

解説

ナノレベルのゲノム実験室

— DNA チップによる生体からのゲノム知識発見 —

中谷明弘 森下真一

東京大学医科学研究所

はじめに

「DNA チップ」という言葉から、DNA コンピューティングや遺伝的アルゴリズムに関係するデバイス、もしくは、何らかの生体の機構を模した新種のマイクロプロセッサを連想される読者も多いかもしれない。しかしDNA チップとは、多数のDNA 配列をガラスやシリコンといったチップ上に規則的に配置したバイオ分野向けの情報獲得デバイスである。

DNA チップが医学や生物学の分野で近年注目されるようになった背景には、ゲノムプロジェクトの進展と深い関係がある。近年、大腸菌、酵母、線虫の全DNA 配列が文字列として解読され、ヒトのDNA 配列もここ3~5年のうちに解読されると予想されている。文字列としての理解の後に最も関心が寄せられるのが、DNA 配列上にコードされた各遺伝子機能の解明である。そのためには細胞内で特定の遺伝子が発現しているか否かという状況証拠を観測することが重要になる。DNA チップはこのような計測技術の1つである。

■ 遺伝子発現観測によるゲノム知識発見

DNA チップの解説の前に、観測から科学的知見をどのように導こうとするか紹介しよう。たとえば、ある特徴的なDNA 配列が、正常な細胞では発現しない一方で、HIV

ウイルスに感染した細胞においては発現するという事実を観測したとする(図-1)。するとそのDNA 配列は病気に関連する遺伝子をコードしている可能性が高い¹⁾。

また、単一ではなく複数の遺伝子の発現量を、脳・心臓・肺・肝臓等の臓器において、胎児から成長するまで時間軸に沿って観測し、発現パターンの類似性に基づいて遺伝子をグループ化(クラスタリング)してみる。すると、同じグループに属する遺伝子は、類似の制御構造のもとで発現がコントロールされていると予想できる。そこで、各遺伝子の発現をコントロールしているプロモータという領域から共通の部分配列を抽出するという、研究のロードマップが考えられる。また、同じグループ内において、機能が既知の遺伝子から、機能が未知の遺伝子の性質を類推できる可能性もある^{2), 3)}。

あるいは、酵母ではある遺伝子を人為的に破壊したり、または過剰に発現させることが試みられているが、このような操作が他の遺伝子の発現に影響を及ぼすか否かが観測できたとする。すると、遺伝子間の影響関係(遺伝子ネットワーク)の解明という深い生物学的知識が得られる可能性がある⁴⁾。

■ DNA チップと情報処理

以上の例のように、遺伝子発現量の観測は、医科学的知見を広げる手段として欠かせない。DNA チップはこのような観測手段として現在脚光を浴びている技術であり、情報処理技術に大きく依存している。DNA チップという名前は、このデバイスが、ガラスやシリコンといった材質のチップ上に核酸配列(DNA やRNA)を規則的に配置したものであることに由来する。マイクロプロセッサやインクジェットプリンタ向けに開発された微細加工技術が応用され、米国を中心に開発が急速に進んでおり、一部は実用化されている。

DNA チップ業界では時として“massively parallel”という形容詞でDNA チップを修飾するが⁵⁾、これは微細な反応多数を並列に実行させることが可能であることを意味している。具体的には、細胞中における数千から数万種類の異なる遺伝子の発現量を同時に観測することが

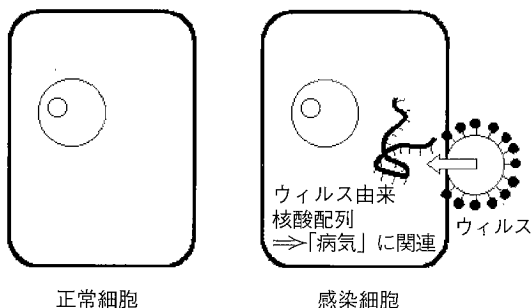


図-1 疾病に特徴的な核酸配列

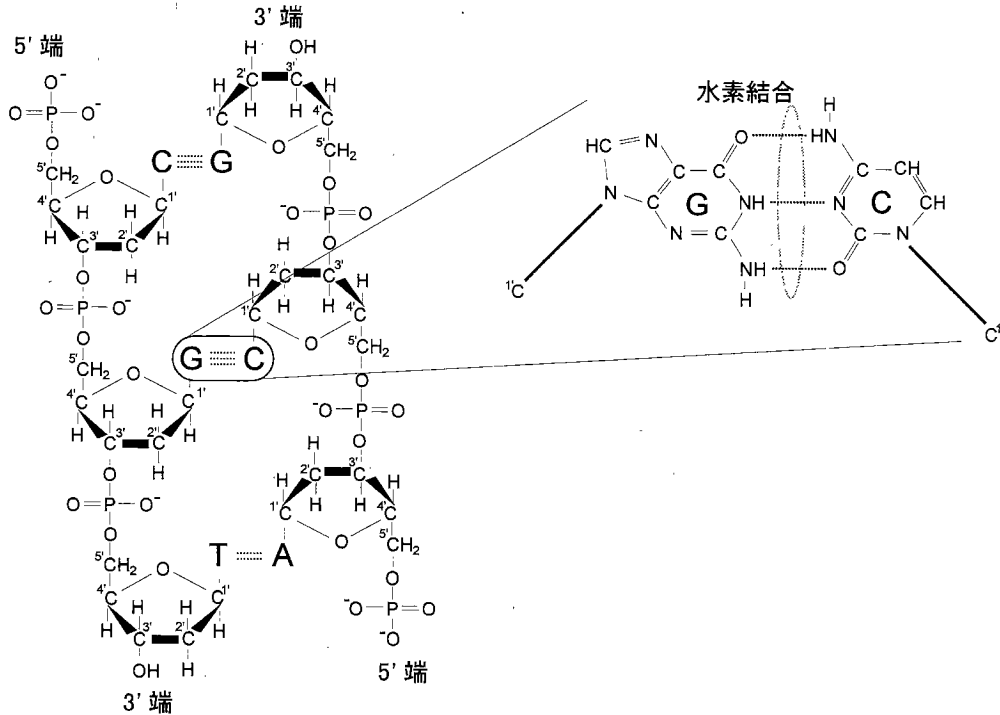


図-2 2本鎖DNA

可能である。多量の観測データが得られた後は、遺伝子の機能を予測・発見するため情報処理が重要であり、クラスタリング・アソシエーションルール・プリアンネットワークを効率的かつ高精度で生成する技術が開発されつつある。このような背景を踏まえて、本稿ではDNAチップの構造と用途について説明する。

遺伝情報の獲得

核酸の構造

核酸分子はヌクレオチドと呼ばれる繰り返し単位が共有結合によって多数連なった高分子ポリマーである。この繰り返し単位は、五炭糖を中心に、塩基とリン酸基から構成される。DNAでは、五炭糖としてデオキシリボースが、塩基として、アデニン (A)、グアニン (G)、シトシン (C)、チミン (T) が用いられる。ひと続きのDNA配列 (1本鎖DNA) は、五炭糖とリン酸基を省略して上記4種の塩基の配列として表すことができる。この配列には方向性がある、ヌクレオチドに含まれる五炭糖の5員環中炭素の番号を付加して表現する。たとえば、図-2中に示した左側の1本鎖DNAは、DNA配列を五炭糖の5番炭素から3番炭素に辿る方向に3つの塩基がCGTの順に出現しているので、5'-CGT-3'と表す。このとき、配列の両端をそれぞれ5'端、3'端と呼ぶ。遺伝的情報 (遺伝子) は、上記の塩基を文字として記述されていると考えることができる。ヒトの場合、30億塩基相当の長さのDNA上に、6.5~8万程度の遺伝子が散在しているといわれる。この30億塩基のうち、遺伝子を記述した部分列

が選択的に読み出されて該当する遺伝子の機能を発現する。

ハイブリダイゼーション

核酸配列は、その構成塩基間 (G-C, A-T) に塩基対 (図-2に示すように1塩基対は何本かの水素結合によりなる) を形成し、分子全体としてエネルギー的に安定な3次元の立体構造をとる。この塩基対は、同一分子内および複数の分子間で形成される。たとえば、2本の互いに相補的なDNA配列は梯子状に塩基対を形成し (2本鎖DNA化)、さらに、3次元的にねじれて螺旋構造をとって安定化する。

特に2本の核酸配列間に連続した塩基対を形成することをハイブリダイゼーションという。図-2では、5'-CGT-3'と3'-GCA-5'間のハイブリダイゼーションの様子 (3対の塩基対を生成) を示している。

遺伝情報の流れ

細胞内の遺伝情報の流れの概略は以下のようである。DNA上に塩基配列として規定されている遺伝情報 (遺伝子) がRNAに転写され、このRNA上に転写された情報を翻訳してさまざまな機能を持つタンパク質が合成される。DNA上に規定されている遺伝情報は、この流れに従って発現する。たとえば、脳細胞の中では、細胞を脳細胞たらしめる遺伝子が発現しており、癌化した細胞内では、それに相応する遺伝子が発現している。遺伝子は、単純なON/OFF的な情報のみではなく、複数の遺伝子が互いに関係しあうことによって構造化した情報も規定している。たとえば、酵母が外部環境の変化に応じて、挙動

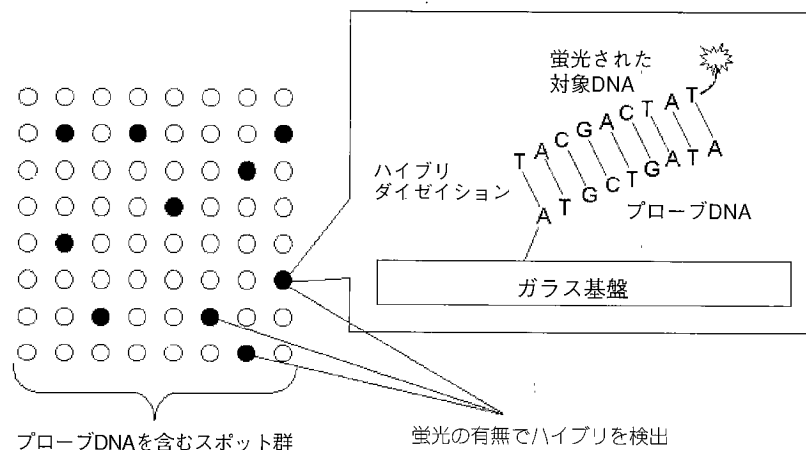


図-3 ハイブリダイゼーション

を変化する場合には、複数の遺伝子が時系列的に発現する。遺伝情報の転写と核酸配列同士のハイブリダイゼーションという2つの性質を用いれば、細胞内に発現している遺伝子をその遺伝子に相当するDNA配列（対象DNA）とそれと相補的なDNA配列（プローブDNA）とのハイブリダイゼーションを介して検出することが可能となる。

例として、対象DNAを5'-GGCCCGGG-3'とすると（この配列が試料中に存在するか否かが未知）、プローブDNAとして塩基配列3'-CCGGGCC-5'を用意して、同プローブDNAにハイブリダイズしたものが存在したか否かを判定して、対象DNAの存在を検出する。この手法は、たとえば、疾病の出現を特徴付ける塩基配列が既知である場合に、同塩基配列をプローブDNAとして準備し、患者から採集したDNAを対象DNAとして用いることによってその疾病の有無の判定に使用可能である。

PCRによる配列コピーの増幅

相補的な核酸配列同士がハイブリダイゼーションを起こすことを利用して、核酸配列の特定の領域（数千塩基対程度）のコピーを大量に生成する方法（PCR: Polymerase Chain Reaction）が広く用いられている。2本鎖DNAは加熱することによって、2本の1本鎖DNAにすることができる（熱変性）。この状態で、増幅したい領域の両端それぞれに相補的な1組の長さ20程度の塩基配列（プライマー）を過剰に加えて冷却すると、プライマーは増幅領域端部分と2本鎖を形成する（この操作をアニーリングという）。これに、DNA合成酵素（DNAポリメラーゼ）を作用させると、プライマーとの2本鎖部分からDNA伸長反応が起こる。これらの処理を1サイクルとして繰り返すことによって、特定の領域（たとえば検出対象となるDNA領域）のコピーを、繰り返し回数の指数関数個だけ生成することができる。ごく微量のDNA片に対しても感度よく適用できることが特徴である（これはノーベル賞の受賞対象となった技術）。

ハイブリダイゼーションの手順

ハイブリダイゼーションを起こさせてそれを検出する手順の概略は以下の通りである。

- (1) 対象DNAの準備: ハイブリダイゼーションに先立って、対象DNAのコピーを、PCRによって多数作成する。また、対象DNAに蛍光物質を取り込ませて印をつける（これを標識という）。蛍光物質としては、ビオチン（ビタミンH）などが用いられる。ビオチンの存在（特定の条件下で緑色の蛍光を発する）を判定することによって、対象DNAの存在を検出する。標識には、放射性同位体元素が用いられることもある。
- (2) ハイブリダイゼーション: プローブDNAと対象DNAを加熱やアルカリ処理によって1本鎖化した後、再び、2本鎖を形成する状態に戻すことによって行う。このとき、再形成された2本鎖を構成する2本のDNA鎖は、プローブDNA、対象DNAそれぞれから由来するものが使われ得るため、プローブDNAに相補的な塩基配列を検出することが可能になる。
- (3) ハイブリダイゼーションの検出: CCDカメラや写真フィルムの感光を用いてハイブリダイゼーションがアクティブになっている領域を2次元画像として検出する。

ハイブリダイゼーションを用いた従来手法

核酸同士がハイブリダイゼーションを起こすという性質を利用して対象DNAを観測する手法は従来から用いられている。たとえば、ドットプロットと呼ばれる手法は、ナイロン膜のような担体上に対象DNAを含んだ試料の水滴を配置し、それとプローブDNAとの間に生成させたハイブリダイゼーションを検出することによって、対象DNAに関する情報を獲得することが可能である。また、サザンプロットと呼ばれる手法においても、ゲル電気泳動によって分離（DNA片は酵素によって切断された長さに従って分離）した対象DNAをナイロン膜等に写し取り、これをプローブDNAとハイブリダイゼーションさせて検出することが可能である。一般的にはこれらの手法

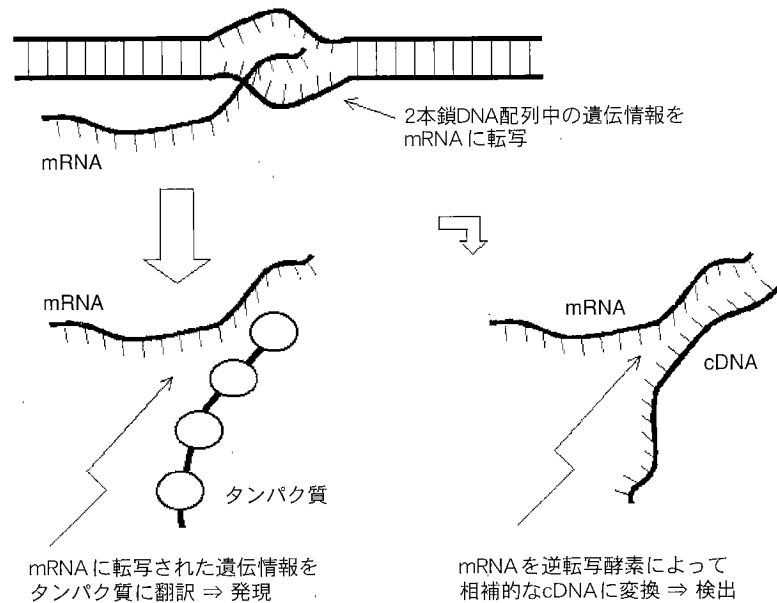


図-4 遺伝情報の流れ

は、使用するDNA片数を増幅する方法（前述のPCR）に由来して、比較的長い核酸（1万塩基対程度以下）同士のハイブリダイゼーションを行うという特徴がある。

DNAチップ『超並列』ハイブリダイゼーション

上記の従来方法では、対象DNAは既知でなければならないという制限があるため、核酸情報の有無の「診断」に用いることはできても、新たな核酸情報そのものの「発見」に用いることは難しい。これに対して、DNAチップが実現する圧倒的な並列度のハイブリダイゼーションは、従来手法の高集積化を実現するのみではなく、未知の対象DNAに関する情報を収集することを可能にする。

たとえば、配列が未知の8塩基の対象DNAを検出しようとする場合、長さ8塩基のDNA配列が取り得るすべての種類（ 4^8 種類）のプローブDNAを用意する。このすべてのプローブDNAに未知の対象DNAを含んだ試料を適応してハイブリダイゼーションの有無を調べることによって、未知のDNA配列を解読できる。あるいは、全遺伝子の配列が知られている場合は、そのプローブDNA配列を配置したDNAチップを設計してもよい。

仮に人間の全遺伝子を特徴付けるのに十分なプローブDNA群を配置したDNAチップが実現するならば、たとえば、ある疾病を発病しているグループと、発病していないグループに同DNAチップを適応した結果を解析することによって、疾病の原因遺伝子を同定することが可能になると期待される。実際に、2000年代初頭には、当初の予定を大幅に前倒してヒトの全DNA配列が同定されるともいわれている。このような背景もあって、DNAチップが実現する低コスト、高効率な手法への期待は大きい。

■ DNAチップの分類

チップの分類に関しては厳密な定義はないが、本稿では、現在DNAチップと呼ばれるものをプローブDNAの構成の仕方によって、次のように分類することにする。

- (1) プローブ配置型: チップ基板上にPCR等によって単離したプローブDNA（2本鎖DNA）を含んだ試料を配置したもの。プローブDNAの長さは数百から数千塩基程度。
- (2) プローブ合成型: このタイプは、ガラスやシリコンの基板上にDNAの伸長反応を用いてプローブDNA（1本鎖DNA）を「生やした」もの。プローブDNAの長さは数十塩基程度。

これらとは別に、頻繁に用いられる用語として「マイクロアレイ」というものがある。これは、数cm四方程度の基板上に、多数の試料を含んだ水滴をグリッド状に配置したものであり、上記の分類に従うと、プローブ配置型に分類される。従来から、10～数10cm四方のナイロン等の膜上に水滴を配置する手法が用いられているが、こちらと比較して、「マイクロ」なアレイと呼ばれるようである。

■ DNAチップの実例と動作原理

本節では、いくつかのDNAチップを例に用いてその動作原理を説明する。

◎スタンフォード大Brownらのシステム

プローブ配置型として注目されたのが、スタンフォード大のBrownらのグループである。同グループは、酵母の約6千個の全遺伝子に対応するDNA配列を含むチップの作成に成功している（多数のコントロールスポットも含めて110枚のチップ上に総計約70万スポット配置）。酵

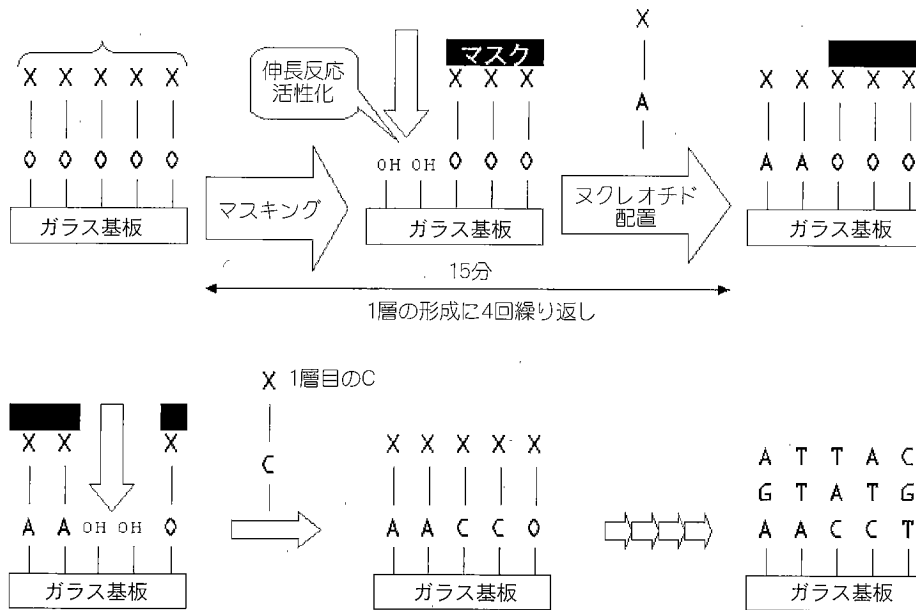


図-5 プロブDNAの伸長

母の各遺伝子に対応するDNA配列は、そのDNA配列のみを個別に切り出して多数のコピーを作成する(PCRとゲル電気泳動を用いる)。これらを含んだ水滴をガラスプレート上に数百 μm 間隔でグリッド状に配置してチップを作成する(図-3)。詳細については、文献3), 6), 7)を参照。このグリッド状に配置された全遺伝子のコピーを用いることによって、酵母の各遺伝子が規定している形質がどのように関連しあっているかを調べることができる。

前述のように、染色体上DNA内に含まれる遺伝情報が発現するとき、その対応するDNA配列部分は相補的な塩基配列を持つmRNAに転写される。したがって、各環境において酵母細胞内に存在するmRNAの発現を検出することによって、そこで発現している遺伝情報を間接的に調べることができる(図-4)。

Brownらは、異なる環境下の酵母からmRNAを採取し、これに相補的なDNA配列(complementary DNA: cDNA)を酵素を用いて生成し、それぞれ別な色で標識した上で、上記の全遺伝子のコピーとのハイブリダイゼーションを行っている。

たとえば、2つの異なる環境から採取したmRNAを元に生成したcDNAをそれぞれ赤と緑で標識する。これらを混合してハイブリダイゼーションを行うと、それぞれの環境に固有に発現している遺伝子があれば、赤または緑に蛍光するスポットがチップ上に検出される。両環境に共通に発現する遺伝子があれば、発現量に応じて赤と緑の混色として検出される(同量の発現の場合黄色となる)。

© Synteni社 GEM array

同様の水滴のスポットによるプローブ配置型のシステムは、ほかにも存在している。たとえば、Synteni社のGEM array (GEM: Gene Expression Microarray)⁸⁾は、ガラス表面上にcDNA(500~5000塩基)のセット

をグリッド状に配置して、これと細胞から採取したmRNAとのハイブリダイゼーションを検出する。1つのcDNAが1つのヒト遺伝子に相当するようなcDNAのセットを配置したものを提供している。1チップ上に1万種程度のcDNAを配置することが可能である。

このシステムはその名前の通り、遺伝子の発現量を測定することを目的としている。2種の蛍光物質を使い、他方をコントロールとして用いることで発現量を測定する。たとえば、正常/感染細胞それぞれから採取した試料それぞれをチップ上のcDNAとハイブリダイズさせ、その蛍光の強度の比によって、発現量を測定する。

ただし、ハイブリダイゼーションの強度による発現量の測定は簡単ではないといわれている。それは、ハイブリダイズする塩基配列のパターンによって、そのエネルギー的安定度が異なるためである。たとえば、G-C結合の含有量が高い2本鎖は、A-T結合の含有量が高いそれに比べてエネルギー的に安定するため、遺伝子の発現量とハイブリダイゼーションの強度の観測は必ずしも一致しないという問題による。この問題は、ハイブリベースの手法には共通のものであり、前述のBrownらのシステムによる発現量の測定結果に対しても指摘されることもある。

© Affymetrix社 GeneChip

このチップ⁹⁾はプローブ合成型に分類される。ガラス基板上に配置されるターゲットDNAは、DNA配列伸長反応をマスクングによって制御しながら層状に生成される。

プローブDNAはガラス基板上に配置される。チップはfeatureと呼ばれるグリッド状の区画に分割されており、同一feature内は同じプローブDNAが配置される。

各featureに任意のプローブDNAを生やすことが可能である。プローブDNAは、半導体チップ製造に用いられる光リソグラフィ技術を用いて層状に生成する(図-5)。

その際に、プローブDNAの5'端は配列伸長反応を一時停止するように加工される、ただし、この部分は光の照射によって配列伸長反応を再開するようになっている。この状態で、マスクを介して特定のfeature群のみに水銀灯の光を照射することによって、特定のfeatureのプローブDNAの伸長反応を活性化させる。この操作を1塩基分の伸長に4回(A,C,G,Tに相当)繰り返すことによって、1層分を完成する(1層の生成に1時間程度を要するといわれる)。以降、上記の操作を繰り返す。N塩基長のプローブDNAを生成するためには、マスキングをしての光の照射とヌクレオチドの配置を4N回繰り返す。この過程を複数のチップ分を同一基板上でグリッド状に行って1チップ分を切り出すことによってチップ製造の効率を上げている。文献10)には、49から400枚のチップを同一基板上で作成していると説明されている。チップの集積度は、たとえば、文献11)で用いられているチップで1.64cm四方のチップ内に65000個以上のfeature(各々長さ25塩基のプローブDNAを 10^7 本程度含む)を配置することが可能なオーダーとされている。

上記のプローブDNAの配置はDNAチップのユーザが行うわけではない。現在のところ、特定の対象DNA(またはその配列変異)の検出用にプローブDNAが配置されたチップがカートリッジ化された形で販売されている。このカートリッジ(縦7cm横4cm程度のプラスチック樹脂製)は、プローブDNAが配置されたチップ本体のほかに、ハイブリダイゼーションの結果をスキャナで読み取るための透明樹脂製の窓、対象DNAを含んだ試料を注入するためのコネクタを備える。

課題

Affymetrix社では、カスタムチップの作成も請け負っているが、ユーザ自らが研究室レベルでプローブDNAの配置の設計とチップの作成を行うようなDNAチップ作成キットは提供されていないようである。また、前述のBrownらと同様の目的の実験がGeneChipを用いて行われており、同社からは、GeneChip Ye6100という製品(酵母の約6千個の遺伝子を載せたGeneChip)が販売されている。

Brownらが酵母の全遺伝子の配置に百枚以上のチップを使ったのに対して、この製品は、1.28cm四方のチップ4枚組で構成されているコンパクトさが特徴である。GeneChipでは、プローブDNAとして使用するDNA配列長が高々25程度であることから、長いDNA配列を検出するには工夫が必要になってくる点を考慮しなければならない。対象DNAをズレながら覆うように長さ20程度のプローブDNAを複数用意して、得られる情報に冗長性を加えようとするとも行われる。

まとめ

細胞内における遺伝子の発現量を観測する装置としてのDNAチップを紹介し、周辺の情報処理技術を紹介した。本文中でも指摘したように、DNAチップの観測量の精度をいかに改善してゆくかは今後の大きな課題として残っている。観測されたか否かの2値情報だけで知識発見を行ってゆく場合には、現状のDNAチップからでも十分満足な情報が得られる。しかし観測量の精度に敏感な処理、たとえばクラスタリングを行う場合には、コストがより高くなるものの精度が保証されている他の方法で観測するのが望ましい。著者らは大阪大学の久保研究室と共同で、ヒトおよびマウスの遺伝子を発現量に従ってクラスタリングおよびアソシエーション解析するプロジェクトを進めているが、従来型のゲル泳動を使った手法により発現量を観測している²⁾。ヒトやマウスでは酵母とは異なり、遺伝子破壊等の外部からの刺激を細胞に対して施すのが困難な場合が多いため、発現量をいかに高精度に観測するかが知識発見の鍵となる。

このようにDNAチップに対して慎重な考え方はあるものの、大量の観測を低コストで実現する方法としてDNAチップに対する期待は大きい。観測量の精度が改善されていった場合には、医学および生物学の実験で欠かせない基本的道具として定着することが予想される。

謝辞 大阪大学細胞生体工学センターの久保先生との議論は大変有益でした。ここに深謝いたします。

参考文献

- 1) Kozal, M.J., Shah, N., Shen, N., Yang, R., Fucini, R., Merigan, T.C., Richman, D.D., Morris, D., Hubbell, E., Chee, M. and Gingeras, T.R.: Extensive Polymorphisms Observed in HIV-1 Clade B Protease Gene Using High-Density Oligonucleotide Arrays, *Nature Med.*, Vol.2, pp.753-759 (1996).
- 2) 北, 川本, 森下, 大久保: 1000の神経関連遺伝子の時・空的発現様式による分類, 日本分子生物学会年会講演要旨集, p.223 (1998).
- 3) Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P.O. and Herskowitz, I.: The Transcriptional Program of Sporulation in Budding Yeast, *Science*, Vol.282, pp.699-705 (1998).
- 4) Akutsu, T., Kuhara, S., Maruyama, O. and Miyano, S.: A System for Identifying Genetic Networks from Gene Expression Patterns Produced by Gene Disruptions and Overexpressions, *Genome Informatics 1998*, pp.151-160 (1998).
<http://giw.ims.u-tokyo.ac.jp/giw98/cdrom/papers.html>
- 5) Fodor, S.P.A.: Massively Parallel Genomics, *Science*, Vol.277, pp.393-395 (1997).
- 6) DeRisi, J.L., Iyer, V.R. and Brown, P.O.: Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale, *Science*, Vol.278, pp.680-686 (1997).
- 7) <http://cmgm.stanford.edu/pbrown/>
- 8) <http://www.synteni.com/>
- 9) <http://www.affymetrix.com/>
- 10) *Nature Genet.*, Vol.14, pp.367-370 (1996).
- 11) Winzler, E.A., Richards, D.R., Conway, A.R., Goldstein, A.L., Kalman, S., McCullough, M.J., McCusker, J.H., Stevens, D.A., Wodicka, L., Lockhart, D.J. and Davis, R.W.: Direct Allelic Variation Scanning of the Yeast Genome, *Science*, Vol.281, pp.1194-1197 (1998).

(平成11年1月26日受付)