The Genetic Contribution to Systemic Lupus Erythematosus

Lindsey A. Criswell, M.D., M.P.H.

Abstract
We are currently witnessing an explosion of information about the genetic contribution to complex human diseases such as systemic lupus erythematosus (SLE). These genetic discoveries have profound implications for efforts to more fully characterize etiologic pathways and mechanisms. Further, these etiologic insights should lead to improved diagnostic and prognostic tools and inform the development of more specific therapies for SLE and related conditions. The article summarizes the evidence supporting a role for genetic factors in SLE, highlights the clinical and genetic complexity of the disease, reviews the genes that have established contributions to the risk of SLE and specific disease manifestations, and discusses the recent data emerging from genomewide association studies of SLE.

Systemic lupus erythematosus (SLE) is the prototypic systemic autoimmune disease characterized by autoantibody production and involvement of multiple organ systems. A striking feature of the disease is the clinical heterogeneity, resulting in variable disease manifestations and outcome. Thus, individual patients vary in terms of the specific autoantibodies produced and the presence of skin, joint, hematologic, neurologic, renal, and other organ manifestations.

Similar to other systemic rheumatic diseases, and in part due to this clinical heterogeneity, there is no single diagnostic test that can establish a diagnosis of SLE. Indeed, SLE shares features with other systemic autoimmune diseases, and classification of disease status is based on a set of defined classification criteria.\(^1\)\(^2\) This clinical complexity likely reflects variation in underlying etiologic factors, and thus efforts to identify causative genes or other risk factors must account for this phenotypic heterogeneity.

Etiologic Factors in SLE
Although efforts to identify genetic risk factors have been hampered by these and other complexities, it has been clear for decades that genetic factors contribute importantly to disease risk. However, it is also clear that other factors are important. For example, several lines of evidence, including work in animal models of lupus, relatively lower rates of female predominance in childhood and late-onset human SLE, and disease flares during pregnancy highlight the importance of hormonal factors.\(^3\) There is also growing evidence that drugs and specific factors in the environment, such as exposure to tobacco smoke\(^4\) and infectious agents, such as the Epstein-Barr virus,\(^5\) also contribute to SLE risk.

Importance of Genetic Factors in SLE
Several lines of evidence document the importance of genetic factors in SLE. For example, the disease exhibits familial clustering, with 10% to 12% of SLE patients having an affected first-degree relative. Differences in the concordance rate for disease among monozygotic (MZ) and dizygotic (DZ) twins also support an important role for genes. The available data indicate a concordance rate for SLE between 24% and 69% for MZ, compared to 2% to 9% concordance for DZ twins.\(^6\)\(^7\) However, the lack of complete concordance for SLE among MZ twins also highlights the importance of nongenetic factors.

More recently, assessment of the overall genetic contribution to complex human diseases such as SLE often is expressed in terms of a parameter known as \(\lambda_s\), where the “s” refers to “sibling.” This parameter is simply the ratio of...
the risk to siblings of an affected individual divided by the background population prevalence of the disease.\(^8\) Larger values of \(\lambda_s\) are interpreted to indicate a greater genetic contribution to disease. Table 1 displays estimates of \(\lambda_s\) for six autoimmune diseases, including SLE. These figures indicate that, overall, the genetic contribution to SLE is relatively large, at least compared to the other human autoimmune diseases listed. Presumably, this will translate into an increased ease in defining the specific genes that influence disease risk; however, that will depend in part on the total number of genes that contribute to \(\lambda_s\), as well as other considerations, such as genetic heterogeneity. Nonetheless, these data are encouraging in terms of the likely success of efforts to define the genetic contribution to SLE.

### SLE as a Genetically Complex Trait

Last, before considering the specific genes that have been identified through prior and recent genetic linkage and association studies of SLE, it is important to understand the status of SLE as a genetically complex trait and the implications of that complexity for efforts to define the specific genetic determinants. Most human diseases are now recognized to be genetically complex. In brief, this means that there are at least several, and, possibly very many, disease predisposing genetic loci, as well as an important role for nongenetic factors. Further, genetically complex diseases do not exhibit straightforward Mendelian modes of inheritance (e.g., dominant, recessive, etc.). As well, there is a lack of correspondence between the inherited genotype(s) and the resultant phenotype(s), which may be due to incomplete penetrance of predisposing loci or to phenocopies, that is disease that develops in the absence of apparent genetic risk factors. Genetically complex diseases also typically exhibit genetic heterogeneity, which may relate to differences in clinical features (i.e., the specific phenotype) or ethnic background.

There are a number of implications of this complexity for efforts to define the genetic contribution to SLE and other complex diseases. First, we must be prepared to search for multiple genetic risk factors for disease, each one of which may have only a modest impact on disease risk. We must also anticipate epistasis, that is, interactions between genes or between genes and environmental factors as potentially important determinants of disease risk. We must also pay close attention to the specific clinical features present, as well as ethnic background, in the design and interpretation of genetic studies. Indeed, the striking clinical heterogeneity and ethnic differences in SLE risk and severity suggest that these issues will be critically important in the design and interpretation of genetic studies. Together, these complexities translate into the need for extremely large sample sizes to ensure adequate power to overcome these obstacles to gene discovery efforts in SLE and related diseases.

Figure 1 illustrates the challenges we face as a result of these genetic complexities. This hypothetical example shows three individual susceptibility genes, each one of which, on its own, increases the risk of disease only modestly (i.e., a one- to two-fold increase in risk, shown in Figure 1 as \(\lambda\)). Only when two or more of these genetic risk factors occur together in an individual is the risk of

### Table 1 Population Frequencies and Familial Clustering of Human Autoimmune Diseases*

<table>
<thead>
<tr>
<th>Autoimmune Disease</th>
<th>Population Frequency (%)</th>
<th>Sibling Risk (%)</th>
<th>(\lambda_s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psoriasis</td>
<td>2.8</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>1.0</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>0.4</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>Graves’ disease</td>
<td>0.5</td>
<td>7.5</td>
<td>15</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>0.1</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>0.05</td>
<td>1.5</td>
<td>30</td>
</tr>
</tbody>
</table>

*Adapted from Vyse and colleagues\(^{48}\) and Alarcon-Segovia and coworkers.\(^{49}\) \(\lambda_s\) refers to the ratio of disease risk among siblings of an affected individual divided by the population frequency of the disease.\(^8\) Higher values of \(\lambda_s\) indicate a greater total genetic contribution to disease risk.
disease increased more substantially.

**Contribution of HLA Region Genes to SLE**

Like most human autoimmune diseases, genes within the human leukocyte antigen (HLA) region on the short arm of chromosome 6 (6p21.3) exhibit strong association with the risk of SLE and production of specific autoantibodies that are commonly present in SLE. In spite of the strength of association observed for this genomic region, the challenge has been in determining which of the many genes within this region are primarily responsible for the disease associations. Not only are there literally hundreds of genes in this region, many of which have obvious or suggested immune-related functions, but the region is also characterized by a high degree of linkage disequilibrium, which results in long stretches of DNA encompassing many genes that are inherited together as a unit (referred to as a haplotype). Determining which genes on such extended haplotypes are primarily responsible for the genetic association can be difficult or problematic.

Historically, interest in this region for SLE has focused on the highly polymorphic HLA class I and II genes that encode membrane glycoproteins that present peptides for recognition by T lymphocytes, as well as genes within the HLA class III region, particularly tumor necrosis factor (TNF) and complement component C4 gene loci. Indeed, inherited deficiency of complement genes, particularly C4A (null) alleles, has long been recognized as a strong, albeit rare, genetic risk factor for SLE. More recently, work by Graham and colleagues, involving an analysis of approximately 50 microsatellite genetic markers across the HLA region among a set of 780 SLE families, highlighted the importance of HLA class II haplotypes involving the HLA-DRB1 and -DQB1 loci, particularly those corresponding to serologic types HLA-DR2 and DR3.

A more recent analysis of 314 UK SLE families involving 68 single nucleotide polymorphisms (SNPs) across the region demonstrated the presence of two distinct and independent association signals, one involving the class II HLA-DRB1, -DQA1, and -DQB1 loci (serologic type DR3) and another in the class III region. Of interest, the class III signal was strongest for a marker in the SKIV2L gene and appears to exclude the TNF-308 promoter polymorphism. Additional work will be required to more precisely define the gene or genes responsible for the complex HLA association with SLE.

**Progress Identifying Non-HLA Genetic Risk Factors for SLE**

Although we have not yet fully characterized the contribution of HLA region genes to SLE, it is clear that genes outside this region also contribute to disease risk. This may be particularly true for non-Caucasian populations, given the overall weaker evidence of linkage and association to the HLA region among these populations. In spite of our incomplete understanding of underlying disease mechanisms in SLE, there has been no shortage of candidate genes hypothesized or implicated in the risk of disease or specific disease manifestations. However, with a few exceptions (reviewed below) past efforts to identify genetic risk factors through traditional genetic association or linkage studies have been relatively unsuccessful, likely due to false negative results from inadequately powered studies, false positive results due to population admixture (a form of confounding due to unaccounted for differences in ancestry between case and control study populations), variation in clinical and ethnic characteristics of patients within and between studies, and true negative results of studies focused on the wrong candidate genes. In these respects, the experience and results of genetic association studies in SLE, until very recently, have been similar to lack of progress in other genetically complex human diseases, including other autoimmune and rheumatic diseases.

In spite of such disappointing results in previous studies, a handful of genes are now well established risk factors for SLE, based on extensive work over the past decade. Further, very recently, the culmination of tremendous advances in molecular genetic and statistical methodology now allow for the performance of much more comprehensive screens of the genome, in order to identify most of the common variants that predispose to complex diseases such as SLE. These studies, typically referred to as genomewide association studies, or scans (GWAS), are now being completed in SLE and other autoimmune diseases. And even at this very early stage, it is clear that we are on the brink of a much more complete delineation of the genes that contribute to SLE and related phenotypes. In the remainder of this article, genes that were established as risk factors for SLE prior to the recent GWAS are summarized, and findings from the first published GWAS in SLE are highlighted.

**FcγR 2A and 3A Genes in SLE**

Fc receptors for immunoglobulin G (FcγR) mediate clearance of immune complexes and have been strongly implicated in the pathogenesis of SLE and lupus nephritis. Thus, functional polymorphisms of the genes that encode these receptors, particularly FcγR 2A and 3A, have been the focus of many genetic studies in SLE. Substantial evidence, including several meta-analyses, confirms a role for these genes in disease risk although the magnitude of association is relatively modest, consistent with expectations for a genetically complex disease. For example, a meta-analysis of 17 studies, which included 3114 SLE patients and 2580 controls, indicated that the presence of two copies of the R131 variant of the FcγR2A gene is associated with a modest increase in risk of SLE (odds ratio (OR) = 1.3). Another meta-analysis, which included 481 SLE patients with the antiphospholipid antibody syndrome (APS), 1420 non-APS SLE patients, and 1665 healthy controls, suggests that the risk of APS among individuals with two copies of the R131
variant have an even higher risk for developing APS (OR = 1.65 for APS vs. healthy controls).\textsuperscript{15}

Last, a meta-analysis of the V/F158 polymorphism of the FcγR3A gene, which included a total of 1154 patients with lupus nephritis and 1261 SLE patients without nephritis, from 11 studies, suggests that the F158 variant is associated with an increased risk of lupus nephritis (OR = 1.47, p = 0.006, for individuals with 2 copies of the risk allele).\textsuperscript{16}

**PTPN22 in SLE**

In 2004, Bottini and colleagues\textsuperscript{17} first reported the association of a functional polymorphism (the missense SNP R620W) of the intracellular protein tyrosine phosphatase type N22 (PTPN22) with human autoimmune disease. The PTPN22 gene encodes the cytoplasmic lymphoid specific phosphatase Lyp, which is a powerful inhibitor of T-cell activation. That study, involving North American and Italian patients with type 1 diabetes, quickly led to replication of the association in other autoimmune diseases, including RA\textsuperscript{18} and SLE.\textsuperscript{19} The R620W polymorphism results in an arginine to tryptophan substitution at a key residue in the P1 proline-rich motif that binds Lyp to the SH3 domain in Csk. The R620W substitution disrupts binding of Lyp to Csk.\textsuperscript{17} In addition to the functional nature of the polymorphism, the report of subtle autoimmune features in a PTPN22 knockout animal model\textsuperscript{20} stimulated interest in PTPN22 among investigators of human autoimmune disease. This gene has now been examined in a number of SLE genetic studies and, although findings have been somewhat mixed, overall the evidence suggests that this variant, and particularly inheritance of a double dose of the 620W variant, contributes to the risk of SLE, with an odds ratio of association as high as 3 to 4 for the presence of two copies of the risk variant.\textsuperscript{21} A meta-analysis of PTPN22 studies in SLE has recently been completed by Lee and coworkers.\textsuperscript{22} Also of interest is the clear evidence of a lack of association with multiple sclerosis and certain other autoimmune diseases, which has led some to speculate that this gene plays a primary role in autoantibody production.

**PDCD1 (PD1) in SLE**

Prokunina and associates\textsuperscript{23} first reported an association of the programmed cell death 1 gene (PD1) with SLE\textsuperscript{23} in a large collection of European and Mexican patients. Their interest in this gene related to work in animal models indicating that deletion of the gene resulted in an SLE-like phenotype, findings from a linkage study performed in Swedish and Icelandic families that highlighted the genomic region containing PD1 (2q37), and several functions of the gene that made it a good candidate for involvement in SLE pathogenesis.\textsuperscript{23} Their initial work in SLE highlighted an SNP in the fourth intron of the gene due to evidence that the polymorphism altered the RUNX1 binding site, which is consistent with an important role for apoptosis in SLE.\textsuperscript{24}

Since the initial report, not all studies of PD1 in SLE have yielded consistent results, suggesting at least some population differences in the genetic association,\textsuperscript{25,26} as well as the importance of specific disease features.\textsuperscript{27,28} Consistent with findings in a PD1 knockout animal model, variation in PD1 appears to be associated with lupus nephritis, at least among Northern European populations.\textsuperscript{29-31} It is also possible that additional variants in the region contribute to disease risk and may explain some of the discrepant findings.

**Multiple IRF5 Polymorphisms Contribute to SLE Risk**

Recent work by Graham and colleagues provides an elegant example of how multiple polymorphisms within a single genomic region may interact in complex ways to influence disease risk. Several lines of evidence implicate the interferon pathway in SLE pathogenesis, and thus the report in 2005 by Sigurdsson and coworkers\textsuperscript{32} of genetic association with the interferon regulatory factor 5 (IRF5) gene among Swedish and Finnish SLE cases and controls quickly led to attempts at replication in other populations. Graham and associates\textsuperscript{33} published the results of a meta-analysis involving a total of 2250 SLE cases and 2855 controls, yielding compelling statistical evidence of association with the same SNP highlighted in the original report (rs2004640), with $p = 4.2 \times 10^{-21}$ and OR = 1.47. Of interest is subsequent work by Graham and colleagues\textsuperscript{34} indicating that three functional polymorphisms within the IRF5 genomic region interact in complex ways to influence SLE risk. For example, depending upon the specific alleles present on a 3-marker haplotype, the risk of SLE may be increased, decreased, or neutral. It is likely that future work involving resequencing and more detailed analysis of multiple variants across disease-associated regions will yield similar complex genetic association signals for SLE and related diseases.

**STAT4 and the Risk of SLE**

As described above for PTPN22, in which a single missense SNP has now been established as a genetic risk factor for multiple autoimmune diseases, shared genetic risk factors between multiple autoimmune diseases has now emerged as an increasingly common phenomenon. Another example is provided by recent work demonstrating association between the signal transducer and activator of transcription 4 (STAT4) gene and both RA and SLE. In brief, as part of follow-up work to a genome wide linkage analysis performed in a large collection of multicase RA families,\textsuperscript{35} 13 candidate genes located within a linkage peak on chromosome 2q were screened for association with RA. The strongest evidence of association was within the STAT1-STAT4 region. Subsequent fine mapping of the region in a large set of RA cases and controls from the U.S. and Sweden defined a SNP haplotype in the third intron of STAT4 with strong statistical evidence of association with RA (OR = 1.32, $p = 2.81 \times 10^{-7}$).\textsuperscript{36} The same set of SNPs defining the RA-associated haplotype were then tested in several independent
SLE case series and healthy controls, yielding even stronger evidence of disease association (OR = 1.55, p = 1.87 x 10^-8). Furthermore, homozygosity for the risk alleles was associated with more than doubling the risk for SLE and a 60% increased risk for RA.

STAT4 encodes a transcription factor that transmits signals induced by several key cytokines, including IL-12 and IL-23, and differentiation of both the Th1 and Th17 proinflammatory T-cell lineages requires STAT4-dependent cytokine signaling. Work in STAT4-deficient mice also implicates an important role for STAT4 in autoimmunity. Thus, although additional work will be required to fully characterize the association of this gene with the spectrum of human autoimmune diseases, knowledge that genetic variation in the STAT4 gene contributes to the development of SLE and RA provides new targets for the development of treatment strategies for these chronic, debilitating diseases.

Genomewide Association Studies (GWAS) in SLE

During the past year, there has been an explosion of information about new genes or genomic regions that contribute to complex human diseases, including several autoimmune diseases. This information has emerged largely from GWAS, investigations which have typically involved characterization of large samples of cases and controls (or families) with hundreds of thousands of SNP markers genomewide. In addition to the substantial clinical, informatics, and financial resources required for performance of these studies, sophisticated and complex statistical methods are also required in order to adequately account for problems related to multiple testing and population admixture. However, appropriate tools and methods are now available that allow for the generation of reliable and compelling data from these huge experiments.

Two high-density and one low-density GWAS in SLE have recently been published, yielding definitive data about several new risk genes, as well as valuable confirmatory data related to several previously identified genes, including STAT4 and IRF5. Although results of GWAS in SLE and other human autoimmune diseases have significantly advanced our understanding of the genetic contribution to these diseases, it is also clear that even larger sample sizes and denser or different genetic marker sets will be required to allow us to fully benefit from this new and exciting methodology.

All of the recently published SLE GWAS utilized case control study designs and subjects of European ancestry, although the study by Harley and colleagues restricted the analysis to female cases from families with a history of SLE or related conditions in an attempt to decrease heterogeneity related to gender and select for individuals with a heavy burden of genetic risk factors. They utilized a genomewide set of ~300,000 SNP markers and analyzed 720 female SLE cases and 2337 female healthy controls in the initial phase of their experiment. The investigators then genotyped an additional set of 1846 SLE cases and 1825 controls (all female) for SNP, with strong evidence of association in the first stage. The replication of findings in a second set of cases and controls is a crucial element of the design to try to distinguish true from false positive findings. Their final analysis revealed three new genes or genomic regions with definitive evidence of association with SLE (p < 10^-8) and several other loci associated with somewhat lower levels of statistical significance (10^-4 < p < 10^-6). Although these thresholds for statistical significance may appear to be excessively rigorous, they reflect the very large multiple testing burden inherent to such studies.

The top four associated novel genetic loci in the study by Harley and coworkers were integrin-alphaM (ITGAM, 16p11.2), KIAA1542 (11p15.5), PXK (3p14.3) and 1q25.1. The HLA region revealed the strongest evidence of association overall (OR = 2.36, p = 1.7 x 10^-52 for marker rs313139). Work by Nath and colleagues published concurrently demonstrates that the ITGAM association is present in patients of both European and African descent, and that the likely causal variant is a nonsynonymous SNP (rs1143679) that converts the normal arginine at amino acid position 77 to a histidine (R77H).

The high-density GWAS study by Hom and associates also employed a two-stage design, with 1311 SLE cases and 3340 controls characterized for over 500,000 SNP markers genomewide. They replicated their top two novel association results in a set of 793 Swedish SLE cases and 857 Swedish controls. Overall, the top three genetic loci in their study were the HLA region, and the recently identified STAT4 and IRF5 loci. In addition, two novel loci met genomewide levels of significance (p < 10^-8), ITGAM and genetic variation in the region upstream from the gene encoding B lymphoid tyrosine kinase (BLK) and C8orf13 (chromosome 8p23.1). Functional studies suggested that genetic variation in the 8p23.1 region is associated with altered levels of messenger RNA in B-cell lines. Clearly, however, additional genetic and functional work will be required to clarify which variant(s) within each of the associated regions explains association with disease and to define the underlying molecular mechanisms.

Last, the low-density GWAS recently published by Kozyrev and colleagues suggest that functional variants in the B-cell gene BANK1 (B-cell scaffold protein with ankyrin repeats) are associated with SLE. Their study provides strong genetic association results for a non-synonymous SNP (rs10516487, R61H, combined p = 3.7 x 10^-10; OR = 1.38) and evidence that several BANK1 variants affect regulatory sites and key functional domains. Thus, they hypothesize that BANK1 variants may contribute to the sustained B cell-receptor signaling and B-cell hyperactivity observed in SLE.

Figure 2, from the recent Hom and coworker study,
nicely illustrates the data generated from the GWAS work. Results for all of the SNPs are shown along the x-axis and are color coded according to chromosome number. The y-axis indicates the degree of statistical significance for the association result for each SNP based on a comparison of SLE cases versus controls. Of note is that the large number of SNPs with association results in the range of $p = 10^{-4}$ to $10^{-6}$. Clearly, the top few hits highlighted in these initial GWAS in SLE represent just the tip of the iceberg, and it is very likely that many more

Figure 2 Five major loci associated with SLE in a genomewide association study of 1,311 SLE cases and 3,340 controls of European ancestry characterized for 502,033 SNP variants. Results for all of the SNPs are shown along the x-axis and color coded according to chromosome number. The y-axis indicates the degree of statistical significance (as represented by $-\log_{10} p$ value) for the association result for each SNP, based on a comparison of SLE cases versus controls.

Figure 3 A model of the pathogenesis of SLE that implicates the products of disease-associated polymorphic genes. This proposed model recognizes the essential roles of both the innate and adaptive immune responses as well as environmental triggers of SLE. (Reproduced with permission from: Crow MK. Collaboration, genetic associations, and lupus erythematosus. N Engl J Med. 2008;358(9):956-61. Copyright 2008 Massachusetts Medical Society. All rights reserved.)
genes will be identified among the hundreds of association signals that are strong but not definitive on the basis of these initial experiments. As a first step, a meta-analysis of these and other GWAS in SLE will help to distinguish the true from false association signals. It is also likely, however, that additional cases and controls will be required to complete this task, particularly for the genes associated with relatively smaller impacts on SLE risk. Further, stratification or re-analysis of these and future GWAS datasets according to more homogeneous phenotypes (e.g., SLE characterized by the presence of nephritis or specific autoantibodies) may also reveal new loci more strongly associated with these specific phenotypes, or variation in risk for certain loci according to specific SLE-related phenotypes.

Progress in Defining a Comprehensive Model of SLE Pathogenesis Based on the Underlying Genetic Determinants

A model of SLE pathogenesis is illustrated in Figure 3 that implicates the products of disease-associated polymorphic genes, including many of the genes highlighted in this review. This proposed model recognizes the essential roles of both the innate and adaptive immune responses, as well as endogenous or environmental triggers of SLE. Recent genetic discoveries significantly advance our understanding of disease mechanisms in SLE. Together, these findings support important roles for B-cell receptor signaling pathways and mechanisms that regulate the adhesion of inflammatory cells to the vasculature. During the coming years, as additional genetic and functional studies further define the genetic architecture of SLE and relevant underlying molecular mechanisms, additional targets for novel therapies and improved diagnostic and prognostic tests will be identified, substantially improving our ability to diagnose and treat this potentially devastating systemic autoimmune disease.

Disclosure Statement

The author has no financial or proprietary interest in the subject matter or materials discussed, including, but not limited to, employment, consultancies, stock ownership, honoraria, and paid expert testimony.

References