DNALoopR: A Novel High-Performance Machine Learning Predictor to Identify Genome-Wide 3D DNA Interactions in Cancer

Naidu, Prathik

Studying the 3D topology of the genome, in which DNA interacts to form loops, has become a promising field for building a better understanding of tumor biology. Although identifying 3D genome alterations can help determine the underlying regulation of cancer-associated biological processes, current lab-based approaches to determine these interactions are costly, time-consuming, and prone to error. In this research, a machine learning software, called DNALoopR, was developed to more efficiently search for DNA interactions. ChIP-Seq data from the MCF7 (breast), HeLa (cervical), and K562 (leukemia) cancer cell lines were used to engineer features that capture the biology behind loop formation. The prediction framework utilized an ensemble approach that incorporated k-means clustering and background noise adjustment algorithms to improve performance. After training and cross-validation, DNALoopR achieved an accuracy of 92.8% with a low false positive rate of 4.5%, significantly outperforming existing methods. The tool was validated based on prioritization of features important for loop formation, comparisons with literature-mined experimental results, and confirmation using somatic mutation data. Furthermore, DNALoopR was applied to liver cancer data and identified previously undiscovered mechanisms behind the disruption of well-known cancer-associated genes and pathways such as CDKN2A and RAS. Ultimately, this project conducted the most comprehensive computational characterization of the 3D genome to date and presents the first-ever software for genome-wide prediction of high-resolution DNA interactions. DNALoopR can provide deeper insights into the genetic basis of oncogenesis and assists in the discovery of targetable pathways for next-generation cancer therapeutics.

Awards Won:

Intel ISEF Best of Category Award of $5,000

Serving Society Through Science: First Award of $1000

Association for the Advancement of Artificial Intelligence: Honorable Mention