

Identification and Pattern Analysis of SNPs Involved in Colorectal Cancer

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

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 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 SNPs 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, by harnessing the power of developmental biology tools. In addition, complex relationships and correlations hidden in large data-sets have been dug and analyzed here, by deploying various data-mining (bioinformatics) techniques. The report discusses the scope of such a combinatorial approach, by identifying some potential candidate targets of therapy, in translational research and clinical medicine interventions.

Keywords: *Wnt*-signaling; SNPs; Predisposition; Datamining; Oncogenomics; Pharmacogenomics

Introduction

The *Wnt* signaling pathway

The *Wnt* signaling pathway is an evolutionary conserved signaling pathway, which controls many events during embryogenesis. At the cellular level, the *Wnt* pathway regulates morphology, proliferation, motility and cell fate. Also during tumorigenesis, *Wnt* signaling pathway plays a central role and an inappropriate regulation of this pathway is notably observed in several human cancers [1]. *Wnt* signaling pathway is essential to many biological processes like stem cell renewal/differentiation [2] and numerous studies of this pathway over the last few years have led to the identification of several critical components. Nevertheless, many of the mechanisms involved in activation or inactivation of this particular pathway still remain to be elucidated.

Colorectal cancer

Colon cancer is carcinoma (cancer; epithelial in origin) of the large intestine (colon), the lower part of the digestive system. Rectal cancer is a carcinoma of the last several inches of the colon. Together, they are often referred to as colorectal cancers. Most cases of colon cancer begin as small, noncancerous (benign) clumps of cells called adenomatous polyps or “adenomas”. Over time, some of these polyps become colon cancers (carcinomas) following a multigene-multistep process [3]. More than half a million people worldwide die of Colorectal Cancer (CRC), making CRC, the second largest cause of death by cancer, owing to its genetic epidemiology [4]. A high percentage of tumors show activating mutations in *beta-catenin* or *axin*. Mutations in *APC* and *beta-catenin*, a key mediator of *Wnt* signaling are found in wide majority of sporadic

colon cancers [5,6]. These molecules are crucial components of the *Wnt* signaling pathway (see Figure 1). Inactivating mutations of the *Wnt* signaling pathway inhibitor *APC* (a tumor-suppressor gene) are associated with the colon cancer susceptibility syndrome, whereas, *Familial Adenomatous Polyposis-FAP* [8] is a dominant form of mutation/prominent genetic alteration, which is present in early premalignant lesions in the intestine, such as aberrant crypt foci and small adenomas. Thus, the chances that colon cancer can indeed be cured or stalled, needs effective measures to inhibit/block this particular stage of development (*ie.*, early polyps).

Single nucleotide polymorphisms (SNPs)

A single-nucleotide polymorphism (SNP, pronounced as: *snip*) is a DNA sequence variation occurring when a single nucleotide — A, T, C, or G — in the genome (or other shared sequence) differs between members of a species (or between paired chromosomes in an individual). Numerous mutations can be found in the cancer cells/tissues, but the punch-line would be to develop new treatments and therapies that can identify the variants, which in fact cause the disease

***Corresponding author:** Dr. Jyoti Bhojwani, Bio-Informatics Lab, School of Computer Science and Information Technology, School of Life Sciences, DAVV/Indore University, Indore-452001, India, E-mail: jbhojwani2005@gmail.com

Received July 18, 2013; Accepted August 16, 2013; Published August 19, 2013

Citation: Patidar P, Bhojwani J (2013) Identification and Pattern Analysis of SNPs Involved in Colorectal Cancer. J Stem Cell Res Ther 3: 144. doi:10.4172/2157-7633.1000144

Copyright: © 2013 Patidar P, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

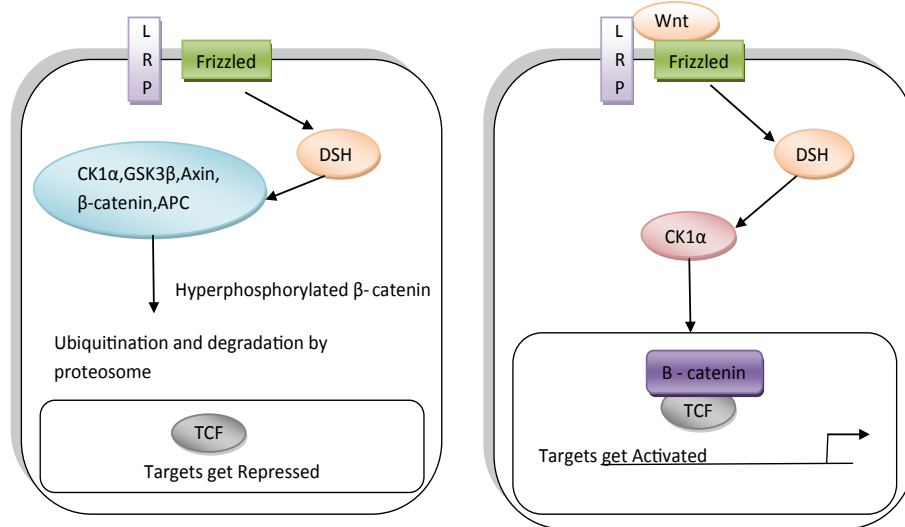


Figure 1: A simple schematic representation of *Wnt*-pathway with or without *Wnt* signal (a transmembrane protein). In the absence of *Wnt* signal (*left panel*), action of the destruction complex (CK1 α , GSK3 β , APC, Axin) creates a hyperphosphorylated β -catenin, which is a target for ubiquitination and degradation by the proteasome. Binding of *Wnt* ligand to a Frizzled/LRP-5/6 receptor complex (*right panel*) leads to stabilization of hypophosphorylated β -catenin, which interacts with TCF/LEF (T-cell Factor/Lymphoid Enhancer-binding Factor) proteins in the nucleus to activate transcription and trigger tumorigenesis. Dishevelled (Dsh) is a family of proteins involved in canonical and non-canonical *Wnt*-signalling pathways. Dsh is a cytoplasmic phospho-protein that acts directly downstream of Frizzled receptors. In the canonical pathway, CK1 α , GSK3 β , APC, and Axin act as negative regulators and all other components act towards positive regulation. It is widely accepted now that constitutive activation of *Wnt* signaling caused by mutations in the components of this pathway is responsible for initiation of Colorectal Cancer [7].

[9]. Genome-Wide Association Studies (GWAS) scans the genome for identifying these genetic variants (SNPs) that are prevalent in patients with specific disease [10].

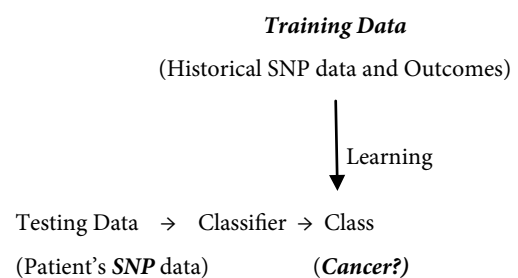
Thus, SNPs serve as potential and invaluable markers (landmarks in the genome) for association-based approaches to discover the genetic components bearing complex traits. Furthermore, where large sample sizes are required, their bi-allelic nature is amenable to high throughput automated genotyping. SNPs have been of particular interest, also to the evolutionary biologists. The genomic areas with extreme SNP density hold specific functions *i.e.*, may have specific gene and structural content [11]. In addition, SNPs are useful genetic markers for family-based linkage studies of Mendelian diseases [12], involving population histories and/or genetics and personalization of medicine [13]. Previous SNP-based efforts have focused on (1) candidate genes for common diseases, (2) genes with expressed sequence tags, or (3) genomic sequences. SNP profiles and patterns may therefore find immense utility in identifying a comprehensive collection of genes that contribute to the development and susceptibility of complex diseases such as cancer [14]. Mining of SNPs from either EST Databases [15] or DNA-Sequence Databases [16] have been the most prevalent approaches in such ventures.

Machine learning

Human genome analysis and the development of high-throughput techniques have provided us with a wide array of complicated biological data. Because of unwieldiness of these data, traditional statistical methods haven't performed efficiently on this kind of analysis. Machine learning algorithms carry the potential for mining significant information from relatively large, noisy, and complicated data.

Machine learning (ML) essentially is the study and computer modelling of learning processes, including acquisition of new declarative knowledge, organization of new knowledge into general

effective representations, and discovery of new facts through observation and experimentation. These programs are advantageous in many cases, where the input/output pairs can be specified, but the concise relationship between the input/output pairs is not usually known. Machine learning programs can thus help in extracting the complex relationships and obscure correlations in relatively large datasets [17] with the intention of uncovering hidden patterns (a process sometimes referred to as: data-mining). Machine Learning deals with computer programs that learn from and improve with experience by training data [18]. However, learners designed to recognize patterns in training data SNPs with realistic clinical data are scarce.



Complex diseases, in nature, arise from the effects of multiple genes which often interact with each other to produce the symptomatic traits. SNP markers are most commonly used to identify such genes for complex disease diagnosis. An increasingly popular and promising way to utilize SNP data for complex disease diagnosis [19] is to employ Artificial Neural Networks (ANN). ANN* is a computer-based algorithm that can be trained to recognize and categorize the complex patterns.

[*ANN is a mathematical model or computational model that is inspired by the structure and/or functional aspects of biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist

approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Modern neural networks are non-linear statistical data modeling tools. They are usually used to model complex relationships between inputs and outputs or to find patterns in data.]

Materials and Methods

Tissue sampling and expression profiling

The paraffin embedded archived tissue biopsies of colorectal cancer patients (45 cases) were obtained from the Veterans Affairs Hospital, Long Beach, CA (USA). The details of these cases are as follows:

Number of cases analyzed: (45 patients)

- **Adenocarcinomas: 25 (Cancer patients suffering from colorectal cancer-Advanced Stage - III-IV)**
- **Tubular adenomas: 10 (Patients with polyposis – Stage-I)**
- **Tubulovillous adenomas: 10 (Patients with polyposis – Stage-I/II)**

To examine the protein-expression *in situ*, the archived tissue-biopsies (paraffin-embedded sections) were analyzed by immune-labeling using fluorescence-based histochemical methods and processing of normal/cancerous/polyp tissues was carried out following a standard protocol [20] with slight modifications like, retrieval of antigen by pressure-cooker treatment in dilute citric acid buffer ([21]; see *Appendix-I*) prior to treatment with primary and secondary antibodies. For the results described in the report, the following antibodies (Santa Cruz Biotechnology, Santa Cruz, CA) were used at various dilutions. HDlg (*anti-human, monoclonal mouse*) was used at 7ug/ml, HCK1e (*polyclonal mouse*) at 10ug/ml and HAPC (*polyclonal rabbit - C-20*) at 8ug/ml. (“H” stands for antibodies raised in various animals here, against human antigen/s: H-Dlg, CK1e, APC).

Combinations (presented here): Primary antibodies:

1. HDlg (*mouse*) + HAPC (*rabbit*)
2. HCK1e (*mouse*) + HAPC (*rabbit*)

Secondary antibodies (fluorophore-tagged):

For combination 1: *Green tag-FITC-conjugated (*anti-mouse*)

*Red tag-Cy3-conjugated (*anti-rabbit*)

For combination 2: *Red tag-Cy3-conjugated (*anti-mouse*)

*Green tag-FITC-conjugated (*anti-rabbit*)

The processed sections were mounted in Vectashield (a standard anti-quenching mountant) and preserved in dark at 4°C (to prevent bleaching of fluorophores) for further observation after undergoing the immune-histochemical regimen. The protein localization of various genes in question (in several permutation/s and combination/s: see *Results* section), were viewed using MRC-1024, BioRad/Nikon Diaphot-200 Confocal microscope and Laser Sharp image analyzer software (BioRad Microscience Division, Cambridge, MA). For each case, the intensity of various proteins (in question) was compared with case-matched (from same patient) normal intestinal tissue, which served as a control. Each case was independently scored three-times in a single-blind manner to reduce the scoring bias. A mean calculated for 3 such scores (for each case) indicated whether there is a significant

loss of the protein or no change at all (as compared to control) for that particular permutation/combination.

Data mining

- Nucleotide sequences for *APC*, *CK-1 (Casein Kinase 1)*, *β-catenin* and *H-dlg* were retrieved from taxonomy browser of NCBI. (<http://www.ncbi.nlm.nih.gov/sites/entrez>).
- Expressed Sequence Tags (ESTs) sequences for the same were obtained from dbEST (<http://www.ncbi.nlm.nih.gov/dbEST>).
- Protein structures for *β-catenin*, *CK-1 (Casein Kinase)* and *H-Dlg* were obtained from PDB (Protein Data Bank). (www.pdb.org).

Softwares used

QualitySNP: This tool was used to detect reliable SNPs and insertions/deletions (indels) in EST data, both with and without quality files [22]. It basically uses three filters for the identification of reliable SNPs. Filter 1 screens for all potential SNPs and identifies variation between or within genotypes. Filter 2 is the core filter that uses a haplotype-based strategy to detect reliable SNPs. Clusters with potential paralogs as well as false SNPs caused by sequencing errors are thus located. Filter 3 screens SNPs by calculating a confidence score, based upon sequence redundancy and quality. *QualitySNP* is an efficient tool for SNP detection, storage and retrieval in diploid as well as polyploid species. It is available for running on Linux or UNIX systems. at <http://www.bioinformatics.nl/tools/snpweb>.

FastSNP: FASTSNP (Function Analysis and Selection Tool for Single Nucleotide Polymorphisms) web server was used to efficiently identify and prioritize high-risk SNPs [23] according to their phenotypic risks and putative functional effects. A unique feature of *FastSNP* is that the functional effect information used for SNP prioritization is always up-to-date, since *FastSNP* extracts the information from 11 external web-servers at query time using a team of web wrapper agents. Moreover, *FastSNP* is extendable by simply deploying more Web wrapper agents. It prioritizes SNPs according to 13 phenotypic risks and putative functional effects, such as changes to the transcriptional level, pre-mRNA splicing, and protein structure and so on.

FastSNP is available at <http://fastsnp.ibms.sinica.edu.tw>

Results

Expression profiling of potential “early” and “late” acting candidate molecules in *Wnt* signaling pathway

Upon observing the tissue biopsies for spatio-temporal protein localization patterns for various potential candidate molecules (like; *p53*, *h-Dlg*, *CKI*, *EGFr*, *ZO-1* etc) that are implicated in colorectal cancer, the most prominent and noticeable results were those of APC-CSNK1e interactions presented here (CSNK1e is mentioned as CKI for ease; Figure 2). The results from hDlg-APC interaction are also reported here (*as control*) and for comparison (Figure 3). It was consistently observed in all the analyzed samples (for the 20 acquired cases of polyposis), that, in the early-to-late polyp formation stages, when cytoplasmic component of APC begins to disappear (which is supposed to be a characteristic feature of APC mutant cells, seen in culture studies (*in-vitro*) by Rosin-Arbesfeld et al. [24]; see *illustration* below), CKI protein also gets concomitantly lost from the cells, which are mutant for APC (Figure 2). Following a similar pattern, the same was recapitulated in the aberrantly developing epithelial cells of the colonic polyps (shown here, Figure 2B). Apart from that, in the normal cells, APC and CKI were seen to be strongly co-localized *in-*

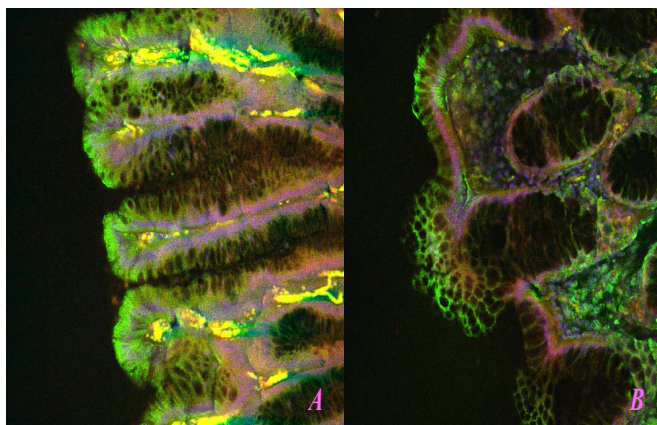


Figure 2: Expression profiles of APC and CKI in a Tubulovillous case during intermediate-late polyp formation: CKI is concomitantly lost due to loss of APC.

This is a comparison between a near-normal section of villi (A) and an abnormally developing villi (B) from the same patient. The abnormal polyp tissue (B) exhibits a con-committant loss of CKI protein expression due to the loss of cytoplasmic APC in growing cancerous cells. Note the transition of CKI protein expression to a patchy discontinuous appearance in the basal cell-membranes of the growing polyps in panel B, in contrast to a continuous basal membrane and cytoplasmic localization of the protein while APC is still tends to be cytoplasmic. The section in panel B suggests loss of this cytoplasmic component of APC in the mutant cells within a mixed population of normal and cancerous tissues in growing polyps.

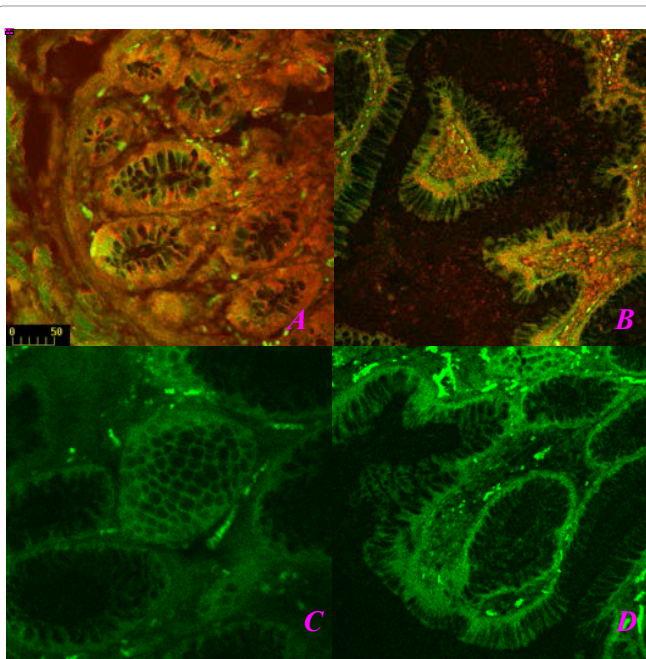


Figure 3: Expression profiling of potential “early” and “late” acting candidate molecules in *Wnt* signaling pathway.

The protein localization of APC and hDLg proteins in (A) normal colonic crypts and (B) polyp tissue. Note the normal basolateral membrane localization of hDLg in the disintegrating polyp structure (B), which is also evident in normal crypts (A). Panel (C) shows an abnormal intestinal crypt exhibiting loss of apico basal polarity and the panel (D) depicts abnormally developing villi from the colon of this tubulovillous case giving rise to early polyps. Note that hDLg is mislocalized due to the architectural disturbances in the dysplastic cells of crypts and villi in these premalignant tissues, which lose their apico-basal polarity and not due to the loss or degradation of APC protein *per se* (B), which tends to invariably accumulate in the lumen of the gut and the crypts. Figures A and C are acquired at 35X whereas; Figures B and D are at 20X.

in vivo (Figure 2A). This clearly indicates a strong functional correlation between the two molecules and suggests a temporal early requirement of CKI for the APC tumor suppressor gene. In contrast, when APC-hDLg protein interaction was closely monitored, there was no obvious change in the hDLg expression *per se*. Dlg expression [*disc large (dlg)*], initially identified in *Drosophila*, which belongs to MAGUK family of membrane associated proteins; [25]] thus, seemed to be regulated or perturbed only as a result of/ or following the architectural changes in abnormal dysplastic tissues *per se* (Figure 3C and 3D) of growing polyps, rather than the loss of the cytoplasmic APC protein in APC mutant cells* (which is a prominent characteristic feature of APC mutant cells* - See illustration below) of the intestinal villi (see, Figure 3D). However, both CKI and hDLg were seen to be continually lost from the cancerous tissues in the later stages of development (ie. in carcinomas, *not shown here*) due to the loss of cell components, cytoskeletal alterations and damage. Thus, *hDLg* was considered to be a late-acting gene in contrast to CKI, which is an early-acting genetic factor (and thus a potential target gene) temporally interacting and/or associated with the APC gene during polyp formation, for the initiation of CRC disease (Figure 4), owing to its role as a binding partner, in the *Wnt*-signaling pathway.

Pattern analysis of “Region Specific SNPs” residing in 5q21-23, 22q13, 3q21, corresponding to APC, CKI and HDLG genes

These strikingly significant observations on sub-cellular localization of the genes in question provided us with a good start-point to explore and analyze the SNPs associated with and within these genes. Quality SNP tool was used to detect reliable SNPs and insertions/deletions (indels) in *EST* data of various genes in question. Mainly focusing on the *Wnt* signaling pathway members, we obtained data with and without quality files. The quality data was filtered using the described standard method (see, *Materials and methods*). Using the FASTSNP (Function Analysis and Selection Tool for Single Nucleotide Polymorphisms), which allows the users to efficiently identify and prioritize high-risk SNPs according to their phenotypic risks and putative functional effects; we charted out the data for the related SNPs (Tables 1 and 2).

Upon keen observation and analyses of the pattern of distribution of these SNPs, we found that, the total number of SNPs retrieved in *HDlg* region was the highest, which probably accounts for all the “noise” and does not carry much weight in mutational cause-effect, also incidentally correlated to the functionally persistent condition during the early stages of development of CRC shown by the wet lab data presented here (Figure 3). Although, the total number of exonic SNP’s obtained in APC region was very high (44) as expected, but only three significant (3* in Table 1) non-synonymous SNPs were found, which actually contribute towards any significant change in terms of being non-sense/mis-sense type. APC being the actual cause of initiation of polyps should have ideally bore a larger number of them, but it doesn’t seem to be the case here. However, a strikingly higher number of exonic SNP’s (even more than that of APC) for the *CSNK1ε* came as a big surprise! A significant cause-effect value (ie. 2*/1917 for *CSNK1ε* as against 2*/266 for *HDlg* in Table 1.) on 22q13 chromosome was seen, which further suggests that *Casein Kinase-I* is one of the most significant and functionally relevant candidate genes amongst the analyzed group here.

SNP distribution data

Further careful analyses of pattern distribution of SNPs within the *Casein Kinase* family, yielded a second potential candidate in this gene-set: *CSNK1δ*, which is a sister-gene of *CSNK1ε* and incidentally

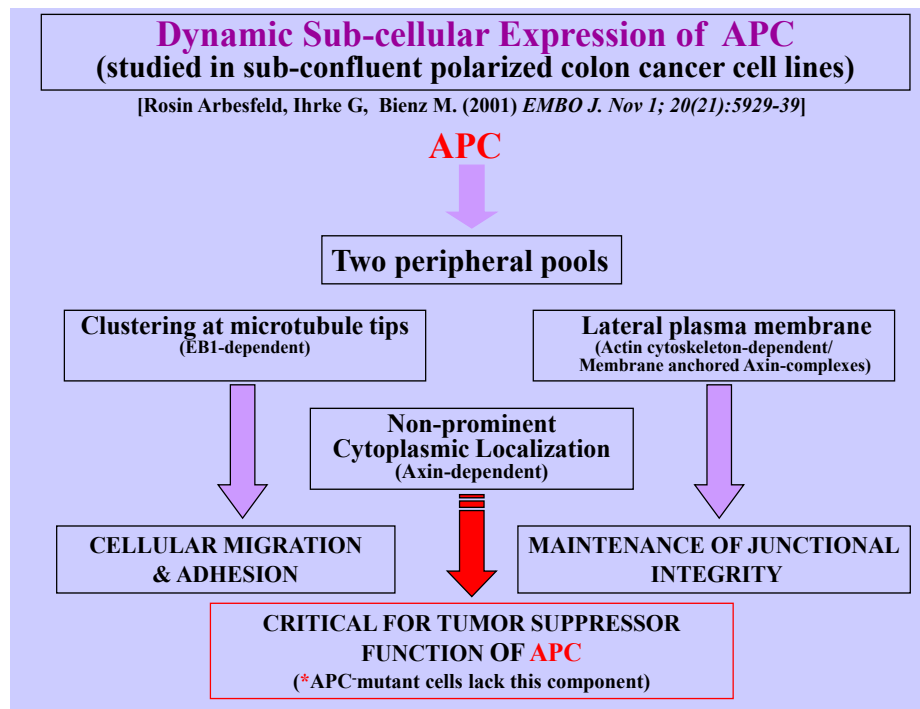


Figure 4: Dynamic Sub-cellular Expression of APC.

S.No	Gene Name	Chromosomal position	Total SNPs retrieved	Total SNPs in exonic region/s	Mis-sense/Non-synonymous SNPs
1.	APC	5q21-22	761	44	3*
2.	CSNK1ε	22q13	266	54	2*
3.	HDLG1	3q21	1917	12	2*

Table 1: *SNPs for all the three genes (APC, CSNK1ε and HDLG), with their significant mutant "effects". Note that, the significance of 2*/266 (CK1ε) rather than, 2*/1917 (HDlg) carries more weight in "cause- effect" value.

Gene	Gene Location	Total SNPs	SNPs in exonic region	Coding /non-synonymous
CSNK1ε	22q13	266	54	2
CSNK1δ	17q25	187	40	11
CSNK1α	5q32	364	29	1
CSNK1γ	15Q22	902	50	2

Table 2: Comparative SNP distribution pattern in the CSNK1 family.

happens to share ~98% amino acid homology with it in the kinase domain [26]. Out of the 40 SNPs retrieved for CSNK1δ in the exonic region, 11* SNPs cause significant effect in terms of generating mis-sense function of the protein (Table 2). Thus, one can easily assess putative functional relevance of these potential candidates depending on their pattern of distribution and the resultant/derived "cause-effect" value. The chromosomal region 22q13, in which CSNK1ε resides, is the richest in the number of exonic SNPs in this data-set and therefore, again tops this chart, in bearing a significant functional relevance to the early initiation-phase of CRC disease.

Discussion

Genotypic diversity is presumed to underlie the heritable phenotypic differences observed as variation in drug response, susceptibility to disease, and other complex traits. It is therefore conceivable, that cataloging DNA sequence variation/s in the genome will presumably provide genuine invaluable tools to relate genotypes with complex phenotypes and hence enable discovery of the causative genetic factors and mechanisms, underlying various complex diseases,

such as cancer [27,13]. Common types of sequence variations in humans include single nucleotide polymorphisms (SNPs), insertions and deletions of a few nucleotides, and variation in the repeat number of a motif (i.e., mini- and microsatellites). SNPs have a relatively high abundance in the genome, occurring at a density of ~1 per kb when any two genomes are compared, and they also possess a relatively low mutation rate. Thus, they serve as a reliable platform in which to search for such complex gene functions and diseases. Precise SNP research thus provides a fundamental understanding to many polygenic diseases, paving a promising way to discovering new therapeutic targets. However, the race amongst pharmaceutical companies today is to apply a new genomics approach to identify "novel targets" and validate these targets in the most efficient fashion. We have a very crucial clue from "pattern analysis" reported here (on the basis of significant "cause-effect" value, see: Tables 1 and 2) of the retrieved "region-specific" SNP's. Our results clearly point to the fact, that although a particular chromosomal region/stretch is richer than others, in effect, only the SNPs that cause a significant/dramatic change in terms of hampering the function of the protein in question, really matter. These

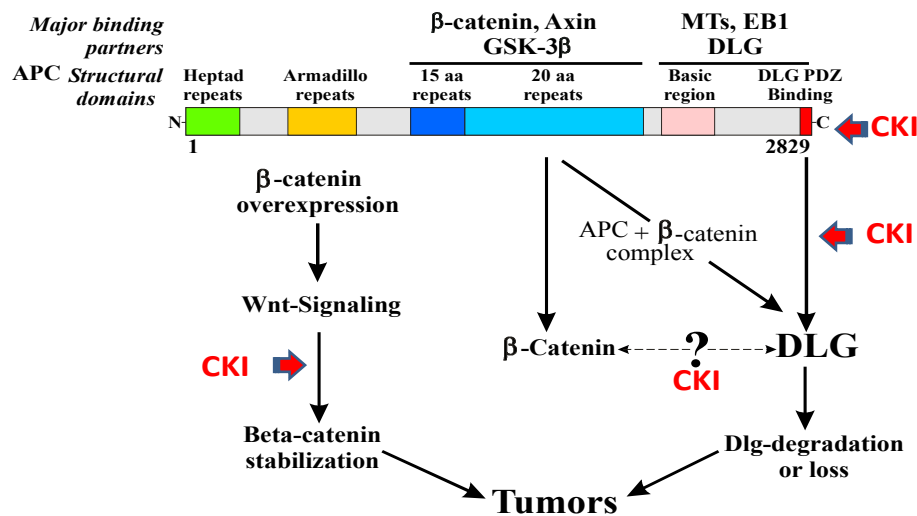


Figure 5: CKI emerges as a potential target/candidate for pharmacogenomics study and also as potential connecting link between APC and H-Dlg.

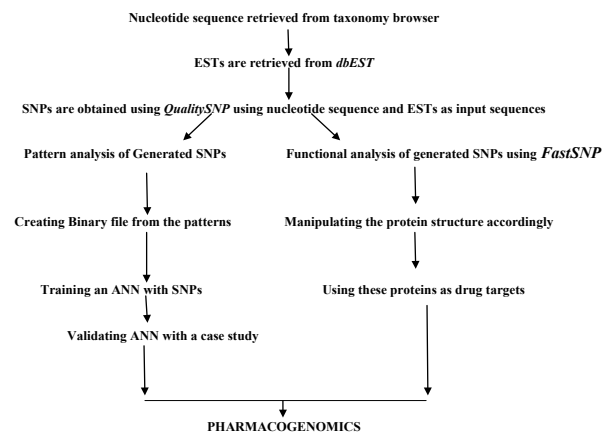
SNPs actually result in dysfunctional/non-functional protein due to the associated frame-shift, nonsense mutations etc. Here, *CKIε* (*epsilon*) and *δ* (*delta*) regions have emerged such susceptible regions over *HDlg*. Thus, CKI can be suggested as a “potential candidate” for such a study. *CSNK1ε* (CKI) for which the expression profiling was also conducted (see; Figure 2 and illustration below), incidentally resides on an SNP-rich chromosome 22 q [11]. This chromosome can thus be diligently exploited for further analysis in context to the cancer-disposition genes and oncogenomics. The SNP-rich regions in the entire human genome are presumed to have specific function/s (hot-spots) unlike the SNP-poor regions (cold-spots), and therefore could act as “potential genetic targets” for therapeutic interventions (Figure 5).

We have already created a simplistic Colorectal Cancer (CRC) model, using some relevant “Systems Biology” approaches now (*in prep.*). It is also quite known now that a promising way to utilize SNP data for complex disease diagnosis is to employ Artificial Neural Networks (ANN), and we also intend to train and design an “ANN model” for determining the predisposition of this disease by harnessing the power of this tool in our forthcoming ventures. Our ANN (Artificial Neural Network) model will hopefully predict such predispositions (see Figure below), owing to the very encouraging results presented here. A trained ANN model is mostly helpful in knowing the probability of disease occurrence in an individual/s implicating susceptibility/predisposition of that particular disease (in this context; colorectal cancer). We would basically focus on the molecules of the 4 major alternative signaling pathways defined for colorectal cancer, namely; the *Wnt*, *P53*, *TGF-beta* and *K-Ras* [28], like the targeted approaches deployed earlier by Pierotti et al. [29], and/or other so-called emerging pathways in CRC [30,31]. *Wnt* pathway has been of particular interest for us because of its implication in stem cell renewal and differentiation, *per se* [2]. The pathways that bear relevance both in carcinogenesis and stem cell biology would thus be of prime interest, for such a translational-research venture. A corollary of this ongoing project would be to identify precisely those SNPs which are associated with obvious biological effects in response to chemical drugs (deploying *chemi-informatics*). Thus, this particular SNP effort would, in part, serve as the bedrock of *pharmacogenomics* (in this context; also, *oncogenomics*), which is an emerging field for personalized medicine:

the right drug, in the right dose, to the right person, at the right time (*i.e.*, Tailor-made/Customized drugs).

Acknowledgements

The authors wish to extend their heartfelt thanks to all the friends and colleagues for invaluable inputs from time to time, during the course of this project. The wet-lab data presented in this report was extracted from an ambitious post-doctoral project conducted by JB under her “Training Grant on Carcinogenesis.” (CRI, NIH grant #CA09054-25) at UC-Irvine, CA, USA.



References

- Spink A, Xu JL (2000) Selected results from a large study of Web searching: the Excite study. Information Research 6.
- Reya T, Clevers H (2005) Wnt signaling in stem cells and cancer. Nature 434: 843-850.
- Vogelstein B, Kinzler KW (1993) The multistep nature of cancer. Trends Genet 9: 138-141.
- Hemminki K, Mutanen P (2001) Genetic epidemiology of multistage carcinogenesis. Mutat Res 473: 11-21.
- Polakis P (2000) Wnt signaling and cancer. Genes Dev 14: 1837-1851.
- Giles RH, van Es JH, Clevers H (2003) Caught up in the Wnt storm: Wnt signaling and cancer. Biochim Biophys Acta 1653: 1-24.
- Sancho E, Battle E, Clevers H (2004) Signaling pathways in intestinal development and cancer. Annu Rev Cell and Dev Biol 20: 695-723.
- Kinzler KW, Nilbert MC, Su LK, Vogelstein B, Bryan TM, et al. (1991) Identification of FAP locus genes from 5q21. Science 253: 661-665.

9. Erichsen HC, Chanock SJ (2004) SNPs in cancer research and treatment. Br J Cancer 90: 747-751.
10. Earl J, Greenhalf W (2010) Single-nucleotide polymorphism (SNP) analysis to associate cancer risk. Methods Mol Biol 576: 171-196.
11. Dawson E, Chen Y, Hunt S, Smink LJ, Hunt A, et al (2001) A SNP resource for Human Chromosome 22: Extracting dense clusters of SNPs from the genomic sequence. Genome Res 11: 170-178.
12. Kruglyak L (1997) The use of a genetic map of biallelic markers in linkage studies. Nat Genet 17: 21-24.
13. Housman D, Ledley FD (1998) Why pharmacogenomics? Why now?. Nat. Biotechnol 16:492-493.
14. Frank B, Hoffmeister M, Kloop N, Illig T, Chang-Claude J, et al. (2010) Single nucleotide polymorphisms in Wnt signaling and cell death pathway genes and susceptibility to colorectal cancer. Carcinogenesis 31: 1381-1386.
15. Picaoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, et al. (1999) Mining SNPs from EST databases. Genome Res 9: 167-174.
16. Van oeveren J, Janssen A (2009) Mining SNPs from DNA sequence data; computational approaches to SNP discovery and analysis. Methods Mol Biol 578: 73-91.
17. Tomita Y, Tomida S, Hasegawa Y, Suzuki Y, Shirakawa T, et al. (2004) Artificial neural network approach for selection of susceptible single nucleotide polymorphisms and construction of prediction model on childhood allergic asthma. BMC Bioinformatics 5:120.
18. Mitchell TM (1997) Machine Learning. McGrawHill Inc, New York, USA.
19. Mutoh H, Hamajima N, Kobayashi T, Honda H (2006) Exhaustive combination search method for SNPs combination Related to *Helicobacter pylori* infection susceptibility Using Artificial Neural Network.
20. Harlow E, Lane D (1988) Antibodies: A Laboratory Manual. Cold Spring Harbor, NY. Cold Spring Harbor Laboratory Press.
21. Norton AJ, Jordon S, Yeomans P (1994) Brief, high-temperature heat denaturation. J Pathol 173: 371-379.
22. Tang J, Vosman B, Voorrips RE, van der Linden CG, Leunissen JA. (2006) QualitySNP: pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species. BMC Bioinformatics 7: 438.
23. Yuan HY, Chiou JJ, Tseng WH, Liu CH, Liu CK, et al. (2006) FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. Nucleic Acids Res 34: W635-W641.
24. Rosin A, Ihrke G, Bienz M (2001) Actin-dependent membrane association of the APC tumour suppressor in polarized mammalian epithelial cells. EMBO J 20: 5929-5939.
25. Woods DF, Bryant PJ (1991) The discs-large tumor suppressor gene of Drosophila encodes a guanylate kinase homolog localized at septate junctions. Cell 66: 451-464.
26. Fish RW (1995) File system for a clustered processing system. US-Patents: NCR Corporation, Dayton, Ohio, USA.
27. Collins FS, Guyer MS, Chakravarti A (1997) Variations on a Theme: Cataloging Human DNA Sequence Variation. Science 278: 1580-1581.
28. Smith G, Carey FA, Beattie J, Wilkie MJ, Lightfoot TJ, et al. (2002) Mutations in APC, Kirsten-ras and p53 – Alternative genetic pathways to colorectal cancer. Proc Natl Acad Sci USA 99: 9433-9438.
29. Pierotti MA, Negri T, Tamborini E, Perrone F, Pricl S, et al. (2009) Targeted Therapies: The Rare Cancer Paradigm. Mol Oncol 4: 19-37.
30. Haydon HR, Jass JR (2002) Emerging pathways in colorectal cancer development. Lancet Oncology 3: 83-88.
31. Luu HH, Ruiven R, Haydon RC, Rayburn E, Kang Q, et al. (2004) Wnt/ β -catenin signaling pathway as novel cancer drug targets. Curr Cancer Drug Targets 4: 653-671.