Research article

Prediction and assessment of deleterious and disease causing nonsynonymous single nucleotide polymorphisms (nsSNPs) in human *FOXP4* gene: An *in-silico* study

Md. Mostafa Kamal [a,d,*], Shamiha Tabassum Teeya [a], Md. Mahfuzur Rahman [b], Md. Enamul Kabir Talukder [c,d], Sonia Sarmin [e], Tanveer A. Wani [f], Md. Mahmudul Hasan [a,**]

[a] *Department of Nutrition and Food Technology, Jashore University of Science and Technology, Jashore, 7408, Bangladesh*
[b] *Department of Genetic Engineering & Biotechnology, Bangabandhu Sheikh Mujibur Rahman Maritime University, Dhaka, 1216, Bangladesh*
[c] *Department of Genetic Engineering & Biotechnology, Jashore University of Science and Technology, Jashore, 7408, Bangladesh*
[d] *Laboratory of Computational Biology, Biological Solution Centre, Jashore, 7408, Bangladesh*
[e] *BIRTAN-Bangladesh Institute of Research and Training on Applied Nutrition, Jhenaidah, 7300, Bangladesh*
[f] *Department of Pharmaceutical Chemistry, College of Pharmacy, King Saud University, Riyadh, 11451, Saudi Arabia*

ARTICLE INFO

ABSTRACT

In humans, *FOXP* gene family is involved in embryonic development and cancer progression. The *FOXP4* (Forkhead box protein P4) gene belongs to this *FOXP* gene family. *FOXP4* gene plays a crucial role in oncogenesis. Single nucleotide polymorphisms are biological markers and common determinants of human diseases. Mutations can largely affect the function of the corresponding protein. Therefore, the molecular mechanism of nsSNPs in the *FOXP4* gene needs to be elucidated. Initially, the SNPs of the *FOXP4* gene were extracted from the dbSNP database and a total of 23124 SNPs was found, where 555 nonsynonymous, 20525 intronic, 1114 noncoding transcript, 334 synonymous were obtained and the rest were unspecified. Then, a series of bioinformatics tools (SIFT, PolyPhen2, SNAP2, PhD SNP, PANTHER, I-Mutant2.0, MUpro, GOR IV, ConSurf, NetSurfP 2.0, HOPE, DynaMut2, GeneMANIA, STRING and Schrodinger) were used to explore the effect of nsSNPs on FOXP4 protein function and structural stability. First, 555 nsSNPs were analyzed using SIFT, of which 57 were found as deleterious. Following, PolyPhen2, SNAP2, PhD SNP and PANTHER analyses, 10 nsSNPs (rs372762294, rs141899153, rs142575732, rs376938850, rs367607523, rs112517943, rs140387832, rs373949416, rs373949416 and rs376160648) were common and observed as deleterious, damaging and diseases associated. Following that, using I-Mutant2.0 and MUpro servers, 7 nsSNPs were found to be the most unstable. GOR IV predicted that these seven nsSNPs affect protein structure by altering the protein contents of alpha helixes, extended strands, and random coils. Following DynaMut2, 5 nsSNPs showed a decrease in the ΔΔG value compared with the wild-type and were found to be responsible for destabilizing the corresponding protein. GeneMANIA and STRING network revealed interaction of FOXP4 with other genes. Finally, molecular dynamics simulation analysis revealed consistent fluctuation in RMSD and RMSF values, Rg and hydrogen bonds in the mutant

---

* Corresponding author. Department of Nutrition and Food Technology, Jashore University of Science and Technology, Jashore, 7408, Bangladesh.

** Corresponding author.

E-mail addresses: rmkmostafa@gmail.com (Md.M. Kamal), hasanm_agb@yahoo.com (Md.M. Hasan).

proteins compared with WT, which might alter the functional and structural stability of the corresponding protein. As a result, the aforementioned integrated comprehensive bioinformatic analyses provide insight into how various nsSNPs of the *FOXP4* gene change the structural and functional properties of the corresponding protein, potentially proceeding with the pathophysiology of human diseases.

1. Introduction

Forkhead box protein P4 (FOXP4) protein is encoded by the *FOXP4* gene [1]. This gene also known as hFKHLA plays a crucial role in oncogenesis [2]. *FOXP4* gene belongs to subfamily P of the forkhead box (FOX) transcription factor family which also consists of three different proteins FOXP1 (3p14.1), FOXP2 (7q31), and FOXP3 (Xp11.23). All of subfamily P of the forkhead box are comprised of immune disorders, embryonic maturity, and cancer progression [3]. Those proteins also show multifarious and influential roles in the development and organogenesis for the regulation of metabolism and the immune system [4]. *FOXP4* gene is precisely associated with FOXP1 and FOXP2 with 54 % and 60 % identity, respectively. *FOXP4* gene forms a large multidomain transcriptional repressor with FOXP1 and FOXP2 [5]. It is a transcription factor of b-catenin, raises the transcription of b-catenin and leads to the malicious progress of ESCC [6]. This gene also plays a vital role in the expansion of tumors in the kidney and larynx [7]. As a transcriptional factor, this gene is also involved in the triggering of several genes that play a vital role in various biological activities for cancer progression [8].

Single nucleotide polymorphism (SNP) is the replacement (variation) of a single nucleotide in deoxyribonucleic acid [9]. It can occur in the coding regions of the gene and make a change in amino acid (nonsynonymous), or can be silent without changing in amino acid (synonymous), or can also occur in noncoding regions (5′UTR, 3′UTR, and introns) [10]. In the human genome, SNPs are the most common cause of genetic alteration. They make up nearly 90 % of human genome polymorphism [11]. Mutations can largely affect the function of corresponding proteins, hence analysis of possible modifications in protein tertiary structure can reveal new evidence for their effect on phenotype. The nsSNPs are one of the major causes of the functional diversity of proteins. The nsSNPs also known as missense variations, which change the amino acid sequence of the protein, are particularly important as they result in almost 50 % of the known genetic variations related to human inherited diseases [12].

Currently, computational methods are widely used for the analysis of possible candidates. The present study performed extensive screening for the most damaging nsSNPs in the *FOXP4* gene and identified the possible pathogenic nsSNPs. This study aimed to sort out the functionally important nsSNPs in the *FOXP4* gene. Additionally, MD simulation analysis was conducted to differentiate the stable and unstable nature of the native protein and nsSNPs.

Following these, the *FOXP4* gene was subjected to functional and structural analysis (multi-level way) to determine their pathogenicity. Briefly, this study involves the following steps: 1) Retrieving of nsSNPs and protein sequence, 2) Functional effect prediction based on sequence homology, 3) Stability prediction of nsSNPs on protein, 4) Prediction of FOXP4 nsSNPs effect on the secondary and tertiary structure using GOR IV, 5) Evolutionary conservation analysis using ConSurf Server, 6) Surface accessibility prediction of *FOXP4* gene nsSNPs, 7) Prediction of structural effects of FOXP4 mutants using HOPE Server, 8) Prediction of destabilizing effects upon nsSNPs of *FOXP4* gene, 9) Prediction of alterations in protein stability and interaction upon nsSNPs, 10) Interaction of FOXP4 with other Genes and Proteins and 11) Molecular dynamic simulation analysis.

This is the first inclusive in silico analysis that conceals the total nsSNPs of *FOXP4* gene. Henceforth, our findings could benefit the precision medicine discipline by acknowledging disease treatment based on these genomic changes.

2. Material and methods

2.1. Retrieving of nsSNPs and protein sequence

The National Center for Biotechnology Information, dbSNP database was used to retrieve the nsSNPs of the *FOXP4* gene (Accessed March 2024). Sequence of the wild type FOXP4 protein (Q8IVH2) was obtained from UniProtKB database. The nsSNPs could change the amino acid of any protein and cause disease pathogenesis [13]. Therefore, nsSNPs have been selected for downstream analysis. Here, numerous bioinformatics tools have been used to analyze the effect of nsSNPs on the corresponding proteins.

2.2. Functional effect prediction based on sequence homology

2.2.1. SIFT

The Sorting Intolerant from Tolerant (SIFT) is a bioinformatics web server that was used to identify deleterious nsSNPs from tolerated for corresponding protein. The higher the tolerance index, the lower the functional impact of amino acid substitution is likely to have. The nsSNPs having prediction score of ≥0.05, are considered as tolerated and the same having the score of ≤0.05 are regarded as deleterious. The rsIDs of the corresponding gene were submitted as input for the analysis [14].

2.2.2. PolyPhen2

The Polymorphism Phenotyping v2 (PolyPhen2) web server was used to analyze the nsSNPs which can affect protein functionality of the corresponding protein. A position-specific independent count (PSIC) score has been calculated for each variant. PolyPhen2 sorts

out the score into three categories, (1) benign (0.00–0.45), (2) possibly damaging (0.45–0.95), and 3) probably damaging (0.95–1). The nsSNPs having HumDiv and HumVar both similar scores were chosen for further analysis [15].

2.2.3. SNAP2

Screening for Deprecated Polymorphism v2 (SNAP2) is a bioinformatics web server that was used to predict the functional effect of sequence variants on protein. On the heat map, a score of +100 (dark red) or −100 (dark blue) indicates the amino acid substitution is pathogenic or neutral respectively. The prediction score ranges from 0 (very low reliability) to 9 (very high reliability) [16].

2.2.4. PhD SNP

The Predictor of Human Deleterious Single Nucleotide Polymorphisms (PhD SNP) web server utilized to predict disease-associated substitutions based on the support vector machine (SVM) method. For each mutation, the PhD SNP predicts an output score (from 0 to 1) representative probability of this nsSNP is associated with the disease. The method considers 0.5 as the threshold above which nsSNP is predicted to be associated with disease [17].

2.2.5. PANTHER

Protein Analysis Through Evolutionary Relationship (PANTHER) was utilized to assess the functional impact of amino acid substitutions on the biological function. This server is based on the Hidden Markov Model (HMM) to estimate the probability that nsSNPs variants can affect the phenotype of a protein according to its evolutionary origin [18].

2.3. Stability prediction of nsSNPs on protein

2.3.1. I-Mutant2.0

I-Mutant2.0 web server was utilized to estimate the potential effects of nsSNPs on the structural reliability of the protein and free energy change Delta Delta G (DDG). Accordingly, stability will increase when DDG is > 0 kcal/mol and decrease when DDG is < 0 kcal/mol. We have inputted the protein FASTA sequence, location AAS, and variant residue [19].

2.3.2. MUpro

The MUpro server was utilized to estimate the stability of mutated proteins due to the substitution of amino acids following both neural networks and support vector machines (SVM). The decrease in protein stability is predicted if the confidence score is < 0, and increases if the confidence score is > 0 [20].

2.4. Prediction of FOXP4 nsSNPs effect on secondary and tertiary structure using GOR IV

GOR IV is a web server, utilized to determine protein secondary structure upon nsSNPs. For each secondary structure at each amino acid location, it provides probability values. GOR IV employs all potential pair frequencies across a range of amino acid residues. GOR IV provides a sequence and predicted secondary structure in rows (including E = extended, H = helix, beta-strand, and C = coil) [21].

2.5. Evolutionary conservation analysis using ConSurf server

ConSurf server was utilized to address the evolutionary pattern of conservation at each amino acid site. By using the Bayesian calculation method, this server analyzes close sequence homologs and estimates the phylogenetic relationship. For each residue of the corresponding protein, the degree of conservation is estimated on a scale of 1–9 and categorized as 1) variable (1–4), 2) intermediate (5–6) and 3) conserved (7–9) [22].

2.6. Surface accessibility prediction of nsSNPs

The NetSurfP 2.0 server was utilized to determine the surface accessibility of the corresponding protein. This tool predicts surface accessibility, secondary structure, and structural disorder of any respective amino acid position by analyzing the amino acid sequence. Here, 25 % was set for relative surface accessibility (RSA) as the threshold. Greater than 25 % of RSA was considered as exposed [23].

2.7. Prediction of structural effects of FOXP4 mutants using HOPE server

Have (y)Our Protein Explained (HOPE) web server utilized to estimate the structural and functional effects of point mutations. It provides the 3D structural visualization of mutated proteins and gives the results by using the UniProt and DAS prediction servers. Inputted the protein sequence, wild-type, and new amino acids for the prediction [24].

2.8. Prediction of destabilizing effects upon nsSNPs of FOXP4 gene

2.8.1. mCSM

The mCSM web server is an artificial intelligence method created to anticipate the consequences of mutations in corresponding proteins. This is also taking into account the structural features of the proteins. Particularly, mCSM is employed to assess how missense

variations impact the stability of FOXP4 protein. This server predicts changes in the energy required for protein folding (ΔΔG) caused by mutations. This prediction sorts into two categories, either stabilizing or destabilizing which depend on their respective ΔΔG values.

2.8.2. SDM

SDM (Site-Directed Mutagenesis) web server is a knowledge-based SVM algorithm predictor. This server is used to predict the stability of the respective protein. This prediction is based on conformationally constrained, environment-specific substitution tables derived from amino acid replacements that are tolerated within protein families.

2.8.3. DUET

DUET is also a web server that was used to predict the effect of nsSNPs on the stability of the corresponding protein. It is an integrated computational approach and its calculation is based on secondary structure. Its utilized algorithms include SVM regression with a Radial Basis Function kernel and RSA. It helps to estimate the consequence of a single mutation on protein stability and its function. This web server is utilized for applying and considering incorporated mutations and their impact on the structure of proteins by computational methodology.

2.9. Prediction of alterations in protein stability and interaction upon nsSNPs

The DynaMut web server was used for prediction of the alterations in the stability of protein and interaction upon nsSNPs. This server performed prediction by evaluating flexibility analysis and protein dynamics. We performed a single mutation prediction feature in this analysis. A mutation list and the wild-type structure in PDB format were inputted [25].

2.10. Interaction of FOXP4 with other genes and proteins

2.10.1. GeneMANIA

GeneMANIA is a flexible and user-friendly web interface for generating hypotheses about gene function, analyzing gene lists and prioritizing genes for functional analyses. GeneMANIA finds related genes to input genes, using a very large set of functional association data. Association data include protein and genetic interactions, pathways, co-expression, co-localization and protein domain similarity. It was used to find new members of a pathway, additional genes [26].

2.10.2. STRING

The predicted version of protein-protein interaction information was analyzed using the STRING server. The protein-protein interaction outcomes are important to analyze since the mutated protein continuously affects the other protein during the diseased condition. This helps to study the mechanism of the diseased condition for targeting the source protein and other corresponding proteins [27].

2.11. Molecular dynamic simulation analysis

To evaluate the structural stability of the protein, a 50 ns MD simulation was performed using the "Desmond v6.3 Program" in
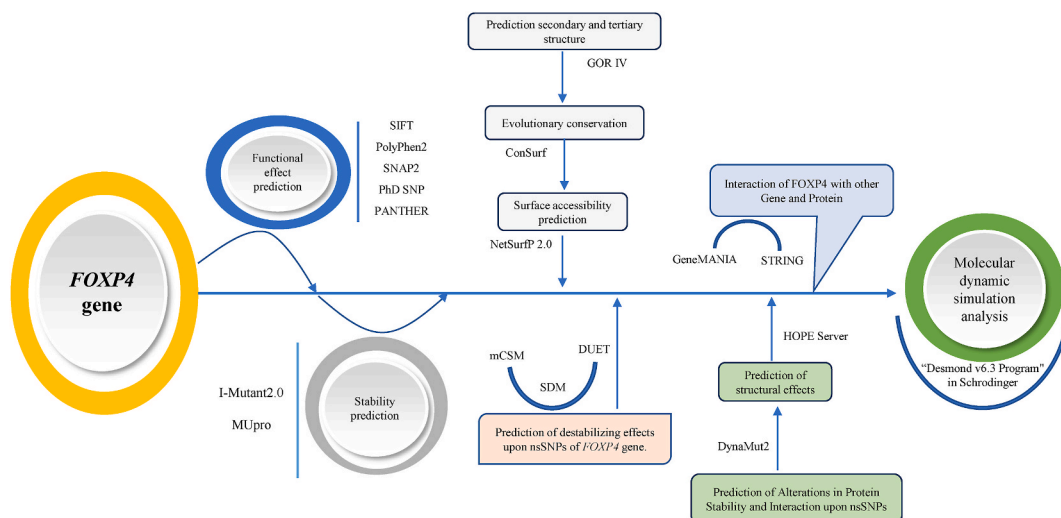

Fig. 1. The complete workflow of nsSNPs analysis.

Schrodinger 2020-3 under the Linux framework [28]. The simulation was examined by the three-site transferrable intermolecular potential (TIP3P) water model [29]. An orthorhombic box shape with a 10 Å distance from the center was used to maintain a specific volume, and Na^+ and Cl^- were added to neutralize the whole system with a salt concentration of 0.15 M. An OPLS3e force field was applied [30]. The protein structure system was further minimized using a natural time and pressure (NPT) ensemble at a constant pressure of 101325 Pa's and a temperature of 300 K. RMSD (root means square deviation), RMSF (root means square fluctuation), Rg (radius of gyration), and hydrogen bonds were analyzed to evaluate the stability and dynamic characteristics of the protein.

3. Results

The complete workflow is summarized in Fig. 1.

3.1. Retrieving of nsSNPs and protein sequence

The nsSNPs of the *FOXP4* gene were extracted from the dbSNP database. Here, about 23124 SNPs, where 555 missense (2.4 %), 20525 introns (88.76 %), 1114 non-coding transcript (4.82 %), and 334 synonymous (1.44 %) were obtained. Then the missense variants were further analyzed to explore the most deleterious.

3.2. Functional effect prediction based on sequence homology

Firstly, 555 missense nsSNPs were analyzed using SIFT, where 57 nsSNPs were found deleterious with score less than ≤0.05. Then, those nsSNPs were analyzed using PolyPhen2. During Polyphen2 analysis, HumDiv and HumVar both scores were computed. After SIFT and PolyPhen2 prediction 10 nsSNPs to be deleterious and probably/possibly damaging and benign, respectively. Finally, those nsSNPs were further analyzed using SNAP2, PhD SNP and PANTHER (Table 1). However, all 10 nsSNPs were observed as non-neutral in SNAP2 analysis. Following that, among all 10 nsSNPs, 5 and 6 SNPs were associated with disease when analyzed by PhD SNP, and PANTHER respectively. In all five analyses, the 10 nsSNPs (rs372762294, rs141899153, rs142575732, rs376938850, rs367607523, rs112517943, rs140387832, rs373949416, rs373949416 and rs376160648) were common and observed as deleterious, damaging and diseases associated (Table 1 and Supplementary Table 1).

3.3. Stability prediction of nsSNPs on protein

I-Mutant2.0 and MUpro server estimate the free energy change DDG value and the Confidence score of the 10 candidates nsSNPs mediated mutated proteins. Among these, a number of 7 nsSNPs guided proteins were found to be the most unstable considering DDG and Confidence score (Table 2).

3.4. Prediction of FOXP4 nsSNPs effect on secondary and tertiary structure using GOR IV

GOR IV tool predicts the effect of nsSNPs on protein secondary and tertiary structure. Following this analysis, the contents of the normal secondary and tertiary structure of the FOXP4 protein, which include 237 alpha helixes, 58 extended strands, and 385 random coils differed from the contents of structures with the seven amino acids substitution. The results showed that the amino acid substitution of the seven nsSNPs effects protein structure by altering the protein contents of alpha helixes, extended strands, and random coils (Table 3).

3.5. Evolutionary conservation analysis using ConSurf server

A colorimetric conservation score was created for FOXP4 protein by the ConSurf server. The program identifies highly conserved

Table 1
Functional effect prediction of non-synonymous SNPs using numerous bioinformatics tools.

SL	SNP ID	Amino acid change	SIFT	PolyPhen2		SNAP2	PhD SNP	PANTHER
				HumDiv	HumVar			
1	rs372762294	G19V	D	Pro. D	Pos. D	E	Dis	NA
2	rs141899153	Q156H	D	Pro. D	Pro. D	E	Dis	Dis
3	rs142575732	R288W	D	Pro. D	Pro. D	E	Dis	Dis
4	rs376938850	T330I	D	Pos. D	Benign	E	N	Dis
5	rs367607523	E353D	D	Pro. D	Pro. D	E	Dis	Dis
6	rs112517943	Q355H	D	Pos. D	Pos. D	E	N	Dis
7	rs140387832	R373W	D	Pro. D	Pos. D	E	Dis	Dis
8	rs373949416	D667H	D	Pos. D	Pos. D	E	N	NA
9	rs373949416	D669H	D	Pos. D	Benign	E	N	NA
10	rs376160648	L679P	D	Pos. D	Pos. D	E	N	NA

D: Deleterious, E: Effect N: Neutral, Dis: Disease, Pro. D: Probably damaging, Pos. D: Possibly damaging, NA: Unspecified.

Table 2
Stability prediction of nsSNPs on protein using I-Mutant2.0 and MUpro.

Sl	SNP ID	Amino Acid Change	I-Mutant2.0		MUpro	
			Stability	DDG (Kcal/mol)	Prediction	Confidence Score
1	rs141899153	Q156H	Decrease	−0.46	Decrease	−0.71089341
2	rs376938850	T330I	Decrease	−0.38	Decrease	−0.23065386
3	rs112517943	Q355H	Decrease	−1.35	Decrease	−0.64577427
4	rs140387832	R373W	Decrease	−0.14	Decrease	−0.78766219
5	rs373949416	D667H	Decrease	−0.6	Decrease	−0.74309817
6	rs373949416	D669H	Decrease	−0.77	Decrease	−1
7	rs376160648	L679P	Decrease	−1.13	Decrease	−1

structural and functional amino acid regions that are essential for biological function. From the analysis, we have found that Q156H, T330I, Q355H, R373W, D667H, D669H, L679P all are exposed residue (Fig. 2A).

3.6. Surface accessibility prediction of nsSNPs

The NetSurfP 2.0 predicted the surface accessibility of individual amino acids of FOXP4 protein by a percentage score. The threshold was 25 %, which means the amino acid residues having a score of more than 25 % were predicted as exposed, and scores less than 25 % were buried. Among the 7 positions having all positions (Q156H, T330I, Q355H, R373W, D667H, D669H, L679P) scored more than 25 % and it indicate these residues are exposed. (Fig. 2B and Supplementary Table 2).

3.7. Prediction of structural effects of FOXP4 mutants using HOPE server

HOPE server guides how mutation affects the structural variation of a protein based on size, charge, hydrophobicity, and spatial structure compared to the wild type. Among the 7 predicted substitutions, 6 mutants (Q156H, T330I, Q355H, R373W, D667H, and D669H) were found to be bigger than the wild residue's consequences localization on the surface of the wild protein. However, L679P mutant residues were smaller than the wild residues, which caused a space in the core of the wild protein and loss of hydrophobic interactions. Therefore, mutation of the residue can interrupt inter/intra-molecular interactions of the protein (Table 4).

3.8. Prediction of destabilizing effects upon nsSNPs of FOXP4 gene

The mCSM anticipates the consequences of nsSNPs of the corresponding protein and considers the structural features of the proteins. This is employed to assess how the mutation impacts the stability of FOXP4 protein. Among 7 nsSNPs, 5 nsSNPs (Q156H, T330I, Q355H, R373W and L679P) are found destabilizing having the ΔΔG score <0. Consequently, the SDM server predicted the stability of nsSNPs of the corresponding protein and found upstream 5 nsSNPs as reduced stability. Lastly, the DUET web server also predicts the effect of nsSNPs on the stability of the corresponding protein. Here, 4 nsSNPs (Q156H, Q355H, R373W and L679P) are found to destabilize the corresponding protein having the ΔΔG score <0 (Table 5).

3.9. Prediction of alterations in protein stability and interaction upon nsSNPs

The DynaMut2 server was used to calculate general dynamic traits of the highest deleterious nsSNPs selected from the previous analysis steps. DynaMut2 represented the predictions for ΔΔG stability value among the WT and mutant protein. From 7 nsSNPs, 5 nsSNPs showed a decrease in the ΔΔG Stability value compared to the wild-type and were found to be responsible for destabilizing the

Table 3
Prediction of FOXP4 nsSNPs effect on secondary and tertiary protein structure using GOR IV.

| Parameter	Normal protein	Q156H	T330I	Q355H	R373W	D667H	D669H	L679P
Alpha helix	237 (34.85 %)	237 (34.85 %)	241 (35.44 %)	237 (34.85 %)	236 (34.71 %)	236 (34.71 %)	236 (34.71 %)	236 (34.71 %)
310 helix	0	0	0	0	0	0	0	0
Pi helix	0	0	0	0	0	0	0	0
Beta bridge	0	0	0	0	0	0	0	0
Extended strand	58 (8.53 %)	58 (8.53 %)	58 (8.53 %)	58 (8.53 %)	58 (8.53 %)	58 (8.53 %)	58 (8.53 %)	58 (8.53 %)
Beta turn	0	0	0	0	0	0	0	0
Bend region	0	0	0	0	0	0	0	0
Random coil	385 (56.62 %)	385 (56.62 %)	381 (56.03 %)	385 (56.62 %)	386 (56.76 %)	386 (56.76 %)	386 (56.76 %)	386 (56.76 %)
Ambiguous states	0	0	0	0	0	0	0	0
Other states	0	0	0	0	0	0	0	0
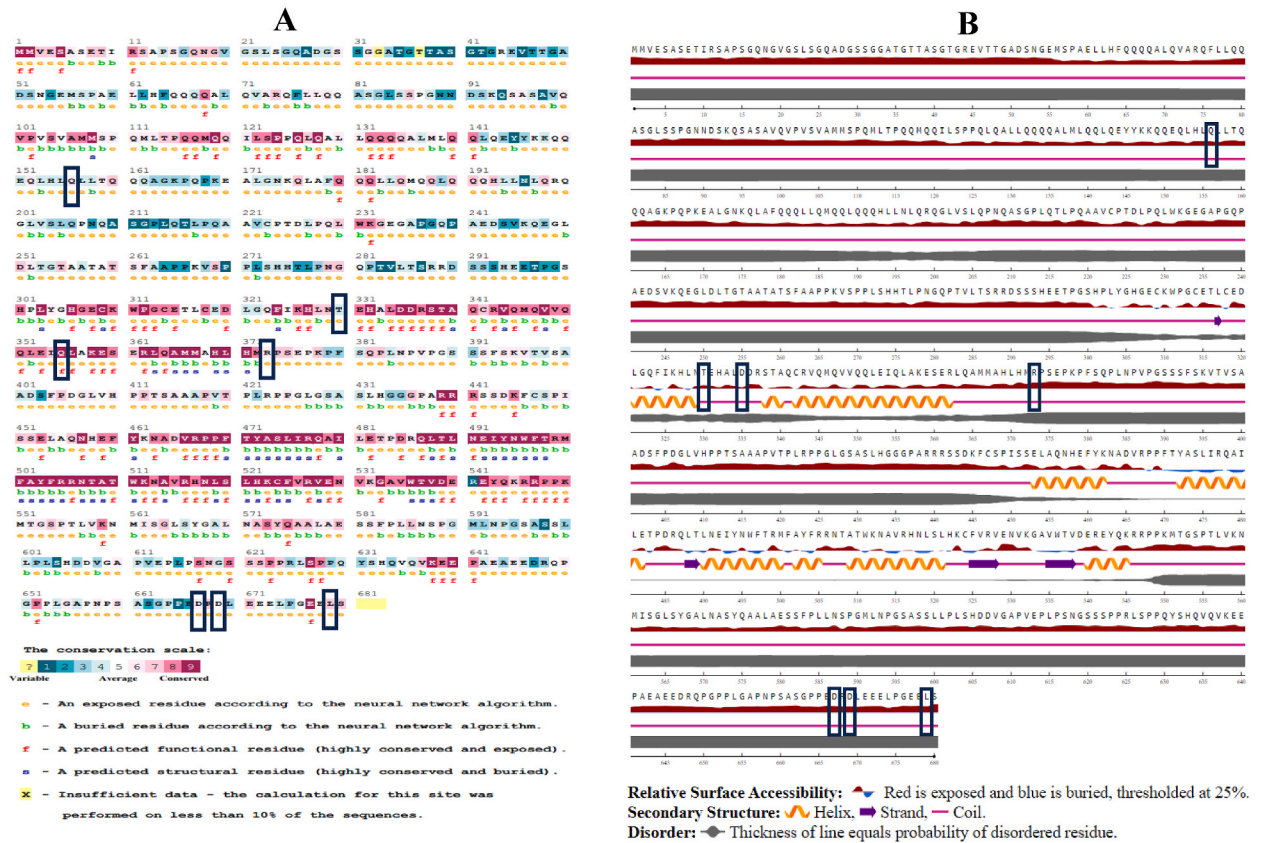
Fig. 2. (A) Prediction of evolutionary conserved amino acid residues by ConSurf server Conservation score is represented as the color coding bars. (B) Surface accessibility prediction of nsSNPs of the FOXP4 protein by NetSurfP 2.0. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

protein. The free energy ($\Delta\Delta G$) Stability values for Q156H, T330I, Q355H, R373W and L679P mutants were found −0.35, −0.030, −0.56, −0.4 and −1.530 kcal/mol respectively (Table 6).

3.10. Interaction of FOXP4 with other genes and proteins

3.10.1. GeneMANIA

Gene-gene interaction network revealed a correlation among genes for the *FOXP4* gene. A composite gene-gene functional interaction network has been built by GeneMANIA. GeneMANIA results showed that the *FOXP4* gene was associated with 20 other genes. It interacts more significantly with *FOXP1, 2* genes, *MKI67* (marker of proliferation Ki-67), *GATAD2B* (GATA zinc finger domain containing 2B), and others genes (Fig. 3A, Supplementary Table 3).
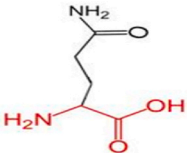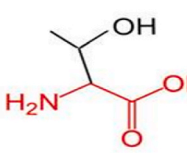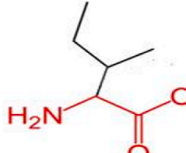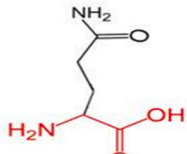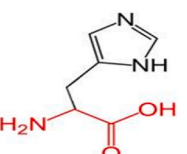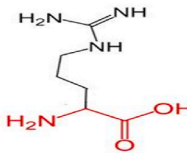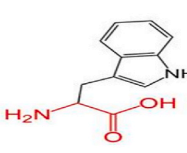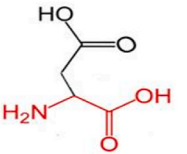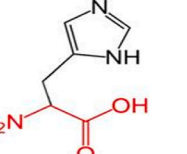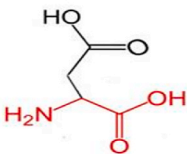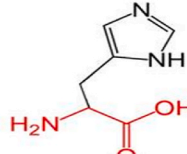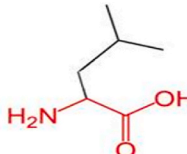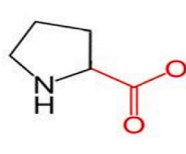
3.10.2. STRING

Proteins work in a way that synergizes and interacts with other proteins to carry out cell signaling and other cellular functions. Therefore, amino acid variations in one protein can affect other proteins in the network. FOXP4 functional partners as predicted using the STRING tool. Forkhead box protein P4; Transcriptional inhibitors inhibit lung-specific expression. The nasal tip box protein P2 (FOXP2) plays a role in the specification and differentiation of the lung epithelium. Heart and neural crest derivatives-expressed protein 1 (HAND1) is a transcription factor that plays an essential role in both trophoblast giant cell differentiation and in cardiac morphogenesis (By similarity). Forkhead box protein P3 (FOXP3), C-terminally processed; Transcriptional regulators are important for the development and repressor function of regulatory T (Treg) cells (Fig. 3B and Supplementary Table 4) [31].

3.11. Molecular dynamic simulation analysis

A 50 ns MD simulation analysis of the predicted deleterious mutants Q156H, T330I, Q355H, R373W, and L679P and FOXP4 native protein was conducted to observe the deviation from the structural stability of native protein along with mutant protein in an artificial environment. Mutants Q156H, T330I, Q355H, R373W, and L679P showed considerable fluctuations after the simulation run to the native protein FOXP4, as shown in Fig. 4A. The RMSD values for the native protein were observed ranging from 4.12 Å to 37.79 Å while

Table 4
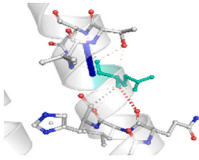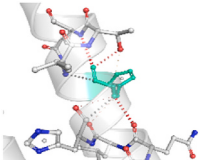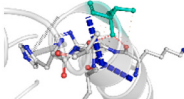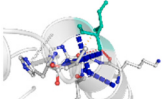Structural effects prediction of FOXP4 nsSNPs using Project HOPE server.
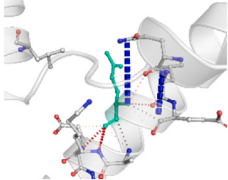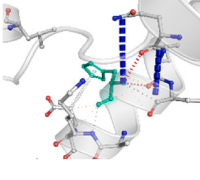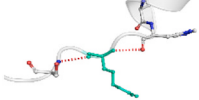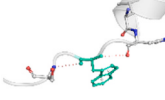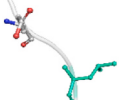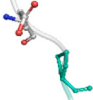
Sl	SNP ID	Amino acid change	Wild Type	Mutant
1	rs141899153	Q156H		
2	rs376938850	T330I		
3	rs112517943	Q355H		
4	rs140387832	R373W		
5	rs373949416	D667H		
6	rs373949416	D669H		
7	rs376160648	L679P		

4.31 Å to 37.41 Å for the Q156H mutant type, 4.18 Å to 27.83 Å for the T330I mutant type, 4.09 Å to 32.87 Å for the Q355H, 4.02 Å to 30.14 Å for the R373W and 4.04 Å to 29.54 Å for the L679P mutant type as shown in Fig. 4A. The average RMSD values of native FOXP4, and mutant Q156H, T330I, Q355H, R373W, and L679P are 31.48 Å, 30.73 Å, 24.80 Å, 27.92 Å, 26.42 Å, and 26.15 Å, respectively. The fluctuations of mutant proteins were more noticeable than wild-type proteins throughout the simulation run. Overall, the RMSD plot showed that the native protein was stable along the simulation run time and also demonstrated that the mutant protein has a major impact on the structural confirmation of FOXP4 protein. The RMSF of protein residue was counted to observe how these changes impact the local flexibility. In addition, RMSF analysis demonstrated considerable fluctuations between native and mutant structures throughout the simulation run. From the RMSF plot, the native protein and five mutated proteins specified by proteins Q156H, T330I, Q355H, R373W, and L679P had the highest peak fluctuation with (PRO168, ASP251, SER381, PHE404, SER627, and

Table 5
Prediction of destabilizing effects upon nsSNPs of *FOXP4* gene.

Sl	SNP ID	Amino Acid Change	mCSM		SDM		DUET	
			Prediction	ΔΔG (kcal/mol)	Prediction	ΔΔG (kcal/mol)	Prediction	ΔΔG (kcal/mol)
1	rs141899153	Q156H	Destabilizing	−0.745	Reduced stability	−0.79	Destabilizing	−0.705
2	rs376938850	T330I	Destabilizing	−0.048	Reduced stability	−1.5	Stabilizing	0.434
3	rs112517943	Q355H	Destabilizing	−0.791	Reduced stability	−0.77	Destabilizing	−0.648
4	rs140387832	R373W	Destabilizing	−0.495	Reduced stability	−0.15	Destabilizing	−0.477
5	rs373949416	D667H	Stabilizing	0.438	Stabilizing	0.56	Stabilizing	0.584
6	rs373949416	D669H	Stabilizing	0.447	Stabilizing	0.8	Stabilizing	0.643
7	rs376160648	L679P	Destabilizing	−0.593	Reduced stability	−1.44	Destabilizing	−0.61

Table 6
Prediction of alterations in protein stability and interaction upon nsSNPs using DynaMut2 server.

SL	Amino acid variants	ΔΔG ENCoM	Prediction	Wild type	Mutant type
01	Q156H	−0.35 kcal/mol	Destabilizing		
02	T330I	−0.030 kcal/mol	Destabilizing		
03	Q355H	−0.56 kcal/mol	Destabilizing		
04	R373W	−0.4 kcal/mol	Destabilizing		
05	L679P	−1.530 kcal/mol	Destabilizing		

ALA657), (LYS232, PRO271, SER390, GLY428, ALA464, ARG506, and GLU666), (THR35, ALA98, ALA163, GLN206, GLY212, PRO266, and LEU422), (GL235, PRO282, SER392, GLY564, SER588, SER604, PRO657, GLU672, and SER680), (LYS267, SER598, LEU601, LEU603, ASP669, LEU679, and SER680), and (THR216, ASP243, THR297, PRO425, LEU600, ASP607, PRO664, andLEU670) respectively. The average RMSF values are 11.08 Å, 10.65 Å, 7.18 Å, 8.7 Å, 7.85 Å, and 8.9 Å correspondingly (Fig. 4B). In a nutshell, the mutant form of the FOXP4 protein showed higher RMSF than the native FOXP4 protein. The Rg measures how the atoms are distributed around the axis of a protein. It is an important indicator for predicting the structural activity of macromolecules and for assessing changes in the compactness of the protein structure. The native protein and five mutants (Q156H, T330I, Q355H, R373W, and L679P) revealed Rg ranges of 35.70 Å to 45.682 Å, 31.21 Å to 45.65 Å, 34.92 Å to 45.98 Å, 31.21 Å to 45.65 Å, 31.46 Å to 45.74 Å and 31.07 Å to 45.732 Å, respectively. Average fluctuation of these SNPs was 39.48 Å, 36.48 Å, 37.08 Å, 36.48 Å, 34.24 Å and 34.91 Å, respectively (Fig. 4C). The mutated protein structures were unstable in 50 ns simulations with greater fluctuation differences from lowest to highest, suggesting that the mutated proteins significantly alter the conformation of the respective protein's active site. The number of hydrogen bonds can help to characterize a protein. Therefore, hydrogen bond numbers were calculated from initial to final times during the 50 ns simulation run to observe each hydrogen bond. All the proteins formed several hydrogen bonds ranging between 300 and 500 occur simultaneously until 50 ns simulation time (Fig. 4D). These results indicated that the native protein has a higher specificity while the mutated protein has a higher fluctuating range of hydrogen bonds denoted the structural instability.
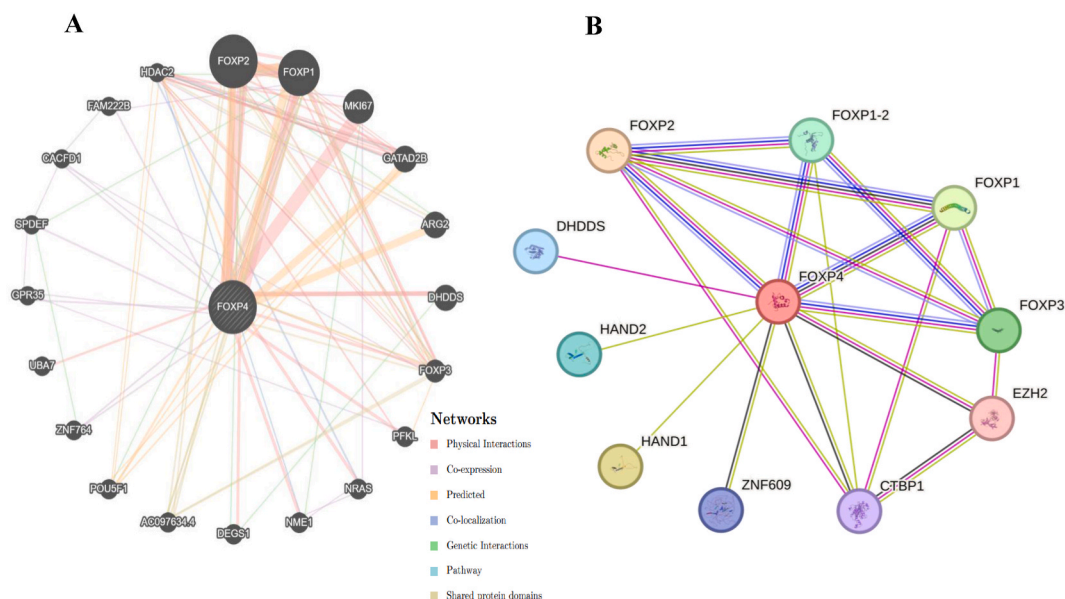
Fig. 3. Interaction of FOXP4 with other genes and proteins. (A) Gene-gene interaction network of *FOXP4* gene predicted by GeneMANIA. (B) Protein-protein interaction of *FOXP4* gene by STRING tool.
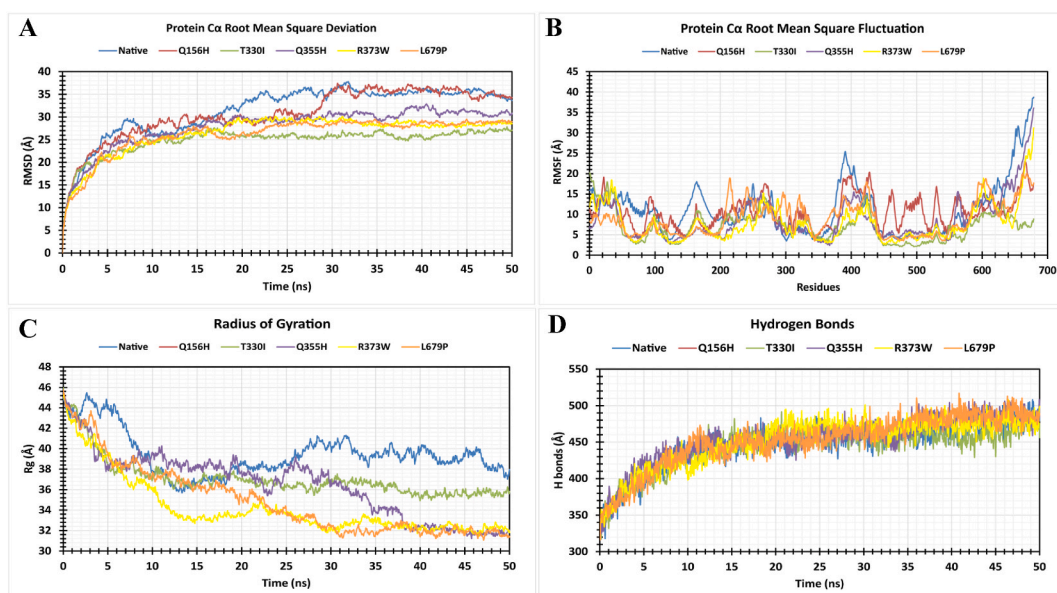
Fig. 4. Results of 50 ns molecular dynamics simulation by Schrodinger. (A) RMSD values of native and mutant structures. (B) RMSF values of native and mutant structures. (C) Calculation of Rg over the entire simulation. (D) Calculation of H-bonds represented as a time-depend. The color scheme is as follows: native (blue), Q156H mutant (red), T330I mutant (green), Q355H mutant (violet), R373W mutant (yellow), and L679P mutant (orange). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

4. Discussion

Forkhead box protein 4 (FOXP4) is a protein-coding gene. This gene belongs to the P subfamily of the Forkhead Box (FOX) transcription factor family [32]. It was first identified in the Drosophila forkhead gene [33] and the mammalian Hnf3 transcription factors were found to contain highly similar domains [34]. The FOXP subgroup of transcription factors includes four different proteins: FOXP1, FOXP2, FOXP3 and FOXP4. The FOXP subfamily is perhaps the most widely characterized subfamily of the fork superfamily, playing important roles in vertebrate development and homeostasis [35–37]. Mutation in the *FOXP4* gene has previously been reported in children with growth retardation, laryngeal hypoplasia, feeding difficulties, and ventricular septal defect [38]. SNPs play an

important role in gene and protein-regulated diseases in humans [39]. Several in silico studies of polymorphisms to predict nsSNPs as harmful or neutral have been performed [40–43].

In this study, we used *in-silico* structural and functional analyses to identify potential nsSNPs utilizing numerous computational bioinformatics tools to predict the deleterious SNPs. Initially, the SNPs of the *FOXP4* gene were extracted from the dbSNP database and a total 23124 SNPs were found, where 555 were nonsynonymous, 20525 intron, 1114 non-coding transcript, and 334 synonymous were obtained and the rest were unspecified. Then, to classify the most deleterious nsSNPs of the corresponding gene, five different prediction servers were used; SIFT, Polyphen2, SNAP2, PhD SNP and PANTHER. In all five analyses, the 10 nsSNPs (G19V, Q156H, R288W, T330I, E353D, Q355H, R373W, D667H, D669H and L679P) were common and observed as deleterious, damaging and diseases associated (Table 1). Then, the nsSNP was observed as a decrease in protein stability using the I-Mutant2.0 and MUpro. Among these, a number of 7 nsSNPs guided proteins were found to be the most unstable considering ΔΔG and Confidence score (Table 2). GOR IV tool was used to predict the effect of nsSNPs on protein secondary and tertiary structure. Following this analysis, the contents of the normal secondary and tertiary structure of the FOXP4 protein, which include 237 alpha helixes, 58 extended strands, and 385 random coils differed from the contents of structures with the seven amino acids substitutions (Table 3). Mutations affect the helix-helix interactions and disordered regions of corresponding proteins by changing their properties [44]. Hence, mutations at these positions might alter the structural and functional properties of the mutant protein. A colorimetric conservation score was created for FOXP4 protein by the ConSurf server. From the analysis, we have found that Q156H, T330I, Q355H, R373W, D667H, D669H, L679P all are exposed residues. Among them Q355H residues were found as highly conserved for playing potential biological functions (Fig. 2A). The conserved residue encounters mutation, it is more destructive compared to the less conserved one [45].

Following NetSurfP 2.0 prediction, Among the 7 positions having all positions (Q156H, T330I, Q355H, R373W, D667H, D669H, L679P) scored more than 25 % which means all are exposed (Fig. 2B and Supplementary Table 2). NetSurfP tool-based surface accessibility analysis revealed that these SNPs have potential to change surface accessibility and secondary structures following changing coil and helix structures. Significant changes in absolute surface accessibility in mutant protein might alter binding affinity of the corresponding mutant protein. HOPE provides 3D structural images of mutated proteins. Among the 7 predicted substitutions, Q156H, T330I, Q355H, R373W, D667H, and D669H mutants were found to be bigger than the wild residue's consequences localization on the surface of the wild protein. However, L679P mutant residues were smaller than the wild residues (Table 4). This space could influence the protein function, characteristics or reactivity [46]. The mCSM, SDM and DUET web server predicts the effect of nsSNPs on the stability of the corresponding protein. Following this analysis, Q156H, T330I, Q355H, R373W and L679P nsSNPs are found to destabilize the corresponding protein having the ΔΔG score <0 (Table 5). Those point mutations might play a role in disease progression. DynaMut2 tools provide Gibbs free energy change (ΔΔG) values for each protein structure; The free energy change during the unfolding of the kinetically stable protein is described by this value ΔΔG. From 7 mutants, 6 mutants showed a decrease in the ΔΔG Stability value compared to the wild-type and were found to be responsible for destabilizing the protein (Table 5). Small changes in proteins can have large phenotypic outcomes. Changes in Gibbs Free energy of folding can occur due to a myriad of factors related to incorporating these properties. Those point mutations might play a role in disease progression and pathogenesis [25]. GeneMANIA analysis revealed that *FOXP4* gene significantly interacts with 20 different genes [47–66]. *FOXP4* gene showed more significant interactions with *FOXP1, 2* genes, *MKI67* (marker of proliferation Ki-67), *GATAD2B* (GATA zinc finger domain containing 2B) than others genes. Here, *FOXP1* promotes cancer stem cell [47], and *FOXP2* is relevant to the human ability to develop language [48]. *MKI67* gene plays a vital role in the tumor microenvironment (TME) and congenital immunity [49]. ATP-dependent chromatin remodeling and histone deacetylation [50]. *FOXP4* gene promotes tumor formation and progression [51]. Hence, nsSNPs in *FOXP4* gene interacting with these genes might result serious health hazards. STRING tool mediated analysis resulted that FOXP4 is tightly interacting with 10 proteins [47,48,51,53,54,67–71]. Among 10 proteins, five were common interacting proteins (FOXP2, FOXP1, FOXP3, FOXP1-2 and DHDDS) in STRING and GeneMANIA analyses indicate that FOXP4 precisely interacted with the mentioned genes (Supplementary Table 3 and Table 4). An aqueous environment allows proteins to be dynamic entities. The MD simulation revealed how nsSNPs affect the stability, residual fluctuation, and compactness of the corresponding protein. We conducted a 50 ns MD simulation to examine the effect of mutation on FOXP4 protein dynamics. The protein structure is stabilized when the RMSD and RMSF values of a protein are within 1–3 Å [72]. RMSD of the Q156H, T330I, Q355H, R373W, and L679P mutants fluctuated consistently indicating that all of these mutants are likely to form an unstable structure in the physiological environment exhibited in Fig. 4A. Furthermore, the higher average RMSD value of native protein FOXP4 and mutant Q156H, T330I, Q355H, R373W, and L679P indicates a loss of stability for the mutants. According to the RMSF analysis, all mutant has the ability to modify the overall and local flexibility of corresponding protein, which affects stability and may therefore have an impact on function or interactions with other proteins (Fig. 4B). Compared to the WT, a higher Rg value represents the disassociation of the respective protein (Fig. 4C). However, the increased number of hydrogen bonds causes the structural unsteadiness [72]. Therefore, continuous fluctuations in RMSD and RMSF values, high Rg value and hydrogen bonds in mutant proteins compared to WT explore the mechanism by which these nsSNPs alter the structure and function of the native FOXP4 protein.

In the Chinese Population, rs1983891 of the *FOXP4* gene associated with Prostate Cancer [73]. In Multi-Ethnic Study of Atherosclerosis Study, rs2894439 of *FOXP4* gene is associated with interstitial lung disease [74]. A study in 3173 Japanese patients has found rs2495239 of *FOXP4* gene were newly identified as loci associated with EGFR mutation-positive lung adenocarcinoma [75]. A study on Japan and five other Asian countries, revealed that the rs7741164 of *FOXP4* gene is associated with a higher risk of lung cancer risk among never-smoking Asian females [76]. For instance, rs2494938 was associated with lung, non-cardia gastric and esophageal squamous cell carcinoma (ESCC) in Han Chinese [77], rs10484761 was associated with ESCC in a GWAS in Chinese [78] and rs1983891 in the intron of FOXP4 was associated with prostate cancer in Japanese [79]. Another study also found that rs9367106 of FOXP4 associated with both long COVID and severe COVID-19 [80].

However, the nsSNPs (rs141899153, rs376938850, rs112517943, rs140387832 and rs376160648) (Q156H, T330I, Q355H, R373W and L679P) respectively have not yet been studied, so they need to be studied with special consideration to the ethnic diversity of the different populations easily suffer from different diseases. The present study predicted high-risk nsSNPs in *FOXP4* gene which might be the important candidates in the pathology and disease progression. MD simulation analysis, continuous fluctuations in RMSD and RMSF values, Rg and hydrogen bonds also confirmed that these five nsSNPs are associated with alter the stability and structural flexibility of corresponding protein. However, further in vitro and in vivo studies are needed to investigate and establish the role of these nsSNPs in various diseases.

5. Conclusion

In this prediction and assessment of deleterious and disease causing nsSNPs study, we performed a series of computational bio-informatics prediction tools analysis for the *FOXP4* gene. Among the 555 nsSNPs, it is figured out that 6 nsSNPs are the most dele-terious, diseases associated, decrease stability and altered stability and interaction of the corresponding protein, which is predicted using numerous prediction tools. Hence, we concluded that 6 nsSNPs (Q156H, T330I, Q355H, R373W, D667H, and L679P) might be important candidates in the pathology and disease progression. Following DynaMut2, 5 nsSNPs showed a decrease in the $\Delta\Delta G$ value compared with the wild-type and were found to be responsible for destabilizing the corresponding protein. Finally, molecular dy-namics simulation analysis revealed continuous fluctuations in RMSD and RMSF values, Rg and hydrogen bonds in mutant proteins (Q156H, T330I, Q355H, R373W, and L679P) compared with WT, which might alter the functional and structural stability of the corresponding protein. Therefore, the key findings of the present study provide a scientific overview for the researchers to determine the significant role in these nsSNPs of these corresponding proteins. The effect of these nsSNPs in the *FOXP4* gene on protein structure and function needs to be confirmed by experimental investigations. Multiple in-silico approaches are used to provide quick and inexpensive screening that may help direct additional laboratory studies.

Funding statement

Data availability statement

Data included in article/supplementary material/referenced in article.

CRediT authorship contribution statement

Md. Mostafa Kamal: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing – original draft. **Shamiha Tabassum Teeya:** Formal analysis, Methodology, Writing – original draft. **Md. Mahfuzur Rahman:** Writing – review & editing. **Md. Enamul Kabir Talukder:** Formal analysis, Writing – original draft. **Sonia Sarmin:** Writing – review & editing. **Tanveer A. Wani:** Formal analysis, Resources. **Md. Mahmudul Hasan:** Conceptualization, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e32791.

References

[1] P. Xue, S. Huang, X. Han, C. Zhang, L. Yang, W. Xiao, J. Fu, H. Li, Y. Zhou, Exosomal miR-101-3p and miR-423-5p inhibit medulloblastoma tumorigenesis through targeting FOXP4 and EZH2, Cell Death & Differentiation 2021 29 (1 29) (2021) 82–95, https://doi.org/10.1038/s41418-021-00838-4.

[2] M. Liu, X. Shi, J. Wang, Y. Xu, D. Wei, Y. Zhang, K. Yang, X. Wang, S. Liang, X. Chen, F. Yang, L. Sun, X. Zhu, C. Zhao, L. Zhu, L. Tang, C. Zheng, Z. Yang, Association of FOXP4 gene with prostate cancer and the cumulative effects of rs4714476 and 8q24 in Chinese men, Clin. Lab. 10 (2015), https://doi.org/10.7754/Clin.Lab.2015.150313.

[3] J.H. Kim, J. Hwang, J.H. Jung, H.J. Lee, D.Y. Lee, S.H. Kim, Molecular networks of FOXP family: dual biologic functions, interplay with other molecules and clinical implications in cancer progression, Molecular Cancer 2019 18 (1 18) (2019) 1–19, https://doi.org/10.1186/S12943-019-1110-3.

[4] B.C. Jackson, C. Carpenter, D.W. Nebert, V. Vasiliou, Update of human and mouse forkhead box (Fox) gene families, Hum. Genom. 4 (2010) 345–352, https://doi.org/10.1186/1479-7364-4-5-345/FIGURES/2.

[5] M.E. Santos, A. Athanasiadis, A.B. Leitão, L. Dupasquier, É. Sucena, Alternative splicing and gene duplication in the evolution of the FoxP gene subfamily, Mol. Biol. Evol. 28 (2011) 237–247, https://doi.org/10.1093/MOLBEV/MSQ182.

[6] Y. Niu, G. Wang, Y. Li, W. Guo, Y. Guo, Z. Dong, LncRNA FOXP4-AS1 promotes the progression of esophageal squamous cell carcinoma by interacting with MLL2/H3K4me3 to upregulate FOXP4, Front. Oncol. 11 (2021) 773864, https://doi.org/10.3389/FONC.2021.773864/BIBTEX.

[7] T. Ma, J. Zhang, Upregulation of FOXP4 in breast cancer promotes migration and invasion through facilitating EMT, Cancer Manag. Res. 11 (2019) 2783–2793, https://doi.org/10.2147/CMAR.S191641.

[8] G. Zhang, G. Zhang, Upregulation of foxP4 in HCC promotes migration and invasion through regulation of EMT, Oncol. Lett. 17 (2019) 3944–3951, https://doi.org/10.3892/ol.2019.10049.

[9] J.N. Gaaib, Prediction of deleterious non-synonymous single nucleotide polymorphisms (Nssnps) of human TLR7 gene, Gaaib Iraqi Journal of Science 63 (2022) 2444–2452, https://doi.org/10.24996/ijs.2022.63.6.11.

[10] T. Ahmad, M.A. Valentovic, G.O. Rankin, Effects of cytochrome P450 single nucleotide polymorphisms on methadone metabolism and pharmacodynamics, Biochem. Pharmacol. 153 (2018) 196–204, https://doi.org/10.1016/J.BCP.2018.02.020.

[11] J.E. Lee, J.H. Choi, J.H. Lee, M.G. Lee, Gene SNPs and mutations in clinical genetic testing: haplotype-based testing and analysis, Mutat. Res. Fund Mol. Mech. Mutagen 573 (2005) 195–204, https://doi.org/10.1016/J.MRFMMM.2004.08.018.

[12] M. Krawczak, E. V Ball, I. Fenton, P.D. Stenson, S. Abeysinghe, N. Thomas, D.N. Cooper, Human gene mutation database—a biomedical information and research resource, Hum. Mutat. 15 (2000) 45–51, https://doi.org/10.1002/(SICI)1098-1004(200001)15:1.

[13] J. Thusberg, M. Vihinen, Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods, Hum. Mutat. 30 (2009) 703–714, https://doi.org/10.1002/HUMU.20938.

[14] N.L. Sim, P. Kumar, J. Hu, S. Henikoff, G. Schneider, P.C. Ng, SIFT web server: predicting effects of amino acid substitutions on proteins, Nucleic Acids Res. 40 (2012), https://doi.org/10.1093/NAR/GKS539.

[15] I. Adzhubei, D.M. Jordan, S.R. Sunyaev, Predicting functional effect of human missense mutations using PolyPhen-2, Curr Protoc Hum Genet Chapter 7 (2013), https://doi.org/10.1002/0471142905.HG0720S76.

[16] M. Hecht, Y. Bromberg, B. Rost, Better prediction of functional effects for sequence variants, BMC Genom. 16 (2015) 1–12, https://doi.org/10.1186/1471-2164-16-S8-S1/FIGURES/4.

[17] E. Capriotti, P. Fariselli, PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants, Nucleic Acids Res. 45 (2017) W247–W252, https://doi.org/10.1093/NAR/GKX369.

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

[19] E. Capriotti, P. Fariselli, R. Casadio, I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure, Nucleic Acids Res. 33 (2005), https://doi.org/10.1093/NAR/GKI375.

[20] C.L. Worth, R. Preissner, T.L. Blundell, SDM—a server for predicting effects of mutations on protein stability and malfunction, Nucleic Acids Res. 39 (2011) W215, https://doi.org/10.1093/NAR/GKR363.

[21] J. Garnier, J.F. Gibrat, B. Robson, [32] GOR method for predicting protein secondary structure from amino acid sequence, Methods Enzymol. 266 (1996) 540–553, https://doi.org/10.1016/S0076-6879(96)66034-0.

[22] H. Ashkenazy, E. Erez, E. Martz, T. Pupko, N. Ben-Tal, ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids, Nucleic Acids Res. 38 (2010), https://doi.org/10.1093/NAR/GKQ399.

[23] M.S. Klausen, M.C. Jespersen, H. Nielsen, K.K. Jensen, V.I. Jurtz, C.K. Sønderby, M.O.A. Sommer, O. Winther, M. Nielsen, B. Petersen, P. Marcatili, NetSurfP-2.0: improved prediction of protein structural features by integrated deep learning, Proteins: Struct., Funct., Bioinf. 87 (2019) 520–527, https://doi.org/10.1002/PROT.25674.

[24] H. Venselaar, T.A.H. te Beek, R.K.P. Kuipers, M.L. Hekkelman, G. Vriend, Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces, BMC Bioinf. 11 (2010) 1–10, https://doi.org/10.1186/1471-2105-11-548/FIGURES/5.

[25] C.H.M. Rodrigues, D.E.V. Pires, D.B. Ascher, DynaMut2: assessing changes in stability and flexibility upon single and multiple point missense mutations, Protein Sci. 30 (2021) 60–69, https://doi.org/10.1002/PRO.3942.

[26] K. Zuberi, M. Franz, H. Rodriguez, C.T. Lopes, G.D. Bader, Q. Morris, GeneMANIA prediction server 2013 update, Nucleic Acids Res. 41 (2013) W115–W122, https://doi.org/10.1093/NAR/GKT533.

[27] D. Szklarczyk, R. Kirsch, M. Koutrouli, K. Nastou, F. Mehryary, R. Hachilif, A.L. Gable, T. Fang, N.T. Doncheva, S. Pyysalo, P. Bork, L.J. Jensen, C. Von Mering, The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest, Nucleic Acids Res. 51 (2023) D638–D646, https://doi.org/10.1093/NAR/GKAC1000.

[28] R.R. Imon, M.E. Kabir Talukder, S. Akhter, M.S. Islam, F. Ahammad, K.M. Anis-Ul-Haque, M. Moniruzzaman, M. Afroze, M. Khan, M.A. Hena Mostofa Jamal, T. A. Wani, M.J. Uddin, M.M. Rahman, Natural defense against multi-drug resistant Pseudomonas aeruginosa: Cassia occidentalis L. in vitro and in silico antibacterial activity, RSC Adv. 13 (2023) 28773–28784, https://doi.org/10.1039/d3ra03923d.

[29] P. Mark, L. Nilsson, Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K, J. Phys. Chem. A 105 (2001) 9954–9960, https://doi.org/10.1021/jp003020w.

[30] K. Roos, C. Wu, W. Damm, M. Reboul, J.M. Stevenson, C. Lu, M.K. Dahlgren, S. Mondal, W. Chen, L. Wang, R. Abel, R.A. Friesner, E.D. Harder, OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules, J. Chem. Theor. Comput. 15 (2019) 1863–1874, https://doi.org/10.1021/acs.jctc.8b01026.

[31] M.G. Rabby, M.H. Rahman, M.N. Islam, M.M. Kamal, M. Biswas, M. Bonny, M.M. Hasan, In silico identification and functional prediction of differentially expressed genes in South Asian populations associated with type 2 diabetes, PLoS One 18 (2023) e0294399, https://doi.org/10.1371/JOURNAL.PONE.0294399.

[32] A. Teufel, E.A. Wong, M. Mukhopadhyay, N. Malik, H. Westphal, FoxP4, a novel forkhead transcription factor, Biochim. Biophys. Acta Gene Struct. Expr. 1627 (2003) 147–152, https://doi.org/10.1016/S0167-4781(03)00074-5.

[33] D. Weigel, Gerd, J. Irgens, F. Kiittner, E. Seifert, H. Jlckle, The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo, Cell 57 (1969) 645–658.

[34] E. Lai, V.R. Prezioso, E. Smith, O. Litvin, R.H. Costa, J.E. Darnell, HNF-3A, a Hepatocyte-Enriched Transcnpuon Factor of Novel Structure is Regulated Transcriptionally, 1990.

[35] W. Shu, H. Yang, L. Zhang, M.M. Lu, E.E. Morrisey, Characterization of a new subfamily of winged-helix/forkhead (fox) genes that are expressed in the lung and act as transcriptional repressors, J. Biol. Chem. 276 (2001) 27488–27497, https://doi.org/10.1074/jbc.M100636200.

[36] M.M. Lu, S. Li, H. Yang, E.E. Morrisey, Foxp4: a novel member of the Foxp subfamily of winged-helix genes co-expressed with Foxp1 and Foxp2 in pulmonary and gut tissues, Mech. Dev. 119 (2002) S197–S202, https://doi.org/10.1016/S0925-4773(03)00116-3.

[37] S. Hori, T. Nomura, S. Sakaguchi, Control of regulatory T cell development by the transcription factor Foxp3, J. Immunol. 198 (2017) 981–985, https://doi.org/10.1126/SCIENCE.1079490/SUPPL_FILE/HORI.SOM.PDF.

[38] W.L. Charng, E. Karaca, Z. Coban Akdemir, T. Gambin, M.M. Atik, S. Gu, J.E. Posey, S.N. Jhangiani, D.M. Muzny, H. Doddapaneni, J. Hu, E. Boerwinkle, R. A. Gibbs, J.A. Rosenfeld, F. Cui, F. Xia, K. Manickam, Y. Yang, E.A. Faqeih, A. Al Asmari, M.A.M. Saleh, A.W. El-Hattab, J.R. Lupski, Exome sequencing in

mostly consanguineous Arab families with neurologic disease provides a high potential molecular diagnosis rate, BMC Med. Genom. 9 (2016) 1–14, https://doi.org/10.1186/S12920-016-0208-3/FIGURES/4.

[39] M.R.M. Hussain, N.A. Shaik, J.Y. Al-Aama, H.Z. Asfour, F. Subhani Khan, T.A. Masoodi, M.A. Khan, N.S. Shaik, In silico analysis of Single Nucleotide Polymorphisms (SNPs) in human BRAF gene, Gene 508 (2012) 188–196, https://doi.org/10.1016/J.GENE.2012.07.014.

[40] M.A. Hasan, F.T. Hakim, M.T. Islam Shovon, M.M. Islam, M.S. Islam, M.A. Islam, The investigation of nonsynonymous SNPs of human SLC6A4 gene associated with depression: an in silico approach, Heliyon 7 (2021) e07815, https://doi.org/10.1016/J.HELIYON.2021.E07815.

[41] M.M. Kamal, M.N. Islam, M.G. Rabby, M.A. Zahid, M.M. Hasan, In silico functional and structural analysis of non-synonymous single nucleotide polymorphisms (nsSNPs) in human paired box 4 gene, Biochem. Genet. (2023) 1–24, https://doi.org/10.1007/S10528-023-10589-1/METRICS.

[42] A.K. Agrahari, E. Pieroni, G. Gatto, A. Kumar, The impact of missense mutation in PIGA associated to paroxysmal nocturnal hemoglobinuria and multiple congenital anomalies-hypotonia-seizures syndrome 2: a computational study, Heliyon 5 (2019) e02709, https://doi.org/10.1016/j.heliyon.2019.e02709.

[43] L.T. Rabi, D.Z. Valente, E. de Souza Teixeira, K.C. Peres, M. de Oliveira Almeida, N.E. Bufalo, L.S. Ward, Potential new cancer biomarkers revealed by quantum chemistry associated with bioinformatics in the study of selectin polymorphisms, Heliyon 10 (2024) e28830, https://doi.org/10.1016/J.HELIYON.2024.E28830.

[44] S.S. Ahmed, Z.T. Rifat, R. Lohia, A.J. Campbell, A.K. Dunker, M.S. Rahman, S. Iqbal, Characterization of intrinsically disordered regions in proteins informed by human genetic diversity, PLoS Comput. Biol. 18 (2022) e1009911, https://doi.org/10.1371/JOURNAL.PCBI.1009911.

[45] M.N.I. Bappy, A. Roy, M.G.R. Rabbi, N. Jahan, F.A. Chowdhury, S.F. Hoque, E.H. Sajib, P. Khan, F.M.A. Hossain, K.M.A. Zinnah, Scrutinizing deleterious nonsynonymous SNPs and their effect on human POLD1 gene, Genet. Res. 2022 (2022) e61, https://doi.org/10.1155/2022/1740768.

[46] O. Strnad, V. Vilém Šustr, B. Kozlíková, J. Sochor, Real-time Visualization of Protein Empty Space with Varying Parameters, 2013, pp. 65–70.

[47] E. Jung Choi, E. Jin Seo, D. Kyoung Kim, S. In Lee, Y. Woo Kwon, I. Ho Jang, K.-H. Kim, D.-S. Suh, J. Ho Kim, FOXP1 functions as an oncogene in promoting cancer stem cell-like characteristics in ovarian cancer cells, n.d. www.impactjournals.com/oncotarget.

[48] W. Enard, M. Przeworski, S.E. Fisher, C.S.L. Lai, V. Wiebe, T. Kitano, A.P. Monaco, S. Pääbo, Molecular evolution of FOXP2, a gene involved in speech and language, Nature 418 (2002) 869–872, https://doi.org/10.1038/nature01025.

[49] S. yi Wu, P. Liao, L. yu Yan, Q. yi Zhao, Z. yu Xie, J. Dong, H. tao Sun, Correlation of MKI67 with prognosis, immune infiltration, and T cell exhaustion in hepatocellular carcinoma, BMC Gastroenterol. 21 (2021), https://doi.org/10.1186/s12876-021-01984-2.

[50] C. Abad, M.C. Robayo, M. del M. Muñiz-Moreno, M.T. Bernardi, M.G. Otero, C. Kosanovic, A.J. Griswold, T.M. Pierson, K. Walz, J.I. Young, Gatad2b, associated with the neurodevelopmental syndrome GAND, plays a critical role in neurodevelopment and cortical patterning, Translational Psychiatry 2024 14 (1 14) (2024) 1–13, https://doi.org/10.1038/s41398-023-02678-x.

[51] A. Liu, J. Pang, G. Xiong, Q. Liu, L. Chen, Forkhead Box P4 promotes the proliferation of cells in colorectal adenocarcinoma, Oncologie 25 (2023) 543–552, https://doi.org/10.1515/oncologie-2023-0009.

[52] P.-J. Lin, F.-Y. Chiu, S.-C. Wang, C.-Y. Li, The oncogenic role of ARG2 in hepatocellular carcinoma, 10.1200/JCO.2020.38.15_suppl.E16713 38, https://doi.org/10.1200/JCO.2020.38.15_SUPPL.E16713, 2020.

[53] S. Galosi, B.H. Edani, S. Martinelli, H. Hansikova, E.A. Eklund, C. Caputi, L. Masuelli, N. Corsten-Janssen, M. Srour, R. Oegema, D.G.M. Bosch, C.A. Ellis, L. Amlie-Wolf, A. Accogli, I. Atallah, L. Averdunk, K.W. Barañano, R. Bei, I. Bagnasco, A. Brusco, S. Demarest, A.S. Alaix, C. Di Bonaventura, F. Distelmaier, F. Elmslie, Z. Gan-Or, J.M. Good, K. Gripp, E.J. Kamsteeg, E. MacNamara, C. Marcelis, N. Mercier, J. Peeden, S. Pizzi, L. Pannone, M. Shinawi, C. Toro, N. E. Verbeek, S. Venkateswaran, P.G. Wheeler, L. Zdrazilova, R. Zhang, G. Zorzi, R. Guerrini, W.C.C. Sessa, D.J. Lefeber, M. Tartaglia, F.F. Hamdan, K.A. Grabińska, V. Leuzzi, De novo DHDDS variants cause a neurodevelopmental and neurodegenerative disorder with myoclonus, Brain 145 (2022) 208–223, https://doi.org/10.1093/brain/awab299.

[54] C.H. Kim, FOXP3 and its role in the immune system, Adv. Exp. Med. Biol. 665 (2009) 17–29, https://doi.org/10.1007/978-1-4419-1599-3_2.

[55] C. Zheng, X. Yu, Y. Liang, Y. Zhu, Y. He, L. Liao, D. Wang, Y. Yang, X. Yin, A. Li, Q. He, B. Li, Targeting PFKL with penfluridol inhibits glycolysis and suppresses esophageal cancer tumorigenesis in an AMPK/FOXO3a/BIM-dependent manner, Acta Pharm. Sin. B 12 (2022) 1271–1287, https://doi.org/10.1016/J.APSB.2021.09.007.

[56] L.H. Castilla, NRAS palmitoylation and oncogenic fitness, Blood 135 (2020) 1725–1726, https://doi.org/10.1182/BLOOD.2020005720.

[57] K. Adam, J. Lesperance, T. Hunter, P.E. Zage, The potential functional roles of NME1 histidine kinase activity in neuroblastoma pathogenesis, Int. J. Mol. Sci. 21 (2020), https://doi.org/10.3390/IJMS21093319.

[58] M. Alsanafi, R.D.R. Brown, J. Oh, D.R. Adams, F. Torta, N.J. Pyne, S. Pyne, Dihydroceramide desaturase functions as an inducer and rectifier of apoptosis: effect of retinol derivatives, antioxidants and phenolic compounds, Cell Biochem. Biophys. 79 (2021) 461, https://doi.org/10.1007/S12013-021-00990-1.

[59] W. Yang, X. Cui, D. Sun, G. Sun, Z. Yan, M. Wei, Z. Wang, W. Yu, POU5F1 promotes the proliferation, migration, and invasion of gastric cancer cells by reducing the ubiquitination level of TRAF6, Cell Death Dis. 14 (2023), https://doi.org/10.1038/S41419-023-06332-8.

[60] A. Fadda, N. Syed, R. Mackeh, A. Papadopoulou, S. Suzuki, P.V. Jithesh, T. Kino, Genome-wide regulatory roles of the C2H2-type zinc finger protein ZNF764 on the glucocorticoid receptor, Sci. Rep. 7 (2017), https://doi.org/10.1038/srep41598.

[61] M. Lin, Y. Li, S. Qin, Y. Jiao, F. Hua, Ubiquitin-like modifier-activating enzyme 7 as a marker for the diagnosis and prognosis of breast cancer, Oncol. Lett. 19 (2020) 2773–2784, https://doi.org/10.3892/OL.2020.11406/DOWNLOAD.

[62] A.E. Mackenzie, G. Milligan, The emerging pharmacology and function of GPR35 in the nervous system, Neuropharmacology 113 (2017) 661–671, https://doi.org/10.1016/J.NEUROPHARM.2015.07.035.

[63] K.C. Bao, F.F. Wang, The role of SPDEF in cancer: promoter or suppressor, Neoplasma 69 (2022) 1270–1276, https://doi.org/10.4149/NEO_2022_220529N571.

[64] J.C. Rudd, S. Maity, J.A. Grunkemeyer, J.C. Snyder, S. Lovas, L.A. Hansen, Membrane structure and internalization dynamics of human Flower isoforms hFWE3 and hFWE4 indicate a conserved endocytic role for hFWE4, J. Biol. Chem. 299 (2023) 104945, https://doi.org/10.1016/j.jbc.2023.104945.

[65] S. Spiegler, B. Kirchmaier, M. Rath, G.C. Korenke, F. Tetzlaff, M. Van De Vorst, K. Neveling, A. Acker-Palmer, A.W. Kuss, C. Gilissen, A. Fischer, S. Schulte-Merker, U. Felbor, FAM222B is not a likely novel candidate gene for cerebral cavernous malformations, Mol. Syndromol. 7 (2016) 144–152, https://doi.org/10.1159/000446884.

[66] K.M. Miller, J.V. Tjeertes, J. Coates, G. Legube, S.E. Polo, S. Britton, S.P. Jackson, Human HDAC1 and HDAC2 function in the DNA-damage response to promote DNA nonhomologous end-joining, Nature Structural & Molecular Biology 2010 17 (9 17) (2010) 1144–1151, https://doi.org/10.1038/nsmb.1899.

[67] M.N. Islam, M.G. Rabby, M.M. Hossen, M.A. Kamal, M.A. Zahid, M. Syduzzaman, M.M. Hasan, In silico functional and pathway analysis of risk genes and SNPs for type 2 diabetes in Asian population, PLoS One 17 (2022) e0268826, https://doi.org/10.1371/JOURNAL.PONE.0268826.

[68] R.M. George, A.B. Firulli, Hand factors in cardiac development, Anat. Rec. 302 (2019) 101–107, https://doi.org/10.1002/AR.23910.

[69] S. Wang, J. Wu, Z. Wang, Z. Gong, Y. Liu, Z. Wang, Emerging roles of Circ-ZNF609 in multiple human diseases, Front. Genet. 13 (2022) 837343, https://doi.org/10.3389/FGENE.2022.837343/BIBTEX.

[70] S. Vijayalingam, U.R. Ezekiel, F. Xu, T. Subramanian, E. Geerling, B. Hoelscher, K.K. San, A. Ganapathy, K. Pemberton, E. Tycksen, A.K. Pinto, J.D. Brien, D. B. Beck, W.K. Chung, C.A. Gurnett, G. Chinnadurai, Human iPSC-derived neuronal cells from CTBP1-mutated patients reveal altered expression of neurodevelopmental gene networks, Front. Neurosci. 14 (2020), https://doi.org/10.3389/fnins.2020.562292.

[71] J. Huang, H. Gou, J. Yao, K. Yi, Z. Jin, M. Matsuoka, T. Zhao, The noncanonical role of EZH2 in cancer, Cancer Sci. 112 (2021) 1376, https://doi.org/10.1111/CAS.14840.

[72] R. Alam, R. Rahman Imon, M. Enamul, K. Talukder, S. Akhter, M.A. Hossain, F. Ahammad, M.M. Rahman, GC-MS analysis of phytoconstituents from Ruellia prostrata and Senna tora and identification of potential anti-viral activity against SARS-CoV-2. https://doi.org/10.1039/d1ra06842c, 2021.

[73] Q.Z. Long, Y.F. Du, X.Y. Ding, X. Li, W. Bin Song, Y. Yang, P. Zhang, J.P. Zhou, X.G. Liu, Replication and fine mapping for association of the C2orf43, FOXP4, GPRC6A and RFX6 genes with prostate cancer in the Chinese population, PLoS One 7 (2012) e37866, https://doi.org/10.1371/JOURNAL.PONE.0037866.

[74] A. Manichaikul, X.Q. Wang, L. Sun, J. Dupuis, A.C. Borczuk, J.N. Nguyen, G. Raghu, E.A. Hoffman, S. Onengut-Gumuscu, E.A. Farber, J.D. Kaufman, D. Rabinowitz, K.D.H. Stukovsky, S.M. Kawut, G.M. Hunninghake, G.R. Washko, G.T. O'Connor, S.S. Rich, R.G. Barr, D.J. Lederer, Genome-wide association study of subclinical interstitial lung disease in MESA, Respir. Res. 18 (2017) 1–11, https://doi.org/10.1186/S12931-017-0581-2/FIGURES/3.

[75] K. Shiraishi, Y. Okada, A. Takahashi, Y. Kamatani, Y. Momozawa, K. Ashikawa, H. Kunitoh, S. Matsumoto, A. Takano, K. Shimizu, A. Goto, K. Tsuta, S. I. Watanabe, Y. Ohe, Y. Watanabe, Y. Goto, H. Nokihara, K. Furuta, A. Yoshida, K. Goto, T. Hishida, M. Tsuboi, K. Tsuchihara, Y. Miyagi, H. Nakayama, T. Yokose, K. Tanaka, T. Nagashima, Y. Ohtaki, D. Maeda, K. Imai, Y. Minamiya, H. Sakamoto, A. Saito, Y. Shimada, K. Sunami, M. Saito, J. Inazawa, Y. Nakamura, T. Yoshida, J. Yokota, F. Matsuda, K. Matsuo, Y. Daigo, M. Kubo, T. Kohno, Association of variations in HLA class II and other loci with susceptibility to EGFR-mutated lung adenocarcinoma, Nature Communications 2016 7 (1 7) (2016) 1–7, https://doi.org/10.1038/ncomms12451.

[76] Z. Wang, W.J. Seow, K. Shiraishi, C.A. Hsiung, K. Matsuo, J. Liu, K. Chen, T. Yamji, Y. Yang, I.S. Chang, C. Wu, Y.C. Hong, L. Burdett, K. Wyatt, C.C. Chung, S. A. Li, M. Yeager, A. Hutchinson, W. Hu, N. Caporaso, M.T. Landi, N. Chatterjee, M. Song, J.F. Fraumeni, T. Kohno, J. Yokota, H. Kunitoh, K. Ashikawa, Y. Momozawa, T. Daigo, T. Mitsudomi, Y. Yatabe, T. Hida, Z. Hu, J. Dai, H. Ma, G. Jin, B. Song, Z. Wang, S. Cheng, Z. Yin, X. Li, Y. Ren, P. Guan, J. Chang, W. Tan, C.J. Chen, G.C. Chang, Y.H. Tsai, W.C. Su, K.Y. Chen, M.S. Huang, Y.M. Chen, H. Zheng, H. Li, P. Cui, H. Guo, P. Xu, L. Liu, M. Iwasaki, T. Shimazu, S. Tsugane, J. Zhu, G. Jiang, K. Fei, J.Y. Park, Y.H. Kim, J.S. Sung, K.H. Park, Y.T. Kim, Y.J. Jung, C.H. Kang, I.K. Park, H.N. Kim, H.S. Jeon, J.E. Choi, Y.Y. Choi, J.H. Kim, I.J. Oh, Y.C. Kim, S.W. Sung, J.S. Kim, H. Il Yoon, S.S. Kweon, M.H. Shin, A. Seow, Y. Chen, W.Y. Lim, J. Liu, M.P. Wong, V.H.F. Lee, B.A. Bassig, M. Tucker, S.I. Berndt, W.H. Chow, B.T. Ji, J. Wang, J. Xu, A.D.L. Sihoe, J.C.M. Ho, J.K.C. Chan, J.C. Wang, D. Lu, X. Zhao, Z. Zhao, J. Wu, H. Chen, L. Jin, F. Wei, G. Wu, S.J. An, X.C. Zhang, J. Su, Y.L. Wu, Y.T. Gao, Y.B. Xiang, X. He, J. Li, W. Zheng, X.O. Shu, Q. Cai, R. Klein, W. Pao, C. Lawrence, H.D. Hosgood, C. F. Hsiao, L.H. Chien, Y.H. Chen, C.H. Chen, W.C. Wang, C.Y. Chen, C.L. Wang, C.J. Yu, H.L. Chen, Y.C. Su, F.Y. Tsai, Y.S. Chen, Y.J. Li, T.Y. Yang, C.C. Lin, P. C. Yang, T. Wu, D. Lin, B. Zhou, J. Yu, H. Shen, M. Kubo, S.J. Chanock, N. Rothman, Q. Lan, Meta-analysis of genome-wide association studies identifies multiple lung cancer susceptibility loci in never-smoking Asian women, Hum. Mol. Genet. 25 (2016) 620–629, https://doi.org/10.1093/HMG/DDV494.

[77] G. Jin, H. Ma, C. Wu, J. Dai, R. Zhang, Y. Shi, J. Lu, X. Miao, M. Wang, Y. Zhou, J. Chen, H. Li, S. Pan, M. Chu, F. Lu, D. Yu, Y. Jiang, J. Dong, L. Hu, Y. Chen, L. Xu, Y. Shu, S. Pan, W. Tan, B. Zhou, D. Lu, T. Wu, Z. Zhang, F. Chen, X. Wang, Z. Hu, D. Lin, H. Shen, Genetic variants at 6p21.1 and 7p15.3 are associated with risk of multiple cancers in Han Chinese. https://doi.org/10.1016/j.ajhg.2012.09.009, 2012.

[78] C. Wu, Z. Hu, Z. He, W. Jia, F. Wang, Y. Zhou, Z. Liu, Q. Zhan, Y. Liu, D. Yu, K. Zhai, J. Chang, Y. Qiao, G. Jin, Z. Liu, Y. Shen, C. Guo, J. Fu, X. Miao, W. Tan, H. Shen, Y. Ke, Y. Zeng, T. Wu, D. Lin, Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations, Nature Genetics 2011 43 (7 43) (2011) 679–684, https://doi.org/10.1038/ng.849.

[79] R. Takata, S. Akamatsu, M. Kubo, A. Takahashi, N. Hosono, T. Kawaguchi, T. Tsunoda, J. Inazawa, N. Kamatani, O. Ogawa, T. Fujioka, Y. Nakamura, H. Nakagawa, Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population, Nature Genetics 2010 42 (9 42) (2010) 751–754, https://doi.org/10.1038/ng.635.

[80] Y.S. Luo, K. Zhang, Z.S. Cheng, Absence of association between a long COVID and severe COVID-19 risk variant of FOXP4 and lung cancer, Front. Genet. 14 (2023), https://doi.org/10.3389/FGENE.2023.1258829/PDF.