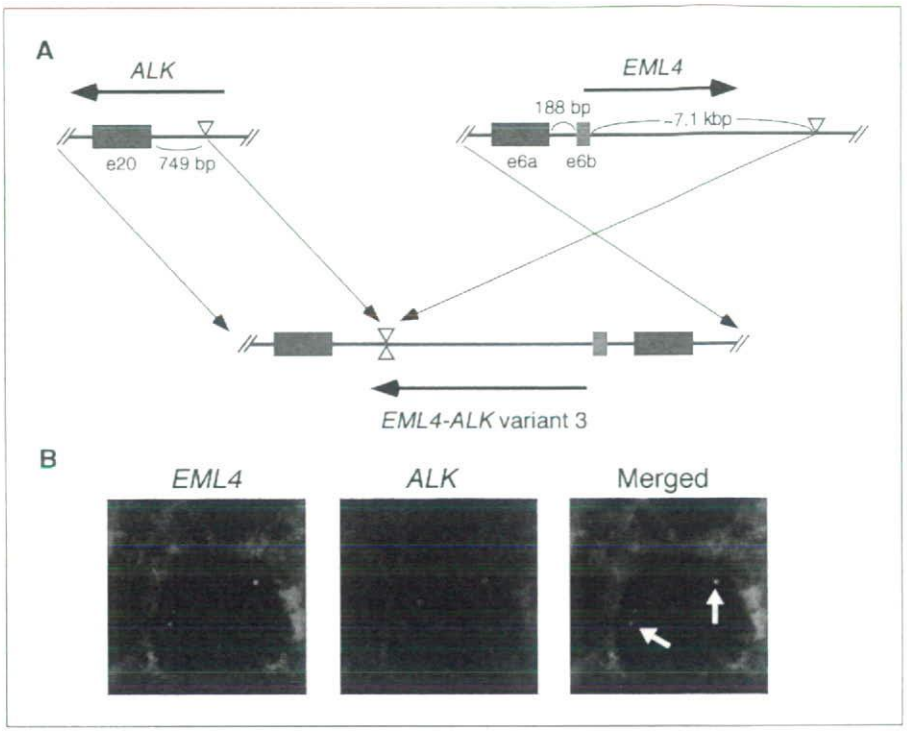

Figure 2. Chromosomal rearrangement responsible for generation of *EML4-ALK* variant 3. *A*, schematic representation of the chromosomal rearrangement underlying the generation of *EML4-ALK* variant 3. Exon 6b of *EML4* is located 188 bp downstream of exon 6a. In NSCLC specimen ID no. 7969, *EML4* is 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

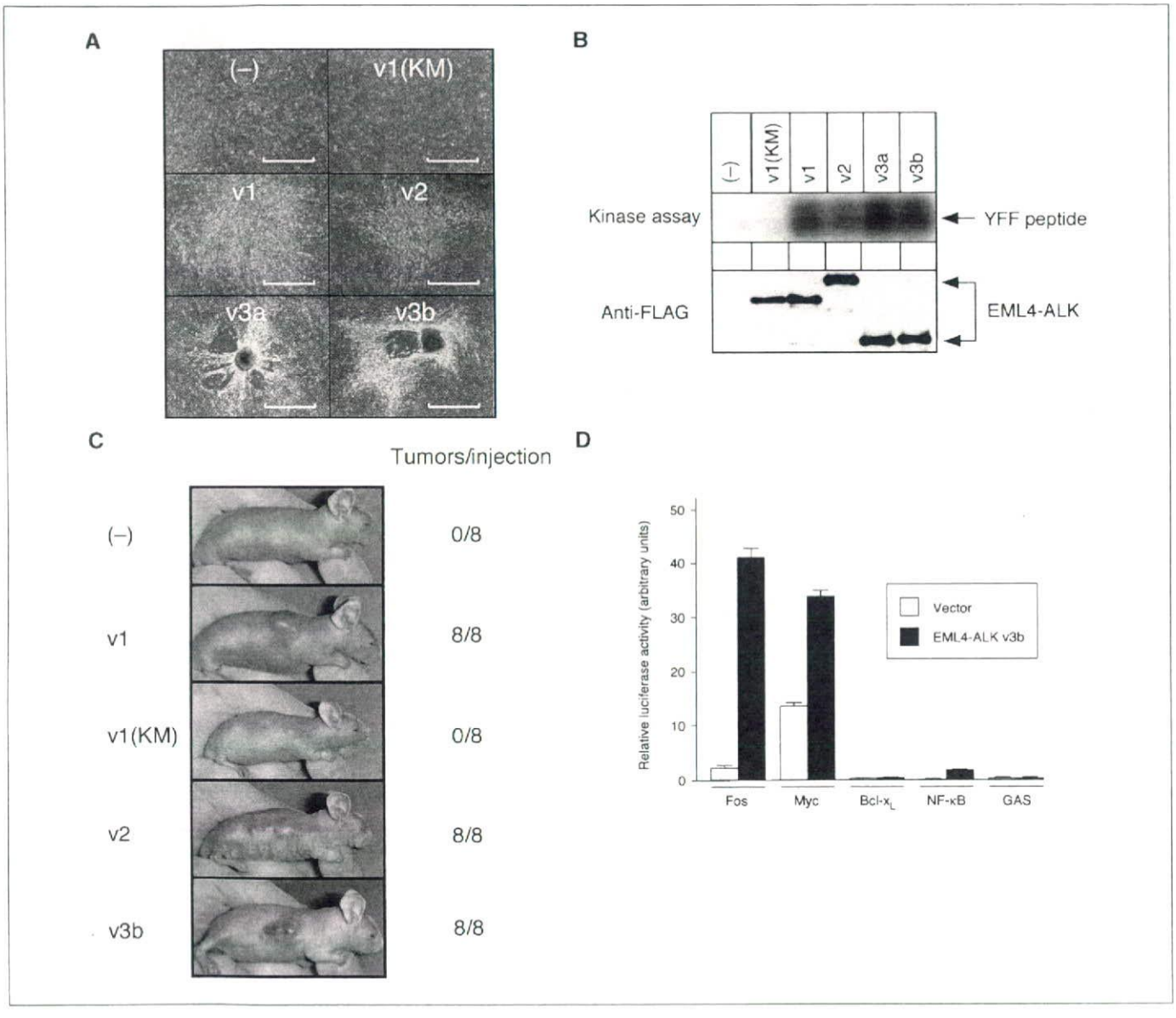

Figure 3. Transforming potential of EML4-ALK variants. *A*, focus formation assay. Mouse 3T3 fibroblasts were transfected with the empty expression plasmid [(–)] or with plasmids for wild-type (*v1*) or K589M mutant [*v1(KM)*] forms of variant 1, variant 2 (*v2*), variant 3a (*v3a*), or variant 3b (*v3b*) 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.

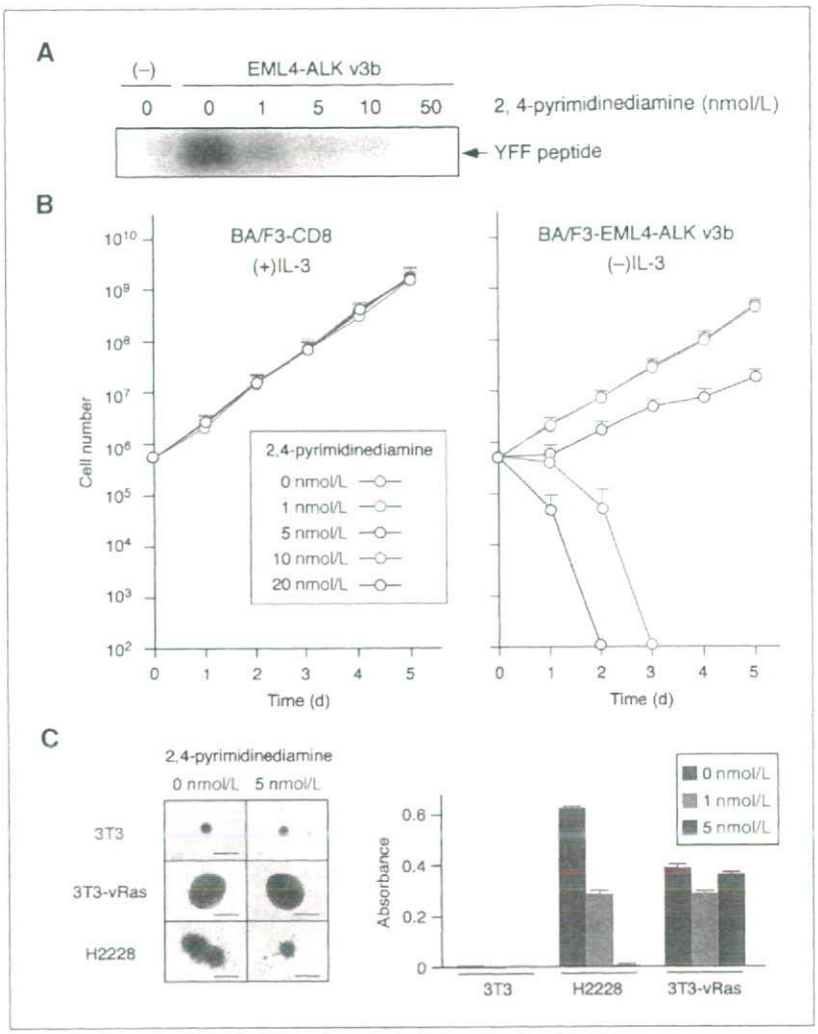

Figure 4. Essential role of EML4-ALK kinase activity in malignant transformation. *A*, lysates of HEK293 cells expressing FLAG-tagged EML4-ALK variant 3b (*v3b*) were divided into five equal portions, and each portion was subjected to immunoprecipitation with antibodies to FLAG. The immunoprecipitates 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 μg 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 ⩾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 38 858 qualified reads for the 13 study subjects. BLAST analysis then isolated 32 867 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.

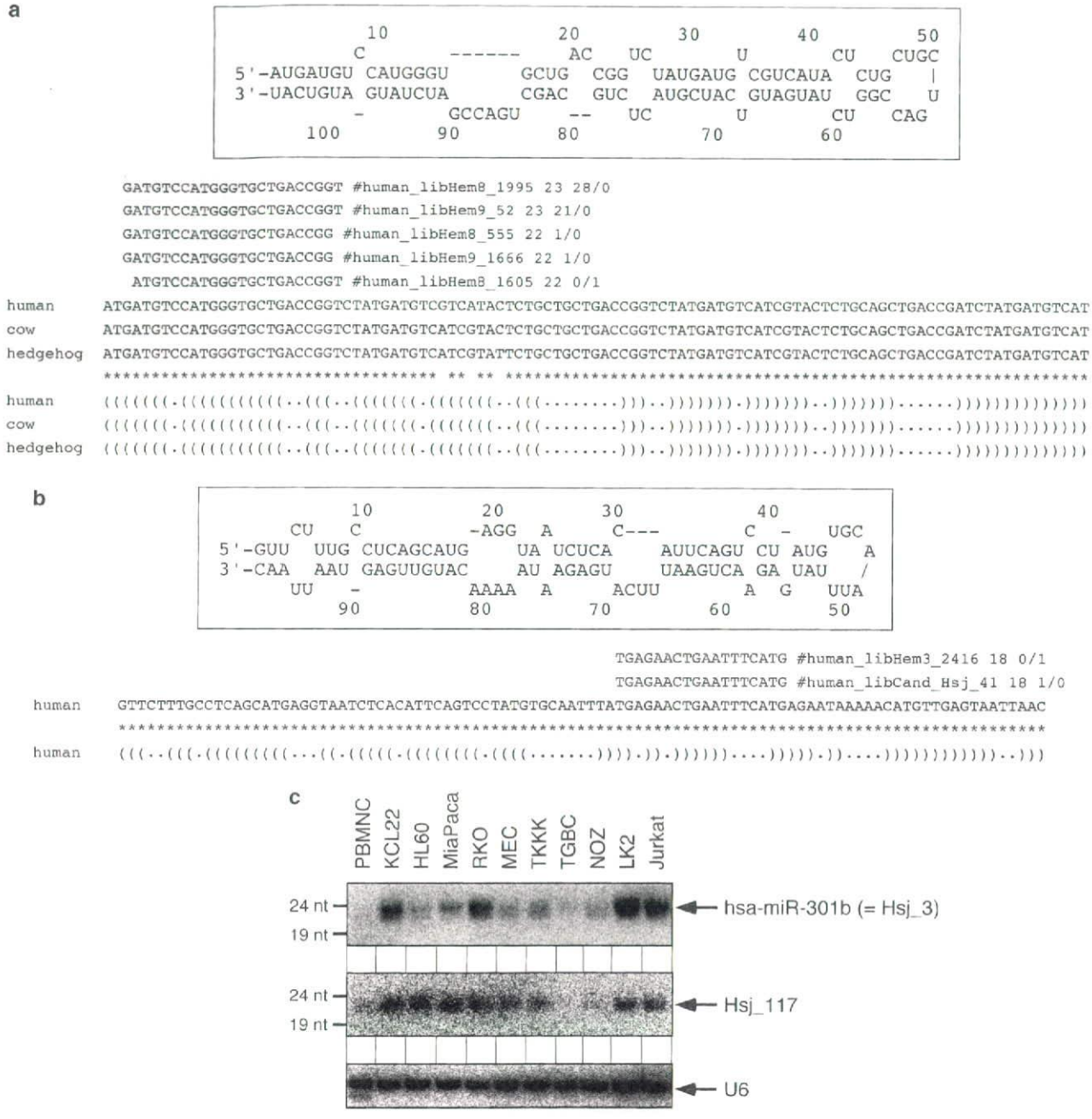

Figure 1 Nucleotide sequence and expression of novel miRNA candidates. Nucleotide sequences (red) of genes for the predicted novel miRNAs Hsj_376 (**a**) and Hsj_41 (**b**) are aligned with genomic sequences of human and other species. Asterisks indicate conserved nucleotides. Possible base-pairing schemes for the respective miRNA precursors are shown in the upper insets. (**c**) Small-RNA fractions (800 ng per lane) purified from the indicated cell lines with the use of a mirVana RNA isolation kit (Ambion, Austin, TX, USA) were subjected to northern blot analysis with 'locked' nucleic acid probes for the candidate miRNAs Hsj_3 or Hsj_117 or for U6 small nuclear RNA (internal control). Hjs_3 has been very recently deposited into the miRBase database as hsa-miR-301b. The positions of 24- and 19-nt size markers are indicated on the left. miRNA, microRNA.

species, given that some miRNAs are species-specific or have arisen recently during evolution.[6]

We isolated an unexpectedly large number ($n = 170$) of independent candidates for novel miRNAs among 296 sequence reads (Supplementary Table 1, Supplementary Data). The proportion of reads for such novel candidate miRNAs among all miRNA reads ranged from 1.7 to 9.5% per sample (mean, 4.7%). Of the 170 candidates, 19 were identified in at least two samples, supporting the notion that they are *bona fide* miRNAs. The surrounding genome sequence for one such candidate (designated Hsj_376) is conserved among human, cow and hedgehog (Figure 1a). Hsj_376 was found in two acute myeloid leukemia samples (corresponding to a total of 52 reads) in our data set and folds into a single hairpin (Figure 1a). In contrast, we obtained only one read for a candidate miRNA (Hsj_41) whose surrounding genome sequence also folds into a single

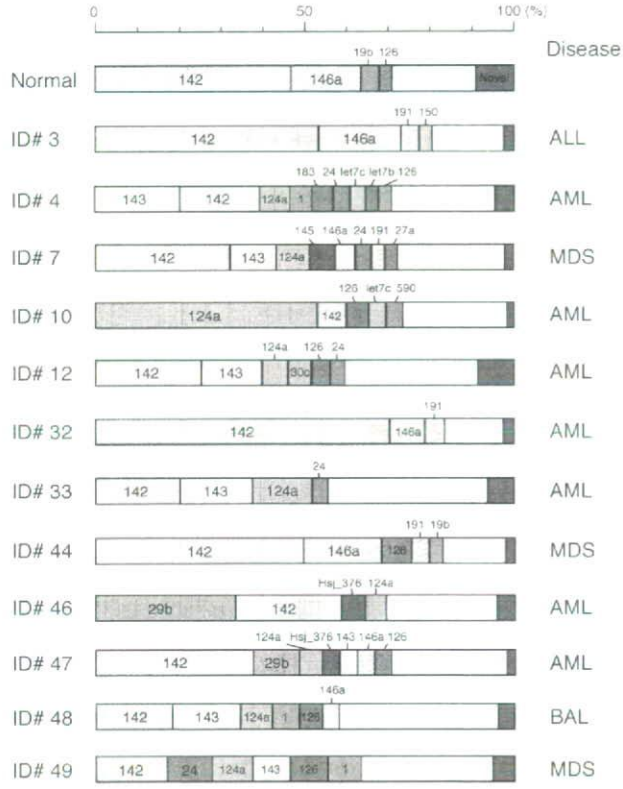

Figure 2 Expression profiles of miRNAs in CD34$^+$ specimens. The percentage contribution of each miRNA to the total miRNA population was calculated for each study subject. Abundant miRNAs are represented as color-coded, with candidates for novel miRNAs shown in red. The disease type of each individual is also indicated on the right. ALL, acute myeloid leukemia; AML, acute lymphoid leukemia; MDS, myelodysplastic syndrome; miRNA, microRNA.

hairpin structure (Figure 1b). However, this read was independently identified in our experiments performed both in Japan and in the Netherlands. The nucleotide sequence of all the miRNA candidates and their flanking sequences are presented in Supplementary Data.

The genomic sequences for some of the candidate miRNAs mapped in the vicinity ($\leqslant$20 kbp) of those for other miRNAs in the human genome. For example, the gene for one candidate (Hsj_360) and hsa-miR-560 are present on the long arm of chromosome 2 separated by a distance of $\sim$1 kbp (Supplementary Figure 1). In this instance, the genome sequences for the two miRNAs are not conserved in other species, indicative of recent evolution.

Expression of some of the candidate miRNAs was confirmed by northern blot analysis with small RNA fractions isolated from a variety of human cancer cell lines, including KCL22 (chronic myeloid leukemia), HL60 (acute myeloid leukemia), MiaPaCa (pancreatic carcinoma), RKO (colorectal carcinoma), MEC (cholangiocarcinoma), TKKK (intrahepatic bile duct carcinoma), TGBC (gallbladder carcinoma), NOZ (gallbladder carcinoma), LK2 (lung squamous cell carcinoma) and Jurkat (T-cell leukemia) (Figure 1c).

The relative expression profile of miRNAs was then calculated for each sample as shown in Figure 2. Whereas some miRNAs, such as miR-124a, miR-142, miR-143 and miR-146a, were expressed in different types of leukemia, most miRNAs were expressed in a sample-specific manner. For instance, miR-29b was abundant in only two samples (ID nos. 46 and 47), with the reads for this miRNA accounting for $<$1% of all miRNA reads in each of the other specimens. Similarly, the novel miRNA candidate Hsj_376 was abundant in the same two samples but not in the others. Both hsa-miR-183 and hsa-miR-590 were detected in only single samples (ID nos. 4 10, respectively).

To examine further the similarities and differences in the miRNA profiles among the study subjects, we performed a hierarchical clustering analysis for the subjects based on the expression patterns of all known and novel miRNAs (Figure 3a). Leukemia specimens with a normal karyotype were clustered in the same branch, indicative of a relative homogeneity of these samples, at least with regard to miRNA expression. Nevertheless, the healthy volunteer was placed in a different branch, suggesting that leukemic blasts with a normal karyotype possess a miRNA profile distinct from that of nonleukemic CD34$^-$ cells with a normal karyotype.

We further attempted to identify miRNAs whose expression level was significantly linked to blast karyotype. Application of Student's *t*-test to the miRNA expression data with a Benjamini and Hochberg false discovery rate [7] of $<$0.05 resulted in the isolation of six miRNAs (hsa-miR-29c, hsa-miR-124a, hsa-miR-150, hsa-miR-183, hsa-miR-382 and hsa-miR-590). Hierarchical clustering of the study subjects based on the expression profiles of these 'karyotype-associated miRNAs' revealed that the healthy volunteer was again placed apart from the leukemic patients with a normal karyotype.

In conclusion, application of the mRAP procedure to CD34$^+$ leukemic blasts yielded 7487 reads for potential miRNA clones. We previously showed that mRAP readily allows the isolation of $>1\times10^6$ miRNA concatamers from $\leqslant1\times10^4$ cells and is thus suitable for miRNA profiling of clinical specimens.[5] Indeed, mRAP functioned well with the small number of purified specimens in the present study, with the result that sequencing capacity, rather than specimen quantity, is likely to be the limiting factor for the size of the final data set in most studies.

Although, in the present study, the total number of sequence reads per sample (average = 2989 reads) was not high, we were able to discover a relatively large number (n = 170) of novel miRNA candidates from our sequence reads. Candidates for novel miRNAs continue to be identified, making it likely that the total number of human miRNAs has not yet reached saturation.[8] Our results show that CD34$^+$ leukemic blasts express a wider range of miRNAs than previously appreciated and that overall miRNA expression profiles generally reflect blast karyotype. Such karyotype-specific miRNAs may play a role in the malignant transformation of blasts of the corresponding karyotype, a possibility that needs to be confirmed by analysis of a large number of samples.

It is possible that some of the miRNA candidates identified in our study are not genuine miRNAs but rather degradation products of RNA or DNA. We believe, however, that a substantial proportion of the candidate miRNAs are indeed novel miRNAs because (i) many of them were identified in different samples in different laboratories (in Japan and in the Netherlands), (ii) many of them (together with the surrounding sequences in the genome) are conserved across various species and (iii) the expression of some of them was confirmed by northern blot analysis.

We have identified 170 novel miRNA candidates in, and demonstrated a high level of diversity in miRNA profiles among, leukemic blasts. Our data thus suggest that the miRNA

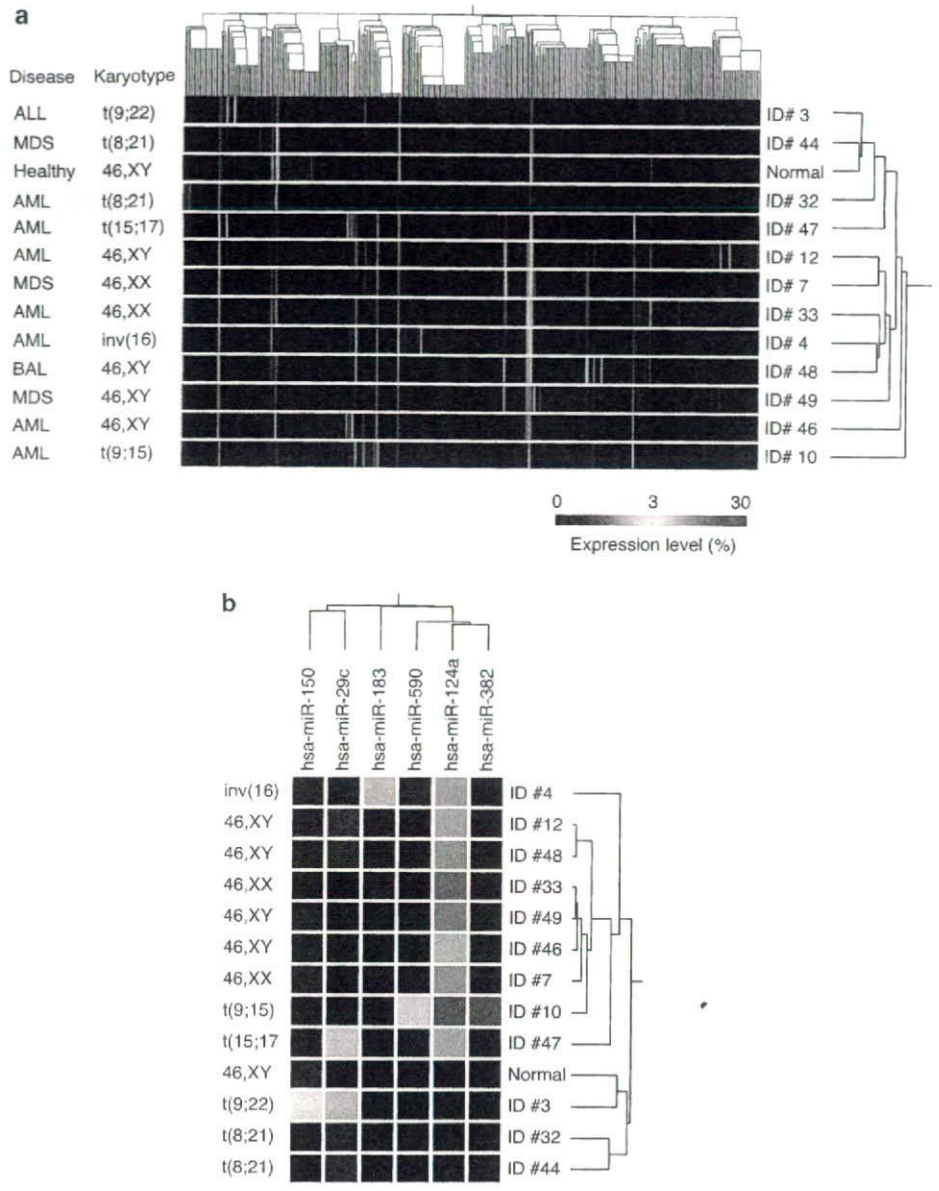

Figure 3 Hierarchical clustering of the study subjects based on miRNA expression profiles. (**a**) Subject tree generated by two-way clustering analysis with the expression profiles of all known and novel miRNAs. Each row corresponds to a separate sample, and each column to a miRNA whose expression is color-coded according to the indicated scale. The disease type and karyotype of each subject are shown at the left. (**b**) Six karyotype-associated miRNAs identified with Student's *t*-test and a false discovery rate of <0.05 were used for two-way clustering analysis as in (**a**). ALL, acute myeloid leukemia; AML, acute lymphoid leukemia; BAL, biphenotypic acute leukemia; MDS, myelodysplastic syndrome; miRNA, microRNA.

repertoire of human leukemias has not yet been exhausted, and they should provide a framework for future studies in this regard.

Note added in proof

Hsj_117 and Hsj_360 have the miRBase accession numbers hsa-miR-590 and hsa-miR-663b, respectively.

Acknowledgements

This study was supported in part by a grant for Third-Term Comprehensive Control Research for Cancer from the Ministry of Health, Labor, and Welfare of Japan as well as by a grant for Scientific Research on Priority Areas 'Applied Genomics' from the Ministry of Education, Culture, Sports, Science and Technology of Japan. The authors declare no competing financial interests.

S Takada[1], Y Yamashita[1], E Berezikov[2], H Hatanaka[1], S-i Fujiwara[1], K Kurashina[1], H Watanabe[1], M Enomoto[1], M Soda[1], YL Choi[1] and H Mano[1,3]
[1]*Division of Functional Genomics, Jichi Medical University, Shimotsukeshi, Tochigi, Japan;*
[2]*Hubrecht Institute, Utrecht, The Netherlands and*
[3]*CREST, Japan Science and Technology Agency, Saitama, Japan*
E-mail: hmano@jichi.ac.jp

References

1 Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004; **116**: 281–297.
2 Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E *et al.* Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci USA* 2002; **99**: 15524–15529.
3 He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S *et al.* A microRNA polycistron as a potential human oncogene. *Nature* 2005; **435**: 828–833.
4 Nelson PT, Baldwin DA, Scearce LM, Oberholtzer JC, Tobias JW, Mourelatos Z. Microarray-based, high-throughput gene expression profiling of microRNAs. *Nat Methods* 2004; **1**: 155–161.
5 Takada S, Berezikov E, Yamashita Y, Lagos-Quintana M, Kloosterman WP, Enomoto M *et al.* Mouse microRNA profiles determined with a new and sensitive cloning method. *Nucleic Acids Res* 2006; **34**: e115.
6 Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, Cuppen E *et al.* Diversity of microRNAs in human and chimpanzee brain. *Nat Genet* 2006; **38**: 1375–1377.
7 Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 2003; **19**: 368–375.
8 Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH, Cuppen E. Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* 2005; **120**: 21–24.

Supplementary Information accompanies the paper on the Leukemia website (http://www.nature.com/leu)

Fusion of *ZMIZ1* to *ABL1* in a B-cell acute lymphoblastic leukaemia with a t(9;10)(q34;q22.3) translocation

Leukemia (2008) **22**, 1278–1280; doi:10.1038/sj.leu.2405033; published online 15 November 2007

The *ABL1* gene has been found to be fused to four identified partner genes in haematological malignancies. It is rearranged with *BCR* by the t(9;22)(q34;q11.2) translocation in more than 95% of chronic myeloid leukaemia and in over 25% of adult B-cell acute lymphoblastic leukaemia.[1,2] It is rearranged with *TEL* (also known as *ETV6*) in rare cases of chronic myeloid leukaemia and acute leukaemia.[3] In T-cell acute lymphoblastic leukaemia, *ABL1* can be fused with *NUP214* (a gene located in 9q34) on episomes[4] or with *EML1* by a t(9;14)(q34;q32) translocation.[5] Moreover, a recent publication described a t(1;9)(q24;q34) translocation in a B-cell acute lymphoblastic leukaemia case with a putative RCSD1-ABL1 fusion without molecular confirmation.[6] Here we report the cytogenetic and molecular analysis of a t(9;10)(q34;q23) translocation, from a case of B-lineage ALL, with recombination of *ABL1* to a new partner gene, *ZMIZ1* (zinc-finger MIZ-type containing 1).

The patient is an 18-month-old Japanese girl, with no personal or familial medical history other than bronchiolitis and an episode of atopic dermatitis treated with dermocorticosteroids in November 2006. In December 2006, at the age of 14 months, she presented with pallor, asthenia and fever. Physical examination was found normal. Laboratory investigation showed an abnormal white blood cell count ($2.2\,G\,l^{-1}$ with neutropenia $0\,G\,l^{-1}$) and non-regenerative anaemia (Hb 3.5 g per 100 ml). Bone marrow analysis showed heterogeneous density with 3–10% of immature cells. Immunophenotyping analysis of the bone marrow sample did not reveal aberrant surface marker expression. No karyotypic or molecular abnormality was detected at the time. Blood culture was positive for alpha-haemolytic streptococcus. The girl was treated for her septicaemia and transfused. The neutropenia ($1.4\,G\,l^{-1}$) persisted for 3 months. In April 2007, she presented with fever. Clinical examination was normal. Blood cell count showed bicytopaenia (neutrophils $0\,G\,l^{-1}$, Hb 6.6 g per 100 ml). Bone marrow examination showed 90% of CD19+, CD10+, CD34+ CD13−, CD33− lymphoblasts, corresponding to a diagnosis of B-cell acute lymphoblastic leukaemia II, according to the immunological EGIL (European Group for the Immunological Characterization of Acute Leukemias) classification. The patient was then treated abroad, according to the standard risk

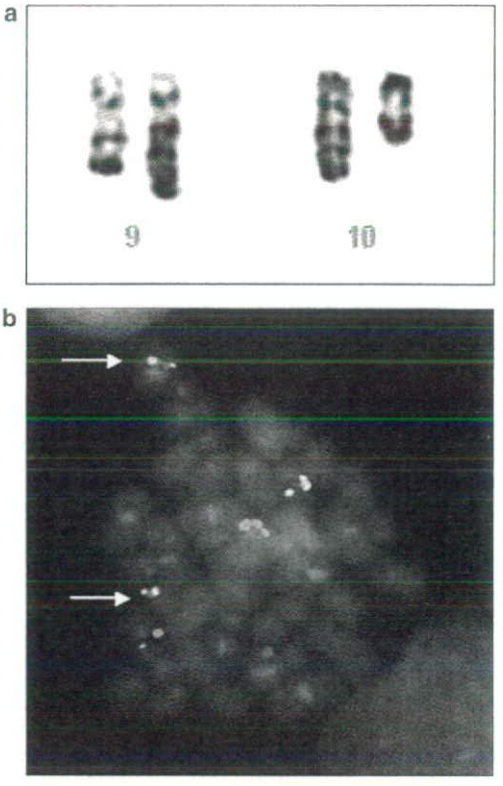

Figure 1 Cytogenetic analysis of patient bone marrow cells. (**a**) Partial karyotype showing the derivative chromosomes 9 and 10. (**b**) FISH using LSI *bcr* (green)/*ABL1* (red) dual-colour probe and the RP11-946M14 BAC clone labelled in coumarin (blue) revealed one normal red signal, one normal blue signal and two red signals fused to two blue signals on derivative chromosomes (arrows).

Leukemia (2008) 22, 1891–1898

www.nature.com/leu

ORIGINAL ARTICLE

High-resolution analysis of chromosome copy number alterations in angioimmunoblastic T-cell lymphoma and peripheral T-cell lymphoma, unspecified, with single nucleotide polymorphism-typing microarrays

S-i Fujiwara[1,2], Y Yamashita[1], N Nakamura[3], YL Choi[1], T Ueno[1], H Watanabe[1], K Kurashina[1], M Soda[1], M Enomoto[1], H Hatanaka[1], S Takada[1], M Abe[4], K Ozawa[2] and H Mano[1,5]

[1]Division of Functional Genomics, Jichi Medical University, Tochigi, Japan; [2]Division of Hematology, Jichi Medical University, Tochigi, Japan; [3]Department of Pathology, Tokai University School of Medicine, Kanagawa, Japan; [4]Department of Pathology, Fukushima Medical University, Fukushima, Japan and [5]CREST, Japan Science and Technology Agency, Saitama, Japan

Angioimmunoblastic T-cell lymphoma (AILT) and peripheral T-cell lymphoma, unspecified (PTCL-u) are relatively frequent subtypes of T- or natural killer cell lymphoma. To characterize the structural anomalies of chromosomes associated with these disorders, we here determined chromosome copy number alterations (CNAs) and loss of heterozygosity (LOH) at >55 000 single nucleotide polymorphism loci for clinical specimens of AILT ($n=40$) or PTCL-u ($n=33$). Recurrent copy number gain common to both conditions was detected on chromosomes 8, 9 and 19, whereas common LOH was most frequent for a region of chromosome 2. AILT- or PTCL-u-specific CNAs or LOH were also identified at 21 regions, some spanning only a few hundred base pairs. We also identified prognosis-related CNAs or LOH by several approaches, including Cox's proportional hazard analysis. Among the genes that mapped to such loci, a poor prognosis was linked to overexpression of *CARMA1* at 7p22 and of *MYCBP2* at 13q22, with both genes being localized within regions of frequent copy number gain. For a frequent LOH region at 2q34, we also identified IKAROS family zinc-finger 2 cDNAs encoding truncated proteins. Our data indicate that AILT and PTCL-u consist of heterogeneous subgroups with distinct transforming genetic alterations.

Leukemia (2008) **22,** 1891–1898; doi:10.1038/leu.2008.191; published online 17 July 2008
Keywords: T-cell lymphoma; chromosome copy number alterations; loss of heterozygosity; IKZF2

Introduction

Angioimmunoblastic T-cell lymphoma (AILT) and peripheral T-cell lymphoma, unspecified (PTCL-u) are relatively frequent subtypes of T- or natural killer (T/NK) cell lymphoma.[1] Although specific chromosomal translocations[2] and viral infections[3,4] have been associated with subsets of T/NK cell lymphoma, the molecular pathogenesis of these disorders remains obscure in most cases. Furthermore, given that PTCL-u is diagnosed on the basis of patients not having other specific subtypes of PTCL,[5] it likely consists of heterogeneous subgroups of lymphoma. Prognosis of AILT and PTCL-u is generally poor, with a 5-year survival rate of ~30%,[1] and standard treatment strategies for these conditions remain to be established. Characterization of the intrinsic genetic aberrations responsible for these two subtypes of T/NK cell lymphoma and the development of new classification schemes based on such molecular pathogenesis are thus important clinical goals.

In addition to nucleotide mutations and epigenetic abnormalities, structural changes of chromosomes, or chromosomal instability, are important in cancer development.[6] Gene amplification may promote the oncogenic activity of a subset of proto-oncogenes, such as *MYC*, *ERBB2* and *CCND1*. Conversely, deletion or truncation of tumor suppressor genes may underlie inactivation of their function. Furthermore, loss of heterozygosity (LOH) is frequently observed in the tumor genome; this condition is characterized by the deletion of one allele of a gene either without (copy number (CN) = 1) or with (CN = 2, referred to as uniparental disomy) duplication of the remaining allele. Regions of the genome affected by LOH have been thought to harbor mutated or epigenetically silenced tumor suppressor genes. However, recent evidence indicates that these regions may also harbor activated oncogenes, as demonstrated for mutated *JAK2* in myeloproliferative disorders.[7]

Comparative genomic hybridization (CGH) has been applied to assess chromosome copy number alterations (CNAs) in AILT/PTCL-u. Renedo *et al.*[8] found the most common CN gain on X chromosome in T-cell non-Hodgkin's lymphoma. The same approach for PTCL-u with Zettl *et al.*[9] identified recurrent CN gains on chromosome 7q22-qter, and recurrent CN losses on 5q, 6q, 9p, 10q, 12q and 13q. Array-based CGH with a resolution of >100 kb has also been used to examine AILT/PTCL-u, revealing recurrent CN gains of 11p11–q14, 19 and 22q in AILT, and of 8, 17 and 22q in PTCL-u.[10] Some inconsistency among these data may reflect the genetic heterogeneity in AILT/PTCL-u, and a low-resolution power in conventional or array-based CGH failed to pinpoint the genes essential to these CNAs.

Microarrays originally developed for typing of single nucleotide polymorphisms (SNPs) are now being applied to assess CNAs. Given that SNP-typing arrays are able both to assess heterozygosity or homozygosity along entire chromosomes and to determine the DNA quantity for each chromosome separately,[11] such arrays are able to measure chromosome CN and LOH simultaneously. Furthermore, the recent development of high-density SNP-typing arrays has allowed such measurements to be made at a resolution of <100 kb.

To identify characteristic genomic aberrations for AILT or PTCL-u in a high resolution, we have collected fresh specimens of AILT ($n=40$) and PTCL-u ($n=33$) and subjected them to hybridization with Affymetrix Mapping 50K Hind 240 microarrays (Affymetrix, Santa Clara, CA, USA). Application of

Correspondence: Professor Dr H Mano, Division of Functional Genomics, Jichi Medical University, 3311-1 Yakushiji, Shimotsuke, Tochigi 329-0498, Japan.
E-mail: hmano@jichi.ac.jp
Received 7 November 2007; revised 21 May 2008; accepted 17 June 2008; published online 17 July 2008

bioinformatics to the resulting large data set revealed several novel genomic imbalances and candidate genes that may contribute to the pathogenesis of these two lymphomas.

Patients and methods

Clinical samples

Lymphoma specimens (70 from enlarged lymph nodes; 3 from extranodal tumors) were obtained from 73 patients (40 with AILT, 33 with PTCL-u) who attended Jichi Medical University Hospital or Fukushima Medical University Hospital between 1985 and 2004. The pathology of the specimens was reevaluated on the basis of the revised classification scheme of the World Health Organization (WHO).[5] All 73 specimens, which conformed with the WHO classification of AILT or PTCL-u, were positive for the pan T-cell marker CD3 and negative for the monoclonal integration of human T-cell leukemia virus—I proviral DNA (data not shown). Mean age at diagnosis was 63 years (range, 19–89) and 67% of the patients were men. Most patients had been treated with cyclophosphamide-, doxorubicin-, vincristine- and prednisone-based regimens. Clinical characteristics of the study subjects are summarized in Supplementary Table 1. Informed consent was obtained according to a protocol approved by the ethics committees of Jichi Medical University and Fukushima Medical University Hospital. As normal controls, CD4-positive cells were isolated with the use of CD4 MicroBeads and a Mini-MACS isolation column (Miltenyi Biotec, Auburn, CA, USA) from peripheral blood mononuclear cells of healthy volunteers.

SNP-typing arrays

Genomic DNA was extracted from the lymphoma specimens with the use of a QIAamp DNA Mini kit (Qiagen, Valencia, CA, USA), digested with *Hind*III, ligated to the Adaptor-Hind (Affymetrix) and subjected to hybridization with Mapping 50K Hind 240 arrays (Affymetrix). SNP genotype calls were subsequently determined with GDAS software version 3.0 (Affymetrix) with a confidence score threshold of 0.05. Chromosome CN and the LOH likelihood score at each SNP site were calculated from the hybridization signal intensity and the SNP call with the use of CNAG 2.0 software (http://www.genome.umin.jp).[12] Only CNAG data for autosomes were analyzed, and known copy number variation (CNV) loci[13,14] were excluded from the analysis. We considered chromosome CNAs or LOH reliable only when $\geqslant 2$ contiguous SNP probes yielded the same data. The mean probe signal intensity at diploid chromosomes was inferred from the data of control samples (in which most chromosomes would be expected to be diploid). Chromosome CN and LOH likelihood score data for all autosomal SNP sites are available on request.

Quantitative RT and real-time PCR analysis

Total RNA was isolated from specimens with the use of an RNeasy Mini column (Qiagen) and was subjected to reverse transcription (RT) with PowerScript reverse transcriptase (Clontech, Palo Alto, CA, USA). The amount of specific cDNAs was quantitated by real-time polymerase chain reaction (PCR) analysis with a QuantiTect SYBR Green PCR Kit (Qiagen). The amplification protocol consisted incubations at 94 °C for 15 s, 60 °C for 30 s and 72 °C for 60 s. The incorporation of the SYBR Green dye into the PCR products was monitored in real time with an 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.

The relative abundance of the cDNAs of interest was calculated from the C_T value for each cDNA and that for *ACTB* cDNA. The primer sequences for RT–PCR are shown in Supplementary Table 2.

Nucleotide sequencing

For mutational screening of IKAROS family zinc-finger 2 (*IKZF2*) cDNA, RT–PCR was performed on a subset of lymphoma cDNAs with PrimeSTAR DNA polymerase (Takara Bio, Shiga, Japan) and the primers 5′-AGATCTCCCGACAGAGCTGGA-3′ and 5′-GGTGGGATTGTAAGTGCGGTATT-3′. Amplified PCR products were cloned into the pT7Blue-2 vector (EMD Biosciences, Madison, MI, USA) for nucleotide sequencing. To detect a cDNA for a short isoform of IKZF2, we performed RT–PCR with the primers 5′-ACCTCAAGCACACCCAATGGAC-3′ and 5′-CATCAGCTCAGCCTCCTTCTCA-3′. The resultant *IKZF2* cDNA sequences were compared with the published human *IKZF2* sequence (GenBank accession nos. NM_016260 and NM_001079526).

Statistical analysis

Changes in chromosome CN or gene expression level were evaluated by Student's *t*-test. Hierarchical clustering of the data set was performed with GeneSpring 7.0 software (Agilent Technologies, Santa Clara, CA, USA). Overall survival was estimated by the Kaplan–Meier method and was compared with the logrank test. Multivariate analysis of survival was performed with the Cox proportional hazard model (stepwise regression approach). Unless indicated otherwise, a *P*-value <0.05 was considered statistically significant.

Results

Recurrent chromosome CNAs

Chromosome CN was computationally inferred at 55 700 SNP sites for all autosomes in 73 specimens of AILT or PTCL-u. Hierarchical clustering of all subjects on the basis of these CN profiles revealed that three quarters of the specimens had relatively stable chromosomes, whereas the remaining one quarter had CNAs of various sizes (Figure 1a). Common chromosome gain (CN $\geqslant 3$ in $\geqslant 2$ cases), for example, was identified at 28 243 SNP loci, whereas common chromosome loss (CN $\leqslant 1$ in $\geqslant 2$ cases) was detected at 6479 loci. The prognosis of study subjects with such CNAs (Figure 1a) was significantly worse than that of those without them (Figure 1b), indicative of linkage between these CNAs and the transformation process for AILT or PTCL-u.

In addition, frequent CNAs were readily identified in our data set. Highly recurrent chromosome amplification (CN $\geqslant 4$ in $\geqslant 20$ cases) was apparent at three distinct regions of 8q, 9p and 19q (Table 1). These regions were as small as 175 bp encompassing three contiguous SNP loci at 8q24.11 or 290 bp encompassing another three SNP loci at 19q13.43, demonstrating the high resolution of the SNP array-based CN analysis. Frequent copy number loss (CN of $\leqslant 1$ in $\geqslant 4$ cases), on the other hand, was identified at two distinct regions of 3q and 9p (Table 1). Our CN data further revealed homozygous deletion at these two regions in some individuals (CN = 0 in seven cases at 3q and in three cases at 9p).

Despite the similarity in the profiles for recurrent CNAs between AILT and PTCL-u (Table 1), we examined whether there might be disease-specific CNAs for either of these disorders. Application of Student's *t*-test to the CN profiles for loci with frequent CNAs (those in ⩾10% of subjects) resulted in the isolation of thirteen regions with a disease-dependent CNA (Supplementary Table 3).

Effects of CNAs on gene expression

To examine the relation between CNAs and gene expression, we performed quantitative RT–PCR analysis for genes that mapped within recurrent CNAs. The recurrent loss at 9p21.3 (Table 1) contains the genes for two important inhibitors of cyclin-dependent kinases, *CDKN2A* and *CDKN2B*, which are deleted or epigenetically silenced in a variety of cancer cells.[15] A decrease in DNA content at 9p21.3 was associated with a reduced level of expression of the genes that mapped to this locus: *CDKN2A*, *CDKN2B* and *MTAP* (left panel of Figure 2a).

In addition to the recurrent CNAs shown in Table 1, we also detected a frequent gain in CN at a locus of 7p22.3–22.2 in 30 out of the 73 patients; this locus contains the gene for caspase recruitment domain membrane-associated guanylate kinase protein 1 (*CARMA1*, GenBank accession no. NM_032415). Overexpression of *CARMA1* has been demonstrated in B-cell lymphoma and adult T-cell leukemia or lymphoma.[16,17] As demonstrated in the right panel of Figure 2a, an increase in CN for *CARMA1* was associated with an increase in the amount of the corresponding mRNA, albeit with a marginal statistical significance ($P=0.053$).

We then examined whether the altered expression of these genes influenced the survival of the affected individuals. Consistent with previous results for other hematologic malignancies,[18,19] our data revealed a negative impact of a reduced level of *CDKN2A* expression on the clinical outcome of AILT or PTCL-u (Figure 2b). In addition, individuals with AILT or PTCL-u showing an increase in *CARMA1* expression had a poorer prognosis than did those without such an increase.

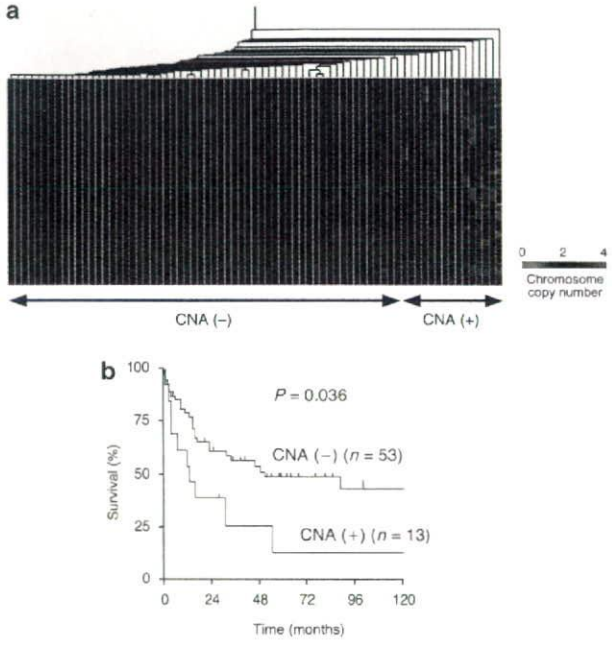

Figure 1 Chromosome copy number alterations (CNAs) in the genome of angioimmunoblastic T-cell lymphoma (AILT) or peripheral T-cell lymphoma, unspecified (PTCL-u). (**a**) Hierarchical clustering analysis of the study subjects ($n=73$) on the basis of the inferred copy number (CN) for all autosomal single nucleotide polymorphism (SNP) sites in the lymphoma specimens. CN is color coded according to the indicated scheme. SNP sites are ordered on the basis of their physical position from top to bottom. The patients could be subdivided into those with or without CNAs as indicated at the bottom. (**b**) The survival of the two groups of patients classified on the basis of the absence or presence of CNAs was compared by Kaplan–Meier analysis, with the *P*-value calculated by the logrank test.

Recurrent LOH

We next calculated the LOH likelihood score[12] at each SNP site. By direct sequencing of some of the genomic regions with a high LOH likelihood score, we determined that a score of ⩾20 was likely to be a reliable indicator of the presence of LOH (data not shown). We therefore used this value as a threshold for LOH in the following analyses.

Many ($n=42\,926$) of the 55 700 SNP sites were found to have an LOH likelihood score of ⩾20 in ⩾2 patients in our cohort. Among these SNP sites, common LOH (LOH likelihood score of ⩾20 in ⩾10% of cases) was apparent at 3093 loci distributed throughout most chromosomes (Figure 3). The most frequent region of LOH (LOH likelihood score of ⩾20 in 22 samples) was an ~440-kb region at 2q32.3 that includes 13 contiguous SNP loci.

We also screened for genomic loci whose LOH status was significantly linked to the diagnosis of AILT or PTCL-u. With a threshold *P*-value 0.001 (Student's *t*-test), we identified eight regions (each consisting of ⩾2 contiguous SNP loci) that mapped to four distinct chromosomes (Table 2). All of these

Table 1 Recurrent CNAs in AILT or PTCL-u specimens

CNA type	*Chromosome*	*Nucleotide position*	*Mapped genes*	*Affected no. of samples*		
				Total (n = 73)	*AILT (n = 40)*	*PTCL-u (n = 33)*
Gain (CN of ⩾4 in ⩾20 cases)						
	8	118 306 152–118 306 326	No genes	20	9	11
	9	10 638 555–10 722 021	No genes	36	24	12
	19	61 752 129–61 752 418	*ZFP28*	20	13	7
Loss (CN of ⩽1 in ⩾4 cases)						
	3	170 709 305–170 709 392	*MDS1*[a]	8	4	4
	9	21 762 317–22 072 375	*MTAP1*, *CDKN2A*, *CDKN2B*	4	1	3

Abbreviations: AILT, angioimmunoblastic T-cell lymphoma; CN, copy number; CNA, copy number alterations; PTCL-u, peripheral T-cell lymphoma, unspecified.
[a]The region with a frequent CN loss at the *MDS1* locus is distinct from the reported CNV region within *MDS1*.[13,14]

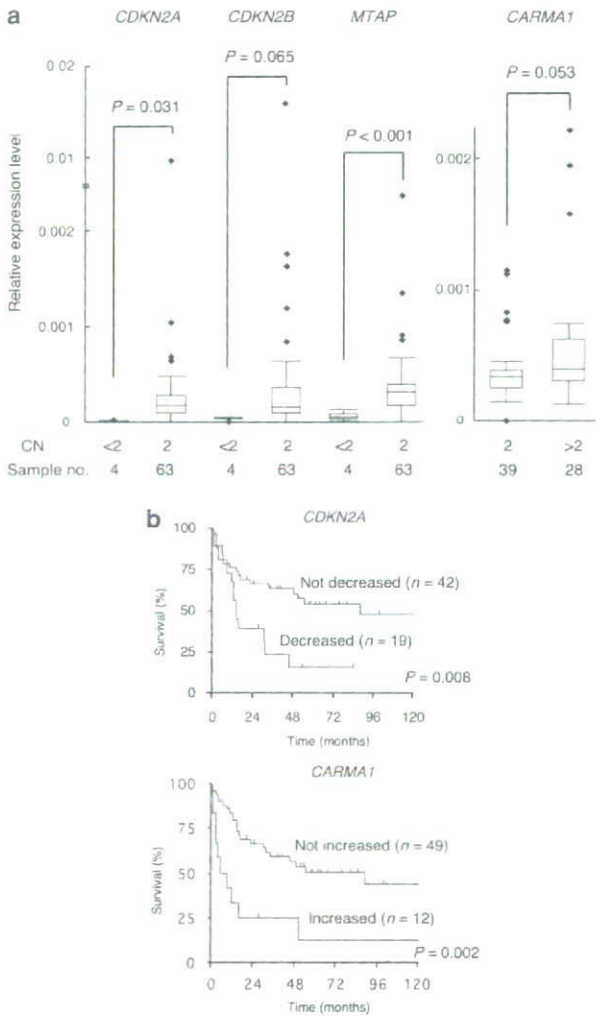

Figure 2 Influence of copy number alterations (CNAs) on candidate gene expression. (**a**) Expression levels of *CDKN2A*, *CDKN2B*, *MTAP* or caspase recruitment domain membrane-associated guanylate kinase protein 1 (*CARMA1*) relative to that of *ACTB* are shown in a box plot for the subjects (in the current cohort for single nucleotide polymorphism (SNP)-typing) with or without CNAs for the corresponding genes. The difference in expression level for each comparison was evaluated by Student's *t*-test. (**b**) The prognosis of patients with or without a reduced level of *CDKN2A* expression (*CDKN2A*/*ACTB* cDNA ratio of <0.0001) was compared by Kaplan–Meier analysis, with the *P*-value calculated by the logrank test (upper panel). Prognosis was similarly compared between individuals with or without an increased level of *CARMA1* expression, with such an increase defined as a relative expression level of more than the mean + 1.0 s.d. of that in the subjects without the copy number (CN) gain at the *CARMA1* locus (lower panel).

disease-associated LOH loci had a normal chromosome CN of 2, indicative of uniparental disomy at these loci.

Recently, aberrant expression of CD10 antigen has been reported for AILT cells.[20] We thus examined whether there are CNA/LOH, in our data set, related to such CD10-postive lymphoma cells. Immunohistostaining for CD10 was conducted among 64 cases in our cohort, revealing 21 cases positive for CD10 (Supplementary Table 1). Statistical analysis to detect CNAs associated with CD10-positive cases have identified one region of ~220 kb at chromosome 7 containing *GPR37* and

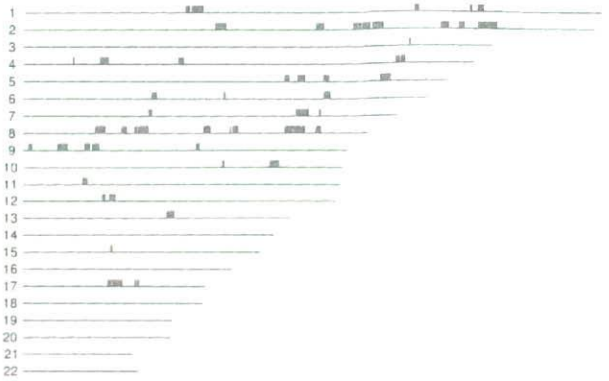

Figure 3 Distribution of recurrent loss of heterozygosity (LOH). Single nucleotide polymorphism (SNP) loci with a recurrent LOH (LOH likelihood score of ≥ 20 in $\geq 10\%$ of subjects) are indicated by blue bars in chromosome views. Chromosome numbers are shown at the left.

Table 2 Comparison of LOH likelihood profiles between AILT and PTCL-u

Chromosome	*Nucleotide position*	*Mapped genes*	*P-value*
2	31 171 920–31 217 236	*GALNT14*	$<5.4 \times 10^{-4}$
2	140 830 794–140 912 732	*LPR1B*	$<1.6 \times 10^{-4}$
2	141 385 148–141 387 720	*LPR1B*	$<3.2 \times 10^{-4}$
8	10 667 288–10 689 316	*PINX1*	$<9.1 \times 10^{-4}$
8	19 844 621–19 908 967	*LPL*	$<9.8 \times 10^{-4}$
11	80 723 837–80 727 426	No genes	$<1.9 \times 10^{-4}$
11	80 760 044–80 785 374	No genes	$<9.1 \times 10^{-4}$
12	57 759 034–57 781 381	No genes	$<9.9 \times 10^{-4}$

POT1 genes (Student's *t*-test, $P<0.001$), whereas a similar analysis for the LOH status found nine distinct regions on chromosomes 1, 4, 5, 6, 7 and 18 (Supplementary Table 4).

Novel isoforms of IKZF2

To isolate additional candidate genes for AILT or PTCL-u, we performed nucleotide mutation screening of known cancer-related genes located within the identified LOH regions. Extensive cDNA sequencing for these genes revealed a cDNA for a novel isoform of IKZF2, also known as Helios, in a subset of subjects. *IKZF2* maps to chromosome 2q34, for which a high LOH likelihood score (≥ 20) was identified in seven specimens (data not shown).

IKZF2 belongs to the IKAROS family of transcriptional factors, which are important regulators of lymphocyte development,[21,22] and short isoforms of IKZF2 have been reported for the malignant cells of adult T-cell leukemia or lymphoma[23] and T-cell acute lymphoblastic leukemia.[24] In our cohort, RT–PCR amplification of the entire coding region of the *IKZF2* mRNA detected a product in five of the seven study subjects with LOH at the *IKZF2* locus (Figure 4a). One of these products (from patient ID no. 1) was ~1.3 kb in size and apparently smaller than the others. Nucleotide sequencing of this cDNA revealed that it did not contain exons 3 and 4 of *IKZF2* (Figures 4b, c) and therefore encodes a protein that lacks 145 amino acids (including the first three zinc-finger domains) compared with the wild-type protein and has a Thr-to-Met substitution at amino-acid position 45 (the exon 2–5 boundary) (Figure 4b).

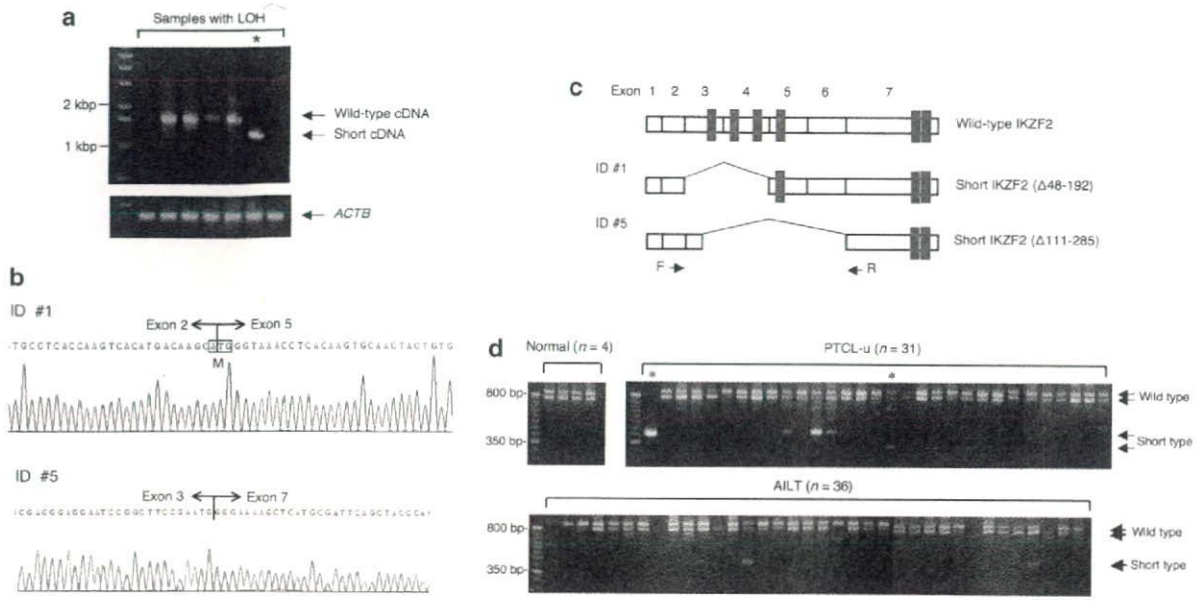

Figure 4 Identification of cDNAs for short isoforms of IKAROS family zinc-finger 2 (IKZF2). (**a**) Reverse transcription (RT)–PCR amplification of the entire coding region of *IKZF2* mRNA (upper panel) or of a portion of *ACTB* mRNA (lower panel) from seven subjects with loss of heterozygosity (LOH) at the *IKZF2* locus. The PCR products were fractionated by electrophoresis on a 0.8% agarose gel. The left lane contains DNA size markers (1-kb ladder). Whereas the predicted ~1.7-kb product of wild-type *IKZF2* cDNA was apparent in four specimens, a single ~1.3-kb product was identified in one patient (red asterisk, ID no. 1). (**b**) Sequence analysis of the exon 2–5 boundary of the *IKZF2* cDNA isolated from patient ID no. 1 or of the exon 3–7 boundary of that isolated from patient ID no. 5. M, methionine. (**c**) Schematic representation of the structure of the wild-type IKZF2 protein and that of the short isoforms identified in patients ID nos. 1 and 5. The positions of zinc-finger domains (red boxes) and of PCR primers (arrows) for amplification of exons 2–7 of *IKZF2* cDNA are also indicated. (**d**) RT–PCR amplification of exons 2–7 of *IKZF2* cDNA from patients with peripheral T-cell lymphoma, unspecified (PTCL-u) or angioimmunoblastic T-cell lymphoma (AILT) or from T cells of normal controls. The products were fractionated by electrophoresis on a 3% agarose gel, with DNA size markers (50-bp ladder) included in the leftmost lanes. Patients ID nos. 1 and 5 are indicated by the red and blue asterisks, respectively. The positions of products corresponding to wild type and short forms of IKZF2 are indicated on the right.

We next examined whether the mRNA for this novel isoform of *IKZF2* identified in the present study was also present in other patients or healthy individuals with the use of RT–PCR to amplify exons 2 to 7 of *IKZF2* cDNA (Figure 4c). Full-length cDNAs for *IKZF2* variant 1 (GenBank accession no. NM_016260) and variant 2 (NM_001079526) were detected in all normal T cells and in most of the lymphoma samples (Figure 4d). However, the cDNA for the short isoform was identified as a 414-bp product in seven patients (four with PTCL-u, three with AILT), including the one in whom this isoform was initially identified. This latter individual (patient ID no. 1) did not yield RT–PCR products corresponding to wild-type *IKZF2* cDNAs. Given that LOH was apparent at the *IKZF2* locus in this patient, the lymphoma cells likely harbor only a single *IKZF2* allele, which produces the truncated mRNA.

A novel cDNA fragment of 321 bp was further detected in the PTCL-u sample from patient ID no. 5 (Figure 4d). Nucleotide sequencing of this product revealed it to encode an IKZF2 protein that lacks 175 amino acids corresponding to a portion of exon 3 and all of exons 4 to 6 (Figures 4b, c).

Prognosis-related CNAs or LOH

To examine whether any of the CNAs or LOH regions identified in our data set are related to clinical outcome, we searched for prognosis-associated changes with several approaches. Given the many recurrent (at various frequencies) CNAs or LOH regions in the data set, we first examined whether some of these changes (observed in ≥ 5 samples) were preferentially present in patients who died within a year after diagnosis compared with those who survived for >1 year. For these potentially outcome-related genomic regions, prognosis was then compared between the individuals with or without each CNA or LOH site with the logrank test. CN gain at chromosomes 2 or 5 was found to be linked to poor prognosis (Supplementary Table 5; Supplementary Figure 1a). In addition, LOH at chromosomes 8 or 9 also had a negative impact on survival (Supplementary Table 5).

We next directly searched for genomic imbalances linked to poor prognosis by applying Cox's proportional hazard regression analysis to the chromosome CN profile for SNP loci with CNAs in $\geq 10\%$ of subjects, resulting in the isolation of 12 regions with a *P*-value <0.05 (Supplementary Table 6; Supplementary Figure 1). The β-score in the Cox analysis for all these regions was positive, indicating that CN gain at any of them is linked to a poor prognosis.

Further, Cox's analysis of CNAs only for the AILT data identified a CN gain at 13q22.3 that was significantly related ($P=0.025$) to poor clinical outcome (data not shown). This region spans only two contiguous SNP sites (corresponding to a distance of 237 bp) that map to a position 30 kb upstream of the MYC-binding protein 2 gene (*MYCBP2*, GenBank accession no. NM_015057). An increase in chromosome copy number at these SNP loci was further confirmed by quantitative PCR analysis (data not shown).

MYCBP2 is a large protein that binds specifically to Myc,[25] but whether it is involved in the transformation process of AILT is unknown. We detected a trend of an increase in the level of *MYCBP2* expression in the specimens of AILT patients with a CN gain at this locus compared to that in those without such a gain (Figure 5a), but without a statistical significance (Student's *t*-test,

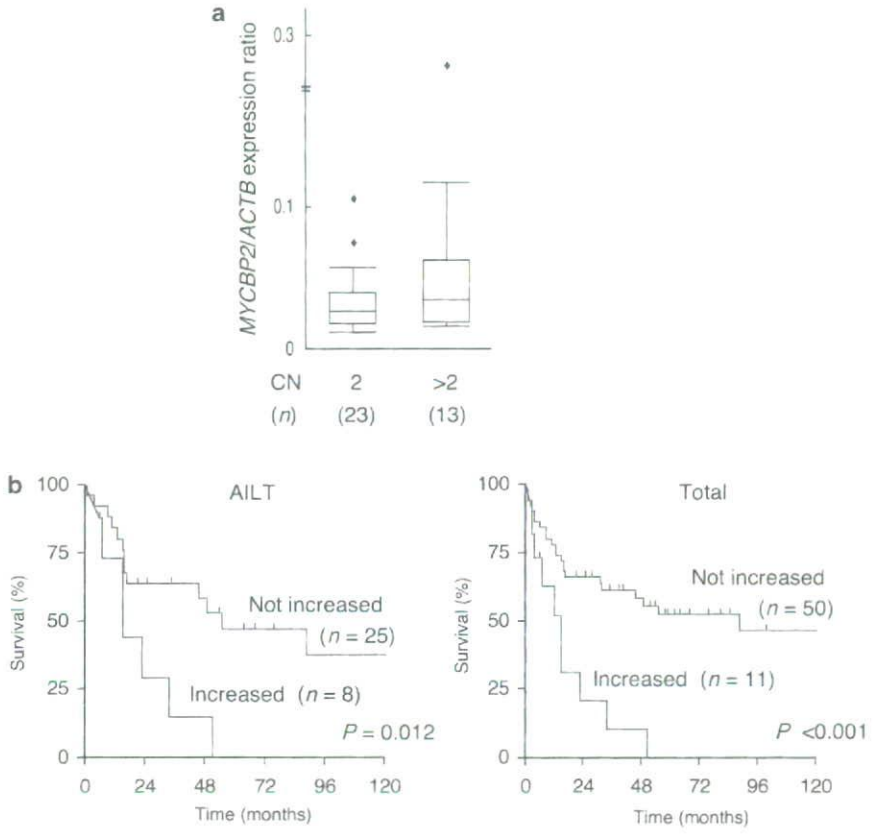

Figure 5 Poor prognosis associated with increased MYC-binding protein 2 (*MYCBP2*) expression. (**a**) The level of *MYCBP2* expression relative to that of *ACTB* expression is shown in a box plot for angioimmunoblastic T-cell lymphoma (AILT) patients ($n = 36$) with or without a copy number (CN) gain at the *MYCBP2* locus. (**b**) Comparison of the prognosis of AILT (left panel) or all (right panel) patients with or without an increased level of *MYCBP2* expression (defined by a relative expression level of more than the mean + 1.0 s.d. of that in the corresponding subjects without the CN gain at the *MYCBP2* locus). The *P*-values were calculated by the logrank test.

$P = 0.23$). Moreover, comparison of the survival of AILT patients with or without an increased expression of *MYCBP2* revealed a significantly worse prognosis for the former (Figure 5b). A poor prognosis for patients with increased *MYCBP2* expression was also apparent for the entire AILT and PTCL-u cohort (Figure 5b). These results suggest that *MYCBP2* is a candidate for the transformation-associated gene that maps to the 13q22.3 locus.

We also applied Cox's regression analysis to the LOH likelihood score data. For the PTCL-u data set, one region of ~9.2 Mbp at chromosome 8 (nucleotide positions: 41 865 249–51 050 357) was identified as being significantly related to poor clinical outcome (Supplementary Figure 1c).

Discussion

We have determined chromosome CN as well as LOH likelihood throughout the AILT or PTCL-u genome and have identified several novel recurrent and prognosis-related changes, some of which affect candidate genes for lymphomagenesis. In contrast to previous genomic analyses of T-cell lymphoma with CGH,[8–10,26] our study based on SNP-typing arrays was able to identify many small regions (<100 kb) with CNAs or LOH in the genome of lymphoma cells.

We detected CNAs or LOH regions at a similar frequency in AILT and PTCL-u, consistent with previous data showing a common chromosome gain at 11q13 in the two disorders[10] or common CNAs in PTCL-u and ALK-negative anaplastic large-cell lymphoma.[9] Few previous studies have described LOH in AILT or PTCL-u, possibly in part because of a high frequency of uniparental disomy in these conditions, as revealed in our data set (especially in the regions of recurrent LOH), that may be undetectable by conventional array-based CGH.

As tumor cell proportion in affected lymph nodes varies substantially among AILT/PTCL-u specimens, we tried to examine if such tumor cell proportion affects the CNA/LOH data. As shown in Supplementary Table 1, samples were classified into three subgroups; specimens with tumor cells composing <30% of lymph nodes were assigned to the 'L' group, whereas those with 30~60% were to the 'M' group, and those with ⩾60% to the 'H' group. We then tested whether such tumor cell proportion was linked to the detection of chromosome unstable cases where aberrant chromosome copy numbers (other than 2) were found in >10% of the SNP probes. Although detection of such cases was not statistically different between the L and M subgroups, CNA-positive specimens in the H group was significantly more frequent than that in the M or L group (Fisher's exact test, $P = 0.002$ for each comparison). Therefore, it is possible that tumor cell proportion in the specimens has significantly affected some parts of data set. However, it should be noted that (1) the high sensitivity of the CNA/LOH-calculation algorithm allows the detection of such changes among tumor samples contaminated with 70–80% of normal cells[27] and (2) tumor cell proportion in the affected lymph nodes

of AILT/PTCL-u may remain stable throughout stage progression.[28,29] Therefore, different frequency of CNA-positive cases may be an intrinsic property to the L/M/H subgroups of AILT/PTCL-u. Large-scale CGH studies with purified AILT/PTCL-u tumor cells (by using laser-capture microdissection system, for instance) would help to address these issues.

One of the goals of our study was to identify novel disease-associated genes in AILT or PTCL-u. Among the recurrent CNA loci, we identified *CARMA1* as a candidate gene for mediating the contribution of a 7p22 gain to lymphomagenesis. CARMA1 interacts with BCL10 and MALT1 and thereby mediates activation of nuclear factor (NF)-κB induced by stimulation of the T- or B-cell receptor.[10] Given that NF-κB frequently is activated and contributes to carcinogenesis in many tumor types,[30] activation of this transcription factor as a result of CARMA1 overexpression may also be important in disease progression and poor outcome of AILT or PTCL-u.

In addition, from a recurrent LOH locus, we identified cDNAs for novel short isoforms of IKZF2. Forced expression of full-length *IKZF2* cDNA results in inhibition of T-cell development at an early stage.[31] A cDNA encoding a truncated, dominant-negative form of IKZF2 with impaired DNA-binding activity has been associated with adult T-cell leukemia or lymphoma[23] and T-cell acute lymphoblastic leukemia.[24] The development of T-cell lymphoma in mice expressing this dominant-negative form of IKZF2 provided further support for its clinical relevance.[31] Although the structure of our isoform of IKZF2 is different from that of the previously described short one, which lacks the second to fourth zinc-finger domains,[23] both forms have only one zinc-finger domain in the N-terminal DNA-binding region of the protein. Given that isoforms of IKAROS family members with fewer than two N-terminal zinc-finger domains act in a dominant-negative manner,[32] both short isoforms of IKZF2 likely function as inhibitors of IKAROS family proteins. Given the transforming potential of a previously identified truncated form of IKZF2,[31] our data support the direct involvement of IKZF2 in transformation for a subset of AILT or PTCL-u.

In addition, an increased level of *MYCBP2* expression was associated with reduced survival time in AILT patients. The region of MYC that mediates binding to MYCBP2 is essential for the transactivation activity of MYC and is frequently mutated in Burkitt's and AIDS-related lymphomas.[25,33] Given that such mutations in MYC impair its ubiquitination and degradation,[33] overexpression of MYCBP2 may similarly hinder the access of ubiquitination enzymes or the proteasome to MYC and thereby promote its accumulation.

As shown in Supplementary Table 1, we have examined the expression of NK cell markers in the tumor cells. Expression of cell-surface CD56 was, for instance, tested with immunohistostaining procedures among 64 samples, and was found in only three cases. Similarly, presence of Epstein–Barr virus genome was analyzed among 65 cases, leading to the isolation of only two cases carrying the genome. It was thus difficult to draw statistically meaningful conclusions for CNA/LOH related to these small subgroups.

RefSeq genes may not be the only potential players in carcinogenesis. Large noncoding RNAs, for example, contribute to methylation of the genome,[34] whereas short noncoding RNAs, such as microRNAs, are implicated in regulation of cell growth and differentiation.[35] Such transcripts, despite their inability to synthesize proteins, may thus be involved in the development of AILT or PTCL-u. Given that the identification and annotation of these noncoding RNAs are still at an early stage,[36,37] many loci identified in our study may contain genes for as yet undiscovered noncoding RNAs, and these transcripts may participate in carcinogenesis.

In conclusion, our study has provided a large-scale, detailed analysis of CNAs and LOH in AILT and PTCL-u, and has identified candidates for lymphomagenesis-related genes. Our data set should prove to be a useful platform for further definition, from the viewpoint of chromosome abnormalities, of these clinical entities.

Acknowledgements

This study was supported in part by a grant for Third-Term Comprehensive Control Research for Cancer from the Ministry of Health, Labor, and Welfare of Japan as well as by a grant for Scientific Research on Priority Areas 'Applied Genomics' from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

Disclosure/Conflict of interest

The authors declare no competing financial interests.

References

1 Rudiger T, Weisenburger DD, Anderson JR, Armitage JO, Diebold J, MacLennan KA *et al.* Peripheral T-cell lymphoma (excluding anaplastic large-cell lymphoma): results from the Non-Hodgkin's Lymphoma Classification Project. *Ann Oncol* 2002; **13**: 140–149.
2 Fischer P, Nacheva E, Mason DY, Sherrington PD, Hoyle C, Hayhoe FG *et al.* A Ki-1 (CD30)-positive human cell line (Karpas 299) established from a high-grade non-Hodgkin's lymphoma, showing a 2;5 translocation and rearrangement of the T-cell receptor beta-chain gene. *Blood* 1988; **72**: 234–240.
3 Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc Natl Acad Sci USA* 1980; **77**: 7415–7419.
4 Harabuchi Y, Yamanaka N, Kataura A, Imai S, Kinoshita T, Mizuno F *et al.* Epstein–Barr virus in nasal T-cell lymphomas in patients with lethal midline granuloma. *Lancet* 1990; **335**: 128–130.
5 Jaffe ES, Harris NL, Stein H, Vardiman JW (eds). *Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues.* IARC Press: Lyon, 2001, 225–229.
6 Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature* 1998; **396**: 643–649.
7 Kralovics R, Passamonti F, Buser AS, Teo SS, Tiedt R, Passweg JR *et al.* A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med* 2005; **352**: 1779–1790.
8 Renedo M, Martinez-Delgado B, Arranz E, Garcia M, Urioste M, Martinez-Ramirez A *et al.* Chromosomal changes pattern and gene amplification in T cell non-Hodgkin's lymphomas. *Leukemia* 2001; **15**: 1627–1632.
9 Zettl A, Rudiger T, Konrad MA, Chott A, Simonitsch-Klupp I, Sonnen R *et al.* Genomic profiling of peripheral T-cell lymphoma, unspecified, and anaplastic large T-cell lymphoma delineates novel recurrent chromosomal alterations. *Am J Pathol* 2004; **164**: 1837–1848.
10 Thorns C, Bastian B, Pinkel D, Roydasgupta R, Fridlyand J, Merz H *et al.* Chromosomal aberrations in angioimmunoblastic T-cell lymphoma and peripheral T-cell lymphoma unspecified: a matrix-based CGH approach. *Genes Chromosomes Cancer* 2007; **46**: 37–44.
11 Matsuzaki H, Dong S, Loi H, Di X, Liu G, Hubbell E *et al.* Genotyping over 100 000 SNPs on a pair of oligonucleotide arrays. *Nat Methods* 2004; **1**: 109–111.
12 Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A *et al.* A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res* 2005; **65**: 6071–6079.

13 Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD *et al.* Global variation in copy number in the human genome. *Nature* 2006; **444**: 444–454.
14 Komura D, Shen F, Ishikawa S, Fitch KR, Chen W, Zhang J *et al.* Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome Res* 2006; **16**: 1575–1584.
15 Drexler HG. Review of alterations of the cyclin-dependent kinase inhibitor INK4 family genes p15, p16, p18 and p19 in human leukemia-lymphoma cells. *Leukemia* 1998; **12**: 845–859.
16 Oshiro A, Tagawa H, Ohshima K, Karube K, Uike N, Tashiro Y *et al.* Identification of subtype-specific genomic alterations in aggressive adult T-cell leukemia/lymphoma. *Blood* 2006; **107**: 4500–4507.
17 Nakamura S, Nakamura S, Matsumoto T, Yada S, Hirahashi M, Suekane H *et al.* Overexpression of caspase recruitment domain (CARD) membrane-associated guanylate kinase 1 (CARMA1) and CARD9 in primary gastric B-cell lymphoma. *Cancer* 2005; **104**: 1885–1893.
18 Takasaki Y, Yamada Y, Sugahara K, Hayashi T, Dateki N, Harasawa H *et al.* Interruption of p16 gene expression in adult T-cell leukaemia/lymphoma: clinical correlation. *Br J Haematol* 2003; **122**: 253–259.
19 Schmitt CA, McCurrach ME, de Stanchina E, Wallace-Brodeur RR, Lowe SW. INK4a/ARF mutations accelerate lymphomagenesis and promote chemoresistance by disabling p53. *Genes Dev* 1999; **13**: 2670–2677.
20 Attygalle A, Al-Jehani R, Diss TC, Munson P, Liu H, Du MQ *et al.* Neoplastic T cells in angioimmunoblastic T-cell lymphoma express CD10. *Blood* 2002; **99**: 627–633.
21 Kelley CM, Ikeda T, Koipally J, Avitahl N, Wu L, Georgopoulos K *et al.* Helios, a novel dimerization partner of Ikaros expressed in the earliest hematopoietic progenitors. *Curr Biol* 1998; **8**: 508–515.
22 Hahm K, Cobb BS, McCarty AS, Brown KE, Klug CA, Lee R *et al.* Helios, a T cell-restricted Ikaros family member that quantitatively associates with Ikaros at centromeric heterochromatin. *Genes Dev* 1998; **12**: 782–796.
23 Tabayashi T, Ishimaru F, Takata M, Kataoka I, Nakase K, Kozuka T *et al.* Characterization of the short isoform of Helios overexpressed in patients with T-cell malignancies. *Cancer Sci* 2007; **98**: 182–188.
24 Nakase K, Ishimaru F, Fujii K, Tabayashi T, Kozuka T, Sezaki N *et al.* Overexpression of novel short isoforms of Helios in a patient with T-cell acute lymphoblastic leukemia. *Exp Hematol* 2002; **30**: 313–317.
25 Guo Q, Xie J, Dang CV, Liu ET, Bishop JM. Identification of a large Myc-binding protein that contains RCC1-like repeats. *Proc Natl Acad Sci USA* 1998; **95**: 9172–9177.
26 Melendez B, Diaz-Uriarte R, Cuadros M, Martinez-Ramirez A, Fernandez-Piqueras J, Dopazo A *et al.* Gene expression analysis of chromosomal regions with gain or loss of genetic material detected by comparative genomic hybridization. *Genes Chromosomes Cancer* 2004; **41**: 353–365.
27 Yamamoto G, Nannya Y, Kato M, Sanada M, Levine RL, Kawamata N *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* 2007; **81**: 114–126.
28 Niitsu N, Okamoto M, Nakamine H, Aoki S, Motomura S, Hirano M. Clinico-pathologic features and outcome of Japanese patients with peripheral T-cell lymphomas. *Hematol Oncol* 2008; e-pub ahead of print.
29 Attygalle AD, Kyriakou C, Dupuis J, Grogg KL, Diss TC, Wotherspoon AC *et al.* Histologic evolution of angioimmunoblastic T-cell lymphoma in consecutive biopsies: clinical correlation and insights into natural history and disease progression. *Am J Surg Pathol* 2007; **31**: 1077–1088.
30 Jost PJ, Ruland J. Aberrant NF-kappaB signaling in lymphoma: mechanisms, consequences, and therapeutic implications. *Blood* 2007; **109**: 2700–2707.
31 Zhang Z, Swindle CS, Bates JT, Ko R, Cotta CV, Klug CA. Expression of a non-DNA-binding isoform of Helios induces T-cell lymphoma in mice. *Blood* 2007; **109**: 2190–2197.
32 Sun L, Heerema N, Crotty L, Wu X, Navara C, Vassilev A *et al.* Expression of dominant-negative and mutant isoforms of the antileukemic transcription factor Ikaros in infant acute lymphoblastic leukemia. *Proc Natl Acad Sci USA* 1999; **96**: 680–685.
33 Bahram F, von der Lehr N, Cetinkaya C, Larsson LG. c-Myc hot spot mutations in lymphomas result in inefficient ubiquitination and decreased proteasome-mediated turnover. *Blood* 2000; **95**: 2104–2110.
34 Chang SC, Tucker T, Thorogood NP, Brown CJ. Mechanisms of X-chromosome inactivation. *Front Biosci* 2006; **11**: 852–866.
35 Carrington JC, Ambros V. Role of microRNAs in plant and animal development. *Science* 2003; **301**: 336–338.
36 Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N *et al.* The transcriptional landscape of the mammalian genome. *Science* 2005; **309**: 1559–1563.
37 Takada S, Berezikov E, Yamashita Y, Lagos-Quintana M, Kloosterman WP, Enomoto M *et al.* Mouse microRNA profiles determined with a new and sensitive cloning method. *Nucleic Acids Res* 2006; **34**: e115.

Supplementary Information accompanies the paper on the Leukemia website (http://www.nature.com/leu)

Review Article

Non-solid oncogenes in solid tumors: *EML4–ALK* fusion genes in lung cancer

Hiroyuki Mano[1]

Division of Functional Genomics, Jichi Medical University, 3311-1 Yakushiji, Shimotsukeshi, Tochigi 329-0498, Japan

(Received July 22, 2008/Revised August 13, 2008/Accepted August 14, 2008/Online publication November 20, 2008)

It is generally accepted that recurrent chromosome translocations play a major role in the molecular pathogenesis of hematological malignancies but not of solid tumors. However, chromosome translocations involving the e26 transformation-specific sequence transcription factor loci have been demonstrated recently in many prostate cancer cases. Furthermore, through a functional screening with retroviral cDNA expression libraries, we have discovered the fusion-type protein tyrosine kinase echinoderm microtubule-associated protein like-4 (EML4)–anaplastic lymphoma kinase (ALK) in non-small cell lung cancer (NSCLC) specimens. A recurrent chromosome translocation, inv(2)(p21p23), in NSCLC generates fused mRNA encoding the amino-terminal half of EML4 ligated to the intracellular region of the receptor-type protein tyrosine kinase ALK. EML4–ALK oligomerizes constitutively in cells through the coiled coil domain within the EML4 region, and becomes activated to exert a marked oncogenicity both *in vitro* and *in vivo*. Break and fusion points within the *EML4* locus may diverge in NSCLC cells to generate various isoforms of EML4–ALK, which may constitute ~5% of NSCLC cases, at least in the Asian ethnic group. In the present review I summarize how detection of *EML4–ALK* cDNA may become a sensitive diagnostic means for NSCLC cases that are positive for the fusion gene, and discuss whether suppression of ALK enzymatic activity could be an effective treatment strategy against this intractable disorder. (*Cancer Sci* 2008; 99: 2349–2355)

Chromosome translocation is the most prevalent form of somatic changes in the cancer genome, occupying nearly three-quarters of all genetic change analyzed in cancer cells.[1] Such translocations may lead to the generation of novel fusion genes at the ligation points of chromosomes, or may juxtapose growth-promoting genes to aberrant promoter or enhancer fragments, resulting in dysregulated expression of the genes. In either case, such fusion genes or wild-type genes with altered expression may participate directly in the malignant transformation of cells that harbor chromosome translocations.

An archetypal example of such tumor-related translocations is t(9;22), which gives rise to the *breakpoint cluster regio (BCR)-Abelson murine leukemia viral oncogene homolog 1 (ABL1)* fusion gene in chronic myeloid leukemia (CML) and acute lymphoblastic leukemia (ALL).[2] Ligation to BCR constitutively elevates the protein tyrosine kinase (PTK) activity of ABL1, and forced expression of *BCR–ABL1* in the hematopoietic system induces CML and ALL in mice,[3,4] proving that BCR–ABL1 plays a pivotal role in the pathogenesis of such leukemias. Also, molecular detection of *BCR–ABL1* and the development of compounds to suppress BCR–ABL1 enzymatic activity have significantly changed the way we diagnose and treat individuals with CML. Because reverse transcription (RT)–polymerase chain reaction (PCR) can detect *BCR–ABL1* fusion transcripts in almost 100% of individuals with CML, even among those without the characteristic t(9;22) translocation, molecular detection of *BCR–ABL1* has become a standard technique used to diagnose CML. Given the very high sensitivity of PCR, such a strategy is also effective to follow up the tumor burden of leukemia in the patients.[5]

Further, the chemical compound STI571, which suppresses ABL1 kinase activity, has substantially prolonged the survival of patients at the chronic phase of CML, and achieved a higher probability of complete cytogenetic response among them compared to that with the previous treatment regimens.[6,7] Therefore, if translocation-mediated fusion genes encode activated enzymes with direct oncogenic potential, targeting such enzymes could provide a feasible approach to treat individuals harboring the corresponding fusion genes.

However, these fusion-type oncogenes have been reported frequently only in hematological malignancies, and not in solid tumors (especially in epithelial tumors).[8] It has been therefore widely assumed that balanced chromosome cytogenetic aberrations (and the resulting fusion genes) may be rare in the latter conditions. As shown in Table 1, for instance, the incidence of new cases with solid tumors in the USA is ~11 times larger than for those with hematological malignancies.[9] However, the number of recurrent balanced cytogenetic aberrations (RBA) in solid tumors ($n = 125$) is only a quarter of that in hematological malignancies ($n = 495$) worldwide,[10] suggesting that RBA are indeed characteristic of hematological malignancies and that these two tumor types may occur through distinct transformation mechanisms.

However, such a notion has been challenged recently by Mitelman *et al.* who have demonstrated that the number of fusion genes may simply be a function of the number of cases with an abnormal karyotype in both hematological malignancies and solid tumors.[11,12] A correlation between the number of patients with an abnormal karyotype and that of fusion genes is constant throughout all types of cancers ($R^2 = 0.82$, $P < 0.001$). It is therefore possible that infrequent reports of fusion genes in solid tumors (especially in epithelial tumors) may have been attributable to technical difficulties in obtaining clear karyotyping data or to the complex chromosome rearrangements in solid tumors.

If this is the case, many more fusion-type oncogenes may await discovery in solid tumors. Indeed, evidence in support of this prediction has been provided recently both by our discovery of the fusion-type PTK echinoderm microtubule-associated protein like-4 (EML4)–anaplastic lymphoma kinase (ALK), associated with lung cancer,[13–15] and by the detection of recurrent *e26 transformation-specific sequence (ETS)* fusion genes in prostate cancer.[16–18]

[1]E-mail: hmano@jichi.ac.jp

doi: 10.1111/j.1349-7006.2008.00972.x

Table 1. Number of recurrent balanced cytogenetic aberrations (RBA) in cancer

Cancer type	No. RBA†	Annual no. new cases in USA†
Leukemia	336	44 240
Lymphoma	159	71 380
Solid tumors	125	1 329 300

†Calculated from data reported previously.[9,10]

How to screen for oncogenes

Given the marked therapeutic efficacy of STI571 in CML, chemical inhibitors of epidermal growth factor receptor (EGFR) in lung cancer with activated EGFR,[19] and specific antibodies to HER2 in breast cancers with amplification of the *HER2* locus,[20] it is necessary to identify pivotal oncogenes in every cancer and to target these 'Achilles' heels'[21] for developing effective treatment strategies. We therefore tried to establish a functional screening system for transforming genes among a wide variety of cancer specimens.

The focus formation assay with 3T3 or RAT1 fibroblasts[22] has been used extensively to screen for oncogenes from clinical specimens. In such screening, genomic DNA is extracted from samples and transfected into the recipient fibroblasts. Because protein products of oncogenes can interfere with contact inhibition in fibroblasts, cell clones that have received oncogenes may keep growing even after the cells become confluent in culture. Such piled-up foci of cell clones (transformed foci) can be readily identified by visual inspection and subjected to the recovery of incorporated oncogenes. Application of such technologies has indeed succeeded in the isolation of a variety of transforming genes, such as mutated RAS family genes, activated RAF family genes, and a number of activated PTK genes.[23]

However, we have noticed that this type of screening system has a strong tendency to isolate the same sets of genes among different types of cancer (i.e. activated RAS family proteins and guanine nucleotide exchange factors). This could be due to an intrinsic property of the assay system, which isolates genes overriding growth inhibition mediated by cell-to-cell contact in fibroblasts. Another reason may be related to promoter specificity (Fig. 1a). In the screening systems where genomic DNA is used for transfection, any oncogene is controlled transcriptionally in the recipient cells by its own promoter and enhancer fragments. Therefore, if a promoter fragment of a given oncogene is active only in a tissue-specific manner (hematopoietic cell-specific, for instance), that gene would not be transcribed in fibroblasts, and thus could not be captured in the assay. Therefore, a genomic DNA-mediated screening system can only identify oncogenes with promoter fragments that are active in the recipient cells.

To overcome this limitation, it would be desirable to express every oncogene using an exogenous promoter fragment that allows abundant expression in any type of assay cell. For this purpose, we have developed a method to construct retrovirus-based cDNA expression libraries,[13,24–29] which can express any incorporated cDNA using a strong promoter fragment, long-terminal repeat (LTR), of the retroviral genome (Fig. 1b). Our system is so sensitive that we can generate libraries from small quantities of clinical specimens (such as $<1 \times 10^5$ cells). Further, given the high infection efficiency of retrovirus to dividing cells, any type of functional assay can be conducted in any proliferating cell with retroviral libraries.[30,31]

Discovery of a fusion-type PTK, EML4–ALK

Lung cancer remains the leading cause of cancer death, with an estimated ~1.3 million deaths worldwide each year.[32] Although

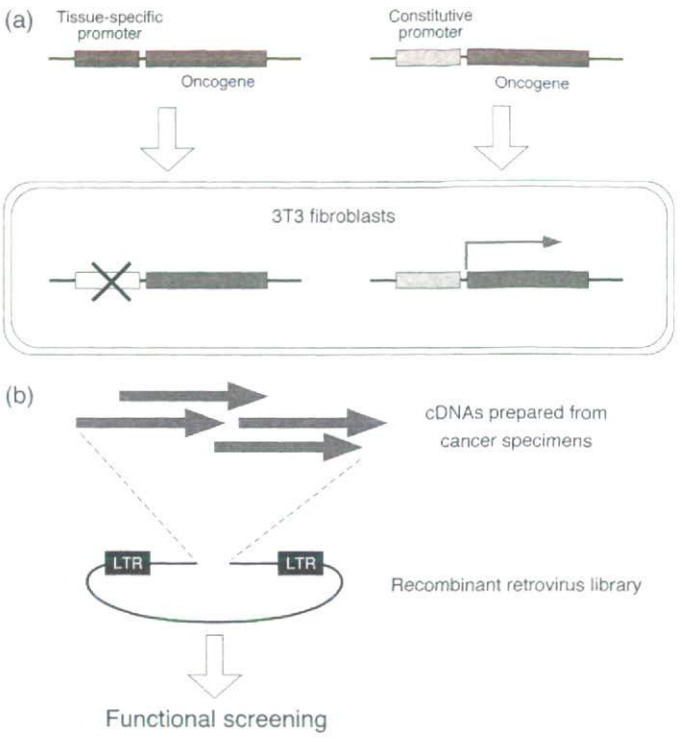

Fig. 1. Development of retroviral cDNA expression libraries for oncogene screening. (a) Although oncogenes controlled by a constitutive promoter or enhancer are expressed when introduced into fibroblasts, those controlled by a tissue-specific promoter-enhancer (e.g. specific to the hematopoietic system) are not transcribed in fibroblasts and will therefore not be detected by such functional screening. (b) To overcome this limitation, we synthesized cDNA from small quantities of clinical specimens, inserted them into a retroviral plasmid, and generated recombinant retroviral libraries. Theoretically, any type of dividing cell can be infected with such libraries, with the incorporated cDNA being expressed at a high level in the recipient cells under the control of the viral long terminal repeat (LTR).

activated EGFR has been identified in non-small cell lung cancer (NSCLC), the major subtype of lung cancer, and specific inhibitors against EGFR provide effective treatment modalities, this type of genetic mutation is found preferentially in non-smokers, young women, and the Asian ethnic group.[33] For other NSCLC patients who are not eligible for the anti-EGFR treatments, there are currently few effective treatments to improve their outcome, unless cancer cells are completely removed by surgery.[34]

We have therefore chosen NSCLC as the target of our retroviral screening system. First, among our consecutive panel of NSCLC specimens, we examined the presence of known transforming genes in lung cancer; that is, mutated *KRAS* and mutated *EGFR*. To raise a retroviral library, from the specimens negative for either mutation we chose a sample of lung adenocarcinoma resected from a 62-year-old man with a smoking history. A total of $>1.4 \times 10^6$ independent retroviral clones (with a mean cDNA size of 1.81 kb) were obtained from the specimen, and were used to infect 3T3 cells for the focus formation assay.

Dozens of transformed foci were readily identified in the assay, from which retroviral insert cDNA was rescued by PCR. Surprisingly, nucleotide sequences of the 5′ and 3′ parts of one cDNA corresponded to two different genes; one for microtubule-associated EML4,[35] and the other for the receptor-type PTK ALK.[36] Nucleotide sequencing of the cDNA revealed that the cDNA was a fusion between exons 1–13 of *EML4* and exons 20–29 of *ALK* (transcript ID ENST00000389048 in the Ensembl database; http://www.ensembl.org/index.html), thus encoding a

doi: 10.1111/j.1349-7006.2008.00972.x

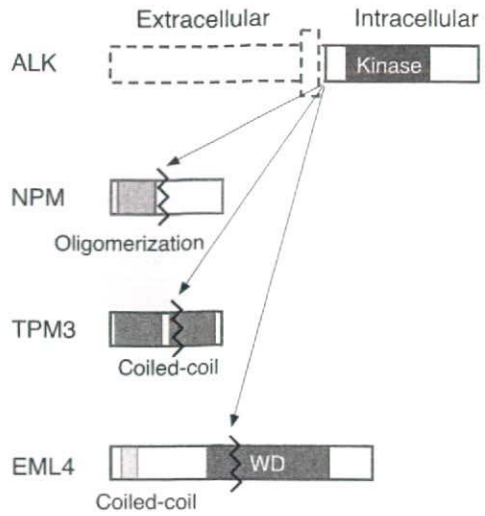

Fig. 2. Anaplastic lymphoma kinase (ALK) fusion proteins. Chromosome rearrangements result in the generation of fusion genes that encode an oligomerization domain (including the coiled coil domain) of nucleophosmin (NPM), tropomyosin (TPM3) or echinoderm microtubule-associated protein like-4 (EML4) fused to the intracellular region of the receptor-type protein tyrosine kinase ALK. WD, WD-repeat domain.

fusion-type PTK between the amino-terminal half of EML4 and the intracellular region of ALK (Fig. 2).[13]

Anaplastic lymphoma kinase was originally identified in anaplastic large cell lymphoma with t(2;5), as a fusion protein to nucleophosmin (NPM).[37,38] NPM–ALK plays an essential role in the lymphomagenesis of this subtype, and is a promising target for therapeutic compounds,[39,40] as is the case for BCR–ABL. In addition to NPM–ALK, ALK kinase may be fused, albeit at a lower frequency, to different partner proteins in the lymphoma through various chromosome translocations, giving rise to TRK-fused gene (TFG)–ALK, 5-aminoimidazole-4-carboximide ribonucleotide formyltransferase/IMP cyclohydrolase (ATIC)–ALK, clathrin, heavy chain (CLTC)–ALK, and others.[41] Subsequently, a non-hematological neoplasm, inflammatory myofibroblastic tumor (IMT) was also shown to harbor ALK fusion proteins such as TPM3–ALK, tropomyosin (TPM)4–ALK, and CLTC–ALK.[41] However, any recurrent translocation involving the *ALK* locus had not been reported for epithelial tumors before our discovery of *EML4–ALK*.

Interestingly, the *ALK* part of our *EML4–ALK* cDNA starts from exon 20 of *ALK*, which is also the fusion point in the vast majority of the other *ALK* fusion cDNA molecules, suggesting the presence of a common fragile locus within intron 19 of the *ALK* gene (Fig. 2). Both the *EML4* and *ALK* genes are mapped closely to the short arm of human chromosome 2 in opposite directions (Fig. 3a). Therefore, a chromosome segment encompassing the *EML4* and *ALK* loci has to become inverted to produce *EML4–ALK* cDNA. We have indeed succeeded to amplify by PCR a genome fragment from a NSCLC specimen that contained the fusion point between the *EML4* and *ALK* genes.[13] In this adenocarcinoma, the *EML4* gene was disrupted at a position 3.6 kb downstream of exon 13, and inverted to become ligated to a position ~300 bp upstream of *ALK* exon 20, proving the presence of inv(2)(p21p23) in the cancer cells.

Therefore, despite the previous notion that epithelial tumors seldom carry fusion-type oncogenes, we discovered an example of a fusion-type PTK with marked oncogenic activity in lung cancer, generated through a chromosome translocation, as is the case for BCR–ABL1, NPM–ALK, and translocation, ETS, leukemia (TEL)–Janus kinase 2 (JAK2) in hematological malignancies.[42] Because mutated EGFR and KRAS have been

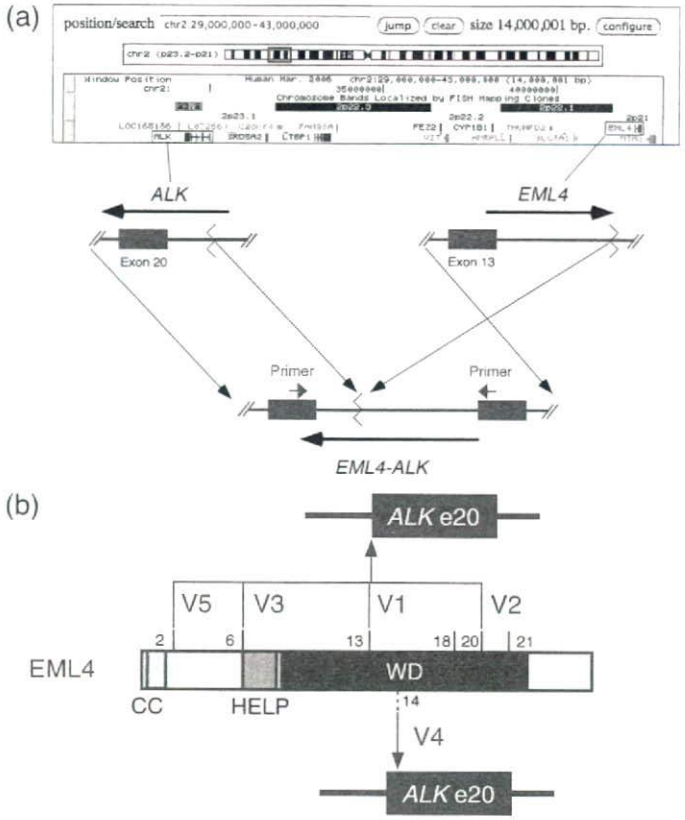

Fig. 3. Diversity in the fusion points between *echinoderm microtubule-associated protein like-4 (EML4)* and *anaplastic lymphoma kinase (ALK)*. (a) *EML4* and *ALK* map in opposite orientations to the short arm of human chromosome 2 (shown in the genome browser of the University of California, Santa Cruz; http://genome.ucsc.edu/cgi-bin/hggateway). Intron 13 of *EML4* is ligated to intron 19 of *ALK* through a chromosome rearrangement, inv(2)(p21p23), generating the *EML4–ALK* (variant 1) fusion gene. Primers flanking the fusion point can be used for the molecular detection of *EML4–ALK*-positive tumors by polymerase chain reaction. Arrows indicate the direction of transcription. (b) Exon boundaries of *EML4* for possible in-frame fusion to exon 20 of *ALK* are shown as vertical bars together with the exon numbers at the corresponding positions in the EML4 protein. Reverse tramscription–polymerase chain reaction screening has identified variants (V) 1, 2, 3, and 5 of *EML4–ALK*, in which exons 13, 20, 6, and 2, respectively, of *EML4* cDNA are fused to exon 20 (e20) of *ALK* cDNA. Unexpectedly, another in-frame fusion was identified in variant 4 cDNA, in which exon 14 of *EML4* was fused via an 11-bp sequence of unknown origin to the nucleotide at position 50 of *ALK* exon 20. CC, coiled coil domain; HELP, hydrophobic EMAL-like protein domain; WD, WD-repeat domain.

found recurrently in NSCLC cells, it is of clinical relevance whether EML4–ALK coexists in cancer cells with active EGFR or KRAS. Interestingly, the presence of EML4–ALK seems to be mutually exclusive to that of EGFR or KRAS mutations in NSCLC,[13,15,43,44] albeit with some exceptions.[45] Therefore, it is likely that EML4–ALK-positive lung cancer forms a subgroup among NSCLC, distinct from that positive for mutated EGFR or KRAS.

Molecular detection of *EML4–ALK*-positive NSCLC

One of the main reasons for the poor prognosis in lung cancer is the lack of sensitive detection methods that can capture tumor cells at early clinical stages (where tumors may be surgically removed). Although pathological examination of sputa and other clinical specimens is used routinely for the diagnosis of lung cancer, reliable detection with such systems usually requires that cancer cells occupy at least a small percentage of the total cells in these specimens. Therefore, patients diagnosed with this