CNV detection using RF Classifier

A PROJECT REPORT
Submitted By

Prasoon Goswami, Roll No - 12616002034

Under the supervision of
Prof. Rituparna Sinha
Department of Information Technology

In partial fulfilment for the award of the degree
Of
BACHELOR OF TECHNOLOGY
In
INFORMATION AND TECHNOLOGY

HERITAGE INSTITUTE OF TECHNOLOGY, KOLKATA
An autonomies Institute under
Maulana Abul Kalam Azad University of Technology
Formerly Known as a
West Bengal University of Technology
July 2020

ACKNOWLEDGEMENT

I would take this opportunity to thank Prof. Pranay Chaudhury, Principal Heritage Institute of Technology for providing me with all the necessary facilities to make our project work a success.

I would like to thank our Head of the Department Prof. Siuli Roy for her kind assistance as and when required.

I will be thankful to Prof. Rituparna Sinha my project coordinator for constantly supporting and guiding me for giving me invaluable insights. Her guidance and her words of encouragement motivated me to achieve my goal and impetus to excel.

I thank my Faculty members and Laboratory assistants at the Heritage Institute of Technology for paying a pivotal and decisive role during the development of the project. Last but not least I thank all my friends for their cooperation and encouragement that they have bestowed on me.

HERITAGE INSTITUTE OF TECHNOLOGY, KOLKATA
An autonomous Institute under
Maulana Abul Kalam Azad University of Technology
Formerly Known as a
West Bengal University of Technology

BONAFIDE CERTIFICATE

Certified that this project report "CNV DETECTION USING RF CLASSIFIER" is the bonafide work of PRASOON GOSWAMI, who carried out under my supervision.

SIGNATURE
PROF. SIULI ROY
HEAD OF THE DEPARTMENT
Department of Information Technology

Heritage Institute of Technology

SIGNATURE
PROF. RITUPARNA SINHA
PROJECT GUIDE
Department of Information Technology

Heritage Institute of Technology

SIGNATURE
EXTERNAL EXAMINER

ABSTRACT

Copy number variants (CNV) are associated with phenotypic variation in several species. However, properly detecting changes in copy numbers of sequences remains a difficult problem, especially in lower quality or lower coverage next-generation sequencing data. Here, inspired by recent applications of machine learning in genomics, we describe a method to detect duplications and deletions in short-read sequencing data. In low coverage data, machine learning appears to be more powerful in the detection of CNVs than the gold-standard methods of coverage estimation alone, and of equal power in high coverage data. We also demonstrate how replicating training sets allows more precise detection of CNVs, even identifying novel CNVs in two genomes previously surveyed thoroughly for CNVs using long-read data.

List of tables, graphs and figures

FIGURES

TABLES

Table of Contents