**Evaluation of Colorectal Cancer (CRC) Epidemiology**  
**A Pharmacogenomic Approach**

HimaVynshavi A M1, Lakshmi Anand C2, Deepak O M3, P K Krishnan Namboori4

1Center for Computational Engineering & Networking (CEN), Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, INDIA.  
2Department of Sciences, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, INDIA.  
3Head of the Research Group, Computational Chemistry Group (CCG), Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, INDIA.

**ABSTRACT**

**Background:** The population-wise variation in proneness of Colorectal Cancer (CRC) has been studied in the manuscript. A population wise analysis of responsiveness towards colorectal cancer is carried out with genetic, epigenetic, metagenomic and environmental factors associated with APC mutation mainly responsible for CRC among eight different populations.  

**Methods and Material:** The APC mutation has been obtained using the human gene mutation database-HGMD and the ‘international cancer genome consortium-ICGC’ Data Portal. The epigenetic factors affecting colon cancer have been identified through EpiGRAPH tool. The ‘human oral microbiome database (HOMD) and ‘comparative toxicogenomics database (CTD)’ are used to find the metagenomic factors affecting CRC.  

**Results:** Variants of APC gene from the selected ethnic classes chosen from Argentina, France, Germany, India, Poland, Romania, UK and USA were characterized, where the chromosome positions 112102966–112177228 are found to be affected. It has been found that among epigenetic factors: chromosome organization, population variation, and evolutionary history are highly promising features for the prediction of DNA methylation.  

**Conclusions:** The chromosome positions 112102966–112177228 are found to be the most prone region for APC mutation. Chromosome organization, population variation, and evolutionary history are highly promising epigenetic features for the prediction of DNA methylation and further mutation. The consumption of spices, coconut oil, fish (in coastal areas), dairy products and reduced intake of red meat may be the reasons for less incidence rate of CRC among the Indian population.

**Key words:** CRC, APC, Epigenetics, Metagenomic, Environmental Factors, genetic signature.

**Key message:** The manuscript includes an exhaustive pharmacogenomic analysis of population-wise variation towards proneness of colorectal cancer (CRC). The observed lesser proneness of the Indian populations towards CRC may be due to the unique food habit in the region with the higher consumption of diary products, spices, coconut oil, fish (in coastal areas) and lesser consumption of red meat.

**Correspondence:**  
Dr. P K Krishnan Namboori,  
Associate Professor and Head of the Research Group, Computational Chemistry Group (CCG), Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, INDIA.  
Phone: +91422 2685000  
Fax: +91 422 2656274  
Email: n_krishnan@cb.amrita.edu  
DOI: 10.5530/jyp.2017.9.7

**INTRODUCTION**

Recently, cancer has become a major leading cause of death in the global population.1 Out of various types of cancers, colorectal cancer (CRC) stands as a prominent form in the world.2 However, the proneness of the disease has been identified as varying among various populations. It has been found that among different populations, Asian traits especially the Indians have been branded as least affected.3 Among Indian traits, CRC has been recognized as the 10th most dominant form of cancer claiming about 4% of lives.4  

The population-wise variation of the disease is always a matter of concern to the areas of pharmacogenomic analysis and designing of precision drugs. The genetic, epigenetic, metagenomic and environmental factors are the major features contributing to population wise variation of the disease.

The backing mutation leading into CRC has been identified as Adenomatous polyposis coli (APC).5 About 1547 APC mutations have been identified in the global populations. The ‘Single Nucleotide Polymorphism (SNP), can be projected out as the most efficient genetic signature behind proneness of the mutation.6

About 847 attributes, put in 10 classes, epigenome and chromatin structure, transcriptome, population variation, regulatory regions, DNA sequence, repetitive DNA, DNA structure, evolutionary history, genes and chromosome organization have been noted down affecting the epigenetic variations of global populations using ‘Epigraph,’ a web based SVM tool. The prediction of involvement of each class of attributes can be set up by making a machine learning approach using ‘Support Vector Machine (SVM)’ with the application of kernel magic.7

The APC can also be supported by metagenomic factors. It has been found that the bacterium *Lactobacillus casei*, present in banana fruits and dairy products helps in preventing APC mutation.8 Hence, it is always advisable to prevent depletion of these bacteria by providing a symbiotic environment in the digestive tract. Besides the metagenomic components, the involvement of other environmental factors such as presence of mutagens, possibility of UV radiation etc. have to be excavated in the analysis.9 All the above features contribute to ethnic variations in the responsiveness of the mutation.
Table 1: Major APC mutations

<table>
<thead>
<tr>
<th>Sl. no.</th>
<th>ID (ICGC)</th>
<th>DNA change</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MU2186971</td>
<td>chr5:g.112175951-&gt;A</td>
<td>insertion of &lt;=200bp</td>
</tr>
<tr>
<td>2</td>
<td>MU54981</td>
<td>chr5:g.112162891C&gt;T</td>
<td>single base substitution</td>
</tr>
<tr>
<td>3</td>
<td>MU58552</td>
<td>chr5:g.112175423C&gt;T</td>
<td>single base substitution</td>
</tr>
<tr>
<td>4</td>
<td>MU61050</td>
<td>chr5:g.112128143C&gt;T</td>
<td>single base substitution</td>
</tr>
<tr>
<td>5</td>
<td>MU62829</td>
<td>chr5:g.112175480G&gt;C</td>
<td>single base substitution</td>
</tr>
<tr>
<td>6</td>
<td>MU63447</td>
<td>chr5:g.112116592C&gt;T</td>
<td>single base substitution</td>
</tr>
<tr>
<td>7</td>
<td>MU64682</td>
<td>chr5:g.112128191C&gt;T</td>
<td>single base substitution</td>
</tr>
<tr>
<td>8</td>
<td>MU65639</td>
<td>chr5:g.112174631C&gt;T</td>
<td>single base substitution</td>
</tr>
<tr>
<td>9</td>
<td>MU70498</td>
<td>chr5:g.112175480G&gt;C</td>
<td>single base substitution</td>
</tr>
<tr>
<td>10</td>
<td>U70250</td>
<td>chr5:g.112175639C&gt;T</td>
<td>single base substitution</td>
</tr>
</tbody>
</table>

Table 2: The genetic signatures of the chosen ethnic groups

<table>
<thead>
<tr>
<th>Sl.no.</th>
<th>Ethnicity</th>
<th>Genetic signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Argentina</td>
<td>rs1801155, rs121913224, rs1801166, rs397514031, rs587782293, rs663224458, rs587782557, rs663225355, rs663224820, rs199531187, rs76658325</td>
</tr>
<tr>
<td>2</td>
<td>France</td>
<td>rs587781490, rs587781490, rs587776685</td>
</tr>
<tr>
<td>3</td>
<td>Germany</td>
<td>rs139196838, rs663223335, rs74953290, rs587779352</td>
</tr>
<tr>
<td>4</td>
<td>India</td>
<td>rs587782305, rs1801155</td>
</tr>
<tr>
<td>5</td>
<td>Poland</td>
<td>rs397515734</td>
</tr>
<tr>
<td>6</td>
<td>Romania</td>
<td>rs775126020, rs12516245, rs12516846, rs12520108, rs12523402, rs139159621, rs7557733, rs2909961, rs35031194, rs35031194, rs371918485, rs372355890, rs389363, rs3909896, rs58053486, rs61212392, rs6885311, rs6887538, rs73022563, rs74940478, rs57280259, rs78919815, rs387906231, rs587776684, rs587779798, rs121908771, rs587781580, rs587782130, rs587781392, rs2909961, rs62619935, rs397515734</td>
</tr>
<tr>
<td>7</td>
<td>UK</td>
<td>rs137854574, rs876658355</td>
</tr>
<tr>
<td>8</td>
<td>USA</td>
<td>rs137854574, rs876658355</td>
</tr>
</tbody>
</table>

MATERIALS AND METHODS

In the pharmacogenomic perspective, the responsiveness towards any disease depends up on the mutation, genes involved in the mutation, gene-gene interaction and the cumulative effect of this interaction, epigenetic factors, environmental factors and metagenomic factors. The present manuscript is a comprehensive approach to identify and evaluate the epidemiological factors associated with the CRC.

The APC mutation has been characterized using the ‘Human Gene mutation database-HGMD’ and the ‘International Cancer Genome Consortium-ICGC Data Portal.’ The relevance of the mutation in causation of CRC can be justified by computing the frequency and rate of transformation in different populations.

The ‘Single Nucleotide Polymorphism-SNP’ has been identified as the most appropriate genetic signature towards responsiveness of the disease. The SNPs can be located using NCBI database gene mapping and evolutionary biology, the National Center for Biotechnology Information (NCBI and Gene Cards). The non-synonymous, pathogenic and deleterious SNPs can be located with the help of the online sup-
ports SIFT and POLYPHEN. The population wise genetic signatures behind promen of the disease have been generated using the popular DNA variant databases, Leiden Open variation database (LOVD), UCSC Genome Browser and Indian Genome variation database (IGVdb). The CpG islands have been located using the ‘Database of CpG islands and Analytical Tool (DBCAT).’ The DNA methylation possibilities have been studied using the epigenome browser, EpiGRAPH. The various epigenetic factors have been analyzed using linear and non-linear kernels of Support Vector Machine (SVM) to compute the relevance of each attribute in making the predictions. The metagenomic factors affecting CRC have been identified using ‘human oral microbiome database (HOMD)’ and comparison toxicogenomics database (CTD).

RESULT

Variation and Mutation study

The APC mutation has been recognized as a major transformation responsible for CRC with a total of 1547 variations, out of which highly prominent 10 mutations have been included in Table 1. The non-synonymous, pathogenic and deleterious SNPs have been generated with the help of SIFT and POLYPHEN for the APC gene from the selected ethnic classes chosen from Argentina, France, Germany, India, Poland, Romania, UK and USA (Table 2).

Epigenetic study

The attributes with high mean accuracy, chromosome organization, population variation and evolutionary history, have been found to be the promising features for the prediction of DNA methylation and further mutation (Figure 1).

Environmental study

It has been found that consumption of Linoelic acid, Oleic acid and Lauric acid play a major role in preventing CRC. Similarly, most of the spices have been recognized as keeping antioxidant, anti-proliferative and carcinogen blocking properties.

Metagenomic study

The involvement of microbiota, especially Streptococcus bovis, Fusobacterium nucleatum, Fusobacterium mortiferum, Fusobacterium necrophorum, Campylobacter leptum, Campylobacter coccoides and Faecalibacterium prausnitzii causing CRC has been emphasized. Similarly, Lactobacillus casei, found in dairy products and banana, has been identified as supporting proper digestion and thereby reducing ulcers in digestive tract. The over
consumption of alcohol, beverages and cigarette smoking are considered as risk factors for CRC. The major mutations leading into APC mutation and CRC include DDT, poly chlorinated biphenols, permuterin, exposure to ASBESTOS and irradiation of the CERVIX UTERI.

DISCUSSION

In APC, the chromosome positions 112102966-112177228 are found to be the region corresponding to the population wise variations. All the ethnicity based genetic signatures are found to be in the above region. 847 attributes in 10 classes have been included in the analysis, where the linear kernel SVM has been found to be the most efficient method to carry out the epigenetic study. This clearly supports the relevance of population wise pharmaco genomic analysis for the variation in CRC. The reported compounds induce apoptosis in carcinoma cells by increased production of 'Reactive Oxygen Species (ROS)', leading into environmental stress and ultimately causing cell death.19

Oleic acid suppresses the over expression of HER2 gene, which helps in invasive progression and metastasis in human cancer.24 Fish oil and olive oil decrease the proliferation and early down regulation of COX2 followed by reduced BCL2 expression.23 The consumption of food containing ROS and cancer chemo preventive agents may be an ingenious method for controlling CRC.

The environmental factors such as pollutants, chemicals and asbestos are found to be major risk factors of APC mutations. It has been found that the dietary habits reserve a major role in the proneness of CRC. The consumption of red meat may be a major reason for the DNA damage causing APC mutation. The 'heme' content present in red meat has been reported as responsible for the production of N-Nitroso compounds and DNA damage resulting in uncontrolled cell division.26 Ethnicity plays an important role in the disease proneness. It has been found that the African-American populations are more prone to the disease than the Asians, especially the Indians. The variation in environmental factors such as pollutants, food habits etc. may be the reason behind this. The large consumption of dairy products, spices, middle chain fatty acid containing food such as coconut oil, fish and less consumption of red meat may be the factors supporting the Indian population to be less prone to CRC.27

CONCLUSION

The responsiveness of CRC to different ethnic groups has been studied in the manuscript. The chromosome positions within the range 112102966-112177228 are found to be markers of the population wise variations of APC mutations. The genetic signatures behind the proneness of the disease to different ethnic groups have been computed. Among various epigenetic factors, chromosome organization, evolutionary history and population variation are found to be the principal attributes responsible for DNA methylation. The metagenomic effect of Lactobacillus caseirich in dairy products has been identified as useful in suppressing CRC. The overconsumption of red meat can be considered as a major risk factor for cancer. The consumption of spices, coconut oil, fish (in coastal areas), dairy products and reduced intake of red meat may be the reasons for less incidence rate of CRC among Indian population.

ACKNOWLEDGEMENT

Authors hereby acknowledge the management and officials of Amrita Vishwa Vidyapeetham for their support.

CONFLICT OF INTEREST

No conflict of interest are declared.

ABBREVIATIONS USED


REFERENCES

16. Fokkema IFA, LOVD: Leiden Open Variation Database. Leiden University Medical Center; 2016.

**Article History:** Submission Date: 21-06-16; Revision Date: 18-07-16; Accepted Date: 07-08-16.