JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 11 Issue 01
2026

Research Article
https://doi.org/10.56979/1101/2026/1067

Computational Assessment of nsSNPs Associated with FGFR2 Gene

Hira Saleem[1*], and Anum Munir[2]

[1]N.F.C I.E.T, Multan, 60000, Pakistan.
[2]University of Lahore, Lahore, 42000, Pakistan.
[*]Corresponding Author: Hira Saleem. Email:hirasaleem0715@gmail.com

Received: September 10, 2025 **Accepted:** March 01, 2026

Abstract: The fibroblast growth factor receptor 2 (FGFR2) controls cell proliferation, differentiation, angiogenesis, and wound healing. Single nucleotide polymorphisms (SNPs) of the FGFR2 gene are linked with Pfeiffer syndrome, Crouzon syndrome, and Jackson-Weiss syndrome. Missense mutations in FGFR2 have also been reported in breast, gastric, and lung cancer. Objective: This study aimed to systematically analyze missense SNPs (nsSNPs) in FGFR2 using an integrative computational pipeline to identify variants with the strongest predicted pathogenic impact. Methods: FGFR2 gene data and protein sequence were retrieved from the NCBI dbSNP [build 153 (March 2020)] and UniProt database [release 2020_02; UniProt ID: P21802]. A total of eleven bioinformatics tools—SIFT, PolyPhen-2, PROVEAN, SNAP2, SNPs&GO, PANTHER, PhD-SNP, PMut, I-Mutant3.0, ConSurf server, and Project HOPE—were used to examine the deleterious potential of nsSNPs. The interaction of FGFR2 with different genes was analyzed using GeneMANIA. Results: We investigated 28,027 total SNPs from dbSNP, of which 701 were coding variants: 294 synonymous and 407 non-synonymous. Among the latter, 393 were missense mutations, 7 frameshift mutations, and 7 nonsense mutations. Only missense nsSNPs were retained for further analysis. Stepwise filtering identified 90 consensus-deleterious variants (≥ 6/8 predictors), 54 extremely damaging variants (all 8 predictors), and 38 unstable variants ($\Delta\Delta G \leq -0.5$ kcal/mol). A final set of 24 highly conserved and damaging variants (ConSurf score ≥7) was prioritized. Conclusion: A total of 24 nsSNPs (L757S, G690R, P666S, R664W, D644N, Y616C, L572F, L550P, L550V, I547M, V514M, G502E, G502R, E489K, R450C, P286S, C278Y, G271R, P263L, P256S, D225E, S224P, E219G, and Y105C) were predicted to have the most damaging and disease-causing effects on FGFR2 protein function and structure. Thus, the early prediction of FGFR2 gene functions could aid in disease prognosis. The results of our study provide beneficial information for devising early diagnostic and therapeutic measures.

Keywords: FGFR2 Gene; Computational Assessment; nsSNPs; Mutations; Damaging; Craniosynostosis; Breast Cancer; In-Silico

1. Introduction

The fibroblast growth factor receptor (FGFR2) plays an essential role in diverse biological processes, including development, proliferation, angiogenesis, differentiation, and wound healing [1]. The FGFR family consists of four members (FGFR1–FGFR4), which encode transmembrane tyrosine kinase receptors involved in cellular signaling and regulation. These signals promote cell growth, proliferation, and molecular variation. FGFR2 belongs to the fibroblast receptor family and exhibits a highly conserved sequence across members and during development. Structurally, FGFR2 contains an extracellular region with three immunoglobulin-like domains, a transmembrane segment, and an intracellular tyrosine kinase domain, and is located on chromosome 10.

Importantly, exons IIIa (chr10:123,XXX–123,YYY, hg38) and IIIc (chr10:124,XXX–124,YYY, hg38) encode parts of the third immunoglobulin-like domain within FGFR2, which are crucial for ligand binding.

These exons represent conservation hotspots for mutations associated with craniosynostosis syndromes such as Crouzon, Pfeiffer, and Jackson–Weiss [2,3]. Mutations in exon IIIc are specifically linked with Crouzon and Jackson–Weiss syndromes [3], while other FGFR2 variants are implicated in Apert syndrome [4]. For instance, more than 99% of Apert syndrome cases are caused by two recurrent mutations (S252T and P253A), which alter amino acids in FGFR2 and disrupt receptor function [5].

Single nucleotide polymorphisms (SNPs) are the most common form of genetic variation in the human genome, arising from substitution of a single nucleotide (A, T, C, or G). They are typically bi-allelic in nature and occur frequently across both coding and non-coding regions [6]. Coding SNPs are classified as synonymous (silent) or non-synonymous (nsSNPs). nsSNPs, also known as missense SNPs, can directly alter amino acid sequences, leading to changes in protein structure, stability, hydrophobicity, or charge. Such deleterious effects can impair protein function and contribute to pathogenic phenotypes. Thus, nsSNPs are particularly important for understanding disease mechanisms and identifying biomarkers.

In FGFR2, SNPs have been strongly associated with syndromic craniosynostosis, including Crouzon, Apert, and Pfeiffer syndromes [2–5,7]. Moreover, genetic variants in FGFR2 have been linked to breast cancer susceptibility [8]. Overexpression or mutation-driven activation of FGFR2 can promote oncogenic processes such as uncontrolled proliferation, cell migration, and angiogenesis, underscoring its clinical significance in both developmental syndromes and cancer.

Although not all SNPs are deleterious, those with functional consequences may serve as valuable biomarkers for disease susceptibility and drug response. Considering that approximately 500,000 SNPs fall within coding regions of the human genome, experimental validation of each is not feasible. Therefore, computational approaches are essential for prioritizing and identifying pathogenic nsSNPs. When combined with sequence and structural data, these methods can achieve prediction accuracies of 80–85% [9–12].

In this study, we investigate functional variations in the coding region of FGFR2 using bioinformatics approaches to identify pathogenic nsSNPs. This classification is critical for understanding genetic determinants of disease, improving interpretation of structural and functional features of FGFR2, and supporting pharmacogenomic studies for targeted drug discovery.

2. Materials and Method

2.1. Dataset assortment:

The FGFR2 protein sequence (UniProt accession ID: P21802) was retrieved from the UniProtKB database (Release 2020_02, accessed in March 2020). Gene-related SNP data (RefSeq accession ID: NP_000132.3) were collected from the NCBI dbSNP database (Build 153, accessed in March 2020), including residue changes. A total of 393 nonsynonymous SNPs (nsSNPs) were screened for this study. Interactions of the FGFR2 gene with other genes were examined using GeneMANIA (accessed in March 2020).

2.2. Investigation of the FGFR2 Gene's Interactions in GeneMANIA

GeneMANIA is a database that finds the function and sets of your favourite gene, using a variety of association data. Association data includes protein interactions, co-localization, pathways, co-expression, and protein domain similarity. GeneMANIA can also put in service to invent new associates, discover further genes that you have lost in screening or identify novel genes such as protein kinase [13]. Accessible at: http://www.GeneMANIA.org

2.3. Recognition of the damaging Non-Synonymous Single-Nucleotide Polymorphisms:

We analyzed four bioinformatics tools (SIFT, PolyPhen-2, PROVEAN, and SNAP2) to assess the functional effects (damaging or tolerated) of nsSNPs in FGFR2. Variants predicted to be damaging in these tools were further evaluated using SNPs&GO, PhD-SNP, PMut, and PANTHER to investigate their disease-association potential. Protein stability effects were assessed with I-Mutant3.0, while evolutionary conservation was analyzed with ConSurf Server. Structural and functional consequences of prioritized mutations were further explored using Project HOPE. To integrate predictions and resolve conflicts across tools, a consensus-based filtering approach was applied: Variants predicted as deleterious by ≥6 of 8 functional tools (SIFT, PolyPhen-2, PROVEAN, SNAP2, SNPs&GO, PhD-SNP, PMut, PANTHER) were retained. These variants were further filtered by stability analysis in I-Mutant3.0 ($\Delta\Delta G < -0.5$ kcal/mol) and

conservation analysis in ConSurf (score ≥ 7). Only variants meeting all these criteria were considered high-confidence deleterious nsSNPs. **Figure 1.** shows a schematic overview of the integrative computational pipeline used in this study.

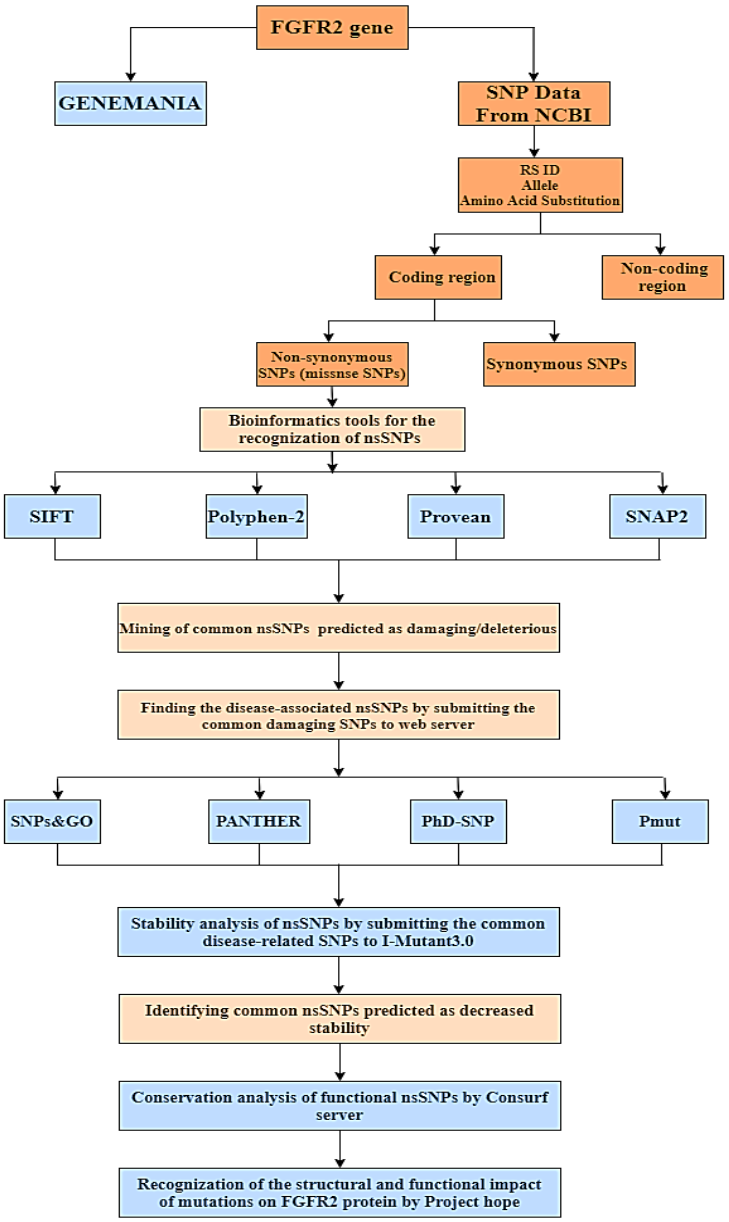

Figure 1. Pictorial Illustration of bioinformatics software used in our study

2.3.1. SIFT

SIFT(Sorting intolerant from tolerant; https://sift.bii.a-star.edu.sg/;version 5.2.2, 2019) is a sequence-homology tool that anticipates whether an amino acid substitution disturbs the protein function. Prediction is built on the gradation of amino acid conservation [14]. SIFT makes investigation that is built on many procedures, and it takes the homology categorizations by using the Swiss-Prot and TrEMBL. Ponders the location at which the variation happened and the amino acid variation type. Having a query sequence, the SIFT selects interrelated proteins and gets an arrangement with the query of these proteins. The chromosome and its position, coordinates, orientations (forward or backward), alleles (C, T, G, A) of the rs-ID were submitted as an input to the server for prediction. On the basis of this, SIFT computes the possibility that the amino acid residue is damaged or tolerated. The prediction was specified as a TI (tolerance index) score which varies from 0 (neutral) to 1, and TI score < 0.05 were foreseen to be damaged, whereas TI score >0.05 were considered to be tolerated.

2.3.2. Polyphen-2

Polyphen-2 software (Polymorphism Phenotyping v2; http://genetics.bwh.harvard.edu/pph2; version 2.2.2, 2018). All nsSNPs yielded to the SIFT tool were also preceded to Polyphen-2. It does not only rely on the homology sequence for SNP predictions as per their modelling also built on structural information [15]. By using evolutionary considerations, it was applied to examine the functional outcome of amino acid substitutions on protein function and structure [16]. Polyphen-2 used the Bayesian classifier for the prediction analysis. It determines the PSIC score and then calculates the change between the two variants. The result was obtained in the form of probability score which categorizes the mutations as 'benign' (0.00 to 0.31), 'possibly damaging' (0.7 to 0.95), and 'probably damaging' (0.95 to 1), Protein accession ID **P21802** along with substitution position, wild and mutant amino acids of nsSNPs were submitted to the software as an input.

2.3.3. PROVEAN.

PROVEAN (Protein Variation Effect Analyzer; http://provean.jcvi.org/index.php; version 1.1.5, 2020) is a suitable tool for refining the functionally imperative sequence variants to recognize the nonsynonymous variants. The performance of the PROVEAN tool is analogous to the SIFT or PolyPhen-2 tool. It is used to anticipate whether an amino acid substitution has an impact on the protein's structure and function. It examines the nsSNPs as natural or deleterious. It is founded on an alignment-based score [17]. By doing the comparison of a query sequence to a protein sequence with and without an amino acid variation, the score computes the change in between these. The score is also premised on the versions of protein sequence variants assembled from the NCBI database. The distribution of a PROVEAN score is a set of UniProt protein. The score is set as 2.5 for the classification of variants as deleterious or neutral. If the score was under the score of -2.5, the variant was foreseen to have a "Deleterious' effect and if the score was more than -2.5, the variant was foreseen to have a "Neutral" effect. Protein sequence in Fasta format along with amino acid substitutions was submitted to the software.

2.3.4. SNAP2

SNAP2 (Screening of non-acceptable Polymorphisms2;http//rostlab.org/services/snap2web/; release 2019) is a freely available tool and a trained classifier which is based on a "neural network". It recognizes the nsSNPs effect on protein function [18]. By taking a range of sequences, snap2 gives the prediction score of every mutant and then differentiates between the nsSNPs as the neutral and non-neutral effect.. FASTA sequence of the protein and a mutants list was submitted to the tool as an input query which then caters to the scores of each substitution. The essential input for the calculation is the data occupied from an automatically produced multiple sequence alignment.

2.4. Recognization of disease-associated nsSNPs:

The nsSNPs present in the coding region may also lead to disease-related phenomena. The nsSNPs which were recognized as damaging in the previous tools (stated above) were also submitted to SNPs&GO, PhD- SNP and PMut tools observe the disease-related SNPs which are present in the FGFR2 coding region.

2.4.1. SNPs&GO

SNPs &GO (Single nucleotide polymorphism & Gene Ontology; http://snps.biofold.org/snps-and-go; release 2019) [19]. SNPs &GO is a support vector machine formed on an accurate process to precisely predict, from protein sequence, whether or not a mutation is disease-associated by manipulating the protein annotations. SNPs& GO assembles in a different context in which data is obtained from protein sequence as encrypted in the Gene Ontology terms. The protein sequence and SWISS Prot code from UniproktB in FASTA format were submitted to the software as an input. After that, mutations were submitted in the WT+POS+NEW format where WT is the wild residue; POS is the position of the residue and NEW is the new residue after mutation. The result is categorized as Disease or Neutral. The reliability index has a value of more than 5 shows the disease-associated effect. The PhD-SNP and PANTHER results were also given in the output.

2.4.2. PhD-SNP

PhD-SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms; http://snps.biofold.org/phd-snp/phd-snp.html; release 2019) is a web server that recognizes the disease-

related variation on single point mutations, which is built on supervised training algorithms. It gives a prediction by obtaining the aligned sequences, conservation index, and the frequencies of residues [20–22]. The FASTA format of protein sequence and the list of wild and mutant residues were provided to PhD-SNP for the analysis. The server provided the prediction of neutral or disease-related polymorphism.

2.4.3. PMut :

PMut (http://mmb.irbbarcelona.org/PMut; release 2017) is a web server that is used to recognize the disease-related mutations. PMUT method is created on a neural network classifier that is trained on a dataset extracted from SwissProt. PMut released in 2005 which allows the user to guess the pathological predictions, hot-spot locations on protein sequences, to obtain the whole repository of pre-calculated predictions and to create and authorize the new predictors. The output is shown as an index ranging from 0 to 1. The indexes > 0.5 were predicted to have a "Disease" effect. [23]. UniProt ID of FGFR2 (P27797) along with the position of wild and the mutant residue was submitted as an input.

2.5. Analysis of nsSNPs effect on Protein Stability by I-Mutant3.0:

I-Mutant3.0 web server; (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi; release 2009) is a support vector tool for the analysis of protein stability on single-site mutations, starting from the protein sequence or structure. It predicts the value of energy change (ΔΔG or DDG) (increase or decrease), in association with a reliability index ranging from 0 to 10, where 0 reliability is the minimum and 10 reliability is the maximum reliability. A DDG < 0 indicates the decrease in the stability of protein, whereas a DDG > 0 indicates an increase in the stability of protein [24]. The FASTA sequence of protein recovered from UniProt and the residues changes was used as an input to estimate the impact on mutation on the stability of the protein. The entries were set at a temperature of 25°C and pH 7.0. The result of the output of the predicted energy change (DDG) categorizes into three groups: largely stable (ΔΔG>0.5 kcal/mol), neutral (-0.5≤ ΔΔG≤0.5 kcal/mol) and largely unstable (ΔΔG < −0.5 kcal/mol)

2.6. Analyzing evolutionary conservation of nsSNPs :

ConSurf server (http://consurf.tau.ac.il; release 2020) is a bioinformatics tool that is used for the interpretation of the conservation of amino acid position in DNA/RNA premised on the phylogenetic relationship between homologous sequences [25]. The tool predicts the conservation score built on the association between the protein and its homologs sites [26]. An amino acid conservation degree was projected using 50 homologous sequences. The score between 1 to 4 is measured as variable, 5–6 is intermediate, and 7 to 9 conserved region. The protein FASTA sequence from UniProt (accession ID: P21802) was inserted to recognize the evolutionary conservation of amino acid.

2.7. Evaluating the Consequence of nsSNPs on the 3D Structure of the Protein:

Project Hope program (http://www.cmbi.ru.nl/hope/; release 2020) is an online tool to explore the 3D structures of protein by gathering the structural information from different websites, which includes calculations on protein 3D coordinates, predictions by DAS services and sequence annotations from the UniProt database. [27]. The protein sequence of the gene and the wild and mutant residues were given to the project hope to analyze the structural and functional impact of mutations. Project Hope makes a report with figures, simulations and text.

3. Results

3.1. FGFR2 gene function and activities demonstrated by GeneMANIA.

In this study, the FGFR2 gene was found to have an association with twenty other different genes. FGFR2 protein has many activities such as fibroblast growth factor receptor signalling pathway, organ growth, renal system development, growth factor activity, heparin-binding, mesonephros development, tube development, sulfur compound binding, lung development, tube morphogenesis, respiratory tube development, muscle organ development, mammary gland morphogenesis, organ formation, cardiac muscle cell proliferation, metabolic alcohol process, striated muscle cell proliferation, gland morphogenesis, reproductive structure development, cardiac muscle tissue growth, heart growth, regulation of angiogenesis, protein localization to the cell surface and Ureteric bud morphogenesis (**Table**

1.). The physical interaction, co-expression, pathways and shared protein domain of this gene were demonstrated by GeneMANIA and revealed in **Figure 2.**

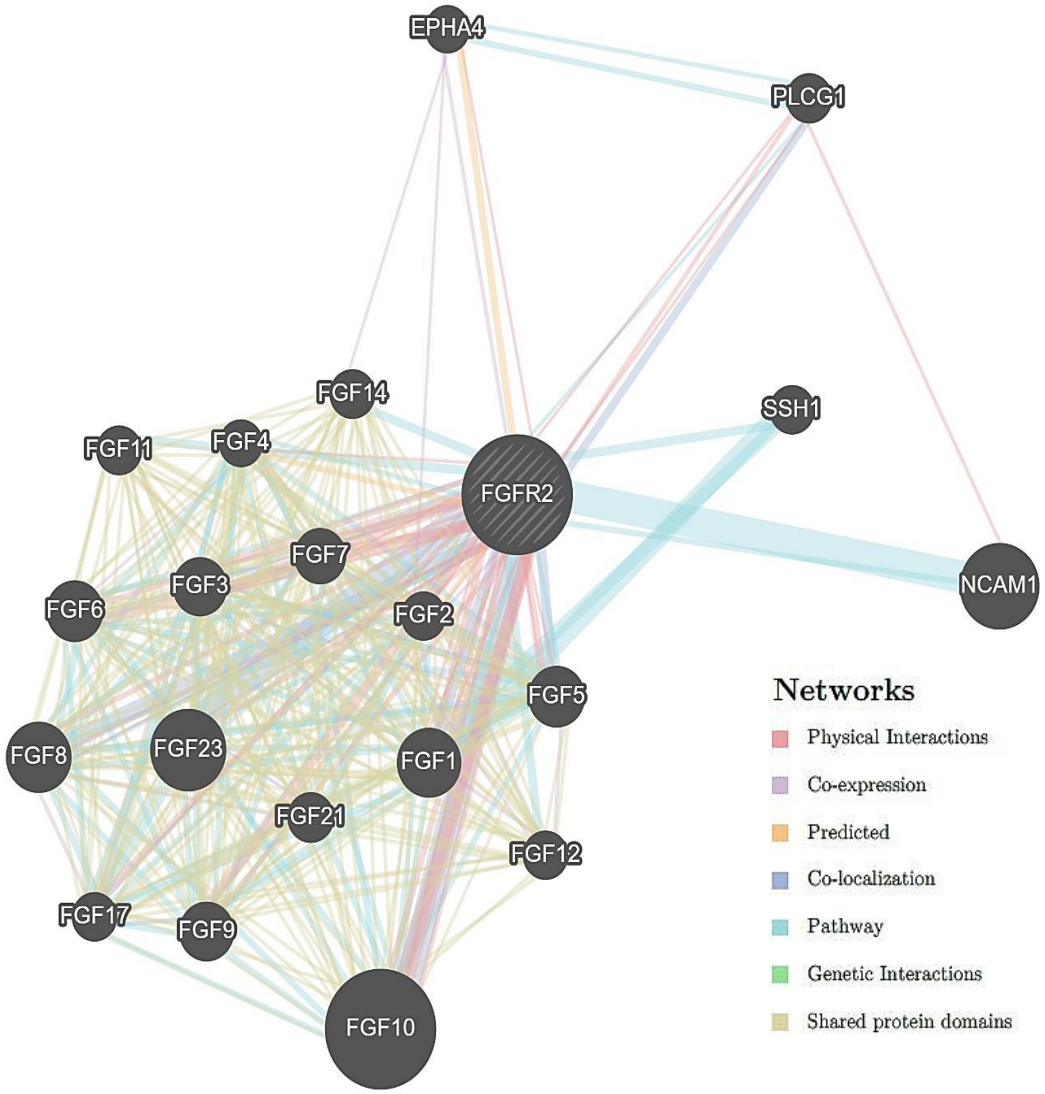

Figure 2. Physical interaction and co-expression of this gene with other genes.

Table 1. Demonstrates the FGFR2 gene functions and their presence in the network and genome.

Function	*FDR	Genes in network	Genes in genome
fibroblast growth factor receptor signaling pathway	8.667777261511095e-21	14	170
organ growth	0.000031989424612167816	4	33
renal system development	0.0001330632770989612	5	125
growth factor activity	0.00020163877509202843	4	54
heparin binding	0.0003219176629799635	4	63
mesonephros development	0.00033015183203449424	3	16
tube development	0.00042334603246629935	5	169
sulfur compound binding	0.0009357592074695067	4	88
lung development	0.003793036636680663	3	41

tube morphogenesis	0.0038719250301931696	4	134
respiratory tube development	0.004165860395552133	3	43
muscle organ development	0.00895048503794816	4	172
mammary gland morphogenesis	0.015376824570093539	2	11
organ formation	0.03210423448560348	2	17
cardiac muscle cell proliferation	0.034557999208576044	2	18
alcohol metabolic process	0.034557999208576044	4	265
striated muscle cell proliferation	0.034557999208576044	2	18
gland morphogenesis	0.034557999208576044	2	18
reproductive structure development	0.03514654753657123	3	102
cardiac muscle tissue growth	0.04450016898276371	2	21
central nervous system neuron development	0.04843419803759439	2	22
heart growth	0.06600672972795917	2	26
regulation of angiogenesis	0.06646869137657999	3	131
protein localization to cell surface	0.07999044146327977	2	29
ureteric bud morphogenesis	0.08981746266857912	2	31

*False discovery rate is greater than or equal to the probability that this is a false positive

3.2. SNPs retrieval from dbSNP database:

We utilized dbSNP for the retrieval of SNPs in the human FGFR2 gene. There were a total of 28,027 SNPs: 352 (1.25%) occurred in the 3'UTR, 212 (0.75%) in the 5' UTR, 26,772 (95.5%) in the intron region, 701(2.5%) were found in the coding region, among them 294 were synonymous and 407 non-synonymous SNPs that includes of 393 missense mutations, 7 frameshift mutations, and 7 nonsense mutations (**Figure 3**.). We selected only 393 missense non-synonymous coding SNPs for our investigation.

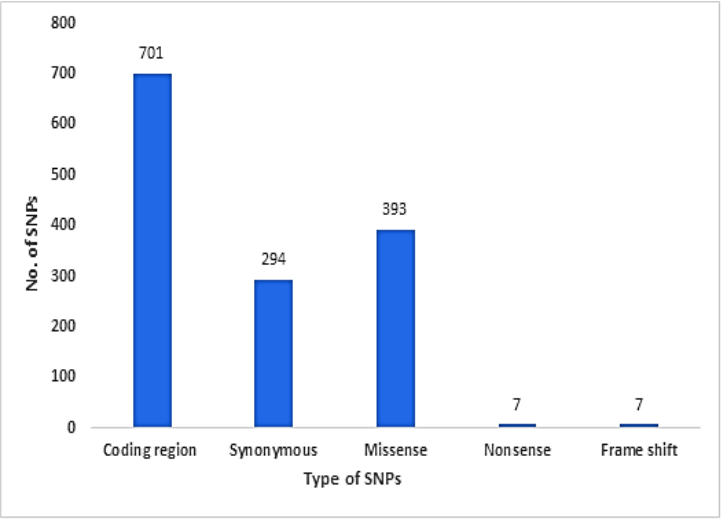

Figure 3. Representation of the partitioning of the coding region SNPs for the FGFR2 gene.

3.3. Functional analysis of nsSNPs.

A total of 393 nsSNPs of the FGFR2 gene from NCBI were found on the coding region were acquiesced to the SIFT, Polyphen-2, PROVEAN, and SNAP2 software to recognize the functional mutations. Based on

our SIFT prediction analysis, 152 nsSNPs were found to be Damaging (TI scores ≤ 0.05), 236 nsSNPs were forecasted to be tolerated (TI ≥ 0.05) and 5 nsSNPs were not found. Moreover, 67 polymorphisms demonstrated a score of 0.00 which was highly deleterious. Polyphen-2 results predicted that 63 nsSNPs were possibly Damaging with a score of (0.7–0.95), 115 nsSNPs were predicted to be 'probably Damaging' with a score of (0.95–1), and 215 nsSNPs were considered as benign. 191 nsSNPs were predicted to be deleterious by PROVEAN and 202 nsSNPs were neutral. The SNAP2 tool predicted that 140 nsSNPs were affected and 253 nsSNPs were neutral. Out of 393 nsSNPs, 90 nsSNPs were predicted to be deleterious commonly by these four tools SIFT, Polyphen-2, PROVEAN and SNAP2.

3.4. Identification of Disease-Associated Mutations:

To confirm the predictions obtained from four previous tools, Snp&Go, Panther, PhD SNP, and PMut were utilized. Out of 90 nsSNPs that predicted to be damaged with SIFT, Polyphen2, PROVEAN, and SNAP2; SNP & GO predicted the results of 85 nsSNPs as disease-causing mutations while the remaining were neutral. The results of the Panther tool were also given. In association with the 90 nsSNPs predicted, 81 nsSNPs were predicted to be disease-causing and the 8 nsSNPs were the neutral effect, the server was unable to predict the disease-causing capacity of 1 nsSNPs. PhD-Snp predicted 62 nsSNPs to be associated with the disease while 28 nsSNPs were neutral. PMut predicted 83 nsSNPs to be associated with the disease while the remaining were the neutral effect (**Figure 4**.). Finally, out of 393 nsSNPs, 54 nsSNPs were found to be extremely Damaging by (SIFT, Polyphen2, PROVEAN, SNAP2, Snp&Go, Panther, PhD-Snp, and PMut) servers.

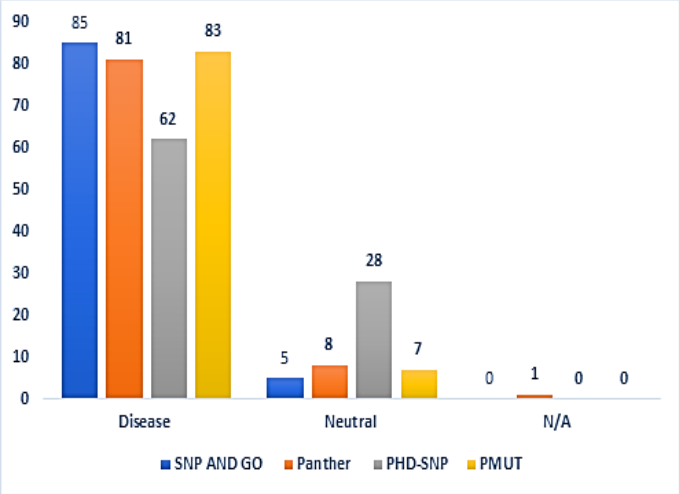

Figure 4. Graphical illustration of disease mutations by Snp&Go, Panther, PhD-Snp, and PMut.

3.5. Prediction Effect of nsSNPS on the FGFR2 Protein Stability:

All the nsSNPs which were predicted as disease-causing were also observed by the I- mutant server to predict the consequence of the nsSNPs on the protein stability. The results were classified as: 38 nsSNPs (C808F ,L757S ,E731K ,S702L ,G690R ,P666S ,R664W ,D644N ,Y616C ,G583R ,L572F ,Y561S ,L550P ,L550V ,I547M ,M537K ,D522Y ,V514M ,G502E ,G502R ,Q494R ,E489K ,R450C ,P286S ,C278Y ,G271R ,P263L ,P256 S ,S239F ,D225E ,S224P,E219G ,R210G ,L192Q ,R178H ,R178S ,R165W ,Y105C) predicted a decrease in stability of the FGFR2 protein and 16 SNP (C808W ,C790Y ,S619F ,Q556P ,N546T ,G542R ,D530N ,D506Y ,K485Q ,L483P ,E475K ,L258F ,V233M ,L216F ,G195E ,G195R) predicted an increase in stability of the FGFR2 protein.

3.6. Evolutionary Conservation Prediction of nsSNPs:

Some variations that lead to disease-causing mutations exist in conserved regions also. The nsSNPs that situated at conserved regions inclines to be more harmful than nsSNPs; located at non-conserved areas. The ConSurf server did the analyses of the 38 nsSNPs that foreseen to be damaged by the nine tools (stated above).Out of the 38 nsSNPs, the 30 nsSNPs (L757S ,E731K ,S702L ,G690R ,P666S ,R664W ,D644N ,Y616C ,L572F ,G583R ,L550P ,L550V ,I547M ,M537K ,D522Y ,V514M ,G502E ,G502R ,Q494R ,E489K ,R450C ,P286S

,C278Y ,G271R ,P263L ,P256S ,D225E ,S224P ,E219G ,Y105C) were located in a highly conserved region with the UniRef90 protein database whereas the 8 nsSNPs (C808F ,Y561S ,S239F ,R210G ,L192Q ,R178H ,R178S ,R165W) were situated in a non-conserved regions.

Figure 5. shows the result of the ConSurf server. The output comprises of nine colour codes (1-9) based on the conservation scores. The purple colour shows the conserved regions and the blue colour specifies variability. ConSurf predicted the residues to be exposed (e), buried (b), functional (highly conserved and exposed; f), or structural (highly conserved and buried; s). The purple boxes represent. C808F, L757S, E731K ,S702L ,G690R ,P666S ,R664W ,D644N ,Y616C ,G583R ,L572F ,Y561S ,L550P ,L550V ,I547M ,M537K ,D522Y ,V514M ,G502E ,G502R ,Q494R ,E489K ,R450C ,P286S ,C278Y ,G271R ,P263L ,P256S ,S239F ,D225E ,S224P ,E219G ,R210G ,L192Q ,R178H ,R178S ,R165W ,Y105C mutations, respectively.

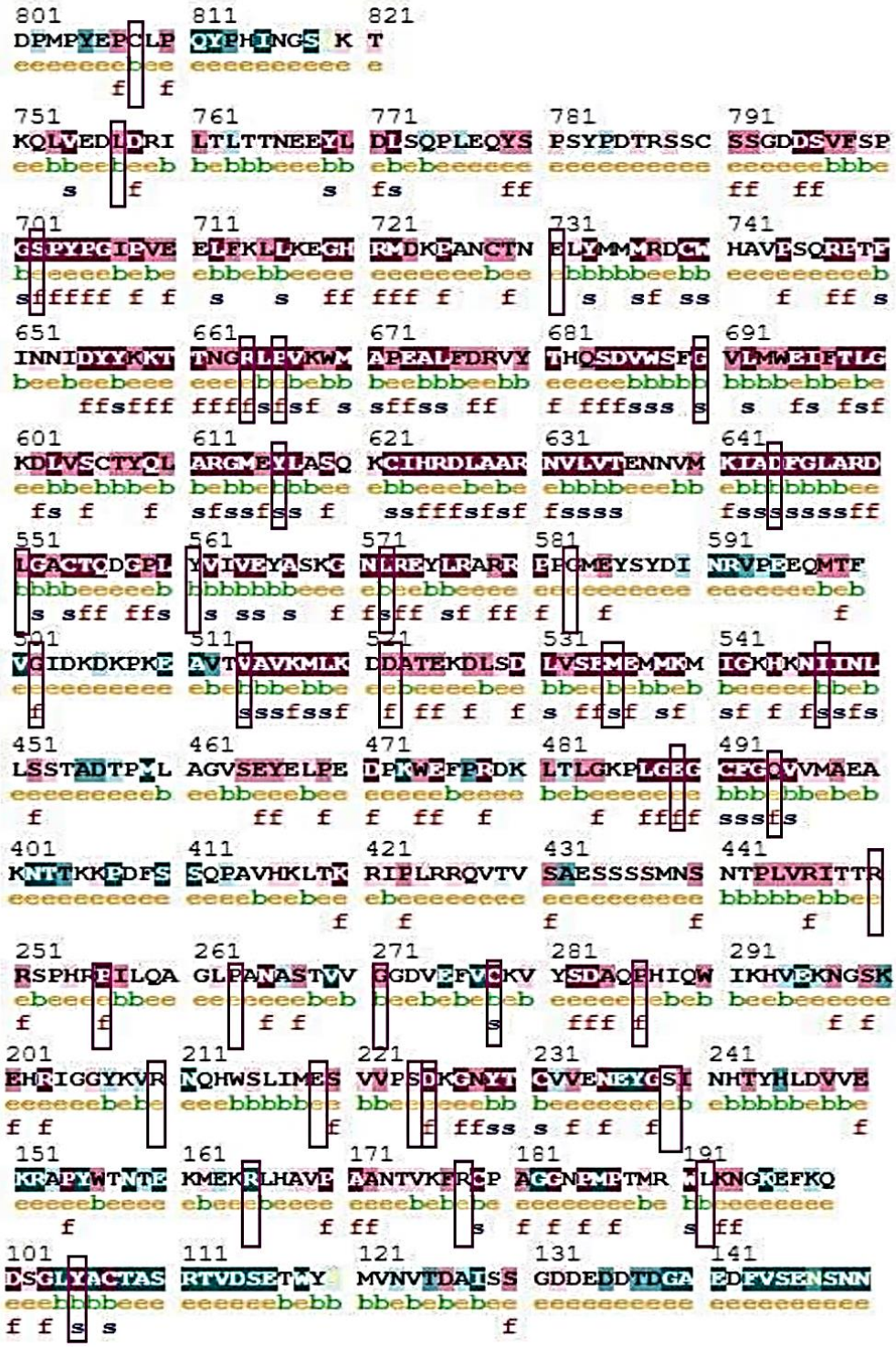

Figure 5. Displays the results foreseen by the ConSurf by the UniRef90.

3.7. Result of mutations on protein by using Project Hope:

From dbSNP build 155, a total of 28,027 FGFR2 variants were retrieved, of which 701 were identified as coding variants, including 294 synonymous and 407 non-synonymous mutations. Among the non-synonymous variants, 393 were missense, 7 were frameshift, and 7 were nonsense mutations. For the present study, only the missense variants were considered for further analysis. A stepwise filtering strategy was applied to refine the dataset. Out of the 393 nsSNPs, 90 were initially predicted to be deleterious by at least six out of eight functional prediction tools. Within this group, 54 variants were consistently predicted as damaging across all eight tools (SIFT, PolyPhen-2, PROVEAN, SNAP2, SNPs&GO, PANTHER, PhD-SNP, and PMut). Further stability analysis with I-Mutant3.0 identified 38 of these variants as structurally unstable ($\Delta\Delta G \leq -0.5$ kcal/mol). ConSurf analysis then revealed that 30 of these unstable variants were located in highly conserved residues (ConSurf score ≥ 7). Based on conservation and functional prediction, a final set of 24 nsSNPs was prioritized as the most deleterious, including L757S, G690R, P666S, R664W, D644N, Y616C, L572F, L550P, L550V, I547M, V514M, G502E, G502R, E489K, R450C, P286S, C278Y, G271R, P263L, P256S, D225E, S224P, E219G, and Y105C.

ConSurf analysis highlighted that these conserved residues are critical for maintaining the structural and functional integrity of FGFR2. Mapping of the variants onto protein domains showed that several were distributed across the extracellular Ig-like domains D2 and D3 (residues ~140–368), including P286S, C278Y, G271R, P263L, P256S, D225E, S224P, E219G, and Y105C. No highly conserved variants were observed in the transmembrane region (residues ~369–391). In contrast, the majority of damaging mutations were clustered in the intracellular tyrosine kinase domain (residues ~481–768), including L757S, G690R, P666S, R664W, D644N, Y616C, L572F, L550P, L550V, I547M, V514M, G502E, G502R, E489K, and R450C. This enrichment within the kinase domain supports their potential role in disrupting catalytic activity and downstream signaling.

Project HOPE provided further structural and physicochemical interpretation of these variants. The analysis showed that multiple substitutions involved alterations in residue size, charge, and hydrophobicity, which can collectively impair protein folding, stability, or interactions. In many cases, such as rs1296874666 (L757S) and rs1389986512 (G522E), the mutant residue differed significantly in size or hydrophobicity from the wild-type residue, leading to destabilization of buried structural elements. In rs1224606327 (E731K), located within the tyrosine kinase domain, the substitution of a negatively charged glutamate by a positively charged lysine introduced a charge reversal, disrupting native electrostatic interactions essential for kinase activity. This resolves the earlier contradiction, emphasizing that the mutation introduces a deleterious gain of charge rather than simply removing one. Similarly, mutations such as R450C and D644N resulted in loss of charge and potential disruption of hydrogen bonds or salt bridges, while G690R and M537K introduced buried charged residues that are likely to impair protein folding. Collectively, these findings suggest that the 24 prioritized variants exert pathogenic effects through destabilization of local structure, alteration of domain-specific interactions, and impairment of FGFR2's signaling function.

4. Discussion

Craniosynostosis is a genetic disease related to the mutations in the FGFR2 gene. It is characterized by the premature closure of skull bones, syndactyly of toes and fingers and cleft lip and palate and anomalies of hand [28]. Germinal missense mutations of the FGFR2 gene include craniosynostosis disorders, such as Apert Syndrome, Crouzon Syndrome, Pfeiffer Syndrome and Jackson-Weiss Syndrome have one of the phenotypes that have been stated in FGFR2 mutant patients. These missense mutations produce skeletal syndromes due to the activation of FGFR2 signaling. Somatic missense mutations of FGFR2 occur in different types of cancers such as uterus, breast, lung, gastric ovarian cancer. nsSNPs have the highest effect on protein structure and function between the different mutations. A total of 28,027 SNPs has been reported in fibroblast receptor protein, and a very small amount of work has been performed related to the computational analysis of the FGFR2 for the analysis of the effect of these SNPs, their disease associations, their functional effect or their protein stability impact. This study designed to fulfil this gap, as a novelty extensive in-silico study was carried out to study the influence of the 393 reported nsSNPs by using different tools (stated above). Each tool aided for mining useful conclusions as they catered with the required input. The approaches used in this research concentrate on the connection between the mutations

and their impacts on the protein. A bunch of different tools used makes the outcomes more assured as each tool runs on a different algorithm.

The mutation rs1224606327 (E731K) has been reported as a novel pathogenic mutation in the FGFR2 gene in six Korean patients with craniosynostosis [29] but in this study, it was found to be not damaging by Project Hope. Another mutation rs776587763(C278Y) and rs779326224(P263L) has been previously reported in 30 craniosynostosis patients with 5 Pfeiffer and 25 Crouzon and patients [30]. The loss of C278Y contributes to Pfeiffer and Crouzon syndromes. P263L mutation exists in between the IgII and IgIII domains region, resulting in Apert syndrome. This research also revealed the result of nsSNP (Y105C) which has been already reported in the child of craniosynostosis which was further identified as Crouzon syndrome through experiments [31]. It was the first reported mutation in the first immunoglobulin-like loop (Y105C) and has a clinical significance by the clinical testing method. In this study, the mutation I547M (rs1375821471) was found to pathogenic by using in-silico approaches and this mutation has been reported as the somatic missense mutation and causes uterus cancer.

5. Conclusions

Using computational tools has become a significant method nowadays for the identification of diseases associated with SNP. In this present research, we studied the impact of SNPs on the structure and function of the FGFR2 gene by using different tools like SIFT, PolyPhen-2, PROVEAN, SNAP2, SNPs&GO, PANTHER, PhD-SNP, PMut, I-Mutant3.0, ConSurf Server and Project Hope. From 393 nsSNPs, 24 nsSNPs with positions rs1296874666(L757S), rs1311258996(G690R), rs868564661(P666S), rs113014479(R664W), rs138631412(D644N), rs762545440(Y616C), rs371854567(L572F), rs755350933(L550P), rs1215036637(L550V), rs1375821471(I547M), rs1262068931(V514M), rs1173224369(G502E), rs1384270958(G502R) ,rs1400354610(E489K) ,rs536181987(R450C) ,rs778094013(P286S) ,rs776587763(C278Y) ,rs183250272(G271R ,rs779326224(P263L ,rs1364891245(P256S) ,rs537758598(D225E) ,rs1216426276(S224P) ,rs1390042185(E219G) ,rs1434545235(Y105C) were foreseen to be highly deleterious mutations in the FGFR2 gene. These mutations may affect the physicochemical characteristics of the protein which includes size, charge, hydrophobicity, and also disturbs the stability of the protein and leads to the loss of protein's structure and function. It also might be essential for the identification of several disorders such as Crouzon, Apert, Jackson-Weiss and Pfeiffer Syndrome, breast, gastric, ovarian, lung, head and neck cancers. Some of these mutations such as rs113014479, rs1375821471, rs776587763, rs779326224 and rs1434545235 have been reported earlier and were confirmed by in silico approaches (stated above) but in this research, some new SNPs were also predicted as disease-associated. Computational tools are very vigorous, particularly when they presented with the right information and utilized by professionals. Still, these tools have the restrictions, which means that the data they give us needs verification using many approaches like functional studies. So, further researches and wet lab investigation are recommended to measure the impinging of these variations to support our results. As a result, we expect our analysis will give beneficial evidence to the researchers to do more study in the future which can be a smash for the identification of different diseases related to FGFR2 and prepare precise medicines according to the human genome.

References

1. Turner N, Grose R. Fibroblast growth factor signalling: From development to cancer. Nat Rev Cancer 2010; 10: 116–129.
2. Meyers GA, Day D, Goldberg R, et al. FGFR2 exon IIIa and IIIc mutations in Crouzon, Jackson-Weiss, and Pfeiffer syndromes: Evidence for missense changes, insertions, and a deletion due to alternative RNA splicing. Am J Hum Genet 1996; 58: 491–498.
3. Gorry MC, Preston RA, White GJ, et al. Crouzon syndrome: Mutations in two spliceoforms of FGFR2 and a common point mutation shared with jackson - weiss syndrome. Hum Mol Genet 1995; 4: 1387–1390.
4. Malcolm S, Reardon W. Fibroblast growth factor receptor-2 mutations in craniosynostosis. In: Annals of the New York Academy of Sciences. Blackwell Publishing Inc., pp. 164–170.
5. Chen P, Zhang L, Weng T, et al. A Ser252Trp Mutation in Fibroblast Growth Factor Receptor 2 (FGFR2) Mimicking Human Apert Syndrome Reveals an Essential Role for FGF Signaling in the Regulation of Endochondral Bone Formation. PLoS One 2014; 9: e87311.
6. Nachman MW. Single nucleotide polymorphisms and recombination rate in humans. Trends in Genetics 2001; 17: 481–485.
7. Azoury SC, Reddy S, Shukla V, et al. Fibroblast growth factor receptor 2 (FGFR2) mutation related syndromic craniosynostosis. International Journal of Biological Sciences 2017; 13: 1479–1488.
8. Mazhar A, Jamil F, Bashir Q, et al. Genetic variants in FGFR2 and TNRC9 genes are associated with breast cancer risk in Pakistani women. Mol Med Rep 2016; 14: 3443–3451.
9. Thomas PE, Klinger R, Furlong LI, et al. Challenges in the association of human single nucleotide polymorphism mentions with unique database identifiers. BMC Bioinformatics 2011; 12: S4.
10. Chen X, Sullivan PF. Single nucleotide polymorphism genotyping: Biochemistry, protocol, cost and throughput. Pharmacogenomics Journal 2003; 3: 77–96.
11. Chasman D, Adams RM. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: Structure-based assessment of amino acid variation. J Mol Biol 2001; 307: 683–706.
12. Ferrer-Costa C, Orozco M, De La Cruz X. Sequence-based prediction of pathological mutations. Proteins Struct Funct Genet 2004; 57: 811–819.
13. Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res 2010; 38: W214-20.
14. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003; 31: 3812–3814.
15. B R, C GPD. Path to Facilitate the Prediction of Functional Amino Acid Substitutions in Red Blood Cell Disorders – A Computational Approach. PLoS One 2011; 6: e24607.
16. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013; 0 7: Unit7.20.
17. Narayana Swamy A, Valasala H, Kamma S. In silico evaluation of nonsynonymous single nucleotide polymorphisms in the ADIPOQ gene associated with diabetes, obesity, and inflammation. Avicenna J Med Biotechnol 2015; 7: 121–127.
18. Hecht L, Wass J, Kelly L, et al. SNAP-Ed Steps to Health Inspires Third Graders to Eat Smart and Move More. J Nutr Educ Behav 2013; 45: 800–802.
19. Calabrese R, Capriotti E, Fariselli P, et al. Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat 2009; 30: 1237–1244.
20. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 2006; 22: 2729–34.
21. Capriotti E, Fariselli P, Calabrese R, et al. Predicting protein stability changes from sequences using support vector machines. Bioinformatics 2005; 21 Suppl 2: ii54-8.
22. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research 1997; 25: 3389–3402.
23. Ferrer-Costa C, Lluis Gelpí J, Zamakola L, et al. PMUT: a web-based tool for the annotation of pathological mutations on proteins. Bioinforma Appl NOTE 2005; 21: 3176–3178.
24. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 2005; 33: W306-10.

25. Glaser F, Pupko T, Paz I, et al. ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics 2003; 19: 163–164.
26. Pupko T, Bell RE, Mayrose I, et al. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. Bioinformatics 2002; 18 Suppl 1: S71-7.
27. Venselaar H, te Beek TAH, Kuipers RKP, et al. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics; 11. Epub ahead of print 2010. DOI: 10.1186/1471-2105-11-548.
28. Lin Y, Ai S, Chen C, et al. Ala344Pro mutation in the FGFR2 gene and related clinical findings in one Chinese family with Crouzon syndrome. Mol Vis 2012; 18: 1278–1282.
29. Park J, Park O-J, Yoon W-J, et al. Functional characterization of a novel FGFR2 mutation, E731K, in craniosynostosis. J Cell Biochem 2012; 113: 457–464.
30. Kress W, Collmann H, Büsse M, et al. Clustering of FGFR2 gene mutations in patients with Pfeiffer and Crouzon syndromes (FGFR2-associated craniosynostoses). Cytogenet Cell Genet 2000; 91: 134–137.
31. Pulleyn LJ. Spectrum of craniosynostosis phenotypes associated with novel mutations at the fibroblast growth factor receptor 2 locus. Eur J Hum Genet 1996; 4: 283–291.