

EDITORIAL

Shared susceptibility variations in autoimmune diseases: a brief perspective on common issues

Genes and Immunity (2009) 10, 1–4; doi:10.1038/gene.2008.92

Three articles^{1–3} in the present issue of the journal expand on an emerging theme in autoimmunity genetics, the overlap in genetic effects of common variants in disparate diseases. The articles also raise other common issues in our approach to both autoimmune genetics and the genetics of other complex diseases: overlap in cases and controls, population stratification, correction for multiple comparisons (thresholds for significance) and combined versus individual publications.

First, the practical issue of overlap between studies. Multiple studies in autoimmune diseases conducted over the last year have used both case and control sets that in part overlap. This phenomenon while not new to type 1 diabetes (T1DM), where international efforts have led to widely available patient samples, has been extended to recent high-profile studies including rheumatoid arthritis (RA),^{4–7} systemic lupus erythematosus (SLE),^{8,9} and in the present issue of the journal, multiple sclerosis (MS). It is often difficult to combine such studies even when there is open communication among research groups due to the practical necessity of appropriately distributing credit to individual investigators and groups. However, such overlap affects interpretation of the results, that is, the concordance of results is not an independent confirmation and it may be difficult for outside groups to combine results in, for example, meta-analyses. If the genotyping results can be combined (that is, they are using the same allele genotyping definitions), the statistical issues posed by this problem can be readily addressed. While the combined single publications may be preferable for journals and the readers in the scientific community, we believe that the present reports are worthy of three individual articles, because they highlight and report results from different populations and different overlaps in genetic etiology among autoimmune diseases.

The very large overlap between the IMSGC study² and that from Hafler *et al.*¹ (see Table 1 footnote in IMSGC study), indicates that the results of these two studies cannot be considered as replicates. In fact, the combined results for the two groups showed only a marginal gain in significance for the CD226 SNP (single-nucleotide polymorphism) rs763361 (overall P -value $1.1E^{-08}$ compared to $5.4E^{-08}$ in the IMSGC study²) (S Sawcer, personal communication). In contrast, the two groups (IMSGC² and Zoledziewska *et al.*³) reporting the association of CLEC16A polymorphisms did not have overlapping subjects. For these studies, the cumulative results were assessed by combining P -values, as the SNPs analyzed did not overlap. The combined P -value $5.0E^{-19}$ using Fishers method (S Sawcer, personal communication) does provide additional confidence in this association. Another approach that can be taken for

combining results from independent studies that use different SNPs is to infer SNP genotypes that are common to both studies²⁸ but this may be difficult when a unique population (in this case the Sardinian population) may have different undefined haplotypes, making imputation problematic.

Two of the present studies raise another issue regarding overlap: the use of a single control group for two different disease studies. In the studies reported by Hafler *et al.*¹ and the IMSGC,² the population control subjects overlap with those used in the study of T1DM. Similarly, use of the same common controls may be a partial concern in evaluating the strength of association between STAT4 and primary Sjogrens syndrome, where the controls overlapped with those used in RA and SLE studies.²⁹ These studies raise the question of how can we adjust our thresholds for significant results to account for these overlaps in which the controls for multiple studies are not independent? This issue may become increasingly problematic as large genotyping sets become publicly available and are used in many studies. In fact, at present several thousand population sets are potentially available from multiple sources (for example, iControlDB (<http://www.illumina.com/pages.ilmn?ID=231>) and dbGaP (<http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login>)). If the population groups are large enough then perhaps the issue might become moot, as the genotype frequencies may accurately reflect the population and have little fluctuation when issues of population structure and substructure are addressed. For this approach to be valid, however, studies should begin to adopt a common means for scoring individuals according to genetic background. For example, the recent publication of Nelson *et al.*³⁰ described the development of a large collection of controls that identified eight eigenvectors for defining major ethnic ancestries as well as more minor ancestry, such as the north–south European cline. Variation within Europe can be further subdivided,³¹ and ongoing studies are most likely to provide additional reference populations and ancestry marker sets when genome-wide studies are not being performed.³² In the short term, some care will be necessary in considering these issues. If comparison of shared and non-shared control population genotypes shows no substantive difference, there is some assurance that the results do not just reflect some demographic or genotyping artifact among certain control subject collections. Thankfully, for the present studies, similar results were obtained from multiple independent sample groups lending support that the findings are real.

Another analytic issue raised by these studies is the often discussed question of appropriate statistical thresholds. For a ‘candidate’ SNP(s) study what is a reasonable P -value. For example, how do we assess the P -value of $6.7E^{-5}$ reported in the Sardinian study? The Hapmap consortium³³ found that within each 500 kb

Table 1 Partial list of genetic associations implicated in multiple autoimmune diseases^a

Gene	Association and estimation of odds ratios in European populations ^b				Selected references
	MS	T1DM	RA	SLE	
PTPN22	No	Yes 2.0	Yes 1.6	Yes 1.4	10–14
STAT4	No	Possible 1.6	Yes 1.3	Yes 1.5	4,8,9,13,15,16
CD226	Yes 1.13	Yes 1.3	No evidence	No evidence	1,2,14
CLEC16A	1.2	Yes 1.25	No evidence	No evidence	2,3,14
TNFAIP3	No evidence	No evidence	Yes 1.3	Yes 2.2	17–20
IL2RA	Yes 1.25	Yes 1.3	Suggestive 1.3	No evidence	21–23
CTLA4	No	Yes 1.2	Yes 1.2	Weak	24–27

Abbreviations: MS, multiple sclerosis; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; T1DM, type 1 diabetes.

^aThis is a partial list of genes and autoimmune disease associations. For many of these genes additional autoimmune diseases have evidence for association.

^bThe odds ratios are based on our interpretation of the literature using a combination of published meta-analyses and large studies (for example, genome-wide association tests). In some cases, the associations remain controversial (for example, PTPN22 in SLE), as some large studies have not shown convincing evidence for association when population substructure is adequately controlled.⁸

region, there were the equivalent of 150 independent allele-based tests in Caucasian and Asian populations and about 350 independent tests in Yorubans. Correcting for the number of independent tests leads to a thresholds of around $5E^{-8}$ for Caucasian and Asian populations and $2.5E^{-8}$ for Yorubans. These highly conservative thresholds may be appropriate for genome-wide analyses but seem excessively conservative when there is excellent motivation to test specific SNPs. Adjusting for the total number of SNPs that have been tested would be one easily applied alternative approach to limit excess false positives, but extracting the total number of SNPs that have been tested from investigators can be difficult, in part, because of shifting priorities in laboratories. This problem is further compounded by studies performed by multiple different laboratories and the bias towards only reporting positive results, which particularly affects smaller studies.³⁴ Applying a false-discovery paradigm is an alternative approach, which seems appealing, although this approach has not yet been widely adopted^{35,36} and studies should report false discovery rates along with significance levels. Another alternative approach would be to weight the previous information and obtain a posterior probability of association allowing for the cost of false-negative and false-positive discoveries, using a Bayesian approach.³⁷ This approach has an advantage of incorporating uncertainty about the reliability of previous information.

Finally, how does the overlap in susceptibility alleles in different autoimmune diseases define risk or insight into the pathogenesis of myriad diseases? A wealth of studies has clearly shown that allelic variations at the same or closely linked loci are critical genetic risk factors for multiple autoimmune diseases. A partial list of the more cogent overlaps for non-major histocompatibility 'genes' is shown in Table 1. For some, the same haplotype or even same putative causative amino acid variation appears to be implicated (for example, haplotypes for STAT4 in SLE and RA, and the PTPN22 Arg620Trp variation in T1DM and RA). However, for other implicated genes (or small genomic intervals) the story is much less clear. This includes TNFAIP3, where it appears that at least one component of the RA risk is different from the SLE risk haplotype.^{17–19}

The shared risk factors appear to have unique overlaps between different autoimmune diseases. For example, the PTPN22 Arg620Trp that is shared between several autoimmune diseases including T1DM and RA has been shown not to be a risk factor in MS, whereas in the present studies CD226 and CLEC16A variations are shared risk factors between MS and T1DM. Interpreting these relationships may also be further complicated by ethnic differences in the frequency and/or risk of particular variants for these diseases.

Moving forward, we will need much clearer definition of the risk associated with each of the genetic variations in different autoimmune diseases within different population groups. This information can also further draw on the emerging integration of expression information and molecular pathways. A combination of an understanding of how genetic variation affects these pathways will most likely provide a clearer understanding of pathophysiology of these complex diseases. In this regard, a recent review providing a diagram of potential synthesis of SLE and RA molecular mechanisms/pathways may be instructive.²³ Arguably, even the modest effect (amount of genetic variation explained) of many of the emerging genetic risk factors will be critical in our understanding of the etiopathogenesis of these diseases. Some caution is of course needed, as in many cases, the actual gene(s) affected by specific sequence or haplotype variations is not yet clear. This may be the case for the STAT4 association in RA and SLE, where the present paucity of functional information has not excluded that the associated sequence variation could be affecting the transcription of the closely linked STAT1 gene even though the responsible haplotype resides within the STAT4 genomic region.

The present studies in the journal provide additional information that may facilitate an understanding of the intersection of molecular pathways resulting in MS and T1DM. However, at present the limited understanding of the role of the CD226 and CLEC16A variants precludes strong speculations. For CD226, the Gly307Ser mutation can explain the present associations and may allow focused experimental studies to determine the altered mechanism in T cells, B cells or natural killer cells that presumably predisposes individuals for MS and/or

T1DM. For CLEC16A, the lack of a clear functionally relevant variation does not at present exclude the possibility that the functional SNP(s) could, in possible analogy to the situation discussed for STAT4, be important in the regulation of the closely located MHC2TA gene. Thus, the difficult step of defining the functional mechanisms by which the variations modify risk is a critical bottleneck in developing cogent hypothesis to explain the common roles of specific gene variations in immunity. However, the common genes for autoimmune diseases will most likely provide important insights into the complex interactions that result in aberrant immunologic activity causing autoimmune disease, and this is most likely to be a recurring theme in Genes and Immunity.

MF Seldin¹ and CI Amos²

¹Rowe Program in Genetics, Departments of Biochemistry and Medicine, University of California Davis, Davis, CA, USA and ²Department of Epidemiology, MD Anderson Cancer Center, University of Texas, Houston, TX, USA

References

- Hafler JP, Maier LM, Cooper JD, Plagnol V, Hinks A, Simmonds MJ *et al.* CD226 Gly307Ser is associated with multiple autoimmune diseases. *Genes Immun* 2009; **10**: 5–10.
- IMSGC T. The expanding genetic overlap between multiple sclerosis and type 1 diabetes. *Genes Immun* 2009; **10**: 11–14.
- Zoledziewska M, Costa G, Pitzalis M, Cocco E, Melis C, Moi L *et al.* Variation within the CLEC16A gene shows consistent disease association with both multiple sclerosis and type 1 diabetes in Sardinia. *Genes Immun* 2009; **10**: 15–17.
- Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW *et al.* STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N Engl J Med* 2007; **357**: 977–986.
- Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B *et al.* TRAF1-C5 as a risk locus for rheumatoid arthritis—a genome-wide study. *N Engl J Med* 2007; **357**: 1199–1209.
- Chang M, Rowland CM, Garcia VE, Schrodi SJ, Catanese JJ, van der Helm-van Mil AH *et al.* A large-scale rheumatoid arthritis genetic study identifies association at chromosome 9q33.2. *PLoS Genet* 2008; **4**: e1000107.
- Kurreeman FA, Rocha D, Houwing-Duistermaat J, Vrijmoet S, Teixeira VH, Migliorini P *et al.* Replication of the tumor necrosis factor receptor-associated factor 1/complement component 5 region as a susceptibility locus for rheumatoid arthritis in a European family-based study. *Arthritis Rheum* 2008; **58**: 2670–2674.
- Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S *et al.* Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N Engl J Med* 2008; **358**: 900–909.
- Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL *et al.* Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci. *Nat Genet* 2008; **40**: 204–210.
- Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC *et al.* A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 2004; **75**: 330–337.
- Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M *et al.* A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat Genet* 2004; **36**: 337–338.
- Lee YH, Rho YH, Choi SJ, Ji JD, Song GG, Nath SK *et al.* The PTPN22 C1858T functional polymorphism and autoimmune diseases—a meta-analysis. *Rheumatology (Oxford)* 2007; **46**: 49–56.
- Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL *et al.* Risk alleles for multiple sclerosis identified by a genome-wide study. *N Engl J Med* 2007; **357**: 851–862.
- Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V *et al.* Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 2007; **39**: 857–864.
- Zervou MI, Mamoulakis D, Panierakis C, Boumpas DT, Goulielmos GN. STAT4: a risk factor for type 1 diabetes? *Hum Immunol* 2008.
- Martinez A, Varade J, Marquez A, Cenit MC, Espino L, Perdignes N *et al.* Association of the STAT4 gene with increased susceptibility for some immune-mediated diseases. *Arthritis Rheum* 2008; **58**: 2598–2602.
- Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J *et al.* Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 2007; **39**: 1477–1482.
- Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM *et al.* Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat Genet* 2008.
- Musone SL, Taylor KE, Lu TT, Nititham J, Ferreira RC, Ortmann W *et al.* Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. *Nat Genet* 2008.
- Thomson W, Barton A, Ke X, Eyre S, Hinks A, Bowes J *et al.* Rheumatoid arthritis association at 6q23. *Nat Genet* 2007; **39**: 1431–1433.
- Weber F, Fontaine B, Cournu-Rebeix I, Kroner A, Knop M, Lutz S *et al.* IL2RA and IL7RA genes confer susceptibility for multiple sclerosis in two independent European populations. *Genes Immun* 2008; **9**: 259–263.
- Brand OJ, Lowe CE, Heward JM, Franklyn JA, Cooper JD, Todd JA *et al.* Association of the interleukin-2 receptor alpha (IL-2Ralpha)/CD25 gene region with Graves' disease using a multilocus test and tag SNPs. *Clin Endocrinol (Oxf)* 2007; **66**: 508–512.
- Qu HQ, Montpetit A, Ge B, Hudson TJ, Polychronakos C. Toward further mapping of the association between the IL2RA locus and type 1 diabetes. *Diabetes* 2007; **56**: 1174–1176.
- Greve B, Simonenko R, Illes Z, Peterfalvi A, Hamdi N, Mycko MP *et al.* Multiple sclerosis and the CTLA4 autoimmunity polymorphism CT60: no association in patients from Germany, Hungary and Poland. *Mult Scler* 2008; **14**: 153–158.
- Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G *et al.* Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 2003; **423**: 506–511.
- Plenge RM, Padyukov L, Remmers EF, Purcell S, Lee AT, Karlson EW *et al.* Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. *Am J Hum Genet* 2005; **77**: 1044–1060.
- 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* 2006; **15**: 3195–3205.
- Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007; **316**: 1341–1345.
- Korman BD, Alba MI, Le JM, Alevizos I, Smith JA, Nikolov NP *et al.* Variant form of STAT4 is associated with primary Sjogren's syndrome. *Genes Immun* 2008; **9**: 267–270.

- 30 Nelson MR, Bryc K, King KS, Indap A, Boyko AR, Novembre J *et al.* The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am J Hum Genet* 2008; **83**: 347–358.
- 31 Tian C, Plenge RM, Ransom M, Lee A, Villoslada P, Selmi C *et al.* Analysis and application of European genetic substructure using 300K SNP information. *PLoS Genet* 2008; **4**: e4.
- 32 Liu MF, Lin LH, Weng CT, Weng MY. Decreased CD4+CD25+bright T cells in peripheral blood of patients with primary Sjogren's syndrome. *Lupus* 2008; **17**: 34–39.
- 33 International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005; **437**: 1299–1320.
- 34 Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet* 2001; **29**: 306–309.
- 35 Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet* 2006; **7**: 781–791.
- 36 Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA* 2008; **299**: 1335–1344.
- 37 Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* 2007; **81**: 208–227.
- 38 Xavier RJ, Rioux JD. Genome-wide association studies: a new window into immune-mediated diseases. *Nat Rev Immunol* 2008; **8**: 631–643.