Editorial

Genome-wide association studies in systemic lupus erythematosus: a perspective

Deborah S Cunninghame Graham

Molecular Genetics and Rheumatology, Fifth Floor Commonwealth Building, Room 5N8b, Division of Medicine, Imperial College, Hammersmith Campus, Du Cane Rd, London W12 ONN

Corresponding author: Deborah S Cunninghame Graham, deborah.cunninghame-graham@imperial.ac.uk

Published: 9 July 2009
This article is online at http://arthritis-research.com/content/11/4/119

Arthritis Research & Therapy 2009, **11:**119 (doi:10.1186/ar2739)

See related research by Suarez-Gestal *et al*., http://arthritis-research.com/content/11/3/R69

Abstract

Genome-wide association studies (GWAS) have been shown to be a powerful way of identifying novel susceptibility genes in systemic lupus erythematosus (SLE), as demonstrated by a series of publications in the past year. Lupus has been a late-comer to the GWAS community, being preceded by success stories for the GWAS approach in other autoimmune diseases, including type I diabetes, ankylosing spondylitis, rheumatoid arthritis, Crohn's disease and ulcerative colitis. The paper by Suarez-Gestal and colleagues seeks to exploit the wealth of data available from a total of four GWAS in SLE, three in European-American populations and one in a Swedish population. The authors describe replication of ten lupus susceptibility alleles in a Spanish SLE case-control study.

Suarez-Gestal and colleagues [1] selected single variants from either systemic lupus erythematosus (SLE) genome-wide association studies (GWAS) or large candidate-gene-based association studies. Three of the markers tested were identified in all three European-American GWAS. Four more were discovered in a single GWAS; either the International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN) studies [2-4] or a Swedish study [5]. The remaining three variants were found in a smaller candidate-gene or regional association study. This study by Suarez-Gestal and colleagues represents an important step forward in SLE genetics because it uses a Spanish population to confirm most of the predominantly European-American associations. It appears that the SLE genetics community have learned valuable lessons in GWAS design from the experiences of other complex trait geneticists where replication studies were less consistent and yielded a multiplicity of false positives and negatives [6]. The lack of replication in other complex diseases may reflect a number of flaws in their GWAS design, including inadequate sample sizes, population stratification, poor quality control of genotyping and substandard matching of cases and controls [7]. Independent replication is an important step in endorsing novel susceptibility genes because it both confirms the validity of the GWAS design as well as verifies the contribution of a given locus to disease pathogenesis.

Suarez-Gestal and colleagues confirmed nine of the ten previous associations. This paper represents a major step forward for SLE because each of the loci containing these single nucleotide polymorphisms (SNPs) can now be examined in more detail to determine biological mechanisms and to gain a greater understanding of their roles in the pathogenic process of SLE. The remaining variant, in the coding region of *LY9*, a gene that forms part of the signalling-lymphocyte-activation molecule (*SLAM*) locus [8], did not replicate. Possible reasons for this are population-specific differences in allele frequency or haplotype structure between the northern and southern European populations [9] or, alternatively, too little power to detect it. The advent of HAPMAP phase 3, with increased numbers for the original four populations, increases the reliability for the tag-SNP selection as well as adds an additional seven population groups for better correlation between the population used for tag-SNP selection and a wider number of study cohorts. This would minimise the effect of population-specific differences in haplotype structure by allowing for the presence of multiple risk haplotypes carrying a particular risk allele(s) while minimising the type I or II errors due to population-specific differences in allele frequency for single variants.

The ideal replication study design would be to perform a meta-analysis of the existing SLE GWAS data and then select the variants showing the strongest association. Suarez-Gestal and colleagues chose four variants, each identified by a single GWAS, since there is no publically available meta-analysis for all the current GWAS. The lack of

GWAS = genome-wide association study; SLE = systemic lupus erythematosus; SNP = single nucleotide polymorphism.

reproducibility between GWAS may reflect an overall lack of power for each individual study, since each published GWAS was of intermediate size (1,100 cases or less) and there was some sample duplication between them. The other factor that may contribute to the diversity of strongly associated SNPs between different studies is the heterogeneity of the disease phenotype, which may reflect the underlying genetic heterogeneity. Consequently, there is clear need for a much larger additional GWAS, in both Europeans and non-Europeans, with clearly defined clinical criteria and at a much greater density of markers. This scale of experiment, which is proving successful for other complex diseases, would offset the loss of power associated with the heterogeneity of lupus.

In complex autoimmune diseases such as lupus, a large number of variants are expected to make a small contribution to the overall genetic risk. Furthermore, there may be multiple susceptibility alleles within a given gene, so that the global risk for a given locus will be a combination of the individual risks for each susceptibility allele in that gene. Further complexity arises because not only do particular individuals carry different combinations of risk alleles, but the genotyping chips do not carry every susceptibility allele for a given locus. Interpretation of whether a particular locus is associated with disease may, therefore, depend on the linkage disequilibrium between the genotyped SNP and the functional allele(s). Patterns of linkage disequilibrium between associated SNPs in different studies will have to be taken into consideration before either claiming replication or the lack of it.

The replication of the variants in the paper by Suarez-Gestal and colleagues therefore represents the tip of the iceberg because it has confirmed the association of a number of common risk alleles of moderate disease risk identified from intermediate-sized GWAS. To test rarer mutations (minor allele frequency <0.1%), such as those that will be generated by the 1000 Genomes Project, it may be necessary to both design custom chips, since even the latest generation of GWAS chips [10] will not carry rarer mutations, and also genotype them in very large populations. However, we anticipate that there will be a larger number of genes with a smaller effect size (odds ratio 1.1 to 1.2), so that we will need an increased number of samples in the study cohort to gain sufficient power to find a significant association. Hence, GWAS funded by the Wellcome Trust Case Control Consortium include studies in the range of 5,000 to 10,000 samples; a study of this size is needed in SLE. These larger population sizes will give sufficient power for the analysis of sub-phenotypes. Nevertheless, the data presented by Suarez-Gestal and colleagues provide a core series of independently validated loci that, together with additional targets generated by larger GWAS, can be used to piece together the key pathways involved in lupus pathogenesis, with each pathway constructed of a number of interacting proteins making an individual contribution to disease susceptibility.

Competing interests

The author declares that they have no competing interests.

References

1. Suarez-Gestal M, Calaza M, Pullmann R, Ros JO, Sebastiani GD, Ruzickova S, Santos MJ, Papasteriades C, Marchini M, Skopouli FN, Saurez A, Blanco FJ, D'Alfonso S, Bijl M, Carreira P, Migliaresi S, Gonzalez A; for the European Consortium of SLE DNA Collections: **Replication of recently identified systemic lupus erythematosus genetic factors: a case–control study.** *Arthritis Res Ther* 2009, **11:**R69.
2. Nath SK, Guthridge JM, Cobb BL, Mirel DB, Marion MC, Williams AH, Divers JD, Wang W, Frank SG, Namjou B, Gabriel SB, Lee AT, Gregersen PK, Behrens TW, Taylor KE, Fernando M, Zidovetzki R, Gaffney PM, Edberg JC, Rioux JD, Ojwang JO, James JA, Merrill JT, Gilkeson GS, Seldin MF, Yin H, Baechler EC, Li QZ, Wakeland EK, Bruner GR, *et al.*: **Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci.** *Nat Genet* 2008, **40:**152-154.
3. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, Lee AT, Chung SA, Ferreira RC, Pant PV, Ballinger DG, Kosoy R, Demirci FY, Kamboh MI, Kao AH, Tian C, Gunnarsson I, Bengtsson AA, Rantapaa-Dahlqvist S, Petri M, Manzi S, Seldin MF, Ronnblom L, Syvanen AC, Criswell LA, Gregersen PK, Behrens TW: **Association of systemic lupus erythematosus with *C8orf13-BLK* and *ITGAM-ITGAX*.** *N Engl J Med* 2008, **358:** 956-961.
4. Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM, Burtt NP, Guiducci C, Parkin M, Gates C, Plenge RM, Behrens TW, Wither J, Rioux JD, Fortin PR, Cunninghame Graham DS, Wong AK, Vyse TJ, Daly MJ, Altshuler D, Moser KL, Gaffney PM: **A genome-wide association scan identifies tumour necrosis factor alpha inducible protein 3 (TNFAIP3/A20) as a susceptibility locus for systemic lupus erythematosus.** *Nat Genet* 2008, **40:**1059-1061.
5. Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, Sanchez E, Gunnarsson I, Svenungsson E, Sturfelt G, Jonsen A, Truedsson L, Pons-Estel BA, Witte T, D'Alfonso S, Barizzone N, Danieli MG, Gutierrez C, Suarez A, Junker P, Laustrup H, Gonzalez-Escribano MF, Martin J, Abderrahim H, Alarcon-Riquelme ME: **Functional variants in the B-cell gene *BANK1* are associated with systemic lupus erythematosus.** *Nat Genet* 2008, **40:**211-216.
6. Ioannidis JPA: **Non-replication and inconsistency in the genome-wide association setting**. *Hum Hered* 2007, **64:**203-213.
7. Neale BM, Purcell S: **The positives, protocols and perils of genome-wide association.** *Am J Med Genet* 2008, **147B:**1288-1294.
8. Cunninghame Graham DS, Vyse TJ, Fortin PR, Montpetit A, Cai Y-C, Lim S, McKenzie T, Farwell L, Rhodes B, Chad L, CaNIOS GenES Investigators, Hudson TJ, Terhorst C, Sharpe AH, Greenwood CMT, Wither J, Rioux JD: **Association of *LY9* in UK and Canadian SLE Families.** *Genes Immun* 2008, **9:**1-10.
9. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. **Principal components analysis corrects for stratification in genome-wide association studies.** *Nat Genet* 2006, **38:** 904-909.
10. Li M, Li C, Guan W: **Evaluation of coverage variation of SNP chips for genome-wide association studies.** *Eur J Hum Genet* 2008, **16:**635-643.