

ゲノム機能解析ツールとしての DNAマイクロアレイ

油谷浩幸



マイクロアレイ 発現プロファイル 染色体変異 遺伝子多型 タイピングアレイ

●はじめに

2003年の4月にヒトゲノム配列解読完了が宣言され、ヒト遺伝子は3万あまりと予測されるが、ヒトとマウスの比較でも“要素”となる遺伝子の数には大きな違いはなく、ゲノム配列からRNAあるいは蛋白質が発現してくる時間的・空間的な制御機構の解明が求められている。一方、生物学のセントラルドグマであるDNA→RNA→蛋白質という生命情報の各要素を迅速に網羅的に大量取得することが可能となり、生命科学研究にパラダイムシフトをもたらしている。DNAマイクロアレイは網羅的な解析技術としてRNA量の変動の解析(トランスクリプトーム)に用いられるのみならず、多型解析やゲノムの機能解析まで臨床から基礎研究においてさまざまな応用が可能である。

I. マイクロアレイ技術

遺伝子発現を比較することは、従来ディフュゼンシャル・ディスプレイやサブトラクション法を用いて行なわれていたわけであるが、DNAチップ技術では3万個をこえる遺伝子の発現レベルを同時解析可能である。マイクロアレイの原理は、基板上的DNAまたはオリゴヌクレオチドの配列(プローブ)と標識した試料RNA(ターゲット)とのハイブリダイゼーションによりRNA発現量を検出するが、プローブの高集積度化や感度の向上により検出精度や再現性も向上し、包括的な発現情報解析の強力な手法として広く使用されている。従来、スライドガラスにcDNAをプリントしたスタンフォード方式

と、半導体技術を利用した合成DNAアレイ(例: Affymetrix社のGeneChip)の2方式が主体であったが、ゲノム配列情報が解読され、おおよその遺伝子領域が同定されたこととなり、合成DNAによるアレイへとシフトしつつある。前者は研究目的に従い比較的簡便にカスタムアレイを作製することができる利点がある一方、大規模アレイを作製する際のクローンあるいはオリゴの管理や異なるアレイ間のデータの標準化に、依然課題が残る。個々の研究室で作製する場合には、ライブラリーからのクローンの選択、プラスミドDNA調製、インサートDNAの増幅の各工程は案外に煩雑なものであり、最終的に増幅されたDNA断片が本来のクローン自体と同一であるかの検証が必要となる。

Affymetrix社のDNAチップ(GeneChip™)は照射化学合成技術を用いてアレイ上でオリゴヌクレオチドの合成を行なうが、当初はペプチド合成技術を有するAffymax社の1部門としてスタートしたり、ゲノム計画の成果としてEST(expressed sequence tag)データの蓄積も進んだことから(本増刊号II-2 浅水・中村の項参照)、網羅的なアレイの作製が可能となりDNAチップとして世界中で広く用いられている(図1)。Agilent社はインクジェット方式で*in situ*合成を行なう方式のアレイを開発しており、カスタムアレイの作製に広く用いられている。同社は2種類の蛍光を用いて2種類の試料を同時測定することを推奨している。従来の2検体比較のような実験には適していると考えられるが、多数検体についてデータを収集する際には共通な対照検体を準備せねばならないことが難点となる。Nimblegen社は光リソグ

Hiroyuki Aburatani, 東京大学先端科学技術研究センターゲノムサイエンス分野 E-mail: haburata-tky@umin.ac.jp
Functional genomics analysis on DNA microarray

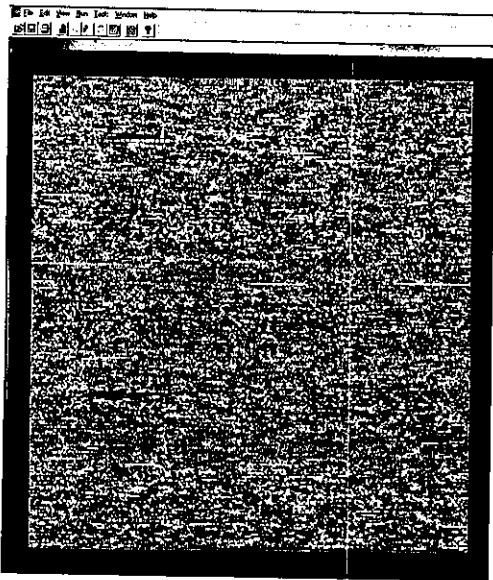


図1 GeneChipのスキャン画像

現行の発現解析用アレイU133 2.0では11 μm 四方ごとに1プローブが合成されており、アレイ全体でおよそ130万種類のオリゴの合成が可能である。シグナルは共焦点スキャナーを用いて1.6 μm の解像度で読み取る。[口絵11 (p.1459) 参照]

ラフ技術を用いながらも、ミラーを用いて照射区域を特定することにより高密度オリゴアレイを *in situ* 合成する技術を開発している。光マスクを必要としないため、カスタム合成を少ない枚数から低コストで行なうことが可能である²⁾。興味深いことに、通常の *in situ* 合成は3'から5'へと行なわれるが、最近5'から3'への合成が可能となり、アレイ上での種々の酵素反応が可能となりつつある³⁾。このほかに製品化されている網羅的な発現解析用のアレイとしては、Amersham社(商品名CodeLink)はあらかじめ合成されたオリゴをインクジェットで3次元マトリックス上にプリントする方式を採用しており、高い再現性を実現している。ABI社、日立ソフトはピンでプリントしているが、ABI社のアレイは60 merの検出用オリゴに加えて24 merの品質管理用オリゴを同時にスポットすることにより、後者からの蛍光シグナルを用いてアレイ間の標準化を目指している。ABI社のアレイはCelera社のゲノムデータに基づいた配列情報を利用しているほか、検出系に化学発光を利用することにより感度の改善を目指している。Illumina社は表面にオリゴを合成したビーズを用いて多型解析や発現解析に利用されている。ビーズを光ファイバーの先端に固定することにより、高感度な検出が可能であるほか、ビーズをアレイ上に配置することにより、通常のスキャナーでの読

み取りも可能としている。なお、ビーズを特定するための配列コードがオリゴ配列に含まれており、あらかじめデコーディングによる配列の同定が必要とされ、オリゴ配列のデザインがやや煩雑である。Combimatrix社(www.combimatrix.com)、Nanogen社(www.nanogen.com)は電気化学的合成技術を生かしたアレイを開発している。

ハイブリダイゼーションシグナルの検出には核酸標識におもに蛍光色素を用いられているが、高感度を得るべく化学発光や共鳴光散乱(RLS)光⁴⁾も利用される。後者は金属超微粒子の光学物性である共鳴光散乱現象を利用しており、米国のGenicon社が開発を進めている。Qiagen社のSensiChipは同技術に加え、PWG(planar waveguide, 平面導波路)技術を併用し背景光によるノイズを遮断することで高感度化を実現している。従来、2種類のターゲットRNAに対して、Cy3-dUTPとCy5-dUTPなど励起波長の異なる蛍光物質による標識を行ない、同時にハイブリダイゼーションを行なうことにより発現量を比較することが広く行なわれてきた。蛍光色素による標識効率および安定性の違いやデータの標準化において、対照検体を共通にすることの困難などから、最近では固定DNA量の再現性の向上とも相まって、単色蛍光による測定系が好まれつつある。量子ドットによる検出系も応用されている(www.qdots.com)。

検出装置については、現在はもっぱら共焦点レーザースキャナーあるいはCCDカメラが用いられているが、とりわけ前者は高価であることと、スキャンに要する時間が長いことが難点となりつつある。最新の高密度アレイでは1枚をスキャンするために10分以上要するようになり、検体数の増加に対して何らかの将来的対応が必要である。また、1枚のアレイから生みだされる情報量は100 MBをこえるサイズになっており、データの保管および処理についても新たな情報処理技術の導入が求められている。

II. DNAチップ解析の実際

1. ゲノム多型解析

SNP解析への応用は1回のアレイ解析により5万カ所のSNPをタイピングできるようになり、解析コストの面でも他の測定法に対して競争力を有するようになり

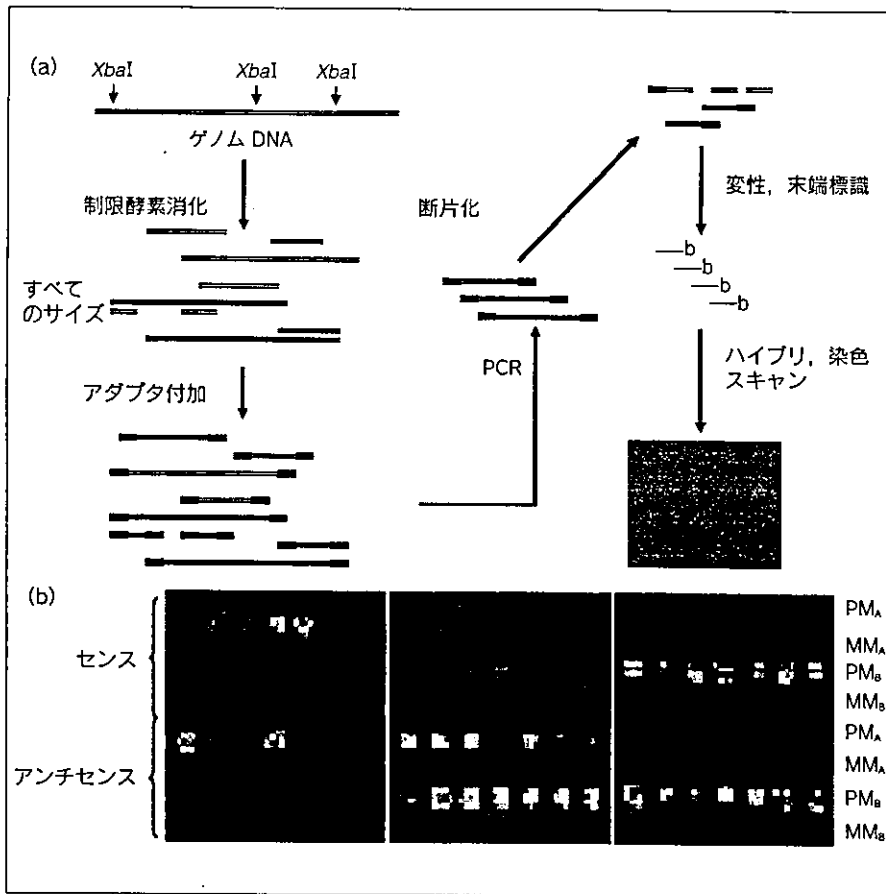


図2 アレイによるゲノムタイピング

全ゲノム中より特定の制限酵素断片をPCR増幅し、多型を認識するプローブを合成したアレイにハイブリダイゼーションを行なう。(a)概略図、(b)スキャン画像。PMはperfect match(完全一致)、MMはmismatch(ミスマッチ)のプローブ配列を意味している。(文献5より許可を得て転載) [口絵12(p.1460)参照]

つつある。GeneChip (Affymetrix) は公共データベースおよびPerlegen社が同定したSNP情報に基づいて設計されている。1セット2枚のタイピングアレイを用いることにより、ゲノム全体から10万SNP、およそ20 kbに1 SNPのタイピングが可能である(図2)。ゲノムワイドに有用な多型マーカーを、関連解析により疾患関連遺伝子を探索している現状では、多数のマーカーを高密度にタイピングする必要があるが、ハプロタイプが決定され将来的に解析すべきマーカー数が絞られた場合には、アレイによるタイピングもスクリーニングの有力手段であると思われる。家系解析の場合は、同時にゲノムワイドに1万のマーカーをタイピングできれば、200 kbの解像度で解析できるので連鎖解析に十分有効である^{5,6)}。症例ごとにタイピングできるので実際の解析のうえでも実用的である。

一方、がん細胞では、がん化に先んじてゲノムの不安定化が生じていることが知られている。塩基レベルあるいは染色体レベルでのゲノム複製の確実性が失われることにより、変異が蓄積し、がん化に至る。個人の全ゲノ

ム配列を決定することも将来的には可能になると期待されるが、現時点ではコスト的に非現実的である。コピー数に関してはアレイCGH(comparative genomic hybridization)やDNAチップを用いてかなり正確な遺伝子のコピー数の解析が可能となりつつあり^{7,8)}、SNP測定とあわせることにより、ゲノム全体について増幅領域や欠失領域の判定などの染色体変異の解明が期待される(図3)。BACクローンによるタイピングアレイ(アレイの高密度化)についても、最近報告がなされた⁹⁾。32,433個のBACクローンを用いて、アレイCGHによるゲノム異常の検出を行なっている。HER2などの治療標的分子の増幅の判定への臨床応用も可能である。

2. 配列解析

A. シークエンシング

がん関連遺伝子の変異を解析するためにシークエンシングは有効な手段である。悪性黒色腫においてのBRAF遺伝子の高頻度な変異については英国のがんゲノムプロジェクトの成果であり¹⁰⁾、網羅的な解析の有効性を物

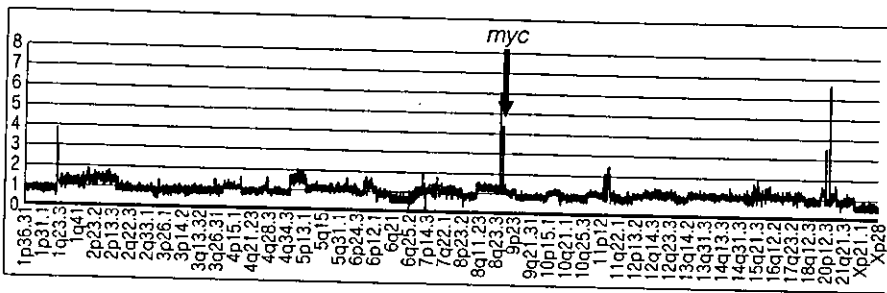


図3 Genomic Imbalance Map
ゲノムアレイを用いてがん細胞ゲノム全体のコピー数解析を行なった結果である。myc遺伝子などの既知の増幅領域をはじめとする染色体異常が一目の下に明らかである。

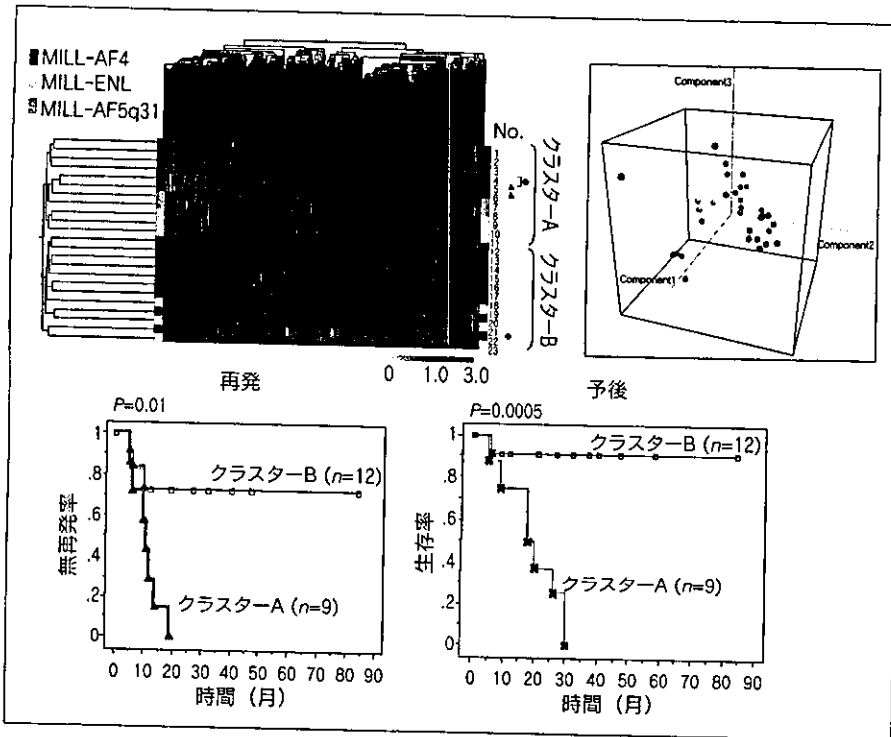


図4 アレイ解析によるMLL転座小児白血病における予後予測
[口絵13 (p.1460) 参照]

語る例である。すでにp53遺伝子¹¹⁾やp450遺伝子については配列決定用にマイクロアレイが用いられている。

B. タイリングアレイ

アレイの高密度化により、ゲノム配列を網羅的に配置することが現実的になりつつある。現在使われている高解像度アレイは、5 μm四方ごとに異なる配列を合成できるため、1枚あたり600万種類のDNAを合成可能であるので、ヒトゲノム配列のほぼすべてが14枚のアレイに納められる。応用範囲はきわめて広く、転写領域の同定¹²⁾、スプライシングバリエントの検出に用いることができるほか、転写因子結合部位 (ChIP-chip解析)^{13,14)} や複製開始点¹⁵⁾ の解析に用いることが試みられている。現在、米国のポストゲノム計画としてENCODEプロジェクトが進められており、全ゲノム配列の1%について詳細な解析が行なわれている。

3. 発現解析

解析に必要な検体量も、1遺伝子あたりに換算すると従来法よりはるかに少量でよく、数万のデータを同時に取得できる利点があり、比較的高発現な遺伝子については他の測定法との相関も良好である¹⁶⁾。転写のネットワークの解析や臨床検体の解析などに広く用いられている (図4)。1枚のアレイで数万に及ぶほぼすべての転写産物を解析することが可能となったものの、低発現レベルの遺伝子についてはまだまだ信頼性が低い。ゲノム配列中の類似配列の排除や2次構造の予測により、さらに特異性の高いプローブ配列の選択が進められているが、実際にハイブリ実験によるプローブ性能の実証も、今後いっそう重要となるであろうと思われる。プローブ配列の長さについては、個々の遺伝子ごとに実験データとの比較検証が必要である。また、現行のアレイではプロー

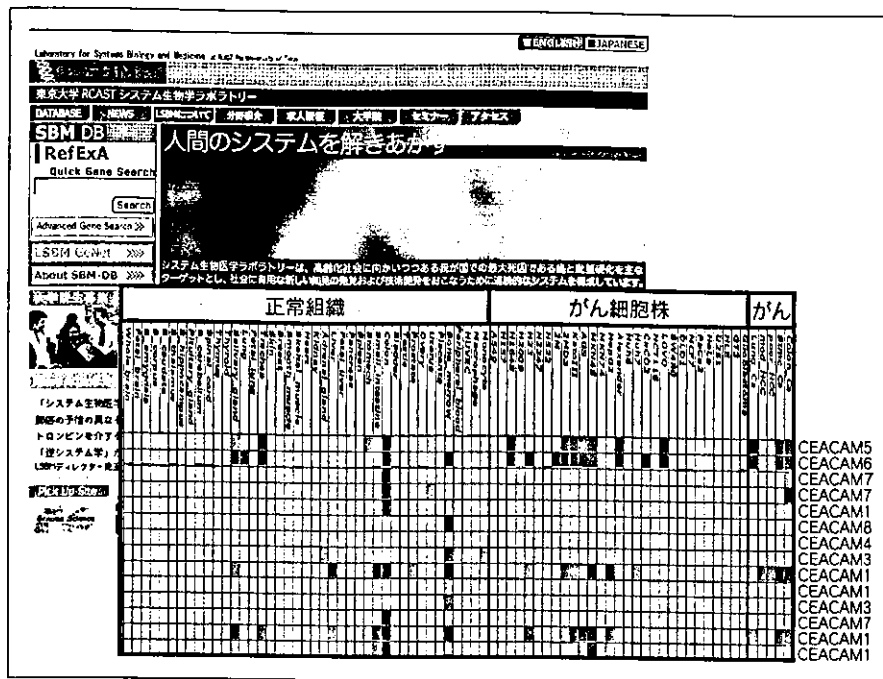


図5 システム生物学データベース (SMB-DB)

ブの選択やRNAの標識法による制約から、スプライシングバリエントを検出することは困難である。そもそもどれだけのバリエントが存在するかの同定を含めて、RNAの全長を解析するための新たな実験手法の開発が今後の課題であり、前述のゲノムタイリングアレイは有力な解析手段と期待される。ただし、短いエクソンやエクソン-イントロンの境界の検出には、さらに高密度なアレイの開発が必要とされ、スプライシング変異産物を網羅的に検出するためには、エクソンどうしをまたぐようなプローブデザインが必要とされる。

微量な臨床試料を解析するために、増幅法や高感度の検出法が開発されている。T7 RNAポリメラーゼを用いる標識法では、1 μg程度、ランダムプライマーによるリニアな増幅を加えると、20 ng程度のトータルRNAからも比較的再現性のよいデータが得られる。現行の標識法のほとんどは、mRNAのポリ(A)配列を利用してオリゴdTプライマーによりcDNAを合成するため、DNAプローブの位置が3'末端よりにデザインされることが重要である。また、検体RNAの品質がよいことが求められることはいうまでもなく、28S/18SのrRNAの量比が1以上であること、増幅されたcRNAの長さが最低数百塩基はいることが望ましいので、Bioanalyzer (Agilent) などにより確認することが望ましい(本増刊号II-8 馬場の項参照)。GeneChipではβア

クチンやG3PDH遺伝子の5'側と3'側のシグナル値を比較することにより、5'側シグナルが極端に低下している場合は標識反応が不良であった可能性が高いという判断を下すようになっている。

臨床検体を解析する際には、必ずしも品質のよいRNAが入手できるとは限らず、稀少な症例では新鮮な凍結標本が得られないこともしばしばであるため、パラフィン包埋標本からRNA抽出を行ない、アレイ解析に供することも試みられている。この場合には標識されるcRNAが短いため、それを認識するためのプローブの位置がきわめて重要である。Affymetrix社はArcturus社と共同で、ほと

んどの遺伝子について、ポリ(A)配列より300塩基以内にプローブをデザインしたX3Pアレイを開発しており、通常タイプのアレイと比較的よい相関を示している。パラフィン固定時の条件にもよるが、発現プロファイル解析に使用される検体についての制約が軽減され、発現解析の応用の可能性が広がるものと思われる。

4. 情報解析

データの正規化をはじめとするマイクロアレイデータの解析手法一般については既報に詳しい^{17,18)}。通常、測定値を対数変換して取り扱うことが多く、おおむね対数正規分布する。基本的に絶対量を測定できる系ではないので、細胞あるいは組織のRNAの中の相対的な発現をモニタリングしていることを理解する必要がある。ハウスキーピング遺伝子のセットを用いて、実験間の比較を行なうことも試みられているが、なかなか問題点も多い¹⁹⁾。マイクロアレイが将来的に研究目的のみならず臨床の場合において活用される場合には、標準検体の設定や実験ごとにスパイクする試料を共通化することも必要と考えられる。

得られた発現プロファイル解析データのデータベース化については、数年来検討されているが、解析プラットフォームの相違やプローブ配列の違いといった本質的な面のみならず、実験プロトコル、検体採取法、スキャニ

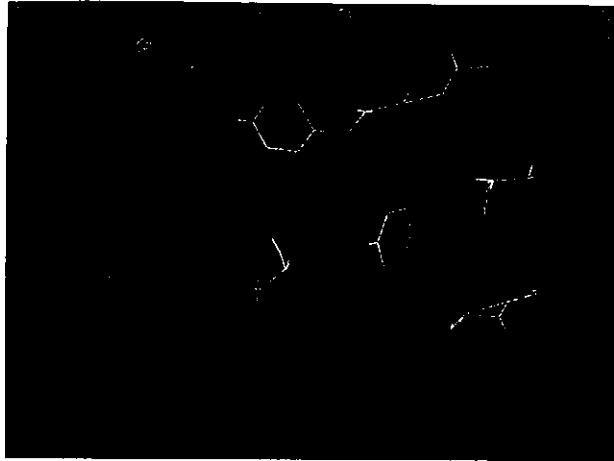
ング条件などの技術的な違いによっても解析データが異なることがあるため、複数の施設間での解析データを相互比較することは容易ではない。とくに2色の蛍光を用いる方式の場合には、対照に用いる検体の選択も重要な因子となる。NCBI, EBIなどで公共データベース GEO (gene expression omnibus), ArrayExpressが公共のデータ登録に用いられており、学術論文にアレイデータを用いる場合には MIAME (minimum information about a microarray experiment)²⁰⁾に準拠することが求められるようになってきた。筆者らの研究室でもヒト正常組織、初代培養細胞、がん細胞株など80検体について、最大4万個の遺伝子を解析した結果をデータベース化して公開している(システム生物医学データベース <http://www.lsbm.org>) (図5)。

●おわりに

大規模な生命情報に基づいての生命現象を、システムとして取り扱う「システム生物学」は生物医学研究において従来の個別遺伝子研究からのパラダイムシフトをもたらすものと考えられる。DNAアレイ解析技術はそのなかで、機能ゲノム解析においての大きな役割を担っていくものと期待され、今後さらに再現性、感度においてのいっそうの向上が望まれる。

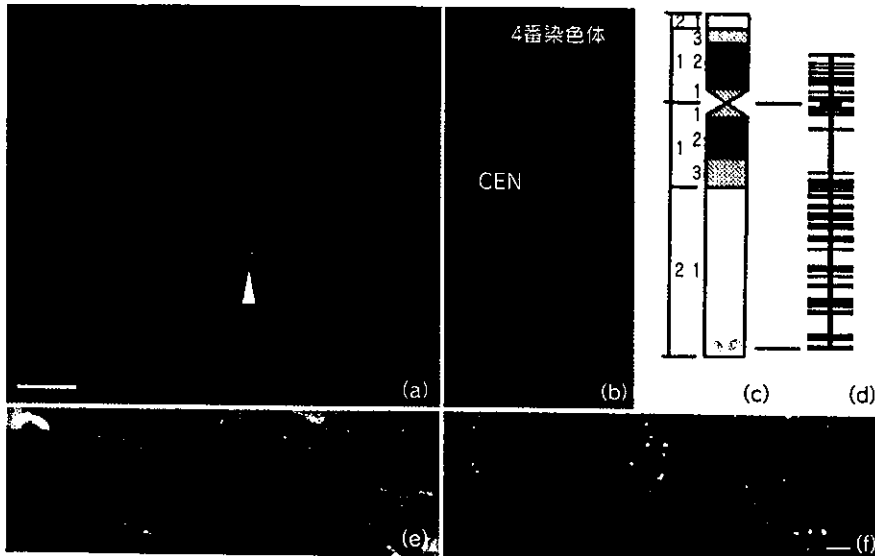
文 献

- 1) Fodor, S. P., Read, J.L., Pirrung, M. C., Stryer, L., Lu, A. T., Solas, D. : *Science*, **251**, 767-773(1991)
- 2) Nuwaysir, E. F., Huang, W., Albert, T. J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J. P., Ballin, J., McCormick, M., Norton, J., Pollock, T., Sumwalt, T., Butcher, L., Porter, D., Molla, M., Hall, C., Blattner, F., Sussman, M.R., Wallace, R.L., Cerrina, F., Green, R. D. : *Genome Res.*, **12**, 1749-1755(2002)
- 3) Albert, T. J., Norton, J., Ott, M., Richmond, T., Nuwaysir, K., Nuwaysir, E. F., Stengele, K. P., Green, R. D. : *Nucleic Acids Res.*, **31**, e35(2003)
- 4) Bao, P., Frutos, A. G., Greef, C., Lahiri, J., Muller, U., Peterson, T. C., Warden, L., Xie, X. : *Anal. Chem.*, **74**, 1792-1797(2002)
- 5) Kennedy, G. C., Matsuzaki, H., Dong, S., Liu, W. M., Huang, J., Liu, G., Su, X., Cao, M., Chen, W., Zhang, J., Liu, W., Yang, G., Di, X., Ryder, T., He, Z., Surti, U., Phillips, M. S., Boyce-Jacino, M. T., Fodor, S. P., Jones, K. W. : *Nat. Biotechnol.*, **21**, 1233-1237(2003)
- 6) Matsuzaki, H., Loi, H., Dong, S., Tsai, Y. Y., Fang, J., Law, J., Di, X., Liu, W. M., Yang, G., Liu, G., Huang, J., Kennedy, G. C., Ryder, T. B., Marcus, G. A., Walsh, P. S., Shriver, M. D., Puck, J. M., Jones, K. W., Mei, R. : *Genome Res.*, **14**, 414-425(2004)
- 7) Lucito, R., Healy, J., Alexander, J. *et al.* : *Genome Res.*, **13**, 2291-2305(2003)
- 8) Bignell, G. R., Huang, J., Greshock, J., Watt, S., Butler, A., West, S., Grigorova, M., Jones, K. W., Wei, W., Stratton, M. R., Futreal, P. A., Weber, B., Shaper, M. H., Wooster, R. : *Genome Res.*, **14**, 287-295(2004)
- 9) Ishkanian, A. S., Malloff, C. A., Watson, S. K., DeLeeuw, R. J., Chi, B., Coe, B. P., Sniijders, A., Albertson, D. G., Pinkel, D., Marra, M. A., Ling, V., MacAulay, C., Lam, W.L. : *Nat. Genet.*, **36**, 299-303(2004)
- 10) Davies, H., Bignell, G. R., Cox, C. *et al.* : *Nature*, **417**, 949-954(2002)
- 11) Ahrendt, S. A., Hu, Y., Buta, M., McDermott, M. P., Benoit, N., Yang, S. C., Wu, L., Sidransky, D. : *J. Natl. Cancer Inst.*, **95**, 961-970(2003)
- 12) Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G., Tammana, H., Gingeras, T. R. : *Genome Res.*, **14**, 331-342(2004)
- 13) Cawley, S., Bekiranov, S., Ng, H. H., Kapranov, P., Sekinger, E. A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A. J., Wheeler, R., Wong, B., Drenkow, J., Yamanaka, M., Patel, S., Brubaker, S., Tammana, H., Helt, G., Struhl, K., Gingeras, T. R. : *Cell*, **116**, 499-509(2004)
- 14) Martone, R., Euskirchen, G., Bertone, P., Hartman, S., Royce, T. E., Luscombe, N. M., Rinn, J. L., Nelson, F. K., Miller, P., Gerstein, M., Weissman, S., Snyder, M. : *Proc. Natl. Acad. Sci. USA*, **100**, 12247-12252(2003)
- 15) Katou, Y., Kanoh, Y., Bando, M., Noguchi, H., Tanaka, H., Ashikari, T., Sugimoto, K., Shirahige, K. : *Nature*, **424**, 1078-1083(2003)
- 16) Ishii, M., Hashimoto, Si., Tsutsumi, S., Wada, Y., Matsushima, K., Kodama, T., Aburatani, H. : *Genomics*, **68**, 136-143(2000)
- 17) 井原茂男・堤 修一・油谷浩幸 : *ゲノム医学*, **3**(2), 71-80(2003)
- 18) 加藤菊也・石井 信 : *蛋白質 核酸 酵素*, **48**, 2300-2309(2003)
- 19) Lee, P. D., Sladek, R., Greenwood, C. M., Hudson, T. J. : *Genome Res.*, **12**, 292-297(2002)
- 20) Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P., Holstege, F. C., Kim, I. F., Markowitz, V., Matese, J. C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., Vingron, M. : *Nat. Genet.*, **29**, 365-371(2001)



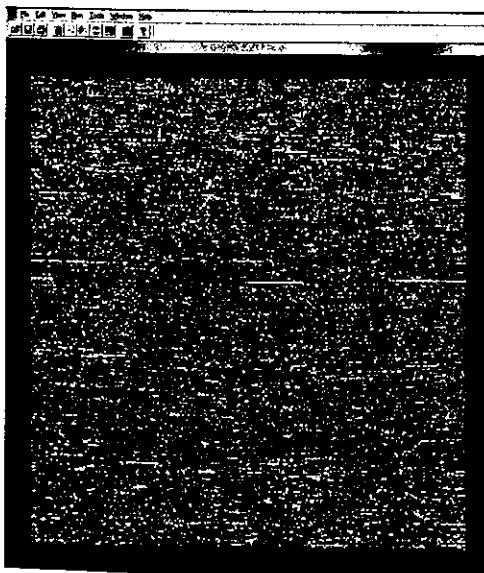
口絵9 Oの実行画面

電子密度図にあわせてモデルを構築する [松浦孝範・中川敦史, p.1825]



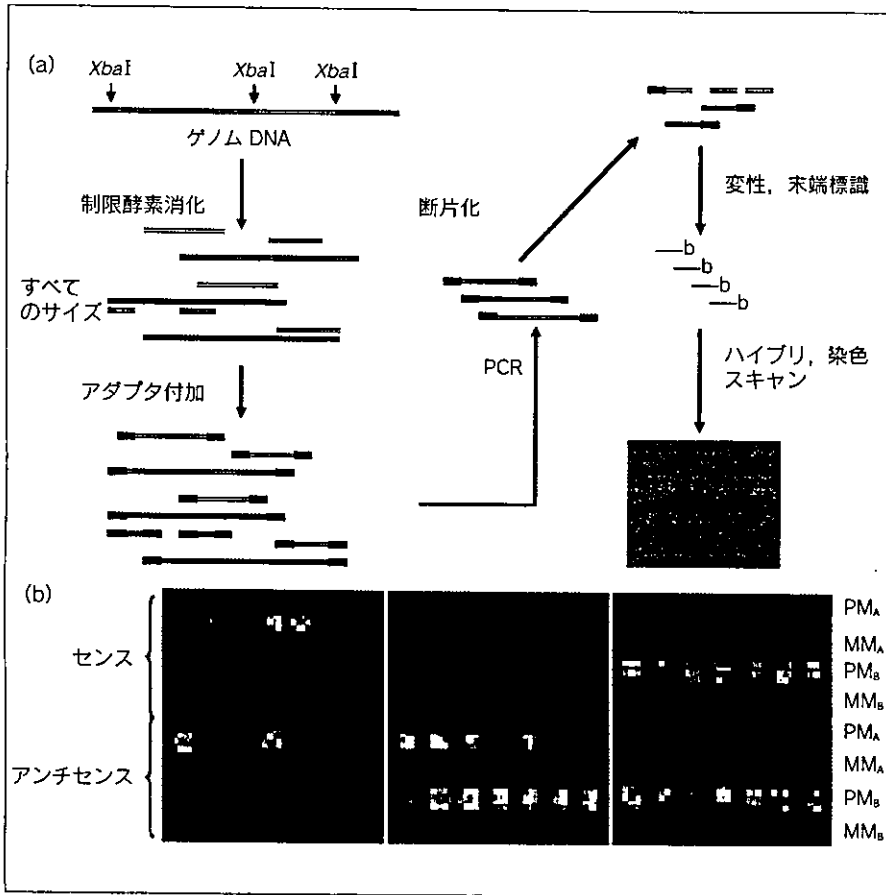
口絵10 FISHによるイネ4番染色体上へのRFLPマーカー (1.29 kb) のFISHによるマッピング

(a) RFLPマーカーのFISHによる検出, (b) 4番染色体の拡大, (c) 定量的染色体地図上への遺伝子のマッピング, (d) 遺伝地図上の遺伝子の位置, (e) DNAファイバー, (f) DNAファイバー上での縦列型反復配列 (赤) とテロメア (緑) の可視化. バーは10 μ mを示す. [佐藤修正・近江戸伸子・中村保一・田畑哲之, p.1829]



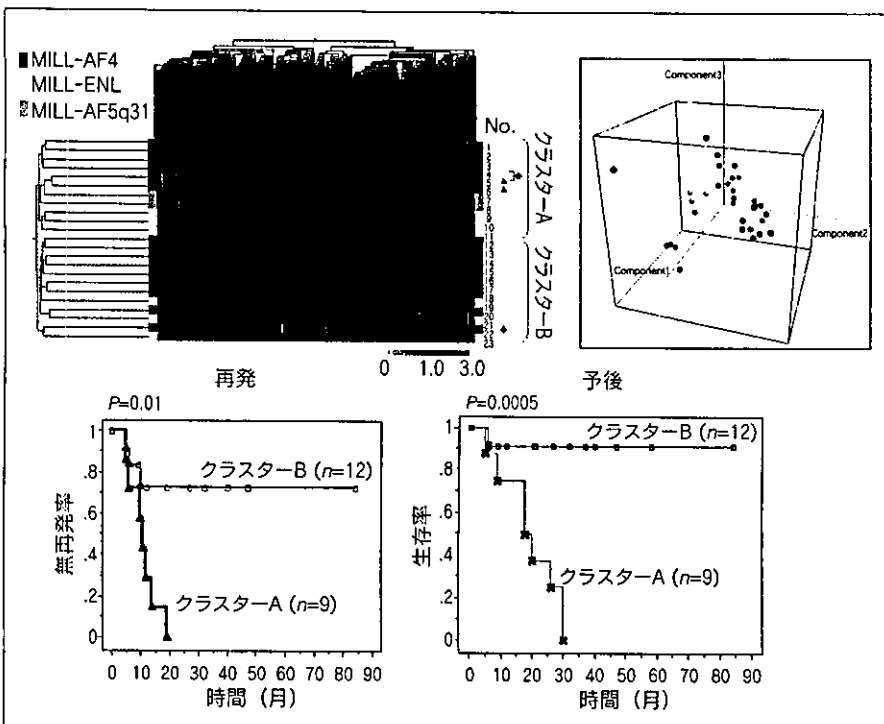
口絵11 GeneChipのスクリーン画像

現行の発現解析用アレイU133 2.0では11 μ m四方ごとに1プローブが合成されており, アレイ全体でおおよそ130万種類のオリゴの合成が可能である. シグナルは共焦点スキャナーを用いて1.6 μ mの解像度で読み取る. [油谷浩幸, p.1854]



口絵12 アレイによるゲノタイピング

全ゲノム中より特定の制限酵素断片をPCR増幅し、多型を認識するプローブを合成したアレイにハイブリダイゼーションを行なう。(a) 概略図、(b) スキャン画像。PMは perfect match (完全一致)、MMは mismatch (ミスマッチ) のプローブ配列を意味している。(文献5より許可を得て転載) [油谷浩幸, p.1855]



口絵13 アレイ解析によるMLL転座小児白血病における予後予測 [油谷浩幸, p.1856]

6 DNA チップのがん治療研究への応用

油谷 浩幸*

Key Sentences

- ① SNP 解析アレイを用いてのゲノムワイドな遺伝子増幅や LOH などの染色体変異の解明が期待される。
- ② ゲノムタイリングアレイにより、新規転写産物の同定や転写因子結合部位 (ChIP-chip 解析) や複製開始点の解析にも用いることができる
- ③ 分子レベルで新規な癌の分類・診断が行われ、治療への応答性および予後に関してより正確な予測に基づいた治療法の選択、すなわち「個別化医療」の実現が期待される。
- ④ 4 万個の遺伝子を発現解析した結果がデータベース化されて公開されている。

Key Words

マイクロアレイ, タイリングアレイ, ChIP-chip 解析 転写ネットワーク, 遺伝子発現データベース

はじめに

2003 年の 4 月にヒトゲノム配列解読完了が宣言され、ヒト遺伝子は 3 万あまりと予測されるが、ヒトとマウスの比較でも「要素」となる遺伝子の数には大きな違いはなく、ゲノム配列から RNA あるいは蛋白が発現してくる時間的、空間的な制御機構の解明が求められている。一方、生物学のセントラルドグマである DNA → RNA

→タンパクという生命情報の各要素を高速に網羅的に大量取得することが可能となり、生命科学研究にパラダイムシフトをもたらしている。DNA チップによる RNA の変動の解析 (トランスクリプトーム) は網羅的な解析技術として用いられるのみならず、臨床から基礎研究にまで様々な応用が可能である。

DNA chip technology in cancer research

* ABURATANI Hirofuki : 東京大学先端科学技術研究センターゲノムサイエンス分野教授

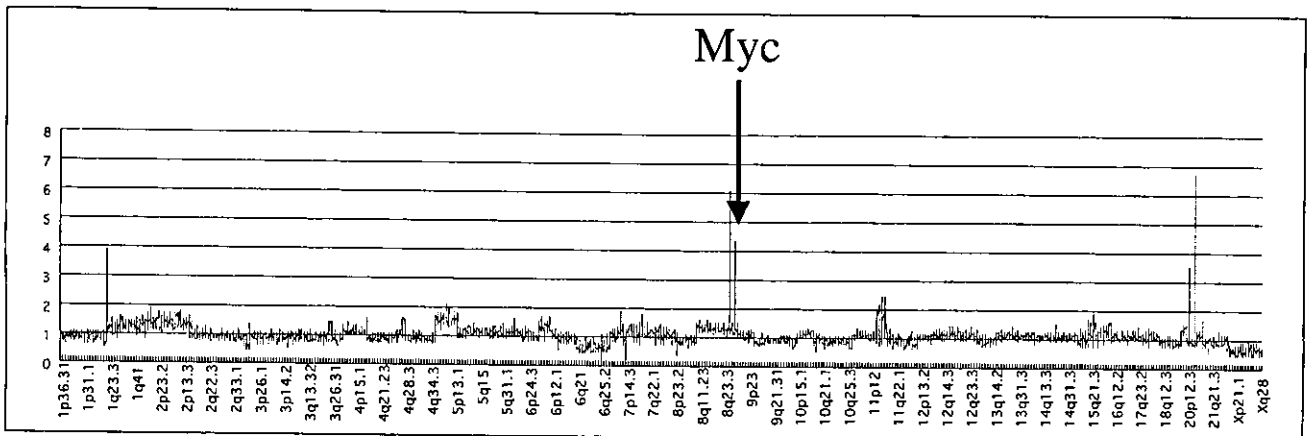


図1 DNAチップによるゲノムコピー数解析

1番染色体から性染色体まで1万箇所に対するプローブが合成されたDNAチップを用いて癌細胞DNAのゲノムのコピー数を解析した。1コピーの違いやmyc遺伝子を含む増幅部位の存在が検出される(石川, 油谷, 未発表データ)。

DNAチップ解析の実際

1. ゲノム多型解析

SNP解析への応用は1回のアレイ解析により5万箇所のSNPをタイピングできるようになり、解析コストの面でも他の測定法に対して競争力を有するようになりつつある。GeneChip (Affymetrix)は公共データベースおよびPerlegen社が同定したSNP情報に基づいて設計されている。1セット2枚のタイピングアレイを用いることにより、ゲノム全体から10万SNP、およそ20kbに1SNPのタイピングが可能である。ゲノムワイドに有用な多型マーカーを関連解析により疾患関連遺伝子を探索している現状では、多数のマーカーを高密度にタイピングする必要があるが、ハプロタイプが決定され将来的に解析すべきマーカー数が絞られた場合には、アレイによるタイピングもスクリーニングの有力手段であると思われる。家系解析の場合は、同時にゲノムワイドに1万のマーカーをタイピングできれば、200kbの解像度で解析できるので連鎖解析に十分有効である¹⁾。症例毎にタイピングできるので実際の解析の上でも実用的である。

一方、癌細胞ではがん化に先んじてゲノムの不安定化が生じていることが知られている。塩基レベルあるいは

染色体レベルでのゲノム複製の確実性が失われることにより、変異が蓄積しがん化に至る。個人の全ゲノム配列を決定することも将来的には可能になると期待されるが、現時点ではコスト的に非現実的である。コピー数に関してはアレイCGHやDNAチップを用いてかなり正確な遺伝子のコピー数の解析が可能となりつつあり²⁾³⁾、SNP測定と合わせることでゲノムワイドな染色体変異の解明が期待される(図1)。ゲノム全体について増幅領域や欠失領域の判定を行うことができることから、後述するHER2などの治療標的分子の増幅の判定への臨床応用も可能である。

2. 配列解析

a. シークエンシング

癌関連遺伝子の変異を解析するためにシークエンシングは有効な手段である。悪性黒色腫におけるBRAF遺伝子の高頻度な変異についても英国の癌ゲノムプロジェクトの成果であり⁴⁾、網羅的な解析の有効性を物語る例である。すでにp53遺伝子⁵⁾やp450遺伝子については配列決定用にマイクロアレイが用いられている。

b. タイピングアレイ

アレイの高密度化により、ゲノム配列を網羅的に配置することが現実的になりつつある。現在すでに1枚のアレイに600万種類のDNAを合成可能であるのでヒト

ゲノム配列のほぼ全てが14枚のアレイに納められる。応用範囲は極めて広く、転写領域の同定、スプライシングバリエーションの検出⁶⁾に用いることができるほか、転写因子結合部位(ChIP-chip解析)⁷⁾や複製開始点⁸⁾の解析にも用いることが試みられている。現在米国のポストゲノム計画としてENCODEプロジェクトが進められており、全ゲノム配列の1%について詳細な解析が行われている。

3. 発現解析

DNAチップには生命情報の解析手法としてのみならず、医療における応用も期待されている。腫瘍組織や疾病に罹患した組織の遺伝子発現プロファイルすなわち遺伝子転写の状態の全貌を俯瞰することにより、分子レベルで新規な癌の分類⁹⁾・診断がおこなわれ(図2)¹⁰⁾、治療への応答性および予後に関してより正確な予測にもとづいた治療法の選択、すなわち「個別化医療」の実現が期待される。その一方、ゲノム配列からの新規な治療標的分子の同定も進められている。欧米では乳癌に代表されるように癌の治療効果や予後判定のために数千人レベルでの解析も計画されている¹¹⁾。薬理ゲノミクスの立場から化学療法剤の開発において、癌細胞株パネルを用いたプロファイリングにより感受性予測への応用¹²⁾が始められているほか、創薬のプロセスの初期に必要である新規創薬標的分子の同定あるいは検証に用いられている。進行性乳癌や前立腺癌では遺伝子増幅などによりHER2遺伝子産物が過剰産生され、癌の増悪因子となっていることが知られており、抗HER2ヒト化モノクローナル抗体(ハーセプチン[®])はタキソールなどの化学療法剤との併用により生存期間、病勢進行までの期間を延長し、奏効率が向上したことが報告されている¹³⁾。特にHER2過剰発現レベルの著明に高い例では、約45%の生存期間の延長が認められている。

a. 新たな疾患単位の発見

従来の分類では異質な症例が混在する集団について、発現プロファイル解析が分子レベルでの分類や遺伝子変異の推定に有効なことがある。クラスタ分析や主成分分析などの教師なし特徴抽出の手法が用いられ、データ全

体を俯瞰するのに有用である。図2-Aは小児白血病検体をマイクロアレイデータを主成分分析したものである。MLL転座を有する白血病は他の転座を有する白血病とは分けられることが認められるものの、融合の相手遺伝子による識別は明かではない。そこでMLL転座白血病群のみのクラスタ分析を行ったところ2群に分かれることが認められた(図2-B)。さらに臨床情報との相関を検討したところ、無再発期間あるいは生存期間と相関することが認められた。予後不良群に対しては強力な治療法を選択することが求められる。

b. アレイデータによる疾患の分類、識別遺伝子の抽出

有効群と無効群、罹患群と健常群というような2群間の比較においては様々な「教師付き学習」のアルゴリズムが提唱され、それなりによい成績を収めており、その詳細は既報に詳しい¹⁴⁾¹⁵⁾。一方、3群以上の比較あるいはあらかじめ分類すべきクラス数が不明の合はまた適当なアルゴリズムがない。また、教師付き特徴抽出を行う場合にはいずれかのクラスに強制的に分類してしまうことから「偽陽性」の問題がある。1施設において解析に十分な症例数を収集することは臨床的に容易ではなく、解析コストの制約から症例数が限られることもある。同様な症例における解析結果をメタ解析する必要がある¹⁶⁾。

c. 転写ネットワークの推定

クラスタ解析により発現パターンが共通な遺伝子各々の上流に共通の転写因子結合部位が抽出できれば、遺伝子制御と機能との関係がわかり、遺伝子間相互作用あるいは転写制御ネットワークを推定することが可能である。酵母ではプロモータ配列が転写開始点の近傍に存在すること、酵母自体が転写で制御される単細胞生物であることなどから、このようなアプローチが極めて有効であった。ヒトを含む高等動物での問題点は、プロモータ配列どころか遺伝子の転写開始点の同定が不十分であり、スプライスバリエーションの存在もアレイ解析のデータのノイズとなる。様々な細胞種が混在する生体組織の解析も当然ながらデータのばらつきが増えるので、マイクロダイセクションやFACSなどにより特定の細胞種を

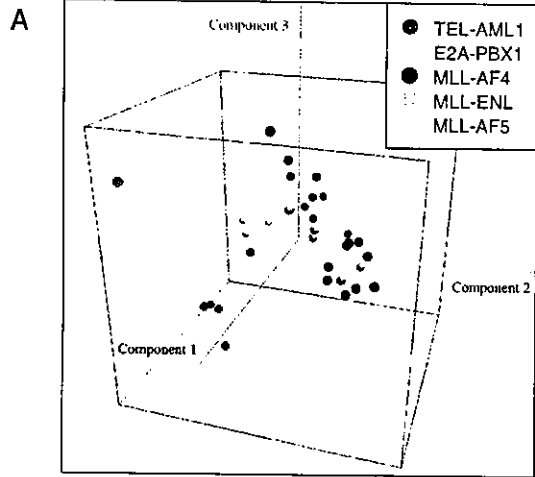
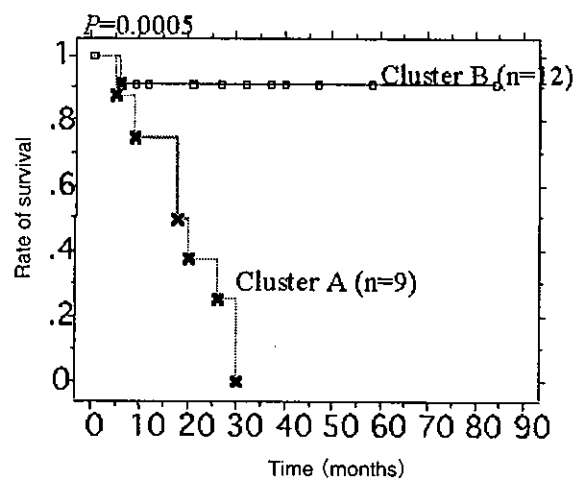
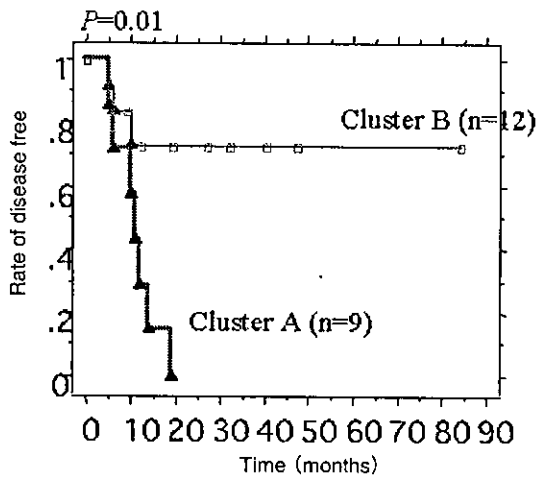
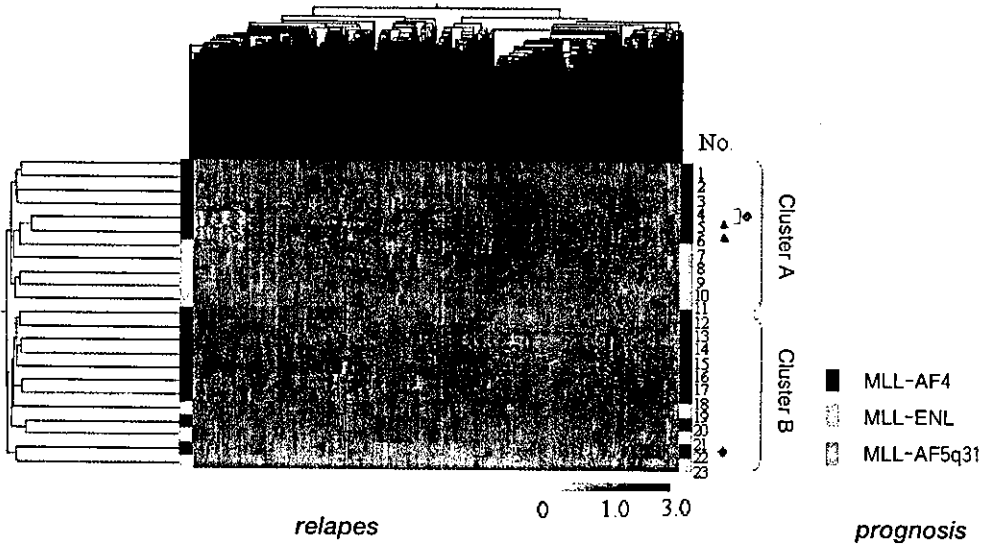


図2 アレイ解析による小児白血病における新たな病態の同定

A : 主成分分析による小児急性リンパ性白血病の分類。
 B : MLL 転座を有する急性リンパ性白血病患者より採取した骨髄について発現プロファイルをクラスタ解析したところ 2 群に分かれ、2 群間で無再発期間及び予後に有意な違いが認められた。

B



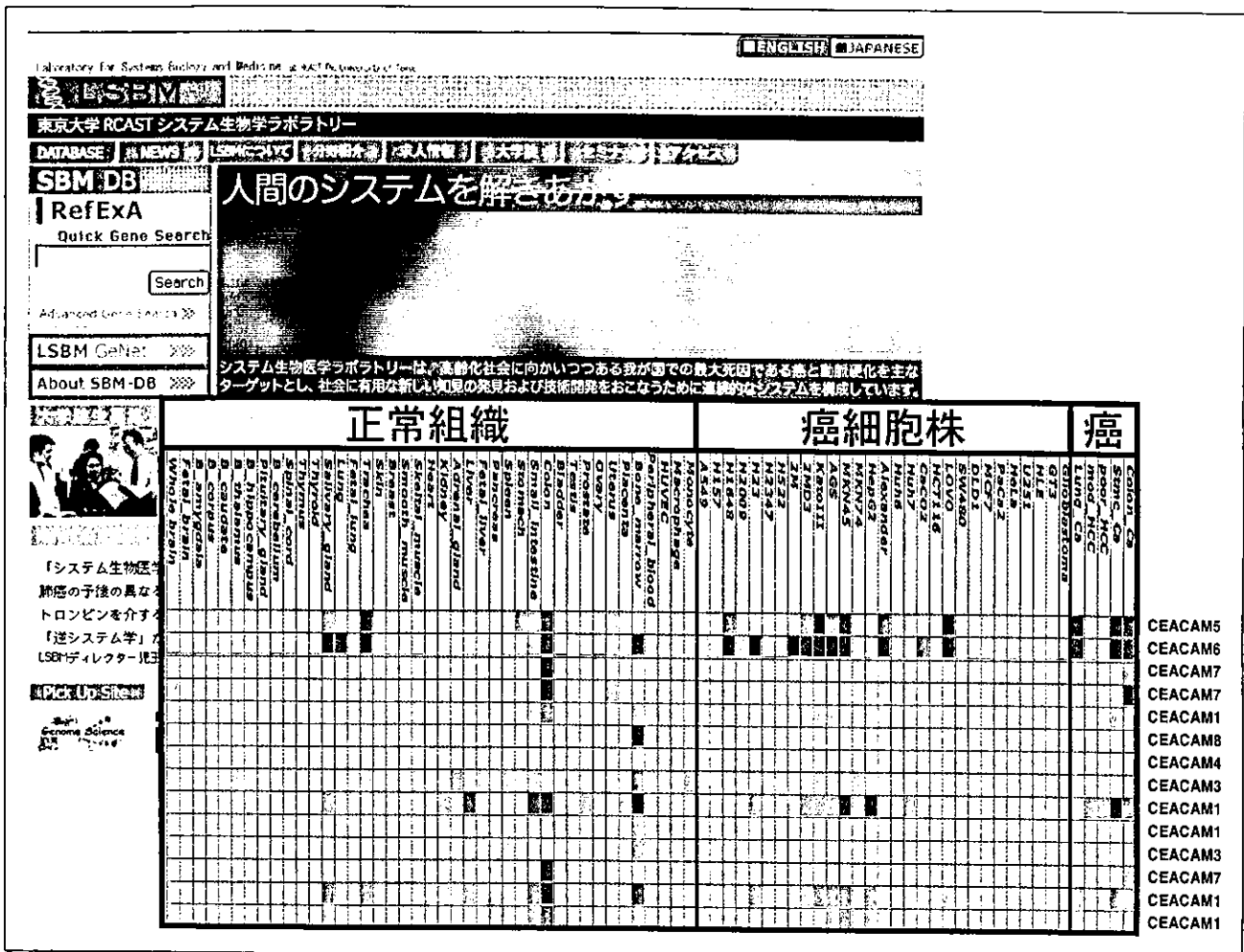


図3 遺伝子発現データベース SMB-DB

GeneChipにより測定したヒト正常組織、初代培養細胞、癌細胞株について2万遺伝子の発現データについて、システム生物医学ラボラトリーのホームページ(<http://www.lsbm.org>)からアクセスできる。遺伝子機能アノテーションについてはAffymetrix社により提供されるもの以外にもNCBIなどへのリンクがはられている。

選別することも必要である。時系列解析や *in vitro* の培養細胞系において、薬剤による介入実験をうまくデザインすることが成功の鍵になる。

d. 発現プロファイルデータベース

得られた発現プロファイル解析データのデータベース化については数年来議論されているが、解析プラットフォームの相違やプローブ配列の違いといった本質的な面のみならず、実験プロトコル、検体採取法、スキニング条件などの技術的な違いによっても解析データが異なることがあるため、複数の施設間での解析データを相互比較することは容易ではない。特に2色の蛍光を

用いる方式の場合には対照検体の選択も重要な因子となる。学術論文にアレイデータを用いる場合にはMIAME (Minimum information about a microarray experiment)¹⁷⁾に準拠することが求められるようになってきた。NCBI, EBIなどで公共データベース GEO (Gene Expression Omnibus), ArrayExpress が提案されている一方、筆者らの研究室でもヒト正常組織、初代培養細胞、癌細胞株など80検体について2万個の遺伝子を解析した結果をデータベース化して公開を行っているので御利用いただけたら幸いである(システム生物医学データベース <http://www.lsbm.org>) (図3)。

おわりに

DNAチップ解析が発現プロファイル解析のみにとどまらず、配列解析、多型解析、染色体解析、さらに転写因子結合部位などのゲノム機能解析に広く用いられることを概説した。

文 献

- 1) Kennedy GC, Matsuzaki H, Dong S, et al : Large-scale genotyping of complex DNA. *Nat Biotechnol* 21 (10) : 1233-1237, 2003
- 2) Lucito R, Healy J, Alexander J, et al : Representational oligonucleotide microarray analysis : a high-resolution method to detect genome copy number variation. *Genome Res* 13 (10) : 2291-2305, 2003
- 3) Bignell GR, Huang J, Greshock J, et al : High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res* 14 (2) : 287-295, 2004
- 4) Davies H, Bignell GR, Cox C, et al : Mutations of the BRAF gene in human cancer. *Nature* 417 : 949-954, 2002
- 5) Ahrendt SA, Hu Y, Buta M, et al : p 53 mutations and survival in stage I non-small-cell lung cancer : results of a prospective study. *J Natl Cancer Inst* 95 (13) : 961-970, 2003
- 6) Kampa D, Cheng J, Kapranov P, et al : Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res* 14 (3) : 331-342, 2004
- 7) Cawley S, Bekiranov S, Ng HH, et al : Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* 116 (4) : 499-509, 2004
- 8) Katou Y, Kanoh Y, Bando M, et al : S-phase checkpoint proteins Top1 and Mrc1 form a stable replication-pausing complex. *Nature* 424 : 1078-1083, 2003
- 9) Golub TR, et al : Molecular classification of cancer : class discovery and class prediction by gene expression monitoring. *Science* 286 (5439) : 531-537, 1999
- 10) van de Vijver MJ, et al : A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347 (25) : 1999-2009, 2002
- 11) Branca M : Genetics and medicine. Putting gene arrays to the test. *Science* 300 : 238, 2003
- 12) Reinhold WC, Kouros-Mehr H, Kohn KW, et al : Apoptotic susceptibility of cancer cells selected for camptothecin resistance : gene expression profiling, functional analysis, and molecular interaction mapping. *Cancer Res* 63 (5) : 1000-1011, 2003
- 13) Slamon DJ, et al : Use of Chemotherapy plus a Monoclonal Antibody against HER2 for Metastatic Breast Cancer That Overexpresses HER2. *N Engl J Med* 344 : 783-792, 2001
- 14) 井原茂男, 堤 修一, 油谷浩幸 : トランスクリプトーム解析のためのバイオインフォマティクス ゲノム医学 3 (2) : 71-80, 2003
- 15) 加藤菊也, 石井 信 : 遺伝子発現プロファイルのデータ解析. 蛋白質核酸酵素 48 (16) : 2300-2309, 2003
- 16) Moreau Y, Aerts S, De Moor B, et al : Comparison and meta-analysis of microarray data : from the bench to the computer desk. *Trends Genet* 19 (10) : 570-577, 2003
- 17) Brazma A, Hingamp P, Quackenbush J, et al : Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 29 (4) : 365-371, 2001

1. マイクロアレイの癌治療研究への応用

油谷浩幸

マイクロアレイは網羅的な解析技術としてRNAの変動の解析（トランスクリプトーム）に用いられるのみならず、多型解析やゲノムの機能解析までさまざまな解析に応用されている。SNP解析アレイは家族集積例の遺伝解析への応用に加えて、ゲノムワイドな遺伝子増幅やLOHなどの染色体変異の解明が期待される。また、ゲノムタイリングアレイは新規転写産物の同定や転写因子結合部位（ChIP_chip解析）や複製開始点、後生的変異の解析にも用いられる。発現プロファイルから分子レベルで新規な癌の分類・診断が行われ、治療への応答性および予後に関してより正確な予測に基づいた治療法の選択、すなわち「個別化医療」の実現が期待される。

はじめに

ゲノムプロジェクトの成果としてヒトゲノム配列（ハブアレイあたり30億塩基対）が明らかにされ、遺伝子あるいはタンパク質といった人体を構成する要素がほぼ明らかにされた。ヒト遺伝子は3万あまりと予測されるが、マウス、ラットとの比較でも「要素」となる遺伝子の数には大きな違いはない。ゲノム配列か

らRNAあるいはタンパクが発現してくる時間的、空間的な制御機構の解明が求められている。ゲノム配列中のタンパク質をコードする配列は1.5%程度であるが、ヒトとマウスの間で相同性がみられる領域は5%程度認められ、非翻訳領域の配列により器官あるいは発生時期それぞれに特異的な転写の制御が行われている。癌細胞においてもこれらの転写制御の破綻が細胞増殖の異常をもたらしている。

一方、少量の生体試料からでもDNA、RNA、タンパクという生命情報の各要素を高速に網羅的に大量取得し解析を行えるようになったことが生命科学研究にパラダイムシフトをもたらしており、Clinical Genomicsの実現が可能になった。なかでもマイクロアレイは、網羅的な解析技術としてRNAの変動の解析（トランスクリプトーム）に用いられるのみならず、多型解析やゲノムの機能解析までさまざまな解析に応用されている。これらの情報に基づいて疾患の「個性」についての理解が加速され、診断あるいは治療に有効なバイオマーカーの開発が期待される。

【キーワード&略語】

マイクロアレイ、LOH解析、タイリングアレイ、ChIP_chip解析、転写ネットワーク、遺伝子発現データベース

ChIP: chromatin immunoprecipitation (クロマチン免疫沈降)

LOH: loss of heterozygosity (ヘテロ接合性の消失)

ROH: retention of heterozygosity (ヘテロ接合性の保持)

SNP: single nucleotide polymorphism (一塩基多型)

Application of microarray technology in cancer research

Hiroiyuki Aburatani: Genome Science Division, Research Center for Advanced Science and Technology, University of Tokyo (東京大学先端科学技術研究センターゲノムサイエンス分野)

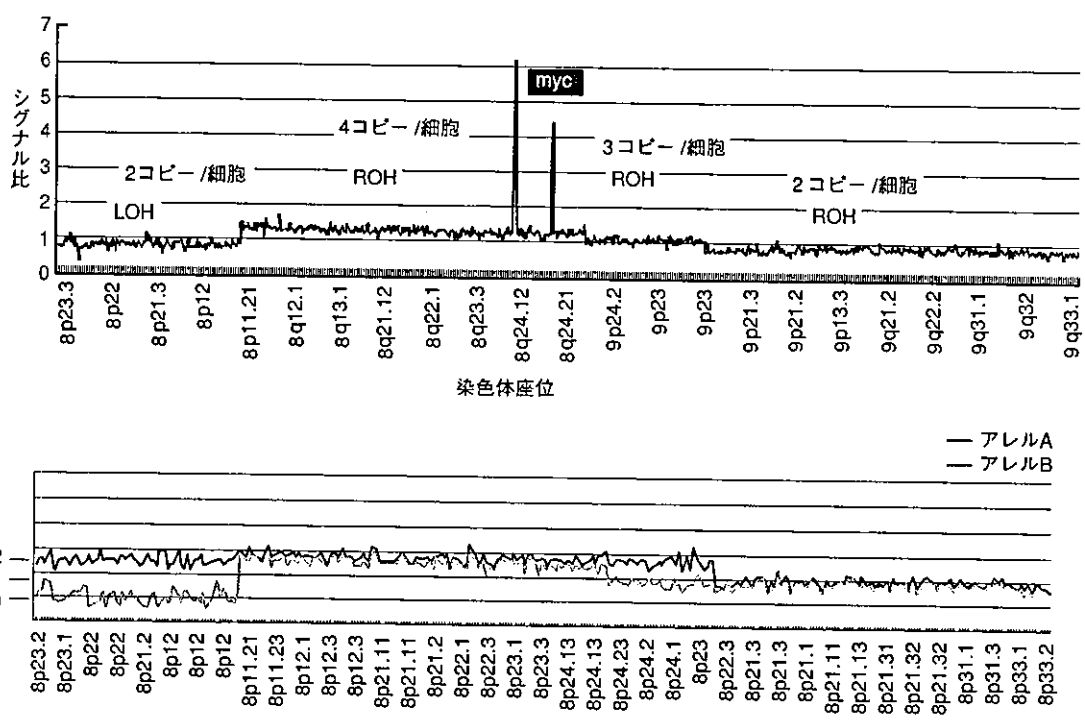


図1 DNAチップによるアレル別のゲノムコピー数解析
1番染色体から性染色体まで1万箇所に対するプローブが合成されたDNAチップを用いて癌細胞DNAのゲノムのコピー数を解析した。myc遺伝子を含む増幅部位の存在が検出されるほか、アレル別にコピー数を測れることからLOHの状態も判定できる(石川, 油谷, 未発表データ)

■ マイクロアレイによるゲノム多型解析

1) 多型解析

SNP解析への応用によって1回のアレイ解析により5万箇所のSNPをタイピングできるようになり、ゲノム全体から10万SNP、およそ20 kbに1 SNPのタイピングが可能である。解析コストの面でもほかの測定法に対して競争力を有するようになりつつある。Affymetrix社製のタイピングアレイは公共データベースおよびPerlegen社が同定したSNP情報に基づいて設計されている。ゲノムワイドに有用な多型マーカーから関連解析により疾患関連遺伝子を探索している現状では多数のマーカーを高密度にタイピングする必要があるが、ハプロタイプが決定され将来的に解析すべきマーカー数が絞られた場合にはアレイによるタイピングもスクリーニングの有力手段であると思われる。家族集積例などの家系解析の場合は、同時にゲノムワイドに1万のマーカーをタイピングできれば、200 kb

の解像度で解析できるので連鎖解析に十分有効であるし¹⁾、症例ごとにタイピングできるので実際の解析の上でも実用的である。

2) 染色体異常解析

一方、癌細胞では癌化に先んじてゲノムの不安定化が生じていることが知られている。塩基レベルあるいは染色体レベルでのゲノム複製の確実性が失われることにより、変異が蓄積し癌化に至る。個人の全ゲノム配列を決定することも将来的には可能になると期待されるが、現時点ではコスト的に非現実的である。コピー数に関してはBAC(大腸菌人工染色体)クローンを配列したBACマイクロアレイを用いたアレイCGHやDNAチップを用いてかなり正確な遺伝子のコピー数の解析が可能となりつつあり^{2) 3)}、SNP測定と合わせるによりアレルごとにゲノムワイドな染色体変異の解明が期待される(図1)。さらに高密度オリゴヌクレオチドアレイの場合は高解像度であることが長所であり、小さな欠失領域や重複した配列の同定が

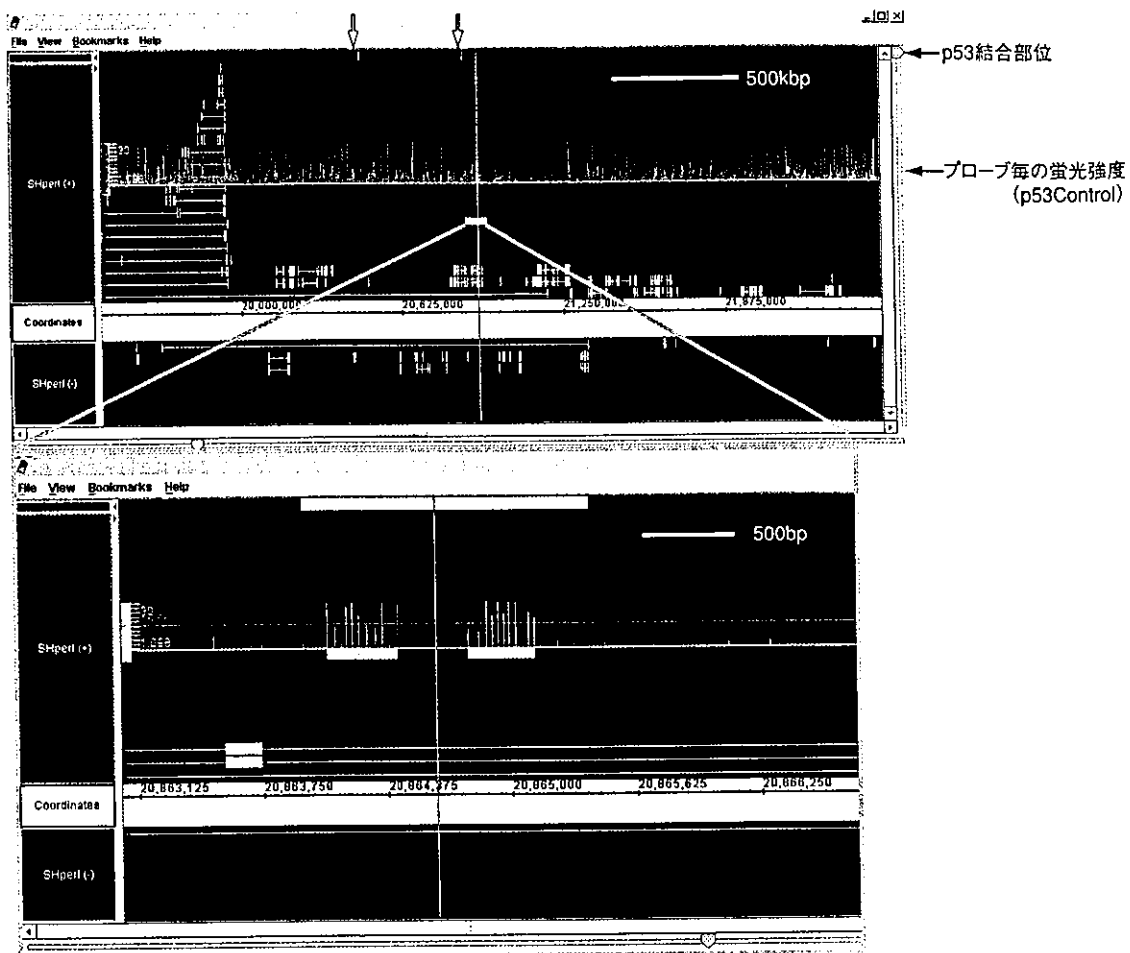


図2 ChIP_chip解析データのブラウザー表示 (巻頭写真1-1参照)
p53による結合配列データをゲノムブラウザー内に表示したものである。連続する10前後のプロープのシグナル強度が有意に高いことから結合の有無を判定する

可能である。少量の検体から短時間でゲノム全体について増幅および欠失領域の判定を行うことができることから、現行のFISH法やBACマイクロアレイによる診断に替わる技術として期待される。現在Herceptin®などの分子標的医薬は標的分子の増幅を確認することが投与に際しての条件とされており、HER-2などの治療標的分子の増幅の判定への臨床応用が考えられる。

2 配列解析

1) シークエンシング

癌関連遺伝子の変異を解析するためにシークエンシングは有効な手段である。悪性黒色腫におけるBRAF

(v-raf murine sarcoma viral oncogene homolog B1) 遺伝子の高頻度な変異の検出についても英国の癌ゲノムプロジェクトの成果であり⁴⁾、網羅的な解析の有効性を物語る例である。すでにp53遺伝子変異⁵⁾やp450遺伝子多型については配列決定用にマイクロアレイが用いられている。最近ではIressa[®]の有効症例にはEGFR (epidermal growth factor receptor) 遺伝子に変異を生じている例が多いことが報告された^{6) 7)}。発癌関連遺伝子について配列変異の有無を調べることが重要であることを物語る。

2) タイリングアレイ

アレイの高密度化により、ゲノム配列を網羅的に配

置ることが現実的になりつつある。応用範囲はきわめて広く、転写領域の同定、スプライシングバリエーションの検出⁸⁾に用いることができるほか、転写因子結合部位 (ChIP_chip解析^{*1)}⁹⁾や複製開始点¹⁰⁾の解析にも用いることが試みられている。bleomycin 刺激後に p53 の結合配列の解析から、第21, 22染色体上に48カ所の結合部位が同定された。われわれも、adriamycin などのほかの薬剤による刺激でもそれらの結合部位がChIP後に濃縮されることを定量PCRおよびチップ解析で再現性よく確認することができた (図2, 未発表データ)。p53のゲノムへの結合と近傍に位置する遺伝子の転写制御との関係について今後解析を進めていく予定である。

3 発現解析

マイクロアレイには医療における応用も期待されており、腫瘍組織や疾病に罹患した組織の遺伝子発現プロファイルすなわち遺伝子転写の状態の全貌を俯瞰することにより、分子レベルで新規な癌の分類¹¹⁾・診断が行われ (図3)¹²⁾、治療への応答性および予後に関してより正確な予測に基づいた治療法の選択、すなわち「個別化医療」の実現が期待される。欧米では乳癌に代表されるように癌の治療効果や予後判定のために数千人レベルでの解析も計画されている¹³⁾。薬理ゲノミクス (1章-3参照) の立場から化学療法剤の開発において癌細胞株パネルを用いたプロファイリングにより感受性予測への応用¹⁴⁾が始められているほか、創薬のプロセスの初期に必要な新規創薬標的分子の同定あるいは検証に用いられている。進行性乳癌や前立腺癌では遺伝子増幅などによりHER-2遺伝子産物が過剰産生され、癌の増悪因子となっていることが知られており、抗HER-2ヒト化モノクローナル抗体

(Herceptin[®]) はタキソールなどの化学療法剤との併用により生存期間、病勢進行までの期間を延長し、奏効率が向上したことが報告されている¹⁵⁾。特にHER-2過剰発現レベルの著明に高い例では、約45%の生存期間の延長が認められている。

1) 新たな疾患単位の発見

従来の分類では異質な症例が混在する集団において、発現プロファイル解析を用いることは分子レベルでの分類や遺伝子変異の推定に有効な場合があるほか、転移をきたす腫瘍の予測にも用いられている¹⁶⁾。クラスター分析や主成分分析などの教師なし特徴抽出^{*2)}の手法が用いられ、データ全体を俯瞰するのに有用である。図3Aは小児白血病検体のマイクロアレイデータを主成分分析したものである。MLL転座^{*3)}を有する白血病は他の転座を有する白血病とは識別されることが認められるものの、予想に反して融合遺伝子による識別は明らかではなかった。そこでMLL転座白血病群のみのクラスター分析を行ったところ2群に分かれることが認められた (図3B)。さらに臨床情報との相関を検討したところ、無再発期間あるいは生存期間と相関することが認められ、予後不良群に対しては早期より強力な治療法を選択することが求められる。

2) アレイデータによる疾患の分類、識別遺伝子の抽出

有効群と無効群、罹患群と健常群というような2群間の比較においてはさまざまな「教師つき学習」のアルゴリズムが提唱され、それなりにより成績を収めており、その詳細は既報に詳しい¹⁷⁾¹⁸⁾。一方、3群以上の比較あるいはあらかじめ分類すべきクラス数が不明の場合はまだ適当なアルゴリズムがない。また、教師つき特徴抽出を行う場合にはいずれかのクラスに強制的に分類してしまうことから「疑陽性」の問題があ

※1 ChIP_chip 解析

クロマチン免疫沈降法 (chromatin immunoprecipitation: ChIP) は個々のタンパクとゲノムDNAとの結合を検出する方法として用いられてきた。DNAチップ技術の高密度化により、全ゲノム配列を網羅的に合成、配置したタイリングアレイが使用可能となりつつある。DNA結合配列の同定やクロマチン構造の変換を解析するために用いられる。

※2 教師つき (なし) 特徴抽出

マイクロアレイデータのような高次元のデータ空間をマイングする機械学習の手法として、教師つき学習 (supervised learning) と教師なし学習 (unsupervised learning) に分

けられる。前者はサンプルに分類名を与えることにより、識別するための変数の組み合わせを知ろうとするのに対し、後者は全てのサンプルにわたり共通の制御を受ける遺伝子群の発見を目的とする。

※3 MLL 転座

Mixed-Lineage Leukemia (MLL) 遺伝子は Trithorax-related chromatin-modifying protooncogene をコードしており、ポリコム群 (PcG) 遺伝子と協調して高次クロマチン構造の制御に動き、Hox 遺伝子群の活動を制御している。MLL 遺伝子が位置する 11q23 は造血器腫瘍、とりわけ乳児白血病における染色体転座の主な切断点の1つである。

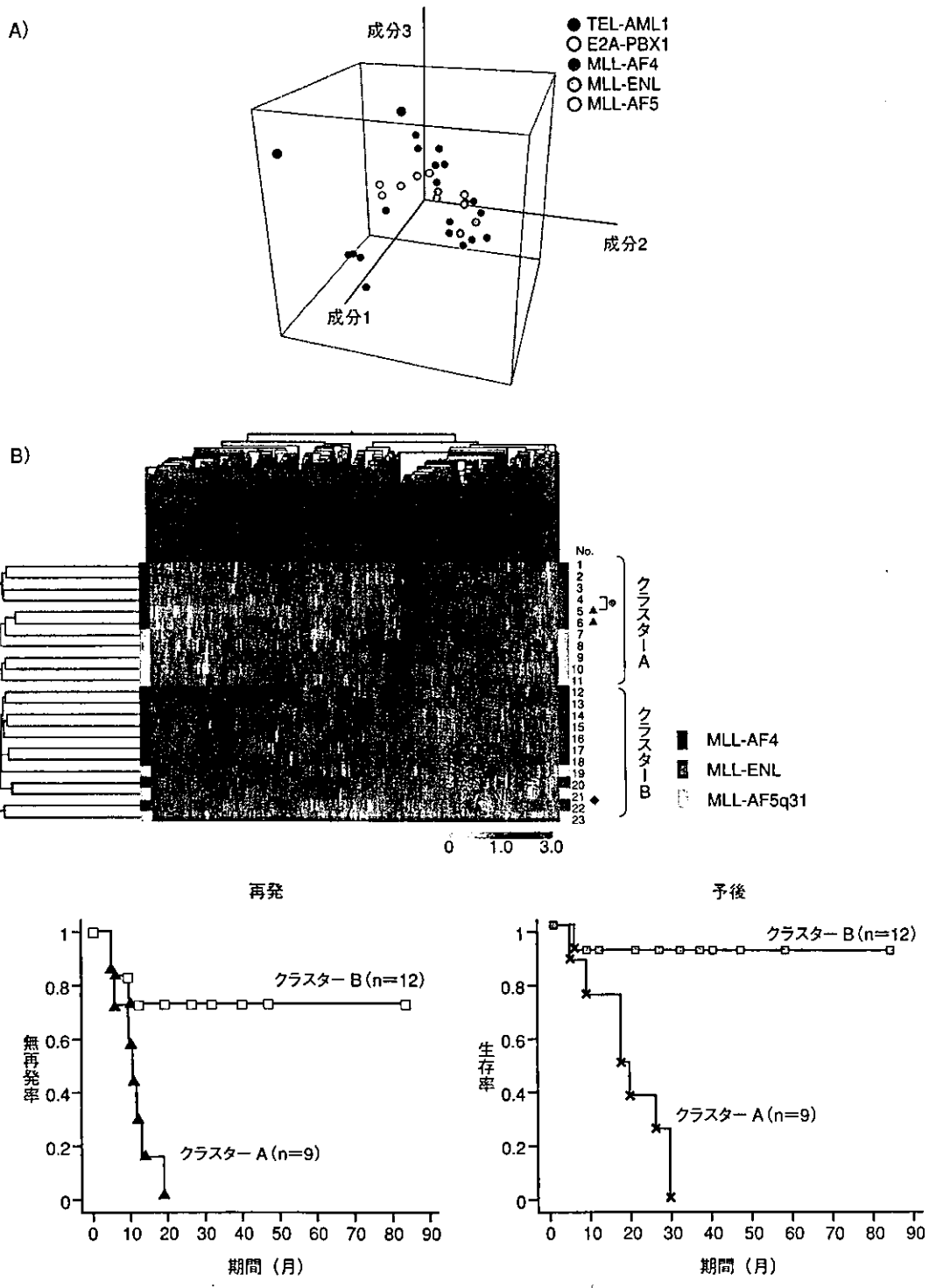


図3 アレイ解析による小児白血病における新たな病態の同定 (巻頭写真1-2参照)
 A) 主成分分析による小児急性リンパ性白血病の分類, B) MLL転座を有する急性リンパ球性白血病患者より採取した骨髄について発現プロファイルをクラスタ解析したところ, 2群に分かれ, 2群間で無再発期間 ($P = 0.01$) および予後 ($P = 0.0005$) に有意な違いが認められた (文献21より引用改変)

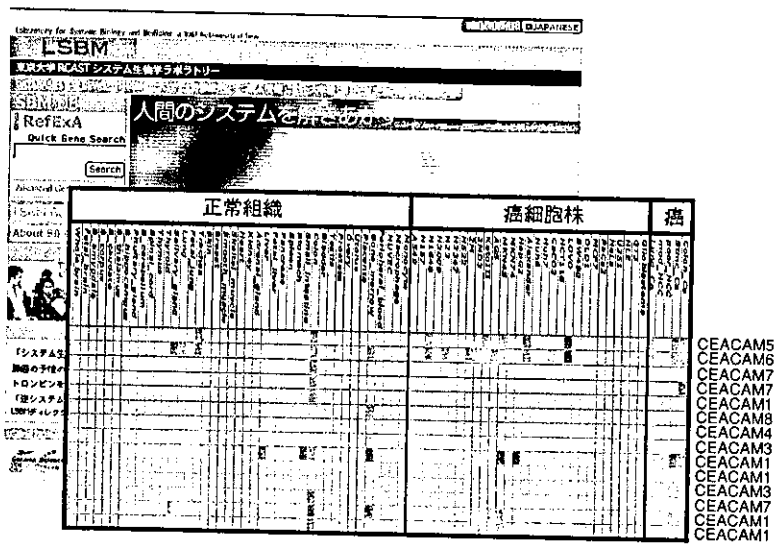


図4 遺伝子発現データベース SMB-DB
GeneChipあるいはCode Linkアレイにより測定したヒト正常組織、初代培養細胞、癌細胞株について2万遺伝子の発現データに、システム生物医学ラボラトリーのホームページ (<http://www.lsbm.org/>) からアクセスできる。遺伝子機能アノテーションについてはAffymetrix社により提供されるもの以外にもNCBIなどへのリンクがはられている

る。1施設において解析に十分な症例数を収集することは臨床的に容易ではなく、解析コストの制約から症例数が限られることもある。同様な症例における解析結果をまとめて統計的手法で再解析する（メタ解析）必要がある¹⁹⁾。

3) 転写ネットワークの推定

クラスター解析により発現パターンが共通な遺伝子のおおの上流に共通の転写因子結合部位が抽出できれば、遺伝子制御と機能との関係がわかり、遺伝子間相互作用あるいは転写制御ネットワークを推定することが可能である。酵母ではプロモーター配列が転写開始点の近傍に存在すること、酵母自身が転写で制御される単細胞生物であることなどから、このようなアプローチがきわめて有効であった。ヒトを含む高等動物での問題点は、プロモーター配列どころか遺伝子の転写開始点の同定が不十分であり、スプライスバリエーションの存在もアレイ解析のデータのノイズとなる。さまざまな細胞種が混在する生体組織の解析も当然ながらデータのばらつきが増えるので、マイクロダイセクションやフローサイトメトリーなどにより特定の細胞種を選別することも必要である。時系列解析や*in vitro*の培養細胞系において薬剤による介入実験をうまくデ

ザインすることが成功の鍵になる。

4) 発現プロファイルデータベース

得られた発現プロファイル解析データのデータベース化については数年来議論されているが、解析プラットフォームの相違やプローブ配列の違いといった本質的な面のみならず、実験プロトコール、検体採取法、スキャニング条件などの技術的な違いによっても解析データが異なることがあるため、複数の施設間での解析データを相互比較することは容易ではない。特に2色の蛍光を用いる方式の場合には対照検体の選択も重要な因子となる。学術論文にアレイデータを用いる場合にはMIAME (minimum information about a microarray experiment)²⁰⁾に準拠して公開することが求められるようになり、NCBI、EBIなどで公共データベースGEO (Gene Expression Omnibus)、Array-Expressが提案されている。われわれの研究室でもヒト正常組織、初代培養細胞、癌細胞株など80検体について2万個の遺伝子を解析した結果をデータベース化して公開を行っているのでご利用いただけたら幸いである (システム生物医学データベース: <http://www.lsbm.org/>) (図4)。