

HIGHLIGHTS

WEB WATCH

SARS resource

• <http://www.ncbi.nlm.nih.gov/genomes/SARS/SARS.html>

With the recent confirmation of a new outbreak of severe acute respiratory syndrome (SARS), the possibility that this disease will become a major killer is a real concern. Modern diagnostic tests, antiviral agents and vaccines are all needed to combat the disease, and these require sequence data and the functional dissection of the genome of the SARS coronavirus.

To this aim, in May 2003, the National Center for Biotechnology Information (NCBI) set up the SARS coronavirus resource to provide "...the most recent sequences, annotations, and analysis...". And with the most recent update in February 2004, to include data from the Chinese SARS Molecular Epidemiology Consortium, this really does seem to be the case.

On the web site, there is the option to view pre-computed global alignments of the genomes of various isolates. The complete sequence data of both genes and proteins are also available, along with their pre-computed analysis. Of course, with the availability of the BLAST alignment tool, there is also the option to carry out your own analysis.

Combine all of this with links to the most recent SARS publications and other useful SARS-related information — from organizations such as the Center for Disease Control and Prevention (CDC), the World Health Organization (WHO) and Medline Plus — and this really is a web site that is worth bookmarking by those who want to keep up to date with the field.

Natalie Wilson

FURTHER READING The Chinese SARS Molecular Epidemiology Consortium. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* 29 Jan 2004 (doi:10.1126/science.1092002)



HUMAN GENETICS

Perfect match

Some very simple strategies for matching cases and controls can deal with the persistent bugbear of association studies — structured populations.

In structured populations, subgroups can differ in both their susceptibility to a disease and a suite of other unrelated characters. If disease cases and controls are selected at random, the more-susceptible population subgroup will be more common in cases than it is in the population as a whole, whereas the less-susceptible subgroup will be similarly over-represented in the controls. The trouble is that when the cases and controls differ in their composition with respect to these population subgroups, the genes underlying the unrelated characters that differ between subgroups can be spuriously associated with the disease. Now, David Hinds and colleagues show that several simple matching strategies can eliminate stratification from population samples.

The Mexican population that they studied is a classic case of admixture: indigenous Indians and Caucasians of Spanish European ancestry have both made substantial contributions to the present-day populace, most of whom regard themselves as 'Mestizo' — mixed ancestry. As Caucasians tend to be taller than Indians, the authors expected that any unmatched association study of height — the model complex trait — would be confounded with ancestry.

This intuitive assessment of the system proved to be spot on: many of the 275 SNPs typed from the genome of 707 individuals differed in allele frequencies among the

self-assessed ancestral categories. Moreover, a genotype-based analysis of population structure in the sample defined two distinct subpopulation clusters — A and B — which correspond to groups with largely European or Indian ancestry, respectively. These clusters hopelessly confounded the predefined 'case' and 'control' groups of tall and short individuals: many SNPs were spuriously associated with height merely because of the differences in ancestry between these groups.

So, in a system that is so plagued by confounding population stratification, could a simple matching scheme remove the possibility of spurious associations?

The authors tried two matching strategies: in one, they removed individuals from the tall and short groups with the largest proportions of confounding ancestry until the average proportions in these groups were equal, whereas in the other, individuals were reselected for the tall and short groups after adjusting the value of their height to take into account differences in ancestry. Both strategies largely removed the telltale signs of population stratification in the data.

Simple matching strategies such as these, based on a limited set of genotyped SNPs, could eventually be the standard prelude to a large-scale genotyping effort on pooled DNA. Such an approach would seem to provide the best of both worlds: the confidence that the whole-genome association study has been controlled for population stratification without the need for genotyping millions of SNPs individually.

Nick Campbell

References and links

ORIGINAL RESEARCH PAPER Hinds, D. A. *et al.* Matching strategies for genetic association studies in structured populations. *Am. J. Hum. Genet.* **74**, 317–325 (2004)

FURTHER READING Sham, P. *et al.* DNA pooling: a tool for large-scale association studies. *Nature Rev. Genet.* **3**, 862–871 (2002)

WEB SITE

Perlegen Sciences: <http://www.perlegen.com>