

2. 特集：ライフサイエンス・医療分野の注目動向

ヒトゲノム解読を巡る国際解析チームとセセラ社の動向及びわが国の今後の動き

ライフサイエンス・医療ユニット 茂木伸一、庄司真理子
 蛸原弘子、長谷川明宏

2.1 はじめに

ヒトゲノムとは、ヒトの遺伝情報の全体像、つまり生命の設計図とも言える。通常、ヒトゲノムとは、細胞の核にある22種類の常染色体と2種類の性染色体に含まれるDNAを指している。DNAは4種類の塩基(アデニン、グアニン、シトシン、チミン)から構成される物質で、ヒトゲノムは約30億個の塩基対から成る。この塩基配列のうち遺伝子(遺伝情報をコードするもの)に相当するのはたった3%と言われている。

ヒトゲノムの全配列解析は1980年代半ばに提唱されたが、当時の解析技術で全配列を解析するには膨大な時間を要することなどから、この研究の実現性は乏しいと考えられていた。しかし、徐々にヒトゲノムに関する動きが活発となり、1990年にはヒトゲノムの全配列解析を行う国際ヒトゲノム解析計画が米国主導により本格的に開始された。当初、この計画は15年間で進行予定であったが、ゲノム解析装置の能力向上やベンチャー企業の参入等により前倒しされていった。そして2001年2月に、国際ゲノム解析チーム及び米国ベンチャー企業のセセラジェノミクス社から、それぞれヒトゲノム配列解析結果が報告された。

解析されたゲノム配列を利用した研究をポストゲノム研究と言う。ヒトゲノムに関するポストゲノム研究は、新薬の開発や新たな診断・治療法の確立など、産業に直接結びつく要素が高く、民間企業も交えた激しい世界的競争となっている。

本稿では、ヒトゲノム配列解析に関する国際的な動向、注目すべきベンチャー企業のセセラジェノミクス社の動き、わが国のポストゲノム研究の動向について概要を報告する。

2.2 ヒトゲノム配列解析に関する国際的動向

2001年2月、国際ゲノム解析チームはネイチャーに、セセラジェノミクス社はサイエンスに、それぞれ

ヒトゲノム配列解析の論文を発表した(図表1)。

図表 1 ヒトゲノム配列解析組織の紹介

学術誌名 題名	ネイチャー 2001年2月15日号 「ヒトゲノム配列の最初の解析」	サイエンス 2001年2月16日号 「ヒトゲノムの塩基配列」
発表者	国際ゲノム解析チーム ・米国(ワシントン大学、エネルギー省合同ゲノム研究所、ベイラー医科大学、マサチューセッツ工科大学ホワイトヘッド研究所など) ・英国(サンガーセンターなど) ・仏(ジェノスコプなど) ・日本(理化学研究所、慶應義塾大学など) ・独など	セセラジェノミクス社 (米国ベンチャー企業)

2.2.1 国際ヒトゲノム解析チームの配列解析

国際ゲノム解析チームの各研究機関が、公的配列データベース GenBank に登録したヒトゲノム配列のうち完成度の高い配列(Finished human sequence)の配列量を図表2に示した。

一本のヒト染色体全体の塩基配列を解読した世界初の成果として、1999年12月に日英米の共同研究チーム(慶應義塾大学、英国サンガーセンター、米国オクラホマ大学とワシントン大学)が第22番染色体の配列解読を終了した。この22番染色体、及び慶應義塾大学や理化学研究所などが分担して解析した第21番染色体(2000年5月解読終了)の配列の解析精度は99.99%と非常に高く、世界的にも評価されている。第22番染色体にはパーキンソン病の原因遺伝子や自己免疫疾患に関係する遺伝子等があり、第21番染色体にはダウン症に関係する遺伝子やアルツハイマー病に関係する遺伝子等がある。これらの疾患関連遺伝子を含むゲノム配列を決定したことは、その発現機序研究などの進展に大きく貢献すると期待されている。

しかし、その他の染色体の中には、解析困難でドラフト配列と呼ばれる領域がまだ残っている。国際ゲノム解析チームの配列解析は現在も進行しており、2003年に全配列を高い精度で決定することを目標としている。

図表 2 ヒトゲノム配列解析組織が GenBank に登録した完成度の高い配列

順番	研究機関	配列量 (キロ塩基)
1	The Sanger Centre (英)	284,353
2	Washington University Genome Sequencing Center (米)	175,279
3	US DOE Joint Genome Institute (米)	78,486
4	Baylor College of Medicine Human Genome Sequencing Center (米)	53,418
5	Genoscope (仏)	48,808
6	Whitehead Institute, Center for Genome Research (米)	46,560
7	Department of Genome Analysis, Institute of Molecular Biotechnology (独)	17,788
8	理化学研究所 (日本)	16,971
9	University of Washington Genome Center (米)	14,692
10	慶應義塾大学 (日本)	13,058
	その他	92,614
	合計	842,027
	(ヒトゲノム全体)	約 3,200,000

ネイチャー2001年2月15日号 868 ページ 表3の数値 (Finished human sequence)をもとに科学技術動向研究センターで作成

2.2.2 セラジェノミクス社のヒトゲノム配列解析

セラジェノミクス社は染色体ごとに配列を決定するのではなく、24種類全ての染色体を同時に解析する全ゲノムショットガン方式を採用した。同社の全ゲノムショットガン方式では、まずヒトゲノムを超音波で物理的に断片化して、大きさがそろった断片(2キロまたは10キロ塩基)を含むライブラリーを構築した。そして各断片の両端にある数百塩基対の配列を決定し、その配列をつなぎ合わせることで、ゲノムの塩基配列を決定した。同社によると99.96%という高い精度でゲノムの95%の領域をカバーするものである。

2.2.3 ヒトゲノム配列へのアクセス

国際ゲノム解析チームのヒトゲノム配列情報へ

のアクセスは無償であるが、セラジェノミクス社のヒトゲノム配列情報を利用するためには同社との契約が必要である。

(1) 国際ゲノム解析チームのヒトゲノム配列

国際ゲノム解析チームでは、各研究機関がヒト染色体の領域を分担して解析し、得られた配列情報を即時に GenBank などの公的データベースに登録し無償で公開している。これらのヒトゲノム配列情報には、第21番染色体や第22番染色体などの精度の高い配列とその他のドラフト配列の領域が含まれている。高精度の全配列情報を利用するには、2.2.1 で記述したとおり国際ゲノム解析チームの今後の解析を待つ必要がある。

(2) セラジェノミクス社のヒトゲノム配列

研究者・研究機関・企業などは、セラジェノミクス社と契約を結ばない限り、同社の配列情報を自由に利用することはできない。例えば、オーストラリアでは国レベルでセラジェノミクス社と契約を結び、国の研究機関が同社のヒトゲノム配列を利用している。わが国においては、大学・民間企業の中に個々にセラジェノミクス社と契約を結んで同社の配列情報を利用しているところがある。

2.3 セラジェノミクス社のゲノム配列解析実績と今後の戦略

2001年3月にセラジェノミクス社のベンター社長が来日し、講演会等で同社のゲノム配列解析と今後の戦略について次のように述べた。

(1) ゲノム配列解析の実績

1998年に設立されたセラジェノミクス社は、これまでにショウジョウバエ、ヒト、マウス(129SvJ, DBA/2, A/Jの3系統)のゲノム配列を決定した。現在はイヌとラットのゲノム配列を解析している。短期間に数多くの成果があがった背景には、①新しい自動シーケンス解析装置(ABI3700)を300台導入したこと、②全ゲノムショットガン配列解析ソフトを開発したこと、③最新のコンピュータを導入したことがあげられる。すなわち、ハードとソフトの両面で時代の流れに乗ることができた。

(2) ヒトゲノム配列解析状況

ヒトゲノムには進化の過程が記録されている。ヒトゲノム配列を調べると、染色体間で配列が類似

した領域が数多くあり、これらは共通の起源から進化の過程を経て生じたものと考えられる。

また、ゲノム配列は個人ごとに異なる部分もあるが、別人のゲノム配列を比較することにより、これまでに約300万箇所の一塩基多型(SNPs: single nucleotide polymorphism、DNAの一つの塩基が他の塩基に置き換わっているもの)を見出した。最終的に400万箇所くらいになるであろう。これらのうち、タンパク質をコードしている領域にあるものは1%以下であり、アミノ酸が変異するものはその数分の一である。疾患との関連ということでは、非コード領域の個体差が病気に対する抵抗力や医薬品の効き方などに影響することも知られているので、タンパク質コード領域以外の部分も重要である。

(3) ポストゲノム戦略

セララジェノミクス社が設立したタンパク質解析施設(プロテオミクス・ファクトリー)では、質量分析装置を用いて1日に100万検体のタンパク質の配列構造を解析することができる。同社は、ヒトの約3万の遺伝子から転写・翻訳・翻訳後修飾を経て約25万種類のタンパク質が生じると考えており、タンパク質の異常と疾患とを結びつけることができるものと期待している。対象疾患としてがんを考慮しており、がんの診断をすること及びがん特異的なワクチンを作ることを目標としている。

なお報道によれば、同社はSNPs解析技術を持つ日本のバイオベンチャー企業に資本参加し、日本人のSNPsと疾患との関連などの解析を目指している。

2.4 わが国のポストゲノム研究に関する動向

ヒトゲノムにおけるポストゲノム研究は、図表3に示したように、医療への貢献、生命科学の進歩、諸産業への貢献など、広い範囲に波及するものである。

こうしたポストゲノム研究の重要性については、2001年3月に閣議決定された第2期科学技術基本計画においても明確に位置付けられており、同計画には、わが国において研究開発の重点分野と位置付けられているライフサイエンス分野の中でも、ポストゲノム研究は特に重点的・戦略的に取り組むべきものであるとされている。

一方、わが国のポストゲノム研究に関する具体

的方向を示すものとして、2000年12月には、「ポストゲノム戦略の推進について」(科学技術会議政策委員会ポストゲノムの戦略的推進に関する懇談会の報告)がまとめられている。この報告では、わが国において緊急の取り組みを要するポストゲノム研究の分野は、ヒトゲノム多様性・疾患遺伝子解析、タンパク質構造機能解析、バイオインフォマティクス、ゲノム機能解析であるとしている。また同報告は、このようなポストゲノム研究を推進する上では、わが国が世界に先行している「技術」、例えば、タンパク質のアミノ酸配列をコードするmRNAの全配列情報を含んでいる完全長cDNAの作製技術(東京大学医科学研究所菅野氏らによる方法及び理化学研究所林崎氏らによる方法)や、無細胞タンパク質発現技術(理化学研究所横山氏らによる多数のタンパク質試料の効率的な発現・調製技術)、さらには先行している「施設」として、大規模NMRパーク(理化学研究所ゲノム科学総合研究センター)、大型放射光施設SPring-8(財団法人高輝度光科学研究センター)等を活用して、緊急な取り組みを行うことにより、諸外国の先手を打つ対応も可能であると述べている。

平成13年度予算においては、「生命科学の21世紀に向けたバイオ施策」として、文部科学省、厚生労働省、農林水産省及び経済産業省により983億円を計上しているうち、ヒトゲノム解析を突破口とした5大疾患の克服を目指した①ヒトゲノム構造・多様性解析(ヒト完全長cDNA構造・機能解析、SNPs研究など)及び②ヒトゲノム機能解析(タンパク質構造・機能解析、バイオインフォマティクスなど)の研究には総額606億円(前年度444億円)が充てられており、当該研究開発を一層推進することとしている。

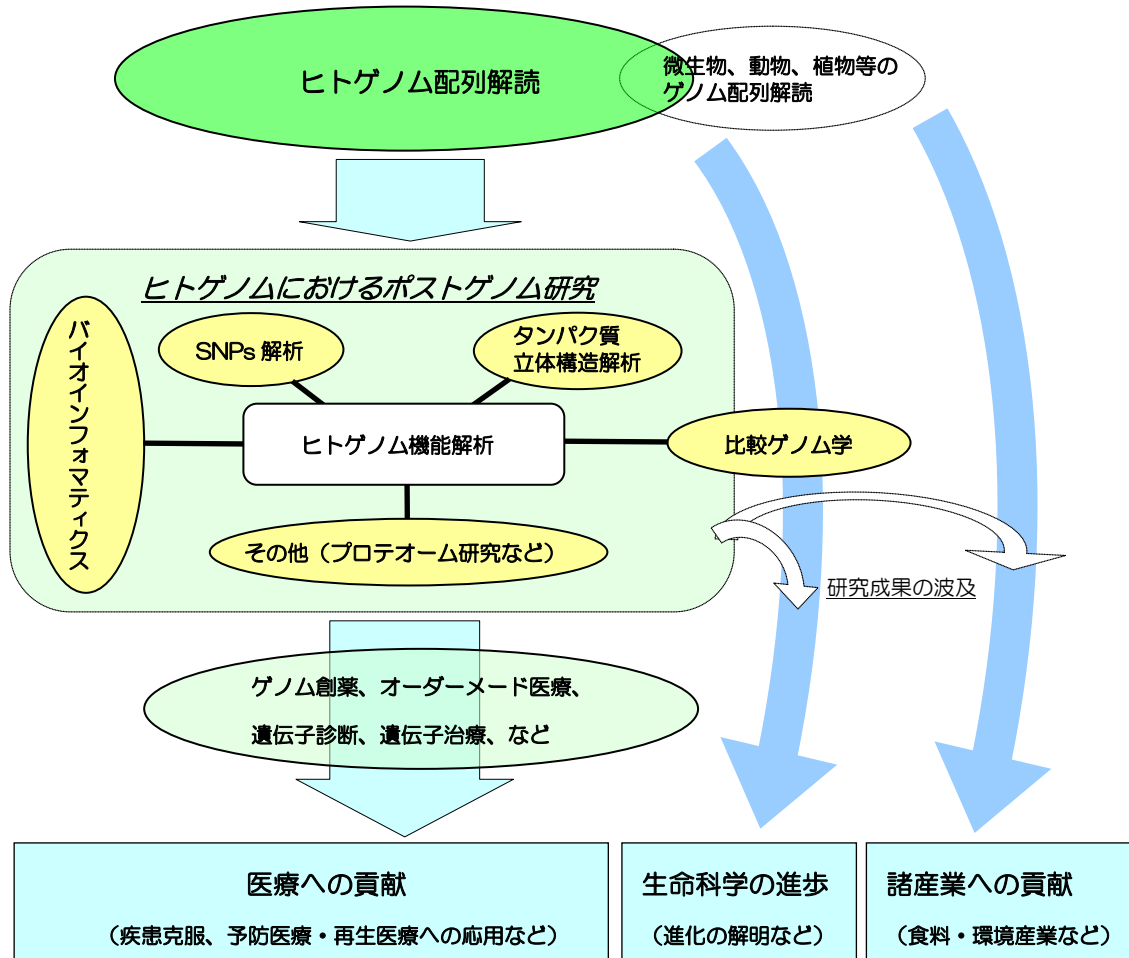
ここでは、タンパク質立体構造の研究、日本人のSNPsの研究及びバイオインフォマティクスの動向について述べる。

(1) タンパク質立体構造の研究

ゲノムから転写翻訳されて生じるタンパク質の機能を推定する普遍的な方法はまだ見出されていない。しかし、機能に密接に関連するタンパク質の立体構造を解明するプロジェクトが理化学研究所などにおいて進められている。

これまでに解明されたタンパク質の立体構造は2,000~3,000種類で、そのほとんどが米国で実施された。

図表 3 ヒトゲノムにおけるポストゲノム研究



科学技術動向研究センター作成

日米欧の国際研究チームは今後5～10年間で10,000種類以上の構造を解明することを計画しており、わが国は理化学研究所ゲノム科学総合研究センターを中心として今後5年間で約3,000種類以上、米国は今後5年間で約5,000種類程度の解明を目標としている。国際ゲノム解析チームにおけるヒトゲノム配列解析では、全体の数パーセントであったわが国の貢献度が、タンパク質の立体構造の解析においては、より大きくなるものと期待される。

2001年4月には、日・米・英が主催した国際構造ゲノム科学会議が米国で開かれ、10カ国の代表者によりタンパク質解析に関する国際協力の枠組みが議論された。ここでは、公開原則のもとに共通の手順で研究の進み具合を報告し合うことが決められた。また立体構造データの公開時期については、このデータが病気の診断や新薬の開発に直結しやすいことから、各国の特許制度の違

いにより不公平が生じないように考慮され、解明時点から最大6カ月まで余裕を与えることが合意された。

(2) 日本人の SNPs の研究

ヒトゲノム配列を個人ごとに比較すると多くの部位で異なっている。この違いを遺伝子多型と呼び、①SNPs、②挿入・欠失型多型(部分的に塩基の挿入や欠失があるもの)、③VNTR(variable number of tandem repeat)多型・マイクロサテライト多型(2から数十の塩基配列が繰り返している部位の繰り返し回数が個人ごとに異なるもの。繰り返しの単位が数塩基から数十塩基のものをVNTR多型、2から4塩基程度のものをマイクロサテライト多型と呼ぶ)などがある。

ヒトゲノム中にVNTR多型とマイクロサテライト多型がそれぞれ数千箇所と数万箇所あるのに対して、SNPsは数が多く、400万箇所程度(セララジェ

ノミクス社のベンター社長)や300万~1,000万箇所(中村祐輔著「先端のゲノム医学を知る」羊土社2000年)あると言われている。またSNPsは高速・大量の検出技術が実用化されつつあることから、疾患関連遺伝子を探索するために利用しやすい。このため、多型の中でも特にSNPsの研究に重点がおかれている。

こうした中、東京大学医科学研究所で行われた一部のSNPsの解析結果から、日本人集団は比較的孤立して存在していた集団であることが示唆され、日本人に特有の疾患関連SNPsが存在する可能性が出てきている。また、東京大学医科学研究所では、SNPsの情報をデータベース化してインターネット上で一般に公開しているほか、ファルマスニップコンソーシアム(理化学研究所、東京女子医科大学及び製薬企業43社からなる)では、日本人一般集団(健康な成人ボランティア)約1,000人のDNAを収集し、165種類の薬物代謝関連遺伝子多型に関する解析を行うなど、将来のオーダーメイド医療につながる研究を進めている。

ヒトゲノム配列のうち個人ごとに異なる部分の情報、オーダーメイド医療すなわち遺伝的体質などを正確にとらえて有効な治療法を提供することへの手がかりになるものと期待されている。その一方、個人の遺伝情報をどのように取り扱うのか問題がある。2001年3月に文部科学省、厚生労働省及び経済産業省は、ヒトゲノム・遺伝子解析研究全般に関する具体的な指針となる「ヒトゲノム・遺伝子解析研究に関する倫理指針」を共同でまとめた。同指針では、試料提供者に事前に研究の趣旨などを十分に説明し文書で合意を得ること(インフォームド・コンセント)や個人情報の保護など、研究者及び研究機関の遵守すべき責務等が提示された。また、日本人類遺伝学会など遺伝医学関連8学会は、ヒトの遺伝子検査のあり方について共同のガイドライン案をまとめた。

(3) バイオインフォマティクス

ヒトゲノム解析の急速な進展に伴い、膨大な配列情報を取り扱うことが求められている。塩基配列情報のみならず、アミノ酸配列から推定されるタンパク質の構造や機能、代謝経路の推測等を可能とするソフトウェアを開発することが今後のポストゲノム研究の発展に必須である。そのための基盤となるデータベースを構築して、各種データを相互に関連づける役割を担う、バイオインフォマティクスの研究の更なる発展が急務である。特に生物

学と情報科学の両方に通じている専門家の育成については、研究者の間では以前から重要とされていた課題である。

わが国においても政府の具体的施策として、平成13年度科学技術振興調整費のうちの人材養成に関する公募課題(全2分野)の1分野として、バイオインフォマティクスが取り上げられた。

また、最近の注目すべき動きとして、2001年3月下旬から4月上旬にかけて、生命科学に必要な情報技術を修得し、新領域を拓く人材の育成を目指す「情報生物学適塾」(塾長・松原謙一奈良先端科学技術大学院大学教授、後援 文部科学省他)が国際高等研究所で開かれた。このように研究者の間でもこの領域の重要性が認識されている。

米国では、専門分野にかかわらず生物学の履修を必修とする大学もあり、今後のバイオインフォマティクスの進展のための施策を検討する上での参考となろう。

25 おわりに ヒトゲノム配列をめぐる今後の動き

(1) 国際ゲノム解析チームによるヒトゲノム配列解析は進行中

現在、ヒトゲノム配列解読は、国際ゲノム解析チームのFinishingグループ(米・英・日・仏)が2003年までに全配列を高精度で決定することを目標に進行している。このFinishingグループは、担当する染色体の解読に責任をもつCoordinating Centerと、できる範囲で協力するParticipating Centerとから成っている。

わが国では、Coordinating Centerとして理化学研究所が米国マサチューセッツ工科大学ホワイトヘッド研究所と共同で第11番染色体及び第18番染色体の解読を進めている。Participating Centerとして慶應義塾大学が第8番染色体の解読に携わっている。

(2) ゲノム研究を疾患克服に活かすために

ゲノム研究を疾患克服に活かすためには、SNPsなどのゲノム配列情報と臨床データを結びつけて解析することが必要である。例えば、アソシエーション(関連)法といって、数百・数千人単位の患者集団と正常集団との比較によって病因関連遺伝子を突き止める手法がある。このような研究を進めるには、より多くのボランティアによる血液試料の提供が必要となる。

わが国においては、まだこのような手法による研究の環境整備が後れている。個々の研究機関レベルでこのような試料を揃えることは困難であり、研究者によっては外国からこのような試料を入手しているところもある。

今後わが国において、国レベルでインフォームド・コンセントなど所定の手続きを経た多数の試料と臨床データを管理するリソースセンターを設置することが、ポストゲノム研究に必須な研究基盤の整備として急務である。

(3) ポストゲノム研究の推進

昨今のヒトゲノム研究では、米国が世界的な主導を握っている状況である。その背景には、積極的なベンチャー企業の設立や特許戦略などがある。例えば、2000年のバイオベンチャー企業は、わが国が約150社であるのに対し、米国では約1,500社であるとされている。(大石正道著「ヒトゲノムのしくみ」日本実業出版社 2001年)

わが国においても、2.4で述べたような政府主導型の様々なプロジェクトが立ち上がっている。また、製薬会社などの民間企業でも、ゲノム創薬をターゲットとする投資拡大の推進、バイオインフォマティクス分野の強化や技術協力を目的とする業務提携などが急速に進んでいる。理化学研究所などの公的研究機関と民間企業との共同研究も進められており、公的資金で得られた研究結果が産業界に活用されることも期待される。

ポストゲノム研究は、これからの国民生活・経済社会に大きなインパクトをもたらすとともに、倫理面をはじめとして社会と深く関わるものである。その推進に当たっては、研究を取り巻く社会情勢や国内外の研究環境の変化に対応して、機動的かつ柔軟に研究資金及び人材を投入し、研究を加速していくことが求められている。

解説

ヒトゲノム配列解読への 日本の貢献度

報道等においては「国際ゲノム解析チームにおける日本の貢献度は“6%”」という数値がよく用いられている。これは、概要解読が終了したと世界的に発表が行われた2000年6月25日現在で、概要解読を行った全データのうち、理化学研究所(188,056 キロ塩基)、慶應義塾大学(20,105 キロ塩基)、東海大学、癌研究会(2機関計 11,783 キロ塩基)の解析データ合計の全データ量(3,934,884 キロ塩基)に占める割合(5.59%)に基づいている。

ネイチャーに論文発表があった2001年2月現在では、概要解読の全データ量はヒトゲノムの約1.4倍にあたる4,338,224 キロ塩基に大幅に増加している。この背景として、外国の研究機関において、SNPs探索のために分担以外の領域について解読を進め、これらが登録されていることがあげられる。

なお、図表2に示した数値は、概要ではなく完成データを示している。