ANIMAL SCIENCES GROUP
WAGENINGEN UR

Including copy number variation in association studies

M. P. L. Calus[1], D. J. de Koning[2], and C. S. Haley[2,3]

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 (CNP) are different from SNP loci because they have higher mutation rate and can have more than 2 alleles. For CNP 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):

Locus type	Mutation rate	(Max.) Number of alleles
SNP	10^{-4}	2
CNP2	10^{-9}	2
CNPm	10^{-9}	No restriction

Models

To investigate the 2nd objective, the following equations were deterministically derived (Note: h^2 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 R^2 to explain variation at the CNP locus is:

$$R^2 = h^2$$

For a model including a SNP and a CNP phenotype, the R^2 to explain variation at the CNP locus is (SNPg is SNP genotype, CNPg is CNP genotype):

$$R^2 = \frac{(1-h^2)\times r^2(SNPg, CNPg) + h^2}{1 - h^2 r^2(SNPg, CNPg)}$$

Results

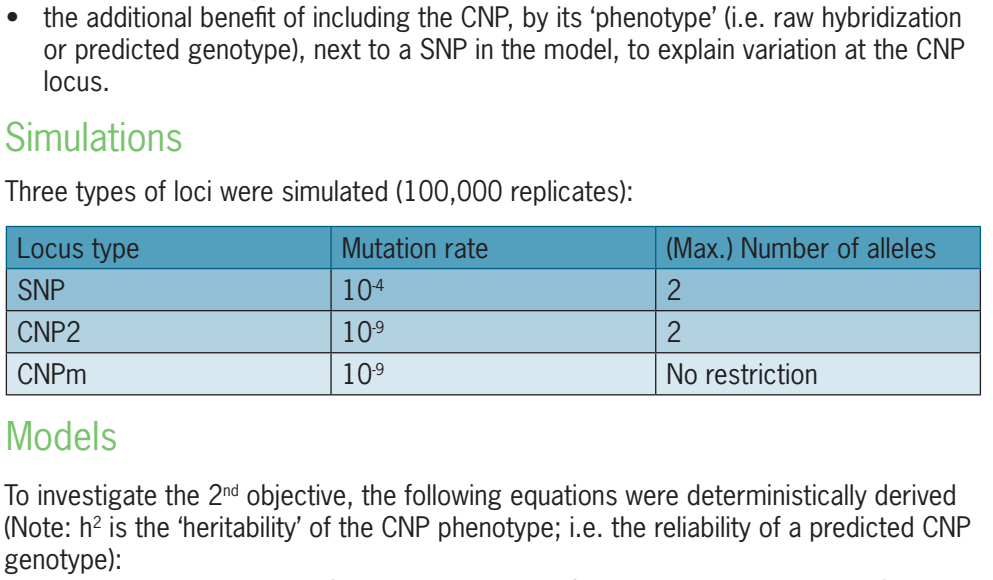

*Figure 1: CNP phenotypes, explaining 25% of the variation in CNP genotypes, plotted against the CNP genotypes for 500 individuals and one CNP locus. → **derivation of discrete CNP genotypes proves difficult.***

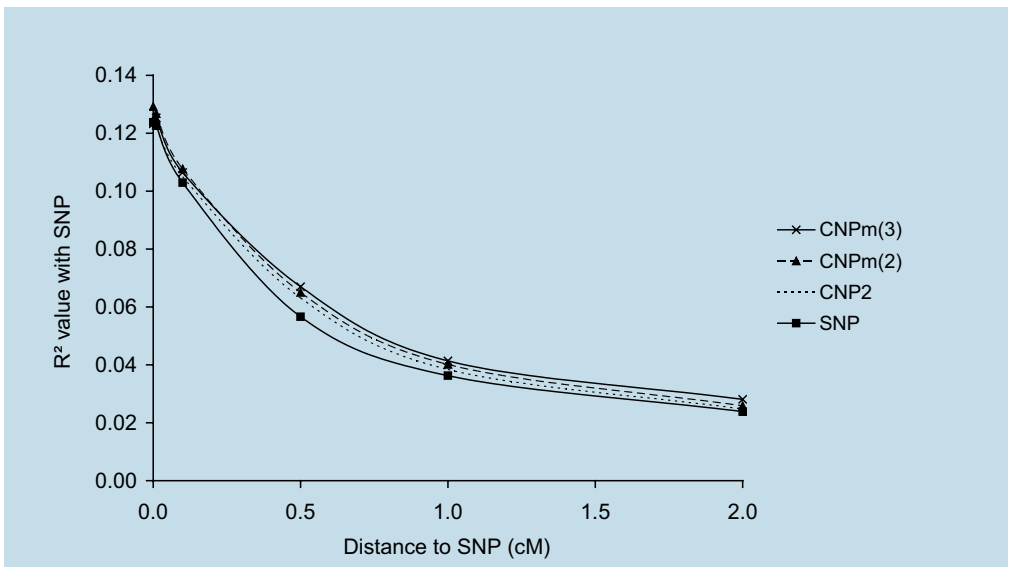

*Figure 2: R^2 value (representing LD) between a SNP and: a SNP, CNP2, CNPm with 2 or 3 segregating alleles → **LD is similar between all types of loci.***

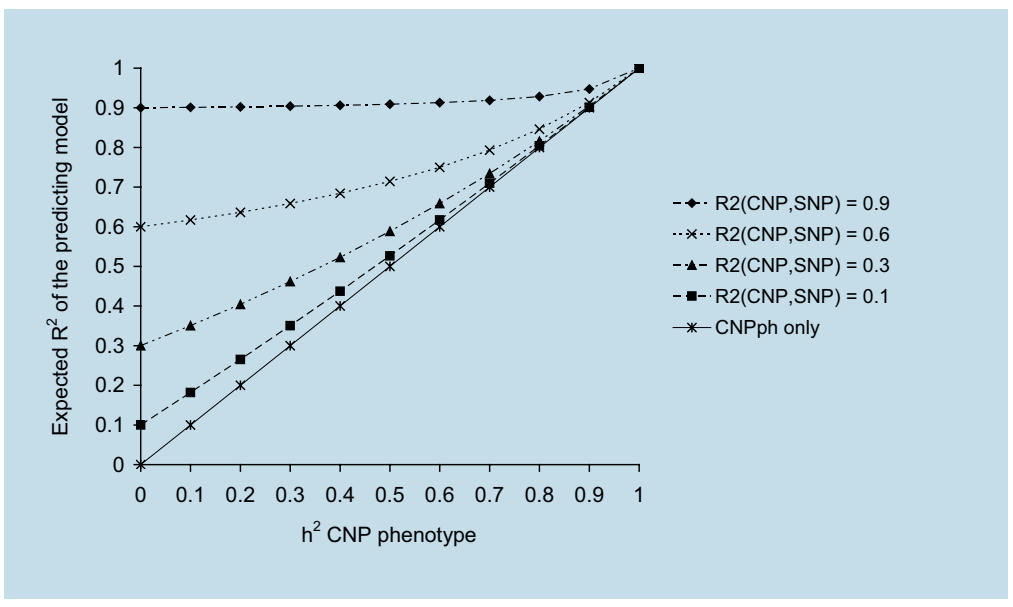

*Figure 3: Deterministically predicted R^2 values obtained for models including CNP phenotypes and SNP genotypes assuming different R^2 values between CNP and SNP loci (R2(CNP,SNP)=…), or only CNP phenotypes (CNPph only), as a function of h^2 of the CNP phenotypes. → **including CNP phenotypes increases model R^2.***

Conclusions

- 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.

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 R(D)SVS, University of Edinburgh, Roslin, EH25 9PS, UK
[3] MRC Human Genetics Unit, Western General Hospital. Crewe Road, Edinburgh, EH4 2XU, UK

6542_Calus 2009