The Journal of Pathology: Clinical Research
J Pathol Clin Res September 2022; **8:** 470–480
Published online 14 July 2022 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cjp2.287

ORIGINAL ARTICLE

Paediatric *BCOR*-associated sarcomas with a novel long spliced internal tandem duplication of *BCOR* exon 15

Jian Yuan Goh[1,2†], Chik Hong Kuick[1†], Masahiro Sugiura[1], Sze Jet Aw[1], Manli Zhao[3], Hongfeng Tang[3], Sandini Gunaratne[4], Fucun Zhu[5], Lin Cai[5], Bin Tean Teh[6,7], Paul S Thorner[8] and Kenneth Tou En Chang[1,2*]

[1]*Department of Pathology and Laboratory Medicine, KK Women's and Children's Hospital, Singapore*
[2]*Pathology Academic Clinical Programme, SingHealth Duke-NUS Medical School, Singapore*
[3]*Department of Pathology, The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, PR China*
[4]*Department of Pathology, Lady Ridgeway Hospital for Children, Colombo, Sri Lanka*
[5]*Department of Pathology, Fuzhou Children's Hospital of Fujian Province, Fuzhou, PR China*
[6]*Laboratory of Cancer Epigenome, National Cancer Centre Singapore, Singapore*
[7]*Cancer and Stem Cell Biology Programme, Duke-NUS Medical School, Singapore*
[8]*Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada*

**Correspondence to: Kenneth Tou En Chang, Department of Pathology and Laboratory Medicine, KK Women's and Children's Hospital, 100 Bukit Timah Road, Singapore 229899. E-mail: kenneth.chang.t.e@singhealth.com.sg*

[†]*These authors contributed equally to the study.*

Abstract

Clear cell sarcoma of the kidney (CCSK) and primitive myxoid mesenchymal tumour of infancy (PMMTI) are paediatric sarcomas that most commonly harbour internal tandem duplications (ITDs) of exon 15 of the *BCOR* gene, in the range of 87–114 base pairs (bp). Some cases, instead, have *BCOR–CCNB3* or *YWHAE–NUTM2* gene fusions. About 10% of cases lack any of these genetic alterations when tested by standard methods. Two cases of CCSK and one PMMTI lacking the aforementioned mutations were analysed using Archer FusionPlex technology. Two related *BCOR* exon 15 RNA transcripts with ITDs of lengths 388 and 96 bp were detected in each case; only the 388 bp transcript was identified when genomic DNA was sequenced. *In silico* analysis of this transcript revealed acceptor and donor splice sites indicating that, at the RNA level, the 388-bp transcript was likely spliced to form the 96-bp transcript. The results were confirmed by Sanger sequencing using primers targeting the ITD breakpoint. This novel and unusually long ITD segment is difficult to identify by DNA sequencing using typical primer design strategies flanking entire duplicated segments because it exceeds the typical read lengths of most sequencing platforms as well as the usual fragment lengths obtained from formalin-fixed paraffin-embedded material. As diagnosis of CCSK and PMMTI may be challenging by morphology and immunohistochemistry alone, it is important to identify mutations in these cases. Knowledge of this novel *BCOR* ITD is important in relation to primer design for detection by sequencing, and using RNA versus DNA for sequencing.

Keywords: clear cell sarcoma of kidney; primitive myxoid mesenchymal tumour of infancy; *BCOR*; internal tandem duplication; splicing

Received 24 February 2022; Revised 14 May 2022; Accepted 22 June 2022

No conflicts of interest were declared.

Introduction

Clear cell sarcoma of the kidney (CCSK) is a rare malignancy but the second most common paediatric renal tumour after Wilms tumour, accounting for up to 5% of all primary renal tumours in childhood [1]. The tumour typically presents at 2–4 years of age and there is a striking male predominance. There are no known associations with any genetic predisposition syndrome, nor are there familial cases. Multiagent chemotherapy and radiation therapy have achieved a 5-year overall survival of 75–90% [1], but the tumour can metastasise to brain and bone, or recur, in which case the outcome is poor [2,3]. As the treatment and

prognosis differ from Wilms tumour, accurate diagnosis of CCSK is important, yet challenging on biopsy alone, due to the different histological appearances this tumour may take. The so-called classic pattern is one of the ovoid cells with monomorphic nuclei and clear cytoplasm, arranged in broad trabeculae, separated by fibrovascular septa forming a distinctive arborising pattern and containing thin-walled capillaries. This pattern is the most common and present in most cases, at least focally. More problematic, however, are variant histologies that include cellular, myxoid, sclerosing, epithelioid, palisading, and spindle cell patterns. The cellular pattern can be confused with other blastemal malignancies (i.e. Wilms tumour and neuroblastoma); the epithelioid pattern mimics epithelial differentiation and the spindle cell pattern may resemble other soft tissue sarcomas [4,5]. While immunohistochemistry is helpful, it does not entirely solve the problem. Nerve growth factor receptor and cyclin D1 are sensitive immunohistochemical markers of CCSK [6–8], but neither is specific. Nerve growth factor receptor can label tumours of neural phenotype, and cyclin D1 can be also positive in other tumours occurring in the region of the kidney including congenital mesoblastic nephroma, epithelial Wilms tumour, and neuroblastoma [6,9]. EZH2 is highly expressed, but this is also the case in a wide variety of other tumours [9].

Primitive myxoid mesenchymal tumour of infancy (PMMTI) can be considered as the soft tissue counterpart of CCSK [10], affecting young children, especially infants, and typically occurring on the trunk, extremities, head, and neck [11]. Local recurrence and metastatic spread occur, each in 44% of patients, with an overall 5-year survival of 34% [12]. The histological features are reminiscent of CCSK, with round to spindle-shaped cells possessing uniform round-to-oval nuclei, situated in abundant myxoid stroma, separated by cellular fibrous septa containing a vascular network. Some cases show microcysts or rosettes. Given the age of the patient and the soft tissue location, PMMTI is most often confused with infantile fibrosarcoma [10,11,13]. As with CCSK, PMMTI shows strong and diffuse nuclear immunoreactivity for cyclin D1 [12].

A major advance in the diagnosis of CCSK and PMMTI has come from molecular genetic studies. The majority of cases of CCSK and PMMTI have an internal tandem duplication (ITD) involving varying segments of exon 15 of the *BCOR* (BCL-6 co-repressor) gene [10,12–17]. A small proportion of CCSK have, instead, a *YWHAE-NUTM2* gene fusion [18], but this genetic change has not been reported for PMMTI. There are also case reports of other genetic changes, such as *BCOR-CCNB3* gene fusion [8]. There remain, however, about 10% of cases that do not show any of these mutations by standard methods (see Detection of *BCOR* ITDs in Materials and methods section) typically used in diagnostic laboratories [19], and thus the important diagnosis of CCSK or PMMTI cannot be confirmed in these patients. Immunohistochemistry for BCOR has been proposed as a possible substitute for performing molecular genetic analysis of the *BCOR* gene, as both CCSK and PMMTI show diffuse and strong immunoreactivity [12,16,20]. However, BCOR immunostaining lacks sensitivity, and tumours with proven ITDs can have negative staining [8], making a molecular approach the preferred one for confirmation of a histological diagnosis of CCSK and PMMTI.

In this report, we looked at that 10% of cases that do not have *BCOR* ITD or a gene fusion. From a collection of 32 cases of CCSK and PMMTI in our pathology archives, there were 3 cases, 2 CCSK and 1 PMMTI, which lacked a molecular diagnosis. The aim of this study was to interrogate those three cases, in order to determine whether there is a yet unrecognised gene involved in CCSK/PMMTI, or whether these cases conform to already known changes. It turned out that the latter hypothesis was correct, namely all three cases had *BCOR* ITD, but in the novel form of a long ITD of *BCOR* exon 15, which undergoes subsequent splicing.

Case histories

Patient 1 is a 1-year-old male who presented with a left renal mass measuring 10 cm in greatest dimension. A core biopsy was performed, and the initial diagnosis was Wilms tumour. Neo-adjuvant chemotherapy for Wilms tumour was administered. A diagnosis of CCSK was made after radical nephrectomy was performed. The patient received additional adjuvant anthracycline-based chemotherapy. The patient relapsed 1 year later, and had surgical resection followed by chemotherapy. This was complicated by sepsis. The child died 3 years after the initial diagnosis.

Patient 2 is a 2-year-old female who presented with a renal mass measuring 4 cm in greatest dimension. Primary nephrectomy was performed. Details of tumour stage were not available. There was tumour recurrence 13 months later at the renal bed with metastases to the anterior abdominal wall and greater omentum. The child underwent surgical resection followed by chemotherapy and remains well at 2 years' follow-up.

Patient 3 is a 5-month-old male who presented with a right neck mass measuring 5.5 cm in greatest

dimension. The mass was resected with positive resection margins. The family declined chemotherapy. Several months later, the patient had widespread metastasis and died shortly thereafter.

Materials and methods

Pathology and immunohistochemistry

Thirty-one cases of CCSK and one case of PMMTI were collected from the archives at the Department of Pathology and Laboratory Medicine, KK Women's and Children's Hospital. All cases were inpatient specimens or referral cases, and all cases were reviewed by a paediatric pathologist (KTEC). The details of some of the CCSK cases have been previously published [6,8]. Twenty-eight cases of CCSK had typical ITDs involving *BCOR* exon 15 and one case of CCSK had a *BCOR-CCNB3* fusion. Three cases (two CCSK and one PMMTI) had no molecular genetic change detected, and these three cases form the basis of this report. Formalin-fixed paraffin-embedded (FFPE) tumour was available for these three cases. Immunohistochemistry was performed for BCOR (mouse monoclonal antibody, dilution 1:100, clone C-10, sc-514 576; Santa Cruz, Dallas, TX, USA) and cyclin D1 (rabbit monoclonal antibody, dilution 1:50, clone SP4; Thermo Fisher Scientific, Waltham, MA, USA) following technical protocols described previously [6,8]. Ethics approval was obtained (SingHealth CIRB 2020/3000).

Gene fusion detection by NanoString nCounter

The NanoString nCounter assay comprises a customised CodeSet of gene-specific DNA probes covering 174 specific gene fusion variants across 24 sarcoma types [21]. The assay includes probes targeting the specific *YWHAE-NUTM2* and *BCOR-CCNB3* fusions. The technical protocol has been previously described [8,21].

Detection of BCOR ITDs

Our standard method for detection of *BCOR* ITDs is by DNA-based polymerase chain reaction (PCR) and Sanger sequencing. Two sets of primer pairs are used that are capable of identifying the full range of *BCOR* exon 15 ITDs described to date [8]. After this approach failed to detect ITDs in our three cases, we hypothesised that there might exist an ITD that is too large to be detected by this approach. We therefore performed a customised anchored multiplex PCR targeted next-generation sequencing (NGS) assay (Archer® FusionPlex, Bolder, CO, USA) that utilised RNA extracted from FFPE tumour sections as starting material, and that included primers to exons 6, 7, 8, 12, 14, and 15 of *BCOR*. In brief, total RNA was extracted from FFPE tissue sections from each of our three cases and quantified with a fluorometer. One hundred and fifty nanograms of RNA was used for library preparation with the Archer® FusionPlex Pan-Solid Tumour panel. The prepared library was sequenced using an Illumina MiniSeq sequencer (San Diego, CA, USA). FASTQ data obtained were analysed with the Archer Analysis (version 6.2.3) online portal. Two novel ITDs of different size were detected, identical for all three cases (see Results section for details).

Confirmation of *BCOR* ITD results

To confirm the results of the Archer FusionPlex assay, we performed PCR followed by Sanger sequencing. For detection of the *BCOR* ITD from RNA samples, 50 ng of total RNA isolated from FFPE sections was reverse transcribed into cDNA and used for qPCR on the GoTaq 1-step reverse transcription (RT)-qPCR system (Promega, Madison, WI, USA) [8]. For detection of the *BCOR* ITD from genomic DNA samples, 100 ng of genomic DNA isolated from FFPE sections was used for qPCR on the same GoTaq qPCR system.

Primer pairs were designed to confirm the breakpoint of the ITDs. The primer sequences used to amplify the shorter transcript breakpoint are:

forward primer: 5′-GATCTGGCCTCAGACAACTAC-3′
reverse primer: 5′-TACAGAGGAGCCCAGCA-3′.

Primer sequences used to amplify the longer transcript breakpoint are:

forward primer: 5′-CCAGTGATCTGGCCTCAG-3′
reverse primer: 5′-TACAGAGGAGCCCAGCA-3′.

PCR products from RT-qPCR and qPCR were purified with ProNex Size Selective Purification System (Promega) following the manufacturer's protocol. Five nanograms of purified PCR product was used for Sanger sequencing with SeqStudio Genetic Analyzer (Thermo Fisher, Waltham, MA, USA). Results were visualised and analysed with Chromas software.

Results

Histological features

All three tumours were histologically compatible with their respective diagnoses of CCSK and PMMTI

(Figure 1). These cases were typical examples of these two tumours, in both clinical and pathological aspects. Case 1 had characteristic classical histological features of a CCSK (Figure 1A), while case 2 had a predominantly myxoid histological appearance (Figure 1C). As befitting a PMMTI, case 3 was extensively myxoid with scattered microcysts (Figure 1E). All three cases showed diffuse and strong nuclear immunoreactivity for cyclin D1 (not shown) and BCOR (Figure 1B,D,F).

Molecular pathology findings

All three cases lacked *YWHAE-NUTM2* and *BCOR-CCNB3* gene fusions (results not shown). *BCOR* exon 15 ITDs were not detected using methodology capable

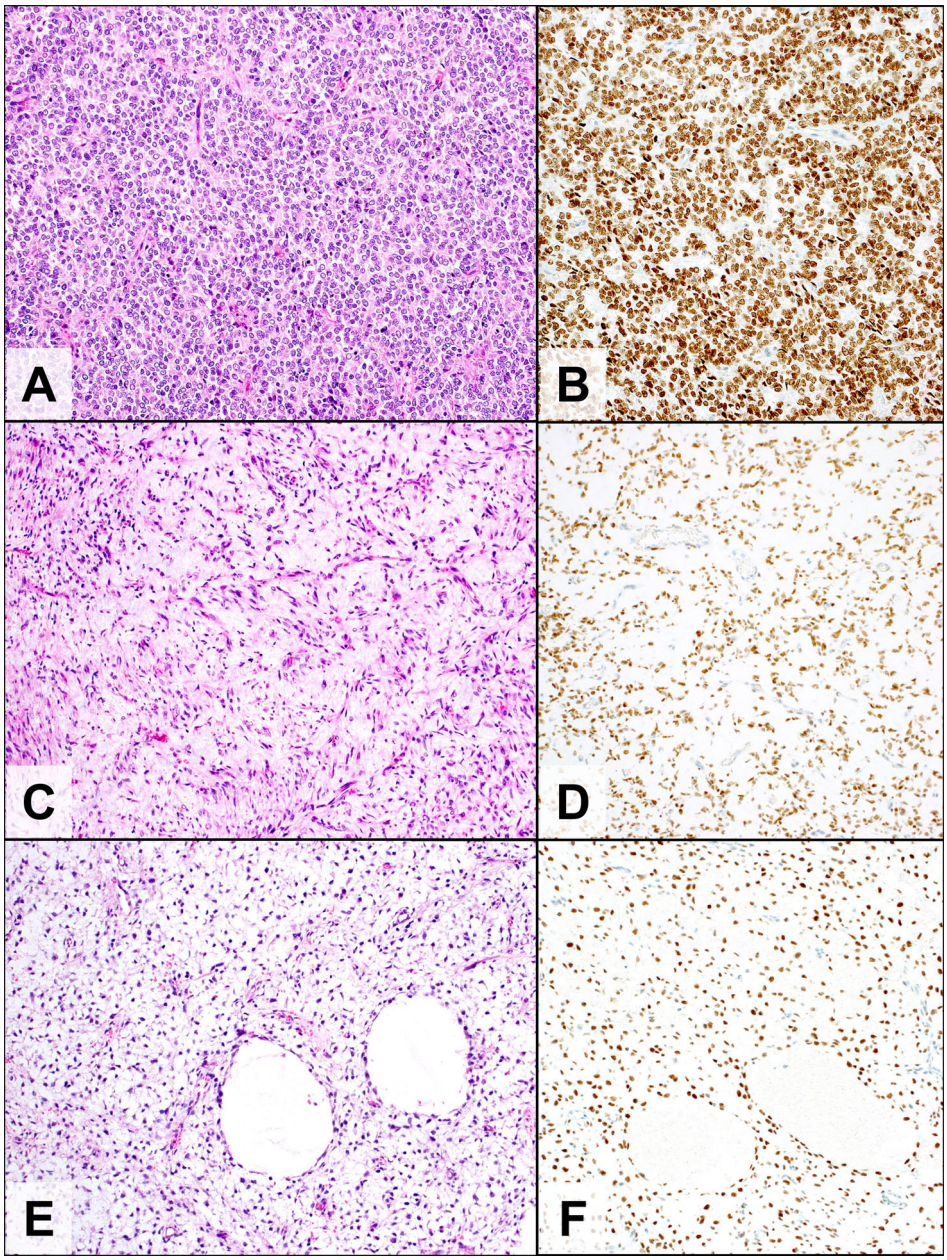


Figure 1. Histological features and BCOR immunohistochemistry (A, C, and E: haematoxylin and eosin, original total magnifications ×200; B, D, and F: immunohistochemistry, original total magnifications ×200). (A and B) Case 1 is a CCSK with classic histology (A) while (C and D) case 2 is a CCSK with a predominantly myxoid histological appearance (C). (E and F) Case 3 is a PMMTI that is extensively myxoid with scattered microcysts (E). All three cases show diffuse and strong nuclear immunoreactivity for BCOR.

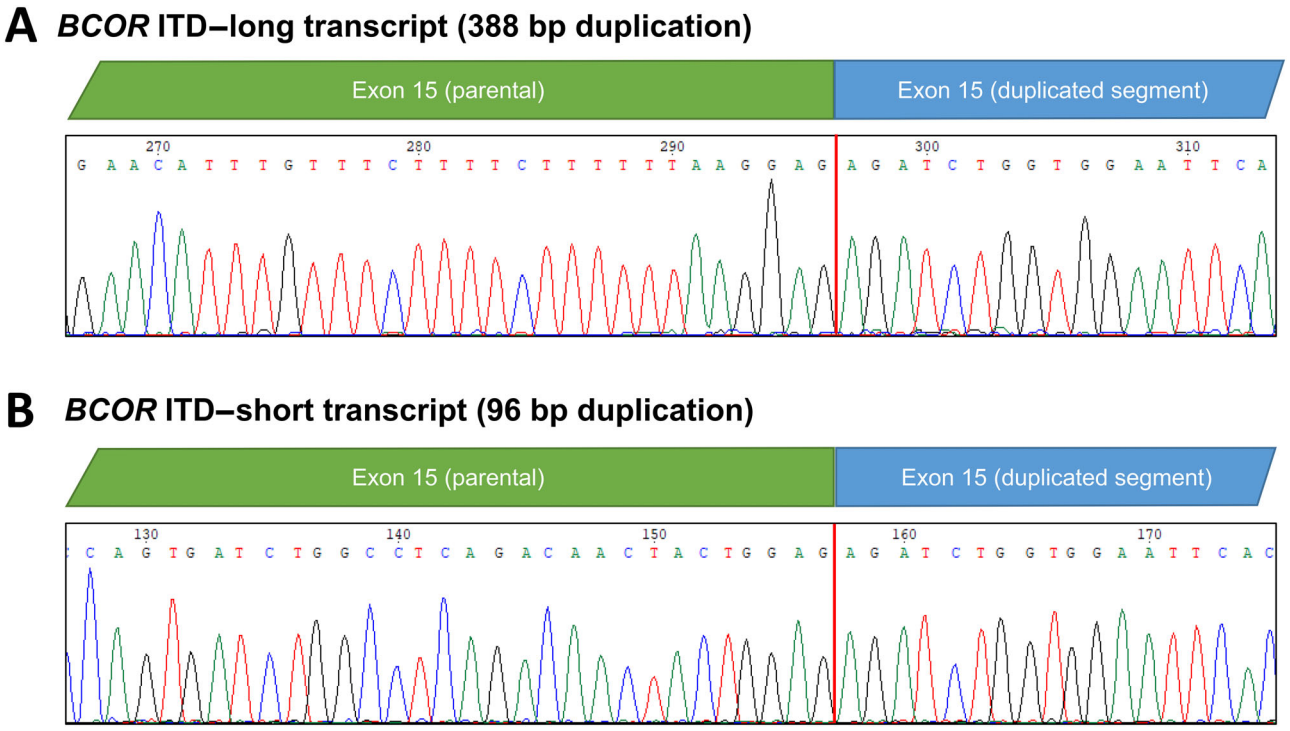
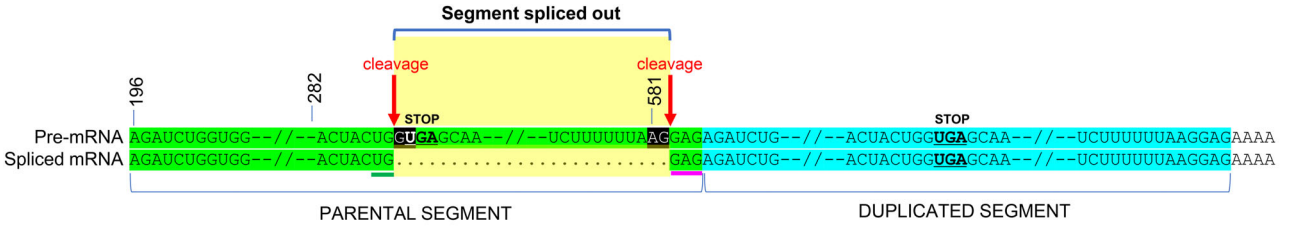


Figure 2. Two novel *BCOR* ITD transcripts with different duplication size, (A) 388 bp and (B) 96 bp, were detected by anchored multiplex PCR analysis (Archer FusionPlex pan-solid tumour assay). RT-PCR and Sanger sequencing show the breakpoint sequence for the two transcripts. The red line indicates the junction between the duplicated segments. The parental segment is upstream (left of the red line) and marked with the green bar, while the duplicated segment is downstream (right of the red line) and marked with the blue bar. The numerical figures (270, 280, etc.) indicate nucleotide numbers in *BCOR* exon 15.



Prediction Tool	Threshold	DONOR SITE PREDICTION			ACCEPTOR SITE PREDICTION		
		Position	Score	Site Sequence	Position	Score	Site Sequence
NNSPLICE	≥0.4	282	0.98	Acuacug|GUgagcaa	560	0.99	uuguuucuuuucuuuuuuaAG|gagagaucugguggaauuca
ESE Finder	≥6.67	274	9.63	cucagacaacuacug|GUgagcaagcuggac	566	9.97	cuuuucuuuuuuaAG|gagagaucuggugga

Figure 3. Alternative splicing of the pre-mRNA *BCOR*-ITD (388 bp) transcript (top panel) and *in silico* prediction of natural splice sites in the long ITD (388 bp) sequence (bottom panel) using splice site prediction tools NNSPLICE and ESE Finder. The thresholds are values above which a score for a given sequence is considered to be significant as a donor or acceptor splice site. The actual scores obtained can be seen above the threshold scores, indicating that these sequences are predicted by the prediction tools to have a high likelihood of being donor and acceptor sites. The parental and duplicated segments are marked in green and blue, respectively. The cleavage sites are marked with red arrows. The segment that is spliced out is highlighted in the yellow box. The stop codon uracil-guanine-adenine (UGA) is underlined. The numerical figures (196, 282, and 581) indicate nucleotide numbers in *BCOR* exon 15. Following splicing, the mRNA sequence becomes UG-GAGA (the dash represents the site of splicing).

of detecting all ITD variants described to date [8]. Archer FusionPlex utilising total RNA extracted from FFPE tumour samples identified two unique yet related *BCOR* exon 15 ITD transcripts in all three cases (Figure 2). The longer transcript was 388 base pairs (bp) in length and extended beyond a stop codon into the 3′ untranslated region of the mRNA. The shorter transcript was 96 bp in length with a 5′ segment identical to the longer transcript. The longer (388 bp) transcript was present in lower abundance (~11%) compared to the shorter (96 bp) transcript (~83%). To validate these results, RT-PCR flanking the breakpoint sequences was performed using the same RNA extracted from the three cases. Sanger sequencing of the products confirmed the ITD sequences for both the longer and shorter transcripts. Next, to determine if the sequences corresponding to both the longer and shorter transcripts were present at the DNA level, we performed PCR and Sanger sequencing using genomic DNA extracted from the same FFPE tumour samples. At the DNA level, only the sequence corresponding to the longer RNA transcript could be identified in the three cases.

To explain this result, we hypothesised that the shorter transcript identified in the RNA-based assays might have been processed from the longer transcript. To test this hypothesis, we analysed the exon 15 sequence of the long ITD sequence for splice sites,

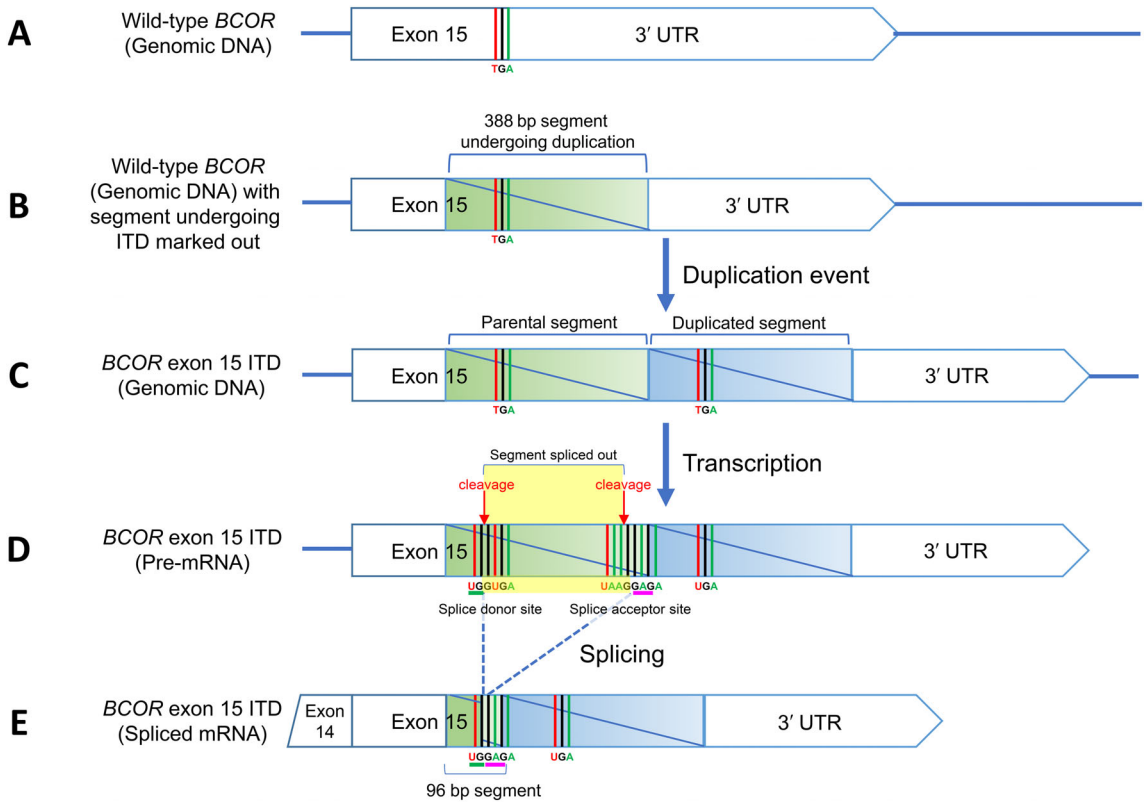


Figure 4. Summary schematic diagram of the duplication event and mRNA splicing to form the final 96 bp ITD transcript. (A) Representation of wild-type *BCOR* exon 15 DNA with the stop codon (TGA) indicated, followed by the 3′ untranslated region (UTR). (B) Identical to (A), with the 388-bp segment that undergoes ITD marked in green. The differential shading of green (darker on the left or 5′ end, and lighter on the right or 3′ end) and the oblique line are simply markers for 5′/3′ orientation. (C) *BCOR* genomic DNA with the occurrence of the ITD. The parental segment is marked in green, and the duplicated segment is marked in blue. Note that the stop codon (TGA) is duplicated in the duplicated segments. (D) Representation of pre-mRNA following transcription of (C), and the parental and duplicated segments are transcribed without sequence change. The stop codon is now represented as UGA in RNA. The splice donor (UGGUGA) and acceptor sites (UAAGGAGA) are indicated and correspond to those depicted in Figure 3. The cleavage sites are marked with the red arrows. The segment that is spliced out is highlighted in the yellow box. Note the sequences of the splice donor site viz. UG [cleavage] GUGA and splice acceptor site viz. UAAG [cleavage] GAGA. (E) Representation of the spliced mRNA. Following splicing, the sequence becomes UG-GAGA (the dash represents the site of splicing). Note that splicing occurs only in the parental segment and not in the duplicated segment.

Table 1. Likelihood of published *BCOR* ITD detection methods to detect the novel long spliced ITD described in this paper

Study	Nucleic acid used for testing	Primer design strategy*	Amplicon length (bp) in *BCOR* ITD from genomic DNA	Detectable with DNA?†	Detectable with RNA?‡
Marino-Enriquez *et al.* 2018 [36]	DNA	1	587	No	Amplicon of 295 bp
Yoshida *et al.* 2018 [37]	DNA	1	654, 545	No	Amplicon of 253 bp
Ueno-Yokohata *et al.* 2015 [17]	DNA and RNA	1	954, 950	No	No
Roy *et al.* 2015 [16]	DNA and RNA	1	640	No	Amplicon of 348 bp
Kenny *et al.* 2016 [19]	DNA and RNA	1	660, 623	No	Amplicon of 331 bp
Astolfi *et al.* 2015 [15]	RNA	1	609	No	Amplicon of 317 bp
Kao *et al.* 2016 [10]	DNA and RNA	1	654, 545	No	Amplicon of 253 bp
Chiang *et al.* 2017 [38]	DNA	1	654, 545	No	Amplicon of 253 bp
Paret *et al.* 2016 [39]	RNA	1	1006	No	No
Wong *et al.* 2018 [8]	DNA and RNA	1	623	No	Amplicon of 331 bp
Appay *et al.* 2017 [40]	DNA	1	660	No	Amplicon of 368 bp
Al-Ibraheemi *et al.* 2021 [14]	DNA and RNA	2	588	No	Amplicon of 296 bp
Bouchoucha *et al.* 2022 [35]	DNA and RNA	1	646	No	Amplicon of 354 bp

*All studies utilise primer design strategy 1 (see Figure 5) with the exception of the study by Al-Ibraheemi *et al* [14] that uses primer design strategy 2.
†This novel long ITD will not ordinarily be detectable using DNA extracted from clinical FFPE material as amplicon lengths exceed fragment lengths usually present in FFPE material.
‡Using RNA extracted from clinical FFPE material, this novel ITD may be detected if RNA quality is sufficiently good with fragment lengths exceeding the stated amplicon sizes.

using *in silico* online prediction tools NNSPLICE [22] and ESE Finder 3.0 [23,24]. Both prediction tools identified new acceptor splicing sites that were created by the duplicated DNA sequence close to the junction between parental and duplicated segments, along with an internal donor splice site present in the ITD sequence (Figure 3). We found that the longer ITD transcript had a segment spliced out to generate the shorter ITD transcript that was detected only in the RNA samples. Furthermore, as a result of the splicing process, the original thymine-guanine-adenine (TGA) stop codon present in the parental segment of the long ITD transcript was removed, allowing translation of the ITD sequence, with the short ITD transcript present as the mature mRNA. The longer ITD transcript therefore serves as precursor mRNA. This process is schematically summarised in Figure 4.

Discussion

The initial breakthrough on the molecular basis of CCSK was the identification of a translocation fusing exons 1–5 of *YWHAE* (tyrosine 3-monooxygenase/

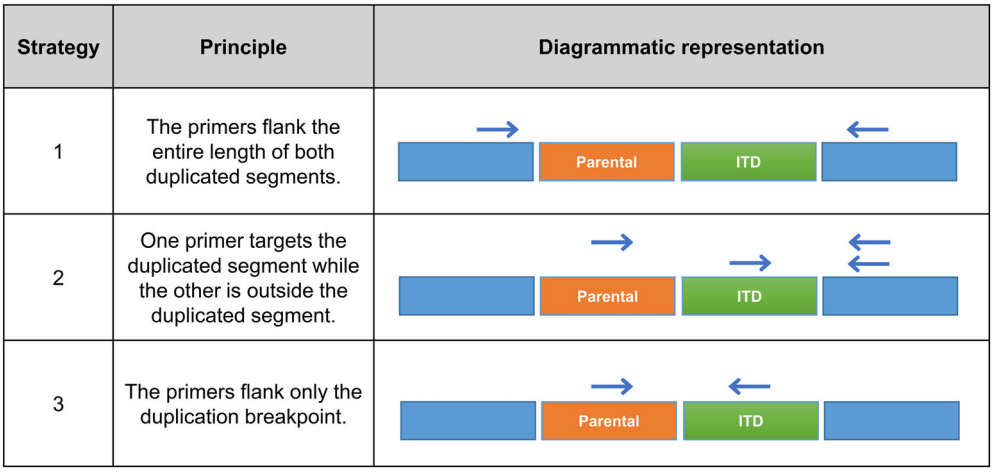


Figure 5. Primer design strategy for the detection of *BCOR* exon 15 ITD by PCR and Sanger sequencing. Three possible primer design strategies are possible (strategies 1, 2, and 3). The diagrammatic representation indicates the positions of the primers (depicted by the blue arrows) in relation to the parental and duplicated segments.

tryptophan 5-monooxygenase activation protein e) on chromosome 17 to exons 2–7 of *NUTM2* (NUT family member 2A) on chromosome 10 [18]. This fusion was found in 12% of examined cases [18]. Subsequent studies on CCSK [8,15–17] found no additional cases, making the proportion of the reported CCSK cases with this gene fusion closer to 5%. This gene fusion has not yet been identified in PMMTI [10,12,13]. Isolated cases of CCSK have been found with other gene fusions, including *BCOR-CCNB3* [8,25] and *IRX2-TERT* [26]. Other sporadic mutations in CCSK include *EGFR* gene amplification and point mutation [27]. It follows that the majority of cases of CCSK have some other genetic change, and that change was found to involve the *BCOR* (BCL-6 co-repressor) gene, which encodes a component of a variant polycomb repressive complex (PRC). The specific mutation type is an ITD involving exon 15 [8,14–17]. ITD is an uncommon structural gene mutation in which a segment of a coding region of a gene is duplicated in an end-to-end manner. The suggested mechanism is genomic instability followed by erroneous DNA repair [8]. ITDs were first identified affecting the *FLT3* gene in patients with acute myeloid leukaemia [28]. ITD involving *EGFR* has been reported in a case of CCSK [29] and is also a consistent finding in the classic type of congenital mesoblastic nephroma [30], another rare type of paediatric renal tumour.

In CCSK, different sequences of ITDs have been found, with overlapping breakpoints, ranging in size from 87 to 114 bp, and sometimes with 1–11 bp inserted between the duplicated sequences [8,14–17]. The sequence alterations are always in frame, adding 30–38 amino acids to the C terminus of the BCOR protein. This region of the protein involves the PCGF (polycomb group RING finger) Ub-like fold discriminator (PUFD) domain that is important for the functioning of the BCOR complex as an epigenetic modifier. It is postulated that the ITDs induce aberrant methylation and modification in gene expression [15,17]. In ITD-positive tumours, there is increased expression of mutant *BCOR* transcripts and increased expression of PRC2-regulated genes [16]. As the *BCOR* gene is X-linked, males with CCSK express only the mutant gene, while female patients express both mutant and wild type, explaining the male predominance seen in CCSK [16,17]. *BCOR* ITDs are not seen in other childhood renal tumours, such as Wilms tumour and congenital mesoblastic nephroma [15–17]. More recently, however, this same genetic change has been found in PMMTI [10,12–14], providing a pathogenetic link between CCSK and PMMTI, and giving rise to the concept that PMMTI is the soft tissue counterpart of CCSK. *BCOR* ITDs have also been reported in high-grade endometrial stromal sarcoma of the adult uterus [31] and one type of embryonal tumour of the central nervous system, previously referred to as 'primitive neuroectodermal tumour' but now renamed as 'central nervous system embryonal tumour with *BCOR* internal tandem duplication' [32,33]. The novel splicing phenomenon occurring within the context of an unusually long ITD in our cases of CCSK and PMMTI has not been described to date in these other tumours.

In CCSK, the proportion of cases with *BCOR* ITDs is in the range of 85% [8,16], and cases with *YWHAE-NUTM2* or *BCOR-CCNB3* fusion likely account for another 5%. These genetic changes are mutually exclusive [19], so that leaves ∼10% of cases with no genetic abnormality detected by the technology used in these studies. The 10% estimate is comparable to the 12% of cases determined in one study to possess neither a *YWHAE–NUTM2* fusion nor *BCOR* ITD [19]. There is no correlation of genetic abnormality with phenotype, so the mutation cannot be presumed for any case that falls into this 10% group. In our group of 32 tumours, we had 3 cases without a molecular diagnosis. We had first used our standard approach for detection of *BCOR* ITDs, which is DNA-based PCR followed by Sanger sequencing, using two sets of primer pairs that can identify the full range of *BCOR* exon 15 ITDs described to date [8]. This approach failed to detect *BCOR* exon 15 ITDs in our three cases and raised the question whether there might be a yet unrecognised gene involved in CCSK, or whether one of the known genetic alterations was involved, but in a novel way, such that standard approaches would not identify it. We hypothesised that there might exist an ITD that is too large to be detected by this approach. ITDs are challenging to identify as they commonly involve an intermediate nucleic acid segment length that is, on the one hand, too large for sequencing-based techniques and, on the other hand, too small for array-based or cytogenetic techniques. Identification of ITDs with traditional NGS shotgun sequencing methods requires specific optimised bioinformatic approaches to reliably identify ITDs with robust allele frequency estimation [34]. Despite the introduction of long-read sequencing, conventional tumour tissue samples are usually FFPE, which results in fragmentation of the DNA and RNA extracted from such samples, limiting the utility of this technology as a diagnostic tool.

We used a customised anchored multiplex PCR targeted NGS assay (Archer® FusionPlex) which utilised RNA extracted from FFPE tumour sections as

starting material. The logic for this approach came from an *in silico* analysis examining, first, the capability of published primer sets to detect an ITD depending on whether DNA or RNA was used as the starting nucleic acid material (Table 1). Second, we assessed whether the shorter DNA and RNA fragment lengths characteristic of FFPE samples would be suitable as the starting material. We determined that several of the primer designs would be able to detect a novel ITD using RNA but not using genomic DNA, should the ITD be considerably larger than those previously described (87–114 bp). Furthermore, nucleic acid fragments from FFPE samples will rarely be of sufficient length for ITD sequences longer than 500 bp to be detected. Therefore, detection of any novel ITD would require either RNA-based techniques such as the Archer FusionPlex anchored multiplex PCR method or primer designs that flank the specific breakpoint regions. On this basis, we elected to screen our three cases with Archer FusionPlex anchored multiplex PCR that included primers to exon 15 of *BCOR*.

RNA from any positive cases would then be subjected to RT-PCR using primers flanking the breakpoint sequences, followed by Sanger sequencing. Primer design for this purpose involves several considerations (Figure 5). The first primer design strategy flanks the entire length of both duplicated segments and includes the start and end sites of both segments with the ITD breakpoint in the middle. However, this strategy requires amplification of at least a 250 bp segment, a length that may be challenging for FFPE samples, which tend to be degraded. The second primer design strategy provides a compromise, allowing amplification of a smaller fragment, but still requiring either the forward or reverse primer to be outside the duplicated region. The third primer design strategy only amplifies the duplication breakpoint and can utilise both DNA and RNA extracted from FFPE samples. However, a knowledge of the breakpoint sequence would be necessary to identify the precise ITD. Thus, for the confirmation of *BCOR* ITDs, we elected to use two sets of primers following strategy 2, using RNA extracted from the three FFPE tumour samples. PCR products can then be gel purified and Sanger sequenced to detect the exact ITD duplication region.

This approach proved to be successful. We were able to identify a novel *BCOR* exon 15 ITD that is 388 bp in length and that spans a stop codon. Following transcription to RNA, this pre-mRNA undergoes splicing to produce a shorter 96 bp-long mature mRNA that harbours a modified ITD. This particular sequence has not been previously reported in previous studies [8,14–17]. The fact that this ITD was found in three of three cases tested suggests it may be a recurrent mutation in CCSK and PMMTI, thereby possibly accounting for a significant proportion of the missing 10% of cases in which mutations are not identified by standard testing techniques. If so, closer to 95% of CCSKs would have *BCOR* ITDs. Additional cases without identified mutations will need to be studied by the RNA-based approach used in this study, in order to confirm this finding. The question then arises why some studies have found 100% of cases with *BCOR* ITDs [12–14]. These studies included 12, 13, and 5 cases, respectively; hence it is possible that, with these numbers, such studies did not include cases from this 10% group, just on a sampling basis. Only one study used targeted RNA sequencing with custom Archer FusionPlex kit [14], while another study used whole transcriptome sequencing [35]. This approach theoretically would have detected the mutations seen in our three cases, if such cases had been included in these two series.

The identification and description of this novel *BCOR* ITD is of clinical significance. The diagnosis of CCSKs or PMMTIs is not always straightforward on morphological or immunophenotypic features, and the identification of a *BCOR* ITD in the setting of an atypical tumour will be important to support a diagnosis on which subsequent patient management and therapeutic decisions rest. The novel *BCOR* ITD sequence in our study eluded detection by DNA PCR with Sanger sequencing. In cases where molecular testing is pursued, it is important to also consider using RNA as the starting material since we now know of the existence of longer segment ITDs that will not be identified by standard DNA testing methods. Equally important is to ensure that primer designs capture the full range of *BCOR* ITDs including the novel ITD described herein. Given the possibility of unexpectedly long ITD segments as described, in the situation when targeted PCR is employed, an alternative primer design strategy targeting the duplication breakpoints rather than entire lengths of the duplicated segments and using RNA rather than DNA as starting material for testing will be required to detect such ITDs. We provide details of a suggested RT-PCR protocol in the Supplementary materials and methods.

Acknowledgements

This study was funded by the Viva-KKH Paediatric Brain and Solid Tumour Programme based at KK

Women's and Children's Hospital, Singapore, supported by VIVA Foundation for Children with Cancer.

Author contributions statement

KTEC conceived and supervised all aspects of the project. JYG and CHK performed the experiments and analysed the data. MS, MZ, HT, SJA, SG, FZ and LC analysed and curated the data. BTT and PST analysed the data and supervised the experiments and project. JYG, PST and KTEC wrote the paper with input, review and approval of all authors.

References

1. Gooskens SL, Furtwängler R, Vujanic GM, *et al*. Clear cell sarcoma of the kidney: a review. *Eur J Cancer* 2012; **48**: 2219–2226.
2. Furtwängler R, Gooskens SL, van Tinteren H, *et al*. Clear cell sarcomas of the kidney registered on International Society of Pediatric Oncology (SIOP) 93-01 and SIOP 2001 protocols: a report of the SIOP Renal Tumour Study Group. *Eur J Cancer* 2013; **49**: 3497–3506.
3. Gooskens SL, Furtwängler R, Spreafico F, *et al*. Treatment and outcome of patients with relapsed clear cell sarcoma of the kidney: a combined SIOP and AIEOP study. *Br J Cancer* 2014; **111**: 227–233.
4. Argani P, Perlman EJ, Breslow NE, *et al*. Clear cell sarcoma of the kidney: a review of 351 cases from the National Wilms Tumor Study Group Pathology Center. *Am J Surg Pathol* 2000; **24**: 4–18.
5. Aw SJ, Chang KTE. Clear cell sarcoma of the kidney. *Arch Pathol Lab Med* 2019; **143**: 1022–1026.
6. Jet Aw S, Hong Kuick C, Hwee Yong M, *et al*. Novel karyotypes and cyclin D1 immunoreactivity in clear cell sarcoma of the kidney. *Pediatr Dev Pathol* 2015; **18**: 297–304.
7. Mirkovic J, Calicchio M, Fletcher CD, *et al*. Diffuse and strong cyclin D1 immunoreactivity in clear cell sarcoma of the kidney. *Histopathology* 2015; **67**: 306–312.
8. Wong MK, Ng CCY, Kuick CH, *et al*. Clear cell sarcomas of the kidney are characterised by BCOR gene abnormalities, including exon 15 internal tandem duplications and BCOR-CCNB3 gene fusion. *Histopathology* 2018; **72**: 320–329.
9. Karlsson J, Valind A, Jansson C, *et al*. Aberrant epigenetic regulation in clear cell sarcoma of the kidney featuring distinct DNA hypermethylation and EZH2 overexpression. *Oncotarget* 2016; **7**: 11127–11136.
10. Kao YC, Sung YS, Zhang L, *et al*. Recurrent BCOR internal tandem duplication and YWHAE-NUTM2B fusions in soft tissue undifferentiated round cell sarcoma of infancy: overlapping genetic features with clear cell sarcoma of kidney. *Am J Surg Pathol* 2016; **40**: 1009–1020.
11. Alaggio R, Ninfo V, Rosolen A, *et al*. Primitive myxoid mesenchymal tumor of infancy: a clinicopathologic report of 6 cases. *Am J Surg Pathol* 2006; **30**: 388–394.
12. Antonescu CR, Kao YC, Xu B, *et al*. Undifferentiated round cell sarcoma with BCOR internal tandem duplications (ITD) or YWHAE fusions: a clinicopathologic and molecular study. *Mod Pathol* 2020; **33**: 1669–1677.
13. Santiago T, Clay MR, Allen SJ, *et al*. Recurrent BCOR internal tandem duplication and BCOR or BCL6 expression distinguish primitive myxoid mesenchymal tumor of infancy from congenital infantile fibrosarcoma. *Mod Pathol* 2017; **30**: 884–891.
14. Al-Ibraheemi A, Putra J, Tsai HK, *et al*. Assessment of BCOR internal tandem duplications in pediatric cancers by targeted RNA sequencing. *J Mol Diagn* 2021; **23**: 1269–1278.
15. Astolfi A, Melchionda F, Perotti D, *et al*. Whole transcriptome sequencing identifies BCOR internal tandem duplication as a common feature of clear cell sarcoma of the kidney. *Oncotarget* 2015; **6**: 40934–40939.
16. Roy A, Kumar V, Zorman B, *et al*. Recurrent internal tandem duplications of BCOR in clear cell sarcoma of the kidney. *Nat Commun* 2015; **6**: 8891.
17. Ueno-Yokohata H, Okita H, Nakasato K, *et al*. Consistent in-frame internal tandem duplications of BCOR characterize clear cell sarcoma of the kidney. *Nat Genet* 2015; **47**: 861–863.
18. O'Meara E, Stack D, Lee CH, *et al*. Characterization of the chromosomal translocation t(10;17)(q22;p13) in clear cell sarcoma of kidney. *J Pathol* 2012; **227**: 72–80.
19. Kenny C, Bausenwein S, Lazaro A, *et al*. Mutually exclusive BCOR internal tandem duplications and YWHAE-NUTM2 fusions in clear cell sarcoma of kidney: not the full story. *J Pathol* 2016; **238**: 617–620.
20. Kao YC, Sung YS, Zhang L, *et al*. BCOR overexpression is a highly sensitive marker in round cell sarcomas with BCOR genetic abnormalities. *Am J Surg Pathol* 2016; **40**: 1670–1678.
21. Chang KTE, Goytain A, Tucker T, *et al*. Development and evaluation of a pan-sarcoma fusion gene detection assay using the NanoString nCounter platform. *J Mol Diagn* 2018; **20**: 63–77.
22. Reese MG, Eeckman FH, Kulp D, *et al*. Improved splice site detection in Genie. *J Comput Biol* 1997; **4**: 311–323.
23. Smith PJ, Zhang C, Wang J, *et al*. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum Mol Genet* 2006; **15**: 2490–2508.
24. Cartegni L, Wang J, Zhu Z, *et al*. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 2003; **31**: 3568–3571.
25. Argani P, Kao YC, Zhang L, *et al*. Primary renal sarcomas with BCOR-CCNB3 gene fusion: a report of 2 cases showing histologic overlap with clear cell sarcoma of kidney, suggesting further link between BCOR-related sarcomas of the kidney and soft tissues. *Am J Surg Pathol* 2017; **41**: 1702–1712.
26. Karlsson J, Lilljebjörn H, Holmquist Mengelbier L, *et al*. Activation of human telomerase reverse transcriptase through gene fusion in clear cell sarcoma of the kidney. *Cancer Lett* 2015; **357**: 498–501.
27. Little SE, Bax DA, Rodriguez-Pinilla M, *et al*. Multifaceted dysregulation of the epidermal growth factor receptor pathway in clear cell sarcoma of the kidney. *Clin Cancer Res* 2007; **13**: 4360–4364.

28. Nakao M, Yokota S, Iwai T, *et al*. Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. *Leukemia* 1996; **10**: 1911–1918.
29. Santiago T, Clay MR, Azzato E, *et al*. Clear cell sarcoma of kidney involving a horseshoe kidney and harboring EGFR internal tandem duplication. *Pediatr Blood Cancer* 2017; **64**: e26602.
30. Zhao M, Yin M, Kuick CH, *et al*. Congenital mesoblastic nephroma is characterised by kinase mutations including EGFR internal tandem duplications, the ETV6-NTRK3 fusion, and the rare KLHL7-BRAF fusion. *Histopathology* 2020; **77**: 611–621.
31. Juckett LT, Lin DI, Madison R, *et al*. A pan-cancer landscape analysis reveals a subset of endometrial stromal and pediatric tumors defined by internal tandem duplications of BCOR. *Oncology* 2019; **96**: 101–109.
32. Sturm D, Orr BA, Toprak UH, *et al*. New brain tumor entities emerge from molecular classification of CNS-PNETs. *Cell* 2016; **164**: 1060–1072.
33. Ferris SP, Velazquez Vega J, Aboian M, *et al*. High-grade neuroepithelial tumor with BCOR exon 15 internal tandem duplication – a comprehensive clinical, radiographic, pathologic, and genomic analysis. *Brain Pathol* 2020; **30**: 46–62.
34. Wang TY, Yang R. Scan ITD: detecting internal tandem duplication with robust variant allele frequency estimation. *Gigascience* 2020; **9**: giaa089.
35. Bouchoucha Y, Tauziède-Espariat A, Gauthier A, *et al*. Intra- and extra-cranial BCOR-ITD tumours are separate entities within the BCOR-rearranged family. *J Pathol Clin Res* 2022; **8**: 217–232.
36. Mariño-Enriquez A, Lauria A, Przybyl J, *et al*. BCOR internal tandem duplication in high-grade uterine sarcomas. *Am J Surg Pathol* 2018; **42**: 335–341.
37. Yoshida Y, Nobusawa S, Nakata S, *et al*. CNS high-grade neuroepithelial tumor with BCOR internal tandem duplication: a comparison with its counterparts in the kidney and soft tissue. *Brain Pathol* 2018; **28**: 710–720.
38. Chiang S, Lee CH, Stewart CJR, *et al*. BCOR is a robust diagnostic immunohistochemical marker of genetically diverse high-grade endometrial stromal sarcoma, including tumors exhibiting variant morphology. *Mod Pathol* 2017; **30**: 1251–1261.
39. Paret C, Theruvath J, Russo A, *et al*. Activation of the basal cell carcinoma pathway in a patient with CNS HGNET-BCOR diagnosis: consequences for personalized targeted therapy. *Oncotarget* 2016; **7**: 83378–83391.
40. Appay R, Macagno N, Padovani L, *et al*. HGNET-BCOR tumors of the cerebellum: clinicopathologic and molecular characterization of 3 cases. *Am J Surg Pathol* 2017; **41**: 1254–1260.

SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods

Suggested protocol for *BCOR* ITD detection from FFPE tissue