BIOINFORMATIC APPROACHES FOR CANDIDATE GENE SELECTION IN INFLAMMATORY BOWEL DISEASES

Mitja Mitrovič,[1, 2] *Uroš Potočnik*[1, 2]

[1] Center for Human Molecular Genetics and Pharmacogenomics, Medical Faculty
[2] Laboratory for Biochemistry, Molecular Biology and Genomics, Faculty for Chemistry and Chemical Engeneering, University of Maribor, Slovenia; uros.potocnik@uni-mb.si

Introduction

The human inflammatory bowel diseases (IBD), usually classified into Crohn's disease (CD) and ulcerative colitis (UC), are complex diseases where environmental factors and several genes play role. In addition to genes coding for metabolizing enzymes and drug transporters, genes involved in risk and disease pathogenesis could predispose to disease progression and therapy outcome. Recently, mutations in NOD2/CARD15 gene on chromosome 16 were associated with IBD. Linkage analysis in IBD families suggested at least 8 other chromosomes to be involved in IBD and IBD genes within these chromosomal candidate regions need to be identified.
The aim of our study was to design and evaluate the efficiency of different bioinformatic approaches for identification of genes associated with IBD.

Methods

We used several bioinformatics approaches including literature data mining tools, microarray expression data analysis and pathway analysis to create candidate gene lists. The different inclusion criteria for genes to be listed were: 1. chromosomal location within IBD linkage region; 2. up or downregulation in colon samples from IBD patients; 3. proven role in pathways known to be involved in pathogenesis of autoimmune response; 4. positive selection using literature data mining tools SNP3D (http://www.snps3d.org/), GeneCards (http://www.genecards.org) and Bitola (http://www.mf.uni-lj.si/bitola). The candidate gene lists were compared and best candidate genes were scored according to their appearances on candidate gene lists. The final gene list was evaluated with disease association studies in patients. We used GDPinfo query tool for association studies (http://apps.nccd.cdc.gov/genomics/GDPQueryTool/SearchByDisease.asp) and PubMed for comprehensive search of all genes tested for association in IBD patients and controls.

Results

Eleven independent linkage studies reviled 10 (IBD1-IBD10) confirmed candidate regions on 9 different chromosomes ranged from 6,6 Mbp (IBD 2) to 28,5 Mbp (IBD6), having from 24 (IBD2) to 646 (IBD6) candidate genes within significant LOD score. Four microarray expression studies reviled 536 of 22 000 analysed genes up or down regulated in IBD patients, 25 differently expressed genes were replicated in all four studies. We used several different search profiles with literature data mining tools and found 745 candidate genes, 17 of which selected with more than 80 % of search profiles used. Fifteen genes involved in dendritic cell migration proved to be impaired in IBD patients were listed on the pathway candidate gene list. Comparison of all candidate gene list reviled 10 genes fulfilling more than 80 % and 25 additional genes fulfilling more than 50 % of all selection criteria. Twenty of 35 genes on the final candidate list were already tested in IBD patients and 14/20 genes showed positive association with IBD. In addition 15 novel candidate genes were identified which need to be tested in IBD patients. Genes confirmed in several independent association studies include genes associated with Crohn disease (NOD2/CARD15, OCTN1, OCTN2), with ulcerative colitis (MDR1/ABCB1) or with both diseases (DLG5, HLA, ICAM-1, TLR4).

Conclusions

Described bioinformatics approaches were able to efficiently identify known candidate genes associated with IBD. In addition novel potential candidate genes were identified which need to be tested in patients. The identification of novel susceptibility genes will lead to better understanding disease pathophysiology, discovery of new therapeutic targets and in better disease management.