



Deep intronic variation in splicing regulatory element of the *ERCC8* gene associated with severe but long-term survival Cockayne syndrome

Audrey Schalk¹ · Géraldine Greff¹ · Nathalie Drouot² · Cathy Obringer³ · Hélène Dollfus^{3,4} · Vincent Laugel^{3,5} · Jamel Chelly^{1,2} · Nadège Calmels¹

Received: 31 March 2017 / Revised: 19 July 2017 / Accepted: 31 August 2017 / Published online: 8 February 2018

© European Society of Human Genetics 2018

Abstract

Cockayne syndrome is an autosomal recessive multisystem disorder characterized by intellectual disability, microcephaly, severe growth failure, sensory impairment, peripheral neuropathy, and cutaneous sensitivity. This rare disease is linked to disease-causing variations in the *ERCC6* (*CSB*) and *ERCC8* (*CSA*) genes. Various degrees of severity have been described according to age at onset and survival, without any clear genotype-phenotype correlation. All types of nucleotide changes have been observed in CS genes, including splice variations mainly affecting the splice site consensus sequences. We report here the case of two brothers from a consanguineous family presenting a severe but long-term survival phenotype of Cockayne syndrome. We identified in the patients a homozygous deep intronic nucleotide variation causing the insertion of a cryptic exon in the *ERCC8* (*CSA*) transcript, by modifying intronic regulatory elements important for exon definition. The pathogenesis of the nucleotide variant NG_009289.1(NM_000082.3):c.173+1119G>C was validated in vitro with a reporter minigene system. To our knowledge, these are the first Cockayne patients described with this kind of disease-causing variation, though molecular mechanism underlying early onset symptoms and unexpected slow raise of progression of the disease remain to be elucidated.

Introduction

Cockayne syndrome (CS) (Orpha number 191, OMIM 216400 and 133540) is an autosomal recessive multisystem disorder with an estimated incidence of 2.7 per million live births in the Western Europe [1]. It belongs to the family of nucleotide excision repair (NER)-related disorders together with xeroderma pigmentosum and trichothiodystrophy. CS is characterized by progressive and severe growth failure, microcephaly, intellectual disability, cerebellar ataxia, spasticity, peripheral demyelinating neuropathy, sensory impairment, cutaneous photosensitivity, dental decay, and recognizable facial appearance with deep sunken eyes [2, 3]. The clinical spectrum of CS encompasses a wide range of severity that includes the classical or “moderate” form CS type I, the early-onset cases with congenital signs of the disease, named CS type II and the late-onset cases who may only be affected in late childhood or even adulthood, named CS type III. The mean age at death was 8.4 years but varies according to the CS type [3].

CS cells show a specific defect in transcription-coupled DNA repair (TC-NER), a subpathway of NER involved in

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41431-017-0009-y>) contains supplementary material, which is available to authorized users.

✉ Nadège Calmels
nadege.calmels@chru-strasbourg.fr

- ¹ Laboratoire de Diagnostic Génétique, Institut de Génétique Médicale d'Alsace (IGMA), Hôpitaux Universitaires de Strasbourg, 1 place de l'hôpital, Strasbourg 67091, France
- ² Département Médecine translationnelle et neurogénétique, Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS U7104, INSERM U964, Université de Strasbourg, Illkirch 67400, France
- ³ Laboratoire de Génétique Médicale – INSERM U1112, Institut de Génétique Médicale d'Alsace (IGMA), Faculté de médecine de Strasbourg, 11 rue Humann, Strasbourg 67000, France
- ⁴ Service de Génétique Médicale, Institut de Génétique Médicale d'Alsace (IGMA), Hôpitaux Universitaires de Strasbourg, 1 avenue Molière, Strasbourg 67200, France
- ⁵ Service de Pédiatrie, Hôpitaux Universitaires de Strasbourg, 1 avenue Molière, Strasbourg 67200, France

the removal of UV-induced DNA lesions in actively transcribed genes [4, 5]. Since the early 1980s this cellular testing using the particular sensitivity of CS cells to UV-light is used to confirm the clinical diagnosis of CS.

The two major genes responsible for the disorder, *ERCC6* (*CSB*) and *ERCC8* (*CSA*), were identified in the mid-1990s [6, 7]. The encoded proteins, CSB and CSA are respectively a DNA-dependent ATPase and a WD40 repeat component of an E3-ubiquitin-ligase complex. CSB protein is believed to assist RNA polymerase II in dealing with transcription blocks caused by DNA damage, but others roles have been suggested, including the repair of oxidative damage in DNA and roles in mitochondrial DNA metabolism (reviewed in Karikkineth et al., [8]). CSA may be involved in the targeted ubiquitination of proteins such as CSB [9]. A role for the CS proteins in expression of neuronal genes and thereby in neuronal differentiation has been recently suggested [10].

ERCC6 (*CSB*) and *ERCC8* (*CSA*) genes account for about two thirds of the CS cases for *ERCC6* (*CSB*) and one third for *ERCC8* (*CSA*), respectively [3, 11]. Although it is usually acknowledged that *ERCC6* (*CSB*) and *ERCC8* (*CSA*) patients are clinically indistinguishable, it seems that a larger number of *ERCC6* (*CSB*) patient are affected by more severe forms of the disease, whereas disease-causing variations in *ERCC8* (*CSA*) appear to be preferentially associated with the classical form type I [11]. The complex genotype-phenotype correlations for *ERCC6* (*CSB*) and *ERCC8* (*CSA*) variations still require to be clarified.

All types of disease-causing variants have been detected in CS patients: missense variations, nonsense variations, frameshift small deletions or insertions, large deletions, as well as splicing variations [2, 11]. The majority of the described splicing variations affects the splice site consensus sequences and usually involves one of the six bases surrounding the exon/intron junction. A mid-intronic variation creating cryptic exon has only been described once in a CSA patient (patient CS333VI in Laugel V et al., [11]). In this case the homozygous deep intronic substitution was shown to activate a cryptic splice site within the intron.

We describe here a deep-intronic *ERCC8* (*CSA*) substitution resulting in the insertion of a new cryptic exon in the mRNA of two CS affected siblings, not by activating a cryptic splice site but by modifying intronic regulatory elements important for exon definition.

Materials and methods

Patients and samples

The reported cases are two brothers born from healthy Turkish parents. They have been already mentioned in a

previous study as patient CS466VI and CS466VI_2 [11]. Mother and father are first cousins once removed (the father of the two patients married the daughter of his first cousin).

Patient 1 (CS466VI) (Fig. 1a) was the second child of the couple. He was born at term. His birth weight was 2650 g (-2 SD), length was 48 cm (-2 SD) and his head circumference was 33 cm (-3 SD). The patient's weight reached a plateau at the age of 2 years. His body weight was 7.9 kg (-3.5 SD), size was 78 cm (-2.75 SD), and head circumference was 41.5 cm (-7 SD). The patient displayed severe psychomotor delay. He was able to sit with support but was never able to walk and did not develop any speech. Ophthalmological examination revealed poor visual contact and persistent pendular nystagmus. Hearing reactions seemed to be normal. The patient had no skin abnormality. Neurological examination at the age of 7 revealed progressive spasticity of lower limbs flexor muscles. At the age of 14, brain magnetic resonance imaging (MRI) showed white matter atrophy and abnormalities in myelination (Fig. 1b). A gastrostomy was performed at the age of 19 years. At the age of 20, type II diabetes was discovered. Ophthalmological examination revealed bilateral cataracts, atrophy of the optic nerve and retinal dystrophy. The development of numerous dental caries required the removal of all teeth at the age of 15. At the last clinical assessment, the patient was 23 years old and was bedridden with poor social interactions. He passed away at 23 years of age.

Patient 2 (CS466VI_2) (Fig. 1c) was the third child of the couple. Pregnancy was marked by intrauterine growth retardation and vermian hypoplasia on antenatal ultrasound. The patient was born at 40 weeks gestation by caesarean section due to signs of fetal distress. Birth weight was 2490 g (-3 SD), size was 48 cm (-1 SD), and head circumference was 30.5 cm (-3 SD). His thumbs were adducted in his palms. The patient presented severe growth retardation: at 4 months of age, his weight was 3.7 kg (-3 SD), size 54 cm (-3 SD) and head circumference 33 cm (< -4 SD). Sensorial examination revealed bilateral deafness, erratic movements of both eyes and poor visual contact. Neurological examination showed progressive spasticity of lower limbs flexor muscles. Brain MRI confirmed the cerebellar vermis hypoplasia. Furthermore the patient presented severe eczematous lesions. The implementation of a gastrostomy was decided at the age of 11 years. At the last examination, the patient was 17 years old. His weight was stable at 22 kg (-5 SD) and his microcephaly was still getting worse (44.5 cm, -7 SD).

Written consent for genetic testing was obtained from the parents, for themselves and for their affected and healthy children. Peripheral blood samples were taken from all family members for DNA extraction. Primary fibroblasts cultures were obtained from patients skin biopsies.

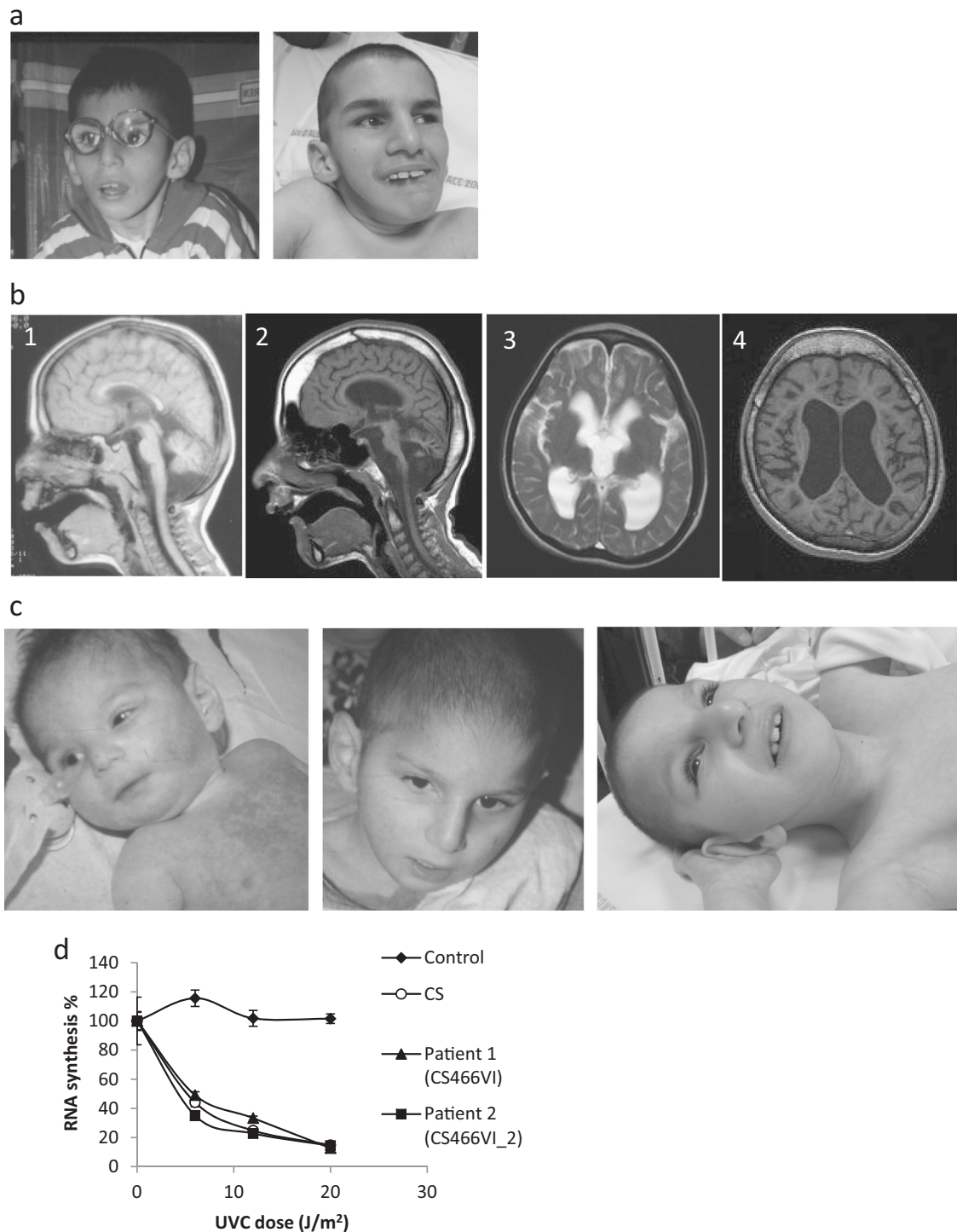


Fig. 1 **a** Patient 1 (CS466VI) at 8 and 16 years old. **b** Brain magnetic resonance imaging of patient 1 (CS466VI) (1–2) Sagittal picture at 4 years (1) and 14 years old (2) showing a gradual atrophy of cerebellum and brain stem. (3–4) Axial T2 (3) and Flair-weighted (4) pictures at 14 years old showing white matter atrophy and abnormal myelination. **c** Patient 2 (CS466VI_2) at 1 month, 5 years, and 16 years old. Picture

publication has been authorized by the family. **d** Recovery of RNA synthesis (RRS) measured 23 h after UV irradiation in fibroblasts from patients 1 (CS466VI) and 2 (CS466VI_2), compared to control cell line (individual non affected with CS) and CSA-mutated cell line (CS) (CS patient homozygous carrier of the NM_000082.3(ERCC8): c.316C>T p.Gln106* variation). (Color figure online)

DNA repair functional assays

Recovery of RNA synthesis (RRS), evaluating the transcription-coupled repair pathway (TC-NER), was performed on primary fibroblasts cultures using fluorescent non-radioactive assay as described previously [12].

Western blot analysis

Fibroblasts whole-cell extracts were separated on SDS/8% polyacrylamide gel and transferred to Hybond nitrocellulose membrane (Amersham Biosciences, Little Chalfont, United Kingdom) in blotting buffer (Tris 25 mM, glycine 0.2 mM, methanol 20%). After blocking (60 min in 5% milk), the blot was incubated overnight with the anti-CSA antibody (dilution 1:1000) purchased from Abcam (product code ab96780, Cambridge, United Kingdom) and 45 min with anti-goat secondary antibody (dilution 1:5000), then revealed by ECL western blotting detection system (Amersham Biosciences) [13].

Sanger sequencing analysis

Genomic DNA was extracted from whole blood and from fibroblast cultures using standard procedure (Qiagen, Hilden, Germany). RNA was extracted from fibroblast using TRI Reagent and was then reverse transcribed by SuperScript II reverse transcriptase (Invitrogen, Carlsbad, California, United States). Complementary DNA was amplified by overlapping fragments covering the whole *ERCC8 (CSA)* transcript. (Supplementary data 1). PCR products obtained from patients and from an unaffected control were analyzed by electrophoresis on a 2% agarose gel or by using the LapChip GX (Caliper) microelectrophoresis system.

ERCC8 (CSA) genomic coding sequence as well as *ERCC8 (CSA)* transcript were studied by PCR amplification and Sanger sequencing. A primer pair has been used for the amplification of the full length *ERCC8 (CSA)* cDNA. Because of the large size of the amplicon, two other primers were used to allow the cDNA sequencing in four overlapping *ERCC8 (CSA)*. Primer sequences are available in Supplementary data 1. PCR products were analyzed on conventional agarose gel or on the LapChip GX (Caliper) microelectrophoresis system and then purified (Illustra ExoProStar S, VWR or Qiaquick PCR purification kit). Sequence reaction were performed with the Big Dye v1.1 sequencing standard kit (Life Technologies Inc., Carlsbad, California, United States), purified by Big Dye Xterminator purification kit (Life Technologies Inc.) and then analyzed on a capillary sequencer (3130 XL or 3500 Applied Biosystems). Sequences of the patients were compared to the reference *ERCC8 (CSA)* sequence

(GRCh37, NM_0000082.3) using Sequence Pilot software (JSI). Exons are numbered as in NG_009289.1.

Minigene construction and expression

Cellular assay to study the expression of a minigene was used to test the splicing effects of the variations found in the patients in intron 2 of the *ERCC8 (CSA)* gene.

The intronic sequence of interest (chr5: g.60.223.150_60.223.725/c.173+966_173+1541) was amplified by PCR using patient (double variant) and control (wild-type) DNA blood sample. The TA Cloning kit (Life Technologies Inc.,) was used for directly inserting the PCR product into the pCR2.1 plasmid vector. Single variant constructs were obtained from the wild-type-TA plasmid by in vitro mutagenesis (Supplementary data 1 for the primers used). Each of the four genomic fragments (wild-type, double variant and single variants) were then subcloned into the pSPL3B vector, between two exons of rabbit β globin gene, using *Not I* and *Sac I* digestion/ligation. All constructs were verified by sequencing. Minigenes carrying either the wild-type sequence or the double variant or the single variants were transfected separately into Neuro2a cells, primary human fibroblasts and HeLa cells by lipofection using lipofectine2000 transfect reagent (Life Technologies Inc.) according to the manufacturer's instruction. The splicing patterns corresponding to the wild-type and to the mutant alleles were then compared by reverse transcriptase-PCR analysis of the RNA from transfected cells. Total cellular RNAs (isolated using the TRIzol Reagent) were treated with DNase and then reverse transcribed using transcriptase (Life Technologies Inc.). The complementary DNA were amplified with primers specific to rabbit β globin sequences and the cryptic exon (Supplementary data 1). The PCR products were resolved on a 1% agarose gel and sequenced using the same primers.

In silico splicing prediction

To identify and evaluate the strength of potential 5' and 3' splice sites, on the partial sequence of intron 2, we have used the Splice Site finder (SSF, <http://www.umd.be/searchSpliceSite.html>), MaxEnt (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html) [14], BDGP Splice site prediction by Neural Network (BDGP/NNSplice: http://www.fruitfly.org/seq_tools/splice.html) [15] and Human Splicing Finder (HSF, <http://www.umd.be/HSF/>) [16] predictors via the Alamut. Visual 2.8 software (Interactive biosoftwares). By default configuration, scores are significant when they are upper than 70, 0, 0.4, and 65, respectively.

To predict the location of serine-arginine (SR)-protein-specific putative ESE (Exonic Splicing Enhancers) in the

sequence of our patients, a web-based program called “ESE finder3.0” (<http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi>) has been used. This program scores any sequence for the presence of consensus motifs thought to serve as binding sites for the specific SR splicing factors SRSF1, SRSF2, SRSF5, and SRSF6.

Exome sequencing

Exome sequencing (ES) of patient 1 and 2 was performed using a solution exome capture kit (SeqCap EZ MedExome Target Enrichment Kit, Roche, USA) and Illumina HiSeq sequencing platforms (Illumina, Inc., San Diego, California, USA) to generate paired end reads sequences. Sequence analysis pipeline included alignment with BWA-mem against the hg19 genome reference, GATK for base recalibration, indel realignment and the GATK Unified Genotyper for variant calling. Variants calling were generated in the standard VCF version and filtered for common variants between the two affected brothers, and excluding known common variants in the dbSNP131 database, 1000 genomes, Exome Variant Server, the Exome Aggregation Consortium (ExAC) and a local ES databases. Subsequent variant annotation and ranking was performed using VaRank v1.3.5 [17] configured with Alamut Batch (Interactive biosoftware). Variant Call Format (VCF) files are available on request.

Results

Molecular characterization of the disease-causing variation

In order to characterize disease-causing events in two CS brothers born from consanguineous parents, the first step of our diagnostic strategy has been to test DNA repair activity using functional assays on the patients’ fibroblasts. Both patients presented a severely decreased RRS (recovery of RNA synthesis) (Fig. 1d), a pattern in accordance with the defect of the TC-NER pathway in CS.

In order to identify the disease-causing variation, DNA sample of our patients has been analyzed by Sanger sequencing to look for nucleotide variation in the two main genes involved in CS, *ERCC6* (*CSB*) (NM_000124.3) and *ERCC8* (*CSA*) (NM_000082.3). This assay did not identify any disease-causing variant neither in the coding sequence nor in the intron/exon boundaries of the genes.

Considering the relevant clinical presentation, the consanguinity of the patients and the DNA repair deficiency, RT-PCR of both *ERCC6* (*CSB*) and *ERCC8* (*CSA*) mRNA, followed by an agarose electrophoresis of the amplified products, have been performed. Whereas a normal profile was obtained for all *ERCC6* (*CSB*) cDNA products (data

not shown), the amplification of the whole *ERCC8* (*CSA*) cDNA in the patients has revealed a very weakly expressed normal transcript and multiple abnormal transcripts of a higher size of a few hundreds base pairs (Fig. 2a), suggesting the inclusion in the mature transcript of an additional sequence. Overlapping RT-PCRs of the *ERCC8* gene have allowed us to localize the inserted sequence between exon 2 and 3. To characterize the nucleotide change underlying this event, sequencing of the purified exon 2-exon 3 RT-PCR fragment was performed and demonstrated the coexistence of only two isoforms: a normal transcript and a single mutated transcript carrying a frameshift insertion of 348 bp between exon 2 and 3 (Supplementary data 2). BLAST analysis allowed to map the inserted sequence in the NCBI Reference Sequence NG_009289.1 of the *Homo sapiens* chromosome 5, which includes the *ERCC8* (*CSA*) gene, with 99% homology. The insert of 348 bp corresponds to a segment of the intron 2 that starts at position +1085 downstream to the 3’ end of exon 2 (r.173_174ins173+1085_173+1432). The junction between exon 2 and the insert has been checked by sequencing the PCR product obtained after amplification of the patient cDNA with one forward primer in exon 2 and one reverse primer in the insert (Fig. 2b, c). The same check has been performed at the junction between the inserted fragment and exon 3. The insertion interrupts the reading frame by introducing a premature stop codon after the Arginine 57 (p.Tyr58*) (Fig. 2c). Western blot analyzes indicated the complete absence of the CSA protein in both patients (Fig. 2d).

Bioinformatics predictions of splice sites at the insert boundaries revealed the presence of a 5’ (acceptor) splice site at c.173+1085 and a 3’ (donor) one at c.173+1432 with significant scores for all tools (see Materials and Methods section). Due to the presence of these natural splicing sites flanking the inserted segment on genomic DNA, our first assumption was to search a nucleotide variation at the level of these splice sites. We undertook genomic DNA analysis of intron 2 in the patients by PCR amplification and sequencing of a DNA fragment corresponding to position c.173+966 to c.173+1541. This analysis showed no change at the level of c.173+1085 donor and c.173+1432 acceptor sites but the presence in the two brothers of two homozygous variations inside the insert, NG_009289.1 (NM_000082.3):c.173+1119G>C (Chr5(GRCh37):g.60223572C>G) and NG_009289.1(NM_000082.3):c.173+1158A>G (Chr5(GRCh37):g.60223533T>C) (Fig. 2e). The first c.173+1119G>C variant is absent from public data bases such as dbSNP142 (The Database of Single Nucleotide Polymorphisms (<http://www.ncbi.nlm.nih.gov/SNP/>). Accessed: Build ID: 142) and is reported in the Genome Aggregation Database (GnomAD, <http://gnomad.broadinstitute.org/>) [18] with an extremely low frequency of

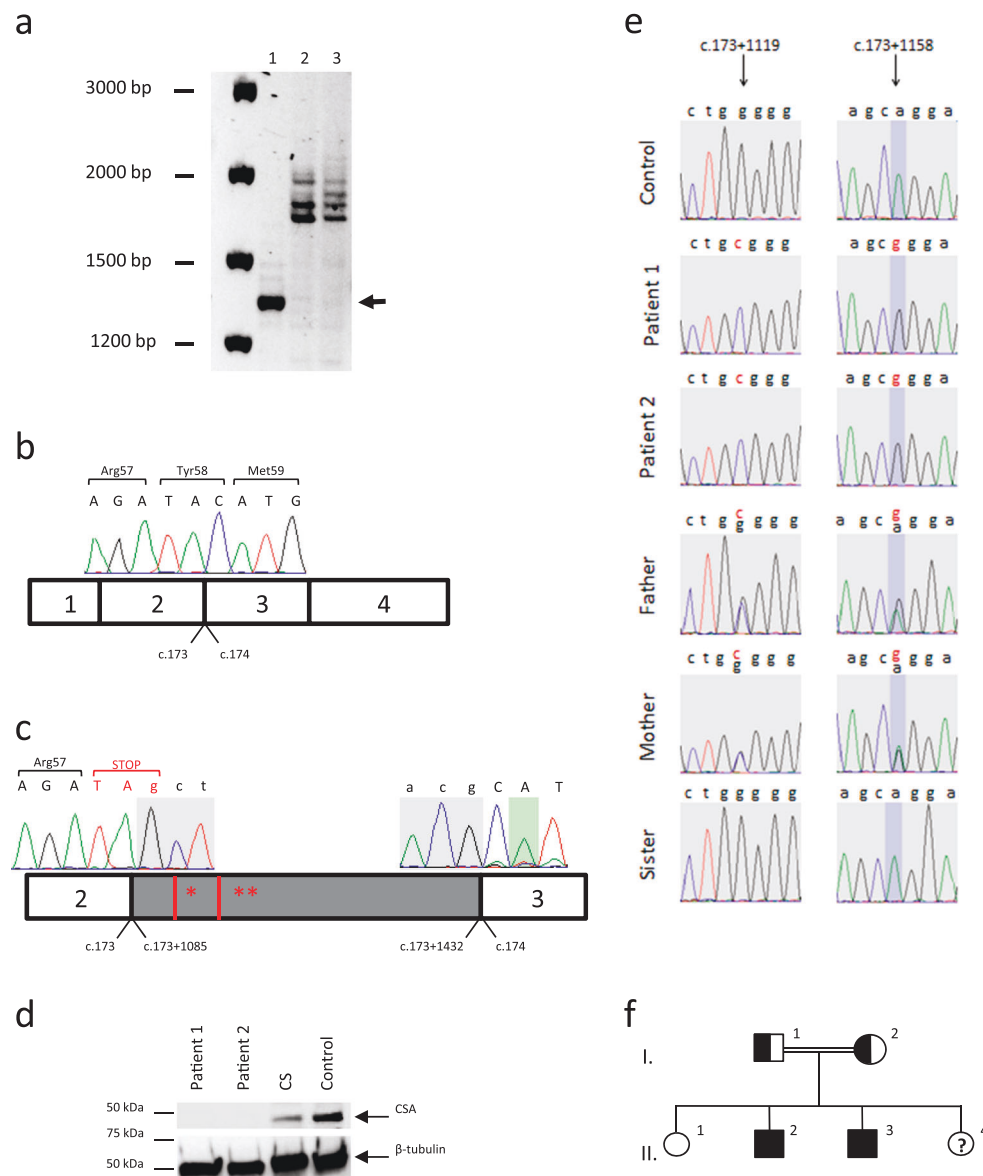


Fig. 2 Study of the *ERCC8* (*CSA*) transcript and *CSA* protein in primary fibroblast cultures. **a** RT-PCR amplification products of the full *ERCC8* (*CSA*) transcript in normal control (lane 1), patients 1 and 2 (lane 2 and 3, respectively). The arrow indicates the size of the normal transcript (1324 bp). **b–c** Schematic drawing and cDNA sequencing chromatograms of exons boundaries in normal control **b** and in patient 1 **c**. DNA sequencing has been done on PCR products obtained from an exonic and an intronic primer to avoid the sequence overlap of the normal and mutant transcripts. The gray box illustrates the 348 bp insert from intron 2 (c.173+1085_173+1432). The corresponding amino acid is indicated above the nucleotide codon (intronic bases in lower case). The insertion interrupts the reading frame by

introducing a premature stop codon. Localization of c.173+1119G>C (*) and c.173+1158A>G (**) variants are mentioned. **d** Western blot analysis of *CSA* protein (43 kDa) in primary fibroblast cultures. CS corresponds to a Cockayne patient compound heterozygous for the variations NM_000082.3(*ERCC8*):c.356 C>A p.Ser119* and NG_009289.1(NM_000082.3):c.618-1G>A. The control cell line comes from individual unaffected with CS. β -tubulin (50 kDa) is used as a loading control. **e** DNA sequencing chromatograms at position c.173+1119 and c.173+1158 of *ERCC8* (*CSA*) intron 2 for patients 1 and 2, their parents and healthy sister, compared to a wild-type control. **f** Pedigree of family. The individual with “?” was not tested

6.617e-5 (2 alleles on 30,224, no homozygous). The second c.173+1158A>G variant is a known variant (rs4647068) with no functional effect reported with a frequency of 0.02510 in the Genome Aggregation Database (759 alleles on 30,236 with 10 homozygous reported). This deep intronic region was not covered in the ExAC project (The

Exome Aggregation Consortium (<http://exac.broadinstitute.org>). Accessed: December 2016). The segregation of these variations has revealed that both parents were heterozygous carrier of the variants (Fig. 2e and 2f), whereas one healthy sister of the patients was homozygous for the wild-type alleles.

Exome sequencing

As several clinical features at birth are uncommon for long-lived CS patients, we performed exome sequencing in both patient 1 and 2 to look for any additional disease-associated variant that might influence the phenotype. Analysis of exome sequencing data was restricted to variations shared between the two brothers and was conducted in three independent steps. First, comparison of the variant file with the ClinVar database revealed that patients 1 and 2 are heterozygous carriers of disease-associated known variant in two genes implicated in autosomal recessive diseases (NPC2 involved in Niemann-Pick disease type C and F11, involved in Congenital factor XI deficiency) (Supplementary data 3). Second, no additional variation was found in genes involved in the NER pathway (*DDB2* (*XPE*); *ERCC1*; *ERCC2* (*XPD*); *ERCC3* (*XPB*); *ERCC4*; *ERCC5* (*XPG*); *ERCC6* (*CSB*); *ERCC8* (*CSA*); *GTF2H5* (*TTDA*); *MPLKIP* (*TTDN1*); *PCNA*; *POLH*; *RNF113A*; *SMARCA11*; *UVSSA* (*KIAA1530*); *XPA* and *XPC*). Finally, analysis of homozygous variants located in OMIM genes did not reveal any convincing variation that might influence the phenotype of the patients.

Minigene construct and expression

In order to further investigate the effects of both c.173+1119G>C and c.173+1158A>G variations on the splicing we used a cellular assay to study the expression of a reporter minigene system with exon-trapping model [19]. The *ERCC8* (*CSA*) cryptic exon with flanking sequences was subcloned in pSPL3B vector, between two exons of the rabbit β -globin gene. The wild-type and double variants construct were obtained from patient and control DNA, respectively, whereas single variant constructions were obtained by in vitro mutagenesis. The splicing patterns corresponding to the wild-type and to the mutant alleles were compared by reverse transcriptase-PCR analysis of the RNA from transfected cells. Figure 3 shows the RT-PCR products obtained from HeLa transfected cells by using a forward primer (SD6) in the β -globin first exon and a reverse primer (ECR) within the inserted *ERCC8* sequence to better visualize and highlight the amplification of the abnormal splicing species (which were faintly visible when using the β -globin primers SD6 and SA2). The same amplification was performed at the other boundary using the ECF and SA2 primers (data not shown). However, no transcript was visibly amplified with the wild-type construct indicating that the c.173+1085 donor and the c.173+1432 acceptor splice sites are not functional and could not promote the insertion of the cryptic exon within the transcript. On the contrary, transfection with the double variant vector has led to the production of a chimeric β -globin-*ERCC8*

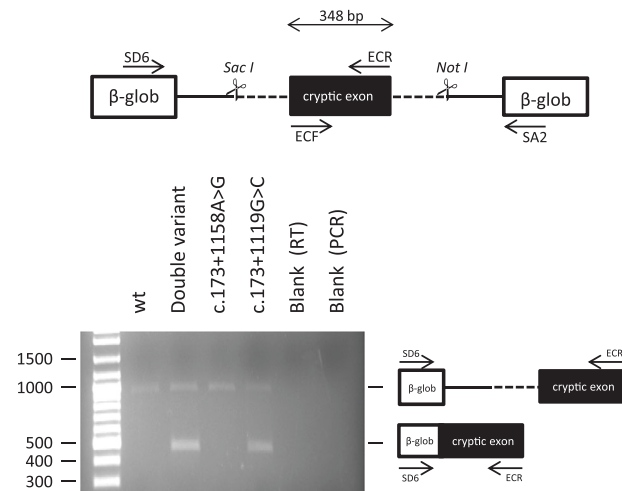


Fig. 3 In vitro analysis of the effect of *ERCC8* (*CSA*) c.173+1119G>C and c.173+1158A>G variants using a minigene assay. The genomic DNA segment tested is shown schematically above the panel: white boxes represent β -globin exons of the minigene vector and gray box illustrates the 348 bp insert from intron 2 (c.173+1085_173+1432) cloned into the vector. The lower panel represents migration patterns of RT-PCR products (primers SD6 and ECR) from HeLa cells transfected with wild-type, double-mutant, c.173+1158A>G, and c.173+1119G>C variants. The two last lines are negative controls (RT control and PCR control, respectively). The PCR products were obtained by using a forward primer (SD6) in the β -globin first exon and a reverse primer (ECR) within the inserted *ERCC8* sequence to highlight the amplification of the abnormal splicing species (which are faintly visible when using the SD6 and SA2 β -globin primers). The 1000 bp band has been sequenced and corresponds to plasmid DNA or to non-spliced transcript. A schematic representation of the PCR products is shown on the right of the picture

RT-PCR product. Sanger sequencing of this amplicon confirmed the inclusion of the cryptic *ERCC8* exon in the transcript, starting at the position c.173+1085. At the other boundary, PCR obtained with the ECF/SA2 primers showed that the cryptic exon ended at the position c.173+1432, as observed in patient's RNA. Then single variant constructs were used to decipher the specific role of each variant. The expression results obtained with the c.173+1119G>C variant were the same as the ones observed with the double variants, whereas the c.173+1158A>G results mimicked the wild-type results. These data support that the c.173+1119G>C variant is a disease-associated variant whereas the c.173+1158A>G variant does not affect function. Similar results were obtained by transfection in other cell lines (Neuro2A and human primary fibroblasts, data not shown).

Discussion

We report here the case of a deep intronic nucleotide variation causing the insertion of a cryptic exon in the *ERCC8*

(CSA) transcript, by modifying intronic regulatory elements important for exon definition (Fig. 4). This variation was identified in two brothers in whom the diagnosis of Cockayne syndrome was clinically suspected and confirmed by DNA repair deficiency. In the absence of variation in the coding sequence of the two main genes involved in Cockayne syndrome, a RT-PCR followed by a sequencing analysis of both transcripts have been performed. Multiple largest *ERCC8* (CSA) transcripts were amplified in both patients when studying the full *ERCC8* (CSA) cDNA, and both patients do not have exactly the same profile. However only two isoforms were detected after exon 2- exon 3 amplification and sequencing: the normal transcript and a mutated transcript carrying an insert of 348 bp of intron 2 (c.173+1085_173+1432) between exons 2 and 3. We hypothesize that the multiband profile observed after whole-gene amplification could be explained by heteroduplex formation between the isoform with the cryptic exon and isoforms corresponding to the wild-type and alternatively spliced transcripts. However the possibility of different splicing products based on the different combinations of splicing acceptor and donor sites cannot be excluded. Bioinformatics prediction of splice sites at the insert boundaries in intron 2 has revealed a 5' splice site (5'ss) at c.173+1085 and a 3' one at c.173+1432. The patients are moreover homozygous carrier of two differences within this insert, an unknown variation c.173+1119G>C and a known variant with no functional effect, c.173+1158A>G reported with a frequency of 2.5%. The contribution of c.173+1119G>C variation on splicing alteration and the neutral effect of the c.173+1158A>G common variant have been demonstrated in vitro with a reporter minigene system.

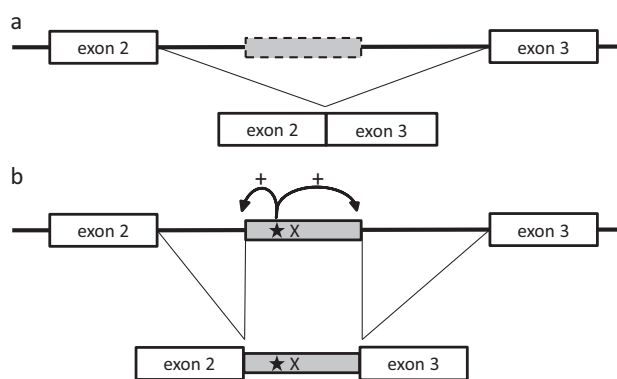


Fig. 4 Schematic representation of the *ERCC8* (CSA) disease-causing mechanism in the family. **a** Normal splicing between exon 2 and 3 of the *ERCC8* (CSA) gene. The dashed box represents the cryptic exon in intron 2. **b** In the patients, the deep intronic variation c.173+1119G>C (star) creates an ESE (exonic splicing enhancer) site, activating cryptic splice sites and leading to the insertion of the cryptic exon between exon 2 and 3. The c.173+1158A>G (cross) does not have any effect. Drawings are not at scale

The c.173+1119G>C change, by modifying intronic splicing regulatory elements, activates the potential splice sites flanking the intronic sequence, leads to intronic retention, frameshifting and the generation of a premature termination codon at the very end of exon 2 that prematurely interrupt the WD-repeat CSA protein (Fig. 4). This variant has been submitted to the ClinVar database (accession SCV000537862). Despite the presence of a small quantity of normal transcript, no full-length protein was detected by western blot. The sensitivity of the antibody and the western blot assay could be too low to detect small amount of normal CSA protein in fibroblast cultures.

In this study, the minigene delivery strategy, rather than the use of in silico bioinformatic predictors, has been implemented to validate the effect of the variations on splicing. Variations affecting canonical splice sites or creating an exonic cryptic site can now be predicted with high confidence by in silico tools that evaluate the strength of splice sites [20]. However, identifying bona fide functional splicing regulatory elements such as ESE (exonic splicing enhancers) remains a challenging issue as splicing regulatory signals are intricate, often overlapping and context-dependent [21, 22]. Cellular assays and functional studies are still required for the validation of bioinformatic predictions. In the present case, the prediction of binding sites of serine/arginine splicing factors (SRSF) has been studied using the web-based “ESE finder 3.0” program and the results illustrate the difficulty to predict the impact of nucleotide changes potentially affecting regulatory elements (Supplementary data 4).

The minigene strategy used in this study also has several drawbacks. Inclusion of the cryptic exon into the β -globin transcript is not very efficient as overlapping PCR between the β -globin exon and the cryptic exon was needed to detect it. However as the splicing effect of the variants has been observed in vivo on the patients’s fibroblasts, the minigene assay using different cell lines (Neuro2A, HeLa and human primary fibroblasts) was only used here in a qualitative manner to discriminate the functional effect between the two variants c.173+1119G>C and c.173+1158A>G. This test is not quantitative at all and does not reflect what’s happen in the different tissues of the patients.

Descriptions of CS cases due to splice disease-causing variations are rare: 19 splice variants have been identified in *ERCC6* and 12 in *ERCC8* (the Human Genome Mutation Database (<https://portal.biobase-international.com/hgmd/pro/start.php>) and personal data). Most of these changes affect the splice site consensus sequences (29 variations out of 31) and usually involve one of the six bases surrounding the exon/intron junction. An *ERCC6* disease-causing variation has been described at 26 bp from the junction (c.2599-26A>G) (patient CS177VI, [11]), creating a new acceptor site and leading to partial insertion of intron 13 in

the transcript (r.2598_2599ins2599-25_2599-1). Mid-intronic variation creating a cryptic exon has only been described once in a CSA patient (patient CS333VI, [11]). In this case the homozygous deep intronic substitution c.173+1046A>G was shown to activate a cryptic splice site within intron 2 of the CSA gene leading to partial insertion of intron 2 in the transcript (r.173_174ins173+977_173+1045). The intron involved in this patient is the same as the one in the present family. Contrary to mechanisms described in CS333VI, where deep intronic variant activate an imperfect splice site, our nucleotide variants are located within the inserted sequence and consequently its disease-causing mechanism should be different. We postulated in our case that the nucleotide variants affect a critical *cis*-regulatory elements resulting in a cryptic exon insertion that interrupts the reading frame of the mature mRNA. To our knowledge, it is the first Cockayne patient described with this kind of nucleotide change. A similar disease-causing mechanism has been described in several genetic diseases such as Fabry disease [23], X linked Alport syndrome [24], afibrinogenemia [25], cblE type of homocystinuria [26], or Duchenne muscular dystrophy [27].

Various levels of severity have been described in CS [2]. Type II CS patients typically show symptoms at birth and an average lifespan of 5 or 6 years. Type I or moderate CS patients usually experience the first signs of the disease at the end of the first year of life and mean life expectancy is around 16 years. Type III CS patients may only manifest the first symptoms after several years of life and death occurs at a mean age of 30 years. In this study both affected patients presented an early onset of the disease (at birth for patient 1 and during neonatal period for patient 2), as well as a severe clinical phenotype, contrasting with their relatively long survival (23 years for patient 1, at least 17 years for patient 2). A similar clinical profile has been observed in the CS333VI patient mentioned above [11]: this CSA patient, homozygous for an *ERCC8* (*CSA*) deep-intronic change that activates an imperfect splice site, showed an early onset of symptoms (psychomotor delay diagnosed at 4–5 months of age), as well as severe involvement, contrasting with an age at death of 17 years. The paucity of CS cases with early onset and prolonged survival prevent from a robust genotype-phenotype correlation study but some issues could be hypothesized. First, a recent large cohort study of 102 CS patients [3] demonstrated that *ERCC8*(*CSA*) disease-causing variations have a borderline significant association with longer survival, compared to *ERCC6*(*CSB*) variations, confirming the tendency that *CSA* variations seem to be linked preferentially to the moderate form of the disease [2]. Despite the early onset, the reported cases linked to the *CSA* gene are in accordance with this observation.

Splice variations have not been constantly associated with milder phenotypes in former genotype-phenotype correlations studies [11, 28]. However the presence of at least one splice variant in all CSB patients previously described with mild type III disease [11] supports the hypothesis that splicing events could have less deleterious consequences compared to non-splicing variants. Moreover the type and degree of aberrant splice, and its corollary regarding residual expression of wild-type isoforms, may vary between tissues and organs and/or during development. This may also contribute to the different effects of *CSA* variations on the severity of the CS symptoms. However this hypothesis is difficult to study due to the scarcity of such cases and to the difficulty to have access to representative tissues in space and time.

Finally, we cannot exclude the possibility in our patients that another variation in a different gene might have contributed to the severe developmental problems, which are unusual in otherwise relatively mild CS. Exome sequencing performed on both brothers has excluded potential disease-causing variations in OMIM gene, including genes involved in the NER pathway. However, we cannot exclude the presence of a variation in a potential modifier gene that might account for the severe clinical features at birth.

Finally the present cases share similarities with the unelucidated Japanese patient described more than 10 years ago by Komatsu et al. [29]: multiple splicing *CSA* variants were identified in a CS patient with late-onset and prolonged survival form, but no disease-causing variation could be found at the genomic level. This observation may also be explained by un-identified variation affecting regulatory splicing sequences, by either weakening or strengthening exon recognition. As in the Japanese patient, multiple mRNA species have been observed in patients 1 and 2 and the relative intensities of each mRNA species are slightly different between the two brothers (Fig. 2a). The complexity of splicing regulation machinery and the combined effects of numerous *cis* and *trans* acting elements may explain alternative splicing in patients' cells as well as intrafamilial variability.

In summary, this report strongly supports the disease-causing effect of *ERCC8* (*CSA*) deep intronic variation NG_009289.1(NM_000082.3):c.173+1119G>C and describes a new molecular mechanism in CS: in this new paradigm for CS, the variant affects a regulatory splicing sequence leading to an activation of splice sites and incorporation of a cryptic exon which interrupts the reading frame of the mRNA. It underlines the importance of transcript analysis, every time that the clinical presentation is highly relevant and genomic analysis is normal. The description of this additional patient with early-onset symptoms and unexpectedly slow rate of progression of the disease, associated with deep-intronic variation, raises

the possibility of genotype-phenotype correlation, though additional cases are required to further support this hypothesis. Moreover, these cases confirm the variability of CS and highlight the difficulty of classification in distinct severity groups.

Acknowledgments We would like to thank the family involved in this study. We also thank Dr Amelie Piton for providing pSPL3B vector and for scientific discussions. We also thank Dr Jean Muller for bioinformatic advice and Mrs. Sinthuja Pachchek for bioinformatic support in exome sequencing.

Conflict of interest The authors declare that they no competing interest.

References

- Kleijer WJ, Laugel V, Berneburg M, et al. Incidence of DNA repair deficiency disorders in western Europe: Xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. *DNA Repair*. 2008;7:744–50.
- Laugel V. Cockayne syndrome: the expanding clinical and mutational spectrum. *Mech Ageing Dev*. 2013;134:161–70.
- Wilson BT, Stark Z, Sutton RE, et al. The Cockayne Syndrome Natural History (CoSyNH) study: clinical findings in 102 individuals and recommendations for care. *Genet Med*. 2016;18:483–93.
- Mayne LV, Lehmann AR. Failure of RNA synthesis to recover after UV irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma pigmentosum. *Cancer Res*. 1982;42:1473–8.
- Schmickel RD, Chu EH, Trosko JE, Chang CC. Cockayne syndrome: a cellular sensitivity to ultraviolet light. *Pediatrics*. 1977;60:135–9.
- Henning KA, Li L, Iyer N, et al. The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIH. *Cell*. 1995;82:555–64.
- Troelstra C, van Gool A, de Wit J, Vermeulen W, Bootsma D, Hoeijmakers JH. ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. *Cell*. 1992;71:939–53.
- Karikkineth AC, Scheibye-Knudsen M, Fivenson E, Croteau DL, Bohr VA. Cockayne syndrome: clinical features, model systems and pathways. *Ageing Res Rev*. 2017;33:3–17.
- Groisman R, Kuraoka I, Chevallier O, et al. CSA-dependent degradation of CSB by the ubiquitin–proteasome pathway establishes a link between complementation factors of the Cockayne syndrome. *Genes Dev*. 2006;20:1429–34.
- Wang Y, Chakravarty P, Raney M, et al. Dysregulation of gene expression as a cause of Cockayne syndrome neurological disease. *Proc Natl Acad Sci*. 2014;111:14454–9.
- Laugel V, Dalloz C, Durand M, et al. Mutation update for the CSB/ERCC6 and CSA/ERCC8 genes involved in Cockayne syndrome. *Hum Mutat*. 2010;31:113–26.
- Calmels N, Greff G, Obringer C, et al. Uncommon nucleotide excision repair phenotypes revealed by targeted high-throughput sequencing. *Orphanet J Rare Dis*. 2016;11:26.
- Laugel V, Dalloz C, Stary A, et al. Deletion of 5' sequences of the CSB gene provides insight into the pathophysiology of Cockayne syndrome. *Eur J Hum Genet*. 2008;16:320–7.
- Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol J Comput Mol Cell Biol*. 2004;11:377–94.
- Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in genie. *J Comput Biol*. 1997;4:311–23.
- 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–e67.
- Geoffroy V, Pizot C, Redin C, et al. VaRank: a simple and powerful tool for ranking genetic variants. *PeerJ*. 2015;3:e796.
- Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–91.
- Burn TC, Connors TD, Klinger KW, Landes GM. Increased exon-trapping efficiency through modifications to the pSPL3 splicing vector. *Gene*. 1995;161:183–7.
- Houdayer C, Caux-Moncoutier V, Krieger S, et al. 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*. 2012;33:1228–38.
- Chasin LA. Searching for splicing motifs. *Adv Exp Med Biol*. 2007;623:85–106.
- Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*. 2008;14:802–13.
- Ishii S, Nakao S, Minamikawa-Tachino R, Desnick RJ, Fan J-Q. Alternative splicing in the α -galactosidase a gene: increased exon inclusion results in the fabry cardiac phenotype. *Am J Hum Genet*. 2002;70:994–1002.
- King K, Flinter F, Nihalani V, Green P. Unusual deep intronic mutations in the COL4A5 gene cause X linked Alport syndrome. *Hum Genet*. 2002;111:548–54.
- Davis RL, Homer VM, George PM, Brennan SO. A deep intronic mutation in FGB creates a consensus exonic splicing enhancer motif that results in afibrinogenemia caused by aberrant mRNA splicing, which can be corrected in vitro with antisense oligonucleotide treatment. *Hum Mutat*. 2009;30:221–7.
- Homolova K, Zavadakova P, Doktor TK, Schroeder LD, Kozich V, Andresen BS. The deep intronic c.903+469T>C mutation in the MTRR gene creates an SF2/ASF binding exonic splicing enhancer, which leads to pseudoexon activation and causes the cbIE type of homocystinuria. *Hum Mutat*. 2010;31:437–44.
- Trabelsi M, Beugnet C, Deburgrave N, et al. When a mid-intronic variation of DMD gene creates an ESE site. *Neuromuscul Disord*. 2014;24:1111–7.
- Mallery DL, Tanganelli B, Colella S, et al. Molecular analysis of mutations in the CSB (ERCC6) gene in patients with Cockayne syndrome. *Am J Hum Genet*. 1998;62:77–85.
- Komatsu A, Suzuki S, Inagaki T, Yamashita K, Hashizume K. A kindred with Cockayne syndrome caused by multiple splicing variants of the CSA gene. *Am J Med Genet*. 2004;128A:67–71.