ALK/3T3 nodules from the lungs (Fig. 4*B*) and again demonstrating the high efficacy of the ALK inhibitor.

Discussion

We have shown here that the EML4-ALK fusion kinase plays an essential role in lung tumorigenesis. Hundreds of adenocarcinoma nodules developed simultaneously within a few weeks after birth in all independent lines of *EML4-ALK* transgenic mice examined. Given that the promoter fragment of *SPC* becomes active only at a late stage of gestation (21), a short period of *EML4-ALK* expression appears to be sufficient for full transformation. Although we did not examine *TP53* and *RB1* for possible abnormalities in the adenocarcinoma nodules of the transgenic mice, with both of these genes being frequently inactivated in human lung cancers (22), it is likely that only one (or at most a few) additional genetic event is required to generate cancer in EML4-ALK-expressing alveolar epithelial cells.

The expression level of EML4-ALK protein in the adenocarcinoma nodules of the transgenic mice was low. Given that the abundance of *EML4-ALK* mRNA in these nodules was found to be greater than that in human *EML4-ALK*-positive NSCLC specimens (data not shown), the expression of EML4-ALK protein appears to be suppressed in the mouse lung epithelial cells, possibly through translational or posttranslational mechanisms. The development of adenocarcinoma even at this low level of protein expression further reinforces the transforming activity of EML4-ALK.

Given the rapid development of NSCLC induced by EML4-ALK, the tumor cells are likely dependent for growth on the tyrosine kinase activity of the fusion protein. Such "oncogene addiction" (23) provides a potential target for the development of treatment strategies. We therefore tested whether inhibition of the enzymatic activity of EML4-ALK might reduce the tumor burden in the transgenic mice. The ALK inhibitor examined proved to be a promising candidate for the treatment of EML4-ALK-positive tumors. Furthermore, given the high sensitivity of the tumors in the transgenic mice to the ALK inhibitor, these animals provide a model system with which to examine the in vivo activity of other compounds or reagents targeted to ALK.

Many of the large tumors in the lungs of the transgenic mice changed to bullae or cysts after treatment with the ALK inhibitor, as revealed both by CT scanning (Fig. S3*A* and Movie S3 and Movie S5) and by pathology examination (Fig. S3*B*). Such a change was not described for treatment of activated EGFR-positive NSCLC in mouse models or humans with EGFR inhibitors (5, 6). A rapid induction of cell death by the ALK inhibitor in the transgenic mice may have triggered a collapse of the tumor burden within each nodule, thereby giving rise to bullae or cysts. Indeed, pathology examination revealed that a large tumor in 1 transgenic mouse (no. 250) became filled with necrotic tissue after treatment (Fig. S3*C*). However, the bullae cysts and necrotic tissue were still surrounded by remaining cancer cells (Fig. S3 *B* and *C*). Similarly, the lining tissue of some bullae cysts in the treated mice appeared to have a high density in CT scans (Fig. S3*A*), suggesting that peripheral cancer cells may survive in the nodules. Furthermore, small foci of cancer cells could be identified in the lungs of transgenic mice in the treatment cohort (Fig. 3*C*). Together, these various observations indicate that the current treatment protocol with the ALK inhibitor did not entirely eliminate tumor cells from the transgenic mice. Indeed, in a separate experiment transgenic mice treated with the 2,4-pyrimidinediamine for 25 days were examined 3 months after cessation of drug administration. Tumors of various sizes regrew in these mice (Fig. S4), indicative of the presence of surviving EML4-ALK-positive cancer cells in the animals after 25 days of drug treatment. Given that we have not tried other protocols or compounds, it remains unknown whether a total cure might be achieved by treatment for a longer period or with a higher dose of the same inhibitor or with a more potent compound. It is also possible that inhibition of additional signaling pathways, such as those mediated by phosphoinositide 3-kinase, mammalian target of rapamycin, or other protein tyrosine kinases (5, 24), may be required for a complete cure.

Despite the rapid growth of multiple tumors in the lungs of the transgenic mice, we failed to detect distant metastasis of such cancer cells in animals killed for analysis or in those that died within the total observation period of 6 months. However, we conclude that the tumors that developed in these mice had malignant characteristics on the basis of the following observations: (*i*) Histological analysis indicated that most tumors were noninvasive papillary adenocarcinomas, with some of them further showing obvious fibrosis and destruction of alveolar structures (Fig. 2*B*), a marker of invasion in human lung adenocarcinoma. (*ii*) Subcutaneous transplantation of tumor nodules that developed in the transgenic mice into the shoulder of *nu/nu* mice resulted in the growth of tumors at 6 of 8 injection sites in the recipient animals (Fig. S5*A*). (*iii*) Tumors that developed in the transgenic mice were shown to keep growing for at least 62 days in in vitro culture (Fig. S5*B*).

It is likely that expression of EML4-ALK (and probably other accompanying genetic changes) alone is not sufficient to render the cancer cells metastatic. It remains to be determined whether additional abnormalities in other oncogenes or tumor suppressor genes, such as *KRAS* or *LKB1* (25, 26), may lead to the generation of metastatic tumors in *EML4-ALK* transgenic mice.

Our present results have reinforced the importance of *EML4-ALK* in the pathogenesis of NSCLC in humans and have provided experimental support for the treatment of such intractable tumors with ALK inhibitors. Given that variants of *EML4-ALK* other than the variants 1 and 2 described in our original study (12) are now being identified (20, 27–29), it will be important to characterize all possible isoforms of *EML4-ALK* in humans to identify precisely the subgroup of patients who are candidates for future treatment with ALK inhibitors. Further to this goal, it will also be important to clarify the genetic changes that accompany the *EML4-ALK* fusion event as well as the downstream targets of EML4-ALK action in human NSCLC.

Materials and Methods

Generation of Transgenic Mice. A cDNA fragment encoding FLAG epitope-tagged EML4-ALK variant 1 (12) was ligated to the *SPC* promoter as well as to splicing and polyadenylation signals (Fig. 1*A*). The expression cassette was injected into pronuclear-stage embryos of C57BL/6J mice (PhoenixBio), and the copy number of the transgene was examined by Southern blot analysis with DNA from the tail of founder animals. All animal procedures were performed with the approval of the scientific committee for animal experiments of Jichi Medical University.

For detection of mRNAs derived from *EML4-ALK* and the glyceraldehyde-3-phosphate dehydrogenase gene (*GAPDH*), total RNA was isolated from the organs of transgenic mice with the use of an RNeasy Mini kit (Qiagen) and was subjected to reverse transcription with SuperScript III reverse transcriptase (Invitrogen) and an oligo(dT) primer. Both reverse transcription and subsequent PCR analysis for each gene were performed as described previously (12).

For analysis of EML4-ALK protein in mice, organ homogenates were prepared with an Nonidet P-40 lysis buffer and subjected to immunoprecipitation with mouse monoclonal antibodies to FLAG (Millipore). The resulting precipitates were then subjected to immunoblot analysis with the same antibodies and a SuperSignal chemiluminescence kit (Pierce Biotechnology).

Pathology Examination. For immunohistochemical staining of EML4-ALK in EML4-ALK/3T3 cells, paraffin-embedded sections were depleted of paraffin with xylene, rehydrated with a graded series of ethanol solutions, and stained with mouse monoclonal antibodies to ALK (ALK1; Dako) at a dilution of 1:20 and with an EnVision+DAB system (Dako). The sections were subjected to heat-induced antigen retrieval with Target Retrieval Solution, pH 9.0 (Dako), before exposure to the antibodies. For detection of EML4-ALK in transgenic mice, cryostat sections were fixed with 4% paraformaldehyde in 0.1 M sodium phosphate buffer (pH 7.4) for 10 min, treated with Target Retrieval Solution, pH 9.0, and immunostained with the monoclonal antibodies to ALK and the EnVision+DAB system.

Treatment with ALK Inhibitor. For the experiments based on i.v. administration of EML4-ALK/3T3 cells, the cells (2×10^5) were injected into the tail vein of 4-week-old *nu/nu* mice (Clea Japan). An inhibitor specific for the tyrosine kinase activity of ALK [example 3-39 in the patent application: Garcia-Echeverria C, *et al.*, inventors; Novartis AG, Novartis Pharma GmbH, IRM LLC, applicants (24 Feb 2005). 2,4-Pyrimidinediamines useful in the treatment of neoplastic disease and in inflammatory and immune system disorders. PCT WO 2005016894] was synthesized by Astellas Pharma and was orally administered each day at a dose of 10 mg/kg to the injected mice or to *EML4-ALK*-transgenic mice. Sequential examination of lung tumors was performed with an X-ray CT apparatus for experimental animals (LCT-100; Aloka).

1. Schiller JH, *et al.* (2002) Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N Engl J Med* 346:92–98.
2. Paez JG, *et al.* (2004) EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy. *Science* 304:1497–1500.
3. Pao W, *et al.* (2004) EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci USA* 101:13306–13311.
4. Shigematsu H, *et al.* (2005) Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst* 97:339–346.
5. Li D, *et al.* (2007) Bronchial and peripheral murine lung carcinomas induced by T790M-L858R mutant EGFR respond to HKI-272 and rapamycin combination therapy. *Cancer Cell* 12:81–93.
6. Lynch TJ, *et al.* (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 350:2129–2139.
7. Kris MG, *et al.* (2003) Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: A randomized trial. *J Am Med Assoc* 290:2149–2158.
8. Hatanaka H, *et al.* (2007) Transforming activity of purinergic receptor P2Y, G-protein coupled, 2 revealed by retroviral expression screening. *Biochem Biophys Res Commun* 356:723–726.
9. Fujiwara S, *et al.* (2007) Transforming activity of purinergic receptor P2Y, G protein coupled, 8 revealed by retroviral expression screening. *Leuk Lymphoma* 48:978–986.
10. Choi YL, *et al.* (2007) Identification of a constitutively active mutant of JAK3 by retroviral expression screening. *Leuk Res* 31:203–209.
11. Besse B, Ropert S, Soria JC (2007) Targeted therapies in lung cancer. *Ann Oncol* 18(suppl 9):ix135–ix142.
12. Soda M, *et al.* (2007) Identification of the transforming *EML4-ALK* fusion gene in non-small-cell lung cancer. *Nature* 448:561–566.
13. Mishra A, Weaver TE, Beck DC, Rothenberg ME (2001) Interleukin-5-mediated allergic airway inflammation inhibits the human surfactant protein C promoter in transgenic mice. *J Biol Chem* 276:8453–8459.
14. Duan W, *et al.* (2002) Lung-specific expression of human mutant p53–273H is associated with a high frequency of lung adenocarcinoma in transgenic mice. *Oncogene* 21:7831–7838.

ACKNOWLEDGMENTS. We thank Koichi Hagiwara (Saitama Medical University, Saitama, Japan) for kindly providing the human *SPC* promoter fragment as well as Keiko Shiozawa and Tomoyo Kakita for technical assistance. This study was supported in part by a grant for Research on Human Genome Tailor-made from the Ministry of Health, Labor, and Welfare of Japan; by Grants-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science, and Technology of Japan; and by grants from the Japan Society for the Promotion of Science, the National Institute of Biomedical Innovation, the Princess Takamatsu Cancer Research Fund, the Takeda Science Foundation, the Uehara Memorial Foundation, the Smoking Research Foundation, the Vehicle Racing Commemorative Foundation, and the NOVARTIS Foundation (Japan) for the Promotion of Science.

15. Zhao B, *et al.* (2000) Transgenic mouse models for lung cancer. *Exp Lung Res* 26:567–579.
16. Inamura K, *et al.* (2008) *EML4-ALK* fusion is linked to histological characteristics in a subset of lung cancers. *J Thorac Oncol* 3:13–17.
17. Chiarle R, Voena C, Ambrogio C, Piva R, Inghirami G (2008) The anaplastic lymphoma kinase in the pathogenesis of cancer. *Nat Rev Cancer* 8:11–23.
18. McDermott U, *et al.* (2007) Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumor cell line profiling. *Proc Natl Acad Sci USA* 104:19936–19941.
19. Galkin AV, *et al.* (2007) Identification of NVP-TAE684, a potent, selective, and efficacious inhibitor of NPM-ALK. *Proc Natl Acad Sci USA* 104:270–275.
20. Choi YL, *et al.* (2008) Identification of novel isoforms of the EML4-ALK transforming gene in non-small cell lung cancer. *Cancer Res* 68:4971–4976.
21. Korfhagen TR, *et al.* (1990) Cis-acting sequences from a human surfactant protein gene confer pulmonary-specific gene expression in transgenic mice. *Proc Natl Acad Sci USA* 87:6122–6126.
22. Testa JR, *et al.* (1997) Advances in the analysis of chromosome alterations in human lung carcinomas. *Cancer Genet Cytogenet* 95:20–32.
23. Sharma SV, Settleman J (2007) Oncogene addiction: Setting the stage for molecularly targeted cancer therapy. *Genes Dev* 21:3214–3231.
24. Engelman JA, *et al.* (2007) MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* 316:1039–1043.
25. Ji H, *et al.* (2007) LKB1 modulates lung cancer differentiation and metastasis. *Nature* 448:807–810.
26. Huang CL, *et al.* (1998) Mutations of p53 and K-ras genes as prognostic factors for non-small cell lung cancer. *Int J Oncol* 12:553–563.
27. Rikova K, *et al.* (2007) Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* 131:1190–1203.
28. Koivunen JP, *et al.* (2008) EML4-ALK fusion gene and efficacy of an ALK kinase inhibitor in lung cancer. *Clin Cancer* Res 14:4275–4283.
29. Takeuchi K, *et al.* (2008) Multiplex reverse transcription-PCR screening for EML4-ALK fusion transcripts. *Clin Cancer Res* 14:6618–6624.

Vol 455|16 October 2008 doi:10.1038/nature07399

nature

LETTERS

Oncogenic mutations of ALK kinase in neuroblastoma

Yuyan Chen[1,2,3]*, Junko Takita[1,2,3]*, Young Lim Choi[4]*, Motohiro Kato[1,3], Miki Ohira[5], Masashi Sanada[2,3,6], Lili Wang[2,3,6], Manabu Soda[4], Akira Kikuchi[7], Takashi Igarashi[1], Akira Nakagawara[5], Yasuhide Hayashi[8], Hiroyuki Mano[4,6] & Seishi Ogawa[2,3,6]

Neuroblastoma in advanced stages is one of the most intractable paediatric cancers, even with recent therapeutic advances[1]. Neuroblastoma harbours a variety of genetic changes, including a high frequency of *MYCN* amplification, loss of heterozygosity at 1p36 and 11q, and gain of genetic material from 17q, all of which have been implicated in the pathogenesis of neuroblastoma[2–5]. However, the scarcity of reliable molecular targets has hampered the development of effective therapeutic agents targeting neuroblastoma. Here we show that the anaplastic lymphoma kinase (ALK), originally identified as a fusion kinase in a subtype of non-Hodgkin's lymphoma (NPM–ALK)[6–8] and more recently in adenocarcinoma of lung (EML4–ALK)[9,10], is also a frequent target of genetic alteration in advanced neuroblastoma. According to our genome-wide scans of genetic lesions in 215 primary neuroblastoma samples using high-density single-nucleotide polymorphism genotyping microarrays[11–14], the *ALK* locus, centromeric to the *MYCN* locus, was identified as a recurrent target of copy number gain and gene amplification. Furthermore, DNA sequencing of *ALK* revealed eight novel missense mutations in 13 out of 215 (6.1%) fresh tumours and 8 out of 24 (33%) neuroblastoma-derived cell lines. All but one mutation in the primary samples (12 out of 13) were found in stages 3–4 of the disease and were harboured in the kinase domain. The mutated kinases were autophosphorylated and displayed increased kinase activity compared with the wild-type kinase. They were able to transform NIH3T3 fibroblasts as shown by their colony formation ability in soft agar and their capacity to form tumours in nude mice. Furthermore, we demonstrate that downregulation of *ALK* through RNA interference suppresses proliferation of neuroblastoma cells harbouring mutated *ALK*. We anticipate that our findings will provide new insights into the pathogenesis of advanced neuroblastoma and that *ALK*-specific kinase inhibitors might improve its clinical outcome.

To identify oncogenic lesions in neuroblastoma, we performed a genome-wide analysis of primary tumour samples obtained from 215 neuroblastoma patients using high-density single-nucleotide polymorphism (SNP) arrays (Affymetrix GeneChip 250K *Nsp*I) (Supplementary Table 1). Twenty-four neuroblastoma-derived cell lines were also analysed (Supplementary Table 2). Interrogating over 250,000 SNP sites, this platform permits the identification of copy number changes at an average resolution of less than 12 kilobases (kb)[13,14].

Analysis of this large number of samples, consisting of varying disease stages, permitted us to obtain a comprehensive registry of genomic lesions in neuroblastoma (Supplementary Figs 1 and 2). A gain of chromosomes, often triploid or hyperploid (defined by mean copy number of >2.5), was a predominant feature of neuroblastoma genomes in the lower stages. Ploidy generally correlated with the clinical stage, where non-hyperploid cases were significantly associated with stage 4 disease ($P = 4.13 \times 10^{-5}$, trend test) (Supplementary Fig. 3 and Supplementary Table 3). 17q gains, frequently in multiple copies (3≤ copy number <5), were a hallmark of the neuroblastoma genome[4] and were found in most neuroblastoma cases. Copy number gains tended to spare chromosomes 3, 4, 10, 14 and 19 (Supplementary Figs 2 and 3). Notably, these chromosomes often had copy number losses including 1p (22.8%), 3p (8.8%), 4p (5.1%), 6q (7.0%), 10q (9.8%), 11q (19.5%), 14q (3.7%), 19p (7.4%) and 19q (5.1%), implicating the pathogenic role of 'relative' gene dosages.

After excluding known copy number variations, we identified a total of 28 loci undergoing high-grade amplifications (copy number ≥5) (Supplementary Table 4). These lesions fell into relatively small genomic segments, having a mean size of 361 kb, which accelerated the identification of gene targets in these regions (Supplementary Table 4 and Supplementary Fig. 4). The candidate gene targets included *TERT* (5p15.33), *HDAC3* (5q31.3), *IGF2* (11p15.1), *MYEOV* (11q13.3), *FGF7* (15q21.1) and *CDH13* (16q23.3). However, many of them were not recurrent but found only in a single case. Although the recurrent lesions were mostly explained by the amplification of *MYCN* at 2p24, as found in 50 out of 215 (23%) of the primary cases, we identified another peak of recurrent amplification at 2p23 (Fig. 1a), which consisted of amplicons in five primary cases and in one neuroblastoma-derived cell line, NB-1 (Supplementary Fig. 5). This peak was located at the centromeric margin of the common copy number gains in chromosome 2p, which was created by copy number gains in 109 samples mostly from non-hyperploid stage 4 cases. The minimum overlapping amplification was defined by the amplicons found in the NB-1 cell line (Supplementary Fig. 5) and contained a single gene, the anaplastic lymphoma kinase (*ALK*), which has previously been reported to be overexpressed in neuroblastoma cases[15]. Although five of the six samples showing *ALK* amplification also had *MYCN* amplification, one primary case (NT056) lacked a *MYCN* peak and the amplification was confined to the *ALK*-containing locus. In interphase fluorescent *in situ* hybridization (FISH) analysis of NB-1, *MYCN* and *ALK* loci were amplified in separate amplicons (Fig. 1b), indicating that the 2p23 amplicons containing *ALK* were unlikely to represent merely 'passenger' events of *MYCN* amplification but actively contributed to the pathogenesis of neuroblastoma.

Because an oncogene can be activated by gene amplification and/or mutation, to search for possible mutations we performed DNA heteroduplex formation analysis[16] and genomic DNA sequencing for the exons 20 to 28 of *ALK*, which encompass the juxtamembrane and kinase domains (Supplementary Table 5). In total, we identified eight nucleotide changes in 21 neuroblastoma samples, 13 out of 215

[1]Department of Pediatrics, [2]Cell Therapy and Transplantation Medicine, [3]Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, Tokyo 113-8655, Japan. [4]Division of Functional Genomics, Jichi Medical University, Toguchi 329-0498, Japan. [5]Division of Biochemistry, Chiba Cancer Center Research Institute, Chiba 260-8717, Japan. [6]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Saitama, 332-0012, Japan. [7]Division of Hematology/Oncology, Saitama Children's Medical Center, Saitama 339-8551, Japan. [8]Gunma Children's Medical Center, Shibukawa 377-8577, Japan.
*These authors contributed equally to this work.

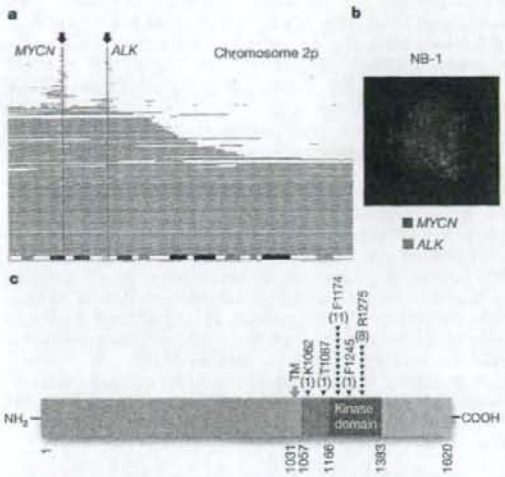

Figure 1 | Common 2p gains/amplifications and *ALK* mutations in neuroblastoma samples. **a**, Recurrent copy number gains on the 2p arm. High-grade amplifications are shown by light-red horizontal lines, whereas simple gains are shown by dark-red lines. Two common peaks of copy number gains and amplifications in the *MYCN* and *ALK* loci are indicated by arrows. The cytobands in 2p are shown at the bottom. **b**, Interphase FISH analysis of NB-1 showing high-grade amplification of *MYCN* (red) and *ALK* loci (green). The amplified *MYCN* locus appears as a single large signal. **c**, Distribution of the eight *ALK* mutations found in 21 neuroblastoma samples. The positions of the mutated amino acids are indicated by black (primary samples) and red (cell lines) arrowheads. The number of mutations at each site is shown at the top of the arrowheads. TM, transmembrane.

(6.1%) primary samples and 8 out of 24 (33%) cell lines, which resulted in seven types of amino acid substitutions at five different positions (Table 1 and Supplementary Fig. 6). They were not found in either the genomic DNA collected from 50 healthy volunteers or in the SNP databases at the time of preparing this manuscript. In fact, somatic origins of missense changes were confirmed in 9 out of 13 primary cases, for which DNA was obtained from the peripheral blood or the tumour-free bone marrow specimens (Supplementary Fig. 6). On the other hand, T1087I (ACC>ATC), found in case NT126, had a germline origin and thus it could not be determined whether the T1087I change was a rare non-functional polymorphism or represented a pathogenic germline mutation. For other changes found in three primary cases (NT128, NT217 and NT218) and cell lines, normal DNA was not available but they were likely to represent oncogenic mutations because they were identical to common somatic changes (F1174L or R1275Q) or shown to have oncogenic potential in functional assays (K1062M).

Most mutations occurred within the kinase domain (20 out of 22 or 91%), which clearly showed two mutation hotspots at F1174 and R1275 (Fig. 1c). A neuroblastoma-derived cell line, SJNB-2, had a homozygous *ALK* mutation of R1275Q, which was probably due to uniparental disomy of chromosome 2 (Supplementary Fig. 7a). Another case (NT074) harboured two different mutations, F1174L and R1275Q, but it remains to be determined whether both are on the same allele. ALK mutations within the kinase domain occurred at amino acid positions that are highly conserved across species and during molecular evolution (Supplementary Figs 8 and 9). According to the conserved structure of other insulin receptor kinases we predicted that F1174 is located at the end of the Cα1 helix, whereas the other two are on the two β-sheets: before the catalytic loop (β6, F1245) and within the activation loop (β9, R1275) (Supplementary Fig. 7b, c)[17]. Thus, conformational changes due to amino acid substitutions at these positions might be responsible for the aberrant activity of the mutant kinases.

Table 1 | *ALK* mutations/amplifications in neuroblastoma samples

Sample	Age (months)	Stage	*MYCN**	Clinical outcome	Mutations/ amplifications	Nucleotide substitution	Origin of mutations
NT126	99	4	–	Dead	T1087I	ACC>ATC	Germ line
NT218	8	1	–	Alive	F1174L	TTC>TTG	ND
NT074	34	3	+	Dead	F1174L	TTC>TTA	Somatic
					R1275Q	CGA>CAA	
NT160	12	4	+	Dead	F1174L	TTC>TTA	Somatic
NT217	24	4	+	Dead	F1174L	TTC>TTA	ND
NT190	48	4	+	Alive	F1174L	TTC>TTA	Somatic
NT060	163	3	–	Alive	F1174C	TTC>TGC	Somatic
NT162	28	4	+	Dead	F1174V	TTC>GTC	Somatic
NT195	24	4	+	Alive	F1245L	TTC>TTG	Somatic
NT055	6	3	–	Alive	R1275Q	CGA>CAA	Somatic
NT128	8	4	–	Dead	R1275Q	CGA>CAA	ND
NT164	54	4	+	Dead	R1275Q	CGA>CAA	Somatic
NT200	133	4	–	Dead	R1275Q	CGA>CAA	Somatic
SCMC-N5†	-	-	+	-	K1062M	AAG>ATG	ND
SJNB-4†	-	-	+	-	F1174L	TTC>TTA	ND
LAN-1†	-	-	+	-	F1174L	TTC>TTA	ND
SCMC-N2†	-	-	+	-	F1174L	TTC>TTA	ND
SK-N-SH†	-	-	–	-	F1174L	TTC>TTA	ND
SJNB-2†‡	-	-	+	-	R1275Q	CGA>CAA	ND
LAN-5†	-	-	+	-	R1275Q	CGA>CAA	ND
TGW†	-	-	+	-	R1275Q	CGA>CAA	ND
NT204	12	1	+	Alive	Amplification	-	-
NT056	11	3	–	Dead	Amplification	-	-
NT071	36	3	+	Alive	Amplification	-	-
NT165	19	4	+	Dead	Amplification	-	-
NT169	7	4	+	Dead	Amplification	-	-
NB-1†	-	-	+	-	Amplification	-	-

ND, not determined.
* Presence (+) or absence (–) of *MYCN* amplification in FISH analysis. All cases where there was an absence of *MYCN* amplification (–) were also checked for possible *MYCN* mutations by sequencing of all *MYCN* exons, but no *MYCN* mutations were identified.
† Cell lines.
‡ Homozygous mutation.

ALK mutation highly correlated with *MYCN* amplification ($P = 1.55 \times 10^{-4}$, Fisher's exact test; Supplementary Table 6) where 14 out of 21 mutations coexisted with *MYCN* amplification. Regardless of the status of *MYCN* amplification, 12 of the 13 mutations were found in patients with advanced stage neuroblastoma (Table 1). However, whereas *MYCN* amplification and stage 4 were significant risk factors for poor survival, the mutation/amplification status of *ALK* was not likely to have a major impact on survival (Supplementary Fig. 10 and Supplementary Table 7), although the statistical power of the current analysis was largely limited in order to detect a marginal hazard.

To evaluate the impact of ALK mutations on kinase activity, we generated Flag-tagged constructs of ALK and its mutants, F1174L and K1062M, which were stably expressed in NIH3T3 cells, and examined their phosphorylation status and *in vitro* kinase activity. The ALK mutants stably expressed in NIH3T3 cells were phosphorylated according to western blot analysis using an antibody specific for phosphorylated ALK (anti-pY1604) and a PY20 blot after anti-Flag immunoprecipitation of the mutant kinases (Fig. 2a), whereas the wild-type kinase was not phosphorylated. The immunoprecipitated ALK mutants also showed increased tyrosine kinase activity *in vitro* when compared with wild-type ALK. This was shown using both a universal substrate for tyrosine kinase (poly-GluTyr) and the synthetic YFF peptide[18], which was derived from a sequence of the activation loop of *ALK* (Fig. 2b, c). In accordance with these findings, downstream molecules of ALK signalling including AKT, STAT3 and ERK[15] were activated in cells expressing mutant ALK, as shown by their increased phosphorylation (Fig. 2d).

Next, we investigated the oncogenic potential of these mutants. NIH3T3 cells stably expressing mutant kinases showed increased colony formation in soft agar compared with the wild-type protein (Fig. 3a and Supplementary Fig. 11). The tumorigenicity of these ALK mutants was further assayed by injecting 1.0×10^7 NIH3T3 cells into nude mice. The NIH3T3 cells transfected with the ALK mutants showed focus-forming capacity and developed subcutaneous tumours (6 out of 6 inoculations) 21 days after inoculation, whereas the mock and wild-type ALK-transfected cells did not (0 out of 6 inoculations) (Fig. 3b, c). Finally, we examined the effect of ALK inhibition on the proliferation of neuroblastoma-derived cell lines. RNA interference (RNAi)-mediated *ALK* knockdown resulted in reduced cell proliferation of SK-N-SH cells harbouring the F1174L mutation, but the effects were less clear in wild-type ALK-expressing LAN-2 cells (Fig. 3d, e). Of particular interest is a recent report that 5 out of 17 neuroblastoma-derived cell lines, including SK-N-SH and NB-1, frequently showed high sensitivity to the specific ALK inhibitor TAE684 (ref. 19).

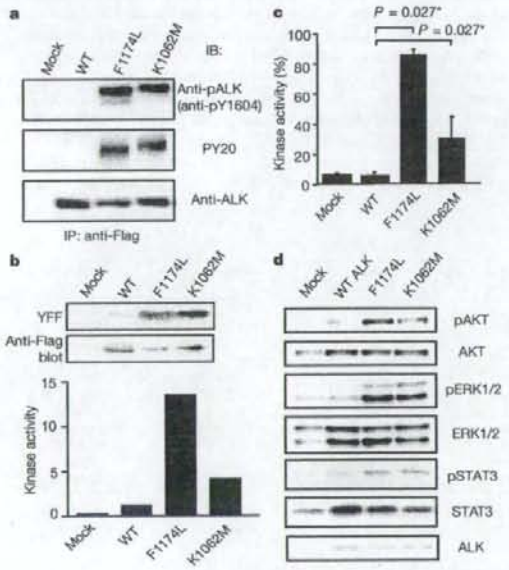

Figure 2 | Kinase activity of ALK mutants and their downstream signalling. **a**, Stably expressed ALK and its mutants (F1174L and K1062M) were immunoprecipitated with an anti-Flag antibody and subjected to western blot analysis with anti-pY1604 (upper panel) or PY20 (middle panel). An anti-ALK blot of precipitated kinases is also displayed (bottom panel). **b**, *In vitro* kinase assay for wild-type ALK kinase and its mutants using the synthetic YFF peptide as a substrate, where kinase activity is expressed as relative values to that for wild-type kinase based on the densities in the autoradiogram. **c**, Kinase activity was also assayed for the poly-GluTyr peptide. Significantly different measurements are indicated by asterisks with *P* values. Bars show mean (±s.d.) in three independent experiments. **d**, Western blot analyses of NIH3T3 cells expressing wild-type and mutant ALK for phosphorylated forms of AKT (pAKT), ERK (pERK1/2) and STAT3 (pSTAT3). The total amount of each molecule is also displayed (AKT, ERK1/2, and STAT3) together with an anti-ALK blot (ALK).

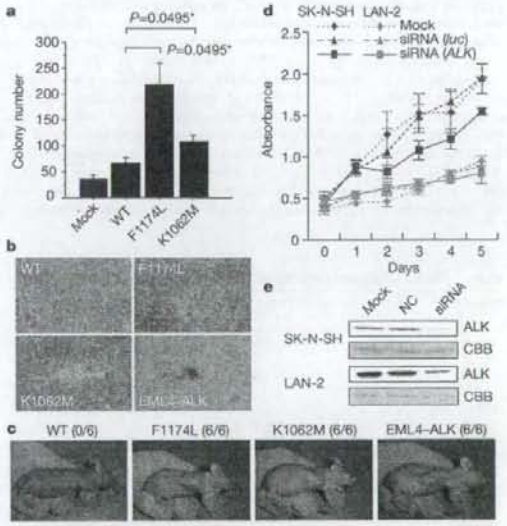

Figure 3 | Oncogenic role of ALK mutations. **a**, Colony assays for NIH3T3 cells stably expressing wild-type as well as mutant ALK (F1174L and K1062M). The average numbers of colonies in triplicate experiments are plotted and standard deviation is indicated. Results showing statistically significant differences as compared with experiments using wild-type ALK are indicated by asterisks with *P* values. **b**, **c**, NIH3T3 cells were transfected with wild-type and mutant ALK (F1174L, K1062M and EML4–ALK) and subjected to a focus forming assay (**b**) as well as an *in vivo* tumorigenicity assay in nude mice (**c**). **d**, Effect of RNAi-mediated ALK knockdown on cell proliferation in neuroblastoma cell lines expressing either the F1174L mutant (SK-N-SH) or wild-type ALK (LAN-2). Cell growth was measured using the Cell Counting Kit-8 after knockdown experiments using *ALK*-specific siRNAs (siRNA *ALK*), control siRNAs (siRNA *luc*), or mock experiments, where absorbance was measured in triplicate and averaged for each assay. To draw growth curves, the mean ± s.d. of the averaged absorbance in three independent knockdown experiments is plotted. **e**, Successful knockdown of ALK protein was confirmed by anti-ALK blots (ALK) using Coomassie brilliant blue G-250 (CBB) staining as loading controls. NC, control siRNA; siRNA, *ALK* siRNA.

Through the genome-wide analysis of genetic lesions in neuroblastoma, we identified novel oncogenic *ALK* mutations in advanced neuroblastoma. Combined with the cases having a high-grade amplification of the *ALK* gene, aberrant ALK signalling was likely to be involved in 11% (16 out of 151) of the advanced neuroblastoma cases. Because ALK kinase has been shown to be deregulated only in the form of a fusion kinase in human cancers, including lymphoma and lung cancer, the identification of oncogenic mutations in *ALK* not only increases our understanding of the molecular pathogenesis of advanced neuroblastoma, but also adds a new paradigm to the concept of 'ALK-positive human cancers' in that the mutated *ALK* kinases themselves might participate in human cancers. Our results again highlight the power of genome-wide studies to clarify the genetic lesions in human cancers[20–22]. Given that *ALK* mutations are preferentially involved in advanced neuroblastoma cases having a poor prognosis, our findings implicate that *ALK* inhibitors may improve the clinical outcome of children suffering from intractable neuroblastoma.

METHODS SUMMARY

Genomic DNA from 215 patients with primary neuroblastoma and 24 neuroblastoma-derived cell lines was analysed on GeneChip SNP genotyping microarrays (Affymetrix GeneChip 250K *Nsp*I). After appropriate normalization of mean array intensities, signal ratios were calculated between tumours and anonymous normal references in an allele-specific manner, and allele-specific copy numbers were inferred from the observed signal ratios based on the hidden Markov model using CNAG/AsCNAR software[13,14]. *ALK* mutations were examined by DNA heteroduplex analysis and/or genomic DNA sequencing[16]. Full-length cDNAs for mutant *ALK* were isolated by high-fidelity PCR and inserted into pcDNA3 and pMXS. The expression plasmids were transfected into NIH3T3 cells using Effectene Transfection Reagent (Qiagen) or by calcium phosphate methods[3]. Western blot analysis of mutant ALK kinases, *in vitro* kinase assays, and tumour formation assays in nude mice were performed as previously described[3]. This study was approved by the ethics boards of the University of Tokyo and of the Chiba Cancer Center Research Institute.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 June; accepted 28 August 2008.

1. Maris, J. M., Hogarty, M. D., Bagatell, R. & Cohn, S. L. Neuroblastoma. *Lancet* **369**, 2106–2120 (2007).
2. Maris, J. M. *et al.* Loss of heterozygosity at 1p36 independently predicts for disease progression but not decreased overall survival probability in neuroblastoma patients: a Children's Cancer Group study. *J. Clin. Oncol.* **18**, 1888–1899 (2000).
3. Attiyeh, E. F. *et al.* Chromosome 1p and 11q deletions and outcome in neuroblastoma. *N. Engl. J. Med.* **353**, 2243–2253 (2005).
4. Bown, N. *et al.* Gain of chromosome arm 17q and adverse outcome in patients with neuroblastoma. *N. Engl. J. Med.* **340**, 1954–1961 (1999).
5. Brodeur, G. M., Seeger, R. C., Schwab, M., Varmus, H. E. & Bishop, J. M. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* **224**, 1121–1124 (1984).
6. Shiota, M. *et al.* Anaplastic large cell lymphomas expressing the novel chimeric protein p80NPM/ALK: a distinct clinicopathologic entity. *Blood* **86**, 1954–1960 (1995).
7. Morris, S. W. *et al.* Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* **263**, 1281–1284 (1994).
8. Fujimoto, J. *et al.* Characterization of the transforming activity of p80, a hyperphosphorylated protein in a Ki-1 lymphoma cell line with chromosomal translocation t(2;5). *Proc. Natl Acad. Sci. USA* **93**, 4181–4186 (1996).
9. Soda, M. *et al.* Identification of the transforming *EML4–ALK* fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566 (2007).
10. Rikova, K. *et al.* Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131**, 1190–1203 (2007).
11. Kennedy, G. C. *et al.* Large-scale genotyping of complex DNA. *Nature Biotechnol.* **21**, 1233–1237 (2003).
12. Matsuzaki, H. *et al.* Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nature Methods* **1**, 109–111 (2004).
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. Yamamoto, G. *et al.* Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am. J. Hum. Genet.* **81**, 114–126 (2007).
15. Osajima-Hakomori, Y. *et al.* Biological role of anaplastic lymphoma kinase in neuroblastoma. *Am. J. Pathol.* **167**, 213–222 (2005).
16. Donohoe, E. Denaturing high-performance liquid chromatography using the WAVE DNA fragment analysis system. *Methods Mol. Med.* **108**, 173–187 (2005).
17. Hu, J., Liu, J., Ghirlando, R., Saltiel, A. R. & Hubbard, S. R. Structural basis for recruitment of the adaptor protein APS to the activated insulin receptor. *Mol. Cell* **12**, 1379–1389 (2003).
18. Donella-Deana, A. *et al.* Unique substrate specificity of anaplastic lymphoma kinase (ALK): development of phosphoacceptor peptides for the assay of ALK activity. *Biochemistry* **44**, 8533–8542 (2005).
19. McDermott, U. *et al.* Genomic alterations of anaplastic lymphoma kinase may sensitize tumors to anaplastic lymphoma kinase inhibitors. *Cancer Res.* **68**, 3389–3395 (2008).
20. Garraway, L. A. *et al.* Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117–122 (2005).
21. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758–764 (2007).
22. Kawamata, N. *et al.* Molecular allelokaryotyping of pediatric acute lymphoblastic leukemias by high-resolution single nucleotide polymorphism oligonucleotide genomic microarray. *Blood* **111**, 776–784 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. P. Koeffler for critically reading and editing the manuscript. We also thank M. Matsumura, Y. Ogino, S. Ichimura, S. Sohma, E. Matsui, Y. Yin, N. Hoshino and Y. Nakamura for their technical assistance. This work was supported by the Core Research for Evolutional Science and Technology, Japan Science and Technology Agency and by a Grant-in-Aid from the Ministry of Health, Labor and Welfare of Japan for the third-term Comprehensive 10-year Strategy for Cancer Control.

Author Contributions Y.C., Y.L.C. and J.T. contributed equally to this work. M.K. and M.Sa. performed microarray experiments and subsequent data analyses. Y.C. and J.T. performed mutation analysis of *ALK*. Y.C., Y.L.C., J.T., M.So., L.W. and H.M. conducted functional assays of mutant ALK. A.N., M.O., T.I., A.K. and Y.H. prepared tumour specimens and were involved in statistical analysis. A.N., Y.H., H.M., J.T. and S.O. designed the overall study, and S.O. and J.T. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information The nucleotide sequences of *ALK* mutations detected in this study have been deposited in GenBank under the accession numbers EU788003 (K1062M), EU788004 (T1087I), EU788005 (F1174L; TTC/TTA), EU788006 (F1174L; TTC/TTG), EU788007 (F1174C), EU788008 (F1174V), EU788009 (F1245L) and EU788010 (R1275Q). The copy number data as well as the raw microarray data will be accessible from http://www.ncbi.nlm.nih.gov/geo/ with the accession number GSE12494. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.O. (sogawa-tky@umin.net) or Y.H. (hayashiy-tky@umin.ac.jp).

METHODS

Specimens. Primary neuroblastoma specimens were obtained during surgery or biopsy from patients who were diagnosed with neuroblastoma and admitted to a number of hospitals in Japan. In total, 215 primary neuroblastoma specimens were subjected to SNP array analysis after informed consent was obtained from the parents of each patient. The patients were staged according to the International Neuroblastoma Staging System[23]. The clinicopathological findings are summarized in Supplementary Table 1. Twenty-four neuroblastoma-derived cell lines were also analysed by SNP array analysis (Supplementary Table 2). The SCMC-N2, SCMC-N4 and SCMC-N5 cell lines were established in our laboratory[24,25]. The SJNB series of cells and the UTP-N-1[26] cell line were gifts from A. T. Look and A. Inoue, respectively. The other cell lines used were obtained from the Japanese Cancer Resource Cell Bank (http://cellbank.nibio.go.jp/).

Microarray analysis. High molecular mass DNA was isolated from tumour specimens as well as from the peripheral blood or the bone marrow as described previously[24]. The DNA was subjected to SNP array analysis using Affymetrix GeneChip Mapping 50K and/or 250K arrays (Affymetrix) according to the manufacturer's suggested protocol. The scanned array images were processed with Gene Chip Operation software (GCOS)[13], followed by SNP calls using GTYE. Genome-wide copy number measurements and loss of heterozygosity detection were performed using CNAG/AsCNAR algorithms[14], which enabled an accurate determination of allele-specific copy numbers.

Confirmation of SNP array data. FISH and/or genomic PCR analysis confirmed the results of SNP array analyses as described previously[13]. PCR primer sets were designed to amplify several adjacent fragments inside and outside of the homozygously deleted regions in tumour samples.

Mutation analysis. Mutations in the *ALK* gene were examined in 239 neuroblastoma samples, including 24 cell lines, by denaturing high-performance liquid chromatography (DHPLC) using the WAVE system (Model 4500; Transgenomic) according to the manufacturer's suggested protocol[16]. The samples showing abnormal conformations were subjected to direct sequencing analysis using an ABI PRISM 3100 Genetic Analyser (Applied Biosystems). Using direct sequencing, mutation analysis of *MYCN* was also performed in seven cases with *ALK* alterations but not *MYCN* amplification. The primer sets used in this study are listed in Supplementary Table 5.

Transforming potential of *ALK* mutants. Total RNA was extracted from SJNB-1 (wild type), SCMC-N2 (F1174L) and SCMC-N5 (K1062M) cells as described previously[24]. First-strand cDNA was synthesized from RNA using Transcriptor Reverse Transcriptase and an oligo (dT) primer (Roche Applied Science). The resulting cDNA was then amplified by PCR using the KOD-Plus-Ver.2 DNA polymerase (Toyobo) and the primers sense 5′-TCAGAAGCTTTACCAAGGACTGTTCAGAGC-3′ and antisense 5′-AATTGCGGCCGCTACTTGTCATCGTCGTCCTTGTAGTCGGGCCCAGGCTG GTTCATGC-3′, thereby introducing a HindIII site at the 5′ terminus and a NotI site and a Flag sequence at the 3′ terminus. The HindIII–NotI fragments of *ALK* cDNA were subcloned into pcDNA3 to generate expression plasmids. After resequencing to confirm that they had no other mutations, the *ALK* plasmids were used for transfection into NIH3T3 cells using Effectene Transfection Reagent (Qiagen) according to the suggested manufacturer's protocol. The transfected NIH3T3 cells were selected in 800 μg ml^{-1} G418 for 2 weeks to obtain stably expressing clones.

To evaluate the phosphorylation status of ALK mutants, the cell lysates of stable clones were immunoprecipitated with antibodies to Flag (Sigma) and the resulting precipitates were subjected to western blot analysis with the antibody specific to pTyr 1604 (Cell Signaling Technology) of ALK and the generic anti-phosphotyrosine antibody (PY20). The *in vitro* kinase activity of ALK mutants was measured using a non-radioactive isotope solid-phase enzyme-linked immunosorbent assay using the Universal Tyrosine Kinase Assay kit (Takara) according to the manufacturer's suggested protocol. We also performed the *in vitro* kinase assay with the synthetic YFF peptide (Operon Biotechnologies) as described previously[18]. For anchorage-independent growth analysis, 1×10^3 stably transfected NIH3T3 cells were mixed in 0.3% agarose with 10% FBS-DMEM and plated on 0.6% agarose-coated 35-mm dishes. After culture for 28 days, the colonies of >0.1 mm in diameter were counted. The quantification of the colonies was from three independent experiments. To investigate the downstream signalling of ALK, western blot analysis was performed using the anti-ERK1/2, anti-phospho-ERK1/2, anti-AKT, anti-phospho-AKT, anti-STAT3 and anti-phospho-STAT3 antibodies (Cell Signaling Technology)[15].

The cDNA mutant of *ALK* was also inserted into the pMX5 plasmid and the constructs were introduced into NIH3T3 cells by the calcium phosphate method as described previously[9]. The cells were then either cultured for 21 days or injected subcutaneously at six sites in three nude mice.

Inhibition of ALK through RNAi-mediated knockdown. To suppress the expression of the ALK protein, two different pairs of *ALK* siRNAs (*ALK* siRNA1 and *ALK* siRNA2) were obtained (Qiagen)[15]. The sequences were 5′-GAGUCUGGCAGUUGACUUCdTdT-3′ for *ALK* siRNA1 and 5′-GCUCCGGCGUGCCAAGCAGdTdT-3′ for *ALK* siRNA2. A siRNA, targeting a sequence in firefly (*Photinus pyralis*) luciferase mRNA (*luc* siRNA), was used as a negative control (Qiagen)[15]. The sequences of *luc* siRNA were as follow: sense 5′-CGUACGCGGAAUACUUCGAdTdT-3′ and antisense 5′-UCGAAGUAUUCCGCGUACGdTdT-3′. Gene knockdown was achieved in SK-N-SH and LAN-2 cells using HiPerFect transfection reagent following the manufacturer's suggested instructions (Qiagen). To assess the effect of ALK knockdown on cell growth, these cells were seeded in 96-well plates at a concentration of 8.0×10^3 cells per well 24 h before transfection and assayed using the Cell Counting Kit-8 (Wako).

Statistical analysis. The significance of the correlation between *MYCN* amplification and *ALK* mutation was tested according to the conventional 2×2 contingency table using Fisher's exact test. The significance of the differences in kinase activity between wild-type and mutant ALK kinases was examined by the Mann–Whitney *U*-test based on the measured percentage activity of kinases in the precipitates of the corresponding samples. The significance of the differences in colony formation between wild-type and mutant ALK kinases was also examined by the Mann–Whitney *U*-test. The size of the hazards from possible risk factors, including International Neuroblastoma Staging System stages, *MYCN* status and *ALK* mutation/amplification were estimated by Cox regression analysis assuming a proportional hazard model using Stata software. Correlation between ploidy and clinical stage was tested by nptrend test.

23. Smith, E. I., Haase, G. M., Seeger, R. C. & Brodeur, G. M. A surgical perspective on the current staging in neuroblastoma—the International Neuroblastoma Staging System proposal. *J. Pediatr. Surg.* 24, 386–390 (1989).
24. Takita, J. *et al.* Allelotype of neuroblastoma. *Oncogene* 11, 1829–1834 (1995).
25. Takita, J. *et al.* Absent or reduced expression of the caspase 8 gene occurs frequently in neuroblastoma, but not commonly in Ewing sarcoma or rhabdomyosarcoma. *Med. Pediatr. Oncol.* 35, 541–543 (2000).
26. Takita, J. *et al.* Allelic imbalance on chromosome 2q and alterations of the caspase 8 gene in neuroblastoma. *Oncogene* 20, 4424–4432 (2001).

Chromosome copy number analysis in screening for prognosis-related genomic regions in colorectal carcinoma

Kentaro Kurashina,[1,2] Yoshihiro Yamashita,[1] Toshihide Ueno,[1] Koji Koinuma,[2] Jun Ohashi,[3] Hisanaga Horie,[2] Yasuyuki Miyakura,[2] Toru Hamada,[1,2] Hidenori Haruta,[1,2] Hisashi Hatanaka,[1] Manabu Soda,[1] Young Lim Choi,[1] Shuji Takada,[1] Yoshikazu Yasuda,[2] Hideo Nagai[2] and Hiroyuki Mano[1,4,5]

[1]Division of Functional Genomics and [2]Department of Surgery, Jichi Medical University, Tochigi 329-0498; [3]Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo 113-0033; [4]CREST (Core Research for Evolutional Science and Technology), Japan Science and Technology Agency, Saitama 332-0012, Japan

(Received March 9, 2008/Revised May 12, 2008/Accepted May 13, 2008/Online publication June 28, 2008)

Colorectal carcinoma (CRC) remains the major cause of cancer death in humans. Although chromosomal structural anomaly is presumed to play an important role in the carcinogenesis of CRC, chromosomal copy number alterations (CNA) and loss of heterozygosity (LOH) have not yet been analyzed extensively at high resolution in CRC. Here we aim to identify recurrent CNA and LOH in human CRC with the use of single nucleotide polymorphism-typing microarrays, and to reveal their relevance to clinical outcome. Surgically resected CRC specimens and paired normal mucosa were obtained from a consecutive series of 94 patients with CRC, and both of them were subjected to genotyping with Affymetrix Mapping 50K arrays. CNA and LOH were inferred computationally on every single nucleotide polymorphism site by integrating the array data for paired specimens. Our large dataset reveals recurrent CNA in CRC at chromosomes 7, 8, 13, 18, and 20, and recurrent LOH at chromosomes 1p, 4q, 5q, 8p, 11q, 14q, 15q, 17p, 18, and 22. Frequent uniparental disomy was also identified in chromosomes 8p, 17p, and 18q. Very common CNA and LOH were present at narrow loci of <1 Mbp containing only a few genes. In addition, we revealed a number of novel CNA and LOH that were linked statistically to the prognosis of the patients. The precise and large-scale measurement of CNA and LOH in the CRC genome is efficient for pinpointing prognosis-related genome regions as well as providing a list of unknown genes that are likely to be involved in CRC development. (*Cancer Sci* 2008; 99: 1835–1840)

Colorectal carcinoma (CRC) remains the fourth most prevalent cancer and the second highest cause of cancer death in the USA.[1] The life expectancy of individuals with CRC is mainly dependent on the clinical stage when CRC is detected, and the current chemotherapeutic regimens can only marginally improve the prognosis of advanced cases.[2] To achieve better outcomes for such individuals, it would be desirable to identify and target cellular molecules involved in the carcinogenesis of CRC.

A variety of genetic alterations take place, in a defined order, during the development of CRC.[3] In addition to nucleotide sequence mutations and epigenetic abnormalities of genes, structural changes of chromosomes and chromosomal instability (CIN) are known to play a major role in the carcinogenesis of CRC.[4] Gene amplification may induce oncogenic activity in a subset of protooncogenes, such as *MYC*, *MYCN*, *ERBB2*, and *CCND1*. In contrast, deletion or truncation of tumor-suppressor genes may confer inactivation of their function. These chromosomal copy number alterations (CNA) can be as large as numerical anomaly of entire chromosomes, or as small as segmental amplification or deletion of <10 kb.

Further, loss of heterozygosity (LOH) of the genome is frequently present in cancer cells, where one allele of a chromosome is deleted (chromosome copy number of one) or the remaining allele is further duplicated (chromosome copy number of two), referred to as uniparental disomy (UPD). It has been hypothesized that such regions likely harbor mutated or epigenetically silenced tumor-suppressor genes. However, recent evidence indicates that these regions may also carry activated oncogenes, as demonstrated for mutated *JAK2* in myeloproliferative disorders.[5]

In addition to the conventional array-based comparative genomic hybridization (CGH) technique,[6,7] microarrays developed originally for single nucleotide polymorphism (SNP) typing are now being applied to CIN investigation.[8–10] The main advantage of the latter system over the former is that it readily screens CIN at very high resolution in an allele-specific manner. The SNP arrays are thus able to screen for both CNA and LOH throughout the genome.

A few studies have been conducted recently for the SNP array-based CIN analysis of CRC,[11–14] but the interpretation of such data may be hampered by the small number of clinical specimens and the lack of paired normal samples for the analysis (especially in the cases of LOH examination).

Here we have collected CRC and paired normal specimens from a total of 94 individuals with CRC, and hybridized each DNA to Affymetrix Mapping 50K Xba 240 microarrays (Affymetrix, Santa Clara, CA, USA), which are able to examine CNA and LOH at a mean resolution of 47.2 kb. Application of bioinformatics to these large datasets has identified a number of novel prognosis-related regions in the CRC genome.

Materials and Methods

Preparation of genomic DNA. Primary tumors and paired colonic mucosal specimens (as normal controls) were surgically resected and frozen from a total of 94 individuals with sporadic CRC (from January 2002 to March 2003 at Jichi Medical University Hospital). The clinical characteristics of these study subjects are summarized in Suppl. Table S1. Informed consent was obtained from each subject according to the protocols approved by the ethics committees of Jichi Medical University. Genomic DNA was extracted from the samples with the use of the QIAamp DNA Mini Kit (Qiagen, Valencia, CA, USA) according to manufacturer's instructions. The microsatellite instability (MSI) status of each tumor was determined on the basis of the analysis of nine microsatellite repeat loci as described previously.[15]

[5]To whom correspondence should be addressed. E-mail: hmano@jichi.ac.jp

doi: 10.1111/j.1349-7006.2008.00881.x

Hybridization with SNP-typing arrays. Each DNA sample (250 ng) was digested with *Xba*I, ligated to Adaptor-Xba (Affymetrix), amplified by polymerase chain reaction (PCR), and subjected to hybridization with Mapping 50K Xba 240 arrays (Affymetrix). SNP genotyping calls were generated using GDAS software version 3.0 (Affymetrix) with a confidence score threshold of 0.05. Chromosome copy number and allele-specific copy number at each SNP site were calculated from the hybridization signal intensity for both CRC and paired normal mucosa specimens with the use of CNAG software (http://www.genome.umin.jp).[8] Only the CNAG data for autosomes were analyzed in the present study. Genotype-call data and original CEL files are available at the Gene Expression Omnibus website (http://www.ncbi.nlm.nih.gov/geo) under the accession number GSE11417, and CNAG output data are available upon request. We considered chromosome copy number changes or LOH data reliable only when contiguous SNP probes presented the same data.

Quantitative real-time PCR. RNA was isolated from the samples with the use of an RNeasy Mini column (Qiagen) and was used to synthesize cDNA with PowerScript reverse transcriptase (Clontech, Palo Alto, CA, USA). Portions of genomic DNA or cDNA were subjected to PCR with the QuantiTect SYBR Green PCR Kit (Qiagen). The amplification protocol comprised incubations at 94°C for 15 s, 60°C for 30 s, and 72°C for 60 s. Incorporation of the SYBR Green dye into the PCR products was monitored in real time with the ABI PRISM 7700 sequence detection system (Applied Biosystems, Foster City, CA, USA), thereby allowing determination of the threshold cycle (C_T) at which exponential amplification of products begins. To quantitate the genomic DNA, the C_T values for genomic DNA corresponding to the glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene and the target regions were used to calculate the abundance of target regions relative to that of *GAPDH* DNA. The primer sequences used for PCR were: 5′-AGGACATTTGTAATCAGTATCTGTG-3′ and 5′-AGGGCAGTCAATAAGCTAAGGAA-3′ for period homolog 3 (*PER3*); 5′-CTCAACTTCCTTGAGCACCTCCTG-3′ and 5′-TACCTTGGACAGCTTGCTCTGTTG-3′ for invasion inhibitory protein 45 (*IIP45*); 5′-ACTGGTGCTCTCACTGTCCAAAAC-3′ and 5′-CGCAGAGTAGACATCCTGGGTAAA-3′ for FAT tumor suppressor homolog (*FAT*); 5′-AGCGAATGGAAGTTAAATTTTGG-3′ and 5′-TGCATCTGTCCTAACTCACTCCT-3′ for breast cancer cell 2 (*BRCC2*); 5′-AGAGACTGTATTGCAGGGTGAAGA-3′ and 5′-CTTTCCATTATATGTCCCGACTCC-3′ for v-maf musculoaponeurotic fibrosarcoma oncogene homolog K (*MAFK*); 5′-CTACTCTCTTGCCAGCATTTTCAC-3′ and 5′-ACCTAAGCCTTATCCACACCTCAC-3′ for protein tyrosine phosphatase, non-receptor type 1 (*PTPN1*); 5′-GCTCATAGCCCTGCCTTCCT-3′ and 5′-GGTCCCCAAAACGCACACTC-3′ for *CSMD1*; and 5′-CTGACCTGCCGTCTAGAAAAACCT-3′ and 5′-CAGGAAATGAGCTTGACAAAGTGG-3′ for *GAPDH*.

Similarly, the relative quantity of cDNA was calculated using the C_T value of PCR for each cDNA and that for the *GAPDH* cDNA. The primer sequences for reverse transcription (RT)-PCR were: 5′-CGGTTTTCTCACAACACATTAGCA-3′ and 5′-ACTGGAAGGTGGGAAATCAATAGG-3′ for *PER3* cDNA; 5′-CTGGAACTCAGGCAGCAGACAAG-3′ and 5′-GACTCCTGGGGGAGAACAGCATT-3′ for *IIP45* cDNA; 5′-GTGAGTAATCCGCGCTGTTCTTT-3′ and 5′-CAGTAGTTGGGACACTGGAAATGG-3′ for *FAT* cDNA; 5′-GACAGATTTCGCCCATTATTCAGG-3′ and 5′-TGTTTCTCTGCACAATTTGAACCA-3′ for *BRCC2* cDNA; 5′-GCCATATACCACTCTCCCTTCCAC-3′ and 5′-TGGAGTGTGCCTTGATTTCATACA-3′ for *CSMD1*; and 5′-GTCAGTGGTGGACCTGACCT-3′ and 5′-TGAGCTTGACAAAGTGGTCG-3′ for *GAPDH* cDNA. The primer sets for *MAFK* and *PTPN1* cDNA were the same ones used for genomic amplification of the corresponding genes.

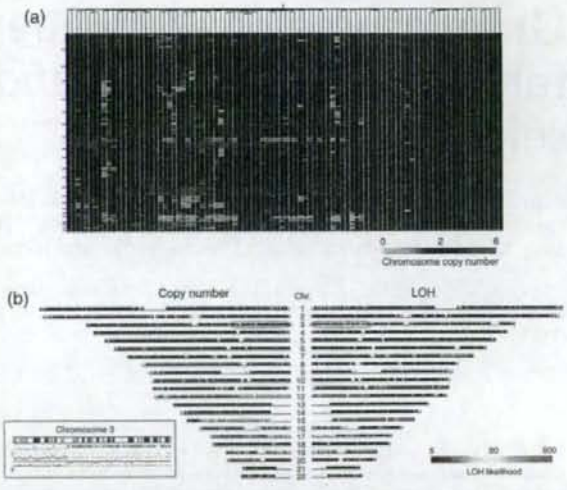

Fig. 1. Chromosomal copy number alterations and loss of heterozygosity (LOH) in the colorectal carcinoma genome. (a) The study subjects ($n = 94$) were subjected to a hierarchical clustering analysis based on the inferred copy number for all autosomal single nucleotide polymorphism (SNP) sites. Copy number is color coded according to the indicated scheme at the bottom. SNP sites are ordered by their physical position from top to bottom, and the borders between chromosomes are indicated by small bars at the left. (b) Chromosome copy number (left panel) and LOH likelihood score (right panel) are demonstrated for patient ID#002 in a chromosome view in a symmetrical manner. Copy number value is color coded as in (a), and LOH likelihood score is colored according to the scheme indicated at the bottom. Chromosome numbers are shown at the center. The allele-specific copy number data for the 3p region (indicated by a blue circle) is demonstrated in the inset as pink and green lines. Below the cytoband figure, the positions of SNP sites with a hetero- or discordant-call are indicated by green or pink bars, respectively.

Statistical analysis. Hierarchical clustering of the dataset and Student's *t*-test were carried out using GeneSpring 7.0 software (Agilent Technologies, Santa Clara, CA, USA), and survival analyses were carried out with SAS software (version 8.0.2: SAS Inc., Cary, NC, US) and the 'Survival' package in R version 2.6.0 (http://www.R-project.org.). The *q*-values for the false discovery rate were calculated directly from the ordered *P*-values above using the '*Q*-value' software (http://genomics.princeton.edu/storeylab/qvalue/) developed by Storey *et al.*[16] with parameters defined by Jones *et al.*[17]

Results

Frequent CAN. Genomic DNA was extracted from both CRC specimens and normal mucosa obtained from the same study subjects ($n = 94$). Both data were integrated into the CNAG software to infer chromosome copy number at every SNP site for each CRC sample. Incorporation of the data for paired normal mucosa markedly increased the accuracy of the calculation; the mean probe-signal intensity at diploid chromosomes in CRC was inferred from the data of control samples (where the majority of the chromosomes were expected to be diploid). Chromosome copy number data at each SNP probe site ($n = 57\,290$ for all autosomal SNP) was thus calculated for all CRC specimens, and a hierarchical clustering analysis for the study subjects was conducted based on the overall CNA profile. As shown in Figure 1a and Suppl. Fig. S1, approximately one-quarter of the subjects (the right side branch in the figure) had stable chromosomes, but the remaining samples had

doi: 10.1111/j.1349-7006.2008.00881.x

Table 1. Frequent regions of chromosomal copy number alterations or loss of heterozygosity (LOH) in colorectal carcinoma patients

Change	Chromosome	Nucleotide position	Mapped RefSeq gene	GenBank accession no.
Gain (chromosome copy no. ≥5 in ≥15 subjects)				
	6	16,176,003–16,176,549	None	
	8	70,887,465–71,089,425	*SLCO5A1*	NM_030958.1
	20	31,768,314–31,919,527	*PXMP4*	NM_007238.4
			ZNF341	NM_032819.3
			CHMP4B	NM_176812.3
Decrease (chromosome copy no. ≤1 in ≥35 subjects)				
	18	60,114,744–61,522,755	None	
	18	64,600,350–65,380,261	*CCDC102B*	NM_024781.1
			DOK6	NM_152721.2
	18	67,791,010–68,366,009	*CBLN2*	NM_182511.2
Homozygous deletion (common in two subjects)				
	3	60,393,402–60,490,818	*FHIT*	NM_002012.1
	20	14,796,659–15,040,864	*C20orf133*	NM_080676.5
LOH (common in ≥55 subjects)				
	5	108,765,615–112,484,272	*APC*	NM_000038.3
	17	5,265,130–8,883,455	*TP53*	NM_000546.3
			XAF1	NM_017523.2
			DVL2	NM_004422.2
	17	11,076,427–12,490,201	Others	

frequent CNA of various sizes. For instance, gross amplification was found commonly in chromosomes 7, 8q, 13, and 20, whereas large deletions of chromosomes were identified in 8p and 18.

Further, in-depth analysis of the dataset identified amplifications of various magnitudes at various frequencies. For instance, a high-grade amplification of the genome (copy number of five or greater) was found at three different loci in the genome of ≥15 subjects (Table 1), the size of which ranged from 547 to 201 961 bp. Surprisingly, amplification of one of these loci at chromosome 8q was found among as many as 25 patients (the most common, highly amplified region in our dataset). As expected, low-grade amplifications of the genome were found more commonly; a region of ~2.7 Mbp at chromosome 20q was, for example, amplified to four or more copies in more than 30 subjects, and this grade of amplification was also identified at many loci throughout the genome. For instance, genome regions with a copy number of four or greater in ≥10% of the patients were mapped to choromosmes 7p, 8q, 13, 20q, and others, comprising a total of 1921 SNP sites (3.4% of all sites).

Similarly, a decrease in chromosome copy number ($n \leq 1$) was also frequently identified throughout the genome; three distinct loci had such decreases in ≥35 subjects (Table 1). Further, a less-frequent decrease (found in ≥10% of patients) was mapped to chromosomes 1p, 5q, 8p, 14q, 17p, and others, comprising 3899 SNP sites in total (6.8% of all sites).

In our dataset, common homozygous deletions were unexpectedly rare. Only two loci demonstrated a chromosome copy number of zero in two individuals (Table 1). Interestingly, one such loci on chromosome 3 is known to be a common fragile region containing the fragile histidine triad gene (*FHIT*, GenBank accession no. NM_002012.1), a putative tumor suppressor.[18] The other homozygous deletion site at chromosome 20 spans 244 206 bp containing only one unknown gene, *C20orf133* (GenBank accession no. NM_080676.5).

Frequent LOH. With the SNP-typing array platform, we can carry out SNP genotyping by comparing the signal intensity between two alleles, which reflects the DNA amount of each allele. In the present study, with a moving window for 21 contiguous SNP, allele-specific copy number decreases were examined to identify LOH regions. Three most common LOH loci (found in 55 cases) were thus mapped to chromosomes 5 and 17 (Table 1). Other frequent LOH (found in ≥20% of patients) were identified on chromosomes 1p, 4q, 5q, 8p, 11q, 14q, 15q, 17p, 18, and 22. Less-frequent LOH were seen in two large loci (chromosomes 10 and 16), which contain a tumor necrosis factor receptor superfamily member (*FAS*, GenBank accession no. NM_000043.3) and ataxin-2 binding protein 1 (*A2BP1*, GenBank accession no. NM_018723.2).

Genome regions with UPD may contain tumor-suppressor genes (where both alleles carry a mutated, inactivated tumor-suppressor gene) or oncogenes (where cancer cells have two copies of a mutated and activated oncogene). CRC may have UPD at specific loci as demonstrated by Andersen *et al.*[19] In our dataset, we readily identified UPD regions that were characterized by a chromosome copy number of two and an LOH likelihood score of ≥50 defined by the CNAG software. In the data for patient ID# 002, for instance, a very high LOH likelihood score was inferred on chromosomes 3p and 15q (right panel of Fig. 1b). Although the latter region of the genome had a decreased copy number (left panel), the former was supposed to be diploid, indicating the presence of UPD. SNP array-based analysis can measure in detail changes in copy number in an allele-specific manner. With such analysis, as shown in the inset in Figure 1b, one allele at 3p was indeed amplified to a copy number of two (pink line), but the other allele was deleted (green line) in the same region, thus confirming the presence of UPD. Similar UPD was also identified on chromosomes 5q, 8p, 11, 14, 15, 17p, and 18q in our dataset.

Verification of the CNA data. The inferred copy number of chromosomes was then verified by quantitative real-time PCR. First, four genes (*BRCC2*, GenBank accession no. NM_001001786.1; *FAT*, GenBank accession no. NM_005245.3; *IIP45*, GenBank accession no. NM_021933.2; and *PER3*, GenBank accession no. NM_016831.1) mapped to independent loci with a frequent copy number loss were chosen to measure DNA quantity. The amount of DNA of each gene relative to that of *GAPDH* was examined in the patients with an inferred copy number of two and those with a copy number of one. As shown in the left panel of Figure 2a, the relative DNA amount of each gene was decreased to 0.7-0.5 in the patients with the copy number loss; 0.65 ± 0.52 (mean ± SD), 0.56 ± 0.26, 0.65 ± 0.22, and 0.73 ± 0.33 for *BRCC2*, *FAT*, *IIP45*, and *PER3*, respectively. The correlation coefficients between inferred copy number by

Table 2. Prognosis-related regions of copy number alterations (CNA) or loss of heterozygosity (LOH) in colorectal carcinoma patients

Change	Chromosome	Position	Size (Mbp)	*P*-value	*q*-value	Mapped RefSeq gene	GenBank accession no.
CNA	5	113,733,368–117,078,267	3.34	<0.001	0.095	*SEMA6A* and others	NM_20796.3 and others
	5	121,427,436–122,773,632	1.35	<0.001	0.095	*LOX* and others	NM_002317.3 and others
	5	123,233,993–126,057,451	2.82	<0.0005	0.095	Others	
	5	142,509,574–142,681,049	0.17	<0.001	0.095	Others	
	5	160,137,590–160,786,796	0.65	<0.001	0.098	Others	
	5	162,659,919–162,863,325	0.20	<0.0005	0.095	*CCNG1* and others	NM_004060.3 and others
	6	109,126,299–109,435,125	0.31	<0.0005	0.095	*SESN1* and others	NM_014454.1 and others
	10	110,674,541–111,338,259	0.66	<0.0005	0.095	None	
	18	51,345,876–52,332,836	0.99	<0.0005	0.095	*TCF4* and others	NM_003199.2 and others
	18	53,401,262–55,536,937	2.14	<0.0005	0.095	*RAX* and others	NM_013435.2 and others
LOH	16	4,858,366–6,679,934	1.82	<0.0001	0.046–0.224	*UBN1* and others	NM_002705.4 and others
	16	7,010,644–7,608,397	0.60	<0.0005	0.181	*A2BP1*	NM_018723.2

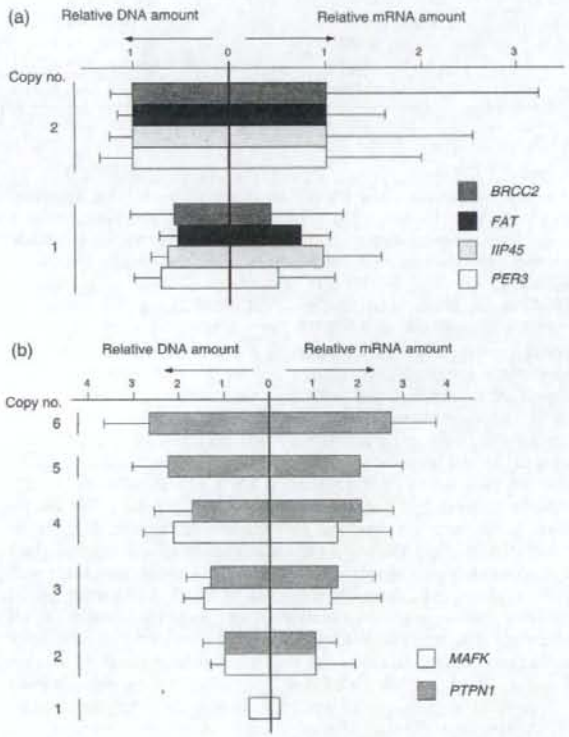

Fig. 2. Verification of copy number changes. (a) The DNA quantities of *PER3*, *IIP34*, *FAT*, and *BRCC2* (relative to that of *GAPDH*) were measured by real-time polymerase chain reaction in the subjects with inferred copy number two ($n = 2$) or one ($n = 1$). The mean + SD value for each gene was normalized to the corresponding mean value for the group with diploid chromosomes, and is shown in the left panel. The mRNA amount for each gene (relative to that of GAPDH) was also quantitated by real-time reverse transcription–polymerase chain reaction and is shown in a similar way. (b) The relative DNA (left panel) or mRNA (right panel) of *MAFK* and *PTPN1* was calculated as in (a).

array hybridization and DNA quantification by PCR were 0.219, 0.383, 0.216, and 0.314, respectively.

For the same gene set, we also examined how copy number changes affect mRNA level. Quantitative real-time reverse transcription-PCR was used to quantify the relative amount of each mRNA to that of GAPDH (right panel of Fig. 2a). Similar to the chromosome copy number, the mean mRNA level in the samples with copy number loss was decreased compared to the level in those without the loss; 0.47 ± 075, 0.80 ± 0.29, 0.97 ± 0.64, and 0.54 ± 0.69 for *BRCC2*, *FAT*, *IIP45*, and *PER3*, respectively. However, a relatively large SD in each mRNA amount indicates that transcriptional level was also influenced significantly by other factors such as epigenetic regulation and transcriptional factors.

We also measured the DNA amount of two genes (*MAFK*, GenBank accession no. NM_002360.3, and *PTPN*, GenBank accession no. NM_002827.2) that showed various levels of copy number amplification in our dataset. As shown in the left panel of Figure 2b, the calculated DNA amount of *MAFK* relative to that of *GAPDH* by quantitative PCR paralleled the copy number inferred from SNP arrays. Similarly, DNA quantity measured by real-time PCR for *PTPN1* generally followed the inferred copy number ($n = 2$–6). Again, the mRNA from each gene was quantified by real-time reverse transcription-PCR, revealing that amount of DNA significantly affects mRNA level (right panel).

Prognosis-related CNA and LOH. To directly search for CNA and LOH linked to the survival of patients, we utilized Cox's proportional-hazard regression analysis[20] coupled with the false-discovery rate correction on the copy number profile of all autosomal SNP sites.

From the CNA dataset, several loci at chromosomes 5, 6, 10, and 18 were proved to be significantly related to prognosis ($P < 0.001$ and $q < 0.1$) (Table 2). For all loci except one at chromosome 10, chromosomal loss had a negative impact on the outcome of the patients (Fig. 3a). One locus at chromosome 5 contains the cyclin G1 gene (*CCNG1*, GenBank accession no. NM_004060), which belongs to the cyclin gene superfamily. In contrast to the other cyclins, expression of *CCNG1* is stable throughout the cell cycle, and becomes activated in mouse cells by exposure to ionizing radiation, which also induces cell cycle arrest.[21] Further, disruption of WT1 function is linked to the downregulation of *CCNG1* expression.[22] These data together indicate a pro-apoptotic role for *CCNG1*, and our discovery of a relationship between loss of *CCNG1* and poor prognosis may imply a function of *CCNG1* as a tumor suppressor in CRC.

In addition to CNA analysis, we further searched for prognosis-related LOH with the following approach. There were many recurrent LOH regions in our dataset at various frequencies. We thus examined whether some of those recurrent alterations (observed in five or more samples) were preferentially present in the patients who died of CRC compared to those who survived in our observation period. For these potentially outcome-related genomic regions, prognosis was compared statistically between the two subject groups by the log-rank test with the false-discovery

doi: 10.1111/j.1349-7006.2008.00881.x

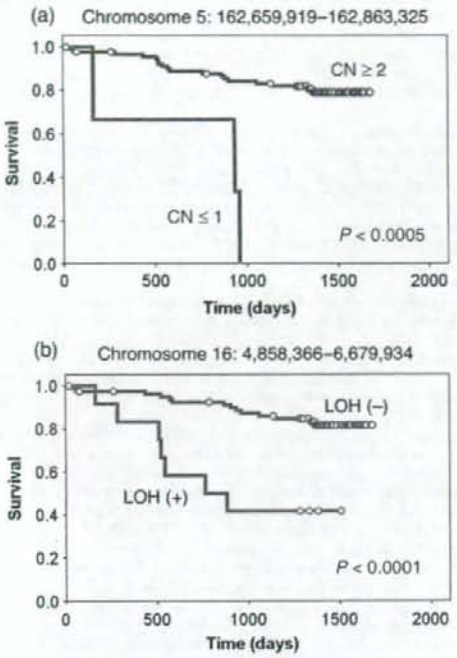

Fig. 3. Prognosis-related copy number (CN) loss and loss of heterozygosity (LOH). The survival of the subjects with or without copy number loss of a locus at (a) chromosome 5 or (b) chromosome 16 was compared using Kaplan–Meier analysis. The *P*-value for each comparison was calculated using the log rank test.

rate correction. We finally isolated two loci of LOH where the presence of LOH was related to a short survival time ($P < 0.0005$) (Table 2; Fig. 3b). One such prognosis-related LOH locus contains the ubinuclein 1 gene (*UBN1*, GenBank accession no. NM_016936). Because *UBN1* associates physically with *AP1* and interferes with its DNA-binding activity,[23] *UBN1* may also function to suppress tumor development.

Discussion

We have here calculated chromosome copy number as well as LOH likelihood throughout the genome of 94 CRC specimens. Together with the clinical information for the study subjects, we identified many loci whose DNA quantity or LOH is associated with the survival and various characteristics of CRC subjects. Some of the RefSeq genes mapped on such loci are well-known cancer-related genes. One frequent LOH was mapped to a genomic region of approximately 235 kb only containing the *MCC* gene, which had already been shown to be prone to somatic mutations and deletions in CRC and other cancers.[24,25] Overexpression of *MCC* suppresses the G_1 to S transition of the cell cycle, whereas such activity is lost for an *MCC* mutant identified in CRC,[26] supporting the tumor-suppressor activity of *MCC*.

In addition to the analysis presented in the present manuscript, our large dataset can also be utilized to characterize other aspects of CRC. CRC may be subdivided into microsatellite-stable cancer and MSI-high cancer. Comparison of our copy number data between the two subgroups has identified a locus of only 56 kb long, the copy number of which was statistically different between the subgroups ($P < 0.001$). This region contains only one RefSeq gene, ribosomal protein S6 kinase 90-kDa 5 (*RPS6KA5*, GenBank accession no. NM_004755.2), discovering another unexpected linkage between MSI and mitogen-activated protein kinase (MAPK) functions. A similar comparison of our data between the CRC with or without lymph node metastasis has identified two distinct loci in the genome (Suppl. Table S2). Also, a narrow genomic region was identified, the LOH of which is linked to the presence of liver metastasis ($P < 0.001$). However, that locus does not contain any RefSeq genes. Given the high resolution of SNP-typing arrays for CNA and LOH analysis, many genomic regions identified in this manuscript are <100 kb and contain only a few RefSeq genes per locus (Tables 1,2). Thus, our analysis is highly useful in narrowing down the list of genes associated with various characteristics of CRC.

Copy number alterations of CRC specimens have been studied with bacterial artificial chromosome array-based CGH,[13,27,28] and large segmental changes of chromosomes in such reports and publicly available databases match well with those identified in our study (see, for example, http://www.cghtmd.jp/CGHDatabase/tumor?lang=en). Although SNP-typing array-based CNA and LOH analyses have been reported recently for CRC, information for genes involved directly in such CNA and LOH is scarce.[11–14] Lips *et al.* examined the LOH status of paraffin-embedded CRC specimens ($n = 4$) and found recurrent LOH at chromosomes 5q, 17p, 18, and 20,[14] the former three of which were indeed identified in our study. However, Gaasenbeek identified LOH at the *TP53* locus in MSI-positive CRC.[13] In our cohort, however, there was only one MSI-positive case among 55 cases with LOH at *TP53*, whereas four were positive for MSI among 39 individuals without LOH at the locus, indicating no significant linkage between MSI and LOH at *TP53* (Fisher's exact test, $P = 0.186$).

It should, however, be noted that the RefSeq genes may not be the sole players in carcinogenesis. Long non-coding RNA is known to be involved in methylation of the genome,[29] and short non-coding RNA such as microRNA may be involved directly in cell growth and differentiation.[30] These transcripts, despite their inability to synthesize proteins, may thus contribute to the characteristics of CRC. As the discovery and annotation of these non-coding RNAs is still in its infancy,[31,32] many loci identified through our analysis may contain yet-undiscovered non-coding RNA, and these transcripts, not protein-coding mRNA, may play an important role in carcinogenesis as well. Indeed, one of the loci linked to lymph node metastasis has no RefSeq genes but only one non-coding RNA (Suppl. Table S2).

Our analysis provides a large-scale, accurate CNA and LOH dataset together with detailed information of clinical characteristics (including survival information in Suppl. Table S1) for the subjects. These data may become a framework for further analysis on structural alterations of the cancer genome in CRC.

Acknowledgments

The present study was supported in part by a Grant-in-Aid for Third-Term Comprehensive Control Research for Cancer from the Ministry of Health, Labor, and Welfare of Japan, and by a grant for 'High-Tech Research Center' Project for Private Universities: Matching Fund Subsidy, from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (2002-06) to HM.

References

1 Jemal A, Siegel R, Ward E *et al.* Cancer statistics, 2006. *CA Cancer J Clin* 2006; **56**: 106–30.

2 Portier G, Elias D, Bouche O *et al.* Multicenter randomized trial of adjuvant fluorouracil and folinic acid compared with surgery alone after resection of colorectal liver metastases: FFCD ACHBTH AURC 9002 trial. *J Clin Oncol* 2006; **24**: 4976–82.

3 Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990; **61**: 759–67.
4 Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature* 1998; **396**: 643–9.
5 Kralovics R, Passamonti F, Buser AS *et al.* A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med* 2005; **352**: 1779–90.
6 Jiang JK, Chen YJ, Lin CH, Yu IT, Lin JK. Genetic changes and clonality relationship between primary colorectal cancers and their pulmonary metastases – an analysis by comparative genomic hybridization. *Genes Chromosomes Cancer* 2005; **43**: 25–36.
7 Kleivi K, Teixeira MR, Eknaes M *et al.* Genome signatures of colon carcinoma cell lines. *Cancer Genet Cytogenet* 2004; **155**: 119–31.
8 Nannya Y, Sanada M, Nakazaki K *et al.* A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res* 2005; **65**: 6071–9.
9 Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* 2004; **20**: 1233–40.
10 Redon R, Ishikawa S, Fitch KR *et al.* Global variation in copy number in the human genome. *Nature* 2006; **444**: 444–54.
11 Andersen CL, Wiuf C, Kruhoffer M, Korsgaard M, Laurberg S, Orntoft TF. Frequent occurrence of uniparental disomy in colorectal cancer. *Carcinogenesis* 2007; **28**: 38–48.
12 Tsafrir D, Bacolod M, Selvanayagam Z *et al.* Relationship of gene expression and chromosomal abnormalities in colorectal cancer. *Cancer Res* 2006; **66**: 2129–37.
13 Gaasenbeek M, Howarth K, Rowan AJ *et al.* Combined array-comparative genomic hybridization and single-nucleotide polymorphism–loss of heterozygosity analysis reveals complex changes and multiple forms of chromosomal instability in colorectal cancers. *Cancer Res* 2006; **66**: 3471–9.
14 Lips EH, Dierssen JW, van Eijk R *et al.* Reliable high-throughput genotyping and loss-of-heterozygosity detection in formalin-fixed, paraffin-embedded tumors using single nucleotide polymorphism arrays. *Cancer Res* 2005; **65**: 10 188–91.
15 Miyakura Y, Sugano K, Konishi F *et al.* Extensive methylation of *hMLH1* promoter region predominates in proximal colon cancer with microsatellite instability. *Gastroenterology* 2001; **121**: 1300–9.
16 Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003; **100**: 9440–5.
17 Jones HE, Ohlssen DI, Spiegelhalter DJ. Use of the false discovery rate when comparing multiple health care providers. *J Clin Epidemiol* 2008; **61**: 232–40.
18 Zanesi N, Fidanza V, Fong LY *et al.* The tumor spectrum in FHIT-deficient mice. *Proc Natl Acad Sci USA* 2001; **98**: 10 250–5.
19 Andersen CL, Wiuf C, Kruhoffer M, Korsgaard M, Laurberg S, Orntoft TF. Frequent occurrence of uniparental disomy in colorectal cancer. *Carcinogenesis* 2006; **28**: 38–48.
20 Cox DR. Regression models and life tables. *J R Stat Soc* 1972; **34**: 187–220.
21 Sugihara T, Magae J, Wadhwa R *et al.* Dose and dose-rate effects of low-dose ionizing radiation on activation of Trp53 in immortalized murine cells. *Radiat Res* 2004; **162**: 296–307.
22 Wagner KJ, Patek CE, Miles C, Christie S, Brookes AJ, Hooper ML. Truncation of WT1 results in downregulation of cyclin G1 and IGFBP-4 expression. *Biochem Biophys Res Commun* 2001; **287**: 977–82.
23 Aho S, Buisson M, Pajunen T *et al.* Ubinuclein, a novel nuclear protein interacting with cellular and viral transcription factors. *J Cell Biol* 2000; **148**: 1165–76.
24 Kinzler KW, Nilbert MC, Vogelstein B *et al.* Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers. *Science* 1991; **251**: 1366–70.
25 Cawkwell L, Lewis FA, Quirke P. Frequency of allele loss of DCC, p53, RBI, WT1, NF1, NM23 and APC/MCC in colorectal cancer assayed by fluorescent multiplex polymerase chain reaction. *Br J Cancer* 1994; **70**: 813–18.
26 Matsumine A, Senda T, Baeg GH *et al.* MCC, a cytoplasmic protein that blocks cell cycle progression from the G_0/G_1 to S phase. *J Biol Chem* 1996; **271**: 10 341–6.
27 Fijneman RJ, Carvalho B, Postma C, Mongera S, van Hinsbergh VW, Meijer GA. Loss of 1p36, gain of 8q24, and loss of 9q34 are associated with stroma percentage of colorectal cancer. *Cancer Lett* 2007; **258**: 223–9.
28 Jones AM, Douglas EJ, Halford SE *et al.* Array-CGH analysis of microsatellite-stable, near-diploid bowel cancers and comparison with other types of colorectal carcinoma. *Oncogene* 2005; **24**: 118–29.
29 Chang SC, Tucker T, Thorogood NP, Brown CJ. Mechanisms of X-chromosome inactivation. *Front Biosci* 2006; **11**: 852–66.
30 Carrington JC, Ambros V. Role of microRNAs in plant and animal development. *Science* 2003; **301**: 336–8.
31 Carninci P, Kasukawa T, Katayama S *et al.* The transcriptional landscape of the mammalian genome. *Science* 2005; **309**: 1559–63.
32 Takada S, Berezikov E, Yamashita Y *et al.* Mouse microRNA profiles determined with a new and sensitive cloning method. *Nucleic Acids Res* 2006; **34**: e115.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Hierarchical clustering tree in Figure 1a is demonstrated with subject ID indicated at the bottom.

Table S1. Clinical characteristics of the study subjects

Table S2. Chromosomal copy number alterations (CNA) and loss of heterozygosity (LOH) related to clinical characteristics of colorectal carcinoma

Please note: Blackwell Publishing are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

doi: 10.1111/j.1349-7006.2008.00881.x

Identification of Novel Isoforms of the *EML4-ALK* Transforming Gene in Non–Small Cell Lung Cancer

Young Lim Choi,[1] Kengo Takeuchi,[3] Manabu Soda,[1,2] Kentaro Inamura,[3] Yuki Togashi,[3] Satoko Hatano,[3] Munehiro Enomoto,[1,2] Toru Hamada,[1] Hidenori Haruta,[1] Hideki Watanabe,[1] Kentaro Kurashina,[1] Hisashi Hatanaka,[1] Toshihide Ueno,[1] Shuji Takada,[1] Yoshihiro Yamashita,[1] Yukihiko Sugiyama,[2] Yuichi Ishikawa,[3] and Hiroyuki Mano[1,4]

Divisions of [1]Functional Genomics and [2]Pulmonary Medicine, Jichi Medical University, Tochigi, Japan; [3]Department of Pathology, The Cancer Institute, Japanese Foundation for Cancer Research, Tokyo, Japan; and [4]CREST, Japan Science and Technology Agency, Saitama, Japan

Abstract

The genome of a subset of non–small-cell lung cancers (NSCLC) harbors a small inversion within chromosome 2 that gives rise to a transforming fusion gene, *EML4-ALK*, which encodes an activated protein tyrosine kinase. Although breakpoints within *EML4* have been identified in introns 13 and 20, giving rise to variants 1 and 2, respectively, of *EML4-ALK*, it has remained unclear whether other isoforms of the fusion gene are present in NSCLC cells. We have now screened NSCLC specimens for other in-frame fusion cDNAs that contain both *EML4* and *ALK* sequences. Two slightly different fusion cDNAs in which exon 6 of *EML4* was joined to exon 20 of *ALK* were each identified in two individuals of the cohort. Whereas one cDNA contained only exons 1 to 6 of *EML4* (variant 3a), the other also contained an additional 33-bp sequence derived from intron 6 of *EML4* (variant 3b). The protein encoded by the latter cDNA thus contained an insertion of 11 amino acids between the EML4 and ALK sequences of that encoded by the former. Both variants 3a and 3b of EML4-ALK exhibited marked transforming activity *in vitro* as well as oncogenic activity *in vivo*. A lung cancer cell line expressing endogenous variant 3 of EML4-ALK underwent cell death on exposure to a specific inhibitor of ALK catalytic activity. These data increase the frequency of *EML4-ALK*–positive NSCLC tumors and bolster the clinical relevance of this oncogenic kinase. [Cancer Res 2008;68(13):4971–6]

Introduction

Lung cancer is the leading cause of cancer deaths in the United States, with >160,000 individuals dying of this condition in 2006 (1). The efficacy of conventional chemotherapeutic regimens with regard to improving clinical outcome in lung cancer patients is limited. Activating mutations within the epidermal growth factor receptor gene (*EGFR*) have been identified in non–small-cell lung cancer (NSCLC), the major subtype of lung cancer (2, 3), and chemical inhibitors of the kinase activity of EGFR have been found to be effective in the treatment of a subset of NSCLC patients harboring such mutations. However, these somatic mutations of *EGFR* are prevalent only among young women, nonsmokers, and Asian populations (3, 4).

We recently identified a novel transforming fusion gene, *EML4* (echinoderm microtubule–associated protein–like 4)-*ALK* (anaplastic lymphoma kinase), in a clinical specimen of lung adenocarcinoma from a 62-year-old male smoker (5). This fusion gene was formed as the result of a small inversion within the short arm of chromosome 2 that joined intron 13 of *EML4* to intron 19 of *ALK* (transcript ID ENST00000389048 in the Ensembl database[5]). The EML4-ALK protein thus contained the amino-terminal half of EML4 and the intracellular catalytic domain of ALK. Replacement of the extracellular and transmembrane domains of ALK with this region of EML4 results in constitutive dimerization of the kinase domain of ALK and a consequent increase in its catalytic activity (5).

Whereas this *EML4-ALK* fusion gene was detected in 3 of 75 individuals with NSCLC, we further identified another isoform of *EML4-ALK* in two patients of the same cohort (5). In these two individuals, intron 20 of *EML4* was disrupted and joined to intron 19 of *ALK*, with the fusion protein thus consisting of the amino-terminal two thirds of EML4 and the intracellular domain of ALK. This larger version of EML4-ALK was referred to as variant 2, with the original smaller version being termed variant 1. A total of 5 of the 75 (6.7%) patients in the cohort were thus positive for *EML4-ALK*.

Given that detection of *EML4-ALK* cDNA by the PCR would be expected to provide a highly sensitive means for diagnosis of lung cancer, and given that inhibition of the catalytic activity of EML4-ALK may be an effective approach to treatment of this disorder, we have examined whether other isoforms of EML4-ALK are associated with NSCLC. We now describe a third isoform of EML4-ALK (variant 3) that is smaller than variants 1 and 2.

Materials and Methods

PCR. This study was approved by the ethics committees of Jichi Medical University and The Cancer Institute of the Japanese Foundation for Cancer Research. Total cDNA of NSCLC specimens was synthesized with PowerScript reverse transcriptase (Clontech) and an oligo(dT) primer from total RNA purified with the use of an RNeasy Mini RNA purification kit (Qiagen). Reverse transcription-PCR (RT-PCR) to amplify the fusion point of *EML4-ALK* variant 3 mRNA was done with a QuantiTect SYBR Green kit (Qiagen) and the primers 5′-TACCAGTGCTGTCTCAATTGCAGG-3′ and 5′-TCTTGCCAGCAAAGCAGTAGTTGG-3′. A full-length cDNA for *EML4-ALK*

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

Requests for reprints: Hiroyuki Mano, Division of Functional Genomics, Jichi Medical University, 3311-1 Yakushiji, Shimotsukeshi, Tochigi 329-0498, Japan. Phone: 81-285-58-7449; Fax: 81-285-44-7322; E-mail: hmano@jichi.ac.jp.

doi:10.1158/0008-5472.CAN-07-6158

[5] http://www.ensembl.org/index.html

variant 3 was amplified from total cDNA of a NSCLC specimen (ID no. 2075) with PrimeSTAR HS DNA polymerase (Takara Bio) and the primers 5′-ACTCTGTCGGTCCGCTGAATGAAG-3′ and 5′-CCACGGTCTTAGGGATCCCAAGG-3′; PCR was done for 35 cycles of 98°C for 10 s and 68°C for 6 min. The fusion point of *EML4-ALK* in the genome was amplified by PCR with genomic DNA of NSCLC specimens, PrimeSTAR HS DNA polymerase, and the primers 5′-GGCATAAAGATGTCATCATCAACCAAGG-3′ and 5′-AGCTTGCTCAGCTTGTACTCAGGG-3′. The nucleotide sequences of the *EML4-ALK* variant 3a and 3b cDNAs have been deposited in DDBJ/EMBL/GenBank under accession nos. AB374361 and AB374362, respectively.

Fluorescence *in situ* hybridization. Fluorescence *in situ* hybridization (FISH) analysis of the fusion gene was done with archival pathology specimens and with bacterial artificial chromosomes containing genomic DNA corresponding to *EML4* or *ALK* and their flanking regions as probes. In brief, surgically removed lung cancer tissue was fixed in 20% neutral buffered formalin, embedded in paraffin, and sectioned at a thickness of 3 μm. The sections were placed on glass slides and processed with a Histology FISH Accessory Kit (DakoCytomation) before hybridization with the *EML4* and *ALK* probes and examination with a fluorescence microscope (BX61, Olympus).

Transforming activity of EML4-ALK variant 3. Analyses of the function of EML4-ALK variant 3 were done as described previously (5). In brief, the cDNA for EML4-ALK variant 3a or 3b was fused with an oligonucleotide encoding the FLAG epitope tag and then inserted into the retroviral expression plasmid pMXS (6). The resulting plasmids as well as similar pMXS-based expression plasmids for EML4-ALK variant 1, variant 1 (K589M), or variant 2 were individually introduced into mouse 3T3 fibroblasts by the calcium phosphate method for a focus formation assay and assay of tumorigenicity in nu/nu mice. The same set of EML4-ALK proteins was expressed in HEK293 cells and assayed for kinase activity *in vitro* with the YFF peptide (7).

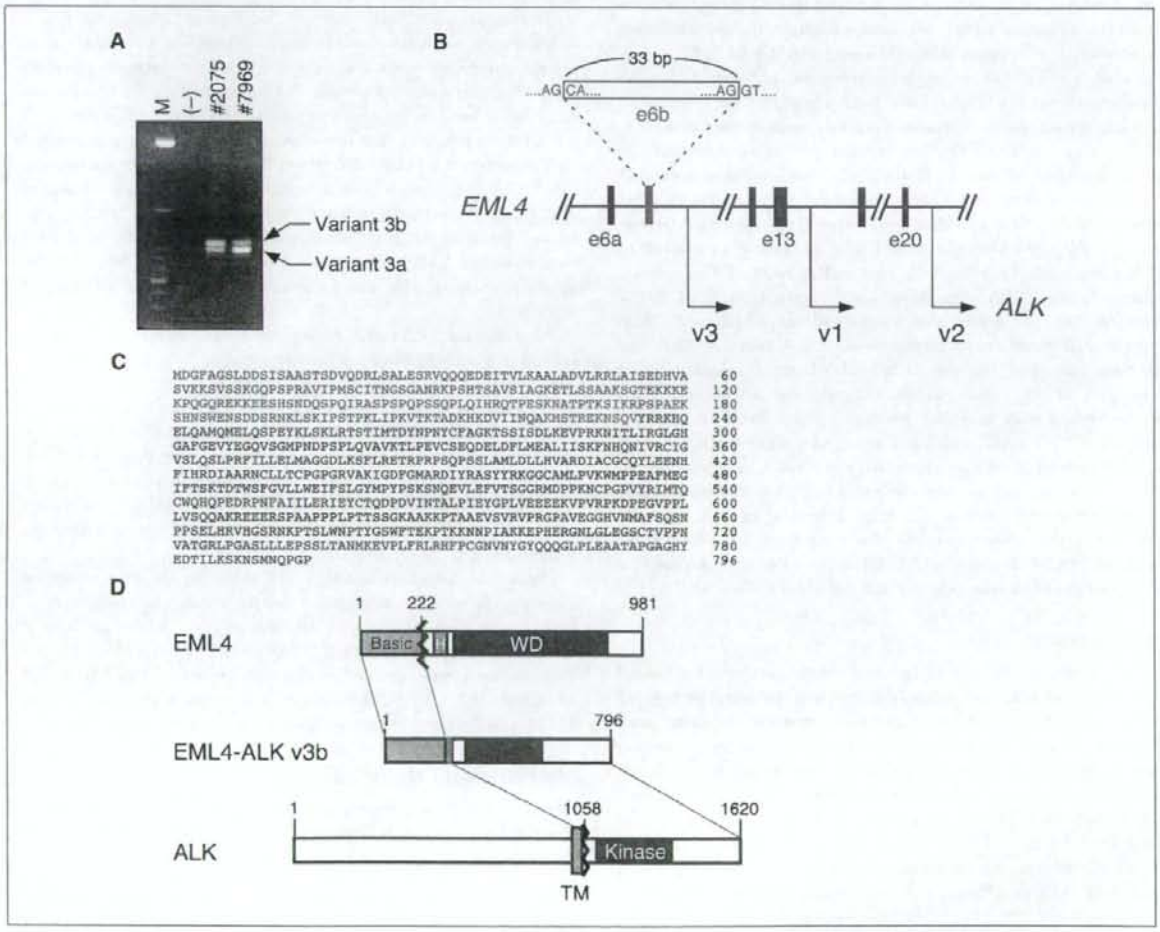

Figure 1. Identification of *EML4-ALK* variant 3. *A*, detection of fusion cDNAs linking exon 6 of *EML4* to exon 20 of *ALK* by RT-PCR analysis. Two RT-PCR products of 548 bp (corresponding to variant 3b) and 515 bp (corresponding to variant 3a) were detected by agarose gel electrophoresis with total RNA from two NSCLC specimens (tumor ID nos. 2075 and 7969). *Lane (−)*, no-template control; *lane M*, size markers (50-bp ladder). *B*, genomic organization of *EML4*. Intronic sequences downstream of exons (*e*) 6, 13, and 20 of *EML4* are fused to intron 19 of *ALK* to generate variants (*v*) 3, 1, and 2 of *EML4-ALK*, respectively. Exon-intron boundary sequences as well as the size of exon 6b are indicated. *C*, predicted amino acid sequence of EML4-ALK variant 3b. *Blue*, *green*, and *red*, amino acids corresponding to exons 1 to 6a of *EML4*, exon 6b of *EML4*, and *ALK*, respectively. Amino acid number is indicated on the right. *D*, fusion of an amino-terminal portion of EML4 [which consists of a basic region (*Basic*), HELP domain (*H*), and WD repeats] to the intracellular region of ALK (containing the tyrosine kinase domain) generates EML4-ALK variant 3b. *Green*, the region of the fusion protein encoded by exon 6b of *EML4*. *TM*, transmembrane domain.

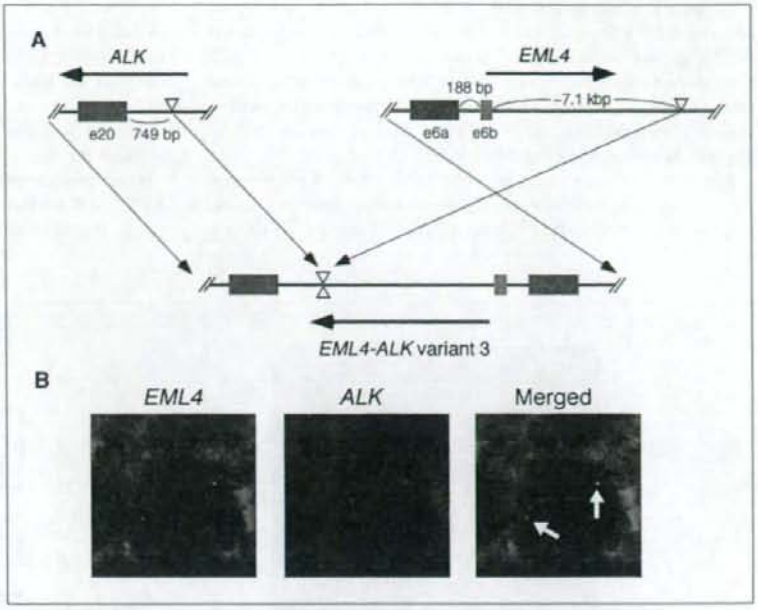

Figure 2. Chromosomal rearrangement responsible for generation of *EML4-ALK* variant 3. *A*, schematic representation of the chromosomal rearrangement underlying the generation of *EML4-ALK* variant 3. Exon 6b of *EML4* is located 188 bp downstream of exon 6a. In NSCLC specimen ID no. 7969, *EML4* is disrupted at a position ~7.1 kbp downstream of exon 6b and is ligated to a position 749 bp upstream of exon 20 of *ALK*, giving rise to the *EML4-ALK* (variant 3) fusion gene. *Horizontal arrows*, direction of transcription. *B*, FISH analysis of a representative cancer cell in a histologic section of lung adenocarcinoma (ID no. 7969) with differentially labeled probes for *EML4* (*left*) and *ALK* (*center*). Two fusion signals (*arrows*) and a pair of green (corresponding to *EML4*) and red (corresponding to *ALK*) signals are present in the merged image (*right*).

The cDNA for FLAG-tagged EML4-ALK variant 3b was also inserted into pMX-iresCD8 for the expression of both EML4-ALK and mouse CD8 (8), and the resulting recombinant retroviruses were used to infect mouse BA/F3 cells (9). CD8-positive cells were then purified with the use of a miniMACS magnetic bead–based separation system (Miltenyi Biotec) and cultured in the absence or presence of mouse interleukin-3 (IL-3; Sigma) or 2,4-pyrimidinediamine (Example 3-39, a specific inhibitor of ALK enzymatic activity that was developed by Novartis[6] and synthesized by Astellas Pharma).

Mouse 3T3 fibroblasts and NCI-H2228 lung cancer cells (both from American Type Culture Collection) as well as 3T3 cells expressing v-Ras were plated in 96-well spheroid culture plates (Celltight Spheroid, Sumilon) at a density of 1×10^3 per well. Cell growth was examined with the WST-1 Cell Proliferation Reagent (Clontech) after culture for 5 d with 2,4-pyrimidinediamine.

Luciferase reporter assays. The promoter fragments of *Fos*, *Myc*, and *Bcl-x$_L$* genes were ligated to a luciferase cDNA to generate pFL700 (10), pHXLuc (11), and pBclxl-Luc (12) reporter plasmids, respectively. Luciferase cDNA ligated to the DNA binding sequence for nuclear factor κB (NF-κB) or to the GAS sequence was obtained from Stratagene. HEK293 cells were transfected with these various reporter plasmids together with the expression plasmid for EML4-ALK variant 3b or the empty vector, as described previously (13). The pGL4 plasmid (Promega) for expression of *Renilla* luciferase was also included in each transfection mixture. After culture of the cells for 2 d, luciferase activity in cell lysates was measured with a Luciferase Assay system (Promega).

Results and Discussion

Detection of *EML-ALK* variant 3. The EML4-ALK variant 1 and 2 proteins are produced as a result of genomic rearrangements that lead to the juxtaposition of exons 13 and 20 of *EML4*, respectively, to exon 20 of *ALK*. It is theoretically possible that exon 2, 6, 18, or 21 of *EML4* also could undergo in-frame fusion to exon 20 of *ALK*. We therefore examined whether transcripts of any such novel *EML4-ALK* fusion genes are present in NSCLC cells by RT-PCR analysis with primers that flank each putative fusion point (data not shown). With the primer set for amplification of the *EML4* (exon 6)-*ALK* (exon 20) fusion cDNA, we detected a pair of PCR products in two individuals with lung adenocarcinoma (Fig. 1*A*). Although one of the patients (tumor ID no. 7969) had a smoking index of 540, the other patient (tumor ID no. 2075) had never smoked. Nucleotide sequencing of each PCR product from both patients revealed that the smaller product of 515 bp corresponded to a fusion cDNA linking exon 6 of *EML4* to exon 20 of *ALK*, whereas the larger product of 548 bp contained an additional sequence of 33 bp that was located between these exons of *EML4* and *ALK* and which mapped to intron 6 of *EML4* (Fig. 1*B*). The larger cDNA would thus be expected to encode a fusion protein with an insertion of 11 amino acids between the EML4 and ALK sequences of the protein encoded by the smaller cDNA.

Although we did not detect human mRNAs or expressed sequence tags containing this cryptic exon of *EML4* in the nucleotide sequence databases, it is likely that this exon is physiologic and functional because (*a*) the fusion cDNA containing this exon was identified in two independent patients and in amounts no less than those of the corresponding cDNA without it (Fig. 1*A*); (*b*) the intron-exon boundary sequence for this exon conforms well to the AG-GU rule for mRNA splicing (Fig. 1*B*); and (*c*) *EML4* cDNAs or expressed sequence tags containing this exon were detected in the sequence databases for other species (for instance, GenBank accession no. AK144604 corresponding to a mouse *EML4* cDNA). We thus refer to this cryptic exon as exon 6b and to the original exon 6 as exon 6a (Fig. 1*B*). The novel isoforms of *EML4-ALK* transcripts containing exons 1 to 6a or 1 to 6b of *EML4* were also designated variants 3a and 3b, respectively.

[6] Patent information: Garcia-Echeverria C, Kanazawa T, Kawahara E, Masuya K, Matsuura N, Miyake T, et al., inventors; Novartis AG, Novartis Pharma GmbH, IRM LLC, applicants. 2,4-Pyrimidinediamines useful in the treatment of neoplastic disease, inflammatory and immune system disorders. PCT WO 2005016894. 2005 Feb 24.

To isolate a full-length cDNA for EML4-ALK variant 3, we performed RT-PCR with total cDNA of a positive specimen (ID no. 2075) and with a sense strand primer targeted to the 5′ untranslated region (UTR) of *EML4* mRNA and an antisense strand primer targeted to the 3′ UTR of *ALK* mRNA. One-step PCR analysis yielded cDNA products for both *EML4-ALK* variants 3a and 3b (Fig. 1*C*; Supplementary Fig. S1).

The EML4 protein contains an amino-terminal basic domain followed by a hydrophobic echinoderm microtubule–associated protein–like protein (HELP) domain and WD repeats (14). Given that exons 1 to 6 of *EML4* encode the basic domain, the proteins encoded by the variant 3 cDNAs contain the entire basic domain of EML4 directly linked to the catalytic domain of ALK (Fig. 1*D*). The fact that the basic domain was found to be essential for both the self-dimerization and oncogenic activity of EML4-ALK (5) suggested that the variant 3 isoforms likely also possess transforming activity.

Chromosome rearrangement responsible for generation of *EML4-ALK* variant 3. To show the presence of a chromosome rearrangement responsible for the generation of *EML4-ALK* variant

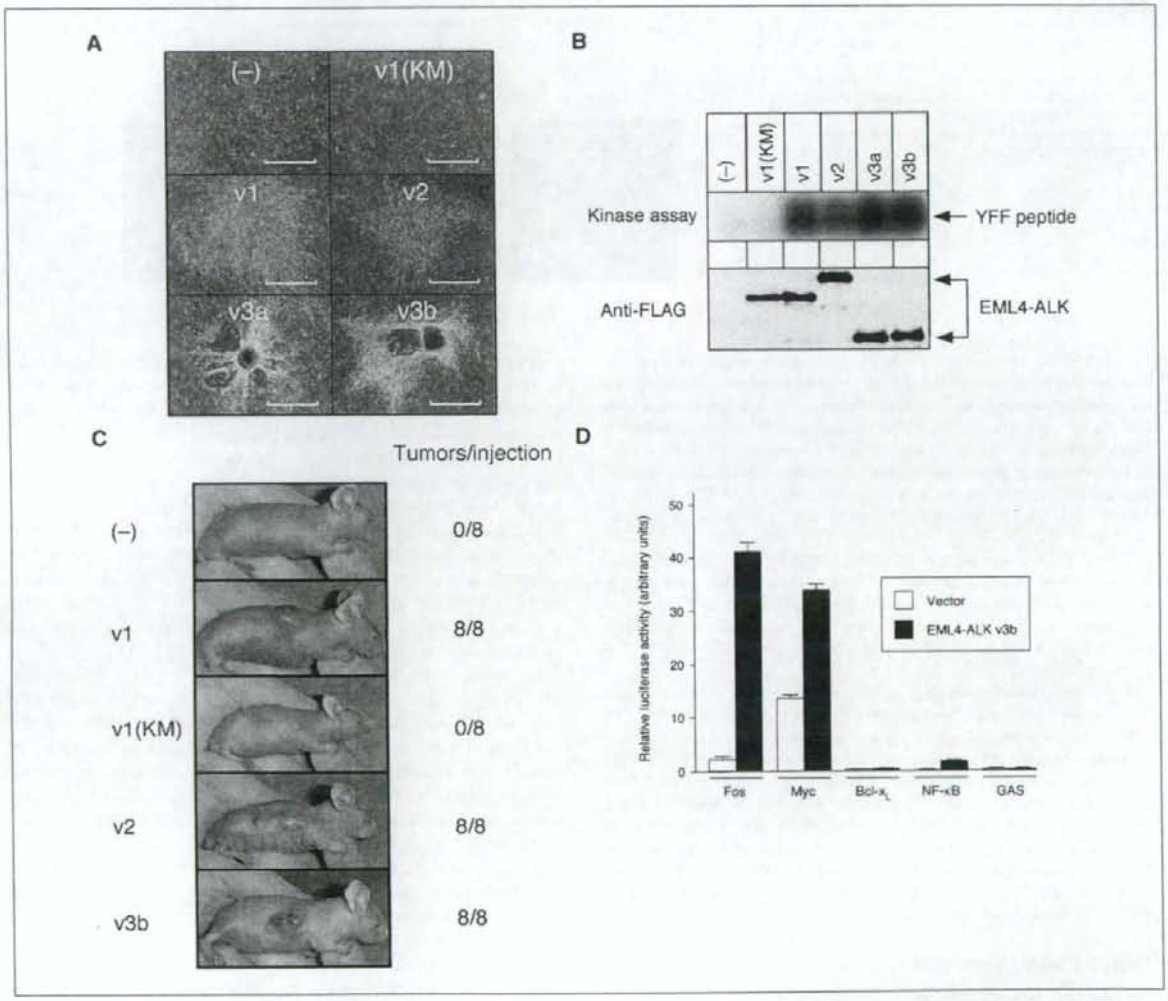

Figure 3. Transforming potential of EML4-ALK variants. *A*, focus formation assay. Mouse 3T3 fibroblasts were transfected with the empty expression plasmid [(–)] or with plasmids for wild-type (*v1*) or K589M mutant [*v1(KM)*] forms of variant 1, variant 2 (*v2*), variant 3a (*v3a*), or variant 3b (*v3b*) of FLAG-tagged EML4-ALK. The cells were photographed after culture for 18 d. *Bar*, 1 mm. *B*, *in vitro* kinase assay. HEK293 cells expressing the various FLAG-tagged variants of EML4-ALK were lysed and subjected to immunoprecipitation with antibodies to FLAG, and the resulting precipitates were assayed for kinase activity with the synthetic YFF peptide (*top*) or subjected to immunoblot analysis with antibodies to FLAG (*bottom*). *C*, *in vivo* assay of tumorigenicity. 3T3 cells expressing the indicated EML4-ALK variants were injected s.c. into nu/nu mice, and tumor formation was examined after 20 d. The number of tumors formed per eight injections is indicated on the right. *D*, analysis of EML4-ALK signaling with luciferase-based reporter plasmids. HEK293 cells were transfected with an expression plasmid for EML4-ALK variant 3b (or with the empty vector) together with reporter plasmids containing the promoter fragment of *Fos*, *Myc*, or *Bcl-x*$_L$ gene; the DNA binding sequence for NF-κB; or the GAS sequence. Cells were cultured for 2 d, lysed, and assayed for luciferase activity. The activity of firefly luciferase was normalized by that of *Renilla* luciferase. *Columns*, mean of three experiments; *bars*, SD.

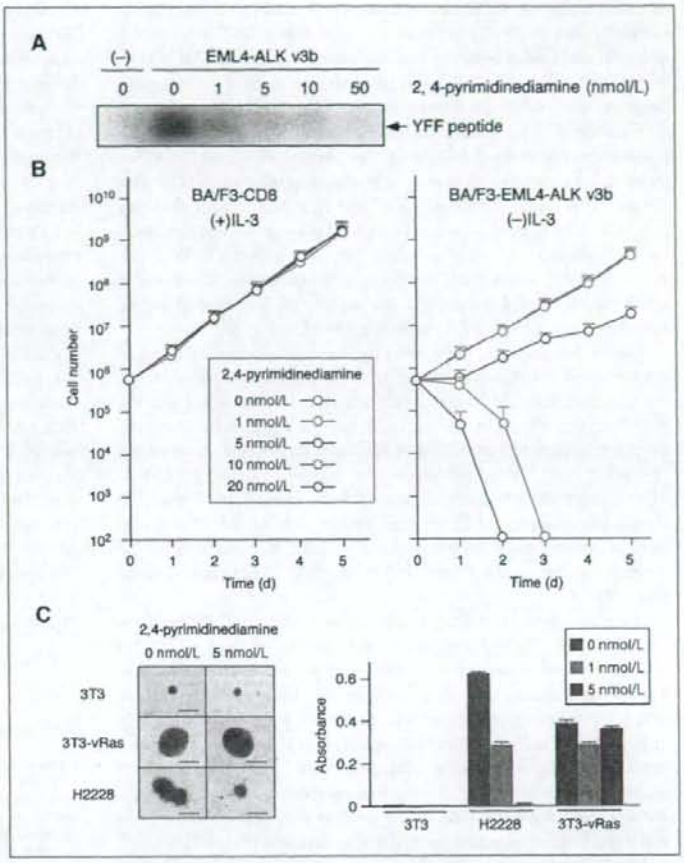

Figure 4. Essential role of EML4-ALK kinase activity in malignant transformation. *A*, lysates of HEK293 cells expressing FLAG-tagged EML4-ALK variant 3b (*v3b*) were divided into five equal portions, and each portion was subjected to immunoprecipitation with antibodies to FLAG. The immunoprecipitates were washed with kinase buffer [10 mmol/L HEPES-NaOH (pH 7.4), 50 mmol/L NaCl, 5 mmol/L $MgCl_2$, 5 mmol/L $MnCl_2$, 0.1 mmol/L Na_3VO_4] containing 0, 1, 5, 10, or 50 nmol/L of 2,4-pyrimidinediamine and then incubated for 30 min at room temperature for assay of kinase activity with the YFF peptide in the continued absence or presence of 2,4-pyrimidinediamine. The same amount of lysate of cells transfected with the empty vector was also subjected to immunoprecipitation and assayed as a negative control (–). *B*, mouse BA/F3 cells expressing CD8 alone were cultured in the presence of IL-3 (1 ng/mL) and the indicated concentrations of 2,4-pyrimidinediamine (*left*). BA/F3 cells expressing both CD8 and EML4-ALK variant 3b were cultured with the indicated concentrations of 2,4-pyrimidinediamine but without IL-3 (*right*). Cell number was counted at the indicated times. *Points*, mean of three separate experiments; *bars*, SD. *C*, mouse 3T3 fibroblasts expressing (or not) v-Ras or NCI-H2228 cells were cultured in a spheroid culture plate for 2 d, after which 2,4-pyrimidinediamine was added to the culture medium at a concentration of 0, 1, or 5 nmol/L. The cells were photographed after culture for an additional 5 d (*left*). *Bar*, 4 mm. Cell number in each well was also assessed at the same time with the use of the WST-1 assay (*right*). *Columns*, mean of three wells from a representative experiment; *bars*, SD.

3, we attempted to amplify the fusion point between the two genes from the genome of positive NSCLC cells. PCR with primers targeted to regions flanking the putative fusion point yielded a product of ~8 kbp with the genomic DNA of tumor ID no. 7969 (data not shown). Our failure to detect an unambiguous PCR product with genomic DNA of tumor ID no. 2075 may indicate that the breakpoint in intron 6 of *EML4* in this specimen is too distant from exon 6 to be readily amplified by PCR (intron 6 of *EML4* is >16 kbp). Nucleotide sequencing of the PCR product for tumor ID no. 7969 revealed that intron 6 of *EML4* was disrupted at a position −7.1 kbp downstream of exon 6b and was joined to a point 749 bp upstream of exon 20 of *ALK* (Fig. 2*A*).

We also confirmed the chromosome rearrangement involving *EML4* and *ALK* by FISH analysis of cells from tumor ID no. 7969 (Fig. 2*B*) and tumor ID no. 2075 (data not shown) with differentially labeled probes for the two genes. Both genes map to the short arm of chromosome 2 within a distance of −12 Mbp. The tumor cells exhibited fusion signals (corresponding to *EML4-ALK*) in addition to a pair of isolated green and red signals (corresponding to the two genes on the normal chromosome 2). The chromosome rearrangement involving the *ALK* locus was further verified with a different set of fluorescent probes (Supplementary Fig. S2).

Transforming activity of EML4-ALK variant 3. To compare the transforming potential of variants 1, 2, 3a, and 3b of EML4-ALK, we introduced expression plasmids for each variant into mouse 3T3 fibroblasts for assay of focus formation. No transformed foci were detected for cells transfected with the empty plasmid or with a plasmid for a kinase-inactive mutant (K589M) of EML4-ALK variant 1 (5) in which Lys^{589} in the ATP binding site of the catalytic domain is replaced with Met (Fig. 3*A*). In contrast, variants 3a and 3b of EML4-ALK each exhibited marked transforming activity that was not less than that of variant 1 or 2. To examine directly the tyrosine kinase activity of EML4-ALK variants, we subjected HEK293 cells expressing each of these variants to an *in vitro* kinase assay with a synthetic YFF peptide (7). Again, both variants 3a and 3b exhibited marked kinase activity that was not less than that of variant 1 or 2 (Fig. 3*B*). Similarly, in a tumorigenicity assay with nude mice, 3T3 cells expressing EML4-ALK variant 3b formed large subcutaneous tumors at all injection sites (Fig. 3*C*). Consistent with our previous observations (5), cells expressing variant 1 or 2 of EML4-ALK also formed tumors.

To examine the intracellular signaling pathways activated by EML4-ALK, we linked the luciferase cDNA to the promoter fragment of *Fos*, *Myc*, or *Bcl-x_L* gene (10–12); the DNA binding sequence for NF-κB; or the GAS sequence [a target site of the transcription factors signal transducers and activators of transcription (STAT)-1 and STAT3; ref. 15]. The resulting constructs were then introduced into HEK293 cells together with an

expression plasmid for EML4-ALK variant 3b. EML4-ALK variant 3b markedly activated the promoters of the *Fos* and *Myc* genes (Fig. 3*D*), consistent with the transforming potential of EML4-ALK. In contrast, although STAT3 has been shown to be a downstream target of the NPM-ALK fusion protein (16), EML4-ALK did not activate the GAS sequence, suggesting that STAT3 is unlikely to be a major target of EML4-ALK, as was shown in an EML4-ALK–positive lung cancer cell line by a proteomics approach (17). The distinct subcellular localizations of the two ALK fusion proteins [EML4-ALK in the cytoplasm (5) and NPM-ALK in both the nucleus and cytoplasm (18)] may account for this difference. Whereas EML4-ALK did not activate the *Bcl-x_L* gene promoter, it induced a small but significant increase in the activity of the NF-κB binding sequence ($P = 1.86 \times 10^{-4}$, Student's t test).

Several compounds have recently been identified as specific inhibitors of the kinase activity of ALK and as potential drugs for the treatment of lymphoma positive for *NPM-ALK* (19). We examined the effects of one such inhibitor, 2,4-pyrimidinediamine, on the transforming potential of EML4-ALK. We first determined the effect of this inhibitor on the kinase activity of EML4-ALK variant 3b immunoprecipitated from transfected cells. 2,4-Pyrimidinediamine inhibited the kinase activity of EML4-ALK in a concentration-dependent manner, with a concentration of 1 nmol/L reducing the kinase activity to <50% of the control value (Fig. 4*A*).

We also introduced EML4-ALK variant 3b and CD8 (or CD8 alone) into the IL-3–dependent hematopoietic cell line BA/F3 (9) and then purified the resulting CD8-positive cell populations. 2,4-Pyrimidinediamine, even at a concentration of 20 nmol/L, did not affect the IL-3–dependent growth of BA/F3 cells expressing only CD8 (Fig. 4*B*), indicating that this agent does not inhibit mitogenic signaling mediated by Janus kinase in BA/F3 cells. Expression of EML4-ALK rendered BA/F3 cells independent of IL-3 for growth, but the cells expressing the fusion protein also rapidly underwent cell death on exposure to 2,4-pyrimidinediamine (Fig. 4*B*).

Finally, we examined the effect of 2,4-pyrimidinediamine on lung cancer cells that express endogenous EML4-ALK variant 3. The human lung cancer cell line NCI-H2228 expresses EML4-ALK variants 3a and 3b (data not shown) and forms spheroids in a three-dimensional spheroid culture system (Fig. 4*C*; ref. 20). Whereas 3T3 fibroblasts are unable to form such spheroids, expression of v-Ras in these cells results in the formation of large spheroids in culture. Whereas 2,4-pyrimidinediamine did not affect the proliferation of 3T3 cells expressing v-Ras in this system, it inhibited the growth of NCI-H2228 cells in a concentration-dependent manner (Fig. 4*C*). These data thus indicate that EML4-ALK is essential for the growth of cancer cells expressing this oncokinase.

In conclusion, we have identified novel isoforms of *EML4-ALK* in two patients with NSCLC. A chromosome inversion within 2p was shown to connect intron 6 of *EML4* to intron 19 of *ALK* and to be responsible for the generation of fusion cDNAs connecting exons 1 to 6a or exons 1 to 6b of *EML4* to exon 20 of *ALK*. Given that fusion cDNAs with or without exon 6b of *EML4* were each present in the two patients, EML4-ALK variant 3a and 3b proteins are likely to be coexpressed in NSCLC cells. Although RT-PCR analysis to detect *EML4-ALK* may provide a highly sensitive means to detect lung cancer, it is important that all variant forms of the fusion gene be assayed with appropriately designed primer sets. Given that all the identified variants possess prominent transforming activity, the newly revealed increased incidence of *EML4-ALK* fusion in NSCLC further increases the importance of the fusion gene as a therapeutic target for this intractable disorder.

Disclosure of Potential Conflicts of Interest

K. Takeuchi: Consultant, DAKO. The other authors disclosed no potential conflicts of interest.

Acknowledgments

Received 11/8/2007; revised 3/3/2008; accepted 4/22/2008.

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

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.

We thank Takashi Aoki and Yasunobu Sugiyama for technical assistance.

References

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2006. CA Cancer J Clin 2006;56:106–30.
2. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 2004;350:2129–39.
3. Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004;304:1497–500.
4. Shigematsu H, Lin L, Takahashi T, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst 2005;97:339–46.
5. 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.
6. Onishi M, Kinoshita S, Morikawa Y, et al. Applications of retrovirus-mediated expression cloning. Exp Hematol 1996;24:324–9.
7. 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.
8. Yamashita Y, Kajigaya S, Yoshida K, et al. Sak serine/threonine kinase acts as an effector of Tec tyrosine kinase. J Biol Chem 2001;276:39012–20.
9. Palacious R, Steinmetz M. IL-3 dependent mouse clones that express B-220 surface antigen, contain Ig genes in germ-line configuration, and generate B lymphocytes *in vivo*. Cell 1985;41:727–34.
10. Hu Q, Milfay D, Williams LT. Binding of NCK to SOS and activation of ras-dependent gene expression. Mol Cell Biol 1995;15:1169–74.
11. Takeshita T, Arita T, Higuchi M, et al. STAM, signal transducing adaptor molecule, is associated with Janus kinase and involved in signaling for cell growth and c-myc induction. Immunity 1997;6:449–57.
12. Grillot DAM, Gonzalez-Garcia M, Ekhterae D, et al. Genomic organization, promoter region analysis, and chromosome localization of the mouse bcl-x gene. J Immunol 1997;158:4750–7.
13. Fujiwara S, Yamashita Y, Choi YL, et al. Transforming activity of purinergic receptor P2Y, G protein coupled, 8 revealed by retroviral expression screening. Leuk Lymphoma 2007;48:978–86.
14. Pollmann M, Parwaresch R, Adam-Klages S, Kruse ML, Buck F, Heidebrecht HJ. Human EML4, a novel member of the EMAP family, is essential for microtubule formation. Exp Cell Res 2006;312:3241–51.
15. Wesoly J, Szweykowska-Kulinska Z, Bluyssen HA. STAT activation and differential complex formation dictate selectivity of interferon responses. Acta Biochim Pol 2007;54:27–38.
16. 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.
17. Rikova K, Guo A, Zeng Q, et al. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. Cell 2007;131:1190–203.
18. Duyster J, Bai RY, Morris SW. Translocations involving anaplastic lymphoma kinase (ALK). Oncogene 2001;20:5623–37.
19. 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.
20. Kunita A, Kashima TG, Morishita Y, et al. The platelet aggregation-inducing factor aggrus/podoplanin promotes pulmonary metastasis. Am J Pathol 2007;170:1337–47.

H Tamai[1], Y Shioi[1], H Yamaguchi, M Okabe, S Wakita, T Mizuki, K Nakayama, K Inokuchi, K Tajika and K Dan
Department of Hematology, Nippon Medical School, Tokyo, Japan
E-mail: s6056@nms.ac.jp
[1]*These two authors contributed equally to this work.*

References

1 Taksin AL, Legrand O, Raffoux E, de Revel T, Thomas X, Contentin N *et al.* High efficacy and safety profile of fractionated doses of Mylotarg as induction therapy in patients with relapsed acute myeloblastic leukemia: a prospective study of the alfa group. *Leukemia* 2007; **21**: 66–71.
2 Mrozek K, Bloomfield CD. Chromosome aberrations, gene mutations and expression changes, and prognosis in adult acute myeloid leukemia. *Hematology 2006 Education program book.* American Society of Hematology: Washington, DC, 2006, 169–177.
3 Garrido SM, Bryant E, Appelbaum FR. Allogeneic stem cell transplantation for relapsed and refractory acute myeloid leukemia patients with 11q23 abnormalities. *Leuk Res* 2000; **24**: 481–486.
4 Larson RA, Sivers EL, Stadtmauer EA, Lowenberg B, Estey EH, Dombret H *et al.* Final report on the efficacy and safety of gemtuzumab ozogamicin (Mylotarg) in patients with CD33 positive acute myeloid leukemia in first recurrence. *Cancer* 2001; **104**: 1442–1452.
5 Muñoz L, Nomdedéu JF, Villamor N, Guardia R, Colomer D, Ribera JM *et al.* Acute myeloid leukemia with MLL rearrangements: clinicobiological features, prognostic impact and value of flow cytometry in the detection of residual leukemic cells. *Leukemia* 2003; **17**: 76–82.

MicroRNA expression profiles of human leukemias

Leukemia (2008) **22**, 1274–1278; doi:10.1038/sj.leu.2405031; published online 8 November 2007

MicroRNAs (miRNAs) are small noncoding RNAs of 20–24 nucleotides (nt) that negatively regulate the translation of target mRNAs through incomplete base-pairing with their 3′-untranslated regions.[1] Evidence indicates that miRNAs play an important role in the development of human cancers including leukemias, with one of the most well-characterized examples being association of miR-15a and miR-16a with chronic lymphocytic leukemia. Almost half of chronic lymphocytic leukemia patients harbor a chromosome deletion that encompasses 13q14, a region that includes the genes for miR-15a and miR-16a, and the abundance of these miRNAs is reduced in chronic lymphocytic leukemia cells with the chromosome deletion.[2] Several other miRNAs, such as miR-155 and miR-17-92, have also been implicated in the pathogenesis of lymphoma.[3] It is therefore important that the entire miRNA repertoire of clinical specimens be characterized and compared among various hematologic malignancies.

Reliable assessment of the global expression profiles of miRNAs, especially for the small amounts of clinical specimens available, is not straightforward, however. Microarray-based detection of miRNAs is prone to the generation of false-positive data that may result from mishybridization of probes, although improvements have recently been developed for this technology.[4] A large-scale cloning strategy would be an ideal approach to reliable estimation of the expression level of miRNAs, provided that a sufficient number of clones were to be analyzed. However, conventional methods for isolation of miRNAs require $>10\,\mu g$ of total RNA, which is not always obtainable from clinical specimens.

We recently developed a sensitive method, mRAP (micro RNA amplification profiling)[5] that readily allows the isolation of miRNA clones from $\leqslant 1 \times 10^4$ cells. To examine the miRNA expression profiles for leukemias with mRAP, we first purified $CD34^+$ cells from individuals ($n=12$) with *de novo* acute myeloid leukemia, acute myeloid leukemia secondary to myelodysplastic syndrome, acute lymphoid leukemia or biphenotypic acute leukemia (Table 1). Column affinity-chromatography to isolate $CD34^+$ cells yielded 10–50% of the input cells with a purity of $\geqslant 90\%$ as judged by flow cytometry (data not shown). As a normal control, we also purified a $CD34^+$ cell fraction from bone marrow mononuclear cells of a healthy volunteer. Then mRAP procedure was applied to 1.1×10^6–1.0×10^8 of the purified $CD34^+$ cells from each individual in order to obtain short RNA clones.

Sequencing and computer filtering[5] of the mRAP amplicons identified a total of 38858 qualified reads for the 13 study subjects. BLAST analysis then isolated 32867 reads that match the human genome sequence (ncbi 36 assembly), among which 2054 reads were mapped to transfer RNA genes, 2720 to ribosomal RNA genes and 9474 to repetitive sequences. From the remaining sequences, we identified 7191 reads corresponding to 143 independent known miRNAs (Supplementary Table 1). We further searched for candidate sequences corresponding to novel miRNAs whose surrounding genome sequences (of ~100 nt) potentially fold into a hairpin structure with a single notch. In this analysis, we did not exclude miRNA candidates that were not detected in the genomes of other

Table 1 Clinical characteristics of the study subjects

ID (no.)	*Age (years)*	*Sex*	*Sample origin*	*Disease*	*Karyotype*
3	64	M	PB	ALL	46,XY,t(9;22)
4	45	M	BM	AML (M4)	46,XY,inv(16)
7	78	F	BM	MDS-derived AML	46,XX
10	21	F	PB	AML (M0)	46,XX,t(9:15)
12	58	M	BM	AML (M2)	46,XY
32	43	M	BM	AML (M2)	46,XY,t(8;21)
33		M	PB	AML (M1)	46,XY
44	71	M	PB	MDS-derived AML	46,XY,t(8:21)
46	61	M	PB	AML (M2)	46,XY
47	61	M	BM	AML (M3)	46,XY,t(15;17)
48	29	M	PB	BAL	46,XY
49	58	M	PB	MDS-derived AML	46,XY

Abbreviations: ALL, acute myeloid leukemia; AML, acute lymphoid leukemia; BAL, biphenotypic acute leukemia; MDS, myelodysplastic syndrome; BM, bone marrow; F, female; M, male; PB, peripheral blood.