Relationship Between Genetic Risk and Age of Diagnosis in Systemic Lupus Erythematosus

Daniela Dominguez1, Sylvia Kamphuis2, Joseph Beyene3, Joan Wither4, John B. Harley5, Irene Blanco6, Catarina Vila-Inda6, Hermine Brunner7, Marissa Klein-Gitleman8, Deborah MacCurdy9, Dawn M. Wahezi10, Thomas Lehman11, Marija Jelusic12, Christine A. Peschken13, Janet E. Pope14, Dafna D. Gladman15, John G. Hanly16, Ann E. Clarke17, Sasha Bernatsky18, Christian Pineau19, C. Douglas Smith20, Susan Barr21, Gilles Boire22, Eric Rich23, and Earl D. Silverman24, for the Canadian Network for Improved Outcomes in SLE (CaNiOS) and APPLE Investigators

ABSTRACT.

**Objective.** Specific risk alleles for childhood-onset systemic lupus erythematosus SLE (cSLE) vs adult-onset SLE (aSLE) patients have not been identified. The aims of this study were to determine if there is an association (1) between non-HLA–related genetic risk score (GRS) and age of SLE diagnosis, and (2) between HLA-related GRS and age of SLE diagnosis.

**Methods.** Genomic DNA was obtained from 2001 multiethnic patients and genotyped using the ImmunoChip. Following quality control, genetic risk counting (GRCS), weighted (GRWS), standardized counting (GRSCS), and standardized weighted (GRWS) scores were calculated based on independent single-nucleotide polymorphisms from validated SLE loci. Scores were analyzed in a regression model and adjusted by sex and ancestral population.

**Results.** The analyzed cohort consisted of 1540 patients: 1351 females and 189 males (675 cSLE and 865 aSLE). There were significant negative associations between all non-HLA GRS and age of SLE diagnosis: P = 0.011 and r^2 = 0.175 for GRWS; P = 0.008 and r^2 = 0.178 for GRSCS; P = 0.002 and r^2 = 0.176 for GRWS (higher GRS correlated with lower age of diagnosis.) All HLA GRS showed significant positive associations with age of diagnosis: P = 0.049 and r^2 = 0.176 for GRCS; P = 0.022 and r^2 = 0.176 for GRSCS; P = 0.011 and r^2 = 0.177 for GRWS (higher GRS correlated with higher age of diagnosis).

**Conclusion.** Our data suggest that there is a linear relationship between genetic risk and age of SLE diagnosis and that HLA and non-HLA GRS are associated with age of diagnosis in opposite directions.

Key Indexing Terms: age of onset, genetic risk score, SLE, SNP
Systemic lupus erythematosus (SLE) is a chronic multisystem autoimmune disease with a peak incidence in females during childbearing years. SLE tends to be more severe in men, patients with younger age of onset, and specific genetic ancestries. Multiple genes have been implicated in its pathogenesis, with > 90 genes/loci associated with predisposition to SLE. Although the majority of SLE susceptibility variants across ancestral populations are in the same gene, the associated single-nucleotide polymorphisms (SNPs) may differ or convey different risks for the development of SLE. SLE-associated genetic risk variants are located in both HLA and non-HLA regions, with alleles within the HLA region showing the strongest disease association.

For polygenic diseases such as SLE, it is accepted that a genetic risk score (GRS) provides better information on the genetic contribution to the development of autoimmune diseases than investigating a single SNP. There are 2 popular models for the calculation of polygenic risk scores: (1) a counting score that is a simple sum of risk and protective alleles present in an individual; and (2) a weighted score that takes into account the effect size of the SNP. Although there have been previous publications of polygenic risk scores in SLE, few examined if a risk score can be used as a predictor of age of disease onset and only 1 examined a non-HLA GRS over a large multiethnic population.

To our knowledge, this is the first study to examine the influence of HLA and non-HLA genes separately in a large combined pediatric and adult SLE cohort across a multiethnic population.

The aims of this study were to determine if there is an association between (1) GRS and age of SLE diagnosis for HLA genes, and (2) between GRS and age of SLE diagnosis for non-HLA genes.

METHODS

SLE cohort. This is a multicenter, multinational genetic study of patients with both childhood-onset SLE (cSLE) and adult-onset SLE (aSLE). SLE was defined as SLE diagnosed at age ≤ 18 years and aSLE as age of diagnosis ≥ 18 years. Genomic DNA was collected from 2001 patients from 17 centers within North America and 1 from Europe (Supplementary Table 1, available with the online version of this article) who fulfilled > 4 of 11 American College of Rheumatology classification criteria for SLE, with an age of disease diagnosis range of 1–82 years. The following clinical variables were available in 1979 of the 2001 patients: date of diagnosis, date of birth, age at diagnosis, and sex.

Determination of ancestral population. Determination of ancestral population.

There are no competing interests for the principal investigator or any coinvestigators. There was no financial or other commercial support for any of the work associated with this manuscript in any way.

Address correspondence to Dr. E.D. Silverman, Division of Rheumatology, Hospital for Sick Children, 555 University Ave., Toronto, ON M5G 1X8, Canada. Email: earl.silverman@sickkids.ca.

Full Release Article. For details see Reprints and Permissions at jrheum.org. Accepted for publication October 1, 2020.

Ethics. The study was approved by the coordinating center (Hospital for Sick Children Research Ethics Board [REB] number 100007761). In addition, ethics was obtained from the REBs at each center, as were data and material transfer agreements. All patients gave informed consent for genetic and clinical data used in this study.

Genotyping. All of the 1979 DNA samples with clinical information were sent for genotyping using the Illumina Infinium genotyping chip (Illumina, Inc.) at 1 of 2 centers: HudsonAlpha Genomics Services Laboratory and Cincinnati Children’s Hospital Medical Center-Harley Laboratory. Quality control (QC) of the genotype data was conducted using SNP & Variation Suite v8 software (Golden Helix Inc.) for each of the 4 different genotyping runs. Poor-quality samples with low call rates (< 95%), sample contamination, mixed samples or duplication, and close relatedness to another sample in the study were excluded. Relatedness between subjects was estimated by identity by descent analysis (IBDA). The total number of samples after QC was 1773/1979 samples genotyped.

SNP selection. By literature review of genome-wide association studies (GWAS), meta-GWAS, and candidate gene and replication studies, we found 106 SLE-associated SNPs, of which 93 SNPs were present on the Immunochip (Supplementary Table 2, available with the online version of this article). These 93 SNPs represented 39 different HLA-associated loci. All SNPs with a call rate < 0.99, minor allele frequency < 0.05, and Hardy-Weinberg Equilibrium P < 0.001 were excluded. After QC, there were 68 SNPs in 33 different loci: 58 in non-HLA regions and 10 in the HLA complex (Supplementary Table 3). Of the 13 SNPs that were not present on the Immunochip, we identified 4 alleles (in the HLA region) using SNP Annotation and Proxy Search (SNAP) online tool version 2.2 (Broad Institute, Harvard University), which resulted in a total of 72 SNPs (58 non-HLA and 14 HLA). SNPs that were located in the same gene were tested for pairwise high linkage disequilibrium (LD) using the SNAP online tool. In the case of pairs of SNPs that met the threshold criteria (r² > 0.8) in a specific gene for LD, the one with the lower OR reported in the literature was excluded. A further 14 SNPs were eliminated, leaving 58 independent SNPs (48 non-HLA and 10 HLA) from 33 SLE-susceptibility loci (Supplementary Table 3).

GRS calculation. ORs for all SNPs were classified as protective (OR < 1) or risk (OR > 1) by ancestral population (Supplementary Table 3, available with the online version of this article).

Four types of GRS were calculated for the statistical analysis: (1) genetic risk counting scores (GRCS), (2) genetic risk weighted scores (GRWS), (3) genetic risk standardized counting score (GRSCS), and (4) genetic risk standardized weighted score (GRSWS). All scores were calculated separately for HLA and non-HLA SNPs and analyzed in the total population.

The GRCS was an additive genetic model based on the presence of the risk or protective allele and was determined as the sum of those alleles present in each individual (Supplementary Table 3, available with the online version of this article); a simple count of the risk alleles minus the protective alleles.

The GRWS accounts for the relative effect of each risk/protective allele by using the OR. The GRWS was calculated as the sum of the natural logarithm of the OR of each risk/protective allele present. The OR of a risk allele was positive and OR of a protective allele was negative (Supplementary Table 3, available with the online version of this article).

Considering the large variation in the number of HLA and non-HLA SNPs identified in each individual ancestral population, we standardized the maximum GRCS and GRWS to 10 for each population to produce standardized GRCS (GRSCS) and GRWS (GRSWS) using only the SNPs available for the individual ancestral population. This allowed for comparisons across ancestral populations in order to weight each population equally (Figure 1).

Determination of ancestral population. Principal component analysis (PCA) was used to determine the ancestral identity of each patient. We first ran the analysis in the whole population, comparing the first 2 principal components.
components against reference samples of known ethnicities from the phase III International HapMap Project (HapMap3; www.sanger.ac.uk/resources/downloads/human/hapmap3.html). Samples that were outliers from the calculated clusters were dropped from the study. Multidimensional outlier detection (MOD) analysis was performed for each ancestral population individually until we did not detect any outliers. PCA/MOD analysis eliminated 131 patients, and a further 102 patients were eliminated as they were identified as having South Asian ancestry (an ancestry without any applicable gene studies). Therefore, the population analyzed was 1540.

Statistical analysis. Since sex and ancestral population are suggested to influence genetic susceptibility to SLE, we first determined the distribution of both across all age groups. We also determined the association of these factors with GRS and age of SLE diagnosis in each model. We used linear regression analysis to determine if GRS varied by age of SLE diagnosis for the whole cohort. As the age of diagnosis distribution did not follow a normal distribution, the natural logarithm of the age of diagnosis was used for statistical analysis as it followed a normal distribution.

Both age of SLE diagnosis (dependent variable) and GRS (predictor variable) were analyzed as continuous variables. A P value of < 0.05 was considered statistically significant. The percent variation in the dependent variable explained by the predictor is quantified using the adjusted r² statistic in each model. The effects of sex and ancestral population were tested for interactions in the final model. SNP & Variation Suite v8 software (Golden Helix Inc.), R version 3.1.2 statistical package (R Foundation for Statistical Computing), and StatPlus:mac (AnalystSoft) were used for the statistical analyses.

RESULTS

Descriptive statistics. The analyzed cohort consisted of 1540 patients: 1351 females and 189 males; 1094 were of White (71.0%), 196 of Black (12.7%), 129 of Hispanic (8.4%), and 121 of Asian (7.9%) ancestry. Mean age of diagnosis in the total population was 25.3 years (SD 14.2) and median age was 21.0 years (Table 1). The mean age at diagnosis for females at 25.8 (SD 14.3) years was statistically significantly higher than that of males at 21.5 (SD 13.1, P = 4.16 × 10⁻³). The percentage of cSLE patients in the total male population (59.3%) was higher than the percentage of cSLE patients in the total female population (41.7%), with lower absolute number (112 males vs 563 females).

Non-HLA GRS. We initially determined the association of sex and ancestral population with GRS and age of SLE diagnosis. We found that for all non-HLA GRS, there were statistically significant associations between sex and age of SLE diagnosis. As a result, sex was included in our statistical model. There was a statistically significant association between sex and GRSWS (P = 0.013, regression coefficient = 0.180), but not between sex and the other non-HLA GRS. Significant associations were seen with age of diagnosis and specific ancestries for GRS (P < 2 × 10⁻¹⁶ for White and P = 0.033 for Hispanic ancestry) and GRSCS (P < 2 × 10⁻¹⁶ for White and P = 0.001 for Black ancestry), whereas for GRWS and GRSWS, the only statistically significant association with age of diagnosis was in White ancestry (P < 2 × 10⁻¹⁶ for GRWS and GRSWS). Therefore, the final statistical model for GRS and GRSCS included sex and all 4 ancestral populations while the final model for GRWS and GRSWS included sex and White vs non-White ancestry as covariates.

The final linear regression models for all non-HLA GRS, except the GRCs, showed a statistically significant negative association with age (Table 2). These models explained similar percentages of the variance of the genetic contribution (GRWS 17.8%, GRSCS 17.5%, and GRSWS 17.6%; Supplementary Figures 1–3, available with the online version of this article).

HLA GRS. There were statistically significant associations between sex and age of SLE diagnosis for all 4 HLA GRS. Therefore, sex was included in our statistical model. However, there were no statistically significant associations between sex and any of the 4 HLA GRS. All the ancestral populations showed statistically significant associations with age of SLE diagnosis for all 4 risk scores. However, in contrast to the results for non-HLA GRS, the final linear regression models of all 4 HLA GRS with age of SLE diagnosis, which included sex and all 4 ancestral populations as covariates, showed statistically significant positive associations with age (Table 3). All 4 models explained almost identical percentages of the variance of the genetic contribution (17.6% or 17.7%; Supplementary Figures 4–7, available with the online version of this article).
It has been suggested that the genetic contribution to the development of SLE likely differs between cSLE and aSLE. However, candidate gene studies have not identified any genes that were specific or unique to cSLE. No GWAS has been performed in cSLE, although a study has suggested that unique SNPs were found in a Korean cSLE population. For these reasons, it was necessary to apply a different genetic approach to better understand the genetics of SLE across all ages. In our study, we used polygenic risk scores to better understand the genetic association with age of SLE diagnosis. Our data have suggested that there is a linear relationship between genetic risk and age of SLE diagnosis, and that HLA and non-HLA GRS influence age of SLE diagnosis differently.

**DISCUSSION**

It has been suggested that the genetic contribution to the development of SLE likely differs between cSLE and aSLE. However, candidate gene studies have not identified any genes that were specific or unique to cSLE. No GWAS has been performed in cSLE, although a study has suggested that unique SNPs were found in a Korean cSLE population. For these reasons, it was necessary to apply a different genetic approach to better understand the genetics of SLE across all ages. In our study, we used polygenic risk scores to better understand the genetic association with age of SLE diagnosis. Our data have suggested that there is a linear relationship between genetic risk and age of SLE diagnosis, and that HLA and non-HLA GRS influence age of SLE diagnosis differently.

**Table 1. Demographic characteristics of patients.**

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
<th>White</th>
<th>Black</th>
<th>Hispanic</th>
<th>Asian</th>
<th>Total Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of diagnosis, yrs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>25.81 (14.3)</td>
<td>21.48 (13.1)</td>
<td>28.65 (14.36)</td>
<td>17.58 (9.84)</td>
<td>18.48 (11.025)</td>
<td>14.50 (7.49)</td>
<td>25.28 (14.2)</td>
</tr>
<tr>
<td>Minimum</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>4.3</td>
<td>1</td>
</tr>
<tr>
<td>Maximum</td>
<td>82.14</td>
<td>64.55</td>
<td>82.14</td>
<td>54.98</td>
<td>56.76</td>
<td>54.40</td>
<td>82.14</td>
</tr>
<tr>
<td>cSLE (&lt; 18 yrs), n (%)</td>
<td>563 (41.67)</td>
<td>112 (59.25)</td>
<td>329</td>
<td>144</td>
<td>90</td>
<td>112</td>
<td>675 (43.83)</td>
</tr>
<tr>
<td>aSLE (≥ 18 yrs), n (%)</td>
<td>788 (58.32)</td>
<td>77 (40.74)</td>
<td>765</td>
<td>52</td>
<td>39</td>
<td>9</td>
<td>865 (56.16)</td>
</tr>
</tbody>
</table>

aSLE: adult-onset SLE; cSLE: childhood-onset SLE; SLE: systemic lupus erythematosus.

**Table 2. Linear regression analyses of the non-HLA genetic risk scores (GRS).**

<table>
<thead>
<tr>
<th></th>
<th>Genetic Risk Counting Scores</th>
<th>Genetic Risk Weighted Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GRCS</td>
<td>GRSCS</td>
</tr>
<tr>
<td>Final model P value*</td>
<td>0.103</td>
<td>0.008</td>
</tr>
<tr>
<td>Slope*</td>
<td>−0.066</td>
<td>−0.044</td>
</tr>
<tr>
<td>Adjusted $r^2$</td>
<td>0.175</td>
<td>0.178</td>
</tr>
</tbody>
</table>

* Final statistical model of the genetic risk weighted scores includes sex and White vs non-White ancestry as covariates; final model of the genetic risk counting scores includes sex and all 4 ancestral populations as covariates. Values in bold are statistically significant. Slope represents the rate of change in $y$ (age of diagnosis) as $x$ (non-HLA GRS) changes. GRCS: genetic risk counting score; GRSCS: genetic risk standardized counting score; GRWS: genetic risk weighted score; GRSWS: genetic risk standardized weighted score.

**Table 3. Linear regression analyses of the HLA genetic risk scores (GRS).**

<table>
<thead>
<tr>
<th></th>
<th>Genetic Risk Counting Scores</th>
<th>Genetic Risk Weighted Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GRCS</td>
<td>GRSCS</td>
</tr>
<tr>
<td>Final model P value*</td>
<td><strong>0.049</strong></td>
<td><strong>0.022</strong></td>
</tr>
<tr>
<td>Slope*</td>
<td>0.014</td>
<td>0.013</td>
</tr>
<tr>
<td>Adjusted $r^2$</td>
<td>0.176</td>
<td>0.176</td>
</tr>
</tbody>
</table>

* Final model of all HLA GRS includes sex and all 4 ancestral populations as covariates. Slope represents the rate of change in $y$ (age of diagnosis) as $x$ (HLA GRS) changes. Values in bold are statistically significant. GRCS: genetic risk counting score; GRSCS: genetic risk standardized counting score; GRWS: genetic risk weighted score; GRSWS: genetic risk standardized weighted score.
and non-HLA SNPs in determining the association of GRS and age of SLE onset. We suggest that the contribution of non-HLA SNPs may be more important in the development of SLE in the younger patient, whereas the contribution of HLA SNPs may be more important in the development of the disease in later years. Previous publications have used different GRS (counting, weighted, or both) to predict the risk of multiple autoimmune diseases. In SLE, there have been 7 publications that used polygenic risk scores to determine risk of SLE but only 4 examined them as a predictor of age of disease onset. These 4 studies used either a counting score and/or weighted score. We therefore took the approach to examine multiple GRS in order to determine which gave the most robust results. We found that GRSWS were optimal. These scores were more robust than additive scores as they are not affected by sample size and strength of marker interactions, and they account for differences in effect size.

Previous studies in SLE have shown conflicting results regarding sex and GRS. An initial study using a weighted score showed that men had a higher genetic risk than women (largely the result of HLA SNPs). However, a second study using a weighted score but smaller cohort and different HLA SNPs (only 13 SNPs were shared) did not replicate the finding. Since the main difference between the 2 previous studies was the different HLA SNPs included, we covered all of the HLA SNPs used in both investigations (either with the same SNP or with an SNP in high LD). A small study of 75 cSLE patients, using only 7 SNPs and none in the HLA region, did not find a significant difference in the GRCs between the sexes. We found a significant association between sex and age of SLE diagnosis in the analysis of all the GRS (HLA and non-HLA) for all ages; therefore, we used sex as a covariant because it was a potential confounder. When we controlled for the variation in sex between ages, only the HLA GRSWS showed a statistically significant association with sex: male patients showed an increase in their non-HLA GRSWS compared to female patients. Although this difference was low (regression coefficient = 0.180), it was statistically significant and replicated what was found in the largest study. More investigations are needed to confirm these conclusions.

It is well-recognized that SNPs associated with susceptibility differ across ancestral populations. In both cSLE and aSLE populations, SLE is more prevalent in non-White populations (Hispanic, Black, Native American, and Asian). Thus, in the calculation of the weighted score, each SNP was weighted according to its effect in the population studied. Moreover, when we analyzed the associations between age of SLE diagnosis and ancestral populations, we found statistically significant associations for all the GRS. Therefore, in the final linear regression models of all GRS with age of SLE diagnosis, we included ancestral populations as covariates. This is the first time that effects of genetic ancestry on GRS have been addressed, to our knowledge. One limitation of our study may be that all genotyping was performed on the Immunochip platform. This platform was designed for use in European populations and therefore is less informative in other ethnic populations, with poorer coverage of SNPs associated with the development of SLE in non-White populations. In addition, there is increasing evidence that rare variants may be important in the development of SLE, but only a few of these rare variants are present on the Immunochip. Although we were able to examine only 10/21 (47.6%) HLA SNPs validated in SLE GWAS, this is much greater coverage than in previous studies that used polygenic risk scores. Regarding the SNP coverage in each ancestral population, most of the non-HLA and HLA SNPs validated by GWAS and meta-GWAS for White and Asian populations were analyzed. However, for Black and Hispanic populations, there were no published GWAS and therefore, candidate gene studies were used. The resulting differences in the number of SNPs covered in each population were overcome by the process of GRS standardization, but it is still likely that some relevant SNPs in Black and Hispanic groups were missed. However, by standardizing the genetic score, the effect of any single locus will differ between ethnicities. Finally, there were no publications in South Asian populations and therefore, we could not include this population. However, our coverage of both HLA and non-HLA SNPs is the largest published to date in each ancestral population, to our knowledge. Another limitation of our study is the possibility of unmeasured confounders that can affect genetic scores and age of diagnosis of SLE. This could be the case in our models: because our dependent variable is age of SLE diagnosis and not age of SLE onset, there may be differential bias (e.g., time to diagnose in cSLE vs aSLE patients, males vs females; access to care, socioeconomic factors). However, we strongly believe that our results are still valid as a starting point for future investigations.

The present study is the first to show that there are different effects of non-HLA and HLA GRS on age of SLE diagnosis in a multiethnic population, to our knowledge. Specifically, non-HLA GRS showed that the higher the number of cSLE-associated non-HLA SNPs, the younger the age of SLE diagnosis. Conversely, the higher the HLA GRS, the older the age of SLE diagnosis. These results were consistent across all methods of estimating their effect. We suggest that the non-HLA GRSWS may be the most robust score to use as it has shown the highest degree of statistical significance. However, for HLA risk, all the risk scores performed well. Overall, GRS explained 18% of the variance of age of SLE onset. We suggest that future studies use standardized GRS, which accounts for the variability in the distribution of the scores across populations, examines the effect of sex on risk scores, and determines the effect of HLA and non-HLA risk scores separately. These findings emphasize the complexity of the influence of genetic risk on the age of SLE onset.

ACKNOWLEDGMENT

REFERENCES


31. Han S, Kim-Howard X, Deshmukh H, Kamatani Y, Viswanathan...


