Multiple Single Nucleotide Polymorphisms analysis of candidate genes in Inflammatory Bowel Diseases by using RLS classifiers

Annese V (1), Latiano A (1), Palmieri O (1), D'Addabbo A (2), Maglietta R (2), Liuni S (3), Pesole G (3,4), Ancona N (2)

(1) U.O. Gastroenterologia, Ospedale CSS-IRCCS, San Giovanni Rotondo, Italy
(2) Istituto di Studi sui Sistemi Intelligenti per l'Automazione, CNR, Via Amendola 122/D-I, 70126 Bari, Italy
(3) Istituto di Tecnologie Biomediche-Sezione di Bari,CNR, Via Amendola 122/D, 70126 Bari Italy
(4) Dipartimento di Biochimica e Biologia Molecolare - Università di Bari, Via E. Orabona 4, 70126 Bari, Italy

Motivation

Crohn's disease (CD) and Ulcerative Colitis (UC) are related chronic inflammatory bowel disorders (IBD). The aetiology of IBD is still elusive, but recent studies have suggested that environmental and immunogenetic factors play an important role in the their development. Many susceptibility regions on different chromosomes have been recently pointed out and the correlation between some Single Nucleotide Polymorphisms (SNPs) and IBD has been investigated. In particular, three variants (R702W, G908R, L1007fs) in CARD15 gene on chromosome 16q12 have been associated with CD in Caucasian population. Moreover, some SNPs in DLG5 (chr 10q23) and OCTN1-2 (chr 5q31) genes have also been identified as correlated to IBD. More conflicting data have been reported for the TNF? and MDR1 genes. Since several genes could be involved, each with a limited impact and a possible gene-gene interaction, the traditional methods of analysis such as linkage and association studies could be ineffective. In order to fully understand the relation between many genetic markers and polygenic diseases, new multivariate methods have to be considered. Such methods have to take into account mutual interactions between multiple SNPs in the estimation of the correlation between genotype and phenotype. Statistical learning theory provides valuable methods which are able to jointly consider multiple SNPs and to detect their correlation with pathology. In this paper we quantify the correlation between a set of SNPs and IBD by using the prediction accuracy of classifiers trained on a finite number of examples.

Methods

We have used Regularized Least Square (RLS) classifiers to measure the correlation existing between a set of SNPs and CD or UC. To this end, we have considered 127 CD case, 127 UC cases and 127 healthy controls. We have tested 12 SNPs on 6 different genes: R702W, G908R and L1007fs on CARD15 gene, -857C>T and -308G>A on TNFα gene, C3435T and G2677T/A on MDR1 gene, SLC22A4 and SLC22A5 on OCTN genes, rs124869 on DLG5 gene, and IGR2096Ms1 and IGR2198a on chromosome 5.

Results

We showed that RLS classifiers with second degree polynomial kernels have prediction accuracy of 58.6% in the classification of CD samples versus healthy controls. Moreover, we found a prediction of 50.8%, obtained with linear RLS, in the classification of UC examples versus healthy controls, indicating that the selected SNPs set is more correlated with CD than UC.

Contact email: ancona@ba.issia.cnr.it