Supplementary information

Supplementary note

Gene fusions other than *KIF5B-RET*

Sanger sequencing of the reverse transcription (RT)-PCR products led to the identification of five novel fusion transcripts other than *KIF5B-RET* (**Supplementary Table 1**). All these five were fusions of non-kinase genes, which mostly result in protein truncations, suggesting that they were passenger mutations. Therefore, they were not investigated further.

Structure of breakpoints for chromosome inversion

DNA sequences encompassing breakpoints revealed no significant homology. The breakpoints were joined without any nucleotide overlaps or insertions in case BR0020 (**Supplementary Fig. 1d**), while insertions (BR1001 and BR1003) or overlaps (BR1002 and BR0030) were identified in other cases. In case BR1004, joining was associated with the insertion of a 349-bp DNA fragment (**Supplementary Fig. 1d**). These data suggest that the *KIF5B-RET* fusions were produced through illegitimate repair of DNA double strand breaks through non-homologous end joining, consistent with many other chromosomal translocations identified in human cancers[1].

LADC with *RET* expression in the absence of *RET* fusion

A proportion (22%, 48/222) of LADCs without the *KIF5B-RET* fusion also

showed higher *RET* expression than that observed in non-cancerous lung tissues. Six such cases (**Supplementary Table 5**) were subjected to RNA sequencing, but a *RET* fusion to genes other than *KIF5B* was not detected; neither somatic/germline *RET* mutations nor an increased copy number of the *RET* locus were identified (**Supplementary Fig. 1f and 5**). Therefore, the mechanisms underlying increased *RET* expression in the fusion-negative cases are unknown and require further investigation. Association of *RET* expression in the absence of *RET* fusion in LADCs with their sensitivity to RET TKIs is also worth investigating.

Status of *EGFR* mutations and *KIF5B-RET* fusions in LADCs which responded to vandetanib

Vandetanib, a TKI against RET and other tyrosine kinases, including EGFR and VEGFR, has been shown to be effective for non-small cell lung carcinoma[2,3]. In light of the present results, the observed anti-tumor effect might be explained not only by suppression of EGFR and/or VEGFR, but also by suppression of KIF5B-RET. To address this issue, three cases of LADCs of responders to the vandetanib treatment in a previous clinical trial[4], for which tumor DNAs was available, were subjected to somatic *EGFR* mutation screening. All these cases were positive for *EGFR* mutation. Due to the lack of remaining DNAs and RNAs, *RET* fusions could not be examined. Therefore, the relationship between *RET* fusions and responsiveness to vandetanib in LADC *in vivo* in the human remains unknown.

KIF5B-RET fusions in human lung cancer cell lines

We screened 200 commonly-used human lung cancer cell lines[5] and all were negative for this fusion (data not shown); therefore, the oncogenic properties of the fusion remain un-validated in human cells. Functional analysis by exogenous expression of *KIF5B-RET* cDNAs in immortalized lung epithelial cell lines and examination of proliferation/ invasion will be significant to assess the tumorigenic potential of *KIF5B-RET* fusions.

Supplementary Materials and Methods

Samples

The Japanese cohort was comprised of 319 LADC patients undergoing resection at the National Cancer Center Hospital between 1997 and 2008. Tissue specimens were provided by the National Cancer Center Biobank, Japan. The US (UMD) cohort was comprised of 80 LADC patients recruited from hospitals in the Metropolitan Baltimore area between 1987 and 2009[6]. The Norwegian cohort was comprised of 34 LADC patients recruited from Haukeland University Hospital, Bergen, Norway, from 1988 to 2003[6]. All tumors were pathologically diagnosed according to the tumor-node-metastasis classification of malignant tumors[7,8]. Total RNA was extracted from grossly dissected, snap frozen tissue samples using TRIzol (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions and its quality was examined using a model 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). All samples showed RNA Integrity Numbers > 6.0. Genomic DNA was also extracted from the Japanese tissue samples using a QIAamp DNA Mini kit (Qiagen, Valencia, CA, USA). The study was approved by the institutional review boards of the institutions involved in the present study.

RNA sequencing

cDNA libraries for RNA sequencing were prepared using the mRNA-Seq Sample Prep Kit (Illumina, San Diego, CA, USA) according to the manufacturer's standard protocol. Briefly, poly-A(+) RNA was purified from 2 μg of total RNA for each sample, fragmented by incubation at 94°C for 5 min in

fragmentation buffer, and used for double-stranded cDNA synthesis. After ligation of the PE adapter, a fraction of 250–300 bp (insert size: 150–200 bp) was gel-purified and amplified with 15 cycles of PCR. The resulting libraries were subjected to the paired-end sequencing of 50-bp reads on the Genome Analyzer IIx (GAIIx) (Illumina).

Detection of fusion transcripts

Fusion transcript candidates that were supported by >20 paired-end reads and >10 junction reads in each LADC sample were selected (**Supplementary Table 1**). To detect paired-end reads derived from fusion transcripts, the recently described algorithm developed for analysis of genomic structural alteration[9] was modified and used. First, paired-end reads with the same sequences were removed, since they were deduced to be generated during the PCR amplification process. Both reads of the remaining paired-end reads were mapped on human reference RNA sequences deposited in the RefSeq database (File: human.rna.fna from ftp://ftp.ncbi.nih.gov/refseq, Date: Sep 20, 2010) using the BOWTIE program (version 0.12.5) allowing two or less mismatches. 'Proper' paired-end reads in which both reads were mapped on the same RNA sequences, with proper spacing and orientation, were removed. Paired-end reads containing a read with multiple hits on several genomic loci were also removed. Then, the remaining paired-end reads were assembled to 'clusters', and the resulting 'paired-clusters' were screened to select those indicating the presence of fusion transcripts, as follows: (i) clusters, which included reads aligned within the maximum insert distance, were separately constructed from

the forward and reverse alignments on reference RNA sequences; (ii) clusters in which the distance between the reads at both ends was greater than the maximum insert distance were discarded; (iii) paired-clusters in which one cluster was allocated in the forward orientation the other was allocated in the reverse orientation were selected; (iv) paired-clusters that included at least one paired-end read perfectly matched to reference RNA sequences were selected; and (v) paired-clusters mis-assembled due to nucleotide variations were removed (For this purpose, paired-end reads contained in paired-clusters were re-aligned to reference RNA sequences using the BLASTN program. If one end sequence was aligned to a region of paired-clusters and the other end was aligned on the same RNA sequences with proper spacing and orientation, the paired-cluster was removed. An expectation value of 1,000 was used as a cutoff value.). Then, paired-clusters consisting of >20 paired-end reads in each LADC sample, for which paired-end reads did not appear in any of three noncancerous lung tissues, were picked up. When mapped on human genome sequence, paired-clusters that were mapped within a gene region or a neighboring-gene region were not investigated further due to the possibility that they were alternatively spliced or read-through transcripts that have not been deposited in the RefSeq database. Junction reads encompassing the fusion boundaries were searched using the MapSplice (version 1.14.1) software, as follows: a DNA sequence was made by joining two genome DNA sequences comprising a read cluster region and a 300-bp adjacent region for each gene, and junction reads were searched against this DNA sequence using the MapSplice software.

RT-PCR, genomic PCR and Sanger sequencing

Total RNA (500 ng) was reverse-transcribed to cDNA using Superscript III Reverse Transcriptase (Invitrogen). cDNA (corresponding to 10 ng total RNA) or 10 ng genomic DNA was subjected to PCR amplification using KAPA Taq DNA Polymerase (KAPA Biosystems, Woburn, MA, USA). The reactions were carried out in a thermal cycler under the following conditions: 40 cycles of 95°C for 30 sec, 60°C for 30 sec and 72°C for 2 min, with a final extension for 10 min at 72°C. The gene encoding glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) was amplified to estimate the efficiency of cDNA synthesis. The PCR products were directly sequenced in both directions using the BigDye Terminator kit and an ABI 3130xl DNA Sequencer (Applied Biosystems, Foster City, CA, USA). The PCR primers used in the present study are shown in **Supplementary Table 6**.

Analysis of *EGFR, KRAS, HER2* and *ALK* mutations

Genomic DNAs from LADC tissues were analyzed for somatic mutations in the *EGFR* and *KRAS* genes using the high-resolution melting method as previously described[10]. Genomic DNA was also subjected to mutation analysis for exon 20 of the *HER2* gene using a PCR-direct sequencing method, as previously described[11]. Total RNAs from the same tissues were examined for expression of *ALK/EML4* or *ALK/KIF5B* fusion transcripts using a multiplex reverse transcription-PCR method[12].

Genome copy number analysis

The genome copy number for the LADCs was determined using GeneChip Human Mapping 250-K SNP arrays (Affymetrix, Santa Clara, CA, USA) and the Copy Number Analyzer for Affymetrix GeneChip Mapping Array (CNAG) software[13], as previously described[14].

Microarray experiments and data processing

In total, 228 cases were subjected to expression profiling. Total RNA (100 ng) was labeled using a Two-Cycle Target Labeling and Control Reagents (Ambion, Austin, TX, USA) and analyzed using Human Genome U133A Plus 2.0 arrays (Affymetrix). The obtained data were normalized using the MAS5 algorithm and the mean expression level of 54,675 probes was adjusted to 1,000 for each sample.

Immunohistochemistry

Immunohistochemistry was performed on tissue microarray sections. Sections 4-μm thick were deparaffinized and heat-induced epitope retrieval was performed using a targeted retrieval solution (Dako, Carpinteria, CA, USA). The slides were then treated with 3% hydrogen peroxide for 20 min to block endogenous peroxidase activity, followed by washing in deionized water for 2–3 min. The slides were then incubated with primary antibodies against RET (1:250, clone 3454_1; Epitomics, Burlingame, CA, USA), TTF-1 (1:100, clone 8G7G3/1; Neomarker, Fremont, CA, USA) or Napsin A (1:400, TMU-Ad02, IBL, Gunma, Japan) for 1 hr at room temperature. For immunohistochemical

staining of thyroglobulin, heat-induced epitope retrieval was performed using citrate buffer (0.02 mol/L; pH 6.0). The slides were then incubated with primary antibodies against thyroglobulin (1:100, DAK-Tg6, DAKO, Carpinteria, CA) for 1 hr at room temperature. Immunoreactions were detected using the Envision-Plus system (Dako) for TTF-1 and the EnVision-FLEX and LINKER (Dako) systems for RET, respectively. The reactions were visualized with 3,3'-diaminobenzidine followed by counterstaining with hematoxylin. Nuclear staining of > 10% of tumor cells was considered positive for TTF-1 and cytoplasmic staining was considered positive for RET.

Fluorescent *in situ* hybridization

To identify *RET* rearrangements, Fluorescent in situ hybridization (FISH) was performed on formalin-fixed, paraffin-embedded tumors by using a break-apart probe for *RET* (GSP laboratory, Kawasaki, Japan; TexRed-labeled 5'*RET* probe and FITC-labeled 3'*RET* probe; **Supplementary Fig. 2**).

Materials for functional analysis of *KIF5B-RET*

KIF5B-RET variant 1 cDNA was obtained by PCR amplification against cDNA for the BR1001 tumor using KOD-PLUS Taq polymerase (Toyobo, Osaka, Japan). The product was subcloned into a pLenti-V5-DEST plasmid (Invitrogen). The KIF5B-RET-KD (kinase dead mutant corresponding to S765P (ref)) expression plasmid was constructed using the Quick Change Site-Directed Mutagenesis Kit (Invitrogen). Mutant *KRASV12* cDNA was also subcloned into a pLenti-V5-DEST plasmid. The integrity of the inserted cDNA was verified by

Sanger sequencing of the plasmid. Lentiviruses expressing *KIF5B-RET* or mutant *KRASV12* were produced according to the manufacture's instruction. A pcDNA3 plasmid expressing wild-type *RET* was obtained from Dr. Masahide Takahashi of Nagoya University. Vandetanib (ZD6474) was purchased from Selleck (Houston, TX, USA). Stock solution (50 mM) was made in 100% DMSO and diluted with culture media.

Constitutive activation of RET kinase by fusion to KIF5B

NCI-H1299 lung cancer cells were transfected using a Lipofectamine 2000 reagent (Invitrogen) for 24 hr with 0.8 μg of empty, wild-type *RET* or *KIF5B-RET* expression plasmids with 10% FBS, and then serum starved for another 24 hr. After 2 hr treatment with DMSO or 5 μM vandetanib, the cells were treated with 10% FBS for 10 min. Whole cell lysates were subjected to immunoblot analysis as below.

Immunoblot analysis of RET protein

Whole cell lysates from cell lines and cancerous and non-cancerous lung tissues were subjected to SDS-PAGE followed by blotting onto a polyvinylidene difluoride membrane. The membranes were blocked overnight with TBS containing 0.1% Tween 20 and 0.5% BSA, and probed with the primary antibodies, anti-RET (3454_1; Epitomics) and anti-phospho-Ret (Tyr905) (#3221; Cell Signaling, Danvers, MA, USA). After washing with TBS containing 0.1% Tween 20, the membranes were incubated with horseradish peroxidase-conjugated anti-mouse or anti-rabbit secondary antibodies and

treated with an enhanced chemiluminescence reagent (GE Healthcare, Tokyo, Japan). Intensities of RET and phospho-RET signals were calculated using LAS3000 (Quansys Biosciences, West Logan, UT, USA).

Transforming activity of *KIF5B-RET*

NIH3T3 fibroblast cells were infected with empty, *KIF5B-RET* or *KRASV12* –expressing lentiviruses, and were treated with Blastcidin (4 μg/ml) for 2 weeks. Mass-cultured Blastcidin-resistant cells were seeded on 24-well plates (1,000 cells/well) in 0.35% agar in complete medium with or without vandetanib (0.25, 0.5 and 1.0 μM) on a base layer of 0.5% agar. The expression of exogenous KIF5B-RET, KIF5B-RET-KD and KRASV12 proteins and phosphorylation of RET-Tyr905 were examined by immunoblot analysis. The compound solution was added to the top layer every 3 days. Colonies were counted 21 days later.

References

1. Mani, R.S. & Chinnaiyan, A.M. Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. *Nat Rev Genet* **11**, 819-829 (2010).
2. Herbst, R.S., Heymach, J.V. & Lippman, S.M. Lung cancer. *The New England journal of medicine* **359**, 1367-1380 (2008).
3. Natale, R.B.*, et al.* Phase III trial of vandetanib compared with erlotinib in patients with previously treated advanced non-small-cell lung cancer. *J Clin Oncol* **29**, 1059-1066 (2011).
4. Tamura, T.*, et al.* A phase I dose-escalation study of ZD6474 in Japanese patients with solid, malignant tumors. *J Thorac Oncol* **1**, 1002-1009 (2006).
5. Gazdar, A.F., Girard, L., Lockwood, W.W., Lam, W.L. & Minna, J.D. Lung cancer cell lines as tools for biomedical discovery and research. *J Natl Cancer Inst* **102**, 1310-1321 (2010).
6. Saito, M.*, et al.* The Association of MicroRNA Expression with Prognosis and Progression in Early-Stage, Non-Small Cell Lung Adenocarcinoma: A Retrospective Analysis of Three Cohorts. *Clin Cancer Res* **17**, 1875-1882 (2011).
7. Travis, W.D., Brambilla, E., Muller-Hermelink, H.K. and Harris, C.C. . Pathology and Genetics, Tumours of Lung, Pleura, Thymus and Heart. *In Travis, W.D., Brambilla, E., Muller-Hermelink, H.K. and Harris, C.C. (eds.), World Health Organization Classification of Tumors;* , 1-344 (2004).
8. Travis, W.D., Colby, T.V., Corrin, B., Shimosato, Y., Brambilla, E. (ed.) *World Health Organization. International Histological Classification of Tumors: Histological typing of lung and pleural tumors*, (Springer-Verlag, Heidelberg, Germany, 1999).
9. Totoki, Y.*, et al.* High-resolution characterization of a hepatocellular carcinoma genome. *Nature genetics* **43**, 464-469 (2011).
10. Takano, T.*, et al.* Epidermal growth factor receptor gene mutations and increased copy numbers predict gefitinib sensitivity in patients with recurrent non-small-cell lung cancer. *J Clin Oncol* **23**, 6829-6837 (2005).
11. Shigematsu, H.*, et al.* Somatic mutations of the HER2 kinase domain in lung adenocarcinomas. *Cancer Res* **65**, 1642-1646 (2005).
12. Takeuchi, K.*, et al.* Multiplex reverse transcription-PCR screening for EML4-ALK fusion transcripts. *Clin Cancer Res* **14**, 6618-6624 (2008).
13. Nannya, Y.*, et al.* A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res* **65**,

6071-6079 (2005).

14. Iwakawa, R.*, et al.* MYC Amplification as a Prognostic Marker of Early-Stage Lung Adenocarcinoma Identified by Whole Genome Copy Number Analysis. *Clin Cancer Res* (2011).

Supplementary Table 1. Lung adenocarcinoma cases subjected to RNA-sequencing

No.	Sample	Tumor or normal	Sex	Gene mutation	Age	Smoking (Pack-years)	Pathological stage	No. total read	Gene fusion* (No. paired-end reads/junction reads)
1	BR0009	Tumor	Male	*ALK*	30	Ever-smoker (30)	IIB	30,067,759	EML4-ALK: E13; A20 (variant 1) (91/64)
2	BR0052**	Tumor	Female	*ALK*	38	Ever-smoker (7)	IIA	27,841,176	EML4-ALK: E13; A20 (variant 1) (60/67)
3	BR0003	Tumor	Female	*EGFR*	60	Never-smoker	IIB	33,358,341	TMEM209-DPP6 (61/16), ZNF862-WDR91*** (24/22
4	BR0044	Tumor	Male	*EGFR*	68	Ever-smoker (10)	IIB	32,262,234	
5	BR0005	Tumor	Male	*KRAS*	59	Ever-smoker (62)	IIB	23,297,267	
6	BR0016	Tumor	Female	*KRAS*	75	Never-smoker	IIB	27,188,879	
7	BR0001	Tumor	Male	None	68	Ever-smoker (86)	IB	23,467,018	MTAP-CDKN2BAS*** (23/37)
8	BR0004	Tumor	Male	None	62	Ever-smoker (46)	IIB	33,540,967	
9	BR0006	Tumor	Female	None	62	Ever-smoker (38)	IB	21,386,586	
10	BR0012	Tumor	Male	None	65	Ever-smoker (92)	IIB	28,465,957	
11	BR0013	Tumor	Female	None	58	Never-smoker	IB	28,740,939	
12	BR0014	Tumor	Male	None	52	Ever-smoker (68)	IIB	21,036,216	
13	BR0015	Tumor	Female	None	49	Ever-smoker (20)	IIB	21,125,603	
14	BR0019**	Tumor	Female	None	54	Never-smoker	IIB	28,868,572	
15	BR0020	Tumor	Male	None	57	Never-smoker	IIB	25,404,815	KIF5B-RET: K15; R12 (variant 1) (30/12)
16	BR0026	Tumor	Male	None	58	Ever-smoker (41)	IIB	23,458,513	
17	BR0027	Tumor	Male	None	68	Ever-smoker (49)	IIA	22,344,197	
18	BR0029	Tumor	Male	None	53	Ever-smoker (70)	IA	31,534,640	FAM3C-CADPS2*** (21/20)
19	BR0031	Tumor	Male	None	54	Never-smoker	IA	32,324,822	CDC42-TMCO4 (41/28)
20	BR0032	Tumor	Female	None	61	Never-smoker	IA	22,693,962	
21	BR0033	Tumor	Male	None	69	Ever-smoker (102)	IIB	29,321,549	
22	BR0034	Tumor	Male	None	66	Ever-smoker (45)	IA	29,069,205	
23	BR0035	Tumor	Male	None	61	Ever-smoker (42)	IIB	25,683,757	
24	BR0036**	Tumor	Male	None	66	Ever-smoker (37)	IB	21,790,422	
25	BR0037	Tumor	Male	None	63	Ever-smoker (40)	IB	31,571,318	
26	BR0038	Tumor	Female	None	47	Never-smoker	IA	26,967,632	
27	BR0040	Tumor	Male	None	69	Ever-smoker (105)	IIA	27,417,202	
28	BR0041	Tumor	Male	None	46	Ever-smoker (27)	IIB	29,761,895	
29	BR0043	Tumor	Male	None	68	Ever-smoker (48)	IIB	26,223,934	
30	BR0045	Tumor	Female	None	64	Never-smoker	IIB	24,191,784	
31	BR0052	Normal	Female	-	38	Ever-smoker (7)	IIA	28,563,803	-
32	BR0036	Normal	Male	-	66	Ever-smoker (37)	IB	33,523,557	-
33	BR0019	Normal	Female	-	54	Never-smoker	IIB	29,088,007	-

*Gene fusions represented by >20 paired-end reads are shown.

**Corresponding non-cancerous lung tissue RNA of these cases was also subjected to RNA sequencing.

***Out of frame fusions.

Supplementary Table 2. Characteristics of lung adenocarcinoma cases

Variable	Japan							USA	Norway
	All (%)	Mutation type							
		EGFR [a] (%)	*KRAS* [a] (%)	*ALK* [b] (%)	*RET* [b] (%)	*HER2* [c] (%)	None (%)		
Total	319	169	30	11	6	6	97	80	34
Age (mean ± SD; years)	61.7 ± 8.6	61.0 ± 7.9	62.3 ± 9.1	53.5 ± 13.2	57.0 ± 15.2	62.3 ± 6.9	63.7 ± 8.2	64.1± 9.7	64.2± 11.7
Sex									
Male (%)	158 (49.5)	68 (40.2)	19 (63.3)	2 (18.2)	3 (50.0)	2 (33.3)	64 (66.0)	42 (52.5)	19 (55.9)
Female (%)	161 (50.5)	101 (59.8)	11 (36.7)	9 (81.8)	3 (50.0)	4 (67.7)	33 (34.0)	38 (47.5)	15 (44.1)
Smoking habit									
Never-smoker (%)	157 (49.2)	92 (54.4)	12 (40.0)	7 (63.6)	6 (100.0)	4 (67.7)	36 (37.1)	5 (6.3)	2 (5.9)
Ever-smoker (%)	162 (50.8)	77 (45.6)	18 (60.0)	4 (36.4)	0 (0.0)	2 (33.3)	61 (62.9)	73 (91.3)	31 (91.2)

[a]Mutations detected by high-resolution melting assay.

[b]Fusions detected by RT-PCR.

[c]Mutations detected by direct sequencing.

Supplementary Table 3. Lung cancer cases subjected to RET fusion screening

Histological type	Japanese cohort (NCC)						USA cohort						Norway cohort					
	No. poisitive	/	No. cases screened	(	%	)	No. poisitive	/	No. cases screened	(	%	)	No. poisitive	/	No. cases screened	(	%	)
Adenocarcinoma	6	/	319	(	1.9	)	1	/	80	(	1.3	)	0	/	34	(	0.0	)
Squamous cell carcinoma	0	/	205	(	0.0	)	-	/	-	(	-	)	0	/	29	(	0.0	)
Large cell carcinoma	-	/	-	(	-	)	-	/	-	(	-	)	0	/	17	(	0.0	)
Small cell carcinoma	0	/	20	(	0.0	)	-	/	-	(	-	)	-	/	-	(	-	)

Supplementary Table 4. Characteristics of lung adenocarcinoma with *HER2* mutation

No.	Sample	Country	Sex	Age	Smoking (Pack-years)	*HER2* mutation type	Oncogene mutation*	Pathological stage	Pathological findings
1	BR1105	Japan	Female	68	Never-smoker	G776V,Cins	None	IB	Moderately differentiatedADC
2	BR1107	Japan	Female	66	Never-smoker	YVMA 776-779 ins	None	IA	Moderately differentiatedADC
3	BR1207	Japan	Male	66	Smoker (45)	YVMA 776-779 ins	None	IA	Moderately differentiated ADC
4	BR1211	Japan	Female	52	Never-smoker	YVMA 776-779 ins	None	IB	Moderately differentiated ADC
5	BR1234	Japan	Male	55	Smoker (36)	YVMA 776-779 ins	None	IA	Well differentiated ADC
6	BR1330	Japan	Female	67	Never-smoker	G776V,Cins	None	IA	Moderately differentiated ADC

**EGFR* and *KRAS* mutations; and *ALK* and *RET* fusions.

NT: not tested.

Supplementary Table 5. Characteristics of lung adenocarcinoma with *RET* expression in the absence of *RET* fusion

No.	Sample	Sex	Age	Smoking (Pack-years)	Oncogene mutation	Pathological stage	Pathological findings
1	BR0012	Male	65	Ever-smoker (92)	None	IIB	Poorly differentiated ADC
2	BR0005	Male	59	Ever-smoker (62)	KRAS	IIB	Poorly differentiated ADC
3	BR0015	Female	49	Ever-smoker (20)	None	IIB	Poorly differentiated ADC
4	BR0031	Male	54	Never-smoker	None	IA	Well differentiated ADC
5	BR0043	Male	68	Ever-smoker (48)	None	IIB	Moderately differentiated ADC
6	BR0032	Female	61	Never-smoker	None	IA	Well differentiated ADC

Supplementary Table 6. PCR Primers

No.	Name	Location	Sequence	Use
KIF5B-RET fusion				
1	KIF5B-RET-F1	KIF5B exon 15	AGGAAATGACCAACCACCAG	RT-PCR and sequencing
2	KIF5B-RET-R1	RET exon 12	TCCAAATTCGCCTTCTCCTA	RT-PCR and genomic PCR
3	KIF5B-int15-F1	KIF5B intron 15	CCATAAGTGAAATGATTGGAAC	Genomic PCR
4	KIF5B-int15-F2	KIF5B intron 15	GATTTGTATGTTGCAGTAGCTG	Genomic PCR
5	KIF5B-ex16-F1	KIF5B exon 16	GGAGTTAGCAGCATGTCAGC	Genomic PCR and sequencing
6	KIF5B-ex23-F1	KIF5B exon 23	GCTCACTAAAGTGCACAAACAG	Genomic PCR and sequencing
7	KIF5B-ex24-F1	KIF5B exon 24	GAAGAGGGCATTCTGCACAG	Genomic PCR
8	RET-int11-R3	RET intron 11	GGAGGCTCCAGGATACTCGG	Genomic PCR
9	RET-int7-R1	RET intron 7- exon 8	CCTCCTCGGCCACATCTG	Genomic PCR
10	KIF5B-F-orf2438	KIF5B exon 22-23	AGAGTGCTGAGATTGATTCTG	Sequencing
11	KIF5B-int15-F3.5	KIF5B intron 15	CCCGAGTAGCTAGGATTACA	Sequencing
12	RET-int11-R0.5	RET intron 11	ATGACAGGTGTGGTCACAGC	Sequencing
13	RET-int11-R1	RET intron 11	TATCCACACATTGGGCCCAC	Sequencing
14	RET-int7-R2	RET intron 7	ATGGCAGCTGTGTCAGCATG	Sequencing
RET expression				
1	RET-F-orf2154	RET exon 12	ATTCCCTCGGAAGAACTTGG	RT-PCR
2	RET-R-orf2364	RET exon 13	GATGACATGTGGGTGGTTGA	RT-PCR

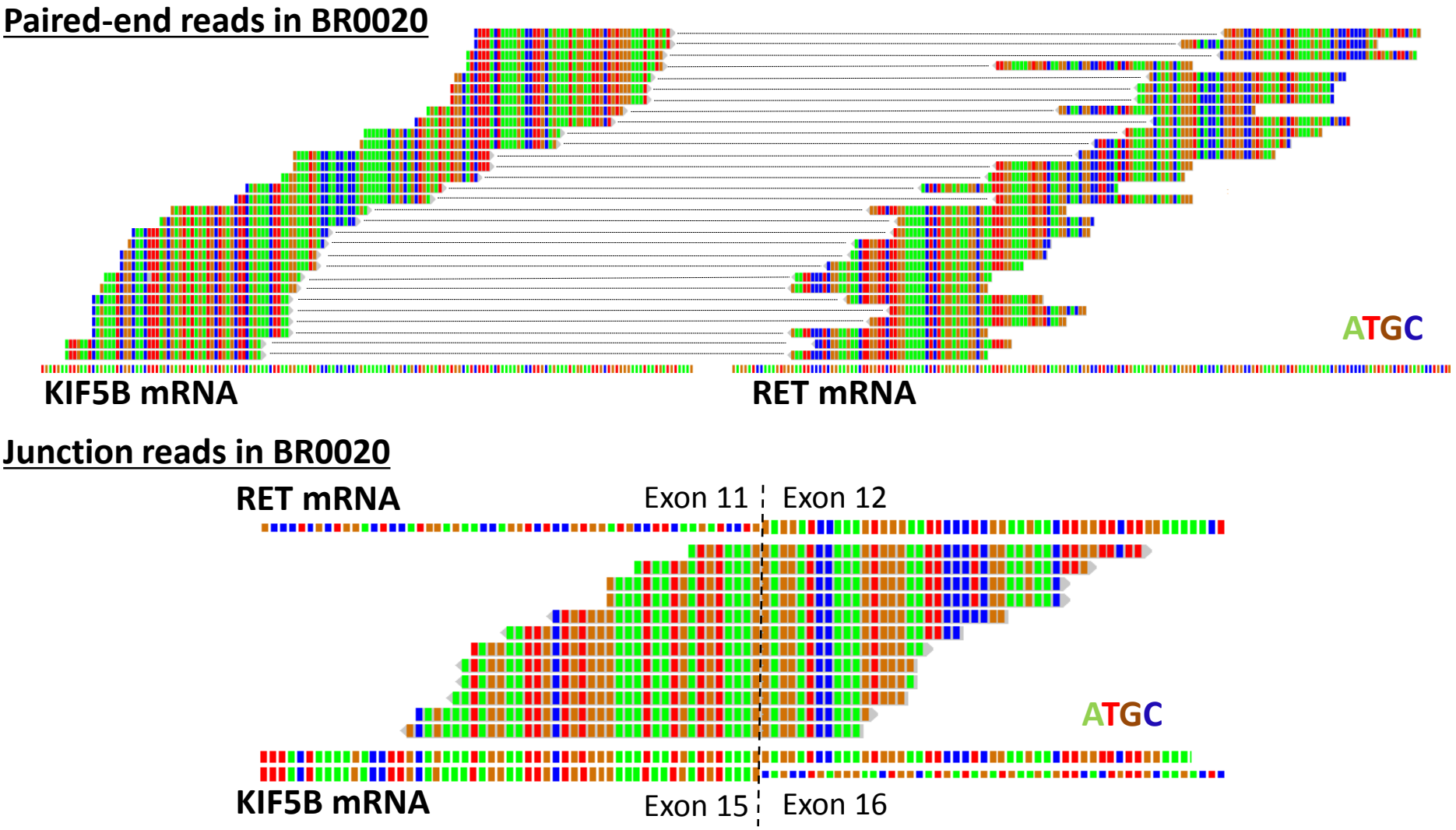

Figure S1. Detection of KIF5B-RET fusion. (a) Paired-end (upper) and junction (lower) reads for the *KIF5B-RET* fusion transcripts. Nucleotides are indicated by different colors. cDNA fragments 150–200 bp in size were subjected to paired-end sequencing (50 bp from both ends were sequenced). The junction reads span the junctions within the gene fusion.

Nature Medicine doi:10.1038/nm.2644

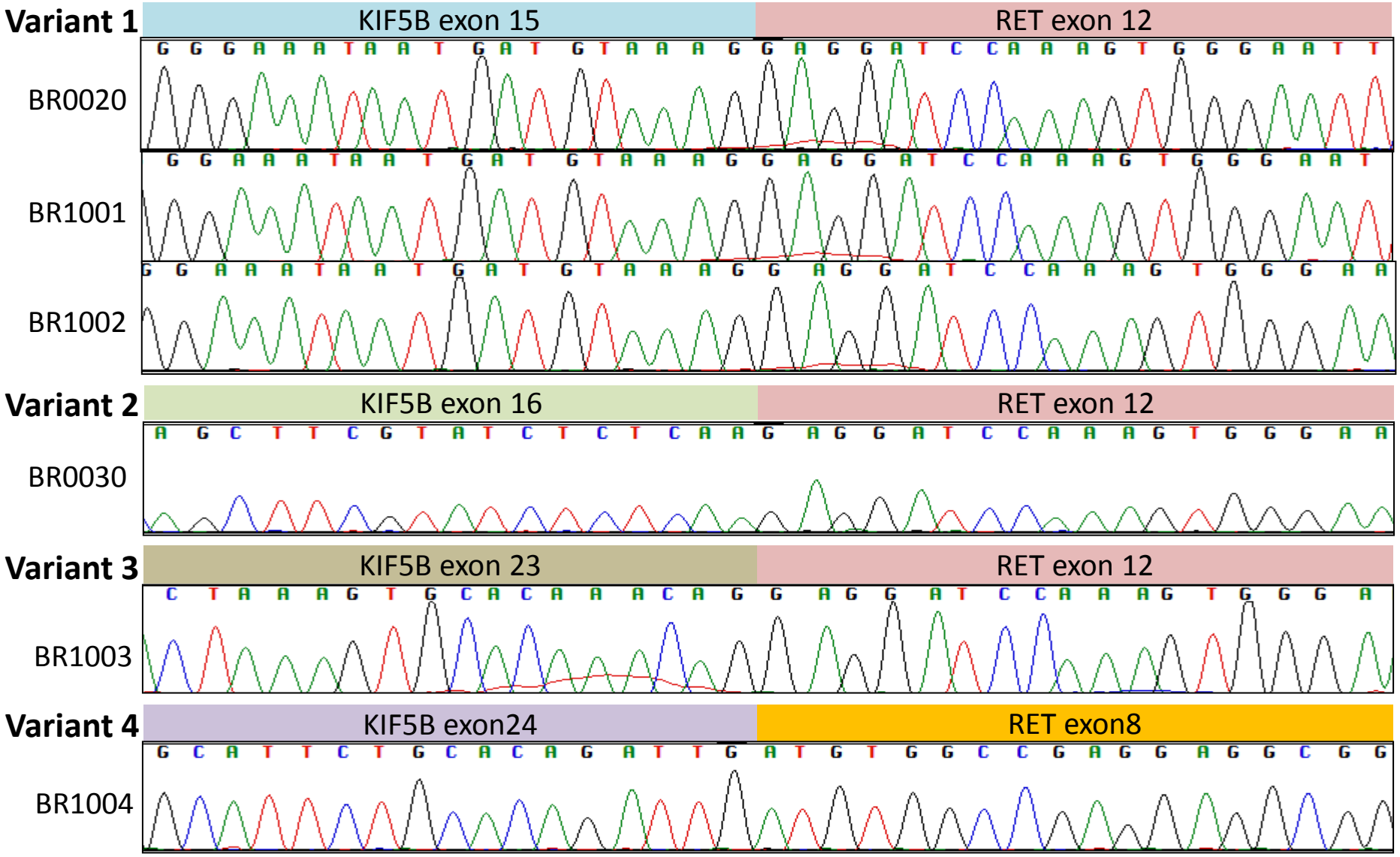

Figure S1. (b) Electrophoregram for Sanger sequencing of cDNAs for *KIF5B-RET* fusion transcripts. RT-PCR products amplified by the KIF5B-RET-F1 and KIF5B-RET-R1 primers were directly sequenced by the KIF5B-RET-F1 (BR0020, BR1001, BR1002 and BR0030) and KIF5B-F-orf2438 (BR1003 and BR1004) primers (Supplementary Table 6).

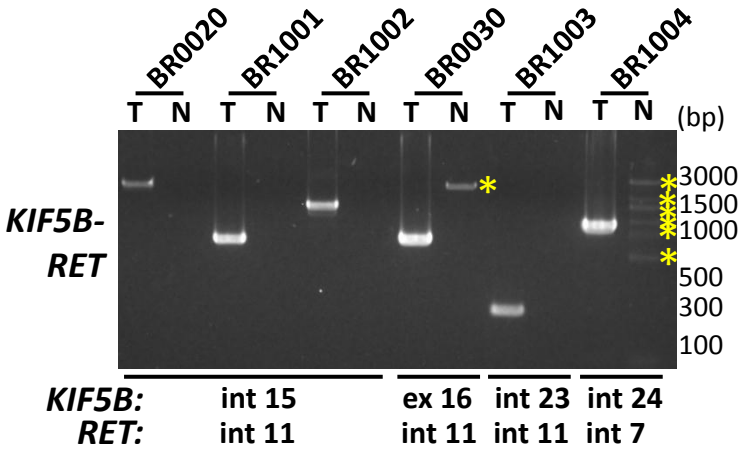

Figure S1. (c) Detection of *KIF5B-RET* fusions by genomic PCR. The locations of the PCR primers used to amplify the DNA fragments containing the breakpoint junctions are indicated below. int, intron; ex, exon. Non-specific bands observed in non-cancerous lung tissues from BR0030 and BR1004 are indicated by asterisks. These bands were detected in most non-cancerous and fusion-negative tumor samples. Direct sequencing of the PCR products confirmed that they were non-specific.

Nature Medicine doi:10.1038/nm.2644

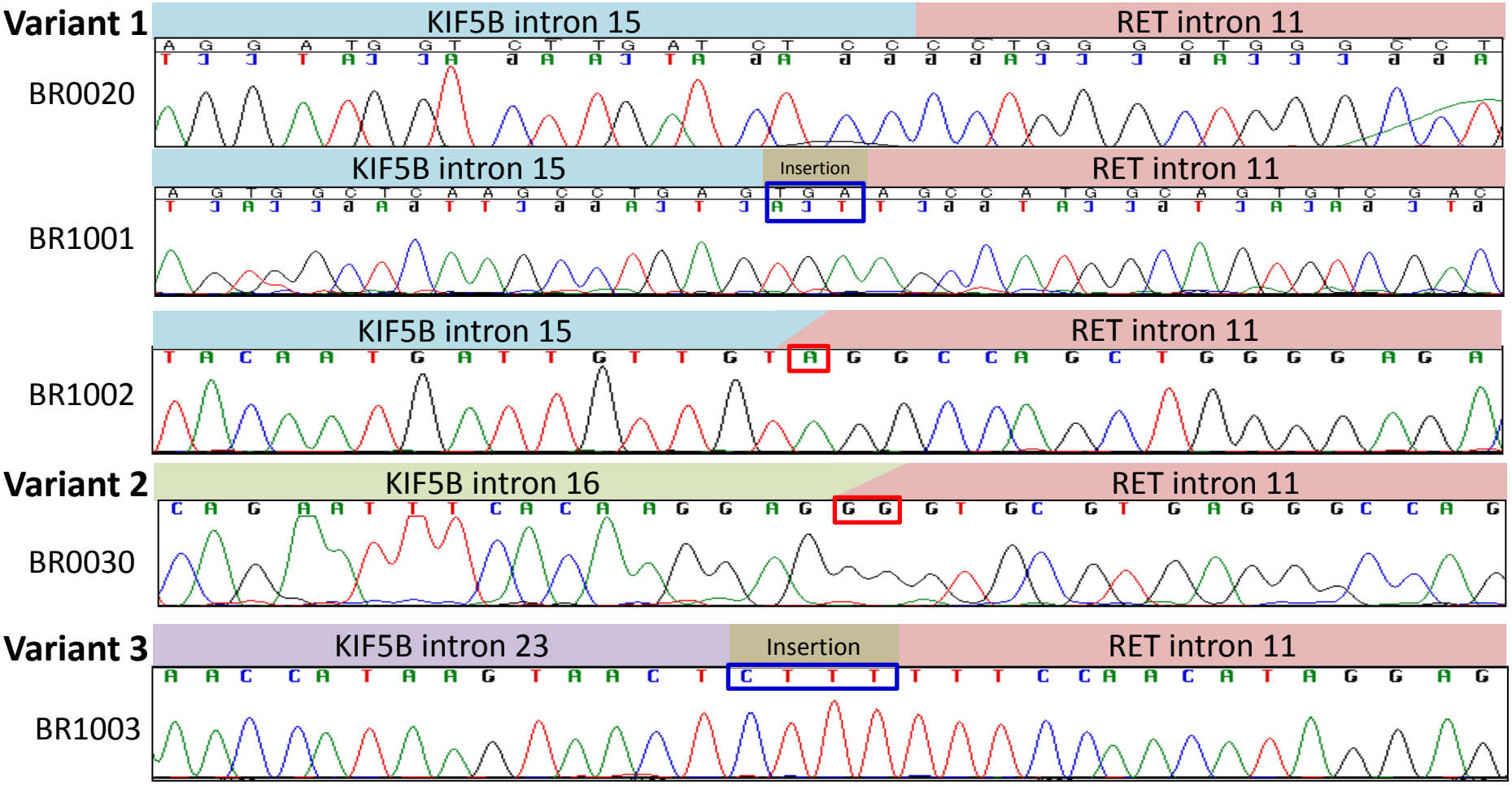

Figure S1. (d) Electrophoregram for Sanger sequencing of genomic fragments encompassing *KIF5B-RET* breakpoint junctions. PCR products were directly sequenced. Primers used for amplification and sequencing for each sample are as follow. BR0020: KIF5B-int15-F1/KIF5B-RET-R1 and RET-int11-R0.5, BR1001: KIF5B-int15-F1/KIF5B-RET-R1 and RET-int11-R1, BR1002: KIF5B-int15-F2/RET-int11-R3 and KIF5B-int15-F3.5, BR0030: KIF5B-ex16-F1/KIF5B-RET-R1 and KIF5B-ex16-F1; and BR1003: KIF5B-ex23-F1/KIF5B-RET-R1 and KIF5B-ex23-F1. Overlapped nucleotides and inserted nucleotides at breakpoint junctions are indicated by red and blue boxes.

Variant 4 BR1004

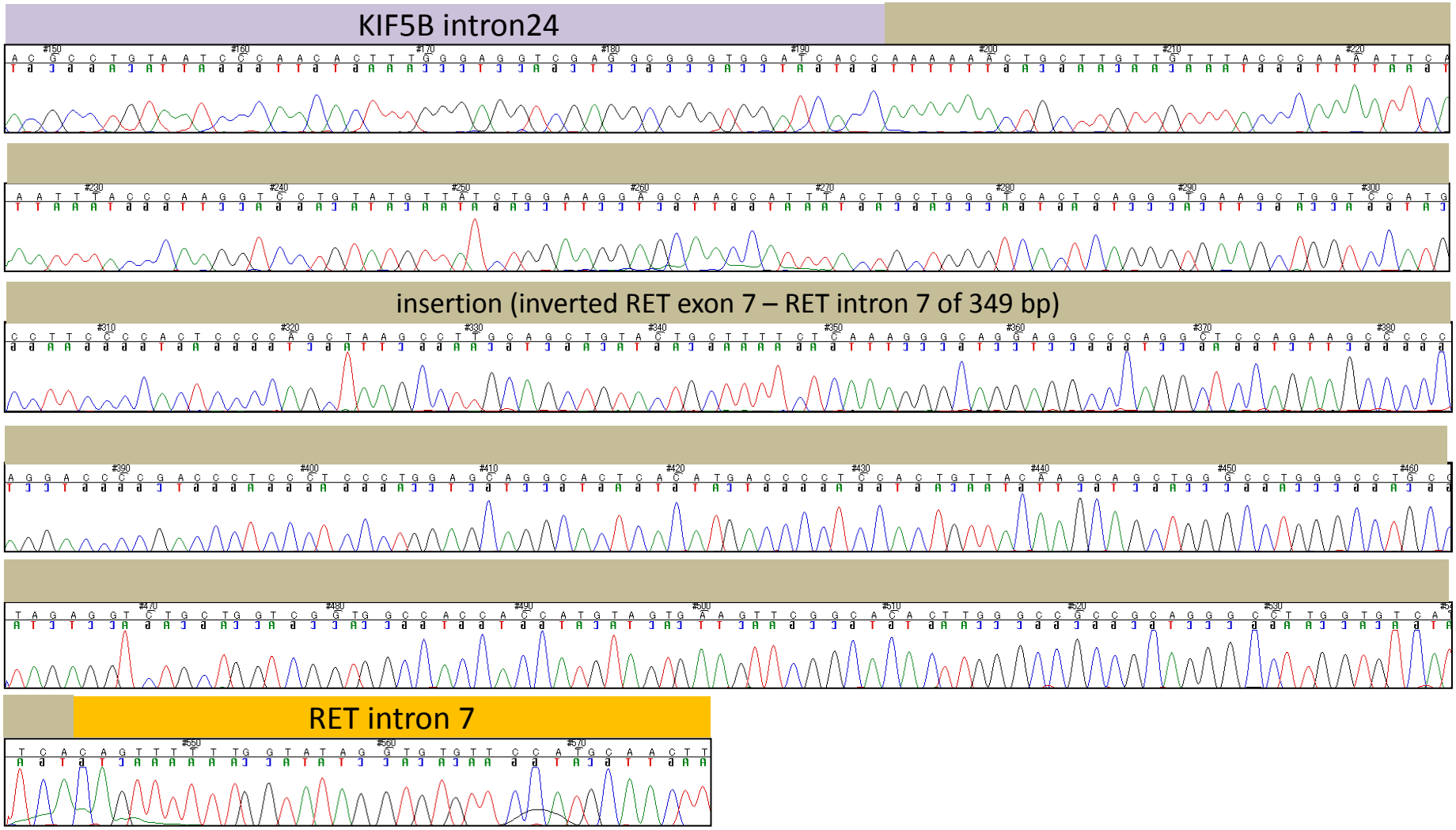

Figure S1. (d, continued) Electrophoregram for Sanger sequencing of genomic fragments encompassing the *KIF5B-RET* breakpoint junction of BR1004. PCR products amplified by the KIF5B-ex24-F1 and RET-int7-R1 primers were directly sequenced by the RET-int7-R2 primer. A genomic fragment of 349 bp encompassing RET exon 7 – RET intron 7 was inserted at the breakpoint junction.

Nature Medicine doi:10.1038/nm.2644

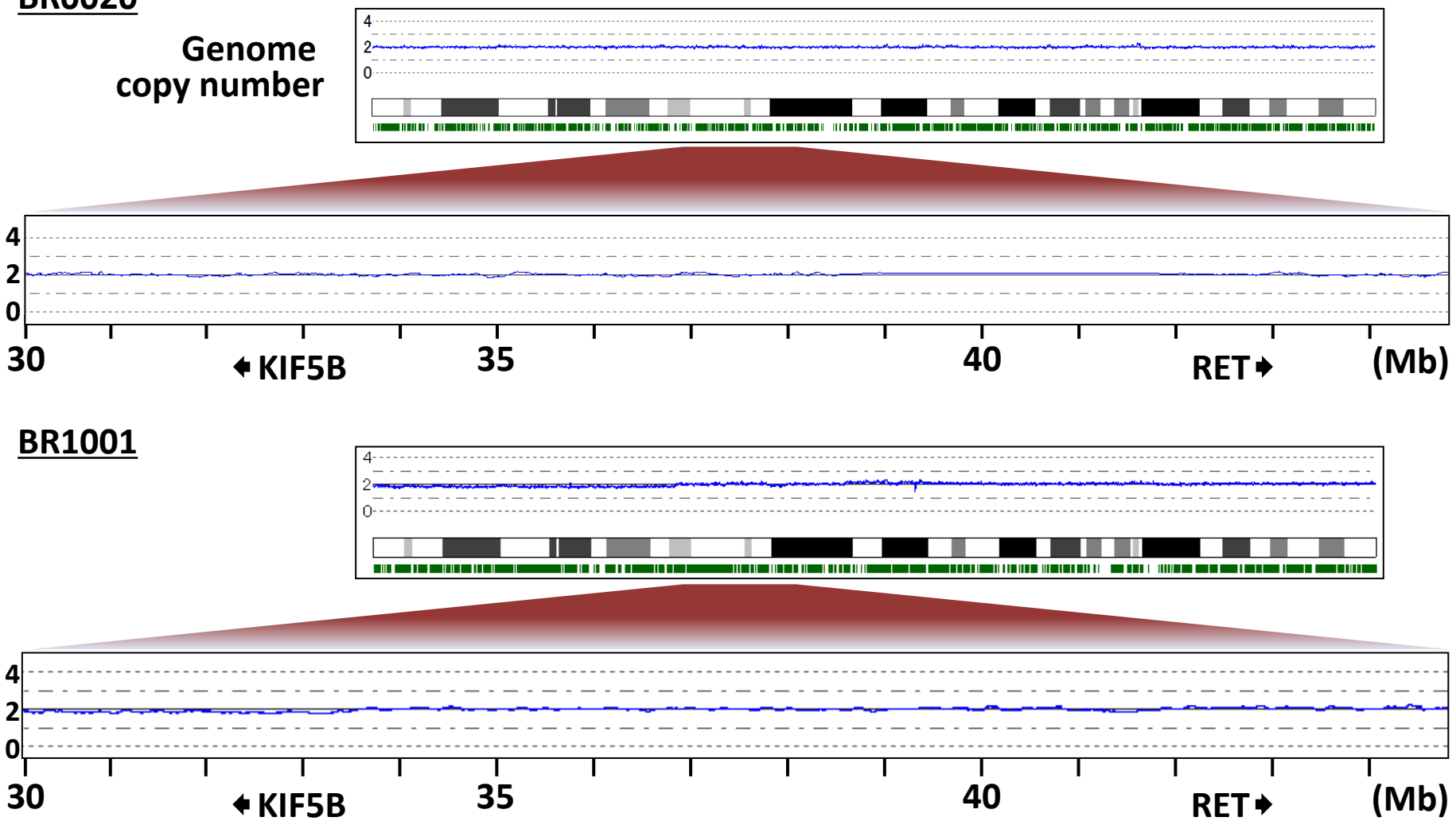

Figure S1. (e) The occurrence of chromosomal inversion was supported by the lack of genome copy number changes at these two loci. The genome copy number of chromosome 10 was examined in 4 of the 6 cases with *KIF5B-RET* fusion. Copy number was deduced by the CNAG program analysis. The location and direction of the *KIF5B* and *RET* genes on the reference genome are indicated by arrows.

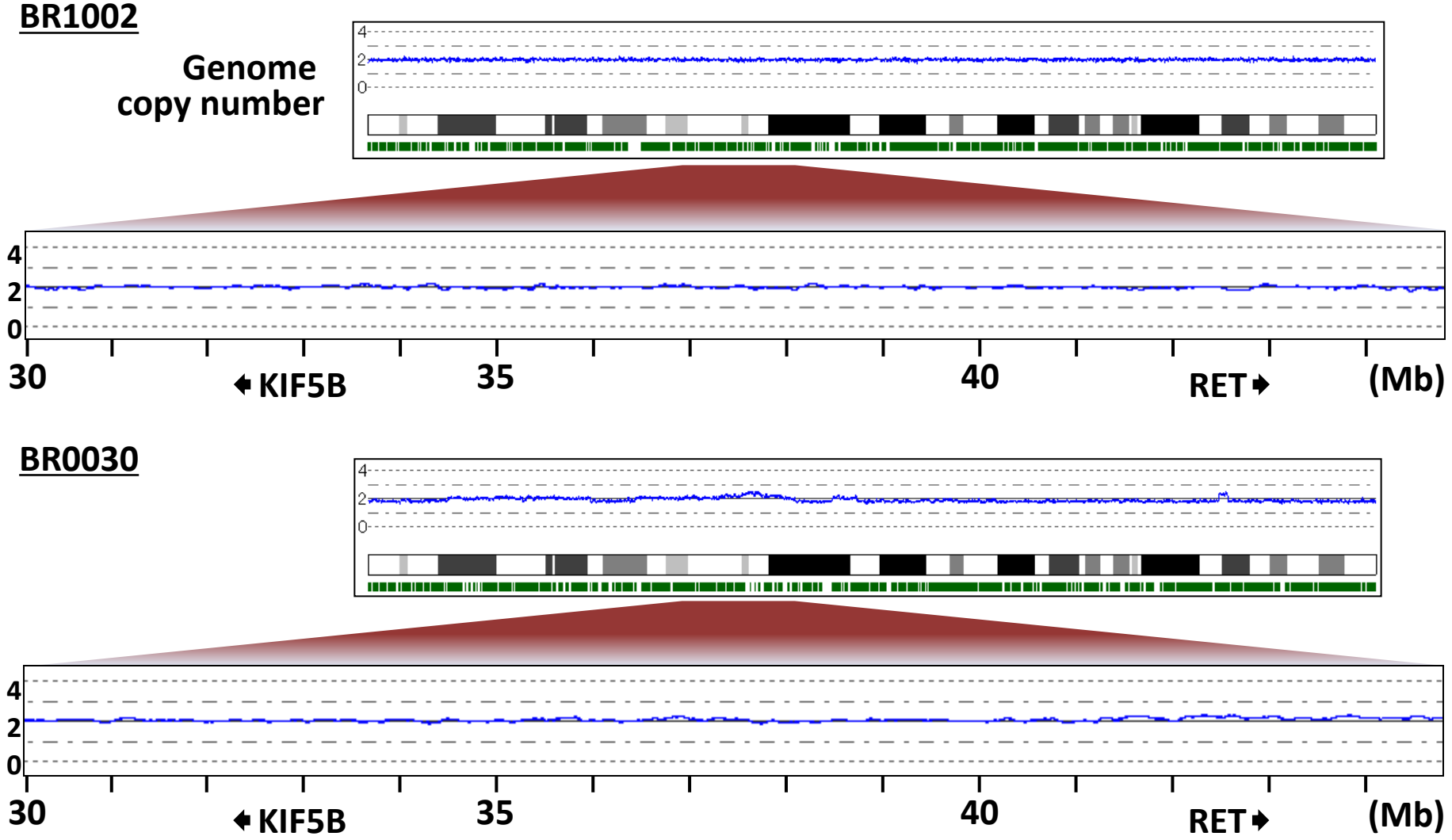

Figure S1. (e) (Continued)

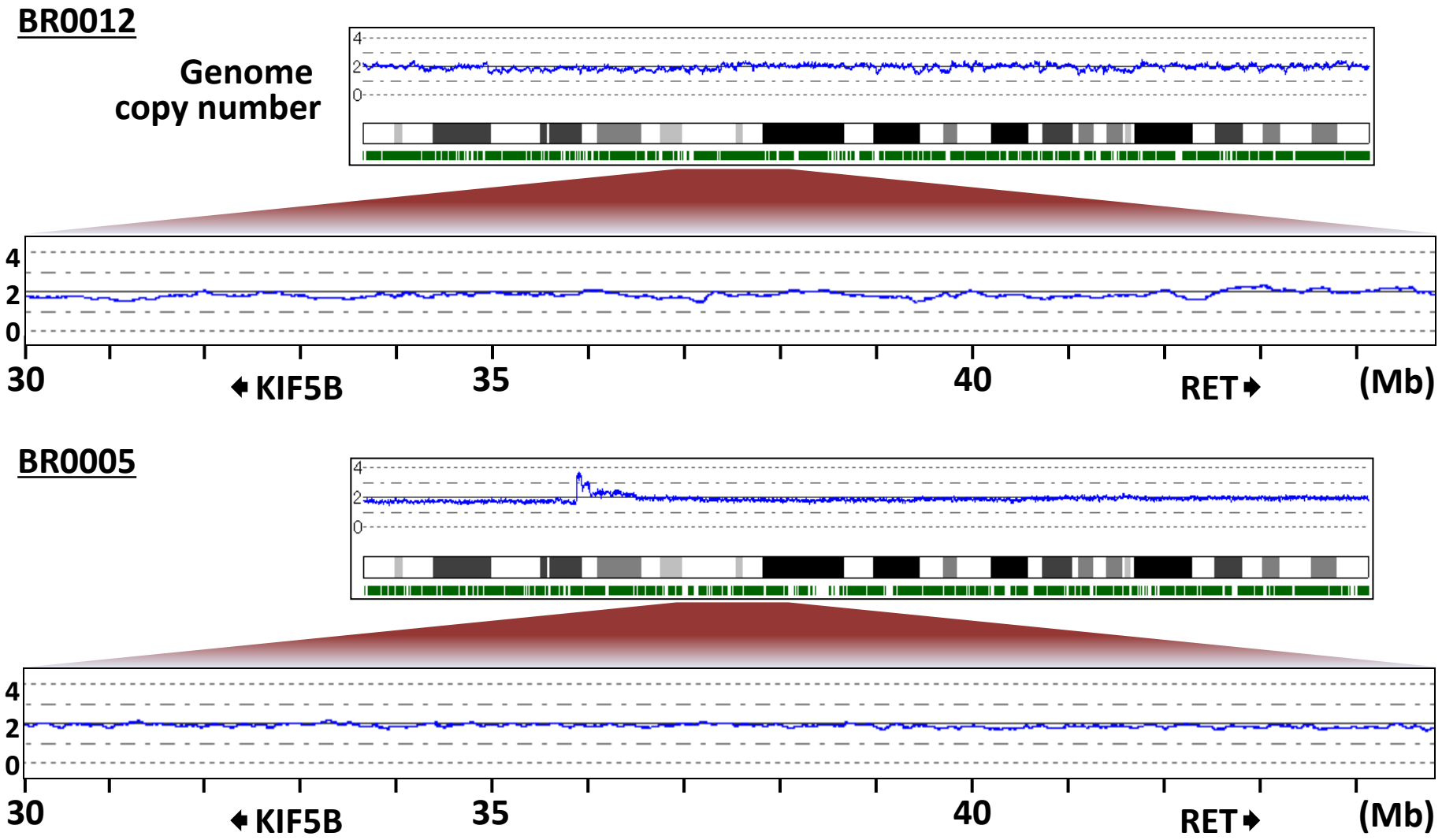

Figure S1. (f) The genome copy number of chromosome 10 was examined in two cases with high levels of *RET* expression but without *RET* fusion. Copy number was deduced by the CNAG program analysis. The location and direction of the *KIF5B* and *RET* genes are indicated by arrows.

Nature Medicine doi:10.1038/nm.2644

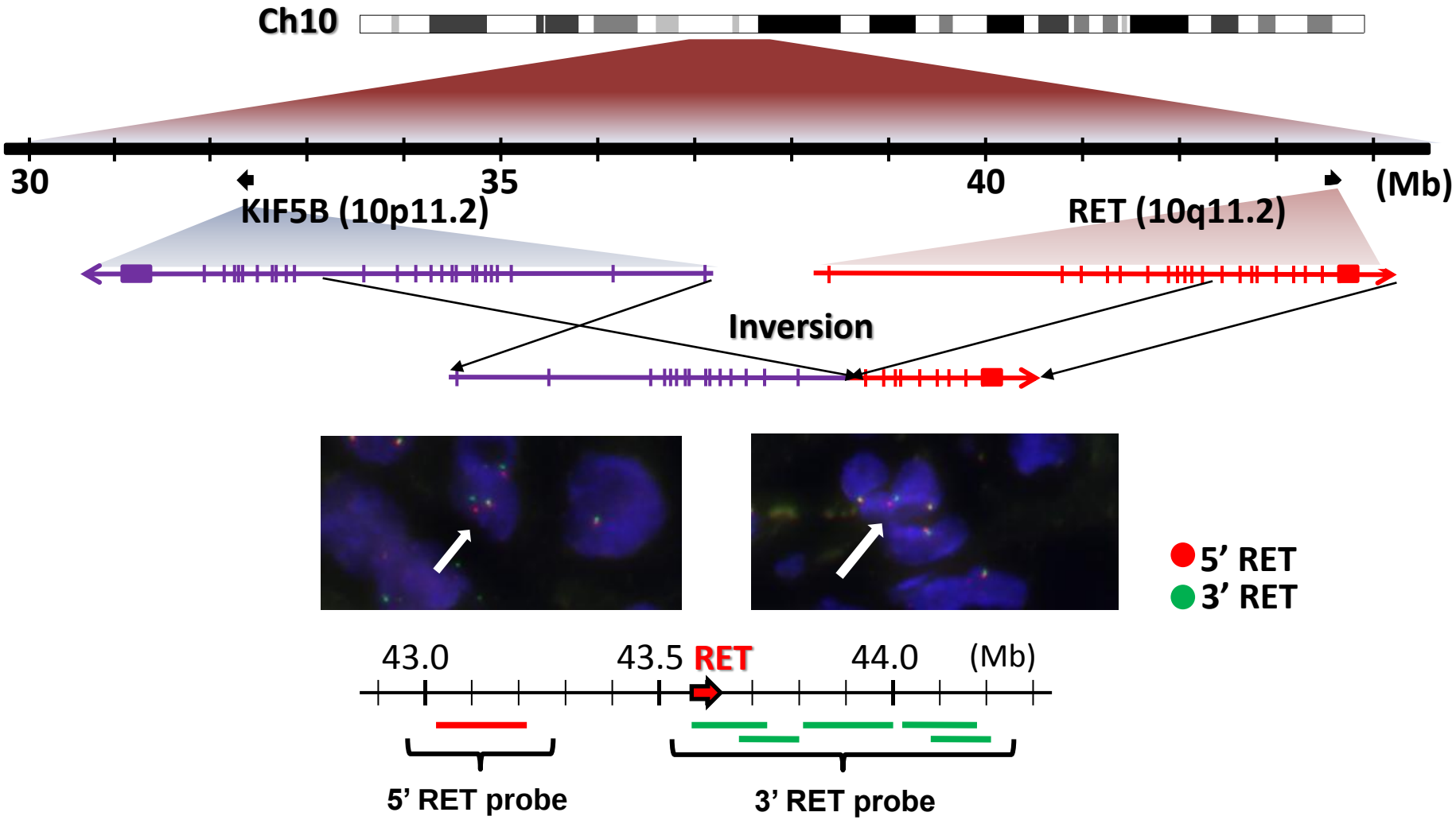

Figure S2. Chromosome aberration causing KIF5B-RET fusion. Chromosome inversion, a deduced chromosomal rearrangement responsible for KIF5B-RET fusion (variant 1, BR0020) (upper). Fluorescent in situ hybridization (FISH) reveals a split of red and green probes that flank the *RET* translocation site in a *KIF5B-RET* fusion positive tumor (arrows in middle). BAC clones used to make 5'- and 3'-RET probes are shown below.

Nature Medicine doi:10.1038/nm.2644

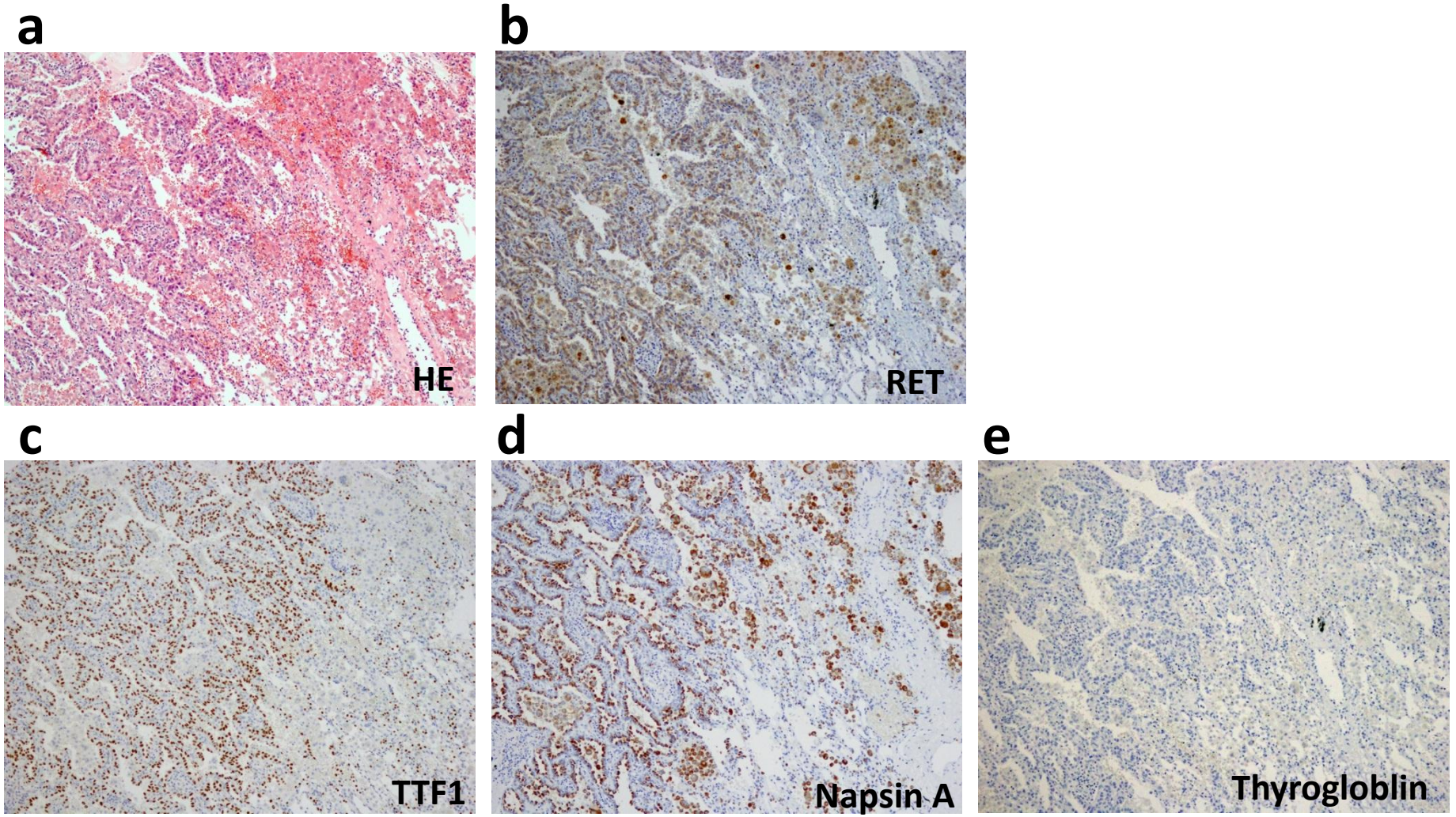

Figure S3. Histological characterization. Representative histology images showing KIF5B-RET fusion-positive LADC (BR1003). **(a)** Hematoxylin-eosin staining. Clara- or type II-like tumor cells are growing along the thickened alveolar walls and show lepidic or papillary growth (left half of image). The right half shows noncancerous lung cells. **(b)** Staining of tumor cells with an antibody recognizing the C-terminal region of RET shows a granular cytoplasmic pattern. **(c and d)** Nuclear staining of TTF1 and cytoplasmic staining of Napsin A in both cancerous and non-cancerous areas. **(e)** Thyroglobulin is negative in both in cancerous and non-cancerous areas. (Magnification ×10).

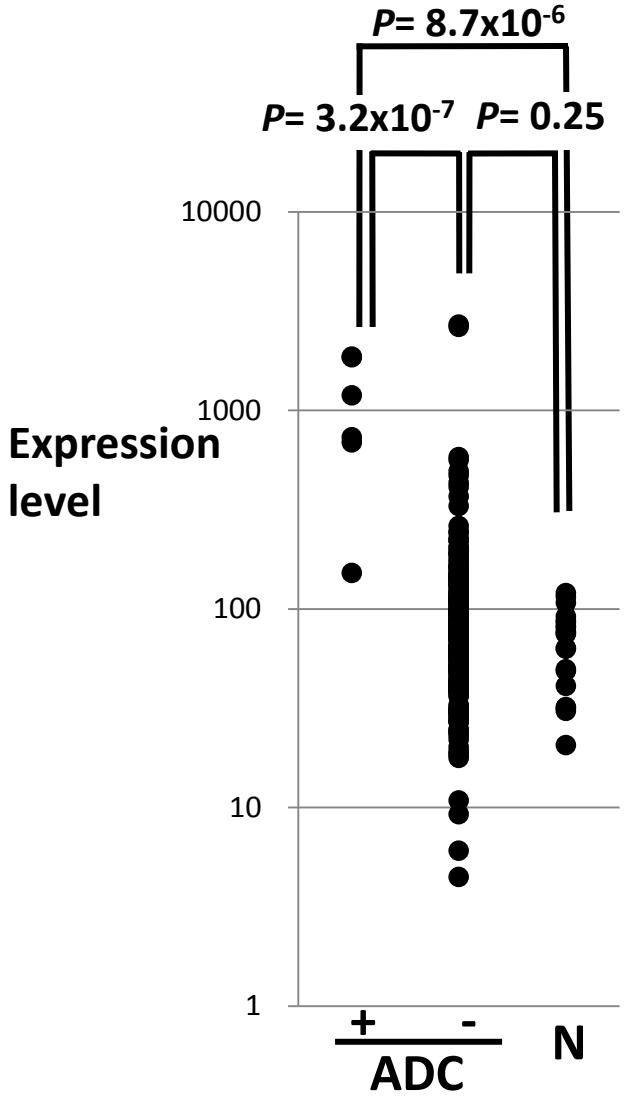

Figure S4. RET mRNA expression. Expression levels of *RET* in 228 LADCs and 20 non-cancerous lung tissues assessed by U133Aplus2.0 microarray analysis. Results of a RET probe (211421_s_at) are shown. P values by U-test for differences in expression levels are shown. *RET* expression in fusion-positive LADCs was higher than that in non-cancerous lung tissues and that in fusion-negative LADCs.

Nature Medicine doi:10.1038/nm.2644

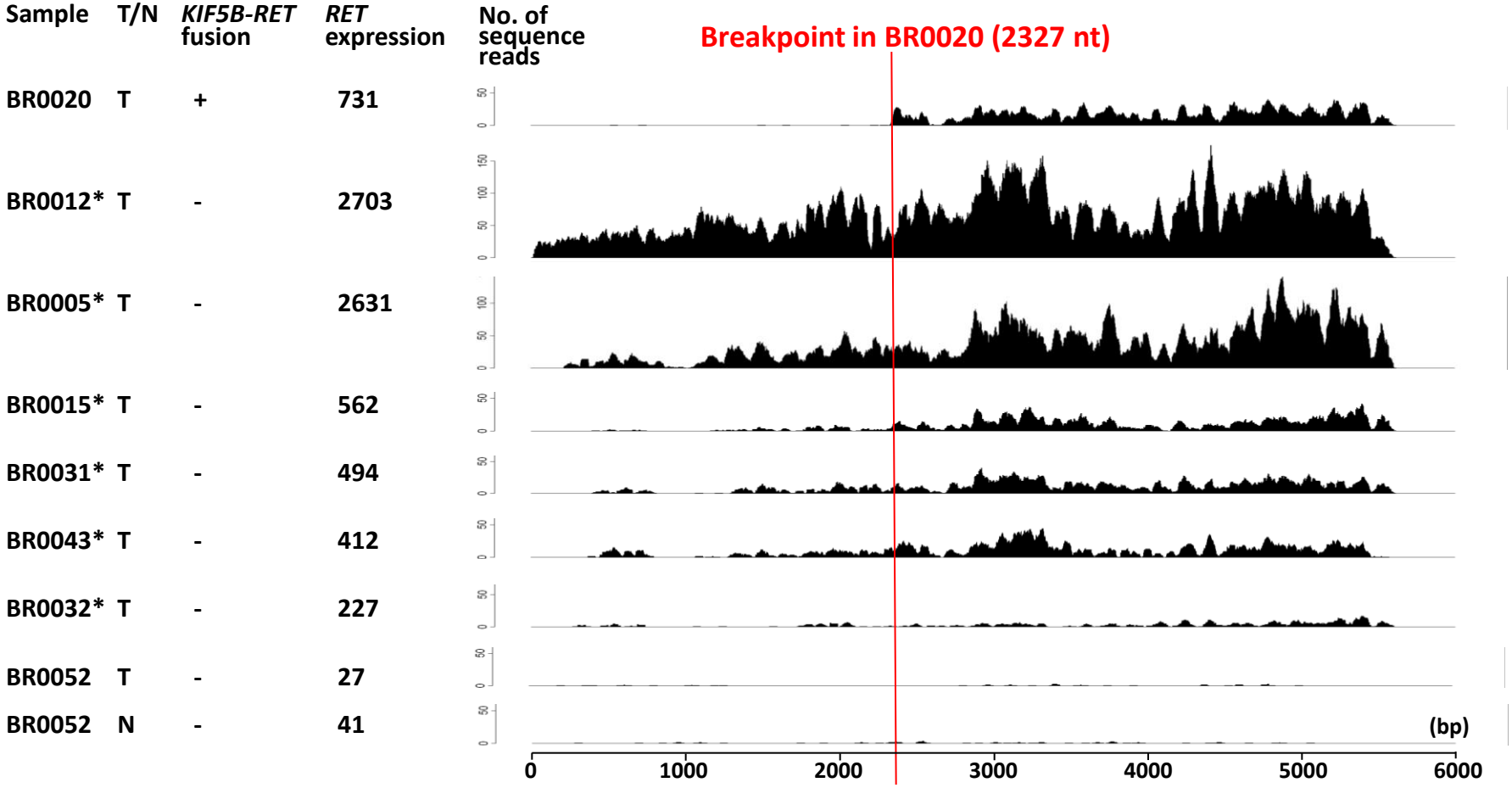

Figure S5. Distribution of sequence reads. Distribution of sequence reads from LADCs and a noncancerous lung tissue shown along the *RET* transcript (NM_020975.4). Sequence reads from a *KIF5B-RET* fusion positive sample, BR0020, were exclusively mapped downstream to the fusion point, while six samples (marked by *) without fusion but with *RET* expression showed sequence reads spanning over the whole *RET* transcript. No somatic and germline mutations were detected. Expression levels of *RET* assessed by an oligonucleotide microarray analysis are shown on the left.

Nature Medicine doi:10.1038/nm.2644

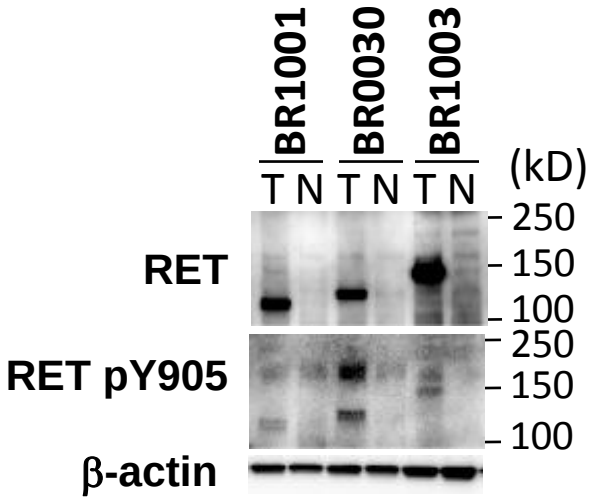

Figure S6. Expression and phosphorylation of KIF5B-RET proteins in lung adenocarcinoma specimens with *KIF5B-RET* fusions. Whole cell lysates of lung adenocarcinoma (T) and noncancerous lung tissues (N) were subjected to immunoblotting with anti-RET C-terminal and anti-phosphotyrosine 905 antibodies. To prepare whole-cell lysates, tissues were homogenized in a tissue homogenizer and stored at -80^{o}C. The cell homogenates were lysed with chilled IP150 buffer (50 mM Tris-HCl (pH 7.2), 150 mM NaCl, 2 mM MgCl2, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT, 1 mM Na3VO4, and 5 mM NaF) containing 0.1% NP40 and a protease inhibitor cocktail comprising 10 μg/ml pepstatin A, 10 μg/ml antipain, 10 μg/ml chymostatin, 10 μg/ml leupeptin, 10 μg/ml E-64, and 10 μg/ml PMSF. After 1 h on ice, the lysates were centrifuged for 15 min in a microcentrifuge at 4^{o}C and the supernatants were collected and boiled in SDS sample buffer.

Nature Medicine doi:10.1038/nm.2644

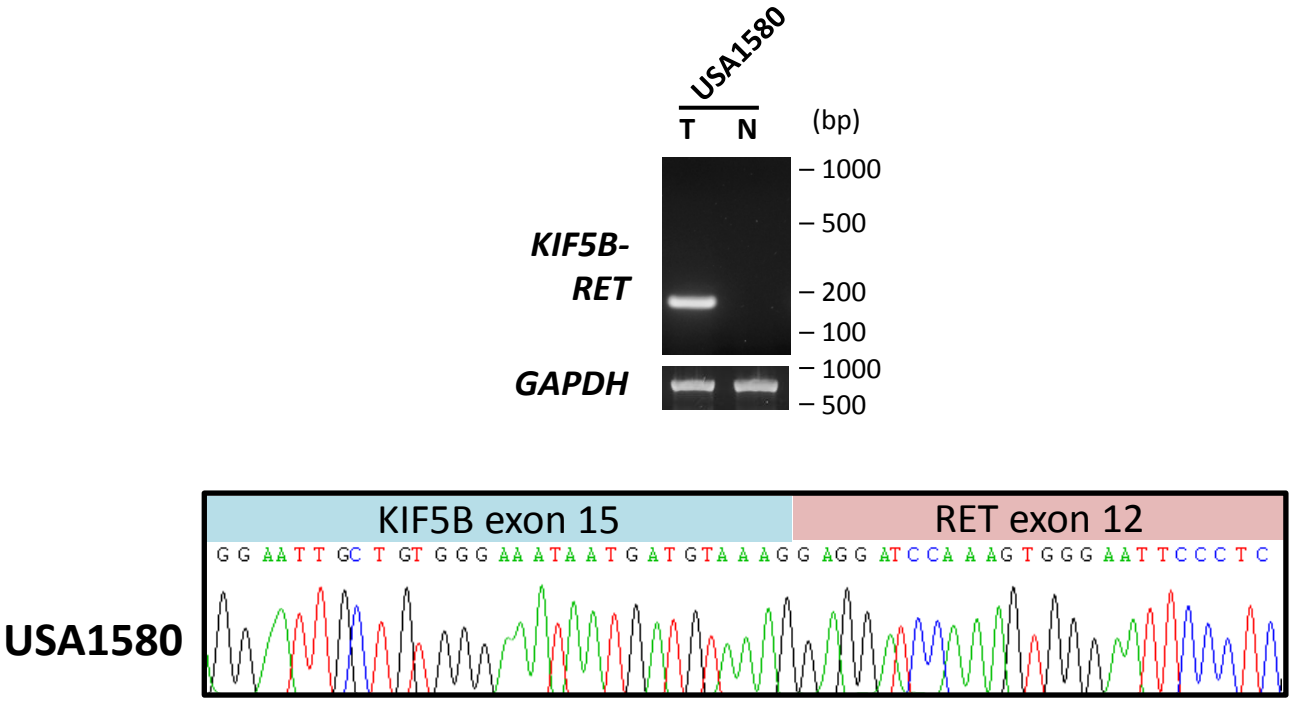

Figure S7. *KIF5B-RET* fusion in LADC of a non-Asian patient. Detection of *KIF5B-RET* fusion transcripts (variant 1) by RT-PCR in a LADC case of USA/European cohorts.

Nature Medicine doi:10.1038/nm.2644

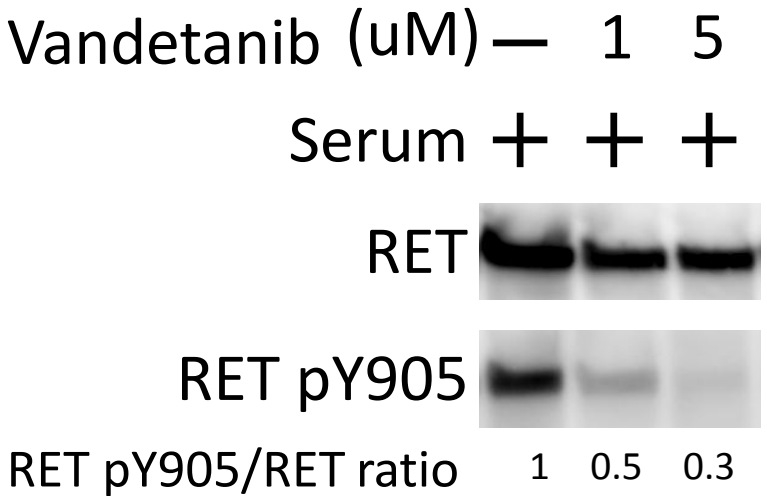

Figure S8. Suppression of RET kinase activity of KIF5B-RET protein by vandetanib. H1299 lung cancer cells were transfected for 24 hr with 0.8 ug of empty-vector, wild-type *RET* or *KIF5B-RET* expression plasmids with 10% FBS, and then serum starved for another 24 hr. After 2-hr treatment with DMSO; or 1 or 5 μM vandetanib, the cells were treated with 10% FBS for 10 min. Whole cell lysates were subjected to immunoblotting with anti-RET C-terminal and anti-phosphotyrosine 905 antibodies. Ratios of phosphoRET to RET signals with respect to wild-type RET after the serum treatment are shown below.

Nature Medicine doi:10.1038/nm.2644

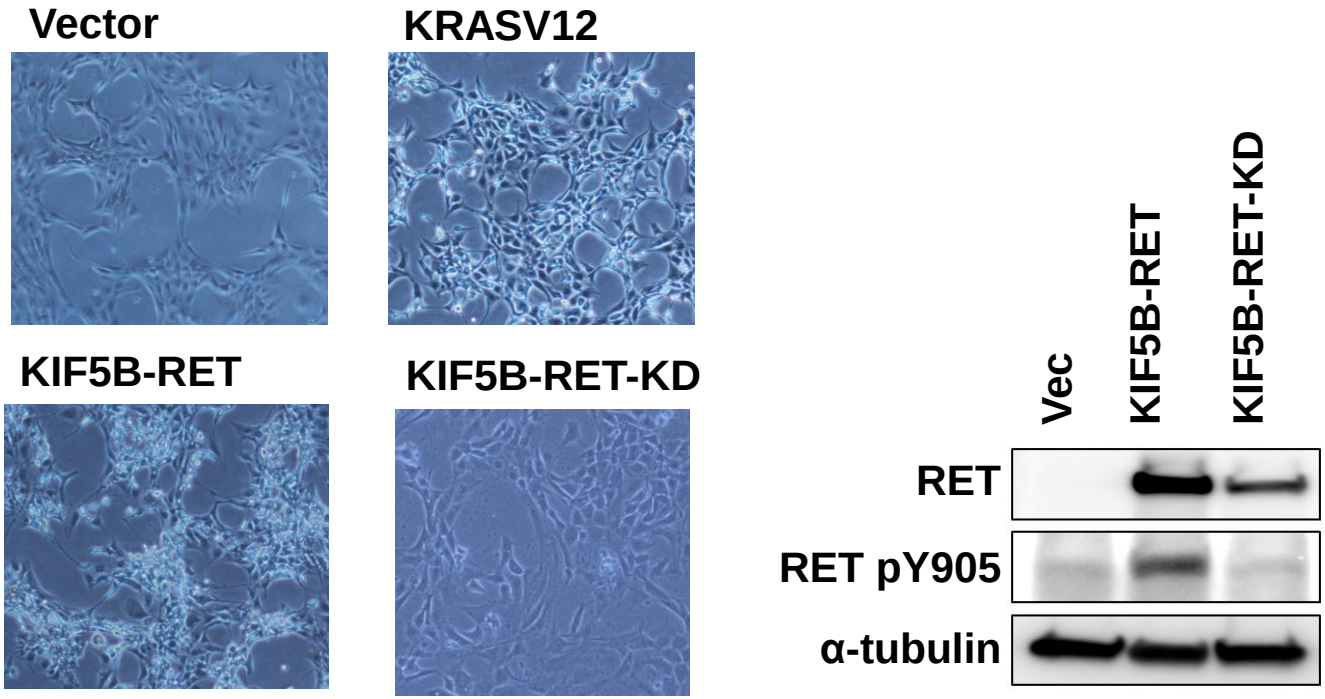

Figure S9. Transformation of NIH3T3 cells. (Left) Morphological transformation of NIH3T3 cells by the expression of exogenous *KRASV12*, *KIF5B-RET* or *KIF5B-RET (KD)*. Cells were photographed under a phase-contrast light microscope (× 150) under identical conditions. (Right) Reduced phosphorylation of the Y905 residue in kinase-dead mutants of the KIF5B-RET protein in NIH3T3 cells. Whole NIH 3T3 cell lysates from cells cultured with serum were subjected to immunoblotting with anti-RET C-terminal and anti-phosphotyrosine 905 antibodies.

Nature Medicine doi:10.1038/nm.2644