Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/jptcp.v31i5.7127

MUTATION SPECTRUM ANALYSIS OF *ITGA3* GENE ASSOCIATED WITH NEPHROTIC SYNDROME

Quratul Ain Nazir[1], Wasim Shehzad[2], Muhammad Asad Ali[3], Muhammad YasirZahoor[4*]

[1]Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences Lahore, Pakistan. Email address: qurat.ain.nazir@gmail.com
[2]Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences Lahore, Pakistan. Email address: wasim.shehzad@uvas.edu.pk
[3]Department of Microbiology, University of Veterinary and Animal Sciences Lahore, Pakistan. Email address: asad.ali@uvas.edu.pk
[4*]Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences Lahore, Pakistan. Email address: yasir.zahoor@uvas.edu.pk

***Corresponding author:** Muhammad Yasir Zahoor
*Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences Lahore, Pakistan. Email address: yasir.zahoor@uvas.edu.pk

Abstract:
Nephrotic syndrome is a renal disorder that affects the Glomerular Filtration Barrier (GFB). ITGA3 protein may have a pivotal function in the intricate interaction between cells, morphogens, and the extracellular matrix (ECM) that leads to the development of the kidneys (nephrogenesis). This study involves a detailed analysis of the missense mutations in the ITGA3 gene through various in silico and bioinformatics tools. Conservation score of each mutation was measured to determine the level of conservation and hence severity of the mutations with regards to the protein structure and function. 3D structural formation, phylogenetic analysis and RNA expression profiles were also constructed. Three highly deleterious mutations (G125R, R143H, P680A) were identified through these in silico tools. Molecular dynamics simulations (MDS) analysis was performed for understanding the dynamic behavior of wild type and mutant proteins. This data aims to facilitate future studies on ITGA3 protein and its role in the development of the Nephrotic syndrome, along with the implication of the mutations on the structure and function of the ITGA3 protein. This study also gives an insight on the detrimental effect on the function of the kidneys due to these mutations that ultimately results in the development of nephrotic syndrome.

Keywords: Nephrotic syndrome, ITGA, Nephrogenesis, in silico, Bioinformatics

Background

Nephrotic syndrome is a renal disorder that results in protein leakage in the urine. This causes the occurrence of symptoms such as hypo-tension, hypercoagulation and life-threatening infections. The glomerular filtration barrier (GFB) is responsible for the filtration of the blood in kidneys [1]. Damage to this barrier may result in many renal disorders including nephrotic syndrome. This may be caused by immune complex deposition, phospholipase antibody production, or the development of alloantibodies [2]. Mutations in numerous podocyte proteins may target the functioning of the podocyte via various pathological mechanisms by modifying the slit diaphragm structure,

perturbing the delicate podocyte cytoskeleton, destroying cell-matrix connections, or obstructing critical signaling pathways [3].

Manifestational operational changes of particular integrins within the glomerulus can play a role in the impairment of podocytes and the disruption of the glomerular barrier. Mutations or dysregulation of *ITGA3, ITGA4, ITGA5, ITGA6,* and *ITGA8* have been linked to different types of glomerular diseases that are associated with nephrotic syndrome. The *ITGA3* gene has been linked to a condition known as generalised junctional epidermolysis bullosa with respiratory and renal involvement (JEB-RR), as well as congenital interstitial lung disease, nephrotic syndrome, and epidermolysis bullosa [4].

The genes belonging to the integrin α (*ITGA*) subfamily are known to have a significant impact on the development and progression of different types of disorders including cancer. Integrins are classified as heterodimeric surface receptors, comprising of non-covalently linked α and β subunits. Currently, our understanding suggests that the integrin family encompasses a total of 18α and 8β members [5]. Numerous studies have demonstrated that integrins possess the ability to act as signalling molecules across the cell membrane in both directions. This includes "inside-out signalling," which occurs when extracellular stimulation prompts the binding of intracellular linin and kindlin to the cytoskeleton, resulting in the extracellular domain adopting a high affinity state. Additionally, there is "outside-in signalling," a complex process whereby the heterodimeric adhesion receptors of integrins facilitate cell adhesion to the extracellular matrix (ECM) and subsequently activate integrins to interact with the cytoskeleton. This activation leads to the initiation of various intracellular signalling pathways, which in turn enhance the binding of activated integrin ligands and enable the perception of the intracellular environment [6]. The expression of Integrin α3β1 is prevalent in the epithelial tissues of the skin, lungs, and kidneys [7].

While integrins play a crucial role as receptors for extracellular matrix (ECM) proteins and are widely present throughout kidney development. *ITGA3* is predominantly recognized as a passive stabilizer of glioblastoma multiforme (GBM) rather than an active participant in nephrogenesis. In recent studies, it has been demonstrated that genetic alterations occurring within the human *ITGA3* gene are responsible for the development of a complex condition known as NEP syndrome (Nephrotic syndrome, Epidermolysis bullosa, and Pulmonary disease).

The deficiency of ITGA3 manifests as an early developmental defect characterized by the dysregulation of multiple crucial pathways involved in nephrogenesis. This dysregulation ultimately gives rise to the renal hypodysplasia/CAKUT phenotype. The transcripts that were found to be down regulated in the kidney, specifically in the differentiated proximal tubules, include genes such as *GATM, AGXT2, GSTA1, SLC3A1, SLC13A1, SLC7A9, SLC2A2*, and *AQP11* [8].

Several approaches have been employed previously to explore the etiology of these genes and their impacts on the disease predisposition. These include in vitro and in vivo studies but these methods frequently need numerous technical and financial resources, manpower and are also quite time-consuming. Numerous insilico techniques can be used for the prediction of different SNPs and their effects on the functions and structure of proteins [9].

The present study involves a detailed analysis of the reported missense mutations in the *ITGA3* gene through various in silico and bioinformatics tools. The severity of the mutations was analyzed along with the conservation score of each mutation to determine which mutations fall on the conserved regions and are consequently more detrimental to the protein structure and function. The 3D structure of the mutant forms of the protein was also investigated to determine the changes in the structure due to these mutations. A phylogenetic analysis was performed to determine the relationship between different *ITGA* proteins. Molecular dynamics simulations were performed for the most deleterious mutations to identify the dynamic behavior of the mutations. The present study demonstrates the potential of using computational methods in predicting the effect of deleterious SNPs on protein structure.

Methodology:

Data collection and retrieval of nsSNPs

The data about the reported missense mutations (nsSNPs) in ITGA3 gene was obtained from Human Gene Mutation Database (HGMD) [10] on July 22, 2023. The protein sequence of ITGA3 was obtained through UniProt [11] Database (UniProt ID: P26006). The protein structure was predicted using the Phyre2 tool, homology modeling technique was used.

Analysis of mutations

The functional effects of the missense SNPs on the protein were predicted using the nine sequence-based servers and tools. The effect of the amino acid substitution on the structure and function of protein was predicted by using PANTHER (Protein Analysis Through Evolutionary Relationships) [12] and PolyPhen-2 (Polymorphism Phenotyping v2) [13]. Consensus methods such as Meta-SNP [14] and PredictSNP [15] were also used. SIFT (Sorting Intolerant From Tolerant) [16], MAPP (Multivariate Analysis of Protein Polymorphism) [17], PMut [18], SNAP (Synonymous Non-synonymous Analysis Program) [19], PhD-SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms) [20] and SNPs&GO [21] were also used for predicting the pathogenicity and disease-association of the missense SNPs in silico.

The genetic tolerance of these missense SNPs was analyzed through Metadome [22], which is a web server designed to visually analyze genetic tolerance through lens of protein domains.

Mutation Protein Stability Prediction

The stability changes induced on the protein due to the missense SNPs were analyzed using five structure-based tools. The impact of the missense SNPs on the protein stability and dynamics was predicted using I-Mutant 2.0 [23], INPS-MD (Impact of Non-synonymous mutations on Protein Structure- Multi Dimension) [24], MUpro (Multi-objective Optimization Approach for Protein Stability Prediction) [25], CUPSAT (Cologne University Protein Stability Analysis Tool) [26], and DUET [27].

Determination of conservation sites

The evolutionary conservation profile of the *ITGA3* gene was predicted using the online tools ConSurf [28] and PDBsum [29]. Also, the multiple sequence alignment (MSA) was performed using the Clustal Omega, MAFFT, T-Coffee, and MUSCLE [30] for ITGA sequences of different species that includes Homo sapiens, Mus musculus, Cricetulus griseus, and Bos taurus having the Uniprot IDs P26006, Q62470, P17852, F1MMS9 respectively. The visualization of the sequence was carried out using the tool WebLogo [31].

Prediction of conserved domains in ITGA3

The domains of *ITGA3* were determined using SMART (Simple Modular Architecture Research Tool) [32]. It is a web-based resource dedicated to the identification and annotation of protein domains, mainly focusing on signaling domains.

Prediction of secondary structure and transmembrane protein helix

The current study used the tool SOPMA (Self-Optimized Prediction Method with Alignment) [33] to analyze the secondary structure of the protein. The SOPMA analysis methods are based on the homologue method – it takes into account information from an alignment of sequences belonging to the same family. It has nearly 70% accuracy for a three-state description [33].

3D Protein Structural Analysis

Phyre-2 (Protein Homology/AnalogY Recognition Engine) [34] was used to generate the 3D structure of the proteins. PyMOL [35] molecular graphics software was used for examining the

models produced by Phyre-2 and to determine the changes in the protein structure due to these mutations.

Prediction of post-translational modification sites

Post-translational modification is responsible for the correct folding and transport of proteins. To analyze the PTM sites in ITGA3, NetPhos 3.1 [36] tool was used. It is a web-based tool that uses Artificial Neural Networks (ANNs) to predict serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins.

Prediction of network interaction

To analyze how *ITGA3* gene interacts with other proteins and genes, STRING (Search Tool for Retrieval of Interacting Genes/Proteins) [37] database and GeneMANIA [38] was utilized respectively. STRING is a large database of known and predicted protein-protein signaling interactions. Through these tools, the proteins and genes can be prioritized for functional experiments. The importance of this analysis in the context of current study as when ITGA3 is affected by the missense mutations – the interaction network and associated functions in which the network is involved, will also be affected.

Phylogeny analysis

The phylogenetic analysis is done to thoroughly understand the evolution of different proteins through genetic modifications. So, the NCBI tool (blastp) [39] was used to carry out the phylogenetic tree analysis of the *ITGA* family. This technique presents the results in the form of tree, that helps in the comparison of the degree of relationship between all the sequences present in the set. The results of the phylogenetic analysis are helpful in classifying whether there are subset(s) of the sequences in BLAST output that can be grouped as a family.

RNA expression profile

The online tool UCSC genome browser [40] was used to find out the RNA expression profiles, specifically RNA-seq data. This helps in determination of functional characterization to elucidate the roles of genes in physiology and stress adaptation.

Molecular dynamics simulation

The structure of ITGA3 was subjected to molecular dynamics simulations (MDS) using GROMACS version 2019.4. The protein parameters were generated using the gromos54a7 force field. The cubic simulation box was built using the Gmxeditconf tool. Processed setup was first vacuum minimized for 1500 steps using the steepest descent algorithm. Solvation was performed with the simple point-charge water model using the gmx solvate tool. The gmxgenion tool was used to electro-neutralize the system. Next, steric clashes were removed, and the structure was optimized through energy minimization. After this, the system was equilibrated in two steps. In the first step, the temperature of the system was stabilized through 100 ps of NVT equilibration, where the system was heated up to 300 K. In the second step, the pressure and density of the system were stabilized through 100 ps of NPT ensemble. Each resultant structure from the NPT equilibration phase was subjected for a final production run, for a simulation time of 100 ns.

Moreover, the trajectory analysis was done using various GROMACS analysis tools. The gmxrms and gmxrmsf tools were used to calculate the root-mean square deviation (RMSD) and root-mean-square fluctuations (RMSF) respectively, of the wild type protein and its mutants.

Results and Discussion

SNPs for the protein ITGA3 were identified from HGMD; it identified 51 SNPs in the protein. Among them, 27 SNPs were in the coding region, 7 were missense SNPs, 14 were synonymous, 3 were frameshift, and 3 were start/lost SNPs. Whereas, the other mutations were found in the non-coding region of the protein. This study focuses on the missense SNPs.

Following the analysis of the FASTA sequence from the UniProt database, the length of the protein was found to be 1051 amino acids, located in the cell membrane.

Analysis of mutations

A total of 7 missense SNPs was retrieved from the HGMD database dated July 22, 2024. All of the seven missense SNPs were evaluated using the structure-based and sequence-based tools to explore their deleterious properties. The results of these tools are presented in Table 1 and Figure 1.

The PolyPhen-2 and PANTHER are the software tools that help in predicting whether a particular amino acid substitution has an impact on the biological function of the protein. PolyPhen-2 calculates the PSIC (position-specific independent score) for each input variable. This tool uses Nave Bayes approach for the determination of the implication of changes in allele. PANTHER is a tool capable of identifying and classifying the functions of different gene products by utilizing a curated database of the gene or protein families. Both the tools predicted 5 missense SNPs as deleterious. Tools like Meta-SNP and PredictSNP are also responsible for the detection of disease-associated SNPs. However, these tools integrate pre-existing tools like SIFT, PhD-SNP, PANTHER etc. to improve their prediction accuracy. The SIFT method is used to assess the influence of amino acid replacement on the protein function. Meta-SNP, PredictSNP, and SIFT predicted 6, 6, 5 missense SNPs as deleterious, respectively. SNAP is a program that works on neural networks. It used for the prediction of pathogenicity of Single Amino Acid Variants (SAVs). It has a score ranges from -100 (strongly neutral) to +100 (strongly pathogenic). SNPs&GO is a tool that also utilizes an SVM-based classifier to predict mutations that are likely to be the insurgence of diseases in humans. MAPP (Multivariate Analyzer of Protein Polymorphism) is another computational tool that helps in the prediction of pathogenicity of missense variants. It works by aligning various physiochemical parameters with orthologous sequences to determine the likely impacts of amino acid alterations. PhD-SNP is a support vector machine (SVM) based predictor of human deleterious SNPs. The PMut tool is a neural network-based classifier that allows rapid and accurate pre diction of pathological characteristics of SNPs with a success rate of approximately 80 % in humans. SNAP, SNPs&GO, MAPP, PhD-SNP, and PMUT predicted 5, 4, 4, 6, and 5 missense SNPs as deleterious, respectively. These 7 missense SNPs were further analyzed to predict their structural stability using the structure-based tools.

Protein stability prediction tools analyze the effects of mutations on the balance between a protein's folded and unfolded forms. They consider factors that influence stability of the protein, such as hydrogen bond networks, electrostatic forces, van der Waals contacts and burial propensities of amino acids. I-Mutant 2.0 is a support vector machine (SVM)- based tool for automatically predicting protein stability changes upon single point mutations. It gathers features such as solvent accessibility, secondary structure, phi/psi angles and others. It reports the Gibbs Free Energy Change (DDG). The DDG value is calculated by subtracting the unfolding Gibbs free energy value of the mutated protein from the unfolding Gibbs free energy value of the wild-type protein (Kcal/mol). Positive values imply decreased stability, negative values mean increased stability and zero means little to no change in stability. INPS-MD is a webs server created to predict the impact of missense SNPs on protein stability. CUPSAT predicts protein stability changes upon point mutations using amino acid atom potentials and torsion angles to assess the amino acid environment of the mutation site. This prediction model is also capable of distinguishing the amino acid environment based on the individual solvent accessibility and its secondary structure specificity. DUET is another tool responsible for predicting the effects of mutations on protein stability. However, DUET is an integrated computational approach utilizing mutation Cutoff Scanning Matrix (mCSM) and Site-Directed Mutator (SDM) to generate a consensus-based prediction by combining the results of these 2 methods in an optimized predictor using SVM. MUpro is a software tool designed to predict protein stability changes upon mutation. It uses Support Vector Machine regression to calculate the Gibbs Free Energy Change (DDG) upon mutation, reflecting the impact on protein stability. It is known to achieve excellent agreement with the experimental data, particularly for destabilizing mutations. The range of score is 1 to -1. Less than 0 score

indicates decrease in protein stability and vice versa. I-Mutant, INPS-MD, CUPSAT, DUET, and Mupro predicted 7, 7, 4, 6, and 7 missense SNPs as structurally deleterious mutations, respectively. Out of seven missense SNPs, three SNPs (G125R, R143H, P680A) were identified to be deleterious to the structural stability through all of the five above-mentioned structure-based tools. The results of these tools are summarized in Table 2.

Table 1: Results obtained from sequence-based tools for ITGA3 variants

Amino acid change	PANTHER	Polyphen-2	SIFT	PhD-SNP	Meta-SNP	PredictSNP	SNAP	SNP&GO	MAPP	PMut
G125R	Disease	probably damaging	deleterious	disease	disease	deleterious	disease	disease	deleterious	disease
R143H	Disease	probably damaging	deleterious	disease	disease	deleterious	disease	disease	neutral	disease
R274Q	Disease	probably damaging	deleterious	disease	disease	deleterious	neutral	disease	neutral	neutral
A349S	Neutral	possibly damaging	tolerated	neutral	neutral	neutral	neutral	neutral	neutral	neutral
R463W	Disease	probably damaging	deleterious	disease	disease	deleterious	disease	disease	deleterious	disease
R463W	Neutral	possibly damaging	tolerated	disease	disease	deleterious	disease	neutral	deleterious	disease
P680A	Disease	probably damaging	deleterious	disease	disease	deleterious	disease	neutral	deleterious	disease

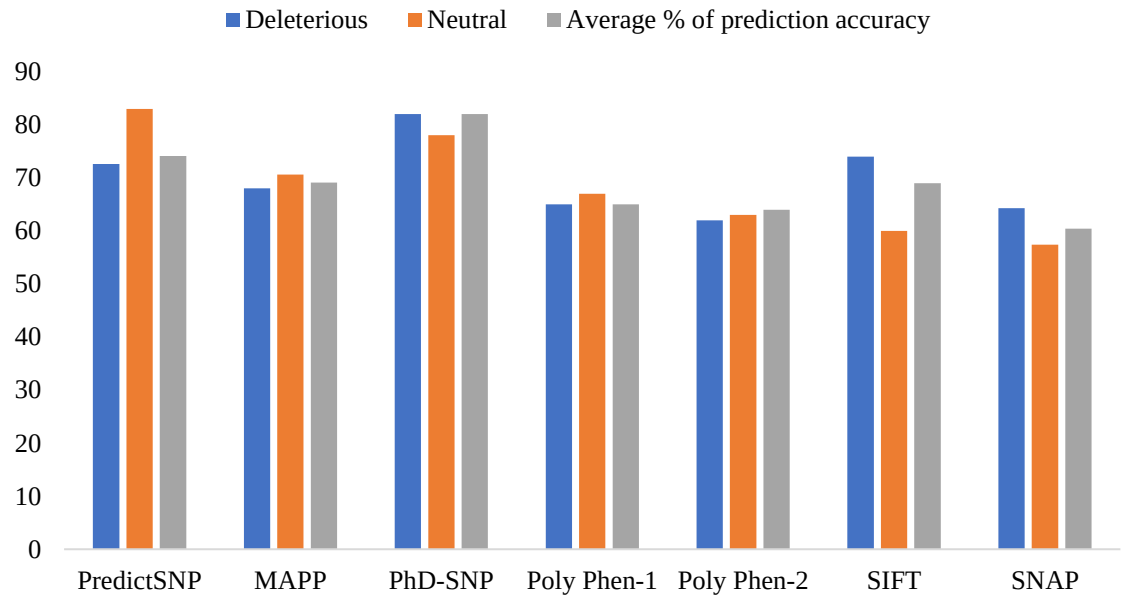


Figure 1: Percentage of prediction pathogenicity of ITGA3 missense variants. PhD-SNP predicted 80% variants as deleterious, followed by PredictSNP and others

Table 2: Prediction of protein stability of mutations in ITGA3 gene by structure-based tools

Amino acid change	I-Mutant	DUET			CUPSAT	MUpro		
		mCSM	SDM	DUET		SVM	Neural Netwok	INPS-MD
G125R	decrease	destabilizing	destabilizing	Destabilizing	unfavorable	decrease	increase	decrease
R143H	decrease	destabilizing	stabilizing	destabilizing	favorable	decrease	decrease	decrease
R274Q	decrease	destabilizing	destabilizing	destabilizing	unfavorable	decrease	decrease	decrease
A349S	decrease	destabilizing	destabilizing	destabilizing	unfavorable	decrease	decrease	decrease
R463W	decrease	destabilizing	stabilizing	destabilizing	favorable	decrease	decrease	decrease
R628P	decrease	stabilizing	destabilizing	stabilizing	unfavorable	decrease	increase	decrease
P680A	decrease	destabilizing	destabilizing	destabilizing	unfavorable	decrease	decrease	decrease

Furthermore, MetaDome web server was used to analyze the mutation tolerance at each position in the ITGA3 protein. MetaDome analyses the mutation tolerance at each position in a human protein. It enhances the analysis of the gene of interest by parallel analysis of all homologous domains in the whole human genome. This helps in identification of regions where certain mutations might be tolerated versus those that are expected to cause disruption due to evolutionary constraints. The result obtained from the MetaDome web server for the analysis of gene ITGA3 with the transcript ENST00000320031.8, shown in Figure 2. The tolerance landscape depicts a missense over synonymous ratio calculated as a sliding window over the entirety of the protein. The missense and synonymous variation are annotated from the gnomAD dataset and the landscape provides some indication of regions that are intolerant to missense variation. The domains are presented as purple blocks; also the missense variants are selected and denoted by the arrow. The results predicted that four of the missense mutations are located at the domain of protein and are intolerant, whereas, other three missense mutations range from tolerant to neutral.

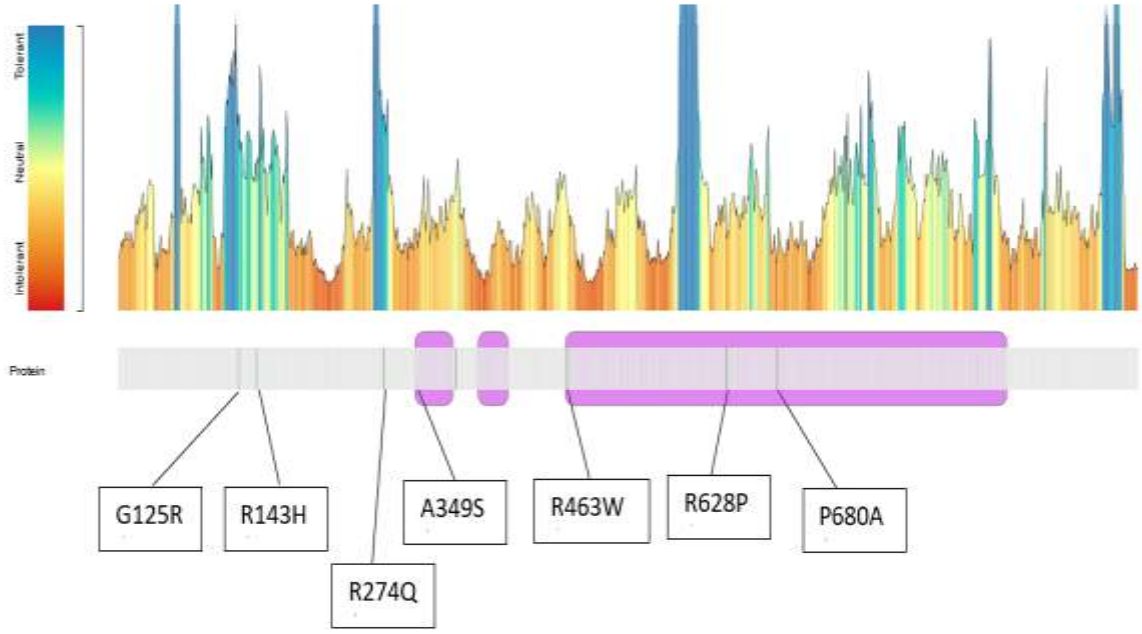


Figure 2: Tolerance visualization of ITGA3 through MetaDome web server with relative positions of selected missense variants

Table 3: Characterization of ITGA3 selected residues position on the basis of tolerance score and presence in protein domain using MetaDome server

Residue	Protein domain	Tolerance prediction	Tolerance score (dn/ds)	Clinvar variants at position	Related variants	
					GnomAD*	ClinVar*
Gly125	-	Tolerant	1.24	0	-	-
Arg143	-	Neutral	0.82	0	-	-
Arg274	-	Tolerant	1.24	0	-	-
Ala349	-	Neutral	0.77	0	-	-
Arg463	PF08441	Neutral	0.71	1	24	0
Arg628	PF08441	Intolerant	0.45	1	11	0
Pro680	PF08441	Intolerant	0.32	0	12	0

Determination of conservation sites

The conservation level of residues gives an idea about the level of damage that deleterious mutations can impart on the protein structurally as well as functionally. A deleterious mutation at a highly conserved residue will very likely be detrimental in nature. The evolutionary conservation profile of the *ITGA3* gene was predicted using the online tool ConSurf. This tool uses the Bayesian technique for detecting the structural and functional residues and to assess the evolutionarily

conserved amino acid residues in the protein. This data was used to examine the possibility of high-risk nsSNP in the protein *ITGA3* to cause damage. The conservation scores for the residues can be between 1 and 9. A score of "1" for a residue signifies that it is extremely variable and conversely, a score of "9" for a residue signifies that it is extremely conserved.

The ConSurf results showed that two of the selected mutations, G125R and R143H were located at the highly conserved regions – with a conservation score of 9. P680A was at the residue 680 and had a conservation score of 7. 416 of the 1051 residues were found in conserved region having the conservation score of greater than or equal to 7; 172 among them were highly conserved with the conservation score equal to 9. Conservation results for chain A of ITGA3 have been structurally represented in Figure 3. To visualize the conservation status of the missense SNPs, the PDBsum database was used, shown in Figure 4. PDBsum is a web-based database providing the pictorial summary of the key information on each macromolecular structure deposited at the Protein Data Bank (PDB). Each protein chain has a direct link to the SAS (Sequence Annotated by Structure) database. It scans the sequence of particular chain of amino acid residues against the database of all sequences in the PDB. The result is a list of all other chains in the PDB that are similar at the sequence level. The SAS database displays different annotations of the resultant multiple-sequence alignment and also represents the superposed structure of protein in 3D. The protein's wiring diagram is annotated with the red lines, helices and arrows depicting the secondary structures – coil, α-helix and β-strand, respectively.

To further confirm the findings of ConSurf server, Multiple sequence alignment (MSA) was performed and it was found that G125R, R143H, and P680A were all present at highly conserved regions as per all four MSA tools (i.e. MUSCLE, MAFFT, T-COFFEE, CLUSTAL) used in the current study. This can also be visually observed in the logo created using the WebLogo tool, where the conservation of sequence is indicated by the overall height of the stack at a given position, while the relative frequency of each amino acid is indicated by the height of the amino acid symbol within the stack at the given position.

Furthermore, the green color indicates a neutral amino acid for hydrophobicity, black color indicates a hydrophobic amino acid, and blue color indicates a hydrophilic amino acid. It was observed that the glycine at position 125, arginine at the position 143, and proline at position 680 are all relatively tall, also have no residue variations amongst all the ITGA3 sequences of different species and except for arginine (pos. 143) the two amino acid residues are neutral, whereas arginine is hydrophilic.

Figure 3: *The color grade indicates the degree of the conservation status of amino acid residues. The color grade rises (1 is highly variable and 9 is a highly conserved site) – predictions of nsSNP in ITGA3 showing conservation scores*

***Figure 3**:*

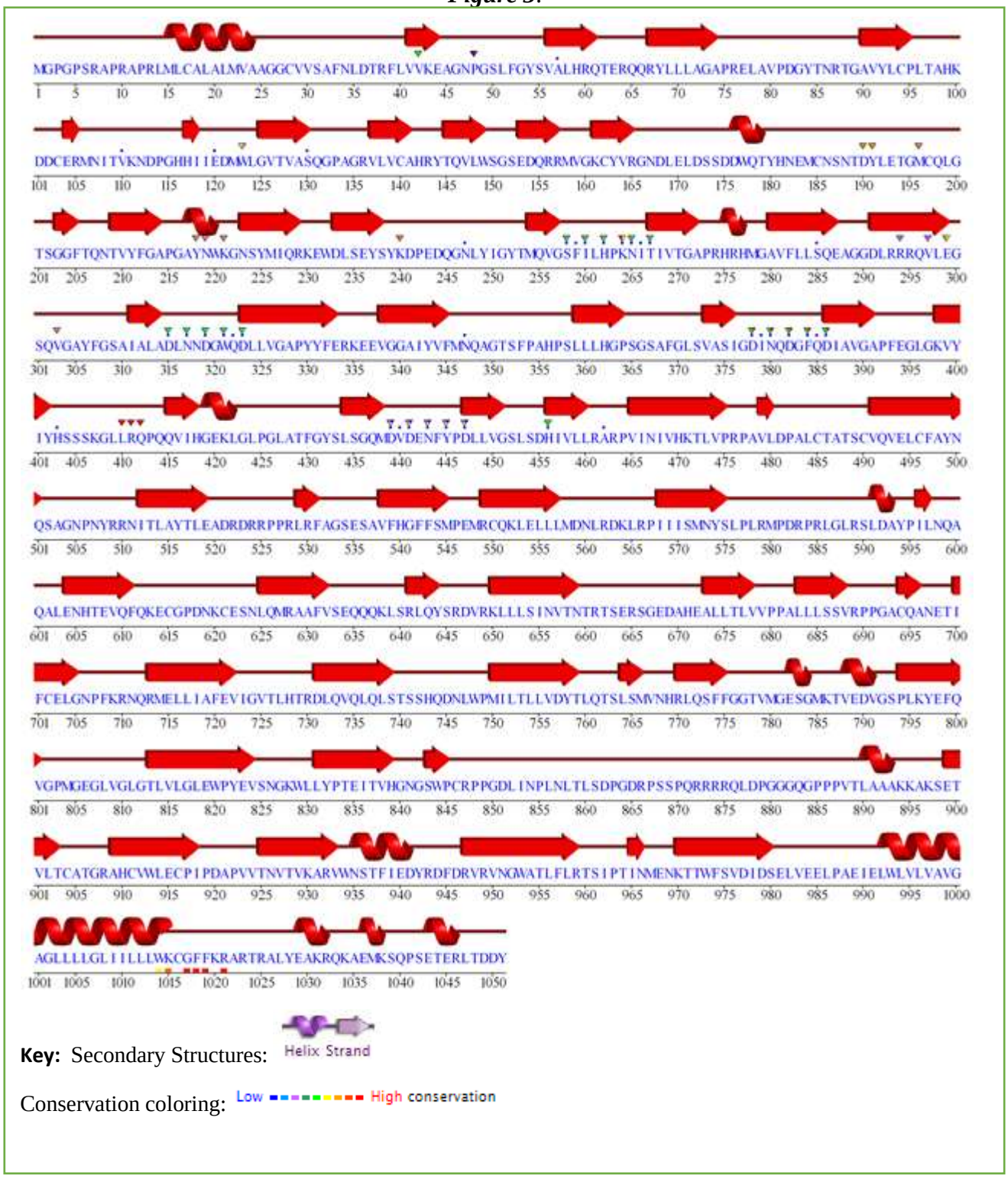


Figure 4: Conservation of ITGA3 Protein by PDBsum, the residues of interest are highlighted with black arrows, conservation score ranges from 1-9 mentioned with colors

3D Protein Structural Analysis:

Interpro projected that *ITGA3* proteins include a large domain conserved across species. Integrin_alpha-2 is the domain that contains amino acids 462-916.

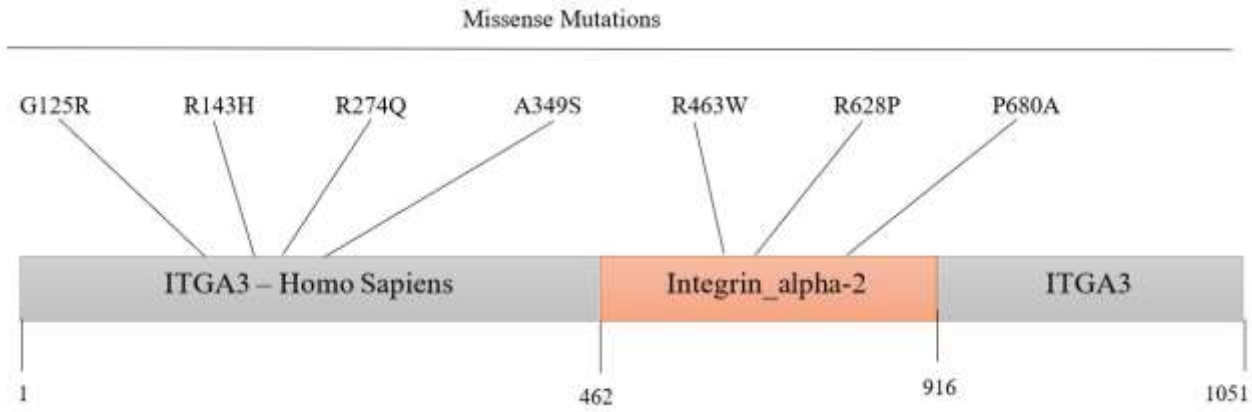


Fig. II: *ITGA3* mutations classified into different domains. The orange color indicates conserved region across species.

Later, the schematic representation of the wild type of *ITGA3* and its' mutants were produced using the online tool Phyre2. The predicted models were examined using the software PYMOL. Each nsSNP in the *ITGA3* protein was assigned the potential 3D structure (Figure IV).

Phylogeny Analysis:

Figure III represents the phylogenetic analysis of ITGA family. Different proteins of ITGA family were presented in the form of a phylogenetic tree.

Prediction of protein domains in ITGA3:

SMART (Simple Modular Architecture Research Tool) server was used to predict the protein domains. SMART is an online analysis tool for the identification and annotation of protein domain. The domains of the protein are searched by entering the ID or sequence of the protein from the Uniprot or Ensembl database. The result of the SMART tool is displayed in Figure 5. Grey line indicates the predicted coiled coil regions, the pink lines represent segments with low complexity, and the blue bar indicates a transmembrane helix. Also, the tool merges the domain prediction and intrinsic features of protein into a single line output.

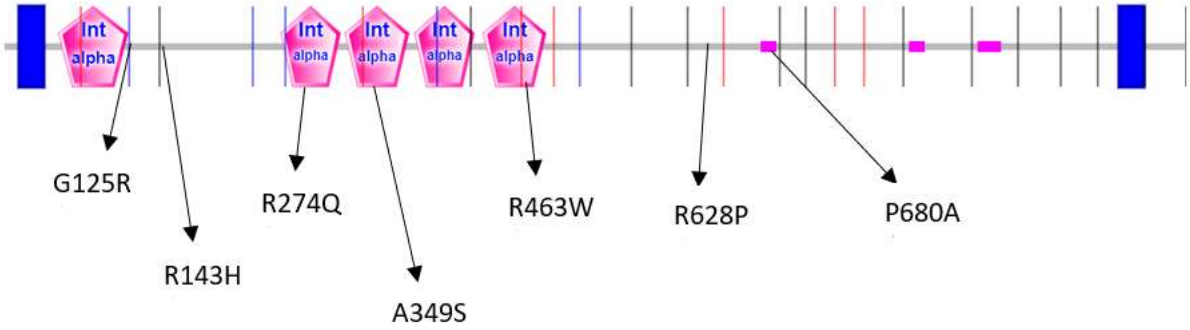


Figure 5: Domains in ITGA3 protein. SMART database was used to retrieve domain data and residues of interests were labelled

Prediction of secondary structure and transmembrane protein helix

The anticipation of secondary structure of ITGA3 gene was done through SOPMA (self-optimized prediction from multiple alignment) server. This software applies a combination of multiple sequence alignments and consensus prediction to increase the accuracy of assigning amino acids to secondary structure classes, i.e. alpha-helix, beta-strand and coil. The secondary structure of protein ITGA3, predicted by SOPMA, are summarized in the Figure 6. It was found that 176 amino acids (16.75%) made alpha helix, 300 amino acids (28.54%) made extended strand and 49 amino acids

(4.66%) made beta turns. The highest number of secondary structures was found to be random coils, i.e. 526 amino acids (50.05%). The structure of three selected mutations were observed, glycine at position 125 was the part of extended strand, arginine at 143 formed the helix, and proline at 680 was the part of coiled region. Also, the secondary structure of protein is represented in graphical form, with blue, red, green and purple lines represents the percentages of helix, sheet, turn, and random coil in the protein respectively, Figure 6b.

a)

SOPMA Labelling:

Alpha Helix Hh	176 is 16.75%	3_{10} helix	Gg 0 is 0.00%	Pi helix	Ii 0 is 0.00%
Beta bridge Bb	0 is 0.00%	Extended strand	Ee 300 is 28.54%	Beta turn	Tt 49 is 4.66%
Bend region Ss	0 is 0.00%	Random coils	Cc 526 is 50.05%		
Ambiguous state ?	0 is 0.00%	Other states	0 is 0.00%		

b)

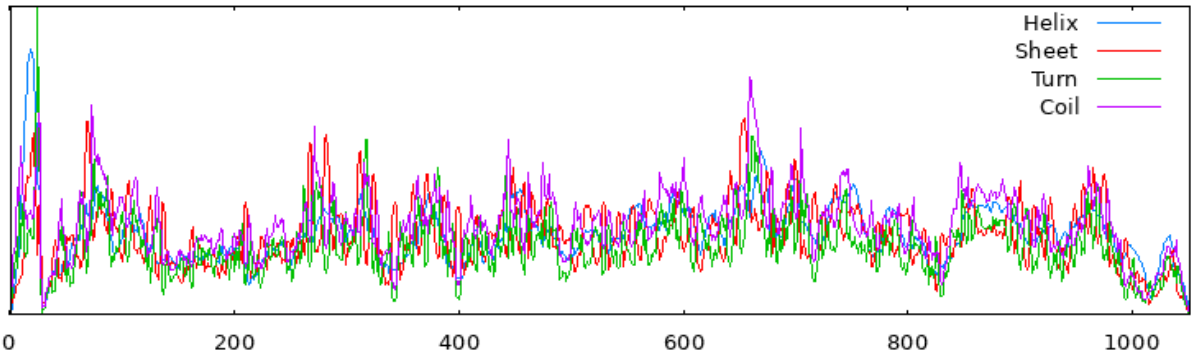


Figure 6: Prediction of ITGA3 protein secondary structure by SOPMA. a) ITGA3 alpha helix, extended strand, beta turns, and random coiled region, b) Graphical representation of secondary structure along the amino acid scale

3D protein structure analysis

The 3D modeling of the ITGA3 protein was visualized in the one of the visualization software, PyMOL. It is a python-based, cross-platform molecular graphics tool and has been widely used for the 3D visualization of nucleic acids, proteins, and other molecules. The main functions of PyMOL includes visualization and analysis enhancement, protein-ligand modeling, molecular simulations, and drug screening. The 3D structures of three selected missense mutations were superimposed with the wildtype structure of protein using the PyMOL software. The change in the structure was observed, as shown in Figure 7.

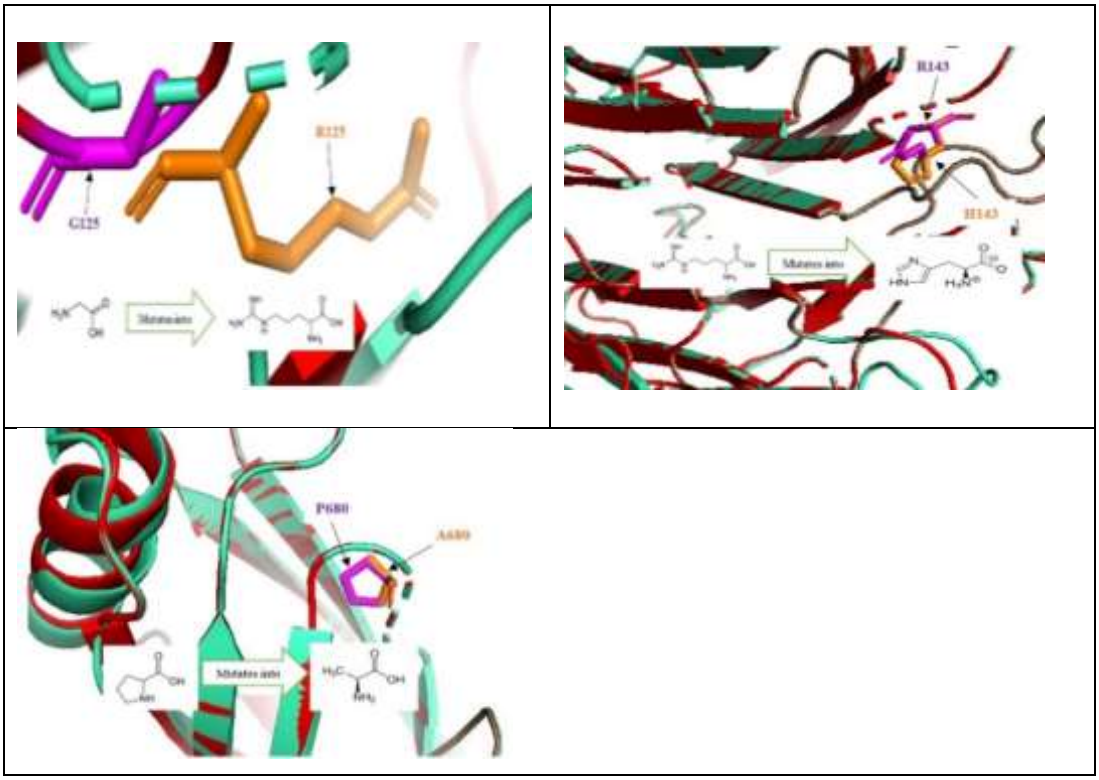


Figure 7: 3D visualization of wild-type amino acid residues (purple) and mutated amino acid residues (orange) for selected mutants through the PyMOL

Prediction of post-translational modification sites

Post-translational modification (PTM) of proteins refers to the chemical changes that occur after a protein has been produced. Some of the most well-known post-translational modifications include

phosphorylation, methylation, acetylation and ubiquitination. Mutations at PTM sites can drastically alter the PTM event, subsequently changing the protein's function. Mutations at adjacent sites can also cause steric hindrance or introduce repulsive forces that impede PTM installation or removal, eventually disturbing protein homeostasis (Narayan et al., 2016). NetPhos 3.1a was used to determine the phosphorylation sites in ITGA3 protein. It produces the neural network predictions for serine, threonine or tyrosine residues in the protein sequences. NetPhos processes protein sequences and generates predictions based on trained ensembles of neural networks. Red, green, blue line indicates the total count of serine, threonine and tyrosine residues that participated in the phosphorylation of the protein, as shown in Figure 8. The changes in the length of phosphorylation sites were observed in the selected mutations. Mutation at position 125 and 680 caused the changes in the length of serine at the respective positions, and the mutation at position at 143 observed the length change of tyrosine. This can help in predicting the effect of missense mutations on the post-translational modification sites.

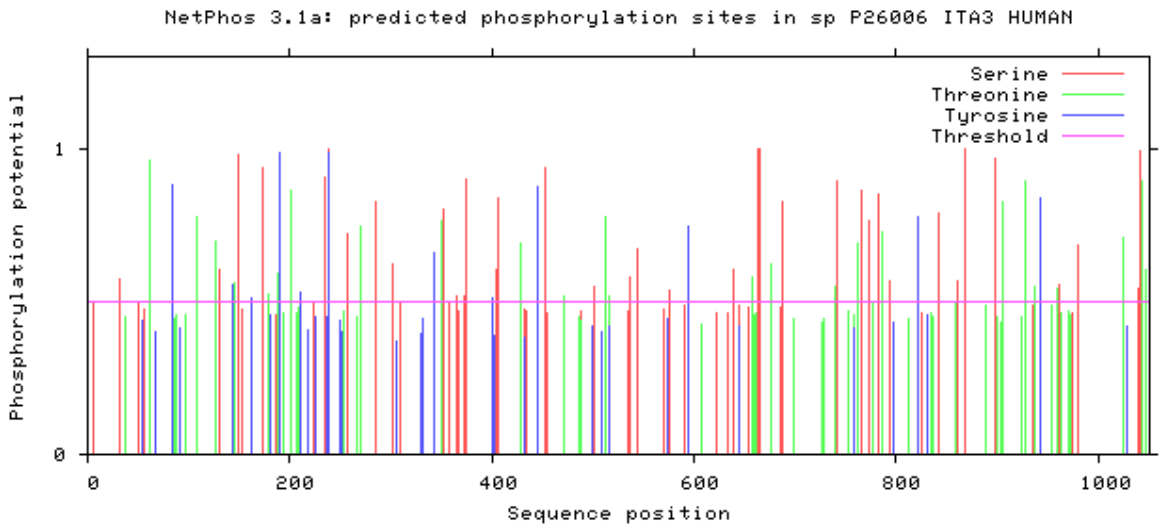


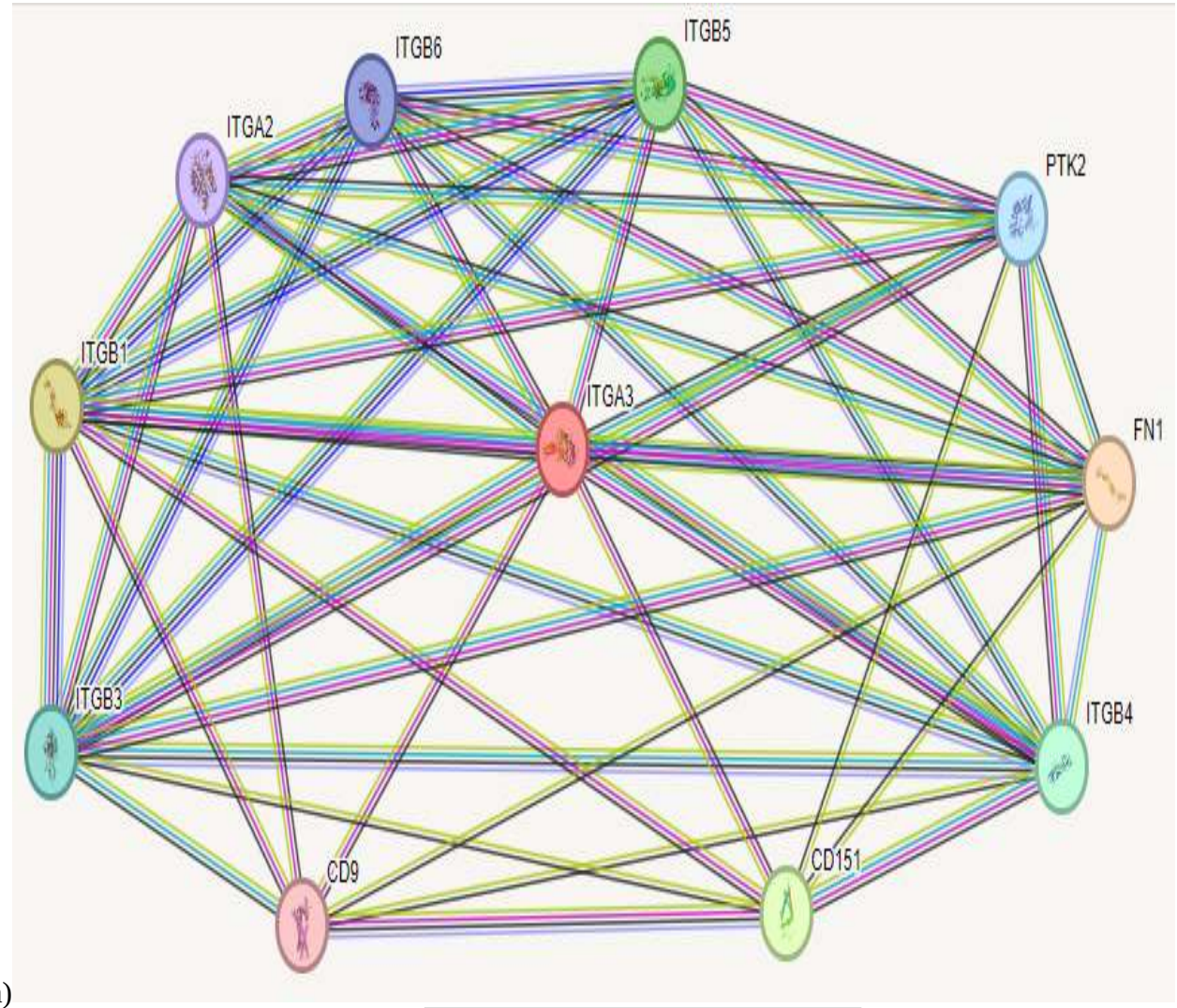
Figure 8: Phosphorylation site prediction using NetPhos3.1 in wildtype protein

Prediction of Network interaction

In every biological system, various genes and their proteins coordinate and interact with each other to carry out specific functions. Mutations in a gene can alter these interactions which results in increased susceptibility to a disease. The protein-protein interaction network of ITGA3 was attained using the STRING database. This is a robust database and web resource focused on collecting, assessing and integrating protein-protein interaction information. The input name was given as "ITGA3" and "Homo sapiens" was selected as the organism. The output represents the proteins and interactions in the form of nodes and edges, respectively. The STRING database found that the ITGA3 gene interacts with ten other proteins, i.e. ITGA2, ITGB1, ITGB3, CD9, ITGB6, ITGB5, CD151, ITGB4, FN1 and PTK2, as shown in Figure 9 (a). All of the predicted functional partners of ITGA3 and their corresponding confidence scores have been summarized in Table 4. All these proteins together contribute to the collective functions with ITGA3. These scores indicate the confidence (i.e. how likely an interaction is to be true) estimated by STRING with the help of available evidence. The STRING analysis revealed that the biological processes in which the network is mainly involved include the skin morphogenesis, fibronectin binding, integrin binding, positive regulation of fibroblast migration, adhesion of calcium-independent cell-matrix, laminin binding, and others – all of the biological processes had a strength value greater than 2.

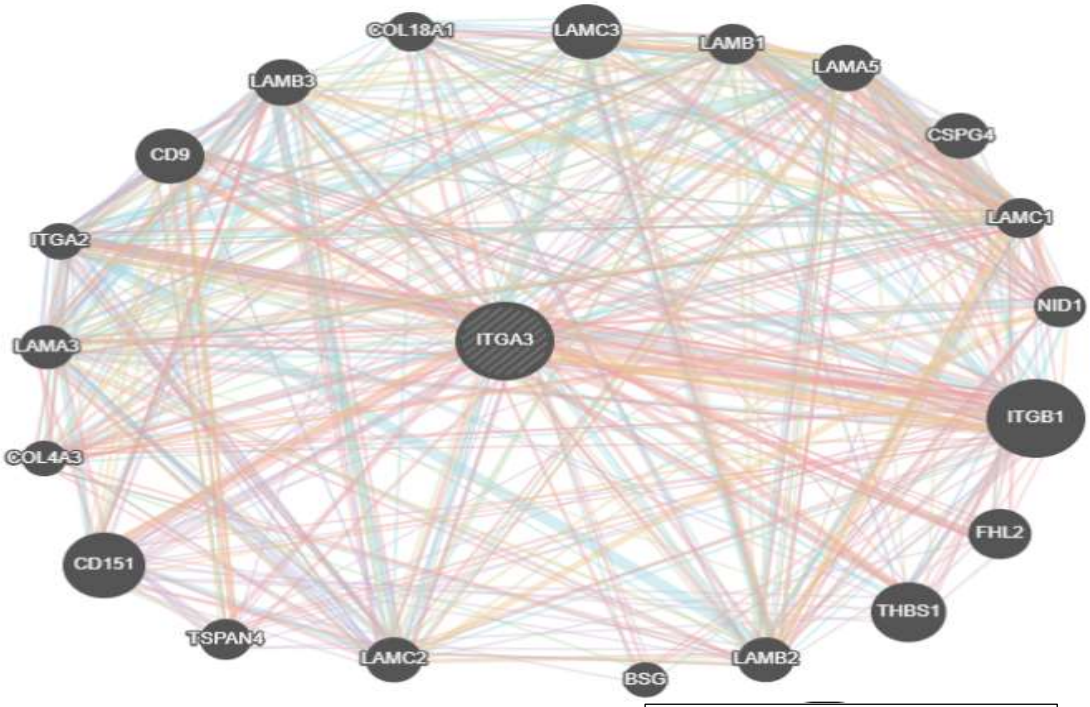
The gene-gene interaction of ITGA3 was studied using the GeneMANIA. GeneMANIA is a versatile tool that identifies functionally related genes and predicts gene functions. It is a flexible user-friendly tool for generating hypotheses about gene function, analyzing gene lists and prioritizing genes for functional assays. GeneMANIA finds genes that are likely to share function

with the query gene based on their interactions. The output represents the interaction in different colored lines linking the genes with each other. The GeneMANIA found that ITGA3 gene interacts with 20 other genes, i.e. LAMB3, CD9, ITGA2, LAMA3, COL4A3, CD151, TSPAN4, LAMC2, BSG, LAMB2, THBS1, FHL2, ITGB1, NID1, LAMC1, CSPG4, LAMA5, LAMB1, LAMC3 and COL18A1. It found 20 related genes, having a total of 597 links, shown in Figure 9 (b).



(a)

Nodes:
Represent proteins
Edges:
From curated databases
Experimentally determined
Gene neighborhood
Gene fusions
Gene co-occurrence
Textmining
Co-expression
Protein homology



(b)

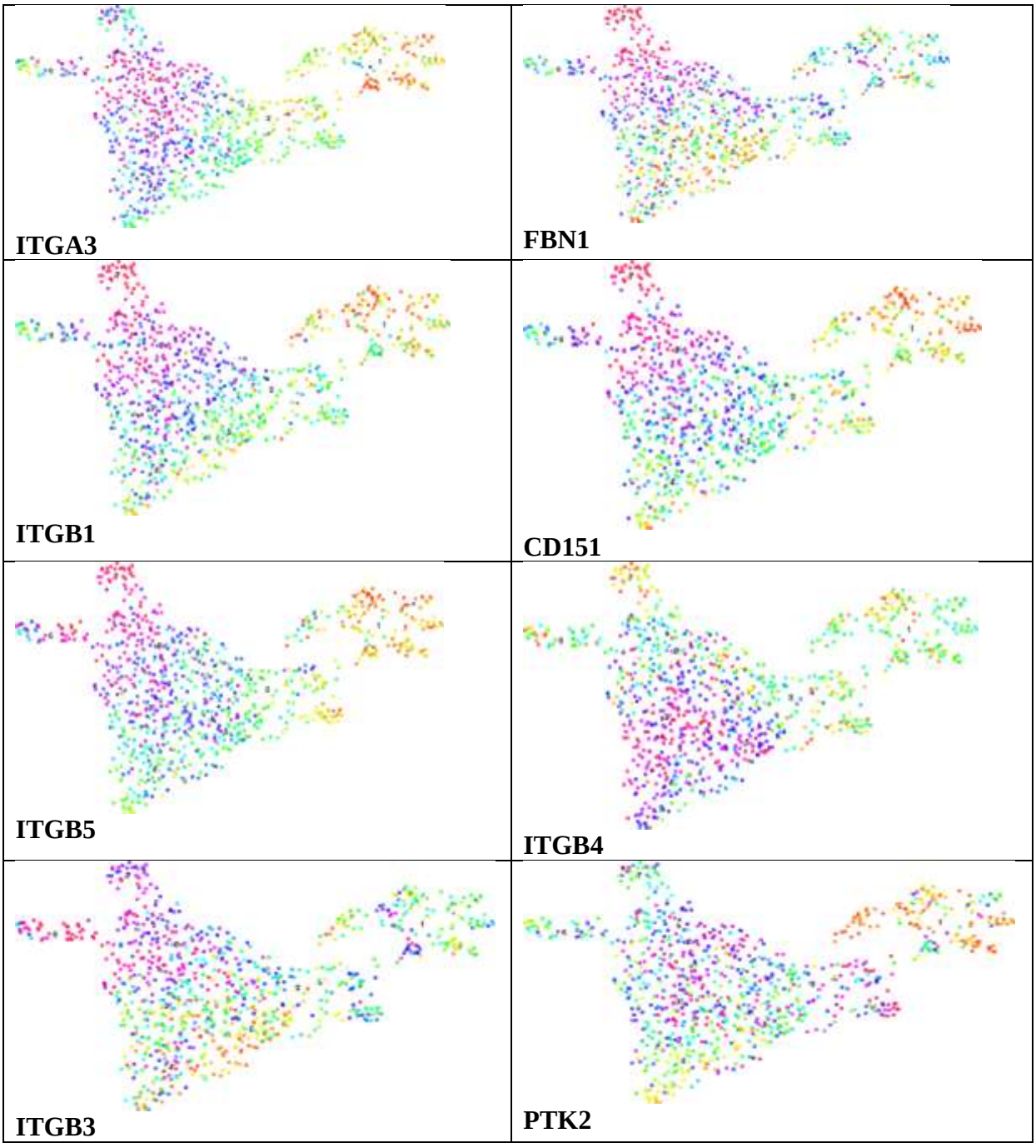
Figure 9: Protein to Protein and gene to gene interaction of ITGA3 by a), STRING database, b) GeneMANIA, respectively

Table 4: Prediction of molecular interaction of ITGA3 with other proteins using STRING database; C=Co-expression, E=Experiments, D=Databases, T=Text mining, H=Homology

Predicted functional parameters	Prediction for specific action	Active interaction sources	Confidence scores
FBN1	bind cell surfaces and various compounds including collagen, fibrin, heparin, DNA, and actin	C,E,D,T	0.999
ITGB1	receptors for collagen	C,E,D,T	0.999
CD151	Essential for the proper assembly of the glomerular and tubular basement membranes in kidney	C,E,T	0.993
ITGB5	receptor for fibronectin, recognizes the sequence R-G-D in its ligand	C,E,D,T	0.986
ITGB4	receptor for laminin, plays a critical structural role in the hemidesmosome of epithelial cells	C,E,D,T	0.982
ITGB3	receptor for cytotactin, fibronectin, laminin, matrix metalloproteinase-2, osteopontin, osteomodulin, prothrombin, thrombospondin, vitronectin and von Willebrand factor	C,E,D,T	0.979
PTK2	Non-receptor protein-tyrosine kinase that plays an essential role in regulating cell migration, adhesion, spreading, reorganization of the actin cytoskeleton	C,D,T	0.978
ITGB6	receptor for fibronectin and cytotactin, recognizes the sequence R-G-D in its ligands.	C,E,D,T	0.977
ITGA2	receptor for laminin, collagen, collagen C-propeptides, fibronectin and E-cadherin, recognizes the proline-hydroxylated sequence G-F-P-G-E-R in collagen	C,D,T,H	0.972
CD9	Integral membrane protein associated with integrins, which regulates different processes, such as sperm-egg fusion, platelet activation and aggregation, and cell adhesion	C,E,T	0.969

RNA expression profiles

The Treehouse Childhood Cancer group was used for determining the RNA expression profile of ITGA3; it is the research arm of the UCSC Genomics Institute. The UCSC Cell Browser interactively presents the samples in Treehouse dataset that are positioned according to their RNA profiles. The purpose of Treehouse Childhood Cancer Initiative is to carry out the evaluation of utility of the comparative gene expression analysis for pediatric cancer patient(s) that are difficult-to-treat. The collection "Treehouse Cancer Compendium" contains eight datasets. RNA expression is measured for each of the associated interactor protein along with its overall expression. Level of expression in each tissue has differentiation on the base of its color. Highest expression is shown in red and pink colors while the lowest expression with orange color and all the other colors falls in between, as shown in figure 10. The proteins ITGA3, FBN1, ITGB1, C151, ITGB5 and ITGB3 displayed the highest expression in central nervous system, pleura, thyroid, skin and kidneys. Whereas ITGB4, ITGB6, ITGA2 and CD9 showed highest expression in kidneys, liver, respiratory system and digestive system. All the proteins, however, showed expression in kidneys.



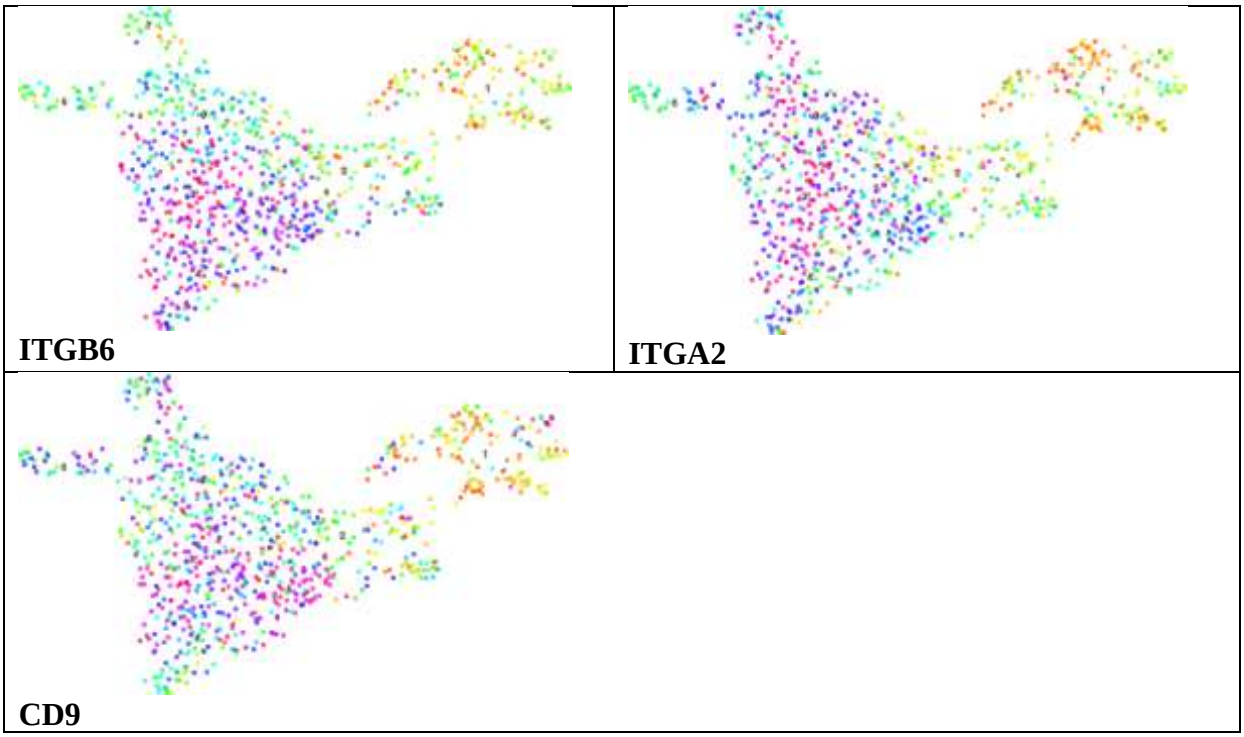


Figure 10: Visualization of ITGA3 single-cell RNA expression through RNA-seq, with its interactors FBN1, ITGB1, CD151, ITGB5, ITGB4, ITGB3, PTK2, ITGB6, ITGA2, and CD9, retrieved from UCSC cell genome browser. Pink color dots show highest while orange red color shows the lowest frequency of ITGA3 genes. The numbers represent the following organs in the body

0 = pleura and thyroid
1 = haematopoietic and lymphoid tissue
2 = small intestine and autonomic ganglia
3 = kidneys, liver, urinary tract, lungs and endometrium
4 = salivary gland, oesophagus and upper aerodigestive tract
5 = large intestine and prostate
6 = skin
7 = central nervous system

Molecular dynamics simulation

Molecular dynamics simulations were performed to identify the conformational changes between the wild type protein and its mutants. After generating four systems (Wild Type, G125R, R143H, and P680A), MDS was performed for 100 ns using GROMACS. Various above-mentioned GROMACS functions were used for trajectory analysis.

Stability analysis

RMSD was calculated for the wild type protein and its mutant for predicting their respective stabilities. The RMSD value was used for measuring the difference between the backbones of the wild-type protein and mutants. The stability of the protein relative to its conformation can be determined by the RMSD deviations produced during MDS. Smaller RMSD deviations imply that the protein structure is more stable. RMSD of the protein backbone atoms was plotted against the time to assess its variations in structural confirmation. The wild-type protein had an average RMSD value of 2.1 nm (Figure 11a). Meanwhile, the mutants G125R, R143H, and P680A had average RMSD values of 3.5, 1.55, and 2.2 nm (Figure 11b-d), respectively. Mutant G125R was noticeably less stable as compared to the wild type protein. Furthermore, it was also observed that the G125R mutant showed a large fluctuation in RMSD after 10 ns when compared to the wild-type protein and

other mutants. Furthermore, the R143H mutant seemed to be more stable in comparison with wild-type protein – as its average RMSD value is less than the wild type protein.

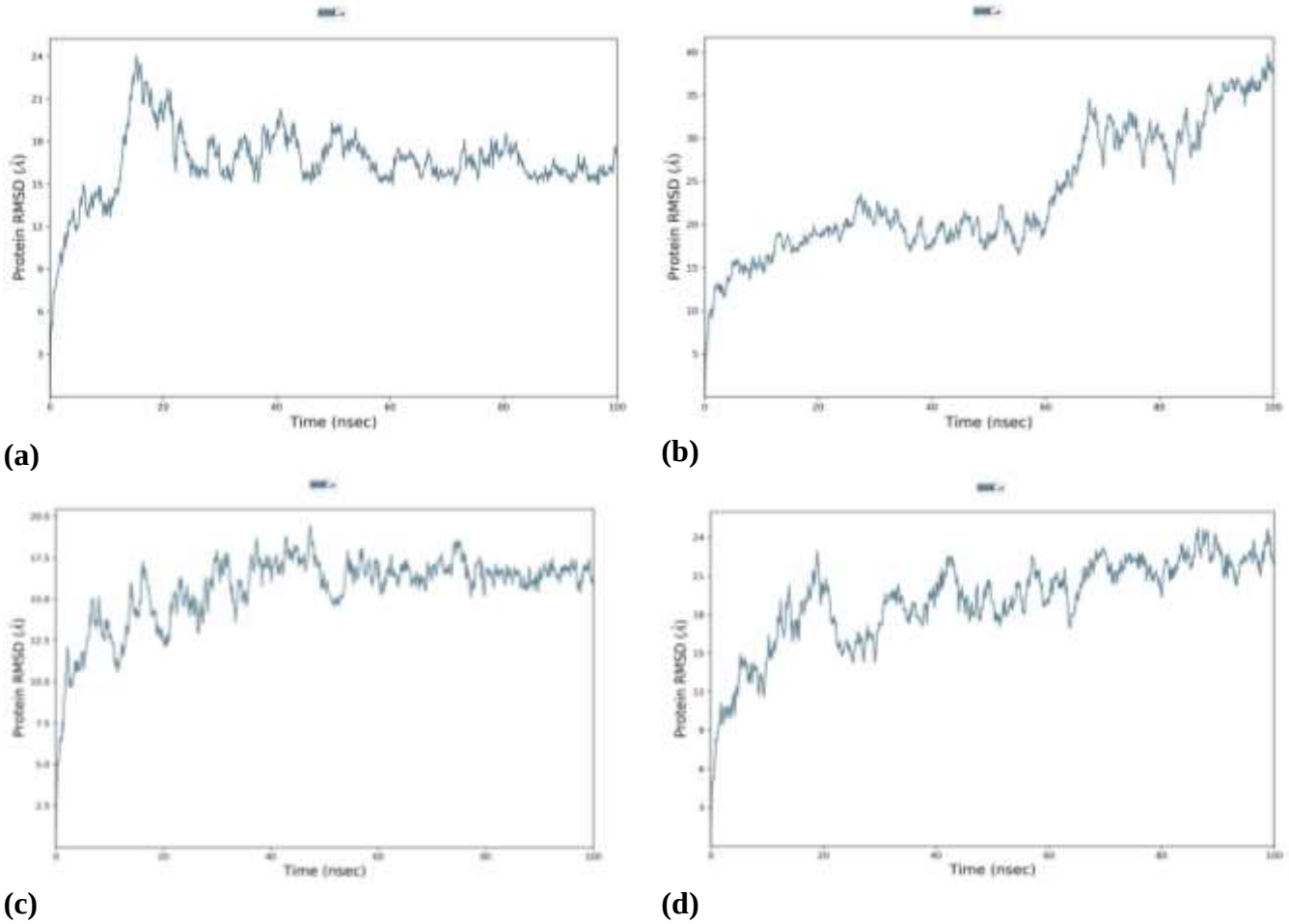


Figure 11: RMSD values from Molecular dynamic simulation of wild-type and mutant proteins

Flexibility analysis

RMSF is also a crucial parameter for examining the flexibility and stability of complex systems during simulations. High RMSF values indicate greater flexibility during the MDS. Hence, RMSF values were calculated to determine the structural flexibility of the wild-type protein and its mutant. The average value of RMSF for the wild-type protein was found to be 0.9 nm (Figure 12a). Meanwhile, the mutants G125R, R143H, and P680A had average RMSF values of 1.2, 0.4, and 0.6 nm respectively – Figure 12b-d.

The G125R and P680A mutants showed a peak of 0.20 and 0.21 nm, respectively at the very first residue. All mutants, especially G125R, peak from residue positions 100 to 200. This region mainly consists of β-strands. G125R also shows high RMSF values from residue positions of 450 to 500, and 750 to 800.

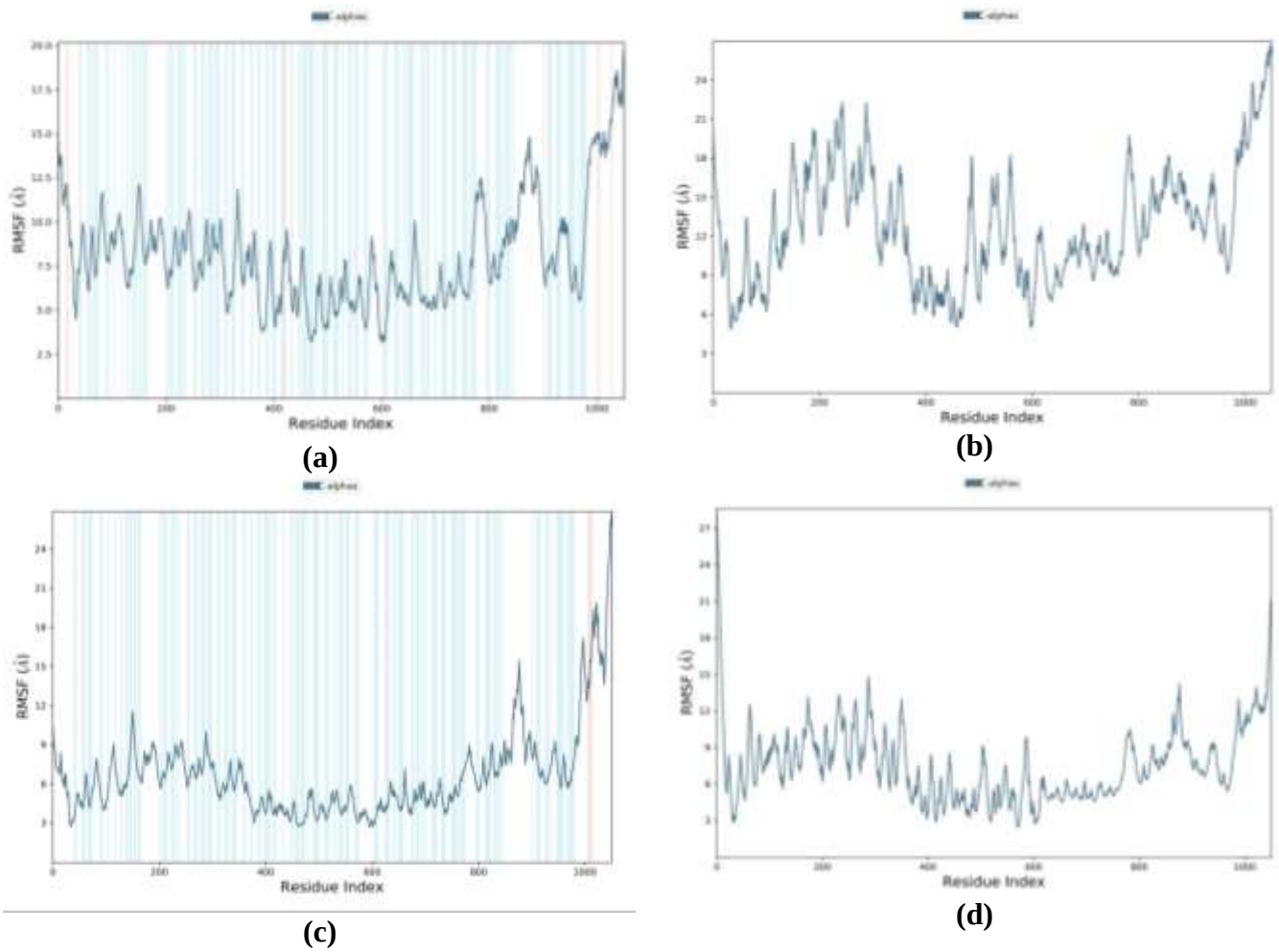


Figure 12: RMSF values from Molecular dynamic simulation of wild-type and mutant proteins

Conclusion:

The *ITGA3* gene plays an important role in the formation of integrin and anchoring proteins that play a vital role in the development of Extracellular matrix of the kidneys. In this study, the nsSNPs present in the gene have been analyzed. As a result, out of 7 missense mutations in ITGA3 (reported in HGMD database) the three of them – G125R, R143H, P680A – were identified as highly deleterious. Our comprehensive bioinformatic analysis sheds light on the intricate relationship between missense mutations in the *ITGA3* gene and their impact on protein structure and function, particularly in the context of nephrotic syndrome. By elucidating these molecular mechanisms, our study provides a crucial foundation for future research aimed at developing targeted therapies and personalized treatment approaches for individuals affected by nephrotic syndrome. Furthermore, our findings underscore the importance of integrating bioinformatics into clinical practice, offering valuable insights that can ultimately improve patient care and outcomes. We can harness the power of bioinformatics to advance our understanding of genetic diseases and pave the way for more effective interventions in future.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets analyzed during the current study are available in the HGMD repository. The datasets used during this study are available from the corresponding author on reasonable request. All data generated or analyzed during this study are included in this published article.

The datasets analyzed during the current study are not publicly available due to the restriction of the access to the HGMD web portal but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

The present study was funded by the Higher Education Commission Pakistan.

Authors' contribution

Q.N. and M.Y.Z. have worked on the main manuscript. W.S. and M.A.A have worked on the compilation of data

Acknowledgements

Not applicable

References:

1 Kalluri R (2006) Proteinuria with and without renal glomerular podocyte effacement. J Am Soc Nephrol 17(9): 2383-2389

2 McCloskey O, Maxwell AP (2017) Diagnosis and management of nephrotic syndrome. Pract 261(1801): 11-15

3 Gigante M, Caridi G, Montemurno E, Soccio M, d'Apolito M, Cerullo G, Gesualdo L (2011) TRPC6 mutations in children with steroid-resistant nephrotic syndrome and atypical phenotype. Clin J Am Soc Nephrol 6(7): 1626-1634

4 Margadant C, Charafeddine RA, Sonnenberg A (2010) Unique and redundant functions of integrins in the epidermis. FASEB J 24(11):4133-4152

5 Humphries JD, Byron A, Humphries MJ (2006) Integrin ligands at a glance. J Cell Sci 119(19): 3901-3903

6 Shen B, Delaney MK, Du X (2012) Inside-out, outside-in, and inside–outside-in: G protein signaling in integrin-mediated cell adhesion, spreading, and retraction. Curr Opin Cell Biol 24(5):600-606

7 Goodman SL, Picard M (2012) Integrins as therapeutic targets. Trends Pharmacol Sci 33(7):405-412

8 Sneha P, Doss CG (2016) Molecular Dynamics: New Frontier in Personalized Medicine. Adv Protein Chem Struct Biol 102: 181-224. https://doi.org/10.1016/bs.apcsb.2015.09.004

9 Adzhubei D, Jordan M, Sunyaev SR (2013) Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. Curr Protoc 76:20

10 Stenson PD, et al. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat.* 2003;**21**:577–581

11 The UniProt Consortium, UniProt: the universal protein knowledgebase in 2021, Nucleic Acids Res. 49 (2021) D480–D489, https://doi.org/10.1093/nar/gkaa1100.

12 Paul D. Thomas, Michael J. Campbell, Anish Kejariwal, Huaiyu Mi, Brian Karlak, Robin Daverman, Karen Diemer, Anushya Muruganujan, Apurva Narechania. 2003. **PANTHER: a library of protein families and subfamilies indexed by function.** *Genome Res.*, 13: 2129-2141 doi: 10.1101/gr.772403. PMID: 12952881; PMCID: PMC403709

13 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248–249
14 Capriotti E, Altman RB, Bromberg Y. Collective judgment predicts disease-associated single nucleotide variants. BMC Genomics. 2013;14 Suppl 3(Suppl 3):S2. doi: 10.1186/1471-2164-14-S3-S2. Epub 2013 May 28. PMID: 23819846; PMCID: PMC3839641
15 Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, Brezovsky J, Damborsky J. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput Biol. 2014 Jan;10(1):e1003440. doi: 10.1371/journal.pcbi.1003440. Epub 2014 Jan 16. PMID: 24453961; PMCID: PMC3894168
1616 Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003 Jul 1;31(13):3812-4. doi: 10.1093/nar/gkg509. PMID: 12824425; PMCID: PMC168916
17 Stone E.A., Sidow A. Physicochemical constraint violation by missense substitution mediates impairment of protein function and disease severity. Genome Research (2005) 15 978-986
18 López-Ferrando V, Gazzo A, de la Cruz X, Orozco M, Gelpí JL. PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. Nucleic Acids Res. 2017 Jul 3;45(W1):W222-W228. doi: 10.1093/nar/gkx313. PMID: 28453649; PMCID: PMC5793831
19 Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI "SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap." Bioinformatics 2008; 24(24):2938-9 https://doi.org/10.1093/bioinformatics/btn564 PMID: 18974171 PMCID: PMC2720775
20 Capriotti E, Fariselli P. PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. Nucleic Acids Res. 2017 Jul 3;45(W1):W247-W252. doi: 10.1093/nar/gkx369. PMID: 28482034; PMCID: PMC5570245
21 Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. BMC Genomics. 2013;14 Suppl 3(Suppl 3):S6. doi: 10.1186/1471-2164-14-S3-S6. Epub 2013 May 28. PMID: 23819482; PMCID: PMC3665478
22 Wiel L, Baakman C, Gilissen D, Veltman JA, Vriend G, Gilissen C. MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. Hum Mutat. 2019 Aug;40(8):1030-1038. doi: 10.1002/humu.23798. Epub 2019 Jun 18. PMID: 31116477; PMCID: PMC6772141.
23 Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 2005 Jul 1;33 W306-10. doi: 10.1093/nar/gki375. PMID: 15980478; PMCID: PMC1160136.
24 Savojardo C, Fariselli P, Martelli PL, Casadio R. INPS-MD: a web server to predict stability of protein variants from sequence and structure. Bioinformatics. 2016 Aug 15;32(16):2542-4. doi: 10.1093/bioinformatics/btw192. Epub 2016 Apr 10. PMID: 27153629.
25 Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins. 2006 Mar 1;62(4):1125-32. doi: 10.1002/prot.20810. PMID: 16372356.
26 Parthiban V, Gromiha MM, Schomburg D. CUPSAT: prediction of protein stability upon point mutations. Nucleic Acids Res. 2006 Jul 1; 34:W239-42. doi: 10.1093/nar/gkl190. PMID: 16845001; PMCID: PMC1538884.
27 Pires DE, Ascher DB, Blundell TL. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. Nucleic Acids Res. 2014 Jul; 42:W314-9. doi: 10.1093/nar/gku411. Epub 2014 May 14. PMID: 24829462; PMCID: PMC4086143.
28 Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules.

Nucleic Acids Res. 2016 Jul 8;44(W1):W344-50. doi: 10.1093/nar/gkw408. Epub 2016 May 10. PMID: 27166375; PMCID: PMC4987940.

29 Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: Structural summaries of PDB entries. Protein Sci. 2018 Jan;27(1):129-134. doi: 10.1002/pro.3289. Epub 2017 Oct 27. PMID: 28875543; PMCID: PMC5734310.

30 Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019 Jul 2;47(W1):W636-W641. doi: 10.1093/nar/gkz268. PMID: 30976793; PMCID: PMC6602479.

31 Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004 Jun;14(6):1188-90. doi: 10.1101/gr.849004. PMID: 15173120; PMCID: PMC419797.

32 Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res. 2000 Jan 1;28(1):231-4. doi: 10.1093/nar/28.1.231. PMID: 10592234; PMCID: PMC102444.

33 C. Geourjon, G. Deléage, SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, *Bioinformatics*, Volume 11, Issue 6, December 1995, Pages 681–684, https://doi.org/10.1093/bioinformatics/11.6.681

34 Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015 Jun;10(6):845-58. doi: 10.1038/nprot.2015.053. Epub 2015 May 7. PMID: 25950237; PMCID: PMC5298202.

35 Schrödinger, L., & DeLano, W. (2020). *PyMOL*. Retrieved from http://www.pymol.org/pymol

36 Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics. 2004 Jun;4(6):1633-49. doi: 10.1002/pmic.200300771. PMID: 15174133.

37 Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ, von Mering C. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 2021 Jan 8;49(D1):D605-D612. doi: 10.1093/nar/gkaa1074. Erratum in: Nucleic Acids Res. 2021 Oct 11;49(18):10800. PMID: 33237311; PMCID: PMC7779004.

38 Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010 Jul;38(Web Server issue):W214-20. doi: 10.1093/nar/gkq537. PMID: 20576703; PMCID: PMC2896186.

39 Mark Johnson, Irena Zaretskaya, Yan Raytselis, Yuri Merezhuk, Scott McGinnis, Thomas L. Madden, NCBI BLAST: a better web interface, *Nucleic Acids Research*, Volume 36, Issue suppl_2, 1 July 2008, Pages W5–W9, https://doi.org/10.1093/nar/gkn201

40 Olena Morozova, Yulia Newton, Melissa Cline, Jingchun Zhu, Katrina Learned, Josh Stuart, Sofie Salama, Robert Arceci, David Haussler. Treehouse Childhood Cancer Project: a resource for sharing and multiple cohort analysis of pediatric cancer genomics data. [abstract]. In: Proceedings of the 106th Annual Meeting of the American Association for Cancer Research; 2015 Apr 18-22; Philadelphia, PA. Philadelphia (PA): AACR; Cancer Res 2015