
Article:
Hoff, A.M., Alagaratnam, S., Zhao, S. et al. (2016) Identification of Novel Fusion Genes in Testicular Germ Cell Tumors. Cancer Research, 76 (1). pp. 108-116. ISSN: 0008-5472

https://doi.org/10.1158/0008-5472

Published in final edited form as:
Cancer Res. 2016 January 1; 76(1): 108–116. doi:10.1158/0008-5472.CAN-15-1790.

Identification of novel fusion genes in testicular germ cell tumors

Andreas M. Hoff[1,2], **Sharmini Alagaratnam**[1,2], **Sen Zhao**[1,2], **Jarle Bruun**[1,2], **Peter W. Andrews**[3,4], **Ragnhild A. Lothe**[1,2], and **Rolf I. Skotheim**[1,2,†]

[1]Department of Molecular Oncology, Institute for Cancer Research, Oslo University Hospital-Norwegian Radium Hospital, Oslo, Norway

[2]Centre for Cancer Biomedicine, University of Oslo, Oslo, Norway

[3]Department of Biomedical Science, University of Sheffield, Western Bank, Sheffield, United Kingdom.

[4]Centre for Stem Cell Biology, University of Sheffield, Western Bank, Sheffield, United Kingdom

Abstract

Testicular germ cell tumors (TGCT) are the most frequently diagnosed solid tumors in young men ages 15 to 44 years. Embryonal carcinomas (EC) comprise a subset of TGCTs that exhibit pluripotent characteristics similar to embryonic stem (ES) cells, but the genetic drivers underlying malignant transformation of ECs are unknown. To elucidate the abnormal genetic events potentially contributing to TGCT malignancy, such as the existence of fusion genes or aberrant fusion transcript expression, we performed RNA sequencing of EC cell lines and their non-malignant ES cell line counterparts. We identified eight novel fusion transcripts and one gene with alternative promoter usage, *ETV6*. Four out of nine transcripts were found recurrently expressed in an extended panel of primary TGCTs and additional EC cell lines, but not in normal parenchyma of the testis, implying tumor-specific expression. Two of the recurrent transcripts involved an intrachromosomal fusion between *RCC1* and *HENMT1* located 80 Mbp apart and an interchromosomal fusion between *RCC1* and *ABHD12B*. *RCC1-ABHD12B* and the *ETV6* transcript variant were found to be preferentially expressed in the more undifferentiated TGCT subtypes. *In vitro* differentiation of the NTERA2 EC cell line resulted in significantly reduced expression of both fusion transcripts involving *RCC1* and the *ETV6* transcript variant, indicating that they are markers of pluripotency in a malignant setting. In conclusion, we identified eight novel fusion transcripts that, to our knowledge, are the first fusion genes described in TGCT and may therefore potentially serve as genomic biomarkers of malignant progression.

Keywords

TGCT; RNA-seq; ddPCR; fusion genes; stem cells

[†]**Corresponding author** Rolf I. Skotheim, Department of Molecular Oncology, Institute for Cancer Research, Oslo University Hospital-Radiumhospitalet, P. O. Box 4950 Nydalen, 0424 Oslo, Norway. rolf.i.skotheim@rr-research.no, phone: +47 2278 1727 .

Conflict of interest
The author(s) declare that they have no conflict of interest.

Europe PMC Funders Author Manuscripts

Introduction

Testicular germ cell tumors (TGCTs) are the most common cancer in young men ages 15-44 years (1). Although it is a highly treatable cancer type, exemplified by a 10-year net survival rate of 98 % in England and Wales (2), the disease affects men in their prime and treatment can lead to substantially increased morbidity, including cardiovascular disease, reduced fertility and secondary cancers (3). Histologically, there are two main subtypes of TGCTs, seminomas and non-seminomas. Both are thought to develop from the pre-invasive stage termed intratubular germ cell neoplasia (IGCN; also known as carcinoma *in situ).* Non-seminomas are further divided into the pluripotent embryonal carcinomas (EC) and more differentiated subtypes, with either somatic (teratoma) or extra-embryonic differentiation (yolk sac tumors, YST, and choriocarcinomas)(4).

EC cells are highly similar to embryonic stem (ES) cells, derived from the inner cell mass of the blastocyst stage embryo (5). Both cell types exhibit pluripotent characteristics phenotypically and in gene expression profiles (6,7). Upon extended passaging *in vitro*, ES cells have been shown to acquire genetic changes similar to those seen in malignant transformation *in vivo* of TGCTs and EC, including gain of genetic material from chromosomes 12, 17, and X (8). Gain of chromosome arm 12p, often as an isochromosome, i(12p), is found in virtually all cases of TGCT (9,10). Crucially, despite these similarities, EC cells are malignant in character, whereas ES cells are not. Comparative studies between the two cell types may therefore be useful for characterization of cancer-specific differences in a pluripotent context (5,11). One such study revealed that several transcription factors located on 12p are overexpressed in EC cells as compared to ES cells (6). Although 12p material is gained in virtually all cases of TGCT, no clear genetic driver for TGCT malignant transformation has been pinpointed (12,13).

Recently, whole-exome sequencing studies have revealed that the number of non-synonymous mutations in coding regions of the TGCT genome are few, on a scale similar to that of pediatric cancers (14–16). A number of pediatric cancers with a low mutational load are frequently found to harbor fusion genes with oncogenic properties. Examples are *MLL* rearrangements in acute lymphoblastic leukemia (17), and subtypes of sarcomas classified by distinct chromosomal translocations (18). In Ewing sarcoma, fusions involving *EWSR1* are pathognomonic, while the mutation rate is low, estimated at 0.15/Mb of coding sequence (19). In this study, we have performed RNA sequencing of EC cell lines and their non-malignant counterpart, ES cell lines. Application of a fusion gene analysis pipeline led to the identification of nine novel fusion genes and transcripts, to our knowledge the first described in TGCT.

Material and Methods

Cell lines and patient samples

Three EC cell lines (2102Ep, 833KE, and NTERA2) and 2 ES cell lines (H9 and Shef3) were subjected to RNA sequencing. The EC and ES cell lines were established in the lab of Peter W. Andrews, University of Sheffield, where they also were grown and sorted for expression of the pluripotency marker SSEA3 as previously described (11). The extended

experimental validation panel consisted of four categories of samples: 1) 2 additional EC (Tera 1 and NCCIT) and 2 additional ES (Shef6 and Shef7) cell lines (n=4), 2) NTERA2 and 2102Ep cells treated with all-*trans* retinoic acid (RA) for 0, 3 and 7 days to induce differentiation, as previously described (n=6) (20,21). 3) Thirty-five testicular tissue samples including 5 normal testicular parenchyma, 6 premalignant IGCN and 24 primary TGCTs, all with only one histological subtype each; EC (n=8), seminoma (n=7), choriocarcinoma (n=1), YST (n=4), and teratoma (n=4). 4) Twenty normal tissues from miscellaneous sites of the body were used for exploration of cancer-specificity of the novel transcripts (adipose, bladder, brain, cervix, colon, esophagus, heart, kidney, liver, lung, ovary, placenta, prostate, skeletal muscle, spleen, stomach, testes, thymus, thyroid and trachea; FirstChoice Human Normal Tissue Total RNA). These were each a pool of RNA from at least three individuals, with the exception of one individual sample from the stomach (Ambion, Applied Biosystems by Life Technologies, Carlsbad, CA, USA).

DNA isolated simultaneously from the cell pellets was tested and authenticated by STR fingerprinting using the AmpFLSTR Identifiler PCR Amplification Kit (Applied Biosystems). Profiles positively matched with those reported in the literature for 2102Ep (7), and obtained from the European Collection of Cell Cultures (ECACC; for 833KE), ATCC (for NTERA2, NCCIT, and TERA1), the Wisconsin International Stem Cell Bank (H9), and the UK Stem Cell Bank (Shef3, Shef6, and Shef7). The biobank is registered according to Norwegian legislation (no. 953; Biobank Registry of Norway) and the project has been approved by the National Committee for Medical and Health Research Ethics (S-05368 and S-07453b).

External data for *in silico* validation

Paired-end RNA sequencing data from the Illumina Human Body Map v2 dataset, consisting of 16 non-malignant miscellaneous tissue types, was analyzed as an additional source of normal controls (ArrayExpress accession ID E-MTAB-513 and European Nucleotide Archive study accession ID ERP000546).

Paired end RNA-sequencing of EC and ES cell lines

Library construction was performed using the standard Illumina mRNA library preparation protocol (Illumina Inc, San Diego, CA, USA), including poly-A mRNA isolation, fragmentation, and gel-based size selection. Shearing to about 250 bp fragments was achieved using the Covaris S2 focused ultrasonicator (Covaris Inc, Woburn, MA, USA). Paired-end sequencing, 76 bp from each end, was performed according to protocol on a Genome Analyzer IIx (Illumina Inc.).

Fusion transcript identification

To identify fusion transcripts specific for EC, we used the fusion detection algorithm deFuse v. 0.6.1 (22) with hg19 sequence reference from UCSC and Ensembl release 69 annotation. To enrich for true positive fusion transcripts specifically expressed in EC cells, several heuristic filtering steps of the initial fusion breakpoint candidates were performed, some adapted from the recommended procedures of the original publication of deFuse (22), namely: 1) Only the nominated fusion breakpoint candidates with a probability score greater

than 0.5 were considered. 2) Breakpoints nominated in EC that were also found in ES cell lines or tissues of the Human Body Map v2 were removed to enrich for malignancy-specific candidates and remove systematic technical artifacts. 3) We removed candidates that had more than 5 multi-mapping spanning reads, or a ratio of multi-mapping spanning reads greater than 25 %. 4) To filter out candidates nominated due to homologous or repeat sequences, we removed candidates that had a deFuse homology score greater than 10 and candidates having either of the following three criteria: cDNA adjusted-, genome adjusted- or EST adjusted percent identity greater than 0.1. In an effort to enrich for functionally interesting fusion candidates, we applied 4 functional filters of which candidates needed to pass 3 in order to proceed. 1) Both gene partners are annotated as protein-coding genes. 2) The number of split reads and spanning reads supporting the fusion breakpoint sequence is greater than or equal to 3 or 5, respectively. 3) The fusion breakpoint includes 5′ UTR or coding parts of at least one of the partner genes. 4) The distance between the two partner genes is greater than 30 kb. We also used an additional fusion finder algorithm, SOAPfuse v. 1.26 (23), and included all fusion breakpoints that were picked up by both deFuse and SOAPfuse independently. For the remaining fusion transcripts, breakpoint alignments were evaluated in the UCSC genome browser and the Integrative Genomic Viewer (IGV). In general, we removed chimeric breakpoint sequences likely to derive from read-through transcripts, and those not aligning to conserved exon to exon boundaries.

Validation of fusion transcript breakpoints by reverse-transcriptase PCR and Sanger sequencing

Selected fusion transcript candidates were validated with reverse transcription PCR (RT-PCR) in the RNA-sequenced cell lines and in an extended validation panel. Primers were designed to the fusion transcript breakpoint sequences as detected by deFuse by using the Primer3 web application (24). All primer sequences used in this study are listed in Supplementary Table S1. Briefly, reverse transcription was performed using the high-capacity reverse transcription kit according to protocol (Applied Biosystems by Life Technologies, CA, USA). From 50 ng of starting cDNA template, a PCR protocol was initiated with 15 minutes of HotStarTaq DNA polymerase activation at 95°C, followed by 30 thermal cycles of denaturation for 30 seconds at 95°C, primer annealing for 1 minute at optimal primer melting temperatures (Supplementary Table S1), and extension for one minute at 72°C. After the last cycle, a final extension step was performed at 72°C for 10 minutes. The PCR products were separated by electrophoresis at 200 V for 30 minutes on a 2 % agarose gel and visualized using ethidium bromide and UV light.

To ensure specific amplification of the breakpoint sequences, PCR products from the cell lines that were nominated by RNA-seq to harbor the individual fusion transcripts were sequenced by Sanger sequencing. PCR products that showed a single nucleotide band on the agarose gel were sequenced directly from both sides using forward and reverse primers. Prior to sequencing, the PCR products were purified using Illustra ExoStar 1-step cleanup (GE Healthcare, Little Chalfront, UK). The cycle sequencing reactions were performed using the BigDye Terminator v.1.1 cycle sequencing kit (Applied Biosystems, Foster City, CA, USA) following manufacturer's recommendations. The sequencing products were purified using BigDye Xterminator (Applied Biosystems) before being analyzed by capillary

electrophoresis using the ABI 3730 DNA Analyzer (Applied Biosystems). The resulting sequences were analyzed using the Sequencing Analysis v.5.3.1 software.

The quantity of fusion transcripts assessed by TaqMan real-time PCR

Several fusion transcripts confirmed by regular RT-PCR were recurrent, however with varying nucleotide band intensities between samples, as observed by agarose gel electrophoresis. We performed TaqMan quantitative RT-PCR (qRT-PCR) to quantify the relative expression of these fusion transcripts. Primers and MGB-probes were designed with the Primer Express v.3.0 software (Applied Biosystems) to cross the fusion transcript boundaries (Supplementary Table S1). Two endogenous control assays targeting *ACTB* and *GUSB* were analyzed in all samples to normalize for input template amounts. The qRT-PCR reactions were performed in reaction volumes of 10 μl, with 15 ng of template cDNA, TaqMan universal mastermix II with uracil-N-glycosylase (Applied Biosystems) and final primer and probe concentrations of 0.9 μM and 0.2 μM, respectively. The PCR reactions were run in triplicate on an ABI 7900HT fast real-time PCR system (Applied Biosystems). Expression levels were reported as the median cycle threshold (C_T) of the triplicates and normalized to median C_T values of the endogenous controls. A threshold value at $C_T = 35$ was set for all assays as positive expression.

Assessment of DNA-level fusions with multiplexed droplet digital PCR

Droplet digital PCR (ddPCR) takes advantage of oil/water emulsion, separating PCR reagents and template into thousands of nano-liter sized droplets. Subsequent thermal cycling by traditional fluorescent PCR specifically amplifies target templates in the droplets. The number of target molecules in a reaction mixture is inferred by counting the number of droplets with and without amplified fluorescent signal. It is also possible to measure two target molecules simultaneously, by multiplexing 2 PCR assays with different fluorescent dyes (FAM and VIC/HEX). Since template molecules distribute randomly into droplets, droplets can be expected to contain one, the other, both or none of the target molecules by chance in a multiplexed assay. However, if two template targets are located in close proximity on the same DNA molecule and thereby linked, these would distribute together in a non-random fashion with a higher number of double positive droplets than expected by chance.

Here, we performed ddPCR to investigate DNA-level linkage of the partner genes of 2 recurrent fusion transcripts, as well as 2 fusion transcripts that each were expressed only in one EC cell line. As proof of concept, we included duplex linkage assays for the known fusion *VTI1A-TCF7L2*, which in the NCI-H508 cell line is known to be formed by a genomic deletion (25). To evaluate the integrity of DNA fragments, and as an additional positive control of the ddPCR linkage approach, we used a custom milepost experiment which measures linkage with assays 1 kb, 10 kb, 50 kb and 100 kb apart. In all experiments, a reference assay with FAM fluorescence was multiplexed with one of the milepost assays with VIC fluorescence. FAM and VIC assays were also designed for the two partner genes of the fusion transcripts. All assays used in the ddPCR linkage experiments are listed in Supplementary Table S2. As control experiments for the linkage assays we performed fragmentation of genomic DNA with the NspI restriction endonuclease. We ensured that

none of the assay's target sequences overlapped with the restriction enzyme target sequence. Each ddPCR experiment was carried out in 22 μl reaction volumes, with final concentrations of 0.9 μM of each primer and 0.25 μM probe, 1x ddPCR supermix (Bio-Rad) and 25 to 50 ng genomic DNA. Droplet generation was performed with 20 μL of the reaction mix, according to the manufacturer's protocol. Droplets were then transferred to a 96-well plate and PCR performed with the following thermal cycling profile: initial enzyme activation at 95°C for 10 minutes, followed by 40 cycles of denaturation at 94°C for 30 seconds and annealing/extension at 60°C for 1 minute. As a final step, the enzyme was deactivated at 98°C for 10 min. The droplets were read using the QX200 droplet reader according to manufacturer's protocol. The data was analyzed using the QuantaSoft software (1.7.4.0917; Bio-Rad). Crosshair gating was used to set a threshold for the four quadrants of droplet populations: double-negative, FAM-positive, VIC-positive and double-positive. QuantaSoft outputs the concentration in molecules/μl for each of the assays. Additionally, a linkage concentration is calculated based on the ratio of double-positive droplets, given in linked molecules/μl. We calculated percent linkage as the concentration of linked molecules divided by the mean concentration of the individual assays transformed to percentage.

Results

Identification of fusion transcripts in EC cell lines from paired-end RNA-seq data

RNA sequencing of the three EC (2102Ep, 833KE, and NTERA2) and two ES (H9 and Shef3) cell lines generated a total of 199 million pairs of 76 bp sequencing reads that passed filtering (Supplementary Table S3).

Fusion transcript analyses of the RNA-seq data resulted in an initial list of 1210 unique fusion breakpoints with a probability score above 0.5. Subsequent heuristic filtering nominated nine fusion transcripts which were considered strong enough for further experimental validation (Figure 1). Briefly, 283 candidate fusions were first removed as they were also detected in ES cell lines or normal human tissues (i.e. external data from the Illumina Human Body Map v2 data set). Further technical filtering of fusion breakpoints with a high ratio of multi-mapping spanning reads and breakpoints with a high degree of homology or breakpoint sequence identity, removed 621 and 130 candidates respectively. To enrich for functionally important breakpoints, we removed breakpoints that did not pass at least three out of four functional filters. As an additional step, we used the SOAPfuse fusion finder to identify a list of 85 potential EC fusion breakpoint candidates. Of these, only 11 overlapped with the initial list of EC specific breakpoints generated by deFuse, where five of these were already kept through the filtering process. The remaining six overlapping candidates were retrieved for evaluation in the final candidate list. After filtering steps, a list of 65 unique candidate fusion breakpoints remained, and was manually curated by viewing alignments in IGV and the UCSC genome browser. Fusion transcripts likely to be generated by polymerase read-through were filtered out, except for a read-through between *CLEC6A* and *CLEC4D* located on chromosome arm 12p found to be recurrent in all three EC cell lines. Fusion candidates where the breakpoint did not match intact conserved exon – exon boundaries were filtered out, except for a breakpoint, *ETV6-RP11-434C1.1,* also located on chromosome arm 12p, which was not strictly a fusion transcript but a transcript produced

from an unannotated alternative promoter. This alternative promoter of *ETV6* was specially considered based on the known oncogenic relevance of the ETS family of transcription factors (26). The final list of transcripts selected for experimental validation consisted of two inter-chromosomal and seven intra-chromosomal fusion transcript candidates (Figure 2, Table 1). Of the seven intra-chromosomal fusions, five included breakpoints with both partner genes located on chromosome arm 12p.

Technical and clinical validation of the fusion transcripts

We performed RT-PCR to validate the presence of the nine nominated fusion transcript breakpoints in the EC and ES cell lines investigated, and for further clinical evaluation in a series of IGCN and TGCTs. All nine nominated fusion transcripts were confirmed by RT-PCR spanning the breakpoint. Successful Sanger sequences were produced from eight of these, confirming breakpoint sequences between the two gene pairs, and all found to use intact exon – exon boundaries (Supplementary Figure S1). The eight fusion transcripts as well as the alternative promoter usage of *ETV6* all had intact open reading frames (ORFs), theoretically encoding functional proteins. The ORFs of five out of nine encode N-terminally truncated proteins of the downstream partner gene, while three out of nine encode the full length coding sequence of the downstream partner. None of the fusion transcripts encode potential hybrid proteins with in-frame coding sequence from both intact partner proteins. However, one of the fusion transcripts, *PPP6R3-DPP3,* encodes an out-of-frame ORF encoding 198 amino acids.

Five of the fusion transcripts were only expressed in the originally nominated EC cell lines, and were thus considered private fusion events. The remaining four candidates were however found to be recurrently expressed in TGCT (Supplementary Table S4), and crucially, not in normal testicular parenchyma. The four recurrent candidates included the read-through between *CLEC6A* and *CLEC4D*, alternative promoter usage of *ETV6,* and two fusion transcripts both involving the first two exons of *RCC1* as an upstream partner gene connected to *HENMT1* and *ABHD12B,* located 80 Mbp apart on chromosome 1 and on chromosome 14 respectively. Expression of these recurrent transcripts was variable, with both strongly positive and weaker bands detected by agarose gel electrophoresis. For more accurate assessment of expression, we used qRT-PCR to quantify the expression level of each fusion transcript. All custom TaqMan assays were found to have efficiencies between 80-90 %. Here, a total of 73 tissue samples and cell lines were tested. None of the recurrent transcripts were expressed in normal testicular parenchyma (n = 6). The read-through between *CLEC6A* and *CLEC4D* was found to be expressed in all subtypes of TGCT, as well as pre-malignant IGCN (6/6) and ES cell lines (3/3). However, only two of the four teratoma tissue samples showed expression of the read-through. The read-through was only detected in one (the placenta) of 20 normal tissues from the human body (Figure 3A). Alternative promoter usage of the *ETV6* gene was predominantly detected in EC tissue samples and cell lines (75 % and 92 %, respectively; 6/8 and 12/13). Alternative promoter usage was also observed in ES cell lines (100 %; 3/3), seminoma (14 %; 1/7), Cc (100 %; 1/1) and YST (50 %; 2/4). None of the 20 included normal tissues expressed the alternative promoter of *ETV6* (Figure 3B). The intrachromosomal fusion transcript *RCC1-HENMT1* was widely expressed in all subtypes of TGCT, IGCN, ES cell lines, and in 6/20 normal tissue types

(spleen, esophagus, trachea, thyroid, thymus and skeletal muscle; Figure 3C). By contrast, the interchromosomal fusion transcript *RCC1-ABHD12B* was found expressed predominantly in EC tissue samples and cell lines (100 %) and in seminomas (86 %; 6/7). *RCC1-ABHD12B* was further detected in 67 % (4/6) IGCN, in one of four teratomas and in one of three ES cell lines. None of the tested tissue samples from sites of the human body showed expression of this fusion transcript (Figure 3D).

***RCC1* involving fusion transcripts and alternative promoter usage in *ETV6* are associated with undifferentiated subtypes of TGCT—**Total RNA isolated from NTERA2 and 2102Ep cell lines treated with RA for 0, 3 and 7 days were included in the validation panel. The NTERA2 cell line has previously been shown to differentiate in culture when treated with morphogens such as RA (21). 2102Ep, on the other hand, does not differentiate upon *in vitro* treatment with RA and remains pluripotent (27). Intriguingly, qRT-PCR analyses of both fusion transcripts involving *RCC1* revealed that expression decreased upon treatment with RA in NTERA2, but not in 2102Ep (Figure 4). The measured $\Delta\Delta C_T$ was −3.9 and −3.0 between 0 and 7 days of RA treatment for *RCC1-ABHD12B* and *RCC1-HENMT1* respectively. Additionally, expression of the *ETV6* transcript involving the alternative promoter was silenced after 7 days of RA treatment ($C_T > 35$; $\Delta\Delta C_T$ of −6.8). Expression of the *CLEC6A-CLEC4D* read-through did not change significantly upon RA treatment in NTERA2. The expression of all fusion transcripts remained unchanged in the 2102Ep cell line upon treatment with RA (Figure 4), in line with this cell line's previously reported nonresponsiveness to RA treatment.

Linkage assays by ddPCR to identify coupling of fusion genes on the DNA-level

Because the two fusion transcripts involving *RCC1* were recurrently expressed, and the two partner genes were not located close to *RCC1*, the possibility of genome-level rearrangements as a mechanism resulting in gene fusion was tested by ddPCR linkage analysis. For another two fusion genes *EPT1-GUCY1A3* and *PPP6R3-DPP3*, expressed in the 833KE and NTERA2 cell lines respectively, DNA copy number data indicated a shift in the vicinity of all four genetic loci (data not shown).

To establish the integrity of DNA to be included in the linkage analyses, and as proof of concept, we performed a ddPCR milepost experiment. Here, multiplexed fluorescent TaqMan assays measured linkage 1 kb, 10 kb, 50 kb and 100 kb apart. DNA isolated by the AllPrep method from the NTERA2 cell line showed highest integrity, with 90 % linkage at 1 kb, 52 % at 10 kb and 11 % at 50 kb (Supplementary Figure S2). No evidence of linkage was seen at 100 kb. DNA from the 833KE cell line and DNA isolated by phenol-chloroform from NTERA2 showed slightly lower linkage scores, indicating more highly degraded DNA (Supplementary Figure S2). After DNA fragmentation with the NspI restriction endonuclease, all linkage was substantially reduced, but some background levels remained (0 – 3.5 %; Supplementary Figure S2). Fusion gene linkage analysis of *VTI1A-TCF7L2*, previously reported to be caused by a deletion in the NCI-H508 cell line, showed that 18.5 % of molecules are linked and contain both fusion partner targets (Figure 5). The interchromosomal fusion *EPT1-GUCY1A3* and the intrachromosomal fusion *PPP6R3-DPP3* showed evidence of DNA-level linkage, with 15 and 18 % linkage rates respectively (Figure

5). However, we found no evidence for DNA-level linkage for the recurrently expressed fusions *RCC1-ABHD12B* and *RCC1-HENMT* (Figure 5). As a control experiment, we found that the DNA-level linkage in all tested samples was lost upon digesting the DNA with the restriction enzyme NspI.

Discussion

In this study, we have identified the presence of fusion genes and transcripts in TGCT. Nine novel fusion genes and transcripts are reported, of which *RCC1-HENMT1, RCC1-ABHD12B, CLEC6A-CLEC4D* and alternative promoter usage of *ETV6* are recurrently expressed in a significant number of clinical TGCT samples. Whereas these were detected only on the transcript level, two DNA-level fusions were identified, *EPT1-GUCY1A3* and *PPP6R3-DPP3*, although these were privately expressed by individual EC cell lines.

The non-synonymous mutation rate in TGCTs has recently been found to be low, on a scale similar to pediatric cancers (14–16). The fact that few genes are recurrently mutated indicates that other molecular mechanisms are responsible for the development of TGCT. The genomes of TGCTs are generally aneuploid, with several recurrent gains and losses of chromosomal material (10,28). Structural rearrangements in these aneuploid tumors are known from cytogenetic banding analyses, and i(12p) is found in the majority of TGCT (9,29,30). Cases without i(12p) often have gain of parts of 12p material, and/or extra copies of the whole of chromosome 12 (31). Gain or loss of chromosomal material not only leads to increase or loss of gene copies and subsequent expression changes, but can also introduce genomic rearrangements that form fusion genes. The nine novel fusion genes and transcripts found in this study, all consisting of intact ORFs which potentially encode full-length or N-terminally truncated proteins, suggest that fusion genes play an important role, and may be drivers for the malignant development of TGCTs. *CLEC6A* and *CLEC4D,* found to be involved in a read-through expressed in a high number of clinical TGCT samples, and *ETV6* with a novel alternative promoter, are all located on chromosome arm 12p. In addition, all genes involved in the *CD9-ANO2*, *TSPAN9-FOXJ2* and *TSPAN9-GUCY2C* fusion transcripts, expressed in the 2102Ep EC cell line, are located on 12p. These findings indicate that 12p is a dynamic region in TGCTs, and that gain of 12p material may be associated with expression of recurrent fusion transcripts and transcript variants.

Both fusion transcripts that involved *RCC1* and the alternative promoter usage of *ETV6* are overrepresented in undifferentiated histological subtypes of TGCT, and show substantially reduced expression in the NTERA2 EC cell line upon RA-induced differentiation *in vitro*. This indicates that they are all associated with the pluripotent phenotype. *ETV6* (ets variant 6) encodes a transcription factor of the ETS family, a family of transcription factors that are frequently involved as fusion gene partners in cancer. *ETV6* itself has repeatedly been reported as a fusion gene partner in hematological malignancies (32). In fact, a chromosomal rearrangement (4;12)(q11-q12;p13) found in cases of acute myeloid leukemia fuses *CHIC2* (exon 1-3) and *ETV6* (exon 2-8) using the same breakpoint of *ETV6* as the one identified here connected to the newly identified alternative promoter sequence (33,34). *RCC1* (Regulator of chromosome condensation 1) has important functions in the cell cycle and acts as a guanine nucleotide exchange factor (GEF) for the RAS homologue RAN (35). However,

the fusion transcripts involving *RCC1* only includes the two first non-coding exons in the 5′ UTR of *RCC1* connected to either *HENMT1* or *ABHD12B* located 80 Mb downstream on chromosome 1 and on chromosome 14, respectively. The long genomic distance between the partner genes suggests that these fusion transcripts are not expressed as a result of a read-through mechanism (36). To investigate if the fusion transcripts involving *RCC1* were caused by genomic rearrangements, we applied linkage analysis using multiplexed fluorescent PCR assays with ddPCR. To our knowledge, this approach has not been used previously to detect rearrangements of genes resulting in fusion genes. However, linkage analysis with ddPCR has been proven successful in showing the arrangement of the Killer-cell immunoglobulin-like receptor gene complex and in chromosomal phasing (37,38). We found no indication of DNA-level linkage for the partner genes of the *RCC1* fusion transcripts. However, the partner genes of *EPT1-GUCY1A3* and *PPP6R3-DPP3*, which also had indications of chromosomal breakpoints from DNA copy number data, were found to be linked at the DNA-level, indicating bona-fide genomic rearrangements in their respective cell lines, 833KE and NTERA2. The absence of linked partner genes of the *RCC1* fusions, implies that these fusion transcripts are expressed as the result of post-transcriptional mechanisms, such as *trans*-splicing (39). However, we cannot rule out chromosomal rearrangements as an underlying cause of the *RCC1* fusion transcripts, since only a proportion of the input DNA in our assay consisted of DNA fragments longer than 50 kb, and none more than 100 kb in length. A chromosomal rearrangement resulting in a fusion gene may include intronic regions that are longer than these DNA fragments. Such rearrangements will be missed by ddPCR linkage analysis. For optimal sensitivity, linkage analysis should be carried out on DNA samples isolated by protocols that maintain long DNA molecules intact.

Expression of the interchromosomal *RCC1-ABHD12B* fusion transcript and transcripts involving the alternative promoter of *ETV6* was not detected in normal testis or other normal tissue samples from 20 different human organs. Also, expression of *CLEC6A-CLEC4D* was only observed in normal tissue from the placenta, indicating that it may be specifically expressed in adult male TGCT. These molecules are therefore highly specific for TGCTs in a stemness setting, and could prove to have important roles for TGCT malignant transformation, as well as biomarkers for TGCT disease. Diagnosis of TGCT through sensitive detection of these molecules in excreted body fluids such as seminal fluid or serum could have clinical potential (40).

In conclusion, to our knowledge, we present here the first fusion genes to be described in TGCT, including recurrent expression of *RCC1* involving fusions and alternative promoter usage of the *ETV6* gene, both associated with the pluripotency phenotype. These transcript variants may be important drivers of malignancy, and could potentially serve as diagnostic markers in the clinic.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support

The study was funded by grants from the Norwegian Cancer Society (A.M. Hoff was financed as PhD student from the PR-2007-0166 grant to R.I. Skotheim), NorStore (for storage of computer files; project NS9013K granted to R.I. Skotheim), the Medical Research Council UK (grant number G0700785 to P.W. Andrews), the Research Council of Norway through its Centers of Excellence funding scheme (project number 179571 granted to R.A. Lothe), and from the KG Jebsen Foundation (granted to R.A. Lothe).

References

1. Znaor A, Lortet-Tieulent J, Jemal A, Bray F. International Variations and Trends in Testicular Cancer Incidence and Mortality. Eur Urol. 2014; 65:1095–106. [PubMed: 24268506]
2. Quaresma M, Coleman MP, Rachet B. 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971-2011: a population-based study. Lancet. 2015; 385:1206–18. [PubMed: 25479696]
3. Haugnes HS, Bosl GJ, Boer H, Gietema JA, Brydøy M, Oldenburg J, et al. Long-Term and Late Effects of Germ Cell Testicular Cancer Treatment and Implications for Follow-Up. J Clin Oncol. 2012; 30:3752–63. [PubMed: 23008318]
4. Woodward, PJ.; Heidenreich, A.; Looijenga, LHJ. Germ cell tumours. In: Eble, JN.; International Agency for Research on Cancer. , editor. Pathol Genet Tumours Urin Syst Male Genit Organs. IARC Press; Lyon: 2006. p. 221-49.Reprint
5. Andrews PW, Matin MM, Bahrami AR, Damjanov I, Gokhale P, Draper JS. Embryonic stem (ES) cells and embryonal carcinoma (EC) cells: opposite sides of the same coin. Biochem Soc Trans. 2005; 33:1526–30. [PubMed: 16246161]
6. Sperger JM, Chen X, Draper JS, Antosiewicz JE, Chon CH, Jones SB, et al. Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. Proc Natl Acad Sci. 2003; 100:13350–5. [PubMed: 14595015]
7. Josephson R, Ording CJ, Liu Y, Shin S, Lakshmipathy U, Toumadje A, et al. Qualification of Embryonal Carcinoma 2102Ep As a Reference for Human Embryonic Stem Cell Research. Stem Cells. 2007; 25:437–46. [PubMed: 17284651]
8. Baker DE, Harrison NJ, Maltby E, Smith K, Moore HD, Shaw PJ, et al. Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. Nat Biotechnol. 2007; 25:207–15. [PubMed: 17287758]
9. Atkin NB, Baker MC. i(12p): specific chromosomal marker in seminoma and malignant teratoma of the testis? Cancer Genet Cytogenet. 1983; 10:199–204. [PubMed: 6616439]
10. Skotheim RI, Lothe RA. The testicular germ cell tumour genome. APMIS. 2003; 111:136–51. [PubMed: 12752254]
11. Alagaratnam S, Harrison N, Bakken AC, Hoff AM, Jones M, Sveen A, et al. Transforming pluripotency: an exon-level study of malignancy-specific transcripts in human embryonal carcinoma and embryonic stem cells. Stem Cells Dev. 2013; 22:1136–46. [PubMed: 23137282]
12. Looijenga LHJ, Zafarana G, Grygalewicz B, Summersgill B, Debiec-Rychter M, Veltman J, et al. Role of gain of 12p in germ cell tumour development. APMIS. 2003; 111:161–71. [PubMed: 12752258]
13. Sheikine Y, Genega E, Melamed J, Lee P, Reuter VE, Ye H. Molecular genetics of testicular germ cell tumors. Am J Cancer Res. 2012; 2:153–67. [PubMed: 22432056]
14. Brabrand S, Johannessen B, Axcrona U, Kraggerud SM, Berg KG, Bakken AC, et al. Exome sequencing of bilateral testicular germ cell tumors suggests independent development lineages. Neoplasia. 2015; 17:167–74. [PubMed: 25748235]
15. Litchfield K, Summersgill B, Yost S, Sultana R, Labreche K, Dudakia D, et al. Whole-exome sequencing reveals the mutational spectrum of testicular germ cell tumours. Nat Commun. 2015; 6:5973. [PubMed: 25609015]

16. Cutcutache I, Suzuki Y, Tan IB, Ramgopal S, Zhang S, Ramnarayanan K, et al. Exome-wide Sequencing Shows Low Mutation Rates and Identifies Novel Mutated Genes in Seminomas. Eur Urol. 2015; 68:77–83. [PubMed: 25597018]
17. Andersson AK, Ma J, Wang J, Chen X, Gedman AL, Dang J, et al. The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. Nat Genet. 2015; 47:330–7. [PubMed: 25730765]
18. Osuna D, de Alava E. Molecular pathology of sarcomas. Rev Recent Clin Trials. 2009; 4:12–26. [PubMed: 19149759]
19. Brohl AS, Solomon DA, Chang W, Wang J, Song Y, Sindiri S, et al. The Genomic Landscape of the Ewing Sarcoma Family of Tumors Reveals Recurrent STAG2 Mutation. PLoS Genet. 2014; 10
20. Skotheim RI, Lind GE, Monni O, Nesland JM, Abeler VM, Fosså SD, et al. Differentiation of human embryonal carcinomas in vitro and in vivo reveals expression profiles relevant to normal development. Cancer Res. 2005; 65:5588–98. [PubMed: 15994931]
21. Andrews PW. Retinoic acid induces neuronal differentiation of a cloned human embryonal carcinoma cell line in vitro. Dev Biol. 1984; 103:285–93. [PubMed: 6144603]
22. McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MGF, et al. deFuse: An Algorithm for Gene Fusion Discovery in Tumor RNA-Seq Data. PLoS Comput Biol. 2011; 7:e1001138. [PubMed: 21625565]
23. Jia W, Qiu K, He M, Song P, Zhou Q, Zhou F, et al. SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. Genome Biol. 2013; 14:R12. [PubMed: 23409703]
24. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 2000; 132:365–86. [PubMed: 10547847]
25. Bass AJ, Lawrence MS, Brace LE, Ramos AH, Drier Y, Cibulskis K, et al. Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. Nat Genet. 2011; 43:964–8. [PubMed: 21892161]
26. Sharrocks AD. The ETS-domain transcription factor family. Nat Rev Mol Cell Biol. 2001; 2:827–37. [PubMed: 11715049]
27. Matthaei KI, Andrews PW, Bronson DL. Retinoic acid fails to induce differentiation in human teratocarcinoma cell lines that express high levels of a cellular receptor protein. Exp Cell Res. 1983; 143:471–4. [PubMed: 6131831]
28. Gilbert D, Rapley E, Shipley J. Testicular germ cell tumours: predisposition genes and the male germ cell niche. Nat Rev Cancer. 2011; 11:278–88. [PubMed: 21412254]
29. Castedo SMMJ, Jong B de, Oosterhuis JW, Seruca R, Idenburg VJS, Dam A, et al. Chromosomal Changes in Human Primary Testicular Nonseminomatous Germ Cell Tumors. Cancer Res. 1989; 49:5696–701. [PubMed: 2551494]
30. Castedo SMMJ, Jong B de, Oosterhuis JW, Seruca R, te Meerman GJ, Dam A, et al. Cytogenetic Analysis of Ten Human Seminomas. Cancer Res. 1989; 49:439–43. [PubMed: 2910461]
31. Kraggerud SM, Skotheim RI, Szymanska J, Eknæs M, Fosså SD, Stenwig AE, et al. Genome profiles of familial/bilateral and sporadic testicular germ cell tumors. Genes Chromosomes Cancer. 2002; 34:168–74. [PubMed: 11979550]
32. De Braekeleer E, Douet-Guilbert N, Morel F, Le Bris M-J, Basinko A, De Braekeleer M. ETV6 fusion genes in hematological malignancies: A review. Leuk Res. 2012; 36:945–61. [PubMed: 22578774]
33. Cools J, Bilhou-Nabera C, Wlodarska I, Cabrol C, Talmant P, Bernard P, et al. Fusion of a Novel Gene, BTL, to ETV6 in Acute Myeloid Leukemias With a t(4;12)(q11-q12;p13). Blood. 1999; 94:1820–4. [PubMed: 10477709]
34. Cools J, Mentens N, Marynen P. A new family of small, palmitoylated, membrane-associated proteins, characterized by the presence of a cysteine-rich hydrophobic motif. FEBS Lett. 2001; 492:204–9. [PubMed: 11257495]
35. Hadjebi O, Casas-Terradellas E, Garcia-Gonzalo FR, Rosa JL. The RCC1 superfamily: from genes, to function, to disease. Biochim Biophys Acta. 2008; 1783:1467–79. [PubMed: 18442486]
36. Frenkel-Morgenstern M, Lacroix V, Ezkurdia I, Levin Y, Gabashvili A, Prilusky J, et al. Chimeras taking shape: Potential functions of proteins encoded by chimeric RNA transcripts. Genome Res. 2012; 7:1231–42. [PubMed: 22588898]

37. Roberts, C h; Jiang, W.; Jayaraman, J.; Trowsdale, J.; Holland, MJ.; Traherne, JA. Killer-cell Immunoglobulin-like Receptor gene linkage and copy number variation analysis by droplet digital PCR. Genome Med. 2014; 6:20. [PubMed: 24597950]

38. Regan JF, Kamitaki N, Legler T, Cooper S, Klitgord N, Karlin-Neumann G, et al. A Rapid Molecular Approach for Chromosomal Phasing. PLoS ONE. 2015; 10:e0118270. [PubMed: 25739099]

39. Gingeras TR. Implications of chimaeric non-co-linear transcripts. Nature. 2009; 461:206–11. [PubMed: 19741701]

40. Favilla V, Cimino S, Madonia M, Morgia G. New advances in clinical biomarkers in testis cancer. Front Biosci. 2010; 2:456–77. [PubMed: 20036893]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

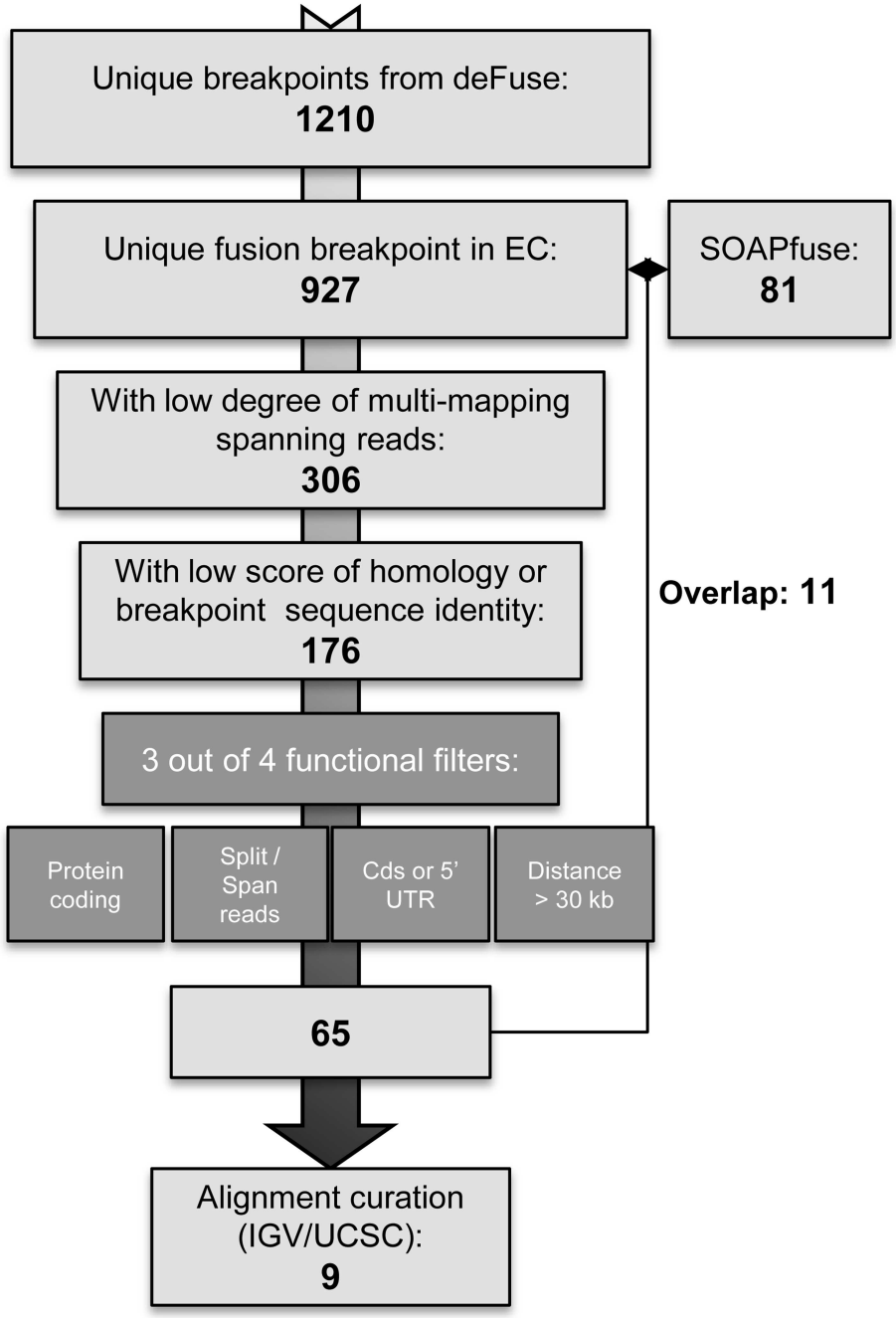

Figure 1. Filtering pipeline of nominated fusion transcripts

The identified fusion transcripts were filtered in a successive manner, resulting in nine final fusion transcripts that were experimentally validated.

Europe PMC Funders Author Manuscripts

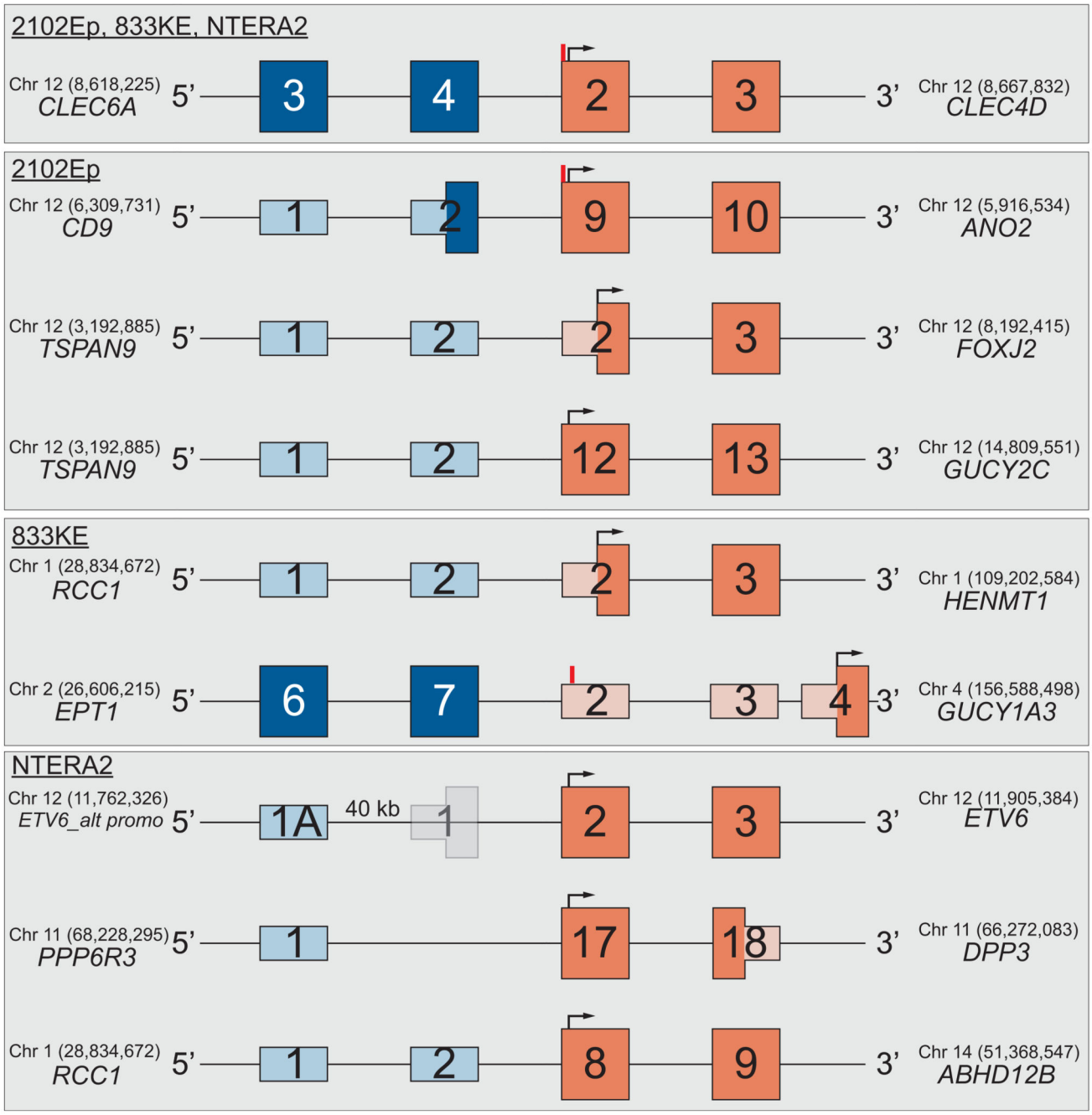

Figure 2. TGCT fusion transcripts identified by RNA-sequencing

All the nine novel transcripts have fusion breakpoints at intact exon-exon boundaries, except for the *ETV6* gene, where a new alternative promoter (exon 1A) was connected to exon 2. The breakpoint boundaries are indicated between upstream partner gene (blue) and downstream partner gene (orange). Full height of boxes of solid color represent predicted coding regions of original partner genes. Arrows mark the start codons of fusion transcript ORFs identified by the ORF finder at the National Centre for Biotechnology Information.

Red lines mark the stop codons of upstream partner gene ORFs. The genomic coordinates indicate the exact coordinate of the fusion breakpoint in the specific partner gene.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

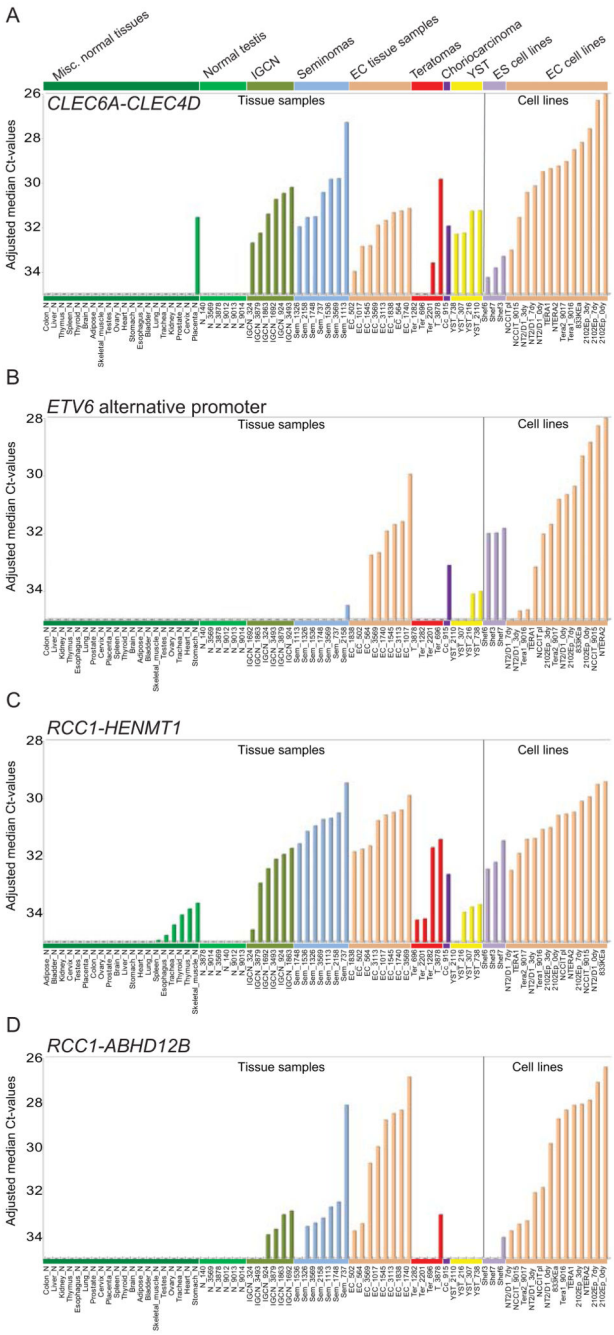

Figure 3. Fusion transcripts and alternative promoter usage of *ETV6* recurrently expressed in TGCT

Quantitative RT-PCR of recurrent fusion transcripts reported in C_T values normalized to median C_T values of endogenous controls, with higher expression corresponding to lower C_T values. Transcripts were considered absent for $C_T > 35$. Samples are grouped together according to histological subtype and ordered with increasing expression **A.** The read-through *CLEC6A-CLEC4D* was expressed in all subtypes of TGCTs and also in the pre-malignant IGCN samples, but not in tissue from normal testis. From normal tissue samples

from 20 different human organs only placenta expressed *CLEC6A-CLEC4D*. **B.** The *ETV6* transcript with an alternative exon 1 was expressed mainly in undifferentiated EC tumor samples and cell lines. It was also found to be expressed in the ES cell lines, in 1/6 seminomas, 1 choriocarcinoma and 2/3 YSTs. None of the samples from normal testis or normal human organs expressed this novel transcript. **C.** The intrachromosomal fusion transcript *RCC1-HENMT1* was expressed in all IGCN and TGCT samples, except for 1 YST. Also, the fusion transcript was detected in 3/3 ES cell lines and in 5/20 samples from normal human tissue. **D.** The interchromosomal fusion transcript *RCC1-ABHD12B* was expressed in 4/6 IGCN samples, undifferentiated subtypes of TGCTs including 5/6 seminomas, and all EC cell lines and samples. Also, 1/3 ES cell lines and 1/4 teratomas expressed the fusion transcript. None of the normal testis samples or normal human tissues expressed the fusion transcript.

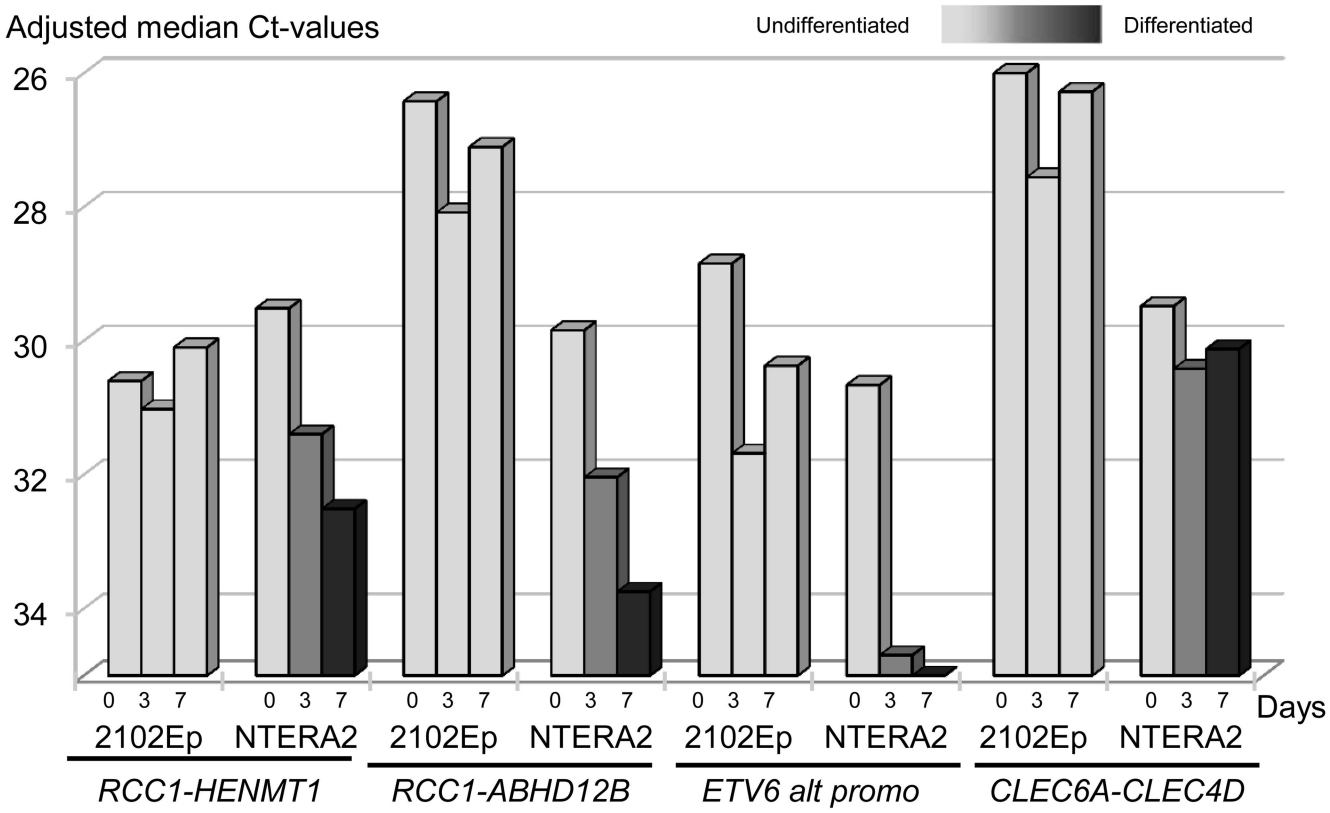

Figure 4. Fusions involving *RCC1* and the *ETV6* alternative promoter transcripts are down-regulated upon treatment with RA

Quantitative RT-PCR results for the fusion transcripts involving *RCC1*, the *ETV6* alternative promoter and *CLEC6A-CLEC4D*, in 2102Ep and NTERA2 cells treated with RA for 0, 3 and 7 days. Expression is reported as C_T values normalized to median C_T values of endogenous controls. The lighter to darker color gradient represents an *in vitro* undifferentiated to differentiated state. No clear patterns of expression are seen in the 2102Ep cell line, except for a general lower level of expression at 3 days of RA treatment. NTERA2 has reduced expression of *RCC1-HENMT1, RCC1-ABHD12B* and *ETV6* alternative promoter after 3 and 7 days.

Europe PMC Funders Author Manuscripts

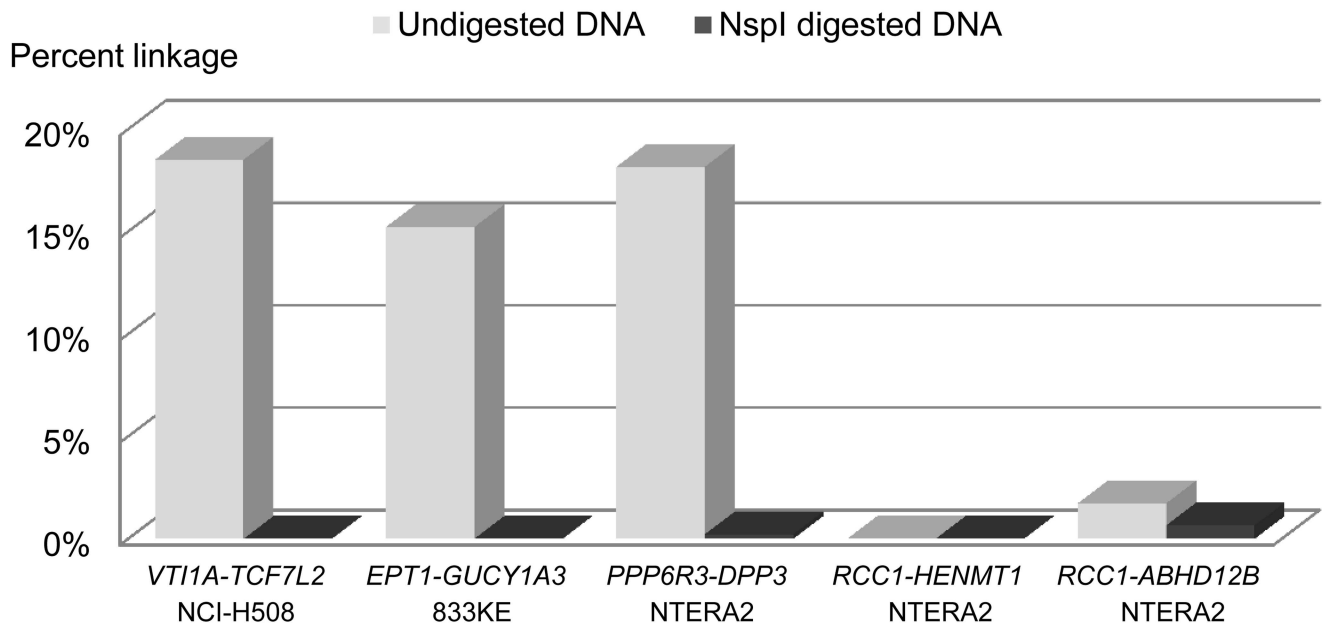

Figure 5. *EPT1-GUCY1A3* and *PPP6R3-DPP3* are chromosomally rearranged

DNA-level linkage of fusion partner genes, reported in percent linkage from ddPCR analysis, confirmed a genomic rearrangement underlying the known *VTI1A-TCF7L2* fusion. The fusions *EPT1-GUCY1A3* and *PPP6R3-DPP3* were shown to be linked on the DNA-level in 833KE and NTERA2, respectively. No DNA-level linkage was detected for the partner genes involved in the *RCC1* fusion transcripts. Linkage was undetected after fragmentation of DNA with the NspI endonuclease.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1
Nominated breakpoints from deFuse analysis of RNA-sequencing data

Nine breakpoints remained after heuristic filtering steps of initial candidates. Of these, *CLEC6A-CLEC4D* was nominated in all three EC cell lines. Breakpoints are listed according to the cell lines in which they were identified and with ascending genomic distance between the two partner genes. Presence of ORFs was determined using the ORF finder at the National Centre for Biotechnology Information (NCBI).

Cell line	Gene A	Gene B	Chromosome bands		Distance (kb)	defuse score	ORF
	CLEC6A	*CLEC4D*	12p13.31	12p13.31	31	0.99	Y
	CD9	*ANO2*	12p13.31	12p13.31	253	0.97	Y
	TSPAN9	*FOXJ2*	12p13.33-p13.32	12p13.31	4,790	0.97	Y
2102Ep	*TSPAN9*	*GUCY2C*	12p13.33-p13.32	12p13.1-p12.3	11,370	0.94	Y
	CLEC6A	*CLEC4D*	12p13.31	12p13.31	31	0.99	Y
	RCC1	*HENMT1*	1p35.3	1p13.3	80,325	0.92	Y
833KE	*EPT1*	*GUCY1A3*	2p23.3	4q32.1		0.97	Y
	CLEC6A	*CLEC4D*	12p13.31	12p13.31	31	0.83	Y
	ETV6	*RP11-434C1.1*[*]	12p13.2	12p13.2	59	0.81	Y
	PPP6R3	*DPP3*	11q13.2-13.3	11q13.2	1,951	0.82	Y
NTERA2	*RCC1*	*ABHD12B*	1p35.3	14q22.1		0.98	Y

[*] *RP11-434C1.1* was nominated as a partner to *ETV6*, located 85kb downstream. However, visual inspection revealed that the breakpoint localized to non-coding regions between these two genes and reflects an alternative promoter of *ETV6*.