

**Combined computational-experimental analyses of *CFTR* exon strength uncover
predictability of exon skipping level**

Abdel Aissat^{1,2}, Alix de Becdelièvre^{1,2,3}, Lisa Golmard^{1,2}, Christian Vasseur³, Catherine
Costa^{1,3}, Asma Chaoui¹, Natacha Martin¹, Bruno Costes^{1,2}, Michel Goossens^{1,2,3}, Emmanuelle
Girodon^{1,3}, Pascale Fanen^{1,2,3}, Alexandre Hinzpeter^{1,2}

¹ Inserm U955, Créteil, 94000, France ; ² Université Paris Est, Créteil, 94000, France ; ³ AP-
HP ; CHU H. Mondor, Génétique, Créteil, France.

Corresponding author: Pascale Fanen, MD, PhD, INSERM U955, Equipe 11,

Hôpital Henri Mondor

Créteil, F-94010, France.

Tel: 33 1 49 81 28 54. Fax: 33 1 48 99 33 45.

E-mail: pascale.fanen@inserm.fr.

This work was supported by grants from Institut National de la Santé et de la Recherche
Médicale (INSERM), Chancellerie des Universités de Paris, and the French Association
«Vaincre la Mucoviscidose».

Abstract

With the increased number of identified nucleotide sequence variations in genes, the current challenge is to classify them as disease-causing or neutral. These variants of unknown clinical significance can alter multiple processes, from gene transcription to RNA splicing or protein function. Using an approach combining several *in silico* tools, we identified some exons presenting weaker splicing motifs than other exons in the Cystic Fibrosis Transmembrane conductance Regulator (*CFTR*) gene. These exons exhibit higher rates of basal skipping than exons harbouring no identifiable weak splicing signals using minigene assays. We then screened 19 described mutations in three different exons, and identified exon-skipping substitutions. These substitutions induced higher skipping levels in exons having one or more weak splicing motifs. Indeed, this level remained under 2% for exons with strong splicing motifs and could reach 40% for exons having at least one weak motif. Further analysis revealed a functional exon splicing enhancer within exon 3 that was associated with the SR protein SF2/ASF and whose disruption induced exon skipping. Exon skipping was confirmed *in vivo* in two nasal epithelial cell brushing samples. Our approach, which point out exons with some splicing signals weaknesses, will help spot splicing mutations of clinical relevance.

Key words

Alternative splicing, exon skipping, unclassified variant, cystic fibrosis, *CFTR*, bioinformatics, *in silico* tools, minigene.

Introduction

With the introduction of high-throughput sequencing technologies, an increasing number of missense, silent, and intronic variants are being detected. Assessing the potential for disease causing of these variations is challenging because of their large number in human genes (Baralle, et al., 2009). The characteristics of a great number of variants, in terms of altered RNA processing or protein function, are often unknown. These nucleotide variations are named unclassified variants (UVs) or variants of unknown significance (VUS).

In vitro functional studies can reveal altered protein function, stability, or maturation. At the RNA level, studying the transcriptional product directly, when accessible, or measuring the effect on RNA splicing efficiency using a minigene approach, can also help to tag UVs as either neutral or disease-causing (Gaildrat, et al., 2010). In addition, *in silico* splice tools can also help detect potential candidate exon-skipping nucleotide substitutions (for review see (Hartmann, et al., 2008; Houdayer, 2011)).

Splicing mechanisms comprise exon recognition within large pre-mRNA molecules and the precise removal of flanking introns, as determined by *cis*-acting elements, exon length, local RNA structure, nucleosome density and rate of pre-mRNA synthesis (Graveley, 2005; Hertel, 2008; Luco, et al., 2010; Tilgner, et al., 2009). Three of the *cis*-acting elements constitute the core splicing signals: the intronic branch point (BP), the acceptor site (AS) (or 3' splice site), including an inconstant upstream polypyrimidine tract (PPT), and the donor site (DS) (or 5' splice site). These core human splice site motifs contain only a part of the information that defines exons (Lim and Burge, 2001; Wang and Burge, 2008) while the rest corresponds to less conserved splicing regulatory elements (SREs). The latter are located within the exon or flanking introns, promoting or inhibiting exon recognition through exonic/intronic splicing enhancers (ESE or ISE) or silencers (ESS or ISS), respectively (for review, see (Black, 2003; Cooper, et al., 2009)). Global approaches have been developed to identify ESE and ESS

motifs generating a comprehensive list of oligonucleotide hexamers (Zhang and Chasin, 2004) or decanucleotides (Wang, et al., 2004). These SREs recruit *trans*-acting splicing factors, serine/arginine rich proteins (SR-proteins) and heterogeneous nuclear ribonucleoproteins (hnRNPs), which activate or suppress splice site recognition or spliceosome assembly (Chasin, 2007). Consequently, nucleotide substitutions located within any of these motifs can hinder proper exon recognition, leading to aberrant RNA processing (Barash, et al., 2010; Cartegni, et al., 2002).

DNA sequence variants of unknown clinical significance were identified among the over 1,900 mutations within the extensively analyzed cystic fibrosis transmembrane conductance regulator gene (*CFTR*; MIM# 602421; GenBank NM_000492.3). Mutations in the *CFTR* gene are responsible for cystic fibrosis (CF, MIM# 219700), the most common severe recessive genetic disease in the Caucasian population (Riordan, et al., 1989). This gene needs to include intact versions of all its 27 exons to be functional, and any mutation affecting its splicing process (about 12% of the described *CFTR* mutations [<http://genet.sickkids.on.ca/CFTR>]) will reduce the amount of functional full-length transcripts. Interestingly, a subset of *CFTR* exons (i.e., exons 10, 13 and 14) undergo partial aberrant skipping in individuals, even with no apparent genomic alterations (Aznarez, et al., 2003; Hull, et al., 1994; Slomski, et al., 1992), suggesting the presence of weaker splicing signals along the *CFTR* sequence. Mutations within these exons increase exon skipping through SRE disruption (Aznarez, et al., 2003; Pagani, et al., 2003a; Pagani, et al., 2003b). It is likely that a number of UVs occurring in other *CFTR* exons could indeed affect splicing. In the present work, we first evaluated the strength of splicing signals of all *CFTR* exons using *in silico* tools and identified those containing weak core splicing elements or inconsistent SRE densities. Next, we compared these results with the basal exon skipping level detectable in several minigene constructs containing individual WT *CFTR* exon and

found that significant basal exon skipping could be observed for those exons having weaker identified splicing signals. The predictability of exon skipping induced by nucleotide variations was then evaluated by comparing *in silico* predictions and experimental minigene results. After analyzing 19 naturally occurring exonic sequence variations in three different exons, our results led to consider some of them as exon-skipping nucleotide substitutions, with consequences on splicing anticipated by *in silico* tools with a correct rate of success (16 out of 19). Interestingly, the level of exon skipping was linked to the strength of splicing motifs, mutations in several exons with weaker splicing signals inducing higher levels of skipping. We further characterized a functional ESE motif in *CFTR* exon 3 that recruited the SR protein SF2/ASF. We were able to confirm exon 3 skipping by direct *in vivo* analysis of *CFTR* mRNA of heterozygous carriers, showing that this ESE motif could be disrupted by nucleotide variations initially considered to be neutral variants or polymorphisms. These results support the hypothesis that exons presenting weak core splicing motifs or SREs are more likely to include exon-skipping nucleotide substitutions that are clinically relevant.

Materials and Methods

Mutation and reference sequence nomenclature

Recommendations of the Human Genome Variation Society (HGVS) (<http://www.hgvs.org/mutnomen/>) were followed for exon numbering and mutations names, using +1 as the A of the ATG translation initiation start site in the GenBank *CFTR* reference sequence NM_000492.3 for the RNA/cDNA. All the studied variants were previously reported to the CFTR database [<http://genet.sickkids.on.ca/CFTR>].

Ethics statement

Informed consent was obtained, and the local ethics committee approved the study.

***In silico* analysis**

The strength of core binding motifs and SRE densities as well as the effect of the different base substitutions were analysed using the free online software packages indicated in the text or in Supplementary Material (Supp. Table S1).

Cloning, plasmid constructs, and mutagenesis

The *CFTR* exons 3, 4, 5, 6, 11, 13, 14, 16, 17, 21 and 23 and intronic flanking regions were amplified using the primers summarized in Supp. Table S2 as previously described (Hinzpeter, et al., 2010), and subcloned into an exon trap vector pET01 (MoBiTec). Mutants were generated by site directed mutagenesis as previously described (Hinzpeter, et al., 2010).

Cell culture and transfection

HEK293, HeLa, HT-29, BEAS-2B, and IB3-1 cells were cultured in DMEM or LHC-8 medium supplemented with 10% fetal calf serum (Invitrogen Life Technologies). Cells were transfected with TurboFect (Thermo Fisher Scientific), according to the manufacturer's instructions.

Hybrid minigene splicing assays

One µg of minigene constructs containing either WT or mutant exon were transfected into the different cell lines, seeded in 6-well plates. After 24 hours, total RNA was purified and analyzed as previously described (Hinzpeter, et al., 2010).

Electrophoretic mobility shift assay

EMSA was performed using 280 ng recombinant SF2/ASF (Abnova) and radiolabeled RNA probe containing *CFTR* exon 3 as indicated in Supplementary Material.

Modulation of SF2/ASF expression

SF2/ASF cDNA was cloned in pcDNA3.1/V5-His-TOPO vector (Invitrogen Life Technologies) using specific oligonucleotides (5'-gccatgctgggaggtggtgt-3'; 5'-ttatgtacgagagcgagatctg-3'). Overexpression of SF2/ASF was obtained by transfection using Lipofectamine 2000 (Invitrogen Life Technologies) according to the manufacturer's instructions. siRNA experiments were performed using 100 pmol ON-TARGETplus siRNA (Thermo Fisher Scientific). Cells were incubated for 48 hours before analysis.

Western blot analysis

After transfection, to compare mRNA expression to protein content in the same sample, cells were lysed using the Allprep kit (Qiagen) following the manufacturer's instructions. Proteins were separated by 12% SDS-PAGE-electrophoresis and transferred onto PVDF membranes before being probed using anti-SF2/ASF (Santa Cruz Biotechnology) or anti-LaminB1 (Abcam) antibodies and revealed through chemiluminescence (GE Healthcare Life Sciences).

Total RNA purification and quantification of splice variants

Nasal epithelial cells were obtained by nasal brushings and *CFTR* transcripts were analyzed as previously described (Hinzpeter, et al., 2010). Semi-quantitative RT-PCR was performed using FAM-labeled primer in exon 1 (5'-ccagcgtgtctcctccttg-3') and a reverse primer in exon 4 (5'-cgatagagcgttctccttg-3').

Results

***In silico* identification of exons prone to harbor nucleotide substitutions promoting exon-skipping**

We postulated that mutations within *CFTR* exons bearing weaker splicing signals could render them more susceptible to skipping. To identify such weak signals, the strength of the core splicing motifs was first calculated for all *CFTR* exons using fourteen different freely available *in silico* tools (Table 1, Supp. Table S3, S4 and Supp. Figure S1). Values for the intronic polypyrimidine tracts and branch points were assigned to the corresponding downstream exons. Analysis of some core splicing motifs could not be performed on exons 1 and 27, the first and last exons of *CFTR*, respectively. Using a similar strategy to that initiated by Steiner and colleagues (Steiner, et al., 2004), strength values for core motifs were compared with the median value for all *CFTR* exons (Table 1, Supp. Table S3, S4 and Supp. Figure S1). A subset of exons had at least one core motif with a score in the outliers under the lower inner fence (*90% Confidence Interval, CI90*) depending on the software used (Table 1). Noteworthy, despite the heterogeneity in the results, several exons presented one of their AS or DS as weak by more than five out of thirteen tools, namely exons 5, 7, 9, 14, 15, 20, 25, and 27, which will be further considered as weak. Other exons displayed weak BP or PPT (2, 7, 10, and 22) however these latter were detected by a unique tool.

Additionally, a low frequency of ESE or high frequency of ESS (Wang, et al., 2005; Zhang, et al., 2009), as well as a higher density of intronic ESS elements in the vicinity of splice sites (Woolfe, et al., 2010), are suggestive of architectural weaknesses in the exon definition and could lead to reduced inclusion. These features were evaluated by calculating the frequency of putative ESE and ESS (and consequently their ratio (ESS/ESE)) within the entire exon or its exonic/intronic boundaries. This was done for all *CFTR* exons by using the EX-SKIP (Raponi, et al., 2011) and the SKIPPY tools (Woolfe, et al., 2010).

A similar analysis between *CFTR* exons revealed an ESE frequency below the lower inner fence (*CI90*) and/or an ESS frequency above the upper inner fence (*CI90*) and/or ESS/ESE ratio above the upper inner fence (*CI90*) for exons 3, 6, 8, 13, 16, 18, and 25 when using EX-SKIP software (Figure 1 and Table S4), for exons 3, 6, 13, 15, 18, 20, and 22 when using SKIPPY software (Figure S2 and Table S4). Additionally, exons 2, 6, 7, 15, 16, 18, and 25 displayed high intronic ESS densities up- or downstream of their splice sites, a feature calculated by SKIPPY software that might influence exon recognition (Figure S2 and Table S4).

Hence, the list of exons presenting weak splicing signals was different according to the software used but when merging all, 17 out of 27 analysed *CFTR* exons harboured at least one putative weak splicing signal, whereas some exons lacked such identifiable weak points, namely exons 1, 4, 11, 12, 17, 19, 21, 23, 24, and 26. These results led us to evaluate basal exon skipping using a minigene approach to correlate *in vitro* exon splicing behaviours with possible weaknesses in their splicing signals highlighted by *in silico* tools.

Basal exon skipping observed in minigene constructs correlates with weaker splicing signals

Basal exon skipping levels were assessed using the data of *CFTR* exons 10, 13, and 14 performed by other groups (Aznarez, et al., 2003; Pagani, et al., 2003a; Pagani, et al., 2003b). We constructed additional minigenes containing WT *CFTR* exons 3, 4, 5, 6, 11, 13, 14, 15, 16, 17, 21, or 23 plus approximately 300 nucleotides of their flanking intronic regions. After transfection in the bronchial epithelial cell line BEAS-2B and total RNA purification, samples were amplified by RT-PCR and analyzed by capillary electrophoresis as illustrated Figure 2A. From this expression system, a single 245 bp transcript was amplified from the empty plasmid, as predicted (Figure 2B). In constructs containing the studied exon, up to two peaks could be detected: a major peak corresponding to the full-length mRNA and a minor peak

corresponding to exon skipping (illustrated for exon 3 in Figure 2C). Basal exon skipping using the different WT constructs was measured (Table 2), ranging from not detectable (exons 6, 11, 16, 21, and 23) to 44% (exon 13). As exon recognition involves multiple signals, we combined *in silico* tools i.e. AS, DS, BP, PPT, SREs (Supp. Tables S5 and S6). When comparing these profiles to the computational analyses, a particular feature emerged. On the one hand, the majority of exons presenting a basal skipping level above 5% in our minigene assays had at least one weak splice signal, namely exons 3, 5, 10, 13, 14, and 15. Exceptions were exons 6 and 16, which presented one weak splice signal (Table 1 and Table 2) but with no skipped transcript detected (Table 2). On the other hand, exons 4, 11, 21, and 23, which showed stronger splice sites, had basal exon skipping under 1% (Tables 1 and 2), with the exception of exon 17 which presented basal skipping ($14\% \pm 1\%$, $n = 3$). When combining *in silico* tools, prediction success rates reached up to 10 out of 13 (Table Supp. S5 and S6). Hence, it appeared that exons presenting at least one weak point in their intrinsic splicing signals detectable by *in silico* tools were prone to basal skipping above 5% *in vitro*.

Predicting the effect of missense alterations on exon skipping

While the consequences of nucleotide substitutions affecting core splicing motifs in their conserved motifs are predicted *in silico* with a good success rate (Houdayer, et al., 2008), prediction of those occurring deeper in the exon is more delicate since SREs are less conserved (Homolova, et al., 2010; Mucaki, et al., 2011) and thus less predictable. One proposed approach consists of comparing the predicted number of ESS and ESE identified within the exon and their ratio for both WT and mutated sequences by the EX-SKIP tool (Raponi, et al., 2011). According to this approach, if the mutated exon has a higher number of ESS, a lower number of ESE or a higher ESS/ESE ratio than the WT exon, it would be considered to be more vulnerable to skipping.

Similarly, SKIPPY software focused on the loss or gain of ESE or ESS, while giving more weight to substitutions which convert an ESE to an ESS for example (Woolfe, et al., 2010). The tendency of a nucleotide substitution to be a splicing-altering variation is calculated with a log odd ratio (LOR); the highest the calculated LOR is, the higher the predicted splicing alteration is.

Consequently, by comparing the results from the splicing pattern of mutations located within exons 10 and 13 tested with a minigene (Pagani, et al., 2000; Pagani, et al., 2003b) and the EX-SKIP predicted skipping susceptibility, we obtained 50% and 100% success rates of prediction, respectively (Supp. Table S7). In parallel, SKIPPY predictions were less correlated, with success rate of prediction reaching only up to 50% (Supp. Table S7).

In light of these results, we wanted to evaluate the presence of exon-skipping nucleotide substitutions in *CFTR* exons presenting weak or strong splicing signals. *In silico* analysis identified exon 3 as having an ESS number and an ESS/ESE ratio in the outliers above the upper inner fence (*CI90*) calculated for all *CFTR* exons (Figure 1 and Supp. Table S4).

Conversely, exon 4 had no apparent weak points, whereas exon 5 displayed a weak donor site (Table 1 and Supp. Figure S1). Therefore, exons 3 and 5 would be more likely to harbor exon-skipping nucleotide substitutions compared with exon 4. To test this hypothesis, we picked natural sequence variations occurring in exons 3, 4, and 5 from the CF mutation database and evaluated their effect on splicing using a minigene strategy.

Hybrid minigene splicing assay identifies exonic mutations that affect splicing

We constructed hybrid minigenes containing exonic mutations in *CFTR* exons 3, 4, or 5 and transfected them, as well as control wild-type constructs, into BEAS-2B cells. The results obtained with the minigene splicing assay revealed exon-skipping nucleotide substitutions (summarized in Supp. Table S8). Increased exon 3 skipping was observed with several nucleotide substitutions ($p < 0.001$), including c.178G>T ($32\% \pm 1.6\%$, $n = 3$), c.220C>T

(39% \pm 0.3%, n = 4), c.223C>T (17% \pm 1.8%, n = 4), c.224G>T (37% \pm 5.2%, n = 4), and c.224G>A (39% \pm 0.3%, n = 4), while c.202A>G (2% \pm 0.3%, n = 3), c.263T>G (3% \pm 0.3%, n = 3), and c.263T>C (4% \pm 0.7%, n = 3) did not increase exon 3 skipping (Supp. Figure S3 and Supp. Table S8).

Some nucleotide substitutions within exon 4, such as c.328G>C (1.8% \pm 0.3%, n = 3), c.366T>A (1.9% \pm 0.6%, n = 3), c.350G>A (1.9% \pm 0.2%, n = 3), significantly increased exon skipping ($p < 0.02$) compared to WT construct (0.7% \pm 0.2%, n = 6), but the latter remained under 2%. Other mutations did not significantly induce exon skipping compared to WT, such as c.419C>T (1.3% \pm 0.3%, n = 3), c.443T>C (0.6% \pm 0.1%, n = 3), and c.472A>C (0.3% \pm 0.1%, n = 3) (Supp. Table S8), while complete exon skipping was observed with both c.274-2A>C and c.489+1G>T mutations, affecting acceptor and donor sites respectively (data not shown).

Compared with the WT construct (5% \pm 0.7%, n = 4), the results obtained with mutations within exon 5 revealed an increase in exon skipping ($p < 0.01$) for mutants c.496A>G (9% \pm 0.4%, n = 3), c.509G>A (18% \pm 2.2%, n = 4) and c.533G>A (22% \pm 0.2%, n = 3), c.547C>A (10% \pm 0.2%, n = 3), and c.574G>A (8% \pm 0.9%, n = 3) (Supp. Table S8).

The best prediction obtained with EX-SKIP for the mutations tested was 100% for exon 3 (8 out of 8), 67% in exon 4 (4 out of 6), and 80% in exon 5 (4 out of 5) (Supp. Table S8). With the SKIPPY software, the best predictions were done with substitutions affecting exon 4 (3 out of 6) (Supp. Table S8).

Nonetheless, even if the negative or positive impact of the substitution on the exon inclusion was correctly predicted by these tools, they are not predictive of the level of exon skipping, which appeared to be correlated to the strength of the splicing signals. Indeed, for exons with at least one weak splicing signal, skipping was present in up to 39% of transcripts for exon 3 and 22% for exon 5 (Supp. Table S8). Similarly, examples of exons with weak splicing

signals reported in the literature were excluded in up to 85% of transcripts carrying mutations in exon 10 and 100% of transcripts carrying mutations in exon 13 (Supp. Table S7). In addition, when the tested mutations within the exon 4 having no detected weak signal induced skipping, this affected less than 2% of transcripts (Supp. Table S8).

Cell-specific effects on exon 3 skipping

We further focused the study on the region of exon 3 containing nucleotides c.220 to c.224, where the polymorphism c.224G>A, which represents 3% of alleles in the European population (dbSNP 135, rs1800076), is located. We evaluated exon 3 splicing behaviour in cell lines of different origin: the lung CF epithelial cell line IB3-1, the colonic cell line HT-29, the embryonic kidney cell line HEK293 and the cervical cancer cell line HeLa. Basal exon skipping with the WT construct ranged from 6% \pm 0.9% (n = 6) in IB3-1 cells to 13% \pm 2.8% (n = 6) in HEK293 cells (Supp. Figure S4 and Supp. Table S9). The results obtained with mutation-containing constructs were globally similar throughout the different cell lines tested; c.220C>T, c.224G>T, and c.224G>A substitutions induced a 2- to 5-fold increase in exon 3 skipping. However, skipping was highest in BEAS-2B epithelial cells and lower in HEK293 and HeLa cells. The c.223C>T substitution presented a more complex pattern, with a significant increase in exon 3 exclusion measured in the colonic and in the two lung epithelial cell lines. By contrast, HeLa cells showed a similar splicing pattern to the WT construct, whereas the substitution had an opposite effect on HEK293 cells, significantly improving exon 3 recognition (two-fold) compared with WT ($p < 0.006$).

***In silico* prediction of a molecular mechanism responsible for exon 3 skipping**

Exon skipping can be the consequence of altered binding of SR-proteins and hnRNPs to the nucleotide stretch. *In silico* analyses, using HSF, ESEfinder3.0, and ASSA, enabled the detection of putative binding site disruption or creation for some of these splicing factors (Supp. Table S10 and S11). Alteration of these binding sites can give clues to the nature of the

molecular mechanism responsible for exon skipping. Indeed, the implication of some of these splicing factors had been previously confirmed by *in vitro* modulation studies for exons 10 and 13 by others (Pagani, et al., 2003a; Pagani, et al., 2003b) (indicated with a # in Supp. Table S10).

Interestingly, the three substitutions associated with a 3- to 5-fold increase in exon 3 exclusion (r.220c>u, r.224g>u, and r.224g>a) reduced the score of the sole predicted SF2/ASF-binding motif located within exon 3 under the default threshold value of 1.867 with ESEfinder3.0 (Figure 3A and Supp. Table S11). Indeed, the SF2/ASF score matrices showed a decrease from 2.7 to 0.99 for both r.220c>u and r.224g>a, and to 0 for r.224g>u. Two substitutions, r.220c>u and r.223c>u, also abolished a potential SRp55-binding motif, while r.224g>u and r.224g>a disrupted two 9G8-binding motifs predicted by HSF (Supp. Table S11). Particularly, the r.223c>u mutation also led to the appearance of a new potential SF2/ASF-binding motif (CUUCGGUG) with a 2.06 score, three nucleotides upstream of the normal SF2/ASF-binding motif (CGGCGAUG) (Figure 3A in bold). This novel SF2/ASF-binding motif is in accordance with the reduced splicing defect associated with this substitution (17%) compared to c.220C>T, c.224G>A, and c.224G>T (37-39%) (Supp. Figures S3 and S4, and Supp. Table S8). Exon 3 exclusion detected in epithelial cell lines associated to the c.223C>T mutation could be caused by the decreased affinity of this new potential binding site compared to WT (from a score of 2.7 to 2.06).

Direct association of SF2/ASF to exon 3 RNA probes

To test whether the SF2/ASF protein could bind to the exon 3 pre-mRNA strand according to the software predictions, direct association between SF2/ASF and exon 3 was assessed by electrophoretic mobility shift assay (EMSA) experiments using a RNA probe containing complete WT or mutant *CFTR* exon 3 (109 nucleotides) and increasing quantities of recombinant purified SF2/ASF protein. The results showed a dose dependent electrophoretic

shift of the RNA probe, indicating a direct association between the WT nucleotide sequence and recombinant SF2/ASF (n = 3, Figure 3B). Results obtained with probes containing the r.224g>a mutation revealed a reduced electrophoretic shift at higher concentrations of SF2/ASF compared to WT probes (n = 3, Figure 3B), compatible with a reduced affinity of the SF2/ASF binding site.

Modulation of SF2/ASF levels

To evaluate the importance of SF2/ASF for exon 3 recognition and to rescue exon skipping, we upregulated its expression in the CF bronchial epithelial cell line IB3-1 transfected with the minigene. The amounts of SF2/ASF were measured by Western blot, normalized by LaminB1 expression, and reported relative to 100% of control. The results illustrated in Figure 3C, obtained 48 hours after transfection, indicated that a 5-fold upregulation slightly but significantly increased exon insertion (from $4.4\% \pm 0.4\%$ to $3.2\% \pm 0.1\%$, n = 6, p < 0.01), whereas a 5-fold downregulation of SF2/ASF did not affect WT exon 3 recognition. Skipping of exon 3 bearing the c.224G>A substitution was more sensitive to SF2/ASF upregulation; a 4-fold overexpression significantly corrected exon 3 splicing, from $27\% \pm 1.0\%$ to $18\% \pm 0.3\%$ (n = 3, p < 0.001). Similar treatment using the minigene containing the c.220C>T substitution resulted in an analogous profile, with a significant correction of the splicing defect after a 3-fold upregulation of SF2/ASF from $18\% \pm 1.0\%$ to $11\% \pm 0.7\%$ (n = 3, p < 0.001) (Supp. Figure S5). The observation that upregulation of SF2/ASF partially corrected the splicing defect associated with both c.224G>A and c.220C>T substitutions indicates that SF2/ASF might play an important role in exon 3 recognition and that the reduced affinity of the binding site could be compensated by overexpression of SF2/ASF.

Direct transcript analysis corroborates *in vivo* skipping of exon 3

CFTR mRNA obtained from nasal epithelial cells of two healthy volunteers heterozygous for the c.224G>A nucleotide variation was subjected to both a quantitative and a semi-

quantitative RT-PCR analysis. For these two samples, no other mutation was detected after genomic analysis of the whole coding sequence and flanking regions. Total *CFTR* transcripts were quantified by qRT-PCR, but this showed no statistically significant differences in the levels of any of the c.224G>A (p.Arg75Gln) heterozygotes compared with three healthy controls (data not shown). Capillary electrophoresis of the two samples revealed skipping of exon 3 in $13\% \pm 1\%$ ($n = 6$) of transcripts, whereas samples obtained from three controls showed skipping in less than 1% of transcripts (Figure 4). These shorter transcripts are predicted to encode a truncated, non-functional 69 amino acid CFTR peptide since a premature termination codon appears in exon 4.

Discussion

The *CFTR* gene is a useful model for the study of genotype/phenotype correlation, CF and related disorders being associated with a large spectrum of symptoms and sequence variations. We tested the predictability of the impact of a subset of exonic nucleotide substitutions on the splicing pattern of the corresponding *CFTR* exon. In the present study, we sought to recognize exons prone to harbor exon-skipping associated mutations. To this purpose, we first calculated using several *in silico* tools the mean strength of the splicing motifs for all 27 *CFTR* exons, and considered them subsequently as having weak splicing signal when one of their calculated score was in the outliers (*CI90*): under the lower inner fence (for core splicing signals and ESE density) or above the upper inner fence (for ESS density and ESS/ESE ratio).

Measuring exon skipping in a subset of WT minigene constructs (n=12) in a second step, led us to observe basal exon skipping above 5% for 6 exons. By comparing *in vitro* and *in silico* results, it appeared that exons presenting weak splicing signals had a basal skipping above 5%. The correlation reached 10 out of 13 when combined *in silico* tools which evaluate the strength of AS, DS, PPT, BP and SREs. Correlation could be improved if taking into account other signals known to influence exon recognition. Indeed, the length of the exons influences their recognition as proposed by the exon definition model (Berget and coworkers), which invokes the necessary communication of 3'- and 5'-splice site complexes across an exon for efficient splicing (Berget, 1995; Robberson, et al., 1990). The requirement for communication across an exon led to the model that splice site recognition is more efficient when exons are small, and exon inclusion decreases with increasing exon size (Berget, 1995). The small length of exon 16 (38 bp, under the *CI90* lower inner fence, Supp. Table S4 and Figure S6), can therefore account for the absence of skipped transcripts. In contrast, the larger exon 14 (724bp, above the *CI90* higher inner fence, Supp. Table S4 and Figure S6) would be less

efficiently recognized, in addition to its weak acceptor site (predicted by 8 out of 12 algorithms, Table 1). Concordantly, this exon has been previously shown to use a cryptic acceptor site (Aznarez, et al., 2003) and results obtained with our minigene model revealing the use of cryptic acceptor sites and exon skipping (data not shown).

Two other discrepancies were observed for exons 6 and 17, which exhibited differential behavior *in silico* and *in vitro*. Therefore, caution should be exercised with *in silico* analyses tools available at present. The presence of basal exon skipping in a minigene construct is relatively relevant for pointing out an exon as having putative strong or weak splicing signals, and when paired to *in silico* analyses facilitates further characterization of these weak points and definition of their appropriate thresholds for an efficient prediction.

This classification of *CFTR* exons splicing signals as either strong or weak, albeit simplistic, could help anticipate the impact of exonic nucleotide substitutions and could be useful to investigate sequence variations as putative splicing mutations. While the occurrence of exon skipping was appropriately anticipated by the *in silico* tool EX-SKIP (success rates from 67% to 100%) with our minigene constructs (Supp. Table S8), its level was not. This latter parameter appeared to be correlated to the strength of the splicing motifs of the exon, as exon skipping levels differed between exons 3 and 5 having weak signals (up to 39%) and exon 4 without any detectable weak signal (up to 1.9%) after nucleotide substitutions. High exon skipping rates were also associated with mutations located within *CFTR* exons 10 and 13, both presenting splicing signal weaknesses (up to 85% and 100%, respectively) (Pagani, et al., 2003a; Pagani, et al., 2003b). On the other hand, SKIPPY skipping tendency did not successfully anticipate the effect of nucleotide substitutions (success rate 12.5% to 50%) (Supp. Table S8). Therefore exon skipping mutations located within exons with low-strength splicing motifs are more likely to have biological consequences and can be predicted with a higher success rate. The nature of the nucleotide substitution itself, its location in such exon

and the *trans*-acting splicing factors which are possibly altered, have also to be taken into account by *in silico* analyses, however, some parameters such as local RNA structure, nucleosome density and rate of pre-mRNA synthesis were not.

Consequently, to finally define the molecular mechanism responsible for increased exon 3 skipping, we looked at the binding of splicing factors putatively altered by a subset of naturally occurring substitutions. ESEfinder3.0 predicted an alteration of the sole potential SF2/ASF binding motif within the exon 3, this uniqueness led us to further explore the involvement of this factor. *In vitro* analysis revealed a functional ESE within exon 3 that was associated with SF2/ASF and whose disruption induced exon skipping. Remarkably, exon 3 skipping was significantly reduced after overexpression of SF2/ASF. An opposite effect was reported for *CFTR* exon 10, in which over-expression of SF2/ASF increased exon skipping via an intron splicing silencer (ISS) sequence located in intron 10 (Pagani, et al., 2000; Pagani, et al., 2003c). These results illustrate the dual effect of SR-proteins such as SF2/ASF, which have been shown to either correct or increase exon skipping depending on the exon under study (Cartegni and Krainer, 2002; Xu, et al., 2005). Great improvement of splicing *in silico* tools has occurred since the first splice sites algorithms by Shapiro and Senapathy (Shapiro and Senapathy, 1987). The postulate that alternatively spliced exons harbor less auxiliary splicing signals than constitutive exons is emerging (Sterne-Weiler, et al., 2011), illustrated by the development of recent multi-analysis tools which compile several algorithms (Woolfe, et al., 2010). However, these computational tools do not take into account the recognition of a motif in a local region, which depends both on the nucleotide constitution and the geometry of the space occupied by the mRNA strand (Buratti and Baralle, 2004; Liu, et al.). ESEfinder and HSF calculations, based on the SELEX method, use the primary nucleotide sequence to score SR binding sites without taking into account RNA secondary structures. These secondary 2D-structures are illustrated for WT and mutated exon 3 showing

possible modification of the local secondary structure in the region r.217-r.227 (Supp. Figure S7), which could have consequences on the binding of SF2/ASF protein. Conversely, these 2D-predictions do not reckon with the steric hindrance of the splicing factors which bind to the mRNA strand and modify the local conformation (Jin, et al., 2011), and deserve to be further explored by *in vitro* structural experiments. In accordance with these assumptions, we observed that full-length exon 3 mRNA probe (109 nucleotides) was more sensitive to substitution by EMSA than a 25 nucleotides probe (data not shown). The larger probe uncovers subtle modifications in RNA secondary structure due to point mutation, which would influence the efficient binding of SF2/ASF protein.

Our data illustrate the importance of SRE motifs in the regulation of exon recognition as 9 mutations (out of 19 tested) induced exon skipping in 10% to 39% of the transcripts. The disease relevance of a reduced level of functional full-length transcripts must be considered for each gene. Nonetheless, these splicing mutations do not abolish wild-type expression completely, and this may lead to milder phenotypes. Concordantly, we identified alternative splicing in *CFTR* exons 3 and 5 associated with mutations that do not cause severe CF but were identified in patients affected by either mild forms of the disease or *CFTR*-related disorders (reviewed in Bombieri, et al., 2011). How reduced *CFTR* channel function is implicated in the onset of these clinical manifestations is still a field in progress.

In conclusion, an approach categorizing exons as either weak or strong could help spot exon skipping mutations reaching clinical relevance. These observations also emphasize the necessity of using a minigene approach to adjust *in silico* analysis (Houdayer, et al., 2012). Before deciphering the ‘splicing code’, *in silico* tools appear, today, to be more useful for providing preventive knowledge of the weak points of each exon than for simple and complete explanation of the consequences of an exonic nucleotide change on splicing (Desmet and Beroud, 2012). Such a preventive analysis could help prioritize samples for

further RNA-based analyses to confirm splicing defects (Houdayer, 2011). The identification of such exon-skipping mutations can aid in the understanding of their associated phenotype variability and can improve the interpretation of genetic test results (Plon, et al., 2008).

Acknowledgements

Abdel Aissat received a doctoral fellowship from the *Ministry of Higher Education and Research* (Ministère de l'Enseignement supérieur et de la Recherche, MESR) and the «Fondation pour la Recherche Médicale».

Disclosure Statement: The authors declare no competing interests.

References

- Aznarez I, Chan EM, Zielenski J, Blencowe BJ, Tsui LC. 2003. 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* 12(16):2031-40.
- Baralle D, Lucassen A, Buratti E. 2009. Missed threads. The impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep* 10(8):810-6.
- Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ. 2010. Deciphering the splicing code. *Nature* 465(7294):53-9.
- Berget SM. 1995. Exon recognition in vertebrate splicing. *J Biol Chem* 270(6):2411-4.
- Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72:291-336.
- Bombieri C, Claustres M, De Boeck K, Derichs N, Dodge J, Girodon E, Sermet I, Schwarz M, Tzetis M, Wilschanski M and others. 2011. Recommendations for the classification of diseases as CFTR-related disorders. *J Cyst Fibros* 10 Suppl 2:S86-102.
- Buratti E, Baralle FE. 2004. Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol Cell Biol* 24(24):10505-14.
- Cartegni L, Chew SL, Krainer AR. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 3(4):285-98.
- Cartegni L, Krainer AR. 2002. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat Genet* 30(4):377-84.
- Chasin LA. 2007. Searching for splicing motifs. *Adv Exp Med Biol* 623:85-106.
- Cooper TA, Wan L, Dreyfuss G. 2009. RNA and disease. *Cell* 136(4):777-93.

- Desmet FO, Beroud C. 2012. Bioinformatics and mutations leading to exon skipping. *Methods Mol Biol* 867:17-35.
- Gaildrat P, Killian A, Martins A, Tournier I, Frebourg T, Tosi M. 2010. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol Biol* 653:249-57.
- Graveley BR. 2005. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* 123(1):65-73.
- Hartmann L, Theiss S, Niederacher D, Schaal H. 2008. Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Front Biosci* 13:3252-72.
- Hertel KJ. 2008. Combinatorial control of exon recognition. *J Biol Chem* 283(3):1211-5.
- Hinzpeter A, Aissat A, Sondo E, Costa C, Arous N, Gameiro C, Martin N, Tarze A, Weiss L, de Beudelievre A and others. 2010. Alternative splicing at a NAGNAG acceptor site as a novel phenotype modifier. *PLoS Genet* 6(10).
- Homolova K, Zavadakova P, Doktor TK, Schroeder LD, Kozich V, Andresen BS. 2010. 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 cbIE type of homocystinuria. *Hum Mutat* 31(4):437-44.
- Houdayer C. 2011. In silico prediction of splice-affecting nucleotide variants. *Methods Mol Biol* 760:269-81.
- Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, Bronner M, Buisson M, Coulet F, Gaildrat P and others. 2012. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat* 33(8):1228-38.
- Houdayer C, Dehainault C, Mattler C, Michaux D, Caux-Moncoutier V, Pages-Berhouet S, d'Enghien CD, Lauge A, Castera L, Gauthier-Villars M and others. 2008. Evaluation

- of in silico splice tools for decision-making in molecular diagnosis. *Hum Mutat* 29(7):975-82.
- Hull J, Shackleton S, Harris A. 1994. Analysis of mutations and alternative splicing patterns in the CFTR gene using mRNA derived from nasal epithelial cells. *Hum Mol Genet* 3(7):1141-6.
- Jin Y, Yang Y, Zhang P. 2011. New insights into RNA secondary structure in the alternative splicing of pre-mRNAs. *RNA Biol* 8(3):450-7.
- Lim LP, Burge CB. 2001. A computational analysis of sequence features involved in recognition of short introns. *Proc Natl Acad Sci U S A* 98(20):11193-8.
- Liu W, Zhou Y, Hu Z, Sun T, Denise A, Fu XD, Zhang Y. Regulation of splicing enhancer activities by RNA secondary structures. *FEBS Lett* 584(21):4401-7.
- Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. 2010. Regulation of alternative splicing by histone modifications. *Science* 327(5968):996-1000.
- Mucaki EJ, Ainsworth P, Rogan PK. 2011. Comprehensive prediction of mRNA splicing effects of BRCA1 and BRCA2 variants. *Hum Mutat* 32(7):735-42.
- Pagani F, Buratti E, Stuani C, Baralle FE. 2003a. Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. *J Biol Chem* 278(29):26580-8.
- Pagani F, Buratti E, Stuani C, Romano M, Zuccato E, Niksic M, Giglio L, Faraguna D, Baralle FE. 2000. Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. *J Biol Chem* 275(28):21041-7.
- Pagani F, Stuani C, Tzetis M, Kanavakis E, Efthymiadou A, Doudounakis S, Casals T, Baralle FE. 2003b. New type of disease causing mutations: the example of the

- composite exonic regulatory elements of splicing in CFTR exon 12. *Hum Mol Genet* 12(10):1111-20.
- Pagani F, Stuani C, Zuccato E, Kornblihtt AR, Baralle FE. 2003c. Promoter architecture modulates CFTR exon 9 skipping. *J Biol Chem* 278(3):1511-7.
- Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, Hogervorst FB, Hoogerbrugge N, Spurdle AB, Tavtigian SV. 2008. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat* 29(11):1282-91.
- Raponi M, Kralovicova J, Copson E, Divina P, Eccles D, Johnson P, Baralle D, Vorechovsky I. 2011. Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in BRCA1 exon 6. *Hum Mutat* 32(4):436-44.
- Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL and others. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 245(4922):1066-73.
- Robberson BL, Cote GJ, Berget SM. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* 10(1):84-94.
- Shapiro MB, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 15(17):7155-74.
- Slomski R, Schloesser M, Berg LP, Wagner M, Kakkar VV, Cooper DN, Reiss J. 1992. Omission of exon 12 in cystic fibrosis transmembrane conductance regulator (CFTR) gene transcripts. *Hum Genet* 89(6):615-9.
- Steiner B, Truninger K, Sanz J, Schaller A, Gallati S. 2004. The role of common single-nucleotide polymorphisms on exon 9 and exon 12 skipping in nonmutated CFTR alleles. *Hum Mutat* 24(2):120-9.

- Sterne-Weiler T, Howard J, Mort M, Cooper DN, Sanford JR. 2011. Loss of exon identity is a common mechanism of human inherited disease. *Genome Res* 21(10):1563-71.
- Tilgner H, Nikolaou C, Althammer S, Sammeth M, Beato M, Valcarcel J, Guigo R. 2009. Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol* 16(9):996-1001.
- Wang J, Smith PJ, Krainer AR, Zhang MQ. 2005. Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. *Nucleic Acids Res* 33(16):5053-62.
- Wang Z, Burge CB. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 14(5):802-13.
- Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. 2004. Systematic identification and analysis of exonic splicing silencers. *Cell* 119(6):831-45.
- Woolfe A, Mullikin JC, Elnitski L. 2010. Genomic features defining exonic variants that modulate splicing. *Genome Biol* 11(2):R20.
- Xu X, Yang D, Ding JH, Wang W, Chu PH, Dalton ND, Wang HY, Bermingham JR, Jr., Ye Z, Liu F and others. 2005. ASF/SF2-regulated CaMKIIdelta alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. *Cell* 120(1):59-72.
- Zhang XH, Arias MA, Ke S, Chasin LA. 2009. Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing. *RNA* 15(3):367-76.
- Zhang XH, Chasin LA. 2004. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev* 18(11):1241-50.

Figure Legends

Figure 1. *In silico* analysis of *CFTR* exons.

CFTR box plot of the mean frequency of exon splicing enhancers (ESE) and exon splicing silencers (ESS) per 100 bp, calculated for all *CFTR* exons with the EX-SKIP tool (left side). *CFTR* box plot of the ESS/ESE ratio, calculated for all *CFTR* exons, is also shown (right side). Exons presenting with ESS scores and ESS/ESE ratio above the upper inner fence (*CFTR*) or with ESE score under the lower inner fence (*CFTR*) are indicated.

Figure 2. Hybrid minigene splicing assays for control constructs.

(A) Schematic representation of the minigene construct produced transcripts and expected size exemplified for *CFTR* exon 3. Examples of capillary electrophoresis analysis of RT-PCR products obtained after 25 cycles. RNA was purified from BEAS-2B cells transfected with empty pET01 plasmid (B), minigene containing WT exon 3 (C). RT-PCR was performed using a Fluorescein amidite (FAM)-labeled forward primer located within the splice donor exon and a reverse primer within the splice acceptor exon of the pET01 plasmid. The corresponding size and relative amount for each peak is indicated. The peak at 245 bp corresponds to exon 3-skipped mRNA, whereas the peak at 354 bp corresponds to full-length mRNA.

Figure 3. *In silico* analysis of exon 3 binding motifs and *in vitro* binding assay.

(A) Score matrix of various SR proteins predicted to bind to exon 3 RNA, as predicted by ESE finder version 3.0 using the default threshold value. The motifs predicted by SF2/ASF are in purple (threshold value of 1.867), SC35 in blue (threshold value of 2.383), SRp40 in green (threshold value of 2.67), and SRp55 in yellow (threshold value of 2.676). Nucleotides r.205-248 of exon 3, encompassing the region of interest, are illustrated. Sequences corresponding to nucleotides r.217-227 are illustrated under histograms with the SF2/ASF-binding motif in bold and the corresponding nucleotide substitution underlined. (B) EMSA

assays performed with WT and mutant *CFTR* exon 3 RNA probes and increasing amounts of recombinant SF2/ASF. (C) SF2/ASF modulation affects exon 3 inclusion. *Upper panel*: semi-quantification of exon 3 exclusion measured for WT sequence or for sequences containing either the c.224G>A mutation, in IB3-1 cells. *Lower panel*: Western blot analysis measuring SF2/ASF and LaminB1 expression in corresponding cells. Mean % of control measures of SF2/ASF normalized to LaminB1 signal.

Figure 4. Direct transcript analysis.

An example of capillary electrophoresis analysis of RT-PCR products obtained after 30 cycles from control samples (n = 3) (A) or c.224G>A/- (p.Arg75Gln/-) heterozygotes (n = 2) (B).

The corresponding size and relative amount of each peak are indicated. The peak size corresponds, respectively, to full-length mRNA (333 bp) and to exon 3-skipped mRNA (224 bp).