Genetic Risk and Longitudinal Disease Activity in Systemic Lupus Erythematosus Using Targeted Maximum Likelihood Estimation

Milena A. Gianfrancesco,
Division of Epidemiology, University of California, Berkeley, CA 94110-7358

Laura Balzer,
Division of Biostatistics, University of California, Berkeley, CA 94110-7358

Kimberly E. Taylor,
Rosalind Russell/Ephraim P. Engleman Rheumatology Research Center, Department of Medicine, University of California, San Francisco, CA, 94143-0500

Laura Trupin,
Rosalind Russell/Ephraim P. Engleman Rheumatology Research Center, Department of Medicine, University of California, San Francisco, CA, 94143-0500

Joanne Nititham,
Rosalind Russell/Ephraim P. Engleman Rheumatology Research Center, Department of Medicine, University of California, San Francisco, CA, 94143-0500

Michael F. Seldin,
Department of Biochemistry and Molecular Medicine, University of California, Davis, CA 95616

Amanda Wheeler Singer,
Division of Epidemiology, University of California, Berkeley, CA 94110-7358

Lindsey A. Criswell[1], and
Rosalind Russell/Ephraim P. Engleman Rheumatology Research Center, Department of Medicine, University of California, San Francisco, CA, 94143-0500

Lisa F. Barcellos[1]
Division of Epidemiology, University of California, Berkeley, CA 94110-7358

Abstract

Systemic lupus erythematous (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

Corresponding Author: Dr. Lisa F. Barcellos, Division of Epidemiology, 324 Stanley Hall, University of California Berkeley, CA 94720-3220, lbarcellos@genepi.berkeley.edu; Phone: 510-642-7814.
[1]Authors contributed equally

CONFLICT OF INTEREST
The authors report no conflicts of interest.

estimate the causal association between a genetic risk score (GRS) comprised of 41 established SLE variants and clinically important disease activity as measured by the validated systemic lupus activity questionnaire (SLAQ) in a multi-ethnic 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 to those with a low GRS across nine timepoints 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 SNP analyses revealed that 12 of the 41 variants were significantly associated with clinically relevant changes in SLAQ scores across timepoints 8 and 9 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.

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 (i.e. greater likelihood of disease given that one's sibling is affected, compared with disease prevalence in the general population).[1] 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.[2,3] Recent genome-wide association studies (GWAS) have also identified more than 40 independent loci related to SLE onset.[4] A weighted genetic risk score (GRS) comprised of these variants was significantly higher on average in SLE cases compared to healthy controls.[5] However, measures of association between genetic variants and SLE are generally modest (with odds ratios [ORs] ranging from 1.1–2.3), gene-environment interactions are poorly understood, and there is little awareness of how genetic profiles impact disease activity.[1, 6]

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.[7] 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. While 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.[8] One study found that a weighted GRS comprised of 22 SLE risk loci demonstrated significant associations with ds-DNA, immunologic disorder, renal disorder, hematologic disorder, and early age at diagnosis.[5] Further understanding of how genetic variants influence disease activity may lead to personalized, targeted treatments for patients with SLE.[1] 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 nine years, accounting for time-dependent confounding of missing visits, in a multi-ethnic cohort of SLE patients. We were also interested in conducting individual 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 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 therapy (31%), cyclophosphamide or chlorambucil (3%), or biologics (1%).

Table 2 shows estimated marginal difference in the expected self-reported SLAQ score for participants with a high GRS (33.0, the median GRS) vs. a low GRS (< 33.0) after controlling for sex, ancestry, disease duration, renal transplant status, dialysis, treatment, depression, smoking, and education. No consistent clinically important difference of SLAQ score between groups was demonstrated throughout the nine years of follow-up period (Figure 1). Additional analyses examining the association of GRS without the *HLA-DRB1*03:01* allele tag SNP on expected SLAQ score showed similar results (data not shown). Analyses excluding renal transplant status or dialysis to further assess any evidence of collinearity of these variables also reflected similar results (data not shown). Analyses of the GRS based on extremes (i.e. the 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 12 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 multi-ethnic 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 comprised of established SLE SNPs and disease activity as measured by the validated self-reported SLAQ. While 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, while 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 *IL1RN* polymorphism was associated with SLE disease characteristics as defined by discoid rash and photosensitivity.[10] 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.[11] Taylor et al. (2008) also concluded that a common polymorphism within *STAT4* was associated with more severe SLE characteristics, including production of antibodies against double-stranded DNA (ds-DNA), nephritis, and age at diagnosis.[8] Status for ds-DNA, which can fluctuate over time with disease activity or severity,[12–14] was previously associated with the *HLA-DRB1*15:01* allele,[12] as well as *ITGAM*, *UBE2L3*, and *HLA-DRB1*03:01*.[5] 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 1–2 copies of a risk allele in *UBE2L3* (rs5754217) at time point 9. An earlier investigation reported significant associations between a weighted GRS comprised of 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.[5] While 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 disease are associated with SLAQ scores.[9, 15] Our results indicate primarily a non-linear and clinically non-significant 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 play 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.[16] 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, multi-ethnic group of individuals, which is important for external validation of our findings. While 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 due to the fact that the GRS was predominately derived from risk alleles associated with SLE in European populations and may underrepresent 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, 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 GWAS in SLE, and pathway analysis may soon inform sub-score 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 (i.e. selection bias) as well as possible given the 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.[16] There was some attrition over the nine-year follow-up period. While 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 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, 12–14] 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 play 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 (LOS) at the University of California, San Francisco (UCSF). 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.[8, 17] Participants were followed from January 2004 – December 2012. Study methodology and details have been reported previously.[18, 19] 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 UCSF approved this study and all participants provided informed consent prior to enrollment.

Data Collection

Participants completed up to nine annual structured telephone interviews comprised of questions related to various events and exposures. Disease activity was measured by a validated comprehensive self-reported measure, the Systemic Lupus Activity Questionnaire (SLAQ). The SLAQ identifies clinical outcomes such as disease symptoms including weight loss, fatigue, fever, skin rash, vasculitis, alopecia, 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.[9, 20] 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 three 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.[21] Quality control procedures included removal of subjects for low genotyping (<5%), sex mismatch, and relatedness (1st degree as determined by identity-by-descent analysis). A GRS was calculated for each individual from 41 established SLE susceptibility loci (Supplementary Table 2)[5,14, 22–29] with 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, rs2187668),[3] the strongest genetic predictor of SLE, to examine if this factor had a major influence on the GRS.

Study data included sex and disease duration (as calculated by 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 (CESD) scale.[30] 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 therapy (methotrexate, cyclosporine, leflunomide, mycophenolate mofetil, tacrolimus, or sulfasalazine)]. Genetic ancestry was represented by the top four principal components determined with EIGENSTRAT using 878 independent ($r^2 < 0.2$) Immunochip SNPs.

Statistical Analyses

L-TMLE was used to estimate the marginal association of the GRS on SLE disease activity during the nine-year observation period.[31] 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.[32] Despite the study's best efforts, cohort members occasionally missed their yearly interviews; however, individuals did not generally miss more than two years in a row. Instead of censoring at the time of the first 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 may 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 dataset, and $_t$ 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 vs. low GRS and were interviewed at time t: $E[Y_t(A=1, _t=1) - Y_t(A=0, _t=1)]$, where $Y_t(A=a, _t=1)$ 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 $_t=1$.[33] The time-dependent covariates (disease duration, depression, treatment, smoking status, dialysis, renal transplant status and education) may affect both the outcome Y_t as well as observation process $_t$. These covariates are also impacted by the exposure GRS A. Therefore, time-dependent confounding is present and standard regression methods are likely to yield biased estimated, even though exposure GRS is randomized within sex-ancestry strata.[16, 34, 35] 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 $Y_t(A=a, _t=1)$ is independent of the observation process at t, given the measured

past, and there is a positive probability of being observed ($_t$=1) within all covariate-measurement histories. Furthermore, by incorporating data-adaptive methods, L-TMLE is able to minimize the potential bias due to 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.[36] 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 (A=1) if he/she had a GRS in the upper 25 percentile and as unexposed (A=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.[37] We restricted the sample to the subset of participants who had no complete missingness (i.e. missing values at all time points) for the following variables: sex, ancestry as measured by EIGENSTRAT principal components, disease duration, depression, treatment, smoking status, genetic risk score, dialysis, renal transplant status, and education (n=903). An assessment of multicollinearity between these variables was conducted by examining variance inflation factors (VIF); all variables were retained (VIFs 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 ($_t$).

Due 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. 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).[38] We created 95% confidence intervals (CIs) and conducted two-sided hypothesis tests controlling the type I error rate at 5% (α=0.05).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Deng Y, Tsao BP. Genetic susceptibility to systemic lupus erythematosus in the genomic era. Nat Rev Rheumatol. 2010; 6:683–692. [PubMed: 21060334]

2. Morris DL, Taylor KE, Fernando MM, Nititham 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; 91:778–793. [PubMed: 23084292]

3. Barcellos LF, May SL, Ramsay PP, Quach HL, Lane JA, Nititham 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. [PubMed: 19851445]

4. Bentham J, Morris DL, Cunningham Graham DS, Pinder CL, Tomblesson 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; 47:1457–1464. [PubMed: 26502338]

5. Taylor KE, Chung SA, Graham RR, Ortmann WA, Lee AT, Langefeld CD, et al. Risk alleles for systemic lupus erythematosus in a large case-control collection and associations with clinical subphenotypes. PLoS Genet. 2011; 7:e1001311. [PubMed: 21379322]

6. Gualtierotti R, Biggioggero M, Penatti AE, Meroni PL. Updating on the pathogenesis of systemic lupus erythematosus. Autoimmun Rev. 2010; 10:3–7. [PubMed: 20863908]

7. Luijten KM, Tekstra J, Bijlsma JW, Bijl M. The systemic lupus erythematosus responder index (sri); a new sle disease activity assessment. Autoimmun Rev. 2012; 11:326–329. [PubMed: 21958603]

8. Taylor KE, Remmers EF, Lee AT, Ortmann WA, Plenge RM, Tian C, et al. Specificity of the stat4 genetic association for severe disease manifestations of systemic lupus erythematosus. PLoS Genet. 2008; 4:e1000084. [PubMed: 18516230]

9. Yazdany J, Yelin EH, Panopalis P, Trupin L, Julian L, Katz PP. Validation of the systemic lupus erythematosus activity questionnaire in a large observational cohort. Arthritis Rheum. 2008; 59:136–143. [PubMed: 18163398]

10. Blakemore AI, Tarlow JK, Cork MJ, Gordon C, Emery P, Duff GW. Interleukin-1 receptor antagonist gene polymorphism as a disease severity factor in systemic lupus erythematosus. Arthritis Rheum. 1994; 37:1380–1385. [PubMed: 7945503]

11. Garred P, Voss A, Madsen HO, Junker P. Association of mannose-binding lectin gene variation with disease severity and infections in a population-based cohort of systemic lupus erythematosus patients. Genes Immun. 2001; 2:442–450. [PubMed: 11781711]

12. Podrebarac TA, Boisert DM, Goldstein R. Clinical correlates, serum autoantibodies and the role of the major histocompatibility complex in french canadian and non-french canadian caucasians with sle. Lupus. 1998; 7:183–191. [PubMed: 9607642]

13. Hanaoka H, Okazaki Y, Satoh T, Kaneko Y, Yasuoka H, Seta N, et al. Circulating anti-double-stranded DNA antibody-secreting cells in patients with systemic lupus erythematosus: A novel biomarker for disease activity. Lupus. 2012; 21:1284–1293. [PubMed: 22740429]

14. Chung SA, Taylor KE, Graham RR, Nititham J, Lee AT, Ortmann WA, et al. Differential genetic associations for systemic lupus erythematosus based on anti-dsdna autoantibody production. PLoS Genet. 2011; 7:e1001323. [PubMed: 21408207]

15. Trupin L, Tonner MC, Yazdany J, Julian LJ, Criswell LA, Katz PP, et al. The role of neighborhood and individual socioeconomic status in outcomes of systemic lupus erythematosus. J Rheumatol. 2008; 35:1782–1788. [PubMed: 18634153]

16. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. Mathematical Modeling. 1986; 7:1393–1512.

17. Thorburn CM, Prokunina-Olsson L, Sterba KA, Lum RF, Seldin MF, Alarcon-Riquelme ME, et al. Association of pdcd1 genetic variation with risk and clinical manifestations of systemic lupus erythematosus in a multiethnic cohort. Genes Immun. 2007; 8:279–287. [PubMed: 17344889]

18. Yelin E, Trupin L, Katz P, Criswell L, Yazdany J, Gillis J, et al. Work dynamics among persons with systemic lupus erythematosus. Arthritis Rheum. 2007; 57:56–63. [PubMed: 17266065]

19. Panopalis P, Julian L, Yazdany J, Gillis JZ, Trupin L, Hersh A, et al. Impact of memory impairment on employment status in persons with systemic lupus erythematosus. Arthritis Rheum. 2007; 57:1453–1460. [PubMed: 18050187]

20. Yazdany J, Trupin L, Gansky SA, Dall'era M, Yelin EH, Criswell LA, et al. Brief index of lupus damage: A patient-reported measure of damage in systemic lupus erythematosus. Arthritis Care Res (Hoboken). 2011; 63:1170–1177. [PubMed: 21584946]

21. Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nat Rev Genet. 2013; 14:661–673. [PubMed: 23917628]

22. 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; 358:900–909. [PubMed: 18204098]

23. 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. Nat Genet. 2008; 40:204–210. [PubMed: 18204446]

24. 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; 41:1234–1237. [PubMed: 19838193]

25. 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; 41:1228–1233. [PubMed: 19838195]

26. Yang J, Yang W, Hirankarn N, Ye DQ, Zhang Y, Pan HF, et al. Elf1 is associated with systemic lupus erythematosus in asian populations. Hum Mol Genet. 2011; 20:601–607. [PubMed: 21044949]

27. Orozco G, Eyre S, Hinks A, Bowes J, Morgan AW, Wilson AG, et al. Study of the common genetic background for rheumatoid arthritis and systemic lupus erythematosus. Ann Rheum Dis. 2011; 70:463–468. [PubMed: 21068098]

28. 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. Nat Genet. 2008; 40:211–216. [PubMed: 18204447]

29. 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. Nat Genet. 2008; 40:1059–1061. [PubMed: 19165918]

30. Eaton, WW.; Muntaner, C.; Smith, C.; Tien, A.; Ybarra, M. Center for Epidemiologic Studies Depression Scale: Review and revision (CESD and CESD-R). In: Maruish, ME., editor. The Use of Psychological Testing for Treatment Planning and Outcomes Assessment. 3rd. Lawrence Erlbaum; Mahwah: 2004. p. 363-377.

31. Petersen M, Schwab J, Gruber S, Blaser N, Schomaker M, van der Laan M. Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. J Causal Inference. 2014; 2:147–185. [PubMed: 25909047]

32. van der Laan MJ, Gruber S. Targeted minimum loss based estimation of causal effects of multiple time point interventions. Int J Biostat. 2012; 8

33. Pearl, J. Causality: Models, Reasoning and Inference. 2nd. Cambridge University Press; New York: 2009.

34. Taubman SL, Robins JM, Mittleman MA, Hernan MA. Intervening on risk factors for coronary heart disease: An application of the parametric g-formula. Int J Epidemiol. 2009; 38:1599–1611. [PubMed: 19389875]

35. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000; 11:550–560. [PubMed: 10955408]

36. van der Laan M, Polley E, Hubbard A. Super Learner. Statistical Applications in Genetics and Molecular Biology. 2007; 6:Article 25.

37. Mikdashi J, Nived O. Measuring disease activity in adults with systemic lupus erythematosus: The challenges of administrative burden and responsiveness to patient concerns in clinical research. Arthritis Res Ther. 2015; 17:183. [PubMed: 26189728]

38. Schwab, J.; Lendle, S.; Petersen, M.; van der Laan, M., editors. ltmle: Longitudinal targeted maximum likelihood estimation. R package version 0.9.3-1. http://CRAN.R-project.org/package=ltmle. Published April 22, 2014
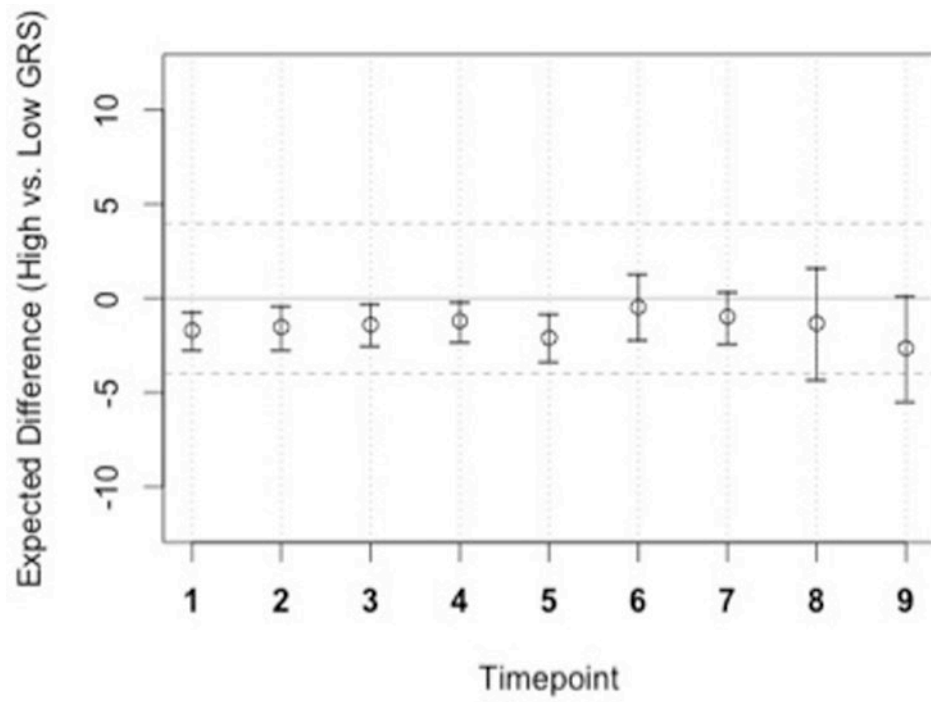
Figure 1. Genetic risk score (GRS) and longitudinal clinically significant disease activity
#Dashed line indicates clinically important difference

Table 1

Characteristics of SLE Participants at Baseline (n= 942)

	Mean ± SD 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 Cholorambucil	27 (2.9)
Other Disease Modifying Therapy	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)

	Mean ± SD or N(%)
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)

n= Number of observations

Table 2

Estimates of the Marginal Association of Genetic Risk Score (GRS) and Disease Activity (SLAQ score) Amongst Individuals Over Study Period[#]

	Expected Difference (High vs. Low GRS) (95% CI)	p-value
L-TMLE		
T1 (n=872)	−1.7 (−2.7, −0.7)	0.001
T2 (n=791)	−1.5 (−2.7, −0.4)	0.01
T3 (n=780)	−1.4 (−2.5, −0.3)	0.01
T4 (n=726)	−1.2 (−2.3, −0.1)	0.03
T5 (n=691)	−2.1 (−3.4, −0.8)	0.001
T6 (n=618)	−0.5 (−2.2, 1.3)	0.61
T7 (n=534)	−1.0 (−2.4, 0.4)	0.17
T8 (n=485)	−1.3 (−4.3, 1.7)	0.39
T9 (n=448)	−2.6 (−5.4, 0.1)	0.06

n= Number of observations without missing outcome at time *t*.

[#]Adjusted for sex, ancestry, disease duration, renal status, renal involvement, dialysis, treatment, depression, smoking, and education.

Table 3

Estimates of the Clinically Important Marginal Associations of Individual SNPs and Disease Activity (SLAQ score) Amongst Individuals Over Study Period[#*]

SNP	Risk allele	Time point 8	Time point 9
Risk allele associated with a clinically important increased SLAQ score			
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)
Risk allele associated with a clinically important decreased SLAQ score			
rs1167796	T	–	−5.90 (−7.65, −4.14)
rs9888739	T	–	−5.27 (−7.11, −3.43)
rs5029937	A	–	−5.58 (−7.61, −3.56)
rs725613	G	–	−5.59 (−7.69, −3.50)

Estimates reflect differences in the expected SLAQ score for participants with 1–2 vs. 0 risk alleles (95% CI).

[#]Adjusted for sex, ancestry, disease duration, renal status, renal involvement, dialysis, treatment, depression, smoking, and education.

[*]All p-values Bonferroni corrected and   0.001