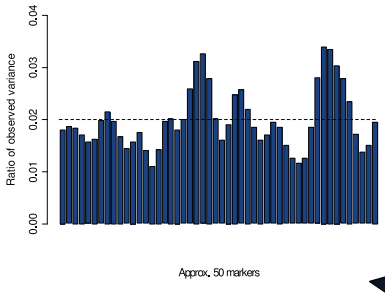
Using KEGG pathways information for genomic partitioning

Stefan McKinnon Edwards, Axel Skarman, Li Jiang, Per Madsen, Guosheng Su, Peter Sørensen
Department of Molecular Biology and Genetics, Faculty of Science and Technology, Aarhus University, 8830 Tjele, Denmark.

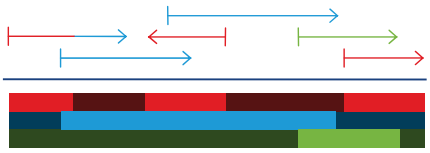
The problem

Estimating single marker effects of a complex trait using 650,000 markers results in marker effects being spread too thin.



A solution

Use prior biological information such as KEGG pathways to partition the genome into two or more non-overlapping sets.



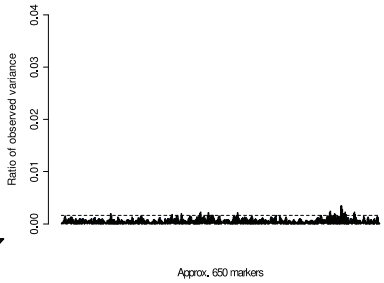
Genes associated with 3 pathways

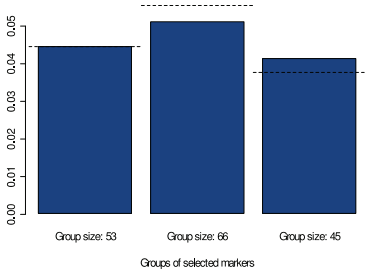
Markers included in group for pathway (light colours)

$$Y = \mathbf{1}\mu + X_1\beta_1 + X_{1^{-1}}\beta_{1^{-1}} + \varepsilon$$

$$Y = \mathbf{1}\mu + X_2\beta_2 + X_{2^{-1}}\beta_{2^{-1}} + \varepsilon$$

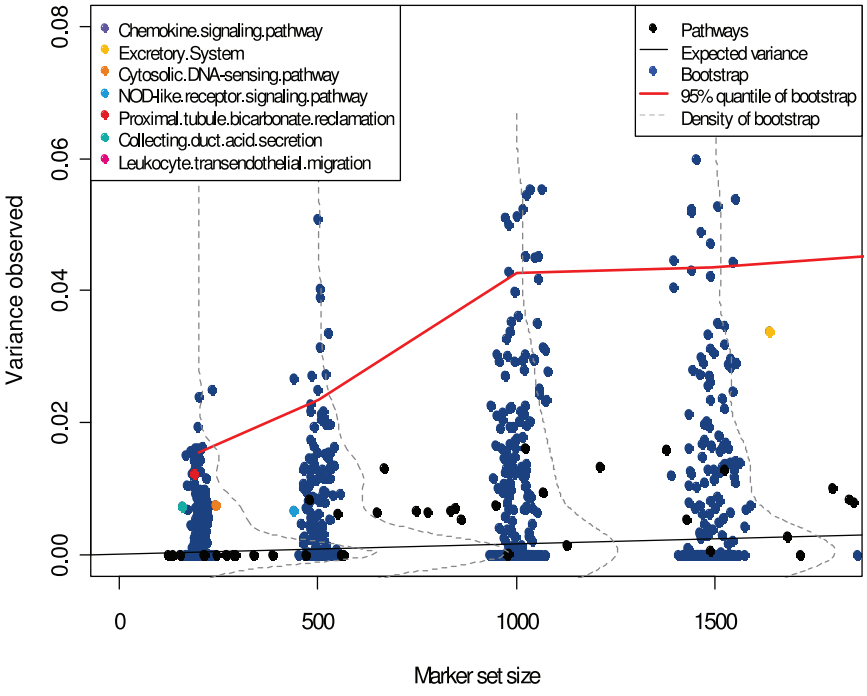
$$Y = \mathbf{1}\mu + X_3\beta_3 + X_{3^{-1}}\beta_{3^{-1}} + \varepsilon$$





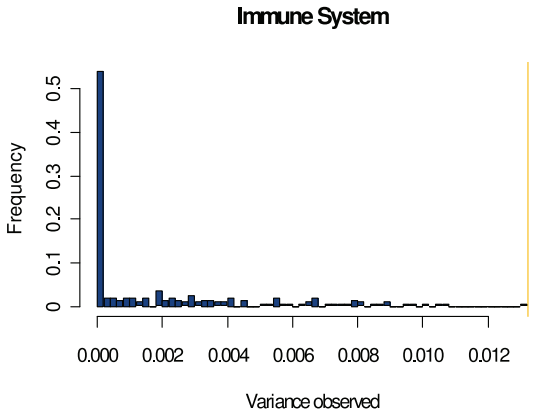
Results

Examination of observed variance on mastitis in dairy cattle on ~650,000 markers.



Test of power

- Null hypothesis: all markers contribute equally to complex trait
- Bootstrapping for estimation of null hypothesis
- Self-contained permutation test for stricter alternative null hypothesis



Conclusion

A large proportion of the genetic variance can be explained by markers located in genes associated with certain pathways.