

IN SILICO ANALYSIS FOR THE PREDICTION OF FUNCTIONAL nsSNPS IN ARID1A GENE OF PI3K-AKT PATHWAY

Submitted in partial fulfilment of the requirements for the award of
Master of Science Degree in Bioinformatics and Data Science

By

NANTHINI M (42739017)



**DEPARTMENT OF BIOINFORMATICS
SCHOOL OF BIO AND CHEMICAL ENGINEERING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade “A++” by NAAC | 12B Status by UGC | Approved by AICTE
JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600119**

APRIL – 2024



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited "A++ Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

DEPARTMENT OF BIOINFORMATICS

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **Nanthini M (42739017)**, who carried out the project entitled "*In Silico analysis for the prediction of functional nssnps in ARID1A gene of pi3k-akt pathway*" under our supervision from **January 2024 to April 2024**.

Internal Guide

Dr. Swetha Sunkar

External Guide

Dr. Swetha Sunkar

Head of the Department

Dr. Jemmy Christy

Submitted for Viva voce Examination held on 20/04/2024

Internal Examiner

External Examiner

DECLARATION

I, **NANTHINI.M (42739017)**, hereby declare that the Project Report entitled "***In Silico analysis for the prediction of functional nssnps in ARID1A gene of pi3k-akt pathway***" done by me under the guidance of **Dr.Swetha Sunkar** at **Sathyabama Institute of Science and Technology, Chennai** and is submitted in partial fulfilment of the requirements for the award of Master of Science degree in Bioinformatics and Data Science.

DATE: 20/04/2024

PLACE: CHENNAI

M. Nanthini

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **the Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. REKHA CHAKRAVATHY**, Dean of Science and Humanities, I am deeply grateful for her mentorship and leadership, which have been a constant source of inspiration. Her dedication to fostering a conducive academic environment has been instrumental in the successful completion of this thesis.

I convey my thanks to **Dr. JEMMY CHRISTY**, Head of the Bioinformatics Department, for providing me the necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. SWETHA SUNKAR** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of Bioinformatics who were helpful in many ways for the completion of the project.

ABSTRACT

Cancer is a harmful diseases involving abnormal cell growth with the potential to invade or spread to other parts of the body. The most common changes occurs in genes so that it will control the cell growth and division due to cancer. PI3K-AKT Pathway is an intracellular signal transduction pathway that promotes metabolism, proliferation, and cell survival. The studies revealed that if mutations in genes of PI3K-AKT pathway becomes dysregulated, it contributes to the development and progression of cancer. AT-rich interaction domain 1A (ARID1A) generally acts as a tumor suppressor gene. In our study, loss of ARID1A function leads to indirectly activating the PI3K-AKT Pathway and was focused onto analyze the potentially deleterious non synonymous single nucleotide polymorphisms. From the dbSNP database, the SNPs of ARID1A gene was retrieved. These SNPs were analyzed using various computational tools viz., PolyPhen-2, I-mutant2.0, dynamut, mutpred2, Panther, SIFT for their pathogenicity and stability analysis to determine whether they alter the protein structure and function. The pathogenicity analysis revealed that 155 mutations were deleterious and further stability analysis revealed 41 mutations decreased the stability. At this point, the ARID domain region was considered (1017-1108 amino acids) where the mutations D1697V, VI817I, L1831V, P1966S, N1986S, R2158Q, H2004Q, S2262L were found to be potentially deleterious. This protein was subjected to homology modeling using SWISS MODEL. This ARID1A protein neither have a binding site nor a compound. Consurf has been performed for further analysis. Selected nsSNPs analyzed with ConSurf revealed that they were positioned in the conserved regions, and the tool predicted that N1986, H2004Q, R2158Q, S2262L were highly conserved. These mutations may involved in PI3K-AKT Pathway and thus indirectly leads to the progression of cancer.

Keywords: In silico analysis; ARID1A ; PI3K-AKT Pathway; nsSNPs; Homology modeling.

TABLE OF CONTENTS

S.NO	TOPICS	PAGE NUMBER
	ABSTRACT	i
	LIST OF CONTENTS	ii
	TABLES	v
	FIGURES	vi
	ABBREVIATIONS	vii
1	INTRODUCTION	1
1.1	CANCER	1
1.2	SIGNS AND SYMPTOMS	1
1.3	ARID1A	2
1.4	ARID1A AND SWI/SNF COMPLEXES	2
1.5	MECHANISMS OF ARID1A REGULATION	3
1.6	TUMOR INITIATION AND SUPPRESSION OF ARID1A	3
1.6.1	TUMOR INITIATION BY ARID1A	3
1.6.2	TUMOR SUPPRESSION BY ARID1A	4
1.7	ROLE OF ARID1A IN CANCER	4
1.7.1	ARID1A IN GYNECOLOGICAL CANCER	4
1.7.2	ARID1A IN BREAST CANCER	5
1.7.3	ARID1A IN HEPATOCELLULAR CARCINOMA	5
1.7.4	ARID1A IN PANCREATIC CANCER	5
1.7.5	ARID1A IN COLON CANCER	5
1.8	PI3K-AKT PATHWAY	6

1.8.1	ROLE OF ARID1A GENE IN PI3K-AKT PATHWAY	6
2	REVIEW OF LITERATURE	8
3	AIMS AND SCOPE OF PRESENT INVESTIGATION	10
4	MATERIALS AND METHODOLOGY	11
4.1	DATA MINING	11
4.2	IDENTIFYING THE NONSYNONYMOUS AND DELETERIOUS SNPS	11
4.3	PREDICTION OF DELETERIOUS NSSNPS BY DIFFERENT BIOINFORMATICS TOOLS	11
4.3.1	SIFT	11
4.3.2	POLYPHEN-2	12
4.3.3	PANTHER	12
4.3.4	MUTPRED2	12
4.3.5	ANALYZING THE EFFECTS OF nsSNPs ON STABILITY OF PROTEIN	13
4.3.6	SNP&GO	13
4.4	HOMOLOGY MODELLING AND STRUCTURAL ANALYSIS:	14
4.5	PHYLOGENETIC CONSERVATION ANALYSIS OF nsSNPs BY CONSURF	14
5	RESULTS AND DISCUSSION	15
5.1	IDENTIFICATION OF FUNCTIONALLY SIGNIFICANT SNPS	15
5.2	EFFECTS OF NSSNPS ON STABILITY OF PROTEIN	18
5.3	EFFECTS OF SELECTED NSSNPS ON PROTEIN STABILITY	21
5.4	ENERGY PARAMETERS OF THE NATIVE AND MUTANT MODELS:	22
5.5	HOMOLOGY MODELING AND STRUCTURAL ANALYSIS OF NATIVE AND MUTANT MODELS	23
5.6	CONSERVATION ANALYSIS BY CONSURF	26

6	SUMMARY & CONCLUSION	28
7	REFERENCE	29

LIST OF TABLES

S.NO	TOPIC	PAGE NO
5.1	Deleterious nsSNPs in ARID1A gene	20
5.2	Effects of nsSNPs on protein stability using SNP&GO and I-MUTANT 2.0	24
5.3	Effects of selected nsSNPs on protein stability using Dynamut.	26
5.4	Energy parameters of the native and mutant models	27
5.5	Pairwise structural alignment results	28
5.6	Conservation analysis by consurf	32

LIST OF FIGURES

S.NO	PARTICULARS	PAGE NO
1	ARID1A in PI3K-AKT pathway	20
2	Homology modeled 3d structure of arid1a gene	24

LIST OF ABBREVIATIONS

ARID1A	AT-rich interacting domain-containing protein 1A gene
SWI/SNF	Switch / sucrose non fermenting
ATP	Adenosine triphosphate
RNA	Ribo nucleic acid
DNA	Deoxyribo nucleic acid
AT	Adenine thymine
BAF	B allele frequency
LXXLL	leucin rich motifs
TF	transcription factors
DDR	DNA damage response
TIME	tumor immune microenvironment
HCC	hepatocellular carcinoma
APC	antigen presenting cells
OCCC	ovarian clear cell carcinoma
PDAC	pancreatic ductal adenocarcinoma
ER	estrogen receptor
PI3K	phosphatidylinositol 3-kinase
AKT	protein kinase B

CHAPTER - 1

INTRODUCTION

1.1 CANCER:

Cancer is one of the main causes of death worldwide, and in the past decade, many research studies have focused on finding new therapies to reduce the side effects caused by conventional therapies. During cancer progression, tumors become highly heterogeneous, creating a mixed population of cells characterized by different molecular features and diverse responsiveness to therapies. This heterogeneity can be appreciated both at spatial and temporal levels and is the key factor responsible for the development of resistant phenotypes promoted by a selective pressure upon treatment administration. Usually, cancer is treated as a global and homogeneous disease and tumors are considered as a whole population of cells. (Dagago et al., 2018).

1.2 SIGNS AND SYMPTOMS:

- A sign is observable by others, such as fast breathing, fever, or abnormal lung sounds detected with a stethoscope, providing external evidence of a disease. On the other hand, a symptom is experienced by the individual, like weakness, aches, or shortness of breath, which may not be apparent to others, reflecting internal sensations related to a disease.
- Cancer often presents with nonspecific symptoms, underscoring the importance of appropriate cancer screening to mitigate risk factors. Screening protocols are tailored to different age groups, and individuals with risk factors like smoking, alcohol consumption, sun exposure, or genetic predispositions should be vigilant about potential cancer symptoms. Regular screenings and heightened awareness of symptoms are vital in early cancer detection and improving treatment outcomes. The best way to fight cancers is only by early detection and by eliminating and decreasing risk factors. (Alwan et al., 2001).

1.3 ARID1A :

- The AT-rich interacting domain-containing protein 1A gene (ARID1A) encodes ARID1A, a member of the SWI/SNF chromatin remodeling complex. ARID1A is located on chromosome 1p36.11 and encodes ARID1A, a core member of the SWI/SNF complex. This complex plays a crucial role in carcinogenesis, and thus ARID1A is viewed as a cancer-inhibiting gene, of which mutation may be involved in onset and progression of several cancers. (Takeda et al., 2015)
- Dysregulation of ATP-dependent chromatin remodeling complexes has been implicated in a variety of cancers in an increasing number of studies. These complexes utilize the energy of ATP hydroxylation to reposition, eject, or exchange nucleosomes, thereby modulating DNA accessibility to other cellular machineries involved in transcription, DNA replication, methylation, and repair. Among different ATP-dependent chromatin remodelers, SWItch /Sucrose Non Fermentable(SWI/SNF) is the most commonly dysregulated in cancer, and ARID1A (AT-rich interactive domain-containing protein 1A) is the most frequently mutated among all genes encoding subunits of SWI/SNF complexes. (Chin Wu et al., 2014)
- Molecular mechanisms for low ARID1A mRNA and protein expression appear to be different among various types of malignant tumors. For example, ARID1A gene mutations result in the low protein expression in ovarian clear cell carcinoma (with mutation rate ~50%); however, in pancreatic cancer, frequent copy number loss (47%) is the main cause. (Zhang et al., 2013)

1.4 ARID1A protein and SWI/SNF complexes:

- ARID1A is a member of a large family of proteins that contain a characteristic ARID domain of approximately 100 amino acids, which binds DNA fragments rich in adenine-thymine (AT) and was originally discovered in 1995 in the Drosophila dead ringer protein and in the murine B cell-specific transcription factor Bright. ARID1A is the largest, non-catalytic BAF subunit, with the main property of

conferring target specificity on the complex and directing the ATPase activity, as the ARID domain binds across to at least 50 bp of specific nucleosomal DNA constituting the origin of chromatin remodeling. Its guiding effect, according to the prevailing theory, is attributed on the one hand to the interactions of the protein with the transcription factors, the hormone nuclear receptors and the p53, p21 (CDKN1A), SMAD3 proteins, via the C-terminal peptide-rich binding loci (LXXLL-leucin rich motifs), while on the other hand to the ARID domain-mediated high affinity between chromatin and SWI/SNF. Recently, ARID1A has also been found to be indirectly involved in the modification of histones, by binding to histone H2B as E3 ligase of ubiquitin. (Pavlidou et al., 2020)

1.5 MECHANISMS OF ARID1A REGULATION:

ARID1A mutations result in a series of changes ranging from cancer inhibition to cancer promotion that could be distinguished as follows:

- Damaging the binding of transcription factors (TFs) to DNA and changing the acetylation and methylation states of histones.
- Affecting DNA damage response (DDR) and a series of subsequent steps, such as DNA repair and cell-cycle checkpoint activation.
- Influencing antitumor immunity and tumor immune microenvironment (TIME), resulting in immune escape.
- Affecting other signaling pathways. In the following sections, we mainly focus on the regulatory mechanisms of ARID1A.

1.6 TUMOR INITIATION AND SUPPRESION OF ARID1A:

1.6.1 TUMOR INITIATION OF ARID1A:

- It is found that 83% of hepatocellular carcinoma (HCC) tumors possess higher ARID1A expression levels than adjacent normal tissues, while in endometrial cancer, the primary tumors have wild-type ARID1A, whereas the metastatic subclones carry deleterious mutations, indicating ARID1A may be required in primary cancers. For example, under the condition of inactivation of APC and PTEN, loss of Arid1a impairs the formation of ovarian cancer, and in an APC-inactivated mouse model of colon

cancer, ARID1A is found to be essential for tumorigenesis. More recently, a study by Sun et al. demonstrated that in multiple in vivo liver cancer models, ARID1A clearly exerts tumor-promoting functions during the early phases of transformation. Thus, ARID1A has a context-dependent oncogenic role in cancer. (Xu Tang et al.,2021)

1.6.2 TUMOR SUPPRESSION OF ARID1A:

- Although ARID1A exerts cancer initiation function in certain contexts, particularly in multiple in vivo liver cancer models where ARID1A clearly presents tumor-promoting functions, the majority of data recently indicate that mutations in ARID1A usually lead to its loss of function, which makes ARID1A to be reconsidered as a tumor suppressor in a broader variety of cancers. Therefore, by discussing the involvement of ARID1A in biological processes linked to tumor suppression such as tumor formation, tumor cell proliferation, tumor cell cycle, and DNA damage repair, we here in summarize the impacts of ARID1A mutations in different cancers, providing a more nuanced understanding of the carcinogenic function of this frequently altered gene. (Xu Tang et al.,2021)

1.7 ROLE OF ARID1A IN CANCER:

1.7.1 ARID1A IN GYNECOLOGICAL CANCER:

- Among gynecological cancers, ARID1A has the highest mutation rate in ovarian epithelial carcinomas (EOC). High-grade serous, low-grade serous, endometrioid, ovarian clear cell and mucinous carcinoma are the five main histological subtypes of EOCs . The ARID1A mutation rates are 40–67% in ovarian clear cell carcinoma (OCCC) and 30–36% in endometrioid ovarian cancer (EnOC) . However, high-grade serous ovarian cancer does not exhibit ARID1A mutations. Furthermore, loss of ARID1A gene expression occurs not only in ovarian cancer but also in adjacent atypical endometriosis, whereas ARID1A expression can be found in distant endometriotic lesions. Surprisingly, the ARID1A status alone does not provide any independent prognostic value in endometriosis-associated ovarian carcinomas. In conclusion, ARID1A mutations exert a significant role in tumor initiation and may transform endometriosis into ovarian carcinoma at the early stage of cancerization.

1.7.2 ARID1A IN BREAST CANCER:

- Breast cancer can be divided into various molecular subtypes according to the expression of the estrogen receptor (ER), progesterone receptor, and/or human epidermal growth factor -receptor 2 (HER2). These subtypes correspond to various clinical outcomes and treatment choices. Mutations in ARID1A are the most common alterations of the SWI/SNF complex in breast cancer. Patients with luminal A or HER2-rich breast cancer have poorer overall survival (OS) and disease-free survival (DFS) rates when the expression of ARID1A levels is reduced.

1.7.3 ARID1A IN HEPATOCELLULAR CARCINOMA:

- The progression of hepatocellular carcinoma (HCC) is inhibited by chromatin remodeling, which is ARID1A-dependent. ARID1A mutation rates in HCC are 10–17%. A previous study indicated that ARID1A levels were significantly decreased relative to the aggressive phenotype of HCCs and associated with overall metastasis and poor prognosis. In HCC, mutations in the PI3K/AKT/mTOR signaling pathway are mutually exclusive with mutations in the SWI/SNF chromosomal remodeling complex. ARID1A protein levels were negatively regulated by mTOR signaling enhancing tumor development in vivo. However, the role of ARID1A is still controversial in HCC.

1.7.4 ARID1A IN PANCREATIC CANCER:

- The ARID1A mutation rates in pancreatic ductal adenocarcinoma (PDAC) are 6–13%. Mutations in ARID1A genes tend to exhibit a shorter DFS and OS in PDAC. Moreover, it was shown that tumor differentiation and TNM stage were correlated with ARID1A expression levels. The effect of somatic mutations in ARID1A showed different clinical outcomes, and patients with ARID1A mutations had shorter median survival times than wild-type individuals.

1.7.5 ARID1A IN COLON CANCER:

- ARID1A mutations are present in 10% of colorectal cancers and, similar to GC, are thought to be caused by mismatch defects. The downregulation of ARID1A is thought to induce colorectal carcinoma possibly through its regulation of proliferation and chemoresistance in colorectal cancer cells. In terms of mechanism, ARID1A usually

targets the SWI/SNF complexes to enhancers, thereby cooperating with transcription factors to promote gene activation. ARID1A loss changes not only the level of histone modifications at enhancers but also their locations.

1.8 PI3K-AKT PATHWAY:

- Activation of the phosphatidylinositol 3-kinase (PI3K)/AKT pathway supports multiple mechanisms responsible for cancer progression, including proliferation, inhibition of apoptosis, cell adhesion and transformation. PI3K activation, e.g., by activating mutations of PIK3CA encoding the catalytic subunit, p110alpha, leads to an activation of AKT, a serine-threonine kinase, which is present in three different isoforms (AKT1-3) in human cancer and leads to increased cellular growth and survival of cancer cells. (Samartzis et al., 2013)
- The PI3K-AKT pathway appears to be a key player in mediating neuronal survival in a range of situations. Several signaling cascades, such as the Ras-mitogen-activated protein kinase (MAPK), the cAMP/protein kinase A (PKA), and PI3K-AKT pathways, are triggered by trophic nutrients such as NGF, insulin-like growth factor I, or BDNF.
- Class I PI3K is the most abundant of the four classes of PI3K. It plays a central role in cell survival, growth, proliferation, autophagy, differentiation, and metabolism by converting phospho-inositide 4,5-bisphosphate to phospho-inositide 3,4,5-triphosphate in various cellular membranes. Class I PI3K enzymes have a catalytic and regulatory subunit. The binding of regulatory subunits to catalytic subunits stabilizes the catalytic subunit proteins and permits the precise regulation of their enzymatic activity. (Hein et al., 2024)

1.8.1 ROLE OF ARID1A IN PI3K-AKT PATHWAY:

- ARID1A can interact with other signaling pathways of which the most prevalent is the PI3K/AKT/mTOR pathway. In both ovarian carcinomas and ECs, ARID1A mutations are associated with PI3K/AKT/mTOR signaling pathway activation. The reduced level of ARID1A promotes cell proliferation via the TGF- β 1/SMAD3 cascade and serves as a prognostic biomarker and therapeutic target for acute myeloid leukemia (AML). (Xuewei et al., 2023)

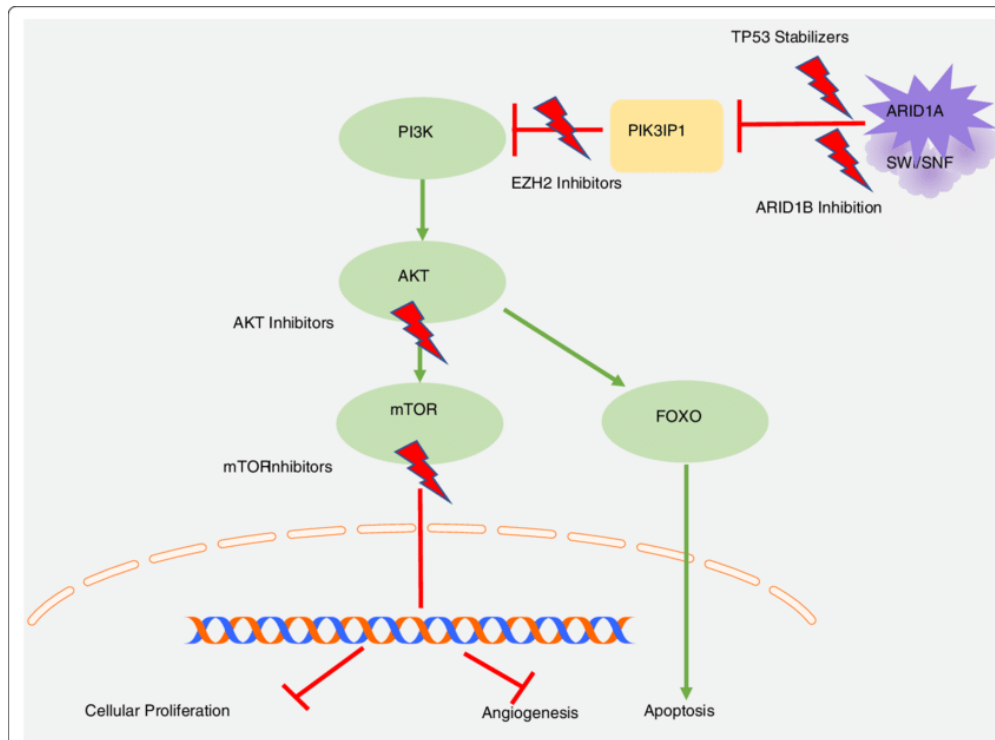


Figure 1: ARID1A in PI3K-AKT pathway

CHAPTER 2

REVIEW OF LITERATURE

- Yadegari & Majidzadeh (2019) in their study address that deleterious variants were mostly enriched in the N- and C-terminal domain of the BRCA1 and BRCA2 C-terminus and utilizing Fold X , they demonstrated that alterations in the function of proteins are not always together with stability alterations.
- Alsheikh & Ameer (2023) investigated that twenty novel SNPs regarded to be the most damaging SNPs that affect structure and function of FLT3 gene using different bioinformatics algorithm. 69 functional classes were predicted in 12 SNPs in the 3'UTR, among them, 31 alleles disrupted a conserved miRNA site and 37 derived alleles created a new site of miRNA.
- Arifuzzaman & Mitra (2020), review highlights a total of four mutations (N19D, A29T, K54N, and S71L) were found to be highly deleterious by all the tools, which are located at highly conserved regions. Furthermore, all four mutants showed structural alterations, and the communities of amino acids for mutant proteins were readily changed, compared to the wild-type.
- Mondal et al.,(2021) conducted a study to investigate a total 349 nsSNPs of CTNNB1 were subjected to extensive sequence and structure based analyses using state-of-art computational tools. The sequence based screening of 349 nsSNPs had revealed 79 highly deleterious nsSNPs of CTNNB1 protein, the positions for 74 nsSNPs were highly conserved. The sequence based stability analysis revealed decreased stability for 275 variants.
- Zhang & Huang (2020),demonstrated a total of 379 nsSNPs of GJA3 were identified. A total of 88 high-risk pathogenic GJA3 nsSNPs were found, including 31 published nsSNPs associated with congenital cataracts and 57 novel nsSNPs predicted by all eight online tools. The 88 high-risk pathogenic mutations, which are related to 67

amino acids in the wild-type sequences, cause a decrease in protein stability according to I-Mutant 3.0, MUpro and INPS. G2 and R33 were predicted to participate in post- translational modification and ligand binding by ModPred, RaptorX Binding and COACH.

- Xu et al., (2023), investigated 10 out of 48 missense SNPs with rsIDs were found deleterious. The substitution of alanine for proline at position 146 pushed several residues toward the center of the protein. Arginine instead of leucine has a minor effect on protein structure and stability.
- Subbiah et al.,(2020) conducted a study to investigate ten nsSNPs of the DEFB1 gene were found to be potentially deleterious. I-Mutant server showed that nsSNPs rs140503947 and rs146603349 decreased stability of the protein, and ConSurf analysis revealed that SNPs were located in conserved regions. The physiochemical properties of the polymorphic amino acid residues and their effect on structure were determined by Project HOPE.
- Hasnain & Shohaib (2020) review the highlights that the bunch of tools including SNP NEXUS, SNAP-2, PhD-SNP, SNPs&GO, I-Mutant, ConSurf, and ModPred were used to predict 23 highly confident damaging and disease causing nsSNPs that could potentially alter the SLC26A4 gene. Moreover, protein structure prediction, protein-ligand docking and Molecular Dynamics simulation were performed to confirm the impact of two evident alterations (Y127H and G334A) on the protein structure and function.
- Saxena et al.,(2021) investigated that four highly deleterious nsSNPs (L36R, Y159C, W162C and L345P) were identified through SNP screening using several sequence-based and structure-based tools. Conservation analysis and oncogenic analysis showed that most of the nsSNPs are at highly conserved residues, oncogenic in nature and cancer drivers. Molecular dynamics simulations (MDS) analysis was performed to understand the dynamic behaviour of native and mutant proteins. PTM analysis revealed that the nsSNP Y159C is at a PTM site and will mostly affect phosphorylation at that site. Based on the overall analyses carried out in this study, L36R is the most deleterious mutation amongst the aforementioned deleterious mutations of GOT1.

CHAPTER 3

AIMS AND SCOPE OF PRESENT INVESTIGATION

AIM:

The aim of this study is to predict the effect of mutations of genes involved in the PI3K-AKT pathway and their structural and functional impact on their corresponding protein.

OBJECTIVES:

- To retrieve the genes information from NCBI database.
- To retrieve the SNPs information like reference sequence number from SNP database and protein sequence from swissprot.
- To identify the nonsynonymous and deleterious SNPs.
- To predict using computational tools include SIFT, POLYPHEN2, PANTHER, MUTPRED2, DYNAMUT, SNP&GO.
- To analyze the effects of nssnps on stability of protein using I – MUTANT 2.0.
- To perform homology modelling and structural analysis using SWISS-MODEL.
- To predict the ligand binding sites.

CHAPTER 4

MATERIALS AND METHODOLOGY

4.1 DATA MINING:

- Collect the information about the genes from National Center of Biotechnology Information(NCBI) database.
- Retrieve the information of SNPs(reference sequence number) and protein Sequence (accession number) of genes from SNP database and swiss prot respectively.

4.2 IDENTIFYING THE NONSYNONYMOUS AND DELETERIOUS SNPs:

- Using different computational tools, coding non synonymous single nucleotide polymorphism(nsSNPs) has been examined.
- These results shows that which genes are potential, deleterious and nonsynonymous that are mutated and leads to a dangerous disease.
- To identify these nsSNPs , computational tools like SIFT, POLYPHEN-2, I-MUTANT 2.0, PANTHER, Dynamut2, Mutpred2 are used.

4.3 PREDICTION OF DELETERIOUS nsSNPs BY DIFFERENT BIOINFORMATICS TOOLS:

4.3.1 SIFT:

- SIFT stands for Sorting Intolerant From Tolerant.
- SIFT tool is used to predict whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids.
- SIFT can be applied to naturally occurring non-synonymous polymorphisms and laboratory-induced missense mutations.
- Reference SNP ids can be used as input.

4.3.2 POLYPHEN-2:

- Polymorphism Phenotyping 2 (PolyPhen-2) is a bioinformatics tool that predicts the impact of a single amino acid substitution on the structure and function of human proteins.
- It is used for predicting Functional Effect of Human Missense Mutations.
- Protein accession number followed by amino acid substitution represented in each row and submitted as query.

4.3.3 PANTHER:

- PANTHER stands for Protein Analysis through Evolutionary Relationship.
- It was designed to classify proteins (and their genes) in order to facilitate high-throughput analysis.
- It was used to classify the proteins on the basis of their evolutionary relationship, molecular functions and interaction with other proteins.
- Analysis was based on substitution position specific evolutionary conservation score, calculated from alignment of different evolutionarily related proteins.

4.3.4 MUTPRED2:

- MutPred2 is a machine learning-based method and software package that integrates genetic and molecular data to reason probabilistically about the pathogenicity of amino acid substitutions.
- It's a computational tool or method used in bioinformatics and genomics to predict the functional effects of genetic mutations.
- MUTPRED2 aims to predict whether a given mutation in a protein-coding gene will affect protein structure, function, or interactions, and to what extent.
- This kind of information is valuable in understanding the molecular basis of genetic diseases, identifying potential drug targets, and guiding personalized medicine approaches.

4.3.5 ANALYZING THE EFFECTS OF *ns*SNPs ON STABILITY OF PROTEIN:

- DynaMut, a tool that incorporates information on protein dynamics and structural environment properties of wild-type residue with our graph-based signatures approach to provide an accurate prediction of mutation effects on stability and dynamics for single and multiple point mutations.
- DynaMut is a user friendly web server that implements two distinct well established normal mode approaches, which can be used to analyse protein dynamics by sampling conformations as well as assess the impact of mutations on protein stability and dynamics.
- I-Mutant 2.0 is a support vector machine (SVM) - based tool used to predict the protein stability changes caused due to SNPs.
- Protein sequence or a structure along with amino acid substitution can be submitted as input.

4.3.6 SNP&GO

- SNP&GO is a web server designed for predicting human disease-related single point protein mutations. This method utilizes support vector machines to analyze protein sequences, protein sequence profiles, and protein function, achieving an accuracy of 82% and a Matthews correlation coefficient of 0.63. SNPs&GO integrates information from protein sequences, profiles, and functions to predict whether a mutation is disease-related or not. It was trained on over 33,000 mutations and tested using a rigorous cross, validation process, demonstrating improved performance with increased input complexity. By incorporating data from the Gene Ontology annotation, specifically the LGO score derived from the GO database, SNPs&GO enhances its ability to distinguish disease-related mutations from neutral ones. The method's performance improves as more information is included in the analysis, highlighting the effectiveness of support vector machines in capturing correlations within diverse datasets. SNPs&GO stands out as one of the top classifiers for predicting the disease relevance of protein mutations based on in-house benchmarking studies.

4.4 HOMOLOGY MODELLING AND STRUCTURAL ANALYSIS:

- Homology modeling was performed, to analyze the quality of predicted models and their possible 3D structure.
- Since, the crystal structures of selected proteins except ARID1A protein were not present in the PDB (Protein data bank) therefore, homology modelling approach with SWISS-MODEL was carried out to predict their possible 3Dstructures.
- Other structural properties like RMSD, free energy, energy minimization etc. were carried out and visualized under the Swiss-PDB viewer, PyMOL.

4.5 PHYLOGENETIC CONSERVATION ANALYSIS OF nsSNPs BY CONSURF:

The extent of evolutionary rate of conservation of amino acids in ARID1A was analyzed using the ConSurf web server. ConSurf calculates conservation scores of the amino acid position of the protein based on the phylogenetic relation between homologous sequences. ConSurf predicts conservation scores ranging from 1 to 9 with grade 9 being the most conserved position.

Chapter 5

RESULTS AND DISCUSSION

5.1 IDENTIFICATION OF FUNCTIONALLY SIGNIFICANT SNPs:

The reference UniProt ID of the protein is O14497 whose length is 2285 amino acids. The SNPs for ARID1A were retrieved using dbSNP. The functional effect of coding region nsSNPs in the genes selected were analyzed using following tools. The Panther tool utilizes sequence homology and the Panther-PSEP method, employing hidden Markov model families to assess the effect of non-synonymous single nucleotide polymorphisms (nsSNPs). The tool categorizes the impact of these SNPs as probably damaging, possibly damaging, or probably benign, based on their evolutionary conservation.

On the other hand, the Polyphen2 tool conducts functional annotation of SNPs by mapping them to gene transcripts, incorporating SwissProt annotations, and constructing conservation profiles. It assigns Position Specific Independent Counts (PSIC) scores to each substitution, with values below 0.05 classified as benign and values above 0.05 categorized as probably damaging or possibly damaging. SIFT tool is based on sequence homology, scores assessment based on position – specific scoring matrices. Score ranges from 0 to 1, where score greater than or equal to ($\geq$) 0.05 is damaging and score lesser than or equal to ($\leq$) 0.05 is tolerated.

Table 5.1: Deleterious nsSNPs in ARID1A gene

SUBSTITUTION	SIFT SCORE	POLYPHEN 2 SCORE	PANTHER SCORE	PANTHER PROBABILITY	MUTPRED2
P392H	0.028	0.999	0.57	Probably damaging	0.296
G14R	0.005	0.994	0.27	Probably benign	0.449
G397R	0.01	0.999	0.5	Possibly damaging	0.476
Q403H	0.015	0.998	0.57	Probably damaging	0.551
P504L	0.007	0.483	0.5	Possibly damaging	0.174

P121L	0.014	0	0.27	Probably benign	0.437
P516L	0.009	0.335	0.57	Probably damaging	0.144
A578V	0.035	0.957	0.5	Possibly damaging	0.135
P697L	0.003	1	0.74	Probably damaging	0.624
P773L	0.002	0.9	0.74	Probably damaging	0.502
P798L	0.003	0.986	0.74	Probably damaging	0.312
G839V	0.022	0.191	0.27	Probably benign	0.406
A900T	0.026	0.984	0.5	Possibly damaging	0.268
N906K	0.042	0.997	0.74	Probably damaging	0.662
R1026C	0.005	1	0.74	Probably damaging	0.818
G1037R	0.044	1	0.57	Probably damaging	0.693
S1123C	0.021	0.998	0.5	Possibly damaging	0.191
S740C	0.026	0.993	0.5	Possibly damaging	0.2
A1136V	0.011	0.997	0.74	Probably damaging	0.328
P1175L	0.004	1	0.74	Probably damaging	0.501
A1190G	0.028	0.435	0.5	Possibly damaging	0.156
Y1226C	0.013	1	0.57	Probably damaging	0.623
D1231V	0.003	0.999	0.74	Probably damaging	0.723
R1262H	0.037	0.998	0.57	Probably damaging	0.307
R1276Q	0.021	0.999	0.74	Probably damaging	0.475
K1370R	0.042	0.997	0.74	Probably damaging	0.293
Y1052C	0.021	1	0.5	Possibly damaging	0.822

P1451H	0.001	0.999	0.5	Possibly damaging	0.41
P1467R	0.013	0.003	0.5	Possibly damaging	0.302
R1551C	0.005	1	0.5	Possibly damaging	0.338
R1593W	0.001	0.999	0.74	Probably damaging	0.287
D1697V	0	0.995	0.86	Probably damaging	0.932
V1817I	0.043	0.558	0.74	Probably damaging	0.096
L1831V	0.029	0.879	0.74	Probably damaging	0.263
R1833C	0.006	1	0.57	Probably damaging	0.478
G1848R	0	1	0.86	Probably damaging	0.766
P1873L	0.013	0.118	0.5	Possibly damaging	0.097
R1879W	0.011	0.978	0.5	Possibly damaging	0.088
I1907F	0.021	0.952	0.74	Probably damaging	0.295
T1917I	0.017	0.037	0.27	Probably benign	0.13
R1918Q	0.03	0.839	0.74	Probably damaging	0.124
P1966S	0	0.999	0.74	Probably damaging	0.569
P1583S	0	0.996	0.57	Probably damaging	0.316
N1986S	0	1	0.86	Probably damaging	0.515
H2004Q	0	0	0.74	Probably damaging	0.738
A2027P	0.014	0.999	0.57	Probably damaging	0.526
R2158Q	0.006	0.999	0.85	Probably damaging	0.372
V2244L	0	0.053	0.74	Probably damaging	0.445
S2262L	0	1	0.57	Probably damaging	0.73

A SIFT score predicts whether an amino acid substitution affects protein function. The SIFT score ranges from 0.0 (deleterious) to 1.0 (tolerated). The score can be interpreted as follows: 0.0 to 0.05 -- Variants with scores in this range are considered deleterious. The PolyPhen-2 score ranges from 0.0 (tolerated) to 1.0 (deleterious). Variants with scores of 0.0 are predicted to be benign. Values closer to 1.0 are more confidently predicted to be deleterious. Panther tool is based on sequence homology and Panther-PSEP method. Scores are based on hidden Markov model families. The effect of the nsSNP is given in terms of probably damaging, possibly damaging and probably benign.

5.2 EFFECTS OF nsSNPs ON STABILITY OF PROTEIN:

The identified deleterious nsSNPs were subjected to stability analysis using I-mutant 2.0. I-mutant determines the effect of given nsSNPs in terms of increasing or decreasing the protein stability. based on the unfolding free energy change (DDG) and the reliability index which can be computed only when the sign of the stability change is predicted and is evaluated from the output of the SVM O as $RI = 20 * \text{abs}(O - 0.5)$. Based on the output, 8 nsSNPs of ARID1A were found to increase the protein stability while 41 nsSNPs of ARID1A were found to decrease the stability. The effects of the selected mutations on protein stability are listed in table 5.2.

Table 5.2: Effects of nsSNPs on protein stability using SNP&GO and I-MUTANT 2.0

POSITION	WILD TYPE	NEW AMINO ACID	STABILITY	RELIABILITY INDEX	SNP&GO
14	G	R	DECREASE	2	NEUTRAL
121	P	L	DECREASE	6	NEUTRAL
392	P	H	DECREASE	1	NEUTRAL
397	G	R	DECREASE	1	NEUTRAL
403	Q	H	DECREASE	4	NEUTRAL
504	P	L	INCREASE	4	NEUTRAL
516	P	L	INCREASE	3	NEUTRAL
578	A	V	DECREASE	3	NEUTRAL
697	P	L	DECREASE	6	DISEASE
740	S	C	DECREASE	0	NEUTRAL
773	P	L	DECREASE	1	NEUTRAL
798	P	L	INCREASE	1	NEUTRAL
839	G	V	INCREASE	6	NEUTRAL
900	A	T	DECREASE	9	DISEASE
906	N	K	DECREASE	3	DISEASE
1026	R	C	DECREASE	3	NEUTRAL
1037	G	R	DECREASE	5	NEUTRAL
1052	Y	C	DECREASE	5	NEUTRAL
1123	S	C	INCREASE	1	NEUTRAL
1136	A	V	DECREASE	1	NEUTRAL
1175	P	L	DECREASE	2	NEUTRAL
1190	A	G	DECREASE	7	NEUTRAL
1226	Y	C	DECREASE	1	NEUTRAL

1467	P	R	DECREASE	6	NEUTRAL
1551	R	C	DECREASE	1	NEUTRAL
1583	P	S	DECREASE	6	NEUTRAL
1593	R	W	DECREASE	5	NEUTRAL
1697	D	V	DECREASE	6	NEUTRAL
1817	V	I	DECREASE	7	NEUTRAL
1831	L	V	DECREASE	8	NEUTRAL
1833	R	C	DECREASE	8	NEUTRAL
1848	G	R	DECREASE	6	NEUTRAL
1873	P	L	DECREASE	4	NEUTRAL
1879	R	W	DECREASE	3	NEUTRAL
1907	I	F	DECREASE	8	NEUTRAL
1917	T	I	INCREASE	4	NEUTRAL
1918	R	Q	DECREASE	3	NEUTRAL
1966	P	S	DECREASE	6	NEUTRAL
1986	N	S	DECREASE	7	NEUTRAL
2004	H	Q	DECREASE	3	NEUTRAL
2027	A	P	INCREASE	3	NEUTRAL
2158	R	Q	DECREASE	9	NEUTRAL
2244	V	L	DECREASE	4	NEUTRAL
2262	S	L	DECREASE	4	NEUTRAL
1231	D	V	DECREASE	0	NEUTRAL
1262	R	H	DECREASE	5	NEUTRAL
1276	R	Q	DECREASE	5	NEUTRAL
1370	K	R	INCREASE	3	NEUTRAL
1451	P	H	DECREASE	6	NEUTRAL

I-MUTANT 2.0 has retrieved results as decrease and increase. Out of 49 mutations, 8 mutations showed increase and remaining 41 mutations showed decrease. SNP&GO retrieved results has disease and neutral. Out of 49 mutations, 3 mutations showed disease and remaining 46 mutations showed neutral. The mutations that showed disease are P697L, A900T, N906K.

5.3 EFFECTS OF SELECTED nsSNPs ON PROTEIN STABILITY:

After using I-MUTANT 2.0, the results that contain 41 decrease mutations has been taken for further analysis. Those 41 mutations has been analyzed. The results are shown in table 5.3.

Table 5.3: Effects of selected nsSNPs on protein stability using Dynamut.

WILD TYPE	POSITION	MUTANT	CHAIN	$\Delta\Delta G$	$\Delta\Delta G$ ENCoM	$\Delta\Delta G$ mCSM	$\Delta\Delta G$ SDM	$\Delta\Delta G$ DUET
SER	2262	LEU	L	1.041 kcal/mol (Stabilizing)	0.338 kcal/mol (Destabilizing)	-0.149 kcal/mol (Destabilizing)	1.450 kcal/mol (Stabilizing)	0.540 kcal/mol (Stabilizing)
HIS	2004	GLN	L	0.264 kcal/mol (Stabilizing)	-0.216 kcal/mol (Destabilizing)	-0.761 kcal/mol (Destabilizing)	-1.450 kcal/mol (Destabilizing)	-0.913 kcal/mol (Destabilizing)
VAL	1817	ILE	L	-0.444 kcal/mol (Destabilizing)	-0.056 kcal/mol (Destabilizing)	-0.323 kcal/mol (Destabilizing)	0.030 kcal/mol (Stabilizing)	0.006 kcal/mol (Stabilizing)
ASP	1697	VAL	L	0.246 kcal/mol (Stabilizing)	0.066 kcal/mol (Destabilizing)	0.132 kcal/mol (Stabilizing)	0.240 kcal/mol (Stabilizing)	0.350 kcal/mol (Stabilizing)
LEU	1831	VAL	L	0.011 kcal/mol (Stabilizing)	-0.042 kcal/mol (Destabilizing)	-0.237 kcal/mol (Destabilizing)	0.040 kcal/mol (Stabilizing)	0.113 kcal/mol (Stabilizing)
PRO	1966	SER	L	-0.143 kcal/mol (Destabilizing)	0.044 kcal/mol (Destabilizing)	-0.622 kcal/mol (Destabilizing)	-0.260 kcal/mol (Destabilizing)	-0.315 kcal/mol (Destabilizing)
ARG	2158	GLN	L	-1.140 kcal/mol (Destabilizing)	-0.656 kcal/mol (Destabilizing)	-0.009 kcal/mol (Destabilizing)	-0.850 kcal/mol (Destabilizing)	-0.118 kcal/mol (Destabilizing)
ASN	1986	SER	L	0.066 kcal/mol (Stabilizing)	0.018 kcal/mol (Destabilizing)	-0.317 kcal/mol (Destabilizing)	-0.440 kcal/mol (Destabilizing)	-0.010 kcal/mol (Destabilizing)

Dynamut tool has retrieved the results by filtering those 41 mutations to 8 mutations. These 8 mutations has retrieved the results that contain mutations's $\Delta\Delta G$, $\Delta\Delta G$ ENCoM, $\Delta\Delta G$ SDM, $\Delta\Delta G$ DUET.

5.4 ENERGY PARAMETERS OF THE NATIVE AND MUTANT MODELS:

Table 5.4: Energy parameters of the native and mutant models

MODEL	TOTAL ENERGY(INITIAL) KCAL/MOL	TOTAL ENERGY(MINIMIZATION) KCAL/MOL
Native model	1151042.01	-105987.04
D1697V	-17391.52	-97948.17
V1817I	-17476.32	-98286.72
L1831V	-17477.74	-98462.97
P1966S	-17444.98	-98099.44
N1986S	-17389.38	-98280.63
H2004Q	-17480.79	-98394.15
R2158Q	-17428.92	-98493.57
S2262L	-17529.01	-98348.40

The total energy of native structure was found to be 1151042. KJ/mol while the energy after minimization was -105987.04KJ/mol. After energy minimization it was found that the total energy of all eight mutant models were lower than the native model. This difference in the total energy after minimization demonstrated that the mutant models were less stable than the native model.

5.5 HOMOLOGY MODELING AND STRUCTURAL ANALYSIS OF NATIVE AND MUTANT MODELS:

Homology modeling approach was used for the prediction of the possible 3D structure for the mutated models of ARID1A, by using SWISS – MODEL individually (Figure 5.5). By using the PDB viewer, RMSD values, TM scores were calculated. The mutant model having high RMSD value shows high variation in structure RMSD values of native and mutant models were provided in the following table 5.

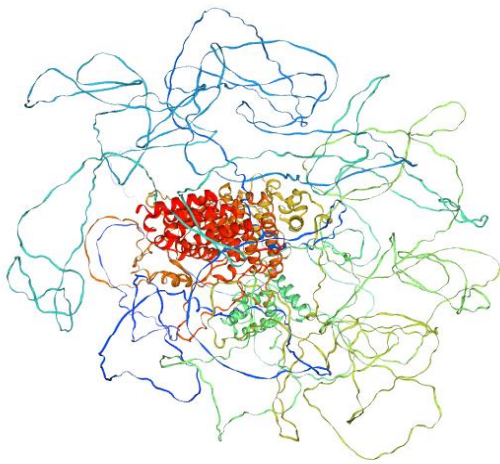
Table 5.5 : Pairwise structure alignment results

MUTATION	Chain	RMSD	TM-score	Identity	Equivalent Residues	Sequence Length	Modelled Residues
D1697V	A	3.9	0.01	4%	88	2285	2285
V1817I	A	3.9	0.01	4%	88	2285	2285
L1831V	A	3.9	0.01	4%	88	2285	2285
P1966S	A	3.9	0.01	4%	88	2285	2285
N1986S	A	3.9	0.01	4%	88	2285	2285
H2004Q	A	3.9	0.01	4%	88	2285	2285
R2158Q	A	3.9	0.01	4%	88	2285	2285
S2262L	A	3.9	0.01	4%	88	2285	2285

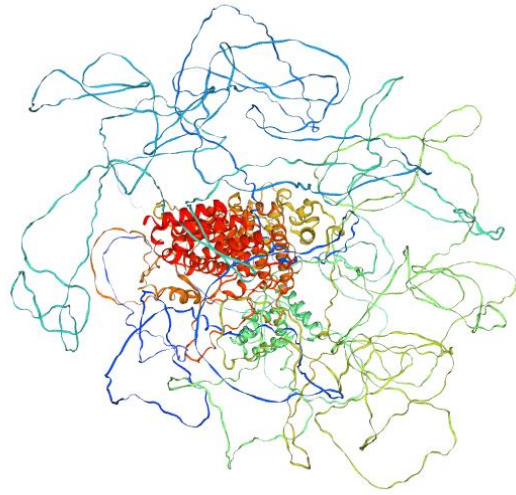
The lower the RMSD, the better the structure alignment between the pair of structures. Residues in loops that are not well aligned are excluded from the RMSD is calculation (i.e., RMSD is calculated only using the residues which can be aligned). The TM-score ranges between 0 and 1, where 1 indicates a perfect match and 0 is no match between the two structures. Scores < 0.2 usually indicate that the proteins are unrelated while those >0.5 generally have the same protein fold.

HOMOLOGY MODELED 3D STRUCTURE OF ARID1A GENE

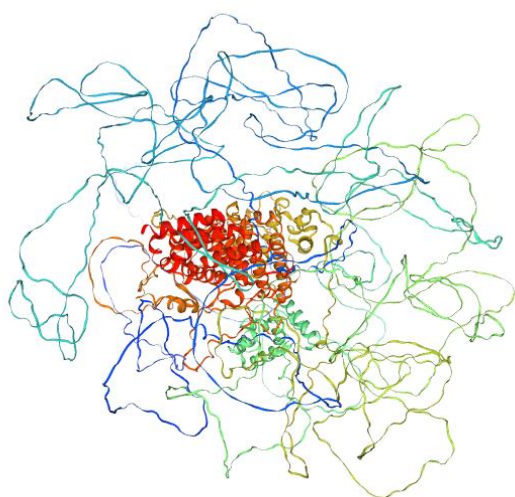
Homology modeling approach was used for the prediction of the possible 3D structure for the mutated ARID1A (D1697V, V1817I, L1831V, P1966S, N1986S, H2004Q, R2158Q, S2262L) by using SWISS – MODEL individually.



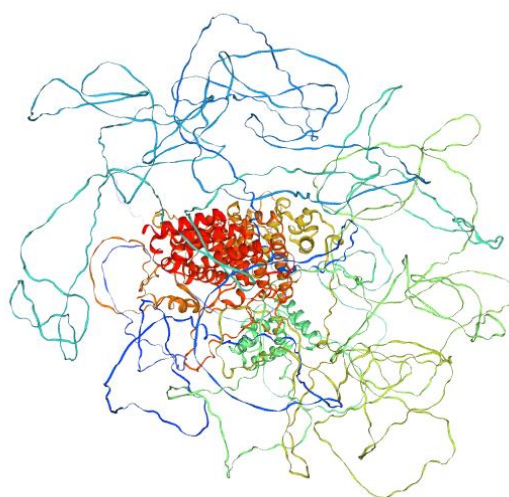
A) D1697V



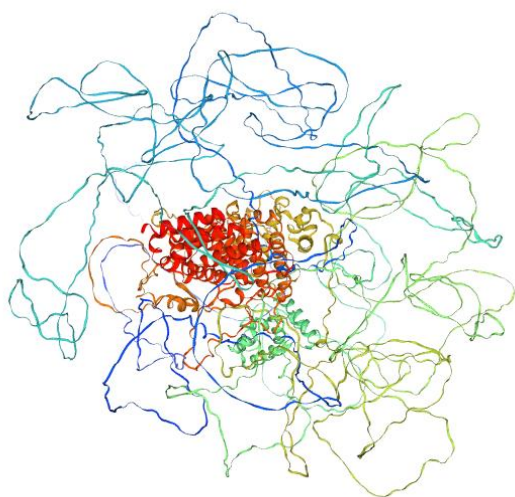
B) V1817I



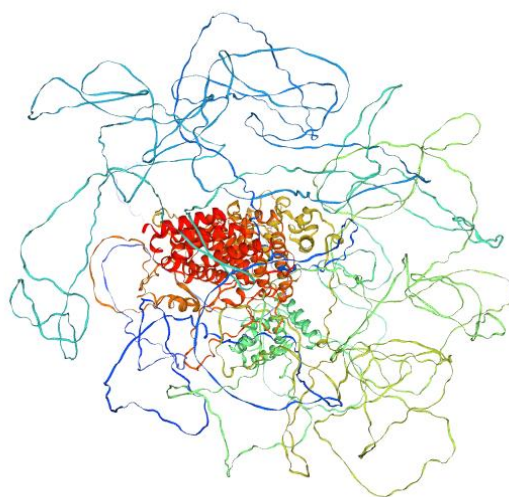
C) L1831V



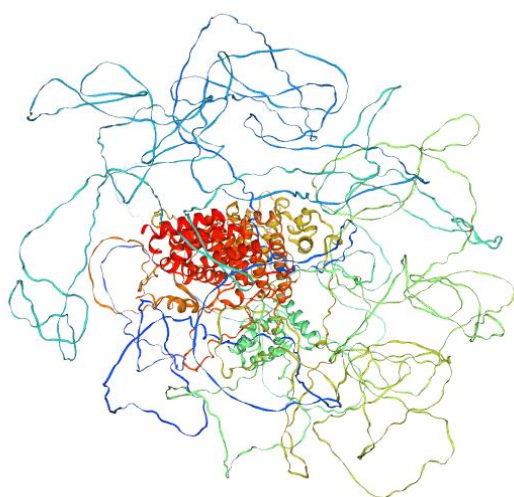
D) P1966S



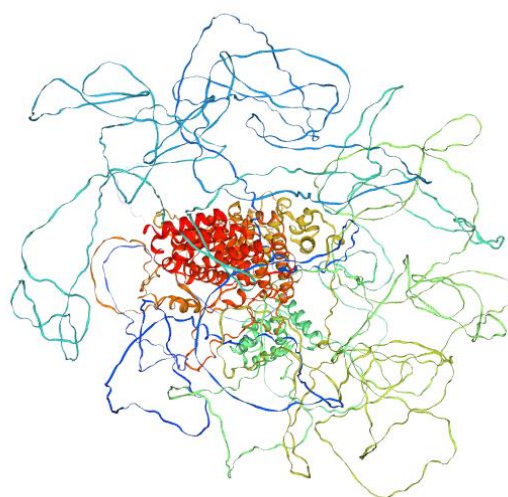
E) N1986S



F) H2004Q



G) R2158Q



H) S2262L

Figure 4.1: 3D STRUCTURE OF ARID1A GENE (A) Mutant protein model(D1697V) (B) Mutant protein model (V1817I) (C) Mutant protein model (L1831V) (D) Mutant protein model (P1966S) (E) Mutant protein model (N1986S) (F) Mutant protein model (H2004Q) (G) Mutant protein model (R2158Q) (H) Mutant protein model (S2262L)

5.6 CONSERVATION ANALYSIS BY CONSURF:

The extent of evolutionary rate of conservation of amino acids in ARID1A was analyzed using the ConSurf web server. ConSurf calculates conservation scores of the amino acid position of the protein based on the phylogenetic relation between homologous sequences. ConSurf predicts conservation scores ranging from 1 to 9 with grade 9 being the most conserved position.

Table 5.6: Conservation analysis by consurf

RS ID	MUTATION	CONSERVATION SCORE
rs375761808	D1697V	7
rs375250508	V1817I	5
rs150534917	L1831V	4
rs370333544	P1966S	8
rs201064943	N1986S	9
rs375639155	H2004Q	9
rs142069738	R2158Q	9
rs375849292	S2262L	9

Specific positions of amino acids that are essential for the correct function of a protein change more slowly than other residues (i.e., are evolutionarily conserved), and deleterious mutations are more frequent in conserved positions. Selected nsSNPs analyzed with ConSurf revealed that they were positioned in the conserved regions, and the tool predicted that N1986, H2004Q, R2158Q, S2262L were highly conserved.

CHAPTER 6

SUMMARY & CONCLUSION

In this study, several web-based algorithmic tools based on sequence and structural conservation were used to identify deleterious nsSNPs of the ARID1A gene. Using dbSNP database, SNPs of ARID1A has been retrieved. Among them, 6601 nsSNPs of ARID1A has been screened. Using some computational tools, 49 has been found to be potentially deleterious. After further analyzes, these 8 mutations (D1697V, V1817I, L1831V, P1966S, N1986S, H2004Q, R2158Q, S2262L) decrease the stability of protein. Homology modelling and structural analyses has been performed for these 8 mutations using swiss model and validated. Consurf server has been performed for those 8 mutations for further analysis. Selected nsSNPs analyzed with ConSurf revealed that they were positioned in the conserved regions, and the tool predicted that N1986, H2004Q, R2158Q, S2262L were highly conserved. These mutations have been involving in PI3K-AKT pathway and thus indirectly leads to the progression of cancer.

CHAPTER 7

REFERENCE

- 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. <https://doi.org/10.1371/journal.pone.0188143>
- Dabhi, B., & Mistry, K. N. (2014). In silico analysis of single nucleotide polymorphism (SNP) in human TNF- α gene. Meta gene, 2, 586-595. <https://doi.org/10.1016/j.mgene.2014.07.005>
- Dagogo -Jack, I., Shaw, A (2018). Tumor heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol 15, 81–94. <https://doi.org/10.1038/nrclinonc.2017.166>
- Dristy, T. T., Noor, A. R., Dey, P., & Saha, A. (2023). Structural analysis and conformational dynamics of SOCS1 gene mutations involved in diffuse large B-cell lymphoma. Gene, 864, 147293. <https://doi.org/10.1016/j.gene.2023.147293>
- Hossain, M. S., Roy, A. S., & Islam, M. S. (2020). In silico analysis predicting effects of deleterious SNPs of human RASSF5 gene on its structure and functions. Scientific reports, 10(1), 14542. <https://doi.org/10.1038/s41598-020-71457-1>.
- Kalia, N., Sharma, A., Kaur, M., Kamboj, S. S., & Singh, J. (2016). A comprehensive in silico analysis of non-synonymous and regulatory SNPs of human MBL2 gene. SpringerPlus, 5(1), 811. <https://doi.org/10.1186/s40064-016-2543-4>
- Kaur, R., Singh, J., & Kaur, M. (2017). Structural and functional impact of SNPs in P-selectin gene: A comprehensive in silico analysis. Open life sciences, 12(1), 19-33. <https://doi.org/10.1515/biol-2017-0003>
- Marín-Martín, F. R., Soler-Rivas, C., Martín-Hernández, R., & Rodríguez- Casado, A. (2014). A comprehensive in silico analysis of the functional and structural impact of

nonsynonymous SNPs in the ABCA1 transporter gene. Cholesterol.639751.
<http://dx.doi.org/10.1155/2014/639751>

- Pavlidou, E. N., & Balis, V. (2020). Diagnostic significance and prognostic role of the ARID1A gene in cancer outcomes. World Academy of Sciences Journal, 2(2), 49-64.
<https://doi.org/10.3892/wasj.2020.37>
- Rajasekaran, R., Doss, C. G. P., Sudandiradoss, C., Ramanathan, K., & Sethumadhavan, R. (2008). In silico analysis of structural and functional consequences in p16INK4A by deleterious nsSNPs associated CDKN2A gene in malignant melanoma. Biochimie, 90(10), 1523-1529.
<https://doi.org/10.1016/j.biochi.2008.05.017>
- Rajasekaran, R., Sudandiradoss, C., Doss, C. G. P., & Sethumadhavan, R. (2007). Identification and in silico analysis of functional SNPs of the BRCA1 gene. Genomics, 90(4), 447-452. <https://doi.org/10.1016/j.ygeno.2007.07.004>
- Ray, M., Mishra, J., Priyadarshini, A., & Sahoo, S. (2019). In silico identification of potential drug target and analysis of effective single nucleotide polymorphisms for autism spectrum disorder. Gene Reports, 16, 100420.
<https://doi.org/10.1016/j.genrep.2019.100420>
- Rotimi, S. O., Peter, O., Oguntade, O., & Rotimi, O. A. (2018). In silico analysis of the functional non-synonymous single nucleotide polymorphisms in the human CYP27B1 gene. Egyptian Journal of Medical Human Genetics, 19(4), 367-378.
- Samartzis, E. P., Noske, A., Dedes, K. J., Fink, D., & Imesch, P. (2013). ARID1A mutations and PI3K/AKT pathway alterations in endometriosis and endometriosis-associated ovarian carcinomas. International journal of molecular sciences, 14(9), 18824-18849. <https://doi.org/10.3390/ijms140918824>

- Takeda, T., Banno, K., Okawa, R., Yanokura, M., Iijima, M., Irie-Kunitomi, H., ... & Aoki, D. (2016). ARID1A gene mutation in ovarian and endometrial cancers. *Oncology reports*, 35(2), 607-613. <https://doi.org/10.3892/or.2015.4421>
- Venkata Subbiah, H., Ramesh Babu, P., & Subbiah, U. (2020). In silico analysis of non-synonymous single nucleotide polymorphisms of human DEFB1 gene. *Egyptian Journal of Medical Human Genetics*, 21, 1-9. <https://doi.org/10.1186/s43042-020-00110-3>
- Wu, R. C., Wang, T. L., & Shih, I. M. (2014). The emerging roles of ARID1A in tumor suppression. *Cancer biology & therapy*, 15(6), 655-664. <https://doi.org/10.4161/cbt.28411>.
- Xu, S., & Tang, C. (2021). The role of ARID1A in tumors: tumor initiation or tumor suppression? *Frontiers in Oncology*, 11, 745187. <https://doi.org/10.3389/fonc.2021.745187>

