Full Length Research Paper

Genome scan meta-analysis in systemic lupus erythematosus strong linkage with loci 6p22.3-p21.1 and 2q31.1-34

Suad AlFadhli1*, Najla AlKharrat2 and Ahmid AlRebai2

1Department of Medical Laboratory Sciences, Faculty of Allied Health Sciences, Kuwait University, Kuwait.
2Centre de Biotechnologie de Sfax, Unit of Bioinformatics, Biostatistics and Signalling, Sfax, Tunisia.

Accepted 15 October, 2009

The genetic contribution to development of systemic lupus erythematosus (SLE) is well established. Several genome scan studies have identified putative susceptibility loci to SLE. However, they have shown high level of inconsistency. Genome search meta-analysis (GSMA) which is a non-parametric method is used to identify genetic regions that rank high on average in terms of linkage statistics across genome scan studies. The validity of GSMA was proven when applied on various complex diseases. We applied the GSMA on 16 genome-wide scans of SLE in various ethnicities published from 1996 - 2008. The SLE GSMA resulted in identifying a total of 4 bins lie above 95% confidence level (P = 0.05) of which 2 bins were above 99% confidence level (P = 0.01); bins 6.2 (6p22.3-p21.1, \( P_{sumrank} = 0.0054 \)), 2.8 (2q31.1-q34) \( P_{sumrank} = 0.0091 \), 16.2 (16p12.3-q12.2) \( P_{sumrank} = 0.0386 \) and 6.1 (6p25.3-p22.3) \( P_{sumrank} = 0.0419 \). The highest summed rank was observed at locus 6p22.3-p21.1 surrounding the HLA region and 2q31.1-q34 which locates various genes that were linked to autoimmunity such as CTLA4. In addition, GSMA identified several other putative regions that may contribute to SLE susceptibility. The application of the GSMA technique to 16 SLE genome-wide linkage studies confirmed linkage to loci 6p22.3-p21.1 and 2q31.1-q34.

Key words: SLE, linkage, genome scan, meta-analysis.

INTRODUCTION

Systemic lupus erythematosus (SLE, also called Lupus; MIM 152700) is a complex systemic autoimmune disease characterized by multiorgan pathology. The severity of the disease, the spectrum of the clinical manifestations and laboratory parameters differ among various ethnic groups where some ethnic groups were shown to have a very early onset of the disease and more severe than others (Hochberg and Petri, 1993; Lawrence et al., 1998). The etiology of SLE is very versatile, involving both environmental and genetic factors and probably also a synergistic relationship between these factors. Genetic factors are likely to play a significant role in susceptibility to SLE in determining the disease expression Wakeland et al. (2001).

The genetic involvement in SLE was studied in various levels; the candidate gene association level, the genome-wide gene expression level and the genome scan linkage level. In the past few years several genes were implicated in human SLE such as; HLA Class II DRB and DQB alleles (Arnett and Reveille, 1992) and early components of the complement cascade (e.g. C1q, C3, C4A, C4B) Sestak et al. (2005). Recently a genome-wide gene expression profiling using microarrays was used by 4 independent groups to identify patterns of gene expression that distinguished most SLE patients from healthy controls (Baechler et al., 2003; Bennett et al., 2003; Han, 2003; Rus, 2002). In addition, several genome scans efforts in families enriched for SLE have identified several putative susceptibility loci (Wakeland et al., 2001; Harley et al., 1998; Cantor, 2004; Gaffney et al., 1998; Gaffney PM et al., 2000; Gray-McGuire et al., 2000; Johansson et al., 2004; Koskenmies et al., 2004; Lindqvist et al., 2000; Moser et al., 1998; Nath et al., 2004a, b; Shai et al., 1999); however, there were always inconsistencies...
among these studies. This could be reasoned for the inadequate studied sample size which reduces the study power and the heterogeneity in the sample genetic background and the analytical method used. Genomescan Meta-analysis (GSMA) is been developed and yielded useful applications for genome scans of several complex diseases (Levinson et al., 2003; Lewis et al., 2003; Van Heel et al., 2004; Demenais et al., 2003; Chiodini BD and Lewis, 2003). GSMA allowed pooling the raw genotype linkage data across several studies, providing greater power to identify regions that showed only weak evidence for linkage in individual studies. The GSMA method is based on several studies of genetically diverse populations. Results from the meta-analysis should direct further studies toward single nucleotide polymorphism association (SNP) studies and positional cloning of SLE development genes.

MATERIAL AND METHODS

The search engines used to screen for the genome scan studies were the NIH PubMed site (http://www.ncbi.nlm.nih.gov/pubmed) and Genetic association database (http://geneticassocia-tiondb.nih.gov/). The following initial keywords were used for the search: “genome-wide screen,” “genome-wide scan,” “genomic scan” and “genome-wide search.” combined with SLE. Then genome scan database of SLE was established using all published studies on families with SLE. This database contained the following information: publication details (author, title, source, year of study); details of study population (ethnic background and studied population); sample-size details (number of probands, individuals, families and sib pairs); genotyping methods (type of markers, number of markers, average polymorphism information content, average spacing of markers); statistical methods and results obtained (All results were included including the positive markers concerning the individual threshold, with localization and marker term; maximum LOD score or Z score, nonparametric LOD score NPL and minimum P value) in addition to the markers with lower statistical value than the threshold. The database was checked by examination of both the discussion section and the reference list of the publications allowing the completeness of the database of genome scans.

All potentially eligible articles (were scrutinized for eligibility and potential overlap of the studied populations. Studies done in the same institution were checked thoroughly. In case of overlap, only the largest study was retained, in order to avoid duplication of data. Eligible studies were those that had performed whole genome scans for strictly defined SLE. Studies with a minimum of 100 microsatellite markers were selected regardless of the statistical analytical method and software employed. In the studies or subsets of subjects with data derived only for specific chromosomes or specific chromosomal regions were excluded. Twenty two genome scan articles were recovered in our search and used to establish the database. However, after applying our exclusion criteria GSMA technique was performed to only 16 SLE genome-wide linkage studies using 10,000 simulations. Data extraction was done independently and discrepancies, if any were carefully sorted out.

RESULTS

Reach results of SLE genome scan

Twenty two genome scan articles were recovered in our search and used to establish the database (Cantor, 2004; Gaffney et al., 1998; Gaffney et al., 2000; Gray-McGuire et al., 2000; Johansson et al., 2004; Koskenmies et al., 2004; Lindqvist et al., 2000; Moser et al., 1998; Shai R et al., 1999; Nath et al., 2004a, b, 2002, 2001; Rao et al., 2001; Namjou B et al., 2002; Namjou B et al., 2005; Scofield RH et al., 2003; Johansson et al. 2006; Xing et al., 2005, 2007, Siegen et al., 2008). The database included Authors, institution in which the study accomplished, year of publication, number and types of pedigrees, number of affected cases, ethnicities, number of markers used, linkage results. Seven articles were excluded (Moser et al., 1998; Tsao, 1997; Nath et al., 2004, 2002, 2001; Namjou et al., 2005; Johansson et al., 2006) as some had overlap in the screened families and others scanned only specific chromosomal region or detailed results were not included in their publication.

GSMA

An average of 342 microsatellite markers was genotyped across the 16 studies. In all studies around 184 markers showed significant linkage with SLE either by p < 0.1 or LOD > 1 or NPL > 2 of which 19% were overlapped between the various genome scan studies. For each chromosome, the summed ranks for each bin, the sumrank p-value and the OrderRank p-value in unweighted and weighted analyses are shown in Table 1.
Table 1. Regions with sumrank < 0.1 and < 0.05 (*a) in both unweighted and weighted GSMA analysis.

<table>
<thead>
<tr>
<th>Unweighted analysis</th>
<th>Weighted analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bin</td>
<td>Cytogenetic location</td>
</tr>
<tr>
<td>2.8</td>
<td>2q31.1-p34</td>
</tr>
<tr>
<td>6.2</td>
<td>6p22.3-p21.1</td>
</tr>
<tr>
<td>6.1</td>
<td>6p25.3-p22.3</td>
</tr>
<tr>
<td>16.2</td>
<td>16p12.3-q12.2</td>
</tr>
<tr>
<td>1.6</td>
<td>1p13.3-q23.3</td>
</tr>
<tr>
<td>1.9</td>
<td>1q32.3-q43</td>
</tr>
<tr>
<td>5.1</td>
<td>5pter-p15.1</td>
</tr>
<tr>
<td>4.7</td>
<td>4q32.1-q34.3</td>
</tr>
<tr>
<td>20.3</td>
<td>20p11.22-q13.13</td>
</tr>
</tbody>
</table>

and Figures 1A and B, respectively. In the unweighted analysis, we observed 9 bins with a summed rank >612 with a significant p-value < 0.1, compared with 4 bins that had significant point-wise sum rank p-value < 0.05. (bin 6.2 sumrank p-value 0.0047, bin 2.8 sumrank p-value 0.0077, bin 16.2 sumrank p-value 0.042 and bin 6.1 sumrank p-value 0.045).

When the ranks from each study were weighted by square root of the affected cases, the summed ranks for all loci identified by unweighted analyses remained the most significant regions with minor differences in the sumrank p-value (bin 6.2 sumrank p-values 0.0054, bin 2.8 sumrank p-value 0.0091, bin 16.2 sumrank p-value 0.0386 and bin 6.1 sumrank p-value 0.0419) and variation in the ranking position (Table 1).

The other regions identified by weighted analysis that had a summed rank of > 612 with a significant sumrank p-value < 0.1 were bin 1.6 (1p13.3-q23.3), bin 1.9 (1q32.3-q43), bin 6.3 (6p21.1-p15), bin 20.3 (20p11.22-q13.13) and bin 2.1 (2p25.3-p25.1) (Table 1). While the unweighted analysis has identified bins 5.1 (5pter-p15.1), bin 4.7 (4q32.1-q34.3) in addition to bins 1.6, 1.9 and 20.3 that had a significant sumrank p-value of < 0.1 (Table 1).

**DISCUSSION**

SLE is one of the complex diseases that have been intensively studied by various groups using genome scan searching for putative susceptibility loci. Although several susceptibility loci have been identified independently, most linkage effects remain inconsistent across these studies. GSMA technique, which is based on several studies of genetically diverse populations, used in pooling the data of the various genome scans has added power to linkage analysis by identifying regions that showed only weak evidence for linkage in individual studies.

Here we have applied GSMA technique to 16 SLE genome-wide linkage studies using 10,000 simulations which confirmed SLE susceptibility linkage to 4 bins representing chromosome 6p25.3-p21.1 (bins 6.1 and 6.2), 2q31.1-34 (bin 2.8) and 16p12.3-q12.2 (bin 16.2) at sumrank p < 0.05. Two of which (bins 6.2 and 2.8) were ranked above 99% confidence level at p < 0.01. Weighted and unweighted analyses gave largely similar results, except that the rank of the bins has shifted in the top 4 ranks. Five more loci were ranked at sumrank p < 0.1. Our result is close to what was recently published by Lee et al. (2005) on meta-analysis for susceptibility loci in systemic lupus erythematosus, which was limited to nine genome scan studies. We replicated their results when performing GSMA to the same 9 genome scans data they used (Cantor, 2004; Gaffney et al., 1998; Gaffney et al., 2000; Gray-McGuire et al., 2000; Johansson et al., 2004; Koskenmies et al., 2004; Lindqvist et al., 2000; Nath et al., 2004b; Shai et al., 1999). However, when including the data from an additional five genome scans (Nath et al., 2004a; Namjou et al., 2002; Scofield et al., 2003; Johansson et al., 2006;
Figure 1. Results from the GSMA, the SumRanks for each bin for the unweighted and the weighted analysis (weighted by the square root of the number of affected cases in each study). Significant levels corresponding to 99% (SumRank p < 0.01), 95% (SumRank p < 0.05), and 90% (SumRank p < 0.1) are shown by horizontal lines.

Xing et al., 2005) the rank of some loci has altered. Although bin 6.2 remains the top ranked in both analysis, bin 2.8 shifted to the second position when it was ranked in the fourth position in Lee’s meta-analysis. While bin 16.2 was in the third rank in our meta-analysis, it had the second rank in Lee’s (Table 2).

Forabosco et al. (2006) has also performed a meta-analysis study on SLE; their results showed strong evidence of linkage in two bins 20.3 and 6.3 representing loci 20p11.22-q13.13 and 6p21.1-p15. We have also identified these two bins with significant linkage at sumrank p < 0.1 in the weighted analysis. Table 2 is used to compare the ranking of the bins that showed evidence of linkage in the three meta-analysis performed on SLE.

The locus that has the strong linkage evidence 6p25.3-p21.1 (bin 6.1 and 6.2) is surrounding the HLA region. Notably, the 2q31.1-34 locus (bin 2.8) identified herein has evidence for linkage to SLE with p-value as 0.0001, 0.0009 and 0.02 in three of the screened genome scan analysis of Scofield et al. (2003), Namjou et al. (2002) and Cantor et al. (2004) respectively and in one of the excluded genome scans of Moser et al. (1998) which showed evidence of linkage at LOD 2.09. This locus was also shown to be linked with other autoimmune diseases such as Osteoarthritis Lee et al. (2006). When we screen this locus for potential genes that could be related to autoimmunity we have found several such as CD28 antigen (MIM;186760), which is Involved in T-cell activation, the induction of cell proliferation and cytokine production and promotion of T-cell survival; Auto antigen La (SSB) (MIM;109090), which was originally defined by its reactivity with auto antibodies from patients with Sjogren syndrome and systemic lupus erythematosus; Engulfment adaptor PTB domain containing (GULP1) (MIM 608165), which is involved in the clearance of cells...
Table 2. Comparison between the three SLE GSMA (weighted) in terms of the ranking of chromosomal region that showed evidence of linkage at p < 0.1.

<table>
<thead>
<tr>
<th>Bin</th>
<th>Cytogenetic band</th>
<th>AlFadhli Psr; (Rank)</th>
<th>Lee and NathPs; (Rank)</th>
<th>Forabosco et al. Psr; (Rank)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2</td>
<td>6p22.3-p21.1</td>
<td>0.00540789 (1)</td>
<td>0.000002 (1)</td>
<td>0.0099 (4)</td>
</tr>
<tr>
<td>2.8</td>
<td>2q31.1-q34</td>
<td>0.00916315 (2)</td>
<td>0.0033 (4)</td>
<td>0.0268 (5)</td>
</tr>
<tr>
<td>16.2</td>
<td>16p12.3-q12.2</td>
<td>0.0386578 (3)</td>
<td>0.00017 (2)</td>
<td>0.0087 (3)</td>
</tr>
<tr>
<td>6.1</td>
<td>6p25.3-p22.3</td>
<td>0.0419368 (4)</td>
<td>0.01588 (6)</td>
<td>NS*</td>
</tr>
<tr>
<td>1.6</td>
<td>1p13.3-q23.3</td>
<td>0.058317 (5)</td>
<td>0.02655 (7)</td>
<td>NS*</td>
</tr>
<tr>
<td>1.9</td>
<td>1q32.3-q43</td>
<td>0.0635302 (6)</td>
<td>0.07023 (8)</td>
<td>NS*</td>
</tr>
<tr>
<td>6.3</td>
<td>6p21.1-q15</td>
<td>0.0776381 (7)</td>
<td>0.01013 (3)</td>
<td>0.0056 (2)</td>
</tr>
<tr>
<td>20.3</td>
<td>20p11.22-q13.13</td>
<td>0.0902157 (8)</td>
<td>0.01254 (5)</td>
<td>0.0044 (1)</td>
</tr>
<tr>
<td>2.1</td>
<td>2p25.3-p25.1</td>
<td>0.098617 (9)</td>
<td>0.07829 (5)</td>
<td>NS*</td>
</tr>
</tbody>
</table>

*NS: non significant.

ACKNOWLEDGMENTS

Wish to thank Ms Eilaf AlShmaly and Bader Al-Temimy for their help. This work was supported by Kuwait University Research administration grant # NM 01/07.

REFERENCES:


