CST analysis in the characterization of cystic fibrosis patients

ID - 169

Boccia Angelo[1,2], Elce Ausilia[2], Tomaiuolo Rossella[1,2], Castaldo Giuseppe[1,2], Paolella Giovanni[1,2]
[1]Università degli Studi di Napoli 'Federico II', Napoli
[2]CEINGE Biotecnologie Avanzate, Napoli

Motivation

Conserved sequence tags (CSTs) are genomic sequences, generally believed to be the result of selective evolutionary pressure for conservation due to a possible functional role, and are therefore considered as a potential resource for the discovery of novel functional elements. The study 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

Selection of CSTs was carried out by using an extension of a pipeline, previously developed in our laboratory (Nucleic Acids Res. 2005; 33 - D505-10). Basically sequences were compared by BLASTZ, and sequences showing at least 70% identity over at least 100 bp were extracted and fully annotated. A protocol was devised for selecting PCR primers, able to cover all CST containing genomic segments while keeping at a minimum the number of multiplex reactions and the synthesized oligonucleotides. Primers selected in this way were synthesized and used to analyze the CFTR genomic region of CF patients and controls.

A PostgreSql database was used to hold data from the analysis, including all the details of the analyzed genomic regions, and additional information concerning the CFTR genotype and phenotype of the source subjects. Information about CFTR wild type sequence, exons and known SNPs were taken from the ENSEMBL (http://www.ensembl.org) web site.

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 the mutations.

Results

A low level analysis of the CSTs contained within the intronic regions of the human CFTR gene was carried out. The study led to the identification of many polymorphic sites within the gene. A specific tool was created for the characterization of the polymorphic sites, where the tested sequences are compared to the wild type ones and polymorphic sites are identified, highlighted and named according to the current standards. In addition, each variation is evaluated on the basis of the list of previously known single nucleotide polimorfisms (SNPs) and, accordingly, classified either as a novel or as an already reported one. The results of this analysis are collected in a database and made accessible through a web interface.

Analysis of 56 intronic CSTs from 70 CF patients and 20 control subjects resulted in the identification of 76 different polymorphic sites, including novel and previously described ones. According to the frequency of mutations, CSTs can be classified in different groups. About a quarter of the CSTs are strongly conserved regions, showing the wild type sequence in all subjects. Other CST regions contain mutation(s) either randomly distributed among the tested subjects, with no correlation to pathology, or limited to specific subsets of subjects. A number of these mutations are probably located on specific pathologic alleles, being typically found in patients carrying one or two alleles containing nonsense or missense mutations. Some mutations tend to concentrate in a subset of patients, carrying alleles characterized by absence of mutations within the coding sequence, and are therefore suggestive of a potential functional role played by the CST containing sequence, possibly related to the disease.

Email: boccia@ceinge.unina.it