

Understanding human diversity

ヒトの多様性をもっとよく知るために

David B. Goldstein & Gianpiero L. Cavalleri



ヒトにおける遺伝的多様性に関するぼう大なデータ集積の初版ともいうべきものが、このほど完成した。今後の課題は、これらのデータを読み解き、ヒト個人間の差異の健康への影響を解明することだが、それには多くの時間が必要だろう。

Nature Vol.437(1241-1242)/27 October 2005

小説家サマセット・モームが医学生だったころ、1本の神経があるはずの場所に見つけられずにいると、解剖学の教官が隠れていた神経を見つけだし、「この世界では教科書通りの正常な例など、めったにないんだよ」と教えてくれたという¹。正常な型といえるヒトなど現実には存在しない、その主な理由の1つは、Nature10月27日号に発表された国際ハップマップコンソーシアムの論文²を見ればよくわかる。

ヒトゲノムには「多型」がおおよそ1000万か所ある。ここでいう多型とは、集団内に保有される遺伝的な変動であって、頻度の低い遺伝子の型でも少なくとも100例に1例の割合で存在するものとする。血縁関係のない2人の人間どうしだと、遺伝的なちがいは数百万か所にのぼり、そのせいで外見も振る舞いも異なってくる。こうした遺伝的な多様性は、われわれ人類の進化史の貴重な遺産ではあるが、その代償も小さくない。遺伝情報および遺伝機構は、恵み深く興味深くもある形にわれわれの差異を作り出すとともに、われわれの健康にも影響を及ぼすからである。

現代の遺伝学研究は、嚢胞性線維症やテイ・サックス病といった単純な家族性遺伝疾患に結びつく遺伝子異常を突き止めることには、すでに大きな成功をおさめている。しかし、こうした異常は比較的まれである。たとえば、がんや心血管疾患や神経変性疾患といったわれわれの多くがかかる遺伝疾患は、もって複合的である。それどころか、これらの頻度の高いありふれた遺伝疾患については、ある厳然たる必然性がみられる。つまり、若くして死にい

たる病に対する治療法が向上していけば、認知症（痴呆）など高齢で発症する疾患が急増していくとみられるのだ。

これらのありふれた疾患は複合的な性質をもつため、今までは遺伝解析による解明がほとんど進んでいなかった。しかし、遺伝的なものがこうした疾患に重要な役割を果たしていることには疑問の余地がない。これは数多くの研究によって実証されている。たとえば、ある集団内のごくありふれた疾患の多様な型のうち40%以上、また、統合失調症など一部疾患のうち70%以上が遺伝的多型から説明づけられることが報告されている³⁻⁵。ややこしいことに、遺伝子の多型によって、ヒトはありふれた疾患にかかりやすくなるが、必ずなるわけではない。また、こうした多型に他の遺伝子の多型や環境と組み合わせることで疾患が起りやすくなる。

1000万か所もの多型のどれが疾患に影響するのかを見極めるには、どうしたらよいのだろうか。幸いなことに、1000万もの部位すべてを直接分析して疾患とそれらとの相関を判定する必要はない。なぜなら、ヒトゲノムにある多型は多くの場合、互いに連動し相関しているからだ。1個の変異が生じたとき、この変異は同じ染色体上にある特定の多型と連動して振る舞う（このように、まとまって遺伝する多型群を「ハプロタイプ」という）。この理由や他の理由から、多型間には統計的に強い相関性がしばしばみられ、そのため1本の染色体上のある部位に特定の多型が存在すれば、別の部位にも特定の多型が存在することを予測できる。すなわち標識用の「タグ」として使えるのである。

ハップマップコンソーシアムの目指した第1のゴールは、多型間のこうした相関性を見つけること（プロジェクト名もここからきている）であり、これはみごとに成功をおさめた。プロジェクトの第1期では、今回報告されたように²、100万か所以上の塩基多型（SNP; スニップと読む）について、4種類のヒト集団の代表者グループの遺伝的構成（遺伝子型）に関するデータを編集・解析した（SNPは、挿入や欠失などとは対照的に、ヒトゲノムにみられる最もありふれた多型の種類の1つである）。その設定目標は、塩基配列上の5000塩基ごとにSNPを1か所ずつ確実に配置していくことだった。第2期ではSNPの被覆率を大幅に高める予定だが、まだ完了していない。

ハップマップ解析グループは、概念的に簡便で信頼できるタグ選別法を用いて、ありふれた多型でゲノム全体の地図を作るのに必要なタグ数が、多型部位総数の10分の1たらずでよいだろうと見積もった。もっと精妙で洗練された方法であれば、さらに良い結果が得られそうである。

ハップマッププロジェクトは多型の相関性を記載しただけでなく、他にも成果をおさめた。3年前の段階では、わかっていた多型の数は170万未満だった。現在その数は、同プロジェクトと関連研究のおかげで800万を超えた。多型の大部分がわかれば、形勢はおおいに有利となる。今やわれわれは、これらのデータに生物情報学のツールを駆使して、考えられる機能の面から多型に優先順位をつけられるようになった。たとえば、タンパク質のアミノ酸配列を変化させる多型や、機能的に重要な（つまり種を超えて保存度が高い）染色体領域に位置する多型、もしくは他のゲノム解析基準を使うのが適切とみられる多型などを重点的に調べることができる。

この成果は、ヒトの疾患や治療に対する反応の個人差を調べるうえで、どんな意味をもつのだろうか。ハップマッププロジェクト以前に、ある医薬の分子標的に一般に起こる高い遺伝的多型が、患者の反応に影響するかどうかを判定しようとしたとする。その場合、十分な数の多型を見つけ出すためには、調査対象の集団を代表する人々で、問題の遺伝子の塩基配列を再度読み取らねばならないことになる。4年前のこと、われわれのグループは抗てんかん薬の標的をコードするSCN1Aという遺伝子について、実際にこの作業を行った。この一般的な多型を突き止めて適切なタグ部位を見つけるのに、2年も要した。今なら、ハップマップのデータを利用して、単純なコンピューター・アルゴリズムを使うだけで同じことが数分でできてしまう（次ページの「ハプロタイプ地図作成の手順」参照）。

ハップマップデータを使ったゲノム全体の相関解析がどれほどの威力をもつかは、同じ号に掲載されたもう1

つの成果（p.1365）を見ればよくわかる⁶。Cheungたちは、家系調査で遺伝的多型によって発現が影響を受けていることがすでにわかっている、27個の遺伝子を解析し直した。そして、これらの遺伝子の多くがゲノム全体の相関解析によって検出されることや、1つの例ではハップマップのデータのおかげで、*in vitro*解析により遺伝子発現を変化させることが明らかである多型の同定を容易にすることを報告している。

次に重大な問題は、ヒトの健康とのかかわりである。これらの解析ツールはまったく新しいものなので、ありふれたヒト疾患への遺伝的な関与をこれまで突き止められなかったからといって、今後もできないとあきらめてかかることはない。しかし同じく、成功するという保証もない。ありふれた疾患に遺伝的特性が関与していることがわかっても、どんな種類の遺伝的多型が原因なのかはわからない。もし、そうした多型が一般的なもの（つまり1000万か所のうちの1つ）なら、新しい解析ツールキットで、疾患と関連する遺伝的多型の特定は大きく加速するはずである。しかし、もし原因となる多型の頻度がこれより低ければ、見つけるのはずっとむずかしくなるだろう。同じように、他の種類の多型（たとえば反復配列要素や挿入や欠失）がSNPのタグによってどの程度うまく把握できるかもわからない。まだ十分な取り組みがなされていないもう1つのややこしい問題は、ハップマッププロジェクトで調べられた4種類の集団が、他のヒト集団の多型をどこまで反映しているかという懸念である。

またこれも認めねばならない現実なのだが、ありふれたヒト疾患のリスク因子を突き止めれば健康状態が改善されることにつながると直接示した証拠は、今のところほとんどない。リスク因子がはっきり突き止められた希少な例でも、それが治療や予防に役立っていないのが普通である。たとえば、APOE遺伝子の多型の1つは老年性アルツハイマー病の強力な発症予測因子だが、この疾患を防ぐために生活スタイルや食事内容をどう変えたらよいのかはわかっていない。最も期待されるのは、リスク因子から治療の新しい道筋がみえてくることだろうが、この面でも成功例はほとんどない。

もっとはっきりいえそうなのは、患者の治療への反応にかかわる遺伝的予測因子の同定が、近い将来の臨床にかかわってくるだろうということだ。多型はこうした反応に大きな影響をもつ可能性があり、これらの影響を見極めることで、代替となる治療法がみえてくるかもしれない。例として、数十年前に導入された薬と新世代の薬を比較した場合でも、異なる抗精神病薬の間で総体的な有効性にはほとんど差がない⁷。ところが、患者の反応は薬剤の種類によって個人差が非常に大きい。たとえば、

新しいほうの薬剤がひどい体重増加を起こすかどうかや、古いほうの薬剤が重度の運動障害を起こすかどうかを予測できるような遺伝子多型を突き止められたら、それに応じて治療の選択肢を考えることができるだろう。ハップマッププロジェクトは、こうした多型の研究を飛躍的におし進めてくれそうだ。

現時点のゲノム科学は、暗中模索してもがく思春期の段階だといえるかもしれない。現在のゲノム解析ツールキットの威力は驚異的である。この数年で、ゲノムの観点から（つまり特定の遺伝子ではなく、ゲノム全体を対象として）特徴をとらえた例がほとんど何もない状態から、多数の生物の全ゲノム塩基配列が得られるまでになり、そして今回、ヒトのありふれた遺伝的な差異を記したほぼ完全なカタログまで得られたのだから。ただし、すぐれた技術能力それ自体は、科学の成熟の証とはなら

ない。ゲノム解析研究の次の段階では、生物学的理解と臨床的有用性の両方にもっと目を向ける必要がある。ゲノム研究者たちは今こそ、関心を技術から応用へと転ずべきである。 ■

デューク大学 (米)

David B. Goldstein & Gianpiero L. Cavalleri

1. Meyers, J. *Somerset Maugham: A Life* (Knopf, New York, 2004).
2. The International HapMap Consortium *Nature* **437**, 1299-1320 (2005); www.hapmap.org
3. Pedersen, N. L., Posner, S. F. & Gatz, M. *Am. J. Med. Genet.* **105**, 724-728 (2001).
4. Sullivan, P. F., Kendler, K. S. & Neale, M. C. *Arch. Gen. Psychiat.* **60**, 1187-1192 (2003).
5. Zdravkovic, S. *et al. J. Intern. Med.* **252**, 247-254 (2002).
6. Cheung, V. G. *et al. Nature* **437**, 1365-1369 (2005).
7. Lieberman, J. A. *et al. N. Engl. J. Med.* **353**, 1209-1223 (2005).

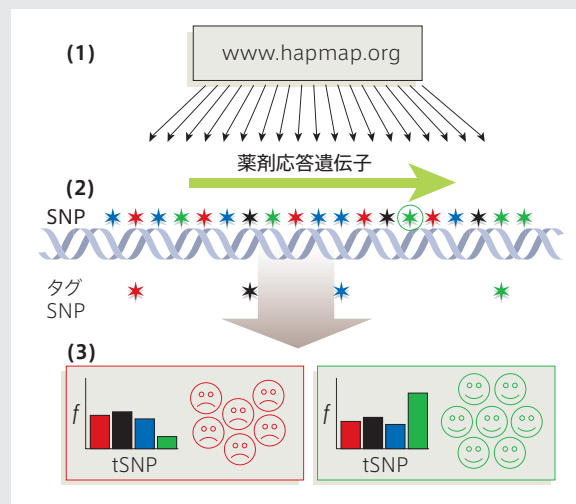
ハプロタイプ地図作成の手順

ヒトの遺伝的多型と健康状態の関係に関心を寄せる研究者たちにとって、ハップマッププロジェクトはまさに天恵である。このデータのおかげで、遺伝子やゲノム領域に対応した信頼できる標識用の一塩基多型（タグ SNP）を、ほんの数分で選び出せる。

(1) 遺伝子型データをダウンロードする。 調べたい1個の遺伝子（または領域）のデータは、ハップマップのウェブサイトから無料で自由にアクセスして入手できる。

(2) タグ SNP を選別するために、データ内の SNP どうしの相関性を解析する。 図では、互いに強く相関する SNP を同じ色で表してある。全 SNP からなるセットは、タグ SNP からなるサブセットに圧縮・省略化できる。図ではタグ SNP を各色で描いてある。

関係するサンプルでは平均すると、1個の SNP には完全に相関する SNP が他に3~10個（この数は集団によってちがう）ある（ただし、同じ集団からとったもう1つのサンプルでもこうなるとは限らない）。したがって、こうした相関する SNP のセットから1つの SNP だけを残して他をすべて除外することで、塩基判別の必要な SNP の数を大きく減らすことができる。もっと無駄なく洗練された方法をとることで、費用対効果はより高くなるだろうし、さまざまな統計的手法もすでに提案されている。総合的にみてどんな方法が最良かは、まだ統一見解が得られていないが、やがて1つか複数の方法が浮上ってきて、全ゲノムを効率よく解析できるタグ SNP セットが得られるだろう。



(3) 遺伝的な相関性。 タグ SNP は、ある集団サンプルでの遺伝子型を表しており、その集団内では、ある薬への反応のよし悪しといった関心がもたれる一部の形質に個人差がある。特定の反応と相関するタグ SNP があるということは、そのタグ SNP と相関する SNP の1つ

がその反応に影響を及ぼしているということである。図の場合、緑色の SNP の1つ（丸を付けたもの）が問題の薬への患者の反応をよくしており、結果として該当するタグ SNP と相関性がある。

D.B.G. & G.L.C.