

【第96回生涯教育講座】

生活習慣病遺伝素因解析の現状

なび か とおる
並 河 徹

キーワード：生活習慣病，遺伝解析，遺伝子，次世代シーケンス

1. はじめに

1980 - 90年代に行われ、2001年に最初のヒトゲノム配列を発表したヒトゲノム計画が終了し、その後、ゲノム上の人種差、民族差の解析が進められてきた。その成果として、ヒトのゲノム上に今や150万個を超える一塩基置換多型 (single-nucleotide polymorphisms, SNP) が同定されている。さらにシーケンス技術の進歩は目覚ましく、今やゲノムの「民族差」から「個人差」へと解析のターゲットが移ってきている。すでに数十万円の費用と数日の期間で誰でも自分のゲノム配列情報を解析できる時代となりいわゆる「1000ドルシーケンス」はほぼ達成できたと考えてよいであろう。

これに対して、これらの技術開発の目的であった生活習慣病の遺伝素因説明は、どちらかといえはあまり進んでいないように見える。

本稿では、生活習慣病の遺伝素因解析の歴史をたどり、現状と今後の展望についてまとめてみたい。

2. 遺伝解析の歴史

a) 単因子遺伝性疾患の遺伝解析—連鎖解析の応用
遺伝様式がはっきりしていて、ひとつの遺伝子の異常で発症することが遺伝学的解析で分かっている病気のうち、その原因遺伝子が不明のものがいくつもあった。たとえば筋ジストロフィーやハンチントン舞踏病などの神経筋変性疾患や、家族性大腸ポリポーシスのような家族性腫瘍疾患である。これらは、フェニルケトン尿症などの先天性代謝異常疾患に比べて発症機序に関する情報が乏しく、したがって発症の鍵となる蛋白質や酵素がわからなかったため、原因遺伝子の探索が困難であった。

1978年春にアメリカ合衆国のユタ州アルタで開かれた小さな研究会が歴史的な転換点となった。この研究会には分子生物学と集団遺伝学の若い研究者が集まっていた。その場で上記のような原因遺伝子不明の遺伝性疾患の解析に、RFLPというヒトゲノムの多型性を利用したマーカーを応用すれば、遺伝子を明らかに出来るのではないかと、というアイデアが生まれたのである¹⁾。後になって、「ポジショナル・クローニング」と呼ばれるようになる、「連鎖解析のヒト遺伝性疾患への応用」というアイデアの生まれた瞬間である。

連鎖解析そのものは非常に古い方法論である。

20世紀初頭に Morgan らはショウジョウバエに放射線を当てて突然変異体を作り、これを交配してその孫世代を調べることで、眼の色、羽の形状などたくさんの「表現型」に關与する遺伝子のゲノム上での相対的な位置を決めることに成功した。これは「連鎖地図」と呼ばれ、異なる表現型を有する個体同志を交配してその表現型の「分離」をもとに連鎖地図を作る方法論を「連鎖解析」と呼んでいた。いまでは一般的な生物学の教科書に載っている方法論である。

原理的にはこれをヒトに応用することは可能である。しかし、現実的にはそれを非常に困難なものにする要素が2つあった。ひとつはショウジョウバエでは突然変異体を人工的に作ることで多数得ることのできた「マーカー」である。ヒトの場合、「マーカー」として使える表現型は非常に限られていて実用的とはいえなかったし、かといってショウジョウバエのように突然変異体を人工的につくるわけにはいかない。RFLPの応用はこの問題を解決した。もうひとつの要素は「交配」である。ショウジョウバエの場合には研究者の好きなだけ交配を掛けて、数千匹におよぶ「孫世代」のバエを調べることができるが、ヒトの場合、無論それは不可能である。これを簡単に解決する方法はなく、研究者と臨床医の地道な共同研究で各種の遺伝性疾患の家系を収集する以外になかった。

1983年の Nature 誌に画期的な論文が掲載された。ハンチントン舞踏病の原因遺伝子が第4染色体の短腕上に存在することがわかったというものである²⁾。ハンチントン舞踏病は原因不明の神経変性疾患で、常染色体優性遺伝することが知られていた。不随意運動で発症し、急速に進行して大脳の萎縮が進み、完全な痴呆状態で死亡する病氣

である。アルタの研究会に参加していた研究者を含むハーバード大学を中心とするチームが、ベネズエラで収集したハンチントン舞踏病家系を用いた連鎖解析で明らかにしたものである。当時著者は駆け出しの研究者で、高血圧の原因遺伝子解明を目指して全く別のテーマで仕事をしていて、この論文を見て、細かいところは良く理解出来なかったにもかかわらず、非常に重要な仕事で高血圧遺伝子解明にも使えるに違いないと興奮したのを覚えている。ちなみにこの論文の第2著者となっている Nancy Wexler は、母親をハンチントン舞踏病で亡くしており、自身も50%の確率で同じ病にかかる可能性があった人であり、この病氣の解明のためにベネズエラの家系収集に尽力した¹⁾。

この後、RFLPの代わりに、より簡便で有用性の高い microsatellite marker が解析に用いられることとなり、1980 - 2000年のほぼ20年の間にそれまで原因不明とされていた数百の遺伝性疾患の原因遺伝子が同定された。この方法論のユニークな点は、原因遺伝子の機能が不明のまま、その遺伝子がある疾患の原因となっていることが分かる点である。実際にポジショナル・クローニングの手法を用いて同定された疾患原因遺伝子には、病氣の名前そのものがつけられているものが多い(ハンチントン舞踏病の原因遺伝子はハンチンチン、筋ジストロフィーの遺伝子はジストロフィン、家族性大腸ポリポーシスの原因遺伝子は、adenomatous polyposis coli の頭文字を取って APC 遺伝子と呼ばれている)。これは発見当時にその遺伝子の機能がわかっていなかったことが主な理由と考えられる。これらの遺伝子の機能はその後の研究で明らかになってきており、特に家族性腫瘍症候群の原因遺伝子として見つかった多くののが

ん抑制遺伝子は、がん化のメカニズムの解明に非常に大きな貢献を果たしている。

このような事情を考えると、おそらく、この分野からいずれノーベル賞が出るのではないかと思われる。問題は誰がその栄誉を独占するか、ということになるだろう。多くの研究者が少しずつ貢献して成し遂げられたことのように見えるため、その中から3人を選ぶというのはかなり難しいように思える。

b) 多因子遺伝性疾患 (生活習慣病) への応用

まだ単因子遺伝性疾患の連鎖解析が盛んに行われていた90年代初めから、同じ手法を高血圧や糖尿病、統合失調症などの多因子遺伝性疾患へ応用する試みが始められた。多因子遺伝性疾患は、複数の遺伝素因に複数の環境因子が複雑に絡み合って発症にいたる疾患群で、単因子遺伝性疾患に比べて患者数ははるかに多い。これらの疾患の素因遺伝子が解明され、それに応じた予防法や治療法の開発が進めば、その恩恵に浴する人の数ははかりしれないと考えられる。

最初に、単因子遺伝性疾患同様、家系の収集と連鎖解析が進められた。しかし、これは思ったような成果を挙げられなかった。主な原因は2つである。まず第一に、単因子遺伝性疾患と異なり、「患者」と「非患者」の区別が曖昧であることが挙げられる。例えば高血圧の場合、血圧が140/90 mmHgを超えると「高血圧症」としているが、それでは139/89 mmHgの人は完全に「健常」といえるのか、と問われれば首をかしげざるを得ない。もともと正規分布をしている血圧という「表現型」の中に無理矢理線を引いて作ったものであるから、この線の直上と直下で本質的な違いがあるとは思えない。第二に、同じ「患者」でも異なる原因遺伝子を有している可能性があることであ

る。このため、単因子遺伝性疾患のときには有効であった、複数の疾患家系を一緒にして分析するという手法が取れないこととなる。

このような問題を解決する方法として、いくつか連鎖解析を改良した方法論が提唱された。これによって、いくつかの疾患で素因遺伝子につながる成果が得られたものの、結局、ヒトゲノム計画終了後に本格化した、SNPを用いた関連解析がおこなわれるまで、大きな成果が得られることはなかった。

c) ヒトゲノムプロジェクトとハップマッププロジェクト

1980年代から、アメリカの国立衛生研究所を中心として世界中の研究所が共同でヒトゲノム配列をすべて解読するという壮大なプロジェクトが実施された。その最初の成果はドラフトシーケンスとして2001年に発表された³⁾。これはあくまでも白人の男性1人分のゲノムを解読したものだったが、これに基づき、いろいろな人種、民族から合計270人分のDNAを集めて、その配列をすべて読むハップマッププロジェクトがスタートした。その目的はヒトゲノムに存在するSNPをなるべく多く同定し、その頻度の民族差の情報を収集することである。これによってすでに150万個を超えるSNPが同定されている。

d) 網羅的全ゲノム関連解析 (genome-wide association studies-GWAS-) の発展

ハップマッププロジェクトの成果を受けて、2007年に初めてのGWASの成果が発表された。

躁鬱病、冠動脈疾患、高血圧症、関節リュウマチ、型、型糖尿病、クローン病の7つの疾患の各2000人の患者とこれらの疾患を持っていない3000人の対照群、合計17000人のDNAを収集し、50万個のSNPを使って遺伝子型を調べたもので

ある。これによって、いくつかの SNP が非常に高い確率でそれぞれの疾患と関連するという結果が得られた⁴⁾。

GWAS の考え方も連鎖解析と同様、古典的なもので、医学研究者にはなじみの深い case-control study を応用したものにすぎない。case-control study では例えば、肺がんの患者と肺がんのない対照群を収集して、過去の喫煙歴（リスク要因への暴露歴）を調べ、その頻度に差があれば喫煙が肺がんのリスクになると結論づける。

GWAS の場合には、「喫煙」の代わりに生まれつき個々人のもつ遺伝子多型 SNP を用いるところが異なるだけである。つまり、ある SNP の特定の遺伝子型が患者群に多ければ、その SNP がその疾患のリスクになると考える。これまでと異なるのは、一度に潜在的なリスク要因として数十万から百万個の SNP を解析するという点であり、そのための統計学的手法に多重解析を考慮した処理法を用いなければならない点のみとあって良い。また、これだけ多数の要因を一度に解析するためには巨大な集団を解析するの必要があり、それをコスト面と作業スピードの面で現実的なものにした技術的進歩がそれを支えた。つまり、ごく古典的かつシンプルな原理に基づく解析方法を、モダンな技術で可能にしたのが GWAS ということになる。

もう一点、GWAS は医学生物学的研究の転換点となるかもしれない要素がある。上述したように、いくら技術が進んでも数万人規模で患者を収集するのは個々の研究者、ひとつの研究室では不可能にちかい。そこで、GWAS を実施している研究室は世界中のたくさんの研究室、研究者とコンソーシアムを形成し、サンプルやデータを共有することでこの種の研究を実現した。ひとつの論

文に載る著者の数は飛躍的に多くなり、上述の最初の GWAS 論文では、著者として、「The Wellcome Trust Case Control Consortium」とあるだけで、このコンソーシアムに参加したメンバーを論文の末尾にほとんど2ページを割いて載せている。これは、医学生物学研究の大型化を意味している。かつて物理学の研究が、小さな実験の結果をもとに個々の研究者が紙と鉛筆だけで行っていた理論物理学から、巨額の研究費を投じて巨大サイクロトロンを建設して行う実験物理学へ移行したのと同じような流れが医学生物学研究でも起こっているかもしれない。

3. 生活習慣病遺伝解析の現状と課題

以上にみてきたように、現在主流となっている研究手法は GWAS と呼ばれる巨大な case-control study である。GWAS は高血圧、糖尿病をはじめ、脳血管障害、心血管障害、統合失調症、自閉症、炎症性腸疾患、各種の自己免疫疾患、乳がんや前立腺がんなどの通常のがんなど、あらゆる「多因子遺伝性疾患」に応用されている。しかし、その中から新たな課題が浮かび上がってきている。

その第1は、見つかった個々の SNP の影響が非常に小さい点である。たとえば高血圧の場合、現在10個程度の SNP がかなり再現性のよいリスク要因としてあがっているが、この SNP はそれぞれ、高血圧を発症する相対的リスクを1.1倍程度（あるいはそれ以下）上昇させるに過ぎない。これを数十個集めても、血圧の個人差の30 - 50%にかかわるといわれている遺伝的要因の数%しか説明できない。

これは他の生活習慣病の GWAS についても同様であり、「missing heritability」（失われた遺

伝性) と呼ばれる重大な問題となっている。いくつかの仮説がその説明を試みているが、これについては今後の実証研究が必要となり、そのためにはおそらく近年実用化された新たなシーケンス技術を使う必要があると思われる。

もう一つの問題も上記と密接にかかわるが、現在の GWAS は頻度の多い一塩基置換という多型のみを対象としている点である。ヒトゲノムの個人差 (多型) を形作る変異としては一塩基置換の他に、重複、欠失などがあり、近年の研究で小さな単位での重複、欠失 (copy number variation-CNV- と呼ばれる) がかなりたくさん存在していることが明らかになってきた。また、より頻度の少ない SNP が生活習慣病のリスクになり得るといふ仮説も提唱されており、このような、これまで研究対象となっていなかった多型が「missing heritability」を説明するのではないかと考えられている。

第3に、これも「missing heritability」と密接に関連するが、現在の GWAS では、遺伝子間や遺伝子と環境因子間での相互作用の解析が出来ないことである。高血圧と食塩摂取やストレス、糖尿病とカロリー過剰摂取や運動不足のように、多くの生活習慣病では環境因子の影響が大きいことが知られている。したがって、SNP がある特定の環境要因に暴露されることではじめて、大きな効果を発揮する可能性もある。生活習慣病の予防や治療を考えるとこのような「遺伝環境相互作用」を示す遺伝子を見つけることは非常に有用なことであると考えられる。

4. これからの多因子遺伝性疾患素因遺伝子研究

a) 次世代シーケンス技術の発展とその応用

現在主流になっている GWAS には上述のよう

な問題点があり、今後はこれをどう解決していくかが課題となる。その鍵を握る技術のひとつが次世代シーケンス技術である。より低頻度の SNP や重複欠失のような SNP 以外の多型を網羅的に解析するために、個々人のゲノム配列をすべて読んでしまうことが考えられる。2008年の Nature 誌に James Watson (DNA 2 重らせん構造の解明でノーベル賞を受賞した研究者) のゲノム配列が発表された⁵⁾。これが、次世代型のシーケンス技術で配列を決定された最初の例ということになるが、今や、数十万円の費用と1週間ほどの時間があれば、ヒトひとりのゲノムをすべて解読することが可能となっている。この技術を応用することで、より頻度の低い SNP や CNV のような別のタイプの多型が疾患発症に及ぼす影響をより網羅的に検討できる時代となった。ただし、全ゲノム配列がわかってしまうと知りたくなかった別の変異などもすべて分かってしまうことになり、この方法論を研究に応用するためには高い倫理的なハードルが存在することも忘れてはならない。

b) エピジェネティックス

遺伝情報は DNA 配列が担っているが、DNA のメチル化であるとか、DNA を巻き取っている蛋白質、ヒストンのアセチル化やメチル化が遺伝子発現に影響を及ぼすことで遺伝子の機能を修飾していることが分かっており、DNA 配列そのものに基づかないこのような調節機構はエピジェネティックスと総称されている。

近年の研究で、驚いたことに、エピジェネティックな変化は世代をまたいで子孫の表現形質に影響を与えることが明らかとなった⁶⁾。これは、いわば「獲得形質が遺伝する」というラマルク説の復権に相当する。むしろ、恒久的な形質は

DNA 配列によって一義的に決まっているのだが、それを微調整する部分で、エピジェネティクスは思った以上に大きな役割を果たしている可能性がある。

高血圧などの生活習慣病における環境因子の役割を考えると、エピジェネティックな変化が環境因子によってもたらされ、次世代次々世代の形質に影響を与え、彼らが親世代と同様の環境で生活することでたえずエピジェネティックな変化が補強され続ける、というメカニズムで高血圧の遺伝性が維持されている可能性も考えられる。これについても次世代シーケンス技術でエピジェネティックな変化を来したゲノム領域を網羅的に調べることで新たな手がかりが得られるかも知れない。

5. おわりに

生活習慣病の遺伝解析の歴史について概観するとともに、今後の方向について予想されることを述べた。生活習慣病遺伝子の解明にはまだしばらく

く時間がかかりそうである。これに対して、本稿では触れなかったが、薬への反応性や副作用の起こりやすさを遺伝子レベルで調べることは近いうちに実用化されるかもしれない。がん組織の遺伝子を調べて、がん化によってどんな遺伝子に変異が起こっているのかを網羅的に調べる研究も進められており、新たな治療の標的を明らかにできることが期待されている。

こうしてみると、テクノロジーの進歩に支えられて巨大化してきた遺伝研究の様子がわかる。テクノロジーの開発には資金が必要であり、一度開発されるとそれを使って利潤を出すことが会社には求められる。研究に必要なだから開発された技術がときには、その技術を活用するための巨大プロジェクト起ち上げを要請する、という構図が無いともいえない。現場で研究に携わる研究者のひとりとして、今後の遺伝研究がどのような方向に発展していくのか、興味を持って見ていきたいと考える。

文 献

- 1) ピショップ, ウォルドホルツ 遺伝子の狩人 牧野賢治, 秦洋一, 瀬尾隆訳, 化学同人 1992年.
- 2) Gusella JF, et al. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 1983; 306: 234-238.
- 3) International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860-921.
- 4) The Wellcome Trust Case Control Consortium. Genome-wide association study of 14000 cases of seven common diseases and 3000 controls. *Nature* 2007; 447: 661-678.
- 5) Wheeler DA et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 2008; 452: 872-876.
- 6) Greer El et al. Transgenerational epigenetic inheritance of longevity in *Caenorhabditis elegans*. *Nature* 2011; 479: 365-372.