CIC-DUX4 expression drives the development of small round cell sarcoma in transgenic zebrafish: a new model revealing a role for ETV4 in CIC-mediated sarcomagenesis

Sarah Watson[1,2,*], Genevieve C. Kendall[3,4,*], Dinesh Rakheja[5], Matthew E. McFaul[6], Bruce W. Draper[6], Franck Tirode[7], Olivier Delattre[2,8] and James F Amatruda[3,4,9,#]

[1] Département d'Oncologie Médicale, Institut Curie, Paris, France

[2] Paris Sciences et Lettres (PSL) Research University, Inserm U830, Equipe Labellisée Ligue Nationale Contre le Cancer, Institut Curie, Paris, France

[3] Department of Pediatrics, UT Southwestern Medical Center, Dallas, TX, USA

[4] Department of Molecular Biology, UT Southwestern Medical Center, Dallas, TX, USA

[5] Department of Pathology, UT Southwestern Medical Center, Dallas, TX, USA

[6] Department of Molecular and Cellular Biology, University of California Davis, Davis, CA, USA

[7] Univ Lyon, Université Claude Bernard Lyon 1, INSERM 1052, CNRS 5286, Research Cancer center of Lyon, Department of Translational Research and Innovation, Centre Léon Bérard, , F-69008, Lyon, France

[8] Unité de Génétique Somatique, Institut Curie, Paris, France

[9] Department of Internal Medicine, UT Southwestern Medical Center, Dallas, TX, USA

*: these authors contributed equally

#Address correspondence to: james.amatruda@utsouthwestern.edu

Abstract

CIC-DUX4 sarcoma is a rare subtype of sarcoma characterized by a devastating prognosis and resistance to conventional therapeutic strategies. So far, only few models of the disease have been reported, and its biological mechanisms remain to be elucidated. We established mosaic transgenic zebrafish expressing the human *CIC-DUX4* fusion under the control of the β-actin promoter. CIC-DUX4 transgenic fish rapidly developed aggressive soft tissue tumors with a high penetrance. RNAseq profiling revealed that fish tumors shared major common targets with human tumors and cell lines, including the overexpression of the Pea3 transcription factors, *etv4* and *etv5*. Tumor development was strongly impaired in *etv4*-deficient zebrafish, implicating Etv4 as a critical effector of CIC-DUX4-mediated oncogenesis. Altogether, we report here the first *in vivo* model of CIC-DUX4 sarcoma in zebrafish, which will represent a major tool for future preclinical research.

Statement of significance

We report here the first successful zebrafish transgenic model of CIC-DUX4 sarcoma. Tumors arising in this model shared significant clinical and biological features with the human disease. This study highlights for the first time the oncogenic properties of CIC-DUX4 *in vivo* and the role of ETV4 in tumor development.

Introduction

CIC-DUX4 sarcomas (CDS) consist of a group of small round cell sarcomas affecting mostly young adults and characterized by the development of aggressive soft tissue tumors with devastating prognosis and resistance to conventional therapeutic strategies (1). From a pathological point of view, CDS present with small to medium-sized, round to ovoid undifferentiated tumor cells, often organized in nodular patterns, and showing variable expression of CD99 (2). Due to their clinical and pathological partial overlap with Ewing sarcomas, CDS have long been classified as "Ewing-like sarcomas" since no evidence of FET-ETS fusion gene could be identified in those tumors. *CIC-DUX4* is a fusion of the transcriptional repressor CIC with the double homeobox protein DUX4. *CIC-DUX4* fusions were first identified in 2006, and later shown to result either from a t(4;19)(q35;q13) or a t(10;19)(q26.3;q13) translocation (3, 4), with the *DUX4* gene being located both on chromosome 4 and chromosome 10 subtelomeric regions. The resulting CIC-DUX4 fusion protein retains most of the wild-type CIC, with its DNA-binding domain fused to the C-terminal portion of DUX4, and it has thus been hypothesized that the chimeric protein would act as an aberrant transcription factor.

Additional studies have shown that CDS are biologically distinct from classical FET-ETS-positive Ewing sarcomas, and characterized by a specific transcriptomic signature (5) which includes expression of the PEA3 transcription factors ETV4, ETV5 and ETV1 (3, 6). ETV4 expression has recently been recognized as a specific diagnostic marker for the identification of CIC-rearranged sarcomas in clinical specimens (7, 8). So far, the mechanisms by which CIC-DUX4 regulates gene expression to promote oncogenesis are still poorly understood, mostly due to the rarity of the disease and the lack of appropriate *in vitro* and *in vivo* models. CIC-DUX4 was shown to be able to bind the promoters of PEA3 transcription factors and activate their expression *in vitro* (3). The

oncogenic properties of the fusion protein have been more recently studied in an *ex vivo* model, showing that *CIC-DUX4* expression could transform mouse mesenchymal progenitor cells and drive the development of small round cell tumors in mice (9). However, the cell of origin of the tumor is still unknown which has hindered the development of robust animal models, and the role of ETV4 overexpression and its contribution in CIC-DUX4-mediated oncogenesis remains to be elucidated.

We report here the successful generation of the first transgenic animal model of CDS. By creating mosaic transgenic animals expressing human *CIC-DUX4* under the control of the *β-actin* ubiquitous promoter, we show that *CIC-DUX4* expression is sufficient to induce the development of small round cell sarcomas in adult zebrafish with a high penetrance. Zebrafish tumors were similar to human CDS both from pathological and biological points of view, with the expression of a common set of genes including Ets transcription factors. By using an *etv4*-deficient fish strain, we demonstrated that Etv4 is crucial to the development of CDS in zebrafish, since *etv4*-deficiency abrogated tumor development. This robust model highlights for the first time the oncogenic properties of CIC-DUX4 *in vivo* and leads the way to the design of new therapeutic strategies.

Results

CIC-DUX4 is a potent oncogene in transgenic zebrafish

To study the oncogenic properties of CIC-DUX4 *in vivo*, we cloned the human *CIC-DUX4* fusion together with the GFP-2A fluorescent reporter under the control of the *β-actin* ubiquitous promoter into the Tol2 transposon system (Figure 1A). The transgene was injected in one-cell-staged wild-type zebrafish embryos to be integrated in the genome in a stable mosaic manner. The embryos were raised to adulthood and monitored for tumor growth. CIC-DUX4 transgenic zebrafish showed the development of aggressive tumors, starting around 50 days post-fertilization (dpf) and affecting more than 30% of injected individuals by 90 dpf, whereas no tumor was observed in control animals injected with GFP-2A alone (Figure 1B). Tumors were mostly developing rapidly from muscle, growing either in the perimuscular tissues or as exophytic masses (Figure 1C), or in the head and brain area (Figure 1D). Pathological examination of tumors revealed two main histotypes. Most tumors developing from muscles showed the presence of proliferative small round tumor cells packed in highly cellular lesions and lacking any specific line of differentiation (Figure 1E). In contrast, tumors developing in the head presented rather with more disperse tumor cells surrounded by a frequent myxoid stroma and exhibiting neural-like differentiation (Figure 1F). The latter tumors were often large and infiltrated adjacent tissues, which made it difficult to determine whether they originated from the brain or from facial structures.

CIC-DUX4 zebrafish tumors are biologically homogeneous

CIC-DUX4 may act either as an oncogene, through a gain-of-function effect via aberrant transactivation of target genes or, and non-exclusively, through a dominant negative effect on CIC's repressor activity. To identify the genes regulated by CIC-DUX4 in zebrafish tumors, we

performed whole transcriptomic analyses of 12 different CIC-DUX4 fish tumors by RNAseq. The transcriptomic signatures were compared to normal muscle from adult fish, and to other fish tumors including four malignant peripheral nerve sheath tumors (arising spontaneously in *tp53*$^{M214K/M214K}$ mutant fish) and six EWSR1-FLI1-driven tumors from our previously-developed zebrafish model of Ewing sarcoma (10). Unsupervised clustering revealed that the zebrafish CIC-DUX4 form a single cluster of tumors that can be further refined in two subgoups, consistent with pathological observations of the two different histological subtypes previously observed (Figure 2A,B). Principal component analysis of the transcriptomic data similarly revealed two subgroups within the CIC-DUX4 cluster, corresponding to the histologic variants (Figure 2C).

CIC-DUX4 zebrafish tumors show biological overlap with human tumors and cell lines and overexpress orthologs of the ETS transcription factors *ETV4*, *ETV5* and *ETV1*

To study the similarity of CIC-DUX4 zebrafish tumors to the human malignancies, we next compared the transcriptomic profiles of CIC-DUX4 fish tumors with those of human primary CDS. We first identified 1412 genes differentially expressed in fish CDS as compared to EF1 and MPNST tumors (Supplementary Table S2), among which 892 were up-regulated and with a defined human orthologue. Matching these 892 genes to the 595 genes specifically expressed in human CDS (6) and having a zebrafish orthologue, highlighted 97 genes commonly up-regulated in both species (Fisher exact test P-value $< 1e^{-12}$).

To further refine the comparison of fish and human tumors, we compared the expression profiles of the CIC-DUX4 fish tumors with the genes regulated by CIC-DUX4 in the IB120 cell line. The IB120 cell line is derived from a primary human CIC-DUX4 tumor. CIC-DUX4 expression was

inhibited *in vitro* by siRNA, and Affymetrix microarrays were performed on siCIC-DUX4 IB120 and siControl IB120 cells to identify the genes regulated by the fusion oncogene (Supplementary Table S1). Gene Set Enrichment Analyses showed a significant concordance between the genes up-regulated in the CIC-DUX4 zebrafish tumors as compared to normal muscle and the genes up-regulated by CIC-DUX4 in the IB120 cells (genes down-regulated by CIC-DUX4 inhibition) (Figure 3A). Reciprocally, genes down-regulated in fish tumors were significantly concordant with the genes down-regulated by CIC-DUX4 (up-regulated upon CIC-DUX4 inhibition) (Figure 3B), showing that the fusion oncogene similarly regulates common targets in fish tumors and in the human primary tumor-derived cell line.

Comparing the 97 genes commonly up-regulated in fish and human tumors with the 636 genes up-regulated by CIC-DUX4 in the IB120 cell line (Supplementary Table S1) and having a known zebrafish orthologue, we identified 26 genes (exact Fisher-test *P*-value: 4e-9) that were overexpressed in all three models (Table 1), which included the PEA3 transcription factors ETV4 and ETV5.

Etv4 expression is necessary for CIC-DUX4 mediated oncogenesis in vivo

ETV4 overexpression has already been well-documented in CDS; however its role in CIC-mediated sarcomagenesis has never been studied *in vivo* so far. To address this question, we took benefit of an *etv4*-deficient fish strain, in which *etv4* has been knocked-out by CRISPR-Cas9 technology (*etv4-/-).* Human CIC-DUX4 was integrated in the genome of wildtype or *etv4-/-* fish using the Tol2 transposon system to generate mosaic transgenic animals, and adult fish were monitored for tumor development. As previously shown, CIC-DUX4-driven tumor development

was highly penetrant in the wild-type background with more than 30% of fish developing tumors by 75 dpf. In contrast, *etv4* deficiency significantly impaired tumor formation with less than 10% of *etv4*$^{-/-}$ fish developing tumors in the same time-frame (Mantel-Cox Log-rank test *P*-value: 0.02) (Figure 4), showing that Etv4 is major mediator of CIC-DUX4-induced sarcomagenesis *in vivo*.

Discussion

Since the first identification of CIC-DUX4 fusion transcript in two human small round cell sarcomas in 2006, more than 150 different cases have been reported in the literature (1, 4, 11). The advances in the techniques of molecular biology and the identification of specific markers to aid in pathological diagnosis have led to a significant increase in the number of reported cases (8, 12-14), and it is now acknowledged that CDS represents the most frequent subtype of the previously-called "Ewing-like" tumors. Despite evidence that CDS are biologically distinct from classical Ewing sarcoma and respond poorly to conventional chemotherapy, most patients are still treated in the same way and included in the same clinical trials as Ewing sarcoma patients.

So far, the functions of *CIC-DUX4* fusion gene have been mostly studied *in vitro*. *CIC* encodes an HMG-box transcriptional repressor homologous to Drosophila *capicua* (15). Although several breakpoints within *CIC* and *DUX4* have been reported in CDS (16), the resulting chimeric proteins always keep the majority of CIC, including its DNA-binding domain, fused together with the C-terminal part of DUX4. In the original study describing the first cases of CDS, *in vitro* experiments showed that CIC-DUX4 could transform NIH3T3 cells (3). Moreover, in U2OS osteosarcoma cells, the chimeric protein acted as a strong transcriptional activator, with the binding of the promoter and the resulting up-regulation of most of its target genes (including ETS transcription factors ETV5 and ETV1). Other *in vitro* models have been reported, including the generation of cell lines generated from patients-derived xenografts (17). More recently, Yoshimoto and colleagues reported an *ex-vivo* model of CDS, with the successful development of small round cell sarcoma in mice xenografted with mouse embryonic mesenchymal cells previously transduced by CIC-DUX4 (9). In this model, tumor latency was of less than a month, the penetrance was 100% and systemic diffusion of the disease frequent, reminiscent of the aggressiveness of human disease.

Our model is the first transgenic animal model of CDS. We demonstrated that CIC-DUX4 expression is sufficient to drive the development of tumors in transgenic zebrafish with a remarkable penetrance for a mosaic model. Moreover, the use of a ubiquitous promoter was chosen due to the fact that CDS cell of origin is unknown; thus the expression of the oncogene under the control of an appropriate tissue specific promoter might even increase tumor penetrance. Transgenic zebrafish tumors arose mainly from muscle or from the head with a very short latency and developed as undifferentiated aggressive malignancies with invasion of local anatomical structures. This clinical and pathological behavior is similar to human CDS, with most tumors developing from limbs or abdominal wall muscles and growing extensively. In our model, tumors arising from the head or the brain presented a rather different pathological aspect, with increased neural differentiation and myxoid stroma. Whether the pathological differences observed within fish tumors reflect the cell of origin of the disease remains to be further studied. Of note, head and neck CDS are not uncommon in humans (18), and several cases of CDS arising from the brain have been reported, sometimes with atypical histology (19, 20). Moreover, the related CIC-NUTM1 fusion is a recurrent genetic lesion in a subset of central nervous system Primitive Neuroectodermal Tumors (CNS-PNETs) (21, 22).

The overexpression of Ets transcription factors has long been identified as a hallmark of CDS, and ETV4 immunostaining is even used in clinical practice to help diagnosis (7). PEA3 transcription factors ETV1, ETV4 and ETV5 are expressed in numerous organs during embryonic development and in adults, and are thought to play various physiological functions involved in morphogenesis of limbs and kidneys, and in motor coordination (23). In oncogenesis, EWSR1-ETV1 and EWSR1-ETV4 fusions have been reported as rare events in classical Ewing sarcoma (24), and *TMPRSS2* fusions with either *ETV1*, *ETV4* or *ETV5* are also observed in prostate carcinoma (25). Moreover,

ETV4 overexpression has been identified as an independent factor of poor prognosis in various types of malignancies, including breast, gastric, colorectal, and lung carcinoma (23). Under physiological conditions, CIC transcriptionally represses the expression of *ETV4* and downstream targets, including genes involved in extracellular matrix remodeling and cell proliferation (26). Importantly, oligodendroglioma harboring *CIC* inactivating mutations present *ETV4* overexpression and are characterized by an aggressive behavior (27). And inactivation of *CIC* in lung and gastric carcinoma result in *ETV4* overexpression and increased risk of metastasis (28). Thus, whether *ETV4* overexpression in CDS is the direct consequence of CIC-DUX4 transcriptional activation or might result from the loss of function of wild-type *CIC* remains to be further studied *in vivo*. We showed however that Etv4 is crucial to tumor development in our model, since CIC-DUX4-mediated tumor growth was abrogated in Etv4-deficient zebrafish. This finding is of major interest for the clinical management of CDS, for which conventional therapies are poorly effective, especially as strategies to pharmacologically target Ets transcription factors are being developed. We thus believe that this first *in vivo* model of CDS will be of particular interest to further study the biological mechanisms of the disease and will constitute a major tool for the elaboration of new targeted therapies for the human disease.

Methods

Zebrafish husbandry
Danio rerio were maintained in an aquatics facility according to industry standards. Vertebrate animal work was carried out according to protocols approved by the Institutional Animal Care and Use Committee at UT Southwestern, an AALAC-accredited institution committee. Zebrafish were originally obtained from the Zebrafish International Resource Center (https://zebrafish.org). The wildtype lines used were AB/TL hybrids. The *tp53* mutant line, *tp53*^{zdf1}, was a kind gift from Tom Look (29). The *etv4* mutant line was generated in the wildtype NHGRI-1 strain.

Generation of a Tol2 enabled CIC-DUX4 expression construct
The human *CIC-DUX4* coding sequence was sub-cloned directly from a patient tumor and integrated into the Gateway expression system (Thermo) by adding 5' and 3' ATT sites (attb2r/attb3) with primers and high-fidelity PCR. Purified PCR products were cloned into a 3' entry clone as previously described (30). The tol2 kit β-actin promoter, 3' SV40 late poly A signal construct, and pDestTol2pA2 destination vector were used for construct generation and expression in zebrafish (31). The GFP viral 2A sequence was a kind gift from Steven Leach (32), and was previously sub-cloned into a middle entry Gateway expression system (33). Tol2 mRNA was synthesized from pCS2FA-transposase from Koichi Kawakami (34). The constructs utilized in this study include: β-actin-GFP2A-pA and β-actin-GFP2A-*CICDUX4*.

CIC-DUX4 integration into zebrafish embryos and characterization of tumor incidence
Zebrafish were injected at the single cell stage with injection mixes containing 50ng/µL Tol2 transposase mRNA, 40ng/µL of BetaActin-GFP2A-pA or BetaActin-GFP2A-*CICDUX4* Tol2 enabled plasmid, 0.1% phenol red, and 0.3X Danieau's buffer. The same microinjection needle was utilized to inject wildtype and etv4-/- mutant zebrafish to compare *CIC-DUX4* tumorigenic capacity. Zebrafish presenting with tumors were screened under the fluorescent microscope to determine if they were GFP positive, were collected, and the presence of malignancies confirmed by H&E staining and visual review by a pathologist.

Zebrafish tumor collection and downstream processing of RNA and histology
Zebrafish with tumors were euthanized and screened under a Nikon SMZ25 fluorescent stereomicroscope to detect GFP expression indicative of CIC-DUX4 expression. Fresh GFP+ tumor tissue was resected and snap frozen in liquid nitrogen. Total RNA was isolated from zebrafish using the TRIzol reagent kit (Invitrogen) and was utilized for RNAseq. The remaining tumor specimen was fixed in 4% paraformaldehyde/1XPBS for 48 hours at 4°C (Fisher). Samples were then de-calcified in 0.5M EDTA for 3-5 days and mounted in paraffin blocks for microtome sectioning. Hematoxylin and eosin (H&E) staining was performed on de-paraffinized slides.

Generation of etv4 mutant allele by CRISPR-Cas9
A single-stranded guide RNA and Cas9 mRNA were prepared as described (35). The guide RNA was designed to target the second of exon of etv4 (5'-<u>aattaatacgactcactata</u>GGACGCAGAATTACCCCCTC*gttttagagctagaaatagc*-3'; T7 promoter underlined, *etv4* target sequence uppercase, overlap to sgRNA scaffold oligo italicized). An injection solution was prepared with 12ng/µL of guide RNA, 30ng/µL Cas9 mRNA, 300mM KCl, and 0.1% phenol red. 2nL of injection solution was injected into the yolk of 1-cell stage

zebrafish embryos. Cutting efficiency of the guide RNA was determined by High Resolution Melt Analysis (HRMA) at 24hpf with 8 injected embryos and 8 uninjected controls using the primers etv4-Fwd (5'-AGAGGTCGCAAGGAAATGGGC-3') and etv4-Rev (5'-AACTTGCTTGTTATTTTACCTTCAGAT-3')(36). Sperm from injected males was screened by HRMA for germline transmission of etv4 alleles with indels. A 14bp frameshift-causing deletion was identified in an injected male founder by Sanger sequencing of genomic DNA. The frameshift allele was confirmed by sequencing an etv4 cDNA from outcrossed embryos.

Cell line and reagents :
IB120 cells are a kind gift from Dr. Jean-Michel Coindre and Frederic Chibon (Institut Bergonié, Bordeaux, France) and were derived from a primary tumor harboring the *CIC-DUX4* fusion. IB120 cells were cultured in Roswell Park Memorial Institute (RPMI)-1640 medium (Life Technologies, Saint Aubin) supplemented with 10% fetal calf serum (Eurobio, Courtaboeuf) and 50U/50µg penicillin/streptomycin (Life Technologies). FlexiTube siRNAs against *DUX4* and control siRNA were purchased from Qiagen (Courtaboeuf). Transfections of siRNAs were performed using the Lipofectamine RNAiMAX transfection reagent (Life Technologies). 150 000 cells were seeded in 6-well plated in 2 mL of antibiotics-free medium and transfected 24 hours later with 20nM of total siRNA. RNA extraction was performed 48 hours after transfection with RNA extraction kits (Macherey) following manufacturer's instructions and then used for Affymetrix microarrays.

Microarray:
HG-U133-Plus2 GeneChip microarrays were performed by the Affymetrix platform of the Institut Curie according to the Affymetrix GeneChip Expression Analysis Technical Manual. Expression profiles were normalized using the gcrma package version 2.34.0 in the R 3.0.2 environment (R Development Core Team, 2012) with Brainarray Entrez gene CDF v18. Quality assessment was based on relative log expression and normalized unscaled standard errors. Raw data were deposited on the Gene expression Omnibus database, as part of the GSE60740 dataset: (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60740).

Paired-end RNA sequencing.
Library constructions were performed following the TruSeq Stranded mRNA LS protocol (Illumina, San Diego, CA, USA) and sequencing was performed on HiSeq 2500 (50nt paired-end) Illumina sequencing machines, with an average of 30 million of reads per sample, by the NGS platform of the Institut Curie. FASTQ files will be deposited on the Gene expression Omnibus database.

Bioinformatic Analysis
Gene expression values were extracted by Kallisto v0.42.5 [Bray NL, Pimentel H, Melsted P, *et al.* Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 2016; **34**: 525-527] with GRCz10 release 90 genome annotation. Clustering and PCA were computed using the R packages Cluster v2.0.3 and FactoMiner v1.31.4, respectively. Gene ontology analyses were performed online (https://david.ncifcrf.gov/) with DAVID v6.7 tool [Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; **37**: 1-13]. Differential expression analysis were performed using Welsh t-tests and the associated P-values were corrected by the Bonferroni procedure. Only the genes with a Bonferroni corrected P-value under 0.05 and an absolute fold change above 2 were

further considered. Orthologues were identified using a combination of annotations from Ensembl (www.ensembl.org) and the Zebrafish Information Network (https://zfin.org) databases.

Acknowledgements:
Supported by grants from the Cancer Prevention and Research Institute of Texas (RP120685), the 1Million4Anna Foundation and the Curing Kids Cancer Foundation to JFA; the National Institute of Child Health and Development (R01 HD081551) to BWD; and the National Institute of General Medical Science (T32 GM007377) to MEM. This work was also supported by grants from the Institut Curie, the Inserm, the European PROVABES (ERA-649 NET TRANSCAN JTC-2011), ASSET (FP7-HEALTH-2010-259348) projects and the following associations: Courir pour Mathieu, Dans les pas du Géant, Les Bagouzamanon, Enfants et Santé, M la vie avec Lisa, Lulu et les petites bouilles de lune, les Amis de Claire, l'Etoile de Martin and the Société Française de lutte contre les Cancers et les leucémies de l'Enfant et de l'adolescent. GCK was supported by a QuadW-AACR Postdoctoral Fellowship for Clinical/Translational Sarcoma Research, a Young Investigator Grant from Alex's Lemonade Stand, and a Hartwell Foundation Postdoctoral Fellowship. High-throughput sequencing has been performed by the NGS platform of Institut Curie, supported by the grants ANR-10-EQPX-03 and ANR10-INBS-09-08 from the Agence Nationale de la Recherche (investissements d'avenir) and by the Canceropôle Ile-de-France.

References

1. Antonescu CR, Owosho AA, Zhang L, Chen S, Deniz K, Huryn JM, et al. Sarcomas With CIC-rearrangements Are a Distinct Pathologic Entity With Aggressive Outcome: A Clinicopathologic and Molecular Study of 115 Cases. Am J Surg Pathol. 2017;41(7):941-9.
2. Graham C, Chilton-MacNeill S, Zielenska M, Somers GR. The CIC-DUX4 fusion transcript is present in a subgroup of pediatric primitive round cell sarcomas. Hum Pathol. 2012;43(2):180-9.
3. Kawamura-Saito M, Yamazaki Y, Kaneko K, Kawaguchi N, Kanda H, Mukai H, 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. 2006;15(13):2125-37.
4. Italiano A, Sung YS, Zhang L, Singer S, Maki RG, Coindre JM, et al. High prevalence of CIC fusion with double-homeobox (DUX4) transcription factors in EWSR1-negative undifferentiated small blue round cell sarcomas. Genes Chromosomes Cancer. 2012;51(3):207-18.
5. Specht K, Sung YS, Zhang L, Richter GH, Fletcher CD, Antonescu CR. Distinct transcriptional signature and immunoprofile of CIC-DUX4 fusion-positive round cell tumors compared to EWSR1-rearranged Ewing sarcomas: further evidence toward distinct pathologic entities. Genes Chromosomes Cancer. 2014;53(7):622-33.
6. Watson S, Perrin V, Guillemot D, Reynaud S, Coindre JM, Karanian M, et al. Transcriptomic definition of molecular subgroups of small round cell sarcomas. J Pathol. 2018;245(1):29-40.

7. Le Guellec S, Velasco V, Perot G, Watson S, Tirode F, Coindre JM. ETV4 is a useful marker for the diagnosis of CIC-rearranged undifferentiated round-cell sarcomas: a study of 127 cases including mimicking lesions. Mod Pathol. 2016;29(12):1523-31.
8. Smith SC, Palanisamy N, Martin E, Almenara J, McHugh JB, Choi EK, et al. The utility of ETV1, ETV4 and ETV5 RNA in-situ hybridization in the diagnosis of CIC-DUX sarcomas. Histopathology. 2017;70(4):657-63.
9. Yoshimoto T, Tanaka M, Homme M, Yamazaki Y, Takazawa Y, Antonescu CR, et al. CIC-DUX4 Induces Small Round Cell Sarcomas Distinct from Ewing Sarcoma. Cancer Res. 2017;77(11):2927-37.
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(1):95-106.
11. Haidar A, Arekapudi S, DeMattia F, Abu-Isa E, Kraut M. High-grade undifferentiated small round cell sarcoma with t(4;19)(q35;q13.1) CIC-DUX4 fusion: emerging entities of soft tissue tumors with unique histopathologic features--a case report and literature review. Am J Case Rep. 2015;16:87-94.
12. Panagopoulos I, Gorunova L, Bjerkehagen B, Heim S. The "grep" command but not FusionMap, FusionFinder or ChimeraScan captures the CIC-DUX4 fusion gene from whole transcriptome sequencing data on a small round cell tumor with t(4;19)(q35;q13). PLoS One. 2014;9(6):e99439.
13. Yoshida A, Arai Y, Kobayashi E, Yonemori K, Ogura K, Hama N, et al. CIC break-apart fluorescence in-situ hybridization misses a subset of CIC-DUX4 sarcomas: a clinicopathological and molecular study. Histopathology. 2017;71(3):461-9.
14. Kao YC, Sung YS, Chen CL, Zhang L, Dickson BC, Swanson D, et al. ETV transcriptional upregulation is more reliable than RNA sequencing algorithms and FISH in diagnosing round cell sarcomas with CIC gene rearrangements. Genes Chromosomes Cancer. 2017;56(6):501-10.
15. Jimenez G, Guichet A, Ephrussi A, Casanova J. Relief of gene repression by torso RTK signaling: role of capicua in Drosophila terminal and dorsoventral patterning. Genes Dev. 2000;14(2):224-31.
16. Loke BN, Lee VKM, Sudhanshi J, Wong MK, Kuick CH, Puhaindran M, et al. Novel exon-exon breakpoint in CIC-DUX4 fusion sarcoma identified by anchored multiplex PCR (Archer FusionPlex Sarcoma Panel). J Clin Pathol. 2017;70(8):697-701.
17. Oyama R, Takahashi M, Yoshida A, Sakumoto M, Takai Y, Kito F, et al. Generation of novel patient-derived CIC- DUX4 sarcoma xenografts and cell lines. Sci Rep. 2017;7(1):4712.
18. Owosho AA, Estilo CL, Huryn JM, Zhang L, Fletcher CDM, Antonescu CR. Head and Neck Round Cell Sarcomas: A Comparative Clinicopathologic Analysis of 2 Molecular Subsets: Ewing and CIC-Rearranged Sarcomas. Head Neck Pathol. 2017;11(4):450-9.
19. Bielle F, Zanello M, Guillemot D, Gil-Delgado M, Bertrand A, Boch AL, et al. Unusual primary cerebral localization of a CIC-DUX4 translocation tumor of the Ewing sarcoma family. Acta Neuropathol. 2014;128(2):309-11.
20. Sturm D, Orr BA, Toprak UH, Hovestadt V, Jones DTW, Capper D, et al. New Brain Tumor Entities Emerge from Molecular Classification of CNS-PNETs. Cell. 2016;164(5):1060-72.
21. Pajtler KW, Witt H, Sill M, Jones DT, Hovestadt V, Kratochwil F, et al. Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups. Cancer Cell. 2015;27(5):728-43.

22. Le Loarer F, Pissaloux D, Watson S, Godfraind C, Galmiche-Rolland L, Silva K, et al. Clinicopathologic Features of CIC-NUTM1 Sarcomas, a New Molecular Variant of the Family of CIC-fused Sarcomas. Am J Surg Pathol. 2018.
23. Oh S, Shin S, Janknecht R. ETV1, 4 and 5: an oncogenic subfamily of ETS transcription factors. Biochim Biophys Acta. 2012;1826(1):1-12.
24. Peter M, Couturier J, Pacquement H, Michon J, Thomas G, Magdelenat H, et al. A new member of the ETS family fused to EWS in Ewing tumors. Oncogene. 1997;14(10):1159-64.
25. Clark JP, Cooper CS. ETS gene fusions in prostate cancer. Nat Rev Urol. 2009;6(8):429-39.
26. Jin Y, Ha N, Fores M, Xiang J, Glasser C, Maldera J, et al. EGFR/Ras Signaling Controls Drosophila Intestinal Stem Cell Proliferation via Capicua-Regulated Genes. PLoS Genet. 2015;11(12):e1005634.
27. Padul V, Epari S, Moiyadi A, Shetty P, Shirsat NV. ETV/Pea3 family transcription factor-encoding genes are overexpressed in CIC-mutant oligodendrogliomas. Genes Chromosomes Cancer. 2015;54(12):725-33.
28. Okimoto RA, Breitenbuecher F, Olivas VR, Wu W, Gini B, Hofree M, et al. Inactivation of Capicua drives cancer metastasis. Nat Genet. 2017;49(1):87-96.
29. Berghmans S, Murphey RD, Wienholds E, Neuberg D, Kutok JL, Fletcher CD, et al. tp53 mutant zebrafish develop malignant peripheral nerve sheath tumors. Proc Natl Acad Sci U S A. 2005;102(2):407-12.
30. Kendall GC, Amatruda JF. Zebrafish as a Model for the Study of Solid Malignancies. Methods Mol Biol. 2016;1451:121-42.
31. Kwan KM, Fujimoto E, Grabher C, Mangum BD, Hardy ME, Campbell DS, et al. The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. Dev Dyn. 2007;236(11):3088-99.
32. Provost E, Rhee J, Leach SD. Viral 2A peptides allow expression of multiple proteins from a single ORF in transgenic zebrafish embryos. Genesis. 2007;45(10):625-9.
33. Kendall GC, Watson S, Xu L, LaVigne CA, Murchison W, Rakheja D, et al. PAX3-FOXO1 transgenic zebrafish models identify HES3 as a mediator of rhabdomyosarcoma tumorigenesis. Elife. 2018;7.
34. Urasaki A, Morvan G, Kawakami K. Functional dissection of the Tol2 transposable element identified the minimal cis-sequence and a highly repetitive sequence in the subterminal region essential for transposition. Genetics. 2006;174(2):639-49.
35. Moreno-Mateos MA, Vejnar CE, Beaudoin JD, Fernandez JP, Mis EK, Khokha MK, et al. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat Methods. 2015;12(10):982-8.
36. Dahlem TJ, Hoshijima K, Jurynec MJ, Gunther D, Starker CG, Locke AS, et al. Simple Methods for Generating and Detecting Locus-Specific Mutations Induced with TALENs in the Zebrafish Genome. PLoS Genet. 2012;8(8):e1002861.

Table 1: The 26 genes similarly regulated in human and fish tumors and modulated by CIC-DUX4 in the IB120 cell line

Zebrafish Ensembl ID (ENSDARG00000)	Zebrafish Symbol	Zebrafish tumors[a]	Human tumors[b]	IB120[c]	Human Ensembl ID (ENSG00000)	Human Symbol	Gene description
105529	vgf	125.21	503.99	-7.91	128564	VGF	VGF nerve growth factor inducible
018303	etv4	66.18	65.27	-18.99	175832	ETV4	ets variant 4
087377	lbh	21.46	2.27	-7.22	213626	LBH	limb bud and heart development
086288	scg3	18.49	13.84	-4.76	104112	SCG3	secretogranin III
068732	spry4	16.54	3.67	-5.19	187678	SPRY4	sprouty homolog 4 (Drosophila)
044688	dusp4	15.74	14.38	-4.95	120875	DUSP4	dual specificity phosphatase 4
069763	etv5a	10.9	19.33	-6.58	244405	ETV5	ets variant 5
098909	SHC3	10.9	22.14	-15.63	148082	SHC3	SHC (Src homology 2 domain containing) transforming protein 3
044511	etv5b	8.49	19.33	-6.58	244405	ETV5	ets variant 5
101959	etv1	6.86	17.02	-6.38	006468	ETV1	ets variant 1
069912	hmga2	6.32	14.73	-3.72	149948	HMGA2	high mobility group AT-hook 2
027657	crhb	5.28	12.05	-13.31	147571	CRH	corticotropin releasing hormone
041449	spred1	4.89	3.23	-3.81	166068	SPRED1	sprouty-related, EVH1 domain containing 1
095732	znf423	4.85	6.4	-3.37	102935	ZNF423	zinc finger protein 423
052494	pcdh18b	4.74	4.8	-3.59	189184	PCDH18	protocadherin 18
055052	map2	4.73	18.56	-2.63	078018	MAP2	microtubule-associated protein 2
024431	CABZ01069595.1	3.42	10.68	-2.66	134954	ETS1	v-ets avian erythroblastosis virus E26 oncogene homolog 1
053990	hmgb2b	3.08	2.59	-4.22	164104	HMGB2	high mobility group box 2
010948	kif11	2.79	2.26	-7.32	138160	KIF11	kinesin family member 11
041411	rad51	2.72	2.76	-4.16	051180	RAD51	RAD51 recombinase
015002	cdh4	2.55	6.88	-3.85	179242	CDH4	cadherin 4, type 1, R-cadherin (retinal)
019367	tgfb3	2.42	4.66	-6.54	119699	TGFB3	transforming growth factor, beta 3
016733	psat1	2.4	3.63	-2.67	135069	PSAT1	phosphoserine aminotransferase 1
091419	nrarpa	2.24	2.54	-4.01	198435	NRARP	NOTCH-regulated ankyrin repeat protein
036446	PRKAR2B	2.22	9.35	-5.59	005249	PRKAR2B	protein kinase, cAMP-dependent, regulatory, type II, beta
014946	angpt2a	2.02	3.05	-42.38	091879	ANGPT2	angiopoietin 2

a. Fold change of gene expression between zebrafish CDS tumors and EF1, MPNST tumors and muscle controls (see table X).; b. Fold change of gene expression between human CDS tumors and a variety of other sarcomas (from Watson et al. 2018); c. Gene expression fold change upon CIC-DUX4 inhibition in the IB120 cell line.

Figure legends

Figure 1: Human *CIC-DUX4* is tumorigenic in zebrafish.

A: Single-cell staged wild-type zebrafish embryos were injected with mosaic GFP2A-tagged human *CIC-DUX4* under the control of the β-actin promoter. B: Tumor incidence detected in β-actin-GFP2A-CICDUX4 (CIC-DUX4) injected zebrafish (n=78) versus β-actin-GFP2A-pA injected controls (n=80) in an AB/TL wild-type genetic background. **: $p<0.001$ by Mantel-Cox Log-rank test. C and D: CIC-DUX4 induced the development of extensive tumors growing either from muscle (C) or in the head of the zebrafish (D). E and F: Representative hematoxylin and eosin stains of CIC-DUX4 induced zebrafish tumors, with either the presence of undifferentiated small round blue tumor cells (E), or tumors with a "neural-like" differentiation and myxoid stroma (F).

Figure 2: CIC-DUX4 tumors form a distinct group in unsupervised clustering analysis.

Whole transcriptome analyses were performed by RNAseq on zebrafish tumors and normal tissue, including 12 CIC-DUX4 tumors (CD4), 6 EWSR1-FLI1 tumors (EF1), 4 malignant peripheral nerve sheath tumors (MPNST), and 2 samples of normal muscle from adult zebrafish. A. Unsupervised clustering showing that CIC-DUX4 tumors form a distinct group with two subgroups. B. Representative histologic images showing small round blue cell (red) and neural-like (blue) histologies. C. Principal component analyses confirmed the separation of zebrafish CDS from the EF1, MPNST or muscles samples and a further dichotomy between the 2 morphologies (highlighted by the blue and red frames, in accordance with panel B). The left panel shows the 2 first components while the right panel shows the 1st and 3rd components (component's contribution to global variation is indicated under brackets).

Figure 3: Cross-species oncogenomics in zebrafish tumors, human tumors, and cell line.

A: Gene Set Enrichment Analyses show a significant concordance between genes up-regulated in CIC-DUX4 tumors versus normal tissue and genes up-regulated by CIC-DUX4 in the IB120 cell line (siCTRL_IB120 positively regulated). B: Gene Set Enrichment Analyses show a significant concordance between genes down-regulated in CIC-DUX4 tumors versus normal tissue and genes down-regulated by CIC-DUX4 (siDUX4_IB120 positively regulated) C: Intersection of genes up-regulated in human and fish tumors and by CIC-DUX4 in the IB120 cell line identified 26 genes in common to the three models, including the PEA3 transcription factors ETV4 and ETV5.

Figure 4: ETV4 is required for CIC-DUX4-mediated oncogenesis.

β-actin-GFP2A-CICDUX4 was injected into AB/TL wildtype (*n*=19) and ETV4-deficient (etv4-/- ; *n*=18) zebrafish embryos, and adult fish were monitored for tumor development. ETV4-deficiency significantly suppressed tumor growth compared to AB/TL wild-type controls. *: *p*=0.02 by Mantel-Cox Log-rank test.

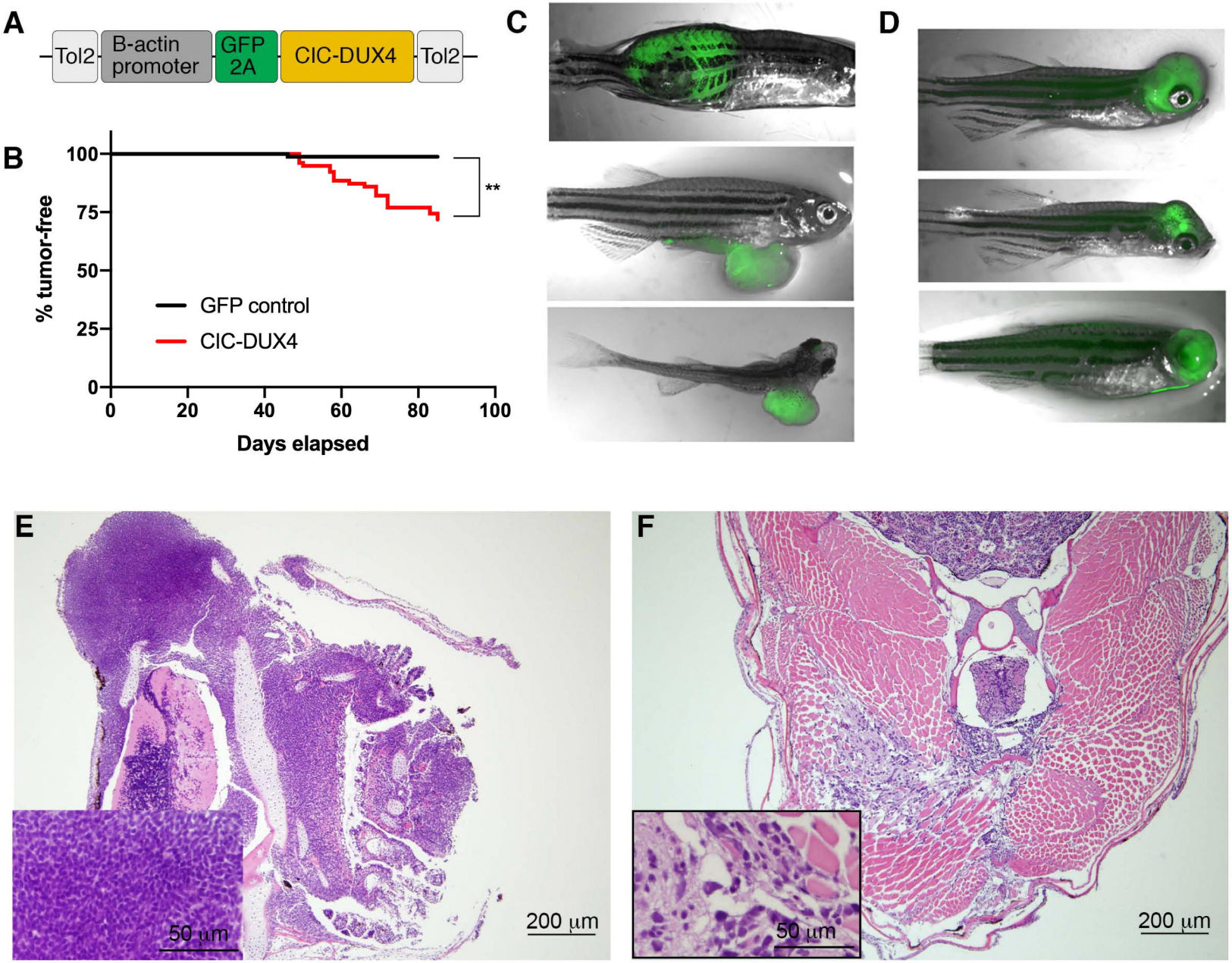
A
Tol2
B-actin promoter
GFP 2A
CIC-DUX4
Tol2
B
% tumor-free
100
75
50
25
0
0
20
40
60
80
100
Days elapsed
GFP control
CIC-DUX4
**
C
D
E
50 µm
200 µm
F
50 µm
200 µm

Figure 1

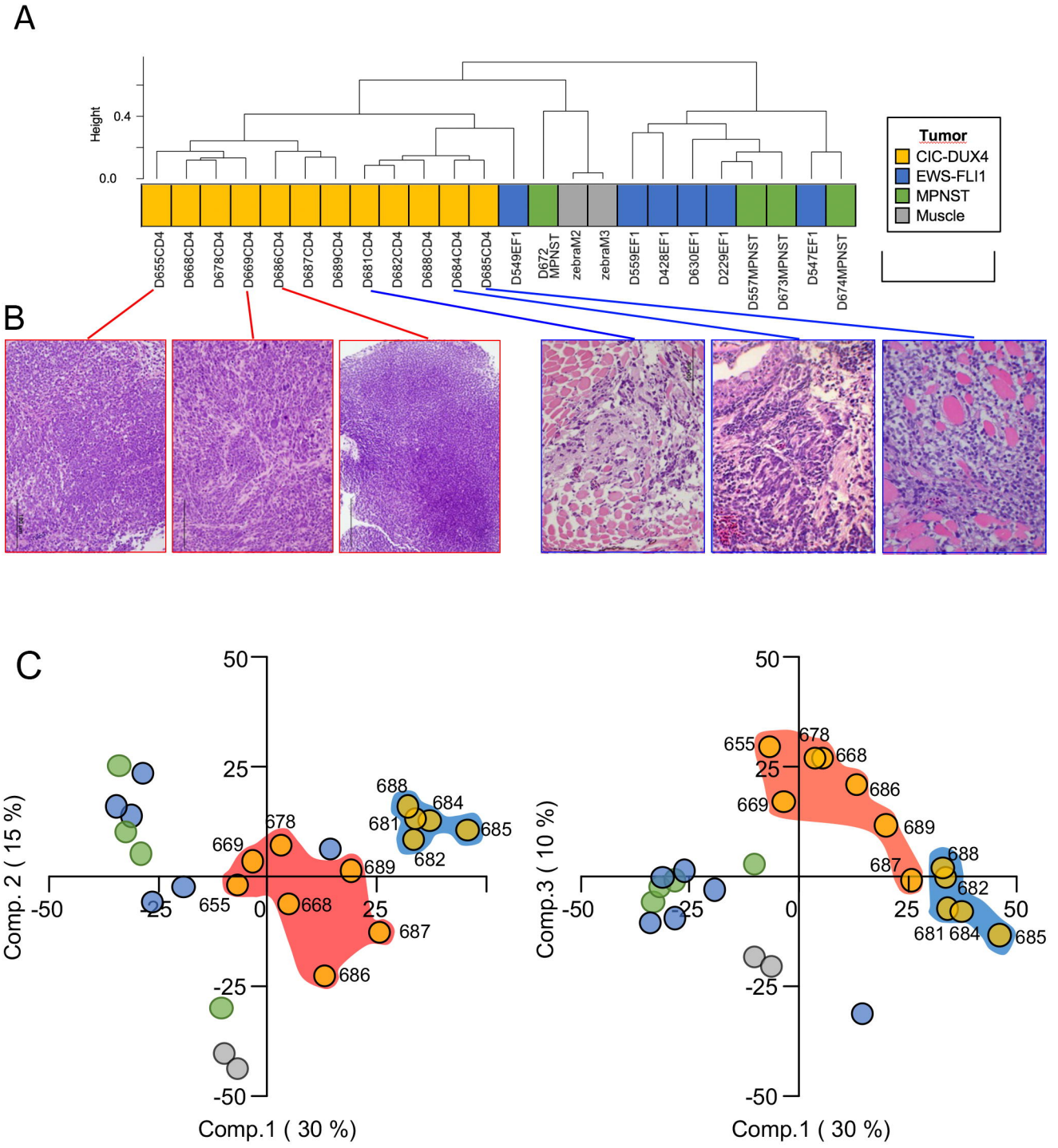
A
Height
0.4
0.0
Tumor
CIC-DUX4
EWS-FLI1
MPNST
Muscle
D655CD4
D668CD4
D678CD4
D669CD4
D686CD4
D687CD4
D689CD4
D681CD4
D682CD4
D688CD4
D684CD4
D685CD4
D549EF1
D672 MPNST
zebraM2
zebraM3
D559EF1
D428EF1
D630EF1
D229EF1
D557MPNST
D673MPNST
D547EF1
D674MPNST
B
C
Comp. 2 (15 %)
Comp.1 (30 %)
Comp.3 (10 %)
Comp.1 (30 %)

Figure 2

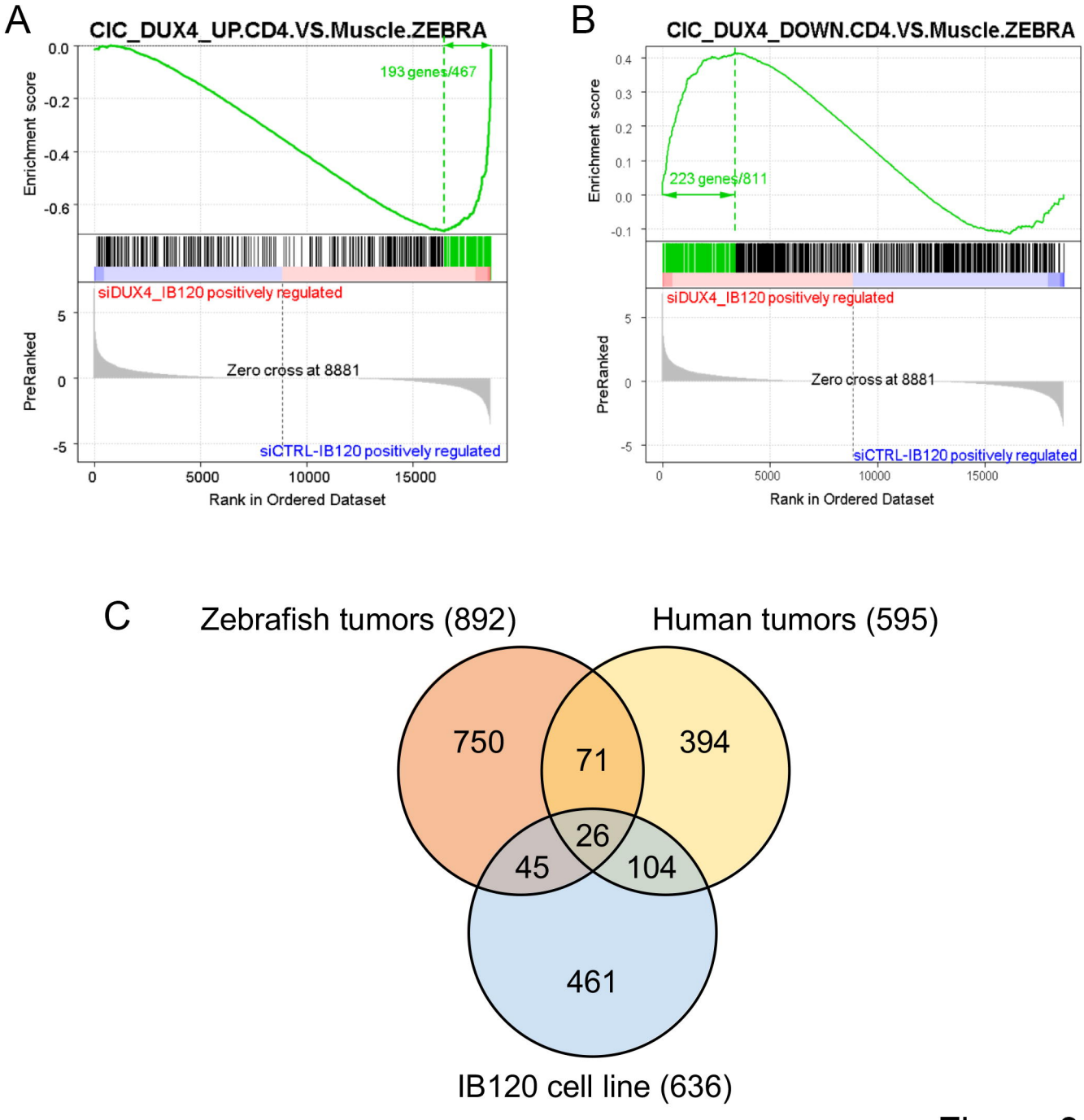
A
CIC_DUX4_UP.CD4.VS.Muscle.ZEBRA
193 genes/467
Enrichment score
siDUX4_IB120 positively regulated
PreRanked
Zero cross at 8881
siCTRL-IB120 positively regulated
Rank in Ordered Dataset
B
CIC_DUX4_DOWN.CD4.VS.Muscle.ZEBRA
223 genes/811
Enrichment score
siDUX4_IB120 positively regulated
PreRanked
Zero cross at 8881
siCTRL-IB120 positively regulated
Rank in Ordered Dataset
C
Zebrafish tumors (892)
Human tumors (595)
750
71
394
26
45
104
461
IB120 cell line (636)

Figure 3

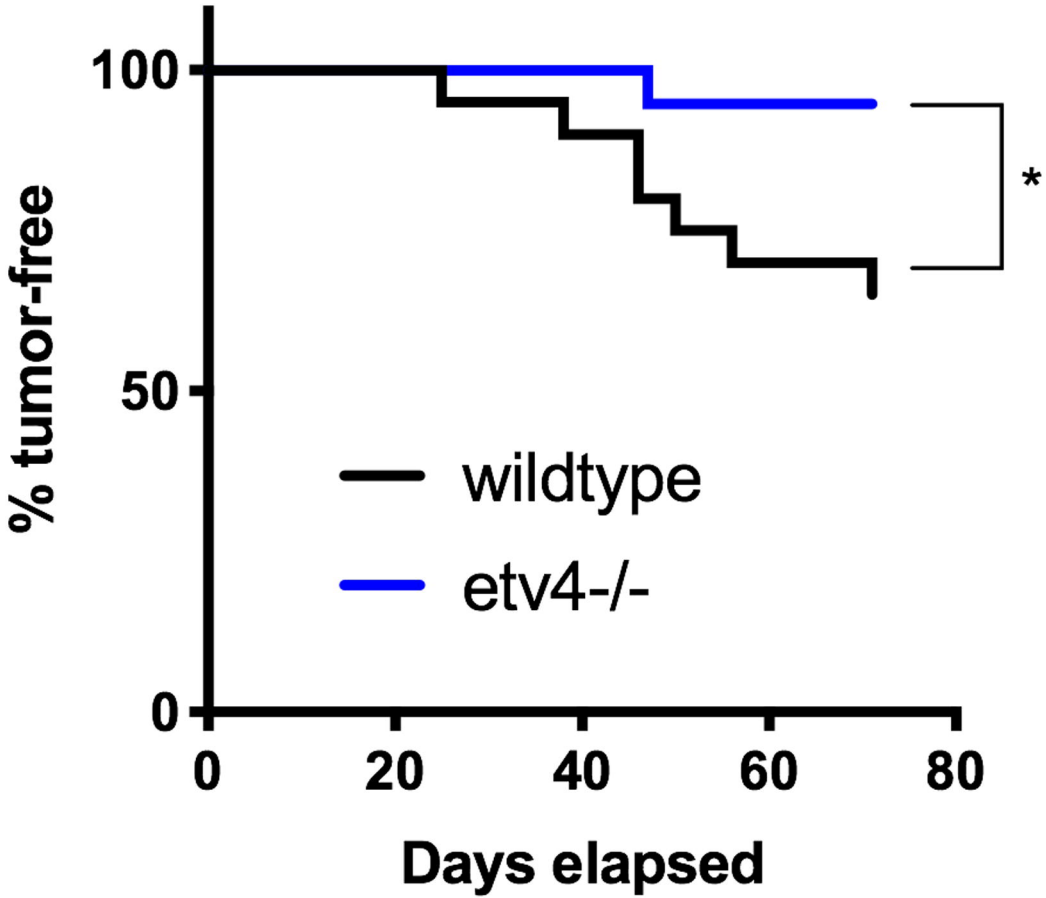

Figure 4