Prediction of functionally significant Single Nucleotide Polymorphism in PTEN tumor suppressor gene

By

Manahil Ghazi

00000274507

Atta-ur-Rahman School of Applied Biosciences (ASAB),

National University of Sciences and Technology (NUST),

Islamabad, Pakistan

2020

Prediction of functionally significant Single Nucleotide Polymorphism in PTEN tumor suppressor gene

A thesis submitted in the partial fulfillment of the requirement for the degree of Masters of Science

In

Industrial Biotechnology

By

Manahil Ghazi

00000274507

Supervisor

Dr. Najam us Sahar Sadaf Zaidi

Atta-ur-Rahman School of Applied Biosciences (ASAB),

National University of Sciences and Technology (NUST),

Islamabad, Pakistan

2020

THESIS ACCEPTANCE CERTIFICATE

Certified that final contents and form of MS/MPhil thesis entitled **"Prediction of functionally significant Single Nucleotide Polymorphism in PTEN tumor suppressor gene"** written by Ms. Manahil Ghazi, Registration No. 00000274507, of ASAB has been vetted by undersigned, found complete in all respects as per NUST Status/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: _____

Supervisor: Dr. Najam us Sahar Sadaf Zaidi

Date: _____

Signature (HOD): _____

Date: _____

Signature (Dean/Principal): _____

Date: _____

CERTIFICATE FOR PLAGIARISM

It is to confirm that MS thesis entitled **"Prediction of functionally significant Single Nucleotide Polymorphism in PTEN tumor suppressor gene"** by Ms. Manahil Ghazi, Registration No. 00000274507 has been examined by me. I undertake that,

- Thesis has significant new work/knowledge as compared to already elsewhere. No sentence, table, equation, diagram, paragraph or section has been copied verbatim from previous work except when placed under quotation marks and duly referenced.
- The work presented is original and own work if the author i.e. there is no plagiarism. No idea, results or words of others have been presented as author's own work.
- There is no fabrication of data or results such that the research is not accurately represented in the records. The thesis has been checked using turnitin (a copy of the originality report attached and found within the limits as per HEC plagiarism policy and instruction issued from time to time.

_____

(Supervisor)

Dr. Najam us Sahar Sadaf Zaidi

Associate Professor

Industrial Biotechnology

ASAB, NUST

DECLARATION

I, Manahil Ghazi, declare that all work presented in this thesis is the result of my own work. Where information has been derived from other sources, I confirm that this has been mentioned in the thesis. The work here in was carried out while I was a post graduate student at Atta-ur-Rahman School of Applied Biosciences, NUST under the supervision of Dr. Najam us Sahar Sadaf Zaidi.

Manahil Ghazi

Master of Science in Industrial Biotechnology

Registration # 00000204376

Dedicated to

My Beloved Family

ACKNOWLEDGMENTS

MOST BENEFICENT AND THE MERCIFUL

Read: in the name of your lord who created man from a clot. Read: and your lord is the most bounteous who taught by the pen. Taught man that which he did not know.

Al Quran

As the time of my MS research comes to an end, I am grateful to Allah Almighty for providing me with a great learning experience in both the scientific area as well as on personal level. The completion of this research could not have been possible without the participation and assistance of so many people whose names may not all be enumerated but their contributions are sincerely appreciated and highly acknowledged.

I would like to express my sincere gratitude to **Dr. Najam us Sahar Sadaf Zaidi**, my supervisor for the continuous support, guidance and motivation throughout my research. This accomplishment would not have been possible without her. Also, I pay my gratitude to my thesis committee members, **Dr. Amjad Ali**, **Dr. Tahir Baig** and **Dr. Muhammad Jawad** for their intellectual support. I would extend my thankfulness to Principal **Dr. Husnain A. Junjua** and Head of Department **Dr. Saadia Andleeb**, for facilitating me with all the necessary support and equipment needed to complete my research.

Last but not the least; I would like to thank my parents, my sister Leenah Ghazi, my friend Muhammad Awais and my research fellows Hassan Farooq and Irum Hassan for all the guidance. I feel blessed to have these guys in my life.

I am at dearth to express my gratitude to my parents and siblings for their unconditional love, care and support. They are my strength and have always guided me in every step of life. May Allah bestow strength and contentment to all these splendid personalities! Ameen.

Manahil Ghazi

Table of Contents

Table of Figures

Table of Tables

Table of Graphs

List of Abbreviation

SNP	Single Nucleotide Polymorphism
nsSNP	Nonsynonymous Single Nucleotide Polymorphism
PHD-SNP	Predictor of human deleterious Single Nucleotide Polymorphism
PTEN	Phosphatase and tensin homolog
PI3K/Akt	Phosphatidylinositol-3-Kinase and Protein Kinase B.
SIFT	Sorting Intolerant from Tolerant
Polyphen-2	Prediction of functional effects of human nsSNPs
Panther	Protein Analysis through evolutionary relationships
MMAC1	Multiple advanced cancers1
PIP3	Phosphatidylinositol 3,4,5-trisphosphate
SNVs	Single nucleotide variants
PROVEAN	Protein Variation Effect Analyzer
LOH	Loss of heterozygosity

ABSTRACT

PTEN (phosphatase and tensin homolog) gene is located at chromosome 10 and is responsible for tumor suppression in a variety of tumors and cancers. PTEN is involved in signal transduction and its abnormal expression level has been associated with a number of diseases including tumor and different type of cancers particularly breast cancer. Studies have shown the correlation of tumor suppressor PTEN gene with brain metastasis in cancer patients as it plays an important role in oncogenic PI3K/Akt pathway and help the tumor cells to survive in brain microenvironment. The mutations in PTEN is the major cause of disturbance in its expression level. Single nucleotide polymorphism present in coding region of proteins (nsSNPs) has the potential to alter the expression level, primary structure as well as function of the protein. Hence, it becomes necessary to differentiate the potential harmful nsSNPs from the neutral ones. Bioinformatics tools are found to be very helping in finding deleterious SNPs. Most of SNPs in human body are common in a population. However, disease-causing variants are mostly private and typically rare and mostly occur in the protein coding region which consists of only 1% of the total genome. In the current study, 6 different bioinformatics tools including SIFT, Polyphen-2, Provean, PHD-SNP and SNPs & GO and Panther were used for the prediction of deleterious SNPs of PTEN involved in disease pathogenesis. By the use of these tools, 80 out of 133 SNPs of PTEN retrieved from dbSNP database were predicted as deleterious and pathogenic. Out of these 80 SNPs, 35 have already been reported in literature, 22 of them are under study and their clinical significance has been uncertain yet, while 23 of them are novel. These results provide a filtered data to explore the effect of uncharacterized SNPs and their association with the disease susceptibility and to design the target dependent drugs for therapeutics.

INTRODUCTION

Single nucleotide polymorphism is a genetic variation in which a single nucleotide in DNA is replaced by another nucleotide. For example, in SNPs the nucleotide cytosine (C) can be replaced by thymine (T), Adenine (A) or Guanine (G). It is also known as SNPs (pronounced "snips"). SNPs are found naturally in a population. At a SNP site, one allele comes from each parent in every individual genotype like AB, AA or BB. At the genomic locus of SNP, there is an appreciable frequency (>1%) of base alternation. SNPs are the common and the most frequently occurring variation in human genome and throughout the genome, it occurs once in every hundred base pairs. There are roughly 4 to 5 million DNA variants particularly SNPs in a human genome (B. Bessenyei *et. al.*,2004).

SNP can occur in a coding region and has no result in amino acid change. There is also a scenario where it occurs in coding and regulatory region and results in amino acid change and because of this, there is a change in gene expression. There is also a case where SNP occurs between the genes. In every case SNP occurring in different regions have a different effect on gene expression. Sometimes the variations are common and occur frequently but sometimes they are unique and occur in many individuals. Till now almost 100 million SNPs are found in different population throughout the world (Bao L, 2005).

Most common SNPs occur between the genes in DNA. These SNPs can be helpful as they can be used as biological markers. Scientists can use them to locate the genes which are associated with different diseases. Although most of the SNP don't affect human health yet some of these can cause genetic changes that can be deleterious. Sometimes they cause fatal and deadly diseases in humans like cancer. These SNPs

can be used for studying human health. These studies include finding out SNPs associated with different diseases such as heart diseases, different type of cancers and diabetes etc. (Bhatnager & Dang, 2018).

There is a research going on to check the response of SNPs to various drugs, and how much they are susceptible to the environmental factors like toxins. SNPs can also be used to track the inheritance of a particular disease that runs in families. SNPs play a direct role when they occur in the regulatory region of the gene, thereby damaging gene function and causing diseases (Capriotti E. C., 2006). SNPs can be bi-allelic, tri-allelic and even tetra-allelic, but the majority are found to be biallelic. SNPs are associated with certain traits; scientists are looking the DNA stretches near particular SNPs to identify a particular gene responsible for a trait (Erichsen & Chanock, 2004).

SNPs are also been used for personalized medication because of the fact that variation in the sequence of a DNA strand or a particular gene lead to a disease and respond to medicines, toxins, vaccine, chemicals, pathogens and other forms of agents (Li, *et al.*, 2015). Scientists are looking for those SNPs that are involved in genetic predisposition of a certain disease. They are comparing different regions of human genome especially the matched cohorts that are with and without a disease. (Katsonis, *et al.,* 2014).

SNP technologies play an important role in identifying genes that are involved in a disease in an individual. They have a medical importance as they are helping in understanding drug response in inter-individual variation. So, they are helping in making personalized genome-based diet by creating a link between genetic make-up and drug response of that individual (Carter,*et.al.,*2014). These personalized medicines are safer to use and are very much effective. SNPs are helping in understanding molecular basis of sequence evolution (Kim,*et al*., 2015).

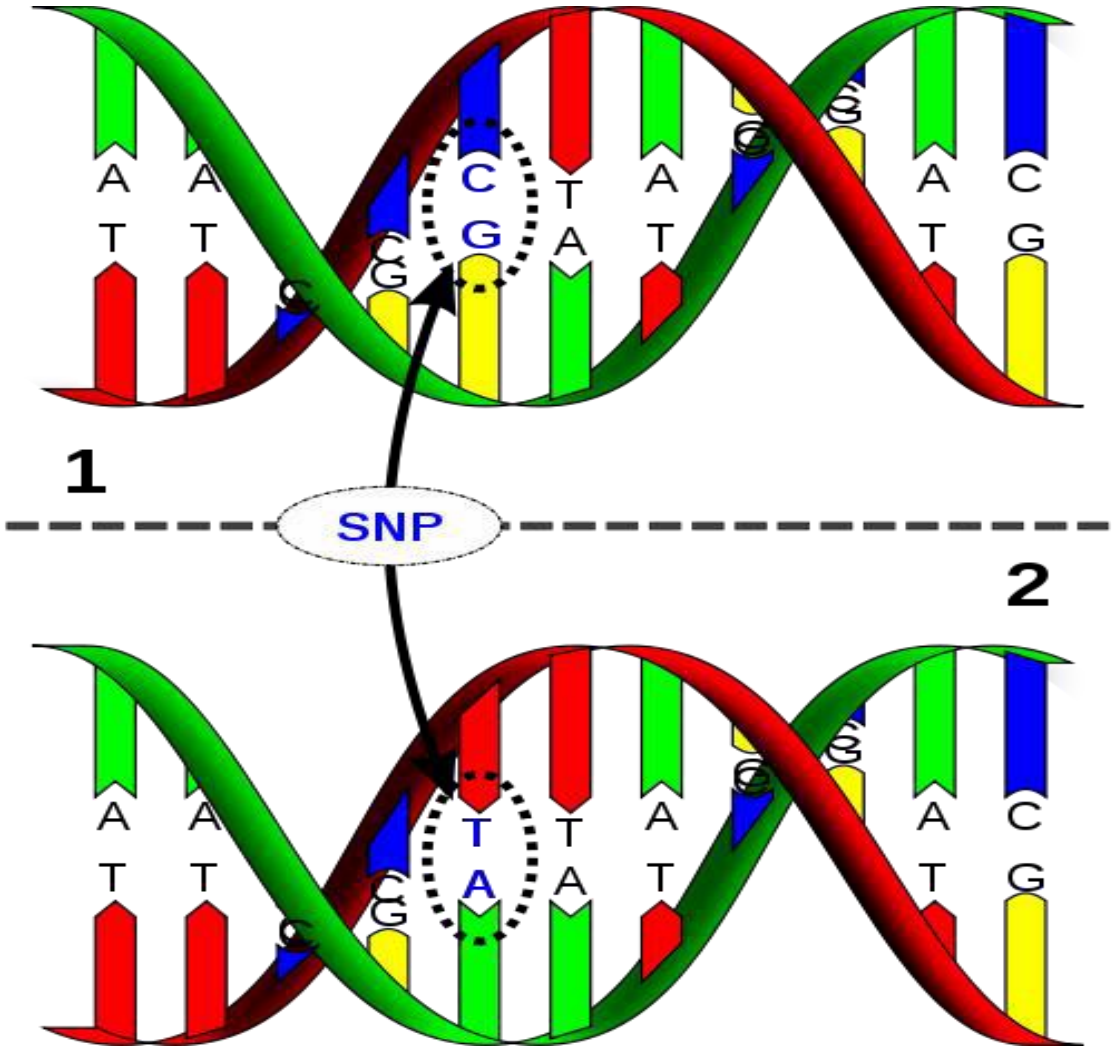


Figure 1.1: Single Nucleotide polymorphism in a DNA strand where thymine (T) paired with Adenine (A) is substituted by Cytosine (C), which is then paired with Guanine (G).

Throughout the gene, the rate of substitution of nucleotide, the site of its substitution and the type is not uniform. There is selection pressure and the mutations under selective pressure are more associated with evolution of human disease (Leslie & Longy, 2016). The other deleterious mutations that change the biological function of protein are not chosen by natural selection. But if substituted nucleotides are fixed by the natural selection in gene pool, they can be a source of several selection advantages as they are most of the time neutral. Some of them maybe deleterious and cause disease. So there is a chance that disease-associated SNPs and evolution are related to each other (Lee *et al.*, 2018).

The evolutionary conserved regions are more frequently associated with disease causing SNPs. Also, a specific allele that is common in a population of one ethnic or geographic group must be rare in another population and geographical region. The important step in evolution is the retention of variants caused by natural selection. The natural theory of evolution says that nature does not select those SNPs that are not present in the coding region of protein (Capriotti E. C., 2006). These SNPs comes under less selective pressure. This type of nucleotide substitution does not cause a change in amino acid and as a result no change in protein function. So most importantly amino acids are very much stable and are less mutable (Capriotti E. C., 2006).

SNPs that are present in coding region of human genes have serious phenotypic effects and are most of the times they are undergo natural selection process. The evolutionary data can be created by comparing the rate of change of non-synonymous to synonymous SNPs in different proteins of several species. It is a measure of selective pressure on amino acid mutations (Bao and Cui 2005).

The pattern of amino acid substitution are the basis of patterns of selection in the genes that codes for protein (Jiang and Zhao,2006). An important example of this is

the pattern of nonpathogenic mutations and the pattern of disease-associated genes distribution in humans. When we trace the mutation data of disease-associated genes as well as phylogenetic lineage of different species, it is found that disease-associated substitution is more common and are more frequent in evolutionarily conserved regions of DNA. That is they are more common in nonrandom distribution as compare to the positions that are undergoing variation (Subramanian and Kumar 2006; Miller and Kumar 2001).

For silent mutation, they are randomly distributed. And because of this random mutation, they have minimal effect on organism. Therefore, silent mutations are not subjected to natural selection (Yang *et al.,* 2017). Most frequent genetic alteration are observed in chromosome 10 at *PTEN* i.e. phosphatase and tensin homolog. PTEN is a tumor suppressor gene in a variety of tumors and breast cancer. PTEN is lost in different tumors and 60% of breast cancer patients. PTEN/PI3K pathway is the main pathway that plays its role in oncogenes (Mashhadi, *et al.,* 2014). PTEN gene is responsible for regulating this pathway and any mutation in PTEN gene leads to deregulation of this pathway and as a result of these mutations cancer and other diseases may result. It is reported that Phosphatase and tensin homolog (PTEN) is mutated in TGF-β or multiple advanced cancers1 (MMAC1) gene (Park *et al.,* 2019).

PTEN is present at chromosome 10q23 and it encodes 403 amino acid protein. This protein is linked with lipid and phosphoinositide 3-phosphatase activity. It regulates the level of phosphatidylinositol 3,4,5-trisphosphate (PIP3). It is cytosolic in general bur small amount of PTEN is also found in the plasma membrane (Li, *et al.,* 2015). PTEN is reported to be second anti-oncogene, so any mutation in it particularly Single nucleotide polymorphism lead to cancer or at least tumor. It is found to exhibit high frequency of mutation in different types of tumors in humans (Smith & Briggs, 2016).

It is also named as named MMAC1 gene or TEP1 gene. Its total length is 200 kb consisting of 8 introns and 9 exons. Single nucleotide polymorphism in PTEN gene can directly result in different diseases including tumors and cancers. It is found to have central role in the normal physiological as well as pathological process of tumor growth (Choi, *et al.,* 2012). Studies have shown its major role in differentiation, invasion as well as metastasis and prognosis of tumor. It is reported that *PTEN* is the only first protein to be discovered to perform activities of both lipid phosphatase and protein phosphatase (Haiman, *et al.*, 2006).

PTEN is known to reduce the level of PIP3, as a result mTOR/AKT signaling pathway is slowed down. So PTEN plays a pivotal role in signal transduction by negative regulation of PI3K signaling pathway. It is reported to be the most frequently mutated gene in human tumors and cancers. Its role in tumor or cancer cell growth, occurrence and development and survival has also been found. The role of PTEN SNPs in cancer research has a major impact in cancer biology. New discoveries related to it will help in targeted therapies for cancer intervention and prevention (Li, *et al*., 2015).

Bioinformatics tools are found to be very helping in finding deleterious SNPs. By whole genome sequencing, it was found that there are approximately 3.7 million single nucleotide variants (SNVs) in the genome of a human. The challenge for the geneticists is to identify the deleterious and disease-causing SNPs. Because most of SNVs in human body are common in a population. However, disease-causing variants are mostly private and typically rare. These also occur in the protein coding region which consists of only 1% of the total genome (Shastry, 2007). For filtering out the common variants of different genes in human body dbSNP is the common data base.

But for finding out diseased and deleterious SNPs some of bioinformatics tools are of much importance. Algorithms like SIFT can help in such search (AbdulAzeez & Borgio, 2016).

Although next-generation sequencing projects have created a wide sequence and variation data for genome, but the tools and softwares of bioinformatics are developed for the computational analysis and prediction of the effect of different sequence variants particularly SNPs on the functional activity of protein. These softwares help in finding out the variants for disease phenotype. The tools help to identify many classes of sequence variations at the level of nucleotide including frame shifts, insertions, deletions and other non-sense mutations that will result in causing diseases, tumors and cancer (Adzhubei I. A., *et al.,* 2010).

SIFT (Sorting intolerant from tolerant) is a bioinformatics tool that helps to predict that the single nucleotide polymorphism of a nucleotide in a gene can affect amino acid substitution and ultimately affect protein expression or not. It sorts out the data based on physical properties of the amino acid being substituted as well as sequence homology. SIFT predicts the SNPs in coding region have an effect on protein function or not (Flanagan, Patch, & Ellard, 2010).

SIFT was first introduced in 2001 but it has been updated many times. It has become a standard tool for finding out and characterization of missense SNPs. It also contains a new feature of insertion/deletion (indel). It also helps in finding frame shifting indels (Hu & Ng, 2013). This tool is used in human genetic research including cancer genetics and in the research for infectious diseases. This algorithm can be used in case studies of different diseases. In general, it can be used for naturally occurring non synonymous polymorphisms. The laboratory induced mutations can also be checked through this tool. It shows accuracy metrics on independent data sets (Forbes *et al*., 2015).

PROVEAN (Protein Variation Effect Analyzer) is another software. Its function is the prediction of the impact of an amino acid substitution on the biological activity of protein. This software helps in filtering out different sequence variants including insertions, deletions and other non-synonymous variants that are very important functionally (Choi & Chan, 2015). PROVEAN is a fast-computational tool comparable to SIFT and give results in the form of scores in pairwise sequence alignment. Its scoring scheme is best for separating deleterious, neutral and disease-causing SNPs from common polymorphisms. It is also helpful in separating deleterious variants from neutral variants (Bao L, 2005).

There are no study from Pakistan that report the SNPs of PTEN using many different software of Bioinformatics. This study will help in finding out the deleterious, disease causing, cancer and tumor causing SNPs of PTEN using different computational tools of bioinformatics. The importance of single nucleotide Polymorphism in PTEN will help in tracing the mutations that cause diseases and will then assist in prescribing early treatment methods to these patients.

Aims and Objectives

The aim is the prediction of damaging single nucleotide polymorphisms (SNPs) in PTEN gene and to see if they are involved in pathogenesis.

Present study was designed to achieve the following objectives:

- To predict the damaging/deleterious SNPs of PTEN gene by using 6 different tools of bioinformatics.
- To predict if a particular SNP in PTEN gene is associated with disease pathogenesis.

LITERATURE REVIEW

2.1 Single Nucleotide Polymorphism

Single Nucleotide polymorphism (SNP) is a sequence variation in DNA. It is caused by the replacement of any nucleotide adenine (A), thymine (T), cytosine (C), or guanine (G) in the genome with one another. As a result a difference arises between the paired chromosomes of an individual. It is the most common variation in genome in different people. SNPs are found in every 300 nucleotides of human genome. As there are a total of three billion nucleotides in the genome of humans so it is estimated that about 10 million SNPs are found in human genome (B. Bessenyei *et al*., 2004).

More than 99% of genome is identical between individuals. Also at least 1% of the population contain same single nucleotide variation. Researchers are focusing on SNPs to find out basis of genetic variation running in human race. Study has been performed to identify important SNPs involved in disease causing. Studies based on SNPs for finding regions of genome that are important in the development of disease. SNPs help in genome wide association studies. SNP can also be used to track an ancestry as they are shared by individual of common descent. The data was collected by the project of International HapMap including data related to location of particular SNPs in human genome (Katsonis, *et al.*, 2014).

2.2 Different types of SNPs

Polymorphism or mutation can be taking place by different ways. One of the forms is the substitution in which one nucleotide is substituted by the other. A situation also other in which the nucleotide itself is deleted from its place and is known as deletion. Sometimes there is an addition of one or more new nucleotide and is known as insertion (Choi, *et al.*, 2012).

Single nucleotide polymorphism can take place in both non-coding as well as coding region of human genome. It can also occur in the region between genes known as intergenic regions. Single nucleotide polymorphism that occur in coding region will not necessarily change the amino acid sequence of coding protein. This is because of the quality of degeneracy of genetic codon (Shastry, 2007). This type of SNP in which there is no change in amino acid take place is known as Synonymous mutations. Often these are called silent mutation. But if the SNP results in the change of amino acid and hence change in polypeptide sequence then this type of mutation is called Non-Synonymous mutations (B. Bessenyei, *et al.*, 2004).

There are two types of non-synonymous mutation. It can be missense or nonsense. In missense a change in nucleotide results in a change of amino acid and hence the change in polypeptide sequence which ultimately affect protein biological function. In nonsense mutation, the substitution of nucleotide result in the change of amino acid into stop codon. SNPs that are not present in protein-coding region sometimes also results in gene splicing (Goswami, 2015).

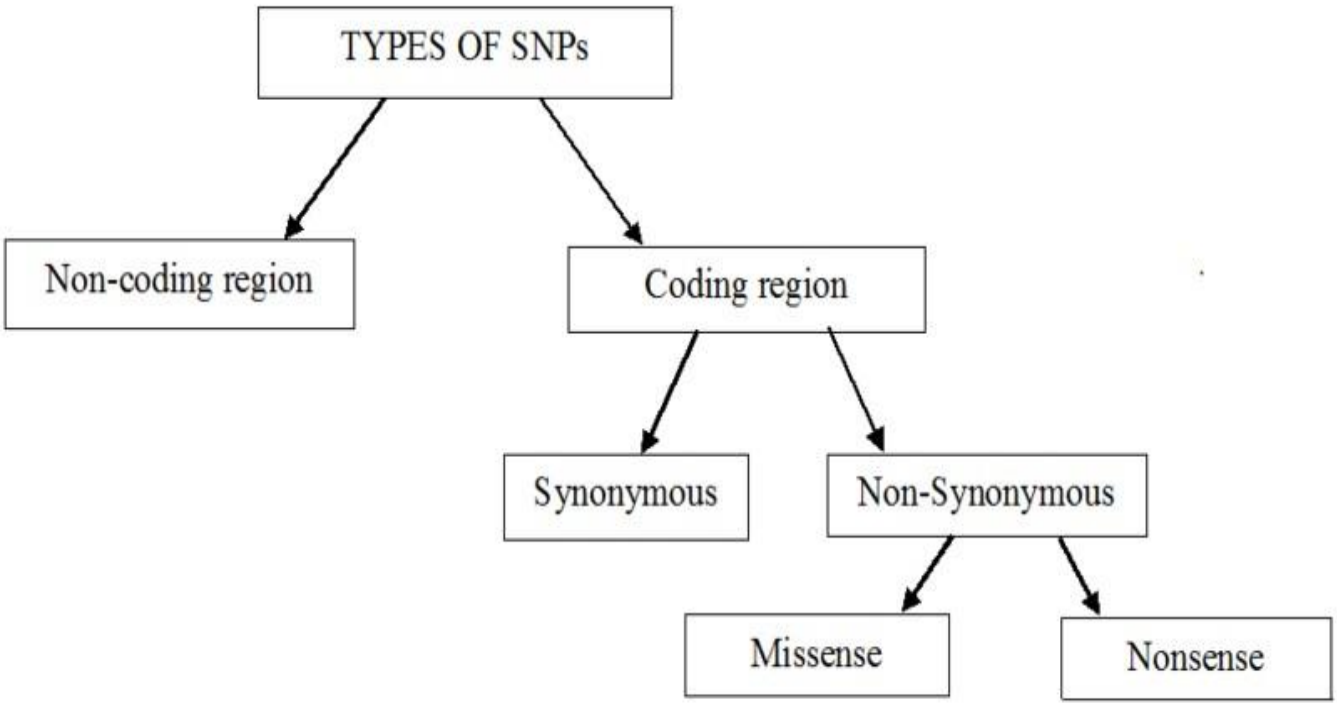


Figure 2.1. Types of Single Nucleotide Polymorphism

2.3 Role of SNPs in evolution

Genetic variations are of great importance when it comes to molecular evolution. The evolution is depending upon balance between mutation caused by environment and mutation caused by natural selection. Most of the time the natural selection maintains the position and type of amino acids among species (Carter, 2014). This is because of the reason that these amino acids are very much critical for the biological function of protein (Katsonis, et al., 2014). This is the only reason that certain amino acids are always highly conserved in a particular set of homologous genes. These conserved regions of amino acids are evolving only under a very strong selective pressure (Ramensky *et al*., 2002).

One of the benefits of natural selection is that most of the deleterious mutations that has a worse impact on biological function of protein are being eliminated effectively from the genetic pool of a population. This strong selection pressure contrary to deleterious SNPs is depending on the biological activity of those gene that code the important transcription and regularity proteins. Therefore, these are under strong selective pressure (Choi, *et al*., 2012).

2.4 Detection and analysis of DNA polymorphism

Single nucleotide polymorphism is the simplest form of DNA variation that occur in a human body. This can result in transition or transversion. It is reported that almost 50% of SNPs are found in the noncoding regions. 25% of SNPs are the cause of mutations that are missense (coding SNPs). And all the remaining 25% of SNPs result in silent mutations. Silent mutations are known as synonymous SNPs and they are not supported by natural selection (Halushka *et al*. 1999).

Non-Synonymous SNPs cause pathology and they are the result of natural selection. Both synonymous and nonsynonymous SNPs affect the activity of Promotor and the stability or conformation of the pre-mRNA. They also change protein's ability to bind with its inhibitor or the substrate. SNPs are responsible for changing the subcellular location of the protein. And the main thing is they are also in charge of susceptibility to disease and the deposition of medicinal drugs and ultimately evolution in genome (Kimchi-Sarfaty *et al.,* 2007).

To identify genetic variation particularly SNPs, many different organizations both public and private are doing remarkable work. They are developing techniques for high-throughput SNP genotyping. The research is going on for the past 20 years. And because of this work done, there is a large collection of available SNPs by the Human Genome Project (Shastry 2002, 2005).

2.5 Epigenetic Changes and Genetic Variation

The instability that is seen in many cancers are due to alterations in genomic sequence, which results in imbalance of alleles. Efforts are made to find out global image of this genomic imbalance using microsatellites and use of SNP markers. It can also be detected by allelotyping of cancers (Capriotti E. C., 2006). The global pattern of this genomic imbalance is also find out on the regions where allelic balance leads to cancer (Lindblad-Toh et al, 2000). The most common allelic imbalance is loss of heterozygosity (LOH) on one or many more regions of the chromosome. Research is being carried out to see the different sites of LOH for the existence of tumor-suppressor genes. Many studies related to lung cancer and prostate cancer have found the allelic imbalance at many sites including SNP for the analysis of LOH (Naidu & Y Suneetha, 2016).

The pattern of loss of heterozygosity can be found out by the analysis of SNPs. These studies have shown the diagnostic and prognostic implications of cancer and tumors because these cancers have a very specific LOH pattern and are related to the expression array profiles for the identification of variants (Forbe *et al*., 2015).

Table 2.1 Disease Associated with Single Nuleotide Polymorphims

Disease	Gene	Disease	Gene
Asthma	EDN1 and NOS1 Chemokine	Lung cancer	MMP1 p53
Arrhythmia	KCN1	Myocardial Infraction	TSP PCS
Blood pressure	TAF1	Migraine	IR
Biliary cirrhosis	MBL	Obesity	PAI1
Bipolar affective disorder	HRT 3A	Ossification	Npps
Colorectal cancer	Cyclin D1	Oxalate stone	E-Cad
Crohn's Disease	MDR1	POAG	Myocilin
Dyslipidemia	Lipase	Rheumatoid arthritis	MIF
Eating disorder	Melanocortin	Systemic sclerosis	Fibrillin1
Esophagel adenocarcinoma	Cyclin D1	Severe sepsis	TNF-α
Hyperbilirubinemia	UGT1A1	Type II diabetes	Syntaxin1A
Idiopathic arthritis	MIF	Ulcerative colitis	MDR1
Idiopathic PD and FTD	Tau	Urinary bladder cancer	Cyclin D1
Knee and hip osteoarthritis	Collagen	Autism	CNP

2.6 SNPs in PTEN are responsible for Cancer

Single nucleotide polymorphism that occur in PTEN gene is responsible for deadly diseases including cancers and tumors. Studies have found that Single nucleotide polymorphism in PTEN is responsible for breast, ovarian, prostrate and liver cancer (Mohammed Arifeen, 2018). PTEN genes have been identified as the major breast cancer-susceptibility genes (Han and Brastianos, 2017).

PTEN plays a major and crucial role in the suppression of the tumor. It regulates negatively an oncogenic pathway known as phosphatidylinositol 3-kinase (PI3K). As a result, PTEN is helping in maintaining the integrity of genome (Yang *et al.*, 2017). This whole pathway starts with binding of many growth factors to the tyrosine receptor kinases. This binding helps in initiating the receptor complex. As a result, the receptor complex activates another pathway known as PI3K pathway. This pathway causes the phosphorylation of PIP2 into PIP3 which then start phosphorylating another pathway called Akt pathway with the help of PDKI. Phosphorylation of Akt activates many of other pathways where it acts as a substrate and its essential target is mTOR pathway. This mTOR pathway is the cancerous pathway that causes in excess synthesis of protein, decrease in the apoptosis as well as it slows down cell cycle proliferation and progression. The main point is the de-phosphorylation of PIP3 back into PIP2 which stops the mTOR pathway. The loss of PTEN is linked with Akt/mTOR pathway activation, but this does not initiate the tumor, it only helps in metastasizing the tumor to distant organs (Phin *et al*, 2013).

PTEN/PI3K pathway is the main oncogenic pathway that plays the crucial role in breast cancer. This pathway is controlled by PTEN gene. This pathway is found to be disturbed in breast cancer because PTEN expression is lowered in almost 25% of

cancers. The low expression of PTEN gene is due to mutations particularly Single nucleotide polymorphism in the gene sequence. The low expression can also be due to chromosome copy number loss, chromosome aberrations or due to epigenetic silencing (Simpson & Parsons., 2001).

2.7 PTEN Gene and its Role in Cancer Brain Metastasis

Brain metastasis is the most important complication that occur during tumors and cancers. The rate of metastasis is totally based on the type of cancer or tumor. Almost 15% of epithelial cancers brain metastasization (Nelson, & Von Deimling., 1997). Studies have shown that the reason of metastasis in some cancers is the low expression of PTEN gene. Some of the times, chromosomal aberrations is the reason of suppression of PTEN gene. PTEN gene is known to be a key tumor suppressor gene that is found in most of primary glioblastomas. With the help of EGFR amplification, studies have found the most frequent alterations in primary glioblastomas is due to loss of PTEN gene expression (Ohgaki *et al.*, 2009). PTEN gene is involved in the PI3K kinase pathway and any mutation that results in low expression of PTEN play a major role in the growth of malignant tumor cells inside brain and therefore they are major target for the therapeutic intervention (Saal *et al*., 2005).

There was a study conducted by Wikman H. which has confirmed there is a loss of chromosome 10q in brain metastases. The study was based on data from a large number of populations. In the people of that region, the potential target gene was found to be the tumor-suppressor gene PTEN. This also plays a crucial role in developing as well as progressing tumors (Nelson, & Von Deimling., 1997).

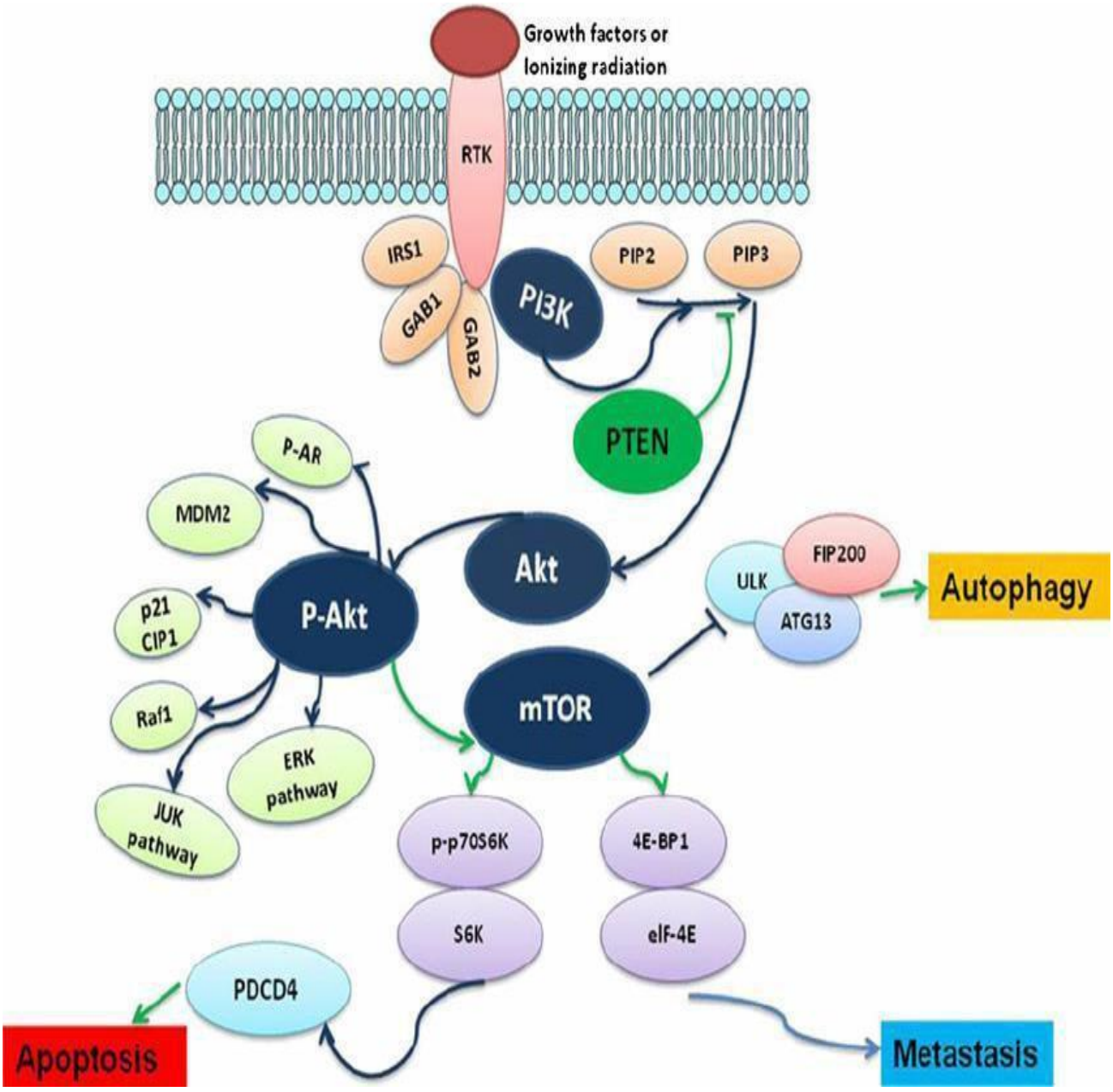


Figure 2.2 Overview of PI3K/AKT/PTEN/mTOR Signaling Pathway in the Regulation of Cancer Metastasis, Apoptosis and Autophagy. (Chang et al., 2014)

2.8 Metastasis and Proliferation in Brain

Polymorphism is one of the causes of brain metastasis. When tumor cells enter in the environment of brain, it results in the metastasization of brain. The chance of brain metastasis in cancer patients is 10% - 15%. The rate of their survival these cancer patients is also very low. Metastasis of brain is the major concern that contribute to morbidity as well as mortality in metastatic cancer patients. From initial tumor site, the vessels starts to distribute. At the secondary site, it first metasitize through haematogenous spread, which ultimately lead to brain metasitization (Azim, Abdel-Malek and Kassem, 2018).

There is an important role of cells present in the brain microenvironment in promoting the colonization and proliferation of metastatic cells (Azim, Abdel-Malek and Kassem, 2018). These cells also help in tumor growth and tumor cell survival (Lin, Bellon and Winer, 2004). In certain cases, brain in the secondary organ of tumor metastasis and for extravasation cancer cells needs the interaction of microvascular endothelium of brain with tumor cells. These brain microvascular endothelium is also involved in blood-brain barrier (Wilhelm *et al.*, 2014).

The tumor cells form metastases in brain with the help of mesenchymal-epithelial transition for the completion of metastatic cascade (Lorger, 2012). The research and analysis single nucleotide polymorphism studies therefore have proved to be a major goal for researchers in finding out metastasis and proliferation of tumor and cancer cells in brain. SNP analysis has also helped in identification of all mechanisms is extremely important that are involved in cancer (Gunasinghe *et al.*, 2012).

2.9 Different tools to check SNPs

For many years, research is being done on knowing the role of genetic changes that results in causing a disease. This effort by scientists and researchers results in the development of many data bases, tools and web sources that contain the information about single nucleotide polymorphisms (SNPs) and the cases in which it result in causing a genetic disease. These software, tools and online resources are designed on the basis of the genomic context. These focus on human annotations and also some of the tools are designed for SNP data of model organisms like fruit fly, chimpanzee and mice. There are many data bases that we can use to find out SNP and disease mutation related data (AbdulAzeez & Borgio, 2016).

2.9.1 PANTHER (protein analysis through evolutionary relationships):

One of the software is PANTHER (protein analysis through evolutionary relationships). It is a classification-based SNP tool that is designed in such a way that it has curated the biological data of different protein and gene families and their relative families based on their function. PANTHER is one of the important category of Gene Ontology Reference Genome Project that is used for the classification of proteins and genes (Adzhubei I. A., et al., 2010).

2.9.2 dbSNP (Single Nucleotide Polymorphism database)

Another database is dbSNP (Single Nucleotide Polymorphism database) includes all the sequence and structure data of SNPs of different genes. Human Gene Mutation Database (HGMD) involves genotype-phenotype based data (Arshad, Bhatti, & John, 2018).

2.9.3 PolyPhen-2 (Polymorphism phenotyping 2):

Another tool that predicts the influence of substitution of amino acid on the structure and function of protein is PolyPhen-2 (Polymorphism phenotyping 2). It predicts by using physical and comparative analysis of SNP and protein. There is a new addition of databases for resequence of the mutation data and sequencing of somatic mutations in different cancers. These databases include SeattleSNPs project. This tool has provided a wealth of human genetic variation data (Adzhubei, Jordan, & Sunyaev, 2013).

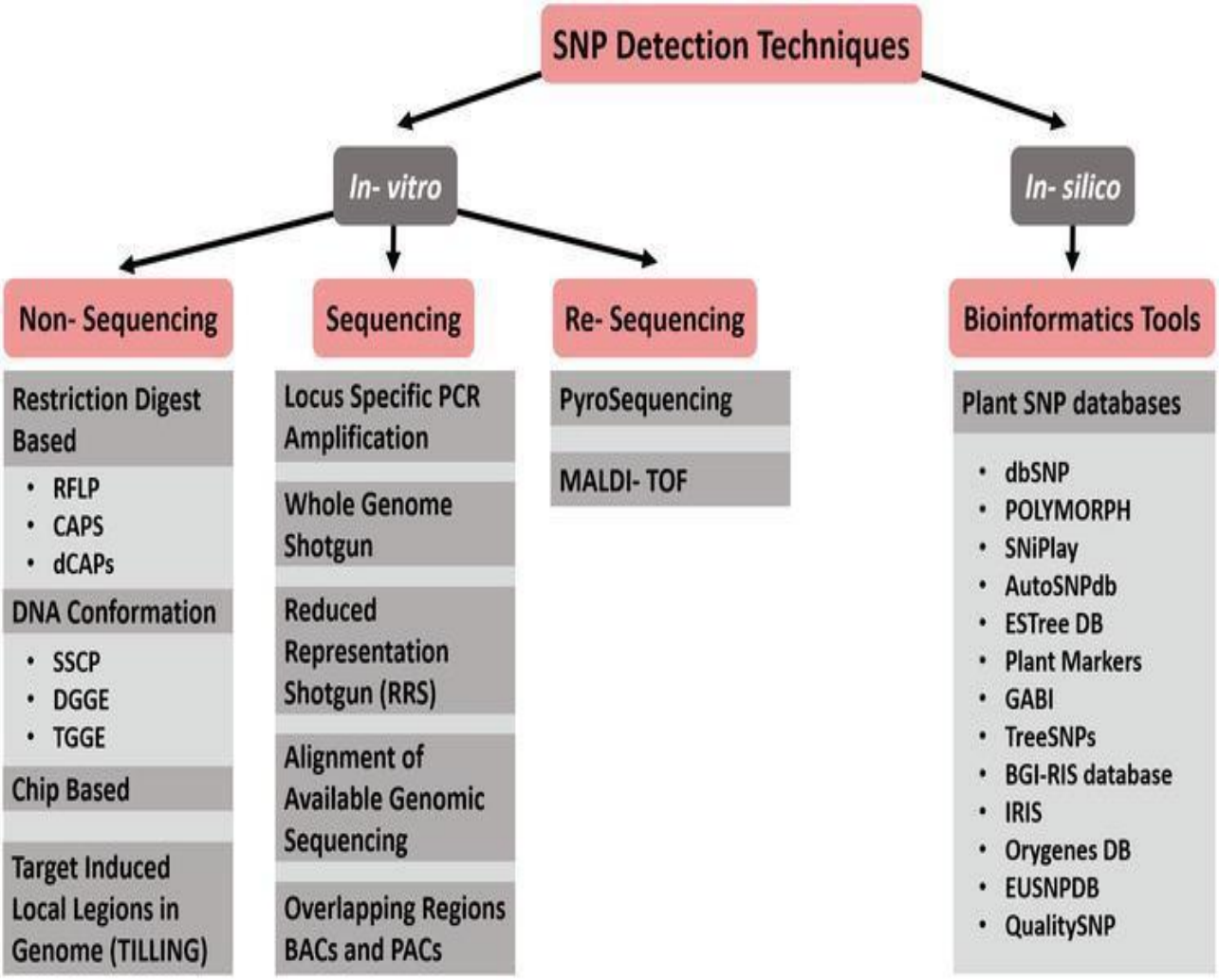


Figure 2.3 Different techniques for Single Nucleotide Polymorphism (SNP) detection

These bioinformatics software contain many different categories. They are designed in such a way that most of the data regarding SNP is curated on de facto central SNP database. The data include whether the SNP is responsible for causing disease or not. It includes if these single points mutations can cause change in phenotype and molecular function of protein or not. These are helping in facing the challenges of identifying functional variants. For narrowing a region for SNP prioritization, the first step is the identification of the candidate disease gene. The next step is the identification of candidate causal SNPs. Other important databases include OMIM (Online Mendelian Inheritance in Man), and PharmGKB (the Pharmacogenetics Knowledge Base) (Miosge, et al., 2015).

2.9.4 PhD-SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms):

Another important software is PhD-SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms). This tool give results by SVM-Sequence classifier on the basis of a decision tree. SVM-Profile is made on the basis of sequence profile information. This tool is used for finding the deleterious effects made by single nucleotide and ultimately single amino acid substitution (Capriotti E. C., 2006). It does that by seeing if amino acid is conserved at a position of the sequence of interest.

Here the results are based on a versatile score based on alignment. This shows that frame shifts and other non-sense mutations have negative and bad effect on the function of protein. Also we can find out if this tool is directly applicable to find out insertions, deletions and other multiple amino acid substitution. The main approach is that its alignment-based score find out any change in the sequence or similarity of our query sequence. And this help in predicting the sequence of protein homolog before

and after the variation in amino acid of our query sequence (Mooney,, Krishnan,, & Evani, 2010).

2.9.5 SNPs & GO:

This is a well-known tool for the prediction of variations that are responsible of causing a disease. The variation occurs in the sequence and structure of protein. It is an algorithm that has been developed by Prof. Rita Casadio of Laboratory of Biocomputing at University of Bologna. SNPs & GO tool is hosted by a server of BioFolD Unit. It is the most accurate method that take the protein sequence as an input and then find out the variation in the inserted sequence of protein is disease causing or not. It do that by exploring the functional annotation of protein. This tool is made by curating information such as evolutionary sequence and function of protein. There is also a SNPs&GO3d latest version in which there are protein 3D structures. Its results are more accurate and reliable than the previous version (Kaman, *et al.*, 2019).

MATERIALS AND METHODS

Current study consists of three main portions. It includes prediction of deleterious and damaging SNPs of PTEN gene. Secondly, predicting the stability of PTEN protein containing SNPs. and the most important part is the prediction of disease causing SNPs of PTEN gene. 133 of total SNPs of PTEN gene from CHR_HG2334_PATCH:87881582 position were selected. 8 different tools of bioinformatics were used for the prediction. These include SIFT, Polyphen-2, Panther, PHD-SNP, SNP and GO, Provean, MUpro and I-Mutant Suits.

3.1 PTEN SNPs selection and data retrieval

133 of total SNPs of *PTEN* gene from CHR_HG2334_PATCH:87881582 position were selected from dbSNP. The single nucleotide polymorphism database is developed by NCBI (National center for biotechnology). It contains a collection of SNPs of all types from different genes. It has a range of missense variants, insertions, deletions, short tandem repeats (STRs), microsatellite.

Here is the list of selected missense SNPs of PTEN gene obtained from dbSNP.

Table 3.1 Selected missense SNPs of PTEN gene

SR.NO.	SNP ID	SR.NO.	SNP ID	SR.NO.	SNP ID
1	rs121909233	46	rs1199933120	91	rs1160370526
2	rs1554890392	47	rs1323866910	92	rs1379620359
3	rs1064795967	48	rs1310956774	93	rs761148721
4	rs876660420	49	rs917927904	94	rs1064794096
5	rs1554890398	50	rs1326819131	95	rs1029309553
6	rs876661244	51	rs587781957	96	rs1316552000
7	rs786201995	52	rs1186471064	97	rs1264797578
8	rs797044910	53	rs1415012784	98	rs1488817615
9	rs398123326	54	rs540063602	99	rs1288670852
10	rs786204912	55	rs930485618	100	rs786204946
11	rs786201506	56	rs1045148218	101	rs1353528657
12	rs786204853	57	rs1064796078	102	rs1252194901
13	rs1012291977	58	rs786204910	103	rs1483165357
14	rs1403295709	59	rs904005027	104	rs572922017
15	rs1043104196	60	rs1221231481	105	rs1417666224
16	rs886047388	61	rs1048912442	106	rs1163742361
17	rs886047389	62	rs575260016	107	rs1394574548
18	rs904599717	63	rs1320222638	108	rs1328640232
19	rs1178630827	64	rs1238691983	109	rs1554898209
20	rs1219491699	65	rs937928726	110	rs376610243
21	rs1451210152	66	rs535471450	111	rs761265816
22	rs886047390	67	rs369849061	112	rs750098228
23	rs1564800954	68	rs1395659213	113	rs1085308044
24	rs1564800966	69	rs1031706678	114	rs889249323
25	rs1285259669	70	rs781647403	115	rs1064793744
26	rs1287849927	71	rs1020319146	116	rs1414611362
27	rs1355651458	72	rs1029342144	117	rs1064793241
28	rs1239421959	73	rs1564801672	118	rs755295390
29	rs1237944553	74	rs1239105602	119	rs1085308047
30	rs1433679388	75	rs1311095241	120	rs753142719
31	rs1564801018	76	rs1219944615	121	rs398123324

32	rs1318588388	77	rs985226639	122	rs753142720
33	rs1359067345	78	rs1064796886	123	rs11202592
34	rs1242894494	79	rs552470098	124	rs749356977
35	rs1564801070	80	rs1448017313	125	rs1554890322
36	rs1196170476	81	rs1282440343	126	rs1554890324
37	rs550595518	82	rs1337121620	127	rs1478570799
38	rs1215644762	83	rs1554825172	128	rs1564801650
39	rs1441812814	84	rs962892260	129	rs1060500109
40	rs1001571902	85	rs911548055	130	rs1554890337
41	rs1554898159	86	rs1258029214	131	rs572685299
42	rs2943772	87	rs974009386	132	rs1564801689
43	rs1401532030	88	rs918537651	133	rs1554890340
44	rs1435766567	89	rs929911235		
45	rs886047393	90	rs1199933120		

3.2 SNPs data retrieval from Ensemble

The nucleotide sequence above and below the flanking site of the region containing our SNP were downloaded from ensemble. Ensemble is central resource for researchers, microbiologists and geneticist for studying and obtaining data related to genes, genomes, SNPs flanking sites, genetic maps of humans and other vertebrates and model organism. It is a genome browser for obtaining genomic data.

3.3 PTEN SNP DNA translation into protein using ExPASy

The nucleotide sequence containing our SNPs was translated into amino acid sequence to check if there is a change in amino acid before and after single nucleotide polymorphism. ExPaSy was used for this purpose. It converts DNA to RNA to protein. It accepts the DNA sequence and convert it into codon form of polypeptide chain. From this we can see if there is a change in amino acid with or without the presence of SNP.

3.4 Applying of different Bioinformatic tools

For the prediction of deleterious, damaging SNPs of PTEN the tools used were SIFT, PolyPhen-2, Panther and Provean. For the prediction of damaging effects of SNPs on PTEN protein stability I-Mutant suits and MuPro softwares were used. Finally we used PHD-SNP, SNP & Go and I-Mutant Suits were used.

3.4.1 SIFT (Sorting Intolerant from Tolerant)

SIFT (sorting intolerant from tolerant) algorithm was used for the prediction of deleterious effects in all of the total 133 SNPs of *PTEN* gene. SIFT tells that if the missense mutation is tolerable or not. It gives a lift of amino acid that are tolerable and deleterious at the position of variation. It also gives a score. SIFT score ranges

from 0.0 to 1.0. The score 0.0 is deleterious for SNPs while moving towards 1.0 it become tolerated. In general a sore from 0.0 to 0.5 is considered as deleterious while from 0.6 to 1.0, it is considered as tolerated.

3.4.2 PolyPhen-2 (Polymorphism phenotyping 2)

PolyPhen-2 (Polymorphism phenotyping) was used for the prediction of deleterious and damaging SNPs from all of the selected 133 missense SNPs of *PTEN* gene. It gave the results in the 3 forms: benign, probably damaging and possibly damaging. It is a structure-based and sequence-based predicting software. It works by comparing wild-type (ancestral, normal) allele with mutant (diseased causing). Benign prediction is considered to be not harmful. Possibly damaging prediction has 100% chance of being deleterious while probably damaging prediction has 75% chance of being deleterious.

3.4.3 PANTHER (protein analysis through evolutionary relationships)

PANTHER (protein analysis through evolutionary relationships) was used for the prediction of deleterious and damaging SNPs from all of the total 133 SNPs of *PTEN* gene. It only predicts those SNPs that are damaging or probably damaging. Sometimes it gives result as invalid substitution for those SNPs which it don't consider to be damaging.

3.4.4 PhD-SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms)

PhD-SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms) was used for the prediction of deleterious and damaging SNPs from all of the selected 133 missense SNPs of *PTEN* gene. It tells that the particular SNP is disease causing or neutral. The output comes in the form of a table. The table has the wild type residue,

the new residue, mutated position and the prediction that if the mutation is disease causing or neutral. There is an RI value for each variation which is evaluated on the basis of support vector machine.

3.4.5 PROVEAN (Protein Variation Effect Analyzer)

PROVEAN (Protein Variation Effect Analyzer) was used for the prediction of deleterious and damaging SNPs from all of the selected 133 missense SNPs of *PTEN* gene. The protein sequence was inserted in the input tab and it searches its homolog using Blast. Based on the homolog blast results, it gave us a particular score to each variation. It gave the prediction in two forms: deleterious and Neutral.

3.4.6 SNPs and GO

SNPs and GO was used for the prediction of deleterious and damaging SNPs from all of the selected 133 missense SNPs of *PTEN* gene. It tells that the particular SNP is disease causing or neutral. This tool has data of 38460 of all important mutations and it works on the cross validation procedure upon insertion of data. It gives results in the form of RI value and the probability score.

3.5 Result Analysis

The data of prediction and their scores were collected through the use of the above mentioned 8 softwares in the form of tables and pie charts. After obtaining data from all these softwares a comparison table and pie chart was made. Then it was checked which SNPS are the most damaging one by the prediction of all these software.

Results

4.1 SIFT (Sorting Intolerant from Tolerant)

SIFT (sorting intolerant from tolerant) algorithm was used for the prediction of deleterious effects in all of the total 133 SNPs of *PTEN* gene. SIFT score ranges from 0.0 to 1.0 from deleterious to be tolerated. Out of 133 SNP of *PTEN* gene checked through the SIFT, 80 showed the results in the range of deleterious and deleterious - low confidence. And 53 showed the results in the range of tolerant and tolerant- low confidence.

The results showed that 62% of our selected PTEN SNPs are deleterious and they can have a deleterious effect. They can cause a fatal disease, tumor or cancer. They can decrease the protein's stability and make changes in it by disturbing the polypeptide chain of protein. Only 38% SNPs are found to be tolerant to this mutation and are not involved in any kind of disease.

The SNP ID, Nucleotide change, type of variation, SIFT (sorting intolerant from tolerant) prediction and score are mentioned in Table no. 4.1. The graph 4.1 is plotted to show the percentage of tolerant and deleterious SNPs. It shows that maximum 62% SNPs are deleterious and other 38% are tolerant.

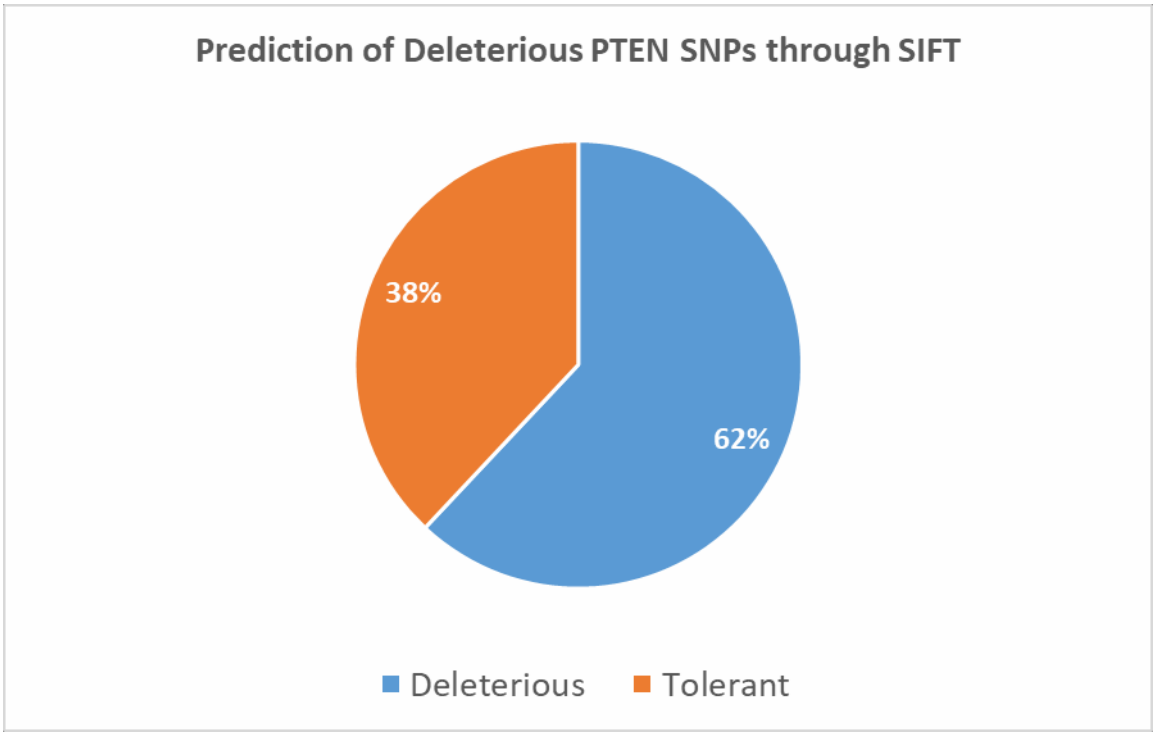
Table 4.1 Prediction of Deleterious PTEN SNPs through SIFT

SR. NO.	SNP ID	ALLELE	AMINO ACID Change	TYPE OF SNP	SIFT prediction	SIFT Score
1	rs886047388	G/A	R20K	missense variant	tolerated	0.92
2	rs1451210152	G/C	R30S	missense variant	tolerated	0.38
3	rs1287849927	G/C	V38L	missense variant	tolerated	0.26
4	rs1318588388	G/T	R48L	missense variant	tolerated	0.07
5	rs1020319146	C/G	P71A	missense variant	tolerated	0.08
6	rs1282440343	T/C	S85P	missense variant	tolerated	0.08
7	rs1337121620	C/T	S85L	missense variant	tolerated	0.42
8	rs911548055	T/C	Y88H	missense variant	tolerated	0.27
9	rs1258029214	A/G	Y88C	missense variant	tolerated	0.09
10	rs886047393	C/T	P92S	missense variant	tolerated	0.61
11	rs1199933120	C/T	P92H	missense variant	tolerated	0.34
12	rs1048912442	A/G	S116G	missense variant	tolerated	0.27
13	rs575260016	C/T	R119G	missense variant	tolerated	0.16
14	rs761148721	G/C	E134D	missense variant	tolerated	0.43
15	rs1029309553	G/T	A137S	missense variant	tolerated	0.51
16	rs1316552000	C/T	A137V	missense variant	tolerated	0.37
17	rs1264797578	G/T	A139S	missense variant	tolerated	0.37
18	rs1483165357	A/C	K152Q	missense variant	tolerated	0.1
19	rs1417666224	C/T	H155Y	missense variant	tolerated	0.06
20	rs1163742361	A/G	Q156R	missense variant	tolerated	0.23
21	rs376610243	C/T	L162F	missense variant	tolerated	0.05
22	rs750098228	T/C	F165Y	missense variant	tolerated	0.53
23	rs889249323	T/C	F165S	missense variant	tolerated	0.49
24	rs1064793744	T/C	F167L	missense variant	tolerated	1
25	rs755295390	C/T	H169Y	missense variant	tolerated	0.64
26	rs753142719	A/G	H169R	missense variant	tolerated	0.18
27	rs749356977	C/G	L171V	missense variant	tolerated	0.19
28	rs1564801650	A/G	T175A	missense variant	tolerated	0.35
29	rs1060500109	C/T	A176V	missense variant	tolerated	0.22
30	rs1554890337	A/G	I178V	missense variant	tolerated	0.11
31	rs1564801672	G/C	E180D	missense variant	tolerated	0.42
32	rs1414611362	C/G	N185K	missense variant	tolerated	0.43
33	rs540063602	A/G	D192G	missense variant	tolerated	0.41

34	rs1064795967	T/G	D192E	missense variant	tolerated	0.08
35	rs1554890324	C/A	P172T	missense variant	tolerated	0.29
36	rs754821870	A/G	Q569R	missense variant	tolerated	0.06
37	rs876661021	A/G	H570R	missense variant	tolerated	0.05
38	rs587782345	A/T	N529D	missense variant	tolerated	0.25
39	rs1564570356	T/A	N529K	missense variant	tolerated	0.2
40	rs1554826026	A/G	Y519C	missense variant	tolerated	0.06
41	rs1554826034	G/T	E525Q	missense variant	tolerated	0.07
42	rs878853932	A/G	E526G	missense variant	tolerated	0.06
43	rs370064195	G/C/T	D470H	missense variant	tolerated	0.35
44	rs1554825550	A/G	E472G	missense variant	tolerated	0.16
45	rs758644748	G/A	D474N	missense variant	tolerated	0.55
46	rs587780007	G/T	S478I	missense variant	tolerated	0.21
47	rs786203858	A/G	I479V	missense variant	tolerated	0.19
48	rs786201507	G/A	R481H	missense variant	tolerated	0.16
49	rs1564568473	A/G	D483G	missense variant	tolerated	0.15
50	rs863224667	A/G	D485G	missense variant	tolerated	0.15
51	rs1554825577	C/A	L489I	missense variant	tolerated	0.38
52	rs1554898209	G/A	G329R	missense variant	tolerated	0.19
53	rs1554900633	A/G	M371V	missense variant	tolerated	0.62
54	rs2943772	C/G	S65C	missense variant	deleterious	0
55	rs985226639	A/T	E77G	missense variant	deleterious	0.01
56	rs552470098	C/T	A79V	missense variant	deleterious	0.01
57	rs1395659213	A/G	Y123C	missense variant	deleterious	0.02
58	rs786204853	G/T	L198F	missense variant	deleterious	0
59	rs11202592	A/T	R170W	missense variant	deleterious	0
60	rs1238691983	C/T	R119C	missense variant	deleterious	0.03
61	rs398123324	A/G	R187G	missense variant	deleterious	0.01
62	rs1064794096	G/T	R188I	missense variant	deleterious	0
63	rs1064796078	A/T	R188S	missense variant	deleterious	0
64	rs786204910	T/G	Y189D	missense variant	deleterious	0
65	rs786201995	T/G	L196V	missense variant	deleterious	0
66	rs797044910	G/T	D197Y	missense variant	deleterious	0
67	rs1320222638	A/G	D197G	missense variant	deleterious	0
68	rs587781957	A/C	N185T	missense variant	deleterious	0.02
69	rs1114167648	A/C	H445P	missense variant	deleterious	0.02
70	rs786204875	G/T	W447L	missense variant	deleterious	0
71	rs587782350	C/T	P419L	missense variant	deleterious	0

72	rs1057519368	T/G	L420S	missense variant	deleterious	0
73	rs1114167664	G/T	G424V	missense variant	deleterious	0
74	rs121909239	A/T	D425V	missense variant	deleterious	0
75	rs1085308051	A/G	E330G	missense variant	deleterious	0
76	rs786202688	A/G	R332G	missense variant	deleterious	0.01
77	rs1114167673	G/T	R332M	missense variant	deleterious	0
78	rs587782603	G/T	G338R	missense variant	deleterious	0
79	rs1554900515	T/A	V339E	missense variant	deleterious	0.04
80	rs397514559	C/A	T340N	missense variant	deleterious	0.01
81	rs1554900534	A/G	S343R	missense variant	deleterious	0
82	rs121909221	T/A	S343R	missense variant	deleterious	0
83	rs121913293	C/T	R346S	missense variant	deleterious	0.03
84	rs121913294	G/T	R346P	missense variant	deleterious	0
85	rs863224666	A/G	H314P	missense variant	deleterious	0.03
86	rs587782360	A/G	I308V	missense variant	deleterious	0.01
87	rs370795352	T/C	I308K	missense variant	deleterious	0.02
88	rs786201044	T/C	C309R	missense variant	deleterious	0.03
89	rs886047389	G/T	G21V	missense variant	deleterious	0.01
90	rs572685299	G/A	V182I	missense variant	deleterious	0
91	rs1564801689	A/G	S183G	missense variant	deleterious	0
92	rs1554890340	C/G	S183R	missense variant	deleterious	0
93	rs876660420	G/T	G193V	missense variant	deleterious	0
94	rs1554890398	G/T	D195Y	missense variant	deleterious	0
95	rs876661244	A/G	D195G	missense variant	deleterious	0
96	rs786204912	T/G	L198V	missense variant	deleterious	0.01
97	rs786201506	T/C	L198S	missense variant	deleterious	0
98	rs1410198544	G/A	V448I	missense variant	deleterious	0
99	rs1554825502	A/G	M443V	missense variant	deleterious	0
100	rs1195369834	G/T	M443I	missense variant	deleterious	0
101	rs1057518538	A/C	Q418P	missense variant	deleterious	0
102	rs1554898217	G/A	V331I	missense variant	deleterious	0
103	rs786202753	A/G	K336R	missense variant	deleterious	0
104	rs146629065	G/C	K337N	missense variant	deleterious	0.03
105	rs1554900530	T/C	I341T	missense variant	deleterious	0
106	rs1564837747	C/G	P342A	missense variant	deleterious	0
107	rs786204865	A/G	Q344P	missense variant	deleterious	0.01
108	rs757498880	A/G	Y349C	missense variant	deleterious	0
109	rs1554898193	A/G	E323G	missense variant	deleterious	0

110	rs9651492	G/C	D326N	missense variant	deleterious	0.01
111	rs1285259669	G/A	G35E	missense variant	deleterious	0.01
112	rs1564801070	C/T	P57S	missense variant	deleterious	0
113	rs1196170476	C/G	P57R	missense variant	deleterious	0
114	rs550595518	C/G	P59Q	missense variant	deleterious	0
115	rs1441812814	C/G	P64A	missense variant	deleterious	0
116	rs1001571902	C/T	P64R	missense variant	deleterious	0
117	rs1401532030	C/G	A66G	missense variant	deleterious	0.03
118	rs1237944553	C/T	P44L	missense variant	deleterious	0.01
119	rs1433679388	C/G	T46R	missense variant	deleterious	0
120	rs1564801018	C/T	R48W	missense variant	deleterious	0.01
121	rs1031706678	C/T	P70T	missense variant	deleterious	0
122	rs1045148218	A/G	E106G	missense variant	deleterious	0.03
123	rs904005027	C/A	P114Q	missense variant	deleterious	0
124	rs1221231481	C/T	P114L	missense variant	deleterious	0
125	rs1160370526	C/G	P124R	missense variant	deleterious	0
126	rs1379620359	C/T	A126V	missense variant	deleterious	0.02
127	rs1085308047	A/C	K186T	missense variant	deleterious	0.01
128	rs750098228	C/T	S164F	missense variant	deleterious	0.03
129	rs572922017	G/A	R154H	missense variant	deleterious	0.01
130	rs575260016	A/C	S117R	missense variant	deleterious	0.03
131	rs1239105602	C/T	A73V	missense variant	deleterious	0.01
132	rs1355651458	C/T	P40L	missense variant	deleterious	0
133	rs786204863	G/T	G338V	missense variant	deleterious	0



Graph 4.1 Prediction of Deleterious PTEN SNPs through SIFT

4.2 PolyPhen-2 (Polymorphism phenotyping 2):

PolyPhen-2 (Polymorphism phenotyping 2) algorithm was used for the prediction of deleterious effects in all of the total 133 SNPs of PTEN gene. PolyPhen-2 score ranges from 0.0 to 1.0 from benign, probably damaging and possibly damaging. Out of 133 SNP of PTEN gene checked through the PolyPhen-2, 86 showed the results in the range of probably damaging and possibly damaging. And 47 showed the results in the range of benign.

The results showed that 65% of our selected PTEN SNPs are damaging and they can have a deleterious effect. They can cause a fatal disease, tumor or cancer. They can decrease the protein's stability and make changes in it by disturbing the polypeptide chain of protein. Only 35% SNPs are found to be benign and tolerant to this mutation and are not involved in any kind of disease.

The SNP ID, Nucleotide change, type of variation, PolyPhen-2 prediction and score are mentioned in Table no. 4.2. The graph 4.2 is plotted to show the percentage of tolerant and deleterious SNPs. It shows that maximum 65% SNPs are deleterious and other 35% are tolerant.

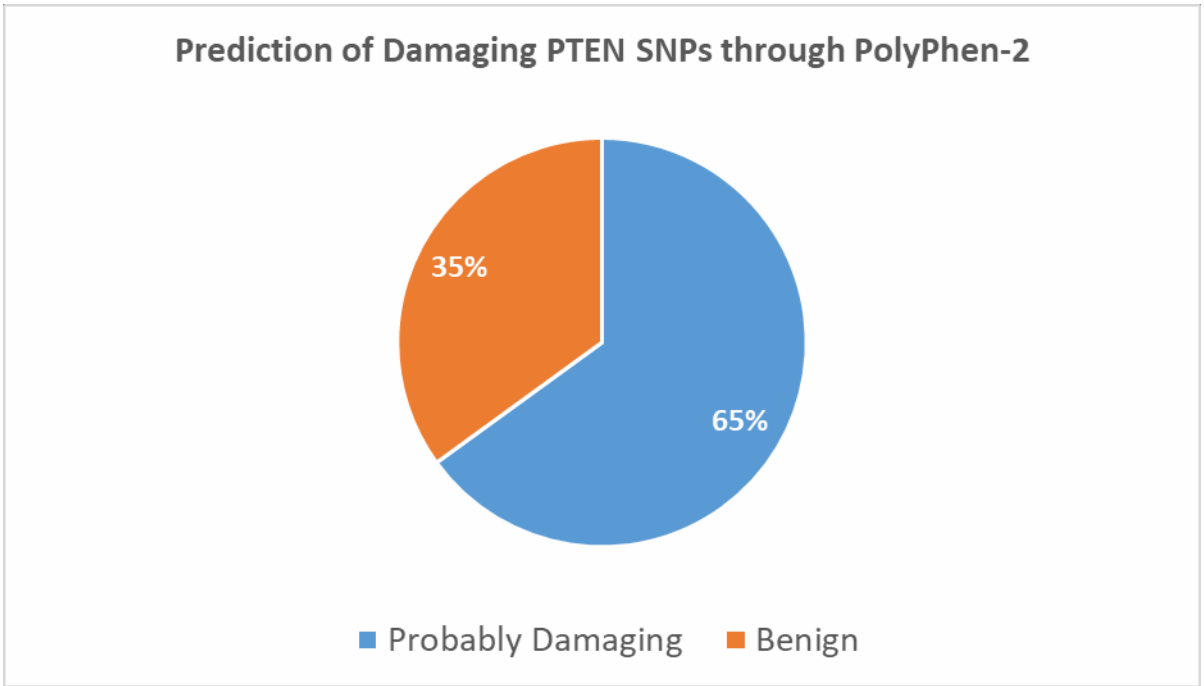
Table 4.2 Prediction of Deleterious PTEN SNPs through Polyphen 2

SR. NO.	SNP ID	ALLELE	AMINO ACID Change	TYPE OF SNP	PolyPhen-2 prediction	PolyPhen-2 Score
1	rs886047388	G/A	R20K	missense variant	benign	0.146
2	rs1451210152	G/C	R30S	missense variant	benign	0.005
3	rs1287849927	G/C	V38L	missense variant	benign	0.205
4	rs1318588388	G/T	R48L	missense variant	benign	0.3
5	rs1020319146	C/G	P71A	missense variant	benign	0
6	rs1282440343	T/C	S85P	missense variant	benign	0.015
7	rs1337121620	C/T	S85L	missense variant	benign	0.005
8	rs911548055	T/C	Y88H	missense variant	benign	0.106
9	rs1258029214	A/G	Y88C	missense variant	benign	0
10	rs886047393	C/T	P92S	missense variant	benign	0.034
11	rs1199933120	C/T	P92H	missense variant	benign	0.194
12	rs1048912442	A/G	S116G	missense variant	benign	0.302
13	rs575260016	C/T	R119G	missense variant	benign	0.097
14	rs761148721	G/C	E134D	missense variant	benign	0.398
15	rs1029309553	G/T	A137S	missense variant	benign	0.082
16	rs1316552000	C/T	A137V	missense variant	benign	0.108
17	rs1264797578	G/T	A139S	missense variant	benign	0.01
18	rs1483165357	A/C	K152Q	missense variant	benign	0.01
19	rs1417666224	C/T	H155Y	missense variant	benign	0.192
20	rs1163742361	A/G	Q156R	missense variant	benign	0.097
21	rs376610243	C/T	L162F	missense variant	benign	0.161
22	rs750098228	T/C	F165Y	missense variant	benign	0.195
23	rs889249323	T/C	F165S	missense variant	benign	0.195
24	rs1064793744	T/C	F167L	missense variant	benign	0.003
25	rs755295390	C/T	H169Y	missense variant	benign	0
26	rs753142719	A/G	H169R	missense variant	benign	0.024
27	rs749356977	C/G	L171V	missense variant	benign	0.101
28	rs1564801650	A/G	T175A	missense variant	benign	0.015
29	rs1060500109	C/T	A176V	missense variant	benign	0.063
30	rs1554890337	A/G	I178V	missense variant	benign	0.3
31	rs1564801672	G/C	E180D	missense variant	benign	0.021
32	rs1414611362	C/G	N185K	missense variant	benign	0.023
33	rs540063602	A/G	D192G	missense variant	benign	0.192

34	rs1064795967	T/G	D192E	missense variant	benign	0.396
35	rs1554890324	C/A	P172T	missense variant	benign	0.179
36	rs754821870	A/G	Q569R	missense variant	benign	0.02
37	rs876661021	A/G	H570R	missense variant	benign	0.593
38	rs587782345	A/T	N529D	missense variant	benign	0.736
39	rs1564570356	T/A	N529K	missense variant	benign	0.511
40	rs1554826026	A/G	Y519C	missense variant	benign	0.018
41	rs1554826034	G/T	E525Q	missense variant	benign	0.948
42	rs878853932	A/G	E526G	missense variant	benign	0.034
43	rs370064195	G/C/T	D470H	missense variant	benign	0.402
44	rs1554825550	A/G	E472G	missense variant	benign	0.048
45	rs758644748	G/A	D474N	missense variant	benign	0.065
46	rs587780007	G/T	S478I	missense variant	benign	0.063
47	rs786203858	A/G	I479V	missense variant	benign	0.01
48	rs786201507	G/A	R481H	missense variant	possibly damaging	0.013
49	rs1564568473	A/G	D483G	missense variant	possibly damaging	0.001
50	rs863224667	A/G	D485G	missense variant	possibly damaging	0.03
51	rs1554825577	C/A	L489I	missense variant	possibly damaging	0.32
52	rs1554898209	G/A	G329R	missense variant	probably damaging	0.883
53	rs1554900633	A/G	M371V	missense variant	possibly damaging	0.007
54	rs2943772	C/G	S65C	missense variant	possibly damaging	0.848
55	rs985226639	A/T	E77G	missense variant	possibly damaging	0.617
56	rs552470098	C/T	A79V	missense variant	possibly damaging	0.691
57	rs1395659213	A/G	Y123C	missense variant	possibly damaging	0.537
58	rs786204853	G/T	L198F	missense variant	probably damaging	0.993
59	rs11202592	A/T	R170W	missense variant	possibly damaging	0.675
60	rs1238691983	C/T	R119C	missense variant	possibly damaging	0.627
61	rs398123324	A/G	R187G	missense variant	possibly damaging	0.786
62	rs1064794096	G/T	R188I	missense variant	probably damaging	0.995
63	rs1064796078	A/T	R188S	missense variant	probably damaging	0.987
64	rs786204910	T/G	Y189D	missense variant	possibly damaging	0.875
65	rs786201995	T/G	L196V	missense variant	probably damaging	1
66	rs797044910	G/T	D197Y	missense variant	probably damaging	0.997
67	rs1320222638	A/G	D197G	missense variant	probably damaging	0.982
68	rs587781957	A/C	N185T	missense variant	possibly damaging	0.593
69	rs1114167648	A/C	H445P	missense variant	probably damaging	0.964
70	rs786204875	G/T	W447L	missense variant	probably damaging	0.998

71	rs587782350	C/T	P419L	missense variant	probably damaging	0.082
72	rs1057519368	T/G	L420S	missense variant	probably damaging	0.439
73	rs1114167664	G/T	G424V	missense variant	probably damaging	0.988
74	rs121909239	A/T	D425V	missense variant	probably damaging	0.997
75	rs1085308051	A/G	E330G	missense variant	probably damaging	0.103
76	rs786202688	A/G	R332G	missense variant	probably damaging	0.996
77	rs1114167673	G/T	R332M	missense variant	probably damaging	0.998
78	rs587782603	G/T	G338R	missense variant	probably damaging	0.982
79	rs1554900515	T/A	V339E	missense variant	probably damaging	0.978
80	rs397514559	C/A	T340N	missense variant	probably damaging	0.967
81	rs1554900534	A/G	S343R	missense variant	probably damaging	0.999
82	rs121909221	T/A	S343R	missense variant	probably damaging	0.999
83	rs121913293	C/T	R346S	missense variant	possibly damaging	0.991
84	rs121913294	G/T	R346P	missense variant	probably damaging	0.998
85	rs863224666	A/G	H314P	missense variant	probably damaging	0.975
86	rs587782360	A/G	I308V	missense variant	probably damaging	0.747
87	rs370795352	T/C	I308K	missense variant	probably damaging	0.997
88	rs786201044	T/C	C309R	missense variant	probably damaging	0.995
89	rs886047389	G/T	G21V	missense variant	probably damaging	0.98
90	rs572685299	G/A	V182I	missense variant	probably damaging	1
91	rs1564801689	A/G	S183G	missense variant	probably damaging	0.986
92	rs1554890340	C/G	S183R	missense variant	probably damaging	0.986
93	rs876660420	G/T	G193V	missense variant	probably damaging	0.97
94	rs1554890398	G/T	D195Y	missense variant	probably damaging	0.983
95	rs876661244	A/G	D195G	missense variant	probably damaging	0.939
96	rs786204912	T/G	L198V	missense variant	probably damaging	0.928
97	rs786201506	T/C	L198S	missense variant	probably damaging	0.982
98	rs1410198544	G/A	V448I	missense variant	possibly damaging	0.015
99	rs1554825502	A/G	M443V	missense variant	probably damaging	0.155
100	rs1195369834	G/T	M443I	missense variant	probably damaging	0.237
101	rs1057518538	A/C	Q418P	missense variant	probably damaging	0.039
102	rs1554898217	G/A	V331I	missense variant	probably damaging	0.063
103	rs786202753	A/G	K336R	missense variant	probably damaging	0.012
104	rs146629065	G/C	K337N	missense variant	possibly damaging	0.924
105	rs1554900530	T/C	I341T	missense variant	probably damaging	0.99
106	rs1564837747	C/G	P342A	missense variant	probably damaging	0.985
107	rs786204865	A/G	Q344P	missense variant	probably damaging	0.997

108	rs757498880	A/G	Y349C	missense variant	probably damaging	0.031
109	rs1554898193	A/G	E323G	missense variant	probably damaging	0.202
110	rs9651492	G/C	D326N	missense variant	probably damaging	0.07
111	rs1285259669	G/A	G35E	missense variant	probably damaging	0.97
112	rs1564801070	C/T	P57S	missense variant	possibly damaging	0.81
113	rs1196170476	C/G	P57R	missense variant	possibly damaging	0.908
114	rs550595518	C/G	P59Q	missense variant	possibly damaging	0.908
115	rs1441812814	C/G	P64A	missense variant	possibly damaging	0.737
116	rs1001571902	C/T	P64R	missense variant	possibly damaging	0.908
117	rs1401532030	C/G	A66G	missense variant	possibly damaging	0.578
118	rs1237944553	C/T	P44L	missense variant	possibly damaging	0.856
119	rs1433679388	C/G	T46R	missense variant	possibly damaging	0.711
120	rs1564801018	C/T	R48W	missense variant	possibly damaging	0.833
121	rs1031706678	C/T	P70T	missense variant	possibly damaging	0.81
122	rs1045148218	A/G	E106G	missense variant	possibly damaging	0.617
123	rs904005027	C/A	P114Q	missense variant	possibly damaging	0.908
124	rs1221231481	C/T	P114L	missense variant	possibly damaging	0.856
125	rs1160370526	C/G	P124R	missense variant	possibly damaging	0.908
126	rs1379620359	C/T	A126V	missense variant	possibly damaging	0.691
127	rs1085308047	A/C	K186T	missense variant	possibly damaging	0.843
128	rs750098228	C/T	S164F	missense variant	possibly damaging	0.692
129	rs572922017	G/A	R154H	missense variant	possibly damaging	0.65
130	rs575260016	A/C	S117R	missense variant	possibly damaging	0.514
131	rs1239105602	C/T	A73V	missense variant	possibly damaging	0.691
132	rs1355651458	C/T	P40L	missense variant	possibly damaging	0.856
133	rs786204863	G/T	G338V	missense variant	probably damaging	0.966



Graph 4.2 Prediction of deleterious SNPs of PTEN through Polyphen-2

4.3 PhD-SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms):

PhD-SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms) was used for the prediction of diseased SNPs out of the selected 133 SNPs of PTEN gene. PhD-SNP simply gives the prediction as diseased or neutral. Out of 133 SNP of PTEN gene checked through the PhD-SNP, 81 showed the results as diseased. And 52 showed the results as neutral.

The results showed that 61% of our selected PTEN SNPs are damaging and they can have a deleterious effect. They can cause a fatal disease, tumor or cancer. They can decrease the protein's stability and make changes in it by disturbing the polypeptide chain of protein. Only 39% SNPs are found to be Neutral to this mutation and are not involved in any kind of disease.

The SNP ID, Nucleotide change, type of variation, PhD-SNP prediction are mentioned in Table no. 4.3. The graph 4.3 is plotted to show the percentage of tolerant and deleterious SNPs. It shows that maximum 61% SNPs are deleterious and other 39% are tolerant.

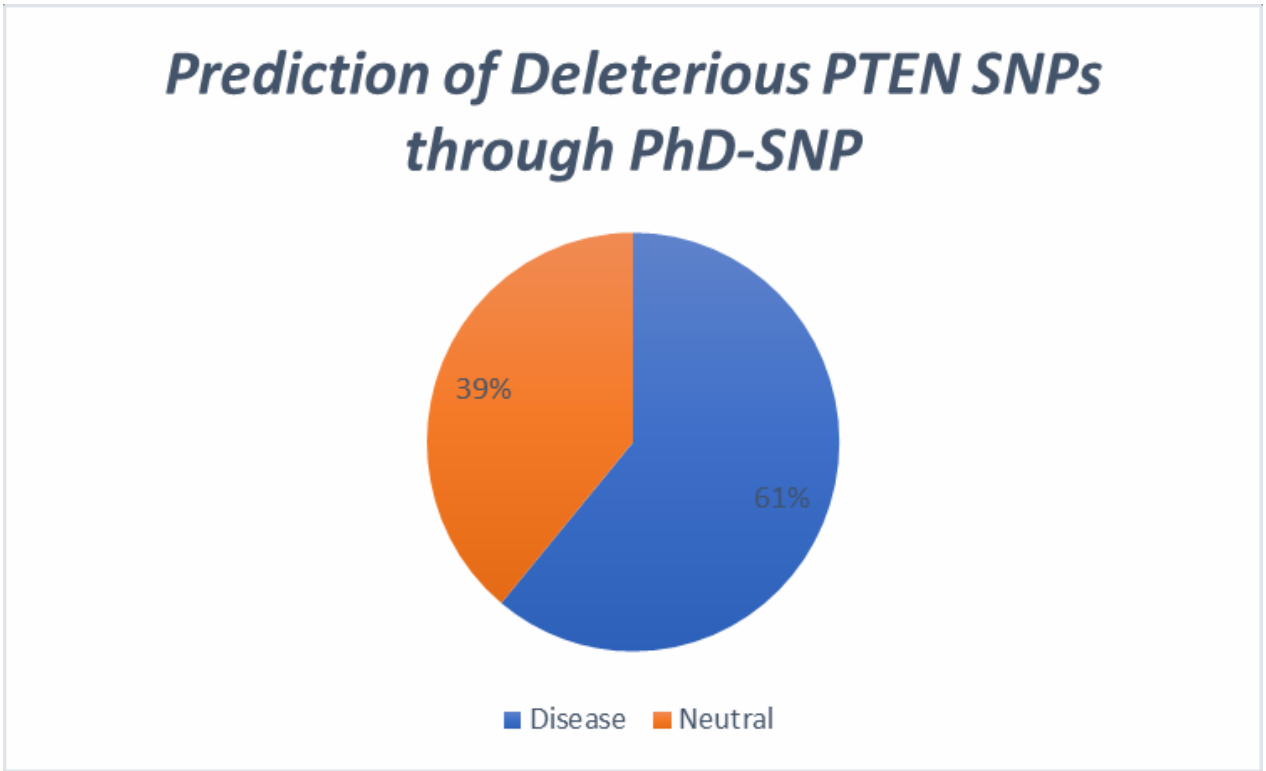
Table 4.3 Prediction of Deleterious PTEN SNPs through PhD-SNP

SR.NO.	SNP ID	ALLELE	AMINO ACID CHANGE	TYPE OF SNP	PHD-SNP prediction
1	rs886047388	G/A	R20K	missense variant	Neutral
2	rs1451210152	G/C	R30S	missense variant	Neutral
3	rs1287849927	G/C	V38L	missense variant	Neutral
4	rs1318588388	G/T	R48L	missense variant	Neutral
5	rs1020319146	C/G	P71A	missense variant	Neutral
6	rs1282440343	T/C	S85P	missense variant	Neutral
7	rs1337121620	C/T	S85L	missense variant	Neutral
8	rs911548055	T/C	Y88H	missense variant	Neutral
9	rs1258029214	A/G	Y88C	missense variant	Neutral
10	rs886047393	C/T	P92S	missense variant	Neutral
11	rs1199933120	C/T	P92H	missense variant	Neutral
12	rs1048912442	A/G	S116G	missense variant	Neutral
13	rs575260016	C/T	R119G	missense variant	Neutral
14	rs761148721	G/C	E134D	missense variant	Neutral
15	rs1029309553	G/T	A137S	missense variant	Neutral
16	rs1316552000	C/T	A137V	missense variant	Neutral
17	rs1264797578	G/T	A139S	missense variant	Neutral
18	rs1483165357	A/C	K152Q	missense variant	Neutral
19	rs1417666224	C/T	H155Y	missense variant	Neutral
20	rs1163742361	A/G	Q156R	missense variant	Neutral
21	rs376610243	C/T	L162F	missense variant	Neutral
22	rs750098228	T/C	F165Y	missense variant	Neutral
23	rs889249323	T/C	F165S	missense variant	Neutral
24	rs1064793744	T/C	F167L	missense variant	Neutral
25	rs755295390	C/T	H169Y	missense variant	Neutral
26	rs753142719	A/G	H169R	missense variant	Neutral
27	rs749356977	C/G	L171V	missense variant	Neutral
28	rs1564801650	A/G	T175A	missense variant	Neutral
29	rs1060500109	C/T	A176V	missense variant	Neutral
30	rs1554890337	A/G	I178V	missense variant	Neutral
31	rs1564801672	G/C	E180D	missense variant	Neutral
32	rs1414611362	C/G	N185K	missense variant	Neutral
33	rs540063602	A/G	D192G	missense variant	Neutral
34	rs1064795967	T/G	D192E	missense variant	Neutral
35	rs1554890324	C/A	P172T	missense variant	Neutral
36	rs754821870	A/G	Q569R	missense variant	Neutral

37	rs876661021	A/G	H570R	missense variant	Neutral
38	rs587782345	A/T	N529D	missense variant	Neutral
39	rs1564570356	T/A	N529K	missense variant	Neutral
40	rs1554826026	A/G	Y519C	missense variant	Neutral
41	rs1554826034	G/T	E525Q	missense variant	Neutral
42	rs878853932	A/G	E526G	missense variant	Neutral
43	rs370064195	G/C/T	D470H	missense variant	Neutral
44	rs1554825550	A/G	E472G	missense variant	Neutral
45	rs758644748	G/A	D474N	missense variant	Neutral
46	rs587780007	G/T	S478I	missense variant	Neutral
47	rs786203858	A/G	I479V	missense variant	Neutral
48	rs786201507	G/A	R481H	missense variant	Neutral
49	rs1564568473	A/G	D483G	missense variant	Neutral
50	rs863224667	A/G	D485G	missense variant	Neutral
51	rs1554825577	C/A	L489I	missense variant	Neutral
52	rs1554898209	G/A	G329R	missense variant	Neutral
53	rs1554900633	A/G	M371V	missense variant	Diseased
54	rs2943772	C/G	S65C	missense variant	Diseased
55	rs985226639	A/T	E77G	missense variant	Diseased
56	rs552470098	C/T	A79V	missense variant	Diseased
57	rs1395659213	A/G	Y123C	missense variant	Diseased
58	rs786204853	G/T	L198F	missense variant	Diseased
59	rs11202592	A/T	R170W	missense variant	Diseased
60	rs1238691983	C/T	R119C	missense variant	Diseased
61	rs398123324	A/G	R187G	missense variant	Diseased
62	rs1064794096	G/T	R188I	missense variant	Diseased
63	rs1064796078	A/T	R188S	missense variant	Diseased
64	rs786204910	T/G	Y189D	missense variant	Diseased
65	rs786201995	T/G	L196V	missense variant	Diseased
66	rs797044910	G/T	D197Y	missense variant	Diseased
67	rs1320222638	A/G	D197G	missense variant	Diseased
68	rs587781957	A/C	N185T	missense variant	Diseased
69	rs1114167648	A/C	H445P	missense variant	Diseased
70	rs786204875	G/T	W447L	missense variant	Diseased
71	rs587782350	C/T	P419L	missense variant	Diseased
72	rs1057519368	T/G	L420S	missense variant	Diseased
73	rs1114167664	G/T	G424V	missense variant	Diseased
74	rs121909239	A/T	D425V	missense variant	Diseased
75	rs1085308051	A/G	E330G	missense variant	Diseased
76	rs786202688	A/G	R332G	missense variant	Diseased

77	rs1114167673	G/T	R332M	missense variant	Diseased
78	rs587782603	G/T	G338R	missense variant	Diseased
79	rs1554900515	T/A	V339E	missense variant	Diseased
80	rs397514559	C/A	T340N	missense variant	Diseased
81	rs1554900534	A/G	S343R	missense variant	Diseased
82	rs121909221	T/A	S343R	missense variant	Diseased
83	rs121913293	C/T	R346S	missense variant	Diseased
84	rs121913294	G/T	R346P	missense variant	Diseased
85	rs863224666	A/G	H314P	missense variant	Diseased
86	rs587782360	A/G	I308V	missense variant	Diseased
87	rs370795352	T/C	I308K	missense variant	Diseased
88	rs786201044	T/C	C309R	missense variant	Diseased
89	rs886047389	G/T	G21V	missense variant	Diseased
90	rs572685299	G/A	V182I	missense variant	Diseased
91	rs1564801689	A/G	S183G	missense variant	Diseased
92	rs1554890340	C/G	S183R	missense variant	Diseased
93	rs876660420	G/T	G193V	missense variant	Diseased
94	rs1554890398	G/T	D195Y	missense variant	Diseased
95	rs876661244	A/G	D195G	missense variant	Diseased
96	rs786204912	T/G	L198V	missense variant	Diseased
97	rs786201506	T/C	L198S	missense variant	Diseased
98	rs1410198544	G/A	V448I	missense variant	Diseased
99	rs1554825502	A/G	M443V	missense variant	Diseased
100	rs1195369834	G/T	M443I	missense variant	Diseased
101	rs1057518538	A/C	Q418P	missense variant	Diseased
102	rs1554898217	G/A	V331I	missense variant	Diseased
103	rs786202753	A/G	K336R	missense variant	Diseased
104	rs146629065	G/C	K337N	missense variant	Diseased
105	rs1554900530	T/C	I341T	missense variant	Diseased
106	rs1564837747	C/G	P342A	missense variant	Diseased
107	rs786204865	A/G	Q344P	missense variant	Diseased
108	rs757498880	A/G	Y349C	missense variant	Diseased
109	rs1554898193	A/G	E323G	missense variant	Diseased
110	rs9651492	G/C	D326N	missense variant	Diseased
111	rs1285259669	G/A	G35E	missense variant	Diseased
112	rs1564801070	C/T	P57S	missense variant	Diseased
113	rs1196170476	C/G	P57R	missense variant	Diseased
114	rs550595518	C/G	P59Q	missense variant	Diseased
115	rs1441812814	C/G	P64A	missense variant	Diseased
116	rs1001571902	C/T	P64R	missense variant	Diseased

117	rs1401532030	C/G	A66G	missense variant	Diseased
118	rs1237944553	C/T	P44L	missense variant	Diseased
119	rs1433679388	C/G	T46R	missense variant	Diseased
120	rs1564801018	C/T	R48W	missense variant	Diseased
121	rs1031706678	C/T	P70T	missense variant	Diseased
122	rs1045148218	A/G	E106G	missense variant	Diseased
123	rs904005027	C/A	P114Q	missense variant	Diseased
124	rs1221231481	C/T	P114L	missense variant	Diseased
125	rs1160370526	C/G	P124R	missense variant	Diseased
126	rs1379620359	C/T	A126V	missense variant	Diseased
127	rs1085308047	A/C	K186T	missense variant	Diseased
128	rs750098228	C/T	S164F	missense variant	Diseased
129	rs572922017	G/A	R154H	missense variant	Diseased
130	rs575260016	A/C	S117R	missense variant	Diseased
131	rs1239105602	C/T	A73V	missense variant	Diseased
132	rs1355651458	C/T	P40L	missense variant	Diseased
133	rs786204863	G/T	G338V	missense variant	Diseased



Graph 4.3 Prediction of Diseased PTEN SNPs through PhD-SNP

4.4 PANTHER (protein analysis through evolutionary relationships)

Panther (protein analysis through evolutionary relationships) was used for the prediction of deleterious effects in all of the total 133 SNPs of PTEN gene. Panther gives us prediction of only those SNPs that are damaging. Those not damaging comes under not predicted category. Out of 133 SNP of PTEN gene checked through the panther, 83 showed the results in the range of probably damaging. And 50 showed the results in the category of non-damaging.

The results showed that 63% of our selected PTEN SNPs are probably damaging and they can have a deleterious effect. They can cause a fatal disease, tumor or cancer. They can decrease the protein's stability and make changes in it by disturbing the polypeptide chain of protein. Only 37% SNPs are found to be benign and tolerant to this mutation and are not involved in any kind of disease.

The SNP ID, Nucleotide change, type of variation and PhD-SNP prediction are mentioned in Table no. 4.4. The graph 4.4 is plotted to show the percentage of tolerant and deleterious SNPs. It shows that maximum 63% SNPs are deleterious and other 37% are tolerant.

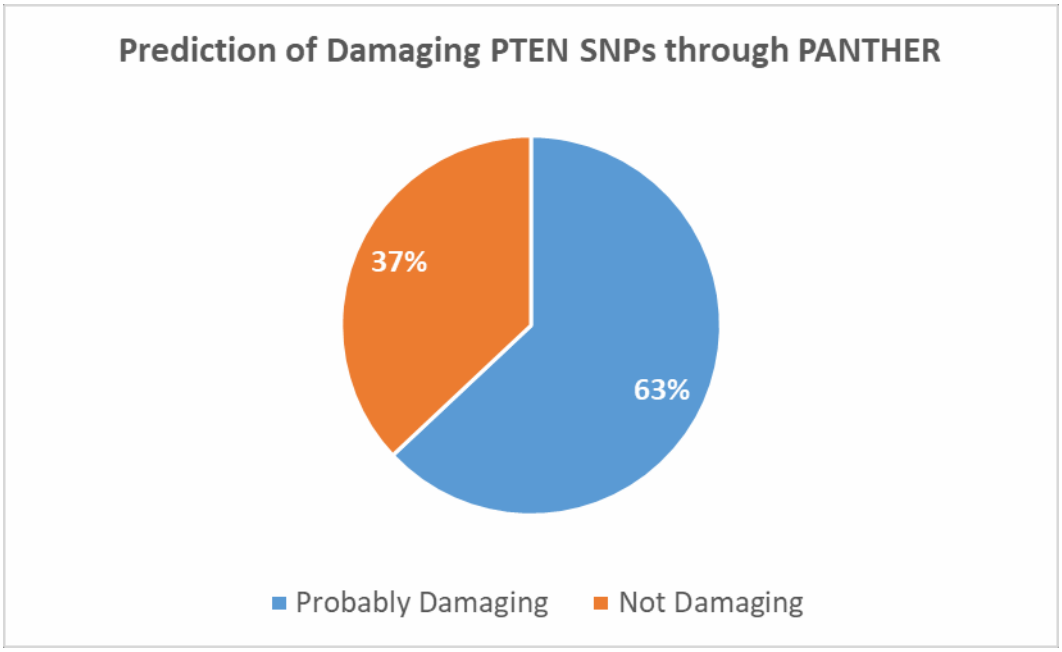
Table 4.4 Prediction of Deleterious PTEN SNPs through Panther

SR.NO.	SNP ID	ALLELE	AMINO ACID CHANGE	TYPE OF SNP	PANTHER prediction
1	rs886047388	G/A	R20K	missense variant	Non- Damaging
2	rs1451210152	G/C	R30S	missense variant	Non- Damaging
3	rs1287849927	G/C	V38L	missense variant	Non- Damaging
4	rs1318588388	G/T	R48L	missense variant	Non- Damaging
5	rs1020319146	C/G	P71A	missense variant	Non- Damaging
6	rs1282440343	T/C	S85P	missense variant	Non- Damaging
7	rs1337121620	C/T	S85L	missense variant	Non- Damaging
8	rs911548055	T/C	Y88H	missense variant	Non- Damaging
9	rs1258029214	A/G	Y88C	missense variant	Non- Damaging
10	rs886047393	C/T	P92S	missense variant	Non- Damaging
11	rs1199933120	C/T	P92H	missense variant	Non- Damaging
12	rs1048912442	A/G	S116G	missense variant	Non- Damaging
13	rs575260016	C/T	R119G	missense variant	Non- Damaging
14	rs761148721	G/C	E134D	missense variant	Non- Damaging
15	rs1029309553	G/T	A137S	missense variant	Non- Damaging
16	rs1316552000	C/T	A137V	missense variant	Non- Damaging
17	rs1264797578	G/T	A139S	missense variant	Non- Damaging
18	rs1483165357	A/C	K152Q	missense variant	Non- Damaging
19	rs1417666224	C/T	H155Y	missense variant	Non- Damaging
20	rs1163742361	A/G	Q156R	missense variant	Non- Damaging
21	rs376610243	C/T	L162F	missense variant	Non- Damaging
22	rs750098228	T/C	F165Y	missense variant	Non- Damaging
23	rs889249323	T/C	F165S	missense variant	Non- Damaging
24	rs1064793744	T/C	F167L	missense variant	Non- Damaging
25	rs755295390	C/T	H169Y	missense variant	Non- Damaging
26	rs753142719	A/G	H169R	missense variant	Non- Damaging
27	rs749356977	C/G	L171V	missense variant	Non- Damaging
28	rs1564801650	A/G	T175A	missense variant	Non- Damaging
29	rs1060500109	C/T	A176V	missense variant	Non- Damaging
30	rs1554890337	A/G	I178V	missense variant	Non- Damaging
31	rs1564801672	G/C	E180D	missense variant	Non- Damaging
32	rs1414611362	C/G	N185K	missense variant	Non- Damaging
33	rs540063602	A/G	D192G	missense variant	Non- Damaging
34	rs1064795967	T/G	D192E	missense variant	Non- Damaging
35	rs1554890324	C/A	P172T	missense variant	Non- Damaging
36	rs754821870	A/G	Q569R	missense variant	Non- Damaging

37	rs876661021	A/G	H570R	missense variant	Non- Damaging
38	rs587782345	A/T	N529D	missense variant	Non- Damaging
39	rs1564570356	T/A	N529K	missense variant	Non- Damaging
40	rs1554826026	A/G	Y519C	missense variant	Non- Damaging
41	rs1554826034	G/T	E525Q	missense variant	Non- Damaging
42	rs878853932	A/G	E526G	missense variant	Non- Damaging
43	rs370064195	G/C/T	D470H	missense variant	Non- Damaging
44	rs1554825550	A/G	E472G	missense variant	Non- Damaging
45	rs758644748	G/A	D474N	missense variant	Non- Damaging
46	rs587780007	G/T	S478I	missense variant	Non- Damaging
47	rs786203858	A/G	I479V	missense variant	Non- Damaging
48	rs786201507	G/A	R481H	missense variant	Non- Damaging
49	rs1564568473	A/G	D483G	missense variant	Non- Damaging
50	rs863224667	A/G	D485G	missense variant	Non- Damaging
51	rs1554825577	C/A	L489I	missense variant	Probably Damaging
52	rs1554898209	G/A	G329R	missense variant	Probably Damaging
53	rs1554900633	A/G	M371V	missense variant	Probably Damaging
54	rs2943772	C/G	S65C	missense variant	Probably Damaging
55	rs985226639	A/T	E77G	missense variant	Probably Damaging
56	rs552470098	C/T	A79V	missense variant	Probably Damaging
57	rs1395659213	A/G	Y123C	missense variant	Probably Damaging
58	rs786204853	G/T	L198F	missense variant	Probably Damaging
59	rs11202592	A/T	R170W	missense variant	Probably Damaging
60	rs1238691983	C/T	R119C	missense variant	Probably Damaging
61	rs398123324	A/G	R187G	missense variant	Probably Damaging
62	rs1064794096	G/T	R188I	missense variant	Probably Damaging
63	rs1064796078	A/T	R188S	missense variant	Probably Damaging
64	rs786204910	T/G	Y189D	missense variant	Probably Damaging
65	rs786201995	T/G	L196V	missense variant	Probably Damaging
66	rs797044910	G/T	D197Y	missense variant	Probably Damaging
67	rs1320222638	A/G	D197G	missense variant	Probably Damaging
68	rs587781957	A/C	N185T	missense variant	Probably Damaging
69	rs1114167648	A/C	H445P	missense variant	Probably Damaging
70	rs786204875	G/T	W447L	missense variant	Probably Damaging
71	rs587782350	C/T	P419L	missense variant	Probably Damaging
72	rs1057519368	T/G	L420S	missense variant	Probably Damaging
73	rs1114167664	G/T	G424V	missense variant	Probably Damaging
74	rs121909239	A/T	D425V	missense variant	Probably Damaging
75	rs1085308051	A/G	E330G	missense variant	Probably Damaging
76	rs786202688	A/G	R332G	missense variant	Probably Damaging

77	rs1114167673	G/T	R332M	missense variant	Probably Damaging
78	rs587782603	G/T	G338R	missense variant	Probably Damaging
79	rs1554900515	T/A	V339E	missense variant	Probably Damaging
80	rs397514559	C/A	T340N	missense variant	Probably Damaging
81	rs1554900534	A/G	S343R	missense variant	Probably Damaging
82	rs121909221	T/A	S343R	missense variant	Probably Damaging
83	rs121913293	C/T	R346S	missense variant	Probably Damaging
84	rs121913294	G/T	R346P	missense variant	Probably Damaging
85	rs863224666	A/G	H314P	missense variant	Probably Damaging
86	rs587782360	A/G	I308V	missense variant	Probably Damaging
87	rs370795352	T/C	I308K	missense variant	Probably Damaging
88	rs786201044	T/C	C309R	missense variant	Probably Damaging
89	rs886047389	G/T	G21V	missense variant	Probably Damaging
90	rs572685299	G/A	V182I	missense variant	Probably Damaging
91	rs1564801689	A/G	S183G	missense variant	Probably Damaging
92	rs1554890340	C/G	S183R	missense variant	Probably Damaging
93	rs876660420	G/T	G193V	missense variant	Probably Damaging
94	rs1554890398	G/T	D195Y	missense variant	Probably Damaging
95	rs876661244	A/G	D195G	missense variant	Probably Damaging
96	rs786204912	T/G	L198V	missense variant	Probably Damaging
97	rs786201506	T/C	L198S	missense variant	Probably Damaging
98	rs1410198544	G/A	V448I	missense variant	Probably Damaging
99	rs1554825502	A/G	M443V	missense variant	Probably Damaging
100	rs1195369834	G/T	M443I	missense variant	Probably Damaging
101	rs1057518538	A/C	Q418P	missense variant	Probably Damaging
102	rs1554898217	G/A	V331I	missense variant	Probably Damaging
103	rs786202753	A/G	K336R	missense variant	Probably Damaging
104	rs146629065	G/C	K337N	missense variant	Probably Damaging
105	rs1554900530	T/C	I341T	missense variant	Probably Damaging
106	rs1564837747	C/G	P342A	missense variant	Probably Damaging
107	rs786204865	A/G	Q344P	missense variant	Probably Damaging
108	rs757498880	A/G	Y349C	missense variant	Probably Damaging
109	rs1554898193	A/G	E323G	missense variant	Probably Damaging
110	rs9651492	G/C	D326N	missense variant	Probably Damaging
111	rs1285259669	G/A	G35E	missense variant	Probably Damaging
112	rs1564801070	C/T	P57S	missense variant	Probably Damaging
113	rs1196170476	C/G	P57R	missense variant	Probably Damaging
114	rs550595518	C/G	P59Q	missense variant	Probably Damaging
115	rs1441812814	C/G	P64A	missense variant	Probably Damaging
116	rs1001571902	C/T	P64R	missense variant	Probably Damaging

117	rs1401532030	C/G	A66G	missense variant	Probably Damaging
118	rs1237944553	C/T	P44L	missense variant	Probably Damaging
119	rs1433679388	C/G	T46R	missense variant	Probably Damaging
120	rs1564801018	C/T	R48W	missense variant	Probably Damaging
121	rs1031706678	C/T	P70T	missense variant	Probably Damaging
122	rs1045148218	A/G	E106G	missense variant	Probably Damaging
123	rs904005027	C/A	P114Q	missense variant	Probably Damaging
124	rs1221231481	C/T	P114L	missense variant	Probably Damaging
125	rs1160370526	C/G	P124R	missense variant	Probably Damaging
126	rs1379620359	C/T	A126V	missense variant	Probably Damaging
127	rs1085308047	A/C	K186T	missense variant	Probably Damaging
128	rs750098228	C/T	S164F	missense variant	Probably Damaging
129	rs572922017	G/A	R154H	missense variant	Probably Damaging
130	rs575260016	A/C	S117R	missense variant	Probably Damaging
131	rs1239105602	C/T	A73V	missense variant	Probably Damaging
132	rs1355651458	C/T	P40L	missense variant	Probably Damaging
133	rs786204863	G/T	G338V	missense variant	Probably Damaging



Graph 4.4 Prediction of Deleterious PTEN SNPs through Panther

4.5 PROVEAN (Protein Variation Effect Analyzer)

Provean (protein variation effect analyzer) was used for the prediction of deleterious effects in all of the total 133 SNPs of PTEN gene. Provean (protein variation effect analyzer) give us prediction in two categories as deleterious or neutral. Out of 133 SNP of PTEN gene checked through Provean, 88 showed the results in the category of deleterious and 45 gave the prediction as Neutral.

The results showed that 66% of our selected PTEN SNPs are deleterious and they can have a deleterious effect. They can cause a fatal disease, tumor or cancer. They can decrease the protein's stability and make changes in it by disturbing the polypeptide chain of protein. Only 34% SNPs are found to be tolerant to this mutation and are not involved in any kind of disease.

The SNP ID, Nucleotide change, type of variation and Provean prediction are mentioned in Table no. 4.5. The graph 4.5 is plotted to show the percentage of neutral and deleterious SNPs. It shows that maximum 66% SNPs are deleterious and other 34% are tolerant.

Table 4.5 Prediction of Deleterious PTEN SNPs through Provean

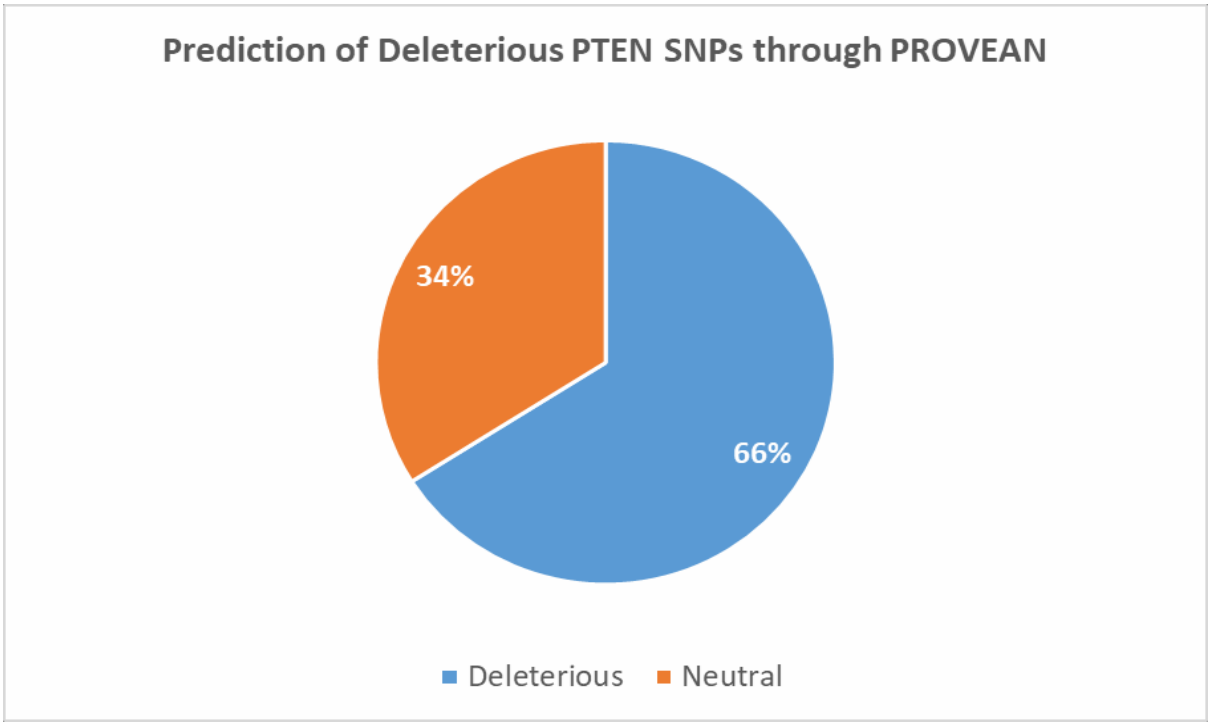
SR.NO.	SNP ID	ALLELE	AMINO ACID CHANGE	TYPE OF SNP	PROVEAN prediction
1	rs886047388	G/A	R20K	missense variant	Neutral
2	rs1451210152	G/C	R30S	missense variant	Neutral
3	rs1287849927	G/C	V38L	missense variant	Neutral
4	rs1318588388	G/T	R48L	missense variant	Neutral
5	rs1020319146	C/G	P71A	missense variant	Neutral
6	rs1282440343	T/C	S85P	missense variant	Neutral
7	rs1337121620	C/T	S85L	missense variant	Neutral
8	rs911548055	T/C	Y88H	missense variant	Neutral
9	rs1258029214	A/G	Y88C	missense variant	Neutral
10	rs886047393	C/T	P92S	missense variant	Neutral
11	rs1199933120	C/T	P92H	missense variant	Neutral
12	rs1048912442	A/G	S116G	missense variant	Neutral
13	rs575260016	C/T	R119G	missense variant	Neutral
14	rs761148721	G/C	E134D	missense variant	Neutral
15	rs1029309553	G/T	A137S	missense variant	Neutral
16	rs1316552000	C/T	A137V	missense variant	Neutral
17	rs1264797578	G/T	A139S	missense variant	Neutral
18	rs1483165357	A/C	K152Q	missense variant	Neutral
19	rs1417666224	C/T	H155Y	missense variant	Neutral
20	rs1163742361	A/G	Q156R	missense variant	Neutral
21	rs376610243	C/T	L162F	missense variant	Neutral
22	rs750098228	T/C	F165Y	missense variant	Neutral
23	rs889249323	T/C	F165S	missense variant	Neutral
24	rs1064793744	T/C	F167L	missense variant	Neutral
25	rs755295390	C/T	H169Y	missense variant	Neutral
26	rs753142719	A/G	H169R	missense variant	Neutral
27	rs749356977	C/G	L171V	missense variant	Neutral
28	rs1564801650	A/G	T175A	missense variant	Neutral
29	rs1060500109	C/T	A176V	missense variant	Neutral
30	rs1554890337	A/G	I178V	missense variant	Neutral
31	rs1564801672	G/C	E180D	missense variant	Neutral
32	rs1414611362	C/G	N185K	missense variant	Neutral
33	rs540063602	A/G	D192G	missense variant	Neutral
34	rs1064795967	T/G	D192E	missense variant	Neutral
35	rs1554890324	C/A	P172T	missense variant	Neutral
36	rs754821870	A/G	Q569R	missense variant	Neutral

37	rs876661021	A/G	H570R	missense variant	Neutral
38	rs587782345	A/T	N529D	missense variant	Neutral
39	rs1564570356	T/A	N529K	missense variant	Neutral
40	rs1554826026	A/G	Y519C	missense variant	Neutral
41	rs1554826034	G/T	E525Q	missense variant	Neutral
42	rs878853932	A/G	E526G	missense variant	Neutral
43	rs370064195	G/C/T	D470H	missense variant	Neutral
44	rs1554825550	A/G	E472G	missense variant	Neutral
45	rs758644748	G/A	D474N	missense variant	Neutral
46	rs587780007	G/T	S478I	missense variant	Deleterious
47	rs786203858	A/G	I479V	missense variant	Deleterious
48	rs786201507	G/A	R481H	missense variant	Deleterious
49	rs1564568473	A/G	D483G	missense variant	Deleterious
50	rs863224667	A/G	D485G	missense variant	Deleterious
51	rs1554825577	C/A	L489I	missense variant	Deleterious
52	rs1554898209	G/A	G329R	missense variant	Deleterious
53	rs1554900633	A/G	M371V	missense variant	Deleterious
54	rs2943772	C/G	S65C	missense variant	Deleterious
55	rs985226639	A/T	E77G	missense variant	Deleterious
56	rs552470098	C/T	A79V	missense variant	Deleterious
57	rs1395659213	A/G	Y123C	missense variant	Deleterious
58	rs786204853	G/T	L198F	missense variant	Deleterious
59	rs11202592	A/T	R170W	missense variant	Deleterious
60	rs1238691983	C/T	R119C	missense variant	Deleterious
61	rs398123324	A/G	R187G	missense variant	Deleterious
62	rs1064794096	G/T	R188I	missense variant	Deleterious
63	rs1064796078	A/T	R188S	missense variant	Deleterious
64	rs786204910	T/G	Y189D	missense variant	Deleterious
65	rs786201995	T/G	L196V	missense variant	Deleterious
66	rs797044910	G/T	D197Y	missense variant	Deleterious
67	rs1320222638	A/G	D197G	missense variant	Deleterious
68	rs587781957	A/C	N185T	missense variant	Deleterious
69	rs1114167648	A/C	H445P	missense variant	Deleterious
70	rs786204875	G/T	W447L	missense variant	Deleterious
71	rs587782350	C/T	P419L	missense variant	Deleterious
72	rs1057519368	T/G	L420S	missense variant	Deleterious
73	rs1114167664	G/T	G424V	missense variant	Deleterious
74	rs121909239	A/T	D425V	missense variant	Deleterious
75	rs1085308051	A/G	E330G	missense variant	Deleterious
76	rs786202688	A/G	R332G	missense variant	Deleterious

77	rs1114167673	G/T	R332M	missense variant	Deleterious
78	rs587782603	G/T	G338R	missense variant	Deleterious
79	rs1554900515	T/A	V339E	missense variant	Deleterious
80	rs397514559	C/A	T340N	missense variant	Deleterious
81	rs1554900534	A/G	S343R	missense variant	Deleterious
82	rs121909221	T/A	S343R	missense variant	Deleterious
83	rs121913293	C/T	R346S	missense variant	Deleterious
84	rs121913294	G/T	R346P	missense variant	Deleterious
85	rs863224666	A/G	H314P	missense variant	Deleterious
86	rs587782360	A/G	I308V	missense variant	Deleterious
87	rs370795352	T/C	I308K	missense variant	Deleterious
88	rs786201044	T/C	C309R	missense variant	Deleterious
89	rs886047389	G/T	G21V	missense variant	Deleterious
90	rs572685299	G/A	V182I	missense variant	Deleterious
91	rs1564801689	A/G	S183G	missense variant	Deleterious
92	rs1554890340	C/G	S183R	missense variant	Deleterious
93	rs876660420	G/T	G193V	missense variant	Deleterious
94	rs1554890398	G/T	D195Y	missense variant	Deleterious
95	rs876661244	A/G	D195G	missense variant	Deleterious
96	rs786204912	T/G	L198V	missense variant	Deleterious
97	rs786201506	T/C	L198S	missense variant	Deleterious
98	rs1410198544	G/A	V448I	missense variant	Deleterious
99	rs1554825502	A/G	M443V	missense variant	Deleterious
100	rs1195369834	G/T	M443I	missense variant	Deleterious
101	rs1057518538	A/C	Q418P	missense variant	Deleterious
102	rs1554898217	G/A	V331I	missense variant	Deleterious
103	rs786202753	A/G	K336R	missense variant	Deleterious
104	rs146629065	G/C	K337N	missense variant	Deleterious
105	rs1554900530	T/C	I341T	missense variant	Deleterious
106	rs1564837747	C/G	P342A	missense variant	Deleterious
107	rs786204865	A/G	Q344P	missense variant	Deleterious
108	rs757498880	A/G	Y349C	missense variant	Deleterious
109	rs1554898193	A/G	E323G	missense variant	Deleterious
110	rs9651492	G/C	D326N	missense variant	Deleterious
111	rs1285259669	G/A	G35E	missense variant	Deleterious
112	rs1564801070	C/T	P57S	missense variant	Deleterious
113	rs1196170476	C/G	P57R	missense variant	Deleterious
114	rs550595518	C/G	P59Q	missense variant	Deleterious
115	rs1441812814	C/G	P64A	missense variant	Deleterious
116	rs1001571902	C/T	P64R	missense variant	Deleterious

117	rs1401532030	C/G	A66G	missense variant	Deleterious
118	rs1237944553	C/T	P44L	missense variant	Deleterious
119	rs1433679388	C/G	T46R	missense variant	Deleterious
120	rs1564801018	C/T	R48W	missense variant	Deleterious
121	rs1031706678	C/T	P70T	missense variant	Deleterious
122	rs1045148218	A/G	E106G	missense variant	Deleterious
123	rs904005027	C/A	P114Q	missense variant	Deleterious
124	rs1221231481	C/T	P114L	missense variant	Deleterious
125	rs1160370526	C/G	P124R	missense variant	Deleterious
126	rs1379620359	C/T	A126V	missense variant	Deleterious
127	rs1085308047	A/C	K186T	missense variant	Deleterious
128	rs750098228	C/T	S164F	missense variant	Deleterious
129	rs572922017	G/A	R154H	missense variant	Deleterious
130	rs575260016	A/C	S117R	missense variant	Deleterious
131	rs1239105602	C/T	A73V	missense variant	Deleterious
132	rs1355651458	C/T	P40L	missense variant	Deleterious
133	rs786204863	G/T	G338V	missense variant	Deleterious

s



Graph 4.5 Prediction of Deleterious PTEN SNPs through Provean

4.6 SNP and GO (Predicting disease associated variations using GO terms)

SNP and GO was used for the prediction of diseased SNPs out of the selected 133 SNPs of PTEN gene. SNP and GO simply gives the prediction as diseased or neutral. Out of 133 SNP of PTEN gene checked through the SNP and GO, 90 showed the results as diseased. And 43 showed the results as neutral.

The results showed that 60% of our selected PTEN SNPs are diseased and they can have a deleterious effect. They can cause any fatal disease, tumor or cancer. They can decrease the protein's stability and make changes in it by disturbing the polypeptide chain of protein. Only 40% SNPs are found to be neutral to this mutation and are not involved in any kind of disease.

The SNP ID, Nucleotide change, type of variation, SNP and GO prediction are mentioned in Table no. 4.6. The graph 4.6 was plotted to show the percentage of neutral and diseased SNPs. It shows that maximum 60% SNPs are diseased and other 40% are neutral.

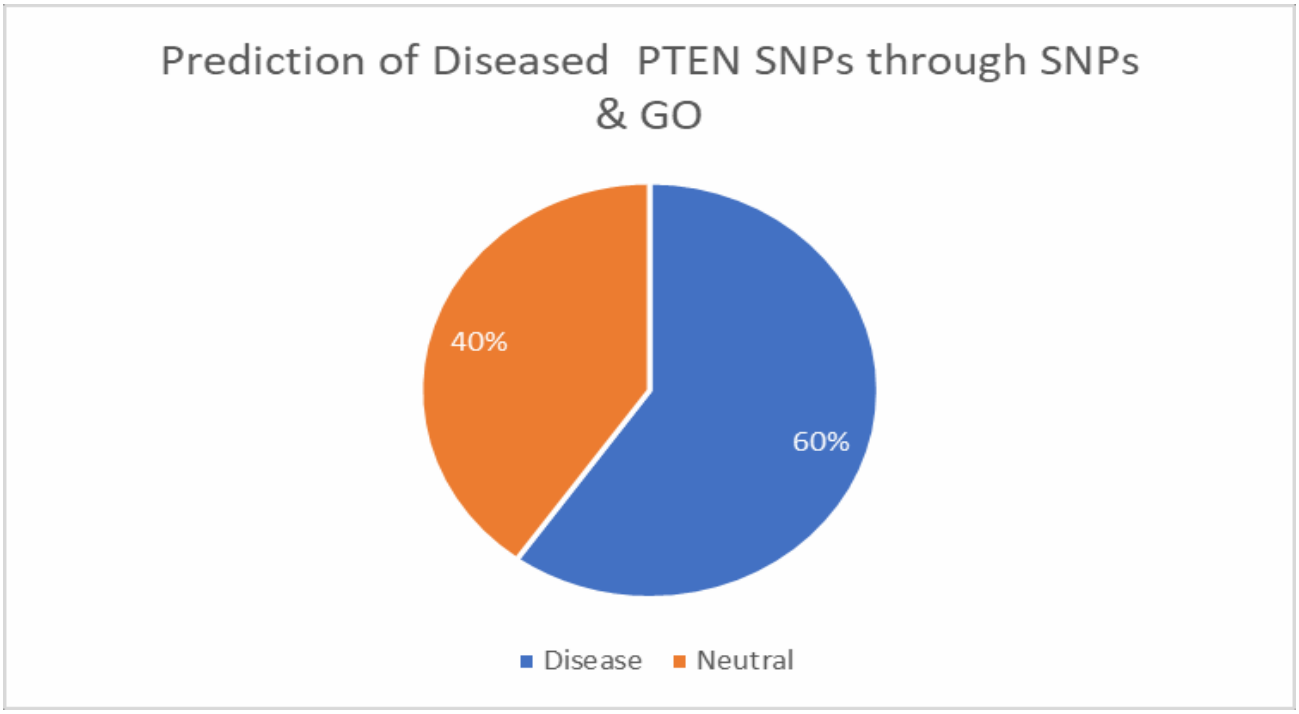
Table 4.6 Prediction of Diseased PTEN SNPs through SNPs & GO

SR.NO.	SNP ID	ALLELE	AMINO ACID CHANGE	TYPE OF SNP	SNP and GO prediction
1	rs886047388	G/A	R20K	missense variant	Neutral
2	rs1451210152	G/C	R30S	missense variant	Neutral
3	rs1287849927	G/C	V38L	missense variant	Neutral
4	rs1318588388	G/T	R48L	missense variant	Neutral
5	rs1020319146	C/G	P71A	missense variant	Neutral
6	rs1282440343	T/C	S85P	missense variant	Neutral
7	rs1337121620	C/T	S85L	missense variant	Neutral
8	rs911548055	T/C	Y88H	missense variant	Neutral
9	rs1258029214	A/G	Y88C	missense variant	Neutral
10	rs886047393	C/T	P92S	missense variant	Neutral
11	rs1199933120	C/T	P92H	missense variant	Neutral
12	rs1048912442	A/G	S116G	missense variant	Neutral
13	rs575260016	C/T	R119G	missense variant	Neutral
14	rs761148721	G/C	E134D	missense variant	Neutral
15	rs1029309553	G/T	A137S	missense variant	Neutral
16	rs1316552000	C/T	A137V	missense variant	Neutral
17	rs1264797578	G/T	A139S	missense variant	Neutral
18	rs1483165357	A/C	K152Q	missense variant	Neutral
19	rs1417666224	C/T	H155Y	missense variant	Neutral
20	rs1163742361	A/G	Q156R	missense variant	Neutral
21	rs376610243	C/T	L162F	missense variant	Neutral
22	rs750098228	T/C	F165Y	missense variant	Neutral
23	rs889249323	T/C	F165S	missense variant	Neutral
24	rs1064793744	T/C	F167L	missense variant	Neutral
25	rs755295390	C/T	H169Y	missense variant	Neutral
26	rs753142719	A/G	H169R	missense variant	Neutral
27	rs749356977	C/G	L171V	missense variant	Neutral
28	rs1564801650	A/G	T175A	missense variant	Neutral
29	rs1060500109	C/T	A176V	missense variant	Neutral
30	rs1554890337	A/G	I178V	missense variant	Neutral
31	rs1564801672	G/C	E180D	missense variant	Neutral
32	rs1414611362	C/G	N185K	missense variant	Neutral
33	rs540063602	A/G	D192G	missense variant	Neutral
34	rs1064795967	T/G	D192E	missense variant	Neutral
35	rs1554890324	C/A	P172T	missense variant	Neutral

36	rs754821870	A/G	Q569R	missense variant	Neutral
37	rs876661021	A/G	H570R	missense variant	Neutral
38	rs587782345	A/T	N529D	missense variant	Neutral
39	rs1564570356	T/A	N529K	missense variant	Neutral
40	rs1554826026	A/G	Y519C	missense variant	Neutral
41	rs1554826034	G/T	E525Q	missense variant	Neutral
42	rs878853932	A/G	E526G	missense variant	Neutral
43	rs370064195	G/C/T	D470H	missense variant	Neutral
44	rs1554825550	A/G	E472G	missense variant	Diseased
45	rs758644748	G/A	D474N	missense variant	Diseased
46	rs587780007	G/T	S478I	missense variant	Diseased
47	rs786203858	A/G	I479V	missense variant	Diseased
48	rs786201507	G/A	R481H	missense variant	Diseased
49	rs1564568473	A/G	D483G	missense variant	Diseased
50	rs863224667	A/G	D485G	missense variant	Diseased
51	rs1554825577	C/A	L489I	missense variant	Diseased
52	rs1554898209	G/A	G329R	missense variant	Diseased
53	rs1554900633	A/G	M371V	missense variant	Diseased
54	rs2943772	C/G	S65C	missense variant	Diseased
55	rs985226639	A/T	E77G	missense variant	Diseased
56	rs552470098	C/T	A79V	missense variant	Diseased
57	rs1395659213	A/G	Y123C	missense variant	Diseased
58	rs786204853	G/T	L198F	missense variant	Diseased
59	rs11202592	A/T	R170W	missense variant	Diseased
60	rs1238691983	C/T	R119C	missense variant	Diseased
61	rs398123324	A/G	R187G	missense variant	Diseased
62	rs1064794096	G/T	R188I	missense variant	Diseased
63	rs1064796078	A/T	R188S	missense variant	Diseased
64	rs786204910	T/G	Y189D	missense variant	Diseased
65	rs786201995	T/G	L196V	missense variant	Diseased
66	rs797044910	G/T	D197Y	missense variant	Diseased
67	rs1320222638	A/G	D197G	missense variant	Diseased
68	rs587781957	A/C	N185T	missense variant	Diseased
69	rs1114167648	A/C	H445P	missense variant	Diseased
70	rs786204875	G/T	W447L	missense variant	Diseased
71	rs587782350	C/T	P419L	missense variant	Diseased
72	rs1057519368	T/G	L420S	missense variant	Diseased
73	rs1114167664	G/T	G424V	missense variant	Diseased
74	rs121909239	A/T	D425V	missense variant	Diseased
75	rs1085308051	A/G	E330G	missense variant	Diseased

76	rs786202688	A/G	R332G	missense variant	Diseased
77	rs1114167673	G/T	R332M	missense variant	Diseased
78	rs587782603	G/T	G338R	missense variant	Diseased
79	rs1554900515	T/A	V339E	missense variant	Diseased
80	rs397514559	C/A	T340N	missense variant	Diseased
81	rs1554900534	A/G	S343R	missense variant	Diseased
82	rs121909221	T/A	S343R	missense variant	Diseased
83	rs121913293	C/T	R346S	missense variant	Diseased
84	rs121913294	G/T	R346P	missense variant	Diseased
85	rs863224666	A/G	H314P	missense variant	Diseased
86	rs587782360	A/G	I308V	missense variant	Diseased
87	rs370795352	T/C	I308K	missense variant	Diseased
88	rs786201044	T/C	C309R	missense variant	Diseased
89	rs886047389	G/T	G21V	missense variant	Diseased
90	rs572685299	G/A	V182I	missense variant	Diseased
91	rs1564801689	A/G	S183G	missense variant	Diseased
92	rs1554890340	C/G	S183R	missense variant	Diseased
93	rs876660420	G/T	G193V	missense variant	Diseased
94	rs1554890398	G/T	D195Y	missense variant	Diseased
95	rs876661244	A/G	D195G	missense variant	Diseased
96	rs786204912	T/G	L198V	missense variant	Diseased
97	rs786201506	T/C	L198S	missense variant	Diseased
98	rs1410198544	G/A	V448I	missense variant	Diseased
99	rs1554825502	A/G	M443V	missense variant	Diseased
100	rs1195369834	G/T	M443I	missense variant	Diseased
101	rs1057518538	A/C	Q418P	missense variant	Diseased
102	rs1554898217	G/A	V331I	missense variant	Diseased
103	rs786202753	A/G	K336R	missense variant	Diseased
104	rs146629065	G/C	K337N	missense variant	Diseased
105	rs1554900530	T/C	I341T	missense variant	Diseased
106	rs1564837747	C/G	P342A	missense variant	Diseased
107	rs786204865	A/G	Q344P	missense variant	Diseased
108	rs757498880	A/G	Y349C	missense variant	Diseased
109	rs1554898193	A/G	E323G	missense variant	Diseased
110	rs9651492	G/C	D326N	missense variant	Diseased
111	rs1285259669	G/A	G35E	missense variant	Diseased
112	rs1564801070	C/T	P57S	missense variant	Diseased
113	rs1196170476	C/G	P57R	missense variant	Diseased
114	rs550595518	C/G	P59Q	missense variant	Diseased
115	rs1441812814	C/G	P64A	missense variant	Diseased

116	rs1001571902	C/T	P64R	missense variant	Diseased
117	rs1401532030	C/G	A66G	missense variant	Diseased
118	rs1237944553	C/T	P44L	missense variant	Diseased
119	rs1433679388	C/G	T46R	missense variant	Diseased
120	rs1564801018	C/T	R48W	missense variant	Diseased
121	rs1031706678	C/T	P70T	missense variant	Diseased
122	rs1045148218	A/G	E106G	missense variant	Diseased
123	rs904005027	C/A	P114Q	missense variant	Diseased
124	rs1221231481	C/T	P114L	missense variant	Diseased
125	rs1160370526	C/G	P124R	missense variant	Diseased
126	rs1379620359	C/T	A126V	missense variant	Diseased
127	rs1085308047	A/C	K186T	missense variant	Diseased
128	rs750098228	C/T	S164F	missense variant	Diseased
129	rs572922017	G/A	R154H	missense variant	Diseased
130	rs575260016	A/C	S117R	missense variant	Diseased
131	rs1239105602	C/T	A73V	missense variant	Diseased
132	rs1355651458	C/T	P40L	missense variant	Diseased
133	rs786204863	G/T	G338V	missense variant	Diseased



Graph 4.6 Prediction of Diseased PTEN SNPs through SNPs & GO

4.7 Comparison of results

An extensive comparison of results of all the tools used for finding deleterious and damaging SNPs of PTEN was carried out. The comparison was necessary to see if all the software gave the same results or vary a lot. It was found that out of 133 SNPs used in the research for finding out damaged SNPs having decreased protein stability, 80 of the SNPs are shown to be confirms deleterious by all the tools used including SIFT, Polyphen-2, Provean, Snps & Go, Panther and Provean. This shows 38% of the SNPs of PTEN used in the study are damaging and deleterious by all means having decreased protein stability. This is confirmed by all the software used in research.

Comparison of all Results of SIFT, POLYPHEN-2, PROVEAN, Snps & GO, PANTHER and Provean is shown in Table 4.7. Out of 80 pathogenic SNPs of PTEN, 35 SNPs are reported to be pathogenic, 22 SNPs are under study and have uncertain clinical significance and 23 are found to be not reported yet and can be considered as novel. The summarized 22 damaged / deleterious and pathogenic novel SNPs of PTEN is given in table 4.8.

Table 4.7 Comparison of Results of SIFT, POLYPHEN-2, PROVEAN, SNPs & GO, PANTHER, Provean

Variant ID	Alleles change	AA change	SIFT	PolyPhen-2	PANTHER	PROVEAN	PHD-SNP	SNP & GO
rs1287849927	G/C	V38L	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1318588388	G/T	R48L	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1020319146	C/G	P71A	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1282440343	T/C	S85P	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1337121620	C/T	S85L	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs911548055	T/C	Y88H	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1258029214	A/G	Y88C	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs886047393	C/T	P92S	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1199933120	C/T	P92H	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1048912442	A/G	S116G	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs575260016	C/T	R119G	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs761148721	G/C	E134D	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1029309553	G/T	A137S	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1316552000	C/T	A137V	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1264797578	G/T	A139S	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1483165357	A/C	K152Q	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1417666224	C/T	H155Y	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1163742361	A/G	Q156R	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs376610243	C/T	L162F	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs750098228	T/C	F165Y	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs889249323	T/C	F165S	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral

rs1064793744	T/C	F167L	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs755295390	C/T	H169Y	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs753142719	A/G	H169R	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs749356977	C/G	L171V	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1564801650	A/G	T175A	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1060500109	C/T	A176V	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1554890337	A/G	I178V	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1564801672	G/C	E180D	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1414611362	C/G	N185K	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs540063602	A/G	D192G	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1064795967	T/G	D192E	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1554890324	C/A	P172T	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs754821870	A/G	Q569R	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs876661021	A/G	H570R	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs587782345	A/T	N529D	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1564570356	T/A	N529K	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1554826026	A/G	Y519C	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1554826034	G/T	E525Q	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs878853932	A/G	E526G	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs370064195	G/C/T	D470H	tolerated	benign	Non-Damaging	Neutral	Neutral	Neutral
rs1554825550	A/G	E472G	tolerated	benign	Non-Damaging	Neutral	Neutral	Diseased
rs758644748	G/A	D474N	tolerated	benign	Non-Damaging	Neutral	Neutral	Diseased
rs587780007	G/T	S478I	tolerated	benign	Non-Damaging	Deleterious	Neutral	Diseased
rs786203858	A/G	I479V	tolerated	benign	Non-Damaging	Deleterious	Neutral	Diseased
rs786201507	G/A	R481H	tolerated	benign	Non-Damaging	Deleterious	Neutral	Diseased

rs1564568473	A/G	D483G	tolerated	benign	Non-Damaging	Deleterious	Neutral	Diseased
rs863224667	A/G	D485G	tolerated	benign	Non-Damaging	Deleterious	Neutral	Diseased
rs1554825577	C/A	L489I	tolerated	benign	Probably Damaging	Deleterious	Neutral	Diseased
rs1554898209	G/A	G329R	tolerated	benign	Probably Damaging	Deleterious	Neutral	Diseased
rs1554900633	A/G	M371V	tolerated	benign	Probably Damaging	Deleterious	Diseased	Diseased
rs2943772	C/G	S65C	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs985226639	A/T	E77G	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs552470098	C/T	A79V	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1395659213	A/G	Y123C	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs786204853	G/T	L198F	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs11202592	A/T	R170W	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1238691983	C/T	R119C	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs398123324	A/G	R187G	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1064794096	G/T	R188I	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1064796078	A/T	R188S	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs786204910	T/G	Y189D	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs786201995	T/G	L196V	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs797044910	G/T	D197Y	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1320222638	A/G	D197G	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs587781957	A/C	N185T	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1114167648	A/C	H445P	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs786204875	G/T	W447L	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs587782350	C/T	P419L	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1057519368	T/G	L420S	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1114167664	G/T	G424V	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased

rs121909239	A/T	D425V	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1085308051	A/G	E330G	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs786202688	A/G	R332G	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1114167673	G/T	R332M	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs587782603	G/T	G338R	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1554900515	T/A	V339E	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs397514559	C/A	T340N	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1554900534	A/G	S343R	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs121909221	T/A	S343R	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs121913293	C/T	R346S	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs121913294	G/T	R346P	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs863224666	A/G	H314P	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs587782360	A/G	I308V	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs370795352	T/C	I308K	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs786201044	T/C	C309R	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs886047389	G/T	G21V	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs572685299	G/A	V182I	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1564801689	A/G	S183G	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1554890340	C/G	S183R	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs876660420	G/T	G193V	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1554890398	G/T	D195Y	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs876661244	A/G	D195G	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs786204912	T/G	L198V	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs786201506	T/C	L198S	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1410198544	G/A	V448I	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased

rs1554825502	A/G	M443V	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1195369834	G/T	M443I	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1057518538	A/C	Q418P	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1554898217	G/A	V331I	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs786202753	A/G	K336R	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs146629065	G/C	K337N	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1554900530	T/C	I341T	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1564837747	C/G	P342A	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs786204865	A/G	Q344P	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs757498880	A/G	Y349C	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1554898193	A/G	E323G	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs9651492	G/C	D326N	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1285259669	G/A	G35E	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1564801070	C/T	P57S	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1196170476	C/G	P57R	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs550595518	C/G	P59Q	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1441812814	C/G	P64A	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1001571902	C/T	P64R	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1401532030	C/G	A66G	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1237944553	C/T	P44L	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1433679388	C/G	T46R	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1564801018	C/T	R48W	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1031706678	C/T	P70T	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1045148218	A/G	E106G	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs904005027	C/A	P114Q	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased

rs1221231481	C/T	P114L	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1160370526	C/G	P124R	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1379620359	C/T	A126V	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1085308047	A/C	K186T	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs750098228	C/T	S164F	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs572922017	G/A	R154H	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs575260016	A/C	S117R	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1239105602	C/T	A73V	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs1355651458	C/T	P40L	deleterious	possibly damaging	Probably Damaging	Deleterious	Diseased	Diseased
rs786204863	G/T	G338V	deleterious	probably damaging	Probably Damaging	Deleterious	Diseased	Diseased

Table 4.8 Novel pathogenic SNPs of PTEN

SNP	Allele	Amino Acid	SIFT		PROVEAN	PolyPhen-2 (HumDiv)		PhD-SNP	SNPs&GO	PANTHER
			Prediction	TI	Prediction (cutoff= -2.5)	Effect	Score	Prediction		
rs1045148218	A/G	E106G	Intolerated	0.03	Deleterious	Probably damaging	0.617	Diseased	Diseased	Probably damaging
rs904005027	C/A	P114Q	Intolerated	0	Deleterious	Probably damaging	0.908	Diseased	Diseased	Probably damaging
rs1221231481	C/T	P114L	Intolerated	0	Deleterious	Probably damaging	0.856	Diseased	Diseased	Probably damaging
rs1160370526	C/G	P124R	Intolerated	0	Deleterious	Probably damaging	0.908	Diseased	Diseased	Probably damaging
rs1379620359	C/T	A126V	Intolerated	0.02	Deleterious	Probably damaging	0.691	Diseased	Diseased	Probably damaging
rs1285259669	G/A	G35E	Intolerated	0.01	Deleterious	Probably damaging	0.97	Diseased	Diseased	Probably damaging
rs1564801070	C/T	P57S	Intolerated	0	Deleterious	Probably damaging	0.81	Diseased	Diseased	Probably damaging
rs1196170476	C/G	P57R	Intolerated	0	Deleterious	Probably damaging	0.908	Diseased	Diseased	Probably damaging
rs550595518	C/G	P59Q	Intolerated	0	Deleterious	Probably damaging	0.908	Diseased	Diseased	Probably damaging
rs1441812814	C/G	P64A	Intolerated	0	Deleterious	Probably damaging	0.737	Diseased	Diseased	Probably damaging
rs1001571902	C/T	P64R	Intolerated	0	Deleterious	Probably damaging	0.908	Diseased	Diseased	Probably damaging
rs1401532030	C/G	A66G	Intolerated	0.03	Deleterious	Probably damaging	0.578	Diseased	Diseased	Probably damaging
rs1237944553	C/T	P44L	Intolerated	0.01	Deleterious	Probably damaging	0.856	Diseased	Diseased	Probably damaging
rs1085308047	A/C	K186T	Intolerated	0.01	Deleterious	Probably damaging	0.843	Diseased	Diseased	Probably damaging
rs750098228	C/T	S164F	Intolerated	0.03	Deleterious	Probably damaging	0.692	Diseased	Diseased	Probably damaging
rs572922017	G/A	R154H	Intolerated	0.01	Deleterious	Probably damaging	0.65	Diseased	Diseased	Probably damaging
rs575260016	A/C	S117R	Intolerated	0.03	Deleterious	Probably damaging	0.514	Diseased	Diseased	Probably damaging

rs1239105602	C/T	A73V	Intolerated	0.01	Deleterious	Probably damaging	0.691	Diseased	Diseased	Probably damaging
rs1355651458	C/T	P40L	Intolerated	0	Deleterious	Probably damaging	0.856	Diseased	Diseased	Probably damaging
rs786204863	G/T	G338V	Intolerated	0	Deleterious	Probably damaging	0.966	Diseased	Diseased	Probably damaging

DISCUSSION

To overcome the rising global burden of diseases including different cancers and tumors, one of the possible cause of it has been predicted in current study. Finding out single nucleotide polymorphism is the best approach to see disease-causing genes within a human body (B. Bessenyei, *et al.,* 2004). This research on SNPs has a major benefit in cancer biology because most of the SNPs in a gene results in the progression of tumor ultimately leading to cancer. The type of disease caused by mutation in gene depends upon the type of mutation, its rate and the site of substituted nucleotide (Arshad, Bhatti, & John, 2018). The selection pressure for all these mutations are not uniform. When there is a strong selective pressure for evolution of new residues then it will significantly involve in the cause of human disease. These single nucleotide polymorphism in different genes affect the biological activity of the protein expressed by mutated genes. These disease associated SNPs are related to evolution (Bao L, 2005).

The current study was carried out involve for predicting the deleterious and diseased SNPs of PTEN. It has been reported to be a tumor suppressor gene located at chromosome 10, has been reported to be frequently mutated in cancer patients (Lee, Chen and Pandolfi, 2018). PTEN has a major role in the inactivation of oncogenic PI3K/mTOR pathway and has also playing its role in the development of metastasis in brain. (Forbes et al., 2015). Harriet and his coworkers in 2016 found that overexpression of PTEN results in slow down of Akt pathway by reducing the Akt kinase activity. This shows that genetic alterations in PTEN results in developing brain metastasis in cancer patients (Frisk et al., 2012). Therefore because of

significant importance of PTEN gene, we chose PTEN gene for our study and found out at how changes in genetic makeup of this gene result in a deadly disease.

As the knowledge of database is growing continuously, the concept for understanding the function of genome is also growing rapidly. Many different changes in the genome has been discovered including different types of mutations (Bhatnager & Dang, 2018). These mutations are characterizes as disease causing mutations or neutral polymorphisms. Although there are many different experimental approaches for the characterization of specific variants, during past few years there is enormous progress in the field of bioinformatics in order to understand the changes in genetics (Bhatnager & Dang, 2018). Current study is based on use of bioinformatics to find out genetic changes. Therefore, we selected 133 PTEN SNPs to find out the most deleterious, deadly and disease causing single nucleotide polymorphisms. We found these deadly SNP with the help of bioinformatics (Song, et al., 2017).

Bioinformatics has many areas in genomic studies. The two most important areas are annotation of genetic variation data with molecular features that are supposed to affect function of protein (Capriotti E. F., 2005). The other area of research is the prediction of mutation that affect the structural and molecular properties of gene and then its expressed protein. In the current study, the molecular functions of single nucleotide polymorphisms (SNPs) is reviewed and described as deleterious and disease causing (Adzhubei I. A., *et al.,* 2010). The study involved the use of online tools that are very useful in predicting deleterious SNPs. with the help of these tools, we worked on sorting out the diseased and damage causing SNPs of PTEN gene from the neutrand tolerant ones (Adzhubei, Jordan, & Sunyaev, 2013).

Many tools have been made for finding out the mutations. Current study involved the use of different software including SIFT, Polyphen-2, Panther, PHD-SNP, Proval

ean and SNP and GO. Many analyses have been made to see substitution of an amino acid for understanding the effects of the substitution (Sim, *et al.,* 2012). These tools are based on mathematical operations designed on the basis of structure, sequence and the function of protein and gene. By the use of these tools, 80 out of 133 SNPs of PTEN retrieved from dbSNP database were predicted as deleterious and pathogenic. Out of these 80 SNPs, 35 have already been reported in literature, 22 of them are under study and their clinical significance has been uncertain yet, while 23 of them are novel. These results provide a filtered data to explore the effect of uncharacterized SNPs and their association with the disease susceptibility and to design the target dependent drugs for therapeutics.

Sorting Intolerant from Tolerant (SIFT, http://blocks.fhcrc.org/sift/ SIFT.html) and Polymorphism Phenotyping (PolyPhen, http:// genetics.bwh.harvard.edu/pph/) were used as an important tool for finding substitutions based on disease-associated human alleles (Flanagan, Patch, & Ellard, 2010). These tools are the most accepted ones and are most authentic. Before choosing a tool for use in prediction and analyses two things must be considered. First is the training sets that are used for prediction and second thing is the classification approach. Recently machine learning approaches are found to be the accurate ones (Song, et al., 2017). Current research efforts are focusing on improving accuracy to predict diseased and damaged SNPs of PTEN gene. It was found that 60% of the SNPs are predicted by all the tools used in current study to be deadly and damaging with decreased protein stability.

So characterizing the substitution of protein's amino acid is the basic step for prediction of effects of genetic variation. The current study helps on predicting the effects (Choi & Chan , 2015) of these variants to understand features important for determining molecular disturbance. SIFT for its performance uses its feature of conservation in a multiple sequence alignment. Also it make use of experimental mutations for training data (Smith & Briggs, 2016). While Polyphen uses the feature of structure of protein and human allele data for its training. Thus, there are now

many resources and tools available for the prediction of functional SNPs and the candidate gene (Adzhubei, Jordan, & Sunyaev, 2013). Also there are tools for the prediction of effects of substitution of amino acid on protein stability (Smith & Briggs, 2016). The current research covers all these areas.

Conclusion and Future Prospects

Conducted study predicted the deleterious SNPs of PTEN gene as a possible cause of tumor, cancer and other deadly diseases with the help of bioinformatics tools. The results concluded that each of the individual tools show that up to 60% of SNPs used in current study are involved in disease cause and 38% of the SNPs are predicted and hence confirmed to be deadly and disease causing by all the 7 tools used in the research. The presence of these SNPs in PTEN causes disturbance in the expression of the PTEN protein and disturb oncogenic pathways. Therefore, we can assume that these SNPs are one of the most dominating cause of cancer, tumor and other genetic diseases. In future, personalized medicine can be suggested according to the SNP that is involved in disease. Moreover, these mutations can be confirmed experimentally by taking blood and tissue sample of the patient. We can also conduct a study to find out the rate of expression of PTEN gene with or without SNPs. Clinicians can be advised to treat the patients accordingly that are found to be at a risk for developing Brain metastasis. The study large number of these variants could help to further validate the results; however, this study might act as cornerstone for future studies. Moreover, the study can further be continued by comparing the expression of PTEN gene in primary breast tumors with those of brain metastatic patients. The experiments can also further be performed on mice models to check effect of PTEN suppression on various targets in brain.

References

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, , A., Bork, P., Sunyaev, S. E. (2010). A method and server for predicting damaging missense mutations. *Nat Methods*, 7(4):248-9.

Bhatnager , R., & Dang, A. (2018). Comprehensive in-silico prediction of damage associated SNPs in Human Prolidase gene. *Scientific RePorTS* , 8:9430 | DOI:10.1038/s41598-018-27789-0.

Gerasimavicius, L., Liu, X., & Marsh, J. A. (2020). Identification of pathogenic missense mutations using protein stability predictors. *Sci Rep*, 10(1):15387.

Goswami, A. M. (2015). Structural modeling and in silico analysis of non-synonymous single nucleotide polymorphisms of human 3β-hydroxysteroid dehydrogenase type 2. *Meta Gene* , 162-172.

Mashhadi, R., Pourmand, G., Mehrsai, A., Pakdel, S., Dialameh , H., Ahmadi, A., Mahboub, R. (2014). Effect of PTEN Gene Mutations and Environmental Risk Factors on the Progression and Prognosis of Bladder Cancer. *Iranian J Publ Health*, Vol. 43, No. 1, Jan 2014, pp. 56-61 .

Naidu, C., & Y Suneetha. (2016). Prediction and Analysis of Breast Cancer Related Deleterious Non-Synonymous Single Nucleotide Polymorphisms in the PTEN Gene. *Asian Pac J Cancer Prev,*, 17 (4), 2199-2203.

AbdulAzeez, S., & Borgio, J. F. (2016). In-Silico Computing of the Most Deleterious nsSNPs in HBA1 Gene. *PLoS ONE 11(1): e0147702. doi:10.1371/*, 1-13.

Adzhubei, I., Jordan, D., & Sunyaev, S. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*, Chapter 7:Unit7.20.

Adzhubei, I., Schmidt, S., Peshkin, L., Ramensky, V., Gerasimova, A., Bork, P., Sunyaev, S. (2010). A method and server for predicting damaging missense mutations. *Nat Methods*, 7(4): 248–249.

Arshad, M., Bhatti, A., & John, P. (2018). Identification and in silico analysis of functional SNPs of human TAGAP protein: A comprehensive study. *Plos One*, 13(1): e0188143.

B. Bessenyei, M. Ma´rka, L. Urba´n , & M. Zeher & I. . (2004). Single nucleotide polymorphisms: aging and diseases. *Biogerontology* , 5: 291–303.

Bao L, Z. M. (2005). identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids*, ;33:480–482.

Bao, L., & Cui, Y. (2005). Prediction of the phenotypic effects of non-synonymous single nucleotide polymorphisms using structural and evolutionary information. *Bioinformatics*, 21(10):2185-90.

Capriotti, E. C. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*, 22:2729-2734.

Capriotti, E. F. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.*, 33 (Web server issue), W306-W310.

Capriotti, E. F. (2005). Predicting protein stability changes from sequences using support vector machines. Bioinformatics,. *Bioinformatics*, 21 (Suppl 2), ii54-ii58.

Carbognin, L., Miglietta, F., Paris, I., & Dieci, M. V. (2019). Prognostic and Predictive Implications of PTEN in Breast Cancer: Unfulfilled Promises but Intriguing Perspectives. *Cancers*.

Carter, H. (2014). Predicting the functional consequences of somatic missense mutations found in tumors. *Methods Mol Biol*, 135-59.

Carter, H., Chen, S., Isik, L., Tyekucheva, S., Velculescu, V., Kinzler, K., Karchin, R. (2009). Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Res*, 69(16):6660-7.

Choi, Y., & Chan , A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, 31(16):2745-7.

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. *Plos One*.

Erichsen, H., & Chanock, S. (2004). SNPs in cancer research and treatment. *British Journal of Cancer*, 90, 747 – 751.

Flanagan, S., Patch, A.-M., & Ellard, S. (2010). Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers*, 14(4):533-7.

Gao, C., Zhuang, J., Zhou, C., Li, H., Liu, C., Liu, L., Sun, C. (2019). SNP mutation- related genes in breast cancer for monitoring and prognosis of patients: A study based on the TCGA database. *Wiley Cancer Medicine*, 8:2303–2312.

Gao, C., Zhuang, J., Zhou, C., Li, H., Liu, C., Liu, L., Sun, C. (2019). SNP mutation- related genes in breast cancer for monitoring and prognosis of patients: A study based on the TCGA database. *Cancer Medicine*, 8; (5) 2303-2312.

Gnad, F., Baucom, A., Mukhyala, K., Manning, G., & Zhang, Z. (2013). Assessment of computational methods for predicting the effects of missense mutations in human cancers. *BMC Genomics*, 14 Suppl 3(Suppl 3):S7.

Haiman, C. A., stram, D. O., Cheng, I., Giorgi, E. E., Pooler, L., Penney, K., Freedman, M. (2006). Common Genetic Variation at PTEN and Risk of Sporadic Breast and Prostate Cancer. *Cancer Epidemiol Biomarkers*, 15(5);1021 – 5.

Hassan, M., Omer, S., Khalf-allah, R., Mustafa, R., Ali, I., & Mohamed, S. (2016). Bioinformatics Approach for Prediction of Functional Coding/Noncoding Simple Polymorphisms (SNPs/Indels) in Human BRAF Gene. *Advances in Bioinformatics*, Article ID 2632917, 15 pages.

Han, C. H. and Brastianos, P. K. (2017) „Genetic characterization of brain metastases in the era of targeted therapy“, Frontiers in Oncology. Frontiers Media S.A. doi: 10.3389/fonc.2017.00230.

Hu, J., & Ng, P. C. (2013). SIFT Indel: predictions for the functional effects of amino acid insertions/deletions in proteins. *PLoS One*, 8(10):e77940.

Kaman, T., Karasakal, Ö. F., Oktay, E. Ö., Ulucan, K., & Konuk, M. (2019). In silico approach to the analysis of SNPs in the human APAF1 gene. *Turkish Journal of Biology*, 371-381.

Katsonis, P., Koire, A., Wilso, S. J., Hsu, T.-K., C Lua, R., Wilkins, A. D., & Lichtarge, O. (2014). Single nucleotide variations: Biological impact and theoretical interpretation. *Protein Sci*, (12): 1650–1666.

Katsonis, P., Koire, A., Wilson, S. J., Hsu, T.-K., Lua, R. C., Wilkins, A. D., & Lichtarge, O. (2014). Single nucleotide variations: biological impact and theoretical interpretation. *Protein Sci*, (12):1650-66.

Kim, Y., Kang, C., Min, B., & Yi, G.-S. (2015). Detection and analysis of disease-associated single nucleotide polymorphism influencing post-translational modification. *BMC Med Genomics*, ;8.

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*, 4(7):1073-81. PMID: 19561590.

Leslie, N., & Longy, M. (2016). Inherited PTEN mutations and the prediction of phenotype. *Seminars in Cell & Developmental Biology*, 30-38.

Lee, Y.-R., Chen, M. and Pandolfi, P. P. (2018) „The functions and regulation of the PTEN tumour suppressor: new modes and prospects“, Nature Reviews Molecular Cell Biology. Nature Publishing Group, 19(9), pp. 547–562.

Li, H.-G., Liu, F.-F., Zhu, H.-Q., Zhou, X., Lu, J., Chang, H., & Hu, J.-H. (2015). Association of PTEN gene polymorphisms with liver cancer risk. *Int J Clin Exp Patho*, 8(11): 15198–15203.

Miosge, L., Field, M., Sontani , Y., Cho, V., Johnson, S., Palkova, A., Andrews, T. (2015). Comparison of predicted and actual consequences of missense mutations. *Proc Natl Acad Sci U S A*, 112(37):E5189-98.

Mooney, S. (2004). Bioinformatics approaches and sources for Single Nucleotide Polymorphism Functional analysis. *Briefings In Bioinformatics*, VOL 6. NO 1. 44–56.

Mooney, S., Krishnan,, V., & Evani, U. (2010). Bioinformatic Tools for Identifying Disease Gene and SNP Candidates. *Methods Mol Biol*, 628: 307–319.

Mohammed Arifeen (2018) Review of breast cancer risks in Pakistan - Pakistan & Gulf Economist. http://www.pakistaneconomist.com/2018/07/23/reviewof-breast-cancer-risks-in-pakistan/ (Accessed: 9 November 2019).

Ng , P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acid Reaearch*, 31(13):3812-4. doi: 10.1093/nar/gkg509.

Ng , P., & Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Research*, 11(5):863-74.

Park, M. K. et al. (2019) „PTEN self-regulates through USP11 via the PI3K-FOXO pathway to stabilize tumor suppression“, Nature communications. Nature Publishing Group, 10(1), p. 636.

S Nakken, & T Rognes, A. (2007). Computational prediction of the effects of non-synonymous single nucleotide polymorphisms in human DNA repair genes. *Neuroscience*, 145(4):1273-9.

Shastry, B. S. (2007). SNPs in disease gene mapping, medicinal drug development and evolution. *Journal of Human Genetics volume 52,* 871–880.

Shastry, B. S. (2007). SNPs in disease gene mapping, medicinal drug development and evolution. *Journal of Human Genetics*, volume 52, pages871–880.

Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & C. Ng, P. (2012 Ju). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, 40(Web Server issue): W452–W457.

Smith, I., & Briggs, J. (2016). Structural Mutation Analysis of PTEN and its Genotype Phenotype Correlations in Endometriosis and Cancer. *Proteins*, 84(11): 1625–1643.

Song, D.-D., Zhang, Q., Li, J.-H., Hao, R.-M., Ma, Y., Wang, P.-Y., & Xie, S.-Y. (2017). Single nucleotide polymorphisms rs701848 and rs2735343 in PTEN increases cancer risks in an Asian population. *Oncotarget*, 96290-96300.

Yang, J.-M. et al. (2017) „Characterization of PTEN mutations in brain cancer reveals that pten mono-ubiquitination promotes protein stability and nuclear localization", Oncogene. Nature Publishing Group, 36(26), p. 3673.