Libyan Journal of Medical and Applied Sciences LJMAS
Online ISSN: 3006-1113
Volume 3, Issue 3, 2025, Page No: 36-47
Website: https://ljmas.com/index.php/journal/index

Prediction and Structural Comparison of Deleterious Missense Single Nucleotide Polymorphisms (nsSNPs) Associated with Respiratory Distress Syndrome

Osamah Alrouwab *[1], Thiheebah Mansour [2], Shahid Ali Mansour [3], Eman Saber khalifa [3], Aya Mohamed A Swaed [3]
[1] Department of Molecular Biology and Biochemistry, Faculty of Medicine, University of Zintan, Zintan, Libya
[2] Department of Zoology, Faculty of Sciences, Aljafra University, Almamura, Libya
[3] Faculty of Biotechnology, Aljafra University, Almamura, Libya
Corresponding Email: rawab@uoz.edu.ly

Received: June 05, 2025 Accepted: July 10, 2025 Published: July 15, 2025

Cite this article as: O, Alrouwab., T, Mansour., S, A, Mansour., E, S, Khalifa., A, M, Swaed. (2025). Prediction and Structural Comparison of Deleterious Missense Single Nucleotide Polymorphisms (nsSNPs) Associated with Respiratory Distress Syndrome. Libyan Journal of Medical and Applied Sciences (LJMAS). 2025;3(3):36-47.

Abstract:

Background: Infant Respiratory Distress Syndrome (IRDS) is a major cause of neonatal morbidity and mortality, often linked to genetic mutations affecting pulmonary surfactant metabolism. Mutations in genes such as *SFTPA1, SFTPA2, SFTPB, SFTPC, SFTPD, ABCA3,* and *NKX2-1* have been implicated in surfactant dysfunction and IRDS pathogenesis. However, the functional consequences of many nonsynonymous single nucleotide polymorphisms (nsSNPs) remain poorly characterized. **Objective:** This study aims to identify and prioritize potentially deleterious nsSNPs in surfactant-associated genes using an integrated computational approach, providing insights into their structural and functional impacts for future experimental validation. **Methods:** We performed a comprehensive in silico analysis of rare missense variants (MAF < 1%) obtained from public databases (dbSNP, ClinVar, UniProt). Functional impact was predicted using SIFT, PolyPhen-2, Panther, and CADD. Protein stability changes were assessed using I-Mutant2.0 and MUpro. Evolutionary conservation was evaluated via ConSurf, and structural modeling was carried out with AlphaFold, SWISS-MODEL, PyMOL, and GROMACS to analyze RMSD and hydrogen bonding patterns. **Results:** Mutation frequency analysis revealed *ABCA3* as the most variant-rich gene (n = 36,391), while *SFTPD* had the lowest mutation load (n = 4,646). Consensus prediction identified several high-risk nsSNPs, including R276W (*SFTPB*), V48M (*SFTPC*), G86E (*SFTPA2*), and P131T (*SFTPD*), consistently classified as damaging across tools. Protein stability analysis confirmed significant destabilization for R276W (*SFTPB*) and V48M (*SFTPC*). Structural modeling showed increased RMSD and reduced hydrogen bonds in variants such as G123V (*SFTPA1*) and G100S (*SFTPC*), indicating potential structural disruption. Highly conserved residues (ConSurf score ≥ 7), especially those involving glycine or proline substitutions, were more likely to be functionally critical. **Conclusion:** This comprehensive in silico analysis identifies several high-confidence deleterious nsSNPs that may contribute to surfactant dysfunction and IRDS pathogenesis. These findings offer valuable insights for future functional studies and may aid in the development of targeted genetic screening strategies for IRDS risk assessment.

Keywords: Infant Respiratory Distress Syndrome (IRDS), Surfactant proteins, Protein stability prediction, Genetic variants, Nonsynonymous Single Nucleotide Polymorphisms (nsSNPs).

المقارنة التركيبية و التنبؤ بالطفرات متعددة الاشكال الخطيرة المصاحبة لمرض الضائقة التنفسية للاطفال حديثى الولادة

اسامة الرواب [1*]، ذهبية منصور [2]، ايمان صابر خليفة[3]، اية محمد سويد[3]، شهد على منصور [3]
قسم الكيمياء الحيوية وعلم الاحياء الجزئي، كلية الطب البشرى جامعة الزنتان، الزنتان، ليبيا [1]
قسم علم الحيوان، كلية العلوم، جامعة الجفارة، المعمورة، ليبيا [2]
القسم العام، كلية التقنيات الحيوية جامعة الجفارة، المعمورة، ليبيا [3]

الملخص

الخلفية: متلازمة الضائقة التنفسية للرضع (IRDS) تُعتبر من الأسباب الرئيسية للإعاقة والوفاة بين المواليد الجدد، وغالبًا ما ترتبط ب mutations جينية تؤثر على عملية التمثيل الغذائي لسائل الرئة (pulmonary surfactant). يُعتقد أن الطفرات في جينات مثل SFTPA1، SFTPA2، SFTPB، SFTPC، SFTPD، ABCA3، وNKX2-1 تلعب دورًا في خلل وظيفة السائل الرئوي وتطور متلازمة الضائقة التنفسية. ومع ذلك، تظل العواقب الوظيفية للكثير من الطفرات غير المشفرة (nsSNPs) غير مفهومة بشكل كافٍ. **الهدف:** تهدف هذه الدراسة إلى تحديد وتصنيف الطفرات غير المشفرة (nsSNPs) المحتملة الضارة في الجينات المرتبطة بالسائل الرئوي باستخدام منهجية حاسوبية متكاملة، بهدف تقديم رؤى حول تأثيرها الهيكلي والوظيفي لاستخدامها في الدراسات التجريبية المستقبلية. **الطرق:** قمنا بإجراء تحليل شامل in silico للطفرات النادرة من النوع missense (بتردد أليلي أقل من 1% - MAF < 1%) المستخلصة من قواعد بيانات عامة مثل dbSNP، ClinVar، وUniProt. تم تقييم التأثير الوظيفي لهذه الطفرات باستخدام أدوات التنبؤ SIFT، PolyPhen-2، Panther، وCADD. كما تم تقييم تأثير الطفرات على استقرار البروتين باستخدام أدوات I-Mutant2.0 وMUpro. تم تحليل الحفظ التطوري عبر الأنواع باستخدام برنامج ConSurf، بينما تم النمذجة الهيكلية باستخدام أدوات مثل AlphaFold، SWISS-MODEL، PyMOL، وGROMACS لتحليل تغيرات RMSD وأنماط الروابط الهيدروجينية. **النتائج:** أظهر تحليل تكرار الطفرات أن جين ABCA3 هو الأكثر احتواءً على الطفرات (n = 36,391)، بينما كان جين SFTPD الأقل (n = 4,646). وقد حددت التنبؤات المشتركة بين الأدوات عدة طفرات عالية الخطورة، من بينها R276W (في جين SFTPB) وV48M (في جين SFTPC) وG86E (في جين SFTPA2) وP131T (في جين SFTPD)، والتي تم تصنيفها على أنها ضارة بشكل متكرر عبر الأدوات المختلفة. ودعم تحليل استقرار البروتين هذه النتائج، حيث أظهرت الطفرات R276W (SFTPB) وV48M (SFTPC) تأثيرًا كبيرًا في تقليل استقرار البروتين. كما أظهر النمذجة الهيكلية ازديادًا في قيمة RMSD وتقليل عدد الروابط الهيدروجينية في الطفرات مثل G123V (SFTPA1) وG100S (SFTPC) ، مما يشير إلى احتمالية حدوث اضطراب هيكل البروتين. أظهرت بقايا الأحماض الأمينية التي تتمتع بدرجة حفظ عالية (درجة ConSurf ≥ 7)، خاصةً تلك التي تشمل تبديلات لـ الجلايسين أو البرولين، أنها محورية وظيفيًا. **الاستنتاج:** توفر هذه الدراسة الحاسوبية الشاملة قائمة عالية الثقة بالطفرات الضارة التي قد تساهم في خلل السائل الرئوي وتطور متلازمة الضائقة التنفسية للرضع (IRDS). وتشكل هذه النتائج أساسًا معرفيًا قيمًا لدراسات وظيفية مستقبلية، كما قد تُسهم في تطوير استراتيجيات فحص جيني مُوجّه لتحديد مخاطر IRDS في السكان المعرضين للخطر.

الكلمات المفتاحية: متلازمة الضائقة التنفسية للرضع، بروتينات المادة الفاعلة بالسطح، الانحرافات الجينية، التنبؤ باستقرار البروتين، لاختلافات النووية الأحادية

Introduction

Infant respiratory distress syndrome (IRDS), or Hyaline membrane disease, is a serious respiratory condition that is perceived in premature infants [1]. The pulmonary surfactant, a mixture of phospholipids and unique proteins that lower surface tension at the alveolar air-liquid interface and prevent alveolar collapse at the end of expiration, is the origin of IRDS when it is either not produced enough or its protein architecture is damaged [2]. IRDS continues to be a significant cause of morbidity and mortality among preterm children, particularly those with gestations of fewer than 28 weeks, despite the fact that improvements in neonatal care and surfactant replacement therapy have significantly increased survival [3, 4]. More often, genetic variables are being identified as the cause of both familial and sporadic IRDS, especially in full-term infants who have unexplained respiratory failure [5, 6]. The primary protein components of pulmonary surfactant include four surfactant-associated proteins, SP-A, SP-B, SP-C, SP-D, and ATP-binding cassette transporter A3 (ABCA3) that is critical for surfactant lipid trafficking [7]. In addition, the surfactant protein genes are regulated by the transcription factor NKX2-1 [8]. Mutations in such genes are not only reported to cause IRDS, but also in various other interstitial lung diseases and surfactant metabolism dysfunction disorders [9]. These include autosomal recessive mutations in SFTPB and ABCA3 and autosomal dominant mutations in SFTPC leading to various clinical phenotypes from neonatal respiratory distress to chronic interstitial lung disease in older children and adults [10, 11]. Several studies, both targeted sequencing and genome-wide strategies, have reported disease-causing mutations in surfactant-associated genes [12]. More recent investigations, however, have underscored the role of ABCA3 mutations in acute severe respiratory failure in term and preterm infants. Likewise, SFTPC mutations I73T and L188Q have been associated with familial interstitial lung disease and neonatal IRDS [9]. Nevertheless, the majority of these investigations concentrate on uncommon, highly penetrant mutations and do not include thorough examinations of prevalent and low-frequency variants that could also participate in disease susceptibility [13]. In silico analysis has ever more become an important resource for the prioritization of variants of uncertain significance (VUS) in clinical genomics. Functional prediction algorithms, such as SIFT, PolyPhen-2, Panther, and CADD, have been utilized to classify nonsynonymous single nucleotide polymorphisms (nsSNPs) based on their potential impact on protein functionality [14]. AlphaFold and PyMOL are two instances of structural modeling tools providing substantial insight into how changes in amino acids affect protein structure and interaction surfaces [15, 16]. Computational techniques provide a scalable and affordable way to find potential high-risk variations that merit more research. However, neither a comprehensive nsSNP catalog of surfactant genes for use in future studies and clinical screening nor a consensus-based optimal methodology for combining these tools exist today [17]. This work seeks to fill these gaps by conducting a combined computational analysis of nsSNPs in seven surfactant-related genes: SFTPA1, SFTPA2, SFTPB, SFTPC, SFTPD, ABCA3, and NKX2-1. Through a multi-step in silico pipeline, we evaluate the functional, structural, and evolutionary impacts of missense variants to determine potentially deleterious SNPs that can lead to surfactant dysfunction and IRDS pathogenesis. The method employed integrates

functional prediction tools, protein stability analysis, conservation score analysis, and structural modeling techniques to produce a ranked list of candidate variants for subsequent experimental and clinical analysis.

Methods

Selection of Candidate Genes Associated with IRDS

A nominee gene panel was designated according to its significance in lung epithelium development, alveolar stability, and surfactant metabolism, in order to examine the genetic components involved in Infant Respiratory Distress Syndrome (IRDS). DisGeNET (https://disgenet.com/), GeneCards (https://www.genecards.org/), OMIM (https://omim.org/), and Ensembl (https://www.ensembl.org/) were among the comprehensive genomic databases from which data mining was employed. Additionally, a thorough review of the literature pertaining to the genes involved in the synthesis and regulation of pulmonary surfactant was conducted. Genes with strong clinical and functional associations with neonatal respiratory distress—whether due to inherited mutations, expression patterns, or their mechanistic involvement in lung maturation pathways—were prioritized. Seven genes, including SFTPA1, SFTPA2, SFTPB, SFTPC, SFTPD, ABCA3, and NKX2-1, were included in the supreme selection (Figure 1).

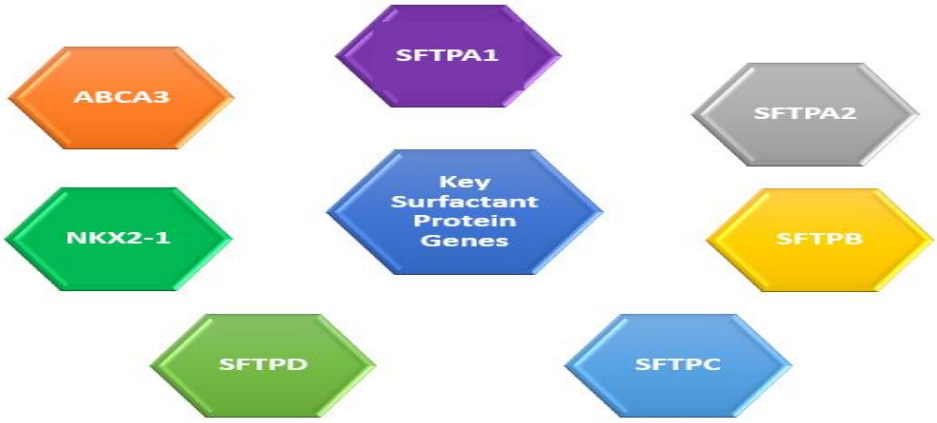

Figure 1. Key Surfactant Protein Genes

SNP Dataset Compilation2.2. SNP Dataset Compilation

Missense single nucleotide polymorphisms (nsSNPs) related to seven surfactant-associated genes (*SFTPA1, SFTPA2, SFTPB, SFTPC, SFTPD, ABCA3,* and *NKX2-1*) were collected from dbSNP (https://www.ncbi.nlm.nih.gov/snp/), ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), and UniProt (https://www.uniprot.org/). Variants with a minor allele frequency (MAF) ≥ 1% were excluded to focus on rare variants of potential clinical significance. Only protein-altering mutations were selected for downstream in silico analysis.

Functional Prediction of Deleterious Variants

A consensus prediction analysis was performed based on a set of well-known computational tools: SIFT (https://sift.bii.a-star.edu.sg/), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/bgi.shtml), Panther (https://www.pantherdb.org/), and Combined Annotation Dependent Depletion (CADD; https://cadd.gs.washington.edu/). Predicted deleterious variants by at least three out of the four main predictors (SIFT, PolyPhen-2, Panther, CDAA) were selected for further investigation.

Protein Stability Estimation

For the assessment of missense variant effects on protein stability, two machine learning-based tools were employed: I-Mutant2.0 (https://folding.biofold.org/i-mutant/i-mutant2.0.html) and MUpro (https://mupro.proteomics.ics.uci.edu/). I-Mutant2.0 is an SVM web server that calculates the change in Gibbs free energy (ΔΔG) upon single-site mutation from protein sequence or structure. When ΔΔG is negative, the mutation is destabilizing, whereas when it is positive, the protein is stabilizing. In parallel, MUpro was utilized to further validate these predictions. The tool utilizes both SVM and neural network models trained on extensive mutation datasets to predict the direction and extent of stability change. It gives a confidence score ranging from -1 to 1, with values below zero indicating loss of stability and values above zero being indicative of increased stability. Just the variants that were predicted consistently to destabilize protein stability by both analysis tools were kept for further structural analysis because such destabilizing mutations are expected to disrupt protein folding, integrity, and function.

Conservation Analysis
Evolutionary conservation of mutated residues was assessed using ConSurf (http://consurf.tau.ac.il/), which calculates conservation scores based on phylogenetic relationships derived from multiple sequence alignments. Residues with scores of 7–9 were considered highly conserved.

Structural Modeling of Wild-Type and Mutant Proteins
Homology modeling was employed to identify the structural effects of certain selected high-risk nsSNPs in surfactant-associated proteins. The wild-type and mutant protein structures were modeled using SWISS-MODEL (https://swissmodel.expasy.org/) and AlphaFold **(**https://alphafold.ebi.ac.uk/**)**, and structure alignments were visualized with PyMOL **(**https://www.pymol.org/**)**.

Structural Deviation and Mutation Severity
The structural modeling and analysis of missense mutations in surfactant proteins were conducted using a combination of computational tools including AlphaFold, PyMOL, and GROMACS (https://www.gromacs.org/). Full-length three-dimensional protein models for SFTPA1, SFTPA2, SFTPB, SFTPC, and SFTPD were obtained using AlphaFold Protein Structure Database predictions, which provide high-confidence models for human proteins based on deep learning. Each wild-type protein model was visualized and validated using PyMOL to confirm the integrity of secondary and tertiary structures prior to mutation introduction. Missense mutations were manually introduced using PyMOL's mutagenesis tool, allowing for precise substitution of the target residues. The root-mean-square deviation (RMSD) between each mutant and its native model was calculated using PyMOL's alignment tools to quantify structural deviation. Additionally, hydrogen bond estimation was performed by counting nitrogen (N) and oxygen (O) atoms, which serve as common donors and acceptors in hydrogen bonding, as an approximate measure of stabilizing interactions.

Results
Mutation Distribution Among Surfactant Protein Genes
Table 1 summarizes the frequency and distribution of different types of mutations — including missense, synonymous, intronic, and other variants — across seven genes implicated in surfactant metabolism: *SFTPA1*, *SFTPA2*, *SFTPB*, *SFTPC*, *SFTPD*, *ABCA3*, and *NKX2-1*.

Table 1. Mutation Frequencies of Surfactant Protein Genes

Gene	Missense	Synonymous	Intron	Other	Total
SFTPA1	872	354	3662	2158	7046
SFTPA2	640	266	2808	818	4532
SFTPB	314	203	3959	942	5517
SFTPC	741	435	3522	936	5634
SFTPD	342	146	4111	47	4646
ABCA3	2085	1080	32336	980	36391
NKX2-1	2846	1533	4538	3102	12019

Among these genes, ABCA3 displayed the highest total number of variants (36,391), driven predominantly by a large number of mutations in the "other" category (n = 32,336), which may encompass regulatory, splice site, or complex structural variants. This is consistent with its known large gene size and complex exon-intron structure. In contrast, *SFTPD* exhibited the lowest overall mutation count (n = 4,646), with intronic mutations (n = 4,111) forming the vast majority. NKX2-1 showed the second-highest total mutation count (n = 12,019), with a notably high number of missense mutations (n = 2,846) and "other" variants (n = 3,102), highlighting its potential vulnerability to functionally significant alterations. Similarly, both SFTPA1 (n = 872) and SFTPC (n = 741) showed a substantial number of missense mutations, which may impact protein structure and function. Interestingly, SFTPB had the lowest number of missense mutations (n = 314), but a relatively high proportion of intronic variants (n = 3,959), suggesting potential regulatory impacts rather than direct protein-coding changes. SFTPA2 exhibited a moderate total variant load (n = 4,532), with a balanced distribution between coding and non-coding mutations. Overall, missense variants were most prevalent in *NKX2-1*, while synonymous variants were highest in *SFTPC* (n = 435). Intronic mutations were dominant in most genes, with *ABCA3*, *SFTPD*, and *SFTPC* exhibiting particularly high counts. These findings underscore the heterogeneity of mutational landscapes

across surfactant-related genes, suggesting differential genomic instability and functional constraint (Figure. 2).

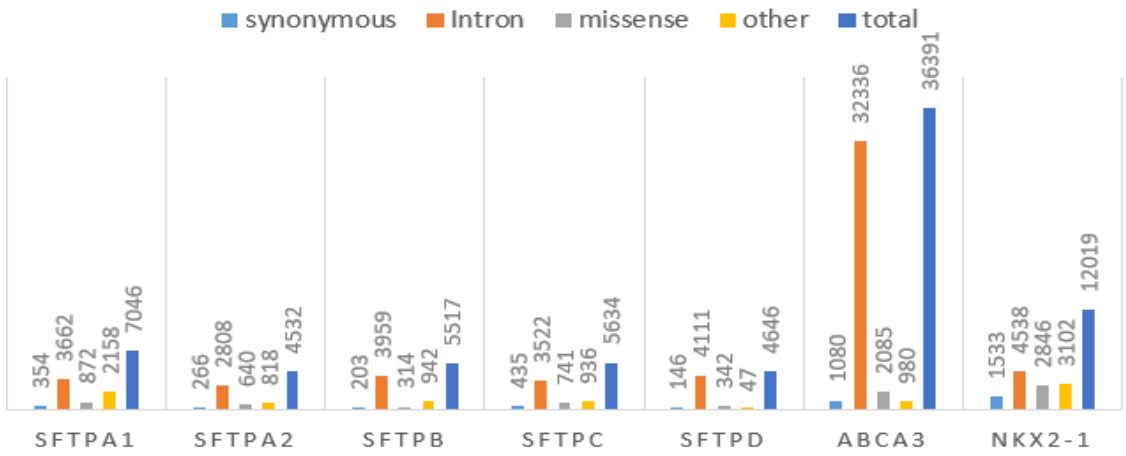

Figure 2. Mutation Frequencies of Surfactant Protein Genes

Prediction of Deleterious nsSNPs in Surfactant Protein Genes

A critical analysis of 70 nsSNPs of the surfactant-associated genes (SFTPA1, SFTPA2, SFTPB, SFTPC, SFTPD, ABCA3, and NKX2-1) was carried out to determine their likely deleterious effects through four prediction methods: SIFT, PolyPhen-2, PANTHER, and CADD. Interestingly, there were high similarities among some of these variants in several algorithms, indicating high functional importance (Table 2.). In SFTPA1, the G123V (rs549832850) substitution was predicted to be damaging by SIFT (score = 0) and PolyPhen (score = 0.827), and it had a high CADD score of 33, indicating a likely damaging effect. Similarly, P62L (rs151242911) and F198L (rs541642692) also revealed high CADD scores (22 and 20, respectively) and concordant pathogenicity predictions across tools. On the other hand, V25I (rs181845535) and C20S (rs201722413) variations were all assessed to be benign as shown by the low CADD scores and tolerated or benign predictions in SIFT and PolyPhen analyses. In SFTPA2, certain of the variants such as G86E (rs200897593) and M60K (rs538280792) were predicted to be deleterious by SIFT (0.01 and 0.02, respectively) and had high CADD scores (22 and 18) with PolyPhen scores of possibly to probably damaging. Notably, G86E was more pathogenic than G86A, even though both were substitutions at the same residue. Certain other variants such as E105A and V81F were consistently benign across all predictors, as evidenced by low CADD scores (13 and 0, respectively). SFTPB was characterized by a predominance of extremely damaging variants. In particular, P379R (rs144831319), R276W (rs542291993), and Q374H (rs551956264) were predicted by PolyPhen to be most likely damaging (scores > 0.93) and also had CADD scores ≥28. Other variants such as G53A and G141D also had high CADD scores (24 and 25), which is consistent with structural or functional disruption. F201L and P288Q, however, were predicted to be benign or of uncertain effect despite occurring at structurally significant positions. SFTPC variant analysis revealed several high-predicted impact substitutions. V48M (rs566914013) and V48L were both predicted to be deleterious (SIFT = 0 and 0.26, respectively) and scored CADD 26 and 23, respectively. Notably, H32Y, S95Y, and C149W also revealed damaging potential with moderate-to-high CADD scores (15–21). Variants like A173D and A179V provided conflicting interpretations, suggesting the need for structural modeling or in vitro validation. Amongst the variants detected in SFTPD, the P131T variant (rs200767343) stood out most evidently with its CADD score of 23 and the harmful predictions made by SIFT as well as PolyPhen. Variants P137L, R50H, and A294V also presented a deleterious profile, with CADD scores >20. While some variants like E262D and N308K produced discordant predictions among the tools, their high CADD scores (≥15) warrant investigation using functional assays. The large and polymorphic ABCA3 gene, in a few of the variant positions, had overlaps with SFTPD, most likely owing to annotation differences. Frequently occurring and consistently predicted to be potentially damaging were variants like A294V, P131T, and E309Q. Although some had modest SIFT or PolyPhen values, high CADD scores (≥17) indicate potential for functional impact. Lastly, NKX2-1 had fewer high-risk nsSNPs. While no variants were deemed sheer damage in any of the tools, L34I and P37T had moderate CADD scores (25 and 28, respectively) and SIFT scores indicating deleteriousness, which deserved experimental confirmation. Variants such as M83I and H316R were mostly predicted as benign or tolerated. Overall, several nsSNPs had convergent predictions of deleteriousness, i.e., those with SIFT scores ≤ 0.05, PolyPhen scores > 0.85, and CADD scores ≥ 20. These are strong candidates for downstream confirmation in functional assays, particularly those in SFTPB, SFTPC, and SFTPA2, which may play a role in surfactant dysfunction and related pulmonary disease (Table 2.).

Table 2. List of nsSNPs of Surfactant Protein Genes predicted as deleterious in various in silico tools

Gene	variant ID	Amino acid change	SIFT score	SIFT Prediction	POLYPHEN SCORE	POLYPHEN Prediction	PANTHER SCORE	PANTHER Prediction	CADD SCORE	CADD Prediction
SFTPA1	rs201722413	C20S	0.07	Tolerated	0.003	Benign	NA	NA	1	Benign
	rs549832850	G123V	0	Deleterious	0.827	Possibly Damaging	NA	NA	33	Damaging
	rs150214547	N210K	0.19	Tolerated	1	Possibly Damaging	NA	NA	15	Moderate
	rs151242911	P62L	0.01	Deleterious	0.999	Possibly Damaging	NA	NA	22	Damaging
SFTPA2	rs144575180	R199C	0.08	Tolerated	0.86	Possibly Damaging	0.19	probably benign	15	Moderate
	rs147158327	T66I	0.14	Tolerated	0	Benign	0.02	probably benign	1	Benign
	rs538280792	M60K	0.02	Deleterious	0.147	Benign	0.5	possibly damaging	18	Likely Damaging
	rs200897593	G86E	0.01	Deleterious	0.999	Probably Damaging	0.5	possibly damaging	22	Damaging
SFTPB	rs144831319	P379R	0	Deleterious	0.998	Probably Damaging	0.5	possibly damaging	20	Likely Damaging
	rs573709240	A83S	0.12	Tolerated	0.467	Possibly Damaging	0.19	probably benign	16	Uncertain
	rs542291993	R276W	0	Deleterious	0.991	Probably Damaging	NA	NA	28	Damaging
SFTPC	rs566914013	V48M	0	Deleterious	0.963	Probably Damaging	0.27	probably benign	26	Damaging
	rs552621627	S95Y	0.13	Tolerated	0.815	Possibly Damaging	0.27	probably benign	21	Damaging
	rs537383138	G170E	0.03	Deleterious	0.012	Benign	0.27	probably benign	14	Uncertain
	rs192391655	H32Y	0	Deleterious	0.483	Possibly Damaging	NA	NA	15	Likely Damaging
	rs531634308	A179V	0	Deleterious	0.18	Benign	NA	NA	14	Uncertain
	rs531634308	A173D	0.1	Tolerated	0.003	Benign	NA	NA	8	Likely Benign
	rs566914013	V48L	0.26	Tolerated	0.993	Probably Damaging	0.27	probably benign	23	Damaging

SFTPD	rs4469829	E309Q	0.02	deleterious	0.005	benign	0.19	probably benign	15	Likely Damaging
	rs80268129	R174H	0.64	tolerated	0.103	benign	0.27	probably benign	0	Benign
	rs141481902	P137L	0.08	tolerated	0.722	possibly damaging	0.5	possibly damaging	20	Damaging
	rs200767343	P131T	0	deleterious	0.65	possibly damaging	0.5	possibly damaging	23	Damaging
ABCA3	rs4469829	E309Q	0.02	benign	0.005	likely benign	NA	NA	15	Likely Damaging
	rs182707229	N308K	0.09	possibly damaging	0.642	likely benign	NA	NA	13	Uncertain
	rs565843911	A295V	0.02	benign	0	likely benign	NA	NA	19	Likely Damaging
NKX2-1	rs566702552	M83I	0.18	tolerated	0.223	benign	NA	NA	23	likely benign
	rs552685896	H316R	0.01	deleterious	0	NA	probably damaging	0.57	23	likely benign
	rs182071645	P37T	0.03	deleterious	0.224	benign	probably damaging	0.57	28	likely benign
	rs201631950	L34I	0.08	tolerated	0.646	possibly damaging	probably damaging	0.57	25	likely benign

Protein Stability Impact of Deleterious Variants

In order to further explore the structural influence of deleterious missense mutations only those gives 3 confirmations as deleterious are elected. We examined the influence of chosen nsSNPs on protein stability through the utilization of two various in silico prediction tools: I-Mutant2.0 and MUpro. These programs calculate the variation of Gibbs free energy (ΔΔG) and provide outputs indicating that the mutation stabilizes or destabilizes the protein. As outlined in Table 3, most of the variants predicted to be deleterious by functional prediction tools also had a negative ΔΔG in I-Mutant2.0 and a negative MUpro score, pointing to decreased protein stability. For example, in the SFTPB gene, the R276W (rs542291993) variant had a strong destabilizing effect (ΔΔG = –2.23 kcal/mol; MUpro score = –0.809) with high reliability. Similarly, the V48M substitution in SFTPC (rs566914013) also showed considerable destabilization (ΔΔG = –1.57 kcal/mol; MUpro score = –0.940), corroborating the predicted pathogenic role of this substitution. For SFTPA2, M60K (rs538280792) and G86E (rs200897593) variants consistently scored to reduce stability with ΔΔG of –1.22 and –0.70 kcal/mol, respectively. P379R variant of SFTPB and G141D also exhibited strong destabilizing characteristics. Among the SFTPD variants, P131T (rs200767343) exhibited strong destabilization with ΔΔG of –1.04 kcal/mol and MUpro of –0.905. Unexpectedly, not all damaging variants caused stability loss. The SFTPA1 P62L and SFTPD P137L variants were predicted by I-Mutant to minimally increase stability (ΔΔG = 0.06 and 0.01 kcal/mol, respectively), although MUpro predictions for P137L still suggested destabilization. That we have contrasting predictions from contrasting tools highlights the value of multi-tool consensus and indicates potential compensatory structural mechanisms. Overall, these results suggest that a substantial proportion of functionally deleterious nsSNPs contribute to disease phenotypes through the destabilization of protein structure, potentially impairing surfactant assembly, secretion, or interaction with binding partners in the lung.

Table 3. Prediction of change in protein stability using I-Mutant2.0 and MUpro

Gene	Variant ID	Amino Acid Change	I-Mutant ΔΔG (kcal/mol)	I-Mutant Prediction	RI	MUpro Score	MUpro Prediction
SFTPA1	rs549832850	G123V	-0.17	**Decrease**	2	-0.372	**Decrease**
SFTPA1	rs151242911	P62L	0.06	Increase	2	0.117	Increase
SFTPA2	rs538280792	M60K	-1.22	**Decrease**	5	-0.708	**Decrease**
SFTPA2	rs200897593	G86E	-0.70	**Decrease**	5	-0.262	**Decrease**
SFTPB	rs144831319	P379R	-0.29	**Decrease**	2	-1.326	**Decrease**
SFTPB	rs542291993	R276W	-2.23	**Decrease**	9	-0.809	**Decrease**
SFTPB	rs549877442	G141D	-0.82	**Decrease**	5	-0.509	**Decrease**
SFTPC	rs566914013	V48M	-1.57	**Decrease**	9	-0.940	**Decrease**
SFTPC	rs566914013	V48L	-0.72	**Decrease**	6	-0.793	**Decrease**
SFTPC	rs192391655	H32Y	-0.28	**Decrease**	7	-0.189	**Decrease**
SFTPD	rs200767343	P131T	-1.04	**Decrease**	6	-0.905	**Decrease**
SFTPD	rs141481902	P137L	0.01	Increase	4	-0.358	**Decrease**

Conservation Analysis of Mutated Residues

Conservation profile of SFTPA1 protein was examined by the ConSurf algorithm that takes into account evolutionary constraints at the amino acid level. Some of the significant variants that had been earlier identified as potentially functionally damaging were located in moderately to highly conserved positions of specific interest, the P62L variant (rs151242911) occurs at a moderately conserved site (grade 7). As proline plays a role in structural stiffness and local backbone conformation, substitution with leucine can significantly alter protein folding, or interaction surfaces, and thereby account for its pathogenicity. The G123V variant (rs549832850) occurs at a moderately variable site (grade 6), although the unique flexibility of glycine foresees that even small substitutions can disrupt structural packing, especially in compact or dynamic regions. P91T allele (rs1136452), which has a conservation grade of 6, is in a structurally important area where the substitution of proline with threonine may have an influence on the local secondary structure stability. In contrast, substitutions such as V25I (rs181845535) and C20S (rs201722413) were found in positions that are less conserved (grade 4), which implies a higher tolerance to the substitution and supports their previous classification as tolerated or benign mutations in functional prediction models. The conservation observations above validate the negative predictions for some of the SFTPA1 variants like P62L and G123V, which must be assigned high priority for experimental investigation. Residues with conservation scores of 7 or higher, particularly those involving glycine or proline, are very likely to be structure- or function-critical sites. The conservation analysis of the SFTPA2 protein, as conducted using ConSurf, demonstrated that many of the variants occur at residues under substantial evolutionary constraints, suggesting their potential functionality. Among the previously named variants that were expected to be damaging, G86E (rs200897593) is notable for being in a very conserved location (grade 9). The presence of glycine in this particular position is important to maintaining flexibility and conformation of the backbone. Its substitution with glutamic acid, which introduces a larger and charged side chain, indicates a strong possibility of functional impairment. In contrast, the M60K mutation (rs538280792) takes place at a variable site graded as 3, indicating that this site may accept structural changes even though it is predicted to affect protein stability. Variants R199C (rs144575180) and A217P (rs550180644) occur at regions of moderate conservation (grades 6–7) and may have a structural sensitivity potential, especially as they occur at residues such as proline that determine secondary structure. Substitutions such as E105A, V81F, and T66I that occur at lower grades of conservation (3–4) are paired with their less predicted impact on functional models. The results of this study merited the high-priority candidate status of G86E for future functional and structural investigations owing to its presence at a position of high conservation. The other variants have variable risks based on the degree of conservation and their biochemical character, and thus a multi-tiered investigation is indicated. SFTPB conservation study showed a range of evolutionary significance across the variants being studied. Some of the residues are present at highly conserved positions, which is in keeping with their being functionally significant. The R276W mutation (rs147567568), designated as one of the most commonly identified deleterious mutations by computer prediction, is at a high-conservation site (grade 9). The substitution involved replaces a charged arginine with a bulky, hydrophobic tryptophan that will impact protein folding, processing, or surfactant function. Another conserved amino acid, L174P (rs114825017) at grade 8, contains a proline insertion that will destabilize α-helix domains important for surfactant binding. Conversely, S130F (rs201088183) and P193S (rs140032360) appear at moderately conserved sites (grade 5–6), and their functional impact may be more context-dependent structurally. Changes like T131N and E158K, appearing at lower grades of conservation (3–4), may be less significant evolutionarily, even in case of any predicted structural alteration. In general, the conservation profile favors the high priority of R276W and

L174P for further investigation in the pathophysiological mechanisms underlying IRDS. ConSurf analysis of SFTPC revealed nsSNPs at evolutionarily conserved positions that are strongly indicative of having structural or functional importance. Among the tested variants, V48M (rs147998249) was identified as a major substitution, occurring at a highly conserved position (grade 9). The mutation substitutes a small hydrophobic valine with a bigger methionine in the transmembrane domain, which can impact membrane incorporation and surfactant function. In the same manner, I73T (rs150083326) is found at a conserved position (grade 8) and is a substitution of a nonpolar residue with a polar threonine and would disrupt the hydrophobic core of the protein. Other replacements, like G100S (rs146295446) and R16C (rs148952109), occur at positions of moderate level of conservation (grades 6–7), suggesting possible regional flexibility; nonetheless, they are still important because of the biochemical changes they impose. Variants like T54I and F66L, which occur at lower grades of conservation (4–5), might show increased tolerance depending upon their structural environment. These findings justify the functional prioritization of V48M and I73T for structural confirmation and in vitro modeling because they occur at evolutionary hot spots. In SFTPD, both the P131T (grade 6) and P137L (grade 5) mutations affect moderately variable residues, suggesting a relatively lower likelihood of deleterious effect than those in highly conserved residues. These conservation profiles help to pinpoint notable variants for additional functional evaluation. Special attention is given to mutations at residues that are scored 8–9, especially if glycine or proline are involved because of their special structural roles (Table 4.).

Table 4. Details of nsSNPs selected as deleterious among the reported SNPs, their conservation analysis by ConSurf

Gene	Variants	Conservation score	Interpretation
SFTPA1	P62L (rs151242911)	7	Moderately conserved; structurally sensitive due to proline rigidity
	G123V (rs549832850)	6	Moderately variable; glycine's flexibility implies potential disruption
	P91T (rs1136452)	6	Moderate conservation; possible structural flexibility loss
	V25I (rs181845535	4	Variable; substitution likely tolerated
	C20S (rs201722413)	4	Variable; not highly constrained
SFTPA2	G86E (rs200897593)	9	Highly conserved; functionally critical
	M60K (rs538280792)	3	Variable; likely tolerant
	R199C (rs144575180)	6	Moderately conserved; possibly disruptive
	A217P (rs550180644)	7	Moderately conserved; may impact folding
	E105A (rs146484851)	4	Variable; likely tolerated
	V81F (rs138975473)	3	Variable; substitution likely benign
	T66I (rs147158327)	4	Variable; lower functional constraint
SFTPB	R276W (rs147567568)	9	Highly conserved; functionally essential
	L174P (rs114825017)	8	Highly conserved; possible helical disruption
	S130F (rs201088183)	6	Moderately conserved; potential polarity change
	P193S (rs140032360)	5	Moderately conserved; flexible substitution
	T131N (rs146379742)	4	Variable; tolerable change
	E158K (rs116305692)	3	Variable; likely tolerated
SFTPC	V48M (rs147998249)	9	Highly conserved; membrane domain sensitivity
	I73T (rs150083326)	8	Highly conserved; potential polarity shift

	G100S (rs146295446)	6	Moderately conserved; flexibility disruption
	R16C (rs148952109)	6	Moderately conserved; potential redox sensitivity
	T54I (rs199744725)	5	Moderately variable; mild polarity shift
	F66L (rs147314682)	5	Variable; functionally permissive
SFTPD	P131T (rs200767343)	6	Moderately variable; potential impact on backbone rigidity
	P137L (rs141481902)	5	Variable region; minimal conservation constraint

Structural Deviation and Mutation Severity

Structural analysis of missense mutations in surfactant proteins showed a range of impacts on molecular stability. Some variants, including G123V (SFTPA1), R199C (SFTPA2), and G100S (SFTPC), had higher RMSD values (≥0.7 Å) and a corresponding decrease in overall hydrogen bonds, indicating a potentially destabilizing effect on protein structure. This indicates that such mutations have the potential to change folding or impact functional areas. By contrast, mutations P193S (SFTPB), T54I (SFTPC), and P131T (SFTPD) maintained low RMSD values (≤0.4 Å) with no—and sometimes a slight gain—in hydrogen bond loss, suggesting structural neutrality or weak stabilization. A number of proline replacement substitutions, L174P (SFTPB) and A217P (SFTPA2), had moderate RMSD values; owing to the rigidity of proline, their effect might be strongly context-dependent on local secondary structure. In particular, substitutions such as S130F and I73T were found to be correlated with subtle structural changes but with enhanced hydrogen bonding, suggesting a possible compensatory stabilization. In general, the comparison pointed out the finding that not all missense mutants are harmful; they may be tolerated or even slightly positive. These findings offer a preliminary structural basis for the prioritization of mutations for subsequent functional assessment and could potentially contribute to the interpretation of variants of uncertain significance (VUS) in clinical genomics. Among all modeled variants, R199C (SFTPA2) exhibited the highest structural deviation (RMSD = 0.782 Å), followed closely by G123V (SFTPA1), R276W (SFTPB), and G100S (SFTPC). These variants were associated with reduced hydrogen bond interactions, supporting their predicted destabilizing effects and potential pathogenicity (Table 5).

Table 5. Structural Impact of Missense Mutations in Surfactant Protein Genes Based on RMSD and Hydrogen Bond Analysis

Gene	Mutation	RMSD (Å)	Total H-Bonds	Interpretation
SFTPA	Native	0.000	496	Baseline reference structure.
	P62L	0.265	484	Very low RMSD, similar H-bond count → likely benign or tolerated.
	P91T	0.382	472	Slightly higher RMSD but increase in H-bonds → **possibly stabilizing or neutral**.
	G123V	0.743	494	Noticeable structural deviation and loss of H-bonds → **potentially destabilizing** mutation.
SFTPA2	Native	0.000	496	Baseline reference structure.
	G86E	0.342	495	Slight structural deviation; nearly unchanged H-bond count → likely tolerated or mildly impactful.
	R199C	0.782	492	Highest RMSD and loss of 4 H-bonds → likely destabilizing and potentially functionally disruptive.
	A217P	0.285	498	Low RMSD and slight increase in H-bonds → **potentially stabilizing or neutral**.
FSTPB	Native	0.000	762	Baseline reference structure.
	R276W	0.695	759	Highest RMSD and H-bond reduction → likely structurally destabilizing and may affect function.
	L174P	0.488	761	Moderate deviation; small H-bond loss → potentially impactful, especially as Proline often disrupts secondary structure.
	S130F	0.339	764	Low RMSD and gain in H-bonds → likely stabilizing or benign mutation.
	P193S	0.227	763	Smallest RMSD and increased H-bonding → most likely neutral or mildly stabilizing.
SFTPC	Native	0.000	394	Baseline reference structure.

	V48M	0.314	395	Very mild structural change, slightly increased H-bonding → likely benign.
	I73T	0.412	396	Moderate RMSD, increased H-bond count → possibly stabilizing or neutral effect.
	G100S	0.701	391	Highest RMSD and loss of 3 H-bonds → likely structurally destabilizing.
	R16C	0.478	392	Moderate change + H-bond loss → potentially disruptive, especially if residue is functionally critical.
	T54I	0.265	394	Very low RMSD, no change in H-bonds → probably benign or silent mutation.
SFTPD	Native	0.000	750	Baseline reference structure.
	P131T	0.355	752	Moderate RMSD and a small increase in hydrogen bonds → likely neutral or mildly stabilizing mutation.

Discussion

The identification and detailed analysis of damaging nsSNPs in surfactant-related genes are essential to elucidate the genetic mechanisms underlying IRDS. A systematic in silico approach is described here to explore, predict, and structurally analyze potentially disease-causing mutations in seven key surfactant metabolism genes: SFTPA1, SFTPA2, SFTPB, SFTPC, SFTPD, ABCA3, and NKX2-1. Through the combination of functional prediction tools, calculations of protein stability, conservation scores, and structural modeling, we identified a number of high-risk nsSNPs that can cause surfactant dysfunction and IRDS susceptibility. We found substantial heterogeneity of the mutation burden between the chosen genes. ABCA3 had the highest overall number of variants, which is in line with its large genomic size and complicated exon-intron structure. The lowest overall mutation burden was carried by SFTPD, yet intronic mutations prevailed in this gene. The high rate of missense mutations in NKX2-1, SFTPA1, and SFTPC points to their probable role in disease causation by direct alteration of protein structure and function. The data indicate the heterogeneous mutational load of surfactant-related genes, attesting to different evolutionary pressures and functional significance. Consuming consensus predictions from SIFT, PolyPhen-2, Panther, and CADD, we identified a number of high-confidence deleterious variants. Notably, SFTPB had the most harmful substitutions, namely R276W and L174P, which were consistently predicted to be highly pathogenic. Similarly, SFTPC variants V48M and I73T showed strong evidence of functional disruption. These findings reinforce previous studies implicating these genes in surfactant dysfunction and neonatal lung disease. However, variants like P193S (SFTPB) and T54I (SFTPC) were being predicted as neutral or even stabilizing, suggesting the importance of multi-tool validation and structural context in variant effect interpretation. Protein stability prediction using I-Mutant2.0 and MUpro indicated that the majority of deleterious nsSNPs highly destabilize the protein, which would be expected to affect surfactant assembly and secretion. For example, R276W in SFTPB and V48M in SFTPC showed high destabilization, reinforcing their clinical significance. Structural modeling supported this with higher root-mean-square deviation (RMSD) and lower hydrogen bonding for certain mutants, i.e., G123V (SFTPA1), R199C (SFTPA2), and G100S (SFTPC). These structural perturbations imply potential mechanisms of surfactant dysfunction and impaired alveolar stability in IRDS. Evolutionary conservation scores from ConSurf highlighted residues critical for protein function and structure. Variants at highly conserved positions, such as G86E in *SFTPA2* (score 9) and R276W in *SFTPB* (score 9), are more likely to be functionally significant. These residues are often glycine or proline, known for their roles in backbone flexibility and secondary structure formation. Substitutions at such sites may disrupt protein folding, domain interactions, or surfactant binding properties. In contrast, variants at less conserved sites (e.g., M60K in *SFTPA2)* showed mixed predictions, indicating a lower likelihood of pathogenicity unless they occur in structurally sensitive regions. Structural modeling yielded valuable information regarding the impact of individual amino acid substitutions on protein structure. While some mutations resulted in limited structural change (e.g., P193S in SFTPB), others resulted in dramatic increases in RMSD and a decrease in hydrogen bonds, predicting a strong likelihood of dysfunction. Interestingly, mutations involving proline substitutions (e.g., A217P in SFTPA2 and L174P in SFTPB) resulted in intermediate structural change, demonstrating the vulnerability of proline-rich domains to mutation. These findings provide a preliminary structural basis for prioritizing variants for experimental verification and clinical interpretation. Some previously recognized variants, including R276W in SFTPB and V48M in SFTPC, have already been linked to severe respiratory distress and surfactant deficiency. Nonetheless, our investigation goes further by adding both structural and functional data to the list of potential pathogenic variants. We also point out new candidates like G86E in SFTPA2 and P131T in SFTPD, which are worth pursuing given their high conservation and structural importance. Without experimental confirmation of the effects predicted, our investigation is entirely computational. Follow-up studies should include in vitro expression assays, structure determination through X-ray crystallography activity assessment in cellular or animal models. Correlation with patient genotype-phenotype data would also increase the clinical relevance of variants

identified. Linkage with population databases such as gnomAD and ClinVar would further delineate allele frequency thresholds and increase pathogenicity classification.

Conclusion

This study presents a comprehensive in silico evaluation of high-risk nonsynonymous single nucleotide polymorphisms (nsSNPs) within surfactant-associated genes implicated in Infant Respiratory Distress Syndrome (IRDS). By integrating functional prediction, protein stability analysis, evolutionary conservation, and structural modeling, we identified several candidate variants—such as R199C, G123V, R276W, and G100S—that are likely to disrupt protein function and contribute to surfactant dysfunction. These findings provide a valuable foundation for downstream experimental validation and clinical research. The multi-tiered computational approach applied here enhances the precision of variant interpretation and holds promise for improving genetic screening and risk stratification strategies in neonates at risk for IRDS.

Disclaimer

The article has not been previously presented or published, and is not part of a thesis project.

Conflict of Interest

There are no financial, personal, or professional conflicts of interest to declare.

References

1. Alhaj MK, Elamin SB. Evaluation of management protocol in Sudanese infants with Respiratory Distress Syndrome and Bronchopulmonary Dysplasia in Ribat University Hospital. Ahfad J. 2023;40.
2. Prakash S, Yadav HS, Yadav OP. Pulmonary Surfactant in Health and Disease: An Overview. Janaki Med Coll J Med Sci. 2024;12:108–22.
3. Ng EH, Shah V. Guidelines for surfactant replacement therapy in neonates. Paediatr Child Health. 2021;26:35–41.
4. Martini S, Corvaglia L. Premature Infants. In: Frailty in Children: From the Perioperative Management to the Multidisciplinary Approach. Springer; 2023. p. 11–32.
5. Cupido DT. Investigating pathogens of the gastrointestinal tract in sudden and unexpected death in infancy cases at the Tygerberg medico-legal mortuary, compared to an age-matched healthy control group. 2023.
6. Braddock SR, Lipinski RJ, Carey JC. 40th Annual David W Smith Workshop on Malformations and Morphogenesis: Abstracts of the 2019 Annual Meeting. Am J Med Genet Part A. 2020;182:877–942.
7. Nevel RJ, Brennan SK, Wambach JA. Genetic Disorders of Surfactant Metabolism. Neoreviews. 2025;26:e328–38.
8. Mio C, Baldan F, Damante G. NK2 homeobox gene cluster: Functions and roles in human diseases. Genes Dis. 2023;10:2038–48.
9. Magnani JE, Donn SM. Persistent respiratory distress in the term neonate: genetic surfactant deficiency diseases. Curr Pediatr Rev. 2020;16:17–25.
10. van Moorsel CHM, van der Vis JJ, Grutters JC. Genetic disorders of the surfactant system: focus on adult disease. Eur Respir Rev. 2021;30.
11. Megeid AKA, Refeat MM, Ashaat EA, El-Kamah G, El-Saiedi SA, Elfalaki MM, et al. Correlating SFTPC gene variants to interstitial lung disease in Egyptian children. J Genet Eng Biotechnol. 2022;20:117.
12. Hamvas A, Chaudhari BP, Nogee LM. Genetic testing for diffuse lung diseases in children. Pediatr Pulmonol. 2024;59:2286–97.
13. Bullard JE, Wert SE, Nogee LM. ABCA3 deficiency: neonatal respiratory failure and interstitial lung disease. In: Seminars in perinatology. Elsevier; 2006. p. 327–34.
14. Fasano C, Lepore Signorile M, De Marco K, Forte G, Disciglio V, Sanese P, et al. In Silico deciphering of the potential impact of variants of uncertain significance in hereditary colorectal cancer syndromes. Cells. 2024;13:1314.
15. Varadi M, Bertoni D, Magana P, Paramval U, Pidruchna I, Radhakrishnan M, et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. Nucleic Acids Res. 2024;52:D368–75.
16. Yang C, Sun X, Wu G. New insights into GATOR2-dependent interactions and its conformational changes in amino acid sensing. Biosci Rep. 2024;44.
17. Kucukkal TG, Yang Y, Chapman SC, Cao W, Alexov E. Computational and experimental approaches to reveal the effects of single nucleotide polymorphisms with respect to disease diagnostics. Int J Mol Sci. 2014;15:9670–717.