Analysis of Conserved Sequence Tags in the molecular diagnosis of cystic fibrosis

Boccia A[2], Elce A[2], Tomaiuolo R[1,2], Castaldo G[1,2], Paolella G[1,2]

[1] Universita` degli Studi di Napoli 'Federico II', Napoli
[2] CEINGE Biotecnologie Avanzate, Napoli

Motivation

Conserved sequence tags (CSTs) are genomic sequences considered as a potential resource for the discovery of novel functional elements, as they are generally believed to be the result of selective evolutionary pressure for conservation. The analysis of such sequences within the context of genes related to genetically transmitted diseases, may result in the identification of novel elements involved in the molecular mechanisms underlying the disease. Cystic fibrosis is a well known monogenic disease, caused by reduced activity of the CFTR chloride channel. Molecular diagnosis is based on the identification of mutations affecting the coding sequence in both alleles. In 10% of CF patients it is not identified a mutation in the exonic sequence. This sub-group of CF patients is a good candidate for CST analysis.

Methods

CSTs were selected by using an extension of a pipeline, previously developed in our laboratory (Nucleic Acids Res. 2005; 33 - D505-10). A protocol was implemented for selecting PCR primers, used to analyze the CFTR genomic region of CF patients and controls. A sql database was used to hold data from the analysis: all the details of the analyzed genomic regions; additional information concerning the CFTR genotype and phenotype of the source subjects; information about CFTR wild type sequence, exons and known SNPs, taken from the ENSEMBL (http://www.ensembl.org) web site; the list of identified polymorphisms. A procedure was created for the identification of haplotypes, based on the algorithm of Haplotyper (Am. J. Hum. Genet. 2002; 70:157–169), which uses a Monte Carlo approach. The web site, used to access the data, has been developed by using the PHP scripting language, and includes the tools used to identify and classify mutations, to identify haplotypes, and to perform some basic statistical analyses.

Results

In this study we carried out a low level analysis of the CSTs contained within the intronic regions of the human CFTR gene, leading to the identification of many polymorphic sites. Each variation was evaluated on the basis of the list of previously known single nucleotide polymorphisms (SNPs) and, where data were available, according to their frequency within the Caucasian population. The analysis originally included 56 intronic CSTs from 70 CF patients and 20 control subjects, resulting in the identification of 60 different polymorphic sites. According to their frequency, some mutations are randomly distributed among the tested subjects, with no correlation to pathology. On the other hand some mutations tend to concentrate in a subset of patients, characterized by absence of mutations within the coding sequence of CFTR, and suggesting a potential role played by the CST, possibly related to the disease. To further characterize this subset of polymorphisms, additional 60 control subjects have been involved in the study and results are currently under evaluation. The study led also to the identification of three novel polymorphic repeats which are now successfully used to improve segregation analysis for cystic fibrosis. Interestingly, about a quarter of the CSTs are strongly conserved regions, with the wild type sequence in all subjects, suggesting a potential functional role at least for some of them. The study included linkage disequilibrium analyses, which involves the identification of haplotypes by using inference methods, in view of a possible correlation with the presence of disease in subjects apparently missing causing mutations.

Contact : boccia@ceinge.unina.it