



WMA

IT の進歩と医療

## 情報医学とゲノム科学によるヘルスケアの変化



Ju Han Kim

国立ソウル大学校医科大学バイオメディカル  
インフォマティクス部門助教授

### はじめに

私は韓国医師会の情報通信技術部長を務めており、また国立ソウル大学校医科大学バイオメディカルインフォマティクス部門に所属している。

さて、「バイオメディカルインフォマティクス、情報科学、およびゲノム科学によるヘルスケアの変化」ということであるが、これはゲノム科学、プロテオミクス、バイオメディカルインフォマティクスのインテリジェントな統合に関するものである。ゲノム科学とプロテオミクスの一大革命の進展についてはご承知のとおりであり、またゲノム科学、プロテオミクス、バイオインフォマティクスの3つの概念がセットで使われているのを耳にされているであろう。

遺伝子と蛋白質は兄弟であり、ゲノム科学とプロテオミクスが組み合わせて使われるには納得できるだろう。これに対して「インフォマティクス」はいわば「ドライ」な科学、言い換ればコンピュータ科学の分野に属するもので、分子生物学やバイオ医学の分野にとっては非常に新しい要素である。しかしながら、昨今は「ウェット」な生物学と「ドライ」な科学や技術

を組み合わせた「インフォマティクス」の登場により、これらを組み合わせた3つの分野がまとめて言及されることが多い。

本日の演題はこの3つの革命的学問分野の統合と情報革命についてであるが、最初に、バイオメディカルインフォマティクスとゲノム科学の統合を推進している要素について述べる。次いでこの分野で起きているパラダイムシフトについて説明し、バイオチップテクノロジー、すなわちDNAチップ、遺伝子チップ、プロテインチップをはじめとするさまざまなバイオチップを扱う分野について触れたいと思う。これらは生物学における大規模な垂直型実験である。また、バイオチップインフォマティクスが臨床にもたらす意義について説明する。さらに、統合的なバイオチップインフォマティクスの抱える課題、こうした複雑な科学技術がどのような方法でバイオメディカルインフォマティクスやゲノム医学と組み合わされるのかを紹介する。そして最後に、新たな医学の台頭について述べて結びに代えたい。

### ■ 遺伝子情報の増大

NIHのバイオ医学刊行物を集めた世界最大規模のデータベースは拡大の様相を呈している。

バイオ医学の全分野について現在約150万件の論文が登録されており、その数は日に日に増大しており、現在では、研究者は遺伝子シークエンスを Gene Bank や、ヨーロッパのバイオ情報科学では Ensemble などのデータベースに登録しているので、ゲノムのシークエンスも文献データベースと同様、データとして扱われている。これは1990年代半ば以降、急速に増大しているが、何を意味するのだろうか。確かに、なかには遺伝子を扱っていない論文も含まれているかもしれないし、1つまたはわずかな数の遺伝子、あるいはその一次解析またはその詳述しか扱っていない論文もあるかもしれないが、意味するところは、シークエンスの数、一次解析も完了していない段階の遺伝子データの量があまりに多く存在するということである。1,000万以上の遺伝子シークエンスが、分析もされずに登録されている状況なのである。

シークエンスを例に述べたが、これは状況のほんの一部にすぎない。シークエンスを得たら、リンクージマップ、物理マップ、多型等によってシークエンスのアノテーションを行う。さらに重要なことに、同じ程度に複雑な発現プロファイルも手に入るようになっている。こうした構造情報も含めるとしたら、データ、すなわちゲノムに関するデータの爆発的増大がどれほどものになるか、予測もつかない。昔の分子生物学のパラダイムにおいては、データそのものが科学のすべてであった。したがって、実験を行い、正しいデータを取得することが科学全体において最も重要だった。

しかし現在では、微生物の全シークエンスの解析をわずか1日か2日で行うことができる。つまり、データはすぐに手に入るので、次の問題は、このデータをバイオ医学に役立つ知識とするにはどう扱い、どう解釈すればよいのかということになる。

もうひとつ、病院情報システムの話を紹介しよう。私の勤務するソウル大学校病院では昨

年(2003年)、1,000床を擁する病院および別の入院施設を擁する病院で、電子医療記録システムを開発した。その結果、完全なペーパーレス、フィルムレスが実現した。臨床情報、カルテ、画像、研究所のデータはすべてコンピュータシステムに入っている。

そしてまた、臨床サイドにおける爆発的なデータの増大にも直面している。この爆発的増大こそ、ゲノム科学とインフォマティクスという革命を推進する中心的要素なのである。もちろん、しばらく前からすでに存在しているデータベースとコンピュータ技術、そしてインターネットにより、全情報が統合的環境におかれるようになったのである。

### III 生命の構成要素はすべてデジタル

セントラルドグマでは、情報はDNAからRNAを経て蛋白質へと伝達されること、RNAポリメラーゼがDNAの情報をmRNAに転写し、tRNAと共にリボソームに伝えられ、蛋白質へと翻訳される。これらはすべてデジタルといえないであろうか。インフォマティクスに携わる者としては、セントラルドグマのプロセス全体が基本的に情報処理だといえるように思う。生命の構成要素はすべてデジタルであると。DNAは糸を引いているものと、蛋白質も糸を引いているものと見なすことができる。ではわれわれは目下、何をすべきなのか。こうしたデジタル情報をマッピングし、組み合わせることでシステム全体を再構築する必要があり、それによって相互作用やネットワーク、システム全体の理解が可能になるのである。

パラダイムシフトに関連して、1995年に出版されたメディカルインフォマティクスの最も有名な教科書(Edward Shortliffe著)から、メディカルインフォマティクスの定義を引用したい。Shortliffeは、メディカルインフォマティクスの分野における代表的な先駆者のひとりである。

「インフォマティクスとは急速に発展しつつ

ある科学的分野の1つで、バイオ医学関連情報のうち、問題解決および意思決定に必要なものについて、その保存・検索・管理の最適化に用いる資源・装置・正式な方法を取り扱う」。

この定義によると、「情報科学は問題解決および意思決定のためにある」とされている。言い換れば、「情報科学は何かを可能にするための技術にすぎない」ということであり、以前から可能であったことを行う意味では役に立つが、まだ実現していないことを行うには役立たない。それは以前の定義であったが、彼は新しい教科書でその定義を変えている。最初の部分は同じだが、問題解決、意思決定、医学的発見と理解のための技術だと言っている。彼はなぜ定義を変えたのであろうか。情報医学に新たにバイオインフォマティクスを加えたのだと私は思う。バイオインフォマティクスの大きな違いは、真の科学の追求に役立つという点である。

#### Ⅳ 情報の統合

ここで、科学と技術と工学の違いについて述べたい。科学とは、事実に関するものであり、事実を1回発見することである。2回目や3回目は意味がない。科学は自然に関するものであり、実際に発見することである。

インフォマティクスは、何かを促進し実現するためのものであり、主に技術と工学の分野に分類される。しかし、DNAがRNAを経て蛋白質に至る生物学的プロセスはすべてが情報処理であり、インフォマティクスは真のバイオ医学的な発見と理解を追及する科学の一部を担うことができる。バイオ医学の複雑な専門知識をすべて管理することは非常に難しい課題であるが、古典的パラダイムは臨床医を中心としている。臨床医は、図書館、コンピュータ、研究所、教科書などを活用する。そこからあらゆる情報を収集し、すべての情報を頭の中で1つにまとめて再構築する。

しかし、バイオインフォマティクスの場合、

こうした情報はすべて集中管理されており、研究者や臨床医はサイバースペースでデータを収集する。つまり、データや情報が複雑なため、1人の人間の頭脳ですべての情報を再構築するのは不可能だということである。情報はコンピュータやサイバースペースで集中管理され、すべての情報がサイバースペースで再構築されている。

臨床情報システムの導入という問題もある。これは単なる病院の情報システムのことではない。オーダー通信システム、研究室情報システム、RIS (radiology information system；放射線医学情報システム)、PACS (picture archiving and communication system；画像保管通信システム)、電子医療記録は統合されたが、異なる部門、図書館、研究データベースにまだ多くの情報システムが存在している。そのため、これらの情報を統合し、科学と実践を結合した臨床情報システムを構築する必要性を提唱する人が増えていている。

情報医学を構成する基本要素を示すA、B、Cがある。Aは“academic”，科学の部分、情報の原理と論理、Bは“bioinformatics”，最近この分野に加えられたバイオインフォマティクスである。Cは“clinical”，臨床インフォマティクスである。このように、理論分子から臨床プラクティス、つまりミクロの世界からマクロの世界へと移行している。インフォマティクスという同じ名称が付けられている理由は、臨床科学、バイオ医学、情報科学、および技術を共有しているからである。

なかには、Aを“administrative”，管理すること、物事を構築すること、Bを“business”，Eヘルス（インターネット通信による医療情報の交換や遠隔治療）などのビジネス関連の活動、Cを“computer”，医学生や医療関係者のコンピュータ技能だと誤った解釈をする人もいる。

もちろん、こうしたことでも重要だが、基本的な要素は、先に述べたA、B、Cの定義である。

さらに重要なことは、メディカルインフォマティクスとゲノム科学を統合する考え方が医学と生物学の両方で十分に発達したことである。生物学の分野にインフォマティクスが加わることで実験生物学から大幅に変化している。そのため生物医学は、情報の伝送と解読や、解読した情報産物の機能に関する基礎学問になっている。

#### IV バイオチップ

バイオチップは、技術をその1枚に凝縮したもので、医療分野で最も関係の深い技術であろう。バイオチップの動作原理は非常に簡単である。DNAがRNAに転写されて蛋白に書き換えられることはご存じのとおりである。DNAの情報をコピーして、mRNAを生成して高速で破壊すると、ある一定の状態で、mRNAの量、すなわちコピー数を、対応する遺伝子の活動として解釈できる。これは従来の分子生物学の事項である。重要な点は、ある状態でmRNAの量を測定できれば、一定の遺伝子の活動が測定できるという点である。

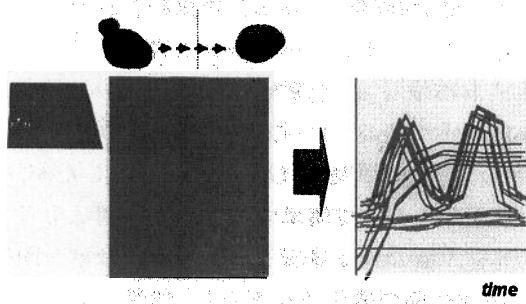
遺伝子は二重らせん構造をしており、遺伝子の量を相補的配列で測定できる。今ではヒトゲノムの全体の完全なマップが判明している。そのため、 $100 \times 100$ 配列で1万個の遺伝子を得て、そのすべてをアンチセンス法で1枚のスライドに置けば、全ゲノムの活動プロファイルが測定できる。これが全貌である。mRNAのコピー数を直接測定することは困難なため、それをフルオレセインで染色して、そのフルオレセインを光学スキャナーで強い光に変える。こうすることで、全ゲノムの全活動プロファイルを非常に簡単な技法で測定することができる。おそらく、現在は医学全体が診断、治療、予後、臨床知識の管理と提供方法に関わっていると思われる。こうした全分野でバイオチップとの関係がすでに確立されている。

#### ■ 予測モデル

ここで1999年、*Science*に掲載された論文を紹介する。マサチューセッツ工科大学(MIT)ホワイトヘッド研究所ゲノムセンターのTodd Golubが、“Molecular Classification of a Cancer(癌の分子分類)”と題した論文を発表したが、その概念はきわめて簡単である。縦の列は患者の症例を、横の列は遺伝子を示す。彼は単一配列で16,000個の遺伝子を測定し、急性リンパ芽球白血病と骨髄芽球性白血病を患う70人の患者から採取した。Golubの考えは、その高い数値の遺伝子をALLに、低い数値の遺伝子をAMLに集めるというものである。彼は16,000個の遺伝子から高い数値と低い数値の遺伝子をそれぞれ25個集めた。彼の考えでは、その遺伝子の働きを知らなくても、このプロファイルを用いれば、予測モデルを分類して構築できるというものであった。新しい症例に直面したら、この簡単なプロファイルを用いて遺伝子をALLかAMLに分類するが、それが診断になる。

もう少し詳しく説明しよう。その翌年、スタンフォード大学の研究者Alizadehが、リンパ腫、びまん性大B細胞型リンパ腫で同様の実験を行った。それは単一の臨床サブタイプである。Alizadehは、単一の病気であるにもかかわらず、2つの特異な病気パターンがみられることに気付いた。実際に2つの臨床サブタイプであるのか疑問をもった。そして、同大学の医の倫理委員会に報告した。これがKaplan-Meier法による生存曲線である。この生存曲線のパターンによって非常に異なる臨床結果が得られることを発見した。臨床的に低リスクのグループでさえこの表示プロファイルに従って分類することができる。今ではこれらの病気の予後を予測している。

重要な医療分野で対応されているその技術の優れた点は、何の遺伝子かを問う必要がないという点である。分子生物学では正確な分子メカ



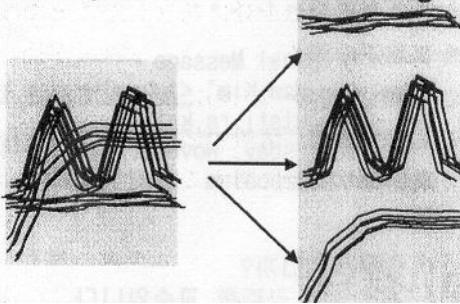
図①

ニズムを理解することは重要である。しかし、われわれは今の段階でそれを知る必要はない。われわれはゲノムタイプを発現型と遺伝子型に関連付け、予測モデルを作ることができる。そしてもちろん、詳細を研究し、分子メカニズムを解明することもできるのである。

## Ⅳ 未知の遺伝子の発現

近代天文学の起源は、古代バビロンに始まるといわれている。バビロニア人は満天の星空の完全な星図を作成した。複雑なデータを見るとき、最初にすべきことはデータをマップし、完全なマップを作成することである。すると、データはランダムでないことに気付く。そのデータには何らかの明確なパターンがあるはずである。バビロニア人は、マッピングして銀河系や太陽系などの存在を発見した。

たとえば、私の同僚の見解だが、満天の星空のような臨床ミクロデータ画像がある。この種の複雑なデータを見るとき、そのデータはランダムではなく、一定の特異パターンが存在することを確信する。有名な実験の1つに、酵母の細胞分裂の実験がある。スタンフォード大学が発芽酵母で実験したものである。2次元のマイクロアレイのチップがあるが、その遺伝子すべてをランダムにプロットしているため、1次元の数値配列になる。10分後、別の実験を行い、



図②

その後も10分ごとに繰り返す。そうすると、2次元のマトリックスが得られる。これを色分けすると、ランダムでないことが分かる。特異パターンがあるはずである。

複雑なシステムを完全にマップしたら、次に行うべきことは、その複雑なデータの中から特異パターンを発見することである。その特異パターンをクラスタリングという技法で抽出できれば、固有の働きがあることを発見できる(図①)。

横に遺伝子を、縦に条件を配置するのが慣例で、このようなクラスターが得られるので、このように抜き出して遺伝子の全知識を収集する。データベースから収集する必要がある。この技術の優れたところは、未知の遺伝子の発現プロファイルも同時に得られる点である(図②)。未知の遺伝子の実験を行うこともできる。マイクロアレイでは、それらをすべて同時に実行するため、未知の遺伝子の発現プロファイルが得られる。

未知の遺伝子の発現とは何だろうか。未知の遺伝子の働きは、この条件で同じパターンに示される遺伝子の働きに関連していると推理できる。これが、私の同僚による実験の全貌である。彼は、二細胞分裂周期でも同じ実験を行った。彼の仮説は、この種の二周期パターンの遺伝子は、二細胞分裂周期に関係している、というも

のだった。彼は、細胞分裂周期に関する800個の新しい遺伝子を発見したと報告した。同時に、酵母の細胞分裂回数に関する遺伝子は142個しか存在しないことが証明された。これは実験の大きな成果である。クラスタリングの概念はきわめて簡単である。マップを作成し、隣接する遺伝子をグループ化して、次の遺伝子を見つけていく。そうすると全体のクラスターを得ることができる。この簡単で強力な概念は、この種の複雑なデータの研究に活用できること、コンピュータ科学の能力を医療分野に導入していることを説明したわけである。

#### ■ インフォマティクスの統合

インフォマティクスの統合は、あらゆる実験を標準化された手法、つまり大規模並列処理で行ってきたため、現在の問題は、こうした世界中のデータをすべて収集し、単一のリポジトリーを構築し、すべてのデータを比較して、さらにバイオ学的システムからの洞察を得ることにある。

2001年、*Nature*は、無料の公衆のリポジトリーを構築することを研究者たちに推奨した。そして*Nature*は最終的に、2002年12月1日以降、説明された全データをCDで提出しなければ、マイクロ実験に関する論文の論評はしないことにした。

そこで、実験データを標準化することと、2次分析、あるいはメタ分析を行い、最終的に全体のシステムを明らかにすることの重要性を述べたい。標準化問題を扱う多くの作業部会が活動し、外部のデータベースにリンクさせるデータモデルを構築した。バイオチップインフォマティクスを構築するには、バイオチップのいわゆる研究の量産化が必要だと思う。

バイオ研究は崩壊寸前といわれている。異なる実験の文献を統合するのは困難である。しかし、この種のゲノム学とプロテオミクスでは、プロセスを小型化、合理化して全体の製造シス

テムと量産化パイプラインを構築する必要がある。別の統合方法として、こうしたパスウェイデータベースを私の同僚の実験や色分けで統合する方法がある。

統合化されたバイオチップインフォマティクスにはどのような意味があるのだろうか。全プロセスを標準化し合理化し、非常に体系的な手法で実験を行う。病院の臨床情報をバイオ実験に統合しなければならない。そうしなければ、スタンフォード大学の生存分析の論文は真実ではなくなる。現在、標準化の問題があるが、すべての統合にバイオ的存在論や通信基準が役立つ。民間のデータベースを調べ、統合バイオ情報医学システムを構築する必要がある。長期的に、こうした医学的、バイオ学的な情報を1つに統合したいのである。

癌研究の分野で1つのビジョンがある。米国国立衛生研究所の癌センターは、21世紀の癌センターは、次の3つの技術で転換が図られるというビジョンを描いている。その技術は、①インターネット、②ゲノム学、③マルチメディア臨床記録である。これに関する研究論文があるが、連携を図りながら作業を推進できる情報インフラとコンポーネントを支持している。

もう1つのビジョンに、臨床試験システムをプラットフォームに統合するというものがある。プロトコルベースのシステムで実行されている診療を現実の診療システムに転換するものである。こうしたことを統合したいのである。

さらに、もう1つのビジョンに、Mayo ClinicとIBMの共同による情報システムの構築がある。第1段階で、臨床情報システムを構築し、臨床データを統合した。第2段階で、そのバイオ実験データの統合を試みている。ソウル大学病院でも、まだ実現はしていないが、病院の情報システムを確立し、バイオ実験をいくつか行った。最終的に、XMLベース、文書ベースのシステムを用いて、この種のレガシー・システム、バイオ実験システム、臨床研究試験システ

ムを統合して、単一病院で全体のシステムの構築を実現させることを計画している。

## 結論

最後に、「オミック」革命に対する私の見解は、既存の技術の大規模並行処理というものである。単一遺伝子の活動を測定する技術が存在するので、それを大規模並行処理し、ゲノム学の配列決定やバイオチップ分析などのゲノム技術を構築した。今ではプロテインチップが作成できる。それは一種の分子情報化、つまり分子レベルの情報収集であり、水平的に統合されている。

水平的な統合とは何か。遺伝子と蛋白質の両方の活動を同時測定することは困難なため、垂直的ではなく水平的に統合することである。MAMSと呼ばれる次世代技術などが必要となるかもしれない。しかし現在、大規模並行処理の活用はきわめて簡単で、データ集約的な科学を構築し、「オミック」革命を起こしている。

それではインフォマティクス革命とは何だろ

うか。分子生物学には機能ゲノム学とバイオチップ学がある。細胞生物学には、E-cellと“*in silico*”生物学がある。生理学には、ニューロインフォマティクスと心血管シミュレーションがある。私は意図的にこれらの学問を1つにまとめた。実際に医学生が学ぶカリキュラムであり、分子生物学、細胞生物学、生理学、解剖学、薬理学、臨床医学、社会医療科学があり、これらはすべてデジタル化されている。インフォマティクス革命は、分子の非常に微細なレベルから社会医療に至るまでのこの複合体を統合する。インフォマティクス革命は、これらの学問を情報科学的に強化した縦型統合といえるかもしれない。

新しい医学の出現により、垂直的から水平的に移行するこの2つの革命の深い絡み合いが必要となる。両方とも重要であり、1つは、分子レベルでの情報収集、もう1つは、情報科学で強化され管理されたバイオメディカルインフォマティクスとゲノム医学である。