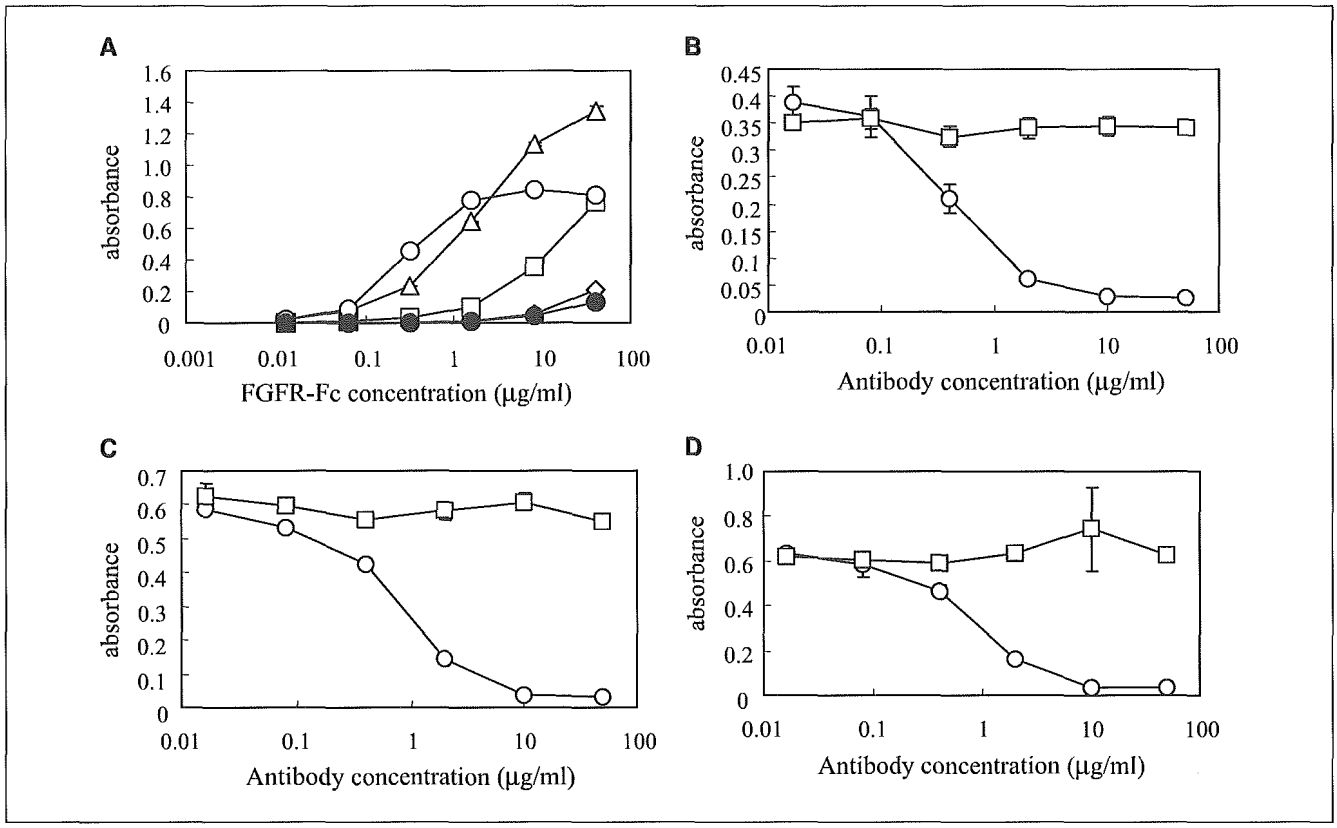


Fig. 3. Binding activity of FGFRs to FGF8b and their inhibition by KM1334. *A*, FGF8b was coated to a ELISA plate at concentration of 5 μg/mL. Then, FGFR extracellular domains fused to human Fc region [FGFR1IIIc-Fc (◇), FGFR2IIIc (□), FGFR3IIIc (△), and FGFR4 (○)] were reacted to the FGF8b at various concentrations. IL5R-human Fc fusion protein was used as a negative control (●). *B-D*, effect of KM1334 (○) and KM511 (□), negative control, on the binding of FGF8b to FGFR-Fc fusion proteins [FGFR2IIIc-Fc (*B*), FGFR3IIIc-Fc (*C*), and FGFR4-Fc (*D*)]. Points, mean for $A_{415\text{-}490}$; bars, SD.

To understand the mechanism by which tumors treated by 400 μg KM1334 regressed rapidly *in vivo*, histologic analysis was done on tumor sections from SC-3 tumor model, and tumor sections were stained for markers of cell proliferation and apoptosis. Figure 6A shows the effect of KM1334 on cumulative BrdUrd incorporation at three time points (7, 24, and 72 hours). Twenty-four hours after treatment with KM1334, there was a remarkable reduction in BrdUrd incorporation persisting until 72 hours, which is suggestive of arrest in G_0-G_1. We next examined the effect of KM1334 on apoptosis over the same time course via TUNEL labeling (Fig. 6B). The number of apoptotic cells increased markedly between 7 and 24 hours and decreased to the level of control by 72 hours. These results suggested that inhibition of cell proliferation and induction of apoptosis were responsible for the dramatic regression of tumor size *in vivo*.

Discussion

To our knowledge, this is the first report that describes mAb against FGF8 for therapeutic use. We showed that an anti-FGF8b murine mAb, KM1334, suppressed SC-3 tumor growth dose-dependently and the maximum dose injection of the antibody caused rapid regression of SC-3 tumors established in nude mice accompanied by inhibition of cell proliferation and induction of apoptosis. KM1334 showed very high specificity to FGF8b and FGF8f out of FGF family members examined and potent blocking activity of FGF8b binding to its three receptors (FGFR2IIIc, FGFR3IIIc, and FGFR4) and FGF8b-induced downstream signaling. These results provide the rationale for further development of KM1334 as a candidate anticancer therapeutic agent against FGF8b-dependent tumors.

Contribution of FGFs, including FGF2, FGF5, FGF8, etc., as autocrine and/or paracrine growth factors in tumorigenesis have been widely studied (31, 42); however, there are only limited reports that showed the potential of anti-FGF ligand neutralizing antibodies as candidates of therapeutic agents. Although several groups reported that anti-FGF2 antibodies showed antitumor activities *in vivo* (43, 44), they were not so potent as observed in this study. We show for the first time that FGF ligands are very potent targets for antibody-based cancer therapy.

Many studies reported up-regulation of FGF8 in various clinical cancers, including prostate and breast cancers, and the strongest mitogenic isoform FGF8b among human FGF8 variants (a, b, e, and f) is considered to be the most important for the progression of those diseases (22, 23, 25, 27). FGF8f has the second strongest mitogenic activity in human FGF8 isoforms (17, 31) and coexpression of FGF8f with FGF8b resulted in worse prognosis than single expression of FGF8b in esophageal carcinoma (45). Therefore, inhibition of both FGF8b and FGF8f activity might be desirable for the therapy

of FGF8-associated cancers. It was already shown that KM1334 specifically recognizes FGF8b among murine FGF8 variants (a-c); however, it was unclear which human FGF8 variants KM1334 recognizes. Among synthetic peptides derived from four human

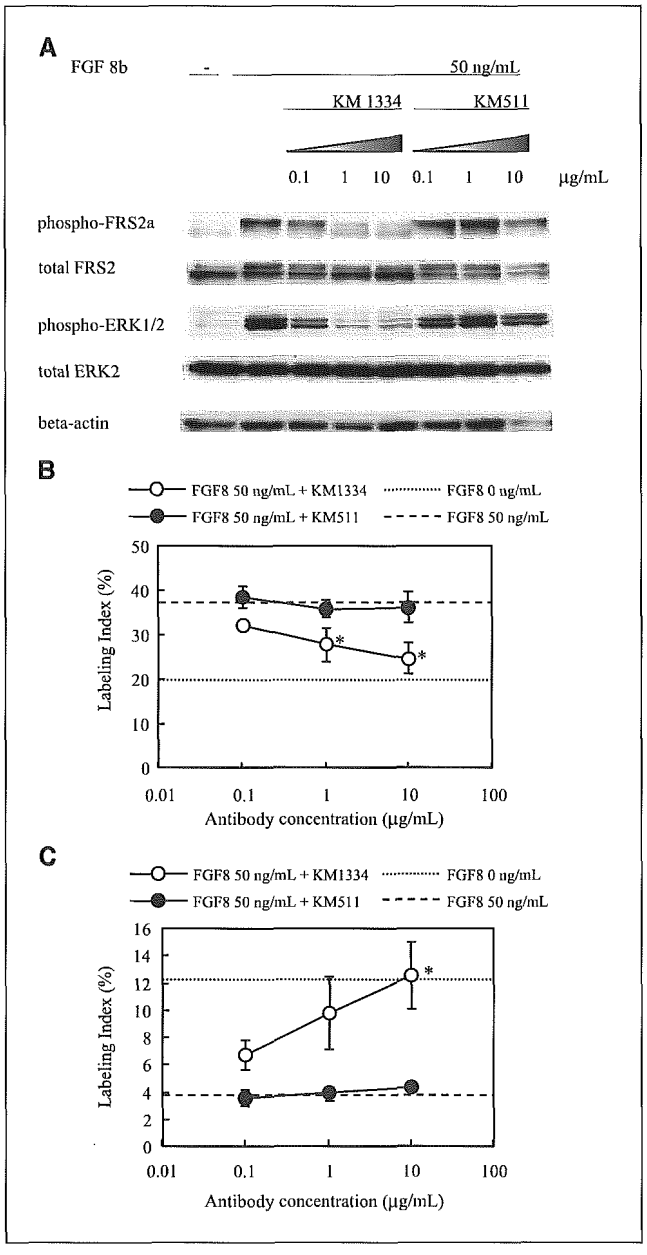


Fig. 4. *In vitro* effect of KM1334 on SC-3 cells stimulated by FGF8b. *A*, Western blot analysis. SC-3 cells were stimulated by 50 ng/mL FGF8b with KM1334 or KM511 (isotype control). After 15 minutes of incubation at 37°C, cells were collected and cell lysates were prepared. Western blot analyses were done using the indicated phosphospecific antibodies to FRS2α (Tyr^{196}) or ERK1/2 (Thy^{202}/Tyr^{204}), and the blots were reprobed with antibodies to total FRS2 or total ERK2. A representative β-actin reprobed blot is shown as a loading control. *B* and *C*, immunocytochemical analysis of SC-3 cells. SC-3 cells were stimulated by 50 ng/mL FGF8b with KM1334 or KM511. After 24 hours of incubation, BrdUrd was added and incubated for additional 1 hour. Uptake of BrdUrd was examined by anti-BrdUrd antibody with an immunocytochemical method (*B*). Results of TUNEL staining were shown in (*C*). Labeling index shows the number of 3,3′-diaminobenzidine-positive cells divided by the total number of nuclear pixels (hematoxylin-positive cells + 3,3′-diaminobenzidine-positive cells) multiplied by 100. Points, mean for three independent areas of 370 × 500 μm; bars, SD. Statistics were done by two-tailed Student's *t* test. *, $P < 0.05$, relative to control (*B* and *C*).

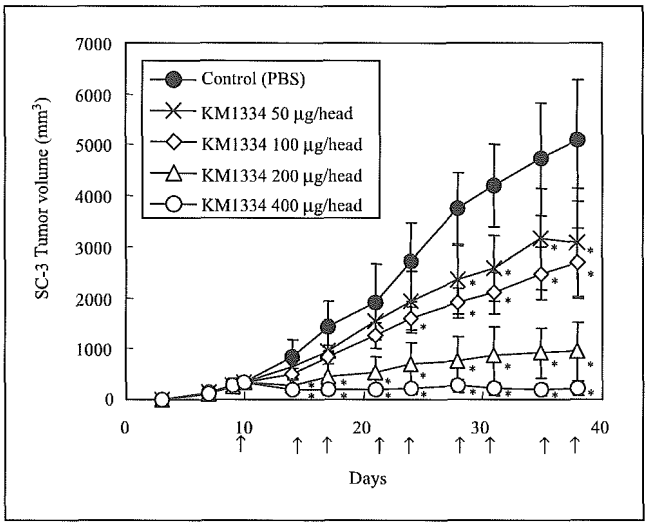


Fig. 5. Antitumor activity of KM1334 against SC-3 tumors established on nude mice. SC-3 cells were transplanted s.c. into adult male nude mice. SC-3 tumors were allowed to grow ~ 300 mm^3; then, KM1334 [50 μg/head (×), 100 μg/head (◇), 200 μg/head (△), or 400 μg/head (○)] and PBS (●), a vehicle control, were given into the mice twice weekly (total of nine times). Tumor volume was measured until day 38. Arrows below the transverse axis, days for antibody treatment. Points, mean for tumor volume ($n = 5$ each group); bars, SD. *, $P < 0.01$, relative to control, two-tailed Student's *t* test.

FGF8 variants, FGF8b-1 (derived from FGF8b and used for the immunogen of KM1334) and FGF8f-1 (derived from FGF8f) were proven to inhibit the binding of KM1334 to its antigen, indicating that KM1334 recognize both FGF8b and FGF8f isoforms (Fig. 1C). Human FGF17 and FGF18 have been isolated recently and form a subfamily with FGF8 in FGFs. Their amino acid sequences corresponding to FGF8b-1 have significant identity to FGF8b (FGF17, 54.1%; FGF18, 20.0%; Fig. 2A); however, KM1334 did not bind to FGF17b and FGF18 (Fig. 2B). These results indicate that KM1334 possesses high specificity enough to target FGF8b and FGF8f selectively.

Although colocalization of FGF8 and its receptors was identified in various neoplastic tissues (27, 28, 45), it was still unknown which type of FGFRs are activated by FGF8. Therefore, an anti-FGF8 neutralizing mAb, which has the ability to inhibit FGF8-induced activation via all possible FGFRs, is desired. Relative mitogenic activities of FGF8 isoforms to each FGFR have already been described (17, 33); however, direct comparison of mitogenic activity or binding activity of FGF8b to FGFRs has never been reported. In this study, we could establish for the first time the ELISA system to detect direct binding of FGF8b to FGFR-Fc fusion proteins. In this system, FGFR4-Fc showed the most sensitive detection limit; however, FGFR3IIIc-Fc gave the highest value of the absorbance indicative of binding activity (Fig. 3A). The affinity of FGFR2IIIc-Fc to FGF8b was lower than that of FGFR3IIIc-Fc and FGFR4-Fc. These results were consistent with previous reports (17, 33). KM1334 inhibited the binding of FGF8b to all FGF8b receptors examined (FGFR2IIIc-Fc, FGFR3IIIc-Fc, and FGFR4-Fc) completely in the ELISA (Fig. 3B-D), which indicates that KM1334 has an ideal character as an inhibitor of FGF8b. It has already been shown that FGF8b-dependent growth of SC-3 cells is mediated by FGFR1IIIc (14, 46). However, in this study, the specific binding of FGFR1IIIc-Fc to FGF8b was not detected in the ELISA. Several

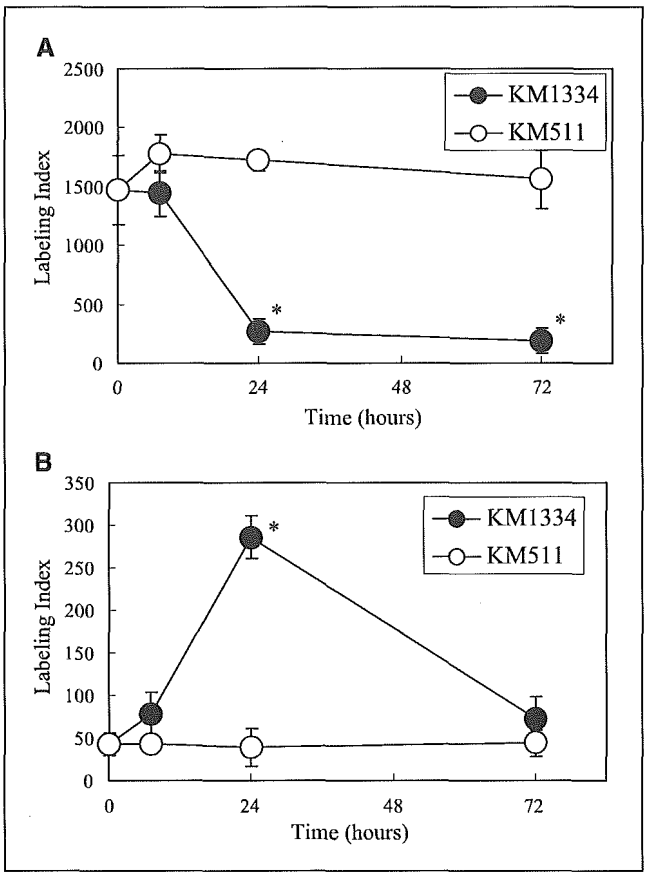


Fig. 6. Histologic analysis of SC-3 tumors treated with KM1334. *A*, uptake of BrdUrd. Mice carrying SC-3 tumors were treated with KM1334 (●) or KM511 (○), negative control. SC-3 tumors were collected before and at 7, 24, and 72 hours after antibody injection. Uptake of the BrdUrd was detected by anti-BrdUrd antibody. *B*, TUNEL staining. The same sample used for the BrdUrd uptake experiment was used for TUNEL staining. Labeling index indicates the number of 3,3′-diaminobenzidine-positive cells per mm^2. Points, mean for three independent areas; bars, SD. *, $P < 0.05$, relative to control, two-tailed Student's *t* test (*A* and *B*).

explanations of this discrepancy were possible. One explanation is that mutations in FGFR1IIIc (47) facilitate the binding between FGFR1IIIc and FGF8b in SC-3 cells. Another explanation, which is considered as more likely than the preceding one, is that concentration of FGF8b has significant effects on the detection of FGFR1IIIc-FGF8b binding. In previous reports, FGF8b could activate FGFR2IIIc, FGFR3IIIc, and FGFR4 at nanomolar concentrations (17, 33). However, FGFR1IIIc needs micromolar range of FGF8b concentration for its activation (17). Therefore, the binding between FGFR1IIIc and FGF8b in the ELISA might be enhanced by increasing the amount of FGF8b coated on the ELISA plate.

FRS2α is essential for the FGF-induced MAPK response (48) and ERK is a member of the MAPK implicated in the regulation of cell proliferation and differentiation. FGF8b treatment with SC-3 cells induced phosphorylation of both FRS2α and ERK1/2 (Fig. 4A), indicating that the FGF8b-induced proliferation of SC-3 cells is mediated by FGFR(s)-FRS2α-ERK1/2 signal transduction pathway. Together with the fact that KM1334 blocked this FGF8b-induced phosphorylation of FRS2α and ERK1/2 in SC-3 cells (Fig. 4A), FGF8b-neutralizing activity of KM1334 on SC-3 cell growth could be explained by the inhibition of FGF8b binding to its receptors and subsequent blockade of FGFR-mediated signal transduction.

Regression of SC-3 tumors by KM1334 treatment was explained by inhibition of cell proliferation and induction of apoptosis (Fig. 6A and B), which were well consistent with *in vitro* results shown in Fig. 4B and C. Together with the fact that that the antitumor effect of KM1334 was equal to castration (data not shown), it was suggested that FGF8b is critical to the androgen-dependent growth of SC-3 cells *in vivo* as well as *in vitro*. These results indicate that KM1334 might become a good therapeutic reagent against FGF8b-dependent tumors. In clinical prostate carcinoma in which androgen ablation therapy was effective, increase of the apoptotic index and decrease of mitotic index of tumors were detected (49, 50). Although the period of treatment was different, the same phenomenon was observed in the regressing SC-3 tumors treated with KM1334 (Fig. 6). Because the expression of FGF8 was identified in hormone-refractory prostate cancer as well as in hormone-responsive cancer (15), KM1334 may also be effective against hormone-refractory prostate cancer. As additional supportive evidence of this possibility, we have recently shown that overexpression of FGF8b in human prostate cancer LNCaP cells gave growth advantage to the cells *in vitro* and *in vivo*, and KM1334 could inhibit tumor growth in xenograft models of FGF8b-transfected LNCaP cells under androgen-dependent and androgen-independent conditions.[5] Further analysis, including additional *in vivo* xenograft studies and future clinical studies, might be needed to fully address the importance and contribution of FGF8b in human cancers.

Because amino acid sequences of mouse and human FGF8b are 100% identical (16), KM1334 could block endogenous FGF8b in mice. Administration of KM1334 caused no apparent toxicity, like weight loss or death to mice (data not shown), suggesting that targeting FGF8b in adult human might not cause severe side effects, although further studies are needed to fully analyze the side effects.

[5] K. Maruyama-Takahashi et al. An anti-fibroblast growth factor (FGF) 8 monoclonal antibody, KM1334, inhibits the growth of FGF8b-overexpressing LNCaP xenografts under androgen-dependent and -independent conditions, submitted for publication.

References

1. Kohler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. Nature (London) 1975;256:495–7.
2. Brekke OH, Sandlie I. Therapeutic antibodies for human diseases at the dawn of the twenty-first century. Nat Rev Drug Discov 2003;2:52–62.
3. Hurwitz H, Fehrenbacher L, Novotny W, et al. Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. N Engl J Med 2004;350:2335–42.
4. Szebenyi G, Fallon JF. Fibroblast growth factors as multifunctional signaling factors. Int Rev Cytol 1999;185:45–106.
5. Draper BW, Stock DW, Kimmel CB. *Zebrafish fgf24* functions with *fgf8* to promote posterior mesodermal development. Development 2003;130:4639–54.
6. Tanaka A, Miyamoto K, Minamino N, et al. Cloning and characterization of an androgen-induced growth factor essential for the androgen-dependent growth of mouse mammary carcinoma cells. Proc Natl Acad Sci U S A 1992;89:8928–32.

7. Ohuchi H, Yoshioka H, Tanaka A, Kawakami Y, Nohno T, Noji S. Involvement of androgen-induced growth factor (FGF-8) gene in mouse embryogenesis and morphogenesis. Biochem Biophys Res Commun 1994;204:882–8.
8. Mahmood R, Bresnick J, Hornbruch A, et al. A role for FGF-8 in the initiation and maintenance of vertebrate limb bud outgrowth. Curr Biol 1995;5:797–806.
9. Crossley PH, Martin GR. The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. Development 1995;121:439–51.
10. Crossley PH, Minowada G, MacArthur CA, Martin GR. Roles for FGF8 in the induction, initiation, and maintenance of chick limb development. Cell 1996; 84:127–36.
11. Crossley PH, Martinez S, Martin GR. Midbrain development induced by fgf8 in the chick embryo. Nature 1996;380:66–8.
12. Heikinheimo M, Lawshe A, Shackleford GM, Wilson DB, MacArthur CA. Fgf-8 expression in the post-gastrulation mouse suggests roles in the development of the face, limbs and central nervous system. Mech Dev 1994;48:129–38.
13. Lewandoski M, Meyers EN, Martin GR. Analysis of Fgf8 gene function in vertebrate development. Cold Spring Harb Symp Quant Biol 1997;62:159–68.
14. Kouhara H, Koga M, Kasayama S, Tanaka A, Kishimoto T, Sato B. Transforming activity of a newly cloned androgen-induced growth factor. Oncogene 1994;9:455–62.
15. Dorkin TJ, Robinson MC, Marsh C, Bjartell A, Neal DE, Leung HY. FGF8 over-expression in prostate cancer is associated with decreased patient survival and persists in androgen independent disease. Oncogene 1999;18:2755–61.
16. Gemel J, Gorry M, Ehrlich GD, MacArthur CA. Structure and sequence of human FGF8. Genomics 1996;35:253–7.
17. MacArthur CA, Lawshe A, Xu J, et al. FGF-8 isoforms activate receptor splice forms that are expressed in mesenchymal regions of mouse development. Development 1995;121:3603–13.
18. MacArthur CA, Lawshe A, Shankar DB, Heikinheimo M, Shackleford GM. FGF-8 isoforms differ in NIH3T3 cell transforming potential. Cell Growth Differ 1995;6:817–25.
19. Ghosh AK, Shankar DB, Shackleford GM, et al. Molecular cloning and characterization of human FGF8 alternative messenger RNA forms. Cell Growth Differ 1996;7:1425–34.
20. Daphna-Iken D, Shankar DB, Lawshe A, Ornitz DM, Shackleford GM, MacArthur CA. MMTV-Fgf8 transgenic mice develop mammary and salivary gland neoplasia and ovarian stromal hyperplasia. Oncogene 1998;17:2711–7.
21. Song Z, Wu X, Powell WC, et al. Fibroblast growth factor 8 isoform B overexpression in prostate epithelium: a new mouse model for prostatic intraepithelial neoplasia. Cancer Res 2002;62:5096–105.
22. Song Z, Powell WC, Kasahara N, van Bokhoven A, Miller GJ, Roy-Burman P. The effect of fibroblast growth factor 8, isoform b, on the biology of prostate carcinoma cells and their interaction with stromal cells. Cancer Res 2000;60:6730–6.
23. Ruohola JK, Viitanen TP, Valve EM, et al. Enhanced invasion and tumor growth of fibroblast growth factor 8b-overexpressing MCF-7 human breast cancer cells. Cancer Res 2001;61:4229–37.
24. Tanaka A, Furuya A, Yamasaki M, et al. High frequency of fibroblast growth factor (FGF) 8 expression in clinical prostate cancers and breast tissues, immunohistochemically demonstrated by a newly established neutralizing antibody against FGF 8. Cancer Res 1998; 58:2053–6.
25. Gnanapragasam VJ, Robinson MC, Marsh C, Robson CN, Hamdy FC, Leung HY. FGF8 isoform b expression in human prostate cancer. Br J Cancer 2003;88:1432–8.
26. Tanaka A, Kamiakito T, Takayashiki N, Sakurai S, Saito K. Fibroblast growth factor 8 expression in breast carcinoma: associations with androgen receptor and prostate-specific antigen expressions. Virchows Arch 2002;441:380–4.
27. Marsh SK, Bansal GS, Zammit C, et al. Increased expression of fibroblast growth factor 8 in human breast cancer. Oncogene 1999;18:1053–60.
28. Valve E, Martikainen P, Seppanen J, et al. Expression of fibroblast growth factor (FGF)-8 isoforms and FGF receptors in human ovarian tumors. Int J Cancer 2000; 88:718–25.
29. McKeehan WL, Wang F, Kan M. The heparan sulfate-fibroblast growth factor family: diversity of structure and function. Prog Nucleic Acid Res Mol Biol 1998;59:135–76.
30. Ornitz DM, Xu J, Colvin JS, et al. Receptor specificity of the fibroblast growth factor family. J Biol Chem 1996;271:15292–7.
31. Powers CJ, McLeskey SW, Wellstein A. Fibroblast growth factors, their receptors and signaling. Endocr Relat Cancer 2000;7:165–97.
32. Yan G, Fukabori Y, McBride G, Nikolaropolous S, McKeehan WL. Exon switching and activation of stromal and embryonic fibroblast growth factor (FGF)-FGF receptor genes in prostate epithelial cells accompany stromal independence and malignancy. Mol Cell Biol 1993;13:4513.
33. Blunt AG, Lawshe A, Cunningham ML, Seto ML, Ornitz DM, MacArthur CA. Overlapping expression and redundant activation of mesenchymal fibroblast growth factor (FGF) receptors by alternatively spliced FGF-8 ligands. J Biol Chem 1997;272:3733–8.
34. Ullrich A, Schlessinger J. Signal transduction by receptors with tyrosine kinase activity. Cell 1990;61: 203–12.
35. Denhardt DT. Signal-transducing protein phosphorylation cascades mediated by Ras/Rho proteins in the mammalian cell: the potential for multiplex signalling. Biochem J 1996;318:729–47.
36. Klint P, Claesson-Welsh L. Signal transduction by fibroblast growth factor receptors. Front Biosci 1999;4: D165–77.
37. Yoshida H, Shitara S. Generation and characterization of monoclonal antibodies to recombinant human granulocyte-colony stimulating factor (G-CSF) and its muteins. Agric Biol Chem 1989;53:1095–101.
38. Yamanishi H, Nonomura N, Tanaka A, et al. Proliferation of Shionogi carcinoma 115 cells by glucocorticoid-induced autocrine heparin-binding growth factor(s) in serum-free medium. Cancer Res 1991;51:3006–10.
39. Soga S, Neckers LM, Schulte TW, et al. KF25706, a novel oxime derivative of radicicol, exhibits *in vivo* antitumor activity via selective depletion of Hsp90 binding signaling molecules. Cancer Res 1999;59: 2931–8.
40. Takaba K, Saeki K, Suzuki K, Wanibuchi H, Fukushima S. Significant overexpression of metallothionein and cyclin D1 and apoptosis in the early process of rat urinary bladder carcinogenesis induced by treatment with *N*-butyl-*N*-(4-hydroxybutyl) nitrosamine or sodium L-ascorbate. Carcinogenesis 2000; 21:691–700.
41. Geran RI, Greenberg NH, MacDonald MM, Schumacher AM, Abbott BJ. Protocols for screening chemical agents and natural products against animal tumors and other biological systems [part 3]. Cancer Chemother Rep 1972;3:1–103.
42. Moroni E, Dell'Era P, Rusnati M, Presta M. Fibroblast growth factors and their receptors in hematopoiesis and hematological tumors. J Hematother Stem Cell Res 2002;11:19–32.
43. Hori A, Sasada R, Matsutani E, et al. Suppression of solid tumor growth by immunoneutralizing monoclonal antibody against human basic fibroblast growth factor. Cancer Res 1991;51:6180–4.
44. Gross JL, Herblin WF, Dusak BA, et al. Effects of modulation of basic fibroblast growth factor on tumor growth *in vivo*. J Natl Cancer Inst 1993;85: 121–31.
45. Tanaka S, Ueo H, Mafune K, Mori M, Wands JR, Sugimachi K. A novel isoform of human fibroblast growth factor 8 is induced by androgens and associated with progression of esophageal carcinoma. Dig Dis Sci 2001;46:1016–21.
46. Sato B, Kouhara H, Koga M, et al. Androgen-induced growth factor and its receptor: demonstration of the androgen-induced autocrine loop in mouse mammary carcinoma cells. J Steroid Biochem Mol Biol 1993;47:91–8.
47. Kouhara H, Kasayama S, Saito H, Matsumoto K, Sato B. Expression cDNA cloning of fibroblast growth factor (FGF) receptor in mouse breast cancer cells: a variant form in FGF-responsive transformed cells. Biochem Biophys Res Commun 1991; 15:31–7.
48. Hadari YR, Gotoh N, Kouhara H, Lax I, Schlessinger J. Critical role for the docking-protein FRS2α in FGF receptor-mediated signal transduction pathways. Proc Natl Acad Sci U S A 2001;98:8578–83.
49. Szende B, Romics I, Torda I, Bely M, Szegedi Z, Lovasz S. Apoptosis, mitosis, p53, bcl(2), Ki-67 and clinical outcome in prostate carcinoma treated by androgen ablation. Urol Int 1999;63:115–9.
50. Szende B, Romics I, Minik K, et al. Repeated biopsies in evaluation of therapeutic effects in prostate carcinoma. Prostate 2001;49:93–100.

〈抄録〉第25回 日本臨床薬理学会年会 2004年9月17～18日 静岡
シンポジウム2（安全性分野）：トキシコゲノミクス―現状と臨床薬理学への応用―

4．日本人組織を用いたトキシコゲノミクス研究

大島康雄* 藤村昭夫*

我々の研究室では日本人由来の組織を用いたトキシコゲノミクス研究を行っている。現状では日本人組織を商業ベースで合法的に入手することはできないために、我々は自治医科大学附属病院で手術を受ける患者さんにご協力いただき、手術時に病変部位と同時にやむを得ず切除される正常組織を研究に利用することとした。動物実験で得られた情報だけではヒトへの外挿が必ずしも十分ではなく、ヒトで最終的に確認できればより好ましい。一方、臨床検体の入手には困難な点が多く、研究計画の自由度は低い。動物実験・細胞株を用いた研究と我々の様にヒト組織を用いた研究は相補的な位置づけとされ、それぞれの役割を担うことが期待される。

1．倫理評価ワーキンググループ：
治療を目的として我々の病院を訪れる患者さんから、研究のための組織を提供していただくにあたり、我々は倫理評価ワーキンググループを立ち上げ、研究計画の審査・インフォームドコンセント取得の手順・検体採取後の病理組織の評価・匿名化の手順・関係書類の保管などにつき詳細な検討を行った。倫理評価ワーキンググループの中には、宗教家・法律家などの学外委員も含まれている。彼らとの討論の中で、医療関係者以外の第三者から誤解を受けやすい点が一つ浮かび上がった。今回我々が研究に用いる検体は、従来の手術方法で切除されてしまう非病変部組織であり、通常廃棄処分される組織部分を研究へ利用させていただく計画であった。しかし、非医療関係者は、病変の治療のためには必要もないのに研究目的のためだけに正常組織を切除するものと誤解されるようであった。このような誤解を受けやすい部分を今後も啓発することによって、日本人組織を研究・開発に利用しやすい社会環境を形成してゆく必要がある。

2．臨床検体取得の現場：
患者さんよりインフォームドコンセントをいただき、附属病院の手術室から臨床検体を取得し、プライマリーカルチャーを作成することができた。我々が試みた範囲ではディスパーゼとトリプシンを併用する方法が安定した良好な結果をもたらすものと思われた[1]。得られた細胞は、上皮性の細胞として矛盾のない形態を示し、腎臓皮質由来の細胞の多くはGlut-2抗原及びγGTP活性を示し、尿細管由来であることが示された。また、肝臓由来の細胞はアルブミン産生能及びCyp3A4抗原の存在から、主に肝細胞であると判断した。

3．臨床検体を用いた遺伝子発現解析の問題点：
出版された論文とともに公開されたデータベース(http://www.ncbi.nlm.nih.gov/geo/)をレビューした結果、171の臨床検体を用いたGeneChipデータのうち29 Chip (17%)ではRNAの質が不良であると判定された。一方で63の非臨床検体を用いたGeneChipデータではRNAの質がすべて良と判定された。このようにすでに公開・出版されているデータですら臨床検体のデータには問題点があることが示された。これは、臨床検体の取り扱いの難しさを示している。同時に、臨床検体を用いた網羅的遺伝子発現解析の実行・データの解釈にはRNAの質に注意すべきであることも示している。
網羅的遺伝子発現解析のもう一つの問題点は、個別の遺伝子発現全てにつきそのデータの信頼性とそれが意味するところを研究者自身が検証しながら研究を進めていくことが困難であることがあげられる。このためチップデータ全体の質を管理することが必要と考えられる。我々が使用している

* 自治医科大学薬理学講座臨床薬理学部門
〒329-0498 栃木県河内郡南河内薬師寺 3311-1

Affymetrix 社の GeneChip システムでは、こうした実験の質の管理を行うために用いることのできる様々なパラメータを実験結果の一部として得ることができる。このようなパラメータにはバックグラウンドノイズ・ハウスキーピングジーンの 3'/5'比・パーセントプレゼントなどがある。こうしたパラメータを活用し、研究の質をコントロールしつつ臨床検体の処理・プライマリーカルチャーの作成を行うことによって、我々はより良質のプライマリーカルチャーを作成することができたと考えている。また、既知濃度のスパイク RNA を用いて検出系の定量性・ダイナミックレンジの検討も行った。その結果、1.5 pM~100 pM での範囲での RNA 濃度の読みとりの直線性は良好であった。

4. 日本人組織を用いる必要性～人種差の克服：

海外で調整済みのプライマリーカルチャーを用いることに比較して、日本人組織を用いた遺伝子発現解析の利点として、理論的には研究結果について人種差を懸念する必要がないこと、海外調整済みのプライマリーカルチャーの利用にかかわる資金や場合によっては知的所有権などが海外へ流出する心配がないこと等がある。もちろんヒト組織であるから、種差も存在しない。さらに、自前でヒト組織を調整するため、組織の質を制御することが可能である。

我々の研究室ではこれまでのところ 11 名分の組織を研究に使用している。この 11 名につきそれぞれ 3 回、薬物等の刺激に未曝露の状態で発現解析を行ったところ、約 44,000 のトランスクリプトの内、統計学的に有意に個人差がある（t-test p 値が 0.01 未満）トランスクリプトは 100 に満たなかった。大多数の遺伝子発現は、薬物未曝露の状態では有意な個人差が見られないと判断された。

5. 現状と限界：

これまで 20 以上の薬物をプライマリーカルチャーへ曝露し、遺伝子発現解析を行った。腎障害をしばしば起こすことによって臨床的に問題になる薬物を用いて、有意に誘導や抑制される遺伝子を同定した。現在その遺伝子発現の確認を進めるとともに、その細胞内での働きを解析している。また、クラスタリングされた遺伝子群に有意に高頻度に出現する転写因子（DNA binding protein）認識配列を検索するシステムを構築した。[2]

本稿で記載した、個人差の研究により得られた情報は、重要な基礎的検討である。先行している実験動物や培養細胞株の研究に追従する形で曝露化合物の数を増加することよりも、データの質を管理するプロセスや、未刺激の日本人プライマリー腎細胞にどのような遺伝子に個人差があるのかを明らかにすることは、日本人の臨床検体を取り扱っている我々が取るべき、より優先順位の高い課題である。これらこそが限られたリソースと与えられた条件のなかで今後の日本人組織を用いた研究を活かし、示唆に富む情報を提供するもである。

[1] Yasuo Oshima, Shinsuke Kurokawa, Akihiko Tokue, Hiroyuki Mano, Ken Saito, Makoto Suzuki, Masashi Imai, and Akio Fujimura. Primary Cell Preparation of Human Renal Tubluar Cells for Transcriptome Analysis. *Toxicology Mechanisms and Methods*, 14:309-316, 2004

[2] Yasuo Oshima, Yusuke Ishida, Ayumi Shinohara, Hiroyuki Mano, Akio Fujimura. Expression Profiling of Gene with Upstream Aml1 Recognition Sequence in Hematopoietic Stem Cell-Like Fractions from Individuals with the M2 Subtype of Human Acute Myeloid Leukemia. Annual Meeting for International Society of Experimental Hematology, New Orleans, LA. USA. Jul 16-20, 2004

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.

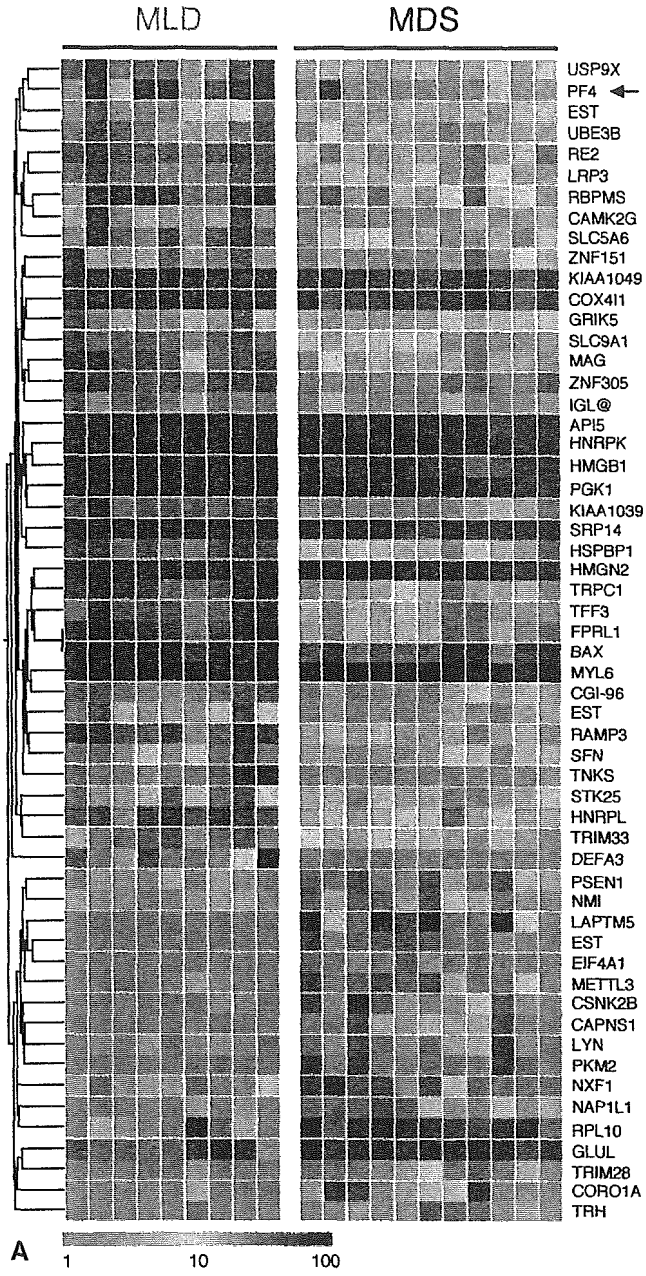
AML with dysplasia and secondary AML evolving from an undiscovered phase of MDS. Although the clinical characteristics of the former have not been fully defined, it has been reported that de novo AML-MLD may be associated with poor prognosis [17,18] and, in some cases, with an increased megakaryopoiesis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

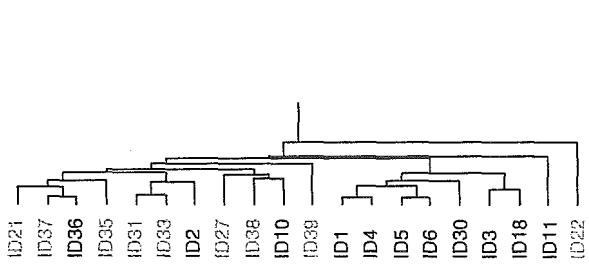
To clarify directly whether de novo AML-MLD is truly a clinical entity distinct from MDS-related leukemia, we searched for differences between the transcriptomes of $AC133^+$ cells derived from the individuals diagnosed with these two conditions. Among the 11 cases of AML-MLD studied, 9 were revealed 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





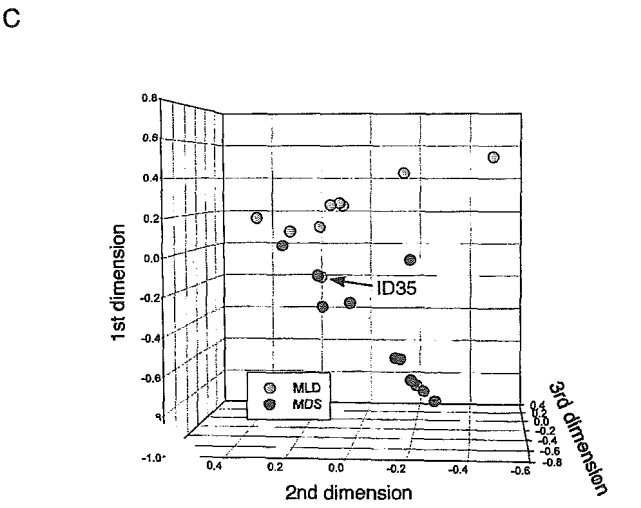


Figure 1. Continued

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).

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

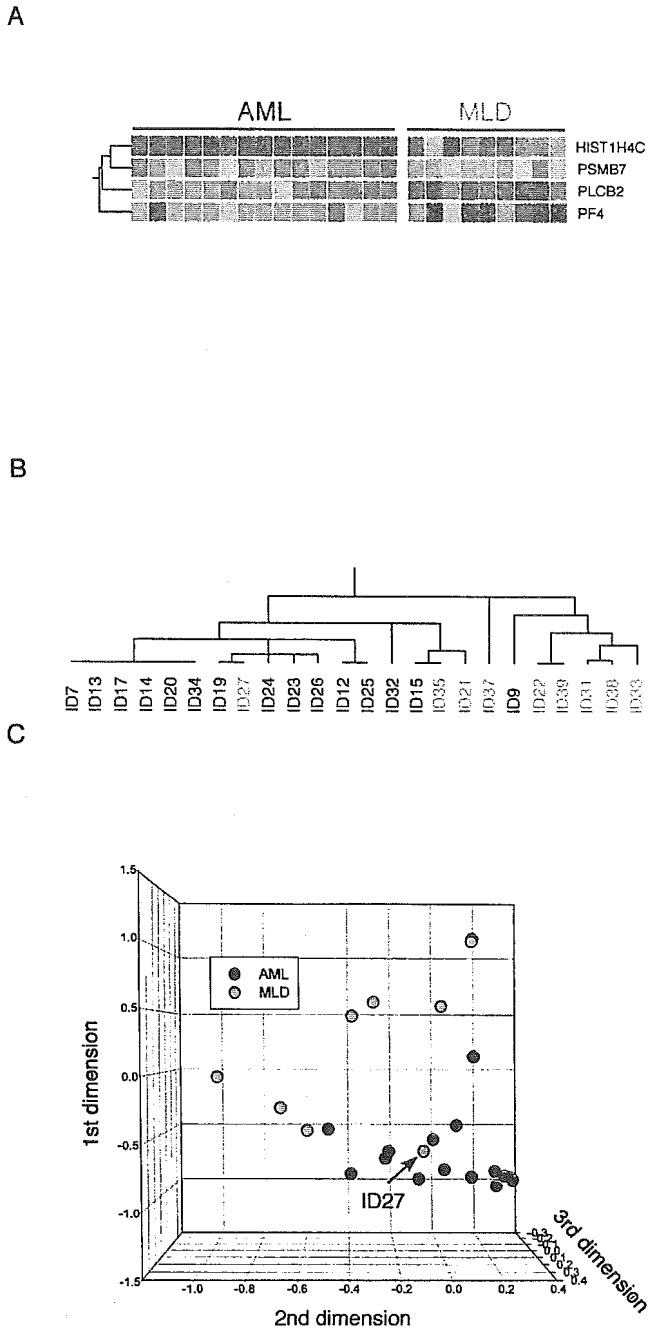


Figure 2. Comparison of gene expression profiles between patients with de novo AML without dysplasia and those with de novo AML-MLD. (**A**): Gene tree for the expression levels of four human genes in $AC133^+$ cells from individuals with de novo AML without dysplasia (AML) or de novo AML-MLD (MLD). (**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

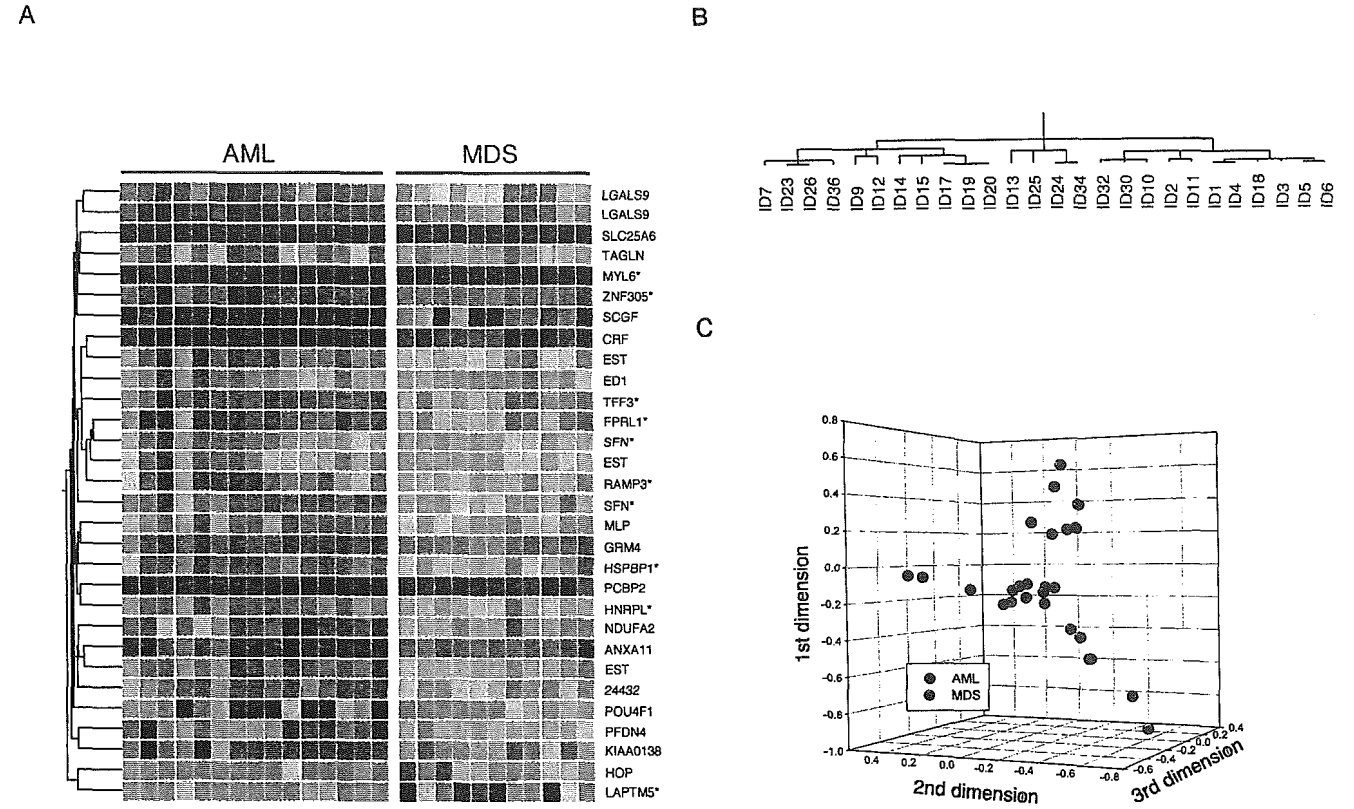


Figure 3. Comparison of gene expression profiles between patients with de novo AML without dysplasia and those with MDS-related leukemia. (**A**): Expression profiles of 30 probe sets (28 genes) that contrast de novo AML without dysplasia (AML) and MDS-related leukemia (MDS). Two independent probe sets 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.

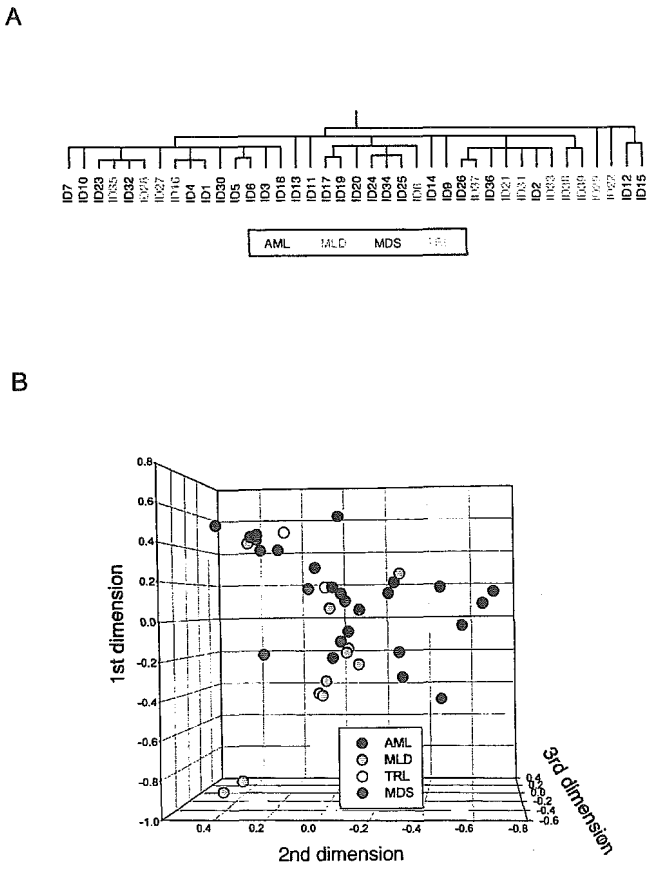


Figure 4. Comparison of all samples. Two-way clustering analysis (**A**) and correspondence analysis (**B**) were performed on all 39 specimens based on the expression profiles of the 80 probe sets whose expression was associated with the clinical diagnosis of de novo AML without dysplasia (AML, blue), 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 $\pm$ 11695 (mean $\pm$ 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 $\pm$ s.d.).

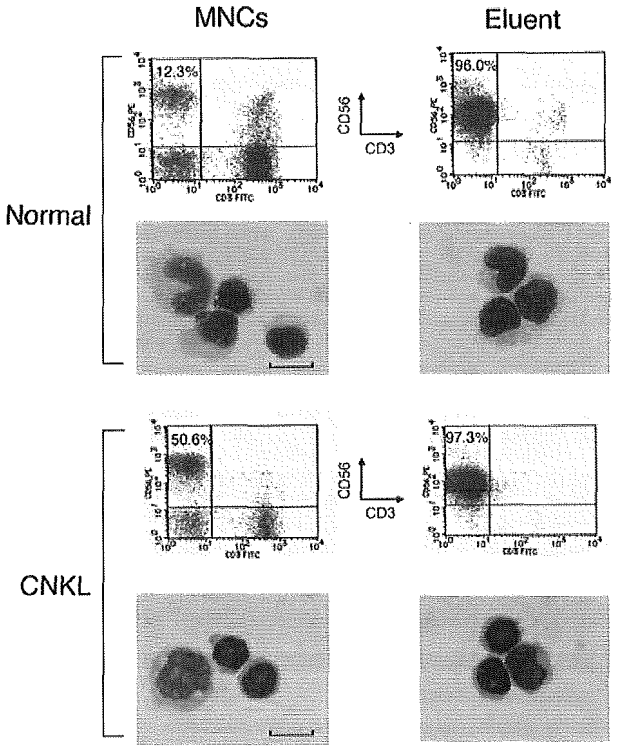


Figure 1 Purification of $CD3^-CD56^+$ fraction. MNCs isolated from PB of a healthy volunteer (normal) and a patient with CNKL were used to purify $CD3^-CD56^+$ fractions (Eluent). Cell surface expression of CD3 and CD56 was monitored in each fraction by flow cytometry, and the proportion (%) of $CD3^-CD56^+$ 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.

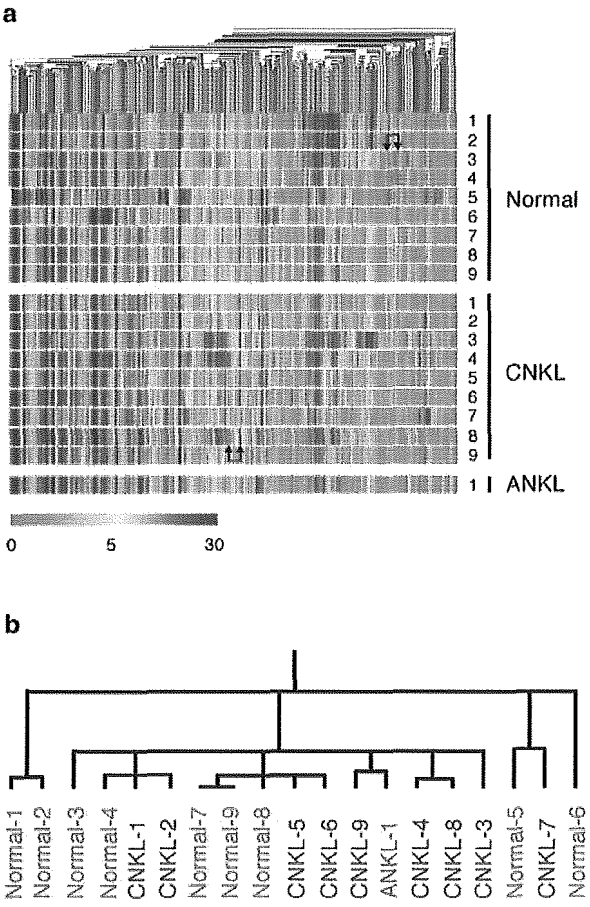


Figure 2 Expression profiles of 6494 genes in NK cell fractions. (a) Hierarchical clustering of 6494 genes on the basis of their expression profiles in $CD3^{-}CD56^{+}$ fractions derived from nine healthy volunteers (Normal), nine individuals with CNKL, and one with ANKL. Each column represents a single 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

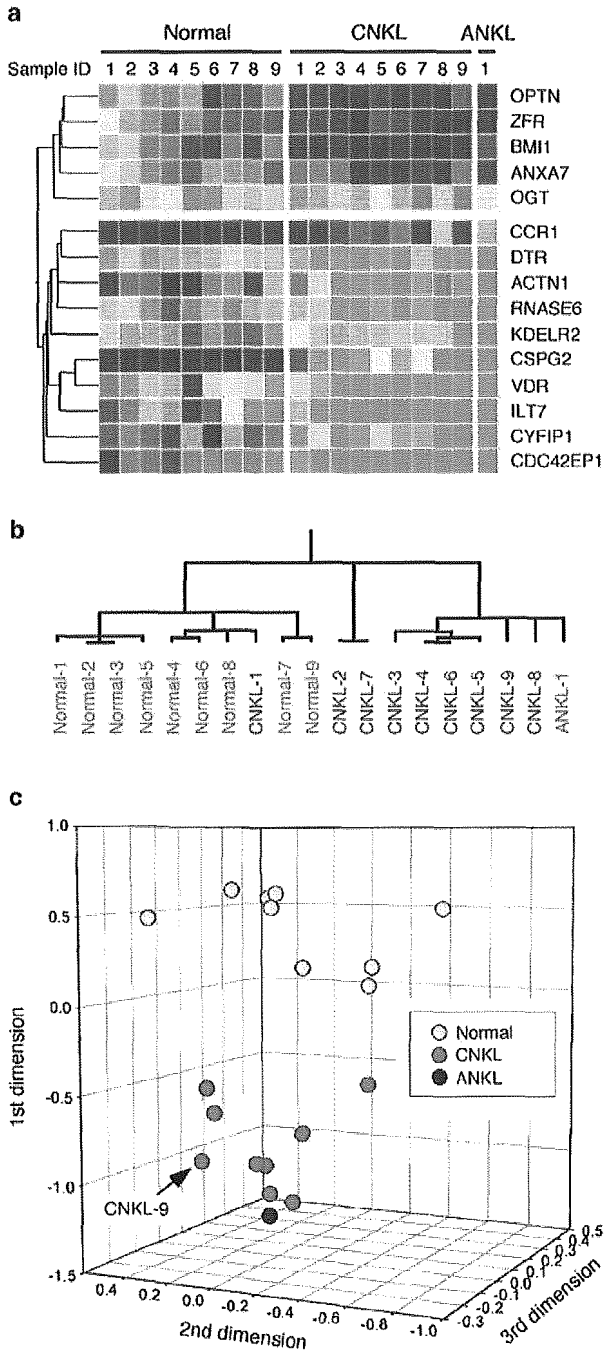


Figure 3 Identification of LDGL-associated genes. (a) Expression profiles of 15 LDGL-associated genes are shown in a dendrogram, color-coded as indicated by the scale in Figure 2a. Each row corresponds to a single gene, and each column to NK cells from healthy individuals (normal) and patients with CNKL 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

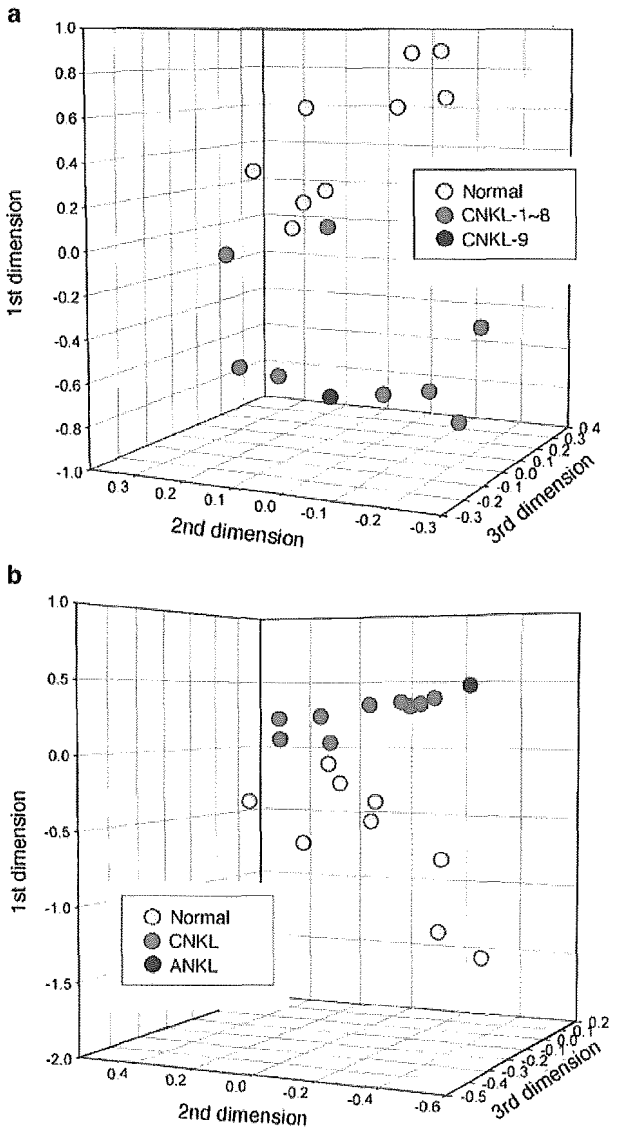


Figure 4 Investigation of the EBV$^+$ samples. (a) We could isolate nine genes, expression of which differentiated normal NK cells (normal) from indolent CNKL ones (CNKL-1–8). The EBV$^+$ CNKL-9 was projected into a virtual space together with the other normal and CNKL specimens, based on the coordinates 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).