

## 遺伝疫学基礎知識

遺伝学の研究は“生命言語”である DNA が発見されて以来進歩し続けています。現在 Human Genome Project としてヒトの遺伝子が全て解き明かされようとしてだけでなく、この情報が特別な遺伝子配列と病気の関係を探るのにも重要な役割を演じていることでしょう。遺伝子の知識とその機能を知る学問が Genomics であり、これを創薬や遺伝子治療に転ずることも Genomics の大きな課題となりつつあります (Pharmacogenetics)。また、遺伝的体質に基き薬物投与を行うことも今後重要な課題でしょう。マウスでは焦点を当てた遺伝子を欠損させたり、あるいは過剰発現させることによってその遺伝子の生態系における機能を検討することができます。しかし、ヒトでこれを行うわけにはいきません。そこで、患者さんと健常人との間にどのような遺伝子の差があるかを検討することにより遺伝子の働きを知ろうとするアプローチもあります (genetic epidemiology)。逆に、genetic epidemiology の知識を駆使すれば個人の将来の疾病発生率を予測し予防することができるかもしれません。また病気から離れて、民俗学にも応用できます (population genetics)。人類はアフリカから発症したのでしょうか、それとも並行して進化してきたのでしょうか？つまり、古典的遺伝学が他の学問と結びついて現代花開こうとしているのです。さらには、遺伝子操作により細胞や組織、延いては固体まで複製できる時代となりました。患者サイドからも必ずや遺伝と病気の関係について質問されることでしょう。もはや遺伝学を無視して医療を行えない時代に入ったのです。

### Human Genome Project

あたかもアメリカゴールドラッシュの時代のように、医学においてもある技術の進歩に伴って新しい金鉱が発見される度に科学者が群がるという歴史を繰り返してきました。北里柴三郎の時代、多くの病気は細菌によるものであり、多くの科学者は病原微生物の発見に努めていました (細菌狩人の時代)。その後ビタミン B 欠乏による脚気で代表される栄養性の疾患が注目を集めました (ビタミン狩人の時代)。ワトソンとクリックが DNA の構造を解明し、PCR という遺伝子を増幅する方法が開発され、最近 20 年は 遺伝子狩人 の時代とも言えましょう。しかしながら海図に載っていない 8 万以上もの遺伝子を探すのは、銀河の星を探すようなもので、非常な苦労がありました。そこでアメリカ議会は、アメリカでは対癌戦略、エイズ撲滅戦略に続き、1990 年“組織的に人の遺伝子を全て明らかにしよう”とする計画 (Human Genome Project) をおよそ 30 兆円投じて開始しました。当初 15 年の予定であったので 2005 年に終了する予定だったのですが、技術の進歩 (positional candidate cloning) がこれを加速し、2001 年春には 90% の下絵ができあがり、2003 年以前に予定を完了するであろうとされています。しかもインターネットを通じて世界の科学者に無料で供給される予定です。それでは人の遺伝子が全て解明されたとすると医学・医療は変わるのでしょうか？

遺伝子の異常によって発生する病気というと、先天代謝異常症のような小児期よりみられる疾患を連想しがちですが、例えばアルツハイマー病やパーキンソン病のような中年以降にみられる疾患の一部も遺伝子の病気と考えられています。その遺伝子はシヌクラインという蛋白の分解に関与しており、この物質の脳細胞への蓄積はアルツハイマー病やパーキンソン病の共通した病態として認められます。ですからこのシヌクラインという蛋白が脳細胞に蓄積しない工夫をすれば病状を改善できるかもしれません。このように遺伝子の解析から病気の病態を理解し新しい治療の可能性がでてきます。更に最近では糖尿病、高血圧、喘息といった一般的な病気の遺伝子解析が注目を集めています。こ

これらの病気は確かに家族歴があることも多く遺伝性を匂わせますが、かといってメンデルの法則に従う単純な遺伝疾患ではありません。比較的希なメンデルの法則に従う遺伝疾患では家系図をもとに研究(linkage analysis)することによって原因遺伝子が次々と発見されてきました。しかし一般的疾患においては linkage analysis で原因遺伝子を特定することは極めて困難であり、1990年代に入って多くの統計学的手法が考え出されました。

更に遺伝子多型といって、遺伝子の塩基配列にバリエーションを含むものが存在します。Human Genome Project と平行して Human Genome Diversity Project も1993年よりはじまりました。最近この遺伝子多型は癌や多くの慢性疾患の素因と関連していることがわかっていました。例えばアンジオテンシノーゲン遺伝子の235番目のアミノ酸にメチオニンとスレオニンの2種類があったとします。このアミノ酸の置換が科学的にどのように蛋白の機能を変えるのかはわかりませんが、スレオニンを持つ人は高血圧をもっていることが多いといった具合です。更にこの遺伝子多型と塩分の多い食事、少ない食事により4つのグループに分類でき、遺伝子と環境因子の相互作用を検討できます。他の例として、TGF $\alpha$  遺伝子多型 TaqI 持っているだけでは何も起こりませんが、母親が喫煙すると口蓋裂を合併する確率が5倍に増えます。これは遺伝子と環境の相互作用の典型例といえます。ですからこれからは、従来のように遺伝子の機能を明らかにするだけでなく、個人差について探求することも必要となります。全ての遺伝子が解明されたからといって1種類の人間の遺伝子では、「どうして鼻の形が違うか」を説明することはできません。理論上60億以上の遺伝子多型の組み合わせが存在するはずで、従来は遺伝子がクローニングされるとこの遺伝子を欠損しているあるいは強く発現しているマウスを作ってその遺伝子の機能が検討されてきました(functional genomics)。この手法は非常に多くの情報を与えてくれます。ところがマウスの遺伝子操作による表現型と人で予想される病気とは必ずしも一致しません。かといって人で遺伝子を操作して実験することはできません。ですから症状のある人の遺伝子を症状の無い人のそれと比較する手法がとられるようになってきたのです(reverse functional genomics)。

この遺伝子多型については医者だけでなく、製薬会社が遺伝子多型に強い興味を示しています。遺伝子多型により薬の副作用を防ごうというものです。何故ならたった1人の死亡例でも、過去に大金を投じてきた薬剤が市場にだせなくなるからです。そのことを考えると、事前に薬剤に対する特異体質がある患者さんを遺伝子多型解析により予知(prediction medicine)投薬を避けるようにした方が安上がりなのです。また人類学者も興味を持っています。通常子供は母親から半分、父親から半分の遺伝子を譲りうけます。この遺伝子多型は代を重ねるごとに混ざっていきませんが、あう一定の人口の中で特殊な遺伝子多型が保たれます(linkage disequilibrium)。このことを利用して、現在あるアフリカ人、ヨーロッパ人、アジア人、オーストラリア人は共通のホモサピエンスから発生したのか、ホモサピエンスがアフリカ人に進化し、そのアフリカ人からヨーロッパ人、アジア人、オーストラリア人が進化したのか研究が行なわれています。

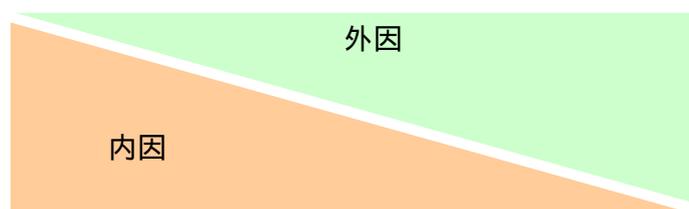
遺伝子診断には倫理的問題を伴います。例えば BRCA という癌抑制遺伝子は家族性の乳癌(卵巣癌)で変異していることが多く、逆にこの遺伝子に変異がある場合何歳までに何パーセントの割合で乳癌(卵巣癌)になるか予測できます。近年乳癌の予防に抗エストロゲン製剤が使用されるようになりましたが、その効果の真偽については未だ結論には至っておりません。このような治療法、予防法の確立していない疾患を遺伝子診断することは倫理的、精神的、社会的問題を含んできます。この家族性乳癌は特にユダヤ人に多くアメリカでは遺伝子診断が先行していますが、やはりこの BRCA 遺伝子

に変異があるということ传达了場合、そのインパクトは癌の告知をされたのに匹敵するとのレポートもあります。また家族に BRCA 遺伝子変異を伴う乳癌を認めた場合、妊娠中胎児診断をして人工中絶するか否かという問題にまで波及しています。もちろんこのような個人情報がリークすれば生命保険や仕事にも影響しかねません。現在遺伝子を調べる際には説明と同意が必要となっています。

最近の注目すべき技術の1つに DNA chip があります。切手サイズのスライドグラスに何千という遺伝子を並べ、その上にある人から分離した遺伝子を流し、遺伝子異常がないかどうかを一度に知ることができるスクリーニング検査です。Human Genome Project が完了すれば、一度に人の遺伝子全てを検査することも可能になるので、遺伝子多型を1つ1つ調査することを考えると遺伝子診断（遺伝子変異、多型、発現）をする上で非常に強力な道具となり得ます。また、何千という遺伝子の mRNA 発現状態を一度に調べることもできます。従来は遺伝子1つ1つを丹念に調べてきましたが、今後は細胞内にある一連の遺伝子群が外界からの刺激によりどのような動きをするのか、あるいは遺伝子の異常な発現パターンが病気や同じ症状でも予後の良し悪しに影響するかもしれない、あるいは臓器毎の発現パターン、さらには個人レベルの遺伝子発現の相違と病気発生、そして最後は民族による発現パターンを検討していかなくてはなりません。特に多くの遺伝子発現変化は細胞機能の調整系や生化学的パスウェイを思わせます。これは始まりであって、Human Genome Project が完了すると同時に膨大なエビデンスの蓄積作業を開始しなくてはならないでしょう。もしも 病気の発生 = 遺伝的素因 + 環境因子 + 行動 + 相互作用 +  $\alpha$  といった公式で現されるようになれば、医療は治療から予防に大きくシフトすることが予想されます。つまり病院には自分の体質を調べに行って、生活習慣、避けるべき環境因子に関する指導を受けて帰るといった時代になるかもしれないのです。

DNA chip を用いれば全ての遺伝子発現を同時に解析できるわけですが、なぜ蛋白でなく mRNA なのでしょう。蛋白発現の多くは mRNA によって規定されています。さらに mRNA は蛋白に比べてはるかに感度がよいからです。最近蛋白チップもありますが、DNA chip の方が汎用されているのはその点にあります。さらに chip による同時解析の利点は先に述べた点だけでなく、類似した機能を有する一連の遺伝子が同時に発現されていれば、結果はより確からしいといえます（クラスタ - 解析）。例えばある系の酵素1つだけが発現されていたとしたら、何か非特異的な反応かもしれません。しかし、同じ生合成系に関与する酵素全ての転写が亢進していれば意味を持ちます。また、機能の判っていない遺伝子も多々あるわけで、生物学的変化と組み合わせたり、同じ動きをする mRNA と似た働きを持つと考えるのが一般的です。更に、DNA chip は従来のノーザン・ブロットチングや RT-PCR と比較しても、同時にたくさん測定するからといって決して偽陽性、偽陰性が増えるわけではなく、むしろ先にクラスタ - で解析できる点からしてもエラーは少ないのです。名前はマイクロアレイですが、遺伝子発現の動きを全体として同時に測ることができる点、マクロの範疇に入ります。これに対してノーザンブロットチングなどで1つ1つの遺伝子発現を検討するのであればマイクロの範疇です。

## 遺伝と疾病発生



先天性代謝疾患  
ダウン症  
ハンチントン舞蹈病

心血管疾患  
癌  
精神病  
糖尿病  
アルツハイマー病  
アレルギー疾患  
高血圧  
感染症

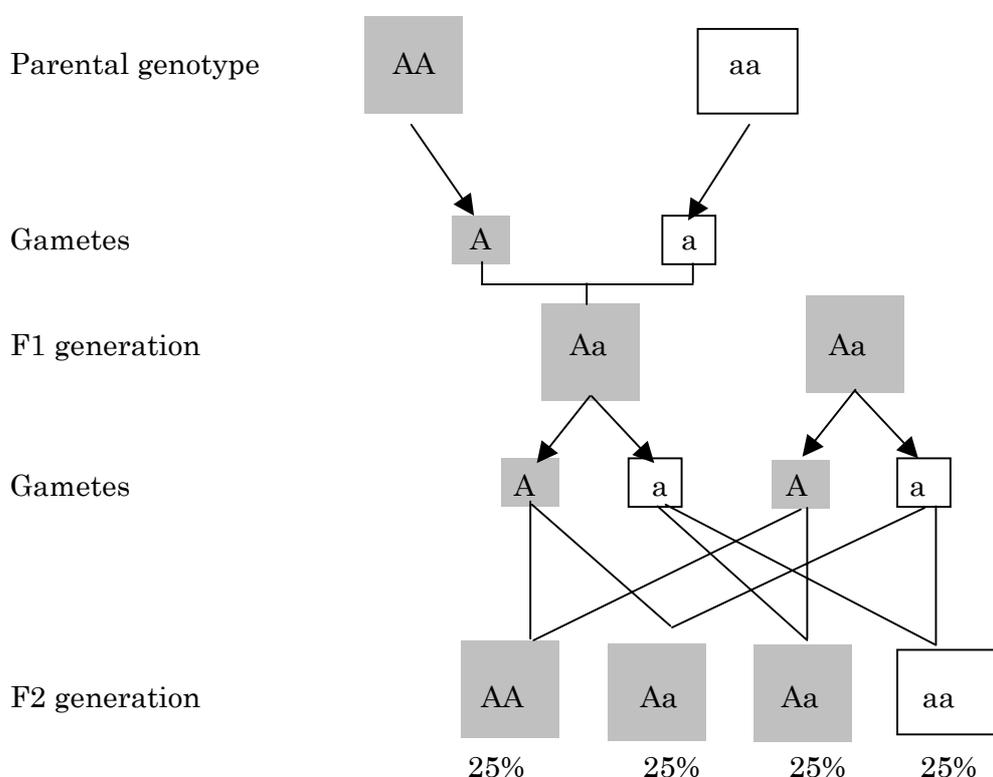
遺伝性疾患というどうしても先天的な病気を思い浮かべてしまいます。しかし、成人病も多かれ少なかれ遺伝性を有します。「親も祖父も糖尿病だった」とか、「両親とも喘息だ」といった話はよくあります。まずは、いわゆるメンデル遺伝の法則に従う疾患について解説し、後にメンデルに従わないタイプの遺伝性疾患について解説したいと思います。

### メンデル第一の法則

ヘブライ人は男性が出血で死亡すると、その死亡した男性に男の子供があればやがて同じように出血で死亡することを知っていました。また家族内で二分脊椎が発生することも記載があります。昔から一部の人はある特定の病気がある家系内で多発する現象には気が付いていたのです。

1865年オーストリアの修道士であった Gregor Mendel はサヤエンドウの種の色、形、植物の高さを含む一連の特徴が親から子へそのまま伝えられることを発表しました。この時点より学問としての遺伝学が始まったと言っても過言ではないでしょう。

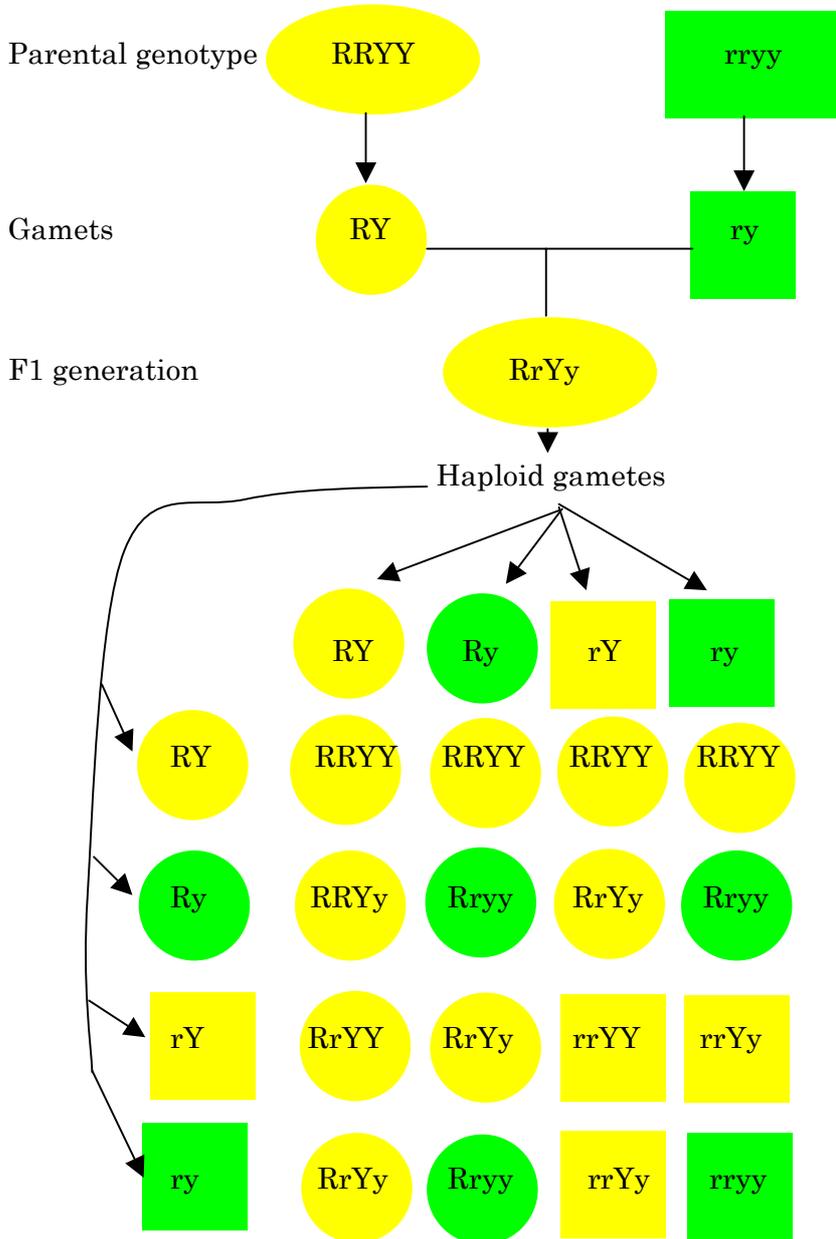
彼は、植物の高さや花の色が孫の世代に 1/4 の確率で出現することを見出し、下図のように考えました。



メンデルは最初偶然にも優性遺伝する因子を選びました。ですからAを含むものは皆同じ表現型(phenotype)となります。よってお祖父さんの代の表現型(aa)が親の代には出現しないのに子供の代に出るという奇妙な現象を認めます。

### メンデル第二の法則

2つ以上の表現型を規定する因子は独立して子供に伝わるというものです。彼は種の色（黄色か緑）と形（丸いかしわしわ）という2種類の形質がどのように伝わるかを調べました。



この画期的発見は1900年代まで陽の目を浴びることはありませんでした。20世紀、同じく植物学者とアルカプトン尿症を研究していた Sir Archibald Garrod により allele と genetic linkage の関係が発表されました。

### Hardy-Weinberg Equilibrium

遺伝学において 20 世紀初頭もう 1 つの重大な発見がありました。進化生物学者達の「何故優性形質や病気の頻度は長い世代を重ねるまで変わらないのだろう」という疑問に端を発します。この疑問に対する答えが Hardy と Weinberg の 2 人によって別々に発表されました。大きな集団において、優性形質の genotype 頻度はランダムな mating により 1 世代後に平衡状態に達することを指します。そしてこの平衡状態は進化を促すような力が加わったとき、民族の移動があったとき、新しい変異が発生したときなどに乱されます。

特に Hardy-Weinberg theory では優性遺伝において allele A と a がそれぞれ頻度  $p$  と  $q$  であるとき、 $p + q = 1$  であり、genotype AA, Aa, aa の頻度はそれぞれ  $p^2$ ,  $2pq$ ,  $q^2$  となります。この公式により病気の頻度だけでなく、キャリアの頻度、penetrance なども判ります。

疾患 A は常染色体劣性遺伝で、400 人に 1 人の頻度で発症します。この疾患 A のキャリアの頻度はどれくらいですか？

$$q = \sqrt{1/400} = 1/20$$
$$2pq = 2(19/20)(1/20) = 0.095 \quad 9.5\%$$

常染色体優性遺伝の疾患 B の allele frequency ( $q$ ) =  $1/100$  です。この人口における病気の頻度はどれくらいですか？

$$p = 1 - 1/100 = 99/100$$
$$q^2 + 2pq = (1/100)^2 + 2(99/100)(1/100) = 0.0199$$

1000 人当たりおよそ 20 人です。

浸透率 (penetrance) が不完全である常染色体優性遺伝の疾患 C があります。ある集団内でこの疾患は 1000 人中 16 人において認められます。もしも正常 allele  $p$  の頻度が 0.99 だとすると予想される浸透率はどれくらいですか？

$$P = 0.99, \quad \text{then } q = 0.01$$
$$\text{Penetrance} \times (q^2 + 2pq) = 0.016$$
$$q^2 + 2pq = 0.0199$$

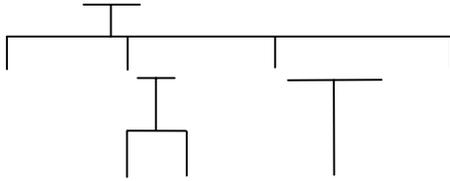
therefore, penetrance = 0.804

浸透率は 80.4% です。

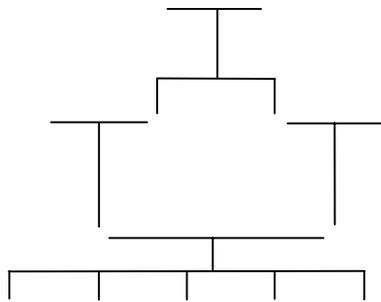
下の家系図をみて遺伝形式を推定してください。病気は単一であり、misclassificationはないものとしてください。

□ : 健康男性、 □ : 健康女性、 ■ : 病気男性、 ■ : 病気女性

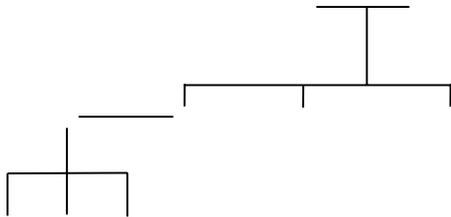
A



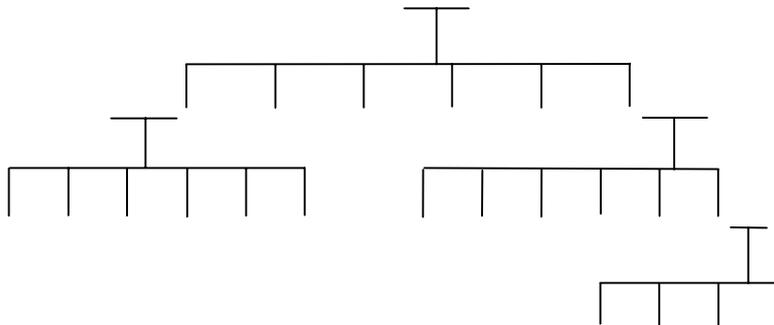
B



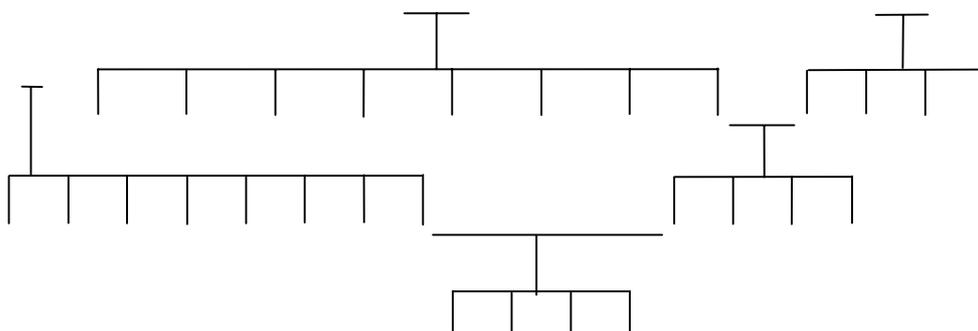
C



D



E



A:常染色体優性遺伝:両親とも病気であることは確率的に少なく片親が病気の場合を示しました。子供は 50%の確率で病気になります。男女差はありません。各世代で病気を持つ人が認められます。病気は乳幼児期致命的でない場合が多く、むしろ成人発症します。

例:軟骨発育不全症、球状赤血球症、ハンチントン舞蹈病

B:常染色体劣性遺伝:代謝疾患等に多く、重症です。1世代のみに認められることも多く、特に近親婚がある場合は要注意です。通常は4人に1人の確率です。男女とも発症し得ます。

C;伴性劣性遺伝:X染色体によって病気が伝播します。通常、女性が保因者となり、病気の遺伝子を受け取った男児が病気になります。確率は50%です。男性は病気か健康のどちらかで保因者になることはありません。女兒は基本的に病気を発症しません。

D:ミトコンドリア病遺伝:ミトコンドリアは全て母親由来です。ですから自分の持っているミトコンドリア遺伝子は遠い先祖の母親から由来しているのです。この図で注意してもらいたい点は男性患者の子供には患者は発生していないことです。そして、女性患者の子供に多く発症しています。

E:この病気は常染色体劣性遺伝で説明がつきます。しかし第4世代は両親が病気なので子供も全員病気になるはずですが、しかし4人とも正常です。どうしてしまったのでしょうか。もし表現型(病気の症状)は同じでも責任遺伝子が大きな家系で異なれば説明がつきます(責任遺伝子が2つ以上ある)。例えば酵素A,Bのどちらかの機能が異常な場合、疾病Eを発症するとします。左の家系は酵素Aの異常、右の家系は酵素Bの異常をもつとします。両家系は疾病Eを持ちますが、原因遺伝子は異なります。常染色体劣性遺伝ですから、第4世代の酵素AとBは共に正常なので、病気にはなりません。

ある遺伝子の存在する場所は locus と呼ばれます。例えば癌抑制遺伝子 p16 の locus は染色体 9p21 であるといった呼び方をします。そして、この遺伝子の 100 番目の遺伝子配列の G が T に変わることがあれば、2 種類の allele が存在することになります。2 つ以上の allele が存在し、その頻度が 1% 以上である場合 polymorphic と呼びます。polymorphism が多岐にわたるかどうかは heterogeneity で示されます。

$$H = 1.0 - \sum p_i^2$$

P: それぞれの allele の頻度

染色体は 2 本ずつありますから、1,2 の 2 種類の allele が存在した場合、genotype は 11, 12, 22 の 3 種類になります。

2 つの allele が存在する genotype について検討します。

genotype	11	12	22	total
観察された人数	35	330	635	1,000

- 1,000 人(2,000 chromosome)の allele 1 の頻度を計算してください。
- Hardy-Weinberg equilibrium を想定して期待される genotype count を計算してください。
- Homozygotes (genotype 12)の半数が genotype 22 と間違っていたとします。
  - この間違いによってどのような genotype の頻度が観察されると思いますか？
  - このデータを用いて allele 1 の頻度を計算してください。
  - この間違っ て算出された genotype の頻度は Hardy-Weinberg equilibrium に従っていますか？
- 最初の正しい genotype に戻ります。Allele 1 を homozygous として持つとすると 7 人中 6 人が 20 歳までに死亡するとします。
  - 20 歳の genotype の頻度はどのようになりますか？20 歳までにこの病気以外の死亡はないものとします。
  - この年齢の人口を調査した場合、どのようなバイアスが含まれることが考えられますか？

$$1. (35 + 35 + 330) / 2000 = 0.2$$

$$2. p = 0.2, q = 0.8$$

$$p^2 = 0.04, 2pq = 0.2 \times 0.8 = 0.32, q^2 = 0.64$$

genotype	11	12	22	total
期待される人数	40	320	640	1,000

3.
  - (1)

genotype	11	12	22	total
間違った人数	35	165	800	1,000

(2)  $(35 + 35 + 165)/1,000 = 0.1175$

(3)  $(\text{Observed} - \text{Expected})^2/\text{Expected} \gg \chi^2_1$

よって H - W には当てはまりません。

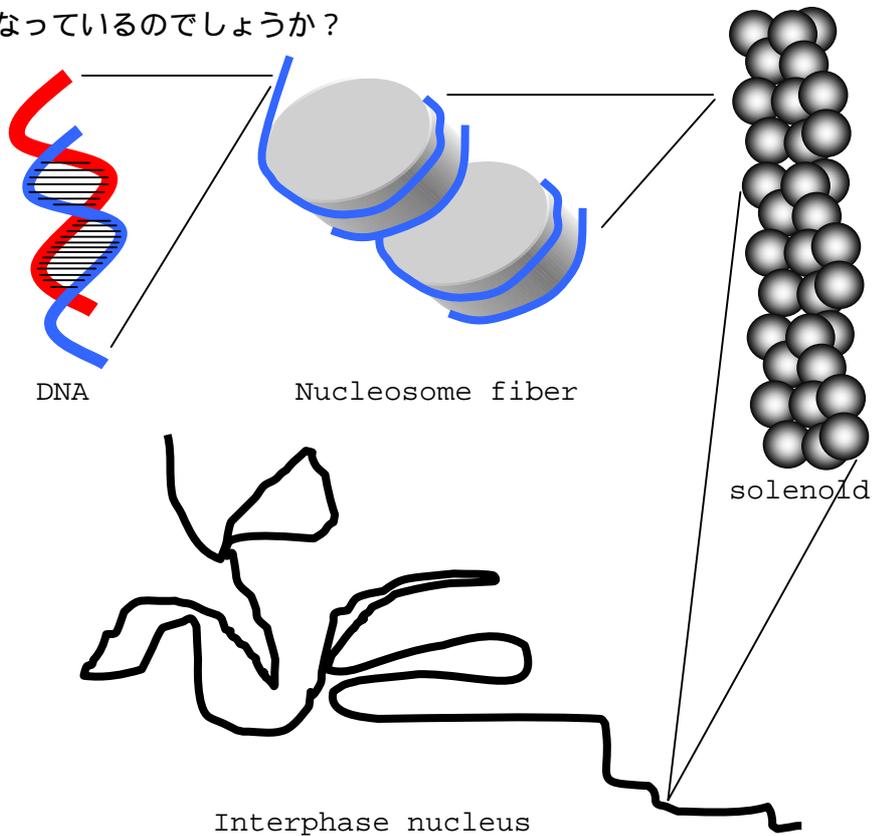
4 (1)

genotype	11	12	22	total
20 歳における分布	5	330	635	970

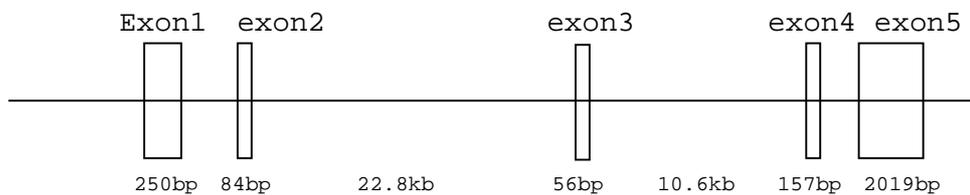
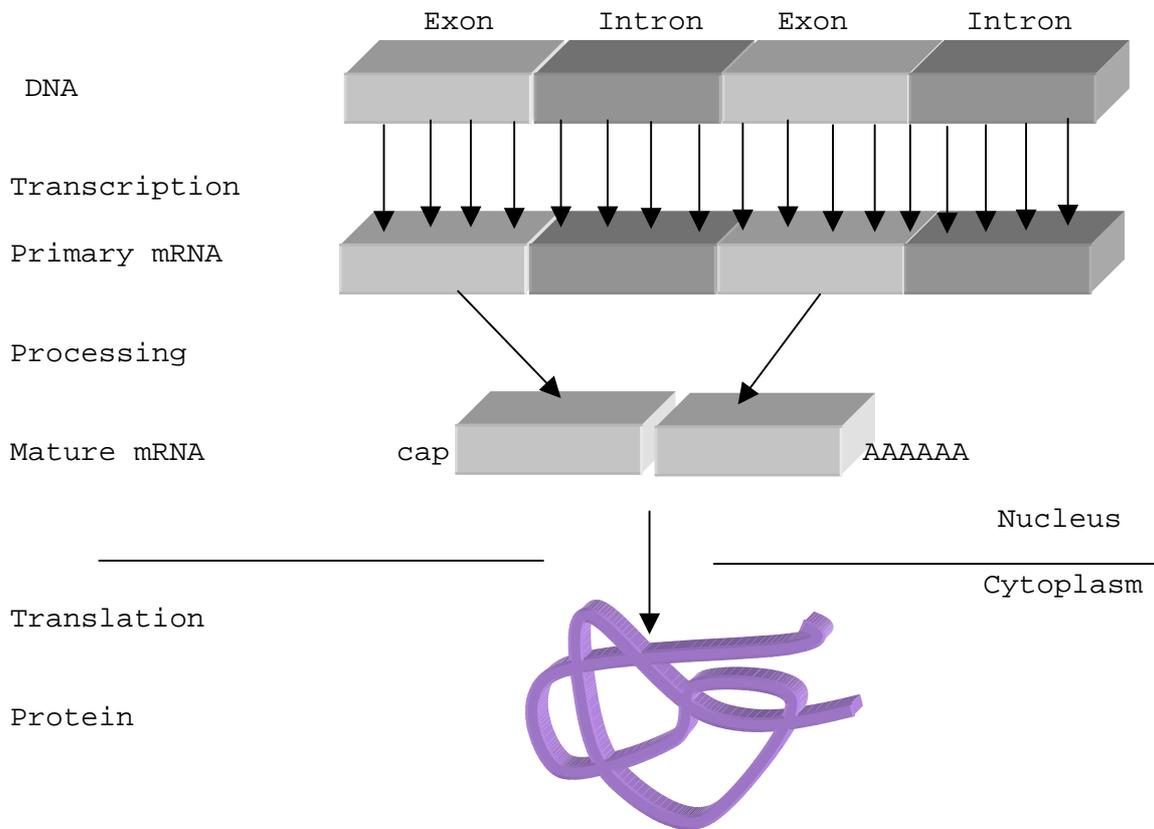
- (4) 20 歳以上の集団から検体を集め解析した場合、本疾患の頻度を明らかに過小評価することになってしまいます。この疾患により死亡がでる前の年齢層を対象にして解析しなくてはなりません。

### 遺伝子と染色体の構造

1953 年ワトソンとクリックにより遺伝子 (DNA) が二重らせん構造をとることが明らかにされたのは有名な話ですが、その 3 年後に Tjio & Levan により染色体が 46 本であることが明らかとされました。染色体は遺伝子の塊なわけですが、その関係はどのようなになっているのでしょうか？



1 つの染色体上には数百から数千の遺伝子がついています。



Exon: coding region; この部の遺伝情報が蛋白の設計図になります。

Intron: non-coding region; exon の間にある mRNA としては写しとられず、よって蛋白を規定するものではありません。個人により variation が強くあるため遺伝子のマーカーとしてしばしば用いられます。

Promoter: 遺伝子の転写を制御します。

遺伝子のどこかの塩基配列がヒトによって異なり、これを Polymorphism と呼びます。一種の遺伝子変異ですが、この polymorphism は大きな異常や病気を起こすことなく存在します。Exon や promoter は特に重要ですから、同部に変異を持つことは少ないのですが、多くの遺伝子に基く表現型の相違は、同部の polymorphism に由来すると考えられています。

遺伝子 polymorphism により 546 番目のアミノ酸のアスパラギンがヒスチジンに置き換わるような場合は A546H と表します。蛋白の機能に決定的な影響を与えなければ病気にはなりません。もしも重要な生理機能を有する蛋白に決定的な変異であれば病

気になるかもしれません。そして、1つの遺伝子において変異しているだけでも大きな影響を示すのであれば優性遺伝になりますし、1つでも正常機能を持っていけば問題ない蛋白であれば劣性遺伝を示すかもしれません。塩基の1つがかけていたり余分にあると、その後の配列は皆1つずつずれてしまうので蛋白機能は失われます(c末端に近ければ問題ないこともあります)。

**Single Nucleotide Polymorphism (SNP; アメリカでは スニップ と発音する)**

およそ 1kb 毎に存在し、全体で 300 万以上の SNPs があると見積もられています。SNPs は病気は素因の多様性に影響すると考えられています。もちろん全く影響しないものから SNP の存在自体で病気を発症させるものまで、その影響力は様々です。通常は1つの塩基対に対し2種類存在します。つまり allele は2種類となることが多いこととなります。

Sequence	AAGG <b>C</b> GTA <b>C</b> T	AAGG <b>G</b> GTA <b>C</b> T
Allele	1	2
Frequency	0.7	0.3
Heterogeneity: Het	= 1 - (0.6 <sup>2</sup> + 0.4 <sup>2</sup> ) = 0.48	

**Variable Number of Tandem Repeats (VNTRs) or Microsatellites**

Intron 部分ではしばしば(10<sup>6</sup>bp毎に存在する)いくつかの塩基の繰り返し配列が観られます。

例

allele 1: .....TACTGGTCCAATACACACACACAGGTCCCCGTAA..... (AC)4  
 allele 2: .....TACTGGTCCAATACACACACACACAGGTCCCCGTAA..... (AC)5  
 allele 3: .....TACTGGTCCAATACACACACACACACAGGTCCCCGTAA..... (AC)6  
 allele 4: .....TACTGGTCCAATACACACACACACACAGGTCCCCGTAA..... (AC)7

つまり何回繰り返し配列があるかで allele を区別します。通常は何種類もあることが多く個人で異なるのでマーカーとしては便利で、しばしば linkage analysis に使用されます。このマーカーを家系図の中で調べることにより、患者と常に一緒にリンクしているマーカーは病気の原因と関係する遺伝子と接近している可能性があります。しかし、時に病気と関連します。代表例は tri-nucleotide repeats 病で、Huntington 舞踏病などが含まれます。しかし、この繰り返し配列がどのように疾患発生に関与しているかは未だ解明されていません。

Genetic marker とは個人個人で違う DNA の印です。このマーカーをどのように使いますか？

1. Systematic genome screen
2. Linkage analysis in families

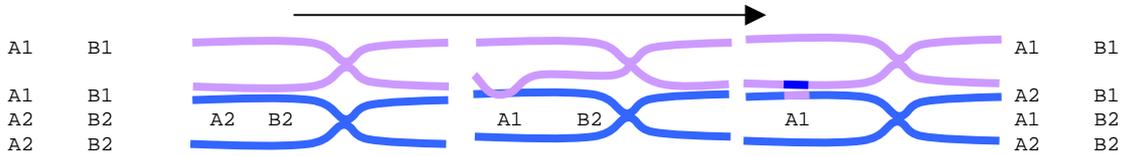
- 3. Linkage disequilibrium
  - a. Population based association analysis
  - b. Family based association analysis

論文を読んでいるとあるマーカーが heterozygosity は 0.75 であり、大きな家系で病気の遺伝子座に近いと書いてありました。これは Microsatellite ですか？それとも SNPs ですか？

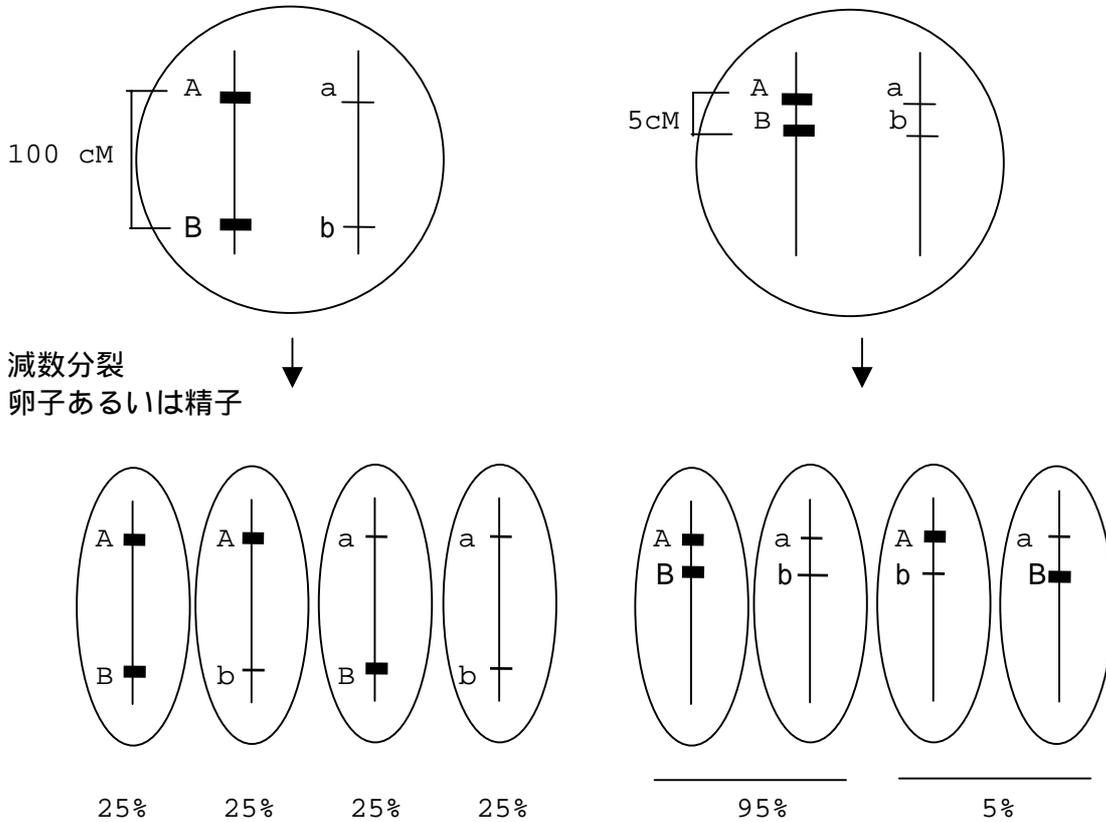
Microsatellite は 1 つの座で何種類ものマーカーがあるのが普通であり、heterogeneity は 0.7 を容易に超えます。しかし、SNPs では 2 種類のマーカーしか存在しないことも多いので、heterogeneity が 0.7 を超えることはまずありません。ですからこの場合 Microsatellite が考えられます。

### 染色体の組換え

染色体はメンデルの言うように独立して遺伝するわけではありません。減数分裂の際に染色体同志がついたり離れたりして組換えを起こします。まず最初の分裂で4本ある染色体が適当に組換えを起こします。組換えはお互いの染色体で同じ場所で発生します。よって組換えによって allele が交換されることになります。



下図のように2つの遺伝子が遠いと組換えを起こしやすく、近いと起こしにくくなります。



上図左のように100%の組換えが生じる染色体上の距離を100cMorgan, 逆に5%しか生じない距離を5cMorganと呼びます。染色体の長さは1番染色体が1番長く305cMあり、22番が一番短く81cMです。

4つの遺伝子座(A, B, C, D)があります。それぞれの組換えの頻度は以下の通りです。

A - B: 0.2086

B - C: 0.1296

C - D: 0.0906

C - A: 0.1067

D - B: 0.1967

(a) Haldane map function を用いて A-B, B-C, C-D, C-A, D-B の map distance を計算してください。

(b) 染色体上の4つの遺伝子座はどのような順番で並んでいますか？

(c) A と D の間の距離はどうか？

Haldane map function:  $x = -1/2 \ln(1 - d)$ ,  $d < 0.5$

$x_{A-B} = -1/2 \ln(1 - 0.2086) = 0.27M = 27cM$

$x_{B-C} = -1/2 \ln(1 - 0.1296) = 0.15M = 15cM$

$x_{C-D} = -1/2 \ln(1 - 0.0906) = 0.10M = 10cM$

$x_{C-A} = -1/2 \ln(1 - 0.1067) = 0.12M = 12cM$

$x_{D-B} = -1/2 \ln(1 - 0.1967) = 0.25M = 25cM$

$x_{B-C} + x_{C-D} = 15cM + 10cM = 25cM = x_{D-B}$  なので、

B, C, D の順番は B-C-D あるいは D-C-B です。

$x_{C-A} < x_{A-B}$  であり  $x_{B-C} < x_{A-B}$  なので可能性のある順番は A-D-C-B あるいは B-C-D-A です。

順番は A-D-C-B ですから、 $x_{C-A}$  は  $x_{A-D}$  と  $x_{C-D}$  の和です。さらに

$x_{A-D} = x_{C-A} - x_{C-D} = 12cM - 10cM = 2cM$  となります。

### 疾患家族内集積測定

家族内に似た病気が多くみられれば遺伝性のものかなと考えるのが普通です。しかし、病気の一般の頻度についても考慮に入れる必要があります。例えばある人口において喘息、アトピー、食事アレルギー - などのアレルギー疾患が 10 人に 1 人みられるとします。もしも血縁親戚内 15 人中 2 人にアレルギー疾患があったとします。あなたは、これを遺伝性である可能性が強いととらえますか？一方多発性硬化症などの希な疾患が 15 人の身内で 2 人認められたらどうでしょう？ K を使って家族内で発症した疾患を評価してみましょう。

$$\lambda_s = K_s / K$$

K: 一般の病気 S 頻度

K<sub>s</sub>: 同胞ペアの病気 S 共有頻度 cf. K<sub>R</sub>: 親戚の病気共有頻度

$\lambda_s = 1$ ; 遺伝性はない

遺伝性のあるものは  $\lambda_s > 1$  となるが、 $\lambda_s > 1$  だからといって遺伝性であるとは限りません。例えば環境因子 A が病気 S の原因であったとき、ある家族が因子 A に暴露されれば病気 S が遺伝病でなくとも  $\lambda_s$  は高くなります。

### 参考

疾患	遺伝型	$\lambda_s$
Cystic fibrosis	Recessive	$0.25 / 0.0004 = 500$
Huntington disease	Dominant	$0.5 / 0.0001 = 5,000$
Multiple sclerosis	Complex	20 - 30
Osteo-arthritis	Complex	4 - 6
Alzheimer disease	Complex	4 - 5
Type I diabetes	Complex	15

$\lambda_s$  だけに絞る必要はありません。

1 卵性双胎	$\lambda_M$	2 allele share
2 卵性双胎・兄弟	$\lambda_s$	0, 1, or 2 allele share
子供・親	$\lambda_0$	Exactly 1 allele share
2 親等	$\lambda_2$	Possible 1 allele share
3 親等	$\lambda_3$	Possible 1 allele share

そして

$$\lambda_M > \lambda_s > \lambda_1 > \lambda_2 > \lambda_3$$

であれば遺伝性を示唆します。例えば 1 卵性双胎で両方が病気である率が 30%、2 卵性双胎で 2 - 4%であったとしますと、遺伝的な病因が強く疑われるわけです。

一方、

$$\lambda_M = \lambda_s$$

であれば、環境因子が強く疑われます。

あなたは友人から質問を受けました。その友人は $\lambda_s = 3.0$  の病気 S について調査しており、同胞の病気 S 発生率は 30%であることを知っています。

1. この病気 S の prevalence はどれくらいですか？
2. この病気 S は遺伝性疾患ですか？
3. もし友人が 200 の病気 S の同胞ペアを集めたとしたら原因遺伝子を見つけることができるでしょうか？ $\lambda_s$  が 2.0 なのに 200 程度のサンプル数で遺伝子を見つけた論文があったからです。

1.

$\lambda_s = 3.0, K_s = 0.3 \quad K = 0.1 \quad 10\%$

2. この病気 S は遺伝性の可能性があります。一般的に言って、遺伝的要素を含有する疾患の $\lambda_s$  は 1.0 を超えます。しかしながら、 $\lambda_s$  が 3.0 前後の軽 - 中等度の疾患において遺伝性を明確に証明することは必ずしも容易ではありません。1 卵性双胎の方が 2 卵性双胎より頻度が高いなどの所見は遺伝性の可能性をさらに強く示唆します。

3. まずは $\lambda_s$  が 3.0 だからといって遺伝性とは限りません。例えば病気 S を引き起こす環境因子や B 型肝炎のような垂直感染しえる感染症であっても兄弟間の $\lambda_s$  は高くなり容易に 3.0 に達してしまいます。また家族内でインフルエンザが多いからといって遺伝病を考える人はいません。仮に遺伝的要素があったとしても、いくつもある遺伝子のうちの 1 つであったり、遺伝子 - 環境因子の相互作用であるかもしれないのです。そのような場合、どの遺伝子から調べたらよいか自体もはっきりしません。このような状態で 200 人集めれば病気がわかったら苦労はありません。

あなたの友人が論文を読んでいて以下の単語の意味がわからないと質問してきました。簡単な例を挙げて説明してください。

- A. Non-penetrant
- B. Genetic heterogeneity
- C. Phenocopy
- D. Variable expressivity

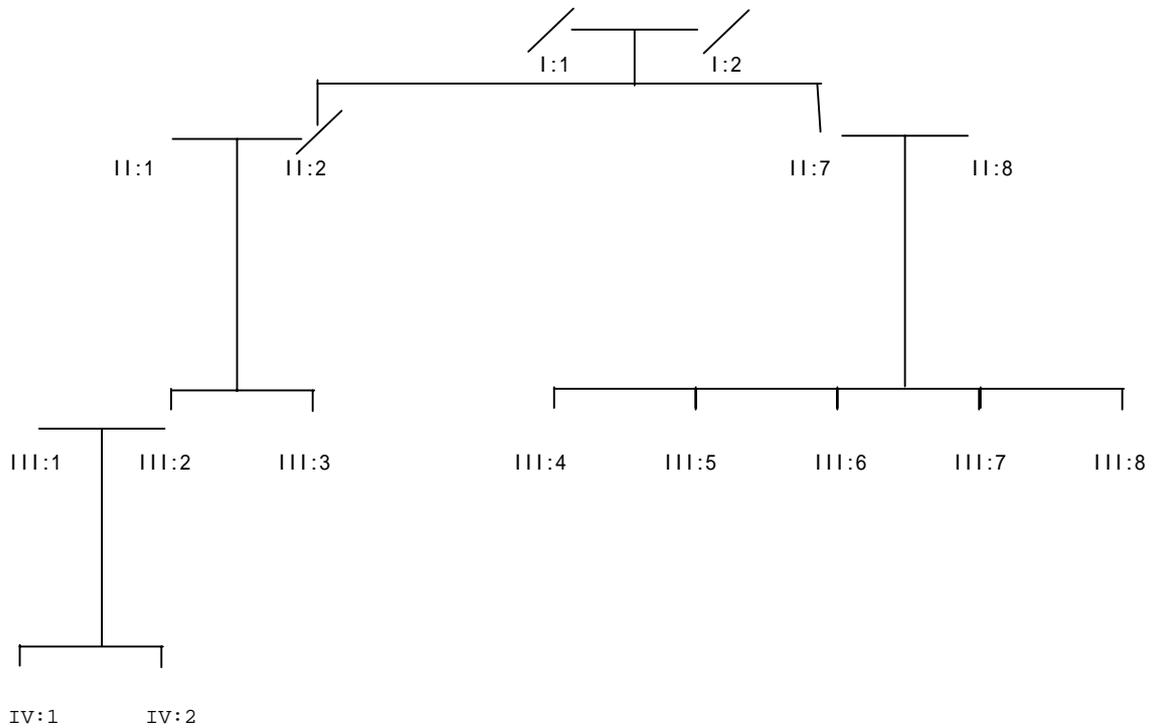
A. Non-penetrant: その疾患の原因の 1 つと考えられている遺伝子を受け継いでいるにもかかわらず、病気の症状を示していない状態。原因が複雑な疾患では non-penetrant はしばしば認められます。

B. Genetic heterogeneity: 異なる遺伝子の変異が同じ (同類の) 病気 (あるいは症状) を引き起こすことを指します。溶血性貧血、精神発達遅滞、糖尿病、小児症などにおいては症状は似ていても原因は様々です。

C. Phenocopy: ある遺伝子異常に引き起こされる病気と同じ症状を呈していながら、その遺伝子の異常を持たない状態。ある遺伝子変異の型をもつと糖尿病になります。しかし、多くの糖尿病はこれらの遺伝子に異常を持ちません。

D. Variable expressivity: ある遺伝子変異のタイプと関連して同じ病気でも症状等が異なる場合を指します。マルファン症候群を引き起こす fibrillin 遺伝子の変異はある家系において眼の異常を引き起こすし、他の家系では心臓の異常を引き起こします。

あなたはある常染色体優性遺伝の疾患の家系に関して7つの polymorphic marker の遺伝子タイピングを終了しました。Linkage analysis の前に以下の点について検討してみてください。



### 遺伝子タイピング解析結果

II:7	2/3	7/9	6/6	8/15	4/8	1/4	8/14
II:8	2/3	6/7	2/3	7/11	3/9	13/13	7/13
III:1	2/6	7/7	3/6	7/9	5/10	4/6	5/9
III:2	2/3	7/7	4/6	10/15	8/10	1/4	4/8
III:4	3/3	7/7	2/6	11/15	3/8	1/13	7/8
III:5	2/3	6/7	3/6	7/15	8/9	1/13	8/13
III:6	3/3	7/7	2/6	11/15	3/8	1/13	7/8
III:7	3/3	7/7	2/6	11/15	3/8	1/13	7/8
III:8	2/2	6/9	3/6	7/8	4/9	4/13	13/14
IV:2	3/6	7/7	3/6	7/15	5/8	1/6	5/8

1. 個々の haplotype を判る範囲で示してください。どちらが母親由来でどちらが父親由来かも示してください。I:1, I:2 についてもできるだけ推測してみてください。
2. どこで遺伝子組換えが生じていますか？
3. 疾患とリンクする haplotype があれば指摘してください。

1.

---

I:1	2 9 6 8 4 4 14	X X X X 10 6 9
I:2	3 7 6 15 8 1 8	X X X X X X X
II:1	2 3 3 7 6 13 6	2 7 4 10 10 4 4
II:2	3 7 6 15 8 1 8	X X X X 10 6 9
II:3	2 9 6 8 4 4 14 Father	3 7 6 15 8 1 8 Mother
II:4	3 7 2 11 3 13 7	2 6 3 7 9 13 13
III:1	2 7 6 7 5 6 5	6 7 3 9 10 4 9
III:2	2 7 4 10 10 4 4 Father	3 7 6 15 8 1 8 Mother
III:3	2 3 3 7 6 13 6	3 7 6 15 R 10 6 9
III:4	3 7 6 15 8 1 8 Father	3 7 2 11 3 13 7 Mother
III:5	2(3) 7 6 15 8 1 8 Father	3(2) 6 3 7 9 13 13 Mother
III:6	3 7 6 15 8 1 8 Father	3 7 2 11 3 13 7 Mother
III:7	3 7 6 15 8 1 8 Father	3 7 2 11 3 13 7 Mother
III:8	2 9 6 8 4 4 14 Father	2 6 3 7 9 13 13 Mother
IV:2	6 7 3 R 7 5 6 5 Father	3 7 6 15 8 1 8 Mother

---

3 7 6 15 haplotype が病気の遺伝子を運んでいる可能性が高いと考えられます。

▶ のところで遺伝子組換えが発生しています。

Linkage の程度をみるのに Maximum Lod Score を用いて判定します。一般的には  $\text{Mas Lod} \geq 3.0$  をもって有意な linkage ありとみなします。その詳細については、本書のカバーすべき範囲を超えるため解説しません。

高血圧と関係しそうな 2 種類の SNPs について検討してみました。

Polymorphism A

	Controls	Cases
Genotype 1/1	42	79
Genotype 1/2	89	109
Genotype 2/2	54	22
Total	185	210

Polymorphism B

	Controls	Cases
Genotype 1/1	10	25
Genotype 1/2	66	78
Genotype 2/2	109	107
Total	185	210

1. case と control で allele の頻度はいくつですか？
2. これらの 2 つの polymorphism が病気と関係しているか調べるにはどのようなテストを行いますか？
3. 相関はどの程度ですか？

1.

Polymorphism A

allele	cases	control
1	267	173
2	153	197
total	420	370

2. この例は親の情報がないので遺伝情報を単なる疫学の exposure の 1 つとして扱わざるを得ません。Case-control study, Odds ratio などで検討します。

3.

```
. cci 267 153 173 197
```

	Exposed	Unexposed	Proportion	
			Total	Exposed
Cases	267	153	420	0.6357
Controls	173	197	370	0.4676
Total	440	350	790	0.5570
	Point estimate		[95% Conf. Interval]	
Odds ratio	1.987193		1.494835	2.641722 (Cornfield)
Attr. frac. ex.	.4967775		.3310298	.6214591 (Cornfield)
Attr. frac. pop	.3158086			
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+				
chi2(1) = 22.54 Pr>chi2 = 0.0000				

Polymorphism B

