"Identification and Pattern Analysis of SNPs Involved in Colorectal Cancer"

Praneti Patidar [*] and Jyoti Bhojwani [*,#]

Bio-Informatics Lab, School of Computer Science and Information Technology*,

School of Life Sciences[#], DAVV/Indore University, Indore-452001, INDIA.

Corresponding author: Dr. Jyoti Bhojwani (jbhojwani2005@gmail.com)

Abstract:

Colorectal cancer (CRC) is the second leading cause of cancer related deaths globally posing a lifetime risk of 80-100% in every individual. Genetics and relevant mechanisms underlying some key signaling pathways like *Wnt, TGF, p53, K-ras* etc. play a detrimental role in governing the predisposition for CRC. A high percentage of colorectal tumors (adenomas and carcinomas) show activating mutations in *beta-catenin* or *axin*, whereas, loss of certain tumor suppressor genes (TSGs), like *APC* cause the initiation of random polyps in the colon. All of these molecules incidentally are critical components of an evolutionarily conserved *Wnt* signaling pathway, which is instrumental at various time points in the development of this disease. Differences in SNP profiles amongst sample groups in the genomic landscape can be recognized through a smart and efficient use of machine learning techniques. The statistics and pattern analyses of these SNP profiles, interestingly provides us with a concrete and logical platform upon which, relative contribution/s of each unique SNP, ranging "from cause to effect" can be significantly assessed. The biological relevance of these SNP variations with respect to cancer prediction and predisposition, however, remains to be resolved, pending a better understanding of the impact of rational control design in SNP studies. Our results emerging from the analyses of significant SNP's reported here, demonstrates the utility of relevant bioinformatics tools and machine learning techniques in discriminating diseased populations based on realistic SNP data. In this study, we have primarily targeted critical members of *Wnt* signaling pathway, which play important developmental role/s during different stages of colorectal cancer, depicting a classical "multigene-multistep nature" of cancer. We have identified and related common genetic variants for the "early-acting" and "late-acting" members of this pathway, that are most prevalent in patients with CRC disease. Complex relationships and correlations hidden in large data sets have been dug and analyzed here, by deploying various data-mining techniques.

Keywords: Wnt-signaling, SNPs, predisposition, datamining, oncogenomics, pharmacogenomics