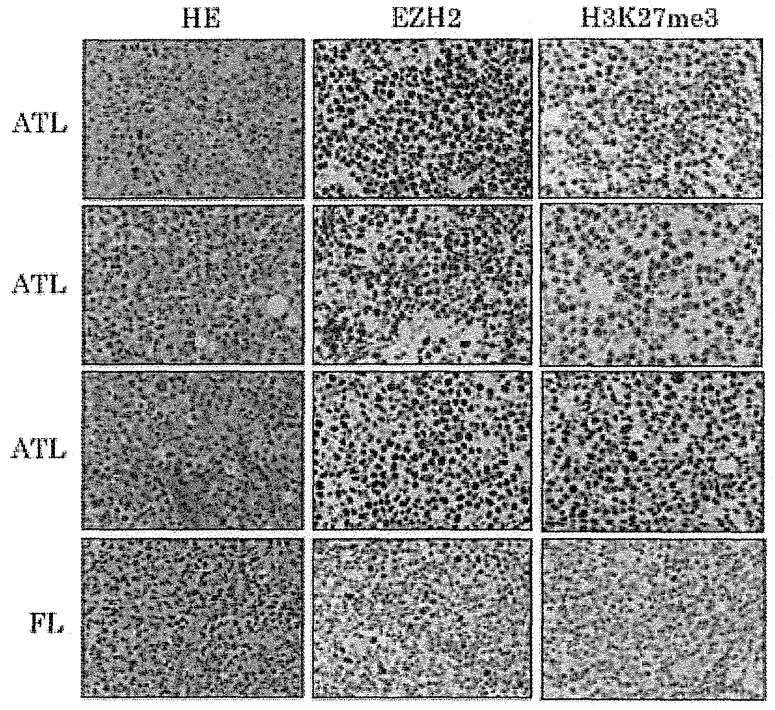
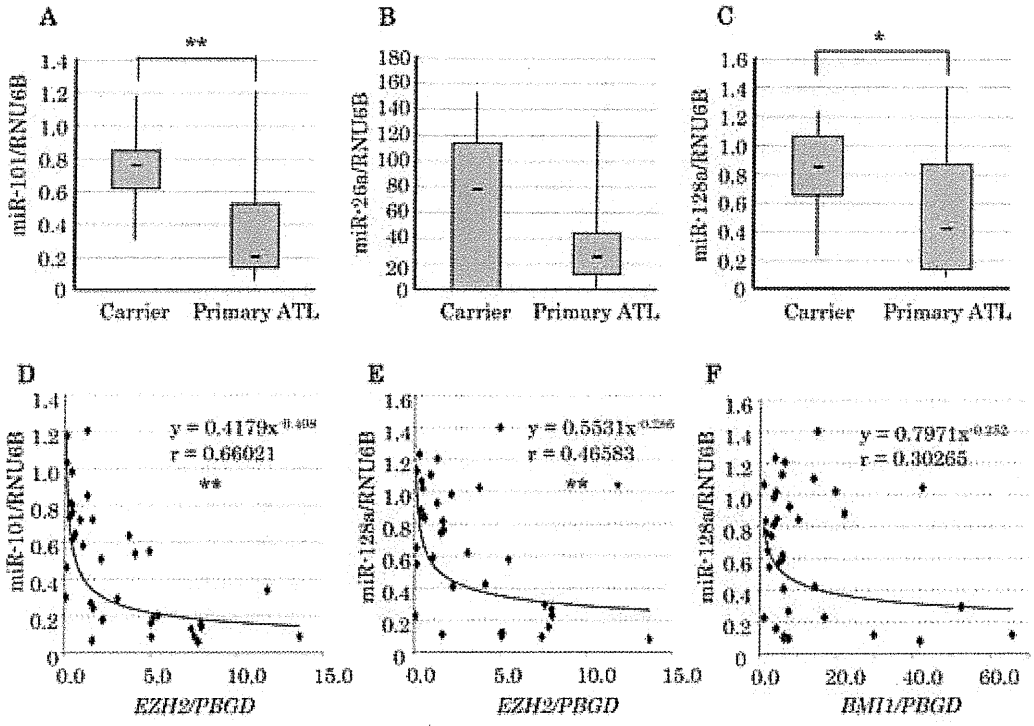
DOI: 10.3324/haematol.2010.028605

Figure 4



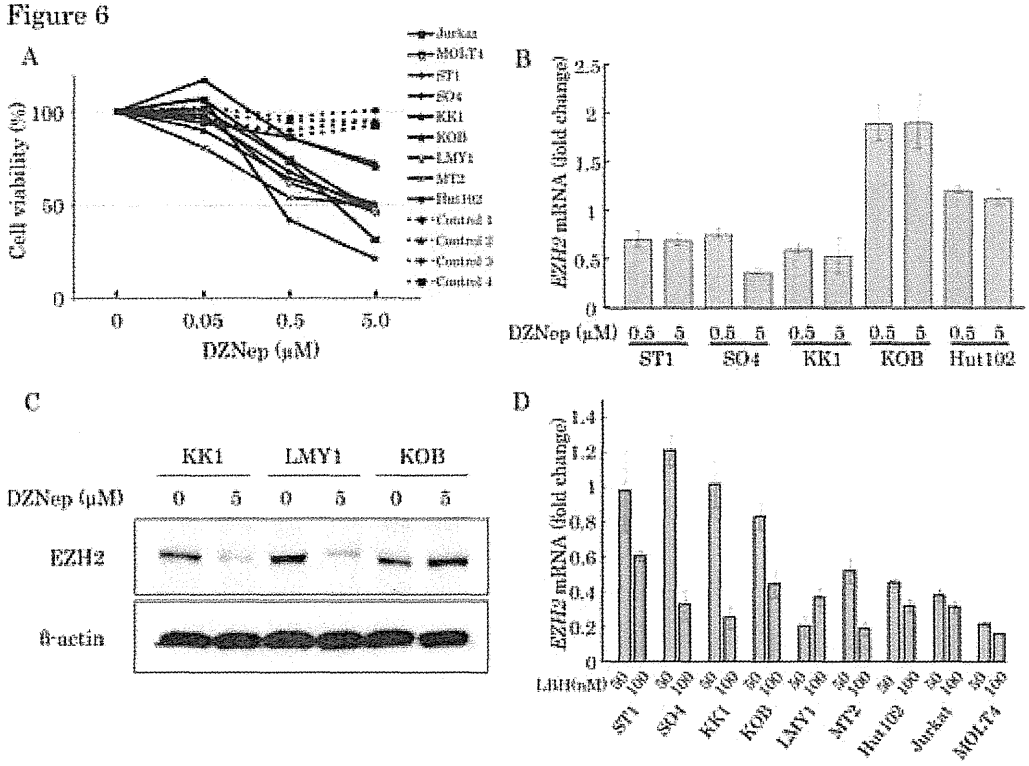
DOI: 10.3324/haematol.2010.028605

Figure 5



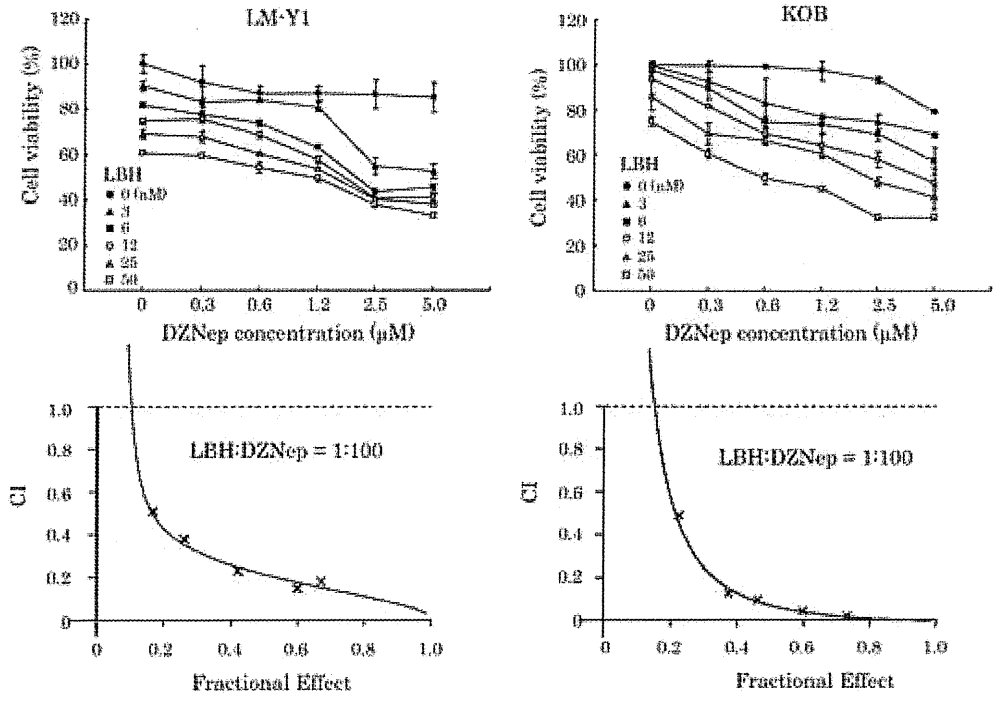
DOI: 10.3324/haematol.2010.028605

Figure 6



DOI: 10.3324/haematol.2010.028605

Figure 6 F



DOI: 10.3324/haematol.2010.028605

Supplementary Appendix

Figure legends for Supplementary Figures

Supplementary Figure 1. RYBP protein expression. Western blot analysis for RYBP protein was performed on primary ATL cells and cells from healthy adults. Most primary ATL samples showed a clear band for RYBP. In contrast, cells from healthy adults lacked the band.

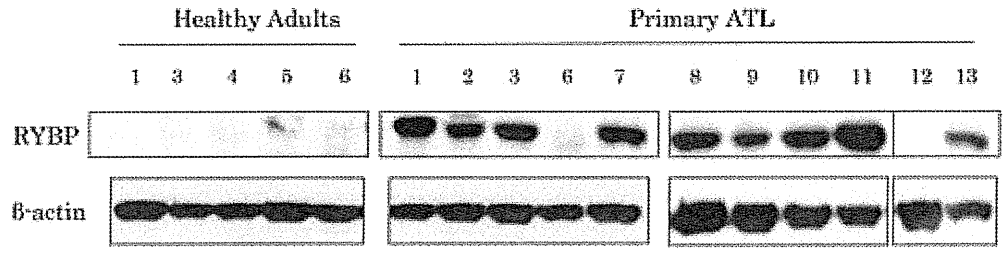
Supplementary Figure 2. Quantitative genomic PCR for miR-101. PCR was performed in two loci, miR-101-1 (chromosome 1p31) and miR-101-2 (chromosome 9p24), in 10 primary ATL samples and cells from 10 HTLV-1 carriers as a control. Both loci were preserved in ATL cells, refuting the possibility that downregulation of miR-101 is caused by genomic loss of the gene.

Supplementary Figure 3. Analysis of 3'-UTR sequence of EZH2 to predict potential target sites for miRNA. In addition to the target sites for miR-101 and miR-26a, there is also a potential target site for miR-128a in the 3'-UTR of EZH2 near one of the miR-101 target sites.

DOI: 10.3324/haematol.2010.028605

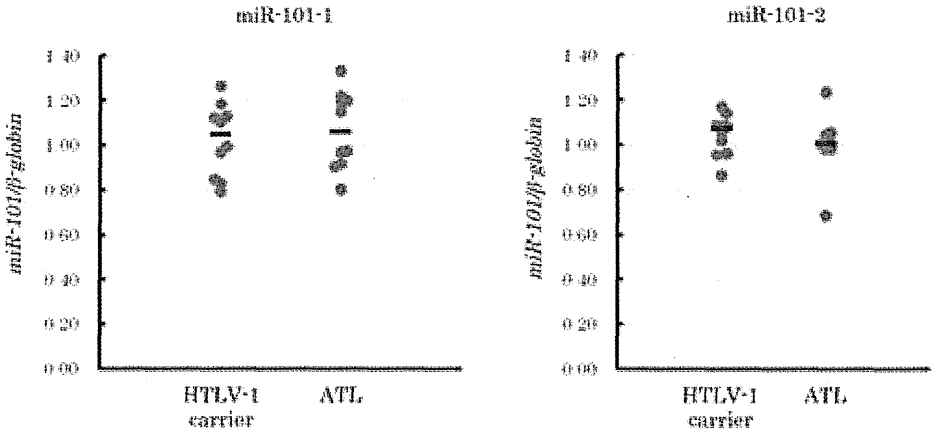
Supplementary Figure 4. Sequence analysis of EZH2. Pyrosequence analysis of EZH2 Try641 was performed in 10 ATL patients and 10 HTLV-1 carriers. Pyrograms of 6 ATL patients are shown. There were no mutations in the examined samples.

Supplementary Figure 1.



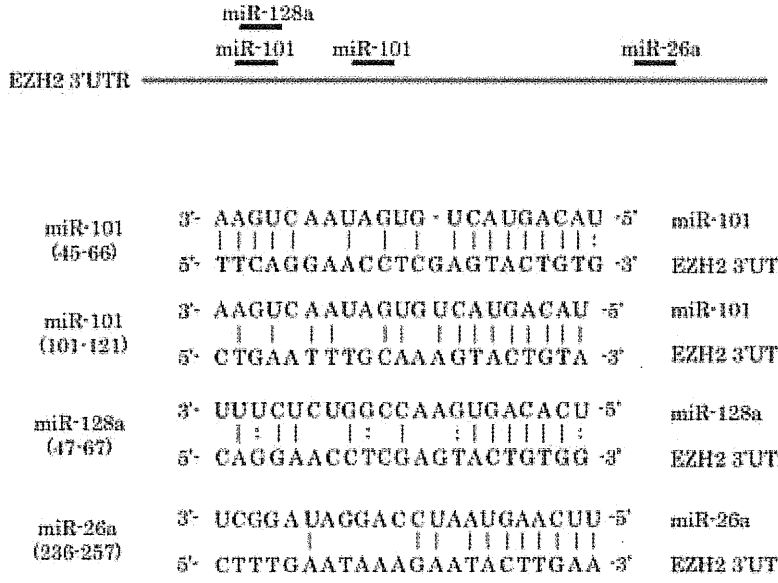
DOI: 10.3324/haematol.2010.028605

Supplementary Figure 2.



DOI: 10.3324/haematol.2010.028605

Supplementary Figure 3

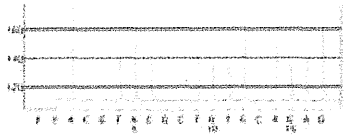


DOI: 10.3324/haematol.2010.028605

Supplementary Figure 4

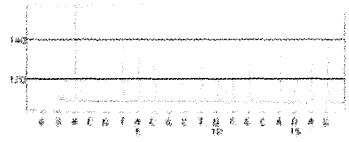
Wild-type sequence: AATACTGTGGAGAGG

ATL patient 1: wild type



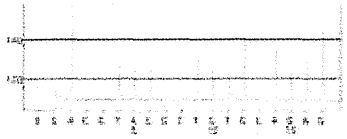
Sequence: AATAC TGTGGAGAGG

ATL patient 4: wild type



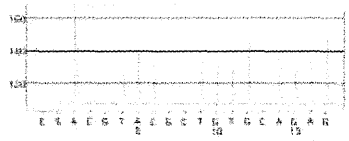
Sequence: AATAC TGTGGAGAGG

ATL patient 2: wild type



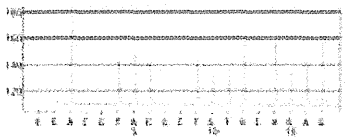
Sequence: AATAC TGTGGAGAGG

ATL patient 5: wild type



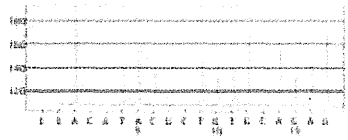
Sequence: AATAC TGTGGAGAGG

ATL patient 3: wild type



Sequence: AATAC TGTGGAGAGG

ATL patient 6: wild type



Sequence: AATAC TGTGGAGAGG

BRIEF REPORT

EML4-ALK Mutations in Lung Cancer That Confer Resistance to ALK Inhibitors

Young Lim Choi, M.D., Ph.D., Manabu Soda, M.D., Ph.D., Yoshihiro Yamashita, M.D., Ph.D., Toshihide Ueno, Ph.D., Junpei Takashima, M.D., Takahiro Nakajima, M.D., Ph.D., Yasushi Yatabe, M.D., Ph.D., Kengo Takeuchi, M.D., Ph.D., Toru Hamada, M.D., Hidenori Haruta, M.D., Ph.D., Yuichi Ishikawa, M.D., Ph.D., Hideki Kimura, M.D., Ph.D., Tetsuya Mitsudomi, M.D., Ph.D., Yoshiro Tanio, M.D., Ph.D., and Hiroyuki Mano, M.D., Ph.D., for the ALK Lung Cancer Study Group

SUMMARY

The EML4 (echinoderm microtubule-associated protein-like 4)–ALK (anaplastic lymphoma kinase) fusion-type tyrosine kinase is an oncoprotein found in 4 to 5% of non–small-cell lung cancers, and clinical trials of specific inhibitors of ALK for the treatment of such tumors are currently under way. Here, we report the discovery of two secondary mutations within the kinase domain of EML4-ALK in tumor cells isolated from a patient during the relapse phase of treatment with an ALK inhibitor. Each mutation developed independently in subclones of the tumor and conferred marked resistance to two different ALK inhibitors. (Funded by the Ministry of Health, Labor, and Welfare of Japan, and others.)

From the Division of Functional Genomics, Jichi Medical University, Tochigi (Y.L.C., M.S., Y. Yamashita, T.U., T.H., H.H., H.M.); the Department of Medical Genomics, Graduate School of Medicine, University of Tokyo, Tokyo (Y.L.C., H.M.); the Department of Internal Medicine, Osaka General Medical Center, Osaka (J.T., Y.T.); the Division of Thoracic Diseases, Chiba Cancer Center, Chiba (T.N., H.K.); the Departments of Pathology (Y. Yatabe) and Thoracic Surgery (T.M.), Aichi Cancer Center Hospital, Aichi; the Pathology Project for Molecular Targets (K.T.) and Division of Pathology (Y.I.), the Cancer Institute, Japanese Foundation for Cancer Research, Tokyo; and Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Saitama (H.M.) — all in Japan. Address reprint requests to Dr. Mano at the Division of Functional Genomics, Jichi Medical University, 3311-1 Yakushiji, Shimotsukeshi, Tochigi 329-0498, Japan, or at hmano@jichi.ac.jp.

N Engl J Med 2010;363:1734-9.

EML4-ALK IS A FUSION-TYPE PROTEIN TYROSINE KINASE THAT IS PRESENT in 4 to 5% of cases of non–small-cell lung cancer and is generated as a result of a small inversion within the short arm of human chromosome 2.[1-3] EML4-ALK undergoes constitutive dimerization through interaction between the coiled-coil domain within the EML4 region of each monomer, thereby activating ALK and generating oncogenic activity. In transgenic mice that express EML4-ALK specifically in lung epithelial cells, hundreds of adenocarcinoma nodules develop in both lungs soon after birth, and oral administration of a specific inhibitor of ALK tyrosine kinase activity rapidly eradicates such nodules from the lungs.[4] These observations reveal the essential role of EML4-ALK in the carcinogenesis of non–small-cell lung cancer harboring this fusion kinase. Furthermore, clinical trials are investigating crizotinib (PF-02341066), an inhibitor of the tyrosine kinase activity of both ALK and the met proto-oncogene (MET), for the treatment of EML4-ALK–positive non–small-cell lung cancer.

In addition to crizotinib, other tyrosine kinase inhibitors have been shown to have pronounced therapeutic activity in patients with cancer. For instance, imatinib mesylate and gefitinib, tyrosine kinase inhibitors for the c-abl oncogene 1 nonreceptor tyrosine kinase (ABL) and epidermal growth factor receptor (EGFR), improve the outcome for patients who have chronic myeloid leukemia that is positive for the BCR (breakpoint cluster region protein)–ABL fusion kinase[5] and patients who have non–small-cell lung cancer that is associated with EGFR activation,[6]

respectively. Unfortunately, however, a fraction of the target tumors are either refractory to corresponding tyrosine kinase inhibitors from the start of treatment or become resistant after an initial response.

In a case of EML4-ALK–positive non–small-cell lung cancer that became resistant to crizotinib after successful treatment for 5 months, we have discovered two de novo mutations in EML4-ALK, each of which confers resistance to the drug.

CASE REPORT

The patient was a 28-year-old man without a history of smoking who had received a diagnosis of lung adenocarcinoma, at a tumor–node–metastasis (TNM) clinical stage of T4N3M1, in April 2008. Given that the tumor did not harbor any EGFR mutations, the patient was treated with conventional chemotherapy. However, his tumor progressed after six cycles of three two-drug combinations. In November 2008, the presence of *EML4-ALK* variant 1 messenger RNA (mRNA)[1] in the tumor was confirmed by means of reverse transcription–polymerase-chain-reaction (PCR) analysis of a sputum sample. At this stage, the patient had large tumor nodules in the hilum of the right lung, multiple enlarged lymph nodes in the mediastinum, atelectasis in the right lung, and a massive effusion in the right pleural cavity (Fig. 1 in the Supplementary Appendix, available with the full text of this article at NEJM.org).

The patient was enrolled in the A8081001 study of crizotinib (ClinicalTrials.gov number, NCT00585195) on November 28, 2008, with oral administration of the drug at a dose of 250 mg twice per day. Within 1 week after the start of crizotinib treatment, his symptoms improved markedly. Although he had a partial response to the treatment, his pleural effusion was not completely eradicated (Fig. 1 in the Supplementary Appendix). After 5 months of treatment, however, the tumor abruptly started to grow again, resulting in a rapid expansion of the pleural effusion and in the development of tumors in both lungs (Fig. 1 in the Supplementary Appendix). The patient was withdrawn from the trial on May 25, 2009, and a sample of the pleural effusion in the right lung was then obtained for molecular analysis.

METHODS

DNA sequencing and characterization of the EML4-ALK mutants are described in detail in the Supplementary Appendix.

RESULTS

Because our patient's tumor resumed growth despite sustained administration of the ALK inhibitor crizotinib, we speculated that it might have acquired secondary genetic changes that confer resistance to the drug. Furthermore, given that resistance to tyrosine kinase inhibitors often results from acquired mutations within the target kinases,[7-9] we first examined the possibility that EML4-ALK itself had undergone amino acid changes.

Molecular analysis was performed on sputum specimens obtained before crizotinib treatment and pleural-effusion specimens obtained after relapse when treatment was stopped. Given that the proportion of tumor cells in the two types of specimens may have differed, we performed deep (high-coverage) sequencing of *EML4-ALK* complementary DNA (cDNA) derived from the specimens, using a high-throughput sequencer (Genome Analyzer II, Illumina) (Fig. 2 in the Supplementary Appendix). The sensitivity of our sequencing system, examined with the use of cDNA corresponding to the Janus kinase 3 (JAK3) amino acid mutation V674A[10] as a control, revealed that the maximum detection sensitivity was no more than one mismatched read per 6.50×10^5 total reads (Table 1 in the Supplementary Appendix).

Using deep sequencing, we detected a known single-nucleotide polymorphism, rs3795850, in the cDNA from the four specimens that were positive for *EML4-ALK* (Table 2 and Fig. 3 in the Supplementary Appendix). In addition, a T→C change at a position corresponding to nucleotide 4230 of human wild-type *ALK* cDNA (GenBank accession number, NM_004304) was detected at a low frequency (8.9%) in the sputum cDNA from our patient. Furthermore, two new alterations, G→A and C→A changes at positions corresponding to nucleotides 4374 and 4493 of wild-type *ALK* cDNA, were detected at frequencies of 41.8% and 14.0%, respectively, in the patient's pleural-effusion cDNA. There were no other recurrent alterations (present in 5% of reads) in the kinase-domain cDNA derived from any of the specimens.

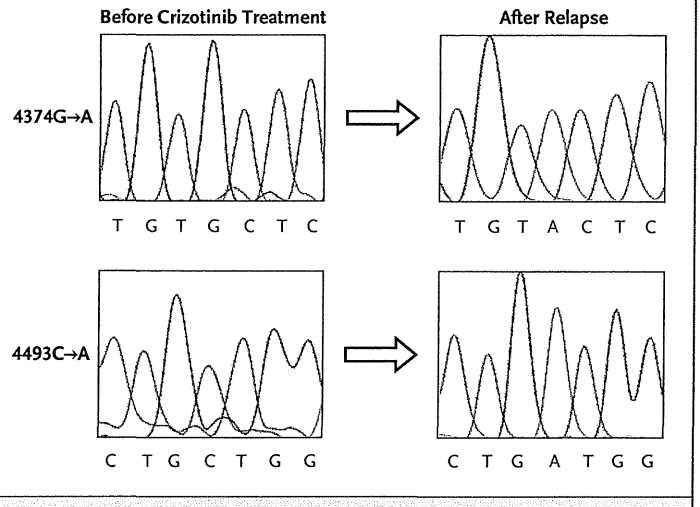


Figure 1. Secondary Mutations within *EML4-ALK*.
Electrophoretograms are shown for *EML4-ALK* cDNA clones prepared from sputum specimens obtained from our patient before crizotinib treatment and from pleural-effusion specimens obtained after relapse. The 4374G→A and 4493C→A mutations are present in the specimens obtained after relapse.

We next attempted to confirm these nucleotide changes by using Sanger sequencing. To rule out the possibility that the mutations had occurred in endogenous wild-type *ALK* rather than in *EML4-ALK*, we performed PCR with a forward primer targeted to *EML4* cDNA so that only the fusion cDNA would be amplified (Fig. 2 in the Supplementary Appendix). We did not detect the 4230T→C change among the 256 fusion cDNA clones derived from the patient's sputum specimens (data not shown), indicating that it was an artifact of the initial PCR or the deep-sequencing step. We did, however, readily confirm both 4374G→A and 4493C→A changes. Among 73 *EML4-ALK* cDNA clones from the patient's pleural-effusion specimens, 34 (46.6%) were positive for 4374G→A and 11 (15.1%) were positive for 4493C→A (Fig. 1). (The remaining 28 [38.4%] were negative for both point mutations.) These rates of detection are similar to those from the deep sequencing of *ALK*, indicating that wild-type *ALK* mRNA was present at a low level in lung tissue, as reported previously.[1]

The PCR analyses covered both nucleotide positions, yet none of the patient's specimens contained both mutations, indicating that each mutation occurred independently. Genomic fragments encompassing the 4374G and 4493C positions were also amplified by means of a PCR assay and were then subjected to nucleotide sequencing, which confirmed the presence of each of the two mutations in the tumor genome (Fig. 4 in the Supplementary Appendix).

The 4374G→A and 4493C→A substitutions result in cysteine→tyrosine (C→Y) and leucine→methionine (L→M) changes at the positions corresponding to amino acids 1156 and 1196, respectively, of wild-type human ALK (Fig. 2 in the Supplementary Appendix). We examined whether such amino acid changes affect the sensitivity of EML4-ALK to ALK inhibitors.

Cells of the mouse interleukin-3–dependent cell line BA/F3 that were made to individually express primary EML4-ALK and secondary mutant EML4-ALK (with the C1156Y or L1196M mutation) were exposed to ALK inhibitors. Crizotinib inhibited the growth of BA/F3 cells expressing primary EML4-ALK, in a concentration-dependent manner (Fig. 2A). In contrast, cells expressing either the C1156Y or L1196M mutant form manifested a markedly reduced sensitivity to the drug. Cells expressing the L1196M mutant form of EML4-ALK were more resistant to crizotinib than were those expressing the C1156Y mutant form (Fig. 2A, and Fig. 5 in the Supplementary Appendix).

We also examined whether cells expressing these EML4-ALK mutants are also refractory to other ALK inhibitors. A 2,4-pyrimidinediamine derivative (PDD) has a median inhibitory concentration for ALK of less than 10 nM,[11] and oral administration of PDD has been shown to eradicate lung-cancer nodules in transgenic mice with EML4-ALK expression.[4] BA/F3 cells expressing EML4-ALK with either the C1156Y or L1196M mutation were markedly less sensitive to PDD than were those expressing the primary EML4-ALK (Fig. 2A). Thus, although these mutations appear to develop during clinical treatment with crizotinib, their generation probably renders EML4-ALK resistant not only to crizotinib but also to other ALK inhibitors. In contrast to the resistance profile for crizotinib, BA/F3 cells expressing the EML4-ALK C1156Y mutant form were slightly more resistant to PDD than were those expressing the L1196M mutant form (Fig. 2A, and Fig. 6 in the Supplementary Appendix), indicating that the resistance profiles for the two mutations may be, in part, inhibitor-dependent, as was previously shown for BCR-ABL mutants.[12]

We examined tyrosine phosphorylation of

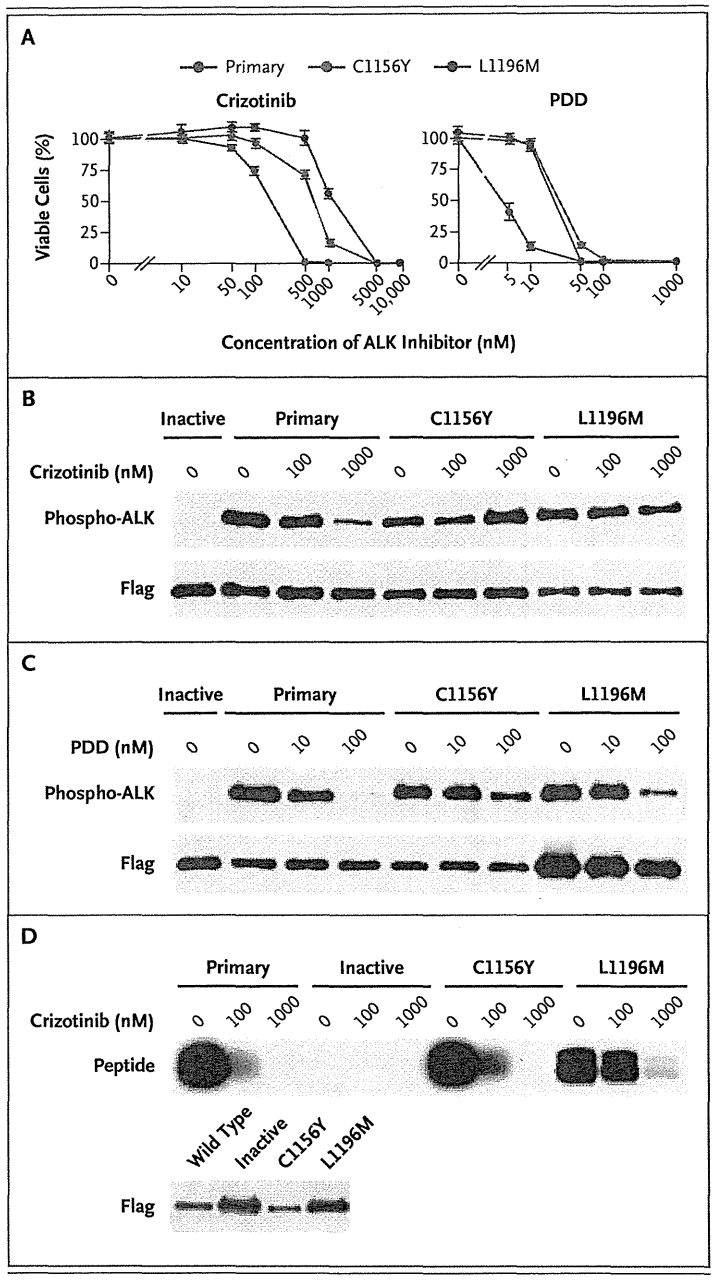


Figure 2. Properties of EML4-ALK with Secondary Mutations.

Panel A shows the percentage of viable BA/F3 cells expressing primary EML4-ALK, EML4-ALK with the C1156Y mutation, or EML4-ALK with the L1196M mutation, after 5×10^5 cells were incubated for 48 hours with the indicated concentration of crizotinib (left) or 2,4-pyrimidinediamine derivative (PDD) (right). Data are expressed as the mean value, from three separate experiments, for the percentage of cells expressing primary EML4-ALK after incubation in the vehicle (dimethyl sulfoxide) only. The I bars indicate standard deviations. Because primary EML4-ALK, EML4-ALK with the C1156Y mutation, and EML4-ALK with the L1196M mutation each abrogate the interleukin-3 dependence of BA/F3 cells, the assays were performed in the absence of the interleukin. Panels B and C show the effect of ALK inhibitors on EML4-ALK and its secondary mutant forms, tagged with the Flag epitope, in BA/F3 cells. Panel B shows the results of exposure to various concentrations of crizotinib for 15 hours, after which EML4-ALK was immunoprecipitated from cell lysates with antibodies against the Flag epitope and the immunoprecipitate was subjected to immunoblot analysis with the use of antibodies specific for ALK phosphorylated at the tyrosine at position 1604 (Phospho-ALK) or for the Flag epitope. Cells expressing an inactive mutant form of EML4-ALK were examined as a negative control. Panel C shows the results of a similar experiment, involving PDD instead of crizotinib. Panel D shows the results of an in vitro kinase assay for Flag-tagged EML4-ALK or its secondary mutants immunoprecipitated from BA/F3 cells with antibodies against the Flag epitope. The immunoprecipitates were incubated with [γ-^{32}P]ATP, a synthetic peptide, and various concentrations of crizotinib (top). Separate immunoprecipitate samples were subjected to immunoblot analysis with antibodies against the Flag epitope (bottom).

EML4-ALK by means of immunoblot analysis, using antibodies specific for ALK phosphorylated at the tyrosine at position 1604. The exposure of BA/F3 cells to crizotinib markedly inhibited the tyrosine phosphorylation of EML4-ALK but did not substantially affect that of the C1156Y and L1196M mutants (Fig. 2B). Exposure to PDD also inhibited the tyrosine phosphorylation of EML4-ALK, in a concentration-dependent manner, with a lesser effect on the mutants (Fig. 2C). The results of an in vitro kinase assay were consistent with these findings, showing pronounced inhibition of the enzymatic activity of primary EML4-ALK with crizotinib, whereas the effect on the C1156Y mutant was less pronounced and the effect on the L1196M mutant was much less pronounced (Fig. 2D).

Figure 3 shows the cysteine at position 1156 (C1156) and the leucine at position 1196 (L1196) of the kinase domain of ALK.[13] C1156 is positioned adjacent to the N-terminal of the predicted helix αC as well as close to the upper edge of the ATP-binding pocket. No activating mutations have been reported at this position in other tyrosine kinases in cancer specimens. L1196 of ALK corresponds to the threonine at position 315 in ABL and at position 790 in EGFR, each of which is the site of the most fre-

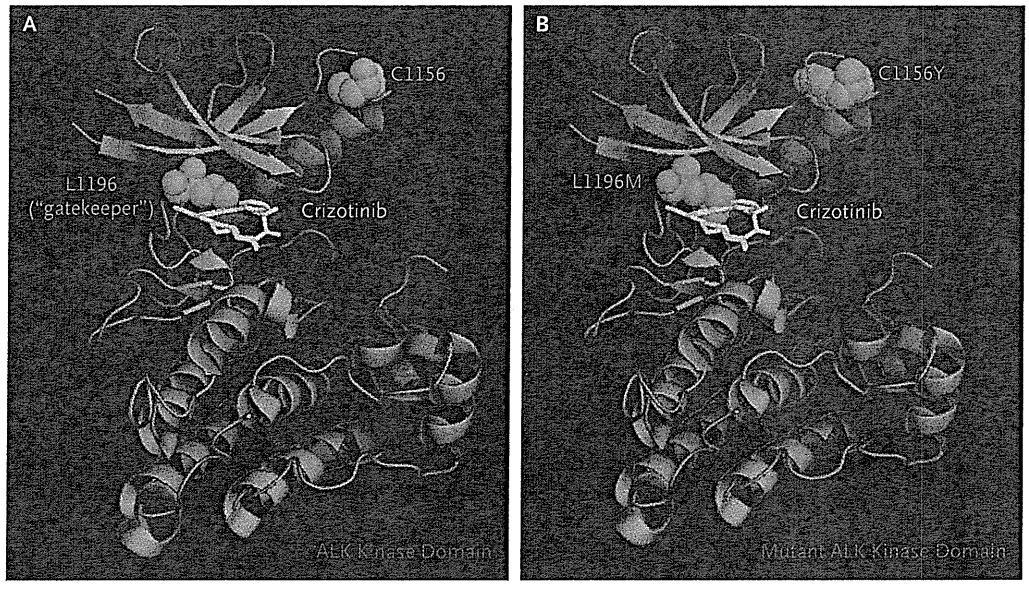


Figure 3. Predicted Crystal Structure of the Kinase Domain of ALK.

Panel A shows the nonmutant kinase domain and the site of binding of crizotinib. Panel B shows the EML4-ALK secondary mutations at positions 1156 and 1196. Adapted from Bossi and colleagues.[13]

quently acquired mutations that confer resistance to tyrosine kinase inhibitors in these kinases (Fig. 7 in the Supplementary Appendix).[14,15] This site is located at the bottom of the ATP-binding pocket (Fig. 3), and the presence of an amino acid with a bulky side chain at this "gatekeeper" position may interfere with the binding of many tyrosine kinase inhibitors.[7,16]

DISCUSSION

We identified two de novo mutations within the kinase domain of EML4-ALK from the tumor of a single patient that confer resistance to multiple ALK inhibitors. Given that we did not detect any *EML4-ALK* cDNA harboring both mutations, we propose that each mutation developed independently in distinct subclones of the tumor. Because we were not able to examine pleural-effusion specimens from the patient before he received crizotinib treatment, we do not know whether the resistant clones were present initially or developed secondarily, during the treatment.

Amino acid substitutions at the gatekeeper position of several tyrosine kinases have been detected in tumors treated with tyrosine kinase inhibitors (Fig. 7 in the Supplementary Appendix).[7-9,17,18] Whereas no mutations at this site have previously been reported for EML4-ALK or ALK, the effects of various artificial amino acid substitutions at the gatekeeper position of nucleophosmin (NPM)–ALK, another fusion-type "oncokinase" form of ALK, were recently examined.[19] The findings were consistent with the results of our analysis of tumor cells in vivo: the introduction of methionine at this position rendered NPM-ALK resistant to ALK inhibitors. It is therefore likely that gatekeeper alterations constitute a universal mechanism for the acquisition of tyrosine kinase–inhibitor resistance in oncogenic tyrosine kinases.

In contrast to gatekeeper substitutions, activating mutations at the position adjacent, on the N-terminal side, to the αC helix (e.g., C1156 in ALK) have not been confirmed for other tyrosine kinases in cancer specimens. Though a T→I change at the corresponding position of EGFR was described in one case of non–small-cell lung cancer, its relevance to drug sensitivity was not examined.[16] The importance of helix αC for allosteric regulation of enzymatic activity has been shown, however, for serine–threonine kinases.[20] A change at C1156 of ALK might therefore interfere allosterically with the binding of tyrosine

kinase inhibitors. Determination of the crystal structure of the ALK kinase domain with the C1156Y or L1196M mutation should shed light on these matters, as well as provide a basis for the development of next-generation ALK inhibitors that may effectively eradicate tumors harboring EML4-ALK with the acquired mutations.

Supported in part by grants from the Ministry of Health, Labor, and Welfare of Japan; the Ministry of Education, Culture, Sports, Science, and Technology of Japan; and the Japan Society for the Promotion of Science.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Y.-J. Bang and the medical staff at Seoul National University Hospital for their support in the treatment of this patient, as well as Y. Togashi and S. Hatano for technical assistance.

REFERENCES

1. Soda M, Choi YL, Enomoto M, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 2007;448:561-6.
2. Mano H. Non-solid oncogenes in solid tumors: EML4-ALK fusion genes in lung cancer. Cancer Sci 2008;99:2349-55.
3. Horn L, Pao W. EML4-ALK: honing in on a new target in non-small-cell lung cancer. J Clin Oncol 2009;27:4232-5.
4. Soda M, Takada S, Takeuchi K, et al. A mouse model for EML4-ALK-positive lung cancer. Proc Natl Acad Sci U S A 2008;105:19893-7.
5. Druker BJ, Talpaz M, Resta DJ, et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med 2001;344:1031-7.
6. Mok TS, Wu YL, Thongprasert S, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 2009;361:947-57.
7. Shah NP, Nicoll JM, Nagar B, et al. Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. Cancer Cell 2002;2:117-25.
8. Kobayashi S, Boggon TJ, Dayaram T, et al. EGFR mutation and resistance of non–small-cell lung cancer to gefitinib. N Engl J Med 2005;352:786-92.
9. Pao W, Miller VA, Politi KA, et al. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. PLoS Med 2005;2(3):e73.
10. Choi YL, Kaneda R, Wada T, et al. Identification of a constitutively active mutant of JAK3 by retroviral expression screening. Leuk Res 2007;31:203-9.
11. Choi YL, Takeuchi K, Soda M, et al. Identification of novel isoforms of the EML4-ALK transforming gene in non-small cell lung cancer. Cancer Res 2008;68:4971-6.
12. Branford S, Melo JV, Hughes TP. Selecting optimal second-line tyrosine kinase inhibitor therapy for chronic myeloid leukemia patients after imatinib failure: does the BCR-ABL mutation status really matter? Blood 2009;114:5426-35.
13. Bossi RT, Saccardo MB, Ardini E, et al. Crystal structures of anaplastic lymphoma kinase in complex with ATP competitive inhibitors. Biochemistry 2010;49:6813-25.
14. Deininger M, Buchdunger E, Druker BJ. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. Blood 2005;105:2640-53.
15. Linardou H, Dahabreh IJ, Bafaloukos D, Kosmidis P, Murray S. Somatic EGFR mutations and efficacy of tyrosine kinase inhibitors in NSCLC. Nat Rev Clin Oncol 2009;6:352-66.
16. Carter TA, Wodicka LM, Shah NP, et al. Inhibition of drug-resistant mutants of ABL, KIT, and EGF receptor kinases. Proc Natl Acad Sci U S A 2005;102:11011-6.
17. Cools J, DeAngelo DJ, Gotlib J, et al. A tyrosine kinase created by fusion of the PDGFRA and FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. N Engl J Med 2003;348:1201-14.
18. Tamborini E, Bonadiman L, Greco A, et al. A new mutation in the KIT ATP pocket causes acquired resistance to imatinib in a gastrointestinal stromal tumor patient. Gastroenterology 2004;127:294-9.
19. Lu L, Ghose AK, Quail MR, et al. ALK mutants in the kinase domain exhibit altered kinase activity and differential sensitivity to small molecule ALK inhibitors. Biochemistry 2009;48:3600-9.
20. Hindie V, Stroba A, Zhang H, et al. Structure and allosteric effects of low-molecular-weight activators on the protein kinase PDK1. Nat Chem Biol 2009;5:758-64.

Oncogene (2010) 29, 3723–3731

www.nature.com/onc

ONCOGENOMICS

Array-based genomic resequencing of human leukemia

Y Yamashita[1], J Yuan[2], I Suetake[3], H Suzuki[4], Y Ishikawa[5], YL Choi[1,6], T Ueno[1], M Soda[1], T Hamada[1], H Haruta[1], S Takada[1], Y Miyazaki[7], H Kiyoi[8], E Ito[9], T Naoe[5], M Tomonaga[7], M Toyota[10], S Tajima[3], A Iwama[2,11] and H Mano[1,6,11]

[1]Division of Functional Genomics, Jichi Medical University, Tochigi, Japan; [2]Department of Cellular and Molecular Medicine, Graduate School of Medicine, Chiba University, Chiba, Japan; [3]Laboratory of Epigenetics, Institute for Protein Research, Osaka University, Osaka, Japan; [4]First Department of Internal Medicine, Sapporo Medical University, Hokkaido, Japan; [5]Department of Hematology and Oncology, Nagoya University Graduate School of Medicine, Nagoya, Japan; [6]Department of Medical Genomics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; [7]Department of Molecular Medicine and Hematology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan; [8]Department of Infectious Diseases, Nagoya University School of Medicine, Nagoya, Japan; [9]Department of Paediatrics, Hirosaki University Graduate School of Medicine, Aomori, Japan; [10]Department of Biochemistry, Sapporo Medical University, Hokkaido, Japan and [11]CREST, Japan Science and Technology Agency, Saitama, Japan

To identify oncogenes in leukemias, we performed large-scale resequencing of the leukemia genome using DNA sequence arrays that determine ~9 Mbp of sequence corresponding to the exons or exon–intron boundaries of 5648 protein-coding genes. Hybridization of genomic DNA from CD34-positive blasts of acute myeloid leukemia ($n=19$) or myeloproliferative disorder ($n=1$) with the arrays identified 9148 nonsynonymous nucleotide changes. Subsequent analysis showed that most of these changes were also present in the genomic DNA of the paired controls, with 11 somatic changes identified only in the leukemic blasts. One of these latter changes results in a Met-to-Ile substitution at amino-acid position 511 of Janus kinase 3 (JAK3), and the JAK3(M511I) protein exhibited transforming potential both *in vitro* and *in vivo*. Further screening for *JAK3* mutations showed novel and known transforming changes in a total of 9 out of 286 cases of leukemia. Our experiments also showed a somatic change responsible for an Arg-to-His substitution at amino-acid position 882 of DNA methyltransferase 3A, which resulted in a loss of DNA methylation activity of >50%. Our data have thus shown a unique profile of gene mutations in human leukemia.

Oncogene (2010) **29,** 3723–3731; doi:10.1038/onc.2010.117; published online 19 April 2010

Keywords: resequencing; AML; JAK3; DNMT3A

Introduction

Leukemias are clonal disorders of hematopoietic stem cells or immature progenitors. Several subtypes of leukemia are associated with disease-specific karyotype anomalies in the malignant blasts. Most cases of acute promyelocytic leukemia a subtype of acute myeloid leukemia (AML), for instance, are associated with a t(15;17) chromosomal rearrangement that results in the production of the PML-RARA fusion-type oncoprotein (Tallman and Altman, 2008). Similarly, another subtype of AML is associated with a t(8;21) rearrangement, resulting in the production of the oncogenic RUNX1-CBFA2T1 protein (Nimer and Moore, 2004).

The karyotype of leukemic blasts is an important determinant of the long-term prognosis of affected individuals. AML with t(15;17), t(8;21) or inv(16) rearrangements thus constitutes a subgroup of leukemias with a 'favorable' karyotype, with a 5-year survival rate of >60%, whereas AML with an 'adverse' karyotype (monosomy 7, monosomy 5 or complex anomalies) has a 5-year survival rate of only <15% (Grimwade *et al.*, 1998). The prognosis of AML with a normal karyotype (constituting ~50% of all AML cases) is substantially worse than that with a favorable karyotype, with a 5-year survival rate of 24% (Byrd *et al.*, 2002), indicating that blasts with a normal karyotype may contain transforming genes generated as a result of (1) sequence alterations, (2) epigenetic abnormalities or (3) small chromosomal rearrangements not detectable by the G-banding technique. Indeed, several genes, including *NPM1* and *KIT*, have been found to be mutated and activated in AML blasts with a normal karyotype (Schlenk *et al.*, 2008).

The identification of transforming genes in AML will require large-scale resequencing of the blast genome. Although a new generation of sequencing technologies is now available, whole-genome resequencing of many samples remains a demanding task (Bentley *et al.*, 2008; Wheeler *et al.*, 2008). Although DNA microarray-based sequencing is suitable for analysis of multiple samples, currently available platforms are limited in the number of nucleotides that each array is able to probe. To overcome such limitations, we have now applied the extra-large arrays ('wafers') manufactured by Perlegen Sciences (Mountain View, CA, USA) (originally developed for typing of

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 6 May 2009; revised 1 January 2010; accepted 17 March 2010; published online 19 April 2010

single-nucleotide polymorphisms) (Patil *et al.*, 2001) to resequencing of the human genome. Our two-step analysis of human leukemia specimens ($n=20$) has identified a novel transforming mutation in the gene for Janus kinase 3 (JAK3) and a hypomorphic mutation in that for DNA methyltransferase 3A (DNMT3A).

Results

Sequencing strategy

Oligonucleotide probes on the sequencing wafer for the first phase of our study were designed to detect nonsynonymous nucleotide changes in the coding exons of the genome. Intronic sequences (GT in the splicing donor sequence AG-GT and AG in the splicing acceptor sequence AG-G) adjacent to coding exons were also interrogated with the wafer to capture splicing anomalies. Genes examined by the wafer included those known to be mutated in cancer and reported in the catalog of somatic mutations in cancer (COSMIC, http://www.sanger.ac.uk/genetics/cgp/cosmic) as of September 2006 ($n=338$) and those related to the regulation of DNA repair ($n=419$), chromatin structure ($n=299$), redox regulation ($n=102$), epigenetic regulation ($n=44$), cell signaling ($n=2490$), protein kinases ($n=314$), gene transcription ($n=797$), cell cycle ($n=297$), apoptosis ($n=312$), DNA replication ($n=144$) or other functions ($n=92$) (Figure 1a). A total of 5648 genes were thus analyzed with the wafer.

To efficiently isolate oncogenes generated by point mutation using our sequencing array, we selected leukemic blasts with a karyotype characterized by few chromosome anomalies and by few copy number variations of chromosomes, as determined by comparative genomic hybridization with single-nucleotide polymorphism-typing arrays (Supplementary Figure S1). We isolated 15 cases of *de novo* AML, 4 cases of AML that developed from myelodysplastic syndrome, and 1 case of myeloproliferative disorder negative for the JAK2(V617F) and MPL(W515L) mutations (Kralovics *et al.*, 2005; Pikman *et al.*, 2006) (Supplementary Table S1).

From each of these 20 individuals enrolled in the study, we purified immature blasts positive for the surface expression of CD34 (leukemic fraction) as well as a paired control fraction of mature T cells positive for the surface expression of CD4. Although monocytes-macrophages may also express a low level of CD4 at the cell surface, our magnetic bead-based purification system preferentially enriched mature T cells with a high level of CD4 expression; contamination of the mature T-cell fraction with monocytes-macrophages was judged to be <9% by flow cytometry (Supplementary Figure S2).

Given the potential presence of substantial numbers of unreported single-nucleotide polymorphisms in the human genome, we adopted a two-step analysis to select somatic changes (Figure 1b). In phase I, genomic DNA was isolated from the CD34$^+$ fraction, subjected to mid-range PCR amplification and hybridized with the wafer to examine ~9 Mbp of nucleotide sequence. In phase II, we constructed a smaller wafer to investigate only the

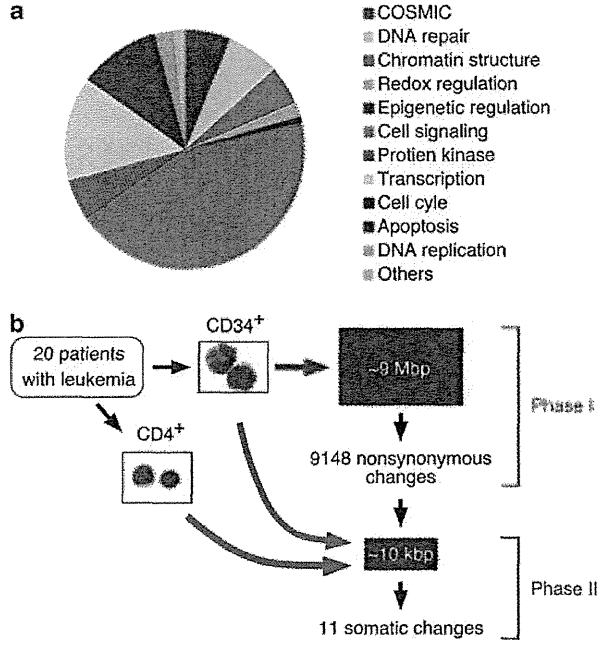


Figure 1 Resequencing of the leukemia genome with wafers. (**a**) Genes interrogated by the phase I wafer ($n=5648$) included those listed in the COSMIC database and those categorized on the basis of function of the encoded protein as indicated. (**b**) CD34$^+$ and CD4$^+$ cell fractions were purified from individuals with leukemia ($n=20$). Genomic DNA of the former fractions was assayed with the phase I wafer including ~9 Mbp of sequence, resulting in the isolation of 9148 nonsynonymous nucleotide changes in 3403 independent genes. The phase II wafer was then constructed to analyze these 9148 changes and was hybridized with genomic DNA from both CD34$^+$ and CD4$^+$ fractions separately. Only 11 mutations were found to be present in the former fraction but not in the latter.

nucleotides shown to be changed in phase I relative to the human reference sequence. Genomic DNA isolated from leukemic blasts and paired control fractions was then analyzed individually with the phase II wafer. We assumed that a nucleotide change was a germline polymorphism if it was observed in both leukemic and control fractions of the same individual, and that it was a somatic mutation if it was observed in the former fraction but not in the latter.

Identification of the JAK3(M511I) mutation

Screening of the leukemic blasts of the 20 individuals for point mutations in phase I yielded 9148 nonsynonymous changes among 3403 independent genes, a frequency similar to that observed in other large-scale resequencing studies performed with capillary sequencers (Sjoblom *et al.*, 2006; Greenman *et al.*, 2007). However, analysis of CD4$^+$ fractions showed that most of these sequence changes were also present in the paired control genome, leaving only 11 nonsynonymous somatic mutations in 11 genes (Supplementary Table S2). Such small number of somatic mutations is in a good agreement with the eight somatic mutations found in AML through whole-genome resequencing using the

Illumina Genome Analyser (Illumina, San Diego, CA, USA) (Ley *et al.*, 2008). All of our 11 somatic changes were confirmed by analysis of both genomic DNA and cDNA of the corresponding specimens with a capillary sequencer (data not shown). These data thus support the necessity of examining paired noncancerous specimens to pinpoint somatic changes in the cancer genome.

One of the gene mutations found only in the $CD34^+$ fractions results in a Met-to-Ile change at amino-acid position 511 of JAK3. A heterozygous *JAK3* mutation responsible for the amino-acid change was confirmed in both genomic DNA and cDNA from the $CD34^+$ fraction, but not in those from the corresponding $CD4^+$ fraction of patient ID JM07 (Supplementary Figure S3), who had *de novo* AML (M1 subtype) and a normal karyotype (Supplementary Table S1). In contrast to JAK2, activating mutations in which are preferentially associated with myeloproliferative disorder, several gain-of-function mutations (such as I87T, P132T, Q501H, A572V, R657Q and V722I) of JAK3 have recently been associated with acute megakaryoblastic leukemia of children (Walters *et al.*, 2006; Sato *et al.*, 2008). Other JAK3 mutations (such as A573V and A593T) were also identified in the same disorder, and an M576L substitution was detected in an adult with acute megakaryocytic leukemia (AML, M7 subtype) (Kiyoi *et al.*, 2007), although the transforming potential of these changes remains unknown.

Given that the M511I mutant of JAK3 has not previously been described and that the relevance of JAK3 to the pathogenesis of adult AML has not been extensively investigated, we first focused on the function of JAK3(M511I). The M511 residue is located in the linker region between the Src homology 2 (SH2) domain and the pseudokinase domain of JAK3 (Figure 2a). The transforming mutation Q501H that is associated with juvenile acute megakaryoblastic leukemia (Sato *et al.*, 2008) is also located in this region. Given that JAK3 is abundant in and has an essential role in the development of lymphocytes (Russell *et al.*, 1995), we examined the expression level of *JAK3* in AML blasts. The gene was expressed at a high level in most AML specimens ($n = 52$), with its expression level being greater than that of *JAK2* in all but three cases (Supplementary Figure S3).

To examine the transforming potential of JAK3(M511I), we introduced the mutant or wild-type protein into the interleukin-3 (IL-3)-dependent mouse cell line 32D (Greenberger *et al.*, 1983). Although 32D cells forced to express wild-type JAK3 underwent rapid apoptosis after withdrawal of IL-3, those expressing JAK3(M511I) continued to grow even in the absence of IL-3, although at a reduced rate compared with that of cells expressing the artificially generated, highly transforming mutant JAK3(V674A) (Choi *et al.*, 2007) (Supplementary Figure S3). 32D cells differentiate into terminal granulocytes in the presence of granulocyte colony-stimulating factor. However, cells expressing the M511I or V674A mutant of JAK3 maintained an exponential rate of growth, without any sign of differentiation, in the presence of granulocyte colony-stimulating factor (Figure 2b, Supplementary Figure S3), supporting the notion that the M511I mutant has transforming potential.

To directly examine the leukemogenic activity of JAK3(M511I), we generated a recombinant retrovirus encoding this mutant and used it to infect murine hematopoietic stem cells. Reconstitution of the bone marrow of lethally irradiated mice with such infected cells resulted in marked lymphocytosis in peripheral blood and enlargement of the spleen in the recipient animals (Figure 2c). The cells in the peripheral blood, spleen and bone marrow of the recipients responsible for these phenotypes manifested a medium-sized, blastic morphology, and flow cytometric analysis revealed them to be $CD8^+$ T cells (Figure 2c). The clonal nature of these proliferating T cells was further confirmed by Southern blot analysis (Supplementary Figure S4), indicative of the development of T-cell acute lymphoblastic leukemia in the recipient mice.

To assess the prevalence of *JAK3* mutations in adult leukemia, we further examined the nucleotide sequence of the entire coding region of *JAK3* cDNA in an additional 266 specimens of leukemic blasts. The coding region of *JAK3* cDNA was successfully amplified by PCR from 83 specimens. We could further identify 4 distinct *JAK3* sequence changes in 8 of these 83 samples: 1 case with G62S, 4 cases with Q501H, 2 cases with R657Q and 1 case with R918C (Figure 2a). Taking into account the 20 cases evaluated in the phase I analysis, we thus identified a total of 9 cases with a mutant form of JAK3 (3.1%) among 286 cases of leukemia (Supplementary Table S3). Our identification of known transforming JAK3 mutants (Q501H and R657Q) originally associated with acute megakaryoblastic leukemia prompted us to determine the prevalence of these two changes in another cohort of AML ($n = 148$), revealing two cases with JAK3(Q501H) and one case with JAK3(R657Q). In addition, analysis of a hematopoietic cell line (KCL22) (Kubonishi and Miyoshi, 1983) established from a patient with chronic myeloid leukemia in BC revealed yet another mutation (L1017M) of JAK3 (Figure 2a).

To directly compare the transforming potential of these various JAK3 mutants, we introduced each protein into the IL-3-dependent mouse B-cell line BA/F3 and examined the growth properties of the resulting transfectants. Whereas all cells expressing the JAK3 mutants proliferated in a similar manner in the presence of IL-3 (data not shown), culture without IL-3 revealed marked differences in the transforming potential among the mutants. JAK3(M511I) was the most efficient oncokinase, with a transforming activity similar to that of JAK3(V674A). The frequent mutants JAK3(Q501H) and JAK3(R657Q) exhibited weaker but still pronounced transforming potential, whereas the remaining mutants (G62S, R918C and L1017M) showed an even lower potential (Figure 2d).

Somatic mutations of DNMT3A

Another somatic mutation identified in the phase II data set was a heterozygous change in *DNMT3A* that results

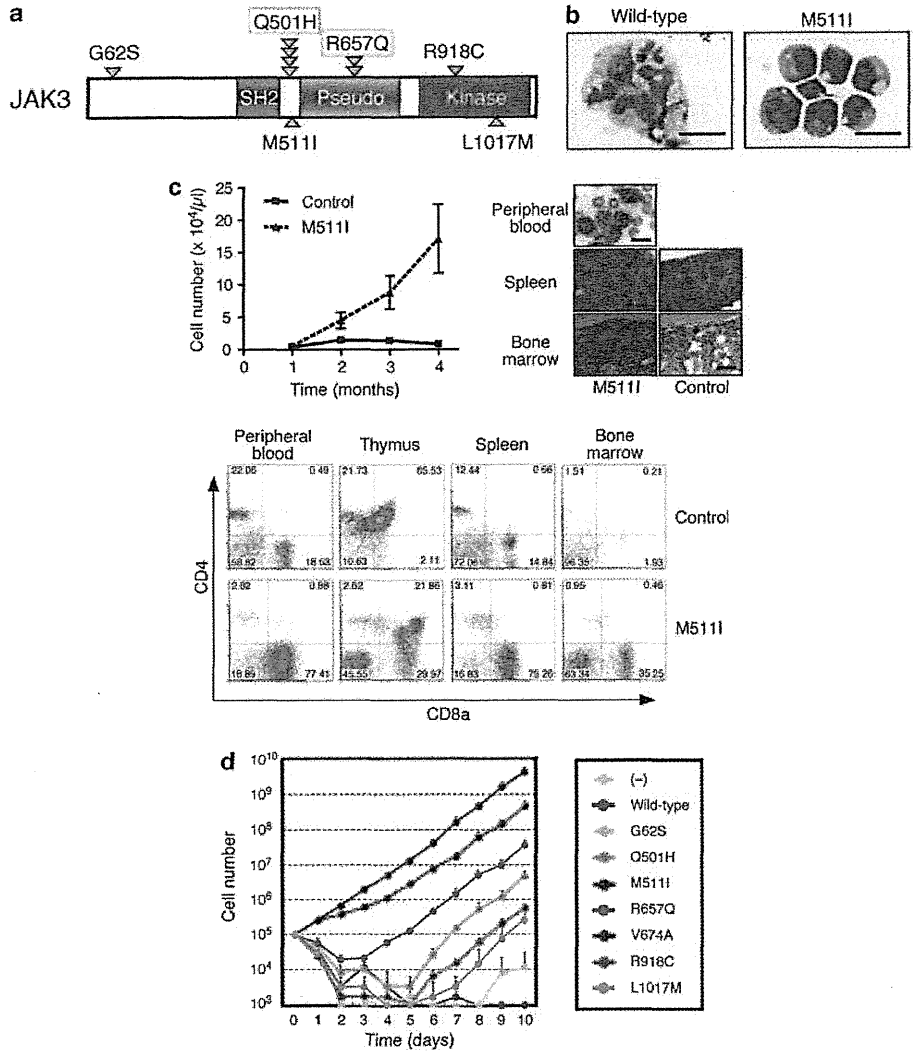


Figure 2 Identification of JAK3 mutants in leukemia. (**a**) Amino-acid substitutions detected in this study are shown relative to the domain organization of JAK3. The mutations M511I (one case) and Q501H (four cases) are located in the linker region between the SH2 and pseudokinase domains of JAK3, whereas G62S (one case), R657Q (two cases) and R918C (one case) are located in the amino-terminal region, the pseudokinase domain and the kinase domain, respectively. The KCL22 cell line also harbors an L1017M mutation within the kinase domain of JAK3. Previously known activating mutations of JAK3 (Q501H and R657Q) are indicated by red rectangles. (**b**) Mouse 32D cells expressing wild-type human JAK3 or the JAK3(M511I) mutant were incubated with G-CSF (0.5 ng/ml) for 14 days, stained with Wright–Giemsa solution and examined by light microscopy. Scale bars, 20 μm. (**c**) C57BL/6 mice were irradiated and then injected intravenously with syngeneic $CD34^{-}KSL$ hematopoietic stem cells infected with a retrovirus encoding JAK3(M511I) or the corresponding empty virus (control). The number of white blood cells in peripheral blood was counted at the indicated times thereafter; data are means ± s.d. for 10 mice in each group (upper left panel). Peripheral blood, spleen and bone marrow isolated from recipient mice 3 months after cell injection were stained with the Wright–Giemsa solution (peripheral blood) or hematoxylin–eosin (spleen and bone marrow) and were then examined by light microscopy (upper right panel); scale bars represent 10, 200 and 50 μm, respectively. Mononuclear cells isolated from peripheral blood, thymus, spleen and bone marrow of recipient mice 3 months after cell injection were subjected to flow cytometric analysis of surface expression of CD4 and CD8a (lower panel). (**d**) Control BA/F3 cells (–) or those expressing the indicated JAK3 mutants were cultured without IL-3 for the indicated times, after which the cell number was determined. Data are means + s.d. of triplicates from a representative experiment.

in an R882H substitution in the encoded protein (Figure 3a, Supplementary Figure S5). DNMT3A, together with DNMT3B, has an essential role in *de novo* methylation of the human genome (Okano *et al.*, 1999), and an aberrant methylation profile (hypermethylation of CpG islands and hypomethylation of other regions) is a hallmark of cancer cells (El-Osta, 2004). Despite a direct linkage between such methylation changes and silencing of tumor-suppressor genes in cancer, the molecular mechanism responsible for such abnormal methylation remains unknown. Our data thus provide the first evidence of somatic mutation of a DNA methyltransferase gene in cancer cells. Mutations in the catalytic domain of DNMT3B have been shown to be

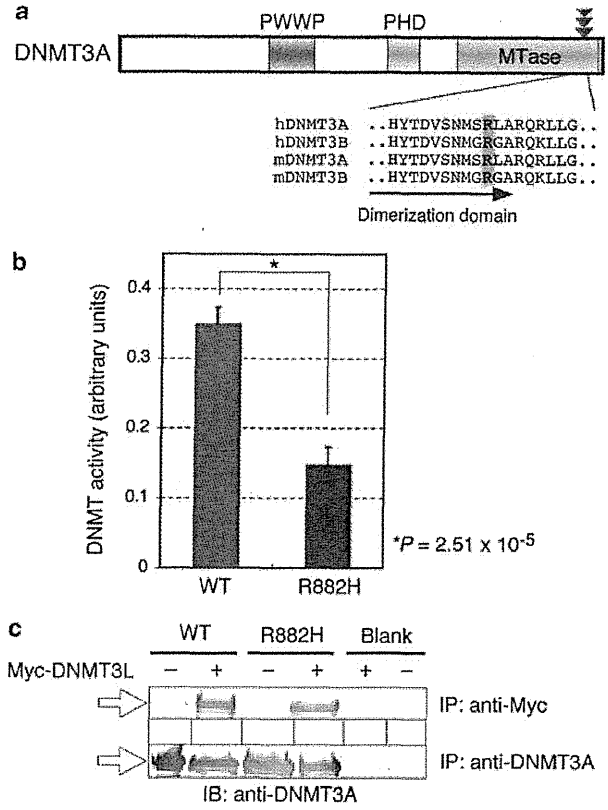


Figure 3 Identification of a DNMT3A mutant in leukemia. (**a**) Domain organization of human DNMT3A showing that the R882 residue found to be mutated in leukemia is conserved among human (h) and mouse (m) members of the DNMT3 family. DNMT3A contains a tetrapeptide PWWP domain, polybromo homology domain (PHD) and methyltransferase (MTase) domain. The R882 residue is located in the homodimerization region present within the MTase domain. (**b**) Wild-type (WT) and R882H forms of DNMT3A were expressed in and purified from insect cells and then subjected to an *in vitro* assay of methyltransferase activity. Data are means + s.d. of triplicates from a representative experiment. The *P*-value was determined by Student's *t*-test. (**c**) Lysates of HEK293 cells expressing Myc epitope-tagged DNMT3L and wild-type or R882H forms of DNMT3A, as indicated, were subjected to immunoprecipitation (IP) with antibodies to Myc or to DNMT3A, and the resulting precipitates were subjected to immunoblot analysis (IB) with antibodies to DNMT3A. The position of DNMT3A (wild-type or mutant) is indicated by an open arrow.

responsible for a hereditary syndrome characterized by ICF (immunodeficiency, instability of the centromeric region of chromosomes and facial anomalies) in humans (Ehrlich, 2003). One of the mutation sites of DNMT3B (R823) associated with the ICF syndrome corresponds to the residue of DNMT3A (R882) shown to be mutated in this study.

The R882 residue of DNMT3A is considered to participate in the homodimerization and activation of the protein (Jia *et al*., 2007) (Figure 3a). To determine whether the R882H mutation affects the catalytic activity of DNMT3A, we expressed mutant and wild-type proteins separately in insect cells, purified them to near homogeneity and subjected them to an *in vitro* assay of methyltransferase activity with a synthetic substrate (Suetake *et al*., 2003). The catalytic activity of DNMT3A(R882H) was <50% of that of the wild-type protein (Figure 3b). DNMT3L acts as a coactivator for the methyltransferase activity of DNMT3A or DNMT3B through its association with the latter proteins (Jia *et al*., 2007). The R882H mutation did not affect the interaction of DNMT3A with DNMT3L in transfected mammalian cells (Figure 3c) or its sensitivity to DNMT3L as examined by the *in vitro* assay of methyltransferase activity (data not shown). These data thus suggested that the R882H mutation directly inhibits the enzymatic activity of DNMT3A.

Screening of another cohort of leukemia cases ($n = 54$) for mutant forms of *DNMT3A* revealed another two patients with a mutation of the same amino acid (R882H in one patient and R882C in the other) (Supplementary Table S4). Therefore, we identified a total of 3 cases with an R882 mutation (4.1%) among 74 cases of leukemia. Screening for mutations of *DNMT3B* failed to detect any somatic changes in the same individuals (data not shown), suggesting that *DNMT3A* is a preferential target in leukemia.

Multistep transformation in leukemia

Although >99% of nucleotide changes in the phase I data were also observed in the paired $CD4^+$ cells, it is unlikely that all of these changes are actually germline polymorphisms because they include established oncogenic mutations. They thus include 190 nucleotide changes previously described in cancer cells (Supplementary Table S5), such as those giving rise to NRAS(Q61H) in patient ID JM17 and to FLT3(D835Y) in patient ID JM08 (Figure 4a). Given that both NRAS(Q61H) and FLT3(D835Y) are well-characterized oncoproteins (Yamamoto *et al*., 2001), it is unlikely that these individuals harbored such nucleotide changes in the germ line. There are at least two possible explanations for these findings. First, it is possible that purification of the $CD4^+$ fraction was not efficient, with the result that this fraction was contaminated by $CD34^+$ cells. However, the *CD4* expression ratio for the $CD4^+$ and $CD34^+$ fractions of each individual was $\geqslant 17.1$ (median = 40.1) (Figure 4b), and contamination of the $CD4^+$ fraction with $CD34^+$ cells at such a level would not likely produce detectable changes in Sanger sequencing outputs (compare, for instance, the signal intensities of the normal and mutant alleles in Figure 4a).

Furthermore, although CD4 expression has been occasionally observed in AML blasts (Schwonzen *et al*., 2007), quantitation of *CD4* and *CD34* mRNA within our purified $CD34^+$ fractions failed to detect a significant level of the former message in the blasts (Supplementary Figure S6). Therefore, it is unlikely that contamination of $CD4^+$ leukemic blasts within the purified, control $CD4^+$ fraction substantially affected the sequencing results in our phase II experiment.

Rather, it is more likely that leukemia may develop in a stepwise manner with a substantial time interval