on the Immunochip. The GRS was calculated by summing the number of independent risk alleles for each locus across the 41 loci; missing genotypes (< 5%) were assumed to be the most frequent genotype. We also calculated a GRS without HLA-DRB1 (DRB1*03:01 allele tag SNP, rs21876668),\(^13\) the strongest genetic predictor of SLE, to examine whether this factor had a major influence on the GRS.

Study data included sex and disease duration (calculated as age at time of interview minus age at diagnosis). Renal transplant and dialysis status at time of interview were self-reported (yes/no), and depression was measured using the Center for Epidemiologic Studies Depression Scale.\(^16\) Treatment was categorized as type of medication at time of interview: oral prednisone or other glucocorticoid; plaquenil; or other (including biologics (etanercept, rituxan, belimumab, anakinra, infliximab, adalimumab or abatacept); cyclophosphamide or chlorambucil; or other disease-modifying therapies (methotrexate, cyclosporine, leflunomide, mycophenolate mofetil, tacrolimus or sulfasalazine)). Genetic ancestry was represented by the top four principal components determined with EIGENSTRAT using 878 independent (r² < 0.2) Immunochip SNPs.

### Statistical analyses

L-TMLE was used to estimate the marginal association of the GRS on SLE disease activity during the 9-year observation period.\(^13\) L-TMLE allows for estimation of the cumulative genetic impact of the baseline exposure (GRS) while accounting for time-varying covariates and non-monotone observation process.\(^14\) Despite the study’s best efforts, cohort members occasionally missed their yearly interviews; however, individuals did not generally miss more than 2 years in a row. Instead of censoring at the time of last missed interview, which would substantially reduce the data set, L-TMLE was used to control for the possibly informative measurement process, that is, to account for missing time points (interviews) that have been influenced by time-varying covariates.

Our primary research question aimed to examine the causal effect of the GRS on clinically important disease activity as measured by SLAQ score at each time point (year) \(t\) while accounting for the observation process. Let \(A\) be an indicator that patient had a GRS above the median in the data set and \(Δ\) be an indicator that the patient was interviewed at time \(t\). Our causal parameter was the expected difference in the counterfactual SLAQ score at time \(t\) if all patients had high versus low GRS and were interviewed at time \(t\): \(E[Y(\Delta = 1) − Y(\Delta = 0)]\), where \(Y(\Delta = a, \Delta = 1)\) is the counterfactual outcome at time \(t\) if possibly contrary-to-fact the patient’s GRS was \(A = a\) and he/she was interviewed at time \(t\). The time-dependent covariates (disease duration, depression, treatment, smoking status, dialysis, renal transplant status and education) may affect both the outcome \(Y\), as well as observation process \(Δ\). These covariates are also influenced by the exposure GRS \(A\). Therefore, time-dependent confounding is present and standard regression methods are likely to yield biased estimates, even though exposure GRS is randomized within sex-ancestry strata.\(^14\)\(^-\)\(^15\)\(^-\)\(^16\) However, more sophisticated methods allow us to estimate the statistical parameter, best approximating our causal parameter of interest.

Our statistical estimand was the longitudinal G-Computation formula,\(^16\) which will equal our causal parameter if the needed assumptions are met. Specifically, we need the sequential randomization and positivity assumptions to hold: that is, at each time \(t\), the counterfactual outcome \(Y(\Delta = a, \Delta = 1)\) is independent of the observation process at \(t\), given the measured past, and there is a positive probability of being observed (\(Δ = 1\)) within all covariate–measurement histories. Furthermore, by incorporating data-adaptive methods, L-TMLE is able to minimize the potential bias because of regression model misspecification and maximize precision of the estimates. In this analysis, Super Learner was used to build the best-weighted combination of candidate algorithms as measured by the cross-validated mean squared error.\(^17\) The library of candidate prediction algorithms in Super Learner included linear regression; linear regression with interactions; the mean; stepwise regression; and stepwise regression with interactions. For comparison, we also examined the association of being in the most extreme quartiles of GRS on average SLAQ score. Specifically, we defined the patient as exposed (\(Δ = 1\)) if he/she had a GRS in the upper 25 percentile and as unexposed (\(Δ = 0\)) if he/she had a GRS in the lower 25 percentile. Again, we adjusted for potentially informative measurement, baseline confounders (sex and ancestry), and accounted for time-dependent confounding.

We tested the association between the GRS (above or below the median and in the extreme quartiles) and disease activity as measured by SLAQ across nine time points: baseline (T1) and eight additional time points (T2–T9). We were specifically interested in SLAQ differences between the two groups that were clinically meaningful (greater than 4.0 points).\(^9\) We examined each time point separately to better control for informative missingness by using time-updated covariates to fit the observation status at time \(t\), and because disease activity in SLE patients varies both between and within individuals over time.\(^14\) We restricted the sample to the subset of participants who had no complete missingness (that is, missing values at all time points) for the following variables: sex, ancestry as measured by EIGENSTRAT principal components, disease duration, depression, treatment, smoking status, GRS, dialysis, renal transplant status and education (\(n = 903\)). An assessment of multicollinearity between these variables was conducted by examining variance inflation factors; all variables had a variance inflation factor (VIF) < 1.2. For missing values of the data that remained in the subset of 903 individuals (< 3% for each variable), information was imputed using last value carried forward. As described above, individuals with missing outcomes (SLAQ score) were accounted for by conditioning on the observation process (\(Δ\)).
Owing to the limitation of a binary GRS and inability to determine whether specific genetic variants are associated with disease activity, we also conducted individual SNP analyses. We defined the patient as exposed (A=1) if he/she carried 1-2 risk alleles and as unexposed (A=0) if he/she carried 0 risk alleles for each SNP, and adjusted analyses for potentially informative measurement, baseline confounders and accounted for time-dependent confounding. Individual SNP analyses were corrected for multiple testing using Bonferroni adjustment. All analyses were conducted with the 'ltmle' package in R v3.2.0 (www.r-project.org). We created 95% confidence intervals and conducted two-sided hypothesis tests controlling the type I error rate at 5% (α = 0.05).

CONFLICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGEMENTS
We thank Sam Lendle and Josh Schwab for their assistance with the L-TMLE software, as well as Maya Petersen, Alex Luedtke and Alice Baker for their thoughtful feedback. We also thank Paula Bronson for her assistance in developing the genetic risk score, and Lily Hoang for her help with this project. This work was supported by the Lupus Foundation of America (to MAG); Rheumatology Research Foundation (to MAG); National Institutes of Health (Grants P60 AR053308 and K24-AR-02175 to LAC); University of California, San Francisco General Clinical Research Center (Grant RO1-AR-44804 to LAC); National Center for Advancing Translational Sciences (Grant UL1-TR-000004 to LAC); and Alliance for Lupus Research (to LAC).

REFERENCES
1 Deng Y, Tsao BP. Genetic susceptibility to systemic lupus erythematosus in the genomic era. Nat Rev Rheumatol 2010; 6: 683–692.
2 Morris DL, Taylor KE, Fernando MM, Nittatham J, Alarcon-Riquelme ME, Barcellos LF et al. Unraveling multiple mhc gene associations with systemic lupus erythematosus: model choice indicates a role for hla alleles and non-hla genes in europeans. Am J Hum Genet 2012; 91: 776–793.
3 Barcellos LF, May SL, Ramsay PP, Quach HL, Lane JA, Nittatham J et al. High-density SNP screening of the major histocompatibility complex in systemic lupus erythematosus demonstrates strong evidence for independent susceptibility regions. PLoS Genet 2009; 5: e1000696.
4 Bentham J, Morris DL, Cunningham Graham DS, Pinder CL, TOMBlesson P, Behrens Owing to the limitation of a binary GRS and inability to determine whether specific genetic variants are associated with disease activity, we also conducted individual SNP analyses. We defined the patient as exposed (A=1) if he/she carried 1-2 risk alleles and as unexposed (A=0) if he/she carried 0 risk alleles for each SNP, and adjusted analyses for potentially informative measurement, baseline confounders and accounted for time-dependent confounding. Individual SNP analyses were corrected for multiple testing using Bonferroni adjustment. All analyses were conducted with the 'ltmle' package in R v3.2.0 (www.r-project.org). We created 95% confidence intervals and conducted two-sided hypothesis tests controlling the type I error rate at 5% (α = 0.05).

CONFLICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGEMENTS
We thank Sam Lendle and Josh Schwab for their assistance with the L-TMLE software, as well as Maya Petersen, Alex Luedtke and Alice Baker for their thoughtful feedback. We also thank Paula Bronson for her assistance in developing the genetic risk score, and Lily Hoang for her help with this project. This work was supported by the Lupus Foundation of America (to MAG); Rheumatology Research Foundation (to MAG); National Institutes of Health (Grants P60 AR053308 and K24-AR-02175 to LAC); University of California, San Francisco General Clinical Research Center (Grant RO1-AR-44804 to LAC); National Center for Advancing Translational Sciences (Grant UL1-TR-000004 to LAC); and Alliance for Lupus Research (to LAC).

REFERENCES
1 Deng Y, Tsao BP. Genetic susceptibility to systemic lupus erythematosus in the genomic era. Nat Rev Rheumatol 2010; 6: 683–692.
2 Morris DL, Taylor KE, Fernando MM, Nittatham J, Alarcon-Riquelme ME, Barcellos LF et al. Unraveling multiple mhc gene associations with systemic lupus erythematosus: model choice indicates a role for hla alleles and non-hla genes in europeans. Am J Hum Genet 2012; 91: 776–793.
3 Barcellos LF, May SL, Ramsay PP, Quach HL, Lane JA, Nittatham J et al. High-density SNP screening of the major histocompatibility complex in systemic lupus erythematosus demonstrates strong evidence for independent susceptibility regions. PLoS Genet 2009; 5: e1000696.
4 Bentham J, Morris DL, Cunningham Graham DS, Pinder CL, TOMBlesson P, Behrens et al. Unraveling multiple mhc gene associations with systemic lupus erythematosus: model choice indicates a role for hla alleles and non-hla genes in europeans. Am J Hum Genet 2012; 91: 776–793.
3 Barcellos LF, May SL, Ramsay PP, Quach HL, Lane JA, Nittatham J et al. High-density SNP screening of the major histocompatibility complex in systemic lupus erythematosus demonstrates strong evidence for independent susceptibility regions. PLoS Genet 2009; 5: e1000696.
4 Bentham J, Morris DL, Cunningham Graham DS, Pinder CL, TOMBlesson P, Behrens et al. Unraveling multiple mhc gene associations with systemic lupus erythematosus: model choice indicates a role for hla alleles and non-hla genes in europeans. Am J Hum Genet 2012; 91: 776–793.
3 Barcellos LF, May SL, Ramsay PP, Quach HL, Lane JA, Nittatham J et al. High-density SNP screening of the major histocompatibility complex in systemic lupus erythematosus demonstrates strong evidence for independent susceptibility regions. PLoS Genet 2009; 5: e1000696.
4 Bentham J, Morris DL, Cunningham Graham DS, Pinder CL, TOMBlesson P, Behrens et al. Unraveling multiple mhc gene associations with systemic lupus erythematosus: model choice indicates a role for hla alleles and non-hla genes in europeans. Am J Hum Genet 2012; 91: 776–793.