Linkage of cytokine genes to rheumatoid arthritis. Evidence of genetic heterogeneity

Sally John, Anne Myerscough, Angela Marlow, Ali Hajeer, Alan Silman, William Ollier, Jane Worthington

Abstract

Objective—To investigate linkage of candidate disease susceptibility genes to rheumatoid arthritis (RA) in affected sibling pairs stratified for specific clinical features.

Method—Two hundred RA affected sibling pairs were genotyped for informative microsatellite markers mapping within or less than 3cM from: INFα, INFγ, INFβ, IL1α, IL1β, IL1R, IL2, IL6, IL5R, IL8R, BCL2, CD40L, NOS3, NRAMP, α-anti-trypsin, and α-anti-chymotrypsin, using fluorescence based automated technology. Linkage was examined by defining allele sharing sibling pairs. This was assessed by maximum likelihood—inheritance by descent methods.

Results—An increase in allele sharing was seen for IL5R in female sibling pairs (LOD 0.91, p = 0.03), for INFγ in sibling pairs with an affected male (LOD 0.96, p = 0.03) and most significantly for IL2 in sibling pairs where one or both were persistently seronegative (LOD 1.05, p = 0.02).

Conclusion—Weak evidence of linkage of RA to IL5R, IFNγ, and IL2 has been detected in clinical subsets of sibling pairs suggesting that RA is a genetically heterogeneous disease.

(Royal Society of Medicine Press)

Rheumatoid arthritis (RA) is a chronic, relapsing inflammatory condition of unknown aetiology. Twin and family studies provide evidence to support the involvement of both genetic and environmental factors in the aetiology of RA. One measure of the size of the genetic component of susceptibility is the degree of familial clustering (λf), which is calculated from the ratio of the risk of RA in siblings of patients and the population prevalence. Estimates of λf for RA vary between 2–10 depending on the sibling recurrence risk and population prevalence values used in the calculation. Using a sibling recurrence risk of 3.9% and a population prevalence of 0.8%, λf of 4.9 is obtained. To date only HLA-DRB1 on chromosome 6 has been confidently identified as being linked to RA susceptibility. A number of alleles of this locus contain the “shared epitope” sequence and are associated with the disease. Analysis of the first 100 families of the ARC National Repository indicates that HLA accounts for approximately 40% (λHLA 1.8) of the total genetic component of susceptibility. Thus the majority of the genetic risk of RA is still to be accounted for, although as HLA is the major disease locus the remaining genetic effect is likely to result from a number of further loci.

Screening of the genome for non-HLA disease susceptibility genes is currently in progress in RA using microsatellite markers for linkage analysis in large numbers of affected sibling pairs (ASP) families. Simulations and experience to date suggest that detection of disease genes with a small (≤1.8) will require many hundreds of families and markers at much closer intervals than the average of 11 cM currently being used. Limited results from French and British studies of 200 RA ASP families detected linkage to HLA but no reproducible effects at any other loci. An alternative approach is to apply the same microsatellite technology to sibling pair families to target likely candidate susceptibility genes. The pathogenesis of RA indicates several molecules, which may be of potential importance in disease susceptibility. For example, many cytokines have been shown to be up regulated in the synovial tissue of RA patients and are produced spontaneously by synovial cells taken from RA patients. To identify the genes that are critical to the development of RA compared with those that are up regulated as part of the inflammatory process, it is necessary to investigate aetiology at a genetic level. Highly polymorphic markers close to gene sequences can be used in family based linkage analysis studies to directly test candidate genes. Investigating candidate genes in this way has the advantage that as the marker selected is genetically very close to the candidate gene, there is a greater chance of detecting even weak linkage.

Many of the proteins implicated in the pathology of RA will be involved as a consequence of the disease process and not a cause of it. One approach to identify molecules involved with the aetiology of RA is to identify genes that are linked to RA. Here we report on linkage analysis of 200 multi-case RA families for 16 potential candidate genes, to test the hypothesis that these genes are linked with RA. Analysis of 200 ASP families using markers within 3cM of candidate genes will permit genetic effects as small as λ of 1.5 to be detected. In view of the clinical and genetic heterogeneity within RA we have stratified the data for age at onset, sex, seropositive disease, and HLA haplotype sharing.

Methods

FAMILY MATERIAL
DNA was available from 200 RA families collected for the ARC National Repository for family material. Pedigrees for the first 100
families have been published and are accessible on the ARC ERU web site (address http://130.88.95.76/natrep.html) together with pedigrees for the second 100 families. All family members were seen and examined according to a standard protocol. Information was also obtained from physician records and rheumatoid factor (RF) measured to classify patients obtained from physician records and rheumatoid arthritis to a standard protocol. Information was also obtained from physician records and rheumatoid factor (RF) measured to classify patients as having RA if they satisfied the 1987 ACR criteria modified for genetic studies. In subjects as having RA if they satisfied the 1987 ACR criteria modified for genetic studies.

In total, 281 affected sib-pairs were included of which 101 sib-pairs were made up from families with between three and seven affected siblings. DNA from both parents was available for 41 families, and from one parent in a further 25 families. Unaffected siblings were genotyped to infer missing parental genotypes. In total 703 individuals were available for genotyping.

**CANDIDATE GENE MARKERS**

A set of 11 markers with reported heterozygosity ranging from 0.57–0.9, which mapped within or less than 3cM from 16 candidate genes has been compiled from published data and data held on the genome data base (GDB) (table 1). The candidate genes include a number of cytokines, interferon α (IFNα), interferon β, the interferon γ gene cluster, interleukin 1α (IL1α), interleukin 1β, interleukin 1 receptor, interleukin 2 (IL2), interleukin 6 (IL6), interleukin 5 receptor (IL5R), and the interleukin 8 receptor (IL8R). The other candidate genes are B-cell lymphoma 2 (BCL2), CD40 ligand (CD40L), natural resistance associated macrophage protein (NRAMP1), endothelial nitric oxide synthase, (NOS3) α, anti-trypsin (PT) and α1-antichymotrypsin (AACT).

**MICROSATELLITE GENOTYPING**

Semi-automated analysis of microsatellite genotypes was performed using a PE Applied Biosystems 377 DNA sequencer with Genescan analysis and Genotyper software (PE Applied Biosystems). Forward polymerase chain reaction (PCR) primers were fluorescently labelled with either 6-FAM, HEX or TAMRA 350 size standards (PE Applied Biosystems). PCR products from two DNA samples were included on every gel as a control to monitor any gel to gel variation.

**DATA ANALYSIS**

Sib-pair analysis methods were used to test for excess allele sharing between sibs at all of the loci tested. Sharing at each locus was quantified by the number of alleles shared identical by descent (IBD), alleles shared IBD have the same genotype and the same ancestral origin. Under the null hypothesis of no linkage IBD sharing of 2, 1 or zero alleles is expected to be in the proportion 1:2:1, an increase in the proportion sharing two alleles indicates linkage. As RA has a comparatively late age at disease onset, parental data are often missing, making it difficult to assign IBD sharing directly. The maximum likelihood score (MLS) method was therefore used to estimate allele sharing. The MLS IBD method, allows for uncertainty of IBD assignment resulting from unknown marker haplotypes or missing parental genotype data by inferring most likely IBD sharing probabilities. MLS IBD was carried out in the SPLINK package, version 1.05, (David Clayton MRC Biostatistics unit, Cambridge). In this analysis the IBD data from multiple affected sibships were given the conservative weighting of 2/n where n is the number of affected sibs.

The transmission disequilibrium test (TDT) is a test of linkage in the presence of an association. Parents heterozygous for a marker are used to test for deviation from random transmission from the parent to an affected offspring for one or more alleles of a candidate marker. A maximum likelihood method of TDT was carried out in TRANSMIT, version 2.1 (David Clayton, MRC Biostatistics unit Cambridge). In the TDT, the proband from each pedigree was used in the analysis.

**Stratification of the data**

Data were analysed as a complete set and for certain markers where some deviation from
random inheritance was seen, a number of stratifications were made to the data set as specific genetic effects may be greater in some subgroups of patients. The following criteria were used for stratification:

1. Age at onset—both siblings having an age of onset of less than 50 years were compared with those where one or both siblings had a later age of onset. It is predicted that stronger genetic effects would be seen in siblings with a younger age at onset.

2. Sex—As there is evidence that male RA is genetically different, female/male and male/male sibling pairs were compared with female/female pairs. There were insufficient male/male pairs for direct comparison.

3. Disease severity—sibling pairs who both have erosions and who were “ever seropositive” were compared with sibling pairs with a less severe disease. The less severe disease group contained 49 sibling pairs with at least one sibling who had no erosions and 47 sibling pairs with at least one sibling who was “persistently seronegative”. There were eight sibling pairs who were both persistently seronegative and only four sibling pairs who both had no erosions.

4. HLA haplotype sharing—as HLA is known to be the major susceptibility locus in RA, sib-pairs were subdivided into two groups by IBD status at HLA. One group of families showing stronger evidence of linkage to HLA was compared with those discordant for HLA background—that is, sharing 1 or zero HLA haplotypes. It would also be interesting to compare siblings who were “shared epitope” positive with those who were “shared epitope” negative, however the numbers of “shared epitope” negative pairs (14) was too small to analyse.

Results

Whole data set

Estimated IBD sharing generated by SPLINK did not generate any significant LOD scores for any of the markers tested (table 3) although a small increase in sharing of two alleles IBD was seen for markers IL2, IL5R, and IFNγ.

Stratification of the data set

RA is an extremely heterogeneous condition, and as it is expected that different genes may contribute to variation in disease expression, the whole data set was stratified as described in the methods. MLS IBD analysis was performed on the subgroups for markers IL5R, IL2, and IFNγ, where some deviation from random sharing had been seen in the whole data set. No significant deviation from random allele sharing was seen in the sibling pairs stratified by age at onset of RA (table 4(A)).

Two LOD scores were significant at p < 0.05 level in the patient group stratified by sex (table 4(B)). IL5R showed an increase in two alleles sharing in the all female pairs (LOD 0.9, p = 0.03) and a similar increase was detected for IFNγ in pairs with a male—that is, male/male and male/female pedigrees (LOD 0.9, p = 0.03).

The data set was also subdivided to test for linkage in sibling pairs with more severe disease. The strongest evidence of linkage was for IL2 in the group with less severe disease (LOD 1.05, p = 0.02 (table 4(C)). Many patients in this group were seronegative and to
Table 4  MLS IBD analysis in patient subgroups for IFNγ, IL5R, and IL2

<table>
<thead>
<tr>
<th>Marker</th>
<th>Patient group</th>
<th>Total IBD assignment 0 : 1 : 2</th>
<th>χ²</th>
<th>Equivalent LOD</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) IFNγ</td>
<td>Early onset</td>
<td>29 : 67 : 37</td>
<td>0.48</td>
<td>0.1</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Late onset</td>
<td>24 : 39 : 38</td>
<td>2.59</td>
<td>0.56</td>
<td>0.08</td>
</tr>
<tr>
<td>IL5R</td>
<td>Early onset</td>
<td>33 : 59 : 40</td>
<td>0.97</td>
<td>0.2</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Late onset</td>
<td>21 : 43 : 28</td>
<td>0.74</td>
<td>0.16</td>
<td>NS</td>
</tr>
<tr>
<td>IL2</td>
<td>Early onset</td>
<td>17 : 59 : 32</td>
<td>2.8</td>
<td>0.6</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Late onset</td>
<td>20 : 42 : 19</td>
<td>0</td>
<td>0</td>
<td>NS</td>
</tr>
<tr>
<td>(B) IFNγ</td>
<td>All female</td>
<td>33 : 64 : 31</td>
<td>0</td>
<td>0</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>female/male, male/male</td>
<td>24 : 43 : 42</td>
<td>4.43</td>
<td>0.96</td>
<td>0.03</td>
</tr>
<tr>
<td>IL5R</td>
<td>All female</td>
<td>26 : 40 : 38</td>
<td>4.20</td>
<td>0.91</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>female/male, male/male</td>
<td>23 : 49 : 23</td>
<td>0</td>
<td>0</td>
<td>NS</td>
</tr>
<tr>
<td>IL2</td>
<td>All female</td>
<td>23 : 52 : 26</td>
<td>0.07</td>
<td>0</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>female/male, male/male</td>
<td>17 : 51 : 27</td>
<td>2.50</td>
<td>0.54</td>
<td>0.08</td>
</tr>
<tr>
<td>(C) IFNγ</td>
<td>Severe disease</td>
<td>21 : 36 : 22</td>
<td>0.16</td>
<td>0.04</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>All female</td>
<td>35 : 72 : 51</td>
<td>1.7</td>
<td>0.38</td>
<td>NS</td>
</tr>
<tr>
<td>IL5R</td>
<td>Severe disease</td>
<td>21 : 34 : 21</td>
<td>1.02</td>
<td>0.2</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>All female</td>
<td>31 : 62 : 39</td>
<td>0.8</td>
<td>0.19</td>
<td>NS</td>
</tr>
<tr>
<td>IL2</td>
<td>Severe disease</td>
<td>16 : 38 : 13</td>
<td>0</td>
<td>0</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>All female</td>
<td>19 : 64 : 41</td>
<td>4.85</td>
<td>1.05</td>
<td>0.02</td>
</tr>
<tr>
<td>(D) IFNγ</td>
<td>HLA concordant</td>
<td>17 : 33 : 23</td>
<td>0.76</td>
<td>0.18</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>HLA discordant</td>
<td>32 : 52 : 38</td>
<td>0.81</td>
<td>0.18</td>
<td>NS</td>
</tr>
<tr>
<td>IL5R</td>
<td>HLA concordant</td>
<td>16 : 30 : 20</td>
<td>0.48</td>
<td>0.1</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>HLA discordant</td>
<td>24 : 41 : 30</td>
<td>1.0</td>
<td>0.24</td>
<td>NS</td>
</tr>
<tr>
<td>IL2</td>
<td>HLA concordant</td>
<td>12 : 35 : 14</td>
<td>0.1</td>
<td>0.02</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>HLA discordant</td>
<td>20 : 50 : 29</td>
<td>1.16</td>
<td>0.25</td>
<td>NS</td>
</tr>
</tbody>
</table>

Test if IL2 is linked specifically to seronegative RA we analysed sibling pairs where one or both were seronegative. Despite the smaller sample size of this subgroup, the LOD score was maintained at 1.07 (p = 0.02) with final total IBD sharing assignments of (0-IBD: 1-IBD: 2-IBD) 9:35:25. No significant deviation from random allele sharing was seen in the families stratified by HLA sharing (table 4(D)). The marker for NRAMP (D2S1471) was also analysed in the HLA subgroups as some weak evidence of linkage had previously been shown in a smaller set of ARC families and in patients discordant for HLA haplotypes. In this larger data set, no deviation from random sharing was observed (data not shown).

TDT ANALYSIS

As IFNγ, IL5R, and IL2 are all markers within the gene sequence, evidence of linkage to these genes would be further supported by demonstration of an association between specific alleles of the microsatellite marker and RA, resulting from linkage disequilibrium with a potential functional polymorphism in the gene. Markers for IL5R, IFNγ, and IL2 were therefore analysed by TDT in the subsets of patients where there was some preliminary evidence of linkage (p≤0.05). No evidence of linkage to a particular allele of IL5R or IFNγ was seen (data not shown). There were 16 alleles observed for the IL2 microsatellite. Table 5 shows TDT analysis of observed and expected transmission of each of these alleles. Two alleles of the IL2 marker that were only 2bp apart (141 and 143bp) appeared to be inherited by affected offspring more often than expected. If the transmission of these two alleles were combined, a χ² of 4.37 (p=0.05 one degree of freedom) was observed.

Discussion

We have investigated the contribution of a number of candidate gene loci to RA susceptibility using linkage analysis methods. Analysis of the data before stratification showed no significant deviation from expected allele sharing for any of the markers tested. Although the results were negative and a major contribution for these genes can be excluded, the actual effect of each locus that can be excluded in this dataset varies depending upon a number of factors. Over 200 sibling pairs have been genotyped in this investigation, but because the markers were not fully informative and also because of an absence of parental information, the amount of information contributed by the families is variable. It is possible to exclude a genetic effect (s=1.5) for the following genes; IFNγ, IL2, IL5R, NRAMP, IL8R, NOS3, and PI. For IFNγ, IL16, and CD40L, which were less informative, a genetic effect with a λs = 1.7 can be excluded. The heterozygosity of BCL2 (0.36), calculated from our data set, was much lower than the reported heterozygosity; therefore, it is only possible to exclude a λs = 1.85. We conclude that none of the genes tested in this investigation has a role as great as HLA, which was calculated to have a λs of 1.8, based on analysis of the first 100 families in the ARC National Repository. A possible reason for failing to detect weak genetic effects in complex disease is heterogeneity. To deal with this, the data were subdivided into a number of more homogeneous subsets and analysis was repeated for IL5R, IFNγ, and IL2, where some deviation from random allele sharing was observed. For each of these genes, an increase in allele sharing (p < 0.05) was observed in a patient subgroup.

One way of supporting the initial finding of linkage is to demonstrate linkage disequilibrium between the marker and a potential functional polymorphism within the gene. TDT detects linkage in the presence of an association. There was some evidence that the larger alleles of the IL2 microsatellite were being transmitted to affected offspring in the patient subgroup with less severe disease, more often than expected. This provides some preliminary evidence that alleles of the IL2 microsatellite may be in linkage disequilibrium with a functional polymorphism in the gene. The
Evidence for linkage was seen most convincingly in seronegative RA. HLA associations have already suggested that patients with seronegative RA may represent a disease with a different genetic background. IL-2 is primarily a Th1 cytokine produced by both naïve and activated CD4+ve cells, and functional polymorphisms within regulatory or coding sequences could lead to changed immune function. Interestingly, IL-2 knockout mice develop severe autoimmune disease and lymphocyte proliferation, suggesting functions of IL-2 are more diverse than previously realised.

The TDT analysis did not support an association with both IFNγ or IL5R alleles and RA, there are a number of explanations. Highly polymorphic microsatellites with many alleles are not always in linkage disequilibrium with less polymorphic markers, for example, functional biallelic point mutations, which are genetically very close. Linkage also extends much further than linkage disequilibrium and the linkage finding may be true but result from linkage, not with the candidate gene, but with a different gene in the same region. It is also possible that these results may be false positives. In the initial analysis on 200 families, no correction for multiple testing was made, as there was an a priori hypothesis for investigating each of the candidate genes. However, for each marker tested in the patient subsets, nine separate analysis tests were performed. It is possible that for the three genes tested in this way that 1–2 false positive results would be seen at a 5% significance level. Replication of these results in further data sets or the demonstration of an association between RA and a functional polymorphism of the genes is required to confirm which, if any of these findings are because of genuine linkage. This study has identified three genes worthy of further study and clearly illustrates the importance of considering genetic heterogeneity when investigating a complex disease such as RA.

Linkage of cytokine genes to rheumatoid arthritis. Evidence of genetic heterogeneity

Sally John, Anne Myerscough, Angela Marlow, Ali Hajeer, Alan Silman, William Ollier and Jane Worthington

doi: 10.1136/ard.57.6.361