

ゲノム

1 細胞機能、発生・分化のプログラム

2 情報の次世代への伝播

- 1) 遺伝学の歴史
- 2) ゲノム構造の基本
- 3) 遺伝学の基礎知識
- 4) 臨床遺伝学研究方法論

遺伝子の概念・歴史

1) 因子としての遺伝子

1865年:メンデル Mendel

初めて仮定の概念として遺伝子が遺伝学に導入された(単因子遺伝)。

1883年:ゴルトン Galton

両親の身長(多因子形質)と子の身長の関係を調査。親子の身長をグラフにプロットする(横軸に親、縦軸に子)と直線が得られ、その勾配は45度より小さく回帰係数は0.325が得られた。子の身長は両親から等量の寄与を受けると考えられるから、期待される回帰係数は0.5である。これらの比は今日でいう遺伝力で、ゴルトンの例では0.65となる。残りの35%は環境要因で身長が決まると考えることができる(多因子遺伝)。

1885年:ダーウィン Darwin

「種の起原」で効果の小さな有利な突然変異の蓄積が生物の進化で重要なことを強調している。

1918年:フィシャー Fisher

多因子形質は多数の単因子と環境要因で説明できることを理論的に示した。

2) 遺伝子の物理

遺伝子の担体としての染色体に関する研究

1912年: 染色体の交叉、遺伝連鎖、染色体の不分離(モルガン Morgan 一派)

1915年: モルガン Morgan 一派はショウジョウバエを用いた多年にわたる系統的研究から、遺伝子は染色体上に並んでいることを明らかにした(遺伝子の染色体理論)。

1927年: 遺伝子が変わることはドブリースが見つけ **mutation** という用語を使った(1901年)が、マラー Muller はX線で突然変異がショウジョウバエで誘発されることを明らかにした。

1956年: ヒト染色体の本数は46である(Tijo & Levan)

1962年: X染色体の不活性化(Lyon)

1974年: 染色質構造、ヌクレオソーム(Kornberg, Olins & Olins)

3) 遺伝子の生化学

- 1928年: グリフィス Griffith は肺炎双球菌株の無細胞抽出液により、遺伝する変化が肺炎双球菌に生じることを観察した(形質転換)。
- 1944年: アヴェリー Avery、マクラウド MacCleod とマクカーティ McCarty はDNAが細菌の遺伝情報を担っていることを解明した。
- 1952年: ハーシェイ Hershey とチェイス Chase が遺伝情報はDNAによって伝えられることを証明した(バクテリオファージの実験)
- 1953年: ワトソンとクリックが**DNAの二重らせん**モデルを提唱した。
- 1977年: ギルバートらが、ヒトを含めた哺乳類の**遺伝子が分断**されていることを発見。
- 1988年: ヒトゲノムプロジェクトの開始

- 1) 遺伝学の歴史
- 2) **ゲノム構造の基本**
- 3) 遺伝学の基礎知識
- 4) 臨床遺伝学研究方法論

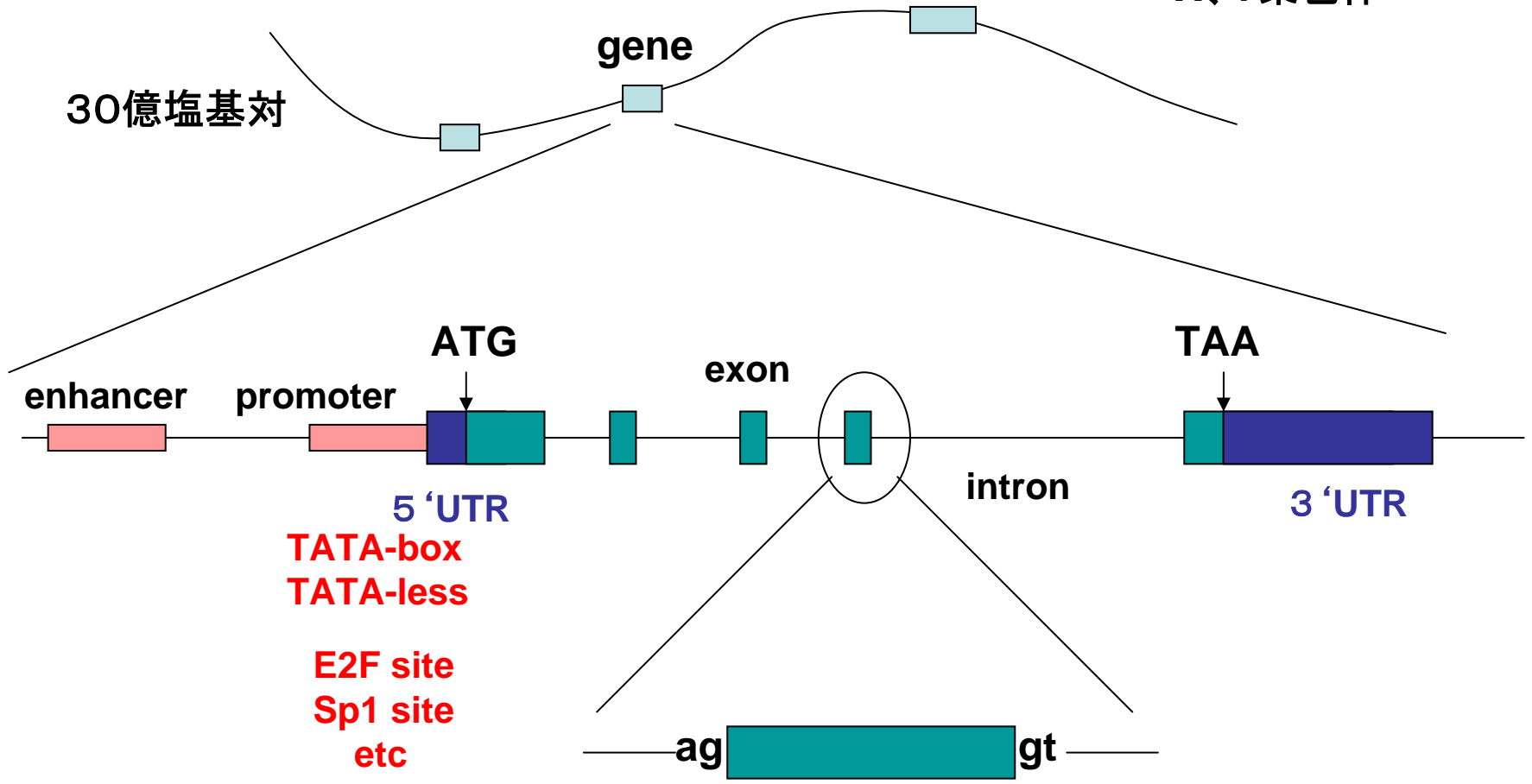
ゲノムの構造

約3万2千個の遺伝子

22常染色体
X、Y染色体

30億塩基対

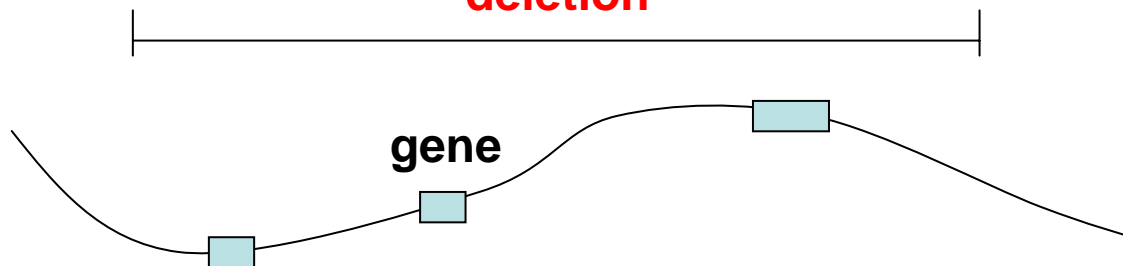
gene



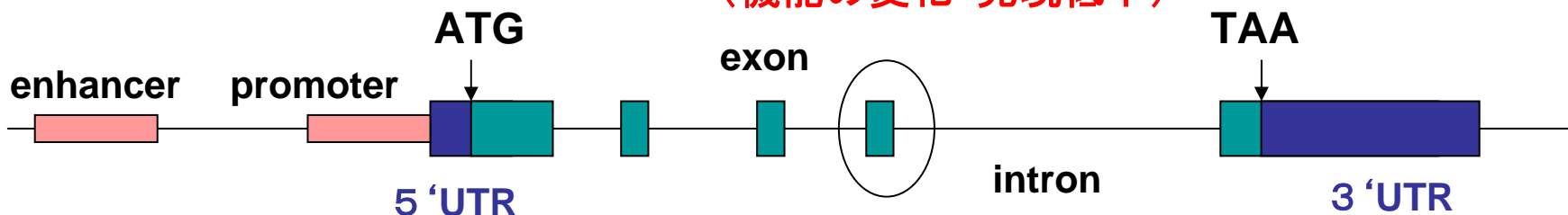
Exon-intron junction (Chambon's rule)

ゲノムの構造と異常

deletion 発現無し



異常タンパク分子の発現
(機能の変化・発現低下)



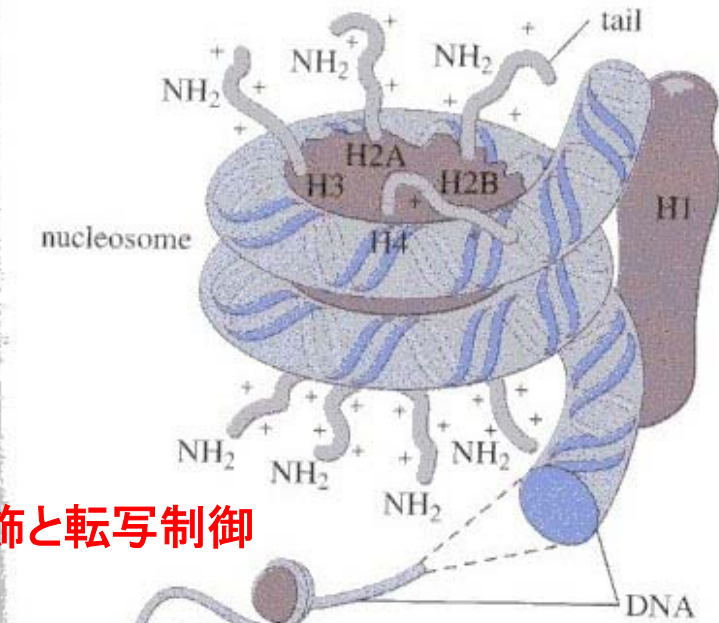
転写抑制
転写活性化

mRNAの安定性の変化

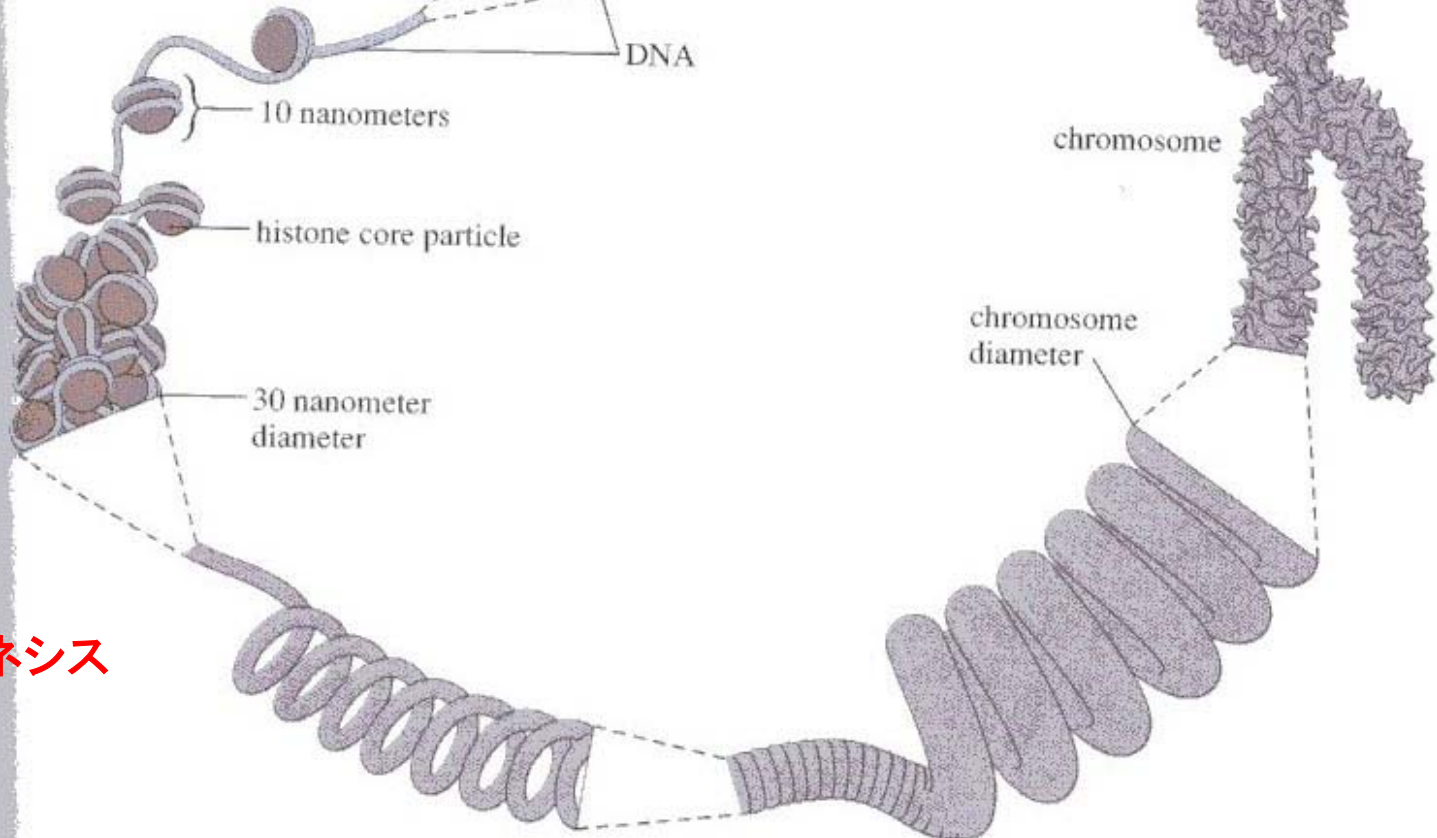
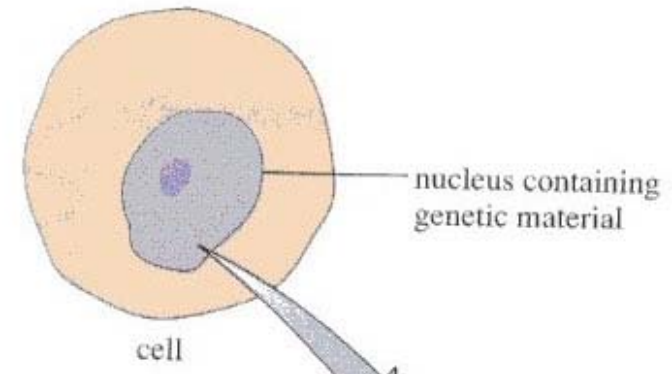
スプライシング異常



ゲノムを立体構造で考えると・・・



ヒストン修飾と転写制御

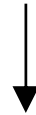


エピジェネシス

既知遺伝子と一致する遺伝子 14,882(17,764)

+

コンピュータで予測された遺伝子 16,896 (~21,350)



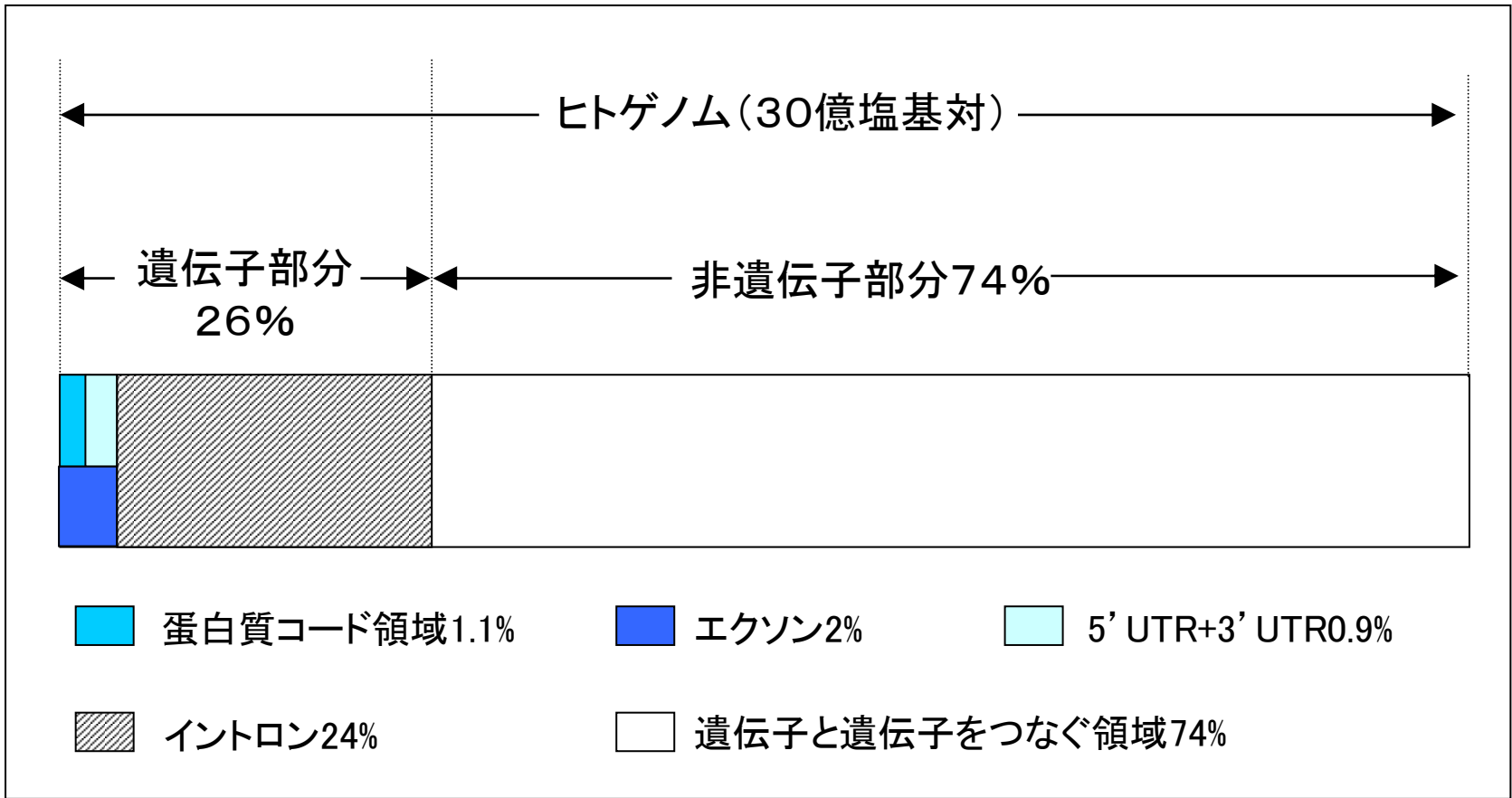
ヒトの蛋白質をコードする遺伝子 31,778
(機能未知は60%)

蛋白質をコードしていないRNA遺伝子 740以上

トランスファーRNA 497

核内低分子RNA 78

ヒトの遺伝子数(国際コンソーシアム)



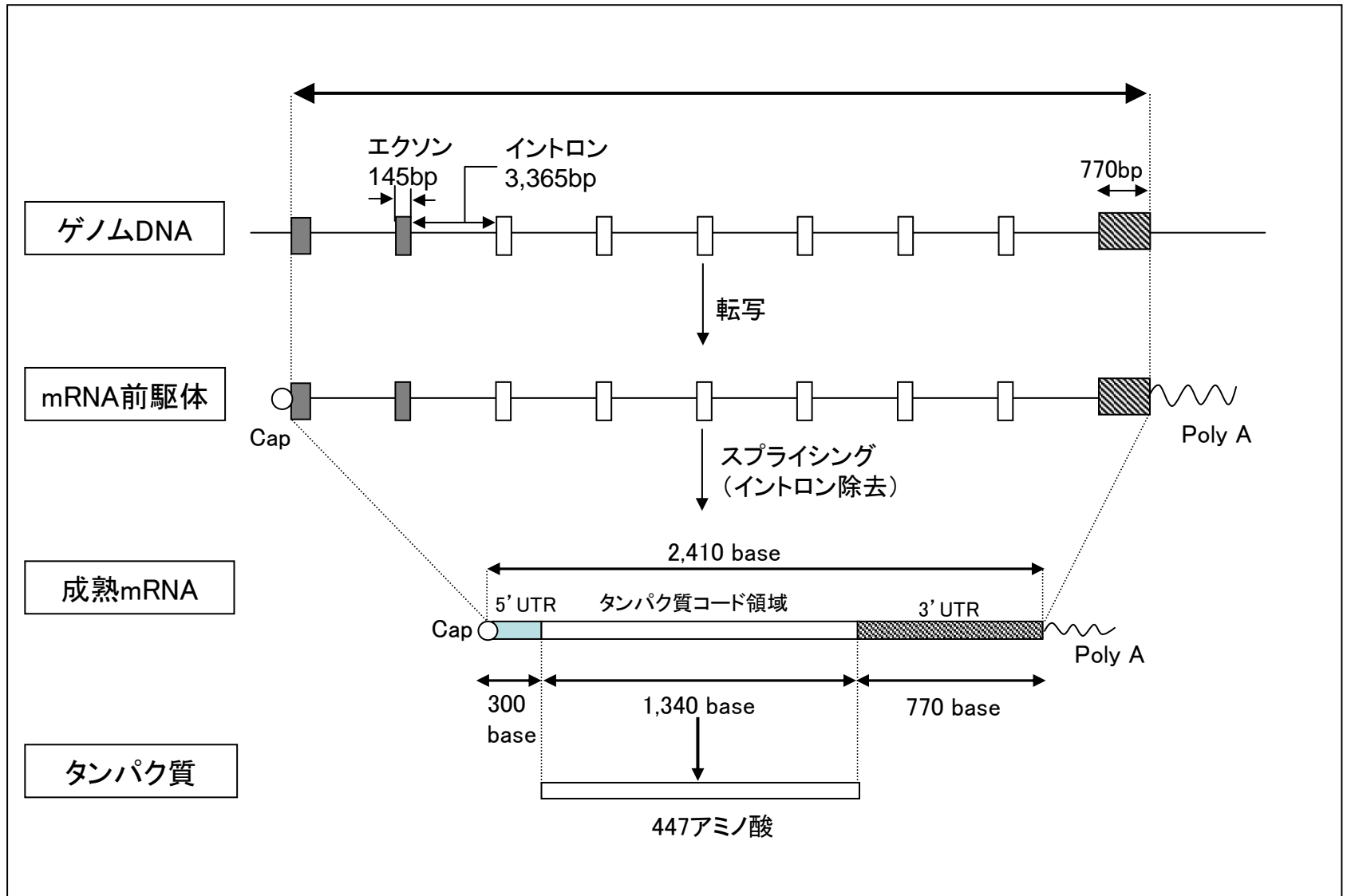
ヒト遺伝子部分がゲノム全体に占める割合

ヒト遺伝子の特徴

	平均値
エクソンの長さ	145bp(48コドン)
エクソン数	8.8個
イントロン数	7.8個
イントロンの長さ	3,365bp
3'UTRの長さ(①)	770bp
5'UTRの長さ(②)	300bp
コード配列(CDS)の長さ(③)	1,340bp(447アミノ酸)
成熟mRNA(①+②+③)	2,410bp
ゲノム上での遺伝子の大きさ	27kbp
コード配列が遺伝子全体に占める割合	約5%
遺伝子の密度	11.1個/Mb
遺伝子が最も密に存在する染色体	19番染色体(26.8個/Mb)
	19番染色体(23個/Mb)(セセラ)
遺伝子が最もまばらに存在する染色体	Y染色体(6.4個/Mb)
	Y染色体(5個/Mb)(セセラ)
	13番染色体(5個/Mb)(セセラ)
遺伝子1個に対して存在するmRNAの種類数	5個

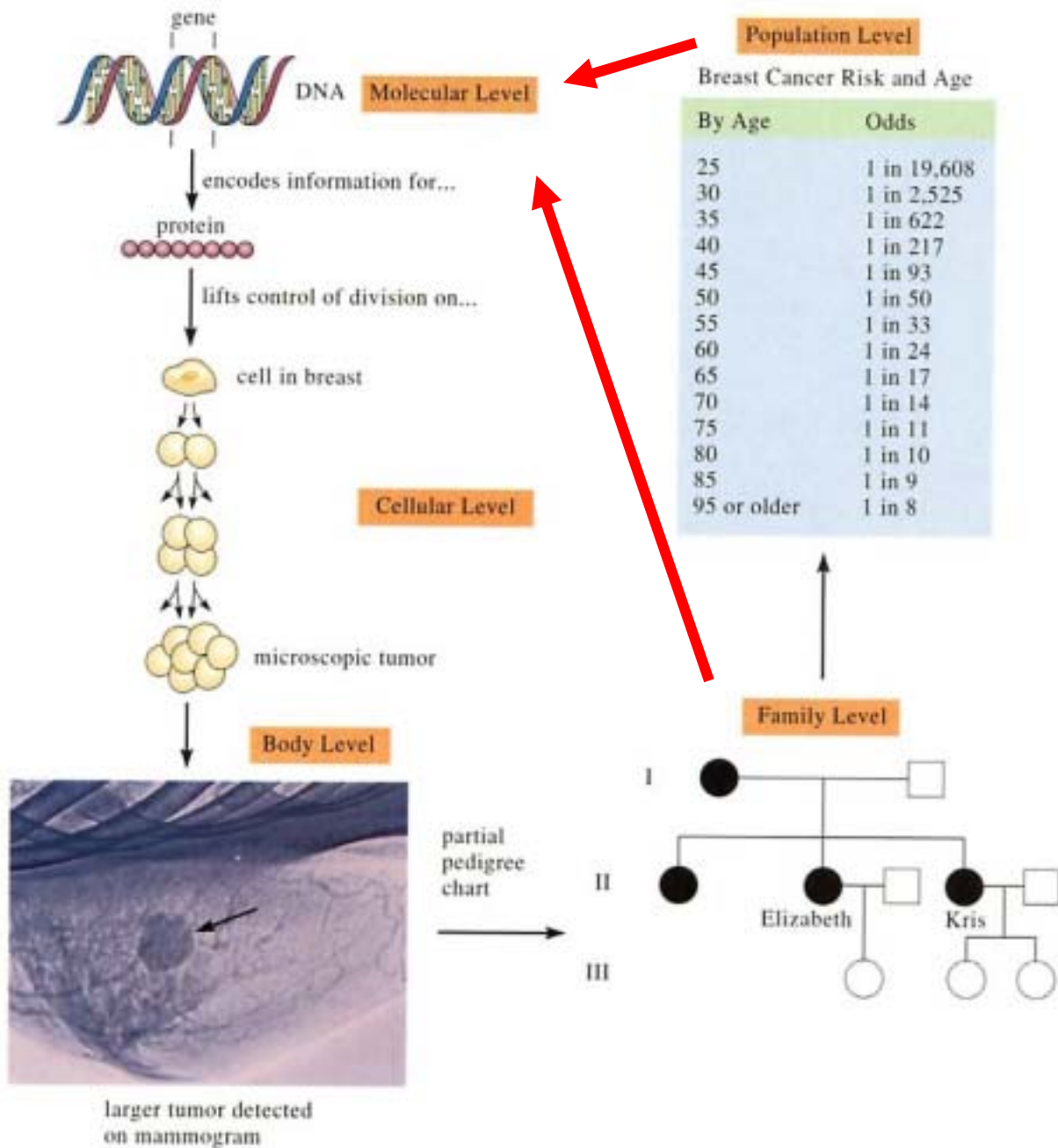
UTRの長さは実際より小さく見積もられていると考えられる

平均値に基づくヒト遺伝子の構造



ヒト遺伝子の大きさの平均値は27kbpで、そのうち9%にあたる2,410baseがエクソンとして成熟mRNAを形成する。タンパク質をコードしているのは遺伝子全体の5%にあたるわずか1,340baseである。ただし、この数字はあくまでも平均値であり、個々の遺伝子の構造は多様である。

- 1) 遺伝学の歴史
- 2) ゲノム構造の基本
- 3) 遺伝学の基礎知識**
- 4) 臨床遺伝学研究方法論



遺伝学的アプローチの特徴

遺伝子
タンパク



形質(病気)

知識を基礎にしたアプローチ
分子(多くの場合はタンパク)の機能を材料
に遺伝子と形質を結びつける。



遺伝子

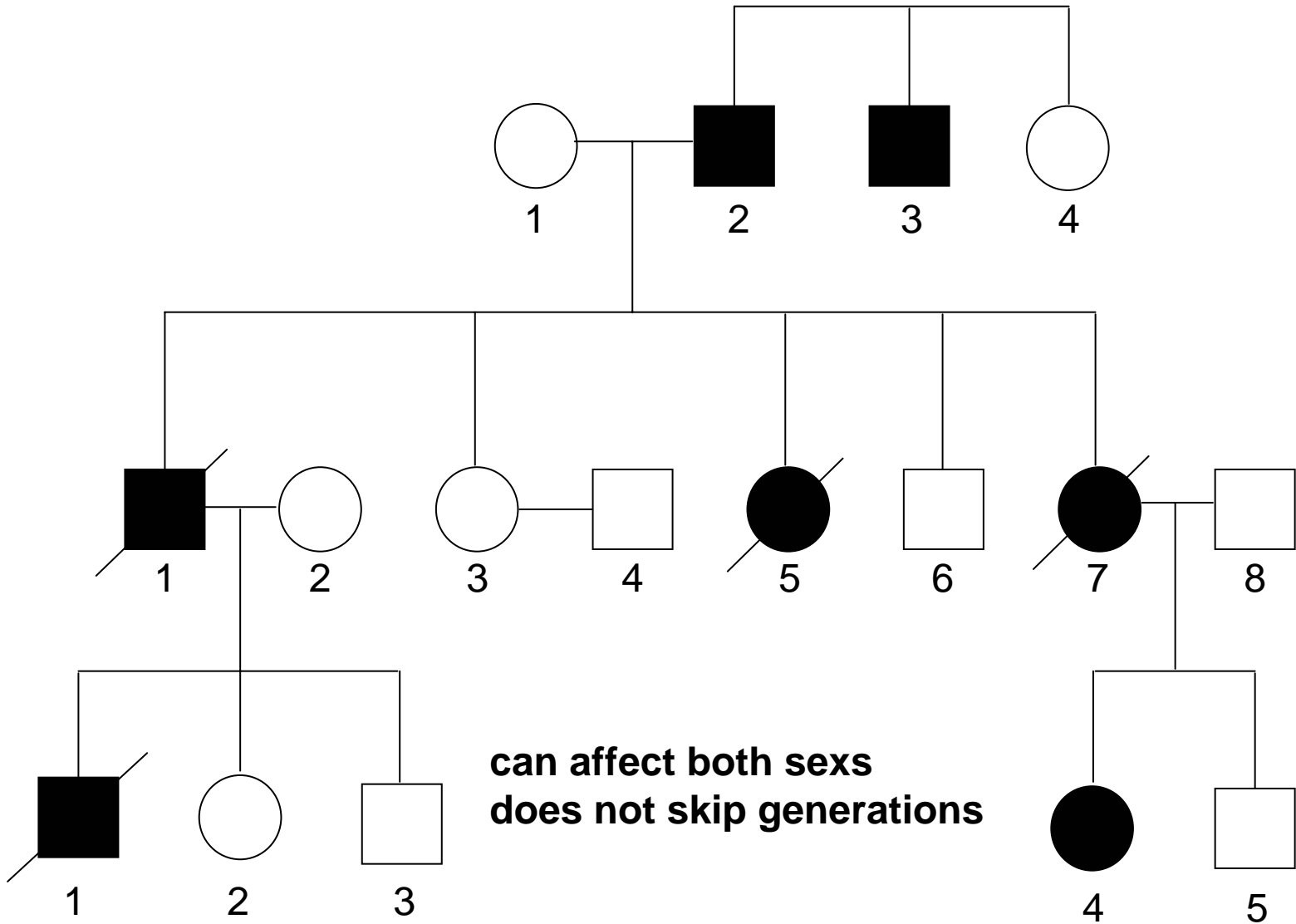


形質(病気)

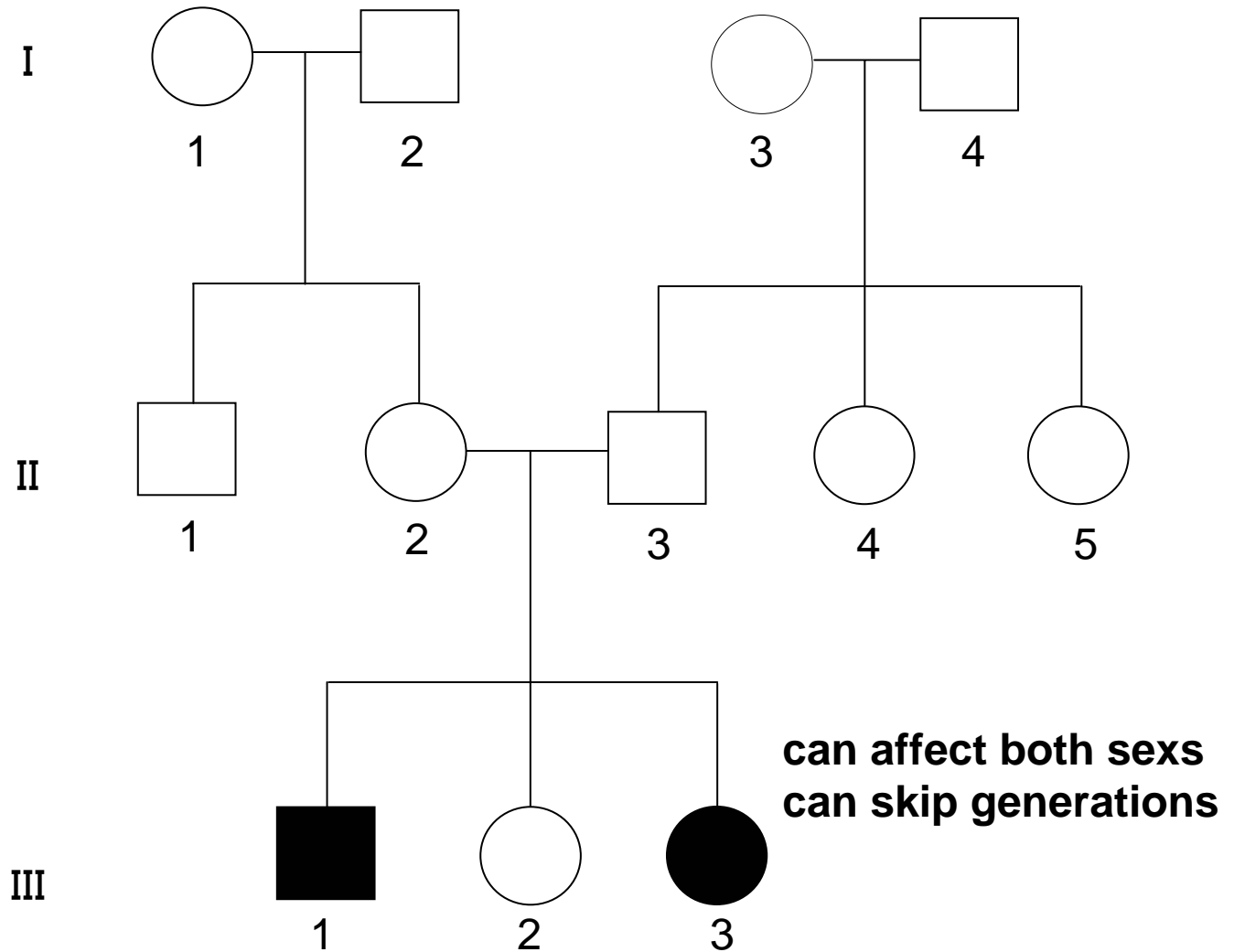
統計学を基礎にしたアプローチ
遺伝的多型を材料に形質と配列(遺伝子)を
結びつける。

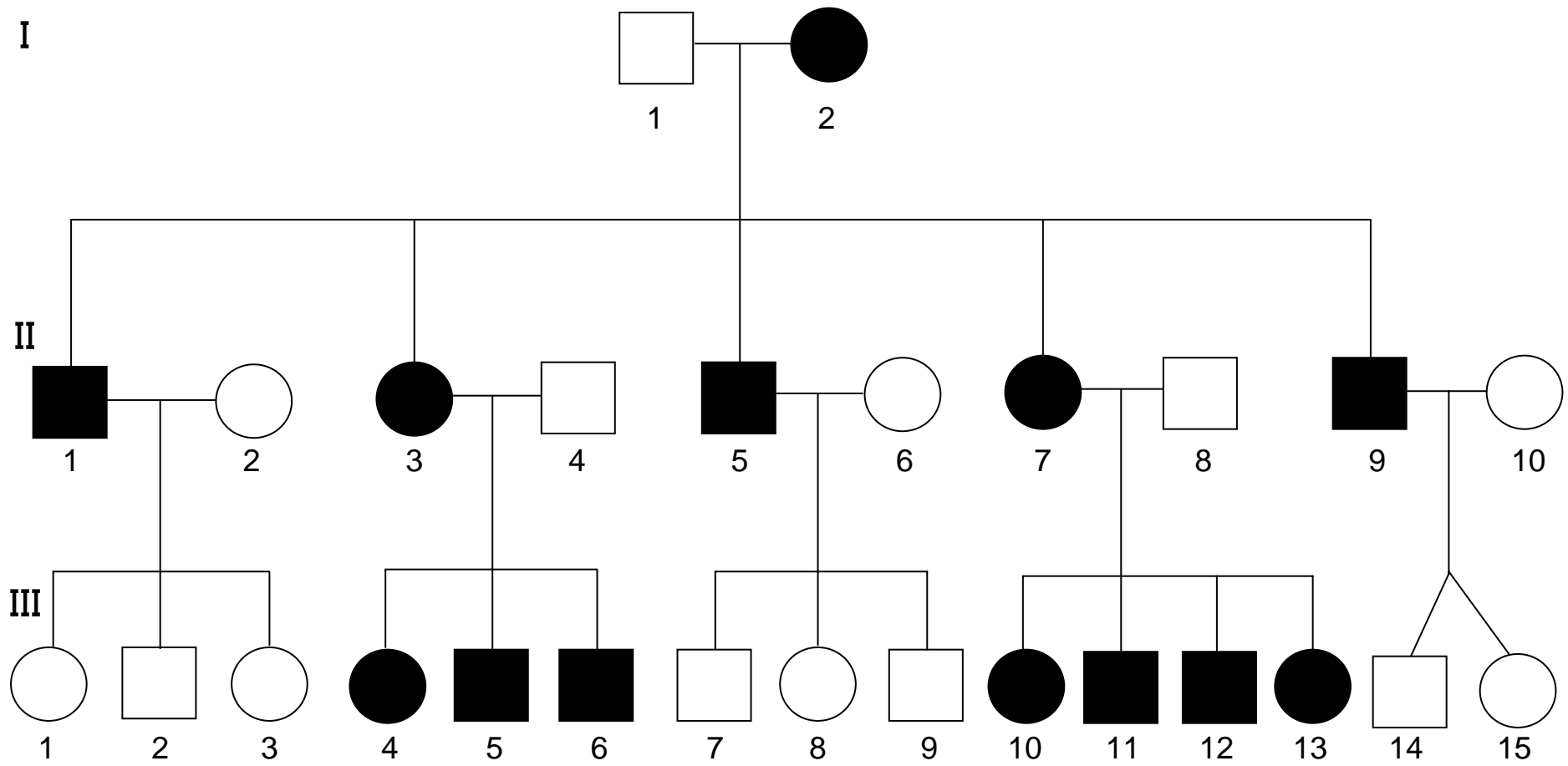
遺伝の本質の一つは、それが確率事象であることである。
決定論では扱うことができず、よって統計学的手法が必要となる。

A pedigree for autosomal dominant trait



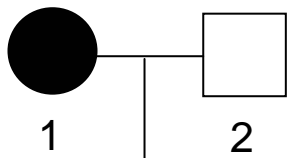
A pedigree for autosomal recessive trait



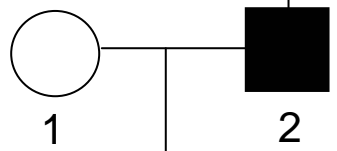


Mothers pass the gene to all offsprings
Fathers do not transmit the gene

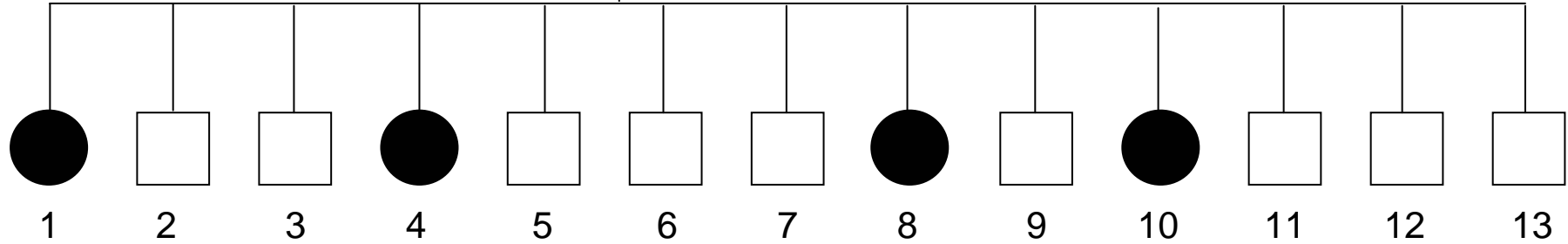
I



II



III



Affected male in the 2nd generation has passed the gene to all of his daughters and none of his sons

遺伝マーカー

マイクロサテライト多型

短い塩基モチーフ（通常 2 - 4 塩基）の反復数の多型

10^{-3} から 10^{-5} の変異率

1 ゲノムあたり 10 万ヶ所以上

1 塩基多型 (SNP)

単一塩基の違い

10^{-8} から 10^{-9} の変異率

1 ゲノムあたり 100 万ヶ所以上

多型：最大頻度を示す対立遺伝子の頻度が 99% 以下

遺伝学の基礎知識

[Phenotype (表現型) と Trait (形質)]

[Locus (遺伝子座、または座位)]

[Genotype (遺伝子型)]

[Allele (対立遺伝子)]

[Penetrance (浸透率)]

[Linkage (連鎖)]

[Recombination (組み換え)]

[Recombination fraction (組み換え割合)]

[Haplotype (ハプロタイプ)]

[Linkage Disequilibrium (連鎖不平衡)]

[Phenotype (表現型) と Trait (形質)]

個体レベルで観察されるものを示す言葉。

形質とは**カテゴリー**のこと。

表現型は**その形質において観察される個々のタイプ**。

「エンドウ豆の形」という形質の中に、「丸い」、「皺がある」という表現型が存在する。「病気であるか無いか」が形質であり、「病気である」、「病気でない」という表現型がある。

しばしば表現型と形質とは同じ意味で用いられる。

表現型、形質は最も我々の実生活と関係するものである。しかし、それは遺伝の本質と一対一に対応するものではない。

[Locus (遺伝子座、または座位)]

一つの交配集団に共通して、多数の遺伝子座が存在すると考える。一つの個体については、一つの遺伝子座に一つの遺伝子型が存在する。

染色体による表現では遺伝子座は交差が起きる最小単位の領域には含まれた部分である。分子生物学的には遺伝子座は一塩基である。

[Genotype (遺伝子型)]

遺伝子型は一つの遺伝子座について、一つの個体に存在する二つの対立遺伝子(常染色体遺伝子座の場合)の**組合せ**である。一方の対立遺伝子は母親から、別の対立遺伝子は父親から(有性生殖の場合)配偶子に運ばれて来る。また、次の世代では、遺伝子型を構成する一方の対立遺伝子を配偶子により子に伝達する。

メンデルの第一法則(優劣の法則)は遺伝子型と表現型の対応に関する法則である。即ち、第一法則は、個体の表現型が遺伝子型の関数であることを主張し、その関数に優性と劣性が(場合によっては共優性)が存在することを示す。現在では表現型は遺伝子型の関数で表される確率分布に従う事象であると解釈される。(即ち、個体の持つ各遺伝子型に**浸透率**が対応し、浸透率の確率で、個体はある表現型をとる)。

メンデルの第一法則によると個体において最も本質的なものは表現型ではなく、遺伝子型であり、遺伝子型は個体のレベルでは完全に安定である。

染色体のレベルでは、遺伝子型は二つの相同染色体上の相同部分に存在する二つの相同部位の**組合せ**である。分子生物学のレベルでは遺伝子型は多型のあるSNP、などの**組合せ**である。

[Allele (対立遺伝子)]

遺伝子型は個体のレベルでは安定であるが、世代を超えると安定ではない。個体を超えて安定して伝えられるのは遺伝子型ではなく対立遺伝子である。

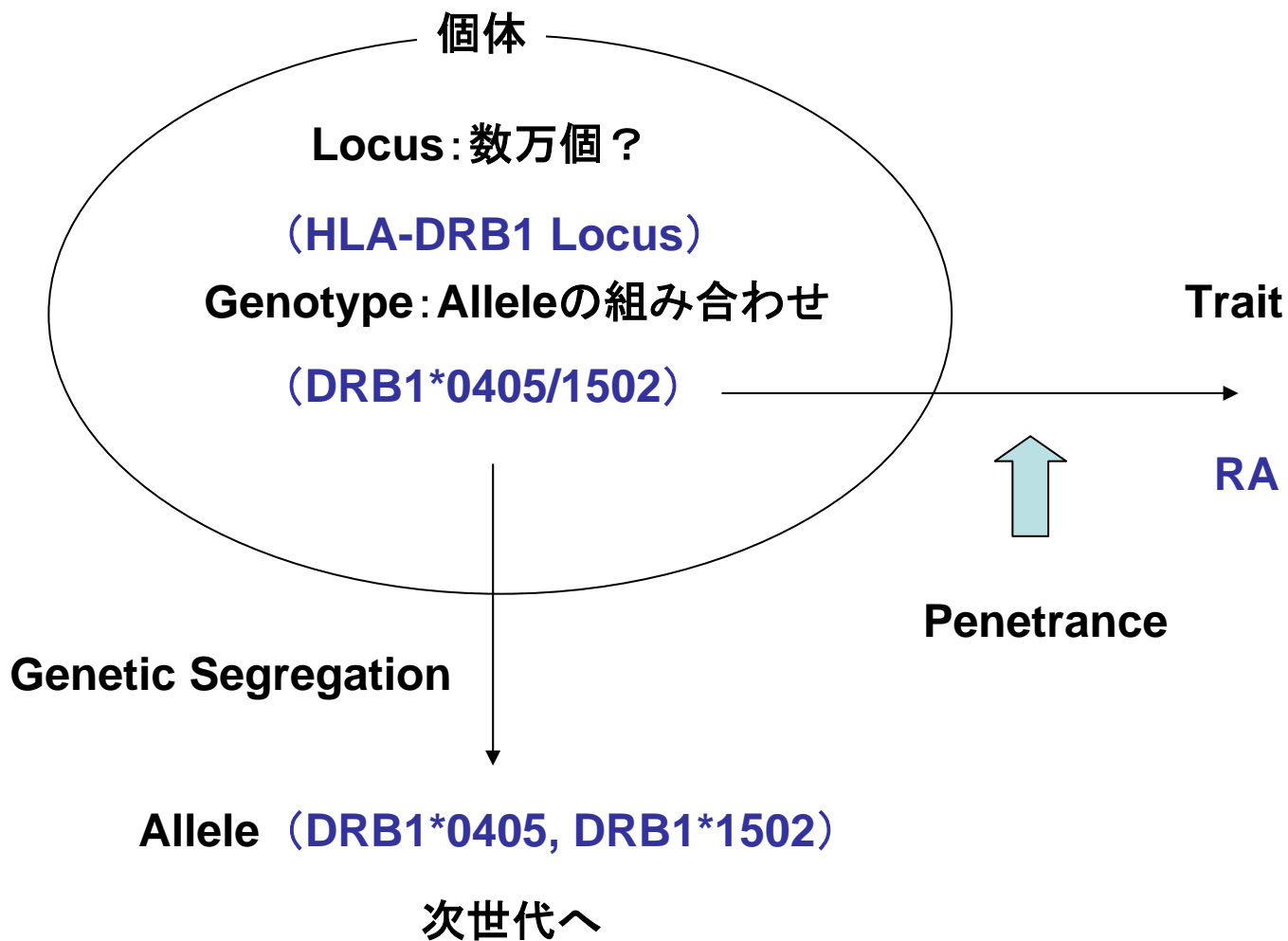
メンデルの遺伝法則の第二法則(分離の法則)は遺伝子型を構成する要素のうち、世代を超えて安定な単位に関する法則である。世代交代の際に遺伝子型を構成する対立遺伝子が一個だけ、同確率で次世代に伝えられることを示している。そのため、世代を超えては対立遺伝子が最小の安定した単位である。

世代交代で対立遺伝子が一個だけ、同確率で伝えられるということは、一つの対立遺伝子が伝えられる確率が $1/2$ という事である。遺伝統計学の成功は、対立遺伝子伝達の事前確率が $1/2$ である事がメンデルの分離の法則により完全に保証されているという事に依存する。

染色体のレベルでは対立遺伝子は片方の染色体上の部位であり、分子生物学では片方の染色体上のSNPなどの多型である。

[Penetrance (浸透率)]

メンデルの第一法則は表現型が遺伝子型に規定されていることを主張するが、遺伝子型が同じ個体がすべて同じ表現型を持つのではない。例えば、ある遺伝子座があり、その対立遺伝子としてAとaがあるとする。Aが正常対立遺伝子、aが病気を起こす対立遺伝子とする。この疾患が優性遺伝で、この遺伝子座のみによって決まるとすると、Aaとaaの個体が患者となる。しかし、Aaの遺伝子型の個体がすべて患者となるのではなく、確率により患者となるかどうかが決まると考えるのが遺伝学の考え方である。**ある遺伝子型の個体がある表現型を取る確率を浸透率(Penetrance)という。**浸透率はそれぞれの遺伝子型に対応した 0-1の値を取る。



[Linkage (連鎖)]

連鎖とはメンデルの第三法則の例外である。メンデルの第三法則(独立の法則)は複数の遺伝子座の関係に関する法則である。異なる座位にある対立遺伝子の子への分配は互いに独立であることを示している。

染色体レベルでは、これは二つの座位が異なった染色体上にあるときにのみ正しい。同じ染色体上の二つの対立遺伝子は物理的に結合しているので、**つながったまま同じ配偶子に入り、子に伝わる**。即ち、第三法則が成立しない。しかし、同じ染色体上に存在する対立遺伝子が完全に結合して世代交代を続けるとは限らない。減数分裂の時に、染色体が交差して結合していた対立遺伝子の組合せが変化することがあるからである(組み換え)。組み換えの可能性は二つの座位間の距離が離れていればいるほど大きい。

遺伝子座の異なる遺伝子において必ずしもメンデルの独立の法則が当てはまらないことを示したのはモルガンである。即ち**連鎖 (Linkage)** の発見である。

しかし、異なった染色体上の遺伝子座では依然として独立法則は正しく、しかも連鎖の概念が成立するためには独立の法則の概念が存在することが前提であることに注意が必要である。独立の法則はメンデルの誤りであるという解釈は間違っている。

[Recombination (組み換え)]

連鎖している二つの遺伝子座の対立遺伝子について、親からもらった配偶子のそれぞれの対立遺伝子が、そのままの組合せで減数分裂の時、同じ配偶子に入り、子供に伝えられるか、別の親の対立遺伝子が同じ配偶子に入るかに関する概念。片親からもらった二遺伝子座の対立遺伝子の組合せがそのまま子に伝えられる場合、組み換えが起こらなかった、それ以外の場合(別々の親から来た対立遺伝子の組合せが子に伝わる)を組み換えが起こったと考える。

染色体レベルの表現では、1本の同一染色体上にある2つの遺伝子座の対立遺伝子、A、Bを考えたとき、奇数回の交差によって、それぞれの対立遺伝子が違う染色体上に存在部位が移ったとき、この2つの遺伝子座間には**組み換えが起こった**とされる。

一方、全く交差が起こらなかったとき、もしくは偶数回の組み換えが起こった場合、最終的には2つの遺伝子間の”状態”は変化しない(やはり同一染色体上にある事になる)。この場合2つの遺伝子間での**組み換えは起こらなかった**とされる。

[Recombination fraction (組み換え割合)]

1回の減数分裂において2座位間で組み換えが起こる確率。確率なので1以下であるのはもちろんだが、**通常 $0 \leq \theta \leq 0.5$ である**。なぜなら、座位間が離れれば交差で組み換えが起きる確率は高くなるものの、もう一度交差が起きて組み換えがもどる可能性も高くなるからである。

$\theta = 0.5$ のとき、自由組み換え(違う染色体に乗っている2つの遺伝子の組み換え)と同じと考えられる。

まれに、 $\theta > 0.5$ となる場合がある。これは、1回の組み換えが起こったときに、2回目の組み換えが抑制される現象(干渉, Interference)が起こる事により生じる。

組み換え割合は遺伝的距離と違って、単純に加減計算することはできない。A、B、Cがこの順に並んでおり、A-B、B-C間の組み換え割合が θ_1 、 θ_2 の時、A-C間の組み換え割合は、どちらかのみ組み換えが起きる確率の和なので、 $\theta_1(1 - \theta_2) + \theta_2(1 - \theta_1)$ である。

[Haplotype (ハプロタイプ)

連鎖が定義されると、遺伝子座1、遺伝子座2の各対立遺伝子の組み合わせが定義できる(Haplotype, ハプロタイプ)。連鎖する複数の座位の対立遺伝子のうち、一つの配偶子に存在するものの組合せがハプロタイプである。

古くからハプロタイプが問題となったのはHLAとRh血液型である。例えば、HLAではA, B, C, DR, DQ, 補体C2などの遺伝子座が連鎖しているため、一つの染色体で考えて、それぞれの遺伝子座のアレルが一塊となって世代を超えて伝えられる。

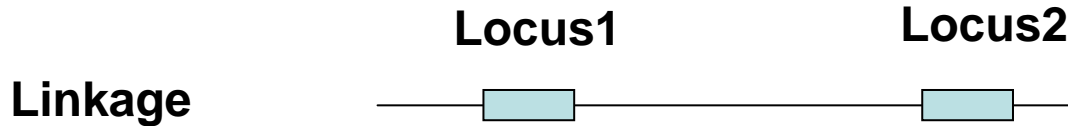
ハプロタイプは減数分裂での組み換えが無い限り保存される。組み換えが起きると新たなハプロタイプが形成され、それが配偶子により次世代に伝えられ、次の組み換えが起きるまで変化しない。

[Linkage Disequilibrium (連鎖不平衡)

連鎖不平衡とは連鎖のある2つ以上の遺伝子座に観察される現象である。

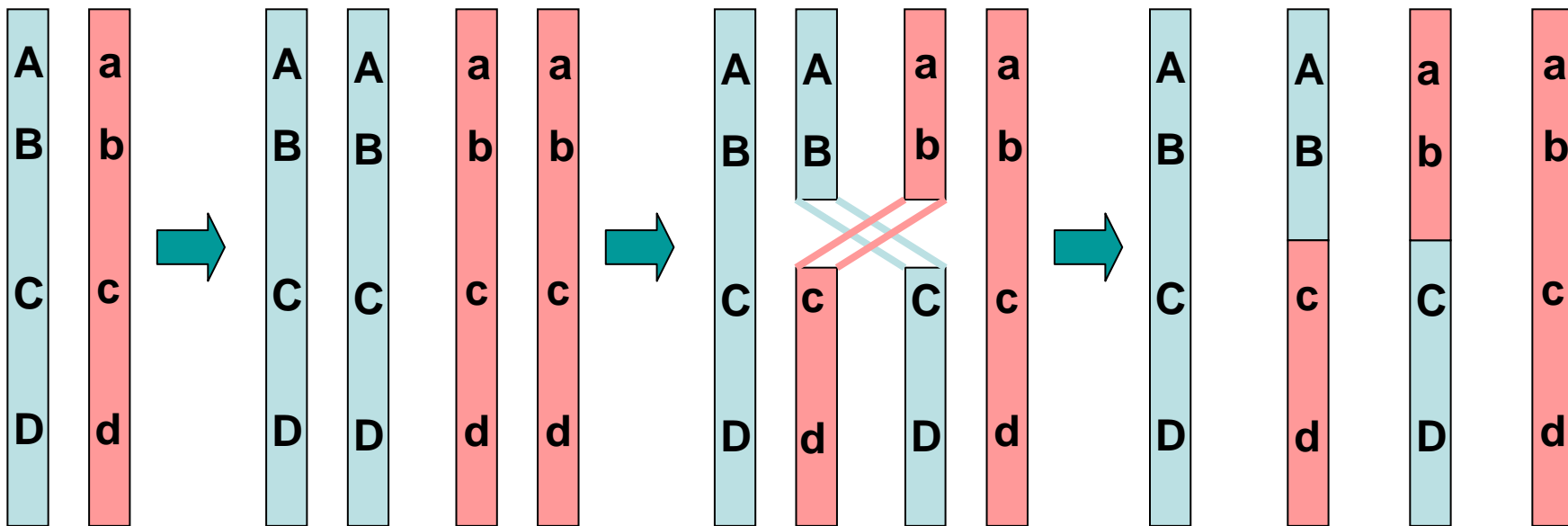
小学校の1年生100名が横一列に手をつなぎ校庭の端から端へ走って移動するとする。最初のうちは列は完全につながっているが、あそこで転びあちらで手が離れ、最後には列はバラバラになるであろう。それでもなお、数人づつの手はつながっている場合が多い。最初の列全体を一つの染色体と考えると最初出発してまもなくの間の現象が連鎖、最後の状態が連鎖不平衡と考えれば良い。

異なる遺伝子座間のアレルの非独立を言う。



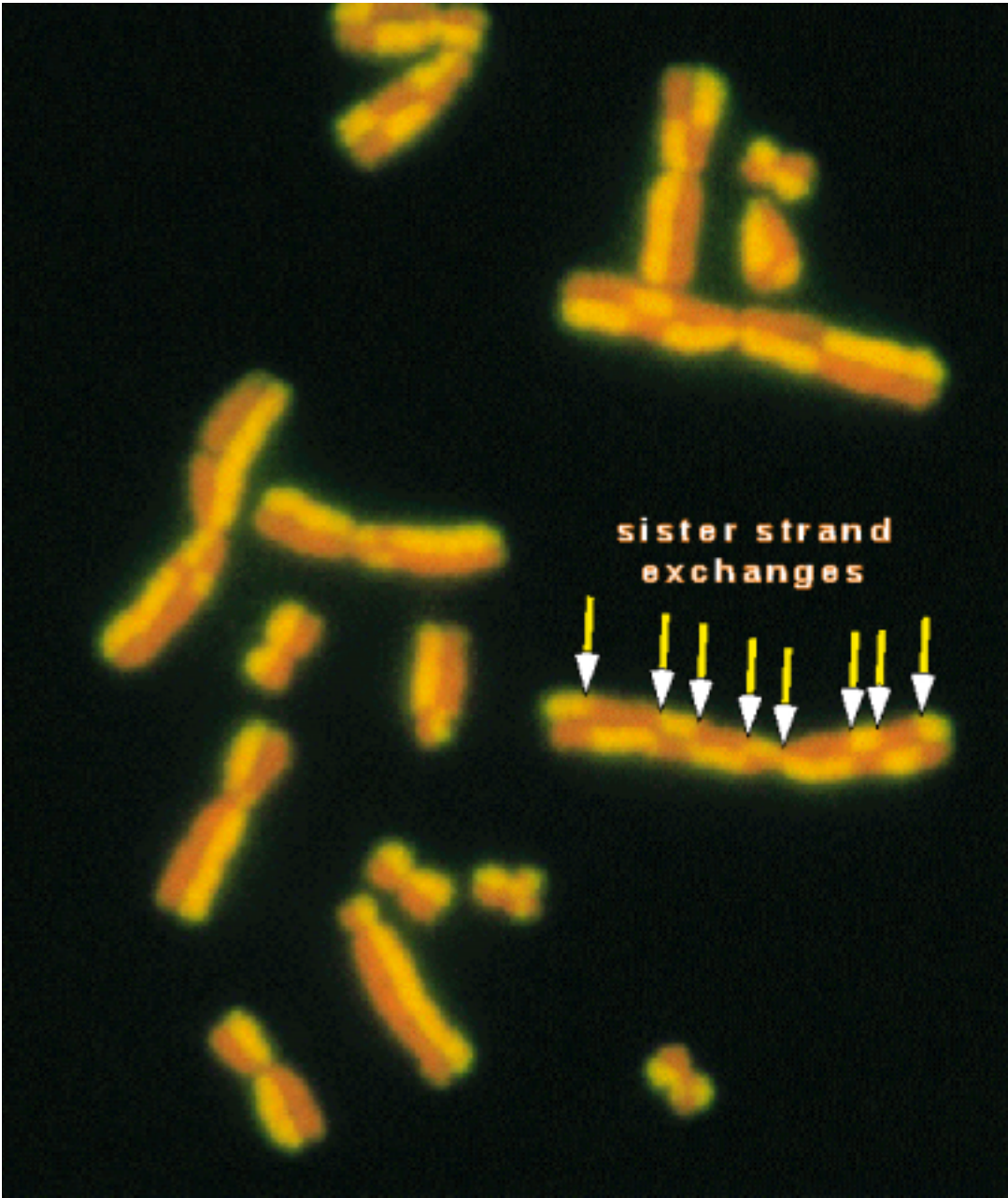
相同染色体

Recombination



配偶子

A-B-C-D haplotypeとa-b-c-d haplotype



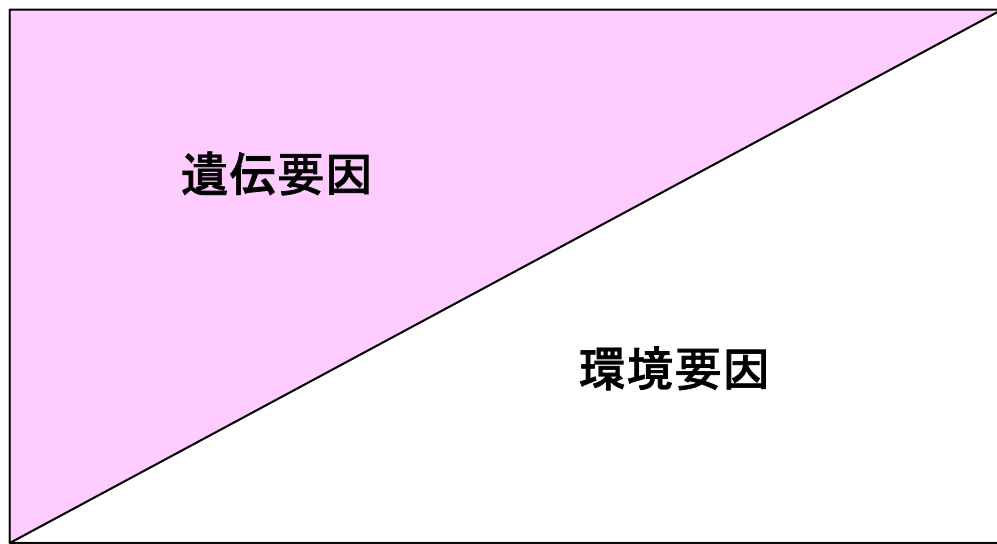
疾患の発症に遺伝要因が関与している証拠

1) 家族集積

2) 双生児研究

3) 同胞罹患率と一般集団罹患率の比較 (λ_s)

4) 遺伝マーカーとの相関



遺伝要因

環境要因

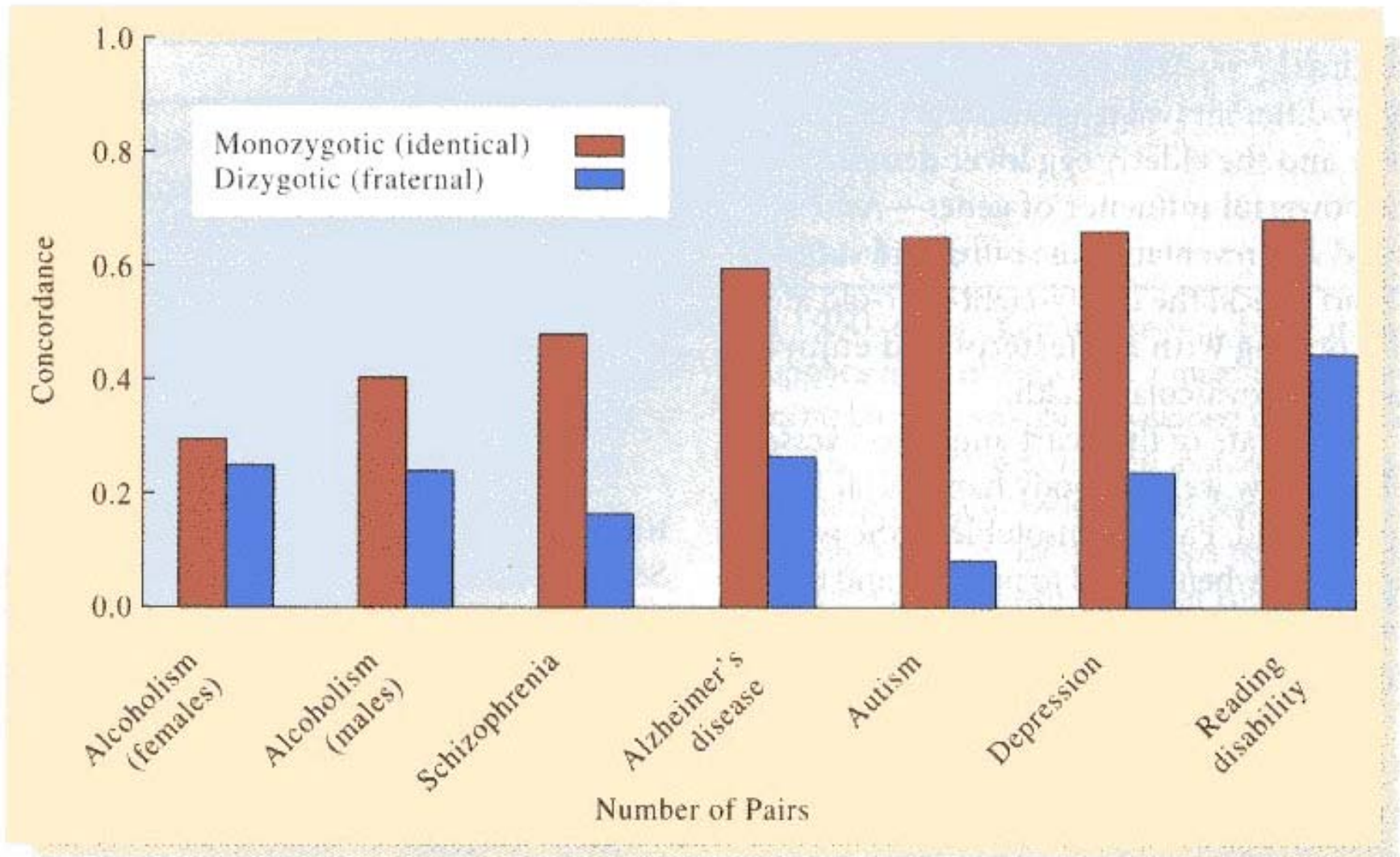
遺伝病
ハンチントン病など

生活習慣病
がん、高血圧、
糖尿病、痴呆、
ぜん息など

外傷など

疾患の成因

Twin study



Population frequencies and familial clustering of human autoimmune and immune-related diseases

Diseases	Population Frequencies (%)	Sibling Risk (%)	λ_s
Psoriasis	2.8	17	6
Rheumatoid arthritis	1.0	8	8
Graves disease	0.5	7.5	15
Type 1 diabetes	0.4	6	15
Ankylosing spondilitis	0.13	7	54
Multiple sclerosis	0.1	2	20
Ulcerative colitis	0.1	1.2	12
Lupus (SLE)	0.1	2	20
Crohn's disease	0.06	1.2	20
Narcolepsy	0.06	1	12
Celiac disease	0.05	3	60
Primary biliary cirrhosis	0.008	0.8	100

Vyse, TJ and Todd, JA, reviewed in Cell, 1996

- 1) 遺伝学の歴史
- 2) ゲノム構造の基本
- 3) 遺伝学の基礎知識
- 4) **臨床遺伝学研究方法論**

臨床遺伝学研究方法論

1) 対象疾患は？

単一遺伝病？

多因子疾患？

連続形質？

不連続形質？

2) 連鎖解析？ 相関解析？

パラメトリック連鎖解析

ノンパラメトリック連鎖解析

罹患者同胞対法

多点解析・単点解析

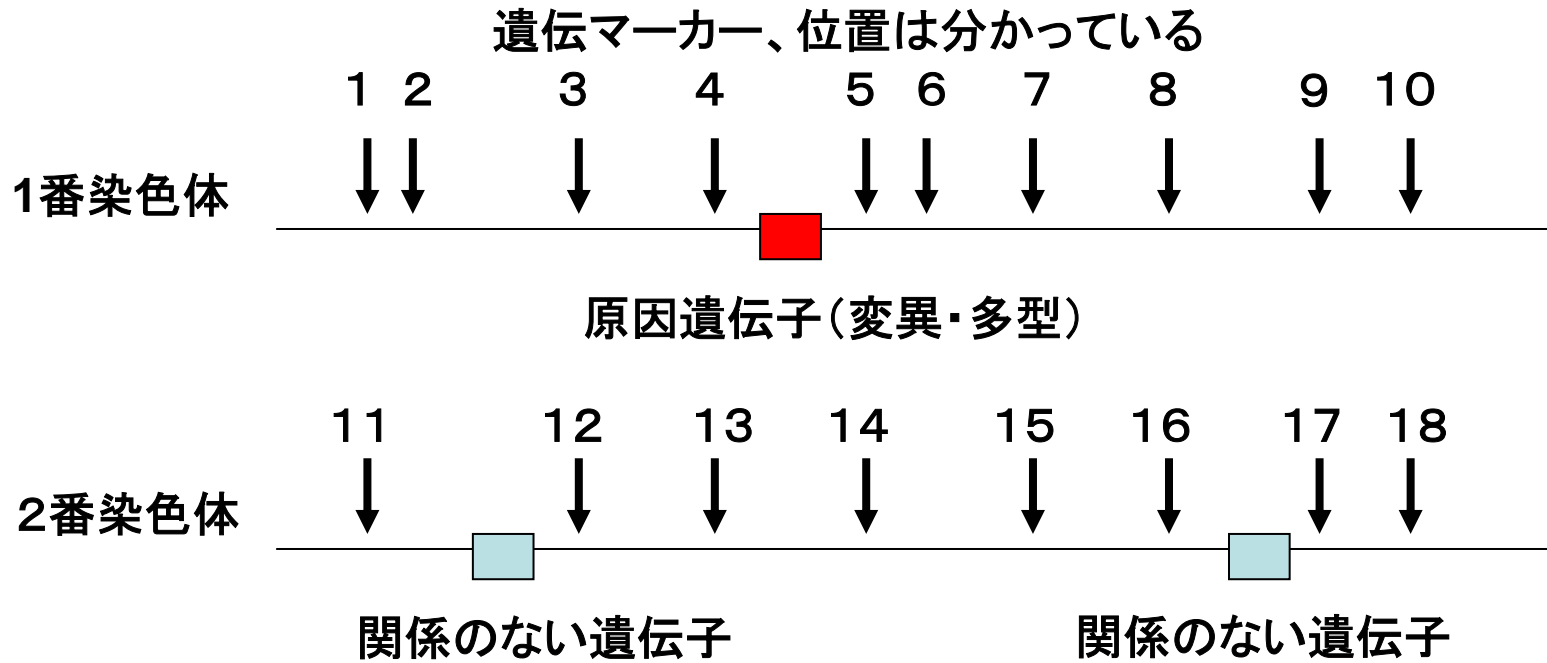
相関解析

患者-対照解析

Transmission Disequilibrium Test

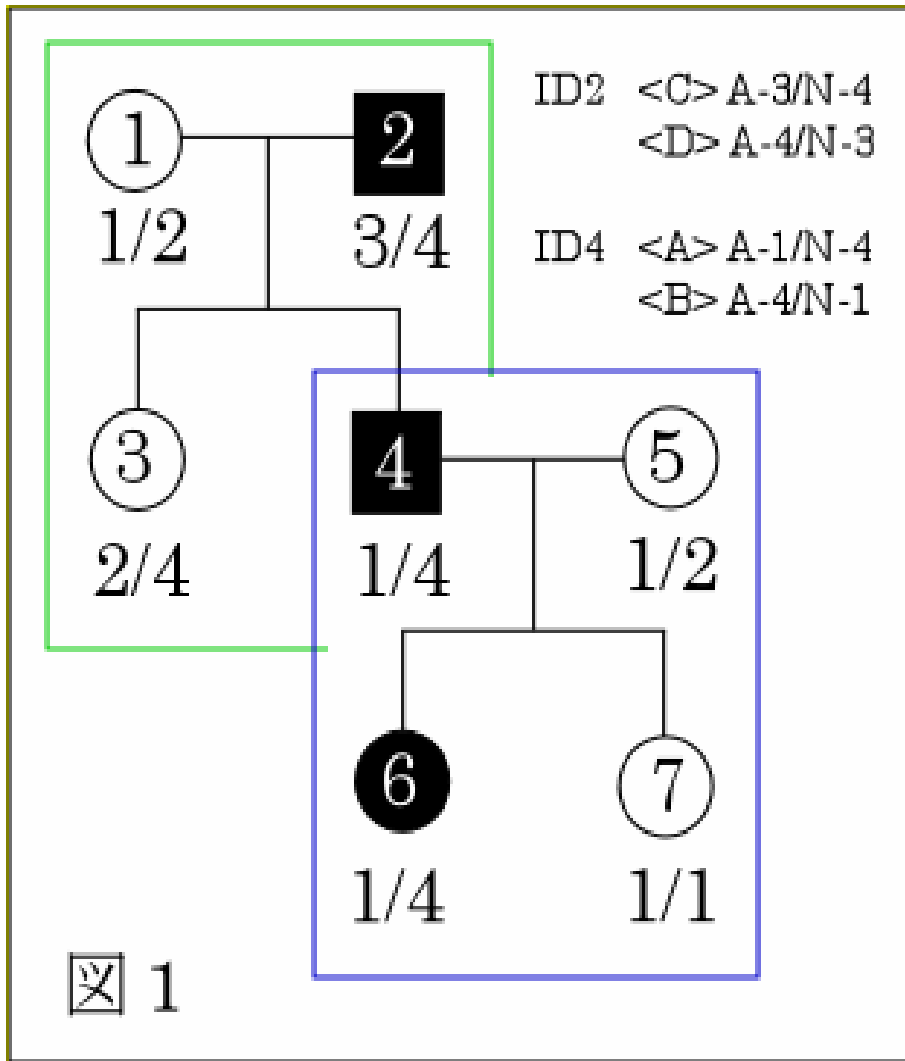
3) 遺伝マーカーは？

マイクロサテライトマーカー、SNP



原因遺伝子は遺伝マーカー1-10と連鎖している。
 原因遺伝子は遺伝マーカー4、5の近傍に存在する。
 遺伝マーカー11以降は原因遺伝子と連鎖していない。

パラメトリック連鎖解析の例



A: 原因遺伝子

N: 正常

AA、ANが発症(浸透率1、常染色体完全優性)

マーカーと原因遺伝子との組み換え率: θ

患者4に注目(ANである)

パターン<A>: A-1とN-4の場合

患者6はA-4、N-1である。

その尤度は、

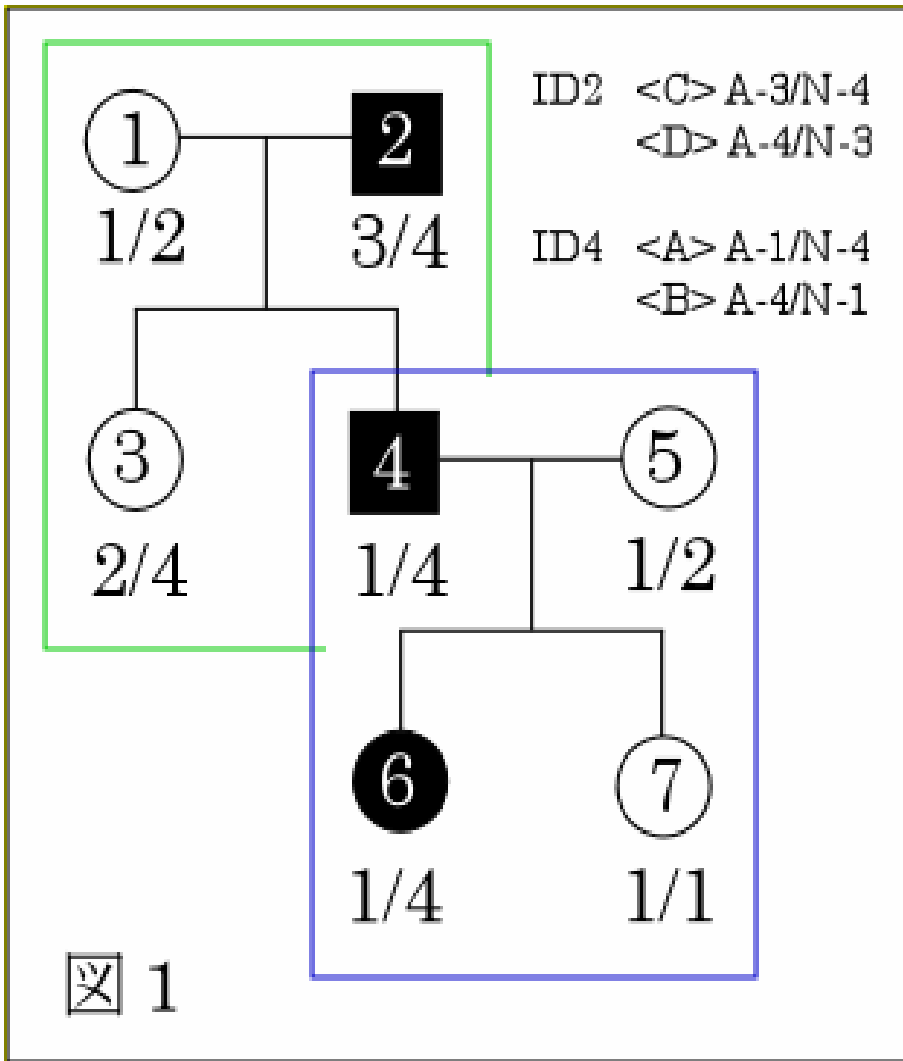
① 父親でのハプロタイプが組み換えを起こす確率: θ

② 父親の組み換えハプロタイプから (A-4) が選ばれる確率: $1/2$

③ 母親から、マーカー対立遺伝子, アレル1 (N-1) が選ばれる確率: $1/2$

により, $(1/2)^2 \theta$ と算出される。

パラメトリック連鎖解析の例



正常7はN-1、N-1である。

その尤度は、

①父親でのハプロタイプが組み換えを起こす確率: θ

②父親の組み換えハプロタイプから (N-1) が選ばれる確率: $1/2$

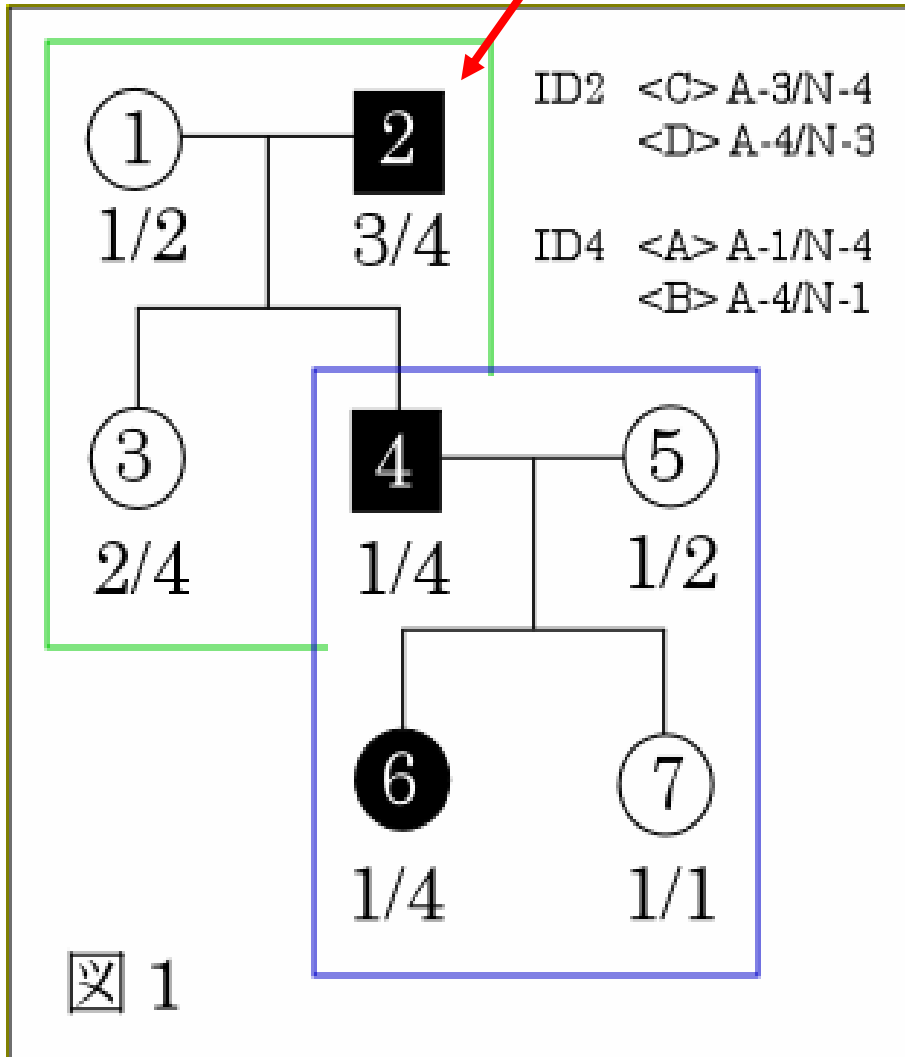
③母親から、マーカー対立遺伝子, アレル1 (N-1) が選ばれる確率: $1/2$ により, $(1/2)^2 \theta$ と算出される。

よってこの核家族の観察データの尤度は
 $(1/2)^2 \theta \cdot (1/2)^2 \theta = (1/2)^4 \theta^2$

パターン: A-4とN-1の場合

$(1/2)^2 \theta \cdot (1/2)^2 \theta = (1/2)^4 (1-\theta)^2$

パラメトリック連鎖解析の例



パターン $\langle C \rangle - \langle B \rangle$ もしくは $\langle D \rangle - \langle B \rangle$ しかありえない。

$\langle C \rangle$

正常3: $(1/2)^2 (1-\theta)$

患者4: $(1/2)^2 \theta$

$\langle C \rangle - \langle B \rangle$ の尤度は、

$$(1/2)^2 (1-\theta) \cdot (1/2)^2 \theta \cdot (1/2)^4 (1-\theta)^2 = (1/2)^8 \theta (1-\theta)^3 \quad \text{①}$$

同様に $\langle D \rangle - \langle B \rangle$ の尤度は

$$(1/2)^8 \theta (1-\theta)^3 \quad \text{②}$$

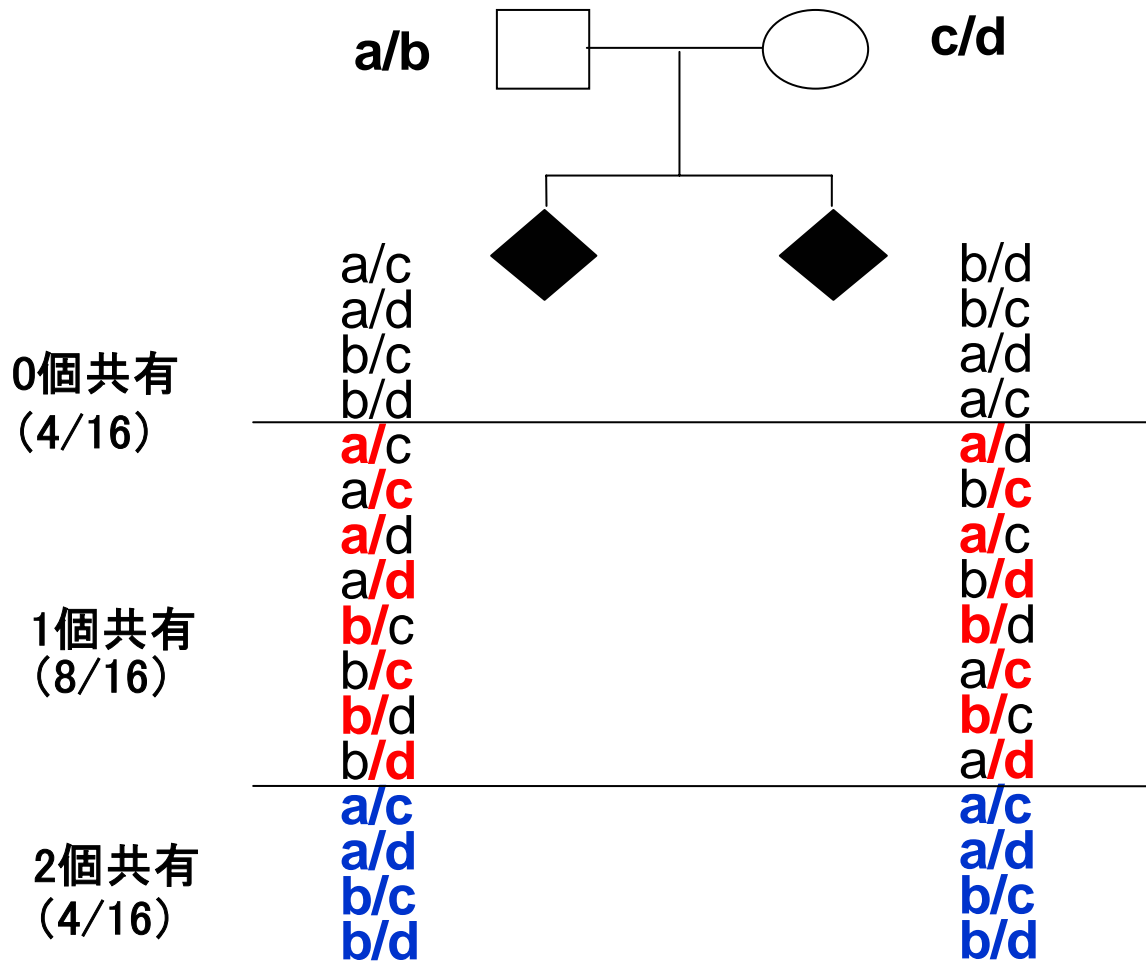
$\langle C \rangle$ となるか $\langle D \rangle$ となるかは等確率

よってこのような家系全体で観察されたデータの尤度は、

$$(1/2) \cdot \text{①} + (1/2) \cdot \text{②} = (1/2)^8 \theta (1-\theta)^3$$

最尤法では通常尤度比(自由組み換え ($\theta = 0.5$) の尤度に対する尤度)を用いて $0 \leq \theta < 0.5$ において尤度比が最大になる θ を求める。この場合は $\theta = 0.25$ の時に尤度比最大となり、1.6875 となる(LOD(Logarithm of odds)値は、0.22724)。

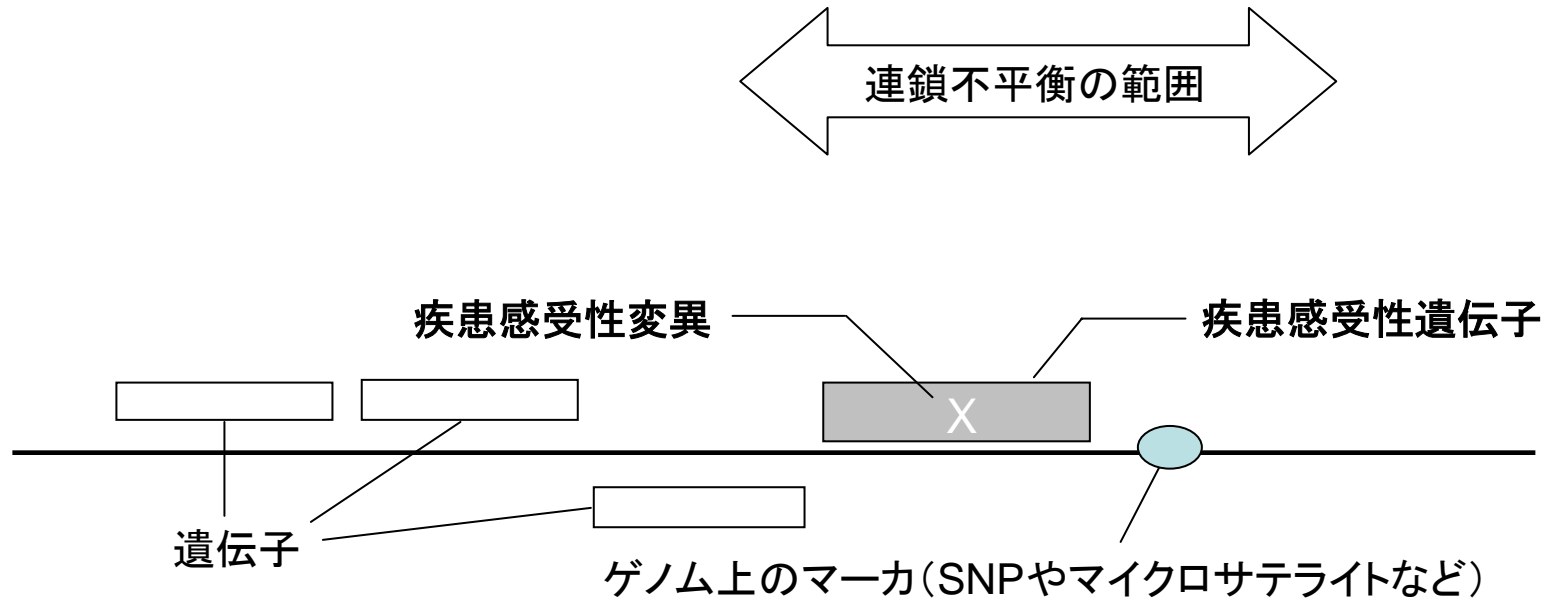
罹患同胞対法の原理



両親の遺伝マーカーの遺伝子型をそれぞれa/b、c/dとすると、その2人の子供の遺伝子型を比較した場合、アレルを0個、1個、2個共有する組み合わせは無作為分配の場合それぞれ、4/16 (25%)、8/16 (50%)、4/16 (25%)となる。

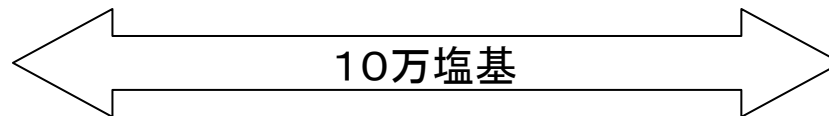
ある遺伝マーカーが疾患の責任遺伝子と連鎖していると、ともに疾患を発症している同胞は、無作為分配に比べてアレルの共有頻度が高くなる。

相関解析を支える考え方(連鎖不平衡)

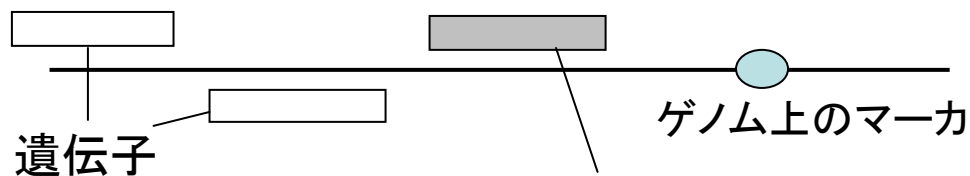


ゲノム上のマーカと疾患感受性原因変異との間に連鎖不平衡が存在する場合、疾患感受性変異を解析しなくとも、マーカ解析で近くに疾患感受性変異をもった遺伝子が存在することがわかる

現人類は5~10万年前
ごろアフリカをでる？



1,000(世代) × 20年 = 2万

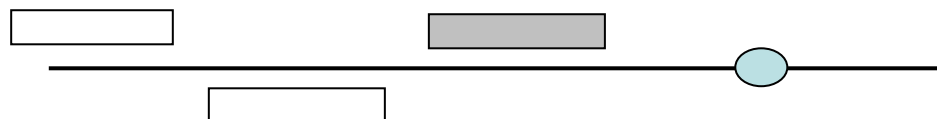


原日本人はマンモスを
追って2万年前ごろ
シベリアから日本へ？

疾患感受性遺伝子

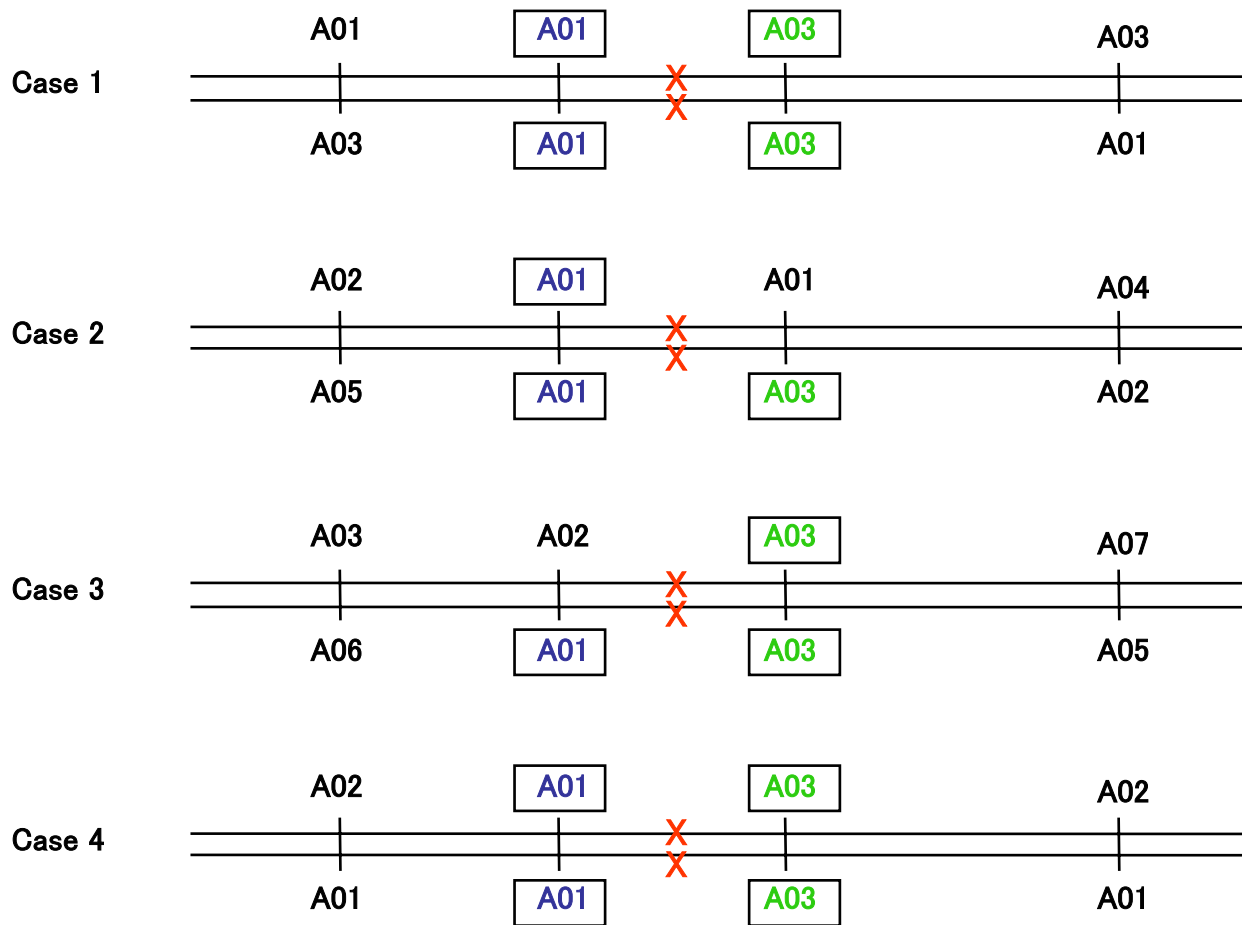


2万年では、10万塩基の
範囲にある遺伝子群は
組み換えがほとんど起こらず、
祖先と同じ組み合わせをもつ



2万年では10万塩基範囲にある遺伝子の組み換えは起こらない

マーカーと疾患遺伝子との相関

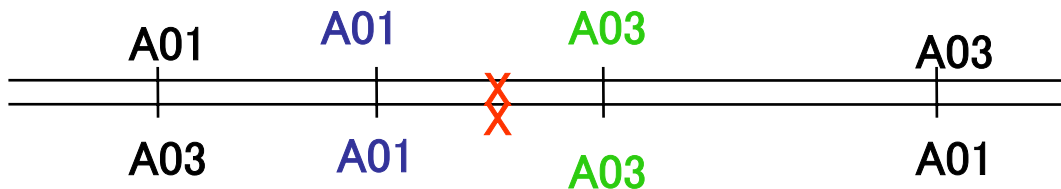


連鎖不平衡

患者群でA01、A03が増加する

マーカーと疾患遺伝子との相関

患者



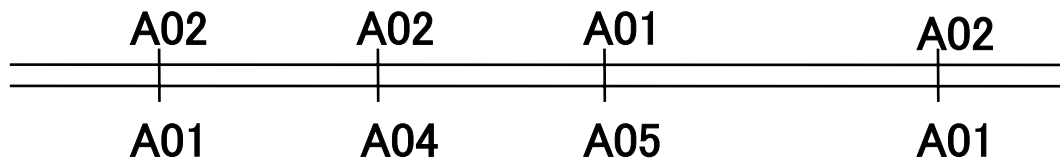
X²乗検定

A01 : A03 患者群 65 / 100

A01 : A03 健常人 2 / 100

P < 0.0001

健常人



各統計解析の特徴

解析の種類	適する対象集団	遺伝的相対危険	構造化問題	陽性領域
Parametric連鎖解析	少数大家系	高い	小さい	広い
Nonparametric連鎖解析	多数小家系	中-低い	小さい	広い
TDT , S-TDT	多数小家系	低い	中	狭い
連鎖不平衡解析	多数の患者と対照者	低い	中-大	狭い
相関分析	多数の患者と対照者	低い	大	狭い
家系内相関分析	多数の患者と家族	低い	中	狭い