Table 1 Gene-specific primers for TAIL-PCR

Primer name	Sequence (5'-3')
ABCC4-TAIL0	CTGGTGGTGGGCGTTTCTGATATTCCC
ABCC4-TAIL1	ACGATGGACTCCAGTCCGGCCTTTGTCGAAC ACAC CACTGAAACAT
ABCC4-TAI L2	CCAGGCGCTTCACATCTCTTGACGTTTCC
ATP6V0A4-TAIL0	TTCCATGTGCCGCTGAACATGGGTTGG
ATP6V0A4-TAIL1	ACGATGGACTCCAGTCCGGCCAAAGATGTTC AAGGACTTGGAGAAGCAG
ATP6VOA4-TAI2	CTGGG1TfATCTCCCGGTAGCTGCCGAC
CDCA2-TAIL0	GCATTGCAGTTTTCCTTCTGCAGCTCC
CDCA2-TAIL1	ACGATGGACTCCAGTCCGGCCTGCTGCAGGG TCAGAGCAGGTTTG
CDCA2-TAIL2	CTTGATG CATATGCAAATCTGGGTCATGACG C
CEP250-TAIL0	GAGCTGGGTCTGTAGTATCCCAGTGG
CEP250-TAIL1	ACGATGGACTCCAGTCCGGCCTCAGTCGTTC CAGTTGTTGGCTG
CEP250-TAIL2	AGCAGTGTCTCCAGGAGGGATACTCTC
MACF1-TAIL0	CGATCATCTAGGAGCCGCTGGAGC
MACF1-TAIL1	ACGATGGACTCCAGTCCGGCCAACCAGCTG AGCAATGGCTCC
MACF1-TAIL2	CCCACAATGCAACAAAGCTTCCTGTAGCTG
RLF-TAIL0	CCATTCCTTCAGTCTCTACAGGAGTCAC
RLF-TAIL1	ACGATGGACTCCAGTCCGGCCAAGGAAGGG GTGTGGAAAAACCCAG
RLF-TAIL2	CTGTCTCAACAGCCAGTAGAAACGGAGG
SLCO4A1-TAIL0	CAGGAGCCCCATGATGAGTATGTAG
SLCO4A1-TAIL1	ACGATGGACTCCAGTCCGGCCACAGCAGAC AGGCCTTGTCGATC
SLCO4A1-TAIL2	GCATTTCCCTGCAGTGGCATGGCC

15 sec at 95°C, 30 sec at 65°C, 2 min at 72°C. The primer pairs used in this experiment were designed to make the amplification product including the breakpoints of the fusion genes. For One Step RT-PCR, TaKaRa One Step SYBR PrimeScript RT-PCR Kit II was used according to the manufacturer's instructions. 1 ng of total RNA from the dissected tumor cells was used as a template in each 20 μl reaction. Thermal cycling was carried out under the following conditions: 30 min at 50°C, 2 min at 94°C followed by 35 cycles of 30 sec at 94°C, 30 sec at 65°C, 1 min at 72°C. The primers for RT-PCR and One step RT-PCR are shown in Table 4.

Table 2 LAD primers and AC1 primer for TAIL-PCR

Primer name	Sequence (5'—3')
LDA1	ACGATGGACTCCAGAGCGGCCGC(G/C/A)N(G/C/A) NNNGGAA
LDA2	ACGATGGACTCCAGAGCGGCCGC(G/C/T/)N(G/C/T) NNNGGTT
LDA3	ACGATGGACTCCAGAGCGGCCGC(G/C/A)(G/C/A)N(G/C/A) NNNCCAA
LDA4	ACGATGGACTCCAGAGCGGCCGC(G/C/T)(G/A/T)N(G/C/T/) NNNCGGT
LDA5	ACGATGGACTCCAGAGAG(A/T)GNAG(A/T)ANCA(A/T)AGG
LDA6	ACGATGGACTCCAGAG(A/T)GTGNAG(A/T)ANCANAGA
AC1	ACGATGGACTCCAGAG

The amplified PCR products were electrophoresed on 1.0% or 2.0% agarose gels, and were purified using GL Sciences MonoFas DNA purification kit I (GL Sciences, Tokyo, Japan). The purified products were sequenced using Applied Biosystems BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies, Carlsbad, California), and the reaction products were purified using Promega Wizard MagneSil Sequencing Reaction Clean-Up System (Promega, Madison, WI). The purified samples were analyzed using Applied Biosystems 3130χ Genetic Analyzer.

Results

Development of fusion gene screening program

To profile the exon expression in fusion genes, SarcomaA which harbors the fusion gene *EWSR1-ATF1*, was used for Exon Array experiments (Figure 1). Exon expression profiles of *EWSR1* and *ATF1* were characterized (Figure 2), and the following features were observed. 1: Probe sets in the exon region had high signal intensity, and probe sets in the intron region had low signal intensity. 2: In some probe sets, all samples had equivalent signal intensity. In other probe sets, all samples had extremely low equivalence. 3: The expression signals vary in each probe set on a gene of one sample. 4: SarcomaA showed a change in the expression level at the breakpoint in comparison with breast cancer cell lines.

Then the fusion gene screening program was developed to detect fusion genes with an exon expression profile similar to that of *EWSR1* and *ATF1*.

The detection performance of the developed program was examined using the Exon Array data of the T-ALL cell lines. The program selected the candidate genes: one gene in ALL-SIL, one gene in BE13, and two genes in LOUCY. NUP214, the partner gene of the known fusion genes, was detected in ALL-SIL and LOUCY. Other known fusion partner genes, ABL1 in ALL-SIL, NUP214 and ABL1 in BE13, SET in LOUCY, were not detected in this case, because the probe sets that could be used in the 5′ or 3′ terminal groups were three or less. Although the NUP214 gene was detected as a candidate gene in ALL-SIL and LOUCY, its exon expression profile was different between the two cell lines. While the expression decreases from the 5′ terminal side to the 3′ terminal side at the breakpoint in ALL-SIL, it was opposite in LOUCY. Thus it was confirmed that gene detection by the program did not depend on the direction of the expression change. Although breakpoints were

Table 3 Thermal conditions for TAIL-PCR

Pre-amplification			Primary TAIL-PCR			Secondary TAIL-PCR		
Step	Temperature (°C)	Time (min:sec)	Step	Temperature (°C)	Time (min:sec)	Step	Temperature (°C)	Time (min:sec)
1	93	2:00	1	94	0:20	1	94	0:20
2	95	1:00	2	65	1:00	2	68	1:00
3	94	0:30	3	72	3:00	3	72	3:00
4	25	2:00	4	To step 1	1 time	4	94	0:20
5	Ramping to 72	0.5°C/s	5	94	0:20	5	68	1:00
6	72	3:00	6	68	1:00	6	72	3:00
7	94	0:30	7	72	3:00	7	94	0:20
8	60	1:00	8	94	0:20	8	50	1:00
9	72	3:00	9	68	1:00	9	72	3:00
10	Go to step7	10 times	10	72	3:00	10	To step 1	7 times
11	94	0:30	11	94	0:20	11	72	5:00
12	25	2:00	12	50	1:00			
13	Ramping to 72	0.5°C/s	13	72	3:00			
14	72	3:00	14	To step 5	13 times			
15	94	0:20	15	72	5:00			
16	58	1:00						
17	72	3:00						
18	Go to step 15	25 times						
19	72	5:00						

detected at a different position in ALL-SIL and LOUCY, they corresponded to the position of reported breakpoints. It was confirmed that the breakpoint was detected accurately by the program (Figure 3).

Candidate genes in breast and pancreatic cancer cell lines

To discover the novel fusion gene in breast and pancreatic cancer cell lines, candidate genes were selected by the proposed methodology. As a result, 20 genes were selected in 24 breast cancer cell lines. Four of the selected genes were excluded from the candidates, because it was thought that the exon expression profiles of these 4 genes were influenced by known transcript isoforms. One gene was excluded, because a similar exon expression profile to the cancer cell line detected by the program was also observed in HMEC. As a result of the evaluation of the 15 remaining genes, 4 most attractive genes were selected as candidate genes in the breast

Table 4 Primers for RT-PCR

Target fusion gene	Primer name	Orientation	Sequence (5'- 3')	Amplicon size
DOCK5-CDCA2	DOCK5-exonl	Forward	GAGGAGCTGTAGCAGCCTTAGTCG	371 bp
	CDCA2-TAIL2	Reverse	CTTGATGCATATGCAAATCTGGGTCATGACGC	
DOCK5-CDCA2	DOCK5-exon1	Forward	GAGGAGCTGTAGCAGCCTTAGTCG	760 bp
	CDCA2-TAIL0	Reverse	GCATTGCAGTTTTCCTTCTGCAGCTCC	
ZMYND8-CEP250	ZMYND8-exon18	Forward	TACATCAGGAGGCMAGCGACA	513 bp
	CEP250-TAIL2	Reverse	GCAGTGTCTCCAGGAGGGATACTCTC	
ZMYND8-CEP250	ZMYND8-exonl5	Forward	GCCGCTTTTACCGAAGGAGACT	1476 bp
	CE P250-exon27	Reverse	GCTGCTGCTCCGTGATATGAGA	
RLF-ZMPSTE24	RLF-TAIL2	Forward	CCCCCAGGCTACTGCTTTATCAAAACTA	445 bp
	ZMPSTE24-exon3	Reverse	CATAACCACAGAACCGTCCAGAAAG	
RLF-ZMPSTE24	RLF-exon1	Forward	GTTGCCTACGCGCTGGTG	2167 bp
	ZMPSTE24-exon10	Reverse	GATGTCCAGGATCTGTGACTGA	

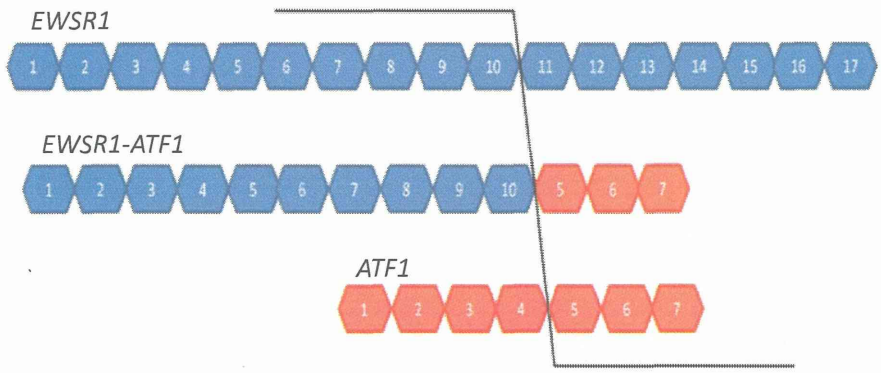

Figure 1 Schema of EWSR1-ATF1 mRNA. EWSR1 exon 1-10 fuse to ATF1 exon 5-7 by in-frame. Boxes with numbers represent the exon regions of the genes.

cancer cell lines. In the 20 pancreatic cancer cell lines, 23 genes were selected by the program. Nine genes of them thought to be influenced by known transcript isoforms, and 3 genes that correspond to two or more RefSeq genes, respectively, were excluded from the candidate genes. As a result of evaluating the 11 remaining genes, the 3 most attractive genes were selected as candidate genes in the pancreatic cancer cell lines. Details are shown in Table 5 and Figures 3, 4, 5, 6, 7, 8, 9 and 10.

Exon expression profiles of all selected gene by the program are shown in Additional file 1 and Additional file 2.

Identification of novel fusion gene

It was attempted to identify unknown counterpart genes using TAIL-PCR from higher expression ends of selected candidate genes. In this research we did not carry out it from lower ends. TAIL-PCR is one of the methods by which an unknown sequence adjacent to an already-known sequence can be efficiently amplified [19]. As a result of fusion gene identification experiments for the 7 candidate genes, gene fusion fragments were acquired for 3 candidate genes. Additionally, the frequency of fusion genes evaluated in cell lines and clinical tissue samples using RT-PCR and One Step RT-PCR.

DOCK5-CDCA2

The upstream sequence of exon 14 of *CDCA2* gene (ENST00000380665) was acquired in breast cancer cell line UACC893. This sequence was part of the exon 1 of *DOCK5* gene (ENST00000276440) (Figure 11A). In addition, the fusion of *DOCK* 5 exon 1 and *CDCA2* exon 14 was confirmed by RT-PCR (Figure 11B). But *DOCK5-CDCA2* fusion mRNA was not detected by RT-PCR in 111 breast cancer clinical tissues.

ZMYND8-CEP250

The upstream sequence of exon 22 of *CEP250* gene (ENST00000356095) was searched for in breast cancer cell line BT474, and was found to be a sequence from exon 16 to exon 19 of *ZMYND8* gene (ENST00000360911) (Figure 12A). The fusion of *ZMYND8* exon 19 and *CEP250* exon 22 was confirmed by RT-PCR (Figure 12B). But *ZMYND8-CEP250* fusion mRNA was not detected by RT-PCR in 111 breast cancer clinical tissues.

RLF-ZMPSTE24

The upstream sequence of exon 5 of *RLF* gene (ENST00000372771) was acquired in pancreatic cancer cell line PA043, and was found to be a sequence from exon 2 to part of exon 5 of *ZMPSTE24* gene (ENST00000372759) (Figure 13A). In addition, the fusion of *RLF* exon 5 and *ZMPSTE24* exon 2 was confirmed by RT-PCR (Figure 13B). *RLF-ZMPSTE24* fusion mRNA was detected by RT-PCR in pancreatic cancer clinical tissue, PA043T (Figure 13C). This tissue was the origin of the cell line PA043 where *RLF-ZMPSTE24* was first identified. The frequency of *RLF-ZMPSTE24* expression in pancreatic cancer patients was 1/58 (1.7%).

Discussion

Here, a method is proposed to detect novel fusion genes using exon array data of tumor samples in combination with a new computational program.

Development of new fusion gene detection program

This computational program is based on the following ideas.

Selection of probe set:

Although a large number of probe sets are designed on Exon Array, it is known that there are some non-functional probes. Technical anomalies may give a false signal for un-functional probe sets due to cross-hybridization, saturation or an inherently weak and non-linear response. Actually, some probe sets for *EWSR1* and *ATF1* were thought to be un-functional probes. To

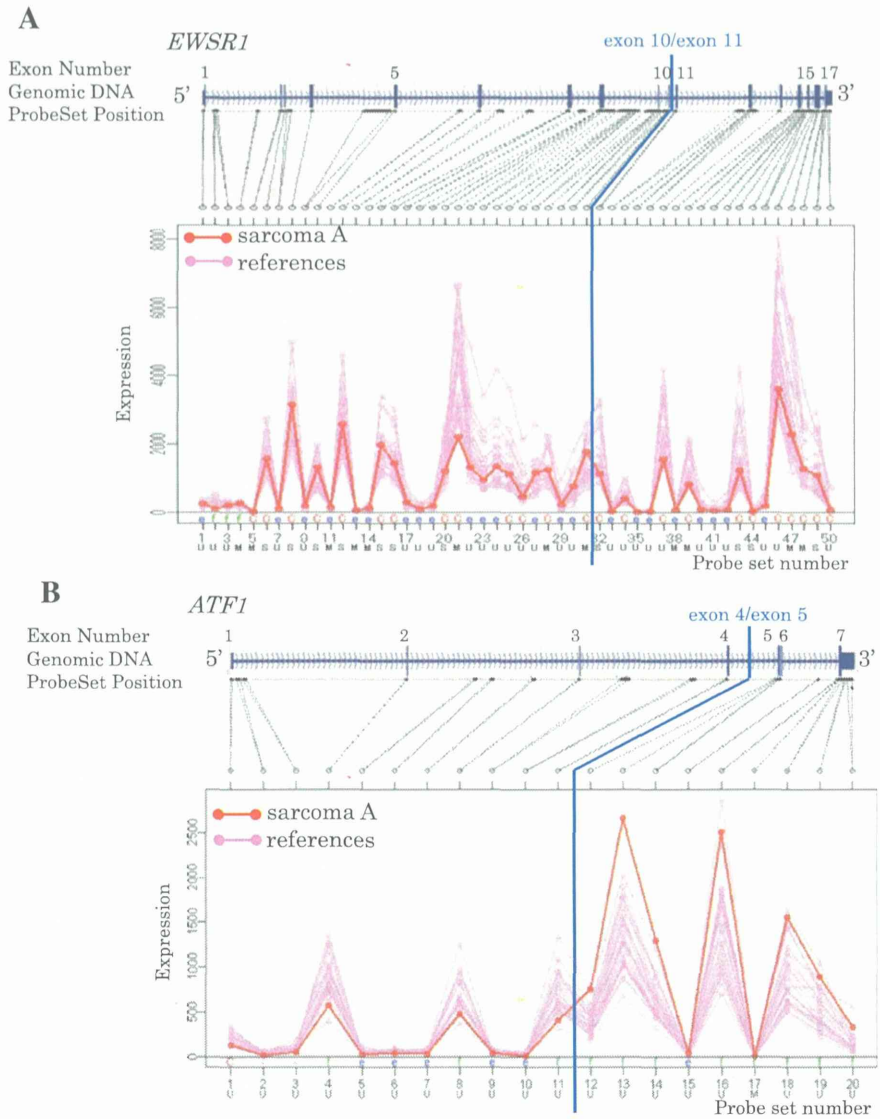

Figure 2 Exon Array data of fusion partner genes. Exon expression profiles of fusion partner genes EWSR1 **(A)** and ATF1 **(B)** are shown by line graphs with target areas of probe sets and genomic DNA structures. SarcomaA cell line and reference samples (breast cancer cell lines) are indicated by red and pink lines, respectively. Known breakpoints are shown by blue lines. Characters at top and bottom of probe set numbers indicate annotations: C = core, e = extended, f = full, U = unique, S = similar, M = mixed.

minimize the effect of a false signal, non-functional probes were removed in step 1, 2, and 3 of the computational program.

Comparison of expression on different probe sets:
Chromosome rearrangements often lead to the altered expression of 5′ or 3′ terminal regions of fusion partner genes by exchange of the transcriptional regulatory elements. The detection of sudden changes in the expression level between neighboring probe sets led to the discovery of breakpoints of fusion genes; however, the signal intensities obtained from different probes cannot be compared directly. Amplification and labeling efficiency are different in each RNA region. The hybridization property of probe sets on the array is also different in each probe set. Because of these biases, the signal intensity and dynamic range differ greatly between probe sets. Each probe set in the same gene has markedly different signal intensity; therefore, a normalizing method is needed to compare the signal intensities generated from different probe sets. On the other hand, signal intensities from different samples on the same probe sets

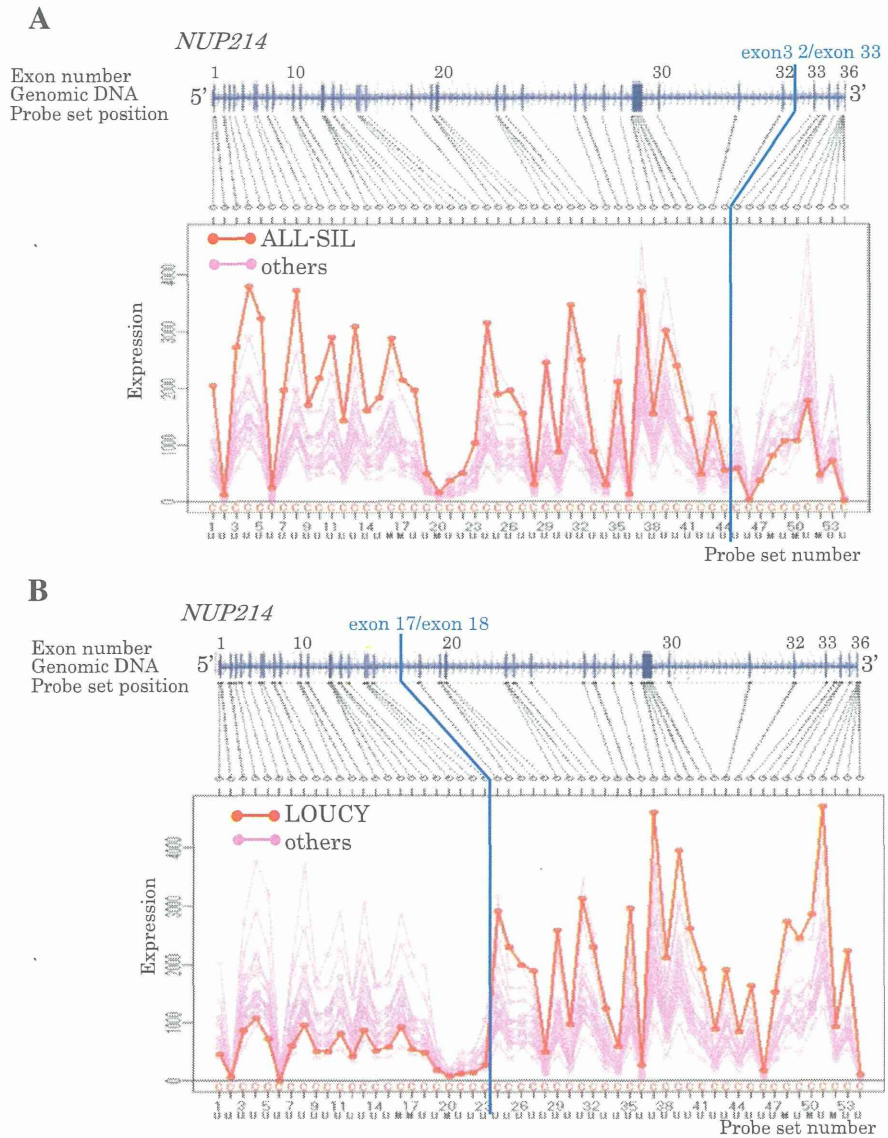

Figure 3 Expression profiles of NUP214. Exon expression profiles of NUP214 (ENST00000359428) are shown by line graphs with target areas of probe sets and genomic DNA structures. Examined cell line, **(A)**: ALL-SIL **(B)**: LOUCY, and reference samples (16 T-ALL cell lines) are indicated by red and pink lines, respectively. Predicted breakpoints are shown by blue lines. Characters at top and bottom of probe set numbers indicate annotations: C = core, U = unique, S = similar, M = mixed.

can be compared because the biases are the same for all samples. In the program, samples were ranked using the signal intensities for each probe set in a gene. The change in rank of a sample implies intragenic exon expression change.

Grouping and average calculation of probe sets:

Many genes have alternative transcript isoforms in vivo. Alternative splicing may contribute to expression differences between neighboring exons (probe sets), leading to a rank change. Moreover, because hybridization reactions on a great number of probes were performed under only one experimental condition in microarray experiments, non-specific cross hybridization cannot be avoided completely. The generated non-specific signals may influence the rank. Thus, rank changes between neighboring probe sets are thought to be observed frequently, and make it difficult to find the breakpoint. In the developed program, probe sets in the gene were divided into 5′ and 3′ terminal groups, and the average ranks of the probe set in each group were compared. The influences of unexpected rank changes were mitigated by this process.

Table 5 Candidate genes

	Transcript cluster ID	Gene symbol	Breakpoint		Examined sample
			Upstream probe set ID	Downstream probe set ID	
Breast	3075381	*ATP6V0A4*	3075407	3075406	DU4475
	3090697	*cDcA2*	3090726	3090727	UACCS93
	3883309	*C'EP250*	3883348	3883349	BT474
	3892812	*SLCO4AJ*	3892835	3892837	MDA-MB-231
Pancreas	3521174	*ABCC4*	3521233	3521248	MA005
	2331505	*MACF1*	2331398	2331419	MA028
	2331771	*RLF*	2331793	2331801	PA043

Exclusion of false positives because of quantitative determination error margin:

When the gene expression level is similar between samples, rank changes might take place at random due to quantitative determination error margins in Exon Array data, influencing the detection of breakpoints. False detection was decreased by monitoring the decentralization of a sample's rank.

The main feature of the program is that expression levels between probe sets can be compared by replacing the expression signal intensity with the rank. In general, expression levels were not compared between probe sets in gene or exon expression analysis by microarray. In this research, the developed program and evaluation of candidates chose seven candidate genes, and three novel fusion genes were identified by TAIL-PCR and RT-PCR; therefore, it is thought that the proposed method is very efficient for fusion gene discovery.

There existed fusion gene detection methods through transcript analysis by microarrays before. However, these methods were restrictive ones for confirmation of known fusion gene or for detecting some known partner genes [20-23].

The detection method for novel fusion genes using Exon Array has been reported by Eva Lin, in addition to this research [24]. Lin et al. detected intragenic expression changes of the *ALK* gene in lung, breast, and colon cancer. Based on their results, fusion gene *EML4-ALK* was identified using 5′RACE (rapid amplification cDNA end). Although fusion gene *EML4-ALK* was originally

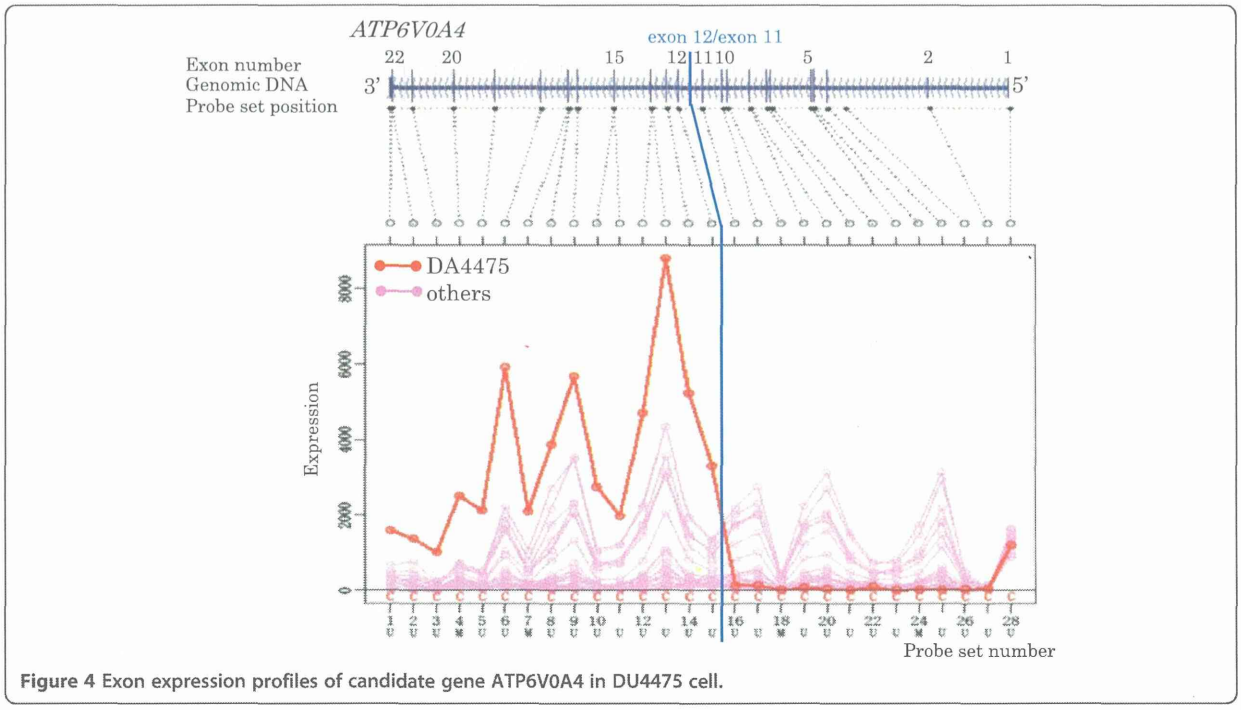

Figure 4 Exon expression profiles of candidate gene ATP6V0A4 in DU4475 cell.

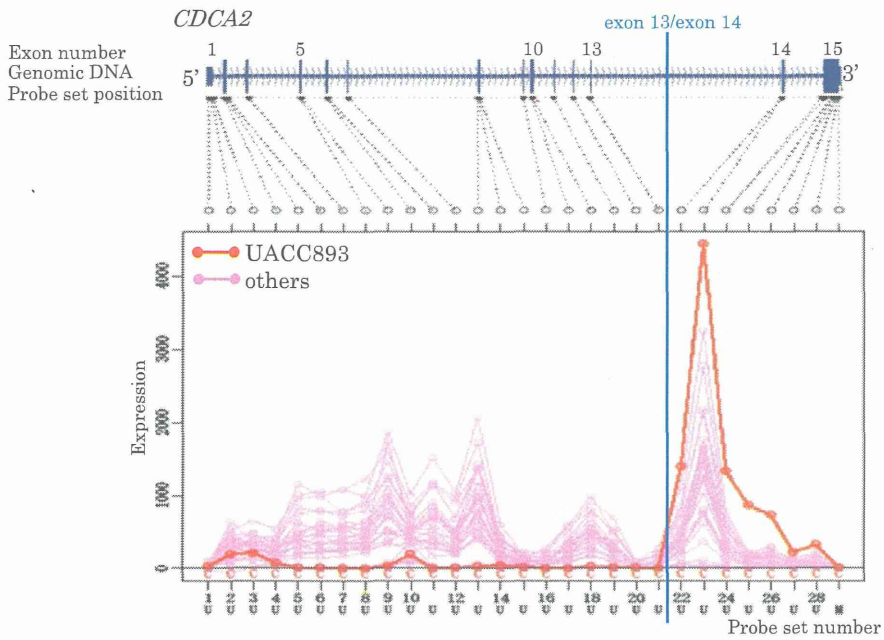

Figure 5 Exon expression profiles of candidate gene CDCA2 in UACC893 cell.

discovered in lung cancer, it had not been discovered in other cancers before their study. Their methods also detect the expression level change between 5′ and 3′ terminal groups of a gene for fusion gene discovery as well as this report. To compare the expression level between probe sets, they developed the following method. First, the mean value and standard deviation of the signal value of each probe set were calculated for all samples. Signal intensity was then standardized by subtracting its mean and dividing by its standard deviation. The

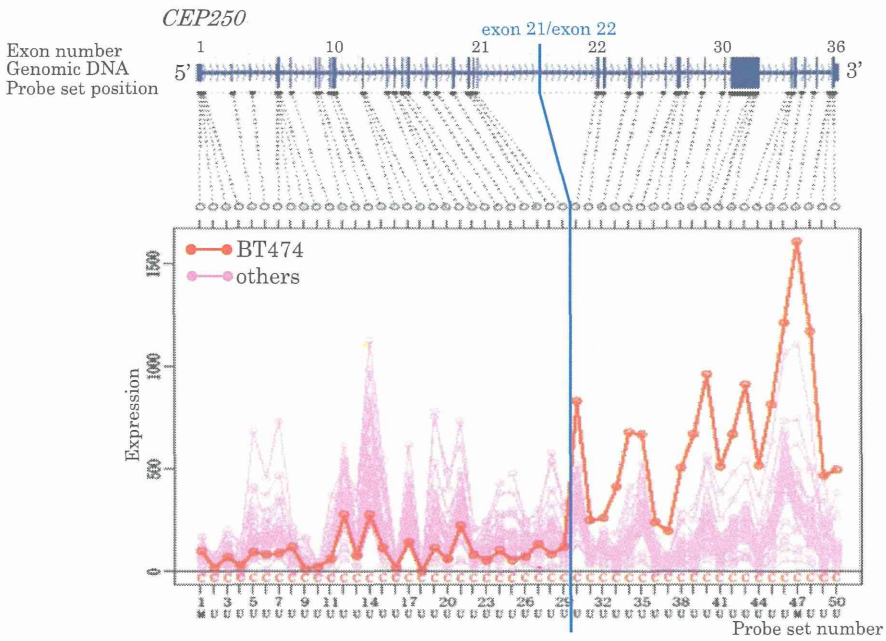

Figure 6 Exon expression profiles of candidate gene CEP250 in BT474 cell.

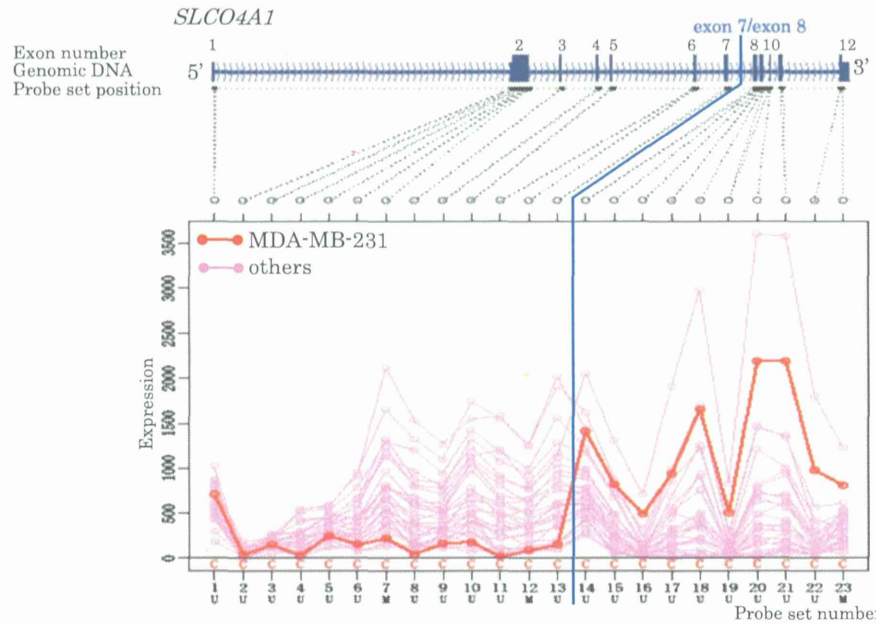

Figure 7 Exon expression profiles of candidate gene SLCO4A1 in MDA-MB-231 cell.

standardized value was used as an index of the expression level of each probe set. The probe sets were then separated in a transcript cluster into 5′ and 3′ terminal groups by one arbitrary point, and the expression level change was monitored between groups by t-test.

Comparing the proposed methodology with Lin's method, a common feature is that signal intensity is normalized based on the relative relation to reference samples, aiming to compare the expression levels of all probe sets in a gene. The most important difference is

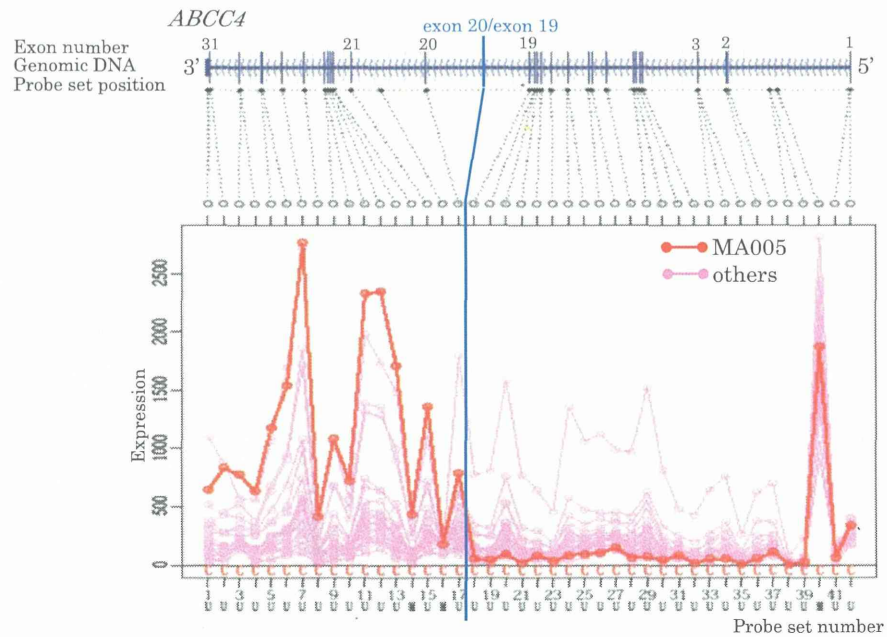

Figure 8 Exon expression profiles of candidate gene ABCC4 in MA005 cell.

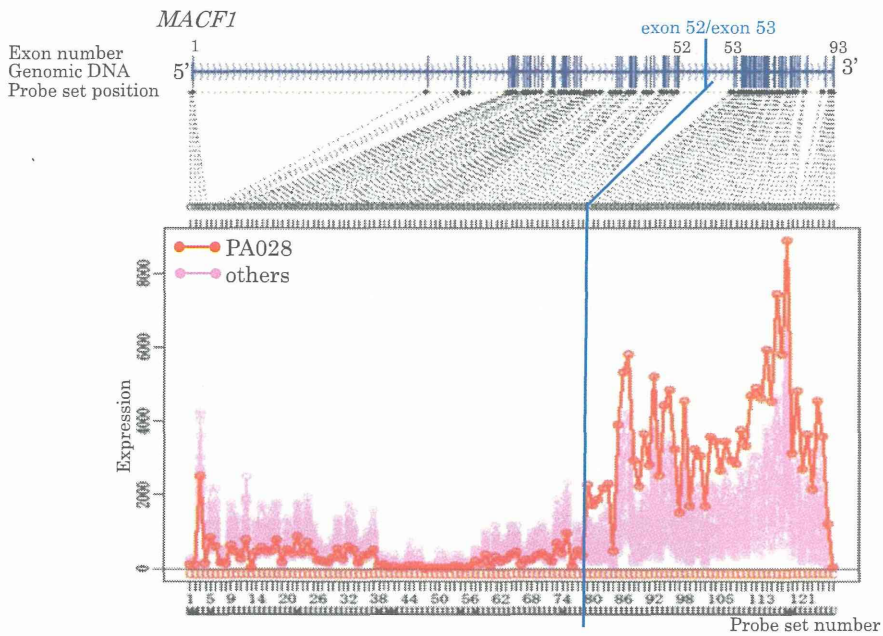

Figure 9 Exon expression profiles of candidate gene MACF1 in PA028 cell.

the strategy of normalizing. In Lin's method, it is thought that normalized values have a fixed quantity, which is an advantage to evaluate whether the magnitude of the change is significant; however, this is influenced easily by outlier intensities, which are generated frequently in microarray experiments. On the other hand, in the developed program, the magnitude of the change is not evaluated appropriately, but it has the advantage that the result is not influenced easily by the outlier value because the expression intensity is converted into the rank.

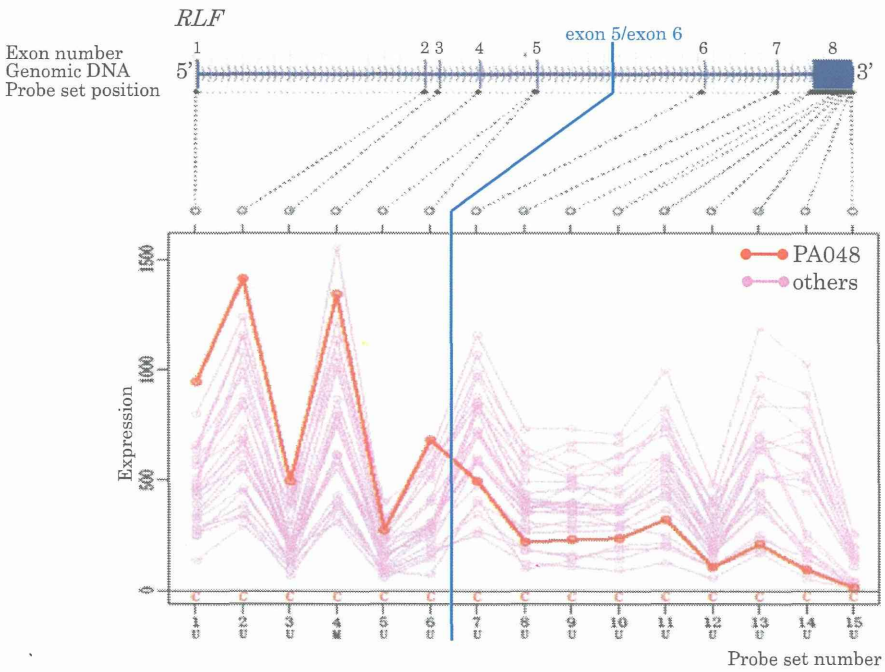

Figure 10 Exon expression profiles of candidate gene RLF in PA043 cell.

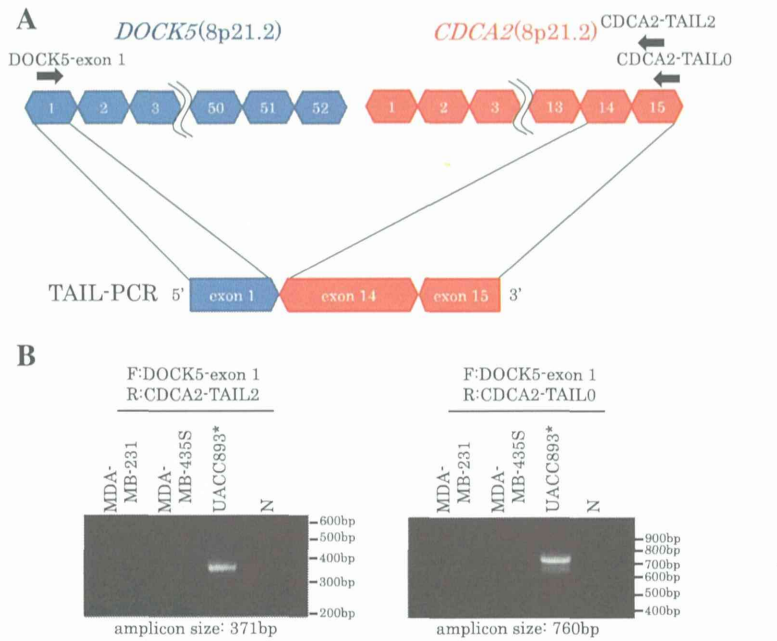

Figure 11 TAIL-PCR detection and RT-PCR confirmation of fusion gene DOCK5-CDCA2. Acquired fusion fragments by TAIL-PCR and exon structures of fusion partners are shown in **(A)** Red blocks are exons of candidate gene, blue blocks are exons of detected genes by TAIL-PCR. Arrows are primers for RT-PCR. RT-PCR confirmations for indicated samples are shown in **(B)** F: forward primer, R: reverse primer, *: detected samples in the program.

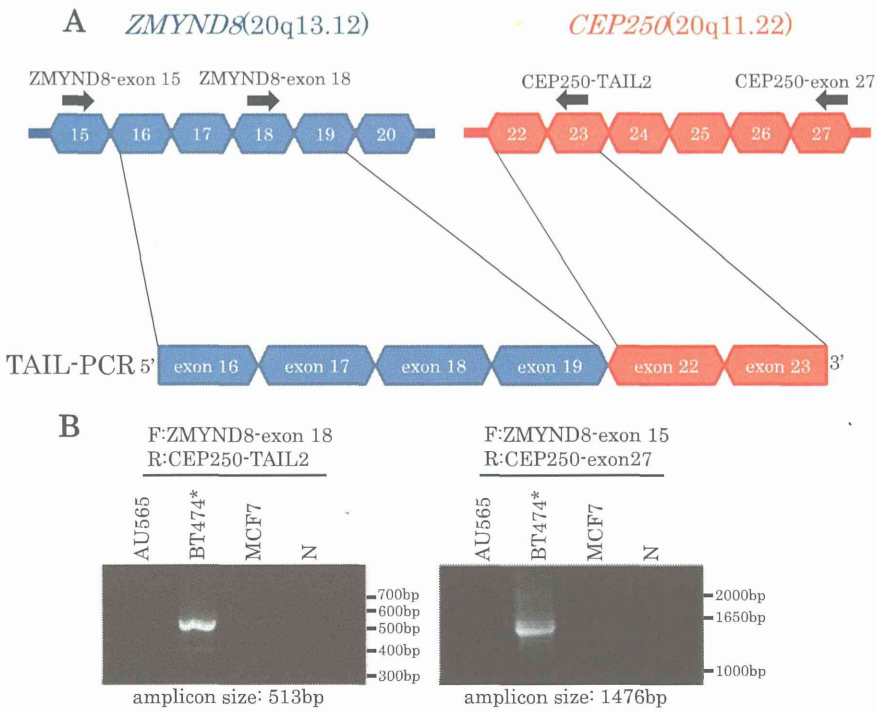

Figure 12 TAIL-PCR detection and RT-PCR confirmation of fusion gene ZMYND8- CEP250. Acquired fusion fragments by TAIL-PCR and exon structures of fusion partners are shown in **(A)** and **(B)** like Figure 11.

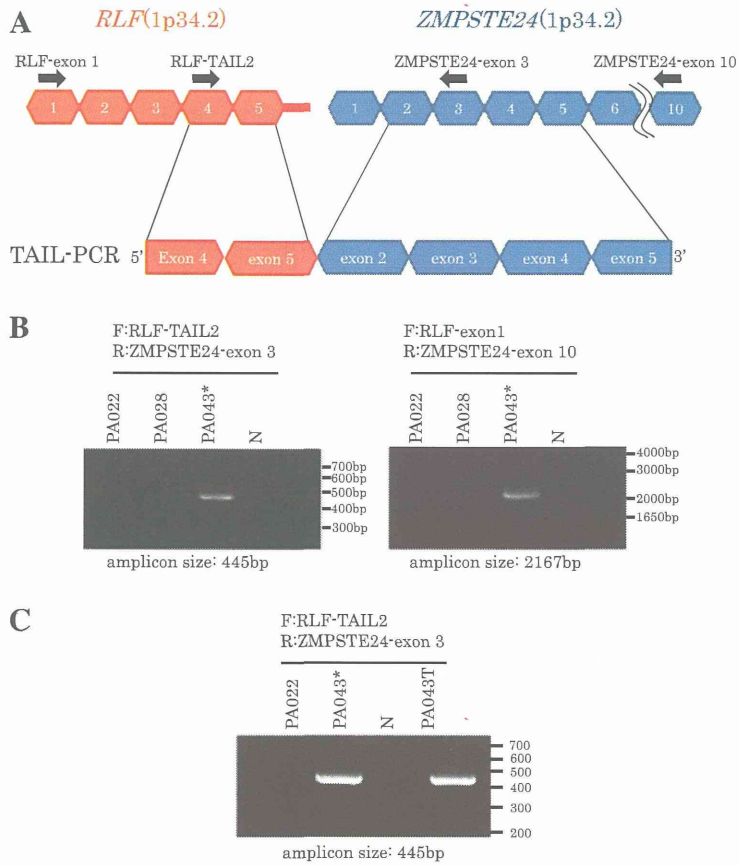

Figure 13 TAIL-PCR detection and RT-PCR confirmation of fusion gene RLFZMPSTE24. Acquired fusion fragments by TAIL-PCR and exon structures of fusion partners are shown in **(A)** and **(B)** like Figure 11, and RT-PCR detections for pancreatic cancer clinical tissue PA043T is shown in **(C)**.

Points to be improved and limitations

The analysis result would possibly change depending on the selection of reference samples, because signal intensities are converted into relative values by comparing with other samples. Lin's method has the same problem. It is thought that ideal reference samples for the program would show moderate variance of the gene expression level. Although cancer cell lines and healthy cells from the same organ were used in this research, further examination is necessary to assess whether this is the best choice. In addition, parameter optimization (degree of rank change, standard deviation and so on) for the reference samples is required.

The following points are limitations of this method, and alternative methods are needed. As this method detects the intragenic expression change in fusion partner genes, the method cannot detect the genes with no significant expression change between exons. Additionally, breakpoint detection from exon array data depends on the genomic position of the probe set. Thus, this method is not able to identify breakpoints on genomic DNA in detail.

Contribution of the fusion genes to cancer

The discovery of fusion genes that contribute to the pathology (tumorigenesis, metastasis etc.) are hoped from the viewpoint of the diagnosis and treatment of cancer. Considering the functional aspect of the fusion gene, it is important to incorporate other information, such as protein domain composition, when prioritizing novel, biologically relevant genomic aberrations [25].

Although three novel fusion genes were identified in this research, their function and contribution to cancer are unclear.

DOCK5-CDCA2:

DOCK5 (dedicator of cytokinesis 5) is a member of the DOCK family of guanine nucleotide exchange factors which function as activators of small G proteins [26]. Although DOCK5 is predicted to activate the small G

protein Rho and Rac, its function and signaling properties are poorly understood. CDCA2 (cell division cycle associated 2) recruits protein phosphatase 1 to mitotic chromatin at anaphase and into the following interphase, regulating the chromosome structure during mitosis [27]. Because *DOCK5* and *CDCA2* show out-of-frame fusion, it is thought that the amino acid sequence of *CDCA2* is disrupted and a premature termination codon appears in *CDCA2* exon 14. The fusion gene might therefore produce a short protein, 42aa (14aa from *DOCK5* exon 1, and 28aa from *CDCA2* exon 14). No functional protein domains have been found so the function of the fusion protein is unclear. Significant chromosome loss and underexpression of *DOCK5* have been reported in osteosarcoma [28]. DOCK5 dysfunction might contribute to tumors.

ZMYND8-CEP250:

ZMYND8 is a member of RACK (receptor for activated C-kinase) family proteins that anchor activated protein kinase C (PKC). ZMYND8 interacts specifically with PKCβI and is predicted to regulate subcellular localization and activity [29]. In addition, ZMYND8 contains a bromo domain, a PWWP domain, and two zinc fingers, and is thought to be a transcriptional regulator. CEP250 is a core centrosomal protein required for centriole-centriole cohesion during interphase of the cell cycle [30], but details of the mechanism are not well known. *ZMYND8-CEP250* is also an out-of-frame fusion gene, so a premature termination codon appears in *CEP250* exon 24 and is likely to express a 1121aa protein (994aa from *ZMYND8* exon 1–19, and 127aa from *CEP250* exon 22–24). The down-regulation of PKCβ1 protein expression has been reported in colon cancer [31]. The PKCβ1 binding site in the C terminal region of ZMYND8 racks in the predicted fusion protein. Formation of the fusion gene may lead to the low activity of PKCB1, and may contribute to cancer, or deregulation of the transcript regulatory network managed by ZMYND8 might cause cancer.

RLF-ZMPSTE24:

RLF is predicted as a transcription factor with zinc fingers from the amino acid sequence. It is reported that RLF forms a fusion gene with the LMYC gene in lung cancer [32]. The fusion gene RLF-LMYC contributes to carcinogenesis by changing the LMYC manifestation of a gene [33]. ZMPSTE24 performs a critical endoproteolytic cleavage step to generate mature lamin A, a major component of the nuclear lamina and nuclear skeleton [34]. Lack of functional ZMPSTE24 results in progeroid phenotypes, including genomic instability in mice and humans [35,36]. *RLF-ZMPSTE24* is an in-frame fusion gene, which may expresses the 704aa protein (270aa from *RLF* exon 1–5, and 434aa from *ZMPSTE24* exon 2–10). The known function domains of *RLF* are not contained in the fusion gene, and no change of *ZMPSTE24* expression level is observed in Exon Array data. Functional change of ZMPSTE24 may induce DNA damage and lead to cancer.

Genomic structure of the fusion genes

RLF and *ZMPSTE24* genes located on chromosome 1, approximately 20 kb apart, have the same orientation. Southern blot analysis with a probe hybridizing to *RLF* intron 5 region showed chromosome rearrangement (data not shown), and a fragment that is part of *RLF* intron 5 fused to a part of *ZMPSTE24* intron 1 was obtained by TAIL-PCR for the upstream region of *ZMPSTE24* exon 2 on genomic DNA (data not shown). Both parts fused in the opposite orientation; therefore, the cause of the gene fusion, *RLF-ZMPSTE24*, might be chromosome inversion with some deletion. *ZMYND8* and *CEP250* genes were located on chromosome 20, approximately 12 Mb apart, in opposite orientation. *DOCK5* and *CDCA2* genes were located on chromosome 8, approximately 50Kb apart, in the same orientation. The mechanisms of gene fusions remain to be revealed.

The proposed method might be applied to not only Exon Array but also the Affymetrx GeneChip Gene 1.0 ST Array (Gene Array) with some improvements. Gene Array, in which each of the 28,869 genes is represented on the array by approximately 26 probes spread along the full length of the gene, is widely used for global gene expression analysis. Using this method for more samples, it is thought that fusion genes can be identified. This is expected to lead to new diagnostic methods and treatment strategies.

Additional files

Additional file 1: Selected genes by the program in 24 breast cancer cell lines.

Additional file 2: Selected genes by the program in 20 pancreatic cell lines.

Competing interests

The authors declare that they have no competing financial interests or other conflicts of interest.

Authors' contributions

YW carried out sample preparation, Exon Array Experiments, TAIL-PCR, manuscript writing and helped to develop the algorithms. MM developed the majority of algorithms, helped to draft the manuscript. MS contributed cell cultue and TAIL-PCR, helped to draft the manuscript. MU and SM contributed statistical support and data processing. KN contributed preparation of clinical samples. TN and YM contributed to study conception, and critical manuscript review. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Dr. T Nakamura (The Cancer Institute, JFCR) for providing sarcoma cell lines. This research was partially supported by the New Energy and Industry Technology Development Organization.

Author details
[1]Genome Center, Japanese Foundation for Cancer Research, 3-8-31 Ariake, Koto-ku, Tokyo 135-8550, Japan. [2]Cancer Institute, Japanese Foundation for Cancer Research, 3-8-31 Ariake, Koto-ku, Tokyo 135-8550, Japan. [3]Department of Cardiovascular Medicine, Tohoku University Graduate School of Medicine, 2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi 980-8575, Japan. [4]Current address: FASMAC Co., Ltd, 3088 Okata, Atsugi City, Kanagawa 243-0041, Japan.

Received: 6 December 2013 Accepted: 4 February 2014
Published: 18 February 2014

References

1. Nowell P, Hungerford D: **A minute chromosome in human chronic granulocytic leukemia.** *Science* 1960, **132**:1497.
2. Rowley JD: **Letter: a new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining.** *Nature* 1973, **243**(5405):290–293.
3. Shtivelman E, Lifshitz B, Gale RP, Canaani E: **Fused transcript of abl and bcr genes in chronic myelogenous leukaemia.** *Nature* 1985, **315**(6020):550–554.
4. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM: **Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer.** *Science* 2005, **310**(5748):644–648.
5. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H: **Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer.** *Nature* 2007, **448**(7153):561–566.
6. Mitelman F, Johansson B, Mertens F: **Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer.** *Nat Genet* 2004, **36**(4):331–334.
7. Shaw AT, Hsu PP, Awad MM, Engelman JA: **Tyrosine kinase gene rearrangements in epithelial malignancies.** *Nat Rev Cancer* 2013, **13**(11):772–787.
8. Speicher MR, Carter NP: **The new cytogenetics: blurring the boundaries with molecular biology.** *Nat Rev Genet* 2005, **6**(10):782–792.
9. Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, Santarius T, Stebbings LA, Leroy C, Edkins S, Hardy C, Teague JW, Menzies A, Goodhead I, Turner DJ, Clee CM, Quail MA, Cox A, Brown C, Durbin R, Hurles ME, Edwards PA, Bignell GR, Stratton MR, Futreal PA: **Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing.** *Nat Genet* 2008, **40**(6):722–729.
10. Lipson D, Capelletti M, Yelensky R, Otto G, Parker A, Jarosz M, Curran JA, Balasubramanian S, Bloom T, Brennan KW, Donahue A, Downing SR, Frampton GM, Garcia L, Juhn F, Mitchell KC, White E, White J, Zwirko Z, Peretz T, Nechushtan H, Soussan-Gutman L, Kim J, Sasaki H, Kim HR, Park SI, Ercan D, Sheehan CE, Ross JS, Cronin MT, Jänne PA, Stephens PJ: **Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies.** *Nat Med* 2012, **18**(3):382–384.
11. Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, Chinnaiyan AM: **Transcriptome sequencing to detect gene fusions in cancer.** *Nature* 2009, **458**(7234):97–101.
12. Wu YM, Su F, Kalyana-Sundaram S, Khazanov N, Ateeq B, Cao X, Lonigro RJ, Vats P, Wang R, Lin SF, Cheng AJ, Kunju LP, Siddiqui J, Tomlins SA, Wyngaard P, Sadis S, Roychowdhury S, Hussain MH, Feng FY, Zalupski MM, Talpaz M, Pienta KJ, Rhodes DR, Robinson DR, Chinnaiyan AM: **Identification of targetable FGFR gene fusions in diverse cancers.** *Cancer Discov* 2013, **3**(6):636–647.
13. Polyak K: **Breast cancer: origins and evolution.** *J Clin Invest* 2007, **117**(11):3155–3163.
14. Simpson PT, Reis-Filho JS, Gale T, Lakhani SR: **Molecular evolution of breast cancer.** *J Pathol* 2005, **205**(2):248–254.
15. Mackenzie RP, McCollum AD: **Novel agents for the treatment of adenocarcinoma of the pancreas.** *Expert Rev Anticancer Ther* 2009, **9**(10):1473–1485.
16. Graux C, Cools J, Melotte C, Quentmeier H, Ferrando A, Levine R, Vermeesch JR, Stul M, Dutta B, Boeckx N, Bosly A, Heimann P, Uyttebroeck A, Mentens N, Somers R, MacLeod RA, Drexler HG, Look AT, Gilliland DG, Michaux L, Vandenberghe P, Wlodarska I, Marynen P, Hagemeijer A: **Fusion of NUP214 to ABL1 on amplified episomes in T-cell acute lymphoblastic leukemia.** *Nat Genet* 2004, **36**(10):1084–1089.
17. Van Vlierberghe P, van Grotel M, Tchinda J, Lee C, Beverloo HB, van der Spek PJ, Stubbs A, Cools J, Nagata K, Fornerod M, Buijs-Gladdines J, Horstmann M, van Wering ER, Soulier J, Pieters R, Meijerink JP: **The recurrent SET-NUP214 fusion as a new HOXA activation mechanism in pediatric T-cell acute lymphoblastic leukemia.** *Blood* 2008, **111**(9):4668–4680.
18. Liu YG, Chen Y: **High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences.** *Biotechniques* 2007, **43**(5):649–650. 652, 654 passim.
19. Liu YG, Whittier RF: **Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking.** *Genomics* 1995, **25**(3):674–681.
20. Jhavar S, Reid A, Clark J, Kote-Jarai Z, Christmas T, Thompson A, Woodhouse C, Ogden C, Fisher C, Corbishley C, De-Bono J, Eeles R, Brewer D, Cooper C: **Detection of TMPRSS2-ERG translocations in human prostate cancer by expression profiling using GeneChip Human Exon 1.0 ST arrays.** *J Mol Diagn* 2008, **10**(1):50–57.
21. Løvf M, Thomassen GOS, Bakken AC, Celestino R, Fioretos T, Lind GE, Lothe RA, Skotheim RI: **Fusion gene microarray reveals cancer type-specificity among fusion genes.** *Genes Chromosomes Cancer* 2011, **50**(5):348–357.
22. Nasedkina T, Domer P, Zharinov V, Hoberg J, Lysov Y, Mirzabekov A: **Identification of chromosomal translocations in leukemias by hybridization with oligonucleotide microarrays.** *Haematologica* 2002, **87**(4):363–372.
23. Skotheim RI, Thomassen GOS, Eken M, Lind GE, Micci F, Ribeiro FR, Cerveira N, Teixeira MR, Heim S, Rognes T, Lothe RA: **A universal assay for detection of oncogenic fusion transcripts by oligo microarray analysis.** *Mole Cancer* 2009, **8**:5.
24. Lin E, Li L, Guan Y, Soriano R, Rivers CS, Mohan S, Pandita A, Tang J, Modrusan Z: **Exon array profiling detects EML4-ALK fusion in breast, colorectal, and non-small cell lung cancers.** *Mol Cancer Res* 2009, **7**(9):1466–1476.
25. Ortiz De Mendibil I, Vizmanos JL, Novo FJ: **Signatures of selection in fusion transcripts resulting from chromosomal translocations in human cancer.** *PLoS One* 2009, **4**(3):e4805.
26. Meller N, Merlot S, Guda C: **CZH proteins: a new family of Rho-GEFs.** *J Cell Sci* 2005, **118**(Pt 21):4937–4946.
27. Trinkle-Mulcahy L, Andersen J, Lam YW, Moorhead G, Mann M, Lamond AI: **Repo-Man recruits PP1 gamma to chromatin and is essential for cell viability.** *J Cell Biol* 2006, **172**(5):679–692.
28. Sadikovic B, Yoshimoto M, Chilton-MacNeill S, Thorner P, Squire JA, Zielenska M: **Identification of interactive networks of gene expression associated with osteosarcoma oncogenesis by integrated molecular profiling.** *Hum Mol Genet* 2009, **18**(11):1962–1975.
29. Fossey SC, Kuroda S, Price JA, Pendleton JK, Freedman BI, Bowden DW: **Identification and characterization of PRKCBP1, a candidate RACK-like protein.** *Mamm Genome* 2000, **11**(10):919–25.
30. Mayor T, Stierhof YD, Tanaka K, Fry AM, Nigg EA: **The centrosomal protein C-Nap1 is required for cell cycle-regulated centrosome cohesion.** *J Cell Biol* 2000, **151**(4):837–846.
31. Dempsey EC, Newton AC, Mochly-Rosen D, Fields AP, Reyland ME, Insel PA, Messing RO: **Protein kinase C isozymes and the regulation of diverse cell responses.** *Am J Physiol Lung Cell Mol Physiol* 2000, **279**(3):L429–L438.
32. Makela TP, Kere J, Winqvist R, Alitalo K: **Intrachromosomal rearrangements fusing L-myc and rlf in small-cell lung cancer.** *Mol Cell Biol* 1991, **11**(8):4015–21.
33. Makela TP, Hellsten E, Vesa J, Hirvonen H, Palotie A, Peltonen L, Alitalo K: **The rearranged L-myc fusion gene (RLF) encodes a Zn-15 related zinc finger protein.** *Oncogene* 1995, **11**(12):2699–704.
34. Young SG, Meta M, Yang SH, Fong LG: **Prelamin A farnesylation and progeroid syndromes.** *J Biol Chem* 2006, **281**(52):39741–39745.
35. Liu B, Wang J, Chan KM, Tjia WM, Deng W, Guan X, Huang JD, Li KM, Chau PY, Chen DJ, Pei D, Pendas AM, Cadinanos J, Lopez-Otin C, Tse HF, Hutchison C, Chen J, Cao Y, Cheah KS, Tryggvason K, Zhou Z: **Genomic instability in laminopathy-based premature aging.** *Nat Med* 2005, **11**(7):780–785.

36. Navarro CL, Cadinanos J, De Sandre-Giovannoli A, Bernard R, Courrier S, Boccaccio I, Boyer A, Kleijer WJ, Wagner A, Giuliano F, Beemer FA, Freije JM, Cau P, Hennekam RC, Lopez-Otin C, Badens C, Levy N: **Loss of ZMPSTE24 (FACE-1) causes autosomal recessive restrictive dermopathy and accumulation of Lamin A precursors.** *Hum Mol Genet* 2005, **14**(11):1503–1513.

doi:10.1186/2043-9113-4-3
Cite this article as: Wada *et al.*: Development of detection method for novel fusion gene using GeneChip exon array. *Journal of Clinical Bioinformatics* 2014 **4**:3.

Tumor and Stem Cell Biology

Cancer Research

BRCA2 Phosphorylated by PLK1 Moves to the Midbody to Regulate Cytokinesis Mediated by Nonmuscle Myosin IIC

Miho Takaoka[1], Hiroko Saito[2], Katsuya Takenaka[1], Yoshio Miki[1,2], and Akira Nakanishi[1]

Abstract

Cytokinesis is the critical final step in cell division. BRCA2 disruption during cytokinesis is associated with chromosome instability, but mechanistic information is lacking that could be used to prevent cancer cell division. In this study, we report that BRCA2 phosphorylation by the mitotic polo-like kinase (PLK1) governs the localization of BRCA2 to the Flemming body at the central midbody, permitting an interaction with nonmuscle myosin IIC (NM-IIC). Formation of an NM-IIC ring-like structure at the Flemming body shows that the IIC-ring relies on its ATPase activity stimulated by interaction with BRCA2 and associated proteins. Notably, inhibiting this binding inactivated the ATPase activity, causing disassembly of the IIC-ring, defective formation of the midbody, and interruption of cytokinesis. An analysis of cancer-associated mutations in BRCA2 at the PLK1-binding site suggests that they may contribute to cytokinetic defects by altering BRCA2 localization. Our findings suggest that BRCA2-dependent IIC-ring formation is a critical step in proper formation of the midbody, offering an explanation for how chromosome instability may arise in breast cancer. *Cancer Res; 74(5); 1518–28. ©2014 AACR.*

Introduction

Germline mutations in the *BRCA2* gene have been reported to increase the risk of developing breast and ovarian cancer. The BRCA2 protein has multiple functions, including DNA double-strand break repair (1–3), and the regulation of centrosome amplification and localization (4, 5). BRCA2 also contributes to the regulation of cytokinesis (6, 7). During anaphase, constriction of the actomyosin ring leads to formation of a cleavage furrow (8–10). Continued furrowing results in the formation of a narrow intercellular bridge, which contains the midbody, consisting of a bundled microtubule and a ring-like structure called the Flemming body within its central portion (11). Several studies have implicated various protein factors in the modulation of cytokinesis by BRCA2 (12–14). Disruption of BRCA2 during cytokinesis leads to disorganization of myosin-II at the cleavage furrow and the intercellular bridge (6).

Mitotic polo-like kinase 1 (Plk1) is a key regulator of mitosis from mitotic initiation to cytokinesis. Plk1 contains a serine/threonine kinase domain followed by the carboxy-terminal polo-box domain (PBD), which binds to phosphopeptides within a consensus motif of S-[pS/pT]-[P/X] (15, 16). The PBD regulates cellular function, the interaction with substrates, and the subcellular localization of Plk1 (17). Plk1 is also required for appropriate localization of substrates (18). Previous studies revealed that Plk1 binds to the N-terminal region of BRCA2 and phosphorylates Ser193, and that this phosphorylation is enhanced as mitosis progresses (19, 20). However, the specific Plk1-binding site within this region of BRCA2 has not been identified. The role of BRCA2 phosphorylation in the maintenance of genome stability also remains unclear.

Nonmuscle myosin-II (NM-II) proteins in humans are hexamers, consisting of a pair of heavy chains and two pairs of light chains that hydrolyze MgATP. They are members of a family of actin-binding motor proteins that play essential roles in cellular processes such as cell division and embryonic development. The NM-II family comprises three isoforms: NM-IIA, NM-IIB, and NM-IIC. These contain different nonmuscle myosin heavy chains (NMHC-IIA, NMHC-IIB, and NMHC-IIC) that are encoded by *MYH* (myosin heavy chain) *9, MYH10*, and *MYH14*, respectively (21, 22). The N-terminal region of the NMHC-II protein consists of a globular head containing the actin-binding region, an ATPase domain, and a Src homology 3 (SH3)-like domain (myosin head; refs. 23, 24). NM-IIC is alternatively spliced both in loop-1 and loop-2. Isoform NM-IIC0 contains no inserts in either of the loops. An 8–amino acid extension in the loop-1 region is present in isoforms NM-IIC1 and NM-IIC1C2. Isoform NM-IIC1C2 displays a 33–amino acid extension in the loop-2 region. The presence of 8 amino acid insert in NM-IIC increases the actin-activated ATPase activity (25). The C-terminal deletion isoform of NM-IIC is already present in the Ensembl database (http://www.ensembl.org/index.html; MYH14-007, Protein ID: ENSP00000469573). Jana

Authors' Affiliations: [1]Department of Molecular Genetics, Medical Research Institute, Tokyo Medical & Dental University, Bunkyo-Ku; and [2]Department of Genetic Diagnosis, the Cancer Institute of JFCR, Koto-Ku, Tokyo, Japan

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Corresponding Author: Yoshio Miki, Department of Molecular Genetics, Medical Research Institute, Tokyo Medical & Dental University, Yushima 1-5-45, Bunkyo-Ku, Tokyo 113-8510, Japan. Phone: 81-3-5803-5825; Fax: 81-3-5803-0242; E-mail: miki.mgen@mri.tmd.ac.jp

doi: 10.1158/0008-5472.CAN-13-0504

and colleagues reported the localization of NM-IIC to the midbody during the process of abscission. In that report, they confirmed that NM-IIC1 is required for cytokinesis (25). Accordingly, we made use of this isoform. In this study, we suggest that BRCA2-dependent IIC-ring formation represents a key step in proper midbody formation, and that a hereditary breast cancer-associated mutation within the Plk1 interaction motif of BRCA2 affects the localization of BRCA2 to the Flemming body, resulting in cytokinesis defects.

Materials and Methods

Detailed descriptions of plasmids, transfections, siRNA treatment, antibodies, immunofluorescence and three-dimensional (3D) reconstitutions, immunoprecipitation, Western blot analysis, binding assays, glycerol density gradient centrifugation, time-lapse microscopy, measurement of ATPase activity, cell-cycle analysis, and statistical analysis are provided in the Supplementary Materials and Methods.

Cell culture

HeLaS3 and COS-7 cells were purchased from the RIKEN GENBANK, and U2OS and A549 were from American Type Culture Collection (ATCC). These cells were cultured in Dulbecco's Modified Eagle Medium containing 10% fetal calf serum (FCS). MCF7 was purchased from ATCC and cultured in Eagle's Minimum Essential Medium containing 10% FCS, 0.01 mg/mL bovine insulin, and 1% sodium pyruvate. The identities of HeLaS3, U2OS, A549, and MCF7 cell lines were confirmed by short tandem repeat (STR) profiling by BEX CO., Ltd and used within 6 months of testing. COS-7 cells (monkey cell line) were not analyzed by STR profiling because STR profiling is method for authentication of human cell lines.

Midbody isolation

Mitotic cells (A549) were released from colcemid treatment (50 ng/mL) for 17 hours by washing twice with fresh medium. After incubating at 37°C for 60 minutes, the cells were collected. Midbody isolation was performed according to the methods of Mullins and McIntosh (26).

Measurement of ATPase activity

COS-7 cells were transfected with plasmids expressing HA-tagged NMHC-IIC and FLAG-tagged BRCA2. After 24 hours, cells (5×10^6) were lysed in lysis buffer [20 mmol/L Tris-HCl (pH 8.0), 100 mmol/L NaCl, 1 mmol/L EDTA, 1% NP-40, and protease inhibitors]. BRCA2 and NMHC-IIC were immunoprecipitated by anti-FLAG or anti-HA antibodies, respectively. Each sample was eluted by 40 μL elution buffer (50 mmol/L glycine, pH 2.8), which was immediately neutralized with 1 mol/L Tris. BRCA2-FLAG and NMHC-IIC-HA were mixed and incubated at room temperature for 1 minute and then placed into a solution for measurement of ATPase activity. F-actin was added immediately before the measurement. ADP production during ATP hydrolysis was measured as a change in NADH concentration. NADH absorbance was measured every 2 seconds at 340 nm and the rate of NADH consumption was interpreted as the ATPase activity.

Results

Localization of BRCA2 to the Flemming body through phosphorylation by Plk1

To analyze the role of phosphorylation of BRCA2 Ser193 by Plk1, we generated an antibody (193pBRCA2) specific for this phosphoprotein and confirmed the specificity of the 193pBRCA2 antibody by performing a competition assay using pS^{193}-Peptides and treatment with protein phosphatase (PP2A; Supplementary Fig. S1A and S1B). This antibody recognized Ser^{193}-phosphorylated BRCA2 (pS^{193}-BRCA2) in anti-BRCA2 (Ab-1) immunoprecipitates from midbody lysates (Supplementary Fig. S1C).We observed its localization during the cell cycle by immunofluorescence microscopy. BRCA2 localized to the midbody (particularly at the Flemming body) during cytokinesis of mitotic HeLaS3 cells (Fig. 1A). Similar findings were observed in A549, MCF7, and U2OS cells (Supplementary Fig. S1D). The phosphorylated BRCA2 also localized to the Flemming body (Fig. 1A). In contrast, BRCA1 did not localize to the Flemming body (Fig. 1A and Supplementary Fig. S1E). This result is consistent with that of a previous study by another group (13). We then isolated the midbodies from HeLaS3 cells (Supplementary Fig. S1F) and compared the extent of BRCA2 phosphorylation in midbody lysates with that in whole-cell lysates (Fig. 1B). A greater fraction of the BRCA2 was phosphorylated in the midbody. Using immunofluorescence microscopy, we showed that Plk1 colocalizes with pS^{193}-BRCA2 at the Flemming body (Fig. 1C). Assays of COS-7 cells transfected with BRCA2-FLAG and Plk1-HA demonstrated that BRCA2 and Plk1 coimmunoprecipitated (Fig. 1D). We located a Plk1-binding motif (S-T-P sequence at codons 76–78) in the N-terminal region of BRCA2. A Thr77-phosphorylated peptide corresponding to this region (a.a. 73–82 of BRCA2) was synthesized for *in vitro* binding analyses. Streptavidin bead–immobilized p-Thr77 peptide precipitated Plk1 from mitotic HeLaS3 cell lysates, whereas unphosphorylated peptide or peptide-free streptavidin beads did not (Fig. 1E). Plk1 was not precipitated from S-phase lysates due to low expression.

This Plk1-binding site is conserved among diverse species (Supplementary Fig. S1G). A hereditary breast cancer-associated missense mutation (T77A) within this site is annotated in the Breast Cancer Information Core (BIC) database (http://research.nhgri.nih.gov/bic/). We constructed plasmids encoding a truncated N-terminal region [BRCA2 (R1)-FLAG (WT):1–157 a.a.] containing either the wild-type Plk1-binding site or this mutant, BRCA2 (R1)-FLAG (T77A; Supplementary Fig. S1H). The mutant protein was unable to interact with endogenous Plk1 in HeLaS3 cells (Supplementary Fig. S1I), indicating that this mutation disrupted the interaction with Plk1. Next, we examined whether overexpression of this truncated region could compete with endogenous BRCA2. Coimmunoprecipitation assays were performed in HeLaS3 cells expressing BRCA2 (R1)-FLAG (WT). It was demonstrated that this region inhibited the binding of endogenous BRCA2 to Plk1 (Supplementary Fig. S1J). Also, the endogenous BRCA2 ceased to be detected at the Flemming body following ectopic expression of this region (Supplementary Fig. S1K). These results suggested a dominant-negative effect of HA-BRCA2 (R1) upon endogenous BRCA2. A distinct HA-tagged BRCA2 region [HA-BRCA2 (R2):

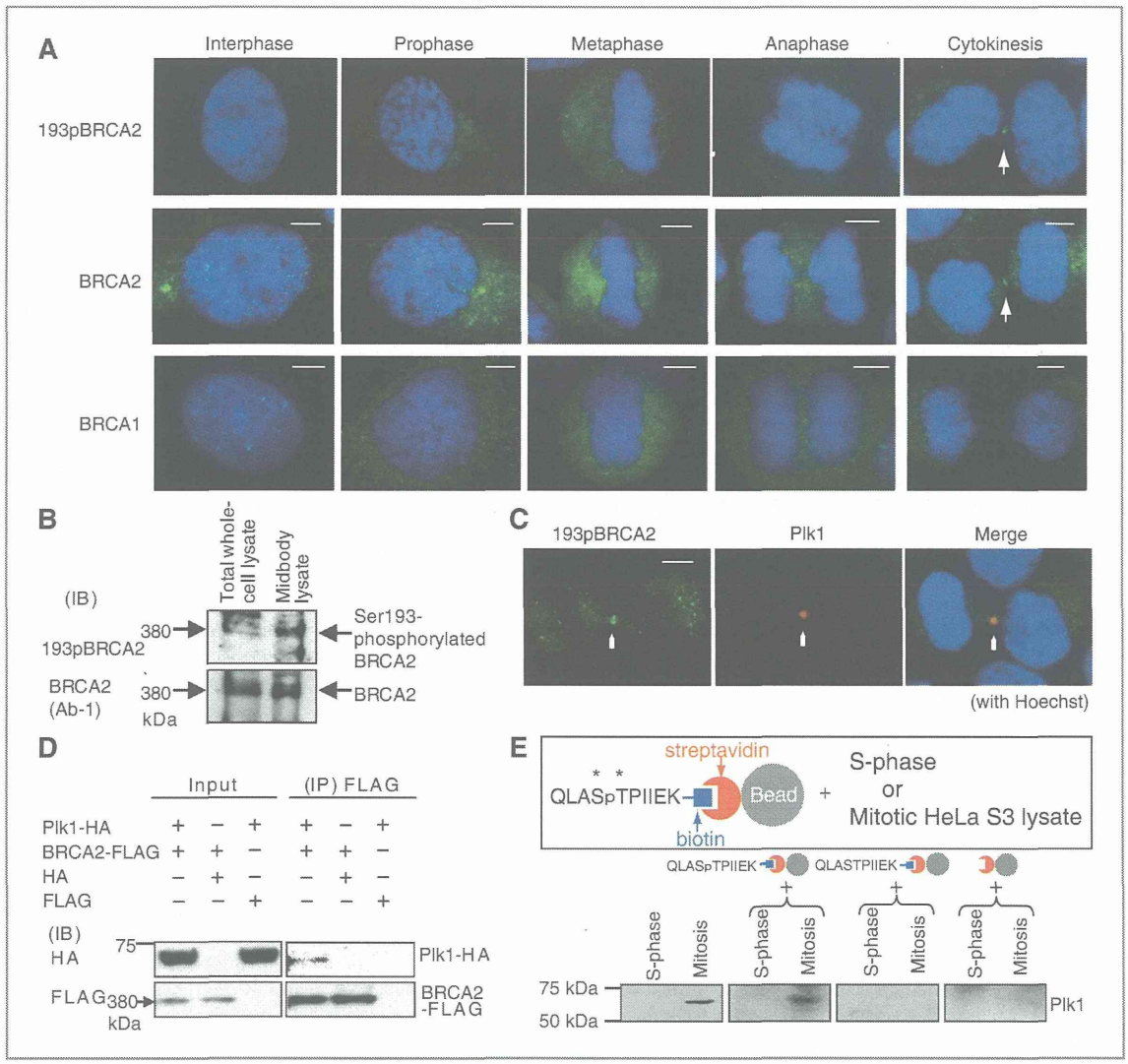

Figure 1. Plk1-phosphorylated BRCA2 localizes to the Flemming body in HeLaS3 cells. A, cells were fixed and stained for Ser193-phosphorylated BRCA2 (pS193-BRCA2), BRCA2, and BRCA1 throughout the cell cycle. Nuclei were detected by Hoechst 33258 stain (blue). Arrows, the Flemming body. Scale bar, 5 μm. B, total whole-cell and midbody lysates were subjected to Western blot analysis using antibodies against pS193-BRCA2 and BRCA2. IB, immunoblot. C, cells were fixed and stained for pS193-BRCA2 (green) and Plk1 (red) in cytokinesis. White arrows, the Flemming body. Scale bar, 5 μm. D, COS-7 cells were transfected with Plk1-HA and/or BRCA2-FLAG expression plasmids. The expression of transgene-derived proteins in input cell lysates and anti-FLAG immunoprecipitates (IP) analyzed using anti-HA or anti-FLAG antibodies are shown. E, N-terminal biotin-fused pThr77 peptides (QLASpTPIIEK) were linked to streptavidin beads and tested for the ability to precipitate Plk1 from S-phase or mitotic HeLaS3 cell lysates. The phosphorylated Thr77 residue and the Ser76 residue crucial for PBD binding are indicated with asterisks (*). Immunoblot analyses show Plk1 coprecipitated with the peptides.

112–685 a.a.] did not affect the localization of the endogenous BRCA2 (Supplementary Fig. S1K). To evaluate further the effect of BRCA2 phosphorylation at Ser193 upon its localization, we expressed either wild-type recombinant BRCA2-FLAG or mutants thereof containing substitutions of Ser193 by alanine (S193A) or glutamate (S193E). Although both the wild-type and the S193E mutant localized to the Flemming body, the S193A mutant did not (Supplementary Fig. S1L). This suggested that phosphorylation of BRCA2 at Ser193 by Plk1 is important for its localization to the Flemming body and that a cancer-associated mutation of BRCA2 might prevent phosphorylation leading to its mislocalization.

Interaction between BRCA2 and NM-IIC

BRCA2 deficiency causes abnormal accumulation of myosin-II, particularly during cytokinesis (6). However, the physiologic role of colocalization of BRCA2 and myosin-II to the midbody remains unclear. We first examined the subcellular

localization of each NM-II isoform in A549 cells during cytokinesis (Fig. 2A). NM-IIC localized to the Flemming body and formed a ring-like structure. In contrast, NM-IIA was diffusively distributed throughout the midbody, and NM-IIB was localized within an area from the basal portion of the midbody to the cytosol of the dividing cell. Similar results were obtained in human mammary epithelial cells (Supplementary Fig. S2A). Further analysis of 3D images by confocal laser microscopy clearly demonstrated that NM-IIC formed part of a unique ring-like structure within the central portion of the midbody (Supplementary Fig. S2B). This structure, which we term the IIC-ring, was approximately 1.5 μm in diameter and colocalized with the Flemming body. To characterize this BRCA2-NM-IIC colocalization further, we analyzed its sedimentation in a glycerol gradient following isolation from A549 cells (Fig. 2B and Supplementary Fig. S2C). The pS193-BRCA2, NM-IIC, Plk1, and MgcRacGAP (a marker of the midbody) were all detected within fraction 2, whereas NM-IIB was detected in fraction 3, and NM-IIA was present in several fractions (1–8). BRCA1 was not found at the midbody and nonphosphorylated BRCA2 was mainly detected within fraction 4 (Supplementary Fig. S2C). These results suggested that pS^{193}-BRCA2 might interact with NM-IIC and MgcRacGAP in addition to Plk1 at the Flemming body. BRCA2-FLAG and NMHC-IIC-HA were coexpressed in COS-7 cells and cell lysates were subjected to immunoprecipitation with anti-FLAG and anti-HA antibodies. NMHC-IIC-HA and BRCA2-FLAG were both detected in the anti-FLAG or anti-HA immunoprecipitates (Fig. 2C). Endogenous BRCA2 immunoprecipitates from A549 cells were shown to contain NMHC-IIC (Fig. 2D and Supplementary Fig. S2D). This result was validated by two different antibodies that recognized distinct epitopes of NMHC-IIC (Fig. 2D) and by mass spectrometric analysis (Supplementary Fig. S2D). Immunoprecipitations from A549 cell lysates using anti-BRCA2 antibody revealed the presence of several protein components in the BRCA2 complex. Polypeptides with apparent molecular weights of 485, 385, 220, 156, and 122 kDa coimmunoprecipitated with BRCA2 (Supplementary Fig. S2D). As a control, we also performed immunoprecipitation with normal rabbit immunoglobulin G (IgG). These polypeptides were digested in gel by trypsin, and the resulting peptides were sequenced by nanoelectrospray mass spectrometry. Besides NMHC-IIC and BRCA2, polycystin-1 (PKD1), ATM, Rho-associated protein kinase 1 (ROCK1), and PARP1 were coimmunoprecipitated by the anti-BRCA2 antibody (Supplementary Fig. S2D). In contrast, NM-IIB did not bind to BRCA2 in this assay. Endogenous MgcRacGAP did not also bind to BRCA2 or NM-IIC (Supplementary Fig. S2E and S2F) despite its colocalization to the Flemming body.

Abnormal cytokinesis following NMHC-IIC gene silencing

To investigate the role of the IIC-ring during cell division, we observed the formation of the midbody 24 hours following siRNA knockdown of NMHC-IIC (Fig. 3A). Midbody structures in most siRNA-control cells were bundled and formed the Flemming bodies (Fig. 3B, top). In contrast, a majority of the cells treated with the NMHC-IIC siRNA retained midbody fibers, that is, exhibited a nonbundled midbody and lacked the Flemming body (Fig. 3B, bottom). The fraction of abnormal cells (exhibiting a non-Flemming body) was 25.2 ± 3.6% in the control cells and 63.5 ± 3.6% in the knockdown cells (Fig. 3B). At 48 hours, we observed a 1.5- to 3-fold increase in the number of multinuclear cells in response to NMHC-IIC knockdown (IIC-1: 57.2 ± 4.5%; IIC-2: 36.3 ± 1.7%) compared with control cells (scrambled control: 21.0 ± 2.0%; no-transfection: 15.8 ± 4.5%; Supplementary Fig. S3A). The NMHC-IIC knockdown cells were observed using time-lapse differential interference microscopy for 48 hours after siRNA treatment. A subpopulation of the cells had fused back together, resulting in binucleated cells. Another subpopulation exhibited catastrophic cell death before cytokinesis completion (Supplementary Fig. S3B). This process is shown in greater detail by time-lapse images (Supplementary Fig. S3C).

Inhibition of IIC-ring formation by BRCA2 gene silencing

To see an effect of BRCA2 suppression on the IIC-ring formation, we silenced BRCA2 using siRNA in A549 cells, which express all three isoforms: IIA, IIB, and IIC (Fig. 3C). As observed in the NMHC-IIC knockdown cells, midbody fibers were present in the absence of the Flemming body in the BRCA2 knockdown cells. Consequently, NM-IIC was distributed throughout the nonbundled midbody, and the IIC-ring was not apparent (Fig. 3D and Supplementary Fig. S3D). A greater number of abnormal midbodies were observed in siRNA-BRCA2–treated cells (siRNA-BRCA2-1: 48.3 ± 2.9%; siRNA-BRCA2-3: 43.3 ± 7.6%) than in control cells (siRNA-luciferase: 21.7 ± 2.9%; no-transfection: 10.0 ± 5.0%; $P = 0.017$; Fig. 3D, right). Similar results were obtained in U2OS cells treated with siRNA targeting BRCA2 (Supplementary Fig. S3E and S3F). To determine whether BRCA2 is important for the formation of midbody structures and progression of cytokinesis, we analyzed the BRCA2 knockdown HeLa S3 cells undergoing cytokinesis. BRCA2 protein levels were reduced after siRNA application (Fig. 3E), and the length of the midbody during cytokinesis in HeLa S3 cells was greater in *BRCA2* siRNA-treated cells than in control cells (siRNA-luciferase or no-transfection; $P < 0.01$; Fig. 3F). Next, we examined whether the length of the midbody in siRNA-BRCA2 cells could be restored by expressing BRCA2-FLAG or BRCA2 (S193A)-FLAG (Supplementary Fig. S3G and S3H). Although BRCA2-FLAG localization was observed in the midbody, this was not seen in BRCA2 (S193A)-FLAG cells (Supplementary Fig. S3H, left). When BRCA2-FLAG was expressed, the length of the midbody was similar to that in the control (siRNA-luciferase). In contrast, the length of the midbody did not recover in cells expressing BRCA2 (S193A)-FLAG ($P = 0.041$; Supplementary Fig. S3H, right), suggesting that it could not localize to the midbody. We did not observe an abnormal midbody or IIC-ring in cells treated with the *BRCA1*-siRNA (Supplementary Fig. S4A–S4E). These results suggest that the formation of the midbody and IIC-ring may require the presence of BRCA2, but not BRCA1, at the Flemming body and that the binding of BRCA2 to NM-IIC is necessary for this process.

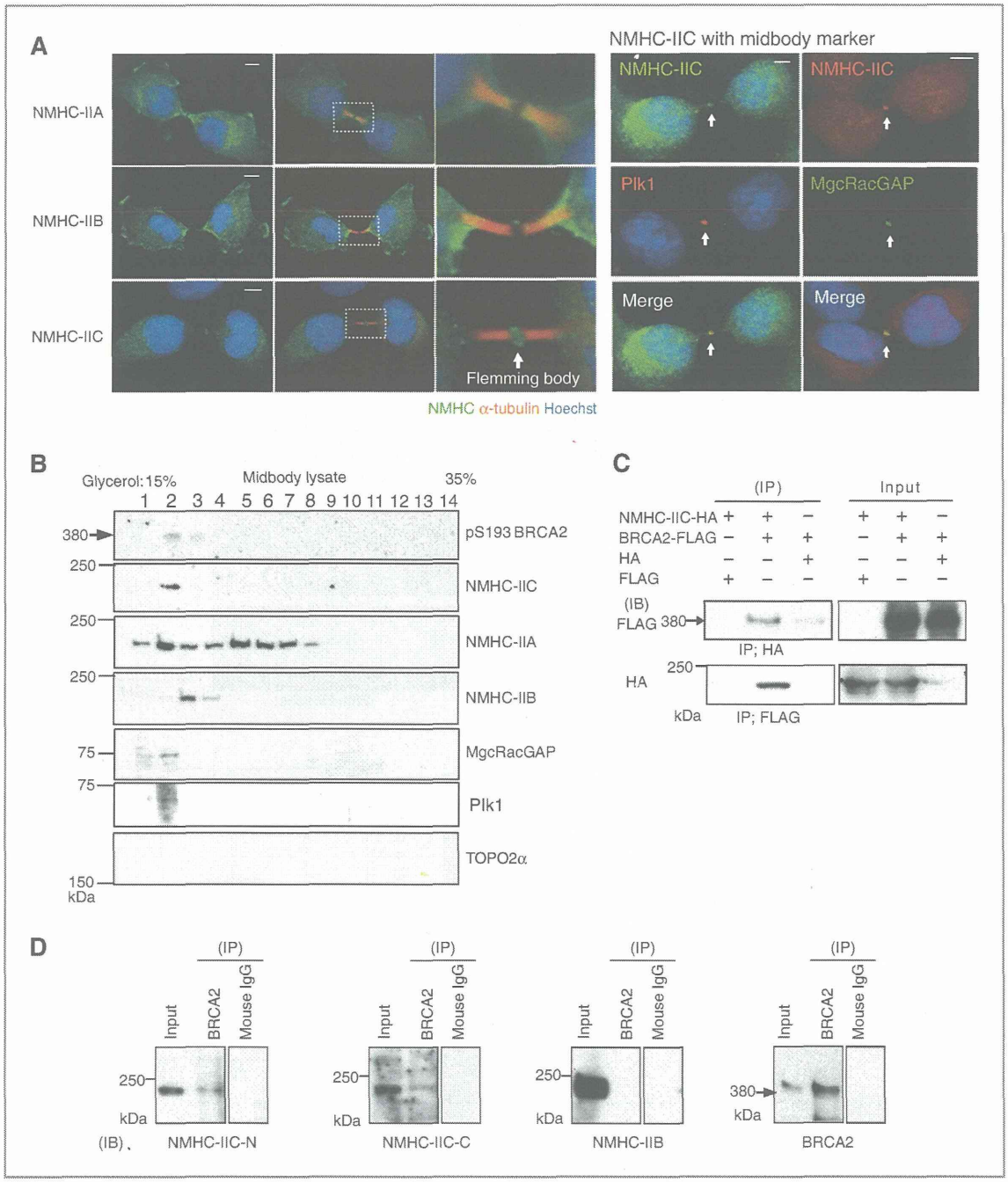

Figure 2. NM-IIC localizes to the Flemming body and interacts with BRCA2. A, immunofluorescence in A549 cells stained for NMHC-IIA (green), IIB (green), IIC (green), and α-tubulin (red). Cells in cytokinesis were also costained for NMHC-IIC (green or red) and Plk1 (red) or MgcRacGAP (green). Scale bar, 5 μm. B, midbody lysates from A549 cells were separated by 15% to 35% (w/v) glycerol gradient centrifugation and subjected to Western blot analysis using the indicated antibodies. The nuclear marker, TOPO2-α, was not detected in any fractions. C, COS-7 cells were transfected with the indicated expression plasmids. The expression of transgene-derived proteins in cell lysates, anti-FLAG immunoprecipitates (IP), and anti-HA immunoprecipitates (IP) revealed by anti-FLAG or anti-HA antibodies are shown. D, A549 cell lysates were immunoprecipitated (IP) with either mouse IgG or anti-BRCA2 antibodies and immunoprecipitates were analyzed using polyclonal antibodies to different epitopes of NMHC-IIC at the N-(NMHC-IIC-N) and C-(NMHC-IIC-C) terminals. Samples were also analyzed by anti-BRCA2 and anti–NMHC-IIB antibodies.

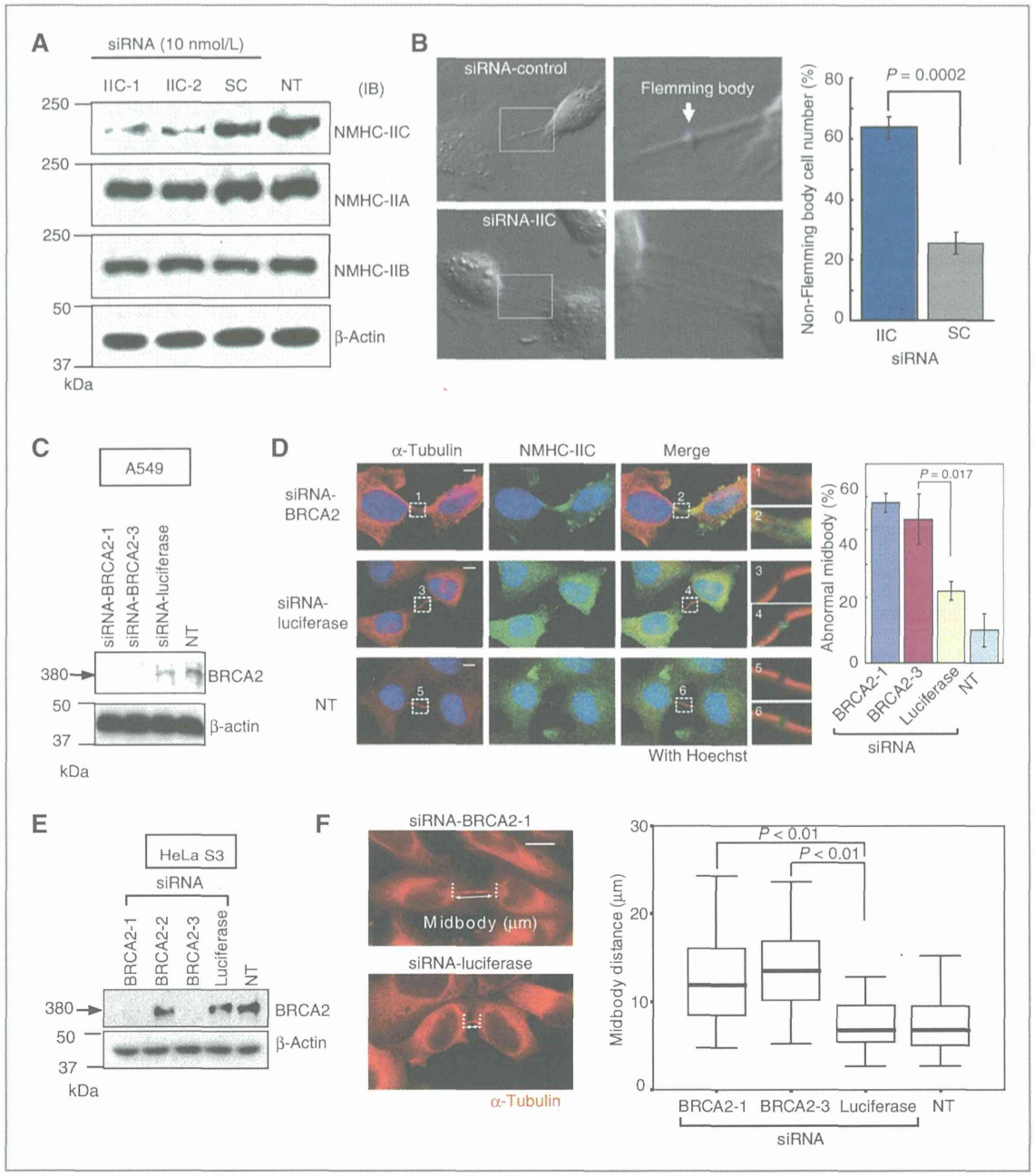

Figure 3. NM-IIC and BRCA2 are essential for midbody formation in A549 cells. A, lysates from A549 cells treated with NMHC-IIC siRNA (IIC-1 or IIC-2) or siRNA-scramble (SC), or no-transfection (NT) were subjected to Western blot analysis using anti–NMHC-IIC, IIA, IIB, and anti–β-actin antibodies. B, NMHC-IIC siRNA (bottom) and siRNA-scramble (top)-treated cells were observed by phase contrast microscopy. The bar graph shows the mean fractions (± SD) of cells showing abnormal midbodies based on three independent experiments ($n = 100$). C, lysates from A549 cells treated with two different BRCA2 siRNAs or siRNA-luciferase or no-transfection (NT) were subjected to Western blot analysis with anti-BRCA2 and anti–β-actin antibodies. D, A549 cells treated with BRCA2 siRNA or siRNA-luciferase or siRNA-no-transfection were fixed and stained with anti–α-tubulin (red) and anti–NMHC-IIC antibodies (green) in cytokinesis. Scale bar, 5 μm. Magnifications of dotted square areas are shown on the right (1–6). The bar graph shows the mean frequencies (±SD) of abnormal midbodies in cells based on three independent experiments ($n = 20$). E, lysates from HeLa S3 cells treated with three different BRCA2 siRNAs or siRNA-luciferase or no-transfection (NT) were subjected to Western blot analysis with anti-BRCA2 and anti–β-actin antibodies. F, HeLa S3 cells treated with BRCA2 siRNA (BRCA2-1) were fixed and stained for α-tubulin (red). White lines, the measured midbody length by Adobe Photoshop CS3 software. Scale bar, 5 μm. The midbody lengths are shown in box plot that signifies the top and bottom quartiles. The medians are represented by bold lines. Nonparametric data were statistically analyzed by the Mann–Whitney *U* test ($n = 81$ in each case, $P < 0.01$).