## RESEARCH ABSTRACT:

## In Silico Prediction of Pathogenic vs. Non-Pathogenic Mutations in the APC Tumor Supression Gene

*By Michael Harrington*

Colon cancer is the second leading cause of cancer in the U.S. and causes over 50,000 deaths a year. The leading cause of sporadic colon cancer has been determined to be mutations in the Adenomatous polyposis coli (APC) gene. Clinical cancer genetics techniques are now common in healthcare as a tool of indentifying risk factors for cancer expression. Testing the genetic code of a patient can yield variances in genetic, but little is known about the effect of genetic variance. The goal of this research is to create a genetic block that will be the basis for determining the risk of deleterious mutations within the APC gene for the purpose of clinical genetic application. The scientific problem is to classify as either "pathogenic" or "not pathogenic" all variations within the APC tumor suppression gene.

The problem will be addressed via the in silico prediction using a compiled database of APC gene samples. They depend on carefully constructed multispecies protein sequence alignments based upon published model species genetic sequences. The model species are readily available due to the publication of many genome projects. The prediction of pathogenic vs. non-pathogenic will be determined by in silco methods based upon evolutionary conservation across the model species. "Missense variance in the APC gene at positions that are evolutionary constrained, "conserved", are often pathogenic, while positions that are not constrained are often neutral."[1]

Once the APC sample database has been compiled the program the sequences must be aligned for comparison. This is accomplished using the Clustalw2 algorithm: "ClustalW2 is a general purpose multiple sequence alignment program for DNA or proteins."[2] The alignment will clearly show which positions are evolutionary constrained and where there is a lack of conservation. The next step is to use to use a second algorithm to run human APC samples against the compiled APC sample database. This goal can be accomplished via Scale-invariant feature transform (or SIFT). "by using related sequences provided by the user. Your query sequence is the sequence that you would like prediction on (i.e. the sequence in which you have introduced the amino acid substitutions)" [3] The SIFT algorithm will identify variation from the model block and identify if it is within a highly conserved region or not.

The results can then be used to compile heat risk assessment maps that can indicate the relative level of risk within the human sequence that is under analysis. There are many published human APC sequences that can be analyzed to test the validity of this research method. In conclusion this research into the cause and effect of genetic variance of the APC tumor gene is at the leading edge of a collaborative effort to address a current deficit in clinical genetic diagnostic methods. This research is novel, and it is exciting because it have far reaching applications beyond the APC gene.

---

1 Sean V. Tavtigian,1_ Marc S. Greenblatt,2 Fabienne Lesueur,1 and Graham B. Byrnes,1 for the IARC Unclassified Genetic Variants Working Groupy . 2008 In Silico Analysis of Missense Substitutions Using Sequence-Alignment Based Methods. Human Mutation, Volume 29, Issue 11, pages 1327-1336

2 http://www.ebi.ac.uk/Tools/msa/clustalw2/

[3] http://sift.jcvi.org/www/SIFT_related_seqs_submit.html