World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:17, No:01, 2023

Predicting Open Chromatin Regions in Cell-Free DNA Whole Genome Sequencing Data by Correlation Clustering

Authors : Fahimeh Palizban, Farshad Noravesh, Amir Hossein Saeidian, Mahya Mehrmohamadi

Abstract : In the recent decade, the emergence of liquid biopsy has significantly improved cancer monitoring and detection. Dying cells, including those originating from tumors, shed their DNA into the blood and contribute to a pool of circulating fragments called cell-free DNA. Accordingly, identifying the tissue origin of these DNA fragments from the plasma can result in more accurate and fast disease diagnosis and precise treatment protocols. Open chromatin regions are important epigenetic features of DNA that reflect cell types of origin. Profiling these features by DNase-seq, ATAC-seq, and histone ChIP-seq provides insights into tissue-specific and disease-specific regulatory mechanisms. There have been several studies in the area of cancer liquid biopsy that integrate distinct genomic and epigenomic features for early cancer detection along with tissue of origin detection. However, multimodal analysis requires several types of experiments to cover the genomic and epigenomic aspects of a single sample, which will lead to a huge amount of cost and time. To overcome these limitations, the idea of predicting OCRs from WGS is of particular importance. In this regard, we proposed a computational approach to target the prediction of open chromatin regions as an important epigenetic feature from cell-free DNA whole genome sequence data. To fulfill this objective, local sequencing depth will be fed to our proposed algorithm and the prediction of the most probable open chromatin regions from whole genome sequencing data can be carried out. Our method integrates the signal processing method with sequencing depth data and includes count normalization, Discrete Fourie Transform conversion, graph construction, graph cut optimization by linear programming, and clustering. To validate the proposed method, we compared the output of the clustering (open chromatin region+, open chromatin region-) with previously validated open chromatin regions related to human blood samples of the ATAC-DB database. The percentage of overlap between predicted open chromatin regions and the experimentally validated regions obtained by ATAC-seq in ATAC-DB is greater than 67%, which indicates meaningful prediction. As it is evident, OCRs are mostly located in the transcription start sites (TSS) of the genes. In this regard, we compared the concordance between the predicted OCRs and the human genes TSS regions obtained from refTSS and it showed proper accordance around 52.04% and ~78% with all and the housekeeping genes, respectively. Accurately detecting open chromatin regions from plasma cell-free DNA-seq data is a very challenging computational problem due to the existence of several confounding factors, such as technical and biological variations. Although this approach is in its infancy, there has already been an attempt to apply it, which leads to a tool named OCRDetector with some restrictions like the need for highly depth cfDNA WGS data, prior information about OCRs distribution, and considering multiple features. However, we implemented a graph signal clustering based on a single depth feature in an unsupervised learning manner that resulted in faster performance and decent accuracy. Overall, we tried to investigate the epigenomic pattern of a cell-free DNA sample from a new computational perspective that can be used along with other tools to investigate genetic and epigenetic aspects of a single whole genome sequencing data for efficient liquid biopsy-related analysis.

Keywords : open chromatin regions, cancer, cell-free DNA, epigenomics, graph signal processing, correlation clustering
Conference Title : ICBMMA 2023 : International Conference on Bioinformatics Models, Methods and Algorithms
Conference Location : New York, United States
Conference Dates : January 30-31, 2023