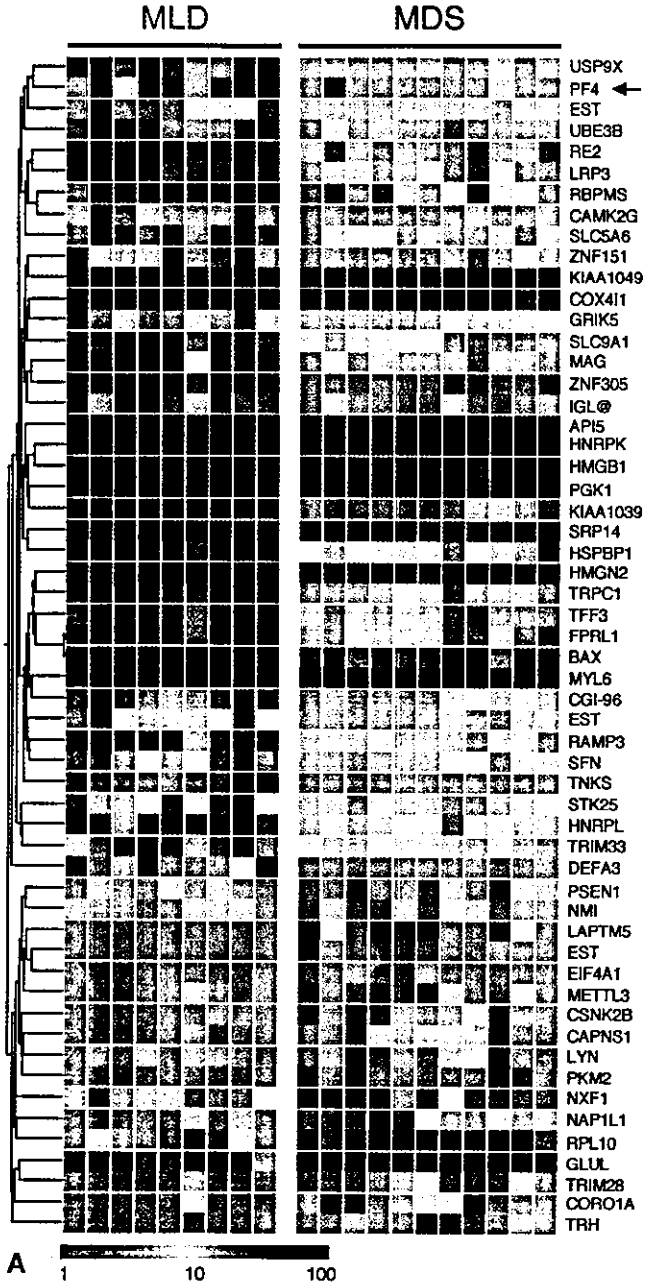

Figure 1. Comparison of gene expression profiles between individuals with de novo AML-MLD and those with MDS-related leukemia. (A): Gene tree for the expression levels (color-coded as indicated by the scale at the bottom) of 56 human genes in AC133$^+$ cells from patients with de novo AML-MLD (MLD) or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

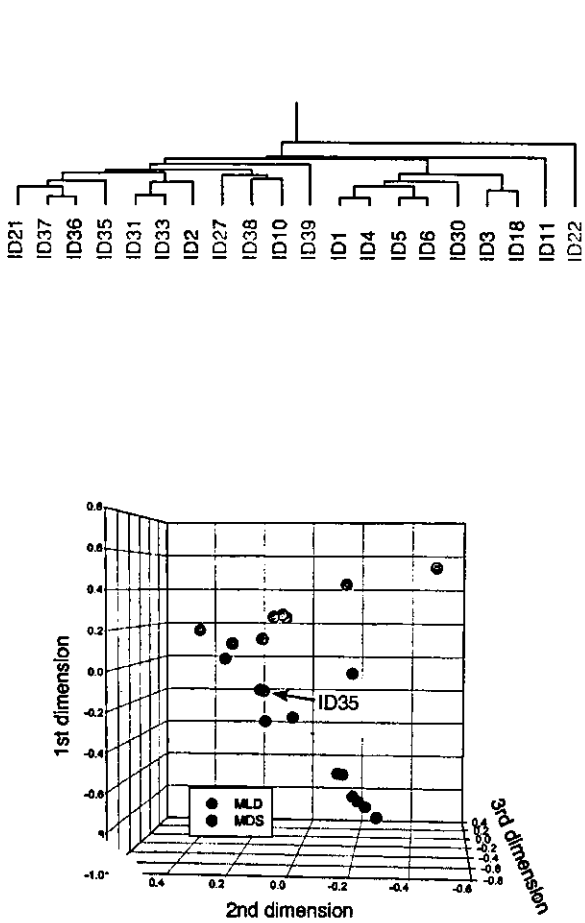

Figure 1. Continued

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

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

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

Comparison of AML without dysplasia and de novo AML-MLD

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

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

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

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

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

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

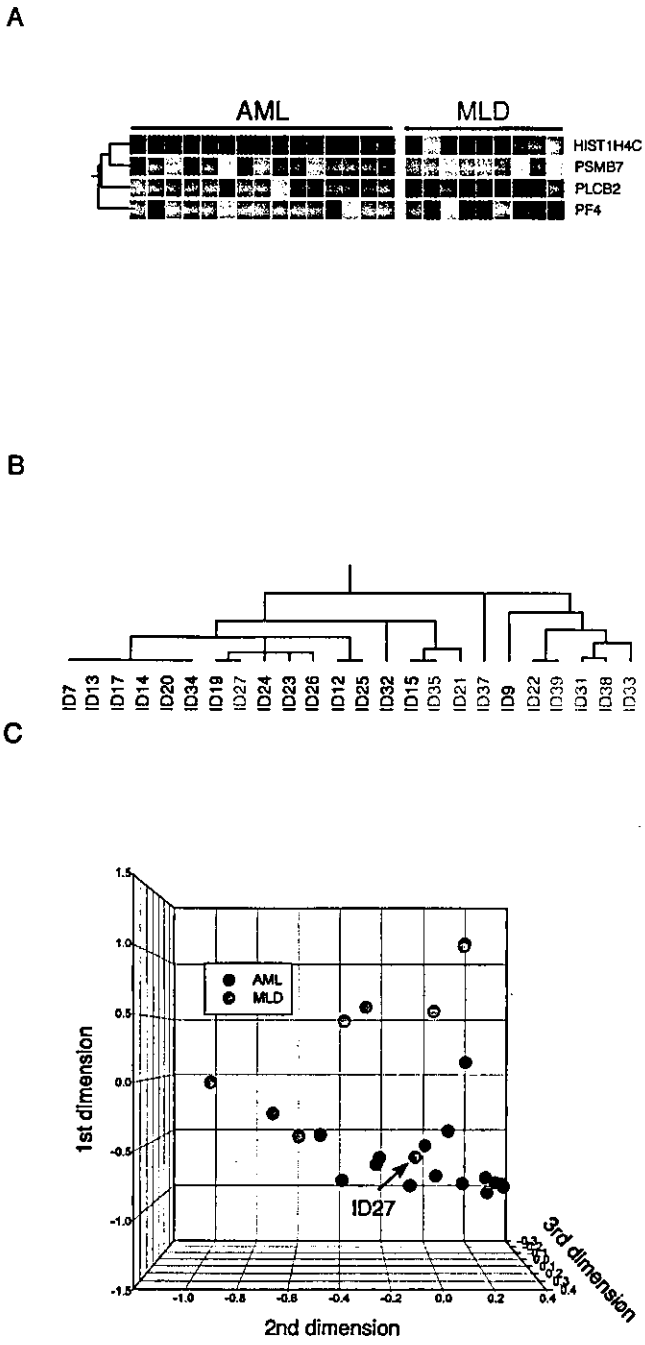

Figure 2. Comparison of gene expression profiles between patients with de novo AML without dysplasia and those with de novo AML-MLD. (A): Gene tree for the expression levels of four human genes in $AC133^+$ cells from individuals with de novo AML without dysplasia (AML) or de novo AML-MLD (MLD). (B,C):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

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

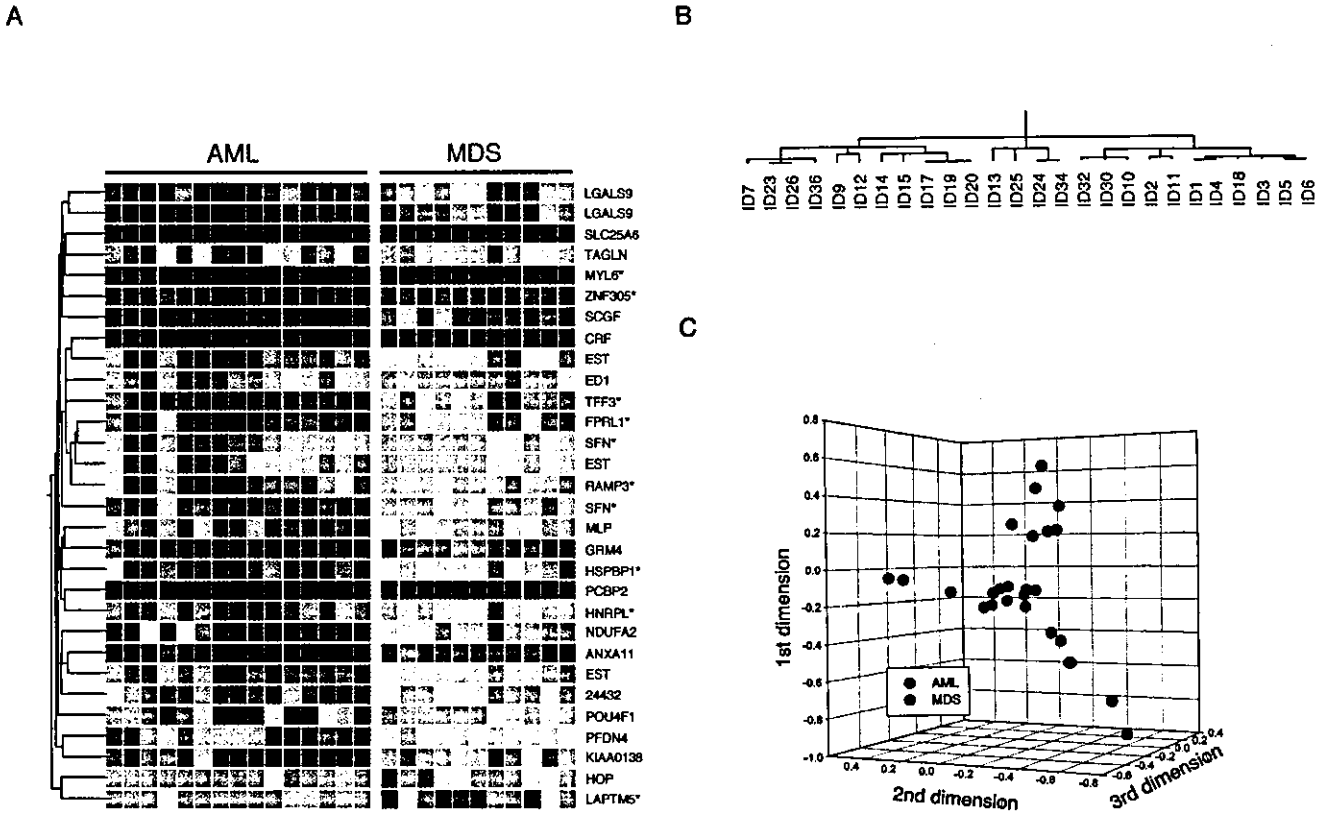

Figure 3. Comparison of gene expression profiles between patients with de novo AML without dysplasia and those with MDS-related leukemia. (**A**): Expression profiles of 30 probe sets (28 genes) that contrast de novo AML without dysplasia (AML) and MDS-related leukemia (MDS). Two independent probe sets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

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

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

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

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

Comparison of whole samples

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

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

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

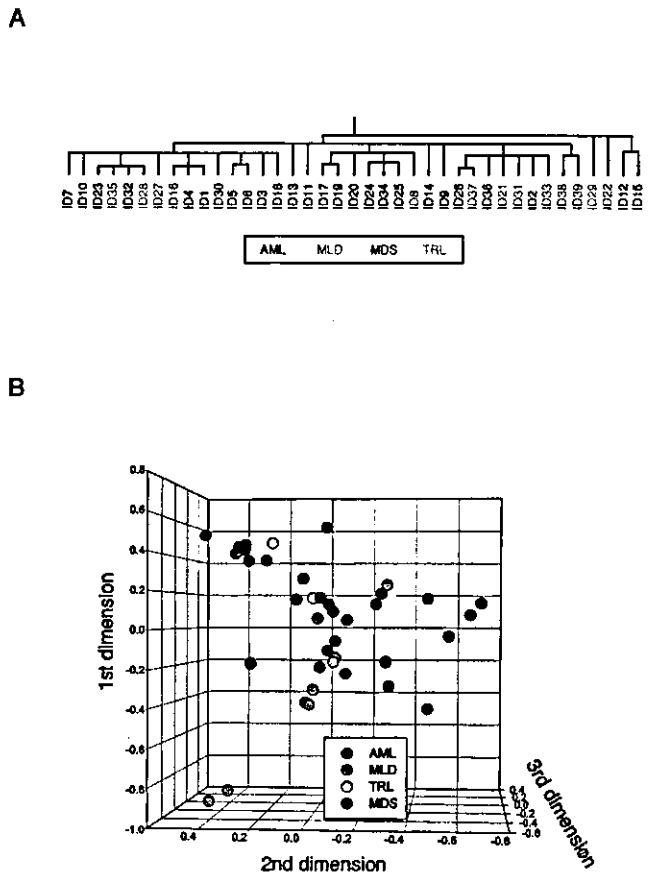

Figure 4. Comparison of all samples. Two-way clustering analysis (A) and correspondence analysis (B) were performed on all 39 specimens based on the expression profiles of the 80 probe sets whose expression was associated with the clinical diagnosis of de novo AML without dysplasia (AML, blue), AML-MLD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

Discussion

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

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

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

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

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

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

Acknowledgments

We thank all patients and physicians who participated in the Blast Bank project. This work was supported in part by grants for the Second-Term Comprehensive 10-Year Strategy for Cancer Control, and for Research on Development of Novel Therapeutic Modalities for Myelodysplastic Syndrome from the Ministry of Health, Labor, and Welfare of Japan; by a grant from Research Foundation for Community Medicine of Japan; by a grant from Sankyo Foundation of Life Science; by a grant from Takeda Science Foundation; and by a grant from Mitsubishi Pharma Research Foundation.

References

1. Jaffe ES, Harris NL, Stein H, Vardiman JW, eds. Pathology and genetics of tumours of haematopoietic and lymphoid tissues. Lyon: IARC Press; 2001.
2. Smith SM, Le Beau MM, Huo D, et al. Clinical-cytogenetic associations in 306 patients with therapy-related myelodysplasia and myeloid leukemia: the University of Chicago series. Blood. 2003;102:43–52.
3. Bain BJ. The bone marrow aspirate of healthy subjects. Br J Haematol. 1996;94:206–209.
4. Brito-Babapulle F, Catovsky D, Galton DAG. Clinical and laboratory features of de novo acute myeloid leukemia with trilineage myelodysplasia. Br J Haematol. 1987;66:445–450.
5. Jinnai I, Tomonaga M, Kuriyama K, et al. Dysmegakaryocytopoiesis in acute leukaemias: its predominance in myelomonocytic (M4) leukaemia and implication for poor response to chemotherapy. Br J Haematol. 1987;66:467–472.
6. Kuriyama K, Tomonaga M, Matsuo T, et al., Japan Adult Leukaemia Study Group (JALSG). Poor response to intensive chemotherapy in de novo acute myeloid leukaemia with trilineage myelodysplasia. Br J Haematol. 1994;86:767–773.
7. Taguchi J, Miyazaki Y, Yoshida S, et al. Allogeneic bone marrow transplantation improves the outcome of de novo AML with trilineage dysplasia (AML-TLD). Leukemia. 2000;14:1861–1866.
8. Miyazato A, Ueno S, Ohmine K, et al. Identification of myelodysplastic syndrome–specific genes by DNA microarray analysis with purified hematopoietic stem cell fraction. Blood. 2001;98:422–427.
9. Hin AH, Miraglia S, Zanjani ED, et al. AC133, a novel marker for human hematopoietic stem and progenitor cells. Blood. 1997;90:5002–5012.
10. Oshima Y, Ueda M, Yamashita Y, et al. DNA microarray analysis of hematopoietic stem cell–like fractions from individuals with the M2 subtype of acute myeloid leukemia. Leukemia. 2003;17:1990–1997.
11. Ohmine K, Ota J, Ueda M, et al. Characterization of stage progression in chronic myeloid leukemia by DNA microarray with purified hematopoietic stem cells. Oncogene. 2001;20:8249–8257.
12. Ueda M, Ota J, Yamashita Y, et al. DNA microarray analysis of stage progression mechanism in myelodysplastic syndrome. Br J Haematol. 2003;123:288–296.
13. Ota J, Yamashita Y, Okawa K, et al. Proteomic analysis of hematopoietic stem cell–like fractions in leukemic disorders. Oncogene. 2003;22:5720–5728.
14. Van Gelder RN, von Zastrow ME, Yool A, et al. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc Natl Acad Sci U S A. 1990;87:1663–1667.
15. Bennett JM, Catovsky D, Daniel MT, et al. Proposed revised criteria for the classification of acute myeloid leukemia. A report of the French-American-British Cooperative Group. Ann Intern Med. 1985;103:620–625.
16. Fellenberg K, Hauser NC, Brors B, et al. Correspondence analysis applied to microarray data. Proc Natl Acad Sci U S A. 2001;98:10781–10786.
17. Goasguen JE, Matsuo T, Cox C, Bennett JM. Evaluation of the dysmyelopoiesis in 336 patients with de novo acute myeloid leukemia: major importance of dysgranulopoiesis for remission and survival. Leukemia. 1992;6:520–525.
18. Kuriyama K, Tomonaga M, Kobayashi T, et al. Morphological diagnoses of the Japan adult leukemia study group acute myeloid leukemia protocols: central review. Int J Hematol. 2001;73:93–99.
19. Dhanasekaran SM, Barrette TR, Ghosh D, et al. Delineation of prognostic biomarkers in prostate cancer. Nature. 2001;412:822–826.
20. Gurney D, Lip GY, Blann AD. A reliable plasma marker of platelet activation: does it exist? Am J Hematol. 2002;70:139–144.
21. Alon U, Barkai N, Notterman DA, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci U S A. 1999;96:6745–6750.
22. Bustin M. Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. Mol Cell Biol. 1999;19:5237–5246.
23. Gong TW, Huang L, Warner SJ, Lomax MI. Characterization of the human UBE3B gene: structure, expression, evolution, and alternative splicing. Genomics. 2003;82:143–152.
24. Jones MH, Furlong RA, Burkin H, et al. The Drosophila developmental gene fat facets has a human homologue in Xp11.4 which escapes X-inactivation and has related sequences on Yq11.2. Hum Mol Genet. 1996;5:1695–1701.

Leukemia (2004) 18, 556–565

www.nature.com/leu

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

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

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

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

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

Introduction

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

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

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

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

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

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

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

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

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

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

Materials and methods

Purification of $CD3^-CD56^+$ cells

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

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

DNA microarray analysis

Total RNA was extracted from the $CD3^-CD56^+$ cell preparations by the acid guanidinium method, and was subjected to two rounds of amplification with T7 RNA polymerase as described.[20] High fidelity of our RNA amplification procedure has been already reported.[18] The amplified cRNA (2 μg) was then converted to double-stranded cDNA, which was used to prepare biotin-labeled cRNA for hybridization with GeneChip HGU95Av2 microarrays (Affymetrix, Santa Clara, CA, USA) harboring oligonucleotides corresponding to a total of 12 625 genes. Hybridization, washing, and detection of signals on the arrays were performed with the GeneChip system (Affymetrix).

Class prediction by the 'weighted distance method'

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

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

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

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

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

All raw array data as well as details of the genes shown in the figures are available as supplementary information at the *Leukemia* web site.

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

Portions of nonamplified cDNA were subjected to PCR with a QuantiTect SYBR Green PCR Kit (Qiagen, Valencia, CA, USA). The amplification protocol comprised incubations at 94°C for 15 s, 60°C for 30 s, and 72°C for 60 s. Incorporation of the SYBR Green dye into PCR products was monitored in real time with an ABI PRISM 7700 sequence detection system (PE Applied Biosystems, Foster City, CA, USA), thereby allowing determination of the threshold cycle (C_T) at which exponential amplification of PCR products begins. The C_T values for cDNAs corresponding to the glyceraldehyde-3-phosphate dehydrogen-

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

Determination of serum level of IFNG protein

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

Results

Purification of NK cells

To directly compare the transcriptome of normal and affected NK cells, we here purified $CD3^-CD56^+$ fractions from PB MNCs of healthy volunteers ($n=9$) as well as of individuals with CNKL ($n=9$) or ANKL ($n=1$). A total of 19 specimens were thus registered into this study. The clinical characteristics of the 10 patients (CNKL-1~9 and ANKL-1) are summarized in Table 1. The LGL count in their PB was 14056 cells/ml$\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 $3.9\times10^5\pm3.4\times10^5$ (mean$\pm$s.d.).

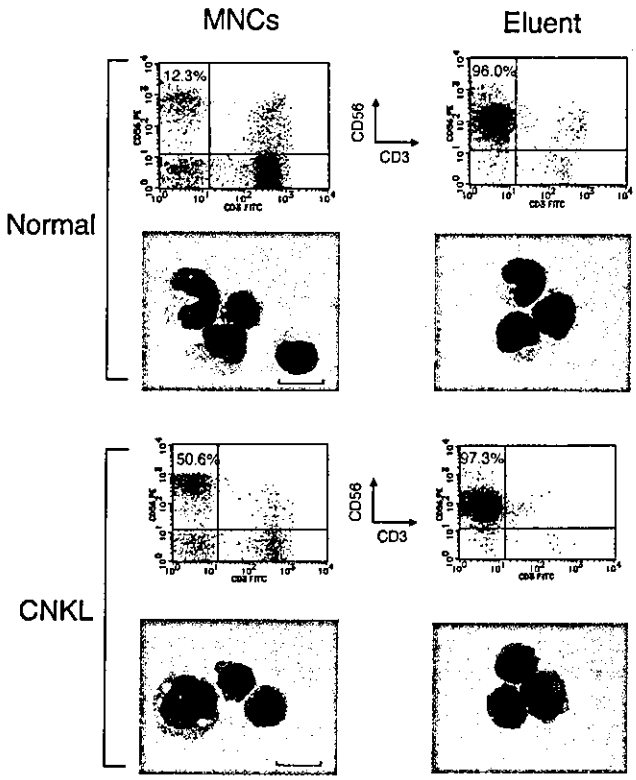

Figure 1 Purification of $CD3^-CD56^+$ fraction. MNCs isolated from PB of a healthy volunteer (normal) and a patient with CNKL were used to purify $CD3^-CD56^+$ fractions (Eluent). Cell surface expression of CD3 and CD56 was monitored in each fraction by flow cytometry, and the proportion (%) of $CD3^-CD56^+$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

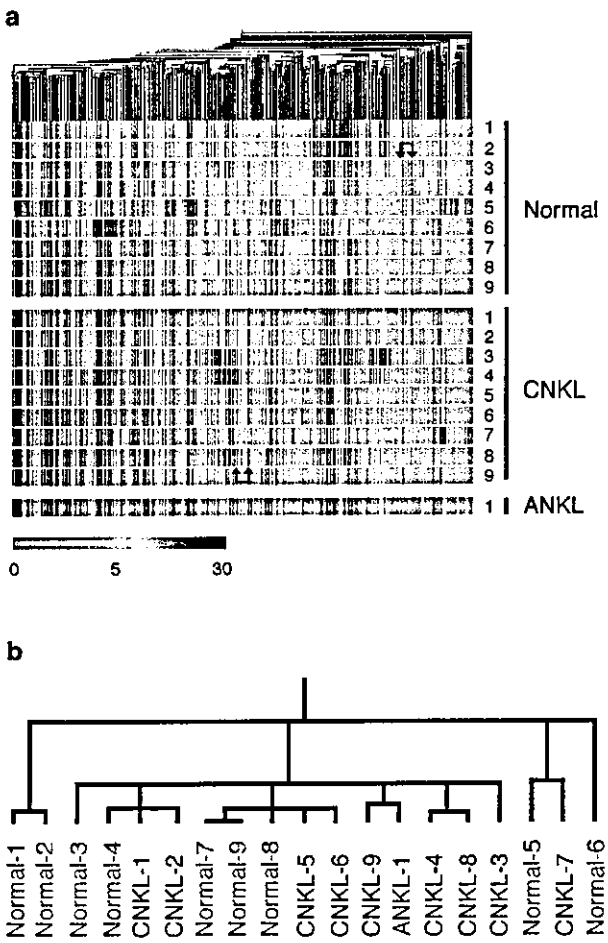

Figure 2 Expression profiles of 6494 genes in NK cell fractions. (a) Hierarchical clustering of 6494 genes on the basis of their expression profiles in $CD3^{-}CD56^{+}$ fractions derived from nine healthy volunteers (Normal), nine individuals with CNKL, and one with ANKL. Each column represents a single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

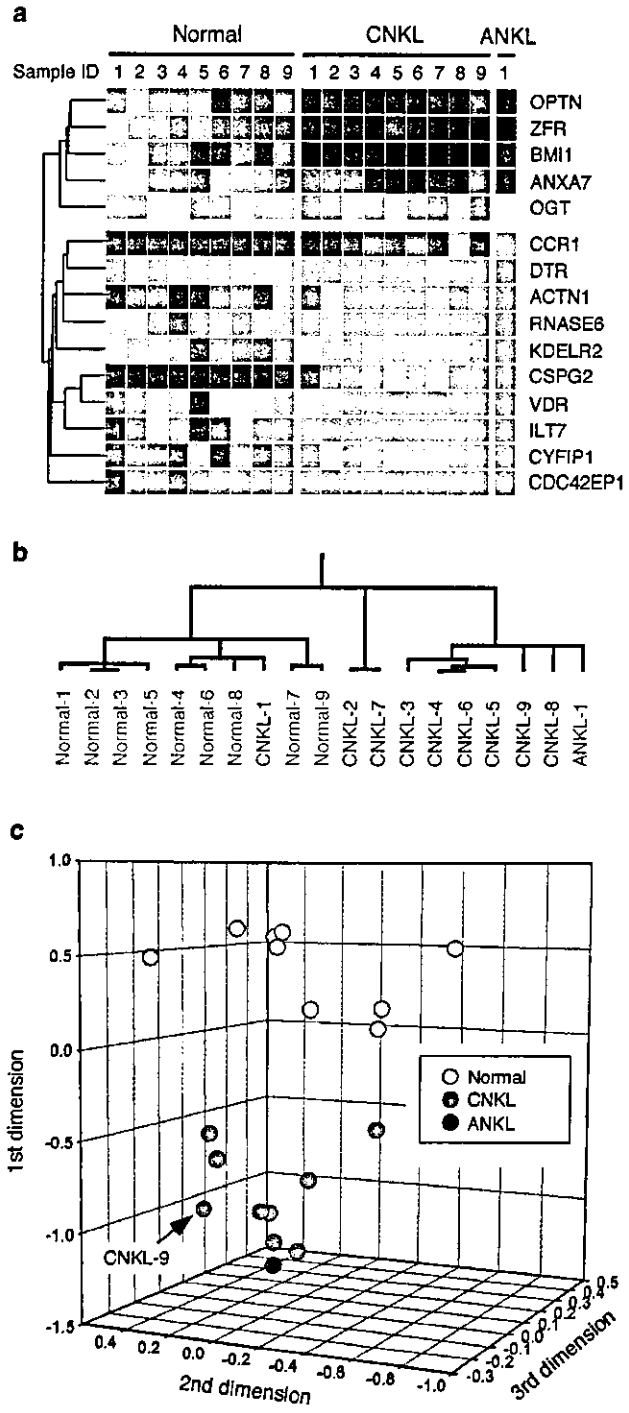

Figure 3 Identification of LDGL-associated genes. (a) Expression profiles of 15 LDGL-associated genes are shown in a dendrogram, color-coded as indicated by the scale in Figure 2a. Each row corresponds to a single gene, and each column to NK cells from healthy individuals (normal) and patients with CNKL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

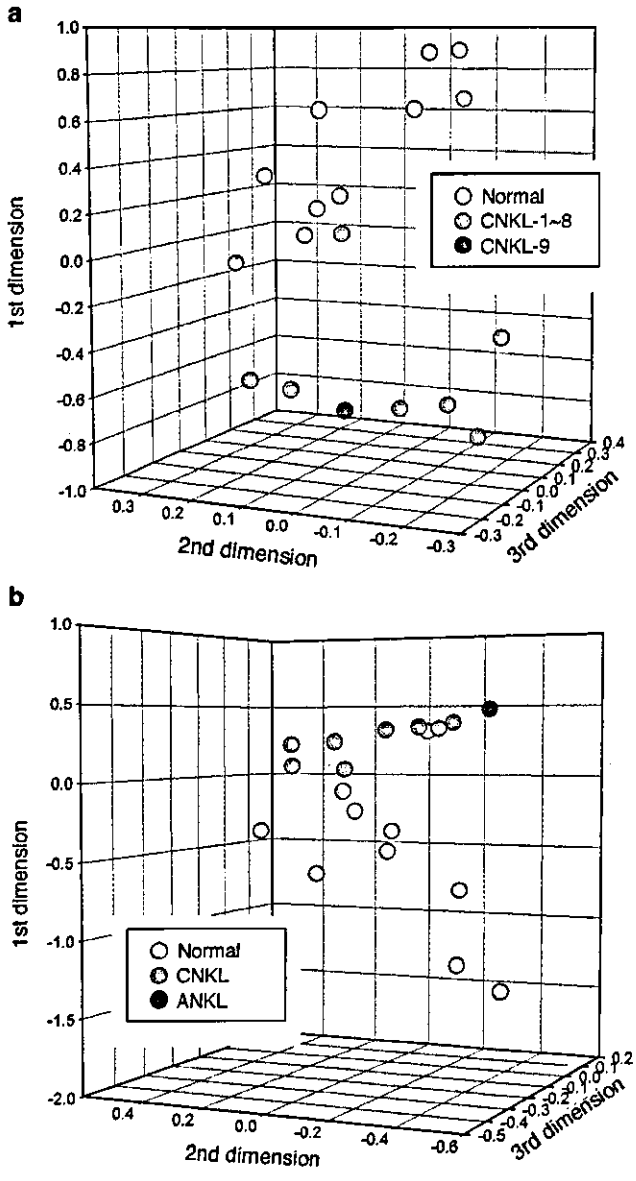

Figure 4 Investigation of the EBV$^+$ samples. (a) We could isolate nine genes, expression of which differentiated normal NK cells (normal) from indolent CNKL ones (CNKL-1–8). The EBV$^+$ CNKL-9 was projected into a virtual space together with the other normal and CNKL specimens, based on the coordinates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

Isolation of single-gene markers for LDGL diagnosis

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

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

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

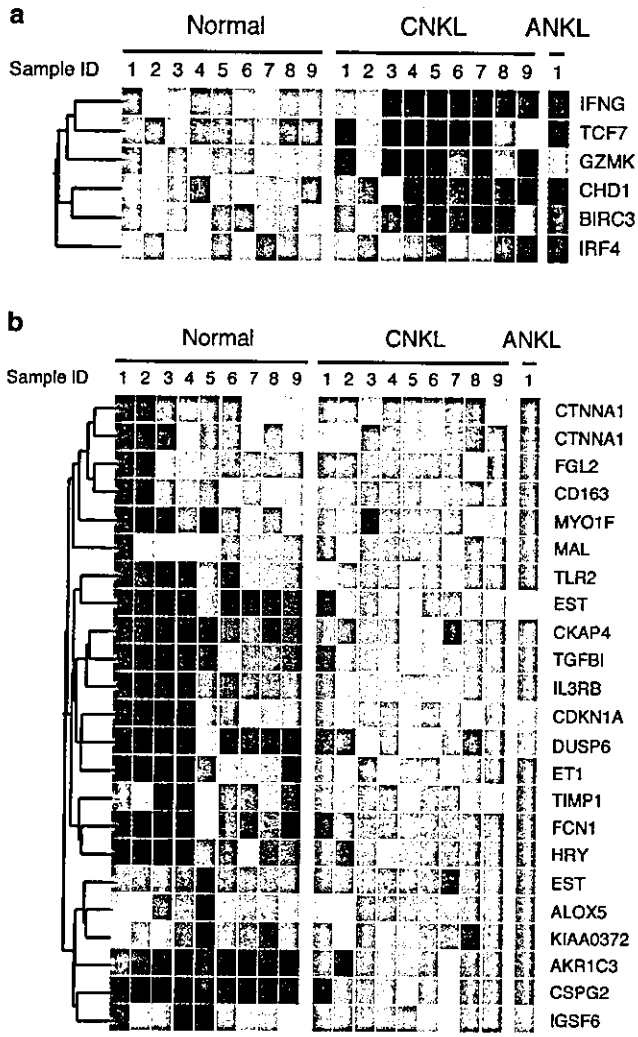

Figure 5 Identification of single-gene markers for LDGL. (a) Dendrogram showing the expression profiles of six genes whose expression intensity was kept suppressed in normal NK cells, but became activated in a part of affected NK cells. Each row represents a single gene, and each column a separate patient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

Confirmation of overproduction of IFNG

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

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

Discussion

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

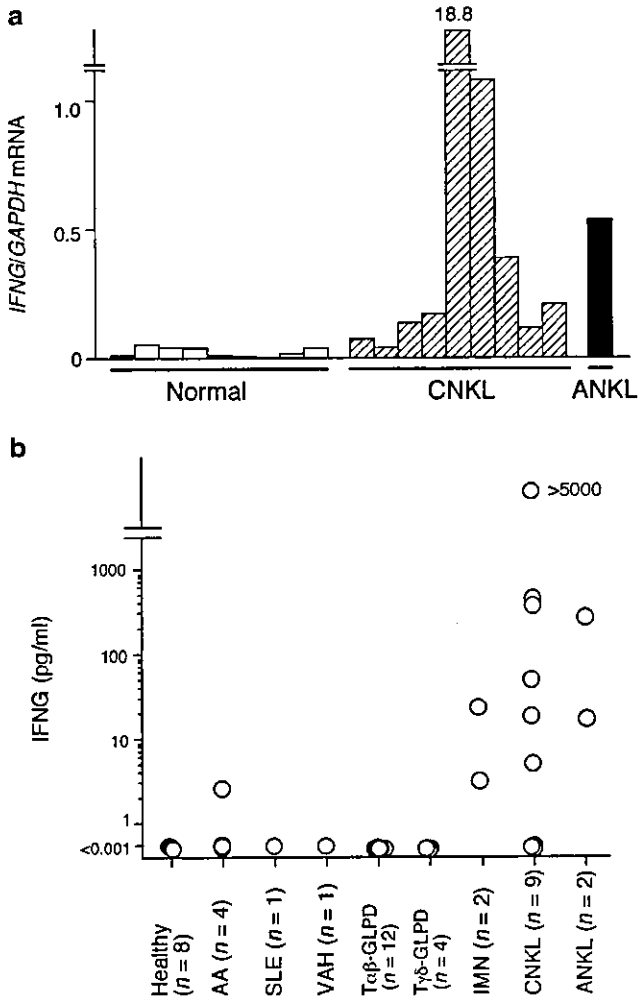

Figure 6 Confirmation of overexpression of IFNG. (a) Quantitation of *IFNG* mRNA in NK cell fractions. Complementary DNA was prepared from the NK cell fractions, and was subjected to real-time RT-PCR analysis with primers specific for the *IFNG* or *GAPDH* genes. The ratio of the abundance of *IFNG* mRNA to that of *GAPDH* mRNA was calculated as 2^n, where n is the C_T value for *GAPDH* cDNA minus the C_T value for *IFNG* cDNA. (b) Sera were obtained from healthy volunteers (healthy) 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 expression level of IFNG protein in the sera was determined by flow cytometry, and shown as pg/ml.

intrinsic for an accurate comparison of the disease status, since simple comparison with PB MNCs would be severely influenced by any changes in the cell composition of PB MNCs in each individual.

However, even by using expression profiles of the purified fractions, similarity of the expression pattern of all gene set failed to clearly separate the affected NK cells from normal ones (Figure 2b), indicating the necessity of a diagnostic system with a 'supervised' algorithm. For this aim, we first extracted gene clusters, expression of which was specific to either normal or affected NK cells. Correspondence analysis on the expression patterns of such 'LDGL-associated genes' has succeeded in the reduction of the number of pattern dimensions into three. To our surprise, projection of all samples into this 3D space clearly demonstrated that the affected NK cells (CNKL/ANKL) were placed clustered at a position separate from that of normal NK cells (Figure 3c). With coordinates in such decomposited dimensions, we then invented a novel class prediction means, 'weighted-distance method'. As expected from the clear separation in the 3D view, the weighted-distance method provided correct prediction for all samples studied. A cross-validation trial for the disease diagnosis also gave a highly accurate prediction rate (77.8%).

Given a high incidence of clonal EBV infection in ANKL cells, EBV is believed to play an essential role in the pathogenesis of this fulminant disorder. From the point of view of gene-expression profile, the NK cells positive for EBV infection (CNKL-9 and ANKL-1) shared gene-expression patterns characteristic to other indolent CNKL cells. For instance, a comparison of normal NK cells and CNKL samples (except EBV^+ CNKL-9) has identified a total of nine CNKL-related genes including those for nuclear matrix protein-2 (GenBank accession number, D50926), cytochrome *C* (D00265), tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, theta polypeptide or 14-3-3 protein tau (X56468), O-linked GlcNAc transferase (AL050366), IFNG, chemokine C–C motif receptor 1 (CCR1; D10925), chondroitin sulfate proteoglycan 2 (X15998), and fibrinogen-like 2 (AI432401). The 3D view of the correspondence analysis for these 'CNKL-associated' genes clearly demonstrated that the EBV^+ CNKL-9 subject was included in the other indolent CNKL group (Figure 4a). Similarly, another EBV^+ subject (ANKL-1) was placed closely with the CNKL samples, according to the expression profiles of the genes which differentiated CNKL cells from normal NK cells (Figure 4b).

These data propose a hypothesis that gene-expression alterations characteristic to an activated, yet indolent, proliferation of NK cells found in CNKL patients also take place in the highly proliferating NK cells in EBV^+ individuals. In other words, a similar mechanism may be utilized for the sustained outgrowth of NK cells both in the individuals with CNKL and ANKL. It should be emphasized, however, that the number of EBV^+ samples ($n=2$) in these analyses was too small to extract any conclusive remarks on the pathophysiology of proliferating EBV^+ NK cells. Nevertheless, we believe that it was interesting to find that EBV^+ NK cells may share a molecular signature with EBV^- LDGL cells.

We finally tried to identify single-gene markers, the presence or absence of which helps the diagnosis of LDGL. A total of 22 genes were shown to be specific to normal NK cells, including those for cyclin-dependent kinase inhibitor 1A (CDKN1A; GenBank accession number U03106) or CIP1, dual-specificity phosphatase 6 (DUSP6; AB013382), toll-like receptor 2 (TLR2; AF051152), tissue inhibitor of metalloproteinase 1 (TIMP1; D11139), and aldo-keto reductase family 1 member C3 (AKR1C3; D17793) (Figure 5b).

CDKN1A is transcriptionally regulated by the activity of p53, and functions as a major effector for antitumor activity of p53, through the suppression of cyclin-dependent kinase activities.[32] Therefore, loss of expression of *CDKN1A* may allow uncontrolled transition at the G_2–M boundary in the cell cycle, and may partially account for the overgrowth of NK cells in LDGL individuals. Similarly, DUSP6 antagonizes MAPK activities *via* dephosphorylation of the latter kinases.[33] Decrease of DUSP6 expression may therefore contribute to overactivation of MAPK and to enhanced mitogenesis.

AKR1C3 catalyzes conversion of aldehydes and ketones to alcohols *in vivo*. Although its role in NK cells is unknown yet, downregulation of its transcription has been also reported in the

LGLs of T cell-type LDGL.[19] Comparison with DNA microarray of $CD4^{-}CD8^{+}$ T-cells between healthy individuals and those with T cell-type LDGL has identified the *AKR1C3* gene as the specific marker to the former. Decrease of *AKR1C3* message was also confirmed by quantitative 'real-time' RT-PCR method in those patients. Transcriptional suppression of *AKR1C3* was thus revealed in affected LGLs for both NK cell- and T cell-type LDGL, and may be a common marker for the diagnosis of LDGL condition.

A number of growth-promoting factors were found in the 'LDGL-specific genes' in Figure 5a. IFNG and BIRC3[34] are, for instance, known to protect NK cells from apoptosis, and IRF4 has an oncogenic activity *in vivo*.[35] Additionally, CHD1 contains an SNF2-related helicase/ATPase domain, and is presumed to be involved in the regulation of chromatin structure and gene transcription as well.[36]

Our data, together with that by Mizuno *et al*[31] suggest that the serum concentration of IFNG protein may be an indicator of NK cell-type LDGL. Direct production of IFNG by affected NK cells may also imply the presence of an autocrine loop for the NK cell growth.

The mechanism by which affected NK cells produce IFNG is still to be revealed. It is known that IL-2, IL-12, IL-15, and IL-18 all activate production of IFNG in NK cells. In our microarray data set, however, none of IL-2, IL-12, and IL-18 were found to be significantly expressed in the subjects (not shown). Although IL-15 was moderately expressed in our NK samples, its expression level did not differ between normal and affected NK cells. In support of this notion, we could not detect significant level of IL-2 protein in the examination of serum level of cytokines (not shown). Therefore, it is currently an open question as to whether activation of *IFNG* transcription in the affected NK cells is a secondary event from the stimulation by other cells such as T cells, or intracellular mechanism of IFNG expression is deregulated in the affected NK cells.

Conclusion

We have characterized the transcriptome of a relatively uncommon disorder, NK cell-type LDGL. Comparison of purified NK cells between healthy and CNKL individuals led to the identification of gene sets which are useful in the expression profile-based differential diagnosis of the disorder. Such disease-associated genes have also provided us insights into the molecular pathogenesis of NK cell-type LDGL. Together with further optimization of statistical methods, increase in the number of both genes and subjects for the analysis would help to define and clarify the clinical entities of NK cell disorders.

Acknowledgements

We thank all patients, healthy volunteers, and physicians who participated in the collection of NK cell depository. We are also grateful to Dr T Miwa for his helpful suggestions.

Supplementary Information

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

References

1 Loughran Jr TP. Clonal diseases of large granular lymphocytes. *Blood* 1993; **82**: 1–14.
2 Semenzato G, Zambello R, Starkebaum G, Oshimi K, Loughran Jr TP. The lymphoproliferative disease of granular lymphocytes: updated criteria for diagnosis. *Blood* 1997; **89**: 256–260.
3 Oshimi K. Granular lymphocyte proliferative disorders: report of 12 cases and review of the literature. *Leukemia* 1988; **2**: 617–627.
4 Loughran Jr TP, Starkebaum G. Large granular lymphocyte leukemia. Report of 38 cases and review of the literature. *Medicine (Baltimore)* 1987; **66**: 397–405.
5 Jaffe ES, Harris NL, Stein H, Vardiman JW (eds). *Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues*. Lyon: IARC Press, 2001.
6 Gelb AB, van de Rijn M, Regula Jr DP, Cornbleet JP, Kamel OW, Horoupian DS *et al*. Epstein–Barr virus-associated natural killer-large granular lymphocyte leukemia. *Hum Pathol* 1994; **25**: 953–960.
7 Tefferi A. Chronic natural killer cell lymphocytosis. *Leuk Lymphoma* 1996; **20**: 245–248.
8 Rabbani GR, Phyliky RL, Tefferi A. A long-term study of patients with chronic natural killer cell lymphocytosis. *Br J Haematol* 1999; **106**: 960–966.
9 Nash R, McSweeney P, Zambello R, Semenzato G, Loughran Jr TP. Clonal studies of CD3-lymphoproliferative disease of granular lymphocytes. *Blood* 1993; **81**: 2363–2368.
10 Garcia-Suarez J, Prieto A, Reyes E, Arribalzaga K, Perez-Machado MA, Lopez-Rubio M *et al*. Persistent lymphocytosis of natural killer cells in autoimmune thrombocytopenic purpura (ATP) patients after splenectomy. *Br J Haematol* 1995; **89**: 653–655.
11 Ohno Y, Amakawa R, Fukuhara S, Huang CR, Kamesaki H, Amano H *et al*. Acute transformation of chronic large granular lymphocyte leukemia associated with additional chromosome abnormality. *Cancer* 1989; **64**: 63–67.
12 Tefferi A, Greipp PR, Leibson PJ, Thibodeau SN. Demonstration of clonality, by X-linked DNA analysis, in chronic natural killer cell lymphocytosis and successful therapy with oral cyclophosphamide. *Leukemia* 1992; **6**: 477–480.
13 Morice WG, Kurtin PJ, Leibson PJ, Tefferi A, Hanson CA. Demonstration of aberrant T-cell and natural killer-cell antigen expression in all cases of granular lymphocytic leukaemia. *Br J Haematol* 2003; **120**: 1026–1036.
14 Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G. Making and reading microarrays. *Nat Genet* 1999; **21**: 15–19.
15 Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. *Nat Genet* 1999; **21**: 10–14.
16 Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K *et al*. Delineation of prognostic biomarkers in prostate cancer. *Nature* 2001; **412**: 822–826.
17 Miyazato A, Ueno S, Ohmine K, Ueda M, Yoshida K, Yamashita Y *et al*. Identification of myelodysplastic syndrome-specific genes by DNA microarray analysis with purified hematopoietic stem cell fraction. *Blood* 2001; **98**: 422–427.
18 Ohmine K, Ota J, Ueda M, Ueno S-i, Yoshida K, Yamashita Y *et al*. Characterization of stage progression in chronic myeloid leukemia by DNA microarray with purified hematopoietic stem cells. *Oncogene* 2001; **20**: 8249–8257.
19 Makishima H, Ishida F, Ito T, Kitano K, Ueno S, Ohmine K *et al*. DNA microarray analysis of T cell-type lymphoproliferative disease of granular lymphocytes. *Br J Haematol* 2002; **118**: 462–469.
20 Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci USA* 1990; **87**: 1663–1667.
21 Fellenberg K, Hauser NC, Brors B, Neutzner A, Hoheisel JD, Vingron M. Correspondence analysis applied to microarray data. *Proc Natl Acad Sci USA* 2001; **98**: 10781–10786.
22 Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D *et al*. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci USA* 1999; **96**: 6745–6750.

23 Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat Genet* 2003; **33**: 49–54.
24 Alkema MJ, Wiegant J, Raap AK, Berns A, van Lohuizen M. Characterization and chromosomal localization of the human proto-oncogene BMI-1. *Hum Mol Genet* 1993; **2**: 1597–1603.
25 Park IK, Qian D, Kiel M, Becker MW, Pihalja M, Weissman IL *et al.* Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* 2003; **423**: 302–305.
26 Lessard J, Sauvageau G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature* 2003; **423**: 255–260.
27 Meagher MJ, Braun RE. Requirement for the murine zinc finger protein ZFR in perigastrulation growth and survival. *Mol Cell Biol* 2001; **21**: 2880–2890.
28 Thornton S, Kuhn KA, Finkelman FD, Hirsch R. NK cells secrete high levels of IFN-gamma in response to in vivo administration of IL-2. *Eur J Immunol* 2001; **31**: 3355–3360.
29 Ross ME, Caligiuri MA. Cytokine-induced apoptosis of human natural killer cells identifies a novel mechanism to regulate the innate immune response. *Blood* 1997; **89**: 910–918.
30 Carson WE, Giri JG, Lindemann MJ, Linett ML, Ahdieh M, Paxton R *et al.* Interleukin (IL) 15 is a novel cytokine that activates human natural killer cells via components of the IL-2 receptor. *J Exp Med* 1994; **180**: 1395–1403.
31 Mizuno S, Akashi K, Ohshima K, Iwasaki H, Miyamoto T, Uchida N *et al.* Interferon-gamma prevents apoptosis in Epstein–Barr virus-infected natural killer cell leukemia in an autocrine fashion. *Blood* 1999; **93**: 3494–3504.
32 El-Deiry WS, Tokino T, Velculescu VE, Levy DB, Parsons R, Trent JM *et al.* WAF1, a potential mediator of p53 tumor suppression. *Cell* 1993; **75**: 817–825.
33 Muda M, Boschert U, Dickinson R, Martinou JC, Martinou I, Camps M *et al.* MKP-3, a novel cytosolic protein-tyrosine phosphatase that exemplifies a new class of mitogen-activated protein kinase phosphatase. *J Biol Chem* 1996; **271**: 4319–4326.
34 Liston P, Roy N, Tamai K, Lefebvre C, Baird S, Cherton-Horvat G *et al.* Suppression of apoptosis in mammalian cells by NAIP and a related family of IAP genes. *Nature* 1996; **379**: 349–353.
35 Iida S, Rao PH, Butler M, Corradini P, Boccadoro M, Klein B *et al.* Deregulation of MUM1/IRF4 by chromosomal translocation in multiple myeloma. *Nat Genet* 1997; **17**: 226–230.
36 Woodage T, Basrai MA, Baxevanis AD, Hieter P, Collins FS. Characterization of the CHD family of proteins. *Proc Natl Acad Sci USA* 1997; **94**: 11472–11477.

Int. J. Cancer: **108,** 237–242 (2004)

Publication of the International Union Against Cancer

MUTATIONS OF *BRAF* ARE ASSOCIATED WITH EXTENSIVE *hMLH1* PROMOTER METHYLATION IN SPORADIC COLORECTAL CARCINOMAS

Koji KOINUMA[1,2], Kazuhisa SHITOH[1], Yasuyuki MIYAKURA[1], Taiji FURUKAWA[1], Yoshihiro YAMASHITA[2], Jun OTA[2], Ruri OHKI[2], Young Lim CHOI[2], Tomoaki WADA[2], Fumio KONISHI[3], Hideo NAGAI[1] and Hiroyuki MANO[2,4]*

[1]*Department of Surgery, Jichi Medical School, Tochigi, Japan*
[2]*Division of Functional Genomics, Jichi Medical School, Tochigi, Japan*
[3]*Department of Surgery, Omiya Medical Center, Jichi Medical School, Saitama, Japan*
[4]*Crest, Japan Science and Technology Agency, Saitama, Japan*

Activating mutations of *BRAF* have been frequently observed in microsatellite unstable (MSI^+) colorectal carcinomas (CRCs), in which mutations of *BRAF* and *KRAS* are mutually exclusive. Previously, we reported that hypermethylation of *hMLH1* might play an important role in the tumorigenesis of right-sided sporadic CRCs with MSI showing less frequency of *KRAS/TP53* alteration. Therefore, we have assumed that *BRAF* mutations might be highly associated with *hMLH1* methylation status rather than MSI status. In this study, mutations of *BRAF* and *KRAS* and their relationship with MSI and *hMLH1* methylation status were examined in 140 resected specimens of CRC. The methylation status was classified into 3 types: full methylation (FM), partial methylation (PM) and nonmethylation (NM). Only FM closely linked to reduced expression of hMLH1 protein. *BRAF* mutations were found in 16 cases (11%), all leading to the production of $BRAF^{V599E}$. As for MSI status, *BRAF* mutations were found in 43% of MSI^+ and 4% of MSI^- cases ($p < 0.0001$). Among the MSI^+ individuals, *BRAF* mutations were more frequent in cases with hMLH1 deficiency (58%) than those with hMSH2 deficiency (0%; $p = 0.02$). Moreover, they were found in 69% of FM, 4% of PM and 4% of NM, revealing a striking difference between FM and the other 2 groups (FM *vs.* PM or NM; $p < 0.0001$). These findings suggest that BRAF activation may participate in the carcinogenesis of sporadic CRCs with *hMLH1* hypermethylation in the proximal colon, independently of KRAS activation.

Key words: *sporadic colorectal cancer;* BRAF *mutation;* hMLH1 *hypermethylation; microsatellite instability*

In the development of colorectal cancer (CRC), it is now widely accepted that some forms of genetic instability lead to the sequential accumulation of genetic alterations and consequently develop carcinomas.[1] RAS activation in the MAP kinase cascade is supposed to constitute a part of the primary events in colorectal carcinogenesis, and the *KRAS* gene mutations have been found in about 30–40% cases of sporadic CRCs.[2–4]

Recently, activating *BRAF* mutations have been found almost invariably in melanoma cells and sometimes in other types of carcinoma, including CRCs,[5–7] implying a function of *BRAF* as a protooncogene. The *RAF* genes are members of MAPK pathway, encoding serine/threonine kinases that integrate the upstream input signals.[8,9] Once recruited at the cell membrane by GTP-loaded RAS, RAF becomes activated and subsequently phosphorylates the downstream kinases, MEKs, which eventually induce transcriptional activation of the target genes.[9]

More recently, frequent *BRAF* mutations and infrequent *KRAS* mutations have been reported in DNA-mismatch repair (MMR)-deficient CRCs.[10] Inactivation of MMR genes incurs instability of genomic microsatellite sequence (microsatellite instability, or MSI), which is found in the majority of patients with hereditary nonpolyposis colorectal cancer syndrome (HNPCC) and in 10–15% of cases of sporadic CRCs.[11–13] Moreover, it was also reported that 70–90% of sporadic CRCs with MSI (MSI^+ CRCs) are associated with hypermethylation of *hMLH1*, one of DNA-MMR genes, and have distinct clinical and pathologic characteristics, *i.e.*, occurrence in older females, location in the proximal colon and histopathology of mucious or poor differentiation.[14–20]

We have previously examined the methylation status of *hMLH1* gene in sporadic CRCs by use of 5 sets of primer spanning the whole CpG sites within its promoter region and have classified the methylation status into 3 subtypes: full methylation, partial methylation and nonmethylation.[21,22] We reported that an extensive methylation, or full methylation, of *hMLH1* promoter was found in about 80% of MSI^+ CRC cases and was highly associated with loss of expression of its gene product. Interestingly, this type of CRC cells are rarely associated with *KRAS* mutations and loss of heterozygosity (LOH) of *TP53* gene.[22] It is therefore possible that extensive methylation of *hMLH1* promoter region may contribute to the carcinogenesis of the right-sided sporadic CRCs, independently of *KRAS/p53* alterations.

From these results, 2 questions may arise. First, does the activation of BRAF, instead of KRAS, take part in the carcinogenesis of CRCs with extensive *hMLH1* methylation? Second, if so, does the BRAF activation have any relationship with the CRCs with partial methylation, although most of which are microsatellite stable (MSI^-), maintain MMR gene expression and show a relatively high incidence of KRAS and p53 alterations?[22]

Additionally, in the melanoma cells, high frequency of mutations of *β-catenin* and *BRAF* has been recognized.[23] Some researchers previously reported that *β-catenin* mutations were more common in MSI^+ CRCs than in MSI^- ones.[19,23–25] However, it has not been elucidated yet whether there are any relationship between the mutations of *β-catenin* and *BRAF* in the CRCs with *hMLH1* hypermethylation.

In this study, we have investigated the frequency of *BRAF* mutation and its relationship with *KRAS* and *β-catenin* mutations in a large consecutive series of sporadic CRCs in regard to both MSI status and degrees of *hMLH1* methylation.

Abbreviations: CRC, colorectal cancer; FM, full methylation; HNPCC, hereditary nonpolyposis colorectal cancer syndrome; LOH, loss of heterozygosity; MAPK, mitogen-activated protein kinase; MAPKKK, mitogen-activated protein kinase kinase kinases; MEK, mitogen-activated protein/extracellular signal-regulated kinase kinase; MGMT, O^6-methylguanine DNA methyltransferase; MMR, mismatch repair; MSI, microsatellite instability; NM, nonmethylation; PM, partial methylation.

*Correspondence to: Division of Functional Genomics, Jichi Medical School, 3311-1 Yakushiji, Minamikawachi-machi, Kawachi-gun, Tochigi 329-0498, Japan. Fax: +81-285-44-7322. E-mail: hmano@jichi.ac.jp

Received 23 May 2003; Revised 4 August 2003; Accepted 25 August 2003

DOI 10.1002/ijc.11523
Published online 27 October 2003 in Wiley InterScience (www.interscience.wiley.com).

MATERIAL AND METHODS

Tumor samples

Tumor samples were obtained from 140 sporadic CRC patients who underwent surgical treatment at the Jichi Medical School Hospital. None of the patients had first-degree relatives with CRC. Informed consents were obtained from all patients, and the ethics committee of the Jichi Medical School approved this study (#02-01). We selected these cases from approximately 380 consecutive series of CRCs previously analyzed for MSI status.[21,26] All the MSI^+ cases were reconfirmed for the MSI status by pentaplex PCR method, whereas MSI^- CRCs were selected so that the gender and tumor site were balanced between the MSI^+ and MSI^- groups (MSI^+, $n = 28$; MSI^-, $n = 112$). The patients were 69 men and 71 women, and their age ranged from 19 to 86 years with a mean of 63 years.

DNA extraction

Genomic DNA was extracted from fresh-frozen samples of tumor by use of QIAamp DNA Mini Kit (Qiagen, Chatsworth, CA) according to the manufacturer's protocol.

BRAF *mutation analysis*

BRAF mutations were analyzed in exons 11 and 15. These exons were chosen because all reported *BRAF* mutations occurred at these regions. PCR was performed with 2–5 ng of genomic DNA as a template by using the same PCR primer as reported previously.[5] PCR condition was as follows: 94°C for 9 min, followed by 35 cycles of 94°C for 1 min, 60°C for 1 min and 72°C for 2 min. The PCR products were purified using a QIAquick spin purification kit (Qiagen), and the purified PCR products were sequenced with BigDye Terminator Cycle Sequencing Ready Reaction kits (PE Applied Biosystems, Foster City, CA), all according to the manufacturers' instructions. Sequencing was performed in both directions using forward and reverse PCR primers. The purified products were run on an ABI 310 PRISM Genetic Analyzer (PE Applied Biosystems). The data were collected and analyzed using the Applied Biosystems sequencing analysis software.

MSI status analysis

MSI was analyzed by using 9 microsatellite repeat loci (3 markers were dinucleotide repeats and 6 were mononucleotide repeats) as described previously.[21] MSI status was stratified as follows according to the criteria of the National Cancer Institute (NCI) workshop.[27] High-frequency MSI (MSI-H) was defined as the alterations of microsatellite repeat were found in more than 40% of examined markers or in 2 or more NCI-recommended markers. Low-frequency MSI (MSI-L) was defined as the alterations in less than 40% or only one NCI-recommended marker. If no alterations of any examined markers were found, tumors were defined as microsatellite stable (MSS). In this study, we defined MSI-H as MSI-positive (MSI^+), and both MSI-L and MSS as MSI-negative (MSI^-), because only the MSI-H phenotype in sporadic CRCs is associated with true MMR defects and distinctive clinicopathologic features.[28,29] For the precision of MSI status, we reexamined all MSI^+ samples by pentaplex PCR method using 5 quasimonomorphic mononucleotide repeats, because this method has been reported to be simpler to use and show higher sensitivity and specificity.[30]

Analysis of methylation status of hMLH1 *promoter region*

Analysis of methylation status of *hMLH1* gene was performed by Na-bisulfite treatment and PCR single-strand conformation polymorphism (SSCP) analysis (BiPS) as described previously.[21] In brief, 5 sets of primers comprising the whole CpG sites within the *hMLH1* promoter region were prepared (Fig. 1), and methylated and unmethylated DNA amplicons were separated through SSCP analysis. When the bands showed mobility shifts, they were cut from the gels and subsequently sequenced directly by use of an ABI 310 PRISM Genetic Analyzer. Primer sequences and PCR conditions were utilized as reported previously.[21] The methylation patterns were defined as full methylation if all the CpG sites within the promoter regions showed methylation; as partial methylation if some CpG sites in the upstream region showed methylation; and as nonmethylation if no CpG sites in the region showed methylation.

KRAS *mutation analysis*

KRAS mutations were analyzed by direct sequencing at codons 12 and 13 of *KRAS* by using its genomic DNA. First, a flanking PCR product of 179 bp was amplified (annealing temperature was 58°C) using the primers 5′-AGGCCTGCTGAAAATGACTGAATA-3′ (sense) and 5′-CTGTATCAAAGAATGGTCCTGCAC-3′ (antisense). The resulting fragment was then used as a template to amplify a 114 bp fragment, including codons 12 and 13 using the primers 5′-AAAATGACTGAATATAAACTTGTGG-3′ (sense) and 5′-CTCTATTGTTGGATCATATTCGTC-3′ (antisense; annealing temperature was 50°C). The PCR product was sequenced by the same method as in the *BRAF* mutation analysis.

β-catenin *mutation analysis*

Mutations in *β-catenin* were analyzed by direct sequencing at its exon 3, in which the majority of mutation hot spots were included. The PCR primers were 5′-GATTTGATGGAGTTGGACATGG-3′ (sense) and 5′-TGTTCTTGAGTGAAGGACTGAG-3′ (antisense; annealing temperature was 63°C). The direct sequencing of the PCR product was performed by the same method as in the *BRAF* mutation analysis.

Immunohistochemical analysis

Immunohistochemical analysis for both hMLH1 and hMSH2 expression was performed on all MSI^+ tumor samples as described previously.[26]

Statistical analysis

Statistical analyses for variable results were performed by Fisher's exact test and Student's *t*-test. Probability values below 0.05 were considered to be statistically significant (StatView J 5.0 software, Abacus Concepts, Berkeley, CA).

RESULTS

Clinicopathologic features of patients with BRAF *mutations*

We identified 16 patients whose CRCs showed *BRAF* mutations (Table I). All the mutations resulted in a V599E substitution in the

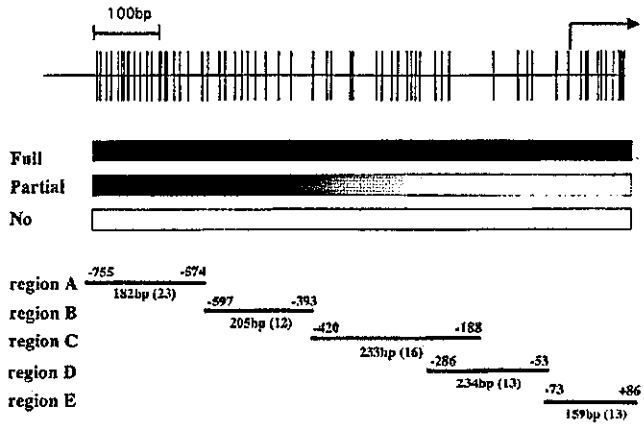

FIGURE 1 – Top: Schematic presentation of the *MLH1* promoter region. Middle: Full methylation (all CpG sites in regions A–E show methylation), partial methylation (some CpG sites in upstream region show methylation) and nonmethylation (no CpG site shows methylation). Bottom: Design of the PCR primers and the PCR products for regions A–E. Their positions relative to the adenine residue at the start codon and the size of the amplified DNA fragments are shown. Numbers in parenthesis indicate the number of CpG sites within each region.

BRAF protein (Fig. 2). None of the cases had *BRAF* mutations in the normal colonic mucosa positioned far away from the cancer area, implying that the *BRAF* mutations should be a somatic event. The mean age of cancer onset in the patients with *BRAF* mutations was older than those without *BRAF* mutations, although the difference was not statistically significant: 73.1 ± 10.5 years compared with 62.5 ± 12.5 years ($p = 0.06$, Student's *t*-test; Table II). Gender distribution was also different between these 2 groups, with females comprising 75% (12/16) of the *BRAF* mutation group and 48% (59/124) of the nonmutation group ($p = 0.06$, Fisher's exact test; Table II). The tumor with *BRAF* mutation cases was more frequently located in the proximal colon (94%; 15/16) than that with nonmutation ones (37%; 46/124; $p < 0.0001$, Fisher's exact test; Table II).

BRAF *mutations and MSI status*

BRAF mutations were found in 43% (12/28) of MSI^+ CRCs and 4% (4/112) of MSI^- CRCs ($p < 0.0001$, Fisher's exact test; Table III).

BRAF *mutations and MMR protein expression*

BRAF mutations were more common in the tumors showing reduced hMLH1 protein expression (58%; 11/19) than those showing reduced hMSH2 expression (0%; 0/6; $p = 0.02$, Fisher's exact test; Table IV).

BRAF *mutations and* hMLH1 *promoter methylation status*

BRAF mutations were found in 69% (11/16) of full methylation, 4% (2/45) of partial methylation and 4% (3/79) of nonmethylation (Table V). The ratio of *BRAF* mutations was statistically significant between full and partial as well as between full and none ($p < 0.0001$, Fisher's exact test).

KRAS *mutations*

KRAS mutations were identified in 38 cases. Two cases with *KRAS* mutations were in MSI^+ (7%; 2/28) and 36 cases were in MSI^- (32%; 36/112; $p = 0.008$, Fisher's exact test; Table III). Regarding the methylation status, *KRAS* mutations were not found

TABLE I – ALL CRC CASES WITH *BRAF* MUTATIONS

Patient no.	Age (yr)	Gender	Site	MSI	*hMLH1* methylation	*BRAF* amino acid	*KRAS*	β-*catenin*	Dukes' stage	Histologic grade
225	83	F	P	+	Full	V599E	Wild	Wild	C	Well
263	86	F	P	+	Full	V599E	Wild	Wild	B	Moderate
268	85	F	P	+	Full	V599E	Wild	Wild	B	Poor
280	83	F	P	+	Full	V599E	Wild	Wild	C	Well
305	74	M	P	+	Full	V599E	Wild	Wild	B	Poor
318	76	F	P	+	Full	V599E	Wild	Wild	B	Well
336	68	M	P	+	Full	V599E	Wild	Wild	B	Mucinous
413	69	F	P	+	Full	V599E	Wild	Wild	B	Well
416	76	F	P	+	Full	V599E	Wild	Wild	B	Mucinous
479	74	F	P	+	Full	V599E	Wild	Wild	B	Moderate
507	64	M	P	+	Full	V599E	Wild	Wild	A	Moderate
274	81	M	D	−	Partial	V599E	Wild	Wild	B	Moderate
328	52	F	P	−	Partial	V599E	Wild	Wild	B	Moderate
293	70	F	P	−	Non	V599E	Wild	Wild	D	Mucinous
384	77	F	P	+	Non	V599E	Wild	Wild	A	Poor
509	51	F	P	−	Non	V599E	Wild	Wild	C	Poor

P, proximal colon; D, distal colon; Full, full methylation; Partial, partial methylation; Non, nonmethylation; Well, well-differentiated adenocarcinoma; Mod, moderately differentiated adenocarcinoma; Poor, poorly differentiated adenocarcinoma; Muc, mucinous carcinoma.

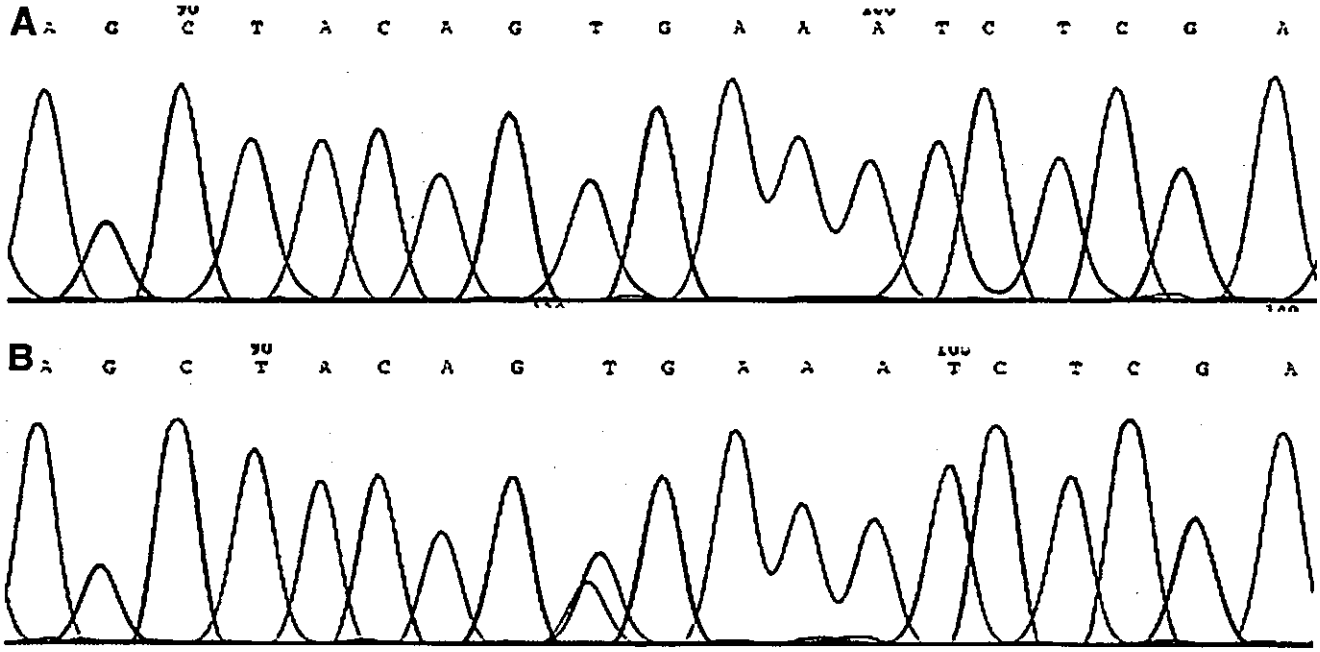

FIGURE 2 – (*a*) Representative sequence chromatographs from *BRAF* exon 15 showing wild type. (*b*) T1796A transversion resulting in a V599E substitution.

TABLE II – CLINICOPATHOLOGIC FEATURES OF *BRAF* MUTATION CASES

	BRAF mutation (n = 16) (%)	Non-*BRAF* mutation (n = 124) (%)	P-value
Age (yr)	73.1 ± 10.5	62.5 ± 12.5	0.06
Gender			0.06
Male	4 (25)	65 (52)	
Female	12 (75)	59 (48)	
Tumor site			< 0.0001
Proximal	15 (94)	46 (37)	
Distal	1 (6)	78 (63)	
Dukes' stage			0.07
A/B	12 (75)	59 (50)	
C/D	4 (25)	59 (50)	
Histologic grade			0.0008
Well/moderate	9 (56)	109 (92)	
Poor/mucinous	7 (44)	10 (8)	

Well, well-differentiated adenocarcinoma; moderate, moderately differentiated adenocarcinoma; Poor, poorly differentiated adenocarcinoma; mucinous, mucinous carcinoma.

TABLE III – MSI STATUS AND MUTATIONS OF *BRAF*, *KRAS* AND β-*CATENIN*

MSI status	*BRAF* mutation (%)	*KRAS* mutation (%)	β-*catenin* mutation (%)
MSI^+	42.9 (12/28)[a]	7.1 (2/28)[b]	7.1 (2/28)
MSI^-	3.6 (4/112)[a]	32.1 (36/112)[b]	2.7 (3/112)
Total	11.4 (16/140)	27.1 (38/140)	3.6 (5/140)

[a]Fisher's exact test, $p < 0.0001$.–[b]Fisher's exact test, $p = 0.008$.

TABLE IV – *BRAF* MUTATION AND MMR PROTEIN EXPRESSION

	Number of cases	*BRAF* mutation (%)
hMLH1-deficient	19	57.9 (11/19)[a]
hMSH2-deficient	6	0 (0/6)[a]

[a]Fisher's exact test, $p = 0.02$.

TABLE V – *hMLH1* METHYLATION STATUS AND MUTATIONS OF *BRAF*, *KRAS* AND β-*CATENIN*

hMLH1 methylation status	*BRAF* mutation (%)	*KRAS* mutation (%)	β-*catenin* mutation (%)
Full methylation	68.8 (11/16)[a,b]	0 (0/16)[c,d]	0 (0/16)
Partial methylation	4.4 (2/45)[a]	37.8 (17/45)[c]	0 (0/45)
Nonmethylation	3.8 (3/79)[b]	26.6 (21/79)[d]	6.3 (5/79)
Total	11.4 (16/140)	27.1 (38/140)	3.6 (5/140)

[a]Fisher's exact test, $P < 0.0001$.–[b]Fisher's exact test, $P < 0.0001$.–[c]Fisher's exact test, $P = 0.003$.–[d]Fisher's exact test, $P = 0.02$.

(0%; 0/16) in any of the full methylation cases, but were found in 17 of the partial methylation patients (38%; 17/45) and 21 of the nonmethylation group (26.6%; 21/79), which was consistent with our previous results[21] (Table V). The ratio of *KRAS* mutations was significantly different between full and partial and between full and none cases ($p = 0.003$ and 0.02, respectively, Fisher's exact test). None of the cases with *BRAF* mutations exhibited *KRAS* mutations simultaneously.

β-catenin *mutations*

β-catenin mutations were found in 5 cases. Two cases were in MSI^+ (7%; 2/28) and 3 cases were in MSI^- (3%; 3/112; $p = 0.26$, Fisher's exact test; Table III). However, none of the cases with full or partial methylation showed *β-catenin* mutations, and there were no cases exhibiting both *BRAF* and *β-catenin* mutations simultaneously (Table V).

DISCUSSIONS

The MAPK pathway plays a crucial role in the signal transduction of many hormones, growth factors and differentiation factors.[31,32] At the level of MAPKKKs, several RAF family members exist, that is, ARAF, BRAF and RAF1 with divergent tissue specificity and upstream regulation.[33] The 3 proteins are thought to have uneven ability to activate MEK, and BRAF has been identified as the major MEK activator.[32]

Recently, *BRAF* activating mutations were observed in some proportion of human carcinomas, especially in melanoma, lung cancer, as well as colon cancer.[5–7] *BRAF* gene was therefore supposed to be a novel protooncogene that might contribute to the tumorigenesis in these types of transformed cells. Interestingly, the mutational spots of *BRAF* gene cluster within the activation segment (exon 15) and the G-loop (exon 11) of the kinase domain, which are highly conserved among serine/threonine kinases throughout evolution.[5] Activating mutations in these hot spots are supposed to increase its kinase activity and subsequently urge to phosphorylate the downstream kinase, MEK. V599 is the major site of point mutations in the BRAF protein and V599E acidic substitution has been commonly found in melanoma, colon cancer and ovarian cancer cells.[5] Intriguingly, the tumors with $BRAF^{V599E}$ showed no *KRAS* mutations simultaneously, although non-V599E cases were sometimes coincident with *KRAS* mutations.[5,10,34] It has been hypothesized therefore that V599E might mimic the phosphorylation of T598 of BRAF that constitutes the natural activation mechanism of this protein. Because of its potent kinase activity, BRAF with this type of mutation might have no need to depend on RAS for the initiation of the MAP kinase pathway activation.

In our series of sporadic CRCs, 28 cases showed MSI^+, in which 19 were with hMLH1 deficiency and 6 were with hMSH2 deficiency. *BRAF* mutations were more frequent in MSI^+ CRCs than in MSI^- CRCs (43%, 12/28 *vs.* 4%, 4/112; $p < 0.0001$). This result was nearly consistent with that in the previous report.[10] Interestingly, *BRAF* mutations were more frequent in hMLH1-deficient cases than hMSH2-deficient ones (58%, 11/19 *vs.* 0%, 0/6; $p = 0.02$). It has been widely accepted that MSI in sporadic CRCs commonly results from epigenetic silencing of *hMLH1* gene, secondary to its promoter methylation, and 70–90% of MSI^+ CRCs indeed show hypermethylation of the *hMLH1* gene.[15,21,35] Moreover, extensive methylation of *hMLH1* promoter is closely correlated with hMLH1 inactivation.[21] Therefore, we have examined the frequency of *BRAF* mutations with regard to the methylation status of *hMLH1* promoter region. Amazingly, *BRAF* mutations were extremely frequent in the cases with full methylation compared to those without full methylation (69%, 11/16 *vs.* 4%, 5/124; $p < 0.0001$). As generally seen in the cases with *hMLH1* methylation, the CRCs with *BRAF* mutations were more frequent in older females, commonly located in the proximal colon, and showed the histopathology of mucinous or poor differentiation. Our data suggest that the activating mutation of *BRAF* may be highly associated with an extensive methylation of *hMLH1* gene.

In a recent study, we proposed that the shift of methylation status from partial to full might be critical in the tumorigenesis of right-sided sporadic CRCs with MSI, because more than half of the cases with full methylation showed partial methylation in their normal mucosa far from the tumor.[21] However, the cancers with partial methylation not yet reaching full methylation showed distinct clinical and biologic features from those with full methylation, with relatively high frequency in the alterations of *KRAS*/*p53*.[22] In this study, *BRAF* mutations were less frequent in the cases with partial methylation compared to those with full methylation (4%, 2/45 *vs.* 69%, 11/16; $p < 0.0001$). We state that partial methylation is not generally the true pathogenic methylation status of *hMLH1* gene.

In our study, all the mutations of *BRAF* resulted in V599E substitutions (T-to-A transversion at nucleotide 1796). Rajago-

palan *et al.*[10] reported that all but one of the 15 *BRAF* mutations in MMR-deficient cases resulted in V599E. O^6-methylguanine DNA methyltransferase (MGMT) is a DNA repair protein and MGMT epigenetic inactivation by its promoter hypermethylation is supposed to cause G-to-A transition mutation in *KRAS* and G:C-to-A:T transition mutation in *p53*. Indeed, 71% cases with *KRAS* and *p53* mutations showed hypermethylation of *MGMT*.[36,37] Therefore, it might be possible that inactivation of an anonymous DNA-repair gene by promoter hypermethylation has an association with A-to-T transition mutation in *BRAF* gene.

Yuen *et al.*[34] have reported that there are many similarities between the phenotypic patterns of CRCs with *KRAS* and *BRAF* mutations. However, they showed that the cases with *BRAF* mutations differ from those with *KRAS* mutations in the Dukes' stage. Consistent with their results, the cases with *BRAF* mutations in our study were more common with Dukes' A/B grades than with Dukes' C/D, although the difference was not statistically significant ($p = 0.07$, Fisher's exact test). Moreover, the patients with *BRAF* mutations were approximately 13 years older than those with *KRAS* mutations (data not shown). Therefore, we speculate that the CRCs with *BRAF* mutations may belong to a clinical entity distinct from one of CRCs with *KRAS* mutations.

In melanoma cells as well as MSI^+ CRC cells, high frequency of *β-catenin* and *BRAF* mutations have been reported.[19,23–25] In this study, *β-catenin* mutations were uncommon in both MSI^+ (7%; 2/28) and MSI^- (3%; 3/112), as reported previously by Jass *et al.*[29] Moreover, none of our cases with full or partial methylation showed *β-catenin* mutations. There were no cases harboring *BRAF* and *β-catenin* mutations simultaneously, implying that *β-catenin* mutation may have no association with *hMLH1* hypermethylation with regard to CRC carcinogenesis.

We previously proposed that extensive methylation of *hMLH1* promoter might play a crucial role in tumorigenesis in the proximal colon.[21] In this study, we additionally demonstrated that the activating mutations of *BRAF* might take part in the carcinogenesis of sporadic CRCs with *hMLH1* hypermethylation in the proximal colon, independently of KRAS activation. However, one question remains to be addressed. At which stage does BRAF activation contribute to malignant transformation of colon epithelial cells? Hyperplastic polyps and serrated adenomas in the right-sided colon show high frequency of *hMLH1* hypermethylation, and these lesions have been presumed to be premalignant lesions of right-sided CRCs with MSI.[29] It would be interesting to examine whether the majority of such hyperplastic polyps and serrated adenomas already have *BRAF* mutations. If most of the CRCs with extensive methylation are associated with $BRAF^{V599E}$, such subtype would be a good target for novel anticancer drugs acting on the MAPK pathway.[38,39]

ACKNOWLEDGEMENTS

The authors thank the staff of the Division of Functional Genomics, Jichi Medical School, for their helpful advice.

REFERENCES

1. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. Cell 1996;87:159–70.
2. Aaltonen LA, Peltomaki P, Leach FS, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, Powell SM, Jen J, Hamilton SR, Petersen GM, Kinzler KW, et al. Clues to the pathogenesis of familial colorectal cancer. Science 1993;260:812–6.
3. Bos JL, Fearon ER, Hamilton SR, Verlaan-de Vries M, van Boom JH, van der Eb AJ, Vogelstein B. Prevalence of ras gene mutations in human colorectal cancers. Nature 1987;327:293–7.
4. Forrester K, Almoguera C, Han K, Grizzle WE, Perucho M. Detection of high incidence of K-ras oncogenes during human colon tumorigenesis. Nature 1987;327:298–303.
5. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, et al. Mutations of the BRAF gene in human cancer. Nature 2002;417:949–54.
6. Pollock PM, Meltzer PS. A genome-based strategy uncovers frequent BRAF mutations in melanoma. Cancer Cell 2002;2:5–7.
7. Brose MS, Volpe P, Feldman M, Kumar M, Rishi I, Gerrero R, Einhorn E, Herlyn M, Minna J, Nicholson A, Roth JA, Albelda SM, et al. BRAF and RAS mutations in human lung cancer and melanoma. Cancer Res 2002;62:6997–7000.
8. Vojtek AB, Der CJ. Increasing complexity of the Ras signaling pathway. J Biol Chem 1998;273:19925–8.
9. Magee T, Marshall C. New insights into the interaction of Ras with the plasma membrane. Cell 1999;98:9–12.
10. Rajagopalan H, Bardelli A, Lengauer C, Kinzler KW, Vogelstein B, Velculescu VE. Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. Nature 2002;418:934.
11. Aaltonen LA, Peltomaki P, Mecklin JP, Jarvinen H, Jass JR, Green JS, Lynch HT, Watson P, Tallqvist G, Juhola M. Replication errors in benign and malignant tumors from hereditary nonpolyposis colorectal cancer patients. Cancer Res 1994;54:1645–8.
12. Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 1993; 75:1027–38.
13. Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, Kane M, Earabino C, Lipford J, Lindblom A, Tannergard P, Bollag RJ, et al. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature 1994;368:258–61.
14. Kane MF, Loda M, Gaida GM, Lipman J, Mishra R, Goldman H, Jessup JM, Kolodner R. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. Cancer Res 1997; 57:808–11.
15. Cunningham JM, Christensen ER, Tester DJ, Kim CY, Roche PC, Burgart LJ, Thibodeau SN. Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. Cancer Res 1998;58:3455–60.
16. Veigl ML, Kasturi L, Olechnowicz J, Ma AH, Lutterbaugh JD, Periyasamy S, Li GM, Drummond J, Modrich PL, Sedwick WD, Markowitz SD. Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. Proc Natl Acad Sci USA 1998;95:8698–702.
17. Kuismanen SA, Holmberg MT, Salovaara R, Schweizer P, Aaltonen LA, de La Chapelle A, Nystrom-Lahti M, Peltomaki P. Epigenetic phenotypes distinguish microsatellite-stable and -unstable colorectal cancers. Proc Natl Acad Sci USA 1999;96:12661–6.
18. Malkhosyan SR, Yamamoto H, Piao Z, Perucho M. Late onset and high incidence of colon cancer of the mutator phenotype with hypermethylated hMLH1 gene in women. Gastroenterology 2000;119:598.
19. Young J, Simms LA, Biden KG, Wynter C, Whitehall V, Karamatic R, George J, Goldblatt J, Walpole I, Robin SA, Borten MM, Stitz R, et al. Features of colorectal cancers with high-level microsatellite instability occurring in familial and sporadic settings: parallel pathways of tumorigenesis. Am J Pathol 2001;159:2107–16.
20. Hawkins N, Norrie M, Cheong K, Mokany E, Ku SL, Meagher A, O'Connor T, Ward R. CpG island methylation in sporadic colorectal cancers and its relationship to microsatellite instability. Gastroenterology 2002;122:1376–87.
21. Miyakura Y, Sugano K, Konishi F, Ichikawa A, Maekawa M, Shitoh K, Igarashi S, Kotake K, Koyama Y, Nagai H. Extensive methylation of hMLH1 promoter region predominates in proximal colon cancer with microsatellite instability. Gastroenterology 2001;121:1300–9.
22. Miyakura Y, Sugano K, Konishi F, Fukayama N, Igarashi S, Kotake K, Matsui T, Koyama Y, Maekawa M, Nagai H. Methylation profile of the MLH1 promoter region and their relationship to colorectal carcinogenesis. Genes Chromosomes Cancer 2003;36:17–25.
23. Sparks AB, Morin PJ, Vogelstein B, Kinzler KW. Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. Cancer Res 1998;58:1130–4.
24. Shitoh K, Furukawa T, Kojima M, Konishi F, Miyaki M, Tsukamoto T, Nagai H. Frequent activation of the beta-catenin-Tcf signaling pathway in nonfamilial colorectal carcinomas with microsatellite instability. Genes Chromosomes Cancer 2001;30:32–7.
25. Mirabelli-Primdahl L, Gryfe R, Kim H, Millar A, Luceri C, Dale D, Holowaty E, Bapat B, Gallinger S, Redston M. Beta-catenin mutations are specific for colorectal carcinomas with microsatellite instability but occur in endometrial carcinomas irrespective of mutator pathway. Cancer Res 1999;59:3346–51.
26. Furukawa T, Konishi F, Masubuchi S, Shitoh K, Nagai H, Tsukamoto T. Densely methylated MLH1 promoter correlates with decreased mRNA expression in sporadic colorectal cancers. Genes Chromosomes Cancer 2002;35:1–10.