**DALMNPred: A machine learning based approach to predict novel lncRNA-miRNA interactions from multi-omics data**

**Introduction**

The identification of ncRNA-mediated regulatory networks is pivotal to understand the mechanistic insights underlying the core biological processes. However, experimental investigation of these regulatory axes is labor intensive, costly and time-consuming. We have designed an integrated machine learning-based pipeline,DALMNPred (Disease-associated LncRNA-miRNA network Prediction), to identify lncRNA-associated competing endogenous RNA (**ceRNA**) networks from multi-omics data. Probe repurposing strategies for extraction of lncRNA signatures on publicly available GEO datasets enabled us to identify differentially expressed lncRNAs (**DElncRNAs**). Further, an optimum set of lncRNAs that could distinguish between high HbF and normal conditions were predicted from these **DElncRNAs** using supervised machine learning and a stepwise selection model. The predicted lncRNA-miRNA-mRNA axes provide novel insights into lncRNA-mediated regulatory mechanisms and possible therapeutic avenues for HbF reactivation. Taking fetal hemoglobin regulation as an example, our study provides novel computational pipeline to unravel the ceRNA networks in various pathophysiological processes. The methodology described here can easily be employed to other complex diseases, and will benefit mechanistic study and lncRNA-mediated drug development.

**Requirements**

python 3.7= <https://www.python.org/downloads/release/python-370/>

R packages= <https://www.rstudio.com/products/rpackages/>

**For differentially expressed lncRNAs identification**

The Significance Analysis of Microarrays (SAM) was performed using the R package. The differentially expressed lncRNAs were imported in Cluster view 3.0 [27] to carry out the Hierarchical Cluster Analysis (HCA).

All the codes for differentially expressed lncRNA analysis are given in File 1.

**Machine learning classification**

Supervised machine learning algorithms, namely Random Forest and Support Vector Machines were used for classification.

All the codes of ML classifications are uploaded in File 2.

**Contact**

Please direct your questions to: Dr. Nishant Chakravorty, [nishant@smst.iitkgp.ac.in](mailto:nishant@smst.iitkgp.ac.in)