Family studies in systemic lupus erythematosus

Most diseases have a genetic basis. How genes are involved varies greatly from case to case and their products may play a major or a secondary role in disease development. Traditionally, the genetics of common diseases has been analysed through association studies. Many of these studies have used only small sets of patients and have rarely attempted replication in independent sets. Their significance is difficult to assess.
as the results are in many instances inconclusive, particularly when other researchers have attempted replication in a different population. Success stories include the association between HLA-B27 and ankylosing spondylitis, and the association of DRβ1*04 alleles with rheumatoid arthritis. The main reason for these successes has been the happy coincidence between the interest in HLA at the time, the fact that HLA is indeed of importance in diseases of immunological pathogenesis, and the strong and unusual linkage disequilibrium of the MHC region at 6p21. However, HLA alone does not account for the overall risk to the development of disease: the presence of other gene effects is required. The combination of genes involved in susceptibility to a given disease may be different for each population, which makes the task of discovering these genes even more difficult. Interaction between genes is to be expected, and these interactions may be extremely complex and unpredictable. Therefore, new strategies to study the genetics of common diseases are required. It has not been until very recently that the fields of genetics and clinical medicine in general and rheumatology in particular have met and begun to be integrated. The process of integration involves close collaboration between geneticists, molecular biologists and clinical practitioners as well as basic immunologists and cell biologists. Only through the integrated study of diseases will we be able to understand their pathogenesis. Progress towards understanding pathogenesis begins with adequate epidemiology and family studies that include both environmental and genetic factors. It is the genetic factors that fall within the scope of the present paper. The use of families for genome-wide genotyping and linkage analysis and for family-based association studies has been seen as an alternative, not yet exhausted, to the case–control studies done in the past. Results obtained from such studies are now yielding a harvest, but much work lies ahead.

On the other hand, new methods are being proposed continually. The reason for this is that case–control studies, although they have several problems, are always more powerful and easier to perform than family studies, due to the difficulty in the collection of material and in gathering information from relatives to patients. This is why several groups have put forward the use of genome-wide analysis of case–control cohorts using closely spaced single-nucleotide polymorphisms (SNPs). However, this approach poses several problems. The degree of linkage disequilibrium (L.D.) between anonymous SNPs and a causal polymorphism or mutation varies enormously along the human genome, and we do not yet have a clear idea of how L.D. looks. To analyse a single polymorphism per gene is insufficient to define if that gene is involved in disease, as SNPs have evolved at different times in human history. The size of the patient group required to reach statistical power with this approach, assuming an 80% linkage disequilibrium with a causal mutation, is in the range of 1000–2000 to give statistical significance at \( P = 0.05 \) for a given SNP; and there is also the problem of multiple testing. The high cost of using SNP methods makes such analysis unrealistic.

Family studies, although less powerful, are more reliable than case-control studies. But, as I will show below, the different designs for genetic analysis are not mutually exclusive, and case–control cohorts may be required for the independent replication of a result in a family study. Families with multiple cases of a disease are used for genetic linkage analysis and positional cloning. Such methods have a long tradition in the study of monogenic disorders, and hundreds of genes and mutations for rare metabolic and cognitive diseases have been discovered in this way. However, most common diseases pose a methodological problem. The penetrance of the disease is less than 100%, environmental factors play an important role in disease expression and genetic heterogeneity is expected to be very high. Therefore, there are a number of factors that should be taken into account when linkage analyses are performed for common diseases. The usual strategy is as follows.

To find genes involved in susceptibility to any disease, initial whole-genome screens are performed using highly polymorphic microsatellite markers. When a region of linkage is found and confirmed we initiate haplotype construction and recombination finding. At this point we have to be aware that only affected individuals can give us information on the recombination sites, as healthy individuals might possess the at-risk genotypes but not express the disease. For systemic lupus erythematosus (SLE), for example, we may expect that men represent such a case. When a critical region is defined, analysis of candidate genes can be done if an obvious candidate gene is present. Such candidate gene analysis is then performed with single-case families or trios and/or with sporadic case–control sets. Support from linkage analysis with multicase families and association analysis with an independent set of sporadic patients is strong evidence for a candidate gene. Care should always be taken to identify the ethnicity of the multicase families and sporadic cases for the association analysis. If no obvious candidate gene is found, a possibility is to analyse the expression of the RNA for the genes in the region as a means of filtering out genes that are expressed in irrelevant tissues in normal individuals. The remaining genes can then be analysed for association. Where statistical genetics ends, functional genetics begins. The relationship between genotype and a functional/dysfunctional role for the polymorphism associated with the disease has to be demonstrated, and this is one of the challenges that molecular geneticists confront. Some of the strategies that can be used at this point include expression analysis, in which differences in pattern of expression between individuals having an associated allele and those without are investigated. Other methods that have previously been used only by cell biologists should also be included. This means that a human genetics laboratory must be prepared to acquire the expertise required for the experiments that will be needed to demonstrate the function of a polymorphism in a given
disease. The final proof may come from animal models, but this topic is outside the scope of this editorial.

SLE is a relatively uncommon disease with a prevalence of 0.05% in Caucasians and with a familial aggregation index of about 30, which is quite high compared with rheumatoid arthritis (I = 5). This fact and results of studies on twin concordance suggest that genes are important in susceptibility to SLE. In recent years, at least four whole-genome screens have been performed for SLE [1–4]. In the main, two strategies have been used, namely sib-pair, non-parametric analysis and extended family-based parametric analysis, and both strategies have produced good results. In general, five regions were detected with significant linkage, and one more region had been found previously by analysing a genomic segment of chromosome 1 syntenic with a mouse SLE locus [5, 6]. These regions are shown in Table 1. In addition, a number of regions were detected that did not reach the genome-wide significance level of LOD Z = 3.30 but were observed in different independent sets of families. An example is a region in chromosome 4p16 [7].

We have concentrated on the SLEB2 locus [8] and performed further analyses of the 2q37 region. Unfortunately, it has been a long task because of the scarcity of genome sequence data in this region and the difficulty in ordering markers due to the high recombination rate of the telomere [8]. Most available maps are incorrect and we have had to repeat some of the work. An interesting candidate gene is located in this region, namely the gene for the immunoreceptor tyrosine-based inhibitory motif-containing molecule PD1 [9]. We have performed fine mapping of the region, identified and analysed new SNPs and constructed haplotypes. We have identified a major haplotype associated with SLE in multicase and transmission/disequilibrium test trios from the Nordic countries. We also identified SNPs within the PD1 gene and found association between one of these polymorphisms and SLE and with nephritis [10].

There is more to be done. We need to investigate whether this polymorphism is functional in any way, as it is located within an intron, or whether it is in linkage disequilibrium with a functional mutation. In addition, we need to investigate if this association occurs in other populations, as we have seen that the ethnic origin of the families should be considered carefully when investigating the genetics of SLE. The above is just one example of the work that can be done in order to understand the genetics of common diseases and to show how family studies are of use in reaching this goal.

M. E. ALARCÓN-RIQUELME
Institute of Genetics and Pathology, Section for Medical Genetics, Rudbeck Laboratories, University of Uppsala, Dag Hammarskjölds väg 20, 751 85 Uppsala, Sweden

References

Table 1. Major regions detected in genome screens for SLE where linkage was supported by the lod score (>3.00)

<table>
<thead>
<tr>
<th>Study</th>
<th>Region</th>
<th>Candidate gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moser et al. [1]</td>
<td>1q23</td>
<td>FcGRIIA</td>
</tr>
<tr>
<td>Tsao et al. [5]</td>
<td>1q42</td>
<td>PARP</td>
</tr>
<tr>
<td>Lindqvist et al. [2]</td>
<td>2q37 (SLEB1)</td>
<td>PDI</td>
</tr>
<tr>
<td>Lindqvist et al. [2]</td>
<td>4p13</td>
<td>?</td>
</tr>
<tr>
<td>Gaffney et al. [3, 4]</td>
<td>6p21</td>
<td>HLA/TNF/C4</td>
</tr>
<tr>
<td>Gaffney et al. [3, 4]</td>
<td>16q13</td>
<td>?</td>
</tr>
</tbody>
</table>