EML4-ALK Fusion Is Linked to Histological Characteristics in a Subset of Lung Cancers

Kentaro Inamura, MD, PhD, Kengo Takeuchi, MD, PhD,* Yuki Togashi, MPharm,* Kimie Nomura,* Hironori Ninomiya, MD,* Michiyo Okui, PhD,* Yukitoshi Satoh, MD, PhD,*† Sakae Okumura, MD,† Ken Nakagawa, MD,† Manabu Soda, MD, PhD,‡ Young Lim Choi, MD, PhD,‡ Toshiro Niki, MD, PhD,§ Hiroyuki Mano, MD, PhD,‡ and Yuichi Ishikawa, MD, PhD**

Introduction: Very recently, we have found a novel fusion product between the echinoderm microtubule-associated protein-like4 (EML4) and the anaplastic lymphoma kinase (ALK) in non-small cell lung cancers (NSCLCs). Tumors featuring EML4-ALK fusion constitute one subtype of NSCLC that might be highly sensitive to ALK inhibitors. Herein, we present results of a first large scale study of EML4-ALK fusion in lung cancers.

Methods: Using reverse transcription-polymerase chain reaction for EML4-ALK fusion mRNA, we investigated 149 lung adenocarcinomas, 48 squamous cell carcinomas, 3 large-cell neuroendocrine carcinomas, and 21 small-cell carcinomas. For EML4-ALK-positive cancers, we further investigated the presence of ALK fusion proteins by immunohistochemistry.

Results: Five of 149 adenocarcinomas (3.4%) showed EML4-ALK fusion mRNA, this being totally lacking in carcinomas of other types (0/72). In all the fusion-positive cases, ALK fusion protein could be detected in the cytoplasm immunohistochemically. The five fusion cases featured two EML4-ALK variant 1 fusions and three variant 2 fusions. Histologically, both variant 1 cases were mixed type adenocarcinomas, showing papillary with bronchioloalveolar components. Interestingly, all three variant 2 cases were acinar adenocarcinomas, the link being statistically significant ($p = 0.00018$). None of the five fusion-positive cases demonstrated any mutations of EGFR or KRAS, pointing to a mutually exclusive relationship ($p = 0.014$). There was no association with smoking habits.

Conclusions: In the present first investigation of EML4-ALK fusion in a large study of lung cancers (5/221), we found an interesting histotype-genotype relationship. Furthermore, we could detect the fusion protein by immunohistochemistry, pointing to possible clinical applications.

*Department of Pathology, The Cancer Institute and †Department of Chest Surgery, The Cancer Institute Hospital, Japanese Foundation for Cancer Research (JFCR), Tokyo, Japan; and ‡Division of Functional Genomics, §Department of Pathology, Jichi Medical University, Tochigi, Japan.

The first two authors contributed equally to this work.

Disclosure: The authors declare no conflict of interest.

Address for correspondence: Yuichi Ishikawa, MD, PhD, Department of Pathology, JFCR Cancer Institute, 3-10-6 Ariake, Koto-ku, Tokyo 135-8550, Japan. E-mail: ishikawa@jfcr.or.jp

ISSN: 1556-0864/08/0301-0013

Keywords: Lung cancer, EML4-ALK, RT-PCR, Immunohistochemistry, Histology.

(*J Thorac Oncol.* 2008;3: 13–17)

Lung cancer is the leading cause of cancer death in men and women worldwide. Identification of activating mutations of the epidermal growth factor receptor (EGFR) is one of the most intriguing recent discoveries in the field of lung cancer research.[1,2] EGFR mutations are present in one subtype of lung adenocarcinoma, and tumors with this mutation have been shown to be highly sensitive to gefitinib (Iressa). The subtype is prevalent in women, and in patients of Japanese and other Asian ethnicity, especially in nonsmokers.[1,3,4] With the move to personalized cancer therapy, we need to understand oncologic biology at the molecular level in individual lesions to be able to treat cancers with specific molecular-targeting strategies.

Very recently, we have found a novel transforming fusion gene resulting from linkage between the echinoderm microtubule-associated protein-like4 (EML4) and the anaplastic lymphoma kinase (ALK) genes in non-small cell lung cancers (NSCLCs).[5] Tumors featuring EML4-ALK fusion constitute one subtype of NSCLC which might be highly sensitive to ALK inhibitors. The fusion gene is formed by a small inversion within chromosome 2p. EML4 on chromosome 2p21 belongs to the family of echinoderm microtubule-associated protein-like proteins, localized in the cytoplasm, and is necessary for correct microtubule formation.[6,7] ALK on chromosome 2p23 codes for a receptor tyrosine kinase and was first identified as a fusion partner of nucleophosmin (NPM) in anaplastic large-cell lymphomas (ALCLs) with a t(2;5) chromosome rearrangement.[8,9] NPM is an RNA-binding protein that transports ribonucleoproteins between the nucleus and cytoplasm and contributes a nuclear localization signal to the NPM-ALK fusion protein.[10] Other chromosome translocations involving the ALK locus have been identified in ALCLs[11,12] as well as in inflammatory myofibroblastic tumors (IMTs).[13] The fusion point of ALK is conserved among all these chimeric tyrosine kinases including EML4-ALK, resulting in fusion of the entire intracellular kinase domain of ALK to the different partners.[14]

Herein, we present a first large scale study of EML4-ALK fusion in lung cancers, including SCLCs. Furthermore, we detail clinicopathologic and genetic features of fusion-positive lung cancers.

PATIENTS AND METHODS

Clinical Samples and RNA Extraction

This study was conducted with clinical samples from 149 lung adenocarcinomas, 48 squamous cell carcinomas, 3 large cell neuroendocrine carcinomas, and 21 SCLCs. Many of these samples were previously examined and reported.[15–21] For example, most adenocarcinomas were examined as to their mRNA levels of PTEN[19]or HOXB2,[21] and some adenocarcinomas were examined as to their let-7 microRNA levels.[20] All were collected with ethical committee approval and informed consent from patients undergoing surgery at the Cancer Institute Hospital, Tokyo, Japan, between May 1995 and August 2004. Histologic diagnosis was according to World Health Organization classifications[22] as well as to differentiation-grading criteria for adenocarcinomas of the Japanese Lung Cancer Society.[23,24] All lesions were grossly dissected and snap-frozen in liquid nitrogen within 20 minutes of removal and stored at −80°C until total RNA extraction and purification using an RNeasy Mini Kit (QIAGEN, Valencia, CA). RNA quality and absence of genomic DNA contamination were checked by formaldehyde agarose gel electrophoresis.

Reverse Transcription-Polymerase Chain Reaction and Sequencing Analysis

Total RNAs were reverse transcribed with random primers and SuperScript III reverse transcription (Invitrogen, Carlsbad, CA). To detect fusion transcripts derived from EML4 and ALK, reverse transcription-polymerase chain reaction (RT-PCR) experiments were carried out with primers Fusion-RT-S (5′-GTGCAGTGTTTAGCATTCTTGGGG-3′) and Fusion-RT-AS (5′-TCTTGCCAGCAAAGCAGTAGTTGG-3′). We used PCR primers 5′-GTCAGTGGTGGACCTGACCT-3′ and 5′-TGAGCTTGACAAAGTGGTCG for the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as an internal control. For PCR of the fusion transcripts, after initial denaturation at 94°C for 10 minutes, 32 cycles each consisting of denaturation at 94°C for 1 minute, annealing at 60°C for 1 minute, and strand elongation at 72°C for 1 minute were performed, followed by a final elongation at 72°C for 10 minutes. For GAPDH, amplification was performed for 35 cycles with denaturation for 1 minute at 94°C, primer annealing for 30 seconds at 58°C, and elongation for 30 seconds at 72°C. PCR was performed using AmpliTaq Gold (Applied Biosystems, Foster City, CA) and amplified fragments were subjected to direct sequence analysis.

Immunohistochemical Analysis

A representative tissue block from each lesion was selected, and 4-μm tissue sections were routinely deparaffinized in xylene and rehydrated through graded ethanols. Immunohistochemical staining was performed using the EnVision + DAB system (DAKO, Carpinteria, CA) and a mouse monoclonal anti-ALK antibody (ALK1, DAKO, 1:20).

DNA Extraction and Mutation Analysis of EGFR and KRAS

Of 149 patients with adenocarcinomas, both EGFR and KRAS data were available for 62 and EGFR data alone for a further 12. DNA extraction and mutation analysis of EGFR and KRAS were performed as described previously.[19]

Analysis of Clinicopathologic Parameters

Survival data were analyzed by the log-rank test using cancer death-specific survival data. We analyzed statistical correlations for the other clinicopathologic features using the Student t test, Fisher exact test, or χ^2 test as appropriate. The two-sided significance level was set at $p < 0.05$.

RESULTS

Using RT-PCR for EML4-ALK fusion mRNA, we investigated the presence of the EML4-ALK translocation in 221 lung cancers (Table 1). Five of 149 adenocarcinomas (3.4%) featured EML4-ALK fusion mRNA, whereas other types of carcinoma were all negative (0/72) (Figure 1). Of the five fusion-positive cases, two had EML4-ALK variant 1 and three had variant 2.[5] The fusions were confirmed by direct sequencing.

Histologically, both the variant 1 cases were mixed type adenocarcinomas, papillary with bronchioloalveolar components (Figure 2*A*). Interestingly, all three variant 2 cases were acinar adenocarcinomas, moderately or poorly differentiated (Figure 2*B*). The link between variant 2 and acinar morphology was statistically significant ($p = 0.00018$, Fisher exact test).

Immunohistochemically, all the five fusion-positive cases showed ALK fusion protein in the cytoplasm (Figure 3) in line with the absence of any nuclear localization signal in the EML4 gene. We cannot rule out the possibility of detecting endogenous ALK protein.

Table 2 summarizes details for clinicopathologic and genetic features of the fusion-positive lung cancers. Genetically, all lacked mutations of EGFR or KRAS ($p = 0.014$ as

TABLE 1. EML4-ALK Fusion and Histology

Histology	Total	EML4-ALK(+)	EML4-ALK(−)
Adenocarcinoma	149	5 (3.4%)	144 (97%)
Subtype			
Adenocarcinoma with mixed subtype	89	2[a] (2.3%)	87 (98%)
Papillary adenocarcinoma	35	0 (0%)	35 (100%)
Acinar adenocarcinoma	18	3[b] (17%)	15 (83%)
Solid adenocarcinoma with mucin	4	0 (0%)	4 (100%)
Bronchioloalveolar carcinoma	3	0 (0%)	3 (100%)
Squamous cell carcinoma	48	0 (0%)	48 (100%)
Large cell neuroendocrine carcinoma	3	0 (0%)	3 (100%)
Small cell carcinoma	21	0 (0%)	21 (100%)

[a] Variant 1; Fisher exact test, $p = 0.66$ (Adenocarcinoma with mixed subtype vs. the other adenocarcinomas).

[b] Variant 2; Fisher exact test, $p = 0.00018$ (Acinar adenocarcinoma vs. the other adenocarcinomas).

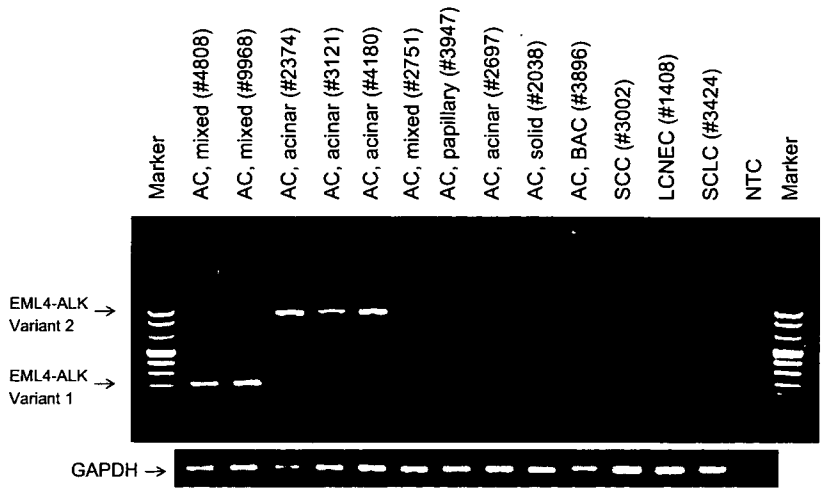

FIGURE 1. RT-PCR for EML4-ALK fusion mRNA. All the 5 fusion-positive cases and fusion negative cases of all the histologic subtypes examined are shown. RT-PCR results for GAPDH mRNA are also included as an internal control. AC, adenocarcinoma; mixed, adenocarcinoma with mixed subtype; papillary, papillary adenocarcinoma; acinar, acinar adenocarcinoma; solid, solid adenocarcinoma with mucin; BAC, bronchioloalveolar carcinoma; SCC, squamous cell carcinoma; LCNEC, large cell neuroendocrine carcinoma; SCLC, small cell lung carcinoma; NTC, no template control.

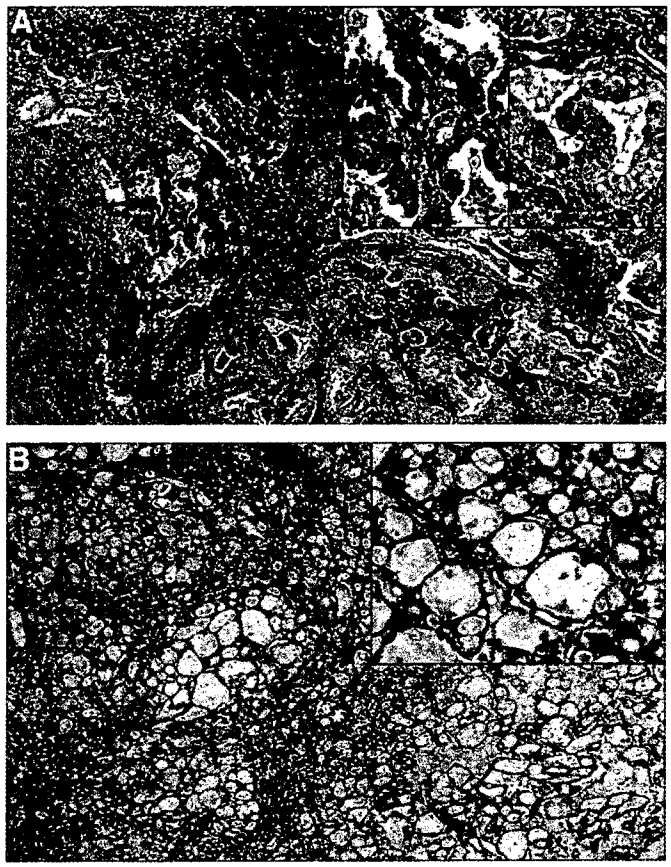

FIGURE 2. Representative examples of histologic features. Both the 2 variant 1 cases were mixed type adenocarcinomas, with papillary and bronchioloalveolar components (*A*). All 3 variant 2 cases were acinar adenocarcinomas (*B*).

compared with expectation). Patients with EML4-ALK fusion-positive tumors were younger than those without by 4 years, though this was not statistically significant. One patient was 43 years old and died 4 months after surgery (variant 2, acinar adenocarcinoma, poorly differentiated, clinical stage-IV, cerebellar metastasis). The other four patients were 58 to 66 years old and are alive now. There was no association between EML4-ALK fusion and smoking habits, although the sample numbers were small ($p = 0.77$).

We analyzed the survival data statistically with the log rank test, but there was no prognostic significance of EML-ALK fusion ($p = 0.84$).

DISCUSSION

In the present first large scale study of a novel EML4-ALK fusion in 221 lung cancers including 21 SCLCs, 5 of 149 adenocarcinomas (3.4%) proved positive for fusion mRNA and fusion protein. Interesting histotype-genotype relationships were observed. Although both variant 1 cases were papillary adenocarcinomas with bronchioloalveolar components, all the three variant 2 cases were of acinar type. Furthermore, none of these lesions had mutations in EGFR or KRAS, pointing to a mutually exclusive relationship.

The ALK gene encodes a transmembrane receptor tyrosine kinase that belongs to the insulin receptor superfamily and is most similar to leukocyte tyrosine kinase.[25] Postnatal ALK expression is normally restricted to a few scattered cells in the nervous system,[26] but chromosomal translocations involving the ALK are characteristic of ALCLs and IMTs. An increasing number of translocation patterns have been identified and other neoplasms with similar changes have been identified, such as large B-cell lymphomas.[27] Very recently, we described a novel subpopulation of NSCLCs with ALK translocations,[5] and it is very likely that other examples will be identified by further searches. ALK-positive ACLCs predominantly affect younger patients and, if treated with chemotherapy, have a more favorable prognosis than their negative counterparts.[28] Similarly ALK-positive IMTs primarily affect younger patients.[13] In this study, ALK-positive lung adenocarcinomas were also found in younger patients when compared with ALK-negative tumors. Especially, one of the five patients was 43 years old, very young for lung adenocarcinomas. Although one example is insufficient for discussion, ALK-positive adenocarcinomas might include younger patients. Although survival data analysis demonstrated no significant difference between ALK-positive

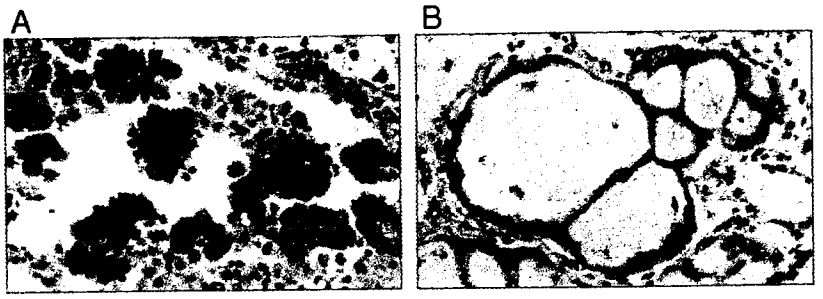

FIGURE 3. Representative examples of immunohistochemical features. Both adenocarcinomas with mixed subtype (*A*) with the variant 1 EML4-ALK fusion and acinar adenocarcinomas (*B*) with the variant 2 show ALK fusion protein in their cytoplasm.

TABLE 2. Relationship between EML4-ALK Fusion and Clinicopathologic and Genetic Features in Lung Adenocarcinomas

		EML4-ALK fusion		
Variables category	**No. samples (%)**	**(+) (n = 5)**	**(−) (n = 144)**	p
Age (yr; mean ± SD)	149	59.4 ± 9.7	63.4 ± 8.7	0.31[a]
Sex				0.87[b]
Male	80 (54%)	2 (40%)	78 (54%)	
Female	69 (46%)	3 (60%)	66 (46%)	
Smoking habit				0.77[b]
Never	65 (44%)	3 (60%)	62 (43%)	
Smoker	84 (56%)	2 (40%)	82 (57%)	
Tumor size				0.40[b]
<30mm	77 (52%)	4 (80%)	73 (51%)	
30 mm ≤	72 (48%)	1 (20%)	71 (49%)	
Differentiation				0.73[c]
Well	48 (32%)	1 (20%)	47 (33%)	
Moderate	62 (42%)	2 (40%)	60 (42%)	
Poor	39 (26%)	2 (40%)	37 (26%)	
EGFR				0.034[b]
Mutation (+)	41 (55%)	0 (0%)	41 (59%)	
Mutation (−)	33 (45%)	5 (100%)	28 (41%)	
KRAS				0.92[b]
Mutation (+)	7 (11%)	0 (0%)	7 (12%)	
Mutation (−)	55 (89%)	5 (100%)	50 (88%)	
EGFR or KRAS				0.014[b]
Mutation (+)	38 (61%)	0 (0%)	38 (67%)	
Mutation (−)	24 (39%)	5 (100%)	19 (33%)	
p-Stage				0.73[b]
I	63 (43%)	2 (40%)	61 (43%)	
II–IV	85 (57%)	3 (60%)	82 (57%)	

Percentages may not total 100, because of rounding.
[a] Student t test.
[b] Fisher exact test.
[c] Yates χ^2 test.

and negative adenocarcinomas, this might be due to the small number of positive cases. Whatever is the cause, for ALK-positive tumors, molecular targeted therapies including ALK inhibitors may be used.

ALK1 antibody, used in the immunohistochemical analysis, detects the cytoplasmic region of the ALK protein and also detects the full-length endogenous ALK protein. When we detect the positive staining of ALK1, three possibilities are considerable: (i) EML4-ALK fusion protein, (ii) endogenous full-length ALK protein, or (iii) ALK fusion protein with another partner. The five EML4-ALK fusion cases immunostained positive for ALK with variable intensity. Endogenous full-length ALK protein might, however, be also detected by immunohistochemistry. Therefore, EML4-ALK fusion should be confirmed by RT-PCR practically, although the immunohistochemistry can be used for the screening purpose.

In conclusion, we here found a minor subpopulation of lung adenocarcinomas featuring EML4-ALK fusion with evidence of histotype-genotype relationships. Furthermore, we could detect the fusion protein by immunohistochemistry, pointing to clinical applications.

ACKNOWLEDGMENTS

Supported financially by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, from the Japan Society for the Promotion of Science, and by grants from the Ministry of Health, Labour and Welfare, the Smoking Research Foundation, the National Institute of Biomedical Innovation and the Vehicle Racing Commemorative Foundation.

The authors thank Ms. Kazuko Yokokawa, Mr. Motoyoshi Iwakoshi, Ms. Miyuki Kogure, and Ms. Tomoyo Kakita for their technical assistance, and Ms. Kimiko Meguro for secretarial work.

REFERENCES

1. Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304: 1497–1500.
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;350:2129–2139.
3. Pao W, Miller V, Zakowski M, 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 U S A* 2004;101:13306–13311.
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;97:339–346.
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; 448:561–566.
6. Eudy JD, Ma-Edmonds M, Yao SF, et al. Isolation of a novel human homologue of the gene coding for echinoderm microtubule-associated protein (EMAP) from the Usher syndrome type 1a locus at 14q32. *Genomics* 1997;43:104–106.
7. Pollmann M, Parwaresch R, Adam-Klages S, et al. Human EML4, a novel member of the EMAP family, is essential for microtubule formation. *Exp Cell Res* 2006;312:3241–3251.
8. Morris SW, Kirstein MN, Valentine MB, et al. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* 1994;263:1281–1284.
9. Shiota M, Fujimoto J, Semba T, et al. Hyperphosphorylation of a novel 80 kDa protein-tyrosine kinase similar to Ltk in a human Ki-1 lymphoma cell line, AMS3. *Oncogene* 1994;9:1567–1574.
10. Chan WY, Liu QR, Borjigin J, et al. Characterization of the cDNA encoding human nucleophosmin and studies of its role in normal and abnormal growth. *Biochemistry* 1989;28:1033–1039.
11. Hernandez L, Pinyol M, Hernandez S, et al. TRK-fused gene (TFG) is a new partner of ALK in anaplastic large cell lymphoma producing two structurally different TFG-ALK translocations. *Blood* 1999;94:3265–3268.
12. Lamant L, Dastugue N, Pulford K, et al. A new fusion gene TPM3-ALK in anaplastic large cell lymphoma created by a (1;2)(q25;p23) translocation. *Blood* 1999;93:3088–3095.
13. Lawrence B, Perez-Atayde A, Hibbard MK, et al. TPM3-ALK and TPM4-ALK oncogenes in inflammatory myofibroblastic tumors. *Am J Pathol* 2000;157:377–384.
14. Duyster J, Bai RY, Morris SW, et al. Translocations involving anaplastic lymphoma kinase (ALK). *Oncogene* 2001;20:5623–5637.
15. Virtanen C, Ishikawa Y, Honjoh D, et al. Integrated classification of lung tumors and cell lines by expression profiling. *Proc Natl Acad Sci U S A* 2002;99:12357–12362.
16. Jones MH, Virtanen C, Honjoh D, et al. Two prognostically significant subtypes of high-grade lung neuroendocrine tumours independent of small-cell and large-cell neuroendocrine carcinomas identified by gene expression profiles. *Lancet* 2004;363:775–781.
17. Inamura K, Fujiwara T, Hoshida Y, et al. Two subclasses of lung squamous cell carcinoma with different gene expression profiles and prognosis identified by hierarchical clustering and non-negative matrix factorization. *Oncogene* 2005;24:7105–7113.
18. Inamura K, Shimoji T, Ninomiya H, et al. A metastatic signature in entire lung adenocarcinomas irrespective of morphological heterogeneity. *Hum Pathol* 2007;38:702–709.
19. Inamura K, Togashi Y, Nomura K, et al. Up-regulation of PTEN at the transcriptional level is an adverse prognostic factor in female lung adenocarcinomas. *Lung Cancer* 2007;57:201–206.
20. Inamura K, Togashi Y, Nomura K, et al. let-7 microRNA expression is reduced in bronchioloalveolar carcinoma, a non-invasive carcinoma, and is not correlated with prognosis. *Lung Cancer* 2007;58:392–396.
21. Inamura K, Togashi Y, Okui M, et al. HOXB2 as a novel prognostic indicator for stage I lung adenocarcinomas. *J Thorac Oncol* 2007;2:802–807.
22. Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC. World Health Organization Classification of Tumours: Pathology and Genetics of Tumours of the Lung, Pleural, Thymus and Heart. Berlin: Springer, 2004.
23. Japan Lung Cancer Society. General Rules for Clinical and Pathologic Record of Lung Cancer [in Japanese], 5th Ed. Tokyo: Kanahara, 1999.
24. Ishikawa Y, Furuta R, Miyoshi T, et al. Loss of heterozygosity and the smoking index increase with decrease in differentiation of lung adenocarcinomas: etiologic implications. *Cancer Lett* 2002;187:47–51.
25. Morris SW, Naeve C, Mathew P, et al. ALK, the chromosome 2 gene locus altered by the t(2;5) in non-Hodgkin's lymphoma, encodes a novel neural receptor tyrosine kinase that is highly related to leukocyte tyrosine kinase (LTK). *Oncogene* 1997;14:2175–2188.
26. Pulford K, Lamant L, Morris SW, et al. Detection of anaplastic lymphoma kinase (ALK) and nucleolar protein nucleophosmin (NPM)-ALK proteins in normal and neoplastic cells with the monoclonal antibody ALK1. *Blood* 1997;89:1394–1404.
27. De Paepe P, Baens M, van Krieken H, et al. ALK activation by the CLTC-ALK fusion is a recurrent event in large B-cell lymphoma. *Blood* 2003;102:2638–2641.
28. Stein H, Foss HD, Durkop H, et al. CD30(+) anaplastic large cell lymphoma: a review of its histopathologic, genetic, and clinical features. *Blood* 2000;96:3681–3695.

Available online at www.sciencedirect.com
ScienceDirect
Leukemia Research 31 (2007) 203–209
Leukemia Research
www.elsevier.com/locate/leukres

Identification of a constitutively active mutant of JAK3 by retroviral expression screening

Young Lim Choi [a], Ruri Kaneda [a], Tomoaki Wada [a], Shin-ichiro Fujiwara [a], Manabu Soda [a], Hideki Watanabe [a], Kentaro Kurashina [a], Hisashi Hatanaka [a], Munehiro Enomoto [a], Shuji Takada [a], Yoshihiro Yamashita [a], Hiroyuki Mano [a,b,*]

[a] *Division of Functional Genomics, Jichi Medical University, 3311-1 Yakushiji, Shimotsukeshi, Tochigi 329-0498, Japan*
[b] *CREST, Japan Science and Technology Agency, Saitama 332-0012, Japan*

Received 25 March 2006; received in revised form 2 May 2006; accepted 2 May 2006
Available online 21 June 2006

Abstract

To identify transforming genes in acute myeloid leukemia (AML) we here constructed a retroviral cDNA expression library from an AML patient, and then used this library to infect a mouse cell line 32Dcl3-mCAT. cDNA inserts of the cell clones which proliferated in the presence of granulocyte colony-stimulating factor were derived from *JAK3* encoding a JAK3 mutant with a valine-to-alanine substitution at codon 674 and two additional amino acid substitutions. The transforming activity of JAK3(V674A) was confirmed by its introduction into 32Dcl3-mCAT. Sequencing of the original *JAK3* cDNA derived from the patient, however, failed to detect the V674A mutation.

Keywords: JAK3; Retrovirus; Acute myeloid leukemia; cDNA expression library

1. Introduction

Acute myeloid leukemia (AML) is a clonal disorder of immature progenitor cells in the hematopoietic system. Chromosomal translocations are present in the AML blasts of 20–30% of individuals with this condition, and the fusion genes generated by such translocations have been shown to possess transforming activity [1,2]. Even with current technology, however, chromosomal abnormalities are not detectable in the blasts of almost half of AML patients [3]. Furthermore, although point mutations in a variety of genes implicated in cell growth or differentiation, such as *RAS*, *FLT3*, *KIT*, and *RUNX1*, are detectable in some such blasts, many cases of AML are not associated with a detectable gene anomaly [4]. Clarification of the transforming events in such AML cases would thus be expected to provide a basis for the development of effective treatments.

Functional screening based on transforming activity is one potential approach to the efficient isolation of tumor-promoting genes in AML. Focus formation assays with mouse 3T3 fibroblasts have indeed proved successful for the identification of oncogenes in human cancer [5]. In such screening assays, genomic DNA isolated from cancer specimens is used to transfect 3T3 cells and the formation of transformed cell foci is then evaluated. There are, however, substantial drawbacks to such screening for AML. First, expression of the exogenous genes in conventional 3T3 screening is driven by the associated promoters or enhancers, so that oncogenes are able to exert transforming effects in 3T3 cells only if their regulatory regions are active in fibroblasts, which is not guaranteed. Furthermore, given that the transcriptome and proteome would be expected to differ markedly between fibroblasts and leukemic blood cells, active oncogenes in the latter cells may not function properly in the former cells even if they are adequately expressed.

We reasoned that these concerns may be overcome through expression of test cDNAs under the control of an ectopic

*. Corresponding author. Tel.: +81 285 58 7449; fax: +81 285 44 7322.
E-mail address: hmano@jichi.ac.jp (H. Mano).

0145-2126/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.leukres.2006.05.006

promoter and with the use of myeloid cells, instead of 3T3 cells, for the assay. We have now constructed a retroviral cDNA expression library from a purified $CD133^+$ stem cell fraction isolated from a patient with AML and used this library to infect the mouse myeloid cell line 32Dcl3 [6]. We then screened for transforming genes that override the differentiation program of 32Dcl3 cells triggered by granulocyte colony-stimulating factor (G-CSF). Furthermore, in preparation of the cDNA library, we took advantage of a polymerase chain reaction (PCR)-based system that preferentially amplifies full-length cDNAs [7]. Through this screening, we isolated a cDNA encoding a constitutively active form of the protein kinase JAK3. However, investigation of *JAK3* cDNA sequence in the original RNA failed to detect the activating mutation, indicating that the mutation arose probably from an experimental artifact.

2. Materials and methods

2.1. Cell culture and clinical samples

32Dcl3 cells were cultured in RPMI 1640 medium (Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum (FBS, Invitrogen) and mouse interleukin (IL) – 3 (20 U/mL; Sigma, St. Louis, MO). To increase the efficiency of retroviral infection, we generated 32D-mCAT cells, which overexpress a receptor (mCAT) for ecotropic retrovirus [8], through infection of 32Dcl3 cells with a recombinant retrovirus containing both mCAT and blasticidin-S resistance genes. For induction of granulocyte differentiation, 32D-mCAT cells were cultured with mouse G-CSF (0.5 ng/mL, Sigma) instead of IL-3. The BOSC23 packaging cell line [9] was maintained in Dulbecco's modified Eagle's medium – F12 (Invitrogen) supplemented with 10% FBS and 2 mM L-glutamine.

$CD133^+$ cells were purified with anti-CD133 MicroBeads and a Mini-MACS magnetic cell separation column (Miltenyi Biotec, Auburn, CA), as described previously [10], from mononuclear cells of bone marrow from a patient with AML, who provided written informed consent.

2.2. Construction of a retroviral library

A retroviral plasmid library was constructed as described previously [7]. In brief, total RNA was extracted from the purified $CD133^+$ cells with the use of an RNeasy Mini column and RNase-free DNase (Qiagen, Valencia, CA), and first-strand cDNAs were then synthesized with PowerScript reverse transcriptase, the SMART IIA oligonucleotide, and CDS primer IIA (all from Clontech, Palo Alto, CA). The resulting cDNAs were then amplified for 20 cycles with 5′-PCR primer IIA and LA Taq polymerase (Takara Bio, Shiga, Japan). The PCR products were ligated to a BstXI adapter (Invitrogen) and then incorporated into the pMX retroviral plasmid [11]. The pMX-cDNA plasmids were introduced into ElecroMax DH10B cells (Invitrogen) by electroporation.

2.3. 32D-mCAT transformation assay

BOSC23 cells (1.8×10^6) were seeded in a 6-cm culture plate and transfected with a mixture comprising 2 μg of retroviral plasmids, 0.5 μg of pGP plasmid (to express *gag-pol* proteins, Takara Bio), 0.5 μg of pE-eco plasmid (to express ecotropic *env* protein, Takara Bio), and 18 μL of Lipofectamine reagent (Invitrogen). Two days after transfection, the culture supernatant was collected and incubated with 32D-mCAT cells (5×10^5) for 24 h in the presence of Retronectin (Takara Bio). The 32D-mCAT cells were then incubated for an additional 24 h with mouse IL-3, washed with cytokine-free medium, and cultured at an initial density of 1×10^4 to 4×10^4 cells/mL in the presence of mouse G-CSF (0.5 ng/mL).

2.4. Recovery of cDNAs from 32D-mCAT cells

Genomic DNA was isolated from 32D-mCAT cells that grew without differentiation (determined by the presence of polymorphonuclear neutrophils as assessed by cytospin preparations stained with the Wright Giemsa solutions) in the presence of G-CSF and was 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. To confirm the transforming activity of the isolated cDNAs, they were again isolated by PCR with 5′-PCR primer IIA and PfuUltra High-Fidelity DNA polymerase (Stratagene, La Jolla, CA) for 30 cycles of 95 °C fro 30 s, 60 °C for 30 s and 72 °C for 4 min. The cDNAs were then individually ligated into the pMX plasmid and used to generate recombinant retroviruses. 32D-mCAT cells were then infected with the resulting viruses and cultured with G-CSF.

2.5. Protein analysis

Wild-type human *JAK3* cDNA (a kind gift of Dr. James N. Ihle) was used to generate mutant cDNAs with the use of a QuickChange site-directed mutagenesis kit (Stratagene, La Jolla, CA). Oligonucleotides used for mutagenesis were 5′-GGGATGGGGGGCTGTACGTAGATGGGGTGGC-3′ and 5′-GCCACCCCATCTACGTACAGCCCCCCATCCC-3′ for JAK3(H463Y), 5′-GAGTGACCCTGGGGCCAGCCCCGC-TGTGTTAAGCC-3′ and 5′-GGCTTAACACAGCGGG-GCTGGCCCCAGGGTCACTC-3′ for JAK3(V674A), and 5′-GATGGGATGTGAGCGGGGTGTCCCCGCCCTCTG-3′ and 5′-CAGAGGGCGGGGACACCCCGCTCACATCC-CATC-3′ for JAK3(D1043G). The wild-type or mutant cDNAs were individually ligated into the pMX-ires-neo retroviral vector in order to generate recombinant retroviruses encoding both JAK and the neomycin resistance gene. 32D-mCAT cells were infected with the resulting viruses and cultured for >2 weeks in the presence of IL-3 and

G418 (1 mg/mL, Invitrogen). JAK3 was immunoprecipitated from neomycin-resistant mass cultures of 32D-mCAT cells with goat polyclonal antibodies specific for this protein (sc-1078; Santa Cruz Biotechnology, Santa Cruz, CA). Half of the resulting precipitates were subjected to immunoblot analysis with a mouse monoclonal antibody to phosphotyrosine (4G-10; Upstate Biotechnology, Charlottesville, VA) or rabbit polyclonal antibodies to JAK3 (sc-513, Santa Cruz Biotechnology). Antibodies used for the immunoprecipitation of STAT1 were obtained from Upstate Biotechnology, those specific for STAT3 were from Cell Signaling Technology (Danvers, MA), and those specific for STAT5 or STAT6 were from Santa Cruz Biotechnology. Immune complexes on immunoblots were visualized with SuperSignal West Pico Chemiluminescent substrate (Pierce, Rockford, IL).

The remaining half of the immunoprecipitates were then washed with kinase buffer [10 mM HEPES–NaOH (pH 7.4), 50 mM NaCl, 5 mM $MgCl_2$, 5 mM $MnCl_2$, 0.1 mM Na_3VO_4] and then incubated for 30 min at room temperature in 30 μL of the kinase buffer containing 0.37 MBq of [γ-^{32}P]ATP (GE Healthcare Bio-Sciences, Uppsala, Sweden) and 1 μg of a synthetic peptide (LLPLDKDYYVVREPGQS) corresponding to amino acids 973–989 of human JAK3 (Operon Biotechnologies, Huntsville, AL).

3. Results

3.1. Screening for transforming genes in 32D-mCAT cells

To isolate transforming genes from a retroviral expression library, we used mouse 32Dcl3 cells, which proliferate in the presence of IL-3 but undergo terminal differentiation to granulocytes in response to G-CSF [6]. Introduction of v-Ha-Ras or v-Abl oncogenes into these cells induces continuous growth even in the presence of G-CSF [12,13], indicating that active oncogenes are able to override the differentiation program of 32Dcl3 cells. We thus reasoned that retroviral transduction of transforming genes present in AML blasts into 32Dcl3 cells might also elicit proliferation in the presence of G-CSF. Given that the efficiency of infection of 32Dcl3 cells with ecotropic retroviruses was low (<2%), we generated 32D-mCAT cells, which overexpress an ecotropic retrovirus receptor and exhibit an infection efficiency of 10–30% (data not shown). We then examined the feasibility of a novel screening system for transforming genes that consists of (1) infection of 32D-mCAT cells with recombinant retroviral cDNA libraries constructed from AML specimens, (2) selection of proliferating 32D-mCAT cells in the presence of G-CSF, and (3) isolation of transforming cDNAs by PCR with a primer targeted to a sequence flanking the cDNA in each retrovirus.

3.2. Screening with a retroviral cDNA expression library from an AML patient

To screen for transforming genes in AML, we prepared a CD133$^+$ hematopoietic stem cell-like fraction from an 84-year old Japanese man with AML (M2 subtype according to the FAB classification). The patient had a clinical history of myelodysplastic syndrome prior to the onset of AML, and his malignant blasts had a karyotype of monosomy 7. Purification of a CD133$^+$ immature cell faction likely enriched for

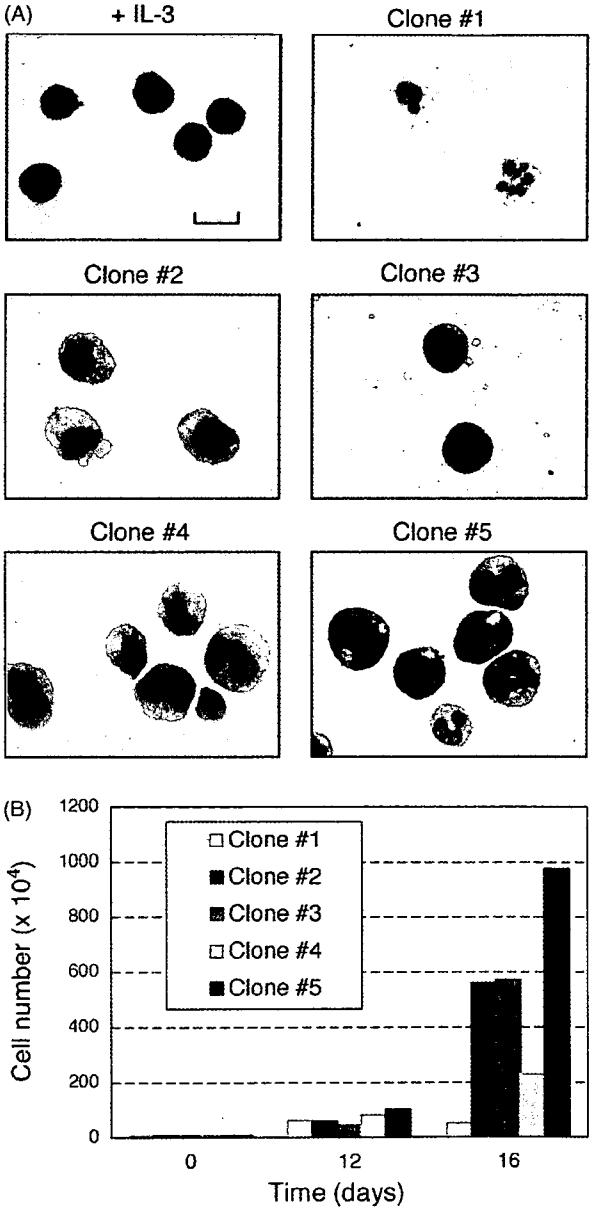

Fig. 1. Isolation of 32D-mCAT clones that proliferate in the presence of G-CSF. (A) 32D-mCAT cells cultured with IL-3 (+IL-3) as well as 32D-mCAT clones infected with the retroviral expression library (clones #1, #2, #3, #4, and #5) and cultured for 16 days with G-CSF were stained with Wright-Giemsa solutions and photographed. Scale bar, 100 μm. (B) Proliferation of 32D-mCAT cells of clones #1 to #5 was determined by counting the cell number after culture with G-CSF for 12 or 16 days. The experiments were repeated twice, and both data were basically identical.

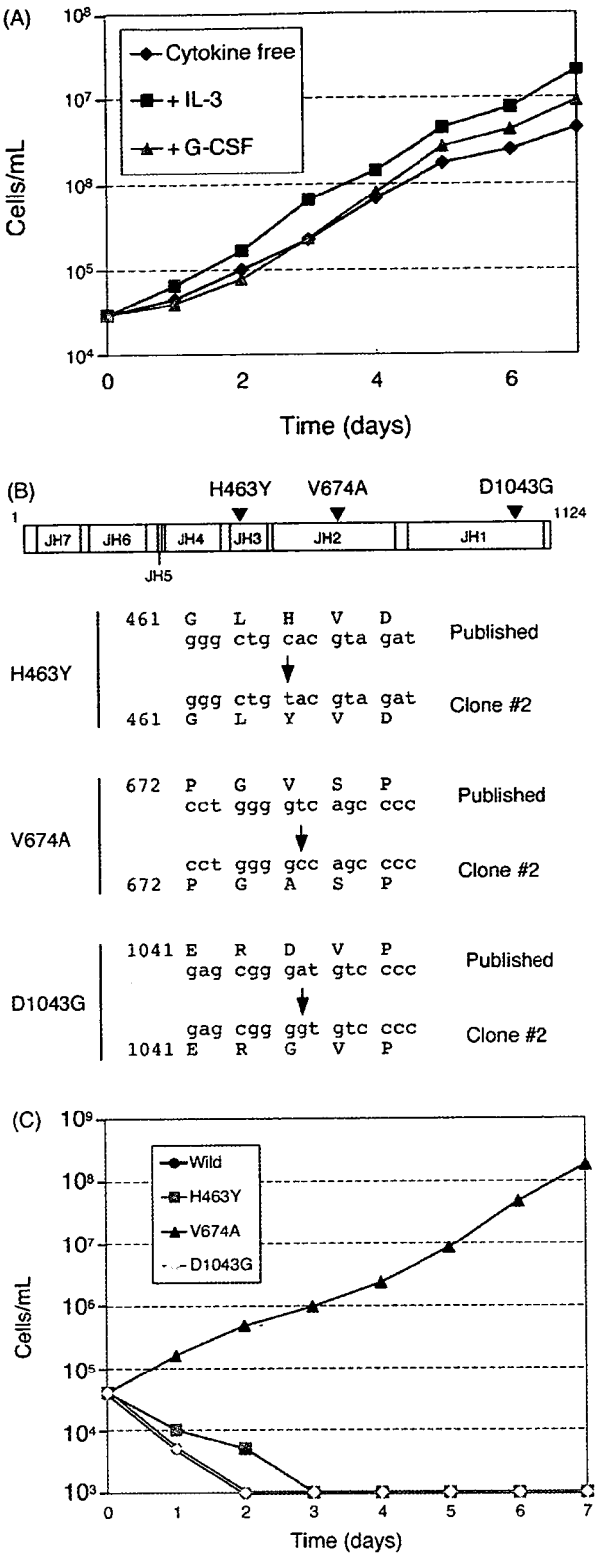

Fig. 2. Expression of JAK3(V674A) abrogates cytokine dependency of 32D-mCAT cell growth. (A) 32D-mCAT cells were infected with recombinant retroviruses containing the *JAK3* cDNA isolated from 32D-mCAT clone #2. The cells were then cultured in the absence or presence of IL-3 or G-CSF, and cell number was counted at the indicated times thereafter. The experiments were repeated twice, and both data were basically identical. (B) Schematic structure of human JAK3 showing the JH domains and the positions of amino acid substitutions encoded by the PCR product obtained from 32D-mCAT clone #2. The corresponding nucleotide changes in the *JAK3* cDNA sequence are shown below. (C) Growth curves for 32D-mCAT cells infected with retroviruses encoding wild-type human JAK3 or the mutants JAK3(H463Y), JAK3(V674A), or JAK3(D1043G) and cultured without cytokine. The experiments were repeated twice, and both data were basically identical.

leukemic clones and helped to eliminate mRNA of normal bone marrow cells [10].

We selectively amplified full-length cDNAs from total RNA of the $CD133^+$ blasts of the patient and ligated them into the retroviral vector pMX. We obtained a total of 8.7×10^6 colony-forming units of independent plasmid clones. To evaluate the quality of the library, we randomly selected 40 clones and examined the incorporated cDNAs. Thirty-four (85%) of the 40 clones contained inserts with an average size of 1.63 kbp, suggesting that the library was of sufficient complexity for the present study.

A recombinant retroviral library was generated from the plasmids and used to infect 32D-mCAT cells. A total of 2.75×10^6 independent retroviral clones was assayed with a total of 8×10^6 32D-mCAT cells. The cells were subsequently cultured for 16 days in the presence of G-CSF. Whereas proliferating 32D-mCAT cells manifested a high nucleus-to-cytoplasm ratio and basophilic cytoplasm (Fig. 1A, +IL-3), incubation with G-CSF for 16 days induced granulocytic differentiation in most of the cells (Fig. 1A, clone #1). Among the 32D-mCAT cells infected with retroviruses, however, we detected four cell clones (clones #2 to #5) that grew rapidly in the presence of G-CSF (Fig. 1A, data not shown). The cells in each clone were expanded and examined for their growth properties in the presence of G-CSF. The proliferation rate (Fig. 1B) as well as differentiation ability (Fig. 1A, data not shown) varied markedly among the clones. Genomic DNA was isolated from each clone and subjected to PCR in order to amplify the retroviral inserts. Nucleotide sequencing of the inserts revealed that those isolated from clones #2 and #3 were derived from human *JAK3* (GenBank accession number, NM_000215), whereas that from clone #5 was derived from *CSF3R* (accession number, NM_000760), which encodes the G-CSF receptor. PCR failed to detect a cDNA insert in cells of clone #4.

3.3. Identification of constitutively active JAK3(V674A)

JAK3 is a cytoplasmic tyrosine kinase that plays an important role in T cell development and function [14,15]. Given that constitutively active forms of JAK3 have not been detected previously, we focused on the *JAK3* cDNAs identified by retroviral expression screening. To confirm the transforming activity of the cDNAs, we ligated them into the pMX-ires-neo plasmid for generation of recombinant retroviruses. The retroviruses were used to infect 32D-mCAT cells, which were then subjected to selection with G418. The cells expressing the *JAK3* cDNA derived from clone #2 grew exponentially in the presence of G-CSF or IL-3 (Fig. 2A).

These cells also grew in the absence of cytokine, showing that the JAK3 derived from clone #2 abrogates cytokine dependency of 32Dcl3 cells. Similar results were obtained with the *JAK3* cDNA isolated from clone #3, and both of the *JAK3* cDNAs (from clones #2 and #3) were identical on both ends (5′- and 3′-termini) (data not shown). We thus chose the *JAK3* cDNA from clone #2 for further analysis.

Complete sequencing of the *JAK3* cDNA from clone #2 revealed that it spans nucleotide positions 20–3460 of the published human *JAK3* cDNA (NM_000215). However, four nucleotides within the isolated cDNA do not match those in the published sequence, and three of these unmatched nucleotides affect the corresponding amino acid sequence of JAK3. A nucleotide change at position 1446 (C to T) results in the replacement of a histidine residue at codon 463 with tyrosine; a nucleotide change at position 2080 (T to C) results in the replacement of valine at codon 674 with alanine; and a nucleotide change at position 3187 (A to G) results in the replacement of aspartic acid at codon 1043 with glycine (Fig. 2B).

To examine the role of these amino acid changes in the transforming activity of the *JAK3* cDNA derived from clone #2, we introduced each mutation individually into the wild-type human *JAK3* cDNA. Each mutant cDNA was ligated into pMX-ires-neo for the production of recombinant retroviruses, which were then used to infect 32D-mCAT cells. Culture of the infected cells in the absence of cytokine revealed that those expressing JAK3(V674A), but not those expressing JAK3(H463Y) or JAK3(D1043G), grew exponentially (Fig. 2C), demonstrating that the amino acid change from valine to alanine at position 674 is responsible for the transforming activity of *JAK3* cDNA from clone #2.

Given that amino acid position 674 is within the JH2 region of JAK3, which negatively regulates the kinase activity of the enzyme [16], the V674A mutation might be expected to affect kinase activity and thereby to confer cytokine independence on 32D-mCAT cell growth. To examine this possibility, we immunoprecipitated JAK3 from 32D-mCAT cells expressing the wild-type or V674A form of the enzyme and subjected the precipitates to immunoblot analysis with an antibody to phosphotyrosine. JAK3(V674A) was constitutively phosphorylated irrespective of cell stimulation with IL-3, and the extent of its phosphorylation was markedly greater than that of wild-type JAK3 (Fig. 3A). Immunoblot analysis of the same samples with antibodies to JAK3 confirmed that the differences in the levels of JAK3 tyrosine phosphorylation were not due to differences in the levels of JAK3 expression. To confirm directly that the kinase activity of JAK3(V674A) is increased, we subjected immunoprecipitates of wild-type JAK3 or JAK3(V674A) to an in vitro kinase assay with a synthetic peptide substrate corresponding to the autophosphorylation site within the kinase domain of JAK3. The kinase activity of JAK3(V674A) was indeed found to be markedly greater than that of wild-type JAK3 and was not influenced by cell stimulation with IL-3 (Fig. 3B).

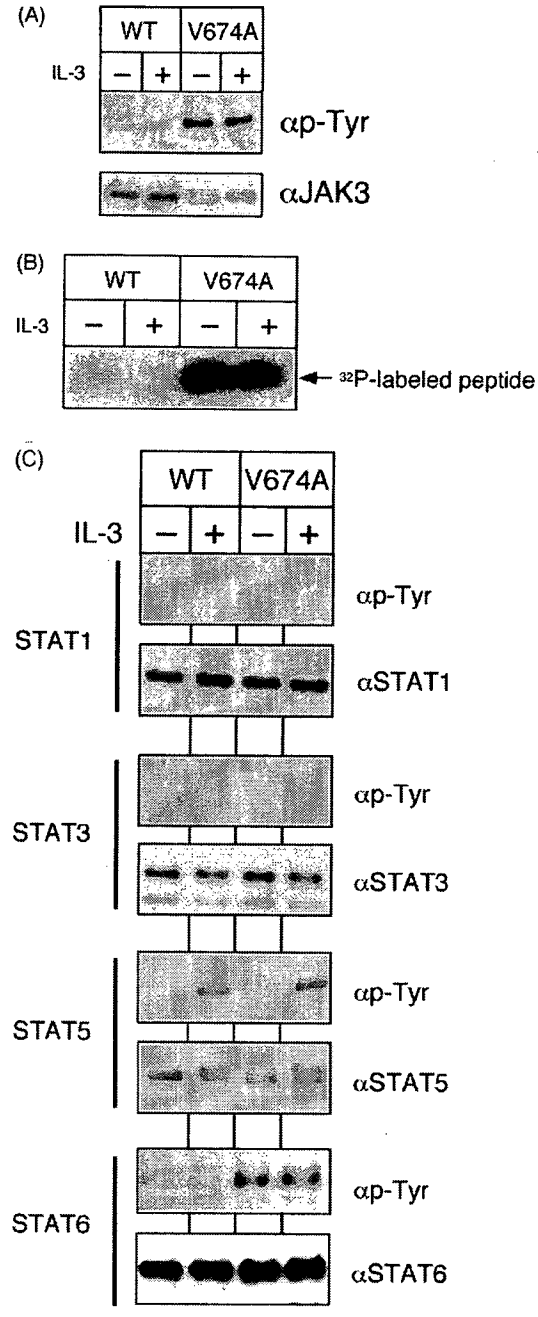

Fig. 3. Constitutive activation of JAK3(V674A). (A) 32D-mCAT cells expressing wild-type JAK3 (WT) or JAK3(V674A) were cultured in the absence of cytokine or FBS for 1 h and then incubated with (+) or without (−) IL-3 for 10 min. Cell lysates were subjected to immunoprecipitation with antibodies to JAK3. Half of the immunoprecipitates were subjected to immunoblot analysis with antibodies to phosphotyrosine (αp-Tyr) or to JAK3 (αJAK3). The experiments were repeated three times, and the data were basically identical. (B) The other half of the precipitates in (A) were subjected to an in vitro kinase assay with a synthetic peptide corresponding to the autophosphorylation site in the kinase domain of JAK3. (C) Cells treated as in (A) were subjected to immunoprecipitation with antibodies to STAT1, STAT3, STAT5, or STAT6, and the resulting precipitates were subjected to immunoblot analysis with antibodies to phosphotyrosine or to the corresponding STAT protein. The experiments were repeated three times, and the data were basically identical.

Signal transducer and activator of transcription (STAT) proteins are important signal transducers for JAK family members [17], and it is known that JAK3 can activate STAT5 and STAT6 in the context of IL-2- and IL-4-signaling systems, respectively. We thus examined the phosphorylation levels of various STATs in 32D-mCAT cells expressing wild-type JAK3 or JAK3(V674A). STAT1, STAT3, STAT5, and STAT6 were immunoprecipitated from such cells and subjected to immunoblot analysis with the antibody to phosphotyrosine. Only STAT6 manifested a higher level of phosphorylation in 32D-mCAT cells expressing JAK3(V674A) than in those expressing the wild-type kinase (Fig. 3C). In addition, STAT6 phosphorylation in the cells expressing JAK3(V674A) was not affected by IL-3, consistent with the lack of effect of this cytokine on the tyrosine phosphorylation of JAK3(V674A). Although STAT5 underwent tyrosine phosphorylation in response to IL-3 stimulation, the magnitude of this effect did not differ between cells expressing wild-type or V674A forms of JAK3 (Fig. 3C).

Finally, to examine whether the nucleotide substitution responsible for the V674A mutation of JAK3 was detectable in the RNA isolated from the AML patient and used for library construction, we amplified a region of *JAK3* cDNA surrounding codon 674 from the unamplified cDNA of the patient and determined its nucleotide sequence. Extensive sequencing failed to detect this nucleotide change, indicating that the activating mutation was artificially generated in the initial PCR step of library construction.

4. Discussion

We have constructed a retroviral cDNA expression library from a patient with AML. The complexity of the library was sufficient to represent most of the transcriptome of the $CD133^+$ cells from which it was prepared. With the use of this retroviral library, we searched for cDNAs whose expression conferred on 32Dcl3 cells the ability to continue proliferating in the presence of G-CSF. Such cDNAs would be expected either (1) to inhibit the differentiation-inducing activity of G-CSF while supporting its mitogenic activity or (2) to possess marked transforming activity. Our screening resulted in the isolation of a cDNA for a constitutively active form of JAK3, which likely conforms to this second possibility.

An activating mutation of *JAK2* was recently identified in blood cells of individuals with myeloproliferative disorders (MPDs) with the exception of chronic myeloid leukemia [18–21]. This JAK2 mutant, JAK2(V617F), possesses an increased kinase activity that is essential for the erythropoietin-independent growth of MPD cell colonies. These observations indicate that some MPD cases are caused by a constitutively active form of JAK2, and that reclassification of MPDs on the basis of transforming events may be required. Amino acid position 617 is localized within the JH2 (pseudokinase) domain of JAK2, which is thought to function as a negative regulatory region for kinase activity [22]. Disruption of JH2 function through point mutations (in addition to that associated with MPDs) has thus been shown to activate JAK2 [23]. The amino acid (valine-674) whose mutation is responsible for the constitutively active form of JAK3 identified in the present study is also located in the JH2 domain of this kinase, emphasizing the importance of this domain in the negative regulation of JAK3. As far as we are aware, oncogenic forms of JAK3 have not previously been described, and mutation of amino acid residues corresponding to valine-674 of JAK3 have not been described for the other JAK family members.

Our inability to detect the corresponding mutation at nucleotide position 2080 in the original *JAK3* cDNA prepared from the AML patient indicates that it was introduced into the retroviral library by the initial PCR step of library construction. This mutation thus does not appear to be related to the molecular pathogenesis of AML in the patient. Similarly, we failed to detect the other two mutations (C1446T and A3187G of *JAK3*) in the original RNA of the patient (data not shown). This observation suggests that care is required to minimize the introduction of artificial mutations in the PCR step for this type of screening and that any mutations identified by such screening should be confirmed by analysis of the original sample. Nevertheless, our identification of an active form of JAK3 demonstrates the potential efficacy of the retroviral library – 32Dcl3 cell screening system for the detection of hematopoietic oncogenes as well as provides insight into the regulation of JAK3 activity.

Acknowledgments

We are thankful to Drs. James N. Ihle and Evan Parganas for their suggestions on the experiments and for their critical reading of this manuscript. 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.

Contributions. Retrovirus library construction and screening was undertaken by Y.L.C., T.W., S.-I.F., M.S., H.W., and H.H. Construction of mutant JAK3 cDNAs was undertaken by K.K. and M.E., and protein analysis was undertaken by R.K., S.T. and Y.Y. The project was designed by H.M.

References

[1] Kalantry S, Delva L, Gaboli M, Gandini D, Giorgio M, Hawe N, et al. Gene rearrangements in the molecular pathogenesis of acute promyelocytic leukemia. J Cell Physiol 1997;173:288–96.
[2] Moore MA. Converging pathways in leukemogenesis and stem cell self-renewal. Exp Hematol 2005;33:719–37.
[3] Mrozek K, Heerema NA, Bloomfield CD. Cytogenetics in acute leukemia. Blood Rev 2004;18:115–36.

[4] Frohling S, Scholl C, Gilliland DG, Levine RL. Genetics of myeloid malignancies: pathogenetic and clinical implications. J Clin Oncol 2005;23:6285–95.
[5] Aaronson SA. Growth factors and cancer. Science 1991;254: 1146–53.
[6] 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;80:2931–5.
[7] 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;29:943–9.
[8] Albritton LM, Tseng L, Scadden D, Cunningham JM. A putative murine ecotropic retrovirus receptor gene encodes a multiple membrane-spanning protein and confers susceptibility to virus infection. Cell 1989;57:659–66.
[9] Pear WS, Nolan GP, Scott ML, Baltimore D. Production of high-titer helper-free retroviruses by transient transfection. Proc Natl Acad Sci USA 1993;90:8392–6.
[10] Ohmine K, Ota J, Ueda M, Ueno S-I, Yoshida K, Yamashita Y, et al. Characterization of stage progression in chronic myeloid leukemia by DNA microarray with purified hematopoietic stem cells. Oncogene 2001;20:8249–57.
[11] Onishi M, Kinoshita S, Morikawa Y, Shibuya A, Phillips J, Lanier LL, et al. Applications of retrovirus-mediated expression cloning. Exp Hematol 1996;24:324–9.
[12] Mavilio F, Kreider BL, Valtieri M, Naso G, Shirsat N, Venturelli D, et al. Alteration of growth and differentiation factors response by Kirsten and Harvey sarcoma viruses in the IL-3-dependent murine hematopoietic cell line 32D C13(G). Oncogene 1989;4:301–8.
[13] Rovera G, Valtieri M, Mavilio F, Reddy EP. Effect of Abelson murine leukemia virus on granulocytic differentiation and interleukin-3 dependence of a murine progenitor cell line. Oncogene 1987;1:29–35.
[14] Witthuhn BA, Silvennoinen O, Miura O, Lai KS, Cwik C, Liu ET, et al. Involvement of the Jak-3 Janus kinase in signalling by interleukins 2 and 4 in lymphoid and myeloid cells. Nature 1994;370:153–7.
[15] Macchi P, Villa A, Giliani S, Sacco MG, Frattini A, Porta F, et al. Mutations of Jak-3 gene in patients with autosomal severe combined immune deficiency (SCID). Nature 1995;377:65–8.
[16] Saharinen P, Silvennoinen O. The pseudokinase domain is required for suppression of basal activity of Jak2 and Jak3 tyrosine kinases and for cytokine-inducible activation of signal transduction. J Biol Chem 2002;277:47954–63.
[17] Ihle JN, Nosaka T, Thierfelder W, Quelle FW, Shimoda K. Jaks and Stats in cytokine signaling. Stem Cells 1997;15(Suppl 1):105–11 [Discussion 12].
[18] Levine RL, Wadleigh M, Cools J, Ebert BL, Wernig G, Huntly BJ, et al. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. Cancer Cell 2005;7:387–97.
[19] Baxter EJ, Scott LM, Campbell PJ, East C, Fourouclas N, Swanton S, et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. Lancet 2005;365:1054–61.
[20] James C, Ugo V, Le Couedic JP, Staerk J, Delhommeau F, Lacout C, et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. Nature 2005;434: 1144–8.
[21] 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;352:1779–90.
[22] Saharinen P, Takaluoma K, Silvennoinen O. Regulation of the Jak2 tyrosine kinase by its pseudokinase domain. Mol Cell Biol 2000;20:3387–95.
[23] Argetsinger LS, Kouadio JL, Steen H, Stensballe A, Jensen ON, Carter-Su C. Autophosphorylation of JAK2 on tyrosines 221 and 570 regulates its activity. Mol Cell Biol 2004;24:4955–67.

Oncogene (2007) 26, 1245–1255

www.nature.com/onc

ONCOGENOMICS

A genomic analysis of adult T-cell leukemia

YL Choi[1], K Tsukasaki[2], MC O'Neill[3], Y Yamada[4], Y Onimaru[2], K Matsumoto[5], J Ohashi[6], Y Yamashita[1], S Tsutsumi[7], R Kaneda[1], S Takada[1], H Aburatani[7], S Kamihira[4], T Nakamura[5], M Tomonaga[2] and H Mano[1,8]

[1]Division of Functional Genomics, Jichi Medical University, Shimotsukeshi, Tochigi, Japan; [2]Atomic Bomb Disease Institute, Nagasaki University Graduate School of Biomedical Science, Nagasaki, Japan; [3]Department of Biological Sciences, University of Maryland, Baltimore, MD, USA; [4]Department of Laboratory Medicine, Nagasaki University Graduate School of Biomedical Science, Nagasaki, Japan; [5]Division of Molecular Regenerative Medicine, Osaka University Graduate School of Medicine, Osaka, Japan; [6]Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; [7]Research Center for Advanced Science and Technology, University of Tokyo, Tokyo, Japan and [8]CREST, Japan Science and Technology Agency, Saitama, Japan

Adult T-cell leukemia (ATL) is an intractable malignancy of CD4$^+$ T cells that is etiologically associated with infection by human T-cell leukemia virus-type I. Most individuals in the chronic stage of ATL eventually undergo progression to a highly aggressive acute stage. To clarify the mechanism responsible for this stage progression, we isolated CD4$^+$ cells from individuals in the chronic ($n=19$) or acute ($n=22$) stages of ATL and subjected them to profiling of gene expression with DNA microarrays containing >44 000 probe sets. Changes in chromosome copy number were also examined for 24 cell specimens with the use of microarrays harboring ~50 000 probe sets. Stage-dependent changes in gene expression profile and chromosome copy number were apparent. Furthermore, expression of the gene for MET, a receptor tyrosine kinase for hepatocyte growth factor (HGF), was shown to be specific to the acute stage of ATL, and the plasma concentration of HGF was increased in individuals in either the acute or chronic stage. HGF induced proliferation of a MET-positive ATL cell line, and this effect was blocked by antibodies to HGF. The HGF-MET signaling pathway is thus a potential therapeutic target for ATL.
Oncogene (2007) **26,** 1245–1255. doi:10.1038/sj.onc.1209898; published online 14 August 2006

Keywords: adult T-cell leukemia; DNA microarray; MET; artificial neural network

Introduction

Adult T-cell leukemia (ATL) is an intractable malignancy of CD4$^+$ T cells that is etiologically associated with infection by human T-cell leukemia virus-type I (HTLV-I) (Uchiyama *et al.*, 1977; Poiesz *et al.*, 1980; Yoshida *et al.*, 1982). Virally encoded proteins such as Tax trigger polyclonal growth of T cells in infected individuals, and there are an estimated 15–20 million such carriers worldwide (Edlich *et al.*, 2000). After a latency period of decades, a small proportion of carriers (~2%) develop ATL. Many ATL patients initially manifest only monoclonal (or oligoclonal) growth of leukemic clones without apparent clinical symptoms, a condition referred to as the chronic or smoldering stages (Shimoyama, 1991). Most individuals in the chronic stage, however, eventually undergo progression to a highly aggressive acute stage (Tajima, 1990). Given that the prognosis of individuals at the acute stage remains very poor, it is important to clarify the molecular mechanism that underlies stage progression.

Homozygous deletion or epigenetic silencing of the gene for the cyclin-dependent kinase inhibitor p16 (Hatta *et al.*, 1995; Yamada *et al.*, 1997; Nosaka *et al.*, 2000) as well as altered expression of other genes related to cell proliferation (Cesarman *et al.*, 1992; Tamiya *et al.*, 1998) have been detected in ATL cells at the acute stage. However, such genetic or epigenetic changes may be infrequent (Matsuoka, 2003), and the transforming events responsible for chronic to acute stage progression remain largely unknown.

DNA microarray analysis allows simultaneous comparison of the expression intensities of tens of thousands of genes. Such analysis of the transcriptomes of ATL cells at the chronic and acute stages might thus be expected to provide insight into the mechanism of stage progression in this disease. With the use of this approach, Sasaki *et al.* (2005) recently compared transcriptomes between normal CD4$^+$ T cells ($n=5$) and mononuclear cells (MNCs) isolated from individuals in the acute stage of ATL ($n=8$). Tsukasaki *et al.* (2004) also compared transcriptomes between MNCs from patients in the chronic or acute stages of ATL ($n=4$ for each). However, the significance of these data may be limited by the small number of study subjects and by the use of unfractionated MNCs that contain various proportions of non-ATL cells.

Correspondence: Dr H Mano, Division of Functional Genomics, Jichi Medical University, 3311-1 Yakushiji, Shimotsukeshi, Tochigi 329-0498, Japan.
E-mail: hmano@jichi.ac.jp
Received 20 March 2006; revised 10 July 2006; accepted 11 July 2006; published online 14 August 2006

In addition to changes in gene expression, ATL cells frequently manifest various karyotype anomalies. Comparative genomic hybridization (CGH) has thus revealed recurrent gains in chromosomes 2p, 3p, 7q and 14q as well as losses in 6q in ATL cells (Ariyama *et al.*, 1999; Tsukasaki *et al.*, 2001). However, CGH or its successor, bacterial artificial chromosome (BAC) array-based CGH, is able to analyse chromosome copy number alterations (CNAs) at a resolution of only several hundred kilobase pairs (Lockwood *et al.*, 2005). High-density oligonucleotide microarrays originally designed for genotyping of single nucleotide polymorphisms (SNPs) have recently been adapted for CNA analysis (Lin *et al.*, 2004; Nannya *et al.*, 2005). With this approach, chromosome copy number is inferred from the signal intensity of SNP probe sets distributed throughout the human genome. For instance, with Affymetrix GeneChip Mapping 100K arrays developed for genotyping of $\sim$100 000 SNPs, it is possible to determine CNAs at a mean resolution of 23.6 kbp, which is substantially greater than that achievable with BAC array-based technologies.

With both microarray-based gene expression profiling and SNP array-based CNA profiling, we have now performed a comprehensive genomic analysis of ATL in order to investigate the mechanism of stage progression from chronic to acute. Given that the $CD4^+CD8^-$ fraction of peripheral blood (PB) cells of individuals with chronic or acute ATL is composed predominantly of ATL cells, we purified this fraction from ATL patients. We then subjected the isolated cells to gene expression profiling with microarrays containing $>$44 000 probe sets and to CNA analysis with microarrays harboring $\sim$50 000 probe sets. The gene expression data indicate that the transcriptomes for the chronic and acute stages of ATL are distinct, and the CNA data reveal frequent amplification or deletion of genomic fragments of various sizes in each ATL stage.

Results

Transcriptomes of ATL cells

To characterize the transcriptomes of ATL cells, we purified $CD4^+$ cells from PB of ATL patients at either the chronic ($n=19$) or the acute ($n=22$) stage. The clinical characteristics of the patients are summarized in Supplementary Table 1. The $CD4^+$ fraction was also purified from healthy volunteers ($n=3$) and was either activated with phytohemagglutinin (PHA) or not.

A simple, one-step column purification with antibodies to CD4 yielded a highly pure $CD3^+CD4^+$ T-cell fraction. For example, whereas the $CD3^+CD4^+$ fraction constituted only 29.1% of PB MNCs of one healthy individual, it constituted 98.8% of the corresponding column eluate (Figure 1a). Similarly, $CD3^+CD4^+$ cells constituted 25.7% of MNCs from one ATL patient at the acute stage, but accounted for 97.5% of cells in the corresponding column eluate (data not shown).

All of the ATL and normal $CD4^+$ cell specimens were then subjected to expression profiling with $\sim$44 000 probe sets (corresponding to $\sim$33 000 transcripts) on Affymetrix HGU133 microarrays. To eliminate from the analysis genes that were transcriptionally silent in the ATL specimens, we first selected probe sets that received the 'Present' call by Microarray Suite 5.0 software (Affymetrix) in at least 30% ($n=13$) of the ATL samples. A total of 15 121 probe sets fulfilled this criterion. On the basis of the similarity of the expression profiles for these probe sets, all 47 samples were subjected to hierarchical two-way clustering (Alon *et al.*, 1999), yielding a dendrogram of the subjects (Figure 1b). All six normal samples, irrespective of PHA stimulation, formed a distinct branch separated from the ATL specimens, indicating that the overall gene expression profiles differed between normal and transformed T cells. However, samples corresponding to patients with chronic or acute ATL were not clearly separated from each other in this tree.

To compare the transcriptomes of ATL cells between chronic and acute stages, we conducted Student's *t*-test on the gene expression intensity for the 15 121 probe sets with the Benjamini and Hochberg false discovery rate (Reiner *et al.*, 2003) of 0.01, leading to the isolation of 84 probe sets (data not shown). To enrich probe sets whose expression level was high in at least one of the stages, we adopted another selection window, effect size (absolute difference in mean expression intensity) (Dhanasekaran *et al.*, 2001). We extensively compared the expression level of given probe sets determined by DNA microarray and by quantitative real-time reverse transcription–polymerase chain reaction (RT–PCR). With our normalization procedure (see Materials and methods), expression of genes with an array data of $\geqslant$100 units (U) was almost always detected by real-time RT–PCR (data not shown). Thus, we chose 100 U as the threshold value for the effect size.

A total of 21 probe sets (corresponding to 21 independent genes) whose expression level contrasted the two clinical conditions were finally identified. Hierarchical two-way clustering analysis of the expression profiles of these stage-associated genes revealed that only two gene were preferentially expressed at the chronic stage, whereas the other 19 genes were preferentially expressed in the acute stage (Figure 1c and Supplementary Table 2). Interestingly, the latter gene cluster contains several genes encoding for growth-related proteins, such as nuclear receptor coactivator 3 (NCOA3, GenBank accession no. NM_006534), heat-shock 60-kDa protein 1 (HSPD1, GenBank accession no. NM_002156) and general transcription factor IIIA (GTF3A, GenBank accession no. BE542815).

Gene expression-based prediction of ATL stage

We next attempted to develop a microarray-based class prediction algorithm for ATL. Among several approaches examined, an artificial neural network (ANN) provided the highest accuracy for prediction (O'Neill and Song, 2003). ANNs are computer-based

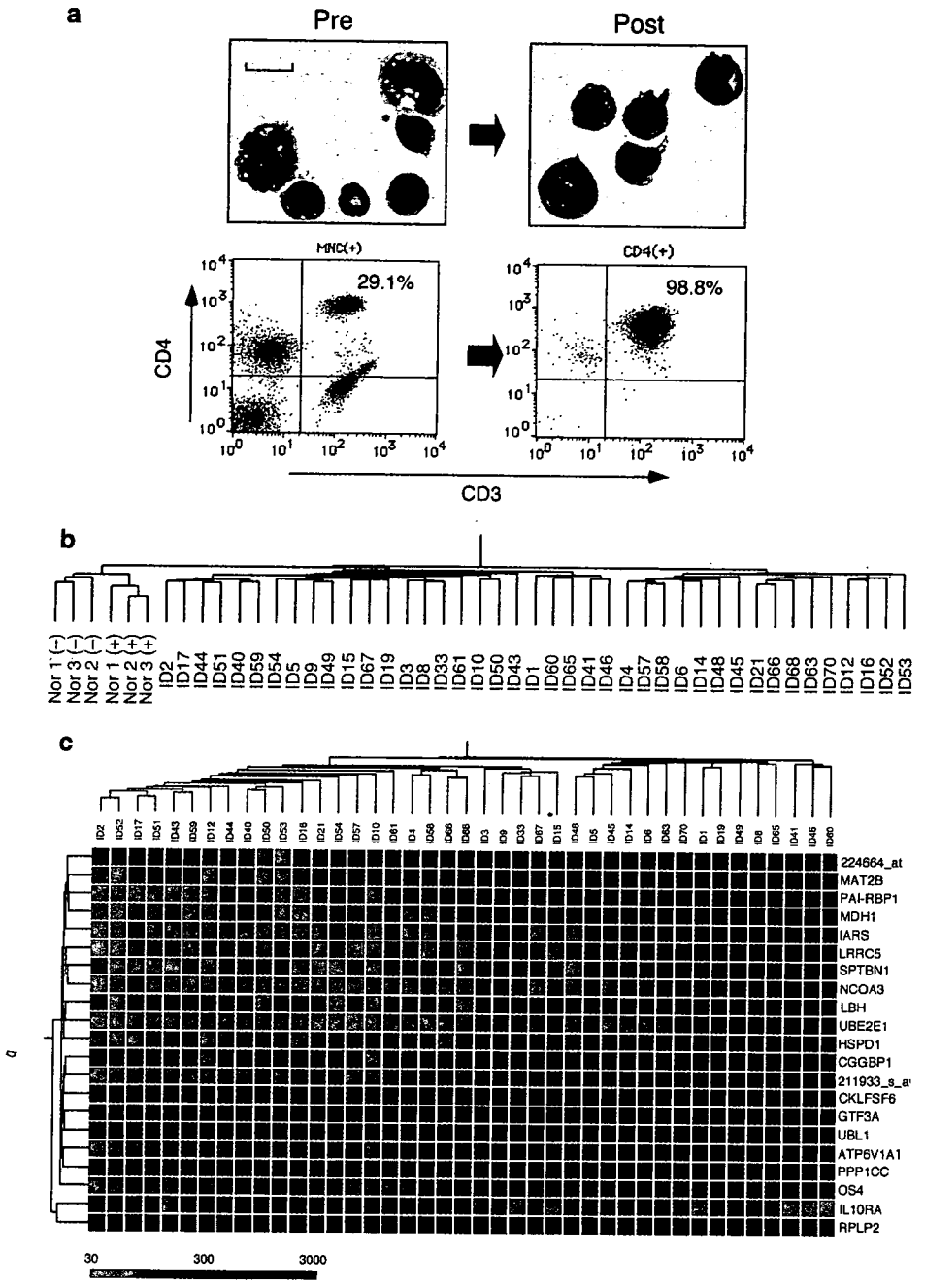

Figure 1 Purification and gene expression profiling of ATL cells. (**a**) MNCs isolated from the PB of a healthy individual were subjected to staining with Wright–Giemsa solution before (Pre) and after (Post) purification by affinity chromatography with antibodies to CD4 (upper panels). Scale bar, 10 μm. The same fractions were also subjected to flow cytometry with antibodies to CD3 and to CD4 (lower panels). The proportion of double-positive cells is indicated. (**b**) Subject tree generated by hierarchical clustering analysis of the expression profiles for 15 121 probe sets. Normal T cells (Nor 1–3) stimulated (+) or not (−) with PHA (8 μg/ml) clustered together, separate from the ATL samples from patients in the chronic (green) or acute (red) stage. (**c**) Subject tree generated by two-way clustering analysis with 21 probe sets that contrasted the two clinical conditions (Student's *t*-test with the Benjamini and Hochberg false discovery rate of 0.01, and effect size of $\geqslant$100 U). Each column corresponds to a separate sample, and each row to a probe set whose expression is color-coded according to the indicated scale. Gene symbols are shown on the right; 224664_at and 211933_s_at are the expressed sequence tag IDs designated by Affymetrix (http://www.affymetrix.com). Annotations and expression intensities for the probe sets are presented in Supplementary Table 2.

algorithms modeled on the structure and behavior of neurons in human brain. Pattern recognition by ANNs is accomplished by training the networks for multiple times in a supervised mode. ANNs adjust continuously their internal weighted connections to reduce the observed errors in matching input to output.

Here, the 15 121 probe sets originally selected in Figure 1b were divided into three nonoverlapping

groups, each of which was used as the input for 10 ANNs (Figure 2a). We performed a 10% crossvalidation rotation with 37 samples, training with 33 samples and testing of the remaining four samples. We then reduced the weight of one input in the first layer (one at a time by 15%), and the network was run again to evaluate the difference in the result from the original output. The same procedure was performed in turn for every input, in order to identify 44 'predictor' genes whose expression markedly influenced the prediction accuracy in each set of ANNs (Figure 2b and Supplementary Table 3). Such predictor set contains only one gene (*UBE2E1*) shared with the stage-associated probe sets shown in Figure 1c. As demonstrated previously, ANN and other approaches (such as *t*-test or clustering analysis) frequently isolate distinct sets of predictor genes (O'Neill and Song, 2003).

Another nine ANNs were then trained and tested with the 44 predictor genes in the same 10% crossvalidation round, yielding one error of prediction for the 37 samples. Finally, the withheld four samples were tested with the trained ANN, resulting in the correct prediction of the class of each. Given that diagnosis of the stage of ATL patients is sometimes problematic, especially when an individual is undergoing stage transition, our analysis offers the possibility of a microarray diagnostic system based on the expression profile of a small number of genes.

Copy number analysis of the ATL genome

To analyse chromosomal gain or loss in ATL cells, we subjected genomic DNA to hybridization with genotyping arrays that represent ~50 000 human SNPs and allow determination of copy number at an average resolution of 47.2 kbp. We first examined whether MNCs and purified $CD4^+$ ATL cells may yield similar CNA profiles by analyzing genomic DNA from such cell fractions of a single individual (patient ID6) at the acute stage of ATL. Flow cytometry revealed that $CD3^+CD4^+$ T cells constituted 58.9 and 98.0% of MNCs and purified $CD4^+$ cells of this individual, respectively (data not shown).

As shown in Figure 3a, gain of chromosomal content ($\geqslant 3n$) was apparent at 1q, 3q, 5p, 7q, 18q and 21q, whereas loss of genomic content ($\leqslant 1n$) was observed at 2p, 12p, 13q, 14q and 18p. In addition to changes affecting such large chromosomal regions, numerous CNAs too small to be detected by conventional methods were apparent at various positions (hospital karyotyping of MNCs from this patient indicated a karyotype of 46,XY). We also identified many chromosomal regions whose copy number differed between the unfractionated MNCs and purified $CD4^+$ cells (Figure 3a). These data indicate that purification of $CD4^+$ cells increases the sensitivity of copy number measurement.

Among the ATL specimens subjected to gene expression profiling, all those for which $CD4^+$ cells were available for preparation of genomic DNA were analysed for CNAs ($n=24$; 15 specimens for the acute stage, nine specimens for the chronic stage). Assessment of copy number revealed frequent anomalies of various sizes, ranging from amplification of an entire chromosome to small deletions spanning only a few probe sets, in the ATL genome (Figure 3b). The most frequent gain or loss in our data set was a small deletion at 14q11.2, which was identified in 22 of the 24 patients tested; the core deleted region spans five probe sets, encompassing as little as 30 857 bp at the locus of *TRD* (encoding T-cell receptor delta locus) and *TRA* (encoding T-cell receptor alpha constant). These deletions likely reflect genomic rearrangement at the T-cell receptor locus in ATL cells and support the high sensitivity of the method.

Further, a high-grade amplification of genome could be found in a region spanning ~14 Mbp at 3p (nucleotide 10 672 576–24 556 563) among the ATL patients, especially at the acute stage. A chromosome copy number of four in this region was inferred for three patients at the acute stage (ID 3, 15 and 70), and that of three was inferred for seven patients. Interestingly, expression level of the genes mapped on this 3p region was significantly higher in the patients with a chromosome copy number of four compared to those with a copy number of two ($P=0.03$, Student's *t*-test), and marginally higher to those with a copy number of three ($P=0.051$) (data not shown).

To confirm the inferred copy numbers in our data set, we subjected genomic DNA at a locus with marked variation in copy number (chromosome 6, nucleotides 16 651 304–16 651 533) to quantitative real-time PCR analysis. Such analysis of the 24 patients, two healthy volunteers (one male, one female), and a cell line (KK-1) (Imaizumi *et al*., 2003) derived from a patient at the acute stage of ATL revealed that the inferred copy number was highly correlated with DNA content measured by PCR (Figure 3c).

Stage-dependent CNAs

To screen for CNA patterns linked to stage progression in ATL, we applied Student's *t*-test ($P<0.01$) to the obtained data set. Subsequent application of a selection window specifying that at least two contiguous probes show the same CNA pattern led to the isolation of 330 probe sets that corresponded to 3p, 3q, 14q and 19p (Figure 3d). Segmental amplification of chromosome 3 was detected only in the ATL patients at the acute stage, consistent with previous results obtained by CGH analysis (Tsukasaki *et al*., 2001; Oshiro *et al*., 2006).

To examine the effect of gene dosage on mRNA abundance, we analysed our gene expression data set for the expression level of genes assigned to a segment (region #1, nucleotides 114 092 369–119 769 881) of chromosome 3 (Figure 3d and e). The mean expression intensity of genes in this region was significantly greater for the patients with a corresponding gain of DNA content than for those without such a gain ($P=0.00015$, Student's *t*-test). Similarly, the expression level of genes on a segment (region #2; nucleotide 8 782 486–12 322 072) of chromosome 19 was greater in cells with a gain of DNA content in this region than in those

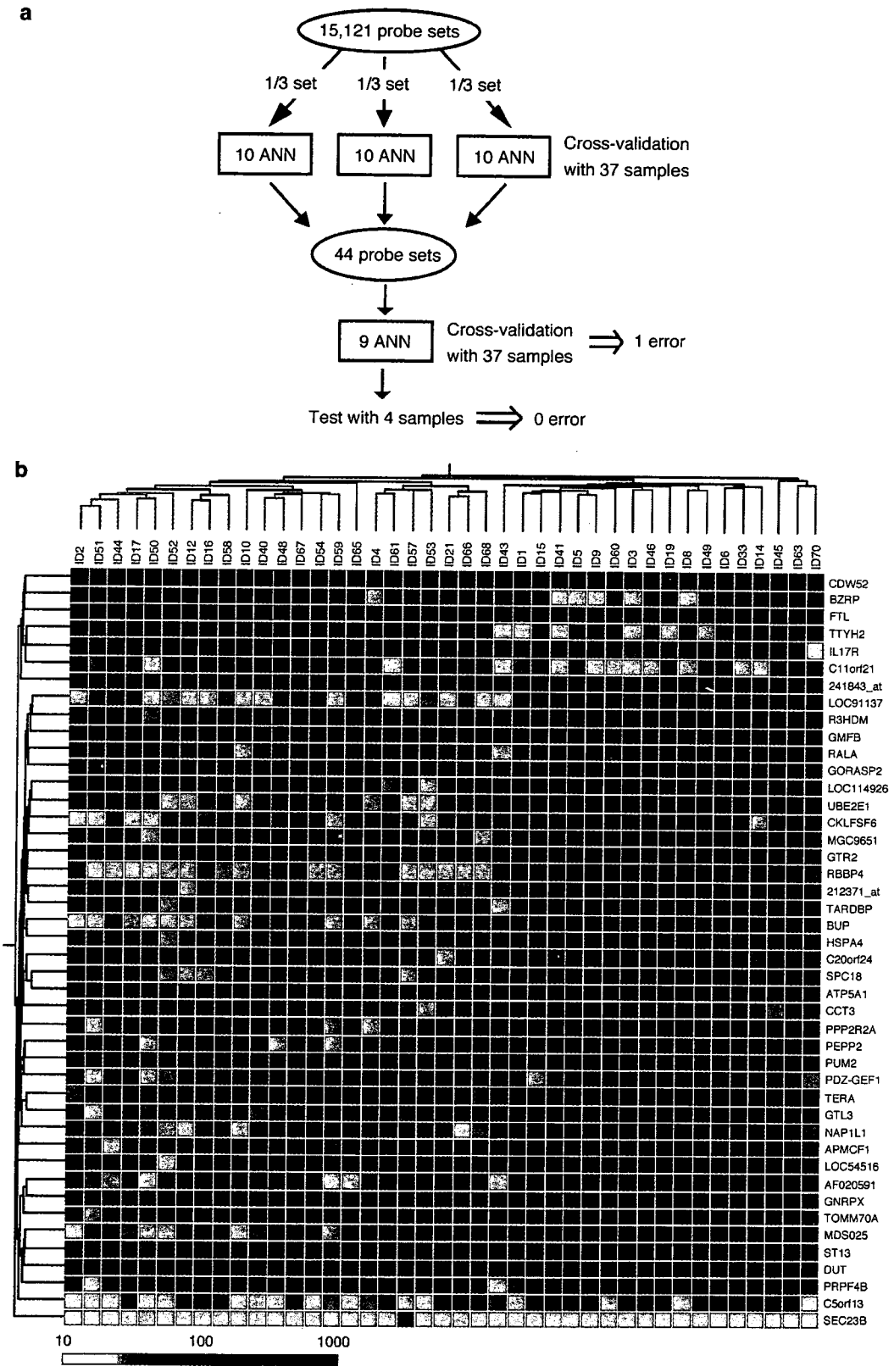

Figure 2 Schematic of the ANN analysis used for class prediction of ATL. (**a**) The 15 121 probe sets originally selected in Figure 1b were divided into three nonoverlapping groups, each of which was used as the input for 10 ANNs. We performed a 10% crossvalidation rotation with 37 samples, training with 33 samples and testing of the remaining four samples. On the basis of the differentiation process with the three sets of 10 ANNs, we selected 44 'predictor' genes whose expression markedly influenced the prediction accuracy in each set of ANNs. (**b**) Subject tree generated by two-way clustering analysis with the 44 predictor genes selected in (**a**) is demonstrated as in Figure 1c. 241843_at and 212371_at are the expressed sequence tag IDs designated by Affymetrix. Annotations and expression intensities for the probe sets are presented in Supplementary Table 3.

without such a gain ($P=0.0357$). These data indicate that gene dosage indeed affects transcript abundance in ATL cells. The large standard deviations apparent in the data shown in Figure 3e, however, suggest that other mechanisms (mediated by transcription factors or epigenetic regulation, for example) have also a large impact on gene expression level.

The hepatocyte growth factor-MET pathway in ATL cells

The long latency period for ATL in HTLV-I carriers suggests that the molecular pathogenesis of ATL and its stage progression might be highly heterogeneous. To identify molecular events that might contribute to transition to the acute stage, we next attempted to isolate 'acute stage-specific genes,' defined by their

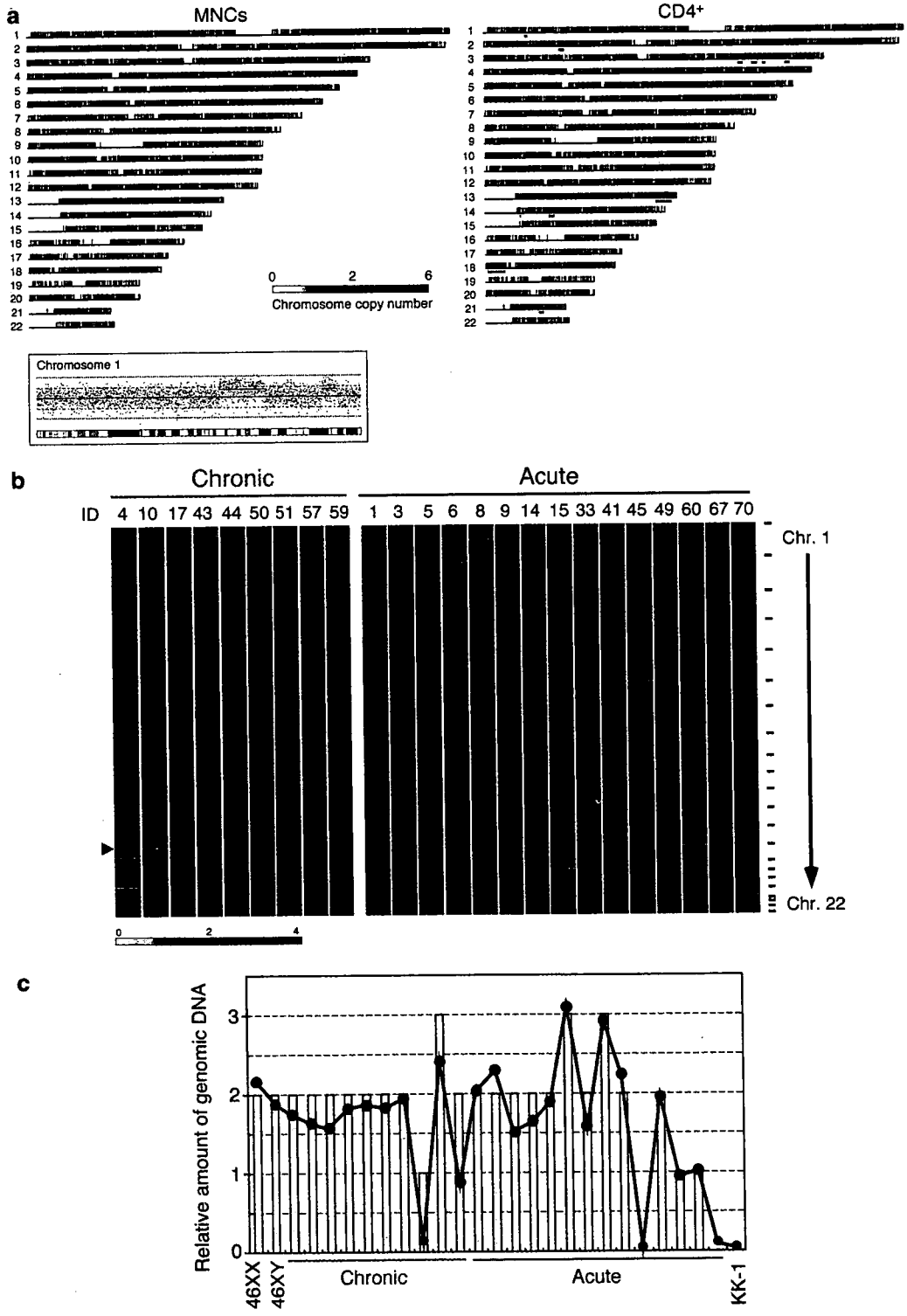

Figure 3 *Continued.*

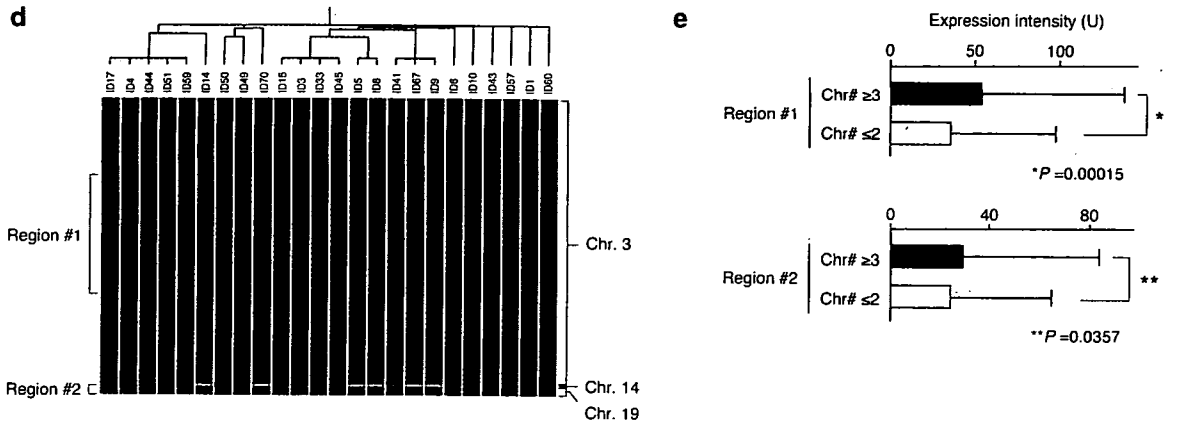

Figure 3 CNA analysis of purified ATL cells. (**a**) Inferred copy number for all SNP sites analysed in MNCs (left) and CD4$^+$ cells (right) isolated from an ATL patient in the acute stage (ID6). Copy number is color-coded according to the indicated scheme. Chromosomal regions with an aberrant copy number detected only in CD4$^+$ cells are indicated by blue underlines. An inset below demonstrates the raw signal ($\log_2$ ratio) for every SNP locus by the array hybridization (red dots) and corresponding inferred copy number (green lines) on chromosome 1 for the MNC sample, along with the chromosome cytobands. (**b**) Inferred copy number for all SNP sites (chromosomes 1–22) in all subjects studied ($n=9$ for chronic stage, $n=15$ for acute stage) shown according to the color scheme at the bottom. SNP sites are ordered by their physical position from top to bottom (shown on the right), and the borders between chromosomes are indicated by small bars. The most frequently deleted region at 14q11.2 is indicated by the arrowhead. (**c**) Inferred copy number (yellow columns) for a region of chromosome 6 (nucleotides 16 651 304–16 651 533) compared with the relative amount (blue line) of genomic DNA corresponding to this region (expressed relative to the amount of *GAPDH* genomic DNA). An ATL cell line (KK-1) and CD4$^+$ cells isolated from a male (46XY) or female (46XX) volunteer were also analysed. (**d**) Subject tree based on inferred copy number of chromosomal regions that showed a statistically significant difference in copy number for at least two consecutive SNP sites between the acute and chronic stages of ATL ($P<0.01$, Student's *t*-test). (**e**) Comparison of expression intensities of the genes assigned to two chromosomal regions (region #1 and region #2) indicated in (**d**) between the subjects with or without a gain in chromosome copy number (Chr#) for the corresponding region. Data are means + s.d. The *P*-values were calculated by Student's *t*-test.

silence (expression level of <10 U) in all normal T cells and chronic ATL specimens and their activation (>100 U) in at least one of the acute-stage samples. We isolated six probe sets that fulfilled such criteria (Figure 4a and Supplementary Table 4).

Among these acute stage-specific genes, we focused on *MET* (GenBank accession no. NM_000245), given that we recently found, in an independent study, that the amount of *MET* mRNA was specifically increased in ATL cells that manifested liver invasion (Imaizumi *et al.*, 2003). *MET* encodes a transmembrane protein tyrosine kinase that is the receptor for hepatocyte growth factor (HGF) (Bottaro *et al.*, 1991). The expression level of *MET* in the study specimens as determined by microarray analysis was highly correlated with that determined by quantitative RT–PCR analysis (Figure 4b), as revealed by a Pearson's correlation coefficient (r) of 0.851 ($P<0.001$). (Also see Supplementary Table 5 for verification of microarray data by RT–PCR.) Flow cytometry revealed that the expression of MET at the cell surface reflected the abundance of the corresponding mRNA in ATL samples (Figure 4c).

The acute stage-specific expression of MET at both the mRNA and protein levels suggested that ATL cells might acquire mitogenic potential as a result of activation of a MET-linked signaling pathway. To examine the possible operation of an HGF-MET autocrine loop, we quantitated HGF mRNA in ATL cells by both microarray and quantitative RT–PCR analyses. No substantial amounts of HGF mRNA were detected in ATL specimens, however (data not shown).

We therefore next measured the plasma concentration of HGF in the study subjects. High levels of HGF were detected in the plasma of ATL patients, especially in that of individuals in the acute stage (Figure 5a), compared with the previously determined values for healthy adults (0.27 ± 0.08 ng/ml, mean $\pm$ s.d.) and some cancer patients (1–2 ng/ml) (Funakoshi and Nakamura, 2003). To test directly whether activation of the HGF-MET signaling pathway is able to promote the proliferation of ATL cells, we examined the MET-positive ATL cell line KK-1. HGF induced both the tyrosine phosphorylation of MET and proliferation in KK-1 cells (Figure 5b and c). The addition of antibodies to HGF (Montesano *et al.*, 1991) could abolish both effects.

Discussion

We have analysed gene expression and CNA profiles in leukemic cell-enriched fractions of individuals with ATL. We found that both types of profile differ markedly between the chronic and acute stages of ATL, and that the level of gene expression is influenced by the copy number of genomic DNA in ATL cells. Although ATL cells have previously been shown to manifest multiple genomic gains or losses (Ariyama *et al.*, 1999; Tsukasaki *et al.*, 2001), our data have revealed that the ATL genome is more unstable than has been appreciated. Similar complex CNAs have been identified by SNP array-based methods for other types

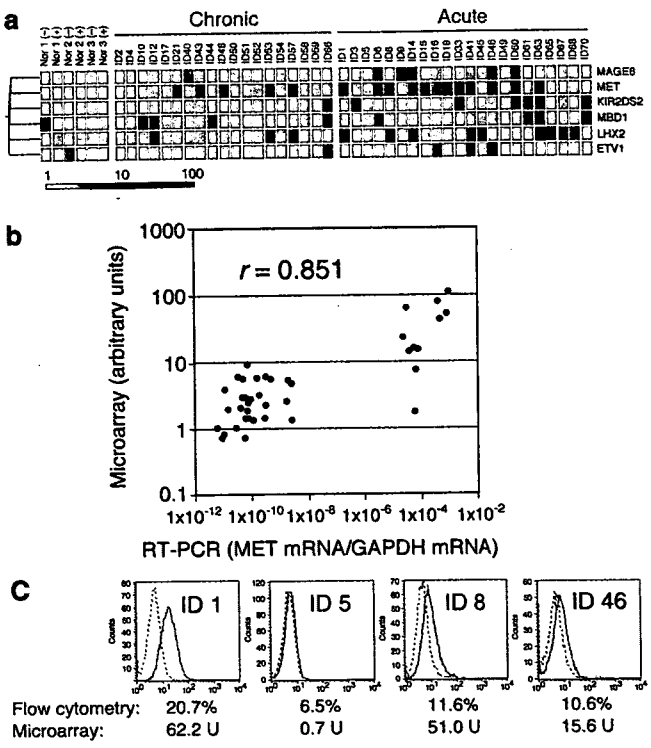

Figure 4 Acute stage-specific expression of *MET* in ATL. (**a**) Expression profiles of the six acute stage-specific probe sets. The expression level of each probe set is colored according to the indicated scale. Annotations and expression intensities for the probe sets are presented in Supplementary Table 4. (**b**) Comparison of the abundance of *MET* mRNA in the study specimens as determined by microarray and RT–PCR analyses. For the latter, the amount of *MET* mRNA is expressed relative to that of *GAPDH* mRNA. Pearson's correlation coefficient (*r*) for the comparison is indicated. (**c**) Cell surface expression of MET was examined by flow cytometry in four ATL samples corresponding to the acute stage. The solid and dashed traces were obtained with antibodies to MET and control antibodies, respectively. The proportion of MET$^+$ cells determined by flow cytometry is indicated together with the corresponding amount of *MET* mRNA determined by microarray analysis.

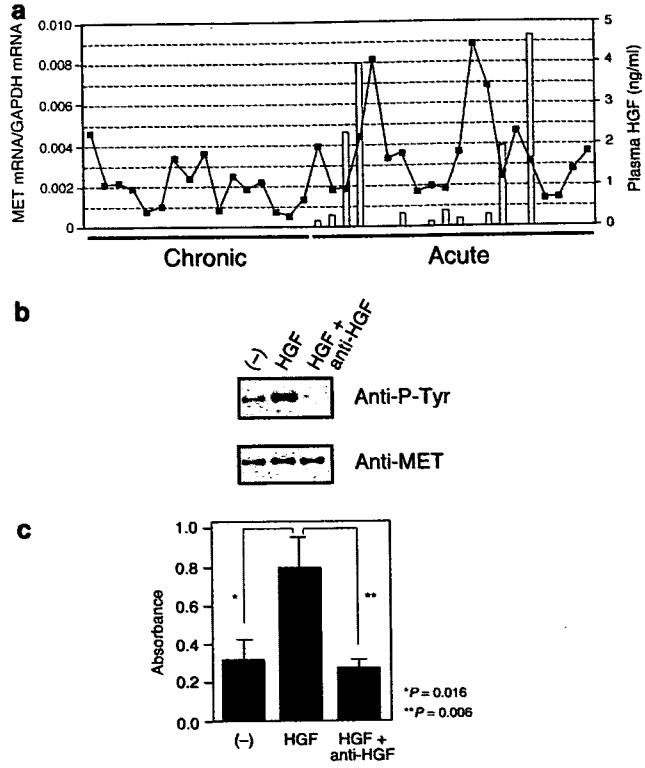

Figure 5 HGF-MET activity induces proliferation in ATL cells. (**a**) Comparison of the plasma concentration of HGF in ATL patients (as measured by enzyme-linked immunosorbent assay) with the corresponding amount of *MET* mRNA in leukemic cells (as determined by RT–PCR). (**b**) Tyrosine phosphorylation of MET in KK-1 cells incubated for 10 min in the absence (–) or presence of human HGF (50 ng/ml), alone or together with antibodies to HGF (10 μg/ml), was examined by immunoblot analysis of cell lysates with antibodies to phosphotyrosine (Anti-P-Tyr). The blot was also probed with antibodies to MET. (**c**) The proliferation of KK-1 cells (3×10^4) incubated for 1 day in the absence (–) or presence of HGF (50 ng/ml), alone or together with antibodies to HGF (10 μg/ml), was evaluated by the MTS assay. Data are expressed in absorbance units and are means + s.d. of triplicates from a representative experiment. The *P*-values for the indicated comparisons were determined by Student's *t*-test.

of cancer (Zhao *et al.*, 2005). We detected 3386.9 $\pm$ 2254.0 (mean $\pm$ s.d.) and 4678.5 $\pm$ 2874.6 SNP sites showing $\geqslant 3n$ ploidy as well as 1039.9 $\pm$ 2310.0 and 927.1 $\pm$ 1137.9 SNP sites with $\leqslant 1n$ ploidy in samples corresponding to the chronic and acute stages of ATL, respectively. The numbers of probe sets showing gain or loss of DNA content did not differ significantly ($P > 0.05$) between the chronic and acute stages of ATL. Given the large numbers of probe sets with an aberrant DNA content in the ATL genome, a large-scale study will likely be required to pinpoint the *bona fide* disease-dependent or stage-dependent CNAs.

Recently, with the use of BAC array-based CGH (with a mean resolution of 1.3 Mbp), Oshiro *et al.* (2006) have compared CNA of MNCs for 17 patients at the acute stage of ATL to that of lymph nodes for 42 cases with the lymphoma type of ATL. The recurrent gain of chromosomes was found at 3/3p among individuals with the acute stage of ATL, which is in good agreement with our results.

We found that an increased plasma concentration of HGF coexists with an increased expression of MET in ATL cells from some individuals at the acute stage of the disease. Together with the demonstrated mitogenic effect of HGF in ATL cells, these data suggest a novel scenario for stage progression in ATL. The mechanism responsible for the increased plasma level of HGF in ATL patients is unclear. Given that ATL is a malignancy of activated mature T cells, the leukemic cells secrete a variety of cytokines, including tumor necrosis factor-α and interleukin-1β (Wano *et al.*, 1987; Yamada *et al.*, 1996). Both of these cytokines are potent inducers of HGF expression in fibroblasts (Matsumoto *et al.*, 1992; Tamura *et al.*, 1993), suggesting that ATL cells may indirectly increase the plasma HGF level through secretion of these cytokines and consequent activation of fibroblasts.

Our data indicate that the plasma concentration of HGF in ATL patients increases before the onset of