A comprehensive *in silico* investigation into the pathogenic SNPs in RTEL1 gene and their biological consequences

By

Rifah Rownak Tanshee

ID- 18336028

A thesis submitted to the Department of Mathematics and Natural Sciences in partial fulfillment of the requirements for the degree of
B.Sc in Biotechnology

Department of Mathematics and Natural Sciences
Brac University
March 2023

© 2023. Brac University
All rights reserved.

Declaration

It is hereby declared that

1. The thesis submitted is my own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, exceptwhere this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any otherdegree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

Student's Full Name & Signature:

Rifah Rownak Tanshee
ID: 18336028

Approval

The thesis/project titled "[Thesis/Project Title]" submitted by

Rifah Rownak Tanshee (18336028) of Spring, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B. Sc in Biotechnology on .

Examining Committee:

Supervisor:
(Member)

_____

Mohammed Sayem
Lecturer, Department of Mathematics and Natural Science, Brac University

Program Director:
(Member)

_____

Dr. Munima Haque,
Associate Professor, Department of Mathematics and Natural Sciences, Brac University

Departmental Head:
(Chair)

_____

A F M Yusuf Haider
Professor and Chairperson, Department of Mathematics and Natural Sciences
Brac University

Abstract

Regulator of Telomere Helicase 1 (RTEL1) is a protein-coding gene that encodes an essential DNA helicase which is thought to be involved in preserving the telomere and genetic stability. Germline mutations in the RTEL1 gene have been clinically associated with Hoyeraal-Hreidarsson syndrome (HH), a more severe version of Dyskeratosis Congenita (DC). Missense mutations are also reported in several other non-communicable diseases, namely high-grade glioma, astrocytomas, glioblastomas, myeloid neoplasms, breast and lung cancers. Despite the fact that various research has sought to link RTEL1 mutations to specific disorders, no thorough investigation on germline missense mutations has been performed yet. In this study, we attempted to investigate functionally and structurally deleterious nonsynonymous or missense SNPs of the RTEL1 gene using an in-silico approach. Initially, out of 1392 missense SNPs reported in the dbSNP database, 43 SNPs were filtered out through 10 bioinformatics-based web servers. With subsequent analysis using 9 in-silico tools, these 43 nsSNPs were further shortened to 13 most deleterious nsSNPs. Following analysis of mutated protein structures, secondary structure, evolutionary conservancy, conservation profile, surface accessibility, domain & cluster, PTM site, and interatomic interaction also revealed the detrimental effect of these 13nsSNPs on RTEL1 protein. In-depth investigation of these mutations through molecular docking demonstrated a striking change in the interaction pattern of DNA with F15L, M25V, Y228C, G706R, and R729C mutant proteins suggesting the more severe consequences of these mutations on protein structure and functionality. Thus, these insights will pave the way for extensive analysis of RTEL1 gene variants in the future along with the advancement of precision medicine and other treatment modalities.

Keywords: RTEL1, telomere, genetic stability, germline mutations, Hoyeraal-Hreidarsson syndrome, Dyskeratosis Congenita, missense, SNPs, bioinformatics tools, molecular docking.

Dedication

My mother, who has always been my source of inspiration, strength, and unconditional love.

Thank you for instilling in me the value of education, hard work, and perseverance. Your selflessness, resilience, and unshakeable faith have taught me to never give up, no matter how challenging the path may seem. I am forever grateful for the sacrifices you have made to provide me with the best possible opportunities in life.

I am honored to dedicate this work to my dear mother. I hope that I have made you proud and that my accomplishments bring a smile to your face.

Acknowledgement

I would like to express my sincere gratitude to everyone who has been the part of my undergraduate journey.

First and foremost, I am grateful to Allah for blessing me with the knowledge, strength, and patience to undertake this journey. I would not have been able to do this endeavor without Allah's guidance and blessings.

I am deeply grateful to my supervisor Mohammad Sayem, for his guidance, encouragement, and unwavering support throughout this thesis project. His expertise, insights, and constructive feedback have been instrumental in shaping my research project and refining my analytical skills.

I would also like to extend my thanks to the chairperson of MNS department for giving me scope of my research and for all the facilities that helped me to complete this overall curriculum.

My sincere thanks also go to all my respected faculty members and my friends who have provided me with invaluable support, both personally and academically. I am grateful for their encouragement, constructive feedback, and willingness to help whenever I needed it.

Lastly, I am indebted to my family for their unwavering love, support, and encouragement throughout my academic journey. Their sacrifices and patience have been a source of strength and inspiration for me, and I am forever grateful for their unconditional love and support.

Thank you for being a part of my journey and for making it all worthwhile. May Allah bless you all abundantly.

Table of Contents

Declaration ii

Approval iii

Abstract iv

Dedication v

Acknowledgement vi

List of Table ix

List of Figures x

List of Acronyms xi

Chapter 1 Introduction 1

1.1 Background 2

Chapter 2 Materials & Method 11

2.1 Data Retrieval 13

2.2 Retrieval of 3D Structure & Quality Checking 13

2.3 Functional Impact Prediction 13

2.4 Structural Impact Prediction 15

2.5 Comparative Modeling & Evaluation of mutated 3D structures 16

2.6 Secondary Structure Analysis 16

2.7 Identification of Domain and Cluster 17

2.8 Prediction of Post Translational Modification Sites 17

2.9 Conservation, Surface Accessibility & Evolutionary Relationship 17

2.10 Interatomic Interaction Prediction 18

2.11 Molecular Docking Analysis 18

Chapter 3 Result 20

3.1 SNP Annotation 21

3.2 Assessment of RTEL1 Protein Structure 21

3.3 Determination of functional consequences of RTEL1 nsSNPs 22

3.4 Determination of Structural Impact of RTEL1 nsSNPs 26

3.4.1 Analysis of protein stability 26

3.4.2 Prediction of Phenotypic Effects 27

3.5 Three Dimensional Structure Prediction for Mutant Proteins 29

3.6 Investigation of the Impact of nsSNPs on Secondary Structure 30

3.7 Identification of Domains and Clusters 32

3.8 Effect of nsSNPs on Post Translational Modification Sites 33

3.9 Analyzation of Evolutionary Relationship of RTEL1 Protein and Conservation Profile & Surface Accessibility of nsSNPs 34

3.9.1 Evolutionary Relationship 34

3.9.2 Analysis of Evolutionary conservation 36

3.9.3 Evaluation of Surface accessibility 39

3.10 Prediction of Interatomic Interaction 40

3.11 Molecular Docking 43

Chapter 4 Discussion 49

Conclusion 55

References 56

List of Table

Table 1: Ramachandran plot parameters of RTEL1 wild-type protein's AlphaFold structure.... 21

Table 2: High-risk nsSNPs anticipated by five web tools. 23

Table 3: High-risk nsSNPs predicted by five web tools. 24

Table 4: Destabilizing nsSNPs identified by seven in silico tools..... 27

Table 5: Damaging nsSNPs predicted by MutPred2 and Project HOPE. 28

Table 6: RMSD value of 13 mutant models generated through PyMOL 29

Table 7: ProMotif information of native and mutant protein..... 30

Table 8: Surface accessibility result of 10nsSNPs from NetSurf2.0 39

Table 9: Analysis of binding affinity and interaction of wild-type and mutant protein with DNA 44

List of Figures

Figure 1: The structure of Human Telomere. 3

Figure 2: Schematic diagram of 18 mutations found in individuals having dyskeratosis congenita (DC) and Hoyeraal–Hreidarsson syndrome (HH) patients. 7

Figure 3: Illustration of RTEL1-deficient cells undergoing malignant transformation..... 8

Figure 4: Illustration showing the workflow of this study..... 12

Figure 5: Ramachandran plot of AlphaFold structure 22

Figure 6: Secondary structures of wild-type and mutated RTEL1 proteins. 32

Figure 7: Domain & cluster information of nsSNPs represented in linear protein model..... 33

Figure 8: Evolutionary relationship analysis 36

Figure 9*:* Evolutionary conservation profile prediction of RTEL1 protein 38

Figure 10: Interatomic interaction of 13nsSNPs..... 43

Figure 11: Graphical representation of molecular docking 48

List of Acronyms

RTEL1 Regulator of Telomere Helicase 1

SNP Single Nucleotide Polymorphism

nsSNPs Non-Synonymous Single Nucleotide Polymorphisms

HR Homologous Recombination

DSB Double Stranded Break Repair

SDSA Synthesis Dependent Strand Annealing

DC Dyskeratosis Congenita

HHS Hoyeraal–Hreidarsson Syndrome

LCLs Lymphoblastoid Cell Lines

CFSs Common Fragile Sites

MiDAS Mitotic DNA Synthesis

pCNA Proliferating Cell Nuclear Antigen

PIP pCNA Interacting Protein

RMSD Root Mean Square Deviation

HHD2 Harmonin Homology Domain 2

UniprotKB Universal Protein Knowledgebase

PDB Protein Data Bank

SuSPect Disease-*Sus*ceptibility-based *S*AV Phenotype Prediction

SIFT Sorting Intolerant from Tolerant

PROVEAN Protein Variation Effect Analyzer

SNP & GO SNP & Gene Ontology

Polyphen2 Polymorphism phenotype v2

PANTHER Protein Analysis Through Evolutionary Relationship

INPS-MD Impact of Non-synonymous mutations on Protein Stability – Multi Dimension

DAS Distributed Annotation System

PTM Post Translational Modification

XPD Xeroderma Pigmentosum Group D

FANCJ Fanconi anemia complementation group J

HB Hydrogen Bond

VdW Van der Waals

Chapter 1

Introduction

1.1 Background

The human genome is not identical among individuals, there are inherent distinctions among racial and ethnic groups, as well as between healthy people and people who are prone to illness. Around 99.9% of the DNA sequences in the human genome are the same around the world, and the remaining 0.1% of the genome contains unique variations for each individual. This variance in each person's DNA is the product of random mutations (Forsberg et al., 2000). Single nucleotide polymorphism (SNP), which is a single base substitution in alleles, is the most prevalent type of mutation among them. SNPs occur in every 1,000 base pairs in the genome (Collins et al., 1998) and can be found in both coding and non-coding regions. It is estimated that SNPs account for 90% of the sequence variation in the human genome, which may operate as a possible genetic marker as well as impose a neutral or detrimental effect on protein function and transcription factor-mediated regulations (Goswami, 2015). Variants in the non-coding region have been demonstrated to have an impact on the function of cis or trans-regulatory elements, UTR, and intron which might disrupt the affinity of transcription factors, various epigenetic factors, alternative splicing, and mRNA stability (Mansur et al., 2018). But the most significant SNPs are thought to be located in the coding region and nearly 50 thousand SNPs have been registered in this region up until now(Collins et al., 1998). The SNPs in the coding region are categorized as Synonymous and Non-synonymous SNPs. Synonymous SNPs are not responsible for causing any alteration in the amino acid sequence of protein while the non-synonymous SNPs, particularly missense SNPs, are to blame for amino acid substitutions in the protein sequence thus altering the activity of the protein. According to earlier research, nsSNPs account for about 50% of the mutations linked to a number of genetic illnesses (Doniger et al., 2008; Radivojac et al., 2010), as well as several autoimmune and inflammatory conditions (Azad et al., 2012; Begovich et al., 2004; Sobieszczyk et al., 2011).

Telomeres, which serve as the termini of linear chromosomes, are made up of certain repeating sequences or tandem (TTAGGG) hexamer DNA repeats along with a crucial single-stranded 3′-overhang at the end (Jain & Cooper, 2010; O'Sullivan & Karlseder, 2010). Shelterin complex, a set of six proteins, uses the repeating telomeric motif as a platform to protect telomeres from fusion

and degradation(de Lange, 2009). In order to protect the chromosome extremities from DNA damage response, telomeres may take on a lariat structure in which the 3′ G-overhang invades the double-stranded telomere to create a T-loop with the help of shelterin complex (Griffith et al., 1999a).

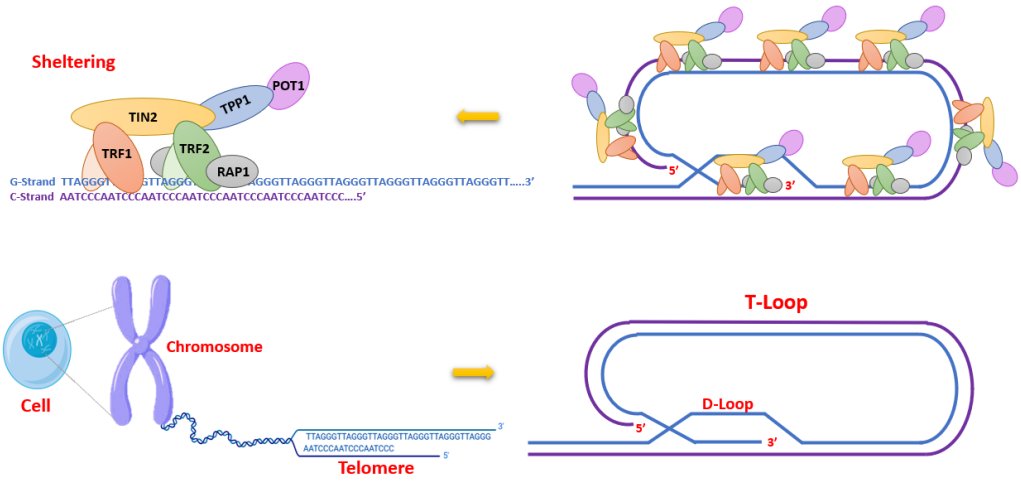


Figure 1: The structure of Human Telomere.

Moreover, the 3'G-rich overhang also promotes the formation of G-quadruplex or G4 DNA that acts as an obstacle for DNA replication machinery(Gilson & Géli, 2007). Additionally, each replicative cell cycle results in telomere shortening due to DNA processing and incomplete DNA replication. This phenomenon is counteracted through telomerase enzyme, which can lengthen this overhang to compensate for losses. As a result of the inability of maintaining the telomere length and protecting it from cellular degrading factors, telomere dysfunctionality, apoptosis, senescence, and genomic instability might occur.

Regulator of telomere elongation helicase 1 (RTEL1) is an essential iron-sulfur (FeS)-containing DNA helicase, which is a member of the DEAH subfamily of the Superfamily 2 (SF2) helicases and also categorized as a RAD3-like helicase with a 5′ to 3′ helicase activity. It is located at chromosome 20q13.33, and contains thirty-five exons. Various isoforms are produced through alternative splicing results in multiple transcript variants and in human the two main isoforms are-

isoform 2 (1219 amino acid) and isoform 6 (1300 amino acid), both differ in C terminal region (LeGuen et al., 2013a). RTEL1 is a multidomain protein, that includes a RAD3-like helicase domain that contains helicase type 2 ATP binding domain and C terminus (DEAD2 and Helicase C2) domains, DEAH box, PCNA interacting motifs or PIP boxes, Harmonin N-like domains and RING-finger domain (Glousker et al., 2015; Vannier et al., 2014). This gene is essential for telomere regulation, DNA repair, and genome stability that interacts with proteins in the shelterin complex to preserve the telomere. In embryonic stem cells, deletion of the mouse Rtel1 gene caused telomere loss and chromosomal abnormalities, indicating the necessity of the gene for both maintenance of telomere and genome integrity (Ding et al., 2004). Furthermore, it has been noted that human RTEL1 interacts with TRF1 and TRF2, pointing to its recruitment to telomeres (Sarek et al., 2015; Sfeir et al., 2009).

DNA secondary structures such as trinucleotide repeats, G-quadruplexes, or the intermediates formed during the 3R process must be processed correctly in order to maintain genome stability and reduce pathological consequences (Vannier et al., 2014). Several studies have suggested the role of RTEL1 as an anti-recombinase that combats harmful recombination and limit the crossover in meiosis. Homologous recombination (HR) is an essential biological activity for DNA double-strand break repair (DSB) that is necessary for the replication of DNA and the development of crossovers during meiosis. Homologous recombination also helps in the formation of T-loop at the end of the telomere (Griffith et al., 1999b; R. C. Wang et al., 2004), an important event that prevents the telomere being recognized as a breakpoint by the DNA repair system and protects it from degradation. During homologous recombination, the displacement loop (D-loop) forms as an intermediary structure due to the invasion process. The T-loop structure of telomere DNA also contains D-loop (de Lange, 2004). The processing mechanism of the secondary structure or the D-loop formed during HR can develop crossover or non-crossover products. The RTEL1 gene maintains the crossover homeostasis through the physical separation of strand invasion events, which encourages non-crossover repair through synthesis-dependent strand annealing (SDSA) and during DNA repair and meiotic recombination procedures, it facilitates the breakdown of D loop recombination intermediates (Barber et al., 2008; Uringa et al., 2010). Additionally, through resolving G-quadruplexes created during telomere replication, mouse Rtel1 has also been linked to disassembling T loops and preventing telomere fragility, which collectively maintains the dynamics and integrity of the telomere (Vannier et al., 2012). Besides, Frizzell et al. demonstrated

the association of RTEL1 in unwinding trinucleotide repeat to prevent triplet repeat-mediated chromosome fragility.

The common expression of the RTEL1 gene is found in the testis, appendix, spleen, endometrium, adrenal, prostate, bone marrow, and 20 other tissues. The mutation in the RTEL1 gene has been linked to a variety of human diseases, including dyskeratosis congenita, Hoyeraal-Hreidarsson syndrome, glioma, glioblastoma, pulmonary fibrosis, bone marrow failure, breast cancer, and other malignancies.

Dyskeratosis congenita (DC) and its phenotypically acute form Hoyeraal–Hreidarsson syndrome (HHS) is a group of inherited diseases characterized by significantly short telomeres and a wide range of clinical manifestations. Due to the defective telomeric biology, the patients with DC are distinguished by the characteristic appearance of nail dystrophy, abnormal skin coloration, and mouth leukoplakia, while some additional symptoms, including intrauterine growth retardation, cerebral hypoplasia, and bone marrow failure, have been seen in patients with HHS. This results in an early manifestation of the diseases as well as poor prognosis in patients with HHS (Lee, 2013). Bone marrow failure is the leading cause of death in DC and HHS, however, mortality from malignancy and pulmonary fibrosis also happens more frequently than usual(Deng et al., 2013). Despite the mutations in the most common genes such as KDC1, TERT, TINF2, TERC, NOP10, NHP2, WRAP53, and CTC1, studies have linked the mutation in the regulation of telomere elongation helicase 1 (RTEL1) gene to the pathogenesis of HHS. In this context, several clinical research has been performed and discovered novel mutations of RTEL1 in patients with HHS. Deng et al. reported a family with compound heterozygous mutations (nonsense: R974X and missense: M492I mutation) in RTEL1, where the patients were observed to have severely shortened telomere and length-independent telomere defects in blood cells and fibroblast cells, respectively. Additionally, they demonstrated the suppression of DNA damage response and the inability of active telomerase for the maintenance of stable telomeres in fibroblast and lymphoblastoid cell lines (LCLs) in patients with RTEL1 mutation. The mutation M492I was identified to reside in an evolutionarily conserved region, and exhibited more severe onsets in patients with this mutation, suggesting an increased risk of cancer predisposition. The nonsense variant R974X was also reported in two sperate studies, denoted as R998X and both studies used

a 1,243-amino acid transcript variant (NM_032957) for their analysis. Ballew et al. documented three mutations in RTEL1 in two independent families, where the Arg1010X mutation was found to have an autosomal dominant inheritance in the first family and Arg998X and Glu615Asp mutations as compound heterozygote autosomal recessive inheritance in the second family. The finding demonstrated the deletion of the pCNA interacting protein (PIP) motif as a result of the mutations Arg1010X and Arg998X, and a high conservation profile of substitution Glu615Asp in the helicase domain. In contrast to the splice variants shown in Deng et al. investigation, the variant Arg998X was discovered to have an alternative 24 aa exon, pointing to the usage of distinct cell lines in these two studies. However, both investigations indicate the negative impact of the mutation on RTEL1 activity and telomere preservation. Furthermore, Walne et al. did a larger investigation on 10 patients from 7 different families with familial HHS and found 11 biallelic autosomal recessive mutations in RTEL1, among which one mutation has been previously documented in the NHLBI Exome variant server. It was shown that people with RTEL1 mutations had telomere lengths that were considerably shorter. Intriguingly, the authors detected increased T-circles in HHS patients with RTEL1 mutation, which is consistent with the results of the study done on mice, whereas Deng et al. observed a decrease in RTEL1deficient cells and an increase in ectopically expressed RTEL1 fibroblast and LCLs cells which had active telomerase level. These findings imply that the manifestation of RTEL1 deficiency depends on the organism and cell type, along with the techniques used in the detection process. Moreover, any discernible variations in T-circle levels were not seen in HHS cases carrying the DKC1 gene, demonstrating that aberrant T-circle formation is only present in HHS patients with RTEL1 mutations (Lee, 2013).

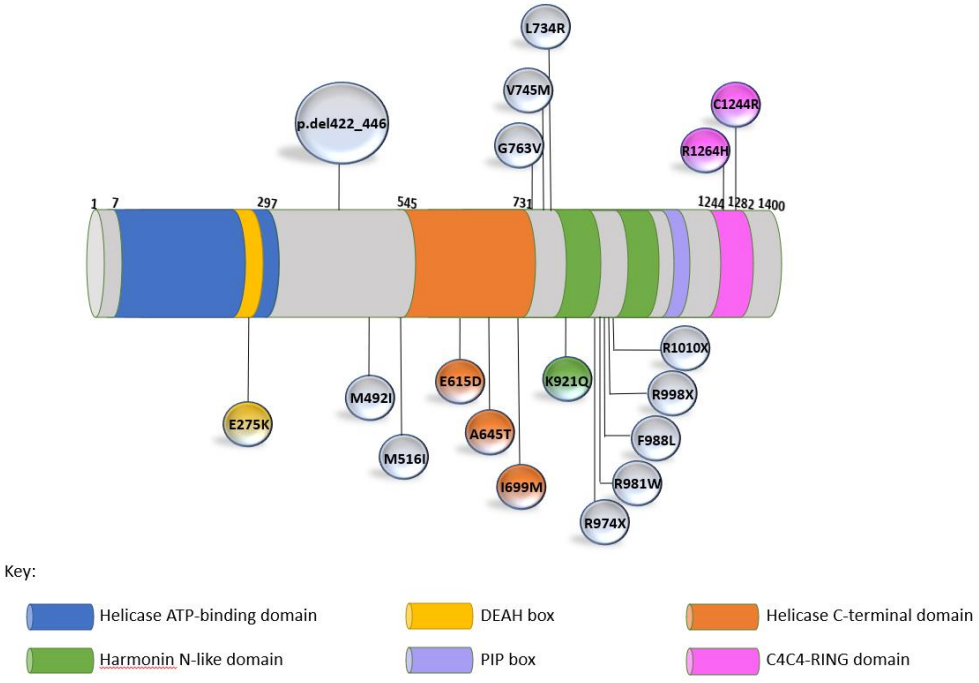


Figure 2: Schematic diagram of 18 mutations found in individuals having dyskeratosis congenita (DC) and Hoyeraal–Hreidarsson syndrome (HH) from different studies. (Adapted from Vannier et al., 2014)

The risk of tumorigenesis or cancer predisposition due to RTEL1 mutations is not only observed in the case of HHS or DC but interestingly it has been also connected to predispositions of brain malignancies like gliomas, astrocytomas, and glioblastomas(Egan et al., 2011; Lin et al., 2021; Liu et al., 2010). The RTEL1 gene has thus been suggested to be a tumor suppressor gene for the emergence of brain malignancies(Wrensch et al., 2009). However, recent studies have also shown that the RTEL1 gene locus is amplified in a number of malignancies, including gastrointestinal and breast tumors (Bai et al., 2000; Muleris et al., 1995). In many cellular circumstances, it is conceivable that either overexpression or downregulation of the RTEL1 gene could lead to the formation of cancer or tumorigenesis in many different ways. But among them, the most prevalent mechanism of cancer formation due to RTEL1 is thought to be defective G4 unwinding.

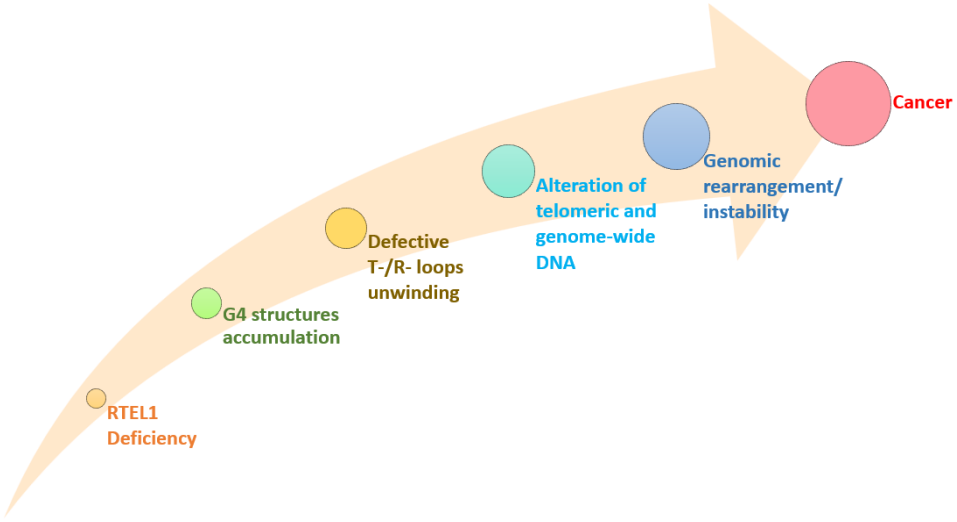


Figure 3: Illustration of RTEL1-deficient cells undergoing malignant transformation. (Adapted from Hassani et al., 2023)

R-loops, a co-occurrence known for its intimate relationship between G4-DNA and RNA structures, happens to increase due to deficient functionality of RTEL1 in cells. A number of separate investigations demonstrated that the regulation of G4-DNA/R-loops is facilitated by RTEL1 and cells with depleted RTEL1, observed to have the inability to unwind G4-DNAs, leading to an increase in R-loops formation which in turns increase the transcription-replication collisions (Hassani et al., 2023). This ultimately leads to genome instability and the emergence of cancer.

DNA replication stress, produced by oncogene activation during tumorogenesis, causes G4/R-loop forming loci for example, common fragile sites (CFSs) and telomeres to remain under-replicated during interphase, which is compensated through mitotic DNA synthesis (MiDAS) (Wu et al., 2020). The mechanism of MiDAS depends on the RTEL1 protein, where the recruitment of RTEL1 to the affected loci is facilitated through SLX4, which in turn assists in attracting RAD52 and POLD3 protein—both essential for MiDAS (Wu et al., 2020). This suggests the necessity of RTEL1 in the maintenance of genomic stability through the resolution of conflicts between the replication and transcription machinery. On the other hand, Takedachi et al. showed SLX4-RTEL1 complex increases the recruitment of proteins to nascent DNA, strongly associated with active RNA pol II, which also facilitates the co-localization of FANCD2/RNA pol II. Therefore, the

interaction of SLX4 and RTEL1 is necessary for replication fork development and abolishment of this interaction has been observed in patients with HHS and cancer.

Functional variations caused by SNPs might have detrimental or beneficial effects on protein structure or function (Capriotti & Altman, 2011). Damage to protein structures and disruption of gene regulation are examples of detrimental impacts (Barroso et al., 1999). Additionally, changes in the protein sequence may ultimately have an impact on changes in the dynamics, translation, hydrophobicity,(Petukh et al., 2015) charge, shape, and inter/intra protein interactions, endangering cells(Bee et al., 2010; Chasman & Adams, 2001; Kucukkal et al., 2015). This information supports the notion that nsSNPs, particularly missense SNPs, are connected to several human disorders. The use of computational methods in recent studies on nsSNPs successfully revealed the possible relevance of mutation in comprehending the molecular pathways of numerous disorders (Adiba et al., 2021; Mondal et al., 2022; Rajendran et al., 2018). Analysis of nsSNPs' structural and functional properties could help with the creation of tailored treatments based on genetic variation. Despite the tools' questionable accuracy, combinatorial usage of a variety of algorithms made it possible to ensure the precise prediction of the effect of a certain mutation. Moreover, the computation analysis is important for primary filtration, as working with a large number of SNP data in laboratory experiments would be expensive and time-consuming.

The mutation in the RTEL1 gene can cause multiple discrepancies not only in telomere biology but also in cellular replication and DNA repair mechanism as well. A broader spectrum of clinical complications in patients with DC and HHS who have inherited RTEL1 mutation has been observed in multiple clinical studies. In addition, the effect of a mutated RTEL1 gene may vary depending on the cell type and the mutation that occurred in the gene. Even though the RTEL1 gene has been the subject of multiple genome-wide association studies, the majority of the RTEL1 SNPs have not yet been thoroughly studied for their potential to cause disease. It is still unclear how nsSNPs affect the RTEL1 protein in terms of disease etiology and to our best knowledge, there hasn't been any comprehensive *in silico* analysis of the RTEL1 gene done till now to detect all the potential nsSNPs associated with the functional and structural change of the protein.

The design and execution of this study are based on the following objectives:

- To identify the most deleterious nsSNPs in the RTEL1 gene.
- To elucidate the impact of genetic variation on the protein's structure and stability.
- To attain molecular-level insights on SNP-mediated protein's functional divergence.

Chapter 2

Materials & Method

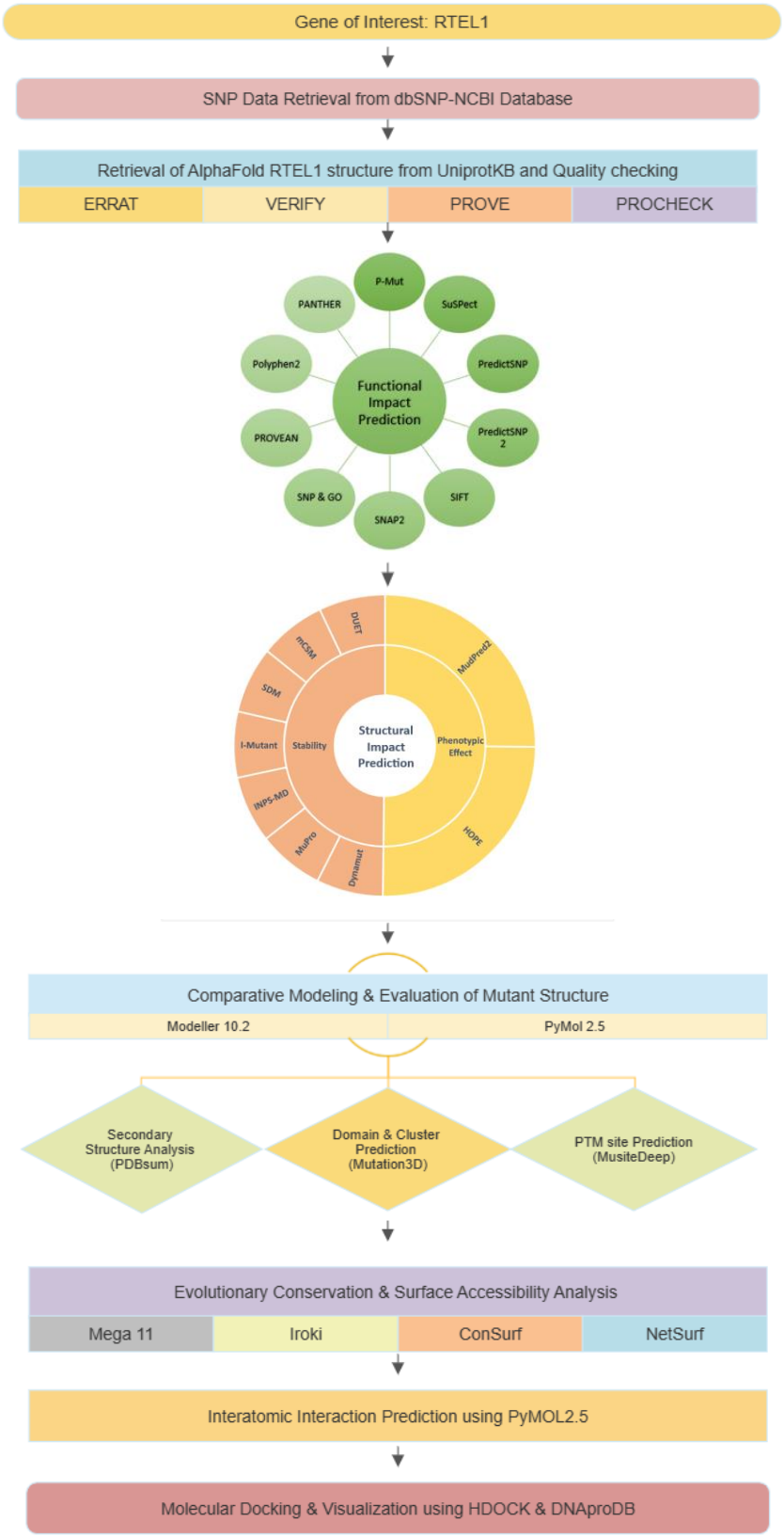


Figure 4: Illustration showing the workflow of this study

2.1 Data Retrieval

The SNP data of the RTEL 1 gene was acquired from the available human GRCh37 genome SNPs in NCBI dbSNP [https://www.ncbi.nlm.nih.gov/snp/?term=] database. Relative data about the RTEL1 gene and the amino acid sequence (FASTA format) of RTEL1 protein were collected from NCBI [https://www.ncbi.nlm.nih.gov/]and UniProtKB (Universal Protein Knowledgebase) [https://www.uniprot.org/] databases, respectively.

2.2Retrieval of 3D Structure & Quality Checking

The AlphaFold structure of the Human RTEL1 protein was retrieved from the UniprotKB (Universal Protein Knowledgebase) [https://www.uniprot.org/] database. The validation of the retrieved structure was checked using SAVES [https://saves.mbi.ucla.edu/] server. Taking the PDB file as input, SAVES returns the output of five programs. The results of ERRAT, VERIFY, PROVE, and PROCHECK Ramachandran plot were analyzed to estimate the validation of the AlphaFold structure of the native protein.

2.3 Functional Impact Prediction

To determine the functional consequences of nsSNPs that were retrieved from the dbSNP database, ten bioinformatics-based web tools i.e., PMut, SuSPect, PredictSNP, PredictSNP2, SIFT, SNAP2, SNP & GO, PROVEAN, Polyphen2, PANTHER were used to ensure the veracity and stringency of the results. SNPs that were commonly identified as deleterious by all of these ten algorithms, were considered as high-risk nsSNPs.

PMut [http://mmb.irbbarcelona.org/PMut/] anticipate the pathological mutations on protein FASTA sequences, where a score >0.5 indicates disease effects of nsSNPs and <0.5 indicates the neutral effects of nsSNPs on the given protein's functionality (López-Ferrando et al., 2017). SuSPect [http://www.sbg.bio.ic.ac.uk/~suspect/] (Disease-Susceptibility-based SAV Phenotype Prediction) webserver predicts single amino acid variants associated with the disease with 82%

accuracy (Yates et al., 2014). PredictSNP [https://loschmidt.chemi.muni.cz/predictsnp/] is a consensus classifier with eight integrated established prediction tools to predict the mutations related to the disease.(Bendl et al., 2014).

PredictSNP2 [https://loschmidt.chemi.muni.cz/predictsnp2/] is a unified web platform with six integrated prediction tools that predicts the pathogenic effect of SNPs in distinct genomic regions (Bendl et al., 2016). SIFT (Sorting Intolerant from Tolerant) [https://sift.bii.a-star.edu.sg/] predicts the impact of an amino acid alteration on protein depending on the sequence homology and physical property of amino acids, where score <=0.05 indicates damaging and >0.05 is tolerant (Sim et al., 2012).

Next, PROVEAN (Protein Variation Effect Analyzer) [https://www.jcvi.org/research/provean] was used for the prediction of the damaging impact of nsSNPs on protein sequence (Choi & Chan, 2015). The PROVEAN score which is generated by averaging the delta alignment scores of variants and reference protein query sequence concerning homology sequence, helps to separate the nsSNPs as deleterious (score <= -2.5) and neutral (score >-2.5) variants. SNAP2 [https://rostlab.org/services/snap/] is another neural network-based web tool that gives prediction scores between -100 to +100 which indicates strong neutral to strong impactful variants (Hecht et al., 2015). SNP & GO (SNP & Gene Ontology) [https://snps.biofold.org/snps-and-go/snps-and-go.html] is an SVM-based classifier that classifies polymorphisms as a neutral variation or disease-associated variation (when probability score >0.5) (Calabrese et al., 2009). Polyphen2 (Polymorphism phenotype v2) [http://genetics.bwh.harvard.edu/pph2/] analyzes the potential effect of amino acid substitution on the function and structure of protein and based on the probabilistic score it provides the result as benign, possibly damaging and probably damaging (Adzhubei et al., 2010). PANTHER (Protein Analysis Through Evolutionary Relationship) [http://www.pantherdb.org/tools/csnpScoreForm.jsp] employs PANTHER-PSEP (Position Specific Evolutionary Preservation) method to distinguish disease-related variants from neutral variants in the human protein. It estimates the likelihood of nsSNPs disrupting protein functionality by calculating the evolutionary preservation of the amino acid residues where a long preservation period indicates greater chances of nsSNPs causing a functional impact on protein (Tang & Thomas, 2016).

2.4 Structural Impact Prediction

The structural impact of nsSNPs on the RTEL1 protein was analyzed using 9 distinct web tools. For the change in stability and phenotypic effect prediction, seven tools (DUET, mCSM, SDM, I-Mutant, INPS-MD, MUpro, Dynamut) and two webservers (HOPE, MudPred2) were used respectively.

DUET [http://biosig.unimelb.edu.au/duet/] predicts the alteration in the stability of protein due to the introduced mutation by combining the SDM and mCSM approaches, therefore, both SDM and mCSM predicted results come together with DUET (Pires et al., 2014). In this tool, the PDB file of the wild-type protein along with the single mutation information were provided as input and the server gave the result of the change in folding free energy or value of ΔΔG in kcal/mol where the negative value indicates destabilization and a positive value indicates stabilization of the structure. MuPro [http://mupro.proteomics.ics.uci.edu/] predicts the effects of a single-site amino acid substitution on the stability of protein with 84% accuracy using protein sequence and mutation information (Cheng et al., 2005). I -Mutant 2.0 [https://folding.biofold.org/i-mutant/i-mutant2.0.html] assess the protein stability change from a given protein sequence and provide information about the state of stability as a decrease or increase in stability upon possible mutation along with Reliability Index. The INPS-MD (Impact of Non-synonymous mutations on Protein Stability – Multi Dimension) [https://inpsmd.biocomp.unibo.it/inpsSuite/default/index] can also predict the stability change of protein from both protein sequence and structure (Savojardo et al., 2016). The FASTA sequence was given as input and ΔΔG value was provided as output, where all the negative values were considered as destabilizing nsSNPs.

The webserver MutPred2 [http://mutpred2.mutdb.org/] uses machine learning-based algorithms that enable the prediction of pathogenicity of amino acid substitutions in proteins with a probabilistic score along with a list of specific alterations of the molecular mechanism (Pejaver et al., 2020). This tool predicts the pathogenicity of a given FASTA sequence of the protein. The effects of harmful nsSNPs on protein structure were examined using the HOPE [http://www.cmbi.ru.nl/hope/home] server. By combining data from numerous sources, such as sequence annotations, tertiary structure, homology models from the Distributed Annotation System (DAS) servers, UniProt database, etc., the Project HOPE server foresees the structural

effects of nsSNPs(Venselaar et al., 2010). The wild-type protein sequence was used as the input, and before starting the analysis, specific mutation sites and targeted mutations were specified.

2.5 Comparative Modeling & Evaluation of mutated 3D structures

The three-dimensional (3D) model of the mutant proteins was obtained through comparative modeling in Modeller 10.2 [https://salilab.org/modeller/] standalone software. The AlphaFold structure of wild-type protein was used as a template for the generation of altered protein structure. Each model was initially optimized using the variable target function approach with conjugate gradients and subsequently refined using molecular dynamics and simulated annealing in Modeller 10.2 (Webb & Sali, 2016). After completion of the 3D model generation, PyMOL 2.5 [https://pymol.org/2/] software was utilized for analyzing the root mean square deviation (RMSD) value of each mutant structure. By superimposing native and mutant structures, this tool forecast the RMSD value, which aids in identifying the closet related structural analog.

2.6 Secondary Structure Analysis

To analyze the secondary structure, all 10 variants sequence along with the native sequence were evaluated using PDBsum [http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html]. Taking sequence as input, the server returns the best match structure list of the given sequence and from there the closest matched structure was selected. This database provides several information of the protein structure across different pages or tabs which can be accessed from the top of the page denoted as TOP, PROTEIN, CLEFT, etc (Laskowski et al., 2018). From the protein tab, the secondary structure information of the protein was collected. Here, only the wiring diagram and structural motif information from the protein tab were selected for analyzing the in-depth information about the secondary structure of all 10 mutant and wild-type structures.

2.7 Identification of Domain and Cluster

Mutation 3D [http://mutation3d.org/] was utilized to assess the arrangements of SNPs on protein models or structures and to look for the functional domain information of the SNP positions (Meyer et al., 2016). Through the complete-linkage clustering procedure, this tool also identifies clusters of amino acid substitutions in protein structure which indicates the positions that have the most impact on the structure of a protein. For the analysis of the functional domain and cluster of the filtered SNPs, the protein symbol along with the mutations was submitted to the Mutation 3D web interface.

2.8 Prediction of Post Translational Modification Sites

MusiteDeep [https://www.musite.net/] was employed to predict the putative PTM sites in RTEL1 protein. Utilizing a deep learning-based algorithm and depending on the confidence threshold, with a default cut-off of 0.5, MusiteDeep predicts and identifies the desired PTM sites in the sequence (D. Wang et al., 2020). The FASTA format of the protein sequence was submitted to the server with the selection of all prediction models.

2.9 Conservation, Surface Accessibility & Evolutionary Relationship

The evolutionary conserved amino acid position in RTEL1 protein was interpreted using ConSurf [https://consurf.tau.ac.il/consurf_index.php] web server (Ashkenazy et al., 2016). In this server, the evolutionary profile is computed by searching for homologous sequences along with multiple sequence alignment, followed by generating a phylogenetic tree using a neighbor-joining algorithm. Moreover, through the Bayesian method (Mayrose et al., 2004)or a maximum likelihood algorithm(Pupko et al., 2002), this tool enumerates a site-specific conservation score from 1 to 9, with 9 representing a highly conserved position. The analysis can be done by using protein structure or protein sequence (Berezin et al., 2004). Here, to anticipate the conservation score and color scheme, the FASTA format of the protein sequence was inserted as input with default parameters.

NetSurfP-2.0 [http://www.cbs.dtu.dk/services/NetSurfP/] is a sequence-based web server that employs convolutional and long short-term memory neural network architecture to predict structural features such as surface accessibility, structural disorder, and secondary structure for each amino acid position (Klausen et al., 2019). To assess the surface accessibility of each amino acid residue of the RTEL1 protein, the protein sequence was run within the default parameter in the NetSurfP-2.0 server.

A phylogenetic tree of the ten closest matches to the human RTEL1 protein, determined by BLASTp search [https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins], was constructed in MEGA11 software using the maximum likelihood technique and a bootstrap parameter of 1000(Tamura et al., n.d.). This enables us to elucidate the evolutionary relationship of the RTEL1 protein. The tree was then visualized using the Iroki web server [https://www.iroki.net/] (Moore et al., 2020).

2.10 Interatomic Interaction Prediction

The interatomic interaction was predicted by implementing several programs of PyMOL 2.5 software [https://pymol.org/2/] which helps to visualize the change in atomic interaction in amino acid residues due to any single mutation. For this, the PDB file of the mutant structures or wild type structure was opened in PyMOL and from the sequence, the specific amino acid residue was selected. The polar contacts of selected residue with other atoms were searched for and the distance between the atoms was measured.

2.11 Molecular Docking Analysis

Using the HDOCK [http://hdock.phys.hust.edu.cn/] web server, molecular docking with telomeric DNA corresponding to PDB ID 1W0U, was performed on the selected most harmful mutant structures, along with the native structure. HDOCK server can perform docking on both the FASTA sequence and pdb file to predict the binding complexes between protein and nucleic acid by following the hybrid docking approach (Yan et al., 2017, 2020). For the input molecule in the

server, protein structure (both wild type and mutant) and DNA structure were provided as receptor molecule and ligand molecule respectively.

From literature, the HHD2 (Harmonin Homology Domain 2) domain of RTEL1 was found to interact directly with DNA (Kumar et al., 2022). Therefore, to specify the binding site, the positions of the HHD2 domain (A1059, V1060, S1061, A1062, Y1063, L1064, A1065, D1066, A1067, R1068, R1069, G1075, S1077, Q1078, L1079, L1080, A1081, A1082, T1084, K1087, D1090, and D1134) mentioned in the literature were used here as a receptor binding site residue and TTAGGG motif and its complementary sequence positions were selected from both strands (chain C and chain D) of DNA for ligand binding site residue. From the provided HDOCK result, docked models were chosen based on the following criteria: smaller docking score, confidence score >=0.5, and smaller RMSD value and subjected to DNAproDB [https://dnaprodb.usc.edu/] to visualize the interaction patterns that each complex formed (Sagendorf et al., 2017, 2020).

Chapter 3

Result

3.1 SNP Annotation

The Single Nucleotide Polymorphism data about the human RTEL1 gene was retrieved from the NCBI dbSNP database. Among the total 20734 SNPs from the search result, 25 are inframe deletions, 17554 are in the intronic region, 1392 are missense (non-synonymous), 2522 are non-coding variants, and 781 are synonymous. For this study, only the nsSNPs or missense SNPs (a total of 1392) were filtered for subsequent analysis.

3.2Assessment of RTEL1 Protein Structure

The tertiary structure of the protein determines its properties and capacity for interacting with ligands. As no crystal structure was found in the protein data bank for human RTEL1 protein, therefore, the AlphaFold structure of RTEL1 protein was taken from UniProt. The structure was validated using the SAVES server, where ERRAT provided 91.1036 for the overall quality factor, Verify-3D revealed that 52.83% of the residues have an average 3D-1D score of 0.2, and PROVE predicted a total of 4.9% buried outliner protein atoms. The Ramachandran plot, which is available in PROCHECK, was utilized to further evaluate the quality of the 3D protein structure. The plot from the AlphaFold model revealed that 93.7% of the residues are in the favored region, 10.9% are in the additional allowed region, 2.0% are in the generously allowed region and 3.4% are in the disallowed region.

Table 1: Ramachandran plot parameters of RTEL1 wild-type protein's AlphaFold structure.

Ramachandran Plot Statistics for AlphaFold Structure	
Residues in most favoured regions (%)	83.70%
Residues in additional allowed regions (%)	10.90%
Residues in generously allowed region (%)	2.00%
Residues in disallowed region (%)	3.40%

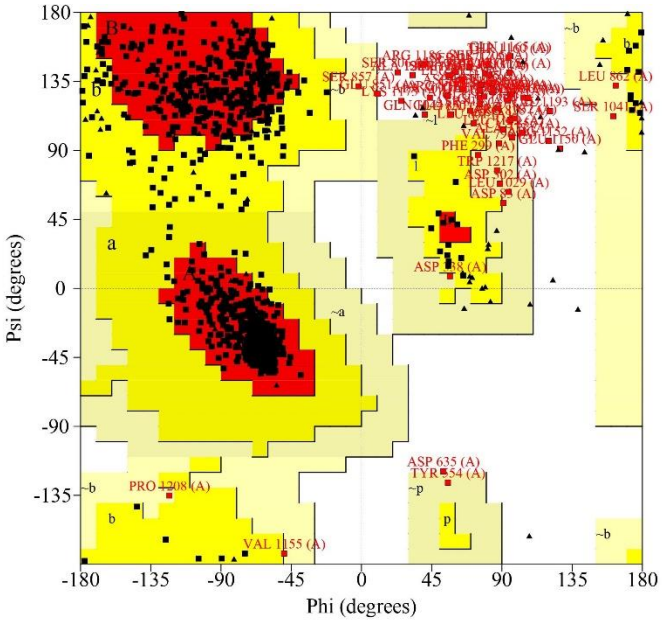


Figure 5: Ramachandran plot of native RTEL1 derived from PROCHECK

The general conclusions drawn from the aforementioned results pointed to the good quality of our protein structure, which allowed it to be used to subsequent investigation.

3.3 Determination of functional consequences of RTEL1 nsSNPs

The functional impact of nsSNPs on RTEL1 has been assessed using a total of ten tools.

SIFT predicted 441 as damaging, where 88 had a low confidence score. Therefore, after eliminating the redundancies, 353 remained as the most functionally detrimental ones. Out of the submitted 1392 nsSNPs, the PROVEAN server identified 489 as potentially harmful. PolyPhen-2 and Panther anticipated 386 and 579 as probably damaging ones respectively. Both of these tools generate three types of results: benign, possibly damaging, and probably damaging; however, only the probably damaging outputs from Polyphen2 and Panther were selected because this indicates the most deleterious SNPs with the highest confidence score. Moreover, SuSPect provides a list of scores ranging from 0-100 for each variant that is likely to be disease-causing and the recommended cutoff is 50 for the most deleterious ones. Therefore, 72 disease-causing variants with a score>=50 were chosen from the SuSPect output.

Table 2: High-risk nsSNPs anticipated by five web tools.

SNP ID	Variants	SIFT	PROVEAN	PolyPhen2	Panther	SuSPect
rs1296968885	F15L	Deleterious	Deleterious	Damaging	Damaging	Disease
rs2089974057	M25V	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1044603913	T44M	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1163455875	T49M	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1407364310	L50P	Deleterious	Deleterious	Damaging	Damaging	Disease
rs778739638	C54F	Deleterious	Deleterious	Damaging	Damaging	Disease
rs554402067	S113P	Deleterious	Deleterious	Damaging	Damaging	Disease
rs963067934	T115S	Deleterious	Deleterious	Damaging	Damaging	Disease
rs746010778	H116P	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1333433189	S140F	Deleterious	Deleterious	Damaging	Damaging	Disease
rs746931551	R141W	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1242481082	R141Q	Deleterious	Deleterious	Damaging	Damaging	Disease
rs772748212	P225L	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1188469323	N227D	Deleterious	Deleterious	Damaging	Damaging	Disease
rs923910999	Y228C	Deleterious	Deleterious	Damaging	Damaging	Disease
rs764165415	D231V	Deleterious	Deleterious	Damaging	Damaging	Disease
rs398123019	E251K	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1454150484	A252V	Deleterious	Deleterious	Damaging	Damaging	Disease
rs748740521	H253Y	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1386490624	P460L	Deleterious	Deleterious	Damaging	Damaging	Disease
rs2090629066	T478N	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1012871786	G480R	Deleterious	Deleterious	Damaging	Damaging	Disease
rs773057452	T481M	Deleterious	Deleterious	Damaging	Damaging	Disease
rs786205700	P484L	Deleterious	Deleterious	Damaging	Damaging	Disease
rs374168761	F559L	Deleterious	Deleterious	Damaging	Damaging	Disease
rs535749230	A621V	Deleterious	Deleterious	Damaging	Damaging	Disease
rs753779060	R624Q	Deleterious	Deleterious	Damaging	Damaging	Disease
rs2090674658	S628G	Deleterious	Deleterious	Damaging	Damaging	Disease
rs766112578	E629K	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1262691904	R639C	Deleterious	Deleterious	Damaging	Damaging	Disease

rs1484865003	R639H	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1262230406	G640D	Deleterious	Deleterious	Damaging	Damaging	Disease
rs367688683	G645D	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1323023332	L646F	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1177091623	P647L	Deleterious	Deleterious	Damaging	Damaging	Disease
rs16983886	K659N	Deleterious	Deleterious	Damaging	Damaging	Disease
rs760108811	G696R	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1176364985	R697Q	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1416515129	R700Q	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1332347156	G706R	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1472657490	P725R	Deleterious	Deleterious	Damaging	Damaging	Disease
rs1555811919	R729C	Deleterious	Deleterious	Damaging	Damaging	Disease
rs2090803986	H960R	Deleterious	Deleterious	Damaging	Damaging	Disease

PredictSNP integrates the results of six (MAPP, PhD-SNP, Polyphen1, Polyphen2, SIFT, SNAP) best-performing tools while PredictSNP2 combines the results of five top tools (CADD, DANN, FATHMM, FunSeq2, GWAVA) and gives a consensus score. Here, only the consensus score from both of these tools was taken into consideration, where PredictSNP and PredictSNP2 identified 309 and 364 as deleterious, respectively. In addition, 280 nsSNPs were found to be pathological in P-Mut, 505 nsSNPs were predicted to be impactful in SNAP2, and 166 nsSNPs were found to be disease-associated in SNP and GO.

Table 3: High-risk nsSNPs predicted by five web tools.

SNP ID	Variants	PredictSNP	PredictSNP2	P-Mut	SNAP2	SNP &GO
rs1296968885	F15L	Deleterious	Deleterious	Disease	Effect	Disease
rs2089974057	M25V	Deleterious	Deleterious	Disease	Effect	Disease
rs1044603913	T44M	Deleterious	Deleterious	Disease	Effect	Disease
rs1163455875	T49M	Deleterious	Deleterious	Disease	Effect	Disease
rs1407364310	L50P	Deleterious	Deleterious	Disease	Effect	Disease
rs778739638	C54F	Deleterious	Deleterious	Disease	Effect	Disease
rs554402067	S113P	Deleterious	Deleterious	Disease	Effect	Disease
rs963067934	T115S	Deleterious	Deleterious	Disease	Effect	Disease

rs746010778	H116P	Deleterious	Deleterious	Disease	Effect	Disease
rs1333433189	S140F	Deleterious	Deleterious	Disease	Effect	Disease
rs746931551	R141W	Deleterious	Deleterious	Disease	Effect	Disease
rs1242481082	R141Q	Deleterious	Deleterious	Disease	Effect	Disease
rs772748212	P225L	Deleterious	Deleterious	Disease	Effect	Disease
rs1188469323	N227D	Deleterious	Deleterious	Disease	Effect	Disease
rs923910999	Y228C	Deleterious	Deleterious	Disease	Effect	Disease
rs764165415	D231V	Deleterious	Deleterious	Disease	Effect	Disease
rs398123019	E251K	Deleterious	Deleterious	Disease	Effect	Disease
rs1454150484	A252V	Deleterious	Deleterious	Disease	Effect	Disease
rs748740521	H253Y	Deleterious	Deleterious	Disease	Effect	Disease
rs1386490624	P460L	Deleterious	Deleterious	Disease	Effect	Disease
rs2090629066	T478N	Deleterious	Deleterious	Disease	Effect	Disease
rs1012871786	G480R	Deleterious	Deleterious	Disease	Effect	Disease
rs773057452	T481M	Deleterious	Deleterious	Disease	Effect	Disease
rs786205700	P484L	Deleterious	Deleterious	Disease	Effect	Disease
rs374168761	F559L	Deleterious	Deleterious	Disease	Effect	Disease
rs535749230	A621V	Deleterious	Deleterious	Disease	Effect	Disease
rs753779060	R624Q	Deleterious	Deleterious	Disease	Effect	Disease
rs2090674658	S628G	Deleterious	Deleterious	Disease	Effect	Disease
rs766112578	E629K	Deleterious	Deleterious	Disease	Effect	Disease
rs1262691904	R639C	Deleterious	Deleterious	Disease	Effect	Disease
rs1484865003	R639H	Deleterious	Deleterious	Disease	Effect	Disease
rs1262230406	G640D	Deleterious	Deleterious	Disease	Effect	Disease
rs367688683	G645D	Deleterious	Deleterious	Disease	Effect	Disease
rs1323023332	L646F	Deleterious	Deleterious	Disease	Effect	Disease
rs1177091623	P647L	Deleterious	Deleterious	Disease	Effect	Disease
rs16983886	K659N	Deleterious	Deleterious	Disease	Effect	Disease
rs760108811	G696R	Deleterious	Deleterious	Disease	Effect	Disease
rs1176364985	R697Q	Deleterious	Deleterious	Disease	Effect	Disease
rs1416515129	R700Q	Deleterious	Deleterious	Disease	Effect	Disease
rs1332347156	G706R	Deleterious	Deleterious	Disease	Effect	Disease
rs1472657490	P725R	Deleterious	Deleterious	Disease	Effect	Disease

rs1555811919	R729C	Deleterious	Deleterious	Disease	Effect	Disease
rs2090803986	H960R	Deleterious	Deleterious	Disease	Effect	Disease

Among 1392 nsSNPs, 43 were deemed to be functionally harmful by all ten different tools and the remaining SNPs were assumed to be neutral in at least one of these tools. So, by taking into consideration only the common variants predicted by all ten tools, 43 nsSNPs (Table 1) were filtered out for further analysis.

3.4 Determination of Structural Impact of RTEL1 nsSNPs

For the determination of the structural impact of nsSNPs on RTEL1 protein, the filtered nsSNPs from the upstream analysis were subjected to nine different tools. Among these nine tools, seven were utilized for the prediction of stability change, and two were used for phenotypic effect prediction.

3.4.1 Analysis of protein stability

The change in the structural stability of RTEL1 protein due to the introduction of point mutations was predicted through seven bioinformatics-based web tools. The 43 deleterious nsSNPs were run to check the structural stability of proteins in the DUET server, which includes the results of both mCSM and SDM. mCSM, SDM, and DUET predicted 36, 30, and 33 nsSNPs as destabilizing for RTEL1 protein, respectively. To increase the accuracy of our predictions of changes in protein stability caused by single AA mutations, all 43 variants were analyzed through I-Mutant, INPS-MD, Mupro, and Dynamut. I-Mutant anticipated 34 and MuPro analyzed 41 as stability-decreasing nsSNPs. Moreover, 40 nsSNPs with a negative ΔΔG score were considered destabilizing in the INPS-MD result. Lastly, by combining the structure or NMA-based prediction (ΔΔG ENCoM) and vibrational entropy change (ΔΔS ENCoM) between mutant and wild-type structures, Dynamut provides the ΔΔG prediction score for each amino acid substitution. Here, 22 nsSNPs were predicted to be destabilizing by Dynamut.

Table 4: Destabilizing nsSNPs identified by seven in silico tools

nsSNPs	mCSM	SDM	DUET	I-Mutant	INPS-MD	MuPro	Dynamut
F15L	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease
M25V	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease
S113P	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease
H116P	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease
R141Q	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease
Y228C	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease
A252V	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease
G480R	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease
R639H	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease
G645D	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease	Decrease

Combining the findings from seven tools, 14 nsSNPs were identified frequently by all of these tools as being extremely detrimental based on their effects on the structural stability of proteins.

3.4.2 Prediction of Phenotypic Effects

The phenotypic effects of 14 functionally damaging SNPs were computed using MutPred2 and Project HOPE. Together with the P-value and probability score, some predictions made using MudPred2 were: loss or gain of allosteric site, catalytic site, helix, relative solvent accessibility, increase in various types of modification such as transmembrane protein, DNA, ligand, metal binding, or ordered interface, etc. Besides that, a MutPred2 score was given, with a cutoff of 0.50, which determines the overall probability of pathogenicity. The score goes from 0 to 1, and as the score rises, it becomes more likely that the SNP-induced alterations can influence the molecular mechanism of disease. Out of 14 total SNPs, 13 were identified as having higher pathogenic potential. Additionally, the mutations were submitted to HOPE for analysis. According to HOPE

results, 9 of the 14 mutant amino acids differed in charge, 2 differed in the level of hydrophobicity, and all of the 14 mutant residues were predicted to differ in size from the wild-type residue. These differences in size, charge, and hydrophobicity can interfere with the interactions of the nearby amino acid residues and also with protein folding. Aside from these, amino acid substitution also has an impact on numerous other attributes. For example, substitutions, where proline or glycine was a wild-type or changed amino acid, may disrupt the conformation by interfering with the rigidity or flexibility of the structure as prolines are rigid and glycine is a more flexible amino acid.

Table 5: Damaging nsSNPs predicted by MutPred2 and Project HOPE.

nsSNPs	MutPred2	HOPE
F15L	Damaging	Damaging
M25V	Damaging	Damaging
S113P	Damaging	Damaging
H116P	Damaging	Damaging
R141Q	Damaging	Damaging
Y228C	Damaging	Damaging
A252V	Damaging	Damaging
G480R	Damaging	Damaging
R639H	Damaging	Damaging
G645D	Damaging	Damaging

Finally, 13 nsSNPs were repeatedly recognized by MutPred2 and Project HOPE web server as being particularly harmful based on their effects on protein phenotype. These SNPs were found to induce a decrease in protein stability along with having a negative impact on the other properties.

3.5 Three Dimensional Structure Prediction for Mutant Proteins

In order to investigate whether the selected nsSNPs cause any alteration in the resultant protein, comparative 3D modeling along with a structural comparison between native and mutant structures was carried out through Modeller 10.2 followed by PyMOL 2.5 software. To generate the sequence for each individual variant, the wild-type amino acid residues located in the selected deleterious SNP positions in the RTEL1 protein sequence were replaced with the mutant amino acid. The mutated protein sequence was then utilized in Modeller 10.2 to develop the 3D structure for each variant using the AlphaFold structure as a template.

Next, the RMSD values of the mutant models were examined in PyMOL 2.5 for the investigation of structural similarity between the native and mutant structures. All the mutant models were observed to have a high RMSD value (Table 6) when superimposed over the native structure. As the larger RMSD value demonstrates greater deviation between wild-type and mutant structures, all 13 nsSNPs were therefore taken into consideration for the following investigation.

Table 1: RMSD value of 13 mutant models generated through PyMOL

Mutated Protein Model	RMSD
F15L	0.471
M25V	0.393
R141Q	0.455
Y228C	0.449
A252V	0.685
G480R	0.437
R639H	0.396
G645D	0.526
R697Q	0.511
R700Q	0.580
G706R	0.418
R729C	0.607
H960R	0.574

3.6 Investigation of the Impact of nsSNPs on Secondary Structure

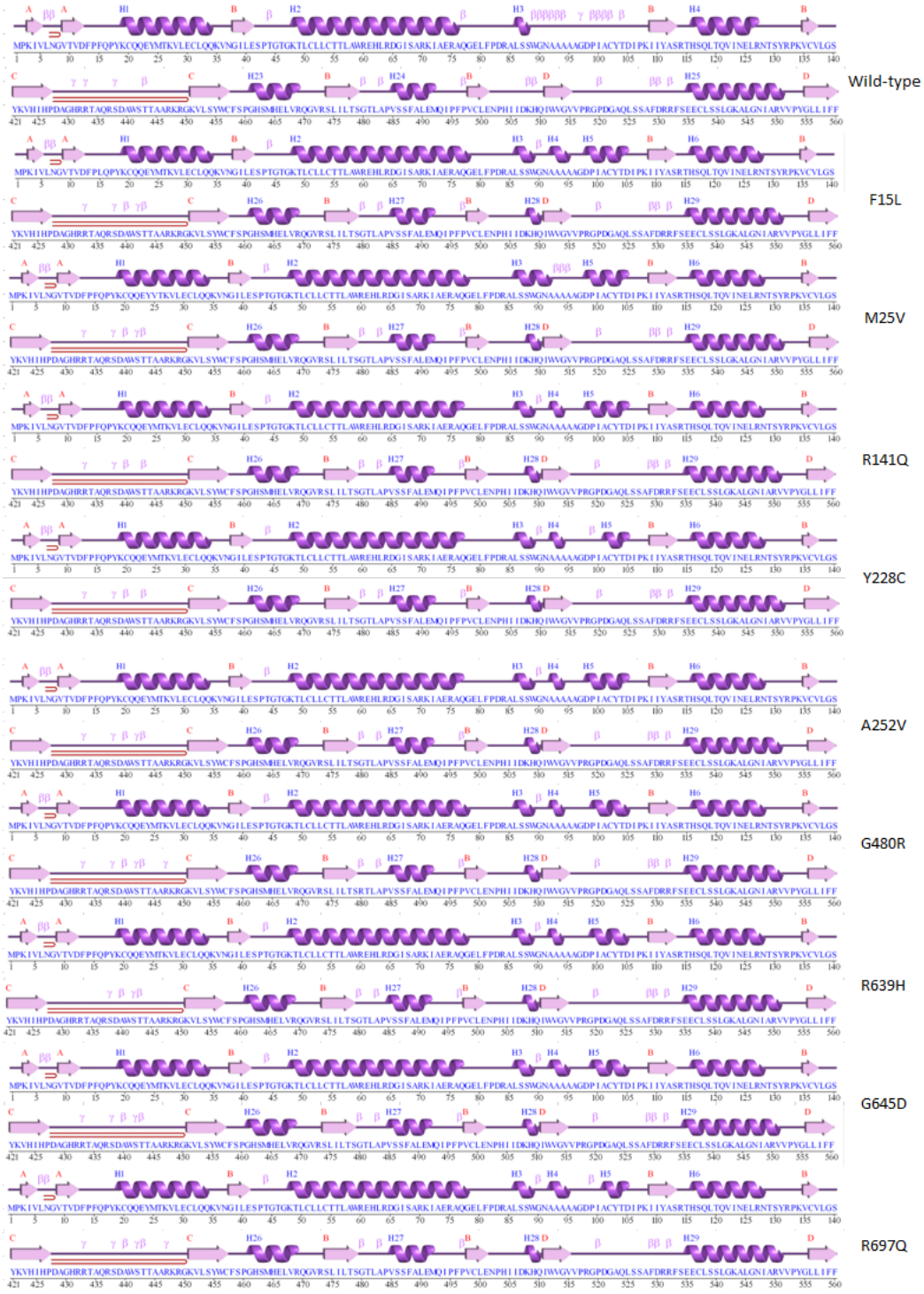
The prediction of secondary structure conformation of RTEL1 and 13 mutants were performed in the PDBsum web tool. From the output of the tool, it was found that both the wild-type and mutant structures have the same number of strands, sheets, beta hairpins, and beta alpha beta units. Apart from the number of helix-helix interactions, which remained the same in the native and M25V mutant structure, the number of helices and helix-helix interactions were increased in mutant structures compared to the native structure. Additionally, in all mutant structures, the amount of beta and gamma turns was reduced which is shown in Table 7.

Table 2: ProMotif information of native and mutant protein

Protein	**Sheets**	**beta alpha beta units**	**beta hairpins**	**strands**	**helices**	**helix-helix interacts**	**beta turns**	**gamma turns**
Wild-type	4	3	3	20	53	74	67	14
F15L	4	3	3	20	58	80	41	7
M25V	4	3	3	20	60	74	43	10
R141Q	4	3	3	20	58	76	40	9
Y228C	4	3	3	20	58	77	42	9
A252V	4	3	3	20	58	75	40	9
G480R	4	3	3	20	57	75	43	9
R639H	4	3	3	20	58	75	41	7
G645D	4	3	3	20	58	79	40	8
R697Q	4	3	3	20	58	77	41	10
R700Q	4	3	3	20	57	76	41	10
G706R	4	3	3	20	58	81	40	8
R729C	4	3	3	20	58	75	42	9
H960R	4	3	3	20	58	77	40	8

Furthermore, in the native structure, positions W89 to D105 had a large number of closely packed beta turns, whereas, in the mutant structures, this varied widely (either entirely absent or 2/3 beta

turns were present). Besides, F15L, M25V, A252V, G480R, R639H, G645D, R697Q, and R700Q mutants showed more tightly packed beta and gamma turns after position A429 than R141Q Y228C, G706R, R729C, and H960R mutants.



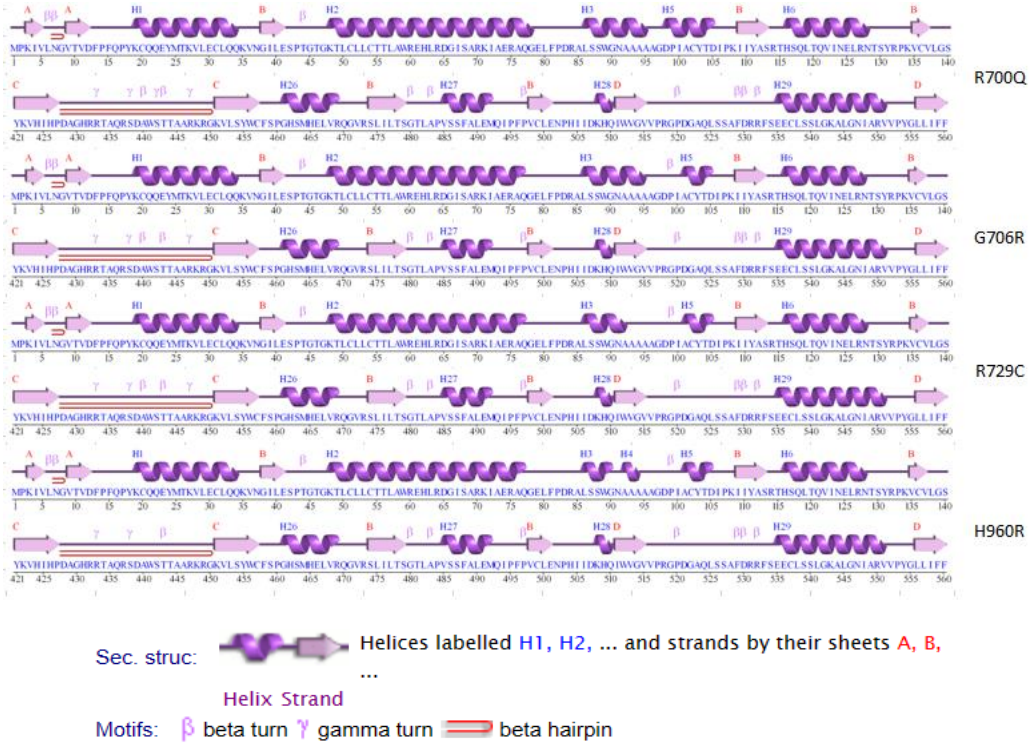


Figure 6: Analysis of the secondary structures of wild-type and mutated RTEL1 proteins using PDBsum. It displays the changes brought on by nsSNPs in terms of alpha helices, beta strands, and other patterns.

3.7 Identification of Domains and Clusters

Mutation 3D was used to predict mutant positions in domains and clusters, and the tool predicted two domains based on the submitted data. Dead 2 domain (111–272) contains R141Q, Y228C, and A252V mutants, and Helicase C2 domain (545–731) contains R639H, G645D, R697Q, R700Q, G706R, and R729C mutants. Moreover, the tool projected three ModBase models. The original model featured two clusters, including R639H, R697Q, R700Q, and G706R in one cluster and M25V, A252V, and G480R in the second cluster. The second model, on the other hand, featured one cluster, which housed R639H, R697Q, R700Q, and G706R. Furthermore, the third model lacks a cluster.

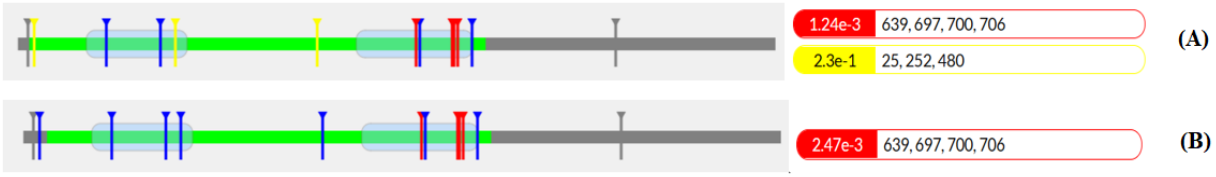


Figure 7: Domain & cluster information of nsSNPs represented in linear protein model. Clusters of (A) ModBase Model 1 and (B) ModBase Model 2 are shown in red and blue boxes (on the right side). Helicase C2 (right) and Dead2 (left) domains are indicated as a light blue transparent box in the highlighted green region of the linear model and the position of amino acid substitutions are portrayed in vertical lines

According to the findings of Mutation3D, seven mutants were found to be part of a cluster, which indicates that these mutations may have the greatest impact on the structure of the protein. Even though the rest of the mutants were not predicted to form clusters, we kept all of them for further analysis as those were predicted to be deleterious in former investigations.

3.8 Effect of nsSNPs on Post Translational Modification Sites

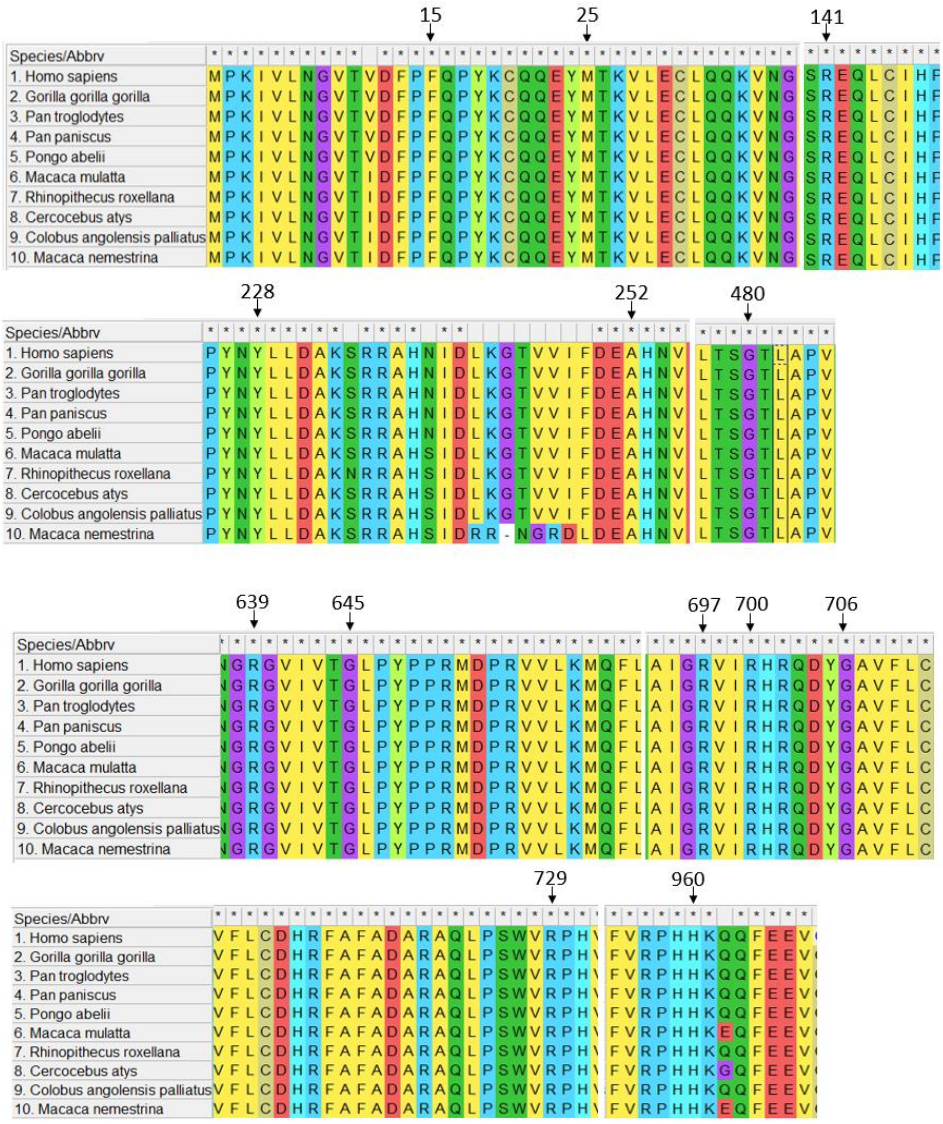
Post Translational Modification (PTM) entails the procedure by which proteins are chemically modified and induce conformational changes that allow the protein to be functional and participate in specific biological activities. To predict the potential PTM sites in RTEL1 and the effects of SNPs on PTM sites, MusiteDeep was used. A total of 8 types of 74 PTM sites were predicted for the protein sequence. Among all the selected deleterious SNPs, only the R639 position was predicted to be in a methylation site. Studies have linked methylation to the fine-tuning of a variety of biological processes, resulting in the formation of numerous diseases (Ramazi & Zahiri, 2021). Thus, amino acid alteration in position 639 can be anticipated to result in PTM impairment, which might disrupt the stability of a protein or have an overall negative impact on it.

3.9 Analyzation of Evolutionary Relationship of RTEL1 Protein and Conservation Profile & Surface Accessibility of nsSNPs

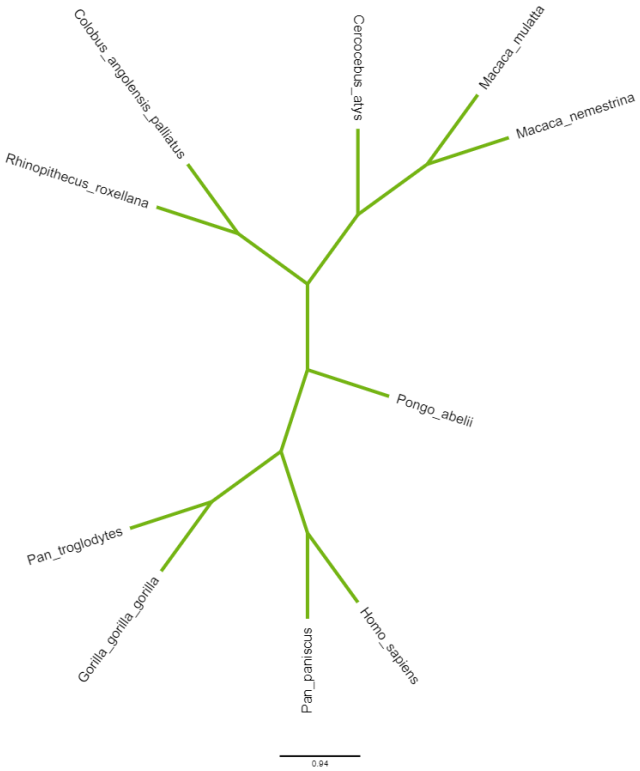
Despite the evolutionary change, amino acid residues essential for a variety of biological functions, including genome integrity, typically persist. Because of this, it is frequently believed that the degree of residue conservation indicates how crucial a location is to preserve the stability and functionality of a protein. In this regard, the conservation profile and surface accessibility of the 13 nsSNPs were analyzed through the ConSurf and NetSurf web tools, along with inspecting the evolutionary relationship of RTEL1 protein by using MEGA 11 software.

3.9.1 Evolutionary Relationship

The MEGA 11 program was used to analyze the conservation of the selected 13 SNP positions in ten different species along with phylogenetic analysis in order to determine the evolutionary relationships between these species. Then the tree was displayed by Iroki to examine evolutionary conservation. According to the findings, all of the amino acid positions are conserved among these ten species. Moreover, Pan paniscus, Pan troglodytes, and Gorilla gorilla are the three species that have been found to share the largest genetic similarity with human RTEL1 protein. So, according to the phylogenetic tree, it can be said that the RTEL1 protein is more conserved in primates.


15
25
141
Species/Abbrv
1. Homo sapiens
2. Gorilla gorilla gorilla
3. Pan troglodytes
4. Pan paniscus
5. Pongo abelii
6. Macaca mulatta
7. Rhinopithecus roxellana
8. Cercocebus atys
9. Colobus angolensis palliatus
10. Macaca nemestrina
228
252
480
639
645
697
700
706
729
960

(A)



(B)

Figure 8: (A) Evolutionary conservancy of 13 nsSNP position analyzed through multiple sequence alignment. (B) Graphical depiction of the evolutionary relationship of human RTEL1 with its closest relatives.

3.9.2 Analysis of Evolutionary conservation

To determine the conserved positions in the amino acid sequence of RTEL1 protein, the ConSurf server was used. Using the Bayesian approach, the ConSurf online browser assessed the degree of conservation of each protein residue along with identified potential structural and functional residues. The result showed twelve structural (buried) residues, two functional (exposed) residues, and one identified as exposed residue out of the fifteen residues that were filtered out from the upstream study, with the highly conserved profile. On the conservation scale of 1-9, thirteen positions exhibit the highest conservation profile with a conservation score of 9, and one position

(F15) has a high level of conservation with a conservation score of 8. However, position R729 scored 7 which still falls under the highly conserved category.

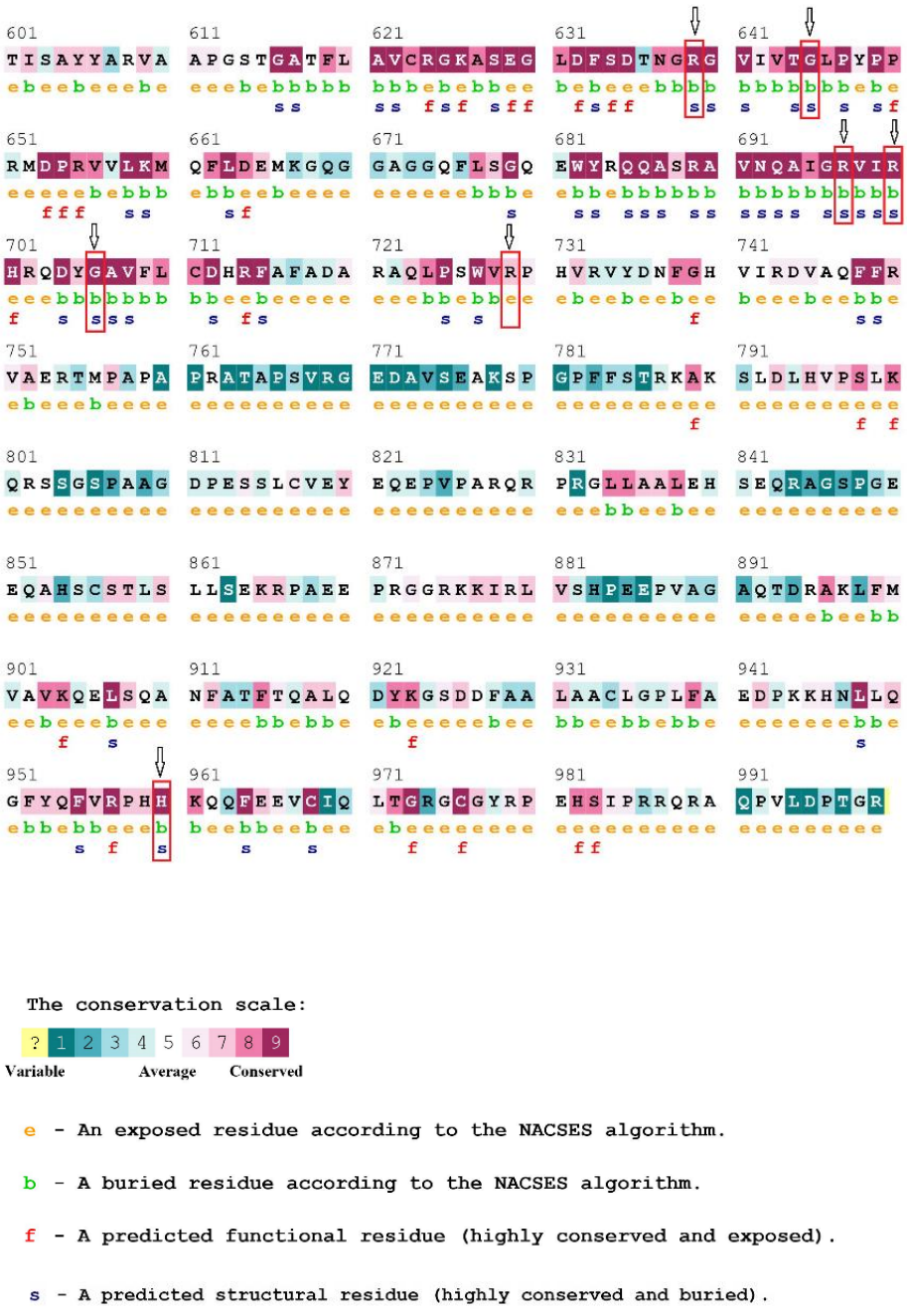


Figure 9*:* Evolutionary conservation profile prediction of RTEL1 protein using ConSurf web server. All of the nsSNPs identified as harmful belonged to highly conserved regions in the RTEL1 protein.

3.9.3 Evaluation of Surface accessibility

With the percentage scores, NetSurfP-2.0 estimated the surface accessibility of each amino acid site of the RTEL1 protein. The relative surface accessibility of each position in the amino acid sequence was predicted at a threshold of 25%, which meant that amino acid residues with scores of more than 25% were expected to be exposed, whilst residues with scores of less than 25% were assumed to be buried. Among thirteen selected positions, - R141, R639, and R729 each received a score of more than 25%. These amino acid residues were therefore anticipated to be in an exposed position, while the remaining 10 locations were expected to be in the buried zone, scoring less than 25%.

Table 3: Surface accessibility result of 10nsSNPs from NetSurf2.0

Class assignment	Amino acid	Sequence name	Amino acid number	Relative Surface Accessibility
B	F	sp_Q9NZ71_RTEL1_HU	15	0.116
B	M	sp_Q9NZ71_RTEL1_HU	25	0.012
E	R	sp_Q9NZ71_RTEL1_HU	141	0.315
B	Y	sp_Q9NZ71_RTEL1_HU	228	0.076
B	A	sp_Q9NZ71_RTEL1_HU	252	0.034
B	G	sp_Q9NZ71_RTEL1_HU	480	0.15
E	R	sp_Q9NZ71_RTEL1_HU	639	0.287
B	G	sp_Q9NZ71_RTEL1_HU	645	0.061
B	R	sp_Q9NZ71_RTEL1_HU	697	0.145
B	R	sp_Q9NZ71_RTEL1_HU	700	0.17
B	G	sp_Q9NZ71_RTEL1_HU	706	0.005
E	R	sp_Q9NZ71_RTEL1_HU	729	0.429
B	H	sp_Q9NZ71_RTEL1_HU	960	0.097

As any modification of amino acids in any highly conserved position is much more harmful than in any non-conserved position as well as the residues in the buried or exposed zone can also potentially hamper the structure of proteins and their interactions. Therefore, the outcomes from

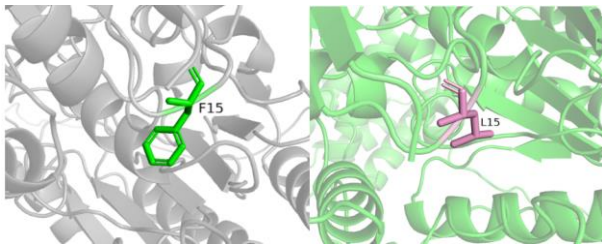
the aforementioned techniques indicated that the thirteen selected nsSNPs may have a significant impact on the RTEL1 protein.

3.10 Prediction of Interatomic Interaction

PyMOL2.5 software was used to evaluate ten mutated proteins chosen from upstream analyses to visualize the changes in interaction patterns with nearby amino acid residues.

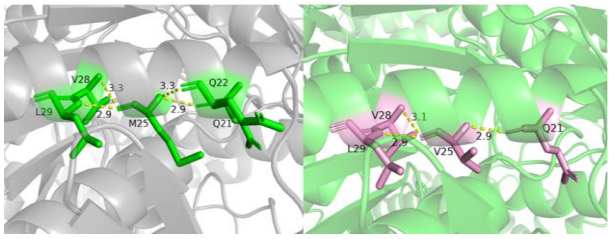
In the case of the substitution of phenylalanine with leucine at position 15, no alteration in interatomic interaction was observed. In the native structure, methionine at position 25 interacts with four nearby residues (Gln21, Gln22, Val28, and Leu29), whereas due to the substitution of methionine with valine, the number of interacting residues decreased to three and the distance remained quite similar to that of the wild-type residue. For the substitution of arginine with glutamine at position 141, only one interaction with Cys145 remained intact in the mutant, with the distance being decreased to 2.8 and the rest of the interactions were completely eliminated. Upon cysteine replacement of tyrosine at position 228, the distances with Pro225 and Asn227 atoms in the mutated RTEL1 protein structure were increased to 3.3 and 3, respectively, whereas the wild-type residue showed no interaction with Asn227. When comparing wild-type and mutant amino acids, it was found that the A252V mutation did not significantly alter the interaction pattern and that the distance between the neighboring residues (Val255 and Thr478) remained nearly unchanged. The distance between Gly480 and the nearby Ser479 residue was 2.9, as shown in Figure. Due to the glycine being replaced with arginine, the distance was reduced to 2.8, and three additional interactions with neighboring Thr44, Gly696 and Gln693 were introduced in the mutant structure. The mutation G645D formed an entirely new interaction with Ser527, Leu646, and Arg714 at a distance of 2.6, which was not seen in the wild-type residue. Moreover, the distance between Arg639 (wild type) and nearby Gly555, Asp635, Asp704, Tyr705, and Ala707 residues was 3, 3.4, 2.7, 2.9, and 2.9, whereas for His639 (mutant), the values were 3 and 2.9 for Gly555 and Ala707, respectively, and the rest of the interactions were not observed to persist in the mutated protein structure. Furthermore, when arginine was replaced with glutamine, the structure relaxed because five of the six interactions observed in the wild-type amino acid (Leu631, Asp632,

Phe633, Gly696, and Arg697) were completely eliminated in the mutant amino acid, while the remaining interaction (Asp704) showed the slightest increase (2.9 to 3.3) in distance. The mutant at position 960, on the other hand, had little effect on the interaction, where the interaction with Tyr922 was canceled out along with minimal fluctuation in other interactions with other neighboring atoms (as shown in the figure). Finally, among the rest 3 mutations, two mutations showed complete elimination of two interactions while the other showed the addition of two new interactions. In the case of R697Q, interaction with Ser628 and Leu631 was eliminated and the interacting distance with both Ala694 and Arg700 increased by at least 1Å. Similarly, contacts with nearby Pro725 and Val732 were unaffected by the cysteine substitution at position 729 in the mutant, and the distance was increased to 3.1 with Pro725 and reduced to 3.2 with Val732. Last but not least, when glycine was switched out for arginine at position 706, two new interactions with Gly638 and Gly640, at distances of 2.7 and 2.6, as well as a minor increase in the interacting distance with atoms comparable to the wild-type, were noticed (shown in Figure).



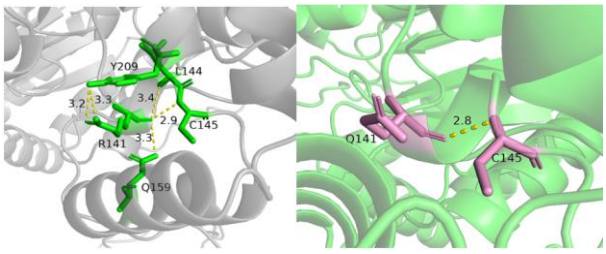
	Distance from nearby Amino acid atoms
Phe15 (Native)	NO INTERACTION
Leu15 (Mutant)	

(A)



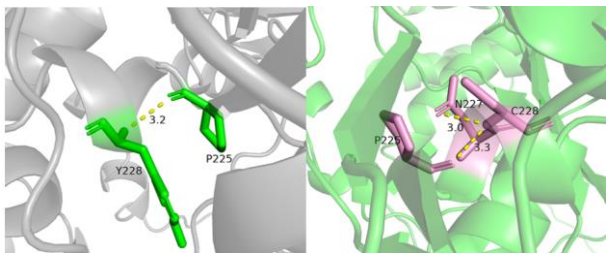
	Distance from nearby Amino acid atoms			
	Gln21	Gln22	Val28	Leu29
Met25 (Native)	2.9Å	3.3Å	3.3Å	2.9Å
Val25 (Mutant)	2.9Å	-	3.1Å	2.9Å

(B)



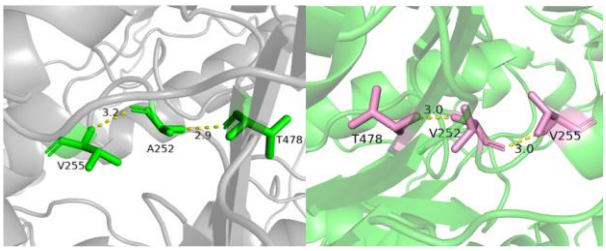
	Distance from nearby Amino acid atoms			
	Leu144	Cys145	Gln159	Tyr209
Arg141 (Native)	3.4Å	2.9Å	3.3Å	3.2Å & 3.3Å
Gln141 (Mutant)	-	2.8Å	-	-

(C)



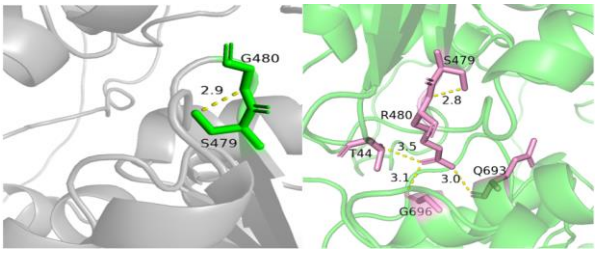
	Distance from nearby Amino acid atoms	
	Pro225	Asn227
Tyr228 (Native)	3.2Å	-
Cys228 (Mutant)	3.3Å	3Å

(D)



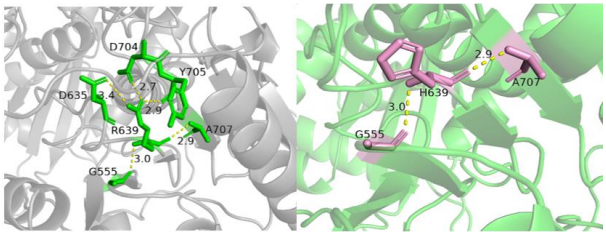
	Distance from nearby Amino acid atoms	
	Val255	Thr478
Ala252 (Native)	3.2Å	2.9Å
Val252 (Mutant)	3Å	3Å

(E)



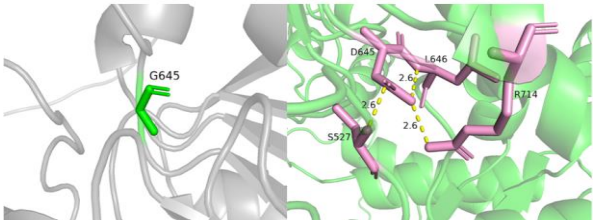
	Distance from nearby Amino acid atoms			
	Ser479	Thr44	Gly696	Gln693
Gly480 (Native)	2.9Å	-	-	-
Arg480 (Mutant)	2.8Å	3.5Å	3.1Å	3Å

(F)



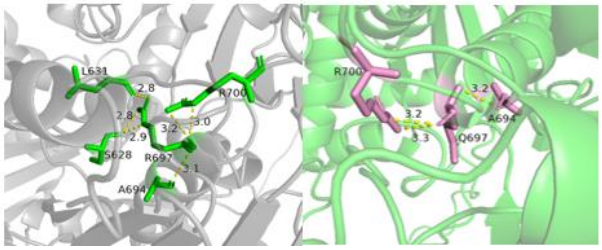
	Distance from nearby Amino acid atoms				
	Gly555	Asp635	Asp704	Tyr705	Ala707
Arg639 (Native)	3Å	3.4Å	2.7Å	2.9Å	2.9Å
His639 (Mutant)	3Å	-	-	-	2.9Å

(G)



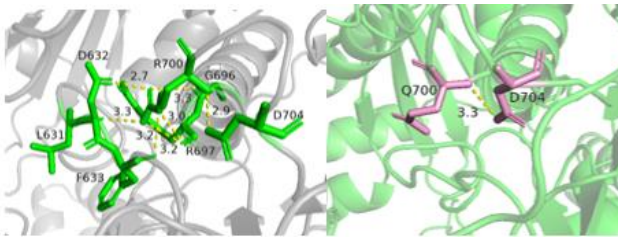
	Distance from nearby Amino acid atoms		
	Ser527	Leu646	Arg714
Gly645 (Native)	-	-	-
Asp645 (Mutant)	2.6Å	2.6Å	2.6Å

(H)



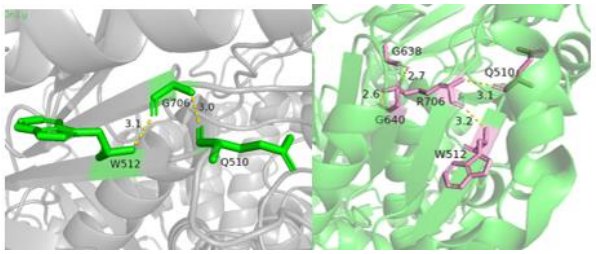
	Distance from nearby Amino acid atoms			
	Ser628	Leu631	Ala694	Arg700
Arg697 (Native)	2.8Å & 2.9Å	2.8Å	3.1Å	3.0Å & 3.2Å
Gln697 (Mutant)	-	-	3.2Å	3.2Å & 3.3Å

(I)



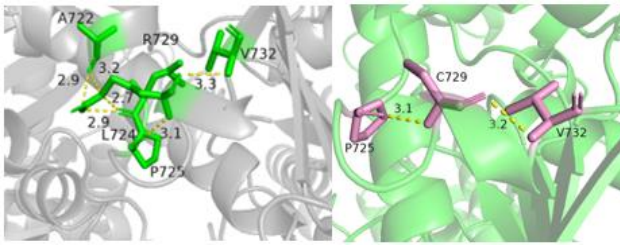
	Distance from nearby Amino acid atoms					
	Leu631	Asp632	Phe633	Gly696	Arg697	Asp704
Arg700 (Native)	3.3Å	2.7Å	3.2Å	3.3Å	3.0Å & 3.2Å	2.9Å
Gln700 (Mutant)	-	-	-	-	-	3.3Å

(J)



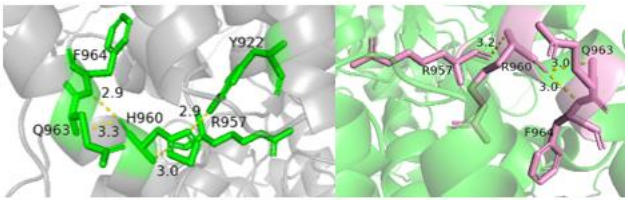
	Distance from nearby Amino acid atoms			
	Gln510	Trp512	Gly638	Gly640
Gly706 (Native)	3.0Å	3.1Å	-	-
Arg706 (Mutant)	3.1Å	3.2Å	2.7Å	2.6Å

(K)



	Distance from nearby Amino acid atoms			
	Ala722	Leu724	Pro725	Val732
Arg729 (Native)	2.9Å & 3.2Å	2.7Å & 2.9Å	3.1Å	3.3Å
Cys729 (Mutant)	-	-	3.1Å	3.2Å

(L)



	Distance from nearby Amino acid atoms			
	Tyr922	Arg957	Gln963	Phe964
His960 (Native)	2.9Å	3.0Å	3.3Å	2.9Å
Arg960 (Mutant)	-	3.2Å	3.0Å	3.0Å

(M)

Figure 10: Analyzation of the effect of nsSNPs on the interaction pattern with the neighboring residues. The distance from nearby amino acid atoms in (A) F15L, (B) M25V, (C) R141Q, (D) Y228C, (E) A252V, (F) G480R, (G) R639H, (H) G645D, (I) R697Q, (J) R700Q, (K) G706R, (L) R729C, and (M) H960R mutant structure are visualized using PyMOL2.5.

3.11 Molecular Docking

Due to RTEL1 being an essential DNA helicase, molecular docking of native and 13 filtered mutant proteins was performed with telomeric DNA. Active residues of the HHD2 domain in

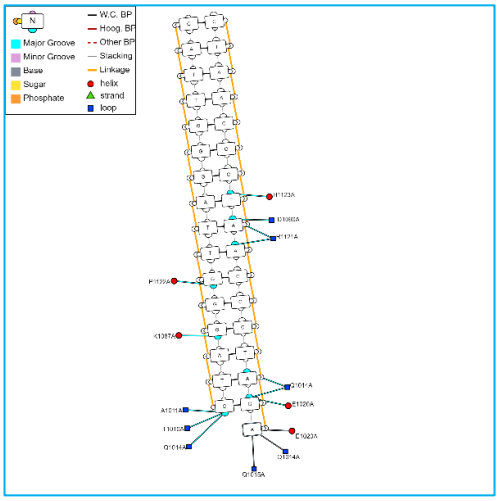
RTEL1 were extracted from the literature and used for specifying the DNA binding site in the HDOCK docking server.

A total of 14 molecular dockings were performed in the HDOCK server, which predicts binding complexes using a hybrid algorithm, to predict binding affinity. Some deviation in the orientation of the molecular complexes has been observed. Eight mutants (F15L, M25V, Y228C, A252V, G480, R639H, R697Q, and R729C) have been predicted to have a less negative docking score when binding with DNA than the wild type, indicating a less stable binding complex. Besides, three of the other mutants showed a more negative docking score, which might result in a more rigid binding complex, leading to a discrepancy in the functionality of proteins. The bound conformations revealed significant differences between the mutant and wild-type molecules when visualized in the DNAproDB web-based tool. All of the mutant proteins deviated from the wild type when binding to DNA, not only in terms of interacting residues but also the number of hydrogen bonds, Van der Waals interactions, and nucleic acid interactions. Additionally, in comparison to the wild type, the DNA has been observed to bind with entirely new residues in the F15L, M25V, Y228C, G706R, and R729C mutant proteins.

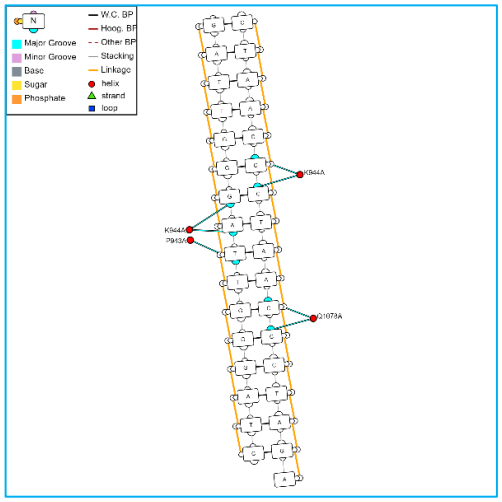
Table 4: Analysis of binding affinity and interaction of wild-type and mutant protein with DNA

Docked Molecules	Interacting Residues	Total HW Count	Total VdW count	Total Nuc.Int. Count	Docking Score	Confidence Score	Ligand rmsd (Å)
AF-DNA	H1123,D1090,R1121, P1122, K1087,Q1014, A1011,T1010, Q1015, E1020	9	70	25	-162.78	0.5636	163.85
F15L-DNA	P943,K944,Q1078	2	40	11	-108.52	0.3037	150.61
M25V - DNA	N947,P887,Q950, K1005,T1007,S1009	1	2	6	-147.45	0.4873	150.56
R141Q - DNA	R433,A1011,K1087, R1121,P1122,D1090, H1123,Q1126	1	45	19	-199.3	0.7283	159.89

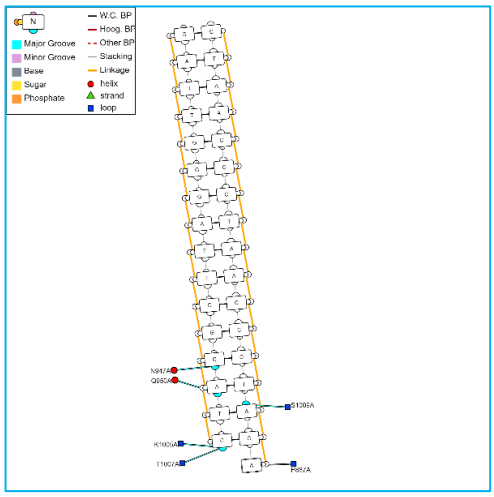
Y228C - DNA	W441,Q291,T287	0	24	12	-159.8	0.5488	147.01
A252V - DNA	A1011,E1020,K1087, Q1014,Q1015, Q1088,R1121	2	57	26	-137.35	0.4371	158.6
G480R - DNA	R895,Q954,Q1088, K1087,R1121, D1089,D1090	0	19	11	-129.01	0.3966	166.48
R639H - DNA	A440,A1101, K1106,W441	0	24	11	-137.36	0.4371	147.05
G645D - DNA	R433,S1009,T1010, A1011,A1012,K1087, Q1088,D1089,D1090, R1121,H1123	5	46	23	-180.79	0.6493	154.45
R697Q- DNA	A440,T443,K1087, Q1088,D1089,D1090, R1121, H1123	3	40	16	-147.01	0.4851	160.21
R700Q- DNA	Q436,R447,K1087, Q1088,D1090,R1121, P1122, H1123, H1124	5	72	20	-177.62	0.6347	156.1
G706R- DNA	E312,E313,T287, A316,K319,W441	0	21	9	-186.84	0.6763	146.61
R729C- DNA	T1007,S1077, Q1078,A1081	1	9	7	-131.14	0.4068	157.55
H960R- DNA	T1010,A1011,Q1014, K1087, Q1088,D1089, D1090,R1121,P1122, H1123	7	82	24	-196.81	0.7183	166.68


W.C. BP
Hoog. BP
Other BP
Stacking
Linkage
helix
strand
loop
Major Groove
Minor Groove
Base
Sugar
Phosphate

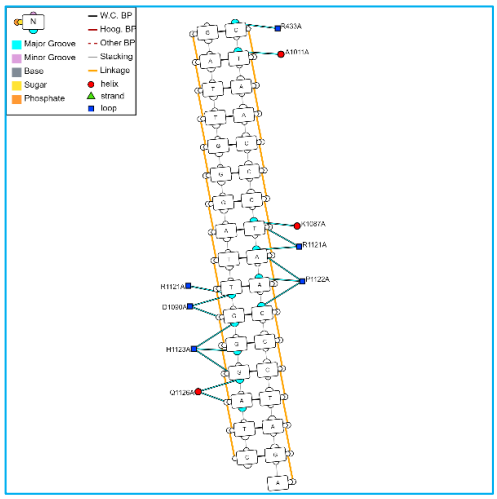
(A)


W.C. BP
Hoog. BP
Other BP
Stacking
Linkage
helix
strand
loop
Major Groove
Minor Groove
Base
Sugar
Phosphate

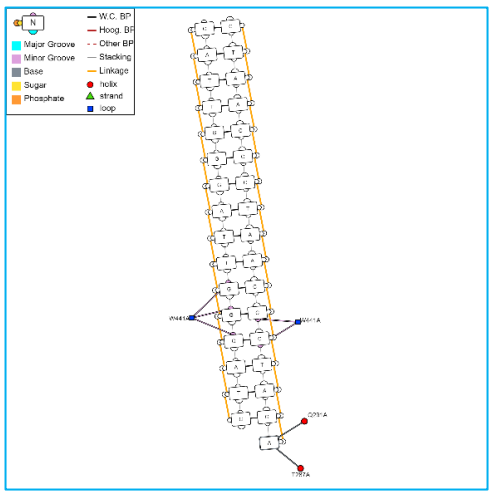
(B)


W.C. BP
Hoog. BP
Other BP
Stacking
Linkage
helix
strand
loop
Major Groove
Minor Groove
Base
Sugar
Phosphate

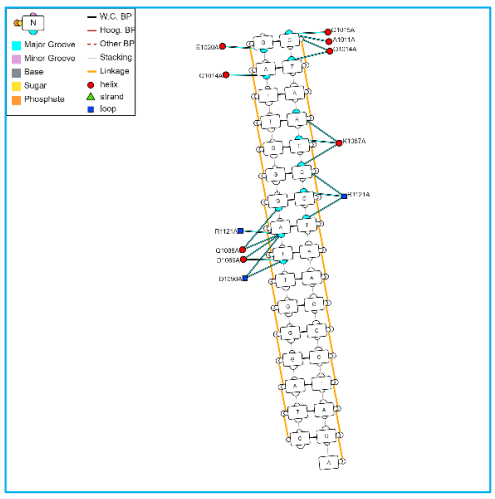
(C)


W.C. BP
Hoog. BP
Other BP
Stacking
Linkage
helix
strand
loop
Major Groove
Minor Groove
Base
Sugar
Phosphate

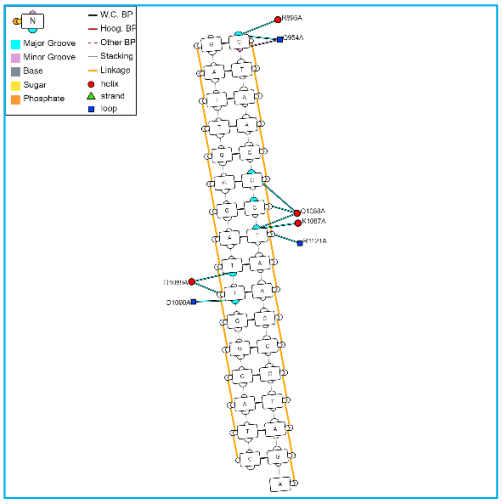
(D)


W.C. BP
Hoog. BP
Other BP
Stacking
Linkage
helix
strand
loop
Major Groove
Minor Groove
Base
Sugar
Phosphate

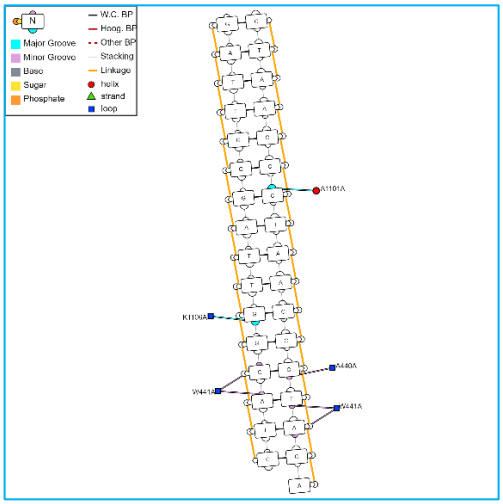
(E)


W.C. BP
Hoog. BP
Other BP
Stacking
Linkage
helix
strand
loop
Major Groove
Minor Groove
Base
Sugar
Phosphate

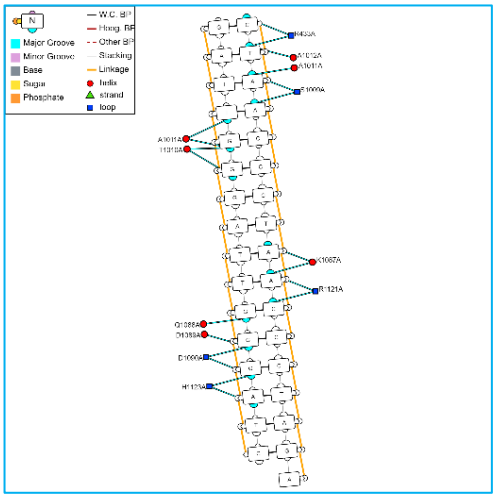
(F)



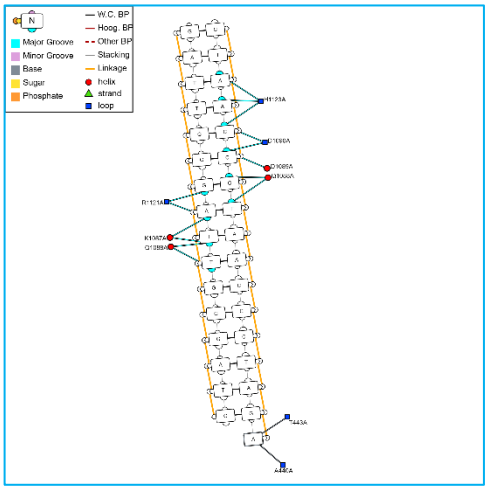
(G)



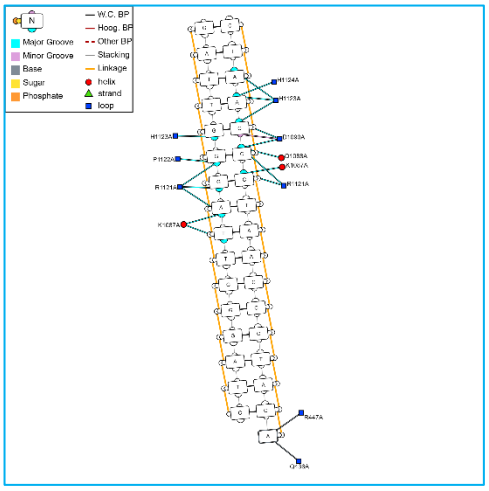
(H)



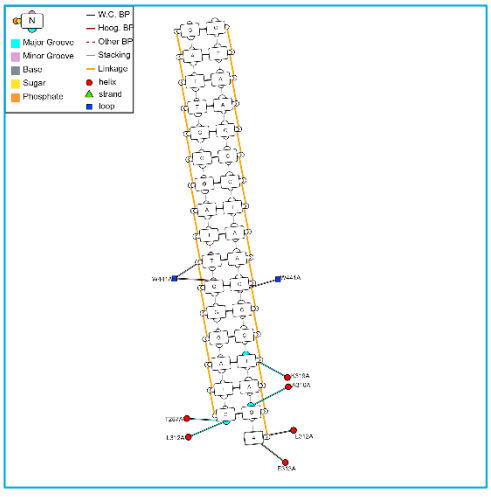
(I)



(J)



(K)



(L)

(M) **(N)**

Figure 11: Graphical representation of molecular docking using DNAproDB. Illustration of docking result of DNA with (A) native, and mutant (B)F15L, (C) M25V, (D)R141Q, (E)Y228C, (F) A252V, (G) G480R, (H) R639H, (I)G645D, (J) R697Q, (K) R700Q, (L) G706R, (M) R729C, and (N) H960R protein are shown above.

Chapter 4

Discussion

The advancement of technology and high throughput sequencing techniques enable us to construct genomic databases over the past few years. These massive, expanding databases are converted into useful information using a variety of bioinformatics tools and techniques (Tenenbaum, 2016; Vamathevan & Birney, 2017). Therefore, the demand for bioinformatics tools and techniques is increasing now. Moreover, in vitro procedures involve a significant number of labor with no assurance of successful results, in this regard cost effective and time-saving computational analyses are preferred prior to work in the laboratory, especially in case of mutagenesis, functional or structural characterization of protein or working with any particular gene of interest in genomic data analysis research. Thus, bioinformatics analysis can help to create in vitro experiments that are specifically targeted, saving time and other resources (Mustafa et al., 2020). The goal of this study was to investigate the functional effect of germ-line SNPs of the RTEL1 gene from available databases. These findings may be useful for future research on the RTEL1 gene, RTEL1-related disorders, and the development of its therapeutics.

The essential helicase, RTEL1 is a member of the superfamily 2 (SF2) helicases, which comprises a RAD3-like DNA helicase, as well as the iron-sulfur (Fe-S) cluster helicase family, that includes xeroderma pigmentosum group D (XPD) and Fanconi anemia complementation group J. (FANCJ) and DEAD/H box DNA helicase (Uringa et al., 2010; Vannier et al., 2014). RTEL1 dissembles recombination intermediates as well as breaks down telomeric loop or T loops and restricts excessive meiotic crossing over. Studies have shown the function of RTEL1 in DNA replication machinery and its association with maintaining the proper DNA replication, stability of replication fork, and maintenance of telomere integrity. Moreover, the role of RTEL1 has been also observed in suppressing the homologous recombination in double-stranded break repair. In addition, according to Vannier et al., RTEL1 may also play a role in telomere protection by breaking up G-quadruplexes, a complex DNA structure that is present at telomeres, and preventing replication. This would prevent telomeres from degradation and the manifestation of a fragile telomere phenotype. The deficiency of RTEL1 in different cell lines has proven the increasing risk of telomere fragility and genomic instability. In humans, mutations in the RTEL1 gene have been proven to cause a genetic rare hereditary disease called Dyskeratosis congenita (DC) and its severe form Hoyeraal–Hreidarsson syndrome (HHS). People with these diseases have shown to have some major clinical signs such as bone marrow failure, accelerated aging, intrauterine growth restriction, developmental defects, microcephaly, cerebellar hypoplasia, immunodeficiency, and

cancer predisposition. Telomeres are excessively short for their ages in every patient with RTEL1 deficiency who has been identified to date, demonstrating the importance of RTEL1 in controlling telomere size in humans(le Guen et al., 2013). Not only the shortening of telomere length, but also an abnormally increasing rate of spontaneous 53BP1 foci (a marker of DNA damage), presence of anaphase bridges, a marker of genomic instability, and telomeric aberrations such as losses of sister telomeres, terminal deletions, etc. were also noted in the cells of three patients with RTEL1 deficiency (LeGuen et al., 2013b). Additionally, an increasing rate of T-circle has been observed in a variety of RTEL1-deficient individuals' cell types (Deng et al., 2013). The discovery of RTEL1 mutations in DC/HH patients highlights the crucial function of RTEL1 in maintaining telomere integrity and genomic stability in humans. It is possible that RTEL1 expression dysregulation or less harmful alterations could contribute to the emergence of malignancies. Moreover, idiopathic pulmonary fibrosis has been linked to heterozygous missense variations in the RTEL1 gene. The role of RTEL1 has also been found in diseases like gliomas, glioblastomas, and breast cancer. Even though RTEL1 mutations and their association with human disorders are commonly identified in a variety of studies, no *in silico* analysis has been performed to date to anticipate harmful nsSNPs of our targeted gene. Therefore, the purpose of this project is to develop a bioinformatics strategy for identifying the most deleterious nsSNPs and their impact on the structure and functionality of the RTEL1 protein.

Our initial classification of nsSNPs was based on how they might affect the structure and functionality of RTEL1 protein. Different bioinformatics tools have different threshold cut-off values for classifying SNPs as damaging or benign, which can occasionally lead to misleading predictions for SNPs with prediction scores that are close to the threshold cut-off value. Therefore, to overcome this limitation, a total of 19 web tools depending on the structural and sequential homology approaches were used to predict functionally and structurally deleterious nsSNPs. It should be noted that for the analysis, we employed the isoform 2 (1219 amino acid) sequence as it is represented as a canonical sequence in the Uniport database. Using ten computational SNP prediction tools—SIFT, PROVEAN, Polyphen-2, PANTHER, SuSPect, PredictSNP, PredictSNP2, P-Mut, SNAP2, and SNP&GO, —we screened out 43 significantly harmful nsSNPs from the 1392 nsSNPs mentioned in the NCBI dbSNP database. Based on the prediction scores produced by these ten web tools, the 43 harmful nsSNPs were chosen. The structural impact of the filtered nsSNPs was analyzed in two categories— mCSM, SDM, Duet, I-Mutant, INPS-MD,

MuPro, and Dynamut was used for the prediction of stability change, whereas Mutpred2 and Project HOPE were utilized for phenotypic effects prediction.

Protein stability, which governs protein conformational shape, determines how well a protein performs its function. Protein misfolding, disintegration or aberrant protein aggregation can occur as a result of any alteration to the stability of the protein (Witham et al., 2011). According to research, amino acid changes that reduce the stability of proteins by a few kcal/mol are accounted for 80% of missense mutations that are linked to diseases (Wang & Moult, 2001). The ΔΔG value that we received as an output from the tools was used to assess the pathogenicity and the consequences of SNPs on the protein's stability. The protein's net free energy balance between its folded and unfolded states is represented by the ΔG value. The folding free energy change, or ΔΔG, separates the mutant from the wild type which measures the effect of mutation on the protein's stability (Zhang et al., 2012). Hence a decline in ΔΔG value implies that the protein is losing its stability. Thus, we concentrated on the effects of the 43 harmful nsSNPs on the stability of the RTEL1 protein. Of these 43 nsSNPs, 14 nsSNPs (F15L, M25V, R141Q, Y228C, A252V, G480R, F559L, R639H, G645D, R697Q, R700Q, G706R, R729C, H960R) were commonly predicted to have negative ΔΔG value by ten web servers, indicating a destabilizing effect on the protein. The phenotypic consequences of these variants were examined through MutPred2 and HOPE where MutPred2 predicted every potential gain, loss, or modification of different molecular properties and HOPE thoroughly examined them. Except for the F559L mutation, all of the mutations were predicted as having a damaging effect on the protein. SNPs with glycine as wild-type residues (G480R, G645D, G706R) are highly conserved due to their small size and less steric hindrance of side chains, a crucial aspect for protein flexibility. The flexibility required for protein function is therefore compromised by its replacement (Parrini et al., 2005). Additionally, conformational flexibility is the primary factor influencing the aggregation tendency of protein, thus any alteration in protein flexibility may increase the likelihood of protein being aggregated and forming fibril (Board et al., 1990; Valerio et al., 2005). Moreover, arginine is a positively charged amino acid, variants where arginine is replaced with amino acids of the neutral or same charge (R141Q, R639H, R697Q, R700Q, R729C) will possibly cause loss of interaction with other molecules, whereas, in case of H960R, it is predicted by HOPE to cause repulsion of ligand or other molecules of similar charges. Apart from these, changes in size and hydrophobicity due to the SNPs may also result in a destabilizing effect on proteins or a potential loss of external

interactions. Because of the disparity in size, M25V is projected to result in an empty space in the core of the protein. Because of the twofold inherent hydrophobicity, cysteine is an essential amino acid for maintaining the stability and structure of proteins. Mutation with the introduction of cysteine has been observed to have a destabilizing effect on membrane protein (Iyer & Mahalakshmi, 2019). HOPE predicted that variants containing mutant cysteine residues (Y228C, R729C) would result in a shift in the hydrophobicity level, which would be detrimental to the protein. This result was also verified through the evaluation of interatomic interactions where all of the mutations have been observed to gain or lose some interactions with nearby atoms due to the substitution of amino acids. The most significant changes were observed in R141Q, Y228C, G480R, R639H, G645D, R700Q, G706R and R729C mutations. Besides, the domain and cluster information of these 13 nsSNPs were identified through Mutation3D. Two domains were identified in the RTEL1 protein where R141Q, Y228C, and A252V mutations are in the Dead_2 domain and R639H, G645D, R697Q, R700Q, G706R, and R729C mutations are in Helicase C_2 domain. Also, a total of 7 mutations (M25V, A252V, G480R, R639H, R697Q, R700Q, G706R) were found to form cluster.

The correct folding of a protein is necessary for the formation of accurate structure which in turn ensures proper functional attributes of the protein. The most advantageous structures from an energetic standpoint are secondary structures, which facilitates further three-dimensional folding, essential for super secondary structures, tertiary structure, motifs, and domains (Khan & Vihinen, 2007). Beta and gamma turns are thought to play a vital role in a significant subset of the secondary structure of proteins and any alteration to these could impair the protein's overall functionality, which, depending on its associated function, could lead to various diseases. In secondary structure analysis, it has been found that all 13 mutations contain fewer beta and gamma turns than the wild type. All of the mutant structures displayed a larger RMSD value when mutant and wild-type structures were superimposed, which justifies the structural deviation resulting from single amino acid substitution in the protein. Additionally, evolutionary conservation of the protein sequence plays an essential role in evaluating the adverse effect of mutation on species. Therefore, using the ConSurf server, first, we identified the evolutionary conservation profile of each amino acid position in the RTEL1 protein, where all of the SNP positions were predicted to be conserved in the protein. For further evaluation, we execute multiple sequence alignments of ten different species using Mega11 software, and the result showed that all 13 positions are homologous in ten

species. The phylogenetic tree also showed the conservation of RTEL1 protein is mostly found in primates, chimpanzees and gorillas are the closest relatives of humans.

The molecular docking investigation of telomeric DNA with native and 13 nsSNPs revealed alterations in binding affinity, which point to a shift in the interaction pattern of the complex. Docking is basically a technique of measuring the optimal orientation of two molecules when they are bonded together to form a stable complex. Usually, the better-orientated the ligand is at the binding pocket of the receptor, the more negative the binding affinity becomes (Kastritis & Bonvin, 2013). Hence, less negative binding affinity demonstrates the change in the binding orientation of the ligand to the receptor molecule, resulting from the substitution of amino acid residues. Out of 13 mutations, 8 mutations —F15L, M25V, Y228C, A252V, G480, R639H, R697Q, and R729C were found to have less negative docking score than the wild-type protein, indicating a less stable binding complex. On the other hand, compared to the complex generated by the wild-type protein, mutations like R141Q, G645D, R700Q, G706R, and H960R revealed a stiffer DNA binding complex with a more negative docking score. Moreover, in the binding pocket, there was a discernible reduction of H-bond and Van der Waals interactions. Interestingly, a remarkable change in the receptor-interacting residues has been observed in F15L, M25V, Y228C, G706R, and R729C mutations, where the DNA was found to bind with a different spectrum of residues than the wild-type. Besides, among these four, three mutations were found to belong to the Dead_2 (Y228C) and Helicase C_2 (G706R, and R729C) domains. Dead_2 domain is a part of RAD3-related DNA binding helicases which are involved in DNA repair, regulation of transcription, and metabolic process of nucleic acid and nucleotide. Whereas the Helicase C_2 domain falls under the C terminal helicase domain which is thought to be necessary for helicase activity (Vannier et al., 2014), and the common phenotypic outcome seen in patients with HHS or DC, particularly short telomeres is predicted to be responsible for the altered activity of C terminal domain (Ballew et al., 2013; Vannier et al., 2014). Therefore, the mutations in these two domains of RTEL1 protein could impose a more deleterious effect. Moreover, according to Mutation 3D analysis, M25V and G706R belonged to the mutation cluster and Kamburov et al. reported that mutations in cancer tissues have a tendency to form clusters in specific positions of protein. Thus, cluster-forming mutations could possibly cause diseases due to the damaging impact on the protein's functionality.

Conclusion

Our study identified 13 significant nsSNPs from the huge SNP database of the RTEL1 gene. These mutations were discovered to have a deleterious impact on the structural and functional properties of the RTEL1 protein which may disrupt the original conformation of the native protein. This extensive study can therefore be very helpful in future research on RTEL1, opening the door to the possibility of looking into potential disease-causing SNPs, and facilitating the identification of potent drugs or pharmacological targets. Hence, to validate these findings, experimental mutational research, genome-wide association studies, and clinical-based studies are further required.

References

Adiba, M., Das, T., Paul, A., Das, A., Chakraborty, S., Hosen, M. I., & Nabi, A. H. M. N. (2021). In silico characterization of coding and non-coding SNPs of the androgen receptor gene. *Informatics in Medicine Unlocked*, *24*, 100556. https://doi.org/10.1016/J.IMU.2021.100556

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*(4), 248. https://doi.org/10.1038/NMETH0410-248

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, *44*(W1), W344–W350. https://doi.org/10.1093/NAR/GKW408

Azad, A. K., Sadee, W., & Schlesinger, L. S. (2012). Innate Immune Gene Polymorphisms in Tuberculosis. *Infection and Immunity*, *80*(10), 3343. https://doi.org/10.1128/IAI.00443-12

Bai, C., Connolly, B., Metzker, M. L., Hilliard, C. A., Liu, X., Sandig, V., Soderman, A., Galloway, S. M., Liu, Q., Austin, C. P., & Caskey, C. T. (2000). Overexpression of M68/DcR3 in human gastrointestinal tract tumors independent of gene amplification and its location in a four-gene cluster. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(3), 1230–1235. https://doi.org/10.1073/PNAS.97.3.1230/ASSET/8832FF8A-12BD-4309-97DD-25963C128FAE/ASSETS/GRAPHIC/PQ0205123006.JPEG

Ballew, B. J., Yeager, M., Jacobs, K., Giri, N., Boland, J., Burdett, L., Alter, B. P., & Savage, S. A. (2013). Germline Mutations of Regulator of Telomere Elongation Helicase 1, RTEL1, In Dyskeratosis Congenita. *Human Genetics*, *132*(4), 473. https://doi.org/10.1007/S00439-013-1265-8

Barber, L. J., Youds, J. L., Ward, J. D., McIlwraith, M. J., O'Neil, N. J., Petalcorin, M. I. R., Martin, J. S., Collis, S. J., Cantor, S. B., Auclair, M., Tissenbaum, H., West, S. C., Rose, A. M., & Boulton, S. J. (2008). RTEL1 Maintains Genomic Stability by Suppressing Homologous Recombination. *Cell*, *135*(2), 261–271. https://doi.org/10.1016/J.CELL.2008.08.016

Barroso, I., Gurnell, M., Crowley, V. E. F., Agostini, M., Schwabe, J. W., Soos, M. A., Maslen, G. L., Williams, T. D. M., Lewis, H., Schafer, A. J., Chatterjee, V. K. K., & O'Rahilly, S. (1999). Dominant negative mutations in human PPARγ associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature 1999 402:6764*, *402*(6764), 880–883. https://doi.org/10.1038/47254

Bee, C., Moshnikova, A., Mellor, C. D., Molloy, J. E., Koryakina, Y., Stieglitz, B., Khokhlatchev, A., & Herrmann, C. (2010). Growth and tumor suppressor NORE1A is a regulatory node between Ras signaling and microtubule nucleation. *The Journal of Biological Chemistry*, *285*(21), 16258–16266. https://doi.org/10.1074/JBC.M109.081562

Begovich, A. B., Carlton, V. E. H., Honigberg, L. A., Schrodi, S. J., Chokkalingam, A. P., Alexander, H. C., Ardlie, K. G., Huang, Q., Smith, A. M., Spoerke, J. M., Conn, M. T., Chang, M., Chang, S. Y. P., Saiki, R. K., Catanese, J. J., Leong, D. U., Garcia, V. E., McAllister, L. B., Jeffery, D. A., … Gregersen, P. K. (2004). A Missense Single-Nucleotide Polymorphism in a Gene Encoding a Protein Tyrosine Phosphatase (PTPN22) Is Associated with Rheumatoid Arthritis. *American Journal of Human Genetics*, *75*(2), 330. https://doi.org/10.1086/422827

Bendl, J., Musil, M., Štourač, J., Zendulka, J., Damborský, J., & Brezovský, J. (2016). PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions. *PLoS Computational Biology*, *12*(5). https://doi.org/10.1371/JOURNAL.PCBI.1004962

Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E. D., Zendulka, J., Brezovsky, J., & Damborsky, J. (2014). PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. *PLOS Computational Biology*, *10*(1), e1003440. https://doi.org/10.1371/JOURNAL.PCBI.1003440

Berezin, C., Glaser, F., Rosenberg, J., Paz, I., Pupko, T., Fariselli, P., Casadio, R., & Ben-Tal, N. (2004). ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics*, *20*(8), 1322–1324. https://doi.org/10.1093/BIOINFORMATICS/BTH070

Board, P. G., Pierce, K., & Coggan, M. (1990). Expression of functional coagulation factor XIII in Escherichia coli. *Thrombosis and Haemostasis*, *63*(2), 235–240. https://doi.org/10.1055/S-0038-1645201/ID/JR_15

Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L., & Casadio, R. (2009). Functional annotations improve the predictive score of human disease-related mutations in proteins. *Human Mutation*, *30*(8), 1237–1244. https://doi.org/10.1002/HUMU.21047

Capriotti, E., & Altman, R. B. (2011). Improving the prediction of disease-related variants using protein three-dimensional structure. *BMC Bioinformatics*, *12 Suppl 4*(Suppl 4). https://doi.org/10.1186/1471-2105-12-S4-S3

Chasman, D., & Adams, R. M. (2001). Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. *Journal of Molecular Biology*, *307*(2), 683–706. https://doi.org/10.1006/JMBI.2001.4510

Cheng, J., Randall, A., & Baldi, P. (2005). *Prediction of Protein Stability Changes for Single-Site Mutations Using Support Vector Machines*. https://doi.org/10.1002/prot.20810

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. https://doi.org/10.1093/BIOINFORMATICS/BTV195

Collins, F. S., Brooks, L. D., & Chakravarti, A. (1998). A DNA polymorphism discovery resource for research on human genetic variation. *Genome Research*, *8*(12), 1229–1231. https://doi.org/10.1101/GR.8.12.1229

de Lange, T. (2004). T-loops and the origin of telomeres. *Nature Reviews Molecular Cell Biology 2004 5:4*, *5*(4), 323–329. https://doi.org/10.1038/nrm1359

de Lange, T. (2009). How telomeres solve the end-protection problem. *Science (New York, N.Y.)*, *326*(5955), 948–952. https://doi.org/10.1126/SCIENCE.1170633

Deng, Z., Glousker, G., Molczan, A., Fox, A. J., Lamm, N., Dheekollu, J., Weizman, O. el, Schertzer, M., Wang, Z., Vladimirova, O., Schug, J., Aker, M., Londoño-Vallejo, A., Kaestner, K. H., Lieberman, P. M., & Tzfati, Y. (2013). Inherited mutations in the helicase RTEL1 cause telomere dysfunction and Hoyeraal-Hreidarsson syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(36), E3408–E3416. https://doi.org/10.1073/PNAS.1300600110/SUPPL_FILE/PNAS.201300600SI.PDF

Ding, H., Schertzer, M., Wu, X., Gertsenstein, M., Selig, S., Kammori, M., Pourvali, R., Poon, S., Vulto, I., Chavez, E., Tam, P. P. L., Nagy, A., & Lansdorp, P. M. (2004). Regulation of murine telomere length by Rtel: An essential gene encoding a helicase-like protein. *Cell*, *117*(7), 873–886. https://doi.org/10.1016/j.cell.2004.05.026

Doniger, S. W., Kim, H. S., Swain, D., Corcuera, D., Williams, M., Yang, S. P., & Fay, J. C. (2008). A Catalog of Neutral and Deleterious Polymorphism in Yeast. *PLoS Genetics*, *4*(8), 1000183. https://doi.org/10.1371/JOURNAL.PGEN.1000183

Egan, K. M., Thompson, R. C., Nabors, L. B., Olson, J. J., Brat, D. J., Larocca, R. v., Brem, S., Moots, P. L., Madden, M. H., Browning, J. E., & Ann Chen, Y. (2011). Cancer susceptibility variants and the risk of adult glioma in a US case-control study. *Journal of Neuro-Oncology*, *104*(2), 535–542. https://doi.org/10.1007/S11060-010-0506-0

Forsberg, L., de Faire, U., Marklund, S. L., Andersson, P. M., Stegmayr, B., & Morgenstern, R. (2000). Phenotype Determination of a Common Pro-Leu Polymorphism in Human Glutathione Peroxidase 1. *Blood Cells, Molecules, and Diseases*, *26*(5), 423–426. https://doi.org/10.1006/BCMD.2000.0325

Frizzell, A., Nguyen, J. H. G., Petalcorin, M. I. R., Turner, K. D., Boulton, S. J., Freudenreich, C. H., & Lahue, R. S. (2014). RTEL1 inhibits trinucleotide repeat expansions and fragility. *Cell Reports*, *6*(5), 827–835. https://doi.org/10.1016/J.CELREP.2014.01.034

Gilson, E., & Géli, V. (2007). How telomeres are replicated. *Nature Reviews. Molecular Cell Biology*, *8*(10), 825–838. https://doi.org/10.1038/NRM2259

Glousker, G., Touzot, F., Revy, P., Tzfati, Y., & Savage, S. A. (2015). Unraveling the pathogenesis of Hoyeraal–Hreidarsson syndrome, a complex telomere biology disorder. *British Journal of Haematology*, *170*(4), 457–471. https://doi.org/10.1111/BJH.13442

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*, *5*, 162–172. https://doi.org/10.1016/J.MGENE.2015.07.007

Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H., & de Lange, T. (1999a). Mammalian telomeres end in a large duplex loop. *Cell*, *97*(4), 503–514. https://doi.org/10.1016/S0092-8674(00)80760-6

Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H., & de Lange, T. (1999b). Mammalian Telomeres End in a Large Duplex Loop. *Cell*, *97*(4), 503–514. https://doi.org/10.1016/S0092-8674(00)80760-6

Hassani, M. A., Murid, J., & Yan, J. (2023). Regulator of telomere elongation helicase 1 gene and its association with malignancy. *Cancer Reports*, *6*(1), e1735. https://doi.org/10.1002/CNR2.1735

Hecht, M., Bromberg, Y., & Rost, B. (2015). Better prediction of functional effects for sequence variants. *BMC Genomics*, *16 Suppl 8*(Suppl 8). https://doi.org/10.1186/1471-2164-16-S8-S1

Iyer, B. R., & Mahalakshmi, R. (2019). Hydrophobic Characteristic Is Energetically Preferred for Cysteine in a Model Membrane Protein. *Biophysical Journal*, *117*(1), 25–35. https://doi.org/10.1016/j.bpj.2019.05.024

Jain, D., & Cooper, J. P. (2010). Telomeric Strategies: Means to an End. *Https://Doi.Org/10.1146/Annurev-Genet-102108-134841*, *44*, 243–269. https://doi.org/10.1146/ANNUREV-GENET-102108-134841

Kamburov, A., Lawrence, M. S., Polak, P., Leshchiner, I., Lage, K., Golub, T. R., Lander, E. S., & Getz, G. (2015). Comprehensive assessment of cancer missense mutation clustering in protein structures. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(40), E5486–E5495. https://doi.org/10.1073/PNAS.1516373112/SUPPL_FILE/PNAS.1516373112.SD01.XLSX

Kastritis, P. L., & Bonvin, A. M. J. J. (2013). On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *Journal of The Royal Society Interface*, *10*(79). https://doi.org/10.1098/RSIF.2012.0835

Khan, S., & Vihinen, M. (2007). Spectrum of disease-causing mutations in protein secondary structures. *BMC Structural Biology*, *7*, 56. https://doi.org/10.1186/1472-6807-7-56

Klausen, M. S., Jespersen, M. C., Nielsen, H., Jensen, K. K., Jurtz, V. I., Sønderby, C. K., Sommer, M. O. A., Winther, O., Nielsen, M., Petersen, B., & Marcatili, P. (2019). NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. *Proteins:*

Structure, Function, and Bioinformatics, *87*(6), 520–527. https://doi.org/10.1002/PROT.25674

Kucukkal, T. G., Petukh, M., Li, L., & Alexov, E. (2015). Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Current Opinion in Structural Biology*, *32*, 18–24. https://doi.org/10.1016/J.SBI.2015.01.003

Kumar, N., Taneja, A., Ghosh, M., Rothweiler, U., Sundaresan, N. R., & Singh, M. (2022). Harmonin homology domain-mediated interaction of RTEL1 helicase with RPA and DNA provides mechanistic insight into its role in DNA repair. *BioRxiv*, 2022.08.08.503141. https://doi.org/10.1101/2022.08.08.503141

Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Science : A Publication of the Protein Society*, *27*(1), 129–134. https://doi.org/10.1002/PRO.3289

le Guen, T., Jullien, L., Schertzer, M., Lefebvre, A., Kermasson, L., de Villartay, J. P., Londoño-Vallejo, A., & Revy, P. (2013). RTEL1, une hélicase de l'ADN essentielle à la stabilité du génome. *Médecine/Sciences*, *29*(12), 1138–1144. https://doi.org/10.1051/MEDSCI/20132912018

Lee, J. W. (2013). Telomere shortening by mutations in the RTEL1 helicase cause severe form of dyskeratosis congenita, Hoyerall-Hreidarsson syndrome. *Clinical Genetics*, *84*(3), 210–210. https://doi.org/10.1111/CGE.12175

LeGuen, T., Jullien, L., Touzot, F., Schertzer, M., Gaillard, L., Perderiset, M., Carpentier, W., Nitschke, P., Picard, C., Couillault, G., Soulier, J., Fischer, A., Callebaut, I., Jabado, N., Londono-Vallejo, A., de Villartay, J. P., & Revy, P. (2013a). Human RTEL1 deficiency causes Hoyeraal-Hreidarsson syndrome with short telomeres and genome instability. *Human Molecular Genetics*, *22*(16), 3239–3249. https://doi.org/10.1093/HMG/DDT178

LeGuen, T., Jullien, L., Touzot, F., Schertzer, M., Gaillard, L., Perderiset, M., Carpentier, W., Nitschke, P., Picard, C., Couillault, G., Soulier, J., Fischer, A., Callebaut, I., Jabado, N., Londono-Vallejo, A., de Villartay, J. P., & Revy, P. (2013b). Human RTEL1 deficiency causes Hoyeraal–Hreidarsson syndrome with short telomeres and genome instability. *Human Molecular Genetics*, *22*(16), 3239–3249. https://doi.org/10.1093/HMG/DDT178

Lin, W. Y., Fordham, S. E., Hungate, E., Sunter, N. J., Elstob, C., Xu, Y., Park, C., Quante, A., Strauch, K., Gieger, C., Skol, A., Rahman, T., Sucheston-Campbell, L., Wang, J., Hahn, T., Clay-Gilmour, A. I., Jones, G. L., Marr, H. J., Jackson, G. H., … Allan, J. M. (2021). Genome-wide association study identifies susceptibility loci for acute myeloid leukemia. *Nature Communications*, *12*(1). https://doi.org/10.1038/S41467-021-26551-X

Liu, Y., Shete, S., Etzel, C. J., Scheurer, M., Alexiou, G., Armstrong, G., Tsavachidis, S., Liang, F. W., Gilbert, M., Aldape, K., Armstrong, T., Houlston, R., Hosking, F., Robertson, L., Xiao, Y., Wiencke, J., Wrensch, M., Andersson, U., Melin, B. S., & Bondy, M. (2010). Polymorphisms of LIG4, BTBD2, HMGA2, and RTEL1 Genes Involved in the Double-

Strand Break Repair Pathway Predict Glioblastoma Survival. *Journal of Clinical Oncology*, *28*(14), 2467. https://doi.org/10.1200/JCO.2009.26.6213

López-Ferrando, V., Gazzo, A., de La Cruz, X., Orozco, M., & Gelpí, J. L. (2017). PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. *Nucleic Acids Research*, *45*(W1), W222–W228. https://doi.org/10.1093/NAR/GKX313

Mansur, Y. A., Rojano, E., Ranea, J. A. G., & Perkins, J. R. (2018). Analyzing the Effects of Genetic Variation in Noncoding Genomic Regions. *Precision Medicine: Tools and Quantitative Approaches*, 119–144. https://doi.org/10.1016/B978-0-12-805364-5.00007-X

Mayrose, I., Graur, D., Ben-Tal, N., & Pupko, T. (2004). Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior. *Molecular Biology and Evolution*, *21*(9), 1781–1791. https://doi.org/10.1093/MOLBEV/MSH194

Meyer, M. J., Lapcevic, R., Romero, A. E., Yoon, M., Das, J., Beltrán, J. F., Mort, M., Stenson, P. D., Cooper, D. N., Paccanaro, A., & Yu, H. (2016). mutation3D: Cancer Gene Prediction Through Atomic Clustering of Coding Variants in the Structural Proteome. *Human Mutation*, *37*(5), 447–456. https://doi.org/10.1002/HUMU.22963

Mondal, A., Paul, D., Dastidar, S. G., Saha, T., & Goswami, A. M. (2022). In silico analyses of Wnt1 nsSNPs reveal structurally destabilizing variants, altered interactions with Frizzled receptors and its deregulation in tumorigenesis. *Scientific Reports 2022 12:1*, *12*(1), 1–18. https://doi.org/10.1038/s41598-022-19299-x

Moore, R. M., Harrison, A. O., McAllister, S. M., Polson, S. W., & Eric Wommack, K. (2020). Iroki: Automatic customization and visualization of phylogenetic trees. *PeerJ*, *8*, e8584. https://doi.org/10.7717/PEERJ.8584/SUPP-4

Muleris, M., Almeida, A., Gerbault-Seureau, M., Malfoy, B., & Dutrillaux, B. (1995). Identification of amplified DNA sequences in breast cancer and their organization within homogeneously staining regions. *Genes, Chromosomes and Cancer*, *14*(3), 155–163. https://doi.org/10.1002/GCC.2870140302

Mustafa, M. I., Murshed, N. S., Abdelmoneim, A. H., Abdelmageed, M. I., Elfadol, N. M., & Makhawi, A. M. (2020). Extensive in Silico Analysis of ATL1 Gene: Discovered Five Mutations That May Cause Hereditary Spastic Paraplegia Type 3A. *Scientifica*, *2020*. https://doi.org/10.1155/2020/8329286

O'Sullivan, R. J., & Karlseder, J. (2010). Telomeres: protecting chromosomes against genome instability. *Nature Reviews. Molecular Cell Biology*, *11*(3), 171. https://doi.org/10.1038/NRM2848

Parrini, C., Taddei, N., Ramazzotti, M., Degl'Innocenti, D., Ramponi, G., Dobson, C. M., & Chiti, F. (2005). Glycine residues appear to be evolutionarily conserved for their ability to inhibit aggregation. *Structure*, *13*(8), 1143–1151. https://doi.org/10.1016/j.str.2005.04.022

Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K. A., Lin, G. N., Nam, H. J., Mort, M., Cooper, D. N., Sebat, J., Iakoucheva, L. M., Mooney, S. D., & Radivojac, P. (2020). Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nature Communications 2020 11:1*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-19669-x

Petukh, M., Kucukkal, T. G., & Alexov, E. (2015). On human disease-causing amino acid variants: statistical study of sequence and structural patterns. *Human Mutation*, *36*(5), 524–534. https://doi.org/10.1002/HUMU.22770

Pires, D. E. V., Ascher, D. B., & Blundell, T. L. (2014). DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Research*, *42*(W1), W314–W319. https://doi.org/10.1093/NAR/GKU411

Pupko, T., Bell, R. E., Mayrose, I., Glaser, F., & Ben-Tal, N. (2002). Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics (Oxford, England)*, *18 Suppl 1*(SUPPL. 1). https://doi.org/10.1093/BIOINFORMATICS/18.SUPPL_1.S71

Radivojac, P., Vacic, V., Haynes, C., Cocklin, R. R., Mohan, A., Heyen, J. W., Goebl, M. G., & Iakoucheva, L. M. (2010). Identification, Analysis and Prediction of Protein Ubiquitination Sites. *Proteins*, *78*(2), 365. https://doi.org/10.1002/PROT.22555

Rajendran, V., Gopalakrishnan, C., & Sethumadhavan, R. (2018). Pathological role of a point mutation (T315I) in BCR-ABL1 protein-A computational insight. *Journal of Cellular Biochemistry*, *119*(1), 918–925. https://doi.org/10.1002/JCB.26257

Sagendorf, J. M., Berman, H. M., & Rohs, R. (2017). DNAproDB: an interactive tool for structural analysis of DNA–protein complexes. *Nucleic Acids Research*, *45*(W1), W89–W97. https://doi.org/10.1093/NAR/GKX272

Sagendorf, J. M., Markarian, N., Berman, H. M., & Rohs, R. (2020). DNAproDB: an expanded database and web-based tool for structural analysis of DNA–protein complexes. *Nucleic Acids Research*, *48*(D1), D277–D287. https://doi.org/10.1093/NAR/GKZ889

Sarek, G., Vannier, J. B., Panier, S., Petrini, H. J., & Boulton, S. J. (2015). TRF2 recruits RTEL1 to telomeres in S phase to promote t-loop unwinding. *Molecular Cell*, *57*(4), 622–635. https://doi.org/10.1016/J.MOLCEL.2014.12.024

Savojardo, C., Fariselli, P., Martelli, P. L., & Casadio, R. (2016). INPS-MD: a web server to predict stability of protein variants from sequence and structure. *Bioinformatics (Oxford, England)*, *32*(16), 2542–2544. https://doi.org/10.1093/BIOINFORMATICS/BTW192

Sfeir, A., Kosiyatrakul, S. T., Hockemeyer, D., MacRae, S. L., Karlseder, J., Schildkraut, C. L., & de Lange, T. (2009). Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell*, *138*(1), 90–103. https://doi.org/10.1016/J.CELL.2009.06.021

Sim, N. L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, *40*(W1), W452–W457. https://doi.org/10.1093/NAR/GKS539

Sobieszczyk, M. E., Lingappa, J. R., & McElrath, M. J. (2011). Host genetic polymorphisms associated with innate immune factors and HIV-1. *Current Opinion in HIV and AIDS*, *6*(5), 427–434. https://doi.org/10.1097/COH.0B013E3283497155

Takedachi, A., Despras, E., Scaglione, S., Guérois, R., Guervilly, J. H., Blin, M., Audebert, S., Camoin, L., Hasanova, Z., Schertzer, M., Guille, A., Churikov, D., Callebaut, I., Naim, V., Chaffanet, M., Borg, J. P., Bertucci, F., Revy, P., Birnbaum, D., … Gaillard, P. H. L. (2020). SLX4 interacts with RTEL1 to prevent transcription-mediated DNA replication perturbations. *Nature Structural & Molecular Biology 2020 27:5*, *27*(5), 438–449. https://doi.org/10.1038/s41594-020-0419-3

Tamura, K., Stecher, G., & Kumar, S. (n.d.). *MEGA11: Molecular Evolutionary Genetics Analysis Version 11*. https://doi.org/10.1093/molbev/msab120

Tang, H., & Thomas, P. D. (2016). PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics (Oxford, England)*, *32*(14), 2230–2232. https://doi.org/10.1093/BIOINFORMATICS/BTW222

Tenenbaum, J. D. (2016). Translational Bioinformatics: Past, Present, and Future. *Genomics, Proteomics & Bioinformatics*, *14*(1), 31–41. https://doi.org/10.1016/J.GPB.2016.01.003

Uringa, E. J., Youds, J. L., Lisaingo, K., Lansdorp, P. M., & Boulton, S. J. (2010). RTEL1: an essential helicase for telomere maintenance and the regulation of homologous recombination. *Nucleic Acids Research*, *39*(5), 1647–1655. https://doi.org/10.1093/NAR/GKQ1045

Valerio, M., Colosimo, A., Conti, F., Giuliani, A., Grottesi, A., Manetti, C., & Zbilut, J. P. (2005). Early events in protein aggregation: Molecular flexibility and hydrophobicity/charge interaction in amyloid peptides as studied by molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics*, *58*(1), 110–118. https://doi.org/10.1002/PROT.20306

Vamathevan, J., & Birney, E. (2017). A Review of Recent Advances in Translational Bioinformatics: Bridges from Biology to Medicine. *Yearbook of Medical Informatics*, *26*(1), 178–187. https://doi.org/10.15265/IY-2017-017/ID/JR017-55

Vannier, J. B., Pavicic-Kaltenbrunner, V., Petalcorin, M. I. R., Ding, H., & Boulton, S. J. (2012). RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell*, *149*(4), 795–806. https://doi.org/10.1016/J.CELL.2012.03.030

Vannier, J. B., Sarek, G., & Boulton, S. J. (2014). RTEL1: Functions of a disease-associated helicase. *Trends in Cell Biology*, *24*(7), 416–425. https://doi.org/10.1016/j.tcb.2014.01.004

Venselaar, H., te Beek, T. A. H., Kuipers, R. K. P., Hekkelman, M. L., & Vriend, G. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach

with life scientist friendly interfaces. *BMC Bioinformatics*, *11*. https://doi.org/10.1186/1471-2105-11-548

Wang, D., Liu, D., Yuchi, J., He, F., Jiang, Y., Cai, S., Li, J., & Xu, D. (2020). MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. *Nucleic Acids Research*, *48*(W1), W140. https://doi.org/10.1093/NAR/GKAA275

Wang, R. C., Smogorzewska, A., & de Lange, T. (2004). Homologous Recombination Generates T-Loop-Sized Deletions at Human Telomeres. *Cell*, *119*(3), 355–368. https://doi.org/10.1016/J.CELL.2004.10.011

Wang, Z., & Moult, J. (2001). SNPs, protein structure, and disease. *Human Mutation*, *17*(4), 263–270. https://doi.org/10.1002/HUMU.22

Webb, B., & Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, *54*, 5.6.1. https://doi.org/10.1002/CPBI.3

Witham, S., Takano, K., Schwartz, C., & Alexov, E. (2011). A missense mutation in CLIC2 associated with intellectual disability is predicted by in silico modeling to affect protein stability and dynamics. *Proteins*, *79*(8), 2444–2454. https://doi.org/10.1002/PROT.23065

Wrensch, M., Jenkins, R. B., Chang, J. S., Yeh, R. F., Xiao, Y., Decker, P. A., Ballman, K. v., Berger, M., Buckner, J. C., Chang, S., Giannini, C., Halder, C., Kollmeyer, T. M., Kosel, M. L., Lachance, D. H., McCoy, L., O'Neill, B. P., Patoka, J., Pico, A. R., … Wiencke, J. K. (2009). Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. *Nature Genetics 2009 41:8*, *41*(8), 905–908. https://doi.org/10.1038/ng.408

Wu, W., Bhowmick, R., Vogel, I., Özer, Ö., Ghisays, F., Thakur, R. S., Sanchez de Leon, E., Richter, P. H., Ren, L., Petrini, J. H., Hickson, I. D., & Liu, Y. (2020). RTEL1 suppresses G-quadruplex-associated R-loops at difficult-to-replicate loci in the human genome. *Nature Structural & Molecular Biology 2020 27:5*, *27*(5), 424–437. https://doi.org/10.1038/s41594-020-0408-6

Yan, Y., Tao, H., He, J., & Huang, S. Y. (2020). The HDOCK server for integrated protein-protein docking. *Nature Protocols*, *15*(5), 1829–1852. https://doi.org/10.1038/S41596-020-0312-X

Yan, Y., Zhang, D., Zhou, P., Li, B., & Huang, S. Y. (2017). HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Research*, *45*(W1), W365–W373. https://doi.org/10.1093/NAR/GKX407

Yates, C. M., Filippis, I., Kelley, L. A., & Sternberg, M. J. E. (2014). SuSPect: Enhanced prediction of single amino acid variant (SAV) phenotype using network features. *Journal of Molecular Biology*, *426*(14), 2692–2701. https://doi.org/10.1016/J.JMB.2014.04.026

Zhang, Z., Miteva, M. A., Wang, L., & Alexov, E. (2012). Analyzing effects of naturally occurring missense mutations. *Computational and Mathematical Methods in Medicine, 2012*. https://doi.org/10.1155/2012/805827