

ORAL PRESENTATION

Open Access

Genome wide association study of nonsynonymous Single Nucleotide Polymorphisms for seven common diseases

Praveen Surendran^{1,2*}, Alice Stanton^{1,2}, Denis Shields¹

From Sixth International Society for Computational Biology (ISCB) Student Council Symposium
Boston, MA, USA. 9 July 2010

Background

Associations of several Single Nucleotide Polymorphisms (SNPs) with common diseases like Coronary Artery Disease (CAD), Crohn's Disease (CD), Hypertension (HT), Bipolar Disorder (BD), Type 1 Diabetes (T1D), Type 2 Diabetes (T2D) and Rheumatoid Arthritis (RA) were identified in a study conducted by the Wellcome Trust Case Control Consortium (WTCCC) (1). WTCCC study compared the effects of genetic variations in 14,000 cases and 3000 shared controls and identified 24 independent associations with the diseases mentioned above using the genotype information of approximately 500,000 directly genotyped SNPs and genotype information simulated at 2.8 million loci studied by the International Hapmap Project(2). We hypothesize that there are more chances of finding association of rare SNPs with diseases by refined analysis of non-synonymous SNPs (nsSNPs) in genome wide association studies. In the present study we analyzed the association of 12,660 nsSNPs using a case control study in the WTCCC population.

Materials and methods

We simulated the genotypes at 10,798 nsSNP loci studied by the Stage 2 HapMap project using the genotype information from WTCCC for all 14,000 individuals studied for seven diseases and in 3000 controls. These simulations were done using the genetic recombination map of the respective regions obtained from the haplotypes of Hapmap European population. We performed these simulations or imputations using

two widely used programs called IMPUTE(3) and MACH. All the genotyped SNPs used to impute missing genotypes passed quality control tests for Hardy Weinberg equilibrium ($p < 10^{-2}$), Minor allele frequency ($MAF < 10^{-2}$), missing genotypes per marker (more than 10%) performed using programs in PLINK genome wide analysis package. Subsequent case control association of 10,798 imputed nsSNPs and 1,862 genotyped nsSNPs was performed using an additive model and genotype model in a frequentist and bayesian framework.

Results

We found 2 nsSNPs associated with BD, 2 with Coronary Artery Disease, 7 with CD, 1 with HT, 22 with RA, 17 with T1D and 2 with T2D. In total, 53 new associations with the seven diseases ($p < 5 \times 10^{-6}$) studied by WTCCC. We also developed a pipeline which summarizes quality control measures which should be considered to minimize false associations in genome wide association studies. In any such large scale genome wide association studies, there are chances of getting false positives which show association at loci imputed using genetic information from Hapmap. This can arise due to the genotype quality of tag SNPs which are in high linkage disequilibrium (LD) in the region where the missing genotype is simulated. Such false associations can be ruled out by visually inspecting the cluster plots of genotyped SNPs which are in high LD in respective regions. A comprehensive quality control will be performed at this stage by visually inspecting cluster plots of all genotyped SNPs which are in high LD with the imputed SNPs associated with each disease which will count out any such influences on new associations identified in our study.

* Correspondence: praveen.surendran@ucd.ie

¹School of Medicine and Medical Science, University College Dublin, Belfield, Dublin 4, Ireland

Full list of author information is available at the end of the article

Author details

¹School of Medicine and Medical Science, University College Dublin, Belfield, Dublin 4, Ireland. ²Department of Clinical Pharmacology, Royal College of Surgeons in Ireland, Dublin-1, Ireland.

Published: 7 December 2010

References

1. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007, **447**(7145):661-678.
2. The International HapMap Project. *Nature* 2003, **426**(6968):789-796.
3. Marchini J, Howie B, Myers S, McVean G, Donnelly P: A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007, **39**(7):906-913.

doi:10.1186/1471-2105-11-S10-O6

Cite this article as: Surendran *et al.*: Genome wide association study of nonsynonymous Single Nucleotide Polymorphisms for seven common diseases. *BMC Bioinformatics* 2010 **11**(Suppl 10):O6.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

