Exon identity influences splicing induced by exonic variants and in silico prediction efficacy

Natacha Martin, Anne Bergougnoux, Nesrine Baatallah, Benoit Chevalier, Jessica Varilh, David Baux, Bruno Costes, Pascale Fanen, Caroline Raynal, Isabelle Sermet-Gaudelus, et al.

Exon identity influences splicing induced by exonic variants and *in silico* prediction efficacy

Natacha Martin[a,1], Anne Bergougnoux[b,c,1], Nesrine Baatallah[d,e], Benoit Chevalier[d,e], Jessica Varilh[c], David Baux[b,c], Bruno Costes[a], Pascale Fanen[a,f], Caroline Raynal[b,c], Isabelle Sermet-Gaudelus[d,e], Emmanuelle Girodon[d,g], Magali Taulan-Cadars[c,*], Alexandre Hinzpeter[d,e,**]

[a] *INSERM, U955, Institut de Recherche Henri Mondor, IMRB, Créteil, France*
[b] *CHU de Montpellier, Laboratoire de Génétique Moléculaire, Montpellier, France*
[c] *Université de Montpellier, Laboratoire de Génétique de Maladies Rares, EA7402 Montpellier, France*
[d] *INSERM U1151, Institut Necker Enfants Malades, INEM, Paris, France*
[e] *Université Paris Descartes, Paris, France*
[f] *Department of Genetics, GH Henri Mondor, APHP, Créteil, France*
[g] *Laboratoire de Génétique et Biologie Moléculaires, Hôpital Cochin, APHP Centre-Université de Paris, Paris, France*

Keywords:
CFTR
Splicing
Minigene
Exon skipping
In silico predictions

A B S T R A C T

Background: Minigenes and *in silico* prediction tools are commonly used to assess the impact on splicing of *CFTR* variants. Exon skipping is often neglected though it could impact the efficacy of targeted therapies. The aim of the study was to identify exon skipping associated with *CFTR* variants and to evaluate *in silico* predictions of seven freely available software.

Methods: *CFTR* basal exon skipping was evaluated on endogenous mRNA extracted from non-CF nasal cells and on two *CFTR* minigene banks. *In silico* tools and minigene systems were used to evaluate the impact of *CFTR* exonic variants on exon skipping.

Results: Data showed that out of 65 *CFTR* variants tested, 26 enhanced exon skipping and that *in silico* prediction efficacy was of 50%-66%. Some *in silico* tools presented predictions with a bias towards the occurrence of splicing events while others presented a bias towards the absence of splicing events (non-detection including true negatives and false negatives). Classification of exons depending on their basal exon skipping level increased prediction rates up to 80%.

Conclusion: This study indicates that taking basal exon skipping into account could orientate the choice of the *in silico* tools to improve prediction rates. It also highlights the need to validate effects using *in vitro* assays or mRNA studies in patients. Eventually, it shows that variant-guided therapy should also target exon skipping associated with variants.

1. Introduction

The *CFTR* gene is extensively studied for diagnostic purposes and over 2000 *CFTR* variants have been identified in Cystic Fibrosis (CF) and CFTR-related disorders (CFTR-RD) [1]. As Next Generation Sequencing (NGS) has been introduced in laboratories for routine diagnosis, the number of variants will probably greatly increase, along with the need of properly addressing them as disease causing or not. While a number of variants have been clearly shown to cause disease, many are still uncharacterized and are referred to as variants of unknown significance (VUS). *CFTR* disease-causing variants have been classified depending on their consequences on protein production, maturation, function, regulation or stability [2]. Around 11% of reported *CFTR* variants are referred to as splicing variants [1] and most of them affect key splicing elements, such as acceptor and donor sites (AS and DS, respectively) or induce

* Corresponding author at: Université de Montpellier, Laboratoire de Génétique de Maladies Rares, 641 avenue du Doyen Gaston Giraud, Montpellier 34093, France.
** Corresponding author at: INSERM U1151, Institut Necker Enfants Malades (INEM), 160 rue de Vaugirard, Paris 75015, France.
E-mail addresses: magali.taulan@inserm.fr (M. Taulan-Cadars), alexandre.hinzpeter@inserm.fr (A. Hinzpeter).
[1] These authors contributed equally to the work and should both be considered first authors.

the inclusion of pseudo-exons in transcripts [3]. If the open reading frame is disrupted, transcripts will be additionally degraded by the nonsense mediated mRNA decay (NMD) pathway [4]. A few studies have shown exon skipping associated to exonic variants [5–9] which can modify splicing regulatory elements (SREs), such as exon splicing enhancers (ESE) and exon splicing silencers (ESS) motifs. This mechanism is not routinely investigated and is probably underestimated.

Alterations in exon splicing can be measured on endogenous mRNAs if available or *in vitro* using minigene systems. Minigenes enable to evaluate the effect of both exonic and intronic variants on exon skipping or the use of cryptic AS or DS. On the other hand, absence of translation of the produced transcripts in minigenes prevents eligibility to NMD. Multiple *in silico* tools have also been developed to predict splicing defects associated to variants. While these tools usually successfully recognize the impact of variants affecting 5' and 3' splice sites, the impact of variants located in branch points (BP), polypyrimidin tracts (ppt-tract) or SREs is harder to assign.

Recently, the knowledge of multiple defects associated to single variants has led to a renewed interest in the era of CF personalized medicine [2,10] because it is of importance to predict efficacy of CFTR modulators and adapt treatments. Increased exon skipping could hinder treatment efficacy by reducing the amount of transcripts targeted by readthrough inducing molecules [9,11] or the amount of protein targeted by CFTR modulators (e.g. correctors or potentiators) [11]. In this context, we evaluated the effect of exonic variants on exon skipping using minigenes. We first compared two *CFTR* minigene banks to endogenous basal exon skipping and alternative splicing in non-CF nasal cells to identify the most relevant system. This minigene was then used to evaluate the effect on splicing of 65 *CFTR* exonic variants and results were used to challenge commonly used *in silico* prediction tools.

2. Material and methods

2.1. Ethical statement

Informed consent for genetic analysis and functional assays was obtained from all subjects and the local ethics committee approved the study (local ethic committee, CPP #2011.12.08ter Sud Méditerranée III; n°ID- RCB: 2011-A01520-41).

2.2. Mutation nomenclature

The recommendations of the Human Genome Variation Society (HGVS, http://www.hgvs.org/mutnomen/) were followed for exon numbering and variant names, using reference sequences NM_000492.3 and NG_016465.3. For readability, traditional names of tested variants were used in the figures.

2.3. Nasal epithelial cell sampling

Nasal epithelial cells from control individuals were sampled and processed as previously described [3]. The analysis of the *CFTR* transcript was performed by fluorescent PCR on cDNA using the Multiplex PCR Kit (Qiagen, Hilden, Germany) and primers listed in Table S1. Fragments size analysis of the fluorescent PCR products was performed using the Genemapper® 4.0 software. Level of basal exon skipping was determined using a relative quantification and the ratio of the alternative peak area (without exon) on the expected full exon inclusion peak area.

The quantitative analysis of *CFTR* transcripts in nasal epithelial cells was performed by qPCR as previously described [12].

2.4. Cell culture

BEAS-2B cells were grown and transfected as previously reported [3,7]. T84 cells were cultivated as previously described and were used as reference for *CFTR* mRNA quantification because high levels of mRNA are detected [13].

2.5. Minigene constructs

pET01 and *pSPL3* minigene constructs containing single WT *CFTR* exons and their flanking intronic sequences have already been published [7,14,15] or were PCR amplified using the primers indicated in Table S2 or in Table S3. Mutagenesis was performed using the QuickChange XL II mutagenesis kit (Agilent, Courtaboeuf, France) following the manufacturer's instructions.

2.6. Hybrid minigene splicing assays

One μg of minigene plasmids were transfected in BEAS-2B cells seeded in six-well plates. After 24 hr, total RNA was purified using Trizol (ThermoFisher Scientific, Courtaboeuf, France) following the manufacturer's instructions and RT-PCR analysis performed as previously described [16].

2.7. In silico *analysis*

In silico analysis was performed using the different freely available software, using default settings.

2.8. Statistical analysis

Mean values are given with standard deviation, 'n' representing the number of repeats. Unpaired t-tests were performed to evaluate the effect of variants compared to WT minigenes. Non-parametric Wilcoxon-Mann-Whitney (*U*-test) was used to define if the minigene system reflects the behavior of the *CFTR* gene meaning to define if the exon skipping obtained for minigene system reflects that quantified in control nasal cells samples with $p \leq 0.05$, $p \leq 0.01$.

3. Results

3.1. CFTR *basal exon skipping levels in human nasal cells and minigenes*

mRNA was obtained from nasal epithelial cells of four healthy volunteers who had no disease-causing mutation as assessed by complete sequencing of the *CFTR* locus with NGS [12]. *CFTR* transcripts were quantified by qRT-PCR revealing similar expression levels (Suppl. Fig. S1). Transcripts were then amplified with overlapping RT-PCRs and fragments analyzed by capillary electrophoresis. Basal exon skipping (>1%) was observed in at least one sample for exons 3, 5, 10, 12, 13, 14, 15 and 22, while alternative transcripts were detected associated to the use of cryptic ASs within exons 8 (c.940_941) and 14 (c.1960_1961/c.2013_2014) (Table 1). Alternative transcripts of exon 14 had been previously reported [5,17]. While not observed in our samples, transcripts containing a truncated exon 22 were previously described associated with the use of a cryptic donor site (c.3518_3519) [5] (Table 1).

These results were compared to data obtained with minigenes of *CFTR* exons 2 to 26 subcloned in either *pET01* or *pSPL3* plasmids. Each minigene contains a single exon and its flanking intronic region (detailed in Tables S2 and S3) corresponding to the *CFTR* reference sequence NG_016465.3. RT-PCR analysis showed that all minigenes produced transcripts containing full length exons, indicating the presence of all core splicing elements. Exons 1 and 27 cannot

Table 1
Basal skipping and alternative transcripts of endogenous *CFTR* and *CFTR* minigenes. Percentage of exon skipping are indicated. Detectable basal exon skipping >1% was considered for comparison between endogenous *CFTR* and minigenes. Alternative transcripts generated by the usage of cryptic acceptor (cAS) or donor (cDS) sites are indicated. *Ref* are results from the literature. Haplotypes for healthy individuals at c.1210-34_1210-6TG(9_13)T(3_9) (TGmTn for simplification) #1: TG11T7/TG11T7; #2: TG10T9/TG10T9; #3: TG11T7/TG11T7; #4: TG11T7/TG11T5; *pET01*-exon10 : TG11T7 ; *pSPL3*-exon10 : TG11T7. To assess if the observed exon skipping by using minigenes are similar to those quantified in control nasal cells samples (CNC), a *U*-test (Wilcoxon - Mann Whitney) was used. No (N) means that values are significantly different with *p*-values that are indicated in brackets with $p \leq 0.05$ or $p \leq 0.01$. Yes (Y) means non-significantly different from the NCN, indicating that values found with minigene system belong to the same group than the control group, reflecting a good correspondence between endogenous *CFTR* and minigenes.

Exon	Control nasal cells					Minigenes		Same group that CNC Yes (Y) or No (N)	
	1	2	3	4	*Ref*	*pET01*	*pSPL3*	*pET01*	*pSPL3*
2	-	-	-	-		-	-	*Y*	*Y*
3	1.3	-	4.5	-	*Skipping*	8 ± 3.2	-	*N (p<0.01)*	*Y*
4	-	-	-	-		-	1.8 ± 0.6	*Y*	*N (p<0.01)*
5	-	1.4	1.4	-		5 ± 0.7	1.8 ± 0.5	*N (p<0.05)*	*Y*
6	-	-	-	-		1.2 ± 1.1	2.3 ± 1.2	*Y*	*Y*
7	-	-	-	-		-	-	*Y*	*Y*
8	-	-	-	-		-	-	*Y*	*Y*
			+ cAS			+ cAS	+ cAS	*Yes*	*Yes*
9	-	-	-	-		-	-	*Y*	*Y*
10	20.8	2.7	20.6	39.3		2.5 ± 0.5	-	*N (p<0.01)*	*N (p<0.01)*
11	-	-	-	-		-	-	*Y*	*Y*
12	1.3	-	1.6	2.7		-	-	*N (p<0.05)*	*N (p<0.05)*
13	5.9	7.3	10.1	10.2		50 ± 1.5	45 ± 3.1	*N (p<0.01)*	*N (p<0.01)*
14	-	-	-	3.4+ cAS	*Skipping+ cAS*	8 ± 2.0+ cAS	64 ± 1.7+ cAS	*N (p<0.05) Yes*	*N (p<0.01) Yes*
15	1.3	5.1	1.3	5.3	*Skipping*	3 ± 0.7	-	*Y*	*N (p<0.05)*
16	-	-	-	-		-	-	*Y*	*Y*
17	-	-	-	-		5 ± 0.6	12 ± 4.7	*N (p<0.05)*	*N (p<0.05)*
18	-	-	-	-		-	-	*Y*	*Y*
19	-	-	-	-		-	-	*Y*	*Y*
20	-	-	-	-		1.5 ± 1.3	10 ± 3.2	*Y*	*N (p<0.05)*
21	-	-	-	-		-	-	*Y*	*Y*
22	1.6	-	-	1.4	*Skipping+ cDS*	- + cDS	4 ± 1.9 + cDS	*Y* *Yes*	*N (p<0.05)* *Yes*
23	-	-	-	-		-	-	*Y*	*Y*
24	-	-	-	-		-	-	*Y*	*Y*
25	-	-	-	-		3 ± 0.7	4 ± 2.2	*N (p<0.01)*	*N (p<0.05)*
26	-	-	-	-		-	-	*Y*	*Y*
						Basal skipping		*17/25*	*15/25*
						Alternative transcripts		*3/3*	*3/3*

be studied with these minigenes due to the lack of flanking exons, preventing their amplification. Additional transcripts were detected in both *pET01* and *pSPL3* minigenes containing exons 8, 14 and 22, as identified in tissues (Table 1) and by other groups using different minigene systems [5,17]. These additional transcripts correspond to the use of cryptic ASs or DSs (Fig. 1A), as confirmed by sequencing and identified by *in silico* analysis (Table S4).

In *pET01*, minigenes containing exons 3, 5, 6, 10, 13, 14, 15, 17, 20 and 25 consistently produced detectable amounts of basal exon skipping greater than 1% (Fig. 1D and Table 1) and exons 4, 5, 6, 13, 14, 17, 20, 22 and 25 in *pSPL3*. *U*-test was used to compare the exon skipping obtained from both minigene systems (*pET01* and *pSPL3*) and events quantified in endogenous *CFTR*. When comparing to the occurrence of endogenous basal exon skipping in nasal cells, *pET01* minigenes mimicked basal skipping in 17 out of 25 exons (binomial probability $p<0.032$). Correlation was not significantly different between *pSPL3* and control nasal cells, with minigenes mimicking endogenous basal exon skipping in 15 out of 25 exons (binomial probability $p<0.097$; Fig. 1D and Table 1). For exons that do not belong to the same group than control nasal cells group (correlation not significant by comparing exon skipping of control samples and minigenes), out of the eight exons skipped in *pET01,* only six were seen in at least in one controls. In *pSPL3*, ten exons showed basal skipping of which three could be detected in controls. Therefore, only 6/8 in *pET01* and 3/10 in *pSPL3* were consistent with the results from the controls. In addition, there were two exons skipped in *pET01* (exons 17 and 25) and three in *pSPL3* (exons 4, 17 and 25) that were not skipped in the controls. Also, skipping levels were not similar. Compared to control samples, higher skipping levels was measured in *pET01* minigenes containing exon 3, 5, 13, 14, 17 or 25 while lower levels were measured with exon 10. Differences could be in part attributed to the strength of upstream DS and downstream AS which is constant in minigenes and variable in endogenous *CFTR* (Suppl. Fig. S2). For exon 13, the weak upstream DS (HSF 6.38) could favor skipping as compared to minigenes. Also, skipping of exon 10 is linked to the length of the ppt-tract, with greater skipping associated to the TG11T5 genotype as compared to TG11T7 and TG11T9. This could be observed in control samples where exon skipping well correlated with the length of the ppt-tract. Finally, distant intronic SREs could be missing in the minigenes due to the limited upstream and downstream intronic regions cloned (Suppl. Tables S2 and S3). Analysis of the variants identified in the four control individuals (n=49, n=100, n=33 and n=99) using Splice AI [18] did not predict alterations in the *CFTR* splicing pattern (data not shown), suggesting that variants found by NGS do probably not affect splicing events.

3.2. Testing exonic variants

As *pET01* minigene system appeared to correlate better to endogenous basal skipping, this system was used to study 65 exonic

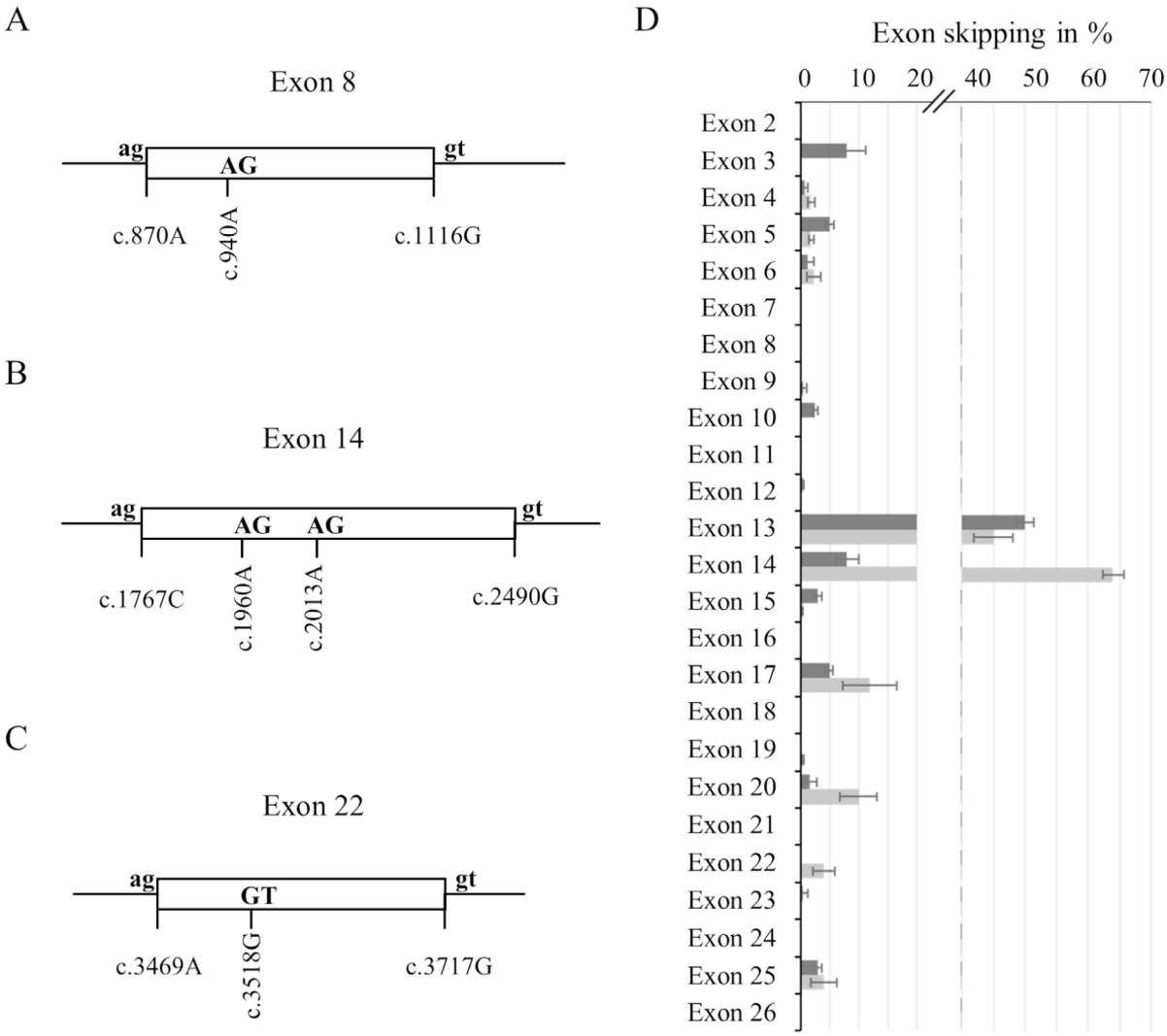

Fig. 1. Characteristics of minigenes containing WT *CFTR* exons.
Alternative splicing generated by minigenes containing *CFTR* exon 8 (A), exon 14 (B) and exon 22 (C). (D) Basal exon skipping levels for indicated exons measured using minigenes in *pET01* (black) and *pSPL3* (grey) (± sd; *n*=3-10).

variants located in 21 *CFTR* exons known as disease-causing or assigned as VUS (Fig. 2 and Suppl. Tables S5A and S5B). Two disease-causing variants favored the production of alternative transcripts that were not identified in the WT minigene. The c.2491G>T, p.Glu831* (E831X) mutation within exon 15 led to the production of an additional transcript missing 3bp due to the generation of a new AS, as previously described [16], and c.2128A>T, p.Lys710* (K710X) within exon 14 favored the production of two novel transcripts corresponding to the use of cryptic ASs (c.2147_2148 and c.2285_2286).

Results also revealed that 26 out of the 65 variants significantly increased exon skipping, ranging from 3% up to 94% (Fig. 2 and Suppl. Table S5A). For mutations within exons 14 and 22, which constitutively produced alternative transcripts, a decrease of transcripts containing the full-length exon was considered as defective (Suppl. Table S5B). The effect of some mutations has been validated from mRNA obtained from nasal brushings of CF patients (Fig. 2, with variants underlined bold, from [5–7,9,16] and our data, summarized in Suppl. Table S6), results showing a single discrepancy observed for c.3472C>T, p.Arg1158* (R1158X).

3.3. In silico *predictions of the effect of exonic variants on exon skipping*

We used results obtained with the mutant minigenes to evaluate the prediction efficacy of six freely available online *in silico* tools based on the calculation of either the ESS/ESE ratio (Skippy [19], EX-SKIP [20], ΔtESRseq [21]), ΔHZ_{EI} (HEXplorer [22]) or $\Delta\Psi$ (Spidex [23]) and MutPred Splice [24]. Analysis was also performed with Splice AI [18], a deep learning-based tool based on artificial intelligence to identify splicing variants (Table S7). *In silico* analysis was performed and results are summarized in Fig. 3, where predictions correlated with results of minigenes experiments are indicated.

Prediction efficacy was globally low, ranging from around 50% to 66% (Fig. 4) and could be associated to either high levels of false negative predictions (MutPred Splice and Splice AI) or false positive predictions (ΔtESRseq, Spidex, HEXplorer and EX-SKIP), while Skippy was associated equally to both (Fig. 4A). Results also revealed that the tools differed in their ability to correctly predict skipping events for variants that have no effect on exon skipping

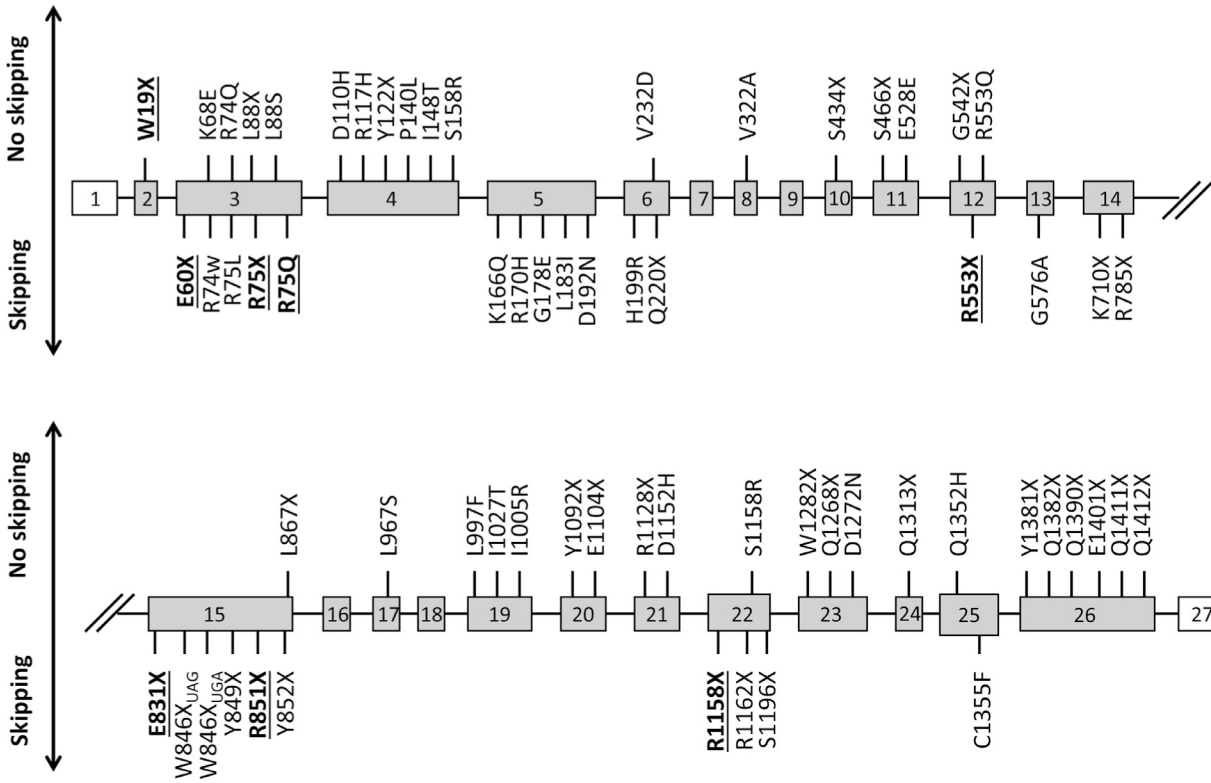

Fig. 2. *CFTR* exonic variants tested using the *pET01* minigene system.
Schematic representation of the 27 exons of *CFTR* with the positions of the 65 tested variants. Variants increasing exon skipping are indicated below the gene while those having no effect are above (n=3 to 10 experiments, unpaired t-test, $p<10^{-2}$). Variants analyzed in patient samples are indicated in bold and underlined. Variant names are indicated according to the traditional nomenclature, in order to facilitate readability.

(Fig. 4B). Indeed, while Skippy, MutPred Splice and Splice AI were more successful in identifying variants that did not affect exon recognition, ΔtESRseq, EX-SKIP, Spidex and HEXplorer were more successful identifying variants that increased exon skipping.

Additionally, prediction efficacy appeared variable from one exon to another. We compared the success rates obtained with *CFTR* exons 3, 4, 5, 15, and 26 in which 5 or more variants were tested. Exon 3 was readily analyzed by six of the seven tools, with success rates ranging from 66% up to 100% while Splice AI had a success rate of 44%. On the other hand, variants within exons 5 and 15 were more successfully analyzed with ΔtESRseq, Spidex, Hexplorer and EX-Skip than with Skippy, MutPred Splice or Splice AI. Exon 4 was poorly analyzed by ΔtESRseq, Spidex and exon 26 was poorly predicted by all software except MutPred Splice and Splice AI (Suppl. Table S8).

These exons differ by their level of basal exon skipping (Suppl. Table S8). We therefore categorized *CFTR* exons based on the level of basal exon skipping, applying a threshold of 0%, 2%, 5% and 10% (Fig. S3). Consequently, this also categorized variants located within these exons, enabling to re-evaluate prediction rates of the 65 tested variants (Fig. 5 and Suppl. Table S9).

For example, while the global success rate of ΔtESRseq was assessed at 58%, this success rate increased to 90% considering the 32 variants located in exons with a basal skipping higher than 2%, but decreased down to 36% for the 33 variants located in exons presenting a basal skipping lower than 2%. Exon categorization affected prediction success rates (Fig. 5 and Suppl. Table S9). These rates were globally increased for ΔtESRseq, Spidex, HEXplorer and EX-Skip for exons presenting a high basal exon skipping levels, with a concomitant decrease in correct predictions for exons presenting low skipping levels. An opposite result was obtained with MutPred Splice and Splice AI, while Skippy appeared only moderately affected by exon categorization.

4. Discussion

Classification of *CFTR* variants is essential to ensure accurate diagnosis and appropriate genetic counseling. Pathogenicity is assumed when variants introduce premature termination codons (PTCs), when an amino acid change is shown to alter protein function or if affecting pre-mRNA processing. In the context of therapies targeting the protein function such as CFTR potentiators or correctors, efficacy could be hindered by reduced amounts of target protein resulting from enhanced exon skipping.

In this work, we first compared two minigene banks to endogenous *CFTR* transcripts from control nasal cells. Analysis was performed with a sensitive method using capillary electrophoresis, which enabled the detection of low level transcripts (with a RFU signal above 50 to eliminate background). *CFTR* mRNA transcripts analyzed from control individuals revealed the presence of alternative transcripts for exons 8, 14, 22 and basal exon skipping greater than 1% for exons 3, 5, 10, 12, 13, 14, 15 and 22. Minigenes also produced alternative transcripts for exons 8, 14 and 22, but at different levels. Regarding basal exon skipping, a level greater than 1% was observed for exons 3, 5, 6, 10, 13, 14, 15, 17, 20 and 25 using *pET01* and exons 4, 5, 6, 13, 14, 17, 20, 22, and 25 using *pSPL3*.

Differences with endogenous *CFTR* transcripts could be attributed to several parameters that are missing in the minigene systems, such as absence of distant SREs. Another factor is NMD that will degrade transcripts lacking out of frame exons in primary cells while not in the minigene system, e.g. exons 3, 6, 8, 12, 14, 16, 17, 18, 19, 21, and 25, reducing apparent skipping levels in pa-

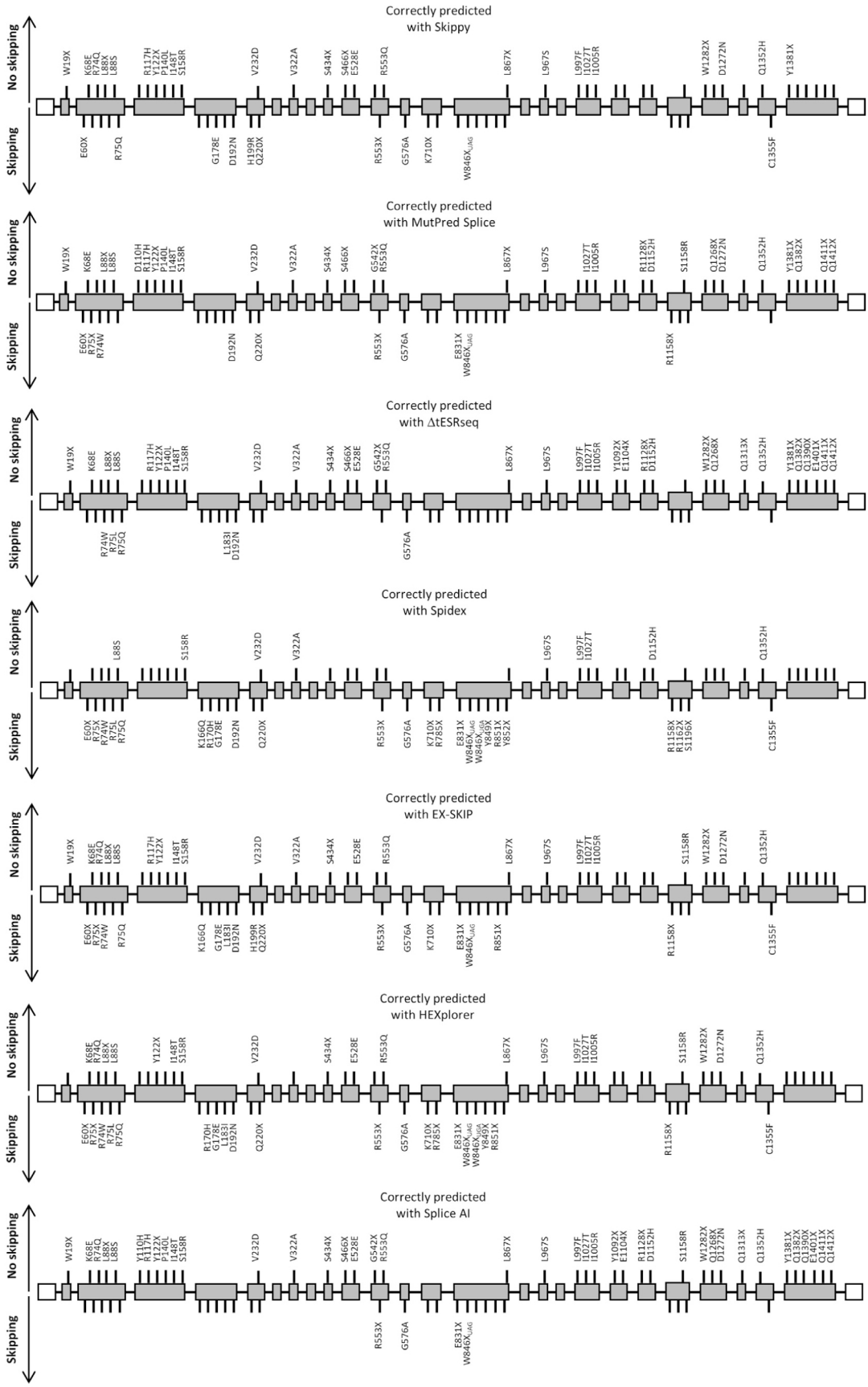

Fig. 3. *In silico* predictions for 65 exonic variants using the six software.
Correctly predicted events for each software are illustrated while incorrectly predicted effects were removed. Variant names are indicated according to the traditional nomenclature, in order to facilitate readability.

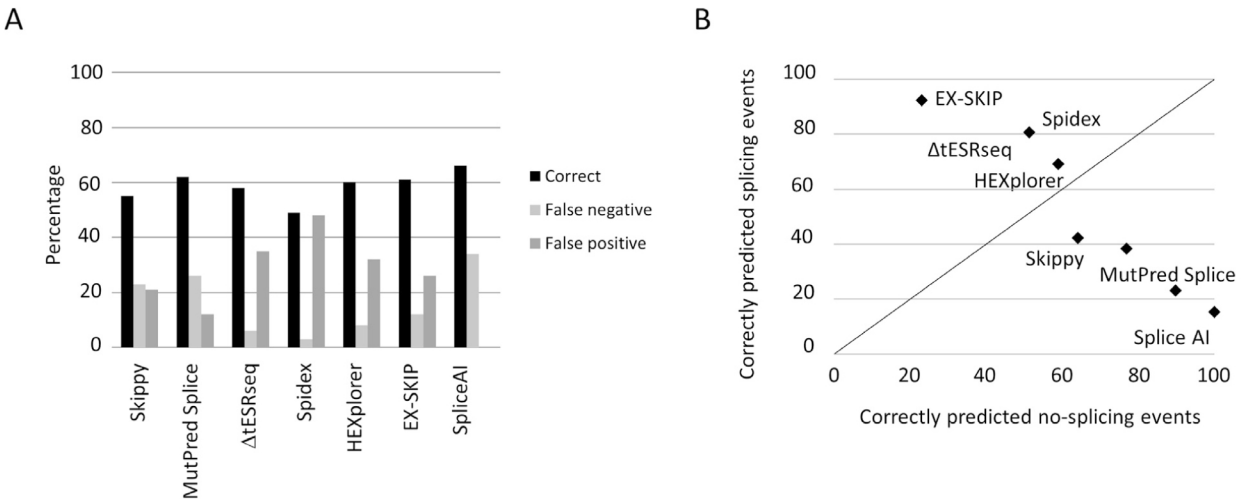

Fig. 4. Evaluation of the prediction efficacy of *in silico* tools.
(A) Percentage of correctly predicted, false negative and false positive events for the indicated *in silico* tools for *CFTR* variants. (B) Plot of the percentage of correctly predicted skipping events *versus* correctly predicted true negative events for the different *in silico* tools using the *CFTR* data set.

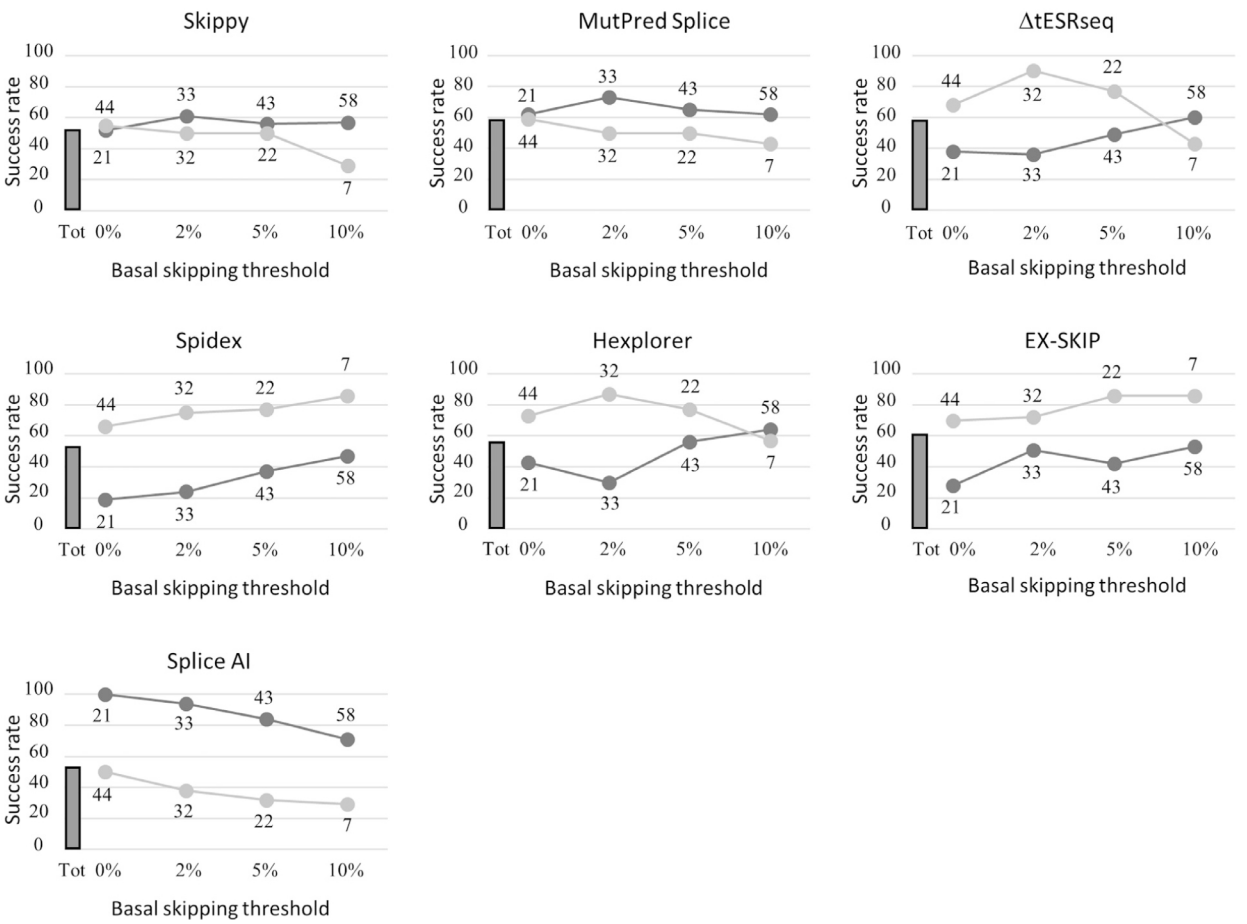

Fig. 5. Success rates for the different software at the considered level of basal exon skipping.
Global rates are indicated as dark vertical bars (Tot); dark grey line: overall success rates for variants located in exons for which basal exon skipping is lower than the indicated value at bottom; light grey line: overall success rates for variants located in exons for which basal exon skipping is higher than the indicated value at bottom. The number of variants falling in either category is indicated at each point (total=65).

tient samples. This could explain the discrepancy observed for exons 6, 17, and 25, which showed basal exon skipping in minigene but none in nasal epithelial cells (Table 1). Also, the strength of the upstream DS and downstream AS could influence exon skipping. While both downstream AS and upstream DS are constant in the minigene, they are of variable strength in *CFTR* pre-RNA [7] (Suppl. Fig. S2). For example, basal skipping of exon 13 in primary cells could be favored by the weak DS of exon 12 and be absent in minigenes having a stronger DS. Despite weaknesses underlined in this

study, minigene systems are robust enough to assess the impact of exonic mutations when patient samples are not available.

The evaluation of 65 *CFTR* exonic variants using minigenes provides substantial data on the effect on exon skipping that could be integrated in databases such as CFTR2 [25] and *CFTR*-France [26]. Out of the 65 *CFTR* exonic variants evaluated, 26 significantly increased exon skipping in *pET01*. Six of these could be confirmed in samples obtained from nasal epithelial cells. Of note, while a good correspondence with the occurrence of exon skipping was observed, the levels of exon skipping were different. Nevertheless, exon-skipping events should be taken into account in the mechanism of pathogenicity of these variants, that impact not only on diagnosis but also on therapeutic purposes.

Interestingly, 22 out of the 26 variants increasing exon skipping were found in exons presenting a basal skipping level greater than 2%, while 30 out of the 39 variants tested that did not induce exon skipping were located in exons presenting a basal skipping lower than 2%. Notable basal exon skipping therefore appeared to be indicative of a greater susceptibility of the exon to exonic skipping mutations. It could be speculated that exons presenting weaker core splicing motifs need additional signals to improve exon recognition such as ESE/ESS motifs. When exons present strong core motifs, alteration of ESE/ESS motifs could be less critical.

The occurrence of splicing defects and in particular exon skipping associated to exonic variants is not routinely searched for. Nonetheless, it is important to document such effects of variants that may have no impact on protein function, such as for c.220C>T, p.Arg74Trp (R74W) [27,28] to understand the mechanisms underlying variant pathogenicity. In the case of known CF-causing variants, the level of enhanced exon skipping may not directly impact on diagnosis and genetic counseling issues but on treatment options [29,30], as for numerous nonsense variants here studied, such as c.178G>T, p.Glu60* (E60X), c.223C>T, p.Arg75* (R75X), c.1657C>T, p.Arg553* (R553X), c.2491G>T, p.Glu831* (E831X) or c.2551C>T, p.Arg851* (R851X).

In silico tools are commonly used to predict effects on splicing, and exon skipping events identified by minigenes were used to challenge them. Overall prediction efficacy was around 50-66% with low scores associated to either high levels of false negative or false positive results. Interestingly, Splice AI which showed the highest success rate (66%) correctly assigned all thirty-nine mutations not affecting exon skipping and identified four out of the twenty-six mutations inducing exon skipping. However, Splice AI unlike to others did not predicted some splicing events, e.g. exon 3 skipping associated with p.Arg75Gln (R75Q) which was observed in nasal epithelial cells (13%) and by using *pET01* minigene (39%)(Suppl. Table S6) [7].

These results were in contrast to Soukarieh et al. who obtained much higher success rates using *CFTR* exon 13 (exon 12 in their study in legacy nomenclature) [8]. Such a discrepancy could be due to different minigene constructs and protocols used. In other respects, only three natural variants among 42 were investigated in their study and consistent results were obtained for c.1727G>C, p.AlaG576Ala (G576A), with 92% and 94% of exon skipping in Soukarieh et al. and with *pET01*, respectively. This mutation showed complete exon skipping in *pSPL3* in a previous study [12].

Importantly, categorizing *CFTR* exons on the basis of their level of basal skipping enabled to enhance prediction efficacy up to 70-100%. Exons presenting a basal exon skipping were better analyzed by deltaESR, EX-Skip, Hexplorer and Spidex while MutPred Splice and Splice AI were found to be better for exons with no basal skipping. Skippy prediction scores were not modified by exon categorization. This could lead to the preferential use of specific *in silico* tools depending on the exon type, namely exons carrying active ESEs involved in exon recognition.

While prediction tools are being constantly improved, this study illustrates the difficulties to predict the effect of sequence variants on exon splicing and highlights the need to validate the predicted effects using *in vitro* assays or, preferably, samples obtained from patients. Eventually, it shows that variant-guided therapy should also target exon skipping associated with variants.

Funding

This work was supported by the French CF Foundation « Vaincre La Mucoviscidose », grant RF20140501147.

Declaration of Competing Interest

Authors declare no conflicts of interests

CRediT authorship contribution statement

Natacha Martin: Data curation, Formal analysis, Investigation, Methodology. **Anne Bergougnoux:** Data curation, Formal analysis, Investigation, Methodology, Writing - review & editing. **Nesrine Baatallah:** Data curation. **Benoit Chevalier:** Data curation, Formal analysis. **Jessica Varilh:** Data curation. **David Baux:** Data curation, Formal analysis, Methodology, Writing - review & editing. **Bruno Costes:** Data curation, Formal analysis, Methodology, Validation. **Pascale Fanen:** Funding acquisition, Writing - review & editing. **Caroline Raynal:** Data curation, Formal analysis, Writing - review & editing. **Isabelle Sermet-Gaudelus:** Funding acquisition, Writing - review & editing. **Emmanuelle Girodon:** Writing - review & editing. **Magali Taulan-Cadars:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Validation, Writing - review & editing. **Alexandre Hinzpeter:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Validation, Writing - original draft, Writing - review & editing.

Acknowledgment

We thank Sylvie Tuffery-Giraud and Julie Miro for their helpful discussions.

References

[1] Cystic fibrosis mutation database, 2018, http://www.genet.sickkids.on.ca/Home.html n.d.

[2] Veit G, Avramescu RG, Chiang AN, Houck SA, Cai Z, Peters KW, et al. From CFTR biology toward combinatorial pharmacotherapy: expanded classification of cystic fibrosis mutations. Mol Biol Cell 2016;27:424–33. doi:10.1091/mbc.E14-04-0935.

[3] Bergougnoux A, Délétang K, Pommier A, Varilh J, Houriez F, Altieri JP, et al. Functional characterization and phenotypic spectrum of three recurrent disease-causing deep intronic variants of the CFTR gene. J Cyst Fibros 2019;18:468–75. doi:10.1016/j.jcf.2018.10.012.

[4] Maquat LE. When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells. RNA 1995;1:453–65.

[5] Aznarez I, Chan EM, Zielenski J, Blencowe BJ, Tsui L-C. Characterization of disease-associated mutations affecting an exonic splicing enhancer and two cryptic splice sites in exon 13 of the cystic fibrosis transmembrane conductance regulator gene. Hum Mol Genet 2003;12:2031–40. doi:10.1093/hmg/ddg215.

[6] 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.

[7] Aissat A, de Becdelièvre A, Golmard L, Vasseur C, Costa C, Chaoui A, et al. Combined computational-experimental analyses of CFTR exon strength uncover predictability of exon-skipping level. Hum Mutat 2013;34:873–81. doi:10.1002/humu.22300.

[8] Soukarieh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frébourg T, et al. Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using *in silico* tools. PLoS Genet 2016;12:e1005756. doi:10.1371/journal.pgen.1005756.

[9] Hinzpeter A, Aissat A, de Becdelièvre A, Bieth E, Sondo E, Martin N, et al. Alternative splicing of in-frame exon associated with premature termination codons: implications for readthrough therapies. Hum Mutat 2013;34:287–91. doi:10.1002/humu.22236.

[10] Veit G, Xu H, Dreano E, Avramescu RG, Bagdany M, Beitel LK, et al. Structure-guided combination therapy to potently improve the function of mutant CFTRs. Nat Med 2018;24:1732–42. doi:10.1038/s41591-018-0200-x.

[11] Pranke I, Bidou L, Martin N, Blanchet S, Hatton A, Karri S, et al. Factors influencing readthrough therapy for frequent cystic fibrosis premature termination codons. ERJ Open Res 2018;4. doi:10.1183/23120541.00080-2017.

[12] 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.

[13] Viart V, Bergougnoux A, Bonini J, Varilh J, Chiron R, Tabary O, et al. Transcription factors and miRNAs that regulate fetal to adult CFTR expression change are new targets for cystic fibrosis. Eur Respir J 2015;45:116–28. doi:10.1183/09031936.00113214.

[14] Bergougnoux A, Viart V, Miro J, Bommart S, Molinari N, des Georges M, et al. Should diffuse bronchiectasis still be considered a CFTR-related disorder? J Cyst Fibros 2015;14:646–53. doi:10.1016/j.jcf.2015.02.012.

[15] Raynal C, Baux D, Theze C, Bareil C, Taulan M, Roux A-F, et al. A classification model relative to splicing for variants of unknown clinical significance: application to the CFTR gene. Hum Mutat 2013;34:774–84. doi:10.1002/humu.22291.

[16] Hinzpeter A, Aissat A, Sondo E, Costa C, Arous N, Gameiro C, et al. Alternative splicing at a NAGNAG acceptor site as a novel phenotype modifier. PLoS Genet 2010;6. doi:10.1371/journal.pgen.1001153.

[17] Hull J, Shackleton S, Harris A. Analysis of mutations and alternative splicing patterns in the CFTR gene using mRNA derived from nasal epithelial cells. Hum Mol Genet 1994;3:1141–6. doi:10.1093/hmg/3.7.1141.

[18] Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting splicing from primary sequence with deep learning. Cell 2019;176:535–48 e24. doi:10.1016/j.cell.2018.12.015.

[19] Woolfe A, Mullikin JC, Elnitski L. Genomic features defining exonic variants that modulate splicing. Genome Biol 2010;11:R20. doi:10.1186/gb-2010-11-2-r20.

[20] Raponi M, Kralovicova J, Copson E, Divina P, Eccles D, Johnson P, et al. Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in BRCA1 exon 6. Hum Mutat 2011;32:436–44. doi:10.1002/humu.21458.

[21] Di Giacomo D, Gaildrat P, Abuli A, Abdat J, Frébourg T, Tosi M, et al. Functional analysis of a large set of BRCA2 exon 7 variants highlights the predictive value of hexamer scores in detecting alterations of exonic splicing regulatory elements. Hum Mutat 2013;34:1547–57. doi:10.1002/humu.22428.

[22] Erkelenz S, Theiss S, Otte M, Widera M, Peter JO, Schaal H. Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. Nucl Acids Res 2014;42:10681–97. doi:10.1093/nar/gku736.

[23] Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RKC, et al. RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. Science 2015;347:1254806. doi:10.1126/science.1254806.

[24] Mort M, Sterne-Weiler T, Li B, Ball EV, Cooper DN, Radivojac P, et al. MutPred splice: machine learning-based prediction of exonic variants that disrupt splicing. Genome Biol 2014;15:R19. doi:10.1186/gb-2014-15-1-r19.

[25] Sosnay PR, Siklosi KR, Van Goor F, Kaniecki K, Yu H, Sharma N, et al. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. Nat Genet 2013;45:1160–7. doi:10.1038/ng.2745.

[26] 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.

[27] Claustres M, Altiéri J-P, Guittard C, Templin C, Chevalier-Porst F, Des Georges M. Are p.I148T, p.R74W and p.D1270N cystic fibrosis causing mutations? BMC Med Genet 2004;5:19. doi:10.1186/1471-2350-5-19.

[28] Fanen P, Clain J, Labarthe R, Hulin P, Girodon E, Pagesy P, et al. Structure-function analysis of a double-mutant cystic fibrosis transmembrane conductance regulator protein occurring in disorders related to cystic fibrosis. FEBS Lett 1999;452:371–4. doi:10.1016/s0014-5793(99)00647-x.

[29] Masson A, Schneider-Futschik EK, Baatallah N, Nguyen-Khoa T, Girodon E, Hatton A, et al. Predictive factors for lumacaftor/ivacaftor clinical response. J Cyst Fibros 2019;18:368–74. doi:10.1016/j.jcf.2018.12.011.

[30] Baatallah N, Bitam S, Martin N, Servel N, Costes B, Mekki C, et al. Cis variants identified in F508del complex alleles modulate CFTR channel rescue by small molecules. Hum Mutat 2018;39:506–14. doi:10.1002/humu.23389.