Combined Computational-Experimental Analyses of CFTR Exon Strength Uncover Predictability of Exon-Skipping Level.

Abdel Aissat, Alix De Becdelièvre, Lisa Golmard, Christian Vasseur, Catherine Costa, Asma Chaoui, Natacha Martin, Bruno Costes, Michel Goossens, Emmanuelle Girodon, et al.

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.

Name of the software	Exon(s) with a weak AS	Exon(s) with a weak DS	Exon(s) with a weak BP	Exon(s) with a weak PPT
1) ESEfinder3.0	9; 12; 14; 20	5; 20; 25	-	-
2) FSplice	7; 9; 20	5; 21	-	-
3) GeneID	7; 15; 20; 27	20; 25	-	-
4) GenSCAN	7; 14; 15; 18; 20; 27	5; 9; 10; 12; 25	-	-
5) H-Bond	*na*	10; 12; 25	-	-
6) Human Splicing Finder	14; 21	5; 6; 10; 17; 20; 25	-	-
7) MaxEntScan	7; 11; 15; 20; 27	6 ; 9	-	-
8) NetGene2	14; 15; 20	5; 6; 9; 12; 25	-	-
9) NNSplice	7; 9; 11; 14; 15; 20	4; 9; 10; 16; 22	-	-
10) SplicePort	13; 14	12; 20; 22	-	-
11) SplicePredictor	7; 11; 20; 27	3; 4; 9; 17	-	-
12) SpliceSiteFrame	14; 20; 27	6; 20; 25	-	-
13) SpliceView	7; 14; 20	5; 7; 20	-	-
14) SROOGLE			2; 22	7; 10
Most represented exons	Exon 20 *(10 out of 12)* Exon 14 *(8 out of 12)* Exon 7 *(7 out of 12)* Exon 15 *(5 out of 12)* Exon 27 *(5 out of 12)*	Exon 25 *(7 out of 13)* Exon 5 *(6 out of 13)* Exon 20 *(6 out of 13)* Exon 9 *(5 out of 13)*		

Table 1. Summary of exons harboring weak core splicing signals according to each splicing tool used for the analyses. Weak splicing signals are defined as those in the outliers inferior or equal to the lower inner fence when comparing the strength values with the median value of all *CFTR* exons (confidence interval of 90%, *CI90*). All the calculated values, with each software, are recapitulated in the Supp. Table S3 and all *CI90* boxplots are illustrated in the supp Fig. S1. *na*: not available. The complete list of the link and publications for each splicing tool are summarized in Suppl. Table S1.

Table 2

Exon number	% of skipping	Skill	Relative strength of splicing signals calculated with:					Ref
			SROOGLE	HSF	MaxEntScan	EX-SKIP	SKIPPY	
3	8 ± 3.2% *(n=7)*	*Weak*	*Strong*	*Strong*	*Strong*	*Weak*	*Weak*	This study
4	0.7 ± 0.5% *(n=6)*	*Strong*	*Strong*	*Strong*	*Strong*	*Strong*	*Strong*	This study
5	5 ± 0.7% *(n=4)*	*Weak*	*Strong*	*Weak*	*Strong*	*Strong*	*Strong*	This study
6	*nd (n=4)*	*Strong*	*Strong*	*Weak*	*Weak*	*Strong*	*Weak*	This study
10	35%	*Weak*	*Weak*	*Weak*	*Strong*	*Strong*	*Strong*	(Pagani, et al., 2003a)
11	*nd (n=6)*	*Strong*	*nd*	*Strong*	*Weak*	*Strong*	*Strong*	This study
13	15% 44 ± 1.5 % *(n=4)*	*Weak*	*Weak*	*Strong*	*Strong*	*Weak*	*Weak*	(Pagani, et al., 2003b) This study
14	*na*	*Weak*	*Strong*	*Weak*	*Strong*	*Strong*	*Strong*	(Aznarez, et al., 2003)
15	7 ± 0.8% *(n=9)*	*Weak*	*Strong*	*Strong*	*Weak*	*Strong*	*Weak*	(Hinzpeter, et al., 2010)
16	*nd (n=4)*	*Strong*	*Strong*	*Strong*	*Strong*	*Weak*	*Weak*	This study
17	14 ± 1% *(n=4)*	*Weak*	*Strong*	*Weak*	*Strong*	*Strong*	*Strong*	This study
21	*nd (n=3)*	*Strong*	*Strong*	*Weak*	*Strong*	*Strong*	*Strong*	This study
23	*nd (n=4)*	*Strong*	*Strong*	*Strong*	*Strong*	*Strong*	*Strong*	This study
Success rate of prediction			58%	62%	38%	54%	54%	

Table 2. Quantification of basal WT exon skipping. The percentage of exon skipping, plus or minus SD, is represented. Skill of each exon was assigned according to its basal skipping percentage using minigene (weak when superior or equal to 5%, strong if under or not detectable). Experiments were repeated three to nine times for each condition. Relative strength of splicing signals of each exon was tagged as weak when at least one splicing signal was predicted as weak by each *in silico* tool and strong if none. *nd*: not detected. *na*: not

available. Success rate of prediction was calculated by confronting *in vitro* skills with

predictions by each *in silico* tool.