# COMPUTATIONAL STUDIES TO INVESTIGATE THE ROLE OF REGULATORY MECHANISMS IN COLORECTAL CANCER (CRC)

A

PROJECT REPORT

*Submitted in partial fulfilment of the requirements for the award of the degree*

*of*

**BACHELOR OF TECHNOLOGY**

**IN**

**BIOTECHNOLOGY**

*Under the supervision*

*of*

**Dr. Tiratha Raj Singh**

**Professor**

*by*

**DECHEN PELDEN (211813)**

**MILAN RAI (211814)**

*to*


**DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS**

**JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY**

**WAKNAGHAT SOLAN-173234**

**HIMACHAL PRADESH INDIA**

**May, 2025**

# DECLARATION

We hereby declare that the work presented in the Project report entitled **"COMPUTATIONAL STUDIES TO INVESTIGATE THE ROLE OF REGULATORY MECHANISMS IN COLORECTAL CANCER (CRC)"** submitted for partial fulfilment of the requirements for the degree of Bachelor of Technology in Biotechnology at the **Jaypee University of Information Technology, Waknaghat** is an authentic record of my work carried out under the supervision of **Dr. Tiratha Raj Singh.** This work has not been submitted elsewhere for the reward of any other degree/diploma. We are fully responsible for the contents of our project report.

DECHEN PELDEN (211813) MILAN RAI (211814)

Department of Biotechnology and Bioinformatics.
Jaypee University of Information Technology
Waknaghat, Solan

# CERTIFICATE

This is to certify that the work which is being presented in the project report titled **"COMPUTATIONAL STUDIES TO INVESTIGATE THE ROLE OF REGULATORY MECHANISMS IN COLORECTAL CANCER (CRC)"** in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Biotechnology submitted to the Department of Biotechnology and Bioinformatics, **Jaypee University of Information Technology, Waknaghat** is an authentic record of work carried out by **Dechen Pelden (211813) and Milan Rai (211814)** during a period from January 2025 to May 2025 under the supervision of **Professor (Dr) Tiratha Raj Singh,** Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat. The above statement is correct to the best of our knowledge.

Dr. Tiratha Raj Singh

Professor

Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology

Waknaghat, Solan

Date: ……..

# ACKNOWLEDGEMENT

The initiation and the working of this project were due to guidance and assistance from a large number of people, and we are grateful for every contribution of the individual minds. We are fortunate to be receiving insightful ideas and help, which supported our project, and we want to thank everyone who gave their time and expertise to this project.

We want to add a few heartfelt words for those individuals who helped in the completion of this project. It is well known that behind every success there stands some person other than the success itself. Firstly, we would like to express our gratitude and special thanks to Dr. Tiratha Raj Singh (Project Supervisor), who gave us this golden opportunity to do a project on our own chosen topic (colorectal cancer), which also helped us in doing a lot of research and came to know about many things. Thanks to his noble ideas and leadership, we were able to draw a clear and proper conclusion for my project.

Secondly, we would like to express our appreciation to Somlata Madam for her help and advice in seeing this project through to completion. We also want to take this opportunity to thank our parents for their unwavering moral and financial support.

Finally, we are thankful to all those people who directly or indirectly helped me in the completion of this project and who were involved in giving suggestions and recommendations

# TABLE OF CONTENTS

# LIST OF ABBREVIATIONS

- CRC- Colorectal Cancer
- TNM -Tumor, Node, Metastasis
- MSA - Multiple Sequence Alignment
- BLAST –Basic Local Alignment Search Tool
- WNT signalling pathway - Wingless-type MMTV integration site signalling pathways
- MAPK signalling pathway - Mitogen-Activated Protein Kinase signalling pathway
- APC - Adenomatous Polyposis Coli
- BRAF- B-Raf Proto-Oncogene, Serine/Threonine Kinase.
- KRAS - Kirsten Rat Sarcoma Viral Oncogene Homolog
- GTP – Guanosine Triphosphate
- GDP - Guanosine Diphosphate
- mCSM - mutation Cutoff Scanning Matrix
- MDS – Molecular Dynamic Simulation

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF GRAPHS

# ABSTRACT

Colorectal cancer (CRC) is listed as the third most common and worldwide killer in human kinds**.** CRC is caused by various genetic along with epigenetic factors. The main challenges of cancer are in its early diagnosis. Even if the cancers are identified, it is already in a later stage where treatment and other therapies remain unsuccessful. To tackle these challenges, this project focuses on finding those factors involved in causing cancer (CRC). This study aims to conduct computational studies to gain a better understanding of the mechanisms involved in CRC. It entails the use of numerous bioinformatics techniques, including the analysis of multiple genes at the nucleotide and protein levels, to determine the pathways and their interactions. It employs Multiple Sequence Alignment (MSA), which uses both nucleotide and amino acid sequences. Our findings demonstrated that the WNT signalling route and MAPK signalling pathway of the APC, BRAF and KRAS genes, linked with the progression and development of CRC in humans. They play an important part in CRC, with mutations causing a deviation from the usual pathway, resulting in uncontrolled growth and division. In the SNP analysis, which included both sequence and structural analysis, we identified 4 mutants for APC (G2227C, R414C, R2228G, and S643P), 4 mutants for BRAF (G464V, I463S, K483Q, and V600E), and 3 mutants for KRAS (G60S, K5E, and N116S). Further analysis, including ConSurf and mCSM analysis, was conducted for the wild type and each mutant**.** The work also contributes to a better knowledge of the mechanisms connected with CRC, as well as future investigations to improve patient outcomes for better treatment.

**Keywords**

Colorectal Cancer, Regulatory Mechanisms, Bioinformatics, Computational Studies, KRAS Signalling Pathway, Therapeutic Targets, Analysis, Single Nucleotide Polymorphisms

# CHAPTER 1: INTRODUCTION

CRC is a malignant tumour that is grown and developed in the colon and rectum, the lower part of the human digestive tract. It usually grows as a small lining on the inner side of the intestine known as polyp; however, it is to be noted in mind that not all polyps are harmful. To consider polyps as cancer-causing, they must meet certain requirements like if there are the presence of more than three polyps and if the length of the polyps is larger than 1 cm.

Changes in bowel habits like diarrhoea or constipation, rectal bleeding or blood in stool, discomfort in the belly, feeling tired and weak and loss of weight are some symptoms of CRC which are not portrayed right from the day when the polyps develop into cancerous cells but only after the second stage of cancer. Sometimes, the above symptoms are even caused by some other issues like infections, haemorrhoids, irritable bowel syndrome, etc. It often holds no symptoms, and to tackle it, screening tests like FIT (Faecal immunochemical testing), stool DNA testing and many more are recommended by the doctors and are even carried out. Because of the above reasons, people fail to detect cancer and even if it is detected, treatment fails on them.

Diagnostics of CRC can be done from the presence of aberrant gene promoter methylation in the plasma or serum of the patients. The following three genes APC, BRAF and KRAS were found associated with CRC and with the aberration in DNA methylation and hydroxymethylation present on the same promoter region. Alongside, miRNAs were also found to be associated with the above three genes. To further have a broader knowledge, SNP analysis and other Bioinformatics tools analysis were carried out.

The study's goal is to explore many mechanisms involved in CRC using computational techniques. We used various data to identify the main regulatory elements, such as transcription factors, microRNAs, and mutations, that are critical for CRC development. Furthermore, two pathways, the WNT pathway and the MAPK pathway, were narrowed down to involve the APC, BRAF and KRAS, which are dysregulated in CRC. This work seeks to provide a thorough picture of the mechanism in CRC so that, in the future, it can improve treatment for CRC patients.

# CHAPTER 2: LITERATURE REVIEW

## 1. Hydroxy methylation

Hydroxy methylation is a process where 5 methyl cytosines (5mc) is oxidized into 5 hydroxy methylcytosines (5hmc). It even includes the oxidation to 5 formyl cytosine (5fC) and 5 carboxlycytosine (5caC) with TET (Ten-eleven translocation) family methylcytosine dioxygenase presence. TET family belongs to 5- oxoglutarate dependent dioxygenase with BEP repair pathways. This is an oxygen-dependent reaction. TET is an epigenetic DNA modifier which plays a role in physiological roles and their functions in health and diseases in humans [1] TET 1 was initially discovered through their involvement in myeloid leukaemia followed by the rest 2. In mammals, the TET family consists of TET1, TET2, and TET3 genes that encode proteins with a double-stranded β-helix-fold and a cysteine-rich region in the catalytic domain [2]. TET gene plays a crucial role in the oxidization of 5mC to 5hmC. Furthermore, it can even oxidize 5-hydroxymethylcytosine iteratively to 5-formyl cytosine (5-fC) and 5-carboxy cytosine (5-caC) [2]. TET expression varies between cells/organs; TET2 and TET 3 are expressed in various tissues, while TET 1 is enriched in ESCs [3]. Its unique characteristics are discussed below:

a. TET 1 is a protein that converts 5mC to 5hmC with limited accessibility of DNA to DNMTs with a strong binding with unmethylated CpG sites via its CXXC domain. It is highly expressed in ESC (embryonic stem cells).
b. TET 2 is a protein that converts 5mC into 5hmC. It is expressed in ESC and hematopoietic cells.
c. TET 3 has a similar role to TET 2 but it is highly expressed in oocytes and zygotes

TET 1 and TET 3 work together and have an N-terminal CXXC DNA-binding domain but TET 2 exists as a separate gene (IDAX or CXXC4), which seems to have lost its during evolution and additionally, it encodes an inhibitor of Wnt signalling [4] TET 1 and 2 are expressed higher along with a high rate of differentiation when downregulated. On the other hand, TET 3 is expressed slowly and is upregulated during differentiation. TET 1 plays the role of conversion of 5mC to 5hmC in both cultured cells and in vitro [5].TET proteins are mammalian homologs of the trypanosome base J binding proteins, JBP1 and JBP2, that contain catalytic motifs typical of Fe(II)- and 2-oxoglutarate (2OG)-dependent dioxygenases and require α-ketoglutarate as a co-substrate for enzymatic activity. These results, collectively, highlight the crucial function of TET-mediated 5-hmC alteration in developmental processes and the likelihood of

dysregulation of 5-hmC being related to disease[6]. The methylcytosine hydroxylase work is further stretched to TET 2 and TET 3, and the removal of TET1 hinders mESC self-renewal and maintenance [7]. The 5-hmC, also known as the "sixth base," was initially found in Purkinje cells, granule cells, and mouse embryonic stem cells (ESCs) as an intermediate of active DNA demethylation [4]**.** 5hmC is found in high-level organizations where the number peaks in the brain and the tissues of the central nervous system [7].

5mC occurs as sole methylation in eukaryotes, where methylation occurs in sequence-specific as well as locus-specific manner. Methylation occurs mainly in CpG sites where there is a presence of a minute of 5mC. 5hmC is the blood biomarker to identify different stages of cancer and distinguish between the stages of cancer. Many numbers of lives could be saved if the cancers they are suffering from are detected at their earlier stages alongside their treatment. To do so, liquid biopsy is the one to detect cancer in the body's fluid like blood and urine. Liquid biopsy is an important approach to detect cancers during its early stage along with the challenges while designing it. Liquid biopsy can even detect the minute biomarkers released by the tumour cells. It could be tallied that from the cell-free DNA (cfDNA) 0.1 to 1 % are represented by circulating tumour DNA (ctDNA) during early diseases. The change in cfDNA, DNA methylation and DNA hydroxy methylation are widely studied and investigated for cancer detection.

5mC and 5hmC both play a distinct role i.e 5mC is found in the heterochromatin and euchromatin and is tagged with gene expression whereas 5hmC, the oxidative product of 5mC, is mainly found in euchromatin and represents a transition state towards the removal of gene repression plus highly transcribed gene bodies. 5mC and 5hmC can bind to Methyl Binding Protein (MBD) and UHRF2 gene respectively. MBD is composed of three different families- the MBD family, the kaiso and kaiso-like proteins and has the assets to comprehend the methylated DNA These proteins can modify and silence their transcription in the nearing chromatin. UHRF is a gene (protein coding) that is ubiquitin-like with PHD and ring finger domains 2. UHRF 1 gene has the affinity to bind to both 5mC and 5hmC but UHRF 2 gene has affinity only with 5hmC. After mitosis, both 5mC and 5hmC are replaced.

5mC and 5hmC have different biomarkers used for identifying cancer. Studies are going on to overcome the limitation of distinguishing between 5mC and 5hmC. With time, a new technology emerged which can quantify and use 5hmC for cancer testing using liquid biopsy. In CRC, the use of 5hmC as biomarkers could not be of much use as it decreases its detection

ability as the cancer progresses to later and metastatic cancers. The chemical composition of mC and hmC vary in their polarity and sizes. mC is hydrophobic while hmC hydrophilic hence, can form hydrogen bonds and is bulkier than the methyl group and therefore needs a larger binding pocket to be accommodated.

## 2. DNA methylation

DNA methylation is accounted as an epigenetic modification because it involves the inheritance of changes by cell division rather than DNA sequence changes [8]**.** DNA methylation plays a role in the modifications of DNA in most of the cancer cells. It adds methyl group at the fifth position of the cytosine with the help of DNA methyltransferase (DNMT), which controls gene expression and stable gene silencing[9]**.** DNA methylation is usually found to occur in the region with low CpG density known as CpG shores, which are close to CpG islands [10]**.** CpG shores is a 2kb long region found on the left and right sides of the CpG islands[11]**.** In 70% of human genes, CpG island comprises 1000bps in length with more than 50% GC content forming a cluster of CpG nucleotides on the promoter regions [12]. The CpG dinucleotide's distance has an impact on its likelihood of being methylated; the greater the distance, the more likely it is to be methylated[13].

The distance between CpG dinucleotides has impact on methylation probabilities; genes with high 5mC content in their promoters are in non-transcriptional state resulting in accumulation of DNA methylation and long-term gene silencing. However, sometimes it accumulates microbes in germ cells and pluripotent cells due to its low CpG density. In comparison with promoter CpG island, methylated cytosines are symmetrically arranged. The cytosines of CpG in CpG island shores have a lesser number of slightly or moderately methylated. Later a study concluded the negative association of gene expression with methylated levels at CpG shore [11]**.**

In the mammalian family, DNMT is broken down into DNMT1, DNMT3A, DNMT3B and DNMT3L grouped into maintenance and denovo DNMTs [9]**.** Along with these and the TET family, all help in DNA methylation. The methylation state of individual CpG sites is not as heritable as the average DNA methylation level of DNA regions. All these enzymes (DNTM) have a C-terminal and an N-terminal.

The C-terminal is known for its catalytic proportion and the N-terminal is a region with highly conserved amino acids and is of variable size with regulatory domains which take part in DNA, chromatin and other proteins. N-terminal and C-terminal regions have 621 amino acids and 500 amino acids, respectively [14].

DNA methylation patterns are established early in the zygote by the de novo DNA methyltransferases DNMT3A and DNMT 3B) and they are conserved during cell divisions by the maintenance DNA methyltransferase DNMT1. These enzymes transfer the methyl group from S-adenosylmethionine to cytosine (usually in a CpG context), producing 5-methylcytosine (5-mC) [15]. It includes two active de novo DNMTs, DNMT 3A and DNMT3B and inactive regulatory factor DNMT 3 like protein DNMT 3L[16]. During embryonic development in mice, DNMT3A and DNMT3B play essential roles. The removal of DNMT3B leads to embryonic death, the removal of DNMT3A results in postnatal death [17] and the removal of DNMT3L is not related to death but results in sterility in male mouse [18].

In mammalian, mutation in DNMT3B gene leads to ICF (Immunodeficiency, Centromere instability, Facial abnormalities) syndrome is a rare autosomal disease [19] and in case of DNMT3A causes overgrowth disorders [20]. DNMT 2 also known as TRDMT1 is the mammalian ‘s smallest DNTM and it has the potential to methylate RNA more than DNA [21][22]. It lacks the regulatory N-terminal region, but its C-terminal domain contains all 10 conserved sequence motifs of DNMTs, including the S-adenosyl-l-methionine-binding motifs [23]**.**

Recent discoveries about the mechanisms of methylation and its regulation have led to the identification of numerous regulatory proteins and enzymes. Scientists are currently investigating the effects of dietary folate and methylene tetrahydrofolate reductase polymorphisms on methylation patterns in normal and cancerous tissues. As methylation occurs early and can be detected in bodily fluids, it has the potential for use in early tumour detection and determining prognoses. Additionally, since DNA methylation is reversible, treatments using drugs such as 5'-azacytidine, decitabine, and histone deacetylase inhibitors are being developed to treat various types of tumours [24]**.** The pattern of DNA methylation in a given cell appears to be associated with the stability of gene expression states [25].

DNA methylation can broadly be classified into hypomethylation and hypermethylation.

### A. Hypomethylation

Is associated with repeated DNA sequences like Long Interspersed Nucleotide Element 1(LINE 1). LINE 1 can make a duplication of itself and move towards the genomic site. It affects the genome structure and stability causing the initiation of the cancer. The cancer risk confirmed by the association between LINE 1 hypomethylation can be determined in the tissue samples both in fresh or frozen fixed paraffin-embedded (FFPE) tissues than in blood samples. LINE 1 hypo methylation was found to cause higher CRC in the early cancer stage.

### A. Hypermethylation

Associated with CpG island of the promoter region. At the cancer initiation, the CpG island undergoes hypermethylation, resulting in the inactivation of the tumour suppressor gene, which causes genomic instability and chromosomal aberrations. Hypermethylation of CpG islands in cancers is associated with transcriptional silencing of the genes in which this change occurs [26]. It marks the presence of tumour suppressor genes (TSGs) for hypomethylation.

Hypomethylation counts more than hypermethylation during carcinogenesis. Hypomethylation and hypermethylation work solely towards the progression of cancers. However, hypomethylation was found more abundant than hyper-methylation. In a study, it was even concluded that the hypomethylation count was mainly in men and hypermethylation in women. Hypomethylation main role in the early progression of the tumour but on the other hand hypermethylation was found to accumulate in much later advanced stages. In that same study, carcinomas showed a higher expression level of hypermethylation than adenomas. The expression level of hypomethylation was the same in both carcinomas and adenomas[27].

Growth-regulatory genes are epigenetically silenced in colonic mucosa in elder individuals, which may increase the risk of cancer associated with ageing[28]. Hypomethylation of ASC/TMS1 was reported to be a late-stage event in CRC, [29] whereas hypermethylation of SOX17 is an early event of CRC. Methylation of hMLH1, p16, DAP-kinase, APC, MGMT, RASSF2A, and Wif-1 were regarded as plasma or serum detection markers [30].

## 3. miRNAs

Pathologic changes result from an abbreviation in protein-coding proto-oncogene and tumour suppressor genes. However, later discoveries and research in the area of miRNA led to a broader understanding of the role of non-protein-coding genes in carcinogenesis. It can alter cell proliferation, apoptosis, metastasis, angiogenesis, and multidrug resistance through interaction with signalling networks. It even affects the genetic and genomic stability [31].

The metastasis cancer cells can acquire the properties of the mesenchymal cell, which is a multipotent stem cell found in the bone marrow; essential for making and repairing skeletal tissues like cartilage, bones, and fats found around the bone marrow, etc. However, it is to be noted that with time and age, MSC get converted into lipid-accumulating fat cells. It occurs by a phenomenon known as epithelial-mesenchymal transition (EMT), which pulls itself out from the primary tissues [32]**.** Despite the presence of physiological barriers, this metastasis can pass its way towards the secondary target tissues through local invasion, intravasation, survival in the circulating system, extravasation, colonisation and MET. The MET process runs a process that runs in different ways compared to the EMT process [33]. Metastasis plays a core role in causing death in CRC patients, due to which the detection of molecular mechanisms is a must and can help prevent and treat CRC. Surgery and chemotherapy are the primary treatments for any cancer [34]**.**

Metastasis plays a core role in causing death in CRC patients, due to which detection of molecular mechanisms is a must and can help prevent and treat CRC. To treat CRC, several treatment options are developed, including surgery, chemotherapy, radiation therapy, and targeted therapy [35]. Surgery and chemotherapy are the primary treatments for any cancer [34]**.** The cellular and bioprocess like inflammation, stress responses, migration, cellular degradation, cell cycle progression, apoptosis, metabolism, MET transcriptional factors and progression of CRC, in all these processes, miRNA is involved and has a hand in chemoresistance, chemosensitivity, resistance to radiotherapy and survival of the tumour cells [36].

miRNAs are 22 nucleotides, a large family of single-stranded, short non-coding endogenous and an important class of posttranscriptional regulators. It is produced through multiphase biogenesis which makes use of RNase II Dicer enzyme in endogenous hairpin transcriptions. When miRNA expression deviates from its normal pathway, it often holds a bidirectional effect, which results in oncogenesis or tumour suppression [37]. Based on their sources, miRNA is of cellular and cell-free circulating miRNA [33]. Almost in all the species, miRNA is highly conserved and protected and is a part of the translational controller in some vertebrates, worms and plants [36]. Transcriptional regulation, epigenetic methylation of miRNA containing loci, miRNA processing pathways, sequestration with long non-coding RNA or miRNA sponge and all hold the right to cause deviation in miRNA expression [38][39]

Each miRNA has a distinct role in colorectal tissue expression. A single miRNA can regulate multiple targets. Despite accounting for approximately 1-3% of the human genome, bioinformatics analysts reveal that miRNA may regulate up to 30% of human genes [40]. Changes within the miRNA binding site hold a higher risk of CRC whereas miRNA from the faeces or the blood can be used as a biomarker for early diagnosis and analysis of CRC progression [41] [42].

With peaking interest in the study of noncoding genomic sequences, it gave the classes of noncoding RNA as antisense RNA, small nucleolar RNA and microRNA in carcinogenesis. Because RNA acts as an intermediary in transmitting genetic information from DNA to protein, it plays a relatively passive function in carcinogenesis. Of all the RNA classes, miRNA got the most attention [43]. Based on studies and the conclusion drawn found that miRNA loci are frequently annotated within chromosomal regions [44].

The biological functions of miRNA to the targeted mRNA are determined by the binding to the 3’ end of the untranslated regions (UTR) resulting in the curtailment of 30% of the expression of a protein-coding gene with a formation of incomplete hybrids at 3’UTR regions of messenger RNA [40]. Most miRNAs contain their promoters and regulators, enabling them to operate as an independent transcriptional unit. miRNA contains both intervening introns and coding exon genes which are transcribed along with their host genes. RNA polymerase II helps transcribe the miRNA gene to primary miRNA (pri-miRNA).

Primary miRNAs are split into pre-miRNAs which are of 70-85 nucleotides and double-stranded strands, one with 22 nucleotides known as mature mRNA and the other as a complementary strand [45]. Primary miRNA contains 5'cap, the 3′ poly(A) tail and sequences flanking the hairpin structure. It is then exported from the nucleus to the cytoplasm by exportin 5. The endonucleases then further process primary miRNA into short double-stranded RNA fragment which consists of mature miRNA strands and their complementary strands [40].

MicroRNP (miRNA with ribonucleoproteins) separates the two strands where one of the strands is changed into active miRNA and the remaining are all digested [45]. Because of its small size, miRNA hardly degrades which increases its stability. It even remains unchanged against chemicals, and physical and environmental pressure apart from the messenger RNA. 5' end of miRNA has a 2-8 nucleotide length known as seed sequence which helps in base pairing with the target mRNA molecules. With its asymmetric thermostability, endogenous mRNA is targeted for silencing by the RNA-induced silencing complex (RISC), which includes mature RNA [46].

How RISC causes translational repression is intricate and involves several mechanisms. RISC is composed of a single standard miRNA, Argonaute (AGO) and GW182 (TNRC6A) and is found to play a main role in the suppression of translation or degradation of the respective cells [47]. These include cap-dependent inhibition of translation initiation, recruitment of eukaryotic translation initiation factor-6 to RISC, degradation of nascent proteins, ribosomal drop-off, and hindrance of the interaction between poly(A)-binding proteins and eukaryotic translation initiation factor-4G post mRNA deadenylation [48]. miRNA targeting specificity depends on AU nucleotide composition and binding site position relative to the stop codon [49].

The biosynthesis of miRNAs consists of two processing steps. to generate a precursor miRNA (pre-miRNA). First, the initial miRNA transcripts (pri-miRNA) are micrpprocessed in the nucleus by a protein complex containing the Drosha enzyme and cofactor DGCR8 to produce precursor miRNA(pre-miRNA) [50]. To create a precursor miRNA (pre-miRNA), the parent miRNA transcript (pri-miRNA) is first processed in the nucleus by the protein complex microprocessor, which is made up of the cofactor DGCR8 and the enzyme Drosha [51].

Following pre-miRNA transport to the cytoplasm comes the second processing step, which is carried out by the enzyme Dicer and its cofactors, TAR (HIV-1) RNA binding protein 2 (TARBP2) and protein kinase EIF2AKS (PRKRA) is induced by the protein activator of the interferon [52]. It has been demonstrated that colorectal adenocarcinoma expression of Drosha, Dicer, and DGCR8 is up-regulated in comparison to non-neoplastic tissue[53].

i. Steps involved in miRNA biosynthesis

Many miRNAs have their own promoters and regulatory elements that allow them to function as independent transcriptional units. miRNA genes contain both intervening introns and coding exon genes that are transcribed together with their host genes. RNA polymerase II is responsible for transcribing the miRNA gene into primary miRNA (pri-miRNA). The precursor miRNA (pre-miRNA) and the parent miRNA transcript (pri-miRNA) are first processed in the nucleus by the protein complex microprocessor, which comprises the cofactor DGCR8 and the enzyme Drosha.

The parent miRNA transcripts into precursor miRNA (pre-miRNA)

1. miRNA transcripts into primary miRNA (pri-miRNA) with the aid of RNA polymerases II.
2. pri-miRNA contains 5'cap, the 3′ poly(A) tail and sequences flanking the hairpin structure. pri-miRNA in the presence of Drosha DGCR8 converts into precursor miRNAA (pre-miRNA). The above two steps take place in the nucleus.
3. Pre-miRNA is then exported to cytoplasm by exportin 5.
4. The endonucleases then further process primary miRNA into short double-stranded RNA fragment which consists of mature miRNA strands and its complementary strands [40]
5. Nucleases Dicer cleavages of the loop to generate the mature miRNA which is later incorporated into RISC.

The first two steps occur in the nucleus and the rest in the cytoplasm.

A single miRNA holds a strong authority as it can impose effects on many transcripts in different species and these effects can even happen in a vice versa manner too.

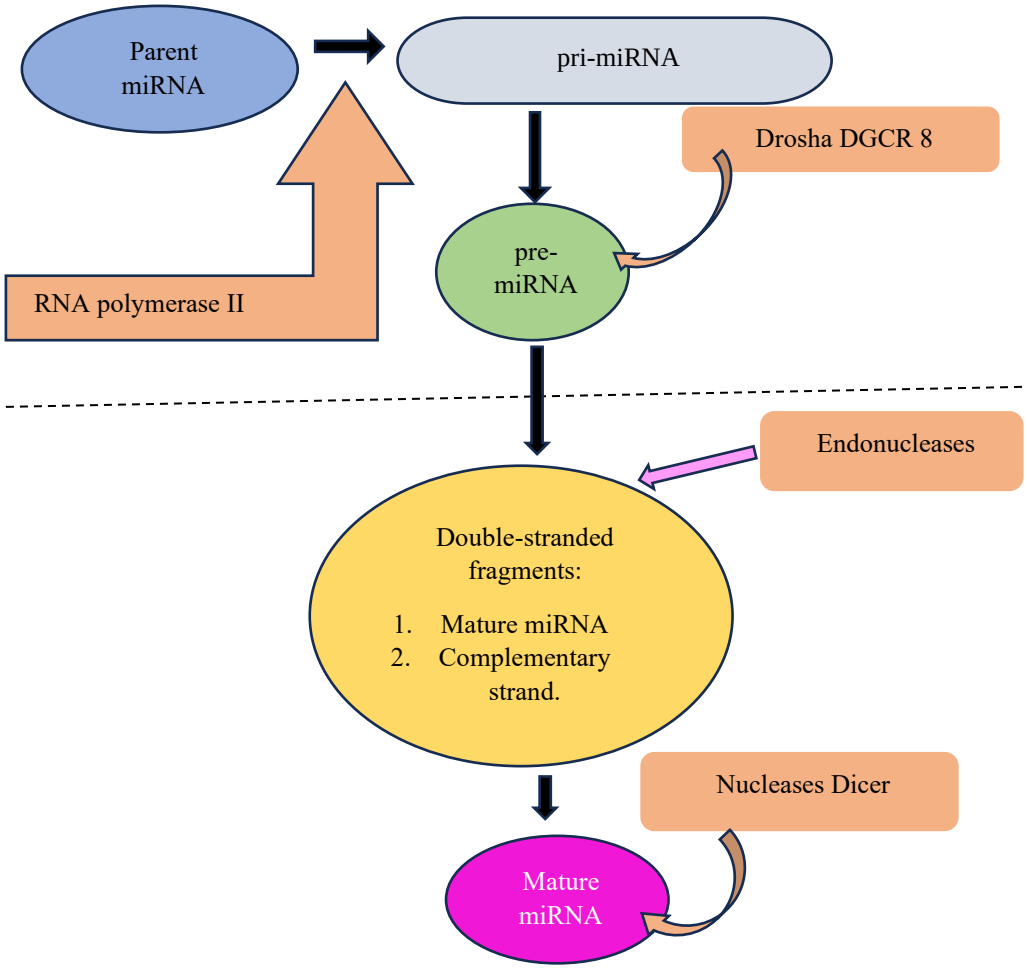



*Fig 1; miRNA synthesis*

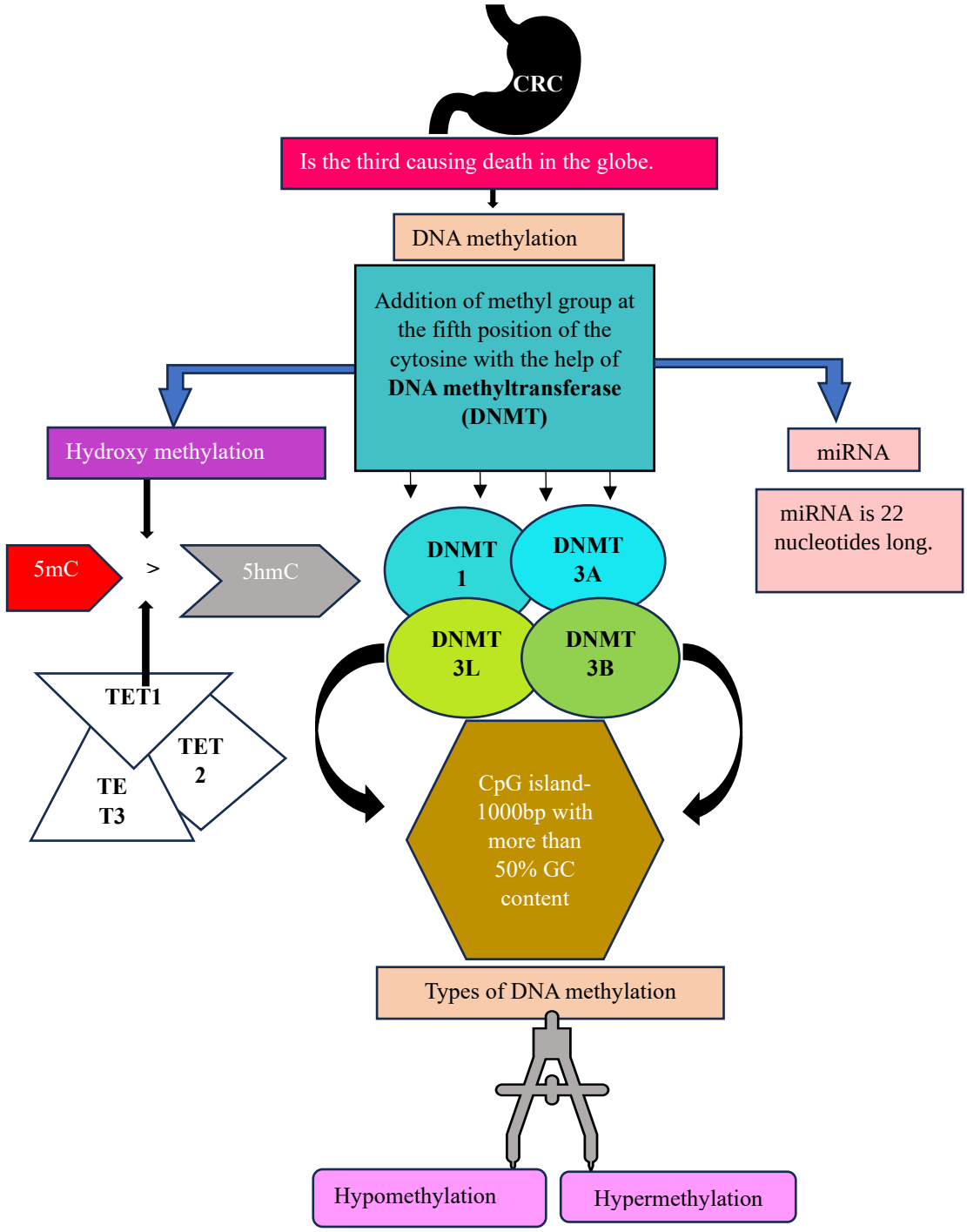



*Fig 2. Flow chart of CRC*

By going through many papers, numerous genes are linked to colorectal cancer. A few examples are listed in Table 1.

***Table 1;*** *The list of CRC-causing genes.*

| Sl.no | Gene | Acronym | Descriptions | Gene length (approx. base pair) | Chromosome no. |
|---|---|---|---|---|---|
| 1. | CNOT2 | CCR4-NOT Transcription Complex Subunit 2. | Encodes subunit of CCR4-NOT complex which regulates mRNA synthesis and degradation. | 26,645 | 12 |
| 2. | RPIA | Ribose 5-Phosphate Isomerase A. | It plays a role in the pentose phosphate pathway, a metabolic pathway related to the production of nucleotides and NADPH. | 6040 | 2 |
| 3. | MSH2 | MutS Homolog 2. | It encodes a mismatch repair protein which holds the role of correction of errors during DNA replication. | 73249 | 2 |
| 4. | MLH1 | MutL Homolog 1. | It encodes mismatch repair proteins which aids in error correction during DNA replication and maintains stability of the genome. | 19,218 | 3 |
| 5. | MSH6 | MutS Homolog 6. | It encodes mismatch repair proteins for error corrections in DNA replication. | 78,692 | 2 |
| 6. | PMS2 | PMS1 Homolog 2. | Mismatch Repair System Component.<br>It encodes for a mismatch repair protein for error correction during DNA replication. | 85,820 | 7 |

| | | | | | |
|---|---|---|---|---|---|
| 7. | APC | Adenomatous Polyposis Coli. | It encodes a tumour suppressor protein regulating cell growth. | 120,510 | 5 |
| 8. | KRAS | Kirsten Rat Sarcoma Viral Oncogene Homolog. | It encodes proteins that belong to the RAS family of oncogenes regulating cell growth and differentiation. | 33,321 | 12 |
| 9. | BRAF | B-Raf Proto-Oncogene, Serine/Threonine Kinase. | It encodes a MARK/ERK signalling pathway protein regulating cell division and differentiation. | 78,814 | 7 |
| 10. | TP53 | Tumour Protein p53. | It encodes a tumour suppressor protein regulating cell division and prevents mutations to prevent tumour formation. | 19,200 | 17 |
| 11. | SMAD4 | SMAD Family Member 4. | It encodes proteins which take part in the TGF beta signalling pathway regulating cell growth, differentiation and apoptosis. | 59,246 | 18 |
| 12. | MGMT | O-6-Methylguanine-DNA Methyltransferase. | It encodes proteins which aids in the removal of alkyl groups of DNA. | 20,191 | 10 |
| 13. | CTNNB1 | Catenin Beta 1. | It encodes proteins that take part in the Wnt signalling pathway regulating cell proliferation and differentiation. | 53,530 | 3 |

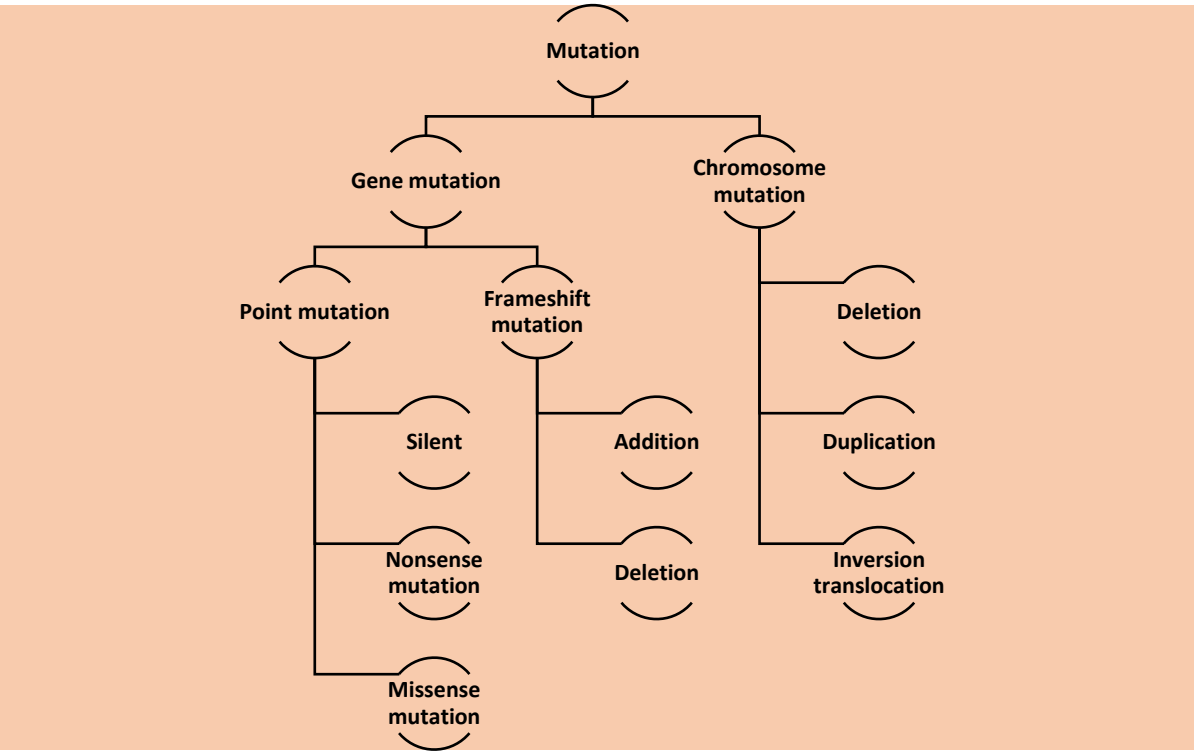
| | | | | | |
|---|---|---|---|---|---|
| 14. | MAL | T-Cell Differentiation Protein. | It encodes a protein that promotes the formation of myelin sheath in neurons. | 6006 | 2 |
| 15. | MYH | Myosin Heavy Chain. | It encodes proteins which are useful in muscle contraction. The gene length and the chromosome location depend on the specific gene family member. | VARIES | VARIES |
| 16. | EPCAM | Epithelial Cell Adhesion Molecule. | It encodes proteins that help in cell adhesion molecules in cell signalling and movement. | 9439 | 2 |
| 17. | NPAS | Neuronal PAS Domain Proteins. | It encodes for those proteins of the transcription factor family, regulating gene expression. The gene length and the chromosome location depend on the specific gene family member. | VARIES | VARIES |
| 18. | SEPT9 | Septin 9. | It encodes for a member of the septin family and plays a role in cytokinesis, cell morphology and many more. | 33,823 | 17 |
| 19. | TAC1 | Tachykinin Precursor 1. | It encodes P precursor substances which helps in neurotransmitters. | 11,316 | 7 |

**Mutation**

Mutations are alterations in an organism's DNA sequence brought either by errors made during DNA replication or by external stimuli such as UV light. Depending on where they occur and which functions they have changed, mutations can have varying effects on health. There are two categories for it: gene mutation and chromosomal mutation

1. **Gene mutation** is a variation in the DNA sequence of a gene that can alter the protein that the gene codes for.

2. **Chromosome mutation** is a change that takes place in an organism's chromosomes, either numerically or structurally. These alterations are caused by mutations that arise during meiosis or by environmental agents such as toxins, radiation, etc.



**Frameshift mutations**

When nucleotide bases are inserted or deleted which are not the multiple of three in a gene is commonly known as a frameshift mutation. Normally, cells read a genetic code composed of three bases to create a protein, making it crucial in any study. Those cancers with frameshift mutations generally display a higher level of genomic instability and are often found associated with advanced TNM (Tumor, Node, Metastasis) stage.

Patients at advanced stages of CRC are mostly being treated with DNA damage-inducing 5-fluorouracil-based chemotherapy which results in a successful survival rate for a certain group of populations [54]. From a study, it was concluded the following; UVRAG gene undergoing mutations in CRC, not only enhances tumour growth but also enhances metastasis and the patients develop sensitivity to chemotherapy. A frameshift mutation in the UVRAG gene gives a shorter form of a protein that nullifies tumour-suppressor functions of wild-type (WT) UVRAG contributing to the cancer cell growth and its invasion. The shorter form of a UVRAG protein holds no negative effect, instead, it can be utilised as a biomarker for the interventions of CRC due to its ability to sensitise the tumours of CRC to chemotherapy.

The CRC rates are influenced by the geographical features of the people living in the area. Often Asia, Africa and some Latin parts of America usually have lower CRC rates than the higher CRC-rated areas like Northern Europe, Australia, New Zealand, US and many more[55]. When those people from lower CRC-rated areas move to high CRC-rated areas, a study concluded that they are at higher risk of developing Cancer [55]. For a cell to change from adenoma to carcinoma, it is knee important to have seven distinct genetic changes, its specific and temporal accumulation of mutations is the most important to be noted [56]. Inactivation of the APC gene and activation KRAS gene are found to be associated with the initiation and progression of CRC [57]. CRC is associated with two instabilities: chromosomal and genetic instability. Genetic instability, also known as microsatellite instability, results in 12 to 15% of CRC cases, which are linked with Mismatch Repair (MMR) deficiency [58].

## 1. APC gene

APC is a complex 312 KDa protein which has many protein-protein interaction pockets like binding sites for β catenin, EBI and axin [59] holds many functions and undergoes many molecular pathways [60]. The complex protein comprises of the scaffold protein axin, GSK3β, casein kinase 1α, β-catenin, and the E3-ubiquitin ligase β-TrCP [59],[60]. When the complex of APC genes is inactivated, it results in the phosphorylation of β-catenin and in its degradation [61]. The APC gene's primary function is to negatively regulate the canonical Wnt signaling pathway by downregulating the β-catenin [60] and besides that, it may have additional nuclear functions, such as DNA repair and cell-cycle control, unrelated to the Wnt pathway [62].

Frameshift mutations in the APC gene are an early event in CRC progression. These mutations give a negative function impacting its tumour-suppressing functions [63]. The loss of its function results in uncontrolled cell proliferation and pro-carcinogenic characteristics [63]. The multifunctional protein APC has roles in Wnt signalling, cell structure, cell polarity and apoptosis [64], [65], [66]. From a study, it was concluded that, for the APC gene, common mutation sites were found at codon 1309 (5bp deletion, c.3927_3931 delAAAGA) and 849 (c.2547_2548delTA p.Asp849fsX62) with small deletion of 5 nucleotides (AAAGA) and 2 nucleotides (AT) respectively. In the first mutation, many polyposis was seen alongside the initial stage of CRC [67].

### Role of WNT signalling pathway

This pathway plays a crucial role in the progression and growth of CRC, with mutations in APC. It is important to carry out processes like proliferation, differentiation and cell maintenance [68].

Most malignancies, including CRC, are connected with this pathway when they stray from the usual pathway. Beta-catenin is implicated in the WNT pathway, which regulates cell growth and division [69], [70]. The typical APC protein keeps this activity updated. Normally, APC aids in the destruction of beta-catenin when it is no longer required, keeping cells from expanding out of control. However, when there are mutations in the APC gene, this control is lost, and the WNT pathway is constantly activated, resulting in uncontrolled cell growth and CRC. APC not only regulates the WNT pathway, but it also ensures that chromosomes are

correctly divided as cells multiply, keeps cells in the right shape and direction, and inhibits superfluous DNA replication [71].

APC gene mutations frequently result in a shortened APC protein that is unable to carry out these essential tasks. They also cause disruptions to the complex, which activates β-catenin-dependent transcription and contributes to oncogenesis [71].The APC gene becomes inactive as a result of mutations. The APC gene is shortened as a result of point mutations or deletions, which cause its inactivation. APC gene deviation leads to the buildup of β-catenin, which is transported to the nucleus and activates genes involved in cell survival and proliferation, ultimately contributing to colorectal cancer (CRC) [72].

APC the frameshift mutation occurs for AAAGA deletion nucleotides and even AT deletion. These mutations are also found to be associated with Familial adenomatous polyposis which results in shorten also known as truncation of the APC protein which all in all impairs its function and leads to the progression of the cancers. Summarised from a study showed 9 frameshift mutations and 4 nonsense mutations. All of these showed deletion or insertion of one or two base pairs most at simple nucleotide repeat sequences particularly within a gene.

C >T transition or a point mutation also occurs at CpG sites. Mutations in those sites include mismatch repair (MMR) deficiency. It then results in the progression of cancer especially Hereditary non-polyposis Colorectal Cancer (HNPCC). For the APC gene, both somatic and germline mutations are possible and are found at 5' half of the sequence which then results in premature and shortened proteins [73],[74],[75]. MMR plays a key role in the occurrence of adenomatous in HNPCC [76],[76] APC gene is also a partner for FAP causing gene which is associated with tumour initiation and plays the role of housekeeper for the cellular proliferation [77]. Inactivation of the APC gene plays a major role in the progression of the cancer, HNPCC.

The conclusion from a study showed, the mutations which occur in codon 463 and 1387 and 1403 and 1578 with shortened protein, are found to be associated with many diseases namely congenital hypertrophy retinal pigment epithelium (CHPPE) and Garden's syndrome respectively [78], [79]. AAAGA deletion mutation is also associated with osteomas, duodenal osteomas, ostemas of the jaw, glandular polyposis, desmoid tumours and many more [80]. A study done for the Singaporean FAP families, pointed out a hotspot for germline mutations of AAATA at codon 1309.

**Inactivation APC gene**

Inactivation of the APC gene accounts for 80-85% of CRC cases, an important event in the progression of CRC. When inactivated, it disrupts the Wnt signalling pathway, which regulates the degradation of β-catenin. When the APC gene deviates from its normal function; it loses its ability, hence, inside the nucleus, β-catenin amasses resulting in cell proliferation and forbids apoptosis [81]. All the processes above are crucial for the initiation of tumours and the transformation of benign tumours into cancerous ones. Apart from its role in Wnt signaling, the protein also plays a hand in maintaining chromosomal stability and cell adhesion [82], [83]. When its functions are disrupted, it results in chromosome instability, further creating a favourable environment for the development of CRC [81], [82]. It usually occurs through routes like mutations; frameshift and point mutation, deletion or epigenetic silencing.

## 2. BRAF

The BRAF is a serine/threonine protein kinase which encodes for B-Raf protein and is found on the 7th chromosome. It plays a guiding role in transmitting chemical signals to the nucleus from the outside of the cell. It is also a part of RAS/MAPK signalling pathway, regulating proliferation, division, growth, migration, apoptosis and other functions. This pathway is filled with RAS small GTPases, which activate the family of proteins, resulting in the phosphorylation and activation of MEK proteins further resulting in cell growth. This activation increases the BRAF activity 10 times more than the wild-type gene.

It is also popular as an oncogene, having immense potential to transmit normal cells to immortal cancerous cells when they undergo mutation. The type of cancer it causes is often tagged with poor prognosis and less effective therapies. The mutation accounts for about 10% of CRC patients [84].

In the BRAF gene, point mutations are more common than frameshift mutations, except in microsatellite instability-high (MSI-H) tumors, where mutations result in the accumulation of repetitive DNA sequences due to defects in DNA mismatch repair. MSI-H tumors also have additional mutations that affect tumor behavior and treatment response [85].

At around 96% of mutation of T1799A transversion in 15$^{th}$ coding gene can be seen in BRAF mutation. Studies have shown the less effectiveness of the treatment options like chemotherapy, anti-epidermal growth factor receptor therapies on CRC caused by mutated BRAF gene. The mechanisms by which BRAF-mutated metastatic colorectal cancer develops resistance to the therapies are complex and varied. Nevertheless, the combination of BRAF inhibitor shows a promising result to the CRC patients [86],[87].

**BRAF and signalling pathways**

RAS/RAF/MEK/ERK signalling pathway, plays an immense role in regulating growth and proliferation of the cells. Mutation in BRAF gene contributes in CRC progression with tumour behaviour and poor treatment response. The mutations upshots BRAF gene to become 10 times more active than the normal gene and driving an uncontrolled cell division [86].

**Activation of BRAF gene**

Deviations in normal BRAF gene results in a truncated gene. V600E substitution is the most frequent mutation seen in CRC. This mutation induces the activation BRAF gene, avoiding the needs for upstream signals like RAS activation resulting in uncontrolled growth and cell divison. Different cancers have shown varying biology and therapeutic responses when V600E substitution mutations are absent [88].

**3. KRAS**

KRAS gene is found at chromosome 12 and produces a protein-encoding for 188 amino acid residues known as K-Ras [89] . It weighs around 21KDa and is known for its GTPase function [89], [90]**.** It's responsible for encoding 4 small cytoplasmic proteins: H-Ras, K-Ras4a, K-Ras4b, and N-Ras.

It is a protein that delivers the outside signals of the cells like mitogens, differentiation factors, transcription factors and cell cycle proteins to the inner cell nucleus. It then relays a message for the cells to grow, divide, develop and differentiate into specific functions. Turning on and off of the KRAS revolves around the GTP molecule; the GTP molecules' attachment turns on and when the GTP is converted to GDP, it turns off.

The most well-known oncogene with a high mutation rate in non-small lung cancer, colorectal cancer, pancreatic ductal adenocarcinoma, and many other cancers, is this gene, which produces the protein K-Ras [91]. In colorectal cancer, KRAS gene undergoes point mutation rather than frameshift mutations. These mutations are important for the progression of CRC through the deregulation of the signalling pathway like RAS/MAPK signalling pathway. Observing frameshift mutations are not so common in KRAS gene however mutations in those genes can lead to unregulated cellular proliferation and progression of the tumours [92].

It is the most well-known oncogene with a high mutation rate in non-small lung cancer, colorectal cancer, pancreatic ductal adenocarcinoma, and many other cancers [93]. It accounts for approximately 25% of all cancers, leading to growth and the spreading to different body parts. KRAS gene is a crystal structure with 6 beta strands and 4 alpha-helical strands which are evolutionary conserved. It further comprises two isoforms namely KRAS-4B and KRAS-4A [94]. From a paper, it was pointed out that mutations in the KRAS gene were found at codon 12 with a G>A transition, often with combined alterations [95].

**Role of RAS/MAPK signalling pathway (Mitogen-Activated Protein Kinase)**

It is an essential mechanism in cells that controls various activities, including survival, differentiation, and proliferation [96]. Transmembrane tyrosine kinase receptors (RTKs) are initially activated by extracellular stimuli. RTKs then transcribe adaptor proteins and catalyse the hydrolysis of GDP to GTP on Ras [97]**.** After it is turned on, it initiates a cascade event in which it sends signals to RAF, which turns on MEK and ERK [96].

In particular, single point mutations are observed in exon 2 codons 12, 13, and exon 3 codon 61 of the Ras gene. This kind of mutation results in set proteins that are permanently activated; the GTP-bound conformation affects the activity of the ATPase. Mutations in the RAS gene, more especially in KRAS, cause the pathway to be abnormally activated, which prevents apoptosis and causes uncontrollable cell growth. Consequently, this regulation encourages the CRC to advance. Because of the KRAS gene mutations, even when the drugs inactivate the EGFR, MAPK signalling pathways stay activated [98].

KRAS mutations are linked to a poor prognosis in 40% of CRC cases. These mutations may cause resistance to particular medicines, making treatment challenging [99]. Scientists are still researching ways to inhibit the MEK and ERK components, though. Treatment outcomes may be improved by combining those inhibitors with CDK4/6 inhibition, as demonstrated in a recent trial[100]. According to a paper, KRAS mutations can be found using a variety of methods, such as direct sequencing, Next Generation Sequencing (NGS), and the Thera Screen KRAS kit. Each strategy has a specific sensitivity range. The NGS approach can detect any clinically relevant KRAS mutation with a sensitivity of 1% to 6% [101].

One innovative and practical method for identifying KRAS mutations is liquid biopsy. Since it requires less time, is less invasive, and provides complete molecular characteristics of the illness, all of which are essential for figuring out the nature of the disease and the most effective course of treatment is concluded to be more rewarding and satisfying [102]. The point mutation G>A at codons 12, 13 and 61 is found in CRC, pancreatic cancer and non-small cell lung cancer (NSCLC). This mutation results in the activation of KRAS gene which in a way results in the progression of cancer characterized by uncontrolled cell proliferation and survival by consistently downregulations of the signalling pathways.

CRC at 40%, pancreatic cancer at 92% and NSCLC at 29% are associated with point mutation G>A which was concluded from a study. It further pointed out the common variants G12D, G12V, G13D, and G12A where G12D and G12V listed as the most common amongst the list. The above substitution shows different effects on KRAS signalling pathway along with distinct effector proteins.

Role of KRAS gene

It plays an important role in cell growth and division. It acts as a protein production guide by acting as a molecular switch that tours around active and inactive states regulating various signalling pathways associated with cell growth. KRAS is associated with signalling pathways like RAS/RAF/MEK/ERK and PI3K/AKT/mTOR signalling pathways.

**KRAS and signalling pathways**

The RAS/RAF/MEK/ERK and PI3K/AKT/mTOR signalling pathways revolve around the KRAS gene, which plays a crucial role in cell division and survival. In colorectal cancer (CRC), mutations in the KRAS gene lead to abnormal activity in this pathway, resulting in tumour growth and progression [103]. Because of mutation in KRAS gene, those pathways are more often activated in CRC. it can account for 40 to 50 % of CRC because of this mutation.

**Activation of KRAS**

Activation KRAS gene accounts for 30 to 50 % of CRC cases affecting the tumours located on the right side. When it does this activity by acting as the molecular switch. This molecular switch is turned on and off by the presence of a GTP molecule, its attachment turns on and when the GTP is converted to GDP, it turns off. To further simplify that, GDP bound is inactive or off state and GTP bound is active or on state**.** In normal scenarios, the molecular switch operates regularly; however, in the presence of mutation, this switch or cycle is disturbed, resulting in the buildup of KRAS in an active state. As a result, accumulated KRAS deviates from its usual function, causing downstream MAPK and PI3K pathways, which are critical for cell proliferation and survival [103]. Those mutations that heighten the Guanosine triphosphate (GTP) binding, upshots the downstream of signalling pathways like MAPK and PI3K pathways [104].

**Hydrolysis of GTP to GDP**

The hydrolysis of GTP to GDP is an essential key for activating and inactivating the KRAS gene. It is assisted by the presence of the regulatory proteins, GTPase activating proteins (GAPs and guanine nucleotide exchange factor (GEFs). When KRAS protein is in an active state, it can bind to the effectors and cause deviation to the various signalling pathways. However, with the hydrolysis process, it switches off or goes to an inactive state without any effect on the signalling pathway. Due to the presence of intrinsic GTPase activity in KRAS, it allows the hydrolyse of GTP to GDP enabling the enzyme to auto inactive the signal propagation [105].

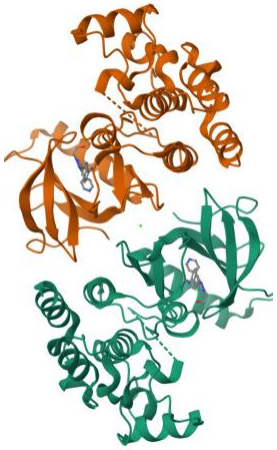
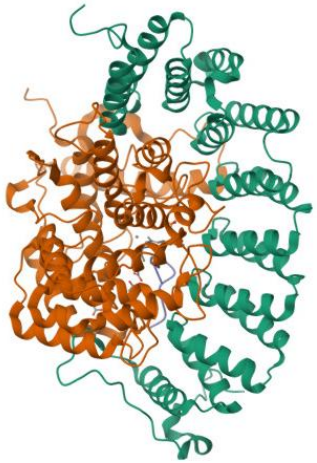
The uninterrupted cycle between active and inactive states of KRAS is important for regulating various signalling pathways in cells. GEFs also help in the converting of GTP to GDP however, it lowers the affinity of KRAS for GDP, when GEF interacts with KRAS enhancing GTP binding and activating the protein [106].

GAP also encourages the GTPase activity of KRAS and helps in hydrolysis of GTP to GDP much better than KRAS alone. This upshot in deactivation of the KRAS gene, ensuring that the pathways are tightly regulated and preventing prolonged activation resulting in the growth and progression of tumours [107].

**Ways to combat abnormalities in the genes:**

a. Lifestyle modification- promoting a healthy lifestyle like maintaining proper diet, doing physical activities, and avoiding various risk factors can help prevent it.
b. Therapeutic approaches – like doing gene therapy to restore gene's functions.
c. Genetic approaches – The use of gene editing technologies such as CRISPR/Cas9 and RNA interference (RNAi) can directly result in the suppression of mutations or detour their route.
d. Combination therapies- The combination of several treatments produces better results for induced carcinogenesis. For example, KRAS inhibitors may be used in addition to signaling pathway inhibitors.

*Table 2; Different genes of APC, BRAF and KRAS*

| | APC | BRAF | KRAS |
|---|---|---|---|
| PDB ID | 1DEB | 4MNF | 1D8D |
| Weight | Crystal structure of the N-terminal coiled-coil domain of the APC. | Crystal structure of BRAF-V600E bound to GDC0879. | Co- crystal structure of Rat protein Farnesyltransferases complexed with a Kras-4B peptide substrate and FPP Analog at 2.0A resolution. |
| Structure |  |  |  |

# CHAPTER 3: MATERIALS AND METHODS

## A. Multiple sequence alignment

A bioinformatics approach called multiple sequence alignment is used to align sequences of about equal length, such as those of DNA, RNA, and proteins. It facilitates pattern recognition and provides an evolutionary relationship visualization for the sequences. FASTA forms of sequences are gathered from NCBI databases and processed using a variety of algorithms, including CLUSTALW, MUSCLE, OMEGA, and many more. Following the completion of sequencing, phylogenetic trees and phylograms are constructed.

In MSA, three or more nucleotide sequences are aligned to find conservation or similarity regions that may indicate structural, functional, or evolutionary links. Understanding gene functions, regulatory components, evolutionary trends, and other topics requires the use of MSA.

A method called **BLASTn** (Basic Local Alignment Search Tool for nucleotides) is used to detect regions of similarity between a query nucleotide sequence and a nucleotide database. Finding sequences that are homologous to the query using BLASTn is frequently done as a first step before using them in MSA. Similarly, **BLASTp** (Basic Local Alignment Search Tool for proteins) finds similar regions by comparing a query protein sequence to a database of protein sequences. It is an effective bioinformatics tool for researching the structure, function, and evolutionary relationships of proteins. In order to identify similarities between a query sequence and sequences in a database, BLASTp aligns protein sequences. Because protein structure and function are conserved, BLASTp operates at the protein level, where evolutionary and functional insights are more noticeable than with nucleotide BLAST (e.g., BLASTn).

Phylogram and phylogenetic tree

A phylogenetic tree and a phylogram depict the evolutionary relationships between various genes, proteins, or species. Despite their similarities, they transmit different kinds of information. An illustration of the evolutionary links between species, genes, or proteins is called a phylogenetic tree. Although it may not always contain details regarding the degree of evolutionary change, it symbolizes the ideas of shared ancestry and divergence.

Branch lengths in a phylogram, a particular kind of phylogenetic tree, are proportionate to the degree of evolutionary change (e.g., genetic distance or number of mutations).

Symbols like *****, **:** , and **.** are frequently used in the alignment output of multiple sequence alignments (MSA) to show the degree of conservation (similarity) among the aligned sequences. Usually, alignment programs like Clustal Omega, MUSCLE, or MAFFT

I. *** (Asterisk**): Denotes total conservation and a place in the alignment where every sequence has the same residue (amino acid or nucleotide). It implies that this residue is essential to the molecule's structure or function (e.g., regulatory motifs in DNA, active site in proteins).

II. **: (Colon):** Denotes a place where all sequences contain comparable residues, while they are not identical; similarity is determined by biological characteristics (e.g., size, charge, polarity). For instance, lysine (K) and arginine (R) are both positively charged, whereas alanine (A) and glycine (G) are seen as being similar because of their tiny size. It implies structural or functional restrictions that permit the substitution of comparable residues.

III. **. (Period):** Denotes a place where all sequences contain weakly similar residues; weak similarity may nevertheless suggest some structural or functional importance, but with less assurance. It might indicate areas with greater variety and less intense evolutionary pressure.

IV. **Blank Space:** Denotes a location where the residues are neither conserved nor comparable. The areas that lack conservation are either less important structurally or functionally, or they might match non conserved areas like protein loops.

### B. Single Nucleotide Polymorphism (SNPs)

SNP is a single base change in a DNA sequence of a genome which results in genetic variations [108].SNP plays an immense role in identifying and studying different diseases and explaining the different responses to the treatment of drugs. Genome-wide association studies (GWAS) have pointed out more than 100 loci that can be associated with CRC [109]. SNPs can be found in both coding(exons) and non-coding (introns) regions and they can be accounted for more than 1% or about 1 in 1000 bases to 1in 100 to 300 bases in a population. SNPs found in coding regions (exons) cause structural changes in protein resulting in the alteration in its function too. With the help of GWAS, it was identified that about 90% of the SNPs were found to be located in non-coding regions [110] causing hindrance in identifying the genes and their functions.

One of the studies done recently concluded that few noncoding SNPs found in the potential regulatory element through long-range genome interaction can regulate the expression of distal genes and the mechanism that happens between noncoding SNPs and gene expression. With a fast process in the field of technology, in the earlier years, more than 1 million SNP assays for alleles that can be fitted in a small single chip which aids in genotyping were developed [111]**.** SNP alleles can broadly be of actual DNA sequence variants and indirect contributors. In the case of actual DNA sequence variant, it can cause structural and functional change and is found to be directly associated with a disease like Norrie disease, cystic fibrosis, etc whereas, on the other hand, the indirect contributor was found to be indirectly associated with a disease like diabetes, cancer, Alzheimer disease with less frequency. SNPs can be of great diversity and can result in the change in function regulation or expression of the proteins [112]**.**

For the betterment of all living and non-living like plants and animals, SNPs are often used as molecular markers. Even in the plant field, the use of SNP as molecular markers is peaking high [113]. The position of the SNP alleles in a DNA sequence can result in deviation in function and expression of the protein [114]**.**

### a. Sequence-based analysis

Tools used are: SIFT, PANTHER, PolyPhen2, PhD SNPs, Meta SNPs and SNP &GO

### b. Structure-based analysis

Tools used are: Dynamut, mCSM, SDM, DUET, CUPSAT and SNP&GO 3D

For both the steps, the data (nsSNPs) was retrieved from NCBI SNP database.

### C. ConSurf (Conservation Surface Mapping)

ConSurf is a bioinformatics tool used to examine the evolutionary rate of each protein in a family because of its high estimation accuracy. It is a bioinformatics tool used to analyze the evolutionary conservation of amino acids in protein structures. It sheds light on the structural and functional relevance of amino acid residues in diverse protein families. This program performs several sequence alignments to obtain the conserved portions of the protein family, which improves the identification of conserved sites during protein analysis.

### D. mCSM (mutation Cutoff Scanning Matrix)

A computational tool to predict mutation effects on protein stability, structure, function and interactions. ΔΔG represents the change in Gibbs Free Energy that this method uses to classify mutations as either stabilizing or destabilizing. A positive ΔΔG value indicates stabilizing, that the mutation stabilizes the protein, enhancing its stability. Conversely, a negative ΔΔG value suggests destabilizing, that the mutation destabilises the protein, which can contribute to instability and may be associated with diseases.

RSA% (Relative Solvent Accessibility Percentage) is a parameter used in this tool to measure the accessibility of protein residues within a solvent. It highlights the local environment of the amino acid residues in the protein and indicates whether they can interact with the solvent. A low RSA% value (around 5%) suggests that the residue is buried and hydrophobic, which can affect the stability of the protein. Alongside, a high RSA% value indicates that the residue is exposed and may be either hydrophobic or hydrophilic, and it is typically associated with less protein instability.

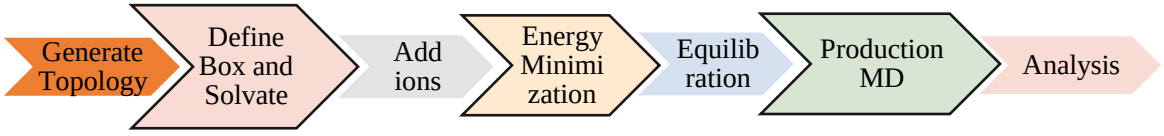
### E. Molecular Dynamic Simulations

It is a computer method for studying the movements and interactions of atoms and molecules. It allows the user to visualize and analyze the molecular system's evolution throughout time. It is widely utilized in a variety of scientific areas, including chemistry and biophysics, and aids in the refinement of 3D structures of proteins and other macromolecules, providing critical information for studies.

**MD Simulation Methodology**

1. Data collection from the PBD database.
2. Mutagenesis for the mutant-type proteins.
3. Use the GROMACS software by following the GROMACS Tutorial for Lysozyme in Water, the following steps are followed:

a. Generate Topology
b. Define Box and Solvate
c. Add ions
d. Energy Minimization
e. Equilibration
f. Production MD
g. Analysis



# CHAPTER 4: RESULTS AND DISCUSSIONS

## A. Multiple Sequence Analysis

## 1. APC

### 1.1 Multiple Sequence Alignment for Nucleotide Sequences

a. By performing blastn, we selected only the homo sapiens and got 8 sequences.

select all 8 sequences selected — GenBank | Graphics | Distance tree of results | MSA Viewer

| Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|
| Homo sapiens APC regulator of WNT signaling pathway (APC), transcript variant 16, mRNA | Homo sapiens | 15911 | 15911 | 100% | 0.0 | 100.00% | 10949 | NM_001407446.1 |
| PREDICTED: Pan paniscus APC regulator of WNT signaling pathway (APC), transcript variant X5,... | Pan paniscus | 15753 | 15753 | 100% | 0.0 | 99.66% | 12675 | XM_008952081.4 |
| PREDICTED: Pan troglodytes APC regulator of WNT signaling pathway (APC), transcript variant X5,... | Pan troglodytes | 15730 | 15730 | 100% | 0.0 | 99.62% | 9690 | XM_009449456.5 |
| PREDICTED: Gorilla gorilla gorilla APC regulator of WNT signaling pathway (APC), transcript varian... | Gorilla gorilla g... | 15681 | 15681 | 100% | 0.0 | 99.51% | 10960 | XM_055387556.2 |
| Homo sapiens APC regulator of WNT signaling pathway (APC), transcript variant 21, mRNA | Homo sapiens | 15616 | 15616 | 98% | 0.0 | 100.00% | 10713 | NM_001407451.1 |
| Homo sapiens APC regulator of WNT signaling pathway (APC), transcript variant 17, mRNA | Homo sapiens | 15610 | 15918 | 100% | 0.0 | 100.00% | 11102 | NM_001407447.1 |
| Homo sapiens APC regulator of WNT signaling pathway (APC), transcript variant 18, mRNA | Homo sapiens | 15610 | 15610 | 98% | 0.0 | 100.00% | 10869 | NM_001407448.1 |
| Homo sapiens APC regulator of WNT signaling pathway (APC), transcript variant 5, mRNA | Homo sapiens | 15610 | 15610 | 98% | 0.0 | 100.00% | 10758 | NM_001354896.2 |
| Homo sapiens APC regulator of WNT signaling pathway (APC), transcript variant 19, mRNA | Homo sapiens | 15610 | 15610 | 98% | 0.0 | 100.00% | 10866 | NM_001407449.1 |
| Homo sapiens APC regulator of WNT signaling pathway (APC), transcript variant 9, mRNA | Homo sapiens | 15609 | 15609 | 98% | 0.0 | 100.00% | 10605 | NM_001354900.2 |
| PREDICTED: Pongo pygmaeus APC regulator of WNT signaling pathway (APC), transcript variant X... | Pongo pygmaeus | 15464 | 15464 | 100% | 0.0 | 99.06% | 9863 | XM_054488715.1 |
| PREDICTED: Pan paniscus APC regulator of WNT signaling pathway (APC), transcript variant X3,... | Pan paniscus | 15463 | 15463 | 98% | 0.0 | 99.67% | 12440 | XM_008952082.4 |
| PREDICTED: Pan paniscus APC regulator of WNT signaling pathway (APC), transcript variant X1,... | Pan paniscus | 15463 | 15463 | 98% | 0.0 | 99.67% | 12551 | XM_008952080.4 |

*Fig 3. 1. 1. a; Showing the results of blastn*

b. Multiple Sequence Alignment

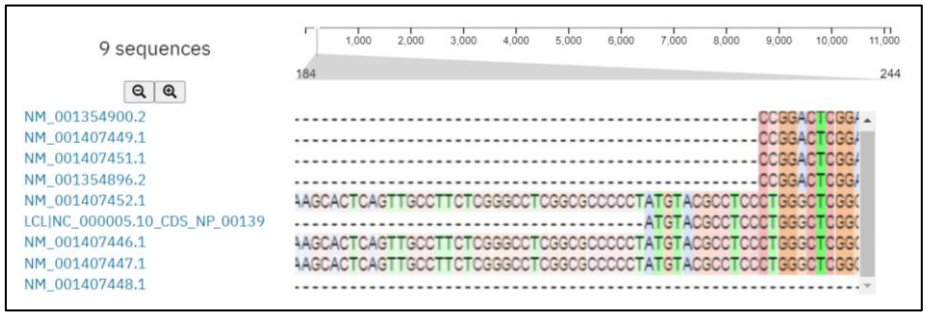
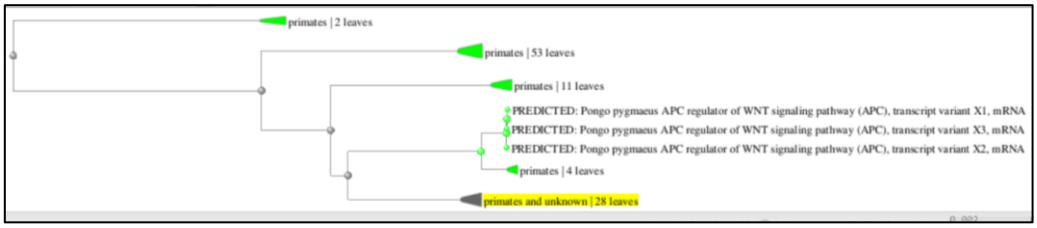



*Fig 3 .1. 1. b; MSA for nucleotide sequences.*

c. Phylogram

Fast minimum evolution




Neighbour-joining methods

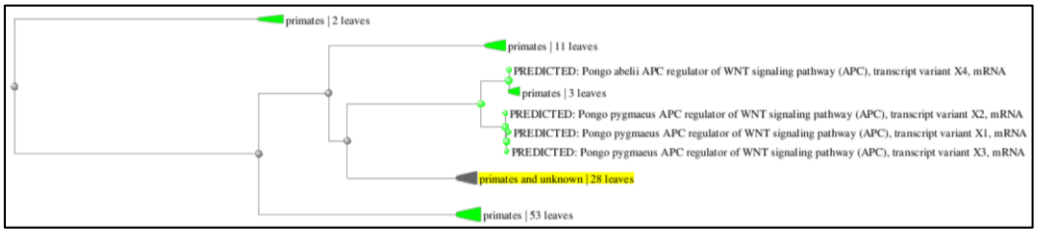



*Fig 3. 1. 1. c; Showing two different phylograms.*

## 1.2 Multiple Sequence Alignment for Protein Sequence

a. By doing Blastp, we got 13 sequences for Homo sapiens.

**Sequences producing significant alignments** Download Select columns Show 100

select all *13 sequences selected* GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

| Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|
| adenomatous polyposis coli protein isoform b [Homo sapiens] | Homo sapiens | 5852 | 5852 | 100% | 0.0 | 100.00% | 2843 | NP_000029.2 |
| APC [Homo sapiens] | Homo sapiens | 5848 | 5848 | 100% | 0.0 | 99.96% | 2843 | AAA03586.1 |
| adenomatous polyposis coli protein isoform c [Homo sapiens] | Homo sapiens | 5846 | 5846 | 100% | 0.0 | 99.37% | 2861 | NP_001341825.1 |
| adenomatous polyposis coli protein isoform X7 [Gorilla gorilla gorilla] | Gorilla gorilla gorilla | 5820 | 5820 | 100% | 0.0 | 99.51% | 2844 | XP_004042399.1 |
| adenomatous polyposis coli protein isoform X9 [Pan paniscus] | Pan paniscus | 5819 | 5819 | 100% | 0.0 | 99.54% | 2844 | XP_003826783.3 |
| polyposis locus-encoded protein [Homo sapiens] | Homo sapiens | 5812 | 5812 | 100% | 0.0 | 99.47% | 2844 | AAA60353.1 |
| adenomatous polyposis coli protein isoform X10 [Pan troglodytes] | Pan troglodytes | 5811 | 5811 | 100% | 0.0 | 99.47% | 2844 | XP_063666622.1 |
| adenomatous polyposis coli protein isoform X4 [Gorilla gorilla gorilla] | Gorilla gorilla gorilla | 5810 | 5810 | 100% | 0.0 | 98.88% | 2862 | XP_018882861.1 |
| adenomatous polyposis coli protein isoform X7 [Pan troglodytes] | Pan troglodytes | 5800 | 5800 | 100% | 0.0 | 98.85% | 2862 | XP_009447734.1 |
| adenomatous polyposis coli protein isoform X4 [Pongo pygmaeus] | Pongo pygmaeus | 5799 | 5799 | 100% | 0.0 | 99.23% | 2844 | XP_063521205.1 |
| adenomatous polyposis coli protein isoform X2 [Pongo pygmaeus] | Pongo pygmaeus | 5792 | 5792 | 100% | 0.0 | 98.60% | 2862 | XP_054344692.1 |
| adenomatous polyposis coli protein isoform X4 [Pongo abelii] | Pongo abelii | 5791 | 5791 | 100% | 0.0 | 99.16% | 2843 | XP_002815830.3 |
| adenomatous polyposis coli protein isoform X6 [Symphalangus syndactylus] | Symphalangus syndactylus | 5788 | 5788 | 100% | 0.0 | 98.95% | 2844 | XP_063468973.1 |

*Fig 3. 1. 2. a; Showing the results of blastp*

b. Multiple sequence analysis

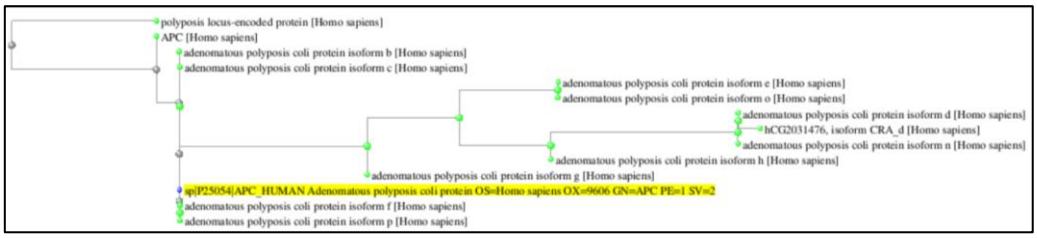
| | | | | |
|---|---|---|---|---|
| Query_3713795 | 1 | MAA | ASYDQLLKQVEALKMENSNLRQELEDNSNHLTKLETEASNMKEVLKQLQGSIEDEAMASSGQIDLLERLKELN | 76 |
| NP_000029.2 | 1 | MAA | ASYDQLLKQVEALKMENSNLRQELEDNSNHLTKLETEASNMKEVLKQLQGSIEDEAMASSGQIDLLERLKELN | 76 |
| AAA03586.1 | 1 | MAA | ASYDQLLKQVEALKMENSNLRQELEDNSNHLTKLETEASNMKEVLKQLQGSIEDEAMASSGQIDLLERLKELN | 76 |
| NP_001341825.1 | 1 | MAA | ASYDQLLKQVEALKMENSNLRQELEDNSNHLTKLETEASNMKEVLKQLQGSIEDEAMASSGQIDLLERLKELN | 76 |
| AAA60353.1 | 1 | MAA | ASYDQLLKQVEALKMENSNLRQELEDNSNHLTKLETEASNMKEVLKQLQGSIEDEAMASSGQIDLLERLKELN | 76 |
| NP_001341828.1 | 1 | MAA | ASYDQLLKQVEALKMENSNLRQELEDNSNHLTKLETEASNMKEVLKQLQGSIEDEAMASSGQIDLLERLKELN | 76 |
| NP_001394381.1 | 1 | MAA | ASYDQLLKQVEALKMENSNLRQELEDNSNHLTKLETEASNMKEVLKQLQGSIEDEAMASSGQIDLLERLKELN | 76 |
| NP_001341827.1 | 1 | MAD | VNLVERATGDERKRRR----------QEVLKQLQGSIEDEAMASSGQIDLLERLKELN | 51 |
| NP_001341826.1 | 1 | MYA[11] | ASVPPSVLGSWSTGGSRSCVRQETKSPGGARTSGHW-ASVWQEVLKQLQGSIEDEAMASSGQIDLLERLKELN | 86 |
| EAW49005.1 | 1 | MYA[11] | ASVPPSVLGSWSTGGSRSCVRQETKSPGGARTSGHW-ASVWQEVLKQLQGSIEDEAMASSGQIDLLERLKELN | 86 |
| NP_001394380.1 | 1 | MAD | VNLVERATGDERKRRR----------QEVLKQLQGSIEDEAMASSGQIDLLERLKELN | 51 |
| NP_001394375.1 | 1 | MYA[11] | ASVPPSVLGSWSTGGSRSCVRQETKSPGGARTSGHW-ASVWQEVLKQLQGSIEDEAMASSGQIDLLERLKELN | 86 |
| NP_001341830.1 | 1 | --- | ----------MASSGQIDLLERLKELN | 17 |
| NP_001341829.1 | 1 | --- | ----------MASSGQIDLLERLKELN | 17 |

| | | | |
|---|---|---|---|
| Query_3713795 | 77 | LDSSNFPGVKLRSKMSLRSYGSREGSVSSRSGECSPVPMGSFPRRGFVNGSRESTGYLEELEKERSLLLADLDKEEKEKD | 156 |
| NP_000029.2 | 77 | LDSSNFPGVKLRSKMSLRSYGSREGSVSSRSGECSPVPMGSFPRRGFVNGSRESTGYLEELEKERSLLLADLDKEEKEKD | 156 |
| AAA03586.1 | 77 | LDSSNFPGVKLRSKMSLRSYGSREGSVSSRSGECSPVPMGSFPRRGFVNGSRESTGYLEELEKERSLLLADLDKEEKEKD | 156 |
| NP_001341825.1 | 77 | LDSSNFPGVKLRSKMSLRSYGSREGSVSSRSGECSPVPMGSFPRRGFVNGSRESTGYLEELEKERSLLLADLDKEEKEKD | 156 |
| AAA60353.1 | 77 | LDSSNFPGVKLRSKMSLRSYGSREGSVSSRSGECSPVPMGSFPRRGFVNGSRESTGYLEELEKERSLLLADLDKEEKEKD | 156 |

*Fig.3. 1.2.b; MSA for protein sequence*

c. Phylogram

Fast minimum evolution




Neighbour-joining method

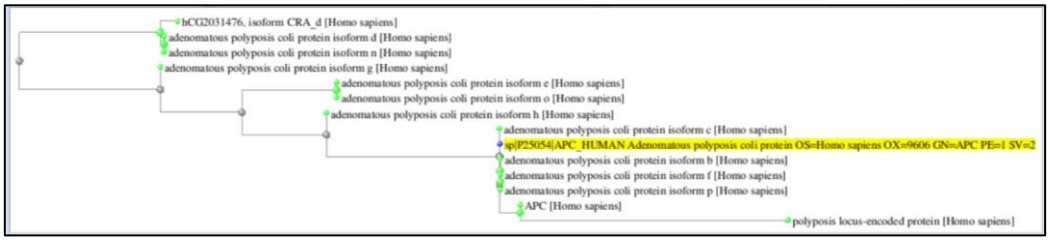


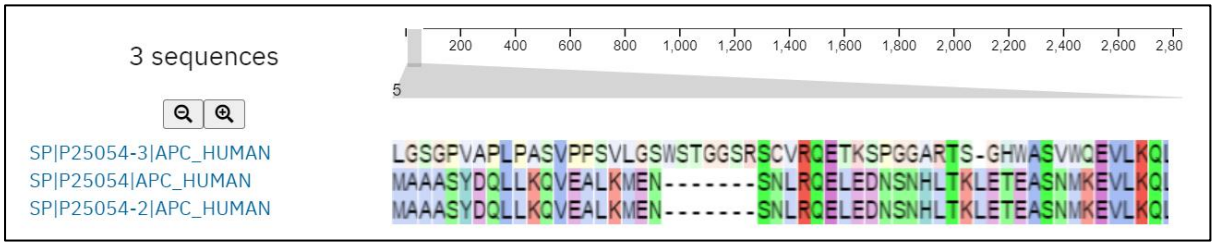

*Fig 3.1.2.c; Showing two different phylograms.*

### 1.3. Multiple sequence alignment for CANONICAL and ISOFROMS for the protein sequence:

i.

```
sp|P25054-3|APC_HUMAN      MYASLGSGPVAPLPASVPPSVLGSWSTGGSRSCVRQETKSPGGARTS-GHWASVWQEVLK
sp|P25054|APC_HUMAN        ----MAAASYDQLLKQVEALKMEN-------SNLRQELEDNSNHLTKLETEASNMKEVLK
sp|P25054-2|APC_HUMAN      ----MAAASYDQLLKQVEALKMEN-------SNLRQELEDNSNHLTKLETEASNMKEVLK
                               :.:..   *  .* .  : .       * :*** :. ..  *.    **  :****

sp|P25054-3|APC_HUMAN      QLQGSIEDEAMASSGQIDLLERLKELNLDSSNFPGVKLRSKMSLRSYGSREGSVSSRSGE
sp|P25054|APC_HUMAN        QLQGSIEDEAMASSGQIDLLERLKELNLDSSNFPGVKLRSKMSLRSYGSREGSVSSRSGE
sp|P25054-2|APC_HUMAN      QLQGSIEDEAMASSGQIDLLERLKELNLDSSNFPGVKLRSKMSLRSYGSREGSVSSRSGE
                           ************************************************************

sp|P25054-3|APC_HUMAN      CSPVPMGSFPRRGFVNGSRESTGYLEELEKERSLLLADLDKEEKEKDWYYAQLQNLTKRI
sp|P25054|APC_HUMAN        CSPVPMGSFPRRGFVNGSRESTGYLEELEKERSLLLADLDKEEKEKDWYYAQLQNLTKRI
sp|P25054-2|APC_HUMAN      CSPVPMGSFPRRGFVNGSRESTGYLEELEKERSLLLADLDKEEKEKDWYYAQLQNLTKRI
                           ************************************************************
```

ii.




iii.

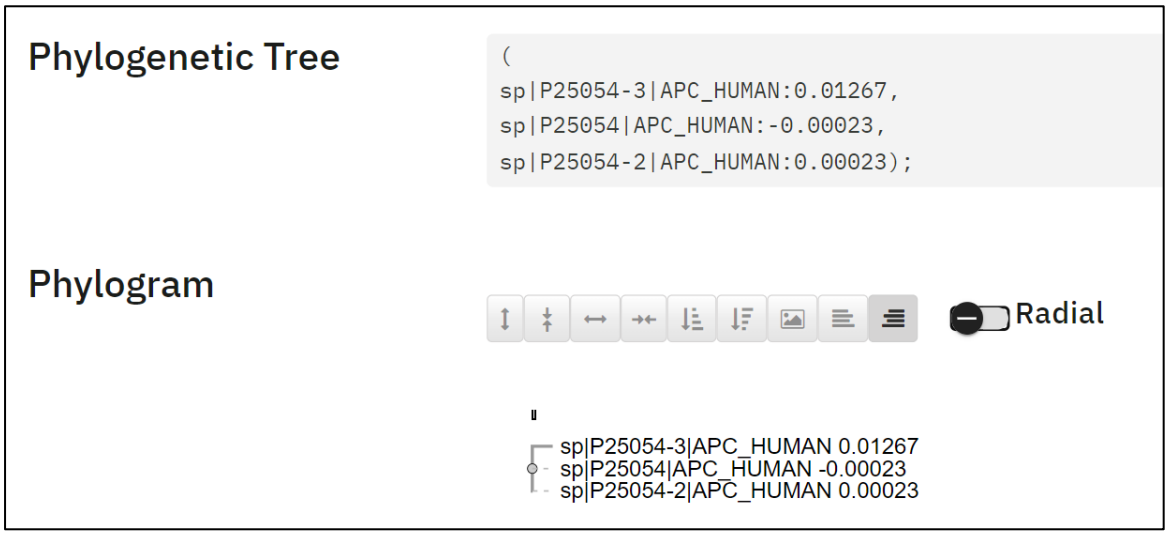
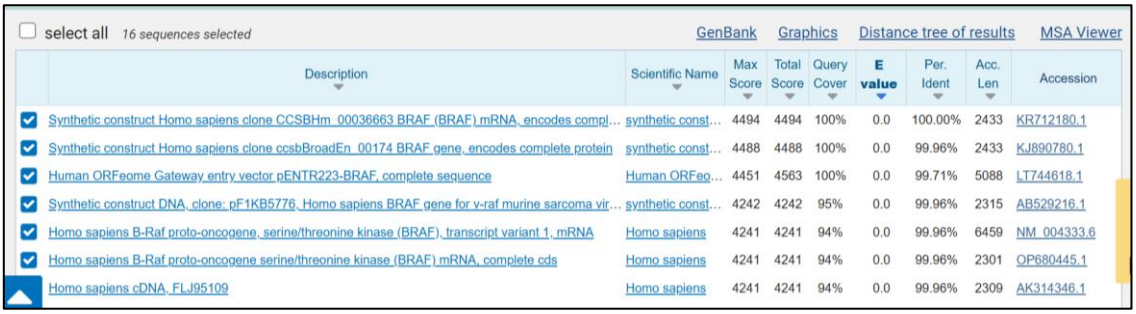



*Fig 3. 1. 3; i, ii, and iii display the MSA, one in colour and one without colour, along with the phylogenetic tree.*

## 2. BRAF

2.1. Multiple Sequence Alignment for Nucleotide Sequences

a. By performing blastn, we selected only the homo sapiens and got 16 sequences.

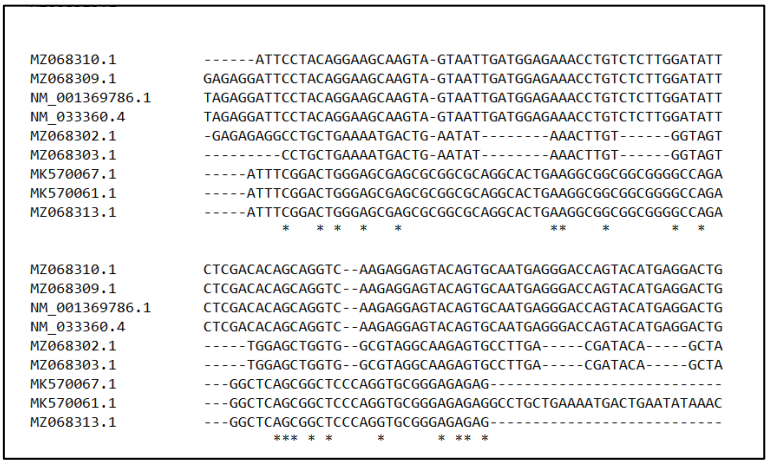


select all 16 sequences selected — GenBank | Graphics | Distance tree of results | MSA Viewer

| Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|
| Synthetic construct Homo sapiens clone CCSBHm_00036663 BRAF (BRAF) mRNA, encodes compl... | synthetic const... | 4494 | 4494 | 100% | 0.0 | 100.00% | 2433 | KR712180.1 |
| Synthetic construct Homo sapiens clone ccsbBroadEn_00174 BRAF gene, encodes complete protein | synthetic const... | 4488 | 4488 | 100% | 0.0 | 99.96% | 2433 | KJ890780.1 |
| Human ORFeome Gateway entry vector pENTR223-BRAF, complete sequence | Human ORFeo... | 4451 | 4563 | 100% | 0.0 | 99.71% | 5088 | LT744618.1 |
| Synthetic construct DNA, clone: pF1KB5776, Homo sapiens BRAF gene for v-raf murine sarcoma vir... | synthetic const... | 4242 | 4242 | 95% | 0.0 | 99.96% | 2315 | AB529216.1 |
| Homo sapiens B-Raf proto-oncogene, serine/threonine kinase (BRAF), transcript variant 1, mRNA | Homo sapiens | 4241 | 4241 | 94% | 0.0 | 99.96% | 6459 | NM_004333.6 |
| Homo sapiens B-Raf proto-oncogene serine/threonine kinase (BRAF) mRNA, complete cds | Homo sapiens | 4241 | 4241 | 94% | 0.0 | 99.96% | 2301 | OP680445.1 |
| Homo sapiens cDNA, FLJ95109 | Homo sapiens | 4241 | 4241 | 94% | 0.0 | 99.96% | 2309 | AK314346.1 |

*Fig 4. 2. 1.a; Showing the results of blastn*

b. Multiple Sequence Alignment

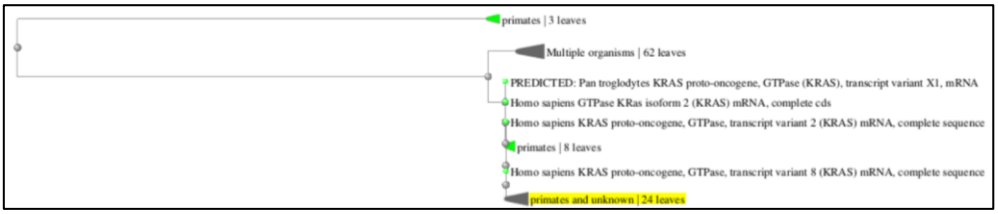


```
MZ068310.1          ------ATTCCTACAGGAAGCAAGTA-GTAATTGATGGAGAAACCTGTCTCTTGGATATT
MZ068309.1          GAGAGGATTCCTACAGGAAGCAAGTA-GTAATTGATGGAGAAACCTGTCTCTTGGATATT
NM_001369786.1      TAGAGGATTCCTACAGGAAGCAAGTA-GTAATTGATGGAGAAACCTGTCTCTTGGATATT
NM_033360.4         TAGAGGATTCCTACAGGAAGCAAGTA-GTAATTGATGGAGAAACCTGTCTCTTGGATATT
MZ068302.1          -GAGAGAGGCCTGCTGAAAATGACTG-AATAT--------AAACTTGT------GGTAGT
MZ068303.1          ---------CCTGCTGAAAATGACTG-AATAT--------AAACTTGT------GGTAGT
MK570067.1          -----ATTTCGGACTGGGAGCGAGCGCGGCGCAGGCACTGAAGGCGGCGGCGGGGCCAGA
MK570061.1          -----ATTTCGGACTGGGAGCGAGCGCGGCGCAGGCACTGAAGGCGGCGGCGGGGCCAGA
MZ068313.1          -----ATTTCGGACTGGGAGCGAGCGCGGCGCAGGCACTGAAGGCGGCGGCGGGGCCAGA
                             *   * *  *   *                **    *       *  *

MZ068310.1          CTCGACACAGCAGGTC--AAGAGGAGTACAGTGCAATGAGGGACCAGTACATGAGGACTG
MZ068309.1          CTCGACACAGCAGGTC--AAGAGGAGTACAGTGCAATGAGGGACCAGTACATGAGGACTG
NM_001369786.1      CTCGACACAGCAGGTC--AAGAGGAGTACAGTGCAATGAGGGACCAGTACATGAGGACTG
NM_033360.4         CTCGACACAGCAGGTC--AAGAGGAGTACAGTGCAATGAGGGACCAGTACATGAGGACTG
MZ068302.1          -----TGGAGCTGGTG--GCGTAGGCAAGAGTGCCTTGA-----CGATACA-----GCTA
MZ068303.1          -----TGGAGCTGGTG--GCGTAGGCAAGAGTGCCTTGA-----CGATACA-----GCTA
MK570067.1          ---GGCTCAGCGGCTCCCAGGTGCGGGAGAGAG----------------------------
MK570061.1          ---GGCTCAGCGGCTCCCAGGTGCGGGAGAGAGGCCTGCTGAAAATGACTGAATATAAAC
MZ068313.1          ---GGCTCAGCGGCTCCCAGGTGCGGGAGAGAG----------------------------
                           *** * *     *      * ** *
```

*Fig 4. 2. 1.b; MSA for nucleotide sequences.*

c. Phylogram

Fast minimum evolution




Neighbour -joining method

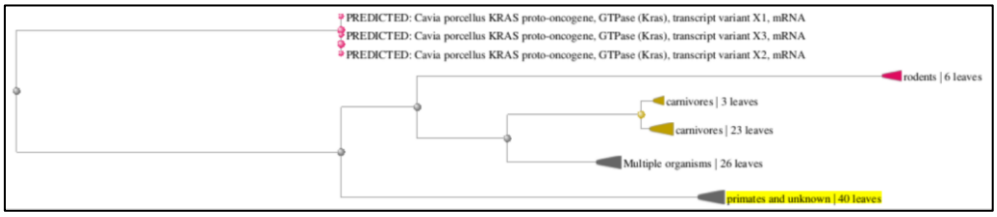



*Fig 4. 1. c; Showing two different phylograms.*

2.2. Multiple Sequence Alignment for Nucleotide Sequences

a. By doing Blastp, we got 3 sequences for Homo sapiens.

select all *3 sequences selected* GenPept Graphics Distance tree of results Multiple alignment MSA Viewe

| Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|
| GTPase KRas isoform a [Homo sapiens] | Homo sapiens | 390 | 390 | 100% | 7e-137 | 100.00% | 189 | NP_001356715.1 |
| GTPase KRas isoform X3 [Cavia porcellus] | Cavia porcellus | 392 | 392 | 100% | 8e-137 | 100.00% | 227 | XP_003470499.2 |
| GTPase KRas isoform X1 [Gorilla gorilla gorilla] | Gorilla gorilla gorilla | 391 | 391 | 100% | 9e-137 | 100.00% | 207 | XP_063550180.1 |
| Chain A, GTPase KRas [Homo sapiens] | Homo sapiens | 390 | 390 | 100% | 9e-137 | 100.00% | 190 | 7VVB_A |
| GTPase KRas isoform X3 [Chlorocebus sabaeus] | Chlorocebus sabaeus | 392 | 392 | 100% | 1e-136 | 100.00% | 227 | XP_007966155.1 |
| PREDICTED: GTPase KRas isoform X1 [Bison bison bison] | Bison bison bison | 390 | 390 | 100% | 1e-136 | 99.47% | 189 | XP_010854295.1 |
| GTPase KRas isoform X3 [Desmodus rotundus] | Desmodus rotundus | 391 | 391 | 100% | 1e-136 | 99.47% | 227 | XP_053777690.1 |
| GTPase KRas isoform X3 [Neophocaena asiaeorientalis asiaeorientalis] | Neophocaena asiaeorientalis asiaeorientalis | 391 | 391 | 100% | 1e-136 | 99.47% | 227 | XP_024588800.1 |
| KRAS isoform 4 [Pongo abelii] | Pongo abelii | 390 | 390 | 100% | 2e-136 | 99.47% | 189 | PNJ03835.1 |

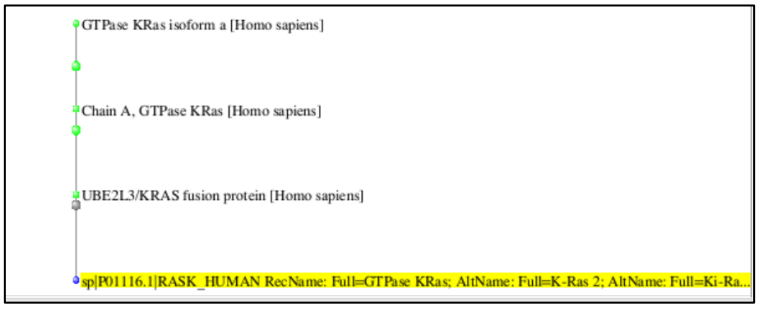
*Fig 4. 2. 2. a; Showing the results of blastp*

b. Multiple sequence analysis


*Fig 4. 2. 2. b; MSA for protein sequence*

c. Phylogram

Fast minimum evolution




Neighbour joining method

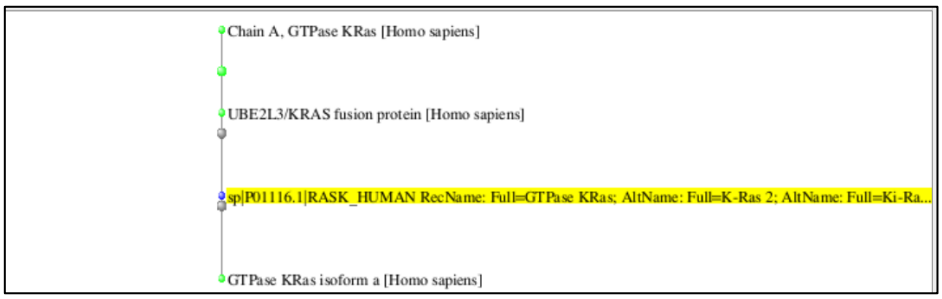
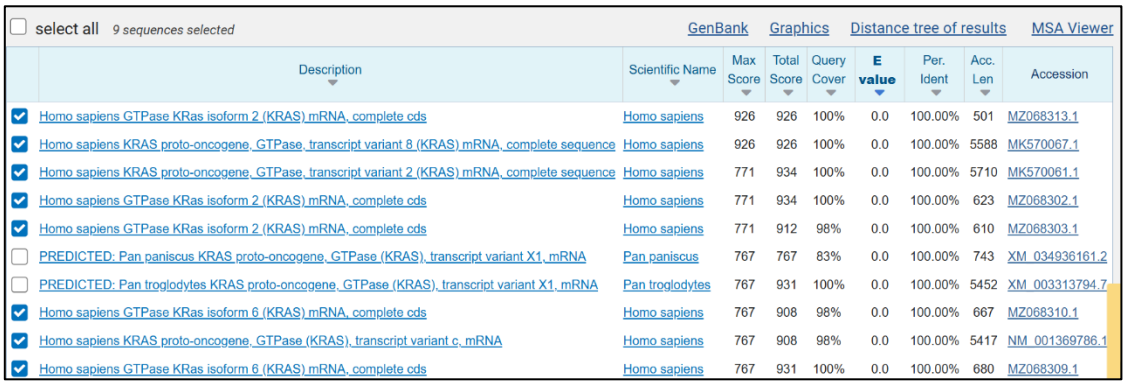



*Fig 4.2.c; Showing two different phylograms.*

## 3. KRAS

### 3.1.Multiple Sequence Alignment for Nucleotide Sequences

a. By performing blastn, we selected only the homo sapiens and got 9 sequences.



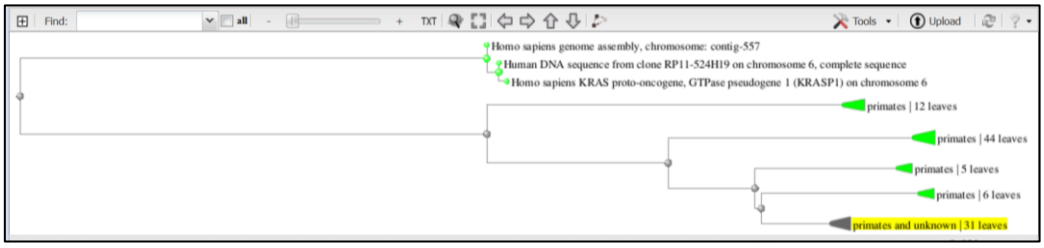
select all 9 sequences selected GenBank Graphics Distance tree of results MSA Viewer

| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | Homo sapiens GTPase KRas isoform 2 (KRAS) mRNA, complete cds | Homo sapiens | 926 | 926 | 100% | 0.0 | 100.00% | 501 | MZ068313.1 |
| ☑ | Homo sapiens KRAS proto-oncogene, GTPase, transcript variant 8 (KRAS) mRNA, complete sequence | Homo sapiens | 926 | 926 | 100% | 0.0 | 100.00% | 5588 | MK570067.1 |
| ☑ | Homo sapiens KRAS proto-oncogene, GTPase, transcript variant 2 (KRAS) mRNA, complete sequence | Homo sapiens | 771 | 934 | 100% | 0.0 | 100.00% | 5710 | MK570061.1 |
| ☑ | Homo sapiens GTPase KRas isoform 2 (KRAS) mRNA, complete cds | Homo sapiens | 771 | 934 | 100% | 0.0 | 100.00% | 623 | MZ068302.1 |
| ☑ | Homo sapiens GTPase KRas isoform 2 (KRAS) mRNA, complete cds | Homo sapiens | 771 | 912 | 98% | 0.0 | 100.00% | 610 | MZ068303.1 |
| ☐ | PREDICTED: Pan paniscus KRAS proto-oncogene, GTPase (KRAS), transcript variant X1, mRNA | Pan paniscus | 767 | 767 | 83% | 0.0 | 100.00% | 743 | XM_034936161.2 |
| ☐ | PREDICTED: Pan troglodytes KRAS proto-oncogene, GTPase (KRAS), transcript variant X1, mRNA | Pan troglodytes | 767 | 931 | 100% | 0.0 | 100.00% | 5452 | XM_003313794.7 |
| ☑ | Homo sapiens GTPase KRas isoform 6 (KRAS) mRNA, complete cds | Homo sapiens | 767 | 908 | 98% | 0.0 | 100.00% | 667 | MZ068310.1 |
| ☑ | Homo sapiens KRAS proto-oncogene, GTPase (KRAS), transcript variant c, mRNA | Homo sapiens | 767 | 908 | 98% | 0.0 | 100.00% | 5417 | NM_001369786.1 |
| ☑ | Homo sapiens GTPase KRas isoform 6 (KRAS) mRNA, complete cds | Homo sapiens | 767 | 931 | 100% | 0.0 | 100.00% | 680 | MZ068309.1 |

*Fig 5.3.1.a; Showing the results of blastn*

b. Phylogram

Fast minimum evolution




Neighbour joining method

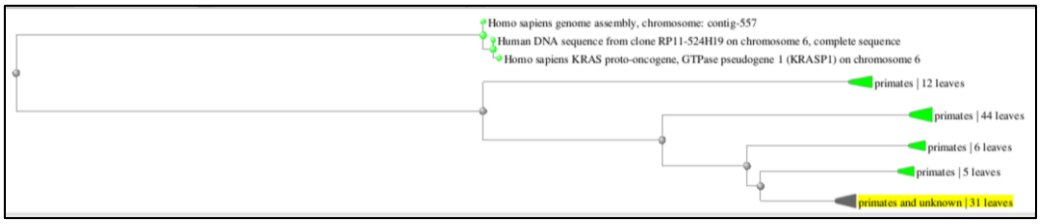
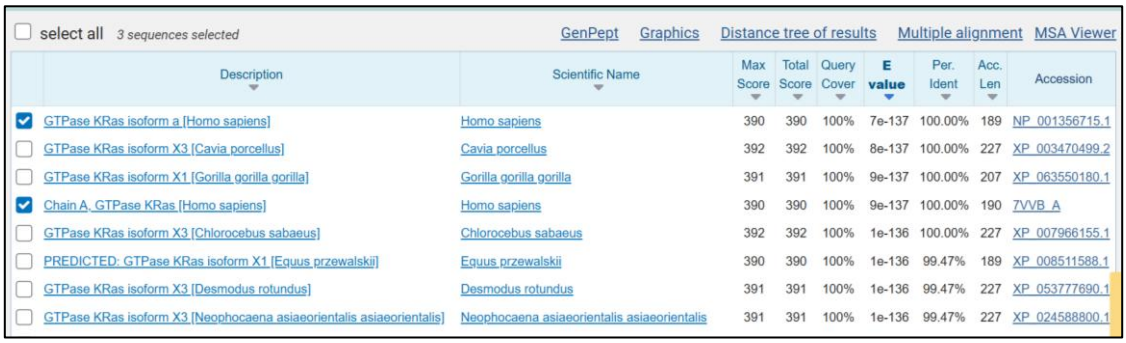



*Fig 5. 3. 1.b; Showing two different phylograms.*

### 3.2.Multiple Sequence Alignment for Protein Sequence

a. By doing Blastp, we got 3 sequences for Homo sapiens.



select all 3 sequences selected GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

| Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|
| GTPase KRas isoform a [Homo sapiens] | Homo sapiens | 390 | 390 | 100% | 7e-137 | 100.00% | 189 | NP_001356715.1 |
| GTPase KRas isoform X3 [Cavia porcellus] | Cavia porcellus | 392 | 392 | 100% | 8e-137 | 100.00% | 227 | XP_003470499.2 |
| GTPase KRas isoform X1 [Gorilla gorilla gorilla] | Gorilla gorilla gorilla | 391 | 391 | 100% | 9e-137 | 100.00% | 207 | XP_063550180.1 |
| Chain A, GTPase KRas [Homo sapiens] | Homo sapiens | 390 | 390 | 100% | 9e-137 | 100.00% | 190 | 7VVB_A |
| GTPase KRas isoform X3 [Chlorocebus sabaeus] | Chlorocebus sabaeus | 392 | 392 | 100% | 1e-136 | 100.00% | 227 | XP_007966155.1 |
| PREDICTED: GTPase KRas isoform X1 [Equus przewalskii] | Equus przewalskii | 390 | 390 | 100% | 1e-136 | 99.47% | 189 | XP_008511588.1 |
| GTPase KRas isoform X3 [Desmodus rotundus] | Desmodus rotundus | 391 | 391 | 100% | 1e-136 | 99.47% | 227 | XP_053777690.1 |
| GTPase KRas isoform X3 [Neophocaena asiaeorientalis asiaeorientalis] | Neophocaena asiaeorientalis asiaeorientalis | 391 | 391 | 100% | 1e-136 | 99.47% | 227 | XP_024588800.1 |

*Fig 5. 3. 2. a; Showing the results of blastp*

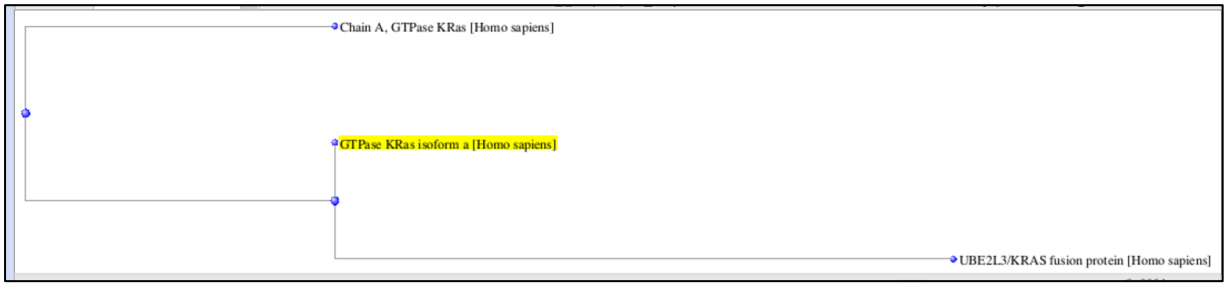
b. Multiple sequence analysis

```
NP_001356715.1   1     -MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEYSAMRDQYMRT  74
7VVB_A           1     GMTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEYSAMRDQYMRT  75
AEA35014.1       1     [106]KMTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEYSAMRDQYMRT  181

NP_001356715.1   75    GEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDVPMVLVGNKCDLPSRTVDTKQAQDLARSYGIPFIETSAKTRQRVED  154
7VVB_A           76    GEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDVPMVLVGNKCDLPSRTVDTKQAQDLARSYGIPFIETSAKTRQRVED  155
AEA35014.1       182   GEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDVPMVLVGNKCDLPSRTVDTKQAQDLARSYGIPFIETSAKTRQRVED  261

NP_001356715.1   155   AFYTLVREIRQYRLKKISKEEKTPGCVKIKKCIIM  189
7VVB_A           156   AFYTLVREIRQYRLKKISKEEKTPGCVKIKKCIIM  190
AEA35014.1       262   AFYTLVREIRQYRLKKISKEEKTPGCVKIKKCIIM  296
```

*Fig 5. 3.2.b; MSA for protein sequence*

c. Phylogram

Fast minimum evolution




Neighbour joining

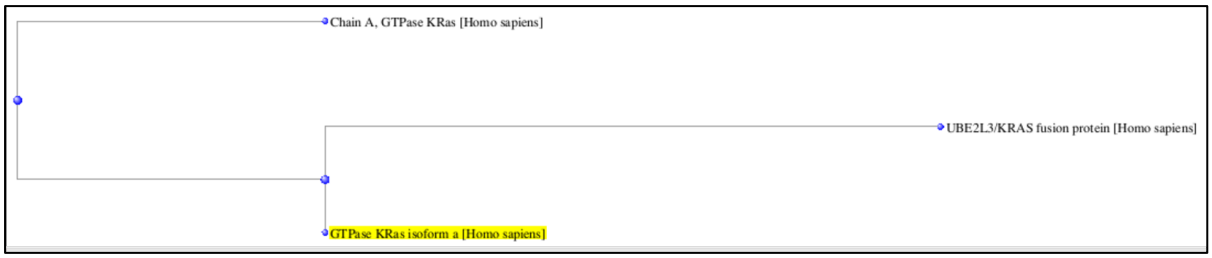



*Fig 5. 3. 2.c; Showing two different phylograms.*

## B. Single Nucleotide Polymorphisms (SNPs)

### a. Sequence-based analysis

#### 1. APC

*Table 3. 1. Sequence-based analysis results for APC*

| Sl.No. | SNP | PROTEIN ID | AMINO ACID CHANGE | SIFT PREDICTION | SIFT SCORE | PANTHER Message | Pdel | PolyPhen | Score range | PhD-SNP | Score | Meta | Score | SNP &GO | Probabilty |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | rs113486158 | ENSP00000257430 | A2648E | DELETERIOUS | 0.01 | probably damaging | 0.57 | Possibly damaging | 0.469 | Disease | 0.668 | Disease | 0.613 | Neutral | 0.252 |
| 2 | rs370955311 | ENSP00000257430 | A604G | DELETERIOUS | 0.012 | probably damaging | 0.85 | Possibly damaging | 0.89 | Disease | 0.571 | Disease | 0.636 | Neutral | 0.475 |
| 3 | rs200339830 | ENSP00000257430 | A766V | DELETERIOUS | 0.03 | probably damaging | 0.57 | Probably damaging | 0.975 | Neutral | 0.201 | Neutral | 0.305 | Neutral | 0.073 |
| 4 | rs138367627 | ENSP00000257430 | C1578G | DELETERIOUS | 0 | probably damaging | 0.85 | Probably damaging | 0.998 | Disease | 0.808 | Disease | 0.666 | Disease | 0.624 |
| 5 | rs148725540 | ENSP00000257430 | D1058G | DELETERIOUS | 0.026 | possibly damaging | 0.5 | Benign | 0.141 | Disease | 0.528 | Neutral | 0.476 | Neutral | 0.149 |
| 6 | rs368327191 | ENSP00000257430 | D1422G | DELETERIOUS | 0.005 | probably damaging | 0.57 | Probably damaging | 0.996 | Neutral | 0.38 | Neutral | 0.34 | Neutral | 0.245 |
| 7 | rs373419559 | ENSP00000257430 | D1570G | DELETERIOUS | 0.013 | probably damaging | 0.57 | Probably damaging | 1 | Disease | 0.637 | Disease | 0.61 | Neutral | 0.297 |
| 8 | rs148275069 | ENSP00000257430 | D1714N | DELETERIOUS | 0.013 | probably damaging | 0.57 | Benign | 0.102 | Neutral | 0.48 | Neutral | 0.387 | Neutral | 0.441 |
| 9 | rs76306073 | ENSP00000257430 | D1841Y | DELETERIOUS | 0.015 | probably damaging | 0.57 | Probably damaging | 1 | Disease | 0.899 | Disease | 0.704 | Disease | 0.787 |
| 10 | rs368724873 | ENSP00000257430 | D2696Y | DELETERIOUS | 0.03 | probably benign | 0.27 | Possibly damaging | 0.671 | Neutral | 0.05 | Neutral | 0.129 | Neutral | 0.032 |
| 11 | rs201093383 | ENSP00000257430 | D285G | DELETERIOUS | 0.027 | probably damaging | 0.57 | Possibly damaging | 0.761 | Neutral | 0.403 | Neutral | 0.335 | Neutral | 0.154 |
| 12 | rs111423620 | ENSP00000257430 | D556V | DELETERIOUS | 0 | probably damaging | 0.85 | Possibly damaging | 0.506 | Disease | 0.799 | Disease | 0.76 | Disease | 0.733 |
| 13 | rs79853077 | ENSP00000257430 | D802Y | DELETERIOUS | 0.021 | probably damaging | 0.57 | Probably damaging | 0.959 | Disease | 0.892 | Disease | 0.634 | Disease | 0.706 |
| 14 | rs202161017 | ENSP00000257430 | E140D | DELETERIOUS | 0.025 | probably damaging | 0.57 | Possibly damaging | 0.479 | Neutral | 0.448 | Neutral | 0.484 | Neutral | 0.379 |
| 15 | rs112961968 | ENSP00000257430 | E1464G | DELETERIOUS | 0.009 | possibly damaging | 0.5 | Benign | 0.002 | Neutral | 0.275 | Neutral | 0.194 | Neutral | 0.058 |
| 16 | rs138098808 | ENSP00000257430 | E536K | DELETERIOUS | 0.022 | probably damaging | 0.57 | Possibly damaging | 0.453 | Neutral | 0.213 | Neutral | 0.257 | Neutral | 0.144 |
| 17 | rs199740875 | ENSP00000257430 | E893K | DELETERIOUS | 0.011 | probably damaging | 0.57 | Possibly damaging | 0.837 | Neutral | 0.417 | Neutral | 0.268 | Neutral | 0.238 |
| 18 | rs142637152 | ENSP00000257430 | G116C | DELETERIOUS | 0.014 | probably damaging | 0.57 | Possibly damaging | 0.784 | Disease | 0.64 | Disease | 0.571 | Disease | 0.522 |
| 19 | rs369999291 | ENSP00000257430 | G116D | DELETERIOUS | 0.013 | probably damaging | 0.57 | Benign | 0.218 | Disease | 0.64 | Disease | 0.618 | Neutral | 0.463 |
| 20 | rs146695342 | ENSP00000257430 | G2187E | DELETERIOUS | 0.02 | probably damaging | 0.57 | Possibly damaging | 0.674 | Neutral | 0.412 | Neutral | 0.219 | Neutral | 0.335 |
| 21 | rs367905430 | ENSP00000257430 | G2227C | DELETERIOUS | 0.002 | probably damaging | 0.57 | Probably damaging | 1 | Disease | 0.791 | Disease | 0.653 | Disease | 0.712 |
| 22 | rs143145868 | ENSP00000257430 | G84V | DELETERIOUS | 0.013 | possibly damaging | 0.5 | Benign | 0.215 | Neutral | 0.302 | Neutral | 0.262 | Neutral | 0.106 |
| 23 | rs200803739 | ENSP00000257430 | I1524T | DELETERIOUS | 0.004 | probably damaging | 0.57 | Benign | 0.127 | Disease | 0.637 | Neutral | 0.456 | Neutral | 0.34 |
| 24 | rs377860 | ENSP00000257430 | K1370Q | DELETERIOUS | 0.045 | probably damaging | 0.57 | Probably damaging | 0.998 | Neutral | 0.248 | Neutral | 0.196 | Neutral | 0.095 |
| 25 | rs201988789 | ENSP00000257430 | K1915N | DELETERIOUS | 0.049 | probably damaging | 0.57 | Possibly damaging | 0.82 | Neutral | 0.299 | Neutral | 0.204 | Neutral | 0.154 |
| 26 | rs185411044 | ENSP00000257430 | L1618P | DELETERIOUS | 0.004 | probably damaging | 0.57 | Probably damaging | 0.963 | Disease | 0.684 | Neutral | 0.475 | Neutral | 0.44 |
| 27 | rs372418435 | ENSP00000257430 | L2039F | DELETERIOUS | 0.025 | probably damaging | 0.85 | Probably damaging | 0.998 | Neutral | 0.199 | Neutral | 0.198 | Neutral | 0.053 |
| 28 | rs372367350 | ENSP00000257430 | L23V | DELETERIOUS | 0.013 | probably damaging | 0.57 | Probably damaging | 0.996 | Neutral | 0.222 | Neutral | 0.38 | Neutral | 0.159 |
| 29 | rs77451514 | ENSP00000257430 | L505F | DELETERIOUS | 0.021 | probably damaging | 0.85 | Possibly damaging | 0.651 | Disease | 0.648 | Disease | 0.666 | Neutral | 0.237 |
| 30 | rs367761032 | ENSP00000257430 | P2222L | DELETERIOUS | 0.002 | probably damaging | 0.74 | Probably damaging | 0.996 | Neutral | 0.458 | Neutral | 0.33 | Neutral | 0.404 |
| 31 | rs377308875 | ENSP00000257430 | P2369S | DELETERIOUS | 0.019 | probably damaging | 0.74 | Probably damaging | 0.999 | Neutral | 0.05 | Neutral | 0.181 | Neutral | 0.099 |
| 32 | rs372305287 | ENSP00000257430 | P2467T | DELETERIOUS | 0.011 | possibly damaging | 0.5 | Benign | 0.371 | Neutral | 0.33 | 0.043 | 0.043 | Neutral | 0.012 |
| 33 | rs371149405 | ENSP00000257430 | P2522T | DELETERIOUS | 0.018 | possibly damaging | 0.5 | Benign | 0.023 | Disease | 0.531 | Disease | 0.575 | Neutral | 0.079 |
| 34 | rs182456139 | ENSP00000257430 | P2681L | DELETERIOUS | 0.028 | probably damaging | 0.57 | Probably damaging | 0.999 | Neutral | 0.487 | Neutral | 0.464 | Neutral | 0.076 |
| 35 | rs74535574 | ENSP00000257430 | Q1429K | DELETERIOUS | 0.009 | probably damaging | 0.57 | Probably damaging | 0.952 | Neutral | 0.121 | Neutral | 0.226 | Neutral | 0.06 |
| 36 | rs141576417 | ENSP00000257430 | Q203E | DELETERIOUS | 0.014 | probably damaging | 0.57 | Possibly damaging | 0.908 | Neutral | 0.328 | Neutral | 0.317 | Neutral | 0.173 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 37 | rs148878262 | ENSP00000257430 | Q2291H | DELETERIOUS | 0.014 | possibly damaging | 0.5 | Benign | 0.002 | Neutral | 0.138 | Neutral | 0.153 | Neutral | 0.019 |
| 38 | rs190988940 | ENSP00000257430 | Q473R | DELETERIOUS | 0.027 | probably damaging | 0.57 | Benign | 0.122 | Neutral | 0.395 | Neutral | 0.364 | Neutral | 0.124 |
| 39 | rs201830995 | ENSP00000257430 | R1171C | DELETERIOUS | 0.048 | probably benign | 0.02 | Benign | 0.004 | Neutral | 0.285 | Neutral | 0.167 | Neutral | 0.147 |
| 40 | rs72541813 | ENSP00000257430 | R1589G | DELETERIOUS | 0.01 | probably damaging | 0.57 | Benign | 0.33 | Neutral | 0.246 | Neutral | 0.205 | Neutral | 0.117 |
| 41 | rs374048423 | ENSP00000257430 | R1589H | DELETERIOUS | 0.004 | probably damaging | 0.57 | Benign | 0.006 | Neutral | 0.187 | Neutral | 0.229 | Neutral | 0.09 |
| 42 | rs34157245 | ENSP00000257430 | R1882T | DELETERIOUS | 0.022 | probably damaging | 0.57 | Benign | 0.086 | Neutral | 0.267 | Neutral | 0.154 | Neutral | 0.137 |
| 43 | rs199648202 | ENSP00000257430 | R2228G | DELETERIOUS | 0.005 | probably damaging | 0.57 | Probably damaging | 0.994 | Disease | 0.703 | Disease | 0.652 | Disease | 0.506 |
| 44 | rs79630786 | ENSP00000257430 | R2505G | DELETERIOUS | 0.036 | probably damaging | 0.57 | Benign | 0.121 | Neutral | 0.423 | Neutral | 0.217 | Neutral | 0.128 |
| 45 | rs199539353 | ENSP00000257430 | R2549C | DELETERIOUS | 0.004 | probably damaging | 0.57 | Probably damaging | 0.998 | Disease | 0.544 | Disease | 0.592 | Neutral | 0.157 |
| 46 | rs199558585 | ENSP00000257430 | R2549H | DELETERIOUS | 0.044 | probably damaging | 0.57 | Probably damaging | 0.998 | Neutral | 0.482 | Neutral | 0.397 | Neutral | 0.132 |
| 47 | rs377665107 | ENSP00000257430 | R332Q | DELETERIOUS | 0.012 | probably damaging | 0.78 | Possibly damaging | 0.894 | Neutral | 0.239 | Neutral | 0.335 | Neutral | 0.145 |
| 48 | rs137854567 | ENSP00000257430 | R414C | DELETERIOUS | 0.001 | probably damaging | 0.85 | Probably damaging | 1 | Disease | 0.779 | Disease | 0.736 | Disease | 0.667 |
| 49 | rs137854580 | ENSP00000257430 | R499G | DELETERIOUS | 0.002 | probably damaging | 0.85 | Probably damaging | 0.949 | Disease | 0.713 | Disease | 0.669 | Neutral | 0.377 |
| 50 | rs139196838 | ENSP00000257430 | R99W | DELETERIOUS | 0.028 | probably damaging | 0.57 | Probably damaging | 0.95 | Neutral | 0.328 | Neutral | 0.496 | Neutral | 0.289 |
| 51 | rs150973053 | ENSP00000257430 | S130G | DELETERIOUS | 0.02 | probably benign | 0.19 | Benign | 0.008 | Neutral | 0.335 | Neutral | 0.324 | Neutral | 0.133 |
| 52 | rs137854578 | ENSP00000257430 | S1395C | DELETERIOUS | 0 | probably damaging | 0.57 | Probably damaging | 0.996 | Neutral | 0.255 | Neutral | 0.258 | Neutral | 0.096 |
| 53 | rs267600319 | ENSP00000257430 | S1400L | DELETERIOUS | 0.002 | probably damaging | 0.57 | Probably damaging | 0.992 | Neutral | 0.168 | Neutral | 0.218 | Neutral | 0.079 |
| 54 | rs377169217 | ENSP00000257430 | S1581C | DELETERIOUS | 0.015 | probably damaging | 0.57 | Probably damaging | 0.971 | Neutral | 0.206 | Neutral | 0.235 | Neutral | 0.103 |
| 55 | rs4987109 | ENSP00000257430 | S1973T | DELETERIOUS | 0 | probably damaging | 0.57 | Probably damaging | 0.995 | Neutral | 0.135 | Neutral | 0.179 | Neutral | 0.035 |
| 56 | rs75207119 | ENSP00000257430 | S2350F | DELETERIOUS | 0.005 | probably damaging | 0.57 | Probably damaging | 0.998 | Neutral | 0.093 | Neutral | 0.164 | Neutral | 0.016 |
| 57 | rs75207119 | ENSP00000257430 | S2350Y | DELETERIOUS | 0.006 | probably damaging | 0.57 | Probably damaging | 0.998 | Neutral | 0.098 | Neutral | 0.164 | Neutral | 0.017 |
| 58 | rs150882838 | ENSP00000257430 | S2398F | DELETERIOUS | 0.018 | probably damaging | 0.57 | Probably damaging | 0.989 | Neutral | 0.286 | Neutral | 0.194 | Neutral | 0.052 |
| 59 | rs199806334 | ENSP00000257430 | S2586I | DELETERIOUS | 0.025 | probably damaging | 0.57 | Probably damaging | 0.991 | Neutral | 0.145 | Neutral | 0.179 | Neutral | 0.03 |
| 60 | rs72541816 | ENSP00000257430 | S2621C | DELETERIOUS | 0.031 | probably benign | 0.27 | Benign | 0.237 | Neutral | 0.235 | Neutral | 0.131 | Neutral | 0.048 |
| 61 | rs374853436 | ENSP00000257430 | S2793N | DELETERIOUS | 0.042 | probably damaging | 0.57 | Probably damaging | 0.992 | Neutral | 0.249 | Neutral | 0.178 | Neutral | 0.045 |
| 62 | rs75870842 | ENSP00000257430 | S535F | DELETERIOUS | 0.001 | probably damaging | 0.78 | Probably damaging | 0.999 | Neutral | 0.317 | Neutral | 0.486 | Neutral | 0.34 |
| 63 | rs373718658 | ENSP00000257430 | S5L | DELETERIOUS | 0.007 | probably damaging | 0.57 | Probably damaging | 0.941 | Neutral | 0.172 | Neutral | 0.222 | Neutral | 0.124 |
| 64 | rs74727182 | ENSP00000257430 | S62F | DELETERIOUS | 0.001 | probably damaging | 0.57 | Benign | 0.248 | Neutral | 0.317 | Neutral | 0.442 | Neutral | 0.231 |
| 65 | rs78349383 | ENSP00000257430 | S643P | DELETERIOUS | 0 | probably damaging | 0.85 | Probably damaging | 0.997 | Disease | 0.739 | Disease | 0.72 | Disease | 0.664 |
| 66 | rs146221748 | ENSP00000257430 | S95F | DELETERIOUS | 0.039 | probably damaging | 0.57 | Probably damaging | 0.929 | Neutral | 0.197 | Neutral | 0.188 | Neutral | 0.134 |
| 67 | rs374892194 | ENSP00000257430 | T1493M | DELETERIOUS | 0.018 | probably damaging | 0.57 | Probably damaging | 0.999 | Neutral | 0.33 | Neutral | 0.368 | Neutral | 0.247 |
| 68 | rs371686531 | ENSP00000257430 | T1847M | DELETERIOUS | 0.002 | probably damaging | 0.57 | Probably damaging | 0.999 | Disease | 0.678 | Disease | 0.652 | Neutral | 0.367 |
| 69 | rs200790804 | ENSP00000257430 | T2820I | DELETERIOUS | 0.033 | probably damaging | 0.57 | Benign | 0.239 | Neutral | 0.087 | Neutral | 0.162 | Neutral | 0.021 |
| 70 | rs371453363 | ENSP00000257430 | T518M | DELETERIOUS | 0.03 | probably damaging | 0.74 | Probably damaging | 0.992 | Neutral | 0.359 | Neutral | 0.364 | Neutral | 0.151 |
| 71 | rs202199891 | ENSP00000257430 | V530A | DELETERIOUS | 0 | probably damaging | 0.85 | Probably damaging | 0.999 | Neutral | 0.271 | Neutral | 0.283 | Neutral | 0.072 |
| 72 | rs147863331 | ENSP00000257430 | V609I | DELETERIOUS | 0 | probably damaging | 0.85 | Probably damaging | 0.911 | Disease | 0.547 | Disease | 0.513 | Neutral | 0.245 |
| 73 | rs186641437 | ENSP00000257430 | Y825C | DELETERIOUS | 0.011 | probably damaging | 0.57 | Probably damaging | 0.99 | Disease | 0.678 | Disease | 0.713 | Disease | 0.589 |

**Result:** D1841Y, D556V, D802Y, G2227C, R414C, S643P, Y825C, G116C and R2228 were found with the highest hits with different tools and further structural based analysis was done.

## 2. BRAF

*Table 3. 2. Sequence-based analysis results for BRAF*

| Sl.No. | | Protein ID | AA substitution | SIFT | PANTHER Message | Pdel | PolyPhen | Score range | PhD-SNP | Score | Meta SNP | Score | SNP&GO | Probabilty |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | **rs55939351** | **ENSP00000288602** | **I592T** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.782** | **Disease** | **0.705** | **Disease** | **0.758** |
| 2 | **rs55939351** | **ENSP00000288602** | **I592K** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **0.991** | **Disease** | **0.913** | **Disease** | **0.804** | **Disease** | **0.891** |
| 3 | rs55939351 | ENSP00000419060 | I199K | DELETERIOUS | NA | | NA | | Neutral | 0.419 | Neutral | 0.321 | NA | |
| 4 | **rs121913225** | **ENSP00000288602** | **F595S** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.913** | **Disease** | **0.834** | **Disease** | **0.818** |
| 5 | rs121913225 | ENSP00000419060 | F202S | DELETERIOUS | NA | | NA | | Neutral | 0.32 | Neutral | 0.419 | NA | |
| 6 | **rs121913337** | **ENSP00000288602** | **D594E** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **0.999** | **Disease** | **0.835** | **Disease** | **0.825** | **Disease** | **0.788** |
| 7 | **rs121913338** | **ENSP00000288602** | **D594V** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.935** | **Disease** | **0.869** | **Disease** | **0.878** |
| 8 | **rs121913338** | **ENSP00000288602** | **D594G** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **0.999** | **Disease** | **0.9** | **Disease** | **0.851** | **Disease** | **0.816** |
| 9 | rs121913340 | ENSP00000288602 | E586K | DELETERIOUS | Probably damaging | 0.74 | Benign | 0.204 | Neutral | 0.438 | Disease | 0.667 | Disease | 0.599 |
| 10 | **rs121913341** | **ENSP00000288602** | **F595L** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.82** | **Disease** | **0.657** | **Disease** | **0.71** |
| 11 | **rs121913348** | **ENSP00000288602** | **G464V** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **0.996** | **Disease** | **0.935** | **Disease** | **0.885** | **Disease** | **0.887** |
| 12 | **rs121913348** | **ENSP00000288602** | **G464A** | DELETERIOUS | **Probably damaging** | **0.86** | **Possibly damaging** | **0.508** | **Disease** | **0.617** | **Disease** | **0.664** | **Disease** | **0.72** |
| 13 | **rs121913348** | **ENSP00000288602** | **G464E** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **0.998** | **Disease** | **0.924** | **Disease** | **0.872** | **Disease** | **0.895** |
| 14 | **rs121913349** | **ENSP00000288602** | **G464R** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **0.985** | **Disease** | **0.959** | **Disease** | **0.871** | **Disease** | **0.865** |
| 15 | **rs121913353** | **ENSP00000288602** | **G466R** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **0.999** | **Disease** | **0.925** | **Disease** | **0.924** | **Disease** | **0.848** |
| 16 | **rs121913357** | **ENSP00000288602** | **G469R** | DELETERIOUS | **Probably damaging** | **0.86** | **Possibly damaging** | **0.815** | **Disease** | **0.913** | **Disease** | **0.853** | **Disease** | **0.86** |
| 17 | **rs121913357** | **ENSP00000288602** | **G469R** | DELETERIOUS | **Probably damaging** | **0.86** | **Possibly damaging** | **0.815** | **Disease** | **0.959** | **Disease** | **0.853** | **Disease** | **0.86** |
| 18 | **rs121913361** | **ENSP00000288602** | **G596R** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.881** | **Disease** | **0.857** | **Disease** | **0.794** |
| 19 | rs121913361 | ENSP00000419060 | G203R | DELETERIOUS | Probably damaging | 0.74 | NA | | Disease | 0.768 | Disease | 0.678 | NA | |
| 20 | **rs121913362** | **ENSP00000288602** | **I592M** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.742** | **Disease** | **0.691** | **Disease** | **0.662** |
| 21 | rs121913364 | ENSP00000288602 | K601E | DELETERIOUS | Probably damaging | 0.86 | Possibly damaging | 0.862 | Neutral | 0.369 | Disease | 0.654 | Disease | 0.759 |
| 22 | **rs121913365** | **ENSP00000288602** | **K601N** | DELETERIOUS | **Probably damaging** | **0.86** | **Possibly damaging** | **0.954** | **Disease** | **0.591** | **Disease** | **0.75** | **Disease** | **0.756** |
| 23 | **rs121913366** | **ENSP00000288602** | **L597R** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.907** | **Disease** | **0.85** | **Disease** | **0.849** |
| 24 | **rs121913366** | **ENSP00000288602** | **L597Q** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.848** | **Disease** | **0.784** | **Disease** | **0.76** |
| 25 | rs121913369 | ENSP00000288602 | L597V | DELETERIOUS | Probably damaging | 0.86 | Possibly damaging | 0.894 | Neutral | 0.451 | Disease | 0.574 | Neutral | 0.384 |
| 26 | **rs121913370** | **ENSP00000288602** | **N581S** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.857** | **Disease** | **0.824** | **Disease** | **0.778** |
| 27 | rs121913371 | ENSP00000288602 | R444W | DELETERIOUS | Probably damaging | 0.85 | Probably damaging | 0.999 | Disease | 0.67 | Disease | 0.735 | Neutral | 0.294 |
| 28 | **rs121913375** | **ENSP00000288602** | **T599I** | DELETERIOUS | **Probably damaging** | **0.86** | **Possibly damaging** | **0.652** | **Disease** | **0.606** | **Disease** | **0.693** | **Disease** | **0.742** |
| 29 | **rs121913376** | **ENSP00000288602** | **V471F** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **0.999** | **Disease** | **0.903** | **Disease** | **0.869** | **Disease** | **0.841** |
| 30 | rs121913376 | ENSP00000288602 | V471I | DELETERIOUS | Probably damaging | 0.86 | Benign | 0.341 | Disease | 0.561 | Disease | 0.633 | Neutral | 0.473 |
| 31 | rs121913378 | ENSP00000288602 | V600L | DELETERIOUS | Probably damaging | 0.85 | Benign | 0.003 | Neutral | 0.248 | Neutral | 0.452 | Neutral | 0.417 |
| 32 | rs121913378 | ENSP00000288602 | V600M | DELETERIOUS | Probably damaging | 0.85 | Benign | 0.398 | Neutral | 0.346 | Disease | 0.625 | Neutral | 0.442 |
| 33 | rs139420557 | ENSP00000288602 | P764L | DELETERIOUS | Probably benign | 0.19 | Benign | 0.002 | Neutral | 0.484 | Neutral | 0.414 | Neutral | 0.167 |
| 34 | rs150050723 | ENSP00000288602 | Q93H | DELETERIOUS | Probably benign | 0.57 | Probably damaging | 0.998 | Neutral | 0.495 | Disease | 0.537 | Neutral | 0.192 |
| 35 | rs180177032 | ENSP00000288602 | R462I | DELETERIOUS | Probably damaging | 0.86 | Possibly damaging | 0.909 | Neutral | 0.424 | Disease | 0.681 | Disease | 0.708 |
| 36 | **rs180177033** | **ENSP00000288602** | **I463S** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.876** | **Disease** | **0.81** | **Disease** | **0.886** |
| 37 | **rs180177034** | **ENSP00000288602** | **A246P** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **0.998** | **Disease** | **0.874** | **Disease** | **0.797** | **Disease** | **0.815** |
| 38 | rs180177036 | ENSP00000288602 | L485F | DELETERIOUS | Probably damaging | 0.86 | Benign | 0.061 | Neutral | 0.5 | Neutral | 0.496 | Neutral | 0.499 |
| 39 | rs180177036 | ENSP00000419060 | L92F | DELETERIOUS | Probably damaging | 0.57 | NA | | Neutral | 0.337 | Neutral | 0.175 | NA | |
| 40 | **rs180177038** | **ENSP00000288602** | **E501K** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.899** | **Disease** | **0.866** | **Disease** | **0.831** |
| 41 | **rs180177039** | **ENSP00000288602** | **E501G** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.787** | **Disease** | **0.841** | **Disease** | **0.714** |
| 42 | **rs180177040** | **ENSP00000288602** | **N581D** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **0.998** | **Disease** | **0.844** | **Disease** | **0.827** | **Disease** | **0.771** |
| 43 | **rs180177040** | **ENSP00000288602** | **N581H** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.878** | **Disease** | **0.851** | **Disease** | **0.794** |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 44 | rs180177041 | ENSP00000288602 | G534R | DELETERIOUS | Probably damaging | 0.86 | Benign | 0.401 | Disease | 0.517 | Disease | 0.674 | Disease | 0.65 |
| 45 | **rs180177042** | **ENSP00000288602** | **D638E** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.832** | **Disease** | **0.778** | **Disease** | **0.757** |
| 46 | **rs368528867** | **ENSP00000288602** | **R719L** | DELETERIOUS | **Probably damaging** | **0.57** | **Possibly damaging** | **0.941** | **Disease** | **0.796** | **Disease** | **0.65** | **Disease** | **0.666** |
| 47 | **rs387906660** | **ENSP00000288602** | **T241M** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.77** | **Disease** | **0.806** | **Disease** | **0.585** |
| 48 | **rs387906660** | **ENSP00000288602** | **T241R** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.87** | **Disease** | **0.784** | **Disease** | **0.797** |
| 49 | **rs387906661** | **ENSP00000288602** | **T241P** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.789** | **Disease** | **0.711** | **Disease** | **0.747** |
| 50 | rs397507462 | ENSP00000288602 | S122Y | DELETERIOUS | Probably damaging | 0.57 | Possibly damaging | 0.845 | Neutral | 0.097 | Neutral | 0.138 | Neutral | 0.021 |
| 51 | **rs397507466** | **ENSP00000288602** | **L245F** | DELETERIOUS | **Probably damaging** | **0.86** | **Possibly damaging** | **0.817** | **Disease** | **0.806** | **Disease** | **0.767** | **Disease** | **0.741** |
| 52 | rs397507466 | ENSP00000288602 | L245F | DELETERIOUS | Probably damaging | 0.86 | Possibly damaging | 0.817 | NA | | NA | | Disease | 0.741 |
| 53 | **rs397507467** | **ENSP00000288602** | **F247S** | DELETERIOUS | **Probably damaging** | **0.57** | **Probably damaging** | **0.985** | **Disease** | **0.858** | **Disease** | **0.74** | **Disease** | **0.689** |
| 54 | **rs397507473** | **ENSP00000288602** | **F468S** | DELETERIOUS | **Probably damaging** | **0.74** | **Possibly damaging** | **0.907** | **Disease** | **0.828** | **Disease** | **0.683** | **Disease** | **0.792** |
| 55 | **rs397507474** | **ENSP00000288602** | **K483Q** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.815** | **Disease** | **0.766** | **Disease** | **0.773** |
| 56 | **rs397507475** | **ENSP00000288602** | **L485S** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **0.997** | **Disease** | **0.779** | **Disease** | **0.747** | **Disease** | **0.741** |
| 57 | rs397507475 | ENSP00000419060 | L92S | DELETERIOUS | Probably damaging | 0.57 | NA | | Disease | 0.662 | Disease | 0.598 | NA | |
| 58 | rs397507476 | ENSP00000288602 | K499N | DELETERIOUS | Probably damaging | 0.86 | Probably damaging | 0.983 | Neutral | 0.357 | Neutral | 0.5 | Neutral | 0.367 |
| 59 | rs397507476 | ENSP00000288602 | K499N | DELETERIOUS | Probably damaging | 0.86 | Probably damaging | 0.983 | Neutral | 0.357 | Neutral | 0.5 | Neutral | 0.367 |
| 60 | **rs397507477** | **ENSP00000288602** | **L505F** | DELETERIOUS | **Probably damaging** | **0.86** | **Possibly damaging** | **0.785** | **Disease** | **0.708** | **Disease** | **0.551** | **Disease** | **0.649** |
| 61 | **rs397507479** | **ENSP00000288602** | **C532Y** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **0.998** | **Disease** | **0.811** | **Disease** | **0.824** | **Disease** | **0.903** |
| 62 | **rs397507481** | **ENSP00000288602** | **H574Q** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.861** | **Disease** | **0.827** | **Disease** | **0.751** |
| 63 | **rs397507483** | **ENSP00000288602** | **G596V** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **1** | **Disease** | **0.886** | **Disease** | **0.852** | **Disease** | **0.886** |
| 64 | rs397507483 | ENSP00000419060 | G203V | DELETERIOUS | Probably damaging | 0.57 | NA | | Disease | 0.763 | Disease | 0.683 | NA | |
| 65 | **rs397507484** | **ENSP00000288602** | **K601I** | DELETERIOUS | **Probably damaging** | **0.86** | **Probably damaging** | **0.997** | **Disease** | **0.564** | **Disease** | **0.654** | **Disease** | **0.759** |
| 66 | rs397509343 | ENSP00000288602 | F247L | DELETERIOUS | Probably damaging | 0.57 | Possibly damaging | 0.81 | Disease | 0.815 | Disease | 0.727 | Neutral | 0.487 |
| 67 | **rs55715359** | **ENSP00000288602** | **L692S** | DELETERIOUS | **Probably damaging** | **0.85** | **Probably damaging** | **0.997** | **Disease** | **0.816** | **Disease** | **0.689** | **Disease** | **0.73** |
| 68 | rs267601317 | ENSP00000288602 | P367L | DELETERIOUS | Probably damaging | 0.86 | Probably damaging | 1 | Disease | 0.591 | Neutral | 0.474 | Neutral | 0.22 |
| 69 | rs372569965 | ENSP00000288602 | R603Q | DELETERIOUS | Probably damaging | 0.86 | Probably damaging | 0.981 | Neutral | 0.223 | Disease | 0.52 | Disease | 0.55 |
| 70 | rs375520366 | ENSP00000288602 | P490S | DELETERIOUS | Probably damaging | 0.86 | Possibly damaging | 0.601 | Neutral | 0.168 | Neutral | 0.426 | Neutral | 0.414 |
| 71 | rs397507485 | ENSP00000288602 | R671Q | DELETERIOUS | Probably damaging | 0.86 | Probably damaging | 0.97 | Neutral | 0.326 | Disease | 0.603 | Disease | 0.576 |
| 72 | rs397507486 | ENSP00000288602 | Q709R | DELETERIOUS | Probably damaging | 0.86 | Benign | 0.412 | Disease | 0.583 | Disease | 0.654 | Disease | 0.538 |
| 73 | **rs397507487** | **ENSP00000288602** | **R719C** | DELETERIOUS | **Probably damaging** | **0.57** | **Probably damaging** | **1** | **Disease** | **0.723** | **Disease** | **0.56** | **Disease** | **0.605** |
| 74 | **rs397507487** | **ENSP00000288602** | **V600E** | DELETERIOUS | **Probably damaging** | **0.86** | **Possibly damaging** | **0.923** | **Disease** | **0.627** | **Disease** | **0.558** | **Disease** | **0.799** |

**Result:** A246P, C532Y, D594E, D594G, D594V, E501G, E501K, F595L, F595S, G464A, G464E, G464V, H574Q, I592K, I592M, I463S, I592T, K601I, K601N, K483Q, L245F, L597Q, L597R, L692S, N581D, N581H, N581S, R719C, R719L, T599I, T241M, T241R, T241P, V600E and V471F were found with the highest hits with different tools and further structural based analysis was done.

## 3. KRAS

*Table 3. 3. Sequence-based analysis results for KRAS*

| Sl.No. | PROTEIN ID | SNP | AA CHANGE | SIFT | SCORE | PANTHER | Score | PolyPhen | Score | PhD-SNP | Score | Meta SNP | Score | SNP &GO | Probability |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ENSP00000256078 | rs373500216 | A134G | DELETERIOUS | 0.016 | NA | | Benign | 0.15 | Disease | 0.781 | Disease | 0.521 | Disease | 0.513 |
| 2 | ENSP00000256078 | rs121913527 | A146T | DELETERIOUS | 0 | NA | | Possibly damaging | 0.47 | Disease | 0.85 | Disease | 0.725 | Disease | 0.646 |
| 3 | ENSP00000256078 | rs121913528 | A59T | DELETERIOUS | 0.024 | Neutral | 0.447 | Benign | 0.429 | Disease | 0.748 | Disease | 0.618 | Disease | 0.525 |
| 4 | ENSP00000256078 | rs104886029 | A59V | DELETERIOUS | 0 | Neutral | 0.497 | Probably damaging | 0.92 | Disease | 0.801 | Disease | 0.744 | Disease | 0.555 |
| 5 | ENSP00000256078 | rs104894360 | D153V | DELETERIOUS | 0.001 | NA | | Benign | 0.313 | Disease | 0.61 | Neutral | 0.476 | NA | |
| 6 | ENSP00000256078 | rs104894360 | D40V | DELETERIOUS | 0.001 | NA | | Possibly damaging | 0.904 | Disease | 0.868 | Disease | 0.642 | NA | |
| 7 | ENSP00000256078 | rs138669124 | F141L | DELETERIOUS | 0.035 | NA | | Benign | 0.062 | Disease | 0.737 | Disease | 0.538 | Neutral | 0.359 |
| 8 | ENSP00000256078 | rs104894362 | F156L | DELETERIOUS | 0.001 | NA | | Possibly damaging | 0.678 | Disease | 0.832 | Disease | 0.658 | Neutral | 0.481 |
| 9 | ENSP00000256078 | rs104894362+25:31 | F43L | DELETERIOUS | 0.005 | NA | | Benign | 0.294 | Disease | 0.847 | Disease | 0.656 | NA | |
| 10 | ENSP00000256078 | rs104894359 | G60R | DELETERIOUS | 0 | Disease | 0.91 | Probably damaging | 1 | Disease | 0.846 | Disease | 0.85 | Disease | 0.752 |
| 11 | ENSP00000256078 | rs104894359 | G60S | DELETERIOUS | 0.001 | Disease | 0.805 | Probably damaging | 1 | Disease | 0.775 | Disease | 0.729 | Disease | 0.718 |
| 12 | ENSP00000256078 | rs387907206 | K147E | DELETERIOUS | 0.001 | NA | | Probably damaging | 0.986 | Disease | 0.829 | Disease | 0.714 | Disease | 0.629 |
| 13 | ENSP00000308495 | rs193929331 | K5E | DELETERIOUS | 0.001 | Neutral | 0.378 | Probably damaging | 0.991 | Disease | 0.689 | Disease | 0.691 | Disease | 0.658 |
| 14 | ENSP00000308495 | rs104894361 | K5N | DELETERIOUS | 0 | Neutral | 0.448 | Probably damaging | 0.998 | Disease | 0.523 | Neutral | 0.458 | Neutral | 0.376 |
| 15 | ENSP00000308495 | rs121913538 | L19F | DELETERIOUS | 0.001 | Neutral | 0.397 | Probably damaging | 0.996 | Disease | 0.854 | Disease | 0.599 | Disease | 0.573 |
| 16 | ENSP00000308495 | rs202247812 | N116S | DELETERIOUS | 0.001 | NA | | Probably damaging | 0.978 | Disease | 0.803 | Disease | 0.684 | Disease | 0.584 |
| 17 | ENSP00000308495 | rs104894366 | P34R | DELETERIOUS | 0 | Disease | 0.92 | Probably damaging | 1 | Disease | 0.835 | Disease | 0.848 | Disease | 0.755 |
| 18 | ENSP00000308495 | rs121913236 | Q22K | DELETERIOUS | 0.001 | Neutral | 0.427 | Probably damaging | 0.949 | Disease | 0.868 | Disease | 0.688 | Disease | 0.742 |
| 19 | ENSP00000308495 | rs121913238 | Q61E | DELETERIOUS | 0.003 | Neutral | 0.133 | Possibly damaging | 0.839 | Disease | 0.716 | Disease | 0.605 | Neutral | 0.252 |
| 20 | ENSP00000308495 | rs17851045 | Q61H | DELETERIOUS | 0.002 | Neutral | 0.277 | Benign | 0.151 | Disease | 0.756 | Disease | 0.719 | Neutral | 0.247 |
| 21 | ENSP00000451856 | rs121913238 | Q61K | DELETERIOUS | 0.013 | Neutral | 0.109 | Benign | 0.105 | Disease | 0.832 | Disease | 0.717 | Neutral | 0.274 |
| 22 | ENSP00000451856 | rs372793780 | R164Q | DELETERIOUS | 0.042 | NA | | Probably damaging | 0.926 | Disease | 0.545 | Neutral | 0.469 | Neutral | 0.393 |
| 23 | ENSP00000452512 | rs372793780 | R51Q | DELETERIOUS | 0.016 | NA | | Benign | 0.417 | Disease | 0.706 | Disease | 0.509 | NA | |
| 24 | ENSP00000452512 | rs104894364 | T58I | DELETERIOUS | 0 | Disease | 0.765 | Probably damaging | 0.996 | Disease | 0.794 | Disease | 0.771 | Disease | 0.664 |
| 25 | ENSP00000452512 | rs104894365 | V14I | DELETERIOUS | 0.006 | Neutral | 0.253 | Possibly damaging | 0.887 | Disease | 0.633 | Disease | 0.752 | Neutral | 0.343 |
| 26 | ENSP00000452512 | rs104894367 | V152G | DELETERIOUS | 0.004 | NA | | Probably damaging | 0.998 | Disease | 0.832 | Disease | 0.751 | Disease | 0.507 |

**Results:** G60S, K5E, N116S, D40V and F156L were found with the highest hits with different tools and further structural based analysis was done.

## b. Structure-based analysis

### 1. APC

*Table 4. 1. Structural-Based analysis results for APC*

| Sl. No. | AA | Structure number | Dynamut | Score | mCSM | Score | SDM | score | DUET | Score | CUPSAT | Torison | Predicted | SNP &GO 3D | Probablity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | D1841Y | D7Y | Stabilizing | 0.107 | Stabilizing | 0.121 | Destabilizing | -0.13 | Stabilizing | 0.146 | Destabilizing | Unfavourable | -0.71 | Disease | 0.653 |
| 2 | D556V | D7V | Stabilizing | 0.118 | Destabilizing | -0.032 | Destabilizing | -0.11 | Stabilizing | 0.257 | Destabilizing | Unfavourable | -0.74 | Disease | 0.644 |
| 3 | **G2227C** | G53C | **Destabilizing** | **-0.555** | **Destabilizing** | **-0.906** | **Stabilizing** | **0.12** | **Destabilizing** | **-0.585** | **Destabilizing** | **Unfavourable** | **-2** | **Disease** | **0.864** |
| 4 | D802Y | D7Y | Stabilizing | 0.118 | Destabilizing | -0.032 | Destabilizing | -0.11 | Stabilizing | 0.257 | Destabilizing | Unfavourable | -0.71 | Disease | 0.644 |
| 5 | **R414C** | R24C | **Stabilizing** | **0.077** | **Destabilizing** | **-0.455** | **Destabilizing** | **-0.83** | **Destabilizing** | **-0.533** | **Destabilizing** | **Unfavourable** | **-1.57** | **Disease** | **0.903** |
| 6 | **S643P** | S21P | **Stabilizing** | **0.297** | **Destabilizing** | **-0.15** | **Destabilizing** | **-0.96** | **Destabilizing** | **-0.101** | **Destabilizing** | **Unfavourable** | **-1.39** | **Disease** | **0.807** |
| 7 | Y825C | Y6C | Stabilizing | 0.009 | Destabilizing | -0.097 | Stabilizing | 0.06 | Stabilizing | 0.174 | Stabilizing | Unfavourable | 0.58 | Disease | 0.859 |
| 8 | **G116C** | G53C | **Destabilizing** | **-0.555** | **Destabilizing** | **-0.906** | **Stabilizing** | **0.12** | **Destabilizing** | **-0.585** | **Destabilizing** | **Unfavourable** | **-2** | **Disease** | **0.864** |
| 9 | **R2228G** | R24G | **Destabilizing** | **-0.516** | **Destabilizing** | **-0.032** | **Destabilizing** | **-0.11** | **Stabilizing** | **0.257** | **Destabilizing** | **Unfavourable** | **-0.81** | **Disease** | **0.861** |

**Results:** G2227C, R414C, R2228G, and S643P were selected with highest hit using different tools.

## 2. BRAF

*Table 4. 2. Structural-based analysis results for BRAF*

| Sl. No. | AA substitution | Structure number | Dynamut | Score | mCSM | Score | SDM | score | DUET | Score | CUPSAT | Torison | Predicted | SNP &GO 3D | Probablity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | A246P | A481P | Stabilizing | 0.541 | Destabilizing | -0.132 | Destabilizing | -1.85 | Destabilizing | -0.159 | Destabilizing | Favorable | -0.74 | Disease | 0.849 |
| 2 | C532Y | C532Y | stabilizing | 1.379 | Destabilizing | -1.566 | Destabilizing | -1.49 | Destabilizing | -1.734 | Stabilizing | Favorable | 4.96 | Disease | 0.785 |
| 3 | D594E | D449E | Destabilizing | -0.51 | Destabilizing | -0.295 | Stabilizing | 0 | Stabilizing | 0.033 | Stabilizing | Unfavorable | 1.93 | Neutral | 0.026 |
| 4 | D594G | D449G | Destabilizing | -0.383 | Destabilizing | -0.441 | Stabilizing | 0 | Stabilizing | 0.239 | Destabilizing | Favourable | -1.74 | Neutral | 0.197 |
| 5 | D594V | D449V | Destabilizing | -0.383 | Stabilizing | 0.029 | Stabilizing | 0 | Destabilizing | -0.23 | Destabilizing | Favourable | -2.73 | Neutral | 0.412 |
| 6 | E501G | E451G | Destabilizing | -0.554 | Destabilizing | -1.254 | Destabilizing | -0.22 | Destabilizing | -0.175 | Destabilizing | Unfavorable | -1.91 | Disease | 0.523 |
| 7 | E501K | E451K | Destabilizing | -1.062 | Destabilizing | -0.399 | Stabilizing | 0.48 | Destabilizing | -1.027 | Destabilizing | Favourable | -0.65 | Disease | 0.684 |
| 8 | F595L | F468L | Destabilizing | -0.447 | Destabilizing | -0.555 | Stabilizing | 0.33 | Destabilizing | -0.383 | Stabilizing | Unfavorable | 0.51 | Neutral | 0.469 |
| 9 | F595S | F468S | Destabilizing | -0.737 | Destabilizing | -1.061 | Destabilizing | -1.79 | Destabilizing | -1.162 | Stabilizing | Unfavorable | 0.71 | Disease | 0.635 |
| 10 | G464A | G455A | Destabilizing | -1.119 | Destabilizing | -0.299 | Destabilizing | -1.53 | Destabilizing | -0.56 | Stabilizing | Unfavorable | 0.13 | Neutral | 0.31 |
| 11 | G464E | G455E | Destabilizing | -1.345 | Destabilizing | -0.981 | Destabilizing | -2.58 | Destabilizing | -1.313 | Destabilizing | Unfavorable | -0.8 | Neutral | 0.379 |
| 12 | **G464V** | **G455V** | **Destabilizing** | **-0.989** | **Destabilizing** | **-0.403** | **Destabilizing** | **-2.06** | **Destabilizing** | **-0.576** | **Destabilizing** | **Unfavorable** | **-3.25** | **Disease** | **0.603** |
| 13 | H574Q | H477Q | Stabilizing | 0.07 | Destabilizing | -1.228 | Destabilizing | -0.56 | Destabilizing | -0.257 | Destabilizing | Unfavorable | -1.58 | Disease | 0.72 |
| 14 | I592K | I452K | Destabilizing | -2.175 | Destabilizing | -1.168 | Destabilizing | -3.23 | Destabilizing | -2.495 | Stabilizing | Unfavorable | 0.62 | Disease | 0.765 |
| 15 | I592M | I452M | Destabilizing | -1.269 | Destabilizing | -0.384 | Destabilizing | -2 | Destabilizing | -1.917 | Destabilizing | Unfavorable | -0.36 | Neutral | 0.32 |
| 16 | **I463S** | **I452S** | **Destabilizing** | **-0.488** | **Highly Destabilizing** | **-2.187** | **Destabilizing** | **-2.09** | **Destabilizing** | **-1.211** | **Destabilizing** | **Unfavorable** | **-4.28** | **Disease** | **0.614** |
| 17 | I592T | I452T | Stabilizing | 0.576 | Destabilizing | -1.814 | Destabilizing | -1.22 | Destabilizing | -0.513 | Destabilizing | Favourable | -2.82 | Disease | 0.651 |
| 18 | K601I | K473I | Stabilizing | 0.05 | Stabilizing | 0.229 | Stabilizing | 1.08 | Stabilizing | 0.754 | Destabilizing | Unfavorable | -0.32 | Disease | 0.764 |
| 19 | K601N | K473N | Destabilizing | -1.147 | Destabilizing | -1.259 | Destabilizing | -0.63 | Destabilizing | -1.119 | Destabilizing | Unfavorable | -3.57 | Disease | 0.594 |
| 20 | **K483Q** | **K473Q** | **Destabilizing** | **-0.279** | **Destabilizing** | **-1.049** | **Destabilizing** | **-0.48** | **Destabilizing** | **-0.855** | **Destabilizing** | **Favorable** | **-4.8** | **Disease** | **0.586** |
| 21 | L245F | L485F | Stabilizing | 0.491 | Destabilizing | -1.399 | Destabilizing | -0.4 | Destabilizing | -0.661 | Destabilizing | Favorable | -0.48 | Neutral | 0.465 |
| 22 | **L597Q** | **L485Q** | **Destabilizing** | **-0.31** | **Destabilizing** | **-1.433** | **Destabilizing** | **-1.33** | **Destabilizing** | **-1.402** | **Destabilizing** | **Unfavorable** | **-2.28** | **Disease** | **0.676** |
| 23 | L597R | L485R | Destabilizing | -0.986 | Destabilizing | -0.877 | Destabilizing | -1.73 | Destabilizing | -2 | Destabilizing | Unfavorable | -0.7 | Disease | 0.68 |
| 24 | L692S | L485S | Stabilizing | 0.444 | Destabilizing | -1.922 | Stabilizing | 0.12 | Destabilizing | -1.23 | Destabilizing | Unfavorable | -0.55 | Disease | 0.664 |
| 25 | N581D | N486D | Destabilizing | -0.47 | Stabilizing | 0.205 | Destabilizing | -0.73 | Destabilizing | -0.05 | Destabilizing | Favorable | -0.06 | Neutral | 0.168 |
| 26 | N581H | N486H | Destabilizing | -0.486 | Destabilizing | -0.516 | Stabilizing | 0.49 | Destabilizing | -0.304 | Destabilizing | Unfavorable | -0.54 | Neutral | 0.294 |
| 27 | N581S | N486S | Destabilizing | -0.295 | Destabilizing | -0.205 | Destabilizing | -0.06 | Stabilizing | 0.514 | Destabilizing | Unfavorable | -0.22 | Neutral | 0.223 |
| 28 | R719C | R462C | Destabilizing | -0.258 | Destabilizing | -0.698 | Destabilizing | -0.05 | Destabilizing | -0.557 | Stabilizing | Favorable | 3.43 | Disease | 0.862 |
| 29 | R719L | R462L | Destabilizing | -0.134 | Stabilizing | 0.203 | Stabilizing | 0.54 | Stabilizing | 0.478 | Destabilizing | Favorable | -0.88 | Disease | 0.817 |
| 30 | T599I | T458I | Destabilizing | -0.251 | Destabilizing | -0.069 | Destabilizing | -0.05 | Stabilizing | 0.305 | Destabilizing | Favourable | -0.99 | Neutral | 0.345 |
| 31 | T241M | T458M | Destabilizing | -0.075 | Destabilizing | -0.063 | Destabilizing | -1.02 | Destabilizing | -0.107 | Destabilizing | Unfavorable | -4.76 | Neutral | 0.29 |
| 32 | T241R | T458R | Destabilizing | -0.059 | Destabilizing | -0.087 | Destabilizing | -0.16 | Stabilizing | 0.032 | Destabilizing | Unfavorable | -2.48 | Neutral | 0.226 |
| 33 | T241P | T458P | Destabilizing | -0.112 | Destabilizing | -0.115 | Stabilizing | 0.64 | Stabilizing | 0.404 | Destabilizing | Unfavorable | -4.56 | Disease | 0.604 |
| 34 | **V600E** | **V459E** | **Stabilizing** | **0.592** | **Destabilizing** | **-0.231** | **Destabilizing** | **-1.05** | **Destabilizing** | **-0.918** | **Stabilizing** | **Unfavorable** | **1.1** | **Disease** | **0.58** |
| 35 | V471F | V459F | Destabilizing | -0.215 | Destabilizing | -0.784 | Destabilizing | -1.51 | Destabilizing | -0.057 | Destabilizing | Favorable | -1.5 | Neutral | 0.292 |

**Results:** G464V, I463S, K473Q and V600E were selected with the highest hit using different tools.

## 3. KRAS

*Table 4. 3. Structural-based analysis results for KRAS*

| Sl.No. | AA CHANGE | Structure number | Dynamut | Score | mCSM | Score | SDM | score | DUET | Score | CUPSAT | Torison | Predicted | SNP &GO 3D | Probablity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | **G60S** | **G83S** | **Destablizing** | **-0.123** | **Destablizing** | **-0.462** | **Destablizing** | **-3.48** | **Destabilizing** | **-0.77** | **Destablizing** | **Unfavourable** | **-3.11** | **Neutral** | **0.187** |
| 2 | **K5E** | **K95E** | **Destablizing** | **-0.276** | **Destablizing** | **-0.74** | **Destablizing** | **-0.08** | **destabilizing** | **-0.408** | **Destablizing** | **Favourable** | **-0.94** | **Neutral** | **0.014** |
| 3 | **N116S** | **N81S** | **Destablizing** | **-0.091** | **Stablizing** | **0.02** | **Destablizing** | **-1.05** | **Stablizing** | **0.103** | **Destablizing** | **Unfavourable** | **-0.12** | **Neutral** | **0.078** |
| 4 | D40V | D59V | Destablizing | -0.216 | Stablizing | 0.137 | Destablizing | -0.11 | Stablizing | 0.258 | Stablizing | Unfavourable | 1.24 | Neutral | 0.152 |
| 5 | F156L | F55L | Stablizing | 0.225 | Destablizing | -0.498 | Stablizing | 0 | Destablizing | -0.313 | Stablizing | Favourable | 1.98 | Neutral | 0.024 |

**Results:** G60S, K5E, and N116S were selected with the highest hit using different tools.

## C. ConSurf

### 1. APC

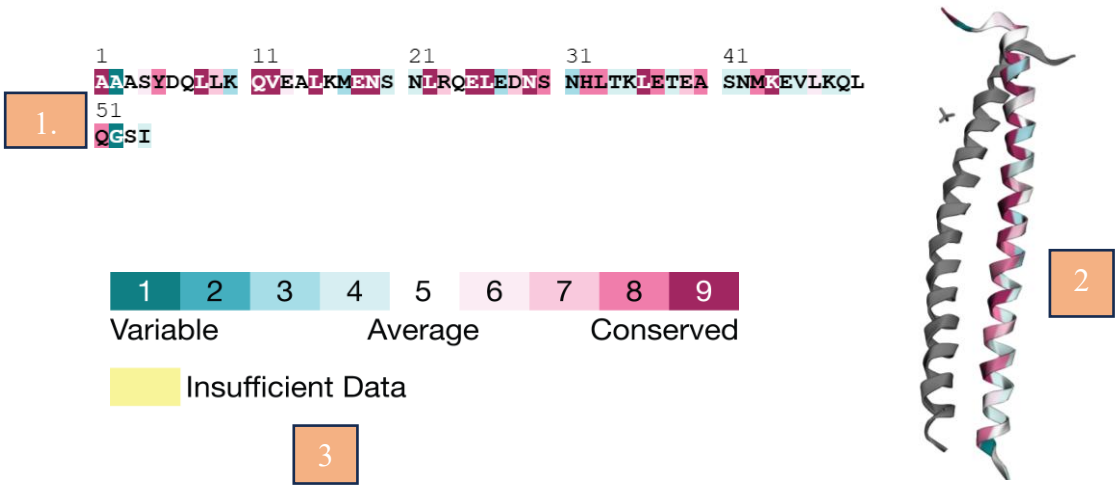



*Fig 6.1. 1.Showing the sequence of APC gene, 2. Structure of APC gene and 3. The scale.*

### 2. BRAF

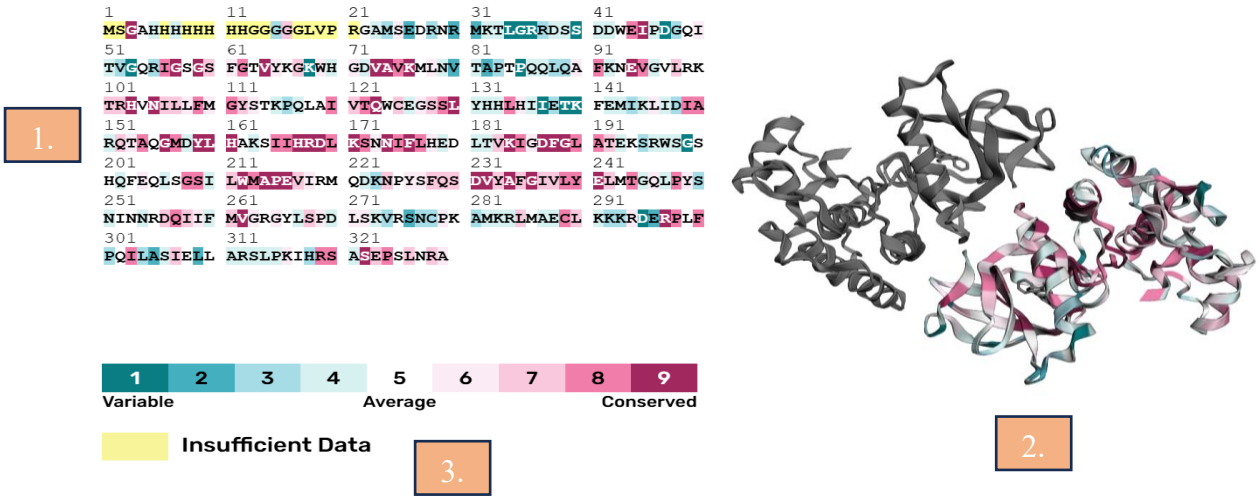



*Fig 6. 2; 1. Showing the sequence of BRAF gene, 2. Structure of BRAF gene and 3. The scale.*

## 3. KRAS

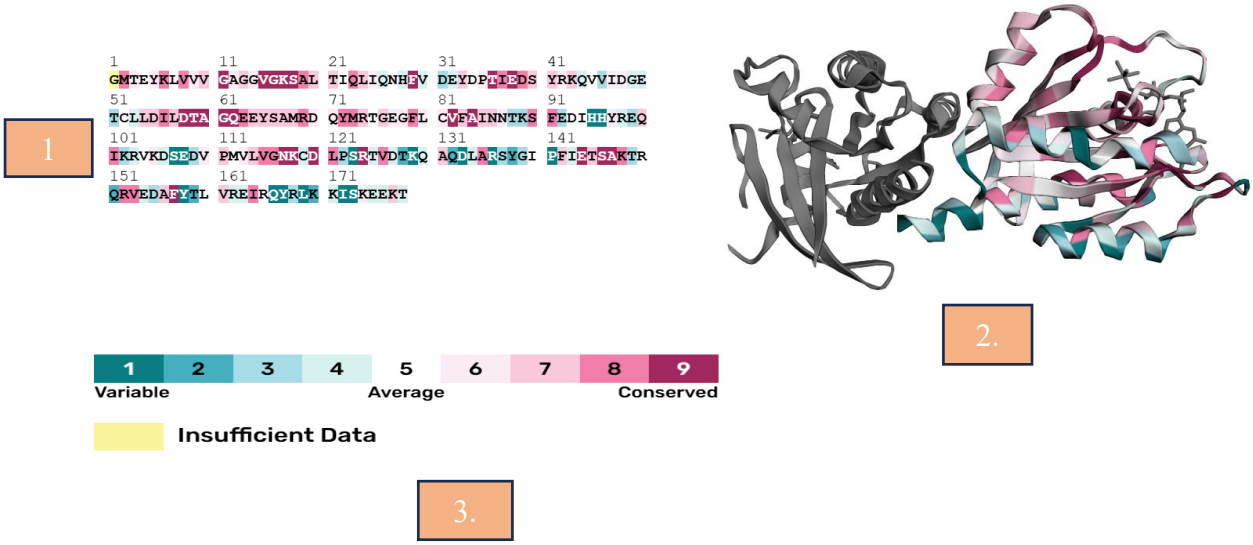


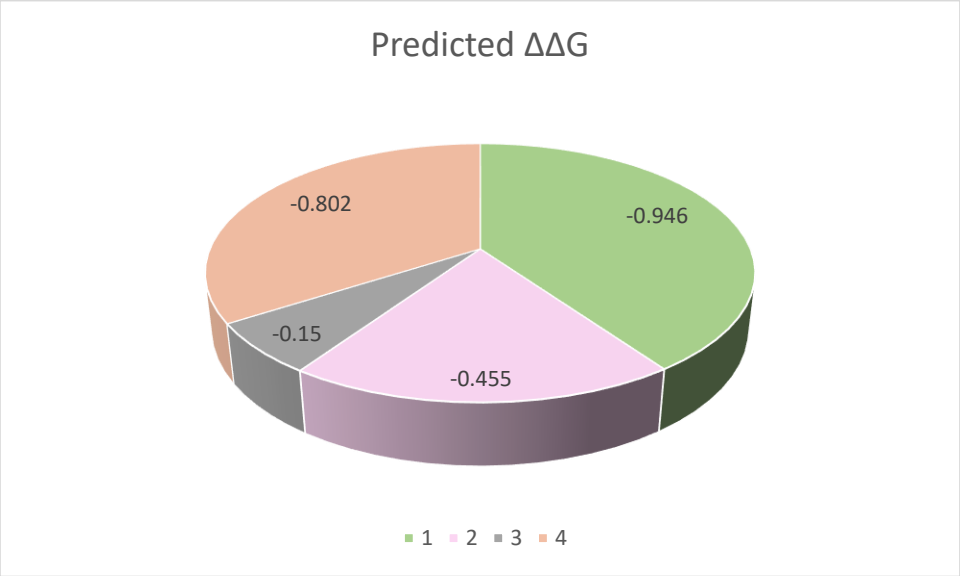

*Fig 6.3; 1. Showing the sequence of KRAS gene, 2. Structure of KRAS gene and 3. The scale.*

## D. mCSM (mutation Cutoff Scanning Matrix)

### 1. APC

*Table. 4.1; Protein Stability Change (ΔΔG) Upon Mutation Results for APC gene.*

| Index | Chain | Wild Residue | Residue Position | Mutant Residue | RSA (%) | Predicted ΔΔG | Outcome |
|---|---|---|---|---|---|---|---|
| 1 | A | G | 53 | C | 114.4 | -0.946 | Destabilizing |
| 2 | A | R | 24 | C | 76.1 | -0.455 | Destabilising |
| 3 | A | S | 21 | P | 69.2 | -0.15 | Destabilizing |
| 4 | A | R | 24 | G | 76.1 | -0.802 | Destabilizing |

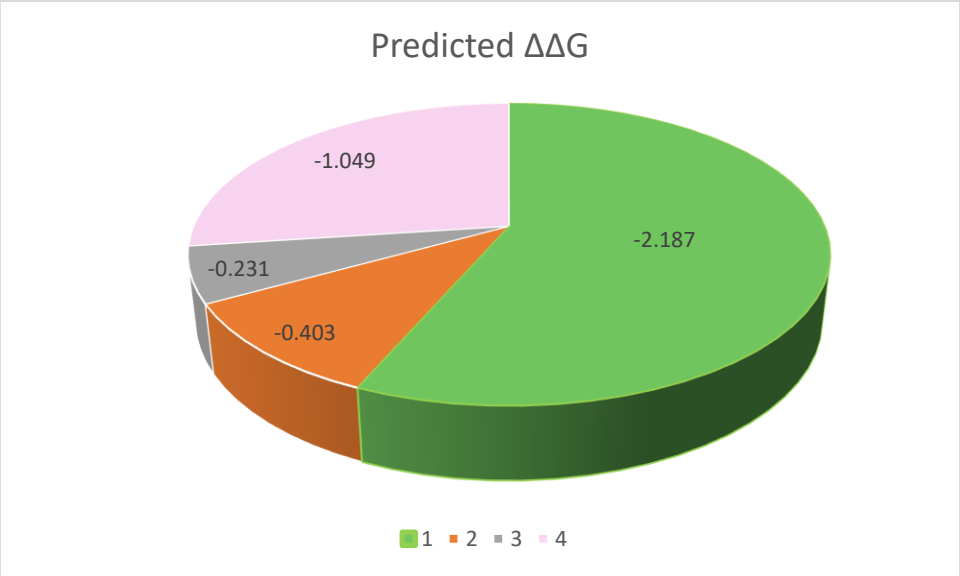



*Pie chart 1. Showing mutations in APC gene*

## 2. BRAF

*Table 4.2; Protein Stability Change (ΔΔG) Upon Mutation Results for BRAF gene.*

| Index | Chain | Wild Residue | Residue Position | Mutant Residue | RSA (%) | Predicted ΔΔG | Outcome |
|---|---|---|---|---|---|---|---|
| 1 | A | I | 452 | S | 2.8 | -2.187 | Highly Destabilizing |
| 2 | A | G | 455 | V | 116.1 | -0.403 | Destabilizing |
| 3 | A | V | 459 | E | 30 | -0.231 | Destabilizing |
| 4 | A | K | 473 | Q | 33 | -1.049 | Destabilizing |

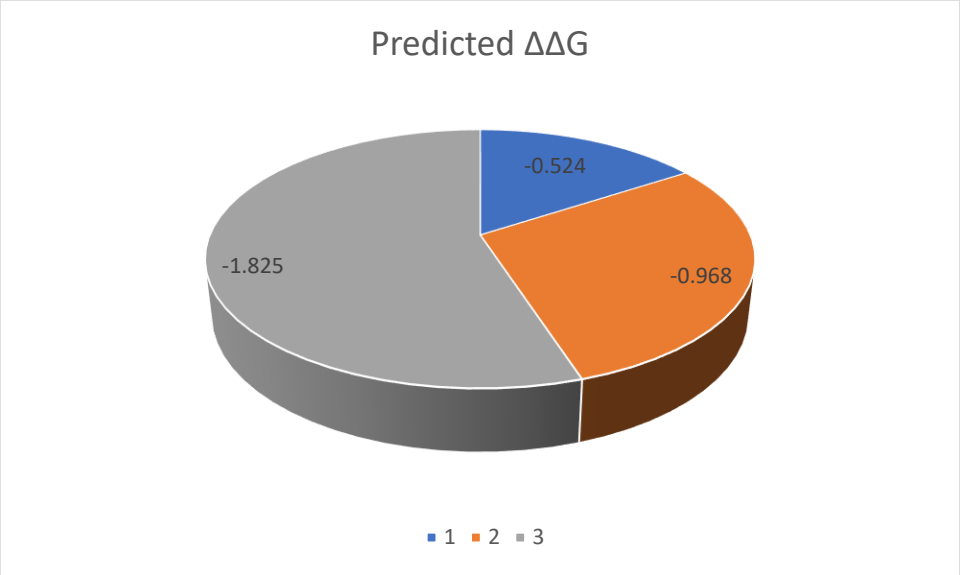



*Pie chart 2. Showing mutations in BRAF gene*

## 3. KRAS

*Table 3.3; Protein Stability Change (ΔΔG) Upon Mutation Results for KARS gene.*

| Index | Chain | Wild Residue | Residue Position | Mutant Residue | RSA (%) | Predicted ΔΔG | Outcome |
|---|---|---|---|---|---|---|---|
| 1 | A | G | 60 | S | 112.4 | -0.524 | Destabilizing |
| 2 | A | K | 5 | E | 31.2 | -0.968 | Destabilizing |
| 3 | A | N | 116 | S | 3.7 | -1.825 | Destabilizing |




*Pie chart 3. Showing mutations in KRAS gene*

## E. Molecular Dynamics Simulations

The results were graphed and visualised using the Visual Molecular Dynamics (VMD) program.

### Wild-type proteins

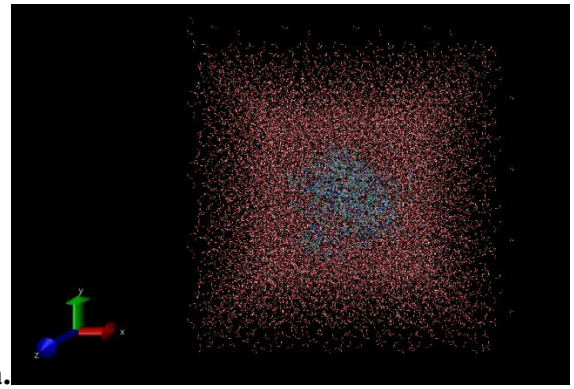
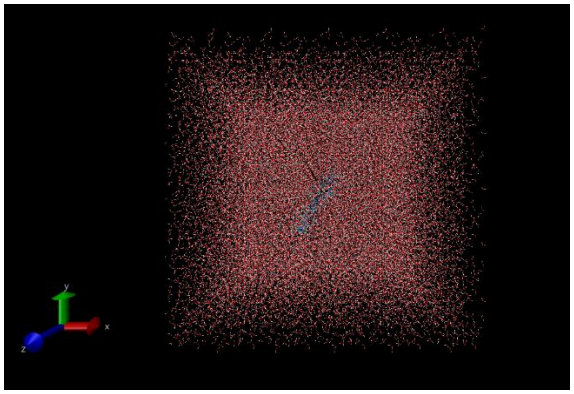
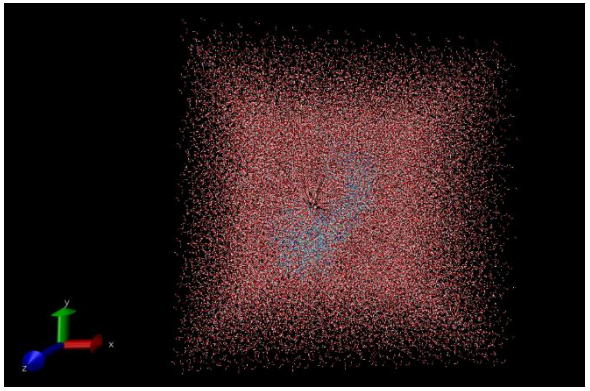
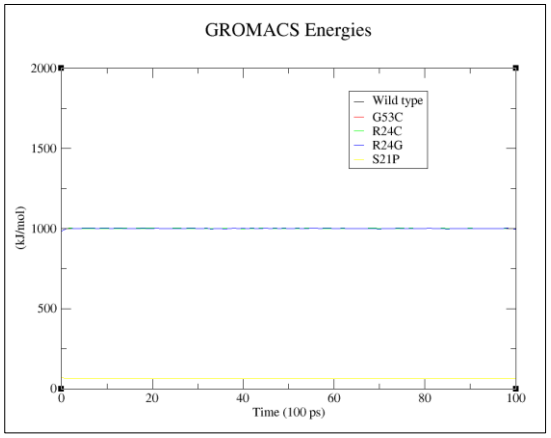
**a.**  **b.**  **c.** 

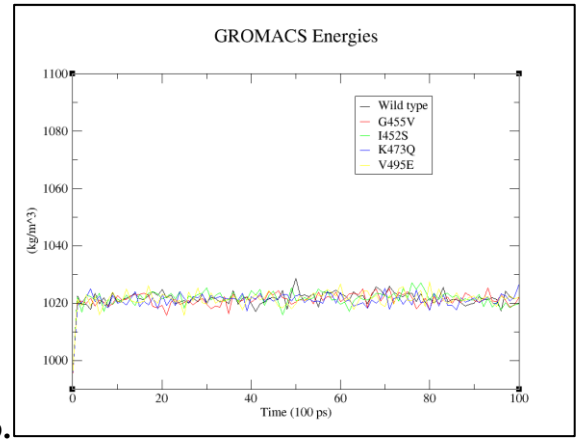
*Fig 7, Showing the three proteins (**a. APC, b. BRAF and C. KRAS**) in the cubic box, the red-colored areas indicate the ions that surround the blue-coloured protein*

## 1. Density

**a.**




**b.**




**c.**

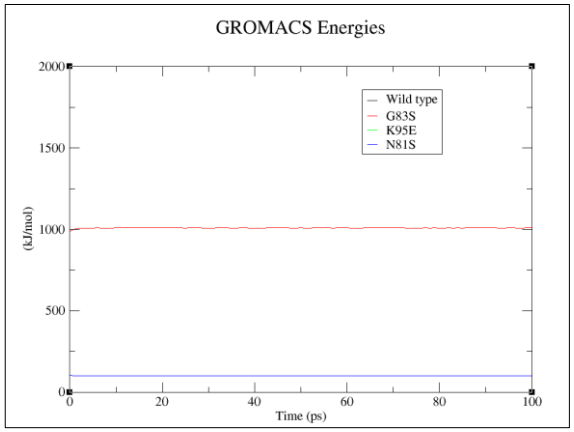



*Fig 8. The Density vs time graph for the wild type and mutant variants for 3 different genes (**a. APC, b. BRAF and C. KRAS**) are shown. The simulation was run for 100 ps. Different proteins are identified by colour coding as indicated above. For **Graph a**., the lines are flat and overlapping, showing the stability with time. For **Graph b**. shows more fluctuations within the range of approximately 1020 to 1040 kg/m³, indicating the system is stable but more dynamic. For **Graph c**., the lines are flat and overlapping, showing the stability with time.*

## 2. Potential energy

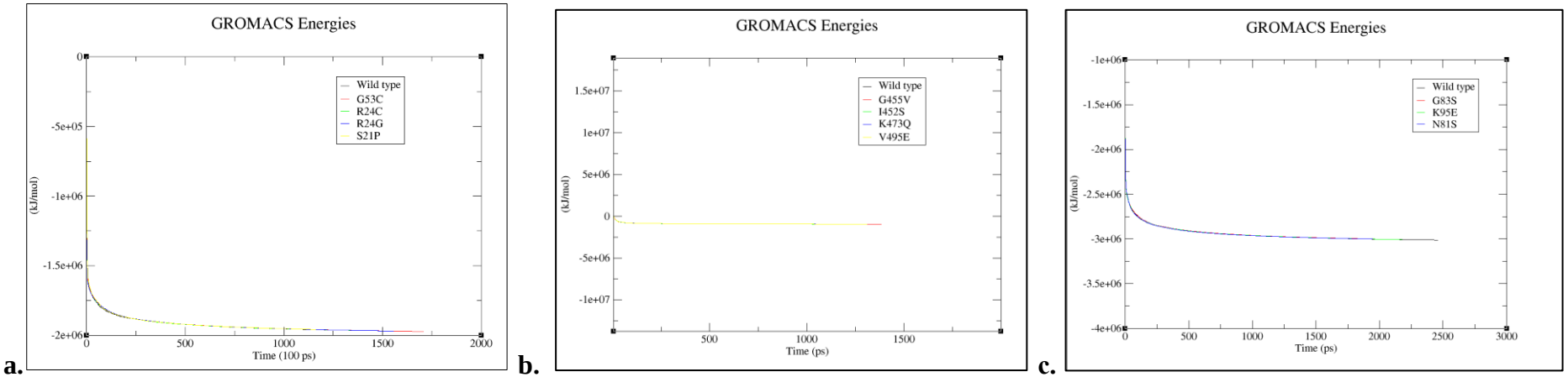



*Fig 9. The Potential energy (KJ/mol) vs time (ps) graph for the wild type and mutant variants for 3 different genes (**a. APC, b. BRAF and C. KRAS)** is shown. The simulation was run for 100 ps. Different proteins are identified by colour coding as indicated above. For **Graphs a and c.** show a sharp drop initially and then stabilise, resulting in the successful minimisation and equilibration but with different energy. For **Graph b**. shows a flat line across the time, indicating small variation between wild type and mutants.*

### 3. Pressure

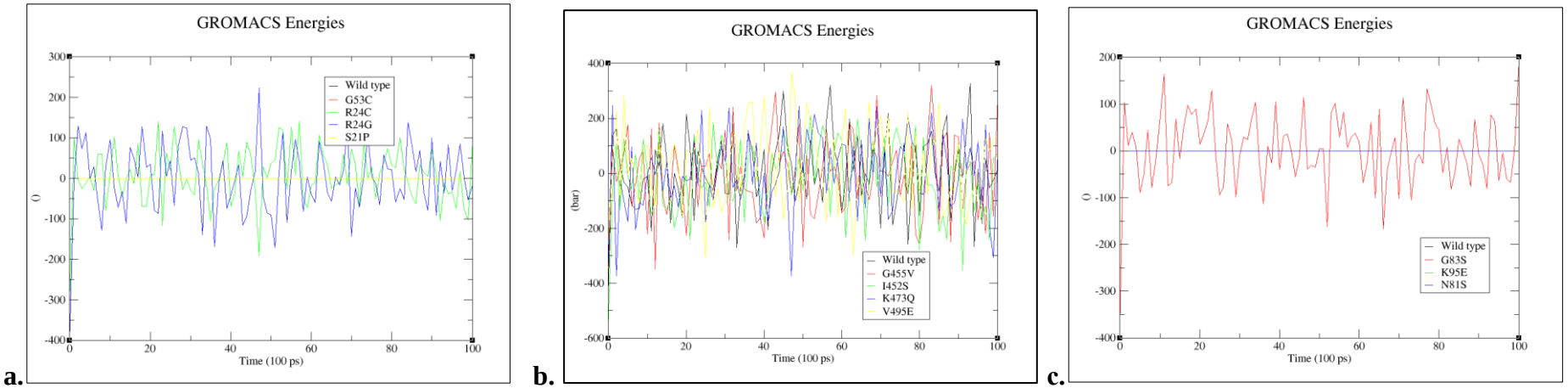



*Fig 10. The Pressure (Bar) vs time (ps) graph for the wild type and mutant variants for 3 different genes (**a. APC, b. BRAF and C. KRAS**) is shown. The simulation was run for 100 ps. Different proteins are identified by colour coding as indicated above. For **Graph a.** shows, other mutants undergo variations with time, but S21P remains at the equilibrium state. For **Graph b.** shows variation for all the wild and the mutants, which oscillates between +400 and -400 bar. For **Graph c**. shows, other mutants undergo variations with time, except N81S, which remains at the equilibrium state.*

## 4. Temperature

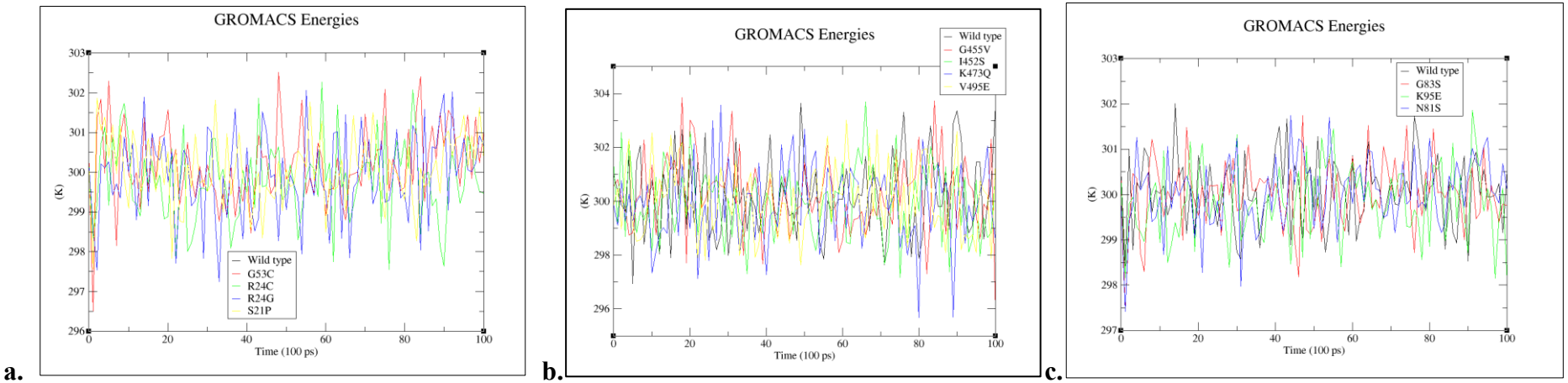



*Fig 11. The Temperature (K) vs time (ps) graph for the wild type and mutant variants for 3 different genes (**a. APC, b. BRAF and C. KRAS**) is, shown. Different proteins are identified by colour coding as indicated above. All three graphs show the fluctuation of the system for a simulation time of 100 ps. The simulation is seen around 300K, where minor deviations from the mean value. All the above shows a stable temperature and good thermal coupling during the simulations.*

.

## 5. Radius of Gyration

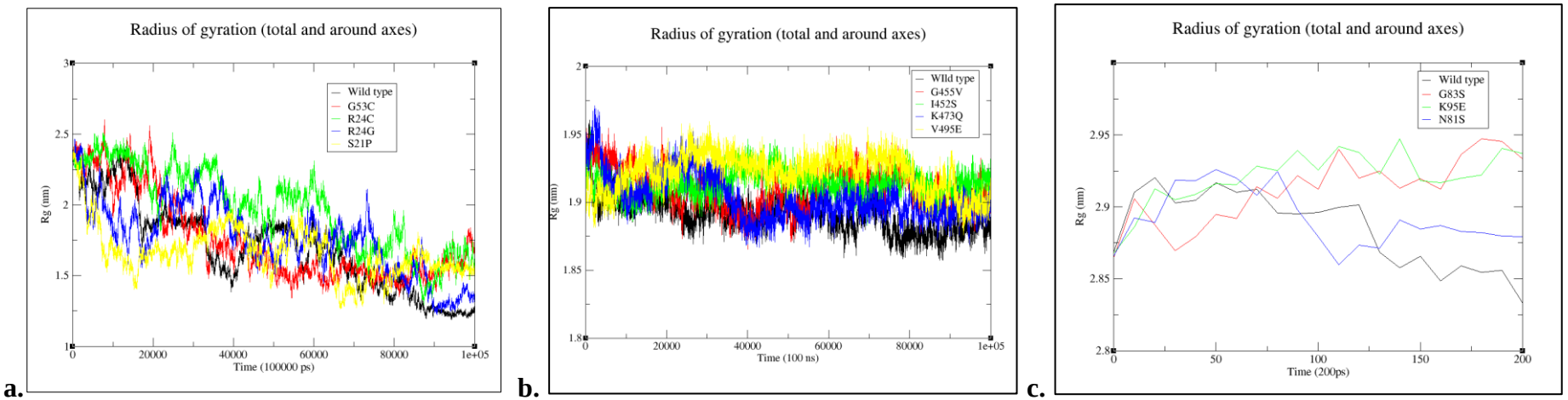



*Fig 12. The Radius of gyration (nm) vs time (ps) graph for the wild type and mutant variants for 3 different genes (**a. APC, b. BRAF and C. KRAS**) is shown. The measurement of a protein's compactness is known as the radius of gyration. All three graphs show the fluctuation of the system for a simulation time of 100ns except for KRAS gene. Different proteins are identified by colour coding as indicated above. For **Graph a.,** all the proteins, wild and mutants, show a decrease in the radius, indicating that with time, all the proteins become compact. For **Graph b.,** all the proteins show a decrease in radius with time, suggesting that all proteins become complacent with time. For **Graph c.,** shows fluctuation within a narrow range with no clear pattern over time. Wild types show a higher value of radius of gyration than the mutant variations. Mutant K95E and N83S show less radius of gyration, giving us the idea that with time, it is becoming more compact.*

.

## 6. Root Mean Square Deviation (RMSD)

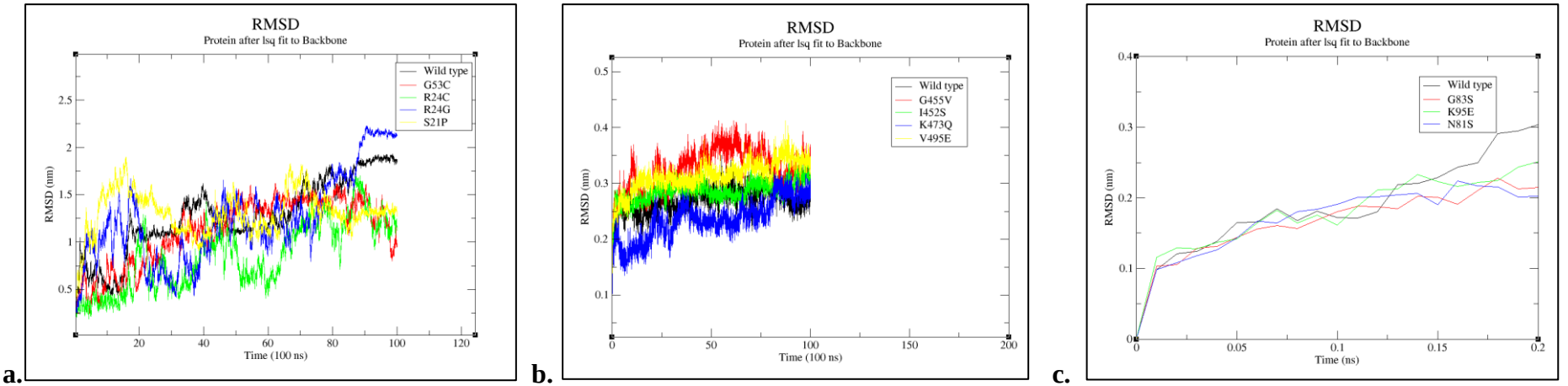



*Fig 13. The RMSD (nm) vs time (ps) graph for the wild type and mutant variants for 3 different genes (**a. APC, b. BRAF and C. KRAS**) is shown. Different proteins are identified by colour coding as indicated above. All three graphs show the fluctuation of the system for a simulation time of 100 ns. For **Graph a.,** all the protein starts from low RMSD values and increase with time. R24G shows high RMSD values, suggesting larger conformational deviations and possible instability as compared with the wild type. For **Graph b**., all the proteins start from low RMSD values and increase with time. V495E shows high RMSD values, suggesting larger conformational deviations and possible instability as compared with the wild type For **Graph c**. shows a sharp increase in the RMSD values initially and followed by less variation in RMSD values with time. G83S and K95E show higher RMSD values than the wild type, suggesting larger conformational deviations and possible instability.*

# CHAPTER 5: CONCLUSION

The computational studies have significantly helped us in understanding the mechanisms underlying CRC. By pointing out the key pathways and their interaction that are taking place during the progression and development of CRC. Inactivation of APC accounts for 80 to 85% of CRC, activation of the BRAF gene accounts for 10 to 15% of CRC and activation of the KRAS gene accounts for 40 to 50% of CRC because of the deviations from the pathways.

The analysis of CRC using the SNP database, which looks into the possible effects of particular mutations using both sequence and structure-based analysis. We investigated the stability, conservation, and structural implications of identified variants using a variety of bioinformatics tools, such as Multiple Sequence Alignment (MSA), GROMACS, mCSM, and ConSurf analysis. Three important CRC-related genes—APC, BRAF, and KRAS—were the focus of our investigation. As a result, we identified potentially impactful mutations including such as G53C, R24C, R24G, and S21P in APC; G455V, I452S, K473Q and V495E in BRAF; and G83S, K95E, and N81S in KRAS. Each of those genes has a varying percentage of causes linked to CRC. The G53C mutation is the most common mutation found in the APC gene, followed by the I452S mutation in the BRAF gene and the N116S mutation in the KRAS gene. Furthermore, a better comprehension of the structural disruptions brought about by genetic changes was made possible by the additional analysis of frameshift mutations.

Future research should concentrate on validating these computational findings through experimental studies and investigating the clinical implications of these regulatory factors in the treatment of CRC. Ultimately, this work contributes to ongoing efforts to enhance patient outcomes and develop more effective strategies for managing CRC. The mutations also claimed to affect the stability of the wild-type protein, as concluded from the MD simulations.

Overall, this study advances our knowledge of the structural alterations caused by mutations in genes linked to CRC and establishes the foundation for further experimental validation and therapeutic investigations.

# CHAPTER 6: REFERENCES


[1] B. Delatte and F. Fuks, "TET proteins: on the frenetic hunt for new cytosine modifications," *Brief Funct Genomics*, vol. 12, no. 3, pp. 191–204, May 2013, doi: 10.1093/bfgp/elt010.

[2] Y.-F. He *et al.*, "Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA," *Science (1979)*, vol. 333, no. 6047, pp. 1303–1307, Sep. 2011, doi: 10.1126/science.1210944.

[3] S. Ito *et al.*, "Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.," *Science*, vol. 333, no. 6047, pp. 1300–3, Sep. 2011, doi: 10.1126/science.1210597.

[4] M. Tahiliani *et al.*, "Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.," *Science*, vol. 324, no. 5929, pp. 930–5, May 2009, doi: 10.1126/science.1170116.

[5] S. Kriaucionis and N. Heintz, "The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain.," *Science*, vol. 324, no. 5929, pp. 929–30, May 2009, doi: 10.1126/science.1169786.

[6] R. Richa and R. P. Sinha, "Hydroxymethylation of DNA: an epigenetic marker.," *EXCLI J*, vol. 13, pp. 592–610, 2014.

[7] S. Ito, A. C. D'Alessio, O. V Taranova, K. Hong, L. C. Sowers, and Y. Zhang, "Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification.," *Nature*, vol. 466, no. 7310, pp. 1129–33, Aug. 2010, doi: 10.1038/nature09303.

[8] J. A. Hackett and M. A. Surani, "DNA methylation dynamics during the mammalian life cycle.," *Philos Trans R Soc Lond B Biol Sci*, vol. 368, no. 1609, p. 20110328, Jan. 2013, doi: 10.1098/rstb.2011.0328.

[9] R. Holliday and J. E. Pugh, "DNA modification mechanisms and gene activity during development.," *Science*, vol. 187, no. 4173, pp. 226–32, Jan. 1975.

[10] A. Portela and M. Esteller, "Epigenetic modifications and human disease," *Nat Biotechnol*, vol. 28, no. 10, pp. 1057–1068, Oct. 2010, doi: 10.1038/nbt.1685.

[11] R. A. Irizarry *et al.*, "The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores," *Nat Genet*, vol. 41, no. 2, pp. 178–186, Feb. 2009, doi: 10.1038/ng.298.

[12] S. Saxonov, P. Berg, and D. L. Brutlag, "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters," *Proceedings of the National Academy of Sciences*, vol. 103, no. 5, pp. 1412–1417, Jan. 2006, doi: 10.1073/pnas.0510310103.

[13] N. Shimoda, T. Izawa, A. Yoshizawa, H. Yokoi, Y. Kikuchi, and N. Hashimoto, "Decrease in cytosine methylation at CpG island shores and increase in DNA fragmentation during zebrafish aging," *Age (Omaha)*, vol. 36, no. 1, pp. 103–115, Feb. 2014, doi: 10.1007/s11357-013-9548-5.

[14] A. Maresca, M. Zaffagnini, L. Caporali, V. Carelli, and C. Zanna, "DNA methyltransferase 1 mutations and mitochondrial pathology: is mtDNA methylated?," *Front Genet*, vol. 6, p. 90, 2015, doi: 10.3389/fgene.2015.00090.

[15] H. Denis, 'Matladi N Ndlovu, and F. Fuks, "Regulation of mammalian DNA methyltransferases: a route to new mechanisms," *EMBO Rep*, vol. 12, no. 7, pp. 647–656, Jul. 2011, doi: 10.1038/embor.2011.110.

[16] T. H. Bestor, "The DNA methyltransferases of mammals.," *Hum Mol Genet*, vol. 9, no. 16, pp. 2395–402, Oct. 2000, doi: 10.1093/hmg/9.16.2395.

[17] M. Okano, D. W. Bell, D. A. Haber, and E. Li, "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.," *Cell*, vol. 99, no. 3, pp. 247–57, Oct. 1999, doi: 10.1016/s0092-8674(00)81656-6.

[18] K. Hata, M. Okano, H. Lei, and E. Li, "Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice.," *Development*, vol. 129, no. 8, pp. 1983–93, Apr. 2002, doi: 10.1242/dev.129.8.1983.

[19] G.-L. Xu *et al.*, "Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene," *Nature*, vol. 402, no. 6758, pp. 187–191, Nov. 1999, doi: 10.1038/46052.

[20] K. Tatton-Brown *et al.*, "Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability.," *Nat Genet*, vol. 46, no. 4, pp. 385–8, Apr. 2014, doi: 10.1038/ng.2917.

[21] M. G. Goll *et al.*, "Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2.," *Science*, vol. 311, no. 5759, pp. 395–8, Jan. 2006, doi: 10.1126/science.1120976.

[22] M. Schaefer and F. Lyko, "Solving the Dnmt2 enigma.," *Chromosoma*, vol. 119, no. 1, pp. 35–40, Feb. 2010, doi: 10.1007/s00412-009-0240-6.

[23] A. Dong, J. A. Yoder, X. Zhang, L. Zhou, T. H. Bestor, and X. Cheng, "Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA.," *Nucleic Acids Res*, vol. 29, no. 2, pp. 439–48, Jan. 2001, doi: 10.1093/nar/29.2.439.

[24] P. M. Das and R. Singal, "DNA Methylation and Cancer," *Journal of Clinical Oncology*, vol. 22, no. 22, pp. 4632–4642, Nov. 2004, doi: 10.1200/JCO.2004.07.151.

[25] C. Plass, "Cancer epigenomics," *Hum Mol Genet*, vol. 11, no. 20, pp. 2479–2488, Oct. 2002, doi: 10.1093/hmg/11.20.2479.

[26] S. B. Baylin, "Aberrant Methylation of Gene Promoters in Cancer--Concepts, Misconcepts, and Promise," *J Natl Cancer Inst*, vol. 92, no. 18, pp. 1460–1461, Sep. 2000, doi: 10.1093/jnci/92.18.1460.

[27] J. Frigola *et al.*, "Differential DNA hypermethylation and hypomethylation signatures in colorectal cancer," *Hum Mol Genet*, vol. 14, no. 2, pp. 319–326, Jan. 2005, doi: 10.1093/hmg/ddi028.

[28] Q. L. A. L. M. S. B. B. J. P. I. N Ahuja, "Aging and DNA methylation in colorectal mucosa and cancer," Dec. 1998.

[29] M. A. Riojas, M. Guo, S. C. Glöckner, E. O. Machida, S. B. Baylin, and N. Ahuja, "Methylation-induced silencing of ASC/TMS1, a pro-apoptotic gene, is a late-stage event in colorectal cancer," *Cancer Biol Ther*, vol. 6, no. 11, pp. 1710–1716, Nov. 2007, doi: 10.4161/cbt.6.11.4829.

[30] B. Bin Lee *et al.*, "Aberrant Methylation of *APC, MGMT, RASSF2A* , and *Wif-1* Genes in Plasma as a Biomarker for Early Detection of Colorectal Cancer," *Clinical Cancer Research*, vol. 15, no. 19, pp. 6185–6191, Oct. 2009, doi: 10.1158/1078-0432.CCR-09-0111.

[31] T. J. R. Harris and F. McCormick, "The molecular pathology of cancer," *Nat Rev Clin Oncol*, vol. 7, no. 5, pp. 251–265, May 2010, doi: 10.1038/nrclinonc.2010.41.

[32] R. Kalluri and R. A. Weinberg, "The basics of epithelial-mesenchymal transition," *Journal of Clinical Investigation*, vol. 119, no. 6, pp. 1420–1428, Jun. 2009, doi: 10.1172/JCI39104.

[33] R. Baghban *et al.*, "Tumor microenvironment complexity and therapeutic implications at a glance," *Cell Communication and Signaling*, vol. 18, no. 1, p. 59, Dec. 2020, doi: 10.1186/s12964-020-0530-4.

[34] A. Mehrgou *et al.*, "Roles of miRNAs in Colorectal Cancer: Therapeutic Implications and Clinical Opportunities," *Adv Pharm Bull*, Jul. 2020, doi: 10.34172/apb.2021.029.

[35] A. D. Roth *et al.*, "Prognostic Role of *KRAS* and *BRAF* in Stage II and III Resected Colon Cancer: Results of the Translational Study on the PETACC-3, EORTC 40993, SAKK 60-00 Trial," *Journal of Clinical Oncology*, vol. 28, no. 3, pp. 466–474, Jan. 2010, doi: 10.1200/JCO.2009.23.3452.

[36] A. Mehrgou, S. Ebadollahi, B. Jameie, and S. Teimourian, "Analysis of subtype-specific and common Gene/MiRNA expression profiles of four main breast cancer subtypes using bioinformatic approach; Characterization of four genes, and two MicroRNAs with possible diagnostic and prognostic values," *Inform Med Unlocked*, vol. 20, p. 100425, 2020, doi: 10.1016/j.imu.2020.100425.

[37] A. Ventura and T. Jacks, "MicroRNAs and Cancer: Short RNAs Go a Long Way," *Cell*, vol. 136, no. 4, pp. 586–591, Feb. 2009, doi: 10.1016/j.cell.2009.02.005.

[38] M. Menigatti *et al.*, "Epigenetic silencing of monoallelically methylated miRNA loci in precancerous colorectal lesions," *Oncogenesis*, vol. 2, no. 7, pp. e56–e56, Jul. 2013, doi: 10.1038/oncsis.2013.21.

[39] M. Cui *et al.*, "Circulating MicroRNAs in Cancer: Potential and Challenge," *Front Genet*, vol. 10, Jul. 2019, doi: 10.3389/fgene.2019.00626.

[40] R. W. Carthew and E. J. Sontheimer, "Origins and Mechanisms of miRNAs and siRNAs," *Cell*, vol. 136, no. 4, pp. 642–655, Feb. 2009, doi: 10.1016/j.cell.2009.01.035.

[41] W. K. K. Wu *et al.*, "MicroRNA dysregulation in gastric cancer: a new player enters the game," *Oncogene*, vol. 29, no. 43, pp. 5761–5771, Oct. 2010, doi: 10.1038/onc.2010.352.

[42] V. P. Dave *et al.*, "MicroRNA amplification and detection technologies: opportunities and challenges for point of care diagnostics," *Laboratory Investigation*, vol. 99, no. 4, pp. 452–469, Apr. 2019, doi: 10.1038/s41374-018-0143-3.

[43] X.-Y. Dong *et al.*, "Implication of snoRNA U50 in human breast cancer," *Journal of Genetics and Genomics*, vol. 36, no. 8, pp. 447–454, Aug. 2009, doi: 10.1016/S1673-8527(08)60134-4.

[44] S. A. Waldman and A. Terzic, "Applications of MicroRNA in Cancer: Exploring the Advantages of miRNA," *Clin Transl Sci*, vol. 2, no. 3, pp. 248–249, Jun. 2009, doi: 10.1111/j.1752-8062.2009.00110.x.

[45] M. Majidinia, S. G. Darband, M. Kaviani, S. M. Nabavi, R. Jahanban-Esfahlan, and B. Yousefi, "Cross-regulation between Notch signaling pathway and miRNA machinery in cancer," *DNA Repair (Amst)*, vol. 66–67, pp. 30–41, Jun. 2018, doi: 10.1016/j.dnarep.2018.04.002.

[46] A. L. Eamens, N. A. Smith, S. J. Curtin, M.-B. Wang, and P. M. Waterhouse, "The *Arabidopsis thaliana* double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes," *RNA*, vol. 15, no. 12, pp. 2219–2235, Dec. 2009, doi: 10.1261/rna.1646909.

[47] D. P. Bartel, "MicroRNAs: Target Recognition and Regulatory Functions," *Cell*, vol. 136, no. 2, pp. 215–233, Jan. 2009, doi: 10.1016/j.cell.2009.01.002.

[48] M. Wakiyama, K. Takimoto, O. Ohara, and S. Yokoyama, "*Let-* 7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system," *Genes Dev*, vol. 21, no. 15, pp. 1857–1862, Aug. 2007, doi: 10.1101/gad.1566707.

[49] A. Grimson, K. K.-H. Farh, W. K. Johnston, P. Garrett-Engele, L. P. Lim, and D. P. Bartel, "MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing," *Mol Cell*, vol. 27, no. 1, pp. 91–105, Jul. 2007, doi: 10.1016/j.molcel.2007.06.017.

[50] P. Pidíková and I. Herichová, "miRNA Clusters with Up-Regulated Expression in Colorectal Cancer," *Cancers (Basel)*, vol. 13, no. 12, p. 2979, Jun. 2021, doi: 10.3390/cancers13122979.

[51] L.-A. MacFarlane and P. R. Murphy, "MicroRNA: Biogenesis, Function and Role in Cancer," *Curr Genomics*, vol. 11, no. 7, pp. 537–561, Nov. 2010, doi: 10.2174/138920210793175895.

[52] K. Murakami *et al.*, "A genome-scale CRISPR screen reveals factors regulating Wnt-dependent renewal of mouse gastric epithelial cells," *Proceedings of the National Academy of Sciences*, vol. 118, no. 4, Jan. 2021, doi: 10.1073/pnas.2016806118.

[53] K. Saito, A. Ishizuka, H. Siomi, and M. C. Siomi, "Processing of Pre-microRNAs by the Dicer-1–Loquacious Complex in Drosophila Cells," *PLoS Biol*, vol. 3, no. 7, p. e235, May 2005, doi: 10.1371/journal.pbio.0030235.

[54] S. He and C. Liang, "Frameshift mutation of *UVRAG* : Switching a tumor suppressor to an oncogene in colorectal cancer," *Autophagy*, vol. 11, no. 10, pp. 1939–1940, Oct. 2015, doi: 10.1080/15548627.2015.1086523.

[55] A. J. McMichael and G. G. Giles, "Cancer in migrants to Australia: extending the descriptive epidemiological data.," *Cancer Res*, vol. 48, no. 3, pp. 751–6, Feb. 1988.

[56] K. W. Kinzler and B. Vogelstein, "Lessons from Hereditary Colorectal Cancer," *Cell*, vol. 87, no. 2, pp. 159–170, Oct. 1996, doi: 10.1016/S0092-8674(00)81333-1.

[57] E. R. Fearon and B. Vogelstein, "A genetic model for colorectal tumorigenesis," *Cell*, vol. 61, no. 5, pp. 759–767, Jun. 1990, doi: 10.1016/0092-8674(90)90186-I.

[58] P. Laurent-Puig, J. Agostini, and K. Maley, "[Colorectal oncogenesis].," *Bull Cancer*, vol. 97, no. 11, pp. 1311–21, Nov. 2010, doi: 10.1684/bdc.2010.1216.

[59] B. Rubinfeld, I. Albert, E. Porfiri, S. Munemitsu, and P. Polakis, "Loss of beta-catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene.," *Cancer Res*, vol. 57, no. 20, pp. 4624–30, Oct. 1997.

[60] R. Nusse and H. Clevers, "Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities.," *Cell*, vol. 169, no. 6, pp. 985–999, Jun. 2017, doi: 10.1016/j.cell.2017.05.016.

[61] P. J. Morin *et al.*, "Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC.," *Science*, vol. 275, no. 5307, pp. 1787–90, Mar. 1997, doi: 10.1126/science.275.5307.1787.

[62] J. Qian *et al.*, "The APC Tumor Suppressor Inhibits DNA Replication by Directly Binding to DNA via Its Carboxyl Terminus," *Gastroenterology*, vol. 135, no. 1, pp. 152–162, Jul. 2008, doi: 10.1053/j.gastro.2008.03.074.

[63] R. Fodde, "The APC gene in colorectal cancer," *Eur J Cancer*, vol. 38, no. 7, pp. 867–871, 2002.

[64] B. M. , & F. J. Z. Boman, "An APC:WNT counter-current-like mechanism regulates cell division along the human colonic crypt axis: a mechanism that explains how APC mutations induce proliferative abnormalities that drive colon cancer development," vol. 3, 2013.

[65] A. C. b e 1, S. D. a b, N. M. b, A. F. S. d, A. M. a c, G. V. Maricarmen Cristofaro a b 1, "Adenomatous polyposis coli (APC)-induced apoptosis of HT29 colorectal cancer cells depends on mitochondrial oxidative metabolism," vol. 1852, no. 9, pp. 1719–1728, 2015.

[66] J. Bellis *et al.*, "The tumor suppressor Apc controls planar cell polarities central to gut homeostasis.," *J Cell Biol*, vol. 198, no. 3, pp. 331–41, Aug. 2012, doi: 10.1083/jcb.201204086.

[67] S. M. H. Kashfi, F. Behboudi Farahbakhsh, M. Golmohammadi, E. Nazemalhosseini Mojarad, P. Azimzadeh, and H. Asadzadeh Aghdaie, "Frameshift Mutations (Deletion at Codon 1309 and Codon 849) in the APC Gene in Iranian FAP Patients: a Case Series and Review of the Literature.," *Int J Mol Cell Med*, vol. 3, no. 3, pp. 196–202, 2014.

[68] Y. Zhang and X. Wang, "Targeting the Wnt/β-catenin signaling pathway in cancer," *J Hematol Oncol*, vol. 13, no. 1, p. 165, Dec. 2020, doi: 10.1186/s13045-020-00990-3.

[69] T. W. Parker and K. L. Neufeld, "APC controls Wnt-induced β-catenin destruction complex recruitment in human colonocytes," *Sci Rep*, vol. 10, no. 1, p. 2957, Feb. 2020, doi: 10.1038/s41598-020-59899-z.

[70] K. Imkeller *et al.*, "Metabolic balance in colorectal cancer is maintained by optimal Wnt signaling levels," *Mol Syst Biol*, vol. 18, no. 8, Aug. 2022, doi: 10.15252/msb.202110874.

[71] W. Hankey, W. L. Frankel, and J. Groden, "Functions of the APC tumor suppressor protein dependent and independent of canonical WNT signaling: implications for therapeutic targeting.," *Cancer Metastasis Rev*, vol. 37, no. 1, pp. 159–172, Mar. 2018, doi: 10.1007/s10555-017-9725-6.

[72] R. Fodde, "The APC gene in colorectal cancer.," *Eur J Cancer*, vol. 38, no. 7, pp. 867–71, May 2002, doi: 10.1016/s0959-8049(02)00040-0.

[73] H. Nagase and Y. Nakamura, "Mutations of the APC (adenomatous polyposis coli) gene.," *Hum Mutat*, vol. 2, no. 6, pp. 425–34, 1993, doi: 10.1002/humu.1380020602.

[74] J. Huang, S. Zheng, and S. Jin, "[APC mutation analysis in sporadic colorectal cancer].," *Zhonghua Zhong Liu Za Zhi*, vol. 18, no. 6, pp. 415–8, Nov. 1996.

[75] R. Gryfe, C. Swallow, B. Bapat, M. Redston, S. Gallinger, and J. Couture, "Molecular biology of colorectal cancer.," *Curr Probl Cancer*, vol. 21, no. 5, pp. 233–300, 1997, doi: 10.1016/s0147-0272(97)80003-7.

[76] H. Iino *et al.*, "DNA microsatellite instability and mismatch repair protein loss in adenomas presenting in hereditary non-polyposis colorectal cancer.," *Gut*, vol. 47, no. 1, pp. 37–42, Jul. 2000, doi: 10.1136/gut.47.1.37.

[77] S. Salahshor, U. Kressner, L. Pâhlman, B. Glimelius, G. Lindmark, and A. Lindblom, "Colorectal cancer with and without microsatellite instability involves different genes.," *Genes Chromosomes Cancer*, vol. 26, no. 3, pp. 247–52, Nov. 1999.

[78] S. Olschwang, A. Tiret, P. Laurent-Puig, M. Muleris, R. Parc, and G. Thomas, "Restriction of ocular fundus lesions to a specific subgroup of APC mutations in adenomatous polyposis coli patients.," *Cell*, vol. 75, no. 5, pp. 959–68, Dec. 1993, doi: 10.1016/0092-8674(93)90539-3.

[79] D. R. Davies *et al.*, "Severe Gardner syndrome in families with mutations restricted to a specific region of the APC gene.," *Am J Hum Genet*, vol. 57, no. 5, pp. 1151–8, Nov. 1995.

[80] X. Cao, K. Weng Eu, F. Seow-Choen, Y. Zao, and P. Yean Cheah, "APC mutation and phenotypic spectrum of Singapore familial adenomatous polyposis patients," *European Journal of Human Genetics*, vol. 8, no. 1, pp. 42–48, Jan. 2000, doi: 10.1038/sj.ejhg.5200397.

[81] L. Zhang and J. W. Shay, "Multiple Roles of APC and its Therapeutic Implications in Colorectal Cancer.," *J Natl Cancer Inst*, vol. 109, no. 8, Aug. 2017, doi: 10.1093/jnci/djw332.

[82] A. S. Aghabozorgi *et al.*, "Role of adenomatous polyposis coli (APC) gene mutations in the pathogenesis of colorectal cancer; current status and perspectives," *Biochimie*, vol. 157, pp. 64–71, Feb. 2019, doi: 10.1016/j.biochi.2018.11.003.

[83] R. Fodde, "The APC gene in colorectal cancer," *Eur J Cancer*, vol. 38, no. 7, pp. 867–871, May 2002, doi: 10.1016/S0959-8049(02)00040-0.

[84] D. Barras, "*BRAF* Mutation in Colorectal Cancer: An Update," *Biomark Cancer*, vol. 7s1, p. BIC.S25248, Jan. 2015, doi: 10.4137/BIC.S25248.

[85] F. Caputo *et al.*, "BRAF-Mutated Colorectal Cancer: Clinical and Molecular Insights," *Int J Mol Sci*, vol. 20, no. 21, p. 5369, Oct. 2019, doi: 10.3390/ijms20215369.

[86] Y. Mori *et al.*, "Somatic mutations of the *APC* gene in colorectal tumors: mutation cluster region in the *APC* gene," *Hum Mol Genet*, vol. 1, no. 4, pp. 229–233, 1992, doi: 10.1093/hmg/1.4.229.

[87] J. H. Strickler, C. Wu, and T. Bekaii-Saab, "Targeting BRAF in metastatic colorectal cancer: Maximizing molecular approaches," *Cancer Treat Rev*, vol. 60, pp. 109–119, Nov. 2017, doi: 10.1016/j.ctrv.2017.08.006.

[88] R. M. Guerrero *et al.*, "Targeting BRAF V600E in metastatic colorectal cancer: where are we today?," *Ecancermedicalscience*, vol. 16, p. 1489, 2022, doi: 10.3332/ecancer.2022.1489.

[89] D. Xu, Q. Shao, C. Zhou, A. Mahmood, and J. Zhang, "*In Silico* Analysis of nsSNPs of Human KRAS Gene and Protein Modeling Using Bioinformatic Tools," *ACS Omega*, vol. 8, no. 14, pp. 13362–13370, Apr. 2023, doi: 10.1021/acsomega.3c00804.

[90] A. E. Karnoub and R. A. Weinberg, "Ras oncogenes: split personalities.," *Nat Rev Mol Cell Biol*, vol. 9, no. 7, pp. 517–31, Jul. 2008, doi: 10.1038/nrm2438.

[91] L. Huang, Z. Guo, F. Wang, and L. Fu, "KRAS mutation: from undruggable to druggable in cancer," *Signal Transduct Target Ther*, vol. 6, no. 1, p. 386, Nov. 2021, doi: 10.1038/s41392-021-00780-4.

[92] S. J. Baker *et al.*, "p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis.," *Cancer Res*, vol. 50, no. 23, pp. 7717–22, Dec. 1990.

[93] U. Kilic *et al.*, "Particular phosphorylation of PI3K/Akt on Thr308 via PDK-1 and PTEN mediates melatonin's neuroprotective activity after focal cerebral ischemia in mice.," *Redox Biol*, vol. 12, pp. 657–665, Aug. 2017, doi: 10.1016/j.redox.2017.04.006.

[94] I. R. Vetter and A. Wittinghofer, "The Guanine Nucleotide-Binding Switch in Three Dimensions," *Science (1979)*, vol. 294, no. 5545, pp. 1299–1304, Nov. 2001, doi: 10.1126/science.1062023.

[95] S. Aissi *et al.*, "TP53 mutations in colorectal cancer from Tunisia: relationships with site of tumor origin, microsatellite instability and KRAS mutations.," *Mol Biol Rep*, vol. 41, no. 3, pp. 1807–13, Mar. 2014, doi: 10.1007/s11033-014-3030-z.

[96] M. Dillon, A. Lopez, E. Lin, D. Sales, R. Perets, and P. Jain, "Progress on Ras/MAPK Signaling Research and Targeting in Blood and Solid Cancers.," *Cancers (Basel)*, vol. 13, no. 20, Oct. 2021, doi: 10.3390/cancers13205059.

[97] I. A. Prior, P. D. Lewis, and C. Mattos, "A comprehensive survey of Ras mutations in cancer.," *Cancer Res*, vol. 72, no. 10, pp. 2457–67, May 2012, doi: 10.1158/0008-5472.CAN-11-2612.

[98] S. Benvenuti *et al.*, "Oncogenic activation of the RAS/RAF signaling pathway impairs the response of metastatic colorectal cancers to anti-epidermal growth factor receptor antibody therapies.," *Cancer Res*, vol. 67, no. 6, pp. 2643–8, Mar. 2007, doi: 10.1158/0008-5472.CAN-06-4158.

[99] Z. Chen, J. C. Otto, M. O. Bergo, S. G. Young, and P. J. Casey, "The C-terminal polylysine region and methylation of K-Ras are critical for the interaction between K-Ras and microtubules.," *J Biol Chem*, vol. 275, no. 52, pp. 41251–7, Dec. 2000, doi: 10.1074/jbc.M006687200.

[100] A. V. Sorokin *et al.*, "Targeting RAS Mutant Colorectal Cancer with Dual Inhibition of MEK and CDK4/6," *Cancer Res*, vol. 82, no. 18, pp. 3335–3344, Sep. 2022, doi: 10.1158/0008-5472.CAN-22-0198.

[101] X. Sagaert, A. Vanstapel, and S. Verbeek, "Tumor Heterogeneity in Colorectal Cancer: What Do We Know So Far?," *Pathobiology*, vol. 85, no. 1–2, pp. 72–84, 2018, doi: 10.1159/000486721.

[102] N. Sasaki and H. Clevers, "Studying cellular heterogeneity and drug sensitivity in colorectal cancer using organoid technology.," *Curr Opin Genet Dev*, vol. 52, pp. 117–122, Oct. 2018, doi: 10.1016/j.gde.2018.09.001.

[103] M. Meng, K. Zhong, T. Jiang, Z. Liu, H. Y. Kwan, and T. Su, "The current understanding on the impact of KRAS on colorectal cancer," *Biomedicine & Pharmacotherapy*, vol. 140, p. 111717, Aug. 2021, doi: 10.1016/j.biopha.2021.111717.

[104] G. Zhu, L. Pei, H. Xia, Q. Tang, and F. Bi, "Role of oncogenic KRAS in the prognosis, diagnosis and treatment of colorectal cancer," *Mol Cancer*, vol. 20, no. 1, p. 143, Dec. 2021, doi: 10.1186/s12943-021-01441-4.

[105] J. C. Hunter, A. Manandhar, M. A. Carrasco, D. Gurbani, S. Gondi, and K. D. Westover, "Biochemical and Structural Analysis of Common Cancer-Associated KRAS Mutations," *Molecular Cancer Research*, vol. 13, no. 9, pp. 1325–1335, Sep. 2015, doi: 10.1158/1541-7786.MCR-15-0203.

[106] T. Pantsar, "The current understanding of KRAS protein structure and dynamics.," *Comput Struct Biotechnol J*, vol. 18, pp. 189–198, 2020, doi: 10.1016/j.csbj.2019.12.004.

[107] Z. Moghadamchargari *et al.*, "Intrinsic GTPase Activity of K-RAS Monitored by Native Mass Spectrometry.," *Biochemistry*, vol. 58, no. 31, pp. 3396–3405, Aug. 2019, doi: 10.1021/acs.biochem.9b00532.

[108] O. Sripichai and S. Fucharoen, "Genetic polymorphisms and implications for human diseases.," *J Med Assoc Thai*, vol. 90, no. 2, pp. 394–8, Feb. 2007.

[109] I. Tomlinson *et al.*, "A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21," *Nat Genet*, vol. 39, no. 8, pp. 984–988, Aug. 2007, doi: 10.1038/ng2085.

[110] J. R. Huyghe *et al.*, "Discovery of common and rare genetic risk variants for colorectal cancer," *Nat Genet*, vol. 51, no. 1, pp. 76–87, Jan. 2019, doi: 10.1038/s41588-018-0286-6.

[111] J. Ragoussis, "Genotyping Technologies for Genetic Research," *Annu Rev Genomics Hum Genet*, vol. 10, no. 1, pp. 117–133, Sep. 2009, doi: 10.1146/annurev-genom-082908-150116.

[112] F. S. Collins, L. D. Brooks, and A. Chakravarti, "A DNA Polymorphism Discovery Resource for Research on Human Genetic Variation: Table 1.," *Genome Res*, vol. 8, no. 12, pp. 1229–1231, Dec. 1998, doi: 10.1101/gr.8.12.1229.

[113] M. P. Costi and R. Zeillinger, "Drug resistance in ovarian cancer: Biomarkers and treatments," *Gynecol Oncol*, vol. 117, no. 2, pp. 149–151, May 2010, doi: 10.1016/j.ygyno.2010.03.006.

[114] L. K. Matukumalli, J. J. Grefenstette, D. L. Hyten, I.-Y. Choi, P. B. Cregan, and C. P. Van Tassell, "Application of machine learning in SNP discovery," *BMC Bioinformatics*, vol. 7, no. 1, p. 4, Dec. 2006, doi: 10.1186/1471-2105-7-4.

**JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT**

**PLAGIARISM VERIFICATION REPORT**

**Date: …… Type of Document (Tick):** PhD Thesis | M.Tech/M.Sc. Dissertation | B.Tech./B.Sc./BBA/Other

**Name:**_____ **Department:**_____ **Enrolment No** _____

**ORCID ID.** | | | | | | | | | | | | | | | | **SCOPUS ID.** _____

**Contact No.**_____ **E-mail.**_____

**Name of the Supervisor:** _____

**Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters):** _____

_____

_____

**UNDERTAKING**

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

- Total No. of Pages =
- Total No. of Preliminary pages =
- Total No. of pages accommodate bibliography/references =

**(Signature of Student)**

**FOR DEPARTMENT USE**

We have checked the thesis/report as per norms and found **Similarity Index** at ..... (%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

**(Signature of Guide/Supervisor)** **Signature of HOD**

**FOR LRC USE**

The above document was scanned for plagiarism check. The outcome of the same is reported below:

<table>
<tr><th>Copy Received on</th><th>Excluded</th><th>Similarity Index (%)</th><th colspan="2">Abstract & Chapters Details</th></tr>
<tr><td rowspan="2"></td><td rowspan="4">• All Preliminary Pages<br>• Bibliography/Images/Quotes<br>• 14 Words String</td><td rowspan="2"></td><td>Word Counts</td><td></td></tr>
<tr><td rowspan="2">Character Counts</td><td rowspan="2"></td></tr>
<tr><td>Report Generated on</td><td rowspan="2">Submission ID</td></tr>
<tr><td rowspan="2"></td><td>Page counts</td><td></td></tr>
<tr><td></td><td>File Size</td><td></td></tr>
</table>

**Checked by**
**Name & Signature** **Librarian**

……

**Please send your complete Thesis/Report in (PDF) & DOC (Word File) through your Supervisor/Guide at plagcheck.juit@gmail.com**

# Project report 1

ORIGINALITY REPORT

| 16% | 11% | 15% | % |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Negin Alizadehmohajer, Shahrzad Zahedifar, Ehsan Sohrabi, Sedighe Shaddel Basir et al. "Using In Silico Bioinformatics Algorithms for the Accurate Prediction of the Impact of Spike Protein Mutations on the Pathogenicity, Stability, and Functionality of the SARS-CoV-2 Virus and Analysis of Potential Therapeutic Targets", Biochemical Genetics, 2022<br>Publication | 2% |
| 2 | docksci.com<br>Internet Source | 2% |
| 3 | Akin Tekcan. "In Silico Analysis of FMR1 Gene Missense SNPs", Cell Biochemistry and Biophysics, 2016<br>Publication | 1% |
| 4 | www.ncbi.nlm.nih.gov<br>Internet Source | 1% |
| 5 | Thelen, Mathias, Cristina Razquin, Isabel Hernández, Ana Gorostidi, Raquel Sánchez-Valle, Sara Ortega-Cubero, Steffen Wolfsgruber, Dmitriy Drichel, Klaus Fliessbach, Tanja Duenkel, Marinella Damian, Stefanie Heilmann, Anja Slotosch, Martina Lennarz, Manuel Seijo-Martínez, Ramón Rene, Johannes Kornhuber, Oliver Peters, Christian Luckhaus, Holger Jahn, Michael Hüll, Eckart Rüther, Jens Wiltfang, Elena Lorenzo, Jordi Gascon, Alberto Lleó, Albert Lladó, Jaume | 1% |