

Functional characterization and phenotypic spectrum of three recurrent disease-causing deep intronic variants of the CFTR gene

A. Bergougnoux, Karine Délétang, A. Pommier, J. Varilh, F. Houriez, J.P. Altieri, Michel Koenig, C. Férec, M. Claustres, G. Lalau, et al.

▶ **To cite this version:**

A. Bergougnoux, Karine Délétang, A. Pommier, J. Varilh, F. Houriez, et al.. Functional characterization and phenotypic spectrum of three recurrent disease-causing deep intronic variants of the CFTR gene. Journal of Cystic Fibrosis, 2019, 18 (4), pp.468-475. 10.1016/j.jcf.2018.10.012 . hal-02400710

HAL Id: hal-02400710
https://hal.science/hal-02400710

Submitted on 25 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S1569199318308968
Manuscript_8f53cf65495283b4b437f1cb473ea031

Functional characterization and phenotypic spectrum of three recurrent disease-causing deep intronic variants of the *CFTR* gene

Bergougnoux A.[a,b], Délétang K[b], Pommier A.[b], Varilh J.[b], Houriez F.[c], Altieri JP.[a], Koenig M.[a,b], Férec C.[d,e,f], Claustres M.[b], Lalau G.[g], Bienvenu T.[c], Audrézet MP.[f], Pagin A.[g], Girodon E.[c], Raynal C.[a] and Taulan-Cadars M.[b].

[a]CHU de Montpellier, Laboratoire de Génétique Moléculaire, Montpellier, France, [b]Université de Montpellier, Laboratoire de Génétique de Maladies Rares EA7402, Montpellier, France, [c]AP-HP, HUPC, Hôpital Cochin, Laboratoire de Génétique et Biologie Moléculaires, Paris, France, [d]Inserm, UMR1078 Génétique, Génomique Fonctionnelle et Biotechnologies, [e]Univ Brest, EFS, IBSAM, Brest, France, [f]CHU de Brest, Laboratoire de Génétique Moléculaire, Brest, France, [g]Service de Toxicologie et Génopathies, Institut de Biochimie et Biologie Moléculaire, Centre Hospitalier Régional Universitaire, Lille, France.

Keywords: deep intronic variants, CFTR gene, large phenotypic spectrum, pseudo-exon inclusion, splicing alteration

Highlights:

- We functionally characterized three deep intronic variants of the *CFTR* gene.
- We determined their phenotypic spectrum by detailed analysis of clinical data.

© 2018 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

- These variants should be included in the routine strategy for patients' diagnosis and carrier screening.

Running Title: *CFTR* recurrent deep-intronic variants in patients with CF and CFTR-RD

Abstract:

Background: The *CFTR* genotype remains incomplete in 1% of Cystic Fibrosis (CF) cases, because only one or no disease-causing variants is detected after extended analysis. This fraction is probably higher in CFTR-Related Disorders (CFTR-RD). Deep-intronic *CFTR* variants are putative candidates to fill this gap. However, the recurrence, phenotypic spectrum and full molecular characterization of newly reported variants are unknown.

Methods: Minigenes and analysis of *CFTR* transcripts in nasal epithelial cells were used to determine the impact on *CFTR* splicing of intronic variants that we previously identified by next generation sequencing of the whole *CFTR* locus. Phenotypic data were collected in 19 patients with CF and CFTR-RD, in whom one of the deep intronic variants has been detected.

Results: Three deep-intronic variants promoted the inclusion of pseudo-exons (PE) in the *CFTR* transcript, hindering the synthesis of a functional protein. The c.2989-313A>T variant, detected in four patients with CF or CFTR-RD from three different families, led to the inclusion of a 118bp PE. The c.3469-1304C>G variant promoted the inclusion of a 214bp-PE and was identified in five patients with CF from four families. Haplotype analysis confirmed that this variant was associated with one CF chromosome of African origin. The most represented variant in our cohort was the c.3874-4522A>G, detected in 10 patients with various phenotypes, from male infertility to CF with pancreatic insufficiency.

Conclusion: These three deep intronic *CFTR* variants are associated with a large phenotypic spectrum, including typical CF. They should be included in CF diagnostic testing and carrier screening strategies.

INTRODUCTION

Cystic Fibrosis (CF) is an autosomal recessive disorder caused by pathogenic variations of the *CFTR* gene (OMIM #602421) that encodes the Cystic Fibrosis Transmembrane conductance Regulator (CFTR) chloride and bicarbonate channel expressed at the apical membrane of epithelial cells. Symptoms of classic CF include (but are not limited to) diffuse bronchiectasis (DB), positive sputum cultures for CF-associated pathogens (especially *P. aeruginosa*), exocrine pancreatic insufficiency, salt loss syndrome, sweat chloride level higher than 59 mmol/L, and obstructive azoospermia caused by congenital and bilateral absence of vas deferens (CBAVD) [1]. Other phenotypes associated with CFTR dysfunction that do not fulfill the CF diagnostic criteria (i.e., presenting only one of the CF clinical features) are designated as CFTR-Related Disorders (CFTR-RD) [2,3]. Most patients with CF carry two severe *CFTR* variants (i.e., variants that always result in CF when in *trans* with another CF-causing variant), whereas patients with CFTR-RD mostly harbour one severe and one moderate variant (variants that do not result in CF but that may cause CFTR-RD when in *trans* with another CF-causing variant), or two moderate variants [4]. The analysis of the 27 exons, their flanking intronic boundaries and the search of *CFTR* large rearrangements allows the identification of around 99% of alleles associated with CF manifestations [5]. Thus, approximately 1% of patients with CF and probably more patients with CFTR-RD are incompletely genotyped. However, for optimal medical care, personalized therapy [6], and genetic counselling to families, the clinical diagnosis must be confirmed by molecular data.

We recently developed an assay that allowed the complete sequencing of the *CFTR* locus in DNA samples from 18 patients with confirmed CF clinical diagnosis and harbouring an incomplete genotype [7]. This previous study provided the proof that small-scale Next-Generation Sequencing (NGS) approaches are suitable for the identification of new *CFTR* variants in non-coding regions. Indeed, we demonstrated that the c.1680-883A>G mutation, detected in three unrelated patients, leads to the creation of a new donor splice site and the activation of a cryptic acceptor splice site, resulting in the inclusion of an additional 53bp pseudo-exon (PE) and the complete loss of wild type (WT) *CFTR* transcripts [7].

Here, we characterized the impact on *CFTR* gene splicing of four deep intronic variants detected with our small-scale NGS approach [6] using minigenes and mRNA analysis of nasal epithelial cell (NEC) samples. We found that the newly identified c.2989-313A>T and two previously reported variants, c.3469-1304C>G (also named 3600+11.5kbC>G [8]) and c.3874-4522A>G (also named 4005+5727A>G [9]), have a pathogenic effect. We then performed microsatellite marker analysis in the different families to define the haplotypes linked to these variants and to associate them, when possible, with specific ethnic groups. Finally, we collected detailed clinical data to determine the phenotypic spectrum of these three pathogenic *CFTR* variants.

METHODS

Patients' clinical data, blood and NEC sample collection, mRNA and DNA extraction

Table 1 and 2 show the patients' demographic, clinical and genotypic data. The clinical information was extracted retrospectively from the chart. The respiratory/digestive signs and symptoms in the table 2 are based on physician's interpretation of the patient's clinical status. DNA was extracted from lymphocytes of peripheral blood samples from 19 patients with CF

or CFTR-RD and one control individual using standard protocols. NEC samples from patients and control were processed as previously described [10,11]. Total RNA from NEC samples was extracted using the RNeasy Plus Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions.

CFTR Genotyping

For all 19 patients included in this study, all *CFTR* exons, their intronic boundaries and several deep intronic regions where known CF mutations have been reported were sequenced, as previously described [12]. Additionally, Multiplex Ligation Probe Assay (MLPA) and Quantitative Multiplex PCR of Short Fluorescent Fragments (QMPSF) were used for assessing the presence of large rearrangements. The whole *CFTR* locus [7] was sequenced in six additional patients. Targeted Sanger sequencing allowed the detection of one of the four deep intronic variants in 13 patients (Table 1). No difficulty was reported for Sanger sequencing except for the c.2989-313A>T variant, located in a region of high sequence homology with other genomic regions. Long-range PCRs (LR-PCR; amplicon of 4520bp, see Table S1) were performed with a forward primer in exon 18 (LR-PCR-IVS18-F) and a reverse primer in exon 19 (LR-PCR-Ex19-R) of *CFTR*, followed by nested PCR (PCR-Seq-IVS18-F and PCR-Seq-IVS18-R primers).

Nomenclature of *CFTR* sequence variations

Variants were named following the international nomenclature recommended by the Human Genome Variation Society (http://www.hgvs.org/mutnomen/), in accordance with the *CFTR* gene numbering where nucleotide +1 in the cDNA reference sequence (GenBank NM_00492.3) corresponds to the A of the ATG translation initiation codon. For convenience,

the legacy name of previously described mutations was added in brackets at their first occurrence.

A detailed description of the methods is in Supplementary Material.

RESULTS

Pathogenicity prediction of the putative disease-causing variants and confirmation by Sanger sequencing

On the basis of the bioinformatics predictions, we selected four putative disease-causing variants identified by sequencing the whole *CFTR* locus [7]: c.53+3158A>G (intron 1), c.2989-313A>T (intron 18), c.3469-1304C>G (intron 21), and c.3874-4522A>G (intron 23). The *in silico* predictions of their impact on splicing are described in Figure 1A and Suppl. Figure 1A-C. We confirmed, or newly identify in some patients, the presence of these variants by targeted Sanger sequencing. We detected c.53+3158A>G and c.2989-313A>T as a complex allele in four patients from three families who also carried the synonymous variant c.627A>G (A209A); c.3469-1304C>G in five patients from four families; and c.3874-4522A>G in ten patients from eight families (Table 1).

Ex vivo CFTR and *in vitro* minigene transcript analysis

As NEC samples were not always available for all patients, we combined the analysis of endogenous *CFTR* transcripts and minigene assays, as described in the Method section (Supplementary material).

Analysis of *CFTR* transcripts in the NEC sample from patient P1 and minigene experiments indicated that the c.53+3158A>G variant and the *cis*-associated synonymous variant c.627A>G (only NEC analysis) did not have any impact on *CFTR* splicing (data not shown).

Conversely, we confirmed the inclusion of the predicted 118bp-PE for the c.2989-313A>T variant (NEC and minigene analyses) (Figure 1B-E) after the creation of a new donor splice site.

Analysis of *CFTR* transcripts in the NEC sample from patient P6 and minigene experiments confirmed that the c.3469-1304C>G variant promoted the inclusion of a 214bp-PE (Suppl. Figure 2A-D).

Analysis of *CFTR* transcripts in the NEC sample from patients P10 (Suppl. Figure 3A), P12 and P14 (Suppl. Figure 3B) indicated that the c.3874-4522A>G variant produced a major transcript with the inclusion of a 125bp-PE. Sanger sequencing of transcripts from the NEC samples of patients P12 and P14 using two different primer pairs (located in *CFTR* exons 21 and 25, and exons 22 and 24, Table S1) showed the presence of the WT sequence and the 125bp-PE transcript, as well as of background sequences that corresponded to minor aberrant transcripts. Moreover, analysis of *CFTR* transcripts in the NEC samples from patient P13 did not reveal any transcript with a PE. To understand the discrepancies of the results obtained on endogenous *CFTR* transcripts, we analysed by fluorescent PCR the transcripts obtained from WT and mutant minigene constructs based on the pSPL3 vector and transfected in Beas-2B cells. In cells expressing the WT minigene (c.3874-4522A), we detected a major normal splicing pattern that accounted for 85% of transcripts, but no transcript containing the 125bp-PE (Suppl. Figure 3C and 3D). In cell expressing the c.3874-4522A>G variant, we found that the inclusion of the 125bp-PE led to aberrant splicing in 76% of transcripts (Suppl. Figure 3D), corresponding to the use of the predicted *de novo* acceptor splice site and of the strongest cryptic donor site, according to the *in silico* analysis (Suppl. Figure 1C).

For each variant, PE inclusion, confirmed by sequencing (Suppl. Fig. 4), led to a premature stop codon and possibly to the production of a truncated protein. These results confirmed that

the three variants c.2989-313A>T, c.3469-1304C>G, and c.3874-4522A>G are disease-causing mutations.

Genotype-Phenotype correlations

To determine the phenotypic spectrum of these variants, we analysed the clinical data of the 19 patients of our cohort. All patients carried a second variant, classified as CF-causing and confirmed to be in *trans* by segregation analysis in nine of them. The familial segregation was partial for two patients and was not performed for the others. According to the criteria described in the Introduction section, 15 patients had CF (i.e., presence of CF clinical symptoms), 3 patients (P2, P15, P18) had CFTR-RD (i.e., monosymptomatic clinical entity) and one patient (P16) could be considered as Cystic Fibrosis Screen Positive, Inconclusive Diagnosis (CFSPID) because of intermediate sweat test and no CF symptoms.

The phenotype associated with the c.2989-313A>T variant was variable. Patient P2, at the age of 32 years only had chronic bronchitis without bronchiectasis or functional airway alteration (FEV_1: 101%), and patient P4 was initially addressed for male infertility at the age of 33 years. Conversely, patients P1 and P3 presented multiple and severe symptoms of CF, although with late onset (Table 2). These observations suggest that this variant is associated with a large phenotypic spectrum, from CF to CFTR, and should be considered as a potentially CF-causing variant for genetic counselling.

All patients with the c.3469-1304C>G variant (n=5) had symptoms from early childhood, including recurrent airway *P. aeruginosa* infections and/or sinusitis, pancreatic insufficiency and/or other digestive symptoms (Table 2). Therefore, c.3469-1304C>G could be considered a CF-causing mutation associated with classic CF.

The c.3874-4522A>G variant was the most frequent in our group of patients (10 out of 19). In these patients, symptom onset was delayed, and some of them had only few symptoms,

suggesting CFTR-RD (Table 2). For instance, patient P18 was infertile (CBAVD), and his sister (P17) had bronchiectasis and allergic bronchopulmonary aspergillosis with normal pulmonary function. The two oldest patients with the c.3874-4522A>G variant (P10 and P19) underwent major lung surgery (lobectomy and double-lung transplant, respectively). The variant c.1523T>G (p.Phe508Cys), which has been associated with CFTR-RD phenotypes [13], was located in *cis* to the c.3874-4522A>G variant in two patients. The large spectrum of phenotypes associated with this mutation was in agreement with the 125bp-PE inclusion rate that varied among patients carrying the c.3874-4522A>G variant. Therefore, it has to be considered as a potentially CF-causing variant for genetic counselling.

Geographic origin of the intronic variants

The identification of the geographic origin of a *CFTR* variant, especially when it is a potential CF-causing mutation, is crucial for CF carrier screening strategies. To this aim, we centralized and analysed all available data about the geographic origin of patients carrying the identified intronic variants. We also assessed and compared the haplotypes at the *CFTR* locus associated with these variants (Suppl. Figure 5).

Although we did not find a specific origin for the c.2989-313A>T variant, we think that a *CFTR* haplotype could be linked to this variant on the basis of the conservation of the first five studied intragenic microsatellite markers in the three families (P1 to P4 families).

We identified the c.3469-1304C>G variant in one patient with CF from Guinea, one patient with CF from French Caribbean Islands, and three patients with CF from Senegal. Conversely, we have not yet found this mutation in patients of Caucasian origin (unpublished data from the four French laboratories involved in this study). All patients carrying this variant had the same haplotype (P5 to P8 families).

Finally, patients with the c.3874-4522A>G variant were from different countries, from France to Middle-East (Iran), and Asia (Laos). Microsatellite markers analysis in the families of patient P14 and of patients P17 and P18 did not highlight any haplotype conservation, which is consistent with the various origins of this previously reported variant [9].

DISCUSSION

Increasing numbers of intronic sequence insertions as PEs in mature transcripts of various genes have been reported in the literature, and their role in human diseases has been largely demonstrated [14-16]. mRNA analysis could be proposed as a routine procedure to directly identify and characterize these splicing variants [11,17], but various factors limit its use, particularly, sample accessibility, low quantity and stability, and the rate of transcript degradation caused by premature stop codons generated by intronic variants [18]. The value of whole-genome sequencing for the identification of variants in genes involved in rare diseases is now recognized [19,20]. Sequence analysis of the whole *CFTR* locus recently revealed the second disease-causing variant in previously unexplored intronic regions [7,21]. Until now, and to our knowledge, only six deep-intronic disease-causing variants have been identified in the *CFTR* gene [7,21-25]. In patients with moderate forms of CF and CFTR-RD, a residual quantity of full-length *CFTR* transcripts remains detectable [13], and this seems to be correlated with the disease severity [26]. Thus, in these milder phenotypes, deep-intronic variants may be even more frequent.

In this work, we fully characterized three disease-causing variants deeply located in introns of the *CFTR* gene that were previously predicted by *in silico* tools to induce aberrant splicing [7]. The splicing efficiency and specificity depend on the natural splice site strength, and the binding of splicing regulatory proteins that orchestrate exon recognition [27]. Disease-causing deep-intronic variants mostly lead to the inclusion of an intronic portion of

the gene that is recognized as a coding sequence (i.e., PE) due to a favourable gene architecture [14]. This usually results from the creation or the activation of a novel donor or acceptor splice site associated with the use of a pre-existing cryptic acceptor or donor splice site, respectively, ignored by the splicing machinery in the WT sequence. Here, we demonstrated the activation of two donor (c.2989-313A>T and c.3469-1304C>G) and one acceptor (c.3874-4522A>G) splice sites that were efficiently predicted by *in silico* tools. However, our criteria to define cryptic exon recognition were "manually" set (i.e., length of the putative cryptic exon <300bp, predicted binding motifs for splicing regulatory proteins, and pre-existing non-canonical strong acceptor or donor splice sites). We also found that some intronic variants predicted to have a strong impact on splicing (e.g., the c.53+3158A>G variant; Suppl. Figure 1A) did not lead to splicing alterations *ex vivo*. Indeed, other parameters may influence exon recognition, thus making *in silico* predictions less reliable and accurate. These parameters include the presence of branch points [28], gene architecture [29], conserved RNA secondary structures [30], presence of Alu sequences in introns that potentially produce Alu-Alu dsRNA structures in mRNA precursors [31], differential GC content between exons and introns [32], and conflicting effects exerted by splicing regulatory proteins. All the events described above can also be combined [17], leading to the complete absence of full-length transcripts. Therefore, predictions should be taken with caution to avoid the exclusion of some candidate disease-causing variants. Alternatively, non-pathological PE inclusion could occur, as it has been described for the a-tropomyosin gene and the ATM serine/threonine kinase gene [15,33] through recruitment of negative factors that repress the inclusion event [34]. Therefore, prediction tools require some improvements to be routinely used for deep intronic regions analysis. As the prediction of the effect of a splice variant remains challenging, mRNA analysis and *in vitro* splicing studies are crucial to determine the pathogenicity of a variant.

However, the combination of these molecular approaches does not allow to precise determination of the residual rate of full-length *CFTR* transcripts [18] and the phenotypic spectrum associated with the variant. For instance, the complete inclusion of a PE due to the c.3469-1304C>G variant was clearly associated with the most severe phenotypes in the cohort, in spite of the fact that this variant was previously reported in two sisters with divergent phenotypes (one with CF, the other with a less severe phenotype) [8]. Moreover, we did not detect any PE inclusion in the NEC sample from the young patient P13 (c.3874-4522A>G variant). Modifier genes, encoding proteins involved in the defence against pathogens [35] or in splicing regulation [36], and possibly an age-related exhaustion of the splicing machinery could contribute to the splicing differences among patients and to the resulting phenotypic variability. These splicing regulation mechanisms, the constraints related to the nucleotide environment that challenge the design of minigene constructs (e.g., shorter donor site, 35bp, included in the c.2989-313 minigene construct due to a G-tract), and nonsense-mediated decay also could contribute to the discrepancies in the PE inclusion level among experiments. Even if some very mild phenotypes were associated with the c.2989-313A>T and c.3874-4522A>G deep intronic variants, especially in younger patients, no asymptomatic compound heterozygote has been reported since 2015, when their detection was included in the routine molecular *CFTR* strategy in the four French laboratories involved in this study. Their absence in healthy individuals suggests the full penetrance of these deep-intronic variants.

Sequencing data could improve our understanding of genotype/phenotype correlations. Epidemiological data are progressively included in core (dbSNP) [37] and locus-specific databases (e.g., CFTR-France [12]); however, they are still lacking for rare intronic *CFTR* variants. A collection of variants in non-coding regions of the *CFTR* gene could help to assess their potential role as genetic factors that modify the phenotype [38].

Efforts are currently made for developing therapies that target disease-causing *CFTR* variants that also affect splicing, highlighting the need to fully determine the impact of intronic *CFTR* variants. A recent study reported the *in vitro* restoration of normal splicing after CRISPR/Cas9 targeted excision of the variant c.3718-2477C>T (3849+10kbC>T) followed by non-homologous end-joining repair [39]. We developed a different approach that involves the use of blocking oligonucleotides specifically designed to avoid aberrant PE inclusion. Indeed, we previously demonstrated that the use of sequence-specific blocker oligonucleotides successfully restore normal splicing, indicating that personalized strategies are possible for patients with CF harbouring deep intronic disease-causing variants [7].

In conclusion, this work demonstrates that novel disease-causing variants can still be discovered in CF and CFTR, thanks to the tailored use of new molecular diagnosis technologies. Due to their unexpected recurrence, targeted screening of the three identified variants should be included in the routine strategy for diagnosis and carrier screening. Moreover, the c.3469-1304C>G variant might come from a common African ancestor and should be specifically screened in these populations. The detection of exonic variants (c.627A>G and c.1523T>G) in a complex allele with c.2989-313A>T and c.3874-4522A>G, respectively, should also lead to the targeted search of these two deep-intronic variants, prospectively and retrospectively. Finally, their recurrence in patients with CF and CFTR-RD should encourage the development of novel target site blocker oligonucleotide combinations to rescue normal splicing and offer new therapeutic options to these patients.

ACKNOWLEDGMENTS

This work was supported by the French Association *Vaincre la Mucoviscidose*. We thank the clinicians for their precious help: Dr Laurence Weiss, Dr Dominique Hubert, Dr Reem Kanaan, Prof Isabelle Fajac, Dr Nadia Nathan, Prof Isabelle Sermet, Dr Valérie David, Dr

Stéphane Dominique, Dr Anne Munck, Prof Michel Roussez, Dr Raphaël Chiron, Dr Yves Billon, Dr Nathalie Wizla, Dr Caroline Thumerelle, Dr Anne Prevotat, and Dr Lidwine Stervinou-Wemeau.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

REFERENCES

[1] Farrell PM, White TB, Derichs N, Castellani C, Rosenstein BJ. Cystic Fibrosis Diagnostic Challenges over 4 Decades: Historical Perspectives and Lessons Learned. J Pediatr. 2017 Feb;181S:S16-S26. doi: 10.1016/j.jpeds.2016.09.067.

[2] Bombieri C, Claustres M, De Boeck K, Derichs N, Dodge J, Girodon E, et al. Recommendations for the classification of diseases as CFTR-related disorders. J Cyst Fibros 2011;10 Suppl 2:S86-102. doi:10.1016/S1569-1993(11)60014-3.

[3] Farrell PM, White TB, Ren CL, Hempstead SE, Accurso F, Derichs N, et al. Diagnosis of Cystic Fibrosis: Consensus Guidelines from the Cystic Fibrosis Foundation. J Pediatr 2017;181S:S4-S15.e1. doi:10.1016/j.jpeds.2016.09.064.

[4] Castellani C, Cuppens H, Macek M, Cassiman JJ, Kerem E, Durie P, et al. Consensus on the use and interpretation of cystic fibrosis mutation analysis in clinical practice. J Cyst Fibros 2008;7:179–96. doi:10.1016/j.jcf.2008.03.009.

[5] Audrézet MP, Munck A, Scotet V, Claustres M, Roussey M, Delmas D, et al. Comprehensive CFTR gene analysis of the French cystic fibrosis screened newborn cohort: implications for diagnosis, genetic counseling, and mutation-specific therapy. Genet Med 2015;17:108–16. doi:10.1038/gim.2014.113.

[6] De Boeck K, Amaral MD. Progress in therapies for cystic fibrosis. Lancet Respir Med 2016;4:662–74. doi:10.1016/S2213-2600(16)00023-0.

[7] Bonini J, Varilh J, Raynal C, Thèze C, Beyne E, Audrezet M-P, et al. Small-scale high-throughput sequencing-based identification of new therapeutic tools in cystic fibrosis. Genet Med 2015;17:796–806. doi:10.1038/gim.2014.194.

[8] Monnier N, Gout JP, Pin I, Gauthier G, Lunardi J. A novel 3600+11.5 kb C>G homozygous splicing mutation in a black African, consanguineous CF family. J Med Genet 2001;38:E4.

[9] Roth EK, Hirtz S, Duerr J, Wenning D, Eichler I, Seydewitz HH, et al. The K+ channel opener 1-EBIO potentiates residual function of mutant CFTR in rectal biopsies from cystic fibrosis patients. PLoS ONE 2011;6:e24445. doi:10.1371/journal.pone.0024445.

[10] Bergougnoux A, Claustres M, De Sario A. Nasal epithelial cells: a tool to study DNA methylation in airway diseases. Epigenomics 2015;7:119–26. doi:10.2217/epi.14.65.

[11] Costa C, Pruliere-Escabasse V, de Becdelievre A, Gameiro C, Golmard L, Guittard C, et al. A recurrent deep-intronic splicing CF mutation emphasizes the importance of mRNA studies in clinical practice. J Cyst Fibros 2011;10:479–82. doi:10.1016/j.jcf.2011.06.011.

[12] Bareil C, Guittard C, Altieri J-P, Templin C, Claustres M, des Georges M. Comprehensive and rapid genotyping of mutations and haplotypes in congenital bilateral absence of the vas deferens and other cystic fibrosis transmembrane conductance regulator-related disorders. J Mol Diagn 2007;9:582–8. doi:10.2353/jmoldx.2007.070040.

[13] Claustres M, Thèze C, des Georges M, Baux D, Girodon E, Bienvenu T, et al. CFTR-France, a national relational patient database for sharing genetic and phenotypic data associated with rare CFTR variants. Hum Mutat 2017;38:1297–315. doi:10.1002/humu.23276.

[14] Dhir A, Buratti E. Alternative splicing: role of pseudoexons in human disease and potential therapeutic strategies. FEBS J 2010;277:841–55. doi:10.1111/j.1742-4658.2009.07520.x.

[15] Liquori A, Vaché C, Baux D, Blanchet C, Hamel C, Malcolm S, et al. Whole USH2A Gene Sequencing Identifies Several New Deep Intronic Mutations. Hum Mutat 2016;37:184–93. doi:10.1002/humu.22926.

[16] Romano M, Buratti E, Baralle D. Role of pseudoexons and pseudointrons in human cancer. Int J Cell Biol 2013;2013:810572. doi:10.1155/2013/810572.

[17] Ramalho A, Beck S, Penque D, Gonska T, Seydewitz H, Mall M, et al. Transcript analysis of the cystic fibrosis splicing mutation 1525-1G>A shows use of multiple alternative splicing sites and suggests a putative role of exonic splicing enhancers. J Med Genet 2003;40:e88. doi:10.1136/jmg.40.7.e88.

[18] Will K, Dörk T, Stuhrmann M, von der Hardt H, Ellemunter H, Tümmler B, et al. Transcript analysis of CFTR nonsense mutations in lymphocytes and nasal epithelial cells from cystic fibrosis patients. Hum Mutat 1995;5:210–20. doi:10.1002/humu.1380050305.

[19] Wooderchak-Donahue WL, McDonald J, Farrell A, Akay G, Velinder M, Johnson P, et al. Genome sequencing reveals a deep intronic splicing ACVRL1 mutation hotspot in Hereditary Haemorrhagic Telangiectasia. J Med Genet. 2018 Sep 22. doi: 10.1136/jmedgenet-2018-105561.

[20] Bagnall RD, Ingles J, Dinger ME, Cowley MJ, Ross SB, Minoche AE, et al. Whole Genome Sequencing Improves Outcomes of Genetic Testing in Patients With Hypertrophic Cardiomyopathy. J Am Coll Cardiol. 2018 Jul 24;72(4):419-429. doi: 10.1016/j.jacc.2018.04.078.

[21] Lee M, Roos P, Sharma N, Atalar M, Evans TA, Pellicore MJ, et al. Systematic Computational Identification of Variants That Activate Exonic and Intronic Cryptic Splice Sites. Am J Hum Genet 2017;100:751–65. doi:10.1016/j.ajhg.2017.04.001.

[22] Chillón M, Dörk T, Casals T, Giménez J, Fonknechten N, Will K, et al. A novel donor splice site in intron 11 of the CFTR gene, created by mutation 1811+1.6kbA-->G, produces a new exon: high frequency in Spanish cystic fibrosis chromosomes and association with severe phenotype. Am J Hum Genet 1995;56:623–9.

[23] Faà V, Incani F, Meloni A, Corda D, Masala M, Baffico AM, et al. Characterization of a disease-associated mutation affecting a putative splicing regulatory element in intron 6b of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. J Biol Chem 2009;284:30024–31. doi:10.1074/jbc.M109.032623.

[24] Highsmith WE, Burch LH, Zhou Z, Olsen JC, Boat TE, Spock A, et al. A novel mutation in the cystic fibrosis gene in patients with pulmonary disease but normal sweat chloride concentrations. N Engl J Med 1994;331:974–80. doi:10.1056/NEJM199410133311503.

[25] Costantino L, Rusconi D, Soldà G, Seia M, Paracchini V, Porcaro L, et al. Fine characterization of the recurrent c.1584+18672A>G deep-intronic mutation in the cystic fibrosis transmembrane conductance regulator gene. Am J Respir Cell Mol Biol 2013;48:619–25. doi:10.1165/rcmb.2012-0371OC.

[26] Chiba-Falek O, Kerem E, Shoshani T, Aviram M, Augarten A, Bentur L, et al. The molecular basis of disease variability among cystic fibrosis patients carrying the 3849+10 kb C-->T mutation. Genomics 1998;53:276–83. doi:10.1006/geno.1998.5517.

[27] Homolova K, Zavadakova P, Doktor TK, Schroeder LD, Kozich V, Andresen BS. The deep intronic c.903+469T>C mutation in the MTRR gene creates an SF2/ASF binding exonic

splicing enhancer, which leads to pseudoexon activation and causes the cblE type of homocystinuria. Hum Mutat 2010;31:437–44. doi:10.1002/humu.21206.

[28] Mercer TR, Mattick JS. Understanding the regulatory and transcriptional complexity of the genome through structure. Genome Res 2013;23:1081–8. doi:10.1101/gr.156612.113.

[29] Fox-Walsh KL, Hertel KJ. Splice-site pairing is an intrinsically high fidelity process. Proc Natl Acad Sci USA 2009;106:1766–71. doi:10.1073/pnas.0813128106.

[30] Amit M, Donyo M, Hollander D, Goren A, Kim E, Gelfman S, et al. Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. Cell Rep 2012;1:543–56. doi:10.1016/j.celrep.2012.03.013.

[31] Shepard PJ, Hertel KJ. Conserved RNA secondary structures promote alternative splicing. RNA 2008;14:1463–9. doi:10.1261/rna.1069408.

[32] Lev-Maor G, Ram O, Kim E, Sela N, Goren A, Levanon EY, et al. Intronic Alus influence alternative splicing. PLoS Genet 2008;4:e1000204. doi:10.1371/journal.pgen.1000204.

[33] Grellscheid S-N, Smith CWJ. An apparent pseudo-exon acts both as an alternative exon that leads to nonsense-mediated decay and as a zero-length exon. Mol Cell Biol 2006;26:2237–46. doi:10.1128/MCB.26.6.2237-2246.2006.

[34] Coles JL, Hallegger M, Smith CWJ. A nonsense exon in the Tpm1 gene is silenced by hnRNP H and F. RNA 2009;15:33–43. doi:10.1261/rna.1225209.

[35] Viel M, Hubert D, Burgel P-R, Génin E, Honoré I, Martinez B, et al. DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. Clin Respir J 2016;10:777–83. doi:10.1111/crj.12288.

[36] Martone J, Briganti F, Legnini I, Morlando M, Picillo E, Sthandier O, et al. The lack of the Celf2a splicing factor converts a Duchenne genotype into a Becker phenotype. Nat Commun 2016;7:10488. doi:10.1038/ncomms10488.

[37] Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308–11.

[38] Vecchio-Pagán B, Blackman SM, Lee M, Atalar M, Pellicore MJ, Pace RG, et al. Deep resequencing of CFTR in 762 F508del homozygotes reveals clusters of non-coding variants associated with cystic fibrosis disease traits. Hum Genome Var 2016;3:16038. doi:10.1038/hgv.2016.38.

[39] Sanz DJ, Hollywood JA, Scallan MF, Harrison PT. Cas9/gRNA targeted excision of cystic fibrosis-causing deep-intronic splicing mutations restores normal splicing of CFTR mRNA. PLoS One 2017;12. doi:10.1371/journal.pone.0184009.

A

c.2989-313A>T

3331bp

315bp

Exon 18

118pb

Exon 19

Control

5’-**ag**ggacatggatgaaattggaaaacatcattctc**ag**taaactatcgcaagaacaaaaaaccaaacaccgcatattctcactcat**ag**gtgggaattgaacaatgagatcacatggacacagg**a**agg-3’

*

3’ss-1: HSF: 85.18; SSF-like: 84.55; MaxEntScan: 7.7; NNSplice: 0.96; GeneSplicer: 5.67

3’ss-2: HSF: 83.35; SSF-like: 71.63; MaxEntScan: -; NNSplice: -; GeneSplicer: -

5‘ss-1: HSF: 84.58; SSF-like: 71.56; MaxEntScan: 7.84; NNSplice: 0.91; GeneSplicer: 2.77

5’ss-2: HSF: -; SSF-like: -; MaxEntScan: -; NNSplice: -; GeneSplicer: -

Mutant

5’-agGGACATGGATGAAATTGGAAAACATCATTCTC**AG**TAAACTATCGCAAGAACAAAAAACCAAACACCGCATATTCTCACTCAT**AG**GTGGGAATTGAACAATGAGATCACATGGACACAG**g**tagg-3’

*

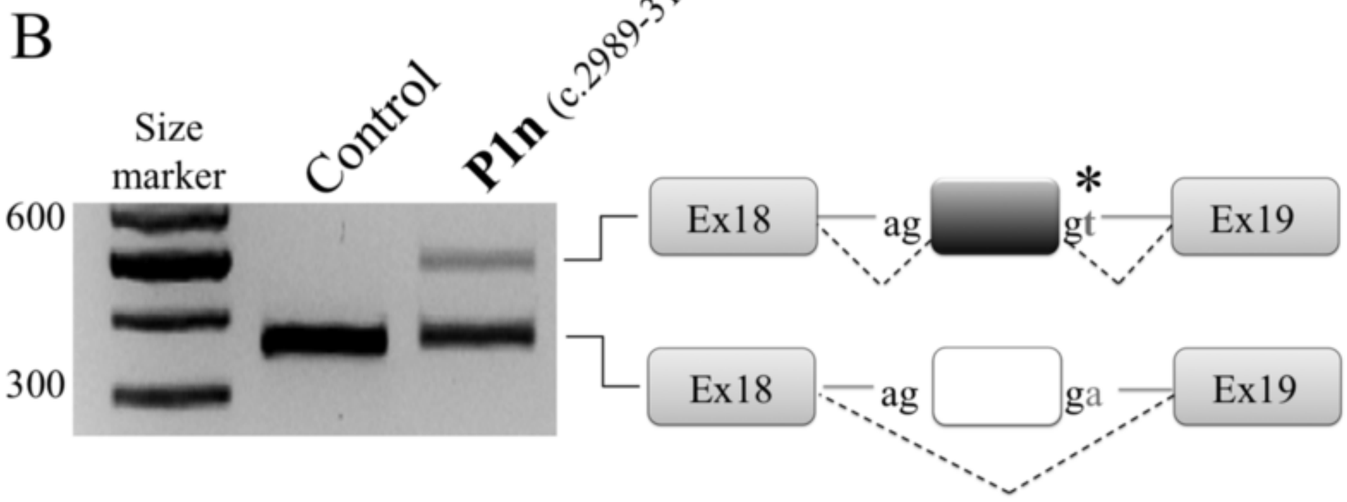
3’ss-1: HSF: 85.18; SSF-like: 84.55; MaxEntScan: 7.7; NNSplice: 0.96; GeneSplicer: 5.67

3’ss-2: HSF: 83.35; SSF-like: 71.63; MaxEntScan: 0.85; NNSplice: -; GeneSplicer: -

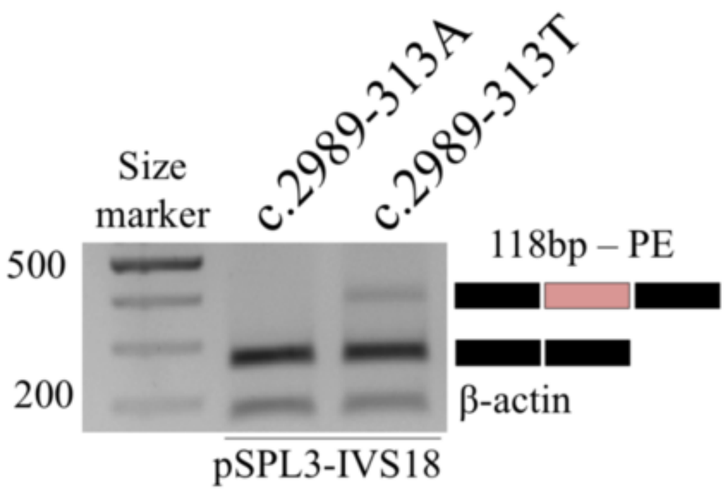
5’ss-1: HSF: 84.58; SSF-like: 71.56; MaxEntScan: 7.84; NNSplice: 0.91; GeneSplicer: 2.77

5’ss-2: HSF: 89.73; SSF-like: 84.90; MaxEntScan: 9.46; NNSplice: 0.93; GeneSplicer: 7.31

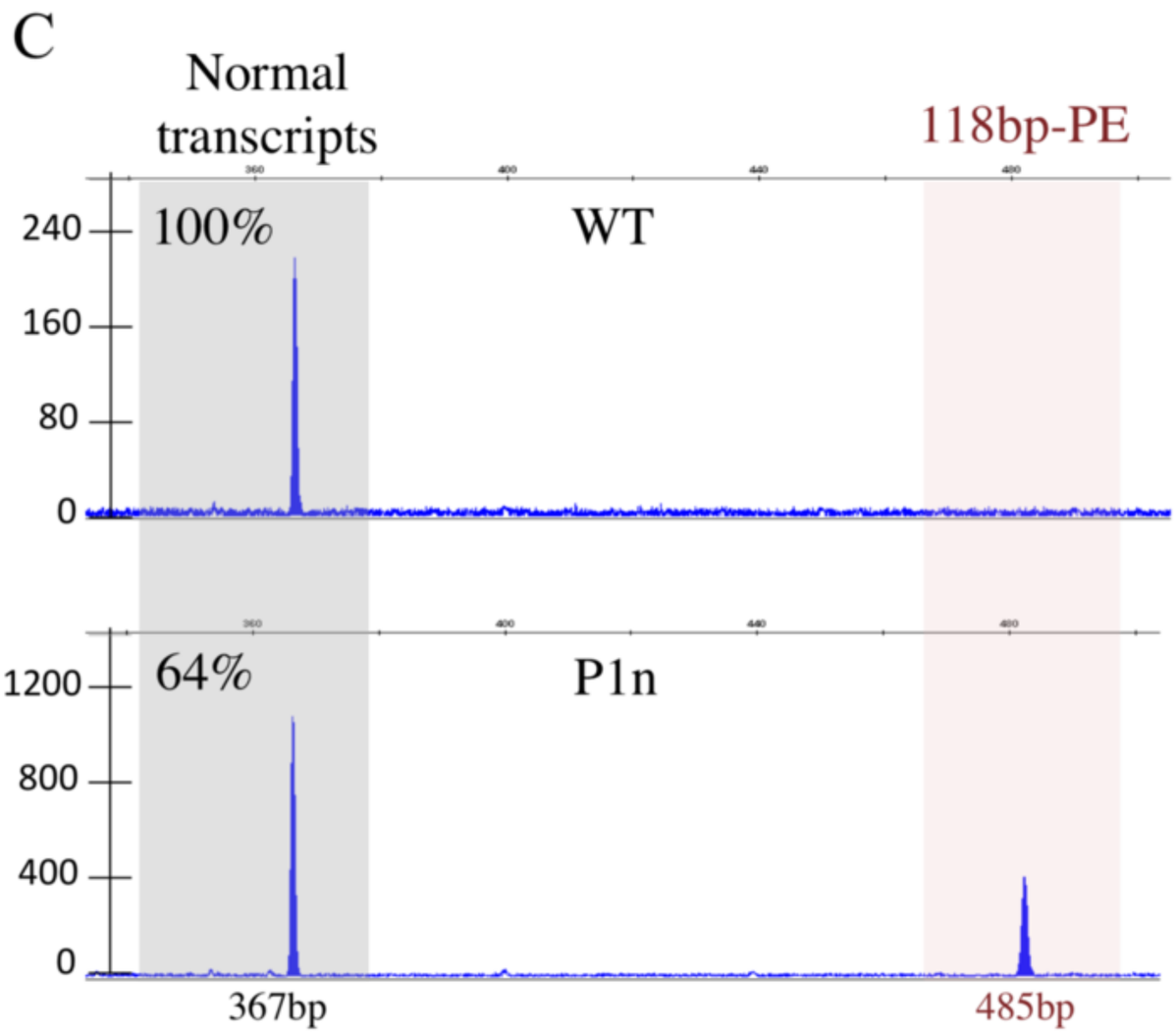
B



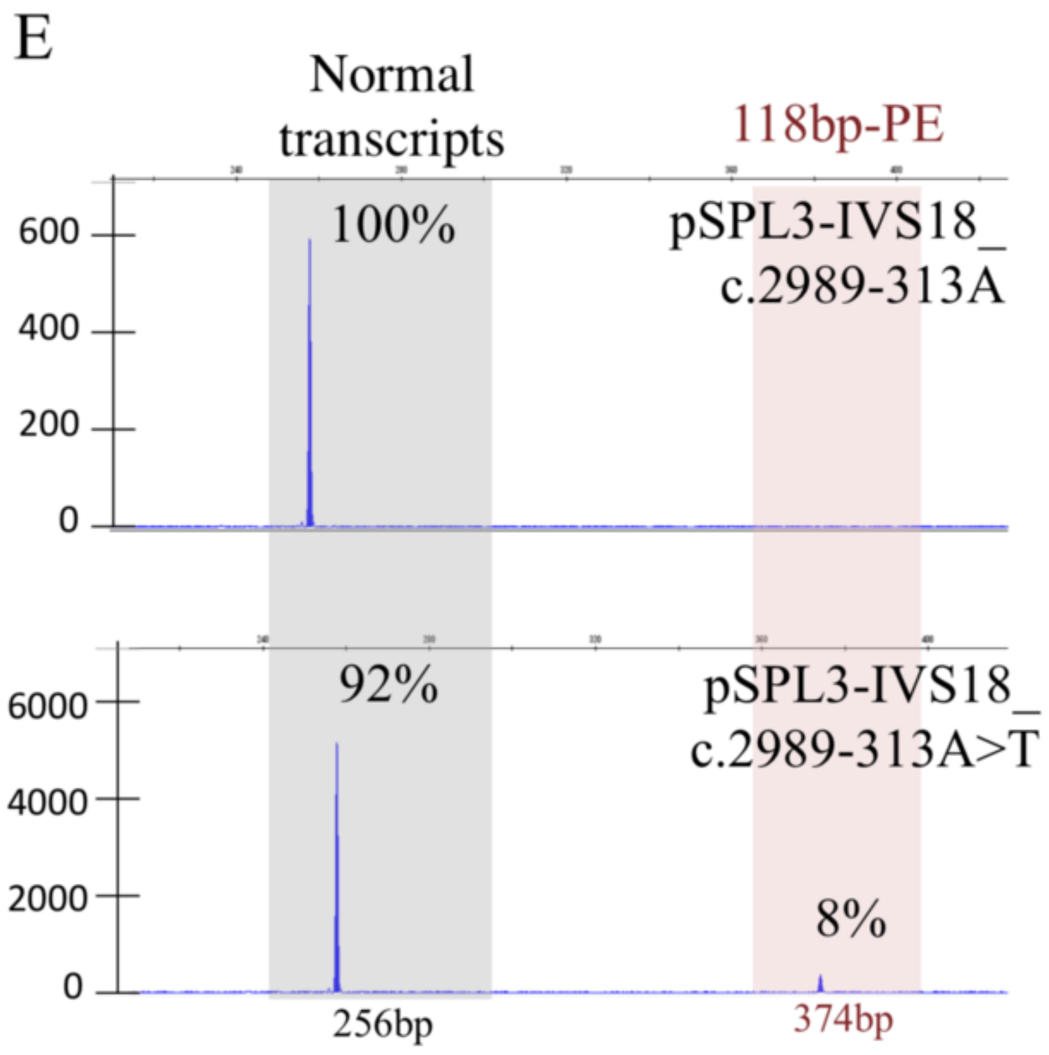
D



C



E



Cases	Genotype	Affected relatives in this study	familial segregation	included in NGS cohort [6]	Paternal origin	Maternal origin	Familial Haplotypes analysis	availability of NEC	Confirmation of PE inclusion in *CFTR* transcripts
P1	c.[53+3158A>G;627A>G;2989-313A>T];[2810dup]	No	Yes	Yes	France	**France (Brittany)**	Yes	Yes (P1n)	Yes (Fig.1)
P2	c.53+3158A>G(;)627A>G(;)2989-313A>T(;)1116+1G>A	Yes (P3's sister)	No	Yes	NC	NC	Yes	No	NA
P3	c.53+3158A>G(;)627A>G(;)2989-313A>T(;)1116+1G>A	Yes (P2's brother)	No	No	NC	NC	Yes	No	NA
P4	c.53+3158A>G(;)627A>G(;)1521_1523del(;)2989-313A>T	No	No	No	NC	NC	Yes	No	NA
P5	c.[3469-1304C>G];[3469-1304C>G]	No	Yes	No	**Guinea Conackry**	**Guinea Conackry**	Yes	No	NA
P6	c.[254G>A];[3469-1304C>G]	No	Yes	Yes	**Guadeloupe Island**	France	Yes	Yes (P6n)	Yes (Suppl Fig.2)
P7	c.[1521_1523del];[3469-1304C>G]	Yes (P8's sister)	Yes	No	**Senegal**	France	Yes	No	NA
P8	c.[1521_1523del];[3469-1304C>G]	Yes (P7's sister)	Yes	No	**Senegal**	France	Yes	No	NA
P9	c.[2989-1698_3469-4460del];[3469-1304C>G]	No	Yes	No	Senegal	**Senegal**	No	No	NA
P10	c.1523T>G(;)1521_1523del(;)3874-4522A>G	No	No	Yes	Spain - Belgium	France	No	Yes (P10n)	Yes (Suppl Fig.3)
P11	c.1521_1523del(;)3874-4522A>G	No	No	Yes	Poland	France (Alsace)	No	No	NA
P12	c.1521_1523del(;)3874-4522A>G	No	No	Yes	France / Vietnam	France (Brittany)	No	Yes (P12n)	Yes (Suppl Fig.3)
P13	c.1798A>G(;)3874-4522A>G	No	No	No	France (Alsace)	France (Alsace), Switzerland	No	Yes (P13n)	No
P14	c.[1521_1523del];[1523T>G;3874-4522A>G]	No	Yes	No	NC	NC	Yes	Yes (P14n)	Yes (Suppl Fig.3)
P15	c.[1585-1G>A];[3874-4522A>G]	Yes (P15's brother)	Yes but partial	No	**France**	Guadeloupe Island	No	No	NA
P16	c.[1585-1G>A];[3874-4522A>G]	Yes (P16's sister)	Yes but partial	No	**France**	Guadeloupe Island	No	No	NA
P17	c.[1521_1523del];[3874-4522A>G]	Yes (P18's sister)	Yes	No	**Laos**	France	Yes	No	NA
P18	c.[1521_1523del];[3874-4522A>G]	Yes (P17's brother)	Yes	No	**Laos**	France	Yes	No	NA
P19	c.1393-1G>A(;)3874-4522A>G	No	No	No	Iran	Iran	No	No	NA

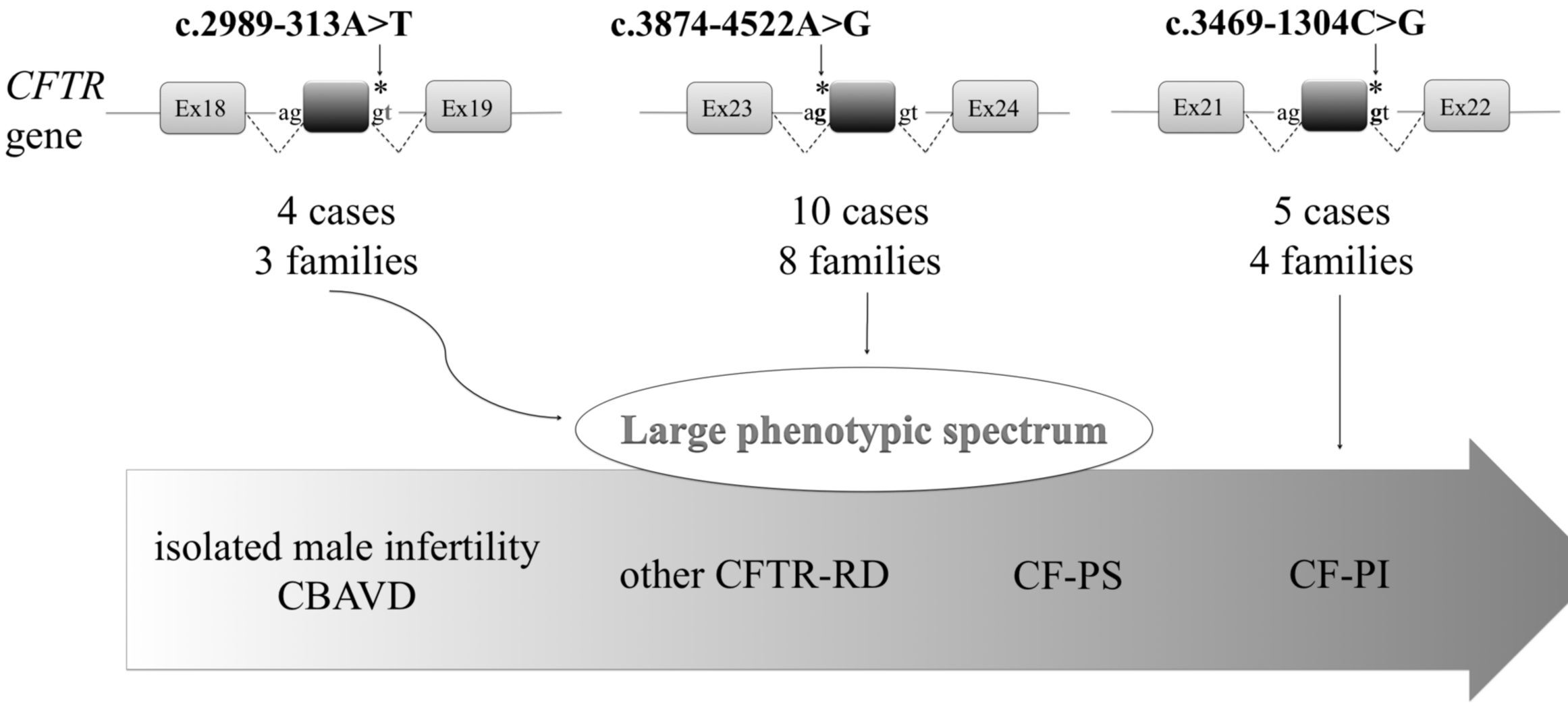
NC: not communicated; ND: Not determined ; Origins in bold letters correspond to the inheritance of deep intronic variants if known (familial segregation)

Table 1. Genotypes, geographical origins, available material and level of molecular analysis from patients included in the study

Cases	Genotype	Gender	DoB	Onset of symptoms (years)	IRT (ng/ml)	Sweat test (mmol/L)	Respiratory symptoms			Digestive manifestations		ENT symptoms	Infertility
							FEV_1 (%)	BD	Other	Pancreatic insuffisiency	Other		
P1	c.[53+3158A>G;627A>G;2989-313A>T] ;[2810dup]	M	1995	14	NA	109 – 110 – 99	49	Yes	Pneumothorax	No	grele occlusion	No	NC
P2	c.53+3158A>G(;)627A>G (;)2989-313A>T(;)1116+1G>A	F	1986	11	NA	130 - 146	101	No	*S. aureus* colonization, Chronic bronchitis,	No		No	NA
P3	c.53+3158A>G(;)627A>G (;)2989-313A>T(;)1116+1G>A	M	1983	14	NA	NC	85	NC	*P. aeruginosa* colonization	No		Polyposis	Yes
P4	c.53+3158A>G(;)627A>G(;)1521_1523del (;)2989-313A>T	M	1968	33	NA	94 – 92 – 108	130	No	Mucoid *P. aeruginosa* and *S. aureus* colonization	No		Polyposis	Yes
P5	c.[3469-1304C>G];[3469-1304C>G]	M	1994	< 1	NA	127	70%	Yes	Mucoid *P. aeruginosa* and *S. aureus* colonization	Yes	liver injury	No	Yes
P6	c.[254G>A];[3469-1304C>G]	F	2007	< 1	NA	105 – 105	NC	No	*Pseudomonas aeruginosa* infection	Yes		NC	NA
P7	c.[1521_1523del];[3469-1304C>G]	F	2009	NBS	170	110 – 112	119	No	*Pseudomonas aeruginosa* infection, bronchial syndrome	Yes		No	NA
P8	c.[1521_1523del];[3469-1304C>G]	F	2011	NBS	181	94 - 97	99	No	*Pseudomonas aeruginosa* colonization	Yes		Yes	NA
P9	c.[2989-1698_3469-4460del]; [3469-1304C>G]	M	2017	1.5	NA	103	NA	No	*Pseudomonas aeruginosa* infection, Bronchial syndrome	Yes	low BMI, hepatic cytolysis	No	ND
P10	c.1523T>G(;)1521_1523del (;)3874-4522A>G	M	1968	childhood	NA	60	29	Yes	*Pseudomonas aeruginosa* colonization, double-lung transplant	No		NC	Yes
P11	c.1521_1523del(;)3874-4522A>G	M	1977	28	NA	NC	103	Yes		Yes		NC	ND
P12	c.1521_1523del(;)3874-4522A>G	M	1979	13	NA	55 – 47	53	Yes		No		Chronic pansinusitis	Yes
P13	c.1798A>G(;)3874-4522A>G	M	2007	4	189	50 – 77 – 21	99	Yes	infant asthma	No	acute pancreatitis	No	ND
P14	c.[1523T>G;3874-4522A>G]; [1521_1523del]	M	1993	17	NA	51 – 60 44 – 49	69	Yes	Suppurative syndrome since childhood	No	very small gallbladder	Yes (Chronic pansinusitis polyposis)	ND
P15	c.[1585-1G>A];[3874-4522A>G]	M	2008	9	positive	52 – 56 – 54	NC	Yes	bronchial hyperreactivity	No		Yes (one sinusitis)	ND
P16	c.[1585-1G>A];[3874-4522A>G]	F	2016	1.5	positive	intermediate	NA	No		No		repeated rhinosinusitis	NA
P17	c.[1521_1523del];[3874-4522A>G]	F	1987	13	NA	78 – 80 87 – 98	98	Yes	*Aspergillus* Bronchopulmonary allergy	No		Yes	NA
P18	c.[1521_1523del];[3874-4522A>G]	M	1992	No symptom at 4	NA	NC	116	No		No		No	Yes
P19	c.1393-1G>A(;)3874-4522A>G	F	1974	9	NA	50-48	59	Yes	Lobectomy, *Aspergillus* abcess	No		Yes	NA

IRT : Immunoreactive Trypsinemia measured at Day 3 of life ; BD : Bronchectasies ; DoB : Date of Birth, NC : not communicated ; NA : not applicable ; ND : Not determined

Table 2. Demographic, clinical and biological data of the 19 patients included in the study.


c.2989-313A>T
c.3874-4522A>G
c.3469-1304C>G
CFTR gene
Ex18
ag
gt
Ex19
Ex23
ag
gt
Ex24
Ex21
ag
gt
Ex22
4 cases
3 families
10 cases
8 families
5 cases
4 families
Large phenotypic spectrum
isolated male infertility
CBAVD
other CFTR-RD
CF-PS
CF-PI