World Cancer Research Journal WCRJ 2019; 6: e1321

CATCHING THE DRIVER MUTATIONS IN EWING SARCOMA TUMOURS: AN IN *SILICO* GENOMIC ANALYSIS

E. A. TORAIH[1,2], A. A. TOREIH[3], M. S. FAWZY[4,5]

[1]Department of Histology and Cell Biology, Genetics Unit, Faculty of Medicine, Suez Canal University, Ismailia, Egypt
[2]Center of Excellence of Molecular and Cellular Medicine, Suez Canal University, Ismailia, Egypt
[3]Department of Orthopedics and Traumatology, Faculty of Medicine, Suez Canal University, Ismailia, Egypt
[4]Department of Biochemistry, Faculty of Medicine, Northern Border University, Arar, Saudi Arabia
[5]Department of Medical Biochemistry and Molecular Biology, Faculty of Medicine, Suez Canal University, Ismailia, Egypt

Abstract – *Objective:* *Ewing sarcoma (EWS) is a rare neuroectodermal-related malignancy affecting bones and soft tissues. The well-known hallmark of genomic alteration in EWS is gene fusion involving the Ewing Sarcoma Breakpoint Region 1* (EWSR1) *gene. However, studies have determined that this is not the sole determinant of tumour transformation and indicated the presence of other mutated genes related to signalling pathways and chromatin-modifying genes.*

Materials and Methods: *This is an* in silico *analysis of the previously published genomic sequencing of 218 EWS patients and 11 cell lines.*

Results: *The presence of frequent deleterious mutations in* EWSR1 *(17%); titin, TTN (16%); stromal antigen 2,* STAG2 *(14%); and tumour protein P53,* TP53 *(9%) was determined. An increased prevalence of the co-occurrence of a few mutated driver genes across tumour samples was significantly noted, namely, between* TP53 *and either* EWSR1 *or* STAG2 *and between* TTN *and complement C3b/C4b receptor 1* (CR1) *or zinc finger homeobox 3* (ZFHX3)*, suggesting their joint contribution to EWS tumour development. Patients carrying the* TP53 *aberration alone or combined with* EWSR1 *or* STAG2 *alterations had much lower survival rates. Functional enrichment analysis highlighted transcription factors, kinases, and hub proteins that could be putative therapeutic targets for EWS in the future.*

Conclusions: *The current analysis provides new insights that can be used as a roadmap for future* in vitro *or* in vivo *work. A systems biology approach will be required that takes into account the genomic and epigenomic landscapes of EWS for risk stratification and future molecular targeted therapy.*

KEYWORDS: *Ewing sarcoma,* EWSR1, TP53, STAG2, TTN, *Molecular analysis.*

LIST OF ABBREVIATIONS: *ADGRV1: adhesion G Protein-coupled receptor V1; ARF: ADP ribosylation factors; ATP7B: ATPase copper transporting beta; ATR: ataxia telangiectasia and Rad3-related protein serine/threonine-protein kinase; BSN: Bassoon presynaptic cytomatrix protein; CR1: complement C3b/C4b receptor 1; EPPK1: epiplakin 1; ERBB2: Erb-B2 receptor tyrosine kinase 2; EWS: Ewing sarcoma; EWSR1: Ewing Sarcoma Breakpoint Region 1; FAT2: FAT atypical cadherin 2; FDA: Food and drug administration; IGF-1: insulin-like growth factor I; KMT2D: lysine methyltransferase 2D; MUC6: Mucin 6; OBSCN: obscurin; RCA1: renal carcinoma associated 1; SPEG: SPEG complex locus; SPEN: spen family transcriptional repressor; STAG: Stromal Antigen 2; TCGA: The cancer genome atlas; P53: TP53Tumor Protein; TTN: Titin; ZFHX3: Zinc Finger Homeobox 3.*

Corresponding Author: Eman A. Toraih, MD; e-mail: emantoraih@gmail.com

INTRODUCTION

Ewing sarcoma (EWS) is a primary malignant bone tumor mostly affecting children, adolescents, and young adults that has an aggressive nature, a high propensity for metastasis, and poor outcomes[1,2]. It accounts for 2.9% of all childhood tumours[3]. It mostly arises in bone, but less than 10% of EWS tumors originate in soft tissues[4]. Histopathologically, EWS is a group of small round blue cell tumours[3], with abundant cytoplasmic glycogen and CD99 marker on the plasma membrane[2] and little insight into the cell of origin[4].

Decades of studies highlighted that the primary genomic aberrations in EWS tumors were gene fusions caused by chromosomal translocations involving Ewing Sarcoma Breakpoint Region 1 (EWSR1) at chromosome 22q12[1,5]. Fused proteins were proposed to orchestrate multiple oncoproteins and to regulate critical genes involved in the EWS tumorigenic phenotype. However, other non-fusion mutations have also been reported. Moreover, mutations in p53 itself were present in approximately 10% to 15% of EWS cases, and other alterations in the p53 pathway were noted in additional cases[5]. There is some evidence from *in vitro* and *in vivo* studies on the inability of transformed *EWSR1* alone to induce EWS and the requirement for additional mutations in cooperating pathways for expression of the full oncogenic phenotype.

To address the above issue, we assembled a dataset comprising all EWS patients and cell lines in The Cancer Genome Atlas (TCGA) and comprehensively analyzed their mutational dataset to identify novel genomic mutations associated with EWS. Advancement in the understanding of the molecular pathogenesis and aggressive biology of EWS disease can be translated into better patient outcomes and used to elucidate novel therapeutic strategies to cure EWS patients.

MATERIALS AND METHODS

Cancer datasets were retrieved from the cBioPortal cancer database version 1.12.2 (http://www.cbioportal.org). The mutation data from two studies were selected and downloaded: Pediatric Ewing Sarcoma (DFCI, Cancer Discov 2014) and Ewing Sarcoma (Institut Cuire, Cancer Discov 2014). In the first study by Crompton et al[6], whole-exome sequencing of 96 paediatric EWS tumour specimens and 11 cell lines was carried out. In contrast, the second study conducted by Tirode et al[2] contained data from whole-genome sequencing of 112 EWS samples and matched germ-line DNA.

Mutually exclusive and co-occurrence analysis was carried out. The odds ratio was computed to indicate the likelihood of whether the events in the two genes were mutually exclusive or occurred simultaneously across the selected cases. Hotspot mutation analysis within protein domains was performed. Overall and disease-free survival analysis was evaluated. Pathway and Gene Ontology (GO) analyses of commonly mutated genes in the two studies were performed using Enrichr browser (http://amp.pharm.mssm.edu/Enrichr/). The genomic chromosomal locations of mutated genes were determined. The transcription factor-gene regulatory network was investigated. A protein-protein interaction network was generated using STRING version 10.5 (https://string-db.org/). Food and Drug Administration (FDA)-approved drugs and microRNAs targeting a mutated gene list were identified.

RESULTS AND DISCUSSION

Study characteristics

The baseline features of the target populations in the two studies are summarized in Table S1. A total of 218 EWS patients and 11 cell lines were sequenced. Both mutation data and copy number alterations were analyzed. Patients were clustered according to their age groups, with a higher frequency among adolescents and older children, representing one-third (32.6%) and one-fifth (20.5%) of patients, respectively. This result is consistent with the finding that EWS is the second most common aggressive bone cancer in children and young adults after osteosarcoma[7]. Male patients were more prevalent (53.5%) than female patients (42.3%) as expected[8]. Tumors were most frequently found in the lower limbs (especially the femur) and pelvic region in both studies. Most lesions were localized at the time of diagnosis, with less than 30% of patients having metastasis. Data for ethnicity and chemotherapy intake were only available in the first study (Fig. 1A-C).

Hip bone Ewing sarcoma tumour patients had the lowest survival rate

Overall and disease-free survival analysis revealed a significantly poorer survival in patients with EWS in the hip-bone ($p < 0.001$) than in other locations. However, stratified analysis revealed nonsignificant differences regarding both overall and disease-free survival by gender (Fig. 1D-G).

TABLE S1. Baseline characteristics of the target populations included in the study.

	Crompton et al[6]	*Tirode et al*[2]	*Overall population*
Study population			
No of samples	107 (100)	112 (100)	
Ewing Sarcoma tumor	96 (89.7)	112 (100)	
Cell lines	11 (10.3)		
No of patients	103 (100)	112 (100)	
Genomic analysis			
CNA data	107 (100)	112 (100)	
Mutation data	105 (98.1)	112 (100)	
Transcriptomic data	0 (0.0)	0 (0.0)	
Age at diagnosis			
Infant (<2y)	6 (5.82)	0 (0.0)	
Early children (>2-6y)	9 (8.73)	9 (8.03)	
Late children (>6-12)	17 (16.5)	27 (24.10)	
Adolescent (>12-18)	24 (23.3)	46 (41.07)	
Adult (>19-40)	3 (2.91)	20 (17.8)	
Elder (>40)	0 (0.0)	1 (0.89)	
NA	44 (42.7)	9 (8.03)	
Gender			
Male	54 (52.4)	61 (54.5)	115 (53.5)
Female	49 (47.6)	42 (37.5)	91 (42.3)
NA		9 (8.0)	9 (4.2)
Tumor status at diagnosis			
Localized	36 (35.0)	63 (56.3)	99 (48.3)
Metastasis	10 (9.7)	36 (32.1)	46 (22.4)
NA	47 (42.7)	13 (11.6)	60 (29.3)
Primary tumor site			
Femur, tibia, fibula, thigh, talus, metatarsal	18 (16.82)	36 (33.6)	54 (25.2)
Pelvis, iliac bone, pubic bone, iliac crest, buttock, sacrum, gluteus muscle	17 (15.88)	24 (22.4)	41 (19.2)
Chest wall, rib, clavicle, thorax, mediastinum	15 (14.01)	13 (12.1)	28 (13.1)
Vertebral body, scapula, paraspinal, parasacral, neck	9 (8.41)	13 (12.1)	22 (10.3)
Humerus, radius, shoulder, metacarpal	2 (1.86)	6 (5.61)	8 (3.7)
Scalp, skull, face, orbit	2 (1.86)	5 (5.60)	7 (3.3)
NA	44 (41.12)	10 (9.34)	54 (25.2)
Overall survival			
Living	13 (12.1)	62 (55.4)	75 (34.9)
Deceased	11 (10.2)	41 (36.6)	52 (24.2)
NA	79 (73.8)	9 (8.0)	88 (40.9)
Disease-free status			
Disease free	–	54 (48.2)	
Recurred / progress	–	45 (40.2)	
NA	–	13 (11.6)	
Ethnicity category			
White/Europe	20 (19.4)	–	
White/Latin America	2 (1.9)	–	
White/North Africa	1 (1.0)	–	
Black/Sub-Saharan Africa	1 (1.0)	–	
NA	79 (76.7)	–	
Chemotherapy first line			
VAC/IE	23 922.3)	–	
GD/IT	1 (1.0)	–	
NA	79 (76.7)	–	
Mutated genes			
No of mutated genes	2074	203	
No of common genes between two studies			88

Crompton study included 96 pediatric EWS tumor specimens of 92 patients and 11 cell lines while that of Tirode enclosed 112 EWS samples. Combination chemotherapy using *vincristine, adriamycin, cyclophosphamide* (VAC) alternating with *ifosfamide* and *etoposide* (IE). Gemcitabine and docetaxel / irinotecan and temozolomide (IT). NA, not applicable.

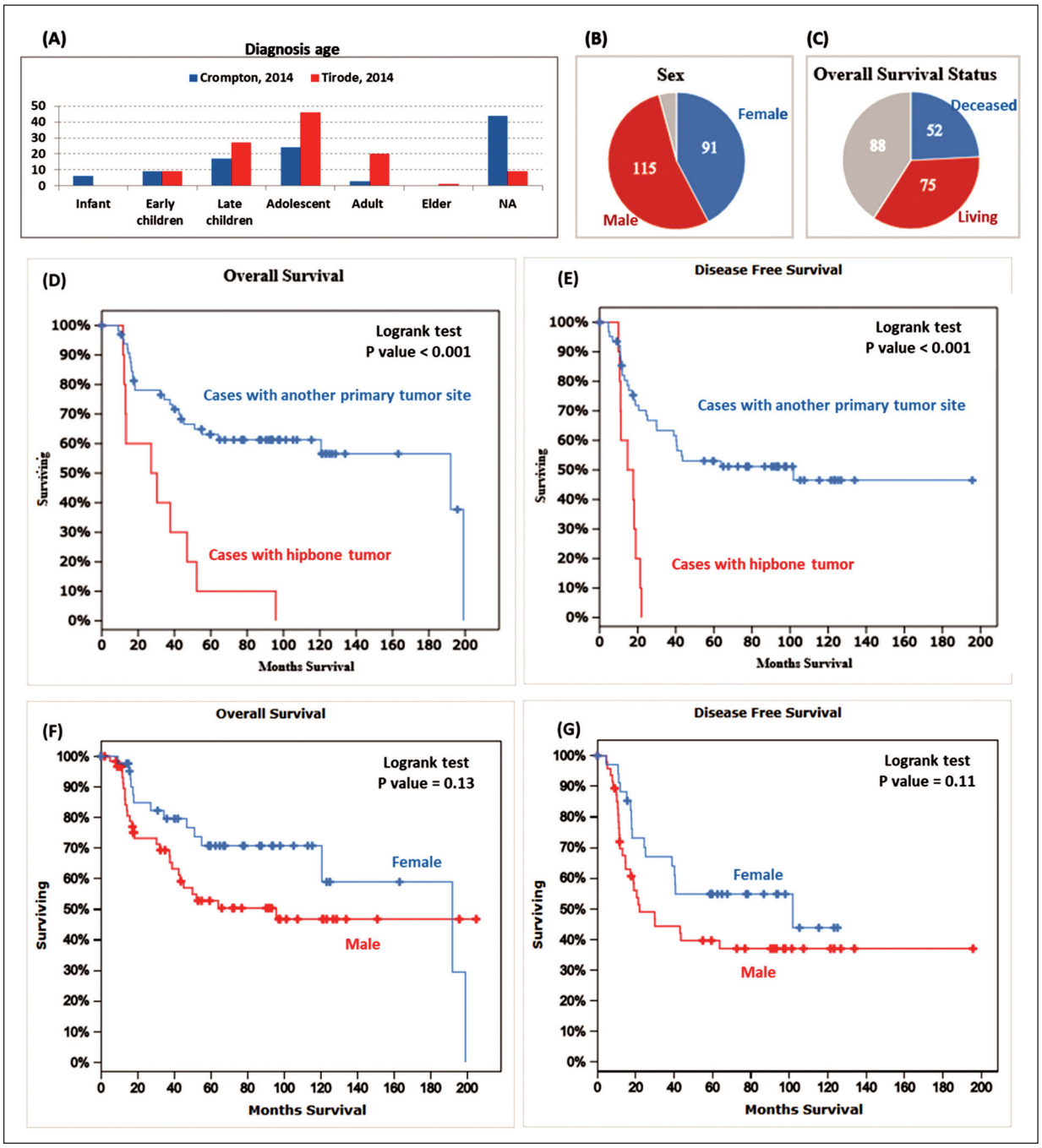

Fig. 1. Summary of the study characteristics of Ewing sarcoma patients. ***A***, Age at diagnosis. ***B***, Gender. ***C***, Vital status. ***D***, Overall survival stratified by primary tumour site. ***E***, Disease-free survival stratified by primary tumour site. ***F***, Overall survival stratified by gender. ***G***, Disease-free survival stratified by gender.

Gene mutation data analysis

In the Crompton et al[6] and Tirode et al[2] studies, the whole-sequenced EWS samples showed alterations in 2074 and 203 genes at variable frequencies, respectively (Fig. 2A). Gene mutations were encountered in a total of 94 (44%) patients in both studies. The most frequently mutated genes in both studies are depicted in Tables S2 and S3. Collectively, 88 genes (4%) were commonly altered (Table S4). Most notably, the most highly significant and prevalent genes mutated in both male and female patients were *EWSR1*, titin (*TTN*), stromal antigen 2 (*STAG2*), and tumor protein 53 (*TP53*), representing an overall prevalence of 17%, 16%, 14%, and 9% of patients, respectively (Fig. 2B-D). Alterations of other genes, such as lysine methyltransferase 2D (*KMT2D*), mucin 6 (*MUC6*), adhesion G protein-coupled receptor V1 (*ADGRV1*), obscurin (*OBSCN*), complement C3b/C4b receptor 1 (*CR1*), and ATPase copper

TABLE S2. Top highly frequent genes altered in the study of Crompton et al[6].

Gene	#Mut	Freq	Cytoband	Gene size (nt)	MutSig Q value
TTN	39	28.57%	2q31.2	281435	1.39E-04
EWSR1	38	36.19%	22q12.2	32517	1.17E-03
STAG2	15	11.43%	Xq25	142097	1.06E-04
MUC6	14	11.43%	11p15.5	1036426	1.35E-05
TP53	13	12.38%	17p13.1	19149	6.79E-04
ADGRV1	12	10.48%	5q14.3	605546	1.98E-05
KMT2D	12	11.43%	12q13.12	41910	2.86E-04
EPPK1	11	9.52%	8q24.3	21725	5.06E-04
DNAH1	10	8.57%	3p21.1	87488	1.14E-04
AHNAK	9	7.62%	11q12.3	113319	7.94E-05
AHNAK2	9	8.57%	14q32.33	41105	2.19E-04
ZFHX3	9	7.62%	16q22.2-q22.3	275751	3.26E-05
BSN	8	7.62%	3p21.31	117061	6.83E-05
NEB	8	5.71%	2q23.3	249154	3.21E-05
NPHP4	8	7.62%	1p36.31	129666	6.17E-05
OBSCN	8	7.62%	1q42.13	173419	4.61E-05
PCDH15	8	7.62%	10q21.1	1826412	4.38E-06
SPEN	8	6.67%	1p36.21-p36.13	92592	8.64E-05
THBS4	8	7.62%	5q14.1	91992	8.70E-05
ATP7B	7	6.67%	13q14.3	79462	8.81E-05
CR1	7	6.67%	1q32.2	145638	4.81E-05
DSP	7	6.67%	6p24.3	45143	1.55E-04
FLG	7	6.67%	1q21.3	23029	3.04E-04
HERC2	7	5.71%	15q13.1	28077673	2.49E-07
HRNR	7	6.67%	1q21.3	12121	5.78E-04
PKHD1	7	5.71%	6p12.3-p12.2	474335	1.48E-05
RPTN	7	6.67%	1q21.3	5634	1.24E-03
SPTA1	7	6.67%	1q23.1	76219	9.18E-05
SYNE1	7	6.67%	6q25.2	515716	1.36E-05
VWF	7	5.71%	12p13.31	175802	3.98E-05
ABCA13	6	5.71%	7p12.3	476074	1.26E-05
ANK3	6	5.71%	10q21.2	707229	8.48E-06
COL18A1	6	5.71%	21q22.3	108584	5.53E-05
COL6A6	6	4.76%	3q22.1	161427	3.72E-05
COL7A1	6	5.71%	3p21.31	31230	1.92E-04
DNAH8	6	5.71%	6p21.2	315470	1.90E-05
DNAH9	6	5.71%	17p12	371738	1.61E-05
DYNC2H1	6	4.76%	11q22.3	370433	1.62E-05
GPR179	6	4.76%	17q12	35983253	1.67E-07
IGSF10	6	5.71%	3q25.1	184876	3.25E-05
LAMA2	6	5.71%	6q22.33	633426	9.47E-06
LLGL2	6	5.71%	17q25.1	50130	1.20E-04
PCDHGA5	6	4.76%	5q31.3	148649	4.04E-05
PDZD2	6	4.76%	5p13.3	471545	1.27E-05
PLEC	6	3.81%	8q24.3	61654	9.73E-05
PRAMEF12	6	5.71%	1p36.21	3059	1.96E-03
PRB2	6	4.76%	12p13.2	4025	1.49E-03
SORL1	6	5.71%	11q24.1	181560	3.31E-05
TCHH	6	4.76%	1q21.3	9138	6.57E-04
TNKS1BP1	6	4.76%	11q12.1	25324	2.37E-04
TULP4	6	5.71%	6q25.3	279637	2.15E-05
XIRP2	6	5.71%	2q24.3	371267	1.62E-05
ABCA1	5	4.76%	9q31.1	147245	3.40E-05
ABCC4	5	4.76%	13q32.1	281722	1.78E-05
ATM	5	4.76%	11q22.3	146619	3.41E-05

TABLE S3. Top highly frequent genes altered in the study of Tirode et al[2].

Gene	#Mut	Freq	Cytoband	Gene size (nt)	MutSig Q value
STAG2	18	16.07%	Xq25	142097	1.27E-04
TP53	8	7.14%	17p13.1	19149	4.18E-04
AP1B1	6	4.46%	22q12.2	60904	9.85E-05
CSMD1	5	4.46%	8p23.2	2059683	2.43E-06
TTN	5	2.68%	2q31.2	281435	1.78E-05
DLGAP2	3	2.68%	8p23.3	1668353	1.80E-06
DMBT1	3	2.68%	10q26.13	83072	3.61E-05
DPP10	3	2.68%	2q14.1	1403395	2.14E-06
EZH2	3	2.68%	7q36.1	77291	3.88E-05
MACF1	3	2.68%	1p34.3	402972	7.45E-06
NUP155	3	2.68%	5p13.2	79823	3.76E-05
PGR	3	2.68%	11q22.1	101058	2.97E-05
PRUNE2	3	2.68%	9q21.2	294845	1.02E-05
RYR2	3	1.79%	1q43	791781	3.79E-06
SEC14L5	3	1.79%	16p13.3	60839	4.93E-05
TMPRSS11B	3	1.79%	4q13.2	19042	1.58E-04
ABI2	2	1.79%	2q33.2	104300	1.92E-05
ACSM5	2	1.79%	16p12.3	31826	6.28E-05
ADAMTS6	2	1.79%	5q12.3	333279	6.00E-06
ADHFE1	2	1.79%	8q13.1	36327	5.51E-05
AFF1	2	1.79%	4q21.3-q22.1	206053	9.71E-06
AGBL1	2	1.79%	15q25.3	951842	2.10E-06
AIFM2	2	1.79%	10q22.1	20668	9.68E-05
ANKHD1	2	1.79%	5q31.3	138043	1.45E-05
AQR	2	1.79%	15q14	113444	1.76E-05
ARMT1	2	1.79%	6q25.1	17844	1.12E-04
BCOR	2	1.79%	Xp11.4	126145	1.59E-05
BHMT	2	1.79%	5q14.1	20510	9.75E-05
BIRC6	2	1.79%	2p22.3	261993	7.63E-06
CDC27	2	1.79%	17q21.32	71636	2.79E-05
CNGB1	2	1.79%	16q21	88777	2.25E-05
COL10A1	2	1.79%	6q22.1	99790	2.00E-05
CR1	2	1.79%	1q32.2	145638	1.37E-05
CREBRF	2	1.79%	5q35.1	82940	2.41E-05
CSMD3	2	1.79%	8q23.3	1214572	1.65E-06
CXCL8	2	1.79%	4q13.3	3211	6.23E-04
DCHS1	2	1.79%	11p15.4	34526	5.79E-05
DIEXF	2	1.79%	1q32.2	29599	6.76E-05
DIRAS1	2	1.79%	19p13.3	6826	2.93E-04
DSC3	2	1.79%	18q12.1	53451	3.74E-05
DSP	2	1.79%	6p24.3	45143	4.43E-05
EPB41L5	2	1.79%	2q14.2	166094	1.20E-05
EPX	2	1.79%	17q22	12451	1.61E-04
EWSR1	2	1.79%	22q12.2	32517	6.15E-05
FAT2	2	1.79%	5q33.1	90731	2.20E-05
GALNT17	2	1.79%	7q11.22	581188	3.44E-06
GMPR	2	1.79%	6p22.3	56970	3.51E-05
GRM4	2	1.79%	6p21.31	136980	1.46E-05
HCN1	2	1.79%	5p12	441169	4.53E-06
HDC	2	1.79%	15q21.2	24105	8.30E-05
HSD3B2	2	16.07%	1p12	8109	2.47E-04
HTT	2	7.14%	4p16.3	169451	1.18E-05
IRX1	2	4.46%	5p15.33	5549	3.60E-04
ISLR2	2	4.46%	15q24.1	13624	1.47E-04
KCNG2	2	2.68%	18q23	103932	1.92E-05

TABLE S4. Common altered genes in Crompton et al[6] and Tirode et al[2] studies.

Genes	Cytoband	Crompton et al[6]		Tirode et al[2]		Total Frequency
		Frequency	p-value	Frequency	p-value	
TTN	2q31.2	39	1.39E-04	5	1.78E-05	44
EWSR1	22q12.2	38	1.17E-03	2	6.15E-05	40
STAG2	Xq25	15	1.06E-04	18	1.27E-04	33
TP53	17p13.1	13	6.79E-04	8	4.18E-04	21
KMT2D	12q13.12	12	2.86E-04	2	4.77E-05	14
EPPK1	8q24.3	11	5.06E-04	1	4.60E-05	12
OBSCN	1q42.13	8	4.61E-05	2	1.15E-05	10
ZFHX3	16q22.2-q22.3	9	3.26E-05	1	3.63E-06	10
CR1	1q32.2	7	4.81E-05	2	1.37E-05	9
CSMD1	8p23.2	4	1.94E-06	5	2.43E-06	9
DSP	6p24.3	7	1.55E-04	2	4.43E-05	9
MACF1	1p34.3	5	1.24E-05	3	7.45E-06	8
SYNE1	6q25.2	7	1.36E-05	1	1.94E-06	8
FAT2	5q33.1	5	5.51E-05	2	2.20E-05	7
PLEC	8q24.3	6	9.73E-05	1	1.62E-05	7
PRUNE2	9q21.2	4	1.36E-05	3	1.02E-05	7
SPEG	2q35	5	8.42E-05	2	3.37E-05	7
CNGB1	16q21	4	4.51E-05	2	2.25E-05	6
DLGAP2	8p23.3	3	1.80E-06	3	1.80E-06	6
FAT1	4q35.2	5	3.59E-05	1	7.19E-06	6
HTT	4p16.3	4	2.36E-05	2	1.18E-05	6
MYH15	3q13.13	4	2.37E-05	2	1.18E-05	6
PCLO	7q21.11	5	1.22E-05	1	2.44E-06	6
PKHD1L1	8q23.1-q23.2	4	2.29E-05	2	1.15E-05	6
RELN	7q22.1	5	9.66E-06	1	1.93E-06	6
AFF1	4q21.3-q22.1	3	1.46E-05	2	9.71E-06	5
BCOR	Xp11.4	3	2.38E-05	2	1.59E-05	5
DMBT1	10q26.13	2	2.41E-05	3	3.61E-05	5
LRP1B	2q22.1-q22.2	4	2.10E-06	1	5.26E-07	5
LRRK2	12q12	4	2.77E-05	1	6.93E-06	5
PRDM9	5p14.2	3	1.43E-04	2	9.53E-05	5
USP6	17p13.2	4	6.83E-05	1	1.71E-05	5
ACSM5	16p12.3	2	6.28E-05	2	6.28E-05	4
ANKHD1	5q31.3	2	1.45E-05	2	1.45E-05	4
ASXL3	18q12.1	3	1.74E-05	1	5.79E-06	4
ATRX	Xq21.1	3	1.07E-05	1	3.55E-06	4
BIRC6	2p22.3	2	7.63E-06	2	7.63E-06	4
CSMD3	8q23.3	2	1.65E-06	2	1.65E-06	4
DCHS1	11p15.4	2	5.79E-05	2	5.79E-05	4
ERF	19q13.2	3	3.95E-04	1	1.32E-04	4
EZH2	7q36.1	1	1.29E-05	3	3.88E-05	4
MGA	15q15.1	3	2.02E-05	1	6.72E-06	4
MUTYH	1p34.1	3	2.57E-04	1	8.56E-05	4
PGR	11q22.1	1	9.90E-06	3	2.97E-05	4
PHIP	6q14.1	2	1.39E-05	2	1.39E-05	4
PRDM2	1p36.21	3	2.38E-05	1	7.92E-06	4
RAD51AP2	2p24.2	2	1.01E-04	2	1.01E-04	4
RRP15	1q41	2	3.79E-05	2	3.79E-05	4
ZNF436	1p36.12	2	8.61E-08	2	8.61E-08	4
EPHA4	2q36.1	2	1.28E-05	1	6.40E-06	3
EPHA7	6q16.1	2	1.11E-05	1	5.57E-06	3
ERBB4	2q34	2	1.72E-06	1	8.60E-07	3

Continued

TABLE S4 (CONTINUED). Common altered genes in Crompton et al[6] and Tirode et al[2] studies.

Genes	Cytoband	Crompton et al[6]		Tirode et al[2]		Total Frequency
		Frequency	p-value	Frequency	p-value	
IRS1	2q36.3	2	2.96E-05	1	1.48E-05	3
KIAA1109	4q27	2	9.50E-06	1	4.75E-06	3
KMT2C	7q36.1	1	3.31E-06	2	6.63E-06	3
MAML2	11q21	2	5.46E-06	1	2.73E-06	3
PDGFRB	5q32	2	4.76E-05	1	2.38E-05	3
PIK3CA	3q26.32	2	2.17E-05	1	1.09E-05	3
PML	15q24.1	2	3.76E-05	1	1.88E-05	3
POLE	12q24.33	2	3.14E-05	1	1.57E-05	3
POLQ	3q13.33	2	1.75E-05	1	8.73E-06	3
SETD2	3p21.31	1	6.77E-06	2	1.35E-05	3
AR	Xq12	1	5.36E-06	1	5.36E-06	2
ARFRP1	20q13.33	1	1.07E-04	1	1.07E-04	2
BAP1	3p21.1	1	1.10E-04	1	1.10E-04	2
BCL11A	2p16.1	1	9.77E-06	1	9.77E-06	2
CCDC6	10q21.2	1	8.48E-06	1	8.48E-06	2
CDKN2A	9p21.3	1	3.66E-05	1	3.66E-05	2
COL11A1	1p21.1	1	4.31E-06	1	4.31E-06	2
ELF4	Xq26.1	1	2.18E-05	1	2.18E-05	2
EPHA2	1p36.13	1	3.15E-05	1	3.15E-05	2
FANCL	2p16.1	1	1.22E-05	1	1.22E-05	2
FGF19	11q13.3	1	1.64E-04	1	1.64E-04	2
GRIN2A	16p13.2	1	2.33E-06	1	2.33E-06	2
IKBKB	8p11.21	1	1.63E-05	1	1.63E-05	2
JARID2	6p22.3	1	3.62E-06	1	3.62E-06	2
MED12	Xq13.1	1	4.18E-05	1	4.18E-05	2
MEF2A	15q26.3	1	6.62E-06	1	6.62E-06	2
MNX1	7q36.3	1	1.72E-04	1	1.72E-04	2
PIK3R1	5q13.1	1	1.16E-05	1	1.16E-05	2
RUNX1	21q22.12	1	3.82E-06	1	3.82E-06	2
SMC1A	Xp11.22	1	2.06E-05	1	2.06E-05	2
TET1	10q21.3	1	7.12E-06	1	7.12E-06	2
TET2	4q24	1	7.47E-06	1	7.47E-06	2
TLL2	10q24.1	1	6.70E-06	1	6.70E-06	2
TLR4	9q33.1	1	7.51E-05	1	7.51E-05	2
TSHZ3	19q12	1	1.35E-05	1	1.35E-05	2
WDR90	16p13.3	1	5.41E-05	1	5.41E-05	2

transporting beta (*ATP7B*) in males and bassoon presynaptic cytomatrix protein (*BSN*), epiplakin 1 (*EPPK1*), zinc finger homeobox 3 transcriptional regulator (*ZFHX3*), spen family transcriptional repressor (*SPEN*), SPEG complex locus (*SPEG*), and FAT atypical cadherin 2 (*FAT2*) in females, showed gender-specific preference.

The results of the top 11 genes observed in EWS patients showed that 53.6% of cases (out of 217 sequenced samples) had an alteration in at least one of the highly frequently mutated genes. In total, 36.9% were mutations, 9.2% were fusion events, and 7.4% were other multiple alterations (Fig. 2E-F). For *EWSR1*, most of the alterations were fusion events, and a few of them were missense mutations. Events associated with the *TTN* gene were mostly in-frame mutations and missense variants with unknown significance. The alterations of *STAG2* were truncating mutations, fusion events, and missense mutations with unknown significance. Patients with the *TP53* mutant exhibited putative driver missense mutations.

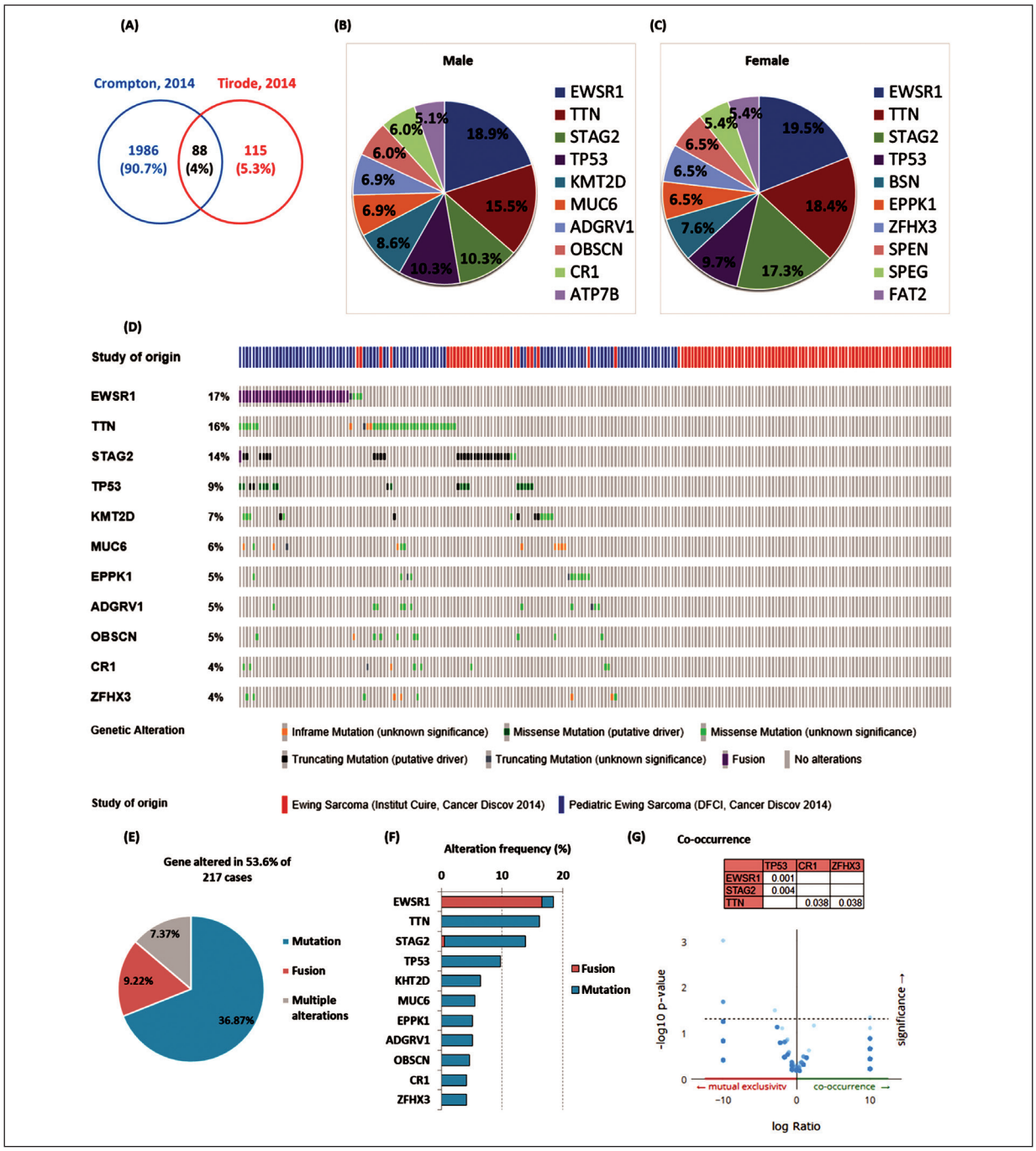

Fig. 2. Mutational analysis of Ewing sarcoma patients. ***A,*** Total number of altered genes with 88 genes commonly mutated in both studies. In Crompton et al[6] and Tirode et al[2] studies, whole-sequenced Ewing sarcoma samples showed alterations in 2074 and 203 genes at variable frequencies, respectively. ***B,*** Most frequently altered genes in male patients. *EWSR1*, *TTN*, *STAG2*, and *TP53* represented more than half (53.3%) of the gene alterations in males. ***C,*** Most frequently altered genes in female patients. *EWSR1*, *TTN, STAG2*, and *TP53* represented two-thirds (64.9%) of gene alterations in males. ***D,*** Graphical summary of genomic alterations in multiple genes across Ewing sarcoma samples. Rows represent genes, and columns represent samples. Colour coding is used to summarize the distinct type of genomic alterations. For *EWSR1*, most of the alterations were fusion, and a few of them were missense mutations. Events associated with the *TTN* gene were mostly in-frame mutations and missense variants with unknown significance. The alterations of *STAG2* were truncating mutations, fusion events, and missense mutations with unknown significance. Patients with the *TP53* mutant exhibited putative driver missense mutations. ***E,*** Types of alterations in the top 11 frequently altered genes in Ewing sarcoma patients. In total, 36.9% were mutations, 9.2% were fusion events, and 7.4% were other multiple alterations. ***F,*** Type of alteration in each gene. ***G,*** Co-occurrence and mutually exclusive analysis. The *p*-values were determined by Fisher's exact test with the null hypothesis, where the frequency of occurrence of a pair of alterations in two genes is proportional to their uncorrelated occurrence in each gene. None of the genes showed a mutually exclusive pattern. However, some genes showed tendency towards co-occurrence, namely, between *TP53* and both *EWSR1* and *STAG2*, as well as between *TTN* and *ZFHX3*.

Mutually exclusive and co-occurrence alteration patterns for identifying driver mutations

One of the main findings encountered in our results was the co-occurrence of a few driver-mutated genes across tumor samples, namely, between *EWSR1* and *TP53* ($p = 0.001$) in 28 male and 20 female patients, between *STAG2* and *TP53* ($p = 0.004$) in 18 male patients, and between *TTN* and *CR1* ($p = 0.038$) and *TTN* and *ZFHX3* ($p = 0.038$) in 20 males and 18 females, suggesting that they may simultaneously contribute to tumor development (Fig. 2G). No significant mutually exclusive gene mutation pattern was observed across tumor samples.

TP53 mutation is a poor survival indicator

EWS tumors with *TP53* aberration alone or combined with *EWSR1* or *STAG2* alterations had a worse prognosis than tumors with wild-type *TP53*. Apart from patients with *TP53* mutational burden, survival analysis of EWS patients with the other top significant mutant genes did not indicate significant differences from cases without their alterations (Fig. 3). This finding was in line with other previously described studies that report the development of sarcomas in *EWSR1-FLI1* transgenic mice upon homozygous deletion of *TP53* and confirm the required cell cycle progression role of p53 in sarcomagenesis[9-11].

Hotspot mutations in commonly altered genes

EWSR1 exhibited 40 types of mutations: 36 fusion alterations (*EWS-FLI1* and *EWS-ERG*), 3 missense mutations (T22I, G284A, and S259N), and one nonsense mutation (G586*). Point mutations in the *TTN* gene occurred at a higher frequency in EWS patients. Various missense mutations (E12125D, K25371N, R18318C, K19231I, V27254M, I24382T, C7655Y, K13566Q, A25090P, R9576I) and in-frame deletions (G25361_I25368del) were observed in the studied patients. On the other hand, a total of 33 mutations were detected in *STAG2*, including one fusion mutation "*STAG2-MAP7D3*", two missense variants "L865R and R422Q" and 30 truncating mutations. Consistent with other types of tumours[12], the majority of *TP53* mutations detected in EWS tumors were located within the DNA-binding domain. Twenty-one mutations were observed, including 17 missense mutations and truncating mutations (premature stop codon, splice aberrations, and frame shift insertion) (Fig. 4).

Structural genomic and proteomic analysis of the top mutated genes

The *EWSR1* gene (at 22q12.2) spans 32.5 kb on the plus strand and forms a transcript of 2654 bp from 7 exons for the canonical form, with a coding sequence of 1971 nucleotides (nt). According to Ensembl, there are 25 transcripts, of which 16 different transcripts code for proteins. *EWSR1* encodes EWSR1, a ubiquitously expressed protein (636 amino acids) that includes several domains, a transactivation domain (TAD) containing multiple degenerate hexapeptide repeats (1-285) (arginine/glycine/proline rich), 3 arginine/glycine rich domains (RG1: aa 300-340, RG2: 454-513, and RG3: aa 559-640), an RNA recognition motif (RRM, aa 300-340), and a RanBP2 type zinc finger motif (aa 518-549). The expression of EWSR1 is affected by the methylation status of its RNA-binding domain. It is found mainly in the nucleus but also in the cytoplasm and plasma membrane (Fig. S1). Localization of EWSR1 in different subcellular compartments reflects a dynamic distribution during the cell cycle, with predominant nuclear localization in interphase cells, perichromosomal localization in prometaphase cells, and cytoplasmic localization in metaphase cells, and association with microtubules in quiescent cells. EWSR1 is a multifunctional protein that is involved in various cellular processes, including gene expression, cell signaling, DNA damage response and RNA processing and transport[13]. Chromosomal translocations forming EWS-chimeric proteins are usually frequent and are involved in tumorigenesis. The t(11;22)(q24;q12) *EWSR1/FLI1* translocation is commonly found in 85% of cases of Ewing tumors, whereas the t(21;22)(q21;q12) translocation with *EWSR1/ERG* is second in frequency and is found in approximately 10% of cases. Other translocations might be partners with ETS family members, such as *EWSR1/FEV*, *EWSR1/ETV1*, *EWSR1/ETV4*, *EWSR1/NFATC2*, and *EWSR1/PATZ1* (http://atlasgeneticsoncology.org) (Fig. S2). Improper expression of fused proteins, such as *EWS-FLI1*, is cytotoxic, especially when combined with *TP53* mutations or any other components of the p53 pathway that allow for stable expression and the growth and survival of tumorigenic cells[5].

The *TTN* gene (at chromosome 2q31.2) is also known as connectin. It contains the largest number of exons (363) discovered in any single gene and the longest single exon (17,106 bp) (https://www.ncbi.nlm.nih.gov/gene/7273). It encodes the TTN protein (34350 aa), the largest known protein, which is composed of 244 individually folded protein domains connected by unstructured peptide sequences: type I fibronectin type III domain (FN; 132 copies) and type II immunoglobulin domain (Ig; 112 copies).

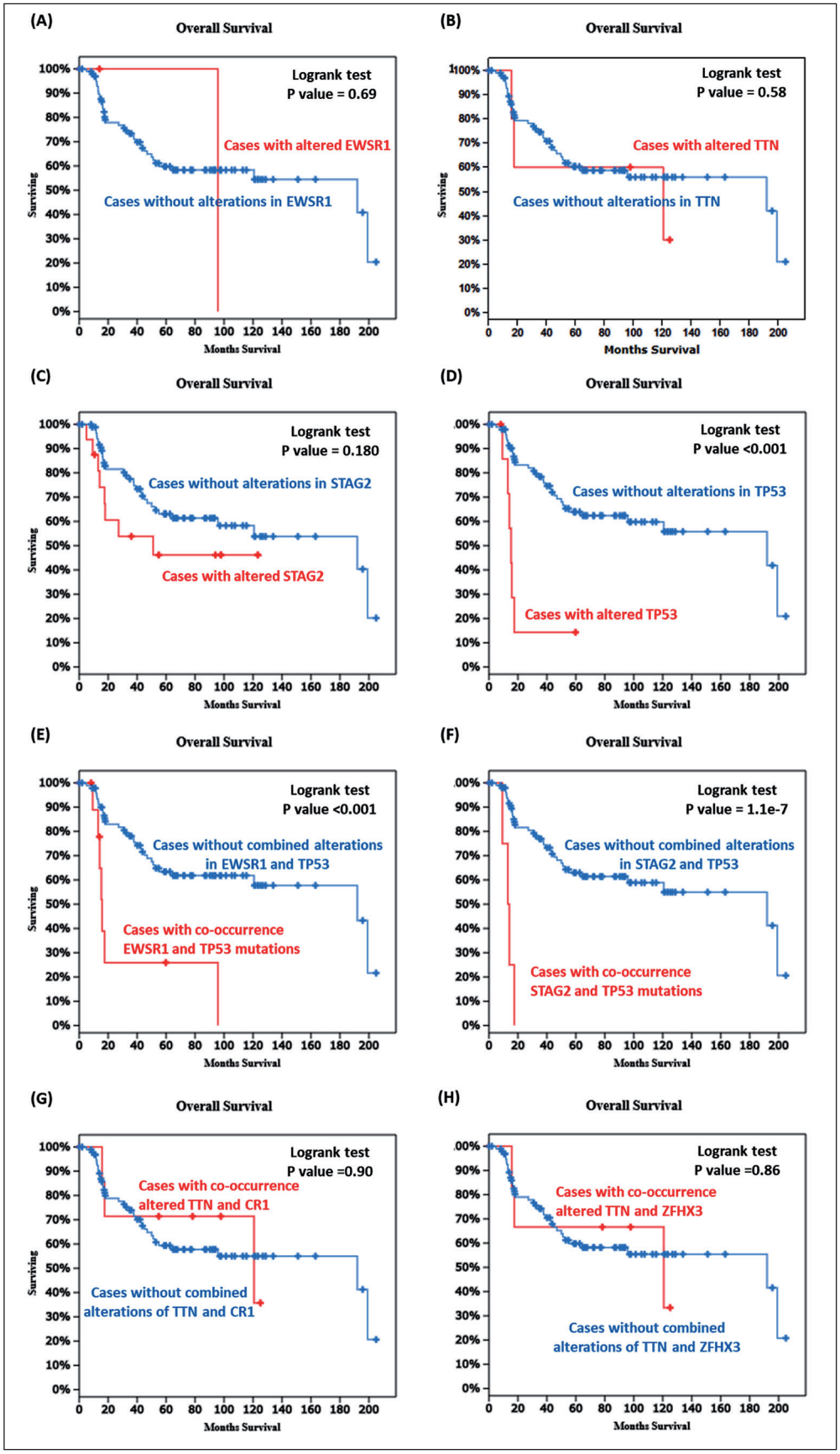

Fig. 3. Disease-specific survival plots. Overall survival according to *(**A**) EWSR1*, *(**B**) TTN*, *(**C**) STAG2*, *(**D**) TP53*, *(**E**) EWSR1* and *TP53*, *(**F**) STAG2* and *TP53*, *(**G**) TTN* and *CR1*, *(**H**) TTN* and *ZFHX3* gene alterations. The results are displayed as Kaplan-Meier plots with *p*-values from a log-rank test.

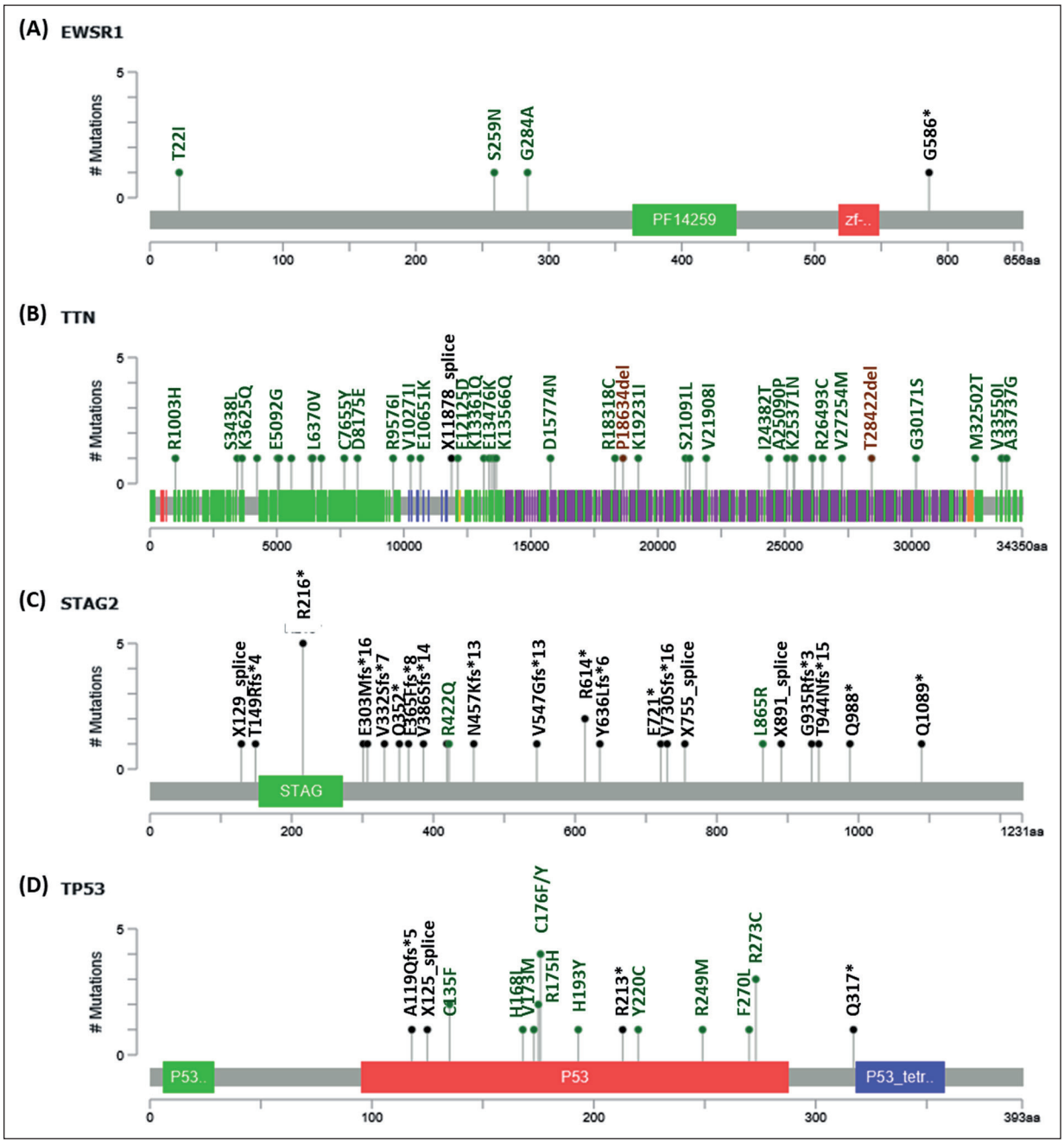

Fig. 4. Distribution of hotspot sites over the commonly mutated genes in Ewing sarcoma tumours. *(**A**) EWSR1*, *(**B**) TTN*, *(**C**) STAG2*, and *(**D**) TP53*. The graphical summary shows the position and frequency of all mutations in the context of Pfam protein domains encoded by the canonical gene isoform. The length of the mutation indicates a higher somatic mutation frequency of the alteration. Mutation diagram circles are coloured with respect to the corresponding mutation types. The most recurrent mutations are labelled in the graphical view. In the case of different mutation types at a single position, the colour of the circle is determined with respect to the most frequent mutation type. Colour codes are as follows: green is for missense mutations, and black is for truncating mutations (nonsense, nonstop, frameshift deletion, frameshift insertion, and splice site), while brown dots are for in-frame mutations (deletions or insertions). Fusion events are not shown in the figure. Protein domain residues are shown below each figure. For *EWSR1*, the two most frequent mutations were *EWS-FLI1* and *G586**. Regarding *TTN*, 6 missense mutations were the most common (E12125D, K25371N, R18318C, K19231I, V27254M, I24382T). In the *STAG2* gene, the N457Kfs*13 frameshift deletion was highly frequent. The two most common *TP53* mutations were the Y220C missense mutation and the R213* nonsense mutation in the DNA-binding domain.

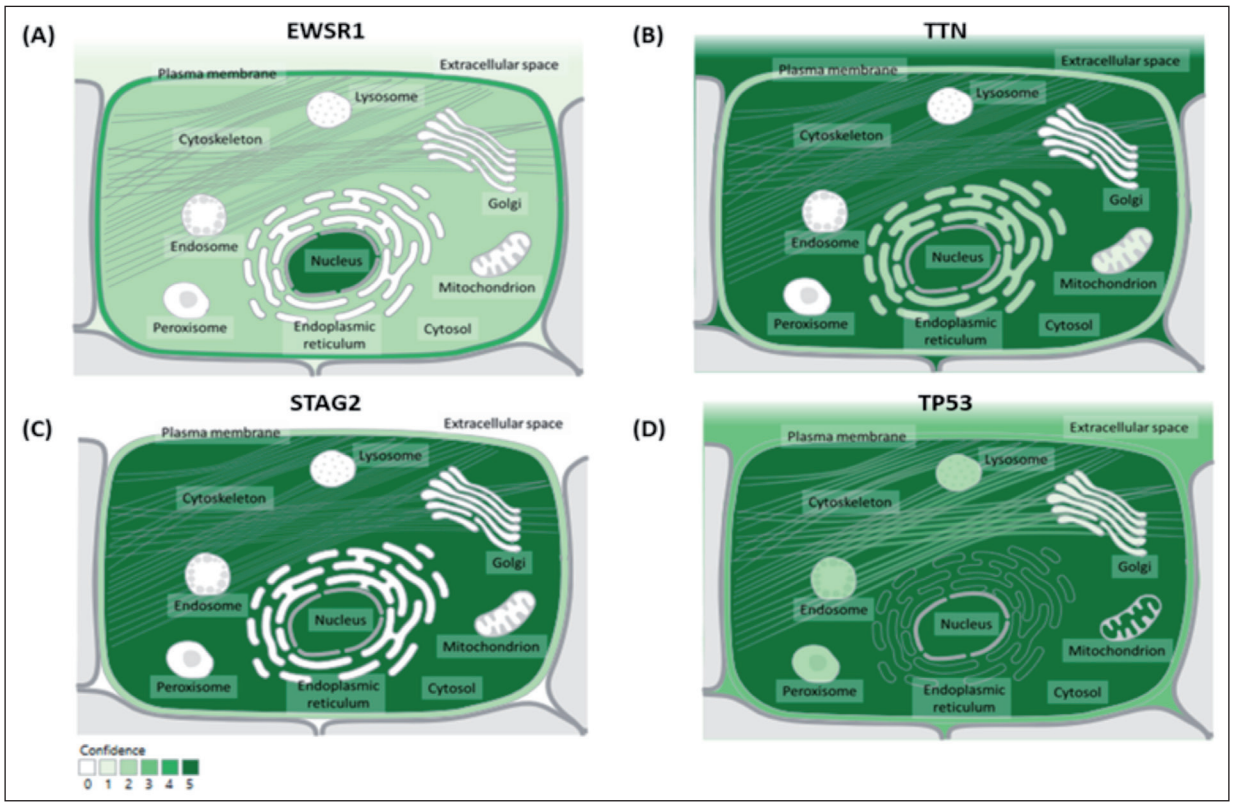

Fig. S1. Subcellular localization of top mutated genes. Ewing Sarcoma Breakpoint Region 1 (EWSR1), titin (TTN), stromal antigen 2 (STAG2), and tumor protein 53 (TP53) [Data source: Compartment database].

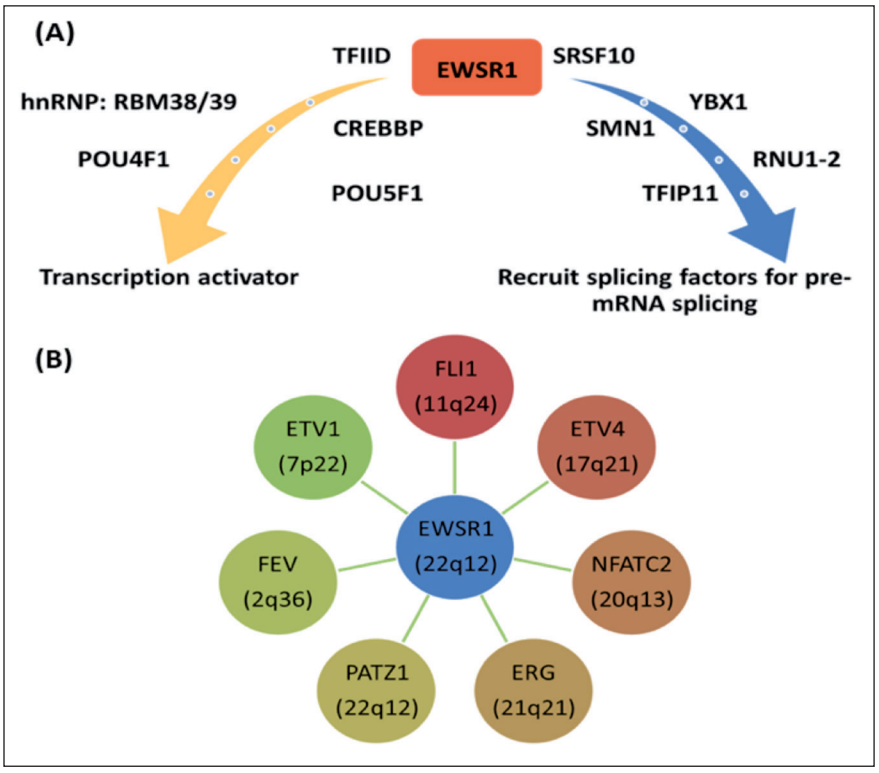

Fig. S2. Mechanism of action of EWSR1 gene. ***A***, Role of EWSR1 as a transcription activator and mRNA splicing. ***B***, Fusion events involving EWSR1 in Ewing sarcoma tumor.

The TTN protein has been identified as a structural protein for chromosomes. It is known to play a role in chromosome condensation and segregation during mitosis and might link the lamina network to chromatin or nuclear actin, or both, during interphase. In addition, it contains binding sites for muscle-associated proteins; therefore, it serves as an adhesion template for the assembly of contractile machinery in muscle cells. It takes part in focal adhesion, the actin cytoskeleton, and calcium signaling pathways (https://www.uniprot.org/uniprot/Q8WZ42).

The *STAG2* gene (at Xq25) consists of 35 exons spanning 142 kb. Transcription of *STAG2* generates 19 different mRNA transcripts; the longest is 6045 bp and contains 34 exons. It is expressed in most tissues and is under-regulated in EWS. It forms a protein subunit in the cohesin complex (SMC1, SMC3, and SCC1), which regulates the separation of sister chromatids during cell division and orchestrates gene expression of transcriptional repressor CTCF (http://atlasgeneticsoncology.org/Genes/GC_STAG2.html). Genetic disruption of cohesin is believed to cause aneuploidy in human cancer[14]. In addition, *STAG2* mutations were previously reported in some solid and hematological tumours[15-17].

The guardian of the genome, the *TP53* gene (at 17p13.1), exhibited the fourth most common alterations. It encodes the p53 protein, which regulates the expression of many target genes with diverse cellular functions in response to stresses. Multiple hotspots were identified in *TP53* genes and were linked to the susceptibility and prognosis of cancer[14].

Protein-protein interaction network

STRING analysis of the top 25 highly frequently mutated genes and clustering revealed few connected nodes with 3 molecular functions and 11 cellular component GO terms that were significantly enriched (Fig. 5).

Functional pathway enrichment analysis

To gain insight into the genetic landscape of EWS tumors, we performed a comprehensive analysis by Enrichr to explore functional annotations of all the mutated genes observed in EWS patients and cell lines. The top KEGG signaling pathways enriched by the gene list, are shown in Table S5. Some of these pathways were the phosphoinositide 3-kinase protein kinase B (PI3K-Akt) signaling pathway, pathways in cancer, focal adhesion, signaling pathways regulating the pluripotency of stem cells, and osteoclast differentiation.

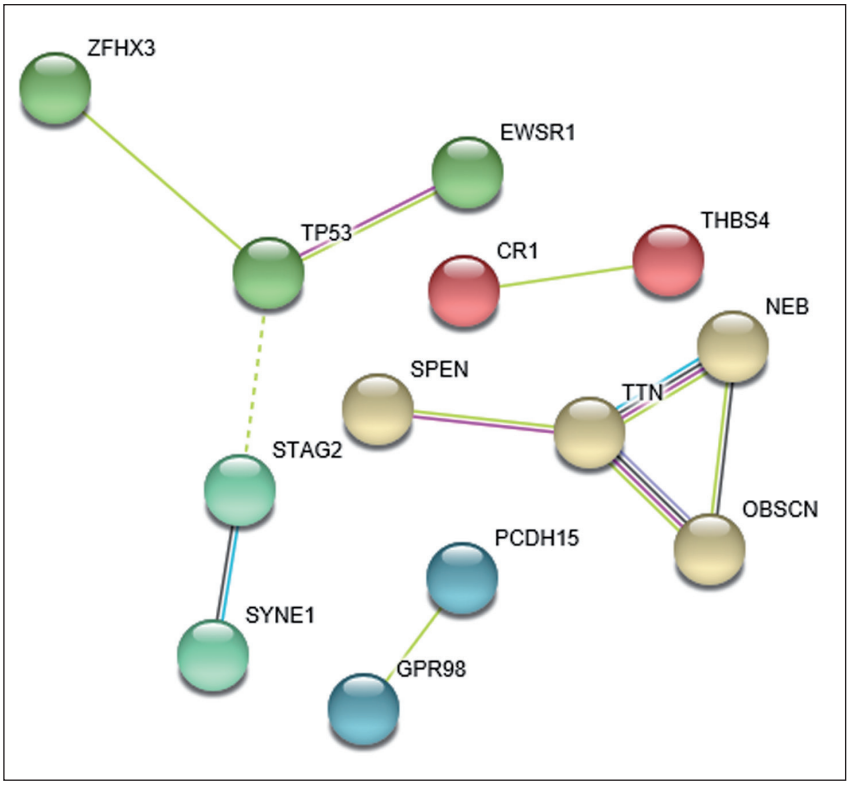

Fig. 5. Protein-protein interaction network of the top 25 highly frequent mutated genes. Physical interactions and functional associations of proteins were investigated using STRING version 10.5. Functional annotation clustering was performed. Proteins are represented as nodes, and the type of association is colour-coded at the edges (pink: experimentally determined, blue: from curated databases, green: text mining). *CR1*: complement C3b/C4b receptor 1 (Knops blood group), *EWSR1*: EWS RNA binding protein 1, *GPR98*: G protein-coupled receptor 98, *ADGRV1*: adhesion G protein-coupled receptor V1, *NEB*: nebulin, *OBSCN*: obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF, *PCDH15*: protocadherin related 15, *SPEN*: spen family transcriptional repressor, *STAG2*: stromal antigen 2, *SYNE1*: spectrin repeat containing nuclear envelope protein 1, *THBS4*: thrombospondin 4, *TTN*: titin, *TP53*: tumour protein p53, *ZFHX3*: zinc finger homeobox 3. The current network enclosed 13 nodes and 10 edges with an average node of 1.54 and an average local clustering coefficient of 0.795. The protein-protein interaction (PPI) enrichment *p*-value was significant (=3.57e-06). Functional enrichment showed significant molecular function terms, such as structural constituents of muscle and structural molecule activity, and cellular component Gene Ontology terms, such as Z disc, I band, intracellular non-membrane-bound organelle, sarcomere, cytoskeleton, and myofibril.

TABLE S5. KEGG signaling pathways for the whole gene set altered in Ewing sarcoma samples.

Pathway term	*KEGG ID*	*p-value*	*Adjusted p-value*	*Genes*
PI3K-Akt signaling pathway	hsa04151	1.33E-12	3.79E-10	IFNA6;CSF3R;IRS1;ITGB4;FLT4;TNC;PTEN;BRCA1;PIK3CB; PIK3CG;IGF1R;IKBKB;RPTOR;FGF6;FGF7;CREB3L3;TNN; CREB3L1;MYC;KDR;RAC1;JAK2;PDGFRB;MAP2K2;SYK; VWF;CHUK;PDPK1;TSC2;PPP2R5A;TSC1;COL2A1;PIK3CA; COL4A1;CCNE1;COL6A2;KIT;ITGA8;COL6A3;COL6A6;ITGA5; SOS1;TP53;TLR4;MET;EPHA2;TLR2;ITGA9;CSF1R;PHLPP1; TNXB;LAMA2;LAMA1;PIK3R1;FOXO3;THBS1;EGFR;THBS4; HSP90B1;NRAS;BCL2L11;RELN;PCK2;MAPK3;LAMB2;NOS3; INSR;LAMB4;FN1;IGF1;NFKB1;MTOR;COL1A1;NR4A1;COL1A2; FGF14;FGF19;MDM2;TEK;FGFR4;IL7R;FGFR3;FGF10
Pathways in cancer	hsa05200	2.60E-11	3.70E-09	RB1;RET;CSF3R;CXCL8;PTEN;CBLB;PIK3CB;BRCA2;GLI3; PIK3CG;CRKL;IGF1R;IKBKB;FGF6;FGF7;MECOM;CDH1; MYC;SUFU;EP300;RAC1;SKP2;PDGFRB;ARHGEF11;MAP2K2; CHUK;DCC;AXIN1;ARNT;AXIN2;RUNX1;AR;PAX8;MSH2; SMO;PIK3CA;COL4A1;MSH3;CCNE1;KIT;ARHGEF1;SOS1; MET;TP53;CSF1R;LAMA2;MAX;LAMA1;ADCY3;ADCY2; PIK3R1;CBL;HIF1A;ADCY6;EGFR;FOXO1;HSP90B1;GNA13; NRAS;STK36;TPR;ERBB2;PLCG2;ABL1;DVL3;MAPK3;CREBBP; SMAD4;CDKN2A;LAMB2;PTCH1;PTCH2;LAMB4;FN1;BRAF; IGF1;MLH1;NFKB1;MTOR;PML;FGF14;APC;FGF19;CCDC6; MDM2;GNAS;CTNNB1;FGFR3;FGF10
Focal adhesion	hsa04510	2.10E-08	1.19E-06	ITGB4;FLT4;TNC;PTEN;PIK3CB;ARHGAP35;PIK3CG;CRKL; MYLK;IGF1R;TNN;KDR;CAPN2;RAC1;PDGFRB;VWF;PDPK1; COL2A1;PIK3CA;COL4A1;COL6A2;ITGA8;COL6A3;COL6A6;T LN2;ITGA5;SOS1;DOCK1;MET;ITGA9;TNXB;LAMA2;LAMA1; PIK3R1;THBS1;EGFR;THBS4;RELN;ERBB2;FLNA;FLNC;MAPK3; LAMB2;LAMB4;FN1;BRAF;IGF1;COL1A1;COL1A2;CTNNB1
Proteoglycans in cancer	hsa05205	6.77E-08	2.76E-06	CBLB;PIK3CB;TNF;PIK3CG;IGF1R;MYC;KDR;PLCE1;RAC1; MAP2K2;PDPK1;ANK2;ANK3;ANK1;TIAM1;SMO;PIK3CA; ARHGEF1;ITGA5;SOS1;TP53;TLR4;MET;TLR2;ITPR2;ITPR3; PIK3R1;CBL;HIF1A;THBS1;EGFR;NRAS;ERBB4;ERBB2;PLCG2; GPC3;FLNA;FLNC;MAPK3;PTCH1;RDX;IGF2;FN1;BRAF;IGF1; MTOR;MDM2;CTNNB1;PTPN6
Transcriptional misregulation in cancer	hsa05202	2.12E-07	7.31E-06	ATF1;CSF1R;CXCL8;KMT2A;MAX;PRCC;DOT1L;PLAT;MLLT1; MLLT3;FOXO1;AFF1;IGF1R;HOXA9;MYC;CCR7;HIST1H3D; MEN1;ZBTB17;MEF2C;TAF15;FUS;ASPSCR1;ETV1;IGF1;PAX5; PML;NFKB1;FLI1;MLF1;RUNX1;ETV6;PER2;NCOR1;PAX8; BCL6;EWSR1;SP1;HIST3H3;MDM2;ATM;TCF3;TP53;MET
Signaling pathways regulating pluripotency of stem cells	hsa04550	3.54E-07	8.40E-06	RIF1;ONECUT1;PIK3R1;PIK3CB;PIK3CG;IGF1R;NRAS;MYC; DVL3;JARID2;JAK2;MAPK3;PCGF6;SMAD4;ZFHX3;MAP2K2; ESX1;ESRRB;AXIN1;LIFR;INHBA;IGF1;HNF1A;AXIN2;KLF4; ACVR2B;ACVR2A;REST;APC;PIK3CA;KAT6A;CTNNB1;TCF3; IL6ST;FGFR4;FGFR3;BMPR1A
Rap1 signaling pathway	hsa04015	1.49E-06	2.83E-05	CSF1R;RGS14;FLT4;ADCY3;ADCY2;PIK3R1;PIK3CB;ITGAL; THBS1;ADCY6;EGFR;PIK3CG;CRKL;IGF1R;FGF6;GRIN2A; NRAS;FGF7;CDH1;KDR;PLCE1;RAC1;MAPK3;PDGFRB; MAP2K2;INSR;MAGI3;BRAF;IGF1;GRIN2B;TIAM1;FGF14; PIK3CA;PARD3;FGF19;KIT;GNAS;CTNNB1;TEK;TLN2;FGFR4; FGFR3;MET;RAPGEF3;EPHA2;RAPGEF4;FGF10
FoxO signaling pathway	hsa04068	5.62E-07	1.14E-05	IRS1;PTEN;FOXO4;PIK3R1;PIK3CB;FOXO3;EGFR;FOXO1; PIK3CG;IGF1R;CCNB3;IKBKB;CCNB2;NRAS;BCL2L11;EP300; SKP2;PCK2;MAPK3;CREBBP;SMAD4;MAP2K2;GADD45B; CHUK;PDPK1;INSR;BRAF;IGF1;PIK3CA;BCL6;CAT;MDM2; ATM;IL7R;SOS1
ECM-receptor interaction	hsa04512	3.27E-07	8.40E-06	TNXB;LAMA2;ITGB4;LAMA1;TNC;THBS1;THBS4;RELN;SV2C; TNN;CD36;VWF;LAMB2;LAMB4;FN1;COL1A1;COL1A2;COL2A1; COL4A1;COL6A2;ITGA8;COL6A3;COL6A6;ITGA5;AGRN;ITGA9
Apoptosis	hsa04210	2.06E-06	3.66E-05	CTSZ;PRF1;ITPR2;CSF2RB;ITPR3;PIK3R1;PIK3CB;TNF; LMNB2; PIK3CG;BBC3;SPTA1;IKBKB;TUBA3E;NRAS;BCL2L11; CTSK;CAPN2;RIPK1;MAPK3;DAXX;PARP3;MAP2K2;PARP4; GADD45B;CHUK;PARP1;PARP2;PDPK1;EIF2AK3;TNFRSF10B; NFKB1;PIK3CA;ATM;TP53

Continued

TABLE S5 (CONTINUED). KEGG signaling pathways for the whole gene set altered in Ewing sarcoma samples.

Pathway term	*KEGG ID*	*p-value*	*Adjusted p-value*	*Genes*
Regulation of actin cytoskeleton	hsa04810	1.21E-05	0.00017	CYFIP1;ITGB4;ARPC1A;PIK3R1;ITGAE;PIK3CB;IQGAP2;ITGAL; IQGAP3;EGFR;ARHGAP35;PIK3CG;CRKL;MYLK;GNA13;FGF6; NRAS;FGF7;RAC1;MYH10;MAPK3;PDGFRB;GSN;MAP2K2; RDX;FN1;BRAF;SSH2;TIAM1;PIKFYVE;FGF14;APC;ABI2; ITGAD;PIK3CA;FGF19;ITGA8;ARHGEF1;ITGA5;FGFR4;SOS1; FGFR3;DOCK1;FGF10;ITGA9
MAPK signaling pathway	hsa04010	1.45E-05	0.000181	ECSIT;TNF;CRKL;IKBKB;RPS6KA4;FGF6;FGF7;MECOM;MYC; RAC1;MAP3K6;MAP3K4;PDGFRB;DAXX;MEF2C;MAP2K2; CHUK;MAPK8IP1;CDC25B;CACNB2;MAPKAPK5;SOS1;TP53; MAX;CACNA1B;CACNA1A;CACNA1D;CACNA1E;EGFR; CACNA1H;STK3;CACNA1I;NRAS;FLNA;FLNC;MAPK3; GADD45B;BRAF;NFKB1;NR4A1;FGF14;FGF19;TAOK2;NF1; PTPN7;TAB1;FGFR4;MAP3K13;PTPN5;FGFR3;FGF10
mTOR signaling pathway	hsa04150	1.25E-05	0.00017	PDPK1;IRS1;PTEN;TSC2;BRAF;TSC1;PIK3CB;IGF1;PIK3R1; TNF;HIF1A;MTOR;PIK3CG;IKBKB;RPTOR;PIK3CA;RRAGC; RICTOR;MAPK3
Longevity regulating pathway-mammal	hsa04211	1.83E-05	0.000209	IRS1;EHMT2;ADCY3;ADCY2;PIK3CB;PIK3R1;FOXO3;ADCY6; FOXO1;PIK3CG;IGF1R;RPTOR;NRAS;CREB3L3;CREB3L1; SESN2;INSR;TSC2;TSC1;IGF1;NFKB1;MTOR;PIK3CA;CAT; TP53
Cell cycle	hsa04110	2.13E-05	0.000234	RB1;PRKDC;BUB1B;CCNB3;CCNB2;MYC;CHEK2;RAD21; CHEK1;CDC27;ABL1;EP300;SKP2;ZBTB17;CREBBP;SMAD4; GADD45B;CDKN2A;SMC1A;CDC25B;STAG1;STAG2;ESPL1; CCNE1;CDC16;MDM2;MCM3;ATM;TP53;ATR
Osteoclast differentiation	hsa04380	0.193664	0.380649	CSF1R;SYK;CHUK;NCF2;PIK3R1;TYK2;PIK3CB;LILRA2;TNF; NFKB1;PIK3CG;IKBKB;PIK3CA;CTSK;PLCG2;TAB1;RAC1; MAPK3

Pathway analysis using a well-known reactome database showed the most significant cell signaling pathways to be extracellular matrix organization, developmental biology, axon guidance, cell cycle, DNA double-strand break, and collagen biosynthesis and modifying enzymes. In another metabolic pathway database, Biocarta, the role of Erb-B2 receptor tyrosine kinase 2 (ERBB2) in signaling transduction and oncology; the role of renal carcinoma associated 1 (RCA1), RCA2 and ataxia telangiectasia and Rad3-related protein serine/threonine-protein kinase (ATR) in cancer susceptibility; insulin-like growth factor I (IGF-1) receptor and longevity; and tumor suppressor ADP ribosylation factors (ARF) inhibiting ribosomal biogenesis, cell cycle G2/M checkpoint, telomerase, cellular ageing, and immortality represented the top significant signaling pathways enriched by the gene list (Fig. S3).

Gene Ontology analysis

To extend our findings from the cohorts, GO analysis of the whole mutated gene set in EWS patients and cell lines using Enrichr was carried out. The top significant GO terms are summarized in Fig. S4.

Comprehensive analysis of mutated genes towards cure

Analysis of chromosome segment location of all mutated gene list did not reveal significant predisposing loci. The human metabolome database (HMDB) was used to explore a significant interaction of many genes with magnesium ($p = 0.015$). Predicted microRNA gene target analysis computationally predicted by TargetScan (Fig. 6A) showed putative targeting of mutated genes by some microRNAs, but none was significant after adjustment of *p*-values using the Benjamini-Hochberg method. Analysis of the transcriptional factor protein-protein interaction network provides comprehensive knowledge involving dysregulated transcriptional regulatory elements that may serve as candidate drug targets for the treatment of EWS, as shown in Fig. 6B. The network included (1) the androgen receptor (AR), which regulates the gene expression of proliferation and differentiation factors in target tissues; (2) catenin (cadherin-associated protein) beta 1 (CTNNB1), a downstream component of the canonical Wnt signaling pathway; (3) signal transducer and activator of transcription 3 (STAT3), which mediates cellular responses to interleukins and other growth factors and is involved in cell

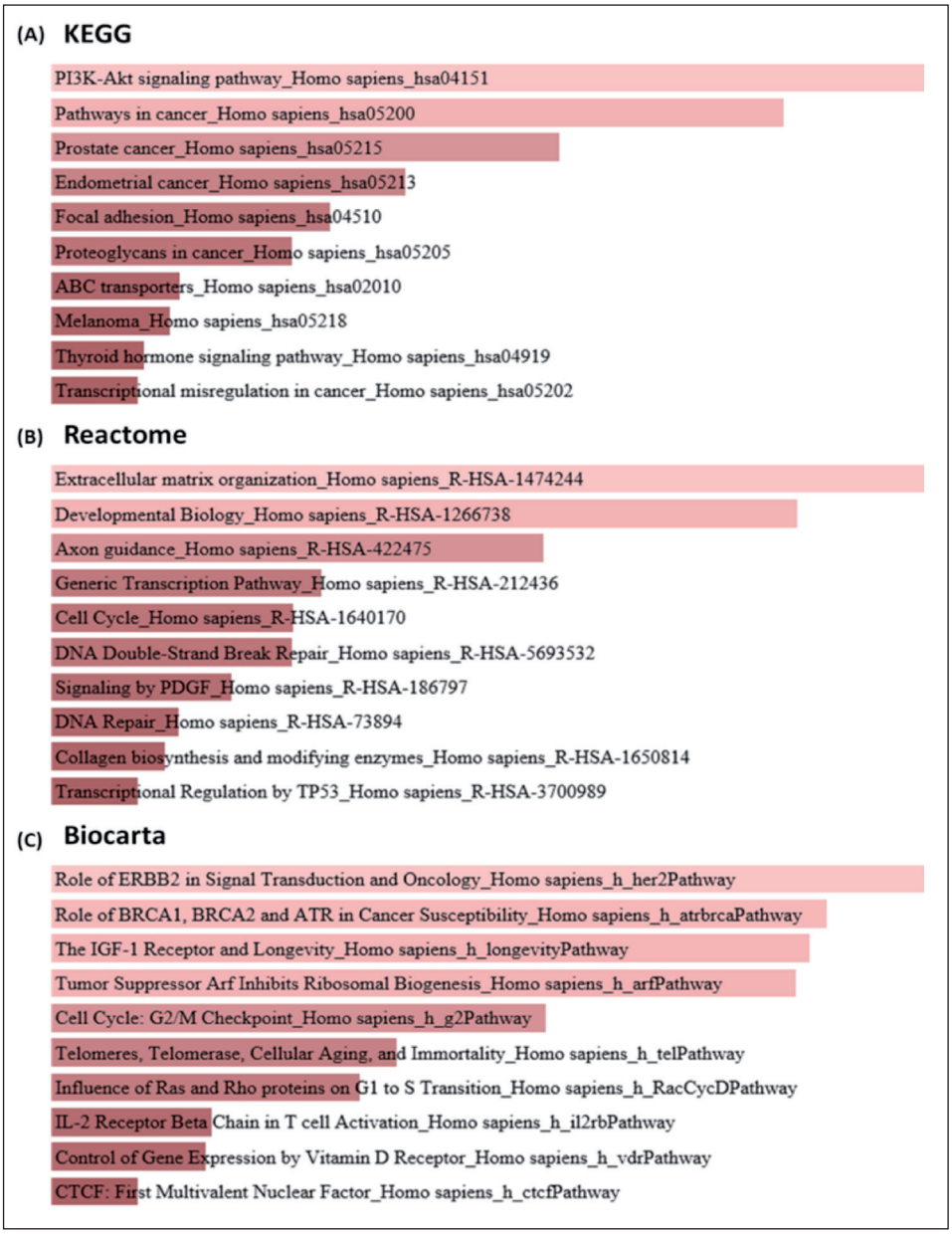

Fig. S3. Pathway enrichment analysis. Bar graph is sorted by combined score method of the p-value and z-score. The length of the bar represents the significance of that specific gene-set or term. In addition, the brighter the color, the more significant that term is.

cycle regulation; (4) V-myc myelocytomatosis viral oncogene homologue (MYC), a potent activator for the transcription of growth-related genes; (5) breast cancer 1 (BRCA1), which plays a central role in DNA repair and transcriptional regulation to maintain genomic stability; (6) SMAD family member 2 (SMAD2) and 3 (SMAD3), intracellular signal transducers and transcriptional modulators that act by binding to the promoter region of many genes; (7) histone deacetylase 2 (HDAC2), which is responsible for the deacetylation of lysine residues on the N-terminal part of the core histones (H2A, H2B, H3 and H4), gives a tag for epigenetic repression and plays an important role in transcriptional regulation, cell cycle progression and developmental events; (8) Sp1 transcription factor (SP1), which can activate or repress transcription in response to physiological and pathological stimuli and regulates the expression of a large number of genes involved in a variety of processes, such as cell growth, apoptosis, differentiation and immune responses; and (9) E1A binding protein p300 (EP300), which functions as histone acetyltransferase and regulates transcription via chromatin remodeling.

The protein-protein interaction hubs (Fig. 6C) enclosed direct protein interactors that have more than 50 protein interactors, while the Kinase Enrichment Analysis (KEA) gene set library identified human kinases and their known substrates collected from the literature. Both databases revealed important inter-connecting target genes that could act as regulators in EWS tumorigenesis, namely, cyclin-dependent kinase 1 (*CDK1*),

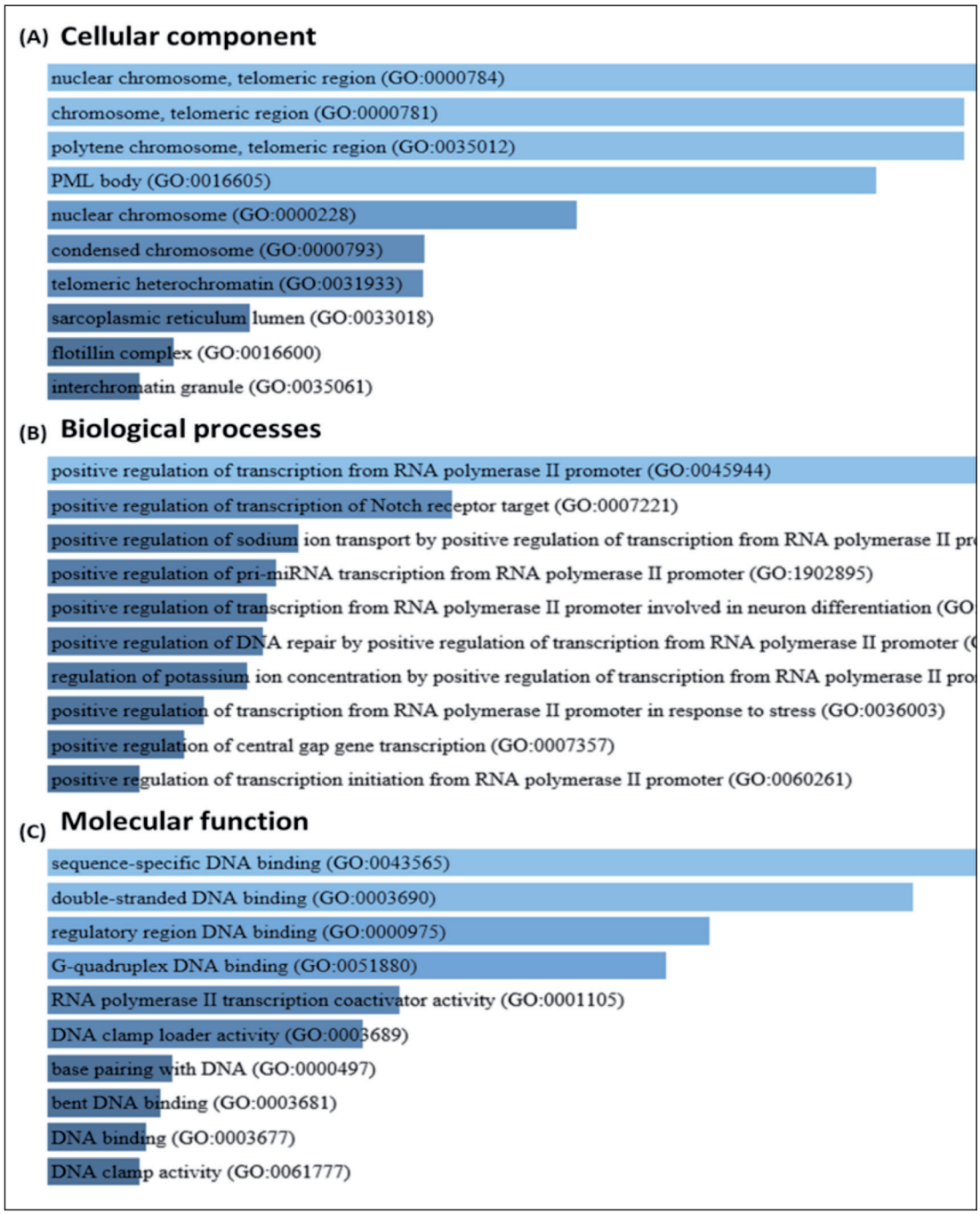

Fig. S4. Gene ontology analysis. Bar graph is sorted by combined score method of the p-value and z-score. The length of the bar represents the significance of that specific gene-set or term. In addition, the brighter the color, the more significant that term is.

cyclin-dependent kinase 2 (*CDK2*), ataxia telangiectasia mutated (*ATM*), mitogen-activated protein kinase 14 (*MAPK14*), glycogen synthase kinase 3 beta (*GSK3B*), and ribosomal protein S6 kinase, 90 kDa, and polypeptide 3 (*RPS6KA3*) (Fig. 6D). These genes play key roles in the DNA damage checkpoint, cell cycle regulation, and protein serine/threonine kinase activity[18-20].

An interactive analysis and visualization of networks that are altered in cancer is represented in Fig. S5. No FDA-approved drugs are yet available to target highly mutated genes identified in EWS samples. Recently, several clinical trials of therapies for soft-tissue sarcomas, including EWS, have been extensively reviewed[21]. Although essential progress has not occurred with molec-

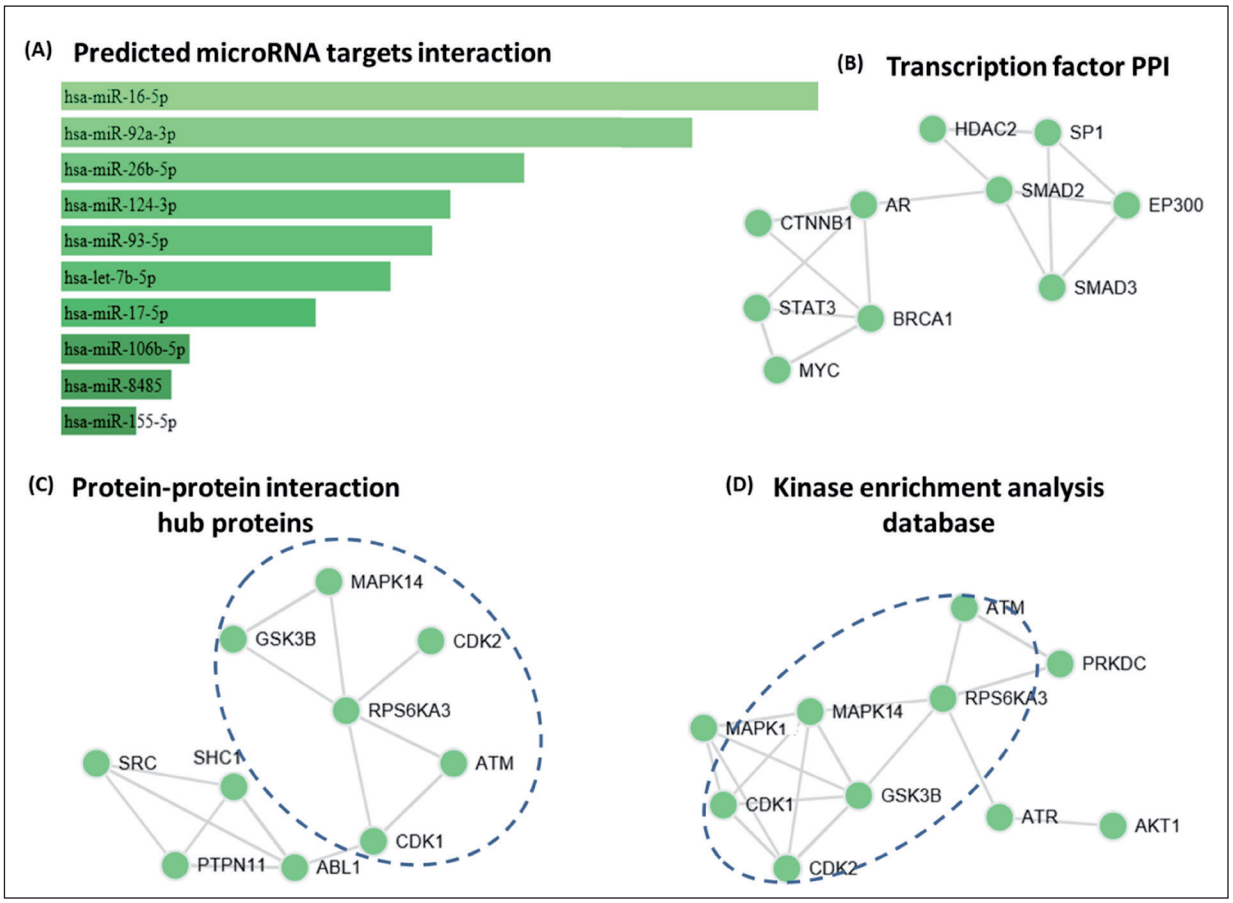

Fig. 6. Enrich functional analysis of the gene list. ***A,*** Computationally predicted microRNAs for target genes. The bar graph is sorted by the combined score method of the *p*-value and z-score. The length of the bar represents the significance of that specific gene set or term. In addition, the brighter the colour, the more significant that term is. ***B,*** Transcription factor protein-protein interaction (PPI) network associated with Ewing sarcoma. ***C,*** Hb protein PPI network. The gene list is compared with the human PPI network of proteins with 50 or more interactions. ***D,*** Kinase enrichment analysis (KEA) network. Each node represents a term, and a link between two nodes means that the two terms have some gene content similarity. Circles indicate the same protein hubs shared in the last two databases.

ularly targeted therapies over the past 30 years, as concluded by Hoang et al[21], increased knowledge about sarcoma molecular biology could lead to new and more effective precision therapeutics testing[22, 23].

CONCLUSIONS

The present *in silico* analyses could provide new insights that can be used as a roadmap for future *in vitro* or *in vivo* work. Taken together, the results show that frequent deleterious mutations in several genes, including *EWSR1*, *TTN*, *STAG2* and *TP53,* were implicated in Ewing sarcomagenesis. Their co-occurrence across tumor samples could confirm their synergistic effects in EWS tumor development. Patients carrying the *TP53* aberration alone or in combination with other mutations had much lower survival rates than those without it. A systems biology approach will be required that takes into account the genomic and epigenomic landscapes of EWS for risk stratification and future molecular targeted therapy.

Ethics approval and consent to participate

The studies involving human tissues were approved by the Ethics Committee of the original authors' institutions

Authors' contributions

EAT, AAT and MSF conceived and designed the experiments. EAT analyzed and interpreted the results. All authors contributed in writing, reading and approval of the final manuscript.

Competing interests

Authors have no competing interests to report.

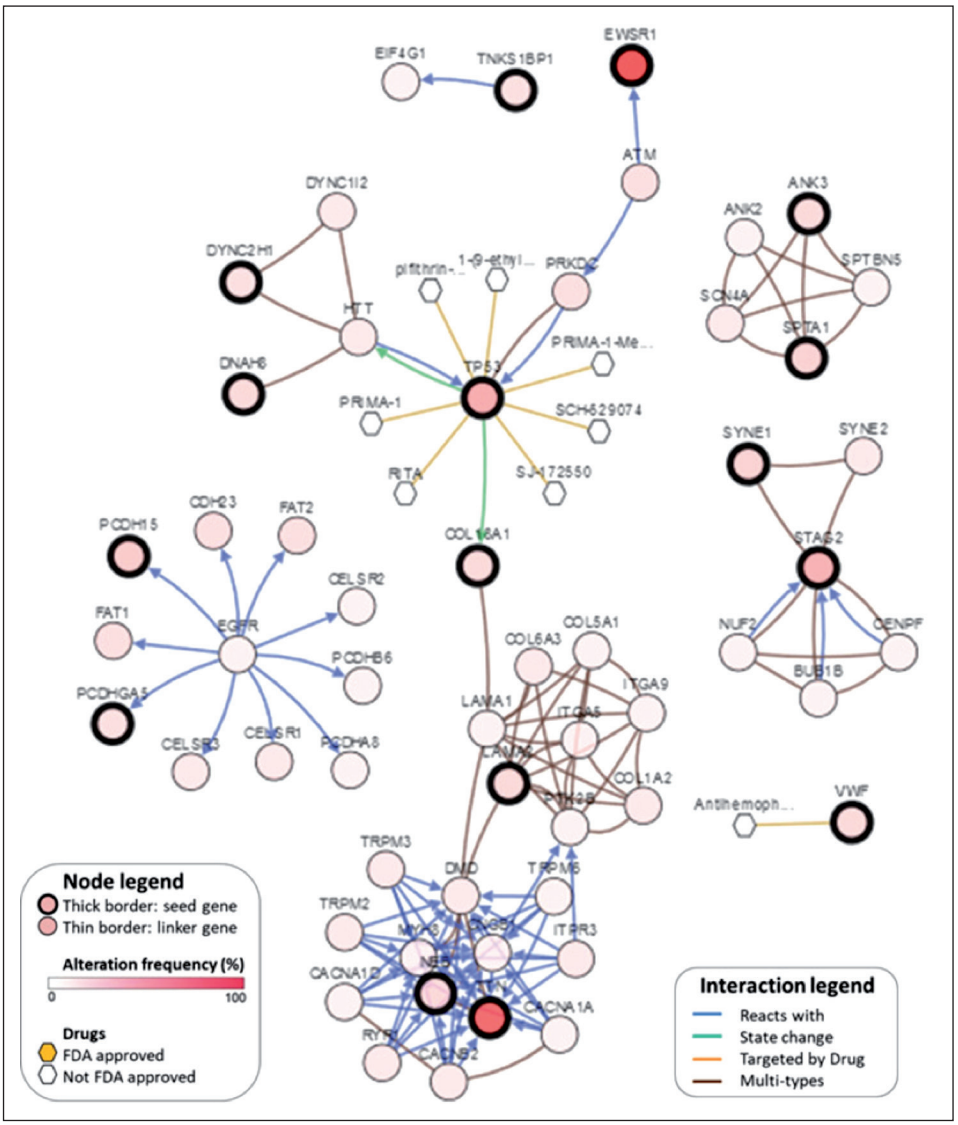

Fig. S5. Drug-gene network analysis in Ewing sarcoma tumor. The network consists of pathways and interactions from the Human Reference Protein Database (HPRD), Reactome, National Cancer Insititue (NCI), and the Memorial Sloan-Kettering Cancer Center (MSKCC) Cancer Cell Map. The network generated contains the 50 neighbor genes with the highest alteration frequency in addition to all query genes. Color of the edges is coded by interaction types which are derived from the BioPAX to SIF inference rules. For example "State Change" indicates that Gene A causes a state change, such as a phosphorylation change, within Gene B. "Targeted by Drug" indicates a drug-target interaction based on gene-centric drug-target information from the following resources: DrugBank, KEGG Drug, NCI Cancer Drugs.

REFERENCES

1. Lynn M, Wang Y, Slater J, Shah N, Conroy J, Ennis S, Morris T, Betts DR, Fletcher JA, O'Sullivan MJ. High-resolution genome-wide copy-number analyses identify localized copy-number alterations in Ewing sarcoma. Diagn Mol Pathol 2013; 22: 76-84.
2. Tirode F, Surdez D, Ma X, Parker M, Le Deley MC, Bahrami A, Zhang Z, Lapouble E, Grossetête-Lalami S, Rusch M, Reynaud S, Rio-Frio T, Hedlund E, Wu G, Chen X, Pierron G, Oberlin O, Zaidi S, Lemmon G, Gupta P, Vadodaria B, Easton J, Gut M, Ding L, Mardis ER, Wilson RK, Shurtleff S, Laurence V, Michon J, Marec-Bérard P, Downing J, Dyer M, Zhang J, Delattre O; St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project and the International Cancer Genome Consortium. Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. Cancer Discov 2014; 4: 1342-1353.
3. Brohl AS, Solomon DA, Chang W, Wang J, Song Y, Sindiri S, Patidar R, Hurd L, Chen L, Shern JF, Liao H, Wen X, Gerard J, Kim JS, Lopez Guerrero JA, Machado I, Wai DH, Picci P, Triche T, Horvai A, Miettinen M, Wei JS, Catchpool D, Llombart-Bosch A, Waldman T, Khan J. The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutation. PLoS Genet 2014; 10: e1004475.
4. Toomey EC, Schiffman JD, Lessnick SL. Recent advances in the molecular pathogenesis of Ewing's Sarcoma. Oncogene 2010; 29: 4504-4516.
5. Lessnick SL, Ladanyi M. Molecular pathogenesis of Ewing sarcoma: new therapeutic and transcriptional targets. Annu Rev Pathol 2012; 7: 145-159.
6. Crompton BD, Stewart C, Taylor-Weiner A, Alexe G, Kurek KC, Calicchio ML, Kiezun A, Carter SL, Shukla SA, Mehta SS, Thorner AR, de Torres C, Lavarino C, Suñol M, McKenna A, Sivachenko A, Cibulskis K, Lawrence MS, Stojanov P, Rosenberg M, Ambrogio L, Auclair D, Seepo S, Blumenstiel B, DeFelice M, Imaz-Rosshandler I, Schwarz-Cruz Y Celis A, Rivera MN, Rodriguez-Galindo C, Fleming MD, Golub TR, Getz G, Mora J, Stegmaier K. The genomic landscape of pediatric Ewing sarcoma. Cancer Discov 2014; 4: 1326-1341.
7. Randall RL, Lessnick SL, Jones KB, Gouw LG, Cummings JE, Cannon-Albright L, Schiffman JD. Is there a predisposition gene for Ewing's sarcoma? J Oncol 2010; 2010: 397632.

8. Jawad MU, Cheung MC, Min ES, Schneiderbauer MM, Koniaris LG, Scully SP. Ewing sarcoma demonstrates racial disparities in incidence-related and sex-related differences in outcome: an analysis of 1631 cases from the SEER database, 1973–2005. Cancer 2009; 115: 3526-3536.
9. Lin PP, Pandey MK, Jin F, Xiong S, Deavers M, Parant JM, Lozano G. EWS-FLI1 induces developmental abnormalities and accelerates sarcoma formation in a transgenic mouse model. Cancer Res 2008; 68: 8968-8975.
10. Leacock SW, Basse AN, Chandler GL, Kirk AM, Rakheja D, Amatruda JF. A zebrafish transgenic model of Ewing's sarcoma reveals conserved mediators of EWS-FLI1 tumorigenesis. Dis Model Mech 2012; 5: 95-106.
11. Javaheri T, Kazemi Z, Pencik J, Pham HT, Kauer M, Noorizadeh R, Sax B, Nivarthi H, Schlederer M, Maurer B, Hofbauer M, Aryee DN, Wiedner M, Tomazou EM, Logan M, Hartmann C, Tuckermann JP, Kenner L, Mikula M, Dolznig H, Üren A, Richter GH, Grebien F, Kovar H, Moriggl R. Increased survival and cell cycle progression pathways are required for EWS/FLI1induced malignant transformation. Cell Death Dis 2016; 7: e2419.
12. Robles AI, Harris CC. Clinical outcomes and correlates of TP53 mutations and cancer. Cold Spring Harb Perspect Biol 2010; 2: a001016.
13. Todorova R. Structure-function based molecular relationships in Ewing's sarcoma. Biomed Res Int 2015; 2015: 798426.
14. Solomon DA, Kim T, Diaz-Martinez LA, Fair J, Elkahloun AG, Harris BT, Toretsky JA, Rosenberg SA, Shukla N, Ladanyi M, Samuels Y, James CD, Yu H, Kim JS, Waldman T. Mutational inactivation of STAG2 causes aneuploidy in human cancer. Science 2011; 333: 1039-1043.
15. Thota S, Viny AD, Makishima H, Spitzer B, Radivoyevitch T, Przychodzen B, Sekeres MA, Levine RL, Maciejewski JP. Genetic alterations of the cohesin complex genes in myeloid malignancies. Blood 2014; 124: 1790-1798.
16. Solomon DA, Kim JS, Waldman T. Cohesin gene mutations in tumorigenesis: from discovery to clinical significance. BMB Reports 2014; 47: 299-310.
17. Hill VK, Kim JS, Waldman T. Cohesin mutations in human cancer. Biochim Biophys Acta 2016; 1866: 1-11.
18. Liao Y, Feng Y, Shen J, Hornicek FJ, Duan Z. The roles and therapeutic potential of cyclin-dependent kinases (CDKs) in sarcoma. Cancer Metastasis Rev 2016; 35: 151-163.
19. Klevernic IV, Morton S, Davis RJ, Cohen P. Phosphorylation of Ewing's sarcoma protein (EWS) and EWS-Fli1 in response to DNA damage. Biochem J 2009; 418: 625-634.
20. Chandhanayingyong C, Kim Y, Staples JR, Hahn C, Lee FY. MAPK/ERK signaling in osteosarcomas, Ewing sarcomas and chondrosarcomas: therapeutic implications and future directions. Sarcoma 2012; 2012: 404810.
21. Hoang NT, Acevedo LA, Mann MJ, Tolani B. A review of soft-tissue sarcomas: translation of biological advances into treatment measures. Cancer Mana Res 2018; 10: 1089-1114.
22. Desai SS, Jambhekar NA. Pathology of Ewing's sarcoma/PNET: current opinion and emerging concepts. Indian J Ortho 2010; 44: 363-368.
23. Grünewald TG, Cidre-Aranaz F, Surdez D, Tomazou EM, de Álava E, Kovar H, Sorensen PH, Delattre O, Dirksen U. Ewing sarcoma. Nat Rev Dis Primers 2018; 4: 5.