提供いただく試料腎組織の採取は、手術で切除された腎臓を処理して使用します。手術そのものは試料の提供の有無にかかわらず最善と判断される方法で常に行いますので、試料提供により加わる危険性は全くありません。試料をいただきましたら個人情報と完全に切り離して解析されますので、解析結果がいかなるものであれ個人情報とリンクしてご提供いただいた方にとって不利益となる可能性はありませんが、解析結果を今後の患者さんご本人の治療へ直接結びつけて役に立てることもできません。研究データが蓄積し、腎臓に障害を引き起こしうる化学物質を早期に発見することができるようになれば、これらのデータが社会に還元され、多くの方の役に立つと言う意味では間接的には患者さんへの利益になると言えます。

(7)　研究計画などを見たいとき

希望があれば、個人情報の保護や研究の独創性の確保に支障を来さない範囲内で、この研究計画の内容を見ることができます。また、遺伝子を調べる方法等に関する資料が必要な場合も用意いたします。

(8)　個人情報の保護

遺伝子解析の結果は、いろいろな問題を引き起こす可能性があるために、他人に漏れないように取扱いを慎重にしています。解析を開始する前に、あなたの腎組織や診療情報からは住所、名前等が削られ、代わりに新しい符号がつけられます。これを匿名化といいます。

あなたとこの符号とを結びつける対応表は、本研究では作成されません。これを連結不可能匿名化といいます。

(9)　試料等又はそれから得られた遺伝情報を他の機関へ提供する可能性

厚生労働省(国立医薬品衛生研究所)、毒性部室長　菅野純　ら　との共同研究となる可能性があります。本研究と並行して行われる、動物や細胞株などを使用した基礎研究は厚生労働省の研究所などで行われ、我々の研究と照らし合わせて、それらは完成した研究成果となります。この場合も、遺伝子発現の情報は、どの患者さん由来の細胞における遺伝子発現であるかは完全に判らない形で情報が提供されます。

(10) 遺伝子解析結果の伝え方

この研究では、どの解析結果がどの患者さんの腎臓由来の情報であったか完全にわからない形で行われますので,解析結果を個々の患者さんやご家族に伝えることは不可能です.

(11) 知的財産権が生じたとき

遺伝子解析の結果として特許権などが生じる可能性がありますが、その権利は、大学や研究者等に属し、あなたには属しません。また、その特許権などにより経済的利益が生じる可能性がありますが、あなたはこれについても権利がありません。

(12) 研究結果の公表

ご協力によって得られた結果は、個人が誰であるか分からないようにした上で、学会や学術雑誌、データベース上などで公に発表されることがあります。また,研究資金の提供者である政府や民間の研究助成団体との契約に従い,それぞれの資金提供者へ報告書を作成提出することになります.この場合も研究へ参加した個人が特定できない様にした上で報告書が作成されます.

(13) 試料等の保存、使用及び廃棄の方法

提供いただいた試料である非罹患部腎臓組織およびこれより抽出いたしました核酸は、自治医科大学臨床薬理学において厳重に保管し、本研究のために使用されます。一方,病理組織は通常の手術標本として自治医科大学附属病院病理診断部で臨床診断のために処理され,通常の方法で保管されます.非罹患部の組織に関しまして,もし、あなたが同意してくだされば、将来の研究のための貴重な資源として研究終了後も保管させていただきます。この場合も、(8)で説明した方法により、誰の試料か分からないようにしたまま、試料を使い切るまで保管します。試料を廃棄する場合は、匿名のまま密封容器に廃棄するか又は焼却処分します。将来、試料を医学研究に用いる場合には、改めて研究計画書を提出し、自治医科大学生命倫理委員会等の承認を受けます。

(14) 試料等をヒト細胞・遺伝子・組織バンクに提供し、一般的に研究資源として分譲する可能性

試料等をヒト細胞・遺伝子・組織バンクに提供し、一般的に研究資源として分譲する予定はありません

(15) 遺伝カウンセリングの利用

本研究に関連した遺伝子解析結果と関連した遺伝カウンセリングは行いません. 疑問点は担当医にご相談ください.

(16) 試料等の提供は無償・無報酬

遺伝子解析は研究費によって行なわれますので、あなたが費用を負担することはありません。また、この研究への協力に対して、あなたへの報酬は支払われません。

この研究の費用は、厚生労働省など公的機関からの研究費や民間の研究助成金などによっています。

(17) 問い合わせ、苦情の受付

この遺伝子解析研究についてのお問い合わせは、研究責任者までご連絡下さい。苦情がある場合は、自治医科大学大学事務部学事課(電話 0285-44-7044)で受け付けます。

平成　　年　　月　　日

研究責任者:自治医科大学[臨床薬理学講座・助手・大島康雄]

〒329-0498　栃木県河内郡南河内町薬師寺 3311-1　電話 0285-58-7388

遺伝子解析研究への協力についての同意書

自治医科大学学長　高久史麿　殿

私は、遺伝子解析研究:研究題目薬物による腎障害の予防に関する研究に関して、から説明文書を用いて説明を受け、その方法、危険性、分析結果のお知らせの方法等について十分理解しました。ついては、次の条件で研究に協力することに同意します。

説明を受け理解した項目(□の中にご自分でチェックの印を付けてください。)

- □ 遺伝子について
- □ 研究の協力は任意で協力しなくても不利益を受けないこと。同意の撤回も文書によって自由にできること。
- □ 研究の目的と方法
- □ 希望により研究計画書等を見ることができること。
- □ 試料等提供者にもたらされる利益と不利益
- □ 個人情報の保護の方法
- □ 遺伝子解析結果の説明の方針
- □ 研究結果の公表
- □ 研究から財産権が生じても試料等提供者には帰属しないこと。
- □ 研究終了後の試料等の取扱の方針
- □ 解析に関する費用負担は無く、試料等の提供に対する報酬の支払いも無いこと。
- □ 希望により遺伝カウンセリングが受けられること。

1　私は上記の項目のすべての□にチェックの印を記入した上で、私の提供する試料(腎組織)等が、本遺伝子解析研究に使用されることに同意します。

本人又は代諾者の署名又は記名・捺印

2　上記１で同意された方は、下記の２－１又は２－２のどちらかを選択し、番号を丸で囲み、署名又は記名・捺印してください。

2−1　提供する試料等を本研究のみに使用し、かつ本研究の終了時には速やかに破棄してください。

2−2　提供する試料等が本研究に使用されるとともに長期間保存され、将来新たに計画・実施される遺伝子の解析を含む医学研究に使用されることに同意します。

本人又は代諾者の署名又は記名・捺印

平成　　年　　月　　日

本人の氏名

住所

電話

本人又は代諾者の署名又は記名・捺印

代諾者の場合は本人との関係

代諾者の住所

電話

説明者の職名・氏名

説明者の署名又は記名・捺印

(患者が未成年の場合は本人と代諾者の2通必要です.患者が痴呆などのため代諾者が同意する場合は本人の同意書は必要ありません)

厚生労働科学研究費補助金(化学物質リスク研究事業)
分担研究報告書
バイオインフォマティクス

分担研究者　篠原　歩　東北大学大学院情報科学科　教授

研究要旨

本プロジェクトを遂行するためには、実験によって得られたデータを維持・管理し、新たな仮説を導き出すために、情報処理の最先端技術を取り入れながら強固かつ柔軟なデータベースシステムを構築することが必須である。本年度は、ヒトのすべての遺伝子を対象として、白血病の発症に関連の深いとされている AML1 binding motif である TGYGGT 配列をはじめとして、アフィメトリクス社のプローブセットの遺伝子の上流配列中の出現位置に関するデータを整理し、データベースとしてまとめあげた。そしてこれをウェブで公開し、インターネットを通じた検索サービスを開始した。

A. 研究目的

白血病の発症に関連の深いとされている AML1 ついて、ヒトのすべての遺伝子を対象として、遺伝子の上流部分に AML1 binding motif が出現する位置をまとめたデータベースを構築する。そしてこのデータベースをウェブで公開し、インターネットを通じて全世界から検索を行えるサービスを提供する。

B. 研究方法

今回の作業では、次のデータセットを用いた。

1. NCBI から取得したデータセット seq_gene.md。
2. アフィメトリクス社が提供している HG-U95Av2 遺伝子チップに関するファイル。これを各遺伝子とその遺伝子の存在する染色体，遺伝子の開始位置等の情報をマッピングするためのデータとして用いる。
3. バイオベース社のモチーフ群。バイオベース社の提供しているモチーフのデータは、あるタンパク質が特定の塩基配列に結合することがわかっているそのモチーフとタンパク質のセットのリストであり、ここに記載された 4,999 個のモチーフを用

いた。

まず初めにアフィメトリクス社の Gene Chip のプローブセット ID に対応するそれぞれの遺伝子を seq_gene.md から検索し，遺伝子の上流配列 1000 塩基を抜き出した。また CDS の開始位置と gene の開始位置の差を抜き出した。CDS の開始位置を 0 とみて，gene の相対的な開始位置を見ている。例えば CDS の開始位置が 2000 で gene の開始位置が 1200 ならば差は-800 ととる。この際、いくつかの例外処理により、HG-U95Av2 の中のすべてのプローブセットに対し、遺伝子を見つけられないものを例外処理して別に示した。例外処理は以下の通りである。

(1)まず HG-U95Av2 の中で遺伝子名と対応がとれないものを除外する。

(2) マッピングファイルで遺伝子名が見つからないものを除外する。

(3) マッピングファイルで複数のヒット箇所が見つかる場合、得られた複数の遺伝子の開始位置と遺伝子コーディング領域の開始位置との差を取り、その差が 10 を越えるものを除外する。

(4) 上流配列の場所が特定できても、その上流配列が得られないものは除外する。

この例外処理によって除外されなかったすべてのプローブセットに対して、その上流配列を 1,000 塩基分取り出して記録しておくことで、データベース検索を高速化している。

次に、与えられたプローブセットの中に有意に現れるモチーフの検出に関して、フィッシャーの正確な確率を用いて各モチーフに対する検定を行った。この実装においては、統計解析ツール R を用いてフィッシャーの正確な確率を計算している。検定のしかたは次の通りである。まず、入力として与えられたプローブセットの集合を P とする。次にバイオベース社の提供するモチーフ群を1つずつ抜き出し、そのモチーフが集合 P の中のプローブセットの上流配列に出現するかどうかを数え上げる。ただしその際、1つのプローブセットの上流配列に複数回出現しても1回とカウントする。こうして、そのモチーフが現れるかどうかに関するクロス集計表を求め、それにたしてフィッシャーの正確な確率を計算し、それが 5%を下回る場合に、そのモチーフは入力として与えられたプローブセット集合 P に対して有意に出現すると判断する。この操作をバイオベース社の提供するすべてのモチーフ群に対しておこない、有意なものをすべて出力する。なお、上流配列の長さに関しては、最大 1,000 までの範囲で、ユーザが任意に指定できるようにしている。

C. 研究結果

当初の目標通り、アフィメトリクス社のプローブセットの上流配列のテーブルセットを作成しデータベースを構築した。HG-U95Av2 のプローブセットの総数 44,760 個のうち、上記の例外処理によってそれぞれ次の個数が除外された。

(1) 遺伝子名と対応がとれないもの 9,855 個。

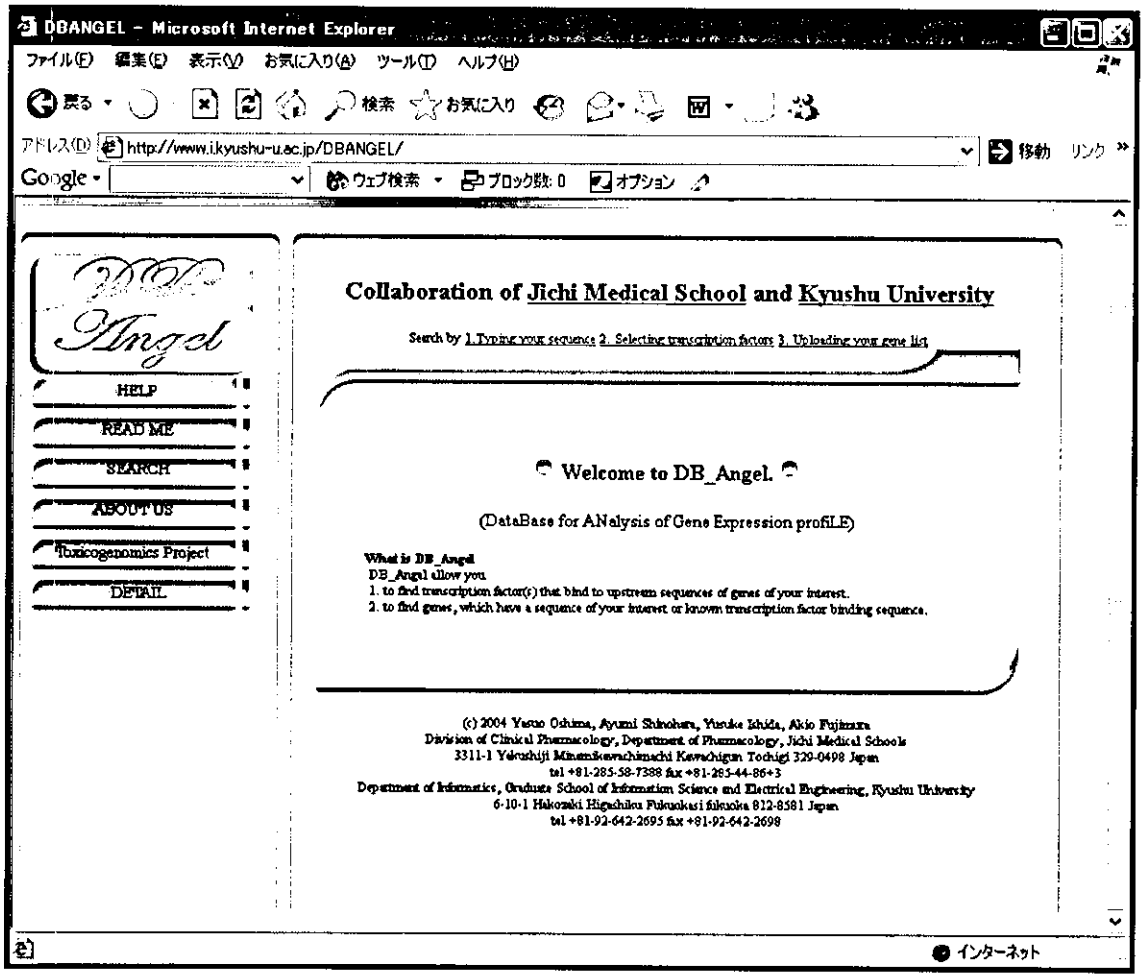
(2) マッピングファイルで遺伝子名が見つからないもの 3,084 個

(3) マッピングファイルで複数のヒット箇所が見つかるもののうち、遺伝子の開始位置と CDS の開始位置との差が 10 を越えるもの 559 個

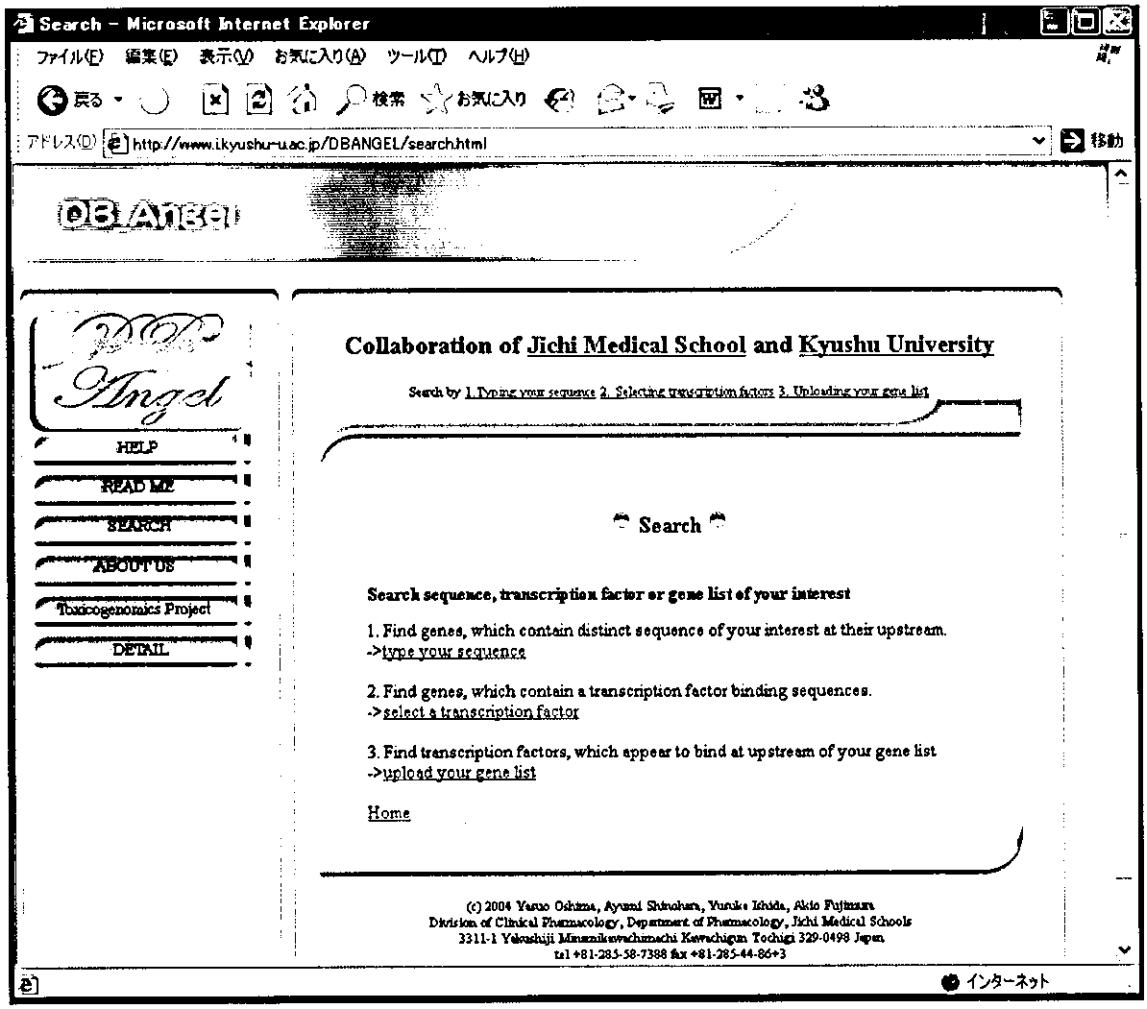
(4) マッピングファイルに遺伝子名は記載されているが上流配列を取れないもの 1個

この処理によって除外されずに残ったプローブセットは 31,281 個であった。なお、例外処理(3)において、複数ヒットするプローブセットは 1,510 個存在した。

こうして得られたものをウェブ上のデータベースとしてまとめあげ、インターネットを通じてどこからでもアクセスできるようにし、検索サービスを開始した。この検索サービスを DBANGLE として、http://www.i.kyushu-u.ac.jp/DBANGLE に公開している(図1、図2)。



DBANGLE



検索画面

D. 考察

利用した基礎データが必ずしも一様な形式をみたしていなかったため、データの変換プログラム作成の際に、例外的な処理を多用せざるを得ず、そのために当初の見積もりよりもシステムの開発に多くの時間を要した。しかしながら、これらの作業はすべてプログラムの中に明示的に記述しているため、今後データベースのバージョンアップなどの際にも、更新作業はそれほど困難なく行えるようになるはずである。このデータベースに対するユーザからの意見を採り入れて、インターフェース部を改良し、機能を向上させていく予定である。

E. 結論

各プローブセットの CDS 領域の上流部分に対して出現位置の範囲を自由に指定して検索できるデータベースを構築し、検索サービスを開始した。このこ

とにより、このデータを遺伝子発現プロファイルのデータと照らし合わせていくことで、疾病の発症に関連する遺伝子配列を特定するための準備が整いつつある。

F. 健康危険情報

なし

G. 研究発表

なし

H. 知的財産権の出願・登録状況

1. 特許取得;なし

2. 実用新案登録;なし

3. その他;なし

研究成果の刊行に関する一覧表

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Choi YL, Moriuchi R, Osawa M, Iwama A, Makishima H, Wada T, Kisanuki H, Kaneda R, Ota J, Koinuma K, Ishikawa M, Takada S, Yamashita Y and Mano H.	Retroviral expression screening of oncogenes in natural killer cell leukemia.	*Leuk. Res.*	in press		2005
Tsutsumi C, Ueda M, Miyazaki Y, Yamashita Y, Choi YL, Ota J, Kaneda R, Koinuma K, Fujiwara S, Kisanuki H, Ishikawa M, Ozawa K, Tomonaga M and Mano H.	DNA microarray analysis of dysplastic morphology associated with acute myeloid leukemia	*Exp. Hematol.*	32	828-835	2004
Choi YL, Makishima H, Ohashi J, Yamashita Y, Ohki R, Koinuma K, Ota J, Isobe Y, Ishida F, Oshimi K and Mano H.	DNA microarray analysis of natural killer cell-type lymphoproliferative disease of granular lymphocytes with purified CD3(-)CD56(+) fractions	*Leukemia*	18	556-565	2004
Koinuma K, Shitoh K, Miyakura Y, Furukawa T, Yamashita Y, Ota J, Ohki R, Choi YL, Wada T, Konishi F, Nagai H and Mano H	Mutations of BRAF are associated with extensive hMLH1 promoter methylation in sporadic colorectal carcinomas	Int J Cancer	108	237-242	2004
Kaneda R, Toyota M, Yamashita Y, Koinuma K, Choi YL, Ota J, Kisanuki H, Ishikawa M, Takada S, Shimada K and Mano H.	High-throughput screening of genome fragments bound to differentially acetylated histones	*Genes Cells*	9	1167-1174	2004
Ohki-Kaneda R, Ohashi J, Yamamoto K, Ueno S, Ota J, Choi YL, Koinuma K, Yamashita Y, Misawa Y, Fuse K, Ikeda U, Shimada K and Mano H.	Cardiac function-related gene expression profiles in human atrial myocytes	*Biochem. Biophys. Res. Commun.*	320	1328-1336	2004
Ohki R, Yamamoto K, Ueno S, Mano H, Misawa Y, Fuse K, Ikeda U and Shimada K.	Transcriptional profile of genes induced in human atrial myocardium with pressure overload	*Int. J. Cardiol.*	96	381-387	2004

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Kano Y, Akutsu M, Tsunoda S, Izumi T, Mori K, Fujii H, Yazawa Y, Mano H and Furukawa Y.	Schedule-dependent synergism and antagonism between pemetrexed and paclitaxel in human carcinoma cell lines in vitro	*Cancer Chemother. Pharmacol.*	54	505-513	2004
He H, Hirokawa Y, Gazit A, Yamashita Y, Mano H, Kawakami Y, Hsieh CY, Kung HJ, Lessene G, Baell J, Levitzki A and Maruta H.	The Tyr-Kinase Inhibitor AG879, That Blocks the ETK-PAK1 Interaction, Suppresses the RAS-Induced PAK1 Activation and Malignant Transformation	*Cancer. Biol. Ther.*	3	96-101	2004
Bai J, Sata N, Nagai H, Wada T, Yoshida K, Mano H, Sata F and Kishi R.	Genistein-Induced Changes in Gene Expression in Panc 1 Cells at Physiological Concentrations of Genistein.	*Pancreas*	29	93-98	2004
Araki H, Katayama N, Yamashita Y, Mano H, Fujieda A, Usui E, Mitani H, Ohishi K, Nishii K, Nasuya M, Minami N, Nobori T and Shiku H.	Reprogramming of human postmitotic neutrophils into macrophages by growth factors.	*Blood*	103	2973-2980	2004
Aoki N, Ueno S-i, Mano H, Yamasaki S, Shiota M, Miyazaki H, Yamaguchi Y, Matsuda T & Ullrich A.	Mutual regulation of protein-tyrosine phosphatase 20 and protein-tyrosine kinase Tec activities by tyrosine phosphorylation and dephosphorylation.	*J. Biol. Chem.*	279	10765-10775	2004
大島康雄, 藤村昭夫	日本人組織を用いたトキシコゲノミクス研究	臨床薬理	36	11-12	2005
Numata A, Shimoda K, Kamezaki K, Haro T, Kakumitsu H, Shide K, Kato K, Miyamoto T, Yamashita Y, Oshima Y, Nakajima H, Iwama A, Aoki K, Takase K, Gondo, H, Mano H, Harada M.	Signal transducers and activators of transcription 3 augments the transcriptional activity of CCAAT/enhancer binding protein a in G-CSF signaling pathway.	J Biol Chem	280	12621-12629	2005
Oshima Y, Tojo A, Fujimura A, Niho Y, Asano S.	Potent receptor-mediated cytotoxicity of granulocyte colony-stimulating factor-Pseudomonas exotoxin, a fusion protein against myeloid leukemia cells.	Biochem Biophys Res Commun	319	582-589	2004
Oshima Y, Tojo A	Gene expression profiles in the cellular response to recombinant Pseudomonas exotoxin A	Cancer Sci	95 (suppl1)	488	2004

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Oshima Y, Komatsu N, Ozawa K, Fujimura A	CML Developed in a Japanese Family Transmitting a Novel Point Mutation in the Thrombopoietin Gene (TPO).	blood	104(11-part 2 of 2 parts)	141b (Abs# 4220)	2004
Oshima Y, Kurokawa S, Tokue A, Mano H, Saito K, Suzuki M, Imai M, Fujimura A	Primary Cell Preparation of Human Renal Tubular Cells for Transcriptome Analysis	**Toxicol Mechanisms and Methods**	14	309-316	2004
Oshima Y, Ishida Y, Shinohara A, Mano H, Fujimura A	Expression Profiling of Gene with Upstream AML1 Recognition Sequence in Hematopoietic Stem Cell-Like Fractions from Individuals with The M2 Subtype of Human Acute Myeloid Leukemia	**Experimental Hematology**	32 (suppl)	34	2004

健康危険情報

なし

ELSEVIER

EXPERIMENTAL HEMATOLOGY

Experimental Hematology 32 (2004) 828–835

DNA microarray analysis of dysplastic morphology associated with acute myeloid leukemia

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

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

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

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

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

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

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

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

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

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

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

0301-472X/04 $–see front matter. Copyright © 2004 International Society for Experimental Hematology. Published by Elsevier Inc.
doi: 10.1016/j.exphem.2004.06.003

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

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

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

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

Patients and methods

Purification of AC133$^+$ cells

BM aspirates were obtained from the study subjects with written informed consent. From each specimen, MNCs were isolated by Ficoll-Hypaque density gradient centrifugation, and were labeled with magnetic bead–conjugated anti-AC133 monoclonal antibody (AC133 MicroBead; Miltenyi Biotec, Auburn, CA, USA). AC133$^+$ HSC-like fractions were then purified through a miniMACS magnetic cell separation column (Miltenyi Biotec), and enrichment of the HSC-like fraction was evaluated by subjecting portions of the MNC and AC133$^+$ cell preparations either to staining with Wright-Giemsa solution or to the analysis of the expression of CD34, CD38, and AC133 by flow cytometry (FACScan; Becton-Dickinson, Mountain View, CA, USA). In most instances, the CD34^{high}CD38^{low} fraction constituted greater than 90% of the eluate of the affinity column.

DNA microarray analysis

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

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

Statistical analysis

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

Results

Comparison of AML-MLD and MDS-related AML

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

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

Table 1. Patient characteristics

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

AML, de novo AML; MLD, AML with multilineage dysplasia; MDS, MDS- associated AML, TRL, therapy-related AML; M, male; F, female.
*Individuals proven not to have a prior history of MDS.

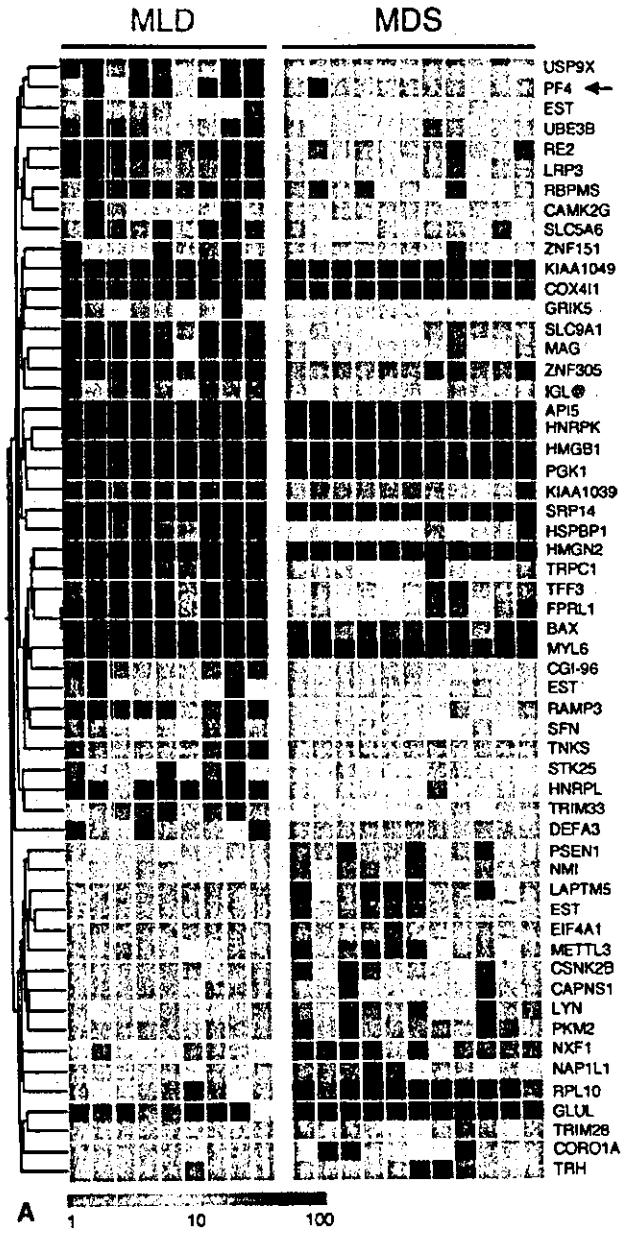
AML with dysplasia and secondary AML evolving from an undiscovered phase of MDS. Although the clinical characteristics of the former have not been fully defined, it has been reported that de novo AML-MLD may be associated with poor prognosis [17,18] and, in some cases, with an increased megakaryopoiesis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

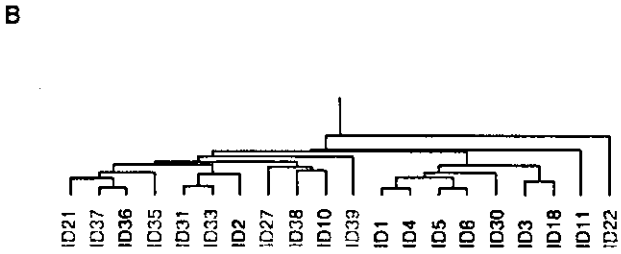
To clarify directly whether de novo AML-MLD is truly a clinical entity distinct from MDS-related leukemia, we searched for differences between the transcriptomes of $AC133^+$ cells derived from the individuals diagnosed with these two conditions. Among the 11 cases of AML-MLD studied, 9 were revealed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

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

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

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





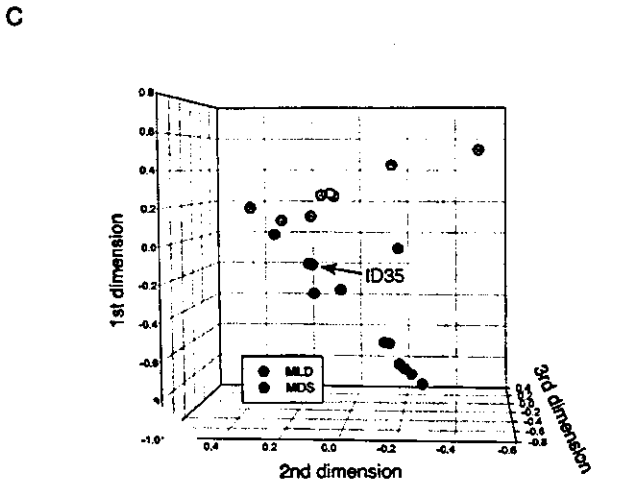


Figure 1. Continued

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

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

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

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

Comparison of AML without dysplasia and de novo AML-MLD

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

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

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

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

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

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

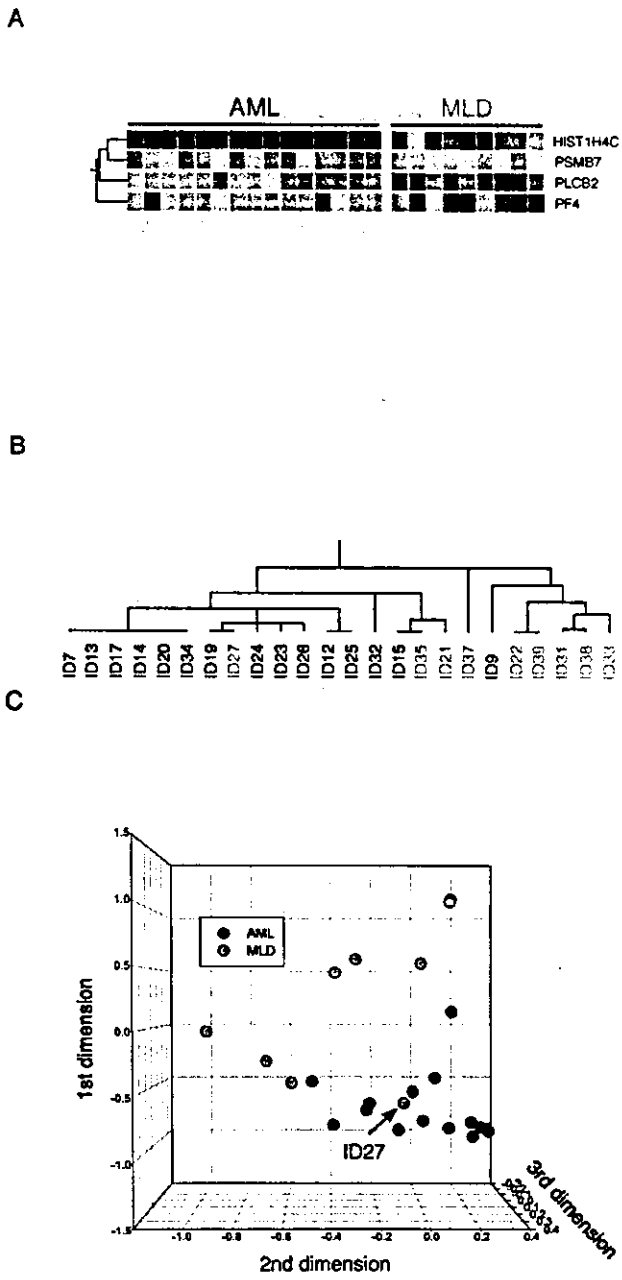


Figure 2. Comparison of gene expression profiles between patients with de novo AML without dysplasia and those with de novo AML-MLD. (A): Gene tree for the expression levels of four human genes in AC133$^+$ cells from individuals with de novo AML without dysplasia (AML) or de novo AML-MLD (MLD). (B,C):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

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

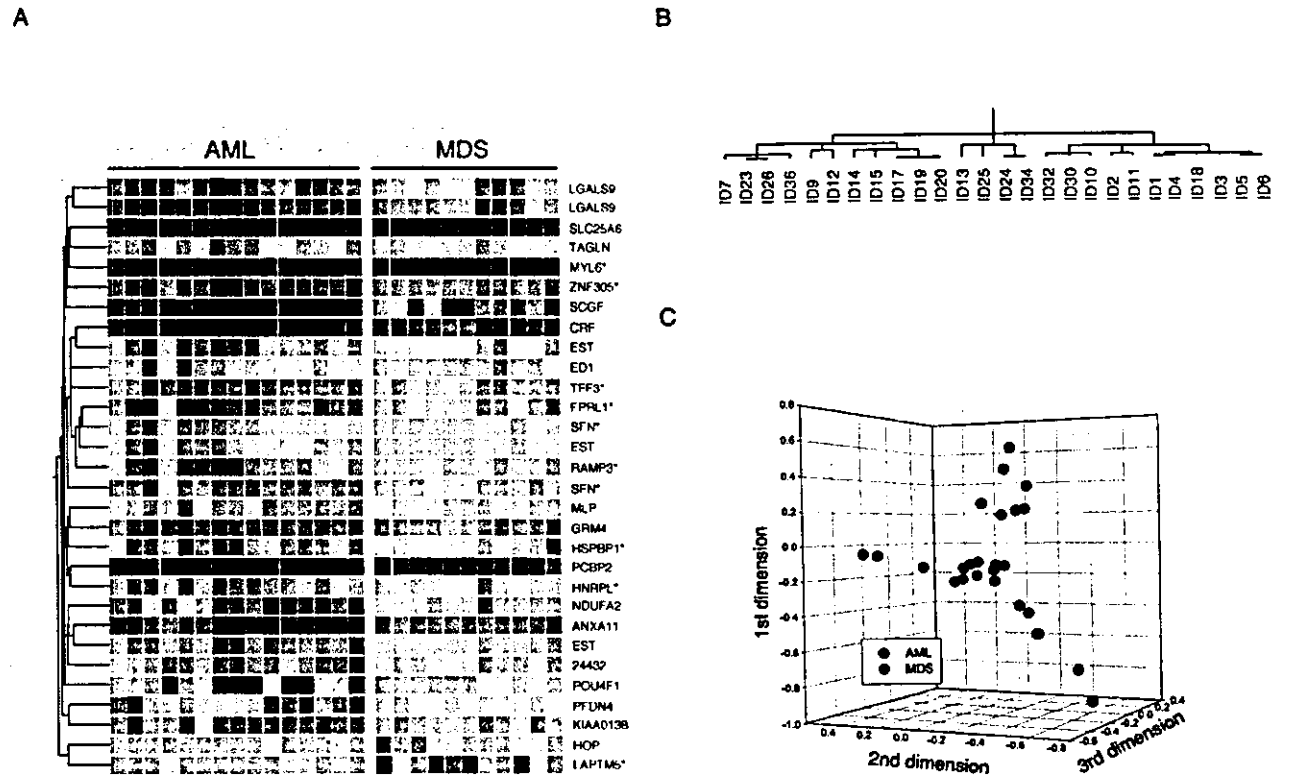


Figure 3. Comparison of gene expression profiles between patients with de novo AML without dysplasia and those with MDS-related leukemia. (A): Expression profiles of 30 probe sets (28 genes) that contrast de novo AML without dysplasia (AML) and MDS-related leukemia (MDS). Two independent probe sets were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

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

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

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

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

Comparison of whole samples

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