

バイオインフォマティクスの発展と東亞保有の関連技術

新事業企画開発部 企画開発グループ 神谷欽也

1990年代初頭に始まったヒトゲノム計画に端を発したヒトゲノムの解読競争は、最終的には公的機関の国際合同チームと民間企業が協調して、21世紀を迎えるのと前後してその終了宣言が出された。しかしその実態は、精度の低い大雑把な配列を決定したのみであった。現在、完全に配列を決定されているヒトゲノムは、21番および22番染色体についてのみである。ここでいうヒトゲノムの解読とは、単にゲノムの全塩基配列を決定することであり、個々の遺伝子がどこにあり、その機能が何であるかがわかったわけではない。実際、これまで全ゲノムが決定された比較的単純な生物種でも、ゲノム中に含まれる遺伝子のうち機能が分かっているものは半分以上に過ぎない。人類を病の困苦から救う万能なイメージのある「ゲノム創薬」という言葉も、医薬品の標的とすべきタンパク質や遺伝子がまだ十分にわかっていないため、実際には手探りで始まったばかりである。

このような中、大量のゲノム配列情報を効率よく解析して研究に生かす目的で、生命科学と情報科学が融合したバイオインフォマティクス（生命情報学）と呼ばれる分野が急速に発展し始めた。生命科学研究の流れの中で「ポストゲノム」時代として幕を開けた21世紀は、バイオインフォマティクスをはじめとする大量情報処理技術により、効率的に有用な遺伝子を見つけることが必要不可欠な時代となっている。

当社でも、世界に認められたバイオインフォマティクス技術を保有している。人類共通の財産であるヒトゲノムの解析に、これらの技術を有効に活用することで社会貢献するとともに、ポストゲノムの潮流に乗り大きなビジネスチャンスを掴めるのではと期待している。

1 ポストゲノム時代の到来

90年代初頭に端を発したゲノム解析は、ヒトゲノム計画との絡みもあり「遺伝子が分かれば、病気を根本的に治療できる医薬品の開発につながる」というゲノム創薬の概念が確立され期待を膨らませた。しかしながら、ゲノム解析が進むにつれ、人間の遺伝子情報がそんなに単純な形では取り扱えない事が分かってきた。たとえば、1つの遺伝子から生成したタンパク質が、その後の修飾を受けることで種々の機能をもつこともその一例である。そこで、もっと遺伝子数の少ない生物や、あるいは実験しやすい哺乳動物のゲノムやプロテオーム^{*1}の解析を行い、そこで得られた情報をヒトに応用しようという動きも出てきた。現時点では、同定した遺伝子の塩基配列にアノテーション^{*2}とよばれる機能などの情報をつけたデータベースの作成を、国家的プロジェクトで推し進めるなどの地道な作業を行っている状況である。このような地道な努力の結果、多くの知見が得られ始め、一部には医薬品開発の糸口が得られたなどの報道を目にするようになった。

このような流れの中で、大量な遺伝子の塩基配列情報、ゲノミクス技術やプロテオミクス^{*3}技術、さらにはバイオインフォマティクス技術を駆使して総合的な解析を行うことが、創薬プロセスを促進するドライビングフォースになると言われている。このバイオインフォマティクスとは、ゲノムの塩基配列からタンパク質の構造・機能にいたるまでの生物情報を扱う総合的な科学で、ライフサイエンスと情報科学が融合する研究分野のことを指す。その範囲は、確率モデル、学習理論、体系的なネットワーク

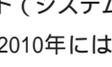
解析、画像処理、並列情報処理などまでも含むといわれ、きわめて数理工学的色彩の濃いものである。

*1：プロテオーム（proteome）とは蛋白質（protein）+ ゲノム（genome）から派生した言葉で、ゲノム或いは組織が発現した蛋白質という意の造語

*2：アノテーションとは、配列情報をもとにして解析された、その遺伝子に関連のある遺伝子やタンパク質、さらには、発現する体内の場所や、多形、疾患等に関する情報のこと。ゲノム情報の場合には、そこから読み取られる遺伝子の配置と塩基配列が主な情報となる。

*3：プロテオミクス（proteomics）とはタンパク質（protein）とゲノミクス（genomics）の造語で、ゲノムからだけでは解明できないタンパク質における各種の相互作用を解析・証明することを意味する。

2 バイオインフォマティクス市場の動向

バイオインフォマティクスに関係するハード、ソフト（システムやデータベースを含む）の市場は、のように2010年には1兆円産業となると予想されている。

現時点で本市場は、システム開発受託、ソフトウェア開発、データベース開発、データベースをもとにした情報解析の受託、ネットワークによるサービス提供の5つに大きく分類できる。これらは、いずれも市場規模が増加すると予想されている。²⁾

現在、国内外を問わず、ベンチャーを含めた多くの企業が本分野に参入をしている。国内では本命である製薬企業の他に、国

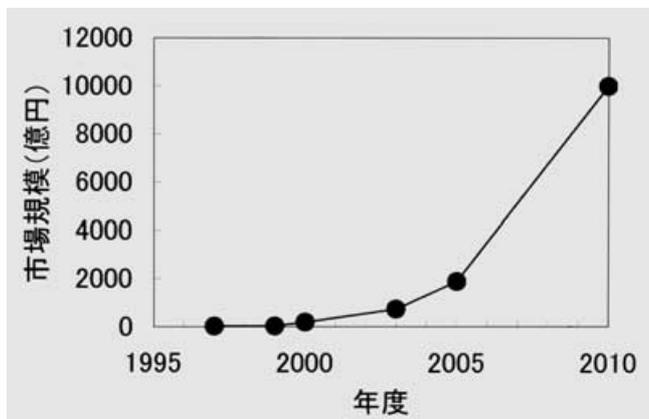


図1 バイオインフォマティクスビジネスの市場規模予想

内IT系企業、化学系企業、商社系企業に至るまで、大中企業本体あるいはその子会社やジョイントベンチャーが種々の内容で市場開拓を行なっている。このように多くの企業がバイオインフォマティクスビジネスに参入しているのは、主に本市場が新しい分野であることに起因しているが、現時点においてはユーザーニーズが多岐に渡っているため、技術やサービスといった商品提供での住み分けをある程度可能にしていることも一因であると考えられている。

また、全体的な研究トレンドとしては塩基配列の解明から有用な遺伝子の解析、蛋白質の相互作用、さらにはそれらのネットワークの解明へと進み、最終的には遺伝子解析から得た個々の部品を集めて人工的な細胞/生物の再構築へと進んでゆくと考えられている。研究対象が変われば、求められる技術やサービス対象顧客も変わるため、本分野ではビジネス上の変化が速い。このため、いかに研究者のニーズ情報をいち早くキャッチし、その変化に柔軟かつすばやく対応できるかという点がきわめて重要である。これらの条件を満たしながらリスク管理しやすい形として、大手企業は子会社をベンチャー化する、あるいは共同出資でベンチャーを設立するという方向でビジネス展開を開始している。

3 当社のバイオインフォマティクス技術とその展開

1998年秋、名古屋大学大学院多元数理科学研究科のグループ(大澤研二助教授、尾畑伸明助教授(現・東北大学)、当社・現・新製品開発研究所 吉田徹彦専門副主幹)は、ロレアル賞を受賞した。ロレアル賞は、色を対象に科学と芸術の掛け橋となるような作品や研究を発表した人を表彰する目的で、化粧品で有名なロレアル・グループが基金を設立して1997年に創設された。あらゆる分野の人々を対象に広い意味での色に関する論文・作品などを募集し、科学と芸術の創造的な出会いに貢献している人を表彰している。これまでに選考委員にノーベル賞受賞者を迎えるなどし、現在では話題性のある国際賞として知られている。

吉田専門副主幹らのグループの受賞理由は、自然界に潜んでいる規則性・不規則性を色彩を用いて顕在化させる独自の方法論の研究を通じ、科学と芸術の双方の発展に寄与したことによ

るものであった。その研究における方法論を基本に出願し、日本においては2001年3月末に特許が成立した(登録特許3149824)。

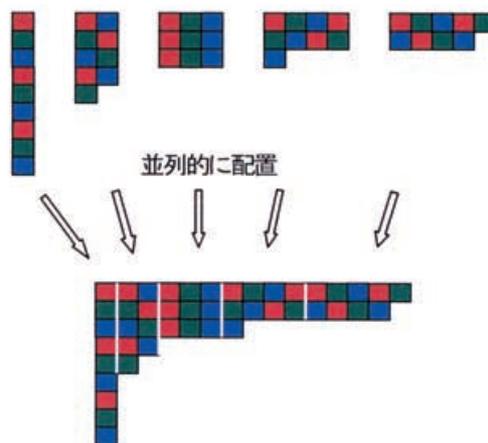
この技術は「二次元色彩パターン化による一次元配列の情報解析法」(通称「二次元色彩パターン化法」あるいは「2D色彩法」というものである。以下に二次元色彩パターン化法の簡単な説明をする。

二次元色彩パターン化の原理は、まずある一次元配列を、数列を1つずつ増加させた二次元配列に変換し、並列的に配置する。これを配列中の記号毎に彩色することを、二次元色彩パターン化と呼ぶ。

例えば、記号 a, b, c で表わされる「abcabcabcabc」のような並びの配列がある場合、下記のように、数列を1つずつ増加させた二次元配列に変換し、次に並列的に配置する。

1列	2列	3列	4列	5列
a	ab	abc	abca	abcabc
b	ca	abc	bcab	cabc
c	bc	abc	c	
a	ab			
b	c			
c				
a				
b				
c				

これを、下記のように配列の記号毎に彩色 (a = 赤、b = 緑、c = 青) し、並列的に配置する。



「二次元色彩パターン化による一次元配列の情報解析法」とは、以上のような処理を行ない、模様として浮き出る情報から、視覚的に一次元配列に潜む規則性を見出すことを意味する。このような方法の解析対象としては、遺伝子を構成している塩基配列、タンパク質を構成しているアミノ酸配列の他、数字列・数値列や単語列、音韻列など、一次元配列として示すことができる情報は全てと言える(表1)。

これまでに本技術の応用として文献や学会にて発表した主な例(3)(4)(5)(6)としては、塩基配列の中に潜むこれまで知られていなかった

表1 2次元色彩パターン化法の応用可能性の例

分類	可能性のある応用分野
自然科学系	遺伝子領域予測、イントロン・エクソン領域予測、蛋白質(翻訳)領域予測(mRNA)、遺伝子やアミノ酸配列(タンパク質)に存在する周期性解析、地震発生予測、火山噴火予測、宇宙からの信号解析、計量・診断
社会科学系	経済(景気)変動予測、株価予測
人文科学系	文学作品の感性情報解析(種々の文学作品の作風解析)
芸術系	種々のデザインの創出、音楽作品の感性情報解析(音楽情報の色彩パターン)

た繰り返し配列(タンデム配列)を発見したことが挙げられる。大腸菌⁷⁾や酵母⁸⁾、さらには全配列が決定されているヒト21番染色体⁹⁾にも応用し、新規な繰り返し配列を多数発見するに至った。ゲノムプロジェクトの研究の結果、ヒトゲノムの半分は、直接生体機能に関与しない繰り返し配列が占め、残りに様々な遺伝子が存在することが分かってきた。また、1~4個の塩基の繰り返し配列であるSTR(Short Tandem Repeats)の多くはゲノム中の遺伝子以外の領域にあり、個人によって繰り返し回数が大きく異なることから多型マーカーとして注目されてきたが、STRの中には蛋白質をコードする領域にあるものがあり、変異により癌抑制遺伝子の不活性化などに関与する場合があることが確認された。さらに、近年ハンチントン病など数多くの遺伝病が、原因遺伝子の中にある3塩基繰り返し配列(トリプレットリピート)が伸長することが原因で発症することもわかってきた。このほか、繰り返し配列との関係に関しては、生物の進化遺伝学や動物の性格・行動遺伝学の面からも研究¹⁰⁾¹¹⁾がなされている。

このように近年繰り返し配列は、遺伝子研究上や疾病研究において重要なものであることがわかってきたが、まだ研究途上の分野でもありその他の繰り返し配列が何を意味するのかは十分に解明されていない。特に、他方法に比べ2次元色彩パターン化法がその発見に有効に役立つと考えられる、もっと塩基数の多い、すなわちもっと長い繰り返し配列や、一部が変異している繰り返し配列については情報が少ない上、研究も始まったばかりで、それらの機能はまだはっきりとしていない。遺伝子の塩基配列の中に潜む繰り返し配列をはじめとした種々の情報を効率的に同定することは、遺伝子の機能を考える上で今後ますます重要なものになると考えられる。本技術によって見いだされた新たな繰り返し配列が研究対象となり、いずれは疾病などとの関係がわかる可能性も大いにあると考えられる。

ここで紹介した技術の他、当社はバイオインフォマティクス分野で有用な技術や知的財産を複数保有している。今後は、これらの技術あるいは知的財産をこれまで以上に社内で活用する方針である。それとともに、今後大きく成長が期待できるバイオインフォマティクス分野においてこれらの技術を広く活用するため、既存ベンチャーとの提携もしくは複数の企業との共同出資によるベンチャー設立を視野に入れた活動を積極的に行う予定である。

人類共通の財産であるヒトゲノムの解析にこれらの技術を有効に活用することで社会貢献するとともに、ポストゲノムの潮流に乗り大きなビジネスチャンスをつかむことを期待している。

引用文献

- 1) バイオテクノロジー基幹技術に関する技術動向調査, 特許庁(2001) および バイオインフォマティクスビジネス/2003, サイリック(2001)などを参考に筆者作成。
- 2) バイオインフォマティクスビジネス/2003, サイリック(2001) p.35.
- 3) T. Yoshida, N. Obata, K. Oosawa, *Genome Informatics* 1997, 8, 360 (1997)
- 4) 吉田徹彦, 大沢健二, 日本生物物理学会第35回年会予稿集, 1997, S128.
- 5) 大沢健二, 吉田徹彦, 尾畑伸明, 日本生物物理学会第36回年会予稿集, 1998, S96.
- 6) 吉田徹彦, 大沢健二, 尾畑伸明, 日本生物物理学会第36回年会予稿集, 1998, S123.
- 7) T. Yoshida, N. Obata, K. Oosawa, *J. Mol. Biol.*, 298, 343 (2000)
- 8) 吉田徹彦, 大沢健二, 尾畑伸明, 第24回日本分子生物学会年会予稿集, 2001, 印刷中.
- 9) 投稿中
- 10) R.P. Ebstein *et al.*, *Nature Genet.*, 12, 78 (1996)
- 11) 都築政起, *化学と生物*, 2001, 656