Evaluation of Colorectal Cancer (CRC) Epidemiology A Pharmacogenomic Approach

HimaVyshnavi A M[1], Lakshmi Anand C[2], Deepak O M[3], P K Krishnan Namboori[*4]

[1]Center for Computational Engineering & Networking (CEN), Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, INDIA.
[2]Computational Chemistry Group (CCG),Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, INDIA.
[3]Department of Sciences, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, INDIA.
[4]Head 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. It has been found that consumption of linoleic acid, oleic acid, and lauric acid play a major role in preventing CRC. **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

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 global 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 productshelps in preventing APC mutation.[8] Hence, it is always advisable to prevent depletion of these bacteria by providing a symbiotic environment in the digestive track. 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

Sl. no.	ID (ICGC)	DNA change	Type
1	MU2186971	chr5:g.112175951->A	insertion of <=200bp
2	MU54981	chr5:g.112162891C>T	single base substitution
3	MU58552	chr5:g.112175423C>T	single base substitution
4	MU61050	chr5:g.112128143C>T	single base substitution
5	MU62829	chr5:g.112175480G>T	single base substitution
6	MU63447	chr5:g.112116592C>T	single base substitution
7	MU64682	chr5:g.112128191C>T	single base substitution
8	MU65639	chr5:g.112174631C>T	single base substitution
9	MU70498	chr5:g.112173917C>T	single base substitution
10	U70250	chr5:g.112175639C>T	single base substitution

**Table 2: The genetic signatures of the chosen ethnic groups**

Sl.no.	Ethnicity	Genetic signature
1	Argentina	rs1801155, rs121913224, rs1801166,rs397514031, rs587782293, rs863224458, rs587782557, rs863225355, rs863224820, rs199531187, rs876658325
2	France	rs587781490, rs587781490, rs587776685
3	Germany	rs139196838, rs863225335, rs74953290, rs587779352
4	India	rs587782305, rs1801155
5	Poland	rs397515734
6	Romania	rs775126020, rs12516245, rs12516846, rs12520108, rs12523402, rs139159621, rs17557733, rs2909961, rs35031194, rs35031194, rs371918485, rs372355890, rs389363, rs3909896, rs58053486, rs61212392, rs6885311, rs6887538, rs73022563, rs74940478, rs75282059, rs78919815
7	UK	rs387906231, rs587776684, rs587779798, rs121908771, rs587781580, rs587781330, rs587781392, rs2909961, rs62619935, rs397515734
8	USA	rs137854574, rs876658355

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'[10] and the 'International Cancer Genome Consortium-ICGC Data Portal.[11] The relevance of the mutation in causing 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[12] gene mapping and evolutionary biology, the National Center for Biotechnology Information (NCBI and Gene Cards).[13] The non-synonymous, pathogenic and deleterious SNPs can be located with the help of the online sup-
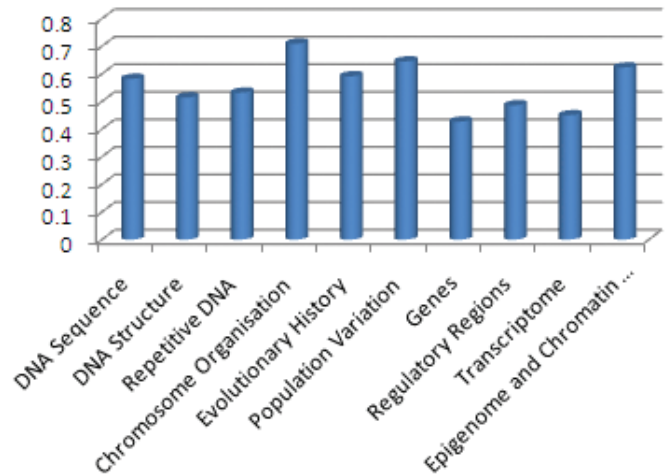
Figure 1: Attributes of getting methylation of APC gene.

ports SIFT[14] and POLYPHEN.[15] The population wise genetic signatures behind proneness of the disease have been generated using the popular DNA variant databases, Leiden Open variation database (LOVD),[16] UCSC Genome Browser[17] and Indian Genome variation database (IGVdb).[18] The CpG islands have been located using the 'DataBase of CpG islands and Analytical Tool (DBCAT).[19] The DNA methylation possibilities have been studied using the epigenome browser, EpiGRAPH.[20] 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)[21] and'comparative toxicogenomics database (CTD).[22]

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 POLYPHENE 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 Linoleic 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 *Faecalibacteriumprausnitzii*in 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 mutagens leading into APC mutation and CRC include DDT, poly chlorinated biphenols, permethrin, 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.[23]

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.[25]

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 casei* rich 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

CONFLICT OF INTEREST

No conflict of interest are declared.

ABBREVIATIONS USED

CRC: Colorectal Cancer; **APC:** Adenomatous polyposis coli; **HGMD:** Human gene mutation database; **ICGC:** International cancer genome consortium; **HOMD:** Human oral microbiome database; **SNP:** Single Nucleotide Polymorphism; **SVM:** Support Vector Machine; **UV:** Ultra Violet; **NCBI:** National Center for Biotechnology Information; **SIFT:** Scale-invariant feature transform; **LOVD:** Leiden Open variation database; **UCSC:** University of California, Santa Cruz; **IGVdb:** Indian Genome variation database; **DBCAT:** DataBase of CpG islands and Analytical Tool; **HOMD:** Human oral microbiome database; **CTD:** Comparative toxicogenomics database; **DDT:** Di chloro di phenyl trichloro ethane (DDT); **ROS:** Reactive Oxygen Species; **HER2:** Human epidermal growth factor receptor 2; **COX2:** Cyclooxygenase-2; **BCL2:** B-cell lymphoma 2.

REFERENCES

1. Behera B, Dash J, Pradhan D, Tripathy G, Pradhan R. Apoptosis and Necrosis of Human Breast Cancer Cells by an Aqueous Extract of Euphorbia hirta leaves. J Young Pharm. 2016;8(3):186–93.

2. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, *et al*. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012: Globocan 2012. Int J Cancer. 2015;136(5):E359–86.

3. Pathy S, Lambert R, Sauvaget C, Sankaranarayanan R. The Incidence and Survival Rates of Colorectal Cancer in India Remain Low Compared With Rising Rates in East Asia: Dis Colon Rectum. 2012;55(8):900–6.

4. Consolidated Report of Hospital Based Cancer Registries 2007-2011. Banglore: National Centre For Disease Informatics And Research National Cancer Registry Programme Indian Council Of Medical Research; 2013.

5. Dr. V. K. Srivastava. Consensus Document For Management of Colorectal Cancer. Ansari Nagar, New Delhi – 110029: Indian Council of Medical Research; 2014.

6. Namboori PKK, Vineeth KV, Rohith V, Hassan I, Sekhar L, Sekhar A, *et al*. The ApoE gene of Alzheimer's disease (AD). Funct Integr Genomics. 2011;11(4):519–22.

7. Sanjay KP, Karthikeyan S, Preethi MI, Krishnan NPK. Prediction of Epigenetic Variations in Alzheimer's disease Identification of Ethnic Variants through Pharmacogenomic Approach. Res J Pharm Biol Chem Sci. 2016;7(4):2742–5.

8. Terra Organic and Natural Foods [Internet]. [cited 2016 Jun 16]. Available from: http://tinyurl.com/c2lgljc

9. Anusooya NJ, Krishnapriya AS, Deepak OM, Aparna N, Namboori PKK. Individual Variation In p53 Responsiveness: A Pharmacogenomic Approach. World J Pharm Pharm Sci. 2014;3(2):2059–69.

10. HGMD® home page [Internet]. [cited 2016 Jan 6]. Available from: http://www.hgmd.cf.ac.uk/ac/index.php

11. Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, *et al*. International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data. Database. 2011;2011: bar026.

12. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, *et al*. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308-11.

13. GeneCards-Human Genes | Gene Database | Gene Search [Internet]. [cited 2016 Mar 10]. Available from: http://www.genecards.org/

14. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31(13):3812–4.

15. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. In: Haines JL, Korf BR, Morton CC, Seidman CE, Seidman JG, Smith DR, editors. Current Protocols in Human Genetics [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2013 [cited 2016 Jun 16]. p. 7.20.1-7.20.41. Available from: http://doi.wiley.com/10.1002/0471142905.hg0720s76

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

17. PubMed entry [Internet]. [cited 2016 Jun 8]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12045153

18. Indian Genome Variation Consortium. The Indian Genome Variation database (IGVdb): a project overview. Hum Genet. 2005;118(1):1–11.

19. Kuo H-C, Lin P-Y, Chung T-C, Chao C-M, Lai L-C, Tsai M-H, *et al*. DBCAT: Database of CpG Islands and Analytical Tools for Identifying Comprehensive Methylation Profiles in Cancer Cells. J Comput Biol. 2011;18(8):1013–7.

20. Bock C, Halachev K, Büch J, Lengauer T. EpiGRAPH: user-friendly software for statistical analysis and prediction of (epi)genomic data. Genome Biol. 2009;10(2):R14.

21. HOMD: Human Oral Microbiome Database [Internet]. [cited 2016 Jun 21]. Available from: http://www.homd.org/

22. Davis AP, Grondin CJ, Lennon-Hopkins K, Saraceni-Richards C, Sciaky D, King BL, et al. The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. Nucleic Acids Res. 2015;43(Database issue):D914-920.

23. Lu X, Yu H, Ma Q, Shen S, Das UN. Linoleic acid suppresses colorectal cancer cell growth by inducing oxidant stress and mitochondrial dysfunction. Lipids Health Dis. 2010;9(1):106.

24. Oleico E, De Acción M. Antitumor effect of oleic acid; mechanisms of action; a review. Nutr Hosp. 2012;6(27):1860–5.

25. Llor X. The effects of fish oil, olive oil, oleic acid and linoleic acid on colorectal neoplastic processes. Clin Nutr. 2003;22(1):71–9.

26. Bastide NM, Pierre FHF, Corpet DE. Heme Iron from Meat and Risk of Colorectal Cancer: A Meta-analysis and a Review of the Mechanisms Involved. Cancer Prev Res (Phila Pa). 2011;4(2):177–84.

27. Sinha R, Anderson D, McDonald S, Greenwald P. Cancer Risk and Diet in India. J Postgrad Me. 2003;(49):222–8.