Exploring epistatic effects between regulatory and protein-coding variation

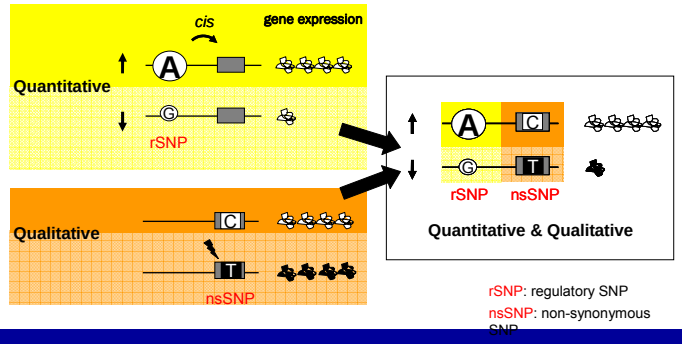
Antigone S. Dimas[1], Barbara E. Stranger[1], Claude Beazley[1], Robert D. Finn[1], Catherine E. Ingle[1], Matthew S. Forrest[1], Matthew E. Ritchie[2], Panos Deloukas[1], Simon Tavaré[2] and Emmanouil T. Dermitzakis[1]

1. Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. 2. Cancer Research UK Cambridge Research Institute, Cambridge, UK.

Abstract

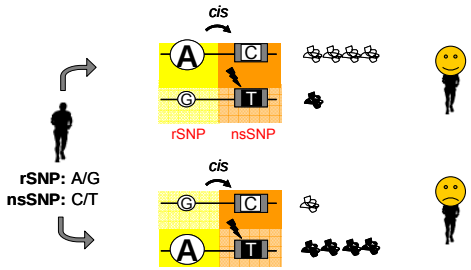
Genome-wide associations have shown a lot of promise in dissecting the genetics of complex traits in humans with single variants, yet a large fraction of the genetic effects is still unaccounted for. Analyzing genetic interactions between variants (epistasis) is one of the potential ways forward. We investigated the abundance and functional impact of a specific type of epistasis, namely the interaction between regulatory and protein-coding variants. Using genotype and gene expression data from the 270 individuals of the HapMap populations, we explored the combined effects of regulatory and coding single nucleotide polymorphisms (SNPs). We predict that over 18% of protein-coding variants can be differentially expressed and that genotypically identical double heterozygotes can have different phenotypic profiles depending on the phasing of the alleles in the two loci. We also demonstrated how regulatory variants can modulate the functional effect of a coding variant in *cis*, and conclude that the differential expression of distinct protein isoforms has an impact on gene expression in *trans*. Given the abundance of both types of variants in human populations, we propose that joint consideration of regulatory and coding variants may reveal additional genetic effects on complex traits and disease.

Interrogating joint quantitative and qualitative variation



rSNP: regulatory SNP
nsSNP: non-synonymous SNP

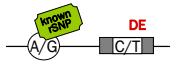
Phasing of regulatory and coding alleles can result in very different phenotypes



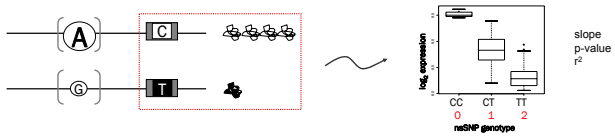
Genotypically identical individuals show distinct phenotypes depending on phasing

2 strategies to discover differentially expressed (DE) nsSNPs in the HapMap populations

1. Identify nsSNPs in genes with known rSNPs[1]



2. Direct association testing between nsSNP genotype and expression levels

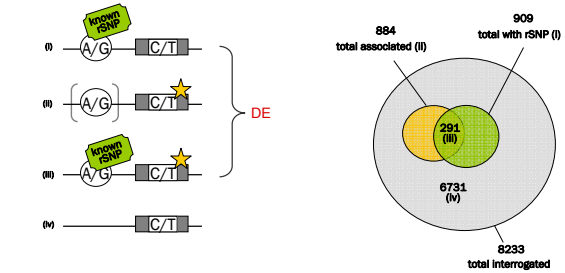


Some nsSNPs will be in linkage disequilibrium with an rSNP

Linear regression to test for association

1. Stranger *et al Nature Genetics* 2007

18.2 % of nsSNPs interrogated are DE

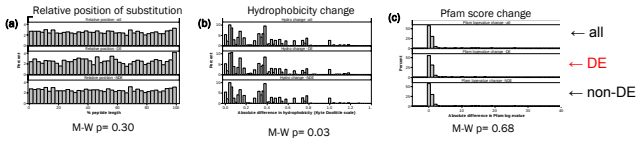


: associated nsSNP

DE and non-DE nsSNPs ~ same properties

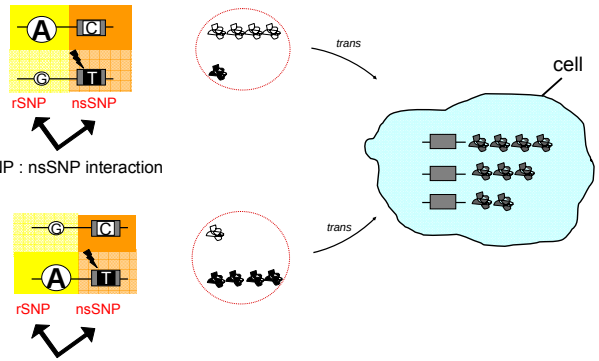
nsSNPs give rise to amino acid substitutions in the protein product

3 aspects of the amino acid substitution explored for DE and non-DE nsSNPs:

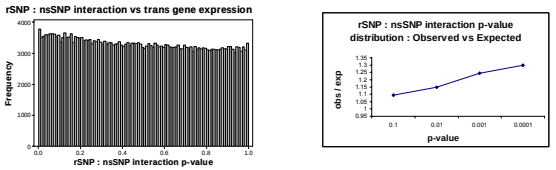


DE and non-DE nsSNPs have similar properties → if a random nsSNP has a functional effect, addition of an rSNP can amplify / mask that effect

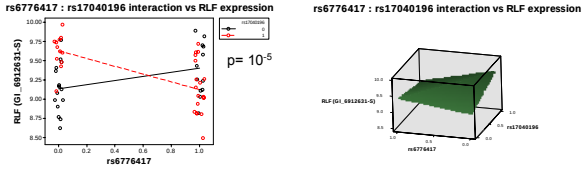
Interacting rSNP and nsSNPs have an epistatic impact on gene expression in *trans*



Multiple linear regression with interaction to test for effect on genome-wide expression arising from an rSNP : nsSNP interaction (CEU pop)



22 rSNP-nsSNP pairs tested against genome-wide expression



Example of rSNP : nsSNP interaction with effect on gene expression

Conclusions

1. Regulatory variants can magnify or mask the functional effect of protein-coding variants → importance of phasing.
2. Epistasis between rSNP-nsSNPs has an impact on gene expression.
3. Considering variants independently may be misleading.
4. Likely to reveal new medically important variants.