

そのような中で重要となるのが医師と患者との関係で、これが現在、大きく変化してきている。例えば、患者はインターネットなどによって多くの医療情報を入手できるようになり、また自分の医療情報をもっと知りたいと思うようになっている。このような状況において、アメリカなどの一部の医療機関では、あなたは自分自身の医療情報のオーナーであると説明し、患者に smart card（前述）を持たせている。まだ、そこに遺伝子データは入れられていないが、これを入れることも可能である。

まとめとなるが、personalized medicine とファーマコゲノミクスは同じコインの両側である。その両方を知っていなければならない。

2. 質疑応答、総合討論

○水嶋部長

SNP の解析は結局、投入するコストと比べて、当たり前データが出ることにならないか。例えば、糖尿病については従来のように家族歴を聞いたほうが効率的である、ということにはならないか。

○村松教授

それはある面で正しいと思う。ただし、例えば親が糖尿病であるため、糖尿病になることを非常に心配している人がいるとすると、SNP を解析することで「あなたは糖尿病になる遺伝子を持っていなかった」と告知できるようになる可能性がある。また逆に、その遺伝子を受け継いでいたら、糖尿病について、より注意してもらうこともできる。

○水嶋部長

例えば喫煙しても肺がんにならないということが明らかになったら、SNP の解析は、健康教育を後退させることになりかねない側面がある。

○村松教授

喫煙については、肺がんだけでなく、COPD、心血管障害、炎症性腸疾患などにも関与しているので、やめたほうがよい。ただし、厳然たる事実として、80 歳、90 歳まで喫煙習慣があっても肺がんにならない人もいる。これは興味深いことでもある。また、それは、どこまで自己責任を認めるかということであり、個人のリスクに対する考え方をどう教育するか、という問題にもなる。

○今井部長

先天性の遺伝疾患であれば、一かゼロ、つまり遺伝子の有無で決まる。しかし、高血圧、糖尿病などは、いろいろなタイプがある。本態性高血圧とひとくくりになっているものでも、例えば ACE、あるいはインスリン抵抗性により動脈硬化が早く進行したというように、いろいろな原因があるだろう。このように現代人の生活習慣病は多因子であり、そのオーダーメイド医療、personalize した治療法を SNP によって明らかにする、ということなのか。

○村松教授

これまで日本では非常に大きなお金が SNP 解析につき込まれたが、そこでは 1 個の遺伝子が非常に効くのではないかと期待されていた。確かに、遺伝子が効いているのはわかった。しかし、リスクのある遺伝子を個々に検討してみると、そのオッズ比は 1.8 程度で、2 に達しないのがほとんどである。それだけに今度は、この遺伝子とこの遺伝子を組み合わせればオッズ比が高くなるかどうかといったように、研究の種は尽きない。ただし、その情報をフィードバックするタイミングなどを考えると、公衆衛生的に本当によいことかどうか、もう一度検討する必要がある。ここで唯一、価値があるとすれば、遺伝子と生活環境因子の相互作用をみていくことである。ここで一番大切なのは、持っている遺伝子が危ないほうだとわかったら、生活環境因子を抑える取り組みをすることが可能だからである。これは非常によいことである。

○遠藤主幹

この研究班の目的は、厚生労働省の研究ではヒトゲノムの分野でどこに焦点を当てるか、検討することである。その 1 つの答えとして、これからはゲノムだけでなく、ゲノムと生活環境因子をみるということか。

○村松教授

これからの SNP 研究その両方がないと、まさに片手落ちである。また、そのために、どのようなシステムをつくるかが課題となる。

○林次長

そのシステムに関することだが、アイスランドに視察に行った。アイスランドでは 100 年分のカルテが残っていて、300 年分の家系図が蓄積している。そこで、法律に基づいて、全国民の血液を採取し、SNP の検査をする。そして、家系図とカルテを突き合わせ、SNP がどのような疾患と関係しているかあらいざらいに分析しよう、と考えていたようである。また、それを実際に行う deCODE 社にも行ったが、現在、同社の研究事業は行き詰まっている。

○村松教授

アイスランドは人口が約 28 万で、すべての国民に 7 桁のアルファベットと数字組み合わせた暗号を対応させるなど、国民の遺伝子のデータを匿名化し、deCODE 社の中に入れて解析することを目指していたが、できなかった。これは先見の明のある方法だと思ったが、結局、1 民間企業では対応しきれなかったということだろう。しかし、日本において 30 万くらいの人口を対象にアイスランドのようなことができれば、本当に興味深いことになる。ただし、家系が 1000 年前までさかのぼれるアイスランドと同じことをしようとしても無理なので、日本の特徴を出すようにしなければならない。

○林次長

日本での遺伝子分野の研究者の大きなグループは、遺伝子の異常な部位をベクターのウイルスに

よって書き直すということに力を入れてきた。しかし、これは、たいていは失敗に終わっている。また、医薬業界の人によると、それは気の遠くなるようなアプローチである。むしろ、村松教授が述べたように、薬剤感受性や副作用との関係を遺伝子の観点から検討する。しかも、それは、新薬の開発ではなく、既存薬の洗い直しに用いる。このようなアプローチをしたほうが医薬業界としてはコスト・エフェクティブで、開発も早くなる、という。これには、なるほどと思った。

もう 1 つの新しい動きとして、山形県鶴岡市の慶應義塾大学先端生命科学研究所を挙げることができる。ここの敷地と建物はすべて、同県が慶應義塾大学に対して無料で提供している。同研究所の研究者からみると、前述の医薬業界の人の考え方も少し遅れているという。例えば、機能の不明なタンパク質もたくさんあるので、その分析をする。これで機能がわかれば、そのタンパク質を合成するなど、治療のほうに持っていく。このような発想のほうが 21 世紀的である、と同研究者は指摘する。また、新聞の報道によると、文部科学省もこれからの研究ではプロテオミクスをコアに据えるという。

このような現状をトータルで考えると、厚生労働省としては遺伝子の研究をどちらの方向に誘導すればよいとお考えになるか。

○村松教授

プロテオミクスは、非常に基盤的なものと創薬に近いものがある。また現在、一般的なプロテオミクスとしては、文部科学省の「タンパク 3000 プロジェクト」があり、構造解析などが行われている。これは基礎の分野であり、厚生労働省が取り組むとすれば、薬剤とダイレクトに関係していたり、機能性食品と関連するプロテオミクス、ゲノミクスがよいのではないかと思う。例えば、漢方薬におけるプロテオミクスである。ただし、薬剤に関しては大手の製薬会社はかなり取り組んでいる可能性があり、そこが取り組んでいないような分野に力を入れたほうがよいだろう。その意味では、予防あるいは栄養・食品のような分野がよいかもしれない。ニュートリションミクス（栄養ゲノム学）という言葉もできており、今後、栄養の分野も非常に重要になってくると思う。

○今井部長

例えば、糖尿病に関連していると思われる 3 つの SNP があり、その 3 つを持っている人は糖尿病になりやすいとして、健診でその 3 つを持っている人を同定できたとする。倫理的な問題は別にして、その人たちを 2 つのグループに分け、一方には健康教室などで徹底的に指導をする。もう一方は、何もしないようにする。これで 3 年ほど経過観察をして、何もしないグループでは 100 人中 80 人が糖尿病を発症したが、徹底的に指導したグループは糖尿病がかなり抑えられた。これは素晴らしいことだが、本当のところ、その 3 つの SNP が関与しているかどうかわからない。そこで、そのデータを臨床のほうにフィードバックし、もう一度検討したところ、3 つの SNP が関係しているように見えたのは偶然であった、ということがわかった。そこで、もう一度、対象者の精度管理を行う。これを社会実験というが、SNP については、そのようなシステムがあと 5-10 年くらい必要だと思う。

○村松教授

そのような取り組みこそが必要だと思う。それは結局、どのようなサンプリングをすればよいのかということである。

○今井部長

わたしは疫学が専門であり、そのようなシステム、ノウハウを提供することが使命であると思う。例えば運動療法を本当に行うなど、精度管理をきちんと実施し、本当のターゲットとなるものを見つきたい。それが今後5-10年の疫学のテーマとなるのではないか。

○林次長

そのような発想の研究テーマは出ているのか。

○今井部長

ゲノム疫学に取り組んでいる人たちは、そのようなことを少し意識している。しかし、実験系の研究者が多く、現場に行って対象の本人と話したりするのはむずかしい。

○村松教授

久山町では、そのようなことが行われている。

○林次長

今井部長が主任研究者を務めると仮定し、そのようなプロジェクトを組むとしたら、どのようなタイプの人材を何人くらい集めればよいのか。

○今井部長

わたしのような疫学を専門とする者と、村松教授のようなゲノム疫学に詳しい人材のコンビネーションが望ましい。

また今後、費用対効果についても検討しなければならない。例えば、病院や診療所などで実際に応用できるのか。このあたりは、わたしも非常に興味を持っている。

これまでブラックボックスだったものに対して一歩踏み込んだのが、ゲノム疫学である。ブラックボックスの中はわからないが、このような結果があって、こちら側にはこのような原因があるようだ、と一歩踏み込んで説明ができるわけである。

ただし、SNPの弱いところは、まだタンパク質との関係がわかっておらず、現象面だけをみていることである。そのため、論文を書いても、レフリーから「これは偶然でしょう」と指摘されることが、よくある。

○村松教授

費用対効果については、非常によくわかっている SNP についてコホートの中で検討するのがよい。

○今井部長

例えば、非常に太りやすい人であることがわかっているならば、その人に対しては食事指導をすればよい。また、心理テストを行い、例えばものぐさな人、問題を先送りする人を見つけるのもよい。人間は、メンタルと体でできている。その両面を調べたうえで指導していけば、費用対効果は非常に良くなると思う。

○水嶋部長

現在、古典的な生活習慣で座標軸を区切っている。これで例えば高脂血症とみなすのはナンセンスである。すなわち、女性の更年期では、コレステロール値が上昇することがある。このような人は、決して心血管疾患のリスクが高いわけではない。しかし、メタボリックシンドロームの診断基準を適用すれば、かなりの人がそれと診断されてしまう可能性がある。このような例において、座標軸の区切り方を探る作業の1つとして SNP があると思う。

取り組んでも効果がない部分が SNP によって最初からわかれば、これは非常に効果がある。例えば糖尿病患者に対しては一律に「運動してください」「甘いものを控えてください」と指導していて、それらの効果がまったくない人も含まれている。この人たちは脱落したり、徒労感がある。将来的には、そのような人に役立つだろう。また、そのモデルケースが出るとよい。

○村松教授

モデルケースが出れば、たいへんなことになると思う。遺伝子を測定し、生活習慣病が改善し、医療費が少なくてすむようになれば、たいへん社会的なインパクトがある。

○水嶋部長

マーケティング的にいうと、対象をどうセグメンテーションするかということである。つまり、どう評価し、ターゲットにするか。その作法として SNP を使うと、従来よりも効率がよい。このような比較が、医療経済的あるいは政策疫学的な観点だと思う。

○今井部長

現在、千葉県によいコホートがあるので、検討している。

○水嶋部長

生活習慣病の新しい対策などで、ハイリスクストラテジー、ポピュレーションストラテジーという2つの観点が打ち出されている。つまり、生活習慣病のリスクは連続性であり、ある人は必ず生活習慣病になり、ある人は絶対にならない、ということではない。また、リスクの高い人については、1人当たりのリスクは高いが、そのハイリスクの集団に入っている人の数は少ない。数からいえば、オッズ比1.2程度、リスクがやや高い程度の人が多い。人数とリスクのかけ算をすれば、集団としては、このオッズ比1.2程度の人たちにおいて生活習慣病になる数が多いことがわかる。厚生労働省としては、より効果的な対策を探らなければならないので、少しでもリスクの高い人たちに対するアプローチをするという視点が必要だと思う。

○今井部長

これまでは、健診で血糖値が高かったら健康指導をするといったように、目の粗いザルを用いていた。ここに SNP を持ってくると、従来よりも効果がある。また、SNP は新しいものだということで、対象も興味を持ってきて、脱落する人も少なくなる。効率性という観点からは、よい。ゲノム疫学によるコホート研究を 5 年くらい続けて本当に効果があることがわかれば、大きく盛り上がり、製薬会社などの取り組みも進むのではないかと。

○林次長

厚生労働省の中でハイストラテジーのアプローチをしているのは、むしろ保険局であろう。つまり、医療費の削減という立場があるので、保険点数の切り下げだけでなく、予防についても考える。しかし、予防についてはアウトカムが出ていないので、当てにできるのかという立場もある。このような状況において、分子疫学的なアプローチ、生活習慣、健康教育などのコンビネーションの中で、SNP に関するプロジェクトをつくる努力をしてみる価値があるのではないかと。

○遠藤主幹

諸外国ではどうなのか。アメリカでは、コホートを使って、分子疫学あるいは遺伝子と環境因子について研究しているのか。例えばフラミンガムスタディでは、どうなっているのか。

○村松教授

最近、フラミンガムスタディにおいてもゲノムの解析が始まっていて、その論文も少しずつ出てきている。ただ、ヒトの研究となると、医療システムまで関係してくる。アメリカは、保険行政はほとんどなきに等しい。このため、分子疫学については、アメリカはあまり手本にならない。また、アメリカの場合、人種問題が非常に大きいので、データの解析をするときでも「白人しか入っていない。黒人もヒスパニックも入れなければ駄目だ」という声が出てくる。

それに関して昨年（2005 年）6 月、初めての黒人特異的な薬剤として BiDil（バイディル）が FDA から認可された。この BiDil は臨床治験では有効性が認められなかったが、黒人だけを対象にすると有効性が出たのである。おそらく、薬剤代謝酵素の関係で、黒人に代謝しにくい人が多く、血中濃度を保つために有効性が出た、と推測される。このように、オーダーメイド医療を行うには、まだ深い隘路がある。

○今井部長

SNP の研究自体は、日米で差はないのか。

○村松教授

故小渕首相が 2000 年、ミレニアムプロジェクトにゲノムプロジェクトを入れたので、この 5 年間でみると、日本はアメリカとヨーロッパを抑えて、世界のトップに来ている。ちなみに、2000 年までは、ヒトゲノムの解析において日本の貢献度は 7% くらいしかなかった。しかし、2000 年以降、日本の貢献度は約 25% と高くなっている。費用対効果でいえば、アメリカと遜色ない。

○今井部長

日本は、ゲノム解析のプレートなどの開発においても遜色ないのか。

○村松教授

そこは、少し差がある。例えば、一番よい機械をつくっているのは今でもアメリカである。また、日本の場合、試薬も外国から多く買っている。ヒュービットジェノミクスという民間会社の立場でいうと、例えばアメリカのロシュ・ダイアグノスティクス社、AGI、アフィメトリクス社などの技術をどうしても使うことになる。

○今井部長

アメリカの場合、ベンチャー企業を買収するなど、対応が早い。

○水嶋部長

そういう意味では、産学連携の仕掛けをするのも1つの方法ではないか。

○村松教授

実際、産学連携はかなりできている。しかし、病院と産業の連携ができていない。

○遠藤主幹

村松教授の東京医科歯科大学には付属病院があるわけだが、どのような病院との連携が必要なのか。

○村松教授

国立病院機構の病院などを念頭に置いている。大学病院の場合、契約をすることによりサンプルの授受などもスムーズにできる。しかし、国立病院機構の病院と共同研究をするのは、非常にむずかしい。

○遠藤主幹

それは制度の問題か。

○村松教授

もともと産学連携という言葉もあったように、大学の場合、知的財産本部のようなものがあり、そこが窓口となって対応している。一方、病院の機能だけの施設の場合、共同研究のうまいルートがあるのかと危惧している。

○遠藤主幹

臨床という視点からは、むしろ大学病院のほうがいろいろな患者が来ているのではないか。大学病院はコモンディーズが少ないので、コモンディーズと SNP の関係という意味でサンプル数が

足りないのか。

○村松教授

大学病院だけでは患者数が足りない。コモンディジーズの人が大学病院に行くかといえば、そうではない。例えば、糖尿病患者であれば東京都済生会中央病院、クローン病患者であれば社会保険中央総合病院、透析患者であれば聖隷佐倉市民病院（旧国立佐倉病院）というように、患者は特徴のある病院を受診することが多い。いわゆる高度先進医療は大学病院に取り組んでもらうとしても、その次の「高度な医療」のレベルを行う病院のほうが、患者も多く、しっかりと患者を診ていると思う。また、それらの病院は、大学病院と違って研究教育の義務もないわけである。

○林次長

わたしが以前、文部科学省の科研費審査委員を務めていたときに気づいたことだが、一部の大学では、協力病院や共同研究者も含めて臨床研究を大規模な全国ネットワークで行うことができる。しかし、他の大学では、そのような広がりを持った臨床研究ができない。

○今井部長

その大学の全国ネットワークも、実際に臨床研究ができているのは環太平洋の特定の地域、特急の止まる都市である。一方、国立病院機構の病院は全国各地にあるので、ゲノム疫学の観点からもよい。きちんと精度管理ができる国立病院機構の病院で臨床研究を行えば、非常に精度の高い SNP が同定できる。

○林次長

それがシステムづくりの1つの大きなエレメントである。

○今井部長

厚生労働省がリーダーシップをとって、そのような形で推進する。また、ポピュレーションストラテジーで行くのであれば、北海道から日本海側も含めるためにも、そのような発想が必要である。

ゲノム研究、ゲノム疫学は、対象を混ぜないといけない。例えば SNP については、関西地区だけで実施した研究ではこういう結果が出たが、関東地区ではそのような結果が出ない、ということもあり得る。いろいろな地域で、対象も数万の単位で研究を実施して初めて、SNP の真実が明らかになることもある。

○林次長

本日は、いろいろとお教えいただき、ありがとうございました。たいへん勉強させてもらい、またディスカッションも幅広い内容で、奥行きもあつた。今後の方向性を考えるうえでの要素を学ぶことができた。今後とも協力をお願いしたい。

II. 分担研究報告

厚生労働科学研究費補助金
分担研究報告書

1. ヒトゲノム研究に関する国際的動向を踏まえた評価のあり方に関する研究

分担研究者 遠藤弘良 国立保健医療科学院 企画調整主幹

研究要旨：

ヒトゲノム研究の評価のあり方に関して、米国NIHにおいて調査、情報収集を行った。その結果、NIHにおけるヒトゲノム研究のpeer review制度の改善、電子化の促進、学際的な研究等の現状が明らかとなり、我が国のヒトゲノム研究事業評価のための示唆を得ることができた。

A. 研究目的

ヒトゲノム研究の評価方法に関する国外の実施状況について調査を行い、我が国の評価事業の参考とする。

B. 研究方法

米国 National Institutes of Health (NIH) の研究評価制度については、すでに平成 14 (2002) 年度厚生労働科学研究「保健医療分野における研究評価のあり方に関する研究」における分担研究として、曾根智史・国立保健医療科学院公衆衛生政策部長（当時）によって調査、報告がなされている。

そこで今回は、ヒトゲノム研究の評価に焦点を絞り、国立ヒトゲノム研究所 (NHGRI: National Human Genome Research Institute) ならびに NIH の CSR (Center for Scientific Review) で関係者に対して面接によって調査した。

また、上記平成 14 (2002) 年度厚生労働科学研究における調査から約 4 年が経過し、NIH の研究評価制度自体も変化しており、この点についても補足的に調査を行った。

C. 研究結果

1. ヒトゲノム研究の評価

1) 応募から一次審査 (peer review)

NIH の研究助成事業、グラント (grant) 交付は一般に、二次にわたる審査を経て決定がなされる。また、その公募は、一般公募 (R01: investigator-initiated research proposals) と指定課題公募 (RFA: request for applications) の 2 つに大きく分かれている。

研究助成の申請書は、すべて CSR で受け付ける。一次審査として、CSR では、申請書の内容に基づいて、CSR の下にある study section で評価、peer review を行うとともに NIH 内のどの institute が助成をすべきかという観点からの仕分けを行う。

NIH という傘の下に、がんや精神の分野など 26 の研究所 (institute) あるいは center が設立している。二次審査は、それらの institute あるいは center が設置している National Advisory Council において実施する。ここでは、研究の細部ではなく、研究の全体的な方向性が各 institute あるいは center の使命に合っているかどうかの観点から審査を行う。

ヒトゲノム研究についての公募も同様で、一般公募と指定課題公募の2つに大きく分かれています。一般公募は、特にテーマを限定せず、研究者の自由なアイデアに基づく研究を受け付けるものである。一方、指定課題公募は、NHGRI が取り組んでもらいたいテーマを提示し、それについての研究を受け付けるものである。

2006年3月現在、NHGRI が公募中の指定課題は次のとおりである。

- ・ Identification of All Functional Elements in Selected Model Organism Genomes
 - ・ A Data Coordination Center for the Model Organism ENCODE Project (modENCODE)
 - ・ Public Consultation to Inform the Design of Possible Large-Scale Studies of Genes and Environment in Common Disease
 - ・ Genome Sequencing Centers
- また、それとは別に、Advanced DNA Sequencing Technology Development として、次のようなテーマについて公募している。
- ・ Near-Term Technology Development for Genome Sequencing
 - ・ Near-Term Technology Development for Genome Sequencing
 - ・ Near-Term Technology Development for Genome Sequencing
 - ・ Revolutionary Genome Sequencing Technologies
 - ・ Revolutionary Genome Sequencing Technologies
 - ・ Revolutionary Genome Sequencing Technologies

研究費の規模は、一般公募の場合、研究者が自由に設定・申請するので、さまざまである。指定公募課題の場合、100,000ドル台の大型の研究と1,000ドル台の小規模

な研究の2種類がある。

応募に対する評価の実施主体は、一般公募と指定課題公募で異なる。一般公募は NIH の CSR で評価する。指定課題公募は、NHGRI 内の Genome Research Review Committee で評価する。

実際には、応募のうち70~75%はCSRで審査を受けている。それ以外、全体の25~30%は、NHGRI 内の Genome Research Review Committee で審査を受けることになる。

CSR での審査は、225 の study section が担当する。ただし、その study section には、ヒトゲノム関係だけでなく、がんや循環器など種々の section がある。それらの study section のうちヒトゲノム関係は下記の9つの section があり、ヒトゲノム関係の一般公募については、そのいずれかにおいて審査されることになる。

- ・ Molecular Genetics
- ・ Genomics, Computational Biology and Technology
- ・ Genetic Variation and Evolution
- ・ Genetics of Health and Disease
- ・ Gene Therapy and Inborn Errors
- ・ Ethical, Legal and Social Implications of Human Genetics
- ・ Genes, Genomics and Genetics Small Business Activities
- ・ Genomics, Genetics, DNA Replication, and Gene Expression Fellowship
- ・ Genes, Genomics and Genetics IRG

例えば、ゲノムと病気との関連についての研究の申請であれば、Genetics of Health and Disease の study section で審査されることになる。

最近の傾向としては、ゲノムだけでなく、ゲノムとそれに関連する分野といったように、multi-disciplinary (学際的) な課題が増加している。そのため、上記の study section が単独で審査をするのではなく、

関係する study section が調整し、合同で審査をするケースも多くなっている。また、そのため、Genes, Genomics, Genetics を統合した study section (Genes, Genomics and Genetics IRG) も設置されている。

2) 二次審査

がん、糖尿病、アルツハイマー病等、特定の疾患に関係した研究については、その疾患に関係する institute または center に送られ、ここで二次審査が行われることになる。それらの機関の中に National Advisory Council があり、ここが実際の審査を担当する。例えばがんの遺伝子に関する研究であれば、NHGRI ではなくて National Cancer Institute (NCI) が審査を担当する。また、その研究が採択された場合、研究費についても同研究所の予算から支援することになる。

NHGRI では、ヒトゲノム関係の主に technology development に関係した研究を審査している。

NHGRI での審査は、National Advisory Council for Human Genome Research (NACHGR) が担当する。NACHGR では、CSR において審査された一般公募、NHGRI 内の Genome Research Review において審査された指定公募の両方を審査することになる。

2003 年 4 月、約 30 億の塩基配列決定することによってヒトゲノムプロジェクトが終了した。それを受けて、NHGRI では新たな mission を策定し、Road Map として示している。NHGRI における二次審査は、その Road Map と mission に合致しているかどうかという観点から審査がなされている。

その Road Map においては、Genomics to Biology、Genomics to Health、Genomics to Society という 3 つのテーマが挙げられている。また、それと横断関係にあるのが、Resources、Technology Development、Computational Biology、Training、

Ethical, Legal and Social Implications (ELSI)、Education という 6 つの領域である。

なお、NHGRI の予算は、NIH の全予算の 2～3% に相当する。

3) 課題

NHGRI でヒトゲノムに関して審査・採用する課題については、次のような特徴がみられる。

第 1 は、technology development でゴールがはっきりしている課題が採用される傾向にあることである。

第 2 は、学際的な課題が増加していることである。

2. 研究評価制度の課題と改革

平成 14 年度の調査（前述）以降、NIH における研究評価制度も変化している。それを踏まえて、現在における NIH の研究評価制度の課題として、次の 4 点を挙げるができる。

1) 評価期間の短縮

NIH の現行のシステムでは、年 3 回の申請サイクルがある。しかし、応募を受け付けてから第一次審査、第二次審査を行い、各 institute あるいは center の director が決定し、採用通知を出すまでに、9 ヶ月かかっている。科学が急速に進歩している現代社会において 9 ヶ月もかかるのは遅すぎるので、これを短縮してほしい、という声が強くなっている。

特に、申請者が第一次審査の後、その結果を受けるまで 2～3 ヶ月の時間を要している。このため、その結果を受けて研究内容を修正し、再申請しようとしても、次の申請サイクルに間に合わない。そこで NIH では、第一次審査の結果のフィードバックを 1 ヶ月半ほど短縮し、現行の半分程度の期間にするような取り組みを試行的に実施している。

それらの期間を短縮するため、申請の電

子化、インターネットを使った受け付けに取り組んでいる。NIHでは、2006年10月からは全面的に電子化に移行する計画がある。

2) 評価委員 (reviewer) の採用

近年、申請が急増しているため、現在約1万5,000人の評価委員を採用して peer review をこなしている。これらの評価委員はCSRでの225の study section に分かれ、申請書の評価に取り組んでいる。評価委員は、申請書の評価のためのデスクワークとして年間に1~3ヶ月ほどかけている。すなわち、申請書は年3回渡される。また通常、1回につき申請書は1件ではなく数件の割り当てがある。1件の申請書でも最低25ページはあるので、1回の審査のために申請書を読むだけでも最低1週間はかかる。それが年3回なので、年間では約1ヶ月かかる。それに一生懸命取り組めば、場合によっては3ヶ月くらいかかることになる。なおかつ、評価委員は年3回、平均2日間の評価会議に出席しなければならない。この評価会議は通常、ワシントンDCで行っている。

NIH の評価委員になるのは、専門家として名誉であり、所属施設内での立場にとっても有利なことが多い。しかし、前述のように大きな負担がある。特に臨床家の場合、日常診療を行いつつ年3回の評価会議に出席するのは、時間的に困難である。このため、臨床家を評価委員として採用するのがむずかしい状況になっている。

そこで、評価委員の負担を少なくするために、申請書の配布、コメント作成、評価委員同士の意見交換などについて電子メールを活用するなど、電子化の推進に取り組んでいる。ただし、電子化を推進するには、秘密の保持、securityの改善といった課題もある。

申請書は最低25ページあり、しかも各申

請者が任意の参考資料を付けることができる。また、参考資料の分量には、特に制限はない。このため実際の申請書は、膨大な書類となっている。そこで、任意の資料が付いて25ページを大幅に超えている現行の申請書を10ページに簡素化することで、評価委員の負担の軽減に取り組んでいる。

また将来のために、若手の評価委員を登用し、育成していこうとしている。

3) 臨床研究 (clinical research) の適切な評価

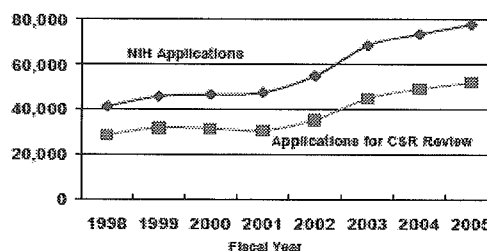
NIH では、実験室などでの基礎医学研究だけでなく、症例研究など臨床研究も重視している。しかし、基礎医学研究と比べると臨床研究の評価は低いという批判が、臨床家の間から出ている。そこで、臨床研究について適切に評価するように、改善に努めている。

4) peer review の透明性、説明責任、一貫性の促進

NIH への申請 (NIH Applications) は、1998年度には約40,000件だったが、2005年度には約80,000件になり、7年間で倍増している(図)。また、それに応じて、CSRがreviewする申請 (Applications for CSR Review) も増加している。一方でNIHの予算は頭打ちになっているため、申請が採用される割合が低下している。

そこで、peer review については、これまで以上に透明性を高め、説明責任を果たすとともに、一貫性を確保・促進しようとしている。

Applications Received for all of NIH and Applications Referred for CSR Review, FY 1998 - 2005



D. 考察

ヒトゲノム研究は最近、学際的な研究が増加している一方、NHGRI は、technology development oriented な研究を高く評価する傾向がある。また、NHGRI はヒトゲノム研究についての Road Map を提示している。しかし研究の方向性は研究者の自由で革新的な発想に任せるべきで、NHGRI が Road Map を示す必要はないのではないか、という意見がある。また NIH 全体としては臨床研究の推進にも力を入れており、NHGRI の technology development を重視する傾向に対しても批判的な意見も聞かれる。国際ヒューマンゲノムプロジェクトが終了し、その後のヒトゲノム研究のあり方について少なくとも NIH の中では、必ずしも意見が一致しているとは言えないとの印象を持った。

評価制度の改革は CSR が中心となり、審査期間の短縮、透明性・公平性の増加、積極的に進めているが、例えば電子化についてみると職員の対応にまだ時間がかかり、また security の問題が必ずしも解決しているわけではないようである。また評価委員臨床家や若手の評価委員の採用増加を推進する必要性の認識があるものの画期的な解決策があるわけはなさそうである。

E. 結論

米国 NIH と我が国における研究支援事業制度のあり方そのものに違いがあるため、ヒトゲノム研究の評価制度についても NIH の評価制度がそのまま我が国にもあてはまるものではない。しかし昨年度から我が国でもヒトゲノム研究事業に導入している peer review については、より透明性・公平性を高めるために NIH の評価制度は参考にすべき点があるといえる。

F. 健康危険情報

なし

G. 研究発表

なし

H. 知的所有権の出願・取得状況（予定を含む。）

なし

A vision for the future of genomics research

A blueprint for the genomic era.

Francis S. Collins, Eric D. Green, Alan E. Guttmacher and Mark S. Guyer on behalf of the US National Human Genome Research Institute*

The completion of a high-quality, comprehensive sequence of the human genome, in this fiftieth anniversary year of the discovery of the double-helical structure of DNA, is a landmark event. The genomic era is now a reality.

In contemplating a vision for the future of genomics research, it is appropriate to consider the remarkable path that has brought us here. The rollfold (Figure 1) shows a timeline of landmark accomplishments in genetics and genomics, beginning with Gregor Mendel's discovery of the laws of heredity¹ and their rediscovery in the early days of the twentieth century. Recognition of DNA as the hereditary material², determination of its structure³, elucidation of the genetic code⁴, development of recombinant DNA technologies^{5,6}, and establishment of increasingly automatable methods for DNA sequencing⁷⁻¹⁰ set the stage for the Human Genome Project (HGP) to begin in 1990 (see also www.nature.com/nature/DNA50). Thanks to the vision of the original planners, and the creativity and determination of a legion of talented scientists who decided to make this project their overarching focus, all of the initial objectives of the HGP have now been achieved at least two years ahead of expectation, and a revolution in biological research has begun.

The project's new research strategies and experimental technologies have generated a steady stream of ever-larger and more complex genomic data sets that have poured into public databases and have transformed the study of virtually all life processes. The genomic approach of technology development and large-scale generation of community resource data sets has introduced an important new dimension into biological and biomedical research. Interwoven advances in genetics, comparative genomics, high-throughput biochemistry and bioinformatics

*Endorsed by the National Advisory Council for Human Genome Research, whose members are Vickie Yates Brown, David R. Burgess, Wylie Burke, Ronald W. Davis, William M. Gelbart, Eric T. Juengst, Bronya J. Keats, Raju Kucherlapati, Richard P. Lifton, Kim J. Nickerson, Maynard V. Olson, Janet D. Rowley, Robert Tepper, Robert H. Waterston and Tadataka Yamada.

are providing biologists with a markedly improved repertoire of research tools that will allow the functioning of organisms in health and disease to be analysed and comprehended at an unprecedented level of molecular detail. Genome sequences, the bounded sets of information that guide biological development and function, lie at the heart of this revolution. In short, genomics has become a central and cohesive discipline of biomedical research.

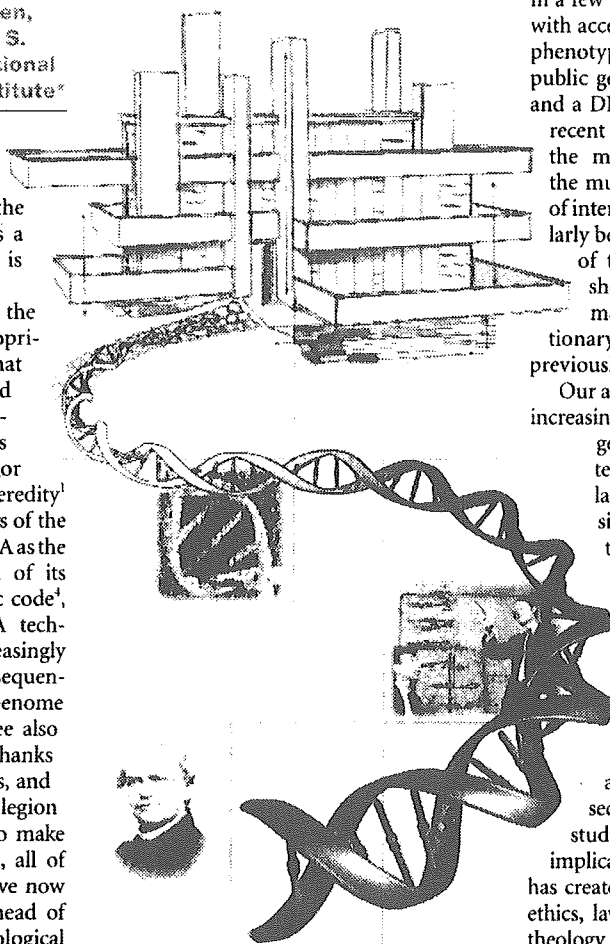
The practical consequences of the emergence of this new field are widely apparent. Identification of the genes responsible for human mendelian diseases, once a herculean task requiring large research teams, many years of hard work, and an uncertain outcome, can now be routinely accomplished

in a few weeks by a single graduate student with access to DNA samples and associated phenotypes, an Internet connection to the public genome databases, a thermal cycler and a DNA-sequencing machine. With the recent publication of a draft sequence of the mouse genome¹¹, identification of the mutations underlying a vast number of interesting mouse phenotypes has similarly been greatly simplified. Comparison of the human and mouse sequences shows that the proportion of the mammalian genome under evolutionary selection is more than twice that previously assumed.

Our ability to explore genome function is increasing in specificity as each subsequent genome is sequenced. Microarray technologies have catapulted many laboratories from studying the expression of one or two genes in a month to studying the expression of tens of thousands of genes in a single afternoon¹². Clinical opportunities for gene-based pre-symptomatic prediction of illness and adverse drug response are emerging at a rapid pace, and the therapeutic promise of genomics has ushered in an exciting phase of expansion and exploration in the commercial sector¹³. The investment of the HGP in studying the ethical, legal and social implications of these scientific advances has created a talented cohort of scholars in ethics, law, social science, clinical research, theology and public policy, and has already resulted in substantial increases in public awareness and the introduction of significant (but still incomplete) protections against misuses such as genetic discrimination (see www.genome.gov/PolicyEthics).

These accomplishments fulfil the expansive vision articulated in the 1988 report of the National Research Council, *Mapping and Sequencing the Human Genome*¹⁴. The successful completion of the HGP this year thus represents an opportunity to look forward and offer a blueprint for the future of genomics research over the next several years.

The vision presented here addresses a different world from that reflected in earlier plans published in 1990, 1993 and 1998 (refs 15-17). Those documents addressed the goals of the 1988 report, defining detailed paths towards the development of genome-



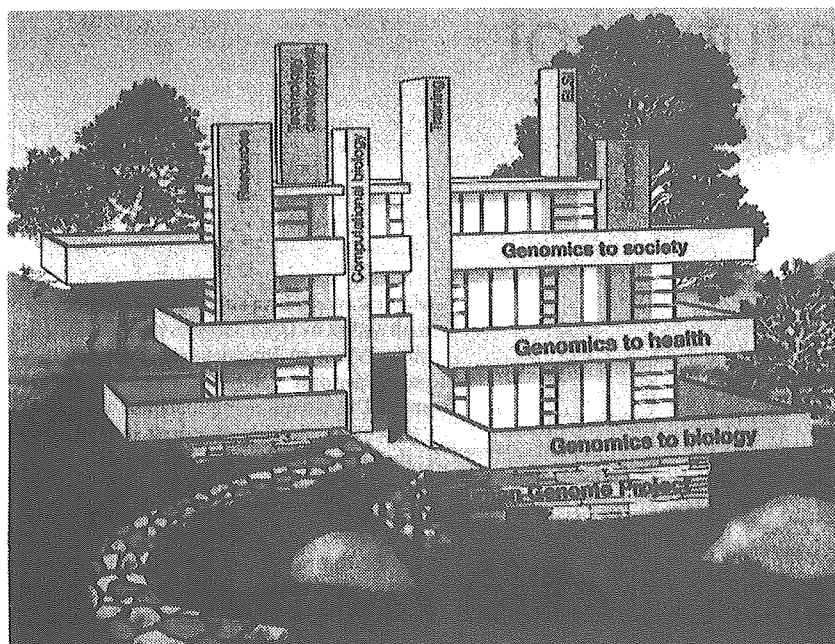


Fig 2 The future of genomics rests on the foundation of the Human Genome Project.

analysis technologies, the physical and genetic mapping of genomes, and the sequencing of model organism genomes and, ultimately, the human genome. Now, with the effective completion of these goals, we offer a broader and still more ambitious vision, appropriate for the true dawning of the genomic era. The challenge is to capitalize on the immense potential of the HGP to improve human health and well-being.

The articulation of a new vision is an opportunity to explore transformative new approaches to achieve health benefits. Although genome-based analysis methods are rapidly permeating biomedical research, the challenge of establishing robust paths

from genomic information to improved human health remains immense. Current efforts to meet this challenge are largely organized around the study of specific diseases, as exemplified by the missions of the disease-oriented institutes at the US National Institutes of Health (NIH, www.nih.gov) and numerous national and international governmental and charitable organizations that support medical research. The National Human Genome Research Institute (NHGRI), in budget terms a rather small (less than 2%) component of the NIH, will work closely with all these organizations in exploring and supporting these biomedical research capabilities. In addition, we envi-

sion a more direct role for both the extramural and intramural programmes of the NHGRI in bringing a genomic approach to the translation of genomic sequence information into health benefits.

The NHGRI brings two unique assets to this challenge. First, it has close ties to a scientific community whose direct role over the past 13 years in bringing about the genomic revolution provides great familiarity with its potential to transform biomedical research. Second, the NHGRI's long-standing mission, to investigate the broadest possible implications of genomics, allows unique flexibility to explore the whole spectrum of human health and disease from the fresh perspective of genome science. By engaging the energetic and interdisciplinary genomics-research community more directly in health-related research and by exploiting the NHGRI's ability to pursue opportunities across all areas of human biology, the institute seeks to participate directly in translating the promises of the HGP into improved human health.

To fully achieve this goal, the NHGRI must also continue in its vigorous support of another of its vital missions — the coupling of its scientific research programme with research into the social consequences of increased availability of new genetic technologies and information. Translating the success of the HGP into medical advances intensifies the need for proactive efforts to ensure that benefits are maximized and harms minimized in the many dimensions of human experience.

A reader's guide

The vision for genomics research detailed here is the outcome of almost two years of intense discussions with hundreds of scientists and members of the public, in more than a dozen workshops and numerous individual consultations (see www.genome.gov/About/Planning). The vision is formulated into three major themes — genomics to biology, genomics to health, and genomics to society — and six crosscutting elements.

We envisage the themes as three floors of a building, firmly resting on the foundation of the HGP (Figure 2). For each theme, we present a series of grand challenges, in the spirit of the proposals put forward for mathematics by David Hilbert at the turn of the twentieth century¹⁸. These grand challenges are intended to be bold, ambitious research targets for the scientific community. Some can be planned on specific timescales, others are less amenable to that level of precision. We list the grand challenges in an order that makes logical sense, not representing priority. The challenges are broad in sweep, not parochial — some can be led by the NHGRI alone, whereas others will be best pursued in partnership with other organizations. Below, we clarify areas in which the NHGRI intends to play a leading role.

BOX 1 Resources



One of the key and distinctive objectives of the Human Genome Project (HGP) has been the generation of large, publicly available, comprehensive sets of reagents and data (scientific resources or 'infrastructure') that, along with other new, powerful technologies, comprise a toolkit for genomics-based research. Genomic maps and sequences are the most obvious examples. Others include databases of sequence variation, clone libraries and collections of anonymous cell lines. The continued generation of such resources is critical, in particular:

- ◆ Genoma sequences of key mammals, vertebrates, chordates, and invertebrates
- ◆ Comprehensive reference sets of coding sequences from key species in various formats, for example, full-length cDNA sequences and corresponding clones, oligonucleotide primers, and microarrays

- ◆ Comprehensive collections of knockouts and knock-downs of all genes in selected animals to accelerate the development of models of disease
- ◆ Comprehensive reference sets of proteins from key species in various formats, for example in expression vectors, with affinity tags and spotted onto protein chips
- ◆ Comprehensive sets of protein affinity reagents
- ◆ Databases that integrate sequences with curated information and other large data sets, as well as tools for effective mining of the data
- ◆ Cohort populations for studies designed to identify genetic contributors to health and to assess the effect of individual gene variants on disease risk, including a 'healthy' cohort
- ◆ Large libraries of small molecules, together with robotic methods to screen them and access to medicinal chemistry for follow-up, to provide investigators easy and affordable access to these tools

The six critically important crosscutting elements are relevant to all three thematic areas. They are: resources (Box 1); technology development (Box 2); computational biology (Box 3); training (Box 4); ethical, legal and social implications (ELSI, Box 5); and education (Box 6). We also stress the critical importance of early, unfettered access to genomic data in achieving maximum public benefit. Finally, we propose a series of 'quantum leaps', achievements that would lead to substantial advances in genomics research and its applications to medicine. Some of these may seem overly bold, but no laws of physics need to be violated to achieve them. Such leaps would have profound implications, just as the dreams of the mid-1980s about the complete sequence of the human genome have been realized in the accomplishments now being celebrated.

I Genomics to biology

Elucidating the structure and function of genomes

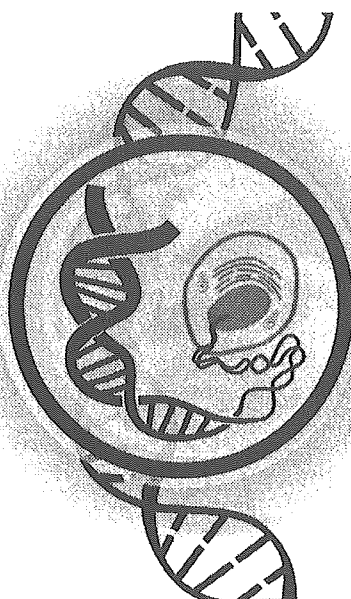
The broadly available genome sequences of human and a select set of additional organisms represent foundational information for biology and biomedicine. Embedded within this as-yet poorly understood code are the genetic instructions for the entire repertoire of cellular components, knowledge of which is needed to unravel the complexities of biological systems. Elucidating the structure of genomes and identifying the function of the myriad encoded elements will allow connections to be made between genomics and biology and will, in turn, accelerate the exploration of all realms of the biological sciences.

For this, new conceptual and technological approaches will be needed to:

- ◆ Develop a comprehensive and comprehensible catalogue of all of the components encoded in the human genome.
- ◆ Determine how the genome-encoded components function in an integrated manner to perform cellular and organismal functions.
- ◆ Understand how genomes change and take on new functional roles.

Grand Challenge I-1 Comprehensively identify the structural and functional components encoded in the human genome

Although DNA is relatively simple and well understood chemically, the human genome's structure is extraordinarily complex and its function is poorly understood. Only 1–2% of its bases encode proteins⁷, and the full complement of protein-coding sequences still remains to be established. A roughly equivalent amount of the non-coding portion of the genome is under active selection¹¹, suggesting that it is also functionally important, yet vanishingly little is known about it. It



probably contains the bulk of the regulatory information controlling the expression of the approximately 30,000 protein-coding genes, and myriad other functional elements, such as non-protein-coding genes and the sequence determinants of chromosome dynamics. Even less is known about the function of the roughly half of the genome that consists of highly repetitive sequences or of the remaining non-coding, non-repetitive DNA.

The next phase of genomics is to catalogue, characterize and comprehend the entire set of functional elements encoded in the human and other genomes. Compiling this genome 'parts list' will be an immense challenge. Well-known classes of functional

elements, such as protein-coding sequences, still cannot be accurately predicted from sequence information alone. Other types of known functional sequences, such as genetic regulatory elements, are even less well understood; undoubtedly new types remain to be defined, so we must be ready to investigate novel, perhaps unexpected, ways in which DNA sequence can confer function. Similarly, a better understanding of epigenetic changes (for example, methylation and chromatin remodelling) is needed to comprehend the full repertoire of ways in which DNA can encode information.

Comparison of genome sequences from evolutionarily diverse species has emerged as a powerful tool for identifying functionally important genomic elements. Initial analyses of available vertebrate genome sequences^{7,11,19} have revealed many previously undiscovered protein-coding sequences. Mammal-to-mammal sequence comparisons have revealed large numbers of homologies in non-coding regions¹¹, few of which can be defined in functional terms. Further comparisons of sequences derived from multiple species, especially those occupying distinct evolutionary positions, will lead to significant refinements in our understanding of the functional importance of conserved sequences²⁰. Thus, the generation of additional genome sequences from several well-chosen species is crucial to the functional characterization of the human genome (Box 1). The generation of such large sequence data sets will benefit from further advances in sequencing technology that yield significant cost reductions (Box 2). The study of sequence variation within species will also be important in defining the functional nature of some sequences (see Grand Challenge I-3).

BOX 2 Technology development



The Human Genome Project was aided by several 'breakthrough' technological developments, including Sanger DNA sequencing and its automation, DNA-based genetic

markers, large-insert cloning systems and the polymerase chain reaction. During the project, these methods were scaled up and made more efficient by 'evolutionary' advances, such as automation and miniaturization. New technologies, including capillary-based sequencing and methods for genotyping single-nucleotide polymorphisms, have recently been introduced, leading to further improvements in capacity for genomic analyses. Even newer approaches, such as nanotechnology and microfluidics, are being developed, and hold great promise, but further advances are still needed. Some examples are:

- ◆ Sequencing and genotyping technologies to reduce costs further and increase access to a wider range of investigators
- ◆ Identification and validation of functional

elements that do not encode protein

- ◆ *In vivo*, real-time monitoring of gene expression and the localization, specificity, modification and activity/kinetics of gene products in all relevant cell types
- ◆ Modulation of expression of all gene products using, for example, large-scale mutagenesis, small-molecule inhibitors and knock-down approaches (such as RNA-mediated inhibition)
- ◆ Monitoring of the absolute abundance of any protein (including membrane proteins, proteins at low abundance and all modified forms) in any cell
- ◆ Improved imaging methods that allow non-invasive molecular phenotyping
- ◆ Correlating genetic variation to human health and disease using haplotype information or comprehensive variation information
- ◆ Laboratory-based phenotyping, including the use of protein affinity reagents, proteomic approaches and analysis of gene expression
- ◆ Linking molecular profiles to biology, particularly pathway biology to disease

feature

Effective identification and analysis of functional genomic elements will require increasingly powerful computational capabilities, including new approaches for tackling ever-growing and increasingly complex data sets and a suitably robust computational infrastructure for housing, accessing and analysing those data sets (Box 3). In parallel, investigators will need to become increasingly adept in dealing with this treasure trove of new information (Box 4). As a better understanding of genome function is gained, refined computational tools for *de novo* prediction of the identity and behaviour of functional elements should emerge²¹.

Complementing the computational detection of functional elements will be the generation of additional experimental data by high-throughput methodologies. One example is the production of full-length complementary DNA (cDNA) sequences (see, for example, www.mgc.nci.nih.gov and www.fruitfly.org/EST/full.shtml). Major challenges inherent in programmes to discover genes are the experimental identification and validation of alternate splice forms and messenger RNAs expressed in a highly restricted fashion. Even more challenging is the experimental validation of functional elements that do not encode protein (for example, regulatory regions and non-coding RNA sequences). High-throughput approaches to identify them (Box 2) will be needed to generate the experimental data that will be necessary to develop, confirm and enhance computational methods for detecting functional elements in genomes.

Because current technologies cannot yet identify all functional elements, there is a need for a phased approach in which new methodologies are developed, tested on a pilot scale and finally applied to the



entire human genome. Along these lines, the NHGRI recently launched the Encyclopedia of DNA Elements (ENCODE) Project (www.genome.gov/Pages/Research/ENCODE) to identify all the functional elements in the human genome. In a pilot project, systematic strategies for identifying all functionally important genomic elements will be developed and tested using a selected 1% of the human genome. Parallel projects involving well-studied model organisms, for example, yeast, nematode and fruitfly, are ongoing. The lessons learned will serve as the basis for implementing a broader programme for the entire human genome.

Grand Challenge 1-2 Elucidate the organization of genetic networks and protein pathways and establish how they

contribute to cellular and organismal phenotypes

Genes and gene products do not function independently, but participate in complex, interconnected pathways, networks and molecular systems that, taken together, give rise to the workings of cells, tissues, organs and organisms. Defining these systems and determining their properties and interactions is crucial to understanding how biological systems function. Yet these systems are far more complex than any problem that molecular biology, genetics or genomics has yet approached. On the basis of previous experience, one effective path will begin with the study of relatively simple model organisms²², such as bacteria and yeast, and then extend the early findings to more complex organisms, such as mouse and human. Alternatively, focusing on a few well-characterized systems in mammals will be a useful test of the approach (see, for example, www.signaling-gateway.org).

Understanding biological pathways, networks and molecular systems will require information from several levels. At the genetic level, the architecture of regulatory interactions will need to be identified in different cell types, requiring, among other things, methods for simultaneously monitoring the expression of all genes in a cell¹². At the gene-product level, similar techniques that allow *in vivo*, real-time measurement of protein expression, localization, modification and activity/kinetics will be needed (Box 2). It will be important to develop, refine and scale up techniques that modulate gene expression, such as conventional gene-knockout methods²³, newer knock-down approaches²⁴ and small-molecule inhibitors²⁵ to establish the temporal and cellular expression pattern of individual proteins and to determine the functions of those proteins. This is a key first step towards assigning all genes and their products to functional pathways.

The ability to monitor all proteins in a cell simultaneously would profoundly improve our ability to understand protein pathways and systems biology. A critical step towards gaining a complete understanding of systems biology will be to take an accurate census of the proteins present in particular cell types under different physiological conditions. This is becoming possible in some model systems, such as microorganisms²⁶. It will be a major challenge to catalogue proteins present in low abundance or in membranes. Determining the absolute abundance of each protein, including all modified forms, will be an important next step. A complete interaction map of the proteins in a cell, and their cellular locations, will serve as an atlas for the biological and medical explorations of cellular metabolism²⁷ (see www.nrcam.uchc.edu, for example). These and other related areas constitute the developing field of proteomics.

BOX 3 Computational biology



Computational methods have become intrinsic to modern biological research, and their importance can only increase as large-scale methods for data generation become more prominent, as the amount and complexity of the data increase, and as the questions being addressed become more sophisticated. All future biomedical research will integrate computational and experimental components. New computational capabilities will enable the generation of hypotheses and stimulate the development of experimental approaches to test them. The resulting experimental data will, in turn, be used to generate more refined models that will improve overall understanding and increase opportunities for application to disease. The areas of computational biology critical to the future of genomics research include:

◆ New approaches to solving problems, such as the identification of different features in a DNA sequence, the analysis of gene expression and

regulation, the elucidation of protein structure and protein-protein interactions, the determination of the relationship between genotype and phenotype, and the identification of the patterns of genetic variation in populations and the processes that produced those patterns

◆ Reusable software modules to facilitate interoperability

◆ Methods to elucidate the effects of environmental (non-genetic) factors and of gene-environment interactions on health and disease

◆ New ontologies to describe different data types

◆ Improved database technologies to facilitate the integration and visualization of different data types, for example, information about pathways, protein structure, gene variation, chemical inhibition and clinical information/phenotypes

◆ Improved knowledge management systems and the standardization of data sets to allow the coalescence of knowledge across disciplines

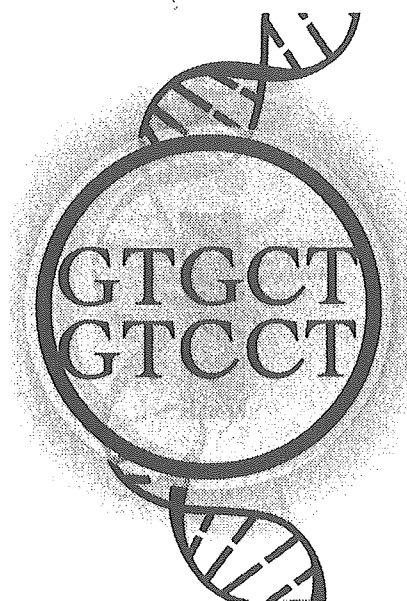
Establishing a true understanding of how organized molecular pathways and networks give rise to normal and pathological cellular and organismal phenotypes will require more than large, experimentally derived data sets. Once again, computational investigation will be essential (Box 3), and there will be a greatly increased need for the collection, storage and display of the data in robust databases. By modelling specific pathways and networks, predicting how they affect phenotype, testing hypotheses derived from these models and refining the models based on new experimental data, it should be possible to understand more completely the difference between a 'bag of molecules' and a functioning biological system.

Grand Challenge I-3 Develop a detailed understanding of the heritable variation in the human genome

Genetics seeks to correlate variation in DNA sequence with phenotypic differences (traits). The greatest advances in human genetics have been made for traits associated with variation in a single gene. But most phenotypes, including common diseases and variable responses to pharmacological agents, have a more complex origin, involving the interplay between multiple genetic factors (genes and their products) and non-genetic factors (environmental influences). Unravelling such complexity will require both a complete description of the genetic variation in the human genome and the development of analytical tools for using that information to understand the genetic basis of disease.

Establishing a catalogue of all common variants in the human population, including single-nucleotide polymorphisms (SNPs), small deletions and insertions, and other structural differences, began in earnest several years ago. Many SNPs have been identified²⁸, and most are publicly available (www.ncbi.nlm.nih.gov/SNP). A public collaboration, the International HapMap Project (www.genome.gov/Pages/Research/HapMap), was formed in 2002 to characterize the patterns of linkage disequilibrium and haplotypes across the human genome and to identify subsets of SNPs that capture most of the information about these patterns of genetic variation to enable large-scale genetic association studies. To reach fruition, such studies need more robust experimental (Box 2) and computational (Box 3) methods that use this new knowledge of human haplotype structure²⁹.

A comprehensive understanding of genetic variation, both in humans and in model organisms, would facilitate studies to establish relationships between genotype and biological function. The study of particular variants and how they affect the functioning of specific proteins and protein pathways will yield important new insights about physiological



processes in normal and disease states. An enhanced ability to incorporate information about genetic variation into human genetic studies would usher in a new era for investigating the genetic bases of human disease and drug response (see Grand Challenge II-1).

Grand Challenge I-4 Understand evolutionary variation across species and the mechanisms underlying it

The genome is a dynamic structure, continually subjected to modification by the forces of evolution. The genomic variation seen in humans represents only a small glimpse through the larger window of evolution, where hundreds of millions of years of trial-and-error efforts have created today's bio-

sphere of animal, plant and microbial species. A complete elucidation of genome function requires a parallel understanding of the sequence differences across species and the fundamental processes that have sculpted their genomes into the modern-day forms.

The study of inter-species sequence comparisons is important for identifying functional elements in the genome (see Grand Challenge I-1). Beyond this illuminating role, determining the sequence differences between species will provide insight into the distinct anatomical, physiological and developmental features of different organisms, will help to define the genetic basis for speciation and will facilitate the characterization of mutational processes. This last point deserves particular attention, because mutation both drives long-term evolutionary change and is the underlying cause of inherited disease. The recent finding that mutation rates vary widely across the mammalian genome¹¹ raises numerous questions about the molecular basis for these evolutionary changes. At present, our understanding of DNA mutation and repair, including the important role of environmental factors, is limited.

Genomics will provide the ability to substantively advance insight into evolutionary variation, which will, in turn, yield new insights into the dynamic nature of genomes in a broader evolutionary framework.

Grand Challenge I-5 Develop policy options that facilitate the widespread use of genome information in both research and clinical settings

Realization of the opportunities provided by genomics depends on effective access to the

BOX 4 Training



Meeting the scientific, medical and social/ethical challenges now facing genomics will require scientists, clinicians and scholars with the skills to understand biological systems and to use that information effectively for the benefit of humankind. Adequate training capacity will be required to address the following needs:

◆ **Computational skills** As biomedical research is becoming increasingly data intensive, computational capability is increasingly becoming a critical skill.

◆ **Interdisciplinary skills** Although a good start has been made, expanded interactions will be required between the sciences (biology, computer science, physics, mathematics, statistics, chemistry and engineering), between the basic and the clinical sciences, and between the life sciences, the social sciences and the humanities. Such interactions will be needed at the individual level (scientists, clinicians and scholars will need to be able to bring relevant issues, concerns and capabilities from different disciplines to bear on

their specific research efforts), at a collaborative level (researchers will need to be able to participate effectively in interdisciplinary research collaborations that bring biology together with many other disciplines) and at the disciplinary level (new disciplines will need to emerge at the interfaces between the traditional disciplines).

◆ **Different perspectives** Individuals from minority or disadvantaged populations are significantly under-represented as both researchers and participants in genomics research. This regrettable circumstance deprives the field of the best and brightest from all backgrounds, narrows the field of questions asked, can lessen sensitivity to cultural concerns in implementing research protocols, and compromises the overall effectiveness of the research. Genomics can learn from successful efforts in training individuals from under-represented populations in other areas of science and health (see, for example, www.genome.gov/Pages/Grants/Policy/ActionPlanGuide).

feature

information (such as data about genes, gene variants, haplotypes, protein structures, small molecules and computational models) by a wide range of potential users, including researchers, commercial enterprises, health-care providers, patients and the public. Researchers themselves need maximum access to the data as soon as possible (see 'Data release', below). Use of the information for the development of therapeutic and other products necessarily entails consideration of the complex issues of intellectual property (for example, patenting and licensing) and commercialization. The intellectual property practices, laws and regulations that affect genomics must adhere to the principle of maximizing public benefit, but must also be consistent with more general and longer-established intellectual property principles. Further, because genome research is global, international treaties, laws, regulations, practices, belief systems and cultures also come into play.

Without commercialization, most diagnostic and therapeutic advances will not reach the clinical setting, where they can benefit patients. Thus, we need to develop policy options for data access and for patenting, licensing and other intellectual property issues to facilitate the dissemination of genomics data.

II Genomics to health

Translating genome-based knowledge into health benefits

The sequencing of the human genome, along with other recent and expected achievements in genomics, provides an unparalleled opportunity to advance our understanding of the role of genetic factors in human health and disease, to allow more precise definition of the non-genetic factors involved, and to apply this insight rapidly to the prevention, diagnosis and



treatment of disease. The report by the US National Research Council that originally envisioned the HGP was explicit in its expectation that the human genome sequence would lead to improvements in human health, and subsequent five-year plans reaffirmed this view¹⁵⁻¹⁷. But how this will happen has been less clearly articulated. With the completion of the original goals of the HGP, the time is right to develop and apply large-scale genomic strategies to empower improvements in human health, while anticipating and avoiding potential harm.

Such strategies should enable the research community to achieve the following:

- ◆ Identify genes and pathways with a role in health and disease, and determine how they interact with environmental factors.
- ◆ Develop, evaluate and apply genome-

based diagnostic methods for the prediction of susceptibility to disease, the prediction of drug response, the early detection of illness and the accurate molecular classification of disease.

- ◆ Develop and deploy methods that catalyse the translation of genomic information into therapeutic advances.

Grand Challenge II-1 Develop robust

strategies for identifying the genetic contributions to disease and drug response. For common diseases, the interplay of multiple genes and multiple non-genetic factors, not a single allele, usually dictates disease susceptibility and response to treatments. Deciphering the role of genes in human health and disease is a formidable problem for many reasons, including impediments to defining biologically valid phenotypes, challenges in identifying and quantifying environmental exposures, technological obstacles to generating sufficient and useful genotypic information, and the difficulties of studying humans. Yet this problem can be solved. Vigorous development of cross-cutting genomic tools to catalyse advances in understanding the genetics of common disease and in pharmacogenomics is needed. Prominent among these will be a detailed haplotype map of the human genome (see Grand Challenge I-3) that can be used for whole-genome association studies of all diseases in all populations, as well as further advances in sequencing and genotyping technology to make such studies feasible (see 'Quantum leaps', below).

More efficient strategies for detecting rare alleles involved in common disease are also needed, as the hypothesis that alleles that increase risk for common diseases are themselves common³⁰ will probably not be universally true. Computational and experimental methods to detect gene-gene and gene-environment interactions, as well as methods allowing interfacing of a variety of relevant databases, are also required (Box 3). By obtaining unbiased assessments of the relative disease risk that particular gene variants contribute, a large longitudinal population-based cohort study, with collection of extensive clinical information and ongoing follow-up, would be profoundly valuable to the study of all common diseases (Box 1). Already, such projects as the UK Biobank (www.ukbiobank.ac.uk), the Marshfield Clinic's Personalized Medicine Research Project (www.mfldclin.edu/pmrp) and the Estonian Genome Project (www.geenivaramu.ee) seek to provide such resources. But if the multiple population groups in the United States and elsewhere in the world are to benefit fully and fairly from such research (see Grand Challenge II-6), a large population-based cohort study that includes full representation of minority populations is also needed.

BOX 5 Ethical, legal and social implications (ELSI)

Today's genomics research and applications rest on more than a decade of valuable investigation into their ethical, legal and social implications. As the application of genomics to health increases along with its social impact, it becomes ever more important to expand on this work. There is an increasing need for focused ELSI research that directly informs policies and practices. One can envisage a flowering of 'translational ELSI research' that builds on the knowledge gained from prior and forthcoming 'basic ELSI research', which would provide knowledge for direct use by researchers, clinicians, policy-makers and the public. Examples include:

- ◆ The development of models of genomics research that use attention to these ELSI issues

for enhancing the research, rather than viewing such issues as impediments

- ◆ The continued development of appropriate and effective genomics research methods and policies that promote the highest levels of science and of protecting human subjects
- ◆ The establishment of crosscutting tools, analogous to the publicly accessible genomic maps and sequence databases that have accelerated other genomics research (examples of such tools might include searchable databases of genomic legislation and policies from around the world, or studies of ELSI aspects of introducing clinical genetic tests)
- ◆ The evaluation of new genetic and genomic tests and technologies, and effective oversight of their implementation, to ensure that only those with confirmed clinical validity are used for patient care