

Current Trends in OMICS (CTO)

Volume 2 Issue 2, Fall 2022


ISSN(P): 2790-8283 ISSN(E): 2790-8291

Homepage: <https://journals.umt.edu.pk/index.php/cto>



Article QR



- Title:** **In Silico Analysis of Functional SNPs in Human CRB2 GENE Associated with Focal Segmental Glomerulosclerosis**
- Author (s):** Imran Sattar¹, Hajira Maqbool¹, Mishel Zainab², Muhammad Shahbaz Aslam³, Iram Gull³, Imran Tipu^{1*}
- Affiliation (s):** ¹University of Management and Technology Lahore, Pakistan
²Mayo Hospital, Lahore
³University of the Punjab Lahore, Pakistan
- DOI:** <https://doi.org/10.32350/cto.22.03>
- History:** Received: November 14, 2022, Revised: November 25, 2022, Accepted: December 05, 2022
- Citation:** Sattar I, Maqbool H, Zainab M, Aslam MS, Gull I, Tipu I. In silico analysis of functional SNPs in human CRB2 GENE associated with focal segmental glomerulosclerosis. *Curr Trend OMICS*. 2022;2(2):00–00. <https://doi.org/10.32350/cto.22.03>
- Copyright:** © The Authors
- Licensing:**  This article is open access and is distributed under the terms of [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)
- Conflict of Interest:** Author(s) declared no conflict of interest



A publication of

The Department of Life Sciences, School of Science
University of Management and Technology, Lahore, Pakistan

In Silico Analysis of Functional SNPs in Human CRB2 GENE Associated with Focal Segmental Glomerulosclerosis

Imran Sattar¹, Hajira Maqbool¹, Mishel Zainab,² Muhammad Shahbaz Aslam³, Iram Gull³, and Imran Tipu^{1*}

¹ Department of Life Sciences University of Management and Technology
Lahore

² Mayo Hospital, Lahore

³ Department of Biochemistry and Biotechnology, University of the Punjab
Lahore

Abstract

The CRB2 gene contributes to the onset hereditary focal segmental glomerulosclerosis, which damages the kidney, particularly the glomerulus, and results in chronic kidney failure. The pathogenicity of CRB2 non-synonymous single nucleotide polymorphisms (nsSNPs) was predicted in this work using a variety of bioinformatics techniques for mutation analysis. For this purpose, 1201 nsSNPs from dbSNP-NCBI were retrieved for the analysis, and 20 were predicted deleterious. The 20 missense variants are G349D, C629S, R534W, G178D, C620Y, C620S, R628C, R633G, R633W, E643A, T841M, R960S, R960C, P1064T, P1064S, N800K, G1088D, T1187P, R1249Q, and R1249P. SIFT, PROVEAN, Mutation Assessor, and PANTHER, four sequence homology-based approaches, suggested that one variation, R960S, was benign and that 19 SNPs were deleterious. Six supervised-based approaches, including SNAP2, MutPred2, SuSPect, PhD-SNP, SNPs&Go, and PMut, predicted 14 SNPs as detrimental in all six applied methods. Hence, 13 variations were identified harmful by one structure-based technique, Poly-Phen, and two consensus-based methods, Meta-SNP and Predict-SNP. The CRB2 protein's mutant and wild-type structures were predicted using a laser. Following PyMol 5 structural analysis, the mutations rs879255250 cysteine at 620 changes to tyrosine, variation rs879255250 glycine at 178 changes to aspartate, and variant rs1322315181 glycine at 178 changes to aspartate arginine changes to proline at position 1249, and cysteine changes to serine at position 620, with variants rs147412276 and rs868484209. Proline at position 1064 alterations to serine were expected to be extremely harmful and resulted in structural changes in the protein, which may be candidates for the FSGS

* Corresponding Author: imran.tipu@umt.edu.pk

etiology. These CBR2 nsSNPs may, thus, be the possibilities for diagnostic genetic screening and therapeutic molecular targeting.

Keywords: CRB2, FSGS, kidney disease, missense variations, nsSNP, missense variations, pathogenic mutation

Introduction

The glomeruli are an essential component of kidneys as they filter and eliminate the blood's waste, however, when scar tissues start to form in these specific areas due to certain pathogens or other factors, it leads to a condition known as focal segmental glomerulosclerosis (FSGS). This is the most serious possible state and frequently results in kidney failure for which the only options are a kidney dialysis or kidney transplant[1]. From 1.4-21 instances per million people are thought to occur annually with FSGS. Approximately 7-10% of children and 20-30% of adults with nephrotic syndrome have FSGS at some point in their lives. FSGS can develop at any age[2]. Males have a 1.5–2 times higher incidence of FSGS in adults than females, making them more likely to have syndrome. Nowadays, FSGS which can come from a multitude of nephrotic syndromes in both children and adults is one of the leading causes of kidney failure [3]. When FSGS progresses as a result of multiple factors, including infections, especially infections of the podocytes, other diseased conditions such as diabetes and sickle cell disease, or even multiple kidney disorders, this type of FSGS is known as secondary FSGS because it was brought on by a secondary cause [4]. Glomeruli are responsible for the effective and efficient filtration of blood, and also responsible for different genes that encode different functions of the kidney, such as those attempting to control the activity of the slit diaphragm, gene of cell membrane proteins, nuclear proteins, cytoskeleton proteins, and other proteins involved in the function of critical organelles such as mitochondria and lysosomes [4, 5]. Numerous genes are connected to FSGS, including those that control the activity of proteins such as transient receptor potential action channel-6 and podocin alpha-actinin, which are crucial for podocyte function. Transient Receptor Protein Cation Channel Subfamily C Member 6 (TRPC6) is a key gene linked to FSGS - 2, which is driven by pathways of primary focal segmental glomerulosclerosis and cell junction organization [6].

The critical gene that is involved in the development of FSGS-3, often referred to as FSGS 3, is CD2 Associated Protein (CD2AP) [7]. Renal

failure and proteinuria are two of their prevalent characteristics, and it is known to affect the heart and kidney tissues. FSGS-4 is a chronic disease that attacks the kidney and cardiac tissues [8]. Stage 5 chronic kidney disease and FSGS are two of the disorder's linked phenotypic outcomes, and APOL1 (Apo lipoprotein L 1) is the main gene implicated in their development. Additionally, the Nephric/Neph 1 signaling pathway in the renal podocytes and the main FSGS pathway are known to be involved in the formation of FSGS 5, and INF2 (Inverted Formin 2) is the gene implicated in its growth [9]. The essential gene in this type is Myo1 E (Myosin 1E), which is sometimes confused with Myo1 C (Myosin 1C) because the mutation in Myo1 C causes focal segmental glomerulosclerosis 6 and it has genetic traits of proteinuria and edema in some patients. However, unlike other types of FSGS, focal segmental glomerulosclerosis 6 is associated with kidney and heart tissues. FSGS 7 is associated with diseases such as Diffuse Mesangial Sclerosis (DMS), a serious condition in which the mesangial matrix starts to worsen without cellular mesangial proliferation, ultimately obstructing the function of the glomeruli and giving rise to the term infantile nephrotic syndrome (INS) in a high percentage of children [10]. Renal hypoplasia and stage 5 chronic kidney disease are the phenotypes of focal segmental glomerulosclerosis, which is caused by the PAX2 (Paired Box 2) gene. In addition, connected pathways and associated tissues are comparable to those of other kinds [11].

Oligomeganephronia and other nephrotic disorders are associated with FSGS-8, with significant characteristics including proteinuria and other nephrotic diseases, this kind of FSGS invades the heart and renal tissues much like other varieties of FSGS do. FSGS-9 shares similarities with FSGS 8 in terms of nephrotic disorders, such as oligomeganephronia and other nephrotic disorders. However, FSGS 9 differs from FSGS 8 in that CRB2 (Crumbs Cell Polarity Complex Component 2) gene plays a role in the onset and progression of the disease, which also affects kidney and heart tissues [12]. A homozygous or compound heterozygous mutation in the CRB2 gene (609720) on chromosome 9q33 causes FSGS9. A biallelic mutation in the CRB2 gene can also result in ventriculomegaly with cystic kidney disease (VMCKD; 219730), a far more serious condition [13].

The crumbs homolog 2 gene is located on the long arm of chromosome 9 in humans. Several proteins of the cell polarity complexes are expressed by this gene; however, its two key isoforms are isoform 1 and isoform 2,

each of which comprises 1176 amino acids. Isoform 1 is a suspected type I transmembrane protein [14]. Additionally, it is known that the human mutation in CRB2 (Crumbs Cell Polarity Complex Component 2) contributes to the onset and development of (FSGS9 and is linked to ventriculomegaly, a condition in which ventricles appear larger during prenatal ultrasound due to the buildup of CSF in empty spaces, and cystic kidney diseases. Clinical analysis of kidney tissues containing mutant CRB2 revealed aberrant glomerulus formation, monocyte foot process abnormalities, and basement membrane insufficiency. The foot process is made up of vesicular-like structures and apical membrane protrusions [15]. The disorders in apical membrane demarcation and protein signaling are associated with membrane polarity defects because the apical membrane, which serves as the first barrier for anion exchange and is affected by phosphorus and other solutes in CRB2-null monocytes, revealed localized projections in Bowman's space [16, 17].

Material and Methods

Retrieval of Protein Sequence

A number of 1285 amino acid sequence of the protein encoded by the CRB2 gene was obtained in FASTA format from NCBI (<https://www.ncbi.nlm.nih.gov/protein>). For structural details, this protein sequence was entered into iTASER. The protein sequence was then mutated at 11 points, notably G at position 178 and 349 into D, R at position 534 into W, C at position 620 into Y, C at position 629 into S, R at position 633 into W, E at position 643 into A, P at position 1064 into S, G at position 1088 into D, T at position 1187 and R at position 11249 into P, before running through iTASER because two predicted harmful mutations occurred at the same site, these two changes, C at position 620 into S and R at position 633 into G, were modified in the protein sequence and ran through iTASER independently.

Retrieval of Missense SNPs

Missense variations of the CRB2 gene were extracted from Ensembl utilizing the dbSNP database (<https://asia.ensembl.org/index.html>). SNPs obtained from Ensembl using various criteria, input the CRB2 protein sequence into Ensembl, which predicted 1201 missense SNPs. Then, additional characteristics were used to distinguish between disease-causing harmful SNPs and missense SNPs with unknown significance.

Identification of Damaging Missense SNPs

Identifying the most dangerous missense SNPs in CRB2 that impact protein structure and function is a difficult undertaking. For this goal, many bioinformatics tools based on distinct methods were employed. There are several computational-based techniques available for predicting the consequences of amino acid substitution on protein function. Based on their technique, tools were classified into four types: homology-based approaches, supervised methods, consensus methods, and structural methods. These computational tools predicted the functional implications of mutations in protein sequences that narrowed down the harmful symptoms.

Sequence Homology-Based Method

Twenty filtered experimental datasets of nsSNPs were classified using these sequence homology-based methods for instance; SIFT, PROVEAN, Mutation Assessor, and PANTHER. With the use of these bioinformatics techniques, it is possible to forecast variations that have significant phenotypic effects, such as those that are functionally detrimental or disease-related. It was determined from the 242 CRB2 protein nsSNPs, which may contribute to focal segmental glomerulosclerosis by analysis (FSGS).

Method Based on Supervised Learning

Three main types of algorithms, including neural network models, random forests, and support vector machines, were used in the supervised learning method. These techniques allow the prediction of the deleteriousness of an amino acid change. Tools in this area, including SNPs and GO, SusPect, MutPred, PMut, PhD-SNP, and SNAP-2, were used to find variations that needed additional screening. With a score productivity of 82%, the SNPs&GO web browser assisted in predicting single point mutation in protein sequence to result in illness state in patients.

Prediction of the Effects of nsSNPs on the Protein Structure by MutPred2

A machine-learning program called Mutpred2 (<http://mutpred2.mutdb.org/index.html>) was utilized to further examine the potential consequences of nsSNPs on structure as a result of amino acid

substitution. To forecast the pathogenicity of amino acid alterations, the MutPred2 tool was combined with genetic and molecular data.

Consensus and Structure-Based Methods

Although consensus-based approaches like Condel and PON-P2 and structure-based tools SDM, Fold-X, and PoPMusic are from separate categories and were intended to be used independently, utilizing the tools required in the PDB four-letter code. There is no PDB code available since CRB2's crystal structure is not in the Protein Data Bank. Therefore, we only employed programs that demanded the protein sequence be in FASTA format. We used these three techniques for future SNP predictions since two consensus-based tools—PRIDICT-SNP and Meta-SNP—fulfilled the criterion, as did one structure-based tool—Poly-Phen2. Alternative analyzers, PhD-SNP, PPH-2, and SNAP, each using decision structures created using various approaches, were utilized to predict amino acid substitutions. Out of the remaining tools, SIFT and PANTHER were only used as transformational data but MAPP was also considered for the differences in physicochemical properties between native and mutant amino acids.

Methods Based on Protein Structure and Function

Our protein structure, which was obtained from the iTaser findings and was in PDB format, was visualized with PyMOL. It was useful for labeling and coloring chains or specific residues, as well as for seeing the amino acid sequence and each amino acid's location. To compare the before and after effects of the mutation at a particular site, we used PyMOL to modify the protein sequence.

Results

Prediction of Missense SNPs by Sequence Homology-Based Methods

SIFT, PROVEAN, MUTATION ASSESSOR, and PANTHER were four homology-based approaches that were used to predict different consequences for various 20 missense SNPs. SIFT identified 06 SNPs as neutral and 14 SNPs as harmful. Therefore, 17 SNPs were labelled as harmful by PROVEAN, whereas 3 were neutral. Moreover, 19 SNPs were labelled as harmful by the mutation assessor, and one was neutral. While PANTHER classified each of the 20 SNPs as harmful.

Table 1. Sequence Homology-Based Techniques, such as SIFT, PROVEAN, Mutation Assessor, and Panther, One May Predict Missense SNPs

Variant ID	Mutations	SIFT	PROVEAN	Mutation Assessor	PANTHER
rs1322315181	G178D	D	D	D	D
rs1564372397	G349D	D	D	D	D
rs759002273	R534W	D	D	D	D
rs879255250	C620Y	D	D	D	D
rs879255250	C620S	D	D	D	D
rs202128397	R628C	D	D	D	D
rs879255252	C629S	D	D	D	D
rs730880377	R633G	D	D	D	D
rs730880377	R633W	D	D	D	D
rs730880300	E643A	D	D	D	D
rs765676223	N800K	D	D	D	D
rs1554784754	T841M	N	D	D	D
rs768010757	R960S	N	N	N	D
rs768010757	R960C	N	N	D	D
rs868484209	P1064T	N	D	D	D
rs868484209	P1064S	N	D	D	D
rs766470795	G1088D	D	D	D	D
rs1554785663	T1187P	D	D	D	D
rs147412276	R1249Q	N	N	D	D
rs147412276	R1249P	D	D	D	D

D stands for deleterious and N for Neutral

The first three techniques scores were used to determine, whether a mutation is harmful or tolerable. N stands for neutral and D for detrimental.

Prediction of Missense SNPs by Supervised Based Methods

These 20 missense SNPs were subsequently assessed using supervised-based techniques, such as SNAP2, PMut, PhD-SNP, SNP&Go, MuTPred-2, and SuSPect. MutPred-2 predicted 14 harmful SNPs and 06 neutral SNPs, PhD-SNP predicted 14 deleterious SNPs, and 06 neutral SNPs, SNP&Go

projected 16 deleterious SNPs and 04 neutral SNPs, SuSPect identified 06 deleterious SNPs and 14 neutral SNPs, and SNAP2 predicted 18 toxic SNPs and 02 neutral SNPs. PMut identified 12 SNPs as neutral mutations and 8 SNPs as detrimental mutations. All 13 SNPs predicted by supervised-based methods were harmful in all applicable 06 SNPs. 07 SNPs, including R628C rs202128397, N800K rs765676223, T841M rs1554784754, R960S rs768010757, R960C rs768010757, and R1249Q rs147412276, were all predicted as neutral in 05 out of 06 supervised-based methods that were used. Therefore, we did not include these 7 SNPs for further analysis, and the remaining 13 missense SNPs were assessed further using consensus-based techniques.

Table 2. Predictions by all supervised-based methods

Variant ID	Mutations	SNAP2	Mutpre d-2	PhD SNP	SNP & GO	SuSPect	PMut
rs1322315181	G178D	D	D	D	D	D	D
rs1564372397	G349D	D	D	D	D	D	D
rs759002273	R534W	D	D	D	D	N	N
rs879255250	C620Y	D	D	D	D	D	D
rs879255250	C620S	D	D	D	D	D	D
rs202128397	R628C	D	N	N	D	N	N
rs879255252	C629S	D	D	D	D	D	D
rs730880377	R633G	D	D	D	D	N	N
rs730880377	R633W	D	D	D	D	N	D
rs730880300	E643A	D	D	D	N	N	N
rs765676223	N800K	D	N	N	N	N	N
rs1554784754	T841M	N	N	N	D	N	N
rs768010757	R960S	D	N	N	N	N	N
rs768010757	R960C	D	N	N	N	N	D
rs868484209	P1064T	N	D	D	D	N	N
rs868484209	P1064S	D	D	D	D	N	N
rs766470795	G1088D	D	D	D	D	D	D
rs1554785663	T1187P	D	D	D	D	N	N
rs147412276	R1249Q	D	N	N	D	N	N
rs147412276	R1249P	D	D	D	D	N	N

For each SNP, different outcomes were predicted by each supervised-based technique. Hence, all 13 SNPs were identified as harmful by all six techniques used. 07 SNPs, including R628C, N800K, T841M, R960S, R960C, and R1249Q, were all or 05 out of 06 employed techniques projected as neutral.

Consensus and Structure-Based Methods

The prediction of missense SNPs can be done using consensus or structure-based techniques, respectively. The majority of the tools used by both approaches utilized PDB format files for predictions, therefore, only one structure-based approach performed the required predictions. The FASTA format of the protein sequence was used by Poly-Phen and the consensus-based techniques Meta-SNP and Predict-SNP. Although consensus and structure-based approaches were intended to be applied independently, PDB four-letter codes were necessary to run consensus-based bioinformatics tools like Condel and PON-P2 and structure-based tools like SDM, Fold-X, and PoPMusic. Our protein of interest, CRB2, is not in the Protein Data Bank. Therefore, we utilized those technologies that needed protein sequences in FASTA format. The condition was answered by two consensus-based tools, PRIDICT-SNP and Meta-SNP, and one structure-based tool, Poly-Phen2, thus we employed all three of them. Ten SNPs were identified by Meta-SNP as harmful, and ten SNPs as neutral.

15 SNPs were identified as harmful and 5 SNPs as neutral by PRIDICT-SNP. The two benign SNPs were classified as neutral since the structural base approach Poly-Phen projected that 18 SNPs were likely to be harmful and only 02 SNPs were benign. Overall, consensus and structure-based prediction approaches identified 12 SNPs, including the mutations G178D rs1322315181, G349D rs1564372397, R534S rs759002273, C620Y rs879255250, R633G rs730880377, 633W rs730880377, E643A rs730880300, and P1064S rs.

The mutations G1088D rs 868484209, T1187P rs 1554785663, R1249P rs 147412276, and G1088D rs 868484209 were expected to be harmful.

For additional analysis, consensus and structure-based approaches that predicted neutral missense SNPs were discarded. PyMol was used which only examined SNPs that were anticipated to be harmful.

Table 3. Predictions of SNPs by consensus and structural-based methods

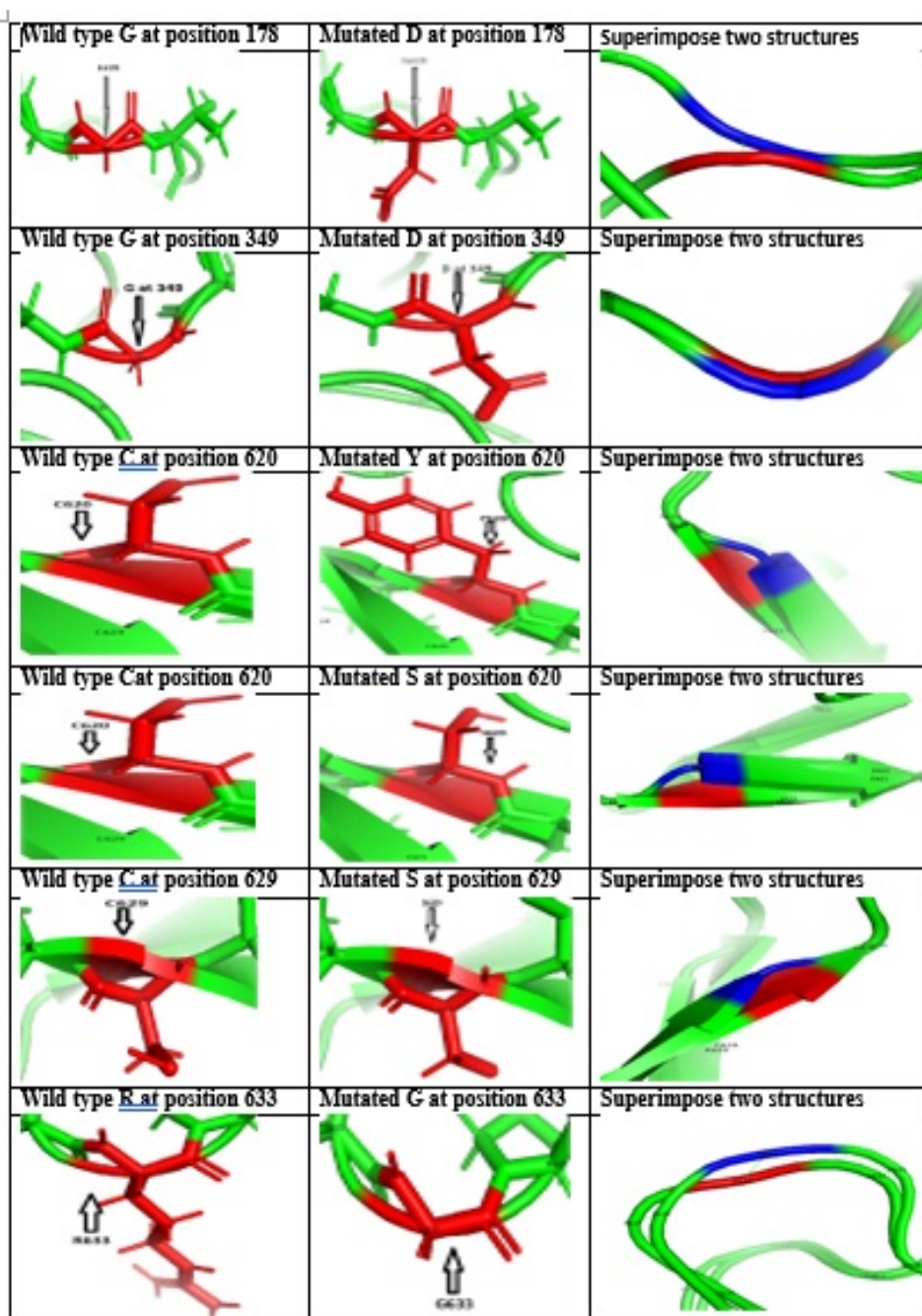
Variant ID	Mutations	Meta-SNP prediction	Predict-SNP prediction	Poly-Phen prediction
rs1322315181	G178D	Disease	Deleterious	Probably damaging
rs1564372397	G349D	Neutral	Deleterious	Probably damaging
rs759002273	R534W	Disease	Deleterious	Possibility damaging
rs879255250	C620Y	Disease	Deleterious	Probably damaging
rs879255250	C620S	Disease	Deleterious	Probably damaging
rs879255252	C629S	Disease	Deleterious	Probably damaging
rs730880377	R633G	Disease	Deleterious	Probably damaging
rs730880377	R633W	Disease	Deleterious	Probably damaging
rs730880300	E643A	Neutral	Deleterious	Probably damaging
rs868484209	P1064T	Neutral	Neutral	Benign
rs868484209	P1064S	Neutral	Deleterious	Probably damaging
rs766470795	G1088D	Disease	Deleterious	Probably damaging
rs1554785663	T1187P	Neutral	Deleterious	Probably damaging

Structure Analysis of Mutant and Wild Type Protein by PyMOL

The PyMOL tool used for a detailed investigation of the structural change was caused by altered amino acids. The results of PyMOL were used to compare the normal protein structure at a certain site to the mutant structure at the same position as shown in Figure 1.

Seven variations were predicted to be neutral by one structure-based technique, poly-Phen, and 13 variants (G178D, R534W, G349D, C620Y, C629S, C620S, R633G, E643A, R633W, P1064S, T1187P, G1088D, and R1249P) to be detrimental by two consensus-based methods, meta-SNP and predict-SNP. In all thirteen used tools, these 13 variations were identified as harmful.

The variation rs1322315181 participates in the coil formation by having both amino acids G and D at position 178 are deeply buried in the same location. The gain of a disulfide bond at position C181 in the transmembrane protein might result in the changes. It potentially changes the function at position 178 by significantly, altering the coil's conformation. Version rs1564372397 at 349, the G becomes a D.

Figure 1. Structural Analysis

Wild type <u>R</u> at position 633	Mutated <u>W</u> at position 633	Superimpose two structures
Wild type <u>E</u> at position 643	Mutated <u>A</u> at position 643	Superimpose two structures
Wild type <u>R</u> at position 534	Mutated <u>W</u> at position 534	Superimpose two structures
Wild type <u>P</u> at position 1064	Mutated <u>S</u> at position 1064	Superimpose two structures
Wild type <u>G</u> at position 1088	Mutated <u>D</u> at position 1088	Superimpose two structures
Wild type <u>T</u> at position 1187	Mutated <u>P</u> at position 1187	Superimpose two structures



Both D and G participated in the development of coils and were highly exposed residues in the structure, which may affect transmembrane proteins, change the disulfide linkage at C344, change the metal binding with, and increase sulfation at Y350. The structure at position 349 following the mutation did not change. Variant rs759002273 converts R to W at position 534, both of which participate in strand formation as buried residues with no discernible modifications to protein structure. Variant rs879255250 switches C to Y, forming strands and deeply buried residues in the protein that may disrupt metal binding, transmembrane protein structure, loss of disulfide linkage, and gain of the catalytic site at C620. Due to the variant rs879255250, both stranded and deeply buried residues in the protein shift from C to S, altering metal binding, a transmembrane protein, and the disulfide bond at C620.

In the variant rs730880377, R at position 633 switches to G, causing both highly exposed and coiling residues in the protein to become disulfide-linked to C631, altering the transmembrane protein, and changing metal binding. As a result of the variant rs730880377, R at position 633 changes to W, and both residues in the protein form a coil and are extensively exposed. This may have changed the transmembrane protein, gained a loop, and lost the disulfide connection at position 629. Variant rs730880300 E at 643 changes to A, both residues in the protein form a coil, and A is substantially exposed, while E is just minimally exposed. This alteration may affect metal binding, gain a disulfide bond at C640, and modify the transmembrane protein. Variant rs868484209 P at position 1064 changes to S, causing both groups in the protein to form a coil and become highly exposed, perhaps altering metal binding, gaining a disulfide bond at C1065, and altering transmembrane protein. Variant rs766470795 G at position 1088 changes to D, causing both residues in the protein to form a coil and become extensively exposed, perhaps resulting in a loop gain, changed transmembrane protein, changed metal interaction, and achievement of

disulfide linkage with C1091. The variant rs1554785663 T at location 1187 transforms to P both residues in the protein form strand and frequently exposed, which might also gain intrinsic disorder, rise strand, modify transmembrane protein, lose disulfide linkage at C1188, affect metal binding, and lose catalytic site at C1182. Variant rs147412276 R at position 1249 changes into P both residues in protein form strand and exposed residues, which may result in loss of helix, the gain of ADP-ribosylation at R1252, loss of pyrrolidine carboxylic acid at Q 1253, and loss of methylation at K1250. These changes in residues lead to changes in protein structure, which in turn affects the protein functions.

Discussion

One of the most typical glomerular abnormalities that result in end-stage kidney disease is focal segmental glomerulosclerosis (FSGS). High-grade proteinuria that is accompanied by lesions of localized and segmental glomerular sclerosis and foot-process effacement is the hallmark of the clinicopathologic condition known as FSGS. Sclerosis develops into a more widespread and global pattern as the condition worsens. Primary (idiopathic) and secondary versions are the two different kinds of FSGS. Primary FSGS's precise etiology is yet unknown. The established etiologies of secondary FSGS include genetic alterations in proteins related to podocytes, infections, and medication toxicity. A member of a family of proteins that make up the crumbs cell polarity complex is encoded by the CRB2 gene. The fetal eye, placenta, lung, and retinal pigment epithelium/choroid of humans are the key tissues where the CRB2 gene is expressed. Recent investigations have suggested that phenotypic diversity contributes to a severe condition in addition to renal impairment caused by the CRB2 mutation. In early embryonic development in mammals, it is believed that this family's members are involved in various cellular activities. The apicobasal polarity of embryonic epithelial cells is controlled by a related protein in *Drosophila*. This gene's mutations have been linked to cystic kidney disease, focal segmental glomerulosclerosis 9, and ventriculomegaly [[16](#), [18](#)].

SNPs are the most common kind of genetic variation in humans, accounting for over 90% of all known mutations. While many of these changes are harmless, certain variations do influence gene expression or the function of translated proteins. Such SNPs have a profound phenotypic influence on the protein, resulting in the emergence of a variety of illnesses.

Due to the rapid expansion of known SNPs, it is becoming increasingly challenging to establish linkages between SNPs and linked disorders. As a result, computationally based prediction techniques have become crucial for the first investigation of SNPs and their prediction for experimental characterization [19].

The majority of these computational techniques are intended to determine whether a certain mutation is benign or pathogenic based on several metrics derived from evolutionary and structural features. These bioinformatics tools rely heavily on machine learning methods to forecast their outcomes based on annotated mutation datasets [20].

The cost-effective, rapid, and practical alternative is thorough *in silico* investigation of mutation consequences. Numerous proteins, including apolipoprotein E3-40, quinone oxidoreductase 1, melanocortin-3-receptor, and connexin 26, contain high-risk nsSNPs that have already been identified using computational methods [21, 22].

By using several filters, we were able to get SNPs from Ensembl. We input the CRB2 protein sequence, which predicted 1201 missense SNPs. Then, additional criteria were used to distinguish between missense SNPs of uncertain significance and those that cause disease or are harmful. We obtained 20 missense SNPs for further investigation after establishing several interest-related factors. Application of homology-based techniques such as SIFT, PROVEAN, Mutation Assessor, and PANTHER was done individually. Panther predicted all 20 SNPs to be harmful, SIFT projected 17 variations to be harmful, PROVEAN predicted 17, Mutation Assessor predicted 19, and SIFT predicted 17 variants to be harmful. Out of the total of four tools applied, just one mutation, R changing into S at position 960 with rs number rs768010757, is projected as neutral in the three homology-based techniques.

These 20 missense SNPs were subsequently assessed using supervised algorithms such as SNAP2, MuTPred-2, PhD-SNP, SNP&Go, SuSPect, and PMut. MutPred-2 predicted 14 deleterious SNPs and 06 neutral SNPs, PhD-SNP predicted 14 deleterious SNPs and 06 neutral SNPs, SNP&Go predicted 16 deleterious SNPs and 04 neutral SNPs, SuSPect predicted 06 deleterious SNPs and 14 neutral SNPs, and PMut predicted 08 deleterious SNPs and 12 neutral mutations. In all of the 06 SNPs that were applied, 13 SNPs were predicted by all supervised-based methods to be harmful. 07

SNPs, including rs202128397 for R628C, rs765676223 for N800K, rs1554784754 for T841M, rs768010757 for R960S, rs768010757 for R960C, and rs147412276 for R1249Q, were predicted as neutral in all or 05 out of 06 supervised-based methods that were used. We, therefore, decided not to continue to analyze these 7 SNPs, and the remaining 13 missense SNPs were instead.

Although PDB four-letter codes were necessary for the operation of consensus and structure-based bioinformatics tools like Condel and PON-P2 as well as SDM, Fold-X, and PoPMusic, these techniques were not deployed as planned [23]. CRB2, our intended protein, is not included in the Protein Data Bank. To employ the technologies, the protein sequence has to be in FASTA format. We used three tools since they all met the requirements: PRIDICT-SNP, Meta-SNP, and Poly-Phen2, which is a structure-based tool. Meta-SNP projected that 10 SNPs would be harmful, whereas the other 10 SNPs would be beneficial. 15 SNPs were deemed deleterious by PRIDICT-SNP, whereas 5 were deemed benign. Polystructural Phen's base approach identified 18 SNPs as potentially harmful and 2 as benign, therefore these 2 benign were considered as neutral. Consensus and structure-based techniques projected that 12 SNPs will result in harmful mutations.

The 1285 amino acid long protein sequence for the CRB2 protein was downloaded from NCBI in FASTA format. The structural details of this protein sequence were checked using iTASER. Then, we made 11 mutations to this protein sequence: G at position 178 became D, G at position 349 became D, R at position 534 became W, C at position 620 became Y, C at position 629 became S, R at position 633 became W, E at position 643 became A, P at position 1064 became S, G at position 1088 became D, T at position 1187 became P, and R at position 11249 became P. We then ran this sequence through iTASER. Because two projected harmful mutations were at the same location, the protein sequences for these two mutations—C at position 620 into S and R at position 633 into G—were modified once more, and then each was subjected to an individual iTASER run. The structural instability and changed conformations of these three mutants as shown by molecular dynamics simulation supported our hypotheses made with the help of bioinformatic techniques. The pathogenicity of the human CBR2 nsSNPs can be accurately predicted by our computational investigation, there is no doubt about it.

Conclusion

The CRB2 gene has a crucial role in the start and progression of FSGS 9. Missense variations account for around 90% of protein-affecting mutations. Using a variety of publically accessible computational techniques, this study identified 20 nsSNPs out of 1201 harmful in the CRB2 gene and predicted 12 nsSNPs namely G178D rs number rs1322315181, R534S rs number 759002273, G349D rs number 1564372397, C620Y rs number 879255250, R633G rs number 730880377, C620Y rs number 879255250, 633W rs number 730880377, P1064S rs number 868484209, E643A rs number 730880300, G1088D rs number 766470795, T1187P rs number 1554785663, and mutation R1249P rs number rs147412276 were predicted as deleterious or as pathogenic according to the consensus and structural based method. These are regarded as extremely harmful since they have a destabilizing effect on the CRB2 protein structure and cause FSGS in people. Additional analysis of these nsSNPs may elucidate the precise molecular processes behind pathogenicity.

References

1. Choi, Y. A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein. Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine, October 7-10, 2012. <https://doi.org/10.1145/2382936.2382989>
2. Daneman A, Navarro OM, Somers GR, Mohanta A, Jarrín JR, Traubici J. Renal pyramids: focused sonography of normal and pathologic processes. *Radiograph*. 2010;30(5):1287-1307. <https://doi.org/10.1148/rg.305095222>
3. Di YM, Chan E, Wei MQ, Liu J-P, Zhou S-F. Prediction of deleterious non-synonymous single-nucleotide polymorphisms of human uridine diphosphate glucuronosyltransferase genes. *Aaps J*. 2009;11(3):469-480. <https://doi.org/10.1208/s12248-009-9126-z>
4. Ebarasi L, Ashraf S, Bierzynska A, et al. Defects of CRB2 cause steroid-resistant nephrotic syndrome. *Am J Hum Genet*. 2015;96(1):153-161. <https://doi.org/10.1016/j.ajhg.2014.11.014>
5. Fogg VC, Liu CJ, Margolis B. Multiple regions of crumbs3 are required for tight junction formation in MCF10A cells. *J Cell Sci*. 2005;118(13):2859-2869. <https://doi.org/10.1242/jcs.02412>

6. Fornoni A, Sageshima J, Wei C. Rituximab targets podocytes in recurrent focal segmental glomerulosclerosis. *Sci Transl Med.* 2011;3(85):85ra46. <https://doi.org/10.1126/scitranslmed.3002231>
7. Frousios K, Iliopoulos CS, Schlitt T, Simpson MA. Predicting the functional consequences of non-synonymous DNA sequence variants--evaluation of bioinformatics tools and development of a consensus strategy. *Genomics.* 2013;102(4):223-228. <https://doi.org/10.1016/j.ygeno.2013.06.005>
8. Gbadegesin R, Hinkes B, Vlangos C, et al. Mutational analysis of NPHS2 and WT1 in frequently relapsing and steroid-dependent nephrotic syndrome. *Pediatr Nephrol.* 2007;22(4):509-513. <https://doi.org/10.1007/s00467-006-0377-y>
9. Gosens I, den Hollander AI, Cremers FP, Roepman R. Composition and function of the Crumbs protein complex in the mammalian retina. *Exp Eye Res.* 2008;86(5):713-726. <https://doi.org/10.1016/j.exer.2008.02.005>
10. Huang T, Zhou Y, Zhang J, et al. The physiological role of Motin family and its dysregulation in tumorigenesis. *J Translation Med.* 2018;16(1):e98. <https://doi.org/10.1186/s12967-018-1466-y>
11. Korbet SM. Treatment of primary FSGS in adults. *J Am Soc Nephrol.* 2012;23(11):1769-1776. <https://doi.org/10.1681/ASN.2012040389>
12. Kriz W, Lemley LM. A potential role for mechanical forces in the detachment of podocytes and the progression of CKD. *J Am Soc Nephrol.* 2015;26(2):258-69. <https://doi.org/10.1681/ASN.2014030278>
13. Kumar P, Mahalingam K. In silico approach to identify non-synonymous SNPs with highest predicted deleterious effect on protein function in human obesity related gene, neuronal growth regulator 1 (NEGR1). *3 Biotech.* 2018;8(11):e466. <https://doi.org/10.1007/s13205-018-1463-0>
14. Musiała A, Donizy P, Augustyniak-Bartosik H, et al. Biomarkers in Primary Focal Segmental Glomerulosclerosis in Optimal Diagnostic-Therapeutic Strategy. *J Clin Med.* 2022;11(12):32-92. <https://doi.org/10.3390/jcm11123292>
15. Okamoto M, Koshino K, Sakai K, et al. A case of recurrent focal segmental glomerulosclerosis (FSGS) involving massive proteinuria (>50 g/day) immediately after renal transplantation. *Clin Transplant.*

- 2011;25 Suppl 23:53-58. <https://doi.org/10.1111/j.1399-0012.2011.01455.x>
16. Sethi S, Glassock RJ, Fervenza FC. Focal segmental glomerulosclerosis: towards a better understanding for the practicing nephrologist. *Nephrol Dial Transplant*. 2015;30(3):375-384. <https://doi.org/10.1093/ndt/gfu035>
 17. Apetri AC, Vanik DL, Surewicz Wk. Polymorphism at residue 129 modulates the conformational conversion of the D178N variant of human prion protein 90–231. *Biochemistry*. 2005;44(48):15880-15888. <https://doi.org/10.1021/bi051455+>
 18. Shabaka A, Ribera AT, Fernández-Juárez G. Focal segmental glomerulosclerosis: State-of-the-Art and clinical perspective. *Nephron*. 2020;144(9):413-427. <https://doi.org/10.1159/000508099>
 19. Shen L. Tight junctions on the move: Molecular mechanisms for epithelial barrier regulation. *Annal New York Acad Sci*. 2012;1258(1):9-18. <https://doi.org/10.1111/j.1749-6632.2012.06613.x>
 20. Tian J, Wu N, Guo X, et al. Predicting the phenotypic effects of non-synonymous single nucleotide polymorphisms based on support vector machines. *BMC Bioinform*. 2007;8(1):e450. <https://doi.org/10.1186/1471-2105-8-450>
 21. Willer CJ, Bonnycastle LL, Conneely KN, et al. Screening of 134 single nucleotide polymorphisms (SNPs) previously associated with type 2 diabetes replicates association with 12 SNPs in nine genes. *Diabetes*. 2007;56(1):256-264. <https://doi.org/10.2337/db06-0461>
 22. Yates CM, Filippis I, Kelley LA, Sternberg MJ, et al. SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features. *J Mol Biol*. 2014;426(14):2692-2701. <https://doi.org/10.1016/j.jmb.2014.04.026>
 23. Yu H, Artomov M, Brähler S, et al. A role for genetic susceptibility in sporadic focal segmental glomerulosclerosis. *J Clin Invest*. 2016;126(3):1067-1078. <https://doi.org/10.1172/JCI82592>