Functional assessment and phenotypic heterogeneity of SFTPA1 and SFTPA2 mutations in interstitial lung diseases and lung cancer

Marie Legendre, Afifaa Butt, Raphaël Borie, Marie-Pierre Debray, Diane Bouvry, Emilie Filhol-Blin, Tifenn Desroziers, Valérie Nau, Bruno Copin, Florence Dastot-Le Moal, et al.

Functional assessment and phenotypic heterogeneity of *SFTPA1* and *SFTPA2* mutations in interstitial lung diseases and lung cancer

Marie Legendre[1,2]*, Afifaa Butt*[1], Raphaël Borie[3], Marie-Pierre Debray[4], Diane Bouvry[5], Emilie Filhol-Blin[2], Tifenn Desroziers[1], Valérie Nau[2], Bruno Copin[2], Florence Dastot-Le Moal[2], Mélanie Héry[1], Philippe Duquesnoy[1], Nathalie Allou[6], Anne Bergeron[7], Julien Bermudez[8], Aurélie Cazes[9], Anne-Laure Chene[10], Vincent Cottin[11], Bruno Crestani[4], Jean-Charles Dalphin[12], Christine Dombret[4], Bérénice Doray[13], Clairelyne Dupin[7], Violaine Giraud[14], Anne Gondouin[12], Laurent Gouya[7], Dominique Israël-Biet[15], Caroline Kannengiesser[16], Aurélie Le Borgne[17], Sylvie Leroy[18], Elisabeth Longchampt[19], Gwenaël Lorillon[7], Hilario Nunes[5], Clément Picard[20], Martine Reynaud-Gaubert[8], Julie Traclet[11], Paul de Vuyst[21], Aurore Coulomb L'Hermine[22], Annick Clement[1,23], Serge Amselem[1,2]**, Nadia Nathan[1,23]**

* and **: both authors contributed equally

1. Sorbonne University, Inserm, Childhood genetic disorders, Paris, France
2. Departement of Genetics, Armand Trousseau Hospital, Assistance Publique Hôpitaux de Paris (APHP), Paris, France
3. Pulmonology Department A, Bichat Hospital; APHP, Université de Paris; Paris, France
4. Radiology Department, Bichat Hospital; APHP, Université de Paris; Paris, France
5. Pulmonology Department, EA 2363, Avicenne Hospital, APHP; Paris 13 University, COMUE Sorbonne Paris Cité, Bobigny, France
6. Pulmonology Department, Felix Guyon Hospital ; Saint Denis de La Reunion, France
7. Pulmonology Department, Saint Louis Hospital ; Université de Paris; Paris, France

8. Pulmonology Department and Lung Transplant Team ; North Hospital – Assistance Publique Hôpitaux de Marseille (APHM), Marseille– MEPHI, IHU Méditerranée Infection, Aix-Marseille University, France
9. Pathology Department, Bichat Hospital; APHP, Université de Paris; Paris, France
10. Pulmonology Department, University Hospital; Nantes, France
11. Pulmonology Department and Coordinating Reference center for rare pulmonary diseases OrphaLung, Hospices Civils de Lyon, Claude Bernard University Lyon 1; Lyon, France
12. Pulmonology Department, UMR-CNRS Chrono-Environnement 6249 ; CNRS and CHU Besançon, France
13. Genetic Department, Felix Guyon Hospital ; Saint Denis de La Reunion, France
14. Pulmonology Department, Ambroise Paré Hospital, APHP, Boulogne Billancourt, France
15. Pulmonology Department, Georges Pompidou European Hospital, APHP ; Université de Paris; Paris, France
16. Genetic Departement, Bichat Hospital; APHP, Université de Paris; Paris, France
17. Pulmonology Department, Larrey Hospital, Toulouse, France
18. Pulmonology Department, Pasteur Hospital, Nice, France
19. Pathology Department, Foch Hospital, Suresnes, France
20. Pulmonology Department, Foch Hospital, Suresnes, France
21. Pulmonology Department, Erasme Hospital, Brussels, Belgium
22. Pathology Department, Armand Trousseau Hospital ; Paris, France
23. Pediatric Pulmonology Department and Reference center for rare lung diseases RespiRare; Armand Trousseau Hospital; Paris, France

Running Title: Interstitial lung diseases due to *SFTPA1* and *SFTPA2* mutations

Take home message 255/256-character (including spaces)

SFTPA1 and *SFTPA2* mutations lead to similar alterations in SP-A secretion and lung tissue expression. They are associated with a highly variable phenotypic expression ranging from incomplete penetrance to severe interstitial lung diseases and lung cancer.

Corresponding author
Dr Nadia Nathan
Pediatric Pulmonology Department
Sorbonne Université and Hôpital Trousseau AP-HP
26 avenue du Dr Arnold Netter, 75012-Paris, France
Tel : +33(1) 44 73 66 18
Fax : +33(1) 44 73 67 18
nadia.nathan@aphp.fr

Abstract word count: 248/250

Text word count: 3813/3000

Tables and Figures: 8/8

Supplementary table: 2

Supplementary figures: 5

ABBREVIATIONS

A549	Adenocarcinomic human alveolar basal epithelial
ACMG	American College of Medical Genetics
AEC	Alveolar epithelial cell
ALAT	Latin American thoracic society
ATS	American thoracic society
CRD	Carbohydrate recognition domain
ER	Endoplasmic reticulum
ERS	European respiratory society
EV	Empty vector
GGO	Ground glass opacities
GRCh37	Genome Reference Consortium Human Build 37
HEK293T	Human embryonic kidney 293 with derived large T antigen
HRCT	High resolution computed tomography
HSP	Hypersensitivity pneumonitis
ILD	Interstitial lung disease
IPF	Idiopathic pulmonary fibrosis
JRS	Japanese respiratory society
MDT	Multidisciplinary team
NSIP	Non-specific interstitial pneumonia
NT	Non transfected
PF	Pulmonary fibrosis
PFT	Pulmonary function tests
PPFE	Pleuroparenchymal fibroelastosis
RespiFIL	National network for rare lung diseases (www.respiFIL.fr)
Rs	Referenced single-nucleotide polymorphism
SP	Surfactant protein
TGF	Transforming growth factor
UIP	Usual interstitial pneumonia
UPR	Unfolded protein response
WT	Wild type

ABSTRACT

Introduction: Interstitial lung diseases (ILD) can be caused by mutations in the *SFTPA1* and *SFTPA2* genes, which encode the surfactant protein complex (SP)-A. Only 11 *SFTPA1/2* mutations have so far been reported worldwide, of which 5 have been functionally assessed. In the framework of ILD molecular diagnosis, we identified 14 independent patients with pathogenic *SFTPA1* or *SFTPA2* mutations. The present study aimed to functionally assess the 11 different mutations identified and to accurately describe the disease phenotype of the patients and their affected relatives.

Methods: The consequences of the 11 *SFTPA1/2* mutations were analyzed both *in vitro* by studying the production and secretion of the corresponding mutated proteins and *ex vivo* by analyzing SP-A expression on lung tissue samples. The associated disease phenotypes were documented.

Results: For the 11 identified mutations, protein production was preserved, but secretion was abolished. The expression pattern of lung SP-A, available in 6 patients, was altered. The family history reported ILD and/or lung adenocarcinoma in 13/14 (93%) families. Among the 28 *SFTPA1/2* mutation carriers, the mean age at ILD onset was 45 [0.6-65] years and 48% of them underwent lung transplantation (mean age 51); 7 carriers were asymptomatic.

Discussion: This study, which expands the molecular and clinical spectrum of SP-A disorders, shows that those pathogenic *SFTPA1/A2* mutations share similar consequences on SP-A secretion in cell models and lung tissue immunostaining, whereas they are associated with a highly variable phenotypic expression of the disease, ranging from severe forms requiring lung transplantation to incomplete penetrance.

KEY WORDS

Surfactant protein A, interstitial lung disease, idiopathic interstitial pneumonia, pulmonary fibrosis, lung cancer, genetics.

INTRODUCTION

Chronic interstitial lung diseases (ILD) include a heterogeneous group of diffuse parenchymal lung disorders that can affect patients of all ages [1–3]. In adult patients, the most severe form of ILD, but also the most frequent, is idiopathic pulmonary fibrosis (IPF). The clinical course of ILD patients is highly variable and unpredictable [4–6]. An unknown part of this heterogeneity implicates genetic factors [7–9]. Pathogenic mutations have been identified in approximately 30% of familial ILD, mainly in genes encoding the telomerase complex in adult patients, and mainly in genes encoding proteins of the surfactant metabolism (*SFTPA1, SFTPA2, SFTPB, SFTPC, ABCA3, NKX2-1*) in pediatric cases [7, 10, 8, 11–13].

Information on *SFTPA1* and *SFTPA2* defects in ILD remains limited. These genes are closely related paralogs containing 6 exons each. They encode the surfactant proteins (SP)-A1 and SP-A2, two collectins that oligomerize in a "bouquet-flower" octadecamer to generate the SP-A complex. SP-A plays a structural role in the surfactant tubular myelin formation, but also an immunomodulatory role through its carbohydrate recognition domain (CRD) that is encoded by the last exon of *SFTPA1/A2* (i.e. exon 6). The molecular epidemiology of *SFTPA1/A2* is still largely unknown. Indeed, over the last few years, only 11 *SFTPA1*/*A2* mutations have been reported in patients with familial forms of ILD and lung cancer. The pathogenicity of these mutations was attested by functional tests for only 5 of them [11, 14–20].

In the present study, performed in the framework of ILD molecular diagnosis, we identified 11 missense *SFTPA1* (n=3) or *SFTPA2* (n=8) class 4 (likely pathogenic) and 5 (pathogenic) mutations - according to the American College of Medical Genetics (ACMG) classification of sequence variations - in 14 independent probands [21]. These findings prompted us to assess the functional consequences of these missense variations by means of *in vitro* studies and to accurately describe the associated disease phenotypes in the 14 involved families.

PATIENTS AND METHODS

Patients

Patients with an ILD and with no telomerase complex mutations, were referred to us in the frame of molecular diagnosis. ILD diagnosis was assessed by expert clinicians of one of the 23 adult centres of the French reference network for rare lung diseases (RespiFIL) that covers the whole of France. Phenotypic information - including age at ILD onset, associated extra-respiratory features, lung cancer association, and family history - was collected. Available high-resolution computed tomography (CT)-scans and histologic samples were centrally reviewed [3].

When a class 4 or class 5 mutation was identified, a genetic counseling was performed and a familial genetic screening proposed to the symptomatic individuals and the asymptomatic major relatives [21]. Each family was discussed within the French multidisciplinary team meeting for genetic ILD [22]. The study was approved by the relevant ethics committees ("Comité de Protection des Personnes") and written informed consent was obtained from all participants or legal representatives. Clinical information was collected in a legally authorized database (CNIL N°681248).

Molecular analyses

SFTPA1 and *SFTPA2* molecular analyses were performed on DNA extracted from peripheral blood leukocytes. The NM_005411.5 and NM_001098668.2 isoforms were used as a reference for *SFTPA1* and *SFTPA2,* respectively. Coding exons and flanking intronic junctions of *SFTPA1* and *SFTPA2* were sequenced with a custom targeted-capture panel (SeqCap EZ Choice, Roche Diagnostics). The library was prepared following manufacturer's instructions. Data were analyzed through an in-house double pipeline based on Bowtie2 and BWA tools.

Reads were visualized with the IGV viewer (Broad Institute). Copy number variation analysis was performed with a depth-ratio comparison between subjects sequenced in the same run. Class 4 (n=10) and class 5 (n=1) mutations were checked by Sanger sequencing using the previously described protocol [19].

Haplotype analyses

Haplotype analyses for families 5, 6, and 11 to 14 were performed by combining *SFTPA2* and flanking *SFTPA1* SNP genotyping by targeted-capture in probands and Sanger sequencing in probands and relatives (*SFTPA1* is the closest gene to *SFTPA2*, 50.5kb downstream on chromosome 10). Phasing was performed through allele segregation within each family and through proband NGS data when contiguous variations were included in at least four common sequencing reads.

Structural and electrostatic changes induced by the SP-A1 and SP-A2 mutations on the CRD

3D structures of wild type (WT) and mutant CRD of SP-A1/A2 were generated using the rat SP-A crystal structure (residues 201-248) from the Protein Data Bank (PDB ID: 3PAR) as a template. Briefly, homology modeling was performed using Modeller software v9.10. Accuracy of output structures was further assessed using Procheck v3.5.4 [23, 24]. Conformational changes were evaluated by a RMSD value comparison between the WT and the mutated SP-A1/A2 domains using SuperPose (Back Bone) v1.0 [25]. A <2 RMSD value is in favor of a very similar protein structure. Electrostatic modeling of the protein domain surface was generated with the use of the PyMOL vacuum electrostatics (v0.99).

Plasmid constructs

The WT *SFTPA1* and WT *SFTPA2* cDNAs, obtained from human universal RNA, were inserted into the pcDNA3.1_V5_His_TOPO expression vector (Invitrogen, Carlsbad, CA). A 27-aminoacid HA-tag and a 24-aminoacid FLAG-tag were subsequently added after the signal peptide to generate the pSFTPA1_WT and pSFTPA2_WT plasmids respectively. Site-directed mutagenesis was performed to generate the mutant plasmids carrying the identified mutations using the Quick Change Site Mutation Mutagenesis kit (New England Biolabs, Ipswich, MA, E0554S). The resulting plasmid constructs were checked by Sanger sequencing of the inserts and cloning site flanking junctions.

Cell culture and transfection

WT and mutant pSFTPA1 and pSFTPA2 expression plasmids were used to assess protein production (cell lysate) and secretion (medium) after transfection in HEK293T cells. 600,000 HEK293T cells were first cultured at 37°C with 5% CO_2 into 35mm wells in complete medium (DMEM (1X) + GlutaMAX ™-I (4.5g/L D-Glucose) (Life Technologies, Paisley, UK 31966-021), 10% fetal bovine serum (FBS), 1% Penicillin/Streptomycin). On day 2, at 80% confluence, the cells were transfected with 1µg of either pSFTPA1_WT, pSFTPA2_WT, the corresponding mutated plasmids or the empty vector by the FuGENE HD (Promega, Madison, WI E2312) method with a 4/1 ratio. At day 4, the cells were refed with complete medium without FBS. Protein production was studied at day 5.

Western blot analyses

Aliquots of the cellular lysates and of the cell culture medium were analyzed by western blotting. The membranes were incubated overnight at 4°C, either with a monoclonal anti-HA peroxidase high affinity antibody (Roche 12013819001®, 1/500) or with a monoclonal anti-FLAG peroxidase high affinity antibody (Sigma-Aldrich A8592®, 1/500). Tubulin was used as

a loading control of protein expression by using an anti-α-tubulin peroxidase-coupled antibody (Cell Signaling (11H10) 9099®, 1/1,000) incubated overnight. LAMC1 secretion was used as a loading control of protein secretion by using an anti-LAMC1 antibody (Sigma-Aldrich HPA001908®, rabbit, 1/1,000) incubated overnight, followed by a 1-hour incubation with an anti-rabbit peroxidase-coupled antibody (Cell Signaling 7074®, 1/5000). Results are representative of three independent experiments.

Immunohistochemistry on lung tissues

Tissue samples from lung biopsies or lung explants were centrally analyzed by optic microscopy coupled with computerized image analysis of stained materials, and studied by haematoxylin and eosin (HE) staining and immunohistochemistry [26]. A mouse SP-A monoclonal antibody (Abcam-32E12, 1/200) was used according to manufacturer's protocol. Histopathological descriptions were provided according to the ATS/ERS/JRS/ALAT Clinical Practice Guideline and SP-A staining was analyzed on the available tissues [3].

RESULTS

Spectrum of *SFTPA1* and *SFTPA2* mutations in 14 independent patients

A total of 11 heterozygous class 4 or class 5 mutations were identified: 3 *SFTPA1* mutations in three probands and 8 *SFTPA2* mutations in 11 probands (**Table 1**). Eight of them have so far never been reported. All are missense mutations identified in exon 6 that encodes the CRD of SP-A1 and SP-A2. All these mutations involve an invariant or highly conserved residue in SP-A1 and SP-A2 and throughout evolution (**Figure 1**). None of them are prone to alter splicing

according to the MaxEntScan tool. Their pathogenicity, as assessed through *in silico* means, is provided in **Table 1** [21]. Three *SFTPA2* mutations have been each characterized in two distinct families: V178M in families 5 and 6, C238S in families 11 and 12, and R242Q in families 13 and 14 **(Figure 2)**. This could be related to a common ancestor to two families, or to recurrent mutational events. Families 5 and 6 originate respectively from China and North Africa, and families 13 and 15 from North Africa and La Réunion Island (**Table 2**). Haplotyping showed that V178M arose on different haplotypes in families 5 (haplotype H2) and 6 (haplotype H4) (**Supplemental Figure 1A**). The R242Q also arose on distinct haplotypes in families 13 (haplotype H11) and 14 (haplotype H14) (**Supplemental Figure 1C**). These data, together with the diverse geographic origins of the families and the fact that those two mutations are G to A transitions involving a CG dimer, strongly support the hypothesis of recurrent mutational events in discrete populations. Regarding the C238S mutation, haplotype analysis of 17 SNPs did not exclude the existence of an ancestor common to families 11 and 12 that both originate from North Africa (Algeria) (**Supplemental Figure 1B**). This variation in not described in the gnomAD database and is a G to C transversion (mutation mechanism independent of CG dimer hotspots). However, the H6 haplotype characterized in both families is made of the most frequent versions of the 17 SNPs in humans, and the occurrence of a recurrent mutational event in a common haplotype cannot be fully rejected.

Structural and electrostatic changes induced by the identified SP-A1 and SP-A2 mutations

As compared to WT SP-A1 and SP-A2 CRD domains, prediction of the 3D structure of the protein carrying the mutations disclosed no major consequences on the overall domain structure: RMSD values < 1.1 between WT and mutated SP-A1 or SP-A2 domains. The

electrostatic modeling of the mutant protein domain surface shows polarity changes for 6 of the mutants, as compared to the corresponding WT SP-A CRD domain (**Supplemental Figure 2**).

Impact of the identified missense variants on SP-A1/SP-A2 production and secretion

As shown on the western blots performed on whole cell lysates, and on the histograms with densitometry ratio from 3 or more independent experiments (**Figure 3**), the expression of each mutant SP-A1 and SP-A2 protein (35-37 KDa) is similar to the normal proteins in HEK293T lysates ($p>0.05$, Student test), thereby indicating that the identified *SFTPA1* and *SFTPA2* missense variations have no effect on protein expression. However, the expression of each mutant SP-A1 and SP-A2 protein is dramatically reduced in the cell medium ($p <0.05$, Student test), thereby demonstrating that these missense variations result in the absence of SP-A1 and SP-A2 secretion and are therefore deleterious.

Table 1: Description of the 11 identified *SFTPA1* and *SFTPA2* mutations

Gene	Mutation name (cDNA)	Mutation name (protein)	Mutation abbreviated name	rs	GnomAD* allele frequency	Conservation**	Prone to alter splicing ✣	Polyphen score HumDiv ✯	Sift score†	Mutation class ✦	First report
SFTPA1	c.532G>A	p.(Val178Met)	V178M	1215316727	1/250994	5/7	No	1	0	4	*Doubková et al. 2019*
	c.631T>C	p.(Trp211Arg)	W211R	_	Not described	7/7	No	1	0	5	*Nathan et al. 2016*
	c.673G>A	p.(Val225Met)	V225M	_	Not described	7/7	No	1	0	4	This study
SFTPA2	c.512A>T	p.(Asn171Ile)	N171I	_	Not described	7/7	No	1	0	4	This study
	c.532G>A	p.(Val178Met)	V178M	371035540	3/250974	5/7	No	1	0	4	*Coghlan et al. 2014*
	c.542A>G	p.(Tyr181Cys)	Y181C	-	Not described	7/7	No	0.99	0.01	4	This study
	c.697T>A	p.(Trp233Arg)	W233R	-	Not described	7/7	No	1	0	4	This study
	c.698G>T	p.(Trp233Leu)	W233L	-	Not described	7/7	No	1	0	4	This study
	c.699G>C	p.(Trp233Cys)	W233C	-	Not described	7/7	No	1	0	4	This study
	c.713G>C	p.(Cys238Ser)	C238S	-	Not described	7/7	No	1	0	4	This study
	c.725G>A	p.(Arg242Gln)	R242Q	759988686	3/251444	7/7	No	1	0.05	4	This study

All identified mutations are located in exon 6 of *SFTPA1* or *SFTPA2*, encoding the carbohydrate domain of the protein.

* GnomAD: includes non-filtered and filtered variants (number of mutated alleles / total number of alleles)

** The conservation has been evaluated through alignment of protein sequences of SP-A (SP-A1 and SP-A2) from species (*Homo sapiens, Pan troglodytes, Canis lupus, Bos taurus, Mus musculus, Rattus norvegicus, Gallus gallus*)

✣ According to the MaxEntScan tool that is based on the maximum entropy principle applied to RNA splicing signals.

✯ Polyphen HumDiv score is compiled from all damaging alleles with known effects on the molecular function causing human Mendelian diseases, present in the UniProtKB database, together with differences between human proteins and their closely related mammalian homologs, assumed to be non-damaging. A score ≥0.957 is considered as probably damaging, a score between 0.453 and 0.956 is considered as possibly damaging and a score ≤0.452 is considered as benign.

† SIFT predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids.

✦ According to the international guidelines: pathogenic (class 5), likely pathogenic (class 4), prior to this study's functional assessment [21].

Abbreviations: GRCh37, Genome Reference Consortium Human Build 37; rs, referenced single-nucleotide polymorphism; Polyphen, Polymorphism Phenotyping v2.

Table 2: Clinical characteristics of the probands with identified *SFTPA1* or *SFTPA2* mutations

Gene	Mutation	Family	Individual *	Gender	Geographic origin	Age at onset (year)	Exposures	Lung cancer	Other clinical features	CT-scan features of ILD	Family history	Treatments	Outcome
SFTPA1	V178M	1 - GM01220	**II.2**	M	North Africa	57	Tobacco	Yes	Seizures, heart disease	UIP	ILD and lung cancer	Pirfenidone at 64 for 3 months then Nintedanib. No oxygen. No treatment for adenocarcinoma	Death at 68
	W211R	2 - PP005	**IV.3**	M	Europe	45	Tobacco, silica	Yes	No	Indeterminate pattern	ILD and lung cancer	Anti-fibrosing therapy.	LT at 57 years-old. Alive at 59
	V225M	3 - GM0537	**II.1**	F	North Africa	49	Pigeons, parakeet, dust	No	No	Indeterminate pattern	ILD	Anti-fibrosing therapy	LT at 54, Alive at 61
			III.1	F		NA	No	No	No	No ILD			Asymptomatic at 32
SFTPA2	N171I	4 - PP375	**IV.1**	F	Europe	32	No	No	Rheumatoid arthritis	UIP and PPFE	ILD		LT at 37. Alive at 40
	V178M	5 - GM01611	**III.1**	F	China	33	No	No	Rheumatoid arthritis	Indeterminate pattern and PPFE	ILD		LT at 45. Alive at 48
		6 - GM02482	**II.1**	F	North Africa	34	Tobacco	No	No	Indeterminate pattern	ILD and lung cancer	No O_2, no treatment	Alive at 37
	Y181C	7 - PP354	**II.6**	M	North Africa	56	Tobacco	No	No	ND	ILD and lung cancer	Anti-fibrosing therapies, cyclophosphamide, steroid pulses.	Death at 63 from acute exacerbation
	W233R	8 - GM03569	**IV.4**	M	North Africa	34	Tobacco, dust, cannabis	Yes	No	Indeterminate pattern	ILD and lung cancer	Anti-fibrosing therapy (nintedanib) and chemotherapy, O_2	Death at 41 from lung fibrosis and adenocarcinoma progression
	W233L	9 - GM02867	**II.1**	M	Europe	45	No	No	No	Indeterminate pattern	ILD and lung cancer		LT at 49, alive at 54

W233C	10 - GM00127	**II.1**	F	Europe	50	Tobacco, birds	Yes	No	Fibrosing NSIP	No	O_2, cyclophosphamide	Death at 64 from lung cancer and sepsis
C238S	11 - GM01161	**III.2**	M	North Africa	46	Tobacco, cement	Yes	No	ND	ILD and lung cancer	Pirfenidone 47	LT at 49, death at 51
	12 - GM00095	**II.1**	M	North Africa	26	No	No	Klinefelter syndrome	HSP	Lung cancer	No treatment	Asymptomatic, Alive at 51
R242Q	13 - PP092	**II.1**	M	North Africa	29	Tobacco	No	No	ND	ND		LT at 37, death at 39 from LT complications.
	14 - GM00077	**II.2**	M	Reunion Island	51	No	No	No	NSIP	ILD	O_2 before LT, no anti-fibrosing therapies	LT at 51, death at 51, 6 weeks later from sepsis

***** According to **Figure 2**

Abbreviations: ILD, interstitial lung disease, UIP, usual interstitial pneumonia; NSIP, non-specific interstitial pneumonia; PPFE, pleuroparenchymal fibroelastosis; HSP, hypersensitivity pneumonitis; LT, lung transplantation

Parenchymal expression pattern of SP-A in the lung tissue of probands

The tissue samples from the lung biopsy or explanted lungs of 4 probands (**Figure 4**) showed "indeterminate for usual interstitial pneumonia (UIP)" pattern. The SP-A expression pattern was characterized by a marked SP-A expression in hyperplastic pneumocytes and a discontinuous expression surrounding the alveolar space. Such expression pattern contrasts with the light expression and thin continuous linear layer of SP-A at the alveolar surface of the control specimen [26]. In all patients, a residual SP-A expression was observed in alveolar macrophages, consistently with the fact that the antibody used in these experiments recognizes both SP-A1 and SP-A2 that are 98% identical. The same features were observed on the lung biopsy of 2 affected siblings (**Supplemental Figure 5**).

Clinical characteristics of the 14 patients and their 14 relatives carrying a *SFTPA1/2* mutation

The detailed phenotypic characteristics of the 14 unrelated probands are provided in **Table 2**. Eight (57%) of them originated from North Africa. Only one proband aged 29 at ILD onset (family 13, II.1) had neither lung cancer nor family history of lung cancer or ILD. Eleven (78%) had a first-degree history of ILD and 9 (64%) had a personal (n=5, 36% assessed by lung biopsy and/or explant analysis) or family (n=7, 50%) history of lung cancer (**Figure 2**). A total of 10 proband chest CT-scans could be centrally reviewed by a thoracic radiologist **(Figure 5)** [3]. Only 2 patients showed a usual interstitial pneumonia (UIP) pattern [3]. Half of cases were classified indeterminate for UIP and 3 cases suggested either nonspecific interstitial pneumonia or hypersensitivity pneumonitis (HSP). Two patients had apical subpleural consolidations evoking pleuroparenchymal fibroelastosis associated to UIP and indeterminate fibrosis (one case each). In total, predominant ground glass opacities (GGO) were observed in 7 out of 10

(70%) patients and lung fibrosis was observed in 7 out of 10 patients (70%). Other inconstant features were basal thickened interlobular septa, micronodules, reticulations, honeycombing and bronchiectasis. The same results were observed on the chest CT-scans of 2 affected siblings (**Supplemental Figure 4**).

The family screening led to identify a *SFTPA1* or *SFTPA2* mutation in 14 more individuals (**Figure 2** and **Supplemental Table 1**).

Altogether, the present study allowed identifying 28 individuals with a *SFTPA1/A2* mutation. The male to female ratio was 0.65. The median age at onset of the ILD was 45 years old [0.6-65]. Five (22%) patients presented an extra-respiratory manifestation, including two patients with rheumatoid arthritis (9%). Ten patients (43%) declared neither tobacco nor professional exposure. Seven relatives remained asymptomatic until a median age of 39 [19-55] and after a median length of follow-up of 5, range [1-8] years. The penetrance of the disease was incomplete as among them, 2 individuals were aged over 40 (42 and 55 years respectively) and harbored normal pulmonary function tests (PFT) and CT-scan. The other asymptomatic relatives had either abnormal PFT and/or minimal interstitial abnormalities on the CT-scan at ages 18, 24, 32, 33, 39 years-old respectively. The first and last available PFT of the patients are provided in **Supplemental Figure 3**. The evolution of the ILD was severe: 5 patients needed anti-fibrosing therapies and 11 (48%) patients received a lung transplantation at a median age of 51 [37-58] years with a median time from diagnosis of 5 years [0-19]. Regarding lung transplantation, one patient with ILD and surgically treated lung cancer (Family 2, IV.3) benefit from an atypical indication: despite a minimally invasive with a lepidic pattern adenocarcinoma on the contralateral lung almost 10 years after the first one, lung transplantation was decided, with a 2-years favorable outcome to date. For all the patients, the lung transplantation was bilateral, in order to prevent lung fibrosis extension and lung cancer development on the native

lung. Nine (32%) patients died at a median age of 51 years old [0.7-68], three of them shortly after lung transplantation (7 days, 6 weeks and 13 months respectively) (**Supplemental Table 1** and **Figure 6**).

DISCUSSION

The present study describes for the first time a large number of functionally assessed pathogenic *SFTPA1* and *SFTPA2* mutations in 14 independent families. Three mutations were identified in *SFTPA1* and 8 in *SFTPA2,* bringing to 5 and 14 respectively the total number of reported *SFTPA1* and *SFTPA2* mutations (**Supplemental Table 2**) [11, 14–19]. Interestingly, all the described mutations are located in exon 6 that encodes the protein CRD, and all but one (p.(Tyr208His)) were found in the heterozygous state.

In this study, the pathogenicity of the identified mutations was assessed by *in vitro* functional studies and *ex-vivo* immunostaining showing that the secretion of all the mutant proteins was similarly abolished and that the protein may be sequestered in the alveolar epithelial cells (AEC)-2. Although part of these results rely on a cell model that overexpresses recombinant proteins, such quite unequivocal presentation of the disease at the cellular level contrasts sharply with the highly variable clinical expression of the disease, which ranges from incomplete penetrance to severe forms requiring lung transplantation.

Phenotypic heterogeneity

The individuals with *SFTPA1/A2* mutations presented with a heterogeneous clinical phenotype, with no evidence of phenotypic differences between patients with *SFTPA1* mutations and those with *SFTPA2* mutations. The penetrance of the disease was incomplete, with the identification of carriers remaining asymptomatic in adulthood (with normal chest CT-scan and pulmonary function tests). The age of onset was variable, from childhood to elderly [19]. However, in the

majority of the patients, the ILD occurred at a younger age (mean 40 years old; SD 14; range [0.6-65]) than the usual age at IPF onset as only one patient experienced a disease onset after 60 years-old. The CT-scan analysis gave highly heterogeneous results in terms of ATS/ERS classification. This probably highlights that this classification is maybe not suitable in such cases. As shown in the current study, and as observed in children with surfactant disorders, the GGO pattern represents the most frequent pattern [27]. This observation differs from the typical UIP pattern that is seen in IPF and could be one of the main arguments – on top of the family form of ILD - for suspecting a surfactant disorder in adults. Of interest, no relationship could be established between the severity of the CT images and the mutations. Moreover, various CT patterns were observed within the same family. Altogether, as it is also the case in other causes of genetic ILD, the variability in disease expression was observed among the different individuals, including among those from a same family [22].

Variability in clinical expression of the disease phenotype is a common feature of most genetic disorders, especially when dominant. In theory, this phenomenon may result from allelic heterogeneity and/or from the influence of environmental or modifying genetic factors [28]. Moreover, post-translational modifications of SP-A proteins could at least partly account for the heterogeneous presentation of the clinical phenotype. The role of viruses has been discussed in pediatric patients who develop fatal pulmonary fibrosis [6]. This hypothesis has also been discussed for mutations in other surfactant genes, especially *SFTPC* and *ABCA3* mutations [29, 30]. In adults, tobacco smoking and occupational exposures are supposed to be the main triggers in adult cases of IPF [6, 31] but the role of viral infections as a risk factor for pulmonary fibrosis has also been reported by several authors and recently confirmed in a meta-analysis [32]. The interactions between SP-A and various viruses have been studied *in vitro* and in *SFTPA* knock-out mouse models. The viral infections appeared to promote lung fibrosis development, whereas

414 the adjunction of a SP-A treatment allowed a beneficial effect on lung fibrosis progression [33].
415 SP-A binds the pathogens oligosaccharides via its CRD domain, induces their opsonisation
416 and/or phagocytosis and promotes a pro-inflammatory response [34]. Alternatively, the CRD
417 binding to the alveolar macrophages could induce an anti-inflammatory response by enhancing
418 apoptotic cells phagocytosis. This CRD ability to regulate pro- and anti-inflammatory responses
419 is known as the "head or tail hypothesis" depending on the orientation of SP-A binding to
420 alveolar macrophages via a CRD-dependent or CRD-independent mechanism [35, 36]. In this
421 context, molecular alterations of SP-A1/A2 CRD could alter its ability to respond to pathogens
422 exposures and to modulate pulmonary inflammation caused by environmental triggers.
423
424 The observation of a SP-A accumulation in the AEC2 observed on lung tissues sections
425 suggests the production of misfolded proteins in the AEC2 that could induce variable levels of
426 endoplasmic reticulum (ER) stress and promote cell death or cell reprogramming [37]. The
427 unsecreted abnormal SP-A protein could thus be responsible for an unfolded protein response
428 (UPR) challenge. UPR and ER stress have been found increased in case of *SFTPC* mutations
429 [37]. Evidence for increased ER stress as well as abnormal oligomerization and increased
430 proteasome degradation has also been reported for mutant SP-A2 proteins (G231V and F198S
431 mutations) [38]. Similarly, *SFTPA1* mutations have been associated with an increased
432 expression of ER stress and necroptosis markers (Y208H mutation) [39]. Thus, there are
433 compelling evidence for a major role of ER stress in *SFTPA1/A2* mutations, which need to be
434 further confirmed on patients' lung tissues.
435
436 **Lung cancer association**
437 IPF has been associated with an increased risk of lung cancer, such association being observed
438 in up to 13.5% of the patients with pulmonary fibrosis [40, 41]. In the present study, the

association of ILD and lung cancer reaches 36% in the probands, and 64% when including family history, which strongly supports a link between *SFTPA1/A2* CRD mutations and lung cancer development. Interestingly, such association has not been reported for *SFTPC* mutations. Given the severity of the *SFTPC*-associated disease that leads to lung transplantation or to death at a young age, the follow-up may not be long enough to lead to tumorigenesis promotion. ER stress, via UPR, is also linked to tumorigenesis [42]. Increased ER stress can alter tissue homeostasis by interfering with cell cycle in many ways, including cell death or cell senescence attenuation (and thus tumorigenesis promotion) [43]. In case of SP-A mutations, these hypotheses still need to be further evaluated. It has also been proposed that an increased expression of SP-A could reduce adenocarcinogenesis in mice [44]. This could be explained by the interactions between SP-A and Transforming growth factor (TGF)-β, a crucial molecule implicated in lung repair and tumor proliferation: SP-A has been shown to inhibit TGF-β1 inactivation [45, 46]. Thus, mutations of SP-A that alter its secretion could prevent SP-A from inhibiting the TGF-β pathway, leading to an impairment of the alveolar healing process [47].

Genetic studies, family screening and genetic counselling

This study suggests that a genetic screening for *SFTPA1/A2* mutations should be proposed to patients with (i) a family form of ILD, especially if associated to a personal or family history of lung cancer; (ii) an early-onset ILD (before 50 years) and no telomerase complex gene mutations. When a mutation is found, in line with the incomplete penetrance and the risk of ILD and lung cancer, a family screening should be offered to the siblings and other relatives, including asymptomatic major relatives. The pre-symptomatic management of individuals with *SFTPA1/A2* mutations is not consensual yet, but it is likely that tobacco smoking and environmental exposures can enhance the disease expression, and that genetic counseling should also include targeted preventive measures. In France, national multidisciplinary team

(MDT) meetings dedicated to genetic forms of pulmonary fibrosis have recently been launched in the frame of the RespiFIL network [22]. Among other issues, prenatal and pre-implantation managements are discussed for each family. Following the MDT report, a personalized genetic counseling is given to the patient and his family. As shown in this study, the phenotypic heterogeneity of the disease is important and could change the families' perception of the disease severity and their expectations in terms of prenatal / pre-implantation management.

Conclusion

Heterozygous *SFTPA1* and *SFTPA2* mutations involving the exon 6 that encodes the CRD of the corresponding proteins contribute to various fibrotic ILDs that can occur in young adults and probably during childhood, and to a higher risk of lung cancer. Their pathogenicity is attested to by impaired protein secretion leading to cytoplasmic retention in the alveolar epithelium. However, the penetrance of the disease phenotype is incomplete and clinical expression of the disease is highly variable with a major proportion of patients, including young adults, presenting a severe disease requiring lung transplantation. As such, family screening and genetic counseling as well as pre-symptomatic management of the relatives are major issues that need to be discussed in the frame of multidisciplinary team meetings.

AUTHORS' CONTRIBUTIONS

NN, ML, AC, SA designed the study; NN, AB, EFB, TD, MH, VN, ACH performed experiments; NN, ML, BC, PD, EFB, TD, MH, AB, ACH, MPD, AC, SA analyzed and interpreted the data; NN and AB wrote the manuscript; ML, SA, AC, BC, MPD, ACH, RB, FDM, CK, LG, VC, JT, BCr, AG, CD, CDo, ACa, ALC, VG, HN, DB, ABe, GL, DIB, PV, CP, EL, MRG, JCD, SL, ALB, NA provided the patients' data; All authors reviewed and approved the manuscript.

ACKNOWLEDGMENTS

We wish to thank the patients and their families for their participation in the study. We also thank the Assistance Publique Hôpitaux de Paris, Sorbonne Université, Paris, France, and the national networks for rare lung diseases: Centre de référence des maladies respiratoires rares (RespiRare), Centre de référence des maladies pulmonaires rares (OphaLung) and Filière de soins pour les maladies respiratoires rares (RespiFIL). The ILD cohort is developed in collaboration with the Rare Cohort Disease (RaDiCo)-ILD project (ANR-10-COHO-0003), the FP7-305653-child-EU project and the COST Action European network for translational research in children's and adult interstitial lung disease (COST-ILD) project (CA16125).

CONFLICT OF INTEREST STATEMENT

Dr Borie declared he has been invited by to national and international meetings, and/or that he received grants and/or personal fees for various missions from Boehringer Ingelheim, Roche, Savapharma, outside the submitted work.

Prof Cottin declared he has been invited by to national and international meetings, and/or that he received grants and/or personal fees for various missions from Actelion, Boehringer Ingelheim, Bayer/MSD, GSK, Novartis, Roche/Promedior, Sanofi, Celgene, Galapagos, outside the submitted work.

The other authors declare no conflict of interest.

FUNDING

This work was supported by grants from the Institut National de la Santé et la Recherche Médicale (INSERM), the Legs Poix from the Chancellerie des Universités (grants 2013 n°1305, 2014 n°1405, 2015 n°1015, 2016 n°2077 and 2017 n°DP2017/1860), Paris, the Société

Française de Pédiatrie - Société Pédiatrique de Pneumologie et d'Allergologie- AstraZeneca, as well as fundings from the patient organizations Respirer c'est Grandir and Belleherbe Association.

REFERENCES

1. Bush A, Cunningham S, de Blic J, Barbato A, Clement A, Epaud R, Hengst M, Kiper N, Nicholson AG, Wetzke M, Snijders D, Schwerk N, Griese M, chILD-EU collaboration. European protocols for the diagnosis and initial treatment of interstitial lung disease in children. *Thorax* 2015; 70: 1078–1084.
2. Kurland G, Deterding RR, Hagood JS, Young LR, Brody AS, Castile RG, Dell S, Fan LL, Hamvas A, Hilman BC, Langston C, Nogee LM, Redding GJ, American Thoracic Society Committee on Childhood Interstitial Lung Disease (chILD) and the chILD Research Network. An official American Thoracic Society clinical practice guideline: classification, evaluation, and management of childhood interstitial lung disease in infancy. *Am. J. Respir. Crit. Care Med.* 2013; 188: 376–394.
3. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, Behr J, Cottin V, Danoff SK, Morell F, Flaherty KR, Wells A, Martinez FJ, Azuma A, Bice TJ, Bouros D, Brown KK, Collard HR, Duggal A, Galvin L, Inoue Y, Jenkins RG, Johkoh T, Kazerooni EA, Kitaichi M, Knight SL, Mansour G, Nicholson AG, Pipavath SNJ, Buendía-Roldán I, et al. Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am. J. Respir. Crit. Care Med.* 2018; 198: e44–e68.
4. Collard HR, Ryerson CJ, Corte TJ, Jenkins G, Kondoh Y, Lederer DJ, Lee JS, Maher TM, Wells AU, Antoniou KM, Behr J, Brown KK, Cottin V, Flaherty KR, Fukuoka J, Hansell DM, Johkoh T, Kaminski N, Kim DS, Kolb M, Lynch DA, Myers JL, Raghu G, Richeldi L, Taniguchi H, Martinez FJ. Acute Exacerbation of Idiopathic Pulmonary Fibrosis. An International Working Group Report. *Am. J. Respir. Crit. Care Med.* 2016; 194: 265–275.
5. Clement A, de Blic J, Epaud R, Galeron L, Nathan N, Hadchouel A, Barbato A, Snijders D, Kiper N, Cunningham S, Griese M, Bush A, Schwerk N, chILD-EU collaboration. Management of children with interstitial lung diseases: the difficult issue of acute exacerbations. *Eur. Respir. J.* 2016; 48: 1559–1563.
6. Nathan N, Sileo C, Thouvenin G, Berdah L, Delestrain C, Manali E, Papiris S, Léger P-L, Pointe HD le, l'Hermine AC, Clement A. Pulmonary Fibrosis in Children. *J. Clin. Med.* 2019; 8.
7. Garcia CK. Idiopathic pulmonary fibrosis: update on genetic discoveries. *Proc Am Thorac Soc* 2011; 8: 158–162.
8. Turcu S, Ashton E, Jenkins L, Gupta A, Mok Q. Genetic testing in children with surfactant dysfunction. *Arch. Dis. Child.* 2013; 98: 490–495.
9. Nathan N, Borensztajn K, Clement A. Genetic causes and clinical management of pediatric interstitial lung diseases. *Curr. Opin. Pulm. Med.* 2018; 24: 253–259.
10. Kropski JA, Lawson WE, Young LR, Blackwell TS. Genetic studies provide clues on the pathogenesis of idiopathic pulmonary fibrosis. *Dis. Model. Mech.* 2013; 6: 9–17.
11. Coghlan MA, Shifren A, Huang HJ, Russell TD, Mitra RD, Zhang Q, Wegner DJ, Cole FS, Hamvas A. Sequencing of idiopathic pulmonary fibrosis-related genes reveals independent single gene associations. *BMJ Open Respir. Res.* 2014; 1: e000057.

12. Borie R, Kannengiesser C, Sicre de Fontbrune F, Gouya L, Nathan N, Crestani B. Management of suspected monogenic lung fibrosis in a specialised centre. *Eur. Respir. Rev.* 2017; 26.
13. Newton CA, Batra K, Torrealba J, Kozlitina J, Glazer CS, Aravena C, Meyer K, Raghu G, Collard HR, Garcia CK. Telomere-related lung fibrosis is diagnostically heterogeneous but uniformly progressive. *Eur. Respir. J.* 2016; 48: 1710–1720.
14. Wang Y, Kuan PJ, Xing C, Cronkhite JT, Torres F, Rosenblatt RL, DiMaio JM, Kinch LN, Grishin NV, Garcia CK. Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis and lung cancer. *Am. J. Hum. Genet.* 2009; 84: 52–59.
15. van Moorsel CHM, Ten Klooster L, van Oosterhout MFM, de Jong PA, Adams H, Wouter van Es H, Ruven HJT, van der Vis JJ, Grutters JC. SFTPA2 Mutations in Familial and Sporadic Idiopathic Interstitial Pneumonia. *Am. J. Respir. Crit. Care Med.* 2015; 192: 1249–1252.
16. Doubková M, Staňo Kozubík K, Radová L, Pešová M, Trizuljak J, Pál K, Svobodová K, Réblová K, Svozilová H, Vrzalová Z, Pospíšilová Š, Doubek M. A novel germline mutation of the SFTPA1 gene in familial interstitial pneumonia. *Hum. Genome Var.* 2019; 6: 12.
17. Takezaki A, Tsukumo S-I, Setoguchi Y, Ledford JG, Goto H, Hosomichi K, Uehara H, Nishioka Y, Yasutomo K. A homozygous SFTPA1 mutation drives necroptosis of type II alveolar epithelial cells in patients with idiopathic pulmonary fibrosis. *J. Exp. Med.* 2019; .
18. Sritharan SS, Gajewska ME, Skytte A-BS, Madsen LB, Bendstrup E. Familial idiopathic pulmonary fibrosis in a young female. *Respir. Med. Case Rep.* 2018; 24: 1–4.
19. Nathan N, Giraud V, Picard C, Nunes H, Dastot-Le Moal F, Copin B, Galeron L, de Ligniville A, Kuziner N, Reynaud-Gaubert M, Valeyre D, Couderc L-J, Chinet T, Borie R, Crestani B, Simansour M, Nau V, Tissier S, Duquesnoy P, Mansour-Hendili L, Legendre M, Kannengiesser C, Coulomb L'Hermine A, Gouya L, Amselem S, Clement A. Germline SFTPA1 mutation in familial idiopathic interstitial pneumonia and lung cancer. *Hum. Mol. Genet.* 2016; 25: 1457–1467.
20. Liu L, Qin J, Guo T, Chen P, Ouyang R, Peng H, Luo H. Identification and functional characterization of a novel surfactant protein A2 mutation (p.N207Y) in a Chinese family with idiopathic pulmonary fibrosis. *Mol. Genet. Genomic Med.* 2020; : e1393.
21. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, ACMG Laboratory Quality Assurance Committee. 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: 405–424.
22. Borie R, Kannengiesser C, Gouya L, Dupin C, Amselem S, Ba I, Bunel V, Bonniaud P, Bouvry D, Cazes A, Clement A, Debray MP, Dieude P, Epaud R, Fanen P, Lainey E, Legendre M, Plessier A, Sicre de Fontbrune F, Wemeau-Stervinou L, Cottin V, Nathan N, Crestani B. Pilot experience of multidisciplinary team discussion dedicated to inherited pulmonary fibrosis. *Orphanet J. Rare Dis.* 2019; 14: 280.
23. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 1993; 26: 283–291.
24. Rullmann JAC. AQUA, Computer Program, Department of NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, The Netherlands. 1996.
25. Maiti R, Van Domselaar GH, Zhang H, Wishart DS. SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res.* 2004; 32: W590–W594.
26. Ochs M, Johnen G, Müller K-M, Wahlers T, Hawgood S, Richter J, Brasch F. Intracellular and intraalveolar localization of surfactant protein A (SP-A) in the parenchymal

609 region of the human lung. *Am. J. Respir. Cell Mol. Biol.* 2002; 26: 91–98.
610 27. Semple TR, Ashworth MT, Owens CM. Interstitial Lung Disease in Children Made
611 Easier…Well, Almost. *Radiogr. Rev. Publ. Radiol. Soc. N. Am. Inc* 2017; 37: 1679–1703.
612 28. Nathan N, Corvol H, Amselem S, Clement A. Biomarkers in Interstitial lung diseases.
613 *Paediatr. Respir. Rev.* 2015; 16: 219–224.
614 29. Glasser SW, Witt TL, Senft AP, Baatz JE, Folger D, Maxfield MD, Akinbi HT,
615 Newton DA, Prows DR, Korfhagen TR. Surfactant protein C-deficient mice are susceptible to
616 respiratory syncytial virus infection. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 2009; 297: L64-
617 72.
618 30. Kaltenborn E, Kern S, Frixel S, Fragnet L, Conzelmann K-K, Zarbock R, Griese M.
619 Respiratory syncytial virus potentiates ABCA3 mutation-induced loss of lung epithelial cell
620 differentiation. *Hum. Mol. Genet.* 2012; 21: 2793–2806.
621 31. Vancheri C, Cottin V, Kreuter M, Hilberg O. IPF, comorbidities and management
622 implications. *Sarcoidosis Vasc. Diffuse Lung Dis. Off. J. WASOG World Assoc. Sarcoidosis*
623 *Granulomatous Disord.* 2015; 32 Suppl 1: 17–23.
624 32. Sheng G, Chen P, Wei Y, Yue H, Chu J, Zhao J, Wang Y, Zhang W, Zhang H-L. Viral
625 Infection Increases the Risk of Idiopathic Pulmonary Fibrosis: A Meta-Analysis. *Chest* 2020;
626 157: 1175–1187.
627 33. Al-Qahtani AA, Murugaiah V, Bashir HA, Pathan AA, Abozaid SM, Makarov E, Nal-
628 Rogier B, Kishore U, Al-Ahdal MN. Full-length human surfactant protein A inhibits influenza
629 A virus infection of A549 lung epithelial cells: A recombinant form containing neck and
630 lectin domains promotes infectivity. *Immunobiology* 2019; 224: 408–418.
631 34. Lopez-Rodriguez E, Pascual A, Arroyo R, Floros J, Perez-Gil J. Human Pulmonary
632 Surfactant Protein SP-A1 Provides Maximal Efficiency of Lung Interfacial Films. *Biophys. J.*
633 2016; 111: 524–536.
634 35. Atochina-Vasserman EN, Beers MF, Gow AJ. Review: Chemical and structural
635 modifications of pulmonary collectins and their functional consequences. *Innate Immun.*
636 2010; 16: 175–182.
637 36. Gardai SJ, Xiao Y-Q, Dickinson M, Nick JA, Voelker DR, Greene KE, Henson PM.
638 By binding SIRPalpha or calreticulin/CD91, lung collectins act as dual function surveillance
639 molecules to suppress or enhance inflammation. *Cell* 2003; 115: 13–23.
640 37. Mulugeta S, Nguyen V, Russo SJ, Muniswamy M, Beers MF. A surfactant protein C
641 precursor protein BRICHOS domain mutation causes endoplasmic reticulum stress,
642 proteasome dysfunction, and caspase 3 activation. *Am J Respir Cell Mol Biol* 2005; 32: 521–
643 530.
644 38. Maitra M, Wang Y, Gerard RD, Mendelson CR, Garcia CK. Surfactant protein A2
645 mutations associated with pulmonary fibrosis lead to protein instability and endoplasmic
646 reticulum stress. *J Biol Chem* 2010; 285: 22103–22113.
647 39. Sano R, Reed JC. ER stress-induced cell death mechanisms. *Biochim. Biophys. Acta*
648 *BBA - Mol. Cell Res.* 2013; 1833: 3460–3470.
649 40. JafariNezhad A, YektaKooshali MH. Lung cancer in idiopathic pulmonary fibrosis: A
650 systematic review and meta-analysis. *PloS One* 2018; 13: e0202360.
651 41. Tzouvelekis A, Spagnolo P, Bonella F, Vancheri C, Tzilas V, Crestani B, Kreuter M,
652 Bouros D. Patients with IPF and lung cancer: diagnosis and management. *Lancet Respir. Med.*
653 2018; 6: 86–88.
654 42. Zhu B, Ferry CH, Markell LK, Blazanin N, Glick AB, Gonzalez FJ, Peters JM. The
655 nuclear receptor peroxisome proliferator-activated receptor-β/δ (PPARβ/δ) promotes
656 oncogene-induced cellular senescence through repression of endoplasmic reticulum stress. *J.*
657 *Biol. Chem.* 2014; 289: 20102–20119.
658 43. Lafont E. Stress Management: Death Receptor Signalling and Cross-Talks with the

659 Unfolded Protein Response in Cancer. *Cancers* 2020; 12.
660 44. Mitsuhashi A, Goto H, Kuramoto T, Tabata S, Yukishige S, Abe S, Hanibuchi M,
661 Kakiuchi S, Saijo A, Aono Y, Uehara H, Yano S, Ledford JG, Sone S, Nishioka Y. Surfactant
662 protein A suppresses lung cancer progression by regulating the polarization of tumor-
663 associated macrophages. *Am. J. Pathol.* 2013; 182: 1843–1853.
664 45. Willems CHMP, Zimmermann LJI, Kloosterboer N, Kramer BW, van Iwaarden JF.
665 Surfactant protein A binds TGF-β1 with high affinity and stimulates the TGF-β pathway.
666 *Innate Immun.* 2014; 20: 192–199.
667 46. Willems CHMP, Zimmermann LJI, Langen RMR, van den Bosch MJA, Kloosterboer
668 N, Kramer BW, van Iwaarden JF. Surfactant protein A influences reepithelialization in an
669 alveolocapillary model system. *Lung* 2012; 190: 661–669.
670 47. Saito A, Horie M, Micke P, Nagase T. The Role of TGF-β Signaling in Lung Cancer
671 Associated with Idiopathic Pulmonary Fibrosis. *Int. J. Mol. Sci.* 2018; 19.
672

673 **FIGURES**

674

675 **Figure 1: SP-A1/SP-A2 mutation localization, SP-A1/SP-A2 conservation and interspecies**
676 **conservation of the SP-A1 and SP-A2 amino-acid impacted by the mutations**

677 Reported *SFTPA1* and *SFTPA2* mutations are respectively in green and red (see also **Table 1**
678 and **Supplemental Table 1**). The mutations identified in this study are in bold. The aminoacids
679 of the highly homologous carbohydrate recognition domain of SP-A1, SP-A2 and of 7 species
680 were aligned (*Homo sapiens, Pan troglodytes, Canis lupus, Bos taurus, Mus musculus, Rattus*
681 *norvegicus, Gallus gallus*). * aminoacids invariant between SP-A1 and SP-A2 and throughout
682 evolution.

683

684 **Figure 2: Genealogical trees of the 14 included families with *SFTPA1* or *SFTPA2***
685 **mutations**

686 The arrows indicate the probands. Individuals with interstitial lung disease (ILD) are indicated
687 by a black symbol. Individuals with lung cancer are indicated by a checkerboard symbol.
688 Symbols that are half black and half checkerboard indicate individuals with ILD and lung
689 cancer. Light grey symbols represent undefined chronic lung diseases reported by the family.
690 Each individual is identified by his generation (Roman numbers) and its number (Arabic

number). +/- indicates the presence of a *SFTPA1* or *SFTPA2* mutation in the heterozygous state; -/- indicates an absence of *SFTPA1* or *SFTPA2* mutation.

Figure 3: Effects of *SFTPA1* and *SFTPA2* variants on protein production and secretion in HEK293T cells

A. Western blot. A 35-37-kDa molecular species consistent with the HA-tagged SP-A1 or FLAG-tagged SP-A2 proteins is observed in lysates of cells expressing the wild-type (WT) or mutant SP-A isoforms. A similar 35-37-kDa molecular species is also observed in the medium of cells expressing the WT proteins, but in the medium of the cells transfected with the mutant constructs of SP-A1 or SP-A2, an absence of secretion is observed. α-tubulin production was used as a loading control of protein expression and LAMC1 secretion was used as a loading control of protein secretion.

B. Normalized protein expression based on densitometry from western blot experiments. Charts show HA-tagged SP-A1 and FLAG-tagged SP-A2 protein expression normalized by α-tubulin (cell lysate expression) and LAMC1 (cell medium expression). Data are means of densitometry ratio (Image J software) from 3 or more independent western blot experiments. Error bars show standard error of the mean.

NT: not transfected; WT: wild type; EV: empty vector; *, $p<0.05$

Figure 4: Lung histopathological features and SP-A expression on lung tissue from patients heterozygous for a *SFTPA1* or *SFTPA2* mutation

Histological examination was performed on lung tissues from 4 index patients (family 5, III.1; family 3, II.1; family 14, II.1; and family 9, II.1) and compared to an adult lung tissue sample. Haematoxylin and eosin (HE) staining was performed and SP-A expression was analyzed by immunohistochemistry. In all the patients, the HE staining (left column, x25) confirmed the

ILD, with however various patterns described according to the 2018 ATS/ERS/JRS/ALAT guideline [3].

Tissue from patient F5, III.1 showed diffuse fibrosis without fibroblastic foci delimitating some unaffected lung parenchyma associated with prominent inflammation with lymphoid nodules consistent with indeterminate for UIP.

Tissue from patient F3, II.1 showed diffuse fibrosis with very few unaffected lung parenchyma with microscopic honey combing, inflammation away from areas of honeycombing, bronchiolar metaplasia and osseous metaplasia consistent with indeterminate for UIP.

Tissue from patient F9, II.1 showed diffuse uniform fibrosis without subpleural or paraseptal predominance and without fibroblast foci and was classified as indeterminate for UIP.

Tissue from patient F14, II.1 showed diffuse fibrosis with microscopic honeycombing and without persistence of unaffected lung. Prominent pneumocyte hyperplasia was observed. This case was classified as indeterminate for UIP.

SP-A immunostaining (right column, x25) documented SP-A expression by hyperplastic pneumocytes without a continuous secreted SP-A linear layer observed in the control individual (insert X200).

Figure 5: High-resolution computed tomography (CT) scans of the probands carrying the *SFTPA1 / SFTPA2* mutations.

The CT-scans were separated into A: predominant ground glass opacities with or without signs of fibrosis; B: lung fibrosis without predominant ground-glass opacities. Other features of ILD were basal thickened septa, micronodules, reticulations, consolidations, honeycombing, and bronchiectasis. The proposed classification is provided for each patient in **Table 2** and **Supplemental Table 1** and includes usual interstitial pneumonia (UIP), non-specific interstitial

pneumonia (NSIP), pleuroparenchymal fibroelastosis (PPFE), hypersensitivity pneumonitis (HSP), indeterminate pattern

Figure 6: Kaplan-Meier survival curves with or without transplantation; and post-transplantation.

A: Survival rate without (n=17) and with (n=11) transplantation of the 28 individuals with identified *SFTPA1/A2* mutations. The age at diagnosis / age at screening was used as a starting point. **B**. Survival rate after transplantation of the 11 patients with identified *SFTPA1/A2* mutations. **A and B**: As all the patients did not complete the same duration of follow-up, the end of follow-up of each patient is represented by a point.

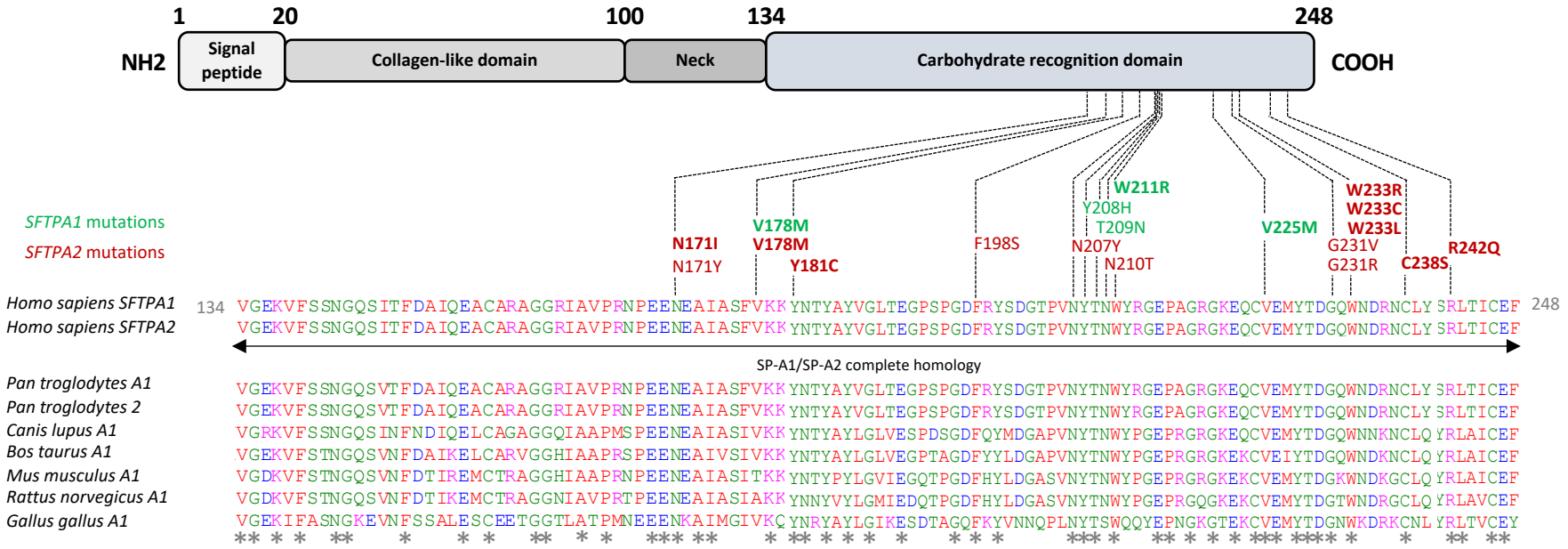
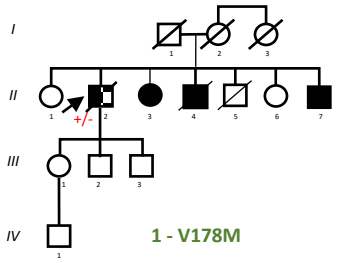


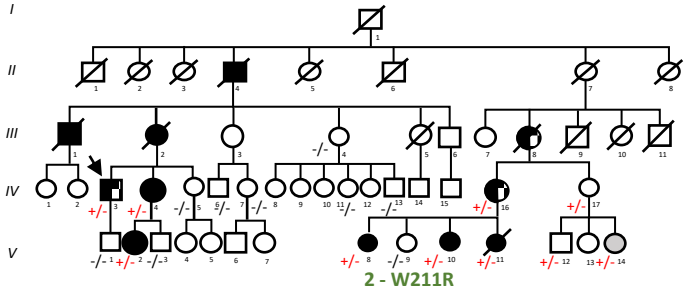
Figure 1: SP-A1/SP-A2 mutation localization, SP-A1/SP-A2 conservation and interspecies conservation of the SP-A1 and SP-A2 amino-acid impacted by the mutations

Reported *SFTPA1* and *SFTPA2* mutations are respectively in green and red (see also **Table 1** and **Supplemental Table 1**). The mutations identified in this study are in bold. The aminoacids of the highly homologous carbohydrate recognition domain of SP-A1, SP-A2 and of 7 species were aligned (*Homo sapiens, Pan troglodytes, Canis lupus, Bos taurus, Mus musculus, Rattus norvegicus, Gallus gallus*). * aminoacids invariant between SP-A1 and SP-A2 and throughout evolution.

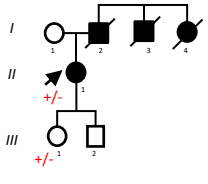
SFTPA1 mutations



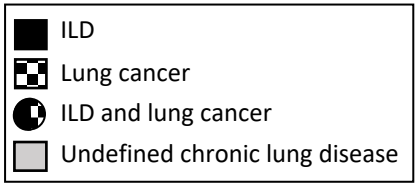
1 - V178M



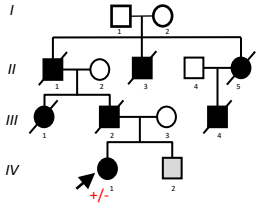
2 - W211R



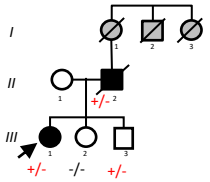
3 - V225M



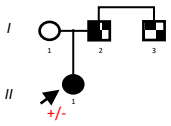
SFTPA2 mutations



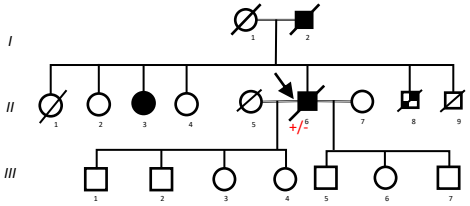
4 - N171I



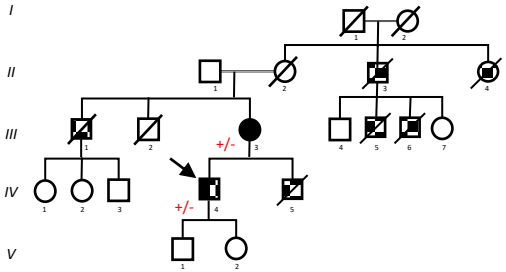
5 - V178M



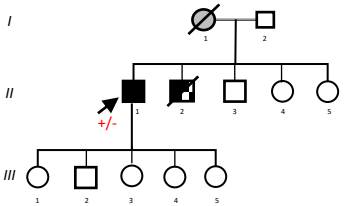
6 - V178M



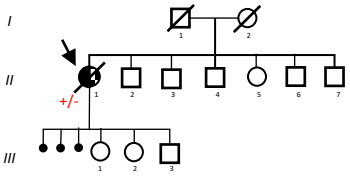
7 - Y181C



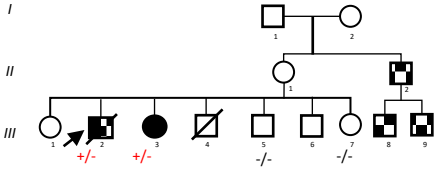
8 - W233R



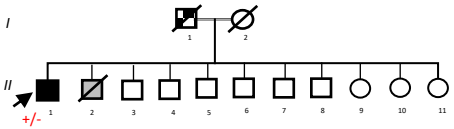
9 - W233L



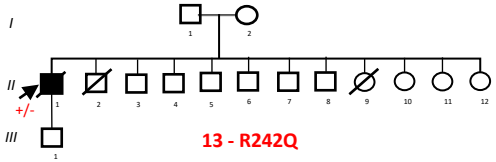
10 - W233C



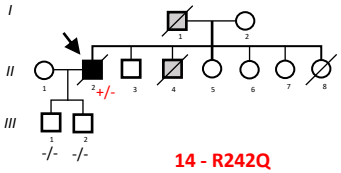
11 - C238S



12 - C238S



13 - R242Q



14 - R242Q

Figure 2: Genealogical trees of the 14 included families with *SFTPA1* or *SFTPA2* mutations

The arrows indicate the probands. Individuals with interstitial lung disease (ILD) are indicated by a black symbol. Individuals with lung cancer are indicated by a checkerboard symbol. Symbols that are half black and half checkerboard indicate individuals with ILD and lung cancer. Light grey symbols represent undefined chronic lung diseases reported by the family. Each individual is identified by his generation (Roman numbers) and its number (Arabic number). +/- indicates the presence of a *SFTPA1* or *SFTPA2* mutation in the heterozygous state; -/- indicates an absence of *SFTPA1* or *SFTPA2* mutation.

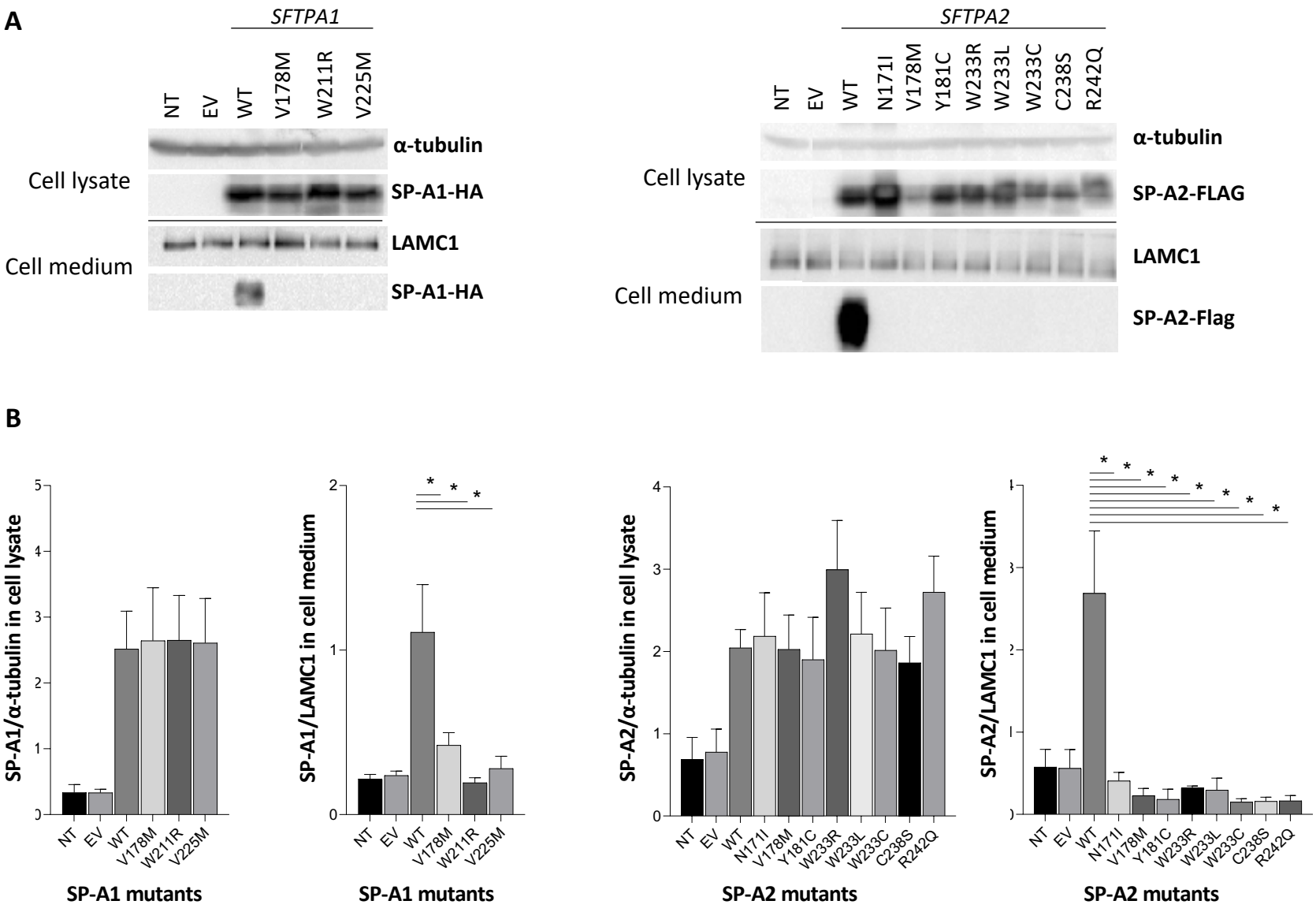

A
SFTPA1
NT
EV
WT
V178M
W211R
V225M
Cell lysate
α-tubulin
SP-A1-HA
Cell medium
LAMC1
SP-A1-HA
SFTPA2
NT
EV
WT
N171I
V178M
Y181C
W233R
W233L
W233C
C238S
R242Q
Cell lysate
α-tubulin
SP-A2-FLAG
Cell medium
LAMC1
SP-A2-Flag
B
SP-A1/α-tubulin in cell lysate
SP-A1 mutants
SP-A1/LAMC1 in cell medium
SP-A1 mutants
SP-A2/α-tubulin in cell lysate
SP-A2 mutants
SP-A2/LAMC1 in cell medium
SP-A2 mutants

Figure 3: Effects of *SFTPA1* and *SFTPA2* variants on protein production and secretion in HEK293T cells

A. Western blot. A 35-37-kDa molecular species consistent with the HA-tagged SP-A1 or FLAG-tagged SP-A2 proteins is observed in lysates of cells expressing the wild-type (WT) or mutant SP-A isoforms. A similar 35-37-kDa molecular species is also observed in the medium of cells expressing the WT proteins, but in the medium of the cells transfected with the mutant constructs of SP-A1 or SP-A2, an absence of secretion is observed. α-tubulin production was used as a loading control of protein expression and LAMC1 secretion was used as a loading control of protein secretion.

B. Normalized protein expression based on densitometry from western blot experiments. Charts show HA-tagged SP-A1 and FLAG-tagged SP-A2 protein expression normalized by α-tubulin (cell lysate expression) and LAMC1 (cell medium expression). Data are means of densitometry ratio (Image J software) from 3 or more independent western blot experiments. Error bars show standard error of the mean.

NT: not transfected; WT: wild type; EV: empty vector; *, $p<0.05$

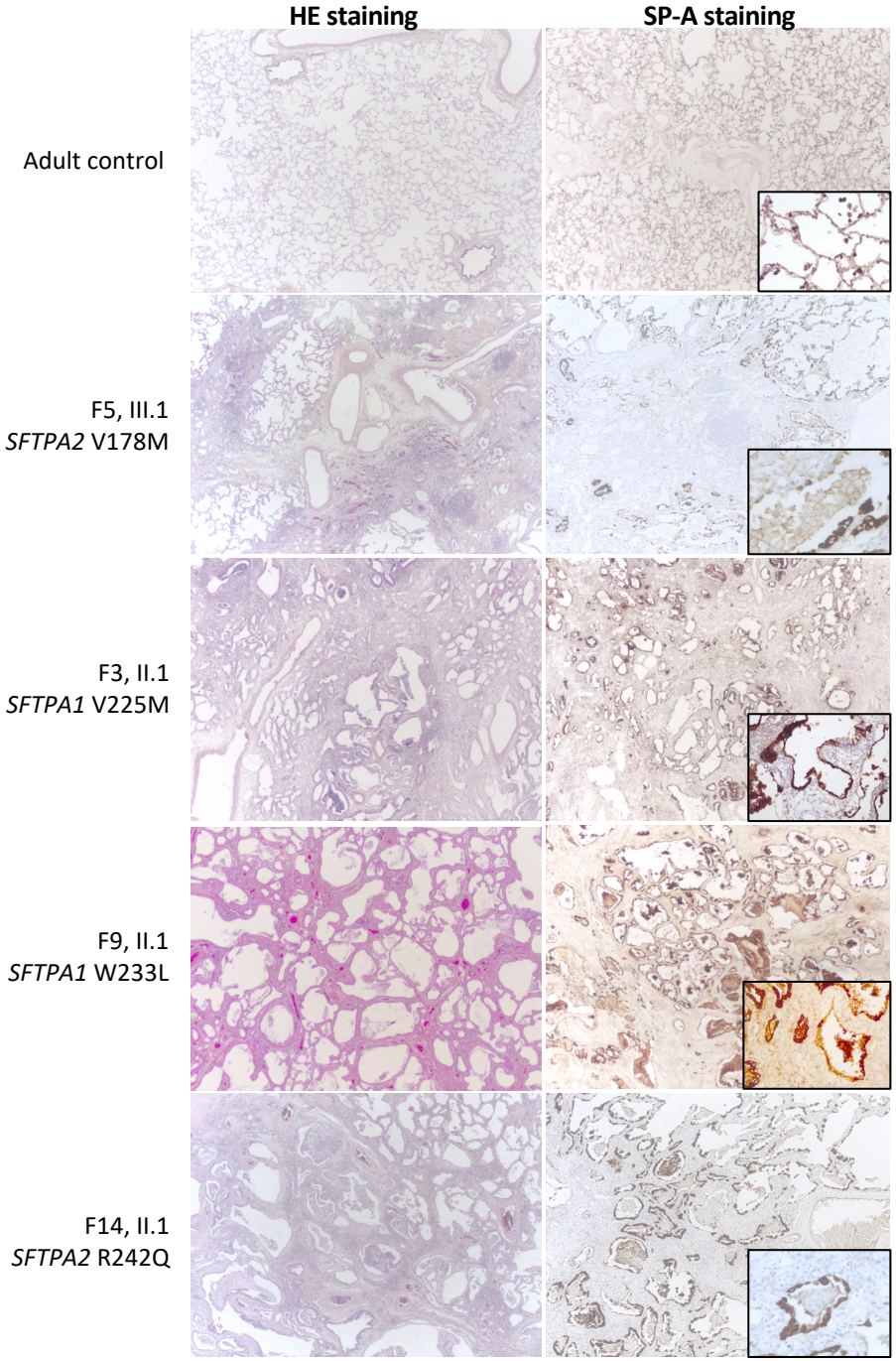

HE staining
SP-A staining
Adult control
F5, III.1
SFTPA2 V178M
F3, II.1
SFTPA1 V225M
F9, II.1
SFTPA1 W233L
F14, II.1
SFTPA2 R242Q

Figure 4: Lung histopathological features and SP-A expression on lung tissue from patients heterozygous for a *SFTPA1* or *SFTPA2* mutation

Histological examination was performed on lung tissues from 4 index patients (family 5, III.1; family 3, II.1; family 14, II.1; and family 9, II.1) and compared to an adult lung tissue sample. Haematoxylin and eosin (HE) staining was performed and SP-A expression was analyzed by immunohistochemistry. In all the patients, the HE staining (left column, x25) confirmed the ILD, with however various patterns described according to the 2018 ATS/ERS/JRS/ALAT guideline [3].

Tissue from patient F5, III.1 shows diffuse fibrosis without fibroblastic foci delimitating some unaffected lung parenchyma associated with prominent inflammation with lymphoid nodules consistent with indeterminate for UIP.

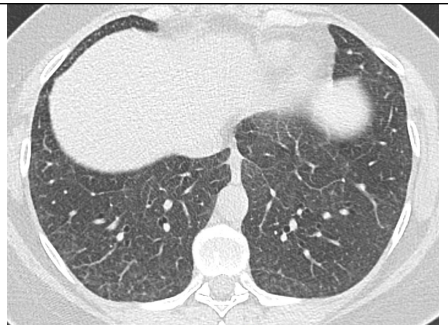
Tissue from patient F3, II.1 shows diffuse fibrosis with very few unaffected lung parenchyma with microscopic honey combing, inflammation away from areas of honeycombing, bronchiolar metaplasia and osseous metaplasia consistent with indeterminate for UIP.

Tissue from patient F9, II.1 shows diffuse uniform fibrosis without subpleural or paraseptal predominance and without fibroblast foci and was classified as indeterminate for UIP.

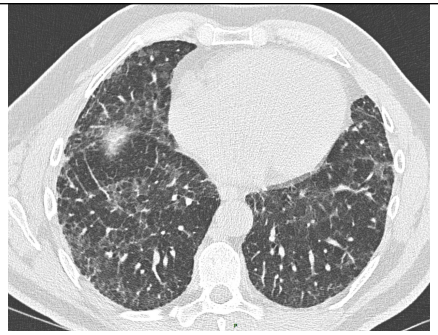
Tissue from patient F14, II.1 shows diffuse fibrosis with microscopic honeycombing and without persistence of unaffected lung. Prominent pneumocyte hyperplasia was observed. This case was classified as indeterminate for UIP.

SP-A immunostaining (right column, x25) documented SP-A expression by hyperplastic pneumocytes without a continuous secreted SP-A linear layer observed in the control individual (insert X200).

A-Predominant GGO (with or without fibrosis)

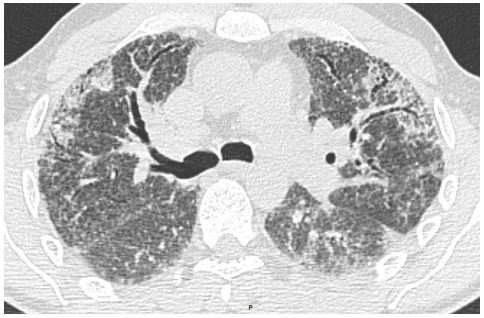


F6, II.1, 34 yo, *SFTPA2* V178M

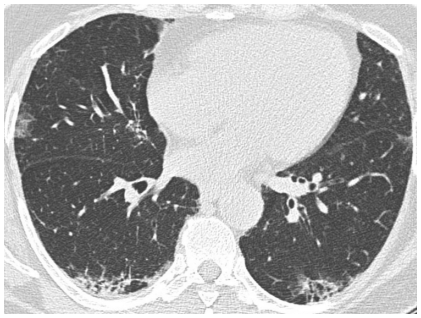


F2, IV.3, 45 yo, *SFTPA1* W211R

F8, IV.4, 34 yo, *SFTPA2* W233R



F9, II.1, 45 yo, *SFTPA2* W233L

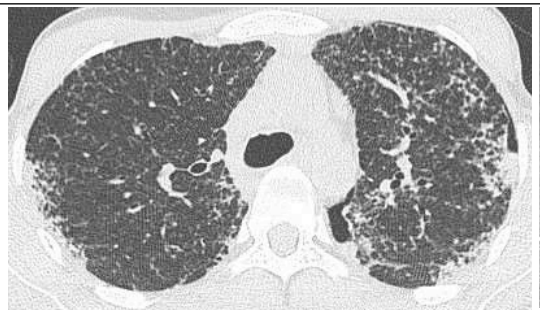


F10, II.1, 50 yo, *SFTPA2* W233C

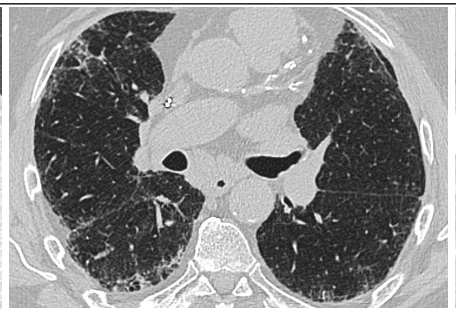
F12, II.1, 26 yo, *SFTPA2* C238S

F14, II.1, 51 yo, *SFTPA2* R242Q

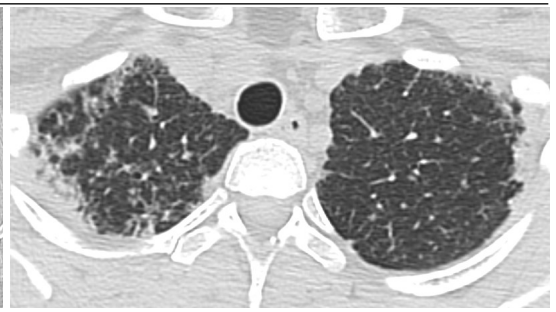
B- Lung fibrosis without predominant GGO



F4, IV.1, 32 yo, *SFTPA2* N171I



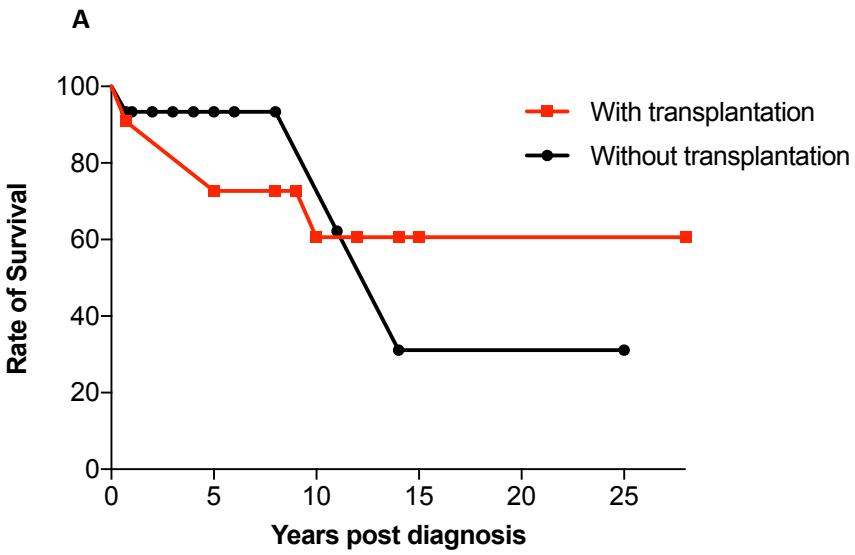
F1, II.2, 64 yo, *SFTPA1* V178M



F5, III.1, 36 yo, *SFTPA2* V178M

Figure 5: High-resolution computed tomography (CT) scans of the probands carrying the *SFTPA1* / *SFTPA2* mutations.

The CT-scans were separated into A: predominant ground glass opacities with or without signs of fibrosis; B: lung fibrosis without predominant ground-glass opacities. Other features of ILD were basal thickened septa, micronodules, reticulations, consolidations, honeycombing, and bronchiectasis. The proposed classification is provided for each patient in **Table 2** and **Supplemental Table 1** and includes usual interstitial pneumonia (UIP), non-specific interstitial pneumonia (NSIP), pleuroparenchymal fibroelastosis (PPFE), hypersensitivity pneumonitis (HSP), indeterminate pattern.


A
100
80
60
40
20
0
Rate of Survival
0
5
10
15
20
25
Years post diagnosis
With transplantation
Without transplantation

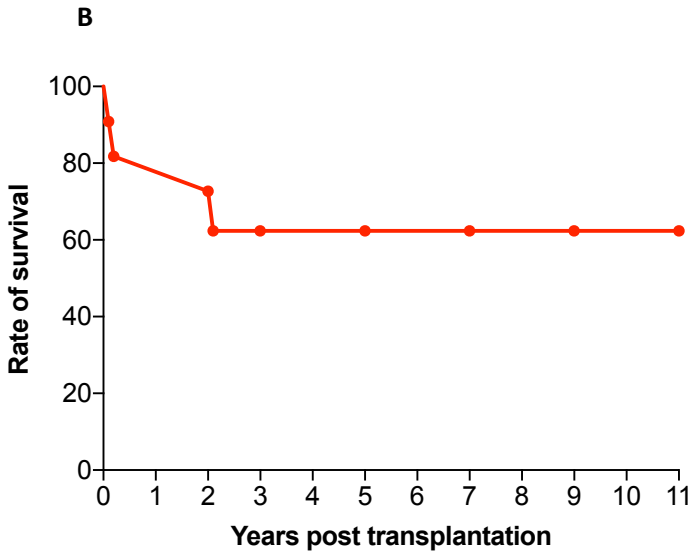

B
100
80
60
40
20
0
Rate of survival
0
1
2
3
4
5
6
7
8
9
10
11
Years post transplantation

Figure 6: Kaplan-Meier survival curves with or without transplantation and post-transplantation.

A. Survival rate without (n=17) and with (n=11) transplantation of the 28 individuals with identified *SFTPA1/A2* mutations. The age at diagnosis / age at screening was used as a starting point. **B.** Survival rate after transplantation of the 11 patients with identified *SFTPA1/A2* mutations. **A and B**: As all the patients did not complete the same duration of follow-up, the end of follow-up of each patient is represented by a point.