



Partitioning of genomic variance using biological pathways

Stefan McKinnon Edwards, Luc Janss, Per Madsen, Peter Sørensen
 Stefan.Hoj-Edwards@agrsci.dk

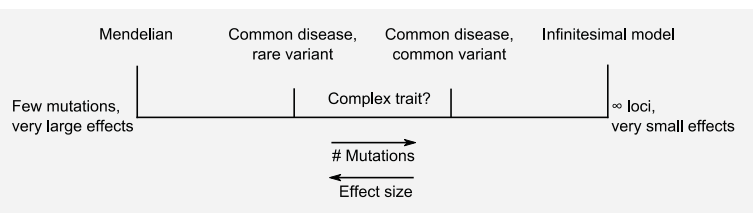
Center for Quantitative Genetics and Genomics, Dept. of Molecular Biology and Genetics, Aarhus University (DK)

Problem

A biological pathway is associated to multiple genes. How can we evaluate multiple genes' joint contribution to a complex trait using linear mixed models?

Introduction

Traditional genome-wide association studies are optimal for traits governed by few mutations with large genetic effects. Traditional linear mixed models (LMM) assume the infinitesimal model, with almost infinite mutations of very small effect. Complex traits is somewhere between these two extremes.



Partitioning by e.g. genes allows LMM to take genetic architecture into account, see figure 2.

What to consider when evaluating the joint contribution to Mastitis of genetic variants in e.g. the Immune System:

- Genes scattered throughout the genome, see figure 3.
- Likelihood ratio distribution not necessarily χ^2 .
- More genes might explain more phenotypic variance, see figure 4.

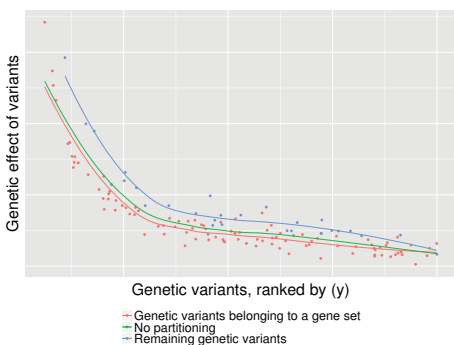


Figure 2: Fitting genetic variants to LMM without partitioning (green) and with partitioning (red and blue).

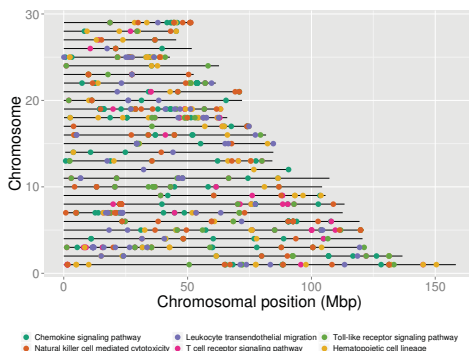


Figure 3: Chromosomal location of genes associated to the KEGG pathway "Immune System".

Data

- 4,497 genotyped bulls.
- 637,951 markers per bull.
- Health and production traits derived from daughters.
- Markers and genes mapped to UMD 3.1 assembly.

Solution

Compare gene set to empirical distributions of random gene sets.

Results

- Likelihood ratio resembles skewed χ^2 distribution.
- Expected proportion of explained phenotypic variance increases when including more genetic variants.
- Genetic variants in some KEGG pathways can account for a significant proportion of explained genetic variance.

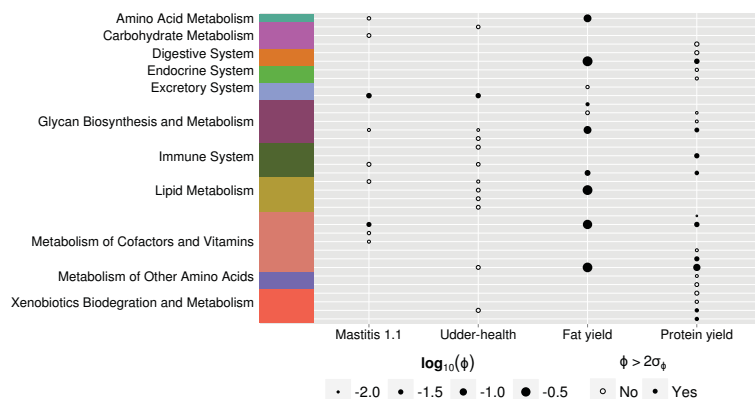


Figure 1: Pathways found significant by empirical likelihood ratio test. ϕ is the proportion of explained genetic variance. $2\sigma_\phi$ is twice the standard deviation of ϕ .

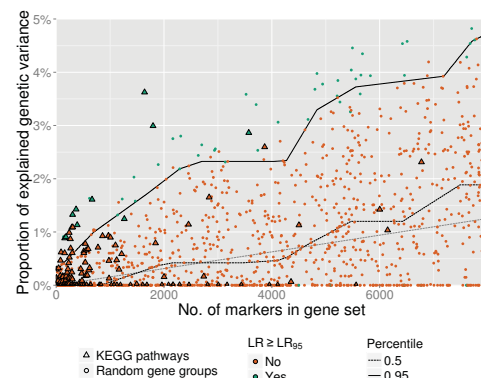


Figure 4: Proportion of explained genetic variance of KEGG pathways vs. randomly sampled gene sets. Green points have a better fit than 95% of random gene sets.

Methods

- Genes sampled uniformly.
- Empirical distributions estimated for up to 50,000 markers.
- Fitting of LMM done by AI-REML algorithm in DMU.

