



Statistical Approach of the Role of the Conserved CSB-PiggyBac Transposase Fusion Protein (CSB-PGBD3) in Genotype-Phenotype Correlation in Cockayne Syndrome Type B

Rayanne Damaj-Fourcade¹, Nicolas Meyer^{2,3}, Cathy Obringer⁴, Nicolas Le May⁴, Nadège Calmels^{4,5} and Vincent Laugel^{4,6*}

¹INSERM CIC 1434, Hôpitaux Universitaires de Strasbourg, Strasbourg, France, ²GMRC, Service de Santé Publique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France, ³ICUBE, UMR 7357, Université de Strasbourg, Illkirch, France, ⁴Laboratoire de Génétique Médicale, Institut de Génétique Médicale d'Alsace, Faculté de Médecine de Strasbourg, Hôpitaux Universitaires de Strasbourg, Strasbourg, France, ⁵Laboratoire de Diagnostic Génétique, Institut de Génétique Médicale d'Alsace, Nouvel Hôpital Civil, Hôpitaux Universitaires de Strasbourg, CRBS, Strasbourg, France, ⁶Service de Pédiatrie Spécialisée et Générale, Unité de Neurologie Pédiatrique, Hôpital de Hautepierre, Hôpitaux Universitaires de Strasbourg, Strasbourg, France

OPEN ACCESS

Edited by:

Corrado Romano,
University of Catania, Italy

Reviewed by:

Victor Faundes,
University of Chile, Chile
Rahim Alhamzawi,
University of Al-Qadisiyah, Iraq

*Correspondence:

Vincent Laugel
vincent.laugel@chru-strasbourg.fr

Specialty section:

This article was submitted to
Genetics of Common and Rare
Diseases,
a section of the journal
Frontiers in Genetics

Received: 27 August 2021

Accepted: 24 January 2022

Published: 17 February 2022

Citation:

Damaj-Fourcade R, Meyer N, Obringer C, Le May N, Calmels N and Laugel V (2022) Statistical Approach of the Role of the Conserved CSB-PiggyBac Transposase Fusion Protein (CSB-PGBD3) in Genotype-Phenotype Correlation in Cockayne Syndrome Type B. *Front. Genet.* 13:762047. doi: 10.3389/fgene.2022.762047

Cockayne syndrome is a rare condition that encompasses a very wide spectrum of clinical severity. Mutations upstream of a transposon called PiggyBac Transposable Element Derived 3 in intron 5 of the *CSB/ERCC6* gene could bring about less severe forms than mutations located downstream of that transposon insertion. Our aim was to study genotype-phenotype correlation by determining whether the position of each mutation of the *CSB/ERCC6* gene has an impact on the phenotype. A hundred and forty-seven Cockayne patients, who had two pathogenic mutations in the *CSB/ERCC6* gene and for whom clinical data was available, were retrospectively selected and included in the study. Data analysis was performed under the Bayesian paradigm. Analysis of the proportion of the different subtypes of Cockayne syndrome according to the position of the mutations was done using an ordinal logistic regression model. Using a vague prior, the risk of developing a more severe subtype when exposed to 2 mutations downstream compared to 2 mutations upstream was 2.0 [0.9–4.5]. Estimations varied through the sensitivity analysis. We could reasonably conclude that a relationship between the number of downstream mutations and the Cockayne syndrome clinical expression exists but it is still difficult to give a precise estimate of this relationship. The real effect could be more complex than the one described in the initial model and other genetic factors might be taken into consideration together with the mutation site to better explain clinical variability.

Keywords: cockayne syndrome, CSB/ERCC6, piggyBac, clinical severity, correlation-regression analysis

Abbreviations: CS, Cockayne syndrome; RRS, recovery of RNA synthesis; OR, odds ratio; 2U, 2 mutations upstream of PiggyBac; 1U1D, 1 mutation upstream and 1 mutation downstream of PiggyBac; 2D, 2 mutations downstream of PiggyBac.

INTRODUCTION

Cockayne syndrome (CS) is an autosomal recessive disorder characterized by growth failure, developmental delay, microcephaly, sensorial alteration, cutaneous photosensitivity, dental anomalies and a recognizable facial appearance (Cockayne, 1946). Its incidence has been estimated at 1/360,000 births in Western Europe (Kleijer et al., 2008).

Several stages of severity have been described previously. They essentially depend on the age at first symptoms and on how quickly the disease develops (Natale, 2011; Laugel, 2013). In the classical form or CS type I, symptoms usually start at the end of the first year or in the second year of life. All the symptoms mentioned above can show up progressively. This type leads to premature death mainly in the second decade of life. CS type II is equivalent to severe congenital forms. Clinical signs are observed at birth, especially congenital microcephaly and cataracts. The evolution is always severe conducting to death before the end of the first decade. CS type III brings together mild juvenile forms. Symptoms are less severe and survival until adulthood is possible. Cerebro-oculo-facio-skeletal (COFS) syndrome is described as an arthrogryposis (Graham et al., 2001). It starts before birth and is associated with a major handicap and death within the first years of life.

Specific sensitivity to ultraviolet (UV) radiation of Cockayne cells has been related to defective DNA repair and transcription processes especially the nucleotide excision repair (NER) subpathway called transcription coupled-repair (TC-NER) (Mayne and Lehmann, 1982). The CS diagnosis can be suggested by clinical criteria (Nance and Berry, 1992; Laugel, 2013; Wilson et al., 2016a). However, diagnosis must be confirmed by assessing TCR Recovery of RNA Synthesis (RRS, decreased in CS) in cultured fibroblasts and searching for mutations in involved genes. Two major genes are responsible for CS: *CSA/ERCC8* and *CSB/ERCC6* (Troelstra et al., 1992; Henning et al., 1995). The *ERCC6* gene is located on chromosome 10q11 and contains 21 exons. A transposon, named PiggyBac Transposable Element Derived 3 (PGBD3), fits into intron 5 of the *CSB/ERCC6* gene. Consequently, in physiological state, the *CSB* gene generates a full-length *CSB* protein and a fusion protein *CSB-PiggyBac* that contains the first 465 amino acids of the *CSB* protein and the PGBD3 transposase (Newman et al., 2008). This chimeric protein is as highly conserved as *CSB* itself and its continued expression is able to reshape the transcriptome in cultured cells (Newman et al., 2008; Gray et al., 2012; Weiner and Gray, 2013). The effect of the chimeric protein is anticipated to be deleterious in the absence of functional *CSB* (absent or truncated) and may thus affect the clinical presentation in CS patients by directly altering the transcriptional program.

The large variety of symptoms and range of severity are a startling aspect of CS. Up to date, this clinical spectrum variability has not been justified by any genotype-phenotype correlation. The presence of a transposon in intron 5 of the *CSB/ERCC6* gene has led to express the following hypothesis: mutations upstream of intron 5 (coding exons 2–5) should generate neither a normal full length *CSB* protein nor a normal *CSB-PiggyBac* fusion

protein, whereas mutations downstream of intron 5 (coding exons 6–21) should impact the *CSB* full length protein but generate a normal chimeric protein which is assumed to have a deleterious effect in the absence of *CSB* full length protein; mutations upstream of intron 5 could thus lead to less severe forms than mutations located downstream of that transposon insertion (Newman et al., 2008; Gray et al., 2012; Weiner and Gray, 2013). This hypothesis has already been considered as a possible therapeutic strategy with splice switching oligonucleotides to potentially reduce the disease severity (Sin et al., 2018).

Our aim was to test this hypothesis and study the genotype-phenotype correlation in a cohort of Cockayne patients to determine whether the position of the mutation on each allele of the *CSB/ERCC6* gene has an impact on the type of CS.

We also aimed to study the effect of the type of variants (protein-truncating variants (PTVs) vs Protein-altering variants (PAVs)) on the phenotype.

MATERIALS AND METHODS

Patients

Our Strasbourg laboratory of genetics has a database for CS from which we retrospectively selected our patients. The database contained 154 Cockayne patients for whom genetic testing confirmed the diagnosis. Cockayne patients from both genders, who had two pathogenic mutations in the *CSB/ERCC6* gene, were included in the analysis. Patients who had pathogenic mutations in other genes within the NER pathway or died *in-utero* with therefore a lack of precise and reliable clinical data were excluded. Absence of clinical data was another exclusion criteria. Our database and fibroblast library (DC-2014-2222) have been registered at the French authority Commission Nationale de l'Informatique et des Libertés and authorized by the Comité de Protection des Personnes. Families had given full consent for genetic screening. In order to increase the number of patients in the main analysis, we searched for cases from the literature for whom both clinical and molecular data were available. We enrolled every publication that reported Cockayne patients. Articles had to be written in English and were eligible if they presented the details of the *CSB* mutations and clinical features of patients. Some of them were extracted through PubMed searches of the terms ["cockayne syndrome [MeSH Terms]" AND "phenotype [MeSH Terms]" AND "mutation [MeSH Terms]"]. We decided not to apply a restriction of publication date. The search was supplemented by our own previous bibliographic survey (Laugel et al., 2010).

Data Collection

Clinical data was collected from specific clinical files, consultation letters or hospitalization reports. It included the following variables: gender, postnatal growth parameters (height, weight, head circumference), dysmorphic signs (enophthalmia, cachexia, bird-like face), sensorial impairment characteristics (hearing loss, cataracts, retinopathy), developmental milestones, neurological symptoms (areflexia, ataxia, spasticity), intellectual disability,

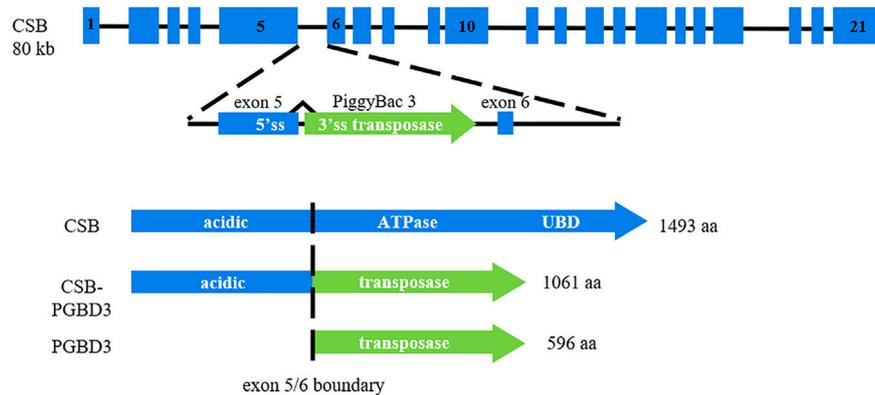


FIGURE 1 | PGBD3 transposon into intron 5 of the *CSB/ERCC6* gene. Alternative splicing of CSB exons 1-5 to the PGBD3 transposase splice acceptor site enables the CSB-PGBD3 fusion protein expression. The primate CSB locus generates two other proteins: full length CSB and solitary PGBD3 transposase. Figure inspired from Gray et al. (2012).

major cutaneous photosensitivity, arthrogryposis, age at first symptoms, age at death or last report. Genetic and molecular data contained the mutations of each patient on cDNA and their position compared with PiggyBac insertion (NM_000124.3: c.1397 + 6912, in intron 5). The position of the mutations was stratified into three groups: two mutations upstream of intron 5 (2U), two mutations downstream of intron 5 (2D), 1 mutation upstream and 1 mutation downstream of intron 5 (1UID) (Figure 1). Clinical and genetic data of patients stemmed from literature were directly extracted from case-reports. The type of variants was stratified into two groups: PTVs and PAVs. Patients who had 1 PTV and 1 PAV were considered in the PAVs group as one of their variants had a more permissive nature.

Patient Classification

Patients were classified into classical clinical subgroups (CS type I, CS type II, CS type III and COFS) by an expert committee without any knowledge of their mutations. The process was done according to classification routinely used (Natale, 2011; Laugel, 2013; Wilson et al., 2016a). COFS syndrome, the most severe subtype of the clinical spectrum, was defined by its own criteria: congenital microcephaly, congenital cataracts and/or microphthalmia, arthrogryposis, severe developmental delay, severe growth failure and facial dysmorphism (Laugel et al., 2008a). Furthermore, information concerning patients' psychomotor development helped to classify them. The type II (severe CS) subgroup included patients who had their first symptoms before the age of 3 months. They could not walk and had an extremely restricted verbal communication. Type I (moderate CS) patients had generally acquired sitting position and could walk independently later on. They were able to associate words, formulate short sentences and they had good interactions with their peers. Type III (mild CS) patients could walk, run, and talk in an elaborated way. They often had learnt to read and write. For reported cases from the literature, the classification already established by the authors was taken into account except for two patients for whom classification was not mentioned in the original article (Wilson et al., 2016b). They were consequently classified by our expert

committee based on available clinical data and pictures. Type II and COFS patients were finally gathered together within the same category, hereinafter referred to as the "Type II" class, as they represent the most severe phenotypes encountered.

Statistical Analysis

Qualitative variables were described as counts and percentages. Age was described with mean (standard deviation, minimum, maximum).

Data analyses were performed under the Bayesian paradigm. Bayesian methods were used for their internal logical consistency for probabilistic inference and their ability to explicitly include previous data and expert knowledge in the data analysis (Dunson, 2001). Results are all expressed with their respective 95% credibility interval (CI) on posterior distribution (Ferreira et al., 2020). The major part of the parameters was estimated using Markov chain Monte Carlo (McMC) algorithms. A burn-in of 5000 iterations, followed by 150 000 iterations with a thinning of 3, was used for the three chains. McMC chain convergence was assessed and checked graphically.

The relationship between the position of the mutations and the type of CS was studied first using a uniform Dirichlet prior distribution $D_i(1, \dots, 1)$ and a Jeffreys prior $D_i(0.5, \dots, 0.5)$. Based on the contingency table obtained with the Jeffreys prior, Odds Ratios (OR) were estimated comparing subtypes and classes of mutations two by two. Next, analysis of the proportion of the different types of CS according to the position of the mutations in the *CSB/ERCC6* gene was done using an ordinal logistic regression model under the assumption that the disease severity would increase with the number of downstream mutations. Prior distributions were defined before conducting the analyses. A sensitivity analysis was led. As part of it, we used three different Gaussian N (mean, variance) priors for parameters of the logistic regression. A low information prior ($\log(\text{OR}) \sim N(\mu = 0, \sigma^2 = 1,000)$) was first used. A more precise prior ($\log(\text{OR}) \sim N(\mu = 0, \sigma^2 = 0.674)$) specified an OR a priori between 0.2 and 5. An optimistic prior ($\log(\text{OR}) \sim N(\mu = 1.099, \sigma^2 = 0.377)$) specified an OR a priori between 0.9 and 10. In this study, Bayesian results

TABLE 1 | Patient characteristics.

Characteristics	Entire cohort (n = 147)		Strasbourg database (n = 85)		Case-reports (n = 62)	
	No	% ⁶	No	% ⁶	No	% ⁶
Sex						
Male	80	58.4	50	59.5	30	56.6
Female	57	41.6	34	40.5	23	43.4
NA ¹	10		1		9	
Alleles						
Homozygous	88	59.9	53	62.4	35	56.5
Heterozygous	59	40.1	32	37.6	27	43.5
Position of the mutations ²						
2 upstream	24	16.3	13	15.3	11	17.7
1 upstream/1 downstream	18	12.2	9	10.6	9	14.5
2 downstream	105	71.4	63	74.1	42	67.7
Fusion protein ³						
Absent	24	16.3	13	15.3	11	17.7
Present	123	83.7	72	84.7	51	82.3
Subtype						
I	54	36.7	31	36.5	23	37.1
I or II ⁴	4	2.7	4	4.7		
I/II ⁵	6	4.1	6	7.1		
II	71	48.3	36	42.4	35	56.4
III	12	8.2	8	9.4	4	6.5

¹.NA, not available.

².Regarding to PiggyBac insertion.

³.Prediction made according to the position of the mutations.

⁴.Patients who could be either in subtype I or subtype II.

⁵.Patients who are at the boundary between subtypes I and II.

⁶.Proportions were calculated without taking into account NA data.

gave the probability that the OR of a severe phenotype would be greater than 1. The effect was judged relevant when the probability that the OR exceeded 1 ($\Pr(\text{OR}>1)$) was larger than 95%. This probability should not be confused with the frequentist *p*-value.

Analysis of the proportion of the different types of CS according to the type of variants was done using an ordinal logistic regression model. As part of the sensitivity analysis, we used two different Gaussian priors for parameters which were a low information prior ($\log(\text{OR}) \sim N(\mu = 0, \sigma^2 = 1,000)$) and a more precise prior ($\log(\text{OR}) \sim N(\mu = 0.661, \sigma^2 = 0.234)$) that specified an OR a priori between 0.75 and 5. The type of variants was then included as a covariable in a multivariate model.

Our secondary outcome, age at onset according to the position of the mutations, was analyzed with a categorical predictor linear model, akin to an ANOVA model. In this model, the parameter prior distribution was $N(0, 100)$.

All analyses were done using JAGS 4.3.0 (rjags, RRID: SCR_017573) and R 4.0.3 (R Project for Statistical Computing, RRID:SCR_001905) softwares, with all the required packages in their latest version at the time of analysis.

RESULTS

In the Strasbourg database, 154 Cockayne patients were available. Patients who had pathogenic mutations in *ERCC1* (n = 2), *ERCC2*

(n = 2), *ERCC3* (n = 1) and *CSA/ERCC8* (n = 51) genes were excluded. Six patients died *in-utero* and were also excluded from the analysis. Clinical data was not available for 7 patients. Eighty-five patients from the Strasbourg database finally met the inclusion criteria to which 62 case-reports from the literature were added (Brumback et al., 1978; Troelstra et al., 1992; Lehmann et al., 1993; Stefanini et al., 1996; Mallery et al., 1998; Colella et al., 1999; Meira et al., 2000; Falik-Zaccari et al., 2008; Hashimoto et al., 2008; Wilson et al., 2016b; Kou et al., 2018; Sanchez-Roman et al., 2018). A total of 147 patients were included in the study, of whom 24 (16.3%) had 2U of PiggyBac, 18 (12.2%) had 1U1D of PiggyBac and 105 (71.4%) had 2D of PiggyBac (**Table 1**). The mean age at first symptoms was 8 months (standard error = 13.4). The values ranged from 0 (birth, minimum) to 72 months (maximum). This variable was missing for 24 patients.

Four patients could either correspond to a type I or a type II and six patients were at the boundary between type I and type II. Those 10 patients were therefore not included in the ordinal logistic regression model. Another patient was also removed from this analysis because he could not be classified according to the current criteria.

The proportions of patients in each subtype/mutations group of the cohort are presented in **Table 2**. Looking closer to the proportions of patients in each mutation group (**Figure 2**), there were more type II patients in the 2D group than in the 2U group and the 1U1D group (59.4 vs 39.1% and 29.4% respectively).

TABLE 2 | Numbers (%) of patients in each subtype/mutations group.

	Type II	Type I	Type III	Total
2U	9 (6.7)	11 (8.1)	3 (2.2)	23 (17.0)
1U1D	5 (3.7)	11 (8.1)	1 (0.7)	17 (12.5)
2D	57 (41.9)	32 (23.5)	7 (5.1)	96 (70.5)
Total	71 (52.3)	54 (39.7)	11 (8)	136 (100)

The first analysis, using a Dirichlet distribution with a Jeffreys prior, showed the following results (Table 3). The risk of developing type II disease as compared to type I was increased by 3.9 (CI: [1.3–12.8]) for patients having 2D in comparison to 1U1D. Patients exposed to 2D in comparison to 2U had a risk multiplied by 2.2 (CI: [0.8–5.9]) to develop a type II as compared to a type I and a risk multiplied by 2.8 (CI: [0.6–12.0]) to develop a type II as compared to a type III. Finally, patients exposed to 2D in comparison to those exposed to at least one mutation upstream (2U and 1U1D) had 2.8 (CI: [1.3–6.3]) more risk to have a type II compared to a type I with a Pr(OR>1) reaching 99.5%.

Ordinal logistic regression using a vague prior (N (0, 1,000)) showed that the risk of having a more severe subtype when exposed to 2D compared to 2U was 2.0 (CI: [0.9–4.5]) with a Pr(OR>1) of 94.1%. That risk when exposed to 1U1D compared to 2U, was of 0.9 (CI [0.3–2.7]) with a Pr(OR>1) equal to 42.6%.

Results obtained with the two other priors are presented in Table 4. Estimations varied through the sensitivity analysis.

Results of the model used to estimate age at first symptoms depending on the mutation position were as follows. Patients with 2U had a mean age at first symptoms of 11.8 [6.7–16.9] months which was higher than the mean age of patients with 2D. Indeed, those ones had 6.5 [3.8–9.3] months at first symptoms. The probability of a difference (Pr (diff) > 0) between 2D and 2U patients reached 96.2%. Patients with 1U1D had a mean age of 11.4 [5.1–17.6] months when first symptoms appeared. Pr (diff) > 0 were of 91.7 and 54.1% comparing respectively 2D to 1U1D and 1U1D to 2U.

We then looked at the potential effect of the mutation type (PAV or PTV). The univariate model aiming to study the effect of the type of variants on the phenotype included 132 patients which corresponded to complete cases. Using a vague prior (N (0, 1,000)), ordinal logistic regression pointed out that the risk of having a more severe subtype with PTVs as compared to PAVs was 0.8 (CI [0.4–1.6]) with a Pr(OR>1) of 25.1%. Results obtained with the prior in favour of the hypothesis that PTVs could induce more severe phenotypes (N (0.661, 0.234)) showed a risk of developing a severe phenotype multiplied by 1.1 (CI [0.6–1.9]) with PTVs as compared to PAVs with a Pr (OR>1) of 60.1%.

The multivariate ordinal logistic regressions (Supplementary Material) which included the position and the type of variants as covariates did not change the conclusions that we could make

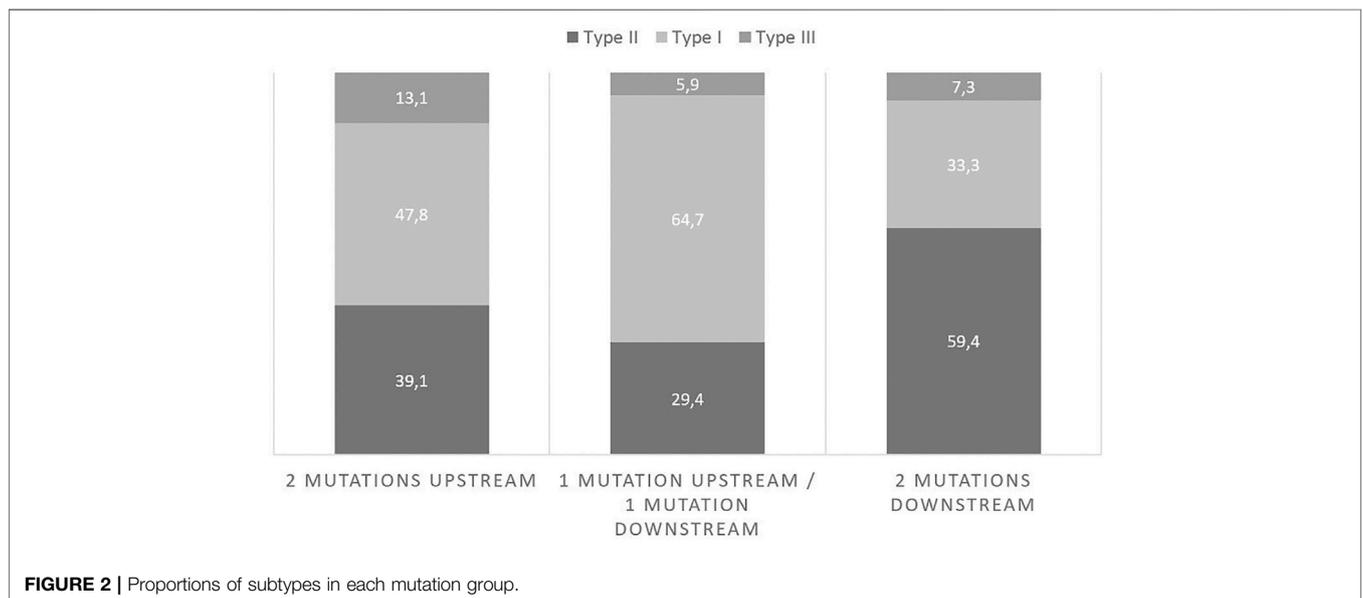


FIGURE 2 | Proportions of subtypes in each mutation group.

TABLE 3 | Risk of having a subtype compared to another depending on the position of the mutations.

	Subtype II vs I	Subtype I vs III	Subtype II vs III
2D vs 1U1D	^a 3.9 [1.3–12.8]; 99.3	0.5 [0.0–2.8]; 22.2	1.8 [0.1–12.6]; 70.5
1U1D vs 2U	0.6 [0.1–2.2]; 20.2	2.7 [0.3–38.8]; 81.8	1.5 [0.2–23.1]; 63.9
2D vs 2U	2.2 [0.8–5.9]; 94.1	1.3 [0.3–5.4]; 62.4	2.8 [0.6–12.0]; 90.1

^aOdds ratios [OR] are presented with their credibility intervals [CI] and probabilities of being higher than 1 (%).

TABLE 4 | Ordinal logistic regression odds estimates. Risk of having a more severe subtype depending on the position of the mutations with three different priors. OR, odds ratio; CI, credibility intervals; Pr(OR>1), probability that the OR is higher than 1 (%). 1: OR_p [CI]¹, odds ratio *a priori* with its credibility interval used to determine prior distribution for the model parameters.

Priors N (μ , σ^2)	Mutations	OR [CI]; Pr(OR>1)
Log (OR)~N (0, 1,000) OR _p [CI] ¹ = 1 [0–8.3*10 ²⁶]	1U1D	0.9 [0.3–2.7]; 42.6
	2D	2.0 [0.9–4.5]; 94.1
	2U	Reference
Log (OR)~N (0, 0.674) OR _p [CI] ¹ = 1 [0.2–5]	1U1D	0.9 [0.4–2.1]; 37.0
	2D	1.7 [0.9–3.5]; 93.6
	2U	Reference
Log (OR)~N (1.099, 0.377) OR _p [CI] ¹ = 3.0 [0.9–10]	1U1D	1.6 [0.7–3.5]; 87.8
	2D	2.6 [1.4–5.0]; 99.8
	2U	Reference

with univariate models as OR and their CI did not fundamentally vary even through the sensitivity analysis that we conducted. Both analyses were unable to show a clear relationship between the mutation type and the CS phenotype.

However, PAVs and PTVs do not seem to be randomly distributed on either side of the PiggyBac insertion and this could have biased the previous analyses, as no PAV could be identified upstream of the PiggyBac insertion. Strikingly, all 2U patients showed only PTVs (23/23), whereas in the 2D group only 55/92 patients were classified in the PTV subgroup. In the 2U group, 14/23 (61%) patients with PTVs had a moderate or mild phenotype, as compared to 24/55 (44%) for the PTVs patients in the 2D group. In the 1U1D group, the upstream mutation was always a PTV. When considering only the 2D group, the distribution of the clinical subgroups was similar between the PAVs and PTVs groups, respectively 64.9% of type II/COFS, 27% of type I, 8.1% of type III with PAVs and 56.4, 36.3, 7.3% with PTVs. The complete distribution of PAVs and PTVs is displayed in **Supplementary Material**.

DISCUSSION

CS encompasses a large spectrum of clinical presentations from the most severe prenatal subtype to the adult-onset subtype. Accumulating clinical data shows that Cockayne syndrome is actually a continuous spectrum of forms of varying severity (Laugel, 2013) and no study has brought to light a clear-cut genotype-phenotype correlation so far. Horibata and others (Horibata et al., 2004) first suggested that CSB truncations generating no functional protein resulted in the mild phenotype of UV sensitive syndrome, whereas more C-terminal truncations might generate inactive protein that could interfere with other processes, thereby resulting in more severe phenotypes. Consistently, Hashimoto *et al* also reported a mild CS phenotype linked to an early truncating mutation (Hashimoto et al., 2008). It led to the “PiggyBac hypothesis” stating that early truncating mutations upstream of the PiggyBac insertion in intron 5 (i.e., upstream of residue 466 in the protein sequence) could be paradoxically associated with the mildest phenotypes and that the CSB-PiggyBac fusion protein could

have a deleterious effect in the absence of the full length CSB protein (Newman et al., 2008; Gray et al., 2012; Weiner and Gray, 2013). The evolutionary conservation of the CSB-PiggyBac fusion protein could further substantiate the fact that it retains a biological effect and experimental data has shown that the fusion protein could indeed inhibit TCR of oxidative damage but facilitates TCR of UV damage (Bailey et al., 2012). However, later reports showed that this hypothesis could not fully account for all reported CS cases especially in some particular genetics isolates with high consanguinity (Laugel et al., 2008b; Falik-Zaccari et al., 2008).

By applying Bayesian ordinal logistic regression with a very low informative prior, we identified that the risk of having a severe phenotype was increased by 2.0 (CI: [0.9–4.5]) in patients with 2D in comparison to patients with 2U. Furthermore, with a prior slightly in favour of our hypothesis (Table 4), the point estimate was larger (2.6 [1.4–5.0]) and Pr(OR>1) reached 99.8%. These results do show a tendency in the expected direction and we could exclude the absence of effect which would have been the case if the OR had been estimated at 1 with a narrow credible interval.

Despite having this high probability of the OR being higher than 1 and a tendency to more severe features when presenting 2D, estimations were quite sensitive to prior choice. A residual uncertainty in the estimates remains which can be due either to the small amount of data available or to a small amount of prior knowledge on the subject, because of a lack of a genetic model to substantiate the relationship. Other genetic factors might, together with the site of the mutations, explain better the variability of the disease severity.

Comparing 1U1D to 2U with the low informative prior, the risk of having a severe clinical presentation was 0.9 [0.3–2.7] times higher. This result was surprising considering that the fusion protein is supposed deleterious in the absence of full length CSB and we would have therefore expected an increased risk of severe phenotype when having at least one mutation downstream of PiggyBac. An optimistic prior slightly modified these results, leading to a risk of developing a severe phenotype increased to 1.6 [0.7–3.5]. In the light of these results, we could not conclude to an effect of 1U1D on severe phenotypes. A rather important level of uncertainty remains. Estimation of the OR when exposed to 2D in comparison to 1U1D showed a risk increased by 3.9 (CI: [1.3–12.8]; Pr(OR>1) = 99.3%) to develop a type II compared to a type I. We would not have thought, *a priori*, to find such an increased risk when comparing those two mutations classes because of expected protein fusion expression for patients having either 1U1D or 2D. Comparing mean age at first symptoms showed that 1U1D patients had a mean age closer to 2U patients than to 2D patients although all CI overlapped. We chose this variable as secondary outcome because age at first symptoms is an information that contributes to ascertain classification and reflects the disease degree of severity. The fact that the 1U1D patients behave in a similar manner as 2U patients, combined with the observation that all upstream mutations are PTVs, might suggest that the deleterious effect of the CSB-PiggyBac fusion protein might have a quantitative threshold and that one allele expressing the fusion

protein is not enough to see this deleterious effect. These findings raised a question that would need further experiments to be answered. We could wonder if patients with 1U1D are more similar to 2U patients or to 2D patients taking into consideration other potential explanative factors.

To sum up, our statistical approach can reasonably confirm a relationship between the position of the mutations on either side of the PiggyBac insertion and the clinical severity. Our data suggest a protective effect when there is at least one mutation upstream of this insertion model and that upstream mutations are all truncating mutations. It is however still difficult to give a precise estimate of this effect but our data suggest that other factors are probably at play.

Other genetic factors should be considered together with the position of the mutations in the understanding of genotype-phenotype correlation. In CS-A patients bearing homozygous mutations, it has been shown previously that missense mutations appear to be more frequently associated with mild phenotypes than protein-truncating mutations (Calmels et al., 2018). Although there are fewer missense mutations than protein-truncating mutations in CS-B patients, the severity of the clinical features might also be partly related to the mutation type in CSB patients. Previous results, comparing splice site mutations to truncating mutations ground the argument that mutation type should be considered (Schalk, 2018). In this work, median severity score for CSA and CSB patients was 7 (confidence interval [3–12]) for patients having 2 splice site mutations whereas it lowered to 4 (confidence interval [0–8]) for patients with 2 truncating mutations. Splice site mutations were associated with less severe phenotypes than truncating mutations. This could be explained by the permissive nature of splice site mutations allowing synthesis of a small amount of residual normal protein. In the present study, the results of the models including PTVs/PAVs do not permit to conclude neither to the presence nor the absence of a global effect of the type of variants. This might be due to more complex and combined effects between different factors and we have observed that the analysis of the PTVs/PAVs impact on the phenotype is clearly biased by the fact that all upstream mutations are PTVs. Indeed, our results suggest that only truncating mutations can be pathogenic when located upstream of the PiggyBac insertion: this might be due to the impact of upstream mutations on the CSB full length protein or on the CSB-PiggyBac fusion protein or on both proteins. On the other hand, when considering only patients with downstream mutations (2D group), the distribution of the clinical profiles was very similar in PTVs and PAVs patients, which was not in favour of a more deleterious effect of truncating mutations downstream of the PiggyBac insertion. More data are needed to clarify the role of the type of mutations in combination with the impact of the position.

In xeroderma pigmentosum (XP) group D gene (*XPD/ERCC2*), Ueda *et al* showed that a given mutation affects specific molecular processes during transcription (Ueda et al., 2009). As a result, different biochemical phenotypes are observed for each *XPD* mutation. They supported Andressoo's biallelic hypothesis in which both alleles contribute to the phenotype (Andressoo et al., 2006). Considering this, it would be interesting

to investigate the role of the second mutation in compound heterozygous CS-B patients who have a common mutation and to see if it has, as suggested in XP patients, a major role in determining the different clinical symptoms. Comparing phenotypes between homozygous and heterozygous patients with a shared mutation and both mutations, either upstream or downstream of PiggyBac, might also be an opportunity to better understand the role of genetic effects on phenotypes.

The main strength of our study is that, to the best of our knowledge, it was the first to include as many Cockayne patients, considering the disease rarity, to study the impact of the mutation position on the CS phenotype. Moreover, the statistical approach used herein to address this problem had not been used so far. A limitation of our study is, from a statistical point of view, the rarity of the disease and the number of patients included, which hindered a more precise assessment of the statistical relationship, despite the use of reasonably informative prior parameter estimates in our models. Prospective western blot studies would also be needed to further assess the actual presence of the CSB protein, of the PiggyBac transposase and of the chimeric PiggyBac-CSB protein in patient samples to confirm the statistical prediction. Due to difficulties to determine which type of CS affects a patient, an overlap of clinical presentations as seen in our cohort, lack of thresholds between groups and potential progression within a group, clinicians from our team recently developed a severity score (Spitz et al., 2021). Using this score in the future could help to study genotype-phenotype correlation since it allows a repeated assessment through all the years of follow-up. Quantitative nature of this assessment seems to fit well to the continuous spectrum described earlier and may represent in a more accurate and reliable way CS evolution of severity.

In conclusion, our statistical approach confirms a likely deleterious effect of the PiggyBac-CSB chimeric protein in the absence of a normal full length CSB protein and that this effect participates in the resulting clinical presentation. Our results also suggest that this is probably not the only factor involved in genotype-phenotype correlation as we would then have expected a more notable effect. We still need more data to get a precise estimation of the risk associated with a downstream mutation and the potential combining effect of the type of mutation. Further data should be collected and added to the present study to reach a better understanding of genotype-phenotype correlations for CSB.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The datasets used and analyzed during the current study are available from the corresponding author on reasonable request. Requests to access these datasets should be directed to VL, vincent.laugel@chru-strasbourg.fr.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Board of Strasbourg University Hospital.

Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

RD-F, VL, and NM wrote the manuscript; RD-F and VL collected and analyzed the data; RD-F and NM performed the statistical analyses; CO, NLM, and NC contributed in collecting genetic and clinical data and in reviewing the manuscript. All authors read and approved the final manuscript.

REFERENCES

- Andressoo, J.-O., Jans, J., de Wit, J., Coin, F., Hoogstraten, D., de Ven, M. v., et al. (2006). Rescue of Progeria in Trichothiodystrophy by Homozygous Lethal Xpd Alleles. *Plos Biol.* 4 (10), e322. doi:10.1371/journal.pbio.0040322
- Bailey, A. D., Gray, L. T., Pavelitz, T., Newman, J. C., Horibata, K., Tanaka, K., et al. (2012). The Conserved Cockayne Syndrome B-piggyBac Fusion Protein (CSB-PGBD3) Affects DNA Repair and Induces Both Interferon-like and Innate Antiviral Responses in CSB-Null Cells. *DNA Repair* 11 (5), 488–501. doi:10.1016/j.dnarep.2012.02.004
- Brumback, R. A., Yoder, F. W., Andrews, A. D., Peck, G. L., and Robbins, J. H. (1978). Normal Pressure Hydrocephalus. *Arch. Neurol.* 35 (6), 337–345. doi:10.1001/archneur.1978.00500300011002
- Calmels, N., Botta, E., Jia, N., Fawcett, H., Nardo, T., Nakazawa, Y., et al. (2018). Functional and Clinical Relevance of Novel Mutations in a Large Cohort of Patients with Cockayne Syndrome. *J. Med. Genet.* 55 (5), 329–343. doi:10.1136/jmedgenet-2017-104877
- Cockayne, E. A. (1946). Dwarfism with Retinal Atrophy and Deafness. *Arch. Dis. Child.* 21 (105), 52–54. doi:10.1136/adc.21.105.52
- Colella, S., Nardo, T., Mallery, D., Borrone, C., Ricci, R., Ruffa, G., et al. (1999). Alterations in the CSB Gene in Three Italian Patients with the Severe Form of Cockayne Syndrome (CS) but without Clinical Photosensitivity. *Hum. Mol. Genet.* 8 (5), 935–941. doi:10.1093/hmg/8.5.935
- Dunson, D. B. (2001). Commentary: Practical Advantages of Bayesian Analysis of Epidemiologic Data. *Am. J. Epidemiol.* 153 (12), 1222–1226. doi:10.1093/ajep/153.12.1222
- Falík-Zaccari, T. C., Laskar, M., Kfir, N., Nasser, W., Slor, H., and Khayat, M. (2008). Cockayne Syndrome Type II in a Druze Isolate in Northern Israel in Association with an Insertion Mutation in ERCC6. *Am. J. Med. Genet.* 146a (11), 1423–1429. doi:10.1002/ajmg.a.32309
- Ferreira, D., Barthoulot, M., Pottecher, J., Torp, K. D., Diemunsch, P., and Meyer, N. (2020). Theory and Practical Use of Bayesian Methods in Interpreting Clinical Trial Data: a Narrative Review. *Br. J. Anaesth.* 125 (2), 201–207. doi:10.1016/j.bja.2020.04.092
- Graham, J. M., Jr., Anyane-Yeboah, K., Raams, A., Appeldoorn, E., Kleijer, W. J., Garritsen, V. H., et al. (2001). Cerebro-oculo-facio-skeletal Syndrome with a Nucleotide Excision-Repair Defect and a Mutated XPD Gene, with Prenatal Diagnosis in a Triplet Pregnancy. *Am. J. Hum. Genet.* 69 (2), 291–300. doi:10.1086/321295
- Gray, L. T., Fong, K. K., Pavelitz, T., and Weiner, A. M. (2012). Tethering of the Conserved piggyBac Transposase Fusion Protein CSB-PGBD3 to Chromosomal AP-1 Proteins Regulates Expression of Nearby Genes in Humans. *Plos Genet.* 8 (9), e1002972. doi:10.1371/journal.pgen.1002972
- Hashimoto, S., Suga, T., Kudo, E., Ihn, H., Uchino, M., and Tateishi, S. (2008). Adult-onset Neurological Degeneration in a Patient with Cockayne Syndrome and a Null Mutation in the CSB Gene. *J. Invest. Dermatol.* 128 (6), 1597–1599. doi:10.1038/sj.jid.5701210
- Henning, K. A., Li, L., Iyer, N., McDaniel, L. D., Reagan, M. S., Legerski, R., et al. (1995). The Cockayne Syndrome Group A Gene Encodes a WD Repeat Protein that Interacts with CSB Protein and a Subunit of RNA Polymerase II TFIIF. *Cell* 82 (4), 555–564. doi:10.1016/0092-8674(95)90028-4

ACKNOWLEDGMENTS

The authors would like to thank the Cockayne syndrome family support groups (“Amy and friends” and “Les P’tits Bouts”) for their help and dedication and the referring clinicians.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.762047/full#supplementary-material>

- Horibata, K., Iwamoto, Y., Kuraoka, I., Jaspers, N. G. J., Kurimasa, A., Oshimura, M., et al. (2004). From the Cover: Complete Absence of Cockayne Syndrome Group B Gene Product Gives Rise to UV-Sensitive Syndrome but Not Cockayne Syndrome. *Proc. Natl. Acad. Sci.* 101 (43), 15410–15415. doi:10.1073/pnas.0404587101
- Kleijer, W. J., Laugel, V., Berneburg, M., Nardo, T., Fawcett, H., Gratchev, A., et al. (2008). Incidence of DNA Repair Deficiency Disorders in Western Europe: Xeroderma Pigmentosum, Cockayne Syndrome and Trichothiodystrophy. *DNA Repair* 7 (5), 744–750. doi:10.1016/j.dnarep.2008.01.014
- Kou, Y., Shboul, M., Wang, Z., Shersheer, Q., Lyu, Z., Liu, P., et al. (2018). Novel Frame Shift Mutation in ERCC6 Leads to a Severe Form of Cockayne Syndrome with Postnatal Growth Failure and Early Death. *Medicine (Baltimore)* 97 (33), e11636. doi:10.1097/md.00000000000011636
- Laugel, V. (2013). Cockayne Syndrome: the Expanding Clinical and Mutational Spectrum. *Mech. Ageing Dev.* 134 (5–6), 161–170. doi:10.1016/j.mad.2013.02.006
- Laugel, V., Dalloz, C., Durand, M., Sauvanaud, F., Kristensen, U., Vincent, M. C., et al. (2010). Mutation Update for the CSB/ERCC6 and CSA/ERCC8 genes Involved in Cockayne Syndrome. *Hum. Mutat.* 31 (2), 113–126. doi:10.1002/humu.21154
- Laugel, V., Dalloz, C., Stary, A., Cormier-Daire, V., Desguerre, I., Renouil, M., et al. (2008). Deletion of 5' Sequences of the CSB Gene Provides Insight into the Pathophysiology of Cockayne Syndrome. *Eur. J. Hum. Genet.* 16 (3), 320–327. doi:10.1038/sj.ejhg.5201991
- Laugel, V., Dalloz, C., Tobias, E. S., Tolmie, J. L., Martin-Coignard, D., Drouin-Garraud, V., et al. (2008). Cerebro-oculo-facio-skeletal Syndrome: Three Additional Cases with CSB Mutations, New Diagnostic Criteria and an Approach to Investigation. *J. Med. Genet.* 45 (9), 564–571. doi:10.1136/jmg.2007.057141
- Lehmann, A. R., Thompson, A. F., Harcourt, S. A., Stefanini, M., and Norris, P. G. (1993). Cockayne's Syndrome: Correlation of Clinical Features with Cellular Sensitivity of RNA Synthesis to UV Irradiation. *J. Med. Genet.* 30 (8), 679–682. doi:10.1136/jmg.30.8.679
- Mallery, D. L., Tanganelli, B., Colella, S., Steingrimsdottir, H., van Gool, A. J., Troelstra, C., et al. (1998). Molecular Analysis of Mutations in the CSB(ERCC6) Gene in Patients with Cockayne Syndrome. *Am. J. Hum. Genet.* 62 (1), 77–85. doi:10.1086/301686
- Mayne, L. V., and Lehmann, A. R. (1982). 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.* 42 (4), 1473–1478.
- Meira, L. B., Graham, J. M., Jr., Greenberg, C. R., Busch, D. B., Doughty, A. T. B., Ziffer, D. W., et al. (2000). Manitoba Aboriginal Kindred with Original Cerebro-Oculo-Facio-Skeletal Syndrome Has a Mutation in the Cockayne Syndrome Group B (CSB) Gene. *Am. J. Hum. Genet.* 66 (4), 1221–1228. doi:10.1086/302867
- Nance, M. A., and Berry, S. A. (1992). Cockayne Syndrome: Review of 140 Cases. *Am. J. Med. Genet.* 42 (1), 68–84. doi:10.1002/ajmg.1320420115
- Natale, V. (2011). A Comprehensive Description of the Severity Groups in Cockayne Syndrome. *Am. J. Med. Genet.* 155 (5), 1081–1095. doi:10.1002/ajmg.a.33933
- Newman, J. C., Bailey, A. D., Fan, H.-Y., Pavelitz, T., and Weiner, A. M. (2008). An Abundant Evolutionarily Conserved CSB-PiggyBac Fusion Protein Expressed

- in Cockayne Syndrome. *Plos Genet.* 4 (3), e1000031. doi:10.1371/journal.pgen.1000031
- Sanchez-Roman, I., Lautrup, S., Aamann, M. D., Neilan, E. G., Østergaard, J. R., and Stevnsner, T. (2018). Two Cockayne Syndrome Patients with a Novel Splice Site Mutation - Clinical and Metabolic Analyses. *Mech. Ageing Dev.* 175, 7–16. doi:10.1016/j.mad.2018.06.001
- Schalk, A. (2018). Les mutations d'épissage dans le syndrome de Cockayne: description d'un cas de mutation intronique profonde et corrélation génotype-phénotype. [Medical doctor thesis]. Strasbourg: University of Strasbourg.
- Sin, Y., Makimura, F., Saijo, M., and Obika, S. (2018). Generation of Splice Switching Oligonucleotides Targeting the Cockayne Syndrome Group B Gene Product in Order to Change the Diseased Cell State. *Biochem. Biophysical Res. Commun.* 500 (2), 163–169. doi:10.1016/j.bbrc.2018.04.015
- Spitz, M. A., Severac, F., Obringer, C., Baer, S., Le May, N., Calmels, N., et al. (2021). Diagnostic and Severity Scores for Cockayne Syndrome. *Orphanet J. Rare Dis.* 16 (1), 63. doi:10.1186/s13023-021-01686-8
- Stefanini, M., Fawcett, H., Botta, E., Nardo, T., and Lehmann, A. R. (1996). Genetic Analysis of Twenty-Two Patients with Cockayne Syndrome. *Hum. Genet.* 97 (4), 418–423. doi:10.1007/bf02267059
- Troelstra, C., van Gool, A., de Wit, J., Vermeulen, W., Bootsma, D., and Hoeijmakers, J. H. J. (1992). ERCC6, a Member of a Subfamily of Putative Helicases, Is Involved in Cockayne's Syndrome and Preferential Repair of Active Genes. *Cell* 71 (6), 939–953. doi:10.1016/0092-8674(92)90390-x
- Ueda, T., Compe, E., Catez, P., Kraemer, K. H., and Egly, J.-M. (2009). Both XPD Alleles Contribute to the Phenotype of Compound Heterozygote Xeroderma Pigmentosum Patients. *J. Exp. Med.* 206 (13), 3031–3046. doi:10.1084/jem.20091892
- Weiner, A. M., and Gray, L. T. (2013). What Role (If Any) Does the Highly Conserved CSB-PGBD3 Fusion Protein Play in Cockayne Syndrome. *Mech. Ageing Dev.* 134 (5-6), 225–233. doi:10.1016/j.mad.2013.01.001
- Wilson, B. T., Lochan, A., Stark, Z., and Sutton, R. E. (2016). Novel Missense Mutations in a Conserved Loop between ERCC6 (CSB) Helicase Motifs V and VI: Insights into Cockayne Syndrome. *Am. J. Med. Genet.* 170 (3), 773–776. doi:10.1002/ajmg.a.37501
- Wilson, B. T., Stark, Z., Sutton, R. E., Danda, S., Ekbote, A. V., Elsayed, S. M., et al. (2016). The Cockayne Syndrome Natural History (CoSyNH) Study: Clinical Findings in 102 Individuals and Recommendations for Care. *Genet. Med.* 18 (5), 483–493. doi:10.1038/gim.2015.110

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Damaj-Fourcade, Meyer, Obringer, Le May, Calmels and Laugel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.