Molecular mechanisms underlying oncogenic *RET* fusion in lung adenocarcinoma

Tatsuji Mizukami, MD,† Kouya Shiraishi, PhD,* Yoko Shimada, M.F.Sc.,* Hideaki Ogiwara, PhD,* Koji Tsuta, MD, PhD,‡ Hitoshi Ichikawa, PhD,§ Hiromi Sakamoto, PhD,§ Mamoru Kato, PhD,¶ Tatsuhiro Shibata, MD, PhD,¶ Takashi Nakano, MD, PhD,† Takashi Kohno, PhD,**

**Division of Genome Biology, §Division of Genetics, and ¶Division of Cancer Genomics, National Cancer Center Research Institute, Tokyo 104-0045, Japan.*
†*Department of Radiation Oncology, Gunma University Graduate School of Medicine, Gunma 371-8511, Japan.*
‡*Division of Pathology and Clinical Laboratories, National Cancer Center Hospital, Tokyo 104-0045, Japan.*

Correspondence to: Prof. Takashi Kohno. Division of Genome Biology, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. Tel: +81-3-3542-2511, Fax: +81-3-3542-0807, Email: tkkohno@ncc.go.jp.

Support: This study was supported in part by Grants-in-Aid for Scientific Research on Innovative Areas (22131006), from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, for the Third-term Comprehensive 10 year Strategy for Cancer Control, from the Ministry of Health, Labor, and Welfare, and for the Program for Promotion of Fundamental Studies in Health Sciences, from the National Institute of Biomedical Innovation (NIBIO), and by Management Expenses Grants from the Government to the National Cancer Center (NCC).

Background: Oncogenic *RET* fusion, caused by an inversion in chromosome 10, was recently identified as a driver mutation for the development of lung adenocarcinoma (LADC). However, the molecular mechanism(s) underlying the rearrangement of the *RET* locus during lung carcinogenesis are unknown.

Patients and methods: Genomic segments containing breakpoint junctions for *RET* fusions were cloned and analyzed by genomic PCR and genome capture sequencing using a next-generation sequencer to identify the mechanisms involved in DNA strand breaks and illegitimate joining of DNA ends. Of the 18 cases studied, 16 were identified by screening 671 LADC cases and two were previously published.

Results: Almost all (17/18, 94%) of the breakpoints in *RET* were located within a 2.0 kb region spanning exon 11 to intron 11, and no breakpoint occurred within 4 bp of any other. This suggested that, as in papillary thyroid carcinoma (PTC), DNA strand breaks formed at non-specific sites within this region trigger *RET* fusion. Just over half of the *RET* fusions in LADC (10/18, 56%) were caused by simple reciprocal inversion, and two DNA-repair mechanisms, namely, non-homologous end joining (NHEJ) and break-induced replication (BIR), were deduced to have contributed to the illegitimate joining of the DNA ends.

Conclusions: Oncogenic *RET* fusion in LADC occurs through multiple pathways and involves the illegitimate repair of DNA strand breaks via mechanisms different from those identified in PTC, where *RET* fusion also functions as a driver mutation.

Keywords: lung adenocarcinoma; molecular target therapy; personalized medicine; *RET*; gene fusion; DNA strand break

Oncogenic fusion of *RET* (rearranged during transfection) tyrosine kinase gene partnered with *KIF5B* (kinesin family member 5B) and *CCDC6 (coiled-coil domain containing 6)* was identified as a novel druggable driver mutation in a small subset (1–2%) of patients with lung adenocarcinoma (LADC)[1-4]. Vandetanib (ZD6474) and cabozantininb (XL184), two FDA (US Food and Drug Administration)-approved inhibitors of the *RET* tyrosine kinase showed therapeutic responses in a few patients with *RET* fusion-positive LADC[5, 6]. Several clinical trials are currently underway to examine the therapeutic effects of RET tyrosine kinase inhibitors, including these two drugs [7, 8]. *RET* fusions are generated by pericentric (includes the centromere, with a breakpoint in each arm) and paracentric (not including the centromere, with both breaks in the same arm) inversions of chromosome 10 (**Figure 1A**). Since the majority of *RET* fusion-positive patients are never-smokers [3, 9, 10], cigarette smoking does not cause a predisposition. Therefore, the mechanism(s) responsible for the rearrangement of the *RET* locus are unknown. Elucidation of such a mechanism(s) may help to identify risk factors that can be modified or other preventive methods that can reduce the incidence of LADC, however, no such mechanism has been identified [8].

Analyzing the breakpoints and structural aberrations in cancer cell genomes is a powerful method of identifying the underlying molecular mechanism(s) responsible, since the breakpoints retain "traces" of the DNA strand breaks and the illegitimate joining of DNA ends [11-13]. In fact, several studies have characterized the structure of the breakpoints responsible for the *ELE1 (also known as RFG, NCOA4 and ARA70)-RET* oncogenic fusion in cases of papillary thyroid cancer (PTC), including post-Chernobyl irradiation-induced cases, to elucidate the mechanism underlying chromosome 10 inversion generating this fusion (**Figure 1A**) [14-17].

Here, we examined the molecular processes underlying chromosome inversions that

generate oncogenic *RET* fusions in LADC by cloning genomic segments containing breakpoint junctions and by comparing their structures with those identified in PTC. The results will increase our understanding of how *RET* fusions are generated, and will also have implications for diagnosis of *RET* fusion positive LADCs.

PATIENTS AND METHODS

Patient samples

Fourteen frozen tissues (13 surgical specimens and a pleural effusion) and two methanol-fixed paraffin-embedded tissues from surgical specimens were obtained from the National Cancer Center (NCC) Biobank. These samples were from patients with LADC who received therapy at the NCC Hospital (Tokyo, Japan) between 1997 and 2012. All frozen samples were confirmed to be positive for *KIF5B-RET* fusion by RT-PCR analysis, according to a previously described method [3]. *CCDC6-RET* fusion was detected by fusion fluorescence *in situ* hybridization (FISH) analysis of paraffin-embedded tissues using *RET*- and *CCDC6*-specific probes (Chromosome Science Labo, Inc; Sapporo, Japan). This study was approved by the institutional review board of the NCC.

Cloning and sequencing of DNAs containing breakpoint junctions

Genomic DNAs were extracted from cancer and noncancerous tissues using the QIAamp DNA Mini Kit or the QIAamp DNA Micro Kit (QIAGEN, Hilden, Germany). Genomic DNA fragments containing breakpoint junctions were amplified by genomic PCR using primers that hybridized within the *KIF5B* and *RET* loci. PCR products specifically amplified in samples of

interest were subjected to direct Sanger sequencing. The primers used are listed in **Supplementary Table 1**.

Genome-capture deep sequencing using a next-generation speed sequencer

Nucleotide sequences of *CCDC6-RET* fusion breakpoints were examined by targeted genome-capture and massively parallel sequencing using an Ion PGM sequencing system and the Ion TargetSeq Custom Enrichment Kit (Life Technologies). One microgram of genomic DNA was subjected to enrichment using the probes listed in **Supplementary Table 2**. The mean depth of sequencing was approximately 1,000.

Analysis of sequence reads obtained by a second generation sequencer

Sequence reads were analyzed using a program developed by the authors. Briefly, reads were mapped to sequences of the *RET* and *CCDC6* genes using the BWA-SW software[18] to detect reads that mapped to both the *RET* and *CCDC6* genes. Breakpoints were extracted from the local alignment results of BWA-SW. The detailed procedure is described in Supplementary Notes. Structures of breakpoint junctions were verified by Sanger sequencing of genomic PCR products.

LOH analysis

Genomic DNAs obtained from cancerous and non-cancerous tissues were subjected to SNP genotyping using the Illumina HumanOmni1 2.5M Chip (Illumina, San Diego, CA, USA). Based on the B-allele frequencies obtained using the Illumina GenomeStudio software, LOH regions in *RET* and surrounding regions were deduced. Representative SNP loci were subjected to analysis of allelic imbalance using the Sequenom MassARRAY system (Sequenom, San Diego, CA, USA).

Analysis of nucleotide sequences

Nucleotide sequence analysis, including search for sequence homology, was performed using the Genetyx-SV/RC Ver 8.0.1. software (Genetyx, Tokyo, Japan). Information about the distribution of repetitive elements, GC contents, conservation, DNA methylation, DNase sensitivity, and histone modification within the *RET* gene was obtained using the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

RESULTS

KIF5B-RET fusion variations in lung adenocarcinoma

In our previous study, six (1.9%) of 319 LADC cases carried *KIF5B-RET* fusions [3]. In this study, we examined *KIF5B-RET* fusion by RT-PCR in a further 352 LADC cases, and found eight additional *KIF5B-RET* fusion-positive cases. In total, 14 (2.1%) of 671 cases were positive for *KIF5B-RET* fusions (cases 1–4 and 7–16 in **Table 1** and **Supplementary Table 3**), and this frequency was consistent with those reported for other cohorts [9, 10, 19].

Among those 14 cases, ten (71%) contained a fusion of *KIF5B* exon 15 to *RET* exon 12 (K15;R12), whereas the remaining four each contained other variants. Thus, K15;R12 is the most frequent variant (**Figure 1B**). The prevalence of the K15;R12 variant (45/60, 75%) was verified in a total of 60 cases, including 46 cases from eight other cohorts published to date [1-4, 9, 10, 19, 20] (**Figure 1B, Supplementary Table 4**). This preference was similar among cohorts from Japan, other Asian countries, and the USA ($P > 0.05$ by Fisher's exact test).

Distribution of breakpoints in the *RET* and *KIF5B* genes

To explore the molecular processes underlying *RET* fusion in LADC, we examined the

location (clustering) of the breakpoints and the structure of the breakpoint junctions; information about the former enabled us to deduce the genomic or chromosomal features that make DNA susceptible to strand breaks, whereas information about the latter enabled us to deduce the mechanism underlying the illegitimate joining of DNA ends by DNA repair pathways.

The locations of the 28 breakpoints in the 14 *KIF5B-RET* fusion-positive cases mentioned above were identified by Sanger sequencing analysis of genomic PCR products and mapped (yellow arrowheads in **Figure 2A, 2B**). The breakpoints in a single Korean case from another study were also identified and mapped (orange arrowheads in **Figure 2A**; case 17 in **Table 1**). Consistent with the predominance of K15;R12 variants, most of the breakpoints were mapped to intron 11 of *RET* and intron 15 of *KIF5B* (**Figure 2,** detailed information in **Supplementary Table 5**).

None of the *RET* and *KIF5B* breakpoints mapped at the same position, and no breakpoint was within 6 bp of another. To further investigate the breakpoint clustering, we mapped breakpoints in three cases of *CCDC6-RET* fusion, a minor fusion variant (cases 5, 6, and 18 in **Table 1** and **Supplementary Table 3**). Two of these cases were primary tumors, diagnosed by break-apart and fusion *FISH*, and their breakpoints were determined by genome-capture deep sequencing of genomic DNAs using a second generation sequencer. The remaining case was a LADC cell line from a Japanese patient, for which the breakpoints had previously been determined by the same method [21]. Two breakpoints and one breakpoint in the *RET* gene were mapped to intron 11 and exon 11, respectively (green arrowheads in **Figure 2**), and no breakpoint was located within 5 bp of another. In total, a 2.0 kb region spanning exon 11 to intron 11 of *RET* and a 5.6 kb region spanning intron 15 of *KIF5B* (10/15, 75%) contained the majority of breakpoints (17/18 [94%] and 10/15 [75%], respectively), and these breakpoints were at least 5 bp from each other. Breakpoints within exon 11 to intron 11

of *RET* and intron 15 of *KIF5B* were not distributed in an evidently biased manner, nor did they exhibit any particular nucleotide sequence or composition (**Supplementary Table 5**). Therefore, DNA strand breaks triggering oncogenic *RET* fusions in LADC occur preferentially in a few defined regions, but at non-specific sites within those regions.

Reciprocal and non-reciprocal inversions causing *RET* fusions

To explore the modes of DNA end joining that give rise to *RET* fusion, we investigated the structures of *RET* fusion breakpoint junctions. To address whether chromosome inversion events were reciprocal, we cloned genomic segments containing reciprocal breakpoint junctions (i.e., *RET-KIF5B* and *RET-CCDC6*) from 17 Japanese cases (**Table 1**). Ten of the seventeen cases, consisting of eight *KIF5B-RET* and two *CCDC6-RET* cases, allowed amplification of reciprocal genomic segments using PCR primers set 1 kb away from the identified *KIF5B-RET* or *CCDC6-RET* breakpoints. This indicated that these fusions were the results of simple reciprocal inversions (cases1-10 in **Table 1**, **Figure 2C**). On the other hand, the remaining seven cases did not allow amplification of genomic segments encompassing the reciprocal breakpoint junctions (cases 11-16 and 18 in **Table 1**). Three of these seven cases, for which corresponding non-cancerous DNA was available, were subjected to loss of heterozygosity (LOH) analysis at the *RET* locus. LOH was detected at a region proximal (N-terminal) to the breakpoints in all three cases (cases 11, 15, and 16 in **Table 1**, **Figure 1A**), indicating non-reciprocal inversion associated with deletion of a copy of the region proximal to the breakpoints. In addition, the inversion in the aforementioned Korean case (case 17) is also non-reciprocal [4]. These data suggested that only a fraction of *RET* fusions (10/18, 56%) are caused by simple reciprocal inversions.

Modes of DNA end joining that give rise to reciprocal inversions

Two major types of DNA repair pathways cause structural variations[11, 12]. The first type is NHEJ of DSBs, which requires very short (a few bp) or no homology, and often inserts a few nucleotides at breakpoint junctions [8, 22, 23]. NHEJ has canonical and non-canonical forms; in the latter, called alternative end joining (alt-EJ), DNA ends are joined using microhomology of a few nucleotides at breakpoints [24]. The second type includes repair pathways that use long (>10 bp) homology at DNA ends, such as break-induced replication (BIR) and non-allelic homologous recombination [12, 25].

Sequence analysis of breakpoint-containing genomic segments in ten reciprocal cases revealed that deletions frequently (8/10, 80%) occur in *RET* and/or its partner locus (i.e., *KIF5B* or *CCDC6*) upon DNA end joining (**Table 1**). This analysis also enabled us to deduce that both types of repair pathways described above are involved in these joining events. In six of the cases (cases 1–6 in **Table 1**)**,** four DNA ends were joined, and in two cases, insertions were observed (representative cases in **Supplementary Figure 1**). The lack of significant homology between the sequences of the *RET* and *KIF5B/CCDC6* breakpoints led us to deduce that DNA end joining was mediated by NHEJ in these six cases: two DSBs formed, one each in *RET* and its partner locus, and the four resultant DNA ends were illegitimately joined by canonical or non-canonical NHEJ (**Figure 3A**).

The remaining four cases (cases 7–10 in **Table 1**) had a distinctive feature. DNA segments of 33–490 bp from either the *RET* or *KIF5B* locus were retained at both the *KIF5B-RET* and *RET-KIF5B* breakpoints, resulting in duplication of these segments. Notably, two regions encompassing the breakpoint in a locus exhibited sequence homology to the duplicated segment of the other locus (representative cases in **Supplementary Figure 2**). This feature led us to deduce that these joining events were mediated by BIR, using both DNA ends generated by DNA single-strand breaks at the *RET* or fusion-partner locus (**Figure 3B**). Specifically, two DNA broken ends generated at the *RET* (or partner locus) annealed with the

DSB sites of the fusion-partner (or *RET*) locus through sequence homology, and were then subjected to ectopic DNA replication. This process left the same DNA segment at both breakpoint junctions, resulting in duplication of the segment.

Speculated mode of DNA end joining giving rise to non-reciprocal inversion

Our study also speculated about the modes of joining involved in the eight remaining cases, which were not likely to have been subjected to simple reciprocal inversion, and are therefore defined here as non-reciprocal (cases 11–18 in **Table 1**). Due to the lack of sequence information from breakpoints in reciprocal counterparts, deletions could not be assessed. The lack of significant homology between the *RET* and *KIF5B/CCDC6* breakpoints suggested the involvement of NHEJ. Consistent with this idea, insertion of a few nucleotides, a common trace of NHEJ, was observed in three cases (cases 11, 15, and 17). A single case (case 16) had an insertion of 349 nucleotides, corresponding to the inverted segment of *RET* exon 7 to intron 7, suggesting the occurrence of an unspecified complex rearrangement mediated by a process other than NHEJ, such as fork stalling and template switching (FoSTeS) (Lee et al., 2007). These results suggest that the predominant molecular process is illegitimate NHEJ repair, in which two DSBs are formed both in the *RET* and partner loci, and one end of the partner locus (the N-terminal part of *KIF5B* or *CCDC6*) and one end of the *RET* locus (the C-terminal part) are joined by NHEJ. However, the remaining two DNA ends were not joined in a simple manner. DNA segments within the DNA ends were either lost or joined with DNA ends other than those at the *RET*, *KIF5B*, and *CCDC6* loci, consistent with the observations of LOH at regions proximal to breakpoints in *RET* (**Table 1**). In fact, in case 17, the 3' part of the *KIF5B* gene was fused to the *KIAA1462* gene, 2.0 Mb away from *KIF5B* [4].

DISCUSSION

In this study, we investigated the molecular mechanisms underlying oncogenic *RET* fusion in LADCs. Distribution of breakpoints made us consider a 2.0 kb segment spanning *RET* exon 11 to intron 11 (and also a 5.6 kb segment spanning *KIF5B* intron 15) as a breakpoint cluster region(s). The breakpoints in these regions were dispersed at intervals larger than 4 bp. The inferred breakpoints do not necessarily indicate the sites of actual DNA breaks because resection of nucleotides from DNA ends sometimes occurs during the DNA repair [23]. In fact, we observed nucleotide deletions in eight of ten LADC cases with reciprocal *KIF5B*/*CCDC6*-*RET* inversions. However, when the locations of putative breakpoints prior to DNA end resection were included, the breakpoint distribution remained scattered. These data strongly suggested that the majority of DNA breaks triggering *RET* fusions occur at non-specific sites in defined regions of a few kb in size. Furthermore, this seems to hold true irrespective of etiology and tumor type: the distribution of breakpoints was not significantly different between ever- and never-smokers, and *RET* exon 11 to intron 11 was also defined as a breakpoint cluster region for *RET* fusions in PTCs, as previously reported[14-17]. The cases shown in **Figure 2** (gray and black arrowheads) include PTCs induced by post-Chernobyl irradiation, in which DNA breaks were presumably caused exclusively by irradiation; the random breakpoint distributions in these PTCs were similar to those of the LADCs we analyzed.

We investigated the DNA end joining pathways that gave rise to *RET* fusions by analyzing the structures of breakpoint junctions. NHEJ was found to be one of the major pathways of DNA end joining. We and others also showed that NHEJ is also prominently involved in interstitial deletions that inactivate tumor-suppressor genes, such as *CDKN2A/p16* and *STK11/LKB1,* in lung cancer [13, 26, 27]. Thus, NHEJ contributes to the occurrence of driver mutations in both tumor-suppressor genes and oncogenes during lung carcinogenesis. Our data also reveal a possible contribution of BIR in DNA end joining to generate reciprocal

inversions. We deduced that BIR occurred from DNA ends, probably generated by DNA single-strand breaks, in the *RET* or partner locus, beginning with annealing with the other locus through nucleotide homologies of tens to hundreds of bp. This process resulted in duplication of breakpoint-flanking DNA segments of tens to hundreds of bp. BIR has recently been implicated in oncogenic *RAF* fusions in pediatric brain tumors [28]. In those cases, the sequence homology used for annealing of DNA ends was on the order of a few bp. Thus, BIR might generate oncogenic fusions frequently, although the detailed process may differ according to tumor type.

Irrespective of the similarities in breakpoint distribution, several processes involved in *RET* fusions differed between LADC and PTC (**Figure 4**). Reciprocal inversion was unlikely to have occurred by BIR in PTC because none of the PTC cases exhibited the duplication of DNA segments that were observed in LADC; therefore, the joining of DNA ends in PTC was likely to have been mediated exclusively by NHEJ [17]. This is plausible because *RET* fusions preferentially occur in PTCs in patients suffering from high-dose radiation exposure, suggesting that DSBs generated at the *RET* or partner loci triggered the chromosome rearrangements that generated *RET* fusions [29]. Repetitive NHEJ repair of abundant DSBs, which occurs in the context of irradiation, may increase the likelihood of illegitimate repair generating *RET* fusion. On the other hand, in LADC, both DSBs and SSBs formed by multiple causes might trigger rearrangements by multiple DNA repair pathways. The high frequency of non-reciprocal inversion also distinguishes LADC from PTC, for previous study revealed that *RET* fusions result from reciprocal inversion in most cases (43/47, 91%) [14, 15]. Frequent non-reciprocal inversion is consistent with the observation that *KIF5B-RET* fusion-positive tumors contain deletions of the 5' part of *RET,* as revealed by FISH staining patterns [1]. The present study provides a molecular basis for such a distinct FISH finding, and will help to define the criteria used to diagnose *RET*-fusion-positive LADC.

Interestingly, FISH analysis also revealed that another driver mutation, *EML4-ALK* fusion, in LADC, caused by a paracentric inversion of chromosome 2, also involves deletion of the 5' region of the *ALK* oncogene locus [30, 31]. Although the structures of breakpoint junctions of *ALK* fusions have not been characterized to the best of our knowledge, these results indicate that a significant fraction of chromosome inversions that cause oncogenic fusions in lung cancer are likely to be non-reciprocal.

Finally, a few issues remain to be elucidated regarding the molecular processes generating oncogenic *RET* fusions. Firstly, although this and previous PTC studies imply that the 2.0 kb region spanning the *RET* exon 11 to intron 11 region is susceptible to DNA strand breaks, the underlying mechanisms remain unknown. For, this region does not exhibit distinctive features known to make DNA susceptible to breaks (**Supplementary Figure 3,** details in Supplementary Notes). Secondly, the etiological factors that cause DNA strand breaks, and the factors that determine reciprocal or non-reciprocal inversion and selection of DNA repair pathways, also remain unknown. The mode of joining and breakpoint distribution was irrespective of smoking history, and, therefore, DNA damage due to smoking is unlikely to be an important factor. The fact that *RET* fusions are more frequent in LADC of never-smokers than in ever-smokers indicates that undefined etiological factors play major roles in the occurrence of *RET* fusions.

ACKNOWLEDGMENTS

We thank Hiromi Nakamura, Isao Kurosaka, Sumiko Ohnami and Sachiyo Mitani of National Cancer Center Research Institute for data analysis and technical assistance. The NCC Biobank is supported by the NCC Research and Development Fund of Japan. SNP array analysis was performed by the genome core facility of the NCC.

TABLE 1. Structure of breakpoint junctions of *RET* fusions in lung adenocarcinoma

No.	Sample name	Fusion partner	Reciprocal/ Non-reciprocal	Deletion in the joining		DNA segment duplication by inversion		Nucleotide overlap at junction		Nucleotide insertion at junction		Mode of DNA end joining	LOH proximal to *RET*	Smoking
				RET	Partner	*RET*	Partner	Partner -*RET*	*RET*- Partner	Partner -*RET*	*RET*- Partner			
1	BR0020	*KIF5B*	Reciprocal	-	-	-	-	-	-	-	-	NHEJ	NT	No
2	L07K201T	*KIF5B*	Reciprocal	14-bp	19-bp	-	-	C	-	-	ATA	NHEJ	NT	Yes
3	349T	*KIF5B*	Reciprocal	1-bp	7-bp	-	-	-	-	A	A	NHEJ	NT	Yes
4	AD08-341T	*KIF5B*	Reciprocal	16-bp	26-bp	-	-	-	-	-	-	NHEJ	NT	No
5	RET-030	*CCDC6*	Reciprocal	52-bp	1021-bp	-	-	-	-	-	-	NHEJ	NT	No
6	RET-024	*CCDC6*	Reciprocal	14-bp	2-bp	-	-	-	-	-	-	NHEJ	NT	Yes
7	AD12-106T	*KIF5B*	Reciprocal	-	573-bp	490-bp	-	-	-	-	-	BIR	NT	Yes
8	BR0030	*KIF5B*	Reciprocal	-	-	-	211-bp	-	-	-	-	BIR	NT	No
9	442T	*KIF5B*	Reciprocal	269-bp	-	-	232-bp	-	-	-	-	BIR	NT	No
10	AD08-144T	*KIF5B*	Reciprocal	5-bp	-	-	33-bp	-	-	-	-	BIR	NT	No
11	BR1001	*KIF5B*	Non-reciprocal					-		AGT		NHEJ	+	No
12	AD09-369T	*KIF5B*	Non-reciprocal					CTC		-		NHEJ (Alt-EJ)	NT	No
13	BR1002	*KIF5B*	Non-reciprocal					A		-		NHEJ	NT	No
14	AD12-001T	*KIF5B*	Non-reciprocal					-		-		NHEJ	NT	Yes
15	BR1003	*KIF5B*	Non-reciprocal					-		CTTT		NHEJ	+	No
16	BR1004	*KIF5B*	Non-reciprocal					-		RET ex 7-int 7 (359-bp)		Complex rearrange	+	No
17	AK55[a]	*KIF5B*	Non-reciprocal					-		GT		NHEJ	NT	No
18	LC-2/ad[b]	*CCDC6*	Non-reciprocal					-		-		NHEJ	NT	Unknown

[a]Ju et al (2012).

[b]Suzuki et al (2013)

Blank: not applicable; NT: not tested

Figure legends

FIGURE 1. *RET* fusions. (**A**) *Upper:* location of the *RET* oncogene and its fusion-partner genes *KIF5B*, *CCDC6*, and *ELE1* on chromosome 10. The *KIF5B-RET* fusion is generated in LADC, whereas the *CCDC6-RET* fusion is generated in LADC and PTC. The *ELE1-RET* fusion is frequent in radiation-induced PTC. *Lower:* LOH analysis. Allelic imbalance at SNP sites proximal and distal to the breakpoints were examined by MassArray analysis in three LADC cases with putative non-reciprocal inversions. Cases 11, 15 and 16 exhibited allelic imbalance (23%, 41%, and 29%, as indicated by arrows) at SNP loci proximal to the breakpoints, consistent with the fact that these samples have 20–40% tumor content. (**B**) Fractions of *KIF5B-RET* fusion variants in lung adenocarcinomas. Fractions comprise the cohort from this study and eight published cohorts. Fractions in patients from Japan, other Asian countries (Korea and China), and the USA are shown below.

FIGURE 2. Breakpoint analysis. (**A**) Distribution of breakpoints in the *CCDC6*, *KIF5B* and *RET* genes.. Yellow arrowheads indicate the locations of breakpoints for *KIF5B-RET* fusions in Japanese cases (cases 1–4 and 7–16 in **Table 1**), whereas the orange arrowhead indicates the breakpoints in a single Korean case (case 17). Green arrowheads indicate the locations of breakpoints of *CCDC6-RET* fusions in three Japanese cases (cases 5, 6, and 18). Arrowheads for ever-smoker LADC cases are hatched. Gray and black arrowheads indicate breakpoints of *RET-ELE1* fusion in 38 radiation-induced post-Chernobyl PTCs and six sporadic PTCs, respectively[14-17]. (**B**) Electropherograms for Sanger sequencing of genomic fragments encompassing *KIF5B-RET* breakpoint junctions. PCR products were directly sequenced. Examples of three fusion patterns (joined without any nucleotide insertions or overlaps, joined

with a nucleotide insertion, and joined with three nucleotide overlap) are shown. Inserted and overlapping nucleotides at breakpoint junctions are indicated, respectively, by the blue and red boxes. (**C**) Electropherogram for Sanger sequencing of genomic fragments encompassing *CCDC6-RET and RET-CCDC6* breakpoint junctions.

FIGURE 3. Deduced processes of reciprocal inversion by NHEJ and BIR. (A) NHEJ. Four DNA ends generated by DSBs at *RET* and a partner locus were directly joined. Often, insertions of nucleotides, (NNN), at breakpoint junctions are observed. (B) BIR. Here, DNA single strand-breaks (SSBs) occur in the *KIF5B* locus and a DSB occurs in the *RET* locus. The two SSBs at the *KIF5B* locus trigger BIR by annealing at two homologous sites in the *RET* locus. BIR results in duplication of a *KIF5B* segment. As a result, the *RET* breakpoints in the *KIF5B-RET* and *RET-KIF5B* fusions are located at the same position (a DSB site), whereas the *KIF5B* breakpoints in these fusions are located at different positions (two SSB sites). ▽, breakpoints for partner-*RET* fusion; ▲, breakpoints for *RET*-partner fusion.

FIGURE 4. Molecular processes underlying *RET* gene fusions in LADC and PTC. Different processes are involved in *RET* fusion in different tumor types. Both reciprocal and non-reciprocal inversions occur in LADC. In LADC, BIR and NHEJ are responsible for DNA end joining in reciprocal inversion, whereas NHEJ is exclusively involved in non-reciprocal inversion. In PTC, reciprocal inversion by NHEJ is dominant.

REFERENCES

1. Takeuchi K, Soda M, Togashi Y, et al. RET, ROS1 and ALK fusions in lung cancer. *Nat Med* 2012;18:378-381.
2. Lipson D, Capelletti M, Yelensky R, et al. Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. *Nat Med* 2012;18:382-384.
3. Kohno T, Ichikawa H, Totoki Y, et al. KIF5B-RET fusions in lung adenocarcinoma. *Nat Med* 2012;18:375-377.
4. Ju YS, Lee WC, Shin JY, et al. A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. *Genome Res* 2012;22:436-445.
5. Gautschi O, Zander T, Keller FA, et al. A patient with lung adenocarcinoma and RET fusion treated with vandetanib. *J Thorac Oncol* 2013;8:e43-44.
6. Drilon A, Wang L, Hasanovic A, et al. Response to Cabozantinib in Patients with RET Fusion-Positive Lung Adenocarcinomas. *Cancer Discov* 2013;3:630-635.
7. Kohno T, Tsuta K, Tsuchihara K, et al. RET fusion gene: Translation to personalized lung cancer therapy. *Cancer Sci* 2013.
8. Shaw AT, Hsu PP, Awad MM, et al. Tyrosine kinase gene rearrangements in epithelial malignancies. *Nat Rev Cancer* 2013;13:772-787.
9. Wang R, Hu H, Pan Y, et al. RET fusions define a unique molecular and clinicopathologic subtype of non-small-cell lung cancer. *J Clin Oncol* 2012;30:4352-4359.
10. Suehara Y, Arcila M, Wang L, et al. Identification of KIF5B-RET and GOPC-ROS1 fusions in lung adenocarcinomas through a comprehensive mRNA-based screen for tyrosine kinase fusions. *Clin Cancer Res* 2012;18:6599-6608.
11. Yang L, Luquette LJ, Gehlenborg N, et al. Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell* 2013;153:919-929.
12. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. *Pathogenetics* 2008;1:4.
13. Kohno T, Yokota J. Molecular processes of chromosome 9p21 deletions causing inactivation of the p16 tumor suppressor gene in human cancer: deduction from structural analysis of breakpoints for deletions. *DNA Repair (Amst)* 2006;5:1273-1281.
14. Nikiforov YE, Koshoffer A, Nikiforova M, et al. Chromosomal breakpoint positions suggest a direct role for radiation in inducing illegitimate recombination between the ELE1 and RET genes in radiation-induced thyroid carcinomas. *Oncogene* 1999;18:6330-6334.
15. Bongarzone I, Butti MG, Fugazzola L, et al. Comparison of the breakpoint regions of ELE1 and RET genes involved in the generation of RET/PTC3 oncogene in sporadic and in radiation-associated papillary thyroid carcinomas. *Genomics* 1997;42:252-259.
16. Minoletti F, Butti MG, Coronelli S, et al. The two genes generating RET/PTC3 are localized in chromosomal band 10q11.2. *Genes Chromosomes Cancer* 1994;11:51-57.
17. Klugbauer S, Pfeiffer P, Gassenhuber H, et al. RET rearrangements in radiation-induced papillary thyroid carcinomas: high prevalence of topoisomerase I sites at breakpoints and microhomology-mediated end joining in ELE1 and RET chimeric genes. *Genomics* 2001;73:149-160.
18. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26:589-595.
19. Cai W, Su C, Li X, et al. KIF5B-RET fusions in Chinese patients with non-small cell lung cancer. *Cancer* 2013;119:1486-1494.
20. Yokota K, Sasaki H, Okuda K, et al. KIF5B/RET fusion gene in surgically-treated adenocarcinoma of the lung. *Oncol Rep* 2012;28:1187-1192.
21. Suzuki M, Makinoshima H, Matsumoto S, et al. Identification of a lung adenocarcinoma cell line with CCDC6-RET fusion gene and the effect of RET inhibitors in vitro and in vivo. *Cancer Sci* 2013.
22. Mahaney BL, Meek K, Lees-Miller SP. Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem J* 2009;417:639-650.
23. Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* 2010;79:181-211.

24. Bennardo N, Cheng A, Huang N, et al. Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. *PLoS Genet* 2008;4:e1000110.
25. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 2007;131:1235-1247.
26. Sasaki S, Kitagawa Y, Sekido Y, et al. Molecular processes of chromosome 9p21 deletions in human cancers. *Oncogene* 2003;22:3792-3798.
27. Matsumoto S, Iwakawa R, Takahashi K, et al. Prevalence and specificity of LKB1 genetic alterations in lung cancers. *Oncogene* 2007;26:5911-5918.
28. Lawson AR, Hindley GF, Forshew T, et al. RAF gene fusion breakpoints in pediatric brain tumors are characterized by significant enrichment of sequence microhomology. *Genome Res* 2011;21:505-514.
29. Hamatani K, Eguchi H, Ito R, et al. RET/PTC rearrangements preferentially occurred in papillary thyroid cancer among atomic bomb survivors exposed to high radiation dose. *Cancer Res* 2008;68:7176-7182.
30. Dai Z, Kelly JC, Meloni-Ehrig A, et al. Incidence and patterns of ALK FISH abnormalities seen in a large unselected series of lung carcinomas. *Mol Cytogenet* 2012;5:44.
31. Yoshida A, Tsuta K, Nitta H, et al. Bright-field dual-color chromogenic in situ hybridization for diagnosing echinoderm microtubule-associated protein-like 4-anaplastic lymphoma kinase-positive lung adenocarcinomas. *J Thorac Oncol* 2011;6:1677-1686.

FIGURE 1.

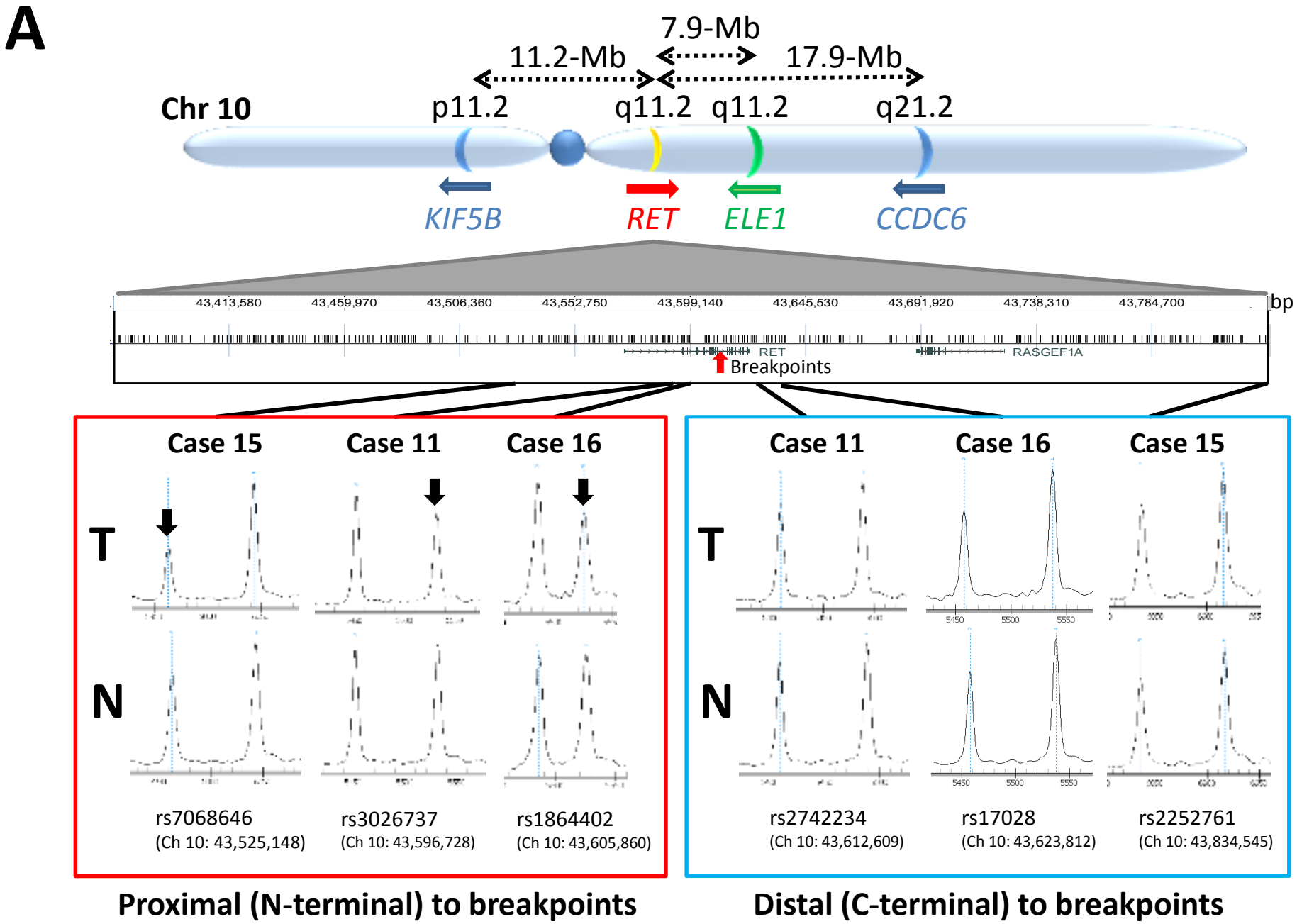

A
7.9-Mb
11.2-Mb
17.9-Mb
Chr 10
p11.2
q11.2
q11.2
q21.2
KIF5B
RET
ELE1
CCDC6
43,413,580
43,459,970
43,506,360
43,552,750
43,599,140
43,645,530
43,691,920
43,738,310
43,784,700
bp
RET
RASGEF1A
Breakpoints
Case 15
Case 11
Case 16
T
N
rs7068646
(Ch 10: 43,525,148)
rs3026737
(Ch 10: 43,596,728)
rs1864402
(Ch 10: 43,605,860)
Case 11
Case 16
Case 15
T
N
5450
5500
5550
rs2742234
(Ch 10: 43,612,609)
rs17028
(Ch 10: 43,623,812)
rs2252761
(Ch 10: 43,834,545)
Proximal (N-terminal) to breakpoints
Distal (C-terminal) to breakpoints

FIGURE 1.

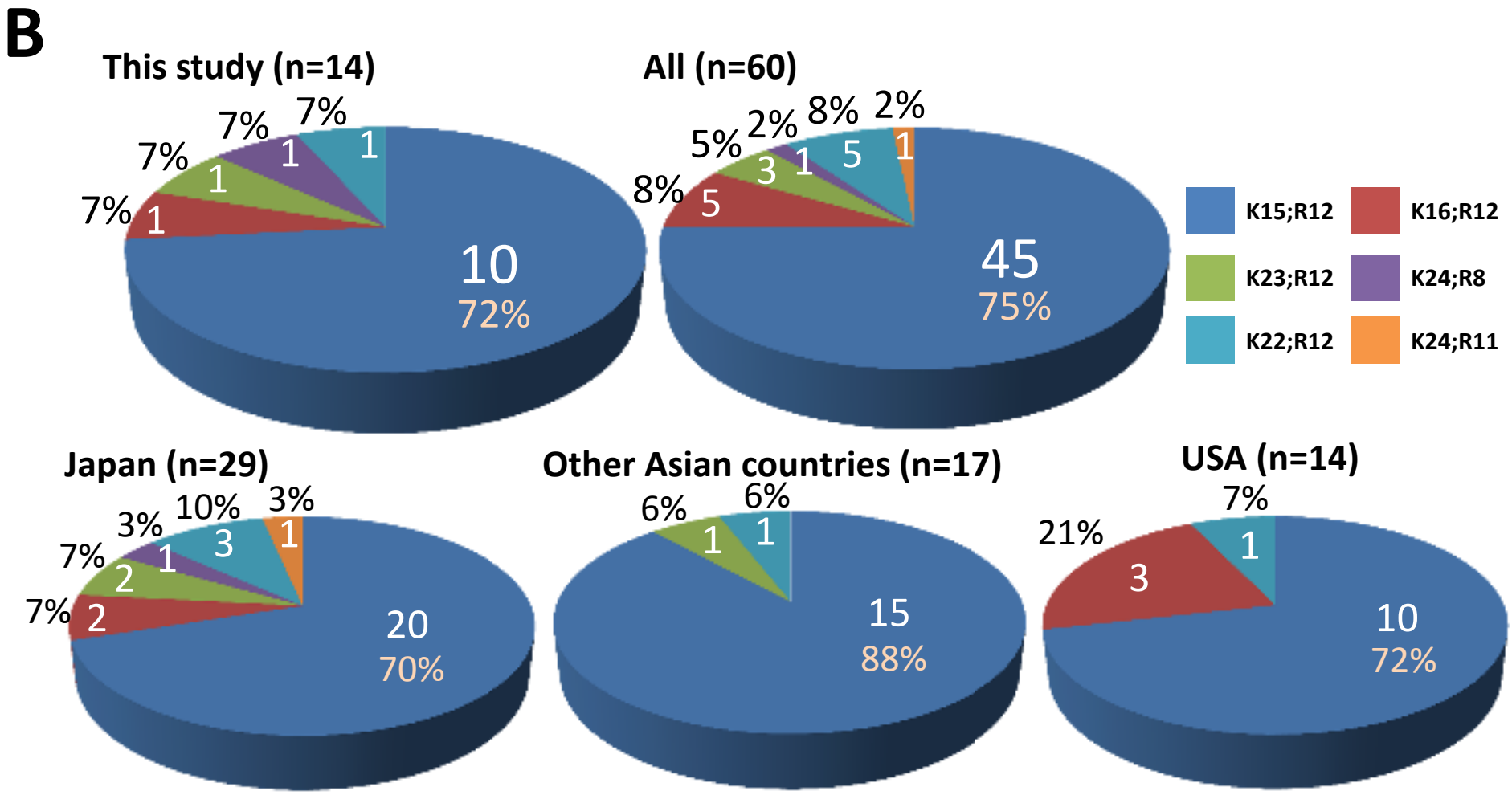

B
This study (n=14)
7%
7%
7%
7%
1
1
1
1
10
72%
All (n=60)
8%
5%
2%
8%
2%
5
3
1
5
1
45
75%
K15;R12
K16;R12
K23;R12
K24;R8
K22;R12
K24;R11
Japan (n=29)
7%
7%
3%
10%
3%
2
2
1
3
1
20
70%
Other Asian countries (n=17)
6%
6%
1
1
15
88%
USA (n=14)
21%
7%
3
1
10
72%

FIGURE 2.

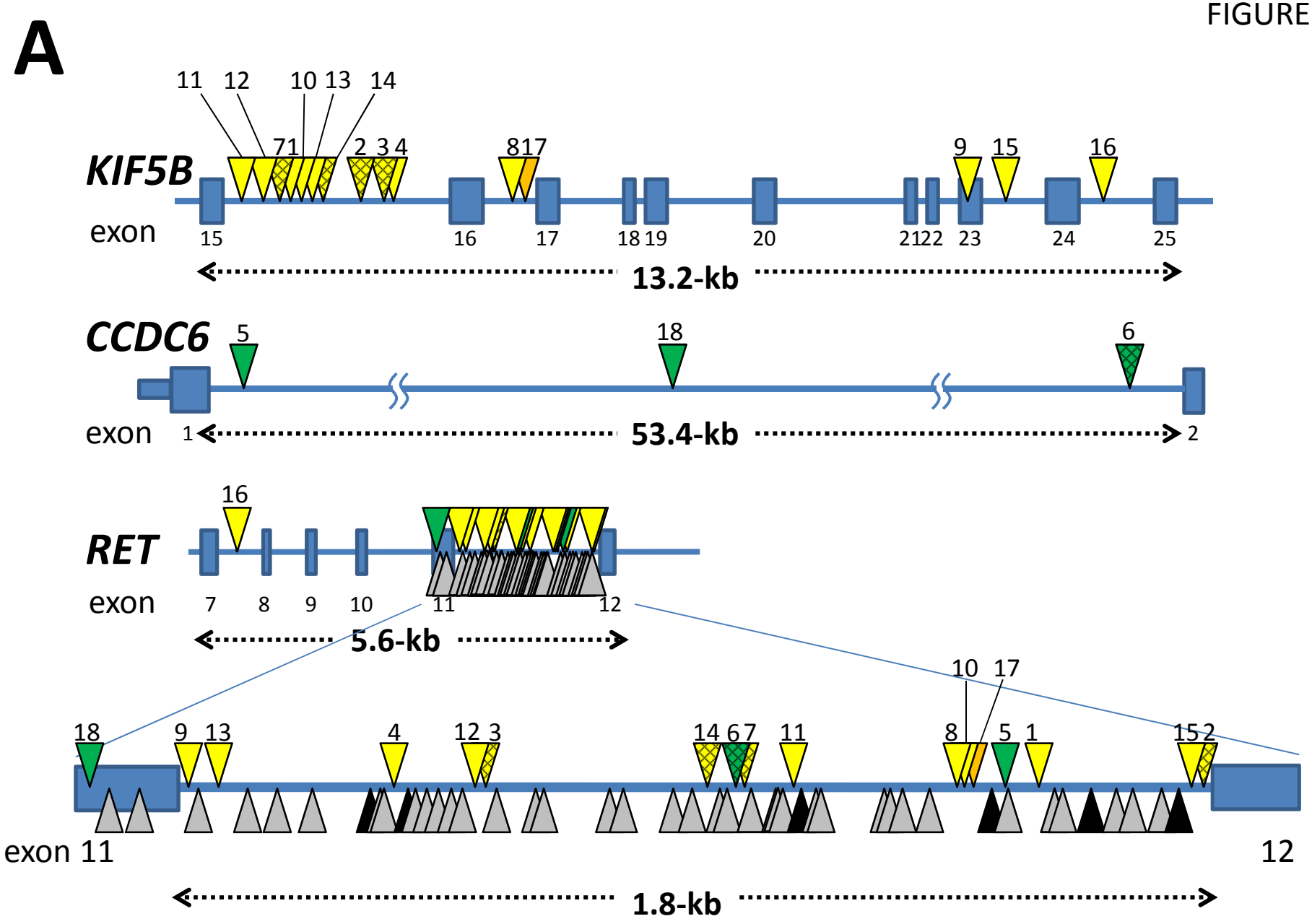

A
KIF5B
11 12 10 13 14
7 1
2 3 4
8 17
9 15
16
exon
15 16 17 18 19 20 21 22 23 24 25
13.2-kb
CCDC6
5
18
6
exon
1
53.4-kb
2
RET
16
exon
7 8 9 10 11 12
5.6-kb
10 17
18
9 13
4
12 3
14 6 7
11
8
5 1
15 2
exon 11
12
1.8-kb

FIGURE 2.

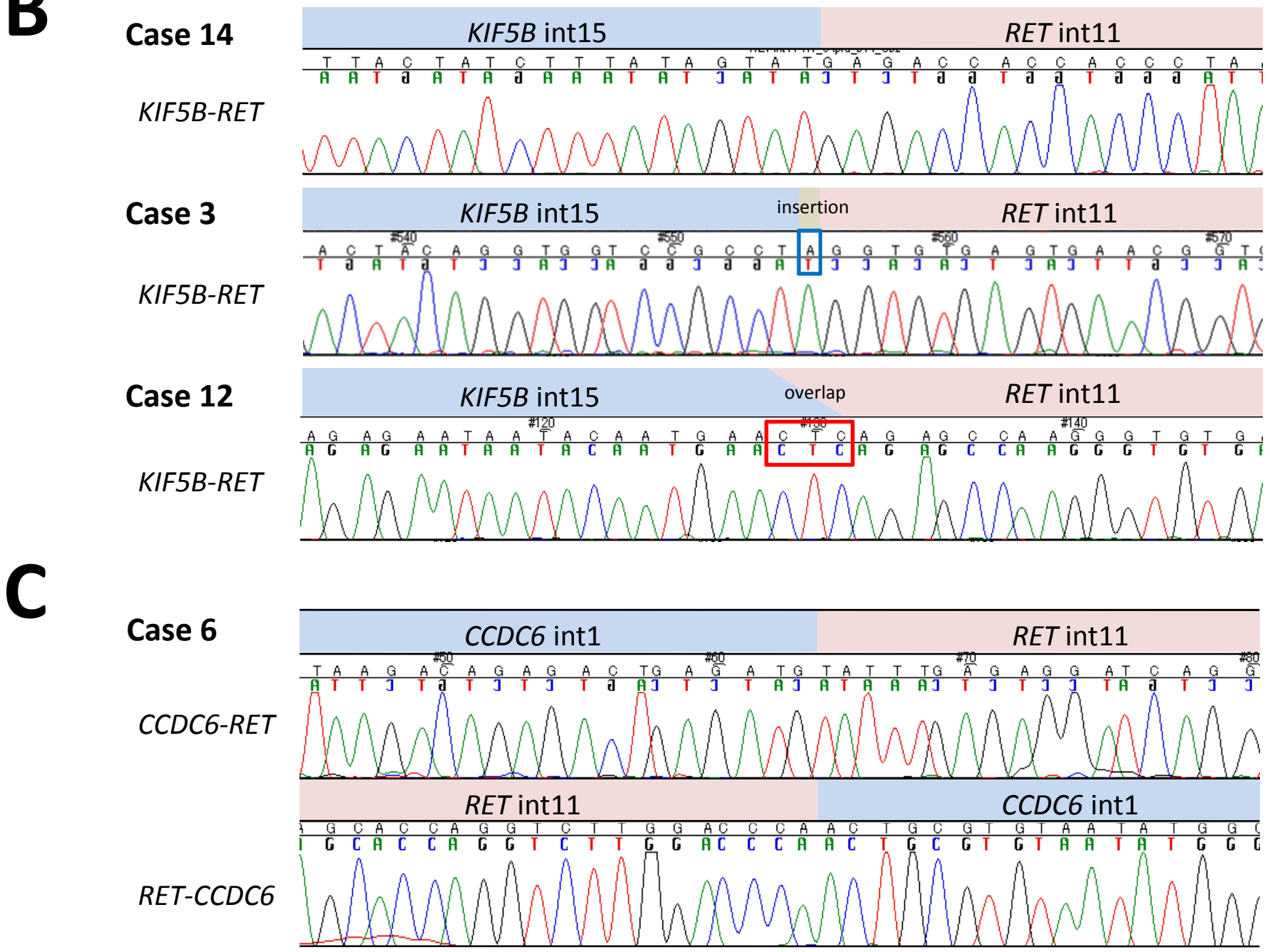

B
Case 14
KIF5B int15
RET int11
KIF5B-RET
Case 3
KIF5B int15
insertion
RET int11
KIF5B-RET
Case 12
KIF5B int15
overlap
RET int11
KIF5B-RET
C
Case 6
CCDC6 int1
RET int11
CCDC6-RET
RET int11
CCDC6 int1
RET-CCDC6

FIGURE 3.

A

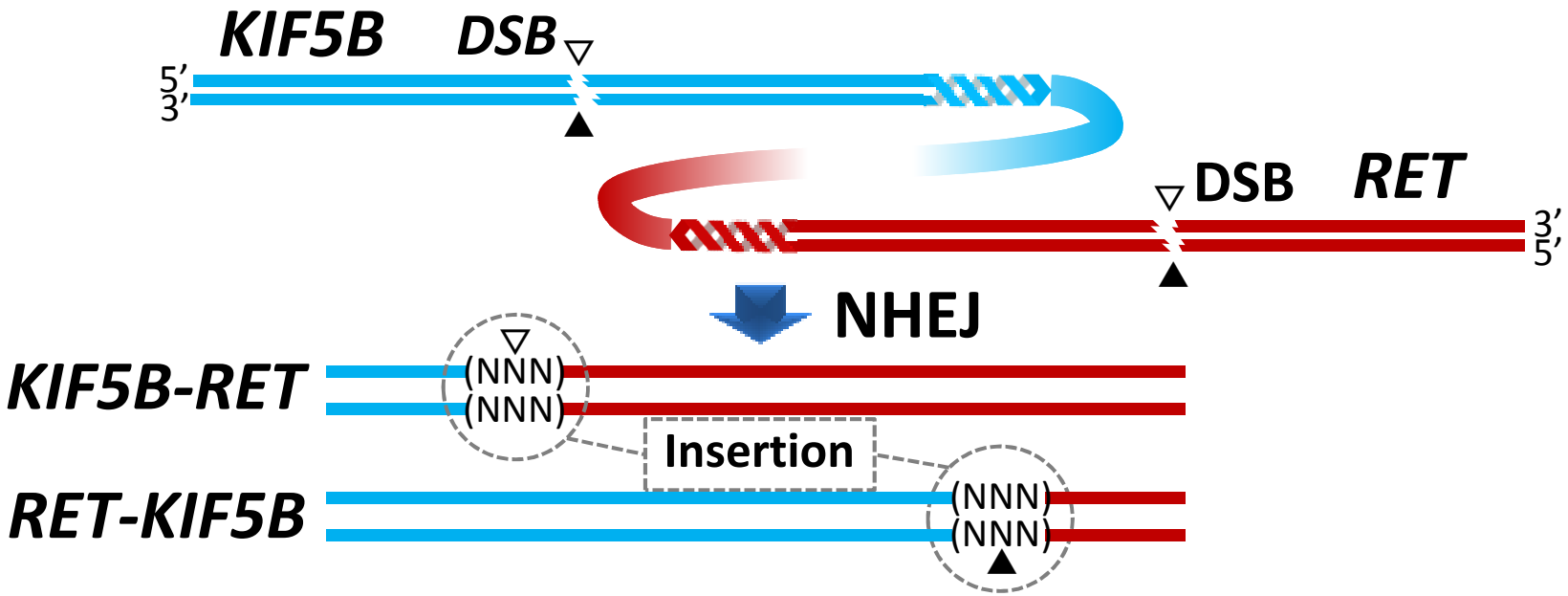

KIF5B
DSB
5’
3’
DSB
RET
3’
5’
NHEJ
KIF5B-RET
(NNN)
(NNN)
Insertion
RET-KIF5B
(NNN)
(NNN)

FIGURE 3.

B

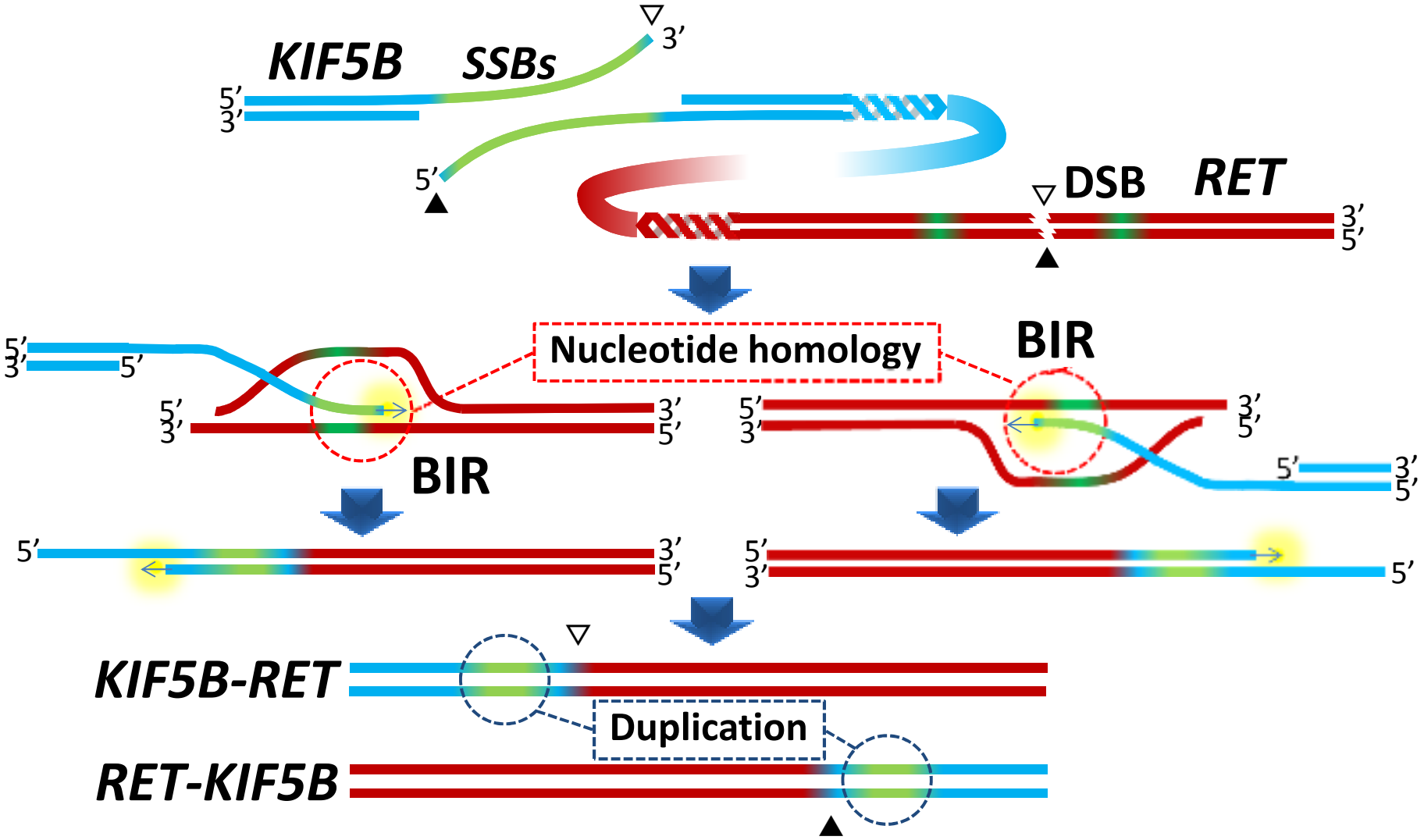
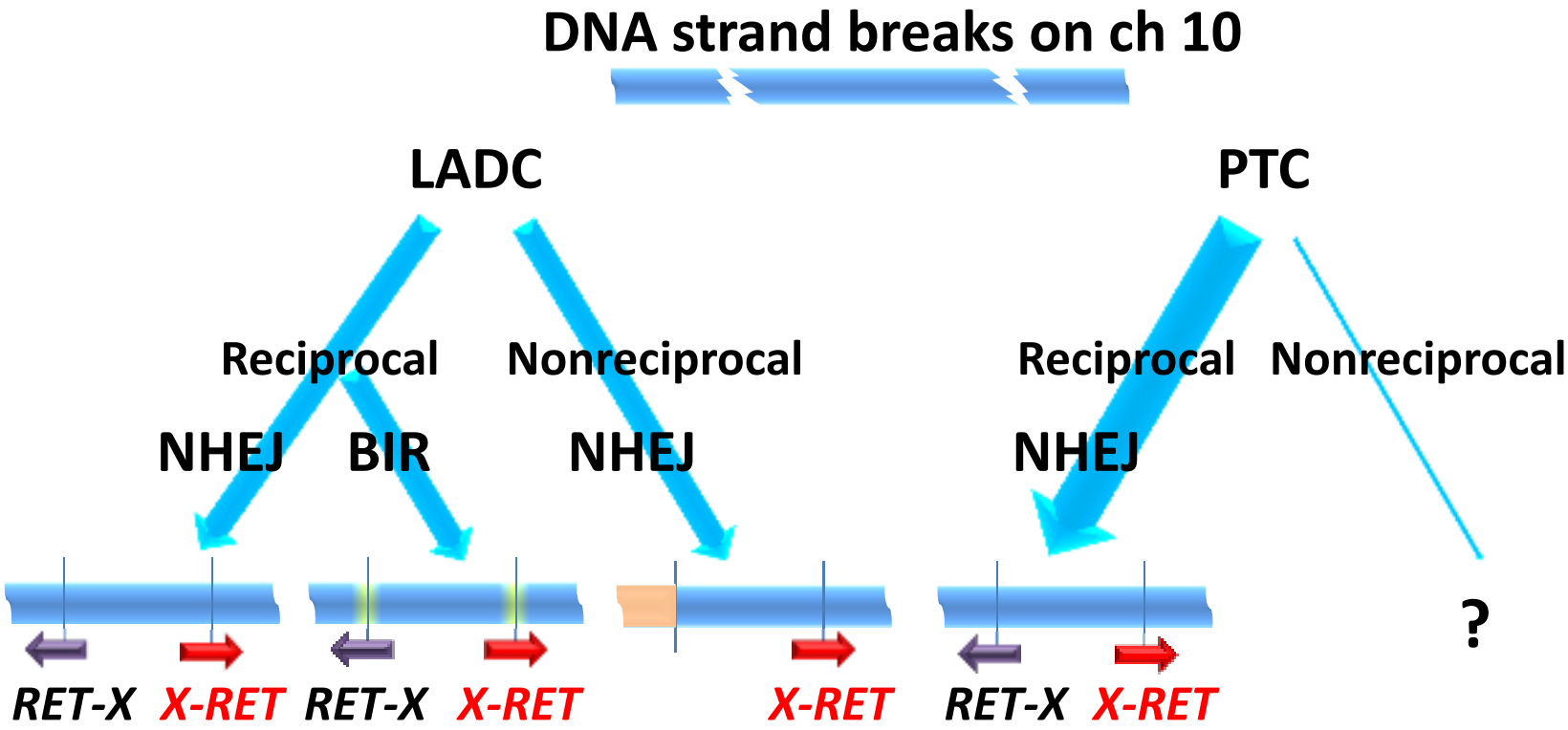
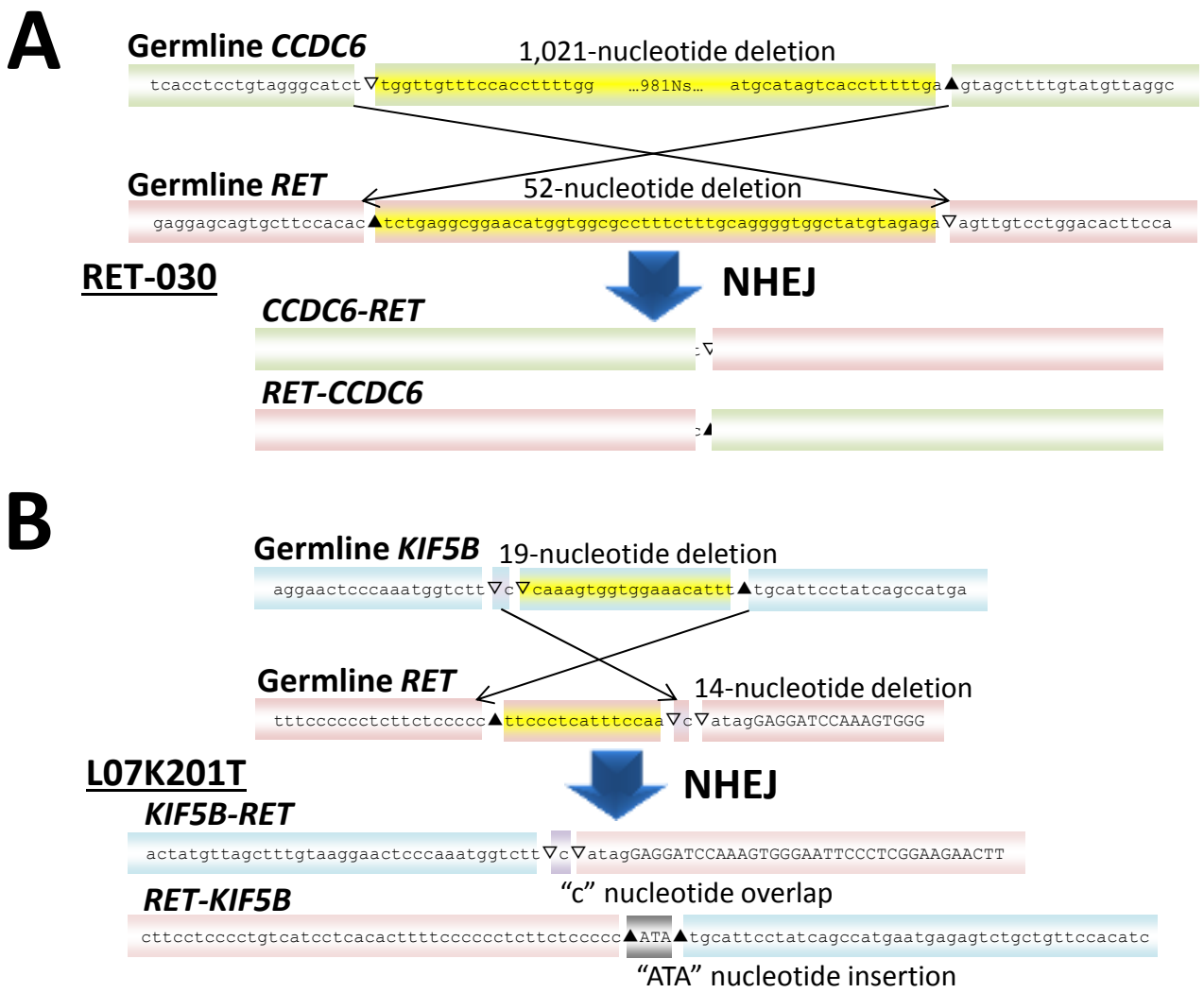
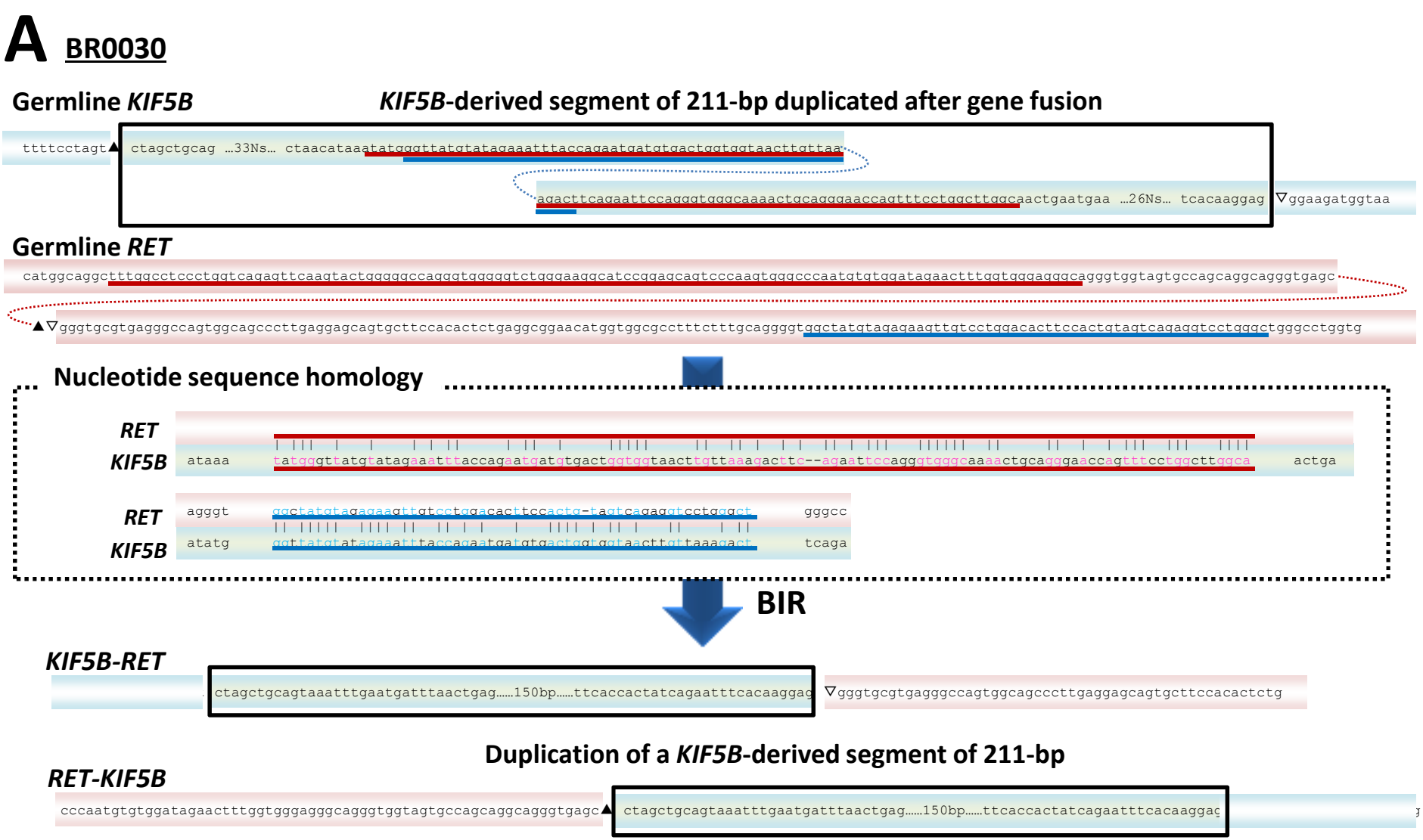

KIF5B
SSBs
5′
3′
3′
5′
DSB
RET
3′
5′
Nucleotide homology
BIR
BIR
KIF5B-RET
Duplication
RET-KIF5B

FIGURE 4.

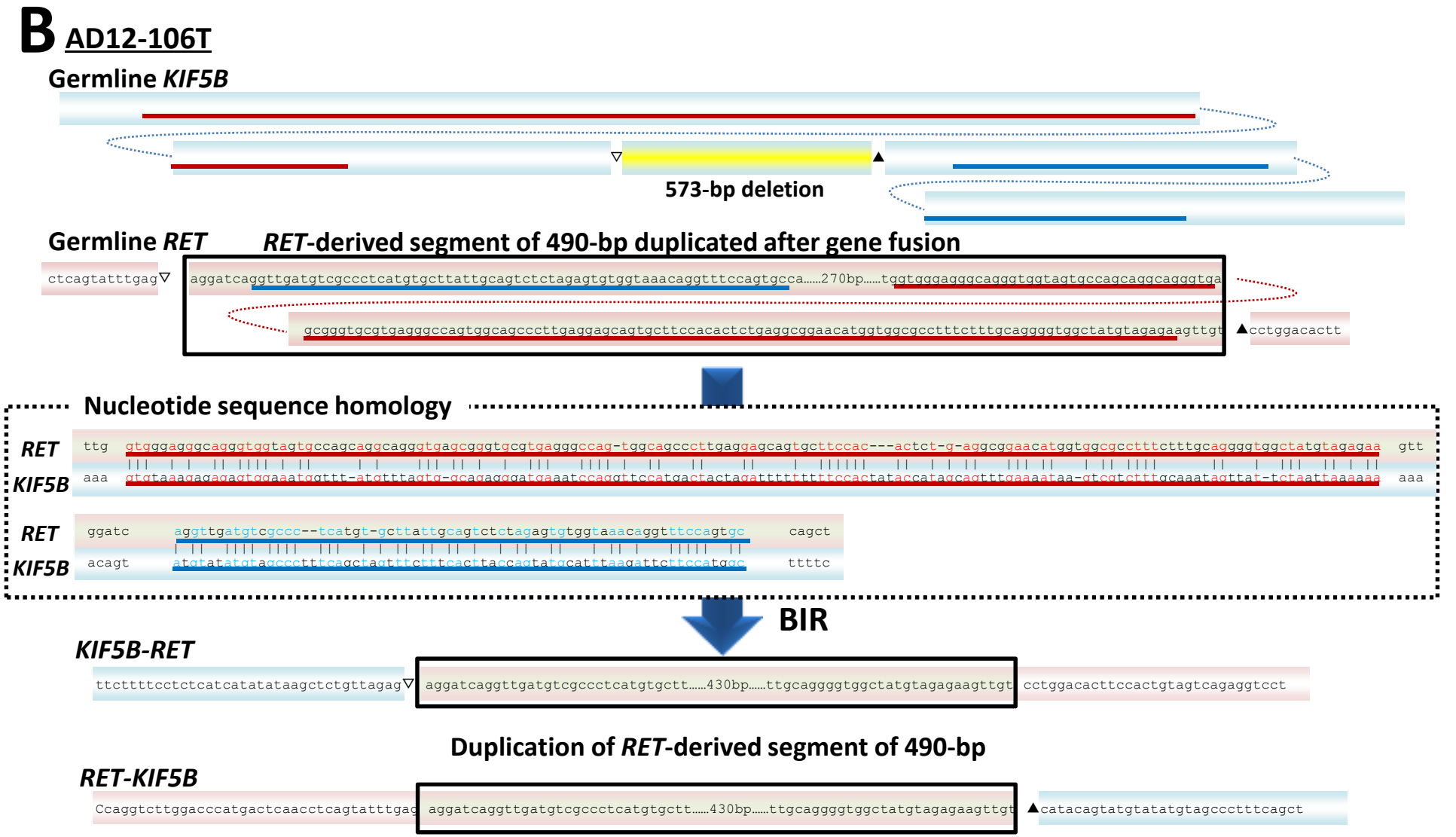

DNA strand breaks on ch 10
LADC
PTC
Reciprocal
Nonreciprocal
Reciprocal
Nonreciprocal
NHEJ
BIR
NHEJ
NHEJ
?
RET-X
X-RET
RET-X
X-RET
X-RET
RET-X
X-RET



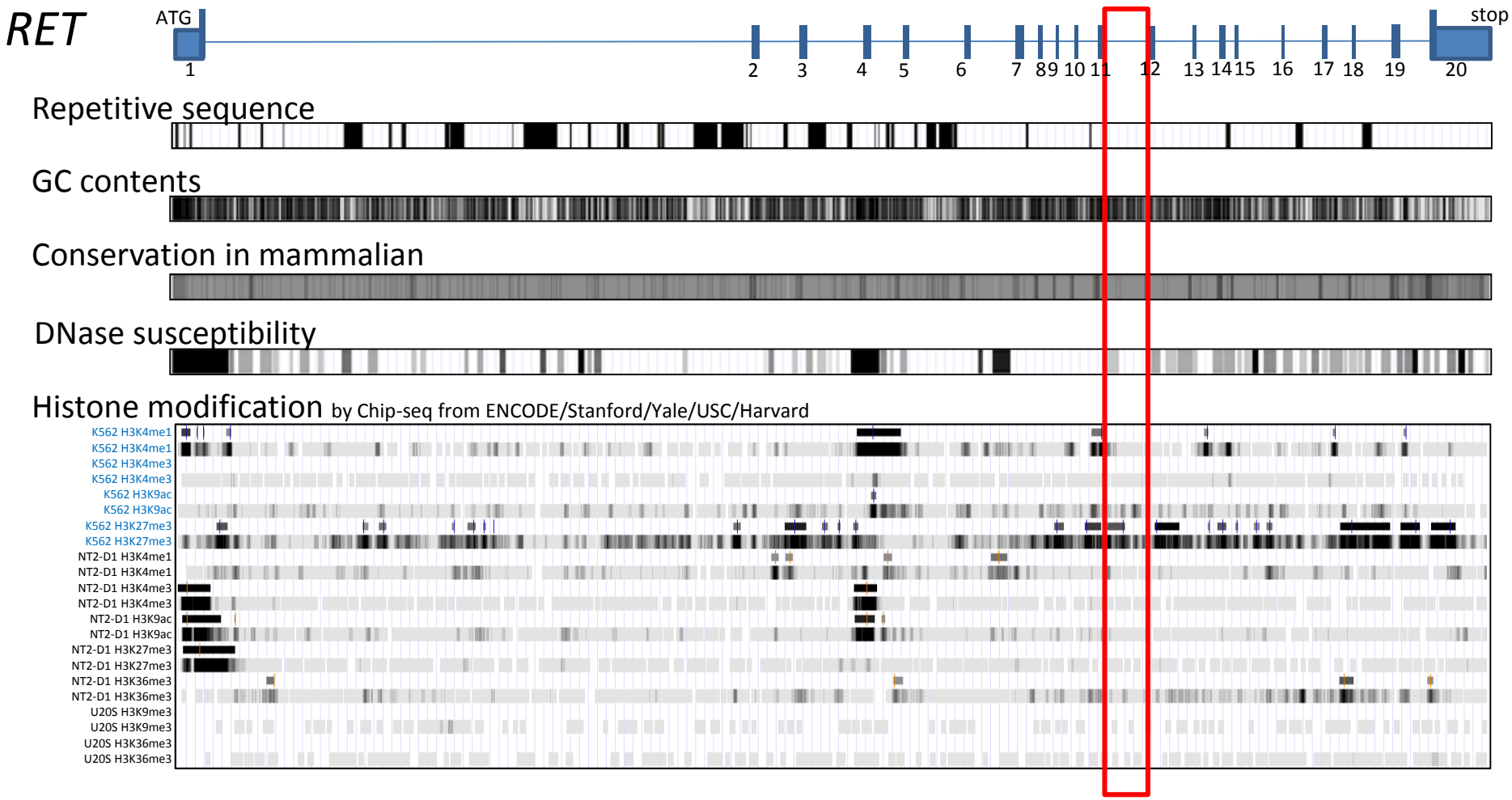
SUPPLEMENTARY FIGURE 1. Breakpoint and junction sequences of two representative cases with reciprocal inversions deduced to have been mediated by NHEJ. The structures of the two resultant fusion DNAs enabled us to deduce the nucleotide deletions in the *RET*, *CCDC6*, and *KIF5B* loci. Nucleotide insertion at the breakpoint junction, a feature of NHEJ, was observed in L07K201T. ▽, breakpoints for partner-*RET* fusion; ▲, breakpoints for *RET*-partner fusion.



SUPPLEMENTARY FIGURE 2. Breakpoint and junction sequences of two representative cases with reciprocal inversions resulting in segment duplication. (**A**) A *KIF5B* segment between two breakpoints (▽, *KIF5B-RET* fusion; ▲, *RET-KIF5B* fusion) has sequence homology with two regions encompassing breakpoints in the *RET* locus. The regions exhibiting homology are underlined in red and blue. The resultant fusion DNAs contain duplications of the *KIF5B* segment.



SUPPLEMENTARY FIGURE 2. Breakpoint and junction sequences of two representative cases with reciprocal inversions resulting in segment duplication. (**B**) A *RET* segment between two breakpoints (▽, *KIF5B-RET* fusion; ▲, *RET-KIF5B* fusion) has sequence homology with two regions encompassing breakpoints in the *KIF5B* locus. The regions exhibiting homology are underlined in red and blue. The resultant fusion DNAs contain duplications of the *RET* segment.



SUPPLEMENTARY FIGURE 3. DNA and chromatin features of the *RET* locus. Locations of repetitive sequences are indicated in black. GC content, extent of conservation, DNase susceptibility (DNase I Hypersensitivity Clusters in 125 cell types from ENCODE), and histone modifications are shown by gray boxes, with darker gray indicating a greater extent or magnitude of each parameter. The data was obtained using the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway). The exon 11 to intron 11 region is indicated by a red box.

Supplementary Notes

Analysis of sequence reads obtained by a second generation sequencer

Sequence reads produced by the Ion PGM sequencer were analyzed by a program developed by the authors. The program can detect gene fusion by searching for sequence reads whose right and left ends map onto two different genes. In more detail, the program takes a BAM file output from the Ion sequencer and performs local alignment between every sequence read from the sequencer and every sequence from a UCSC hg19 database of genes with introns using the BWA-SW program with default parameters. Then, it filters out reads with low mapping quality scores (<20) and screens for reads mapped onto two different genes. It further screens for reads whose spans are entirely mapped, by selecting reads with high proportion (≥0.9) of read bases that map onto two genes. It also screens for reads whose left and right ends map onto two genes, by getting the positions of bases that map onto one gene and the positions of bases that map onto the other gene. It next performs the Wilcoxon test between the two sets of positions and screens for reads with low p-values ($<10^{-5}$). The program extracts breakpoint positions from local alignment results. Detailed information about this program will be published elsewhere. We found reads that mapped onto both the *RET* gene and *CCDC6* gene, thereby detecting *CCDC6-RET* and *RET-CCDC6* fusion breakpoints.

Molecular mechanisms underlying oncogenic *RET* fusion in papillary thyroid cancer (PTC)

RET fusion is a common genetic aberration in PTC patients treated with external beam

radiation [1]. Notably, 50% of pediatric PTC caused by post-Chernobyl exposure to radiation involves *ELE1 (also known as RFG, NCOA4 and ARA70)-RET* fusions[2, 3]. In addition, *RET* fusions were identified in 20% of PTC induced after exposure to a nuclear bomb[4]. Taken together, these observations suggest that DNA strand breaks induced by irradiation trigger chromosome 10 inversions, which result in *RET* fusions.

Previous studies have described the structure of breakpoint junctions in post-Chernobyl and sporadic PTCs[5-8]. Breakpoints were clustered but dispersed within exon 11 to intron 11 of *RET* (gray and black arrowheads in **Figure 2**). No breakpoint was located within 4 bp of another. Even among all the breakpoints of all LADCs (this study) and thyroid carcinomas, no breakpoints were located at the same position. Previous genomic PCR analysis revealed that *RET* fusions result from reciprocal inversion in most cases (43/47, 91%)[5, 6]; in LADC, by contrast, reciprocal inversion accounted for just over half (56%) of *RET* fusions ($P = 0.0021$ by Fischer's exact test). The breakpoint junctions in PTC frequently contained nucleotide insertions, and therefore, joining in the PTC cases was previously deduced as being mediated by NHEJ[8].

Genome/chromatin structure of breakpoint cluster regions

The 2.0 kb region spanning the *RET* exon 11 to intron 11 region lacks repetitive sequence clusters and has an average GC content (**Supplementary Figure 4**). Furthermore, examinations of histone modifications in this region in several kinds of human cells revealed no distinct patterns associated with open chromatin structure; similarly, the DNase I sensitivity of the region, which may reflect accessibility to DNA-damaging agents, is not high. Interestingly, a recent study suggested that *RET*

intron 1 is easy to break during replication through DNA topoisomerase actions[9]. Therefore, this feature might be a cause for the susceptibility.

References

1. Bounacer A, Wicker R, Caillou B, et al. High prevalence of activating ret proto-oncogene rearrangements, in thyroid tumors from patients who had received external radiation. *Oncogene* 1997;15:1263-1273.
2. Klugbauer S, Lengfelder E, Demidchik EP, et al. High prevalence of RET rearrangement in thyroid tumors of children from Belarus after the Chernobyl reactor accident. *Oncogene* 1995;11:2459-2467.
3. Fugazzola L, Pilotti S, Pinchera A, et al. Oncogenic rearrangements of the RET proto-oncogene in papillary thyroid carcinomas from children exposed to the Chernobyl nuclear accident. *Cancer Res* 1995;55:5617-5620.
4. Hamatani K, Eguchi H, Ito R, et al. RET/PTC rearrangements preferentially occurred in papillary thyroid cancer among atomic bomb survivors exposed to high radiation dose. *Cancer Res* 2008;68:7176-7182.
5. Nikiforov YE, Koshoffer A, Nikiforova M, et al. Chromosomal breakpoint positions suggest a direct role for radiation in inducing illegitimate recombination between the ELE1 and RET genes in radiation-induced thyroid carcinomas. *Oncogene* 1999;18:6330-6334.
6. Bongarzone I, Butti MG, Fugazzola L, et al. Comparison of the breakpoint regions of ELE1 and RET genes involved in the generation of RET/PTC3 oncogene in sporadic and in radiation-associated papillary thyroid carcinomas. *Genomics* 1997;42:252-259.
7. Minoletti F, Butti MG, Coronelli S, et al. The two genes generating RET/PTC3 are localized in chromosomal band 10q11.2. *Genes Chromosomes Cancer* 1994;11:51-57.
8. Klugbauer S, Pfeiffer P, Gassenhuber H, et al. RET rearrangements in radiation-induced papillary thyroid carcinomas: high prevalence of topoisomerase I sites at breakpoints and microhomology-mediated end joining in ELE1 and RET chimeric genes. *Genomics* 2001;73:149-160.
9. Dillon LW, Pierce LC, Lehman CE, et al. DNA Topoisomerases Participate in Fragility of the Oncogene RET. *PLoS One* 2013;8:e75741.