Including copy number variation in association studies

M. P. L. Calus¹, D. J. de Koning², and C. S. Haley²,³

Introduction

Copy number polymorphisms (CNPs) are relatively common in the genome and there are clear examples where CNPs affect phenotypic variation. However, it is not clear whether SNPs used in association studies can effectively capture the variation due to CNPs. Copy number polymorphisms (CNPs) are different from SNP loci because they have higher mutation rate and can have more than 2 alleles. For CNPs with >2 alleles, derivation of the CNP genotypes from raw hybridizations is sometimes problematic.

Objectives

To investigate whether SNPs are likely to capture variation caused by CNPs by examining:
- the expected linkage disequilibrium (LD) between a SNP and a CNP locus.
- the additional benefit of including the CNP, by its 'phenotype' (i.e. raw hybridization or predicted genotype), next to a SNP in the model, to explain variation at the CNP locus.

Simulations

Three types of loci were simulated (100,000 replicates):

<table>
<thead>
<tr>
<th>Locus type</th>
<th>Mutation rate</th>
<th>(Max.) Number of alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP</td>
<td>10⁻⁴</td>
<td>2</td>
</tr>
<tr>
<td>CNP₂</td>
<td>10⁻⁹</td>
<td>2</td>
</tr>
<tr>
<td>CNPₘ</td>
<td>10⁻⁹</td>
<td>No restriction</td>
</tr>
</tbody>
</table>

Models

To investigate the 2nd objective, the following equations were deterministically derived (Note: h² is the 'heritability' of the CNP phenotype; i.e. the reliability of a predicted CNP genotype):

For a model including only a CNP phenotype, the R² to explain variation at the CNP locus is:

\[ R² = h² \]

For a model including a SNP and a CNP phenotype, the R² to explain variation at the CNP locus is (SNPg is SNP genotype, CNPg is CNP genotype):

\[ R² = \frac{(1 - h²) \times r²(SNPg, CNPg) + h²}{1 - h²r²(SNPg, CNPg)} \]

Results

- Having a direct measure of CNPs may benefit association studies.
- LD between a SNP and a CNP locus appears to be comparable to LD between two SNP loci despite the higher mutation rate of CNP loci.
- Using the raw hybridizations or predicted genotypes of CNP loci are useful alternatives, even when they explain only 15% of the variation at the CNP locus.

Conclusions

Acknowledgements

SABRE (http://www.sabre-eu.eu/) is acknowledged for financial support of the stay of MC at Roslin Institute. DJK and CSH acknowledge the Institute Strategic Program Grant from the BBSRC to the Roslin Institute.

1 Animal Sciences Group, Wageningen University and Research Centre Animal Breeding and Genomics Centre PO Box 65, 8200 AB Lelystad, The Netherlands Tel. +31 320 238265 e-mail: mario.calus@wur.nl www.asg.wur.nl

2 Division of Genetics and Genomics, Roslin Institute and RIDISVS, University of Edinburgh, Roslin, EH25 9PS, UK

3 MRC Human Genetics Unit, Western General Hospital. Crewe Road, Edinburgh, EH4 2XU, UK