Respiratory Research

RESEARCH Open Access

Two novel genetic variants in the *WFDC2* gene from patients with bronchiectasis

Jeong-Min Kim[1†], Soojin Hwang[2†], Hye-Won Cho[1], Youngjun Kim[1], Dong Mun Shin[1], Eun Lee[3], Myungshin Kim[4], Cheonghwa Lee[5], Jong-Won Kim[5], Hyun-Young Park[6], Beom Hee Lee[2*] and Mi-Hyun Park[1*]

Abstract

Background Bronchiectasis is a chronic respiratory condition characterized by irreversible dilation and damage of the bronchial walls, leading to impaired mucociliary clearance and recurrent infections. Its etiology is diverse; however, genetic factors are critical in its congenital and severe forms. Therefore, we aimed to identify two novel variants of the *WFDC2* gene, known as antiprotease, from patients with bronchiectasis and/or related phenotypes using trio-based whole-genome sequencing analysis.

Methods Patients with bronchiectasis were recruited as trio or quad, and their genomic DNA was isolated. The whole genome sequence was produced and analyzed to find causative genetic variants through an internal pipeline using GATK-DRAGEN-Hail. Variant interpretation and pathogenicity assessment using various in-silico tools were performed to identify causative variants. Clinical characteristics were collected from the patients with identified variants.

Results In this discovery study involving four patients from three families, two novel variants in the *WFDC2* gene were identified and suggested as causative pathogenic variants for bronchiectasis. The first variant (c.291 C > G, p.(Cys97Trp)) is a homozygous variant that was not found in the population genome data. However, the second variant (c.278G > C, p.(Cys93Ser)) was identified in another patient as a heterozygous variant, forming a compound heterozygous state with the first variant. Notably, both variants, located at cysteine residues that are conserved across many species, are crucial in forming disulfide bonds essential for protein structure and function. In-silico analyses classified both variants as pathogenic; they were also identified as likely pathogenic according to the American College of Medical Genetics and Genomic guidelines. Furthermore, in an expansion study, the homozygous variant was also found in two unrelated patients.

Conclusion We identified two novel bi-allelic variants located at cysteine residues in the *WFDC2* gene from patients with bronchiectasis who had previously not received a genetic diagnosis. Therefore, considering prior research on the pivotal role of the WFDC2 protein in the respiratory system, these two novel variants may serve as potential diagnostic markers and therapeutic targets for bronchiectasis.

[†]Jeong-Min Kim and Soojin Hwang contributed equally to this work.

*Correspondence:
Beom Hee Lee
bhlee@amc.seoul.kr
Mi-Hyun Park
mihyun4868@korea.kr

Full list of author information is available at the end of the article

Clinical trial number Not Applicable

Keywords Bronchiectasis, WFDC2, Whole genome sequencing, Genetic variant, Rare disease

Introduction

Bronchiectasis is a chronic respiratory disease characterized by the permanent dilation and damage of the bronchial walls, which leads to impaired mucus clearance and recurrent infections [1–3]. This condition can be due to various factors, including respiratory infection, mucociliary disorder, inflammation, immune deficiency, and idiopathic factors. Genetic factors, particularly cystic fibrosis and primary ciliary dyskinesia, and anatomical abnormalities, such as intraluminal obstructions and congenital cartilage defects, are also contributory factors [4, 5]. However, most bronchiectasis cases are still classified as idiopathic, underscoring the need for future research in this field [6, 7].

Airway inflammation is a critical factor in bronchiectasis [8]. Inflammatory markers are associated with the severity of the disease, clinical outcomes, and treatment targets [9]. Neutrophilic inflammation is a hallmark of the disease, with increased levels of damaging proteases like neutrophil elastase being associated with more severe bronchiectasis [10, 11]. Therefore, new inhibitors for neutrophil protease are proposed as regulators of neutrophilic inflammation in patients with bronchiectasis [12]. Moreover, individuals with severe bronchiectasis demonstrate increased levels of proinflammatory mediators and reduced levels of the anti-inflammatory secretory leukocyte protease inhibitor (SLPI). These observations are consistent with recent research indicating that intravenous antibiotic treatment results in decreased neutrophil elastase and increased SLPI and other antiproteases. This supports a conceptual framework in which severe bronchiectasis correlates with an imbalance of proteases and proinflammatory mediators, leading to the suppression of epithelial anti-inflammatory, antiprotease, and antimicrobial responses [13, 14].

WFDC2 belongs to the whey-acidic-protein (WAP) four-disulfide core (WFDC) family of proteins. This protein family is characterized by a conserved WFDC domain comprising 50 amino acids with eight conserved cysteine residues. WFDC2 comprises 124 amino acids organized into two WFDC domains [15, 16]. WFDC2 was initially identified as human epididymis protein-4; however, it is implicated in sperm maturation. WFDC2 is also upregulated in various cancers, including ovarian carcinoma, and has been suggested as a biomarker for these cancers [17–21]. Additionally, WFDC2 protein is expressed in the oral cavity, respiratory tract, female genital tract, and distal renal tubules [22, 23]. Previous studies have identified WFDC2 as a potential cross-class antiprotease, which can provide defense against microbial virulence with protease activity [24, 25]. In addition, due to its resemblance to SLPI and Elafin, both of which have the WFDC domain, WFDC2 is proposed to contribute to the innate immune defenses within the respiratory system [26]. Experiments with mice showed that the deletion of the *WFDC2* gene led to lung mucosal barrier failure and alveolar cell apoptosis, resulting in severe dyspnea and respiratory failure, ultimately causing death [27, 28]. However, there are no reports on the role of the *WFDC2* gene in bronchiectasis development. Recently, it was reported that recessively inherited deficiency of secreted WFDC2 causes nasal polyposis and bronchiectasis [29].

Therefore, in this study, we identified two novel variants of the *WFDC2* gene, known as antiprotease, from patients with bronchiectasis and/or related phenotypes using trio-based whole-genome sequencing (WGS) analysis. These two variants are suggested to be pathogenic, playing critical roles in protein structure and function through disulfide bonds with other cysteines. Thus, the two novel variants in the *WFDC2* gene identified in this study can be proposed as diagnostic variants in bronchiectasis development.

Methods

Study design and participants

This retrospective study included six patients with bronchiectasis as the main phenotype, who remained genetically undiagnosed. Four patients from three families were recruited for the discovery study, whereas two patients from two families participated in the expansion study. These patients came from singleton proband, duo, trio, and quad families as part of a pilot study of a national project named the 'National Project of Bio Big Data pilot study' [30]. All participants provided written informed consent. Participant enrollment and phenotypic analysis were conducted at each hospital following the project guidelines. Basic information, including age, sex, and ethnicity, as well as clinical data associated with Human Phenotype Ontology (HPO) terms (https://hpo.jax.org/app/), such as symptoms, laboratory findings, computed tomography (CT) scans, bronchoscopy images, transmission electron microscopy (TEM), pulmonary function tests, medical history, and family history, were collected. This study was reviewed and approved by the Institutional Review Board of the Korea National Institute of Health (Approval Number: 2022-02-07-P-A, 2022-09-10-P-A, KDCA-2023-06-06-P-01).

WGS data production and analysis

The generation and analysis of WGS data have been described in a previous report [31]. Briefly, WGS data was

generated in the National Project of Bio Big Data pilot study (https://www.cirn.re.kr). Genomic DNA was isolated from peripheral whole blood and sequenced using the Illumina NovaSeq 6000 platform, followed by alignment to the human reference genome (GRCh38 version). Variant calling was performed using the Genome Analysis Toolkit (GATK, version 4.2.6.1, Broad Institute, MA, USA). Notably, all the processes have been described previously (https://www.kobic.re.kr/kobic/res/ngp).

Variant discovery

Joint genotype calling was performed using the DRAGEN pipeline (Illumina, version 3.8). Therefore, to estimate variant call accuracy, Variant Quality Score Recalibration (VQSR) was applied using GATK (version 4.2.6.1). The process of variant discovery has previously been reported [32]. Furthermore, to identify genetic variants involved in the disease phenotypes, we performed trio-based WGS analysis for trio- and quad-families using the Hail system (version 0.2, https://hail.is) [33]. Variants found in multi-allelic sites, low-complexity regions, and those not passing the VQSR filter were excluded. Genotypes with variant calling depths (DP) < 10 or > 1000 were removed. Variants were selected based on allelic balance (AB) criteria: $0.3 \leq AB \leq 0.7$ for single-nucleotide variant (SNV) heterozygotes, $0.2 \leq AB \leq 0.8$ for Indel heterozygotes, and $AB \geq 0.95$ for homozygotes. Variants were excluded if their call rates were under 0.1 or if the *P*-value of the Hardy-Weinberg equilibrium test was $< 10^{-12}$. Additionally, any calls on the Y chromosome in the female samples and heterozygous calls in non-pseudoautosomal regions in the male samples were removed. The DRAGEN-Hail pipeline identified de novo, compound heterozygous, and homozygous SNV/Indels. The Variant Effect Predictor was used for gene and consequence annotation of variants [34]. Furthermore, to ensure the high quality of the de novo variant, variants with GQ ≤ 25 in the proband and those that were observed in more than 0.1% of the non-neuro subset of gnomAD (GRCh38 v3.1.2) were filtered out. Variants were also excluded if the proband AB was < 0.3, parent AB was > 0.1, or DP ratio (proband read depth / parental read depth) was < 0.3. For inherited heterozygous variants, only those with an internal allele count (AC) < 2 were considered (proband: 0/1, mom: 0/1, dad: 0/0; proband: 0/1, mom: 0/0, dad: 0/1). Furthermore, insertions and deletions of ≤ 50 base pairs (bp) were categorized as Indels. Discovered variants were visualized using Integrative Genomics Viewer (IGV), and true positive variants were processed for variant interpretation and prioritization [35].

Variant interpretation and in silico analysis

Variants identified within coding regions and splicing sequences were prioritized as potential candidate genes based on their pathogenic potential, as assessed by VarSome (Version 11.8, Aug. 2023, http://varsome.com/), according to the American College of Medical Genetics and Genomic (ACMG) variant classification guidelines [36, 37]. The impact of missense variants on protein structure was evaluated using the DynaMut2 protein prediction program (https://biosig.lab.uq.edu.au/dynamut2/), which utilizes three-dimensional structural data from AlphaFold (WFDC2 Human, AF-Q14508-F1) available through Uniprot [38, 39]. Additionally, the pathogenicity of the missense variant was confirmed using the AlphaMissense web resource (https://alphamissense.hegelab.org) [40]. To evaluate the splicing effects of the genetic variants, SpliceAI analysis was performed (https://spliceailookup.broadinstitute.org). The allele frequency and count of discovered variants were more confirmed using gnomAD (version 4.1), 'Korean Variant Archive for a reference database of genetic variations in the Korean population' (KOVA v2, https://www.kobic.re.kr/kova/), and 'Regeneron Genetics Center (RGC) Million Exome Variant Browser' [41, 42]. The amino acid sequences were gathered from the NCBI protein database (NP_006094, XP_005569203, XP_008949604, NP_001069958, NP_001361584, https://ncbi.nlm.nih.gov/protein) and used for multiple sequence alignment by CLUSTALW (https://www.genome.jp/tools-bin/clustalw). Therefore, to establish genotype-phenotype correlations, we collected data from several databases, including OMIM (https://www.omim.org/), ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), IMPC (https://www.mousephenotype.org/), and PubMed (https://pubmed.ncbi.nlm.nih.gov), and reviewed relevant literature sources.

Results

The discovery of causative variants in patients with bronchiectasis through genome sequencing

Overall, six patients with bronchiectasis from five families were selected from the 'National Project of Bio Big Data pilot study' (Fig. 1). The main phenotypes of the patients in this study were bronchiectasis (HPO term, HP:0002110) and recurrent respiratory infections (HP:0002205). The discovery study included quad and trio analyses of three families. Family 1 (F1) was recruited as a quad, comprising two affected brothers (patients 1 and 2; Pt1 and Pt2) and their healthy parents. Families 2 and 3 (F2 and F3) underwent trio analyses, including the proband (patients 3 and 4; Pt3 and Pt4) and their parents. Furthermore, to expand the study, families 4 and 5 underwent proband-only and duo analyses, respectively, with patients 5 (Pt5) and 6 (Pt6).

Two rare variants in the *WFDC2* gene were identified as pathogenic variants through WGS analysis using an internal pipeline. Patients 1, 2, 3, 5, and 6 had a homozygous variant, chr20:45480009:C: G, a missense

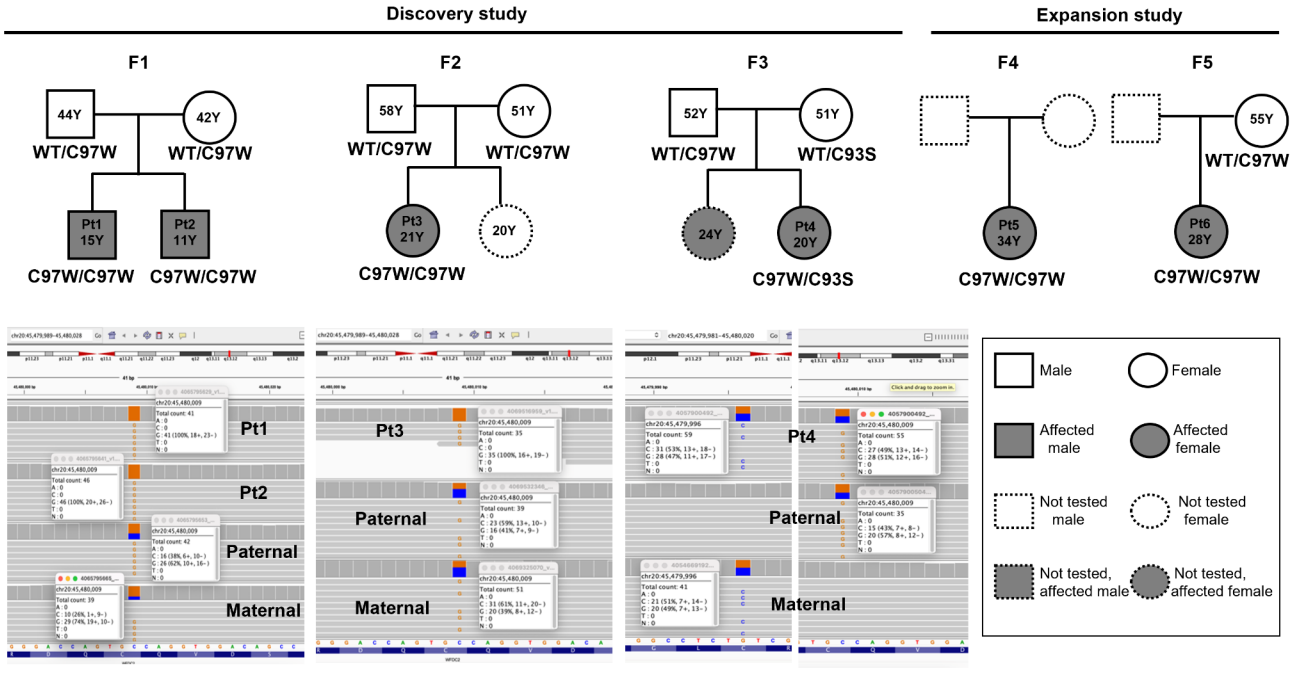

Fig. 1 WFDC2 variants in patients with bronchiectasis were identified through whole-genome analysis. The family pedigree of six patients and the IGV analysis diagram are depicted

variant resulting in the change of cysteine to tryptophan at the 97th amino acid (c.291 C > G, p.(Cys97Trp)) in the WFDC2 protein. However, the other variant, chr20:45479996:G: C, a heterozygous variant, was discovered in patient 4 as one of the compound heterozygous variants. It changes cysteine to serine at the 93rd amino acid (c.278G > C, p.(Cys93Ser)). SpliceAI analysis of this variant did not yield significant results for splicing loss (acceptor/donor delta score = 0.00) or splicing gain (acceptor gain = 0.01, donor gain = 0.00) (https://spliceailookup.broadinstitute.org/). In the discovery study, we also validated these two variants in each patient and their parents through IGV analysis (Fig. 1). These two novel variants in the *WFDC2* gene are classified as 'variant uncertain significance (VUS)' and 'likely pathogenic (LP)' based on the ACMG criteria and alpha missense, respectively (Table 1). Moreover, the classification of these variants was upgraded to LP by adjusting the updated criteria, PP1 and PM3. Additionally, the pathogenicity of those variants was confirmed as 'pathogenic strong' using the Varsome software and individual prediction software, such as CADD, SIFT, and PolyPhen.

Clinical characteristics of patients with bronchiectasis having two novel variants in the *WFDC2* gene

Table 2; Fig. 2 describe the clinical characteristics of patients with these novel variants. Patients 1 and 2 were teenage brothers affected by the homozygous variant (c.291 C > G). They had upper and lower respiratory tract symptoms, including sinusitis, nasal polyps, and bronchiectasis, starting from the neonatal to juvenile period. Patient 3, a 21-year-old female with the same variant, exhibited similar symptoms since infancy. However, Patient 4, a 20-year-old female with an older sister who had similar symptoms (not yet tested), had the compound heterozygous variants (c.278G > C and c.291 C > G) and an additional symptom of dermographism.

Specifically, patient 5, a 34-year-old female with the homozygous variant (c.291 C > G), experienced respiratory symptoms since adolescence and had fertility issues, ultimately giving birth through in vitro fertilization (IVF). Patient 6, a 28-year-old female with the homozygous variant (c.291 C > G), had recurrent respiratory infections and bronchiectasis, with a significant decrease in forced expiratory volume 1 (FEV1) of pulmonary function tests. Notably, Patients 3, 4, 5, and 6, who were adults, tended to have a significant decline in FEV1 and more severe radiologic findings according to the chest CT, with ring shadows and tram-track opacities indicating bronchial dilatation (Fig. 2j, k, l, o, p, r, s). This may be attributed to the longer duration of illness and recurrent infections. Furthermore, the cilia microtubule structures in TEM for patients 1, 3, and 4 were relatively preserved, with or without minimal dynein changes (Fig. 2c, i, n). However, no malignancies were reported in all the patients.

Functional prediction analysis of the novel variants in the *WFDC2* gene

Notably, the two variants are located at cysteine residues in the WAP domain 2, which are pivotal in the

Table 1 Pathogenicity of two novel variants in the *WFDC2* gene from patients with bronchiectasis

Gene	cDNA	Protein	CADD	SIFT	PolyPhen	REVEL	MetaLR	MetaRNN	Classification			Pathogenicity Varsome		
			(PhredScore)	Prediction (Score)	Prediction (Score)	Prediction (Score)	Prediction (Score)	Prediction (Score)	ACMG (Criteria)	Adjusted ACMG (Criteria)	Alpha Missense	In-silico Prediction	Meta scores	Indivi. scores
WFDC2	c. 291 C > G	Cys93Ser	23	Pathogenic Supporting (0)	Possibly Damaging (0.87)	Pathogenic Supporting (0.700)	Uncertain (0.712)	Pathogenic Strong (0.977)	VUS (PM2, PP3)	LP (PM2, PP3, PM3)	LP	PP3_ Pathogenic Strong	9, 2, 0	8, 11, 4
	c. 278G > C	Cys97Trp	22.1	Pathogenic Supporting (0)	Probably Damaging (1)	Pathogenic Moderate (0.802)	Pathogenic Moderate (0.983)	Pathogenic Strong (0.986)	VUS (PM2, PP3)	LP (PM2, PP3, PP1)	LP	PP3_ Pathogenic Strong	14, 0, 0	15, 7, 8

WFDC2 protein (Fig. 3a and b). The cysteine residue at the 93rd amino acid is linked to cysteine at the 114th amino acid, and that at the 97th is linked to cysteine at the 109th amino acid through disulfide bonds. Those two cysteine residues are conserved in many species, including humans, monkeys, cows, and mice. The other fourteen cysteine residues are also conserved in the WFDC2 protein. The score of destabilizing in protein stability, induced by the cysteine change in other amino acids at the 93rd and 97th residues, are −0.03 and −0.7 (kcal/mol), respectively (Fig. 3c). The allele count (AC) and frequency (AF) of the main variant, c.291 C > G (p.C97W), are 26 and 1.6E-05 at gnomAD 4.1_All and 26 and 5.8E-04 at gnomAD 4.1_EAS. However, the numbers of the other variant, c.278G > C (p.(Cys93Ser)), are 5 and 3.1E-06 at gnomAD_All and 5 and 1.1E-04 at gnomAD_EAS (Table S1 in the Supporting Information). In the Korean Variant Archive, KOVA, those numbers are 19 (AC) and 1.8E-03 (AF) for the c.291 C > G variant and 4 and 3.8E-04 for the c.278G > C variant. Furthermore, in the RGC million exome data, the first variant, c.291 C > G, is shown in only one individual with a frequency of 6.08E-07 in all populations and that of 3.33E-05 in the East Asian population (Table S2 in the Supporting Information).

Discussion

In this study, we identified two novel variants, c.291 C > G (p.(Cys97Trp)) and c.278G > C (p.(Cys93Ser)), in the *WFDC2* gene from patients with bronchiectasis. These variants alter conserved cysteine residues within the WAP domain, a critical region for WFDC2 function, and are predicted to be pathogenic based on in silico analyses. The discovery of these variants in multiple unrelated patients with a shared clinical phenotype strengthens the evidence for their role in bronchiectasis. Notably, very recent independent studies, although only archived, have identified the same c.291 C > G (p.(Cys97Trp)) variant in East Asian patients with bronchiectasis, further supporting its pathogenic significance.

Bronchiectasis is a chronic lung disease caused by persistent airway inflammation and impaired mucociliary clearance. While it is often associated with genetic conditions such as primary ciliary dyskinesia, cystic fibrosis, and immune deficiencies, a subset of cases remains unexplained at the genetic level. Our findings add to growing evidence that *WFDC2* may be a novel genetic contributor to bronchiectasis. Recently, one study identified *WFDC2* mutations, distinct from the variants observed in our study, as a novel genetic cause of chronic respiratory diseases, including bronchiectasis [29]. Moreover, during our manuscript's submission and revision process, we found two studies that are archived but have not yet undergone peer review. These studies independently identified the same *WFDC2* genetic variant (c.291 C > G,

Table 2 Clinical characteristics of patients with genetic variants in the *WFDC2* gene

	Discovery study				Expansion study	
Family	**F1**		**F2**	**F3**	**F4**	**F5**
Patient	**Pt1**	**Pt2**	**Pt3**	**Pt4**	**Pt5**	**Pt6**
Age at onset	Neonatal	Juvenile	Infantile	Juvenile	Juvenile	Juvenile
Age at enrollment (years)	15	11	21	20	34	28
Gender	M	M	F	F	F	F
Height (cm)	181.1	156.7	159.8	156	156.3	157
Weight (kg)	87.3	45.7	45.6	44.3	50.1	39
Type of WGS analysis	Quad		Trio	Trio	Proband only	Duo
Mutations	***WFDC2*** c.291 C > G (p.C97W) Homozygous		***WFDC2*** c.291 C > G (p.C97W) Homozygous	***WFDC2*** c.278G > C (p.C93S) Maternal, c.291 C > G (p.C97W) Paternal Compound Heterozygous	***WFDC2*** c.291 C > G (p.C97W) Homozygous	***WFDC2*** c.291 C > G (p.C97W) Homozygous
HPO	HP:0002110 (Bronchiectasis)	HP:0002110 (Bronchiectasis) HP:0002205 (Recurrent Respiratory Infections)	HP:0002110 (Bronchiectasis)	HP:0002110 (Bronchiectasis)	HP:0002110 (Bronchiectasis) HP:0000246 Sinusitis HP:0012265 Ciliary Dyskinesia HP:0000821 Hypothyroidism	HP:0002110 (Bronchiectasis) HP:0002721 (Immunodeficiency)
Clinical symptoms						
Sinusitis	Yes	Yes	Yes	Yes	Yes	Yes
Nasal polyps	Yes	Yes	Yes	Yes	Yes	Yes
Chronic cough	Yes	Yes	Yes	Yes	Yes	Yes
Bronchiectasis	Yes	Yes	Yes	Yes	Yes	Yes
Recurrent respiratory infection	Yes	Yes	Yes	Yes	Yes	Yes
Infertility	N/A	N/A	N/A	N/A	No*	N/A
Malignancy	No	No	No	No	No	No
Others				Dermographism		
Pulmonary function test						
FEV1 (measured %)	106	66	58	66	29	48
FVC (measured %)	125	88	75	78	40	61
FEV1/FVC (%)	78	69	76	83	66	70

N/A, not available; No*, IVF for pregnancy

p.(Cys97Trp)) in Korean and Japanese patients with bronchiectasis and suggested it as a causal genetic factor in the development of bronchiectasis [43, 44]. Patients with bronchiectasis exhibit additional clinical characteristics, such as sinusitis, nasal polyps, and low FEV1, which are associated with bronchiectasis (Table S3 in the Supporting Information). These findings further support the pathogenic role of *WFDC2* variants in patients with bronchiectasis. Therefore, our study identified two novel *WFDC2* variants in patients with bronchiectasis and recurrent respiratory infections, proving that *WFDC2* mutations represent a new genetic cause of bronchiectasis.

WFDC2 gene encodes a protease inhibitor involved in mucosal immunity, lung development, and epithelial protection. Deficiency or dysfunction of this protein could disrupt airway homeostasis, making individuals more susceptible to recurrent infections and progressive lung damage. Previous studies have suggested a link between WFDC2 and respiratory diseases, but direct associations with bronchiectasis have been limited [25, 27, 28, 45]. A recent study identified *WFDC2* mutations (p.(Cys49Arg)) in patients with chronic respiratory diseases, including bronchiectasis, and proposed that loss of WFDC2 function impairs airway defense mechanisms [29]. Additionally, WFDC2 has been implicated in nasal polyposis and infertility, both of which were observed in some of our patients, suggesting a broader spectrum of clinical manifestations. The fact that one of our patients exhibited dermographism also raises the possibility that *WFDC2* mutations may have previously unrecognized systemic effects. Interestingly, WFDC2

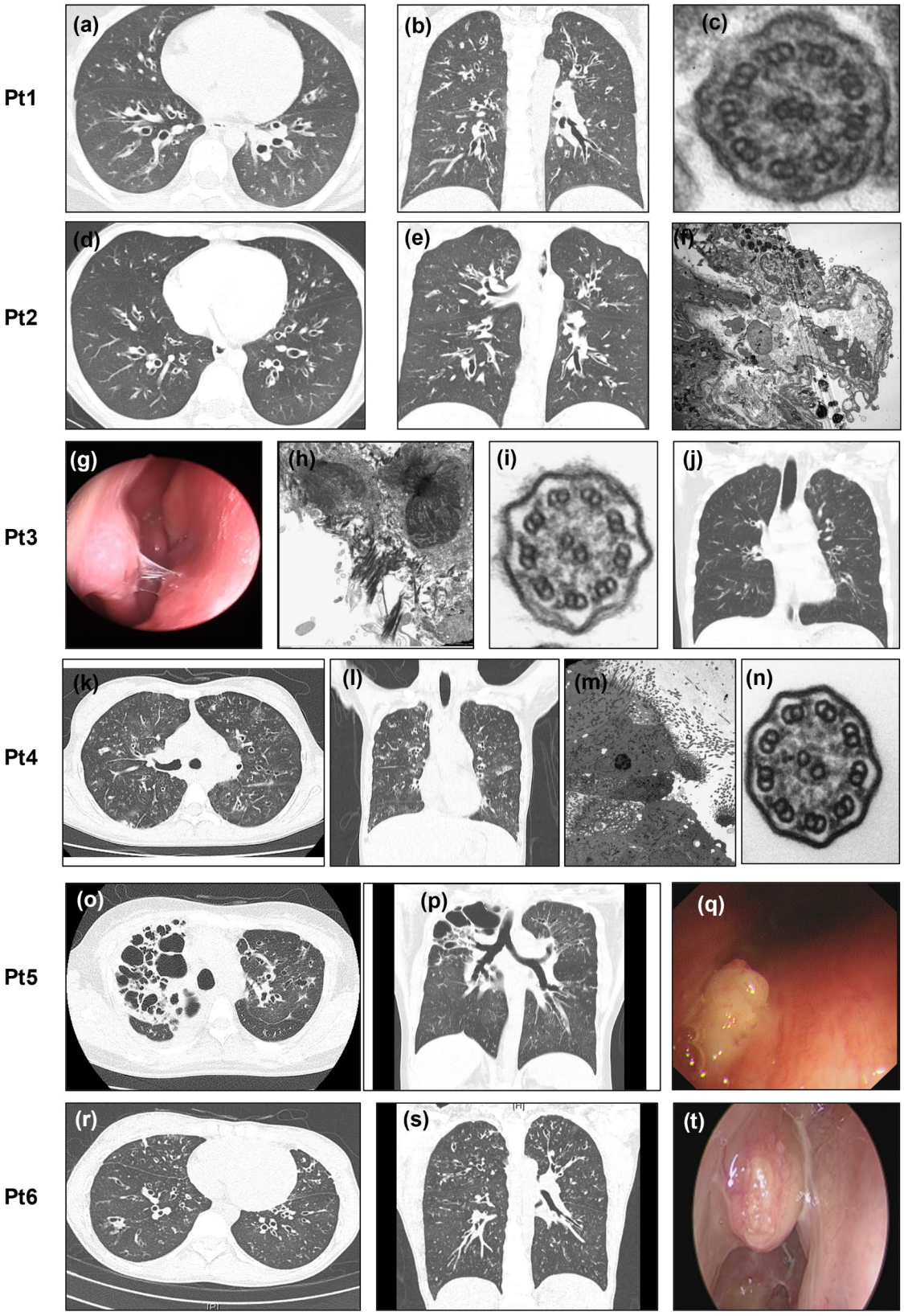

Fig. 2 Clinical manifestations of patients with bronchiectasis associated with WFDC2 mutations. Patient 1: (**a**, **b**) lung CT scans, (**c**) TEM image of microtubules in the bronchial cilia; Patient 2: (**d**, **e**) lung CT scans, (**f**) TEM image of damaged nasal cilia; Patient 3: (**g**) nasal polyp, (**h**, **i**) TEM images of the nasal cilia and microtubules, (**j**) lung CT scan; Patient 4: (**k**, **l**) lung CT scans, (**m**, **n**) TEM image of bronchial cilia and microtubules; Patient 5: (**o**, **p**) lung CT scans, (**q**) bronchial polyp; Patient 6: (**r**, **s**) lung CT scans, (**t**) bronchial polyp

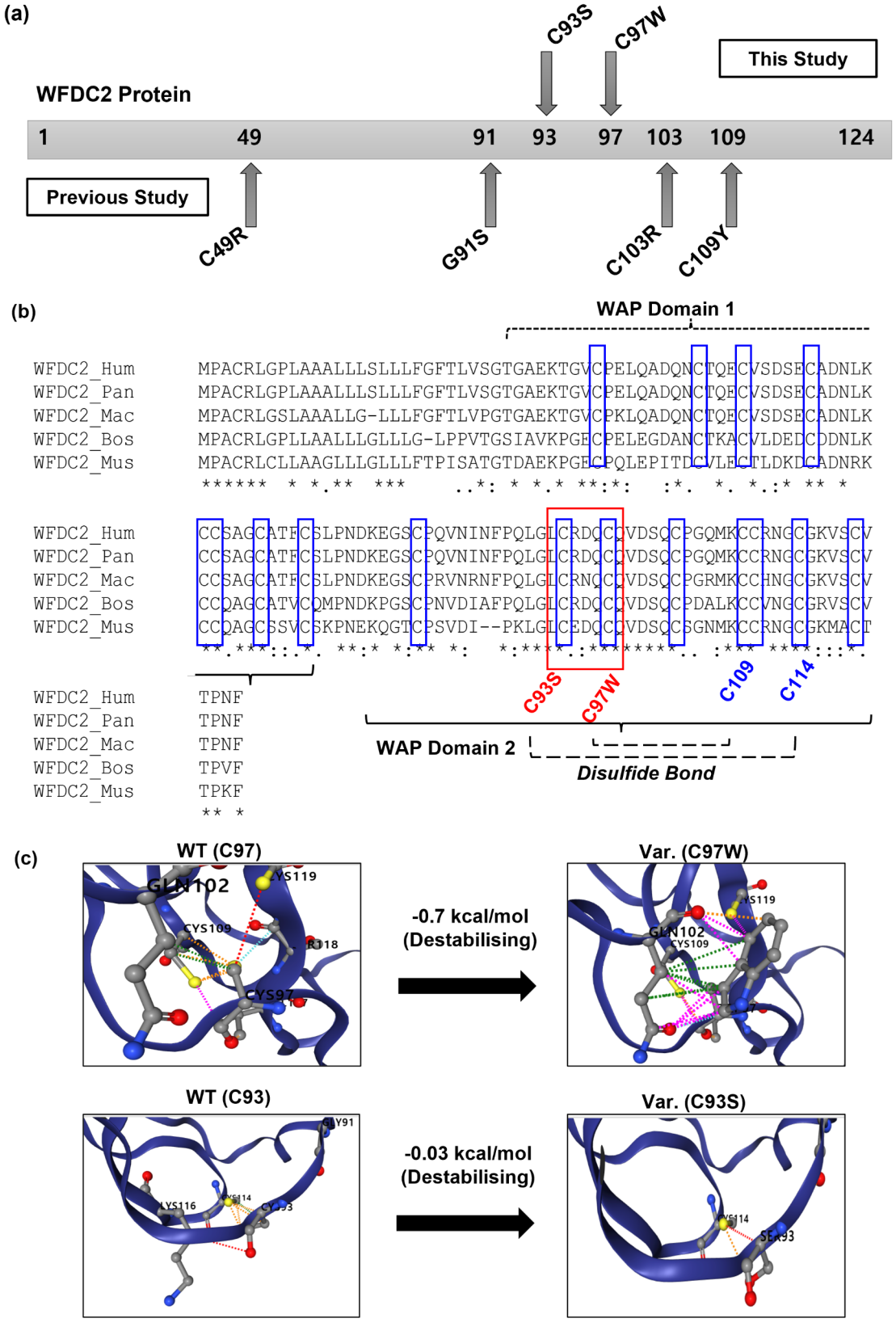

Fig. 3 Conservation of cysteine residues and the change of protein stability in WFDC2. (**a**) Schematic diagram of the pathogenic variants discovered from previous studies and this study. (**b**) The conservation of cysteine residues across all species using ClustalW. Two variants identified in this study (highlighted in red) are linked with other cysteine residues (highlighted in blue) through disulfide bonds. Hum, Human; Pan, Pan paniscus; Mac, Macaca fascicularis; Bos, Bos taurus; Mus, Mus musculus (**c**) In silico prediction of protein stability change using DynaMut2

is widely studied in oncology due to its role as a tumor marker, particularly in ovarian and lung cancers [17, 19–21]. Despite this, none of our patients showed evidence of malignancy, suggesting that pathogenic germline variants in *WFDC2* may primarily contribute to non-cancerous conditions such as bronchiectasis. However, the potential long-term effects of these variants, particularly regarding cancer risk, require further investigation.

One of the strengths of our study is the identification of a homozygous *WFDC2* variant in multiple affected individuals, supporting its recessive inheritance pattern. Additionally, the presence of a compound heterozygous variant in one patient provides further genetic evidence for the role of *WFDC2* mutations in disease development. However, our study had some limitations. Functional validation studies were not conducted, although a recent report in the archive suggests that the c.291 C > G (p.(Cys97Trp) variant disrupts WFDC2 protein folding and secretion [43]. Further research using patient-derived cells or animal models will be essential to confirm the mechanistic impact of these variants. Additionally, while we identified these variants in multiple patients, more extensive studies with more diverse populations are needed to establish the full spectrum of WFDC2-related bronchiectasis.

Conclusions

Our findings provide strong genetic evidence that *WFDC2* mutations contribute to bronchiectasis and recurrent respiratory infections. Given its role in airway defense, further studies are needed to clarify how WFDC2 dysfunction leads to disease and whether targeted therapies can help mitigate its effects. Identifying *WFDC2* as a potential bronchiectasis-associated gene expands our understanding of the genetic landscape of this disease and may pave the way for improved diagnostics and personalized treatment approaches in the future.

Abbreviations

SLPI	Secretory leukocyte protease inhibitor
WAP	Whey-acidic-protein
HPO	Human phenotype ontology
CT	Computed tomography
WGS	Whole-genome sequencing
GATK	Genome analysis toolkit
SNV	Single-nucleotide variant
AB	Allelic balance
ACMG	American college of medical genetics and genomic
KOVA	Korean variant archive
RGC	Regeneron genetics center
FEV1	Forced expiratory volume 1
IVF	In vitro fertilization
TEM	Transmission electron microscopy

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12931-025-03183-z.

Supplementary Material 1: **Table S1**. Allelic information of two *WFDC* genetic variants in population genomic data (gnomAD). **Table S2**. Allelic information of two *WFDC* genetic variants in KOVA and RGC genomic data. **Table S3**. Clinical characteristics of patients with bronchiectasis who have a homozygous c.291 C > G (p.(Cys97Trp)) variant from three study groups

Author contributions

JMK, SH, BHL and MHP designed and organized the study. JMK and SH did formal analysis, interpretated the data and wrote the original manuscrpt. BHL and MHP conceived and supervised the study and revised the manuscript. HWC and YK analyzed the WGS data. DMS constructed the analysis pipeline for WGS data. EL, MK, CL, JWK and HYP contributed to the clinical data. All authors reviewed the manuscript.

Funding

This work was supported by the Korea National Institute of Health intramural research program (2022-NI-060-02, 2025-NI-003-00) and supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Health and Welfare, Ministry of Science and ICT, Ministry of Trade Industry and Energy, Korea Disease Control and Prevention Agency (No. 2020M3E5D7086780, The National Project of Bio Big Data).

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files. Data available on request due to privacy/ethical restrictions.

Declarations

Ethics approval and consent to participate

This study was reviewed and approved by the Institutional Review Board of the Korea National Institute of Health, Korea Disease Control and Prevention Agency (Approval Number: 2022-02-07-2C-A, 2022-09-10-P-A, KDCA-2023-06-06-P-01).

Consent for publication

Written informed consent was obtained from all the participants regarding the use of the genetic analysis and publication of data in this article.

Competing interests

The authors declare no competing interests.

Author details

[1]Division of Genome Science, Department of Precision Medicine, National Institute of Health, Cheongju 28159, Republic of Korea
[2]Department of Pediatrics, Medical Genetics Center, Asan Medical Center, University of Ulsan College of Medicine, Seoul 05505, Republic of Korea
[3]Department of Pediatrics, Chonnam National University Hospital, Chonnam National University Medical School, Gwangju 61469, Republic of Korea
[4]Department of Laboratory Medicine, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul 03397, Republic of Korea
[5]Department of Laboratory Medicine and Genetics, Samsung Medical Center, School of Medicine, Sungkyunkwan University, Seoul 06351, Republic of Korea
[6]National Institute of Health, Cheongju 28159, Republic of Korea

Received: 13 January 2025 / Accepted: 7 March 2025
Published online: 20 March 2025

References

1. Flume PA, Chalmers JD, Olivier KN. Advances in bronchiectasis: endotyping, genetics, microbiome, and disease heterogeneity. Lancet. 2018;392:880–90.
2. Pembridge T, Chalmers JD. Precision medicine in bronchiectasis. Breathe (Sheff). 2021;17:210119.

3. Barker AF. Bronchiectasis. N Engl J Med. 2002;346:1383–93.
4. King PT. The pathophysiology of bronchiectasis. Int J Chron Obstruct Pulmon Dis. 2009;4:411–9.
5. Bush A, Floto RA. Pathophysiology, causes and genetics of paediatric and adult bronchiectasis. Respirology. 2019;24:1053–62.
6. Shoemark A, Ozerovitch L, Wilson R. Aetiology in adult patients with bronchiectasis. Respir Med. 2007;101:1163–70.
7. Pasteur MC, Helliwell SM, Houghton SJ, Webb SC, Foweraker JE, Coulden RA, Flower CD, Bilton D, Keogan MT. An investigation into causative factors in patients with bronchiectasis. Am J Respir Crit Care Med. 2000;162:1277–84.
8. Keir HR, Chalmers JD. Bronchiectasis enters the inflammation era. Respirology. 2022;27:488–9.
9. Oriano M, Amati F, Gramegna A, De Soyza A, Mantero M, Sibila O, Chotirmall SH, Voza A, Marchisio P, Blasi F, Aliberti S. Protease-Antiprotease imbalance in bronchiectasis. Int J Mol Sci. 2021;22:5996.
10. Keir HR, Chalmers JD. Pathophysiology of bronchiectasis. Semin Respir Crit Care Med. 2021;42:499–512.
11. Gramegna A, Amati F, Terranova L, Sotgiu G, Tarsia P, Miglietta D, Calderazzo MA, Aliberti S, Blasi F. Neutrophil elastase in bronchiectasis. Respir Res. 2017;18:211.
12. Gramegna A, Aliberti S, Sibila O, Di Francesco C, Sotgiu G, Perea L, Terranova L, Oriano M, Pilocane T, Saderi L et al. Sputum neutrophil elastase in bronchiectasis: a Southern European cohort study. Eur Respir J. 2020;56:2001702.
13. Sibila O, Perea L, Canto E, Shoemark A, Cassidy D, Smith AH, Suarez-Cuartin G, Rodrigo-Troyano A, Keir HR, Oriano M, et al. Antimicrobial peptides, disease severity and exacerbations in bronchiectasis. Thorax. 2019;74:835–42.
14. Mongkolpathumrat P, Pikwong F, Phutiyothin C, Srisopar O, Chouyratchakarn W, Unnajak S, Nernpermpisooth N, Kumphune S. The secretory leukocyte protease inhibitor (SLPI) in pathophysiology of non-communicable diseases: evidence from experimental studies to clinical applications. Heliyon. 2024;10:e24550.
15. Bingle L, Singleton V, Bingle CD. The putative ovarian tumour marker gene HE4 (WFDC2), is expressed in normal tissues and undergoes complex alternative splicing to yield multiple protein isoforms. Oncogene. 2002;21:2768–73.
16. Bingle CD, Vyakarnam A. Novel innate immune functions of the Whey acidic protein family. Trends Immunol. 2008;29:444–53.
17. Drapkin R, von Horsten HH, Lin Y, Mok SC, Crum CP, Welch WR, Hecht JL. Human epididymis protein 4 (HE4) is a secreted glycoprotein that is overexpressed by serous and endometrioid ovarian carcinomas. Cancer Res. 2005;65:2162–9.
18. Utkarsh K, Kumar A, Aditi, Khan A, Nayyar A, Haque S, Iqbal S. Circulating and non-circulating proteins and nucleic acids as biomarkers and therapeutic molecules in ovarian cancer. Genes Dis. 2023;10:1005–18.
19. Liu W, Liu W, Lin K, Liu Y, Hu H, Yang L. A clinical study of serum human epididymis protein 4 (HE4) in the diagnosis of pancreatic cancer. Indian J Cancer. 2022;59:223–9.
20. Mo D, He F. Serum human epididymis secretory protein 4 (HE4) is a potential prognostic biomarker in Non-Small cell lung cancer. Clin Lab. 2018;64:1421–8.
21. Sai Baba KSS, Rehman MA, Pradeep Kumar J, Fatima M, Raju GSN, Uppin SG, Mohammed N. Serum human epididymis Protein-4 (HE4) - A novel approach to differentiate malignant Frombenign breast tumors. Asian Pac J Cancer Prev. 2021;22:2509–2507.
22. Galgano MT, Hampton GM, Frierson HF Jr. Comprehensive analysis of HE4 expression in normal and malignant human tissues. Mod Pathol. 2006;19:847–53.
23. Hackett NR, Butler MW, Shaykhiev R, Salit J, Omberg L, Rodriguez-Flores JL, Mezey JG, Strulovici-Barel Y, Wang G, Didon L, Crystal RG. RNA-Seq quantification of the human small airway epithelium transcriptome. BMC Genomics. 2012;13:82.
24. Kirchhoff C, Habben I, Ivell R, Krull N. A major human epididymis-specific cDNA encodes a protein with sequence homology to extracellular proteinase inhibitors. Biol Reprod. 1991;45:350–7.
25. Wen Y, Jiang N, Wang Z, Xiao Y. Versatile Whey acidic protein four-disulfide core domain proteins: biology and role in diseases. Front Cell Dev Biol. 2024;12:1459129.
26. Small DM, Doherty DF, Dougan CM, Weldon S, Taggart CC. The role of Whey acidic protein four-disulfide-core proteins in respiratory health and disease. Biol Chem. 2017;398:425–40.
27. Nakajima K, Ono M, Radovic U, Dizdarevic S, Tomizawa SI, Kuroha K, Nagamatsu G, Hoshi I, Matsunaga R, Shirakawa T et al. Lack of Whey acidic protein (WAP) four-disulfide core domain protease inhibitor 2 (WFDC2) causes neonatal death from respiratory failure in mice. Dis Model Mech. 2019;12:dmm040139.
28. Zhang T, Long H, Li J, Chen Z, Wang F, Jiang SW. WFDC2 gene deletion in mouse led to severe dyspnea and type-I alveolar cell apoptosis. Biochem Biophys Res Commun. 2020;522:456–62.
29. Dougherty GW, Ostrowski LE, Nothe-Menchen T, Raidt J, Schramm A, Olbrich H, Yin W, Sears PR, Dang H, Smith AJ et al. Recessively inherited deficiency of secreted WFDC2 (HE4) causes nasal polyposis and bronchiectasis. Am J Respir Crit Care Med. 2024;210:63–76.
30. KISTI. The National Project of Bio Big Data research portal (https://www.cirn.re.kr/index).
31. Kim Y, Kim JM, Cho HW, Park HY, Park MH. Frequency of actionable secondary findings in 7472 Korean genomes derived from the National project of bio big data pilot study. Hum Genet. 2023;142:1561–9.
32. Fu JM, Satterstrom FK, Peng M, Brand H, Collins RL, Dong S, Wamsley B, Klei L, Wang L, Hao SP, et al. Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. Nat Genet. 2022;54:1320–31.
33. Kim JM, Cho HW, Shin DM, Kim OH, Kim J, Lee H, Lee GH, An JY, Yang M, Jo HS, et al. Uncovering potential causal genes for undiagnosed congenital anomalies using an in-house pipeline for trio-based whole-genome sequencing. Hum Genomics. 2025;19:1.
34. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The ensembl variant effect predictor. Genome Biol. 2016;17:122.
35. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 2011;29:24–6.
36. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, Massouras A. VarSome: the human genomic variant search engine. Bioinformatics. 2019;35:1978–80.
37. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. Genet Med. 2015;17:405–24.
38. UniProt C. UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Res. 2023;51:D523–31.
39. Rodrigues CHM, Pires DEV, Ascher DB. DynaMut2: assessing changes in stability and flexibility upon single and multiple point missense mutations. Protein Sci. 2021;30:60–9.
40. Tordai H, Torres O, Csepi M, Padanyi R, Lukacs GL, Hegedus T. Analysis of alphamissense data in different protein groups and structural context. Sci Data. 2024;11:495.
41. Lee J, Lee J, Jeon S, Lee J, Jang I, Yang JO, Park S, Lee B, Choi J, Choi BO, et al. A database of 5305 healthy Korean individuals reveals genetic and clinical implications for an East Asian population. Exp Mol Med. 2022;54:1862–71.
42. Sun KY, Bai X, Chen S, Bao S, Zhang C, Kapoor M, Backman J, Joseph T, Maxwell E, Mitra G et al. A deep catalogue of protein-coding variation in 983,578 individuals. Nature. 2024;631:583–592.
43. Roh JW, Oh J, Kim SJ, Hong JW, Ohk J, Shim HS, Cho H-J, Woo A, Kim SY, Kang YA et al. The founder missense mutation of WFDC2 in Koreans leads to severe respiratory distress accompanied by bronchiectasis and rhinosinusitis. *medRxiv* 2024:2024.11.16.24317083.
44. Masashi Ito KM, Hijikata M, Hasegawa H, Wakabayashi K. Akiko Miyabayashi, Naoto Keicho: severe bronchiectasis and chronic rhinosinusitis due to homozygous WFDC2 variants: the first three cases reported from Japan. Res Square. 2025.https://doi.org/10.21203/rs.3.rs-5949796/v1
45. Parikh K, Antanaviciute A, Fawkner-Corbett D, Jagielowicz M, Aulicino A, Lagerholm C, Davis S, Kinchen J, Chen HH, Alham NK, et al. Colonic epithelial cell diversity in health and inflammatory bowel disease. Nature. 2019;567:49–55.

Publisher's note