Intra- and extra-cranial *BCOR*-ITD tumours are separate entities within the *BCOR*-rearranged family

Y Bouchoucha, A Tauziède-Espariat, A Gauthier *et al. J Pathol Clin Res* DOI: 10.1002/cjp2.255

Supplementary material

Reference numbers refer to the list in the main paper. Supplementary tables are presented as separate files.

Supplementary materials and methods

Text S1. Detailed imaging characteristics of *BCOR*-ITD tumours

Text S2. Detailed histopathological and immunohistochemical analyses

Figure S1. Structure of the *BCOR*-ITD cohort**.**

Figure S2. Examples of typical imaging findings in *BCOR*-ITD sarcomas, at diagnosis

Figure S3. Boxplots showing the mRNA levels of 5 markers in the *BCOR*-ITD family in comparison with other families of sarcomas and brain tumours

Figure S4. Molecular divergence between CNS *BCOR*-ITD and *BCOR*-ITD sarcomas

Figure S5. Dot plot summarising loose (FC>1.5) and stringent (FC>2.5) CNS *BCOR*-ITD and *BCOR*-ITD sarcoma signature distributions over cell types independently analysed by sc/sn RNA-Seq

Table S1. Tumour features at diagnosis, patient by patient.

Table S2. Treatment and prognosis, patient by patient

Table S3. Radiological features of *BCOR*-ITD tumours

Table S4. Immunohistochemistry results

Table S5. Analysis of the differentially expressed genes (DEG) between CNS *BCOR*-ITD and *BCOR*-ITD sarcomas.

Table S6. Analysis of the differentially expressed genes (DEG) between *BCOR*-ITD sarcomas and CNS *BCOR*-ITD

Table S7. List of the top 10 cell types where the gene signatures of CNS *BCOR*-ITD and *BCOR*-ITD sarcomas are enriched, based on two single-cell/single-nuclei RNA-Seq atlases

Supplementary Materials and Methods

Primary antibodies and dilutions

BCOR (1:100, clone C-10, Santa Cruz Biotechnology, Dallas, USA), CD56 (pre-diluted, clone 123C3, Dako, Glostrup, Denmark), Glial Fibrillary Acidic Protein (GFAP) (1:200, clone 6F2, Dako, Glostrup, Denmark), Olig2 (1:500, clone OLIG2, Sigma-Aldrich, Saint-Louis, USA), vimentin (1:800, clone V9, Dako, Glostrup, Denmark), neurofilaments (1:100, clone NF70, Dako, Glostrup, Denmark), synaptophysin (1:150, clone DAK-SYNAP, Dako, Glostrup, Denmark), chromogranin A (1:200, clone LK2H10, Diagnostic BioSystems, Pleasanton, USA), NeuN (1:1000, clone A60, Sigma-Aldrich, Saint-Louis, USA), EMA (1:200, clone GM008, Dako, Glostrup, Denmark), CKAE1AE3 (1:800, clone AE1AE3, Dako, Glostrup, Denmark), CK18 (1:200, clone 6F2, Dako, Glostrup, Denmark), smooth muscle actin (1:4000, clone F5D, Dako, Glostrup, Denmark), desmin (1:200, clone D33, Dako, Glostrup, Denmark), myogenin (1:1000, clone 1A4, Dako, Glostrup, Denmark), CD34 (1:40, clone Qbend10, Dako, Glostrup, Denmark), INI1 (BAF47) (1:50, clone 25/BAF 47, BD Biosciences, Franklin Lakes, USA), EGFR (pre-diluted, clone 3C6, Ventana, Burgess Hill, United Kingdom), YAP1 (1:300, clone YAP (63.7), Santa Cruz Biotechnology, Dallas, USA), GAB1 (1:200, clone EPR375, Abcam, Cambridge, United Kingdom), β-catenin (pure, clone 14, Ventana, Burgess Hill, United Kingdom), p53 (1:5000, clone DO-1, Santa Cruz Biotechnology, Dallas, USA), H3K27me3 (1:2500, polyclonal, Diagenode, Liege, Belgium), PTEN (1:150, clone 6H2.1, Dako, Glostrup, Denmark), SATB2 (1:100, clone SATBA4B10, Zytomed, Berlin, Germany), oestrogen receptors (1:50, clone SP1, Histopathology Ltd., Pecs, Hungary), cyclin-D1 (1:150, clone SP4, Thermo Fischer Scientific, Saclay, France), TLE1 (1:200, clone 1F5, Bio SB, Santa Barbara, USA), NTRK (1:40, clone A7H6R, Cell signaling technology, Danvers, USA), BCL2 (1:200, clone 124, Dako, Glostrup, Denmark), cyclin-B3 (1:400, clone CCNB3, Sigma-Aldrich, Saint-Louis, USA), CD10 (1:40, clone 56C6, Novocastra, Nanterre, France), h-caldesmon (prediluted, clone h-CD, Dako, Glostrup, Denmark), BCL6 (1:30, clone PG-B6p, Dako, Glostrup, Denmark).

Paired-end RNA sequencing

Frozen tissue - Total RNAs were isolated from crushed frozen tumours using a Trizol/Chloroform reagent kit (Life Technologies, Carlsbad, CA, USA). Quantity and quality of RNA were evaluated using NanoDrop (Thermo Fisher Scientific) and Tape Station with Hs RNA Screen Tape (Agilent) with threshold values of DV200 above 50% and RIN >4. Library constructions were performed following the TruSeq Stranded mRNA LS protocol (Illumina, San Diego, CA, USA) from 1 µg of total RNA. Sequencing was performed on either HiSeq 2500 (100 nt paired-end) or NextSeq 500 (150nt paired-end) Illumina platform.

FFPE samples - Total RNA was extracted from macrodissected formalin-fixed paraffin-embedded tumour sections using the FormaPure RNA kit (Beckman Coulter #C19158, Brea, CA, USA). RNase-free DNase set (Qiagen #AM2222, Courtaboeuf, France) was used to remove DNA and RNA quantification was assessed using a NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA). One-hundred nanograms of total RNA were used to prepare libraries with TruSeq RNA UD indexes (Illumina #20022371, San Diego, USA) and TruSeq RNA Exome (Illumina #20020183, San Diego, USA). Twenty-four libraries were pooled at a final concentration of 1.5 nM together with 1% PhiX. Sequencing was performed (paired end, 2 × 75 cycles) with NovaSeq 6000 SP reagent kit (Illumina #20027465) on a NovaSeq 6000 sequencing system (Illumina #20012850).

Identification of the *BCOR*-ITD sequence on FFPE samples

FFPE samples - Gene expression values were extracted using Kallisto version 0.42.5 tool with GENCODE release 23-genome annotation based on GRCh38 genome reference. Kallisto TPM expression values were transformed in $\log_2$(TPM+2), and all samples were normalised together using the quantile method from the R limma package within R (version 3.1.2) environment. Clustering was performed with the R package Cluster version 2.0.3 using the Pearson correlation distance and the Ward clustering method. The fusion transcripts were called with STARFusion, FusionMap, FusionCatcher, ERICSCRIPT, and TopHat-fusion and validated if present in the fusion list of at least 2

algorithms. If clustering analyses demonstrated a sample within the *BCOR*-rearranged group and no fusion could be identified involving *BCOR* then *BCOR*-ITD sequence was searched for using a dedicated home-made perl script. All *BCOR*-ITD were ultimately recently verified using the Arriba fusion tool (https://genome.cshlp.org/content/31/3/448).

Clustering of RNA-seq and DNA methylation data

RNA-seq was aligned using STAR (v2.7.0e) and gene count matrices were generated with GeneCounts. Raw counts were normalised to transcripts per million (TPM). Using the R package Seurat (v3.1.5) the 5000 most variable features were selected for downstream analysis after a variance-stabilising transformation.

For computational power purposes, the DNA-methylation control samples were stratified by histopathological diagnosis and randomly selected (7 samples each, except for specific tumour types: CCSK (11), HGNET-*BCOR* (23), HGNET-MN1 (21), SBRCT-*BCOR* (8), HG-ESS (16), and MRT (17). A normalisation by quantile was applied (in R, using the *preprocessQuantile* function of *minfi* version 1.30.0). We then dropped out low quality samples (mean detection p.value > 0.005), failed probes (one detection p.value > 0.01), X and Y chromosomes, non-specific probes (according to [43]), and probes containing SNPs (*dropLociWithSnps* function). These steps selected 366432 probes out of 452453.

After data scaling, two segregation methods were performed. The first method was a hierarchical clustering based on Euclidean distances with complete linkage, plotted as dendrograms. The second method combines a selection of the most variable gene expressions (or methylation β-values), a principal component analysis (PCA) and the generation of a Uniform Manifold Approximation and Projection (UMAP) 2-dimensional plot using 50 principal components.

List of gene sets of the MSigDB used for GSEA analyses

REACTOME Signalling by Wnt in cancer, REACTOME Signalling by Wnt, KEGG Wnt signalling pathway, BIOCARTA Shh pathway, KEGG Hedgehog signalling pathway, REACTOME signalling by Hedgehog,

BIOCARTA Akt pathway, REACTOME Signalling by NTRKs, REACTOME Signalling by EGFR in cancer, REACTOME signalling by EGFR.

Text S1. Detailed Imaging characteristics of *BCOR*-ITD tumours

All CCSK tumours were relatively large and well-defined masses with a median diameter of 12 cm (range: 4.2-14.6 cm) and a median volume of 720 mL (range: 33-918 mL). They are markedly heterogeneous masses with peripheral solid areas and central areas of necrosis, cysts and haemorrhages. All tumours had a heterogenous sonographic appearance. On non-contrast CT, the lesions were hypo- to isodense to the adjacent normal renal parenchyma. Arterial and portal phase of enhancement showed heterogenous lesions, hypointense as compared to the normal renal cortex. The MR images demonstrated a low signal intensity on T1-weighted images and a high signal intensity on T2-weighted images with diffusion restriction and moderate contrast enhancement. In 2 of 5 tumours, extra-capsular extension was noted. No local lymph node involvement was noted at diagnosis.

The two cases of endometrial sarcomas presented as well-defined heterogenous masses with areas of haemorrhage. The largest tumour displayed also areas of necrosis. By MRI, the lesions demonstrated a low signal intensity on T1-weighted images, a high signal intensity on T2-weighted images with diffusion restriction and moderate to intense contrast enhancement. Regional lymph node involvement and distant metastasis were noted for the largest tumour.

The imaging characteristics of the USTS are non-specific and depend on their anatomical location. Mostly they present as well-defined masses with a heterogeneous appearance.

HGNET-BCOR presented as large intra-axial masses with well-defined borders, restricted diffusion, weak contrast enhancement, frequent central necrosis, haemorrhage and calcifications, intratumoural veins and no leptomeningeal dissemination at time of diagnosis. Detailed description of imaging characteristics with histopathologic correlation of the 10 CNS tumours with BCOR ITD is available in an independent report [44].

Text S2. Detailed Histopathological and immunohistochemical analyses

The histopathological findings were available for 21 cases (9 HGNET-BCOR, 4 CCSK, 2 ESS, and 6 URCS). HGNET-BCOR were all well circumscribed from the brain parenchyma. The most predominant pattern was EPN-like consisting in tumours composed of perivascular pseudorosettes (6/9 cases: P11, 20, 21, 22, 30, and 31)(Figure 2A). The other cases exhibit a compact fascicular growth pattern composed of spindle shaped cells with few perivascular pseudorosettes (3/9 cases: P13, 32, and 33). A myxoid or microcystic background was observed in 7/9 cases (P11, 13, 20, 21, 30, 31, and 33) (Figure 2B), without fibrous stroma. Haemorrhagic modifications were present in 6/9 cases (P11, 13, 20, 21, 30, and 33). Only two tumours presented calcifications (P11 and 33). Mitotic counts ranged from 5 to 62 per 10 high-power fields. Palisading, geographic and calcified necrosis was observed in 6/9 cases (P11, 13, 21, 22, 31, and 33) (Figure 2C). Microvascular proliferation was present only in one case (P13).

URCS were densely cellular, demonstrating a fascicular or diffuse pattern in all cases (Figure 2D). They were composed of monotonous round to ovoid cells (Figure 2E) or spindle cells with some epithelioid component (P7 and 9) or pseudorosettes (P4) (Figure 2F). They did not present microcystic or myxoid modifications, but haemorrhage and a fibrous stroma were observed in respectively 4/6 cases (P1, 6, 7, and 8), and 3/6 cases (P1, 6, and 8). No calcification was present. Mitotic counts ranged from 5 to 41 per 10 high-power fields (1.7-13.7/mm²). Necrosis was present in 3 cases (P4, 8, and 9). Of note, we do not report cases with significant myxoid background, excluding the proposed diagnosis of Primitive Mesenchymal Myxoid Tumour of the Infancy (PMMTI). Accordingly, only one case of URCS expresses BcL6, unlike most reported PMMTI [15].

HG-ESS were highly cellular with a diffuse pattern composed of spindle cells in the 2 cases. Microcystic and haemorrhagic changes were observed in both cases (Figure 2G). One case (P19) presented pseudorosettes. No fibrous stroma nor calcifications were observed. Mitotic counts ranged from 14 to 21 per 10 high-power fields (4.7-7/mm²) and no necrosis was present in both tumours.

CCSK presented as diffuse proliferations (3/4 cases) or with alveolar pattern (P25). No pseudorosette was observed. Microcystic and fibrous changes were present in only one case (respectively P24 and

P25) and no haemorrhage or calcification were noticed. Mitotic counts ranged from 0 to 18 per 10 high-power fields (0-6/mm²). Necrosis was observed in 2/4 cases (P25 and 26).

The immunohistochemical results are summarised in Table 2.

All HGNET-BCOR cases exhibit a diffuse expression of vimentin and CD56. GFAP immunoreactivity was absent in three cases (P20, 21, and 33) and focally observed in the other six cases (Figure 2H). Olig2 was expressed in 5/9 tumours with varying degrees of distribution (focal in P11, partial in P32 and 33, and diffuse in P21 and 30) and absent in four cases (P13, 20, 22, and 31) (Figure 2I). Only one case (P21) was positive for EMA with a focal cytoplasmic pattern. The staining for neurofilaments is positive in few tumour cells in 6/9 cases (P11, 13, 21, 22, 30, and 31) (Figure 2J). Among the other neuronal markers, only NeuN was expressed by tumour cells in all cases except one (P20). CD34, HMB45, LIN28A, CKAE1/AE3, smooth muscle actin, desmin, myogenin, h-caldesmon, and CD45 were never expressed by tumour cells. INI1, BRG1, and H3K27me3 were preserved in all cases. EGFR was strongly and diffusely expressed by all cases (Figure 2K). P53 accumulation was only observed in one case (P13). All cases except one (P20) presented a loss of PTEN. YAP1 and GAB1 were expressed diffusely in most cases. NTRK, CD10 and BCL6 immunopositivity were present in 8/9 cases, 6/9 cases and 5/9 cases respectively. No significant nuclear accumulation of βcatenin and oestrogen receptors were observed in any case. BCOR, SATB2, BCL2, Cyclin D1 and TLE1 were stained in all cases.

CCSK, ESS, and URCS did not express glial and neuronal markers. EGFR overexpression was only present in three cases (1 CCSK: P25, and 2 URCS: P7 and P9). PTEN expression was interpretable in only one URCS and showed a loss of expression. YAP1 was diffusely expressed in those tumours except one URCS (P1), whereas GAB1 was only stained in two cases (1 CCSK: P26, and 1 URCS: P1). No immunopositivity for oestrogen receptors, NTRK, CD34, desmin and h-caldesmon was observed in any case. CD10 was expressed by some tumours (2 CCSK: P25 and P26, 1 ESS: P19, and 2 URCS: P1 and P8). Bcl6 was expressed in only one case of ESS (P19). BCOR, SATB2, Cyclin D1, BCL2 and TLE1 were expressed in all tumours, except one for BCOR and SATB2 (one case of CCSK, P25) (Figure 2L-P).

	URCS n=10	CNS BCOR-ITD n=10	HG-ESS n=4	CCSK n=7	Bone n=2	Total n=33
Clinical	8	10	4	7	2	31
Histology	7	8	2	5	0	22
Radiology	5	10	2	5	1	23
RNA-seq	Frozen 7 FFPE 3	Frozen 8 FFPE 2	Frozen 0 FFPE 4	Frozen 6 FFPE 1	Frozen 1 FFPE 1	Frozen 22 FFPE 11
Methylome	6	9	0	6	0	21

Figure S1. Structure of the *BCOR*-ITD cohort**.** For RNAseq, the FFPE and frozen samples correspond to distinct tumours. There is no overlap between the two sample types

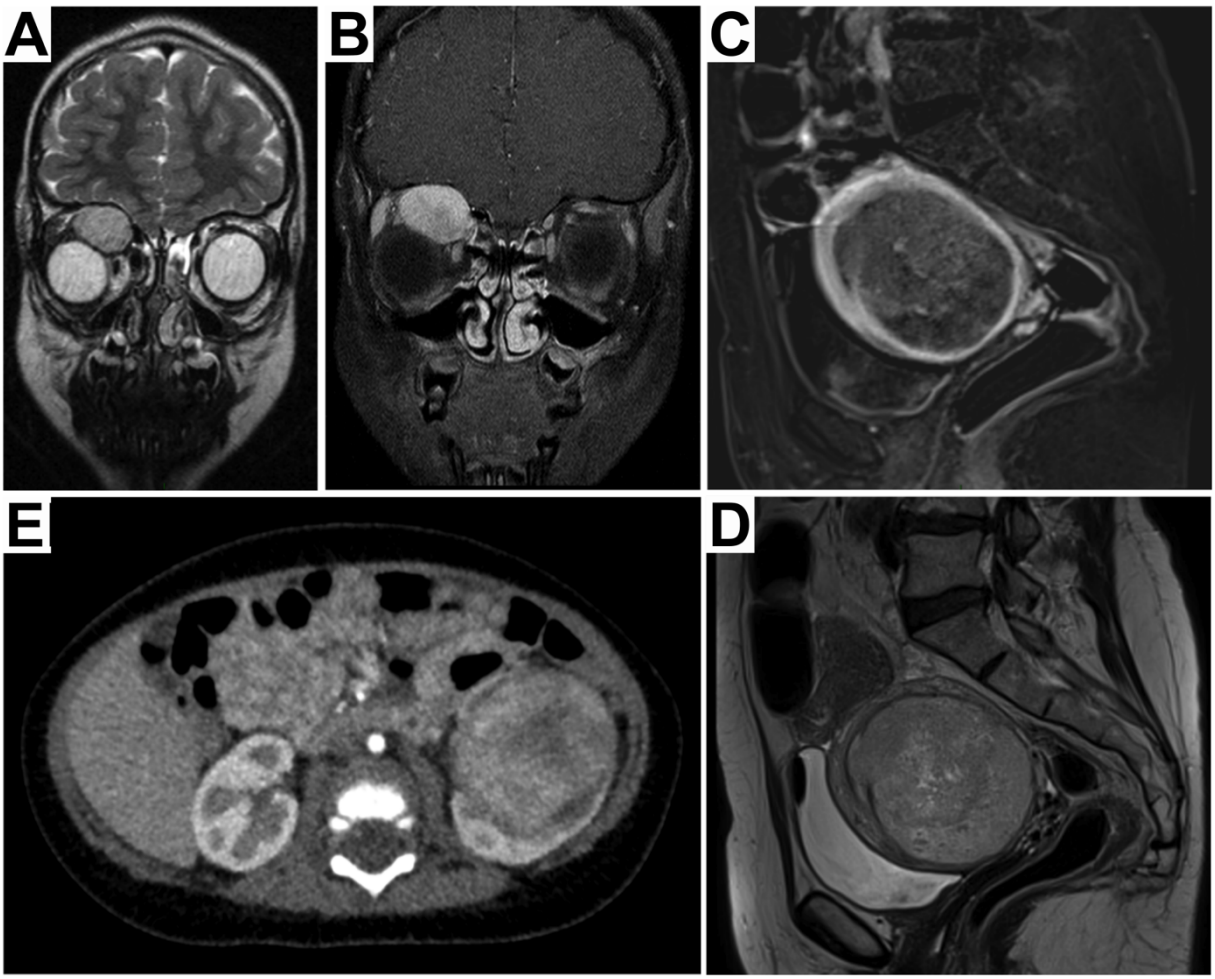

Figure S2. Examples of typical imaging findings in *BCOR*-ITD sarcomas, at diagnosis

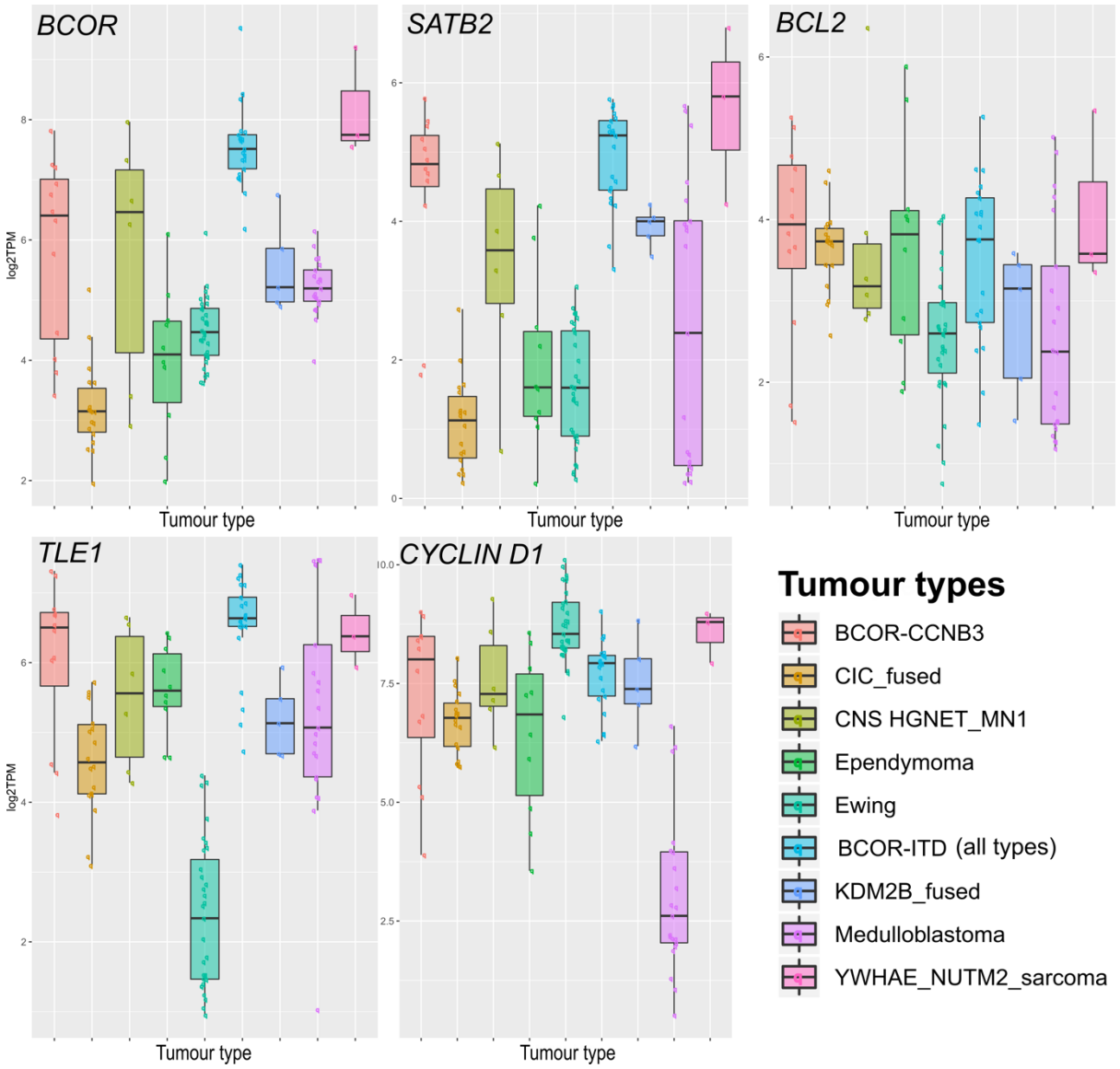

Figure S3. Boxplots showing the mRNA levels of 5 markers in the *BCOR*-ITD family in comparison with other families of sarcomas and brain tumours

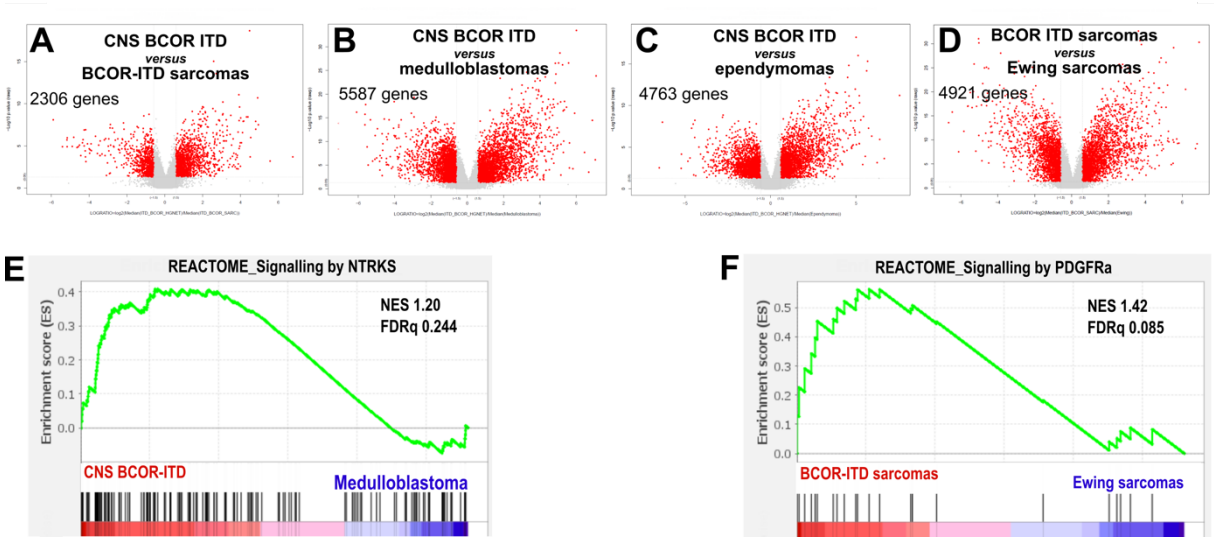

Figure S4. Molecular divergence between CNS *BCOR*-ITD and *BCOR*-ITD sarcomas

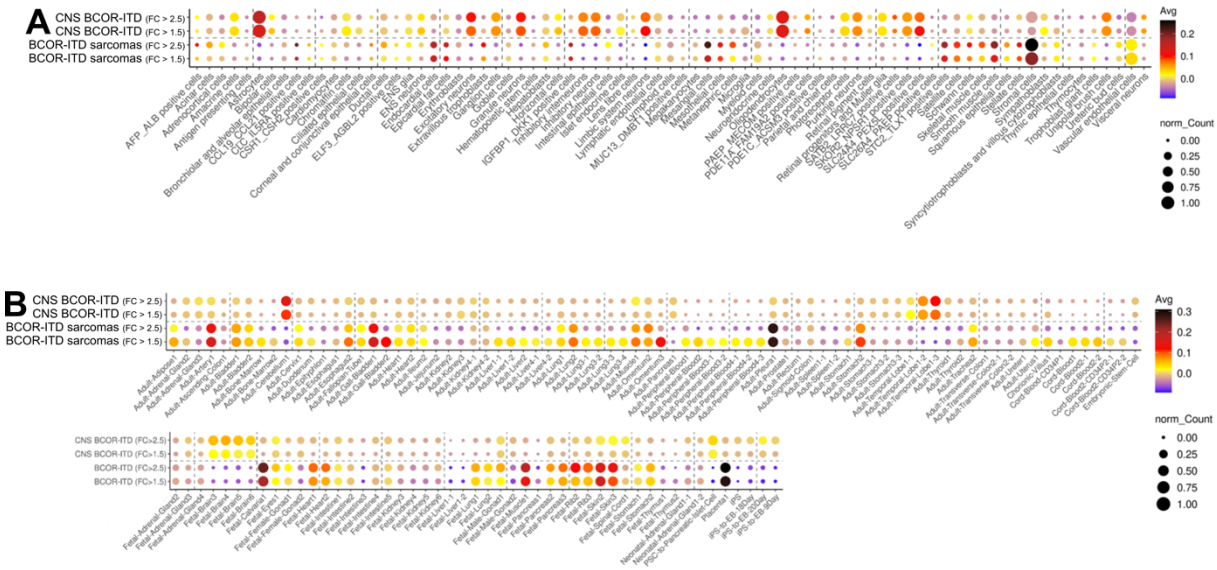

Figure S5. Dot plot summarising loose (FC>1.5) and stringent (FC>2.5) CNS *BCOR*-ITD and *BCOR*-ITD sarcoma signature distributions over cell types independently analysed by sc/sn RNA-Seq