

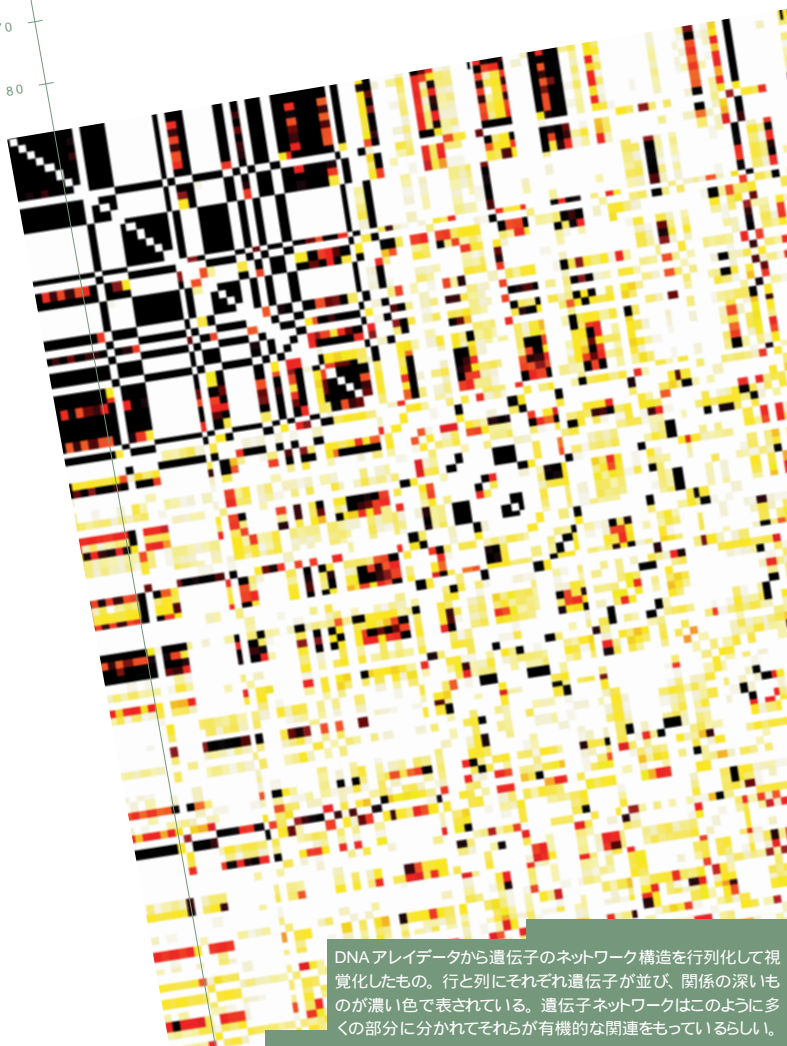
第6回(最終回)

# 遺伝子発現データの解読

## 情報の選別が鍵を握る

遺伝子は常時たんぱく質を作り出しているわけではない。それにはONとOFFの状態があり、必要なときに必要な遺伝子がセントラルドグマによってmRNAに転写し、アミノ酸に翻訳される。このような発現状態を解析すると、多くの遺伝子がネットワークを構成する複雑なトポロジー空間をもつことがわかってきた。それらが解明できれば、個人の特性に合わせたテーラーメイド医療も可能となるだろう。

浅田一憲 株式会社オープンループ、北海道大学大学院医学研究科  
 多田光宏 北海道大学遺伝子病制御研究所、株式会社ジエネティックラボ



DNAアレイデータから遺伝子のネットワーク構造を行列化して視覚化したもの。行と列にそれぞれ遺伝子が並び、関係の深いものが濃い色で表されている。遺伝子ネットワークはこのように多くの部分に分かれてそれらが有機的な関連をもっているらしい。

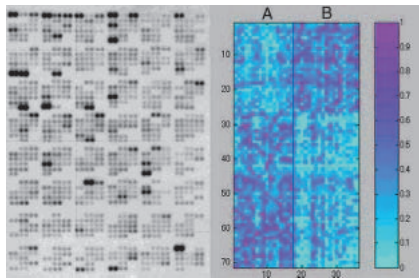
### 遺伝子発現を測定できるDNAアレイ

最近のバイオ技術は、ITと数学の取り込みによって目覚ましく発達してきている。これまで行われていた試験管による多くの実験では考えられなかったような、大量のデータを一挙に取得する方法——ハイスループット解析——が可能になってきた。DNAアレイ(DNAチップ)と呼ばれる技術もその一例で、多数の遺伝子の発現状態を一挙に測定することができる優れたものだ。このDNAアレイは、多数の遺伝子を基板上またはフィルター膜上に配列したもので、調べたいサンプルと反応させると、発現した遺伝子の量、つまりmRNAの量をいちどきに測定できる。これを使えば、1回の実験で数千～数万種の遺伝子発現量がデータとして得られることになる。

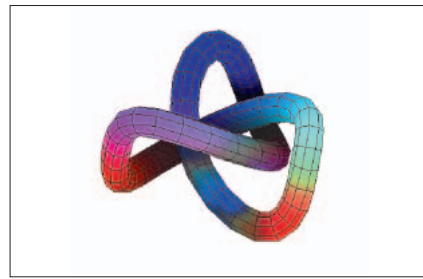
さらにDNAアレイが何万枚もあれば、何万人もの遺伝子発現データをそれぞれ数万遺伝子について測定することもできる。このような測定データを集めて解析することは、医学的に大きな意味をもつ。人による遺伝子発現の違いを明らかにし、生命のメカニズムの解明に役に立つからだ。

### フェノデータとジェノデータ

多くの人の遺伝子発現データを集めて



DNAアレイ原データ 左とそのパターンを可視化したもの。



DNAアレイデータがなす複雑トポロジーの例(想像図)。青っぽいところが たちの良いがん、赤っぽいところがたちの悪いがんの遺伝子発現データ。絡み合う両者は境界が不明瞭なため、従来の統計学的手段ではうまく判別できない。

解析すれば、患者の診療データと遺伝子発現データの間になんらかの相関性を見つけ出せるはずだ。いくつかの遺伝子はがんと密接なかかわりをもつことが知られているが、DNAアレイデータを用いると、今まで知られていないような遺伝子とがんとの関係までも、より大きな規模で明らかにできるだろう。たとえば、多くのがん患者の遺伝子発現データを集めて解析してみると、特定のがんの患者に抗がん剤を投与して効果の有無を見るのに、1~2個ではなく、複数(数十~数百個)の遺伝子の発現量のパターンに特徴があることがわかるという具合だ。これがわかれば、このデータは今後の治療に大変役立つだろう。

遺伝子発現以外にも、SNPsなどの遺伝子配列データも病気の程度やかかりやすさと密接なかかわりをもつ。このように、臨床医学的な現象データ(フェノデータ)と遺伝子データ(ジェノデータ)の相関性を解析することは今後の医療にとって非常に重要になってくる。

#### 大量のデータからパターンを見つける

数種類のサンプルで遺伝子発現量の差異を見るような場合には、データの数字を並べて手作業で検討してもそれほど時間

がかからないが、サンプルが数十にもなると、とてもそんなことをしてはいられない。DNAアレイで測定した膨大な遺伝子発現データを解析するには、コンピュータの力を借りなければならない。そこでハイスーパー解析技術が活躍する。

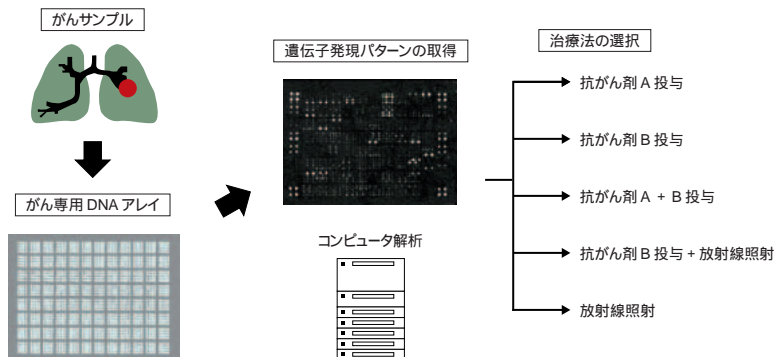
たとえばDNAアレイデータが1万個の遺伝子発現データだとして、それを100×100ピクセルの画像データであると考えてみよう。この画像データの各ピクセルには、高い発現値を示すものは濃い色、低い発現値では薄い色が与えてある。こう考えると、DNAアレイデータは画像パターンと捉えることができる(上左画像)。また、ピクセルでなくて文字が書かれていると想像してもよい。このようなアプローチを行えば、DNAアレイデータには統計学やパターン解析が適用できる。

さて、この文字のパターンを解読しなければならないが、これはそれほど簡単ではない。ここでは仮に「違」と「異」という文字が書かれているとしよう。この2つの文字は意味するところが「ちがう」「ことなる」ということで大体同じだが、パターンはまったく異なっている。それぞれは「遺」や「胃」のほうがパターンはずっと似ているが意味が違うように、実際の遺伝子発現も単純ではない。「この遺伝子が発現したらこ

SNPs.....「Single Nucleotide Polymorphism (一塩基多型)の略。ヒトの個人間の1塩基の違いを表すもの。疾病関連の遺伝子を探したり、病気のリスクを診断したりするのにSNPs情報が利用できると考えられている。

mRNA.....メッセンジャーRNA(messenger RNA)のことで、DNAに保存されている遺伝情報を転写したものである。生物の遺伝情報はDNAに保存されており、この遺伝情報を子孫に残すに際しては、DNA RNA たんぱく質という向きで遺伝情報が流れる。

癌のテーラーメイド医療のイメージ



の病気」と単純に決めることはできず、他の遺伝子や発現制御シーケンスと密接にかかわった複雑なものであるらしい。細胞の似た表現型を作る遺伝子の発現パターンは、複雑なトポロジー空間をなすのである。これは細胞の遺伝子発現ネットワークが単純なものではなく、カオスと隣り合わせの複雑系であることを反映している(前ページ上右画像)。

そればかりではない。DNA アレイデータは細胞の状態の微妙な差、実験での誤差によりパターンが少しずつ異なる。すなわち「文字がかすれたり、汚れたり」しているのだ。この分析には昔からIT分野で使われているパターン解析の方法を用いるのが有効だ。データ空間を幾何学的に考えて分類するk-最近傍法やクラスター分析、あるいはコンピュータに学習を行わせるニューラルネットワーク、遺伝的アルゴリズム、サポートベクターマシン、隠れマルコフモデルなどである。

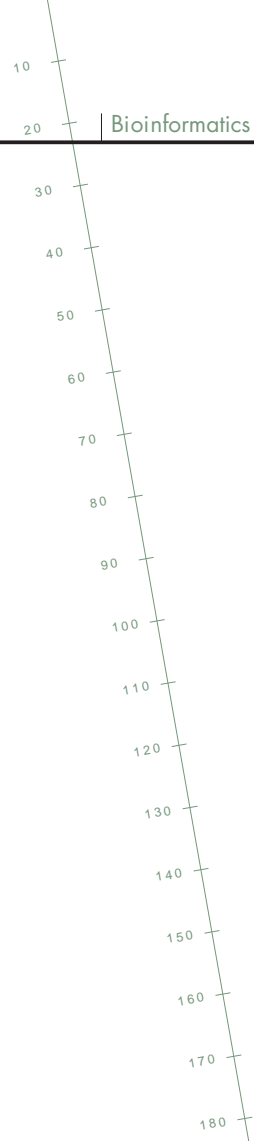
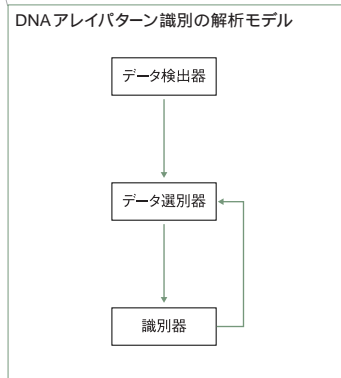
これらのアルゴリズム、特に学習アルゴリズムは、似て非なるもの、あるいは似ていないが同じものが混在する複雑トポロジー空間におけるパターンの認識や判別に威力を発揮する。しかし、これらの強力なアルゴリズムを使っても、得られたデータをすべて入力して解析したのでは、多くの場合ノイズが大きすぎて良好な識別が得られない。そこで我々は、判別にもっとも価値の高い情報のみを選択するために、

判別アルゴリズムを通す前に判別器に渡すデータを取捨選択し、それを用いた判別結果をフィードバックする手法を用いている。文字の認識でいうと、文字全部よりも「へん」や「つくり」などのエッセンスになる部分を自動的に発見して、より効率的で正確な識別を行うアプローチである。

テーラーメイド医療への道

DNA アレイから得られる情報量が非常に多いことから、その中に病気の状態や、治療に対する反応を予測できる情報が含まれていると考えられる。こうした情報を利用すれば、個々の患者の状態に合わせた治療法を的確に選択できる可能性があり、多くの期待が寄せられている。それはお仕着せの画一的な治療ではなく、テーラーメイドの医療の実現である(上図)。大量のデータの中からそうした情報をいかにうまく引き出すかがDNA アレイ解析の課題であり、トランスクリプトームという言葉は遺伝子の発現(転写)を解剖して真実を探るという意味が込められた言葉である。

バイオインフォマティクスは生命の神秘をIT的なアプローチで探ろうとする学問である。生命という大きな恵みが我々の身近なものになるに従って、我々自身も恵みを与えられるような存在になることができるに違いない。





## [インターネットマガジン バックナンバーアーカイブ] ご利用上の注意

このPDFファイルは、株式会社インプレスR&D(株式会社インプレスから分割)が1994年～2006年まで発行した月刊誌『インターネットマガジン』の誌面をPDF化し、「インターネットマガジン バックナンバーアーカイブ」として以下のウェブサイト「All-in-One INTERNET magazine 2.0」で公開しているものです。

<http://i.impressRD.jp/bn>

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容(技術解説、URL、団体・企業名、商品名、価格、プレゼント募集、アンケートなど)は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者(執筆者、写真の撮影者、イラストの作成者、編集部など)が保持しています。
- 著作者から許諾が得られなかった著作物は収録されていない場合があります。
- このファイルやその内容を改変したり、商用を目的として再利用することはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用する際は、出典として媒体名および月号、該当ページ番号、発行元(株式会社インプレス R&D)、コピーライトなどの情報をご明記ください。
- オリジナルの雑誌の発行時点では、株式会社インプレス R&D(当時は株式会社インプレス)と著作権者は内容が正確なものであるように最大限に努めましたが、すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接のおよび間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

このファイルに関するお問い合わせ先

**株式会社インプレスR&D**

All-in-One INTERNET magazine 編集部

[im-info@impress.co.jp](mailto:im-info@impress.co.jp)