STATISTICAL GENETICS

Fitting phenotypes

> An alternative approach to determining gene variants that contribute to a particular trait is to group all SNPs together and ask whether they can predict a phenotype.

Analysing the results of genome-wide association studies is a painstaking effort — each SNP has to pass stringent significance thresholds to be regarded as a respectable candidate. An alternative approach to determining gene variants that contribute to a particular trait is to group all SNPs together and ask whether they can predict a phenotype. One such method, based on a Bayesian approach, has now been used to predict three mouse phenotypes. Similar approaches could be useful in other areas of medical genetics as well as in forensics and artificial selection in livestock.

Bayesian approaches are well suited to the prediction of phenotypes. The aim is not to test hypotheses but to estimate the effect of each SNP and to combine all the SNP effects into a prediction of phenotype that is as accurate as possible. In this paper, the authors have tested the feasibility of using a Bayesian approach called reversible jump Markov chain Monte Carlo (RJMCMC) on genome-wide SNPs to predict three phenotypes in heterogeneous stock mice — coat colour, the percentage of CD8[+] cells, and mean cellular haemoglobin (see the link for a description of how these mice were constructed).

The data came from four generations of mice, over 2,000 animals, and consisted of 10,000 SNPs as well as pedigree and phenotype information. Genetic models were developed based on the full genotypic data but using the phenotypes of only half the animals, and then they were validated by predicting phenotypes in the remaining half of the population. The models incorporated either additive effects only or a mixture of additive and dominance effects (the AD model).

Predictions were successful across all traits — accuracy ranged from 0.4 to 0.9 — with AD models being superior to additive-only models; for example, coat-colour predictions are 81% accurate under the AD model. More accurate predictions were obtained with traits, such as CD8[+] percentage, that are more heritable — that is, for which more of the trait variation between individuals actually depends on genetic factors. Phenotypes were predicted across families but also within families; in the latter case, predictions were enriched by pedigree information and therefore performed better.

Using genome-wide information gave a marked improvement in accuracy over using single SNPs or even entire chromosomes at a time. The high accuracy, computational efficiency and speed of the analysis method (this data set took 15 minutes to analyse) means that it could be adapted for use on additional traits and larger samples, and for other species and applications. This paper builds on previous work by the authors that demonstrated the use of dense SNP genotypes to predict genetic value in livestock and disease risk in humans.

Tanita Casci