Editorial

Powered for Success: Considerations for Using the Candidate Gene Approach in Rheumatic Diseases in the Post-genomics Era

Genetic factors play a substantive role in the susceptibility of ankylosing spondylitis (AS), as evidenced by its high heritability (> 90%) and considerable recurrence risk ratio ($\lambda s = 50–80$)[1]. Powered by 3 AS genome-wide association studies (GWAS), 26 genetic loci have reached genome-wide significance, accounting for about 25% of the overall heritability[1,2,3]. The overwhelming majority of the genetic contribution is provided by the *HLA-B27* variant.

The most efficient method for gene identification at present appears to be association-based studies, which integrate genetic and epidemiological principles. Association-based studies have benefited immensely from the characterization of a large number of single-nucleotide polymorphism (SNP) markers, linkage disequilibrium (LD) data from the HapMap project, and more recently, the 1000 Genomes Project (www.1000genomes.org/), and the development of high-throughput genotyping technologies. The candidate gene approach focuses on associations between genetic variation within prespecified genes of interest and disease phenotypes. The selection of candidate genes is most often based on *a priori* knowledge of the proposed gene function on a particular trait. In this issue of *The Journal*, Nossent, *et al* present results of a cross-sectional and longitudinal study examining the relationship between 2 tumor necrosis factor-α (TNF-α) gene promoter polymorphisms with serum TNF levels and clinical outcomes, in a white Norwegian AS cohort[4].

Multiple lines of evidence support a key role for TNF-α in AS pathogenesis. Despite the lack of consistent association between TNF-α promoter polymorphisms and AS susceptibility[5,6], it is conceivable that variants from this gene are involved in disease expression, such as extra-articular manifestations or disease severity, or with selected endophenotypes such as serum TNF-α levels or pharmacogenetic response. The study by Nossent, *et al* comprised a total of 335 patients with AS. They reported that the TNF-α –308 GA/AA genotype was associated with a reduced risk of anterior uveitis and better spinal function, whereas the TNF-α –238 GA/AA genotype was associated with later age of onset of AS and lower erythrocyte sedimentation rate. They also reported that serum TNF-α levels were not significantly different between these 2 genotype carriers, suggesting that TNF-α genotype does not influence TNF-α production in AS.

Although the above associations are of potential clinical relevance, there is often apprehension or skepticism in "believing" such results, unless they can be replicated in independent cohorts. This view is supported by Hirschhorn, *et al,* who reported that only 3.6% (6/166) of initial association findings were replicated in subsequent studies[7]. Consequently, initial reports of a novel association should be cautiously interpreted, especially in the context of small sample sizes and marginal statistical significance. Although it is well acknowledged that inadequate power in a study raises doubt with respect to negative association, what is often overlooked is the corresponding reduction in the validity of the results that are stated to reach statistical significance[8].

Besides increasing sample size, how can the power of candidate gene studies be increased?

Many factors contribute to the detection power in association studies. Factors beyond a researcher's control include the actual genetic architecture of markers (i.e., dominant/recessive, penetrance), their allele frequency, and their effect size[8]. Potentially modifiable variables include case selection, sample size, marker selection, and the design of association-based studies[8]. In this editorial, we discuss

See TNF pathway in AS, *page 1675*

general strategies to increase statistical power of studies like that of Nossent, *et al*; and where appropriate, make specific reference to that article.

With respect to phenotype, cases can be genetically enriched, extreme phenotypes can be compared, or an endophenotype examined. Cases can be genetically enriched by choosing probands with familial AS. In this approach, it is important that only 1 member of the family is used, as all cases should be independent of each other. Although there are inefficiencies related to sample collection using this approach, this may be outweighed by the genetic enrichment of the case.

The inclusion of extreme phenotypes can enhance statistical power in a study. With this approach, genetic factors are enriched in extreme phenotypes, which provide more informative alleles by maximizing the differences between cases and controls. Because the frequency of alleles that contribute to a trait are enriched in 1 or both phenotype extremes, a modest sample size can be sufficient to detect an association. This strategy has largely been used for quantitatively measurable traits, but can also be used for dichotomous traits[9]. An extreme phenotype design has been demonstrated as an efficient method for complex disease gene mapping[9].

Another approach is to use a quantitatively measured trait related to the disease, which is hypothesized to improve power to detect a genetic effect and often to have a more interpretable outcome[10]. Nossent, *et al* used serum TNF-α levels as an endophenotype and assessed the association of TNF-α –308 and –238 genetic variations with TNF-α levels in patients with AS. Numerous studies investigating the effect of TNF-α genetic variation on serum TNF-α levels have collectively produced conflicting results, with several demonstrating a significant association between the presence of allele A at TNF-α –308 and higher TNF-α production[11,12,13,14], while other studies failed to find such an association, or reported an even lower TNF-α production in the presence of the A allele at position –308[15,16,17,18]. A similar situation was observed for the TNF-α –238 promoter polymorphism[15,18,19,20]. Notably, the influence of TNF-α promoter polymorphisms on *in vitro* investigations are confounded by use of different cell cultures, different stimuli, variation in cell type analyzed, modeling of different diseases, and different detection assays; this may help explain some of the conflicting results.

In some respects, the criteria for selecting an appropriate endophenotype, determining its heritability, and determining an optimal sampling method, remain elusive. Problems with investigating serum TNF-α levels as an endophenotype are demonstrated in the study by Nossent, *et al*[4], where a significant number of patients with AS had undetectable levels of TNF-α in serum compared with controls. Given the percentage of AS patients with very low serum TNF-α levels in that study, it was not surprising that the authors were unable to establish an association between serum TNF-α levels and clinical disease features. In such a situation, statistically transforming serum TNF-α data so that its distribution can be about normal may help improve the statistical power. Also, conflicting results regarding serum TNF-α levels with AS disease activity[16,17] suggest that measurement of serum TNF-α levels is a poor endophenotype for AS studies.

One of the major determinants of statistical power is the allele frequency. Rare allele frequency refers to a population frequency of less than 1%, although some have used 0.5% as the cutoff; whereas low refers to a frequency between 1 to 5%. In the Nossent, *et al* study[4], the minor allele frequency was only 1.5% for TNF-α gene –238 variant and 7% for –308 variant. In certain situations, a very rare allele can be assigned a higher priority than more common markers, as highly deleterious mutations are likely to be subjected to negative selection. However, given the nature of rare variants, a very large number of patients are required to reach statistical significance.

Careful selection of variants that are *a priori* more likely to be causal is another method that can improve statistical power. Priority is generally given to those variants that are most likely to cause disease, for instance nonsense variants are preferred over coding missense non-synonymous non-conservative markers. Selecting disease-causing markers or markers that are in complete LD with a disease-causing variant will help maximize statistical power. It is also prudent to acknowledge that genetic and epigenetic alterations other than SNP within the promoter region of a gene has the capability to adversely affect gene transcription.

Using special populations such as genetic isolates can minimize the effect of genetic heterogeneity. Young isolates with relatively few founders demonstrate particularly extensive LD with very few gaps[21]. These populations are also characterized by environmental and phenotypic homogeneity, restricted geographical distribution, and the presence of exhaustive and detailed records correlating individuals in very well-ascertained pedigrees[22].

The most frequent comment stemming from a review of genetic association studies is the requirement for independent validation. This is often not feasible because of the unavailability of a replication cohort. In GWAS studies, a multistage design is often used, where 1 group is used for discovery while the other group of patients tests either only significant findings or a predetermined proportion of the initial markers. Although this method contains a built-in replication cohort, it reduces the overall power of the study, because the sample size in the respective datasets is smaller than the pooled sample size[23]. Consequently, the optimal design is an independent replication cohort.

A power calculation is an essential requirement for association-based studies for proper interpretation and comparison of results. Researchers do have control over

some aspects of the proposed study design, and these should be optimized before study initiation. We expect in the coming years with improved study designs, larger informative cohorts, and advancements in technology that additional biological markers will be discovered and implemented clinically to better predict response *a priori*.

DARREN D. O'RIELLY, PhD,
Senior Research Scientist;
GUANGJU ZHAI, PhD,
Associate Professor, Discipline of Genetics,
PROTON RAHMAN, MD, FRCPC,
University Research Professor,
Faculty of Medicine,
Memorial University,
St. John's, Newfoundland, Canada.

Address correspondence to Dr. Rahman, 154 Le Marchant Road, 1 South, St. John's, Newfoundland A1C 5B8, Canada. E-mail: prahman@mun.ca

REFERENCES

1. Wellcome Trust Case Control Consortium; Australo-Anglo-American Spondylitis Consortium (TASC). Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 2007;39:1329-37.
2. Australo-Anglo-American Spondyloarthritis Consortium (TASC). Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. Nat Genet 2010;42:123-7.
3. Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, Kochan G, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat Genet 2011;43:761-7.
4. Nossent JC, Sagen-Johnsen S, Bakland G. Tumor necrosis factor-α promoter –308/238 polymorphism association with less severe disease in ankylosing spondylitis is unrelated to serum TNF-α and does not predict TNF inhibitor response. J Rheumatol 2014;41:1675-82.
5. Lee YH, Song GG. Lack of association of TNF-alpha promoter polymorphisms with ankylosing spondylitis: a meta-analysis. Rheumatology 2009;48:1359-62.
6. Li B, Wang P, Li H. The association between TNF-alpha promoter polymorphisms and ankylosing spondylitis: a meta-analysis. Clin Rheumatol 2010;29:983-90.
7. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. Genet Med 2002;4:45-61.
8. Sham PC, Purcell SM. Statistical power and significance testing studies. Nat Rev Genet 2014;15:335-46.
9. Emond MJ, Louie T, Emerson J, Zhao W, Mathias RA, Knowles MR, et al. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. Nat Genet 2012;44:886-9.
10. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. PLoS Comput Biol 2012;8:e1002822.
11. Lu MC, Yang KL, Tung CH, Huang KY, Yu HC, Liu SQ, et al. Higher LPS-stimulated TNF-alpha mRNA levels in peripheral blood mononuclear cells from Chinese ankylosing spondylitis patients with –308G/A polymorphism in promoter region of tumor necrosis factor: association with distinct A33/B58/Cw10 haplotypes. Rheumatol Int 2008;29:189-95.
12. Kroeger KM, Carville KS, Abraham LJ. The –308 tumor necrosis factor-alpha promoter polymorphism effects transcription. Mol Immunol 1997;34:391-9.
13. Louis E, Franchimont D, Piron A, Gevaert Y, Schaaf-Lafontaine N, Roland S, et al. Tumour necrosis factor (TNF) gene polymorphism influences TNF-alpha production in lipopolysaccharide (LPS)-stimulated whole blood cell culture in healthy humans. Clin Exp Immunol 1998;113:401-6.
14. Braun J, Yin Z, Spiller I, Siegert S, Rudwaleit M, Liu L, et al. Low secretion of tumor necrosis factor alpha, but no other Th1 or Th2 cytokines, by peripheral blood mononuclear cells correlates with chronicity in reactive arthritis. Arthritis Rheum 1999;42:2039-44.
15. Pociot F, Briant L, Jongeneel CV, Mölvig J, Worsaae H, Abbal M, et al. Association of tumor necrosis factor (TNF) and class II major histocompatibility complex alleles with the secretion of TNF-alpha and TNF-beta by human mononuclear cells: a possible link to insulin-dependent diabetes mellitus. Eur J Immunol 1993;23:224-31.
16. Rudwaleit M, Siegert S, Yin Z, Eick J, Thiel A, Radbruch A, et al. Low T cell production of TNFalpha and IFNgamma in ankylosing spondylitis: its relation to HLA-B27 and influence of the TNF-308 gene polymorphism. Ann Rheum Dis 2001;60:36-42.
17. Poddubnyy DA, Märker-Hermann E, Kaluza-Schilling W, Zeidler H, Braun J, Listing J, et al. Relation of HLA-B27, tumor necrosis factor-α promoter gene polymorphisms, and T cell cytokine production in ankylosing spondylitis — a comprehensive genotype-phenotype analysis from an observational cohort. J Rheumatol 2011;38:2436-41.
18. de Jong BA, Westendorp RG, Bakker AM, Huizinga TW. Polymorphisms in or near tumour necrosis factor (TNF)-gene do not determine levels of endotoxin-induced TNF production. Genes Immun 2002;3:25-9.
19. Kaluza W, Reuss E, Grossmann S, Hug R, Schopf RE, Galle PR, et al. Different transcriptional activity and in vitro TNF-alpha production in psoriasis patients carrying the TNF-alpha 238A promoter polymorphism. J Invest Dermatol 2000;114:1180-3.
20. Kaijzel EL, van Krugten MV, Brinkman BM, Huizinga TW, van der Straaten T, Hazes JM, et al. Functional analysis of a human tumor necrosis factor alpha (TNF-alpha) promoter polymorphism related to joint damage in rheumatoid arthritis. Mol Med 1998;4:724-33.
21. Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorious H, Bedoya G, et al. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. Nat Genet 2006;38:556-60.
22. Arcos-Burgos M, Muenke M. Genetics of population isolates. Clin Genet 2002;61:233-47.
23. Nsengimana J, Bishop DT. Design considerations for genetic linkage and association studies. Methods Mol Biol 2012; 850:237-62.

J Rheumatol 2014;41:1573–5; doi:10.3899/jrheum.140692