

総 説

ヒトゲノム解析と社会

中 堀 豊

徳島大学医学部公衆衛生学教室

The perspective of human genome project

Yutaka Nakahori

Department of Public Health, School of Medicine, The University of Tokushima, Tokushima

はじめに

ヒトゲノム計画 (Human Genome Project) はヒトのもつ約30億塩基の配列を、全て読んでしまおうという計画である。1991年から米国を中心に本格的に実施され、ヨーロッパ、日本なども加わって国際協調の下に計画が進められている。当初2010年頃完了見込みとされたものが、技術的な革新や、モデル生物での経験の蓄積から、今では2003年頃にヒトの持つ全ての配列を読み終わるであろうとされている。現在、実際に塩基配列が大量決定されつつあるが、それに先だって染色体の遺伝子地図作りや酵母人工染色体を利用したコンテイング作りが行われてきた。ゲノム計画に至るには近年急速に発展した分子生物学の知見があり、またさまざまな新しい技術の開発がある。本稿ではゲノムを理解するためのヒトの細胞遺伝学、分子生物学的な基礎知識を中心に概説し、ゲノム計画の現状、またゲノム計画が社会に及ぼすであろう影響について言及した。

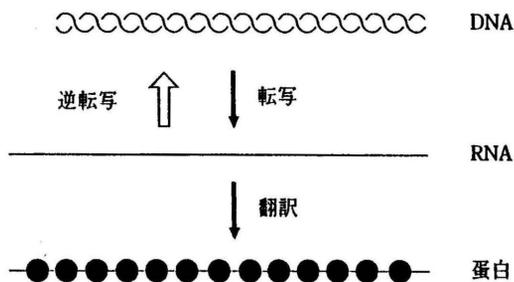
1 遺伝情報の担い手 DNA と RNA

生物はそれぞれ固有の遺伝的プログラムにしたがって形質発現を行い、また自己のコピーを作っている。そのための遺伝情報は核酸 (DNA または RNA) の塩基配列としてそれぞれの生物が固有にもっている。

生物の遺伝情報伝達に使われている DNA はアデニン (A)、チミン (T)、グアニン (G)、シトシン (C) の4種類のみで、これらの核酸の糖の部分の3'炭素と5'炭素がリン酸基を介して1本の鎖としてつながっている。この4つの塩基の並びの違いがそれぞれの生物の遺伝情報の違いとなっている。DNA 分子は物理化学的には1

本鎖であり、塩基どうしが水素結合することによって相補的な2本鎖として安定に存在する。2本鎖 DNA はそれぞれの鎖を鋳型として自己を複製することができるという点で遺伝情報の担い手として理想的なものである。DNA の鎖はそれがつながっている限りは1つの巨大な分子である。したがって、DNA の大きさを表すのには分子量などではなく、塩基の数をを用いて〇〇塩基対 (bp: base pair) または〇〇塩基という表現が使われる。千塩基を 1 Kb (キロベース)、百万塩基を 1 Mb (メガベース) という。また、DNA の鎖の方向性を炭素の番号で5' (ファイブプライムまたはゴダッシュ)、3' というように表す。

図1 遺伝情報の流れ



一般に遺伝情報は DNA から RNA、蛋白へと一方通行であるが、逆転写酵素による RNA から DNA という経路がある。

ヒトをはじめ、ほとんどの生物が DNA を遺伝情報の基本としており、一部のウイルスのみが RNA を遺伝子としてもつ。RNA のまま遺伝子を発現・複製していくウイルスもあるが、レトロウイルスのように DNA 生物に寄生し RNA を一旦 DNA に複製 (逆転写) してから再び固有の RNA を作るものもある。このようなウイルスは通常、逆転写酵素 (Reverse transcriptase) を持ち、

これを使って自身の塩基配列を DNA にうつし変えた後、宿主の複製・発現系を使って自己を再生産する。

このような遺伝情報の流れは図1のようにまとめられる。DNA の遺伝情報は、RNA に転写 (transcription) され、蛋白に翻訳 (translation) されることによって実行に移される。遺伝子の中でも rRNA (リボソーム RNA)、tRNA (トランスファー RNA) やスプライシングに関与する RNA のように蛋白に翻訳されることなく RNA のままで働くものもあるが、多くの遺伝子は mRNA (メッセンジャー RNA) の情報に基づき、蛋白が生成されて初めてその働きを得る。

蛋白はアミノ酸がペプチド結合でつながったものである。生物界に存在するアミノ酸は多数あるが、遺伝子で規定されているアミノ酸は20種類である。RNA の3つの塩基の並び (コドン) に応じて、1つのアミノ酸が規定されている。これは、それぞれのコドンに対応する tRNA が存在し、アミノ酸をつないでいくためである。

蛋白は生体の中でさまざまな働きをしている。例えば、細胞膜の構成成分や細胞を支える線維など構造物質として働くもの、代謝酵素として働くもの、物質の輸送に携わるもの、細胞相互または細胞内のシグナル伝達に関わるものなどである。このような蛋白の働きを通して形質が発現されていく。

真核生物では DNA は核 (染色体)、ミトコンドリア、葉緑体に存在する。植物細胞はこの三者をもつが動物細胞は葉緑体をもたない。ミトコンドリア、葉緑体の起源

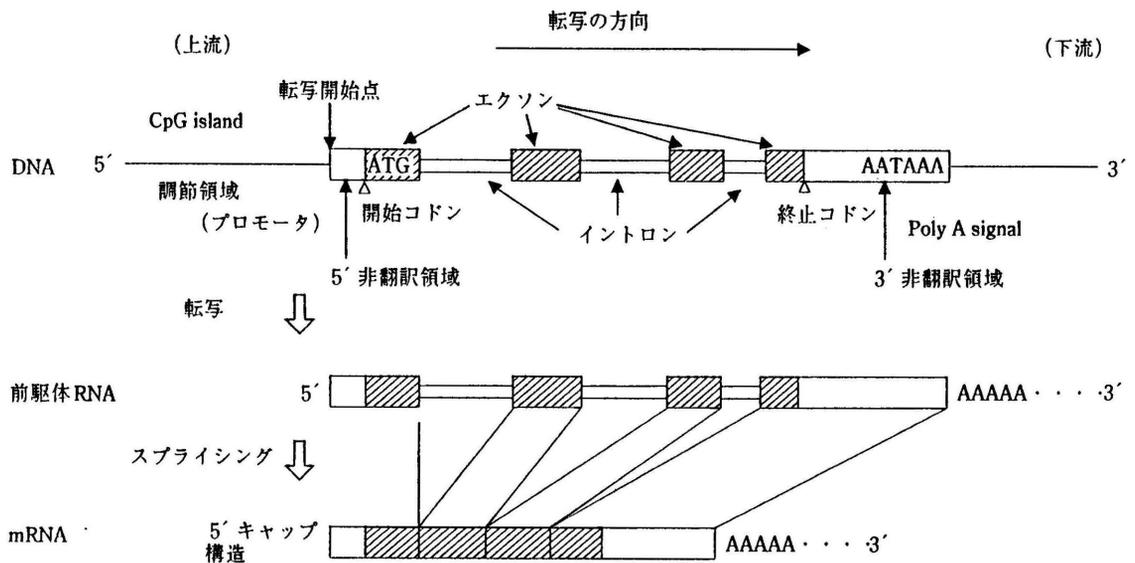
は大型の細胞に共生した特殊な能力 (ミトコンドリアは酸素呼吸、葉緑体は光合成) をもった細胞であると考えられており、これらの複製は染色体上の DNA とは別の支配を受けている。ミトコンドリアや葉緑体上の遺伝子の転写、翻訳は染色体上の遺伝子の産物を使用するか、類似の系を使っているが、染色体の遺伝子とは異なった独自の tRNA を持つため違ったコドンを使用するものがある。ヒトのミトコンドリアは約1万7千塩基対で環状であり、個体により多少の長さの違いがある。ミトコンドリアの DNA は染色体の DNA に比べてはるかに少量であるが、ミトコンドリア DNA の異常による筋疾患や糖尿病も知られている。

2 遺伝子の構造

遺伝子という言葉は一般にさまざまな意味で使われているが、正しくは RNA や蛋白になって働く情報をもった DNA を指す。このような DNA は生物のもつ DNA のうちの一部に過ぎない。遺伝子となって働く部分以外の DNA も含めた生物のもつ DNA 全体をゲノムという。

一般的な遺伝子の構造を図2に示す。一つの遺伝子の遺伝情報は、ヒトでは一続きで存在することの方が稀で、普通は DNA 鎖の上にとびとびに格納されている。遺伝子の発現に際してはまず、DNA が RNA にうつし取られる。RNA 合成酵素は核酸の3'に次の核酸の5'をつなぐという方向にしか働けないので、核酸の鎖を作るときには鎖は5'から3'方向に作られる。遺伝子の5'上流部に

図2 遺伝子の構造



一般的な遺伝子の構造を示した。DNA は転写、スプライシングを経て mRNA となる。

転写を調節する領域があり、さまざまな転写調節因子が直接または間接に DNA に結合し、RNA 合成の量や速度を調節している。正に調節するものをプロモータ、負に調節するものをリプレッサーという場合や、調節領域をまとめてプロモータ領域という場合がある。

遺伝子の調節領域、特にさまざまな臓器で普遍的に発現している遺伝子 (house keeping gene) の場合、にはしばしば CpG island と呼ばれる C と G に富んだ領域がある。ヒトは C のメチル化を通して遺伝子や染色体の活性を負に調節していると考えられているが、メチル化は CG (塩基の並びの意味を出すためリン酸を入れて CpG と書く) と並んだ C に起きる。メチルシトシンはアミノ基を失うことによって T に変化するので、塩基が変わるような突然変異は C が T に変わる方向に多く起こったと考えられている。哺乳類では G や C が A や T に比べて少なく、全 DNA 量の 40% 程度である。常に転写されているような遺伝子の調節領域は、進化の長い時間の中で転写因子が結合し変化を免れてきたと考えられており、このような領域が CpG を多く残して、現在見られる CpG island となったと思われる。遺伝子の 5' 上流の調節領域は転写に際して最も重要な調節をおこなう領域であるが、転写の調節にはこれ以外にも同じ DNA 分子上に存在するさまざまな因子が関与している可能性がある。

RNA 合成酵素は DNA の情報をそのまま RNA にコピーする。この RNA は遺伝子として必要な部分 (エクソン) と介在部分 (イントロン) を含んでいるが、必要な部分だけがつなぎ換えられ (スプライシング)、その他の修飾が加えられて成熟した RNA となる。スプライシングの始まる場所 (ドナー部位)、終わる場所 (アクセプター部位) には共通した配列があり、基本的に全ての遺伝子において同様のメカニズムを使ってスプライシングが行われている。一つの RNA のスプライシングに際して、常に一種類の mRNA ができるのではなく、ドナー部位とアクセプター部位の使い分けによっていろいろな機能分子を作り出しているものがあり

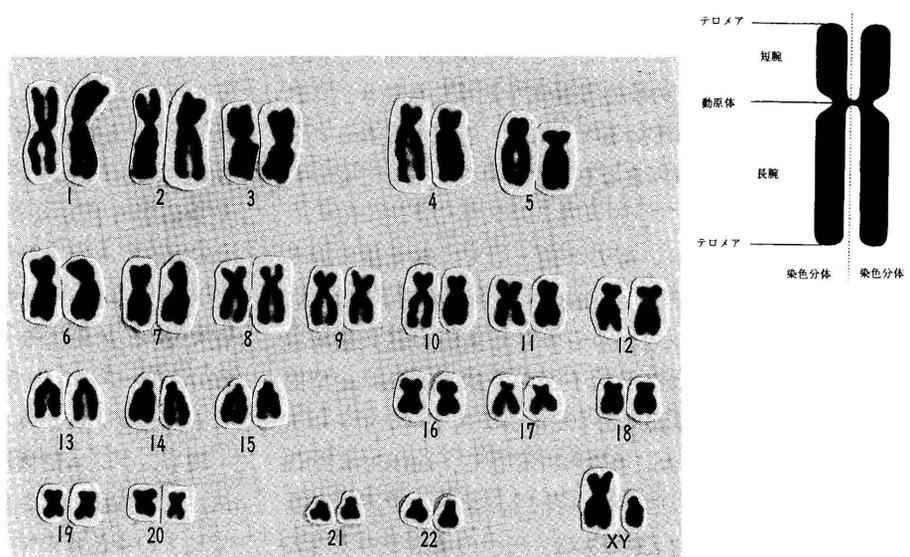
alternative splicing といわれる。一方で、1 番目のエクソンを複数もち、異なる調節領域から転写を開始することにより蛋白産生の組織特異的調節を行っているような遺伝子もある。イントロンは遺伝情報をもたないが、切り出されたイントロンの一部は機能をもっているものがある事が最近分かった。

上述したように、DNA から転写された RNA は RNA として働く場合と、蛋白質に翻訳されてから機能をもつ場合がある。蛋白に翻訳される mRNA には一般に 5' 側と 3' 側には蛋白に翻訳されない領域がある (非翻訳領域)。5' 末端にはキャップ構造といわれる特殊な構造が付加される。また、3' 非翻訳領域の終わり近くに AATAAA という配列 (polyadenylation signal) があり、これが目印となって 3' 末端に複数の A がつけ加えられる (poly A)。

3 染色体の構造

DNA はそのほとんどが染色体の上にある。染色体は細胞の分裂期に顕微鏡で観察可能で図 3 のように見える。染色体は DNA とそれを折りたたむためのヒストン、ヒストン以外の DNA 結合蛋白の 3 つから成り立っている。分裂期に見えるものは既に DNA の複製を終え、今まさに分かれようとしているもので、それぞれの姉妹染色分体が一つの 2 本鎖 DNA にあたる。染色分体では DNA の鎖はきちんと折り畳まれ、染色体の端から端までいったり来たりせずに順序よく並んでいると考えられている。分裂期でもある程度折りたたまれた状態にあるらしく、

図 3 染色体 (ギムザ染色したもの) とその模式図



DNAの鎖としては離れた位置にあるものが、折りたたまれた状態ではごく近い位置に存在することは想像できる。遺伝子の発現異常を来すような染色体の構造異常で、染色体の切断点と遺伝子が100Kbも離れているようなものがあることが明らかになってきた。

ヒトでは1番から22番までとXYをあわせた一倍体(ハプロイド)DNAは約30億塩基とされている。一番大きな1番染色体は全体の9%,一番小さい21番染色体で約2%程度といわれており,単純計算すると一つの染色体は6千万(60Mb)から2億7千万塩基(270Mb)のDNAを含む。動原体には約170塩基の単位がより高次の反復になっていて全体として反復配列が数百Kb以上続いている。テロメアは短い反復単位の繰り返してTTAGGGがテロメア側に向かって並んでいる。テロメアの複製にはテロメラーゼといわれるRNA酵素複合体が関与する。

染色体のバンドはさまざまな処理によって見られるが,どのような染色法によっても基本的な染まりのパターンは類似しており,染色体DNAの一次構造,または高次構造を反映している。現在行われている標準的なG-band分染法では染色体全体で400-850本のバンドが見える。したがって1本のバンドは数Mbから十数Mbに相当する。

ゲノムDNAの量は生物によってまちまちであり,ショウジョウバエはヒトの6%,マウスはヒトよりやや多く,サンショウウオやユリはヒトの数十倍である。遺伝子の数はゲノムサイズと直接的には関係しないと考えられており,ゲノムサイズの大きなものは不要部分が多いと考えられる。

ヒトのもつ遺伝子は約5万から10万といわれている。10万という数字は,ヒトのさまざまな疾患の新生突然変異の頻度から遺伝子の突然変異率を計算した場合,これ以上の遺伝子をもっているとすると,世代を重ねるごとにどんどん変異が積み重なり種が成り立たないと思われる計算値である。

遺伝子のうち小さいものの例として,睾丸決定因子SRYは約600bpである。 α -globinはDNAの長さは0.8Kbであるがイントロンを2つもつため実際の翻訳領域は0.5Kbである。大きな遺伝子としては,Duchenne型筋ジストロフィーの原因遺伝子であるジストロフィン(X染色体上に2Mb以上にわたって存在し,mRNAは16Kbでイントロンは60以上ある(表1)。おおよそ平均的な遺伝子のサイズは数Kbである。一つの遺伝子の平均の

表1 遺伝子のサイズ (Mckusick, Mendelian, Inheritance in Man : Johns Hopkins University Press より改変)

遺 伝 子	遺伝子のサイズ(Kb)	翻訳領域の長さ(Kb)	イントロンの数
小さいもの			
α globin	0.8	0.5	2
SRY	0.6	0.6	0
β globin	1.5	0.6	2
インシュリン	1.7	0.4	2
中位のもの			
コラーゲン1 pro α 1	18.0	5.0	50
アルブミン	25.0	2.1	14
Adenosine deaminase	32.0	1.5	11
第IX因子	34.0	2.8	7
フェニルアラニン水酸化酵素	90.0	2.4	12
大きなもの			
第VII因子	186.0	9.0	26
ジストロフィン	>2,000.0	~16.0	>60

長さを3Kbと考えると5万個の遺伝子がお互いに重ならないければ,延べ1億5千万塩基という値になる。したがって,全ゲノムが30億塩基とすると遺伝子をコードしている部分はゲノム全体の5%ということになる。

残りの部分は,遺伝子ではない不要部分ということになるが,当然さまざまな遺伝子の発現に影響を及ぼしているであろう事は想像できる。このような部分は,反復配列や,偽遺伝子や,過去のウイルス挿入の痕跡などからなる。ヒトの反復配列には直列に集まって存在するサテライトDNAと,ゲノム上のあちこちに散らばっているAlu配列,L1配列などがある。偽遺伝子には遺伝子重複がおこった後,壊れて偽遺伝子となる場合と,一旦RNAとなりスプライスやpoly Aの修飾を受けた後に染色体に逆に入り込んだprocessed pseudogeneがある。

4 ゲノム計画の現状

ゲノム計画としてヒトのもつ約30億塩基の配列を,全て読んでしまうという発想が出てくるには,大きなDNA分子に対するアプローチの方法が確立したことがあった。その一つは,パルスフィールド電気泳動法で,それまでせいぜい数十Kbの大きさまでしか分画できなかったDNAを数百Kbからメガベース単位で分析できるようになった。また,YAC(酵母人工染色体)の登場で,やはり数十Kbの大きさまでしかクローン化できなかったDNAを数百KbからMb単位でクローン化できるようになった。各YACに1Mbのヒト由来DNAをクローン化できるなら,3000個のYACがあれば,ヒトの全ゲノムがカバーできることになる。

現在,ゲノム計画は実際の塩基配列の決定(シーケンス)という段階にさしかかっているが,その前段階としてヒトの染色体のマップづくりが行われた。マップには

遺伝的地図と物理的地図があり、前者はそれぞれの染色体上の座位 (locus) の組換え頻度の地図で、後者は実際の DNA の長さをもって座位間の距離としたものである。染色体の領域によって組換え頻度の高い領域もあるが、大まかにいって遺伝的距離と物理的距離は良く相関し、おおよそ 1 cM (センチモルガン) = 1 Mb と考えてよい。

遺伝的地図の作製に大きな役割を果たしたのは、マイクロサテライト (CA リピート) である。CA リピートは CACACA... という反復が繰り返すもので、ゲノム中、数十 Kb に一つくらい存在する。CA が 15 回を越して繰り返しているものは、多くの場合その繰り返し数に多型があり、家系内の染色体の伝わり方を調べるのに都合がよい。これを大量に準備し、大家系の検体を集めて解析を行う一方、それに対応する YAC をスクリーニングして地図づくりに大きく貢献したのがフランスの Weissenbach から CEPH (Centre de Recherche sur le Genome Humain) のグループである。

また、より重要なのは実際に mRNA になって転写されている遺伝子であるからそれを決めた方がよいという考え方があり、それぞれの組織の cDNA ライブラリーに含まれるクローンを片っ端からシーケンスしていく方法が行われた。NIH の Venter がこの結果を全て特許申請するというで話題をまいたが、結局単に塩基配列を決めただけのクローンはそれだけでは特許にならないことになった。また、この方法だと、組織でよく発現しているものはたくさんのクローンとなっているわけで、同じ cDNA ばかりシーケンスしてしまうという危惧があるが、最近ではマイクロチップとして既知の遺伝子をふるい分けてしまう方法が開発されてきた。これらの塩基配列は EST (Expressed Sequence Tag) として、データベースに登録されている。

ヒトのシーケンスにかかる前に、もう少しゲノムサイズの小さな生物ということで、表 2 に挙げるようなモデル生物や細菌のゲノムシーケンスが計画的に行われた。各国チームが領域を分け合っており、枯草菌などは日本の貢献も大きい。最近、*Haemophilus influenzae* の配列が決まり、病原微生物のシーケンスにおいては当初あまり考えられていなかったような使い方、たとえば細菌の宿主を規定しているものは何かなど、も考えられるようになった。

表 2

全塩基配列が既に決定されたゲノム
古細菌：メタン生成菌
マイコプラズマ
真正細菌：枯草菌、大腸菌、ピロリ菌、インフルエンザ菌
らん藻
真核生物：パン酵母
全塩基配列が決定されつつあるゲノム
古細菌：硫酸還元菌、好塩菌、好熱菌ほか約 10 種類
真正細菌：結核菌、らい菌、ブドウ球菌、肺炎球菌、淋菌、髄膜炎菌、サルモネラ菌、コレラ菌など約 30 種類
梅毒トレポネマ、トラコーマクラミジア
真核生物：マラリア原虫、分裂酵母、コウジカビ、アカパンカビ
シロイヌナズナ
センチュウ、ショウジョウバエ、マウス、ヒト、ほか

5 ゲノム計画と社会

ヒトのゲノムは個人と個人で 1000 塩基に 1 個くらい異なっている (最近では 300-500 塩基に 1 個という話もある)。ところで、現在行われているゲノム計画では誰のゲノムを決めているのだろうか。また、どの研究者・会社がどの染色体のどの場所を決めることになっているのだろうか。

ヒトの場合、微生物やモデル動物と違って一部の染色体を除いて各研究者が全く勝手に好きなどころを決めているというのが現状である。アメリカではある領域をシーケンスするはずになっていた研究者がベンチャー企業へ就職して、計画自体が消えてしまった領域もある。誰の DNA を決めるというのも、全くコンセンサスなく研究室の都合で決められている。

多くの研究者は疾患遺伝子を捕えることに興味がある。したがって頻度の高い疾患・注目度の高い疾患の責任遺伝子があると推定される領域に集中的に研究者人口が集まるというのがヒトゲノム解析の特徴である。このような状況なので、2003 年にヒトの全配列が決まるといわれても本当だろうかと思う余地は十分残されている。ともかく、結果は 5 年待てば分かる。

個人個人の形質の違いとその基盤となる DNA の違いが関連づけられることは、ヒトという生物の研究からいって好ましいことであるが、一方でさまざまな社会問題を引き起こす可能性がある。特に、最近のアメリカではコロンブス 500 周年事業において、白人が先住民を滅ぼしその権利を奪ってしまったという反省が一部にあり、行政も研究者も minority に対する扱いや ethnic group

を話題にすることに異常に神経質である。また、疾患遺伝子を調べてリスクファクターをもった人は保険に加入できないというような社会問題も出てきた。個人と家族のプライバシーの保護という倫理的問題もある。

しかし、アメリカの場合にいう保険は生命保険ではなく、日本でいうところの医療保険に入れないことが特に問題である。日本は世界に冠たる国民皆保険制度の国であるから、あまりこの方面に神経質になりすぎる必要はないと筆者は思っている。それよりも、健康な人の遺伝的背景が明らかになることによって、なんらかの補助手段で皆がそれを見習えるなら、これほど結構なことはない。

おわりに

ヒトゲノムに関する研究について概説した。ヒトのもつDNAの全配列が決まっても（配列から遺伝子を推定はできても確定はできないので）、必ずしも全ての遺伝子が明らかにされるといってもいいものでもない。ヒトゲノムはとてつもないデータの宝庫となるが、そのままでは意味を持たない。これをもとにして、遺伝子がどのように発現し、どのように働いてさまざまな生物機能を担っているのかを解析する時代になる。特に、ヒトにおいては比較の基となるデータができることによって、個体間の遺伝的な違いに基づく、それぞれの個体の表現型の違いが扱えるようになる。