Review Article

Role of Cytokines in Systemic Lupus Erythematosus: Recent Progress from GWAS and Sequencing

John J. Connolly¹ and Hakon Hakonarson¹, ²

¹ Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
² Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA

Correspondence should be addressed to Hakon Hakonarson, hakonarson@chop.edu

Received 2 December 2011; Revised 24 February 2012; Accepted 24 February 2012

Academic Editor: Andrei Surguchov

Copyright © 2012 J. J. Connolly and H. Hakonarson. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Systemic lupus erythematosus (SLE) is a complex autoimmune disorder, known to have a strong genetic component. Concordance between monozygotic twins is approximately 30–40%, which is 8–20 times higher than that of dizygotic twins. In the last decade, genome-wide approaches to understanding SLE have yielded many candidate genes, which are important to understanding the pathophysiology of the disease and potential targets for pharmaceutical intervention. In this paper, we focus on the role of cytokines and examine how genome-wide association studies, copy number variation studies, and next-generation sequencing are being employed to understand the etiology of SLE. Prominent genes identified by these approaches include BLK, FCγR3B, and TREX1.

Our goal is to present a brief overview of genomic approaches to SLE and to introduce some of the key discussion points pertinent to the field.

1. Introduction

Systemic lupus erythematosus (SLE) is a highly heterogeneous autoimmune disorder characterized by the prevalence of autoantibodies directed against double-stranded DNA. Small nuclear RNA-binding proteins, including anti-Ro, anti-La, and anti-RNP are also found in many patients. Worldwide, the prevalence is approximately 52 per 100,000 and may be highest among individuals of Afro-Caribbean descent at 159 per 100,000 (derived from a UK sample) [1]. In the United States, SLE is 2.6 times more common in individuals of African descent at 159 per 100,000 (derived from a UK sample) [1].

SLE is underscored by a range of environmental and genetic risk factors that can affect any part of the organ system, including cardiovascular, hematological, integumentary, musculoskeletal, nervous, renal, and respiratory. In this paper, we focus on three major approaches that are being used to uncover genetic correlates of the disease, genome-wide association studies (GWAS), copy number variation (CNV) studies, and next-generation sequencing (NGS). Our goal is to provide a brief overview of the genomic landscape and to introduce some of the key discussion points in terms of how these approaches inform our understanding of SLE, particularly in relation to cytokines. To date, more than 40 relevant loci have been identified and replicated; many of which directly or indirectly involve cytokines regulation.

Cytokines are important components of immune response and regulation and play an active role in activating, differentiating, and maturing immune cells [6]. An imbalance between pro- and anti-inflammatory cytokines is a well known characteristic of SLE [7]. They are heavily integrated in T-Cell and B-Cell signaling systems and abnormal cytokine levels, particularly interleukins, interferons, and tumor necrosis factors (TNFs), are important hallmarks of SLE.

2. Genome-Wide Association Studies

Genome-wide association studies began to proliferate approximately 6 years ago, coinciding with rapid developments in genomic hardware and software and falling costs of relevant technologies. GWAS use microarrays to tag up to several million single nucleotide polymorphisms (SNPs) at once, which
gives broad coverage of exonic and intronic regions (though coverage is weighted toward the former). When a difference in SNP-frequency is observed between cases and controls, we can infer a difference in the underlying genomic locus, which may affect gene expression or regulation. Because of the large number of comparisons being made, most GWAS require very large numbers of patients and controls and sample sizes in excess of several thousand are the norm.

Table 1 lists the ~40 loci that have been associated with SLE by GWAS. A number of these genes, particularly in the human leukocyte antigen (HLA) complex, were identified before the genome era. Since the publication of the first GWA study into SLE in 2007, this list has expanded rapidly and at least nine GWAS have been published [8–16]. Prominent replicated genes are discussed below. A major benefit of the GWA approach is its applicability to complex disease.

3. HLA Regulation

The HLA system/histocompatibility complex (MHC) has long been associated with SLE. Because of its long-standing importance to immune function [17], the complete HLA/MHC was one of the first multigenic regions of the human genome to be sequenced [18]. It is densely packed with genes that regulate immune functions and all of the GWAS of SLE converge upon the region as the strongest predictor of the association to the disease [12–16]. HLA genes are classified as types I, II, and III, and the region also includes a number of genes that encode tumor necrosis factor α (TNF-α) cytokines, which are discussed separately below. The class II and class III regions have been most closely associated with SLE.

3.1. HLA Class II. Genes belonging to HLA Class II are prominent candidates for SLE susceptibility and are known to play a major role in T-cell immunity. Graham et al. [19] differentiated three haplotypes in the region that are associated with SLE risk. More specifically, the DR beta 1 gene (HLA-DRB1) has been shown to associate with SLE across a range of ethnic populations, though the specific risk to each individual of European ancestry are not associated with the disease. Because of its long-standing importance to immune function, the complete HLA/MHC region is one of the first multigenic regions of the human genome to be sequenced. It is densely packed with genes that regulate immune functions and all of the GWAS of SLE converge upon the region as the strongest predictor of genetic risk. HLA genes are classified as types I, II, and III, and the region also includes a number of genes that encode tumor necrosis factor α (TNF-α) cytokines, which are discussed separately below. The class II and class III regions have been most closely associated with SLE.

Table 1: Loci associated with SLE from GWAS. Table adapted from http://www.genome.gov/gwastudies/. A number of these genes were associated with SLE prior to GWAS and have now been replicated, while others await replication. Intergenic regions are listed in brackets. It is important to note that many genes associated with SLE in individuals of European ancestry are not associated with the disease in Asian populations, with the converse also holding true. Similar differences in susceptibility are certain to be found between other populations for whom the requisite GWAS have not yet been conducted. As Bustamante et al. [3] and others have pointed out, individuals of European ancestry are heavily overrepresented in GWAS, and many more comparison studies are needed to address this sampling bias.

<table>
<thead>
<tr>
<th>Region</th>
<th>Reported gene(s)</th>
<th>Reported OR Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1p31.1</td>
<td>NER1</td>
<td>EU7</td>
</tr>
<tr>
<td>1p32.3</td>
<td>SLCA7</td>
<td>EU1</td>
</tr>
<tr>
<td>1q25.1</td>
<td>TNFSF4</td>
<td>1.22–1.46 AS4, EU6</td>
</tr>
<tr>
<td>1q25.3</td>
<td>LAMC2</td>
<td>EU1</td>
</tr>
<tr>
<td>1q31.1</td>
<td>IL10 [RPS3A9-FAM5C]</td>
<td>1.19–1.3 AS1,9</td>
</tr>
<tr>
<td>2p16.3</td>
<td>[FHSR-RPL7P13]</td>
<td>NR</td>
</tr>
<tr>
<td>2q23.1</td>
<td>RASGRP3</td>
<td>AS4</td>
</tr>
<tr>
<td>2q31.3</td>
<td>[CWC22-FTHL20]</td>
<td>EU1</td>
</tr>
<tr>
<td>2q32.3</td>
<td>STAT4</td>
<td>1.41–1.77 AS4, EU1,5,7</td>
</tr>
<tr>
<td>3p14.3</td>
<td>PXK</td>
<td>EU6</td>
</tr>
<tr>
<td>3q26.31</td>
<td>NAALADL2</td>
<td>EU1</td>
</tr>
<tr>
<td>4q24</td>
<td>BANK1</td>
<td>AS1</td>
</tr>
<tr>
<td>4q25</td>
<td>COL25A1</td>
<td>EU1</td>
</tr>
<tr>
<td>4q28.3</td>
<td>[PCDH18-SLC7A11]</td>
<td>EU5</td>
</tr>
<tr>
<td>5p13.1</td>
<td>GHR</td>
<td>EU7</td>
</tr>
<tr>
<td>5q33.1</td>
<td>TNIP1</td>
<td>1.23–1.27 AS4, EU9</td>
</tr>
<tr>
<td>5q33.3</td>
<td>PTTPG1</td>
<td>EU1</td>
</tr>
<tr>
<td>6p21.31</td>
<td>UHRF1BP1</td>
<td>EU9</td>
</tr>
<tr>
<td>6p21.32</td>
<td>NOTCH4</td>
<td>1.38–2.36 AS1, EU1,6,7</td>
</tr>
<tr>
<td>6p21.32</td>
<td>C6orf10</td>
<td>EU1,6,7</td>
</tr>
<tr>
<td>6p21.32</td>
<td>TNXB</td>
<td>EU1,6,7</td>
</tr>
<tr>
<td>6q21</td>
<td>PRDM1 ATG5</td>
<td>1.20–1.25 AS4, EU9</td>
</tr>
<tr>
<td>6q23.3</td>
<td>TNEAP3</td>
<td>1.72–2.28 AS4, EU5</td>
</tr>
<tr>
<td>7p12.2</td>
<td>IKZF1</td>
<td>1.39</td>
</tr>
<tr>
<td>7p15.1</td>
<td>JAZF1</td>
<td>EU9</td>
</tr>
<tr>
<td>7q32.1</td>
<td>IREFS</td>
<td>1.43–1.92 AS4, EU1,6,7</td>
</tr>
<tr>
<td>8p23.3</td>
<td>FAM16A-BLK C8orf13</td>
<td>1.29–1.45 AS4, EU1,6,7</td>
</tr>
<tr>
<td>9q34.13</td>
<td>NTNG2</td>
<td>EU7</td>
</tr>
<tr>
<td>10q11.23</td>
<td>WDFY4 c10orf64</td>
<td>1.24–1.30 AS3, EU5</td>
</tr>
</tbody>
</table>
rather than SLE in general. This was also true of observed associations with STAT4, IRF5, and ITGAM, which are discussed further below.

### 3.2. HLA Class III and the Complement System

The strongest single genetic risk factors of SLE are complement defects, particularly a complete deficiency of the C1q immune complex which, when present, is associated with a 93% risk of SLE [25]. Related complement defects in C4, C1r/s, and (to a lesser extent) C2, and C3 are firmly established as risk factors, with a prevalence of 75%, 57%, 32%, and 10%, respectively [26, 27]. These deficiencies are different from most candidate variants in that they are strong predictors if present. However, they are collectively rare, found in only 1-2% of cases [28]. Dunckley et al. [29] reported a significant increase in the C4A null allele (C4AQ0) in Chinese, European, and Japanese populations compared to controls. In a study of Japanese SLE patients, two SNPs in the C3 gene, rs7951 and rs2230201, have been also significantly associated with increased disease risk, with the former correlating with lower C3 serum levels [30].

The Class III gene, mutS homolog 5 (MSH5) gene, has been associated with SLE and was in fact the strongest association in the GWAS by Harley et al. [14]. MSH5 facilitates DNA rearrangements, which cause immunoglobulin class switching [31]. Super viralicidic activity 2-like (SKIV2L) is another Class III gene previously identified as an SLE candidate, and a study of 314 trios from the United Kingdom [32] implicated this locus independent of Class II variants. SKIV2L is a putative RNA helicase and is expressed in T cells, B cells, and dendritic cells [32]. The integrin alpha M gene (ITGAM) has also been associated with SLE by a number of studies [9, 11, 14, 15]. Along with the F2 chain (ITGB2) protein, ITGAM forms a leukocyte integrin, complement receptor 3 (CC3), that facilitates adhesion of neutrophils and monocytes to the stimulated endothelium. ITGAM is also a receptor for the degraded CC3—iC3b [33]. Similarly, missense mutations in Fcγ receptors (see below) have been shown to disrupt influence immune complex processing [34].

### 4. Tumor Necrosis Factor α

The tumor necrosis factor α (TNF-α) family of cytokines have been repeatedly associated with SLE, though the specific role of TNF is still debated. As discussed further below, the interleukin, IL-10 is known to suppress TNF-α and the theoretical relationship between the two is illustrated in Figure 1. TNF-α plays an important role in autoimmune regulation and is a primary mediator of the response to infectious organisms. TNF-α is primarily activated by mononuclear phagocytes but is also secreted by NK cells, T cells, and mast cells [35]. Although TNF-α plays a prominent role in fighting infection, excessive levels of this cytokine have been associated with a number of autoimmune diseases including Crohn’s disease, multiple sclerosis, and rheumatoid arthritis [36]. Treatment with anti-TNF medication has been proposed as a treatment for SLE [37] and is an active therapy for other autoimmune diseases, presumably by regulating the production of IFN-α by pDC [38].

#### 4.1. TNFSF4

TNF ligand superfamily 4 as well and its receptor, TNFRSF4, are primarily expressed on activated antigen-presenting cells [39] and activated CD4+ T cells [40], respectively. Graham et al. [41] used family-based and case-control approaches to identify a risk allele upstream of TNFS4 that predisposes to SLE and is correlated with increased TNFSF4 expression. The authors hypothesized that increased expression of TNFSF4 enhances interactions with either antigen-presenting cells or modulating T-cell activation. Recent GWAS [12, 24] have confirmed the association in European and Chinese samples. A recent study by Sanchez et al. [42] which targeted this and 15 other SLE-associated loci found that SLE renal disorder was significantly correlated with risk alleles in ITGAM and TNFSF4. Similar associations were found for ITGAM-disord rash, STAT4-protection from oral ulcers, and IL21- hematological disorder and suggest ways in which genetic profiling may be used to predict SLE etiology.

#### 4.2. TNFAIP3

Graham et al. [13] used whole-genome association to identify a significant relationship between tumor necrosis factor, alpha-induced protein 3 (TNFAIP3), and SLE, with the SNP rs5029939 the primary correlate. Subsequent GWAS [12, 24] replicated this association in Asian populations. A number of polymorphisms in TNFAIP3 have been associated with increased susceptibility to SLE [43, 44]. Recently, Adrianto et al. [45] identified a deletion (T) followed by a transversion (T to A) that resulted in reduced mRNA and A20 expression. By resequencing TNFAIP3 in affected European and Korean samples, the authors were able to compare nine risk variants at a signal locus. They found
that only one (the TT to A polymorphism) showed major overlap with regulatory elements, which was therefore established as the functional polymorphism driving the association between TNFAIP3 and SLE. This is a prime example of how GWAS and NGS approaches can be used in combination to gain a deeper perspective on pathogenesis and is discussed further below.

4.3. TNIP1. TNFAIP3-interacting protein 1 (TNIP1) is a closely related gene, also identified by GWAS as significantly associated with SLE [12, 46, 47] in both European and Chinese populations. Both TNFAIP3 and TNIP1 regulate the NF-κB signaling pathway, which is critical to the immune response and is activated by TNF-α as well as other diverse stimuli [48]. The mechanisms by which TNIP1 and TNFAIP3 regulate NF-κB signaling are not fully understood, though Heyninck and Beyaert [49] found that sequential deubiquitination and ubiquitination of the TNF receptor-interacting protein (RIP) may mediate this process. Gregersen and Olsson [50] discuss a common pathway linking TNFAIP3, TRAF1, TRAF2, CD40, and c-Rel in autoimmunity.

5. Interferon Cytokines

Interferons (IFNs) derive their name from their ability to “interfere” with the host cells of viruses during the infection
process and have been heavily associated with the pathogenesis of SLE. Type I interferons include IFN-α (produced by leukocytes) and IFN-β (produced by fibroblasts), both of which signal through the same receptor. IFN-α links the innate and adaptive immune systems and is therefore a critical component of the autoimmunity. A number of genes involved in IFN regulation have been associated with SLE, and individuals who have received treatment for other disorders using IFN-α supplementation have been known to develop SLE [51], which has also been found to resolve following α American (n pair insertion-deletion proximal to the 5′ gene in a European population. Similar to the approach taken with the IRF5 haplotype above, the Niewold group [66] examined the relationship between this locus, the presence of serum IFN-α, and the presence of autoantibodies. Again, IFN-α was only observed in the presence of specific autoantibodies. Furthermore, when the risk IRF5 and IRF7 genotypes were examined together, an additive effect was observed upon IFN-α that was not found in the autoantibody-negative group.

5.3. IRF8. The GWA replication study by Gateva et al. [24] highlighted a third member of the interferon regulatory factor family, IRF8, which is also an associated risk factor for SLE. A different variant at the same locus has also been associated with multiple sclerosis [67], adding further weight to this region as a potentially important region in relation to autoimmune susceptibility. Hikami et al. [68] found that a functional polymorphism in the 3′ untranslated region of SPI1, known to regulate expression of IRF2, IRF4, and IRF8 [69, 70] is associated with increased risk of SLE.

5.4. STAT4. Kariuki et al. [71] identified an association between a signal transducer and activator of transcription 4 (STAT4) risk allele (the T allele at rs7574865), which is associated with lower serum IFN-α activity and increased sensitivity to IFN-α signaling. Remmers et al. [72] reported an association between signal transducer and activator of transcription 4 (STAT4) and SLE (as well as rheumatoid arthritis). Again, the association with SLE has since been replicated in a number of GWAS in European and Asian populations [12–14, 24]. It represents one of six primary members of the STAT family, all of which play important roles in cytokine signaling. STAT4 is integral to IL-12 signaling in both T and Nk Cells and increases the production of IFN-γ and differentiation of CD4 T Cells [73]. After it binds to the IL-12 receptor, phosphorylated STAT4 forms homodimers, which translocate to the nucleus and initiate transcription of genetic targets, IFN-γ is amongst them [74]. The SNP rs7574865 in the third intron of STAT4 that is the most strongly associated SNP and is also associated with an alternate SLE phenotype that is more severe has a younger age of onset (<30 years) and is characterized by characterized by double-stranded DNA autoantibodies [75]. This SNP also correlates with increased sensitivity to IFN-α [71]. The GWA study by Chung et al. [9] confirms an association between rs7574865 and anti-dsDNA-positive autoantibody production in SLE (P = 2 × 10−20). Indeed associations between STAT4 and this phenotype were found to be stronger than with SLE per se. As such, the authors suggest that STAT4 be considered an autoantibody-propensity locus as opposed to and SLE-susceptibility locus. This claim is also made for the ITGAM and HLA-DR3 loci.

Namjou et al. [76] identified three major STAT4 haplotypes that were highly significant in Europeans and moderately significant in Korean and Hispanic samples. Interestingly, Sigurdsson et al. [76] report that the STAT4 SNP, rs7582694, correlates with the production of anti-double-stranded DNA antibodies and have a multiplicative risk effect of 1.82 with two independent IRF5 risk alleles. This provides
a strong indication that interactions between the two genes contribute to the pathogenesis of SLE. However, a study using 30 tagged SNPs by Abelson et al. [77] revealed no significant interaction effects between SNPs in STAT4 and IRF5 and suggests that an additive model may most closely describe their combined contribution to SLE risk.

6. Interleukins

As a major component of the immune system, interleukins are strongly linked to the pathophysiology of SLE, and genes that encode interleukin proteins have been widely examined as possible SLE susceptibility candidates. Interleukin (IL)-10, which is an important immunoregulator that inhibits T-cell function and suppresses proinflammatory cytokines such as TNFα, IL-1, IL-6, IL-8, and IL-12 [78, 79] is particularly important in this regard.

6.1. IL10. A GWA replication study by Gateva et al. [24] confirmed an association between a SNP (rs3024505) on IL10 and SLE in individuals of European ancestry (P = 3.95 × 10−8). This followup on a number of smaller scale studies that have previously identified an association between SLE and polymorphisms in IL10 [80, 81] (n = 58 and 158 resp.). It is also consistent with a large-scale association study in an Asian sample (n = 554) [82], which used six IL-10 promoter SNPs to identify six haplotypes. The authors found an association between one of the haplotypes and decreased IL-10 production and also observed a dose-dependent effect of the microsatellite IL10. G as a significant risk factor. IL10 is also known to promote B-cell functions by facilitating antibody production, differentiation, and proliferation [83]. Increased production of IL-10 by peripheral B cells has been shown to correlate with SLE severity [84]. Relevant polymorphisms include the SNP rs1800896, which is characterized by a glycine/alanine substitution at position −1082, rs1800871 (−819C/T), and rs1800872 (−592C/A) [4], all of which have been shown to affect IL-10 production [85–87]. Summers et al. [88] showed that the ATA haplotype, specifically the −592A allele, was more frequent in 23 cases of sudden infant death syndrome, suggesting that irregular lower/higher IL-10 production may affect production of inflammatory cytokines.

A number of other genes that encode interleukin proteins have also been associated with SLE. IL18 that has long been touted as an SLE candidate, though a recent dense mapping of the locus failed to find any evidence of a common variant association [89]. A recent CNV study of 938 SLE patients and 1,017 healthy controls by [90] found a higher proportion of copy number amplifications of IL17F, IL21, and IL22 in cases (see below). Elevated expression of IL21 (4q26-q27), which shows homology with genes that encode IL2, IL4, and IL15, has been demonstrated in the sera of SLE cases and mouse models [91]. Further work is needed to clarify the roles of these and other interleukin coding genes in SLE. A thorough review of SLE/interleukins is provided by López et al. [4].

6.2. IRAK1/MECP2. IL-1 receptor-associated kinase 1 (IRAK1) is part of the Toll/interleukin-1 receptor and nuclear NFκB signaling pathway and has been associated with both adult- and pediatric-onset SLE [92]. However, the neighboring locus, methyl-CpG-binding protein 2 (MECP2), which is in strong linkage disequilibrium with IRAK1 and is involved in regulating methylation-sensitive loci [93], has also been proposed as the source of this signal [94], though the two are not necessarily mutually exclusive. GWA by Graham et al. [13] identified the IRAK1-MECP2 locus as significantly associated with SLE risk in Europeans, replicating a previous finding by Sawalha et al. [95].

7. B-Cell Signaling

Although B cells have been strongly associated with SLE, a direct relationship between candidate genes and susceptibility has been difficult to determine. However, GWAS have facilitated the identification of a number of variants involved in B-cell signaling that may predispose to SLE. The most prominent is B lymphocyte kinase (BLK), which is located at 8p23.1. GWAS by Han et al. [12], Harley et al. [14], and a replication study by Gateva et al. [24] identified this locus as strongly associated with SLE risk in European and Asian populations. BLK is a tyrosine kinase that transduces signals downstream of the B-cell receptor and can phosphorylate inhibitory Fc receptors on B cells [52]. A meta-analysis by Fan et al. [96] utilizing 11,000+ cases and 20,000+ controls from European and Asian populations examined the risk alleles of rs13277113 and rs2248932. For the former, BLK mRNA expression was ∼50% lower in A homozygotes than G allele homozygotes. Similarly, at rs2248932, the C allele was associated with lower levels of BLK mRNA, with C homozygotes having ∼30% lower expression levels than T allele homozygotes. Both the A allele of rs13277113 and the T allele of rs2248932 were confirmed as significant risk factors for SLE. Similar to findings in relation to the MHC, STAT4, ITGAM, and IRF5, the GWA study of anti-dsDNA autoantibody production by Chung et al. [9] found evidence of an association between this phenotype and BLK but did not meet threshold criteria for genome-wide significance.

7.1. BANK1. A GWA study by Kozyrev et al. [16] identified a nonsynonymous SNP among a Swedish population in the B-cell scaffold protein with ankyrin repeats 1 (BANK1) gene at rs10516487 (R61H), which was replicated in four independent case-control sets. A study from our group confirmed this association in both European (n = 178) and African American (n = 148) populations [97]. Although a number of GWAS have not found a significant association with this locus, a large study of 1,724 SLE patients and 2,024 healthy controls of African American (n = 148) populations [97]. Although a number of GWAS have not found a significant association with this locus, a large study of 1,724 SLE patients and 2,024 healthy controls of African American descent did replicate BANK1 as an SLE candidate in this cohort (as well as C8orf13-BLK, TNFSF4, KIAA1542, and CTLA4) [98]. Kozyrev et al. [16] identified two other nonsynonymous SNPs at rs17266594 (intronic) and rs3733197 (A383T), which have yet to be replicated as confirmed SLE risk factors. BANK1 has also been associated anti-dsDNA-positive but not anti-dsDNA-negative autoantibody production in SLE [9].
8. T-Cell Signaling: Candidate Genes

Cytokines are usually grouped in accordance with their functional capacity as T helper (Th) Th1, Th2, and Th17. Overproduction of Th1 and Th17 most often results in T-cell hyperactivity, whereas the overproduction of Th2 is linked to hyperactive B cells and humoral responses [35].

8.1. PTPN22. Protein phosphatase nonreceptor type 22 (PTPN22) is known to inhibit T-cell activation [99]. It has been associated with a wide range of autoimmune disorders and initially came to prominence as a candidate gene for type 1 diabetes [100]. Indeed, as Gregersen and Olsson [50] point out, the association between PTPN22 and a wide range of phenotypes (including Graves’ disease [101–103], Hashimoto thyroiditis [104], myasthenia gravis [105], systemic sclerosis [106], generalized vitiligo [107], Addison’s disease [108], alopecia areata [109], juvenile idiopathic arthritis [110–112], and SLE [14, 15]) provided one of the earliest indications of a shared pathophysiology for many autoimmune disorders. In spite of this, however, it is notable that PTPN22 risk alleles are not a universal feature of all autoimmune diseases, and indeed the 1858 T autoimmune risk allele is associated with protection against Crohn’s disease [113]. Kariuki and Niewold [114] point out that this same risk allele is associated with TNF-α-related diseases but not with multiple sclerosis (treated with a related interferon, IFN-β). This suggests a possible relationship between PTPN22 and cytokine profile, albeit at a secondary level. The rs2476601 SNP polymorphism (cysteine to threonine, C1858T) in the LyP protein is significantly associated with the underactivation of both T and B cells and deregulated cytokine production [115, 116].

GWAS by Harley et al. [14] and replication by Gateva et al. [24] identified a positive association between the PTPN22 SNP, rs2476601, and SLE in European populations, but a similar association has not been observed in the Asian GWAS reviewed here. Deng and Tsao [117] point out that this difference may be attributable to more variability in European populations that may be between 2–15% [50]. PTPN22 has also been associated with anti-dsDNA-positive but not anti-dsDNA-negative autoantibody production in SLE [9].

8.2. PPP2R2B. Although not specifically identified by the GWAS listed in Table 1, protein phosphatase 2, regulatory subunit Bβ (PPP2R2B) may be important to SLE pathogenesis. A recent study by Crispin et al. [118] found that the regulatory Bβ that is expressed in resting human T cells is downregulated during T-cell activation and is upregulated by interleukin-2 (IL-2). In a study of SLE patients, the group found that levels of PP2A Bβ were not increased by IL-2 deprivation in 7 of the 14 cases, and this phenomenon was paralleled by resistance to apoptosis. Furthermore, following IL-2 withdrawal in T cells, levels of Bβ in these patients remained unchanged. This contrasts markedly with responses from healthy controls as well as the remaining SLE (nonapoptosis-resistant) patients, who demonstrated an approximate threefold increase in Bβ. It would therefore appear that, at least in a subset of SLE patients, Bβ may be a primary cause of apoptosis resistance in T cells.

9. Fcγ Receptors

The fragment crystallizable (Fc) region is found at the tail of antibody proteins. Fc receptors are involved in clearing the immune complex and include Fcα, Fcε, and Fcγ. The latter has been most closely linked to the SLE susceptibility, with a number of low-affinity Fc receptors for immunoglobulin G (IgG) identified as SLE candidate alleles. These include FCγR2A, FCγR2B, FCγR3A, and FCγR3B, all of which are regulated by cytokines, including IL-4, IL-10, or TGF-β [119]. Cytokine-mediated regulation of FcR expression is cell type-specific; however, IL-4 upregulates FCγR2B expression in myeloid cells but downregulates it in activated B cells [120]. One should bear in mind that, particularly with earlier studies, low coverage of Fcγ receptors by commercial manufacturers (i.e., Affymetrix and Illumina) has been an issue. This is mainly due to the presence of homologous sequence paralogs at this locus [121].

9.1. FCγR2A. The nonsynonymous SNP, rs1801274 in the FCγR2A (Fc fragment of IgG, low affinity IIa, receptor) gene has been associated with reduced clearance of immune complexes, where the (C) allele encodes arginine and the (T) allele encodes the variant histidine (H) [122]. Karassa et al. [123] conducted a meta-analysis of this polymorphism across European, African, and Asian populations and confirmed a positive risk to SLE. In general, however, the relationship has been inconsistent, and a majority of GWAS have not found significant evidence of an association. The exception is the 2008 GWA by Harley et al. [14] which identified an association between rs1801274 but only among European women.

9.2. FCγR2B. FCγR2B is involved in antibody production and macrophage activation. Tsujiya et al. [124] examined a nonsynonymous SNP within the transmembrane domain, which was significantly associated with SLE in Chinese, Japanese, and Thai populations but was found to be rare in Europeans. The SNP (rs1050501) led to an aminoacid substitution within the transmembrane domain at position 232, isoleucine to tryptophan (I232T). The same group found that an FCγR2B promoter polymorphism that has been associated with SLE in Europeans [125] was largely absent in Asians. In a human B-cell line lacking endogenous FCγR2B, Kono et al. [126] demonstrated that I232T was less effective than wild-type at inhibiting B-cell receptor (BCR-) mediated signaling, and that distribution of FCγR2B to detergent-insoluble lipid rafts was disrupted. This is supported by Floto et al. [127], who showed that the lack of inhibition of activatory receptors resulted in unopposed proinflammatory signaling.

9.3. FCγR3A. Nonsynonymous SNPs in FCγR3A have been found to alter binding affinities in the four immunoglobulin
G subclasses. At the SNP rs396991, the (T) allele encodes phenylalanine (F), and the (G) allele encodes the valine variant (F158V) [117]. The low-affinity phenylalanine allele is associated with disrupted immune complex clearance [128], while the high-affinity valine allele was a strong predictor of end-stage renal disease [129]. A copy number variation study by Niederer et al. [130] found significant differences in FCyR3A and FCyR3B CNV profiles between European, East Asian, and Kenyan populations. Reduced copy number and homozygosity of FCyR3B were strongly predictors of SLE susceptibility, but the same association was not reported in relation to FCyR3A.

9.4. FCyR3B. Three allotypic variants have been identified for FCyR3B—NA1, NA2, and SH and are defined by six SNPs. Hatta et al. [131] report an association between NA2 and SLE, but this has not been replicated (see Yuan et al. [132]). Nevertheless, a number of copy number variation studies have converged upon the FCyR3B gene as a potentially important locus for SLE susceptibility. These are reviewed separately below.

10. Interpreting SLE GWAS

All of the discovery SNPs used to identify the genes listed in Table 1 have published odds ratios (ORs) between 1.2 and 2.4, with the majority ranking toward the lower end of this range. This is comparable with GWAS in other autoimmune disorders including Crohn’s disease [133], rheumatoid arthritis [134], and psoriasis [135] and indeed the GWAS field as a whole. While it is important not to downplay their significance, it should be noted that the predictive value of such ratios is relatively low [136] and explains less than 15% [137] of the risk for SLE (as would be expected by GWAS variants all of which are designed to tag LD blocks showing association). We know from twin and family studies that the heritability of SLE is approximately 44–69% [138–140], which means we must consider the problem of missing heritability. There are a number of possible explanations between the discordance between the two figures: (1) SLE is underscored by an even larger number of genes each contributing smaller and smaller proportions of risk variance, (2) that the variants identified by GWAS lose significant power in the process of tagging the causative variants, and/or (3) that the broader SLE phenotype may consist of a several distinct and rare subtypes. It is also likely that epigenetic factors are important elements of the missing heritability, though this is not explored in the current paper.

(1) Gene Networks and Pathway Analysis. The first of these conclusions, that an increasingly larger pool of genes is required to account for the heritability of SLE, is logical but complex. The common disease common variant model is predicated upon the conclusion that the complex diseases are caused by the interactions of a large network of genes, and the number of possible causal loci is only constrained by the number of genes and gene regulators in the human genome. However, as more and more genes are implicated in this network, we run into the law of diminishing returns. Pathway analysis, which leverages existing biological knowledge about gene function to examine how causal factors may interact, is an attractive mechanism for navigating this law. Pathway-based approaches typically examine whether test statistics for a predefined gene-set have concordant (albeit moderate) deviation from chance. Analyses are based on pathway association approaches in gene expression microarray analysis, where examination of groups of related genes has yielded major insights into functional capacity [141, 142].

A recent study from our group [5] adopted the pathways approach to GWAS of Crohn’s disease (CD), which is known to share certain pathophysiological properties with SLE [37]. The study examined enrichment of association signals for genes previously identified as belonging to certain gene pathway networks, as defined by gene ontology, biocarta and KEGG, with careful adjustment for gene size, number of SNPs per each genes, and pathway size. The pathway that most significantly enriched for association signals was the interleukin-12 gene pathway that harbors the cytokines, interleukin 12 and 23 (IL-12/IL-23). These share one cellular receptor subunit and numerous intracellular signaling components previously shown to associate with CD [113], together with multiple other genes associating with CD, for the first time [5]. Indeed, only three genes (IL12B, IL23R, and IL1ZRB2) at two loci (5q23, 1p31) showed genome-wide signals in previous studies [113]. However, three further genes in the IL-12-IL-23 pathway (JAK2, CCR6, and STAT3) were confirmed as candidate genes in replication studies [109], and six more genes (including STAT4 and a number of interleukin-coding genes) supported by association in this pathway have been previously reported as CD susceptibility genes in other association and functional studies [5, 142–147]. Thus, since only three genes in this pathway have surpassed the threshold for genome-wide significance, we begin to develop a much richer picture of the pathophysiology of the disease through the pathway-based approach. This includes related variants that have remained above significance criteria but collectively contribute significantly to the risk variance. Furthermore, given that the strongest candidate gene is sometimes not the most suitable drug target, the pathway approach also opens up possible alternatives for targeted intervention. These interactions are outlined in more detail in Figure 2. Given the shared etiology between many of the autoimmune diseases, it is likely that a similar approach would be productive in extrapolating gene networks in SLE.

(2) Causative Variants and the GWAS Signal. A study from our group substantiates the second conclusion, that at least some of the variants tagged by GWAS may actually be tagging rarer variants that may be responsible for the GWA signal. This phenomenon is known as synthetic association and was confirmed in a recent study by Wang et al. [148], who used sequencing to examine NOD2 as a candidate gene for CD. Three rare variants (nonsynonymous SNPs) in NOD2, rs2066844, rs2066845, and rs2066847, had previously been associated with susceptibility to CD [149, 150], which had additionally been identified by functional assays as
IL-12 heterodimer:
IL-12A (P = 0.04)
IL-12B (P = 8.8e\(-9\))

IL-12RB2
(P = 2.8e\(-17\))

IL-12RB1
(P = 7.7e\(-3\))

JAK2 (P = 6.8e\(-7\))
TYK2 (P = 5.7e\(-3\))
STAT1 (P = 0.17)
STAT4 (P = 0.038)

STAT1 (P = 0.17)
STAT4 (P = 0.038)
TBX21 (P = 0.28)

IFN-γ (P = 0.014)

IL-18R1
(P = 1.2e\(-4\))

IL-18RAP
(P = 2.0e\(-5\))

Cytokine-mediated gut destruction

Immunosuppression

Th1 commitment
Th17 suppression

Th1 T cell
Th17 T cell

STAT3 (P = 3.4e\(-5\))
RO Ryt (P = 0.024)
RO Rd (P = 4.7e\(-4\))

IL-17 (P = 0.012)
IL-17F (P = 0.18)
IL-21 (P = 2.2e\(-4\))
IL-22 (P = 0.17)
IL-26 (P = 0.024)

TNF-α (P = 6.1e\(-3\))

TGF-β

Figure 2: Linking pathways to disease: Crohn’s disease (from Wang et al. [5]). As an example of how biological pathways are involved in disease pathogenesis, the authors illustrate a manually compiled pathway centered on IL-12 and IL-23. For each gene, the most significant P value among SNPs close to the gene (based on GWAS) was annotated. Only three genes at two loci (IL12B at 5q33 and IL23R-IL12RB2 at 1p31) showed genome-wide significant signals, but three further genes (JAK2, CCR6, and STAT3) in the pathway were confirmed as susceptibility candidates in replication studies. Six further genes in this pathway were reported as Crohn’s disease susceptibility genes by other association and functional studies.

potentially causal [151]. Although NOD2 had not previously been shown to harbor common causal variants, a 2007 study by the Wellcome Trust Case Control Consortium (WTCCC) [152] did implicate a common tag SNP (rs17221417, MAF = 29). From HapMap, we know that the first two of the rare NOD2 variants are in complete linkage disequilibrium with rs17221417 (the third variant was not listed). A large-scale meta-analysis of NOD2 estimates the allelic odds ratio (OR) at between 2.2 to 4.1, which, respectively, explain 0.54%, 1.2%, and 3.4% of genetic risk (5.1% in total). For the common tag SNP, on the other hand, the OR is 1.37 and explains on 0.69% of genetic risk for CD. In other words, the GWA signal dramatically underestimates the proportion of explained risk at this locus.

A long range haplotype analysis of the GWA data for all the genes identified above is therefore recommended. Our group previously showed that synthetic associations can cover intervals as long as 2.5-Mb, and include numerous “blocks” of associated variants [136]. As such, followup and interpretation of GWA data merit careful deliberation. Namely, this approach may enrich sample sets for individuals with rare-causative variants, who should be filtered out from the cohort and subsequently sequenced for confirmation

[148]. Anderson et al. [153] make the point that certain GWAS signals are more amenable to the synthetic association effect than others. This includes signals that are found inconsistently between different populations; signals that are not universally common are more likely to have arisen recently or been unequally selected for through population history. A number of the SLE candidate genes reviewed above fall into this category, including ETS1, ITGAM, and BANK1.

(3) Rare Variants. Another source of hidden variance is the existence of rare phenotypes subsumed under the broader SLE umbrella but with a separable genetic basis. The heterogeneity of the SLE phenotype lends itself to this possibility, which is also supported by a number of genome sequencing studies that have emerged in recent years. As mentioned above, complement defects are present in 1-2% of SLE cases and represent the strongest single genetic risk factors of SLE. Additional rare variants associated with SLE include mutations in the three prime repair exonuclease 1 (TREX1) gene, which is the most common known cause of monogenic (i.e., single gene defect) SLE [154, 155] and may be an instructive vehicle by which to explore the application of
sequencing technologies to the disease. Located on chromosome 3p21, TREX1 can metabolize reverse-transcribed DNA and encodes 3′ repair exonuclease 1, the primary 3′ to 5′ exonuclease in humans. It is a regulator of the IFN-stimulatory DNA (ISD) response [156], and mutations of TREX1 can cause Aicardi-Goutières syndrome (AGS) [157], a severe neurological brain disease that is sometimes comorbid with pediatric SLE [158–161]. It is also linked to several other diseases accelerated atherosclerosis, antiphospholipid syndrome, and fetal loss [162].

TREX1 mutations in AGS are predominantly recessive and reduce exonuclease activity. This is the case with a non-synonymous SNP at position 114 that results in an arginine to histidine substitution (R114H) [163, 164]. A large-scale genotyping study of 40 TREX1 SNPs in ~8370 SLE patients and ~7490 controls by Namjou et al. [165] identified nine European patients with heterozygous mutations at this locus, which was also found in five European controls. Among Africans, five cases had a mutation at E266G, but none were found in controls. Moreover, one Asian case, a male with early-onset SLE, was found to have a homozygous R114H mutation. The patient, who also had positive anti-dsDNA antibody, was negative for neurological manifestations, which are found in AGS children with the R114H mutation. A number of other mutations were observed in SLE patients but not respective controls. Among Europeans cases, five heterozygous mutations were observed in SLE patients but not respective controls. Among Europeans cases, five heterozygous mutations were detected at Y305C but none in controls. Y305C is a missense coding mutation located outside of the catalytic domain. Among Africans, five cases had a mutation at E266G, but none were found in controls. Interestingly, this mutation was also present in Europeans but did not differentiate cases from controls. For the SLE group as a whole, the coding mutation frequency was approximately 0.5%.

Importantly, a case-control association study for common SNPs did not identify significant associations between the affected and the unaffected either for the group as a whole, or for respective racial groups. This underscores the point that the association approach is limited in terms of probing rare variants. However, as discussed above in relation to CD, GWA signals at the site of a common tag SNP can prime the locus for followup by sequencing and genotyping. Thus, in the Namjou et al. study, common SNPs (defined by a minor allele frequency of 10% or greater) in Europeans characterized a seizure-associated risk haplotype that was present in 58% of cases, compared to 45% in controls (P = 0.0008, OR = 1.73, 95% CI = 1.25–2.39).

Although untested, one may speculate that many of the loci identified by GWAS may similarly harbor rare variants. Monogenic subtypes of cutaneous lupus erythematosus (a rare cutaneous form of lupus erythematosus) have been observed at the TREX1 locus [166]. Similarly, Qari et al. [167] describe seven Saudi families, where pedigree data is consistent with autosomal recessive Mendelian inheritance. In the last two years, the application of NGS to monogenic and oligogenic disorders has accelerated rapidly, reflecting the increased availability of relevant technologies. In our own group, we have used NGS to identify the causal variants in a number of Mendelian phenotypes including rare forms of Glycogen Storage Disease, familial forms of epilepsy, hemolytic anemia [168], and Ogden syndrome [169].

It remains to be seen what proportion of rare variants will account for hidden variance in SLE. Regardless, their study will continue to play an important role in explicating the pathogenesis of the disorder. Because rare mutations carry large effects, they make resolution of underlying networks distinctly less complex and are also amenable to modeling in other systems. The widespread application of sequencing technologies in the clinic will also help characterize and differentiate SLE subphenotypes. In this vein, De Vries et al. [170] used direct sequencing of exonic TREX1 in 60 patients with neuropsychiatric SLE and identified a novel heterozygous substitution at position 128 (arginine histidine) in one case. Because TREX1 mutations are also linked to AGS, which is characterized by a number of neurological abnormalities, it is possible that this form of SLE shares a common pathogenic mechanism.

11. SLE and Copy Number Variants

The role that rare variants in SLE has is also becoming apparent from a series of copy number variation (CNV) studies, in the past several years. CNVs are insertions, deletions, or inversions in the genome that are universal in the general population and vary in length from many megabases to 1 kilobase or smaller. Although CNVs per se are not associated with any observable phenotype, their presence in genic regions has been associated with a number of major diseases, including autism [171], schizophrenia [172, 173], neuroblastoma [174], and many others. Arguably the most widely known CNV occurs in down syndrome, which is characterized by an extra copy of chromosome 21. The origin of most CNVs is unknown, but causal mechanisms can include replication errors, meiotic recombination, and homologous/nonhomologous repair of double-strand breaks [175].

Recent studies by the 1000 Genomes Consortium [176] and Conrad et al. [177] report that common CNVs are well covered by SNPs in existing arrays and many likely have been indirectly examined in a range of GWAS. The impact of rare CNVs, on the other hand, may be substantial. Pang et al. [178] reexamined data from the Venter genome and identified over 12,000 structural variants spanning more than 40 Mb of sequence. These variants were found in 4,867 genes, which are often large and under negative selection. Because rarer alleles are more likely to have a higher penetrance, these results strongly support the role of CNVs as causal factors in genetic diseases. The study also showed that 24% of CNVs would not be imputed from SNP association alone, which stresses the point that CNVs may be more accurately detected using the NGS approach.

A number of CNV studies have highlighted FcγR3B as an important locus for CNV differences between SLE cases and controls. As outlined above, Fcγ receptors for immunoglobulin G (IgC) are involved in clearing the immune complex, and many members of the Fcγ have been proposed and replicated as SLE candidates. As far back as 2006, Aitman et al. [179] reported an association between low copy number
in FCyR3B and autoimmune glomerulonephritis in a subsample of 30 individuals with SLE. The same group [180] replicated this association in 161 Europeans with SLE-associated glomerulonephritis (control n = 312), as well as an Afro-Caribbean cohort of 134 patients [181]. This second study also examined the broader autoimmune disorder phenotype (n = 1.279) and found that 25 (2%) have no copies of FCyR3B. The same analysis of 862 controls indentified only one individual with no FCyR3B copies. Interestingly, the group did not observe an association between Graves’ or Addison’s diseases, both of which are organ-specific autoimmune disorders.

Niederer et al. [130] also described a strong link between FCyR3B copy number and identified a linkage disequilibrium between FCyR3B and an FCyR2B variant (I232T). Morris et al. [182] confirmed the association between FCyR3B low copy number, which was contingent upon allele; SLE risk was greater for NA1 deletion, than deletion of NA2. The authors also reported a significant correlation between FCyR3B copy number and neutrophil expression in healthy and control participants. FCyR3B is expressed by neutrophils and eosinophils. Neutrophils, which mediate inflammation responses to host injury [183] are attracted by cytokines at the early stages of infection and also release cytokines as part of the inflammatory response [184]. Eosinophils, also an important part of the inflammatory response, are activated by several cytokines, namely, IL-3, IL-5, and GM-CSF [185].

CNVs have been identified at other candidate SLE loci. In a study of 532 Asian patients with SLE and 576 controls, Yu et al. [186] identified CNVs at IL-12B and T-bet as significantly associated with SLE risk. For IL-12B, the frequency of copy number amplification was 63 versus 13 in controls. For T-bet, the respective values were 46 versus 7. The same group [187] identified CNVs at histamine H4 receptor in a cohort of 340 SLE patients (versus 392 controls). These correlated with the presence of antinuclear antibody abnormalities, as well as incidence of arthritis and proteinuria. The group also identified CNV enrichment in interleukins-17F, -21, and -22, in two Chinese SLE cohorts (Yu et al. [90]). Allele frequencies amplifications for cases versus controls were 107 versus 33, 166 versus 16, and 108 versus 19, respectively. Other studies have implicated copy number differences at the TLR7 locus in women with SLE [188] and at the HIN200 locus in UK SLE families and French males [189]. Although the majority of these CNVs are found in only a handful of cases, these results have broad implications for SLE as a whole and may collectively account for a relative large proportion of SLE cases.

12. Conclusions

Taken together, these studies show that SLE is highly heritable, and advances in gene-finding technology in the past decade have rapidly accelerated gene discovery. Over this period, a number of themes have begun to emerge: (1) an ever-increasing catalog of candidate genes that replicate across different studies, (2) a growing list of causal rare variants, and (3) the emergence of monogenic subtypes. Monogenic SLE is particularly interesting from a treatment perspective, as it provides a mechanism for studying the phenotype in model systems and is a more obvious target for drug intervention. In order to capitalize upon these findings, high-quality phenotype data is required. SLE is notoriously heterogeneous and is fractionated in terms of onset, symptoms, and trajectory. The systematic collection of clinical and biomedical data (e.g., cellular, serological, and mRNA transcripts) will dramatically increase our ability to generate testable hypothesis about the contributions of specific genes and gene networks to SLE pathophysiology. Conversely, knowledge of gene function can be used to target treatments and to predict onset.

Ultimately, the primary goal is not to determine the frequency of variation/mutation in cases versus controls but to determine the pathways that lead to disease pathology. This is no simple task, especially when we consider the other major risk factors such as epigenetics, RNA regulatory elements, and environmental exposures. While daunting, the elucidation of these elements will doubtlessly take us closer to developing more effective treatments for SLE targeting selective patients for interventions aimed at restoring the impact from variants within specific molecular pathways and gene networks.

References


24. V. Gateva, J. K. Sandling, G. Hom et al., "A large-scale replication study identifies tnpla1, pmd1, jazf1, uhf1b1p and il10 as risk loci for systemic lupus erythematosus," *Nature Genetics*, vol. 41, no. 11, pp. 1228–1233, 2009.


G. S. LaBerge, D. C. Bennett, P. R. Fain, and R. A. Spritz, “PTPN22 is genetically associated with risk of generalized vitiligo, but CTLA4 is not,” *Journal of Investigative Dermatology*, vol. 128, no. 7, pp. 1757–1762, 2008.


M. F. Seldin, R. Shigeta, K. Laiho et al., “Finnish case-control and family studies support ptpn22 620w polymorphism as a risk factor in rheumatoid arthritis, but suggest only minimal or no effect in juvenile idiopathic arthritis,” *Genes and Immunity*, vol. 6, no. 8, pp. 720–722, 2005.


Submit your manuscripts at http://www.hindawi.com