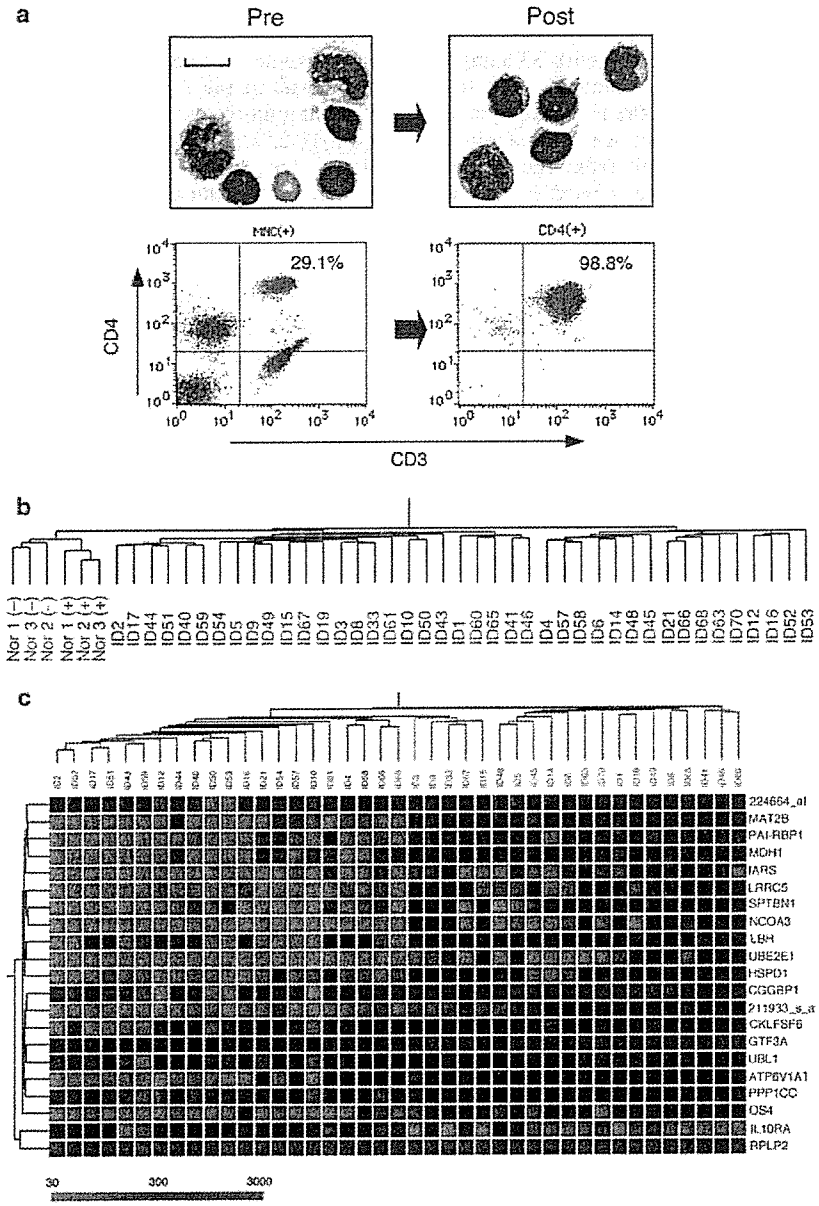

Figure 1 Purification and gene expression profiling of ATL cells. (**a**) MNCs isolated from the PB of a healthy individual were subjected to staining with Wright–Giemsa solution before (Pre) and after (Post) purification by affinity chromatography with antibodies to CD4 (upper panels). Scale bar, 10 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

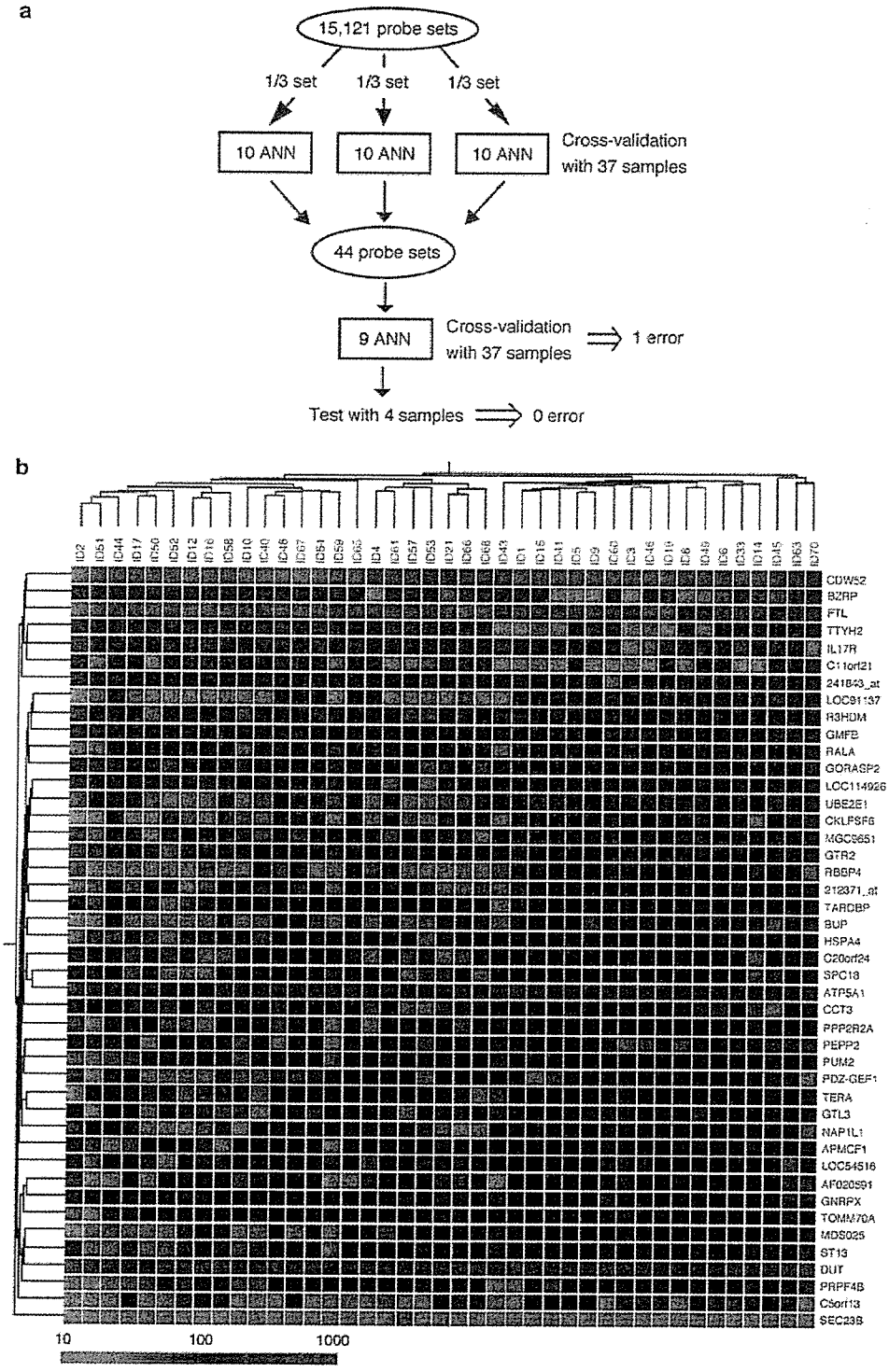

Figure 2 Schematic of the ANN analysis used for class prediction of ATL. (**a**) The 15 121 probe sets originally selected in Figure 1b were divided into three nonoverlapping groups, each of which was used as the input for 10 ANNs. We performed a 10% crossvalidation rotation with 37 samples, training 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

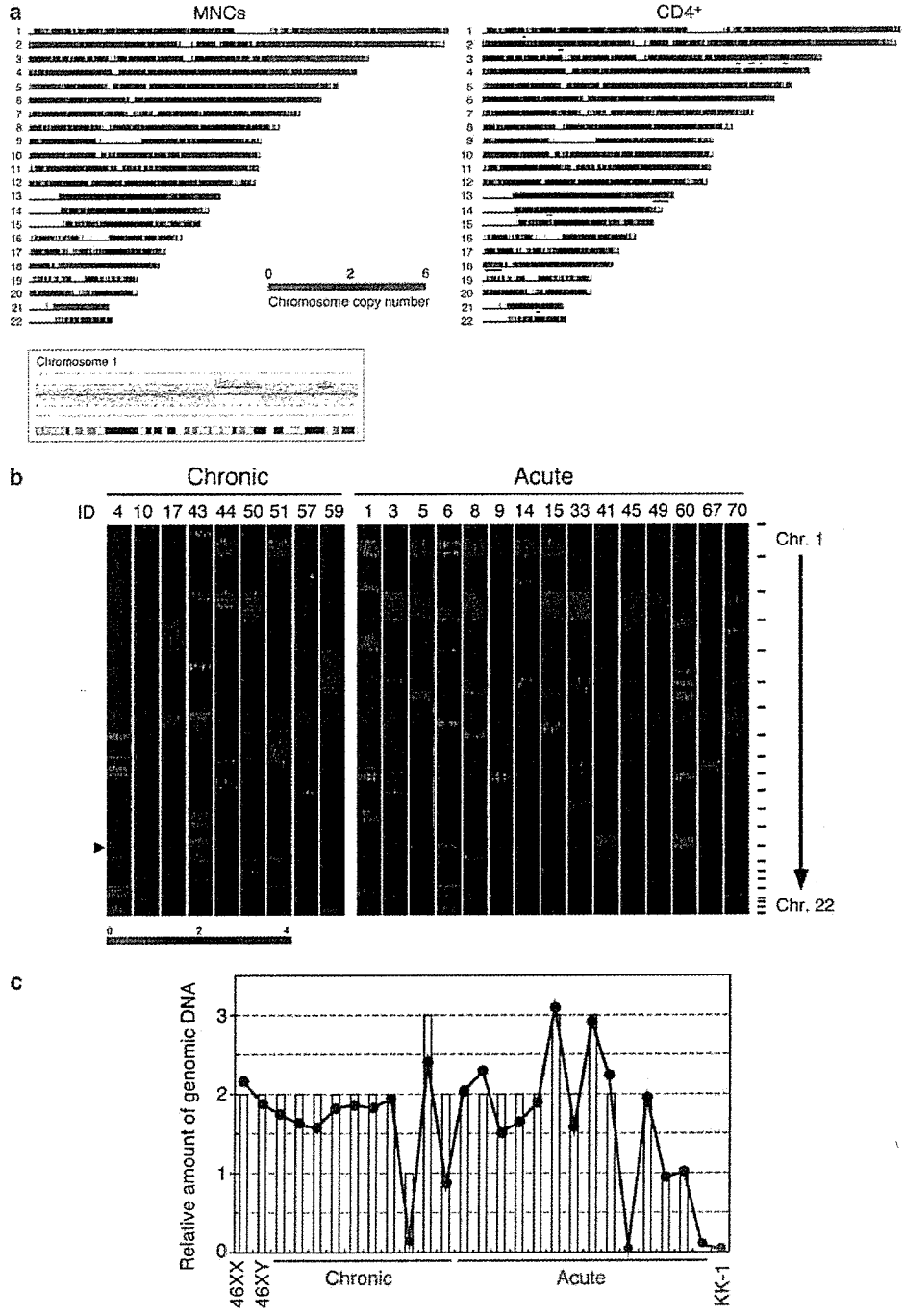

Figure 3 *Continued.*

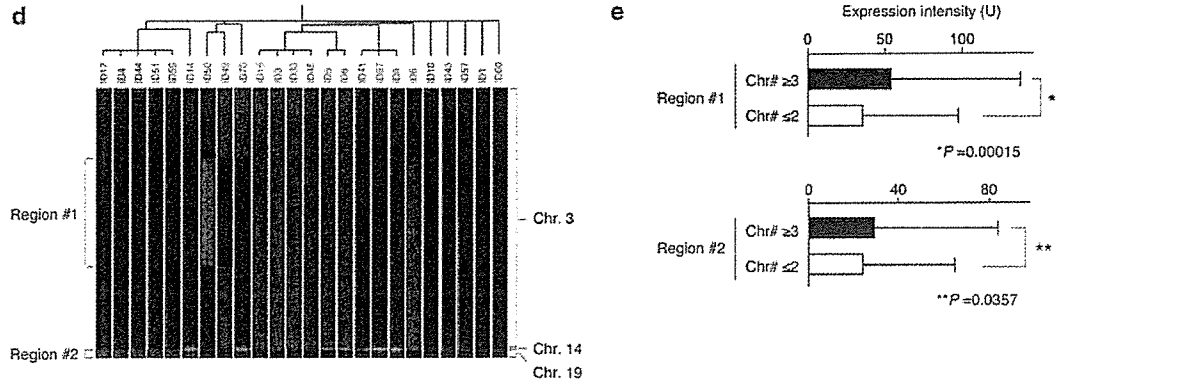

Figure 3 CNA analysis of purified ATL cells. (**a**) Inferred copy number for all SNP sites analysed in MNCs (left) and CD4$^+$ cells (right) isolated from an ATL patient in the acute stage (ID6). Copy number is color-coded according to the indicated scheme. Chromosomal regions with an aberrant copy 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

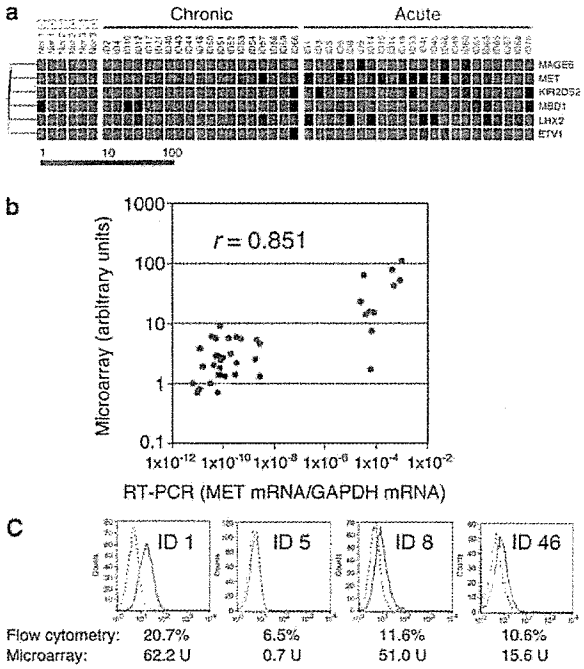

Figure 4 Acute stage-specific expression of *MET* in ATL. (a) Expression profiles of the six acute stage-specific probe sets. The expression level of each probe set is colored according to the indicated scale. Annotations and expression intensities for the probe sets are presented in Supplementary Table 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.

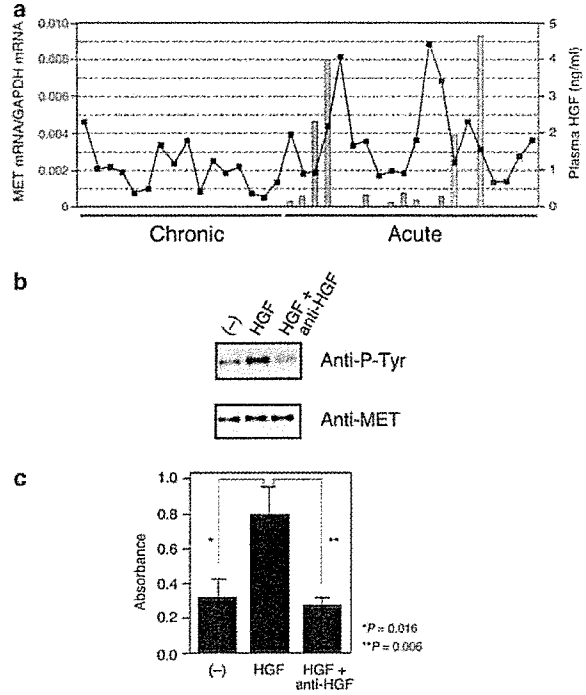

Figure 5 HGF-MET activity induces proliferation in ATL cells. (a) Comparison of the plasma concentration of HGF in ATL patients (as measured by enzyme-linked immunosorbent assay) with the corresponding amount of *MET* mRNA in leukemic cells (as determined by RT–PCR). (b) Tyrosine phosphorylation of MET 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 ± 2254.0 (mean ± s.d.) and 4678.5 ± 2874.6 SNP sites showing $\geqslant 3n$ ploidy as well as 1039.9 ± 2310.0 and 927.1 ± 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

expression of *MET* in the leukemic cells (Figure 5a). The increased concentration of HGF might therefore confer a growth advantage on ATL cells after they upregulate the expression of MET. Given that the JAK-STAT signaling pathway is activated in the leukemic cells of patients with advanced ATL (Migone *et al.*, 1995), it may be relevant that binding sites for STAT1 or STAT3 are present in the promoter regions of five (including *MET*) of the six acute stage-specific genes identified in the present study (Figure 4a). We did not detect a significant difference in DNA content (in our data set) for the *MET* locus between chronic and acute stages of ATL. It is thus possible that JAK-STAT signaling contributes to transcriptional activation of *MET*.

Given that our data are derived from purified ATL cells, they can be further used to characterize ATL in various ways. For instance, we attempted to isolate genes whose expression was linked to the presence of hypercalcemia in the study patients (data not shown); such genes included that for parathyroid hormone-like hormone (GenBank accession no. BC005961), which has been shown to be responsible for many instances of humoral hypercalcemia in individuals with cancer including ATL (Broadus *et al.*, 1988; Motokura *et al.*, 1988).

We have demonstrated the existence of an HGF-MET paracrine loop specific to the acute stage of ATL. Given that ligation of MET by HGF promoted the proliferation of ATL cells, activation of the HGF-MET signaling pathway is a candidate molecular mechanism for stage progression in ATL. Furthermore, our observation that this mitogenic effect was blocked by antibodies to HGF provides potential new targets for ATL therapy.

Materials and methods

Expression profiling

All clinical specimens were obtained with written informed consent, and the study was approved by the ethics committees of both Jichi Medical University and Nagasaki University. The diagnosis of ATL in all cases was based on clinical features, immunophenotypes of leukemic cells, and the monoclonal integration of HTLV-1 proviral DNA into the genome of leukemic cells (Shimoyama, 1991). MNCs isolated from PB were labeled with magnetic bead-conjugated mouse monoclonal antibodies to CD4 (CD4 MicroBeads, Miltenyi Biotec, Auburn, CA, USA). For PHA stimulation, purified $CD4^+$ cells from healthy individuals were incubated for 48 h in Rosewell's Park Memorial Institute media (RPMI) 1640 medium (Invitrogen, Carlsbad, CA, USA) supplemented with 10% fetal bovine serum (FBS) and PHA-P (8 μg/ml) (Sigma, St Louis, MO, USA).

Column fractionation of MNCs, RNA preparation, and hybridization with HGU133A & B microarrays (Affymetrix, Santa Clara, CA, USA) were performed as described previously (Choi *et al.*, 2004). The mean expression intensity of the internal positive control probe sets (http://www.affymetrix.com/support/technical/mask_files.affx) was set to 500 U in each hybridization, and the fluorescence intensity of each test gene was normalized accordingly. All HGU133A & B microarray data are available at the Gene Expression Omnibus web site (http://www.ncbi.nlm.nih.gov/geo) under the accession number GSE1466.

Student's *t*-test with the Benjamini and Hochberg false discovery rate option was performed with GeneSpring 7.0 software (Silicon Genetics). Effect size was defined as an absolute difference in mean expression intensity between a given pair of classes (Dhanasekaran *et al.*, 2001). Education of and prediction by our ANN were performed with NeuralWorks Professional II Plus v.5.3 software (NeuralWare, Carnegie, PA, USA) as described previously (O'Neill and Song, 2003).

Analysis of CNA

Genomic DNA was obtained from purified $CD4^+$ ATL cells ($n=24$) and from MNCs of patient ID6 with the use of a QIAmp DNA Mini Kit (Qiagen, Valencia, CA, USA). Each DNA sample (250 ng) was digested with *Hind*III, ligated to Adaptor-Hind (Affymetrix), amplified by PCR, and subjected to hybridization with Mapping 50K Hind 240 arrays (Affymetrix). Chromosome copy number at each SNP site was inferred from hybridization signal intensity on the arrays with the use of CNAG software (http://www.genome.umin.jp) (Nannya *et al.*, 2005). For a normal reference, we used array data of PB MNCs isolated from four healthy volunteers. Assessment of copy number for all SNP sites is demonstrated in Supplementary Table 6. The raw data of Mapping 50K Hind 240 arrays is available upon request. Statistical analysis of copy number was performed with GeneSpring 7.0. Alterations in the amount of genomic DNA were confirmed by quantitative real-time PCR with an ABI PRISM 7700 sequence detection system (PE Applied Biosystems, Foster City, CA, USA). The oligonucleotide primers were 5′-AGCATGTCCACAAATGGCCTTTGG-3′ and 5′-CAGTTTTCCTGTCATGGGAAAGGG-3′ for a region of chromosome 6, and 5′-CTGACCTGCCGTCTA GAAAAACCT-3′ and 5′-CAGGAAATGAGCTTGACAA AGTGG-3′ for the glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene.

MET expression

ATL cells were stained with rabbit polyclonal antibodies to MET (Santa Cruz Biotechnology, Santa Cruz, CA, USA) or with mouse monoclonal antibodies to CD3 and to CD4 (both from BD Biosciences, San Diego, CA, USA) and were then subjected to flow cytometry with a FACScan instrument (BD Biosciences). The concentration of HGF in plasma was determined with a Quantikine ELISA kit for human HGF (R&D Systems, Minneapolis, MN, USA).

For quantitative RT–PCR analysis of *MET* expression, portions of nonamplified cDNA were subjected to PCR with a QuantiTect SYBR Green PCR kit (Qiagen). The oligonucleotide primers for PCR were 5′-GTCAGTGGTGGACCTG ACCT-3′ and 5′-TGAGCTTGACAAAGTGGTCG-3′ for *GAPDH* cDNA, and 5′-ACTTGTTGCAAGGGAGAAGAC TCC-3′ and 5′-AGCGTTCACATGGACATAGTGCTC-3′ for *MET* cDNA.

KK-1 cell experiments

KK-1 cells were maintained in RPMI 1640 medium supplemented with 10% FBS and recombinant human interleukin-2 (10 ng/ml) (Sigma). For immunoblot analysis, cells were cultured for 48 h without FBS and interleukin-2 and then incubated for 10 min with recombinant human HGF (50 ng/ml) (Sigma) either alone or together with rabbit polyclonal antibodies to HGF (10 μg/ml) (Montesano *et al.*, 1991). The

cells were then lysed and subjected to immunoblot analysis with mouse monoclonal antibodies to phosphotyrosine (4G10, Upstate Biotechnology, Charlottesville, VA, USA) or with rabbit polyclonal antibodies to MET (#05–237, Upstate Biotechnology) as described previously (Yamashita *et al.*, 2001). For assay of cell proliferation, serum-deprived KK-1 cells were cultured for 24 h at a density of 5×10^5/ml with HGF (50 ng/ml) either alone or together with antibodies to HGF (10 μg/ml) and were then mixed with MTS [3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2*H*-tetrazolium, inner salt]. Cell proliferation was measured with a CellTiter 96 Aqueous One Solution Cell Proliferation Assay (Promega, Madison, WI, USA).

Abbreviations

ATL, Adult T-cell leukemia; PCR, polymerase chain reaction; PHA, phytohemagglutinin; ANN, artificial neural network; CNA, copy number alterations; HGF, hepatocyte growth factor.

Acknowledgements

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

References

Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D *et al.* (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci USA* **96**: 6745–6750.

Ariyama Y, Mori T, Shinomiya T, Sakabe T, Fukuda Y, Kanamaru A *et al.* (1999). Chromosomal imbalances in adult T-cell leukemia revealed by comparative genomic hybridization: gains at 14q32 and 2p16–22 in cell lines. *J Hum Genet* **44**: 357–363.

Bottaro DP, Rubin JS, Faletto DL, Chan AM, Kmiecik TE, Vande Woude GF *et al.* (1991). Identification of the hepatocyte growth factor receptor as the c-met proto-oncogene product. *Science* **251**: 802–804.

Broadus AE, Mangin M, Ikeda K, Insogna KL, Weir EC, Burtis WJ *et al.* (1988). Humoral hypercalcemia of cancer. Identification of a novel parathyroid hormone-like peptide. *N Engl J Med* **319**: 556–563.

Cesarman E, Chadburn A, Inghirami G, Gaidano G, Knowles DM. (1992). Structural and functional analysis of oncogenes and tumor suppressor genes in adult T-cell leukemia/lymphoma shows frequent p53 mutations. *Blood* **80**: 3205–3216.

Choi YL, Makishima H, Ohashi J, Yamashita Y, Ohki R, Koinuma K *et al.* (2004). DNA microarray analysis of natural killer cell-type lymphoproliferative disease of granular lymphocytes with purified CD3(−)CD56(+) fractions. *Leukemia* **18**: 556–565.

Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K *et al.* (2001). Delineation of prognostic biomarkers in prostate cancer. *Nature* **412**: 822–826.

Edlich RF, Arnette JA, Williams FM. (2000). Global epidemic of human T-cell lymphotropic virus type-I (HTLV-I). *J Emerg Med* **18**: 109–119.

Funakoshi H, Nakamura T. (2003). Hepatocyte growth factor: from diagnosis to clinical applications. *Clin Chim Acta* **327**: 1–23.

Hatta Y, Hirama T, Miller CW, Yamada Y, Tomonaga M, Koeffler HP. (1995). Homozygous deletions of the p15 (MTS2) and p16 (CDKN2/MTS1) genes in adult T-cell leukemia. *Blood* **85**: 2699–2704.

Imaizumi Y, Murota H, Kanda S, Hishikawa Y, Koji T, Taguchi T *et al.* (2003). Expression of the c-Met proto-oncogene and its possible involvement in liver invasion in adult T-cell leukemia. *Clin Cancer Res* **9**: 181–187.

Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. (2004). dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* **20**: 1233–1240.

Lockwood WW, Chari R, Chi B, Lam WL. (2005). Recent advances in array comparative genomic hybridization technologies and their applications in human genetics. *Eur J Hum Genet* **14**: 139–148.

Matsumoto K, Okazaki H, Nakamura T. (1992). Up-regulation of hepatocyte growth factor gene expression by interleukin-1 in human skin fibroblasts. *Biochem Biophys Res Commun* **188**: 235–243.

Matsuoka M. (2003). Human T-cell leukemia virus type I and adult T-cell leukemia. *Oncogene* **22**: 5131–5140.

Migone TS, Lin JX, Cereseto A, Mulloy JC, O'Shea JJ, Franchini G *et al.* (1995). Constitutively activated Jak-STAT pathway in T cells transformed with HTLV-I. *Science* **269**: 79–81.

Montesano R, Matsumoto K, Nakamura T, Orci L. (1991). Identification of a fibroblast-derived epithelial morphogen as hepatocyte growth factor. *Cell* **67**: 901–908.

Motokura T, Fukumoto S, Takahashi S, Watanabe T, Matsumoto T, Igarashi T *et al.* (1988). Expression of parathyroid hormone-related protein in a human T cell lymphotrophic virus type I-infected T cell line. *Biochem Biophys Res Commun* **154**: 1182–1188.

Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A *et al.* (2005). A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res* **65**: 6071–6079.

Nosaka K, Maeda M, Tamiya S, Sakai T, Mitsuya H, Matsuoka M. (2000). Increasing methylation of the CDKN2A gene is associated with the progression of adult T-cell leukemia. *Cancer Res* **60**: 1043–1048.

O'Neill MC, Song L. (2003). Neural network analysis of lymphoma microarray data: prognosis and diagnosis near-perfect. *BMC Bioinformatics* **4**: 13.

Oshiro A, Tagawa H, Ohshima K, Karube K, Uike N, Tashiro Y *et al.* (2006). Identification of subtype-specific genomic alterations in aggressive adult T-cell leukemia/lymphoma. *Blood* **107**: 4500–4507.

Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC. (1980). Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc Natl Acad Sci USA* **77**: 7415–7419.

Reiner A, Yekutieli D, Benjamini Y. (2003). Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19**: 368–375.

Sasaki H, Nishikata I, Shiraga T, Akamatsu E, Fukami T, Hidaka T *et al.* (2005). Overexpression of a cell adhesion molecule, TSLC1, as a possible molecular marker for acute-type adult T-cell leukemia. *Blood* **105**: 1204–1213.

Shimoyama M. (1991). Diagnostic criteria and classification of clinical subtypes of adult T-cell leukaemia-lymphoma. A report from the Lymphoma Study Group (1984–1987). *Br J Haematol* **79**: 428–437.

Tajima K. (1990). The 4th nation-wide study of adult T-cell leukemia/lymphoma (ATL) in Japan: estimates of risk of ATL and its geographical and clinical features. The T- and B-cell Malignancy Study Group. *Int J Cancer* **45**: 237–243.

Tamiya S, Etoh K, Suzushima H, Takatsuki K, Matsuoka M. (1998). Mutation of CD95 (Fas/Apo-1) gene in adult T-cell leukemia cells. *Blood* **91**: 3935–3942.

Tamura M, Arakaki N, Tsubouchi H, Takada H, Daikuhara Y. (1993). Enhancement of human hepatocyte growth factor production by interleukin-1 alpha and -1 beta and tumor necrosis factor-alpha by fibroblasts in culture. *J Biol Chem* **268**: 8140–8145.

Tsukasaki K, Krebs J, Nagai K, Tomonaga M, Koeffler HP, Bartram CR *et al.* (2001). Comparative genomic hybridization analysis in adult T-cell leukemia/lymphoma: correlation with clinical course. *Blood* **97**: 3875–3881.

Tsukasaki K, Tanosaki S, DeVos S, Hofmann WK, Wachsman W, Gombart AF *et al.* (2004). Identifying progression-associated genes in adult T-cell leukemia/lymphoma by using oligonucleotide microarrays. *Int J Cancer* **109**: 875–881.

Uchiyama T, Yodoi J, Sagawa K, Takatsuki K, Uchino H. (1977). Adult T-cell leukemia: clinical and hematologic features of 16 cases. *Blood* **50**: 481–492.

Wano Y, Hattori T, Matsuoka M, Takatsuki K, Chua AO, Gubler U *et al.* (1987). Interleukin 1 gene expression in adult T cell leukemia. *J Clin Invest* **80**: 911–916.

Yamada Y, Hatta Y, Murata K, Sugawara K, Ikeda S, Mine M *et al.* (1997). Deletions of p15 and/or p16 genes as a poor-prognosis factor in adult T-cell leukemia. *J Clin Oncol* **15**: 1778–1785.

Yamada Y, Ohmoto Y, Hata T, Yamamura M, Murata K, Tsukasaki K *et al.* (1996). Features of the cytokines secreted by adult T cell leukemia (ATL) cells. *Leuk Lymphoma* **21**: 443–447.

Yamashita Y, Kajigaya S, Yoshida K, Ueno S, Ota J, Ohmine K *et al.* (2001). Sak serine/threonine kinase acts as an effector of Tec tyrosine kinase. *J Biol Chem* **276**: 39012–39020.

Yoshida M, Miyoshi I, Hinuma Y. (1982). Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. *Proc Natl Acad Sci USA* **79**: 2031–2035.

Zhao X, Weir BA, LaFramboise T, Lin M, Beroukhim R, Garraway L *et al.* (2005). Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. *Cancer Res* **65**: 5561–5570.

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

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

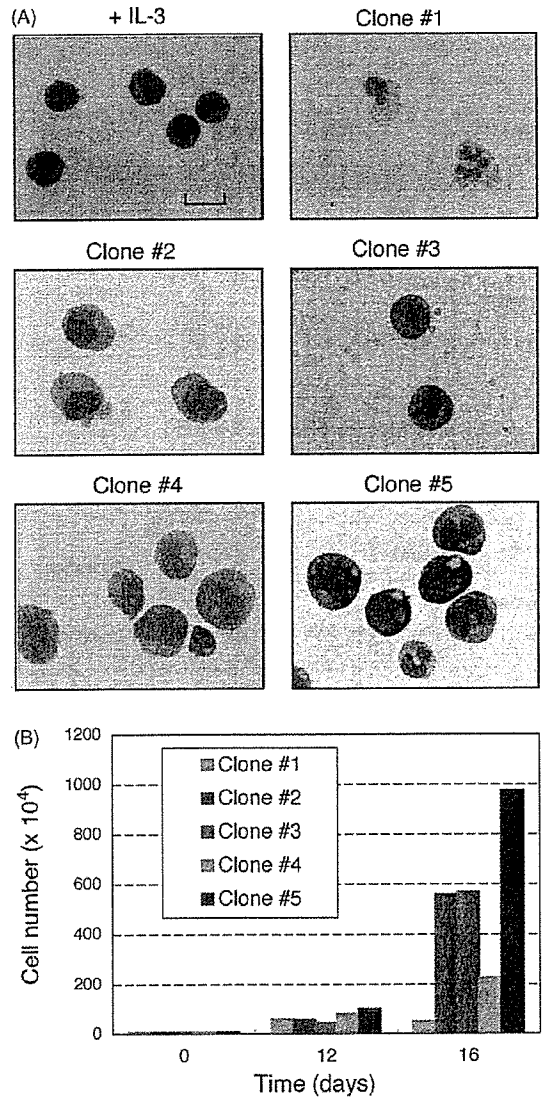

Fig. 1. Isolation of 32D-mCAT clones that proliferate in the presence of G-CSF. (A) 32D-mCAT cells cultured with IL-3 (+IL-3) as well as 32D-mCAT clones infected with the retroviral expression library (clones #1, #2, #3, #4, and #5) and cultured for 16 days with G-CSF were stained with Wright-Giemsa 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.

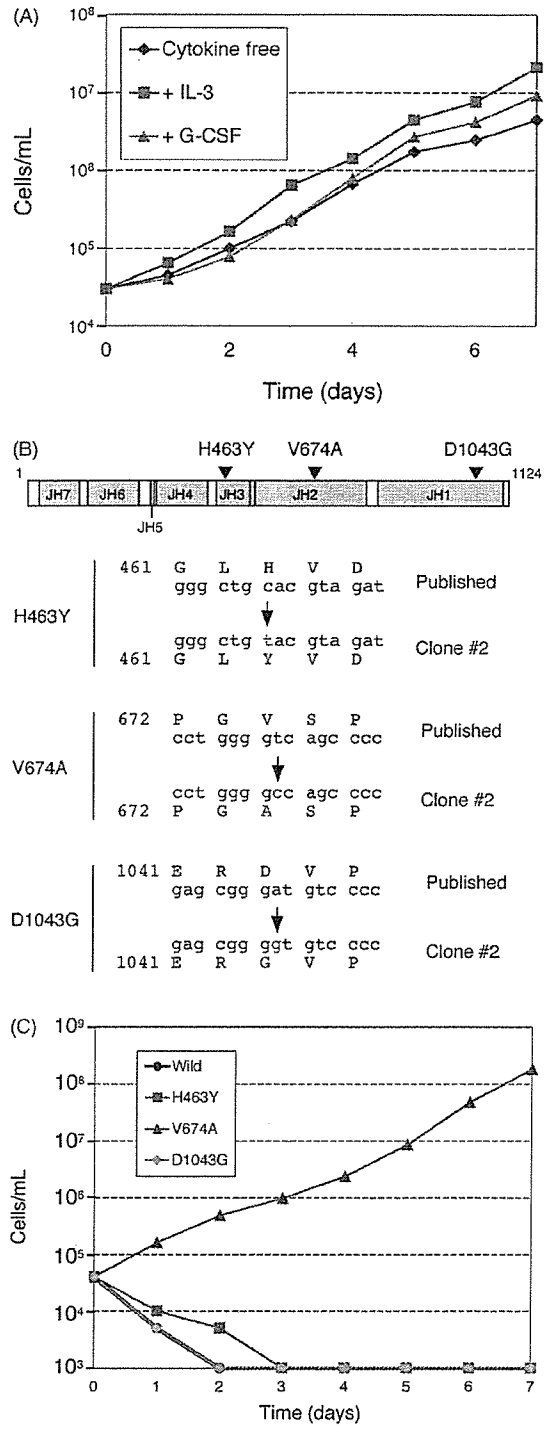

Fig. 2. Expression of JAK3(V674A) abrogates cytokine dependency of 32D-mCAT cell growth. (A) 32D-mCAT cells were infected with recombinant retroviruses containing the *JAK3* cDNA isolated from 32D-mCAT clone #2. The cells were then cultured in the absence or presence of IL-3 or G-CSF, and cell number 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).

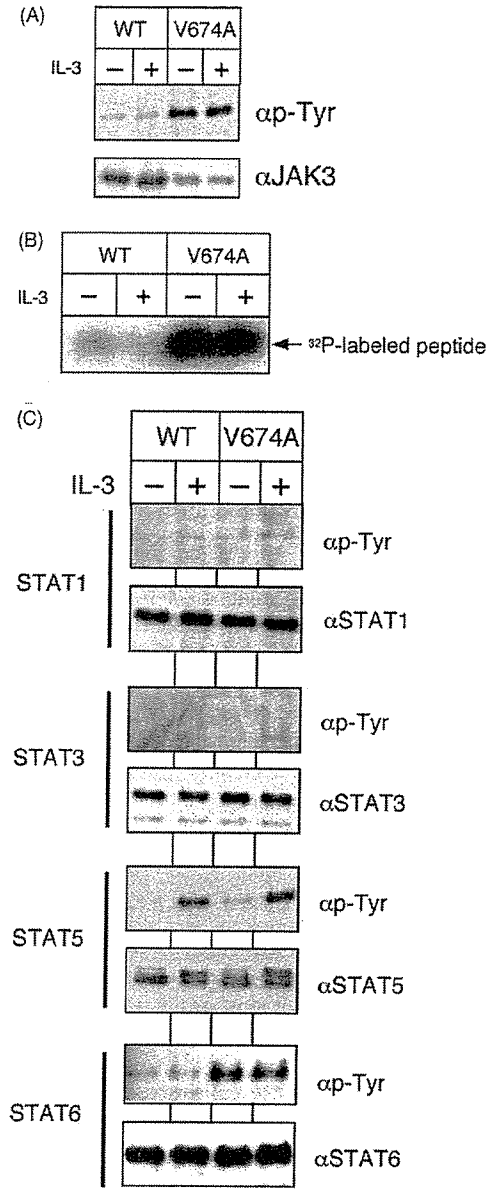

Fig. 3. Constitutive activation of JAK3(V674A). (A) 32D-mCAT cells expressing wild-type JAK3 (WT) or JAK3(V674A) were cultured in the absence of cytokine or FBS for 1 h and then incubated with (+) or without (−) IL-3 for 10 min. Cell lysates were subjected to immunoprecipitation with antibodies to JAK3. 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.

CANCER GENOMICS & PROTEOMICS 3: 169-182 (2006)

Gene Expression Profiles of CD133-positive Fractions Predict the Survival of Individuals with Acute Myeloid Leukemia

YOSHIHIRO YAMASHITA[1], JUN OHASHI[2], YUJI HIRAI[3], YOUNG LIM CHOI[1], RURI KANEDA[1], SHIN-ICHIRO FUJIWARA[1,4], YUKIHIRO ARAI[5], MIYUKI AKUTSU[6], CHIZUKO TSUTSUMI[7], YASUSHI MIYAZAKI[7], KENSUKE USUKI[8], MASANAO TERAMURA[3], KINUKO MITANI[5], YASUHIKO KANO[6], MICHAEL C. O'NEILL[9], AKIO URABE[8], MASAO TOMONAGA[7], KEIYA OZAWA[4] and HIROYUKI MANO[1,10]

Divisions of [1]Functional Genomics and [4]Hematology, Jichi Medical University, Tochigi 329-0498; [2]Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo 113-0033; [3]Department of Hematology, Tokyo Women's Medical University, Tokyo 162-8666; [5]Department of Hematology, Dokkyo University School of Medicine, Tochigi 321-0293; [6]Tochigi Cancer Center, Tochigi 320-0834; [7]Department of Hematology and Molecular Medicine Unit, Nagasaki University, Nagasaki 852-8523; [8]Department of Hematology, Kanto Medical Center NTT EC, Tokyo 141-8625, Japan; [9]Department of Biological Sciences, University of Maryland, Baltimore, Maryland 21250, U.S.A.; [10]CREST, Japan Science and Technology Agency, Saitama 332-0012, Japan

Abstract. *Background: The current classification of acute myeloid leukemia (AML) is based predominantly on the cytogenetic abnormalities and morphology of the malignant blasts and is not always helpful for optimization of the treatment strategy. Gene expression profiles of AML blasts were obtained and a gene expression-based means of predicting the outcome of AML patients was developed. Materials and Methods: CD133-positive hematopoietic stem cell-like fractions were purified from the bone marrow of 99 individuals with AML-related disorders and the expression profiles of ~33,000 human transcripts in these cells were characterized with the use of DNA microarray analysis. Results: The comparison of the expression data between individuals with short- or long-term survival by application of Cox's proportional hazard model led to the identification of four genes, whose expression patterns discriminated between the two groups. The gene expression-based stratification (GES) system, based on a combination of the karyotype approach and the risk index calculated from the expression levels of the four outcome predictor genes, was developed to separate the patients into subgroups with distinct prognoses. Conclusion: DNA microarray analysis of purified fractions provides novel stratification schemes for AML based on the expression profiles of a handful of genes.*

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

Key Words: DNA microarray, prognosis prediction, gene expression profile.

Acute myeloid leukemia (AML) is characterized by clonal growth of immature leukemic blasts in the bone marrow (BM). Although current chemotherapeutic regimens induce an initial complete remission in >70% of affected individuals, the long-term survival of such patients remains poor (5-year survival rate of <30%) (1). Given that leukemic blasts of individuals with AML differ in their abilities to differentiate into cells of the granulocyte, monocyte, erythrocyte, or megakaryocyte lineages, the French-American-British Cooperative Group (FAB) established a classification scheme for AML (M0 to M7) based on blast morphology and differentiation commitment (2). Although some FAB subtypes have proved to be related to good or poor prognosis, the clinical relevance of this classification scheme is complicated by other clinical parameters. A preceding history of myelodysplastic syndrome (MDS) or anticancer treatment, for example, is an important indicator of poor prognosis (3).

One of the most robust predictors of AML prognosis is the blast karyotype (1, 4). A good prognosis is thus predicted from the presence in leukemic clones of t(8;21), t(15;17), or inv(16) chromosomal rearrangements ("favorable" karyotype), whereas -7/7q-, 11q23, or more complex (affecting three or more chromosomes) abnormalities are indicative of a poor outcome ("adverse" karyotype). Other karyotypes are classified as "intermediate." The World Health Organization (WHO) has proposed a classification of AML that separates

individuals with cytogenetic (or molecular) evidence of t(8;21), t(15;17), inv(16)/t(16;16), or 11q23 abnormalities into distinct subcategories (5). However, this classification is of little help for predicting the prognosis of AML patients with a normal karyotype, who constitute ~50% of the AML population. Given that AML patients with a normal karyotype are considered to be at intermediate risk, the corresponding leukemic blasts may harbor heterogeneous minor genomic rearrangements or mutations.

DNA microarray analysis has the potential to provide a stratification scheme for AML based on gene expression profiles and is able to predict the prognosis of each affected individual (6). To facilitate the development of such a genomics approach to the classification of human leukemias, a large-scale cell bank (the Blast Bank) was set up for the storage of CD133 (AC133)-positive hematopoietic stem cell (HSC)-like fractions purified from individuals with a wide range of leukemic disorders (7, 8). CD133 is specifically expressed on HSCs and hemangioblasts that are $CD34^+CD38^-$ (9, 10). There are at least two advantages to the use of such purified immature fractions for the characterization of AML. First, given that the proportion of leukemic blasts within BM varies substantially (20 to almost 100%) among patients and that leukemic blasts possess the ability to differentiate to various extents, a simple comparison of BM mononuclear cells (MNCs) among heterogeneous AML patients is likely to reveal a large number of changes in gene expression that reflect differences either in the percentage of blasts within BM, or in the differentiation ability of the blasts. Analysis of Blast Bank specimens should thus minimize such population-shift effects (7). Second, although leukemic cells in a given patient comprise a mixture of malignant clones at various levels of differentiation, they are thought to be generated from a small number of leukemic stem cells (LSCs), similar to the situation for normal hematopoiesis (11, 12). Characterization of the LSCs should provide insight into the molecular mechanisms of leukemogenesis. Given that such LSCs are exclusively $CD34^+CD38^-$, the Blast Bank may represent a diverse collection of LSC specimens. Cancer stem cells of malignant glioblastoma, but not their progenies, have also been shown to specifically express CD133 (13).

DNA microarray analysis was used here to characterize the expression profiles of ~44,000 probe sets in the $CD133^+$ fractions of 99 adults with AML-related disorders. Statistical analyses of the resulting large data set provided the basis for a new classification of AML that facilitates prediction of the long-term prognosis of affected individuals.

Materials and Methods

Purification of $CD133^+$ cells. Informed consent was obtained from each of the study subjects and the study was approved by the appropriate institutional review boards. Table I shows the clinical characteristics of the study subjects. MNCs were isolated by density gradient centrifugation from BM aspirates of each patient and were subjected to chromatography on a miniMACS column (Miltenyi Biotec, Auburn, CA, USA) with magnetic bead-conjugated monoclonal antibodies to CD133 (AC133 MicroBeads; Miltenyi Biotec) as described previously (7). In most instances, the $CD34^{hi}CD38^{lo}$ fraction constituted >90% of the eluate of the affinity column, as judged by flow cytometry.

Microarray analysis. Total RNA was extracted from the $CD133^+$ cell preparations with the use of an RNeasy Mini column and RNase-free DNase (Qiagen, Valencia, CA, USA). It was then subjected to two rounds of amplification of mRNA with T7 RNA polymerase (14); the high fidelity of the amplification step has been demonstrated previously (15). The resulting cRNAs were labeled with biotin and subjected to hybridization with GeneChip HGU133 A&B microarrays with the GeneChip system (Affymetrix, Santa Clara, CA, USA). The fidelity of the microarray data was confirmed by quantitative RT-PCR analysis.

Method for real-time RT-PCR. Portions of non-amplified cDNA were subjected to PCR with a QuantiTect SYBR Green PCR Kit (Qiagen). The amplification protocol comprised incubations at 94°C for 15 sec, 60°C for 30 sec and 72°C for 60 sec. Incorporation of the SYBR Green dye into the PCR products was monitored in real time with an ABI PRISM 7700 sequence detection system (PE Applied Biosystems, Foster City, CA, USA), thereby allowing determination of the threshold cycle (C_T) at which exponential amplification of products begins. The C_T values for cDNAs corresponding to *GAPDH* and the target genes were used to calculate the abundance of target gene mRNA relative to that of *GAPDH* mRNA. The primer sequences used for RT-PCR were as follows:

FGFR1:
Correlation coefficient between RT-PCR and microarray data=0.746.
Sense primer: 5'-CCACCAGAGTGATGTGTG GTCTTT-3'
Antisense primer: 5'-CATCATCATGTACAGCTC GTTGGT-3'

NRLN1
Correlation coefficient between RT-PCR and microarray data=0.638.
Sense primer: 5'-AATAGACGGAAAATGCTGCA AGGT-3'
Antisense primer: 5'-TGAGGTGGTCTCTCAGTCTCC AGT-3'

ZNF6
Correlation coefficient between RT-PCR and microarray data=0.779.
Sense primer: 5'-ATCTGGTGCAAAAACAGA AAGGTG-3'
Antisense primer: 5'-GGCGGGTTTATGCAGTATTT AACAG-3'

SCGF
Correlation coefficient between RT-PCR and microarray data=0.715.
Sense primer: 5'-TACTACGTCTGCGAGTTC CCCTTC-3'
Antisense primer: 5'-GCCCCTTCAAGGAAAGA CACTAAC-3'

HOXA9
Correlation coefficient between RT-PCR and microarray data=0.586.
Sense primer: 5'-CTCAGGTTGTTTATGAGG GGAAAA-3'
Antisense primer: 5'-ATGAATCTATGCATCCCC GAGAAC-3'

ANGPT1
Correlation coefficient between RT-PCR and microarray data=0.719.
Sense primer: 5'-GGCTGGGGAATGAGTTT ATTTTG-3'
Antisense primer: 5'-AAATCAGCACCGTGTAAG ATCAGG-3'

241376_at (EST)
Correlation coefficient between RT-PCR and microarray data=0.672.
Sense primer: 5'-CAACTCGAAGCTCAAA TACCCTCA-3'
Antisense primer: 5'-ACCGTTTATACACCAACGG TCACA-3'

FLJ13063
Correlation coefficient between RT-PCR and microarray data=0.502.
Sense primer: 5'-AGAGTTCTGCTGTGT CCTCTG-3'
Antisense primer: 5'-CAGGACAGTGCTGAAC CAATG-3'

TSPAN-2
Correlation coefficient between RT-PCR and microarray data=0.479.
Sense primer: 5'-GCAGTTGAAAATTGTG GGAAAGAG-3'
Antisense primer: 5'-CCCACACACAACTAGGA GAAGATG-3'

KIAA0830
Correlation coefficient between RT-PCR and microarray data=0.575.
Sense primer: 5'-TCCAGAGGCATCTGAGG GAAGTAG-3'
Antisense primer: 5'-ATGGCCATGAAGTATGA GATGGTG-3'

DJ79P11.1
Correlation coefficient between RT-PCR and microarray data=0.625.
Sense primer: 5'-CCATCCTGCAGTATAG ATGGGACA-3'
Antisense primer: 5'-GATTCAGGGCATAAGGC AAAACTC-3'

POU4F1
Correlation coefficient between RT-PCR and microarray data=0.643.
Sense primer: 5'-ATGAACAAGCCTGAGC TCTTCAAC-3'
Antisense primer: 5'-GAGAATTTCATCCGCT TCTGCTTC-3'

IGHM
Correlation coefficient between RT-PCR and microarray data=0.593.
Sense primer: 5'-CAGAAGAACATCGGAG ACCAGAGA-3'
Antisense primer: 5'-AACAAGCGTATACACAG CAAAGCA-3'

GAPDH
Sense primer: 5'-GTCAGTGGTGGACCT GACCT-3'
Antisense primer: 5'-TGAGCTTGACAAAGTG GTCG-3'

Statistical analysis. The mean expression intensity of the internal positive control probe sets (http://www.affymetrix.com/support/technical/mask_files.affx) was set to 500 units (U) in each hybridization and the fluorescence intensity of each test probe set was normalized accordingly. All normalized array data are available at the Gene Expression Omnibus web site (http://www.ncbi.nlm.nih.gov/geo) under the accession number GSE1427. All expression values were transformed to logarithms prior to statistical analyses. Hierarchical clustering of the data set and Student's *t*-test were performed with GeneSpring 7.0 software (Silicon Genetics, Redwood, CA, USA). Principal component analysis (16) and survival analyses were performed with the SAS software package (ver. 8.0.2).

Results

Purification of CD133$^+$ HSC-like fractions. The number of CD133$^+$ cells isolated from BM MNCs varied markedly among FAB subtypes of AML. For individuals whose leukemic blasts had a low differentiation capacity (FAB subtypes M0 to M2), for example, the blasts purified by CD133-based affinity chromatography constituted ≥30% with MNCs. In contrast, the blasts purified from individuals with FAB subtypes M3 (characterized by prominent differentiation to the promyelocyte level) or M5 (differentiation to the promonocyte level) constituted <1% of MNCs. In both the latter instances, large promyelocytes (M3) or large promonocytes (M5) constituted a major proportion of the BM MNCs, whereas the corresponding column eluates contained only medium-sized, agranular blasts with a high nucleus-to-cytoplasm ratio (Figure 1A).

The affinity column thus appeared to select for a minor population of cells at a highly immature level of differentiation. To verify that such cells in the column eluates did indeed represent leukemic clones, fluorescence *in situ* hybridization analysis was performed with the cells purified from several individuals. In the case of FAB subtype M3, described above, a t(15;17) was detected in 75% of the purified blasts (data not shown). Similarly, in an AML case characterized by 5q- or a case of chronic myelogenous leukemia characterized by t(9;22), the corresponding abnormality was detected in >90 or 70.7% (8) of cells in the column eluates, respectively. These findings indicate that most of the cells in the column eluates were malignant clones, especially given the only moderate sensitivity of fluorescence *in situ* hybridization.

Gene expression profiles of the Blast Bank specimens. All the specimens were collected from the patients before initiation of chemotherapy. The expression profiles of >44,000 probe

Table I. *Clinical characteristic of the study subjects.*

Blast Bank ID	Age (yr)	Gender	Disease	FAB subtype	Karyotype
ID020	66	Female	AML evolved from MDS	M2	Normal
ID023	34	Female	*de novo* AML	M6	Others>=3
ID026	46	Male	*de novo* AML	M3	Others≤2
ID027	49	Male	AML evolved from MDS	M2	Normal
ID028	62	Male	AML evolved from MDS	M2	Normal
ID032	75	Male	RAEB		Others>=3
ID035	61	Male	*de novo* AML	M2	Others≤2
ID036	74	Male	*de novo* AML	M0	Normal
ID042	67	Female	AML evolved from MDS	M2	Others>=3
ID045	69	Male	RAEB		+8
ID046	84	Male	*de novo* AML	M2	–7
ID051	42	Male	RAEB		Others>=3
ID054	84	Male	AML evolved from MDS	M2	–7
ID062	72	Male	*de novo* AML	M0	Normal
ID063	67	Female	RAEB		Others>=3
ID066	73	Male	RAEB		t(8;21)
ID076	37	Male	*de novo* AML	M2	t(8;21)
ID077	55	Male	RAEB		+8
ID083	64	Female	*de novo* AML	M4	Normal
ID087	66	Male	*de novo* AML	M5	Normal
ID093	53	Female	*de novo* AML	M5	Others≤2
ID098	66	Male	AML evolved from MDS	M0	Normal
ID104	72	Male	*de novo* AML	M0	Others>=3
ID107	48	Female	*de novo* AML	M5	Others≤2
ID109	86	Male	*de novo* AML	M1	Normal
ID127	41	Male	*de novo* AML	M2	Normal
ID139	50	Male	*de novo* AML	M1	Others≤2
ID142	38	Male	*de novo* AML	M2	t(8;21)
ID148	74	Male	AML evolved from MDS	M2	Normal
ID154	49	Male	RAEB		Normal
ID174	51	Male	*de novo* AML	M0	Normal
ID180	47	Male	*de novo* AML	M4	Normal
ID183	50	Female	*de novo* AML	M2	t(8;21)
ID188	59	Male	*de novo* AML	M2	t(8;21)
ID195	61	Female	*de novo* AML	M2	t(8;21)
ID205	39	Male	*de novo* AML	M1	Normal
ID215	52	Male	RAEB		Others>=3
ID226	52	Male	*de novo* AML	M2	t(8;21)
ID227	29	Male	*de novo* AML	M2	Normal
ID234	68	Male	RAEB		Others≤2
ID239	48	Male	*de novo* AML	M2	t(8;21)
ID243	54	Female	RAEB		Others>=3
ID262	70	Female	*de novo* AML	M2	Others≤2
ID265	65	Male	*de novo* AML	M7	Normal
ID266	38	Male	*de novo* AML	M4	Normal
ID267	80	Male	*de novo* AML	M2	Normal
ID269	32	Female	*de novo* AML	M4	Others≤2
ID270	46	Female	*de novo* AML	M1	Others≤2
ID272	57	Female	*de novo* AML	M4	+8
ID274	67	Male	*de novo* AML	M1	–7
ID275	70	Male	AML evolved from MDS	M2	Others>=3
ID276	64	Male	de novo AML	M2	–7, >3
ID277	23	Female	*de novo* AML	M4	inv16
ID278	16	Female	de novo AML	M4	Normal
ID279	68	Male	*de novo* AML	M2	t(8;21)
ID280	68	Male	*de novo* AML	M1	–7
ID288	61	Male	de novo AML	M5	Normal
ID292	74	Male	*de novo* AML	M6	Normal
ID305	63	Male	*de novo* AML	M2	Normal
ID306	43	Male	*de novo* AML	M1	Normal
ID310	75	Male	*de novo* AML	M0	Normal
ID313	88	Female	AML evolved from MDS	M1	Others>=3
ID314	55	Female	*de novo* AML	M5	+8
ID315	45	Female	RAEB		Others>=3
ID316	30	Male	*de novo* AML	M1	Others>=3
ID317	36	Male	*de novo* AML	M2	t(8;21)
ID318	53	Male	*de novo* AML	M2	Normal
ID319	47	Male	*de novo* AML	M6	Others>=3
ID321	68	Female	*de novo* AML	M2	t(8;21)
ID325	49	Male	*de novo* AML	M2	Others>=3
ID326	68		*de novo* AML	M1	Normal
ID329	79	Female	*de novo* AML	M2	Normal
ID338	57	Male	RAEB		Normal
ID339	72	Female	*de novo* AML	M4	inv16
ID347	52	Male	*de novo* AML	M2	Others≤2
ID349	70	Male	*de novo* AML	M6	Others>=3
ID355	69	Male	*de novo* AML	M2	t(8;21)
ID362	59	Female	*de novo* AML	M2	Normal
ID363	67	Male	*de novo* AML	M0	Others>=3
ID375	32	Male	*de novo* AML	M2	Normal
ID376	23	Male	*de novo* AML	M0	Others≤2
ID378	28	Female	*de novo* AML	M2	t(8;21)
ID379	62	Male	*de novo* AML	M2	Normal
ID380	51	Female	*de novo* AML	M6	Others>=3
ID382	49	Female	AML evolved from MDS	M2	Others>=3
ID385	79	Male	RAEB		Normal
ID388	45	Male	*de novo* AML	M5	inv16
ID395	80	Male	AML evolved from MDS	M0	Normal
ID402	63	Male	*de novo* AML	M1	Others≤2
ID409	72	Female	*de novo* AML	M1	Others≤2
ID410	67	Female	RAEB		Normal
ID413	61	Female	*de novo* AML	M2	Normal
ID414	61	Male	*de novo* AML	M0	Normal
ID415	85	Female	*de novo* AML	M2	Others>=3
ID416	42	Female	*de novo* AML	M2	Normal
ID418	57	Female	*de novo* AML	M2	Normal
ID421	81	Female	RAEB		Normal
ID427	62	Male	*de novo* AML	M5	Normal
ID430	66	Female	RAEB		Others≤2

AML=acute myeloid leukemia; MDS=myelodysplastic syndrome; FAB=French-American-British Cooperative Group; RAEB=refractory anemia with excess of blasts.