Systemic lupus erythematosus genetics: what's new?

Benjamin Rhodes & Timothy J Vyse†

†Author for correspondence Imperial College London, Section of Molecular Genetics and Rheumatology, Faculty of Medicine, Du Cane Road, London, W12 0NN, UK Tel.: +44 208 383 2339; Fax: +44 208 383 2379; t.vyse@imperial.ac.uk

'Identifying SLE susceptibility genes is important because they will highlight pathways fundamental to the development of SLE, and may, in turn, become targets for therapeutic intervention.'

The last 6 months has witnessed staggering advances in our understanding of the genetics of systemic lupus erythematosus (SLE). New and important genes have been identified through classical association studies; two independent whole-genome association analyses have been published; we have made inroads into the dissection of complex genetic effects arising from the major histocompatibility complex (MHC); and we are increasingly recognizing that large-scale genomic alteration in the form of copy-number variation (CNV) is an important risk factor for SLE.

It has been known for at least 30 years that SLE has a strong genetic basis, and the results of early family studies are frequently quoted; a disease concordance rate of 2–5% for dizygotic twins, compared with 24–48% for monozygotic twins, and a sibling risk ratio (λ_s) up to 29 [1–3]. The inheritance of SLE is complex and does not follow simple Mendelian rules, which suggests multiple genes contribute to disease susceptibility. Occasional rare but highly penetrant mutations occur, such as complement C1q deficiency and the recently described variation in the TREX1 exonuclease; however, in most cases, SLE risk is likely to be determined by common variants occurring in a number of different genes [4,5]. Each of these genetic variants constitutes only a small additional risk of developing SLE, but if an individual is unfortunate enough to inherit several risk variants together, and be exposed to appropriate environmental triggers, then disease will develop. There are probably 20–50 susceptibility genes conferring a moderate risk of SLE (odds ratio >1.3). There may be many more conferring a smaller risk, but proving their effect would require cohorts of tens of thousands of cases. Identifying SLE susceptibility genes is important because they will highlight pathways fundamental to the development of SLE, and may, in turn, become targets for therapeutic intervention. In addition, genotyping a profile of variants may be of diagnostic or prognostic use.

Two techniques have classically been used to identify susceptibility genes, the linkage study and the association study, and each method has its advantages and disadvantages [6]. This has been extensively reviewed in a recent *Future Rheumatology* article, but, in brief, a linkage scan looks for segregation of particular markers with disease in large families of affected individuals [7]. It has the advantage of being a hypothesis-free technique, but generally has low resolution, with linkage regions often containing several candidate genes. Linkage studies have provided a useful insight into the nature of SLE as a complex trait, but have generally not led to the actual identification of disease candidates; indeed the chromosomal location of the important susceptibility gene *IRF5* (7q23) is not within a recognized linkage region. The association scan looks for a difference in the frequency of a particular genetic marker (typically a single nucleotide polymorphism [SNP]) between an affected individual and an appropriate control; either unrelated matched controls, or unaffected family members. Genetic association is observed only if we type the causal mutation (direct association), or a mutation in strong linkage disequilibrium with it (indirect association), and hence the physical range over which we observe association is small. Until recently, technology has only allowed us to type a handful of markers at once, so by necessity the association study has been hypothesis driven, with genes selected for study on the basis of their biological plausibility. Nonetheless, this technique continues to result in the discovery of major SLE candidate genes.

Key discoveries up to 2007 have been discussed in a number of recent reviews, and include the intracellular signaling molecules IRF5 and STAT4 [7,8]. The early months of 2008 have seen the publication of further genetic associations. Perhaps the most important of these is *ITGAM*, which encodes the α-subunit of $\alpha_M\beta_2$ integrin (also known as complement receptor 3 or Mac-1) [9]. A series of SNPs was associated with SLE in cohorts of European ancestry, but

10.2217/17460816.3.2.103 © 2008 Future Medicine Ltd ISSN 1746-0816

by studying African–American cohorts (where weaker linkage disequilibrium between SNPs was observed) the association signal was narrowed to a single SNP in exon 3. This is a nonsynonymous SNP, although the functional consequence of the argenine to histidine substitution at amino acid position 77 is uncertain. The $\alpha_M\beta_2$ integrin is widely expressed on neutrophils, monocytes/macrophages and dendritic cells (both myeloid/monocyte derived and follicular). It is an important receptor for the C3bi complement fragment and may mediate phagocytocis of C3bi-containing immune complexes and C3bi-coated apoptotic cells. The risk of carrying the associated haplotype was high compared with many in the context of many SLE candidate genes, with an odds ratio across all cohorts of 1.74 ($p = 6.9 \times 10^{-22}$). A second important association is the TNF superfamily gene, *TNFSF4* (also known as OX40 ligand), reported in UK SLE family and case–control cohorts and replicated in the Minnesota (USA) SLE family collection [10]. TNFSF4 is expressed on antigen-presenting cells and when bound to its ligand, TNFRSF4, on $CD4^+$ T cells, it can provide a potent, activating, costimulatory signal. Functional evidence suggests the risk haplotype at *TNFSF4* is associated with increased gene transcription and cell-surface expression of the TNFSF4 protein, which could potentially alter T-cell activation status.

Whole-genome association represents a major leap forward in genetic analysis, and has been made possible by technological advances that allow for the simultaneous typing of 100,000 and more SNPs. This, therefore, combines the strengths and resolution of an association study with the hypothesis-free methodology of a linkage analysis. The results of two independent, large SLE whole-genome association analyses were published in January 2008. The first study from a UK/US collaboration (the International Consortium for Systemic Lupus Erythematosus Genetics [SLEGEN]) used 720 cases and 2337 controls (women of European ancestry), with replication of top hits in two further cohorts of European-ancestry women [11]. Genotyping was by the Illumina HumanHap300 BeadChip, providing successful typing of 317,501 SNPs. Strong association was seen with SNPs in the *MHC* region, and in the already identified *IRF5*. Strong association was also seen in *ITGAM*, complementing the fine-mapping study published side-by-side with this whole-genome analysis. Additional genome-wide significance was observed for markers in *KIAA1542* (a genomic region with homology to a gene encoding an elongation factor), *IRF7* (a member of the Type I interferon signaling pathways) and *PXK* (a Phox homology domain-containing serine–threonine kinase).

'Whole-genome association represents a major leap forward in genetic analysis, and has been made possible by technological advances that allow for the simultaneous typing of 100,000 and more SNPs.'

The second whole-genome association study used 1311 cases and 1783 controls (North Americans of European descent), with replication of top hits in a Scandinavian cohort [12]. Using Illumina HumanHap550 BeadChips, 503,033 SNPs were successfully typed. The strongest genome-wide association was seen for the already-identified *MHC* region, *IRF5* and *STAT4*. Additional strong association was seen for a number of variants surrounding the interval between the B lymphoid tyrosine kinase (*BLK*) gene and the *C8orf13* gene, which are in close proximity to each other, although transcribed in opposite directions. Preliminary evidence suggested that the risk allele for the best associated SNP in this region was associated with decreased expression of *BLK*, but increased expression of *C8orf13*, although further work will be needed to fully delineate this effect. Finally, strong association was again seen for *ITGAM*.

A smaller scan using the 100K Affymetrix chip in 279 Swedish cases and 515 controls led to the identification of the B-cell scaffold protein with ankyrin repeats gene (*BANK1*) as a potential SLE candidate [13]. Follow-up fine-mapping studies in four additional cohorts of European or South American ancestry narrowed the association with three linked SNPs. The risk alleles of these SNPs were associated with the production of a shorter *BANK1* transcript with exon 2 spliced out. It was proposed that one of the associated SNPs altered a branch point sequence, either directly affecting splicing or altering splicing efficiency.

So far we have made only passing reference to the *MHC*. The *MHC* shows the most consistent and strong association with both SLE and a variety of other autoimmune disorders. The *MHC* spans 3.6 Mb of chromosome 6p21.3 and contains over 250 expressed genes. It is divided into the class I and II regions that encode the HLA proteins, and between these lies the class III

region, which contains 58 genes, many of which, including *TNF* and complement components *C2* and *C4*, have an immunomodulatory function. The study of *MHC* genetics has been particularly difficult because of long-range linkage disequilibrium across the whole region. The common haplotype AH8.1 (HLA A1/B8/Cw*07/TNF-308A/C4AQ*0/C4B*1/C2*C/B*fS/DRB1*0301 /DQA1*0501/DQB1*0201) carries a number of genetic variants that have been associated with SLE, including the *TNF* -308A allele, the *C4* complement null allele and *HLA DRB1*0301* (*HLA DR3*) itself [14]. Until late last year, the best studies left this extended haplotype unbroken over a 1-Mb interval, including most of the class II and III regions, making it impossible to prove whether there is one or more than one functional effect from this whole region [15]. At the end of 2007, a study in 314 UK Caucasian SLE trios mapped the *MHC* using four-digit *DRB1* typing combined with a moderate-density SNP map [16]. *HLA DRB1*0301* (but no other *DRB1* type) and 12 other SNPs in the class II and III regions were strongly associated with SLE, but by performing an analysis conditioned on *DRB1*0301*, a second independent signal was arising from a SNP within the class III gene, superkiller viralicidic activity 2-like (*Saccharomyces cerevisiae*) (*SKIV2L*). Using this typing, the *DRB1*0301* haplotype was therefore narrowed to a 180-kb interval containing only three genes, *HLA-B1*, *HLA-DQA1* and *HLA-DQB1*. Reference to the HapMap CEPH dataset suggested that the second signal could be located within a narrow 40-kb interval containing *SKIV2L* and four other genes, complement factor B (*CFB*), RD RNA-binding protein (*RDBP*), dom-3 homolog Z (*Caenorhabditis elegans*) (*DOM3Z*) and serine–threonine kinase 19 (*STK19*). No association was observed between the *TNF* -308A allele and SLE when an analysis conditional on *HLA-DRB1* was performed, suggesting that no independent association arises from this gene. Interestingly, a SNP within *RDBP* was the second most associated marker in the SLEGEN whole-genome analysis, and this latter study also showed evidence for two independent *MHC* effects [11].

By far the most commonly studied human genetic variation is the single-base substitution, known as a SNP, and this has formed the basis of all the studies discussed so far. Interest has naturally focused on SNPs because they are common, they are technologically easy and cheap to study, and understanding the functional consequence of single-nucleotide variation is often relatively easy. Increasingly, other sources of genetic variation are being appreciated, particularly the large-scale duplication or deletion of DNA fragments, known as CNV. It has, of course, been known for some time that genes may exist as duplicated copies, for example, the α-globin gene. However, recent technological advances, particularly the development of the comparative genomic hybridization (CGH) array, have facilitated genome-wide screening for CNV, showing it to be a major source of variation, encompassing up to 12% of the genome [17].

The first suggestion that CNV may be important in the context of lupus was a report in 2006 of CNV at the low-affinity Fc-γ receptor locus *Fcgr3* in the rat and its orthologous human counterpart *FCGR3B* [18]. In a rat model of immune-mediated glomerulonephritis, susceptibility to nephritis was increased by low gene copy number, while in humans, an association between low copy number and lupus nephritis was observed. In a larger follow-up study it was established that the diploid human genome carries between zero and six copies of *FCGR3B*. Carrying less than two copies is a major risk factor for both SLE nephritis (odds ratio: 2.43) and SLE without renal involvement (odds ratio: 2.21), placing *FCGR3B* CNV among the most significant genetic risks for SLE identified to date [19].

> 'It is possible that with these recent advances we have identified all the reasonably common and penetrant mutations predisposing to SLE risk in Caucasian populations.'

Copy-number variation of the *C4* component of complement has also been described as a risk factor for SLE, with low copy number conferring increased risk and high copy number being protective [20]. However, some caution needs to be taken with the interpretation of these results, because the *C4* genes lie within the *MHC* region of extended linkage disequilibrium discussed above. It remains to be determined whether *C4* copy number is in linkage disequilibrium with adjacent single-base mutations; hence, it is quite possible that the observed association is an indirect one. Finally, from mouse studies it has been shown that the SLE-enhancing effects of the *Yaa* region in the BXBS murine model of lupus may, in part, be due to CNV of the Toll-like receptor gene, *Tlr7* [21]. A subsequent study in human SLE confirmed the existence of *TLR7* CNV, but

did not observe an association with SLE, suggesting a more complex role in human lupus [22].

It is possible that with these recent advances we have identified all the reasonably common and penetrant mutations predisposing to SLE risk in Caucasian populations. It is interesting that these genes suggest that multiple immune mechanisms are involved in SLE pathogenesis, particularly the early binding and phagocytosis of immune complexes and apoptotic cells (*ITGAM*, *FCGR3B*, *FCGR3A*, *FCGR2A*, and possibly *MBL* and *CRP*), T-cell activation (*HLA-DRB1*, *TNFSF4*, *CTLA4-ICOS* and *PTPN22*) and type 1 interferon signaling pathways (*IRF5* and *IRF7*). For these genes, future work will move away from purely genetic study into a functional phase as we attempt to understand how dysregulation of these pathways can predispose to autoimmunity. There are also genes highlighted by the genome-wide scans that have either no or poorly understood immunological function (e.g., *KIAA1542* and *PKX*), which will provide a considerable challenge for future study. There is also a great deal more genetic work to be done. The SNP coverage provided by the existing whole-genome scans, while impressive, is certainly not enough to provide a refined genetic signal, and so fine-mapping of the newly associated genes is required to identify the functional variants. Understanding how these variants alter gene expression and predispose to disease is also a challenge. Most associated variants identified so far do not affect the structure of the final protein product, so they must affect gene expression in some other way, perhaps by altering transcription-factor binding or by affecting mRNA stability or splicing. For some genes, such as *IRF5*, there appears to be more than one functional mechanism, adding to the difficulty in study design [23–25].

'The SNP coverage provided by the existing whole-genome scans, while impressive, is certainly not enough to provide a refined genetic signal, and so fine-mapping of the newly associated genes is required to identify the functional variants.'

Technology is advancing rapidly, and within the next few years it may become possible to type every known SNP in a single analysis. The number of individuals included in studies will also grow as investigators expand their cohorts, and further collaborative efforts are made. These two factors will undoubtedly result in the discovery of more genetic associations, as we will have increased both the power to detect modest risk effects and the density of SNP typing to uncover associations missed in the current scans. Furthermore, we will begin to answer more sophisticated questions, such as whether specific genetic loci are associated with specific SLE subphenotypes, or whether there are epistatic interactions between loci.

Most genetic studies to date have been conducted in populations of European ancestry, which represents only a small proportion of the total human population. Further work is required to determine genetic effects in other ethnic groups. Data from the HapMap project reveal quite marked differences in SNP frequency and haplotype structure between populations of different ancestry. The prevalence of SLE in Caucasian populations is actually quite low compared with many other ethnic groups, and genetic analysis of higher-risk groups may reveal novel genetic risk factors. However, human populations have not evolved in complete isolation from each other. Over the last 500 years in particular, human transport has allowed for increased population movement and, inevitably, interbreeding. This results in the problem of genetic admixture, whereby new genetic variants appear very recently on a background genome that may have had many tens of thousands of years of independent evolution prior to this. Nowhere is this effect more prominent than in African–American and African–Caribbean populations, where 10–30% of the genome has a European ancestry, on the background of a largely West African genome [26]. Mouse studies have suggested an interaction between SLE susceptibility genes and genetic background, and it appears that this may also be the situation in admixed human populations. Indeed, the incidence of SLE in African–American and African–Caribbean populations is particularly high, and there is evidence that this incidence may be influenced by the level of genetic admixture. The identification of ancestry-informative markers is therefore important for future genetic study, both because it will allow for better control groups, matched with patients on the basis of their population admixture, but also because an understanding of the mechanism by which susceptibility genes and genetic background interact may tell us something more fundamental about disease pathogenesis.

Financial & competing interests disclosure

The authors have 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.

Bibliography

1. Alarcon-Segovia D, Alarcon-Riquelme ME, Cardiel MH *et al.*: Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. *Arthritis Rheum.* 52, 1138–1147 (2005).
2. Deapen D, Escalante A, Weinrib L *et al.*: A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis Rheum.* 35, 311–318 (1992).
3. Hochberg MC: The application of genetic epidemiology to systemic lupus erythematosus. *J. Rheumatol.* 14, 867–869 (1987).
4. Lee-Kirsch MA, Gong M, Chowdhury D *et al.*: Mutations in the gene encoding the 3´–5´ DNA exonuclease TREX1 are associated with systemic lupus erythematosus. *Nat. Genet.* 39, 1065–1067 (2007).
5. Walport MJ, Davies KA, Botto M: C1q and systemic lupus erythematosus. *Immunobiology* 199, 265–285 (1998).
6. Lander ES, Schork NJ: Genetic dissection of complex traits. *Science* 265, 2037–2048 (1994).
7. Sestak AL: Genetics of systemic lupus erythematosus. *Future Rheumatol.* 3(1), 51–57 (2008).
8. Rhodes B, Vyse TJ: General aspects of the genetics of SLE. *Autoimmunity* 40, 550–559 (2007).
9. Nath SK, Han S, Kim-Howard X *et al.*: A nonsynonymous functional variant in integrin-α(M) (encoded by *ITGAM*) is associated with systemic lupus erythematosus. *Nat. Genet.* 40(2), 152–154 (2008).
10. Graham DS, Graham RR, Manku H *et al.*: Polymorphism at the TNF superfamily gene *TNFSF4* confers susceptibility to systemic lupus erythematosus. *Nat. Genet.* 40, 83–89 (2008).
11. International Consortium for Systemic Lupus Erythematosus Genetics; Harley JB, Alarcón-Riquelme ME, Criswell LA *et al.*: Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in *ITGAM*, *PXK*, *KIAA1542* and other loci. *Nat. Genet.* 40(2), 204–210 (2008).
12. Hom G, Graham RR, Modrek B *et al.*: Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM–ITGAX. *N. Engl. J. Med.* 358(9), 900–909 (2008).
13. Kozyrev SV, Abelson AK, Wojcik J *et al.*: Functional variants in the B-cell gene *BANK1* are associated with systemic lupus erythematosus. *Nat. Genet.* 40(2), 211–216 (2008).
14. Fernando MMA, Stevens CR, Walsh EC *et al.*: Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet.* (2008) (In Press).
15. Graham RR, Ortmann WA, Langefeld CD *et al.*: Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus. *Am. J. Hum. Genet.* 71, 543–553 (2002).
16. Fernando MM, Stevens CR, Sabeti PC *et al.*: Identification of two independent risk factors for lupus within the MHC in United Kingdom families. *PLoS Genet.* 3, E192 (2007).
17. Redon R, Ishikawa S, Fitch KR *et al.*: Global variation in copy number in the human genome. *Nature* 444, 444–454 (2006).
18. Aitman TJ, Dong R, Vyse TJ *et al.*: Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* 439, 851–855 (2006).
19. Fanciulli M, Norsworthy PJ, Petretto E *et al.*: *FCGR3B* copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat. Genet.* 39, 721–723 (2007).
20. Yang Y, Chung EK, Wu YL *et al.*: Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am. J. Hum Genet.* 80, 1037–1054 (2007).
21. Pisitkun P, Deane JA, Difilippantonio MJ, Tarasenko T, Satterthwaite AB, Bolland S: Autoreactive B cell responses to RNA-related antigens due to *TLR7* gene duplication. *Science* 312, 1669–1672 (2006).
22. Kelley J, Johnson MR, Alarcon GS, Kimberly RP, Edberg JC: Variation in the relative copy number of the *TLR7* gene in patients with systemic lupus erythematosus and healthy control subjects. *Arthritis Rheum.* 56, 3375–3378 (2007).
23. Graham DS, Wong AK, McHugh NJ, Whittaker JC, Vyse TJ: Evidence for unique association signals in SLE at the *CD28–CTLA4–ICOS* locus in a family-based study. *Hum. Mol. Genet.* 15, 3195–3205 (2006).
24. Graham RR, Kozyrev SV, Baechler EC *et al.*: A common haplotype of interferon regulatory factor 5 (*IRF5*) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat. Genet.* 38, 550–555 (2006).
25. Graham RR, Kyogoku C, Sigurdsson S *et al.*: Three functional variants of IFN regulatory factor 5 (*IRF5*) define risk and protective haplotypes for human lupus. *Proc. Natl Acad. Sci. USA* 104, 6758–6763 (2007).
26. Wassel Fyr CL, Kanaya AM, Cummings SR *et al.*: Genetic admixture, adipocytokines, and adiposity in black Americans: the Health, Aging, and Body Composition study. *Hum. Genet.* 121, 615–624 (2007).

Affiliations

- *Benjamin Rhodes, Arthritis Research Campaign Clinical Research Fellow*
  Imperial College London, Section of Molecular Genetics and Rheumatology, Du Cane Road, London, W12 0NN, UK
- *Timothy J Vyse, Professor of Rheumatology and Wellcome Trust Senior Research Fellow*
  Imperial College London, Section of Molecular Genetics and Rheumatology, Faculty of Medicine, Du Cane Road, London, W12 0NN, UK
  Tel.: +44 208 383 2339
  Fax: +44 208 383 2379
  t.vyse@imperial.ac.uk