

SCIENTIFIC REPORTS



OPEN

Novel kinase fusion transcripts found in endometrial cancer

Ryo Tamura¹, Kosuke Yoshihara¹, Kaoru Yamawaki¹, Kazuaki Suda¹, Tatsuya Ishiguro¹, Sosuke Adachi¹, Shujiro Okuda², Ituro Inoue³, Roel G. W. Verhaak^{4,5} & Takayuki Enomoto¹

Received: 29 July 2015

Accepted: 20 November 2015

Published: 22 December 2015

Recent advances in RNA-sequencing technology have enabled the discovery of gene fusion transcripts in the transcriptome of cancer cells. However, it remains difficult to differentiate the therapeutically targetable fusions from passenger events. We have analyzed RNA-sequencing data and DNA copy number data from 25 endometrial cancer cell lines to identify potential therapeutically targetable fusion transcripts, and have identified 124 high-confidence fusion transcripts, of which 69% are associated with gene amplifications. As targetable fusion candidates, we focused on three in-frame kinase fusion transcripts that retain a kinase domain (*CPQ-PRKDC*, *CAPZA2-MET*, and *VGLL4-PRKG1*). We detected only *CPQ-PRKDC* fusion transcript in three of 122 primary endometrial cancer tissues. Cell proliferation of the fusion-positive cell line was inhibited by knocking down the expression of wild-type *PRKDC* but not by blocking the *CPQ-PRKDC* fusion transcript expression. Quantitative real-time RT-PCR demonstrated that the expression of the *CPQ-PRKDC* fusion transcript was significantly lower than that of wild-type *PRKDC*, corresponding to a low transcript allele fraction of this fusion, based on RNA-sequencing read counts. In endometrial cancers, the *CPQ-PRKDC* fusion transcript may be a passenger aberration related to gene amplification. Our findings suggest that transcript allele fraction is a useful predictor to find bona-fide therapeutic-targetable fusion transcripts.

Fusion genes resulting from chromosomal rearrangements have an important role in the genesis of some types of tumors and have been recognized as important diagnostic and prognostic parameters in various types of malignant tumors^{1,2}. Fusion genes that are composed of protein kinase genes, such as anaplastic lymphoma receptor tyrosine kinase (*ALK*), ROS proto-oncogene 1, receptor tyrosine kinase (*ROS1*), ret proto-oncogene (*RET*), and fibroblast growth factor receptor (*FGFR*), have attracted significant attention as therapeutic drug targets^{3–6}. Indeed, the echinoderm microtubule-associated protein-like 4 (*EML4*)-*ALK* gene fusion has been detected in a small subset of non-small cell lung cancers⁷, and the efficacy of using *ALK* inhibitors to treat lung cancer patients who harbor the *EML4-ALK* fusion has been evident in clinical trials^{8,9}.

Recent advancements in sequencing technology have enabled comprehensive identification of fusion genes in human cancers, and candidate therapeutic targets have now been reported across many types of tumors^{10,11}. However, the functional evaluation of these fusion genes detected by whole-genome or transcriptome sequencing has been largely insufficient, if attempted at all, and it thus remains difficult to differentiate bona fide therapeutically targetable fusions from among the many candidates.

Endometrial cancer is currently the fourth most common cancer in women, and its incidence is rapidly increasing in Japan (<http://ganjoho.jp/en/index.html>). Although many cases of endometrial cancer are detectable at an early stage and show favorable outcomes, some are resistant to chemotherapy or radiotherapy, leading to a poor prognosis. To investigate the basis of this heterogeneity in endometrial cancer outcomes, several groups have performed genomic or transcriptomic analyses^{12–16}. For example, The Cancer Genome Atlas (TCGA) Research Network divides endometrial cancers into four categories, based on the integrated analysis of genomic, transcriptomic and proteomic data: polymerase ϵ (*POLE*) ultramutated, microsatellite-instability hypermutated, copy-number low, and copy-number high¹². Despite this extensive molecular exploration of endometrial cancer, there is at present no effective therapeutic molecular target for this disease.

¹Department of Obstetrics and Gynecology, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan. ²Department of Bioinformatics, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan. ³Division of Human Genetics, National Institute of Genetics, Mishima, Japan. ⁴Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁵Department of Genome Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. Correspondence and requests for materials should be addressed to K.Y. (email: yoshikou@med.niigata-u.ac.jp)

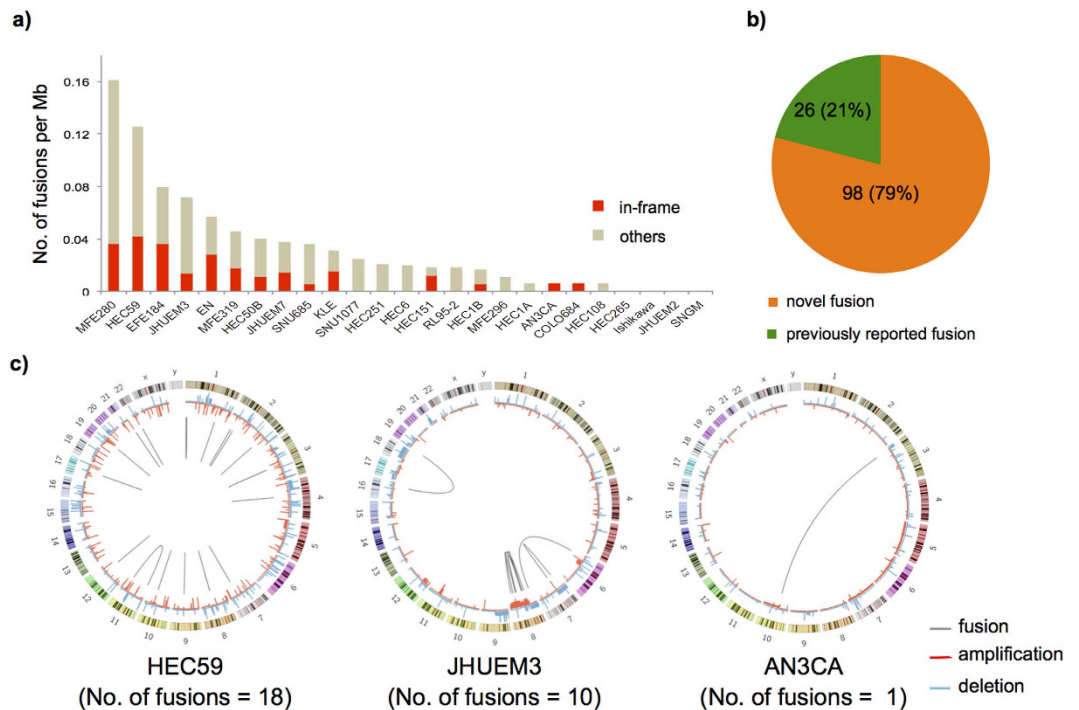


Figure 1. An overview of fusion transcripts found in endometrial cancer cell lines. (a) Histogram of the number of fusion transcripts per Mb in 25 endometrial cancer cell lines. (b) The percentage of novel fusion transcripts in endometrial cancer cell lines. Previously reported fusions were defined on the basis of TCGA Fusion Gene Data Portal (<http://www.tumorfusions.org>). (c) Circos plots indicate the landscape of fusion transcripts in JHUEM3, HEC59, and AN3CA cell lines. The outermost circle represents chromosomes and cytogenetic bands. The next circle represents copy number alterations identified using GISTIC. Green indicates deletion, and red indicates amplification. Each connection with gray lines in an innermost circle indicates fusion transcripts.

We have combined RNA-sequencing data from 25 endometrial cancer cell lines with array data of matched small nucleotide polymorphisms (SNPs) obtained from the Cancer Cell Line Encyclopedia (CCLE, <http://www.broadinstitute.org/ccle>) to define the landscape of gene-fusion transcripts in endometrial cancer cells, with the intention of discovering new potential therapeutic molecular targets for endometrial cancer.

Results

Detection of fusion transcripts. To identify high-confidence fusion transcripts, we applied the Pipeline for RNA sequencing Data Analysis (PRADA, <http://bioinformatics.mdanderson.org/Software/PRADA/>)¹⁷ software program to the CCLE RNA-sequencing data for 25 endometrial cancer cell lines. The detail of RNA sequencing data including such as data quality and origin of cell line was shown in Supplementary Table S1. We detected 287 fusion transcripts supported by at least two discordant read pairs plus one perfectly matched junction-spanning read, with the other end of the read pair mapping to either of the fusion gene partners. To reduce the number of false positive predictions, we filtered the fusion transcripts according to partner gene homology. We used BLASTn (Basic Local Alignment Search Tool-Nucleotides) to determine the homology between partnered genes, and removed 151 hits consisting of gene pairs with high similarity. After removing the overlapping fusion pairs, with 158 kinds of fusion transcripts discovered in 364 normal tissues samples¹¹, we identified 124 cancer-specific fusion transcripts from the 25 endometrial cancer cell lines (Supplementary Table S2).

Landscape of fusion transcripts in 25 endometrial cancer cell lines. The frequency of fusion transcripts per each cell line is shown in Fig. 1a. Four cell lines did not harbor any detectable, expressed fusion transcripts, and 14 cell lines had at least one in-frame fusion transcript that could result in a functional protein. As of May 1, 2015, only 26 of 124 cancer-specific fusion transcripts were previously reported in the TCGA Fusion Gene Data Portal (<http://www.tumorfusions.org>) (Fig. 1b and Supplementary Table S2).

To visualize the fusion-gene profile for each cell line, we used circos plots¹⁸ to represent the fusion events. Based on Genomic Identification of Significant Targets in Cancer 2.0 (GISTIC2)¹⁹ analysis, we also evaluated alterations in the fusion gene copy numbers (Fig. 1c and Supplementary Fig. S1). Overall, intra-chromosomal gene fusions were predominant in the endometrial cancer cell lines. Chromosome 8q harbored 70% of the fusion transcripts in the JHUEM3 endometrial cancer cell line corresponding to broad 8q regional amplification, suggesting the occurrence of chromothripsis²⁰.

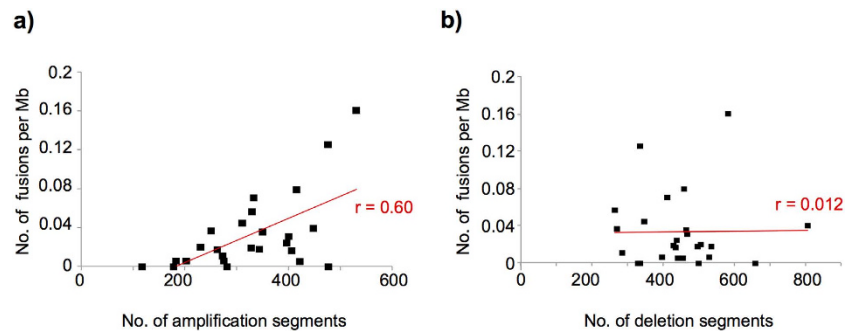


Figure 2. Association of copy number alterations with fusions. Correlation between the number of amplified (a) or deleted (b) segments and fusion transcripts per Mb in 25 endometrial cancer cell lines.

Next, we investigated the association between copy number alterations and fusion transcripts. We found a significant positive correlation between the number of fusion transcripts and the number of amplification segments (Fig. 2a), but not deletion segments (Fig. 2b). At least one of the fused genes was amplified in the cell line in 76 (69%) of the 109 fusion pairs; whereas deletion of one of the fused genes was detected in only 23 (20%) of the 109 fusion pairs. Although the frequency of in-frame or inter-chromosomal fusions was lower than that of out-of-frame or intra-chromosomal fusions, there was no significant association between these fusion types and copy number alterations (Supplementary Fig. S2).

To focus on therapeutically targetable fusion transcripts in endometrial cancer, from the initial list of 124 fusion transcripts, we identified three in-frame kinase fusion transcripts (*CPQ-PRKDC*, *VGLL4-PRKG1*, and *CAPZA2-MET*) that were supported by more than 10 junction-spanning reads, which included an intact kinase domain (Fig. 3a)²¹. The *CPQ-PRKDC* fusion transcript has not been reported previously.

Validation of fusion transcripts. To evaluate the authenticity and validity of our list of gene fusion transcripts, we used three human endometrial cancer cell lines (AN3CA, HEC59 and JHUEM3) that had in-frame kinase fusions. We performed Sanger sequencing of RT-PCR amplification products that spanned fusion junction points to validate all transcripts detected in the three cell lines. Although the RT step has possibility to produce artifactual fusion transcripts, we detected 27 (93.1%) of the 29 fusion transcripts by using Sanger sequencing RT-PCR amplification products (Fig. 3b–d). In particular, all fusion transcripts showing more than 10 junction-spanning reads²¹ were identifiable.

To assess the frequency of occurrence of the three in-frame kinase fusions (*CPQ-PRKDC*, *VGLL4-PRKG1*, *CAPZA2-MET*) in clinical samples, we conducted RT-PCR and Sanger sequencing for 122 primary endometrial cancer tissues. We detected *CPQ-PRKDC* fusion transcripts in three (2.5%) of the 122 primary tumor tissues; whereas we failed to detect *VGLL4-PRKG1* or *CAPZA2-MET* fusion transcripts in any of the clinical samples (Table 1). To rule out congenitally inherited gene fusions, we next conducted RT-PCR using paired normal tissues obtained from patients whose endometrial cancer tissues harbored the *CPQ-PRKDC* fusion transcript, and confirmed that this fusion event was somatic and tumor-specific (Supplementary Fig. S3).

Characterization of the novel *CPQ-PRKDC* fusion transcript. We used microarray data for mRNA expression, downloaded from the CCLE data repository, to examine the expression of *PRKDC* mRNA in endometrial cancer cell lines. The expression of the total *PRKDC* transcript in the *CPQ-PRKDC* fusion-positive cell line was relatively higher than that in the other endometrial cancer cell lines that did not have the *CPQ-PRKDC* fusion (Supplementary Fig. S4a). In addition, we found amplification of *PRKDC* gene in the *CPQ-PRKDC* fusion-positive cell line (Supplementary Fig. S4b). Next we examined the frequency of *PRKDC* gene alterations in TCGA endometrial cancer samples using cBioPortal²² (<http://www.cbioportal.org>). Of 240 TCGA endometrial cancer samples, three samples (1.3%) showed *PRKDC* amplifications with concurrent overexpression. Although 24 samples (10.0%) harbored *PRKDC* somatic mutations, there was no hotspots in *PRKDC* mutations (Supplementary Fig. S4c).

We investigated an association between *PRKDC* and *CPQ-PRKDC* expression and cell proliferation. To knock down the expression of *PRKDC* and *CPQ-PRKDC*, we prepared three kinds of small interfering RNA (siRNA), as follows: *PRKDC*-specific siRNA, dual siRNA and fusion-specific siRNA (Fig. 4a). Quantitative RT-PCR revealed that *PRKDC*-specific siRNA suppressed *PRKDC* expression but not *CPQ-PRKDC* fusion transcript expression, and that dual siRNA knockdown blocked the expression of both the *PRKDC* and *CPQ-PRKDC* transcripts (Fig. 4b,c). Suppression of *PRKDC* expression by *PRKDC*-specific siRNA or by dual siRNA inhibited cell proliferation (Fig. 4d). In addition, after transfection with dual siRNA, caspase 3/7 activity was noticeably higher, compared to the negative siRNA control (Supplementary Fig. S5a), suggesting that a DNA repair defect caused by loss of *PRKDC* expression might be leading to increased apoptosis.

When we transfected fusion-specific siRNA into the cell line expressing the *CPQ-PRKDC* fusion transcript, *CPQ-PRKDC* expression was suppressed, but not the wild-type *PRKDC* expression (Fig. 4e,f). However, knockdown of *CPQ-PRKDC* fusion transcript expression neither inhibited cell proliferation nor elevated caspase 3/7 activity in the fusion-positive cell line (Fig. 4g, Supplementary Fig. S5b).

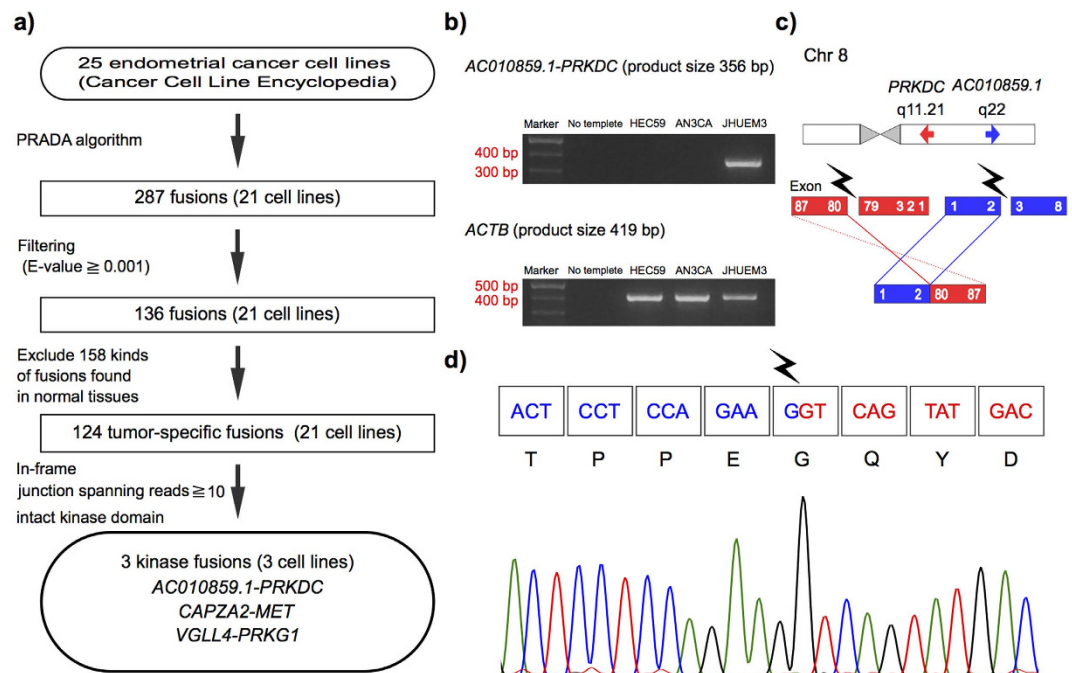


Figure 3. Validation of targetable fusions. (a) Flowchart for discovering targetable fusion transcripts. (b) Identification of *CPQ-PRKDC* fusion transcripts in JHUEM3 using RT-PCR. Marker represents a 100 bp DNA ladder. *ACTB* was used as the internal control for RT-PCR. (c) Schematics of the *CPQ-PRKDC* fusion transcripts. Regions corresponding to *CPQ* or *PRKDC* are shown in red or blue, respectively. (d) Sanger sequencing chromatogram showing the reading frame at the breakpoint and putative translation of the fusion protein in the fusion-positive cell line.

Cell line	5'-Gene	3'-Gene	Locus 5'-Gene	Locus 3'-Gene	5'-Gene copy number*	3'-Gene copy number*	Discordant read (n)	Junction spanning read (n)	Frequency in clinical specimens
JHUEM3	<i>CPQ</i>	<i>PRKDC</i>	8q22.1	8q11.21	2	2	16	44	3/122 (2.5%)
HEC59	<i>CAPZA2</i>	<i>MET</i>	7q31.2	7q31.2	0	1	8	27	0/122 (0%)
AN3CA	<i>VGLL4</i>	<i>PRKG1</i>	3p25.3	10q21.1	-2	1	19	21	0/122 (0%)

Table 1. Three kinase fusion transcripts in endometrial cancer. *Putative copy number was determined by GISTIC2 (-2 = homozygous deletion; -1 = hemizygous deletion; 1 = low level gain; 2 = high level amplification)

In the cell line harboring the *CPQ-PRKDC* fusion transcript, the expression of the fusion transcript was definitely lower than that of the wild-type *PRKDC* transcript. Similarly, the two endometrial cancer tissues expressing the *CPQ-PRKDC* fusion transcript also showed relatively lower expression of the *CPQ-PRKDC* fusion transcript (Supplementary Table S3). Corresponding to the mRNA expression level of fusion transcripts, Western blot did not detect protein expression of the three fusion transcripts (Supplementary Fig. S6).

Differences in transcript allele fraction between endometrial cancer cell lines and clinical samples. To compare the transcript allele fraction (TAF) of the kinase in-frame fusion transcripts between endometrial cancer cell lines and TCGA clinical samples, we used 7,887 fusion transcripts detected in 4,366 TCGA samples across 13 tumor types¹¹. We extracted 82 recurrent in-frame kinase fusions that were supported by more than ten junction-spanning reads and retained an intact kinase domain. Compared to the TAF scores of the recurrent in-frame kinase fusion transcripts, including therapeutically targetable fusion transcripts (Supplementary Table S4). Despite high tumor purity based on DNA copy number alterations²³, the scores of the three kinase in-frame fusions in endometrial cancer cell lines were lower (Fig. 5).

Discussion

Our study of 25 human endometrial cancer cell lines through paired-end RNA-sequencing data analysis has provided a list of high-confidence fusion transcripts and displayed the diversity of fusion transcripts and the association of fusion transcripts with gene amplification events.

TCGA Research Network performed an integrated analysis of genomic, transcriptomic and proteomic data obtained from 373 endometrial cancers to gain a better understanding of the molecular characteristics of this type

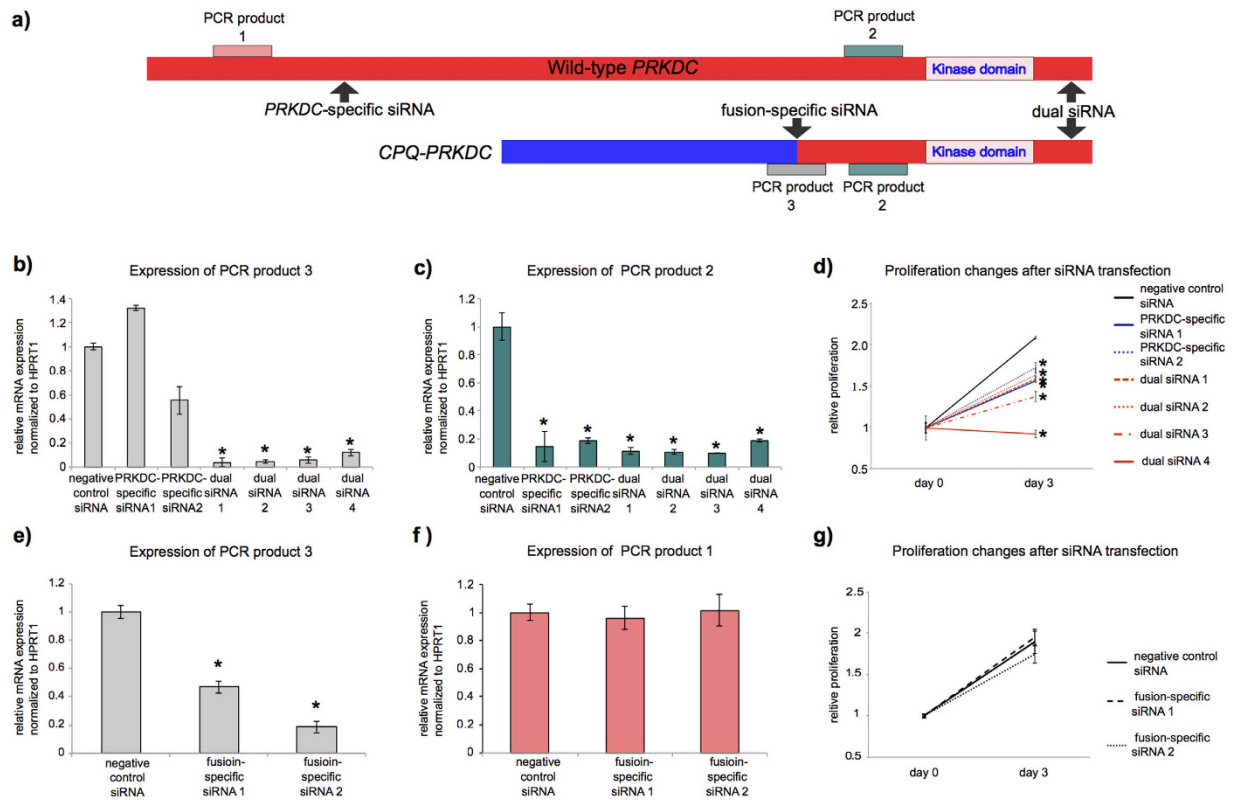


Figure 4. Knockdown of *PRKDC* and *CPQ-PRKDC* by siRNA. (a) Schematics of *CPQ-PRKDC* fusion transcripts and wild-type *PRKDC*. PRKDC-specific, fusion-specific and dual siRNAs were used in the knockdown experiments. Expression levels of the wild-type *PRKDC* (PCR product 1, beige), dual *PRKDC* (PCR product 2, green) and *CPQ-PRKDC* (PCR product 3, pink) were measured using appropriate primers. (b,c) *PRKDC* siRNAs were transfected to JHUEM3 harboring *CPQ-PRKDC* fusion transcript. The mRNA expressions of *CPQ-PRKDC* and dual *PRKDC* were assessed by quantitative real-time RT-PCR. Data represent the mean and error bar of triplicate independent experiments (asterisk denotes $p < 0.01$). (d) The effect of siRNA-mediated down-regulation of *PRKDC* on JHUEM3 cell proliferation. The cell proliferation was assessed by CellTiter-Glo assay. Data represent the mean and error bar of quadruplicate independent experiments (asterisk denotes $p < 0.01$). (e,f) Fusion-specific siRNA were transfected into JHUEM3 harboring an *CPQ-PRKDC* fusion. The mRNA expression levels of *CPQ-PRKDC* and wild-type *PRKDC* were assessed by quantitative real-time RT-PCR. Data represent the mean and error bar of triplicate independent experiments (asterisk denotes $p < 0.01$). (g) The effect of siRNA-mediated down-regulation of *CPQ-PRKDC* on JHUEM3 cell proliferation.

of tumor¹². Although 49 fusion transcripts, including *BCL*-related fusions, were identified in 106 endometrial cancers by using low-pass paired-end whole-genome sequencing data, there were no fusion transcripts that overlapped between the 106 TCGA clinical samples of endometrial cancer and the 25 endometrial cancer cell lines. There was no significant difference in RNA sequencing data quality between CCLE and TCGA (data not shown). This discrepancy might reflect differences in the types of sequencing data (low-pass paired-end whole genome data versus paired-end RNA-sequencing data) and the samples (clinically resected tissue versus cell line). Indeed, two of the three in-frame kinase fusions in the endometrial cancer cell lines were not detected in any of the 122 clinical samples. In this regard, Domcke *et al.*²⁴ reported striking differences in the molecular characteristics between commonly used ovarian cancer cell lines and primary high-grade serous ovarian cancer samples. Genomic alterations in culture-adapted cancer cell lines might not reflect the genomic alterations present in clinically resected samples. Another reason of the discrepancy between cell lines and clinical samples might be the small statistical power to discover new fusions in the small sample set. However, a list of fusion transcripts in cancer cell lines might still be a useful resource for the functional analysis of fusions detected across many tumor types^{21,25}.

Gene fusions involving kinase genes are receiving abundant attention as potential therapeutic molecular targets. Although the frequency of recurrent kinase fusion transcripts is generally lower than that of somatic mutations in malignant solid tumors, specific kinase inhibitors have been shown to have a curative effect on tumors that harbor related kinase gene fusions. For example, the development of imatinib mesylate, the first tyrosine kinase inhibitor targeting the BCR-ABL1 fusion protein, has dramatically improved prognosis in patients with chronic myeloid leukemia²⁶. Today, non-small cell lung cancer patients with an *EMLA-ALK* fusion are routinely treated with ALK inhibitors as first-line chemotherapy^{8,9}, even though this fusion was first reported only eight years ago.

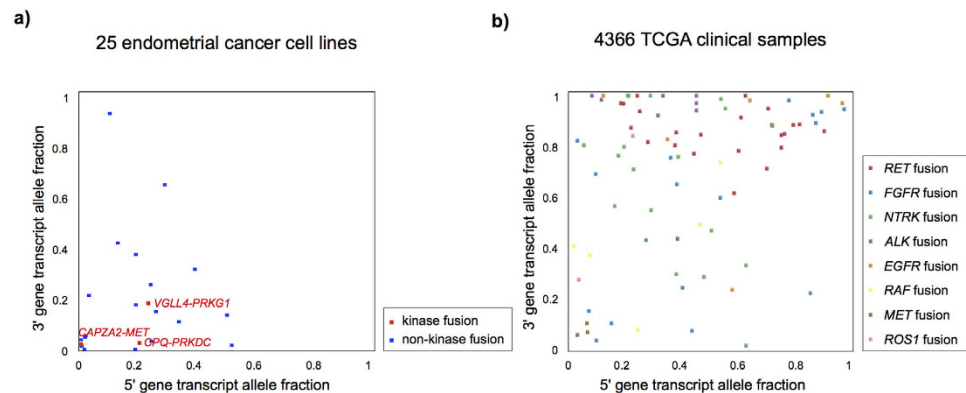


Figure 5. Comparison of transcript allele fraction (TAF) between endometrial cancer cell lines and TCGA clinical samples. (a) TAF value for each fusion transcript in endometrial cancer cell lines. Each plot represents the TAF of fusion transcripts that were supported by more than 10 junction-spanning reads. X- and y-axes represent TAF of 5' and 3' genes. Blue plots and gray squares represent in-frame and out-of-frame fusion transcripts, respectively. The three in-frame kinase fusion transcripts are highlighted in red. (b) TAF for 82 recurrent in-frame kinase fusion transcripts in TCGA clinical samples that were supported by more than 10 junction-spanning reads and that retained the kinase domain.

These success stories have accelerated our efforts to discover novel therapeutically targetable kinase fusions by using whole-genome or transcriptome data analysis.

We found that intra-chromosomal gene fusions predominated in the endometrial cancer cell lines, which is similar to what has been found in other cancers¹⁰. Chromosome 8q harbored 70% of the fusion transcripts in the JHUEM3 endometrial cancer cell line, corresponding to broad 8q amplification in that cell line. This suggests the occurrence of chromothripsis, a phenomenon by which up to thousands of clustered chromosomal rearrangements occur in a single event in localized genomic regions in one chromosome. Chromothripsis is known to be involved in both cancer and congenital diseases¹⁹.

Of particular interest to us was the novel kinase fusion *CPQ-PRKDC*, which we also detected in 2.5% of primary endometrial cancer samples. The DNA-dependent protein kinase catalytic subunit (DNA-PKcs) encoded by *PRKDC* is a critical component of the DNA repair machinery and thus plays an important role in maintaining genome integrity^{27,28}. DNA-PK is also associated with AKT phosphorylation, leading to the activation of the AKT signaling pathway^{29,30}. In addition, recent reports have demonstrated synthetic lethal interaction between *PRKDC* and some genes such as *ATM*³¹, *MYC*³², or *MSH3*³³, suggesting the significance of *PRKDC* as a druggable target. Although the *CPQ-PRKDC* fusion is an attractive candidate for a therapeutic target in endometrial cancer, our knock-down experiments and Western blot analysis demonstrated that inhibiting *CPQ-PRKDC*-positive cell line proliferation was caused by suppression of the wild-type *PRKDC* expression and not suppression of the fusion transcript, which suggests that this fusion transcript might be a passenger alteration.

The TAF provides a quantitative assessment of the relative expression levels of the wild-type and fusion transcripts. The TAF of *PRKDC* was lower than that of other clinically applicable kinase fusions, such as *ALK*^{8,9}, *RET*^{34,35}, *FGFR*^{36,37}, and *ROS1*³⁸ (Fig. 5b). When we calculate the TAF score using RNA-sequencing data from clinical tumor samples, we should consider the tumor purity of clinical samples because mRNA expression derived from non-tumor cells would lead to the underestimation of the TAF of a tumor-specific fusion transcript³⁹. A previous report has demonstrated the association between fusion transcripts and genomic alterations in breast cancer⁴⁰. Kalyana-Sundaram *et al.*⁴¹ have also reported that amplicon-associated *EGFR* or *RPS6KB1* fusion transcripts without a kinase domain represented low TAF values and were passenger aberrations in breast cancer cell lines. Corresponding to the previous work, our study demonstrated that HEC59 harboring *CAPZA2-MET* fusion transcript had high amplification of *MET*, suggesting this fusion was passenger aberration. The information of kinase domain in fusion transcript should also be considered to detect therapeutically targetable fusion genes and Stransky *et al.* have succeeded to discover recurrent *MET* fusions with a putative activating structure in the TCGA dataset¹⁰.

We have described the landscape of fusion transcripts in endometrial cancer by using RNA-sequencing data. Unfortunately, we could not identify therapeutically targetable kinase fusions in a small set of 25 endometrial cancer cell lines. To distinguish fusions that drive tumorigenesis from the more numerous incidental passenger fusions, researchers should weigh the TAF, gene copy number alterations, and retention of functional domains. Our findings suggest that the TAF is a good predictor for determining bona fide targetable fusion transcripts when applying fusion detection methods that use RNA-sequencing data.

Methods

Identification of fusion transcripts. We downloaded the RNA-sequencing data for 25 endometrial cancer cell lines from the CCLE⁴² at the Cancer Genomics Hub (<https://cghub.ucsc.edu>)⁴³. We used PRADA as previously reported¹¹ to detect fusion transcripts in the database. We extracted fusion transcripts characterized by (1) at least two discordant read pairs, (2) at least one junction-spanning read, and (3) no high gene homology

between each fusion gene partner (E-value > 0.001). Next, we extracted tumor-specific fusion transcripts by removing the fusion transcripts that overlapped with 158 fusion transcripts discovered in 364 normal tissue samples.

We also downloaded the copy number data of endometrial cancer cell lines from CCLE. We integrated the fusion transcript data and copy number data to investigate the association between fused genes and copy number alterations.

Samples. The three human cancer cell lines (HEC59, JHUM3, and AN3CA) were purchased from the Japanese Collection of Research Bioresources Cell Bank, RIKEN Cell Bank, and American Type Culture Collection, respectively. All cell lines were cultured in Dulbecco's modified Eagle medium, supplemented with 10% fetal bovine serum, 50 IU/mL penicillin, and 50 µg/mL streptomycin.

Tissues from 122 Japanese patients who were diagnosed with endometrial cancer between 2006 and 2014 at Niigata University Medical and Dental Hospital were included in this study. Fresh frozen samples were obtained from primary tumor tissues during surgery prior to chemotherapy. The histologic characteristics of surgically resected specimens were assessed on formalin-fixed and paraffin-embedded hematoxylin and eosin sections by a gynecologic pathologist. A total of 122 samples (109 endometrioid adenocarcinomas, 5 mixed adenocarcinomas, 4 carcinosarcomas, 2 undifferentiated adenocarcinomas, 1 serous adenocarcinoma and 1 unclassified adenocarcinoma of the uterine corpus) were analyzed. The institutional ethics review board at Niigata University approved this study. All patients provided written informed consent for the collection of samples and subsequent analysis. All experiments were performed in accordance with the approved guidelines.

RT-PCR and Sanger sequencing. Total RNA was extracted from fresh frozen samples using TRIzol (Invitrogen, Carlsbad, CA, USA) and from cell lines using the RNeasy mini kit (Qiagen, Tokyo, Japan). Total RNA (1 µg) was reverse-transcribed into cDNA using Prime Script II Reverse Transcriptase (Takara Bio, Shiga, Japan). cDNA (corresponding to 10 ng total RNA) was subjected to PCR amplification using KAPA Taq DNA polymerase (KAPA Biosystems, Woburn, MA, USA) or Prime Star HS premix (Takara Bio). The reactions were carried out in a thermal cycler under the following conditions: 40 cycles at 95 °C for 30 sec, at 60 °C for 30 sec, and at 72 °C for 1 min, with a final extension at 72 °C for 1 min (KAPA Taq DNA polymerase). Alternatively, PCR was done for 40 cycles at 98 °C for 15 sec, at 68 °C for 1 min, and a final extension at 72 °C for 5 min (Prime Star HS premix).

Beta-actin (*ACTB*) was amplified to evaluate the efficiency of cDNA synthesis. PCR products were extracted and purified by NucleoSpin Gel and PCR Clean-up (Takara Bio) and were sequenced on an ABI 3130xl DNA Sequencer (Applied Biosystems, Foster City, CA, USA) using a BigDye Terminator kit (Applied Biosystems). The PCR primers used in this study (shown in Supplementary Table S5) were designed using Primer3Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>).

siRNA knock-down experiments. The negative control siRNA (with scrambled sequence, SI1022076), *PRKDC*-specific siRNA (SI2663633, SI2224229), and dual siRNA (SI2224236, SI4436698, SI4436705, SI4436684) were purchased from Qiagen (Tokyo, Japan). Fusion-specific siRNA was designed using siDirect (<http://sidirect2.rnai.jp>). The siRNA sequences are shown in Supplementary Table S6. Quantitative real-time RT-PCR was performed using total RNA that was extracted after 72 hours of siRNA transfection. Cell proliferation and apoptosis were measured after 24 and 96 hours of siRNA transfection using the CellTiter Glo assay and the Caspase-Glo 3/7 assay, respectively (Promega, Mannheim, Germany) according to the manufacturer's protocols.

Quantitative real-time RT-PCR. Quantitative real-time RT-PCR was performed with a Thermal Cycler Dice Real-Time System TP800 2.01C (Takara Bio). cDNA (corresponding to 5 ng of total RNA) was subjected to real-time PCR amplification analysis using SYBR Green PCR mix (Applied Biosystems). Gene expression was tested in a set of 15 human housekeeping genes (Takara Bio) and *HPRT1* was selected as having the lowest variation in gene expression. The relative quantification method⁴⁴ was used to measure the amounts of the respective genes normalized to *HPRT1*. All primers used in quantitative real-time RT-PCR are shown in Supplementary Table S7.

Western blotting. Cells were lysed in lysis buffer (50 mM Tris at pH 8.0, 150 mM NaCl, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% SDS and 1 mM EDTA) supplemented with protease inhibitors. The lysates were electrophoresed using a 5–20% real gel plate (Bio Craft, Tokyo, Japan) and transferred to Immobilon-P membranes (Millipore, Darmstadt, Germany). Western blot analyses were performed with anti-*PRKDC* (#4602, Cell Signaling Technology, Tokyo, Japan), anti-*MET* (#3019, Cell Signaling Technology), anti-*PRKGI* (#5592, Cell Signaling Technology), and anti-*GAPDH* (MAB374, Millipore) antibodies.

Statistical analysis. Data were expressed as the mean ± SD. Fisher's exact test and Student t-test were used for evaluation of significance between data groups. P values less than 0.05 indicated statistical significance.

References

1. Mitelman, F., Johansson, B. & Mertens, F. The impact of translocations and gene fusions on cancer causation. *Nature reviews. Cancer* **7**, 233–245, doi: 10.1038/nrc2091 (2007).
2. Mertens, F., Johansson, B., Fioretos, T. & Mitelman, F. The emerging complexity of gene fusions in cancer. *Nature reviews. Cancer* **15**, 371–381, doi: 10.1038/nrc3947 (2015).
3. Shaw, A. T., Hsu, P. P., Awad, M. M. & Engelman, J. A. Tyrosine kinase gene rearrangements in epithelial malignancies. *Nature reviews. Cancer* **13**, 772–787, doi: 10.1038/nrc3612 (2013).
4. Kohno, T. *et al.* RET fusion gene: translation to personalized lung cancer therapy. *Cancer science* **104**, 1396–1400, doi: 10.1111/cas.12275 (2013).

5. Arai, Y. *et al.* Fibroblast growth factor receptor 2 tyrosine kinase fusions define a unique molecular subtype of cholangiocarcinoma. *Hepatology* **59**, 1427–1434, doi: 10.1002/hep.26890 (2014).
6. Singh, D. *et al.* Transforming fusions of FGFR and TACC genes in human glioblastoma. *Science* **337**, 1231–1235, doi: 10.1126/science.1220834 (2012).
7. Soda, M. *et al.* Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566, doi: 10.1038/nature05945 (2007).
8. Shaw, A. T. *et al.* Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *The New England journal of medicine* **368**, 2385–2394, doi: 10.1056/NEJMoa1214886 (2013).
9. Solomon, B. J. *et al.* First-line crizotinib versus chemotherapy in ALK-positive lung cancer. *The New England journal of medicine* **371**, 2167–2177, doi: 10.1056/NEJMoa1408440 (2014).
10. Stransky, N., Cerami, E., Schalm, S., Kim, J. L. & Lengauer, C. The landscape of kinase fusions in cancer. *Nat Commun* **5**, 4846, doi: 10.1038/ncomms5846 (2014).
11. Yoshihara, K. *et al.* The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene*, doi: 10.1038/onc.2014.406 (2014).
12. Cancer Genome Atlas Research, N. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73, doi: 10.1038/nature12113 (2013).
13. Liang, H. *et al.* Whole-exome sequencing combined with functional genomics reveals novel candidate driver cancer genes in endometrial cancer. *Genome research* **22**, 2120–2129, doi: 10.1101/gr.137596.112 (2012).
14. Le Gallo, M. *et al.* Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. *Nature genetics* **44**, 1310–1315, doi: 10.1038/ng.2455 (2012).
15. Zhao, S. *et al.* Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2916–2921, doi: 10.1073/pnas.1222577110 (2013).
16. Liu, Y. *et al.* Clinical significance of CTNNB1 mutation and Wnt pathway activation in endometrioid endometrial carcinoma. *Journal of the National Cancer Institute* **106**, doi: 10.1093/jnci/dju245 (2014).
17. Torres-Garcia, W. *et al.* PRADA: pipeline for RNA sequencing data analysis. *Bioinformatics* **30**, 2224–2226, doi: 10.1093/bioinformatics/btu169 (2014).
18. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome research* **19**, 1639–1645, doi: 10.1101/gr.092759.109 (2009).
19. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome biology* **12**, R41, doi: 10.1186/gb-2011-12-4-r41 (2011).
20. Korbel, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**, 1226–1236, doi: 10.1016/j.cell.2013.02.023 (2013).
21. Kim, H. P. *et al.* Novel fusion transcripts in human gastric cancer revealed by transcriptome analysis. *Oncogene* **33**, 5434–5441, doi: 10.1038/onc.2013.490 (2014).
22. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* **2**, 401–404, doi: 10.1158/2159-8290.CD-12-0095 (2012).
23. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nature biotechnology* **30**, 413–421, doi: 10.1038/nbt.2203 (2012).
24. Domcke, S., Sinha, R., Levine, D. A., Sander, C. & Schultz, N. Evaluating cell lines as tumour models by comparison of genomic profiles. *Nature communications* **4**, 2126, doi: 10.1038/ncomms3126 (2013).
25. Klijn, C. *et al.* A comprehensive transcriptional portrait of human cancer cell lines. *Nature biotechnology*, doi: 10.1038/nbt.3080 (2014).
26. Bjorkholm, M. *et al.* Success story of targeted therapy in chronic myeloid leukemia: a population-based study of patients diagnosed in Sweden from 1973 to 2008. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* **29**, 2514–2520, doi: 10.1200/JCO.2011.34.7146 (2011).
27. Goodwin, J. F. & Knudsen, K. E. Beyond DNA repair: DNA-PK function in cancer. *Cancer discovery* **4**, 1126–1139, doi: 10.1158/2159-8290.CD-14-0358 (2014).
28. Hsu, F. M., Zhang, S. & Chen, B. P. Role of DNA-dependent protein kinase catalytic subunit in cancer development and treatment. *Translational cancer research* **1**, 22–34, doi: 10.3978/j.issn.2218-676X.2012.04.01 (2012).
29. Stronach, E. A. *et al.* DNA-PK mediates AKT activation and apoptosis inhibition in clinically acquired platinum resistance. *Neoplasia* **13**, 1069–1080 (2011).
30. An, J. *et al.* DNA-dependent protein kinase catalytic subunit modulates the stability of c-Myc oncoprotein. *Molecular cancer* **7**, 32, doi: 10.1186/1476-4598-7-32 (2008).
31. Riabinska, A. *et al.* Therapeutic targeting of a robust non-oncogene addiction to PRKDC in ATM-defective tumors. *Science translational medicine* **5**, 189ra178, doi: 10.1126/scitranslmed.3005814 (2013).
32. Zhou, Z. *et al.* Identification of synthetic lethality of PRKDC in MYC-dependent human cancers by pooled shRNA screening. *BMC cancer* **14**, 944, doi: 10.1186/1471-2407-14-944 (2014).
33. Dietlein, F. *et al.* A functional cancer genomics screen identifies a druggable synthetic lethal interaction between MSH3 and PRKDC. *Cancer discovery* **4**, 592–605, doi: 10.1158/2159-8290.CD-13-0907 (2014).
34. Lipson, D. *et al.* Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. *Nature medicine* **18**, 382–384, doi: 10.1038/nm.2673 (2012).
35. Drilon, A. *et al.* Response to Cabozantinib in patients with RET fusion-positive lung adenocarcinomas. *Cancer discovery* **3**, 630–635, doi: 10.1158/2159-8290.CD-13-0035 (2013).
36. Gozgit, J. M. *et al.* Ponatinib (AP24534), a multitargeted pan-FGFR inhibitor with activity in multiple FGFR-amplified or mutated cancer models. *Molecular cancer therapeutics* **11**, 690–699, doi: 10.1158/1535-7163.MCT-11-0450 (2012).
37. Guagnano, V. *et al.* FGFR genetic alterations predict for sensitivity to NVP-BGJ398, a selective pan-FGFR inhibitor. *Cancer discovery* **2**, 1118–1133, doi: 10.1158/2159-8290.CD-12-0210 (2012).
38. Kwak, E. L. *et al.* Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *The New England journal of medicine* **363**, 1693–1703, doi: 10.1056/NEJMoa1006448 (2010).
39. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications* **4**, 2612, doi: 10.1038/ncomms3612 (2013).
40. Stephens, P. J. *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462**, 1005–1010, doi: 10.1038/nature08645 (2009).
41. Kalyana-Sundaram, S. *et al.* Gene fusions associated with recurrent amplicons represent a class of passenger aberrations in breast cancer. *Neoplasia* **14**, 702–708 (2012).
42. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607, doi: 10.1038/nature11003 (2012).
43. Wilks, C. *et al.* The Cancer Genomics Hub (CGHub): overcoming cancer through the power of torrential data. *Database: the journal of biological databases and curation* **2014**, doi: 10.1093/database/bau093 (2014).
44. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods* **25**, 402–408, doi: 10.1006/meth.2001.1262 (2001).

Acknowledgements

We thank the tissue donors and our supporting medical staff for making this study possible. We are grateful to A. Ishida for technical assistance. This work was supported in part by the Kanzawa Medical Research Foundation (to K.Y.).

Author Contributions

R.T., K.Y. and T.E. designed the research plan. K.Y. performed computational analyses. R.T. conducted all experiments. R.T. and K.Y. wrote the main manuscript and prepared all figures and tables. Y.K., K.S., T.I., S.A., S.O., I.I., R.G.W.V. and T.E. discussed the results. All authors reviewed and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Tamura, R. *et al.* Novel kinase fusion transcripts found in endometrial cancer. *Sci. Rep.* 5, 18657; doi: 10.1038/srep18657 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>