CrossMark

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)

Ioannis Panagopoulos^{1,2}*, Ludmila Gorunova^{1,2}, Bodil Bjerkehagen³, Sverre Heim^{1,2,4}

1 Section for Cancer Cytogenetics, Institute for Cancer Genetics and Informatics, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway, 2 Centre for Cancer Biomedicine, Faculty of Medicine, University of Oslo, Oslo, Norway, 3 Department of Pathology, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway, 0 Norway, 4 Faculty of Medicine, University of Oslo, Oslo, Norway, 3 Department of Pathology, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway, 0 Norway, 9 Faculty of Medicine, University of Oslo, Norway, 9

Abstract

Whole transcriptome sequencing was used to study a small round cell tumor in which a t(4:19)(q35:q13) was part of the complex karyotype but where the initial reverse transcriptase PCR (RT-PCR) examination did not detect a CIC-DUX4 fusion transcript previously described as the crucial gene-level outcome of this specific translocation. The RNA sequencing data were analysed using the FusionMap, FusionFinder, and ChimeraScan programs which are specifically designed to identify fusion genes. FusionMap, FusionFinder, and ChimeraScan identified 1017, 102, and 101 fusion transcripts, respectively, but CIC-DUX4 was not among them. Since the RNA sequencing data are in the fastg text-based format, we searched the files using the "grep" command-line utility. The "grep" command searches the text for specific expressions and displays, by default, the lines where matches occur. The "specific expression" was a sequence of 20 nucleotides from the coding part of the last exon 20 of CIC (Reference Sequence: NM_015125.3) chosen since all the so far reported CIC breakpoints have occurred here. Fifteen chimeric CIC-DUX4 cDNA sequences were captured and the fusion between the CIC and DUX4 genes was mapped precisely. New primer combinations were constructed based on these findings and were used together with a polymerase suitable for amplification of GC-rich DNA templates to amplify CIC-DUX4 cDNA fragments which had the same fusion point found with "grep". In conclusion, FusionMap, FusionFinder, and ChimeraScan generated a plethora of fusion transcripts but did not detect the biologically important CIC-DUX4 chimeric transcript; they are generally useful but evidently suffer from imperfect both sensitivity and specificity. The "grep" command is an excellent tool to capture chimeric transcripts from RNA sequencing data when the pathological and/or cytogenetic information strongly indicates the presence of a specific fusion gene.

Citation: Panagopoulos I, Gorunova L, Bjerkehagen B, Heim S (2014) 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 9(6): e99439. doi:10.1371/journal. pone.0099439

Editor: Francesco Bertolini, European Institute of Oncology, Italy

Received February 26, 2014; Accepted May 14, 2014; Published June 20, 2014

Copyright: © 2014 Panagopoulos et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grants from the Norwegian Cancer Society. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: ioannis.panagopoulos@rr-research.no

Introduction

The translocation t(4;19)(q35;q13) was described by Richkind et al [1] as the sole chromosomal aberration in a tumor diagnosed as poorly differentiated extraskeletal mesenchymal sarcoma in a 12-year-old-boy. The authors mentioned that a similar translocation had also been reported as part of complex karyotype in an embryonal rhabdomyosarcoma (RMS) cell line [2] and as part of a three-way translocation t(4;19;12)(q35;q13.1;q13) in an undifferentiated/embryonal RMS [3] and suggested that it might be a recurrent chromosomal aberration in malignant primitive mesenchymal stem cells [1]. Sommers et al [4] described a subcutaneous primitive neuroectodermal tumor/Ewing sarcoma without *EWSR1* rearrangement but with a complex karyotype containing a t(4;19)(q33~35;q13). Kawamura-Saito et al [5] described two cases of Ewing-like sarcoma which had a t(4;19)(q35;q13) in their karyotypes. They also showed that the translocation resulted in

fusion of the capicua transcriptional repressor *CIC* gene on 19q13, which codes for a high mobility group box transcription factor, with the double homeodomain DUX4 gene on 4q35 [5].

DUX4 is located within a D4Z4 repeat array in the subtelomeric region of chromosome arm 4q [6]. A similar D4Z4 repeat array has been identified on chromosome 10 [7]. Each D4Z4 repeat unit has an open reading frame (named DUX4) that encodes two homeoboxes [6]. There is no evidence for transcription of this gene from standard cDNA libraries, but RT-PCR and in vitro expression experiments indicate that the ORF is transcribed [8,9]. The encoded protein is located in the nucleus, induces cell death, and has been reported to function as a transcriptional activator of paired-like homeodomain transcription factor 1 (PITX1) [8,9]. So far, there are roughly 20 reported cases of sarcoma with the t(4;19)(q35;q13) and/or *CIC-DUX4* fusion [1–5,10–15]. In seven other cases with *CIC-DUX4*, the *DUX4* gene involved in the fusion apparently stems from the locus on 10q26 [13,16]. The current data suggest that the CIC-DUX4 fusion defines a subgroup of primitive round cell sarcomas, different from Ewing sarcoma, with distinctive histopathology and rapid disease progression [1-5,10-15].

Recently, whole transcriptome sequencing (RNA-Seq, RNA sequencing) was shown to be an efficient tool in the detection of fusion genes in cancer [17]. In short, extracted RNA from cancer cells is massively sequenced, and then the raw data are analyzed with one or more programs specifically dedicated to the task of detecting fusion transcripts such as ChimeraScan [18], FusionMap [19], and FusionFinder [20]. However, the programs typically identify numerous fusion transcripts making the assessment of which of them are important and which are noise extremely difficult. To overcome this challenge, we and others have used combinations of cytogenetics and RNA-Seq to detect the "primary" fusion genes of neoplasms carrying only one or a few chromosomal rearrangements. A number of fusion genes were found using this approach, among them the recurrent ZC3H7-BCOR in endometrial stromal sarcomas [21], IRF2BP2-CDX1 in a mesenchymal chondrosarcoma [22], and EWSR1-YY1 in a subset of mesotheliomas [23]. In the present study, we performed whole transcriptome sequencing to study a small round cell tumor in which t(4;19)(q35;q13) was part of a complex karyotype. While the fusion gene detection programs ChimeraScan [18], FusionMap [19], and FusionFinder [20] failed to detected the CIC-DUX4 fusion transcript, the "grep" command-line utility captured the cytogenetically indicated CIC-DUX4 fusion gene.

Materials and Methods

Ethics Statement

The study was approved by the regional ethics committee (Regional komité for medisinsk forskningsetikk Sør-Øst, Norge, http://helseforskning.etikkom.no). Written informed consent was obtained from the patient prior to her death. The ethics committee approval included a review of the consent procedure and all patient information has been anonymized and de-identified.

Patient

A 40-year-old female presented with pain in the lower part of the thoracic wall and imaging showed a tumor in thoracic skeletal muscle with extension into the retroperitoneum and costae. The histological diagnosis was small round cell sarcoma (Figure 1). Immunohistochemistry demonstrated positive findings for vimentin, AE1/AE3, and CD99, but was negative for WT1, CD56, synaptophysin, chromogranin, MYF4, SMA desmin, CD3, CD20, CD45, CD79a, TdT, S100, and FLI1. RT-PCR did not show gene fusion consistent with Ewing sarcoma (*EWSR1-ERG/FLI1*) or synovial sarcoma (*SS18-SSX1, 2 or 4*). The patient received preoperative chemotherapy and the resected specimen disclosed a 12 cm large tumor. The patient later developed lung metastasis and a local recurrence and died of sarcoma 10 months after the diagnosis.

Chromosome banding analysis and fluorescence in situ hybridization (FISH)

A sample from the surgically removed tumor was mechanically and enzymatically disaggregated and short-term cultured as described elsewhere [24]. The cultures were harvested and the chromosomes G-banded using Wright stain. The subsequent cytogenetic analysis and karyotype description followed the recommendations of the ISCN [25].

The BAC clone RP11-556K23 (chr19:47422736-47630224), which maps to 19q13.2 and contains the CIC gene, was retrieved



Figure 1. Pathologic examination of the tumor. A) The 12 cm large tumour was localized in the skeletal muscle in the thoracic wall with extension to the retroperitoneum and costae. B) HE-stained slides showed a small round cell tumour. C) Immunexpression of CD99. doi:10.1371/journal.pone.0099439.g001

from the Human genome high-resolution BAC re-arrayed clone set (the "32k set"; BACPAC Resources, http://bacpac.chori.org/ pHumanMinSet.htm). Mapping data for the 32k human re-array are available in an interactive web format (http://bacpac.chori. org/pHumanMinSet.htm, from the genomic rearrays page) and can be obtained by activation of the ucsc browser track for the hg17 UCSC assembly from the "32k set" homepage (http:// bacpac.chori.org/genomicRearrays.php). FISH mapping of the clone was performed on normal controls to confirm their chromosomal location. DNA was extracted and probes were labelled and hybridized according to Abbott Molecular recommendations (http://www.abbottmolecular.com/home.html). Chromosome preparations were counterstained with 0.2 μ g/ml DAPI and overlaid with a 24×50 mm² coverslip. Fluorescent signals were captured and analyzed using the CytoVision system (Leica Biosystems, Newcastle, UK).

High-throughput paired-end RNA-sequencing

Tumor tissue adjacent to that used for cytogenetic analysis and histologic examination had been frozen and stored at -80° C. Total RNA was extracted from the tumor using Trizol reagent according to the manufacturer's instructions (Invitrogen, Oslo, Norway) and its quality was checked by Experion Automated Electrophoresis System (Bio-Rad Laboratories, Oslo, Norway). Three µg of total RNA from the primary tumor were sent for highthroughput paired-end RNA-sequencing at the Genomics Core Facility, The Norwegian Radium Hospital (http://genomics.no/ oslo/). The RNA was sequenced using an Illumina HiSeq 2500 instrument and the Illumina software pipeline was used to process image data into raw sequencing data. Only sequence reads marked as "passed filtering" were used in the downstream data analysis. A total of 100 million reads were obtained. The softwares FusionMap (http://www.omicsoft.com/fusionmap/) [19], Fusion Finder (http://bioinformatics.childhealthresearch.org.au/software/ fusionfinder/) [20], and ChimeraScan (https://code.google.com/p/ chimerascan/) [18] were used for the discovery of fusion transcripts. In addition, the "grep" command (http://en.wikipedia.org/wiki/ Grep) was used to search the fastq files of the sequence data (http://en.wikipedia.org/wiki/FASTQ_format) for CIC sequence (NM_015125 version 3).

FusionMap was run on a PCR with Windows XP professional as the operative system. FusionFinder, ChimeraScan, and "grep" command were run on a PC with Bio-Linux 7 as the operating system [26].

PCR

The primers used for PCR amplification and sequencing are listed in Table 1.

One μ g of tumor total RNA was reverse-transcribed in a 20 μ L reaction volume using iScript Advanced cDNA Synthesis Kit for RT-qPCR according to the manufacturer's instructions (Bio-Rad Laboratories, Oslo, Norway). Initially, the 25 μ L PCR-volume contained 12.5 μ L of Premix Taq (Takara Bio Europe/SAS, Saint-Germain-en-Laye, France), 1 μ L of the synthesized cDNA, and 0.4 μ M of each of the forward CIC-4105F and reverse

DUX4-1538R primers. One μ L of the 1st PCR amplification was used as template in a nested PCR with the forward CIC-4283F and reverse DUX4-1507R primers. For the quality of the cDNA synthesis the primers CIC-4238F and CIC-4958R were used to amplify a CIC cDNA fragment. The PCRs were run on a C-1000 Thermal cycler (Bio-Rad Laboratories) with the following cycling conditions: an initial denaturation at 94°C for 30 sec followed by 35 cycles of 7 sec at 98°C and 2 min at 68°C, and a final extension for 5 min at 68°C.

In subsequent PCR amplifications, PrimeSTAR GXL DNA polymerase was used (Takara Bio). According to the company's information this is a high fidelity polymerase suitable for GC-rich templates that are otherwise difficult to amplify. The 25 μ L PCR volume contained 1× PrimeSTAR GXL Buffer (Takara Bio), 1 μ L of the synthesized cDNA, 200 μ M of each dNTP, 0.4 μ M of each of the forward primer CIC-4377F and the reverse primer DUX4-1151R or 0.4 μ M of each of the primers CIC-4453F and DUX4-1053R. The PCR was run on a C-1000 Thermal cycler (Bio-Rad Laboratories) with an initial denaturation at 94°C for 30 sec, followed by 35 cycles of 7 sec at 98°C, 2 min at 68°C, and a final extension for 5 min at 68°C. Three μ L of the PCR products were stained with GelRed (Biotium, Hayward, CA, USA), analyzed by electrophoresis through 1.0% agarose gel, and photographed.

The rest of the amplified PCR products were purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, VWR International, Oslo, Norway). Direct sequencing (Sanger sequencing) was performed using the light run sequencing service of GATC Biotech (http://www.gatc-biotech.com/en/sanger-services/lightrun-sequencing.html). The BLAST software (http:// blast.ncbi.nlm.nih.gov/Blast.cgi) was used for computer analysis of sequence data. The nucleotide sequence has been deposited in the GenBank with accession number KJ670706.

Results

G-banding analysis yielded the diagnostic karyotype 46,XX,del(2) (q13q23),t(4;19)(q35;q13),ins(11;?)(q11;?),der(20)?t(20;20)(p11;q11) [14]/46,XX[3] (Figure 2A). When metaphase spreads (Figure 2B) were hybridized with the BAC- RP11-556K23, one split signal was seen, indicating that the translocation breakpoint on chromosome 19 was within the BAC (Figure 2C). This clone contains, apart from *CIC*, the genes *GSK3A*, *ERF*, *PAFAH1B3*, *PRR19*, *TMEM145*, *MEGF8*, *CNFN*, and *LIPE* (Figure 2D).

Table 1. Filling is used for FCN antiplineations and sequence	imers used for PCR amplifications and sequencing	cin	۱q
--	--	-----	----

Oligo Name	Sequence $(5' \rightarrow 3')$
CIC-4105F	CGAAGAGCGCTTTGCTGAGTTGCC
CIC-4283F	AGAAGACGCTCCAGCTGCAGCTCG
CIC-4377F	CCGAGGACGTGCTTGGGGAGCTA
CIC-4453F	GGCCCTGGTCATGCAGCTCTTTCA
CIC-4856R	CTCAGGGGTCCCTCACCTGCCTGT
CIC-4958R	CCCAAACTGGAGAGGACGAAATGGC
DUX4-1053R	ACCGAGGAGCCTGAGGGTGGGAG
DUX4-1151R	CTTGAGCGGGCCCAGGCTGTG
DUX4-1507R	CTTCCAGCGAGGCGGCCTCTTC
DUX4-1538R	GCAGAGCCCGGTATTCTTCCTCGC

doi:10.1371/journal.pone.0099439.t001





Figure 2. Cytogenetic and FISH analyses of the tumor. A) Karyogram showing chromosome aberrations del(2)(q13q23), t(4;19)(q35;q13), ins(11;?)(q11;?), and der(20)?t(20;20)(p11;q11); breakpoints are indicated by arrows. B) FISH performed on metaphase spread using BAC RP556K23 (green signal) from 19q13 containing the CIC gene. A part from this probe has moved to the derivative chromosome 4. The der(4), der(19), and the normal chromosomes 4 and 19 are indicated by arrows. C) G- banding of the metaphase spread shown in (B). The der(4), der(19) and the normal chromosomes 4 and 19 are indicated by arrows. D) The location of the BAC RP556K23 on chromosome 19 and the genes found in this region. The data obtained from UCSC Genome Browser (http://genome.ucsc.edu/). doi:10.1371/journal.pone.0099439.g002

The initial PCR with Premix Taq and the primer set CIC-4105F/DUX4-1538R as well as the nested PCR with the primers CIC-4283F/DUX4-1507R failed to amplify any cDNA fragments. However, the primer set CIC-4238F/CIC-4958R amplified a *CIC* cDNA fragment suggesting that the synthesized cDNA was of good quality (Figure 3A). Because of the negative RT-PCR results, whole transcriptome sequencing was performed and the sequencing data were analyzed with FusionMap, FusionFinder, and ChimeraScan which are programs designed to detect fusion genes from high throughput sequencing data [18,19,20].

FusionMap identified 1024 potential fusion transcripts (Table S1) but CIC-DUX4 was not among them. Neither GSK3A, ERF, PAFAH1B3, PRR19, TMEM145, MEGF8, CNFN, nor LIPE, the other genes which are localized on the FISH probe, were found to be partners in the detected fusion transcripts. FusionFinder and ChimeraScan identified 103 and 101 potential fusion transcripts, respectively (Tables S2 and S3), but again CIC-DUX4 was not among them. Neither GSK3A, ERF, PAFAH1B3, PRR19, TMEM145, MEGF8, CNFN nor LIPE, the other genes within the BAC, were found to be partners in the detected fusion transcripts.





Figure 3. RT-PCR results for the expression of CIC-DUX4 in the tumor. A) The initial PCR with Premix Taq and the primer set CIC-4105F/ DUX4-1538R (lane 1) as well as the nested PCR with the primers CIC-4283F/DUX4-1507R (lane 2) did not amplify any cDNA fragments. The primer set CIC-4238F/CIC-4958R (lane 3) amplified a CIC cDNA fragment suggesting the good quality of the synthesized cDNA. Lane 4, Blank, no RNA in cDNA synthesis. B) PCR amplifications using the PrimeSTAR GXL DNA polymerase and the primer combinations CIC-4377F/DUX4-1151R (lane 5) and CIC-4453F/DUX4-1053R (lane 6). Lane 7, Blank, no RNA in cDNA synthesis. M, 1 Kb DNA ladder (GeneRuler, Fermentas). C) Partial sequence chromatogram of the amplified cDNA fragment showing that CIC is fused to DUX4. doi:10.1371/journal.pone.0099439.q003

Since fastq is a text-based format of the sequence data, we decided to use the "grep" command-line utility and search for sequences which contained part of the last exon of CIC (exon 20, nucleotides 4500-5473 in the sequence with accession number NM 015125 version 3). The search terms were (nt 4511-4530) "GCCGCCTTCCAGGCCCGCTA" and "CAGGGGGCCCTGACCCCACC" (nt 4701-4720). The first search term extracted 76 sequences containing CIC cDNA fragments (data not shown). The second search term extracted 22 sequences. Blasting of each of these sequences with the human genomic plus transcript database (http://blast.ncbi.nlm.nih.gov/ Blast.cgi), CIC mRNA reference sequence NM_015125.3, and $DUX4\ {\rm mRNA}$ reference sequence NM_033178.4 showed that 15 out of the 22 were chimeric CIC-DUX4 cDNA fragments (Table 2). The fusion had occurred between nt 4724 of CIC mRNA reference sequence NM_015125.3 and nt 771 of DUX4 mRNA reference sequence NM_033178.4. Using the search term "CCCACCT-CACCGGCAGAGGG" which is composed of 10 nt of CIC (CCCACCTCAC) and 10 nt (CGGCAGAGGG) of DUX4

upstream and downstream of the fusion point, 19 sequences were retrieved, 15 of which were those found with the "CAGGGGGGCCCTGACCCCACC" search term.

To verify the data obtained with the "grep" command, PCR amplifications were performed using the PrimeSTAR GXL DNA polymerase. Both primer combinations, CIC-4377F/DUX4-1151R and CIC-4453F/DUX4-1053R, amplified cDNA fragments (Figure 3B). Sanger sequencing verified that they were *CIC-DUX4* fusion transcripts which had the same fusion point found with the "grep" command (Figure 3C).

Discussion

Our initial negative result for *CIC-DUX4* fusion with RT-PCR prompted us to investigate the tumor using whole transcriptome sequencing. The small round cell tumor had the t(4;19)(q35;q13) translocation as part of its karyotype and in addition a split signal of the BAC RP11-556K23 (mapped on 19q13), which contains *CIC*, features that led us to nevertheless believe strongly that a *CIC-DUX4* fusion must be present. However, also *GSK3A*, *ERF*,

Line	Sequences
1	CCCACTCCCAGCCCCGGGGGGCCCTGACCCCAGCTCAGCTCGGGCTCGGGCCCAGGCCCGGCCCGCCC
2	CACTECEAGECEEG CA 66666CCCT6ACCCACCTCACCTCAE agaggggggtcteceaacetg ceeeggegggggggttttegeetaegeegeegeegeete
ε	CGCCCCCCACTGGCTGCTGCTGCCCCTGCCCCACTCCCAGCCCCG CAGGGGGCCCTGACCCCACCTCACCGGGGGGGGGG
4	CGCTGCTGCCCCTGCCCCACCCCGGCCCCGCCCCGGCCCCGGCCCCGCCCCACCCCGGCGG
Ъ	CCTGCCCCAGCCCAGCCCCG CAGGGGCCCCTGACCTCACCGGCGGGGGGGGGG
9	CACCGCTGCTGCTGCCCCACTCCCAGCCCCCCAGGGGGGCCCCTGACCCCACCTCACcggggggggggggcctcccaacctgccccggcgggggatt
7	TGGCACCGCTTCTGCCCCTGCCCCACTCCCAGCCCCG TGGCACCGCTCTGGCCCCACTCCCAGCCCCG TGGCACCGCCTGGCCCCACTCCCAGCCCCG TGGCACCGCCTGGCCCCACTCCCAGCCCCG TGGCACCGCCTGGCCCCCACTCCCAGCCCCG TGGCACCGCCTTCTGCCCCCACTCCCAGCCCCG TGGCACCGCCTTCTGCCCCCACTCCCAGCCCCG TGGCACCGCCTTCTGCCCCCACTCCCAGCCCCG TGGCACCGCCTTCTGCCCCCACTCCCAGCCCCG TGGCACCGCCTTCTGCCCCCACTCCCAGCCCCG TGGCACCGCCTTCTGCCCCCCCCACTCCCAGCCCCG TGGCACCGCCTTCTGCCCCCACTCCCCCCGCCCCCCCCCC
8	COTGCCCCACCCAGCCCCG CAGGGGCCCCTGACCCCACCGGCGGGGGGGGGG
0	CGCTGCTGCCCTGCCCCAGCCCCGGCCCCGGGGGGCCCTGACCCCACCTCGCAGGGGGGGTCTCCCCAACCTGCCCCGGGCGGG
10	CCTCTCCCTGTACCGCCCCCACTGGCACCGCTGCCCCCCCC
11	CCCAGCCCCG CAGGGGGCCCTGACCCCACCTCAGCTCTGGGCACTCTGGCACGGGCCGGGCTGCCCCGGCCAGCTGCCCCCCGGGGGGCCTGG
12	correccondecondecondectores of the second state of
13	CCCCACTCCCAGCCCCG CA GGGGGCCCTGACCCCACCTCAC cggcagaggggggtctccccaacctgccccgggcgggggggg
14	CTCCCAGCCCCG CAGGGGCCCTGACCCACCTCACcggcagagggggtctcccaacctgccccggcgggggggtttcgcctacgccggccg
15	CCACTGCCAGCCCCG CCCGACCCCCACCTCACCCCGGACTCTGGCACGGCCCGGGCCCGGGCCGCCCCGCCCCCCCC
16	CACTCCCAGCCCCG CA GGGGGCCCTGACCTCACCTCAC
17	CCCCTGCCCCACTCCCAGCCCCG CAGGGGCCCTGACCCCACCTCACcggcagaggggggcctcccaacccgcccccgcgggggggg
18	CCCACTCCCAGCCCCGGCCCCTGACCCCACCTCACCGgggggggggg
19	CCGCTGCTGCCCCTGCCCCAGCCCCGGCCCCGGGCCCCTGACCCCCACCTCAGCCCCGGGCTCTGGGCACGCGGGCCCGGGCCCGGCCCGGCCCCGCCCCCCCC
20	CACCGCTGCTGCCCCTGCCCCAGCCCCGGGGGGGCCCTGACCCCACCTCACCGGGGGGGG
21	GGCACCGCTGCTGCCCCCACTCCCAGCCCCGCCCGGGGGGGG
22	ccAgecccgcGCAGGGCCCTGACCCCACCTCACCggcagaggggggtctccccaacctgccccgggggggggg
The CIC sequences are shown i doi:10.1371/journal.pone.00994	in uppercase letters. DUX4 sequences are shown in bold lowercase letters. 39.t002

		CI	C-4	105	F															
1	cga	aga	gcg	ctt	tgc	tga	gtt	gcc	tga	gtt	tcg	gco	tga	agga	ıggt	cgct	gco	cto	ccc	cacc
	Е	Е	R	F	Α	Е	L	Ρ	Е	F	R	Р	Е	Е	v	L	Р	S	Ρ	т
62	ct	gca	gtc	tct	ggc	cac	ctc	acc	ccg	ggc	cat	cct	ggg	rctc	tta	iccg	rcaa	igaa	gag	gaag
	L	Q	S	L	A	т	S	Ρ	R	A	I	L	G	S	Y	R	к	K	R	K
122	aa	ctc	cac	qqa	cct	qqa	ttc	aqc	acc	cqa	qqa	ccc	cac	ctc	geo	caa	gco	rcaa	gat	gaga
	N	s	т	D	L	D	s	Ā	Р	Ē	D	Р	т	s	P	к	R	ĸ	M	R
		C	IC-	428	3F-	>														
182	aσ	aco	ctc	cao	cta	cao	ctc	aσa	acc	caa	cac	ccc	caa	agad	rtad	caa	ato	icaa	aaa	ggac
	R	R	S	S	<u>с</u>	S	S	e <u>e</u> E	P	N	т	P	к	S	Δ	к	- 9 - 1 C	E	-995 G	,99 D
		•	5	5	U	U	2	-	-		-	-	CT		377	/F->	. Ŭ	-	U	2
242	>+	~++	C 20	~++	+ ~ =	~~~~	+ > 0	200	+ > 0	202	200	cas			act	+ 7		act	- 2 0 2	atat
242	T		m		Lya	D	m	ayy	m	aya E	ayc	E E	<u>ryy</u> e	w	<u>.yc</u>	<u>cuyu</u>	<u>199</u> -	T	aye E	v
	-	Ľ	1	Ľ	D	R	1	G	-	E	A	E	D	v	ц	G	- E		453	- -
202					- 4 -													.C-4	455	2->
302	ga	caa	.ggt	gcc	ατα	CTC	CTC	CCT	gcg	gcg	cac	CCT	gga	icca	.gcg	lccd	laad	-	ιggτ	catg
	D	к.	V	_Р	Y	S	S	г	R	R	т	Ъ	D	Q	R	R	A	Ъ	V	M
	CI	C-4	453	F->	1											_		gre	p″	1
362	ca	gct	ctt	tca	gga	icca	tgg	rctt	ctt	ccc	gtc	ago	ccca	aggo	cad	cago	ccgo	cctt	CCa	aggcc
	Q	L	F	Q	D	н	G	F	F	Ρ	S	Α	Q	A	т	Α	Α	F	Q	A
422	cg	cta	tgc	aga	cat	ctt	tcc	ctc	caa	ggt	ttg	tct	gca	igtt	gaa	igat	ccc	jtga	ıggt	gcgc
	R	Y	A	D	I	F	Ρ	S	к	v	С	L	Q	L	к	I	R	Е	v	R
482	ca	gaa	gat	cat	gca	ggc	tgc	cac	tcc	cac	gga	gca	igeo	ccc	tgg	jago	tga	iggo	tcc	tctc
	Q	к	I	Μ	Q	Α	Α	т	Ρ	т	Е	Q	Ρ	Ρ	G	Α	Е	Α	Ρ	L
542	cc	tgt	acc	gcc	ccc	cac	tgg	cac	cgc	tgc	tge	ccc	tgc	ccc	cac	tcc	cag	JCCC	cg <mark>c</mark>	aggg
	Ρ	v	Ρ	P	Ρ	т	G	т	A	A	A	Р	A	Р	т	Р	S	Р	A	G
		"	gre	p″	2															
602	gg	" ccc	gre tga	P″ .ccc	2 <mark>cac</mark>	ctc	a <mark>cc</mark>	ggc	aga	ggg	ggt	cto	cca	acc	tgc	ccc	ggc	gcg	rcgg	ggat
602	<mark>gg</mark> G	" ccc P	gre <mark>tga</mark> D	P″ .ccc P	2 <mark>cac</mark> T	ctc S	a <mark>cc</mark> P	ggc A	aga E	ggg G	ggt V	cto S	cca Q	acc P	tgo A	cccc P	egge A	egeg R	icgg G	nggat D
602	<mark>gg</mark> G	° CCC P	gre tga D	P" .ccc P	2 <mark>cac</mark> T	s S	a <mark>cc</mark> P	ggc A	aga E	ggg G	ggt V	cto S	cca Q	acc P	tgc A	ecco P (-DU	egge A IX4-	gcg R 105	G G G	ggat D
602 662	<mark>gg</mark> G tto	° P cqc	gre tga D	P" .ccc P .cqc	2 <mark>cac</mark> T cqc	s S	a <mark>cc</mark> P qqc	ggc A tcc	aga E tcc	ggg G qqa	ggt V .cqq	s	cca Q gct	P	tgc A <	P P DU	aggo A X4- ctca	rgcg R 105	G G G G SR	D D
602 662	gg G tto	° P cgc A	gre tga D cta Y	P" .ccc P .cgc A	2 T Cgc A	ccc P	a <mark>co</mark> P ggc A	ggc A tcc P	aga E tcc P	ggg G gga D	ggt V .cgg G	s s ggc A	cca Q cgct	P P Ctc	tgc A <cca< td=""><td>P (-DU acco</td><td>egge A IX4- etca Q</td><td>rgcg R 105 aggc A</td><td>G G 3R tcc P</td><td>ggat D tcgg R</td></cca<>	P (-DU acco	egge A IX4- etca Q	rgcg R 105 aggc A	G G 3R tcc P	ggat D tcgg R
602 662 722	gg G tto F	P CGC A acc	gre b cta Y tcc	P" .ccc P .cgc A .gca	2 T cgc A .ccc	ccc S ccc P	a <mark>cc</mark> P ggc A	ggc A tcc P	aga E tcc P	ggg G gga D	ggt V .cgg G	ggc A	cca Q cgct L	P P Ctc S	tgo A <cca H</cca 	P (-DU acco P ageo	2ggc A 1X4- 2tca Q	R 105 Aggc A	G G G G G G G G G G G G G G G G G G G	D D tcgg R cgccg
602 662 722	gg G tto F tg	° P cgc A gcc P	gre tga D cta Y tcc	P ["] CCC P CGC A :gCa	2 T cgc A .ccc	ccc S P ggg G	a <mark>cc</mark> P ggc A caa	ggc A tcc P aag	aga E tcc P ccg R	ggg G gga D gga E	iggt V .cgg G .gga	ggc A R	ccca Q gct L Jgga	P Ctc S ACCC	tgo A cca H cgca	P (-DU acco P agco R	egge A IX4- etca Q gega	R 105 105 A ggc A G	ICGG G 3R tcc P JCCt L	nggat D tcgg R cgccg P
602 662 722	gg G tto F Ug W	° P cgc A gcc P	gre tga D cta Y tcc P	P" CCC P CGC A :gCa H	2 T cgc A .ccc P	ccc S S S S S S S S S	a <mark>cc</mark> P ggc A caa K	ggc A tcc P aag S	aga E tcc P ccg R x4-	.ggg G gga D gga E 115	iggt V .cgg G ugga D 1R	ggc A R	ecca Q egct L ggga D	P Ctc S accc P	Etgo A CCC H SgCa Q	P (-DU acco P agco R	a A X4- tca Q Jcga D	R 105 105 Aggc A C G	G G G G G G G G C C C C C C C C C C C C	ggat D tcgg R cgccg P
602 662 722 782	gg G tto F tg W	° P cgc A gcc P	gre tga D cta Y tcc P	p" ccc P cgc A cgca H	2 T cgc A .ccc P	ccc S ccc P G G	a <mark>cc</mark> P A Caa K <	ggc A tcc P aag S -DU	aga E tcc P ccg R X4-	ggg G D gga E 115	ggt V cgg G igga D 1R	ggc A R R	cca Q ggt L ggga D	ecto P S Acco P	Etgo A CCC H SgCa Q	P (-DU acco P ageo R	iggo A IX4- itca Q gcga D	R 105 105 A G G	ICGG G G S R S T C C C C C C C C C C C C C C C C C C	D D Etcgg R Egccg P
602 662 722 782	gg G tto F tg W gg	° P cgc A gcc P ccc	gre tga D cta Y tcc P ctg	P ["] CCC P CGC A GCA H	2 T cgc A .ccc P ggt	ccc S ggg G gg <u>c</u>	a <mark>cc</mark> P A Caa K <u>aca</u>	ggc A tcc P aag S -DU	aga E P CCG R X4- tgg G	9999 G D 999a E 115 9000 P	ggt V .cgg G .gga 1R .cgc A	ggc A R stca	cca Q cgct L Jgga D	etc P ctc S acco P cggo	etgo A < <u>cca</u> H gca Q ggca	P C-DU ACCC P Ageo R Cgca	aggo A X4- tca Q Jcga D	rgcg R 105 Aggc A acgc G gcca	Jegg G 3R tec P Ject L agg	ggat D tcgg R cgccg P gggtg
602 662 722 782 842	gg G tto F U gg G G C	v P cgc A gcc P ccc P	gre tga D cta Y ctg ctg C	P ⁷ P cgc A gca H cgc A	2 T cgc A ccc P ggt V	ccc S ggg G ggc A	aca P ggc A caa K aca Q	ggc A tcc P aag S -DU gcc P	aga E tcc P ccg R X4- tgg G	ggg G gga D gga 115 gcc P	cgg G G I I R C G C C C C C C C C C C C C C C C C C	ggc A R tca Q	cca Q gct L Jgga D L ago A	etc P ctc S acco P cggg G	etgo A seca H sgca Q ggco P	P C-DU Cacco P Ageo R Cgca	egge A IX4- etca Q gega D aggg G	gcg R 105 ggc A G gcca Q	ICGG G 3R tcc P JCCt L agg G	nggat D tcgg R tgccg P gggtg V ggtg
602 662 722 782 842	gg G tt F T g W g G c t T	v P cgc A gcc P ccc P tgc	gre tga D cta Y ctg c c gcc	P ["] P cgc A gca H cgc A cgc	2 T cgc A ccc P ggt v cac	ccc S S S S S S S S S S S S S S S S S S	a <mark>cc</mark> P QgC A Caa K <u>aca</u> Q cca	ggc A tcc P aag S -DU ggg ggg	aga E tcc P CCG R X4- tgg G gag	ggg G D gga 115 gcc P tcc	ggt V G Igga 1R icgc A gtg	ggc A ICCC R ICCC R ICCC R	ccca Q ccca D agg A agg A agg A	etc P ctc S acco P G gctg	itgo A icca H igca Q ggca P ggg	P C-DU ACCCO P AGCCO R CGCA Q JCCCO R	iggo A IX4- itca Q Igggo G Igggo	gcg R 105 aggc A acgc G gcca gcca	Jegg G 3R tec P Ject L agg G cca	ggat D tcgg R cgccg P gggtg V ggtc
602 662 722 782 842 902	gg G tt F g G c t L	^w P cgc A gcc P ccc P tgc A	gre tga D cta Y cta P ctg c gcc P	P ["] P cgc A cgc H cgc A cgc P	2 T cgc A ccc P ggt cac T	ccc P ggg G ggc A gtc	aca P ggc A caa K aca Q cca Q	ggc A tcc P aag -DU .gcc P ggg G CCa	aga E tcc P ccg R X4- tgg G gag S	gga gga D gga 115 gcc P tcc	ggt V G JR Cgg D IR Cgc A Sgtg W	scto S A R Etca Q gtg W	Scca Q Sgct D L L J J J J J J J J J J J J J J J J J	etc P ctc S ccc P cggc G gctg W	etgo A <cca H cgca Q ggca P ggg G G</cca 	P C-DU ACCO P AGCO R CGCA R	iggo A ix4- ix2- iztca Q gcga G gggg G gggg G gggg G	gcg R 105 ggc A gcca gcca gcca gcca	G G G G G G G G G G G G G G G G G G G	rggat D Etcgg R Egccg P Iggtg V Iggtg V Iggtc V
602 662 722 782 842 902	gg G tt F g G C t L C A	° CGC P CGC P CCC P CCC P CCC P CCC C C C	gre tga D cta Y tcc P ctg C gcc P ggc A	P ⁷ P cgc A gca H cgc A cgc A cgc A cgc	2 T cgc A ccc P ggt V cac T gtg	ccc S ggg G ggc A ggtc S gga	aca P ggc A ccaa K aca Q cca Q cca P	ggc A tcc P aaag S -DU gcc P ggg G C C a	aga E tcc P ccg R X4- tgg G gag S agc	ggg G gga D gga E 115 <u>gcc</u> P tcc P cgg	ggt V G ggg D 1R cgc A ggg W ggc A	ggc A R tca Q gtg W agc	ccca Q cgct L gggg D L agg G c c c c c c c c c c c c c c c c c c	etc P C C C C C C C C C C C C C C C C C C	etgo A <cca H ccca Q ggca Q ggca P gggg G c tccc P</cca 	P C-DU ACCC P ACCC R CCCA R CCCA R CCCA R	cggc A IX4- tca Q gcga D G ggg G G ggg G G ggc C P	rgcg R 105 Aggc A G Jcca Q Jcca P rcgc A	CCCA CCCA CCCA CCCA CCCA CCCA CCCA CCC	nggat D Etcgg R Ggccg P Nggtg V V Ggtc V C cccg
602 662 722 782 842 902 962	gg G tt F tg G C t L G C C A	x P cgc A gcc P ccc P ccc C C C G C C	gre tga D cta Y ctc P ctg C ggc P ggc A cta	P ["] CCC P CGC A CGC A CGC A CGC	2 T cgc A ccc P ggt v cac T gtg W	ccc S ccc P ggg G ggc S gga S gga C C C	aca P ggc A ccaa K Q cca Q cca Q C C acc P	ggc A tcc P aaag S -DU ggc P ggg G cca Q	aga E tcc P ccc R - tgg G gag S agc A	ggg G D gga E 115 <u>gcc</u> P tcc P cgg G	ggt V G G G G G G G G G G G G G G G G G G	setc S A CCC R R CCC R C C C C C C C C C C C C	ccca Q cgct L Jgga D A gggg G ctcc P	ACCO P CCCC P CCCC P CCCC C C C C C C C C	etge A (ccca H ccca H cgca Q ggca P gggg G c tccc P	P C-DU ACCCC P R CCCCCC R CCCCCCCCCCCCCCCCCCCC	A IX4- IX4- ICG ICG ICG ICG ICG ICG ICG ICG ICG ICG	R 105 Aggc A C C C C C C C C C C C C C C C C C C	G G G G G C C C C C C C C C C C C C C C	ggat D Ecgg R Egccg P ggtg V eggtc V ecccg P
602 662 722 782 842 902 962	gg G ttl F g G ct L g G ct L g G ct	x P cgc A gcc P ccc P tgc C C G cgc	gre tga D cta Y ctcc P ctc C cc gcc P gcc A cctc S	P CCC P CGC A CGC A CGC A CGC A	2 T cgc A cc P ggt V cac T gtg W ctc S	ccc S ccc ggg G ggc A ggc S gga C cgc	aca P ggc A caa K <u>aca</u> Q cca Q cca P gcg	ggc A tcc P aag <u>-DU</u> ggc G G Cca Q gca O	aga E tcc P Ccg R tgg G gag S agc A ggg	ggg G gga D gga E 115 <u>gcc</u> P tcc P cgg G gca	rggt V G G Qgga D IR C C G C Q G Q G C Q Q C Q Q C Q Q C Q Q C Q Q C Q Q C Q Q C C Q Q C C Q Q C C C Q Q C	setc S A acco R R ttcs Q gtg W agco A gca	ccca Q cgct L ggga D G G c c C P agg	acco P ctc S accco P G gctg G gctg W sacco P	etgo A Coca H Cgoa Q G G G G G C C P C C C C C C C C C C C C	CCCC P C-DU ACCCC P Aggco R R CCCC Q CCCC Q CCCC Q CCCC Q CCCCC C CCCCC C C CCCCCC	iggo A IX4- itca Q goga G G gggg G G ggcc P igcc	rgcg R 105 Aggc A C G G G C C C C C C C C C C C C C C C	Jegg G G Etcc P Ject L G G C G G C C C C C C C C C C C C C C	ggat D Ecgg R Egccg P ggtg V eggtc V ecccg P gggcg
602 662 722 782 842 902 962	gg G t F g G C L G C L G C C L G C C C C C C C C C	°CCC P CGC A GCC P CGC A CGC A CGC A	gre ttga D cta Y cta P ctc P ctc P ctc C gcc P gcc A cta S gcc A cta C gcc A cta C C C C C C C C C C C C C	P ["] ccc P cgc A cgc A cgc A cgc A cgc A	2 T cgc A cgc P ggt cac ggt cac ggt cac cac cac cac cac cac cac ca	ccc S ggg G ggc S gga S gga ccc S gga C S	aca P ggc A caaa Q ccaa Q cca Q Q C Q Q R cca	ggc A tcc P aag <u>gcc</u> P ggg G cca gca Q qca	aga E tcc P cc R tggg G s c A ggg G G C C C C C C C C C C C C C C C C	ggg G D gga 115 gcc P tcc G gca Q	ggt V G G 1 <u>R</u> cgg A ggt W ggt A ggt M	setc S A CCC R R CCC Q CCC Q CCC Q CCC Q CCC Q CCC Q CCC Q CCC Q CCC Q CCC Q CCC Q CCC Q CCC Q CCC Q CCC Q CCC A C C C C	ccca Q cgct L ggga D agg G ctcc P agg G ccc	P CCCC P CCCC P CCCC C C C C C C C C C	etgo A Construction Constructio	P C-DU Accor P Agoog R Cogoa Q googa R Cogoa Q Cogoa Q Cogoa Cog	Contractions of the second sec	geog R 105 Aggc A G G G G C P C C C C C C C C C C C C C C	Icegy G 3R Icec P Jcct L agg G C C C C C C C C C C C C C C C C C C	rggat D Eccgg P gggtg V ggtc V ecccg P aggcg A cccta
602 662 722 782 842 902 962 1022	gg G t F g G c t g G c t c A g C t c f C C t	°CCC P CGCC P CCC P CCC CCC CCC CCC CCCCC CCCCCCC	gre tga D cta P cta P ctc P ctc P gcc P gcc S gca S S ga E	P ["] ccc p cgc A cgc A cgc A cgc A cgc A cgc P	2 T cgc A cgc P t C C P t C C P t C C C C P t C C C C C C C C C C C C C	ccc S ggg G gg <u>c</u> S gga E cgc A gcc	aca P GGC CCA CCA CCA CCA CCA CCA CCA CCA CCA	ggc A tcc P aag <u>-DU</u> ggg G G C Q gga Q gtc S	aga E tcc P cc R - t cc R -	ggg gga D gga E 115 <u>gcc</u> P cgg G gca Q act	ggt V G gga D 1R gtg Q G ggc A ggt M cccc	setc S A A C C C C C C C C C C C C C C C C C	ccca Q igct L Jggg G G c C P G G C C C C C C C C C C C C C C C C	P CCCC P CCCC P CCCC C C C C C C C C C	A A C C C C C C C C C C C C C C C C C C	P C-DU P agoo R cgccs Q gccg R cgccs Q cgccs Q cgccs Q cgccs Q cgccs C C C C C C C C C C C C C	agga A X4- 2 2 Q Q G G G G G G G G G G G G G G G G	G G G G G G G G G G C C C C C C C C C C	Jegg G J <u>3R</u> J <u>tec</u> P Ject L J <u>sec</u> C G S G C C P S C C A S G C C A S C C A J C C T L C C A J C C T C C C C A C C C C C C C C C C C C	ggat D cccg P ggtg V ggtc V cccg P ggcg A ggcg A cctg T
602 662 722 782 842 902 962 1022	gg G t F g G c L g A g C t L g C c L g C c L c C	°CCC P CCC P CCC P CCC P CCC C C C C C C	gre tga D cta P cta P cta P cta P cta C C C C C C C C C C C C C	P ["] CCC P CGC A CGC A CGC A CGC A CGC A CGC A CGC A CGC A CGC A CGC A CGC C A CGC A C CGC A C CGC A C CGC A C CGC A C CGC A C C C C	2 Cac T CAC P CC C C C C C C C C C C C C	CCC SCCC SGG GGC GGC GGC GGC GGC GGC CCC CC	aca P GA CAA CAA QA QA QCA QCA QCA QCA QCA QCA	ggc A tcc P aag <u>-DU</u> ggg G G C C Q gga Q C C a C C C C C C C C C C C C C C C	aga E tcc P c R - tgg G gag S agc A ggg tgC A c	ggg G gga D gga E 115 gcc P C gga C g Q a L	ggt V Ggg D 1R cggg A gtg G ggg A ggt M cccc P	setc S aggc R ccc Q gtg W agc A gca Q cct C ctc	ccca Q cgct L Jggga D A gggg G ccgg G C G C C C C C	P etc S ecc P G gctg G gctg W sacc P sacc P L acc	A A CCC2 H CCC2 P CCCC2 P CCCC2 P CCCC2 P CCC2 P CCCC2 P CCCC2 P CCCC2 P CCCC2 P CCCC2 P CCCC2 P CCCC2 P CCCC2 P CCCC2 P CCCC2 P CCCC2 P CCCC2 P CCCC2 P CCCCCCCC	P C-DU Accor P Accor R C C C C C C C C C C C C C	agga A X4- 2 2 Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q	R 105 Aggc A G G G C P C C C C C C C C C C C C C C C	G G G G G G G G G G G G G G G G G G G	rggat D Eccgg P Tggtg V Ggtc V C C C C C C C C C C C C C C C C C C
602 662 722 782 842 902 962 1022	gg G tt F g G ct L c A g D ct L c A	% P cgc gcc P cgc A gcc P cgc A gcc P cgc C <td>gre tga D cta Y ctc P ctc P ctc P ctc S ggc A ctc S gga E ccc</td> <td>P["] CCC P CGC A CGC A CGC A CGC A CGC A CGC A CGC P GGA F</td> <td>2 T CQA CC P GV CC T G V CC T G W C S C A C F</td> <td>Ctc S Ggg G G G G G G C S G G C C C C C P G G C C C C C P G G C C C P G G C C C P G G C C C P G G C C C C</td> <td>aca P ggc Acaa C aca Q c Q c Q c Q C Q C Q C Q C Q C Q C Q C</td> <td>ggc A tcc P ggg G ggg G cca gca Q gtc S gca C</td> <td>$\begin{array}{c} aga\\ E\\ t \\ P \\ C \\ R \\ t \\ g \\$</td> <td>ggg G gga D gga 115 gC P C G g Q t C P C G g Q c L G G Q C C C G C C C C C C C C C C C C C</td> <td>ggt V G G Qga D 1 R cgc A g A g</td> <td>setc S aggc R ccc Q Q g g g g g g g g g c agc C c t c t c t c t c t</td> <td>2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3</td> <td>P etc S ecc P G gctg G gctg W sacc P sacc P sacc P sacc P sacc P S sacc P S sacc P S sacc P S sacc P S sacc S S sacc S S sacc S S sacc S S sacc S S sacc S S sacc S S sacc S S sacc S S sacc S S sacc S S sacc S S S sacc S S S S S S S S S S S S S S S S S S</td> <td>ttgc A Scca H Sgca Q G G G G G G C C P C C C P C C C C C C C</td> <td>P C-DU ACCCC P Agec Q CCCC R CCCCC Q CCCCC Q CCCCCCCCCCCCC</td> <td>A X4- Z Q Q Q Q Q Q Q Q Q Q Q Q Q</td> <td>rgcg R 105 Aggc A G G G G C P C C C P C C C C C C C C C C</td> <td>ICGG G G G G G C C C C C C C C C C C C C</td> <td>ggat D Correction P Correction P Correction P Correction P Correction Correct</td>	gre tga D cta Y ctc P ctc P ctc P ctc S ggc A ctc S gga E ccc	P ["] CCC P CGC A CGC A CGC A CGC A CGC A CGC A CGC P GGA F	2 T CQA CC P GV CC T G V CC T G W C S C A C F	Ctc S Ggg G G G G G G C S G G C C C C C P G G C C C C C P G G C C C P G G C C C P G G C C C P G G C C C C	aca P ggc Acaa C aca Q c Q c Q c Q C Q C Q C Q C Q C Q C Q C	ggc A tcc P ggg G ggg G cca gca Q gtc S gca C	$\begin{array}{c} aga\\ E\\ t \\ P \\ C \\ R \\ t \\ g \\ $	ggg G gga D gga 115 gC P C G g Q t C P C G g Q c L G G Q C C C G C C C C C C C C C C C C C	ggt V G G Qga D 1 R cgc A g A g	setc S aggc R ccc Q Q g g g g g g g g g c agc C c t c t c t c t c t	2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	P etc S ecc P G gctg G gctg W sacc P sacc P sacc P sacc P sacc P S sacc P S sacc P S sacc P S sacc P S sacc S S sacc S S sacc S S sacc S S sacc S S sacc S S sacc S S sacc S S sacc S S sacc S S sacc S S sacc S S S sacc S S S S S S S S S S S S S S S S S S	ttgc A Scca H Sgca Q G G G G G G C C P C C C P C C C C C C C	P C-DU ACCCC P Agec Q CCCC R CCCCC Q CCCCC Q CCCCCCCCCCCCC	A X4- Z Q Q Q Q Q Q Q Q Q Q Q Q Q	rgcg R 105 Aggc A G G G G C P C C C P C C C C C C C C C C	ICGG G G G G G C C C C C C C C C C C C C	ggat D Correction P Correction P Correction P Correction P Correction Correct
602 662 722 782 842 902 962 1022 1082	gg G T F J W G C L C A G C L C A C L C A	°CC P CGC P CGC P CC P CC P CC P CC P CC	gre tga D cta Y ctc P ctc P ctc P ctc S ggc A ctc S ggc E ccc P	P ["] CCC P CGC A CGC A CGC A CGC A CGC A CGC P GGC P GGC E	2 T CQC A CC P G V C T G W C S G C T G W C S G C T T T C A C P T C A C P C A C P C A C P C A C P C A C P C A C P C A C P C A C P C A C P C A C C P C A C C P C A C C P C C A C C P C C A C C P C C A C C P C C A C C P C C A C C P C C A C C P C C A C C P C C A C C P C C A C C P C C A C C P C C A C C P C C C C	C C C C C C C C C C C C C C C C C C C	a CC P GGA CCAA CCAC QCA QCA QCA QCA QCA QCA QCA	ggc A tcc P ggg G ggg G cca gca gcc S gca gca a 4-1	$\begin{array}{c} aga\\ E\\ t P C C R \\ t \frac{1}{2} G G G C \\ g G C R \\ g G C \\ g C \\ g G C \\ g G C \\ g G C \\$	ggg G gga D ggE 115 G P C G g Q C C G g Q L C Q Q C C G G Q Q C C C C C C C C C C C	ggt V G G G G G G G G G G G G G G G G G G	scto S Ggod R R C Ggtg W agod A Ggtg W agod A C C C C C C C	2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	ecto P C C C C C C C C C C C C C C C C C C	A A CCC2 H CCC2 P CCC2 P CCC2 P CCC2 P CCC2 C C C C	P C-DU ACCCC P Ageogram Q C-CCC R C-CCCC R C-CCCC R C-CCCC R C-CCCC R C-CCCC R C-CCCCC R C-CCCCC R C-CCCCC R C-CCCCCCCCCC	2990 A X4- 2097 J J G G G G G G G G G G G G G G G G G	rgcg R 105 Aggc A G G G G C P C C C P C C C C C C C C C C	ICGG G G G G C C C C C C C C C C C C C C	ggat D Ecgg P gggg V gggt V ecccg P eggcg A ccctg L gggag E B
602 662 722 782 842 902 962 1022 1082	gg G t F g G c L g C t L g C c L c C A c L c C	°CC P CGC P CGC P CC P CC P CC P CC P CC	gre tga D ccta P cctg C C C C C C C C C C C C C C C C C C C	P ["] CCC P CGC A CGC A CGC A CGC A CGC A CGC A CGC P C CGC C C C C C C C C C C C C C C	2 T cA cC P g V cC T g W t C S g C A t T	C C C C C C C C C C C C C C C C C C C	a CC P GGA CCAA CCA Q CCA Q CCA Q CCA Q CCA Q CCA Q CCA Q CCA Q CCA Q CCA Q CCA Q CCA Q CCA Q CCA Q CCA Q CCA A CCAA C Q CCAA C A CCAA C C A CCAA C	ggc A tcc P G G G G G G G C C A C C C C C C C C C C	aga E ter P CR $\frac{1}{2}$ $\frac{1}{$	ggg G gga gga 115 gc F cc gga L cc gga L cc gga L cc g g a L cc g g a L	cggt V G G D D C G G G G G G G G G G G G G G G	seto S GGC R R C C C C C C C C C C C C C C C C	CCCa Q ccCa C ccCa C C C C C C C C C C C C C C	etc P C C C C C C C C C C C C C C C C C C	A A CCC2 H CCC2 P CCC2 P CCC2 P CCC2 P CCC2 C C C C	P C-DU Accco P Agcco R C C C C C C C C C C C C C C C C C C	A X4- 2 Q G G G G G G G G G G G G G	In the second se	G G G C C C C C C C C C C C C C C C C C	ggat D Correction P Correction P Correction P Correction Correctio
602 662 722 782 842 902 962 1022 1082	gg G t F g G c t c f g G c t c c L c c c L c c c L c c c t c c c t c c c c	% P care gcc gcc P care gcc P care care gas gas gas gas	greater b cta cto gcto gcto gcc cco gcc cco gcc gcc gcc gcc gcc gc	P ["] CCC P CCC CA CCC CA CCC CA CCC CA CCC CA CCC CA CCC CA CCC CA CCC CA CCC CA CCC CA CCC CA CCC CA CCC CA CCC CA CA	$\begin{array}{c} 2\\ \hline \\ \mathbf{CaC}\\ \mathbf{T}\\ \mathbf{CAC}\\ \mathbf{P}\\ \mathbf{G}\\ \mathbf{V}\\ \mathbf{CC}\\ \mathbf{T}\\ \mathbf{G}\\ \mathbf{W}\\ \mathbf{CS}\\ \mathbf{G}\\ \mathbf{G}\\ \mathbf{F}\\ \mathbf{G}\\ \mathbf{G}\\ \mathbf{F}\\ \mathbf{G}\\ \mathbf{G}$	C C C P G G G G G G G G G G G G G G G G	a CC P ggA caa Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q	ggc A tcc P G G G G G G G C C C C C C C C C C C C	aga E ter p q	ggg G gga gga 115 gc F cc gga L cc gga L cc g g ca L cc g g ca L cc g g a L cc g ga T	ggt V G G D C G G G G G G G G G G G G G G G G	setc S G G C C C C C C C C C C C C C C C C C	Contractions of the second sec	iacc P S iccc P G gctg W iacc P icct I icct I iccc E	tgc A C C C C C C C C C C C C C C C C C C	P -DU -DU -DU -DU -DU -DU -DU -DU	A X4- 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3	R 105 Aggo A G G G C C P C C C C P C C C C C C C C C	G G G G G G G G G G G G G G G G G G G	ggat D tcgg R cgccg P gggtg V oggtc V cccg P ggcg A cctg L gggag E 88 R
602 662 722 782 842 902 962 1022 1082 1142	gg G t F g G c L c A a C L c C A c L c C A c L	° ° P c C P G P C <td>greater tga D cta C C C C C C C C C C C C C</td> <td>P["] CCC P CA CA CA CA CA CA CA CA CA CA CA CA CA</td> <td>2 CACP GACP GVCC GVCC GVCC GVCC GA CCP CCP CCP CCP CCP CCP CCP CC</td> <td>C C C C C C C P C C P C C P C C P C C P C C C P C C C P G G C C C P G G C C C P G G C C C P G G C C C P G G C C C C</td> <td>a P ggc caa Ccaa Q Q Cca Q Cca Q Cca P R Cca P Cca Q Cca P Cca Cca P Cca Cca Cca Cca Cca Cca Cca Cca</td> <td>ggc A tcc P S -DU gcc P ggg G C C a Q gca Q gtc S gca Q 4-1 ccgc A</td> <td>aga E tcc R G gag S c A g G G C A G G C C A C C C R - C C R - C C R - C C R - C C R - C C R C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C C</td> <td>ggg G gga E 115 gc G gga C gga L gga L gga L gga L gga L</td> <td>ggt V G gga D 1R ggt W ggc A ggt Q g g C C C P a c C P gga E</td> <td>setc S G Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q</td> <td>2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3</td> <td>iacc P S iccc P G gctg W iacc P icct I icct L iccc L</td> <td>tgc A C C C C C C C C C C C C C C C C C C</td> <td>P -DU -DU -DU -C -DU -C -C -C -C -C -C -C -C -C -C</td> <td>A A A A A A A A A A A A A A</td> <td>R 105 Aggc A G G G C C P C C C C C C C C C C C C C C</td> <td>G G G G G G G G G G G G G G G G G G G</td> <td>rggat D tcgg R cgccg P gggtg V uggtc V uggtc V ccccg R acctg E 88R accgg R</td>	greater tga D cta C C C C C C C C C C C C C	P ["] CCC P CA CA CA CA CA CA CA CA CA CA CA CA CA	2 CACP GACP GVCC GVCC GVCC GVCC GA CCP CCP CCP CCP CCP CCP CCP CC	C C C C C C C P C C P C C P C C P C C P C C C P C C C P G G C C C P G G C C C P G G C C C P G G C C C P G G C C C C	a P ggc caa Ccaa Q Q Cca Q Cca Q Cca P R Cca P Cca Q Cca P Cca Cca P Cca Cca Cca Cca Cca Cca Cca Cca	ggc A tcc P S -DU gcc P ggg G C C a Q gca Q gtc S gca Q 4-1 ccgc A	aga E tcc R G gag S c A g G G C A G G C C A C C C R - C C R - C C R - C C R - C C R - C C R C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C R - C C C C	ggg G gga E 115 gc G gga C gga L gga L gga L gga L gga L	ggt V G gga D 1R ggt W ggc A ggt Q g g C C C P a c C P gga E	setc S G Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q	2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	iacc P S iccc P G gctg W iacc P icct I icct L iccc L	tgc A C C C C C C C C C C C C C C C C C C	P -DU -DU -DU -C -DU -C -C -C -C -C -C -C -C -C -C	A A A A A A A A A A A A A A	R 105 Aggc A G G G C C P C C C C C C C C C C C C C C	G G G G G G G G G G G G G G G G G G G	rggat D tcgg R cgccg P gggtg V uggtc V uggtc V ccccg R acctg E 88R accgg R
602 662 722 782 842 902 962 1022 1082 1142	gg G t F T W G C L C A G C L C C A C L C C C C C C C C C C C C C	° ° ° <td>gre tga D cta Y ctg C c ggc A ccc P ggc A ggc A ggc</td> <td>P["] CCC P CGC CA CGC P CGC CA CGCC CA CGC CA CGC CA CGC CA CGC CA CGC CA CGC CA CGC CA CGC CA CGC CA CC CGC CA CC CGC CA CC CGC CA CC CGC CA CC CC CA CA</td> <td>2 CACP GVCT GVCT GVCS GACP CACP CACP CACP CACP CACP CACP CACP CACP CACP CACP CACP CACP CACP CACP CACS CACP CACS CAC</td> <td>C C C C C C C C C C C C C C C C C C C</td> <td>a concept of the second second</td> <td>ggc A tcc P aaag G gg G g G g G C ca g g G C ca g g G C ca g G G C Ca S Ca S Ca S Ca S Ca S Ca S Ca</td> <td>aga E cc P cg X4- G ga S c A gg G c A c S C A c C S C C S</td> <td>ggg G gga 115 <u>g</u>c P tc P g g a c L g g Q L L</td> <td>ggt V G G G G G G G G G G G G G G G G G G</td> <td>setc S A C C C C C C C C C C C C C C C C C C</td> <td>2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3</td> <td>etc P ctc S cggc G gctg gct I cct L cct L cct L</td> <td>tgc A Scca H Scca Q G G G G G G G G C P C C P C C C P C C C P C C C C</td> <td>CCCC P CCCC P Ageo Q Q CCCC R CCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCCC</td> <td>2990 A IX4- 210 90 90 0 0 90 0 0 0 0 0 0 0 0 0 0 0 0</td> <td>In the second se</td> <td>G G G G G G G G G G G G G G G G G G G</td> <td>rggat D tcgg R cgccg P gggtg V cccg P gggcg A ccctg L gggag E 8R accgg R</td>	gre tga D cta Y ctg C c ggc A ccc P ggc A ggc A ggc	P ["] CCC P CGC CA CGC P CGC CA CGCC CA CGC CA CGC CA CGC CA CGC CA CGC CA CGC CA CGC CA CGC CA CGC CA CC CGC CA CC CGC CA CC CGC CA CC CGC CA CC CC CA CA	2 CACP GVCT GVCT GVCS GACP CACP CACP CACP CACP CACP CACP CACP CACP CACP CACP CACP CACP CACP CACP CACS CACP CACS CAC	C C C C C C C C C C C C C C C C C C C	a concept of the second	ggc A tcc P aaag G gg G g G g G C ca g g G C ca g g G C ca g G G C Ca S Ca S Ca S Ca S Ca S Ca S Ca	aga E cc P cg X4- G ga S c A gg G c A c S C A c C S C C S	ggg G gga 115 <u>g</u> c P tc P g g a c L g g Q L L	ggt V G G G G G G G G G G G G G G G G G G	setc S A C C C C C C C C C C C C C C C C C C	2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	etc P ctc S cggc G gctg gct I cct L cct L cct L	tgc A Scca H Scca Q G G G G G G G G C P C C P C C C P C C C P C C C C	CCCC P CCCC P Ageo Q Q CCCC R CCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCC Q CCCCCC	2990 A IX4- 210 90 90 0 0 90 0 0 0 0 0 0 0 0 0 0 0 0	In the second se	G G G G G G G G G G G G G G G G G G G	rggat D tcgg R cgccg P gggtg V cccg P gggcg A ccctg L gggag E 8R accgg R

Figure 4. A putative 1208 bp CIC-DUX4 fusion transcript which would have been amplified using the the forward CIC-4105F and reverse DUX4-1538R primers. All the primers used in the study are denoting the primers sequences (in box) together with orientation (arrows). The search sequences "GCCGCCTTCCAGGCCCGCTA" ("grep" 1) and "CAGGGGGCCCTGACCCCACC" ("grep" 2) used as search terms in the "grep" command-line utility are colored yellow and in box. The fusion point between CIC and DUX4 is in red. The part of the protein coded by this CIC-DUX4 fusion transcript fragment is shown under the nucleotide sequence. The nucleotide sequence has been deposited in the GenBank with accession number KJ670706. doi:10.1371/journal.pone.0099439.g004

PAFAH1B3, PRR19, TMEM145, MEGF8, CNFN, and LIPE were present in the BAC bridging the breakpoint and could conceivably be the gene-level target of the chromosomal split. It was therefore surprising that no signs of any CIC-DUX4 were evident when we analyzed the raw sequencing data using ChimeraScan [18], FusionMap [19], and FusionFinder [20], fusion-finder programs that have all been evaluated recently on a synthetic dataset as well as real datasets that included experimentally validated chimeras [27,28]. All three programs produced a plethora of fusion transcripts but none of them contained *CIC* or any of the other 8 genes found in the split RP11-556K23 FISH probe. We then as a last resort decided to search for *CIC* sequences in the whole transcriptome sequencing data set using the "grep" command-line utility. The rationale was: 1) the RNA sequencing data are in fastq format files (filename.fastq) and fastq is a text-based format (http://en.wikipedia.org/wiki/FASTQ_format) and 2) the sequence data can be searched using the "grep" command-line utility (http://en.wikipedia.org/wiki/Grep). The "grep" command-line utility is used for searching text or a file for specific expressions. By default, "grep" displays the lines where matches occur. Our "specific expression" was a sequence of 20 nucleotides from the coding part of the last exon (20) of CIC (Reference Sequence: NM 015125.3) since all the so far reported CIC breakpoints have occurred in that part of the CIC gene [5,12–14]. The sequences obtained by "grep" were blasted against the human genomic plus transcript database (http://blast.ncbi.nlm. nih.gov/Blast.cgi) in order to identify possible chimeric fragments containing part of CIC and part of another gene.

This approach allowed us to obtain from the RNA sequencing fastq file 15 chimeric *CIC-DUX4* cDNA sequences (Table 2) and to map the fusion between the *CIC* and *DUX4* genes precisely. Subsequently, four more chimeric *CIC-DUX4* sequences were identified using a 20-mer sequence containing the fusion point as "specific expression" in the "grep" command-line utility. The fusion occurred between nt 4724 of *CIC* mRNA reference sequence NM_015125.3 and nt 771 of *DUX4* mRNA reference sequence NM_033178.4. This fusion has not been reported before [5,12–14]. *CIC* fusions have been reported at nt 4552, 4579, 4740, 4750 [12–14,16] and for *DUX4* at nt 1071, 1078, and 1145 of the reference sequence with accession number NM_033178.4 [5,12–14].

An explanation for the failure of the initial PCR is that the target *CIC-DUX4* chimeric sequence between CIC-4105F/DUX4-1538R primers was 1208 bp long with 70% CG content (Figure 4). The primer combinations CIC-4377F/DUX4-1151R and CIC-4453F/DUX4-1053R together with a PrimeSTAR GXL DNA polymerase, suitable for GC-rich templates, amplified fragments 546 bp long with 70% CG content and 374 bp long with 69% CG content, respectively (Figures 3B and 4). Sanger sequencing verified that they were *CIC-DUX4* fusion transcripts which had the same fusion point found with the "grep" command-line utility.

References

- Richkind KE, Romansky SG, Finklestein JZ (1996) t(4;19)(q35;q13.1): a recurrent change in primitive mesenchymal tumors? Cancer Genet Cytogenet 87: 71–74.
- Urumov IJ, Manolova Y (1992) Cytogenetic analysis of an embryonal rhabdomyosarcoma cell line. Cancer Genet Cytogenet 61: 214–215.
- Roberts P, Browne CF, Lewis IJ, Bailey CC, Spicer RD, et al. (1992) 12q13 abnormality in rhabdomyosarcoma. A nonrandom occurrence? Cancer Genet Cytogenet 60: 135–140.
- Somers GR, Shago M, Zielenska M, Chan HS, Ngan BY (2004) Primary subcutaneous primitive neuroectodermal tumor with aggressive behavior and an unusual karyotype: case report. Pediatr Dev Pathol 7: 538–545.
- Kawamura-Saito M, Yamazaki Y, Kaneko K, Kawaguchi N, Kanda H, et al. (2006) Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. Hum Mol Genet 15: 2125–2137.
- Gabriëls J, Beckers MC, Ding H, De Vriese A, Plaisance S, et al. (1999) Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene 236: 25–32.
- van Geel M, Dickson MC, Beck AF, Bolland DJ, Frants RR, et al. (2002) Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. Genomics 79: 210–217.
- Dixit M, Ansseau E, Tassin A, Winokur S, Shi R, et al. (2007) DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. Proc Natl Acad Sci U S A 104: 18157– 18162.
- Kowaljow V, Marcowycz A, Ansseau E, Conde CB, Sauvage S, et al. (2007) The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscul Disord 17: 611–623.

Current knowledge about the *CIC-DUX4* fusion holds that in the encoding protein CIC is mostly preserved and retains its HMG-box domain fusion, while DUX4 has lost most of its sequence, including its two DNA-binding homeodomains [5,11,12,14]. As a consequence of the fusion the transcriptional activity of CIC is enhanced, suggesting an abnormal regulation of downstream targets [5]. CIC–DUX4 directly binds the ERM promoter by recognizing a novel target sequence and significantly up-regulates its expression [5]. Mashado et al [16], on the other hand, described an undifferentiated small round cell sarcoma in which *CIC-DUX4* coded for a putative truncated CIC protein. In that case, the last 104 amino acid residues of CIC protein were deleted and *DUX4* contributed a triplet followed by a stop codon. It is not known whether this truncated CIC protein would have resulted in an enhanced transcriptional activity of CIC.

In conclusion, our study showed that the three fusion-finder programs FusionMap [19], Fusion Finder [20], and ChimeraScan [18] generated a plethora of fusion transcripts but not the biologically important and cancer-specific fusion gene, the *CIC-DUX4* chimeric transcript. It was necessary to use the "grep" command-line utility to sift out the latter from the many data produced by the automated algoritms. Cytogenetic, FISH, and clinico-pathologic tumor features hinted at the presence of the said fusion, but it was eventually found only after the manual "grep"function had been used.

Supporting Information

Table S1Fusion transcripts detected using FusionMap.(XLSX)

Table S2Fusion transcripts detected using FusionFinder.(XLSX)

 Table S3
 Fusion transcripts detected using ChimeraScan.

 (XLSX)
 (XLSX)

Author Contributions

Conceived and designed the experiments: IP. Performed the experiments: IP LG BB. Analyzed the data: IP. Contributed reagents/materials/analysis tools: BB SH. Wrote the paper: IP SH.

- Rakheja D, Goldman S, Wilson KS, Lenarsky C, Weinthal J, et al. (2008) Translocation (4;19)(q35;q13.1)-associated primitive round cell sarcoma: report of a case and review of the literature. Pediatr Dev Pathol 11: 239–244.
- Yoshimoto M, Graham C, Chilton-MacNeill S, Lee E, Shago M, et al. (2009) Detailed cytogenetic and array analysis of pediatric primitive sarcomas reveals a recurrent CIC-DUX4 fusion gene event. Cancer Genet Cytogenet 195: 1–11.
- Graham C, Chilton-MacNeill S, Zielenska M, Somers GR (2012) The CIC-DUX4 fusion transcript is present in a subgroup of pediatric primitive round cell sarcomas. Hum Pathol 43: 180–189.
- Italiano A, Sung YS, Zhang L, Singer S, Maki RG, et al. (2012) High prevalence of CIC fusion with double-homeobox (DUX4) transcription factors in EWSR1negative undifferentiated small blue round cell sarcomas. Genes Chromosomes Cancer 51: 207–218.
- Choi EY, Thomas DG, McHugh JB, Patel RM, Roulston D, et al. (2013) Undifferentiated small round cell sarcoma with t(4;19)(q35;q13.1) CIC-DUX4 fusion: a novel highly aggressive soft tissue tumor with distinctive histopathology. Am J Surg Pathol 37: 1379–1386.
- Kajtár B, Tornóczky T, Kálmán E, Kuzsner J, Hogendoorn PC, et al. (2013) CD99-positive undifferentiated round cell sarcoma diagnosed on fine needle aspiration cytology, later found to harbour a CIC-DUX4 translocation: a recently described entity. Cytopathology 25: 129–32.
- Machado I, Cruz J, Lavernia J, Rubio L, Campos J, et al. (2013) Superficial EWSR1-negative undifferentiated small round cell sarcoma with CIC/DUX4 gene fusion: a new variant of Ewing-like tumors with locoregional lymph node metastasis. Virchows Arch 463: 837–842.
- Wang Q, Xia J, Jia P, Pao W, Zhao Z (2012) Application of next generation sequencing to human gene fusion detection: computational tools, features and perspectives. Brief Bioinform 14: 506–19.

- Iyer MK, Chinnaiyan AM, Maher CA (2011) ChimeraScan: a tool for identifying chimeric transcription in sequencing data. Bioinformatics 27: 2903– 2904.
- Ge H, Liu K, Juan T, Fang F, Newman M, et al. (2011) FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. Bioinformatics 27: 1922–1928.
- Francis RW, Thompson-Wicking K, Carter KW, Anderson D, Kees UR, et al. (2012) FusionFinder: a software tool to identify expressed gene fusion candidates from RNA-Seq data. PLoS One 7: e39987.
- Panagopoulos I, Thorsen J, Gorunova L, Haugom L, Bjerkehagen B, et al. (2013) Fusion of the ZC3H7B and BCOR genes in endometrial stromal sarcomas carrying an X;22-translocation. Genes Chromosomes Cancer 52: 610– 618.
- Nyquist KB, Panagopoulos I, Thorsen J, Haugom L, Gorunova L, et al. (2012) Whole-Transcriptome Sequencing Identifies Novel IRF2BP2-CDX1 Fusion Gene Brought about by Translocation t(1;5)(q42;q32) in Mesenchymal Chondrosarcoma. PLoS One 7: e49705.
- Panagopoulos I, Thorsen J, Gorunova L, Micci F, Haugom L, et al. (2013) RNA sequencing identifies fusion of the EWSR1 and YY1 genes in mesothelioma with t(14;22)(q32;q12). Genes Chromosomes Cancer 52: 733–740.
- Mandahl N (2001) Methods in solid tumour cytogenetics. In: Rooney DE, editor. Human cytogenetics: malignancy and acquired abnormalities. New York: Oxford University Press. pp. 165–203.
- Schaffer LG, Slovak ML, Campbell LJ (2009) ISCN 2009: an International System for Human Cytogenetic Nomenclature. Basel: Karger.
- Field D, Tiwari B, Booth T, Houten S, Swan D, et al. (2006) Open software for biologists: from famine to feast. Nat Biotechnol 24: 801–803.
- Carrara M, Beccuti M, Lazzarato F, Cavallo F, Cordero F, et al. (2013) State-ofthe-art fusion-finder algorithms sensitivity and specificity. Biomed Res Int 2013: 340620.
- Carrara M, Beccuti M, Cavallo F, Donatelli S, Lazzarato F, et al. (2013) State of art fusion-finder algorithms are suitable to detect transcription-induced chimeras in normal tissues? BMC Bioinformatics 14 Suppl 7: S2.