



## OPEN ACCESS

## EDITED BY

Paolo Montuschi,  
Catholic University of the Sacred Heart, Italy

## REVIEWED BY

Li Wang,  
Yantai Infectious Diseases Hospital, China  
Qiang Zhang,  
China Medical University, China

## \*CORRESPONDENCE

Sanjoy Podder  
✉ skpzoo2@rediffmail.com

RECEIVED 20 August 2025

REVISED 10 November 2025

ACCEPTED 21 November 2025

PUBLISHED 12 December 2025

## CITATION

Sultana N, Adhikari H, Goswami AM,  
Mondal A, Ganai I, Biswas H, Iqbal A, Das A,  
Moitra S and Podder S (2025) *SCGB1A1*  
rs3741240 variant downregulates CC16:  
a molecular insight into COPD  
pathogenesis in Indian population.  
*Front. Immunol.* 16:1689340.  
doi: 10.3389/fimmu.2025.1689340

## COPYRIGHT

© 2025 Sultana, Adhikari, Goswami, Mondal,  
Ganai, Biswas, Iqbal, Das, Moitra and Podder.  
This is an open-access article distributed under  
the terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# *SCGB1A1* rs3741240 variant downregulates CC16: a molecular insight into COPD pathogenesis in Indian population

Nasima Sultana<sup>1</sup>, Himani Adhikari<sup>1</sup>, Achintya Mohan Goswami<sup>2</sup>,  
Amalesh Mondal<sup>3</sup>, Indranil Ganai<sup>1</sup>, Himani Biswas<sup>4</sup>, Asif Iqbal<sup>5</sup>,  
Aratrika Das<sup>5</sup>, Saibal Moitra<sup>5</sup> and Sanjoy Podder<sup>1\*</sup>

<sup>1</sup>Ecology and Allergology Laboratory, Department of Zoology, The University of Burdwan,  
Burdwan, India, <sup>2</sup>Department of Physiology, Krishnagar Government College, Krishnagar, India,  
<sup>3</sup>Department of Physiology, Katwa College, Katwa, India, <sup>4</sup>Post Graduate Department of Zoology,  
Lady Brabourne College, Kolkata, India, <sup>5</sup>Apollo Multispecialty Hospitals, Kolkata, India

**Background:** Chronic Obstructive Pulmonary Disease (COPD) is a progressive inflammatory lung disorder influenced by environmental and genetic factors. The rs3741240 polymorphism in the *SCGB1A1* gene, which encodes the anti-inflammatory protein CC16, is considered a genetic marker of COPD susceptibility.

**Objective:** This study aimed to investigate the functional significance of the *SCGB1A1* rs3741240 polymorphism in 224 COPD patients and 194 controls in the West Bengal population, India.

**Methods:** Genotyping of rs3741240 was performed using Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP). *SCGB1A1* mRNA levels were quantified in healthy controls and among different genotypes using Real-Time PCR. Protein expression was assessed using Western blotting. *In silico* analyses identified miRNAs and transcription factors that target the *SCGB1A1* promoter using the miRDB, RNA22 v2, and hTFtarget web servers. Structural modeling and docking studies were conducted to analyze miRNA-mRNA-AGO2 and TF-promoter interactions using the I-TASSER, SWISS-MODEL, GalaxyGemini, RNAfold, RNA COMPOSER, HDock, and DNAProDB web servers.

**Results:** The genotype and allele frequencies of the 38AA risk genotype were more prevalent in COPD patients than in controls ( $P = 0.01$  and  $0.03$ , respectively). Patients with 38AA genotype exhibited significantly lower FEV1 and FEV1/FVC ( $P < 0.0001$ ). *SCGB1A1* mRNA and protein expression were significantly reduced in patients bearing the 38AA genotype compared to those bearing the GA and GG genotypes. Bioinformatics analyses suggested that reduced CC16 expression for rs3741240 might be mediated by miRNA hsa-miR-11181-3p and the transcription factors SMAD1, HAND1, and HAND2.

**Conclusion:** This study is the first to reports in an Indian population that *SCGB1A1* rs3741240 is linked to COPD via CC16 downregulation, revealing a key pathogenic mechanism and potential for precision medicine in management.

#### KEYWORDS

COPD, *SCGB1A1*, SNPs, bronchial epithelial cell, immunofluorescence, gene expression, immune regulation, precision immunology

## Introduction

Chronic obstructive pulmonary disease (COPD) is a group of progressive inflammatory lung conditions, primarily chronic bronchitis and emphysema (1). It leads to airflow obstruction and breathing difficulties and is associated with an increased risk of cardiovascular and other pulmonary complications (2). Globally, an estimated 3.23 million people die from COPD each year, making it the third leading cause of death (3). As COPD is both progressive and currently incurable, early diagnosis and the development of effective treatments are essential (4). While long-term exposure to cigarette smoke is the most common cause of COPD, genetic factors play a crucial role in individual susceptibility to COPD (5). One such candidate is the single nucleotide polymorphism (SNP) rs3741240 in the *SCGB1A1* gene, which encodes Clara Cell Secretory Protein (CC16), a key anti-inflammatory molecule in the airway epithelium (6). CC16, a member of the secretoglobulin family, is the most abundant protein found in normal airway secretions and is primarily produced by non-ciliated club cells of the distal airway epithelium (7). Other non-ciliated epithelial cells also contribute to its production at lower levels (8).

Lin and colleagues demonstrated that CC16 promotes epithelial repair and mitigates both inflammation and pyroptosis in PM<sub>2.5</sub>-exposed asthmatic mice (9). Furthermore, a 2022 cohort study of U.S. veterans identified interactions between CC16 levels, smoking status, and lung function, showing that elevated CC16 levels in former smokers were associated with improvements in FEV<sub>1</sub>/FVC ratios (8). Thus, research over several decades suggests that COPD, cigarette smoke, and CC16 are strongly associated with and collectively affect lung function. A longitudinal birth cohort study demonstrated that reduced CC16 levels during childhood are associated with impaired lung function and increased airway hyperreactivity that persists into adulthood (10). Chen et al. (11) reported reduced serum and BALF CC16 levels in COPD patients and cigarette smoke-exposed monkeys. Lacho-Contreras et al. (12) found that decreased CC16 correlates with greater airflow obstruction.

The SNP rs3741240 in the 5' untranslated region/promoter of the *SCGB1A1* gene has been identified as a key regulatory variant significantly associated with CC16 expression levels (13). Furthermore, the *SCGB1A1* rs3741240 polymorphism has been

identified as a major genetic determinant influencing circulating CC16 levels, lung function decline, and susceptibility to COPD (14–16).

Although growing global evidence links *SCGB1A1* variants and CC16 levels to COPD pathogenesis, a significant geographical and ethnic gap persists in this area of research. Most studies investigating the *SCGB1A1* rs3741240 variant and its impact on CC16 expression have been conducted in Western populations, with minimal data available from Indian cohorts, despite India's distinct genetic background, diverse environmental exposures, and high COPD burden.

Therefore, this study aimed to evaluate the relationship between the *SCGB1A1* rs3741240 polymorphism and CC16 downregulation and to examine its potential role in the molecular pathogenesis of COPD in the West Bengal population.

## Materials and methods

### Study subject

This study was approved by the Clinical Research Ethics Committee of the Allergy and Asthma Research Centre, West Bengal, India (CREC-AARC Ref: 62/204). Informed consent was obtained from all the participants. Epidemiological information, including age, sex, weight, height, residence, occupation, and smoking habits, was collected using a standard questionnaire. A total of 447 individuals, both male and female, aged 20 years or older were selected for this study. Among them, 28 participants were excluded due to other respiratory diseases, pregnancy, or failure to provide a written informed consent. Finally, of the 418 selected participants, 224 were clinically diagnosed with COPD.

### Subject classification

The patient groups were further classified into young (20–50 years) and old (>50 years) COPD groups according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria, 2025 (17). Both patient and control participants were classified according to sex (male and female), age (20–50 and >50 years),

smoking status (current smoker, occasional smoker, ex-smoker, and non-smoker), and disease severity (mild, moderate, severe, and very severe).

and serum was separated by centrifugation (1,200 rpm, 10 min). Serum-free samples were stored at 4 °C for genotyping.

Pulmonary function test

Hospital pulmonologists diagnosed COPD based on clinical examination and spirometry (GOLD criteria, 2025) (17). COPD was confirmed by a post-bronchodilator FEV<sub>1</sub>/FVC <0.7. The severity of COPD is classified based on post-bronchodilator FEV<sub>1</sub> as follows: ≥80% (mild, GOLD 1), 50%–79% (moderate, GOLD 2), 30%–49% (severe, GOLD 3), and <30% (very severe, GOLD 4) (GOLD, 2025) (17).

Blood collection

Peripheral blood (2 mL) was collected from both the patients and the control group in EDTA non-vacuum blood collection vials,

DNA extraction and genotyping

Genomic DNA was extracted from blood samples using the QIAamp DNA Blood Mini Kit (Qiagen, Germany) according to the manufacturer’s instructions. SNPs in the *SCGB1A1* gene were identified by polymerase chain reaction (PCR) followed by restriction fragment length polymorphism (RFLP) analysis (Figures 1A, B). The PCR primers, amplicon size, restriction enzymes, and expected RFLP fragment patterns are detailed in Table 1.

DNA sequencing

To validate the RFLP results, representative PCR products from each of the three genotypes for the selected polymorphisms were

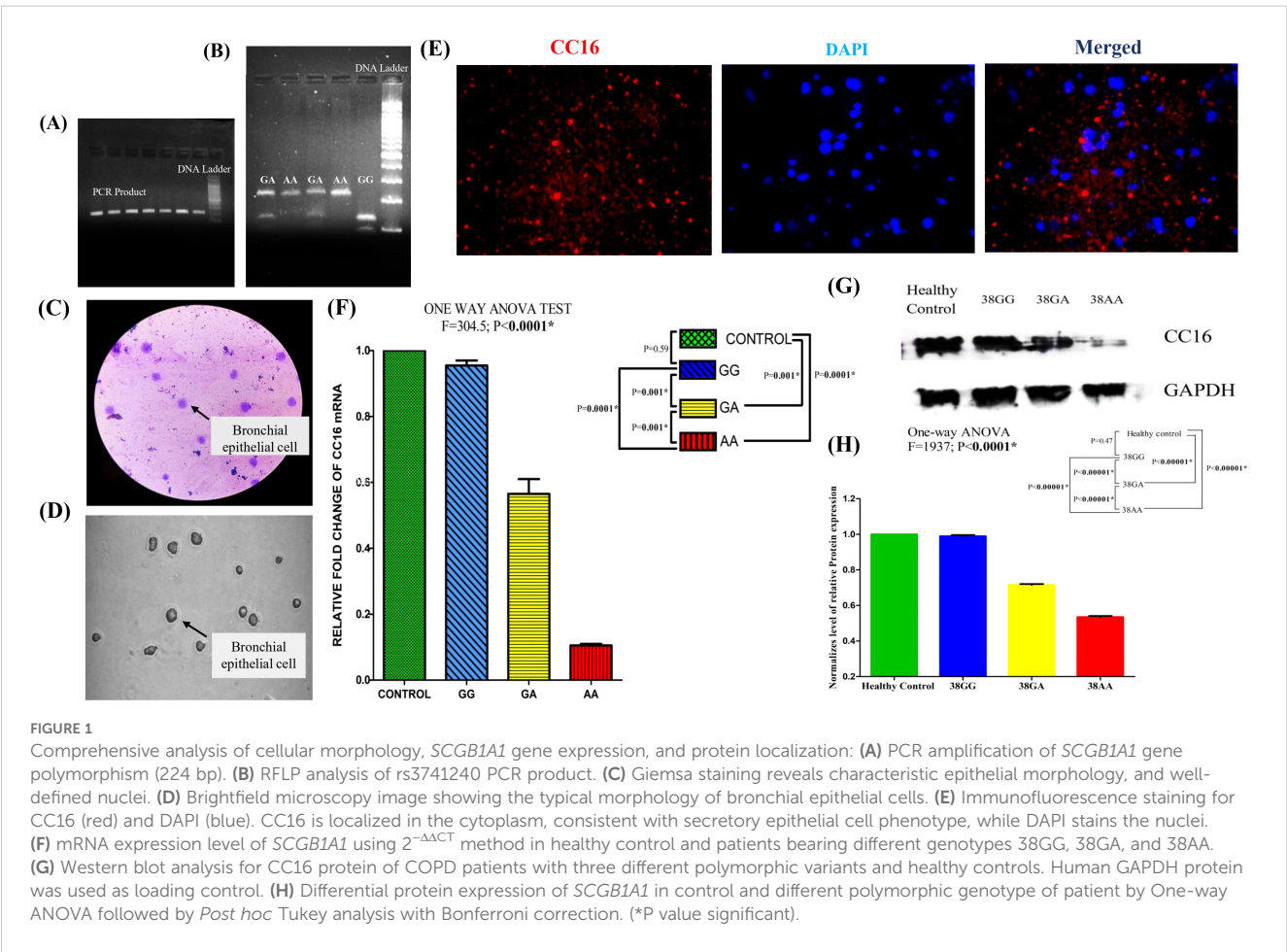


TABLE 1 PCR conditions and restriction fragments of *SCGB1A1* rs3741240 polymorphism.

SNP	Functional position	Primer sequence and PCR product size	PCR condition	Restriction enzyme, site, fragments and genotypes
SCGB1A1 rs3741240	Promoter +38 G>A	Forward: 5'- GCAGTATCTTATGTAGAGCCCT -3' Reverse: 5'- TCTCTGAGCACTACCGGAGCT -3' Product size- 224 bp	95 °C for 5 min, followed by 35 cycles at 95 °C for 30 s, 61°C for 30 s, 72 °C for 1 min, and a final extension at 72 °C for 10 min	Restriction enzyme- Sau 96 I Restriction site- G <sub>1</sub> GNCC (N = A, T, G, C) GG- 129, 95 bp AA- 224 bp GA- 224, 129 and 95 bp

subjected to Sanger sequencing (ABI 3500 Genetic Analyzer) according to the manufacturer's standard protocol.

## Bronchoscopic sampling of bronchial epithelial cells

Bronchial epithelial cells (BECs) were collected from selected participants via bronchoscopy using a sterile cytology brush. Cells were released from the pellet by vigorous agitation, filtered through a 70-µm strainer to remove mucus and debris, and centrifuged at 14,000 rpm for 10 min at 4 °C. The pellet was resuspended in PBS for RNA and protein extraction.

## Giemsa and immunofluorescence staining of BECs

BECs were fixed and stained with Giemsa (Sigma-Aldrich; 1:20) to assess their morphology. The stained cells were mounted with DPX and examined under a light microscope at ×40 magnification. For immunofluorescence staining, BEC pellets were washed with PBS, fixed in 2% paraformaldehyde (20 min, room temperature), and permeabilized with 0.2% Triton X-100 in PBS for 10 min. Cells were incubated overnight at 4 °C with CC16 rabbit polyclonal antibody (ABclonal; A16997; 1:100), washed thrice with PBS, and incubated with PE Donkey anti-rabbit IgG (Biolegend; 406421, 1:200) for 1 h at room temperature in the dark. Nuclei were stained with DAPI (1 µg/mL, 5 min), coverslips were mounted with ProLong<sup>TM</sup> Gold Antifade (Invitrogen), and images were captured using confocal microscopy. ImageJ was used to analyze CC16 localization.

## RNA and protein extraction

Bronchial epithelial cells (BECs) from COPD patients and controls were processed for RNA and protein isolation. BECs suspended in 1× PBS were used for total RNA extraction using TRIzol reagent (Invitrogen, USA) according to the manufacturer's instructions. RNA yield and purity were assessed using a NanoDrop spectrophotometer (Thermo Fisher), and RNA integrity was verified by agarose gel electrophoresis. DNase I (Qiagen) treatment was performed to remove genomic DNA before reverse transcription.

For protein isolation, BECs were lysed in RIPA buffer with protease inhibitor (50×, Promega), incubated on ice (20 min), centrifuged (12,000 rpm, for 10 min at 4 °C), and the supernatants were collected. Protein concentration was measured using the Bradford assay (Bio-Rad).

## Quantitative RT-PCR analysis

Real-time reverse transcription polymerase chain reaction (RT-PCR) was performed using 2 µL of RNA with the Hi-Quanti One Step Probe-Based RT-PCR Kit (HiMedia, India), according to the manufacturer's protocol. *SCGB1A1* was amplified using the following primers: forward 5' ATGAAGATCGCCATCACAA TCA 3' and reverse 5' GAATCTTAAATCTTGCTTACACAG 3'. The RT-PCR conditions were as follows: 95 °C for 1 min, followed by 40 cycles of 95 °C for 10 s and 60 °C for 40 s. Expression was quantified using the  $2^{-\Delta\Delta CT}$  method (18), with GAPDH as the loading control. Each reaction was conducted in triplicate across three biological replicates for the 38GG, 38GA, and 38AA genotypes of rs3741240 in a case-control design.

## Western blot analysis

Western blotting was performed on CC16 protein extracts from BECs of COPD patients and controls, using three representative samples per *SCGB1A1* variant. Proteins (20 µg –40 µg) were separated on 10% SDS-PAGE, transferred to nitrocellulose membranes, and blocked with 5% BSA (1 h, room temperature). The membranes were incubated overnight at 4 °C with primary antibodies: SCGB1A1 Rabbit pAb (ABclonal; A16997, 1:1,000) and anti-GAPDH (Bio Bharati Life Science; BB-AB0060, 1:2,000), followed by HRP-conjugated secondary antibody (1:5,000, 1 h, room temperature). Bands were visualized and quantified using ImageJ software, with CC16 normalized to GAPDH.

## In silico study

### Identification of microRNAs targeting *SCGB1A1* promoter

We used the miRDB web server to identify microRNAs (miRNAs) that target the *SCGB1A1* promoter (19). The query promoter sequence (both wild type and mutant) of *SCGB1A1* was



provided as input, followed by the selection of the 5' UTR from the dropdown list. miRDB predicted the miRNAs targeting the 5' UTR of both the wild-type and mutant *SCGB1A1* promoters. The RNA22 v2 web server was used to obtain miRNA–target interaction (MTI) data for the target miRNAs in both wild-type and mutant *SCGB1A1* promoters (20, 21). The FASTA sequences of miRNAs and both the wild-type and mutant *SCGB1A1* promoters were used as inputs in RNA22 v2 to obtain MTI details.

## 2D and 3D structural prediction of miRNA–mRNA duplexes

We used the RNAfold web server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>) to obtain the optimal secondary structure of both wild-type and mutant miRNA–mRNA duplexes. Dot bracket notation of the miRNA–mRNA duplexes was also obtained from the server. These dot-bracket notations were used as inputs for the RNA COMPOSER web server to obtain the 3D structures of both wild-type and mutant miRNA–mRNA duplexes (22).

## Retrieval of human AGO2 structure and its preparation

The 3D crystal structure of the human AGO2 (hAGO2) protein (PDB ID: 4F3T) was retrieved from the Protein Data Bank. The hAGO2 protein was prepared by removing the heteroatoms and water molecules using BIOVIA Discovery Studio 2020. To address the missing atoms observed in hAGO2, we homology-modeled the protein in I-TASSER using cleaned hAGO2 as a template. The newly modeled hAGO2 was energy-minimized for docking using Swiss-PDBViewer 4.1.0.

## Docking study between miRNA–mRNA duplexes and human AGO2 protein

The HDock web server was used for docking between the miRNA–mRNA duplexes and hAGO2 protein (23). We chose the miRNA–mRNA duplex–HAGO2 complex based on the best docking scores provided by the server algorithm.

## Identification of transcription factors targeting the *SCGB1A1* promoter

We used the hTFtarget web server to identify TFs targeting the selected nucleotide stretch (ACCAGAGACGGGCCAGAGCAT) in both the wild-type and mutant *SCGB1A1* promoters (24).

## Modeling of the DNA bonding domain of predicted TFs and the selected nucleotide stretch of the *SCGB1A1* promoter

The DNA-binding domains of the predicted TFs were modeled using template-guided homology modeling in SWISS-MODEL. The modeled DNA bonding domain of the predicted TFs was dimerized using the GalaxyGemini web server, while both wild-type and mutant *SCGB1A1* promoters (selected nucleotide stretch) were constructed using Discovery Studio (25).

## Docking study between predicted TFs and *SCGB1A1* promoter with selected nucleotide stretch

We used the HDock web server to dock the predicted TFs with both the wild-type and mutant *SCGB1A1* promoters (23).

## DNA–protein interaction analyses

The DNAproDB web server was used for interaction analyses of DNA (*SCGB1A1* promoter)–protein (TF) docked complexes (26).

## Data analysis

Categorical data are presented as numbers with percentages, and continuous data are expressed as median  $\pm$  standard deviation (SD). The Pearson's chi-square test was used to analyze categorical variables. For continuous variables, the independent *t*-test was applied to normally distributed data, and the Mann–Whitney *U* test was used for non-normally distributed data. For variables with non-normal distributions, the interquartile range (IQR) represents the middle 50% of the data and is calculated as the difference between the third (Q3) and first (Q1) quartiles. Data points falling outside the range ( $Q1 - 1.5 \times IQR$ ,  $Q3 + 1.5 \times IQR$ ) were considered as potential outliers. The contingency chi-square test was used to assess the genotype and allele distributions across the study groups. Risk analysis for additive, recessive, and dominant genetic models was performed using odds ratios (OR). The Hardy–Weinberg equilibrium (HWE) was evaluated using the one-degree-of-freedom chi-square goodness-of-fit test. A Generalized Linear Model (GLM) was employed with binomial logit and quasi-Poisson log links. In these models, risk genotypes were treated as dependent variables, whereas disease severity, age, and residence were included as independent predictors. A one-way ANOVA followed by Tukey's Honestly Significant Difference (HSD) test was used to examine the association between FEV<sub>1</sub>, FEV<sub>1</sub>/FVC ratio, and genetic polymorphisms. Fold changes in mRNA and protein expression levels of *SCGB1A1* among controls and patient genotypes were analyzed using one-way ANOVA followed by *post hoc* Tukey analysis with Bonferroni correction. All statistical analyses were performed using GraphPad Prism version 7 (San Diego, CA, USA), with a significance threshold set at  $P < 0.05$ .

# Results

## Study subjects

The demographic and clinical characteristics of the 224 patients and 194 controls are presented in Table 2. Significant differences were observed between the patients and controls in terms of age, residence, smoking habits, and disease severity. Among the 224

TABLE 2 Demographic and clinical characteristics of case and control subjects.

Sl. No	Parameters	Category	Case (n = 224)	Control (n = 194)	P value
1.	Age, n (%)	20–50	76 (33.93%)	119 (61.34%)	0.0001*
		>50	148 (66.07%)	75 (38.66%)	
2.	Sex, n (%)	Males	118 (60.82%)	102 (52.58%)	0.10
		Females	106 (54.64%)	92 (47.42%)	
3.	Family history	Paternal	48 (21.43%)	Nil	–
		Maternal	72 (32.14%)	Nil	
		P + M	25 (11.16%)	Nil	
		Absent	79 (35.27%)	194 (100%)	
4.	Weight, mean $\pm$ SD (kg)	–	63.16 $\pm$ 10.24 (61.81–64.51)	63.89 $\pm$ 10.24 (62.44–65.34)	0.45
5.	Height, mean $\pm$ SD (cm)	–	166.97 $\pm$ 8.95 (165.79–168.15)	167.06 $\pm$ 8.89 (165.80–168.32)	0.92
6.	Residence, n (%)	Urban	44 (19.64%)	88 (45.36%)	0.0003*
		Semi-urban	86 (38.39%)	59 (30.41%)	
		Rural	94 (41.96%)	47 (24.23%)	
7.	Occupation, n (%)	Industry	54 (24.11%)	46 (23.71%)	0.98
		Transport	47 (20.98%)	38 (19.59%)	
		Agriculture	31 (13.84%)	26 (13.40%)	
		Service	57 (25.45%)	47 (24.23%)	
		Unemployed	35 (15.63%)	37 (19.07%)	
8.	Smoking Habit, n (%)	Current Smoker	87 (38.84%)	17 (8.76%)	<0.0001*
		Occasional Smoker	28 (12.50%)	23 (11.86%)	
		Ex-smoker	16 (7.14%)	21 (10.82%)	
		Non-smoker	93 (41.52%)	133 (68.56%)	
10.	Disease severity, n (%)	Mild (GOLD 1)	61 (27.23%)	–	–
		Moderate (GOLD 2)	67 (29.91%)		
		Severe (GOLD 3)	52 (23.21%)		
		Very severe (GOLD 4)	44 (19.64%)		
11.	DLco, mean $\pm$ SD (%)	–	58.56 $\pm$ 25.83 (55.16–61.96)	97.94 $\pm$ 11.23 (96.35–99.53)	<.00001*
12.	FeNO, median (IQR) ppb	–	36	13	<.00001*
13.	FVC, mean $\pm$ SD (Liters)	–	3.69 $\pm$ 0.25 (3.66–3.72)	3.66 $\pm$ 0.25 (3.63–3.69)	0.06
14.	FEV1, mean $\pm$ SD (Liters)	–	2.04 $\pm$ 0.64 (1.96–2.12)	2.90 $\pm$ 0.46 (2.82–2.98)	<.00001*

(\*P value significant).

patients, 33.93% were classified as having Young COPD, while 66.07% had Old COPD. In this study, the majority of patients were from rural areas (41.96%), followed by urban and semi-urban regions. In terms of disease severity, 27.23% had mild COPD, 29.91% had moderate COPD, 23.21% had severe COPD, and 19.64% had very severe COPD. Compared to controls, patients showed significant differences in key pulmonary function parameters, including fractional exhaled nitric oxide (FeNO),

diffusing capacity of the lungs for carbon monoxide (DLco), and forced expiratory volume in 1 s (FEV1).

## Genotyping by PCR-RFLP

Genotyping of the SCGB1A1 rs3741240 variant was confirmed using PCR-RFLP analysis. The PCR product (224 bp) was digested

TABLE 3 Genotypic frequency distribution of SCGB1A1 rs3741240 polymorphism in patients and control.

Genotype	Model	Case (n = 224)			Control (n = 194)			Chi Square Value (P value)	OR (95% CI)			P value		MAF	
I. Overall distribution															
GG	Additive	102 (45.74%)			121 (62.37%)			10.29 (P = 0.01*)	1.00			–		0.21	
GA		91 (40.81%)			67 (34.54%)				1.61 [0.89–2.92]			0.11			
AA		31 (13.90%)			6 (3.09%)				6.43 [1.74–23.70]			0.01*			
GG+GA	Recessive	193 (86.16%)			188 (96.91%)			6.43 (P = 0.01*)	1.00			–			
AA		31 (13.90%)			6 (3.09%)				5.26 [1.46–18.94]			0.01*			
GG	Dominant	102 (45.74%)			121 (62.37%)			4.53 (P = 0.03*)	1.00			–			
GA+AA		122 (54.46%)			73 (37.63%)				1.92 [1.09–3.37]			0.01*			
G		295 (65.85%)			309 (79.64%)			4.97 (P = 0.03*)	1.00			–			
A		153 (34.15%)			79 (20.36%)				2.06 [1.09–3.91]			0.03*			
II. Distribution according to sex															
Genotype		Male (n = 118)		Female (n = 106)		Male (n = 102)		Female (n = 92)		Male		Female		Male	Female
GG	–	57 (48.31%)		45 (42.45%)		64 (62.75%)		57 (61.95%)	24.79 (P = 0.0004*)	1.00		1.00		–	–
GA		48 (40.68%)		43 (40.57%)		33 (32.35%)		34 (36.96%)		1.68 [0.93 - 3.05]		1.64 [0.90-2.96]		0.09	0.10
AA		13 (11.01%)		18 (16.98%)		5 (4.90%)		1 (1.10%)		2.89 [0.94 - 8.87]		25.10 [3.22-195.82]		0.06	0.002*
III. Distribution according to age (Years)															
Genotype		Young COPD case [20–50] (n = 76)		Old COPD case [>50] (n = 148)		20-50 (n = 119)		>50 (n = 75)		20–50		>50		20-50	>50
GG	–	37 (48.68%)		65 (43.91%)		75 (63.03%)		46 (61.33%)	24.37 (P = 0.0004*)	1.00		1.00		–	–
GA		31 (40.79%)		60 (40.54%)		38 (31.93%)		29 (38.67%)		1.65 [0.91–2.98]		1.46 [0.81–2.62]		0.10	0.21
AA		8 (10.53%)		23 (15.54%)		6 (5.04%)		0 (0.00%)		2.82 [0.92–8.68]		45.61 [2.67–780.49]		0.07	0.01*
IV. Distribution according to Residence															
Genotype		Urban (n = 44)	Semi-urban (n = 86)	Rural (n = 94)	Urban (n = 88)	Semi-urban (n = 59)	Rural (n = 47)		Urban	Semi-urban	Rural	Urban	Semi-urban	Rural	
GG	–	22 (50.00%)	39 (45.35%)	41 (43.62%)	55 (62.50%)	34 (57.63%)	32 (68.09%)	30.42 (P = 0.001*)	1.00	1.00	1.00	–	–	–	

(Continued)

TABLE 3 Continued

IV. Distribution according to Residence															
Genotype		Urban (n = 44)	Semi-urban (n = 86)		Rural (n = 94)	Urban (n = 88)	Semi-urban (n = 59)	Rural (n = 47)		Urban	Semi-urban	Rural	Urban	Semi-urban	Rural
GA		17 (38.64%)	36 (41.86%)		38 (40.43%)	29 (32.95%)	24 (40.68%)	14 (29.79%)		1.47 [0.90–2.96]	1.32 [0.74–2.36]	6.14 [1.94–19.71]	0.21	0.35	0.002*
AA		5 (11.36%)	11 (12.79%)		15 (15.96%)	4 (4.55%)	1 (1.69%)	2 (4.26%)		2.73 [0.89–8.37]	7.73 [1.65–36.32]	2.06 [1.12–3.78]	0.08	0.01*	0.02*
V. Distribution according to severity of airflow obstruction in COPD															
Genotype		Mild (GOLD 1) (n = 61)	Moderate (GOLD 2) (n = 67)	Severe (GOLD 3) (n = 52)	Very severe (GOLD 4) (n = 44)	Control		A	B	C	D	A	B	C	D
GG	–	30 (49.18%)	33 (49.25%)	21 (40.39%)	18 (40.90%)	121 (62.37%)	28.00 (P = 0.001*)	1.00	1.00	1.00	1.00	–	–	–	–
GA		26 (42.63%)	28 (41.79%)	20 (38.46%)	17 (38.64%)	67 (34.54%)		1.55 [0.87–2.78]	1.52 [0.85–2.72]	1.73 [0.94–3.19]	1.69 [0.92–3.08]	0.14	0.16	0.07	0.09
AA		5 (8.20%)	6 (8.95%)	11 (21.15%)	9 (20.45%)	6 (3.09%)		3.37 [0.85–13.40]	3.80 [0.97–14.78]	10.85 [3.04–38.77]	10.08 [2.81–36.12]	0.08	0.05	0.0002*	0.0004*

Significant associations are shown in asterisk\*.  
CI, confidence interval; MAF, Minor allele frequency in control group; OR, Odds ratio.  
A- OR against GOLD 1 patient vs control and its corresponding P value.  
B- OR against GOLD 2 patient vs control and its corresponding P value.  
C- OR against GOLD 3 patient vs control and its corresponding P value.  
D- OR against GOLD 4 patient vs control and its corresponding P value.

to yield distinct fragments: GG = 129 bp + 95 bp, GA = 224 bp + 129 bp + 95 bp, and AA = 224 bp. Representative PCR and RFLP gel images are shown in **Figures 1A, B**.

Allele and genotype analysis

The allele and genotype distributions of the rs3741240 polymorphism in our study population are shown in **Table 3**. The SNP did not show a significant deviation from HWE in the control population ( $\chi^2 = 0.54$ ,  $df = 1$ ). In our study, we found significant differences in genotype frequency under the additive ( $P = 0.01$ ), recessive ( $P = 0.01$ ), and dominant ( $P = 0.03$ ) models between the case and control populations. The frequency of the AA risk genotype was significantly higher in the COPD case population than in the control population ( $OR = 6.43$ ;  $P = 0.01$ ). According to sex, age, and residence, and disease severity, the distribution of genotypes in both COPD cases and control subjects demonstrated significant differences between males and females ( $\chi^2 = 24.79$ ;  $P = 0.0004$ ), Young and Old COPD cases ( $\chi^2 = 24.37$ ;  $P = 0.0004$ ), rural, urban, and semi-urban areas ( $\chi^2 = 30.42$ ;  $P = 0.001$ ), as well as mild, moderate, severe, and very severe patients ( $\chi^2 = 28.00$ ;  $P = 0.001$ ), respectively. In female patients, the AA risk genotype was significantly associated with a higher COPD risk than in male patients ( $OR = 25.10$ ;  $P = 0.002$ ). The risk genotype in old COPD cases was significantly associated with a higher COPD risk than in young COPD cases ( $OR = 45.61$ ;  $P = 0.01$ ). Additionally, risk genotypes were associated with a higher risk of COPD in semi-urban and rural areas than in urban areas ( $OR = 7.73$ ;  $P = 0.01$  and  $OR = 2.06$ ;  $P = 0.02$ , respectively). Furthermore, the risk genotype was associated with a higher risk of COPD in patients with severe and very severe disease than in those with mild or moderate disease ( $OR = 10.85$ ;  $P = 0.0002$  and  $OR = 10.08$ ;  $P = 0.0004$ , respectively). The Power of the sample size was also calculated to identify statistically significant risk factors [Power ( $1 - \beta$ ) = 88.63%].

Generalized Linear Model (GLM) analysis indicated that old age, disease severity, and rural residence were positively associated with the rs3741240 AA risk genotype. The detailed findings of the GLM analysis are presented in **Table 4**.

FEV<sub>1</sub> and SNP genotypes

The association of FEV<sub>1</sub> and FEV<sub>1</sub>/FVC ratio with cases and controls bearing the rs3741240 polymorphism is shown in **Figure 2**. One-way ANOVA revealed that a significant difference existed in FEV<sub>1</sub> and FEV<sub>1</sub>/FVC ratio of patients bearing polymorphic genotypes ( $F =$

485.9,  $P < 0.0001$ ;  $F = 166.6$ ,  $P < 0.0001$ , respectively), while those factors were non-significant in the control population ( $F = 2.26$ ,  $P = 0.11$ ;  $F = 1.41$ ,  $P = 0.25$ , respectively). Both FEV<sub>1</sub> ( $P < 0.00001$ ) and FEV<sub>1</sub>/FVC ( $P < 0.0001$ ) were considerably lower in patients with AA risk genotypes.

Localization of CC16 protein in bronchial epithelial cells

To localize the CC16 protein in BECs, Giemsa and immunofluorescence staining techniques were used. Giemsa staining and brightfield images of BECs provided an initial morphological assessment of the bronchial epithelial cell monolayer (**Figures 1C, D**). As shown in **Figure 1C**, cytoplasmic granularity was clearly observed in a subset of cells. Immunofluorescence staining revealed granular and cytoplasmic staining with no significant nuclear signal. The nuclei were counterstained with DAPI to allow a clear distinction between the nuclear and cytoplasmic compartments (**Figure 1E**).

SCGB1A1 mRNA expression in polymorphic variants via qRTPCR analysis

Quantitative real-time PCR analysis revealed that SCGB1A1 mRNA expression was significantly decreased in BECs from COPD patients compared to healthy controls (**Figure 1F**). One-way ANOVA revealed significant differences in expression between patients with polymorphic genotypes and controls ( $F = 304.5$ ;  $P < 0.0001$ ). The *Post Hoc* Tukey test with Bonferroni correction showed significant differences in SCGB1A1 expression between 38GA ( $P = 0.001$ ) and 38AA mutant genotypes ( $P = 0.0001$ ) and the wild-type 38GG genotype in COPD patients. The lowest expression level was observed in patients with 38AA SNPs. No significant difference in SCGB1A1 expression was observed between patients with the wild 38GG genotype and healthy controls ( $p = 0.59$ ), suggesting that SCGB1A1 downregulation is associated with different rs3741240 genotypes in COPD patients.

Assessment of CC16 protein expression by Western blot

Western blot analysis of CC16 protein levels showed a significant reduction in COPD patients with different polymorphic genotypes compared to healthy controls. The intensity of the CC16 band was markedly reduced in COPD patients compared to that in controls (**Figure 1G**).

TABLE 4 Generalized linear model (GLM) analysis to assess the impact of different demographic parameters on risk genotypes of SCGB1A1 rs3741240 (\*P value significant).

Demographic parameter	Estimate	Std. error	t-value	P-value
Disease Severity	0.71	0.10	6.90	<0.00001*
Age	0.04	0.01	3.71	0.0004*
Demographic parameter	Estimate	Std. Error	z-value	P-value
Residence	1.59	0.46	3.46	0.001*



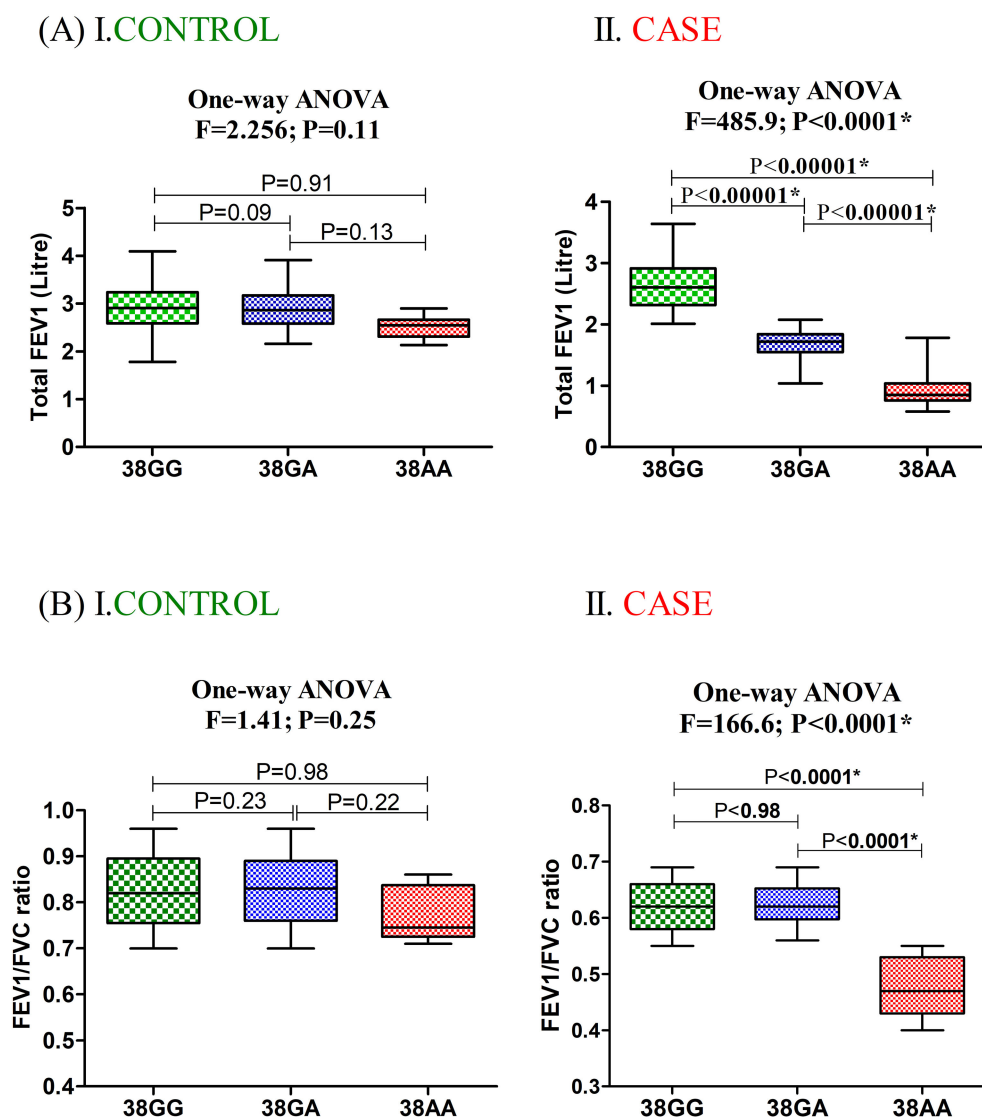


FIGURE 2

Level of (A) FEV1 and (B) FEV1/FVC ratio in case and control population bearing different genotypes for *SCGB1A1* rs3741240. (\* P value significant).

Statistical analysis using one-way ANOVA showed a significant difference in CC16 protein expression among the study groups ( $F = 1937$ ,  $P < 0.0001$ ). *Post hoc* comparisons using Tukey's test with Bonferroni correction demonstrated that patients with the 38AA risk genotype exhibited significantly lower CC16 expression levels than those with the 38GA and 38GG genotypes ( $P < 0.00001$ ) and healthy controls ( $P < 0.00001$ ) (Figure 1H).

### *In silico* study of *SCGB1A1* polymorphic variants

#### Analyses on alteration of miRNA–target interactions due to SNP in *SCGB1A1* promoter

Our study investigated the impact of an SNP in the *SCGB1A1* promoter on miRNA-mediated regulation of *SCGB1A1* expression.

In the current study, we identified 13 miRNAs targeting the 5' UTR of *SCGB1A1* from miRDB (Table 5).

These 13 miRNAs were analyzed using RNA22 v2 to obtain MTI details (Table 6). Among these 13 miRNAs, two miRNAs (hsa\_miR\_11181\_3p and hsa\_miR\_6804\_5p) bound in the vicinity of the mutation site. The binding affinity of these two miRNAs with both wild-type and mutant *SCGB1A1* promoter was measured in terms of their corresponding folding energy in kcal/mol from RNA 22v2. As shown in Table 6, hsa-miR-11181-3p exhibited a stronger binding affinity for the mutant promoter than for the wild-type (WT) promoter, whereas hsa-miR-6804 demonstrated a higher binding affinity for the wild-type *SCGB1A1* promoter than for its mutant counterpart. The predicted dot bracket notations (Table 6) and optimal secondary structures (Figure 3A) of the two miRNA–mRNA duplexes, each corresponding to the WT and mutant promoters, were obtained using the RNAfold web server.

TABLE 5 Details of 13 predicted miRNAs from miRDB.

miRNA name	Target score		Seed location	
	Wild type	Mutant	Wild type	Mutant
hsa-miR-340-3p	82	76	112, 124	112
hsa-miR-6827-3p	80	67	112, 125	112
hsa-miR-654-5p	80	80	98	98
hsa-miR-541-3p	80	80	98	98
hsa-miR-6804-5p	72	72	169	169
hsa-miR-4653-5p	71	72	108, 217	108, 217
hsa-miR-3921	71	72	108, 217	108, 217
hsa-miR-765	63	63	145	145
hsa-miR-5088-5p	62	62	16	16
hsa-miR-2277-3p	61	61	163	163
hsa-miR-500b-3p	59	59	41	41
hsa-miR-1301-5p	54	54	55	55
hsa-miR-11181-3p	50	51	145	145

The 3D structures of the WT and mutant miRNA–mRNA (*SCGB1A1*) complexes were predicted using the RNA COMPOSER server, with the corresponding dot-bracket notations of the miRNA–mRNA duplexes provided as inputs (Figure 3B).

Analyses of molecular docking between mRNA–miRNA duplexes and hAGO2 protein

To investigate the mechanism underlying the alteration in gene expression due to SNP-induced changes in the binding affinity of the targeted miRNAs to the *SCGB1A1* promoter, we performed molecular docking of pre-prepared, cleaned. Energy-minimized human Argonaute 2 (hAGO2) with these two mRNA–miRNA duplexes, each in WT and mutant forms, was used with the HDock web server (Table 7, Figure 4). Consistent with the earlier observation of a higher binding affinity of hsa-miR-11181-3p to the mutant promoter, the results presented in Table 7 also support this finding, showing that the miRNA (hsa-miR-11181-3p)–target (mutant *SCGB1A1* promoter) duplex exhibited a

stronger binding affinity with hAGO2 than its WT counterpart. A similar trend was observed for hsa-miR-6804-5p, consistent with previous results.

Analyses on alteration of transcription factors–target (promoter) interactions due to SNP in *SCGB1A1* promoter

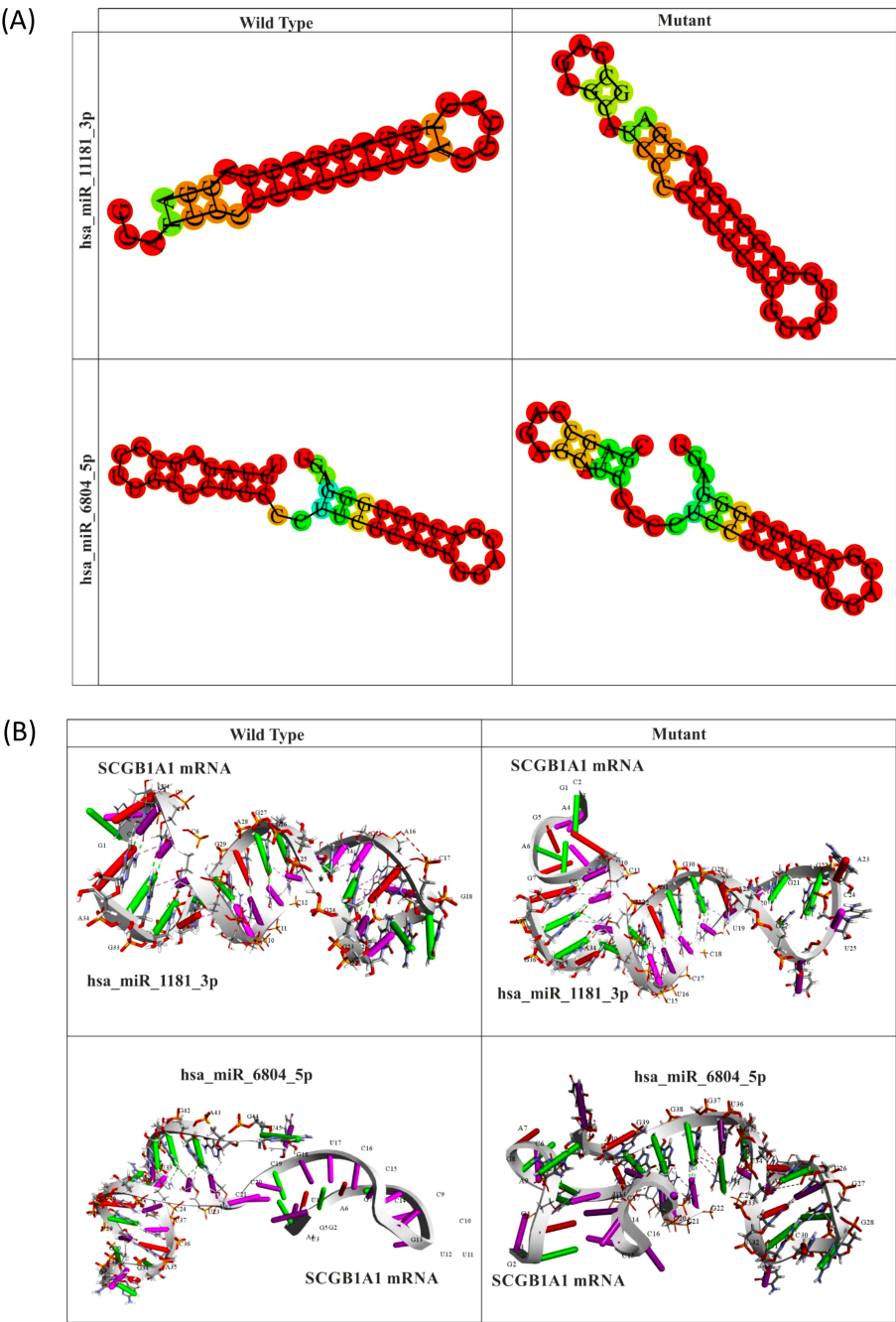
In this study, we have identified three TFs (SMAD1, HAND1, and HAND2) that target the ACCAGAGACGGGCCAGAGCAT sequence of the *SCGB1A1* promoter using the hTFtarget web server. The details of the DNA-binding domains of these three TFs, obtained from UniProt are presented in Table 8. The DNA-binding domains of these three TFs were modeled using template-guided homology modeling. GalaxyGemini dimerized the modeled structures, as dimerization is known to enhance the DNA-binding affinity and specificity of transcription factors (27). The selected nucleotide stretch (ACCAGAGACGGGCCAGAGCAT) of the promoter (both WT and mutant) was constructed using Discovery Studio. To investigate whether an SNP in the promoter region can alter the binding affinity of a transcription factor (TF), we performed molecular docking analyses using the HDock web server. The DNA-binding domains of the three selected TFs were docked to the modeled *SCGB1A1* promoter in both the wild-type and mutant forms (Table 9, Figure 5A). Table 9 shows that the mutant promoter of the *SCGB1A1* gene did not alter the binding affinity of these three TFs. The protein (TFs)–DNAproDB analyzed DNA (*SCGB1A1* Promoter) interactions for both WT and mutant promoters. The docking poses and interacting interfaces of the TF–promoter complexes are shown in Figures 5A, B.

Discussion

Chronic Obstructive Pulmonary Disease (COPD) is a complex, multifactorial respiratory condition characterized by persistent airflow limitation, chronic inflammation, and progressive lung tissue damage (28, 29). In 2015, COPD accounted for approximately 3.2 million deaths worldwide, making it the third leading cause of death among older adults (28). Its chronic nature and high prevalence place a substantial burden on healthcare systems, particularly in the aging population (29). Our study confirmed that COPD was significantly more common in individuals aged >50 than in those aged 20–50 years (Table 3). A

TABLE 6 The MTI details of two miRNAs from RNA22 v2.

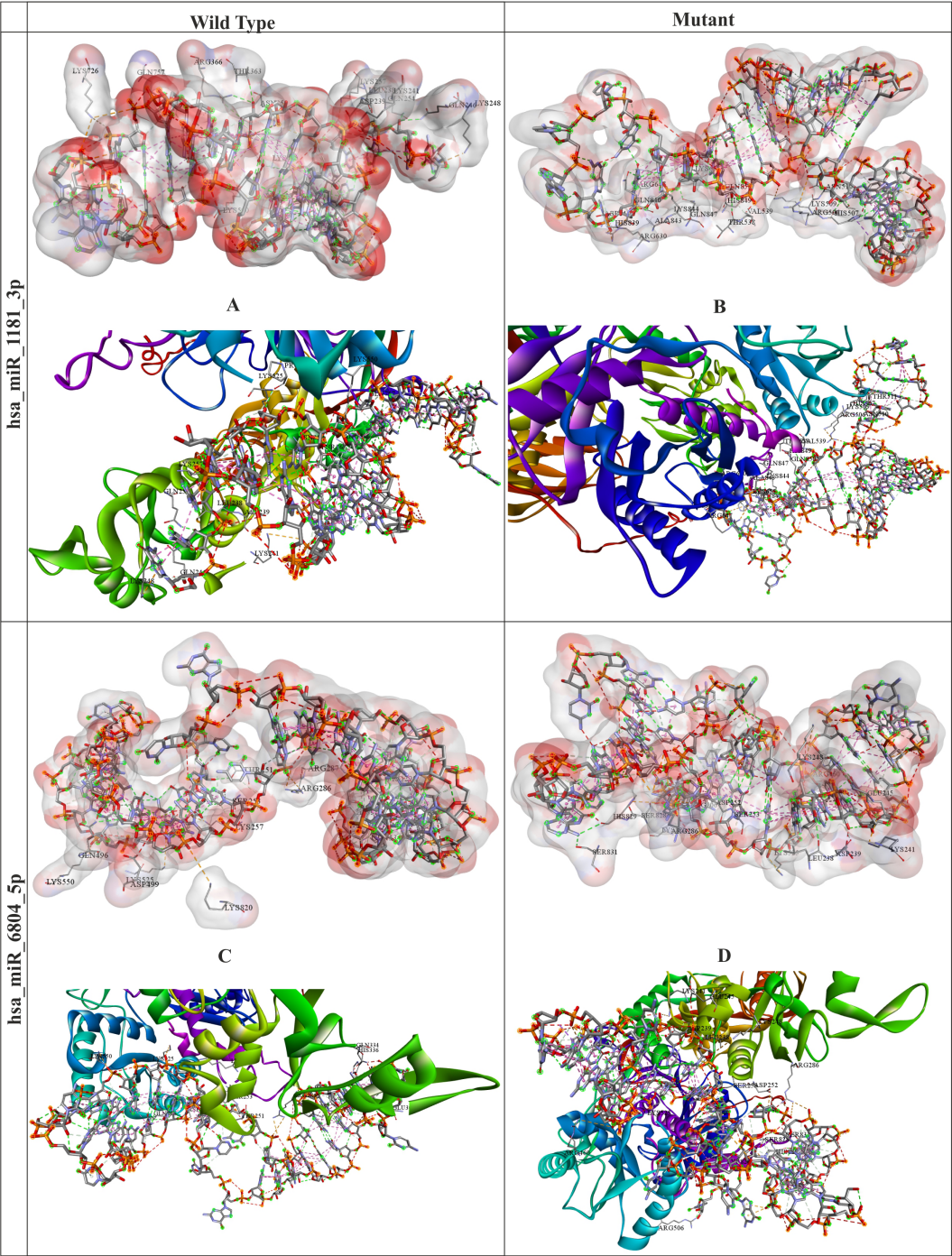
miRNAs targeting SCGB1A1 5' UTR	RNA 22v2				RNAfold	
	Wild type		Mutant		Wild type	Mutant
	Folding energy (Kcal/mol)	p value	Folding energy (Kcal/mol)	p value	Dot bracket notation	Dot bracket notation
hsa_miR_11181_3p	–20	0.0645	–21.3	0.0645	...(((.....)).....))	((.....)).(((.....)).....))
hsa_miR_6804_5p	–13.8	0.0244	–12	0.0645	.(((.....)).....).(((.....)).....) ...	.(((.....)).) .....(((.....)).....))...



**FIGURE 3**  
(A) The optimal secondary structures of both wild type and mutant miRNA-mRNA (SCGB1A1) complexes predicted by RNAfold web server.  
(B) Predicted 3D structures of both wild-type and mutant mRNA-miRNA complexes for the two microRNA-target interactions (MTIs), generated using the RNA COMPOSER server.

**TABLE 7** Docking scores along with confidence scores of WT miRNA-mRNA-AGO2 and mutant miRNA-mRNA-AGO2 complexes for both MTIs.

Model	Wild type		Mutant	
	Docking score	Confidence score	Docking score	Confidence score
hsa_miR_11181_3p VS AGO2	−308.38	0.9596	−340.11	0.9782
hsa-miR-6804-5p VS AGO2	−304.44	0.9564	−300.01	0.9526



**FIGURE 4** Molecular interactions of wild-type and mutant miRNA–mRNA–hAGO2 complexes. Panels (A, B) depict hsa\_miR\_1181\_3p, while panels (C, D) show hsa\_miR\_6804\_5p.

global study by Wang et al. (30) reported that COPD prevalence increases steadily with age, peaking in individuals aged 70–74 years, with the highest mortality rates among those aged 80 and older. Similarly, Weeks et al. (31) found that COPD prevalence rose from 0.4% in those aged 18–24 years to 10.5% in individuals aged 75 years and older.

While environmental factors such as smoking remain the principal risk factor, growing evidence supports the contribution of genetic variants in modulating individual susceptibility to COPD (5). Among these candidates, the *SCGB1A1* gene polymorphism rs3741240 has emerged as a significant contributor to disease progression, primarily through its downregulatory effect on CC16



TABLE 8 Details of DNA binding domains of three TFs.

Information types	SMAD1	HAND1	HAND2
DNA binding Domain	MH1	bHLH	bHLH
Residue position (amino acid) of domain	12-136	96-146	99-151
UniProt ID	Q15797	O96004	P61296

expression, a critical anti-inflammatory and immunomodulatory protein secreted by club cells in the airway epithelium (6).

The *SCGB1A1* gene, located on chromosome 11q12.3, encodes CC16, also known as Clara Cell Secretory Protein or uteroglobin (7). CC16 plays a protective role in the lungs by modulating inflammatory responses, neutralizing oxidative stress, and maintaining epithelial integrity (32). Additionally, epigenetic modifications, such as promoter hypermethylation of the *SCGB1A1* gene in COPD patients, contribute to reduced gene expression and compromise its anti-inflammatory function (8). Reduced CC16 levels have been consistently associated with impaired pulmonary function, increased epithelial permeability, and heightened inflammatory responses in COPD patients (33). In the present study, 64.73% of the participants reported a family history of COPD (maternal, paternal, or both), further supporting a genetic predisposition to the disease (Table 2). Notably, a Mendelian randomization study using *SCGB1A1*-linked SNPs demonstrated that genetically elevated CC16 levels causally reduce the risk of developing COPD and slow its progression (32). We observed, for the first time in an Indian population, a significant association between the *SCGB1A1* rs3741240 polymorphism and COPD risk. Among the genetic variants within the *SCGB1A1* locus, rs3741240, located in the 5' untranslated region (5' UTR), has been the most extensively studied (13). The rs3741240 SNP (G > A, also known as G38A) has been reported to correlate with lower CC16 expression levels at both the mRNA and protein levels, supporting a transcriptional and post-transcriptional impact (34). The present study revealed that COPD patients carrying the rs3741240 AA risk genotype had a significantly higher risk of developing COPD (Table 3). Interestingly, female patients with the risk genotype were at greater risk than male patients. This aligns with the findings of Han et al. (35), who reported that, for the first time in 2000, the number of women dying from COPD in the United States surpassed

the number of male deaths. Furthermore, Tam et al. (36) suggested that female sex hormones influence airway inflammation, mucus production, and other biological processes associated with COPD in women.

The higher COPD risk observed among older individuals with the AA genotype suggests that age-related physiological decline may compound the genetic susceptibility conferred by the *SCGB1A1* variant. This observation is consistent with that of a study by Stone et al. (37), who identified increasing age as a significant risk factor for COPD management. Moreover, the increased prevalence of COPD among individuals with the AA genotype residing in rural and semi-urban areas highlights the potential influence of environmental exposures. This finding supports the previous research by Raju et al. (38), who reported that urban-rural disparities in COPD prevalence may be attributed not only to smoking and indoor air pollution but also to occupational and agricultural exposure. Consistent with this, an Indian hospital-based study by Agarwal et al. (39) reported that 75.4% of rural female biomass fuel users had COPD compared to non-users. Together, these observations emphasize the interplay between genetic predisposition and environmental risk factors in shaping COPD susceptibility in the Indian population. Furthermore, the greater frequency of the AA genotype among patients in advanced GOLD stages (3 and 4) indicates that this variant may contribute not only to COPD susceptibility but also to disease progression and severity.

The association of the 38AA genotype with reduced lung function parameters underscores its potential role in compromising airway integrity, possibly through reduced CC16-mediated epithelial protection. In the control group, individuals with all three genotypes generally exhibited only mild-to-moderate COPD associated with the rs3741240 variant. In contrast, among COPD patients, those harboring the 38AA risk genotype predominantly developed severe disease. In contrast, individuals with heterozygous or major genotypes were more likely to exhibit mild-to-moderate COPD severity (Figure 2).

The localization of CC16 in bronchial epithelial cells (BECs) highlights its potential role in airway homeostasis and epithelial cell function (40). Giemsa staining revealed cytoplasmic granularity in a subset of cells, suggesting active secretion and a secretory phenotype (41). This observation was further supported by immunofluorescence analysis, which confirmed the cytoplasmic localization of CC16 in BECs with a distinct granular pattern,

TABLE 9 Docking and confidence scores for the interactions between wild-type and mutant *SCGB1A1* promoters and the three transcription factors.

Model	Wild type		Mutant	
	Docking score	Confidence score	Docking score	Confidence score
Promoter VS SMAD1	-260.19	0.9006	-257.23	0.8952
Promoter VS HAND1	-303.81	0.9559	-299.94	0.9525
Promoter VS HAND2	-325.42	0.9709	-316.56	0.9655



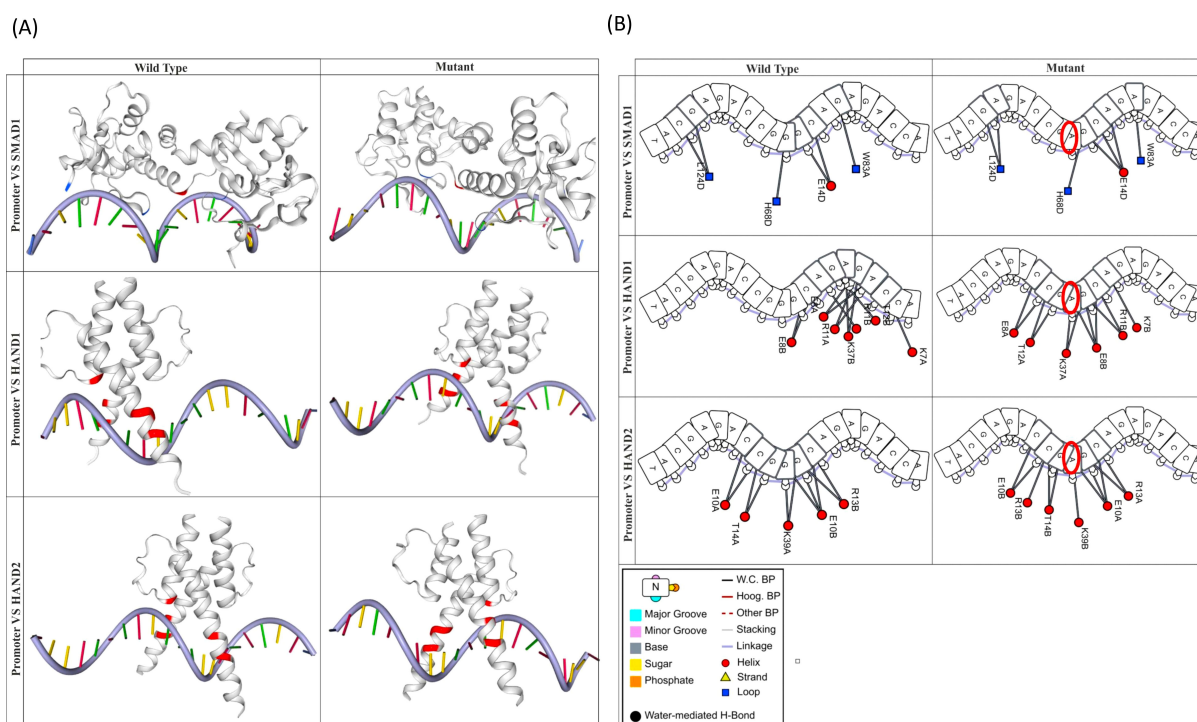


FIGURE 5

(A) Illustration of docking poses of 3 TF-promoter complexes. (B) Interacting interfaces of 3 TF-promoter complexes.

consistent with its known function as a secretory protein (42). The absence of nuclear staining emphasizes the cytoplasmic confinement of CC16 activity (42). Moreover, the downregulation of CC16 removes a critical inhibitory checkpoint in the immune response, facilitating chronic neutrophilic inflammation, oxidative injury, and structural damage to the alveolar walls (41). These findings align with previous reports identifying Clara cells as the primary source of CC16 and support its role as a biomarker of epithelial integrity and inflammation in airway diseases such as COPD (43).

We demonstrated that SCGB1A1 expression was downregulated in individuals carrying the 38GA or 38AA genotypes at rs3741240 compared to the control group with the 38GG genotype, as assessed by RT-PCR and western blot analyses. The presence of the rs3741240 risk allele may serve as a molecular biomarker to identify individuals with heightened susceptibility to COPD, even in non-smokers, suggesting a genotype-driven endotype of this disease. These findings are consistent with those of Li et al. (34), who reported that the A risk allele of rs3741240 was significantly associated with lower SCGB1A1 mRNA expression in both sputum and lung tissues, with regression slopes of  $\beta = -0.81$  ( $P = 0.007$ ) in cross-sectional and  $\beta = -0.69$  ( $P = 0.04$ ) in longitudinal cohorts. Moreover, another study showed a clear allelic imbalance, with excess expression of the G allele over A ( $P < 10^{-6}$ ), suggesting that the A allele leads to reduced transcriptional activity (44). These findings align with protein-level observations: individuals with A

allele genotypes (AG or AA) show lower CC16 levels in various biological fluids, as first reported by Kim et al. and others (13, 45). Kim et al. (13) observed that the minor allele at this SNP was negatively associated with SCGB1A1 mRNA expression levels in sputum samples. In an adult cystic fibrosis cohort, rs3741240 explained up to 19% of the variance in serum CC16; each A allele was associated with a roughly  $-0.63$  SD decrease ( $P \approx 6.2 \times 10^{-18}$ ) (46). This aligns with a broader neonatal/adult population study, in which rs3741240-A correlated with reduced CC16 from birth through early mid-adulthood (47). Therefore, it can be said that the rs3741240 polymorphism is associated with allele-specific downregulation of CC16 at both the transcriptional and translational levels in bronchial brush cells from the population of West Bengal, India. Notably, this also opens avenues for targeted therapeutic strategies, such as gene therapy to restore CC16 expression or recombinant CC16 protein supplementation, to mitigate inflammation and epithelial injury.

In addition to wet-lab experimentation, bioinformatics analyses were employed to reinforce the hypothesis that the rs3741240 SNP may regulate CC16 expression and play a role in COPD pathogenesis. This approach was based on the premise that numerous bioinformatics studies have successfully established associations between specific SNPs and the development of various diseases (48–50). SNPs can modulate gene expression by altering the binding affinity of TFs or miRNAs to target gene sequences (51–53). Therefore, SNP rs3741240 in the SCGB1A1

gene (which encodes CC16) may alter its expression by altering TF or miRNA binding, thereby contributing to disease development. Previous studies have demonstrated that SNPs can influence transcription factor (TF) binding, thereby altering gene expression (54, 55). In the present study, three transcription factors, SMAD1, HAND1, and HAND2, exhibited a slightly better binding affinity for the polymorphic *SCGB1A1* promoter than their wild type counterparts. Since this observation alone does not elucidate the mechanism underlying TF-mediated modulation of CC16 expression, we further extended our bioinformatics investigation to explore miRNA-mediated regulation of CC16 expression affected by the rs3741240 SNP. Accordingly, two miRNAs (hsa\_miR\_11181\_3p and hsa\_miR\_6804\_5p) were predicted to bind in proximity to the rs3741240 locus. Among them, hsa-miR-11181-3p appears to downregulate CC16 expression through AGO2-mediated silencing of a polymorphic variant of *SCGB1A1*. The role of miRNAs in AGO2-mediated gene silencing has been reported in several studies (56, 57). It should be noted that the present study is limited to computational analyses. Therefore, further wet-lab experiments are necessary to confirm TF- and miRNA-mediated silencing of CC16.

Reduced CC16 expression suppresses the epithelial-derived antimicrobial protein Short Palate, Lung, and Nasal Epithelial Clone 1 (SPLUNC1) through impairment of downstream signal transduction mediated by the CC16 receptor, Late Antigen-2 (VLA-2), on epithelial cells. Additionally, diminished CC16-VLA-4 signaling in leukocytes results in decreased leukocyte adhesion to endothelial cells, thereby promoting lung and airway inflammation (58, 59). Similarly, in the present study, the reduced CC16 expression associated with the rs3741240 SNP may contribute to COPD pathogenesis by compromising the respiratory host defense system.

However, the present study had a few limitations. The sample size for specific genotypic subgroups was limited, which may have affected the statistical robustness. Moreover, the study design was correlative and did not include functional assays to confirm the mechanistic role of the rs3741240 variant directly. Although previous Mendelian randomization studies using *SCGB1A1*-linked SNPs have demonstrated that genetically elevated CC16 levels may causally reduce COPD risk and progression, our study provides the first evidence of a significant association between the *SCGB1A1* rs3741240 polymorphism and COPD susceptibility in the Indian population. This causal relationship requires verification through functional validation in larger, multicenter cohorts.

In conclusion, the *SCGB1A1* rs3741240 variant downregulates CC16 expression, contributing to COPD pathogenesis, possibly by promoting unchecked inflammation and epithelial damage. The elucidation of this molecular axis provides valuable insights into disease mechanisms and highlights potential clinical applications. Targeting this pathway, either by restoring CC16 levels or modulating *SCGB1A1* activity, may offer promising avenues for

developing specific personalized medicine-based therapeutic approaches and novel treatment strategies for COPD management.

## Data availability statement

The data presented in the study are deposited in the ClinVar repository, accession number SCV007106175.

## Ethics statement

The studies involving humans were approved by Clinical Research Ethics Committee, Allergy and Asthma Research Centre, West Bengal, India (CREC-AARC Ref: 62/204). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

NS: Data curation, Formal Analysis, Funding acquisition, Methodology, Writing – original draft. HA: Data curation, Formal Analysis, Writing – review & editing. AG: Formal Analysis, Software, Writing – review & editing. AM: Formal Analysis, Software, Writing – review & editing. IG: Data curation, Formal Analysis, Writing – review & editing. HB: Formal Analysis, Software, Writing – review & editing. AI: Resources, Writing – review & editing. AD: Resources, Writing – review & editing. SM: Resources, Writing – review & editing. SP: Conceptualization, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

## Acknowledgments

The author acknowledges fellowship support from the University Grants Commission (Sanction No. 201610048901, dated 29 September 2021). The authors also convey thanks to the patients who voluntarily participated in our study. We are grateful to the lab professionals of the Allergy and Asthma Research Centre, Kolkata, for their technical help. We also thank Dr. Sumit Kumar Hira and Dr. Subha Shankar Mukherjee for their valuable technical support.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial

intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

1. Sultana N, Ganai I, Sultana S, Adhikari H, Laha A, Biswas H, et al. Investigation the role of ADRB2 rs1042713 and rs1042714 polymorphisms among COPD patients of West Bengal population, India. *Hum Gene.* (2025) 45:201446. doi: 10.1016/j.humgen.2025.201446
2. Mayo Clinic. *Chronic obstructive pulmonary disease (COPD)* (2020). Mayo Foundation for Medical Education and Research. Available online at: <https://www.mayoclinic.org/diseases-conditions/copd/symptoms-causes/syc-20353679> (Accessed July 10, 2025).
3. Li HY, Gao TY, Fang W, Xian-Yu CY, Deng NJ, Zhang C, et al. Global, regional, and national prevalence of, and risk factors for, chronic obstructive pulmonary disease: Estimates from the 1990 to 2019 Global Burden of Disease Study. *Respirology.* (2023) 28:29–36. doi: 10.1111/resp.14349
4. Silver SR, Alarcon WA, Li J. Incident chronic obstructive pulmonary disease associated with occupation, industry, and workplace exposures in the Health and Retirement Study. *Am J Ind Med.* (2021) 64:26–38. doi: 10.1002/ajim.23196
5. Adeloye D, Song P, Zhu Y, Campbell H, Sheikh A, Rudan I, et al. Global, regional, and national prevalence of, and risk factors for, chronic obstructive pulmonary disease (COPD) in 2019: a systematic review and modelling analysis. *Lancet Respir Med.* (2022) 10:447–58. doi: 10.1016/S2213-2600(21)00511-7
6. Singh G, Katyal SL. Clara cells and Clara cell 10 kD protein (CC10). *Am J Respir Cell Mol Biol.* (1997) 17:141–3. doi: 10.1165/ajrcmb.17.2.f138
7. Martinu T, Todd JL, Gelman AE, Guerra S, Palmer SM. Club Cell Secretory Protein in Lung Disease: Emerging Concepts and Potential Therapeutics. *Annu Rev Med.* (2023) 74:427–41. doi: 10.1146/annurev-med-042921-123443
8. Gribben KC, Poole JA, Nelson AJ, Farazi PA, Wichman CS, Heires AJ, et al. Relationships of serum CC16 levels with smoking status and lung function in COPD. *Respir Res.* (2022) 23:247. doi: 10.1186/s12931-022-02158-8
9. Lin J, Chen X, Chen Y, Zeng X, Wang F, Luo S, et al. Club cell secretory protein 16 promotes cell proliferation and inhibits inflammation and pyroptosis in response to particulate matter 2.5-induced epithelial damage in asthmatic mice. *J Thorac Dis.* (2025) 17:753–73. doi: 10.21037/jtd-24-1371
10. Zhai J, Insel M, Addison KJ, Stern DA, Pederson W, Dy A, et al. Club Cell Secretory Protein Deficiency Leads to Altered Lung Function. *Am J Respir Crit Care Med.* (2019) 199:302–12. doi: 10.1164/rccm.201807-1345OC
11. Chen M, Xu K, He Y, Jin J, Mao R, Gao L, et al. CC16 as an Inflammatory Biomarker in Induced Sputum Reflects Chronic Obstructive Pulmonary Disease (COPD) Severity. *Int J Chron Obstruct Pulmon Dis.* (2023) 18:705–17. doi: 10.2147/COPD.S400999
12. Lauchó-Contreras ME, Polverino F, Gupta K, Taylor KL, Kelly E, Pinto-Plata V, et al. Protective role for club cell secretory protein-16 (CC16) in the development of COPD. *Eur Respir J.* (2015) 45:1544–56. doi: 10.1183/09031936.00134214
13. Kim DK, Cho MH, Hersh CP, Lomas DA, Miller BE, Kong X, et al. Genome-wide association analysis of blood biomarkers in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* (2012) 186:1238–47. doi: 10.1164/rccm.201206-1013OC
14. Milne S, Li X, Cordero J, Yang C-X, Cho MH, Beaty TH, et al. Large-scale genetic association analysis identifies the SCGB1A1 rs3741240 variant as a major determinant of circulating CC16 levels in COPD patients. *PLoS Genet.* (2013) 9:e1003887. doi: 10.1371/journal.pgen.1003887
15. Guerra S, Stern DA, Zhou M. The SCGB1A1 rs3741240 variant and childhood lung function: longitudinal evidence from the Tucson Children's Respiratory Study. *Eur Respir J.* (2015) 46:69–77. doi: 10.1183/13993003.00422-2015
16. Robin M, Bourdin A, Devouassoux G. The SCGB1A1 rs3741240 polymorphism and COPD susceptibility: findings from the Lovelace Smokers Cohort. *Respir Res.* (2015) 16:132. doi: 10.1186/s12931-015-0288-2
17. Global Initiative for Chronic Obstructive Lung Disease (GOLD). *Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease (2025 Report)*. Available online at: <http://www.goldcopd.org> (Accessed June 12, 2025).
18. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta CT$  method. *Methods.* (2001) 25:402–408. doi: 10.1006/meth.2001.1262
19. Yuhao Chen and Xiaowei Wang. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* (2020) 48:D127–31. doi: 10.1093/nar/gkz757
20. Lohrer P, Rigoutsos I. Interactive exploration of RNA22 microRNA target predictions. *Bioinformatics.* (2012) 28:3322–3. doi: 10.1093/bioinformatics/bts615
21. Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell.* (2006) 126:1203–17. doi: 10.1016/j.cell.2006.07.031
22. Biesiada M, Purzycka KJ, Szachniuk M, Blazewicz J, Adamiak RW. Automated RNA 3D Structure Prediction with RNAComposer. *Methods Mol Biol.* (2016) 1490:199–215. doi: 10.1007/978-1-4939-6433-8\_13
23. Yan Y, Zhang D, Zhou P, Li B, Huang SY. HDock: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* (2017) 45:W365–73. doi: 10.1093/nar/gkx407
24. Zhang Q, Liu W, Zhang HM, Xie GY, Miao YR, Xia M, et al. hTFtarget: A Comprehensive Database for Regulations of Human Transcription Factors and Their Targets. *Genomics Proteomics Bioinf.* (2020) 18:120–8. doi: 10.1016/j.gpb.2019.09.006
25. Lee H, Park H, Ko J, Seok C. GalaxyGemini: a web server for protein homology structure prediction based on similarity. *Bioinformatics.* (2013) 29:1078–80. doi: 10.1093/bioinformatics/btt079
26. Mitra R, Cohen AS, Sagendorf JM, Berman HM, Rohs R. DNAProDB: an updated database for the automated and interactive analysis of protein-DNA complexes. *Nucleic Acids Res.* (2025) 53:D396–402. doi: 10.1093/nar/gkx970
27. Amoutzias GD, Robertson DL, Van de Peer Y, Oliver SG. Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem Sci.* (2008) 33:220–9. doi: 10.1016/j.tibs.2008.02.002
28. GBD. 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet.* (2016) 388:1459–544. doi: 10.1016/S0140-6736(16)31012-1
29. Deng X, Yang X, Gan Z, Huang H, Yang J. Identification of Five NK Cell-Related Hub Genes in COPD Using Single-Cell RNA Sequencing Analysis. *J Inflammation Res.* (2025) 18:2169–83. doi: 10.2147/JIR.S491298
30. Wang Z, Lin J, Liang L, Huang F, Yao X, Peng K, et al. Global, regional, and national burden of chronic obstructive pulmonary disease and its attributable risk factors from 1990 to 2021: an analysis for the Global Burden of Disease Study 2021. *Respir Res.* (2025) 26:2. doi: 10.1186/s12931-024-03051-2
31. Weeks JD, Elgaddal N. Chronic obstructive pulmonary disease in adults age 18 and older: United States, 2023. *NCHS Data Brief.* (2025) 529:1–9. doi: 10.15620/cdc/174596
32. Milne S, Li X, Hernandez Cordero AI, Yang CX, Cho MH, Beaty TH, et al. Protective effect of club cell secretory protein (CC-16) on COPD risk and progression: a Mendelian randomisation study. *Thorax.* (2020) 75:934–43. doi: 10.1136/thoraxjnl-2019-214487

33. Lam DCL, Kwok HH, Yu WC, Ko FW, Tam CY, Lau AC, et al. CC16 levels correlate with cigarette smoke exposure in bronchial epithelial cells and with lung function decline in smokers. *BMC Pulm Med.* (2018) 18:47. doi: 10.1186/s12890-018-0607-7
34. Li X, Guerra S, Ledford JG, Kraft M, Li H, Hastie AT, et al. Low CC16 mRNA Expression Levels in Bronchial Epithelial Cells Are Associated with Asthma Severity. *Am J Respir Crit Care Med.* (2023) 207:438–51. doi: 10.1164/rccm.202206-1230OC
35. Han MK, Postma D, Mannino DM, Giardino ND, Buist S, Curtis JL, et al. Gender and chronic obstructive pulmonary disease: why it matters. *Am J Respir Crit Care Med.* (2007) 176:1179–84. doi: 10.1164/rccm.200704-553CC
36. Tam A, Morrish D, Wadsworth S, Dorscheid D, Man SF, Sin DD. The role of female hormones on lung function in chronic lung diseases. *BMC Womens Health.* (2011) 11:24. doi: 10.1186/1472-6874-11-24
37. Stone RA, Lowe D, Potter JM, Buckingham RJ, Roberts CM, Pursey NJ. Managing patients with COPD exacerbation: does age matter? *Age Ageing.* (2012) 41:461–8. doi: 10.1093/ageing/afs039
38. Raju S, Keet CA, Paulin LM, Matsui EC, Peng RD, Hansel NN, et al. Rural Residence and Poverty Are Independent Risk Factors for Chronic Obstructive Pulmonary Disease in the United States. *Am J Respir Crit Care Med.* (2019) 199:961–9. doi: 10.1164/rccm.201807-1374OC
39. Agrawal A, Garg U, Madan C, Goel S. Association of biomass fuel smoke exposure and COPD in women of rural India: a cross-sectional hospital-based study. *Natl J Integr Res Med.* (2013) 4:11–7. doi: 10.70284/njirm.v4i6.2241
40. Broeckaert F, Bernard A. Clara cell secretory protein (CC16): characteristics and perspectives as lung peripheral biomarker. *Clin Exp Allergy.* (2000) 30:469–75. doi: 10.1046/j.1365-2222.2000.00760.x
41. Hermans C, Bernard A. Lung epithelium-specific proteins: characteristics and potential applications as markers. *Am J Respir Crit Care Med.* (1999) 159:646–78. doi: 10.1164/ajrccm.159.2.9806064
42. Broeckaert F, Clippe A, Knoop B, Hermans C, Bernard A. Clara cell secretory protein (CC16): features as a peripheral lung biomarker. *Ann N Y Acad Sci.* (2000) 923:68–77. doi: 10.1111/j.1749-6632.2000.tb05520.x
43. Shijubo N, Itoh Y, Yamaguchi T, Sugaya F, Hirasawa M, Yamada T, et al. Serum levels of Clara cell 10-kDa protein are decreased in patients with asthma. *Lung.* (1999) 177:45–52. doi: 10.1007/pl00007626
44. Li XX, Peng T, Gao J, Feng JG, Wu DD, Yang T, et al. Allele-specific expression identified rs2509956 as a novel long-distance cis-regulatory SNP for *SCGB1A1*, an important gene for multiple pulmonary diseases. *Am J Physiol Lung Cell Mol Physiol.* (2019) 317:L456–63. doi: 10.1152/ajplung.00275.2018
45. Zhu L, An L, Ran D, Lizarraga R, Bondy C, Zhou X, et al. The Club Cell Marker *SCGB1A1* Downstream of *FOXA2* is Reduced in Asthma. *Am J Respir Cell Mol Biol.* (2019) 60:695–704. doi: 10.1165/rcmb.2018-0199OC
46. Zhai J, Emond MJ, Spangenberg A, Stern DA, Vasquez MM, Blue EE, et al. Club cell secretory protein and lung function in children with cystic fibrosis. *J Cyst Fibros.* (2022) 21:811–20. doi: 10.1016/j.jcf.2022.03.007
47. Voraphani N, Stern DA, Ledford JG, Spangenberg AL, Zhai J, Wright AL, et al. Circulating CC16 and Asthma: A Population-based, Multicohort Study from Early Childhood through Adult Life. *Am J Respir Crit Care Med.* (2023) 208:758–69. doi: 10.1164/rccm.202301-0041OC
48. Mondal A, Paul D, Dastidar SG, Saha T, Goswami AM. In silico analyses of Wnt1 nsSNPs reveal structurally destabilizing variants, altered interactions with Frizzled receptors and its deregulation in tumorigenesis. *Sci Rep.* (2022) 12:14934. doi: 10.1038/s41598-022-19299-x
49. Deng H, Li J, Shah AA, Ge L, Ouyang W. Comprehensive in-silico analysis of deleterious SNPs in APOC2 and APOA5 and their differential expression in cancer and cardiovascular diseases conditions. *Genomics.* (2023) 115:110567. doi: 10.1016/j.ygeno.2023.110567
50. Mondal A, Goswami AM, Saha T. In silico prediction of the functional consequences of nsSNPs in human beta-catenin gene. *Gene Rep.* (2021) 23:101066. doi: 10.1016/j.genrep.2021.101066
51. Mohamed FA, Freude K. Implications of SNP-triggered miRNA dysregulation in Schizophrenia development. *Front Genet.* (2024) 15:1321232. doi: 10.3389/fgene.2024.1321232
52. Ünal P, Lu Y, Bueno-de-Mesquita B, van Eijck CHJ, Talar-Wojnarowska R, Szentesi A, et al. Polymorphisms in transcription factor binding sites and enhancer regions and pancreatic ductal adenocarcinoma risk. *Hum Genomics.* (2024) 18:12. doi: 10.1186/s40246-024-00576-x
53. Antontseva EV, Degtyareva AO, Korbolina EE, Damarov IS, Merkulova TI. Human-genome single nucleotide polymorphisms affecting transcription factor binding and their role in pathogenesis. *Vavilovskii Zhurnal Genet Selektii.* (2023) 27:662–75. doi: 10.18699/VJGB-23-77
54. Leseva M, Santostefano KE, Rosenbluth AL, Hamazaki T, Terada N. E2f6-mediated repression of the meiotic *Stag3* and *Smc1 $\beta$*  genes during early embryonic development requires *Ezh2* and not the *de novo* methyltransferase *Dnmt3b*. *Epigenetics.* (2013) 8:873–84. doi: 10.4161/epi.25522
55. Leung JY, Nevins JR. E2F6 associates with BRG1 in transcriptional regulation. *PloS One.* (2012) 7:e47967. doi: 10.1371/journal.pone.0047967
56. Fadaka AO, Pretorius A, Klein A. MicroRNA Assisted Gene Regulation in Colorectal Cancer. *Int J Mol Sci.* (2019) 20:4899. doi: 10.3390/ijms20194899
57. Behera JK, Bhattacharya M, Mishra P, Mishra A, Dash AA, Kar NB, et al. Regulatory role of miRNAs in Wnt signaling pathway linked with cardiovascular diseases. *Curr Res Pharmacol Drug Discov.* (2022) 3:100133. doi: 10.1016/j.crphar.2022.100133
58. Iannuzo N, Welfley H, Li NC, Johnson MDL, Rojas-Quintero J, Polverino F, et al. CC16 drives VLA-2-dependent SPLUNC1 expression. *Front Immunol.* (2023) 14:1277582. doi: 10.3389/fimmu.2023.1277582
59. Johnson MDL, Younis US, Menghani SV, Addison KJ, Whalen M, Pilon AL, et al. CC16 Binding to  $\alpha_4\beta_1$  Integrin Protects against *Mycoplasma pneumoniae* Infection. *Am J Respir Crit Care Med.* (2021) 203:1410–8. doi: 10.1164/rccm.202006-2576OC