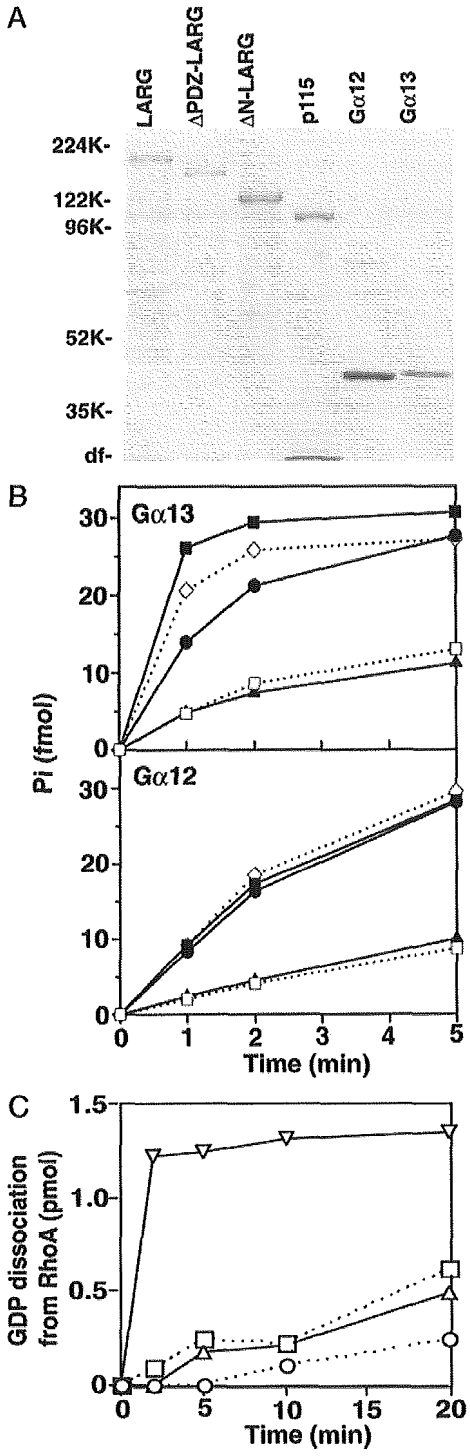


Fig. 2. GAP activity of LARG for Gα12/13 and the regulation of RhoGEF activity of LARG by Gα12/13. (*A*) Coomassie brilliant blue staining of Gα12/13 and RhoGEFs. Purified Gα12, Gα13, LARG, and p115RhoGEF (50 pmol each) were separated by SDS/PAGE and stained by Coomassie brilliant blue. (*B*) Stimulation of GTPase activity of Gα12 and Gα13 by LARG. Hydrolysis of GTP bound to Gα12 or Gα13 was measured at 15°C without (□) or with (■) 25 nM LARG, 25 nM ΔPDZ-LARG (●), 25 nM ΔN-LARG (▲), or 25 nM p115RhoGEF (◇). (*C*) Stimulation of the RhoGEF activity of LARG. Dissociation of GDP from RhoA was measured at 20°C: ○, control; □, 25 nM ΔPDZ-LARG; △, 25 nM ΔPDZ-LARG + 80 nM AlF_4^--activated Gα12; and ▽, 25 nM ΔPDZ-LARG + 80 nM AlF_4^--activated Gα13.

and Gα12 was almost similar to the level with these RhoGEFs and Gα13. We could not detect similar synergistic SRF activation by using p115RhoGEF in the assay. The results suggest that PDZ-RhoGEF or LARG may transduce the signal from both Gα12 and Gα13 to Rho activation. In this study, we focused on the function of LARG in G12/13-mediated signaling.

We examined the biochemical interaction of Gα12/13 with LARG *in vitro* by using purified components. Gα subunits and RhoGEFs were expressed in and purified from Sf9 cells (Fig. 2*A*). As shown in Fig. 2*B*, the constructs of LARG that contain the RGS domain demonstrated GAP activity for Gα12 or Gα13 similar to p115RhoGEF. However, a construct of LARG lacking the RGS domain did not show any GAP activity. The RGS domain of LARG did not have GAP activity for Gαs, Gαi, Gαo, and Gαq (data not shown). Thus, the RGS domain of LARG serves as a specific GAP for Gα12 or Gα13 similar to that of p115RhoGEF.

We also examined the regulation of RhoGEF activity of LARG by Gα12/13. In the case of p115RhoGEF, Gα13, but not Gα12, stimulated its RhoGEF activity (7). As shown in Fig. 2*C*, AlF_4^--activated Gα13 stimulated the RhoGEF activity of LARG. However, AlF_4^--activated Gα12 did not demonstrate RhoGEF activation. Thus, although SRF assays suggested that Gα12-LARG mediated Rho activation in HeLa cells, we could not reconstitute that pathway *in vitro*. The results suggest that additional factors or some modification on Gα12 or LARG will be necessary for activation of Rho through the Gα12-LARG pathway.

Because the involvement of Tec kinase has been reported in the Gα12-mediated pathway, we tested the possibility that Tec tyrosine kinase might be involved in Rho activation through Gα12/13-LARG. As shown in Fig. 3*A*, coexpression of Tec kinase in HeLa cells potently stimulated both Gα12- and Gα13-LARG-mediated SRF activation. However, we did not observe a similar effect of Tec when Gα12/13 or LARG was expressed alone. Coexpression of Tec did not stimulate SRF activation mediated by Gα12/13-p115RhoGEF (data not shown). In addition, a kinase-deficient mutant of Tec (Tec-KD) failed to stimulate the Gα12/13-LARG-mediated SRF activation. GTP-bound Rho pull-down assay also demonstrated that Rho activation by Gα12-LARG in HeLa cells was further stimulated by Tec (Fig. 3*B*). These results suggest that Tec tyrosine kinase regulates Gα12/13-LARG-mediated Rho activation by phosphorylating some component of the pathway.

The interaction of Gα12 with Btk, another member of the Tec family, through its pleckstrin homology–TH domain was recently demonstrated (15). As shown in Fig. 3*C*, we could also observe the interaction between constitutively active Gα12QL and Tec by coimmunoprecipitation. Furthermore, a Tec construct lacking TH domain did not show the stimulatory effect on Gα12-LARG-mediated SRF activation, indicating that the TH domain of Tec is required for its effect on the Gα12/13-LARG pathway (Fig. 3*D*).

We next examined whether Tec can directly phosphorylate Gα12/13 or LARG. Myc-tagged Tec was overexpressed in COS1 cells, immunoprecipitated by anti-myc antibody, and used for *in vitro* phosphorylation assays. As shown in Fig. 4*A*, ΔPDZ-LARG was phosphorylated on tyrosine by Tec. However, p115RhoGEF, Gα12, or Gα13 did not serve as a substrate for Tec. Moreover, the activated Gα12 or Gα13 did not affect the phosphorylation of LARG by Tec. We also examined tyrosine phosphorylation of LARG in cells. A Tec construct with an N-terminal myristoylation signal (mHTec) was targeted to the plasma membrane and exhibited constitutive activity (19). As shown in Fig. 4*B*, ΔPDZ-LARG, but not p115, was tyrosine phosphorylated in HEK293 cells when coexpressed with mHTec. However, we could not detect tyrosine phosphorylation of ΔN-LARG under the same condition. These results suggest that Tec phosphorylates LARG *in vivo* as well as *in vitro*. Furthermore, the phosphorylation site on LARG is likely in

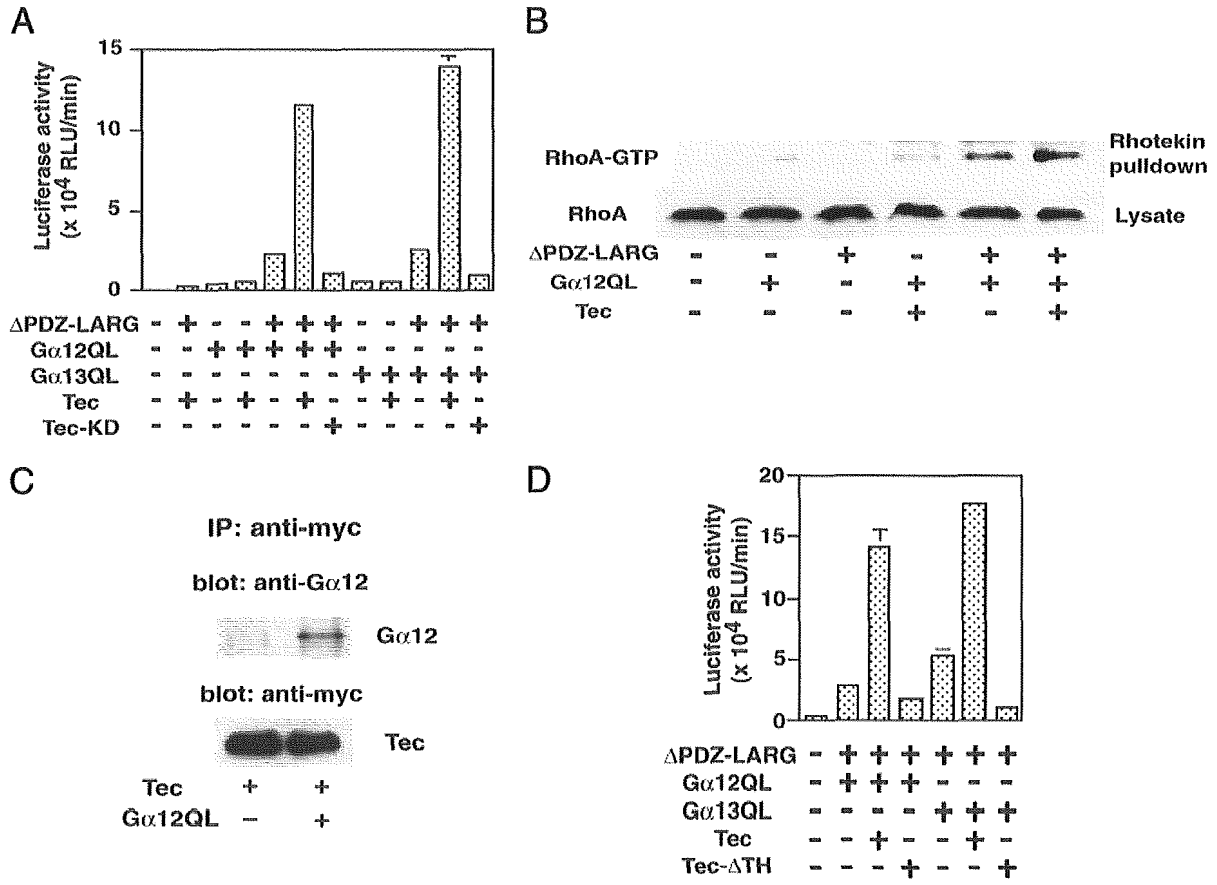


Fig. 3. Effect of Tec for Rho activation by Gα12/13-LARG. (*A*) Activation of SRF activity by Tec. HeLa cells were cotransfected with 0.1 μg of SRE.L-luciferase reporter plasmid with the indicated constructs: 0.01 μg of Gα12QL, 0.01 μg of Gα13QL, 0.1 μg of ΔPDZ-LARG, 0.1 μg of Tec, and 0.1 μg of Tec-KD. SRF activities of cell lysates were measured after 24 h. (*B*) RhoA activation by Gα12-LARG and Tec in HeLa cells. HeLa cells were transiently transfected with the plasmids encoding ΔPDZ-LARG, Gα12QL, or Tec. GTP-bound RhoA in cell lysates was detected using GST-RBD pull-down assay. The result shown is a representative of three separate experiments with similar results. (*C*) Direct interaction of Tec with Gα12. COS1 cells were transfected with myc-tagged Tec with or without Gα12QL. Tec was immunoprecipitated from the lysate, and the immunoprecipitates were analyzed by immunoblotting using Gα12 antibody. (*D*) Requirement of the TH domain of Tec in Gα12/13-LARG signaling pathway. HeLa cells were cotransfected with 0.1 μg of SRE.L-luciferase reporter plasmid with indicated plasmids as described for *A* or with 0.1 μg of Tec-ΔTH. SRF activities of cell lysates were measured after 24 h.

the region including its RGS domain (amino acid residues 307–617 of LARG).

Finally, we examined the effect of phosphorylation of LARG on its RhoGEF activity (Fig. 4 *C* and *D*). LARG was first phosphorylated by preincubation with Tec, and then RhoGEF assays were started by adding RhoA and Gα subunits. The preincubation with Tec did not affect the basal RhoGEF activity of LARG. However, Gα12 stimulated the RhoGEF activity of LARG in a phosphorylation-dependent manner. This stimulation was almost similar to the level by Gα13 without Tec. The stimulatory effect by Gα12 was also dependent on the presence of ATP during preincubation (data not shown), confirming that the phosphorylation of LARG is required for Gα12 to activate Rho. These results indicate that Gα12 can activate Rho through phosphorylated LARG. Tyrosine phosphorylation of LARG also modestly stimulated RhoGEF activity mediated by Gα13 (Fig. 4*D*).

Discussion

The results of this study demonstrated the biochemical mechanism of Rho activation through Gα12. Although Gα12 and Gα13 are similar in amino acid sequences and biochemical properties and both are involved in Rho activation, several studies indicated the functional differences between these two Gα subunits. The most striking difference was demonstrated in Gα13 gene knockout mice, which showed embryonic lethality due to the defect of vascular system formation (26). Gα12 could not rescue the function of Gα13 in these mice. In reconstitution experiments, Gα13 stimulated the GEF activity of p115RhoGEF. However, Gα12 could not stimulate the GEF activity of p115 and competitively inhibited the stimulatory effect of Gα13 (7). In this report, we demonstrated that the tyrosine phosphorylation of LARG is required for Gα12 to activate Rho. Although tyrosine phosphorylation of LARG could further stimulate the effect of Gα13, it was not required for Rho activation by Gα13. These differences in the regulatory mechanisms of Gα12- and Gα13-mediated pathways may be responsible for the different cellular effects induced by Gα12 or Gα13.

Because the TH domain was required for Tec to stimulate Gα12-LARG-mediated SRF activation, this domain may be involved in the interaction with Gα12 similar to the case of Btk. However, we could not detect the activation of Tec kinase by Gα12 *in vitro* (data not shown). The exact mechanism of Gα12 to regulate Tec kinase is currently unclear. However, it is interesting to note that thrombin, which can activate the Gα12/13 pathway, has also been reported to activate Tec in platelets (17). It is possible that activated Gα12 may recruit Tec

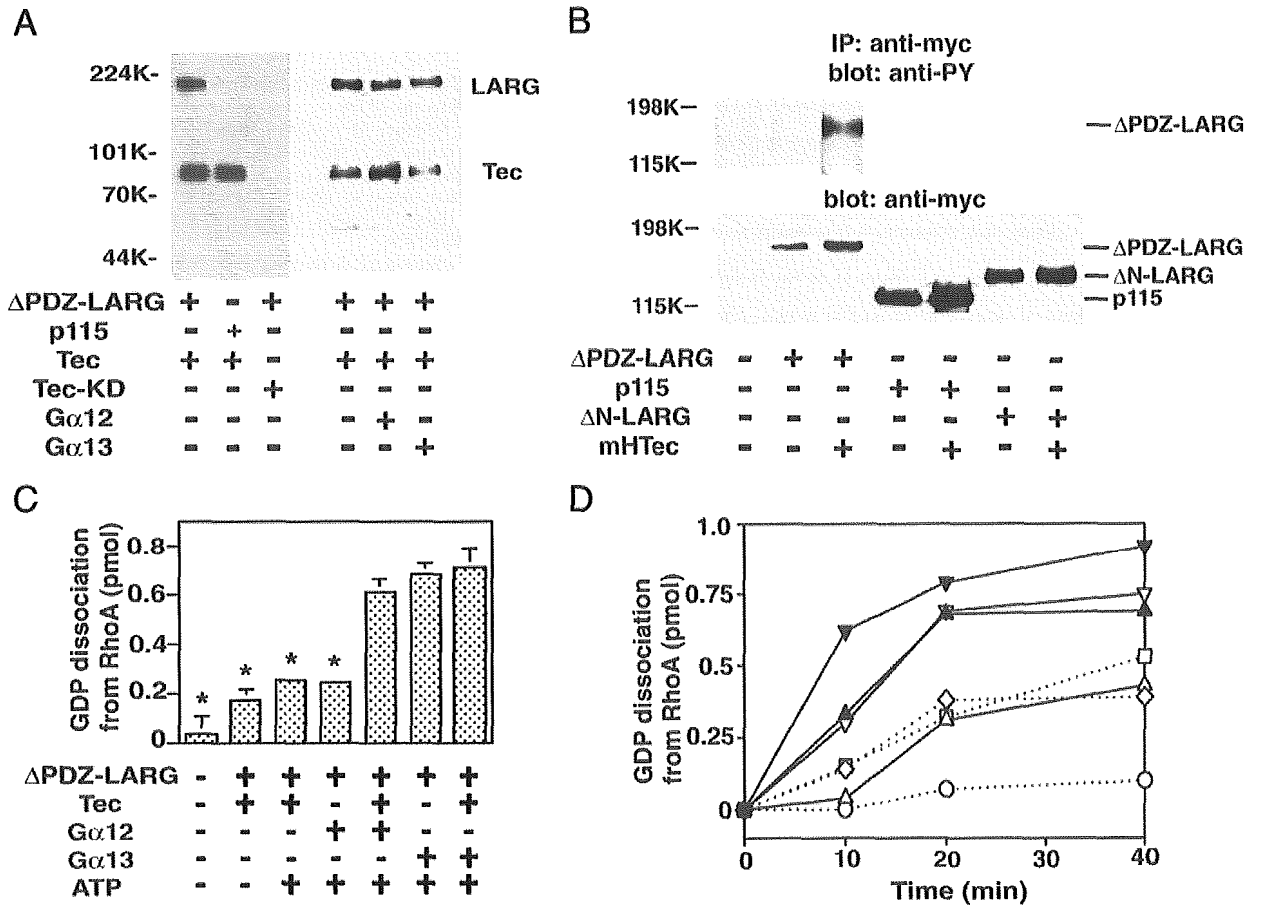


Fig. 4. Tyrosine phosphorylation of LARG by Tec. (*A*) Tyrosine phosphorylation of LARG by Tec *in vitro*. Myc-tagged Tec or Tec-KD was overexpressed in COS1 cells, immunoprecipitated by anti-myc antibody, and used for kinase assays. Recombinant ΔPDZ-LARG or p115RhoGEF (600 nM each) with or without 1 μM AlF_4^--activated Gα12 or Gα13 were incubated with immunoprecipitated Tec in the presence of ATP for 20 min at 30°C. Proteins were separated by SDS/PAGE, followed by immunoblotting with antiphosphotyrosine antibody. (*B*) Tyrosine phosphorylation of LARG by Tec *in vivo*. HEK293 cells were cotransfected with myc-tagged ΔPDZ-LARG, ΔN-LARG, or p115 with or without mHTec. RhoGEFs were immunoprecipitated by anti-myc antibody from the cell lysates, and the immunoprecipitates were separated by SDS/PAGE, followed by immunoblotting using antiphosphotyrosine antibody or anti-myc antibody. (*C*) Stimulation of GDP dissociation from RhoA by Gα12, LARG, and Tec. GDP dissociation from RhoA by ΔPDZ-LARG was assayed for 20 min at 20°C in the presence of the indicated proteins (as described in *Methods*): 10 nM ΔPDZ-LARG, 200 nM AlF_4^--activated Gα12, and 200 nM AlF_4^--activated Gα13. *, $P < 0.01$, significant difference from the data with Gα12 + LARG + Tec. The results shown are from a representative experiment of three such experiments with similar results. (*D*) Time course of GDP dissociation from RhoA. GDP dissociation from RhoA was measured with the indicated proteins: ○, control; □, 10 nM ΔPDZ-LARG; ◇, 10 nM ΔPDZ-LARG + Tec; △, 10 nM ΔPDZ-LARG + 200 nM Gα12; ▽, 10 nM ΔPDZ-LARG + 200 nM Gα13; ▲, 10 nM ΔPDZ-LARG + 200 nM Gα12 + Tec; or ▼, 10 nM ΔPDZ-LARG + 200 nM Gα13 + Tec.

in close proximity of LARG and facilitate the phosphorylation of LARG. Tec is also activated by other stimuli, such as cytokines and growth factors. These signaling pathways will also be able to regulate the Gα12-LARG pathway.

In addition to Tec, the involvement of Pyk2 in the G12/13-RhoGEF pathway has been reported (27). Furthermore, Chikumi *et al.* (28) recently demonstrated that thrombin stimulation activated nonreceptor tyrosine kinase FAK in HEK293 cells and that activated FAK could phosphorylate PDZ-RhoGEF or LARG but not p115RhoGEF. They also demonstrated the enhancement of Rho activation by coexpression of activated FAK and PDZ-RhoGEF *in cells*. They proposed that tyrosine phosphorylation of PDZ-RhoGEF or LARG by FAK might be involved in the activation of Rho. However, its biochemical mechanism remained unclear. We demonstrated here that tyrosine phosphorylation of LARG by Tec does not affect its basal RhoGEF activity, but rather changes its regulation by Gα subunits. It is possible that the activity of PDZ-RhoGEF is also regulated by tyrosine phosphorylation. The modulation of the Gα12/13-RhoGEF pathway by tyrosine kinases may be a widely used mechanism for G protein-coupled receptor-mediated Rho activation.

We thank Dr. T. Nagase for providing cDNAs for *KIAA0380* and KIAA0382 and P. M. Sternweis for the assistance in cloning full-length LARG cDNA. This work was supported in part by the National Institutes of Health and the American Heart Association (T.K.). T.K. is an Established Investigator of the American Heart Association.

1. Hall, A. (1998) *Science* **279**, 509–514.
2. Whitehead, I. P., Campbell, S., Rossman, K. L. & Der, C. J. (1997) *Biochim. Biophys. Acta* **1332**, F1–F23.
3. Gohla, A., Harhammer, R. & Schultz, G. (1998) *J. Biol. Chem.* **273**, 4653–4659.
4. Aragay, A. M., Collins, L. R., Post, G. R., Watson, A. J., Feramisco, J. R., Brown, J. H. & Simon, M. I. (1995) *J. Biol. Chem.* **270**, 20073–20077.
5. Kranenburg, O., Poland, M., van Horck, F. P. G., Drechsel, D., Hall, A. & Moolenaar, W. H. (1999) *Mol. Biol. Cell* **10**, 1851–1857.
6. Kozasa, T., Jiang, X., Hart, M. J., Sternweis, P. M., Singer, W. D., Gilman, A. G., Bollag, G. & Sternweis, P. C. (1998) *Science* **280**, 2109–2111.
7. Hart, M. J., Jiang, X., Kozasa, T., Roscoe, W., Singer, W. D., Gilman, A. G., Sternweis, P. C. & Bollag, G. (1998) *Science* **280**, 2112–2114.

8. Buhl, A. M., Johnson, N. L., Dhanasekaran, N. & Johnson, G. L. (1996) *J. Biol. Chem.* **270,** 24631–24634.
9. Crespo, P., Schuebel, K. E., Ostrom, A. A., Gutkind, J. S. & Bustelo, X. R. (1997) *Nature* **385,** 169–172.
10. Schuebel, K. E., Movilla, N., Rosa, J. L. & Bustelo, X. R. (1998) *EMBO J.* **17,** 6608–6621.
11. Kato, J., Kaziro, Y. & Satoh, T. (2000) *Biochem. Biophys. Res. Commun.* **268,** 141–147.
12. Katoh, H., Aoki, J., Yamaguchi, Y., Kitano, Y., Ichikawa, A. & Negishi, M. (1998) *J. Biol. Chem.* **273,** 28700–28707.
13. Klages, B., Brandt, U., Simon, M. I., Schultz, G. & Offermanns, S. (1999) *J. Cell Biol.* **144,** 745–754.
14. Mao, J., Xie, W., Yuan, H., Simon, M. I., Mano, H. & Wu, D. (1998) *EMBO J.* **17,** 5638–5646.
15. Jiang, Y., Ma, W., Wan, Y., Kozasa, T., Hattori, S. & Huang, X.-Y. (1998) *Nature* **395,** 808–813.
16. Mano, H. (1999) *Cytokine Growth Factor Rev.* **10,** 267–280.
17. Hamazaki, Y., Kojima, H., Mano, H., Nagata, Y., Todokoro, K., Abe, T. & Nagasawa, T. (1998) *Oncogene* **16,** 2773–2780.
18. Nagase, T., Ishikawa, K., Nakajima, D., Ohira, M., Seki, N., Miyajima, N., Kotani, H., Nomura, N. & Ohara, O. (1997) *DNA Res.* **4,** 141–150.
19. Yoshida, K., Yamashita, Y., Miyazato, A., Ohya, K., Kitanaka, A., Ikeda, U., Shimada, K., Yamanaka, T., Ozawa, K. & Mano, H. (2000) *J. Biol. Chem.* **275,** 24945–24952.
20. Kozasa, T. (1999) *G Proteins: Analysis of Technique*, ed. Manning, D. R. (CRC, Boca Raton, FL), pp. 23–38.
21. Ren, X. D. & Schwartz, M. A. (2000) *Methods Enzymol.* **325,** 264–272.
22. Kourlas, P. J., Strout, M. P., Becknell, B., Veronese, M. L., Croce, C. M., Theil, K. S., Krahe, R., Ruutu, T., Knuutila, S., Bloomfield, C. D. & Caligiuri, M. A. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 2145–2150.
23. Fukuhara, S., Murga, C., Zohar, M., Igishi, T. & Gutkind, J. S. (1999) *J. Biol. Chem.* **274,** 5868–5879.
24. Fukuhara, S., Chikumi, H. & Gutkind, J. S. (2000) *FEBS Lett.* **485,** 183–188.
25. Fromm, C., Coso, O. A., Montaner, S., Xu, N. & Gutkind, J. S. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 10098–10103.
26. Offermanns, S., Mancino, V., Revel, J.-P. & Simon, M. I. (1997) *Science* **275,** 533–536.
27. Shi, C. S., Simarajah, S., Cho, H., Kozasa, T. & Kehrl, J. H. (2000) *J. Biol. Chem.* **275,** 24470–24476.
28. Chikumi, H., Fukuhara, S. & Gutkind, J. S. (2002) *J. Biol. Chem.* **277,** 12463–12473.

Oncogene (2003) 22, 5720–5728

www.nature.com/onc

Proteomic analysis of hematopoietic stem cell-like fractions in leukemic disorders

Jun Ota[1,9], Yoshihiro Yamashita[1,9], Katsuya Okawa[1], Hiroyuki Kisanuki[1], Shin-ichiro Fujiwara[1,3], Madoka Ishikawa[1], Young Lim Choi[1], Shuichi Ueno[1,4], Ruri Ohki[1,4], Koji Koinuma[1,5], Tomoaki Wada[1,6], Duane Compton[7], Toshihiko Kadoya[8] and Hiroyuki Mano*[,1]

[1]*Division of Functional Genomics, Jichi Medical School, 3311-1 Yakushiji, Kawachi-gun, Tochigi 329-0498, Japan;* [2]*Pharmaceutical Research Laboratories, Pharmaceutical Division, Kirin Brewery Co. Ltd, Takasaki, Gunma 370-1925, Japan;* [3]*Division of Hematology, Jichi Medical School, Kawachi-gun, Tochigi 329-0498, Japan;* [4]*Division of Cardiology, Jichi Medical School, Kawachi-gun, Tochigi 329-0498, Japan;* [5]*Department of Surgery, Jichi Medical School, Kawachi-gun, Tochigi 329-0498, Japan;* [6]*Department of Gynecology, Jichi Medical School, Kawachi-gun, Tochigi 329-0498, Japan;* [7]*Department of Biochemistry, Dartmouth Medical School, Hanover, NH 03755-3844, USA;* [8]*R&D Center, Product Department, Pharmaceutical Division, Kirin Brewery Co. Ltd, Maebashi, Gunma 371-0853, Japan*

DNA microarray analysis has been applied to identify molecular markers of human hematological malignancies. However, the relatively low correlation between the abundance of a given mRNA and that of the encoded protein makes it important to characterize the protein profile directly, or 'proteome,' of malignant cells in addition to the 'transcriptome.' To identify proteins specifically expressed in leukemias, here we isolated AC133$^+$ hematopoietic stem cell-like fractions from the bone marrow of 13 individuals with various leukemic disorders, and compared their protein profiles by two-dimensional electrophoresis. A total of 11 differentially expressed protein spots corresponding to 10 independent proteins were detected, and peptide fingerprinting combined with mass spectrometry of these proteins revealed them to include NuMA (nuclear protein that associates with the mitotic apparatus), heat shock proteins, and redox regulators. The abundance of NuMA in the leukemic blasts was significantly related to the presence of complex karyotype anomalies. Conditional expression of NuMA in a mouse myeloid cell line resulted in the induction of aneuploidy, cell cycle arrest in G_2–M phases, and apoptosis. These results demonstrate the potential of proteome analysis with background-matched cell fractions obtained from fresh clinical specimens to provide insight into the mechanism of human leukemogenesis.
Oncogene (2003) **22,** 5720–5728. doi:10.1038/sj.onc.1206855

Keywords: acute myeloid leukemia; proteome; AC133; NuMA

*Correspondence: H Mano; E-mail: hmano@jichi.ac.jp
[9]These two authors contributed equally to this work
Received 7 February 2003; revised 23 May 2003; accepted 9 June 2003

Introduction

Annotation of the draft sequence of the human genome has opened up the possibility of applying novel genomic approaches to the characterization of molecular pathogenesis of human disorders (The genome international sequencing consortium, 2001; Venter *et al.*, 2001). Among genomic screening methods, DNA microarray analysis has to date provided the greatest insight into leukemogenesis. This technology readily allows measurement of the expression levels of thousands of genes simultaneously (Duggan *et al.*, 1999). Expression profiling with microarrays has thus made it possible, for example, to distinguish acute myeloid leukemia (AML) from acute lymphoid leukemia (ALL) (Golub *et al.*, 1999), to define novel subgroups of leukemias and lymphomas (Alizadeh *et al.*, 2000; Armstrong *et al.*, 2002), and to identify candidate genes for leukemogenesis (Ohmine *et al.*, 2001; Makishima *et al.*, 2002).

An important concern in the assay of fresh specimens by microarray analysis, however, is that apparent changes in gene expression detected at different stages of carcinogenesis may actually reflect changes in cell composition rather than changes in gene expression *per se*. For example, whereas immature (leukemic) blasts constitute $\geqslant$20% of bone marrow (BM) mononuclear cells (MNCs) in individuals with leukemia, BM MNCs of normal individuals contain only a few percent immature blasts. A simple comparison by microarray analysis between normal and leukemic BM cells would, therefore, likely reveal changes in gene expression only attributable to the expansion of immature blasts in the latter. Indeed, we observed that one of the genes whose expression appeared highly specific for leukemic BM cells, compared with normal BM cells, was that for CD34, simply reflecting the expansion of CD34$^+$ leukemic blasts in the leukemic BM specimen (Miyazato *et al.*, 2001).

To eliminate such population-shift effects, we have purified and stored AC133$^+$ hematopoietic stem cell (HSC)-like fractions from the BM of patients with a wide range of leukemic disorders and deposited them in our 'Blast Bank.' The suitability of such purified fractions for microarray analysis was confirmed by the observation that the CD34 gene was expressed at similar levels in AC133$^+$ cells obtained from normal individuals and in those isolated from leukemic patients (Miyazato *et al.*, 2001). Microarray analysis of Blast Bank samples has also resulted in the identification of new molecular markers for myelodysplastic syndrome (MDS) (Miyazato *et al.*, 2001) and for chronic myeloid leukemia (CML) (Ohmine *et al.*, 2001).

Despite its potential for identifying genes important in leukemogenesis, microarray analysis is not able to provide direct information on the abundance or post-translational modification of proteins. The transcriptional activity on a given gene is thus not always a major determinant of the expression level of the encoded protein. After exclusion of several of the most abundant proteins, a large-scale study (Gygi *et al.*, 1999) of yeast cells determined the correlation coefficient between the amount of an mRNA and the abundance of the corresponding protein to be only $\sim$0.4. Furthermore, only four out of 28 proteins identified in a mouse cell line showed relative levels similar to those of the corresponding mRNAs (Lian *et al.*, 2001). A thorough characterization of leukemogenesis thus requires direct determination of the accompanying changes not only in the amounts of cellular mRNAs but also in protein abundance.

In addition to discrepancies between the amounts of mRNA and protein derived from a given gene, the activities of many proteins are influenced by post-translational modifications such as phosphorylation, cleavage, glycosylation, and redox regulation. Such concerns have highlighted the importance of proteomic approaches that are able to assess changes in the abundance and post-translational modification of proteins on a large scale. One of the main current strategies in proteomics is the combination of two-dimensional gel electrophoresis (2DE) and mass spectrometry (MS). Whereas 2DE allows the screening of hundreds to thousands of proteins for changes in molecular size, isoelectric point (pI), or phosphorylation, MS then allows the identification of protein spots of interest.

However, as in the case of DNA microarray analysis, the population-shift effect is also an important consideration in proteomics. Differentiation of cells is thus accompanied by changes in the expression of a substantial number of proteins (Lian *et al.*, 2001). A proteome comparison between two specimens with different cell compositions would therefore yield pseudopositive data that reflect only the population-shift effect. A proteomic approach to the characterization of leukemogenesis would thus ideally require the purification of background-matched cell populations from fresh specimens of leukemic patients.

We have now prepared protein extracts from the purified AC133$^+$ leukemic blasts of 13 individuals with acute leukemia or related disorders. Screening of these protein samples with 2DE and MS resulted in the identification of 10 proteins that were expressed differentially among the patients.

Results

2DE of Blast Bank samples

Our goal was to identify proteins whose abundance, relative molecular mass, or pI differs markedly among HSC-like fractions isolated from individuals with leukemia. Preliminary studies with cell lines revealed that at least 1×10^6 MNCs were required to yield $\geqslant$100 spots reproducibly on 2DE gels. Given that the AC133$^+$ HSC-like fraction usually constitutes a small proportion of leukemic blasts, it was not always possible to obtain such a large number of AC133$^+$ cells from fresh patient specimens. Indeed, our Blast Bank contained only 13 AC133$^+$ fractions that comprised $\geqslant 2 \times 10^6$ cells. The clinical characteristics of the corresponding patients (five with *de novo* AML, four with MDS-associated AML, three with myeloproliferative disorders (MPDs) in blast crisis (BC), and one with ALL) are summarized in Table 1. All these patients died within 12 months of diagnosis.

Our preliminary studies also revealed that abundant proteins sometimes masked neighboring minor protein spots on the 2DE gels. To minimize such effects, we therefore first fractionated the cell samples into different subcellular components, including nuclear, mitochondrial, microsomal, and cytosolic fractions (Watarai *et al.*, 2000). Each subcellular fraction was then independently compared by 2DE among the 13 patients. The cytosolic fractions consistently yielded $\geqslant$100 spots per gel and were analysed further in the present study. The comparisons of the other subcellular fractions will be described separately.

Identification of differentially expressed proteins

A representative image of a silver-stained gel, for which Melanie III software detected >200 independent spots, is shown in Figure 1a. Comparison of these spots among the gel images for the 13 patients revealed a total of 11 spots that showed a significant difference in abundance (as judged by the criteria described in Materials and methods). Peptide fingerprinting by matrix-assisted laser desorption-ionization time-of-flight (MALDI-TOF) MS of these 11 spots resulted in the identification of the corresponding 10 proteins (Table 2). They include nuclear protein that associates with the mitotic apparatus (NuMA), heat shock 70-kDa protein 5 (HSPA5), heat shock 70-kDa protein 8 (HSPA8), adenosine deaminase (ADA), aldolase A (ALDOA), triose phosphate isomerase 1 (TPI1), glutathione S-transferase pi (GST-pi), superoxide dismutase 2 (SOD2), peptidyl-prolyl isomerase A (PPIA), and heat shock 70-kDa protein 9B (HSPA9B). Two independent spots with different molecular mass and pI values were revealed to

Table 1 Clinical characteristics of the patients subjected to proteome analysis

Patient ID	*Age (years)*	*Sex*	*Disease and subtype*	*Treatment outcome*	*Abnormal chromosomes (⩾3)*
AML #1	65	M	AML, M0	Failure	–
AML #2	71	M	AML, M0	Failure	+
AML #3	65	F	AML, M1	CR	–
AML #4	31	F	AML, M4	Failure	–
AML #5	83	F	AML, M5a	Failure	–
MDS #1	69	M	MDS-associated AML, M2	Failure	–
MDS #2	67	F	MDS-associated AML, M2	Failure	+
MDS #3	68	F	MDS-associated AML, M4	Failure	–
MDS #4	53	M	MDS-associated AML, M2	Failure	+
MPD #1	60	M	MPD, BC	Failure	+
MPD #2	68	F	PV, BC	Failure	+
MPD #3	33	M	CML, BC	Failure	–
ALL #1	23	F	ALL, L1	Failure	–

FAB subtypes for the patients are indicated. Abbreviations: M, male; F, female; CR, complete remission; PV, polycythemia vera

be derived from the same gene product, HSPA8 (Figure 1a).

We then examined whether clinical parameters of the patients were related to the intensity of any of the 11 protein spots. The expression level of NuMA was significantly related to whether the number of abnormal chromosomes was $\geqslant 3$ or <3 ($P=0.017$; Student's *t*-test). The images of the NuMA spots for all 13 patients are shown in Figure 1b. NuMA is a nuclear protein that accumulates in the pericentrosomal region of the mitotic spindle and plays an important role in the assembly of mitotic asters (Compton and Cleveland, 1993; Gaglio *et al.*, 1997; Du *et al.*, 2001).

Since we could obtain only several thousands of $AC133^+$ cells from the BM aspirates of healthy volunteers, it was impossible to assess the protein level of $p240^{NuMA}$ directly in the $AC133^+$ fractions of normal individuals. Instead, here we have quantified the abundance of *NuMA* mRNA by the 'real-time' reverse transcription–polymerase chain reaction (RT–PCR) method. The abundance of *NuMA* mRNA relative to that of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mRNA was examined among the $AC133^+$ HSC-like fractions obtained from a healthy volunteer and those with CML in chronic phase (CP), in addition to the specimens subjected to 2DE (Figure 1c). Although the mRNA level of *NuMA* did not precisely correlate with the corresponding protein level, there was a tendency that its mRNA was abundant in the samples with a high expression of NuMA protein (shown in gray columns).

The proportion of malignant blasts is below 5% in BM MNCs of patients with CML in CP, and, therefore, such BM aspirates could yield only $1\times10^3\sim1\times10^4$ of $AC133^+$ cells, which were not enough for the protein analysis. However, since these patients had only one chromosomal anomaly, t(9;22), they were chosen to be included in this RT–PCR analysis. The expression level of the *NuMA* gene was negligibly low in the HSC-like fractions of a normal individual and most patients with CML in CP, which indirectly supports the low expression of NuMA protein in these specimens.

Aneuploidy associated with NuMA expression

The role of NuMA in mitosis suggested that aberrant induction of NuMA expression in leukemic blasts might contribute to the chromosomal instability apparent in such cells. To examine directly the effects of NuMA overexpression, we transfected mouse myeloid 32D cells (Greenberger *et al.*, 1983) with a vector that confers Zn^{2+}-dependent expression of human NuMA. Cell clones transfected with this vector (32D-NuMA#1 to 32D-NuMA#7) or the corresponding empty vector (32D-MT#1 to 32D-MT#4) were selected by culture in the presence of G418, after which 0.1 mM Zn^{2+} was added to the culture medium. Immunoblot analysis with antibodies to NuMA revealed marked expression of $p240^{NuMA}$ in the 32D-NuMA clones, but not in the 32D-MT clones (Figure 2a). The electrophoretic mobility of NuMA expressed in the 32D-NuMA clones was identical to that of $p240^{NuMA}$ detected in human kidney 293 cells transiently transfected with the pMT-NuMA vector.

The Zn^{2+} dependence of NuMA expression in each 32D-NuMA clone was verified by incubating cells in the absence or presence of $ZnSO_4$. Immunoblot analysis revealed a marked dependence of NuMA expression on the presence of Zn^{2+} for all 32D-NuMA clones (Figure 2b; data not shown). The amount of NuMA in 32D-NuMA#1 cells, for example, incubated overnight in the absence or presence of Zn^{2+} was 12.8 and 59.8 times, respectively, that of endogenous $p240^{NuMA}$ in 32D-MT cells.

Examination of the morphology of 32D transfectants revealed that the induction of NuMA expression in some cells resulted in a marked increase in cell size and in the formation of multiple nuclei (Figure 3a); effects suggestive of the development of aneuploidy. Flow cytometric analysis of DNA content revealed that incubation of 32D-NuMA transfectants with Zn^{2+} for 2 days led to mitotic block (a decrease in the proportion of cells in S phase of the cell cycle, and an increase in the proportion of those in G_2–M); such an effect was not observed in 32D-MT cells (Table 3). Furthermore, induction of NuMA expression in transfected cells was

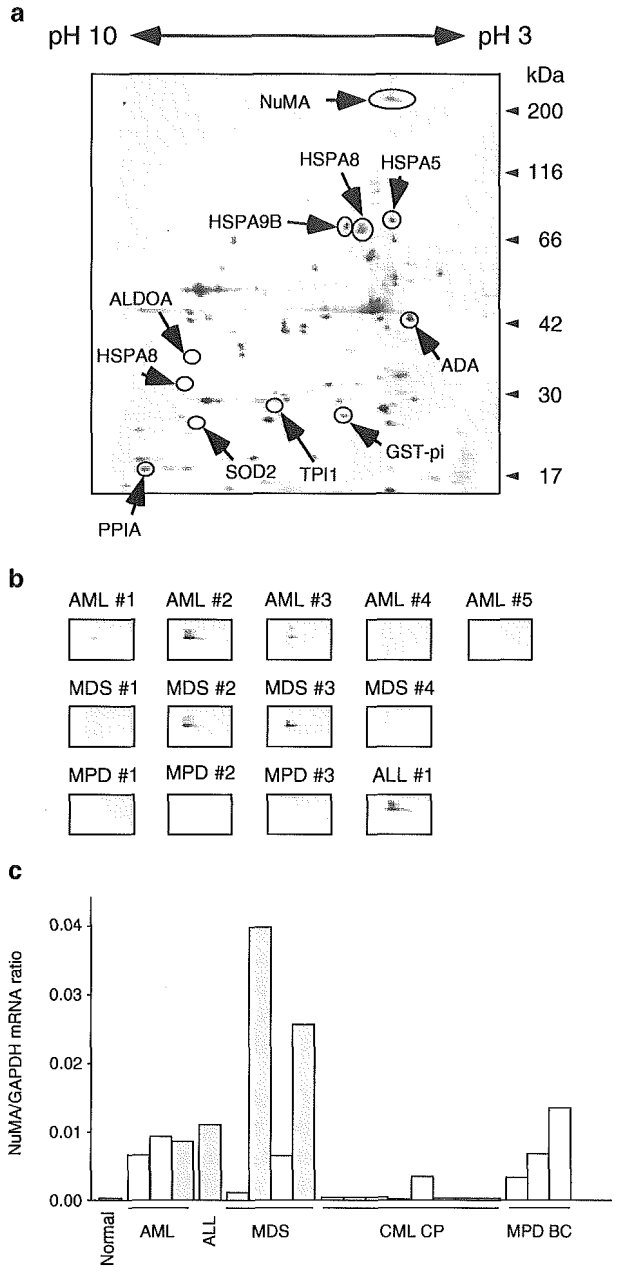


Figure 1 2DE analysis of Blast Bank samples. (**a**) The cytosolic fraction (~10 μg of protein) isolated from AC133$^+$ leukemic blasts was subjected to 2DE on a 7.5–15% gradient gel. The scanned image of the silver-stained gel was then used to detect and compare protein spots. The positions and identities of proteins expressed differentially among leukemic patients are indicated by arrows, and the positions of molecular mass standards (in kilodaltons) are shown on the right. (**b**) Images of the NuMA spots for all 13 patients analysed. (**c**) Complementary DNAs were prepared from the AC133$^+$ blasts of a healthy volunteer (normal) and those with AML, ALL, MDS-associated leukemia (MDS), CML in CP, and MPD in BC, and were subjected to real-time RT–PCR analysis with primers specific for the *NuMA* or *GAPDH* gene. The ratio of the abundance of *NuMA* mRNA to that of *GAPDH* mRNA was calculated as 2^n, where n is the C_T value for *GAPDH* cDNA minus the C_T value for *NuMA* cDNA. The expression levels of *NuMA* mRNA for the patients with abundant expression of NuMA proteins (more than 3.0 U in Table 2) are shown as gray columns

associated with an increase in the proportion of cells with an abnormal (4(n)) DNA content (Table 4). These data thus suggested that overexpression of NuMA perturbs cell cycle progression by inhibiting mitosis and thereby results in the development of aneuploidy.

We also noticed that induction of p240^{NuMA} expression was accompanied by the appearance of cells with apoptotic features (Figure 3a) as well as by marked decreases in both the rate of cell growth (Figure 3b) and cell viability (data not shown). The increase in the prevalence of apoptosis in NuMA-overexpressing cells was confirmed by the detection of an increased extent of internucleosomal DNA fragmentation by agarose gel electrophoresis of genomic DNA (Figure 3c).

In our 2DE analysis, NuMA was identified within the cytoplasmic fractions, while the protein was originally reported to be accumulated in the nucleus. To determine its subcellular localization, p240^{NuMA} fused with enhanced green fluorescent protein (EGFP) (NuMA-EGFP) was expressed in human kidney 293 cells. As shown in Figure 3d, EGFP was expressed diffusely within cells. With regard to NuMA-EGFP, it was localized in the nucleus (but not in the nucleolus) in the majority of transfected cells (NuMA-EGFP#1). There was, however, a fraction of cells that had NuMA exclusively within their cytoplasm (NuMA-EGFP#2–3). Therefore, NuMA can be localized within the cytoplasm in living cells. Since such cells with cytoplasmic NuMA had a round shape, they may be at specific stages of the cell cycle. It would be an intriguing issue to address whether subcellular localization of p240^{NuMA} is regulated in a cell cycle-dependent manner.

Discussion

We have compared the protein profiles of HSC-like fractions derived from individuals with various leukemia-related disorders. Given that the proteome is strongly influenced by cell differentiation (Lian *et al.*, 2002), a simple comparison of BM MNCs from different patients is likely to result in the identification of proteins whose apparent change in expression actually reflects a difference in the cellular composition of the specimens. To prevent such a complication, we isolated highly immature BM cells on the basis of their surface expression of AC133. Comparison of such background-matched fractions should eliminate pseudopositive data that might result from different proportions of leukemic blasts in BM or from differences in cell lineage to which the leukemic blasts are committed (Miyazato *et al.*, 2001).

We identified 10 proteins that were expressed differentially among the leukemic blasts from the 13 patients examined. Three of these proteins (HSPA5, HSPA8 and HSPA9B) belong to the heat shock protein family of 70 kDa (HSP70), and two of them (GST-pi and SOD2) function in redox regulation. HSPA5 (also known as glucose-regulated protein 78 (GRP78) or immunoglobulin heavy chain-binding protein (BiP)), HSPA8 (heat

Table 2 Quantitation of the abundance of proteins expressed differentially among the patients

Protein	*Accession number*	*Patient ID*												
		AML #1	*AML #2*	*AML #3*	*AML #4*	*AML #5*	*MDS #1*	*MDS #2*	*MDS #3*	*MDS #4*	*MPD #1*	*MPD #2*	*MPD #3*	*ALL #1*
NuMA	XP_006005	1.339	8.135	1.350	0.324	0.700	0.492	10.771	4.772	0.460	0.780	0.139	0.177	7.100
HSPA5	P11021	0.378	0.347	0.801	0.391	0.579	0.374	0.424	0.063	0.100	0.396	0.827	0.582	0.478
HSPA8 (p71)	P11142	1.440	0.569	2.554	1.442	0.328	1.107	0.677	0.236	0.180	0.673	0.324	0.793	1.150
ADA	P00813	0.113	0.900	1.859	0.530	0.500	1.007	1.104	1.618	1.178	2.750	1.688	0.236	0.766
ALDOA	P04075	0.264	1.336	0.119	0.164	0.843	0.324	0.970	1.723	0.436	0.144	0.280	0.218	0.630
HSPA8 (p34)	P11142	0.150	0.786	0.081	0.158	0.435	0.490	0.108	0.582	0.000	0.306	0.553	0.071	0.191
TPI1	P00938	0.743	2.061	0.727	0.262	1.254	2.162	0.464	2.990	0.157	0.602	0.439	0.605	0.903
GST-pi	P09211	0.785	0.737	1.312	0.841	2.107	0.685	0.302	1.241	0.290	0.660	1.276	1.077	0.868
SOD2	P04179	0.083	0.283	0.065	0.562	0.290	0.325	0.096	0.341	0.000	0.336	0.143	0.594	0.255
PPIA	P05092	1.473	1.630	0.924	0.147	1.534	2.008	0.905	0.733	1.722	2.006	0.175	0.884	0.567
HSPA9B	P38646	0.253	0.206	0.779	1.688	0.215	0.550	0.231	0.148	0.325	0.445	0.446	0.317	0.175

Normalized spot intensity (arbitrary units) as well as accession numbers for the Entrez protein database (http://www.ncbi.nlm.nih.gov/entrez/) are shown for the differentially expressed proteins

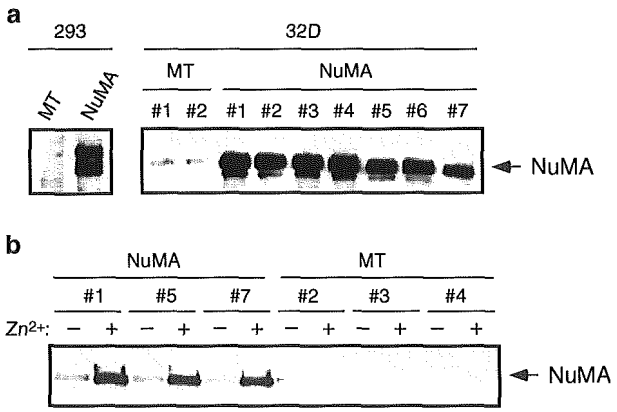
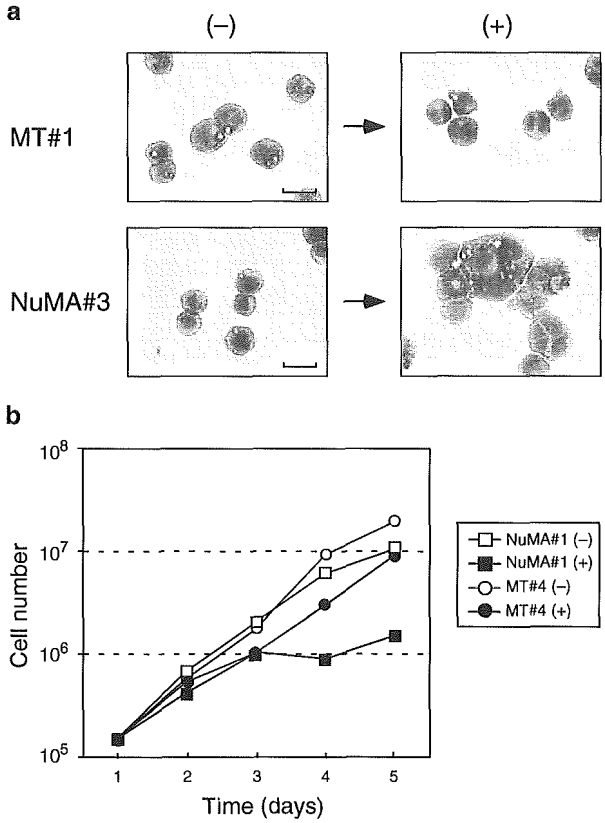


Figure 2 Conditional expression of NuMA in 32D cells. (**a**) Stable transfectants of 32D cells were isolated for pMT-CB6 (MT) or pMT-NuMA (NuMA) vectors. Total cell lysates (10 μg of protein per lane) prepared from cells incubated overnight in the presence of 0.1 mM $ZnSO_4$ were subjected to immunoblot analysis with antibodies to NuMA. Total cell lysates prepared from human kidney 293 cells transiently transfected with pMT-CB6 or pMT-NuMA were similarly analysed. The position of $p240^{NuMA}$ is indicated on the right. (**b**) The transfected 32D clones were cultured overnight with (+) or without (−) $ZnSO_4$ and then analysed for the induction of NuMA expression as in (**a**)

shock cognate protein 70 (HSC70), HSP73, or lipopolysaccharide-associated protein 1 (LAP1)), and HSPA9 (HSP75, mortalin, or GRP75) all function as molecular chaperones. Their expression is induced by a variety of cellular stressors and they are thought to facilitate protein folding and oligomerization. In addition, however, whereas forced expression of one of the two isoforms of HSPA9, HSPA9A (MOT1), induces cell senescence (Kaul *et al.*, 1995), that of HSPA9B (MOT2), which differs from HSPA9A by only two amino acids, promotes cell cycle progression and malignant transformation (Kaul *et al.*, 1998). Although it is not known how MOT proteins induce malignant transformation, the stress-induced tyrosine phosphorylation of these proteins suggests that they function downstream of protein tyrosine kinases. Many molecular chaperones have also been shown to possess antiapoptotic activity (Jaattela, 1999). An increased expression of HSP70 family proteins in leukemic blasts might thus be directly linked to leukemogenesis or to the development of resistance to chemotherapeutic drugs.

The redox state of cells reflects a precise balance between the production of reactive oxygen species and the activity of reducing agents, the latter of which include thiol-based buffers and SOD (Davis *et al.*, 2001). GST functions in cellular detoxification by catalysing the conjugation of reduced glutathione (GSH) to target molecules. In addition, GST-pi has been linked to chemoresistance to doxorubicin (Volm *et al.*, 1992) and to cisplatin (Okuyama *et al.*, 1994), and GST isoforms have been shown to be directly regulated by c-Jun NH_2-terminal kinase (JNK), also known as stress-activated protein kinase (SAPK) (Adler *et al.*, 1999). SOD2, also known as manganese-dependent SOD (MnSOD), is a mitochondrial enzyme that catalyses dismutation of the superoxide anion into O_2 and H_2O_2. Overexpression of this enzyme has been detected in cancer cells (Pang *et al.*, 1997) as well as in neurons of individuals with autosomal recessive parkinsonism (Matsumine *et al.*, 1997; Shimoda-Matsubayashi *et al.*, 1997). It remains to be determined whether an increased expression of GST-pi or SOD2 contributes to leukemogenesis.

NuMA was one of the most abundant soluble proteins in the leukemic blasts from some of the patients analysed in the present study. Given that NuMA is essential for the formation of the mitotic spindle, our observation that the abundance of this protein in leukemic blasts was related to the number of chromosomal abnormalities suggested that aberrant NuMA activity might result in a failure of cells to complete mitosis and lead to the development of chromosomal instability. We indeed demonstrated such effects of NuMA by inducing its expression in mouse 32D cells. However, the forced expression of NuMA also resulted in G_2–M arrest and apoptosis. The mere overexpression of NuMA does not therefore appear to be sufficient for malignant transformation to leukemic cells. Other genetic events that promote cell cycle progression or



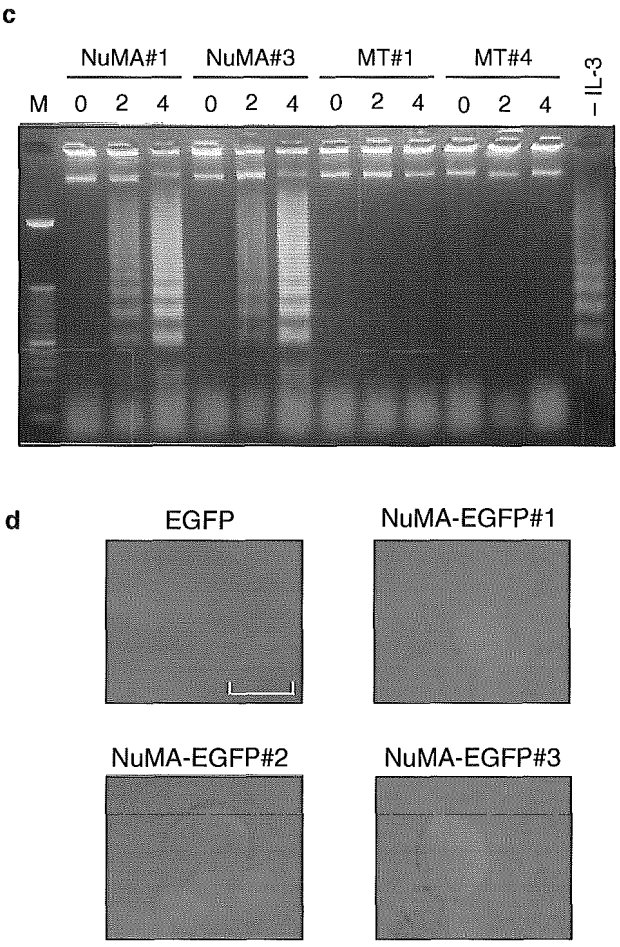


Figure 3 Induction of aneuploidy and apoptosis by expression of $p240^{NuMA}$. (**a**) Mouse 32D cell clones transfected with pMT-CB6 (MT#1) or pMT-NuMA (NuMA#3) were cultured for 4 days in RPMI 1640 medium supplemented with 10% FBS and IL-3 (25 U/ml) either in the absence (−) or presence (+) of 0.1 mM $ZnSO_4$. The cells were then stained with Wright–Giemsa solution and examined by light microscopy. Scale bar; 30 μm. (**b**) Mouse 32D cell clones transfected with pMT-CB6 (MT#4) or pMT-NuMA (NuMA#1) were cultured as in (**a**) in the absence (−) or presence (+) of $ZnSO_4$, and total cell number was determined at the indicated times. (**c**) Mouse 32D cell clones transfected with pMT-CB6 (MT#1 and #4) or pMT-NuMA (NuMA#1 and #3) were cultured as in (**a**) in the presence of IL-3 and $ZnSO_4$ for 0, 2, or 4 days, as indicated. Genomic DNA was then isolated, subjected to agarose gel electrophoresis through a 2% gel, and stained with ethidium bromide. Genomic DNA was also examined for 32D cells in which apoptosis was induced by an overnight deprivation of IL-3 (rightmost lane). Lane M, DNA size markers (50-bp ladder, Invitrogen). (**d**) Human kidney 293 cells transfected with pEGFP (EGFP) or pEGFP/NuMA (NuMA-EGFP#1–3) were subjected to the analysis with a fluorescent microscope. Scale bar; 20 μm

protect cells from apoptosis are likely also required to overcome the proapoptotic activity of NuMA. Chromosomal instability associated with NuMA expression may increase the chance of such genetic events, the occurrence of which may be reflected in changes in the proteome detected in our study.

Current proteomics techniques are limited in their sensitivity for protein detection. In our study, no single 2DE image yielded >300 independent protein spots, which is far fewer than the total number of human proteins predicted from sequencing of the human genome (30 000–40 000 proteins without taking into account the products of alternative RNA splicing). Further improvements in proteomics tools and their application to the direct comparison of protein profiles among background-matched cell fractions prepared from fresh specimens should provide an insight into the intracellular events that underlie malignant transformation in human leukemias.

Figure 3 Continued

Table 3 Cell cycle analysis of NuMA-expressing cells

		Cells (%)		
Clone	*Time with Zn^{2+} (days)*	*G_1*	*S*	*G_2–M*
MT#1	0	29.5	63.1	7.4
	2	27.5	60.5	12.0
NuMA#1	0	31.2	61.1	7.7
	2	33.1	28.9	38.1
NuMA#3	0	30.6	61.0	7.8
	2	31.9	35.2	32.9

Mouse 32D cells transfected with pMT-NuMA (NuMA#1 and #3) or pMT-CB6 (MT#1) were incubated with 0.1 mM $ZnSO_4$ for the indicated times, after which the percentage of cells in G_1, S, or G_2–M phase of the cell cycle was determined by flow cytometry

Table 4 Effect of NuMA induction on the proportion of cells with aneuploidy

Clone	*Time with Zn^{2+} (days)*	*4(n) fraction (%)*
MT#1	0	1.98
	2	1.96
	4	1.35
MT#4	0	1.86
	2	1.11
	4	1.63
NuMA#1	0	1.47
	2	4.87
	4	3.97
NuMA#3	0	1.65
	2	5.59
	4	4.51

Mouse 32D cells transfected with pMT-NuMA (NuMA#1 and #3) or pMT-CB6 (MT#1 and #4) were incubated for the indicated times with 0.1 mM $ZnSO_4$, after which the percentage of cells with an increased DNA contents (4(*n*)) was determined flow cytometry

Materials and methods

Cell culture

Mouse myeloid 32D cells were cultured in RPMI 1640 medium (Invitrogen, Carlsbad, CA, USA) supplemented with 10% fetal bovine serum (FBS) and mouse interleukin-3 (IL-3; 25 U/ml; Invitrogen). Cells transfected with pMT-CB6-based vectors were subjected to selection with the same medium containing G418 (1 mg/ml; Invitrogen). Human kidney 293 cells (American Type Culture Collection) were cultured in DMEM/F12 medium (Invitrogen) supplemented with 10% FBS and 1 mM L-glutamine.

Isolation of Blast Bank samples

With written informed consent, BM MNCs had been isolated by Ficoll-Hypaque density gradient centrifugation from the individuals with various hematopoietic disorders. $AC133^+$ cells were purified from such MNCs and stored in our Blast Bank with the use of an immunoaffinity chromatography for AC133 as described previously (Miyazato *et al.*, 2001). Briefly, MNCs were labeled with magnetic bead-conjugated antibody toward AC133 (AC133 MicroBeads; Miltenyi Biotec, Auburn, CA, USA) in phosphate-buffered saline supplemented with 3% FBS and 2 mM EDTA. The cells were then loaded onto miniMACS magnetic cell separation columns (Miltenyi Biotec), and $AC133^+$ cells were purified according to the manufacturer's instruction. To examine the enrichment of $AC133^+$ cells, portions of the MNC and $AC133^+$ cell preparations were stained with Wright-Giemsa solution or analysed with a FACScan flow cytometer (Becton Dickinson, Mountain View, CA, USA) for the expression of CD34, CD38, and AC133. At the end of March 2003, Blast Bank contained 411 independent $AC133^+$ fractions, including 141, 95, and 60 specimens derived from individuals with de novo AML, MDS, or CML, respectively.

2DE of subcellular fractions

A total of 13 $AC133^+$ cell preparations (2×10^6–1×10^7 cells) were subjected to subcellular fractionation by sequential centrifugation as described previously (Watarai *et al.*, 2000). In brief, cells were suspended in 300 μl of a solution containing 20 mM HEPES-NaOH (pH 7.4) and 0.25 M sucrose, and homogenized by 30 strokes in a Dounce homogenizer. The resulting homogenate was centrifuged at 1000 *g* for 10 min, the resulting pellet was saved as the nuclear fraction, and the resulting supernatant was further centrifuged at 10 000 *g* for 10 min. The pellet from this second centrifugation was saved as the mitochondrial fraction, and the supernatant was centrifuged at 100 000 *g* for 60 min to yield the microsomal fraction (pellet) and cytosolic fraction (supernatant). All subcellular fractions were independently subjected to isoelectric focusing on IPG DryStrips (pH 3–10 NL; Amersham Biosciences, Uppsala, Sweden) with a Multiphor II electrophoresis system (Amersham Biosciences). The separated proteins were then subjected to SDS–PAGE through 7.5–15% gradient gels (Bio-Craft, Tokyo, Japan) and visualized by silver staining (Shevchenko *et al.*, 1996).

Protein identification

Stained gels were scanned with an LAS1000 image analyser (Fujifilm, Tokyo, Japan) to generate a digitized image file. Protein spots in the image were detected with the use of Melanie III software (GeneBio, Geneva, Switzerland) with manual adjustment, and were quantified by the same software. The intensity of each spot on a gel was normalized on the basis of the total signal intensity for that gel. The total number of spots detected in one gel image was 168 ± 59 (mean $\pm$ s.d.). A protein spot was classified as differentially expressed if (1) the intensity of the spot was $\geqslant 0.5$ arbitrary units in at least one of the 13 image files and (2) the spot intensity ratio was $\geqslant 5$ for comparison between any pair of gel images. The 11 spots that fulfilled these criteria were excised from the gels and subjected to protease digestion and peptide fingerprinting by MALDI-TOF MS (Voyager ET STR; Applied Biosystems, Foster City, CA, USA). The determined molecular masses of peptides were compared with an in-house nonredundant protein database maintained by Kirin Brewery Co. Ltd.

Quantitative real-time RT–PCR analysis

Total RNA was extracted from the purified $AC133^+$ cells with the use of RNAzol B (Tel-Test, Inc., Friendswood, TX, USA), and a portion of the RNA was converted to double-stranded cDNA by the SuperScript Choice System (Life Technologies, Gaithersburg, MD, USA). The cDNAs were then subjected to PCR with a QuantiTect SYBR Green PCR Kit (Qiagen, Valencia, CA, USA). The amplification protocol comprised incubations at 94°C for 15 s, 60°C for 30 s (54°C in the case of *NuMA* cDNA), and 72°C for 60 s. Incorporation of the SYBR Green dye into 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 PCR products begins. The C_T values for cDNAs corresponding to *GAPDH* and *NuMA* genes were used to calculate the abundance of *NuMA* mRNA relative to that of *GAPDH* mRNA. The oligonucleotide primers for PCR were 5′-GTCAGTGGTGGACCTGACCT-3′ and 5′-TGAGCTTGACAAAGTGGTCG-3′ for *GAPDH* cDNA, and 5′-AAAATAGCCTCATCAGCAGCTTGG-3′ and 5′-CCAGCTTCTGGCTCTTCTCTGACT-3′ for *NuMA* cDNA.

Conditional expression of NuMA in 32D cells

The full-length cDNA for human NuMA was inserted into the pMT-CB6 vector (kindly provided by T Inaba, Hiroshima University, Hiroshima, Japan), which allows Zn^{2+}-dependent expression of exogenous genes in eucaryotic cells. The resulting pMT-NuMA vector or the empty vector was then introduced into 32D cells by electroporation as described previously

(Yamashita *et al.*, 1996). After culture for 2 weeks in the presence of G418, seven and four independent cell clones were isolated from cells transfected with pMT-NuMA or the empty vector, respectively. For the induction of NuMA expression, transfectants were cultured overnight in the presence of 0.1 mM $ZnSO_4$. Proteins were extracted from the transfectants and subjected to immunoblot analysis as described previously (Yamashita *et al.*, 1996); the latter was performed with rabbit polyclonal antibodies to NuMA (Gaglio *et al.*, 1997). The DNA profiles of cells were examined with a FACScan processor (BD Biosciences, San Jose, CA, USA) and cell cycle distribution was determined with the ModFIT program (BD Biosciences).

Subcellular localization of $p240^{NuMA}$

The *NuMA* cDNA spanning its total coding sequence was amplified by PCR with Pfu DNA polymerase (Stratagene, La Jolla, CA, USA), and inserted into the pEGFP-N1 vector (BD Biosciences), giving rise to pEGFP/NuMA that encodes a NuMA protein tagged C-terminally with EGFP. The human kidney 293 cells were transfected with pEGFP-N1 or pEGFP/NuMA by the calcium phosphate method, and cultured for 48 h. EGFP or NuMA-EGFP was then detected by the IX71 fluorescent microscope (Olympus, Tokyo, Japan) equipped with a digital CCD camera DP50 (Olympus).

Acknowledgments

This work was supported in part by a grant for Research on the Human Genome, Tissue Engineering, and Food Biotechnology and a grant for the Second-Term Comprehensive 10-Year Strategy for Cancer Control from the Ministry of Health, Labor, and Welfare of Japan; by the Science Research Promotion Fund of the Promotion and Mutual Aid Corporation for Private Schools of Japan; and by a grant from the Research Foundation for Community Medicine, Japan. JO is a research resident of the Japan Health Sciences Foundation.

References

Adler V, Yin Z, Fuchs SY, Benezra M, Rosario L, Tew KD, Pincus MR, Sardana M, Henderson CJ, Wolf CR, Davis RJ and Ronai Z. (1999). *EMBO J.*, **18,** 1321–1334.

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson Jr J, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO and Staudt LM. (2000). *Nature,* **403,** 503–511.

Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, Sallan SE, Lander ES, Golub TR and Korsmeyer SJ. (2002). *Nat. Genet.*, **30,** 41–47.

Compton DA and Cleveland DW. (1993). *J. Cell Biol.*, **120,** 947–957.

The genome international sequencing consortium. (2001). *Nature,* **409,** 860–921.

Davis Jr W, Ronai Z and Tew KD. (2001). *J. Pharmacol. Exp. Ther.*, **296,** 1–6.

Du Q, Stukenberg PT and Macara IG. (2001). *Nat. Cell Biol.*, **3,** 1069–1075.

Duggan DJ, Bittner M, Chen Y, Meltzer P and Trent JM. (1999). *Nat. Genet.*, **21,** 10–14.

Gaglio T, Saredi A and Compton DA. (1997). *J. Cell Biol.*, **131,** 693–708.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES. (1999). *Science,* **286,** 531–537.

Greenberger JS, Sakakeeny MA, Humphries RK, Eaves CJ and Eckner RJ. (1983). *Proc. Natl. Acad. Sci. USA,* **80,** 2931–2935.

Gygi SP, Rochon Y, Franza BR and Aebersold R. (1999). *Mol. Cell. Biol.*, **19,** 1720–1730.

Jaattela M. (1999). *Ann. Med.*, **31,** 261–271.

Kaul SC, Duncan EL, Englezou A, Takano S, Reddel RR, Mitsui Y and Wadhwa R. (1998). *Oncogene,* **17,** 907–911.

Kaul SC, Wadhwa R, Matsuda Y, Hensler PJ, Pereira-Smith OM, Komatsu Y and Mitsui Y. (1995). *FEBS Lett.*, **361,** 269–272.

Lian Z, Kluger Y, Greenbaum DS, Tuck D, Gerstein M, Berliner N, Weissman SM and Newburger PE. (2002). *Blood,* **100,** 3209–3220.

Lian Z, Wang L, Yamaga S, Bonds W, Beazer-Barclay Y, Kluger Y, Gerstein M, Newburger PE, Berliner N and Weissman SM. (2001). *Blood,* **98,** 513–524.

Makishima H, Ishida F, Ito T, Kitano K, Ueno S, Ohmine K, Yamashita Y, Ota J, Ota M, Yamauchi K and Mano H. (2002). *Br. J. Haematol.*, **118,** 462–469.

Matsumine H, Saito M, Shimoda-Matsubayashi S, Tanaka H, Ishikawa A, Nakagawa-Hattori Y, Yokochi M, Kobayashi T, Igarashi S, Takano H, Sanpei K, Koike R, Mori H, Kondo T, Mizutani Y, Schaffer AA, Yamamura Y, Nakamura S, Kuzuhara S, Tsuji S and Mizuno Y. (1997). *Am. J. Hum. Genet.*, **60,** 588–596.

Miyazato A, Ueno S, Ohmine K, Ueda M, Yoshida K, Yamashita Y, Kaneko T, Mori M, Kirito K, Toshima M, Nakamura Y, Saito K, Kano Y, Furusawa S, Ozawa K and Mano H. (2001). *Blood,* **98,** 422–427.

Ohmine K, Ota J, Ueda M, Ueno S-I, Yoshida K, Yamashita Y, Kirito K, Imagawa S, Nakamura Y, Saito K, Akutsu M, Mitani K, Kano Y, Komatsu N, Ozawa K and Mano H. (2001). *Oncogene,* **20,** 8249–8257.

Okuyama T, Maehara Y, Endo K, Baba H, Adachi Y, Kuwano M and Sugimachi K. (1994). *Cancer,* **74,** 1230–1236.

Pang XP, Hershman JM and Karsan A. (1997). *Oncol. Res.*, **9,** 623–627.

Shevchenko A, Wilm M, Vorm O and Mann M. (1996). *Anal. Chem.*, **68,** 850–858.

Shimoda-Matsubayashi S, Hattori T, Matsumine H, Shinohara A, Yoritaka A, Mori H, Kondo T, Chiba M and Mizuno Y. (1997). *Neurology,* **49,** 1257–1262.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista

C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A and Zhu X. (2001). *Science,* **291,** 1304–1351.

Volm M, Mattern J and Samsel B. (1992). *Cancer,* **70,** 764–769.

Watarai H, Inagaki Y, Kubota N, Fuji K, Nagafune J, Yamaguchi Y and Kadoya T. (2000). *Electrophoresis,* **21,** 460–464.

Yamashita Y, Miyazato A, Ohya K, Ikeda U, Shimada K, Miura Y, Ozawa K and Mano H. (1996). *Jpn. J. Cancer Res.,* **87,** 1106–1110.

Leukemia (2003) 17, 1990–1997

www.nature.com/leu

DNA microarray analysis of hematopoietic stem cell-like fractions from individuals with the M2 subtype of acute myeloid leukemia

Y Oshima[1], M Ueda[2], Y Yamashita[3], YL Choi[3], J Ota[3], S Ueno[3,4], R Ohki[3,4], K Koinuma[3,5], T Wada[3,6], K Ozawa[2], A Fujimura[1] and H Mano[3,7]

[1]*Division of Clinical Pharmacology, Jichi Medical School, Yakushiji, Kawachigun, Tochigi, Japan;* [2]*Division of Hematology, Jichi Medical School, Yakushiji, Kawachigun, Tochigi, Japan;* [3]*Division of Functional Genomics, Jichi Medical School, Yakushiji, Kawachigun, Tochigi, Japan;* [4]*Division of Cardiology, Jichi Medical School, Yakushiji, Kawachigun, Tochigi, Japan;* [5]*Department of Surgery, Jichi Medical School, Yakushiji, Kawachigun, Tochigi, Japan;* [6]*Department of Gynecology, Jichi Medical School, Yakushiji, Kawachigun, Tochigi, Japan; and* [7]*CREST, JST, Saitama, Japan*

Acute myeloid leukemia (AML) may develop *de novo* or secondarily to myelodysplastic syndrome (MDS). Although the clinical outcome of MDS-related AML is worse than that of *de novo* AML, it is not easy to differentiate between these two clinical courses without a record of prior MDS. Large-scale profiling of gene expression by DNA microarray analysis is a promising approach with which to identify molecular markers specific to *de novo* or MDS-related AML. This approach has now been adopted with AC133-positive hematopoietic stem cell-like fractions purified from 10 individuals, each with either *de novo* or MDS-related AML of the M2 subtype. Sets of genes whose activity was associated with either disease course were identified. Furthermore, on the basis of the expression profiles of these genes, it was possible to predict correctly the clinical diagnosis for 17 (85%) of the 20 cases in a cross-validation trial. Similarly, different sets of genes were identified whose expression level was associated with clinical outcome after induction chemotherapy. These data suggest that, at least in terms of gene expression profiles, *de novo* AML and MDS-related AML are distinct clinical entities.

Leukemia (2003) **17,** 1990–1997. doi:10.1038/sj.leu.2403098
Keywords: DNA microarray; acute myeloid leukemia; myelodysplastic syndrome; *DLK*, M2 subtype

Introduction

Myelodysplastic syndrome (MDS) is a clonal disorder of hematopoietic stem cells (HSCs) that affects mostly the elderly (median age of 70 years). MDS is characterized by two clinical manifestations: (i) cytopenia in peripheral blood despite hyper- or normal cellularity in bone marrow (BM), a condition referred to as ineffective hematopoiesis, and (ii) dysplastic changes in blood cells.[1] It is a multistage syndrome, the early stages of which are termed refractory anemia (RA), RA with ringed sideroblasts, or refractory cytopenia with multilineage dysplasia (RCMD).[2]

Between 10 and 50% of cases of MDS, especially those associated with unfavorable chromosomal abnormalities, eventually undergo malignant transformation into MDS-related acute myeloid leukemia (AML), the outcome of which is poor.[3,4] The blasts of individuals with MDS are, in general, refractory to chemotherapeutic agents, and a cure for this condition is rarely achievable except in cases in which allogeneic BM transplantation is applicable. It is therefore clinically important to be able to differentiate between individuals with *de novo* AML and those with MDS-related AML.

Correspondence: Dr H Mano, Division of Functional Genomics, Jichi Medical School, 3311-1 Yakushiji, Kawachigun, Tochigi 329-0498, Japan; Fax: +81 285 44 7322
Received 9 April 2003; accepted 19 June 2003

However, such differentiation is not always an easy task. Some types of dysplasia are apparent even in BM cells of healthy people and such changes become more prominent in the elderly.[5,6] Chromosomal anomalies often associated with MDS, including those affecting chromosomes 5 and 7, are also manifest in some *de novo* AML blasts and indeed are an indicator of poor prognosis in AML patients. It is therefore often difficult to diagnose correctly elderly individuals with AML and dysplastic changes if knowledge of a preceding history of MDS is not available.

Complicating issues even further, certain cases of *de novo* AML (with the absence of prior MDS history or anticancer treatment) may be associated with prominent blood cell dysplasia. The recent proposal for AML classification by the World Health Organization stipulates that individuals with AML and dysplastic cells be diagnosed with AML with multilineage dysplasia (AML-MLD), irrespective of whether the AML is *de novo* or secondary to MDS.[7] It remains to be determined whether AML-MLD is an amalgamation of clinical entities with distinct mechanisms as well as prognoses.

DNA microarray analysis allows the monitoring of the levels of expression of thousands of genes simultaneously[8] and may therefore help to identify molecular markers that differentiate *de novo* AML from MDS-related AML. However, a simple comparison of BM mononuclear cells (MNCs) between these two conditions may be problematic. The cell composition of BM MNCs differs markedly among individuals. Differences in the gene expression profiles between BM MNCs from a given pair of individuals may thus reflect these differences in cell composition. The elimination of such pseudopositive and pseudonegative data necessitates the purification of background-matched cell fractions from the clinical specimens before microarray analysis.

Given that MDS results from a transformation of HSC clones, HSCs are an appropriate target for purification and gene expression analysis. With the use of an affinity purification procedure based on the HSC-specific surface protein AC133,[9] we have therefore initiated a nationwide project aimed at the purification and storage of HSC-like fractions from individuals with leukemia and related disorders in Japan. To date, more than 400 purified cell fractions have been deposited in this 'Blast Bank.'

To further reduce the influence of differentiation commitment of blasts toward certain lineages, we selected for the present study only those Blast Bank samples with the same phenotype, the M2 subtype according to the classification of the French–American–British (FAB) Cooperative Group.[10] We thus characterized the expression profiles of >12 000 genes in AC133$^+$ Blast Bank samples from 10 patients with *de novo* AML of the M2 subtype as well as from 10 individuals with MDS-related

AML of the M2 subtype. Comparison of the microarray data revealed gene markers of potential clinical utility for the diagnosis and prediction of response to chemotherapy.

Materials and methods

Purification of AC133$^+$ cells

BM aspirates were obtained from subjects with written informed consent and the AC133$^+$ HSC-like fraction was purified from each specimen as described previously.[11,12] In brief, MNCs were isolated by Ficoll–Hypaque density gradient centrifugation, labeled with AC133 MicroBeads (Miltenyi Biotec, Auburn, CA, USA), and subjected to chromatography on a mini-MACS magnetic cell separation column (Miltenyi Biotec). Enrichment of the HSC-like fraction was evaluated by subjecting portions of the MNC and AC133$^+$ cell preparations either to staining with Wright–Giemsa solution or to analysis of the expression of CD34, CD38, and AC133 by flow cytometry (FACScan; Becton Dickinson, Mountain View, CA, USA). In most instances, the CD34^{high}CD38^{low} fraction constituted >90% of the eluate of the affinity column.

DNA microarray analysis

Total RNA was extracted from the AC133$^+$ cell preparations by the acid guanidinium method and was subjected to two rounds of amplification with T7 RNA polymerase as described.[13] A total of 1 μg of the amplified cRNA was then converted to double-stranded cDNA, which was used to prepare biotin-labeled cRNA for hybridization with GeneChip HGU95Av2 microarrays (Affymetrix, Santa Clara, CA, USA) harboring oligonucleotides corresponding to a total of 12 625 genes. Hybridization, washing, and detection of signals on the arrays were performed with the GeneChip system (Affymetrix). To evaluate the fidelity of the RNA amplification procedure, we subjected total RNA and cRNA after one (aRNA1) or two (aRNA2) rounds of amplification from the same patient sample independently to the hybridization with the GeneChip system. The resulting expression intensities yielded Pearson's correlation coefficients of 0.807 for total RNA and aRNA1 and of 0.931 for aRNA1 and aRNA2. In the same experiments, the ratio of the number for genes with 'Present' call and those with 'Absent' call was calculated to be 0.261 for total RNA, 0.303 for aRNA1, and 0.171 for aRNA2.

Statistical analysis

The fluorescence intensity for each gene was normalized relative to the fluorescence value for the 50th percentile gene with a 'Present' or 'Marginal' call (Microarray Suite 4.0, Affymetrix) in each hybridization. Hierarchical clustering of the data set and isolation of diagnosis-related genes were performed with GeneSpring 5.0.3 software (Silicon Genetics, Redwood, CA, USA). All raw array data as well as details of the genes shown in the figures are available as supplementary information at the *Leukemia* web site. The raw array data will also be accessible from the Array-Express web site (http://www.ebi.ac.uk/arrayexpress).

Real-time reverse transcription-polymerase chain reaction (RT-PCR) analysis

Portions of nonamplified cDNA were subjected to PCR with a QuantiTect SYBR Green PCR Kit (Qiagen, Valencia, CA, USA). The amplification protocol comprised incubations at 94°C for 15 s, 60°C for 30 s, and 72°C for 60 s. The annealing temperature was raised up to 63°C only for the amplification of the topoisomerase IIβ (TOPIIB) cDNA. Incorporation of the SYBR Green dye into 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 PCR products begins. The C_T values for cDNAs corresponding to the glyceraldehyde-3-phosphate dehydrogenase (GAPDH), *DLK* and *TOPIIB* genes were used to calculate the abundance of *DLK* or *TOPIIB* mRNA relative to that of *GAPDH* mRNA. The oligonucleotide primers for PCR were 5′-GTCAGTGGTGGACCTGACCT-3′ and 5′-TGAGCTTGACAA AGTGGTCG-3′ for *GAPDH* cDNA; 5′-ATCCTGAAGGTGTC CATGAAAG-3′ and 5′-GCACTTGTTGAGGAAGACGATAC-3′ for *DLK* cDNA; and 5′-AGTTGGAAGAGACAATGCCCTC AC-3′ and 5′-TACCAGGCTCCTTCTTCTCCCTCT-3′ for *TOPIIB* cDNA.

Results

Purification of Blast Bank cells

To demonstrate the utility of the Blast Bank strategy, we previously compared the expression of the CD34 gene between BM MNCs and AC133$^+$ cells by microarray analysis.[11] The expression level of the gene was 519 and 42 322 arbitrary units (U) in a healthy volunteer and a patient with MDS-related AML, respectively, in the analysis of BM MNCs. However, in the analysis of AC133$^+$ cells from the same two individuals, the expression levels were 6 749 546 and 5 543 512 U in the volunteer and patient, respectively. The apparent induction of the CD34 gene in the MNCs of the patient thus actually reflected the expansion of the immature blasts expressing this gene.

From our Blast Bank depository, we selected 20 specimens derived from individuals with AML of the FAB M2 subtype, half of them ($n=10$) with *de novo* AML and the other half ($n=10$) with MDS-related AML. The clinical characteristics of the 20 patients are summarized in Table 1. The M2-specific karyotype anomaly, t(8;21), was apparent in five of the *de novo* AML patients, but in none of those with MDS-related AML. All patients were treated with anthracyclin- and/or cytarabine-based regimens, with most of those with MDS-related AML being refractory to treatment.

For the expression profiling of AC133$^+$ cells to be meaningful, such fractions should contain substantial numbers of leukemic clones. We were able to purify $>1\times10^6$ AC133$^+$ cells from the BM MNCs of our patients with AML M2, whereas the yield of AC133$^+$ cells was only 1000–5000 cells from similar starting numbers of BM MNCs obtained from healthy volunteers (data not shown). The purified HSC-like fractions from individuals with AML thus likely consist predominantly of leukemic HSC-like cells. This conclusion was supported by our observation that, for one individual with MDS-related leukemia, fluorescence *in situ* hybridization analysis revealed a −5q anomaly in only 54% of BM MNCs but in 96% of the corresponding purified AC133$^+$ cells (data not shown).

Table 1 Clinical characteristics of the patients subjected to microarray analysis

Patient	*Disease*	*Response*	*Karyotype*
AML-1	AML	Failure	−7
AML-2	AML	Failure	−7
AML-3	AML	CR	−7
AML-4	AML	CR	t(8;21)
AML-5	AML	CR	t(8;21)
AML-6	AML	Failure	Normal
AML-7	AML	CR	Normal
AML-8	AML	CR	t(8;21)
AML-9	AML	CR	t(8;21)
AML-10	AML	CR	t(8;21)
MDS-1	MDS	Failure	+8
MDS-2	MDS	Failure	+8
MDS-3	MDS	Failure	Other
MDS-4	MDS	Failure	Normal
MDS-5	MDS	Failure	+8
MDS-6	MDS	Failure	+8
MDS-7	MDS	CR	Other
MDS-8	MDS	Failure	Normal
MDS-9	MDS	Failure	Normal
MDS-10	MDS	Failure	+8

AML, *de novo* AML; MDS, MDS-related AML. The Response column indicates the clinical outcome of the first induction chemotherapy; CR, complete remission.

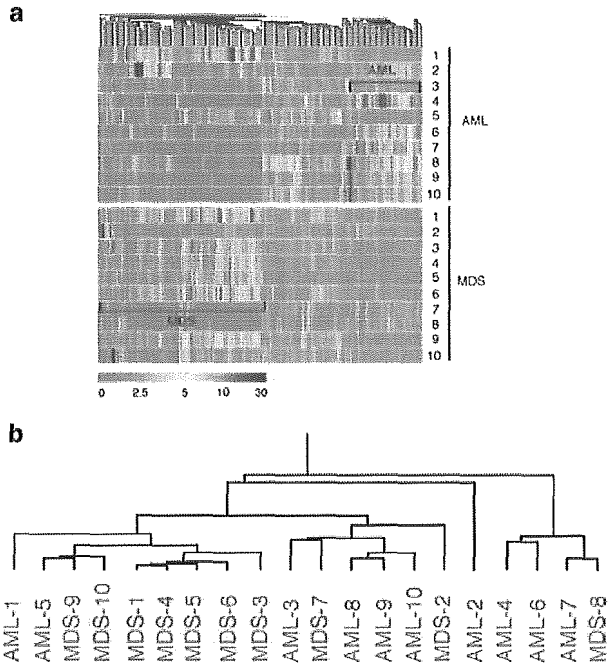


Figure 1 Expression profiles of 6672 genes in leukemic blasts. (a) Hierarchical clustering of 6672 genes on the basis of their expression profiles in Blast Bank samples derived from 10 individuals, each with either *de novo* AML (AML) or MDS-related AML (MDS). Each column represents a single gene on the microarray, and each row a separate patient sample. The positions of large clusters of genes that were expressed preferentially in blasts from patients with *de novo* AML or in those from the patients with MDS-related AML are indicated in blue. (b) Two-way clustering analysis of the *de novo* AML (blue) and MDS-related AML (red) patients based on the similarities in the expression profiles of the 6672 genes shown in (a).

Transcriptome of AML M2 blasts

From the expression data obtained for the >12 000 human genes in the 20 samples analyzed, we first selected genes whose expression received the 'Present' call from the Microarray Suite 4.0 software in at least 10% of the samples, in order to exclude genes that were virtually silent transcriptionally. A total of 6672 genes passed this 'selection window', and their expression profiles in the 20 samples are shown in Figure 1a as a dendrogram, or 'gene tree', in which genes with similar expression profiles (assessed by standard correlation) among the samples are clustered near each other. Clusters of genes that were expressed preferentially in either the *de novo* AML blasts or the MDS-related AML blasts were thus revealed. The genes preferentially expressed in the MDS-related AML blasts were expressed at a substantial level in the blasts from only one (AML-1) of the patients with *de novo* AML.

To evaluate statistically the similarity of the overall gene expression profiles among the 20 samples, we generated another dendrogram, a 'patient tree,' by the two-way clustering method[14] with a separation ratio of 1.0 (Figure 1b). The samples did not cluster into two major disease-specific branches; rather, *de novo* AML and MDS-related AML cases were mixed in several branches.

Identification of disease-associated genes

To identify genes whose expression might allow differentiation between *de novo* AML and MDS-related AML, we first examined those whose expression level differed significantly between the two groups of samples (Welch ANOVA test, $P<0.01$). A total of 574 such genes were identified. Most of these genes, however, were expressed at a low level in both types of specimens, rendering their utility as diagnostic markers uncertain. From these 574 genes, we therefore selected those whose mean expression intensity differed by ⩾5.0 U between the two groups. The resulting 57 'disease-associated' genes are shown in a gene-tree format in Figure 2a; 52 of these genes were specific to MDS-related AML and the remaining five were specific to *de novo* AML. The former group of genes included those for proteins important in regulation of the cell cycle (cyclin D3, GenBank accession number M92287; cyclin I, D50310; CDC10, S72008), in protein synthesis (elongation factor 1α, J04617; eIF4, D30655; elongation factor 1γ-related protein, M55409; elongation factor 2, Z11692; eIF3 p47 subunit, U94855; elongation factor 1α-2, X70940; ribosomal protein S19, M81757; ribosomal protein S4, M58458; ribosomal protein S6, X67309), and in transcriptional regulation (CtBP, U37408; ERF-2, X78992; SP1, X68194). We next performed two-way clustering analysis of the 20 patients based on the expression levels of such 57 disease-associated genes (Figure 2b). The patients clustered into two major branches, one containing only those with *de novo* AML and the other containing mostly those with MDS-related AML.

We also performed 'class prediction' analysis for the samples with GeneSpring software and with the 57 disease-associated genes as the 'class predictor' (http://www.silicongenetics.com/Support/GeneSpring/GSnotes/class_prediction.pdf). In a cross-validation ('drop-one-out') format, the '*k*-nearest-neighbor' samples were counted in Euclidean distance for a 'dropped' sample, and the proportion of neighbor samples from each class was used to calculate the prediction *P*-value for each class (*de novo* AML or MDS-related AML). The resulting class predictions with a *P*-value ratio of <0.2 were determined to be significant. The disease prediction matched the clinical diagnosis for 17

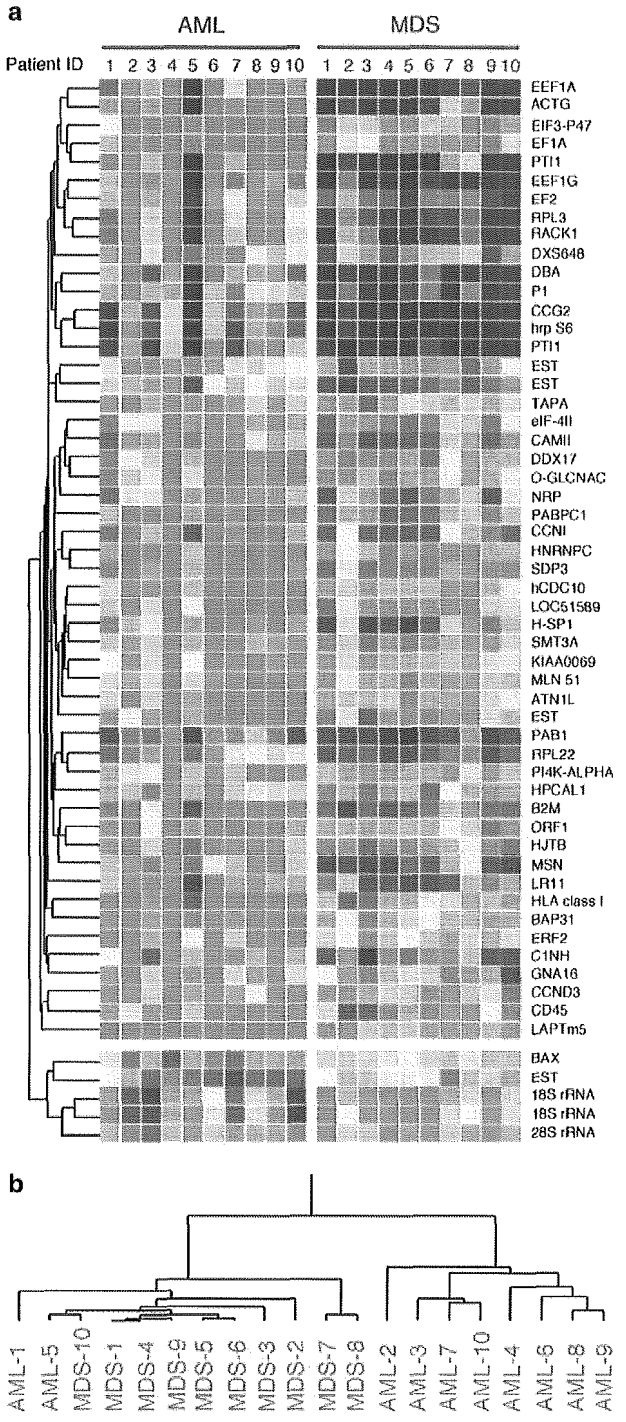


Figure 2 Identification of disease-associated genes. (a) Expression profiles of 57 disease-associated genes in a dendrogram color coded as indicated by the scale in Figure 1a. Each row corresponds to a single gene, and each column to AC133$^+$ cells from a patient with *de novo* AML (AML) or MDS-related AML (MDS). The gene symbols are indicated on the right. The names, accession numbers, and expression intensity data for these genes are available at the Leukemia web site. (b) Two-way clustering analysis of the 20 samples based on the expression levels of the disease-associated genes.

(85%) of the 20 patients; two cases (AML-1 and MDS-8) were not predictable and one case (AML-5) yielded a predicted diagnosis that differed from the clinical one (Table 2).

Table 2 Diagnosis prediction based on gene expression profiles

Patient	*Clinical diagnosis*	*Predicted diagnosis*	*P-value ratio*
AML-1	AML	Unpredictable	0.475
AML-2	AML	AML	0.000
AML-3	AML	AML	0.000
AML-4	AML	AML	0.000
AML-5	AML	MDS	0.000
AML-6	AML	AML	0.000
AML-7	AML	AML	0.000
AML-8	AML	AML	0.000
AML-9	AML	AML	0.000
AML-10	AML	AML	0.000
MDS-1	MDS	MDS	0.000
MDS-2	MDS	MDS	0.019
MDS-3	MDS	MDS	0.000
MDS-4	MDS	MDS	0.000
MDS-5	MDS	MDS	0.019
MDS-6	MDS	MDS	0.000
MDS-7	MDS	MDS	0.13
MDS-8	MDS	Unpredictable	0.255
MDS-9	MDS	MDS	0.000
MDS-10	MDS	MDS	0.000

Prediction of chemosensitivity

Seven of the patients with *de novo* AML and one patient with MDS-related AML (the CR group) underwent complete remission (CR) after the first induction chemotherapy, whereas three patients with *de novo* AML and nine patients with MDS-related AML (the Failure group) did not exhibit a sufficient reduction in the number of immature blasts in their BM in response to such treatment (Table 1). Given that all Blast Bank samples in this study were obtained prior to chemotherapy, we next examined whether it was possible to predict treatment outcome on the basis of the gene expression profiles of the blasts.

For this purpose, we first selected from the 6672 genes expressed in the AC133$^+$ cells of the various patients those whose expression levels differed significantly between the CR and Failure groups (Welch ANOVA, $P<0.01$). The resulting 1353 genes were then examined for those whose mean expression level differed by $\geqslant 5.0$ U between the two groups. A total of 37 'outcome-associated' genes were so identified, 33 of which were preferentially expressed in the Failure group and the remaining four of which were expressed selectively in the CR group (Figure 3a). Unexpectedly, only 16 genes were common to both the disease- and outcome-associated genes.

The genes associated with treatment failure included those for the transcription factor c-FOS (GenBank accession number, V01512), TOPIIB (X68060), INT6 (U62962), and sorting nexin 3 (AF034546). Topoisomerase is a target for anticancer drugs, such as etoposide, that are routinely used in the current treatment of AML.[15] A high level of TOPIIB expression might therefore contribute to drug resistance in the Failure group. To confirm the outcome-dependent expression of *TOPIIB* gene, we examined the mRNA level of *TOPIIB* by the quantitative RT-PCR analysis. As shown in Figure 3b, *TOPIIB* mRNA was abundant only in the leukemic blasts isolated from individuals with drug-resistant AML.

Two-way clustering analysis based on the expression profiles of the 37 outcome-associated genes yielded two major branches: one containing all eight samples in the CR group and two Failure samples, and the other containing the remaining 10 samples in the Failure group (Figure 3c).

We also performed class prediction analysis with the 37 outcome-associated genes as we did with the disease-associated genes. The clinical response of 15 (75%) of the 20 cases was successfully predicted, that of three cases was not predictable, and that of the remaining two cases was predicted incorrectly (Table 3).

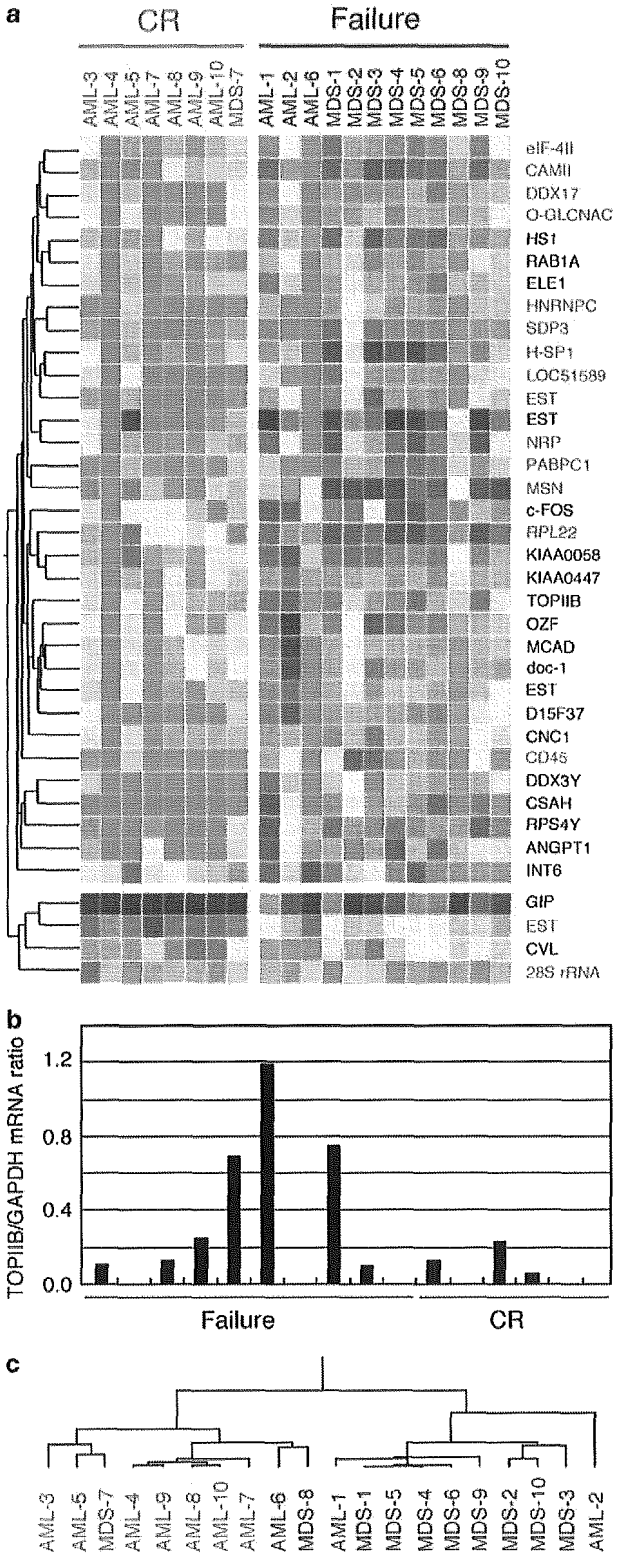


Figure 3 Identification of clinical outcome-associated genes. (a) Expression profiles of 37 outcome-associated genes are shown in a dendrogram color coded as indicated by the scale in Figure 1a. Each row corresponds to a single gene, and each column to AC133$^+$ cells from a patient who underwent CR after induction chemotherapy or in whom treatment failed (Failure). The gene symbols are indicated at the right. The genes also identified in Figure 2a are shown in red. (b) Quantitation of TOPIIB mRNA in AC133$^+$ blasts. Complementary DNA was prepared from the AML M2 blasts of the CR and Failure groups and was subjected to real-time PCR analysis with primers specific for the TOPIIB or GAPDH genes. The ratio of the abundance of TOPIIB mRNA to that of GAPDH mRNA was calculated as 2^n, where n is the C_T value for GAPDH cDNA minus the C_T value for TOPIIB cDNA. (c) Two-way clustering analysis of the 20 samples based on the expression levels of the outcome-associated genes. CR and Failure samples are shown in green and purple, respectively.

Identification of single gene markers for chemosensitivity

The gene sets identified in Figures 2a and 3a may represent candidate genes for the construction of custom-made DNA microarrays for disease diagnosis and prediction of clinical outcome, respectively. Given that the availability of DNA microarray systems is still limited in hospitals, however, it would be beneficial to identify a single gene whose expression (at the mRNA or protein level) could be used as a reliable marker for such purposes. For example, it would be useful to be able to predict the drug sensitivity of leukemic blasts by flow cytometric analysis of a cell surface protein. It is unlikely, however, that determination of the expression of any single gene will allow correct diagnosis or prediction of drug sensitivity for all samples. We therefore attempted to identify individual genes whose high expression level might be sufficient to predict drug resistance, but the absence of which may not necessarily imply drug sensitivity. Such predictor genes would thus be inactive in leukemic blasts from all drug-sensitive patients but be active in samples from at least some proportion of drug-resistant individuals.

Table 3 Prediction of response to chemotherapy

Patient	*Clinical outcome*	*Prediction*	*P-value ratio*
AML-1	Failure	Failure	0.007
AML-2	Failure	Unpredictable	0.366
AML-3	CR	CR	0.029
AML-4	CR	CR	0.029
AML-5	CR	CR	0.187
AML-6	Failure	CR	0.006
AML-7	CR	CR	0.029
AML-8	CR	CR	0.002
AML-9	CR	CR	0.029
AML-10	CR	CR	0.002
MDS-1	Failure	Failure	0.007
MDS-2	Failure	Unpredictable	0.943
MDS-3	Failure	Failure	0.007
MDS-4	Failure	Failure	0.007
MDS-5	Failure	Failure	0.007
MDS-6	Failure	Failure	0.007
MDS-7	CR	CR	0.029
MDS-8	Failure	CR	0.006
MDS-9	Failure	Failure	0.007
MDS-10	Failure	Unpredictable	0.943

We first calculated the mean expression levels of each of the 6672 expressed genes in both the CR and Failure groups. Then, with the use of the GeneSpring software, we searched for genes whose expression profiles were significantly similar, with a minimum correlation of 0.95, to that of a hypothetical 'Failure-specific gene' with a mean expression level of 0.0 U in the CR group and of 100.0 U in the Failure group. From the 2354 such genes identified, we then selected those whose expression value was both <3.0 U in all CR samples and $\geqslant 10.0$ U in at least one of the Failure samples. A total of 42 genes was thus shown to be 'Failure specific' (Figure 4a), of which we examined those with the fewest false-positive results (while allowing false-negative ones). The high level of expression of such genes in leukemic blasts would thus be predictive of treatment failure.

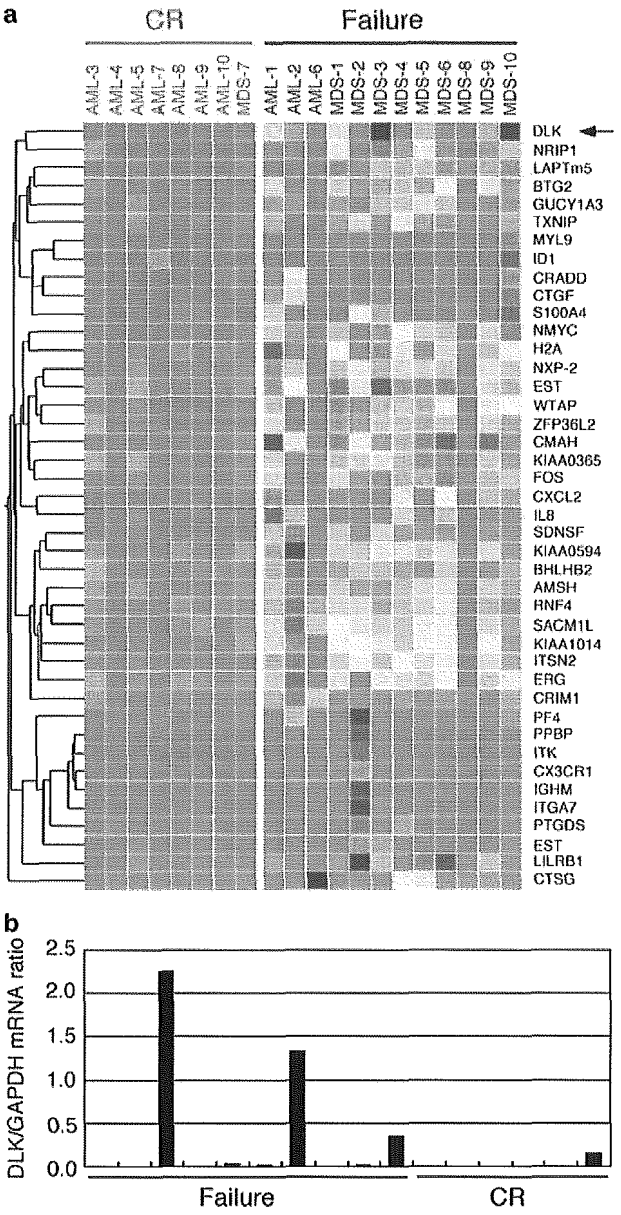


Figure 4 Identification of single gene markers for the prediction of chemoresistance. (a) Dendrogram showing the expression profiles of 42 genes whose expression intensity was <3.0 U in all CR samples and $\geqslant 10.0$ U in at least one sample in the Failure group. The row corresponding to DLK is indicated by an arrow. (b) Quantitation of DLK mRNA in AC133$^+$ blasts. The ratio of the abundance of the DLK mRNA to that of GAPDH mRNA was determined as in Figure 3b.

Among these genes was that for Delta-like (DLK),[16] a membrane protein that is thought to support the self-renewal of HSCs. With a cDNA microarray analysis of Blast Bank samples, we previously showed that *DLK* was transcriptionally active in the AC133$^+$ cells of individuals with MDS-related leukemia but not in those of patients with *de novo* AML.[11] Given the substantial overlap of the Failure group with the MDS-related AML group as well as the limited number of samples analyzed in the present study, it remains to be determined whether a high level of *DLK* expression is related to a specific disease type (MDS-related AML) or to chemosensitivity (treatment failure). However, independent analyses with distinct types of microarray (cDNA array and GeneChip) indicate that *DLK* is specifically activated in the blasts from individuals with overlapping subgroups of leukemia.

To confirm the GeneChip data, we determined the amount of *DLK* mRNA relative to that of *GAPDH* mRNA in the Blast Bank samples analyzed in the present study (Figure 4b). The relative abundance of *DLK* mRNA was significantly correlated with the expression intensity of *DLK* in the GeneChip analysis ($r=0.745$, $P=0.0016$).

Discussion

We have compared the gene expression profiles of AC133$^+$ cells from individuals with two distinct clinical courses of the M2 subtype of AML. Purification of the AC133$^+$ leukemic stem cells from these individuals allowed us to match the specimens as closely as possible prior to microarray analysis. Our study was thus designed to address the question of whether *de novo* AML is a distinct clinical entity from MDS-related AML, and, if so, how distinct?

The overall gene expression profiles did not clearly separate the samples into two major branches corresponding to *de novo* and MDS-related AML. However, two-way clustering analysis of the disease-associated genes resulted in separation of the samples into two groups that closely mirrored the clinical diagnosis. Many of the genes associated with MDS-related AML likely provide growth-promoting or antiapoptotic activities in the MDS blasts. The increased expression levels of genes encoding cyclins, molecules important for protein synthesis, and oncoproteins thus indicate that the proliferative potential of MDS blasts is greater than that of *de novo* AML blasts. A reduced expression level of the gene for the proapoptotic protein BAX[17] in the MDS blasts is also consistent with this notion. The relatively high accuracy of the prediction of clinical diagnosis on the basis of the expression profiles of the disease-associated genes suggests that the products of these genes may play important roles in the pathogenesis.

We also attempted to predict the clinical outcome of the first induction chemotherapy in our study subjects. The genes associated with chemoresistance included those for transcription factors or cofactors such as c-FOS, OZF, SP1, INT6, and ELE1, many of which are implicated in oncogenesis.[18–22] Furthermore, the outcome-associated genes included those for proteins, such as TOPIIB and RNA helicases (DDX17 and DDX3Y),[23] involved in the winding or unwinding of nucleotide strands. Altered expression of these proteins might be linked directly to resistance to certain types of anticancer drugs.

Finally, we attempted to identify single gene markers able to predict chemoresistance of leukemic blasts. Given that the

Failure group includes samples with different karyotypic anomalies, it might be expected that the mechanism of drug resistance of these samples exhibits similar heterogeneity. We therefore searched for genes that were silent transcriptionally in all samples of the CR group but were active in at least some of the samples in the Failure group.

Interestingly, out of the 42 'failure-specific' genes in Figure 4a, five genes encode for chemokines or chemokine receptors, such as proplatelet basic protein (GenBank accession number M54995) or CXC chemokine ligand 7, platelet factor 4 (M25897) or CXC chemokine ligand 4, interleukin 8 (IL8; M17017), chemokine CXC motif ligand 2 (M36820), and chemokine CX3C motif receptor 1 (U20350). Chemokines are a group of small proteins that play a pivotal role in the regulation of the immune system and of cell trafficking as well. Additionally, chemokines can exert pleiotropic effects on target cells, such as mitosis, homing, and even protection against chemotherapeutic reagents. A CC motif chemokine, CCL21, can inhibit, for instance, apoptosis induced by Ara-C.[24] Therefore, it would be possible that chemokine-mediated cell signaling helps to provide drug resistance to the leukemic blasts in our study. Given the high serum levels of IL8 protein in patients with AML and MDS,[25] activation of chemokine/chemokine receptor systems is likely to be relevant to the pathophysiology of leukemic disorders.

Certain types of transcriptional factors and/or their regulators are also activated transcriptionally in the drug-resistant leukemic blasts, including NMYC (GenBank accession number Y00664), basic helix–loop–helix domain containing class B2 (BHLHB2; AB004066), ID1 (X77956), ERG1 (M21535), and c-FOS. Additionally activated were the genes for nuclear proteins, that is, nuclear matrix protein NXP-2 (D50926), Wilms' tumor 1-associated protein 1 (WTAP; X84373), and nuclear receptor interacting protein 1 (NRIP1; X84373). NYMC, ERG1, and c-FOS all mediate mitotic signaling, and ID1 is known to inhibit cell differentiation. WTAP[26] and NRIP1[27] regulate the activity of WT1 and hormone receptors, respectively. Therefore, these gene products may be directly involved in the regulation of cell proliferation and differentiation process.

Conclusion

By narrowing the window of sample selection both to the same FAB subtype of *de novo* and MDS-related AML and to the same level of cell differentiation (as reflected by AC133 expression), we should have minimized population-shift effects that seriously hamper the ability to draw meaningful conclusions from microarray data. Further increases in the number of samples analyzed as well as in the number of genes on the arrays should help to pinpoint genes whose expression levels provide clinically useful information and make it possible to construct custom-made microarrays for the diagnosis of leukemias and the prediction of treatment outcome.

Acknowledgements

We thank all the physicians and patients who participated in the collection of Blast Bank samples. This work was supported in part by grants for Research on the Human Genome, Tissue Engineering, and Food Biotechnology, for the Second-Term Comprehensive 10-Year Strategy for Cancer Control, and for Research on Development of Novel Therapeutic Modalities for Myelodysplastic Syndrome from the Ministry of Health, Labor, and Welfare of Japan; by the Science Research Promotion Fund of the Promotion and Mutual Aid Corporation for Private Schools of Japan; by a grant from Research Foundation for Community Medicine of Japan; by a grant from Sankyo Foundation of Life Science; and by a grant from Takeda Science Foundation. JO is a research resident of the Japan Health Sciences Foundation.

References

1 Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DAG, Gralnick HR *et al.* Proposals for the classification of the myelodysplastic syndromes. *Br J Haematol* 1982; **51**: 189–199.
2 Harris NL, Jaffe ES, Diebold J, Flandrin G, Muller-Hermelink HK, Vardiman J *et al.* World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee Meeting – Airlie House, Virginia, November 1997. *J Clin Oncol* 1999; **17**: 3835–3849.
3 Sole F, Espinet B, Sanz GF, Cervera J, Calasanz MJ, Luno E *et al.* Incidence, characterization and prognostic significance of chromosomal abnormalities in 640 patients with primary myelodysplastic syndromes. Grupo Cooperativo Espanol de Citogenetica Hematologica. *Br J Haematol* 2000; **108**: 346–356.
4 Greenberg P, Cox C, LeBeau MM, Fenaux P, Morel P, Sanz G *et al.* International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood* 1997; **89**: 2079–2088.
5 Bain BJ. The bone marrow aspirate of healthy subjects. *Br J Haematol* 1996; **94**: 206–209.
6 Kuriyama K, Tomonaga M, Matsuo T, Ginnai I, Ichimaru M. Diagnostic significance of detecting pseudo-Pelger–Huet anomalies and micro-megakaryocytes in myelodysplastic syndrome. *Br J Haematol* 1986; **63**: 665–669.
7 Jaffe ES, Harris NL, Stein H, Vardiman JW (eds). *Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues.* Lyon: IARC Press, 2001.
8 Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. *Nat Genet* 1999; **21**: 10–14.
9 Hin AH, Miraglia S, Zanjani ED, Almeida-Porada G, Ogawa M, Leary AG *et al.* AC133, a novel marker for human hematopoietic stem and progenitor cells. *Blood* 1997; **90**: 5002–5012.
10 Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR *et al.* Proposed revised criteria for the classification of acute myeloid leukemia. A report of the French–American–British Cooperative Group. *Ann Intern Med* 1985; **103**: 620–625.
11 Miyazato A, Ueno S, Ohmine K, Ueda M, Yoshida K, Yamashita Y *et al.* Identification of myelodysplastic syndrome-specific genes by DNA microarray analysis with purified hematopoietic stem cell fraction. *Blood* 2001; **98**: 422–427.
12 Ohmine K, Ota J, Ueda M, Ueno S-I, Yoshida K, Yamashita Y *et al.* on of stage progression in chronic myeloid leukemia by DNA microarray with purified hematopoietic stem cells. *Oncogene* 2001; **20**: 8249–8257.
13 Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci USA* 1990; **87**: 1663–1667.
14 Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D *et al.* Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci USA* 1999; **96**: 6745–6750.
15 Valkov NI, Sullivan DM. Drug resistance to DNA topoisomerase I and II inhibitors in human leukemia, lymphoma, and multiple myeloma. *Semin Hematol* 1997; **34**: 48–62.
16 Moore KA, Pytowski B, Witte L, Hicklin D, Lemischka IR. Hematopoietic activity of a stromal cell transmembrane protein containing epidermal growth factor-like repeat motifs. *Proc Natl Acad Sci USA* 1997; **94**: 4011–4016.
17 Chou D, Miyashita T, Mohrenweiser HW, Ueki K, Kastury K, Druck T *et al.* The BAX gene maps to the glioma candidate region at 19q13.3, but is not altered in human gliomas. *Cancer Genet Cytogenet* 1996; **88**: 136–140.
18 Piechaczyk M, Blanchard JM, Marty L, Dani C, Panabieres F, Sabouty SE *et al.* Post-transcriptional regulation of glyceraldehyde-