

ゲノム科学とコンピュータ科学

ゲノム科学の 現状と将来

現在、ゲノム解読が予想を上回るスピードで進み、各分野での応用の機運が非常に高まりつつある。ゲノムの塩基配列の解読に当たっては、膨大な分析結果をコンピュータに蓄積し、その都度、データとの照合を行うというIT技術との連携が必要不可欠なものとなってきている。NTTソフトウェアでも、このゲノム解読のための情報技術(IT)の開発を行っており、今後この分野への貢献が期待されている。21世紀の科学技術の発展のためには欠かすことのできない、このゲノム科学とITの現状について、現場よりレポートする。

はじめに

ゲノム科学に代表される遺伝子関連の研究開発は、近年著しい進歩を遂げ、その成果が産業応用に直接結びつくなど、大きな期待が寄せられている。

生物は細胞の中に染色体DNA(ゲノムDNA)をもっており、ゲノムDNAは遺伝子情報を持つDNA(デオキシリボ核酸)と塩基性タンパク質を主成分とする長い12本のヒモ状の鎖が相互に結合したもの(塩基対)である。ヒトやマウスのゲノムDNAは30億の塩基(文字)からなるといわれており、その塩基対の配列のなかに遺伝情報が隠されている。この文字の配列を解読する作業は、1990年に日米欧政府の国際共同プロジェクト「国際ヒトゲノム計画」で開始され、解読した遺伝情報が人間の染色体のどこに位置づけられているかといった詳細な解析が2005年に終了する予定で進められてきた。既に昨年10月に22番の染色体の解読が完了し、今年の4月には21番の染色体も完了した。しかしながら、一昨年5月に設立された米国のベンチャー企業が大規模な超高

速解析装置を導入して、9月よりヒトゲノムの解析に着手し、2003年までに全てのヒトゲノムの解読を終えると発表して以来、国際共同プロジェクトも目標を2003年までに繰り上げ、今年6月までに「ドラフト」と呼ばれる概略情報の解読が完了する運びとなっている。一方、同ベンチャー企業は今年6月末にはヒトのゲノム解読を終える予定であると発表し、関係者に大きな衝撃を与えている。ゲノムDNAの解読によって、その中に含まれる多様な遺伝子情報が解明され、それをもとに新たな医薬品の開発や遺伝子治療などが可能になる等、産業分野や医療分野へのインパクトは極めて大きいものがあり、これら一連の動きは、如何にゲノムに賭ける期待が大きいかを物語るものである。これまでのDNAに関する経緯を表1に示す。

わが国でのゲノム科学の研究経緯

今、なぜゲノム科学かを理解するために、わが国のこれまでの研究経緯をみておきたい。

わが国でライフサイエンスの研究が開始されたのは1973年で

図1 ゲノム科学の開発テーマとその応用分野



ゲノム科学総合研究センター・ホームページ、パンフレット資料より作成

ある。科学技術庁が宇宙開発、防災に次ぐ新しい科学技術として「ライフサイエンス」研究を取り上げ、理化学研究所に「ライフサイエンス研究推進センター」を設置し、わが国の叡智を集めて研究計画作業に着手したのが始まりである。約2年間かけて作成された計画に基づき理化学研究所では5つのプロジェクトが正式に開始された。「バイオリアクター」「脳思考過程」「生理活性物質の探索」などである。このうち、「バイオリアクター」などの幾つかの研究は10年間継続され、ライフサイエンス研究の基盤が固まった。ちょうどその頃、米国では、遺伝子組換えによって有用物質を作り出す研究が活発に行われ、多くの画期的な成果が得られていた。1953年、米のワトソンと英のクリックによって遺伝子DNAの二重らせん構造が解明されたことをきっかけに、それまで謎とされていた生物の遺伝子の働きが物理化学的に理解されるようになり、また1970年には人為的な方法でDNAが合成され、遺伝子としての役割を果たせる可能性が示された。以来、日米等の研究グループによって実用化に向けての遺伝子合成が進められ、1973年にコーエン、ボイヤーによって遺伝子組換え技術が確立されて遺伝子工学が成立したとされている。わが国

表1 DNA関連の主な発見・研究(抜粋)

1898	独の解剖学者Waldeyerが細胞の核内の物質が特殊な色素を吸着するので、染色体と命名
1909	デンマークの遺伝学者Johansenが遺伝的形質の伝達に關する物質を遺伝子と呼称することを提唱
1910	米の遺伝学者Morganが遺伝子連関の大きな概念を明らかにする
1944	米のAvery, MacLeod, McCartyが実験によりDNAが遺伝情報の担い手であることを確認
1953	米の生物学者Watson、英の生物物理学者CrickによりDNAが二重らせん構造であることを発見
1967	米のKornberg, Sinsheimer, Goulianが完全な合成DNAを生産
1970	米のKhoranaが完全な遺伝子を合成
1972	米のBergらStanford大学の研究チームが遺伝子DNAを科学的に処理して組換え体を作ること成功 - 世界初の遺伝子組換え
1973	米のCohen, Boyerが遺伝子組換え技術を完成
1976	多くの研究者により、ヒトの22個の常染色体のそれぞれに、少なくとも1個以上の遺伝子に対応付けることに成功
1977	日の板倉が大腸菌にインスリンや成長ホルモンをつくらせることに成功 - 遺伝子工学ブームに先鞭
1979	日の谷口が大腸菌内で、インターフェロンを大量に増やすことに成功これまで350の遺伝子について特定の染色体が決定され、250の特定の遺伝子の座が特定の常染色体に決定
1980	米のCline, Salserがヒトにおける最初の遺伝子療法の実験が地中海貧血の治療で実施
1980	日の谷口がインターフェロンの遺伝子構造を解明
1988	米でヒトゲノム解析が国家プロジェクトとしてスタート
1990	米で世界初の遺伝子治療が実施される
1996	英のロズリン研究所でクローン羊が誕生。世界初のほ乳類のクローン動物
1998	日でクローン牛が誕生

出典: DNA発見小史(<http://tag.ahs.kitasato-u.ac.jp/KITASA-2/dnadisco.htm>)

和田博士の提唱から12年後の1998年、科学技術庁は、改めて理化学研究所にゲノム科学総合研究センター（GSC）を設置し、和田博士を所長に招聘してゲノム解析とタンパク質の機能・構造解析等の研究を行う巨大プロジェクトをスタートさせることとなった。

今日、わが国では、かずさDNA研究所やヘリクス研究所でヒトゲノムの解読と有用遺伝子の探索が進められているが、理化学研究所のGSCは、マウスを対象にした遺伝子構造・機能研究およびタンパク質構造・機能研究、ヒトゲノムを対象としたゲノム構造情報研究、動物や植物ゲノムの機能情報研究、ゲノム情報科学研究の6つのグループで研究が活発に進められている（図2）。また今年から遺伝子多型研究センターが設置され、わが国の遺伝子研究も本格化した。このように先端に行く研究構成と、大型施設を完備した点で、世界でも希なる研究センターとして世界から注目を浴びている。

ゲノム科学の重要性

では、なぜ、ゲノム科学が重要なのかを考えてみたい。

ゲノムとは、生物が生命活動を営む上で必要とされる遺伝情報の全てをいう。一般に人間はおよそ60兆個の細胞できているが、個々の細胞の核の中にはゲノムDNAが存在し、そこには4種の核酸塩基によって30億もの文字が書き込まれている。そのDNAの中には親から子に伝わる遺伝子の情報が含まれる。ゲノムDNAの中の約10万ヶ所からmRNA（伝令RNA）に平均2000文字の情報が転写され、その書き写された文字の3つの文字単位に1対1の対応で20種類のアミノ酸の中から1種類のアミノ酸が選ばれて結合され、文字列に沿って順次、タンパク質が合成されていく（図3）。このようにして作られるタンパク質は10万種にも及び、生物の主要な生理活性を担う重要な役割を果たしている。30億の遺伝文字が詳細に解読されれば、その中にあるアミノ酸配列をコードする遺伝子を特定し、有用な生理活性物質であるタンパク質を見つけることができる。ゲノムDNAの塩基配列の解析（シーケンシング）は、このような生物の巧みな仕組みを解明し、そこに隠された情報から人々にとって有用な物質を見出すための重要なステップなのである。

現在、理化学研究所GSCで進められている「理研ゲノム・エン

図3 遺伝子情報からタンパク質が合成される仕組み

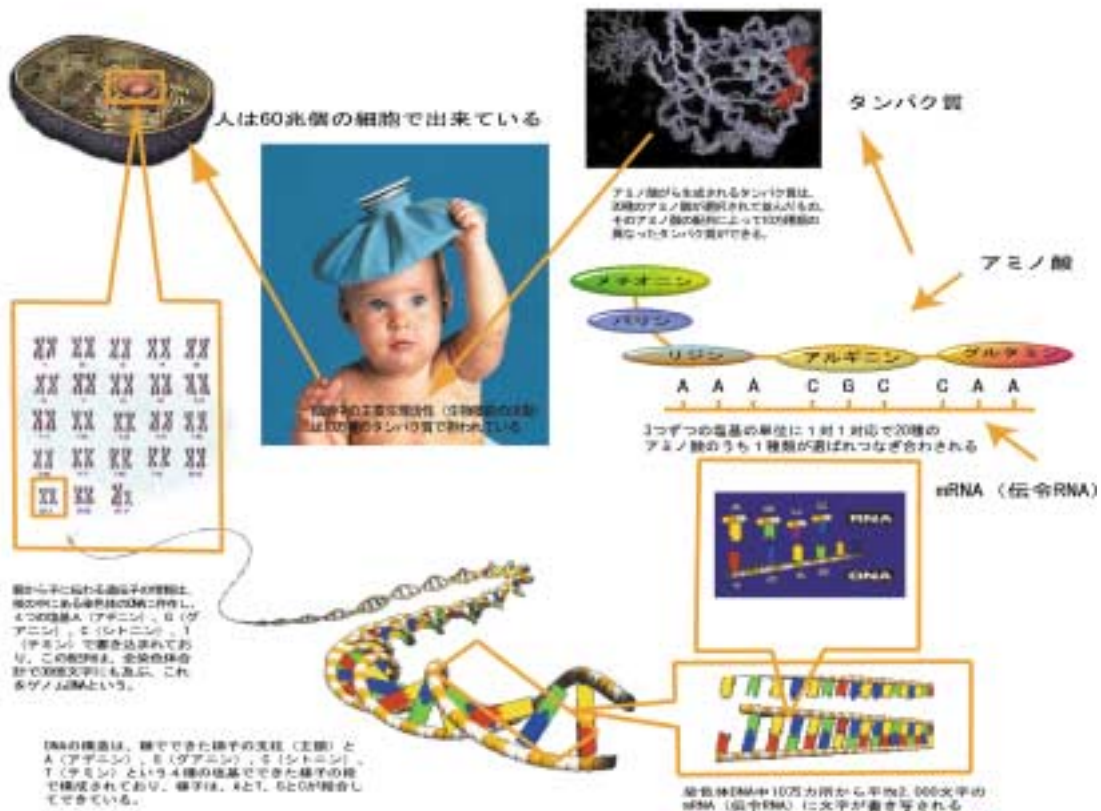
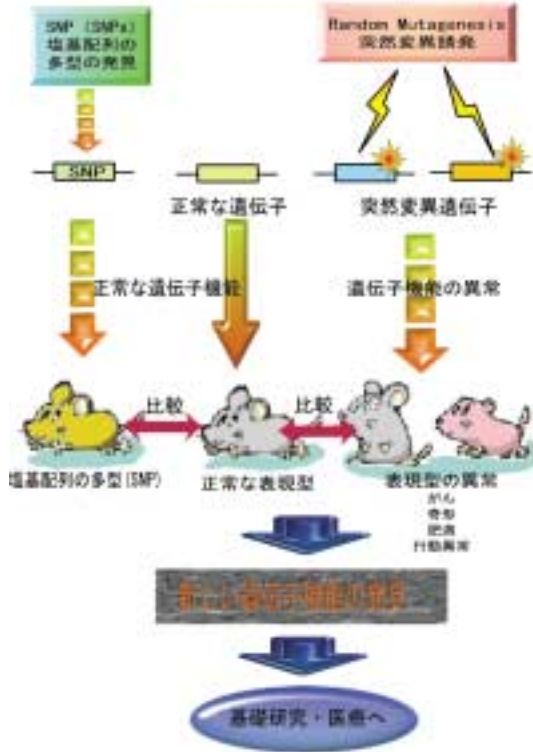


図4 マウスを使用した遺伝子機能の研究



ゲノム科学総合研究センター・ホームページ、パンフレット資料より作成

図4付表 主な生物の塩基対の数・人間との遺伝的重なりの割合

主な生物の塩基対の数

生物	塩基対
ウイルス	10,000
大腸菌	5,000,000
イースト	12,000,000
細菌	100,000,000
ショウジョウバエ	180,000,000
トマト	700,000,000
マウス	3,000,000,000
人間	3,000,000,000

人間との遺伝的重なりの割合

生物	(%)
ウイルス	15
大腸菌	15
イースト	30
細菌	40
マウス	75
牛	90
チンパンジー	98.4
ヒト(他人)	99.9
ヒト(兄弟)	99.95

参照：サイエンス21

「エンサイクロペディア作成計画」は、マウスの全遺伝子の完全長cDNAの塩基配列を決定し、そこから染色体上の位置情報を決定することを目指している。これまで約70,000個のcDNAクローンを収集し、そのうち17,000種類の完全長cDNA配列が決定された。完全長cDNAとは、ゲノムから取り出されたmRNAから逆転写酵素を用いて塩基配列情報を写し取った相補鎖DNAと呼ばれるもので、これを大腸菌の中に入れて増殖させて完全な長さのDNAに復元したものである。これを用いればタンパク質の完全なアミノ酸配列が判る点で極めて重要なものである。理研では、独自に開発した超高速シーケンス装置を用いて完全長cDNAのシーケンス解析が行われ、エンサイクロペディアの作成に一步步近づいている。

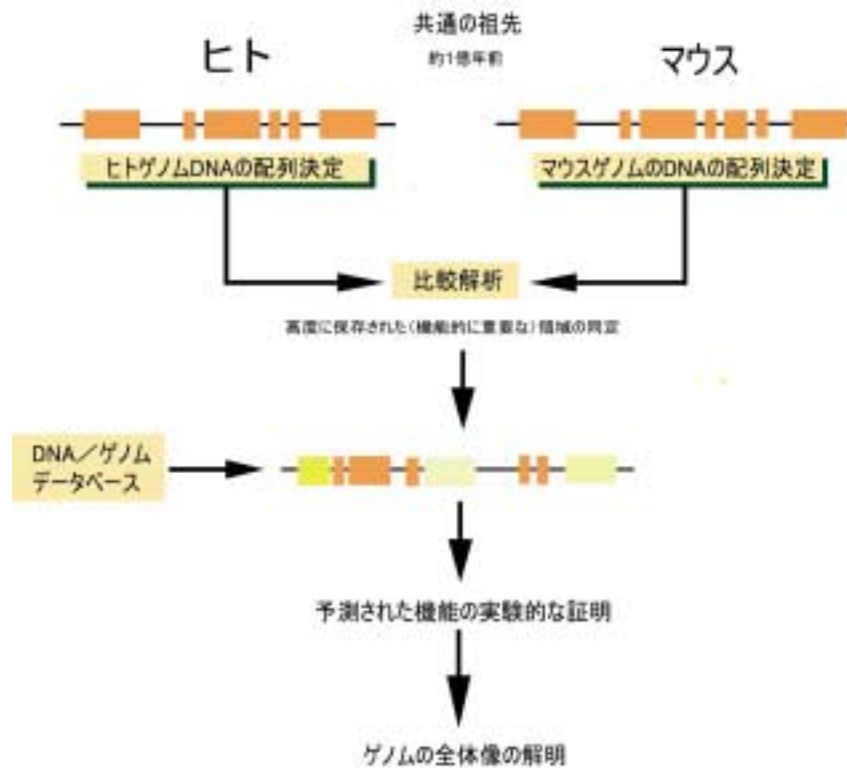
ところで、マウスはヒトと同じ高等動物であり、そのゲノム・サイズや遺伝子の数がほぼ等しく、コード領域も70～80%の相同性をもつことから、そのシーケンス解析の結果を医療や新薬の開発に応用するときに最も扱いやすいモデル動物であることに重要な意味があるのである(図4・図5)。

ゲノム科学の将来

では、ゲノム科学の研究は将来どのような方向に進むであろうか。

一つは完全長cDNAマイクロアレイを用いた遺伝子発現情報の解明である。どの組織で遺伝子が発現しタンパク質を産生するかを明らかにすることによって、創薬のスクリーニングに重要な知見を得ることができる。また、遺伝子間の発現調節のカスケード等を調べることにより、治療などの医療分野に応用が期待される。二つには、完全長cDNAを用いてタンパク質を作成し、その立体構造を解析することである。生体内のタンパク質は、その立体構造によって各種の生理活性をもたらすことが明らかになっているが、その構造解析は今日の大きな技術的課題であり、NMR(核磁気共鳴装置)や大型放射光施設を用いて今後急速に研究を進めることが予定されている。またアミノ酸構造から機能を予測する研究も大量のデータベースとデータ処理システムの進展とともに注目される研究分野である。三つ目は、塩基配列情報の多様性

図5 ゲノムの配列決定からマウスを使用した遺伝子機能の研究



であるSNP(1塩基多型:スニップ、またはスニップスと呼ぶ)の研究である。SNPとは、ヒトなどの個体によって、DNA塩基配列の中の1文字(塩基)が異なった部分があり、それがために個体ごとに異なった生理活性をもたらしていると考えられ、ヒトゲノム30億塩基対の全領域に約300万個のSNPが存在するといわれている。これを探索・開発することは、次世代の個々人に応じたオーダーメイド医療、さらには疾病のリスク予測に基づく予防医学が可能になるなど、医療分野への貢献は計り知れない。また、個人に対応した薬剤投与が可能になり、適切な治療が実現する。米国では、ヒトゲノムの解析が進むにつれ、国を挙げてのSNP研究への取り組みが見られる。わが国でも、昨年6月に科学技術庁、文部省、厚生省、通産省の4省庁が共同でSNP戦略を取り纏め、今年度から国家プロジェクトとして予算化され産官学が一体となって研究がスタートしている。SNP解析で得られたデータ内容やマッピング情報、解析条件、産業に役立つ生物情報などをデータベース化し、その利用システムを開発することで、多くの疾患に関連する遺伝子の探索を行うことができ、さらに疾患に関する遺伝子産物(タンパク質)を特定し、その構造に関する情報を基に創薬研究へと

発展させることが考えられている。また、高度医療機関においてはSNP情報を利用した創薬の臨床への適用や実患者を対象にした投薬効果の事前予測など、医療面においても高度な研究が可能になる。

以上に述べたゲノム科学研究の将来像から明らかなように、当研究は多方面での効用が期待される。直接、間接に最も大きなインパクトが生じるのは医療分野であり、糖尿病、高血圧・動脈硬化症等の多因子遺伝疾患の解明とそれに基づく治療、成人病やガンの発生予測、体質や遺伝的背景の診断、遺伝子治療、遺伝子診断が代表的なものである。また、ゲノム創薬といわれる新しい医薬品のデザインと開発などは、新たな産業の萌芽になる可能性を秘めている(図6)。

ゲノム科学におけるITの重要性

最後に、ゲノム科学とITについて触れておきたい。

ゲノム科学がここまで進歩した裏には、コンピュータ等のITの貢献を見逃すことはできない。膨大なゲノムDNAの情報を解読

する作業や、判明した遺伝情報からタンパク質に関連した領域を特定し、さらに遺伝子間、タンパク質間の関連性の解析や構造と機能の解析を推進する原動力として、ITによる各種の解析システムやソフトウェアの存在は極めて重要であり、今やこれなくしてゲノム科学は成り立たないといっても過言でない。

“バイオインフォマティクス”といわれる総合化された解析技術は、多様なデータベース技術と共に、益々重要性が増している。これまで生物の機能面だけから探求されてきたライフサイエンスもゲノム科学の登場によってより一層ITの支援が不可欠となっており、バイオインフォマティクスの研究もさらなる高度化と発展が求められている。

我々は、昨年からは理化学研究所ゲノム科学総合研究センターの「ゲノム・エンサイクロペディア」作成プロジェクトに従事し、マウスの完全長cDNAの塩基配列データベースの開発と研究者の望む各種の解析ツールの提供、さらに、これらのデータベースとネットワークのセキュリティ・システムの開発を行うなど、バイオインフォマティクスのインフラの整備に尽力してきた。今後ますます複雑・多様化する研究において、パイオ

インフォマティクスの重要性を強く感じ、常に最新のITの提供を行っていく所存である。



技術開発部バイオインフォマティクスセンター



担当部長
香月 祥太郎



担当部長
村田 賢太郎

図6 大規模cDNA解析システムと遺伝子エンサイクロペディアの応用



ゲノム科学総合研究センター・ホームページ、パンフレット資料より作成