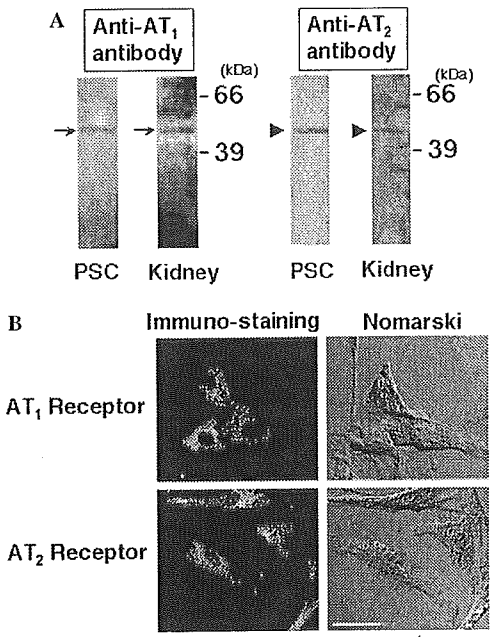


Fig. 1. Presence of AT_1 and AT_2 receptors in PSCs. (A) Western blotting of AT_1 (arrows) and AT_2 (arrow heads) receptors was performed using their specific antibodies. Crude lysate of rat kidney was used as a positive control. Molecular markers are indicated on right. (B) Fluorescence micrographs showing the immunoreactivity of AT_1 and AT_2 receptors in PSCs with corresponding Nomarski images. Immunocytochemistry was carried out with the same first antibodies as those used for Western blotting. Bar: 40 μm.

[^{3}H]thymidine incorporation into PSCs in a dose-dependent manner. The maximum increase was observed at 100 nM. Moreover, losartan, an AT_1 receptor antagonist, inhibited [^{3}H]thymidine incorporation into PSCs enhanced by 100 nM Ang II (Fig. 2B). In contrast, PD123319, an AT_2 receptor antagonist, did not alter [^{3}H]thymidine incorporation into PSCs enhanced by 100 nM Ang II (Fig. 2B). These data indicate that Ang II enhances DNA synthesis in PSCs through AT_1 receptor.

EGF receptor mediates Ang II enhancement of DNA synthesis in PSCs

Ang II receptors are coupled to GTP-binding proteins and possess seven transmembrane domains. Recently, EGF receptor transactivation by GTP-binding protein coupled receptors (GPCRs) has been suggested to be one of the major signaling pathways through which various GPCR-ligands exert their growth-promoting effect on various types of cells [21,22]. We thus hypothesized that Ang II may enhance PSC DNA synthesis through EGF receptor transactivation. To examine this hypothesis, we investigated whether Ang II activates EGF receptor in PSCs. As shown in Fig. 3A, Ang II phosphorylated EGF receptor at its tyrosine

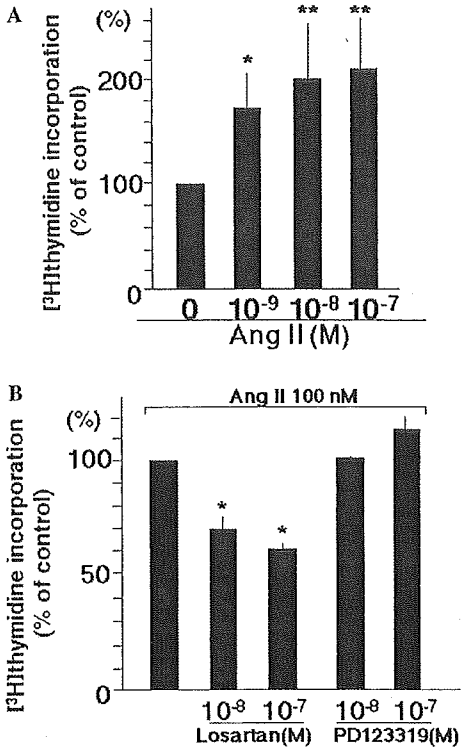


Fig. 2. (A) Effect of Ang II on DNA synthesis in PSCs. Cells were incubated for 48 h with the indicated amounts of Ang II. DNA synthesis was examined by [^{3}H]thymidine incorporation assay. Results are expressed as percent [^{3}H]thymidine incorporation of controls. Values are means ± SEM. The result is representative of four independent experiments with similar results performed in triplicate. (B) Effect of Ang II receptor antagonists on DNA synthesis enhanced by Ang II in PSCs. After 2 h pre-incubation in the presence (losartan; the second and third columns, PD1213319; the fourth and fifth columns) or absence (the first column) of indicated amounts of Ang II receptor antagonists, cells are stimulated for 48 h with 100 nM Ang II. DNA synthesis was examined by [^{3}H]thymidine incorporation assay. Results are expressed as percent [^{3}H]thymidine incorporation of controls. Values are means ± SEM. The result is representative of four independent experiments with similar results performed in triplicate. $^*P < 0.05$, $^{**}P < 0.01$.

residue indicating that Ang II activates EGF receptor. EGF receptor phosphorylation by Ang II was observed at 5 min incubation and maximum phosphorylation was observed at 30 min incubation. Moreover, Ang II promoting effect on PSC DNA synthesis was markedly attenuated by PSC pretreatment with EGF receptor kinase inhibitor AG1478 (Fig. 3B). These data imply that Ang II enhances DNA synthesis in PSCs, at least in part, by transactivating EGF receptor.

Ang II activates ERK through EGF receptor transactivation

We next examined whether Ang II activates ERK, which is downstream of EGF receptor-mediated signaling pathway of cellular growth-promoting stimuli.

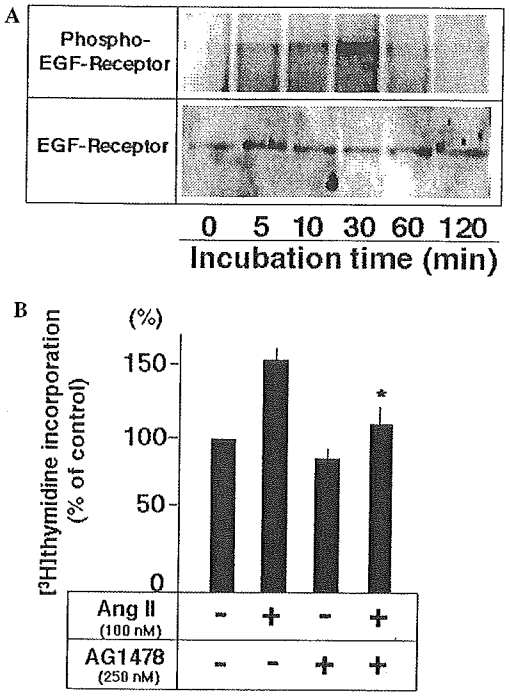


Fig. 3. (A) Effect of Ang II on EGF receptor activation in PSCs. Cells were incubated with 100 nM Ang II for indicated times. The activation of EGF receptor was then determined by Western blotting using anti-phosphorylated EGF receptor antibody (upper panel). Western blotting with anti-EGF receptor antibody (lower panel) was carried out as an internal control. (B) Effect of EGF receptor kinase inhibitor AG1468 on DNA synthesis stimulated with Ang II in PSCs. After 30 min pretreatment with (columns 3 and 4) or without (columns 1 and 2) 250 nM AG1478, cultured PSCs were incubated for 48 h in the presence (lanes 2 and 4) or absence (lanes 1 and 3) of 100 nM Ang II. DNA synthesis was examined by [^{3}H]thymidine incorporation assay. Results are expressed as percent [^{3}H]thymidine incorporation of controls. Values are means ± SEM. The result is representative of four independent experiments with similar results performed in triplicate. *$P < 0.05$ vs. the second column.

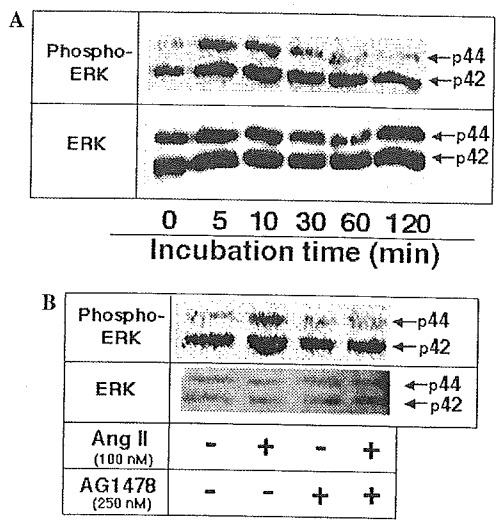


Fig. 4. (A) Effect of Ang II on ERK activation in PSCs. Cells were incubated with 100 nM Ang II for indicated times. The activation of ERK was then determined by Western blotting using anti-phosphorylated ERK antibody (upper panel). Western blotting with anti-ERK antibody (lower panel) was carried out as an internal control. (B) Effect of EGF receptor kinase inhibitor AG1468 on ERK activation stimulated with Ang II in PSCs. After 30 min pretreatment with (lanes 3 and 4) or without (lanes 1 and 2) 250 nM AG1478, cultured PSCs were incubated for 5 min in the presence (lanes 2 and 4) or absence (lanes 1 and 3) of 100 nM Ang II. The activation of ERK was then determined by Western blotting using anti-phosphorylated ERK antibody (upper panel). Western blotting with anti-ERK antibody (lower panel) was carried out as an internal control.

As shown in Fig. 4A, 100 nM Ang II phosphorylated ERK at its tyrosine residue in PSCs, indicating that Ang II activates ERK. Maximum activating effect was observed at 5–10 min incubation. When EGF receptor kinase activity was blocked with AG1478, Ang II failed to activate ERK (Fig. 4B). These data indicate that Ang II activates ERK through EGF receptor transactivation.

Ang II enhances DNA synthesis in PSCs through ERK activation

Knowing that Ang II activates ERK through EGF receptor transactivation, we finally examined whether activated ERK mediates Ang II promoting effect on DNA synthesis in PSCs. For this purpose, we blocked ERK activation by using the MEK1 inhibitor PD98059. As shown in Fig. 5A, pretreatment of PSCs with PD98059 blocked ERK activation by Ang II. Furthermore, Ang II promoting effect on DNA synthesis in PSCs was also attenuated by PSC pretreatment with PD98059 (Fig. 5B). These data indicate that Ang II enhances DNA synthesis in PSCs by activating ERK through EGF receptor transactivation.

Discussion

In this study, we demonstrated that Ang II enhances PSC proliferation through AT_1 receptor. We further elucidated that Ang II augments PSC proliferation by activating ERK through EGF receptor transactivation. These data suggest the participation of Ang II in pancreatic fibrosis by increasing PSC proliferation.

In addition to the traditional action on blood pressure homeostasis, much attention has been directed to Ang II participation in tissue fibrosis. Ang II acts as a growth factor of cardiac fibroblasts (myofibroblasts) and contributes to cardiac remodeling with fibrosis and hypertrophy [23]. In kidney, Ang II induces the proliferation of mesangial cells and fibroblasts, and consequently promotes renal fibrosis [24]. Ang II is also mitogenic for lung fibroblasts and plays a role in pulmonary fibrosis [6]. As to pancreatic fibrosis, Kuno et al. recently reported that angiotensin-converting enzyme inhibitor attenuated pancreatic fibrosis and decreased the number of

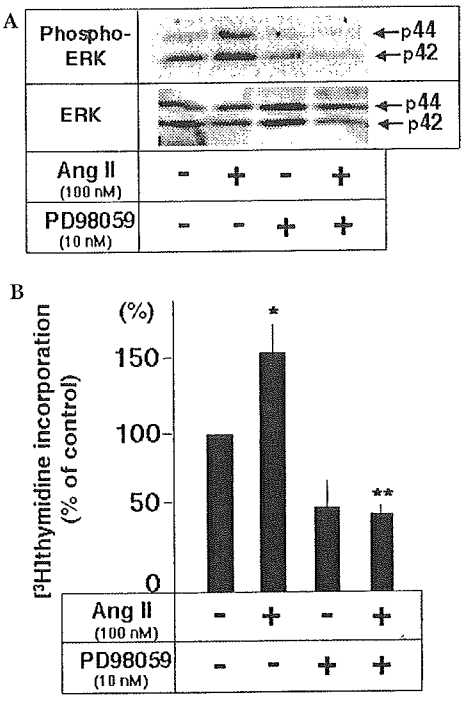


Fig. 5. (A) Effect of MEK inhibitor PD98059 on ERK activation stimulated with Ang II in PSCs. After 2 h pretreatment with (columns 3 and 4) or without (columns 1 and 2) 10 nM PD98059, cultured PSCs were incubated for 5 min in the presence (lanes 2 and 4) or absence (lanes 1 and 3) of 100 nM Ang II. ERK activation was then determined by Western blotting using anti-phosphorylated ERK antibody (upper panel). Western blotting with anti-ERK antibody (lower panel) was carried out as an internal control. (B) Effect of MEK inhibitor PD98059 on DNA synthesis stimulated with Ang II in PSCs. After 2 h pretreatment with (columns 3 and 4) or without (columns 1 and 2) 10 nM PD98059, cultured PSCs were incubated for 48 h in the presence (lanes 2 and 4) or absence (lanes 1 and 3) of 100 nM Ang II. DNA synthesis was examined by [^{3}H]thymidine incorporation assay. Results are expressed as percent [^{3}H]thymidine incorporation of controls. Values are means ± SEM. The result is representative of four independent experiments with similar results performed in triplicate. $P < 0.05$ vs. control, *NS vs. the third column.

activated PSCs in spontaneously occurring chronic pancreatitis in vivo [14]. Since PSCs play a major role in pancreatic fibrosis by synthesizing and secreting various extracellular matrixes such as collagen and fibronectin, their observations suggest the possibility that Ang II may modulate PSC functions. However, its direct evidence had not been demonstrated. Thus, to our knowledge, this is the first evidence of the Ang II regulation of PSCs.

The mechanism of digestive organ fibrosis has been extensively studied on hepatic fibrosis, including Ang II participation. Jonsson et al. [7] revealed that hepatic fibrosis was attenuated by the inhibition of angiotensin-converting enzyme in vivo. Furthermore, Bataller et al. [25] reported that exogenous Ang II induces HSC proliferation in vitro. Taken together, Ang II is assumed to promote hepatic fibrosis by inducing HSC proliferation. Although these reports are consistent with our present data on Ang II action on PSCs, the molecular mechanism of intracellular signaling of Ang II mitogenic effect on HSCs has been still unclear. Thus, we have expanded their studies by elucidating the molecular mechanism of Ang II intracellular signaling in PSCs. We have demonstrated that Ang II activates ERK through EGF receptor transactivation and consequently enhanced PSC proliferation. Warranted is the further study to examine whether similar molecular mechanism underlies the Ang II mitogenic action on HSCs.

Although our current observation with EGF receptor kinase inhibitor AG1478 strongly indicates that Ang II promotes PSC proliferation through EGF receptor transactivation, AG1478 could not completely abolish Ang II enhancement of DNA synthesis in PSCs (Fig. 3B). This might be attributed to the existence of another intracellular signaling pathway of Ang II mitogenic action on PSCs besides EGF receptor transactivation. In this respect, the interplays between GPCRs and TGF-β family signaling have been recently reported. For instance, activin A, a member of TGF-β family, functions as an autocrine inhibitor of DNA synthesis in hepatocytes [26]. We have reported that norepinephrine, a representative GPCR ligand, enhances hepatocyte proliferation by inhibiting activin A signaling with the induction of inhibitory Smad 7 [18]. Since both activin A and TGF-β are also autocrine inhibitors of PSC proliferation [27,28], it is an intriguing open question whether Ang II signaling pathway interacts with TGF-β family signaling in PSCs.

In contrast to EGF receptor kinase inhibitor, MEK inhibitor PD98059 completely abolished Ang II promoting effect on DNA synthesis in PSCs (compare the third and fourth columns in Fig. 5B). This observation suggests that ERK is the major intracellular mediator of Ang II promoting effect on PSC proliferation. Moreover, PD98059 markedly attenuated even basal DNA synthesis in PSCs (compare the first and third columns in Fig. 5B). Since ERK is also a key mediator of mitogenic signals of other growth factors [29], this phenomenon may be attributed to the inhibition of mitogenic effects of other growth factors contained in culture medium used in the present study.

Using Western blotting and immunocytochemsitry, we demonstrated that both AT_1 and AT_2 receptors are present in PSCs (Fig. 1). However, Ang II enhancement of DNA synthesis in PSCs was inhibited by AT_1 receptor blocker losartan, but not by AT_2 receptor blocker PD123319, indicating that AT_1 receptor mediates Ang II stimulation on PSC growth (Fig. 2B). Consistent with our observation, most of Ang II actions well described to date such as vasoconstriction, stimulation of aldosterone release, and promotion of cardiovascular cellular growth are all medicated by AT_1 receptor [30]. As to AT_2 receptor, however, its specific functions have been recently described. AT_2 receptor mediates anti-proliferative

effects in cultured coronary endothelial cells [30,31] and vascular smooth muscle cells [30,32]. Furthermore, AT_2 receptor promotes differentiation in some types of cells [33,34]. Since PSC activation is a kind of differentiation to myofibroblast-like cells and is a cellular function opposite to proliferation, one might speculate that AT_2 receptor could mediate PSC activation. However, Ang II did not exert any effect on PSC activation (Our unpublished data determined with Western blotting using anti-α-SMA antibody). Thus, further studies are needed to elucidate AT_2 receptor function in PSCs.

In conclusion, we have shown that Ang II stimulates PSC proliferation by activating ERK through EGF receptor transactivation. These observations provide new insights into understanding the molecular mechanism of pancreatic fibrosis.

Acknowledgments

We are grateful to Ms. Takako Ishijima for technical and secretarial assistance. This work was supported by Grants-in-Aid from the ministry of Education, Culture, Sports, Science and Technology.

References

[1] M.J. Peach, Renin-angiotensin system: biochemistry and mechanisms of action, Physiol. Rev. 57 (1977) 313–370.
[2] K.T. Weber, Y. Sun, L.C. Katwa, Myofibroblasts and local angiotensin II in rat cardiac tissue repair, Int. J. Biochem. Cell. Biol. 29 (1997) 31–42.
[3] W.F. Ganong, The brain renin-angiotensin system, Annu. Rev. Physiol. 46 (1984) 17–31.
[4] P.S. Leung, H.C. Chan, L.X. Fu, W.L. Zhou, P.Y. Wong, Angiotensin II receptors, AT1 and AT2 in the rat epididymis. Immunocytochemical and electrophysiological studies, Biochim. Biophys. Acta 1357 (1997) 65–72.
[5] G.P. Vinson, E. Saridogan, J.R. Puddefoot, O. Djahanbakhch, Tissue renin-angiotensin systems and reproduction, Hum. Reprod. 12 (1997) 651–662.
[6] R.P. Marshall, R.J. McAnulty, G.J. Laurent, Angiotensin II is mitogenic for human lung fibroblasts via activation of the type 1 receptor, Am. J. Respir. Crit. Care. Med. 161 (2000) 1999–2004.
[7] J.R. Jonsson, A.D. Clouston, Y. Ando, L.I. Kelemen, M.J. Horn, M.D. Adamson, D.M. Purdie, E.E. Powell, Angiotensin-converting enzyme inhibition attenuates the progression of rat hepatic fibrosis, Gastroenterology 121 (2001) 148–155.
[8] P.S. Leung, H.C. Chan, L.X. Fu, P.Y. Wong, Localization of angiotensin II receptor subtypes AT1 and AT2 in the pancreas of rodents, J. Endocrinol. 153 (1997) 269–274.
[9] M. Tahmasebi, J.R. Puddefoot, E.R. Inwang, G.P. Vinson, The tissue renin-angiotensin system in human pancreas, J. Endocrinol. 161 (1999) 317–322.
[10] P.S. Leung, P.O. Carlsson, Tissue renin-angiotensin system: its expression, localization, regulation and potential role in the pancreas, J. Mol. Endocrinol. 26 (2001) 155–164.
[11] P.S. Leung, W.P. Chan, R. Nobiling, Regulated expression of pancreatic renin-angiotensin system in experimental pancreatitis, Mol. Cell. Endocrinol. 166 (2000) 121–128.
[12] S.P. Ip, S.W. Tsang, T.P. Wong, C.T. Che, P.S. Leung, Saralasin, a nonspecific angiotensin II receptor antagonist, attenuates oxidative stress and tissue injury in cerulein-induced acute pancreatitis, Pancreas 26 (2003) 224–229.
[13] W.P. Chan, M.L. Fung, R. Nobiling, P.S. Leung, Activation of local renin-angiotensin system by chronic hypoxia in rat pancreas, Mol. Cell. Endocrinol. 160 (2000) 107–114.
[14] A. Kuno, T. Yamada, K. Masuda, K. Ogawa, M. Sogawa, S. Nakamura, T. Nakazawa, H. Ohara, T. Nomura, T. Joh, T. Shirai, M. Itoh, Angiotensin-converting enzyme inhibitor attenuates pancreatic inflammation and fibrosis in male Wistar Bonn/Kobori rats, Gastroenterology 124 (2003) 1010–1019.
[15] M.G. Bachem, E. Schneider, H. Gross, H. Weidenbach, R.M. Schmid, A. Menke, M. Siech, H. Beger, A. Grunert, G. Adler, Identification, culture, and characterization of pancreatic stellate cells in rats and humans, Gastroenterology 115 (1998) 421–432.
[16] M.V. Apte, P.S. Haber, T.L. Applegate, I.D. Norton, G.W. McCaughan, M.A. Korsten, R.C. Pirola, J.S. Wilson, Periacinar stellate shaped cells in rat pancreas: identification, isolation, and culture, Gut 43 (1998) 128–133.
[17] P.S. Haber, G.W. Keogh, M.V. Apte, C.S. Moran, N.L. Stewart, D.H. Crawford, R.C. Pirola, G.W. McCaughan, G.A. Ramm, J.S. Wilson, Activation of pancreatic stellate cells in human and experimental pancreatic fibrosis, Am. J. Pathol. 155 (1999) 1087–1095.
[18] C. Kanamaru, H. Yasuda, M. Takeda, N. Ueda, J. Suzuki, T. Tsuchida, H. Mashima, H. Ohnishi, T. Fujita, Smad7 is induced by norepinephrine and protects rat hepatocytes from activin A-induced growth inhibition, J. Biol. Chem. 276 (2001) 45636–45641.
[19] H. Ohnishi, N. Ohgushi, S. Tanaka, H. Mogami, R. Nobusawa, H. Mashima, M. Furukawa, T. Mine, O. Shimada, H. Ishikawa, I. Kojima, Conversion of amylase-secreting rat pancreatic AR42J cells to neuronlike cells by activin A, J. Clin. Invest. 95 (1995) 2304–2314.
[20] J. Suzuki, H. Ohnishi, H. Shibata, A. Wada, T. Hirayama, T. Iiri, N. Ueda, C. Kanamaru, T. Tsuchida, H. Mashima, H. Yasuda, T. Fujita, Dynamin is involved in human epithelial cell vacuolation caused by the *Helicobacter pylori*-produced cytotoxinVacA, J. Clin. Invest. 107 (2001) 363–370.
[21] L.M. Luttrell, Y. Daaka, R.J. Lefkowitz, Regulation of tyrosine kinase cascades by G-protein-coupled receptors, Curr. Opin. Cell. Biol. 11 (1999) 177–183.
[22] P.O. Hackel, E. Zwick, N. Prenzel, A. Ullrich, Epidermal growth factor receptors: critical mediators of multiple receptor pathways, Curr. Opin. Cell. Biol. 11 (1999) 184–189.
[23] W. Schorb, G.W. Booz, D.E. Dostal, K.M. Conrad, K.C. Chang, K.M. Baker, Angiotensin II is mitogenic in neonatal rat cardiac fibroblasts, Circ. Res. 72 (1993) 1245–1254.
[24] S.A. Mezzano, M. Ruiz-Ortega, J. Egido, Angiotensin II and renal fibrosis, Hypertension 38 (2001) 635–638.
[25] R. Bataller, P. Gines, J.M. Nicolas, M.N. Gorbig, E. Garcia-Ramallo, X. Gasull, J. Bosch, V. Arroyo, J. Rodes, Angiotensin II induces contraction and proliferation of human hepatic stellate cells, Gastroenterology 118 (2000) 1149–1156.
[26] H. Yasuda, T. Mine, H. Shibata, Y. Eto, Y. Hasegawa, T. Takeuchi, S. Asano, I. Kojima, Activin A: an autocrine inhibitor of initiation of DNA synthesis in rat hepatocytes, J. Clin. Invest. 92 (1993) 1491–1496.
[27] M.L. Kruse, P.B. Hildebrand, C. Timke, U.R. Folsch, W.E. Schmidt, TGFbeta1 autocrine growth control in isolated pancreatic fibroblastoid cells/stellate cells in vitro, Regul. Pept. 90 (2000) 47–52.
[28] N. Ohnishi, T. Miyata, H. Ohnishi, H. Yasuda, K. Tamada, N. Ueda, H. Mashima, K. Sugano, Activin A is an autocrine activator of rat pancreatic stellate cells: potential therapeutic role of follistatin for pancreatic fibrosis, Gut 52 (2003) 1479–1486.
[29] R. Jaster, G. Sparmann, J. Emmrich, S. Liebe, Extracellular signal regulated kinases are key mediators of mitogenic signals in rat pancreatic stellate cells, Gut 51 (2002) 579–584.

[30] D.T. Dinh, A.G. Frauman, C.I. Johnston, M.E. Fabiani, Angiotensin receptors: distribution, signaling and function, Clin. Sci. (Lond.) 100 (2001) 481–492.
[31] M. Stoll, U.M. Steckelings, M. Paul, S.P. Bottari, R. Metzger, T. Unger, The angiotensin AT2-receptor mediates inhibition of cell proliferation in coronary endothelial cells, J. Clin. Invest. 95 (1995) 651–657.
[32] T. Yamada, M. Akishita, M.J. Pollman, G.H. Gibbons, V.J. Dzau, M. Horiuchi, Angiotensin II type 2 receptor mediates vascular smooth muscle cell apoptosis and antagonizes angiotensin II type 1 receptor action: an in vitro gene transfer study, Life Sci. 63 (1998) PL289–PL295.
[33] L. Laflamme, M. Gasparo, J.M. Gallo, M.D. Payet, N. Gallo-Payet, Angiotensin II induction of neurite outgrowth by AT2 receptors in NG108-15 cells. Effect counteracted by the AT1 receptors, J. Biol. Chem. 271 (1996) 22729–22735.
[34] F. Cote, L. Laflamme, M.D. Payet, N. Gallo-Payet, Nitric oxide, a new second messenger involved in the action of angiotensin II on neuronal differentiation of NG108-15 cells, Endocrinol. Res. 24 (1998) 403–407.

ELSEVIER

EXPERIMENTAL HEMATOLOGY

Experimental Hematology 32 (2004) 828–835

DNA microarray analysis of dysplastic morphology associated with acute myeloid leukemia

Chizuko Tsutsumi[a], Masuzu Ueda[b], Yasushi Miyazaki[a], Yoshihiro Yamashita[c], Young Lim Choi[c], Jun Ota[c,d], Ruri Kaneda[c], Koji Koinuma[c], Shin-ichiro Fujiwara[b,c], Hiroyuki Kisanuki[c], Madoka Ishikawa[c], Keiya Ozawa[b], Masao Tomonaga[a], and Hiroyuki Mano[c,d]

[a]*Department of Hematology and Molecular Medicine Unit, Nagasaki University, Nagasaki, Japan; Divisions of* [b]*Hematology and* [c]*Functional Genomics, Jichi Medical School, Kawachigun, Tochigi, Japan;* [d]*CREST, Japan Science and Technology Agency, Saitama, Japan*

(Received 17 February 2004; revised 17 May 2004; accepted 5 June 2004)

Objective. **Acute myeloid leukemia (AML) develops de novo or secondarily to either myelodysplastic syndrome (MDS) or anticancer treatment (therapy-related leukemia, TRL). Prominent dysplasia of blood cells is apparent in individuals with MDS-related AML as well as in some patients with TRL or even with de novo AML. The clinical entity of AML with multilineage dysplasia (AML-MLD) is likely to be an amalgamation of MDS-related AML and de novo AML-MLD. The aim of this study was to clarify, by the use of high-density oligonucleotide microarrays, whether these subcategories of AML are intrinsically distinct from each other.**

Materials and Methods. **The AC133$^+$ hematopoietic stem cell–like fractions were purified from the bone marrow of individuals with de novo AML without dysplasia (n = 15), AML-MLD (n = 11), MDS-related AML (n = 11), or TRL (n = 2), and were subjected to the synthesis of cRNA which was subsequently hybridized to microarray harboring oligonucleotide corresponding to more than 12,000 probe sets.**

Results. **We could identify many genes whose expression was specific to these various subcategories of AML. Furthermore, with the correspondence analysis/three-dimensional projection strategy, we were able to visualize the independent, yet partially overlapping, nature of current AML subcategories on the basis of their transcriptomes.**

Conclusion. **Our data indicate the possibility of subclassification of AML based on gene expression profiles of leukemic blasts.** © **2004 International Society for Experimental Hematology. Published by Elsevier Inc.**

Acute myeloid leukemia (AML) may develop de novo or as a result of either myelodysplastic syndrome (MDS) or anticancer treatment [1]. Given that MDS is characterized by dysplastic changes in differentiated blood cells, individuals with MDS-related leukemia often manifest prominent dysplasia in their blood cells. Therapy-related acute leukemia (TRL) may develop after the administration of alkylating agents, topoisomerase inhibitors, or radiotherapy. The clinical outcome of TRL is generally worse than that of de novo AML [2], and a subset of individuals with TRL also exhibit multilineage dysplasia of blood cells.

A clinical record of a preceding MDS phase is also an indicator of poor prognosis for the individuals with AML. Therefore, to predict the outcome of, and to optimize the treatment for, each AML patient, it would be important to differentiate de novo AML from MDS-related AML and TRL. However, even in the bone marrow (BM) of healthy elderly people, it is not rare to find dysplastic changes (in particular, dyserythropoiesis) in differentiated blood cells [3]. Therefore, the differential diagnosis among such AML-related disorders is not always an easy task in the clinical settings, especially if a prior record of hematopoietic parameters is not available.

Making issues further complicated, prominent dysplasia in blood cells may be found among certain cases of de novo AML, with which prior MDS phases can be excluded [4,5]. It is known that such de novo AML with dysplasia has a poor outcome with conventional chemotherapy, as does MDS-related leukemia [6]. However, Taguchi et al. have argued that the former may be a distinct clinical entity from the

Offprint requests to: Prof. Hiroyuki Mano, M.D., Ph.D., Division of Functional Genomics, Jichi Medical School, 3311-1 Yakushiji, Kawachi-gun, Tochigi 329-0498, Japan; E-mail: hmano@jichi.ac.jp

0301-472X/04 $–see front matter.
doi: 10.1016/j.exphem.2004.06.003

latter based on the finding that the former cases respond far better to allogeneic bone marrow transplantation than the latter one [7]. In the revised classification of AML by the World Health Organization (WHO) [1], an entity of AML with multilineage dysplasia (AML-MLD) has been proposed, which includes both de novo AML with dysplasia and secondary AML from MDS. Whether such amalgamation holds a clinical relevance awaits further studies on this issue.

DNA microarray has made it possible to measure the expression levels in tens of thousands of genes simultaneously, and thus should be a promising tool to discover useful and reliable molecular markers for these AML-related disorders. However, a simple comparison of BM mononuclear cells (MNCs) with DNA microarray is likely to generate a large body of pseudopositive and pseudonegative data, which may only reflect different proportions of blastic cells within BM or the different lineage commitment of leukemic cells [8]. To minimize such "population-shift effect," it should be effective to isolate and compare leukemic blasts at the same differentiation level from AML-related disorders.

Toward this goal, we started the Blast Bank project to purify and store AC133 surface marker [9]–positive hematopoietic stem cell (HSC)–like fractions from patients with a wide range of hematological disorders. Microarray analysis of these Blast Bank specimens has been highly successful in the isolation of molecular markers to differentiate de novo AML from MDS-related leukemia [8,10], and in the identification of genes that may be involved in the stage progression mechanism in chronic myeloid leukemia (CML) [11] or MDS [12]. Further, a proteomics approach with these Bank cells could identify a protein that may be associated with chromosome instability in leukemic cells [13].

We have now determined the expression intensities for more than 12,000 human probe sets in a total of 39 Blast Bank specimens, including those from 15 cases of de novo AML without dysplasia, 11 cases of MDS-related leukemia, 11 cases of AML-MLD, and 2 cases of TRL. The resulting large data set was analyzed to address whether these clinical entities are actually distinct from each other or whether they partially overlap.

Patients and methods

Purification of AC133$^+$ cells

BM aspirates were obtained from the study subjects with written informed consent. From each specimen, MNCs were isolated by Ficoll-Hypaque density gradient centrifugation, and were labeled with magnetic bead–conjugated anti-AC133 monoclonal antibody (AC133 MicroBead; Miltenyi Biotec, Auburn, CA, USA). AC133$^+$ HSC-like fractions were then purified through a miniMACS magnetic cell separation column (Miltenyi Biotec), and 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 the 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 greater than 90% of the eluate of the affinity column.

DNA microarray analysis

Total RNA was extracted from the AC133$^+$ cell preparations by an RNeasy Mini column with RNase-free DNase (both from Qiagen Inc., Valencia, CA, USA), and was subjected to two rounds of amplification of mRNA fractions by T7 RNA polymerase [14]. The high fidelity of the amplification step was confirmed previously [10]. One microgram of the amplified complementary RNA (cRNA) was then converted to double-stranded cDNA by PowerScript reverse transcriptase (BD Biosciences Clontech, Palo Alto, CA, USA), which was used to prepare biotin-labeled cRNA with ENZO BioArray transcript labeling kit (Affymetrix, Santa Clara, CA, USA). Hybridization of the samples with GeneChip HGU95Av2 microarrays was conducted by the GeneChip system (Affymetrix), revealing the expression intensities of 12,625 probe sets in each sample.

The transcriptome of 10 cases each with de novo AML and MDS-related AML has been already reported separately [10], aiming at the comparison between these two clinical conditions with the same differentiation background; the M2 subtype according to the classification of the French-American-British (FAB) Cooperative Group [15].

Statistical analysis

The fluorescence intensity for each gene was normalized relative to the median fluorescence value for all human genes with a "Present" or "Marginal" call (Microarray Suite; Affymetrix) in each hybridization. Hierarchical clustering of the data set and analysis of variance (ANOVA) were performed with GeneSpring 6.0 software (Silicon Genetics, Redwood, CA, USA). Correspondence analysis [16] was performed with the ViSta software (http://www.visual-stats.org) for all genes showing a significant difference. Each sample was plotted in three dimensions based on the coordinates obtained from the correspondence analysis. All array data as well as details of the genes shown in the figures are available as supplementary information at the *Experimental Hematology* web site.

Results

Comparison of AML-MLD and MDS-related AML

Summarized in Table 1 are the clinical characteristics of 39 patients enrolled in this study, including 15 cases with de novo AML without dysplasia, 11 cases with AML-MLD, 11 cases with MDS-related AML, and 2 cases with TRL. The presence of "MLD" was determined according to the definition in the WHO classification [1], by a central review at the Department of Hematology and Molecular Medicine Unit, Nagasaki University, which is also a "central review institute" for the Japan Adult Leukaemia Study Group. It should be noted that favorable karyotypes, t(8;21) and inv(16), were found only in the cases with AML without dysplasia.

According to the WHO proposal of classification, AML-MLD is likely to be an amalgamation of bona fide de novo

Table 1. Patient characteristics

Patient ID	Disease	Age (year)	Sex	Karyotype
1	MDS	79	M	+8
2	MDS	80	M	+8
3	MDS	71	F	Other
4	MDS	44	M	Normal
5	MDS	61	M	+8
6	MDS	69	M	+8
7	AML	83	M	−7
8	MLD	61	M	Other
9	AML	85	M	−7
10	MDS	84	M	−7
11	MDS	57	M	Normal
12	AML	58	M	t(8;21)
13	AML	37	M	t(8;21)
14	AML	84	M	Normal
15	AML	43	M	Normal
16	MLD	41	M	Normal
17	AML	38	M	t(8;21)
18	MDS	69	M	+8
19	AML	49	F	t(8;21)
20	AML	61	F	t(8;21)
21	MLD*	38	M	Normal
22	MLD*	80	M	Normal
23	AML	53	F	−7
24	AML	32	F	Other
25	AML	46	F	Other
26	AML	53	M	Normal
27	MLD*	57	F	+8
28	TRL	59	M	Other
29	TRL	67	M	−7
30	MDS	70	M	Other
31	MLD*	64	M	−7
32	AML	22	F	inv(16)
33	MLD*	16	F	Normal
34	AML	67	M	t(8;21)
35	MLD*	67	M	−7
36	MDS	88	F	Other
37	MLD*	53	M	Normal
38	MLD*	46	M	Other
39	MLD*	50	M	Other

AML, de novo AML; MLD, AML with multilineage dysplasia; MDS, MDS- associated AML, TRL, therapy-related AML; M, male; F, female.

*Individuals proven not to have a prior history of MDS.

AML with dysplasia and secondary AML evolving from an undiscovered phase of MDS. Although the clinical characteristics of the former have not been fully defined, it has been reported that de novo AML-MLD may be associated with poor prognosis [17,18] and, in some cases, with an increased megakaryopoiesis in BM [5].

To clarify directly whether de novo AML-MLD is truly a clinical entity distinct from MDS-related leukemia, we searched for differences between the transcriptomes of $AC133^+$ cells derived from the individuals diagnosed with these two conditions. Among the 11 cases of AML-MLD studied, 9 were revealed not to have prior MDS records, while we could not obtain the clinical information for the other two with regard to their prior MDS history. Therefore, we could not exclude the possibility that the latter cases had evolved from MDS stages. The former nine cases were thus used to measure the difference between de novo AML-MLD and MDS-related secondary AML.

For the expression data set of these 20 subjects, we first set a condition that the expression level of a given gene should receive the "Present" call (from the Microarray Suite 4.0 software) in at least 30% (6 cases) of the samples, aiming to remove transcriptionally silent genes from the analysis. A total of 4851 genes passed this selection window. Toward such genes was then applied a Student's t-test ($p < 0.001$) to extract genes, expression level of which significantly differed between the two classes, de novo AML-MLD and MDS-related AML. However, many of the genes thus identified yet had very low absolute expression levels throughout the samples, even though the ratio of the expression levels between the two classes might be relatively large. To eliminate such "nearly silent" genes and to enrich genes whose expression levels were significantly high in at least one of the classes, we further selected those whose effect size (absolute difference in the mean expression intensities) [19] between the two classes was at least 10 arbitrary units (U).

We could finally identify a total of 56 genes significantly contrasting the two clinical conditions, expression profiles of which are shown in a "gene-tree" format (Fig. 1A). Here genes with similar expression patterns across the samples were clustered near each other. Many of the genes thus identified were preferentially expressed in de novo AML-MLD (upper two-thirds of the tree), while some were so in MDS-related AML (bottom third). Given the association of de novo AML-MLD with dysmegakaryopoiesis in BM, it was of interest to find that the gene for platelet factor 4 (PF4) was preferentially expressed in individuals with this condition. PF4 is a CXC-type chemokine secreted from platelets, and its serum level is known to reflect platelet activities [20]. High production of PF4 from MLD blasts should influence the environment within BM, and may thereby affect megakaryopoiesis.

Were the expression profiles of these 56 genes potent enough to differentiate AML-MLD from MDS-related AML? To examine this possibility, two-way clustering analysis [21] was conducted on the data set to make a "patient tree" among the subjects, based on the standard correlation values with a separation ratio of 1.0 (Fig. 1B). This tree, which reflects the similarity in the expression profiles of the 56 genes among the subjects, showed the presence of a cluster of individuals only with MDS-related AML. However, the large branch at the left contained not only most of the patients with de novo AML-MLD, but also some individuals with MDS-related AML. It was not clear whether the failure in the clear separation of the two clinical categories was due to an inadequacy of the separation power of the clustering method or to an inaccurate clinical diagnosis. Further, it has not been addressed whether de novo AML-MLD should be treated as a single clinical entity distinct

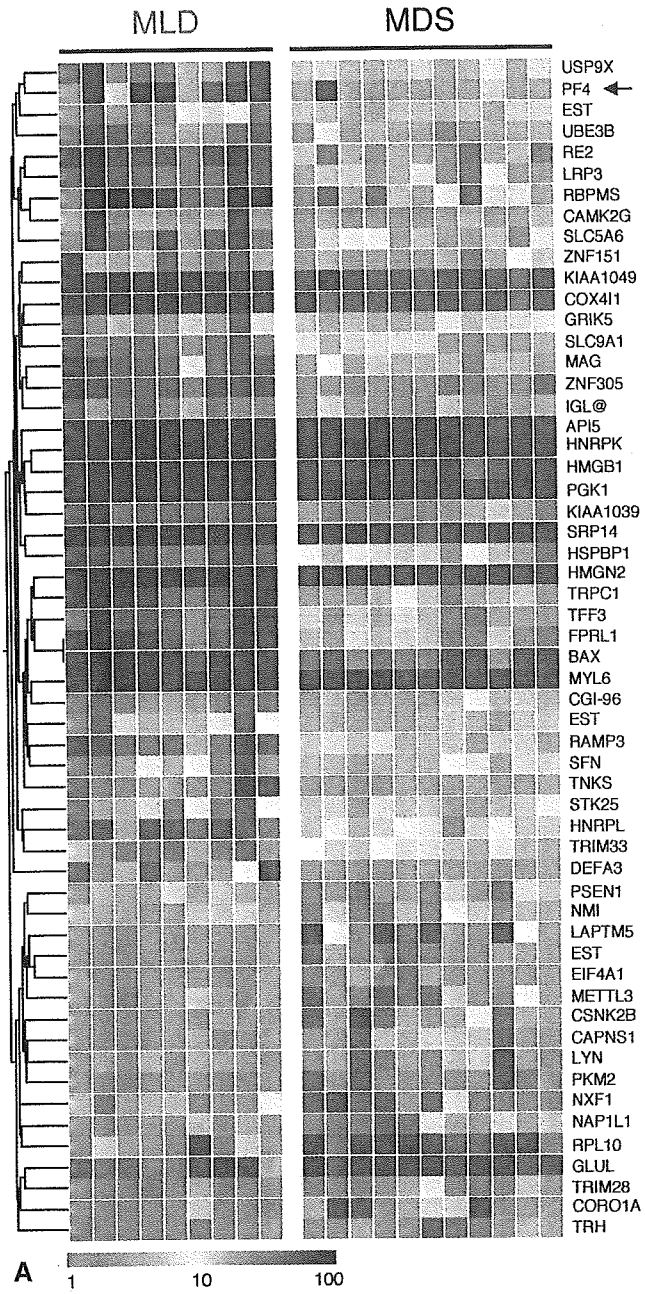


Figure 1. Comparison of gene expression profiles between individuals with de novo AML-MLD and those with MDS-related leukemia. (**A**): Gene tree for the expression levels (color-coded as indicated by the scale at the bottom) of 56 human genes in AC133$^+$ cells from patients with de novo AML-MLD (MLD) or MDS-related leukemia (MDS). Each row corresponds to a single gene and each column to a different patient. The gene symbols are indicated at the right. The position of the *PF4* gene is indicated by an arrow. (**B**): Two-way clustering analysis of the patients with de novo AML-MLD (green) or MDS-related leukemia (red) based on the similarities in the expression profiles of the 56 genes shown in (A). (**C**): Correspondence analysis of the 56 genes identified three major dimensions in their expression profiles. Projection of the specimens into a virtual space with these three dimensions revealed that those from de novo AML-MLD and those from MDS-related leukemia were separated from each other. The arrow indicates a nonconforming specimen (ID 35).

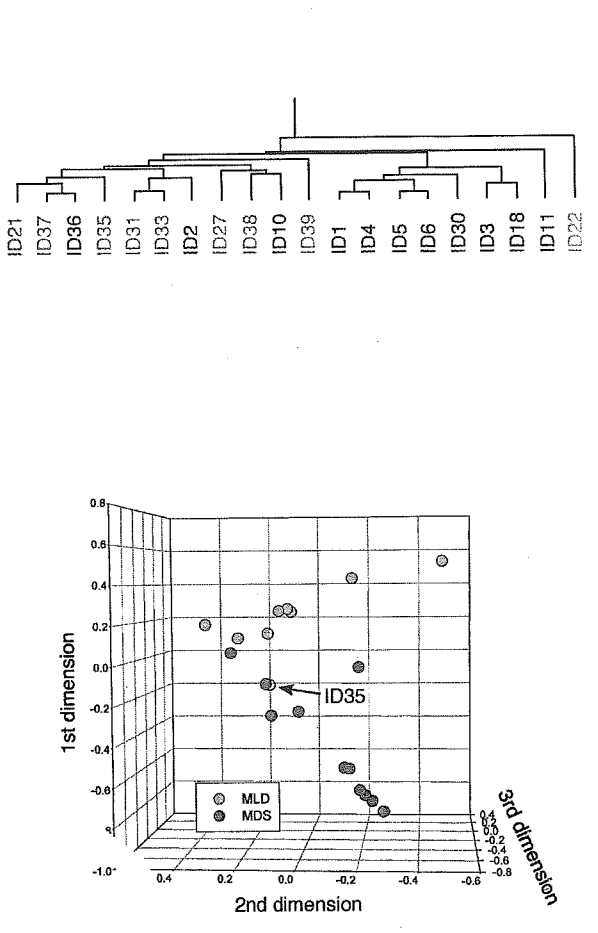


Figure 1. Continued

from MDS-related AML in, at least, the point of view of gene expression profiles.

To address these issues, we tried to visualize the similarity/difference between the two classes. Correspondence analysis is a novel method to decomposite multidimensional data [16]. It enables not only a low-dimensional projection of expression profiles for numerous genes, but measurement of the contribution of each gene to a given extracted dimension and, at the same time, measurement of the contribution of each extracted dimension to the whole complexity. Correspondence analysis was performed on the expression data of the 56 genes in Figure 1A, successfully reducing the complexity of 56 dimensions into 3. On the basis of the calculated three-dimensional (3D) coordinates for each sample, the specimens were then projected into a virtual space (Fig. 1C). It was clear from this figure that most of the samples could be separated into two diagnosis-related groups (whether the coordinate in the first dimension was greater than or equal to 0 or less than 0), supporting the feasibility to set a clinical entity "de novo AML-MLD." Figure 1C also suggests that gene expression profiling could be applied to the differential diagnosis of AML-MLD and MDS-related AML. There was, however, a single patient with AML-MLD

(ID 35) who was misplaced in the MDS group (indicated by an arrow in Fig. 1C).

Comparison of AML without dysplasia and de novo AML-MLD

Similarly, we compared the gene expression profiles between the cases with de novo AML-MLD and AML without dysplastic changes. From the data set of microarray experiments for de novo AML-MLD (n = 9) and de novo AML without dysplasia (n = 15), we selected those whose expression profile received the "Present" call in at least 30% of the cases. Toward such 3608 genes identified, we then applied Student's *t*-test ($p < 0.001$) to extract disease-associated genes between AML-MLD and AML without dysplasia. Further selection with an effect size of at least 10 U led to the identification of four genes whose expression profiles were shown as a gene tree format in Figure 2A.

Similar to the comparison between AML-MLD and MDS-related AML (Fig. 1A), the *PF4* gene was again chosen as a selective marker for AML-MLD. Therefore, among the three classes of AML (de novo AML without dysplasia, de novo AML-MLD, and MDS-related AML), high expression of *PF4* was appreciated only in a single subclass, AML-MLD. It should be also noted that *PF4* was the only gene commonly selected in the comparison of MDS-related AML vs AML-MLD and de novo AML without dysplasia vs AML-MLD.

Two-way clustering of the samples according to the profiles in Figure 2A failed to separate the samples into two major branches corresponding to the clinical diagnosis (Fig. 2B). Three cases of AML-MLD (ID 21, 27, and 35) were misplaced in the large branch of AML without dysplasia, while a patient with AML-MLD (ID 9) was included in the right branch of AML-MLD.

This figure did not clearly tell us how the two conditions are independent or overlapped. Therefore, as in Figure 1C, we tried to construct a 3D view of the samples with the coordinates calculated from correspondence analysis on the four genes. As shown in Figure 2C, the majority of the cases with AML-MLD and AML without dysplasia were separated in the 3D space. In contrast to the prominent separation power of the first dimension in Figure 1C, both of the first and second dimensions in Figure 2C significantly contributed to the separation of the samples. These data indicate that de novo AML without dysplasia can be differentiated from de novo AML-MLD on the basis of gene expression profiles. Again, there was a single subject (ID 27) whose place was incompatible with its clinical diagnosis (indicated by an arrow).

Comparison of de novo AML without dysplasia and MDS-related AML

We have also compared the expression profiles of leukemic blasts between de novo AML (n = 15) and MDS-related leukemia (n = 11). A similar comparison has been previously tried between 10 individuals with de novo AML and

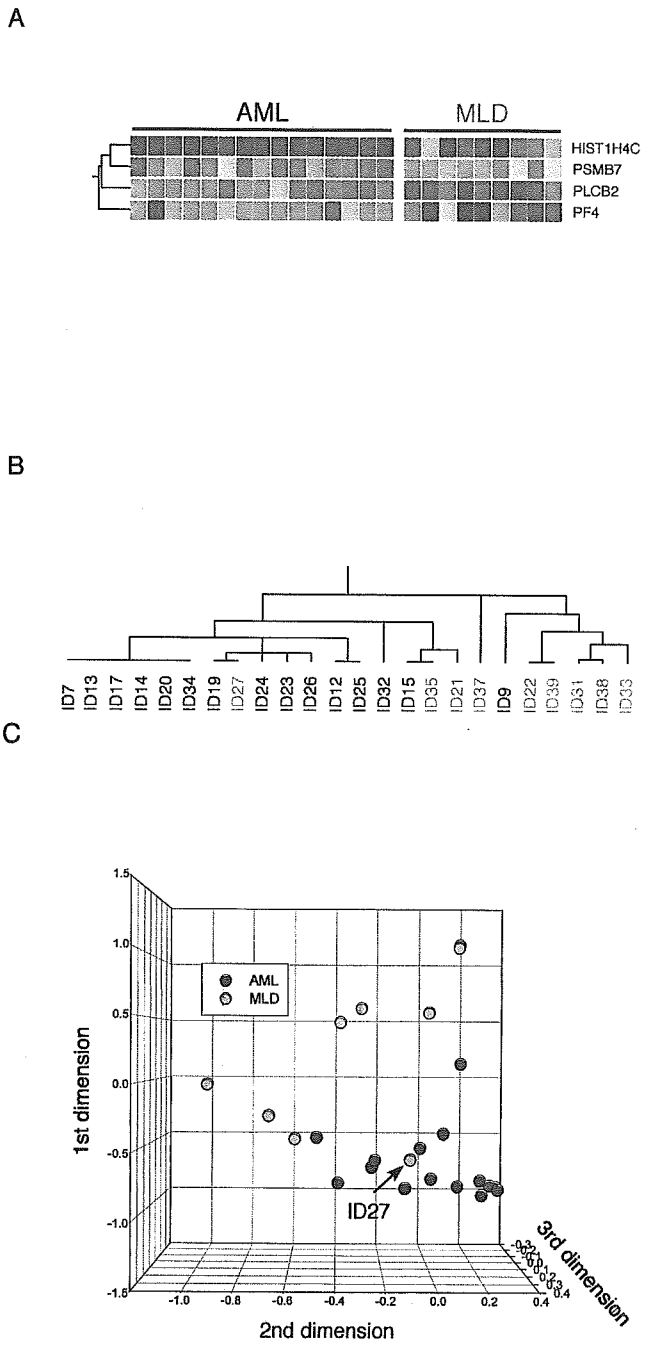


Figure 2. Comparison of gene expression profiles between patients with de novo AML without dysplasia and those with de novo AML-MLD. **(A)**: Gene tree for the expression levels of four human genes in $AC133^+$ cells from individuals with de novo AML without dysplasia (AML) or de novo AML-MLD (MLD). **(B,C)**: Two-way clustering analysis (B) and 3D projection based on correspondence analysis (C) for the patients with de novo AML without dysplasia (blue) or de novo AML-MLD (green) were performed as in Figure 1B and C. The arrow indicates a nonconforming specimen (ID 27).

10 with MDS-related AML matched for the M2 subtype in the FAB classification [10]. In the present study, we identified 30 probe sets (28 genes) whose expression level differed between the two conditions ($p < 0.001$ in Student's *t* test and

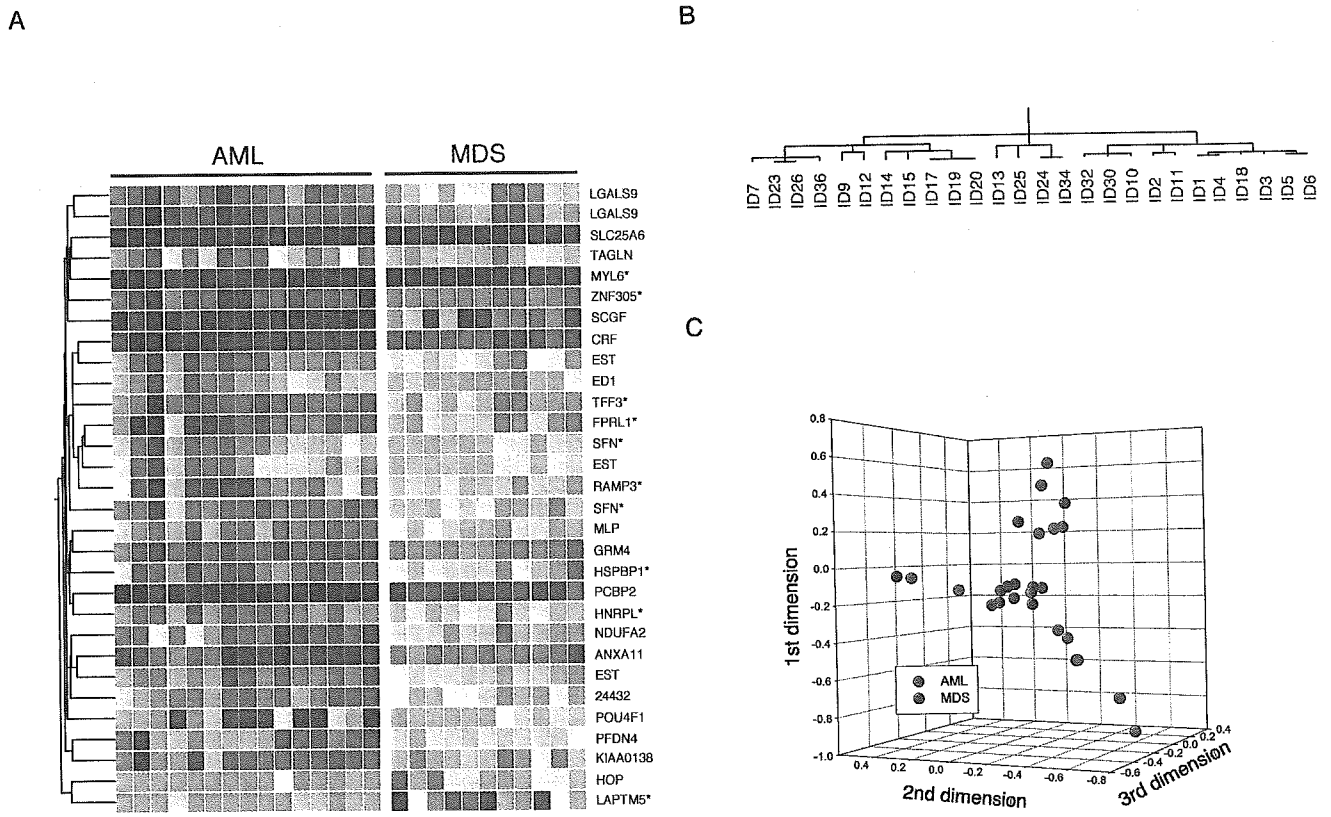


Figure 3. Comparison of gene expression profiles between patients with de novo AML without dysplasia and those with MDS-related leukemia. **(A)**: Expression profiles of 30 probe sets (28 genes) that contrast de novo AML without dysplasia (AML) and MDS-related leukemia (MDS). Two independent probe sets were selected for the *LGALS9* and *SFN* genes. The genes also selected in Figure 1A are indicated by asterisks. **(B,C)**: Two-way clustering analysis (B) and 3D projection (C) of the patients with de novo AML without dysplasia (blue) and those with MDS-related leukemia (red).

an effect size of at least 10 U) (Fig. 3A). Nine of these 28 genes were also among the genes identified in Figure 1A. The gene for lysosomal-associated multispanning membrane protein-5 (LAPTM5) was, for instance, preferentially activated in the MDS-related leukemia but suppressed in AML without dysplasia and AML-MLD. LAPTM5 may be, therefore, a candidate for the novel molecular marker for MDS-related leukemia. All other eight genes were specifically suppressed in MDS-related leukemia compared to AML without dysplasia or AML-MLD.

Two-way clustering analysis of the samples led to generation of three major branches: the left and the center ones composed mostly of the cases of AML without dysplasia (with a misplacement of ID 36), while the right one consisted of cases of MDS-related AML (with a misplacement of ID 32) (Fig. 3B).

To visualize directly the similarity or difference between the two conditions, we also constructed a virtual space with the coordinates obtained from a correspondence analysis on 30 such probe sets (Fig. 3C). Although there was little overlap, in the 3D view, between the AML-MLD and MDS-related AML groups (Fig. 1C), or between the AML without dysplasia and AML-MLD groups (Fig. 2C), in this figure there was a cluster of the samples at the center of the space that contained both individuals with AML without dysplasia and with MDS-related AML. Indeed, the samples in Figure 3C appear to fall into three different groups according to the coordinate for the first dimension. Although the first group (defined by a value of ≥ 0 in the first dimension) and the third group (defined by a value of < -0.3) consisted only of individuals with de novo AML without dysplasia and those with MDS-related AML, respectively, the second group (defined by a value of ≥ -0.3 and < 0) included both types of patients.

Therefore, both the two-way clustering (Fig. 3B) and correspondence analysis (Fig. 3C) indicate that the expression profiles for leukemic blasts of the two clinical entities (AML without dysplasia and MDS-related AML) do not clearly differ from each other. Rather, with regard to transcriptome, the current entities may be partially overlapped.

Comparison of whole samples

Finally, we examined the interrelations among the various subcategories of AML based on the microarray data obtained from all 39 specimens. We combined all the disease-associated genes identified in Figures 1A, 2A, and 3A, and

performed hierarchical clustering analysis of the patients on the basis of the expression profiles of these 78 genes (80 probe sets) (Fig. 4A). Although a few small branches contained only samples from a single clinical entity, the tree failed to group the patients into large diagnosis-related classes.

In contrast to the patient tree, the 3D view of the samples may provide some insight into the independence of each clinical entity (Fig. 4B). There may be three groups of samples in this space, approximately corresponding to the clinical diagnosis (de novo AML without dysplasia, AML-MLD, and MDS-related AML). It is apparent, however, that the central region in this space contained samples of each clinical diagnosis. Clinical history– and cell morphology–based differential diagnosis of AML-related disorders thus remains ambiguous in certain individuals. In this 3D view, the patients with TRL were positioned in close proximity to those with MDS-related AML. The sample size for TRL (n = 2) was too small, however, to draw any conclusion on its relation to the other clinical entities.

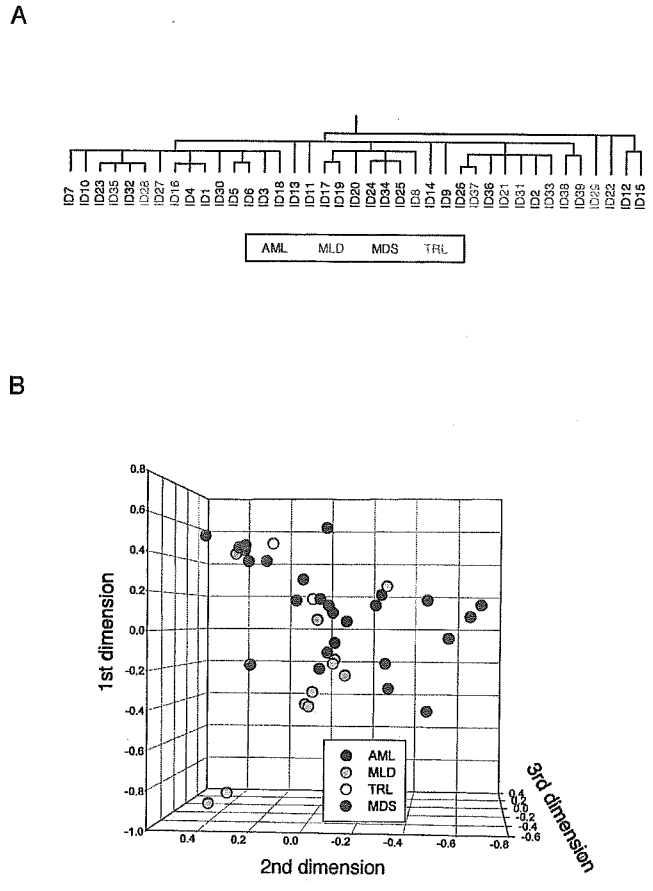


Figure 4. Comparison of all samples. Two-way clustering analysis (**A**) and correspondence analysis (**B**) were performed on all 39 specimens based on the expression profiles of the 80 probe sets whose expression was associated with the clinical diagnosis of de novo AML without dysplasia (AML, blue), AML-MLD (MLD, green), or MDS-related leukemia (MDS, red). The samples from patients with TRL (yellow) were also included in this analysis.

Discussion

We have compared the expression profiles of more than 12,000 human genes among AC133$^+$ HSC-like fractions from 39 individuals with AML-related disorders in an attempt to evaluate the reliability of the current AML classification scheme. We first focused on the relation between de novo AML-MLD and MDS-related leukemia, given that it is currently difficult to discriminate between these two conditions without a clinical history of the patient. Correspondence analysis and 3D projection of the samples suggested that de novo AML-MLD is independent and separable from MDS-related leukemia, at least with regard to gene expression profiles. The 56 genes listed in Figure 1A are thus potential molecular markers for facilitation of the differential diagnosis of the two clinical conditions even in the absence of prior clinical records.

Interestingly, those 56 genes include many related to nuclear functions. All of the high mobility group nucleosomal binding protein 2 (HMGN2), nucleosome assembly protein 1-like 1 (NAP1L1), and high mobility group box 1 (HMGB1) are nuclear proteins that bind to and regulate the structure of double-stranded DNA [22]. On the other hand, heterogeneous nuclear riboprotein L (HNRPL), tunp (or heterogeneous nuclear ribonucleoprotein K [HNRPK]), and nuclear RNA export factor 1 (NXF1) are involved in the maturation process of mRNA. Although the precise relationship of high expression of these genes with dysplastic morphology is not clear yet, our observation may indicate an activated status of nuclear function in the AML-MLD blasts.

Additionally, high expression of ubiquitination-related genes was apparent in de novo AML-MLD. Ubiquitin-protein ligase E3B (UBE3B), for instance, directly catalyzes the transfer reaction of ubiquitin toward the substrates, and thereby regulates their degradation process [23]. Ubiquitin-specific protease 9, X-chromosome (USP9X) is presumed to function as a ubiquitin C-terminal hydrolase [24]. High expression of these genes may reflect a dysregulation in proteasome activity in the leukemic cells.

With our large expression data set (39 samples × 12,625 probe sets), we attempted to isolate genes that characterize the subgroups of AML-related disorders. Also, by applying the correspondence analysis/3D projection approach, we could demonstrate that de novo AML without dysplasia, de novo AML-MLD, and MDS-related AML may have their own expression profile or "molecular signature." However, given the fact that the sample group of each category is partially overlapped in the virtual space, the restricted separation power of the current clinical diagnostic methods may still allow misclassification of patients.

Finally, we should carry in mind a caveat that, despite a large data set (a total of 492,375 data points), the sample

number in each subclass was still small to draw conclusive statements on the identity of AML subcategories. However, our analysis may pave the way for a reorganization of the subcategories of AML in the postgenomic era.

Acknowledgments

We thank all patients and physicians who participated in the Blast Bank project. This work was supported in part by grants 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 a grant from Research Foundation for Community Medicine of Japan; by a grant from Sankyo Foundation of Life Science; by a grant from Takeda Science Foundation; and by a grant from Mitsubishi Pharma Research Foundation.

References

1. Jaffe ES, Harris NL, Stein H, Vardiman JW, eds. Pathology and genetics of tumours of haematopoietic and lymphoid tissues. Lyon: IARC Press; 2001.
2. Smith SM, Le Beau MM, Huo D, et al. Clinical-cytogenetic associations in 306 patients with therapy-related myelodysplasia and myeloid leukemia: the University of Chicago series. Blood. 2003;102:43–52.
3. Bain BJ. The bone marrow aspirate of healthy subjects. Br J Haematol. 1996;94:206–209.
4. Brito-Babapulle F, Catovsky D, Galton DAG. Clinical and laboratory features of de novo acute myeloid leukemia with trilineage myelodysplasia. Br J Haematol. 1987;66:445–450.
5. Jinnai I, Tomonaga M, Kuriyama K, et al. Dysmegakaryocytopoiesis in acute leukaemias: its predominance in myelomonocytic (M4) leukaemia and implication for poor response to chemotherapy. Br J Haematol. 1987;66:467–472.
6. Kuriyama K, Tomonaga M, Matsuo T, et al., Japan Adult Leukaemia Study Group (JALSG). Poor response to intensive chemotherapy in de novo acute myeloid leukaemia with trilineage myelodysplasia. Br J Haematol. 1994;86:767–773.
7. Taguchi J, Miyazaki Y, Yoshida S, et al. Allogeneic bone marrow transplantation improves the outcome of de novo AML with trilineage dysplasia (AML-TLD). Leukemia. 2000;14:1861–1866.
8. Miyazato A, Ueno S, Ohmine K, et al. Identification of myelodysplastic syndrome–specific genes by DNA microarray analysis with purified hematopoietic stem cell fraction. Blood. 2001;98:422–427.
9. Hin AH, Miraglia S, Zanjani ED, et al. AC133, a novel marker for human hematopoietic stem and progenitor cells. Blood. 1997;90:5002–5012.
10. Oshima Y, Ueda M, Yamashita Y, et al. DNA microarray analysis of hematopoietic stem cell–like fractions from individuals with the M2 subtype of acute myeloid leukemia. Leukemia. 2003;17:1990–1997.
11. Ohmine K, Ota J, Ueda M, et al. Characterization of stage progression in chronic myeloid leukemia by DNA microarray with purified hematopoietic stem cells. Oncogene. 2001;20:8249–8257.
12. Ueda M, Ota J, Yamashita Y, et al. DNA microarray analysis of stage progression mechanism in myelodysplastic syndrome. Br J Haematol. 2003;123:288–296.
13. Ota J, Yamashita Y, Okawa K, et al. Proteomic analysis of hematopoietic stem cell–like fractions in leukemic disorders. Oncogene. 2003;22:5720–5728.
14. Van Gelder RN, von Zastrow ME, Yool A, et al. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc Natl Acad Sci U S A. 1990;87:1663–1667.
15. Bennett JM, Catovsky D, Daniel MT, 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.
16. Fellenberg K, Hauser NC, Brors B, et al. Correspondence analysis applied to microarray data. Proc Natl Acad Sci U S A. 2001;98:10781–10786.
17. Goasguen JE, Matsuo T, Cox C, Bennett JM. Evaluation of the dysmyelopoiesis in 336 patients with de novo acute myeloid leukemia: major importance of dysgranulopoiesis for remission and survival. Leukemia. 1992;6:520–525.
18. Kuriyama K, Tomonaga M, Kobayashi T, et al. Morphological diagnoses of the Japan adult leukemia study group acute myeloid leukemia protocols: central review. Int J Hematol. 2001;73:93–99.
19. Dhanasekaran SM, Barrette TR, Ghosh D, et al. Delineation of prognostic biomarkers in prostate cancer. Nature. 2001;412:822–826.
20. Gurney D, Lip GY, Blann AD. A reliable plasma marker of platelet activation: does it exist? Am J Hematol. 2002;70:139–144.
21. Alon U, Barkai N, Notterman DA, 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 U S A. 1999;96:6745–6750.
22. Bustin M. Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. Mol Cell Biol. 1999;19:5237–5246.
23. Gong TW, Huang L, Warner SJ, Lomax MI. Characterization of the human UBE3B gene: structure, expression, evolution, and alternative splicing. Genomics. 2003;82:143–152.
24. Jones MH, Furlong RA, Burkin H, et al. The Drosophila developmental gene fat facets has a human homologue in Xp11.4 which escapes X-inactivation and has related sequences on Yq11.2. Hum Mol Genet. 1996;5:1695–1701.

Leukemia (2004) 18, 556–565

www.nature.com/leu

DNA microarray analysis of natural killer cell-type lymphoproliferative disease of granular lymphocytes with purified CD3$^-$CD56$^+$ fractions

YL Choi[1,2], H Makishima[3], J Ohashi[4], Y Yamashita[1], R Ohki[1], K Koinuma[1], J Ota[1,5], Y Isobe[2], F Ishida[3], K Oshimi[2] and H Mano[1,5]

[1]*Division of Functional Genomics, Jichi Medical School, Kawachigun, Tochigi, Japan;* [2]*Division of Hematology, Department of Medicine, Juntendo University School of Medicine, Tokyo, Japan;* [3]*Second Department of Internal Medicine, Shinshu University School of Medicine, Matsumoto, Nagano, Japan; and* [4]*Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan;* [5]*CREST, JST, Saitama, Japan*

Natural killer (NK) cell-type lymphoproliferative disease of granular lymphocytes (LDGL) is characterized by the outgrowth of CD3$^-$CD16/56$^-$ NK cells, and can be further subdivided into two distinct categories: aggressive NK cell leukemia (ANKL) and chronic NK lymphocytosis (CNKL). To gain insights into the pathophysiology of NK cell-type LDGL, we here purified CD3$^-$CD56$^+$ fractions from healthy individuals ($n=9$) and those with CNKL ($n=9$) or ANKL ($n=1$), and compared the expression profiles of >12 000 genes. A total of 15 'LDGL-associated genes' were identified, and a correspondence analysis on such genes could clearly indicate that LDGL samples share a 'molecular signature' distinct from that of normal NK cells. With a newly invented class prediction algorithm, 'weighted distance method', all 19 samples received a clinically matched diagnosis, and, furthermore, a detailed cross-validation trial for the prediction of normal or CNKL status could achieve a high accuracy (77.8%). By applying another statistical approach, we could extract other sets of genes, expression of which was specific to either normal or LDGL NK cells. Together with sophisticated statistical methods, gene expression profiling of a background-matched NK cell fraction thus provides us a wealth of information for the LDGL condition.

Leukemia (2004) **18,** 556–565. doi:10.1038/sj.leu.2403261
Published online 22 January 2004
Keywords: LDGL; DNA microarray; correspondence analysis

Introduction

Lymphoid cells (10–15%) in peripheral blood (PB) are characterized by the presence of multiple azurophilic granules in pale blue cytoplasm, referred to as large granular lymphocytes (LGLs). Such LGLs originate either from CD3$^+$ T cells or CD3$^-$CD16/56$^-$ natural killer (NK) cells,[1] and sustained outgrowth of LGLs has been designated as lymphoproliferative disease of granular lymphocytes (LDGL),[2] granular lymphocyte-proliferative disorders (GLPD)[3] or LGL leukemia (LGLL).[4]

NK cell-type LDGL can be further subdivided into two distinct categories, that is, aggressive NK cell leukemia (ANKL) and chronic NK lymphocytosis (CNKL).[5] The former is a clonal disorder of NK cells with a very poor outcome. Mono- or oligoclonal Epstein–Barr virus (EBV) genome can be frequently found in an episomal position in these NK cells,[6] suggesting a pathogenetic role of EBV in this disorder. The leukemic NK cells are often refractory to chemotherapeutic reagents, and multiple organ failure is common to ANKL patients.

In contrast, a chronic, indolent course is characteristic to CNKL. Individuals with CNKL are often symptom-free with infrequent fever, arthralgias, and cytopenia, and their NK cells are rarely positive for EBV genome.[7,8] Although the clonality of CNKL cells is still obscure partly due to the limited availability of assessment procedures, one study with X chromosome-linked gene analysis failed to detect clonality in the affected NK cells,[9] suggesting a reactive, rather than a neoplastic, nature of CNKL condition. This hypothesis is further supported by the fact that splenectomy can lead to a sustained elevation of PB NK cell count *in vivo*.[10] However, the hypothesis for the reactive nature of CNKL may be challenged by the fact that some CNKL patients were proved to have a clonal proliferation in NK cells and/or to undergo transformation into NK cell leukemia/lymphoma.[11,12]

Making issues further complicated, the diagnostic criteria for CNKL are not clearly settled yet. Previous reports have proposed the requirement of sustained (>6 months) outgrowth of NK cells in PB (>2.0×10^9 or >0.6×10^9/l)[2,8] for the diagnosis of CNKL. However, NK cell count in the PB of CNKL individuals may fluctuate, and does not always fulfill the criteria. Morice *et al*[13] have reported that affected NK cells may have a restricted expression of a single isoform of killing inhibitory receptors (KIRs), supporting the usefulness of KIR expression as a clonality marker of NK cells.[13] However, these findings yet provide little information for the nature of affected NK cells in the CNKL condition.

DNA microarray enables us to measure the expression level for thousands of genes simultaneously,[14,15] and would be a promising tool to shed light from a new direction on the pathophysiology as well as diagnostic system for LDGL. Gene expression profiling with microarray has, for instance, succeeded in the differential diagnosis between acute myeloid leukemia (AML) and acute lymphoid leukemia (ALL), in extracting novel prognostic markers for prostate cancer,[16] and in the identification of molecular markers for myelodysplastic syndrome (MDS)[17] or chronic myeloid leukemia (CML).[18]

However, simple comparison of tissues or specimens may only yield pseudopositive and pseudonegative data. Although NK cells occupy 10–15% of PB mononuclear cells (MNCs) in healthy individuals, 80–90% of MNCs may be composed of affected NK cells in CNKL patients. If PB MNCs are simply compared between these two groups, any genes specific to NK cells would be considered to be activated in the latter. This misleading result may not reflect any changes in the amount of mRNA per NK cell. To minimize such pseudopositive/pseudonegative data, background-matched NK cell fractions should be purified from healthy individuals as well as LDGL patients prior to microarray analysis. Such approach, referred to as 'back-

Correspondence: Dr H Mano, Division of Functional Genomics, Jichi Medical School, 3311-1 Yakushiji, Kawachigun, Tochigi 329-0498, Japan; Fax: +81-285-44-7322; E-mail: hmano@jichi.ac.jp
This work was supported in part by a grant-in-aid for research on the second-term comprehensive 10-year strategy for cancer control from the Ministry of Health, Labor, and Welfare of Japan, by a grant from Mitsubishi Pharma Research Foundation, by a grant from Takeda Science Foundation, and by a grant from Sankyo Foundation of Life Science.
Received 12 August 2003; accepted 14 November 2003; Published online 22 January 2004

ground-matched population (BAMP) screening',[17] should pinpoint the gene expression alterations truly specific to each condition.

The efficacy of BAMP screening has been already demonstrated by Makishima *et al*[19] in the analysis of $CD4^-CD8^+$ T-cell type LDGL. $CD4^-CD8^+$ fractions were purified from PB MNCs of such LDGL patients and age-matched healthy volunteers, and were subjected to microarray analysis, resulting in the identification of novel molecular markers for T cell-type LDGL.

Analogously, here we isolated $CD3^-CD56^+$ NK cell fractions from healthy volunteers ($n=9$) as well as individuals with CNKL ($n=9$) or with ANKL ($n=1$). By using high-density oligonucleotide microarray, expression profiles for >12 000 human genes were obtained for these purified NK cell specimens. Analysis of the data set with sophisticated statistical methods has clarified that the affected NK cells are clearly distinct from normal ones, at least, with regard to transcriptome.

Materials and methods

Purification of $CD3^-CD56^+$ cells

PB MNCs were isolated by Ficoll–Hypaque density gradient centrifugation from the subjects with informed consent. The cells were incubated with anti-CD3 MicroBeads (Miltenyi Biotec, Auburn, CA, USA), and loaded onto MIDI-MACS magnetic cell separation columns (Miltenyi Biotec) to remove $CD3^+$ cells. The flow-through was then mixed with anti-CD56 MicroBeads (Miltenyi Biotec), and was subjected to a MINI-MACS column for the 'positive selection' of $CD56^+$ cells. Cells bound specifically to the column were then eluted according to the manufacturer's instructions, and stored in aliquots at −80°C.

Enrichment of $CD3^-CD56^+$ NK cell fraction was confirmed in every specimen by subjecting portions of the MNC and column eluates to staining with Wright–Giemsa solution and to the analysis of the cell surface expression of CD3 and CD56 by flow cytometry (FACScan; Becton Dickinson, Mountain View, CA, USA). The $CD3^-CD56^+$ fraction was shown to constitute >90% of each eluate of the affinity column.

DNA microarray analysis

Total RNA was extracted from the $CD3^-CD56^+$ cell preparations by the acid guanidinium method, and was subjected to two rounds of amplification with T7 RNA polymerase as described.[20] High fidelity of our RNA amplification procedure has been already reported.[18] The amplified cRNA (2 μg) 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).

Class prediction by the 'weighted distance method'

The fluorescence intensity for each gene was normalized relative to the median fluorescence value of all human genes on the array in each hybridization. Hierarchical clustering of the data set and isolation of genes specific to the NK cells from healthy individuals (Normal) or to those of patients with LDGL were performed with GeneSpring 5.1 software (Silicon Genetics, Redwood, CA, USA).

In the comparison of normal- and LDGL-$CD3^-CD56^+$ cells, t statistic and effect size (difference in the mean of expression level between normal and LDGL classes)[16] were calculated for each gene. When a gene showed $|t|>3.966$ (corresponding to a significance level of 0.001 in t-test with 17 degrees of freedom) and |effect size|>3, the difference in expression level between two classes was considered statistically significant. The genes showing the significant differences were called as 'informative genes' in this study. Correspondence analysis[21] was then performed with ViSta software (http://www.visualstats.org) for all genes showing a significant difference. Each sample was plotted in three dimensions, based on the coordinates obtained from the correspondence analysis.

To examine whether the informative genes were able to predict the class of the present specimens, we performed class prediction with our 'weighted distance method' (RO *et al*, submitted). This prediction method utilizes the dimensions obtained from correspondence analysis for the informative genes.

Consider a sample X to be predicted from N samples (excluding sample X) in the data set (N_A from class A and N_B from class B). Each sample can be represented by three dimensions, d_1, d_2, and d_3, where d_i denotes the coordinate in the ith dimension for the sample. The weighted distance from sample X to sample Y is defined as $D=\sqrt{\sum_{i=1}^{3}[v_i(d_i^X-d_i^Y)^2]}$, where v_i indicates the contribution of the ith dimension from correspondence analysis, and d_i^X and d_i^Y represent d_i for sample X and sample Y, respectively. Let D_A be the mean value of D from sample X to N_A samples belonging to class A, and D_B be the mean value of D from sample X to N_B samples belonging to class B. When $D_A/(D_A+D_B)<T$, the sample X is assigned class A, and when $D_B/(D_A+D_B)<T$, the sample X is assigned class B, where T is a threshold value. In our analysis, the T value was set to be 0.4. It should be noted that the weighted distance method could be applied to more than two classes.

In a cross-validation trial for the prediction of normal or CNKL class, the entire prediction process mentioned above was repeated for the 18 samples (nine for normal and nine for CNKL). To predict the class of every sample X, the correspondence analysis was carried out for the informative genes obtained from the remaining 17 samples. In this case, the informative genes were selected with a criteria of $|t|>4.073$ (corresponding to a significance level of 0.001 in t-test with 15 degrees of freedom) and |effect size|>3.

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.

'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. 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 dehydrogen-

ase (GAPDH) and interferon-γ (IFNG; GenBank accession number, X13274) genes were used to calculate the abundance of *IFNG* mRNA relative to that of *GAPDH* mRNA. The oligonucleotide primers for PCR were 5′-GTCAGTGGTG-GACCTGACCT-3′ and 5′-TGAGCTTGACAAAGTGGTCG-3′ for *GAPDH*, and 5′-GGGCCAACTAGGCAGCCAACTAA-3′ and 5′-GGAAGCACCAGGCATGAAATCTCC-3′ for *IFNG* cDNA.

Determination of serum level of IFNG protein

Sera were obtained from healthy volunteers and individuals with aplastic anemia (AA), systemic lupus erythematosus (SLE), virus infection-associated hemophagocytic syndrome (VAH), LDGL of $\alpha\beta^+$ T cell, LDGL of $\gamma\delta^+$ T cell, infectious mononucleosis (IMN), CNKL, or ANKL. The serum concentration of IFNG was determined by a flow cytometer with Human Th1/Th2 Cytokine Cytometric Bead Array Kit (BD Biosciences, San Diego, CA, USA) according to the manufacturer's protocols.

Results

Purification of NK cells

To directly compare the transcriptome of normal and affected NK cells, we here purified $CD3^-CD56^+$ fractions from PB MNCs of healthy volunteers ($n=9$) as well as of individuals with CNKL ($n=9$) or ANKL ($n=1$). A total of 19 specimens were thus registered into this study. The clinical characteristics of the 10 patients (CNKL-1 ~ 9 and ANKL-1) are summarized in Table 1. The LGL count in their PB was 14056 cells/ml ± 11695 (mean ± s.d.). The proportion of $CD56^+$ cells in PB MNC was >50% in all affected individuals, indicating a predominant outgrowth of NK cells. All $CD56^+$ fractions in this study were negative for the surface expression of CD3.

The mono- or oligoclonal expansion with regard to EBV infection was confirmed in the NK cells of one CNKL (CNKL-9) and the ANKL (ANKL-1) patients. Importantly, the CNKL-9 patient with monoclonal expansion of EBV^+ NK cells died from leukemic transformation with infiltration into multiple organs at 24 months after the blood sampling. It is, therefore, likely that this patient might have been under a transition process toward ANKL or been at a very early stage of ANKL.

Magnetic bead-based affinity column has succeeded in a substantial enrichment of the NK cell fraction. In one healthy volunteer, for instance, PB MNCs was occupied with 12.3% of $CD3^-CD56^+$ fraction, while the column eluent contained 96.0% of those cells (Figure 1, upper panel). Similar purity of $CD3^-CD56^+$ fraction was also obtained for the patients with CNKL, as demonstrated in the lower panel. The purified cells exhibited a homogenous phenotype of LGL (Figure 1). Successful enrichment of NK cells (>90% purity) was confirmed in every case by flow cytometry and Wright–Giemsa staining of cytospin preparations (not shown). Cell number of the $CD3^-CD56^+$ fractions obtained in each individual was $3.9 \times 10^5 \pm 3.4 \times 10^5$ (mean ± s.d.).

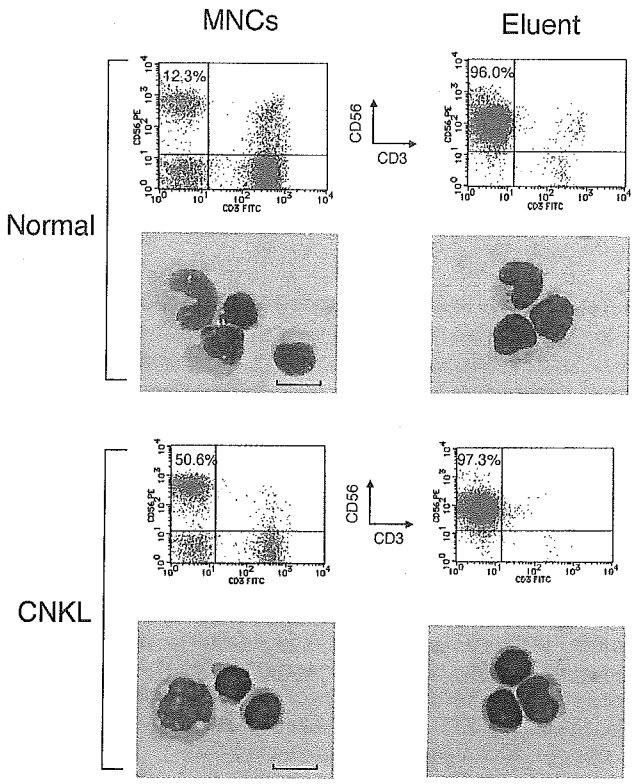


Figure 1 Purification of $CD3^-CD56^+$ fraction. MNCs isolated from PB of a healthy volunteer (normal) and a patient with CNKL were used to purify $CD3^-CD56^+$ fractions (Eluent). Cell surface expression of CD3 and CD56 was monitored in each fraction by flow cytometry, and the proportion (%) of $CD3^-CD56^+$ cells is indicated. Cytospin preparation of each fraction was stained with the Wright–Giemsa solutions. Scale bar, 50 μm.

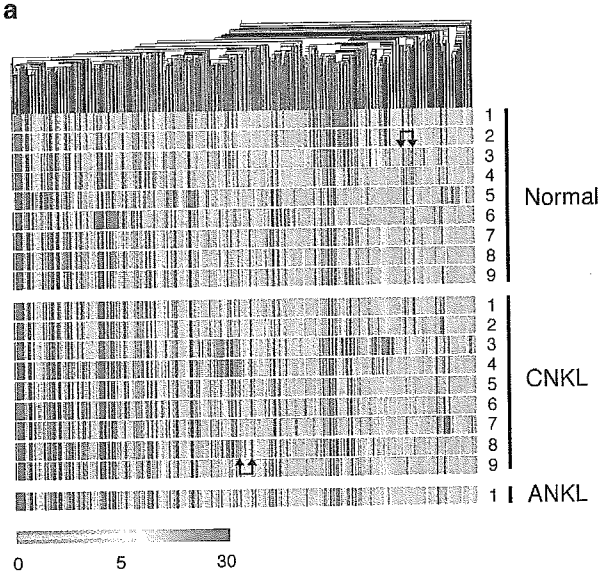
Table 1 Clinical characteristics of the patients with NK cell-type LDGL

Patient	*Disease*	*Age*	*Sex*	*WBC (/mm³)*	*Lymph (/mm³)*	*LGL (/mm³)*	*Hb (g/dl)*	*Plt (×10⁴/mm³)*	*CD3⁺ (%)*	*CD5⁺ (%)*	*CD56⁺ (%)*	*EBV*
CNKL-1	CNKL	70	F	5880	4586	3927	14.6	24.1	13	13	78	–
CNKL-2	CNKL	54	M	10190	7133	6842	14.9	61.8	6	10	94	–
CNKL-3	CNKL	73	F	13780	11024	9721	13.2	24.9	9	7	91	n.d.
CNKL-4	CNKL	51	F	16500	11390	8420	13.1	25	21	23	76	n.d.
CNKL-5	CNKL	55	F	21100	17700	16880	13	17	10	9	79	n.d.
CNKL-6	CNKL	62	M	18000	11880	10252	13.1	39.6	13	13	86	n.d.
CNKL-7	CNKL	34	M	22100	13040	6590	16.4	31.2	35	n.d.	51	–
CNKL-8	CNKL	62	F	45400	43580	22250	11.7	16	3	3	82	n.d.
CNKL-9	CNKL	76	M	14800	13620	10510	11.2	5.8	12	9	75	+
ANKL-1	ANKL	15	F	51700	45760	41360	13.7	20.6	20	20	58	+

F, female; M, male; WBC, white blood cells; Lymph, lymphocytes; LGL, large granular lymphocytes; Hb, hemoglobin; Plt, platelets; n.d., not determined. Monoclonal or biclonal EBV genome was detected in the NK cells of CNKL-9 or ANKL-1 patient, respectively.

BAMP screening of NK cell fractions

Biotin-labeled cRNA was prepared from surface marker-matched NK fractions from the study subjects, and was hybridized with high-density oligonucleotide microarrays (Affymetrix HGU95Av2), providing the expression data for >12 000 human genes. To exclude genes that were virtually silent transcriptionally, we first selected genes whose expression received the 'Present' call from the Microarray Suite 4.0 software (Affymetrix) in at least 10% of the samples. A total of 6494 genes passed this 'selection window,' and their expression profiles in the 19 samples are shown in Figure 2a as a dendrogram, or 'gene tree,' in which genes with similar expression profiles (assessed by standard correlation) among the samples were clustered near each other. In Figure 2a, several clusters of genes that were expressed preferentially in either normal or affected NK cells (shown by arrows) were identified.



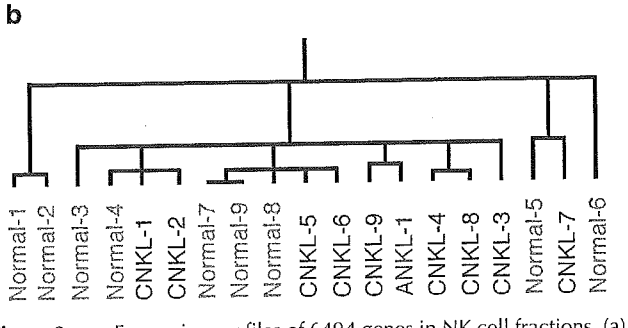


Figure 2 Expression profiles of 6494 genes in NK cell fractions. (a) Hierarchical clustering of 6494 genes on the basis of their expression profiles in $CD3^{-}CD56^{+}$ fractions derived from nine healthy volunteers (Normal), nine individuals with CNKL, and one with ANKL. Each column represents a single gene on the microarray, and each row a separate patient sample. Expression level of each gene is shown color-coded, as indicated by the scale at the bottom. Arrows indicate the positions of clusters of genes that were expressed preferentially in either normal or affected NK cells. (b) Two-way clustering analysis of the healthy individuals (Normal-1–9), CNKL patients (CNKL-1–9), and the ANKL patient (ANKL-1), based on the similarities in the expression profiles of the 6494 genes demonstrated in (a).

To statistically evaluate the similarity of the overall gene expression profiles across the 19 samples, we generated another dendrogram, a 'patient tree,' by the two-way clustering method,[22] with a separation ratio of 0.5 (Figure 2b). The samples did not clearly cluster into disease-specific branches; rather, normal and affected NK samples were mixed in several branches.

Identification of LDGL-associated genes

One of the major goals in this study was to develop expression profile-based diagnostic procedures for the NK cell disorders. For such an approach to be meaningful, an important question to be addressed would be thus to clarify whether affected NK cells share a specific gene expression profile, or 'molecular signature',[23] clearly distinct from that of normal NK cells.

Therefore, we first tried to identify genes whose expression level may efficiently differentiate normal NK cells from LDGL ones. For this purpose, we chose genes whose expression level differed significantly between the two groups of samples (Student's *t*-test, $P<0.001$). However, most of the genes thus identified had a low level of expression throughout all samples, making their usefulness as molecular markers uncertain. From these genes, therefore, we selected those whose mean expression intensity differed by $\geqslant 3.0$ arbitrary units (U) between the two groups. The resultant 15 'LDGL-associated genes' are shown in a gene tree format (Figure 3a); five of them were specific to CNKL/ANKL cells, while the remaining 10 genes were to normal NK cells.

The former group of the genes included those for transcriptional factors involved in the regulation of cell growth and/or apoptosis. B lymphoma Mo-MLV insertion region (BMI1; GenBank accession number, L13689), for example, belongs to the Polycomb type of DNA-binding proteins.[24] Intriguingly, BMI1 is expressed in hematopoietic stem cells (HSCs), and plays an indispensable role in the self-renewal process of HSCs.[25,26] Therefore, abundant expression of *BMI1* gene only in the affected NK cells may be involved in the deregulated outgrowth of the NK cells. Similarly, a zinc-finger protein ZFR (GenBank accession number, AI743507) was shown to protect embryonic cells from apoptosis and provide mitotic activity.[27]

Class prediction by our 'weighted distance method'

We next performed two-way clustering analysis, with a separation ratio of 0.5,. of the 19 specimens based on the expression levels of such 15 LDGL-associated genes. As shown in Figure 3b, the samples clustered into three major branches; one contains mainly normal NK specimens (with an addition of CNKL-1), another is composed solely of two CNKL patients (CNKL-2 and -7), and the other contains only affected NK cells. It should be noted that the ANKL sample was clustered closely with CNKL ones in the third branch.

Do the gene-expression profiles of NK cells differ between healthy individuals and those with NK cell disorders, and, if so, how different? Is such difference large enough to develop an expression profile-based diagnosis system? To address these issues, we performed correspondence analysis[21] to extract three major dimensions from the expression patterns of the 15 LDGL-associated genes. On the basis of the calculated three-dimensional coordinates for each sample, the specimens were then projected into a virtual space (Figure 3c). All normal samples were placed at a position clearly far from that of the

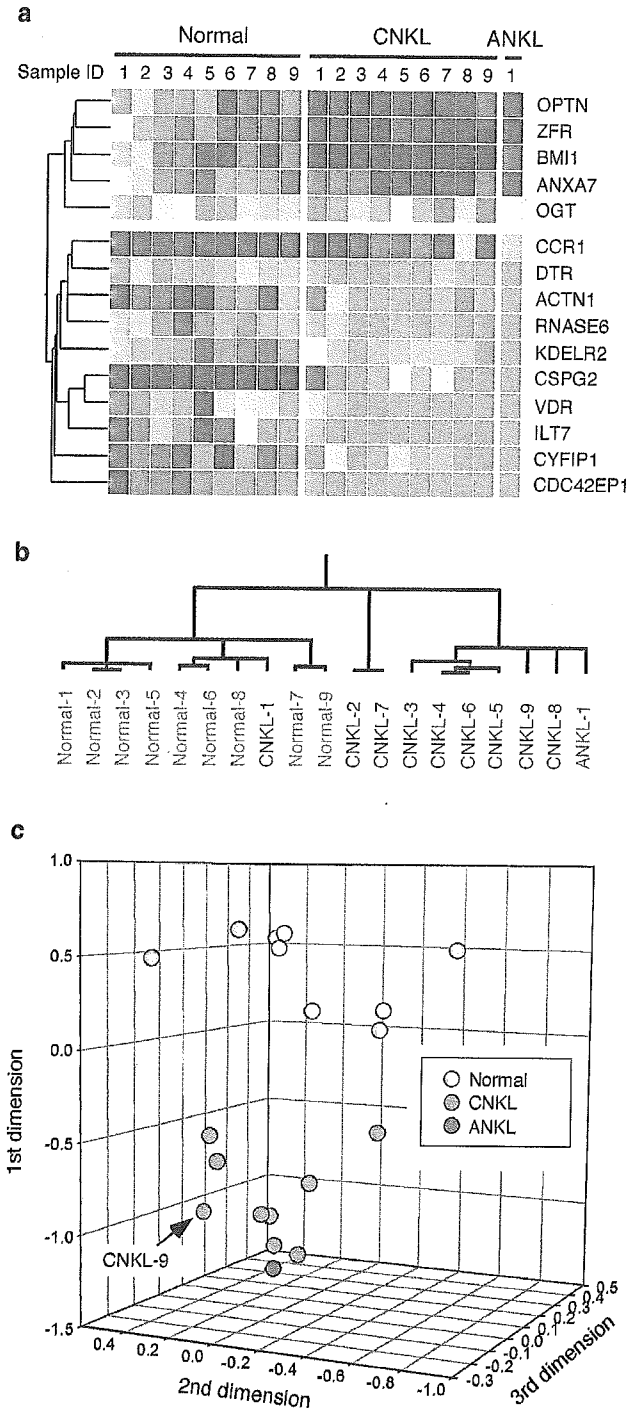


Figure 3 Identification of LDGL-associated genes. (a) Expression profiles of 15 LDGL-associated genes are shown in a dendrogram, color-coded as indicated by the scale in Figure 2a. Each row corresponds to a single gene, and each column to NK cells from healthy individuals (normal) and patients with CNKL or ANKL. 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 19 samples based on the expression levels of the LDGL-associated genes. (c) Correspondence analysis of the LDGL-associated genes identified three major dimensions in their expression profiles. Projection of the specimens into a virtual space with these three dimensions revealed that the specimens from healthy individuals (normal) were clearly separated from those from the patients with CNKL or ANKL. The position of EBV^+ CNKL-9 sample is indicated.

affected NK cells, indicating that all affected NK cells possessed a common molecular signature which was distinct from that of the normal NK cells. Again, here the two samples with clonal EBV infection (CNKL-9 and ANKL-1) were placed closely with the other CNKL specimens.

The clear separation of affected NK specimens from normal ones in Figure 3c also supported the feasibility of an expression profile-based prediction for NK cell disorders. We therefore tried class prediction (normal or LDGL) for each specimen on the basis of the coordinates calculated by the correspondence analysis. The relative 'weighted distances' of a given specimen to the normal or LDGL group (excluding the specimen for the prediction) were calculated, and the specimen was assigned a class when the relative distance to the class was <0.4. As demonstrated in Table 2, our weighted-distance method could correctly predict the class of every sample examined, making the array-based diagnostic procedure of NK cell-type LDGL into reality.

Comparison of 'Normal vs CNKL' by the weighted-distance method

Given the large difference in the clinical course between CNKL and ANKL, there may be gene-expression alterations specific to the latter condition, which characterize its aggressive clinical course. Therefore, it might be appropriate to investigate these two conditions separately. We thus focused on the comparison between normal individuals and those with CNKL, and tried to assign, by the weighted-distance method, either normal or CNKL class to every specimen among nine healthy individuals and nine patients with CNKL.

To accurately measure the prediction power of our weighted-distance method, we conducted a cross-validation trial (i.e., 'drop-one-out' format) for the diagnosis of normal or CNKL class. To predict the class of sample X, 'CNKL-associated genes' were extracted from the comparison of remaining 17 samples

Table 2 Diagnosis by the 'weighted-distance method'

Patient ID	*Clinical diagnosis*	*Distance to normal*	*Distance to LDGL*	*Prediction*
Normal-1	Normal	0.207	0.793	Normal
Normal-2	Normal	0.256	0.744	Normal
Normal-3	Normal	0.192	0.808	Normal
Normal-4	Normal	0.200	0.800	Normal
Normal-5	Normal	0.211	0.789	Normal
Normal-6	Normal	0.229	0.771	Normal
Normal-7	Normal	0.311	0.689	Normal
Normal-8	Normal	0.352	0.648	Normal
Normal-9	Normal	0.391	0.609	Normal
CNKL-1	LDGL	0.729	0.271	LDGL
CNKL-2	LDGL	0.845	0.155	LDGL
CNKL-3	LDGL	0.854	0.146	LDGL
CNKL-4	LDGL	0.877	0.123	LDGL
CNKL-5	LDGL	0.800	0.200	LDGL
CNKL-6	LDGL	0.877	0.123	LDGL
CNKL-7	LDGL	0.735	0.265	LDGL
CNKL-8	LDGL	0.861	0.139	LDGL
CNKL-9	LDGL	0.868	0.132	LDGL
ANKL-1	LDGL	0.842	0.158	LDGL

The relative weighted distance to the normal group (distance to normal) or to LDGL group (distance to LDGL) was calculated, and used to assign a class (normal or LDGL) to each sample.

according to the criteria used in Figure 3a ($P<0.001$ in Student's *t*-test, and |effect size|>3). The number of such CNKL-associated genes ranged from 4 to 10. Correspondence analysis was carried out for the expression profiles of the CNKL-associated genes, and was used to calculate the relative weighted distance of the 'dropped' sample X to either normal or CNKL class. As shown in Table 3, with a *T*-value of 0.4, a clinically matched prediction was obtained for 14 (77.8%) out of 18 cases, while one case (CNKL-2) was unpredictable and three cases (normal-6, normal-8 and CNKL-1) received a prediction incompatible with the clinical diagnosis. Therefore, even in a cross-validation assay, the weighted-distance method could achieve a high accuracy.

For comparison, we also conducted a cross-validation trial of class prediction by using a known prediction algorithm, the '*k*-nearest neighbor method' (http://www.silicongenetics.com/Support/GeneSpring/GSnotes/class_prediction.pdf). Among the 18 samples tested, only 10 samples (55.6%) received correct prediction, indicating the superiority of our weighted distance method.

Since CNKL-9 patient had NK cells with EBV in a clonal episomal form, and had progressed into an ANKL phase in a relatively short period, we questioned if this patient had an atypical molecular signature for CNKL. To visualize the similarity of transcriptome of CNKL-9 sample with that of the other CNKL ones, the result of the cross-validation trial for CNKL-9 is demonstrated as a virtual-space format in Figure 4a. Correspondence analysis of nine genes that most efficiently differentiated normal-1–9 from CNKL-1–8 has identified three major dimensions in their expression pattern, and projection of the CNKL-9 patient together with the other samples in a 3D space indicated that CNKL-9 had an expression profile highly similar to that of the other CNKL subjects at least in the space of these nine highly informative genes.

To confirm the similarity in the gene-expression profile of EBV$^+$ ANKL cells to the CNKL ones, we next carried out correspondence analysis for the ANKL-1 patient. Statistical comparison of transcriptome between Normal-1–9 and CNKL-1–9 subjects identified a total of seven genes, which contrasted the expression profile of normal NK cells from that of CNKL NK cells. As shown in Figure 4b, projection of the ANKL-1 patient into a 3D space constructed from the data of such seven genes demonstrated that the EBV$^+$ ANKL-1 sample was plotted at a neighbor position to those of the CNKL samples. In accordance with the 3D view, the weighted-distance method also concluded that the ANKL-1 sample belonged to the same class with the CNKL ones (data not shown). These analyses unexpectedly suggested that the gene expression profile characteristic to CNKL NK cells is also shared with EBV$^+$ NK cells. It should be noted, however, that additional genetic changes specific to EBV infection may exist, and account for the fulminant clinical character of EBV$^+$ LDGL.

Table 3 Cross-validation of disease prediction

Patient ID	*Clinical diagnosis*	*Prediction*
Normal-1	Normal	Normal
Normal-2	Normal	Normal
Normal-3	Normal	Normal
Normal-4	Normal	Normal
Normal-5	Normal	Normal
Normal-6	Normal	CNKL
Normal-7	Normal	Normal
Normal-8	Normal	CNKL
Normal-9	Normal	Normal
CNKL-1	CNKL	Normal
CNKL-2	CNKL	Unpredictable
CNKL-3	CNKL	CNKL
CNKL-4	CNKL	CNKL
CNKL-5	CNKL	CNKL
CNKL-6	CNKL	CNKL
CNKL-7	CNKL	CNKL
CNKL-8	CNKL	CNKL
CNKL-9	CNKL	CNKL

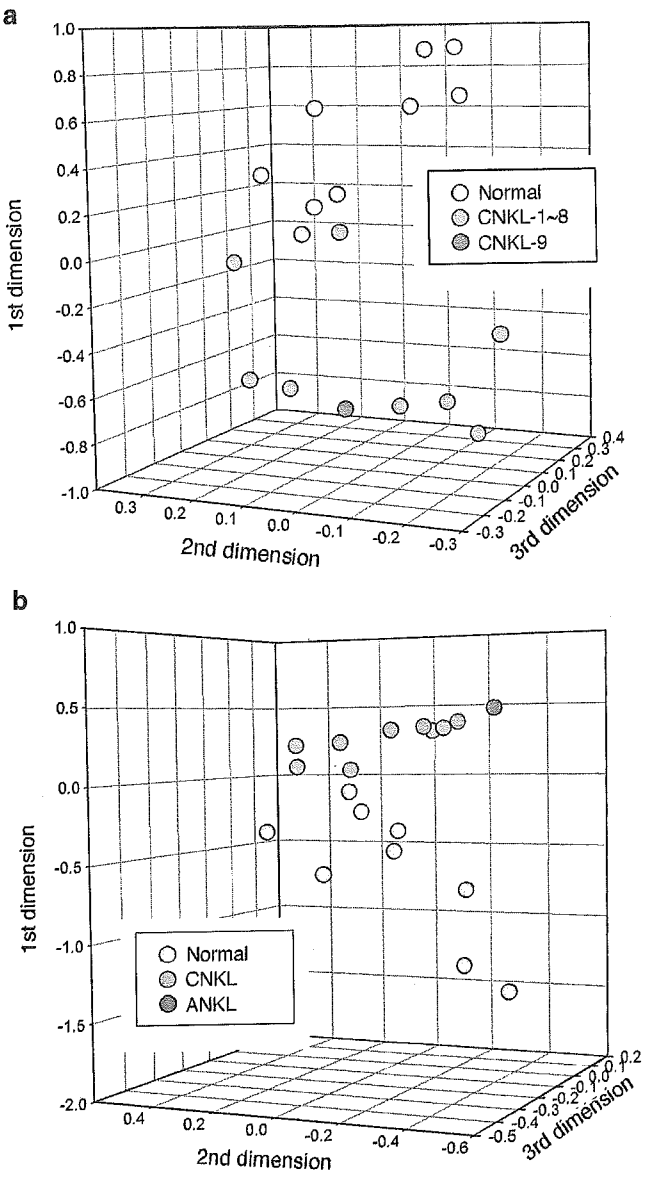


Figure 4 Investigation of the EBV$^+$ samples. (a) We could isolate nine genes, expression of which differentiated normal NK cells (normal) from indolent CNKL ones (CNKL-1–8). The EBV$^+$ CNKL-9 was projected into a virtual space together with the other normal and CNKL specimens, based on the coordinates calculated by the correspondence analysis of such nine genes. (b) A total of seven genes were identified to be differentially expressed between normal NK cells (normal) and CNKL cells. The ANKL-1 sample was projected into the virtual space as in (a).

Isolation of single-gene markers for LDGL diagnosis

The gene set identified in Figure 3a may potentially be the candidate genes to construct custom-made DNA microarrays for disease diagnosis of NK cell disorders. Since availability of DNA microarray systems is still restricted in current hospitals, however, it would be valuable if a high expression of single gene or its product can be used as a reliable marker for such purposes. For instance, it would be highly useful if the serum level of a protein can help to diagnose NK cell disorders. Given the presence of false data even with DNA microarray, it is unlikely that an expression of any single gene can correctly diagnose all samples. Therefore, here we have tried to isolate genes whose high expression may be 'sufficient' to predict the presence of NK cell-type LDGL, but the absence of its expression may not necessarily mean that the NK cells are normal. Such type of predictor genes should be strictly inactivated in all normal NK cells, but become activated in, at least, a part of the NK cells in the LDGL group.

To screen such type of predictors, first, the mean expression value of each gene was calculated for the normal or LDGL group. Then, with the use of GeneSpring software, we searched for genes whose expression profiles were statistically similar, with a minimum correlation of 0.95, to that of a hypothetical 'LDGL-specific gene' that exhibits a mean expression level of 0.0 U in the normal group and 100.0 U in the LDGL group. From such 652 genes identified, we then applied another criteria that gene-expression value should be (i) $\geqslant$60.0 U in, at least, one of the LDGL samples, and (ii) $<$25.0 U in all normal samples. A total of six genes were finally identified to be 'LDGL-specific' (Figure 5a). Here we have tried to extract LDGL-specific genes with minimum false-positive results, while allowing false-negative ones. Therefore, we should confidently tell that the given NK cells are of LDGL if one of the 'LDGL-specific genes' is highly expressed in the specimens.

Conversely, we also tried to identify 'normal-specific genes' through the same approach. Firstly, a total of 1424 genes were identified to be statistically similar to a hypothesized 'normal-specific gene' that has a mean expression value of 100.0 U in the normal group, but of 0.0 U in the LDGL group. Among these genes, those whose expression was kept below 25.0 U throughout the samples in the LDGL group, but became activated at $\geqslant$60.0 U in, at least, one sample in the normal group were selected. We could thus extract a set of 22 genes, expression of which was specific to normal NK cells (Figure 5b).

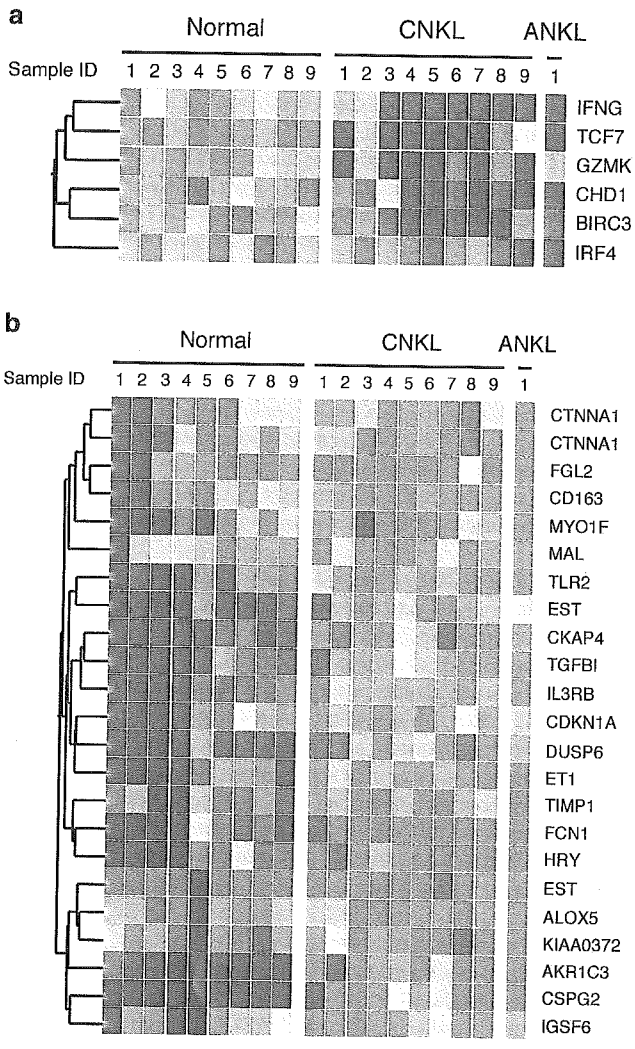


Figure 5 Identification of single-gene markers for LDGL. (a) Dendrogram showing the expression profiles of six genes whose expression intensity was kept suppressed in normal NK cells, but became activated in a part of affected NK cells. Each row represents a single gene, and each column a separate patient sample. Expression level of the genes is shown colored according to the scale in Figure 2a. (b) Expression profiles of 22 'normal-specific genes' are demonstrated as in (a). Two different oligonucleotide sets for the alphaE-catenin (CTNNA1) gene are present on an HGU95Av2 array.

Confirmation of overproduction of IFNG

NK cells become activated and produce IFNG in response to the stimulation with IL-2,[28] IL-12,[29] and IL-15.[30] Under physiological circumstances, however, activated NK cells eventually undergo apoptotic changes to prevent overactivation of the immune system. Interestingly, IFNG itself provides a survival signal onto NK cells, and, therefore, sustained incubation with IFNG of NK cells protects efficiently them from cell death.[31] It was thus provocative to find *IFNG* in our LDGL-specific genes (Figure 5a), indicating a potential role of IFNG in the outgrowth mechanism of NK cells in the LDGL condition.

Here we have confirmed the disease-specific expression of *IFNG* gene by a quantitative 'real-time' RT-PCR assay. As shown in Figure 6a, abundant expression of *IFNG* mRNA was only detected in the purified NK cell fraction of LDGL patients, but not of the normal controls. Furthermore, a high concentration of IFNG protein was also observed in the sera of CNKL/ANKL patients (Figure 6b), proving that overexpression of *IFNG* mRNA in NK cell disorders leads to the systemic elevation of IFNG protein level. Interestingly, overexpression of IFNG protein was also noticed in the patients with IMN. High expression of IFNG in IMN individuals may result from the infection of EBV associated with IMN, or may indicate the activated status of T or NK cells in the condition of IMN.

Discussion

In this manuscript, we tried to clarify whether gene-expression profiling can help to differentiate NK cells of LDGL individuals from those of healthy ones. Toward this goal, we first purified NK cell fractions from study subjects, which are characterized by the absence of cell surface CD3 molecule and the presence of CD56 antigen. Analysis with these isolated NK cells should be