

Pattern analysis and identification of snps involved in colorectal cancer

Jyoti Bhojwani^{1, 2*} and Praneti Patidar¹

¹Bio-Informatics Lab, School of Computer Science and Information Technology, DAVV/Indore University, Indore-452001, India

²Bio-Informatics Lab, School of Computer Science and Information Technology, School of Life Sciences, DAVV/Indore University, Indore-452001, India

Abstract:

Colorectal cancer (CRC) is the term used to refer to cancers located in the colon or rectum that together form the large intestine. Both men and women are affected by colorectal cancer, one of the most prevalent cancers in the United States. The medications and side effects of this disease will affect the quality of your life.

Colorectal cancer (CRC) is the second leading cause of cancer-related deaths with a lifetime risk of 80-100 per cent for each person worldwide. Approximately 1.4 million new cases of colorectal cancer were identified worldwide in 2012, comprising 10 per cent of all new cancer cases. Survival of colorectal cancer depends on the stage at which it is diagnosed, with worse survival in later-stage diagnoses. The five-year survival rate is 90% for diagnosed early-stage colorectal cancers, compared to 13% for those diagnosed late-stage.

In terms of prevalence, colorectal cancer ranks fourth among males in frequency, with about 550 000 new cases worldwide a year, and seventh among females, with an estimated 473 000 new. Among males the age-standardized incidence is 20.1 per 100 000 a year and among females is 14.6. Mortality is roughly half that of about 530 000 deaths in 2002 (males, 280 000; females, 250 000). For an estimated 2.8 million people living within 5 years of diagnosis, colorectal cancer accounts for 11.5 per cent of all prevalent cancer cases, the second most prevalent cancer in males following prostate cancer (1 515 000 cases) and the third most prevalent cancer in females following breast and cervix uteri (1 315 000). The precise cause of colorectal cancer is not clear, although certain risk factors, including diet, cigarette use and heavy alcohol consumption, are strongly related to the disease. Individuals with other genetic cancer syndromes or a family history of colorectal cancer may have a high chance of developing the condition.

The Wnt signaling pathway is a conserved evolutionary signaling pathway that regulates several events during embryogenesis. The Wnt pathway, at the cellular level, controls morphology, proliferation, motility and cell destiny. Wnt signaling pathway also plays a central role during tumorigenesis, and improper regulation of this pathway is found especially in several human cancers. Wnt signaling pathway is important for several biological processes such as stem cell renewal / differentiation and multiple studies of this pathway have resulted in the discovery of several critical components over the past few years. Nevertheless, several of the processes involved in triggering or disabling this specific pathway still remain to be elucidated.

Genetics and relevant mechanisms underlying some key signaling pathways like Wnt, TGF, p53, Kras 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.

A single-nucleotide polymorphism (SNP, pronounced as: snip) is a difference in the DNA sequence that occurs when a single nucleotide — A, T, C, or G — varies in the genome (or other shared sequence) between members of a species (or paired chromosomes in an individual). Numerous mutations can be identified in cancer cells / tissues, but the punch line will be to discover new drugs and therapies that can recognize the variants that actually cause the Genome-Wide Association Studies (GWAS) disease and search the genome and find certain genetic variants (SNPs) that are predominant in particular disease patients.

Thus, SNPs serve as potential and invaluable markers (landmarks in the genome) for association-based approaches to discover the genetic components bearing complex traits. Furthermore, where large sample sizes are required, their bi-allelic nature is amenable to high throughput automated genotyping. SNPs have been of particular interest, also to the evolutionary biologists.

High SNP density genomic areas carry specific roles, i.e. they may have specific gene and structural material. Furthermore, SNPs are valuable genetic markers for family-based Mendelian disease association research, including population history and/or genetics, and medicine personalization. Previous SNP-based efforts focussed on (1) specific disease candidate genes, (2) expressed sequence marker genes, or (3) genomic sequences. Consequently, SNP profiles and patterns that find considerable utility in identifying a comprehensive set of genes that contributes to the development and susceptibility of complex diseases like cancer. The most prevalent approaches in these projects were the mining of SNPs from either EST Databases or DNA-Sequence Databases.

Differences in SNP profiles among sample groups within the genomic landscape can be recognized by using machine learning techniques smartly and efficiently. Interestingly, the statistics and trend analyses of these SNP profiles provide us with a clear and rational basis on which to substantially evaluate the relative contribution / s of each particular SNP, ranging from cause to effect. However, the biological importance of these SNP differences regarding cancer prediction and predisposition remains to be addressed, awaiting a clearer understanding of the influence of the nature of rational control in SNP studies.

Conclusion

Our findings show the usefulness of specific bioinformatics tools and machine learning techniques in distinguishing diseased populations based on practical SNP data. In this research, we specifically addressed essential members of the Wnt signaling pathway, which play significant

developmental roles during various stages of colorectal cancer, illustrating a classic "multigene-multistep form" of cancer. By harnessing the power of developmental biology methods, we have identified and linked specific genetic variants for the "early-acting" and "late-acting" members of this pathway, which are most prevalent in patients with CRC

disease. Furthermore, complex relationships and associations hidden in large data sets were dug and analysed by deploying specific data mining (bioinformatics) technique. This study addresses the application of such a combinatorial approach in translational science and clinical medicine therapies, by defining several possible candidate therapy goals.