As shown in the bottom panel of Fig. 1, tumor formation was readily observed for the cells expressing P2RY2(P46L/R312S), wild-type P2RY2, and v-Ras. These results clearly revealed an unexpected transforming potential of P2RY2 regardless of the amino-acid substitutions identified in our clone.

Discussion

In this study, we have constructed a retroviral cDNA expression library for a CRC cell line RKO. Since 95% (37/39 clones) of the viral plasmids carried cDNA inserts and since the overall clone number was >2 millions, our library should cover nearly all transcriptome in RKO cells.

Purinergic receptors may be subdivided to two families: a P2X family of ligand-gated ion-channel receptors and a P2Y family of G protein-coupled purinoreceptors [8]. Currently, seven or eight members have been reported for the P2X or P2Y subtype, respectively. It is widely recognized that purinergic receptor system may be linked to a wide array of cellular processes, including immune responses, inflammation, pain, platelet aggregation, endothelial-mediated vasodilatation, cell proliferation, and death [9,10].

Interestingly, contrasting effects of the P2 receptors have been documented in some tumor models. Activation of P2 receptors may lead to cell growth in various cancer cells [11,12]. In response to nucleotide activation, P2Y receptors can activate phospholipase C, mobilize intracellular calcium, and alter cAMP levels. On the contrary, P2 receptors may provide signals for anti-proliferation and apoptosis [13,14]. Indeed, ATP, at high concentrations, induces apoptosis through ligation of P2X7 and P2Y1 receptors [15]. Conversely, however, a low concentration of ATP stimulates proliferation. These data suggest that a plethora of functions can be provided by P2 receptors probably in a dependent manner to the receptor subtype, cellular context, stimuli or the concentration of stimuli.

Our data clearly indicate the transforming potential of P2RY2 at least in the mouse fibroblast cell line, suggesting the possible oncogenic activity of P2RY2 in human cancers. Interestingly, Rapaport et al. demonstrated that addition of ATP or ADP in the culture medium modulates the cell cycle of cell lines [16]. Therefore, regulation of P2RY2 or its ligands may provide a novel strategy to develop means to treat cancers.

Acknowledgments

This work was supported in part by grants for Research on Human Genome and Tissue Engineering and 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.

References

[1] A. Jemal, R. Siegel, E. Ward, T. Murray, J. Xu, C. Smigal, M.J. Thun, Cancer statistics, 2006, CA Cancer J. Clin. 56 (2006) 106–130.
[2] E.R. Fearon, B. Vogelstein, A genetic model for colorectal tumorigenesis, Cell 61 (1990) 759–767.
[3] S.A. Aaronson, Growth factors and cancer, Science 254 (1991) 1146–1153.
[4] W.S. Pear, G.P. Nolan, M.L. Scott, D. Baltimore, Production of high-titer helper-free retroviruses by transient transfection, Proc. Natl. Acad. Sci. USA 90 (1993) 8392–8396.
[5] Y.L. Choi, R. Kaneda, T. Wada, S. Fujiwara, M. Soda, H. Watanabe, K. Kurashina, H. Hatanaka, M. Enomoto, S. Takada, Y. Yamashita, H. Mano, Identification of a constitutively active mutant of JAK3 by retroviral expression screening, Leuk. Res. 31 (2007) 203–209.
[6] S. Fujiwara, Y. Yamashita, Y.L. Choi, T. Wada, R. Kaneda, S. Takada, Y. Maruyama, K. Ozawa, H. Mano, Transforming activity of the lymphotoxin-beta receptor revealed by expression screening, Biochem. Biophys. Res. Commun. 338 (2005) 1256–1262.
[7] M. Onishi, S. Kinoshita, Y. Morikawa, A. Shibuya, J. Phillips, L.L. Lanier, D.M. Gorman, G.P. Nolan, A. Miyajima, T. Kitamura, Applications of retrovirus-mediated expression cloning, Exp. Hematol. 24 (1996) 324–329.
[8] G. Burnstock, Introduction: P2 receptors, Curr. Top. Med. Chem. 4 (2004) 793–803.
[9] K. van Kolen, H. Slegers, Atypical PKCzeta is involved in RhoA-dependent mitogenic signaling by the P2Y(12) receptor in C6 cells, FEBS J. 273 (2006) 1843–1854.
[10] L.A. Sellers, J. Simon, T.S. Lundahl, D.J. Cousens, P.P. Humphrey, E.A. Barnard, Adenosine nucleotides acting at the human P2Y1 receptor stimulate mitogen-activated protein kinases and induce apoptosis, J. Biol. Chem. 276 (2001) 16379–16390.
[11] C.J. Dixon, W.B. Bowler, P. Fleetwood, A.F. Ginty, J.A. Gallagher, J.A. Carron, Extracellular nucleotides stimulate proliferation in MCF-7 breast cancer cells via P2-purinoceptors, Br. J. Cancer 75 (1997) 34–39.
[12] A.V. Greig, C. Linge, V. Healy, P. Lim, E. Clayton, M.H. Rustin, D.A. McGrouther, G. Burnstock, Expression of purinergic receptors in non-melanoma skin cancers and their functional roles in A431 cells, J. Invest. Dermatol. 121 (2003) 315–327.
[13] M. Hopfner, K. Maaser, B. Barthel, B. von Lampe, C. Hanski, E.O. Riecken, M. Zeitz, H. Scherubl, Growth inhibition and apoptosis induced by P2Y2 receptors in human colorectal carcinoma cells: involvement of intracellular calcium and cyclic adenosine monophosphate, Int. J. Colorectal Dis. 16 (2001) 154–166.
[14] K. Maaser, M. Hopfner, H. Kap, A.P. Sutter, B. Barthel, B. von Lampe, M. Zeitz, H. Scherubl, Extracellular nucleotides inhibit growth of human oesophageal cancer cells via P2Y(2)-receptors, Br. J. Cancer 86 (2002) 636–644.
[15] R. Coutinho-Silva, L. Stahl, K.K. Cheung, N.E. de Campos, C. de Oliveira Souza, D.M. Ojcius, G. Burnstock, P2X and P2Y purinergic receptors on human intestinal epithelial carcinoma cells: effects of extracellular nucleotides on apoptosis and cell proliferation, Am. J. Physiol. Gastrointest. Liver Physiol. 288 (2005) G1024–G1035.
[16] E. Rapaport, Treatment of human tumor cells with ADP or ATP yields arrest of growth in the S phase of the cell cycle, J. Cell. Physiol. 114 (1983) 279–283.

Leukemia & Lymphoma, May 2007; 48(5): 978–986

informa
healthcare

ORIGINAL ARTICLE: RESEARCH

Transforming activity of purinergic receptor P2Y, G protein coupled, 8 revealed by retroviral expression screening

SHIN-ICHIRO FUJIWARA[1,2], YOSHIHIRO YAMASHITA[1], YOUNG LIM CHOI[1], HIDEKI WATANABE[1], KENTARO KURASHINA[1], MANABU SODA[1], MUNEHIRO ENOMOTO[1], HISASHI HATANAKA[1], SHUJI TAKADA[1], KEIYA OZAWA[2], & HIROYUKI MANO[1,3]

Divisions of [1]Functional Genomics and [2]Hematology, Jichi Medical University, Tochigi 329-0498, Japan, and [3]CREST, Japan Science and Technology Agency, Saitama 332-0012, Japan

(Received 29 December 2006; revised 9 January 2007; accepted 13 January 2007)

Abstract

Biphenotypic acute leukemia (BAL) is a relatively rare subtype of acute leukemia characterized by the presence of both myeloid and lymphoid cell surface antigens. We have now screened for transforming genes in BAL blasts with the use of the focus formation assay with a retroviral cDNA expression library constructed from malignant blasts isolated from a BAL patient. Some of the retroviral inserts recovered from transformed foci were found to encode wild-type purinergic receptor P2Y, G protein coupled, 8 (P2RY8). The oncogenic potential of P2RY8 was confirmed with the in vitro focus formation assay as well as with an in vivo tumorigenicity assay in nude mice. A variety of luciferase-based reporter assays revealed that P2RY8 increased both the trans-activation activities of CREB and Elk-1 as well as the transcriptional activities of the serum response element and enhancer-promoter fragments of the c-Fos and c-Myc genes. Quantitation of *P2RY8* mRNA in $CD34^+$ cells of bone marrow showed that *P2RY8* expression is frequently increased in leukemia patients, especially in those with refractory disease. Our data thus reveal an abundant expression of P2RY8 in leukemic cells and its unexpected role in the pathogenesis of acute leukemia.

Keywords: *Purinergic receptor, P2RY8, biphenotypic acute leukemia, retroviral expression library*

Introduction

Malignant blasts of acute leukemia are classified as of the myeloid or lymphoid lineage on the basis of immunophenotypic features and the presence or absence of gene rearrangement within the T cell receptor or immunoglobulin gene loci. In relatively rare instances, however, such blasts coexpress cell surface antigens of both lineages [1,2], with the underlying condition being referred to as biphenotypic acute leukemia (BAL) according to the revised World Health Organization classification [3]. A scoring system for the diagnosis of BAL has also been proposed by the European Group for the Immunological Characterization of Leukemias [4].

The clinical outcome of BAL is generally poor [5], and it remains controversial as to whether BAL should be treated by regimens for acute myelocytic leukemia (AML) or acute lymphocytic leukemia (ALL). The unfavorable prognosis for BAL may be related to the associated Ph1 positivity, chromosome rearrangements affecting 11q23, and overexpression of P-glycoprotein [1,6]. However, little is yet understood of the molecular pathogenesis of this intractable disorder. Elucidation of oncogenic events in BAL might provide insight into possible new treatment strategies.

The focus formation assay with 3T3 or Rat1 fibroblasts is a functional screening system for transforming genes [7] and has been extensively applied to

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

ISSN 1042-8194 print/ISSN 1029-2403 online © 2007 Informa UK Ltd.
DOI: 10.1080/10428190701225882

various human cancers. However, the conventional format for this assay may not allow expression of all transforming genes in the target cells. Given that the genomic DNA of clinical specimens is introduced directly into the recipient cells, the expression of the exogenous genes is controlled by their own promoters or enhancers. Oncogenes are thus able to exert their effects in the recipient cells only if these regulatory regions are active in fibroblasts, which is not always the case.

Regulation of the transcription of test cDNAs by a promoter known to function efficiently in fibroblasts would be expected to ensure sufficient expression of the encoded proteins in the focus formation assay. We have now constructed a retroviral cDNA expression library from a $CD34^+$ cell fraction isolated from a patient with BAL, and we have tested this library in the focus formation assay with 3T3 cells. For library construction, we took advantage of a polymerase chain reaction (PCR) system that preferentially amplifies full-length cDNAs.

With this assay system, we found that purinergic receptor P2Y, G protein coupled, 8 (P2RY8) possesses transforming activity. We could further demonstrate the relatively abundant expression of *P2RY8* message in leukemic cells. P2RY8 belongs to the G protein–coupled purinergic receptor family, members of which preferentially bind to, and are activated by, adenosine and uridine nucleotides [8,9]. Although the *P2RY8* gene was proved to be disrupted in individuals with X-linked mental retardation [10], its expression profile has been rarely documented in humans. Our data thus reveals an unexpected role of P2RY8 in human leukemia.

Materials and Methods

Clinical specimens and cell lines

Fresh clinical specimens were obtained from individuals with leukemia and healthy volunteers, with informed consent, and the study was approved by the institutional review board of Jichi Medical University. Mononuclear cells were isolated from bone marrow aspirates by centrifugation, labeled with CD34 MicroBeads (Miltenyi Biotec, Gladbach, Germany), and subjected to chromatography on miniMACS magnetic cell separation columns (Miltenyi Biotec). The purity of the resultant $CD34^+$ cell fractions was confirmed by staining with Wright–Giemsa solution and analysis of CD34 expression by flow cytometry (data not shown).

BOSC23 [11] and 3T3 cell lines were obtained from American Type Culture Collection and maintained in Dulbecco's modified Eagle's medium (DMEM)–F12 (Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum (Invitrogen) and 2 mM L-glutamine.

Construction of a retroviral cDNA expression library

A cDNA library was constructed essentially as described previously [12–14]. In brief, total RNA was extracted from the $CD34^+$ cells of a patient with BAL with the use of an RNeasy Mini column and RNase-free DNase (Qiagen, Valencia, CA) and was subjected to first-strand cDNA synthesis with PowerScript reverse transcriptase, the SMART IIA oligonucleotide, and CDS primer IIA (all from Clontech, Palo Alto, CA). The resulting cDNAs were amplified for 20 cycles with 5′ PCR primer IIA and a SMART PCR cDNA synthesis kit (Clontech), with the exception that LA *Taq* polymerase (Takara Bio, Shiga, Japan) was substituted for the Advantage 2 DNA polymerase provided with the kit. The amplified cDNAs were treated with proteinase K, rendered blunt-ended with T4 DNA polymerase, and ligated to the BstXI-adaptor (Invitrogen). Unbound adaptors were removed with the use of a cDNA size-fractionation column (Invitrogen), and the remaining cDNAs were ligated into the BstXI site of the pMXS retroviral plasmid (kindly provided by T. Kitamura, Institute of Medical Science, University of Tokyo). The resulting pMXS-cDNA plasmids were introduced into ElecroMax DH10B cells (Invitrogen) by electroporation.

Focus formation assay

BOSC23 cells (2.0×10^6) were seeded into a 6-cm culture dish and transfected with 2 μg of retroviral plasmids mixed with 0.5 μg of the pGP plasmid, 0.5 μg of the pE-eco plasmid (both from Takara Bio), and 18 μl of Lipofectamine reagent (Invitrogen). After 2 days, polybrene (Sigma, St. Louis, MO) was added to the culture supernatant at a concentration of 4 μg/ml, and the supernatant was then used to infect 3T3 cells for 48 h. The culture medium of the 3T3 cells was then changed to DMEM-F12 supplemented with 5% calf serum and 2 mM L-glutamine, and the cells were cultured for 2 weeks.

Recovery of cDNA from transformed foci

Transformed 3T3 cell clones were harvested with a cloning syringe and cultured independently in 10-cm culture dishes. Genomic DNA was extracted from each clone by standard procedures and then subjected to PCR with 5′ PCR primer IIA and LA *Taq* polymerase for 50 cycles of 98°C for 20 s and 68°C for 6 min. Amplified DNA fragments were purified by gel electrophoresis and ligated into the pT7Blue-2

vector (EMD Biosciences, San Diego, CA) for nucleotide sequencing.

Tumorigenicity assay in nude mice

3T3 cells (2×10^6) infected with a retrovirus encoding P2RY8 were resuspended in 500 μl of phosphate-buffered saline and injected into each shoulder of a nu/nu Balb-c mouse (6 weeks old). Tumor formation was assessed after 3 weeks.

Luciferase reporter assays

PathDetect In Vivo Signal Transduction Pathway *Trans*-Reporting Systems (Stratagene, La Jolla, CA) were used to measure the activity of the transcription factors CREB, Elk-1, and c-Jun. BOSC23 cells (5.0×10^5) were seeded into 6-cm culture dishes and transfected with a mixture of 500 ng of the P2RY8 expression vector, 125 ng of the PathDetect *Trans*-Reporter Activator plasmid (pFA-CREB, pFA-Elk1, or pFA-cJun), 125 ng of the PathDetect *Trans*-Reporter plasmid pFR-Luc, 7 ng of the pGL4 vector (Promega), and 4.5 μl of Lipofectamine reagent. Similarly, the activation of p53, Rb, and E2F proteins was assessed with the use of a Mercury Pathway Profiling System (Clontech); the P2RY8 expression vector (500 ng) and pGL4 (7 ng) were thus introduced into BOSC23 cells together with 125 ng of pp53-TA-Luc, pRb-TA-Luc, or pE2F-TA-Luc vectors.

PathDetect *Cis*-Reporting Systems (Stratagene) were used to evaluate the transcriptional activities of the serum response element (SRE) and an NF-κB binding sequence. BOSC23 cells were transfected with a mixture of 500 ng of the P2RY8 expression vector, 125 ng of the PathDetect *Cis*-Reporter plasmid (pNF-κB-Luc or pSRE-Luc), 7 ng of pGL4, and 4.5 μl of Lipofectamine. The transcriptional activities of enhancer-promoter fragments of the c-Fos [15], c-Myc [16], and Bcl-x_L [17] genes were similarly assessed with luciferase expression vectors controlled by the corresponding enhancer-promoter elements (pFL700, pHXLuc, and pBclxl-Luc, respectively).

After transfection for 2 days, the BOSC23 cells were solubilized with lysis buffer (Promega), and luciferase activities were measured with a Luciferase Assay System (Promega). Firefly luciferase activity was normalized relative to that of the corresponding Renilla luciferase activity.

Quantitative reverse transcription (RT)-PCR analysis

Portions of oligo(dT)-primed cDNAs produced by reverse transcription were subjected to PCR with a QuantiTect SYBR Green PCR kit (Qiagen) and an amplification protocol comprising incubation at 94°C for 15 s, 60°C for 30 s, and 72°C for 60 s. Incorporation of the SYBR Green dye into PCR products was monitored in real time with an ABI PRISM 7900HT sequence detection system (PE Applied Biosystems, Foster City, CA), thereby allowing determination of the threshold cycle (C_T) at which exponential amplification of PCR products begins. The C_T values for cDNAs corresponding to the β-actin gene (*ACTB*) and to *P2RY8* were used to calculate the abundance of the latter mRNA relative to that of the former. The oligonucleotide primers used for PCR were 5′-CCATCATGAAGTGTGA CGTGG-3′ and 5′-GTCCGCCTAGAAGCATTT GCG-3′ for *ACTB* and 5′-TTCCTCTTCACCATC TTCATCCTG-3′ and 5′-CGTGGTAGTAGCTC TTGCCGTAGA-3′ for *P2RY8*.

Results

Screening for transforming genes with the focus formation assay

To screen for transforming genes in BAL, we constructed a cDNA expression library from a $CD34^+$ cell fraction purified from mononuclear cells of bone marrow from a patient with this condition. Full-length cDNAs were preferentially amplified by the SMART protocol (Clontech) and were ligated into the retroviral vector pMXS. We obtained a total of 4.2×10^6 colony-forming units of independent plasmid clones, from which we randomly selected 30 clones and examined the incorporated cDNAs. An insert of ≥500 bp was present in 27 (90%) of the plasmid clones, and the average size of these inserts was 1.94 kbp.

Introduction of the library plasmids into a packaging cell line generated a recombinant ecotropic retroviral library, which was then used to infect mouse NIH 3T3 fibroblasts. After culture of the infected cells for 2 weeks, we identified a total of 67 transformed foci (Figure 1). No foci were observed for 3T3 cells infected with the empty virus. Each transformed focus was isolated, expanded individually, and used to prepare genomic DNA. PCR amplification of the inserts yielded a total of 30 cDNA fragments, each of which was ligated into a cloning vector and subjected to nucleotide sequencing from both ends. Screening of these cDNA sequences against the public nucleotide sequence databases revealed that they showed >95% sequence identity to five independent known genes.

To confirm the transforming ability of the isolated cDNAs, we again ligated them into the pMXS vector and used the corresponding retroviruses to infect

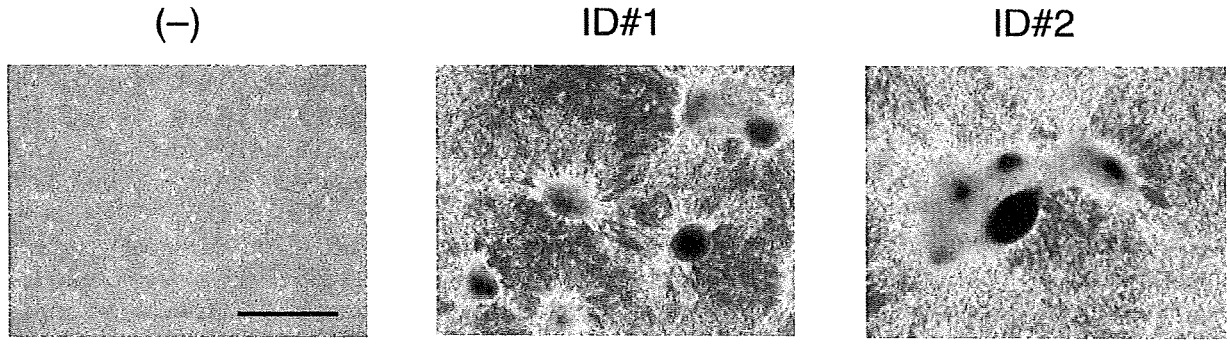

Figure 1. Screening for transforming genes from a patient with BAL. Mouse 3T3 cells were infected with a recombinant retrovirus library or the corresponding empty virus (–). Complementary DNAs for *P2RY8* and *KRAS2* were recovered from 3T3 clones ID No. 1 and ID No. 2, respectively. The photographs show the infected 3T3 cells after culture for 2 weeks. Scale bar, 1 mm.

3T3 cells. Two of the five independent genes, *P2RY8* (GenBank accession no. NM_178129) and *KRAS2* (GenBank accession no. NM_004985), reproducibly induced the formation of transformed foci in 3T3 cells. Further sequencing of the *KRAS2* cDNA revealed that it contained a point mutation that results in a glycine to alanine substitution at codon 12, a change that is known to activate KRAS2 [18]. Given that, as far as we were aware, transforming activity had not previously been described for P2RY8, we focused on P2RY8 in further analyses.

Confirmation of the transforming activity of P2RY8

The complete nucleotide sequence was determined for the *P2RY8* cDNA isolated from 3T3 clone ID No. 1, revealing that the cDNA comprised 4208 bp encoding for a full-length protein of 359 amino acids with a sequence identical to that of P2RY8. Within the protein-coding region, the cDNA sequence differed by only one nucleotide compared with that of the *P2RY8* sequence in the database; the T at nucleotide position 697 in the reported sequence (NM_178129) is replaced with a C in the determined sequence. However, this change does not affect the amino acid sequence of P2RY8 and may therefore reflect a single nucleotide polymorphism. There was no difference in the nucleotide sequence of the 5′ untranslated region between our *P2RY8* cDNA and the deposited one. There were, however, four changes in the 3′ untranslated region of our cDNA compared to that of the reported one; deletion of a A at the position 3599 in the deposited cDNA sequence, a G-to-A transition at the position 3749, deletion of a AA at the positions of 4169–4170, and deletion of a C at the position of 4194. While all of these changes are located at the AT-rich region close to the polyadenylation signal, it is possible that some of these changes may affect the quantity of the protein product.

To confirm that overexpression of wild-type P2RY8 induces transformation of 3T3 cells, we repeated the focus formation assay with a retrovirus generated from the isolated *P2RY8* cDNA. Infection of the cells with the retrovirus resulted in the generation of many foci (data not shown). The transforming potential of P2RY8 was also evaluated in a tumorigenicity assay with athymic nude mice. The animals were thus injected subcutaneously with 3T3 cells infected with the retrovirus encoding P2RY8. Tumor formation was observed at all 10 injection sites. In contrast, no tumors were detected at sites injected with 3T3 cells infected with the empty virus (Figure 2). Injection of nude mice with 3T3 cells infected with a virus encoding v-Ras also induced the formation of subcutaneous tumors at four out of four injection sites (data not shown).

Signal transduction pathways downstream of P2RY8

To identify signal transduction pathways activated by P2RY8 overexpression, we performed a series of luciferase reporter assays. First, the activity of several signaling molecules involved in cell growth or apoptosis was measured by two commercially available systems [Figure 3(A)]. With the use of PathDetect In Vivo Signal Transduction Pathway *Trans*-Reporting Systems (Stratagene), a fusion *trans*-activator protein comprised of the yeast GAL4 DNA-binding domain and the activation domain of either c-Jun, CREB or Elk-1 was expressed in BOSC23 cells, together with the firefly luciferase under the control of the yeast GAL4 binding sequence. Coexpression of P2RY8 resulted in the activation of the *trans*-activator for CREB or Elk-1, which thereby induced the marked production of the luciferase protein [Figure 3(A)].

With the Mercury Pathway Profiling System (Clontech), a plasmid containing the target sequence of p53, Rb, or E2F (each fused to the luciferase cDNA) was introduced into BOSC23 cells with or

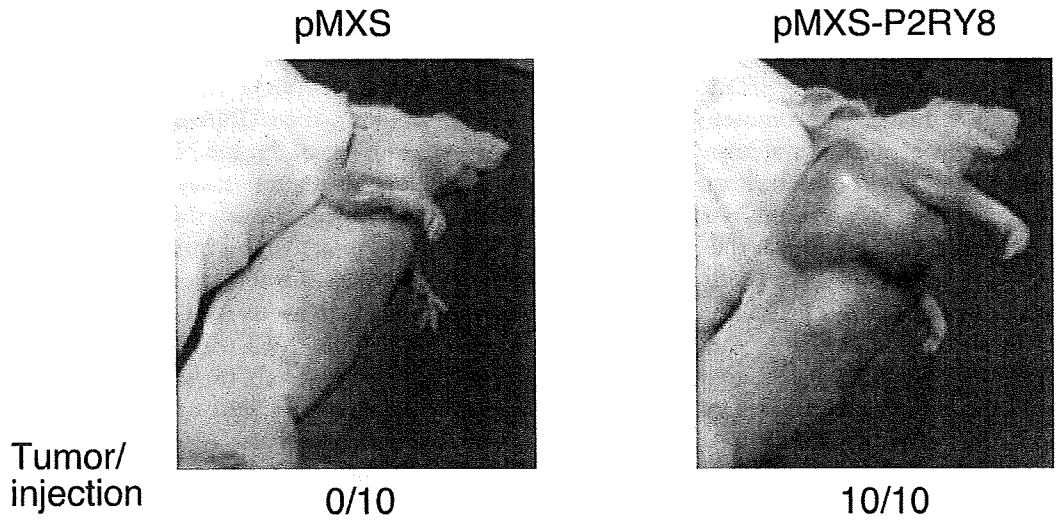

Figure 2. Tumorigenicity assay with nude mice. 3T3 cells infected with a retrovirus encoding P2RY8 (pMXS-P2RY8) or the corresponding empty virus (pMXS) were injected into the shoulder of nu/nu mice. Tumor formation was examined after 3 weeks. The number of tumors formed at the 10 injection sites is indicated.

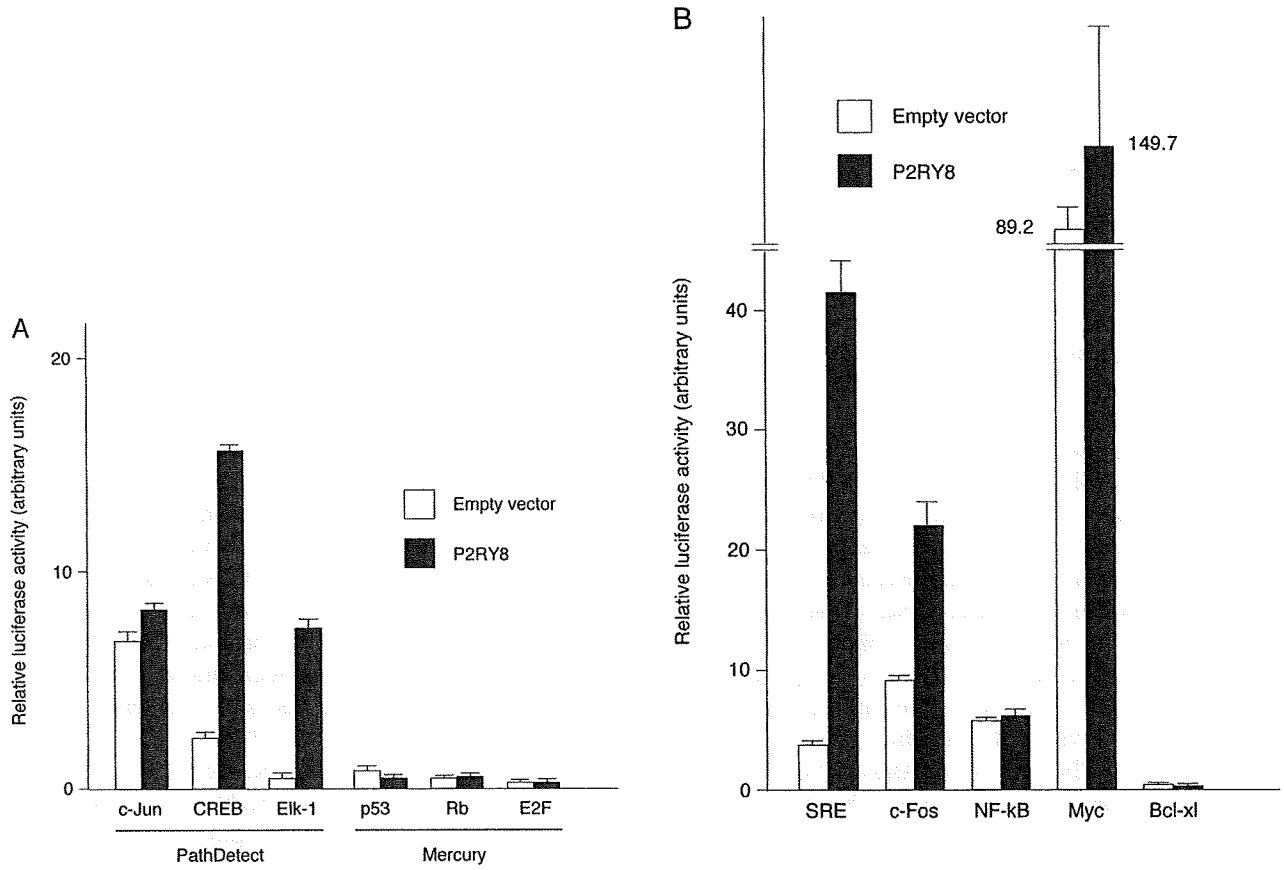

Figure 3. Activation of signal transduction pathways by P2RY8. (A) The trans-activation activities of the transcription factors (c-Jun, CREB, and Elk-1) were assessed with the PathDetect In Vivo Signal Transduction Pathway *Trans*-Reporting Systems (PathDetect) in BOSC23 cells transfected with a P2RY8 expression vector or the corresponding empty vector. Similarly, the activity of p53, Rb, or E2F was measured with the use of Mercury Pathway Profiling System (Mercury). The activity of firefly luciferase in cell lysates was normalized by that of Renilla luciferase. Data are means + SD of values from three experiments. (B) The SRE sequence, the enhancer/promoter fragment of the c-Fos, c-Myc, or Bcl-x_L gene, or the NF-κB binding sequence was fused to the cDNA of firefly luciferase, and was introduced into BOSC23 cells. The activity of firefly luciferase in cell lysates was normalized by that of Renilla luciferase. Data are means + SD of values from three experiments.

without the expression plasmid for P2RY8. As demonstrated in the right panel of Figure 3(A), none of the target sequences were activated by P2RY8.

Next, the activity of enhancer/promoter sequences was measured with *cis* reporting systems. The SRE fragment, the enhancer/promoter sequence for c-Fos, c-Myc, or Bcl-x_L gene, or the NF-κB binding sequence was individually fused to the luciferase cDNA, and was introduced into BOSC23 cells. While coexpression of P2RY8 failed to substantially affect the activity of NF-κB or that of the enhancer/promoter fragments of Bcl-x_L, it resulted in activation of SRE and the enhancer/promoter fragments of the c-Fos and c-Myc genes [Figure 3(B)].

Expression of P2RY8 in leukemia specimens

Given that overexpression of *P2RY8* was found to induce cell transformation, we compared the abundance of *P2RY8* mRNA by quantitative RT-PCR analysis among $CD34^+$ progenitor cells isolated from the bone marrow of healthy volunteers ($n=3$) as well as from individuals with AML developed from myelodysplastic syndrome (MDS) ($n=3$), AML developed from myeloproliferative disorder (MPD) ($n=1$), de novo AML ($n=15$), ALL ($n=4$), or BAL ($n=2$, including the patient used for cDNA library construction).

Overall, the amount of *P2RY8* mRNA in $CD34^+$ progenitor cells was increased in patients with acute leukemia compared with that in healthy volunteers, and this increase was apparent across all the subtypes of leukemia (de novo or secondary AML, ALL, and BAL) examined [Figure 4(A)]. To determine whether an increased expression of *P2RY8* is associated with unfavorable characteristics of leukemia, we compared the amount of *P2RY8* mRNA between untreated patients and those with relapsed or refractory leukemia. The abundance of *P2RY8* mRNA was indeed significantly greater ($P=0.0054$) in $CD34^+$ cells from the latter group of patients than in those from the former group [Figure 4(B)]. To examine further the relation between a high level of *P2RY8* expression and poor clinical outcome, we compared the survival of leukemia patients with a *P2RY8/ACTB* mRNA ratio in blasts of ≥ 0.006 with that of those with a ratio of <0.006. Kaplan–Meier analysis revealed that the patients with a low level of *P2RY8* expression showed a significantly better survival ($P < 0.05$) than did those with a high level of *P2RY8* expression [Figure 4(C)].

Discussion

We constructed a retroviral cDNA expression library with $CD34^+$ cells isolated from the bone marrow of a patient with BAL. Given that the plasmid library was shown to harbor cDNA inserts in $\geq 90\%$ of the clones and that the overall independent clone number was >4 million, the cDNA expression library likely represented most of the mRNAs in the $CD34^+$ cells. A 3T3 focus formation assay with the recombinant viruses generated from the plasmids identified *KRAS2* with an activating mutation and a full-length *P2RY8* cDNA as having transforming activity. The transforming activity of wild-type P2RY8 was confirmed with both the focus formation assay and a tumorigenicity assay in nude mice. We also examined the expression level of *P2RY8* in 3T3 cells by quantitative RT-PCR as in Figure 4(A). Although 3T3 cells infected with an empty virus had no detectable amount of *P2RY8* mRNA, the cells infected with the virus expressing P2RY8 had an abundant expression of *P2RY8* message (*P2RY8/ACTB* mRNA ratio was 48.7 ± 6.1, mean $\pm$ SD).

P2RY8 is presumed to function as a transmembrane receptor for adenosine and uridine nucleotides and to couple with heterotrimeric G-proteins [9]. Upon binding to nucleotides, P2Y receptors trigger activation of heterotrimeric G-proteins and thereby evoke intracellular signaling through adenylyl cyclase, extracellular signal-regulated kinases (ERKs), and/or phospholipase C [8,19]. Purinergic receptor signaling has been shown to be associated with a variety of cellular responses including cell growth [20,21], differentiation [22], migration [23], and apoptosis [24]. In addition, *XlP2Y* (a homolog of *P2RY8* in *Xenopus laevis*) is expressed in the neural plate of developing embryos [25] and *P2RY8* is disrupted in individuals with X-linked mental retardation [10], observations that suggest an important role for *P2RY8* in neural development. Despite the absence of evidence for a direct linkage between *P2RY8* and oncogenic activity, several reports have suggested the potential relationship between purinergic receptors and leukemia. Extracellular ATP, for instance, suppresses the growth of, and induces the differentiation of, HL-60 leukemia cells [26]. Further, *P2X7* message was shown abundant in some leukemic cells [27]. Our current data indicates the possible contribution of another purinergic receptor, P2RY8, to leukemogenesis. However, it should be noted that, in our hands, a forced expression of *P2RY8* in a mouse cell line, 32Dcl3 [28], failed to abrogate an interleukein-3-dependency for cell growth, and to suppress granulocyte colony-stimulating factor-dependent differentiation toward terminal granulocytes (data not shown). Thus, despite the transforming activity, P2RY8 may require other genetic events to confer full properties as leukemia onto blood cells.

Screening for downstream components of P2RY8 signaling in cells with reporter constructs revealed

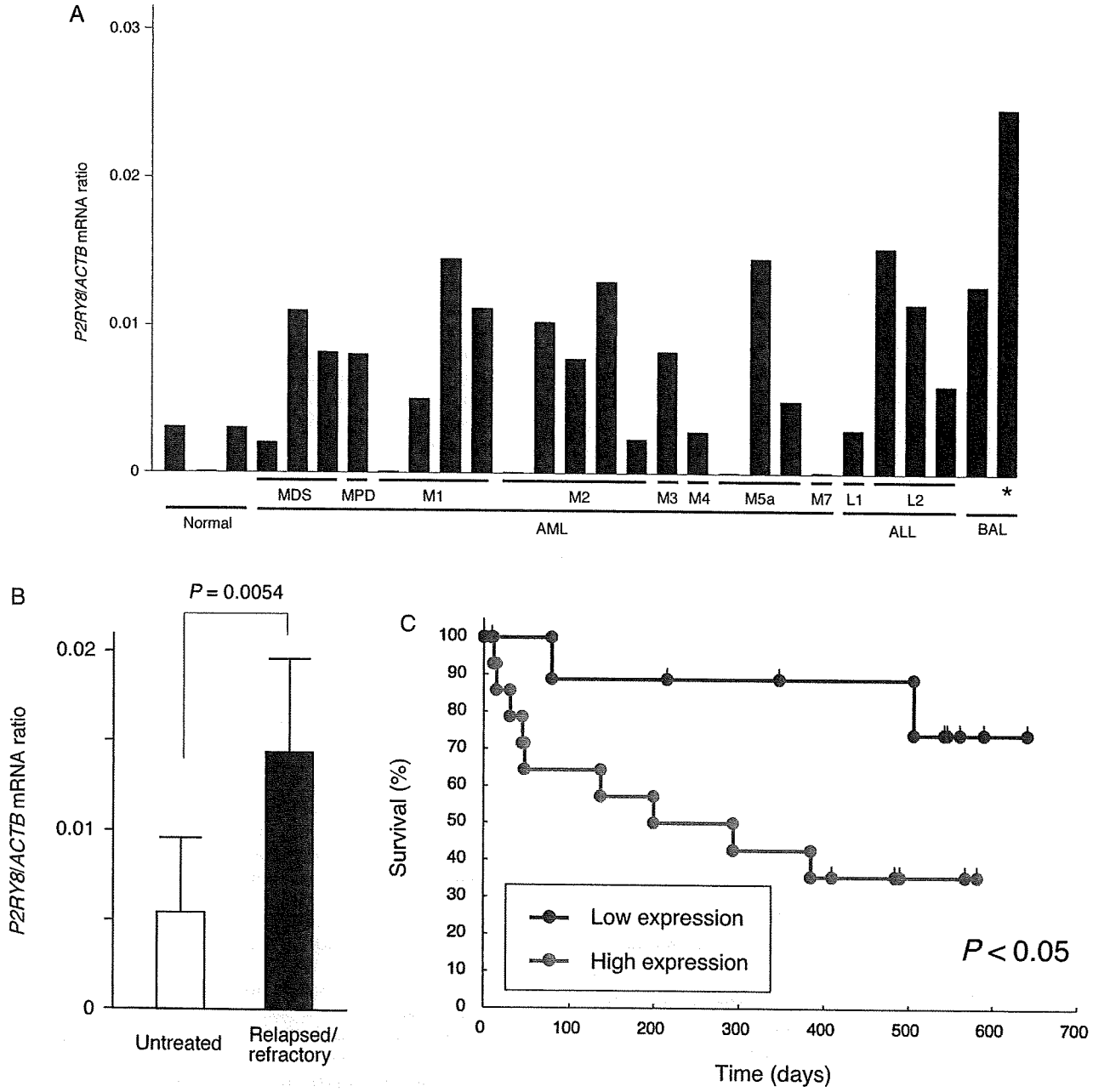

Figure 4. Quantitation of *P2RY8* mRNA in blasts of patients with various types of acute leukemia. (A) The amount of *P2RY8* mRNA in $CD34^+$ cell fractions purified from healthy volunteers (normal) or from patients with MDS-derived AML, MPD-derived AML, de novo AML of various FAB subtypes (M1 to M7), ALL of FAB subtypes L1 or L2, or BAL was determined by real-time RT-PCR analysis. The ratio of the abundance of *P2RY8* mRNA to that of *ACTB* mRNA was calculated as 2^n, where n is the C_T value for *ACTB* cDNA minus the C_T value for *P2RY8* cDNA. The asterisk indicates the patient from whom the cDNA library was constructed. (B) Comparison of the *P2RY8/ACTB* mRNA ratio (mean + SD) between untreated patients ($n = 16$) and those with relapsed or refractory leukemia ($n = 8$). The significance of the difference in the ratio between the two groups was assessed by Student's t-test. (C) Kaplan–Meier plots of the survival of patients with a high level of *P2RY8* expression (*P2RY8/ACTB* mRNA ratio of ≥ 0.006, $n = 15$) compared with that of those with a low level of *P2RY8* expression (*P2RY8/ACTB* mRNA ratio of < 0.006, $n = 10$). The significance of the difference between the two groups was evaluated by the log-rank test.

that the trans-activation activities of CREB and Elk-1 were increased by P2RY8 expression, as were the transcriptional activities of the SRE as well as of enhancer/promoter fragments of the c-Fos and c-Myc genes. Given that both CREB and Elk-1 bind to and activate the promoter of the c-Fos gene and that the SRE is present within this promoter [29,30], these data collectively indicate that c-Fos is a target and effector of P2RY8 signaling. These data are in good agreement with the finding by Muscella et al. that c-Fos can be activated by another purinergic receptor, P2RY6 [19]. Given that the c-Myc gene, an

immediate-early response gene in growth signaling, was also transcriptionally activated by P2RY8, both c-Myc and c-Fos may play an important role in P2RY8-induced malignant transformation.

The second messengers for transformation, acting downstream to P2RY8, are not elucidated yet. The activities of protein kinase C (PKC) were, however, shown essential in the P2RY6-evoked cell proliferation and ERK activation [19]. Since P2RY8 activity may also induce an increase in Ca_i^{2+} [25], it is possible that P2RY8 activates PKC and triggers the machinery for cell proliferation as well.

The finding that overexpression of *P2RY8* results in cell transformation led us to examine whether this gene is overexpressed in leukemic blasts. Quantitative RT-PCR analysis revealed that this was indeed the case, at least for the cells of some patients. A possible contribution of P2RY8 to leukemogenesis was further supported by the observation that patients with a high level of *P2RY8* expression had a poorer prognosis than did those with a lower level. It will be of interest to examine whether the level of *P2RY8* expression is increased in other types of human cancer as well. Especially, given the presumed role of P2RY8 in neural development, it is possible that some brain tumors may have an abundant expression of P2RY8. Our findings implicate the purinergic receptor family in cancer and suggest that P2RY8 is a potential therapeutic target in some types of leukemia.

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, and by a grant for High-Tech Research Center Project for Private Universities: Matching Fund Subsidy (2002–2006) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

References

1. Legrand O, Perrot JY, Simonin G, Baudard M, Cadiou M, Blanc C, et al. Adult biphenotypic acute leukaemia: an entity with poor prognosis which is related to unfavourable cytogenetics and P-glycoprotein over-expression. Br J Haematol 1998;100:147–155.
2. Owaidah TM, Al Beihany A, Iqbal MA, Elkum N, Roberts GT. Cytogenetics, molecular and ultrastructural characteristics of biphenotypic acute leukemia identified by the EGIL scoring system. Leukemia 2006;20:620–626.
3. Jaffe ES, Harris NL, Stein H, Vardiman JW, editors. Pathology and genetics of tumours of haematopoietic and lymphoid tissues. Lyon: IARC Press; 2001.
4. Bene MC, Castoldi G, Knapp W, Ludwig WD, Matutes E, Orfao A, et al. Proposals for the immunological classification of acute leukemias. European Group for the Immunological Characterization of Leukemias (EGIL). Leukemia 1995;9:1783–1786.
5. Killick S, Matutes E, Powles RL, Hamblin M, Swansbury J, Treleaven JG, et al. Outcome of biphenotypic acute leukemia. Haematologica 1999;84:699–706.
6. Carbonell F, Swansbury J, Min T, Matutes E, Farahat N, Buccheri V, et al. Cytogenetic findings in acute biphenotypic leukaemia. Leukemia 1996;10:1283–1287.
7. Aaronson SA. Growth factors and cancer. Science 1991;254:1146–1153.
8. Burnstock G. Introduction: P2 receptors. Curr Top Med Chem 2004;4:793–803.
9. Burnstock G. Pathophysiology and therapeutic potential of purinergic signaling. Pharmacol Rev 2006;58:58–86.
10. Cantagrel V, Lossi AM, Boulanger S, Depetris D, Mattei MG, Gecz J, et al. Disruption of a new X linked gene highly expressed in brain in a family with two mentally retarded males. J Med Genet 2004;41:736–742.
11. Pear WS, Nolan GP, Scott ML, Baltimore D. Production of high-titer helper-free retroviruses by transient transfection. Proc Natl Acad Sci USA 1993;90:8392–8396.
12. Kisanuki H, Choi YL, Wada T, Moriuchi R, Fujiwara SI, Kaneda R, et al. Retroviral expression screening of oncogenes in pancreatic ductal carcinoma. Eur J Cancer 2005;41:2170–2175.
13. Fujiwara S, Yamashita Y, Choi YL, Wada T, Kaneda R, Takada S, et al. Transforming activity of the lymphotoxin-beta receptor revealed by expression screening. Biochem Biophys Res Commun 2005;338:1256–1262.
14. Choi YL, Moriuchi R, Osawa M, Iwama A, Makishima H, Wada T, et al. Retroviral expression screening of oncogenes in natural killer cell leukemia. Leuk Res 2005;29:943–949.
15. Hu Q, Milfay D, Williams LT. Binding of NCK to SOS and activation of *ras*-dependent gene expression. Mol Cell Biol 1995;15:1169–1174.
16. Takeshita T, Arita T, Higuchi M, Asao H, Endo K, Kuroda H, 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;6:449–457.
17. Grillot DAM, Gonzalez-Garcia M, Ekhterae D, Duan L, Inohara N, Ohta S, et al. Genomic organization, promoter region analysis, and chromosome localization of the mouse *bcl-x* gene. J Immunol 1997;158:4750–4757.
18. Shih TY, Weeks MO. Oncogenes and cancer: the p21 ras genes. Cancer Invest 1984;2:109–123.
19. Muscella A, Greco S, Elia MG, Storelli C, Marsigliante S. Differential signalling of purinoceptors in HeLa cells through the extracellular signal-regulated kinase and protein kinase C pathways. J Cell Physiol 2004;200:428–439.
20. Van Kolen K, Slegers H. Atypical PKCzeta is involved in RhoA-dependent mitogenic signaling by the P2Y(12) receptor in C6 cells. FEBS J 2006;273:1843–1854.
21. Wagstaff SC, Bowler WB, Gallagher JA, Hipskind RA. Extracellular ATP activates multiple signalling pathways and potentiates growth factor-induced c-fos gene expression in MCF-7 breast cancer cells. Carcinogenesis 2000;21:2175–2181.
22. Adrian K, Bernhard MK, Breitinger HG, Ogilvie A. Expression of purinergic receptors (ionotropic P2X1-7 and metabotropic P2Y1-11) during myeloid differentiation of HL60 cells. Biochim Biophys Acta 2000;1492:127–138.
23. Bagchi S, Liao Z, Gonzalez FA, Chorna NE, Seye CI, Weisman GA, et al. The P2Y2 nucleotide receptor interacts with alphav integrins to activate Go and induce cell migration. J Biol Chem 2005;280:39050–39057.
24. Sellers LA, Simon J, Lundahl TS, Cousens DJ, Humphrey PP, Barnard EA. Adenosine nucleotides acting at the human P2Y1 receptor stimulate mitogen-activated protein kinases and induce apoptosis. J Biol Chem 2001;276:16379–16390.

25. Bogdanov YD, Dale L, King BF, Whittock N, Burnstock G. Early expression of a novel nucleotide receptor in the neural plate of Xenopus embryos. J Biol Chem 1997;272:12583–12590.
26. Conigrave AD, van der Weyden L, Holt L, Jiang L, Wilson P, Christopherson RI, et al. Extracellular ATP-dependent suppression of proliferation and induction of differentiation of human HL-60 leukemia cells by distinct mechanisms. Biochem Pharmacol 2000;60:1585–1591.
27. Zhang XJ, Zheng GG, Ma XT, Yang YH, Li G, Rao Q, et al. Expression of P2X7 in human hematopoietic cell lines and leukemia patients. Leukemia Res 2004;28:1313–1322.
28. Greenberger JS, Sakakeeny MA, Humphries RK, Eaves CJ, Eckner RJ. Demonstration of permanent factor-dependent multipotential (erythroid/neutrophil/basophil) hematopoietic progenitor cell lines. Proc Natl Acad Sci USA 1983;80:2931–2935.
29. De Cesare D, Jacquot S, Hanauer A, Sassone-Corsi P. Rsk-2 activity is necessary for epidermal growth factor-induced phosphorylation of CREB protein and transcription of c-fos gene. Proc Natl Acad Sci USA 1998;95:12202–12207.
30. Janknecht R, Nordheim A. MAP kinase-dependent transcriptional coactivation by Elk-1 and its cofactor CBP. Biochem Biophys Res Commun 1996;228:831–837.

ARTICLES

Identification of the transforming *EML4–ALK* fusion gene in non-small-cell lung cancer

Manabu Soda[1,2], Young Lim Choi[1], Munehiro Enomoto[1,2], Shuji Takada[1], Yoshihiro Yamashita[1], Shunpei Ishikawa[5], Shin-ichiro Fujiwara[1], Hideki Watanabe[1], Kentaro Kurashina[1], Hisashi Hatanaka[1], Masashi Bando[2], Shoji Ohno[2], Yuichi Ishikawa[6], Hiroyuki Aburatani[5,7], Toshiro Niki[3], Yasunori Sohara[4], Yukihiko Sugiyama[2] & Hiroyuki Mano[1,7]

Improvement in the clinical outcome of lung cancer is likely to be achieved by identification of the molecular events that underlie its pathogenesis. Here we show that a small inversion within chromosome 2p results in the formation of a fusion gene comprising portions of the echinoderm microtubule-associated protein-like 4 (*EML4*) gene and the anaplastic lymphoma kinase (*ALK*) gene in non-small-cell lung cancer (NSCLC) cells. Mouse 3T3 fibroblasts forced to express this human fusion tyrosine kinase generated transformed foci in culture and subcutaneous tumours in nude mice. The *EML4–ALK* fusion transcript was detected in 6.7% (5 out of 75) of NSCLC patients examined; these individuals were distinct from those harbouring mutations in the epidermal growth factor receptor gene. Our data demonstrate that a subset of NSCLC patients may express a transforming fusion kinase that is a promising candidate for a therapeutic target as well as for a diagnostic molecular marker in NSCLC.

Lung cancer remains the leading cause of cancer deaths in western countries[1]. Patients with NSCLC, which accounts for ~80% of lung cancer cases, are often diagnosed at advanced stages of the disease. Given that conventional chemotherapeutic regimens only marginally improve the outcome of such individuals, their median survival time is less than one year after diagnosis (ref. 2). A subset of NSCLCs was recently shown to harbour activating mutations in the epidermal growth factor receptor gene (*EGFR*)[3,4]; such cancers are responsive to gefitinib, a specific inhibitor of the tyrosine kinase activity of EGFR. The efficacy of targeting key 'growth drivers' in cancer treatment is further exemplified by chronic myeloid leukaemia, for which another tyrosine kinase inhibitor, STI571, is highly effective in reducing the number of cancer cells[5]. However, *EGFR* mutations are associated preferentially with NSCLC of non-smokers and Asians[4,6]. Few oncogenes have thus been identified for NSCLC in individuals with a smoking habit, who constitute most cases of the disease.

Retrovirus-mediated complementary DNA expression systems allow expression of the encoded proteins in most of the targeted cells. Through modification of the method used in ref. 7, we have achieved reliable amplification of cDNAs from small quantities of clinical specimens as well as the generation of retroviral libraries for expression of these cDNAs[8–10]. Application of such a cDNA expression library prepared from an NSCLC specimen to a focus formation assay with mouse 3T3 fibroblasts has now led to the identification of a fusion oncogene.

Identification of *EML4–ALK*

To isolate novel transforming genes in NSCLC, we generated a retroviral cDNA expression library from a lung adenocarcinoma specimen surgically resected from a 62-yr-old man with a history of smoking (patient 33). In construction of the library, we used the SMART method (Clontech) for preferential amplification of full-length cDNAs from limited amounts of clinical specimens[7]; this resulted in the production of $>1.4 \times 10^6$ independent plasmid clones. Infection of mouse 3T3 fibroblasts with the recombinant retroviruses that were based on these plasmids led to the formation of many transformed foci, from which insert cDNAs were recovered with the polymerase chain reaction (PCR).

One of the amplified cDNAs comprised 3,926 base pairs (bp) and contained an open reading frame for a protein of 1,059 amino acids (Fig. 1a and Supplementary Fig. 1). The amino-terminal portion (residues 1–496) of the predicted protein is identical to that of human EML4 (GenBank accession number NM_019063), whereas the carboxy-terminal portion (residues 497–1059) is identical to the intracellular domain (residues 1058–1620 of the wild-type protein) of human ALK (GenBank accession number AB209477), suggesting that the cDNA is derived from a fusion product of *EML4* and *ALK* (Fig. 1a, b). EML4 belongs to the family of echinoderm microtubule-associated protein-like proteins[11] and is composed of an N-terminal basic region (isoelectric point, 10.2), a hydrophobic echinoderm microtubule-associated protein-like protein (HELP) domain[12] and WD repeats[13] (Fig. 1b). In the predicted fusion protein, the N-terminal half of EML4 encompassing the basic region, the HELP domain and a portion of the WD-repeat region is fused to the intracellular juxtamembrane region of ALK.

ALK was first identified as a fusion partner of nucleophosmin (NPM) in anaplastic large-cell lymphoma with a t(2;5) chromosome rearrangement[14,15]. Other chromosome translocations involving the *ALK* locus were subsequently identified in the same lymphoma subtype as well as in inflammatory myofibroblastic tumours[16]. The fusion point of ALK is conserved among most of these chimaeric tyrosine kinases, including EML4–ALK (resulting in fusion of the

[1]Division of Functional Genomics, [2]Division of Pulmonary Medicine, [3]Department of Pathology, and [4]Division of General Thoracic Surgery, Jichi Medical University, Tochigi 329-0498, Japan. [5]Research Center for Advanced Science and Technology, University of Tokyo, Tokyo 153-8904, Japan. [6]Department of Pathology, The Cancer Institute, Japanese Foundation for Cancer Research, Tokyo 135-8550, Japan. [7]Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, Saitama 332-0012, Japan.

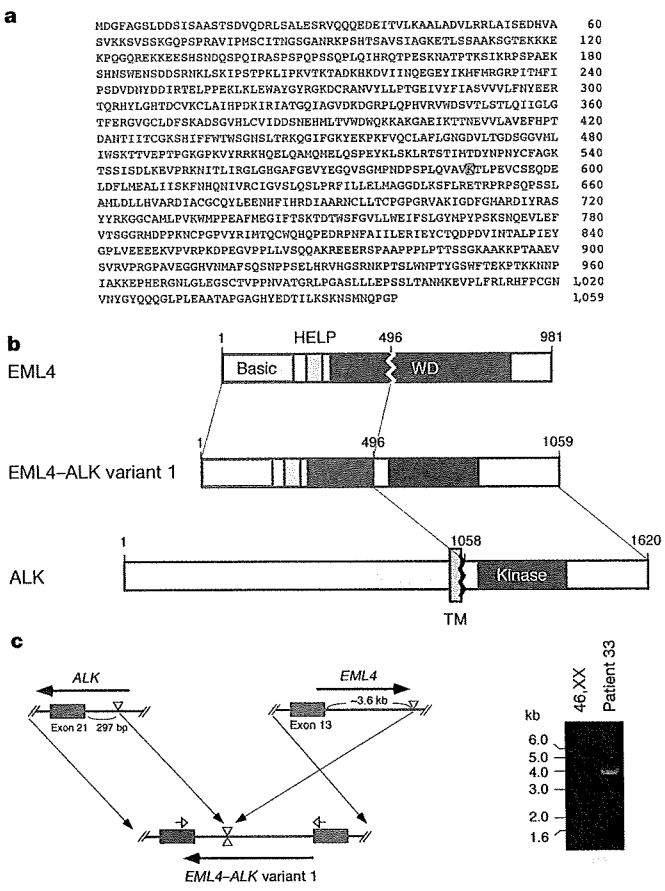

Figure 1 | Gene fusion between *EML4* and *ALK*. **a**, Amino acid sequence of the EML4–ALK protein (variant 1). Residues corresponding to EML4 or to ALK are shown in blue or red, respectively. Lys 589 in the ATP-binding site is boxed. **b**, Fusion of the N-terminal portion of EML4 (comprising the basic region, the HELP domain and part of the WD-repeat region) to the intracellular region of ALK (containing the tyrosine kinase domain). TM, transmembrane domain. **c**, Both the *ALK* gene and the *EML4* gene map to chromosome 2p, but have opposite orientations. In the NSCLC patient 33, *EML4* is disrupted at a position ~3.6 kb downstream of exon 13 and is ligated to a position 297 bp upstream of exon 21 of *ALK*, giving rise to the *EML4–ALK* (variant 1) fusion gene (left panel). Filled and open horizontal arrows indicate the direction of transcription and the positions of the Fusion-genome primers, respectively (Supplementary Fig. 1). PCR with these Fusion-genome primers and genomic DNA of patient 33 generated a single product of ~4 kb (right panel); this product was not detected with control DNA of a healthy female (46,XX).

entire intracellular kinase domain of ALK to the corresponding partner), and the kinase activity of NPM–ALK was shown to be essential for the proliferation of lymphoma cells positive for this construct[17].

Given that *EML4* and *ALK* each map to the short arm of chromosome 2 (2p21 and 2p23, respectively, separated by a distance of ~12 megabases, Mb) but have opposite orientations, either gene might have been inverted to generate the *EML4–ALK* fusion gene (Fig. 1c). To address this issue directly, we amplified the genomic fusion point between *EML4* and *ALK* using genomic DNA of patient 33 as the template. This approach led to the identification of a ~4-kilobase (kb) product (Fig. 1c). Nucleotide sequencing of this genomic fragment revealed that intron 13 of *EML4* is disrupted at a point ~3.6 kb downstream of exon 13 and is inverted to connect to a position 297 bp upstream of exon 21 of *ALK* (Fig. 1c and Supplementary Data), yielding *EML4–ALK* variant 1 (the structure of variant 2 is addressed below). To determine whether the chromosome rearrangement in this specimen is a simple inversion within 2p, we attempted to detect the other connection point between *EML4* and *ALK* in chromosome 2 by PCR amplification of the *ALK–EML4* cDNA or gene. However, neither of the corresponding PCR products was obtained (data not shown). It thus remains undetermined whether the NSCLC cells of patient 33 harbour a simple inv(2)(p21p23) or whether they contain complex chromosome translocations involving 2p.

Transforming activity of EML4–ALK

To confirm the transforming potential of EML4–ALK, we generated expression plasmids for wild-type EML4, wild-type ALK, EML4–ALK, EML4–ALK(K589M) (in which Lys 589 in the ATP-binding site of the kinase domain is replaced with Met), NPM–ALK and v-Ras, and introduced them individually into mouse 3T3 fibroblasts. Transformed foci were readily identified for the cells expressing EML4–ALK, NPM–ALK or v-Ras, but not for those expressing EML4, ALK or EML4–ALK(K589M) (Fig. 2a). Subcutaneous injection of the transfected 3T3 cells into nude mice also revealed that only those expressing EML4–ALK, NPM–ALK or v-Ras formed tumours (Fig. 2a). These data thus showed that EML4–ALK possesses transforming activity that is dependent on its catalytic activity. We also found that fusion to EML4 results in redistribution of the kinase domain of ALK from the cell membrane to the cytoplasm (Supplementary Fig. 2), as revealed by monitoring the fluorescence of the corresponding proteins tagged with enhanced green fluorescent protein.

To identify the domains of EML4 required for the transforming activity of the EML4–ALK fusion protein, we generated expression plasmids for EML4–ALK with internal deletions of the basic domain (ΔBasic, lacking residues 31–140), of the HELP domain (ΔHELP, lacking residues 220–296) or of the WD repeats (ΔWD, lacking residues 305–475) (Fig. 2b). Injection of 3T3 cells expressing the deletion constructs into nude mice revealed that deletion of the WD repeats allowed tumour formation at all injection sites, but that the tumours were smaller than those formed by cells expressing full-length EML4–ALK (Fig. 2b). Tumours were even smaller for cells expressing ΔHELP and were undetectable for those expressing ΔBasic. All domains of EML4 thus seem to contribute to the oncogenic potential of EML4–ALK, with the basic domain being the most important.

Gene fusion often results in activation of tyrosine kinases through oligomerization mediated by the fusion partner. Although little is known regarding the dimerization potential of EML4, our results (Fig. 2b) suggested that the basic domain derived from EML4 may mediate EML4–ALK dimerization. To examine this possibility, we transfected HEK 293 cells with expression vectors for both Myc-epitope-tagged EML4–ALK and Flag-epitope-tagged full-length, ΔBasic, ΔHELP or ΔWD forms of EML4–ALK. Immunoprecipitation of cell lysates with antibodies to Myc and probing of the resulting precipitates with antibodies to Flag revealed that Myc-epitope-tagged EML4–ALK was associated with substantial amounts of each of the Flag-tagged EML4–ALK constructs with the exception of ΔBasic (Fig. 2c). Immunoblot analysis of immunoprecipitates prepared from the same cell lysates with antibodies to Flag confirmed that the various Flag-tagged EML4–ALK constructs were expressed at similar levels. Similar results were obtained in a reciprocal experiment in which anti-Flag immunoprecipitates were probed with anti-Myc (Supplementary Fig. 3). These results thus indicated that the basic domain indeed has an important role in dimerization of EML4–ALK.

To examine directly whether internal deletions affect the enzymatic activity of EML4–ALK, we expressed Flag-tagged full-length or truncated forms of the fusion protein in HEK 293 cells and prepared immunoprecipitates from cell lysates with anti-Flag. The resulting precipitates were then subjected to an *in vitro* kinase assay with the synthetic YFF peptide[18], which is based on the activation loop of the catalytic domain of ALK. Deletion of the basic domain resulted in a marked decrease (~84%) in the catalytic activity of EML4–ALK (Fig. 2d). This low level of kinase activity of the ΔBasic mutant was consistent with its residual ability to dimerize with the full-length

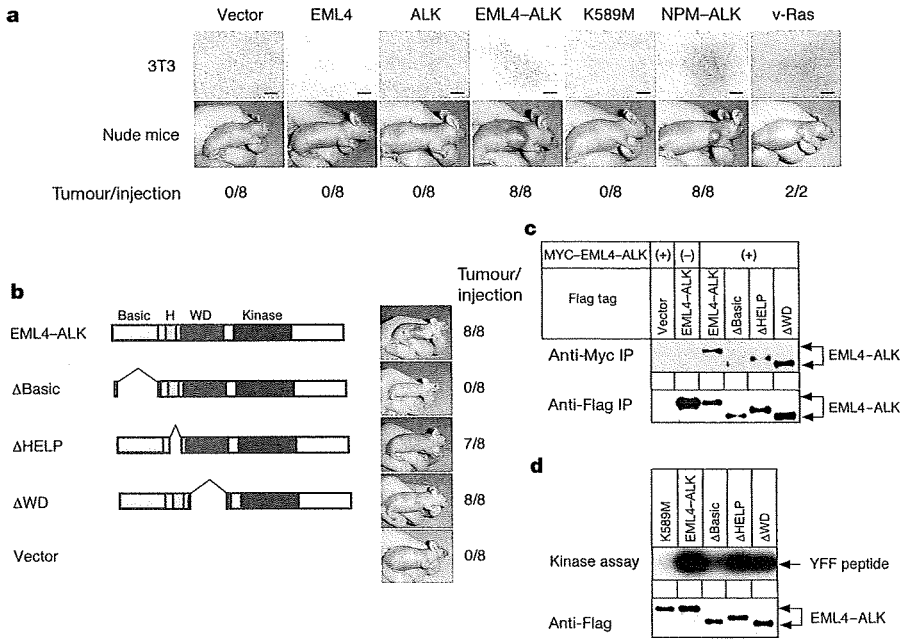

Figure 2 | Transforming activity of EML4–ALK variant 1. **a**, Expression vectors for EML4, ALK, EML4–ALK, EML4–ALK(K589M), NPM–ALK and v-Ras (or the corresponding empty vector) were introduced individually into 3T3 cells, and the cells were photographed after three weeks of culture (upper panels). Scale bars, 100 μm. The same set of transfected cells was also injected subcutaneously into nude mice, and tumour formation was examined after 20 days (lower panels). The number of tumours formed after eight or two injections is indicated. **b**, Schematic representations of EML4–ALK and its deletion mutants are shown on the left. Tumour formation in nude mice was examined as in **a** for 3T3 cells transfected with expression vectors for the indicated forms of EML4–ALK (right). H, HELP domain. **c**, Expression vectors for Flag-tagged EML4–ALK or its deletion mutants were introduced into HEK 293 cells with (+) or without (–) a vector for Myc-epitope-tagged EML4–ALK. Cell lysates were subjected to immunoprecipitation (IP) with antibodies to Myc or to Flag, and the resulting precipitates were subjected to immunoblot analysis with anti-Flag. The positions of EML4–ALK and its mutants are shown on the right. **d**, Expression vectors for Flag-tagged EML4–ALK, EML4–ALK(K589M) or its deletion mutants were introduced into HEK 293 cells. Immunoprecipitates prepared from cell lysates with anti-Flag were subjected to an *in vitro* kinase assay with the synthetic YFF peptide (upper panel) or to immunoblot analysis with anti-Flag (lower panel).

protein (Fig. 2c and Supplementary Fig. 3). Deletion of either the HELP or WD domain reduced the kinase activity of EML4–ALK by ~50% (Fig. 2d), whereas the transforming activity of the ΔHELP mutant was reproducibly lower than that of ΔWD in both the focus formation (data not shown) and tumorigenicity (Fig. 2b) assays. The molecular basis of this discrepancy between kinase and transforming activities remains to be determined.

Detection of *EML4–ALK* in clinical specimens

We next evaluated the frequency of *EML4–ALK* gene fusion in NSCLC. A consecutive panel of NSCLC specimens ($n = 33$) obtained in one hospital was examined for the presence of *EML4–ALK* messenger RNA, wild-type *ALK* mRNA and mutations within the *EGFR* and v-Ki-ras2 Kirsten rat sarcoma viral oncogene homologue (*KRAS*) genes. The *EML4–ALK* fusion mRNA was readily detected by PCR with reverse transcription (RT–PCR) analysis as a 247-bp product in three patients (9.1%), including the patient who served as the source for the retroviral library (Fig. 3a). Sequencing of the PCR products amplified from each of these three patients confirmed the presence of *EML4–ALK* variant 1 cDNA (Supplementary Data); patients 20 and 39 had squamous cell carcinoma and adenocarcinoma of the lung, respectively.

We also amplified the genomic fragments corresponding to exons 18, 19 and 21 of *EGFR* and determined their nucleotide sequences in all 33 patients. This revealed the presence of *EGFR* mutations—all of which were deletions or nucleotide substitutions within exon 19—in six individuals (18.2%) (Supplementary Table 1). Notably, the patient population harbouring *EGFR* mutations did not overlap with that harbouring the *EML4–ALK* fusion gene, showing that *EML4–ALK*-positive cancer is a novel subclass within NSCLC. A *KRAS* mutation (Val 12 to Cys 12 substitution) was detected in two individuals, neither of whom harboured *EGFR* mutations or the *EML4–ALK* fusion gene. Wild-type *ALK* mRNA was detected in 8 of the 33 specimens (24%) of this cohort. A moderate level of *ALK* mRNA in lung cancer specimens has also been detected by serial analysis of gene expression studies (http://cgap.nci.nih.gov/SAGE/AnatomicViewer), although it is not clear whether such profiling of the 3′ end of mRNAs actually detected mRNAs for wild-type *ALK*, *EML4–ALK* or even other *ALK* fusion genes.

To determine whether fusion of *EML4* to *ALK* is specific to NSCLC, we used RT–PCR to attempt to detect the fusion mRNA in cancer specimens from 39 patients with acute myeloid leukaemia, 69 patients with non-Hodgkin's lymphoma, 93 patients with gastric carcinoma and 60 patients with colorectal carcinoma. However, none of these 261 specimens yielded the *EML4–ALK* cDNA (data not shown), indicating that *EML4–ALK* has a high level of specificity to NSCLC.

Further screening for the *EML4–ALK* fusion cDNA with the same RT–PCR primer set in a different cohort of NSCLC patients ($n = 42$) resulted in the identification of a larger PCR product (~1 kb) in another two individuals with lung adenocarcinoma. Nucleotide sequencing of these PCR products revealed that exon 20 of *EML4* was fused to exon 21 of *ALK* (Supplementary Fig. 4 and Supplementary Data), indicative of diversity in the breakpoint region within *EML4*. PCR analysis of genomic DNA from one of these two patients further revealed the breakpoints in *EML4* and *ALK* as well as the formation of both *EML4–ALK* and *ALK–EML4* fusion genes, thus demonstrating the presence of an inv(2)(p21p23) rearrangement in this individual (Supplementary Fig. 4 and Supplementary Data). We here refer to the initially identified *EML4–ALK* gene, in which intron 13 of *EML4* is fused to intron 20 of *ALK*, as variant 1, and to the *EML4–ALK* gene, in which intron 20 of *EML4* is fused to intron 20 of

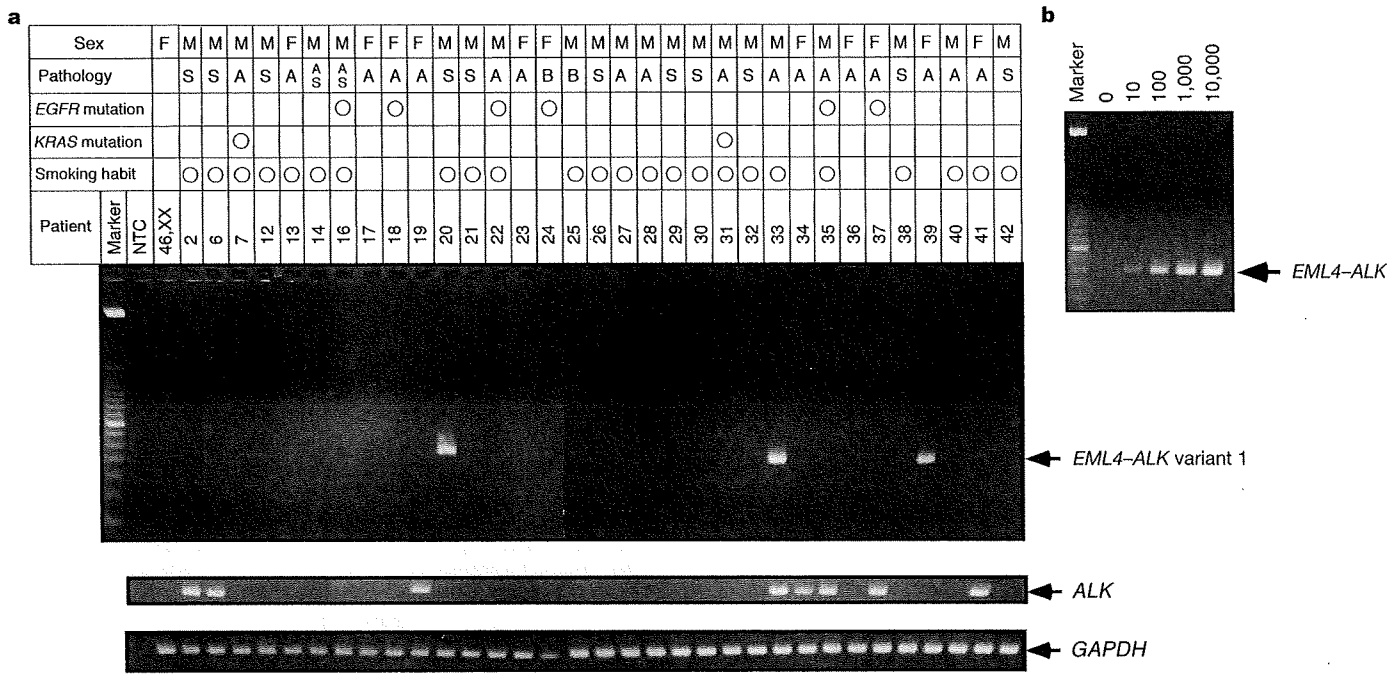

Figure 3 | Screening of NSCLC specimens for *EML4–ALK* variant 1 mRNA. **a**, A consecutive panel of NSCLC specimens ($n = 33$) was subjected to RT–PCR with the Fusion-RT primers (Supplementary Fig. 1). Patients positive for the *EML4–ALK* (variant 1) PCR product are shown in red. Peripheral blood mononuclear cells from a healthy female (46,XX) were also examined as a negative control. RT–PCR for wild-type *ALK* mRNA (with a primer set corresponding to the extracellular domain of ALK) and for glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) mRNA is also shown. Patient characteristics (sex, pathological classification of NSCLC, the presence of *EGFR* or *KRAS* mutations, and smoking habit) are indicated at the top. A, adenocarcinoma; AS, adenosquamous carcinoma; B, bronchioloalveolar carcinoma; F, female; M, male; NTC, no-template control; S, squamous carcinoma. Marker, 50-bp DNA ladder. **b**, Sputum (1 ml) was mixed with 0, 10, 100, 1,000 or 10,000 BA/F3 cells expressing *EML4–ALK* (variant 1) and was then subjected to RT–PCR with the Fusion-RT primer set for detection of *EML4–ALK* mRNA. Marker, 50-bp DNA ladder.

ALK, as variant 2. Multiple variants of the TRK-fused gene (*TFG*)–*ALK* fusion gene associated with anaplastic large-cell lymphoma have also been identified[19].

Given the head-to-head orientation of *EML4* and *ALK* on chromosome 2, RT–PCR with the primers used in our study would not be expected to yield specific products in normal tissues or in any cancers that do not harbour the fusion gene. RT–PCR for *EML4–ALK* mRNA may thus provide a highly sensitive means for detection of lung cancer (with the corresponding chromosomal rearrangement). Although cytological examination of sputum is a reliable method for detection of lung cancer, it is usually only effective at advanced stages. Detection of *EGFR* mutations in sputum is also problematic because NSCLC cells often constitute only a small fraction of cells within the specimen. In contrast, RT–PCR would be expected to detect the presence of only a few cells harbouring the *EML4–ALK* fusion gene among the tens of thousands of non-cancerous cells in sputum. To examine this issue, we established mouse BA/F3 cells[20] that express Flag-tagged EML4–ALK, mixed various numbers of these cells with control sputum, and subjected the mixtures to RT–PCR for detection of *EML4–ALK* mRNA. The fusion mRNA was detected in sputum containing as few as ten BA/F3 cells per ml (Fig. 3b). Early diagnosis of NSCLC (of the *EML4–ALK*$^+$ subtype) by RT–PCR may thus be realistic, as is clinical detection of *Mycobacterium tuberculosis* in sputum by PCR[21]. It is also possible that *EML4–ALK* mRNA would be detected by RT–PCR using pleural effusion, bronchoalveolar lavage, lung biopsy or peripheral blood specimens of patients with lung cancer.

EML4–ALK as a potential therapeutic target

Several small compounds have recently been shown to inhibit the kinase activity of ALK and to suppress the growth of cells expressing NPM–ALK[17,22,23]. To investigate whether use of such ALK inhibitors might be an effective treatment for EML4–ALK$^+$ NSCLC, we expressed Flag-tagged ALK, EML4–ALK or EML4–ALK(K589M) in BA/F3 cells, which are dependent on interleukin-3 (IL-3) for growth. Whereas all transfected cells grew exponentially in the presence of IL-3, only those expressing EML4–ALK proliferated at a similar rate in the absence of IL-3 (Fig. 4a), again confirming the kinase-dependent oncogenic activity of EML4–ALK.

In the presence of IL-3, BA/F3 cells proliferate in a manner dependent on the activity of the tyrosine kinase Janus kinase 2 (JAK2, ref. 24). Addition of a chemical inhibitor (WHI-P154)[22] of ALK to the culture medium affected the IL-3-dependent growth of BA/F3 cells only slightly at concentrations $\geq 5\,\mu M$ (Fig. 4b). Given that WHI-P154 was originally developed as a specific inhibitor of JAK3, it might be expected to show a weak cross-reactivity with JAK2, possibly accounting for the small effect on the JAK2-dependent growth of BA/F3 cells. In contrast, WHI-P154 markedly inhibited the growth of BA/F3 cells expressing EML4–ALK, which do not require IL-3 for proliferation (Fig. 4b). At a concentration of $10\,\mu M$, WHI-P154 rapidly induced the death of these cells in the absence of IL-3. Consistent with these observations, immunoblot analysis revealed that WHI-P154 inhibited the tyrosine phosphorylation of EML4–ALK (on the residue corresponding to Tyr 1604 of wild-type ALK) in a concentration-dependent manner in the transfected BA/F3 cells (Fig. 4c).

Discussion

Using retrovirus-mediated expression screening, we have identified an oncogene, *EML4–ALK*, in a specimen of NSCLC. The 75 NSCLC patients examined in the present study (5 of whom were positive for *EML4–ALK*) were all Japanese. Given that the association of *EGFR* mutations with lung cancer is most prominent in Asian populations[5], it will be important to examine the association of *EML4–ALK* with lung cancer in other ethnic groups. Our data obtained with WHI-P154 suggest that inhibition of the tyrosine kinase activity of

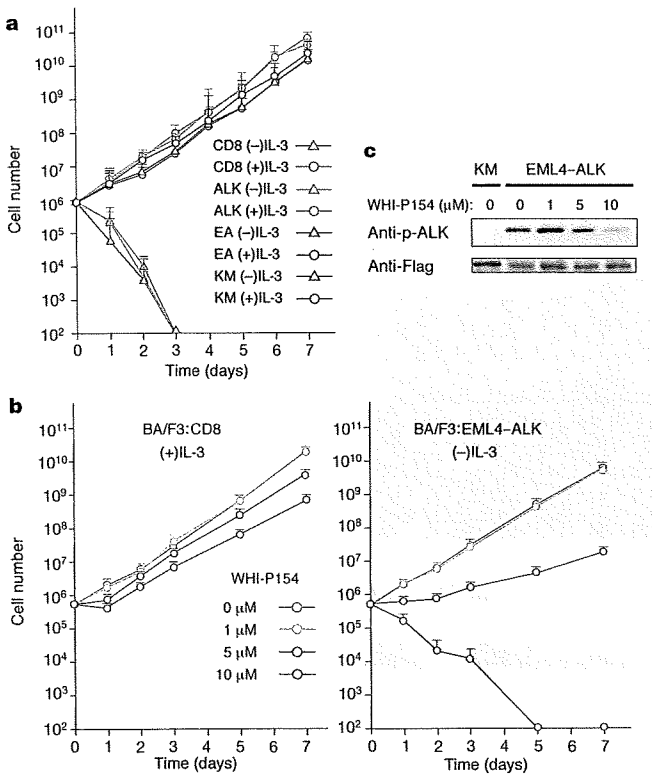

Figure 4 | Inhibition of the growth of BA/F3 cells expressing EML4–ALK variant 1 by a chemical inhibitor of ALK. a, Mouse BA/F3 cells expressing CD8 either alone or together with wild-type ALK, EML4–ALK (EA) or EML4–ALK(K589M) (KM) were cultured in the absence (–) or presence (+) of IL-3 (1 ng ml^{-1}). Cell number was determined at the indicated times. Data are means plus s.d. of values from three separate experiments. **b**, BA/F3 cells expressing CD8 alone were cultured with IL-3 and 0, 1, 5 or 10 μM WHI-P154 (left panel), or those expressing CD8 and EML4–ALK were incubated without IL-3 but with 0, 1, 5 or 10 μM WHI-P154 (right panel). Cell number was determined at the indicated times. Data are means plus s.d. of values from three separate experiments. **c**, BA/F3 cells expressing Flag-tagged EML4–ALK or EML4–ALK(K589M) were incubated with the indicated concentrations of WHI-P154 for 3.5 h, after which total cell lysates (25 μg of protein per lane) were subjected to immunoblot analysis with antibodies to tyrosine-phosphorylated ALK (p-ALK) or to Flag.

EML4–ALK may induce cell death in tumours expressing this fusion protein. Given the lack of apparent phenotypes in *Alk* knockout mice[25], suppression of EML4–ALK function with ALK inhibitors might be expected to be free of severe side effects in NSCLC patients. Furthermore, given that the population of patients who harbour *EGFR* mutations is distinct from that which harbours the *EML4–ALK* fusion gene, ALK inhibitors may provide a means to control NSCLC in the latter population of patients, for whom effective treatments are rarely available.

METHODS SUMMARY

A recombinant retroviral cDNA expression library was constructed as described previously[7–10] from a lung cancer specimen, and was used to infect mouse 3T3 fibroblasts. Transformed foci isolated from the cells after 2 weeks of culture were subjected to extraction of genomic DNA and amplification of retroviral insert cDNAs by PCR. The *EML4–ALK* fusion cDNA was detected by RT–PCR analysis of total RNA from clinical specimens with the Fusion-RT-S (5′-GTGCAGTGTTTAGCATTCTTGGGG-3′) and Fusion-RT-AS (5′-TCTTGCCAGCAAAGCAGTAGTTGG-3′) primers. The *EML4–ALK* variant 1 gene was detected by PCR with genomic DNA of clinical specimens and the Fusion-genome-S (5′-CCACACCTGGGAAAGGACCTAAAG-3′) and Fusion-genome-AS (5′-AGCTTGCTCAGCTTGTACTCAGGG-3′) primers. Expression plasmids for Flag-epitope-tagged ALK, EML4 and EML4–ALK were generated with the retroviral vector pMXS[26]. The kinase-inactive mutant (K589M) of EML4–ALK was constructed by site-directed mutagenesis. Internal deletion mutants of EML4–ALK were also generated by mutagenesis. The expression plasmids were introduced into 3T3 cells by the calcium phosphate method, and the cells were then either cultured for 21 days or injected subcutaneously into nude mice. For analysis of EML4–ALK dimerization, expression vectors for Flag- or Myc-epitope-tagged EML4–ALK or its mutants were introduced into HEK 293 cells, cell lysates were subjected to immunoprecipitation with anti-Flag or anti-Myc, and the resulting precipitates were subjected to immunoblot analysis with the same antibodies. Anti-Flag immunoprecipitates were also subjected to an *in vitro* kinase assay with the synthetic YFF peptide[18]. For BA/F3 experiments, cDNAs for ALK, EML4–ALK or EML4–ALK(K589M) were inserted into the plasmid pMX-iresCD8 (ref. 27) to confer simultaneous expression of the protein of interest and mouse CD8. BA/F3 cells were infected with recombinant retroviruses generated from each plasmid, and the resulting CD8$^+$ cells were purified and incubated with various concentrations of WHI-P154 (EMD Biosciences).

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

Received 15 February; accepted 17 May 2007.
Published online 11 July 2007.

1. Jemal, A. *et al.* Cancer statistics, 2006. *CA Cancer J. Clin.* 56, 106–130 (2006).
2. Schiller, J. H. *et al.* Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N. Engl. J. Med.* 346, 92–98 (2002).
3. Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* 350, 2129–2139 (2004).
4. Paez, J. G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 304, 1497–1500 (2004).
5. Druker, B. J. *et al.* Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.* 344, 1031–1037 (2001).
6. Pao, W. *et al.* EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl Acad. Sci. USA* 101, 13306–13311 (2004).
7. Yoshizuka, N. *et al.* An alternative transcript derived from the *Trio* locus encodes a guanosine nucleotide exchange factor with mouse cell-transforming potential. *J. Biol. Chem.* 279, 43998–44004 (2004).
8. Hatanaka, H. *et al.* Transforming activity of purinergic receptor P2Y, G-protein coupled, 2 revealed by retroviral expression screening. *Biochem. Biophys. Res. Commun.* 356, 723–726 (2007).
9. Choi, Y. L. *et al.* Identification of a constitutively active mutant of JAK3 by retroviral expression screening. *Leuk. Res.* 31, 203–209 (2007).
10. Fujiwara, S. *et al.* Transforming activity of the lymphotoxin-β receptor revealed by expression screening. *Biochem. Biophys. Res. Commun.* 338, 1256–1262 (2005).
11. Pollmann, M. *et al.* Human EML4, a novel member of the EMAP family, is essential for microtubule formation. *Exp. Cell Res.* 312, 3241–3251 (2006).
12. Eichenmuller, B., Everley, P., Palange, J., Lepley, D. & Suprenant, K. A. The human EMAP-like protein-70 (ELP70) is a microtubule destabilizer that localizes to the mitotic apparatus. *J. Biol. Chem.* 277, 1301–1309 (2002).
13. Smith, T. F., Gaitatzes, C., Saxena, K. & Neer, E. J. The WD repeat: a common architecture for diverse functions. *Trends Biochem. Sci.* 24, 181–185 (1999).
14. Morris, S. W. *et al.* Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* 263, 1281–1284 (1994).
15. Shiota, M. *et al.* Hyperphosphorylation of a novel 80 kDa protein-tyrosine kinase similar to Ltk in a human Ki-1 lymphoma cell line, AMS3. *Oncogene* 9, 1567–1574 (1994).
16. Pulford, K., Morris, S. W. & Turturro, F. Anaplastic lymphoma kinase proteins in growth control and cancer. *J. Cell. Physiol.* 199, 330–358 (2004).
17. Galkin, A. V. *et al.* Identification of NVP-TAE684, a potent, selective, and efficacious inhibitor of NPM-ALK. *Proc. Natl Acad. Sci. USA* 104, 270–275 (2007).
18. Donella-Deana, A. *et al.* Unique substrate specificity of anaplastic lymphoma kinase (ALK): development of phosphoacceptor peptides for the assay of ALK activity. *Biochemistry* 44, 8533–8542 (2005).
19. Hernandez, L. *et al.* Diversity of genomic breakpoints in *TFG-ALK* translocations in anaplastic large cell lymphomas: identification of a new TFG-ALK$_{XL}$ chimeric gene with transforming activity. *Am. J. Pathol.* 160, 1487–1494 (2002).
20. 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* 41, 727–734 (1985).
21. Kaul, K. L. Molecular detection of *Mycobacterium tuberculosis*: impact on patient care. *Clin. Chem.* 47, 1553–1558 (2001).
22. Marzec, M. *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.* 85, 1544–1554 (2005).
23. Li, R. *et al.* Design and synthesis of 5-aryl-pyridone-carboxamides as inhibitors of anaplastic lymphoma kinase. *J. Med. Chem.* 49, 1006–1015 (2006).
24. Watanabe, S., Itoh, T. & Arai, K. JAK2 is essential for activation of c-*fos* and c-*myc* promoters and cell proliferation through the human granulocyte-macrophage colony-stimulating factor receptor in BA/F3 cells. *J. Biol. Chem.* 271, 12681–12686 (1996).

25. Duyster, J., Bai, R. Y. & Morris, S. W. Translocations involving anaplastic lymphoma kinase (ALK). *Oncogene* **20**, 5623–5637 (2001).
26. Onishi, M. *et al.* Applications of retrovirus-mediated expression cloning. *Exp. Hematol.* **24**, 324–329 (1996).
27. Yamashita, Y. *et al.* Sak serine/threonine kinase acts as an effector of Tec tyrosine kinase. *J. Biol. Chem.* **276**, 39012–39020 (2001).

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

Acknowledgements We thank R. Moriuchi for suggestions.

Author Contributions M.S. and Y.L.C. contributed equally to this work. M.S., S.-i.F., H.W. and H.H. constructed the cDNA library and screened for transforming genes. Y.L.C. sequenced the *EML4–ALK* cDNA and conducted the experiments with BA/F3 cells. Y.Y. and S.T. searched for *EGFR* and *KRAS* mutations. M.E., S.I., K.K., M.B., S.O., S.T., Y.I. and H.A. performed RT–PCR for *EML4–ALK* transcripts in cancer specimens. T.N., Y. Sohara, Y. Sugiyama and H.M. designed the overall project, and H.M. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information The nucleotide sequences of *EML4–ALK* variant 1 and variant 2 cDNA have been deposited in DDBJ, EMBL and GenBank under the accession numbers AB274722 and AB275889, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.M. (hmano@jichi.ac.jp).

Analysis of chromosome copy number in leukemic cells by different microarray platforms

Leukemia (2007) **21**, 1333–1337. doi:10.1038/sj.leu.2404636; published online 15 March 2007

Changes in the copy number of chromosomes, or copy number alterations (CNAs), are frequently apparent in the genome of cancer cells and may result in the amplification of oncogenes or deletion of tumor suppressor genes.[1] Such CNAs range in size from an entire chromosome to several kilobase pairs, with many of the smaller changes being undetectable by conventional bacterial artificial chromosome-based comparative genomic hybridization (CGH), which has a resolution of several hundred kilobase pairs.[2] This limitation has recently been overcome by the adaptation of high-density oligonucleotide microarrays that were originally developed for typing of single nucleotide polymorphisms (SNPs) to the evaluation of CNAs.[3] Sophisticated software for such analysis, including dChip (http://biosun1.harvard.edu/complab/dchip) and CNAG (http://www.genome.umin.jp/CNAG.html), is now available online.

Although high-density SNP-typing arrays allow determination of CNAs at a resolution of <100 kbp, it is not a simple task to link such data to changes in the number of the corresponding genes. The HGU133 Plus 2.0 (HGU133P2) microarray manufactured by Affymetrix (USA) was designed to quantitate the abundance of $>47\,000$ human transcripts and has been widely used for gene expression profiling. Most of the probe sequences on the array are targeted to exons corresponding to the 3′ untranslated region of each transcript. Although, as far as we are aware, there have been no reports on the use of this microarray for CNA assessment or CGH, we reasoned that HGU133P2 may be a useful platform for evaluation of CNAs because (1) cDNA sequences are one of the most updated and comprehensive information for human genome, and (2) copy number values could be calculated directly for each gene. Furthermore, it would be straightforward to compare the copy number and expression level of a given gene through separate hybridization of genomic DNA and mRNA to the array. For this purpose, we modified a protocol for labeling and hybridizing genomic DNA, provided by Affymetrix (Supplementary Information).

To examine the fidelity of CNA analysis with the HGU133P2 array, we first isolated genomic DNA from mononuclear cells (MNCs) of peripheral blood from a healthy male volunteer (karyotype: 46,XY) and a healthy female volunteer (karyotype: 46,XX). Genomic DNA was also isolated from a human leukemia cell line (KCL22) with trisomy 8 plus t(9;22) and from transformed human B cells with a karyotype of 48,XXXX or of 47,XY,+21 (both from Coriell Cell Repositories, http://ccr.coriell.org). The DNA was fragmented by treatment with DNase I, end-labeled with biotin-conjugated deoxynucleotides and subjected to hybridization with the HGU133P2 array. We first chose 26 354 probe sets that gave reliable signals — those that received the 'Present' call from the GeneChip operating software (GCOS, Affymetrix) and had a signal intensity of $\geqslant 1.0$ — in the control experiment with the normal male genome. This group contained 960 probe sets that mapped to the X chromosome. The signal intensities of these X chromosome-specific probe sets in the experiment with the normal female genome (46,XX) were then normalized by those of the corresponding probe sets in the data for 46,XY. The mean $\pm$ s.d. value of these XX/XY ratios for the X chromosome-specific probe sets was 2.07 ± 1.19 (Figure 1a) and thus matched the predicted value of 2. The intensities of these probe sets for B cells with the

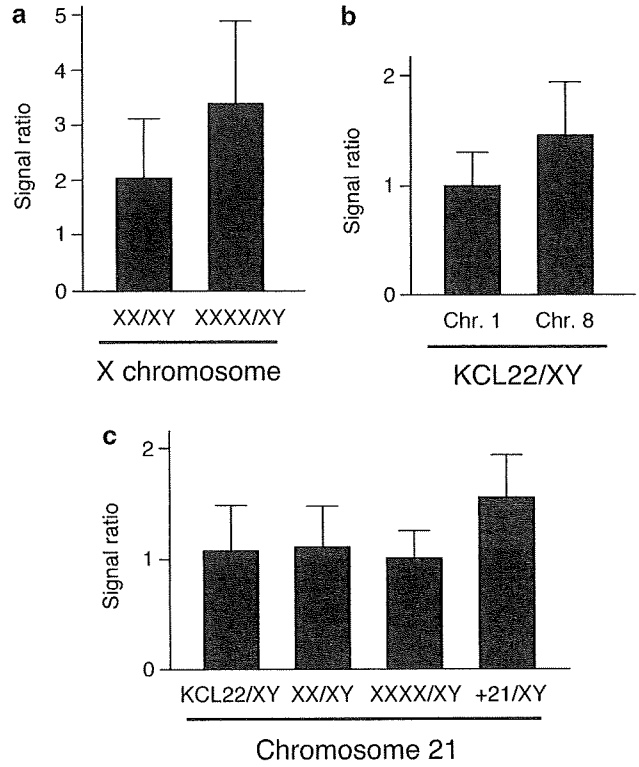

Figure 1 Comparison of chromosome-specific signal intensities obtained by HGU133P2 analysis of genomic DNA isolated from various samples. (**a**) Comparison of signal intensities for the X chromosome. Genomic DNA was purified from MNCs of peripheral blood from a male (XY) or female (XX) volunteer as well as from transformed B cells with a karyotype of 48,XXXX. The DNA was subjected to whole-genome amplification with the REPLI-g kit (Qiagen, Valencia, CA, USA), and the amplification products (100 μg) were digested, labeled with biotin and subjected to hybridization with an HGU133P2 microarray. The signal intensities of the probe sets specific for the X chromosome for the female volunteer or the B cells were normalized by those of the corresponding probe sets for the male volunteer, and the mean + s.d. values were determined. The detailed protocol for CGH analysis with HGU133P2 is described in Supplementary Information. (**b**) Comparison of signal intensities for chromosome 1 or 8. Genomic DNA of KCL22 cells was subjected to hybridization with HGU133P2. The signal intensities of the probe sets specific for chromosome (Chr) 1 or 8 were normalized by those of the corresponding probe sets for the male volunteer, and the mean + s.d. values were determined. (**c**) Comparison of signal intensities for chromosome 21. Genomic DNA of transformed B cells with a karyotype of 47,XY,+21 (+21) was subjected to hybridization with HGU133P2. The signal intensities of the chromosome 21-specific probe sets for these cells, as well as for KCL22, the female volunteer and the B cells with the karyotype 48,XXXX, were normalized by those of the corresponding probe sets for the male volunteer, and the mean + s.d. values were determined.

karyotype 48,XXXX were also normalized by those for 46,XY, giving a mean±s.d. value for the XXXX/XY ratios of 3.34±1.52.

To examine further the linearity of the signal intensity data, we compared the hybridization signals of the probe sets that mapped to chromosome 1 or 8 between the KCL22 cell line and the male control (Figure 1b). The mean±s.d. of the signal intensity ratios (KCL22/XY) for the probe sets on chromosome 8 (1.47±0.516) was well matched to the predicted value of 1.5, whereas that for the probe sets on chromosome 1 (with no abnormality in KCL22) was close to 1.0 (1.01±0.360).

We next compared the signal intensities of the probe sets that mapped to chromosome 21 between the B cells with the karyotype 47,XY,+21, which contain three copies of chromosome 21, and the other samples, all of which contain two copies of this chromosome (Figure 1c). The mean±s.d. of the signal intensities relative to the male control for the cells with trisomy 21 (1.52±0.40) was significantly greater ($P<1.32\times10^{-29}$, Student's *t*-test) than that for the female control (1.13±0.36), for the B cells with 48,XXXX (1.00±0.28) or for KCL22 (1.11±0.41); no significant difference was apparent among the values for the female control, for 48,XXXX and for KCL22.

We then screened for CNAs in leukemic blasts by using HGU133P2 platform and compared the results to the data obtained with a commercially available oligonucleotide-based CGH microarray (Human Genome 44A (CGH44A); Agilent, USA), which harbors >40 000 60-nucleotide oligomers that cover the human genome at a mean resolution of ~35 kbp. The CD34-positive progenitor (leukemic) fraction and the CD34-negative differentiated (control) fraction were isolated from the bone marrow of 23 patients with leukemia (Supplementary Table 1). We were also able to isolate MNCs from the bone marrow of four leukemia patients in complete remission (patient ID nos. 4, 6, 10 and 13), and we thus compared for these individuals the $CD34^+$ fraction obtained before chemotherapy with the $CD34^-$ fraction obtained during complete remission. Although CNA analysis with the HGU133P2 platform was conducted for all subjects, experiments with the CGH44A platform were carried out for 16 patients (ID nos. 4–49).

For the analysis with HGU133P2, genomic DNA isolated from the leukemic and control fractions of each patient was subjected to hybridization with the array separately, and the leukemic fraction/control fraction ratio of signal intensity was calculated for each probe set. For analysis with CGH44A, genomic DNA of the leukemic and control fractions of each patient was labeled with Cy5 and Cy3, respectively, and was mixed before hybridization with the array. To confirm the CGH44A data, we performed a dye-swap experiment for each patient. The influence of outlier probes in the CGH analysis was minimized by transformation of the intensity of each signal with a moving window of 1-Mbp width with the use of CGH Analytics 3.1.8 software (Agilent). Furthermore, gain or loss of chromosome number at given loci was considered significant only if the *z*-score[4] for the corresponding position was ⩾1.0. We considered CNAs identified by CGH reliable only if the original and dye-swap experiments revealed opposite changes each with a *z*-score of ⩾1.0.

The resulting CNA data were highly similar between the HGU133P2 and CGH44A experiments. Nine (~39%) out of the 23 leukemia patients showed a constant copy number of 2 for all chromosomes (data not shown). These results were also confirmed for patient nos. 4, 5 and 6 by hybridization of the corresponding genomic DNA to the Affymetrix 100K SNP-typing array (data not shown). The remaining patients manifested CNAs of various sizes. For instance, both the HGU133P2 and CGH44A analyses revealed a gain of the entire long arm of chromosome 1 and a loss of a short segment of 9q33.1 in patient no. 29 (Figure 2a). The CNAs identified in patient no. 29 by both arrays were not detected by clinical karyotyping (the blasts of the patient were designated 46,XX).

For patient no. 43, both HGU133P2 and CGH44A analyses revealed deletions of 5q, 7q, 9q and 16, a gain of 19q and the same pattern of complex copy number changes at 8p and chromosome 17 (Figure 2b). Again, the clinical karyotyping for this patient (46,XY) failed to detect these various changes.

To confirm the CNAs identified by both HGU133P2 and CGH44A, we measured chromosome copy number in the $CD34^+$ fractions of the leukemia patients with the use of quantitative, real-time polymerase chain reaction (PCR) analysis. The chromosome copy numbers determined by PCR were highly similar to those determined with the arrays for chromosomes 1 and 9q in patient no. 29 and for 8p in patient no. 43 (Figure 3a).

We also prepared single-stranded cDNA from the $CD34^+$ fractions of all subjects except patient no. 72 (RNA with a sufficient quality could not be obtained from this patient), and subjected it to hybridization with HGU133P2 in order to quantitate the amount of each transcript targeted by the array. Given that portions of chromosome 7q were frequently deleted in the leukemia patients, we compared the expression levels of the genes in the deleted regions between the individuals with or without 7q− (Figure 3b). The expression levels of the genes on 7q were significantly lower in the patients with 7q− (patient nos. 28, 43, 44, 60, 61, 70 and 73) than in those without it (18.1±79.5 versus 23.3±87.6 U, mean±s.d.; $P=0.014$, *t*-test).

Additionally, we also examined whether an increase in chromosome copy number affects the expression level of the genes mapped to the chromosomes. As shown in Figure 3c, +8q was observed commonly in five patients (ID nos. 43, 59, 70, 71 and 72). Quantitation of mRNA for the genes mapped to the 8q region has revealed that the expression level of such genes was higher in the individuals with +8q than those without it (42.8±147.3 versus 34.7±114.9; $P=0.022$, *t*-test).

These data indicate that chromosome copy number affects the corresponding level of gene expression; however, given the relatively large s.d. in each data set, other factors (epigenetic changes or interactions with transcriptional factors, for instance) likely have a substantial effect on transcriptional activities. Consistent with this notion, the amount of genomic DNA was not significantly correlated with that of mRNA for the entire collection of probe sets on HGU133P2 (Pearson's correlation coefficient $(r)=6.48\times10^{-4}$, $P=0.881$).

The fluorescence signal intensity for a given gene (probe set) represented on the HGU133P2 array is calculated from those of 11 probes designed for each gene. Given that the physical distance between some of these probes is >10 kbp in the genome, small amplifications or deletions of the genome may be undetectable with HGU133P2. Nevertheless, our data have shown that HGU133P2 is potentially useful for quantitation of chromosome copy number in a gene-oriented manner. Both HGU133P2 and CGH44A array systems have a high resolution, revealing, for instance, a homozygous deletion of a small region of chromosome 9 containing *CDKN2A* and *CDKN2B* in the genome of patient no. 9 (data not shown). The availability of different platforms for CNA assessment will provide insight into different aspects of genomic structural anomalies associated with hematologic malignancies.

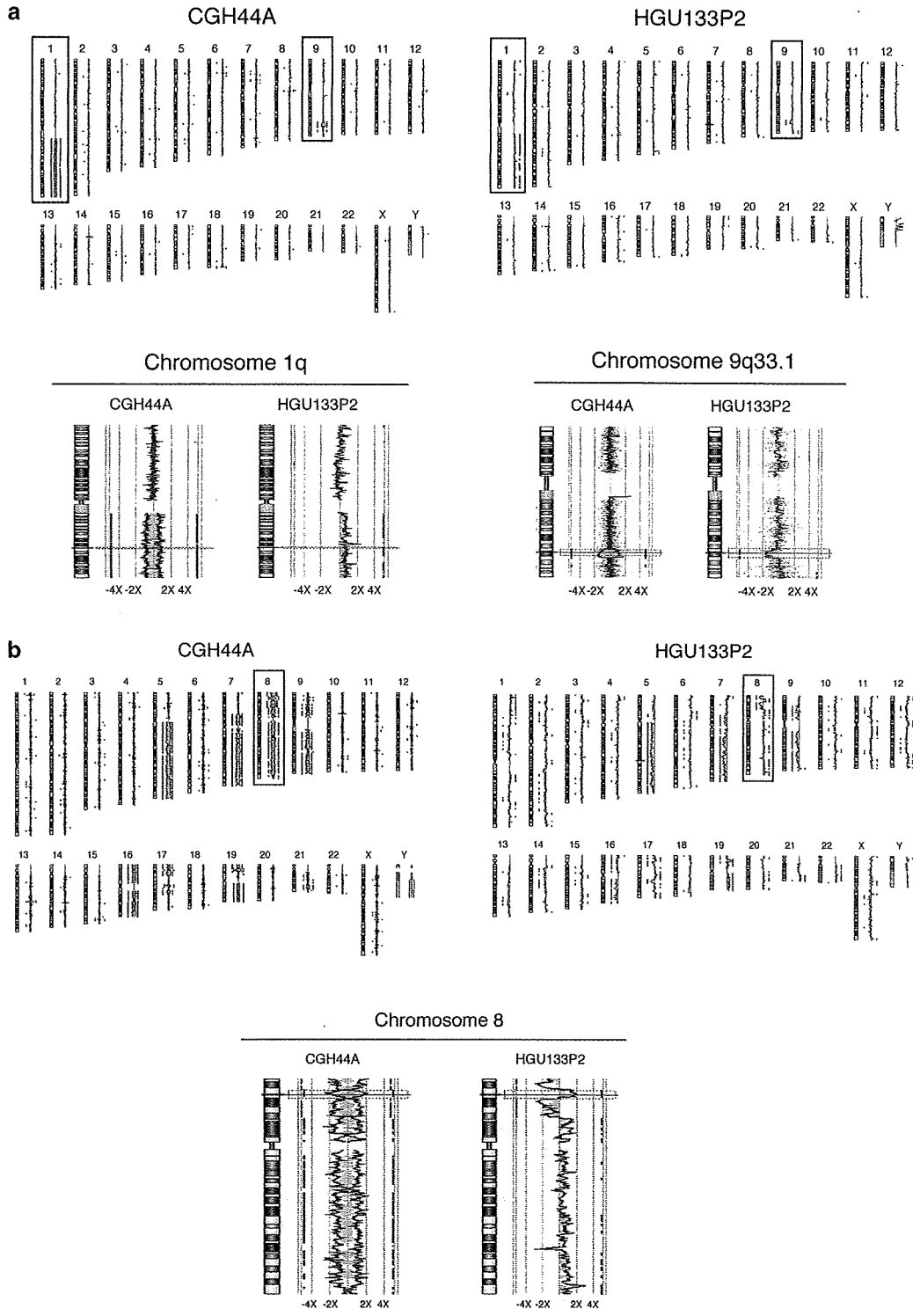

Figure 2 Evaluation of CNAs in leukemia patients with the use of HGU133P2 and CGH44A microarrays. (**a**) Assessment of patient no. 29. $CD34^+$ and $CD34^-$ fractions were isolated from bone marrow with the use of a magnetic bead-conjugated monoclonal antibody to CD34 (CD34 MicroBeads) and a MIDI-MACS cell separation column (both from Miltenyi Biotec, Auburn, CA, USA). For CGH44A analysis, genomic DNA was isolated from $CD34^+$ and $CD34^-$ fractions and labeled with Cy5 and Cy3, respectively; the labeled samples were then mixed and subjected to hybridization with the array.[5] A dye-swap experiment was also performed, and both the original (blue lines) and dye-swap (red lines) data are shown aligned with human chromosome figures by CGH Analytics 3.1.8 (upper panels). Regions with a *z*-score $\geqslant 1.0$ are indicated by side bars and shading (in corresponding colors). The data for chromosome 1 and 9 are shown magnified in the lower panels. For HGU133P2 analysis, genomic DNA isolated from $CD34^+$ and $CD34^-$ fractions was subjected to hybridization with the array separately, and the signal intensities for the former fraction were normalized by those for the latter. Regions with a *z*-score $\geqslant 1.0$ are shown by side bars and blue shading. All data obtained with the CGH44A and HGU133P2 arrays are available at the Gene Expression Omnibus (GEO) web site (http://www.ncbi.nlm.nih.gov/geo) under the accession numbers GSE4659 and GSE4602, respectively. (**b**) Assessment of patient no. 43. Signal intensity data for patient no. 43 were obtained as in (**a**), and the results for chromosome 8 are shown magnified in the lower panel.

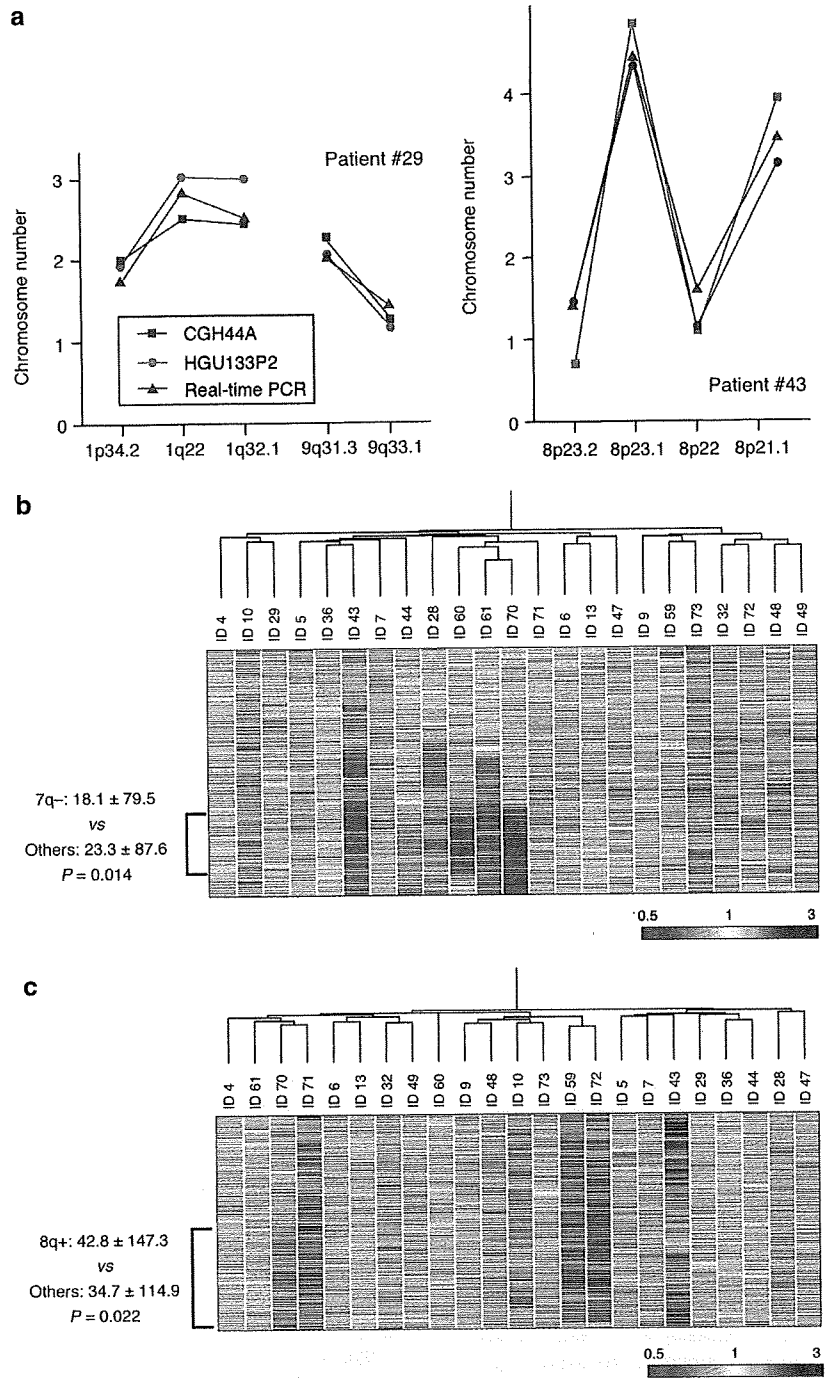

Figure 3 Confirmation of microarray data by quantitative PCR and the effect of DNA copy number on gene expression level. (**a**) Confirmation of microarray data by PCR. The amount of DNA at the indicated chromosomal positions for the $CD34^+$ fractions of patient nos 29 and 43 was determined by real-time PCR and expressed relative to that of DNA corresponding to the glyceraldehyde-3-phosphate dehydrogenase gene. The data are compared with chromosome number determined by CGH44A and HGU133P2 analyses. The sequences of the PCR primers are provided in Supplementary Information. (**b**) Effect of DNA copy number on gene expression level. The leukemic fraction/control fraction DNA ratio was calculated for each probe set on chromosome 7, and the study samples were subjected to hierarchical clustering analysis based on such raw DNA ratio profiles. Each column corresponds to a separate sample, and each row to a probe set whose leukemic fraction/control fraction DNA ratio is color-coded according to the indicated scale. Probe sets are ordered by their physical position in chromosome 7 from top to bottom. mRNA was isolated from the $CD34^+$ fractions of 22 leukemia patients, and the abundance of specific transcripts was quantitated by HGU133P2 array analysis as described previously.[6] The expression level of the genes on a region of the long arm of chromosome 7 was compared between individuals with or without the deletion of the region. The ID numbers of the patients with a loss in chromosome copy number in the region are shown in green. All array data for mRNA quantitation are available at the GEO web site under the accession number GSE4608. (**c**) The leukemic fraction/control fraction DNA ratio was calculated for each probe set on chromosome 8, and the subject tree was constructed as in (**b**). Probe sets are ordered by their physical position in chromosome 8 from top to bottom. The expression level of the genes on the long arm of chromosome 8 was compared between individuals with or without 8q+. The ID numbers of the patients with a gain in chromosome copy number in the region are shown in red.