Heredity (2006) 97, 253

www.nature.com/hdy

NEWS AND COMMENTARY

Quantitative genetics

Wholesale analysis of genes, traits and microarrays

RB O'Hara

Heredity (2006) **97,** 253. doi:10.1038/sj.hdy.6800857; published online 21 June 2006

Statistics is, or at least should be, a service industry; dedicated to helping other researchers with their more disagreeable data analyses. This is becoming more important now, as biologists produce ever more data, which then needs to be mined for useful information that might (with luck) say something about the systems being studied. With the expansion of bioinformatics, there are plenty of new insights to be gained, and hence opportunity for statisticians to develop tools with which to dig these gems of knowledge out from the piles of data being generated. Recently, in *Heredity*, Hoti and Sillanpää (2006) presented a method for analysing quantitative traits using both genotypic and microarray data that has the potential to do just this.

Hoti and Sillanpää (2006) are interested in finding the genes that control a quantitative trait, using a novel approach that combines two conceptually different sources of data. They envision a series of crosses, where a quantitative trait is measured in individual progeny, which are also genotypes at several segregating loci. This is a standard design, which would be appropriate for a traditional QTL analysis. The novelty arises because they additionally propose that, for all the individuals in the cross, measurements are taken of gene expression, using a mircoarray assay that monitors several genes. They then propose that the data from both sources should be combined in one huge regression to find which expressed genes and which marker loci can explain the variation in the phenotype. Using a neat statistical trick, they assign probabilities to whether a locus or gene affects the quantitative trait. Going further, they also allow for the effect of the gene expression to be affected by the genetic background, by including interaction terms between the gene expression and the markers.

Clearly, this technique can be used to analyse the right sort of data, but what would the results mean? In order to answer this question, we should note that genetic loci and gene expression profiles provide qualitatively different types of explanation of the phenotype. Marker data provide information about genetic causes: this will be apparent if a marker is linked to a locus with different alleles that affect a trait. In contrast, gene expression studies provide a more functional explanation. Hence, the conclusions from a study could be messy; although it might be inferred that variation in a trait is caused by variation in gene expression or by segregation at a locus, these are different types of causation.

These conceptual problems arise when interpreting the actual analysis. While variation in gene expression could be entirely caused by environmental variation, it could also be controlled by segregation of a locus that is linked to a marker. Hence, both the genotypic and expression data can explain the trait being studied. This situation leads to the conflict between explanations that can be seen in the statistical analysis. Statistically, the effect is to induce a correlation between the estimated effects of the segregating locus and the level of gene expression, as the more variation is explained by one, the less there is to be explained by the other. Hoti and Sillanpää provide Q-summaries as a tool to diagnose this situation: these are the conditional probabilities that the QTL has an effect, given the gene expression has an effect (or *vice versa*).

The difference in the types of explanation also underlies another aspect of this study. A quantitative trait can obviously be affected by many genes, which can be expressed at different times. So, it may not be affected by the variation in expression at the time that the tissue is sampled for the gene expression analysis. In contrast, the marker data are not constrained by time or place of expression. The sampling strategy for the gene expression analysis clearly has to be thought through carefully, but a marker-based analysis may help by being able to catch variation that has been missed because the differential expression occurred at the 'wrong' time. Even though the markers cannot provide a functional explanation for the variation, they can point to the right bit of the genome to look for answers.

Now we have a shiny new toy, we should be able to play around with it. Luckily, there are many things to try. For example, Hoti and Sillanpää look for possible interactions between marker genotypes and gene expression levels, but it is obvious that interactions can also occur between genes, and between markers (or, to be precise, between loci that are linked to the markers). These extensions are trivial in principle: the problem is that they make the model much bigger and more data would be needed to get good estimates of the effects. All sorts of other effects could also be added (eg measurements in different environments, with different treatments). Another possibility would be to enter the world of genomics, using marker data to explain gene expression in the same analysis. This will get directly to the problem, raised above, of genes linked to markers exerting their effects through the measured gene expression.

These possible extensions to the model will need developing, of course. What makes Hoti and Sillanpää's paper interesting is that it opens up these possibilities. Hopefully, it will encourage researchers in the laboratory to think about carrying out such experiments, and to ask statisticians to develop a model with which to extract information out of the data. Thanks to increased computing power, statisticians are able to develop ever more exciting and sophisticated techniques for analysing data: the challenge for biologists is to use these techniques to their full potential.

RB O'Hara is at the Department of Mathematics and Statistics, University of Helsinki, PO Box 68 (Gusfaf Hällströmin katu 2b), Helsinki, FIN-00014, Finland.

E-mail: bob.ohara@helsinki.fi,
WWW: http://www.RNI.Helsinki.FI/~boh/

Hoti F, Sillanpää MJ (2006). Bayesian mapping of genotype × expression interactions in quantitative and qualitative traits. *Heredity* (doi:10.1038/sj.hdy.6800817).

Editor's suggested reading

Rodríguez-Trelles F (2004). Evolutionary genetics: transcriptome evolution – much ado about nothing? *Heredity* **93**: 405–406.

Brasset E, Vaury C (2005). Insulators are fundamental components of the eukaryotic genomes. *Heredity* **94**: 571–576.

Erickson D (2005). Mapping the future of QTLs. *Heredity* **95**: 417–418.