Bridging the Phenotype-Genotype Gap for disease prognosis

Mathew Palakal (PI), Meeta Pradhan (Co-PI), Shruti Sakhare

{mpalakal, mpradhan, ssakhare }@iupui.edu

Indiana University School of Informatics at IUPUI
Bioinformatics Program
TiMAP: Text information Mining, Analysis and Prediction Laboratory
719 Indiana Avenue
Indianapolis, IN 46202 USA

Abstract

A well-known question we are trying to solve since past two decades is "What is the relationship between genotypes and phenotypes?". Currently, methods such as Genome Wide Association Studies (GWAS) and Gene Regulatory Networks (GRNs) are used to find these phenotype and genotype relationships using statistics and molecular biology respectively. These studies mainly focus on studying limited phenotypes for direct mapping. However it has been reported that disease traits are outcome of many interdependent changes in phenotype. Our study aims to use the extensive clinical and genotype data from publicly available databases to study this interdependency of clinical outcomes and the corresponding changes at gene expression pattern.

The present work of understanding genotype-phenotype relationship across different stages is designed based on the available TCGA data for breast cancer. The clinical features were identified and classified based on the laboratory and other clinical parameters. We selected 60 phenotypes based on their importance reported in literature and these were clustered for their significance for cancer prognosis and their expression at different stages. Multivariate statistical analysis is performed for the outliers from the clusters to identify the interdependency of their expression. An expression profile of these outliers is obtained based on the analysis performed.

The analysis shows the significant phenotypes expressed in different stages of breast cancer. Some of these significant phenotypes are the ones, previously reported for breast cancer prognosis. However, the clustering analysis identified new phenotypes that may play a significant role in breast cancer prognosis. Correlation study for these parameters can then identify relational expression of multiple clinical traits.

Following this study, these genotype features will be analyzed for their SNP, CNV variants for these parameters to bridge the genotype-phenotype gap. By successfully identifying the molecular changes at gene level for such phenotypic diversity of clinical traits it can be made possible to predict the onset of disease at an early stage. Current methodology can then be extended for other disease studies.