Systemic lupus erythematosus (SLE) is an autoimmune disease notable for its complex and varied presentation and its predilection for affecting women in their childbearing years. It is believed that a number of genetic risk factors combine to create susceptibility, and then environmental triggers, such as the Epstein–Barr virus [1], impact the genetically primed host to start the autoimmune cascade. There is the powerful and consistent observation that lupus occurs in families, such that up to 10% of SLE patients have a relative with lupus [2]. Increased concordance in monozygotic twins was observed over 30 years ago [3], and more current twin statistics continue to support this observation, with over 40% concordance in monozygotic twin pairs versus 4% concordance in dizygotic twins [4]. Except in the rare cases of complement deficiency [5], the inheritance pattern of SLE does not follow simple Mendelian rules, which suggests that genetic risk in most lupus patients arises from the combination of a number of relatively common variations in several different genes. The general consensus of those in the field is that the number of ‘lupus genes’ is likely to be in the order of 20–50. The search for the genes that cause SLE has been a subject of active investigation for nearly 20 years by dozens of groups all over the world, generating thousands of papers, many of which have conflicting results. In this review, I will attempt to summarize this body of work.
linkage region, although there are far more candidate genes chosen on the basis of their function than on their location within a linkage region.

A new approach called association scanning is being used by a few studies now in progress. This type of study would not be possible without the recent advances in technology. Nevertheless, both Affymetrix and Illumina have created products that test over 100,000 SNPs distributed across the entire genome. Although the SNP markers still have the limitation that they are never more than 50% heterozygous, and are therefore not as informative as microsatellite markers, in combination they form informative haplotypes that can be used to trace ancestry. Genome scans can be performed on case–control, trio or pedigree collections. Like the linkage studies that precede them, an association scan is hypothesis generating, but the associated regions found will already be fine mapped, in a sense, because the markers used are so densely arrayed.

All three of these study designs share a common problem: they are prone to the production of false-positive results. Therefore, it is important that each finding be replicated in an independent cohort, and the scientific community as a whole recognizes the importance of such replication work. Unfortunately, every cohort is slightly different - they vary in size, ethnicity, selection and matching criteria, and likely a whole host of other variables that may not even be collected or recognized as contributing to this complex disease. Some cohorts are underpowered for the effect they attempt to define, and this can also lead to false negatives. In addition, each investigator has the choice of a large number of potential variants to test within each gene, and often different studies of the same gene have few or no SNPs in common. In part because of this variation, the resulting literature in the field is full of conflicting reports. For example, while there are 16 reports supporting an association with mannose-binding lectin, there are eight that find no association and 14 that find association only with a specific phenotype, such as lupus nephritis. When enough reports have been accumulated, meta-analysis can assist us in determining the true nature of the association, and a meta-analysis of mannose-binding lectin shows that there is a consistent, albeit weak, association with SLE [18]. Unfortunately, 32 of the 48 genes with conflicting reports have less than ten total reports in the literature as of October 2007, such that meta-analysis is impractical. In addition, there are unconfirmed associations, negative reports and associations with phenotype only that remain to be verified. Even among the confirmed associations summarized below, eight of the 20 genes have only been tested in two cohorts. A summary of the nature of the candidate genes in SLE by the types of literature they have generated is given in Table 3.

### Established genes

**Complement**

Work on complement began before the genotyping era, and reports of association with complement deficiency may or may not include genetic information. Nevertheless, it has been established that, in rare cases, complete deficiency of the elements of the classical pathway - C2, C4 and C1q - leads to SLE or lupus-like syndromes [5]. Immune complexes activate complement through these components, which are also important for keeping immune complexes in soluble form and clearing apoptotic bodies [19]. The genetic aspects of C4 deficiency are complicated by both the complex structure of the gene and its location within the HLA region. It has recently been discovered that individuals may carry anywhere between zero and six copies of the C4 cassette, although most carry four, and that decreased copy number is associated with SLE risk [20]. A partial deficiency, also known as C4A*Q0 or C4 null, is also in linkage disequilibrium with

### Table 1. Types of studies used in evaluating genetic risk factors for systemic lupus erythematosus.

<table>
<thead>
<tr>
<th></th>
<th>Traditional linkage study</th>
<th>Traditional association study</th>
<th>Association scan study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort type</td>
<td>Multiplex families</td>
<td>Case-control or trio</td>
<td>Any</td>
</tr>
<tr>
<td>Marker type</td>
<td>Microsatellite</td>
<td>SNP</td>
<td>SNP</td>
</tr>
<tr>
<td>No. markers tested</td>
<td>300–400</td>
<td>One or more</td>
<td>100,000+</td>
</tr>
<tr>
<td>Findings</td>
<td>Linkage</td>
<td>Association</td>
<td>Association</td>
</tr>
<tr>
<td>No. regions</td>
<td>10–15, usually</td>
<td>One or more</td>
<td>100+</td>
</tr>
<tr>
<td>Size of region</td>
<td>20–40 mb</td>
<td>5–10 kb</td>
<td>5–10 kb</td>
</tr>
<tr>
<td>No. using the approach</td>
<td>Five major groups</td>
<td>&gt;50 groups</td>
<td>A few in progress</td>
</tr>
</tbody>
</table>
HLA-DR3 [21], and HLA-DR3 is strongly associated both with autoantibody profile and SLE on its own [22].

HLA region

Although work on the HLA region pre-dates genotyping and linkage in this region is well established, there remains much to be done to define the nature of genetic alteration in the region and its role in autoimmune pathology. First, there are a number of genes with immune functions in tightly linked regions, including not only HLA class I, II and III genes, but also genes encoding complement components C2 and C4, TAP 1 and 2, and TNF-α and -β. Since these genes are so close together, they are often inherited as a unit, a phenomenon known as linkage disequilibrium (LD), and this creates confusion as to which variation within a risk haplotype is truly responsible for disease. For example, the TNF-α -308 variant, which is associated with overexpression, is often found in a haplotype block that includes HLA-B8, C4A*Q0 and HLA-DR3. As such, it has been variably claimed that each of these variants is the ‘real’ cause of increased risk, and unfortunately, there are few studies that attempt to type them all, leading to competing claims rather than clarification work. To add to the confusion, both HLA-DR3 and HLA-DR2 have been found to be associated with disease, and the dual association cannot be explained by ethnic background alone [23]. It is possible that only one of these genes is important in disease, but it is also possible that some combination or combinations of genetic variants in this region is necessary to generate an autoimmune predisposition. Additional work on larger cohorts in which all of these variants are assessed will be necessary to sort out this established association and determine why and how the HLA region genes contribute to lupus pathology.

Meta-analysis

For those genes that have already been tested for the same variant in multiple cohorts, meta-analysis can be used to combine the results at a specific site. This site is usually a SNP (such as the -308 variant of TNF-α), but it could be an insertion/deletion (as in ACE), or even a microsatellite. Meta-analysis takes into account the size of the sample cohorts and the size of the effect for each report. A preliminary comparison is used to determine if the effects are all in the same direction. For example, two reports might both be in favor of association but have opposite risk alleles. In addition, a negative report may be in the same direction as other positive reports but be underpowered to achieve statistical significance on its own. Of the nine genes in Table 4, all but ACE were found to have association in meta-analysis.

Newly discovered associations

Unless it is by coincidence, confirmation work often lags behind an initial report by at least a year or two. In addition, since so many of these associations are likely to be either population-specific or false positives, attempts to confirm may be unsuccessful, and, if underpowered, negative reports should not be published at all. One exception to this pattern has been interferon regulatory factor 5 (IRF5). Following the initial

### Table 2. Confirmed linkage regions.

<table>
<thead>
<tr>
<th>Region</th>
<th>Group(s)</th>
<th>Cohort type</th>
<th>Associated gene(s)</th>
<th>Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q23</td>
<td>OMRF</td>
<td>Extended pedigrees</td>
<td>FCGR2A, FCGR2B, FCGR3A, FCGR3B</td>
<td>EA, AA</td>
</tr>
<tr>
<td>1q31–32</td>
<td>UU</td>
<td>Extended pedigrees</td>
<td></td>
<td>EU</td>
</tr>
<tr>
<td>1q41–43</td>
<td>UCLA, USC</td>
<td>Extended pedigrees</td>
<td></td>
<td>EA, HIS</td>
</tr>
<tr>
<td>2q37</td>
<td>UU</td>
<td>Extended pedigrees</td>
<td></td>
<td>EU</td>
</tr>
<tr>
<td>4p16</td>
<td>OMRF</td>
<td>Sib-pairs</td>
<td>HLA-DR</td>
<td>EA, AA, HIS</td>
</tr>
<tr>
<td>6p11–21</td>
<td>UMN</td>
<td>Sib-pairs</td>
<td></td>
<td>AA</td>
</tr>
<tr>
<td>10q23</td>
<td>OMRF, UCLA</td>
<td>Sib-pairs</td>
<td></td>
<td>EA, HIS</td>
</tr>
<tr>
<td>12q24</td>
<td>OMRF</td>
<td>Sib-pairs</td>
<td></td>
<td>EA, HIS</td>
</tr>
<tr>
<td>16q12–13</td>
<td>UMN, OMRF</td>
<td>Sib-pairs</td>
<td></td>
<td>EA, AA, HIS</td>
</tr>
</tbody>
</table>

*PARP was initially associated in this linkage region [13], but subsequent studies have failed to confirm the association [14–17].

AA: African-American; EA: European-American; EU: European; HIS: Hispanic; OMRF: Oklahoma Medical Research Foundation; UCLA: University of California, Los Angeles, CA, USA; UMN: University of Minnesota, MN, USA; USC: University of Southern California, CA, USA; UU: University of Upsala, Sweden.
Future perspective

If you consider the large body of literature in lupus genetics as a whole, you cannot help but be dismayed that so much has been done and yet so little is truly known. Serious attempts to resolve the conflicts in the literature, to define the scope of each genetic effect, and to determine the interactions of the genetic factors that lead to disease risk must be attempted before true clarity is achieved. Large-scale collaborations that allow for cross-typing of different cohorts for the same markers in the same genes will be critical for determining the relative impact of each genetic variant on disease risk and for generating more complex models of disease that include epistatic interactions and clinical outcome variables. We already know that SLE is a complex disease with a pleomorphic presentation; we should expect that the final equation will also be complex to include multiple genes and encompass multiple phenotypes.

Another area in which we must improve and refine our techniques is in the consideration of racial heritage. The genetic background of the human species is varied, and even though we attempt to group patients by ancestry when matching controls, for example, this is fraught with inherent errors. Each racial subcategory we designate must be broad, so as to include as many samples as possible, but ancestry is not a simple categorical variable; it is a scalar, and will remain so no matter how we try to draw the lines in the sand. For example, African-Americans have admixture of European genes contributing anywhere from 10 to 30% of their ancestry, and this proportion can affect disease risk [49]. Our understanding of the effects of ancestral heritage will improve as we begin to characterize all participants in our genetic studies by their proportion of ancestral genotypes and use these variables as cofactors to combine all of our data. Over the next 10 years, the use of ancestry informative markers to characterize the ancestry of any individual will likely replace the use of questionnaires and self-designated racial categories. This revolution will refine our definitions of race immeasurably and will enable meaningful world-wide collaborations to confirm and further define genetics effects, wherein patient cohorts of different ethnicities will truly be combined for maximum power and not merely run side by side for economic convenience.

Table 3. Classification of lupus candidate genes by types of literature report.

<table>
<thead>
<tr>
<th>Type of gene</th>
<th>No. genes</th>
<th>No. reports*</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causative mutations reported</td>
<td>11</td>
<td>18+</td>
<td>C1q, C2, FasL, DNase 1</td>
</tr>
<tr>
<td>Confirmed associations</td>
<td>20</td>
<td>75+</td>
<td>HLA-DR, IF5, APRIL</td>
</tr>
<tr>
<td>Conflicting reports</td>
<td>48</td>
<td>532</td>
<td>IL10, CTLA4, MBL</td>
</tr>
<tr>
<td>Unconfirmed associations</td>
<td>28</td>
<td>28</td>
<td>STAT4, CR2, IL-21, ICOS</td>
</tr>
<tr>
<td>Phenotypic associations only</td>
<td>28</td>
<td>92</td>
<td>CD38, BDNF, ITGA2</td>
</tr>
<tr>
<td>Negative reports only</td>
<td>53</td>
<td>56</td>
<td>NFkB1, RUNX1, CD40</td>
</tr>
<tr>
<td>Total</td>
<td>188</td>
<td>800+</td>
<td></td>
</tr>
</tbody>
</table>

*Excludes literature prior to 1995 on HLA and complement, which is a body of work of over 300 papers.
New technology has enabled us to gather more data than ever before, but we are currently attempting to lift our heads above the rising tide in the search for understanding. It is not enough to have some data on some of the patients some of the time, we must make efforts to combine and correlate what we know and to fill in the gaps. Much work remains to be done to resolve the apparent conflicts in the literature, as well as to combine the findings we are reasonably certain of into a coherent final picture. There is also much work in progress to define the exact nature of the causative polymorphisms within each risk allele, and the transition from an associated variant with no obvious functional consequence to a meaningful model of autoimmune pathogenesis will be an active area of investigation for many groups over the decade to come. This important work should raise us to a new level of understanding of the pathophysiology of SLE.

Financial & competing interests disclosure
The author has no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending or royalties.

No writing assistance was utilized in the production of this manuscript.

### Table 4. Genes subjected to meta-analysis.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Location</th>
<th>Reports for</th>
<th>Reports against</th>
<th>Meta analysis</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE</td>
<td>17q23</td>
<td>4</td>
<td>9</td>
<td>Negative</td>
<td>[24]</td>
</tr>
<tr>
<td>CTLA4</td>
<td>2q33</td>
<td>9</td>
<td>11</td>
<td>Positive</td>
<td>[25,26]</td>
</tr>
<tr>
<td>FCGR2A</td>
<td>1q23</td>
<td>15</td>
<td>21</td>
<td>Positive</td>
<td>[27-30]</td>
</tr>
<tr>
<td>FCGR3A</td>
<td>1q23</td>
<td>11</td>
<td>9</td>
<td>Positive</td>
<td>[30]</td>
</tr>
<tr>
<td>IL10</td>
<td>1q32</td>
<td>13</td>
<td>9</td>
<td>Positive</td>
<td>[31]</td>
</tr>
<tr>
<td>MBL2</td>
<td>10q11</td>
<td>16</td>
<td>8</td>
<td>Positive</td>
<td>[18,32]</td>
</tr>
<tr>
<td>PTPN22</td>
<td>1p13</td>
<td>8</td>
<td>2</td>
<td>Positive</td>
<td>[33]</td>
</tr>
<tr>
<td>TNF</td>
<td>6p21</td>
<td>19</td>
<td>12</td>
<td>Positive</td>
<td>[34]</td>
</tr>
<tr>
<td>TNFRSF1B</td>
<td>1p36</td>
<td>4</td>
<td>10</td>
<td>Positive</td>
<td>[35]</td>
</tr>
</tbody>
</table>

### Executive summary

**Studies to elucidate the complex genetic causes of systemic lupus erythematosus**
- Linkage studies type microsatellite markers across the whole genome in pedigrees with at least two systemic lupus erythematosus (SLE) patients and define a large linkage region.
- Association studies type SNPs in cases and controls, looking for differences in frequency at a specific base.
- Both linkage and association studies must be confirmed with work in a second, independent cohort. These attempts to confirm reported associations have led to a large body of conflicting literature.

**Established genes**
- Complement deficiencies in C1q, C2 and C4 cause SLE, but these are responsible for less than 1% of lupus cases.
- The HLA region is well known to be associated with disease, but there are a number of genes inherited with HLA that could be responsible for the increased risk. It is possible that a combination of factors is necessary.
- Other genes with both positive and negative reports in the literature have been established through meta-analysis.
- Newly discovered associations with promise include IRF5 and STAT4.

**Future perspective**
- Collaborations that result in the testing of larger cohorts will make confirmation and characterization of the existing candidate genes easier.
- Complex models of disease risk will incorporate multiple genetic elements, as well as clinical and demographic variables.
- A set of markers that more clearly define ancestry will be characterized and used routinely to assess all samples, such that ethnicity can be incorporated as a cofactor, rather than be used as a dividing line.
Bibliography
Papers of special note have been highlighted as either of interest (••) or of considerable interest (•••) to readers.


- Excellent review of the complement deficiencies associated with systemic lupus erythematosus.


- Provides an example of how copy number variations can contribute to disease risk.


Affiliation

• Andrea L. Sestak, MD PhD, Research Assistant Member
Oklahoma Medical Research Foundation,
825 NE 13th St Oklahoma City, OK 73104,
USA
Tel.: +1 405 271 7026
Fax: +1 405 271 4110
sestaka@omrf.org

• Outlines the work in delineating the complex effects at the IRF5 locus.