

## 遺伝病とは

### 1. 単純なメンデル遺伝タイプ（解析が簡単）

- ★ハンチントン舞踏病、 囊胞線維症、デュシェンヌ型筋ジストロフィーは単純な遺伝様式を取る（常染色体優性、常染色体劣性、伴性劣性など）。
- ★遺伝子型と表現型が完全に相関している。変異遺伝子があれば必ず発症する。

### 2. 複雑な多因子遺伝（解析は難しい）

- ★通常のありふれた病気がこれに当たる。がん、喘息、精神分裂病、高血圧、心臓疾患など。疾患にかかる可能性が複数の遺伝子により左右される。
- ★「遺伝病らしさ genetic-ness 」は  $\lambda = (\text{患者の1親等の人が持つリスク}) / (\text{全人口のリスク})$  で計算される。I型糖尿病（若年発症のタイプ）の場合は  $\lambda = 15 (6\%/0.4\%)$
- ★その他の因子、特に他の遺伝子や環境要因も発症のリスクに影響する。

ポジショナルクローニング positional cloning とは（主にヒトの疾患原因）遺伝子を染色体上の位置情報をもとに単離する手法のことである。その遺伝子の構造や機能についての知識はよりどころとしない。この手法は時に（おかしな表現だが）逆行遺伝学 reverse genetics と呼ばれることもある。この手法を用いる状況は疾患遺伝子が何であるのか、何のタンパクを作るのかが予め分かっている状況（サラセミア、フェニルケトン尿症など）とは全く異なるものである。

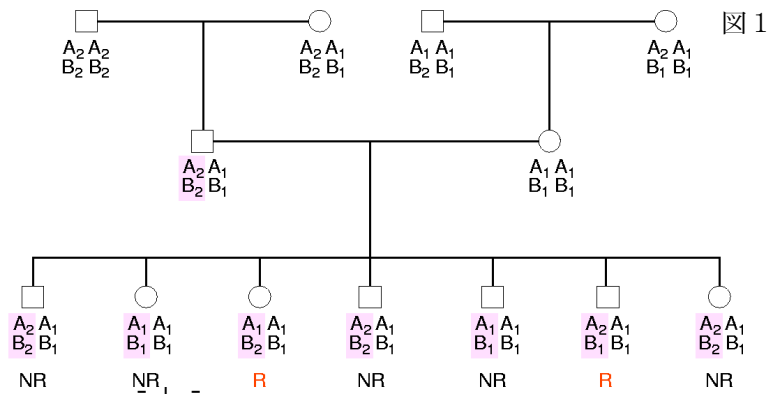
## 連鎖 linkage と組換え recombination

### 家系調査

ヒトの疾患遺伝子の位置を調べるためにはDNA 多型 polymorphism が利用される。DNA 多型とは個人間で違いが認められる塩基配列のことでヒトゲノムの中に多数の場所に存在する。この多型は通常 PCR をもちいた簡単な検出法でタイピングが可能である。常染色体優性、劣性、あるいは伴性遺伝の単純な遺伝病の場合は、多型は遺伝的連鎖を同定するために調べられる。連鎖とはメンデル第三法則（独立の法則）の例外であり、2つの遺伝子が同じ染色体にのっている場合には必ずしも独立の法則がなりたないことがノーベル賞学者モルガンにより示された。染色体上で多型の位置が疾患遺伝子に近い場合、減数分裂の際の組換えが起こる確率が低いので連鎖は保たれることになる。逆に疾患遺伝子と離れている場合には、連鎖が認められない。連鎖の度合いを計算するためには、同じ親から生まれた子供の中で組換えのある子 (R, recombinant) と組換えのない子 (NR, non-recombinant) の数を数えればよい。組換え率 recombination fraction (RF,  $\theta$  theta) は  $R/(R+NR)$  で計算され、これが 0.5 より有意に小さい場合には連鎖ありとみなされる。

図1は家系内での組換えの存在を示している。A と B は2つの遺伝子座であり、それぞれ1と2 というアレル allele（対立遺伝子、あるいは、多型のタイプ）がある。両親（第2世代）においては個々の遺伝子座のどのアレルが祖父母から伝わってきたかが見てとれる。たとえば父親ではアレル A2 と B2 が同じ染色体上にあり、A1 と B1 が同じ染色体上にある。この2つの多型マーカーの関係を位相 phase

という。3番目と6番目の子供には組み換わった染色体 (A1/B2 と A2/B1 アレル) を父親から受け取っている。その他の子供は組み換わっていない染色体 (A1/B1 か A2/B2 アレル) を受け取っている。組換え

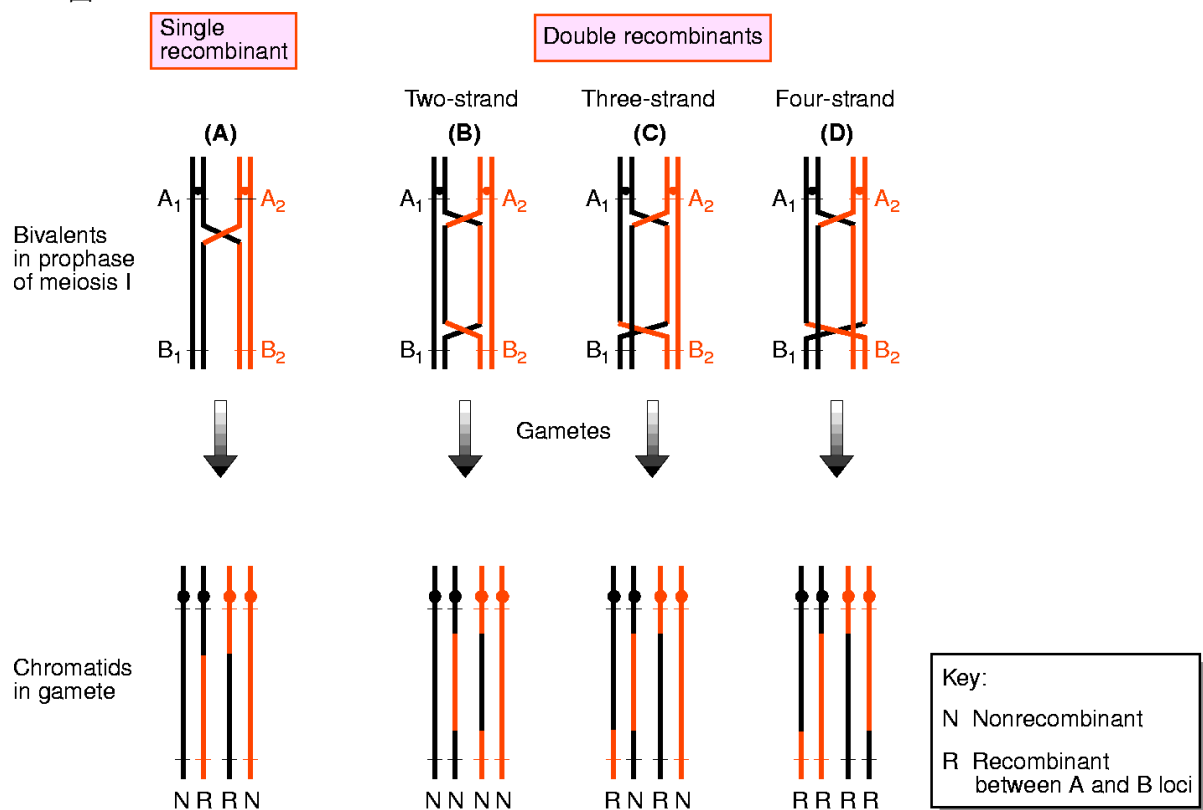


率 (RF,  $\theta$ ) は  $2/7 = 0.28$ となる。

同じ染色体上にある、2つの遺伝子座の間の組換え率 (RF,  $\theta$ ) は実際の DNA 上の距離と関係がある。遺伝学的に計測される距離はセンチモルガン centiMorgans (cM) という単位で表される。5 cM 以下の場合、1% の RF は 1 cM に相当し、さらに1 Mb の DNA に相当する。連続した遺伝子座間の RF は単純に足すことができない。というのもどんなに離れていても2つの遺伝子座間の RF は 0.5 を越えることがないからである。その理由は減数分裂の際に起こることを考慮すれば明らかである。

図2は減数分裂の際、離れたマーカー間で1回もしくは2回の組換えが起きる様子を示している。これらすべてのパターンの組換えがランダムに起きる結果、子孫には組換え型、非組換え型の両方が現れる。RF と遺伝的距離の関係はマッピング関数 (Haldane 関数、Kosambi 関数など) により表される。RF と異なり遺伝的距離は加減算可能でありヒトゲノムの遺伝的長さは 3000 cM となっている。

図2



### ヒト DNA における多型

ヒト DNA には個人間で異なる場所が何百万とある。それらには以下の種類がある。

- ★制限酵素断片長多型 restriction fragment length polymorphism (RFLP) は制限酵素部位に生じた突然変異に由来する。サザン・ブロッティングや PCR 産物の制限酵素処理により検出される。
- ★繰り返し配列 short tandem repeat polymorphism (STR)/variable number of tandem repeat (VNTR) (マイクロサテライト microsatellite ともいう) は (CA)<sub>n</sub> などの dinucleotide あるいは tri-, tetra-nucleotide などの繰り返し配列であり、繰り返しの回数が個人間で異なる場合がある。繰り返し部位を挟む PCR により検出される。
- ★1塩基多型 single nucleotide polymorphism (SNP) は DNA のコード領域にも非コード領域にも存在し、表現型に影響することもしないこともある。RFLP は SNP の特殊形ともいえる。

最初の2つは主に非コード領域 DNA に存在する。これらは全ヒト染色体の連鎖マップを構築するために用いられた。最後の SNP は現在注目されているマーカーであり、多くの遺伝子に存在する SNP が複雑な疾病の感受性と関連しているだろうと言われている。

「ゲノムスキャン genome scan 」は遺伝疾患を持つ複数の家系の遺伝子型をゲノム全体にわたって数百の DNA 多型 (遺伝的マーカーともいわれる) を用いて解析することをさしている。この様な作業は30億塩基対もある巨大なヒトゲノムを扱う際には避けては通れない。数百のマーカーを用いれば、その中に探し求める遺伝子と遺伝的連鎖が見出されるものが一つや二つあることが十分に期待できる。そしてようやく、疾患遺伝子が2つのマーカーで挟まれるゲノムのごく小さな一部分、たとえば、たった数百万塩基対の領域にあるに違いないということが分かるわけである。

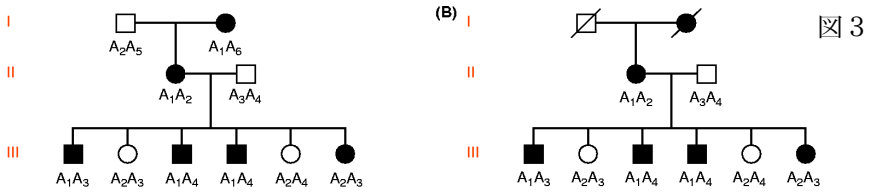
## ロッド・スコア lod score

### 2点解析 two-point mapping (単点解析)

マウスやショウジョウバエなど実験動物を対象にする場合は多数の子供を作らせて組換え型、非組換え型を計数し、組換え率 (RF,  $\theta$ ) を正確に計算することは容易である。しかし、ヒト家系を対象とする場合、子供の数は限られている。そこで、統計的に有意な連鎖の証拠を得るためには、多くの家系から情報を集め、確率論的な推測をしなくてはならない。時には家系の一部の人に関する情報しか得られないこともある。このためコンピュータを駆使した複雑な数学的手法を用いてロッド・スコア lod score を計算するのである。ロッド・スコアとは得られた家系の情報から、さまざまな RF の値を想定した場合の連鎖の可能性の程度を示す統計学的指標である。観察された遺伝子型の組み合わせが RF = 0.5 (連鎖無し) のときに起こる場合の確率  $L(0.5)$  に対する、同じ状況が RF ( $\theta$ ) が0から0.5までの値をとるときの確率  $L(\theta)$  の比 (odds) を常用対数 ( $\log$  of odds  $\rightarrow$  lod) で示したもので、RF の関数  $Z(\theta) = \log_{10} \frac{L(\theta)}{L(0.5)}$  として表現される。その最大値が確率論的に‘もっともらしい’ロッド・スコアであり、その時の RF値が遺伝的距離を反映している。ロッド・スコアが3以上であれば (つまり、非連鎖の場合の1000倍以上の確率であれば) 連鎖ありと見なされ、-2以下であれば (非連鎖の場合の100分の1以下の確率であれば) 連鎖がないと考えられる。値が-2と3の間であれば結論的なことはいえず、より多くのデータが必要であることを意味している。

図3はロッド・スコアの計算方法を単純な2つの例で示している。この家系では常染色体優性遺伝の疾患があり、A という DNA 多型マーカーについて A1, A2, A3 ... などとタイピングがなされている。パネル(A) では罹患している母親の両親 (祖父母) の情報があるため、母親におけるマーカーの位相を以下のように推測することができる。つまり、母親は疾患遺伝子と同じ染色体にある A1 アレルを祖母から受け継いだにちがいない。すると、6人の子供のうち5人は非組換え型で1人が組換え型と考えられる。そうした場合に種々の RF に対して計算されるロッド・スコアは下のテーブルのようになる。グラフで示したものが図4にある。パネル(B) では祖父母の情報がないため、マーカー (A)

の位相が不明である。そこで疾患遺伝子が A1 と連鎖している場合、A2 と連鎖している場合の2つの可能性を考慮する必要がある。その結果が図3下のテ



Lod score  $Z = \log_{10}[(1-\theta)^5 \theta / (1/2)^6]$   
 $= 6 \log_{10}(2) + 5 \log_{10}(1-\theta) + \log_{10}(\theta)$

$\theta$	0	0.1	0.2	0.3	0.4	0.5
Z	$-\infty$	0.577	0.623	0.509	0.299	0

Lod score  $Z = \log_{10}[1/2 (1-\theta)^5 \theta / (1/2)^6 + 1/2 (1-\theta)^5 / (1/2)^6]$

$\theta$	0	0.1	0.2	0.3	0.4	0.5
Z	$-\infty$	0.276	0.323	0.222	0.076	0

ブルと図4に示してあるが、見てのとおり、位相が分からない場合にはロッド・スコアが小さくなり、(A)の場合ほど確かなことが言えないことになるのである。

図4

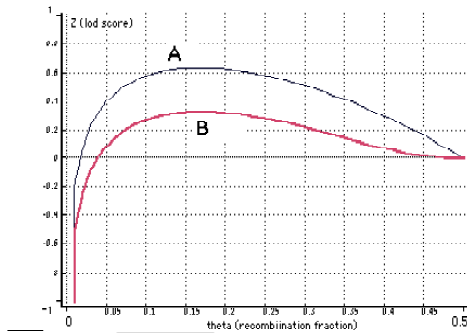


図5

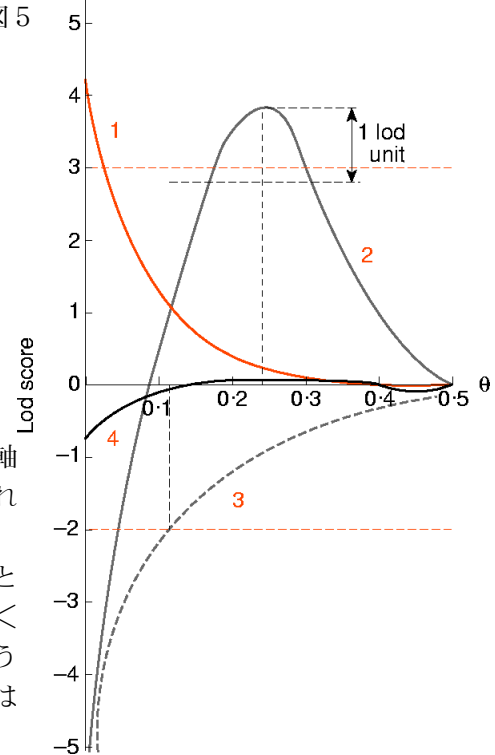


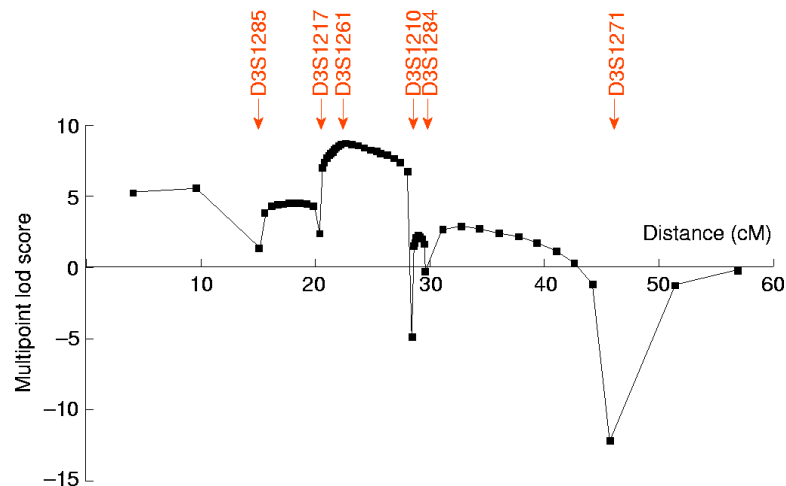
図5には RF ( $\theta$ ) を横軸にロッド・スコア (Z) を縦軸にしたグラフを示している。グラフ1は組換え型が見られない家系の場合で、 $\theta$  が0の場合に最高の値を示している。グラフ2は確率をもっとも高くなるのが  $\theta = 0.23$  のときであることを示している。グラフ3の場合は、 $\theta < 0.12$  では  $Z < -2$  となるので  $\theta < 0.12$  となるような強い連鎖の可能性が否定されることになる。グラフ4はデータが不十分で結論が得られない場合である。

### 多点解析 multi-point mapping

上の場合は疾患遺伝子と一つの多型マーカー、合わせて2点間の連鎖を考慮しただけであるが、ゲノム・スキャンを行なう場合には、染色体上での位置が分かっている多数のマーカーとの間で連鎖の可能性を計算するのである。そして、どの2つのマーカーの間に疾患遺伝子があるのかを知るのが目的である。この場合には多点連鎖解析が行われる。2点解析の場合より複雑な処理が必要でこれもコンピュータを必要とする。

図6は3番染色体に位置するマーカー群に沿って疾患遺伝子との連鎖の度合いを示したものである。横軸は染色体上の遺伝的距離 (cM)、縦軸は各点におけるロッド・スコアを示している。このグラフで最も高いピークを示す D3S1261 のあたりに疾患遺伝子がある可能性が最も高いと考えられる。

図6

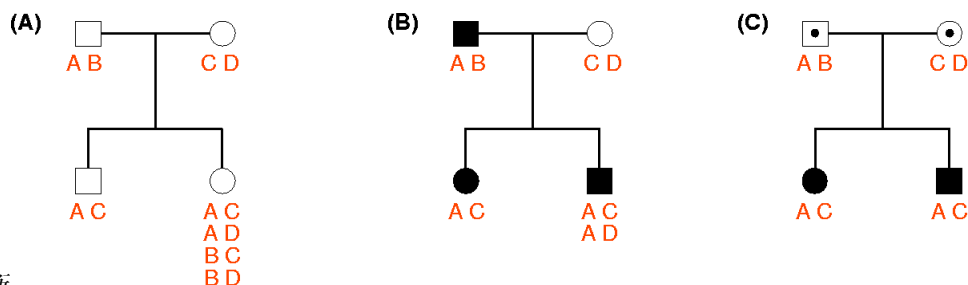


### 罹患同胞対解析 *affected sib-pair analysis*

複雑な多因子が絡む疾患の場合には単純なメンデルの法則が当てはまらず、上に述べたような連鎖解析を行なうことはできない。この場合には同胞対解析が行なわれる。2人以上の罹患した同胞（兄弟姉妹）がいるような家系を収集する。これらの同胞と両親についてゲノム全体をカバーする多型マーカーを用いてタイピングする。もし両親の染色体が子供にランダムに分配されるなら、2人の同胞があるアレルを0、1、2個共有する確率の比率は1/4 : 1/2 : 1/4となるはずである。しかし、そのアレルが疾患遺伝子と連鎖している場合にはその比率は1/4 : 1/2 : 1/4から隔たってくるであろう。この隔たりの度合いを統計的に検定し、有意差を得るためには多くの罹患同胞対を要する。

図7は罹患同胞対解析の例を示している。パネル(A)はランダムな分配の場合。第1子がACの遺伝子型の場合、第2子の遺伝子型がAC, AD, BC, BDである場合が同じ確率で起こりえる。その場合の共有されたアレルの数は2, 1, 1, 0個である。パネル(B)では罹患した同胞が2あるいは1個のアレルを共有し、パネル(C)では2個のアレルを共有している。パネル(C)のような場合は疾患遺伝子がマーカーに近い場合には頻繁に見受けられると思われる。

図7



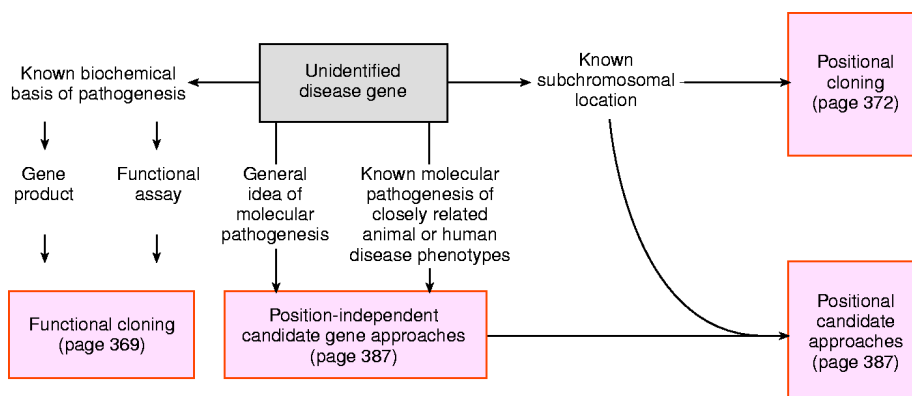
### 連鎖不平衡

単因子、多因子疾患共に扱うことができる方法に連鎖不平衡を調べる方法がある。遺伝子型は2つのアレルにより規定されるので、集団内での自由な結婚がある場合には、遺伝子型の頻度はアレルの頻度により規定される平衡状態に達する(Hardy-Weinberg 平衡という)。しかし、近親婚や亜集団間での排他的な結婚などがあるとこの平衡状態からのずれが生じ、遺伝子型の頻度がアレル頻度を反映しないことが起こる。このずれのことを連鎖不平衡という。調査対象である患者の突然変異が共通の祖先である1個人に生じた突然変異に由来する場合にのみ、連鎖不平衡の解析から疾患遺伝子に迫ることができる。この場合、疾患を持つ亜集団でのアレルの頻度は、その位置が疾患遺伝子に非常に近ければ、集団全体での頻度と異なっているはずである。この方法は候補遺伝子について近傍のマーカーを用い患者群と対象群で頻度の差を調べる際に広く用いられている。

### ポジショナル・クローニング positional cloning

図8に疾患遺伝子を同定するための戦略をまとめた。

図8



1. 疾患の生化学的な基礎が明らかである場合（生化学反応の経路上のある酵素が欠損しているなど）、その酵素をコードする遺伝子をクローニングすることは容易である。これを**機能的クローニング functional cloning** という。
2. しかし、疾患の基礎が明かでない場合も多い。その疾患の病理学的な特徴がある場合や、同様の症状を示す原因が既知のモデル動物やヒト疾患がある場合には、候補となる遺伝子について、変異があるかどうかを検索することができる。これを**候補遺伝子アプローチ**という。
3. もし、連鎖解析から遺伝子の位置しか分からない場合には、**ポジショナル・クローニング**と呼ばれる手法をとる。
4. **ポジショナル候補遺伝子アプローチ**は2と3との組み合わせである。ゲノム上の位置が連鎖解析から得られ、疾病の特徴から遺伝子の性質について予想できる場合、この方法がとられるであろう。ヒトゲノムプロジェクトが終了しすべての遺伝子がマップされている状況ではこの手法がもっとも重要なものとなるであろう。

#### 戦略を立てる際に考慮すべき点

- ★遺伝様式は？単純か（その様式は？）複雑か？
- ★どのようにして検索するか？候補となる遺伝子があるのか？
- ★疾患家系を調べることができるのか？できる場合には、連鎖解析をはじめ、多型マーカーが疾患と挙動を一にするかどうか調べればよい。そしてゲノムの候補領域が狭められていくのである。

#### プロジェクトの進行段階

1. 家系の収集。できるだけ多く集め、患者の臨床データと血液サンプル (DNA) をとる。
2. 全ゲノム上均等に分布する 400-500のマーカーを用いて家系構成人員の遺伝子型をしらべる。PCR を用いた自動処理システムが望ましい。
3. 連鎖解析を行ない（コンピュータで）疾患遺伝子が存在する候補領域を決定する。
4. その領域に含まれる遺伝子をデータベースから検索する。
5. 候補遺伝子の塩基配列を患者群と対象群で比較し、疾患特異的変異を同定する。
6. こうして同定された候補遺伝子の関与を他の方法により確認する。ノックアウトマウスや生化学的解析など。

#### 疾患遺伝子の情報を用いてできること

- ★DNA 検査により発症前（あるいは発症後）の診断が可能。
- ★その疾患の病態についてより詳細な理解が可能。（嚢胞線維症の遺伝子は塩素イオンチャネルであったなど）
- ★あらたな治療薬剤の考案。遺伝子治療も不可能ではない。

#### 参考サイト

1. Lectures by Professor Duncan Shaw (University of Aberdeen, UK)  
(<http://www.abdn.ac.uk/~gen155/lectures/gn3801.htm>): Lecture 1
2. 形質マッピングホームページ（東京女子医科大学附属 膠原病リウマチ痛風センター 遺伝統計学研究部門作成）(<http://www.genstat.net/>)
3. 連鎖解析の原理（北海道立畜産試験場・家畜生産部・育種科作成）  
(<http://www.agri.pref.hokkaido.jp/sintoku/beef/DNA%20terms/Linkage.html>)  
(分子生物学 木下 和生が参考サイト1を邦訳し、加筆した。)