Abstract

Variant identification is a fundamental part in the analysis of genetic diseases. Variants are the alterations which occur in the arrangement of nucleotide in the DNA sequence. Genetic diseases are caused by variations occurring in genes which may cause change in protein, affecting the survival and adaptation of an individual. A number of computational techniques are applied to identify these variant. Precise diagnosis of genetic diseases is important for proper treatment of patients and to determine explicit prevention strategies. Introduction of next generation sequencing (NGS) techniques in the past have made large number of DNA sequences easily available. This has made variant identification using NGS data a area of interest. This paper briefly discussed the analysis steps followed for NGS data analysis. This paper later explains in detail a few approaches that are used for identifying variants such as Support vector machine based approach, Machine learning based approach, MOSAIK: hash-base approach, Bayesian statistical based approach, JointSLM based approach.

References
60. Kvitek, S. G. 2013. Whole genome, whole population sequencing reveals that loss of
signaling networks is the major adaptive strategy in a constant environment. PLoS Genet.
9(11), 1003972.
sequencing data by using JointSLM algorithm. Nucleic acids research. 068.
depth of coverage. Genome. 19, 1586-1592.
63. McCarroll, M. K. 2008. Integrated detection and population-genetic analysis of SNPs

Index Terms

Computer Science

Artificial Intelligence

Keywords

Variants, Variations, Mutations, Genetic Disease, Variant Identification.