

## Identification of Disease Causing DNA Motifs in Human DNA Using Clustering Approach

**Authors :** G. Tamilpavai, C. Vishnupriya

**Abstract :** Studying DNA (deoxyribonucleic acid) sequence is useful in biological processes and it is applied in the fields such as diagnostic and forensic research. DNA is the hereditary information in human and almost all other organisms. It is passed to their generations. Earlier stage detection of defective DNA sequence may lead to many developments in the field of Bioinformatics. Nowadays various tedious techniques are used to identify defective DNA. The proposed work is to analyze and identify the cancer-causing DNA motif in a given sequence. Initially the human DNA sequence is separated as k-mers using k-mer separation rule. The separated k-mers are clustered using Self Organizing Map (SOM). Using Levenshtein distance measure, cancer associated DNA motif is identified from the k-mer clusters. Experimental results of this work indicate the presence or absence of cancer causing DNA motif. If the cancer associated DNA motif is found in DNA, it is declared as the cancer disease causing DNA sequence. Otherwise the input human DNA is declared as normal sequence. Finally, elapsed time is calculated for finding the presence of cancer causing DNA motif using clustering formation. It is compared with normal process of finding cancer causing DNA motif. Locating cancer associated motif is easier in cluster formation process than the other one. The proposed work will be an initiative aid for finding genetic disease related research.

**Keywords :** bioinformatics, cancer motif, DNA, k-mers, Levenshtein distance, SOM

**Conference Title :** ICBBE 2018 : International Conference on Biomaterials and Biomedical Engineering

**Conference Location :** Mumbai, India

**Conference Dates :** February 22-23, 2018