



CHARGE syndrome: a recurrent hotspot of mutations in *CHD7* IVS25 analyzed by bioinformatic tools and minigene assays

Marine Legendre^{1,2} · Montserrat Rodriguez - Ballesteros¹ · Massimiliano Rossi³ · Véronique Abadie⁴ · Jeanne Amiel⁵ · Nicole Revencu⁶ · Patricia Blanchet⁷ · Frédéric Brioude⁸ · Marie-Ange Delrue⁹ · Yassamine Doubaï¹⁰ · Abdelaziz Sefiani¹⁰ · Christine Francannet¹¹ · Muriel Holder-Espinasse¹² · Pierre-Simon Jouk¹³ · Sophie Julia¹⁴ · Judith Melki¹⁵ · Sébastien Mur¹⁶ · Sophie Naudion⁹ · Jennifer Fabre-Teste¹⁷ · Tiffany Busa¹⁸ · Stephen Stamm¹⁹ · Stanislas Lyonnet⁵ · Tania Attie-Bitach⁵ · Alain Kitzis^{1,2} · Brigitte Gilbert-Dussardier^{1,2} · Frédéric Bilan^{1,2}

Received: 13 December 2016 / Revised: 17 August 2017 / Accepted: 29 August 2017 / Published online: 18 December 2017
© European Society of Human Genetics 2018

Abstract

CHARGE syndrome is a rare genetic disorder mainly due to *de novo* and private truncating mutations of *CHD7* gene. Here we report an intriguing hot spot of intronic mutations (c.5405-7G > A, c.5405-13G > A, c.5405-17G > A and c.5405-18C > A) located in *CHD7* IVS25. Combining computational *in silico* analysis, experimental branch-point determination and *in vitro* minigene assays, our study explains this mutation hot spot by a particular genomic context, including the weakness of the IVS25 natural acceptor-site and an unconventional lariat sequence localized outside the common 40 bp upstream the acceptor splice site. For each of the mutations reported here, bioinformatic tools indicated a newly created 3' splice site, of which the existence was confirmed using pSpliceExpress, an easy-to-use and reliable splicing reporter tool. Our study emphasizes the idea that combining these two complementary approaches could increase the efficiency of routine molecular diagnosis.

Introduction

CHARGE syndrome ((CS), MIM#214800) refers to a rare autosomal dominant polymalformative syndrome due to haploinsufficiency of the gene encoding chromodomain helicase DNA-binding protein 7 (*CHD7*, MIM*608892). Most *CHD7* mutations are truncating [1] while missense mutations are present in 8% of cases. Intronic variations not affecting the canonical dinucleotides of the acceptor or donor splice sites are difficult to interpret. Bioinformatic prediction tools exist to evaluate their putative impact on the splicing mechanism but require experimental procedures to be confirmed. *In vitro* experiments like minigene assays

could be a reliable option. pSpliceExpress [2] is an efficient vector allowing fast generation of splicing reporter constructs that use direct recombination between PCR products of the patient's genomic DNA and a modified Exontrap vector. In this work, we report a series of 22 CS patients carrying one of the following nucleotide variations located in *CHD7* IVS25: c.5405-7G > A (rs398124321), c.5405-13G > A (rs1131690787), c.5405-17G > A (rs794727423) and c.5405-18C > A (rs199981784). The four variations were predicted to create a new 3' splice site (3'ss). Accumulation of such variations within IVS25 was surprising because most *CHD7* mutations are private in CS and only a few recurrent nonsense or complex frameshift mutations have previously been reported.

Marine Legendre and Montserrat Rodriguez - Ballesteros contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41431-017-0007-0>) contains supplementary material, which is available to authorized users.

✉ Frédéric Bilan
frederic.bilan@univ-poitiers.fr

Extended author information available on the last page of the article

Material and methods

Patients and molecular genetics data

We studied by routine molecular diagnosis [3, 4] 23 patients with clinical CS according to Verloes' updated diagnostic criteria. *CHD7* NM_017780.2 and NG_007009.1 were used

respectively for nucleotide reference and exon numbering. All variants reported here are submitted in ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>) and could be viewed using the following accession number: SCV000579497 (c.5405-2A > G), SCV000579496 (c.5405-7G > A), SCV000575887 (c.5405-13G > A), SCV000579495 (c.5405-17G > A) and SCV000575888 (c.5405-18C > A). CS patients described in this study and their corresponding variants were also submitted in CHD7 database (<https://molgenis51.gcc.rug.nl/>) under the following references: M1345 (c.5405-2A > G), patient 1131; M184 (c.5405-7G > A), patients 1119 to 1130; M1346 (c.5405-13G > A), patient 1118; M41 (c.5405-17G > A), patients 1111 to 1117 and M1352 (c.5405-18C > A), patients 1109 to 1110.

Software prediction tools

All software is freely available and listed in Supplementary table 1. A brief description of each program and the settings used for splicing analysis is detailed in Supplementary material and methods.

Generation of minigene reporter

We used pSpliceExpress construct, according to the authors experimental procedure [2] (see Supplementary material and methods for details). The different steps of minigene generation are summarized in Fig. 1a. PCR primers are listed in Supplementary table 2.

Minigene product analysis using fluorescent capillary electrophoresis

RT-PCR fragments were obtained with E2F and E3R dye labeled primers using a limited number of cycles so that the DNA yield generated in the exponential step could be comparable between different samples. After PCR, amplicons were differentiated in size by capillary electrophoresis using an ABI PRISM 3130 DNA sequencer (ThermoFisher Scientific, Courtaboeuf, France). Data were visualized using Genemapper software (ThermoFisher Scientific).

Lariat RT-PCR to determine the branch point

We performed nested lariat RT-PCR according to Gao *et al* [5] (see Supplementary material and methods for details) to amplify a fragment spanning the 2'-5' phosphodiester bond at the branch-point. Lariat RT-PCR was performed using primers C and D for the first round and primers A and B for the second round (Fig. 2b).

Results

Clinical findings

All 23 probands in our series fulfilled the diagnostic criteria of CS defined by Verloes [6] (Supplementary table 3). Eighteen had typical CS. In the two familial cases (cases 3a and 6a, Supplementary figure 1, Supplementary clinical data and Supplementary table 4), index cases had a typical CS and inherited c.5405-17G > A variation from their mildly affected mothers that failed to fulfill Verloes' diagnostic criteria.

In silico splice site prediction

We used 8 freely available bioinformatic tools particularly MaxEntScan [7] and Human Splicing Finder [8] (HSF). As depicted in Supplementary table 5, several tools did not detect the native 3'ss. c.5405-7A > G was clearly revealed as a splicing mutation by HSF only. Other mutations were more easily detected by all software.

Minigene product splicing assays

Minigene analysis was performed by RT-PCR (Fig. 1a) and the amplification products were analyzed by fluorescent capillary electrophoresis (Fig. 1b). The four intronic mutations showed a higher amplicon size than the one obtained for the wild type sequence (391 bp) ranging from 396 to 407 bp. The 261 bp amplicon, corresponding to an alternative splicing between exon 2 and exon 3 of rat *Insulin* gene, could be observed in each RT-PCR sample. Direct sequencing of each RT-PCR product confirmed prediction software analysis (Fig. 1c). The 4 IVS25 variations introduced respectively 5, 11, 15, and 16 intronic bp at 5' end of exon 26. Electropherograms did not show any trace of wild type exon 26 sequence. Three of these variations, c.5405-7G > A, c.5405-13G > A and c.5405-18C > A, led to frameshift mutations which could be named respectively p.(Gly1802fs), p.(Tyr1803fs) and p.(Tyr1803fs). Only c.5405-17G > A variation induced an in-frame five amino acid insertion, p.(His1801_Gly1802insAspGlyHisGlyThr). Furthermore c.5405-2A > G variation (patient 23) disrupting native 3'ss is characterized by a shorter *CHD7* exon 26 (Fig. 1b), suggesting the use of a cryptic 3'ss localized within exon 26, a finding confirmed by DNA sequencing.

Branch point analysis

Branch-point (BP) position was obtained from a genome wide study providing the first map of splicing BP in the human genome [9]. Through exploration, we found a putative BP at chr8:g.61762997A, corresponding to c.5405-

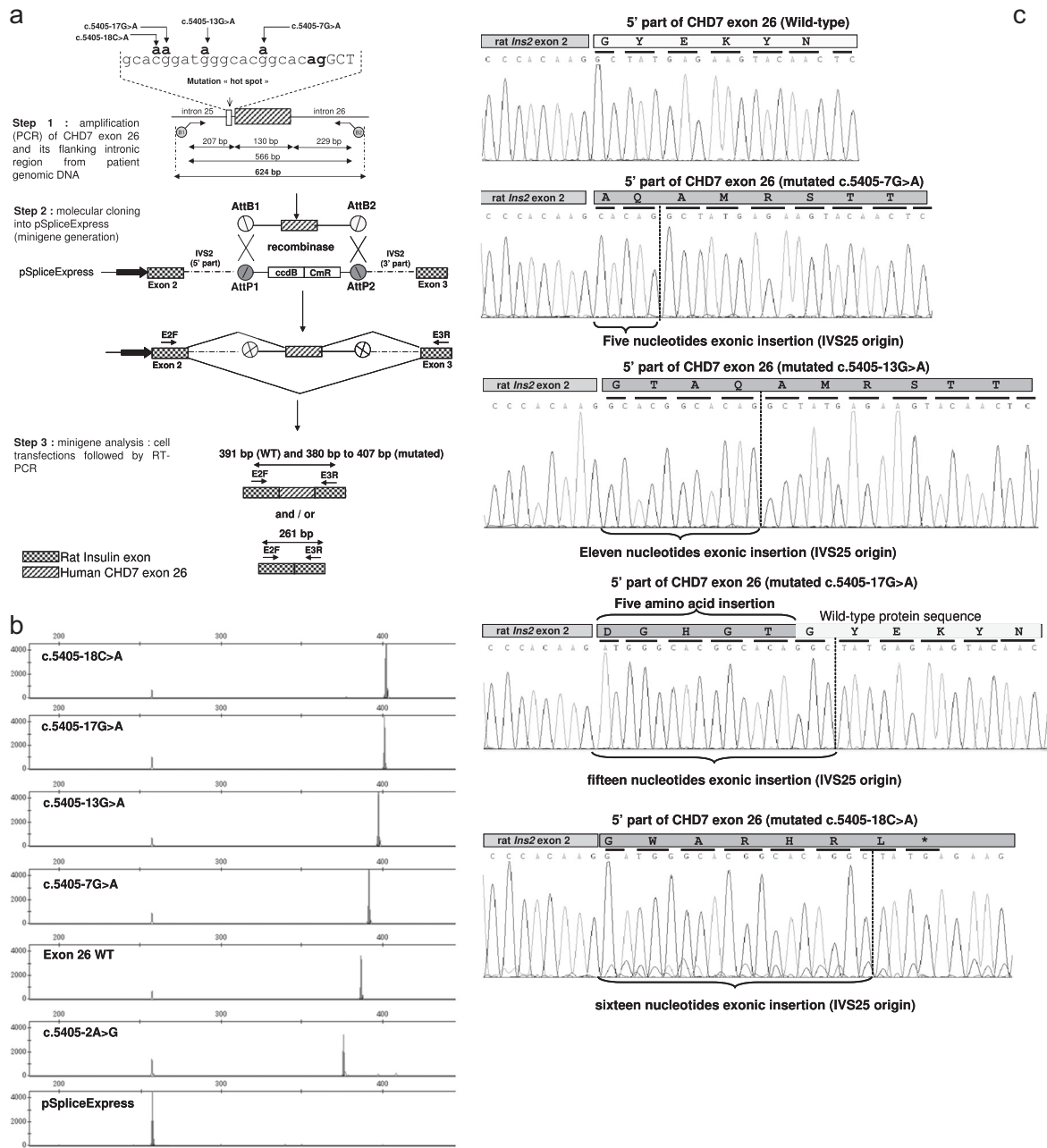


Fig. 1 CHD7 IVS25 recurrent mutation analysis by minigene assays. A- Schematic representation of CHD7 exon 26 and its flanking regions. The last 21 nucleotides of IVS25 DNA sequence are indicated in lower cases and the consensus acceptor splice site in bold. The first nucleotides of exon 26 appear in upper case. Double arrows indicate different lengths in base pair. Simple arrows show the localization of PCR primers used for minigene generation (B1 and B2 refer to the attB1 and attB2 DNA sequences which were added in 5' end of forward and reverse primers respectively). Overview of minigene assays using pSplice Express. Step 1 Amplification of the region of interest (amplicon of 624 bp). Two primers forward and reverse are used to amplify a part of the genomic DNA that harbors exon of interest (hatched rectangle) and its flanking intronic region (around 200 bp). The primers have recombination sites (AttB1, AttB2) that are indicated by circles. Step 2: Construction of the splicing reporter using pSpliceExpress. The PCR fragment is recombined in vitro with pSpliceExpress vector. The vector contains Cm and ccdB selection markers that are used to isolate recombined clones. The inserted

DNA is flanked by two constitutive rat insulin exons, indicated by checkered rectangle. The transcript is driven by a RSV LTR promoter (full black arrow) and the subcloned genomic fragment flanked by attL sites, which are generated by the recombination of attB and attP sites. Step 3: Analysis of the splicing reporter. The minigene construct is transfected into HeLa cells. The RNA generated is determined by RT-PCR, using E2F and E3R primers (indicated by small arrows). The mRNA structures, indicated below the gene structure are expected to be generated by the construct. The range of amplicon length is indicated above double arrows. B- Minigene analysis of all recurrent IVS25 mutations by RT-PCR by fluorescent capillary electrophoresis. 380 to 407 bp PCR products correspond to CHD7 exon 26 minigenes, 261 bp amplicon corresponds to an alternative splicing between exon 2 and exon 3 of rat Insulin gene. C- cDNA organization and DNA sequencing of each RT-PCR amplicon obtained from each minigene analysis. Letters in bold indicate, for each minigene, the CHD7 protein sequence encoded by exon 26

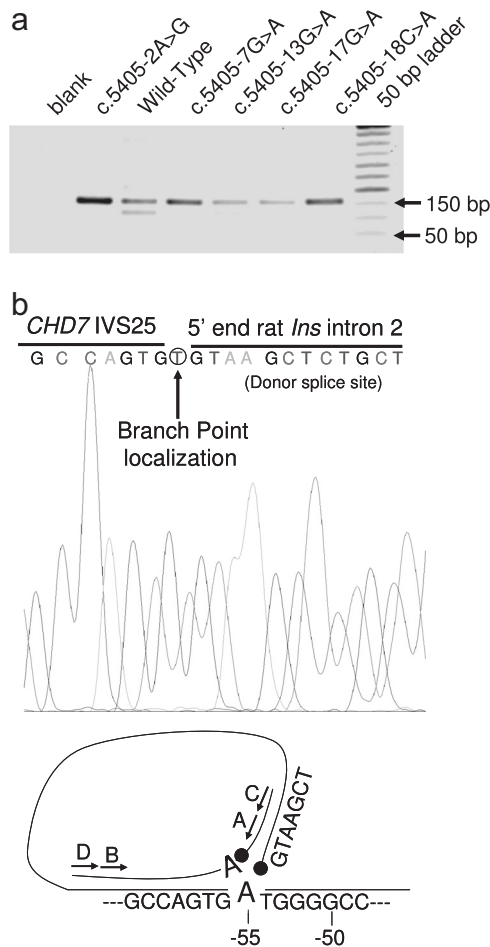


Fig. 2 *CHD7* IVS25 branch point determination by lariat RT-PCR and minigene assays. **A** Nested lariat RT-PCR and minigene assays. Amplicons of 159 bp correspond to the relevant branch point sequence. The smaller amplicon that can be observed lane 3 (Fig. 2a) corresponds to an irrelevant product of splicing. **B** DNA sequencing of 159 bp amplicon obtained from nested lariat RT-PCR. A misincorporated “A” nucleotide (corresponding to the circled nucleotide T) indicates the branch point position. Arrows on the schematic explanation for lariat PCR indicate PCR primer localization: first round of lariat RT-PCR was performed with D and C, respectively located 163 nucleotide upstream *CHD7*-IVS25 3’ss and 97 bp upstream IVS2 rat *Ins* 5’ss. Nested PCR was performed using primers B and A, located respectively 125 bp upstream IVS25 3’ss and 38 bp downstream rats *Ins* 5’ss

55A. To confirm this finding obtained by RNA-seq, we used the lariat RT-PCR method [5]. We obtained a major amplicon of 159 bp (Fig. 2a) of which the DNA sequencing indicates a misincorporated “A” nucleotide at the BP position (Fig. 2b), a finding supporting the notion that c.5405-55A is the natural BP. Using the same experiment, we demonstrated that each IVS25 variation does not affect BP localization (Fig. 2a). To confirm these results, we carried

out minigene assays that selectively disrupted 4 putative BP positions predicted by the SVM-BP finder on-line tool [10]: c.5405-55A > G, c.5405-67A > G, c.5405-98A > G and c.5405-108A > G. As depicted in Fig. 3, the weakest peak signal intensity corresponding to the exon 26 inclusion was obtained for c.5405-55A > G minigene in comparison with the wild-type IVS25 and other constructs. Concomitantly, alternative splicing generating rat *Ins* exon 2 / exon 3 appeared weaker for these minigenes than for the c.5405-55A > G construct. These data strengthen the notion designating c.5405-55A as the native BP.

Discussion

Pathogenic effects of variants

In this work, all IVS25 variations have a *de novo* origin or co-segregate with the disease, data consistent with variants affecting *CHD7* function. Most of bioinformatic programs indicate that IVS25 3’ss has a very weak splicing signal. This particular configuration could be the first factor explaining the hot spot of mutation toward *CHD7* IVS25. The 4 intronic variations were predicted as “pathogenic”, notably by HSF. The use of pSpliceExpress [2] confirmed the effect on the mRNA processing and indicated that IVS25 variants activated a new 3’ss without any alternative splicing. Surprisingly, these splicing mutations were found in 22 index CS cases among 460 *CHD7* mutated patients from the French cohort (5%). So, IVS25 3’ss is the first real hot spot of mutations not related to its genomic size. In this hot spot, c.5405-7G > A and c.5405-17G > A are inexorably recurrent.

CHD7 IVS25 belongs to the “AG independent” intron class and harbors a distant branch point (dBP)

Native BP localization outside the classical 40-50 nucleotide range upstream the 3’ss may be the second explanation for the hot spot of mutations toward *CHD7* IVS25. Introns with dBP are expected to be vulnerable to mutations introducing a new “AG” sequence within the AG Exclusion Zone (AGEZ) [11] area, a phenomenon accounting for almost 42% of intronic mutations [12]. One characteristic of dBPs is their “AG” independent splicing mechanism. Sustaining this hypothesis, the disruption of native 3’ss highlighted the use of an exonic cryptic 3’ss instead of exon skipping. The IVS25 cryptic 3’ss used was the second “AG” dinucleotide downstream native 3’ss. This result was not surprising since the nucleotide preceding the “AG” has a striking influence upon competition between closely spaced AGs (the order of competitiveness is CAG ≥ TAG > AAG > GAG). Here, the first “NAG” trinucleotides is “GAG” while

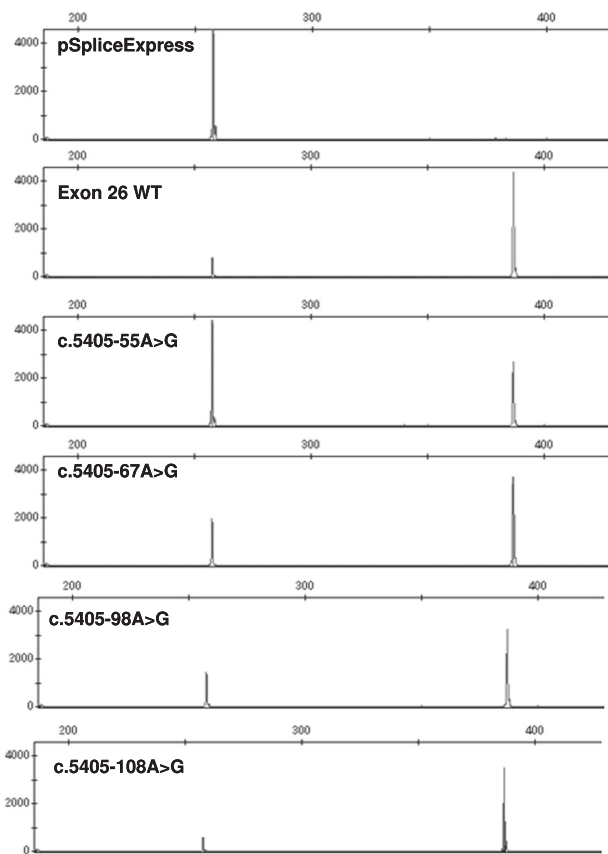


Fig. 3 Selective disruption of branch point candidates using minigene assays. RT-PCR from each minigene is analyzed by fluorescent capillary electrophoresis. Lower amplicon (261 bp) corresponds to rat insulin minigene (exon 2 plus exon 3) and upper amplicon (391 bp) indicates full-length minigene containing the *CHD7* sequence

the second one is “AAG”, leading to a more favorable configuration. Accordingly, all IVS25 mutations presented in this work created 3’ss in a similar favorable context.

Our study provided molecular evidence about an atypical configuration of the splicing core elements of the 3’ IVS25 sequence. There are no strict genotype/phenotype relations in CS and we have shown that splicing variants are responsible for variable phenotypes even within the same family or even if the open reading frame is conserved. For confirmation of a molecular diagnosis, our work demonstrates that pSpliceExpress is a reliable and easy-to-use tool to study nucleotide variation on the splicing mechanism and

confirms the importance of AGEZ screening in genetic diseases.

Acknowledgements We thank B. Manière and Z. Colo-Safi for their technical assistance. We are thankful to national PHRC “Etude clinique, neuropsychologique et moléculaire du syndrome CHARGE” for collaboration and grant. We also wish to thank Jeffrey Arsham, an American medical translator, for reading and reviewing our original English-language manuscript.

Conflict of interest The authors declare no conflict of interest

References

- Janssen N, Bergman JEH, Swertz MA, *et al.* Mutation update on the *CHD7* gene involved in CHARGE syndrome. *Hum Mutat.* 2012;33:1149–60.
- Kishore S, Khanna A, Stamm S. Rapid generation of splicing reporters with pSpliceExpress. *Gene.* 2008;427:104–10.
- Bilan F, Legendre M, Charraud V, *et al.* Complete screening of 50 patients with CHARGE syndrome for anomalies in the *CHD7* gene using a denaturing high-performance liquid chromatography-based protocol: new guidelines and a proposal for routine diagnosis. *J Mol Diagn JMD.* 2012;14:46–55.
- Sanlaville D, Etchevers HC, Gonzales M, *et al.* Phenotypic spectrum of CHARGE syndrome in fetuses with *CHD7* truncating mutations correlates with expression during human development. *J Med Genet.* 2006;43:211–7.
- Gao K, Masuda A, Matsuura T, Ohno K. Human branch point consensus sequence is yUnAy. *Nucleic Acids Res.* 2008;36:2257–67.
- Verloes A. Updated diagnostic criteria for CHARGE syndrome: a proposal. *Am J Med Genet A.* 2005;133A:306–8.
- Yeo and Burge. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol.* 2004;11:377–94.
- Desmet F-O, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 2009;37:e67.
- Mercer TR, Clark MB, Andersen SB, *et al.* Genome-wide discovery of human splicing branchpoints. *Genome Res.* 2015;25:290–303.
- Corvelo A, Hallegger M, Smith CWJ, Eyraas E. Genome-wide association between branch point properties and alternative splicing. *PLoS Comput Biol.* 2010;6:e1001016.
- Gooding C, Clark F, Wollerton MC, Grellscheid S-N, Groom H, Smith CWJ. A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones. *Genome Biol.* 2006;7:R1.
- Vorechovský I. Aberrant 3’ splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Res.* 2006;34:4630–41.

Affiliations

Marine Legendre^{1,2} · Montserrat Rodriguez - Ballesteros¹ · Massimiliano Rossi³ · Véronique Abadie⁴ · Jeanne Amiel⁵ · Nicole Revenu⁶ · Patricia Blanchet⁷ · Frédéric Brioude⁸ · Marie-Ange Delrue⁹ · Yassamine Doubaj¹⁰ · Abdelaziz Sefiani¹⁰ · Christine Francannet¹¹ · Muriel Holder-Espinasse¹² · Pierre-Simon Jouk¹³ · Sophie Julia¹⁴ · Judith Melki¹⁵ · Sébastien Mur¹⁶ · Sophie Naudion⁹ · Jennifer Fabre-Teste¹⁷ · Tiffany Busa¹⁸ · Stephen Stamm¹⁹ · Stanislas Lyonnet⁵ · Tania Attie-Bitach⁵ · Alain Kitzis^{1,2} · Brigitte Gilbert-Dussardier^{1,2} · Frédéric Bilan^{1,2}

- ¹ Service de Génétique, Centre de Référence Anomalies du Développement de l'Ouest, CHU Poitiers, France
- ² EA3808 CiMoTheMA Université Poitiers, Poitiers, France
- ³ Service de génétique, Centre de Référence Anomalies du Développement, Hospices Civils de Lyon et INSERM U1028, CNRS UMR5292, Centre de Recherche en Neurosciences de Lyon, GENDEV Team, Université Claude Bernard Lyon 1, Bron, France
- ⁴ Service de Pédiatrie Générale, Hôpital Necker Enfants-Malades, AP-HP, Paris, France
- ⁵ Département de Génétique, Hôpital Universitaire Necker-Enfants Malades, AP-HP, Institut Imagine, UMR-1163 INSERM-Université Paris Descartes, Paris, France
- ⁶ Center for Human Genetics, Cliniques universitaires St Luc, Université catholique de Louvain, Brussels, Belgium
- ⁷ Département de Génétique Médicale, Hôpital Arnaud de Villeneuve, CHU Montpellier, France
- ⁸ Sorbonne Universités, UPMC Univ Paris 06, INSERM UMR_S938, Centre de Recherche Saint Antoine and AP-HP, Hôpitaux Universitaires Paris Est, Hôpital Trousseau, Service d'Explorations Fonctionnelles Endocriniennes, Paris, France
- ⁹ Service de Génétique Médicale, GH Pellegrin, CHU Bordeaux, France
- ¹⁰ Département de Génétique Médicale, Institut National d'Hygiène, Centre de Génomique Humaine, Faculté de Médecine et de Pharmacie de Rabat, Mohammed V University in Rabat, Rabat, Morocco
- ¹¹ Service de Génétique Médicale, CHU Estaing, Clermont-Ferrand, France
- ¹² Service de Génétique Clinique Guy Fontaine - Hôpital Jeanne de Flandre, CHRU Lille, France
- ¹³ Département Génétique & Procréation, Hôpital Couple-Enfant, CHU Grenoble, France
- ¹⁴ Service de Génétique Médicale, Hôpital Purpan, CHU Toulouse, France
- ¹⁵ CHU Bicêtre, Unité de Génétique Médicale and UMR-1169, Inserm, Le Kremlin Bicêtre, France
- ¹⁶ Clinique de médecine néonatale, Hôpital Jeanne de Flandre, CHU Lille, France
- ¹⁷ Département de génétique médicale, Hôpital Robert Debré, APHP, Paris, France
- ¹⁸ Département de Génétique Médicale, Hôpital d'enfants de la Timone, Marseille, France
- ¹⁹ Department of Molecular and Cellular Biochemistry, University of Kentucky, Lexington, USA