Multiplex Reverse Transcription-PCR Screening for *EML4-ALK* Fusion Transcripts

Kengo Takeuchi,[1] Young Lim Choi,[3] Manabu Soda,[3] Kentaro Inamura,[1] Yuki Togashi,[1] Satoko Hatano,[1] Munehiro Enomoto,[3] Shuji Takada,[3] Yoshihiro Yamashita,[3] Yukitoshi Satoh,[2] Sakae Okumura,[2] Ken Nakagawa,[2] Yuichi Ishikawa,[1] and Hiroyuki Mano[3,4]

Abstract

Purpose: EML4-ALK is a fusion-type protein tyrosine kinase that is generated by inv(2)(p21p23) in the genome of non–small cell lung cancer (NSCLC). To allow sensitive detection of *EML4-ALK* fusion transcripts, we have now developed a multiplex reverse transcription-PCR (RT-PCR) system that captures all in-frame fusions between the two genes.

Experimental Design: Primers were designed to detect all possible in-frame fusions of *EML4* to exon 20 of *ALK*, and a single-tube multiplex RT-PCR assay was done with total RNA from 656 solid tumors of the lung ($n = 364$) and 10 other organs.

Results: From consecutive lung adenocarcinoma cases ($n = 253$), we identified 11 specimens (4.35%) positive for fusion transcripts, 9 of which were positive for the previously identified variants 1, 2, and 3. The remaining two specimens harbored novel transcript isoforms in which exon 14 (variant 4) or exon 2 (variant 5) of *EML4* was connected to exon 20 of *ALK*. No fusion transcripts were detected for other types of lung cancer ($n = 111$) or for tumors from 10 other organs ($n = 292$). Genomic rearrangements responsible for the fusion events in NSCLC cells were confirmed by genomic PCR analysis and fluorescence *in situ* hybridization. The novel isoforms of EML4-ALK manifested marked oncogenic activity, and they yielded a pattern of cytoplasmic staining with fine granular foci in immunohistochemical analysis of NSCLC specimens.

Conclusions: These data reinforce the importance of accurate diagnosis of EML4-ALK–positive tumors for the optimization of treatment strategies.

Authors' Affiliations: [1]Division of Pathology, The Cancer Institute, [2]Department of Thoracic Surgical Oncology, Thoracic Center, Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo, Japan; [3]Division of Functional Genomics, Jichi Medical University, Tochigi, Japan; and [4]CREST, Japan Science and Technology Agency, Saitama, Japan

Received 4/19/08; revised 6/23/08; accepted 7/3/08.

Grant support: Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and grants from the Japan Society for the Promotion of Science; from the Ministry of Health, Labor, and Welfare of Japan; from the National Institute of Biomedical Innovation; from the Smoking Research Foundation; and from the Vehicle Racing Commemorative Foundation.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

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

K. Takeuchi and Y.L. Choi contributed equally to this work.

Current address for Y. Satoh: Department of Thoracic Surgery, Kitasato University School of Medicine, Kanagawa 228-8520, Japan.

The nucleotide sequences of the *EML4-ALK* variant 4, 5a, and 5b cDNAs have been deposited in DDBJ/EMBL/Genbank under the accession numbers AB374363, AB374364, and AB374365, respectively.

Requests for reprints: Kengo Takeuchi, Division of Pathology, The Cancer Institute, Japanese Foundation for Cancer Research, Tokyo 135-8550, Japan. Phone: 81-3-3520-0111; Fax: 81-3-3570-0558; E-mail: kentakeuchi-tky@umin.net.

doi:10.1158/1078-0432.CCR-08-1018

Chromosome rearrangement is a major mechanism giving rise to transforming potential in human cancers, especially in hematologic malignancies (1). A balanced translocation between chromosomes 9 and 22, for instance, generates an activated protein tyrosine kinase, BCR-ABL, that plays an essential role in the pathogenesis of chronic myeloid leukemia (2). The gene for another protein tyrosine kinase, ALK, is fused to those for NPM1 or other partner proteins in anaplastic lymphoma and soft tissue tumors, resulting in an increase in the kinase activity of ALK (3).

Mitelman et al. have suggested that chromosome translocations, in addition to being common in hematologic malignancies, are not rare in epithelial tumors (4, 5). These researchers also proposed that the genetic mechanisms underlying oncogenesis might not differ fundamentally between hematologic and epithelial malignancies, and that the current apparent difference in the frequency of chromosomal translocations between these two types of cancer is likely to disappear with the advent of new and more powerful investigative tools.

Consistent with this notion, recurrent chromosome rearrangements involving genes for ETS transcriptional factors have been identified in many cases of prostate cancer and may contribute to the hypersensitivity of prostate cancer cells to androgens (6, 7). In addition, we recently discovered another

Translational Relevance

EML4-ALK is a fusion-type protein-tyrosine kinase generated through a recurrent chromosome rearrangement, inv(2)(p21p23), observed in non–small cell lung cancer (NSCLC). Because both *EML4* and *ALK* genes are mapped to the short arm of chromosome 2 in opposite orientations, PCR with primer sets flanking the fusion points of the two genes would not produce any specific products from cells without inv(2)(p21p23). Reverse transcription (RT)-PCR for the fusion point would, therefore, become a highly sensitive and accurate means to detect tumors positive for *EML4-ALK*. Such analyses may detect small amounts of cancer cells in sputa from individuals with NSCLC at early clinical stages. Because several isoforms have been already reported for *EML4-ALK*, it is mandatory to detect all isoforms of the fusion kinase in a sensitive and reliable way. Toward this goal, we here developed a single-tube multiplex RT-PCR screening system to capture all possible isoforms of EML4-ALK. Examination of various tumor samples (n = 656) with our multiplex RT-PCR has indeed identified 11 specimens positive for the variants of EML4-ALK only among lung adenocarcinoma (n = 253). Our system, thus, paves a way for a sensitive molecular detection of this intractable disorder at early curable stages.

recurrent chromosome translocation in non–small cell lung cancer (NSCLC; ref. 8), a major cause of cancer deaths in humans. A small inversion within the short arm of chromosome 2, inv(2)(p21p23), was found to be present in <10% of NSCLC cases and to give rise to a novel fusion-type tyrosine kinase, EML4-ALK, that exhibited marked transforming activity *in vitro* (8). Transgenic mice that specifically express EML4-ALK in lung epithelial cells were also found to develop hundreds of adenocarcinoma nodules in both lungs at only a few weeks after birth, and such nodules disappeared rapidly in response to oral administration of a specific inhibitor of the catalytic activity of ALK.[5] These data thus indicate that EML4-ALK plays a pivotal role in malignant transformation in lung cancer, and they suggest that chemical compounds that inhibit the tyrosine kinase activity of EML4-ALK may provide an effective treatment for EML4-ALK–positive lung cancer. The selection of suitable drugs for individuals with lung cancer will thus require accurate determination of the absence or presence of the *EML4-ALK* fusion gene in biopsy specimens.

Given that *EML4* and *ALK* map in opposite orientations within the short arm of chromosome 2, reverse transcription-PCR (RT-PCR) analysis with primers designed to amplify the fusion points of *EML4-ALK* transcripts would not be expected to yield specific products from normal cells or cancer cells without inv(2)(p21p23). Such analysis should thus provide a highly reliable and sensitive means to detect *EML4-ALK* in clinical specimens. Given that sputum has been shown to be a suitable specimen for such molecular diagnosis of *EML4-ALK* positivity (8), detection of *EML4-ALK*–positive cells by RT-PCR analysis of sputa may be effective for the identification of lung cancer at early clinical stages. The accurate diagnosis of *EML4-ALK*–positive tumors, however, will require that all isoforms of *EML4-ALK* are detected.

The fusion of intron 13 or 20 of *EML4* to intron 19 of *ALK* gives rise to variant 1 or 2 of *EML4-ALK*, respectively (8). We have recently discovered another isoform (variant 3) of *EML4-ALK* in which intron 6 of *EML4* is ligated to intron 19 of *ALK* (9). Theoretically, in addition to such fusion of exons 6, 13, and 20 of *EML4*, an in-frame fusion to exon 20 of *ALK* can occur with exons 2, 18, or 21 of *EML4*. Given that the amino-terminal coiled-coil domain of EML4 is responsible for the dimerization and constitutive activation of EML4-ALK (8) and that exon 2 of *EML4* encodes the entire coiled-coil domain, all of these possible fusion genes would encode EML4-ALK proteins containing the coiled-coil domain and therefore likely produce oncogenic EML4-ALK kinases.

To establish a highly sensitive and accurate PCR-based screening system for *EML4-ALK*–positive cancer, we have now developed a high-throughput multiplex RT-PCR assay for the detection of all potential *EML4-ALK* in-frame fusion transcripts. Among a consecutive series of lung adenocarcinoma specimens (n = 253) as well as other solid tumor samples (n = 403), we have now identified a total of 11 lung adenocarcinoma specimens positive for *EML4-ALK*, two of which harbor previously unidentified fusion mRNAs.

Materials and Methods

Clinical samples and RNA extraction. This study was done with clinical samples from 253 lung adenocarcinomas, 90 other NSCLCs (71 squamous cell carcinomas, 7 adenosquamous carcinomas, 7 large cell carcinomas, 2 pleomorphic carcinomas, and 3 large cell endocrine carcinomas), 21 small cell lung carcinomas, 50 breast carcinomas, 46 renal cell carcinomas, 48 colon carcinomas, 13 prostate carcinomas, 29 urothelial carcinomas, 33 gastric carcinomas, 10 uterine carcinomas, 9 hepatocellular carcinomas, 8 pancreatic carcinomas, and 46 malignant fibrous histiocytomas. All specimens were collected with the approval of the ethical committee at the Cancer Institute Hospital (Tokyo, Japan) and with the informed consent of individuals undergoing surgery from May 1995 to July 2003. The NSCLC cases were consecutive and spanned a period of 19 mo. Histologic diagnosis of NSCLC was made according to the WHO classification (10). All lesions were grossly dissected, rapidly frozen in liquid nitrogen, and stored at -80°C until RNA extraction with an RNeasy Mini Kit (Qiagen). RNA quality and the absence of contamination with genomic DNA were verified by formaldehyde-agarose gel electrophoresis.

Multiplex RT-PCR analysis and nucleotide sequencing. Total RNA was subjected to RT with random primers and SuperScript III reverse transcriptase (Invitrogen). For detection of *EML4-ALK* fusion cDNAs, multiplex PCR analysis was done with AmpliTaq Gold DNA polymerase (Applied Biosystems), the forward primers EML4 72F (5′-GTCAGCTCTTGAGTCACGAGTT-3′) and Fusion-RT-S (5′-GTGCAGTGTTTAGCATTCTTGGGG-3′), and the reverse primer ALK 3078RR (5′-ATCCAGTTCGTCCTGTTCAGAGC-3′). The *GAPDH* cDNA was amplified by PCR with the primers 5′-GTCAGTGGTGGACCT-GACCT-3′ and 5′-TGAGCTTGACAAAGTGGTCG-3′. For amplification of *EML4-ALK* fusion cDNAs, the samples were incubated at 94°C for 10 min and then subjected to 35 cycles of denaturation at 94°C for 1 min, annealing at 64°C for 1 min, and polymerization at 72°C for 1 min. For amplification of *GAPDH* cDNA, the samples were subjected to 35 cycles of 94°C for 1 min, 58°C for 30 s, and 72°C for 30 s. Virtual gel electrophoresis of multiplex RT-PCR products was done with a 2100 Bioanalyzer (Agilent Technologies).

[5] M. Soda et al., submitted for publication.

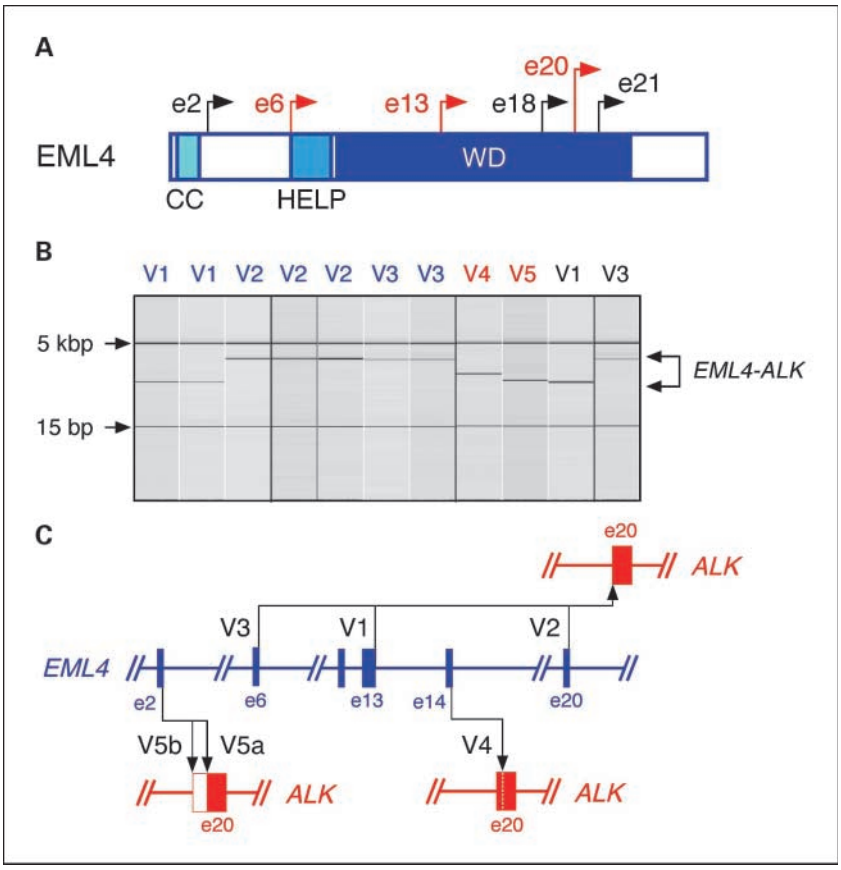

Fig. 1. Identification of *EML4-ALK* variants 4 and 5. *A,* schematic representation of the structure of EML4. The corresponding positions of exons (*e*) that can theoretically be fused in-frame to exon 20 of *ALK* are indicated by arrows, with known fusion points being denoted in red. CC, coiled-coil domain; HELP, hydrophobic EMAP (echinoderm microtubule-associated protein) – like protein domain; WD, WD repeats. *B,* virtual gel electrophoresis of multiplex RT-PCR products derived from lung adenocarcinoma specimens. Seven samples (*blue*) were known to harbor *EML-ALK* variants (*V*) 1, 2, or 3, whereas four samples were newly detected by multiplex RT-PCR. Two of the latter four specimens yielded PCR products corresponding to the newly identified variants 4 and 5. The positions of the fusion products of *EML4-ALK* are indicated on the right, and those of DNA size standards (5 kbp and 15 bp) are shown on the left. *C,* fusions between exons of *EML4* and *ALK*. Fusion of exons 6, 13, or 20 of *EML4* to exon 20 of *ALK* gives rise to variants 3, 1, and 2 of *EML4-ALK*, respectively. In addition, nucleotide sequencing of the PCR products shown in *B* revealed that exon 14 or 2 of *EML4* was fused to exon 20 of *ALK* in the cDNAs for *EML4-ALK* variants 4 and 5, respectively.

The primers used for direct amplification of the fusion points of individual cDNAs were 5′-AGGAGAGAACTCAGCGACACTACC-3′ and 5′-TCCACGCTCAAAAGTGCCAAGTCC-3′ for variant 4 and 5′-GCTTTCCCCGCAAGATGGACGG-3′ and 5′-AGCTTGCTCAGCTTG-TACTCAGGG-3′ for variant 5. Full-length cDNAs for *EML4-ALK* variants were amplified with PrimeSTAR DNA polymerase (Takara Bio) and the primers 5′-ACTCTGTCGGTCCGCTGAATGAAG-3′ and 5′-CCACGGTCTTAGGGATCCCAAGG-3′.

Fluorescence* in situ *hybridization analysis. Surgically resected lung cancer tissue was fixed in 20% formalin, embedded in paraffin, sectioned at a thickness of 4 μm, and placed on glass slides. The unstained sections were processed with a Histology FISH Accessory Kit (Dako), subjected to hybridization with fluorescently labeled bacterial artificial chromosome clone probes for *EML4* and *ALK* (GSP Laboratory) or for genomic regions upstream and downstream of the *ALK* break point (Dako), stained with 4,6-diamidino-2-phenylindole, and examined with a fluorescence microscope (BX51; Olympus).

Immunohistochemical analysis. Unstained paraffin-embedded sections were depleted of paraffin with xylene, rehydrated with a graded series of ethanol solutions, and then subjected to heat-induced antigen retrieval with Target Retrieval Solution pH 9.0 (Dako) before immunohistochemical staining with a mouse monoclonal antibody to ALK (ALK1, Dako) at a dilution of 1:20. Immune complexes were detected with the use of an EnVision+DAB system (Dako) with minor modifications.[6]

Transforming potential of EML4-ALK proteins. Protein analysis of EML4-ALK variants was done as described previously (8). In brief, the *EML4-ALK* variant 4, 5a, or 5b cDNAs were fused with an oligonucleotide encoding the FLAG epitope tag and inserted into the retroviral expression plasmid pMXS (11). The resulting plasmids and similar pMXS-based expression plasmids for EML4-ALK variant 1, variant 1(K589M), variant 2, variant 3a, and variant 3b were individually introduced into HEK293 cells. Lysates of the transfected cells were subjected to immunoprecipitation with antibodies to FLAG, and the resulting precipitates were subjected either to immunoblot analysis with the same antibodies or to an *in vitro* kinase assay with the YFF peptide (12). Mouse 3T3 fibroblasts were also infected with recombinant retroviruses for each of the EML4-ALK variants or wild-type ALK and were then cultured for 12 d for a focus formation assay. The same set of 3T3 cells was injected s.c. into nu/nu mice, and tumor formation was examined after 20 d.

Results

Multiplex RT-PCR screening for EML4-ALK fusion transcripts in lung adenocarcinoma. As described above, exons 2, 6, 13, 18, 20, and 21 of *EML4* may participate in an in-frame fusion to exon 20 of *ALK* (Fig. 1A). To identify all possible *EML4-ALK* fusion cDNAs in a single-tube experiment, we designed a mixture of two sense primers (one targeted to exon 2 and the other to exon 13 of *EML4*) and a single antisense primer (targeted to exon 20 of *ALK*) and did multiplex RT-PCR with these primers and total cDNA preparations from tumor specimens. The exon 2 primer for *EML4* would be expected to generate a PCR product of 458 bp with the exon 2 (*EML4*)-exon 20 (*ALK*) fusion cDNA or of 917 bp with the exon 6-exon 20 fusion cDNA (variant 3). In addition, the exon 13 primer for *EML4* would be expected to generate PCR products of 432, 999, 1,185, or 1,284 bp with the exon 13-exon 20 (variant 1), exon 18-exon 20, exon 20-exon 20 (variant 2), and exon 21-exon 20 fusion cDNAs, respectively.

[6] K. Takeuchi et al., manuscript in preparation.

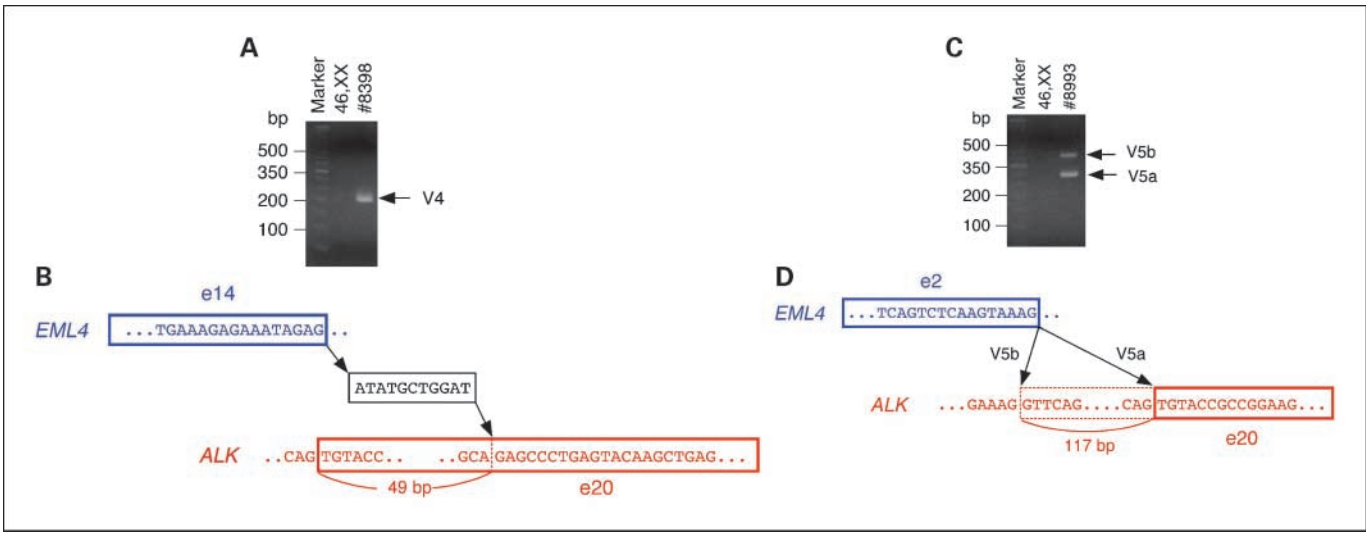

Fig. 2. Structure of *EML4-ALK* variant 4 and 5 cDNAs. *A,* RT-PCR amplification of the fusion point of *EML4-ALK* variant 4 mRNA in NSCLC specimen ID no. 8398 as well as in peripheral blood mononuclear cells of a female volunteer (46,XX). A PCR product of 203 bp corresponding to *EML4-ALK* variant 4 was specifically amplified from the tumor cells. The left lane contains DNA size standards (50-bp ladder). *B,* nucleotide sequencing of the PCR product in *A* revealed that exon 14 of *EML4* (*blue*) was connected to an 11-bp cDNA fragment of unknown identity (*black*), which was ligated in turn to the nucleotide at position 50 of exon 20 of *ALK* (*red*). *C,* RT-PCR amplification of the fusion point of *EML4-ALK* variant 5 mRNA in NSCLC specimen ID no. 8993 as well as in peripheral blood mononuclear cells of a female volunteer (46,XX). Two specific products of 415 and 298 bp were obtained, corresponding to variants 5b and 5a, respectively. The left lane contains DNA size standards (50-bp ladder). *D,* nucleotide sequencing of the PCR products in *C* revealed that exon 2 of *EML4* was fused either to exon 20 of *ALK*, generating the variant 5a cDNA, or to a position 117 bp upstream of exon 20 of *ALK*, generating the variant 5b cDNA.

Virtual gel electrophoresis of the multiplex RT-PCR products (Fig. 1B) revealed that 11 samples (4.35%) were positive for *EML4-ALK* cDNA among a consecutive series of 253 lung adenocarcinoma specimens, including those examined in our previous studies (8, 9, 13). All of the specimens previously shown to harbor *EML4-ALK* (two cases with variant 1, three with variant 2, and two with variant 3) were faithfully detected with our multiplex RT-PCR system. No specific PCR products were obtained for other types of lung cancer (n = 111) or other solid tumors (n = 292). Nucleotide sequencing of the PCR products for the newly identified positive cases revealed that one specimen was positive for variant 1 and another for variant 3 of *EML4-ALK*, but that the remaining two specimens harbored previously unidentified variants (Fig. 1B and C). Exon 14 of *EML4* was ligated to a position within exon 20 of *ALK* in the product from tumor ID no. 8398 (designated variant 4), whereas exon 2 of *EML4* was ligated to exon 20 of *ALK* in the product from tumor ID no. 8993 (designated variant 5).

Structure of EML4-ALK variant 4 cDNA. To verify the presence of novel *EML4-ALK* variants in the cancer cells, we first did direct RT-PCR analysis for the cDNA of tumor ID no. 8398 with a new set of primers encompassing the putative fusion point of variant 4. This analysis showed the presence of the fusion cDNA (Fig. 2A). Nucleotide sequencing of the PCR product revealed that exon 14 of *EML4* was fused to an unknown sequence of 11 bp, which in turn was connected to the nucleotide at position 50 of exon 20 of *ALK* (Fig. 2B). (We failed to detect a region of the human genome (build 36) homologous to the 11-bp connecting sequence in a BLAST search.[7]) Although exon 14 of *EML4* is not expected to produce an in-frame fusion to exon 20 of *ALK*, insertion of the unknown 11-bp sequence and its ligation to a position within the *ALK* exon allows an in-frame connection between the two genes. Fusion cDNAs in which the point of connection is located within, rather that at the 5′ terminus of, exon 20 of *ALK* have also been described for *MSN-ALK* (14) and *MYH9-ALK* (15).

We further examined whether a full-length cDNA encoding such an unexpected EML4-ALK variant could be isolated from the cancer cells. For this purpose, we designed a sense primer targeted to the 5′ untranslated region of *EML4* cDNA as well as an antisense primer targeted to the 3′ untranslated region of *ALK* cDNA. Direct RT-PCR analysis with this primer set yielded a single PCR product of ~3.4 kbp with total cDNA of tumor ID no. 8398 (Supplementary Fig. S1A). Complete nucleotide sequencing of the PCR product revealed that the cDNA contained an open reading frame for 1,097 amino acids comprising residues 1 to 547 of human EML4, residues 1,075 to 1,620 of human ALK, and 4 amino acids of unknown origin between these two sequences (Supplementary Fig. S1B). The isolation of a full-length cDNA containing the 11-bp insert indicated that the variant 4 protein was likely expressed in the cancer cells.

Structure of EML4-ALK variant 5 cDNAs. We similarly investigated the presence of variant 5 mRNA in the cells of tumor ID no. 8993. Direct RT-PCR analysis to amplify the fusion point of this variant cDNA yielded two independent products of 298 and 415 bp (Fig. 2C). Nucleotide sequencing of each product revealed that the former contained exon 2 of *EML4* and exon 20 of *ALK*, as expected, whereas in the latter, exon 2 of *EML4* was connected to a position within intron 19 of *ALK* located 117 bp upstream of exon 20 (Fig. 2D). These fusion constructs were designated variants 5a and 5b, respectively.

Although no mRNAs or expressed sequence tags in the nucleotide sequence database were found to contain the

[7] http://www.ncbi.nlm.nih.gov/genome/seq/blastgen/blastgen.cgi?taxid=9606

117-bp sequence of intron 19 of *ALK*, the human genome sequence surrounding the 5′ terminus of this 117-bp sequence is AG-GT (Fig. 2D), which conforms to the consensus sequence for a splicing acceptor site. To show that such a cryptic exon is indeed involved in the production of an oncogenic kinase, we attempted to detect full-length cDNAs for variants 5a and 5b from total cDNA of tumor ID no. 8993. A doublet of PCR products of ~2.0 kbp was obtained (Supplementary Fig. S1A), and nucleotide sequencing of these products revealed that they indeed encode EML4-ALK variant 5a and 5b proteins (Supplementary Fig. S1C). Genomic PCR and fluorescence *in situ* hybridization (FISH) analyses further revealed that the cells of tumor ID no. 8993 harbor a single *EML4-ALK* fusion gene, suggesting that variant 5a and 5b mRNAs are generated by alternative splicing of the primary transcript of this single fusion gene (see below).

Detection of the EML4-ALK fusion genes by FISH. To confirm the rearrangements involving the *ALK* locus in the specimens harboring variants 4 and 5 of *EML4-ALK* cDNA, we did FISH analysis with tissue sections. We first designed a FISH-based "fusion assay" for *EML4* and *ALK* genes. Bacterial artificial chromosome fragments encompassing the entire genes were fluorescently labeled green and red, respectively. An overlapping signal for both probes was readily identified in a merged image for the tumor cells harboring variants 4 or 5 of *EML4-ALK* (Fig. 3A). To confirm further the breakage of the *ALK* locus, we did an "*ALK* split assay" with bacterial artificial chromosome fragments encompassing the 5′ or 3′ regions of the locus and labeled green and red, respectively. In this assay, the normal *ALK* locus would be expected to yield an overlapping signal, whereas a pair of separate green and red signals would indicate genomic breakage within *ALK*. As expected, a proportion of cells of tumor ID no. 8398 or no. 8993 in the histologic sections generated one overlapping signal and one pair of split signals (Fig. 3B), suggesting that these tumor cells each have at least one normal and at least one rearranged *ALK* locus.

These data, together with genomic PCR analysis (data not shown), thus indicated that the cells of each of these tumors harbor one normal chromosome 2 and a chromosome 2 with an inv(2)(p21p23) rearrangement. The other *EML4-ALK* cDNA-positive specimens (variants 1 to 3) in this cohort showed a similar FISH labeling profile, consistent with the presence of the corresponding *EML4-ALK* rearrangements (data not shown).

***Detection of EML4-ALK proteins* in situ.** To detect EML4-ALK proteins in the cancer cells, we did immunohistochemical analysis with the ALK1 monoclonal antibody to ALK (16). The cytoplasm of tumor cells harboring *EML4-ALK* variant 1 (ID no. 9034), variant 4 (ID no. 8398), or variant 5 (ID no. 8993) manifested a diffuse pattern of immunoreactivity with fine granular concentrations (Fig. 3C). No normal pulmonary epithelial cells or lymphocytes in the sections of these specimens reacted with the antibody.

Transforming activity of EML4-ALK variants. We prepared expression plasmids for FLAG epitope–tagged EML4-ALK variants 1, 2, 3a, 3b, 4, 5a, and 5b, the predicted molecular sizes of which are 118,356; 146,913; 87,613; 88,874; 122,541; 71,046; and 74,867 Da, respectively. Each of these proteins, as well as a kinase-inactive mutant of EML4-ALK variant 1 (8), was expressed independently in HEK293 cells, immunoprecipitated, and subjected to immunoblot analysis with antibodies to FLAG. Each cDNA generated an EML4-ALK protein of the expected molecular size (Fig. 4A). The same immunoprecipitates were subjected to an *in vitro* kinase assay with the synthetic peptide YFF (12). Each variant protein (with the exception of the kinase-inactive mutant of variant 1) was shown to possess protein tyrosine kinase activity, with that of variants 3a, 3b, and 5b being most prominent (Fig. 4A).

To examine the transforming potential of the EML4-ALK variants, we transfected mouse 3T3 fibroblasts with the corresponding expression plasmids and then cultured the cells for 12 days. Transformed foci were readily detected for the cells expressing the variants of EML4-ALK but not for cells overexpressing wild-type ALK (Fig. 4B). Furthermore, s.c. injection of the transfected 3T3 cells into the shoulder of nude mice revealed that those expressing the various EML4-ALK isoforms, but not those overexpressing wild-type ALK, formed large tumors *in vivo* (Fig. 4B).

Discussion

We have done multiplex RT-PCR analysis to detect all possible isoforms of *EML4-ALK* transcripts in NSCLC cells, and unexpectedly identified two novel subtypes of the fusion event. This finding was supported by detection of the corresponding fusion genes by genomic PCR and FISH

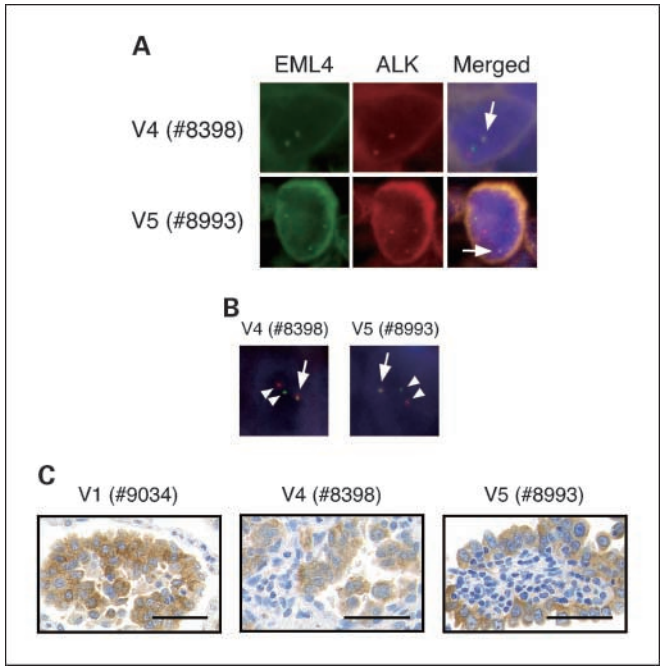

Fig. 3. FISH and immunohistochemical analyses of NSCLC specimens. *A,* FISH analysis of representative cancer cells in sections of lung adenocarcinoma harboring *EML4-ALK* variant 4 (ID no. 8398) or variant 5 (ID no. 8993). Each section was subjected to hybridization with differentially labeled probes for *EML4* (*left*) or for *ALK* (*center*). A fusion signal (*arrow*) and a pair of green (*EML4*) and red (*ALK*) signals are present in each merged image (*right*). *B,* the same clinical specimens as in *A* were subjected to FISH analysis with differentially labeled probes for the 5′ (*green*) or 3′ (*red*) regions of the *ALK* locus. A pair of split signals (*arrowheads*) and an overlapping signal (*arrow*) indicate the rearranged and normal *ALK* loci, respectively. *C,* immunohistochemical analysis of NSCLC specimens positive for *EML4-ALK* variants 1 (ID no. 9034), 4 (ID no. 8398), or 5 (ID no. 8993) with a monoclonal antibody to ALK. A pattern of diffuse staining with fine granular foci was apparent in the cytoplasm of all three tumors. Scale bars, 50 μm.

analyses and by that of the encoded proteins by immunohistochemical analysis in the NSCLC cells. Together with the previously isolated variants (8, 9), we have to date identified a total of seven distinct isoforms of EML4-ALK (variants 1, 2, 3a, 3b, 4, 5a, and 5b). Given that each of these isoforms possesses marked transforming activity, they all likely play an important role in the development of NSCLC. Our failure to detect *EML4-ALK* cDNA in the other solid tumors (n = 313) examined suggests that *EML4-ALK* may be an oncogene specific to NSCLC, especially to lung adenocarcinoma.

In our multiplex RT-PCR analysis, a sense primer targeted to exon 2 of *EML4* was designed to detect fusion events involving exon 2 or 6 of *EML4*, and PCR products of the expected sizes were indeed obtained with NSCLC specimens positive for such fusion events (variants 5 and 3, respectively). The other sense primer was targeted to exon 13 of *EML4* and was designed to detect fusion events involving exon 13, 18, 20, or 21 of *EML4*. Given that we were able to readily amplify a specific product of 1185 bp corresponding to the fusion event involving exon 20 of *EML4* (variant 2), it is likely that all possible fusions giving rise to PCR products up to this size would have been detected in our cohort. It should be noted, however, that a possible fusion between exon 21 of *EML4* and exon 20 of *ALK* would be expected to generate a PCR product of 1,284 bp. Although the size difference between the 1,185- and 1,284-bp products is small (99 bp), it is still possible that our multiplex RT-PCR analysis failed to efficiently amplify the longer product and that there may be as-yet-undetected fusion events for *EML4-ALK* in our cohort.

All EML4-ALK isoforms manifested a similar subcellular distribution profile despite marked differences in the size and domain structure of the EML4 portions of these chimeric proteins. In addition, the intracellular signaling systems activated by EML4-ALK may be shared among variants 1 to 5 (Supplementary Fig. S2). The EML4 portion of variant 5 comprises only the coiled-coil domain. This domain of EML4 may therefore play an essential role not only in the dimerization and activation of EML4-ALK isoforms (8) but also in tethering EML4-ALK to specific subcellular components. The pattern of subcellular immunostaining for EML4-ALK (cytoplasmic staining with fine granular foci) was distinct from that for other ALK fusion proteins associated with other malignancies (17, 18), suggesting that the subcellular localization of ALK fusion kinases varies substantially. The first such fusion kinase to be identified, NPM-ALK, preferentially phosphorylates STAT3, which is thought to participate in mitogenic signaling by NPM-ALK (19–21). Five ALK fusion kinases (NPM-ALK, TFG-ALK, ATIC-ALK, TPM3-ALK, and CLTC-ALK) were shown to differ markedly in their abilities to transform 3T3 fibroblasts, to phosphorylate STAT3 and AKT, and to activate phosphoinositide 3-kinase (17). Furthermore, a proteomics approach to identify tyrosine-phosphorylated proteins failed to detect marked phosphorylation of STAT3 in NSCLC specimens positive for EML4-ALK (22). It is therefore likely that each ALK fusion kinase exerts its effects through fusion-specific (although possibly partially overlapping) downstream pathways. In addition, we detected slight differences in catalytic and transforming activities among the variants of EML4-ALK (Fig. 4). These differences are likely due to the different portions of EML4 present in the different variants, which may affect dimerization affinity or the recruitment of substrates.

In addition to *EML4-ALK*, NSCLC cells harbor other potent oncogenes such as mutant versions of *EGFR* or *KRAS*. These three oncogenes, however, were found to be mutually exclusive

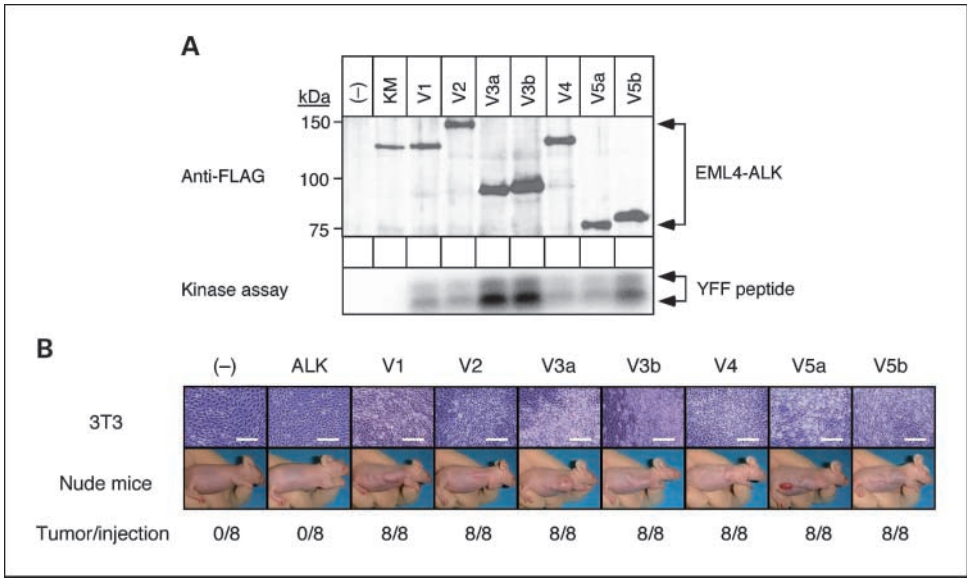

Fig. 4. Transforming potential of EML4-ALK variants. *A,* HEK293 cells expressing FLAG-tagged variant 1, 2, 3a, 3b, 4, 5a, or 5b of EML4-ALK were lysed and subjected to immunoprecipitation with antibodies to FLAG. The resulting precipitates were then either subjected to immunoblot analysis with antibodies to FLAG (*top*) or assayed for kinase activity with the synthetic YFF peptide (*bottom*). Cells transfected with the empty vector (-) or with a vector for a kinase-inactive mutant (*KM*) of EML4-ALK variant 1 were also analyzed. The positions of molecular size standards (kDa) and of EML4-ALK proteins are indicated on the left and right of the top panel, respectively. *B,* mouse 3T3 fibroblasts were transfected with expression plasmids for wild-type ALK or FLAG-tagged EML4-ALK variants, or with the empty plasmid (-), and were photographed after culture for 12 d (*top*). Scale bars, 200 μm. Alternatively, the transfected cells were injected s.c. into the shoulder of nu/nu mice and tumor formation was examined after 20 d (*bottom*). The number of tumors formed per eight injections is indicated at the bottom.

in our previous NSCLC cohort (8, 13), suggesting that EML4-ALK–positive NSCLC is a distinct subclass of lung cancer. Given that a selective inhibitor of the kinase activity of ALK rapidly induces cell death in EML4-ALK–positive cancer cells both *in vitro* (8, 9) and *in vivo*,[8] determination of the presence or absence of *EML4-ALK* in a given tumor may in the future inform the choice of treatment strategy for NSCLC. The demonstration of the existence of multiple isoforms of *EML4-ALK* transcripts will necessitate optimization of the detection systems so that all isoforms are detected with a high accuracy and sensitivity.

[8] M. Soda et al., submitted for publication.

Note Added in Proof

During our revision process, a novel EML4-ALK fusion variant was reported by Koivunen et al. (Clin Cancer Res 2008;14:4275–83). They have designated it as variant 4, which is different from our variant 4 in the present study.

Disclosure of Potential Conflicts of Interest

K. Takeuchi is a consultant providing advisory services to Dako for their antibodies.

Acknowledgments

We thank Kazuko Yokokawa, Motoyoshi Iwakoshi, Miyuki Kogure, and Tomoyo Kakita for technical assistance as well as Yuki Takano for help in preparation of the manuscript.

References

1. Aplan PD. Causes of oncogenic chromosomal translocation. Trends Genet 2006;22:46–55.
2. Sherbenou DW, Druker BJ. Applying the discovery of the Philadelphia chromosome. J Clin Invest 2007;117:2067–74.
3. Pulford K, Morris SW, Turturro F. Anaplastic lymphoma kinase proteins in growth control and cancer. J Cell Physiol 2004;199:330–58.
4. Mitelman F, Johansson B, Mertens F. Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. Nat Genet 2004;36:331–4.
5. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer 2007;7:233–45.
6. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005;310:644–8.
7. Tomlins SA, Laxman B, Dhanasekaran SM, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature 2007;448:595–9.
8. Soda M, Choi YL, Enomoto M, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 2007;448:561–6.
9. Choi YL, Takeuchi K, Soda M, et al. Identification of novel isoforms of the EML4-ALK transforming gene in non-small cell lung cancer. Cancer Res. In press 2008;68:4971–6.
10. Travis WD, Elisabeth B, Muller-Hermelink HK, Harris CC, editors. Pathology and genetics of tumours of the lung, pleura, thymus and heart. Lyon: IARC Press; 2004.
11. Onishi M, Kinoshita S, Morikawa Y, et al. Applications of retrovirus-mediated expression cloning. Exp Hematol 1996;24:324–9.
12. Donella-Deana A, Marin O, Cesaro L, et al. Unique substrate specificity of anaplastic lymphoma kinase (ALK): development of phosphoacceptor peptides for the assay of ALK activity. Biochemistry 2005;44:8533–42.
13. Inamura K, Takeuchi K, Togashi Y, et al. EML4-ALK fusion is linked to histological characteristics in a subset of lung cancers. J Thorac Oncol 2008;3:13–7.
14. Tort F, Pinyol M, Pulford K, et al. Molecular characterization of a new ALK translocation involving moesin (MSN-ALK) in anaplastic large cell lymphoma. Lab Invest 2001;81:419–26.
15. Lamant L, Gascoyne RD, Duplantier MM, et al. Non-muscle myosin heavy chain (MYH9): a new partner fused to ALK in anaplastic large cell lymphoma. Genes Chromosomes Cancer 2003;37:427–32.
16. Pulford K, Lamant L, Morris SW, et al. Detection of anaplastic lymphoma kinase (ALK) and nucleolar protein nucleophosmin (NPM)-ALK proteins in normal and neoplastic cells with the monoclonal antibody ALK1. Blood 1997;89:1394–404.
17. Armstrong F, Duplantier MM, Trempat P, et al. Differential effects of X-ALK fusion proteins on proliferation, transformation, and invasion properties of NIH3T3 cells. Oncogene 2004;23:6071–82.
18. Duyster J, Bai RY, Morris SW. Translocations involving anaplastic lymphoma kinase (ALK). Oncogene 2001;20:5623–37.
19. Wan W, Albom MS, Lu L, et al. Anaplastic lymphoma kinase activity is essential for the proliferation and survival of anaplastic large-cell lymphoma cells. Blood 2006;107:1617–23.
20. Marzec M, Kasprzycka M, Ptasznik A, et al. Inhibition of ALK enzymatic activity in T-cell lymphoma cells induces apoptosis and suppresses proliferation and STAT3 phosphorylation independently of Jak3. Lab Invest 2005;85:1544–54.
21. Galkin AV, Melnick JS, Kim S, et al. Identification of NVP-TAE684, a potent, selective, and efficacious inhibitor of NPM-ALK. Proc Natl Acad Sci U S A 2007;104:270–5.
22. Rikova K, Guo A, Zeng Q, et al. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. Cell 2007;131:1190–203.