A new subtype of bone sarcoma defined by BCOR-CCNB3 gene fusion

Gaëlle Pierron, Franck Tirode, Carlo Lucchesi, Stéphanie Reynaud, Stelly Ballet, Sarah Cohen-Gogo, Virginie Perrin, Jean-Michel Coindre, Olivier Delattre

A new subtype of bone sarcoma defined by *BCOR-CCNB3* gene fusion

Gaëlle PIERRON[1*], Franck TIRODE[2*], Carlo LUCCHESI[2*], Stéphanie REYNAUD[1], Stelly BALLET[1], Sarah COHEN-GOGO[2], Virginie PERRIN[2], Jean-Michel COINDRE[3], Olivier DELATTRE[1,2,**]

* These authors contributed equally to this work

[1] Institut Curie, Unité de Génétique Somatique – Centre Hospitalier, 26 rue d'Ulm 75248 PARIS CEDEX 05

[2] Unité 830 INSERM/Institut Curie – Centre de Recherche, 26 rue d'Ulm 75248 PARIS CEDEX 05

[3] Institut Bergonié, 229 Cours de l'Argonne 33076 BORDEAUX CEDEX

** to whom correspondence should be addressed at:

Unité 830 INSERM/Institut Curie – Centre de Recherche, 26 rue d'Ulm 75248 PARIS CEDEX 05

tel : +33 1 56 24 66 79

Fax : +33 1 56 24 66 30

email : olivier.delattre@curie.fr

Abstract

The identification of subtype-specific translocations has revolutionized diagnostics of sarcoma and provided new insight into oncogenesis. We used RNA-Seq to investigate samples diagnosed as small round cell tumors of bone, possibly Ewing sarcoma, but lacking the canonical *EWSR1-ETS* translocation. A new fusion was observed between the BCL6 co-repressor (*BCOR*) and the testis specific cyclin B3 (*CCNB3*) genes on chromosome X. RNA-Seq results were confirmed by RT-PCR and cloning the tumor-specific genomic translocation breakpoints. 24 *BCOR-CCNB3*-positive tumors were identified among a series of 594 sarcomas. Gene profiling experiments indicate that *BCOR-CCNB3*-positive cases are biologically distinct from other sarcomas, particularly Ewing's sarcoma. Finally, we show that CCNB3 immunohistochemistry is a powerful diagnostic marker for this group of sarcoma and that over-expression of *BCOR-CCNB3* or of a truncated *CCNB3* activates S-phase in NIH3T3 cells. Thus the intrachromosomal X fusion described here represents a new subtype of bone sarcoma caused by a novel gene fusion mechanism.

Text

Sarcomas, tumors derived from the mesenchymal tissues account for approximately 2% of human cancers and constitute a very heterogeneous group of tumors that can be divided into more than 100 different subtypes on the basis of clinical, pathological, immunohistological and genetic features. In adolescents and young adults, osteosarcoma and Ewing's sarcoma (ES) are the two predominant bone sarcomas. ES is genetically characterized by gene fusions between *EWSR1* and *ETS* members[1-3]. In addition, recently, *EWSR1-NFATC2* fusions were observed in a minority of cases[4]. In osteosarcoma, various point mutations and copy number variations have been described but presently, no recurrent gene fusion has been reported[5].

To identify novel fusion transcripts in small round cell sarcomas we selected four cases that clinically and morphologically would be classified in the ES family of tumors yet lacked the pathognomonic gene fusions that characterize ES. They were analysed using RNA-Seq[6-7] with paired-ends sequencing on a SOLiD platform. For each sample, a mean number of $123x10^6$ paired-end (PE) reads were mapped to the reference genome (hg19). Three analysis programs, Genomatix, FusionSeq[8,9] and an in-house software, that differed in their principles, were used to search for novel gene fusions. Strong fusion predictions were found in two cases. In one case, a *FUS-FEV* putative fusion was observed and subsequently confirmed by RT-PCR (data not shown). Although not included in the gene fusion screening panel used[10], this fusion was previously described in one case of ES[11]. In the second notable case, a total of 20

distinct and high quality PE fragments were observed linking the *BCOR*[12] exon 15 and the *CCNB3*[13] exon 5 located at chrXp11.4 and chrXp11.22, respectively (Fig. 1a and b). Analysis of exon coverage showed that all *BCOR* exons were well represented (median base coverage around 2000) with a notable decrease of the coverage at the end of the *BCOR* coding sequence (median base coverage around 300) (Fig. 1c and supplementary Fig. 1). The decreased coverage at this position was specific for this tumor since it was not observed in the alignment of the three other cases (supplementary Fig. 1). For *CCNB3*, virtually no read was observed in the 4 first exons whereas the coverage of exon 5 to 12 was comparable to *BCOR* (Fig. 1c).

To confirm these findings, specific primers designed in *BCOR* exon 15 (BCOR.1) and *CCNB3* exon 5 (CCNB3.1) were used for RT-PCR experiments. The expected fragment was detected in the tumor RNA whereas no specific amplification was observed in peripheral blood cells RNA from the same patient (Fig. 1d). Sequencing of this fragment identified an in-frame fusion between the last codon of *BCOR* and *CCNB3* exon 5 (Fig. 1e).

To further characterize the breakpoint at the genomic level, FISH experiments and long-range PCR experiments were conducted. Two BACs flanking *BCOR* were labeled with FITC and two BACs flanking *CCNB3* were labeled with rhodamine then hybrized on nuclei from either normal cells (LL) or from case 330T. As shown on Fig. 2a, the green or red doublets observed in normal cells were clearly disrupted in 330T nuclei with juxtaposition of green and red spots (usually appearing as yellow spots) therefore documenting the chromosome X inversion. Moreover, long range PCR

using the BCOR.1 and CCNB3.1 primers detected a specific PCR product that was observed in the tumor but not in the matched blood DNA therefore demonstrating that the rearrangement was somatically acquired (Fig. 2b). Sequencing of this fragment indicated that the fusion lies at chrX position 39,910,434 and 50,050,261 within the *BCOR* and *MID1IP1* genes intergenic region and within the *CCNB3* intron 4, respectively (Fig. 2c). Interestingly, the TGA stop codon of *BCOR* is included within a putative GGTGAG donor splice site sequence. Thus, the most likely hypothesis to account for this fusion mRNA is that this cryptic donor site is spliced with the acceptor site of *CCNB3* exon 5, which is confirmed by cDNA junction sequence (Fig. 1e). This splicing is incomplete leading to both *BCOR-CCNB3* and wild type *BCOR* expression (Supplementary Fig. 1 & 2).

To assess the prevalence across a large collection of sarcomas, the afore-described RT-PCR assay was used to screen a series of 594 sarcomas that lacked *EWSR1-ETS*, *EWSR1-NFATC2*, *EWSR1-WT1*, *PAX3* or *7-FOXO1,* the most frequent sarcoma fusions that are routinely investigated in our diagnostic laboratory. A total of 24 *BCOR-CCNB3* positive cases were identified (Table 1). When tumor DNAs were available, long-range PCR confirmed the presence of genomic fusions of variable sizes (Fig. 2b & d). Interestingly, all tumors but five were localized to bone and all occurred in adolescents or young adults. In most cases the pathological report was consistent with an undifferentiated, small round cell sarcoma, suggestive of the ES family of tumors. However, in contrast to ES, strong membrane positivity of CD99 was absent in approximately half of cases (Table 1).

To further investigate the relationship of the *BCOR-CCNB3*-positive group with other sarcomas, 10 cases for which good quality RNA were available were profiled for gene expression by microarray. Resulting profiles were compared with data from osteosarcoma, rhabdomyosarcoma, synovialsarcoma, ES, neuroblastoma, malignant rhabdoid tumors and desmoplastic round cell tumors. Using a variety of unsupervised multivariate statistical factorization techniques including correspondence analysis (COA) and principal component analysis (PCA), the *BCOR-CCNB3* group clusters distinctly from all other groups (Fig. 3a). Given the clinical and morphological similarities with the ES group, the expression profile of the *BCOR-CCNB3* group was specifically compared to *EWSR1-ETS*-positive ES and our findings confirmed the striking difference between the two groups with more than 3000 probesets showing highly significant differences of expression levels (Fig. 3b). The *BCOR-CCNB3*/ES comparison was used for a gene set enrichment analysis. Two of the most significantly different gene sets corresponded to Ewing specific gene sets including the Zhang_Targets_of_EWSR1_FLI1_fusion and the Riggi_Ewing_sarcoma_progenitor signatures (Fig. 3 c & d; Supplementary Fig. 3 & 4). In particular, most of the well-known EWSR1-FLI-1 target genes[3], like *NR0B1*, *CAV 1 & 2*, *PRKCB*, *NKX2-2*, *IGFBPs*, *CCK*, *MYC*, which have been shown to play critical roles in ES oncogenesis were not expressed in the *BCOR-CCNB3* tumors.

Finally, SNP6.0 arrays detected no alteration in 11 cases out of 18 studied cases. In the other cases, the only recurrent abnormalities were chromosome 17p and 10q deletions, each observed in two cases. Neither deletion nor amplification of the

fusion partners was observed. Importantly, the most frequent copy number abnormalities of ES[14], i.e. chromosome 8 gain that is observed in approximately 50 % of ES cases, chromosome gain of 1q and 12 or chromosome 16q loss that are each observed in approximately 25% of ES cases were not recurrent characteristics of *BCOR-CCNB3* positive cases. These data indicate that the classical ES and *BCOR-CCNB3* tumors do not share common copy number changes. Altogether, these data show that a new group of bone sarcomas is characterized by a recurrent *BCOR-CCNB3* fusion as a result of a chromosome X paracentric inversion. Despite remarkable clinical and pathological similarities with the group of ES, gene profiling and SNP-array experiments indicate that this group of tumors is biologically distinct from ES and, in particular, does not share the *EWSR1-ETS* signature of ES. We propose to identify this new group as a distinct type of Ewing-like tumors. To further investigate key oncogenic pathways that may characterize the *BCOR-CCNB3*-positive cases, a differential analysis was conducted between the *BCOR-CCNB3*-positive group and other tumor types (see Fig. 5; to avoid over-representation of any tumor type in the analysis the comparison was made using 10 cases randomly selected among each of the 11 tumor types)(supplementary Tables 1-3). Gene set enrichment analyses (GSEA) as well as DAVID pathway analyses indicated that morphogenesis and development, in particular of the skeletal system, showed up very strongly due to the over-expression of numerous homeobox genes. Even more importantly, activation of the WNT and SHH signaling pathways also appeared highly significant in this analysis (Fig. 4a, supplementary Tables 4-6).

It is notable that the newly characterized rearrangement includes two important genes on the X chromosome. *BCOR* encodes a ubiquitously expressed transcriptional repressor that associates with the BCL6 oncoprotein and with a variety of histone modifying enzymes suggesting that it acts as suppressor of gene expression through epigenetic mechanisms[12,15]. Of particular interest regarding its involvement in bone tumors is the observation that *BCOR* regulates mesenchymal stem cell function by epigenetic mechanisms[16]. *BCOR* has been associated with human diseases. Indeed, mutations of *BCOR* are responsible for the inherited oculofaciocardiodental and Lenz microphtalmia syndromes[17]. Moreover, *BCOR* has recently been described as a fusion partner of *RARalpha* in acute promyelocytic leukemia and as a putative tumor suppressor gene in a subset of acute myeloid leukemia[18,19].

CCNB3 exhibits a pattern of expression highly restricted to the testis and is an early meiotic cyclin that is expressed in leptotene and zygotene phases of spermatogenesis[13,20]. This strongly suggests that ectopic expression of *CCNB3* as a result of the *BCOR* gene fusion may constitute a key oncogenic event in this newly characterized type of Ewing-like tumors. This was further investigated by BrdU staining after over-expressing *BCOR-CCNB3* or *deltaCCNB3*, a truncated version of *CCNB3* corresponding to the portion involved in the fusion gene, in mouse NIH3T3 fibroblasts. As compared to the empty vector, both expression vectors led to a two-fold increase of the amount of cells in S-phase (Fig. 4b and supplementary Fig. 6). Read counts (Figure 1c) and microarray experiments that show similar intensities for

the expression of 5' *BCOR* and 3'*CCNB3* moieties (Fig. 4a and supplementary Fig. 2) strongly suggest that *BCOR-CCNB3* is expressed as a full transcript and that the truncated *CCNB3* is not expressed *per se*. Though cell cycle analyses clearly indicate that the CCNB3 portion of the fusion is potentially sufficient to mediate cell cycle effects, it is also expected that oncogenesis relies on additional phenotypic effects of the full size fusion protein.

Finally, in the series of close to 1000 tumor samples profiled in our laboratory including a spectrum of pediatric tumors, adult sarcomas and brain tumors, only the *BCOR-CCNB3*-positive cases exhibited a very high expression level of the *CCNB3* probesets indicating the outstanding specificity of this marker (Fig. 5a). This specificity was further tested using a CCNB3 specific antibody. A total of 43 tumor samples was investigated including *BCOR-CCNB3*-positive cases, Ewing sarcomas, small round cell desmoplastic tumors, rhabdomyosarcomas, synovial sarcomas. In contrast to the 25 *BCOR-CCNB3*-negative cases that stained completely negative, the 18 *BCOR-CCNB3*-positive cases that could be tested exhibited strong nuclear CCNB3 staining (Fig. 5b). This therefore highlights the potential of a simple CCNB3 immunohistochemistry assay to be used as a diagnostic test for this subgroup of sarcoma.

In conclusion, our study highlights the power of next generation sequencing, particularly RNA-Seq, and microarray experiments for the refined sub-classification of the highly heterogeneous group of sarcomas. Our results have uncovered a distinct gene fusion that shed new light on the development of two tumors with

similar phenotypes, Ewing and Ewing-like tumors. In this regard, we can infer distinct mechanisms of oncogenesis, an observation of critical significance for both diagnostic and possibly, therapeutic interventions in the coming age of personalized medicine. Possible relationships of this group of tumors with *CIC-DUX4*-positive cases will also be important to consider[21]. Further studies will be required to model the *BCOR-CCNB3* fusion in cell and animal models in order to decipher the downstream consequences of this fusion, evaluate the respective contributions of the *BCOR* and *CCNB3* portions of the fusion and gain insights into the underlying oncogenic processes.

Accession Numbers

The *BCOR-CCNB3* sequence has been deposited in GeneBank (accession number JN813375) and Affymetrix expression data in Gene Expression Omnibus (GSE34800)

Acknowledgements

We thank Stephen Chanock for fruitful discussions and critical reading of the manuscript. We are indebted to Virginie Raynal, Patricia Legoix-Ne, Alain Nicolas, David Gentien, Séverine Lair, Alban Lermine, Emmanuel Barillot for critical technical help. We thank the following individuals for contributing cases and paraffin blocks or for supplying clinical information: J Champigneulle, A Croue, P Dechelotte, JM Guinebretière, C Jeanne-Pasquier, A Moreau, JM Picquenot, D Ranchère-Vince, X Sastre-Garau, P Terrier, MC Vacher-Lavenue, V Verkaere, L Guillou. We thank Emmanuel Louis, Samuel Thoraval and Thierry Scarcez from the Life Science Company for help in applying the SOLiD technology. We also thank Nancy Bretschneider, Claudia Gugenmus, Mattias Scherf and Martin Seifert from the Genomatix company for the analysis of data.

This work was supported by grants from the Ligue Nationale Contre le Cancer (Equipe labellisée and CIT program), the Région Ile de France, the INCa (n°2008-044, 0627 and ZP09-027-EPI), the European Union (EET pipeline), and the following associations: Courir pour Mathieu, Dans les pas du Géant, Olivier Chape, Les Bagouzamanon and les Amis de Claire.

Author contributions

SB performed the NGS analysis; GP, SR, and SB characterized the initial case and screened the sarcoma patients; CL and FT performed all the bioinformatics analyses; JMC performed pathological review of the cases and CCNB3 immunohistochemistery experiments; SCG contributed the analysis of clinical files from patients, and with VP to the cloning and cell cycle analyses of *BCOR-CCNB3* and *deltaCCNB3* cDNAs; GP, FT and OD planned and supervised the work and all authors contributed writing the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Table 1. Clinical and pathological data of patients harboring *BCOR-CCNB3* Ewing-like tumors

Case no.	Age at diagnosis (yrs)	Gender	Tumor site	Pathology[a]	CD99[b]	Follow-up[c]
330	15	M	femur	SRCS	M	MR/DOD
985	6	F	neck*	SRCS	M	CR
986	13	F	femur	SRCS	M	U
987	14	M	toe	FCS	C	MR/DOD
988	13	M	femur	SRCS	U	MR
991	6	F	vertebra	SRCS	C	U
992	15	M	tibia	SRCS	-	CR
993	13	M	clavicle	SRCS	M	U
994	11	M	para-vertebral*	SRCS	C	U
995	26	F	chest wall*	SRCS	-	U
996	12	M	pubis	FCS	C	CR
997	10	M	talus bone	SRCS	M	MR/DOD
1005	13	M	sacrum	U	U	U
1006	22	M	pelvis*	SRCS or SV	M	DOD
1007	19	M	rib	SRCS	M	CR
1008	14	M	iliac crest	SCO	C	CR
1009	17	F	sacrum	U	-	U
1010	22	M	iliac crest	SCO	M	CR
1011	8	F	vertebra	Granulocytic sarcoma	U	DOD
1012	19	F	abdominal wall*	Sarcoma or MPNST	C	CR
1013	13	M	femur	FCS	C	DOD
1014	13	F	sacrum	MPNST or FCS or PNET	C	OT
1015	20	M	iliopubic branch	U	U	U
1016	9	M	femur	ESFT/ PNET	U	DOD

a: SRCS, small round cell sarcoma; FCS, Fusiform cell sarcoma; SV, Synovialosarcoma; MPNST, malignant peripheral nerve sheet tumor; SCO, small cell osteosarcoma; ESFT, Ewing sarcoma family of tumors; PNET, Primitive neuroectodermal tumor; U, Unknown.
b: M, membrane positivity; C, cytoplasmic positivity; -, negative; U, unknown.
c: MR, metastatic relapse; DOD, dead of disease; CR, complete remission; LR, local relapse; OT, ongoing treatment; U, unknown status.
* soft tissue tumour

Legends to figures

Figure 1. RNA-Seq identification of the *BCOR-CCNB3* fusion.

a – Scheme of the paired-end coverage of the junction between *BCOR* and *CCNB3* sequences using IGVtools. The transcription orientation is reported in the 5′-3′ direction from *BCOR* exon 15 to *CCNB3* exon5. The junction between both genes is represented by the vertical dotted line. Each PE read is represented by a horizontal segment consisting of an orange box on the left end (the 50bp F3 mate) and a grey box on the right end (the F5 25bp mate) connected by a line representing the unsequenced part of the PE fragment. Duplicated reads have been removed according to the criteria that only PE reads having F3 mates starting at distinct positions are conserved. F3 reads were aligned 5′ to 3′ along the *BCOR* sequence.

b – A scheme of chromosome X with positions and strand orientations of the *BCOR* and *CCNB3* genes.

c - Read coverage of the *BCOR* and *CCNB3* exons (BioScope alignement). The base coverage corresponding to each intron or exon sequence is indicated on top of the genomic organization of *BCOR* and *CCNB3* genes (the Y axis corresponds to the scale of the base coverage). The median base coverage including duplicates of *BCOR* exons is ~2000 (maximum 3939 in exon 4), comparable to *CCNB3* coverage (median ~2500, maximum 5394 on exon 5).

d - RT-PCR detection of a specific 290 base-pair fragment in tumor RNA (T) but not in blood RNAs (C) from the same patient (case 330). CTL: PCR negative control. L: 100bp ladder.

e - Structure of the *BCOR-CCNB3* cDNA. The three ankyrin domains of *BCOR* are indicated in purple and the two cyclin domain of *CCNB3* in green. The sequence of the *BCOR-CCNB3* junction together with the encoded peptide sequence is reported below. The complete *BCOR-CCNB3* nucleotide and peptide sequences are shown in Supplementary Fig. 5.

Figure 2. Characterization of the *BCOR-CCNB3* genomic rearrangement and demonstration of its recurrence in small round cell sarcomas

a - A chromosome X paracentric inversion is documented by FISH experiments. BACs RP11-91I16 and RP11-66502 were labeled with FITC. BACs CTD-2107N14 and RP11-576P23 were labeled with rhodamine. The arrows indicate position and orientation of the *BCOR* and *CCNB3* genes. In normal cells (lymphoblastoid cell line, LL), the two red spots (or the two green spots) are in close vicinity and appear either as a single spot or as unicolor doublets. In *BCOR-CCNB3*-positive cells (330T cells), following the chromosome X paracentric inversion, the two red and green spots are split and appear as bicolored doublets. Scale bar represent 1μm.

b - Left panel: long-range PCR amplification of a specific fragment in the tumor DNA (T) but not in the constitutional DNA (C) of case 330. Right panel: long range PCR of various sizes on genomic DNAs from RT-PCR-positive *BCOR-CCNB3* cases. CTL, PCR negative control. L: 1Kb ladder.

c - Genomic organization of the *BCOR-CCNB3* fusion gene in case 330. A focus on the breakpoint is shown below. *BCOR* and *CCNB3* sequences are indicated in orange and

black respectively. The TGA stop codon of *BCOR* is underlined. As can be seen from the comparison of the genomic and cDNA sequences (Fig. 1e), the *BCOR* TGA stop codon is embedded in a GGTGAG sequence that is used as a donor site spliced with the *CCNB3* exon 5 acceptor site.

d - Position of the breakpoints within the *BCOR 3′ UTR* or *BCOR/MID1IB1* intergenic region and within the *CCNB3* intron 4 region in different *BCOR-CCNB3*-positive samples.

Figure 3. *BCOR-CCNB3* positive cases are distinct from other small round cell tumors and particularly from Ewing's sarcoma.

a- An unsupervised COA multivariate analysis comparing *BCOR-CCNB3*-positive cases with desmoplastic small round cell tumor (DSRCT), Ewing's sarcoma (ES) and malignant rhabdoid tumor cases (MRT). Ostesarcoma, neuroblastoma, rhabdomyosarcoma and synovial sarcoma are not shown for the clarity of the figure. Ewing's sarcoma profiled in the same experiment as *BCOR-CCNB3* are indicated in green and Ewing's sarcomas profiled on the same platform but in previous experiments are indicated in blue. This indicates that the distinct clustering of *BCOR-CCNB3* positive cases and Ewing sarcomas is not linked to a platform bias.

b- A volcano plot comparing *BCOR-CCNB3* positive cases with *EWSR1-ETS* Ewing's sarcoma positive cases. Expression levels were compared using the Welch t-test. Probesets with Bonferroni adjusted Welch two sample t-test *P* values < 0.01 and log

ratio of mean values between the BCOR-CCNB3 and ES groups > 3 are indicated in red (3203 probesets present these characteristics).

c- GSEA on genes ranked differentially between *BCOR –CCNB3* positive cases and *EWSR1-ETS* positive Ewing's sarcomas. Bars indicate each gene of the Zhang_targets_of_EWSR1_FLI1_fusion signature. A strong anti-correlation is observed between *BCOR-CCNB3* overexpressed genes and the *EWSR1-ETS* signature.

d- Heatmap of the *EWSR1-FLI1* signature in *BCOR-CCNB3* positive cases. The genes were ranked using signal-to-noise metric. A complete annotation of this figure is provided in supplementary Fig.3. Similar results were obtained with the gene set RIGGI_Ewing_sarcoma _progenitor_up signature (Supplementary Fig. 3 & 4).

Figure 4. Functional consequences of *BCOR-CCNB3* expression.

a- GSEA on genes ranked differentially between *BCOR-CCNB3*-positive cases and other sarcomas or pediatric tumors (10 tumors randomly chosen in each of the 11 tumor groups shown in Fig. 5a) identifies signatures of Hedgehog and WNT pathways activation (supplementary Table 6).

b- Increased S-phase associated with *BCOR-CCNB3* expression. NIH3T3 cells were transfected with *BCOR-CCNB3* or truncated *CCNB3* (deltaCCNB3) and cell cycle was analysed after BrdU staining (n=3; Welch two sample t-test *P* values=3.3 10^{-3} and 2.4 10^{-3} as compared to empty vector, respectively).

Figure 5. Expression of CCNB3 across a series of tumors.

a - Boxplots of the intensity of the 231481_at probeset (log scale) across the *BCOR-CCNB3* positive cases (10 cases), desmoplastic small round cell tumors (DRSCT, *EWSR1-WT1* positive, 32 cases), Ewing sarcomas (ES , *EWSR1-ETS* positive, 74 cases), gliomas (17 cases), medulloblastomas (Mb, 52 cases), *SMARCB1* deficient Malignant rhabdoid (MRT, 35 cases), neuroblastomas (Nb, 64 cases), osteosarcomas (OS, 35 cases), retinoblastomas (Rb, 39 cases), rhabdomyosarcomas (RMS, 121 cases), synovial sarcomas (Synov, 34 cases), and other sarcomas (288 cases). Normal includes 353 normal tissues (GSE3526) of which three testis samples corresponding to the upper outliers. The sarcomas that express intermediate levels of *CCNB3* tested negative for the *BCOR-CCNB3* fusion. The expression level of *BCOR* is reported in supplementary Fig.2.

b – Hematoxylin-Eosin- staining (HES), immunohistochemistry for CD99 and CCNB3 in an *EWSR1-FLI-1*-positive Ewing sarcoma as compared to two *BCOR-CCNB3*- positive cases. Scale bar (black line) is 20µM for all panels except for CCNB3 staining of *BCOR-CCNB3*-positive cases where it is 40 µM to highlight the negative staining of stromal components (white arrows).

Online methods

RNA extraction

Tumors were snap frozen in liquid nitrogen. After crushing tumors, total RNA was isolated using the Trizol reagent kit (Invitrogen, Carlsbad, Ca) based on the GITC phenol chloroform extraction method. RNAs were quantified by Qubit (Invitrogen, Carlsbad, Ca) and Nanodrop ND1000 (ThermoFisher Scientific, Waltham, MA) before quality assessment with the Agilent 2100 Bioanalyzer (Santa Clara, Ca).

RNA library construction

PolyA mRNAs were purified from 10 µg of total RNA using NucleoTrap® mRNA (Macherey Nagel, Düren, Deutschland) according to the manufacturer's protocol.

Following the SOLiD ™ Total RNA Seq Kit instructions, 100 ng of polyA RNAs were fragmented by incubation with RNAse III for 10 min in a 10 µl reaction volume containing 1 X RNase III buffer supplied with the enzyme. Fragmented RNAs were then purified using the RiboMinus™ Concentration Module (Invitrogen, Carlsbad, Ca). The yield and size distribution of the fragmented RNA were assessed using the Quant-iT™ RNA assay kit with the Qubit® fluorometer (Invitrogen, Carlsbad, Ca) and the RNA 6000 Pico Chip Kit with the Agilent® 2100 Bioanalyzer (Agilent, Santa Clara, Ca). A total of 50 ng of fragmented RNAs were hybridized and ligated with the SOLiD™ adaptator mix and reverse transcribed according to supplier instructions.

The isolated cDNA were size selected around 200 bp using Novex® pre-cast gel products. The cDNA were then amplified according to the SOLiD ™ Total RNA Seq Kit. The yield and size distribution of the cDNA were assessed using the Quant-iT™ HS DNA assay kit with the Qubit® fluorometer (Invitrogen, Carlsbad, Ca) and the High Sensitivity DNA Assay Chip Kit on the Agilent® 2100 Bioanalyzer.

Emulsion PCR

Templated beads were generated for sequencing using standard manufacturer protocols. Briefly, an aqueous phase was prepared from the SOLiD ePCR kit containing AmpliTaq Gold DNA Polymerase UP, buffer, MgCl2, dNTPs, amplification primers and library template. The aqueous phase was then introduced to an oil phase in an ULTRA-TURRAX Turbo Drive (IKA) to create a water-in-oil emulsion. The emulsion was then transferred to a 96 well plate and PCR amplified using the recommended PCR conditions. After PCR amplification, emulsions were broken using butanol, and the beads were washed, enriched, and treated with terminal transferase before quantification and deposition onto a slide for sequencing.

Whole Transcriptome paired end sequencing

Templated beads were deposited on two quadrants of a slide per sample. Massively parallel ligation sequencing was carried out using Applied Biosystems SOLiD System (V4 chemistry) following supplier instructions.

NGS analysis

Apart from Genomatix analysis NGS raw data was aligned in color space using the color-code error-correction capability of LifeTech's BioScope environment and a set of human genome reference sequences associated to the hg19 version of Human Genome. The distribution of insert size, excluding adaptors was 116bp (90% confidence interval 60-163bp). NGS analysis was performed using three independent Paired-End (PE) Gene Fusion analysis pipelines: Genomatix NGS, by Genomatix Software GmbH computational biology company (http://www.genomatix.de), FusionSeq, by the Gerstein academic laboratory (http://rnaseq.gersteinlab.org/fusionseq/)[1] and an in-house pipeline. All three methods, although quite different in conception, detected successfully the presence of a novel *BCOR/CCNB3* chimeric fusion in one sample (330T) and the known *FUS/FEV* fusion in another sample.

Genomatix NGS pipeline

Alignment and analysis of the NGS datasets were performed directly by Genomatix bioinformaticians (http://www.genomatix.de). Raw data were aligned against the human genome (NCBI_build 37) and transcriptome (ElDorado 12-2010; containing transcript annotation from RefSeq, GeneBank, Ensembl and crossmappings) using the two step mapping strategy from the Genomatix Mining Station. Data were mapped with low stringency allowing one point mutation in the seed search (deep) and 85% overall alignment quality in the subsequent step. In addition, global and local spliced alignment was performed to find split reads covering exon boundaries and fusion events. PE statistics were calculated

automatically (mean insert size, standard deviation, strand orientation and pileup distribution) and subsequently used for gene fusion detection.

Gene fusions within and across chromosomes were derived from uniquely mapped PE reads in four steps including filtering, clustering, inclusion of splice-junction information, and scoring. The filtering removed read pairs if one or both reads did not uniquely map to the genome, if the reads aligned within the expected distance, or if any pair was contained in an artificial pileup. All PE that pass the filters and were mapped to transcripts from different loci were considered as gene fusion candidates. After clustering reads mapping to proximal positions, global spliced-alignments were included as an additional line of evidence to determine the exact breakpoint. In the last step, the gene fusion candidates were ranked. The two scores are the number of spliced reads and the number of PE fragments that support any gene fusion candidate (supplementary Table 7).

FusionSeq pipeline

NGS raw data was aligned in color space using the color-code error-correction capability of LifeTech's BioScope environment. Overall the alignment was un-gapped allowing 2 color code mismatches against the reference sequence for each PE read. Local clipping of <5 bases at the start or the end of the PE reads were allowed to maximize the number of aligned reads. In first instance, the F3 and F5 PE mates were aligned independently against three hg19 reference libraries: a library of non-coding sequences (784 non-coding sequences delivered with the BioScope software (including SOLiD™ 4 adapter sequences, rRNAs, tRNAs, Single-base-repeat

sequences, including Poly-A, Poly-T…) to be discarded during alignment, the human genome reference library and a canonical splice junction library, built from RefSeq annotations (.gtf file available via UCSC) that reconstitute the junction sequence between two consecutive exons.

The results of F3 and F5 read alignments were then merged and pairing was performed. No quality filter was applied in advance and the resulting BAM file held the complete set of aligned PE reads.

The BAM file was then passed to FusionSeq following the authors instructions (http://rnaseq.gersteinlab.org/fusionseq/)[9]. The fusion transcript detection module employs the UCSC known Genes set that contains 66,803 isoforms. First, poor quality and orphan reads were removed. Then PE reads that mapped to the same gene were considered as part of the normal transcriptome. PE reads that mapped to different genes were selected as potential candidate fusion transcripts. The complete list of fusion proposals was then passed through a "filtration cascade" to remove spurious candidate fusion genes. We applied in the order the "MitochondrialFilter" and "RepeatMaskerFilter" procedures. The "AnnotationConsistencyFilters" was also applied to remove Fusion Gene candidates where one of the partners was annotated as "collagen", "clone", "ribosomal" or "pseudogene", which were supposed to have elevated sequence homology with genes of the same class. The reduced list of fusion proposals was then classified by FusionSeq according to the type of genomic region involved in a fusion candidate (exons, introns, splice junctions or UTR regions), the absolute number of Exon-to-Exon PE reads encompassing the junction and the

following three statistical indicators: normalized number of Supportive Paired-End Reads (SPER), Difference between the Analytically calculated and Expected SPER (DASPER) and the Ratio of Empirically computed SPERs (RESPER). We extracted from this conspicuous data-set the information concerning the Exon-to-Exon junctions. We used the absolute number of PE fragments encompassing the fusion junction and the DASPER indicator to rank the fusion proposals. This procedure returned a list of 16 fusion proposals that are reported in supplementary Table 8. Probable False-positive candidates were further filtered out based on 1-mapping quality (MAPQ>=20), 2- non redundancy (only one PE was retained for groups of PE reads with F3 and F5 mates starting at identical positions), 3- length of F3 and F5 reads (only 50 and 25bp length mates were retained), 4- consistency of insert size assuming an intronic breakpoint (insert size in the range 60 to 163 was considered normal) and 5- unambiguous mapping (PE reads having homologous sequences in exome and genome were discarded)(Supplementary Table 8).

Inhouse pipeline

A set of in-house scripts was written in bash, awk and R language for Linux system. Raw data was aligned in color space using BioScope ungapped alignment modules but a CDS library for each gene was used instead. CDS annotations were extracted from the UCSC RefSeq GTF annotation file and a fasta seq reference library was built by concatenating the sequence of each exon in their genomic order for each gene. F3 and F5 PE reads were aligned independently by allowing 2 color code mismatches for each read. Finally, the Read-Pairing module of Bioscope was used to

create a PE BAM file. A first read quality filter was applied to remove PE reads whose "mapq" score from samtools quality indicator was lower than 20. We then separated PE reads contained within a single gene from those encompassing different genes by simply checking the gene partners, orientation and expected insert size associated in each PE fragment. We picked the top 10 fusion gene proposals with the higher number of PE read encompassing the junction (supplementary Table 9). False positive candidates were further filtered out using the five criteria aforementioned for the FusionSeq pipeline.

RT-PCR and long range genomic PCR experiments

BCOR3.1 and CCNB3.1 primers (Supplementary Table 10) were used to amplify cDNA and genomic sequences.

Microarray data

Experimental procedures for GeneChip microarray were performed according to the Affymetrix GeneChip Expression Analysis Technical Manual (Affymetrix, Santa Clara, CA) using HG-U133-Plus2 arrays. All microarray data were simultaneously normalized using the gcrma package version 2.22.0 in R 2.12.0 environment[22] and quality assessment was based on Relative Log Expression and Normalized Unscaled Standard Errors (http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/brettschneider). COA analysis was performed using made4[23] and rgl (developed by Daniel Adler and Duncan Murdoch) R packages. Gene set enrichment analysis[24] was done by comparing Ewing-like and ES cases using Signal / Noise ratio algorithm. Ewing-like versus ES comparison was performed using the Welsh t-test

and associated P values were corrected for multiple testing with the Bonferroni method. SAM analyses were performed with the samr R package[22,25] using the 10 *BCOR-CCNB3*-positive samples with a random selection of 10 samples in each 11 other tumor types. Ten different analyses were performed with different samplings. Deltas were chosen as first value with 90th percentile FDR below 0.01. DAVID analyses[26] were performed online (http://david.abcc.ncifcrf.gov/home.jsp) using EASE value<10^{-5}, count>10, Fold Enrichment>2 and Bonferroni *P* value<10^{-2} parameters.

FISH experiments

FISH analysis were performed on 4 μm frozen tissue section fixed 10 min in 1/3 v/v Acetic Acid/Ethanol bath. The slides were immersed in 2X SSC pH 7.0 for 30 min at 37°. The sections were then sequentially dehydrated in alcohol (70%, 90% and 100%) for 2 min each and air dried. The *BCOR* probe was composed of a set of two flanking BAC clones (CHORI): RP11-91I16 (telomeric) and RP11-665O2 (centromeric) both biotin labeled (Biotin Nick translation mix, Roche) and Avidin-Fluorescein revealed (Roche). The *CCNB3* flanking probes were CTD- 2107N14 (telomeric) and RP11-576P23 (centromeric) both digoxigenin labeled (DIG Nick translation mix, Roche) and anti-digoxygenin-Rhodamine Fab Fragment detected (Roche). All BAC positions were assessed by blat after SP6 and T7 end sequencing.

The probes were validated on normal metaphase (male patient), each locus showing an unambiguous unique green or red signal localized on X chromosome.

Immunohistochemistry

4µm-thick paraffin sections were cut and mounted on glass slides (Superfrost+ Menzel Glazer). Subsequently the preparations were entirely treated on automate VENTANA-Benchmark-XT®. The primary antibody was a CCNB3 polyclonal anti rabbit antibody from Sigma Aldrich (St Quentin Fallavier France, Ref HPA000496) and was used at a 1:200 dilution.

BCOR-CCNB3 fusion cDNA cloning

cDNA from case 330T was PCR amplified with Phusion DNA polymerase (Finnzymes, Fisher Scientific, Illkirch, France) and EcoRV_BCOR_fw and XhoI_CCNB3_rev (Supplementary Table 10) primers. EcorV/Xho1 digested fragment was then ligated into pCDNA3.1(+) vector (Invitrogen, Cergy-Pontoise, France) between EcoRV and XhoI restriction sites. *DeltaCCNB3* was PCR amplified from cDNA obtained from testis mRNA (OriGene Technologies, USA) using EcoRV_DCCNB3_fw and XhoI_CCNB3_rev primers (Supplementary Table 10) and cloned into pcDNA3.1(+) vector. The complete cDNA sequences were checked. Cell-cycle analyses were performed by BrdU incorporation as previously described[27].

REFERENCES

1- Delattre, O. *et al*. Gene fusion with an ETS domain caused by chromosome translocation in human tumors. *Nature*. **359**, 162-165 (1992).

2- Delattre, O. *et al*. The Ewing family of tumors: a subgroup of small round cell tumors defined by specific chimeric transcripts. *New Engl J Med*. **331**, 294-299 (1994).

3-Toomey, E.C., Schiffman, J.D., Lessnick SL. Recent advances in the molecular pathogenesis of Ewing's sarcoma. *Oncogene*. **29**, 4504-16 (2010).

4-Szuhai, K. *et al*. The NFATc2 gene is involved in a novel cloned translocation in a Ewing sarcoma variant that couples its function in immunology to oncology. *Clin Cancer Res.* **15**, 2259-68 (2009).

5- Broadhead, M.L.,Clark, J.C., Myers, D.E., Dass, C.R. The molecular pathogenesis of osteosarcoma: a review. Sarcoma. 2011, 959248 (2011).

6- Wilhem, B.T. et *al*. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**, 1239-43 (2008)

7-Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. **5**, 621-8 (2008).

8- Maher, C.A. *et al*. Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc Natl Acad Sci U S A*. **106**, 12353-8. (2009)

9- Sboner, *et al.* FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. Genome Biol. 11, R104 (2010)

10- Peter, M., Gilbert, E., Delattre, O. A multiplex real-time pcr assay for the detection of gene fusions observed in solid tumors. *Lab Invest*. **81**, 905-12 (2001).

11- Ng, T.L et *al.* Ewing sarcoma with novel translocation t(2;16) producing an in-frame fusion of FUS and FEV. *J Mol Diagn*. **9**, 459-63 (2007)

12- Huynh, K.D., Fischle, W., Verdin, E., Bardwell, V.J. BCoR, a novel corepressor involved in BCL-6 repression. *Genes Dev*. **14**, 1810-23 (2000).

13- Gallant, P., Nigg, E.A. Identification of a novel vertebrate cyclin: cyclin B3 shares properties with both A- and B-type cyclins. *EMBO J*. **13**, 595-605 (1994).

14- Armengol, G. *et al*. Recurrent gains of 1q, 8 and 12 in the Ewing family of tumours by comparative genomic hybridization. *Br J Cancer* **75**,1403-9 (1997)

15- Gearhart, M.D., Corcoran, C.M., Wamstad, J.A., Bardwell, V.J. Polycomb Group and SCF Ubiquitin Ligases Are Found in a Novel BCOR Complex That Is Recruited to BCL6 Targets *Mol Cell Biol*, **26**, 6880-6889 (2006)

16- Fan, Z. *et al*. BCOR regulates mesenchymal stem cell function by epigenetic mechanisms. *Nat Cell Biol*. **11**, 1002-9 (2009).

17- Ng, D. *et al*. Oculofaciocardiodental and Lenz microphthalmia syndromes result from distinct classes of mutations in BCOR. *Nat Genet.* **36**, 411-6 (2004).

18- Yamamoto, Y. *et al*. BCOR as a novel fusion partner of retinoic acid receptor alpha in a t(X;17)(p11;q12) variant of acute promyelocytic leukemia. *Blood* **116**, 4274-83 (2010).

19- Grossmann *et al*. Whole genome sequencing identifies somatic mutations of BCOR in acute myeloid leukemia with normal karyotype. *Blood* **118**, 6153-63 (2011).

20- Nguyen, T.B. *et al*. Characterization and expression of mammalian cyclin b3, a prepachytene meiotic cyclin. *J Biol Chem*. **277**, 41960-9 (2002).

21- Kawamura-Saito *et al*. Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35; q13) translocation. *Hum. Mol. Genet.* **15**, 2125-37 (2006).

22- The R manual. Venables, W.N. & Smith, D.M. R Development Core Team, ed. (2005). R: A language and environment for statistical computing. (Vienna, Austria.).

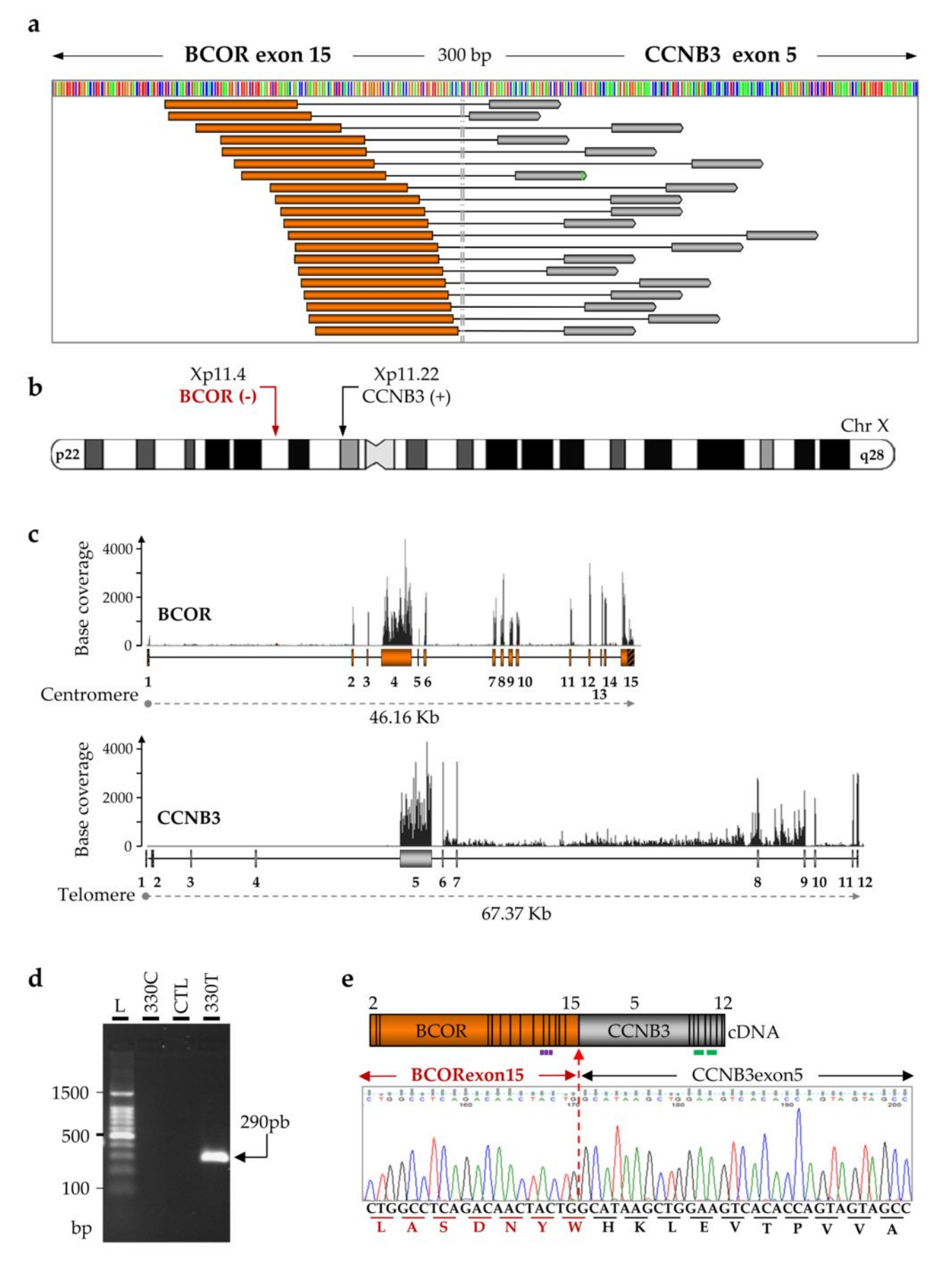
23- Culhane, A. C., Thioulouse, J., Perriere, G., and Higgins, D. G. MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics (Oxford, England) 21, 2789-2790 (2005)

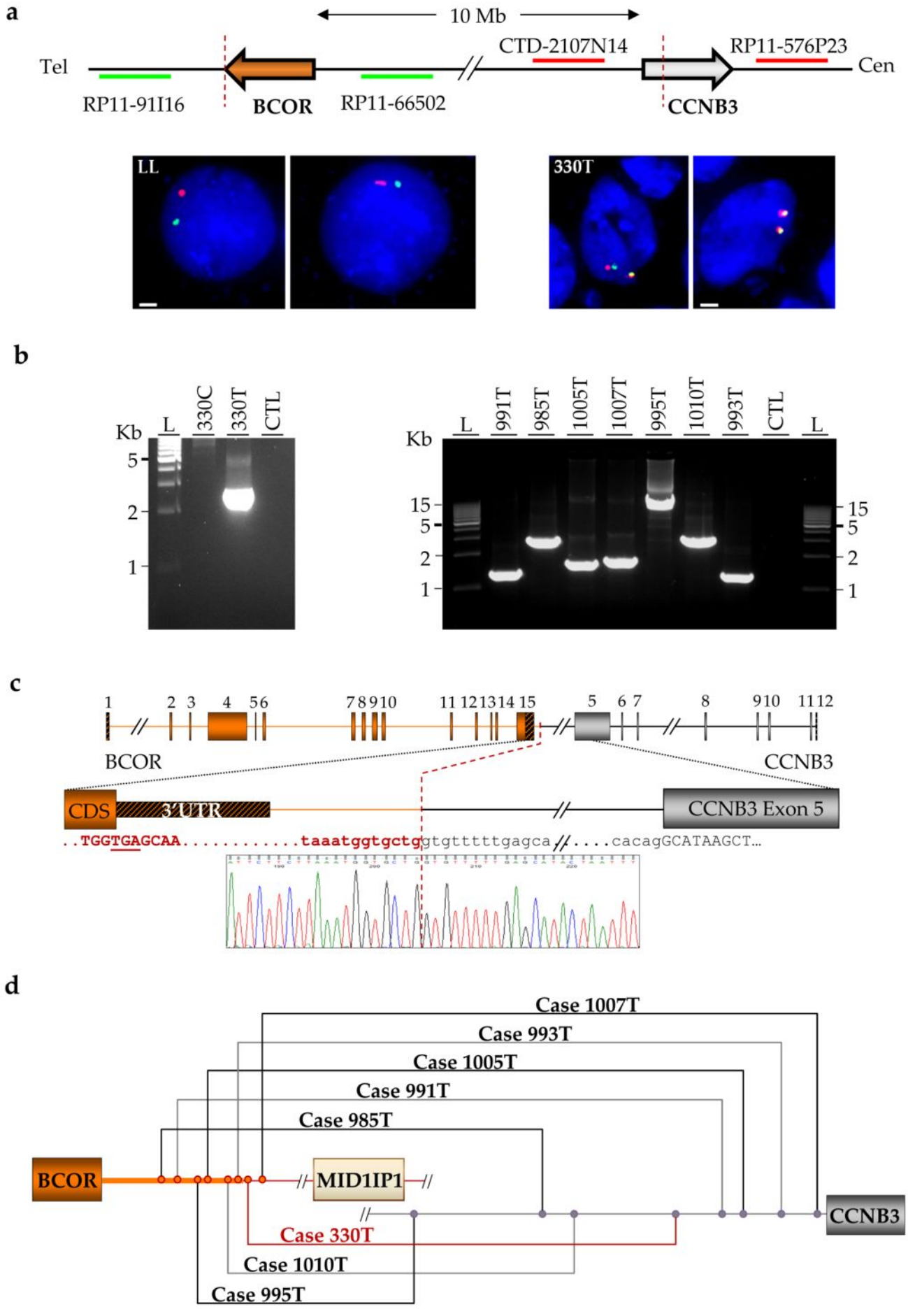
24- Subramanian, A., *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide experssion profiles, *Proc Natl Acad Sci U S A* **102**, 15545–50 (2005)

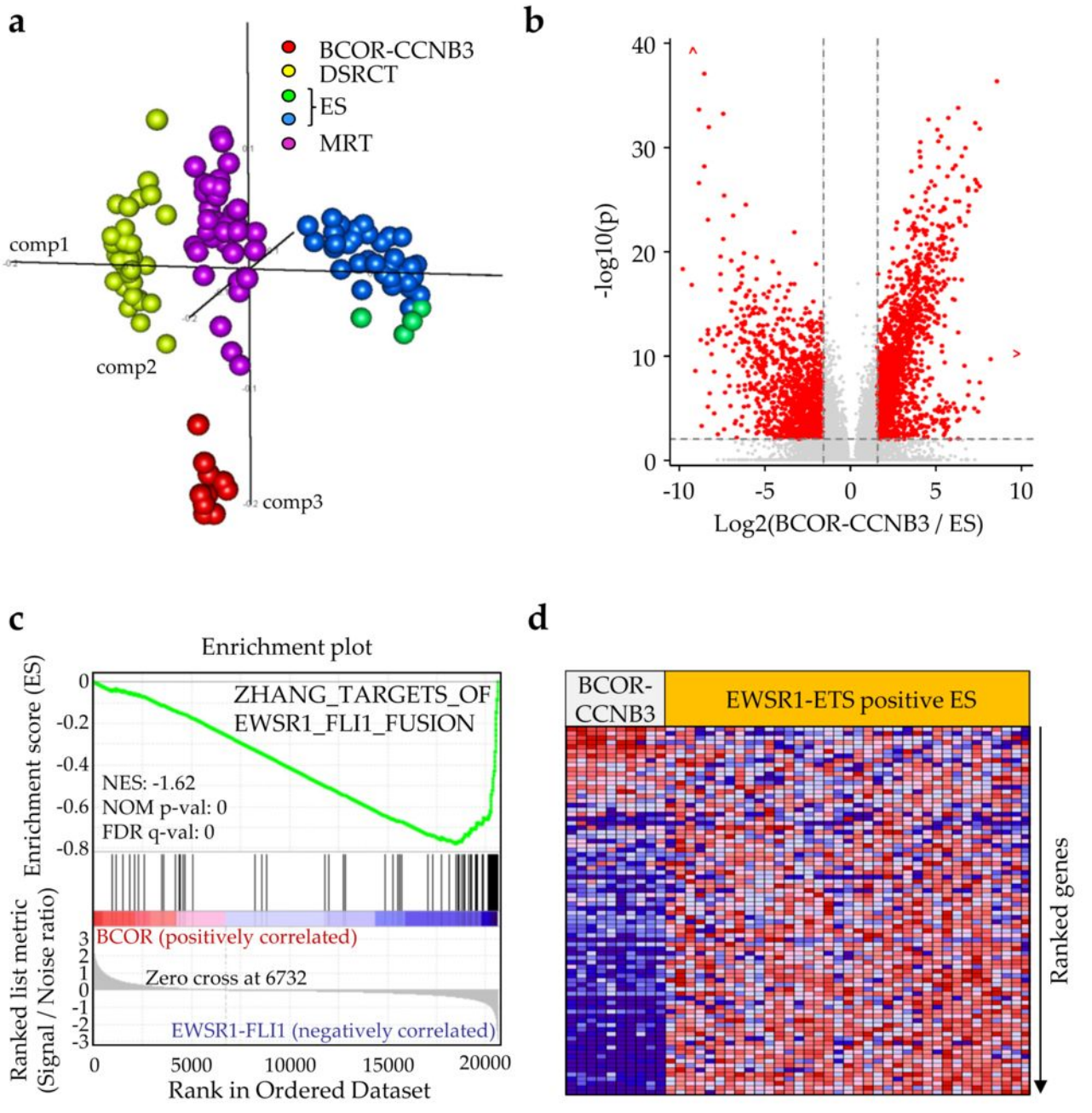
25- Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5116-21 (2001)

26- Huang DW, *et al*. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1-13 (2009)

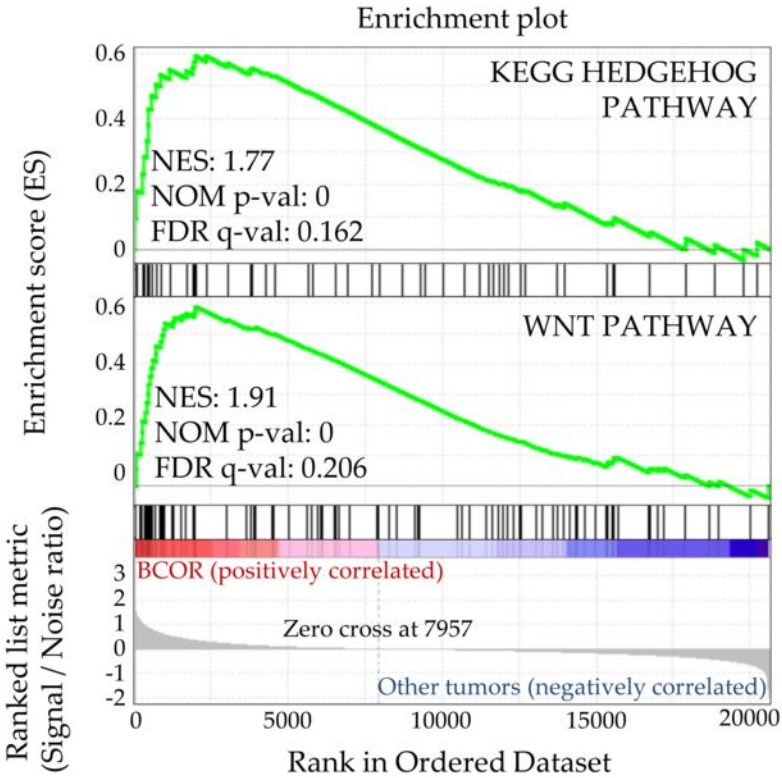
27- Dauphinot *et al.* Analysis of the expression of cell cycle regulators in Ewing cell lines: EWS-FLI-1 modulates p57KIP2and c-Myc expression. *Oncogene* **20**, 3258-65 (2001)


a
BCOR exon 15
300 bp
CCNB3 exon 5
b
Xp11.4
BCOR (-)
Xp11.22
CCNB3 (+)
Chr X
p22
q28
c
Base coverage
4000
2000
0
BCOR
1
2 3
4
5 6
7 8 9 10
11 12
14 15
13
Centromere
46.16 Kb
Base coverage
4000
2000
0
CCNB3
1 2
3
4
5
6 7
8
9 10
11 12
Telomere
67.37 Kb
d
L
330C
CTL
330T
1500
500
100
bp
290pb
e
2
15
5
12
BCOR
CCNB3
cDNA
BCORexon15
CCNB3exon5
CTGGCCTCAGACAACTACTGGCATAAGCTGGAAGTCACACCAGTAGTAGCC
L A S D N Y W H K L E V T P V V A

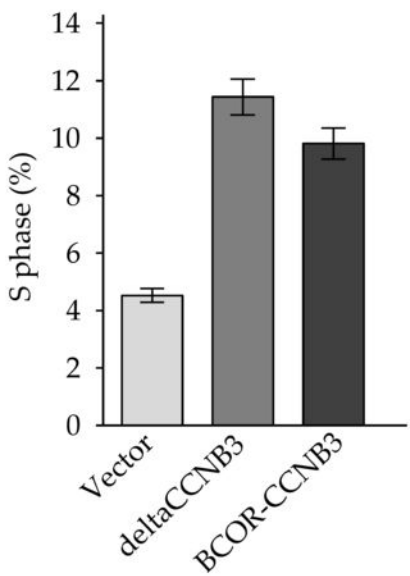

a
10 Mb
Tel
Cen
CTD-2107N14
RP11-576P23
RP11-91I16
BCOR
RP11-66502
CCNB3
LL
330T
b
Kb
L
330C
330T
CTL
5
2
1
991T
985T
1005T
1007T
995T
1010T
993T
15
c
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 5 6 7 8 9 10 11 12
BCOR
CCNB3
CDS
3'UTR
CCNB3 Exon 5
..TGGTGAGCAA.....taaatggtgctggtgttttgagca//....cacagGCATAAGCT...
d
Case 1007T
Case 993T
Case 1005T
Case 991T
Case 985T
BCOR
MID1IP1
CCNB3
Case 330T
Case 1010T
Case 995T

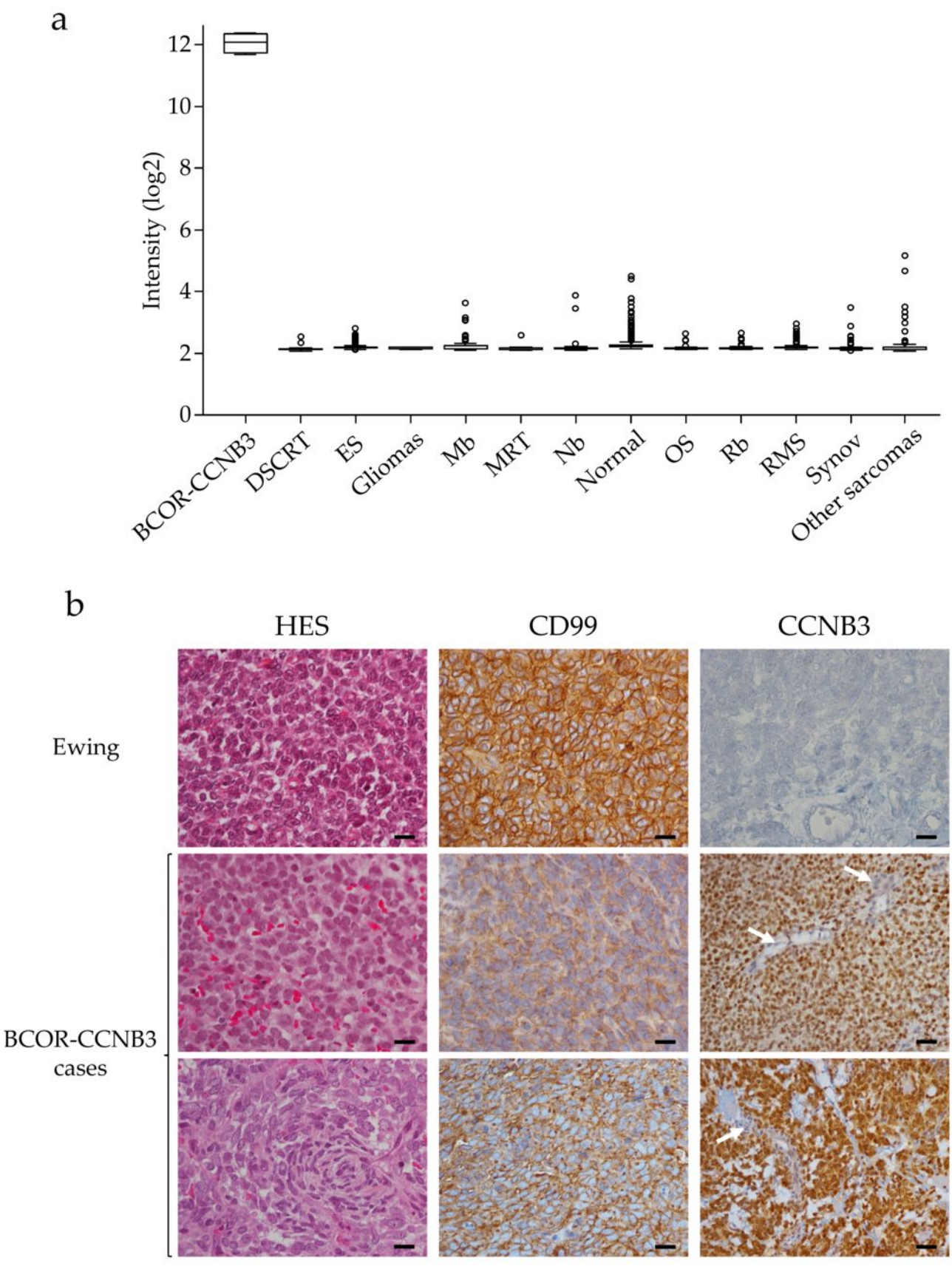

a
BCOR-CCNB3
DSRCT
ES
MRT
comp1
comp2
comp3
b
-log10(p)
Log2(BCOR-CCNB3 / ES)
c
Enrichment plot
Enrichment score (ES)
ZHANG_TARGETS_OF
EWSR1_FLI1_FUSION
NES: -1.62
NOM p-val: 0
FDR q-val: 0
BCOR (positively correlated)
Zero cross at 6732
EWSR1-FLI1 (negatively correlated)
Ranked list metric
(Signal / Noise ratio)
Rank in Ordered Dataset
d
BCOR-
CCNB3
EWSR1-ETS positive ES
Ranked genes

a


Enrichment plot
KEGG HEDGEHOG PATHWAY
NES: 1.77
NOM p-val: 0
FDR q-val: 0.162
WNT PATHWAY
NES: 1.91
NOM p-val: 0
FDR q-val: 0.206
Enrichment score (ES)
BCOR (positively correlated)
Zero cross at 7957
Other tumors (negatively correlated)
Ranked list metric (Signal / Noise ratio)
Rank in Ordered Dataset

b

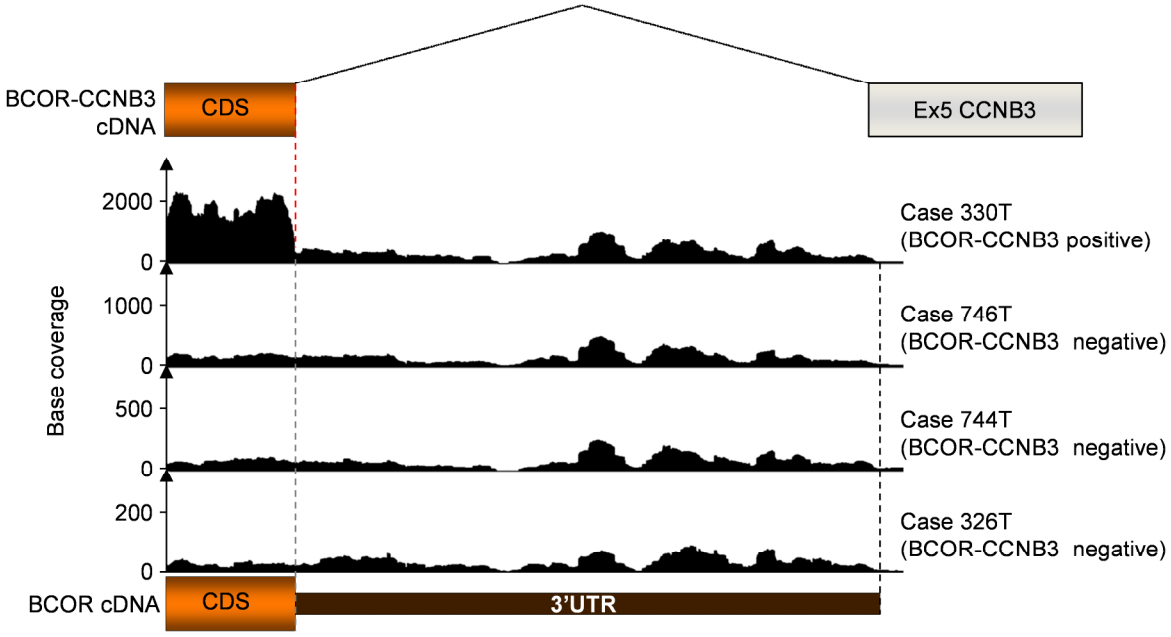

S phase (%)
Vector
deltaCCNB3
BCOR-CCNB3


a
Intensity (log2)
0
2
4
6
8
10
12
BCOR-CCNB3
DSCRT
ES
Gliomas
Mb
MRT
Nb
Normal
OS
Rb
RMS
Synov
Other sarcomas
b
HES
CD99
CCNB3
Ewing
BCOR-CCNB3 cases

A new subtype of bone sarcoma defined by *BCOR-CCNB3* gene fusion

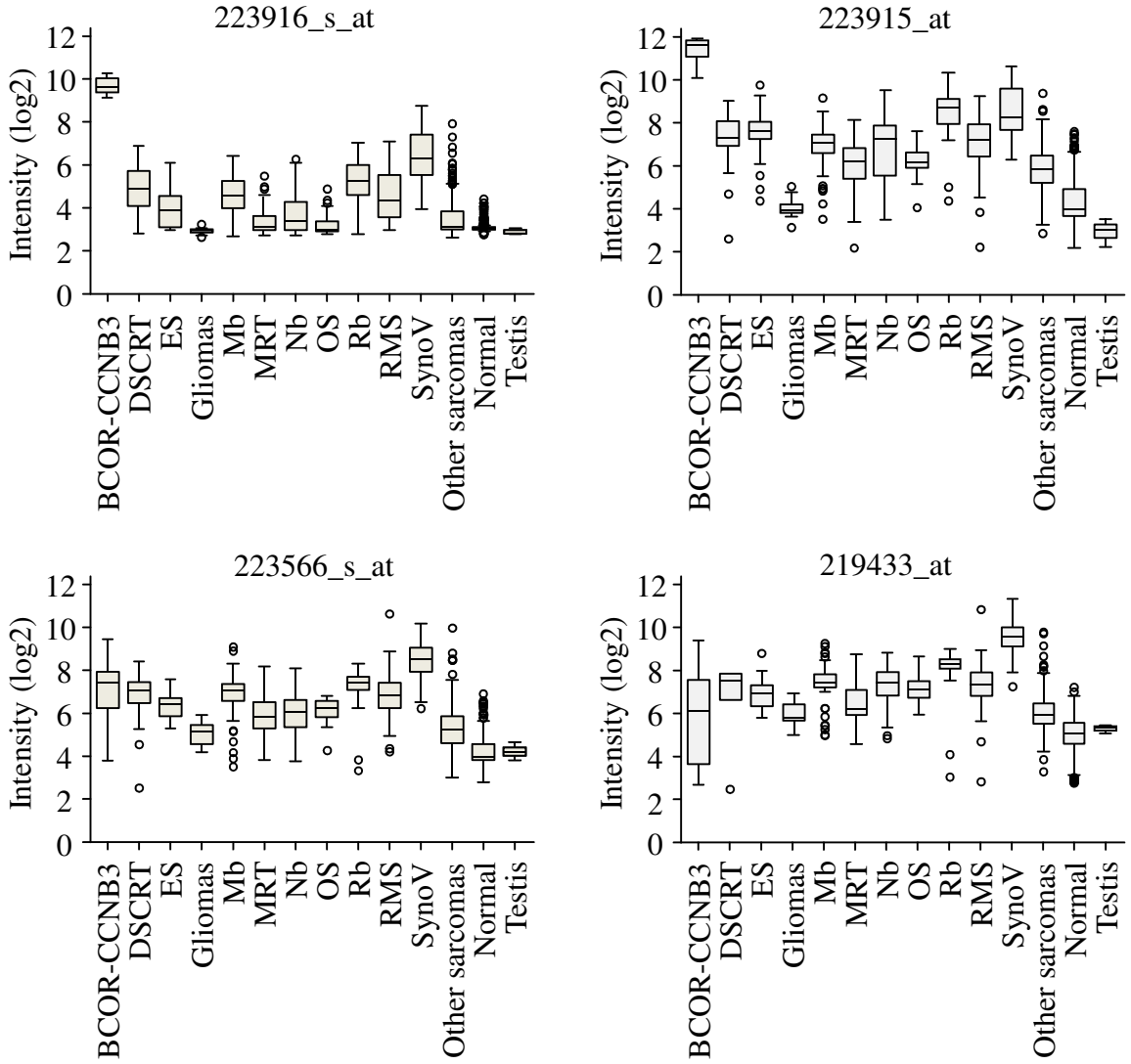
Gaëlle PIERRON*, Franck TIRODE*, Carlo LUCCHESI*, Stéphanie REYNAUD, Stelly BALLET, Sarah COHEN-GOGO, Virginie PERRIN, Jean-Michel COINDRE and Olivier DELATTRE

Supplementary Figure 1
Decreased coverage at the stop codon position of BCOR in the 330T sample



Zoom of the 3′ of the BCOR gene. The orange box indicates the coding sequence and the orange/black hatched line the 3′ UTR. In case 330T a dramatic decrease of the number of reads mapping to the genomic BCOR sequence is observed at the end of the coding sequence (CDS) of exon 15. This decrease is not observed in the three other cases that were sequenced (grey line). Thus, in the 330T sample, the coding part of BCOR is expressed at higher level than the 3′UTR due to its fusion with CCNB3. Yet, there are still reads mapping to the BCOR 3′ end which have to be generated from the same allele since this is a tumor from a male patient with a single chr X, as demonstrated by SNP 6.0 arrays. Hence, the presence of 3′ BCOR UTR reads indicates that the wild type BCOR is also expressed in the 330T case as a result of incomplete usage of the crytic splice site provided by the BCOR STOP codon. This is further supported by microarray data that show that probesets covering the coding sequence of BCOR are expressed at higher levels in BCOR-CCNB3-positive cases as compared to other tumor types whereas probesets covering the 3′ BCOR UTR demonstrate an expression level similar to other tumors (supplementary figure 2).

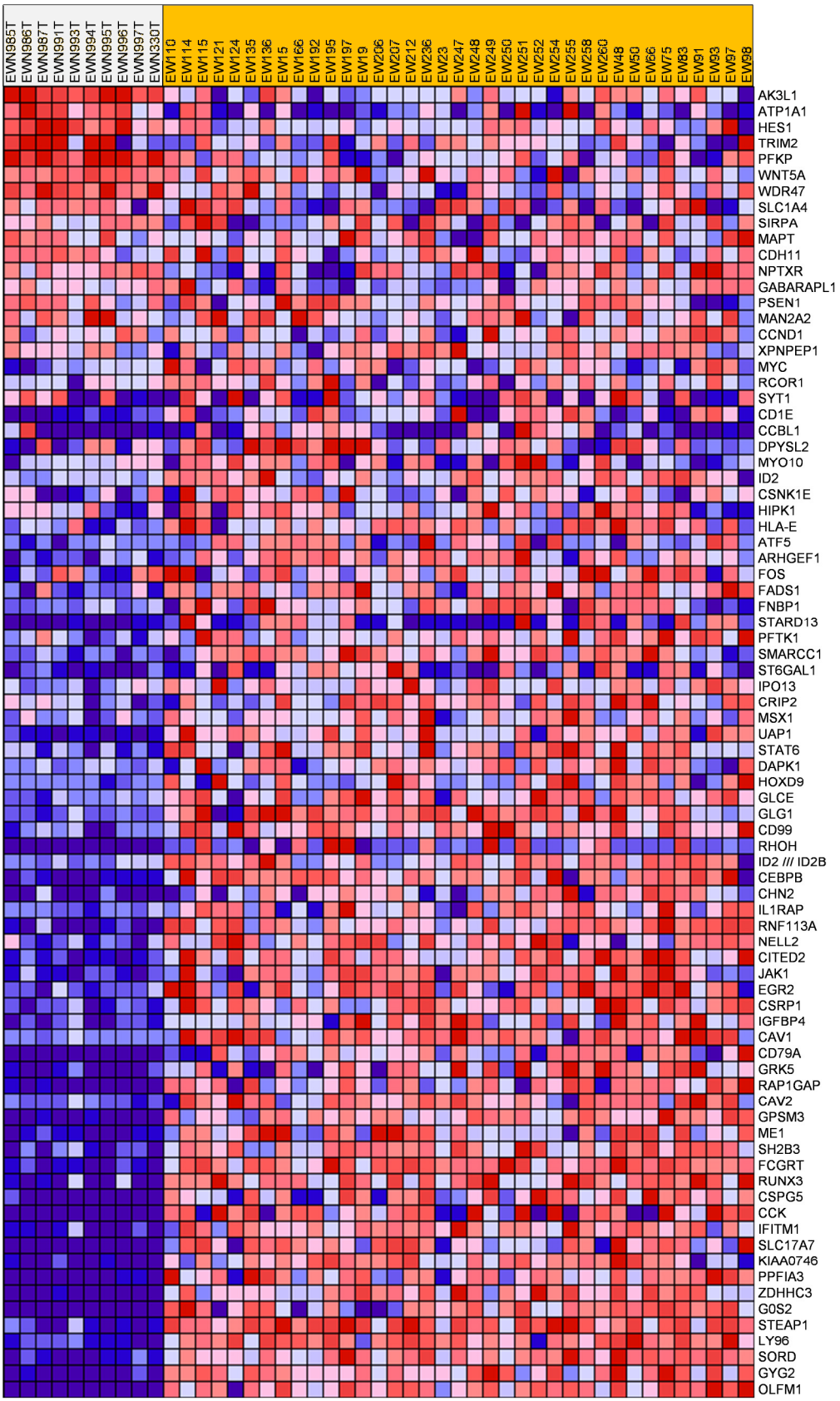
Supplementary Figure 2
Expression of BCOR across a variety of normal tissues and cancers



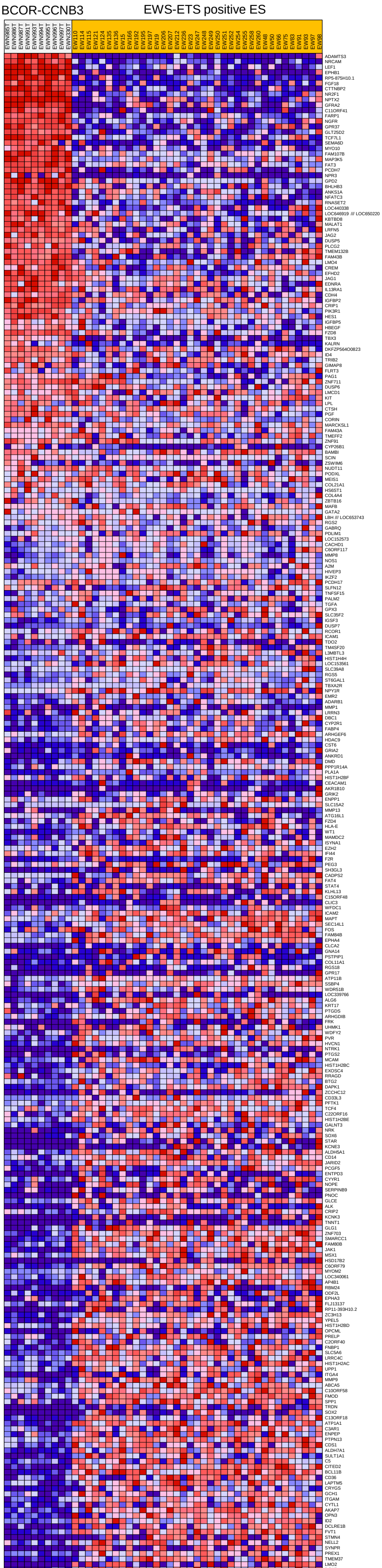
The BCOR gene is covered by four distinct probesets. Two of them, 223916_s_at and 223915_at include probes located along the coding sequence, and demonstrate a higher signal in BCOR-CCNB3-positive cases than in other tumors or normal tissues. These probesets interrogate both the wild-type BCOR and the BCOR-CCNB3 fusion. In contrast, 219433_at and 223566_s_at probesets are located within the BCOR 3' UTR. They indicate a signal similar in BCOR-CCNB3-positive cases and in other tumors or normal tissues. Only the wild type BCOR is evaluated by these probesets.

Supplementary Figure 3

Heat map of the genes included in the Zhang_Targets_of_EWSR1_FLI1 fusion signature



Supplementary Figure 4
Heat map of the genes included in the Riggi_Ewing_sarcoma_progenitor signature



Supplementary Figure 5: Complete nucleotide and peptide sequences of BCOR-CCNB3

Homo sapiens BCL6 corepressor - cyclin B3 (BCOR-CCNB3), CDS

```
   1 atgctctcagcaaccccccctgtatgggaacgttcacagctggatgaacagcgagagggtc
  61 cgcatgtgtggggcgagcgaagacaggaaaatccttgtaaatgatggtgacgcttcaaaa
 121 gccagactggaactgagggaagagaatcccttgaaccacaacgtggtggatgcgagcacg
 181 gcccataggatcgatggcctggcagcactgagcatggaccgcactggcctgatccgggaa
 241 gggctgcgggtcccgggaaacatcgtctattctagcttgtgtggactgggctcagagaaa
 301 ggtcgggaggctgccacaagcactctaggtggccttgggttttcttcggaaagaaatcca
 361 gagatgcagttcaaaccgaatacacccgagacagtggaggcttctgccgtctctggaaaa
 421 cccccaaatggcttcagtgctatatacaaaacaccgcctggaatacaaaaaagtgctgta
 481 gccacagcagaagcgctgggcttggacaggcctgccagcgacaaacagagccctctcaac
 541 atcaatggtgctagttatctgcggctgccctgggtcaatccttacatggagggtgccacg
 601 ccagccatctacccttcctcgactcgccaaataagtattcactgaacatgtacaaggcc
 661 ttgctacctcagcagtcctacagcttggcccagccgctgtattctccagtctgcaccaat
 721 ggggagcgctttctctacctgccgccacctcactacgtcggtccccacatcccatcgtcc
 781 ttggcatcacccatgaggctctcgacaccttcggcctccccagccatcccgcctctcgtc
 841 cattgcgcagacaaaagcctcccgtggaagatgggcgtcagccctgggaatcctgttgat
 901 tcccacgcctatcctcacatccagaacagtaagcagcccagggttccctctgccaaggcg
 961 gtcaccagtggcctgccgggggacacagctctcctgttgcccccctcgcctcggccgtca
1021 ccccgagtccaccttcccacccagcctgctgcagacacctactcggagttccacaagcac
1081 tatgccaggatctccacctctccttcagttgccctgtcaaagccatacatgacagttagc
1141 agcgagttccccgcggccaggctctccaatggcaagtatcccaaggctccggaagggggc
1201 gaaggtgcccagccagtgcccgggcatgcccggaagacagcggttcaagacagaaaagat
1261 ggcagctcacctcctctgttggagaagcagaccgttaccaaagacgtcacagataagcca
1321 ctagacttgtcttctaaagtggtggatgtagatgcttccaaagctgaccacatgaaaaag
1381 atggctcccacggtcctggttcacagcagggctggaagtggcttagtgctctccggaagt
1441 gagattccgaaagaaacactatctcctccaggaaatggttgtgctatctatagatctgaa
1501 atcatcagcactgctccctcatcctgggtggtgcccgggccaagtcctaacgaagagaac
1561 aatggcaaaagcatgtcgctgaaaaacaaggcattggactgggcgataccacagcagcgg
1621 agttcatcatgcccgcgcatgggcggcaccgatgctgtcatcactaacgtttcagggtca
1681 gtgtcgagtgcaggccgcccagcctccgcatcacccgcccccaatgccaatgcagatggc
1741 accaaaaccagcaggagctctgtagaaaccacaccatccgttattcagcacgtgggccag
1801 cccccggccactcctgccaagcacagtagcagcaccagcagcaagggcgccaaagccagc
1861 aacccagaaccgagtttcaaagcaaacgagaacggccttccaccaagctctatatttctg
1921 tctccaaatgaggcattcaggtccccaccaattccctaccccaggagttacctcccttac
1981 ccagcccctgagggcattgctgtaagtcccctctccttacatggcaaaggacctgtctac
2041 cctcacccagttttgttacccaatggcagtctgtttcctgggcaccttgccccaaagcct
2101 gggctgccctatgggcttcccaccggccgtccagagtttgtgacctaccaagatgccctg
2161 gggttgggcatggtgcatcccatgttgataccacacacgcccatagagattactaaagag
2221 gagaaaccagagaggagatcccggtcccatgagagagcccgttacgaggacccaaccctc
2281 cggaatcggttttccgagattttggaaactagcagcaccaagttacatccagatgtcccc
2341 accgacaagaacctaaagccgaaccccaactggaatcaagggaagactgttgtcaaaagc
2401 gacaagcttgtctacgtagaccttctccgagaagaaccagatgctaaaactgacacaaac
2461 gtgtccaaacccagctttgcagcagagagtgttggccagagcgctgagccccccaagccc
2521 tcagttgagccggccctgcagcagcaccgtgatttcatcgccctgagagaggagttgggg
2581 cgcatcagtgacttccacgaaacttatactttcaaacagccagtcttcaccgtaagcaag
2641 gacagtgttctggcaggtaccaacaaagagaacctagggttgccagtctcgactccattc
2701 ctggagccacctctggggagcgatggccctgctgtaacttttggtaaaacccaagaggat
2761 cccaaaccattttgtgtgggcagtgccccaccaagtgtggatgtgacccccacctatacc
2821 aaagatggagctgatgaggctgaatcaaatgatggcaaagttctgaaaccgaagccatct
2881 aagctggcaaagagaatcgccaactcagcgggttacgtgggtgaccgattcaaatgtgtc
2941 actaccgaactgtatgcagattccagtcagctcagccgggagcaacgggcattgcagatg
3001 gaaggattacaagaggacagtattttatgtctacccgctgcttactgtgagcgtgcaatg
3061 atgcgcttctcagagttggagatgaaagaaagagaaggtggccacccagcaaccaaagac
3121 tccgagatgtgcaaattcagcccagccgactgggaaaggttgaaaggaaatcaggacaaa
```

3181 aagccaaagtcggtcaccctggaggaggccattgcagaacagaacgaaagtgagagatgc
3241 gagtatagtgttggaaacaagcaccgtgatccctttgaagccccagaggacaaagatctt
3301 cctgtggagaagtactttgtggagaggcagcctgtgagcgagcctcccgcagaccaggtg
3361 gcctcggacatgcctcacagccccaccctccgggtggacaggaaacgcaaagtctcaggt
3421 gacagcagccacactgagaccactgcggaggaggtgccagaggacccctctgctgaaagcc
3481 aaacgccgacgagtctctaaagatgactggcctgagagggaaatgacaaacagttcctct
3541 aaccacttagaagacccacattatagtgagctgaccaacctgaaggtgtgcattgaatta
3601 acagggctccatcctaaaaaacaacgccacttgctgcaccttagagaacgatgggagcag
3661 caggtgtcggcagcagatggcaaacctggccggcaaagcaggaaggaagtgacccaggcc
3721 actcagcctgaggccattcctcaggggactaacatcactgaagagaaacctggcaggaaa
3781 agggcagaggccaaaggcaacagaagctggtcggaagagtctcttaaacccagtgacaat
3841 gaacaaggcttgcctgtgttctccggctctccgcccatgaagagtctttcatccaccagt
3901 gcaggcggcaaaaagcaggctcagccaagctgcgcaccagcctccaggccgcctgccaaa
3961 cagcagaaaattaaagaaaaccagaagacagatgtgctgtgtgcagacgaagaagaggat
4021 tgccaggctgcctccctgctgcagaaatacaccgacaacagcgagaagccatccgggaag
4081 agactgtgcaaaaccaaacacttgatccctcaggagtccaggcggggattgccactgaca
4141 ggggaatactacgtggagaatgccgatggcaaggtgactgtccggagattcagaaagcgg
4201 ccggagcccagttcggactatgatctgtcaccagccaagcaggagccaaagcccttcgac
4261 cgcttgcagcaactgctaccagcctcccagtccacacagctgccatgctcaagttcccct
4321 caggagaccacccagtctcgccctatgccgccggaagcacggagacttattgtcaataag
4381 aacgctggcgagacccttctgcagcgggcagccaggcttggctatgaggaagtggtcctg
4441 tactgcttagagaacaagatttgtgatgtaaatcatcgggacaacgcaggttactgcgcc
4501 ctgcatgaagcttgtgctaggggctggctcaacattgtgcgacacctccttgaatatggc
4561 gctgatgtcaactgtagtgcccaggatggaaccaggcctctgcacgatgctgttgagaac
4621 gatcacttggaaattgtccgactacttctctcttatggtgctgaccccaccttggctacg
4681 tactcaggtagaaccatcatgaaaatgacccacagtgaacttatggaaaagttcttaaca
4741 gattatttaaatgacctccagggtcgcaatgatgatgacgccagtggcacttgggacttc
4801 tatggcagctctgtttgtgaaccagatgatgaaagtggctatgatgttttagccaacccc
4861 ccaggaccagaagaccaggatgatgatgacgatgcctatagcgatgtgtttgaatttgaa
4921 ttttcagagacccccctcttaccgtgttataacatccaagtatctgtggctcaggggcca
4981 cgaaactggctactgctttcggatgtccttaagaaattgaaaatgtcctcccgcatattt
5041 cgctgcaattttccaaacgtggaaattgtcaccattgcagaggcagaattttatcggcag
5101 gtttctgcaagtctcttgttctcttgctccaaagacctggaagccttcaaccctgaaagt
5161 aaggagctgttagatctggtggaattcacgaacgaaattcagactctgctgggctcctct
5221 gtagagtggctccaccccagtgatctggcctcagacaactactggcataagctggaagtc
5281 acaccagtagtagcctctactaccgtggtaccaaacattatggagaaaccactcattcta
5341 gacatatccaccacctccaaaacacccaacactgaggaggcatctctcttcagaaagcca
5401 ttagttttaaaggaggaacccactattgaggatgaaacccttatcaataagtcattatct
5461 ttaaaaaagtgctcaaatcatgaggaggtgtccttactggaaaagctacagcccctgcag
5521 gaggagagtgacagtgatgatgcgtttgttatagagccaatgacttttaagaagacacat
5581 aaaactgaggaggcagccatcaccaagaagacattatccttaaagaagaagatgtgtgca
5641 agtcagcggaagcagtcctgccaggaagagtcgttggctgtgcaggatgtcaatatggaa
5701 gaggattccttctttatggagtcaatgagttttaagaagaagcctaaaactgaggagtca
5761 atccccacccataagttatcatctttaaagaagaaatgtaccatttatgggaagatatgc
5821 cactttaggaagccaccagtattgcagacaaccatctgtggagcaatgtcctccattaag
5881 aagcctaccactgagaaggagacacttttccaagagctatctgtattgcaagagaaacac
5941 accactgagcatgagatgtccatcttgaagaaatcattggccttgcagaagaccaacttt
6001 aaagaggattcccttgttaaggagtcgttagcctttaagaagaagcctagcactgaggag
6061 gcaatcatgatgccagtaatattgaaggagcagtgcatgactgaggggaagaggtcccgt
6121 ctgaagccattagtattgcaggagatcacctctggagagaagtcgctcattatgaagcca
6181 ttgtccattaaagaaaagccatctactgagaaggagtccttttcccaggaaccatctgca
6241 ttgcaaaagaagcacaccactcaggaggaggtttccatcttaaaggagccctcgtccttg
6301 ctaaagtctccaactgaggagtcacctttgatgaggctttggcttttacaaagaagtgt
6361 accattgaggaggcacccccaccaagaagcctttaattttaaagaggaagcatgccact
6421 caggggacaatgtcccacttgaagaaaccactaatattacagaccacctctggagaaaag
6481 tcacttattaaggagccactgccctttaaagaagaaaaagtgtctttaaagaaaaagtgt
6541 accacacaagagatgatgtccatctgtccagaactgttggactttcaggatatgattggt
6601 gaagataagaattctttctttatggagccaatgtcatttaggaagaaccctacaactgag
6661 gagacagtacttaccaagacatcgttgtctttacaggaaaagaaaattactcaggggaag
6721 atgtcccacttaaagaagccactggtcttgcagaagatcacttctgaggaggagtcattc
6781 tataagaagctgttgccctttaagatgaaatctacaacggaagaaaagttcctctcccag

```
6841 gaaccatctgcattgaaagagaagcataccaccttgcaggaagtgtccctctcaaaagag
6901 tcattggccatccaagagaaggctaccactgaggaggaattctctcaggaactattttca
6961 ttgcatgttaagcataccaacaaaagtgggtccctcttccaggaggctttggtcttgcaa
7021 gagaagactgatgccgaagaggattccttgaagaacttgttggctttgcaggagaaaagc
7081 accatggaagaagagtcccttatcaataagctattggctctgaaggaggagctttctgct
7141 gaggcagccacaaacatacagacacaattatctttaaagaagaagtccacttctcatgga
7201 aaagtgttcttcctgaagaagcagttggctttgaatgagaccatcaatgaagaggagttc
7261 cttaataagcagccactggccttggaggggtatcccagcattgcggaggggagaccctc
7321 ttcaagaagcttttggccatgcaggaggagcccagcattgagaaggaagctgtcctcaag
7381 gagcccactattgacacagaagctcactttaaggaacctttggccttgcaggaggagccc
7441 agcactgagaaggaggctgtcctcaaggagcccagtgttgacacagaagctcactttaag
7501 gaaactttggccttgcaggagaagcccagcattgagcaggaggccctctttaagcgacac
7561 tcagctttgtgggagaagcccagcactgagaaggagaccatcttcaaggagtctttggac
7621 ttgcaagagaagcccagcattaagaaagagaccctcctcaaaaagccattagccttgaag
7681 atgtctaccatcaatgaggcagtcctcttcgaagatatgatagctctgaatgagaaaccc
7741 accactgggaaggagttgtccttcaaggagccattagccttacaagagagtcccacctac
7801 aaggaagacacctttctcaaaacattgttggtcccccaagttggaaccagcccaaatgtg
7861 tctagcactgcccctgaatccataaccagcaagtccagcattgctaccatgaccagtgtg
7921 ggcaaatctggtaccatcaatgaggcattcctcttcgaagatatgataactctgaatgag
7981 aaacccaccactgggaaggagttgtccttcaaggagccattggccttacaagagagtccc
8041 acctgcaaggaagacacctttctggaaacattcttgatcccccaaattggaaccagccca
8101 tatgtgtttagcaccacccctgaatccataacagagaagtccagcattgcaaccatgacc
8161 agcgtgggcaagtccaggaccaccaccgagtccagtgcatgtgaatctgcttctgataaa
8221 cctgtctcaccacaggccaagggaacaccaaaggagataaccccacgggaagatattgat
8281 gaggacagcagtgatccaagtttcaacccaatgtatgccaaggaaatcttcagttacatg
8341 aaagagagagaggaacagtttatacttacagattacatgaacaggcagattgaaatcacc
8401 agtgacatgagggccattcttgtggactggttggtggaggtgcaggtgtcctttgagatg
8461 acccatgagaccctgtacttggcagtgaagctggtggatctctacctaatgaaggcagta
8521 tgcaagaaggataagttacaactccttggtgccactgcctttatgattgcagcaaaattt
8581 gaggagcacaactcacctcgtgtggatgactttgtgtacatctgtgatgataattatcag
8641 cgatctgaggtactcagcatggaaatcaacatcctgaacgtcctcaaatgtgacattaac
8701 attcccatcgcctaccattttctgcgcagatatgctaggtgtatccacaccaacatgaag
8761 acactgaccttgtcccgctacatctgcgagatgaccctgcaggaataccactatgtccag
8821 gagaaggcttccaagctagctgctgcctccttactcctggccctctacatgaagaagctc
8881 ggatactgggttcccttcctggagcattacagtggctacagtatctctgagcttcacccc
8941 ttggtcagacagctgaacaaactgctgactttcagttcttacgatagtctcaaggctgtg
9001 tattacaagtattctcacccggtcttctttgaagtcgccaaaatccctgccttggatatg
9061 ttgaagctggaggagattttgaactgtgattgtgaggctcagggcctggtactctag
```

BCOR-CCNB3 amino acid sequence

```
   1 MLSATPLYGNVHSWMNSERVRMCGASEDRKILVNDGDASKARLELREENPLNHNVVDAST
  61 AHRIDGLAALSMDRTGLIREGLRVPGNIVYSSLCGLGSEKGREAATSTLGGLGFSSERNP
 121 EMQFKPNTPETVEASAVSGKPPNGFSAIYKTPPGIQKSAVATAEALGLDRPASDKQSPLN
 181 INGASYLRLPWVNPYMEGATPAIYPFLDSPNKYSLNMYKALLPQQSYSLAQPLYSPVCTN
 241 GERFLYLPPPHYVGPHIPSSLASPMRLSTPSASPAIPPLVHCADKSLPWKMGVSPGNPVD
 301 SHAYPHIQNSKQPRVPSAKAVTSGLPGDTALLLPPSPRPSPRVHLPTQPAADTYSEFHKH
 361 YARISTSPSVALSKPYMTVSSEFPAARLSNGKYPKAPEGGEGAQPVPGHARKTAVQDRKD
 421 GSSPPLLEKQTVTKDVTDKPLDLSSKVVDVDASKADHMKKMAPTVLVHSRAGSGLVLSGS
 481 EIPKETLSPPGNGCAIYRSEIISTAPSSWVVPGPSPNEENNGKSMSLKNKALDWAIPQQR
 541 SSSCPRMGGTDAVITNVSGSVSSAGRPASASPAPNANADGTKTSRSSVETTPSVIQHVGQ
 601 PPATPAKHSSSTSSKGAKASNPEPSFKANENGLPPSSIFLSPNEAFRSPPIPYPRSYLPY
 661 PAPEGIAVSPLSLHGKGPVYPHPVLLPNGSLFPGHLAPKPGLPYGLPTGRPEFVTYQDAL
 721 GLGMVHPMLIPHTPIEITKEEKPERRSRSHERARYEDPTLRNRFSEILETSSTKLHPDVP
 781 TDKNLKPNPNWNQGKTVVKSDKLVYVDLLREEPDAKTDTNVSKPSFAAESVGQSAEPPKP
 841 SVEPALQQHRDFIALREELGRISDFHETYTFKQPVFTVSKDSVLAGTNKENLGLPVSTPF
 901 LEPPLGSDGPAVTFGKTQEDPKPFCVGSAPPSVDVTPTYTKDGADEAESNDGKVLKPKPS
 961 KLAKRIANSAGYVGDRFKCVTTELYADSSQLSREQRALQMEGLQEDSILCLPAAYCERAM
1021 MRFSELEMKEREGGHPATKDSEMCKFSPADWERLKGNQDKKPKSVTLEEAIAEQNESERC
1081 EYSVGNKHRDPFEAPEDKDLPVEKYFVERQPVSEPPADQVASDMPHSPTLRVDRKRKVSG
```

1141 DSSHTETTAEEVPEDPLLKAKRRRVSKDDWPEREMTNSSSNHLEDPHYSELTNLKVCIEL
1201 TGLHPKKQRHLLHLRERWEQQVSAADGKPGRQSRKEVTQATQPEAIPQGTNITEEKPGRK
1261 RAEAKGNRSWSEESLKPSDNEQGLPVFSGSPPMKSLSSTSAGGKKQAQPSCAPASRPPAK
1321 QQKIKENQKTDVLCADEEEDCQAASLLQKYTDNSEKPSGKRLCKTKHLIPQESRRGLPLT
1381 GEYYVENADGKVTVRRFRKRPEPSSDYDLSPAKQEPKPFDRLQQLLPASQSTQLPCSSSP
1441 QETTQSRPMPPEARRLIVNKNAGETLLQRAARLGYEEVVLYCLENKICDVNHRDNAGYCA
1501 LHEACARGWLNIVRHLLEYGADVNCSAQDGTRPLHDAVENDHLEIVRLLLSYGADPTLAT
1561 YSGRTIMKMTHSELMEKFLTDYLNDLQGRNDDDASGTWDFYGSSVCEPDDESGYDVLANP
1621 PGPEDQDDDDDAYSDVFEFEFSETPLLPCYNIQVSVAQGPRNWLLLSDVLKKLKMSSRIF
1681 RCNFPNVEIVTIAEAEFYRQVSASLLFSCSKDLEAFNPESKELLDLVEFTNEIQTLLGSS
1741 VEWLHPSDLASDNYWHKLEVTPVVASTTVVPNIMEKPLILDISTTSKTPNTEEASLFRKP
1801 LVLKEEPTIEDETLINKSLSLKKCSNHEEVSLLEKLQPLQEESDSDDAFVIEPMTFKKTH
1861 KTEEAAITKKTLSLKKKMCASQRKQSCQEESLAVQDVNMEEDSFFMESMSFKKKPKTEES
1921 IPTHKLSSLKKKCTIYGKICHFRKPPVLQTTICGAMSSIKKPTTEKETLFQELSVLQEKH
1981 TTEHEMSILKKSLALQKTNFKEDSLVKESLAFKKKPSTEEAIMMPVILKEQCMTEGKRSR
2041 LKPLVLQEITSGEKSLIMKPLSIKEKPSTEKESFSQEPSALQKKHTTQEEVSILKEPSSL
2101 LKSPTEESPFDEALAFTKKCTIEEAPPTKKPLILKRKHATQGTMSHLKKPLILQTTSGEK
2161 SLIKEPLPFKEEKVSLKKKCTTQEMMSICPELLDFQDMIGEDKNSFFMEPMSFRKNPTTE
2221 ETVLTKTSLSLQEKKITQGKMSHLKKPLVLQKITSEEESFYKKLLPFKMKSTTEEKFLSQ
2281 EPSALKEKHTTLQEVSLSKESLAIQEKATTEEEFSQELFSLHVKHTNKSGSLFQEALVLQ
2341 EKTDAEEDSLKNLLALQEKSTMEEESLINKLLALKEELSAEAATNIQTQLSLKKKSTSHG
2401 KVFFLKKQLALNETINEEEFLNKQPLALEGYPSIAEGETLFKKLLAMQEEPSIEKEAVLK
2461 EPTIDTEAHFKEPLALQEEPSTEKEAVLKEPSVDTEAHFKETLALQEKPSIEQEALFKRH
2521 SALWEKPSTEKETIFKESLDLQEKPSIKKETLLKKPLALKMSTINEAVLFEDMIALNEKP
2581 TTGKELSFKEPLALQESPTYKEDTFLKTLLVPQVGTSPNVSSTAPESITSKSSIATMTSV
2641 GKSGTINEAFLFEDMITLNEKPTTGKELSFKEPLALQESPTCKEDTFLETFLIPQIGTSP
2701 YVFSTTPESITEKSSIATMTSVGKSRTTTESSACESASDKPVSPQAKGTPKEITPREDID
2761 EDSSDPSFNPMYAKEIFSYMKEREEQFILTDYMNRQIEITSDMRAILVDWLVEVQVSFEM
2821 THETLYLAVKLVDLYLMKAVCKKDKLQLLGATAFMIAAKFEEHNSPRVDDFVYICDDNYQ
2881 RSEVLSMEINILNVLKCDINIPIAYHFLRRYARCIHTNMKTLTLSRYICEMTLQEYHYVQ
2941 EKASKLAAASLLLALYMKKLGYWVPFLEHYSGYSISELHPLVRQLNKLLTFSSYDSLKAV
3001 YYKYSHPVFFEVAKIPALDMLKLEEILNCDCEAQGLVL*

Color legend (SMART domains):
red: Ankyrin repeats
green: Cyclin-like
The first nucleotide/peptide of CCNB3 is indicated in blue

Supplementary Figure 6. Impact of BCOR-CCNB3 or DeltaCCNB3 on cell cycle phases. Mean +/- standard deviation of triplicates is shown.