

# ヒト疾患遺伝子

Gerardo Jimenez-Sanchez\* ,Barton Childs\* & David Valle\*†

\* Department of Pediatrics, McKusick-Nathans Institute of Genetic Medicine, and † Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

ヒトゲノムの全塩基配列の解読により、疾患の原因となるすべての遺伝子の同定が促進されるだろう。我々は、疾患遺伝子とその産物を機能別に分類することにより、ヒトの病気の一般原則を明らかにできると考えている。我々はすでに、論文などで発表されたほぼ1000個の疾患遺伝子について機能別分類を行い、遺伝子産物の機能と発症年齢や遺伝様式などの疾患の性質との間に、意表をつくような関係を見いだしている。複合的形質にかかわるものを含めて疾患遺伝子の説明が進めば、さらに高度な解析ができるようになる。そうして得られた結果により、疾患に関する知識はますます深まり、医学と生物学とが重なり合う部分はさらに拡大していくと思われる。

疾患遺伝子とその産物を機能に基づいて分類すれば疾患について包括的に考察できるだろうという考え<sup>1,2</sup>を検証するために、今回、まず疾患遺伝子の一覧表を作成し、再分類を行った。そのために、まず『遺伝疾患の代謝および分子基礎』<sup>3</sup>の第7版から269個の遺伝子を選んだ。次に、「疾患遺伝子地図」と、ヒトの疾患および関連遺伝子の資料であるオンライン版<sup>4</sup>ヒトにおけるメンデル遺伝(OMIM)<sup>4</sup>(www.ncbi.nlm.nih.gov)中の「疾患地図」から対立形質の変異体を検索したところ、疾患遺伝子の総数は増大し、923個になった。この集団中には、1つの遺伝子が原因で起こる単一遺伝子病の原因遺伝子(全体の97%)、および複合的形質に対して感受性を増大する遺伝子が含まれている。非遺伝性の癌のような、体細胞性の遺伝子疾患だけにかかわる遺伝子や、ミトコンドリアゲノムの遺伝子は集団から除外した。

## 機能別分類

まず、各遺伝子産物タンパク質の機能に基づいて、疾患遺伝子を分類した(補完資料参照)。我々のとった分類方法は、ヒトゲノム解読国際共同研究体(IHGSC)が今回のヒトゲノム概要配列の注釈に用いているもの<sup>4</sup>とは、2つの点で異なっている。第一に、我々はタンパク質の生物学的な位置づけについては考慮せず、タンパク質の機能自体だけに注目した。これに対してIHGSCは、遺伝子オントロジー計画<sup>5</sup>が採用している分類法によっており、こちらは機能の3つの側面、つまり生化学的活性、それがかかわる生体内過程、細胞内での局在部位の3つをまとめて評価している。第二に、機能を表す我々の命名法は大体において病理的性質に着目しているのに対し、IHGSCの命名はほぼ完全に、モデル生物で機能がわかっているタンパク質との配列相同性に基づくものである。本論文ではまた、発症年齢、遺伝様式、頻度、重症度、疾病にかかわる組織の範囲、奇形との関連などを含む臨床症状に関連した性質によっても、各疾患遺伝子を評価している。

923個の疾患遺伝子が指令するタンパク質を機能別に分類した結果を図1aに示した。機能別分類で最大群となったのは酵素に対する遺伝子群で、全体の31.2%を占める。このグループは次に大きい分類群のほぼ2倍にあたる。第二のグループは、タンパク質機能の調節因子(modulator)と命名した。これは13.6%を占めており、ここにはタンパク質機能の安定化あるいは活性化、またタンパク質の折りたたみなど、別のタンパク質に影響を与えるタンパク質が含まれる。これら以外の残りの12群はいずれも、全体の10%以下を占めるにすぎない。こ

の疾患遺伝子群中に酵素を指令するものが多数含まれることは、ヒトの遺伝性疾患の研究が主として代謝異常に関するものを扱ってきたという歴史的な傾向を反映している可能性もある<sup>2</sup>。これに対し、ポジショナルクローニングにより同定された114個の遺伝子(<http://genome.nhgri.nih.gov/clone/>で更新済み)のうち、酵素に対するものはわずかに15%にすぎず、この集団のデータには偏りが存在する可能性がある(補完資料参照)。実際、酵素に関連するタンパク質ドメインは、ショウジョウバエ(*Drosophila*)でこの種のモチーフをもつとされた8360個のタンパク質のうちの27%で見つかっている<sup>6</sup>。この結果は、高等真核生物では酵素の遺伝子の割合が25~30%である可能性を示しており、ここで問題にしている疾患遺伝子群中での酵素遺伝子の占める割合に近いのかもしれない。

## 遺伝子の機能と疾患の性質

疾患遺伝子群について、遺伝子産物の機能とその遺伝子がかかわる疾患の発症年齢との相関を確かめる解析を行ったところ(図1)、いくつか興味深い結果が得られた。まず、機能別分類すべてについて、タンパク質に対応する遺伝子のかかわる疾患は、一生を通じてどの時期にでも出現する。見たところ唯一の例外は、50歳以降に発症する疾患であったが(図1f)、この群に含まれる遺伝子の検体数は少なく、もっと数が増えればタンパク質の機能についてほかと変わらない一般の分布を示すかもしれない。第二に、出生前に発症する遺伝性疾患の原因遺伝子中では、転写因子の遺伝子が群を抜いて多かった(図1b)。転写因子の異常に起因する疾患数がこのように多いのは、発生を協調的に進行させるという、これらのタンパク質のもつ重要な役割をおそらく反映している。したがって、転写因子の遺伝子が、奇形という表現型にかかわる遺伝子の30%以上を占めるのも意外な結果ではない(補完資料参照)。

生後1年間に発症する疾患のうち、酵素遺伝子の欠陥を原因とするものの割合は並はずれて高い(47%; 図1c)。このこともまた、生物学的な予想と臨床で得られたデータの両方とよく一致している。発生途上の胎児は、胎盤を介して母親の恒常性(ホメオスタシス)を維持する代謝系と接触している。したがって、酵素の欠失が原因の先天的異常をもつ胎児は、出生時にはふつう異常が見られず、自身の代謝系が必要になり恒常性を維持する系の欠損が明らかになった後に症状が現れてくる<sup>7</sup>。発症年齢が上がるにつれて、全疾患遺伝子のうちで酵素遺伝子が占める割合は低下してくる(図1d-f)が、50歳以降に発症する疾患は明らかに例外で、酵素遺伝子の割合は増え33%を超えてい

る。しかし、この分類域にあてはまる疾患の数は少なく(18)、この年齢域で出現する複合的形質にかかわる遺伝子についての知見は少ない。この年齢域で発症する疾患遺伝子で、酵素に対するもの6個のうちの3個は、疾患の真の原因遺伝子ではなく、疾患のリスクを増大する類の変異体だとされている。

我々は疾患遺伝子を機能に基づいて分類し、発症頻度、遺伝様式、発症年齢、平均余命の短縮などを含めた疾患の性質についても比較した。図2は、上位4つの機能別分類群について比較した結果を示している。分類とは関係なく、今回解析されたほとんどの遺伝子について、それが引き起こす病気は発症が「まれ」か、あるいは「きわめてまれ」だとわかった。これはある意味で、一般的な複合的形質に影響を与える遺伝子についてはまだ初歩的な知識しかないことを表しているのだろう。遺伝様式の比較を行ったところ、酵素の遺伝子が原因となる異常は主として劣性だが、タンパク質の機能を修飾する因子や受容体の遺伝子が原因である疾患は、劣性と優性がほぼ同じくらいの数だった。一方、転写因子をコードしている遺伝子が原因の疾患は優性の場合が多かった。タンパク質にはさまざまな機能があるが、発生と体内恒常性の維持に対するタンパク質との関連性について、従来の知見と今回の結果とはよく一致している。

おもしろいことに、4つの機能分類群のそれぞれについて、発症年齢のピークは異なっていた。転写因子の遺伝子群ではピークは出生前にあり、酵素の遺伝子群では生後1年以内、受容

体の遺伝子群では1歳から思春期にかけて、タンパク質機能を修飾するものの遺伝子群では成人期初期であった。このような相関関係が見られることは、我々の機能による分類が妥当性をもつことの生物学的な証拠であり、疾患の素因がほかにもあることを示唆している。おそらく、小児期、特に思春期は最も成長が速く、さまざまな細胞や組織の間で情報伝達が活発に行われる時期なので、受容体の異常による疾患はこの時期に最も現れやすいのだろう。機能修飾に働くタンパク質にかかわる異常が高齢になってから現れがちなのも、同じような理由によるのかもしれない。つまり、このような遺伝的欠陥があっても体内の恒常性を維持する系は完全に破壊されるわけではなく、生体の要求するところと系の応答がうまく調和しないかたちをとるため、症状の現れ方はゆるやかになる。最後に、機能的分類と平均余命の短縮との間にははっきりとした関係が見られなかったことを述べておく。このことは、この2つが実際に相関しないことを表している可能性もあるが、関係を見いだすにはさらに検体数を増やすか、あるいは疾患の重症度をもっとうまく特徴づける必要があるだろう。

ヒトゲノムの機能についての注釈づけがもっと進み、ヒト疾患遺伝子を全般的に一覧できるようになれば、生物学と医学の統合化が大幅に進展するに違いない。疾患関連遺伝子についての解明が進んでいけば、すべての生物種に共通の核心的といべき重要な遺伝子群、あるいはヒトに特異的なタンパク質を指

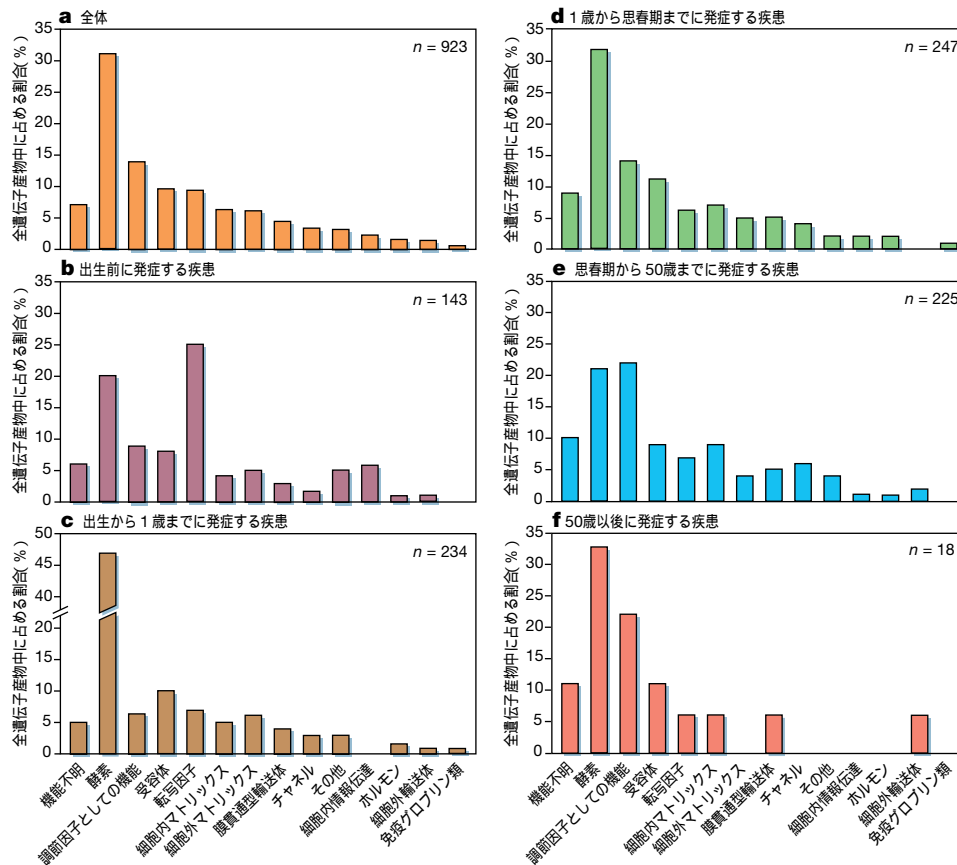


図1 疾患遺伝子産物タンパク質がもつ機能。(a)すべての疾患遺伝子について機能別に分類。(b-f)疾患を表現型が現れる代表的年齢別に分類したうえで、機能別分類群との関係を示した。出生前に発症する疾患群と出生後に発症する疾患群と比べると、転写因子の遺伝子の占める割合が異なっている(前者が25%、後者が6%;  $\chi^2 = 49.4$ ,  $P < 0.001$ )。同様に、酵素を指令している疾患原因遺伝子で、生後1歳までに発症する疾患の原因となっているものの割合(47%)は、発症年齢がこれと違う疾患の原因遺伝子で酵素を指令しているものの割合(25.8%;  $\chi^2 = 35.8$ ,  $P < 0.001$ )とは異なっている。

令している遺伝子群がそれぞれの程度疾患にかかわっているか<sup>8</sup>,塩基配列のもつ特徴 進化における保存とか多型といったような性質 が疾患の性質とどうかかりをもつか,また,タンパク質の機能が臨床治療という成果とどのように関係するのか<sup>9</sup>,などのもっと複雑な問題への取り組みが可能になるだろう。

#### 引用文献

- Childs, B. & Valle, D. Genetics, biology and disease. *Annu. Rev. Genomics Hum. Genet.* 1, 1-19 (2000).
- Jimenez-Sanchez, G., Childs, B. & Valle, D. in *The Metabolic and Molecular Bases of Inherited Disease* (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 167-174 (McGraw-Hill, New York, 2001).
- Antonarakis, S. E. & McKusick, V. A. OMIM passes the 1,000 disease gene mark. *Nature Genet.* 25, 11 (2000).
- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 409, 860-921 (2001).
- The Gene Ontology Consortium. Gene ontology: Tool for the unification of biology. *Nature Genet.* 25, 25-29 (2000).

- Rubin, G. M. et al. Comparative genomics of the eukaryotes. *Science* 287, 2204-2215 (2000).
- Brusilow, S. W., Valle, D. & Arn, P. H. in *Current Therapy in Neonatal Perinatal Medicine* (ed. Nelson, N. M.) 164-169 (B. C. Decker, Philadelphia, 1989).
- Chervitz, S. A. et al. Comparison of the complete protein sets of worm and yeast: Orthology and divergence. *Science* 282, 2022-2028 (1998).
- Treacy, E., Childs, B. & Scriver, C. R. Response to treatment in hereditary metabolic disease: 1993 survey and 10-year comparison. *Am. J. Hum. Genet.* 56, 359-367 (1995).

補充資料は (<http://www.nature.com>) で閲覧でき,印刷版はネイチャーのロンドン編集部から入手可能。

#### 謝辞

OMIMの扱いに関してJ. Ambergerに,統計的な扱いに関する助言を与えてくれたことでJ.A. Escamillaに,論文原稿の作成についてはS. Muscelliにそれぞれ感謝する。D.V. はハーワードヒューズ研究所の研究者である。

#### [原文]

Jimenez-Sanchez, G. et al. Human disease genes.

*Nature* Vol.409 No.6822, pp. 853-855 (15 February 2001)

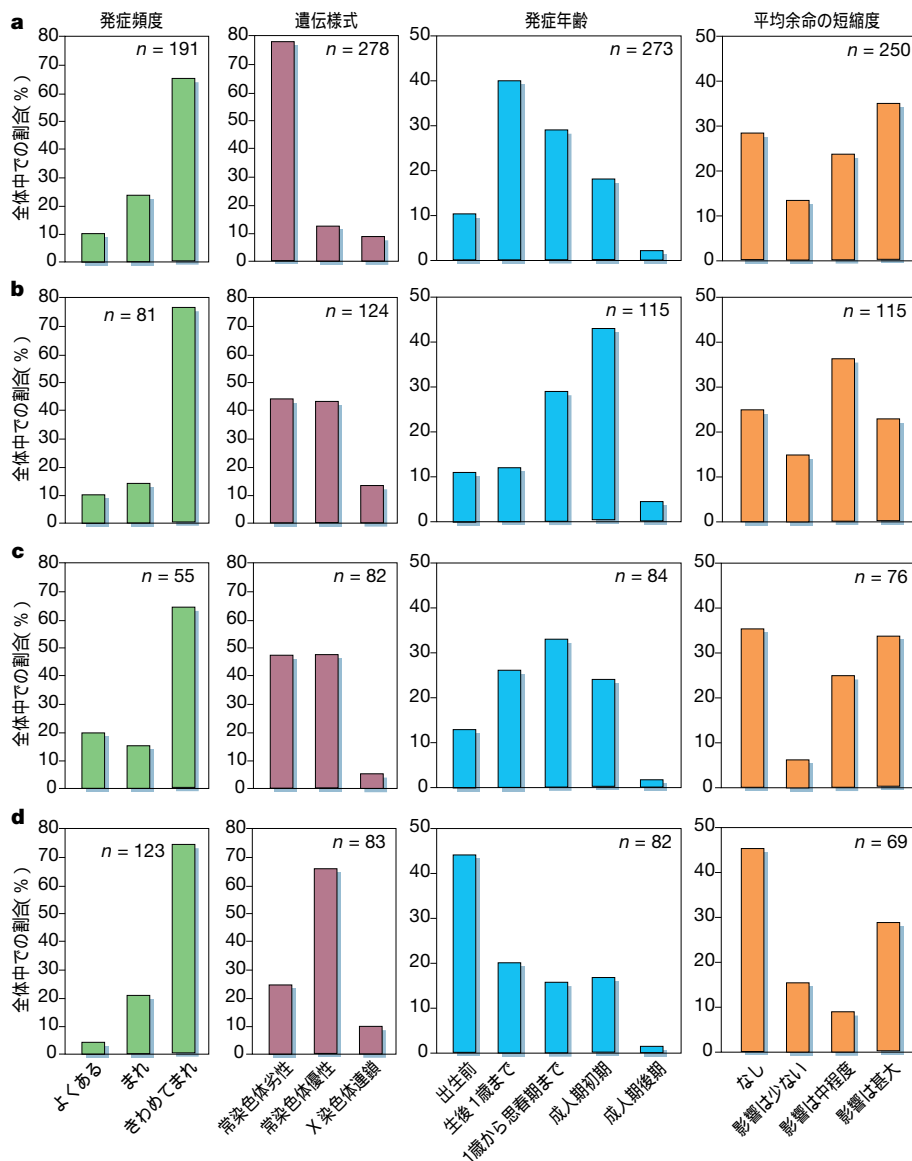


図2 疾患遺伝子を,産物タンパク質の機能に基づいて分類し,それらの遺伝子が関与している疾患の性質との関連を調べた。(a)酵素を指令する疾患遺伝子。(b)タンパク質の機能を修飾する因子を指令する疾患遺伝子。(c)受容体を指令する疾患遺伝子。(d)転写因子を指令する疾患遺伝子。疾患の性質については図の上部に示した。成人期初期とは思春期から50歳未満,成人期後期は50歳以上をさす。