CLIQUES FOR IDENTIFICATION OF GENE SIGNATURES FOR COLORECTAL CANCER ACROSS POPULATION

Meeta P. Pradhan, **Kshithija Nagulapalli**, Mathew J. Palakal
IU School of Informatics, Indiana University–Purdue University Indianapolis, Indianapolis, Indiana 46202

Introduction: Colorectal cancer (CRC) is one of the most common cancers diagnosed worldwide. Studies have correlated CRC with dietary habits and environmental conditions. We developed a novel network based approach where cliques and their connectivity profiles explained the variation and similarity in CRC across four populations- China, Germany, Saudi Arabia and USA.

Methods: Networks generated after data preprocessing were analyzed individually based on topological and biological features. Using greedy algorithm, cliques of various sizes were identified in each network and size 7 cliques were further analyzed based on their clique connectivity profile (CCP). Our algorithm considered the interaction of cliques based on two parameters: (i) Identification of common (links) genes; (ii) CliqueStrength. The cliques were evaluated by two conditions (a) Maximum number of common genes across cliques and highest CliqueStrength and (b) Minimum number of common genes across cliques and highest CliqueStrength.

Results: Large numbers of genes are found to be common between USA, China and Germany. Highly scored nodes based on topological parameters are TP53, SRC, ESR1, SMAD3, GRB2, CREBBP, EGFR, SMAD2, and CSN2KA1. Signal transduction, protein phosphorylation etc., were found to be important GO biological processes. Number of unique size 7 cliques identified in all the population is 650. 49 common cliques identified included genes- EGFR, GRB2, PIK3R1, PTPN6, BRCA1, SMAD2, TP53, CSN2 etc. We found 20 cliques that are uniquely identified for USA, 10 for Germany and one for China. Cliques include genes that are both well studied, less-studied in CRC; but are targets in other cancers.

Conclusion: With CCP, we were able to identify commonality, uniqueness and divergence among the populations. Furthermore, comparing all cliques (their CCP) as gene-signatures across populations can help to identify efficient drug targets. Results were consistent with other studies and demonstrate the power of cliques to study CRC across populations.