

the *c-fos* gene is strongly upregulated by MITF in primary melanocytes (McGill *et al.*, 2002). In agreement with this notion, STAT3 was frequently phosphorylated in several human melanoma cell lines (Figure 3a). However, while B16 and SKMel cells are negative for STAT3 phosphorylation, the same cells do express *c-fos*. Apparently, STAT3-independent mechanism of *c-fos* induction is present in these melanoma cells. Finding such mechanisms of *c-fos* expression in these cells may provide a new clue for understanding the role of *c-fos* for the generation of melanomas.

Promoter analysis and ChIP assay indicated that STAT3 and MITF can bind to the *c-fos* promoter and activate this promoter independently. Reporter gene analysis suggested that the effect of MITF and STAT3 on the *c-fos* promoter are additive rather than cooperative (Figure 5b). This is consistent with our observation that there is no direct physical interaction between MITF and STAT3. Nevertheless, high levels of *c-fos* expression and transformation of NIH-3T3 cells are dependent on both STAT3 and MITF. In addition to promoter activation, another mechanism of *c-fos* expression, such as stabilization of *c-fos* mRNA, may present in the cooperative effect of STAT3 and MITF. Furthermore, a previous report demonstrated that wt- and mi-MITF bind to the c-Fos protein. In particular, mi-MITF prevents c-Fos from being transported to the nucleus, and this may be a reason for the osteopetrosis of mi/mi mutant mice (Sato *et al.*, 1999). wt-MITF was also shown to bind to c-Jun and enhance the transactivation of the MMCP-7 gene, while complexes of mi-MITF and c-Jun were predominantly found in the cytoplasm and suppressed transactivation (Ogihara *et al.*, 2001). Thus, MITF could be involved in AP-1 activation not only by inducing *c-fos* but also by interacting with AP-1 itself. The fact that a dominant-negative AP-1 (SupJunD) suppressed colony formation of transformants expressing both MITF and STAT3C suggests that AP-1 can be a therapeutic target of melanoma.

In addition to *c-fos*, we found several interesting genes that are cooperatively induced by STAT3 and MITF (Figure 4a). GTP cyclohydrolase-1 and mast cell proteases are probably involved in melanocyte and mast cell functions. PMEL17 is also known as an MITF-inducible pigment cell-specific gene and a melanosomal matrix protein, which may function as a structural protein in melanogenesis (Kobayashi *et al.*, 1994). JE and A7 are chemokines that are induced by immunoglobulin (Ig) E plus antigen stimulation in mast cells. (Burd *et al.*, 1989; Kulmburg *et al.*, 1992; Ong *et al.*, 1993). Therefore, STAT3/MITF target genes are strongly linked to the function of mast cells and melanocytes, which constitutively express MITF. This finding implies that STAT3 (or other STATs) may be involved in mast cell and melanocyte functions by inducing genes in cooperation with MITF. It is interesting that IL3, which activates STAT5, is a mast cell growth factor *in vitro*. The functions of STATs in mast cells and melanocytes are under investigation.

Materials and methods

Cell culture and transfection

NIH3T3 and HEK293 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% CS. STAT3C-transformed 3T3 (STAT3C-3T3) cells (Yoshida *et al.*, 2002) were cultured in DMEM supplemented with 10% CS containing 0.8 mg/ml G418. PLAT-E, a packaging cell line (Morita *et al.*, 2000) was maintained in DMEM with 10% FCS containing blasticidin S (10 mg/ml) and puromycin (1 mg/ml). Melanoma cell lines were cultured in the following medium containing 10% FCS, B16 and B16F10 in RPMI 1640, G361 and SKMel28 in MEM, HMV-II in Ham's F12, and MMac in DMEM. MC9, a mast cell line, was cultured in RPMI 1640 supplemented with 5% FCS, 50 μ M 2-mercaptoethanol, and 6 ng/ml mouse recombinant IL-3. Cells were transfected by the calcium phosphate method with Cell Pfect (Amersham) or by the lipofection method with FuGENE 6 (Roche). For retrovirus-mediated gene expression, NIH3T3 cells were infected with the retroviruses produced by PLAT-E as reported (Sasaki *et al.*, 2003). B16, B16F10, G361, and SKMel 28 cell lines were kindly provided by Dr Yonemitsu (Kyushu University, Japan).

Library screening

Retroviruses containing the human HeLa retroviral cDNA library (CLONTECH, Palo Alto, CA, USA) were produced from PLAT-E and 1×10^6 STAT3C-3T3 cells were infected with 10 ml of the retroviral supernatant containing 10 mg/ml polybrene. After 48 h of virus transduction, STAT3C-3T3 cells were seeded in soft-agar medium. Within 3 weeks, two colonies were picked up and expanded, and genomic DNA was isolated. Integrated cDNAs were recovered by PCR using a pair of primers (FWD, 5'-AGCCCTCACTCCTTCTCTAG, and REV, 5'-ATGGCGTACTTAAGCTAGCTTGC-CAAACCTAC) and sequenced.

Colony formation in soft-agar

The MITF-infected 3T3 and STAT3C-3T3 cells were seeded into 35-mm dishes in suspensions of 0.36% Agar noble (Difco) in DMEM supplemented with 10% CS on top of a bed of 0.72% Agar noble in the same complete medium. The MITF-infected STAT3C-3T3 cells were inoculated with a retrovirus of dominant-negative AP-1 or an empty vector (Ui *et al.*, 2000), selected with puromycin, and plated into soft-agar medium. The cultures were incubated for 21 days and then the colonies were counted and photographed.

Construction of MITF

The cDNAs of wt-, Δ N-, and mi-MITF were subcloned into a bicistronic retrovirus vector pMX-IRES-EGFP (Nosaka *et al.*, 1999) or a pcDNA3 expression vector.

Retrovirus of dominant-negative AP-1

The control virus (pBabe-IRESpuro) and the SupJunD-1 virus (pBabe-SupJunD-1-IRESpuro) as a dominant-negative AP-1 were generated as described (Ui *et al.*, 2000). After infection, STAT3C-3T3 cells were cultured with 2 mg/ml puromycin.

High-density oligonucleotide microarray analysis

RNA was extracted by standard methods. Cells were lysed directly in their Petri dishes in TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and total RNA was isolated according to

the manufacturer's instructions. cRNA preparation and microarray hybridization were carried out according to the supplier's instructions (Affymetrix, Santa Clara, CA, USA), using Genechip HG-U95Av2. Scanned output files were analysed by the probe level analysis package, Microarray Suite MAS 5.0 (Affymetrix, Santa Clara, CA, USA). The signal for each of these genes was determined from the 'probe set' in use for this gene and by the probe level analysis method provided by Affymetrix software.

Northern Blotting analysis and RT-PCR

Total RNA (10 µg) extracted from cells using TRIzol was evaluated by Northern blotting analysis with digoxigenin-labeled antisense RNA of *c-fos* and G3PDH labeled with the DIG RNA Labeling Kit (Roche). To determine the microarray data, 1 µg of total RNA was reverse-transcribed with the Reverse Transcribed Kit (Roche) and RT-PCR was performed with primers as follows: mouse MITF, 5'-GGAACAGCAAC-GAGCTAAGG and 5'-TGATGATCCGATTACACCAGA; human MITF, 5'-AGAACAGCAACGCGCAAAAGAAC and TGATGATCCGATTACCAAATCTG; mouse GTP cyclohydrolase, 5'-GGCTGCTTACTCGTCCATTC and AG GTGATGCTCACACATGGA; mouse JE, AGGTCCCTGT CATGCTTCTG and TCTGGACCCATTCCCTTCTTG; mouse PMEL17 CAGGGGTCTAACTGCTGGAG and TT CGGAGGTTTAGGACCAGA; mouse MRP8, GGAAAT CACCATGCCCTCTA and TGGCTGTCTTTGTGAGAT GC; mouse MARC, TCTGTGCTGCTGCTCATAG and CTTTGGAGTTGGGGTTTCA; and mouse MCPL, GCA CTTCTCTTGCCTTCTGG and TGTGCAGCAGTCATCA CAAA.

Western blotting analysis

Cells were lysed in a lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5% or 1% NP-40, 1 mM EDTA, 1 mM vanadate, 50 mM NaF, 1 mM DTT, 0.01 mM APMSF) and centrifuged at 12000 g for 10 min. The supernatants were resolved on SDS-PAGE of 10% gels, blotted, and immunostained with an anti-phospho STAT3 (Tyr705) polyclonal antibody (Cell Signaling Technology) and an anti-STAT3 polyclonal antibody (Santa Cruz Biotechnology).

Luciferase assay

An APRE-luciferase reporter gene for STAT3 activity and a reporter gene construct containing MITF-responsive mMCP-6 (mouse mast cell protease) luciferase have been described previously (Morii et al., 1996; Yasukawa et al., 1999). A *c-fos* promoter-reporter gene consisting of the *c-fos* promoter region containing SIE, SRE, and several potential MITF-binding motifs (Tsujiura et al., 1996) is also described (Shibuya et al., 1994; Kawahara et al., 1995; Kim et al., 1998). Luciferase assays were carried out using the dual-luciferase reporter system (Promega). The expression plasmid of STAT3-C in pRcCMV was kindly provided by Dr JE Darnell Jr (Bromberg et al., 1999).

Nuclear extract preparation and oligonucleotide-binding assay

293T cells (1.5×10^7) were transfected with Myc-ΔN-MITF and STAT3C were stimulate with LIF for 6 h. Cells were collected and resuspended in 0.4 ml of buffer A (10 mM HEPES-KOH (pH 7.8), 10 mM KCl, 0.1 mM EDTA, 0.4%

NP-40, 1 mM DTT, 0.5 mM APMSF, 2 µg/ml leupeptin, and 1 mM Na₃VO₄). After a brief vortexing and centrifugation, the supernatant was discarded and the nuclei-containing pellet was resuspended in 0.05 ml of buffer C (50 mM HEPES-KOH (pH 7.8), 420 mM NaCl, 5 mM MgCl₂, 0.1 mM EDTA, 2% glycerol, 1 mM DTT, 0.5 mM APMSF, 2 µg/ml leupeptin, and 1 mM Na₃VO₄) at 4°C for 30 min. The suspension was pelleted by centrifugation and the supernatants were collected and stored at -80°C until use. Oligonucleotide-conjugated beads were prepared using the DNA-binding protein purification kit (Roche) according to the manufacturer's protocol. In brief, the sense and antisense oligonucleotide of the human *c-fos* promoter sequence, including the SIE and SRE regions (5'-C AGTTCCTCGTCAATCCCTCCCCCTTACACAGGATGTC CATATTAGGACATCTGC-3') were annealed, and they were ligated with the washed streptavidin magnetic particles for 30 min at 25°C. After ligation was carried out, the particles were washed using a magnetic separator and mixed with 100 µg of the nuclear extract for 30 min at 25°C. Then, the particles were washed and the DNA-binding proteins were eluted. The eluates were pelleted with trichloroacetic acid solution, and analysed by immunoblotting with anti-STAT3 and anti-Myc antibodies.

ChIP assay

For ChIP assay, 243T cells (5×10^6) were fixed with 1% formaldehyde for 10 min at 37°C, washed with PBS, and lysed in ChIP lysis buffer (Upstate Biotechnology, Lake Placid, NY, USA). DNA was sonicated by pulsing five times. Anti-acetylated histone H4 (Upstate), anti-STAT3 (Sigma), or anti-MITF monoclonal antibody (Oncogene), and anti-MITF polyclonal antibody (Santa Cruz Biotechnology) was added (5 µg-20 µg per immunoprecipitation) and incubated overnight. Protein A-agarose beads (Upstate Biotechnology) were added for 1 h and then washed once each with a low-salt buffer, high-salt buffer, and LiCl buffer. They were then washed twice with a TE buffer. The beads were eluted with 0.1 M NaHCO₃ and 1% SDS, and crosslinks were reversed at 65°C. DNA was ethanol precipitated in the presence of 20 µg glycogen. PCR was carried out using primers specific to the promoter region of mouse *c-fos* (FWD, 5'-TCTGCCTTTCCTCGCTCC, and REV, 5'-GGCCGTGGAAACCTGCTGAC) for NIH3T3, and human *c-fos* (FWD, 5'-CCCACCTCGGGAACAAGGG, and REV, 5'-ATGAGGGGTTTCGGGGATGG) for HMV-II, or (FWD, 5'-TCTCATTCTGCGCCGTCC, and REV, 5'-GGCCGTGGAAACCTGCTGAC) for G361.

Acknowledgements

We thank Dr T Yoshida (Kurume University) for STAT3C-3T3, Dr Yonemitsu (Kyushu University) for melanoma cell lines, H Meguro for oligonucleotide microarray analysis, Y Kawabata for technical assistance, and N Arifuku for manuscript preparation. This work was supported by special grants-in-aid from the Ministry of Education, Science, Technology, Sports, and Culture of Japan, the Japan Health Science Foundation, the Human Frontier Science Program, the Japan Research Foundation for Clinical Pharmacology, Haraguchi Memorial foundation, and the Uehara Memorial Foundation. AJ is supported by the fellowship from Japan Society for Promotion of Science.

References

- Bowman T, Garcia R, Turkson J and Jove R. (2000). *Oncogene*, **19**, 2474-2488.
- Bromberg J and Darnell Jr JE. (2000). *Oncogene*, **19**, 2468-2473.

- Bromberg JF, Wrzeszczynska MH, Devgan G, Zhao Y, Pestell RG, Albanese C and Darnell Jr JE. (1999). *Cell*, **98**, 295–303.
- Burd PR, Rogers HW, Gordon JR, Martin CA, Jayaraman S, Wilson SD, Dvorak AM, Galli SJ and Dorf ME. (1989). *J. Exp. Med.*, **170**, 245–257.
- Carreira S, Liu B and Goding CR. (2000). *J. Biol. Chem.*, **275**, 21920–21927.
- Darnell Jr JE. (1997). *Science*, **277**, 1630–1635.
- Hatakeyama M, Kawahara A, Mori H, Shibuya H and Taniguchi T. (1992). *Proc. Natl. Acad. Sci. USA*, **89**, 2022–2026.
- Hemesath TJ, Price ER, Takemoto C, Badalian T and Fisher DE. (1998). *Nature*, **391**, 298–301.
- Kamaraju AK, Bertolotto C, Chebath J and Revel M. (2002). *J. Biol. Chem.*, **277**, 15132–15141.
- Kawahara A, Minami Y, Miyazaki T, Ihle JN and Taniguchi T. (1995). *Proc. Natl. Acad. Sci. USA*, **92**, 8724–8728.
- Kim DW, Cheriya V, Roy AL and Cochran BH. (1998). *Mol. Cell. Biol.*, **18**, 3310–3320.
- Kim DS, Hwang ES, Lee JE, Kim SY, Kwon SB and Park KC. (2003). *J. Cell Sci.*, **116**, 1699–1706.
- Kitamura Y, Morii E, Jippo T and Ito A. (2002). *Mol. Immunol.*, **38**, 1173.
- Kitamura T, Onishi M, Kinoshita S, Shibuya A, Miyajima A and Nolan GP. (1995). *Proc. Natl. Acad. Sci. USA*, **92**, 9146–9150.
- Kobayashi T, Urabe K, Orlow SJ, Higashi K, Imokawa G, Kwon BS, Potterf B and Hearing VJ. (1994). *J. Biol. Chem.*, **269**, 29198–29205.
- Kulmburg PA, Huber NE, Scheer BJ, Wrann M and Baumruker T. (1992). *J. Exp. Med.*, **176**, 1773–1778.
- Levy C, Nechushtan H and Razin E. (2002). *J. Biol. Chem.*, **277**, 1962–1966.
- Lufei C, Ma J, Huang G, Zhang T, Novotny-Diermayr V, Ong CT and Cao X. (2003). *EMBO J.*, **22**, 1325–1335.
- Mansky KC, Marfatia K, Purdom GH, Luchin A, Hume DA and Ostrowski MC. (2002). *J. Leukocyte Biol.*, **71**, 295–303.
- Matsui T, Kinoshita T, Hirano T, Yokota T and Miyajima A. (2002). *J. Biol. Chem.*, **277**, 36167–36173.
- McGill GG, Horstmann M, Widlund HR, Du J, Motyckova G, Nishimura EK, Lin YL, Ramaswamy S, Avery W, Ding HF, Jordan SA, Jackson IJ, Korsmeyer SJ, Golub TR and Fisher DE. (2002). *Cell*, **109**, 707–718.
- Morii E, Tsujimura T, Jippo T, Hashimoto K, Takebayashi K, Tsujino K, Nomura S, Yamamoto M and Kitamura Y. (1996). *Blood*, **88**, 2488–2494.
- Morita S, Kojima T and Kitamura T. (2000). *Gene Therapy*, **7**, 1063–1066.
- Nakayama K, Kim KW and Miyajima A. (2002). *EMBO J.*, **21**, 6174–6184.
- Nosaka T, Kawashima T, Misawa K, Ikuta K, Mui AL and Kitamura T. (1999). *EMBO J.*, **18**, 4754–4765.
- Nyormoi O and Bar-Eli M. (2003). *Clin. Exp. Metastasis*, **20**, 251–263.
- Ogihara H, Morii E, Kim DK, Oboki K and Kitamura Y. (2001). *Blood*, **97**, 645–651.
- Ong EK, Griffith IJ, Knox RB and Singh MB. (1993). *Gene*, **134**, 235–240.
- Sasaki A, Inagaki-Ohara K, Yoshida T, Yamanaka A, Sasaki M, Yasukawa H, Koromilas AE and Yoshimura A. (2003). *J. Biol. Chem.*, **278**, 2432–2436.
- Sato M, Morii E, Takebayashi-Suzuki K, Yasui N, Ochi T, Kitamura Y and Nomura S. (1999). *Biochem. Biophys. Res. Commun.*, **254**, 384–387.
- Shibahara S, Takeda K, Yasumoto K, Udono T, Watanabe K, Saito H and Takahashi K. (2001). *J. Invest. Dermatol. Symp. Proc.*, **6**, 99–104.
- Shibuya H, Kohu K, Yamada K, Barsoumian EL, Perlmutter RM and Taniguchi T. (1994). *Mol. Cell. Biol.*, **14**, 5812–5819.
- Stark GR, Kerr IM, Williams BR, Silverman RH and Schreiber RD. (1998). *Annu. Rev. Biochem.*, **67**, 227–264.
- Tachibana M. (1997). *Pigment Cell Res.*, **10**, 25–33.
- Tsujimura T, Morii E, Nozaki M, Hashimoto K, Moriyama Y, Takebayashi K, Kondo T, Kanakura Y and Kitamura Y. (1996). *Blood*, **88**, 1225–1233.
- Turkson J, Bowman T, Garcia R, Caldenhoven E, De Groot RP and Jove R. (1998). *Mol. Cell. Biol.*, **18**, 2545–2552.
- Udono T, Yasumoto K, Takeda K, Amae S, Watanabe K, Saito H, Fuse N, Tachibana M, Takahashi K, Tamai M and Shibahara S. (2000). *Biochim. Biophys. Acta*, **1491**, 205–219.
- Ui M, Mizutani T, Takada M, Arai T, Ito T, Murakami M, Koike C, Watanabe T, Yoshimatsu K and Iba H. (2000). *Biochem. Biophys. Res. Commun.*, **278**, 97–105.
- Vachtenheim J, Novotna H and Ghanem G. (2001). *J. Invest. Dermatol.*, **117**, 1505–1511.
- Yasukawa H, Misawa H, Sakamoto H, Masuhara M, Sasaki A, Wakioka T, Ohtsuka S, Imaizumi T, Matsuda T, Ihle JN and Yoshimura A. (1999). *EMBO J.*, **18**, 1309–1320.
- Yoshida T, Hanada T, Tokuhisa T, Kosai K, Sata M, Kohara M and Yoshimura A. (2002). *J. Exp. Med.*, **196**, 641–653.
- Yu CL, Meyer DJ, Campbell GS, Lerner AC, Carter-Su C, Schwartz J and Jove R. (1995). *Science*, **269**, 81–83.
- Zanocco-Marani T, Bateman A, Romano G, Valentinis B, He ZH and Baserga R. (1999). *Cancer Res.*, **59**, 5331–5340.

DNA チップの臨床応用

油谷 浩幸

東京大学国際・産学共同研究センター (教授)

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→蛋白という生命情報の流れのそれぞれを高速に網羅的に大量取得することが可能となり、生命科学にパラダイムシフトをもたらしている。なかでも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ている。

DNAチップ技術

遺伝子発現を比較することは従来ディフアレンシャル・ディスプレイやサブトラクション法を用いて行われていたわけであるが、DNAチップ技術では3万個を超える遺伝子の発現レベルが同時解析可能である。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方式が主流であったが、ゲノム配列情報が解読され、おおよその遺伝子配列が明らかとなったこともあり、合成DNAによるアレイへとシフトしつつある。Agilent社はインクジェット方式で*in situ*合成を行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便利である。その他、Amersham社(商品名CodeLink)、ABI社、日立ソフトが網羅的な発現解析用のアレイを製品化している。

基板上の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。解析に必要な検体量も1遺伝子当たりに換算すると従来法よりはるかに少量で済む。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 μ g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20ng程度のトータルRNAからも比較的再現性の良いデータが得られる。現行の

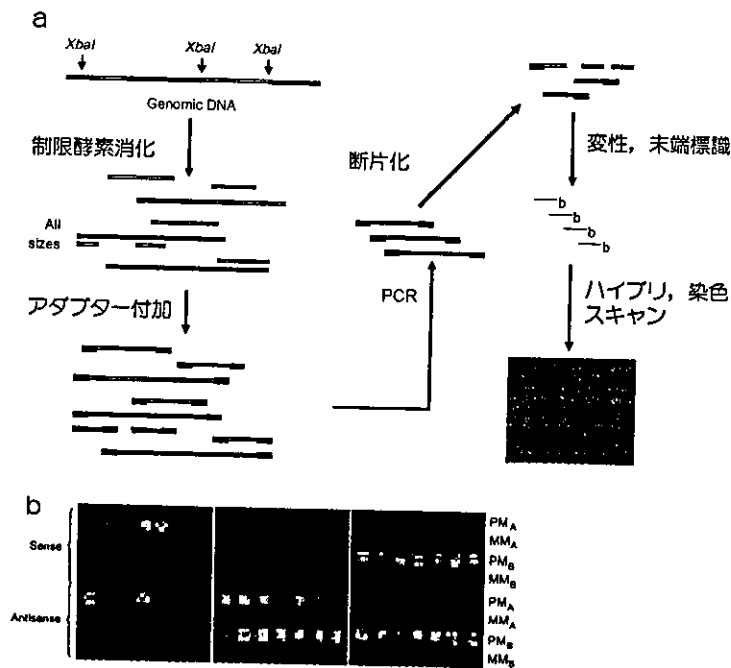


図 1. DNA チップによるゲノタイピング

a : XbaI 断片の増幅と biotin 標識

b : アレル特異的な合成 DNA プロープへのハイブリダイゼーションのイメージ

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 プロープの位置が 3' 末端寄りにデザインされていることが重要である。また、検体 RNA の品質が良いことが求められることはいうまでもなく、リボソーム RNA の 28S/18S 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あることが望ましい。1 枚のアレイでほぼすべての転写産物を測定することが可能となったものの、低発現な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二次構造の予測によりさらに特異性の高いプロープ配列の選択が進められているが、ハイブリダイゼーション実験による性能の実証も今後一層重要であろうと思われる。また、現行のアレイではプ

ロープの選択や RNA の標識法による制約から、サブ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手法の開発が今後の課題である。

DNA チップ解析の臨床応用

1. ゲノム解析

一塩基多型 (SNP) 解析への応用は他の測定法に比較して解析コストが高くなるため、アレイによるタイピングはあまり普及していない。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

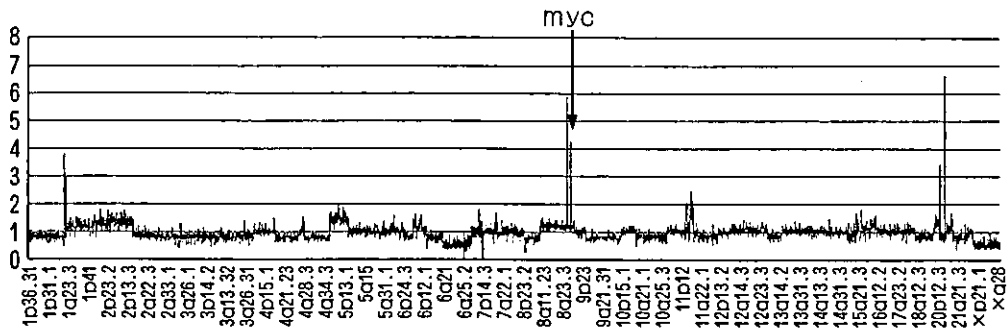


図 2. DNA チップによるゲノムコピー数解析

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)。

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競争力を有すると考えられる。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¹⁾(図 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と考えられる。

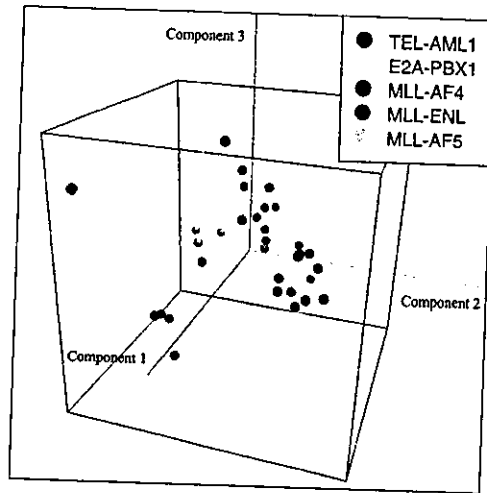
一方、癌細胞では癌化に先んじてゲノムの不安定化が生じていることが知られている。塩基レベルあるいは染色体レベルでのゲノム複製の確実性が失われることにより、変異が蓄積し癌化に至る。個人の全ゲノム配列を決定することも将来的には可能になると期待されるが、現時点ではコスト的に非現実的である。コピー数に関してはアレイ CGH や DNA チップを用いてかなり正確な遺伝子のコピー数の解析が可能となりつつあり²⁾、SNP 測定と合わせることでよりゲノムワイドな染色体変異の解明が期待される(図 2)。

2. 発現解析

DNA チップ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³⁾・診断が可能となり⁴⁾(図 3)、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一方、新規な治療標的分子の同定も進められている。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。細菌やウイルスなどの病原体を同定する感染症診断のほか、薬剤の有効性、副作用の判定にも用いることが可能である。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⁵⁾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 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 ヒトモノクロナル抗体(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⁶⁾。とくに 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



b

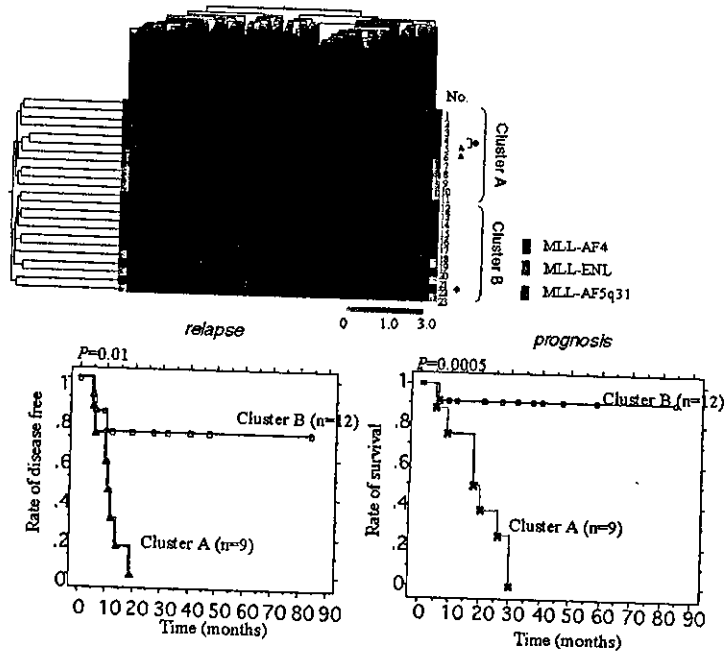


図 3. アレイ解析による小児白血病における新たな病態の同定

- 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ー解析したところ 2 群に分かれ、2 群間で無再発期間および予後に有意な違いが認められた。

糖尿病や高血圧症などのいわゆる生活習慣病は多因子性疾患であり、ゲノムおよびプロテオーム解析を駆使して疾患感受性に関する遺伝子の探索が進められている。将来的にも一つの遺伝子検査のみでは確実な診断は困難と考えられるが、これらの生活習慣病に関する治療薬は巨大な市場を有するだけに研究機関のみならず製薬企業も遺伝子の同定を競って進めており、10~20年後には一般のクリニックにおいても生活習慣の指導や薬剤の処方にもゲノム情報が応用されることが期待される。

■ DNAチップ解析の実際

統計処理を行う前にデータを実験ごとに標準化を行う必要がある。標準化に際して中間値(あるいは平均値)を算出する際には、GeneChip解析の場合には上下2%の遺伝子は除外している。アレイ全体のシグナルで標準化する場合と特定の普遍的に発現する遺伝子セットを用いる場合がある。RNA発現量に差が大きな検体間の比較には注意を要する。発現値を対数変換することでほぼ正規分布に近くなる。後述のクラスター分析の際の距離尺度にユークリッド距離を用いる場合には、対数変換をしていないと発現量の大きな遺伝子の寄与が大きくなるので注意が必要である。発現値の低い遺伝子の測定値は再現性が低いため、一般にあらかじめ解析から除外する必要があるが⁷⁾、真に変動している遺伝子をも除外してしまう危険を伴う。GeneChipでは1遺伝子あたり複数のプローブがあるので、robust multi-array average(RMA)という複数のアレイの比較からデータを標準化する手法も広く使われている⁸⁾。

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。クラスター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3a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。mixed lineage leukemia(MLL)遺伝子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らかではない。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強力な治療法を選択することが求められる。近傍法やSupport Vector Machine(SVM)法、ニューラルネットワークを用いて予後予測アルゴリズムを作成することが可能である。なお、症例数に対して常に遺伝子数が多いという状況であるので、常にoverfittingの危険を伴うことを忘れてはいけない。予測法の評価にはleave-one-outなどのcross validationによる評価を行う。

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「教師付き学習」のアルゴリズムが提唱され、それなりに良い成績を収めており、その詳細は既報に詳しい^{9,10)}。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「疑陽性」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¹¹⁾。

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

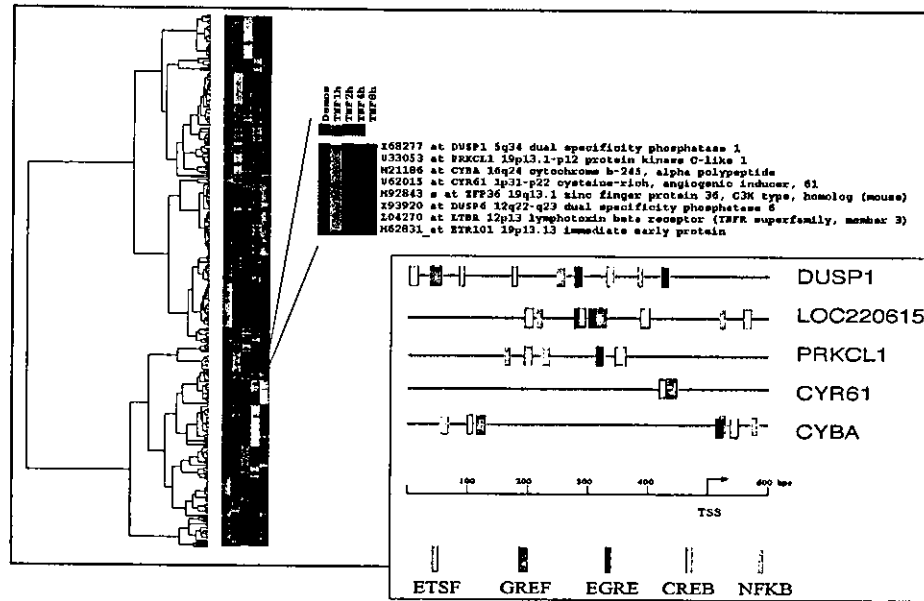


図 4. 転写ネットワークの推定

共通の発現パターンを示す遺伝子群を制御する転写因子の結合部位をプロモータ配列から抽出することができる。HUVECをTNF刺激1時間後に誘導される遺伝子群に共通する転写因子としてNF- κ Bが認められる。

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分かり、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、スプライスバリエント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 fluorescence-activated cell sorter (FACS) 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。例としてヒト臍帯静脈内皮細胞 (HUVEC) を tumor necrosis factor (TNF) 刺激後の発現プロファイルの時系列データのクラスター解析を示す。1時間後に発現誘導され、その後には発現が低下した遺伝子群において nuclear factor- κ B (NF- κ B) の結合配列が共通に存在することが分かる (図 4)。この実験系に種々の阻害剤を添加することにより制御機構の解明が可能である。たとえばプロテアソーム阻害剤である MG-132 を投与すると細胞質に NF- κ B と結合して存在する inhibitor of κ B (I κ B) の分解が起こらず、NF- κ B が核へと移行できないため、ほとんどの遺伝子の転写誘導が抑えられるのに対して、転写誘導が維持される遺伝子も見いだされ、I κ B 経路以外の誘導経路の存在が示唆される。概日時計の解析におい

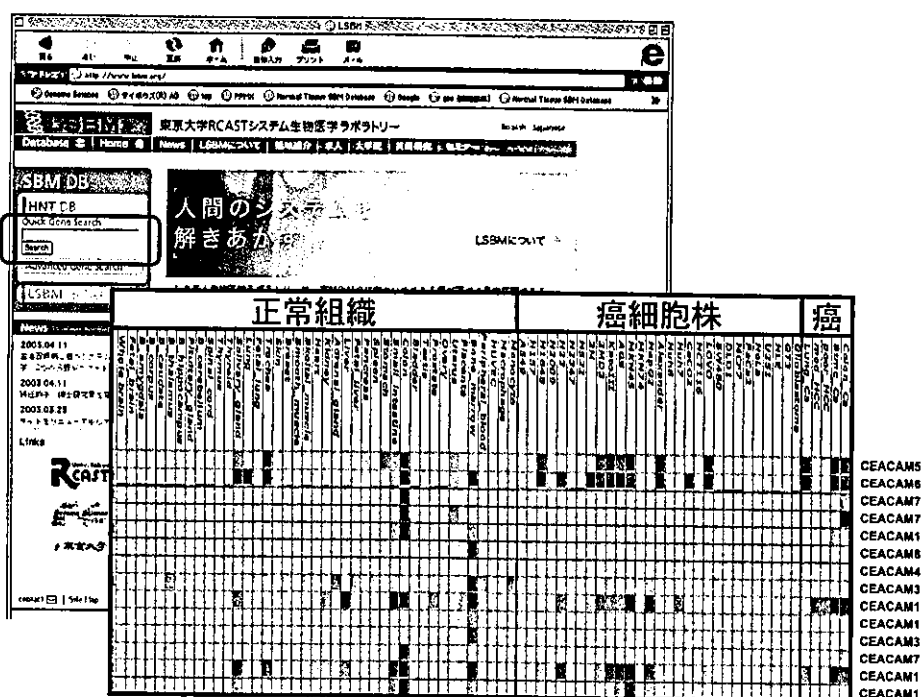


図 5. 遺伝子発現データベース SMB-DB

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。

ても発現プロファイル解析が有効であり, 概日振動を来す遺伝子群の抽出, 制御配列の同定が進められている¹²⁾。

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 minimum information about a microarray experiment (MIAME)¹³⁾ に準拠することが求められるようになってきた。NCBI (National Center for Biotechnology Information), EBI など公共データベース GEO (Gene Expression Omnibus), Array Express 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 (システム生物学データベース <http://www.lsbm.org>) (図 5)。

文 献

- 1) Kennedy GC *et al* : Large-scale genotyping of complex DNA. *Nat Biotechnol* 21 : 1233, 2003.
- 2) Lucito R *et al* : Representational oligonucleotide microarray analysis : a high-resolution method to detect genome copy number variation. *Genome Res* 13 : 2291, 2003.
- 3) Golub TR *et al* : Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286 : 531, 1999.
- 4) van de Vijver MJ *et al* : A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347 : 1999, 2002.
- 5) Reinhold WC *et al* : Apoptotic susceptibility of cancer cells selected for camptothecin resistance : gene expression profiling, functional analysis, and molecular interaction mapping. *Cancer Res* 63 : 1000, 2003.
- 6) 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, 2001.
- 7) Hastie T *et al* : 'Gene shaving' as a method for identifying distinct sets of genes with similar expression patterns. *Genome Biol* 1 : RESEARCH0003, 2000.
- 8) Irizarry RA *et al* : Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4 : 249, 2003.
- 9) 井原茂男ほか : 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 : 71, 2003.
- 10) 加藤菊也ほか : 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 : 2300, 2003.
- 11) Moreau Y *et al* : Comparison and meta-analysis of microarray data : from the bench to the computer desk. *Trends Genet* 19 : 570, 2003.
- 12) Ueda HR *et al* : A transcription factor response element for gene expression during circadian night. *Nature* 418 : 534, 2002.
- 13) Brazma A *et al* : Minimum information about a microarray experiment (MIAME) -toward standards for microarray data. *Nat Genet* 29 : 365, 2001.

ゲノム発現からのアプローチ

油谷浩幸

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部門

ポイント

- RNA やタンパク質などの生物情報を大量かつ高速に取得することが可能となり、生命科学のパラダイムの変化が起こりつつある。
- トランスクリプトームの解明にはさまざまなバリエーションの転写産物の同定、ゲノムのメチル化やコピー数の変化の網羅的解析も重要である。
- タンパク質発現を網羅的に解析する技術開発が必要である。
- 共発現する遺伝子を制御する転写因子をゲノム配列から推定できる。

OMICSによる疾病解析

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ているが、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あるいはタンパク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→タンパク質という生命情報の流れのそれぞれを高速に網羅的に大量取得することが可能となり、生命科学にパラダイムシフトをもたらしている。ゲノム配列が決定された現在、SNP (single nucleotide polymorphism; 一塩基多型) に代表されるその多様性と疾患との相関が注目されている一方で、網羅的な解析技術としてDNAチップなどによるRNAの変動の解析(トランスクリプトーム)や、機能を担う分子であるタンパク質を扱うのがプロテオーム解析である。

1. ゲノミクスとエピゲノミクス

配列情報およびその多様性については本章“遺伝学的アプローチ”に譲る。配列そのものの変異ではなく後成的変異であるエピジェネテ

イクスは、RNAあるいはタンパク質発現の調節機構として重要である。癌細胞では一般にゲノム全体が低メチル化状態にある一方、一部の遺伝子のプロモーター領域が過剰メチル化により不活化されていることが知られているほか、老化現象においてもエピジェネティックな変化が注目されている。たとえば、*p16*、*E-cadherin*などの癌抑制遺伝子のメチル化による不活化が報告されており、さらにヘテロ接合性の消失(loss of heterozygosity; LOH)と同時に生じると当該遺伝子は双方のコピーが不活化されることになる。ゲノムのメチル化の全体像を解析できるようにすることが重要である¹⁾。

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²⁾、SNP測定と合わせることによりゲノムワイドな染色体変異の解明が期待される(図

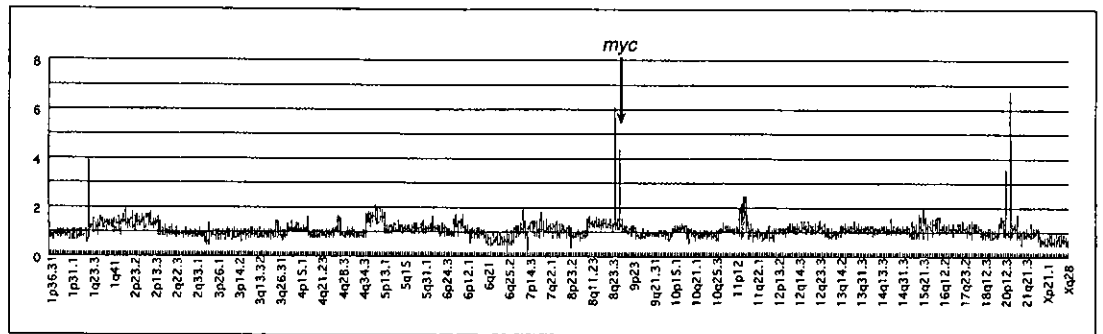


図1 DNAチップによるゲノムコピー数解析 (Genome Imbalance Map)

1番染色体から性染色体まで1万ヶ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)

1). 発現解析データからも変異のある染色体領域が推定される³⁾。

2. トランスクリプトーム

トランスクリプトームの全体像はまだまだ不明な点が多い。ゲノム中に存在する遺伝子の同定は、コンピュータによる予測あるいはEST (expressed sequence tag) などの実験により得られた転写産物の配列情報などを総合して進められている。配列情報に基づいて機能アノテーションが進められてはいるが、いまだに機能が不明の遺伝子も少なくない。限られた細胞種だけに発現する場合やきわめて発現量の低い遺伝子では、その存在を実証することも容易ではない。また、RNAiという現象が普遍的であることが示されるにつれて、タンパク質をコードするRNA以外に注目されているのが、microRNAをはじめとする non-coding RNAである。加えて1つの遺伝子座 (locus) から複数種の転写産物が作られることはまれではなく、発生時期、組織などに特異的な転写産物が認められる。

疾病罹患組織と正常組織における遺伝子発現を比較することは、従来ディファレ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われていたわけであるが、DNAチップ技術では

3万個を超す遺伝子の発現レベルを同時解析可能である。RNAの発現量とタンパク質の発現量には必ずしも良好な相関が認められるわけではなく、発現プロファイル解析の有する網羅性と転写制御をモニタリングできる有用性から、プロテオーム解析とは補完し合う解析技術であると考えられる。

DNAチップ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⁴⁾・診断が可能となり⁵⁾ (図2)⁶⁾、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一方、新規な治療標的分子の同定も進められている。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。細菌やウイルスなどの病原体を同定する感染症診断のほか、薬剤の有効性、副作用の判定にも用いることが可能である。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⁷⁾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。

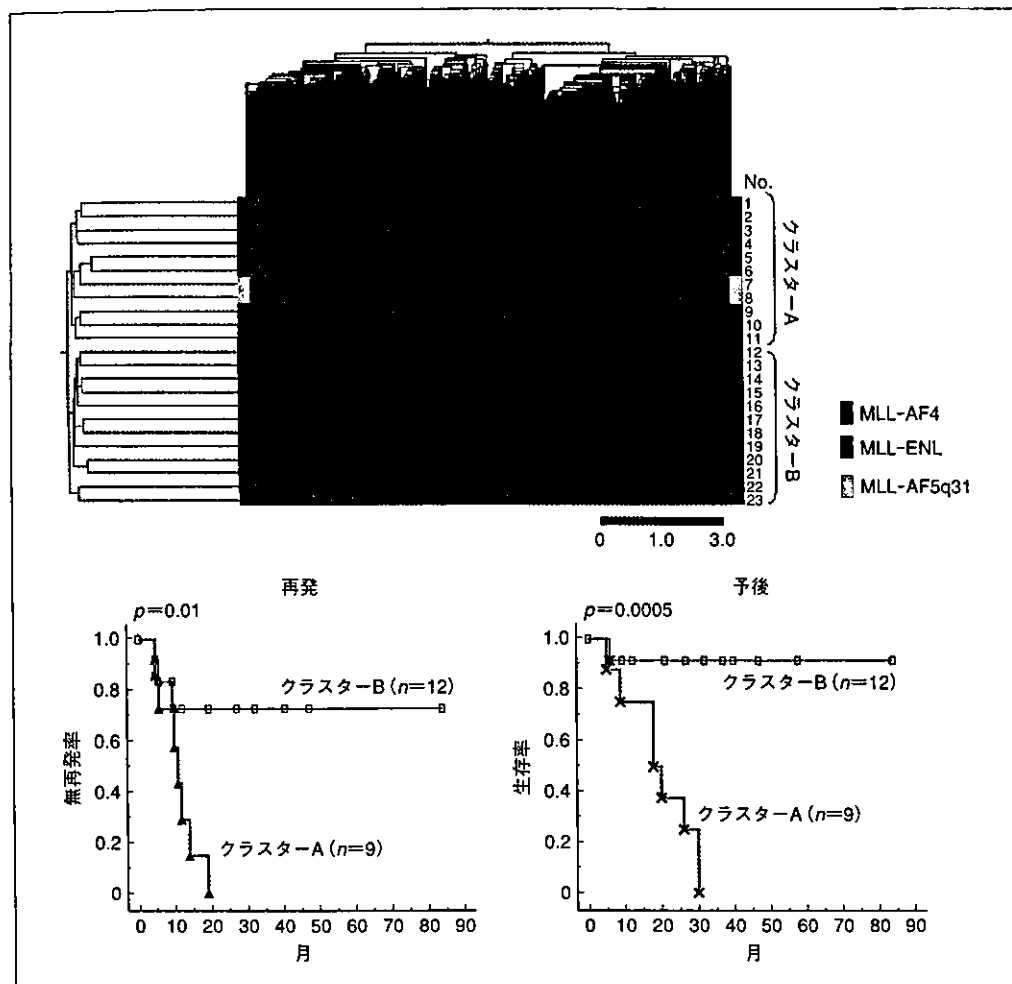


図2 アレイ解析による小児白血病における新たな病態の同定
MLL 転座を有する急性リンパ球性白血病患者より採取した骨髄について、発現プロファイルをクラスター解析したところ2群に分かれ、2群間で無再発期間および予後に有意な違いが認められた⁶⁾。

糖尿病や高血圧症などのいわゆる生活習慣病は多因子性疾患であり、ゲノムおよびプロテオーム解析を駆使して疾患感受性に関する遺伝子の探索が進められている。将来的にも1つの遺伝子検査のみでは確実な診断は困難と考えられるが、これらの生活習慣病に関する治療薬は巨大な市場を有するだけに研究機関のみならず製薬企業も遺伝子の同定を競って進めており、10~20年後には一般のクリニックにおいても生活習慣の指導や薬剤の処方ゲノム情報が応用されることが期待される。

3. プロテオミクス

プロテオミクスは、血清などの臨床試料のタンパク質プロファイリング、あるいはタンパク質相互作用解析に応用すべく、ポストゲノム時代の解析技術として遺伝子発現解析とともに近年注目されている⁸⁾。タンパク質を網羅的に検出する技術としては、従来の二次元電気泳動や質量分析があげられる。タンパク質についても、異なる蛍光色素標識を用いた二次元電気泳動によるプロファイリングが可能である(例: Amersham社のDIGE)。質量分析ではSEL-

DI-TOF を用いた検出系 (例 : Ciphergen 社の ProteinChip™) も限られた目的には有効であるが, MS/MS 技術によりアミノ酸配列決定も行えるようになっており, 現状は発現量が多いせいぜい 2~3 千種類のタンパク質を検出できるのみである. より微量検体の定量的解析へ向けて, 今後のさらなる技術改良が望まれる. タンパク質解析には量的変動に関する情報に加えて, そのリン酸化などの修飾や分解などのプロセシングあるいは他の分子との相互作用解析が重要であるため, 特異的に結合する抗体などのプローブや相互作用の検出法の開発が望まれる. これらのゲノムおよびプロテオーム解析のためには, 臨床試料の系統的なバンク化も待たれるところである.

アレイ解析技術

1. DNA チップ

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 方式が主体であるが, 異なる測定方式も開発されている. DNA チップでは, 基板上の 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. ゲノム配列情報が解読され, おおよその遺伝子配列が明らかとなったこともあり, 合成 DNA によるアレイへと変遷しつつある. ただし EST 配列の決定が遅れている生物種のアレイ作製には当面, cDNA アレイも有用である.

プローブの集積度や感度を上昇させる技術の進歩, ゲノム配列情報の蓄積, スキャナーの性能向上とともに測定結果の信頼性にも格段に向上がみられる. 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. 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

りはるかに少量ですむ. データ解析には臨床情報の収集, および大量の情報処理にはインフォマティクス研究者との連携も不可欠である.

1) 合成 DNA アレイ

GeneChip™ は光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うが, 当初はペプチド合成技術を有する Affymax 社の一部門としてスタートした⁹⁾. ゲノム計画の成果として EST データの蓄積も進んだことから, 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. 1 つの遺伝子あたり 11~20 対のプローブのシグナル強度の差の平均値 (average difference) を遺伝子の発現強度とするので, 絶対量としての精度, 再現性も良好であるが¹⁰⁾, アレイごとにフォトマスク作製の手間とコストがかかるのが難点である. 選択するプローブ配列を改良することによりデータの信頼度を向上させることが期待できるが, それでも新旧のアレイ間でのデータの互換には課題が残る. Febit 社 (独) や NimbleGen 社 (米) が光リソグラフィ法を用いた *in situ* 合成アレイを作製している.

2) マイクロアレイ

マイクロアレイはピンヤインクジェット方式によりスライドガラスやフィルターに PCR 増幅した DNA や合成 DNA を固定化する. 基板についても最近では通常のポリリジン以外でコーティングしたスライドガラスや, ポリアクリルアミドや多孔性基質に DNA を固定化して検出感度を上昇させたものが開発されている. 通常 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い, 同時にハイブリダイゼーションを行うことにより比較するが, スポット量を一定に制御することにより, 単色標識による測定も可能となった. 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

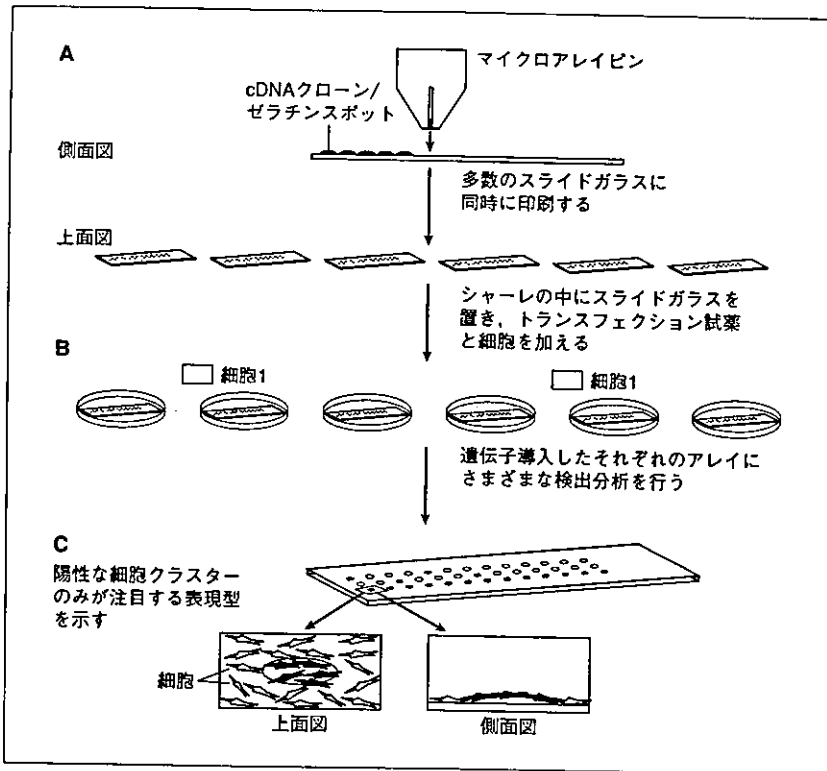


図3 細胞アレイ

(Wu RZ, et al: *Trends Cell Biol* 12, 485-8, 2002¹⁸⁾ より)

際のクローン管理や異なるアレイ間のデータ共有時の標準化に依然課題が残る。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。オリゴヌクレオチドの合成コストの低下に伴い、長鎖のオリゴヌクレオチドを用いたマイクロアレイがcDNAアレイ同様に用いられるようになり、感度面では改善された¹¹⁾。Agilent社はインクジェット方式で*in situ*合成を行う方式のアレイを開発している。一方、Illumina社は、オリゴヌクレオチドを合成したビーズを光ファイバーの先端に固定することにより検出するシステムを開発している。

1枚のアレイで5万に及ぶ転写産物を測定することが可能となったものの、低発現な遺伝子

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二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、実験による経験的なプローブ性能の実証も今後は重要であろう。また、現行のアレイではプローブの選択や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を同定する解析手法の開発も求められている。ゲノム配列を連続的に合成したゲノムタイリングアレイも試作されており、新規転写産物の同定¹²⁾や転写因子結合部位の同定(ChIP-chip)¹³⁾に利用されている。

2. プロテインチップ

前述のSELDI技術を用いたProteinChipTM以外に、多数のタンパク質の測定には抗体アレイ

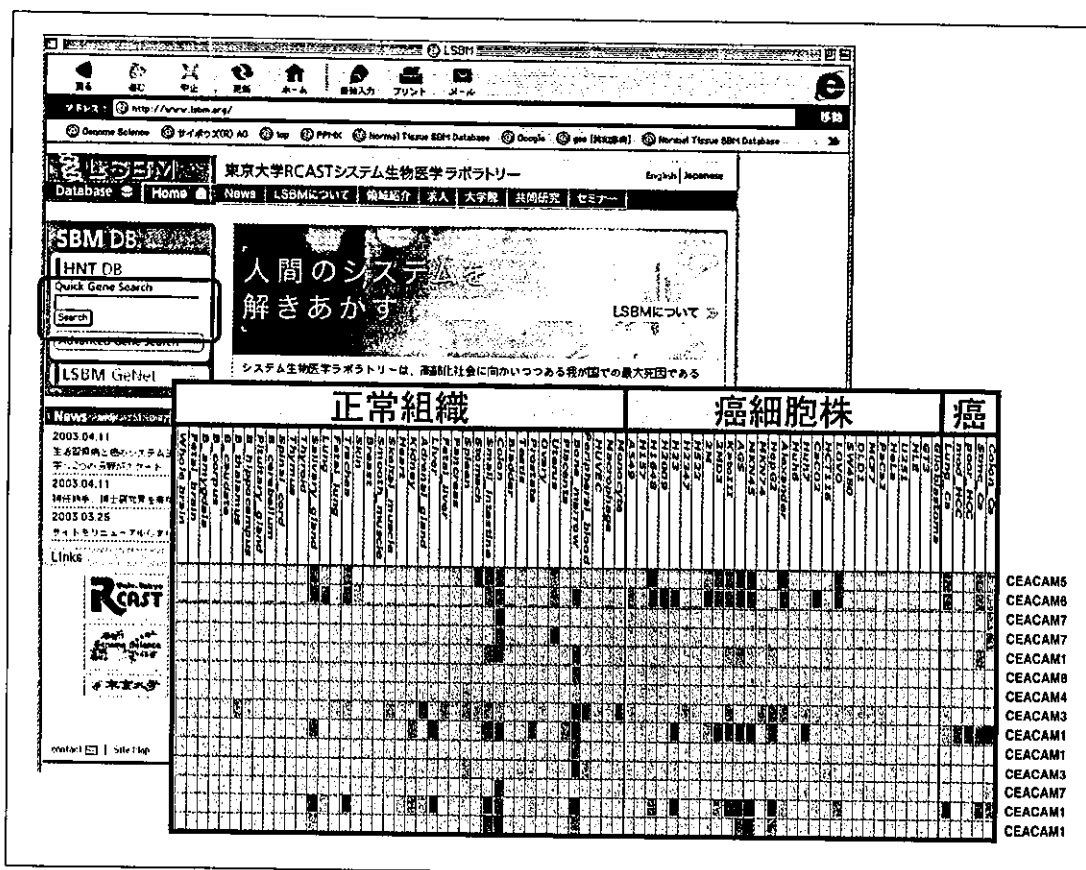


図4 遺伝子発現データベース SBM-DB

GeneChip™により測定したヒト正常組織、初代培養細胞、癌細胞株の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²⁰⁾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 など公共データへのリンクが張られている。

イも注目されている¹⁴⁾。2つの試料タンパク質を異なる蛍光色素で標識することにより、相対的な測定を行うことができ、数百種類のモノクローナル抗体を搭載したアレイが現在製品として利用可能である。しかしながら、抗体の特異性、感度などの性能に依存するところが大きく、網羅的な測定への最大の問題点である¹⁵⁾。

3. その他のアレイ技術

1) 組織アレイ

パラフィン包埋組織切片から多数の組織検体を1枚のスライドガラス上に配列したものであり、異なるタンパク質を連続切片で解析するこ

とも可能である^{16,17)}。免疫組織染色により多数症例を解析するために利用されているが、読み取り、診断のプロセスの自動化が困難なこと、組織片が小さいことが問題となる。

2) 細胞アレイ

DNA や RNA を固定した上に細胞培養を行うことにより、多数のタンパク質を同時に発現させたり、RNAi を同時に評価することが可能となる (図3)^{18,19)}。

生物情報、疾患情報の統合

ゲノミクスをはじめとするトランスクリプト

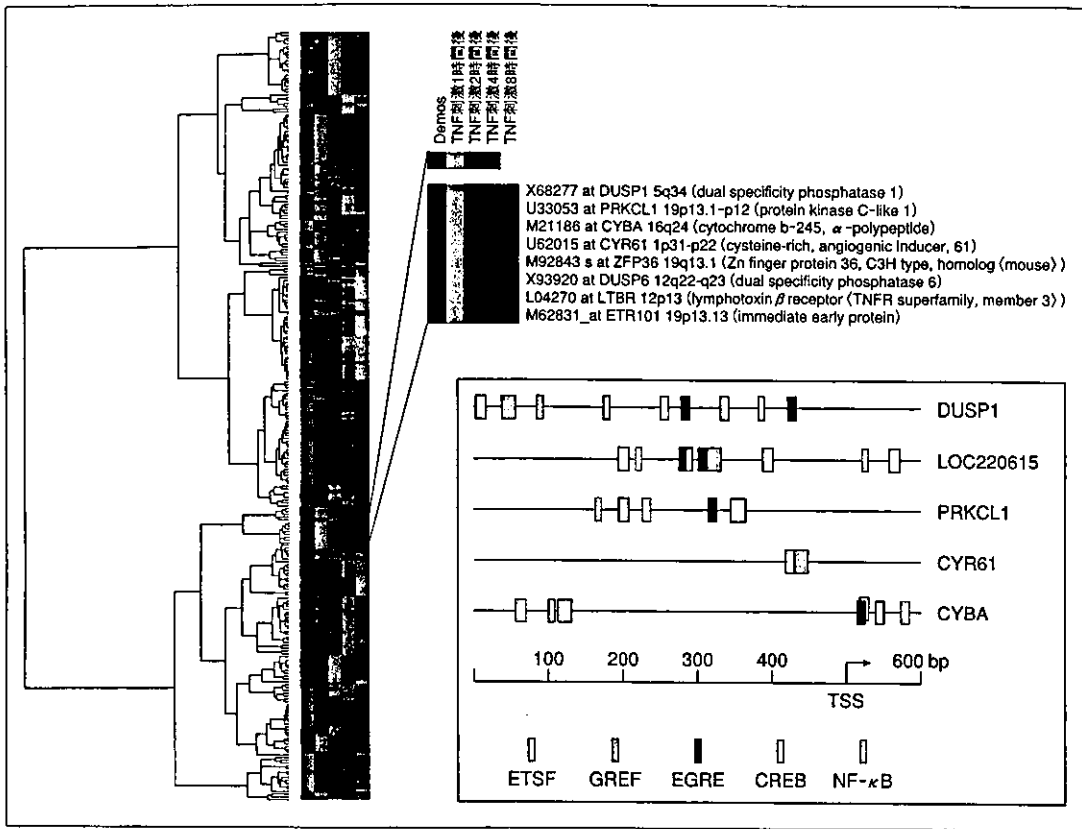


図5 転写ネットワークの推定

共通の発現パターンを示す遺伝子群を制御する転写因子の結合部位をプロモーター配列から抽出することができる。HUVEC細胞をTNF刺激1時間後に誘導される遺伝子群に共通する転写因子として、NF-κBが認められる。

ーム、プロテオーム、メタボローム解析などのいわゆる“OMICS”と呼ばれる生命情報の網羅的な解析が生み出す大量の測定データを統合処理する分野として、バイオインフォマティクスの重要性が高まっている。

1. 情報の統合

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。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。NCBI (National Center for Biotechnology Information), EBI (European Bioinformatics Institute) などで公共データベース GEO (Gene Expression Omnibus), ArrayExpress が提案されているが、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している(図4)²⁰⁾。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、遺伝子発現プロファイル解析技術はそのなかで大きな役割を担っ

ていくものと期待される。

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, 罹患者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“教師付き学習”のアルゴリズムが提唱され, それなりに良い成績を収めており, その詳細は既報に詳しい^{21, 22)}。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”の問題が残る。

3. パスウェイ解析

遺伝子セットの生物・医学的な意味づけが重要である。通常は, 遺伝子のクラスター解析の

結果, 発現プロファイルが類似した遺伝子のそれぞれの上流に共通の転写エレメントが抽出できれば, 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タンパク質間相互作用が転写制御ネットワークの形で構築できる。遺伝子の破壊や過剰発現実験を系統的に行った変化から, 遺伝子間の因果的連鎖を推論することも有効である。そのため, 発現プロファイル解析の結果をプロモーターからの転写因子結合配列の抽出 (図5), 遺伝子間の相互作用ネットワークの解析, あるいはパスウェイ抽出へとつなげることになる。このとき, 選ばれた遺伝子の生物・医学的意味づけを知るために文献情報を活用しなければならない。そこで Gene Ontology のデータの参照, PubMed 検索など, 文献のなかから必要な情報の抽出を研究者自ら行わなければならない。この文献検索の部分をいかに効率良く行うかが今後の大きな課題である。

■文献, Web site

- 1) Shi H, Wei SH, Leu YW, et al. Triple analysis of the cancer epigenome : An integrated microarray system for assessing gene expression, DNA methylation, and histone acetylation. *Cancer Res* **63**, 2164-71 (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**, 2291-305 (2003)
- 3) Kano M, Nishimura K, Ishikawa S, et al. Expression imbalance map : A new visualization method for detection of mRNA expression imbalance regions. *Physiol Genomics* **13**, 31-46 (2003)
- 4) Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer : Class discovery and class prediction by gene expression monitoring. *Science* **286**, 531-7 (1999)
- 5) van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347**, 1999-2009 (2002)
- 6) Tsutsumi S, Taketani T, Nishimura K, et al. Two distinct gene expression signatures in pediatric acute lymphoblastic leukemia with *MLL* rearrangements. *Cancer Res* **63**, 4882-7 (2003)
- 7) 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**, 1000-11 (2003)
- 8) Wulfsberg JD, Liotta LA & Petricoin EF. Proteomic applications for the early detection of cancer. *Nat Rev Cancer* **3**, 267-75 (2003)
- 9) Fodor SP, Read JL, Pirrung MC, et al. Light-directed, spatially addressable parallel

- chemical synthesis. *Science* **251**, 767-73 (1991)
- 10) Ishii M, Hashimoto S, Tsutsumi S, et al. Direct comparison of GeneChip and SAGE on quantitative analysis. *Genomics* **68**, 136-43 (2000)
 - 11) Kane MD, Jatkoe TA, Stumpf CR, et al. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* **28**, 4552-7 (2000)
 - 12) 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**, 331-42 (2004)
 - 13) 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**, 499-509 (2004)
 - 14) Haab BB, Dunham MJ & Brown PO. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol* **2**, RESEARCH0004 (2001)
 - 15) Michaud GA, Salcius M, Zhou F, et al. Analyzing antibody specificity with whole proteome microarrays. *Nat Biotechnol* **21**, 1509-12 (2003)
 - 16) Kallioniemi OP, Wagner U, Kononen J & Sauter G. Tissue microarray technology for high-throughput molecular profiling of cancer. *Hum Mol Genet* **10**, 657-62 (2001)
 - 17) Howbrook DN, van der Valk AM, O'Shaughnessy MC, et al. Developments in microarray technologies. *Drug Discov Today* **8**, 642-51 (2003)
 - 18) Wu RZ, Bailey SN & Sabatini DM. Cell-biological applications of transfected-cell microarrays. *Trends Cell Biol* **12**, 485-8 (2002)
 - 19) Mousses S, Kallioniemi A, Kauraniemi P, Elkahoul A & Kallioniemi OP. Clinical and functional target validation using tissue and cell microarrays. *Curr Opin Chem Biol* **6**, 97-101 (2002)
 - 20) システム生物学データベース. <http://www.lsbm.org>
 - 21) 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**, 71-80 (2003)
 - 22) 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**, 2300-9 (2003)