

Value of Homogenous Populations for Gene Identification in Complex Rheumatic Diseases



Rheumatoid arthritis (RA), like most other complex rheumatic diseases, has multiple genetic and environmental determinants contributing to its susceptibility. Evidence for the genetic contribution to RA can be seen in population based and twin studies, as well as molecular investigations from candidate gene and linkage analysis¹. Despite extensive ongoing research, few candidate genes have been identified and validated in RA from large mixed populations. The lack of replication of initial associations may be a result of false-positive reportings (from multiple testing, population stratification, and genotyping errors), false-negative findings (due to small effect size coupled with inadequate sample size of validation studies), or population-specific differences². In this issue of *The Journal*, Oen, *et al* set out to characterize the RA phenotype, estimate familial incidence of RA, and evaluate the association of various cytokine genes implicated in the pathogenesis of RA in the North American Native (NAN) population from Manitoba and Northwest Ontario, Canada³.

Homogenous populations, like the NAN, are now the focus of intensive interest for gene identification in complex diseases⁴. Although the utility of such populations for this purpose remains unproven, it is widely assumed that homogenous populations offer advantages in gene identification by mitigating some of the above limitations. The benefits in pursuing gene identification studies in such populations include an increase in allele and locus homogeneity, as well as the often-discussed potential increase in linkage disequilibrium. The former is of importance given the etiologic heterogeneity that characterizes complex diseases. Thus the detection of a significant signal in RA from a mixed population is more challenging, as modest genetic signals may be overlooked in outbred populations, unless one assembles very large cohorts. Meanwhile, in homogenous populations there may be an increased signal to noise ratio to identify genes of modest risk, due to the relative genetic and envi-

ronmental homogeneity that often exists within such populations⁵.

We share Oen and colleagues' enthusiasm in investigating the genetic determinants of RA in the NAN population, as there appears to be a greater genetic burden of RA in this population. This is based on the high prevalence of RA, an earlier age of onset, greater severity, as well as higher rate of seropositivity [including rheumatoid factor, shared epitope (SE), and antinuclear antibody] in the NAN RA population. Specifically, the Chippewa and Blackfoot Indians have a 5-fold increased rate of RA compared to the Caucasian North American and European populations (reviewed by Peschken and Esdaile⁶). The ancestral history and migratory route of the NAN population likely has bearing on the increased prevalence of RA. The ancestors of NAN appeared to have originated from northeast Asia; patterns of migration of the NAN are likely to have created population bottlenecks, resulting in a small number of founding chromosomes⁷. Thus the increased prevalence of RA in selected NAN populations is likely due to genetic drift, resulting from the change in allele frequencies associated with founder effects. For these reasons, the NAN population represents a unique resource for identification of RA related genes.

It should be noted, however, that within the NAN population there is likely to be heterogeneity due to multiple distinct waves of migration from different founders, together with later admixture between NAN groups. NAN subpopulations differ in age, size of the genetic bottleneck, and in expansion rate. Clinical heterogeneity can be seen, for example, by comparing Amerind Indians, who have increased rates of RA and connective tissue disease, with the Na-Dene Indians and Eskimos, who have high rates of spondyloarthropathies⁶. The clinical heterogeneity among the NAN groups is mirrored in molecular studies. For instance, the specific SE alleles bearing association in the

See Familial seropositive RA in North American Native families:
effects of shared epitope and cytokine genotypes, *page 983*

Tlingit, Yakima, and Pima (HLA-DRB1*1402) differ from those associated in the Chippewa (DRB1*04). Thus it is important to know as much as possible about the population structure and genealogical history to maximally utilize such populations for novel gene discoveries. Further, even within apparent homogenous populations, documenting self-reported ethnic background (for instance, the majority of the patients in Oen's study were Cree or Ojibway and had 4 NAN grandparents) does not necessarily mean that no substructure exists within this group. This point is illustrated by a recent Icelandic study that investigated population substructure within this relatively homogenous genetic isolate. The investigators concluded that even in a homogenous population, various sampling strategies are required to take account of substructure, since there were small variations in allele frequencies by geographic region. Hence self-reported ethnicity is not sufficient for inferring the presence or absence of population substructure⁸. Despite these caveats, we feel a more homogenous genetic background in a population will mean less molecular heterogeneity.

Oen's report shows a higher familial prevalence of RA in the NAN population. Specifically, they noted that the prevalence of multiplex families among the RA probands whose families were studied was 50% (14 of 28 families); a lower bound for prevalence among relatives of all probands is therefore 17% (14/82). Although the high familial prevalence of RA is not unexpected given the epidemiological (younger age of onset, greater severity of disease) and molecular clues (higher frequency of the SE), this prevalence estimate should still be viewed with caution. There is a potential for bias in such studies, as families with multiple affected family members, younger age of onset, and more severe disease are also more likely to present to a treating physician and thus be included in genetic studies. Similarly, the suggestion of genetic anticipation must be interpreted with extreme caution, since apparent decreases in age at onset are likely due to sampling bias⁹. Thus in the absence of larger confirmatory population based studies, it is premature to translate this result into the clinical setting (i.e., for genetic counselling).

The present investigators previously reported a high prevalence of SE in the Manitoba and Northwestern Ontario NAN population. Even though the risk of SE is 5-fold greater in those with RA from the NAN population, the high population prevalence of SE suggests an increased frequency of "protective" genes in unaffected individuals. To search for these "protective genes," unaffected family controls were used. As noted by Oen, *et al*, the benefit of using unaffected family members for controls in association based studies is the reduction of bias due to population stratification, in particular when a matched analysis is used. However, this familial design results in (over)matching on the gene of interest as well as on the hidden subpopulations, and thus necessitates larger sample sizes to investigate an

association¹⁰. Oen, *et al* therefore also performed an unmatched analysis and compared their cases to unrelated unaffected individuals, but these approaches may show some spurious associations due to hidden population substructure. An alternative approach may have been to maintain the robustness regarding population stratification by adjustment of test statistics based on a series of unlinked markers¹¹. Nevertheless, examining discordant pairs of siblings or relatives is a powerful approach for linkage or association analysis when the disease prevalence is very high, as in this case¹²⁻¹⁴. Further, family based controls are an efficient design choice when the focus is on testing for gene-environment interactions¹⁵. Hence, family based controls may be the best choice here for investigating interactions between SE and the other markers.

Finally, in Oen and colleagues' study, the only significant protective genotype noted was with the interleukin 10 promoter -1082 G allele. There was a decreased frequency of the -1082 G allele among RA probands compared to their unaffected relatives and unaffected individuals of unrelated probands. The validation of the above association is critical, as the generalizability of findings from homogenous populations is a commonly cited concern behind a lack of replication. With respect to population variation and replication, rare variants are most likely to be population-specific and at the same time least replicable. Meanwhile, common alleles are more likely to be found globally.

In summary, the NAN population is a valuable resource for identifying RA related genes, given the high genetic burden of this disease, coupled with the reduced allelic diversity compared with more outbred populations.

PROTON RAHMAN, MD, MSc, FRCPC,
Associate Professor of Medicine,
Division of Rheumatology,
Memorial University of Newfoundland,
St. John's, Newfoundland;
CELIA GREENWOOD,
Project Director,
Program in Genetics and Genomic Biology,
Hospital for Sick Children,
Toronto, Ontario, Canada.

Address reprint requests to Dr. P. Rahman, St. Clare's Mercy Hospital,
1 South - 154 LeMarchant Road, St. John's, Newfoundland A1C 5B8.
E-mail: prahman@mun.ca

REFERENCES

1. Barton A, Ollier W. Genetic approaches to the investigation of rheumatoid arthritis. *Curr Opin Rheumatol* 2002;14:260-9.
2. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003;33:177-82.
3. Oen K, Robinson DB, Nickerson P, et al. Familial seropositive rheumatoid arthritis in North American Native families: effect of shared epitope and cytokine genotypes. *J Rheumatol* 2005;32:983-91.

4. Laitinen T. The value of isolated populations in general studies of allergic diseases. *Curr Opin Allergy Clin Immunol* 2002;2:379-82.
5. Rahman P, Jones A, Curtis J, et al. The Newfoundland population: a unique resource for genetics investigation of disease. *Hum Mol Genet* 2003;12 Supp 12:R167-72.
6. Peschken CA, Esdaile JM. Rheumatic diseases in North America's indigenous peoples. *Semin Arthritis Rheum* 1999;28:368-91.
7. Williams RC, Steinberg AG, Gershowitz H, et al. Gm haplotypes in Native Americans: evidence for three distinct migrations across the Bering land bridge. *Am J Phys Anthropol* 1985;66:1-19.
8. Helgason A, Yngvadottir B, Hrafnkelsson B, Gulcher J, Stefansson K. An Icelandic example of the impact of population structure on association studies. *Nat Genet* 2005;37:90-5.
9. Petronis A, Kennedy JL, Paterson AD. Genetic anticipation: fact or artifact, genetics or epigenetics? *Lancet* 1997;350:1403-4.
10. Risch NJ. Searching for genetic determinants in the new millennium. *Nature* 2000;405:847-56.
11. Freedman ML, Reich D, Penney KL, et al. Assessing the impact of population stratification on genetic association studies. *Nat Genet* 2004;36:388-93.
12. Abecasis GR, Cookson WO, Cardon LR. The power to detect linkage disequilibrium with quantitative traits in selected samples. *Am J Hum Genet* 2001;68:1463-74.
13. Rogus JJ, Krolewski AS. Using discordant sib pairs to map loci for qualitative traits with high sibling recurrence risk. *Am J Hum Genet* 1996;59:1376-81.
14. Boettcher SA. Optimal designs for linkage disequilibrium mapping and candidate gene association tests in livestock populations. *Genetics* 2004;166:341-50.
15. Witte JS, Gauderman WJ, Thomas DC. Asymptotic bias and efficiency in case-control studies of candidate genes and gene-environment interactions: basic family designs. *Am J Epidemiol* 1999;149:693-705.