Human Genetics Certificate
BeSHG course
10 Nov 2023, ULB

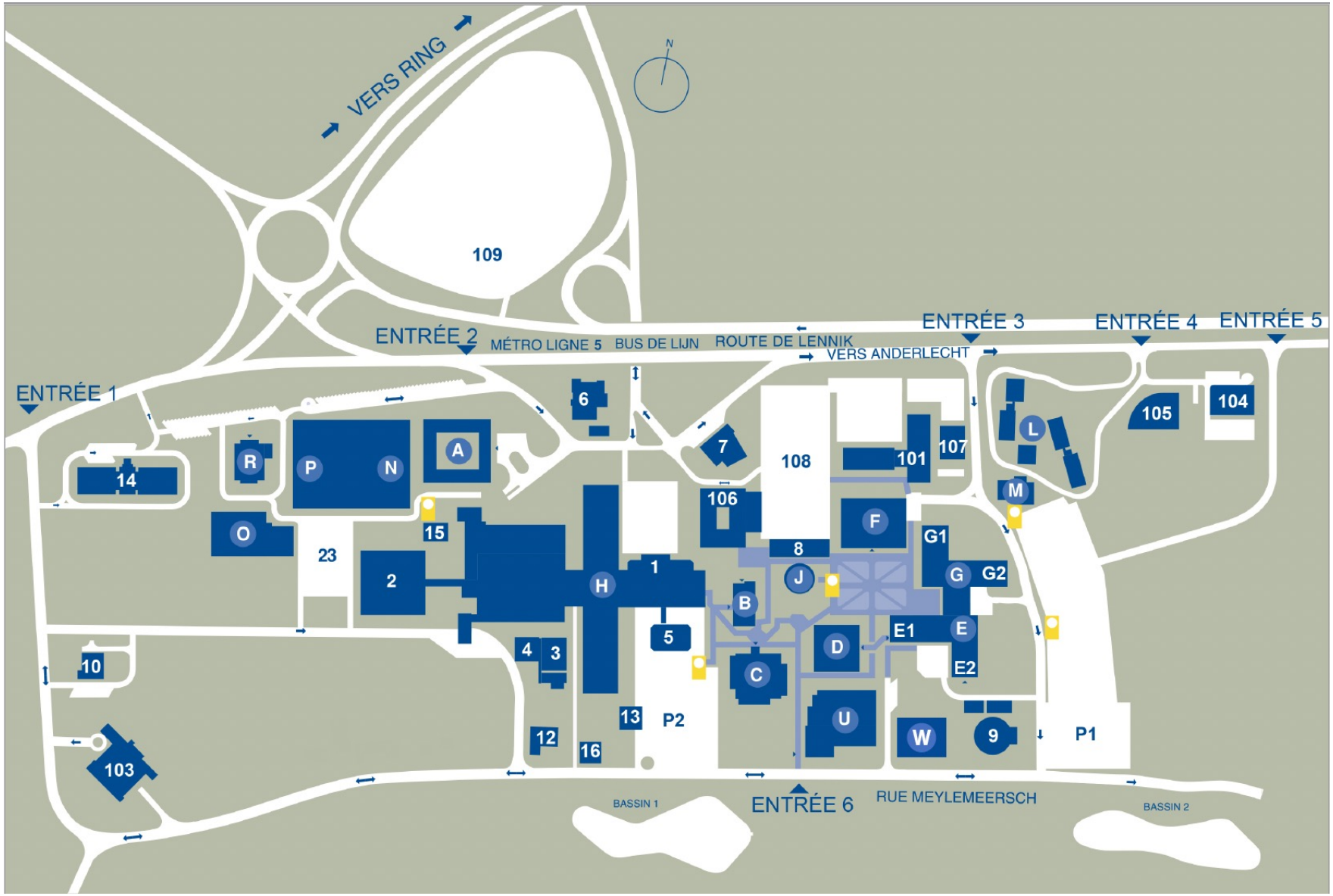
Erasme ULB campus

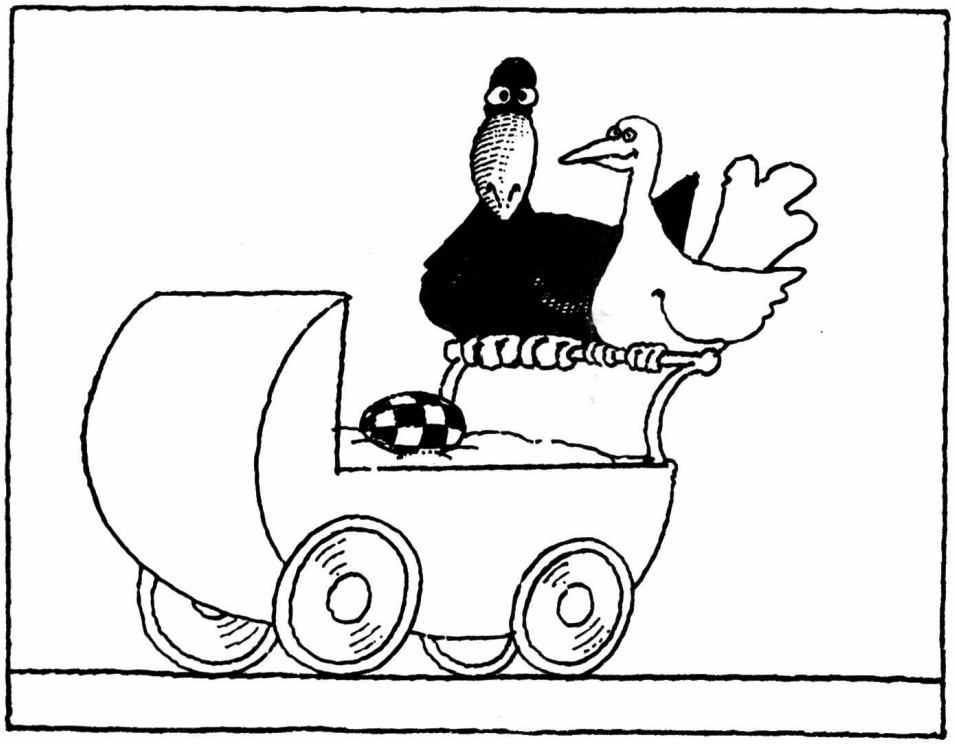
808 Lennik St, B-1070

- 09:00-10:30 Patterns of Single Gene Inheritance (Ch.7, T&T)
- 10:30-11:00 *Break*
- 11:00-12:00 *Lunch*
- 12:00-13:30 Continued + mutations and polymorphisms (Ch.4)
- 13:30-14:00 *coffee break*
- 14:00-15:00 Genetic variation in populations (Ch.9)
- 15:00-15:30 *coffee break*
- 15:30-17:30 Epigenetics

Bldg C, Room C4.121
(level 4, salle Gillet)

Bldg N, Room N3.114


VERS RING
N
109
ENTRÉE 2
MÉTRO LIGNE 5
BUS DE LIJN
ROUTE DE LENNIK
VERS ANDERLECHT
ENTRÉE 3
ENTRÉE 4
ENTRÉE 5
ENTRÉE 1
6
7
108
101
107
104
105
L
M
14
R
P
N
A
106
F
G1
G
G2
O
15
23
8
J
2
H
1
B
E1
E
5
4
3
D
E2
10
C
U
W
9
P1
13
P2
12
16
103
BASSIN 1
ENTRÉE 6
RUE MEYLEMEERSCH
BASSIN 2



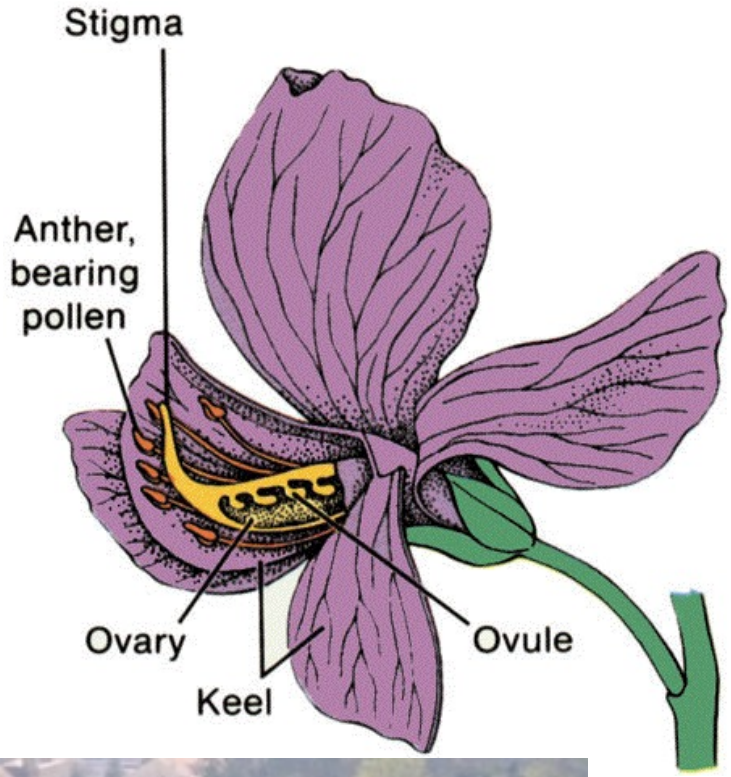
Part 1

PATTERNS OF SINGLE-GENE INHERITANCE

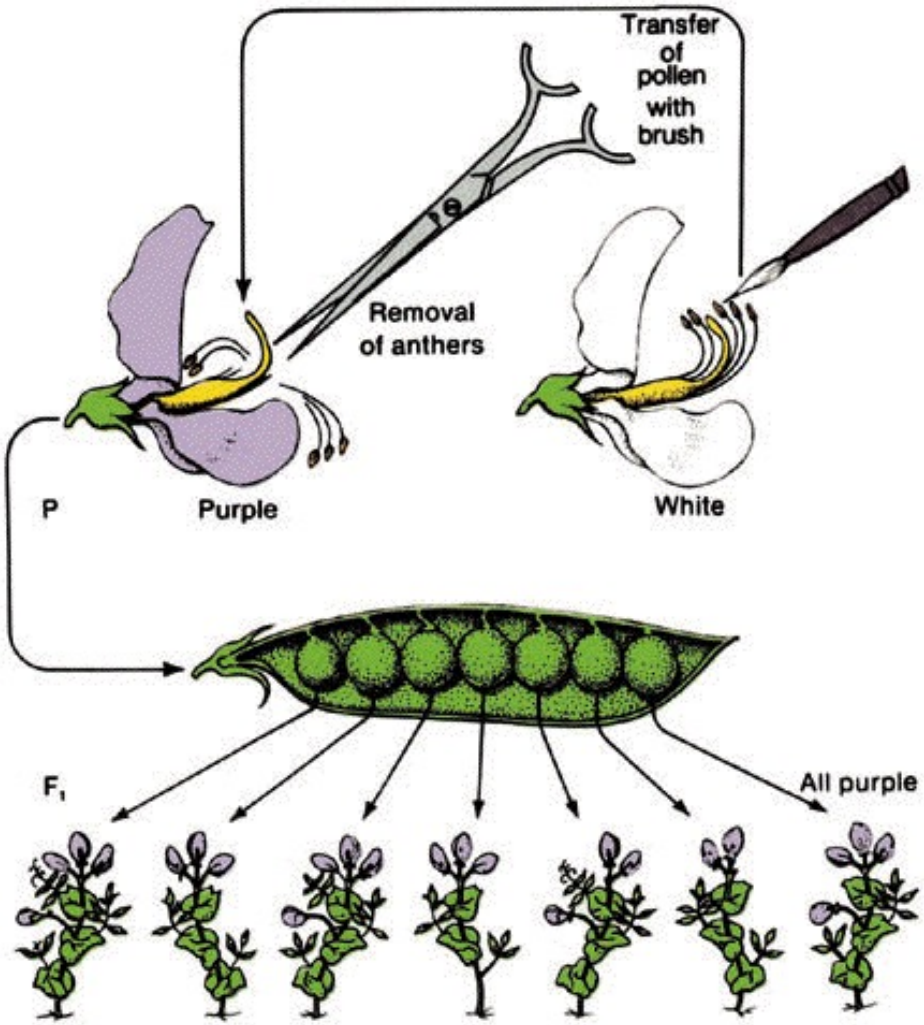
Genetics implies variation

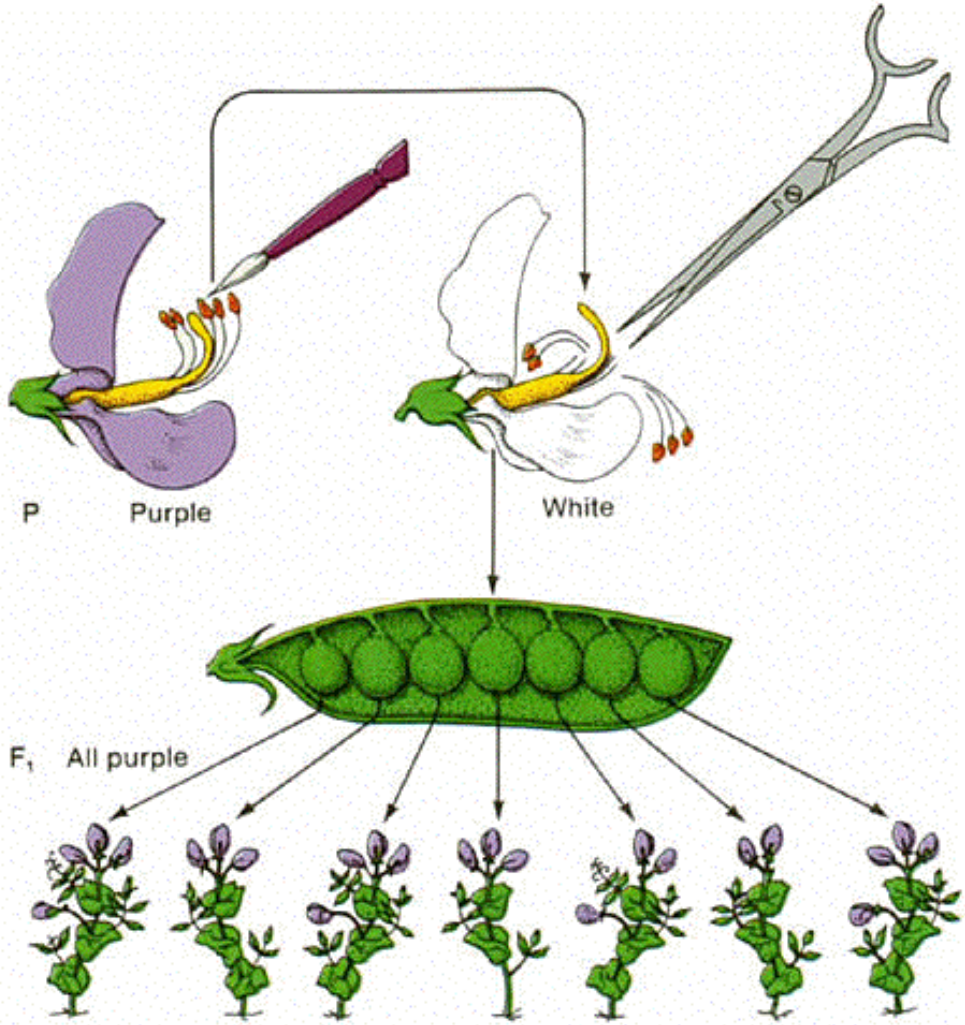
- Genetics = study of inheritance of characters (= traits = features)
- No genetics if no variation

Gregor Mendel (1822-1884)

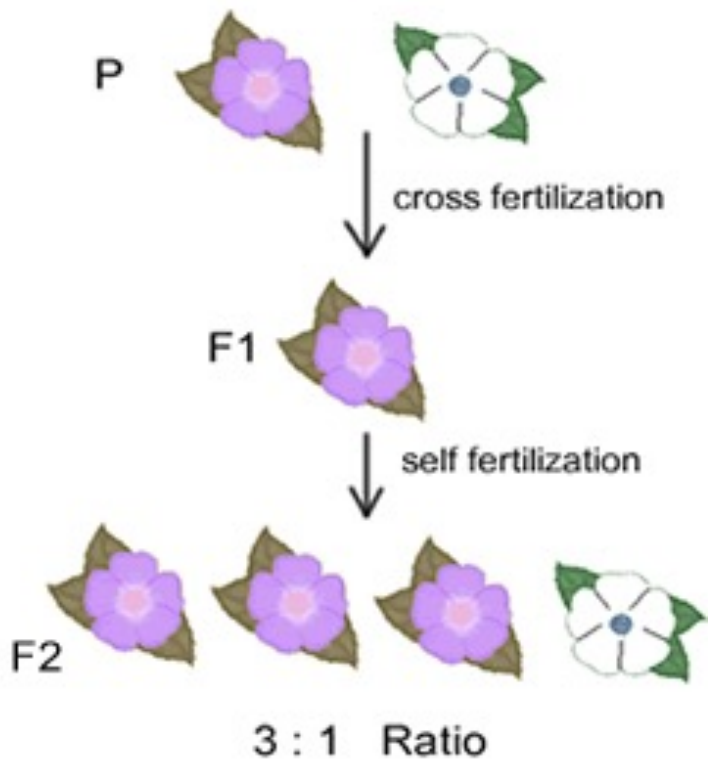


Artificial cross-pollinization



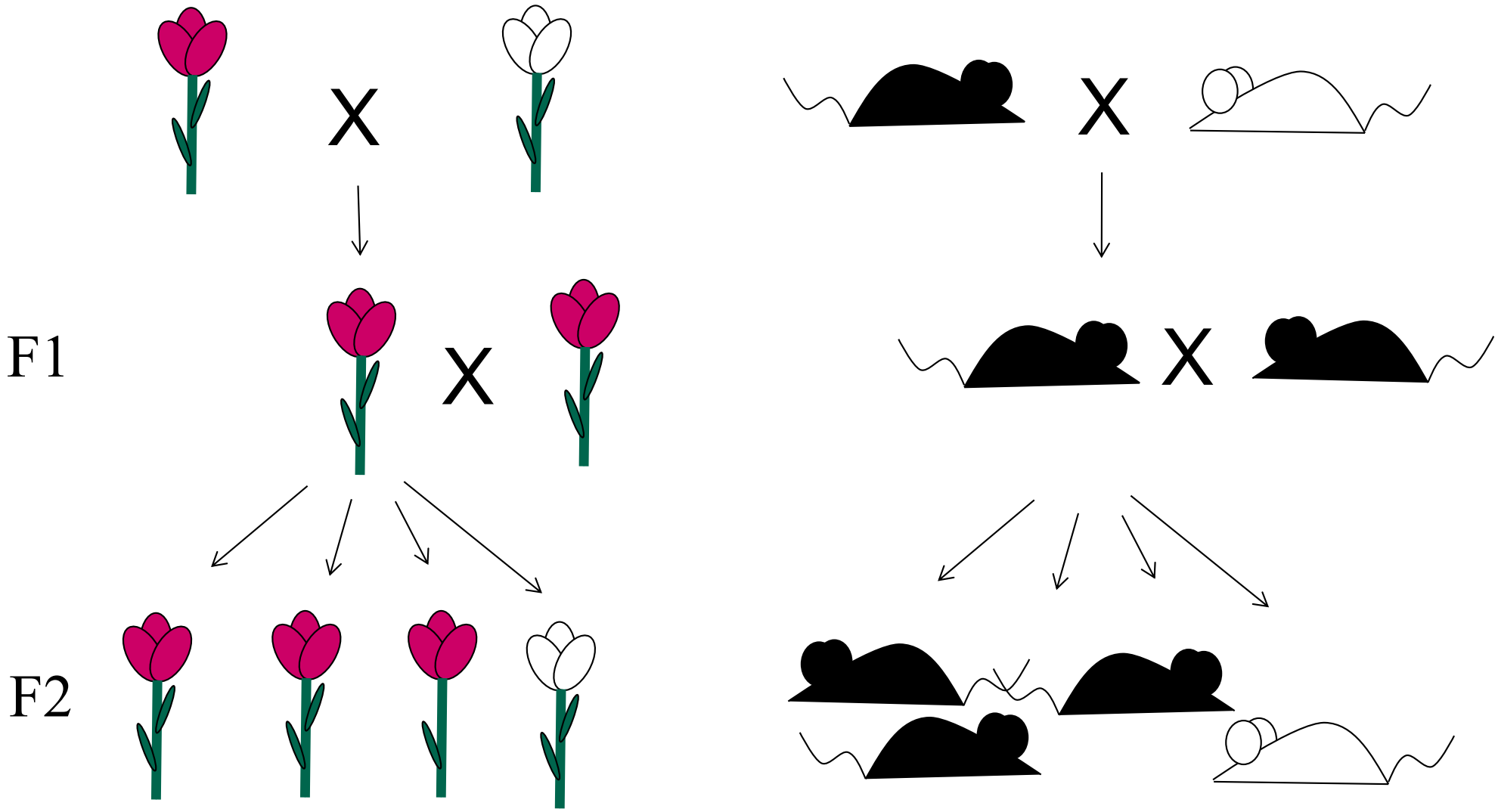


Cross 2 pure strains that differ for 1 character (monohybrid cross)

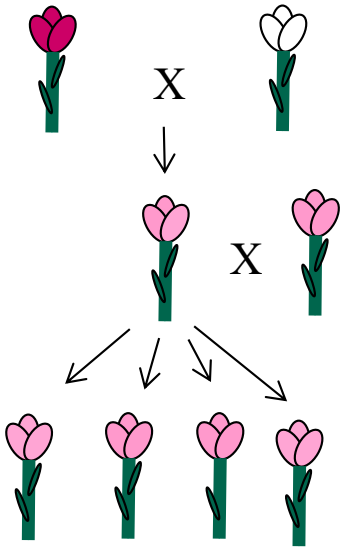


- All purple: not simple dilution
- White character re-appears

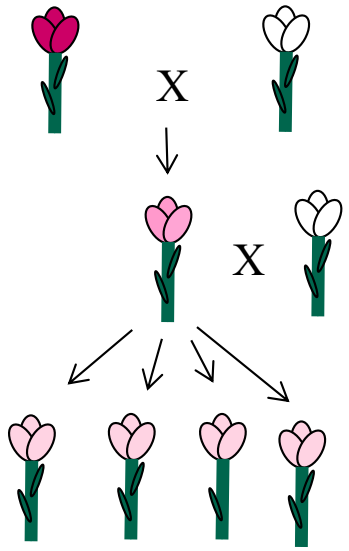
Fixed proportions of phenotypes in offspring



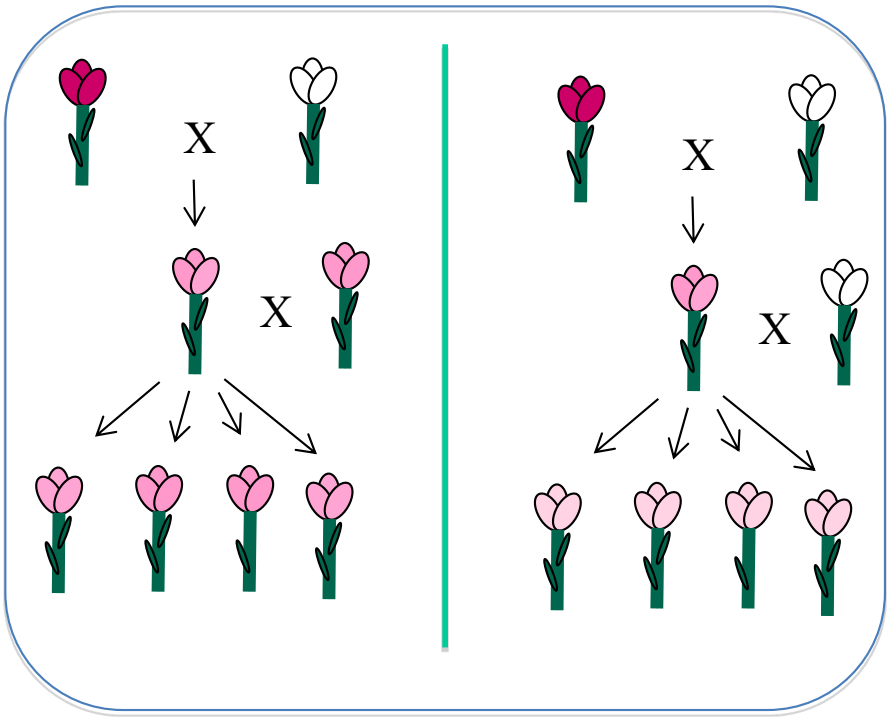
Prior Hypothesis: trait dilution



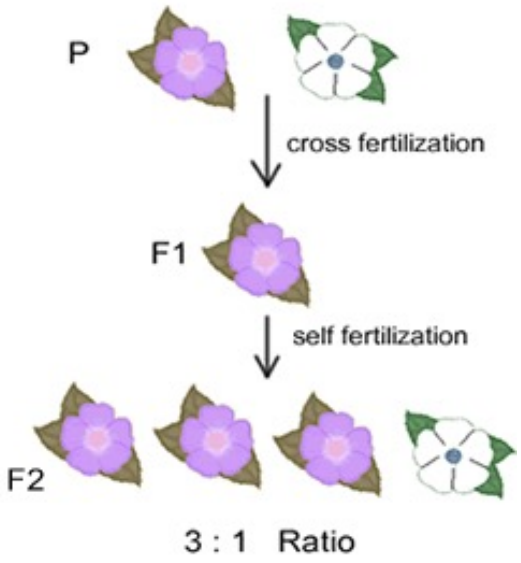
Prior Hypothesis:
trait dilution



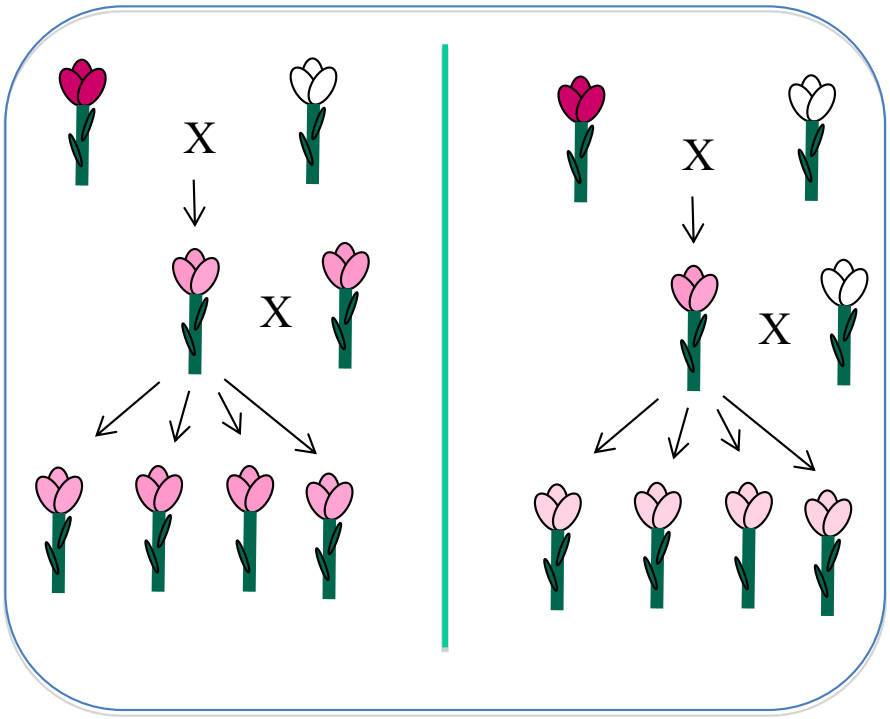
Prior Hypothesis: trait dilution



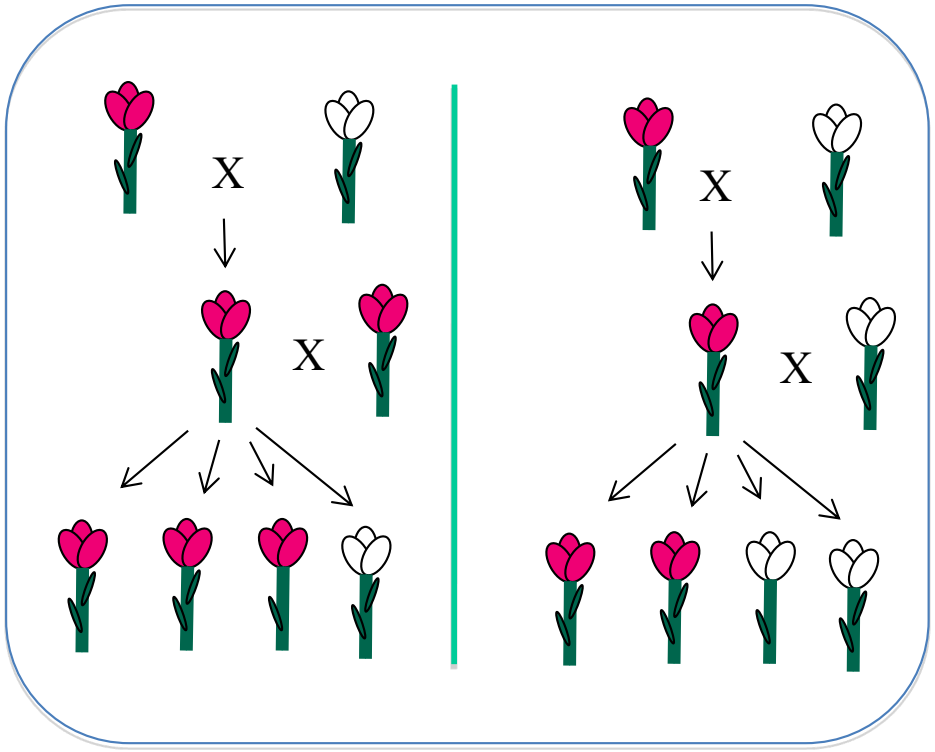
Cross 2 pure strains that differ for 1 character (monohybrid cross)



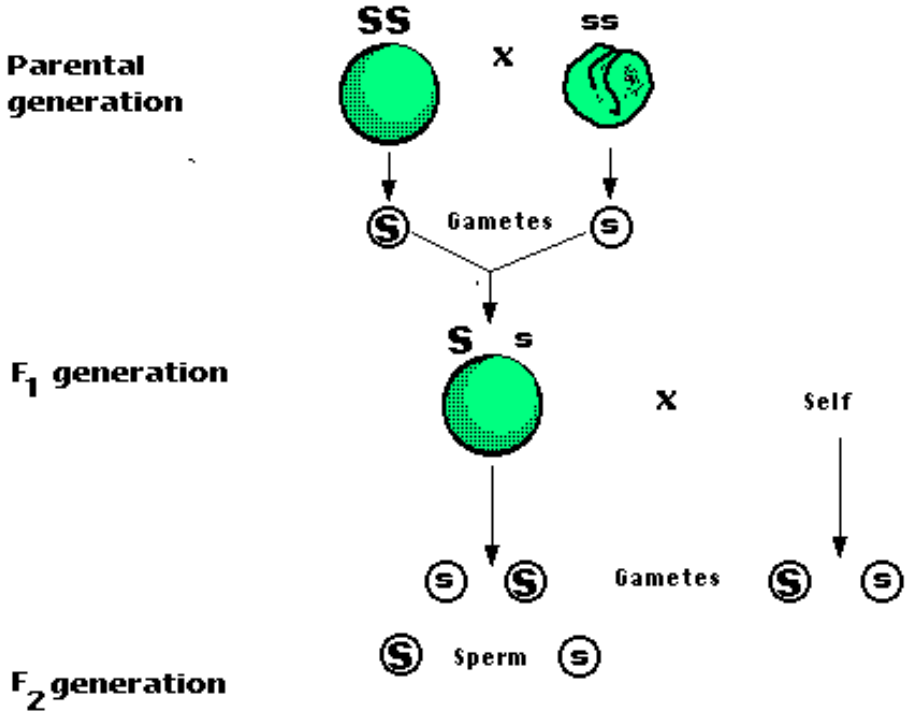
Prior hypothesis:
trait dilution

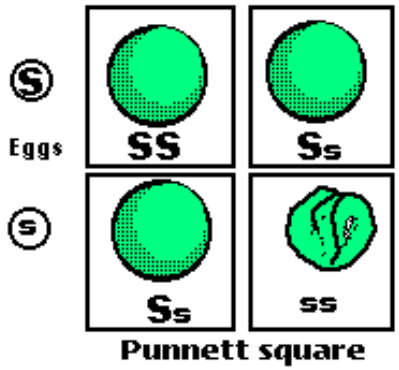


Experimental
evidence



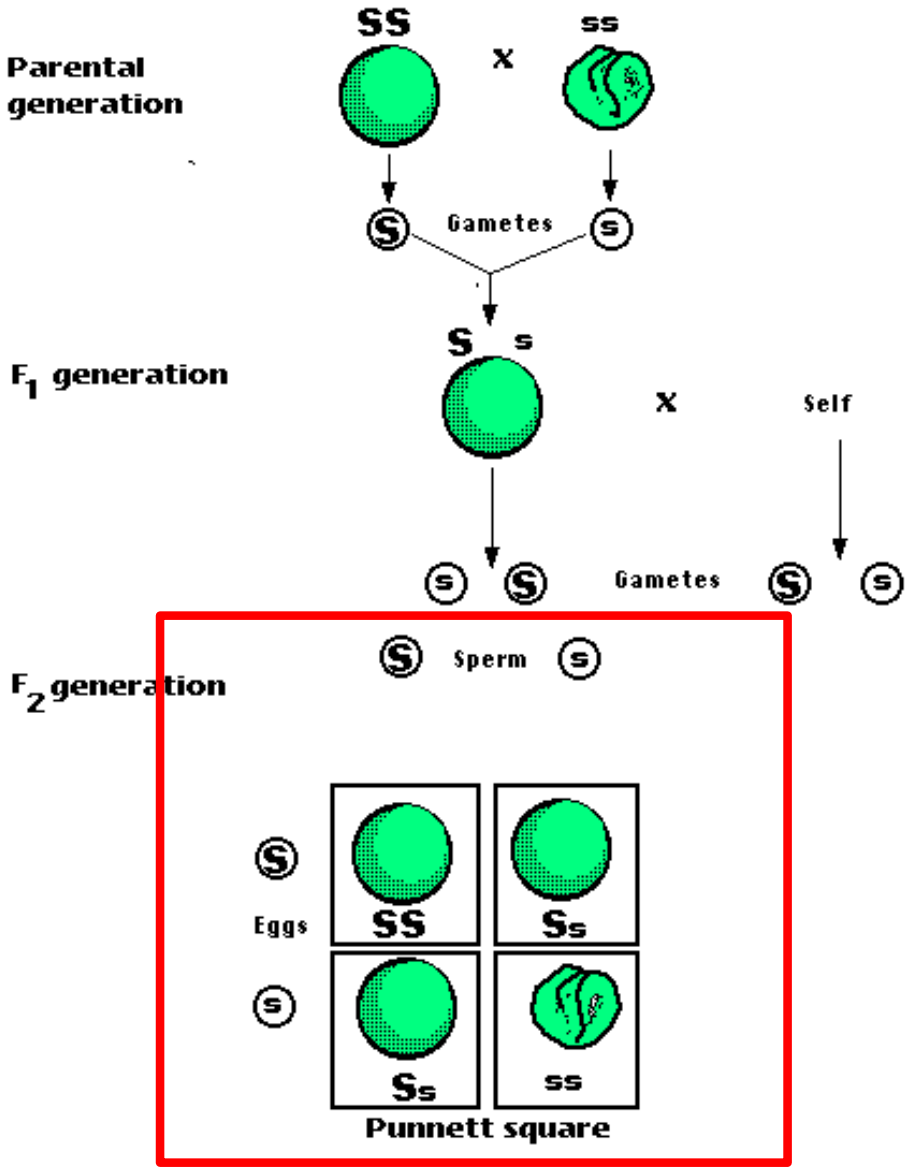
Cross 2 pure strains that differ for 1 character (monohybrid cross)





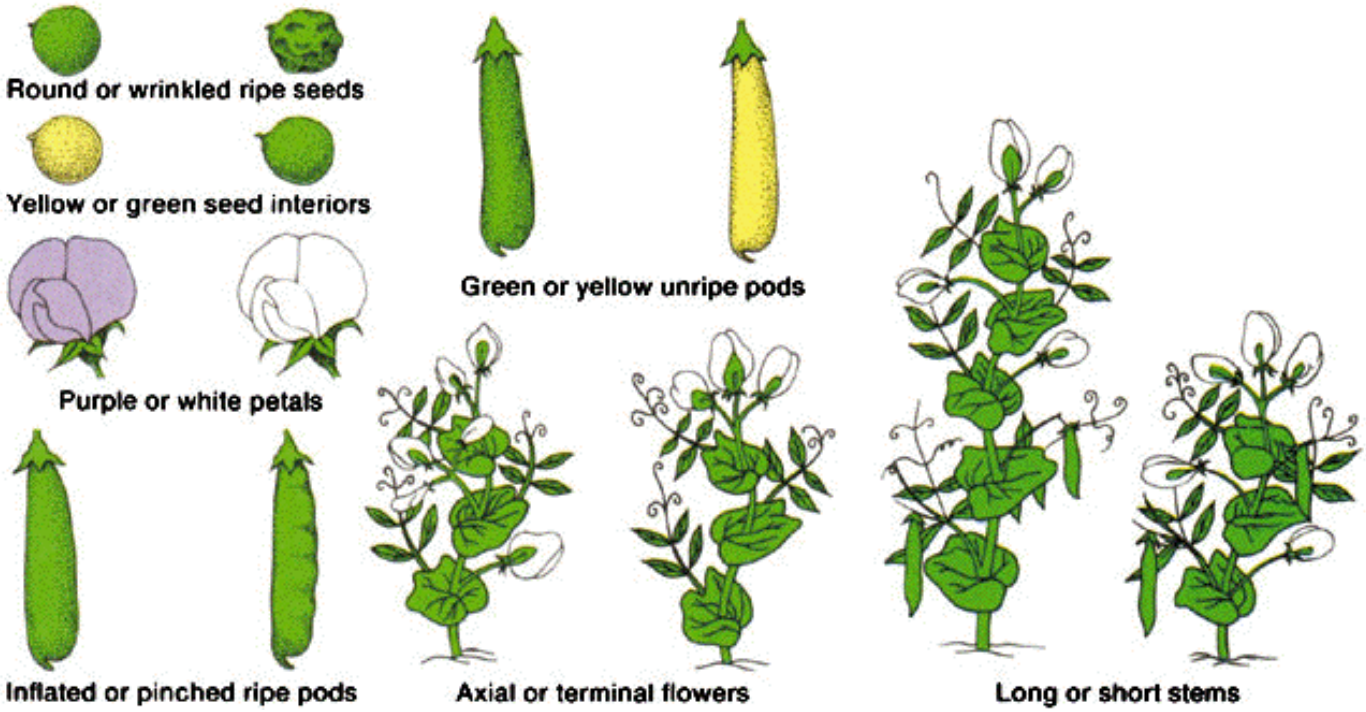
- Smooth or wrinkled
- All F1 individuals are smooth
  (= filia 1)
  S character is dominant, wrinkled is récessif
- But wrinkled character reappears in F2 !
  25% of F2 individuals are wrinkled
- Best explained by INDEPENDANT SEGREGATION of 2 *allelomorphic* variants of one hereditary factor:
  S or s

Cross 2 pure strains that differ for 1 character (monohybrid cross)

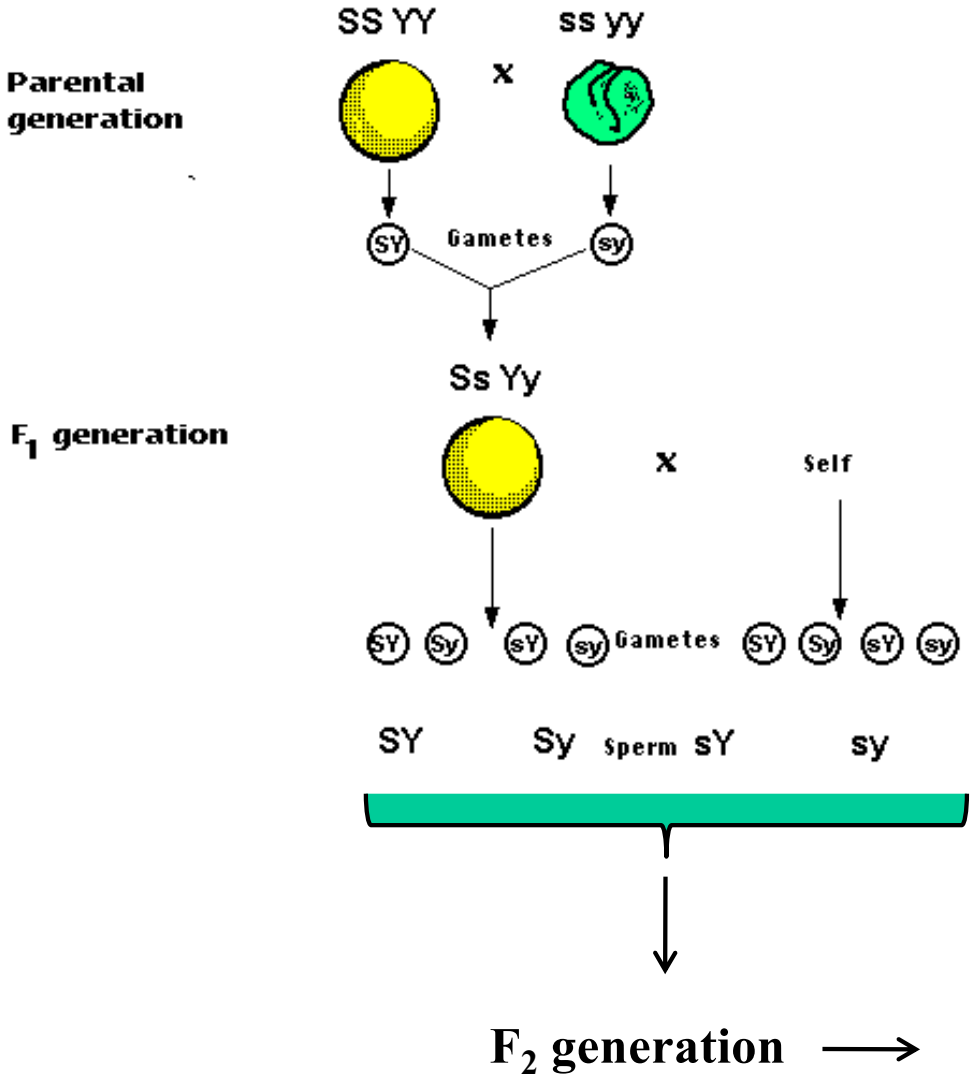


- Smooth or wrinkled
- All F1 individuals are smooth
  (= filia 1)
  S character is dominant, wrinkled is récessif
- But wrinkled character reappeasr in F2 !
  25% of F2 individuals are wrinkled
- Best explianed by INDEPENDANT SEGREGATION of 2 allelomorphic variants of one hereditary factor:
  S or s

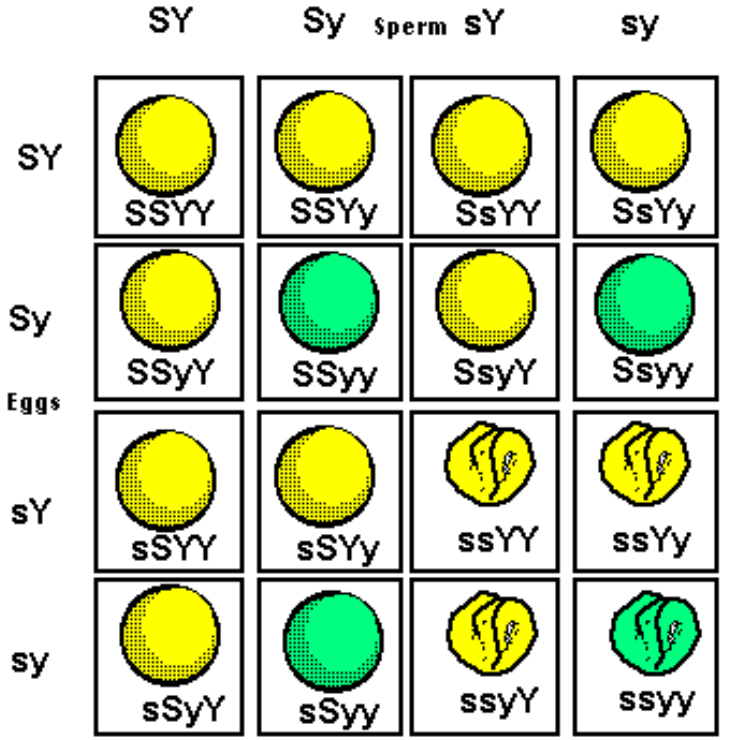
The 7 character differences studied by Mendel



Dihybrid cross (2 characters)

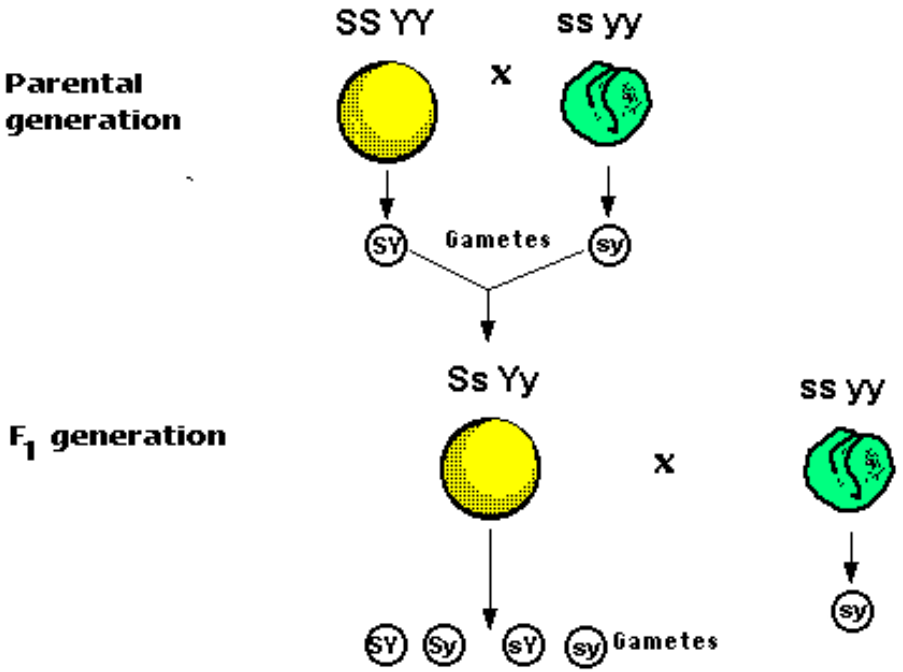


- Independent assortment of hereditary character
- Ratio 9:3:3:1 of phenotypes



Punnett square

Backcross



	SY	Sy	sY	sy
sy	sSyY	sSyy	ssyY	ssyy
	1/4 Smooth Yellow	1/4 Smooth green	1/4 rough Yellow	1/4 rough green

- Cross F1 hybrid (SsYy) with a double recessive homozygote (ssyy)
- Unmasks the genotype of the F1 hybrid
- With independent assortment of hereditary factors, expect phenotypes in the following proportions: .25/.25/.25/.25

Mendel's observations

- Uniformity of hybrids in first generation (F1)
- Independent segregation of several couples of characters in second generation (F2)
  - « purity of gametes: each contain only one hereditary factor for one character » = one allele of each gene
- Independent disjunction of characters in F2

Mendel's laws

1. Law of **segregation**
   – Each gamete contains one or the other of two allelomorphic factors (alleles)
   later found to fit meiotic separation of pat and mat chromosomes

2. Law of **independent assortment**
   – Pairs of alleles from different genes enter gametes independetly of one another

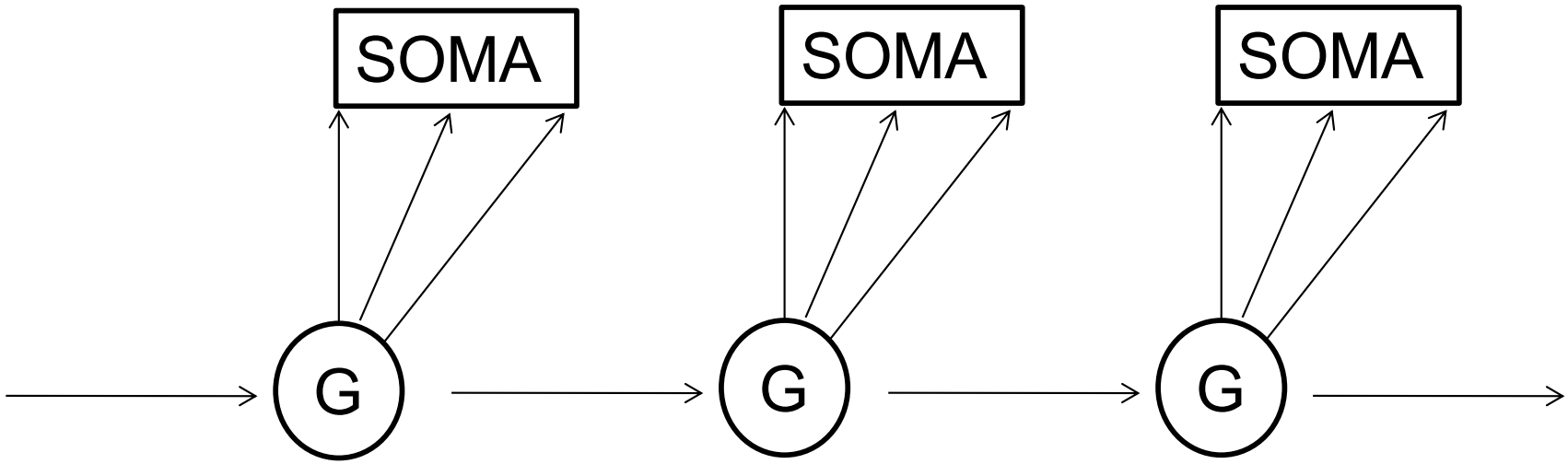
   except if genes closely located on same chromosome (linkage)

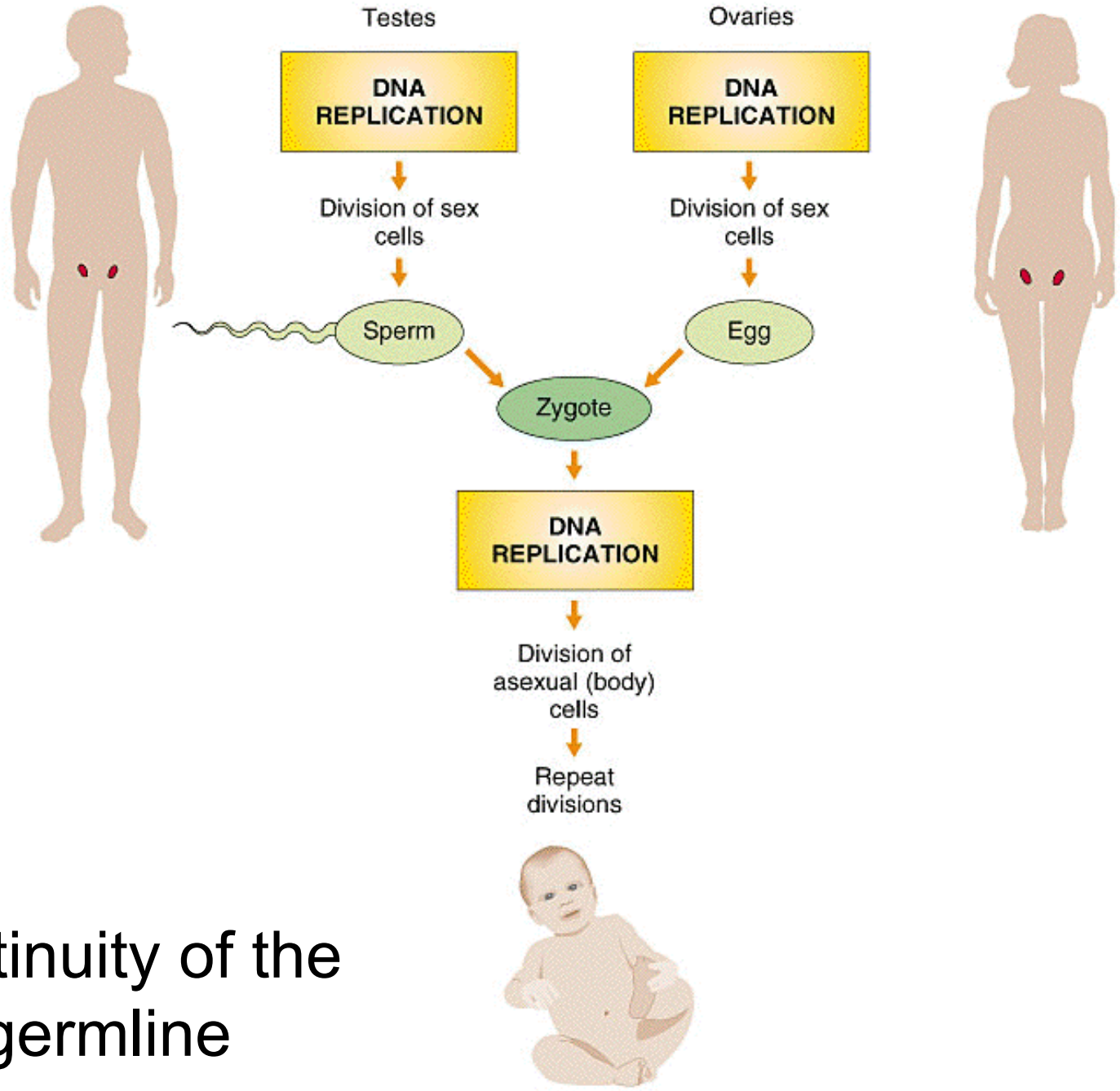
Independant segregation of the two alleles of each gene

Loci, genes, alleles, mutation

- Locus = position in genome
  - gene, or contiguous genes (HLA locus), or SNP, any piece of DNA
- Alleles = alternative variants at one locus
  - Prevailing allele = wild type
- Variant = any change in an allele
- Mutation = change in an allele causing a change in phenotype
- Genotype = individual set of alleles at one locus, or several loci, or whole genome *(music score)*
- Phenotype = observable expression of a genotype *(concert)*
  - Morphological
  - Clinical
  - Cellular
  - Biochemical
  - …
- Pleiotropy = diversity of phenotypic effects

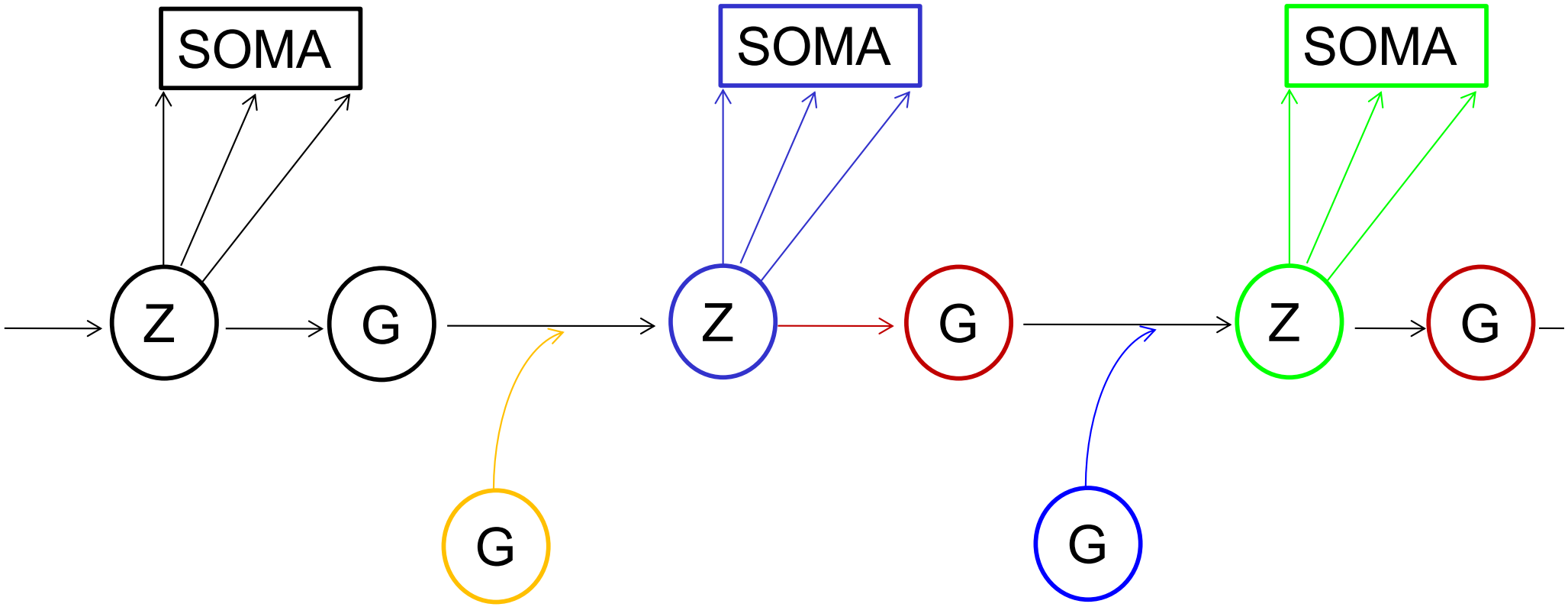
The germ-line



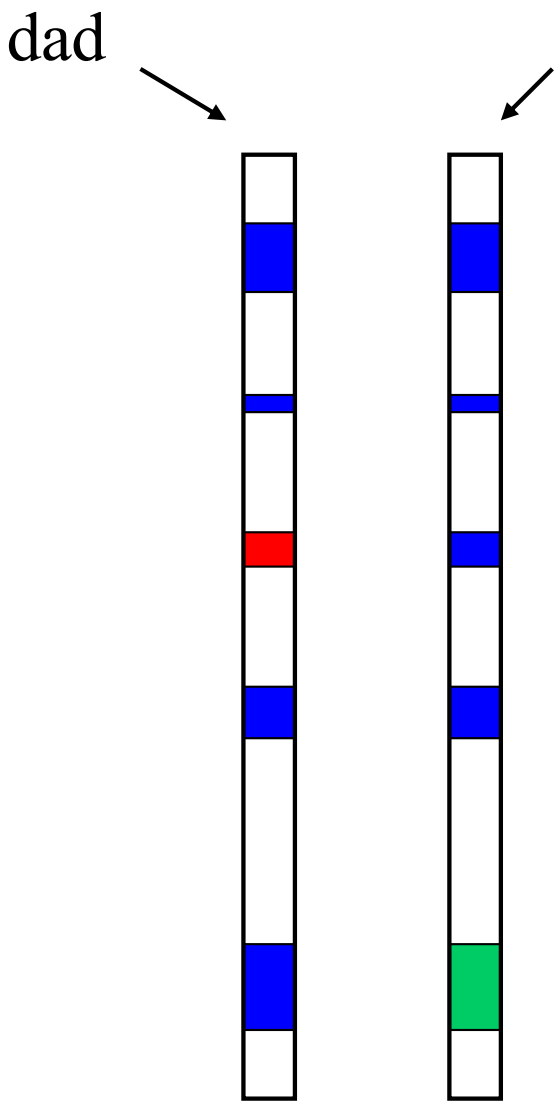


Continuity of the germline

Sexual reproduction

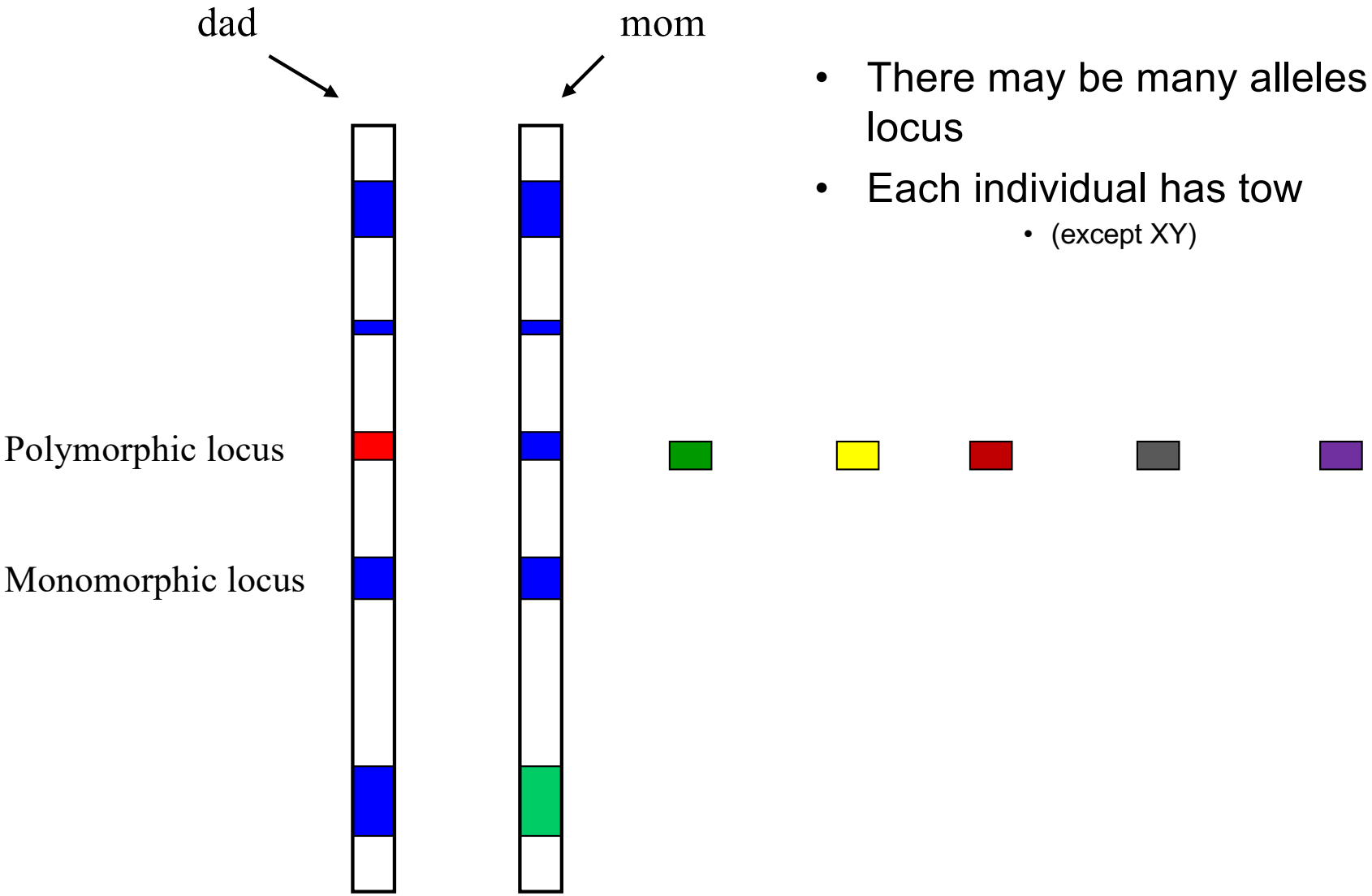


The genome: 20.000 genes, 2 copies each



- 2 alleles of each gene
- Many variants > Allele polymorphism
  - SNPs: Single Nucleotide Polymorphisms
  - CNPs: Copy Number Polymorphisms
  - other
- Unique combination in each person
- Largely encodes our physical characters (height, weight, coulor, blood groups…) and partly our psychologycal/ behavourial charachters
- No clear limit between polymorphism (normal) and mutation (pathogenic)
- Good our bad in different contexts

Locus: 2 alleles per individual ; Pool of alleles in population



- There may be many alleles per locus
- Each individual has tow
  - (except XY)

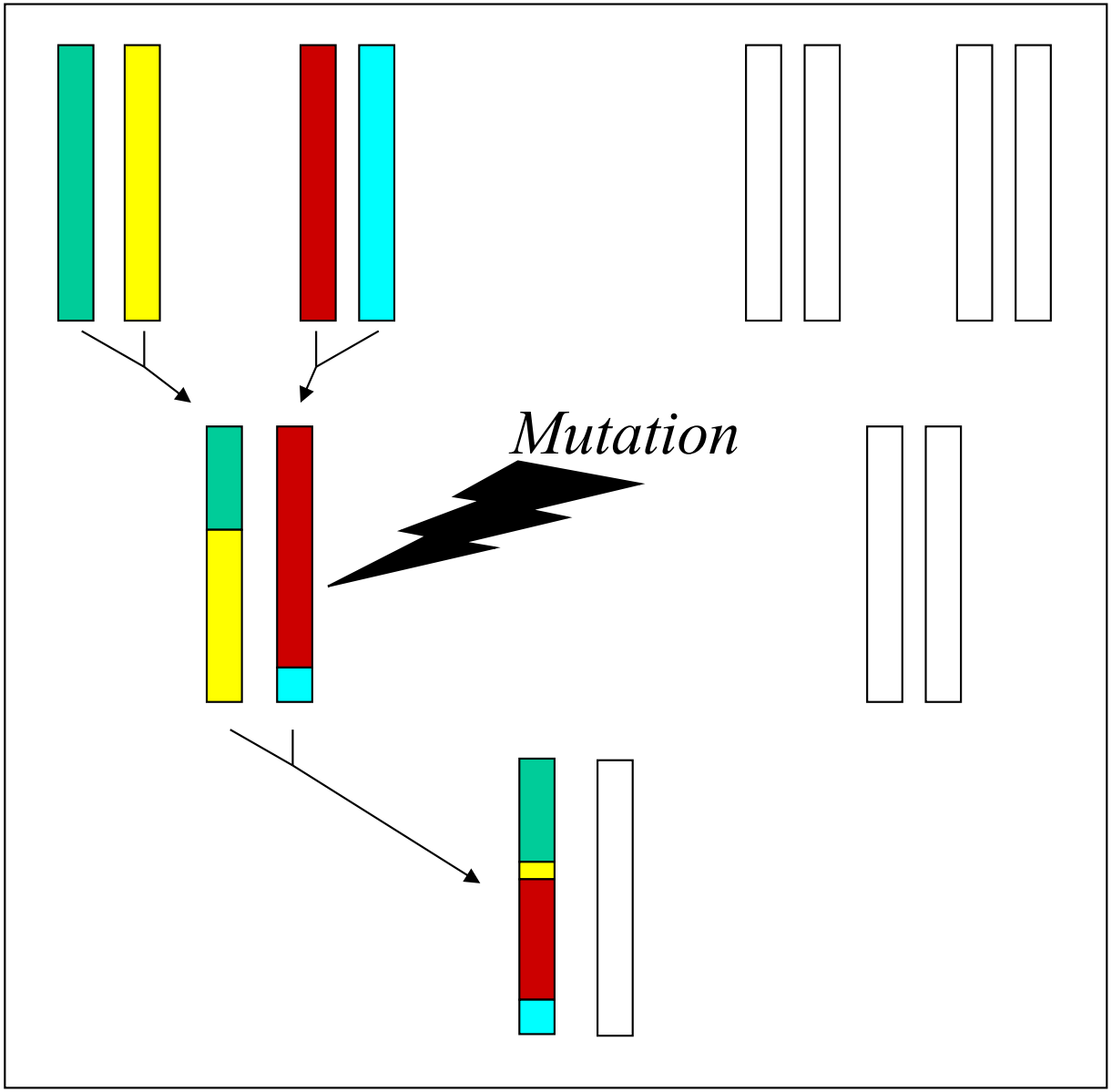
Synteny, linkage, LD

- Synteny = location on the same chromosome
  = pieces of one colinear DNA molecule

- Linkage = synteny close enough for transmission together in >50% gametes

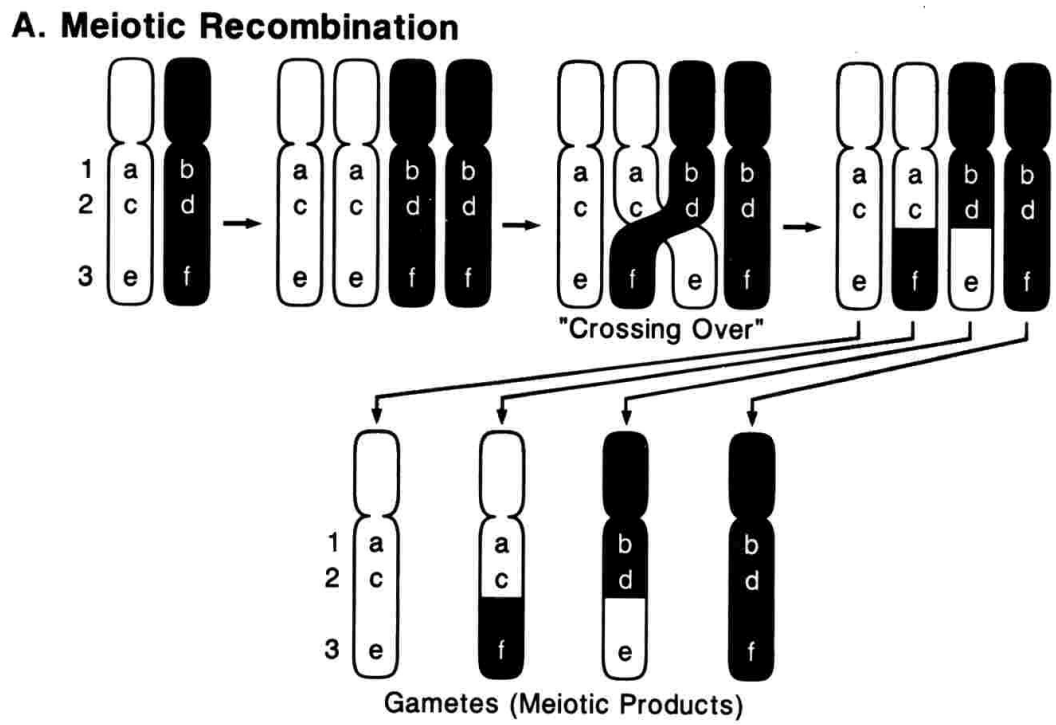
- LDisequilibrium = association of particular alleles at linked loci
  < close linkage / recent ancestor

Meiosis produces diversity by assembly



Ultimate origin of diversity = mutation

Meiotic Recombinations (crossing-overs)



- 1 cM = distance between loci that are separated in 1% gametes

  =>genetic distance, genetic map
  - (genetic linkage map)

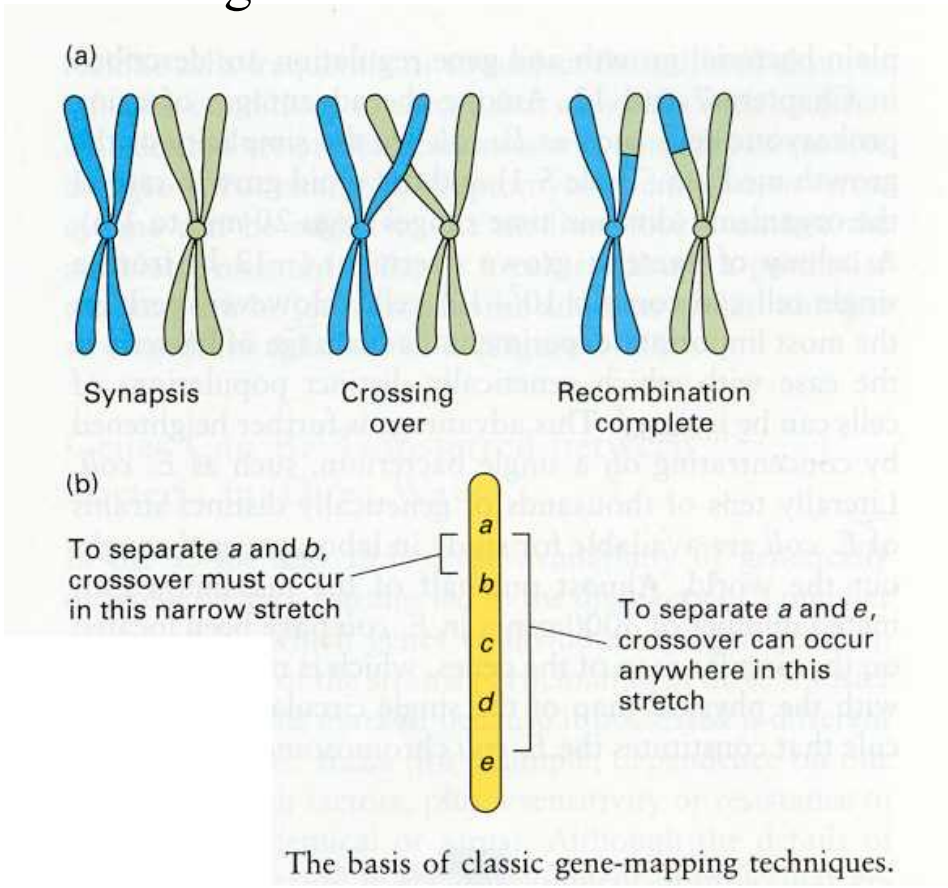
- 1 cM ≈ 1Mb. 3000 cM, 3 Gb.

- Recombinations mix the alleles => equilibrium

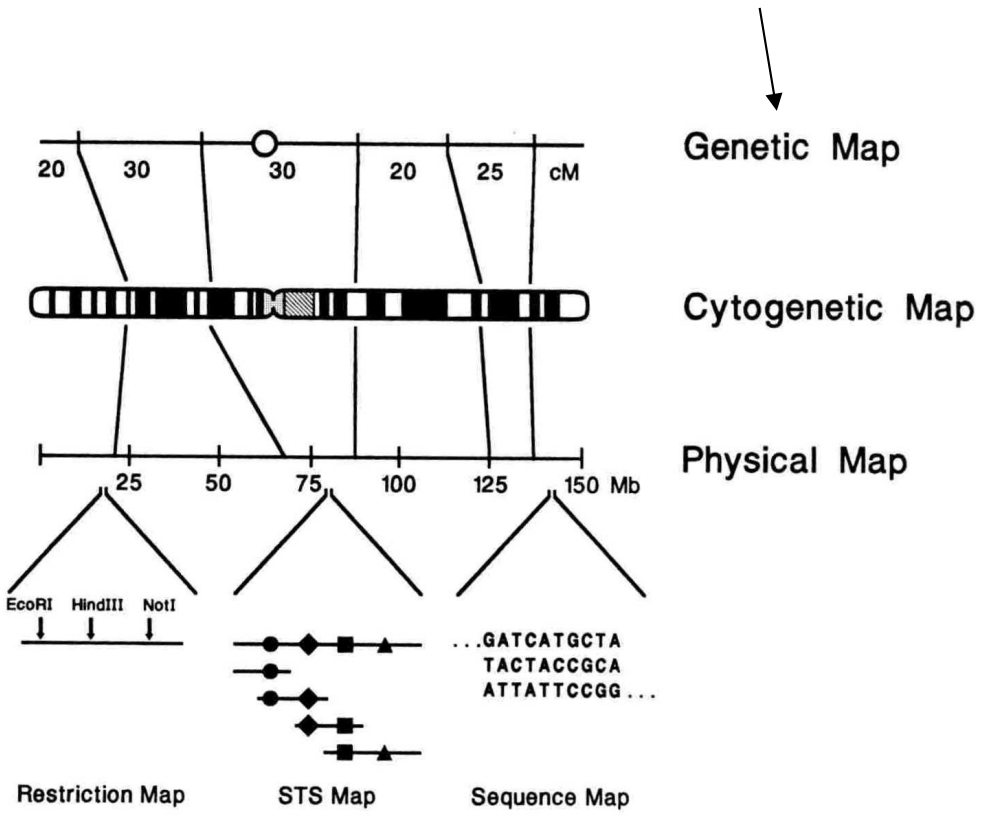
- At least 1 Cr-ov per chromosomal arm

Genetic linkage map (cM), physical map (Mb)

1% recombinant gametes <=> 1cM → 1 centimorgan (1 cM)



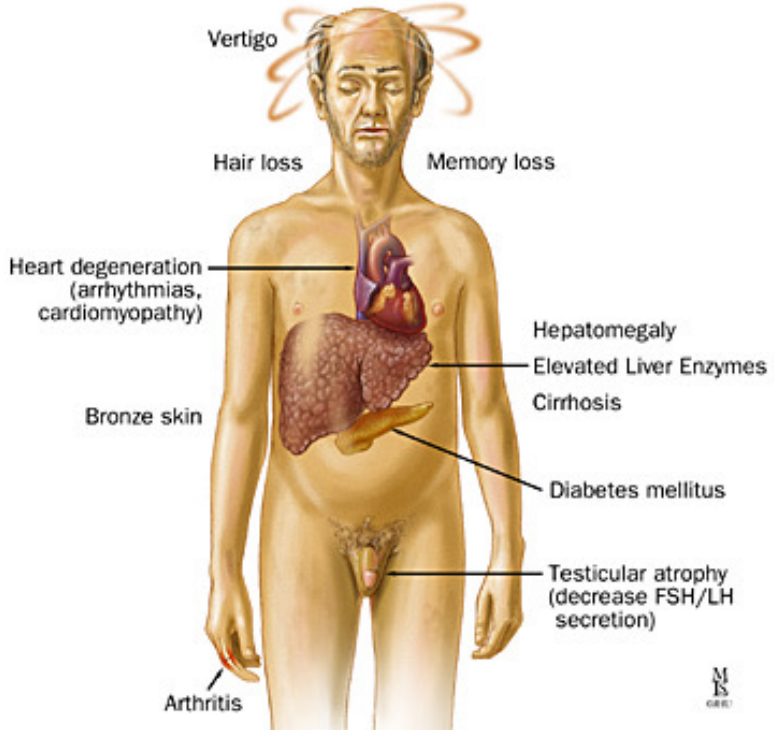
The basis of classic gene-mapping techniques.



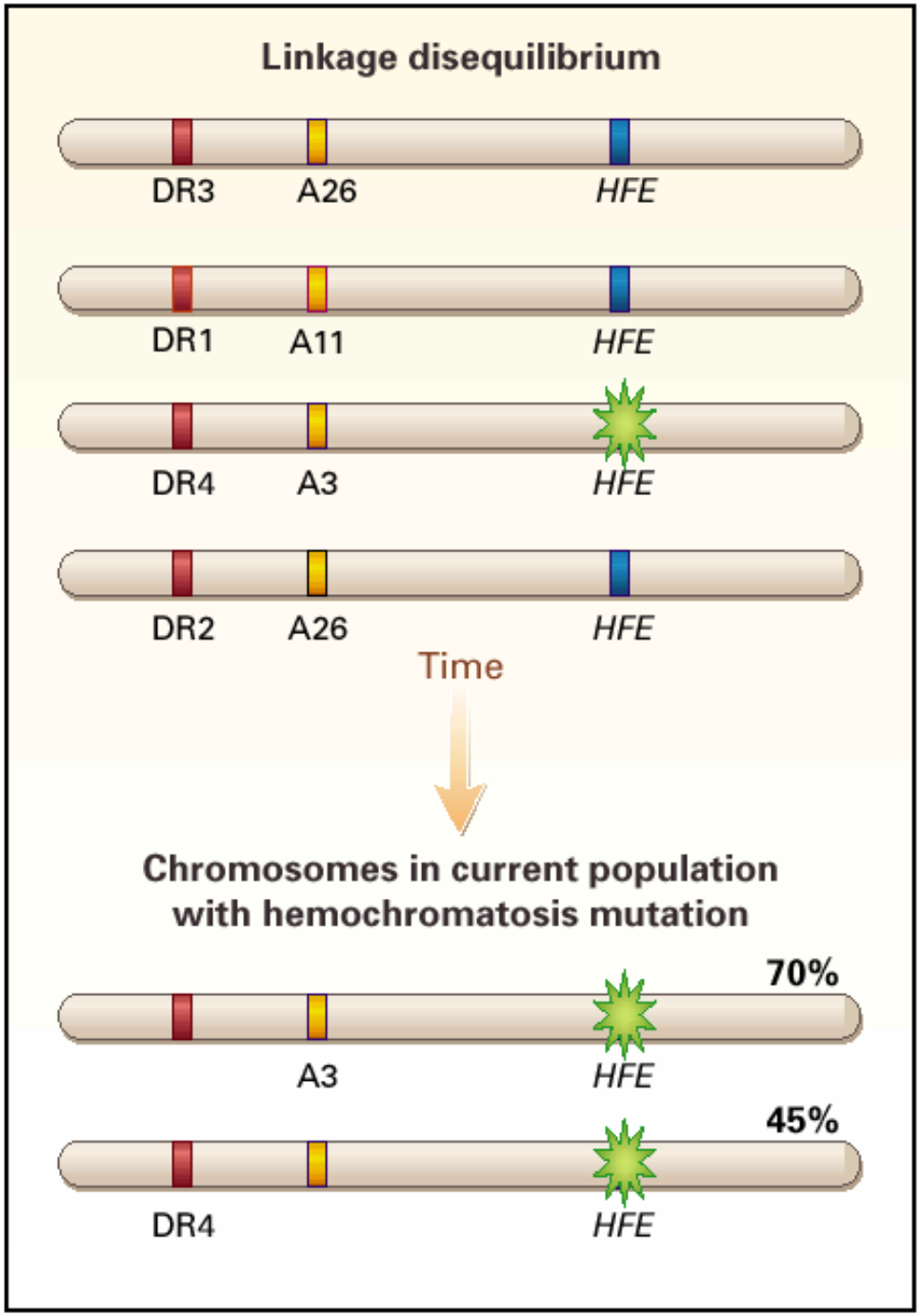
1 cM ≡~1 Mb (1.000.000 bp)

Hemochromatosis: excessive avidity for iron

- Genetic basis, essentially AR, HFE gene
- HFE gene linked to HLA-A gene



- Mutated allele HFE* C282Y which causes hemochromatosis is associated with allele HLA-A3



Linkage disequilibrium LD

- HLA-A3 :
  - General population : 15%
  - Hemochromatosis : 70%

- Mutation appeared 1! x not too long ago

Most common HFE mutation appeared 70 generations ago in a celtic, HLA-A3 subject

Linkage disequilibrium(LD)

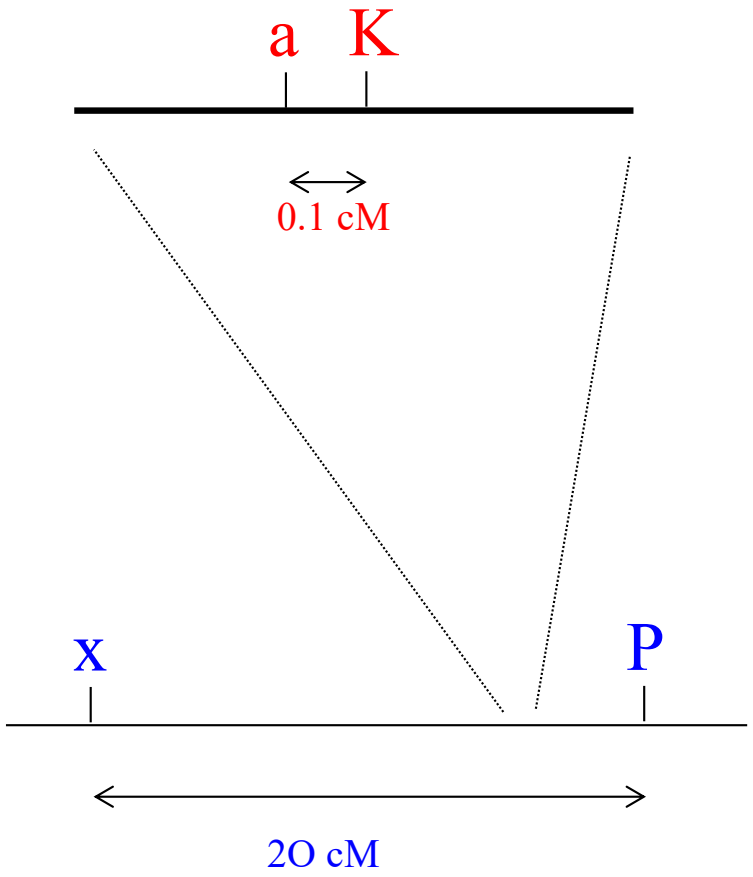
- Locus K,L,M,N
  Locus a,b,c,d

  a remains with K

  f(a,K) >> f(a) . f(K)

- Locus P,Q,R
  Locus x, y

  f(x,P) = f(x) . f(P)



Genetic characters / disorders (traits)

- Single-gene (monofactorial) — Mendelian: fixed proportions in offspring

  + mtDNA: maternal-inherited

- Chromosomal

- Complex

Phenotype inheritance from single gene

- AD — Htz mutation
- AR — Bi-allelic mutation
- X-linked — Hemizygous mutation
- maternal — mtDNA mutation

NECESSARY and SUFFICIENT to cause disease

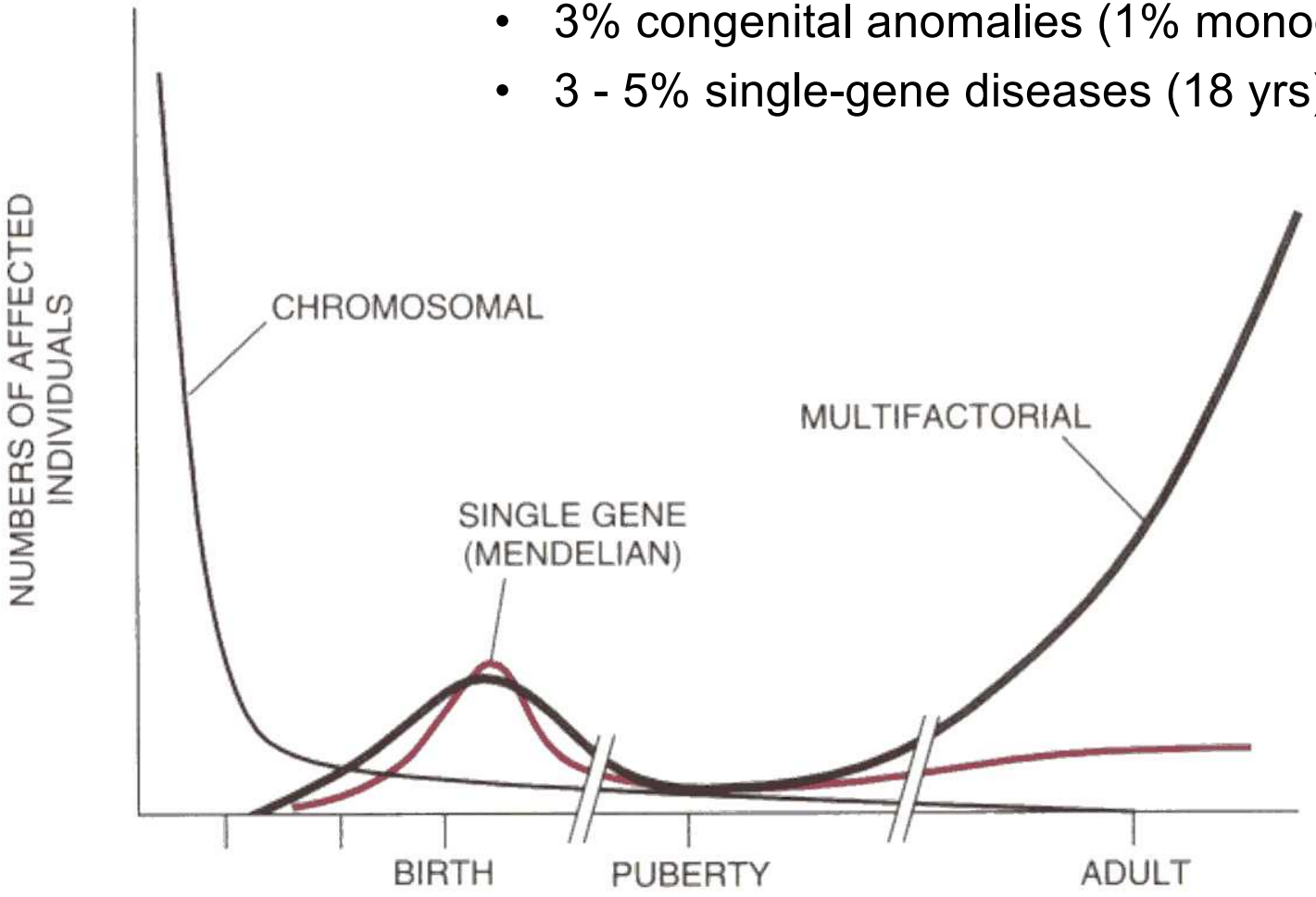
>5000 diseases. See http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim

Impact of genetic diseases

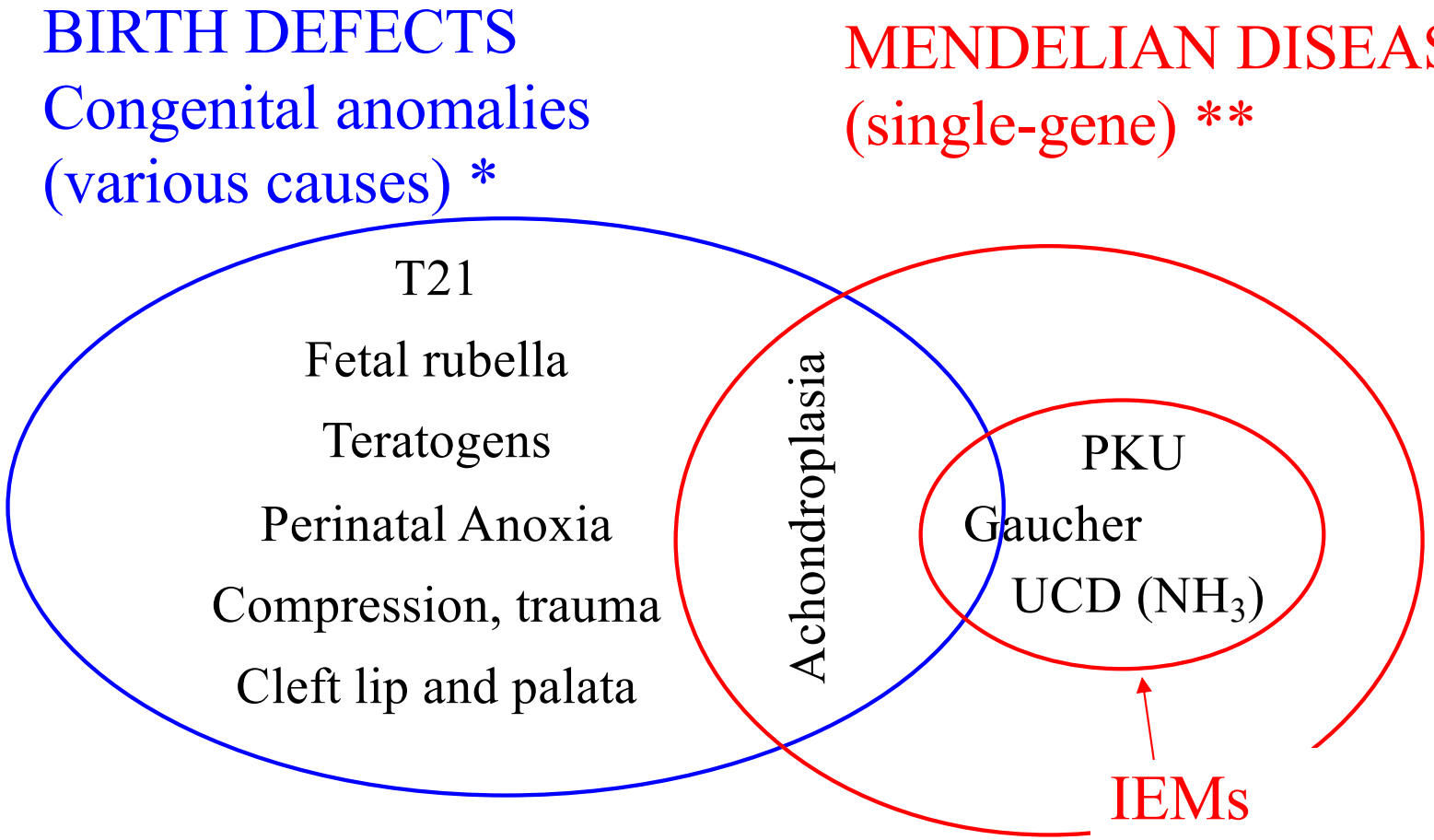
Incidence

Prevalence:

- 3% congenital anomalies (1% monogenic)
- 3 - 5% single-gene diseases (18 yrs)



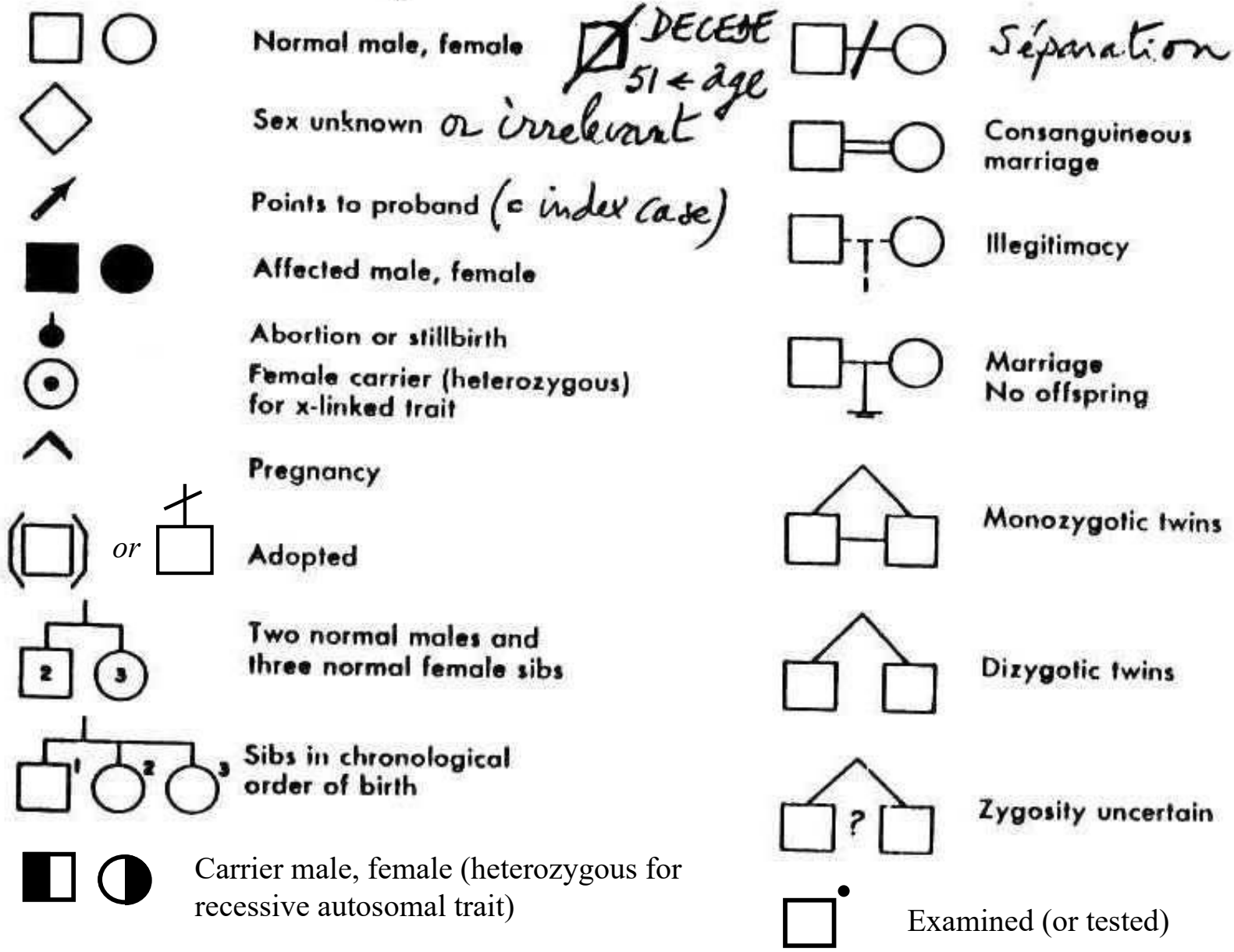
Births defects and Inborn Errors of Metabolism



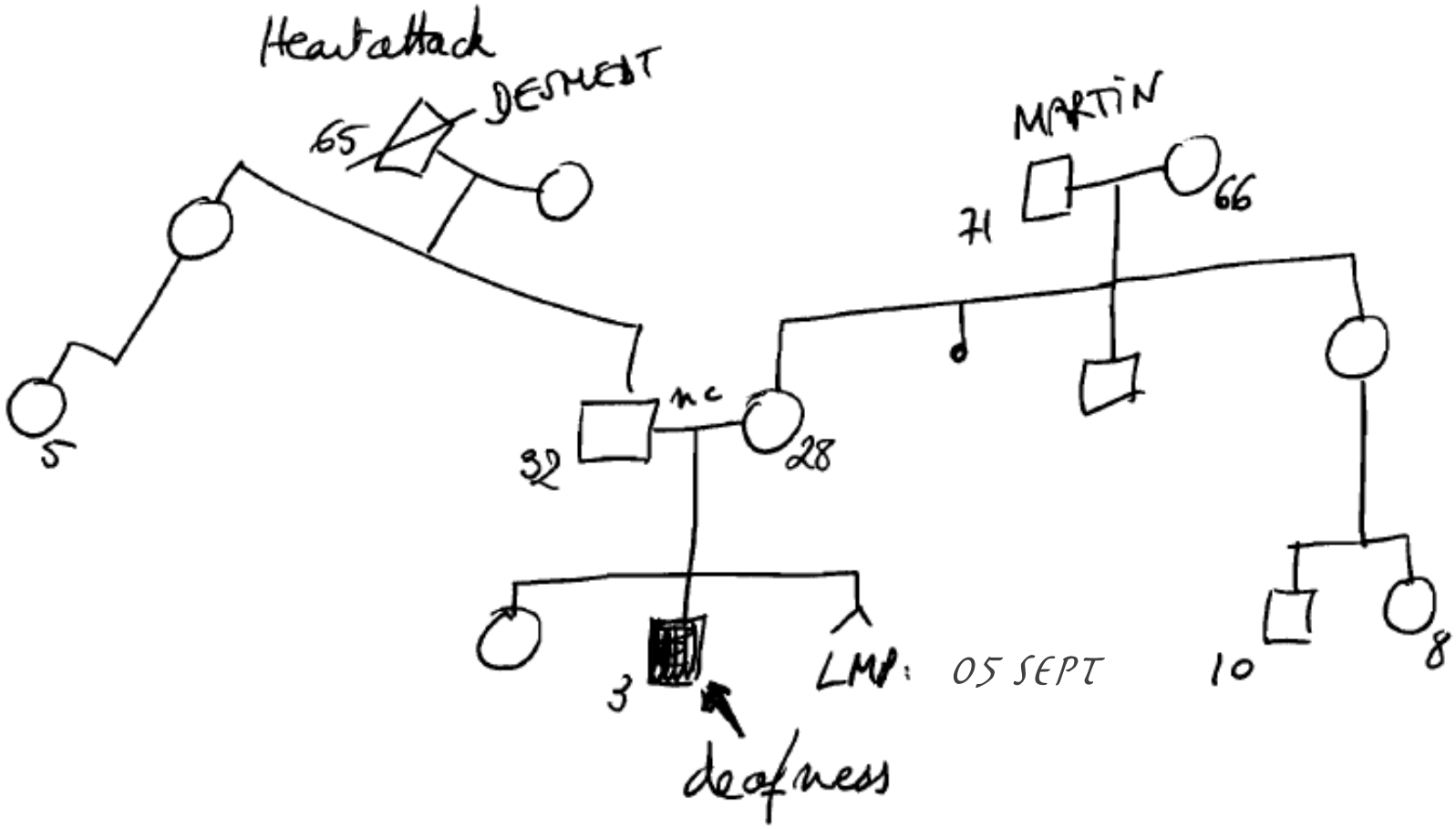
* 3% [live] newborns

** 5 % general population

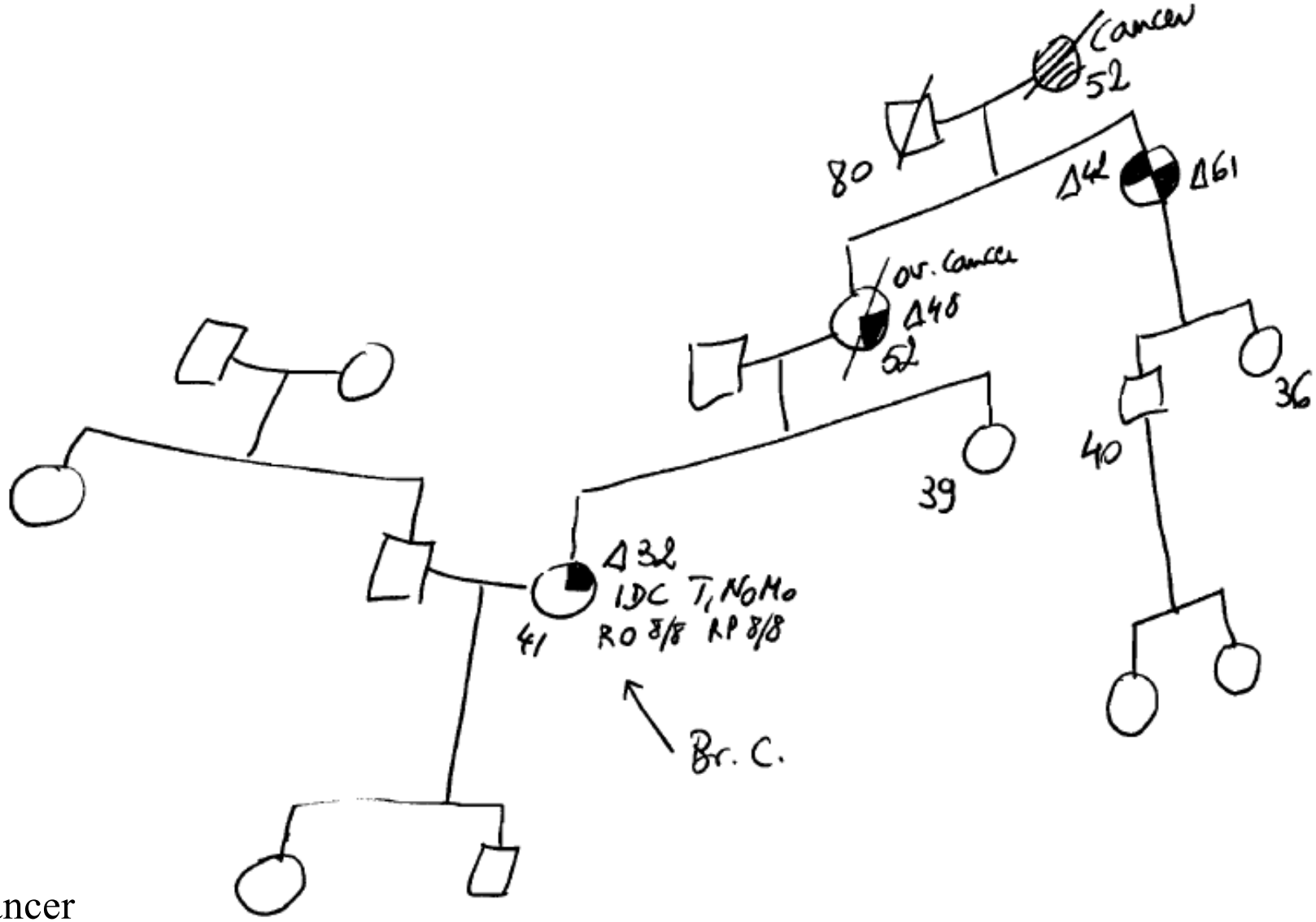
Symbols in pedigree charts



Working pedigree



Working pedigree (breast cancer)



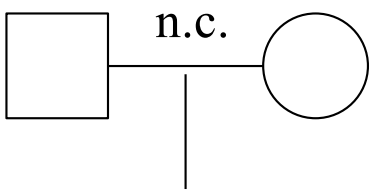
Breast cancer

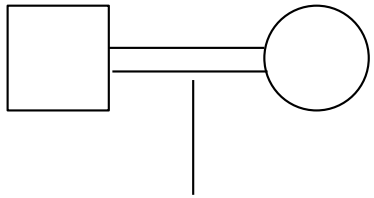
Ovarian cancer

consanguinity

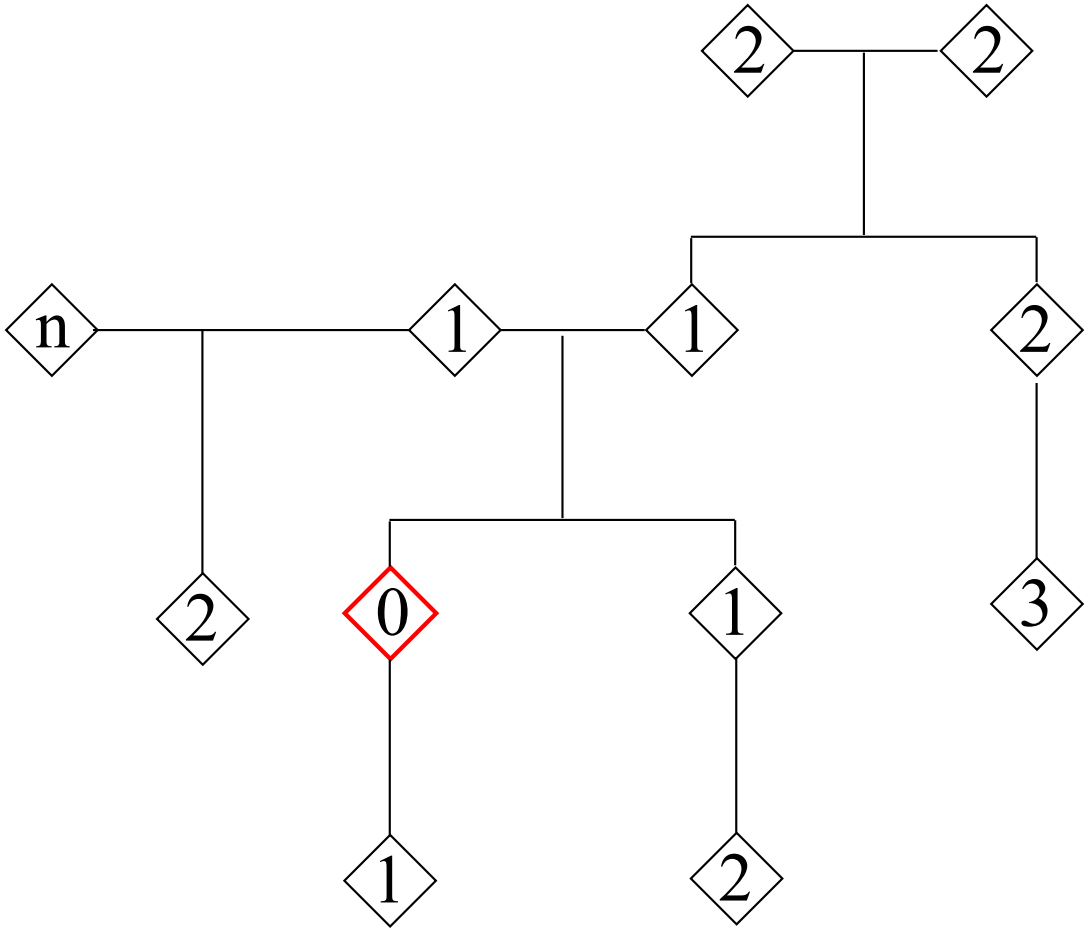
- Always enquire specifically about consanguinity
  - Are your parents cousins? (first cousins, …)
  - Are your grandparents (!) « cross-related » ?

- Annotate the pedigree, also if not consanguineous : ‘ n.c. ’





Degrees of relationship

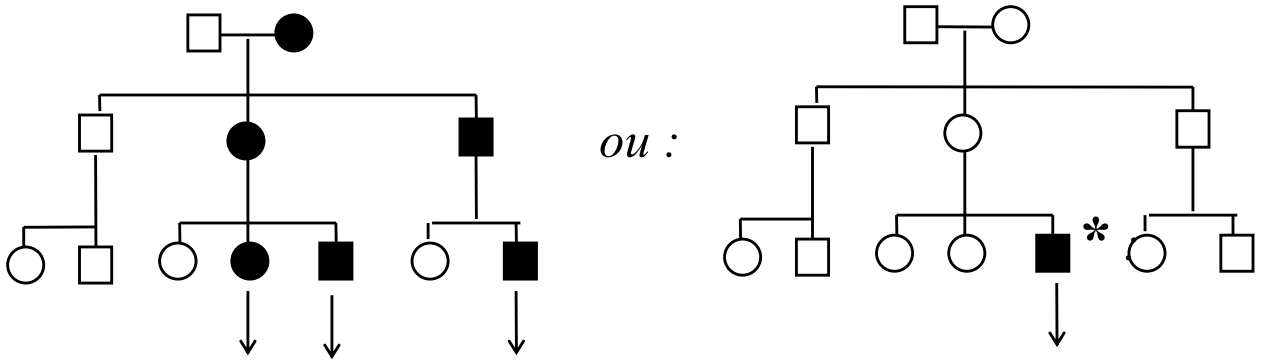


d°	relat	common alleles
1		.50
2		.25
3		.125

Patterns of single gene inheritance

AUTOSOMAL DOMINANT

AD phenotype: vertical transmission.
Genotype: hts mutation in autosome.



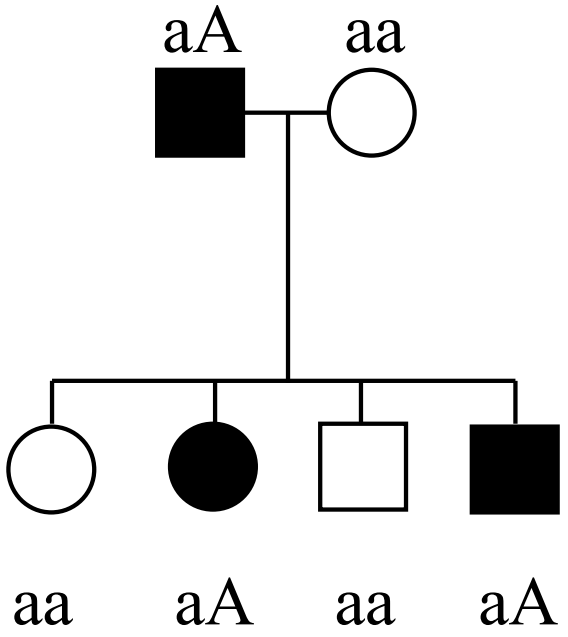
- Risk in each offspring = 50%
- M et F equally affected, equally transmitting
- Male to male transmission possible
- * : Neomutation (fresh mutation): AD disease starts here.
  - No ethnical prevalence (rare exceptions).
  - Increased mean paternal age.

Punnett square
Probability of genotype in offspring

IN FAMILY

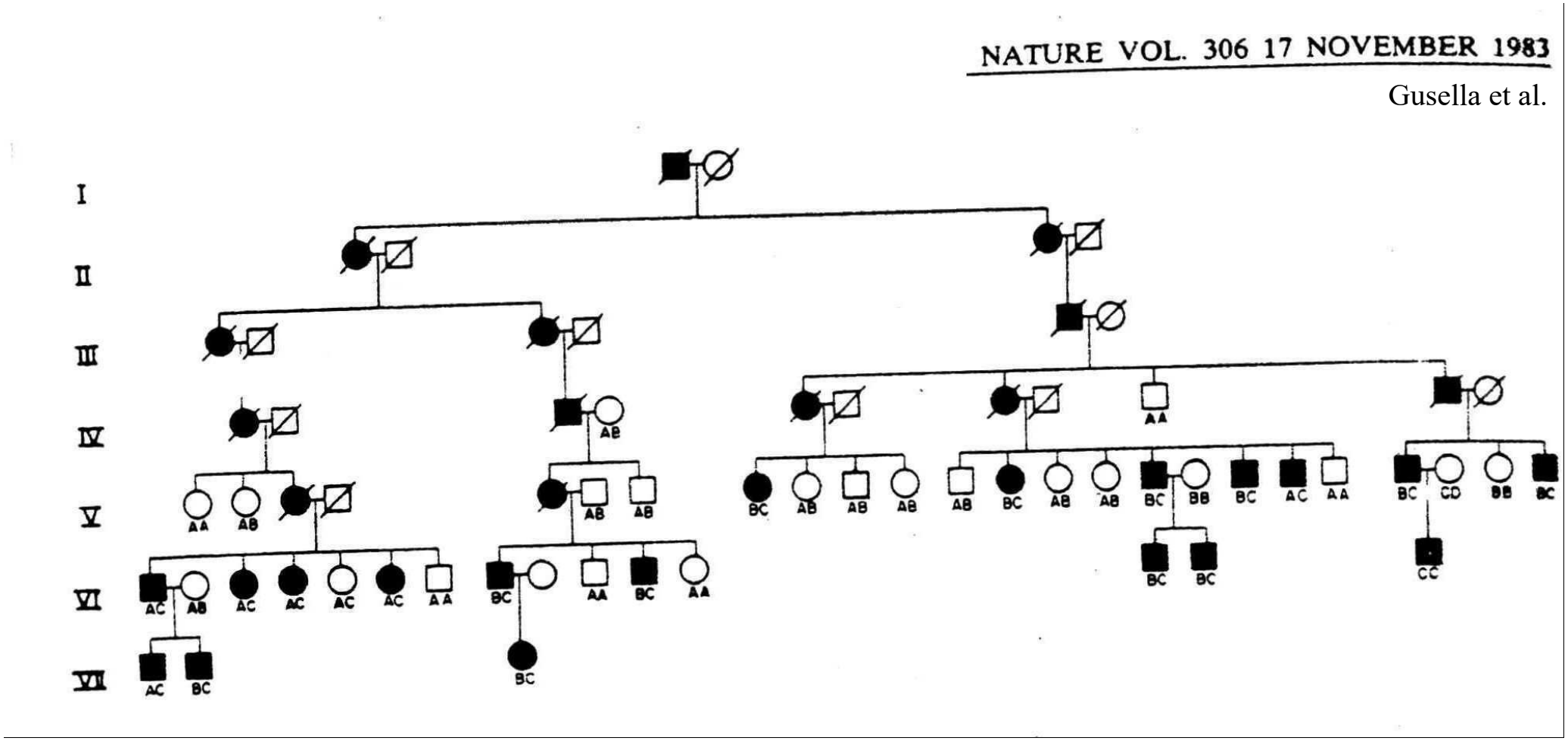
Male gametes

Female gametes	a	A
a	aa	aA
a	aa	aA



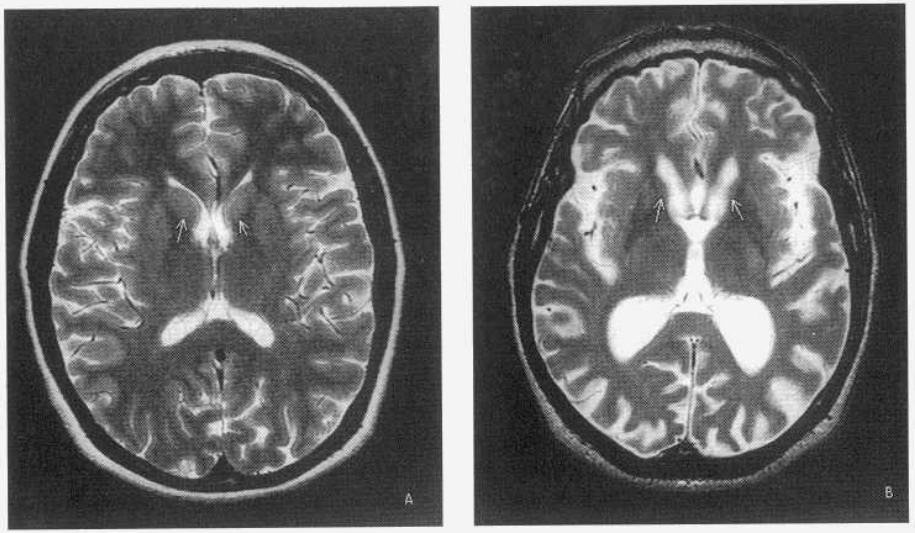
½
heterozygote
(affected)

Huntington : autosomal dominant inheritance



=> One single genetic factor causes disease (monofactorial, genetic).

Huntington disease



(a) Normal volunteer
(Courtesy of Dr M. Lowry, Hull, UK.)

(b) Huntington's disease

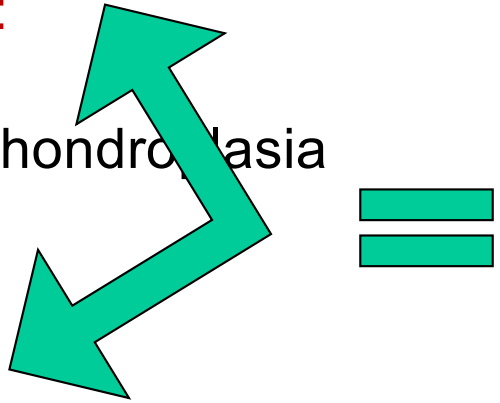
- Neurons in striatum (caudate nucleus) degenerate
- ↓ GABA

AD = approximation

- Dominance: phenotype independent of 2nd allele

- Semi-dominance (incomplete dominance):

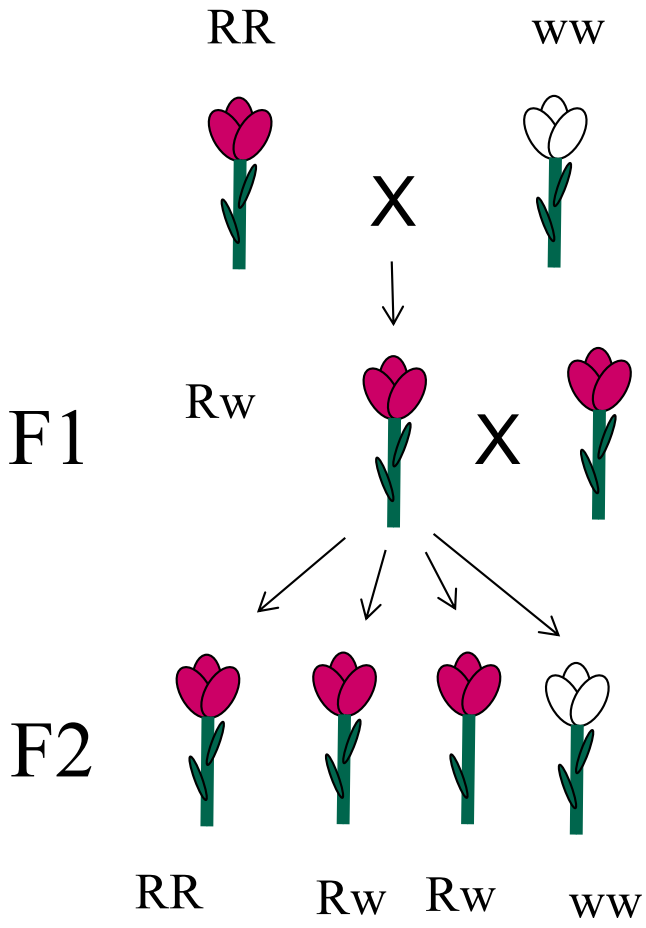
  hmz expresses trait more than htz : eg, achondroplasia



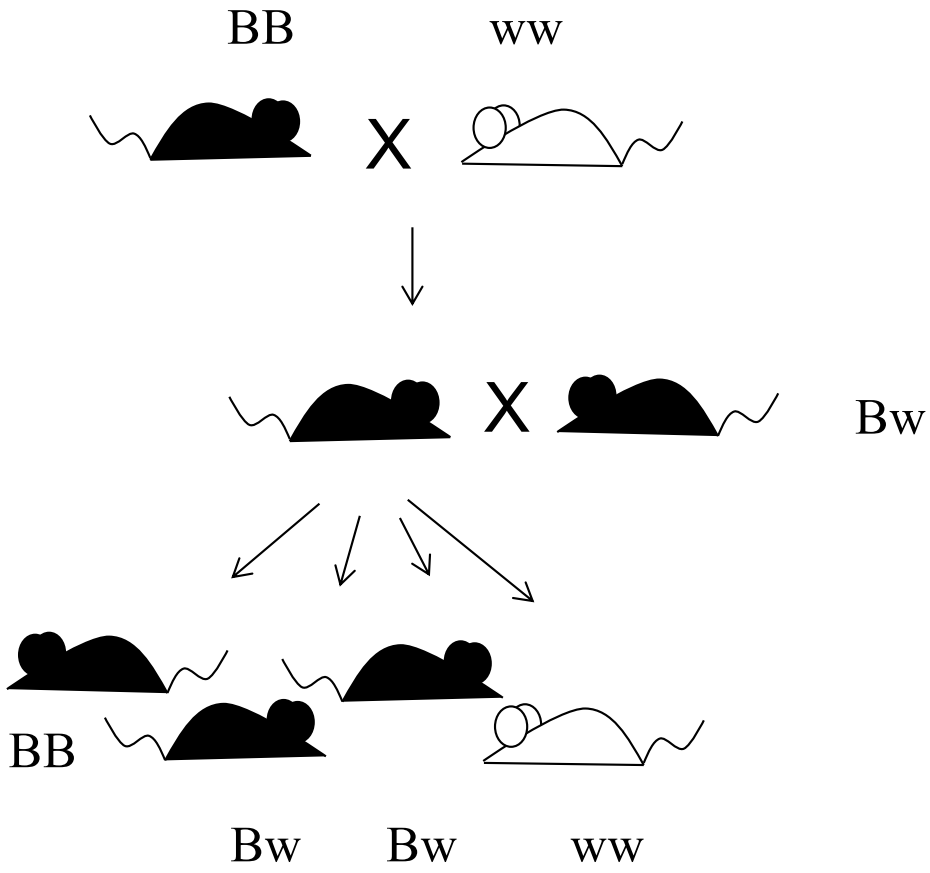
- Co-dominance:

  both alleles expressed : eg, ABO blood group

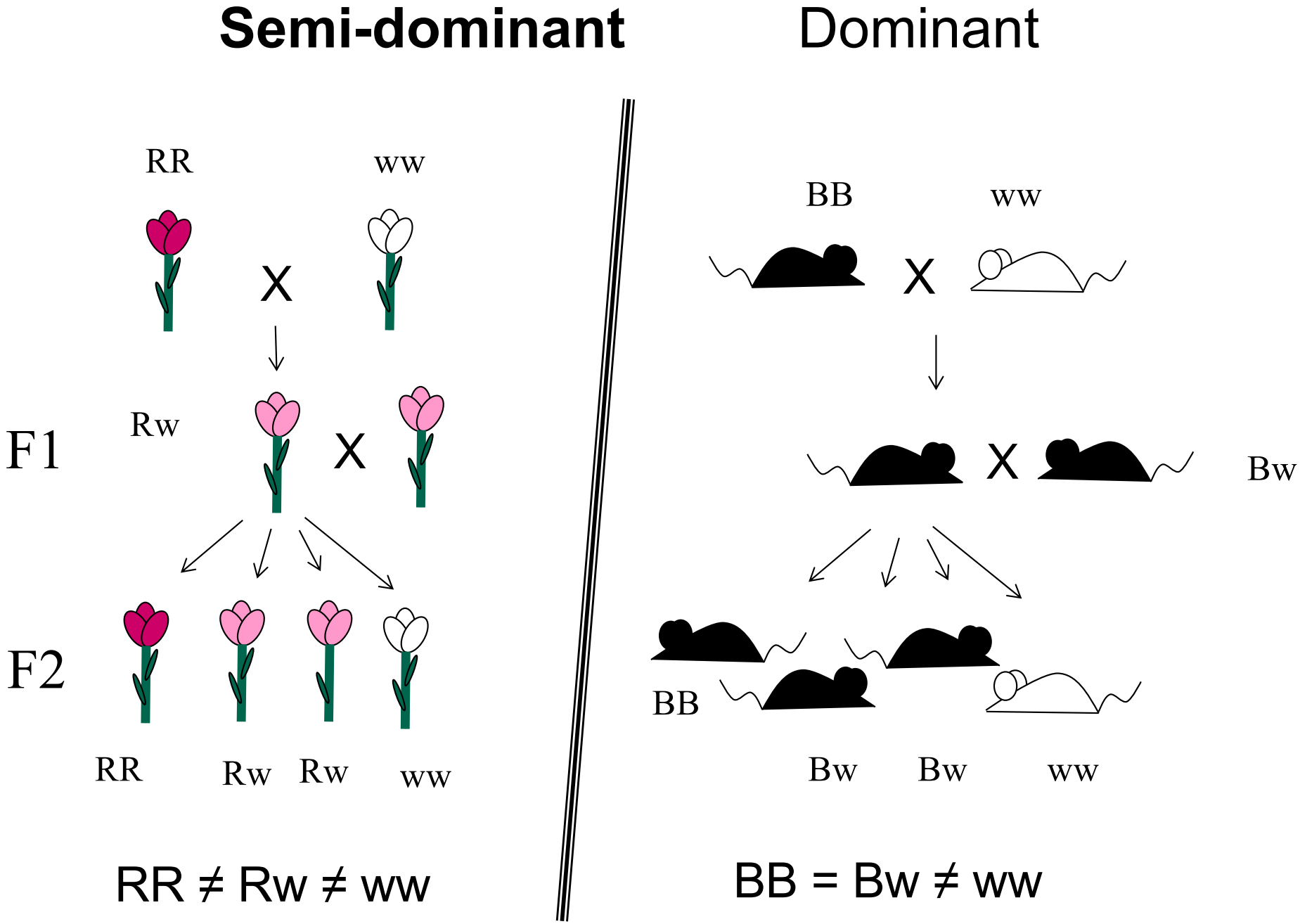
Dominant



RR = Rw ≠ ww

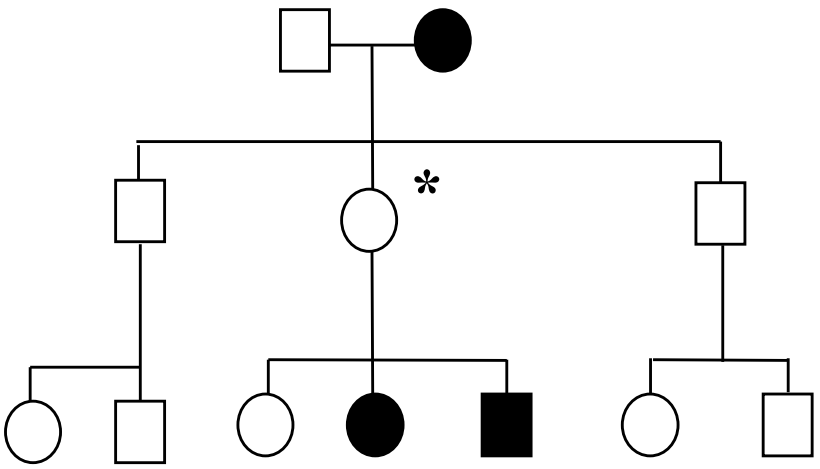


BB = Bw ≠ ww


Semi-dominant
Dominant
RR
ww
X
F1
Rw
X
F2
RR
Rw
Rw
ww
RR ≠ Rw ≠ ww
BB
ww
X
X
Bw
BB
Bw
Bw
ww
BB = Bw ≠ ww

Clinical variability of genetic phenotypes

- **PENETRANCE**: % of mutation carriers who express phenotype
- **EXPRESSIVITY**: clinical severity of the phenotype.



Incomplete penetrance:

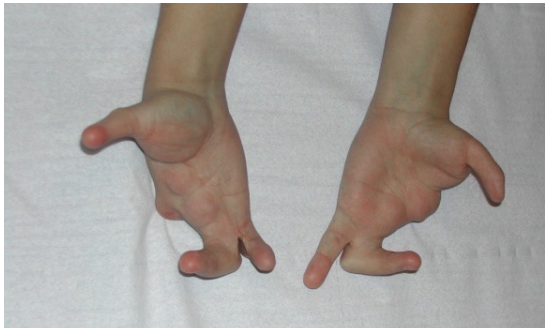
* « non-penetrant » subject :

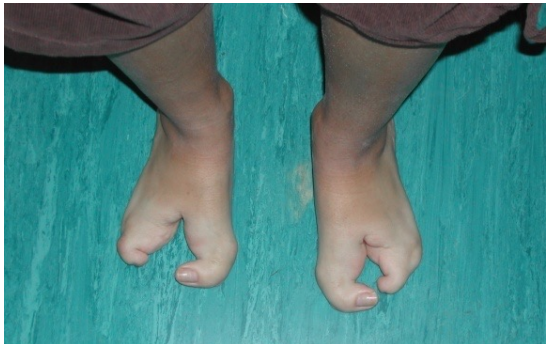
=> Age-related penetrance

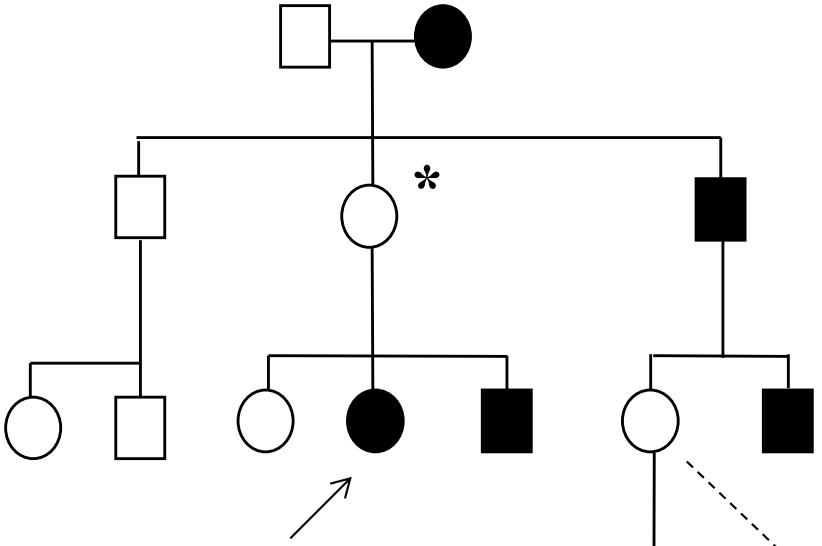
=> Sex-related penetrance

Incomplete penetrance

PENETRANCE: % of mutation carriers who express phenotype

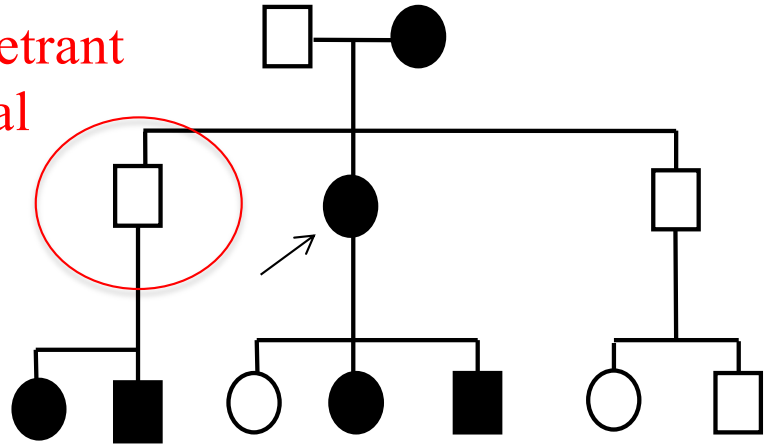






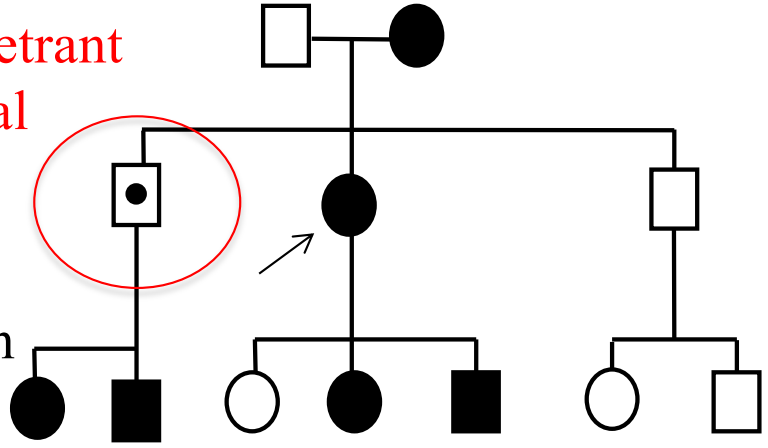
Incomplete penetrance

Non penetrant individual



Non penetrant individual

Dot in symbol = carrier of dominant mutation



Age-related penetrance

MEN2

- Medullary Thyroid Carcin
  Pheochromocytoma
  hyperPTH
- Some have MTC < 15 yrs
- 30% have no sign at 70 yrs

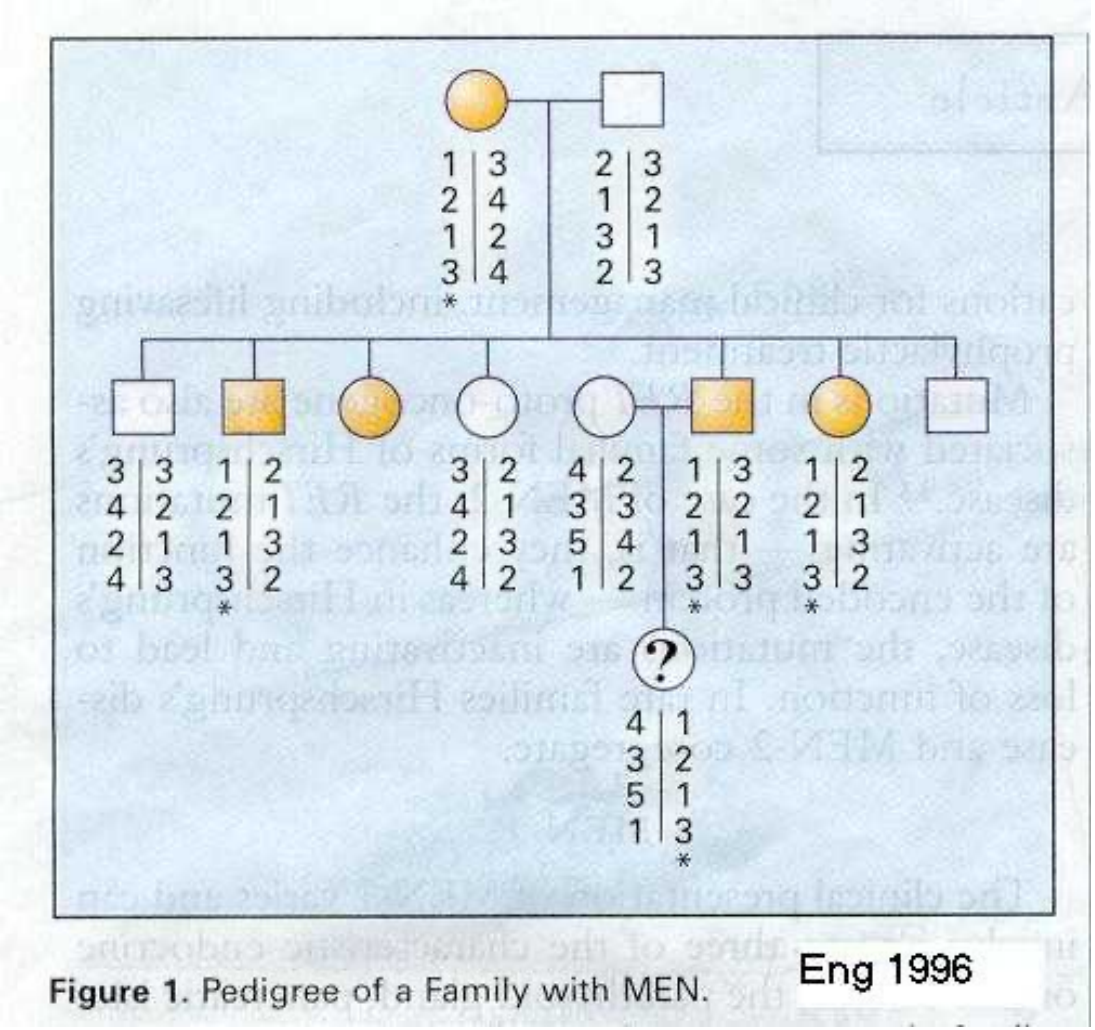
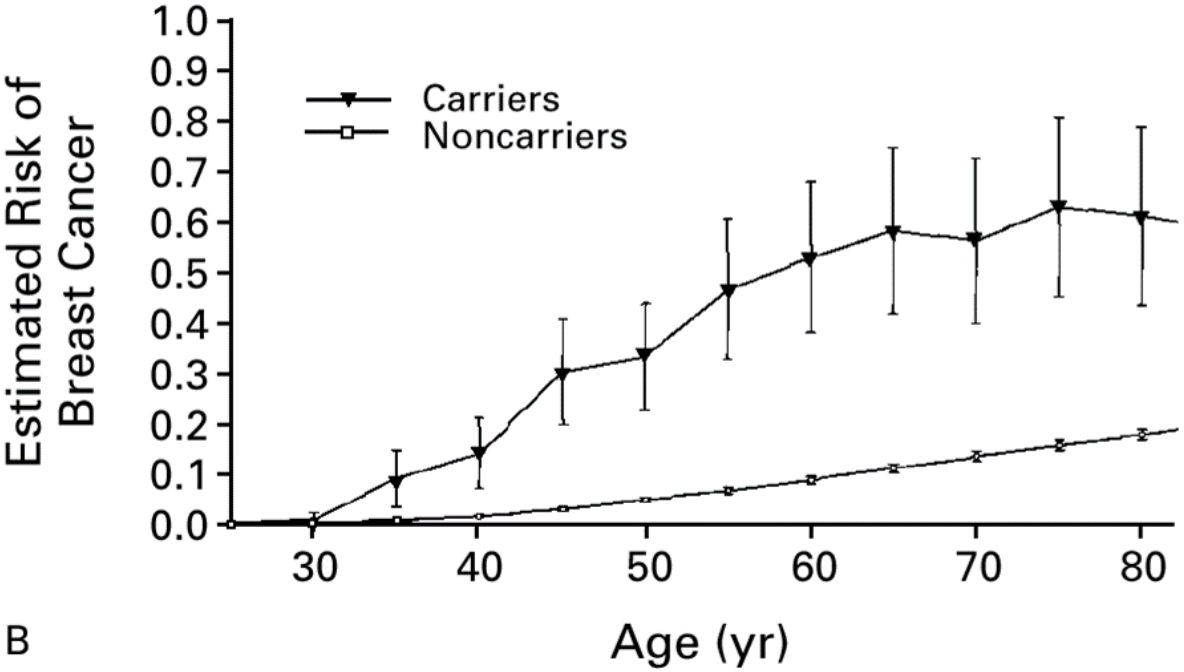


Figure 1. Pedigree of a Family with MEN.

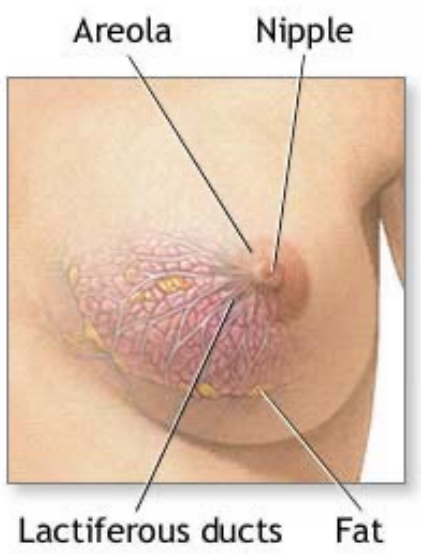
Sex- and age- related penetrance

Penetrance, BRCA1 – linked breast cancer



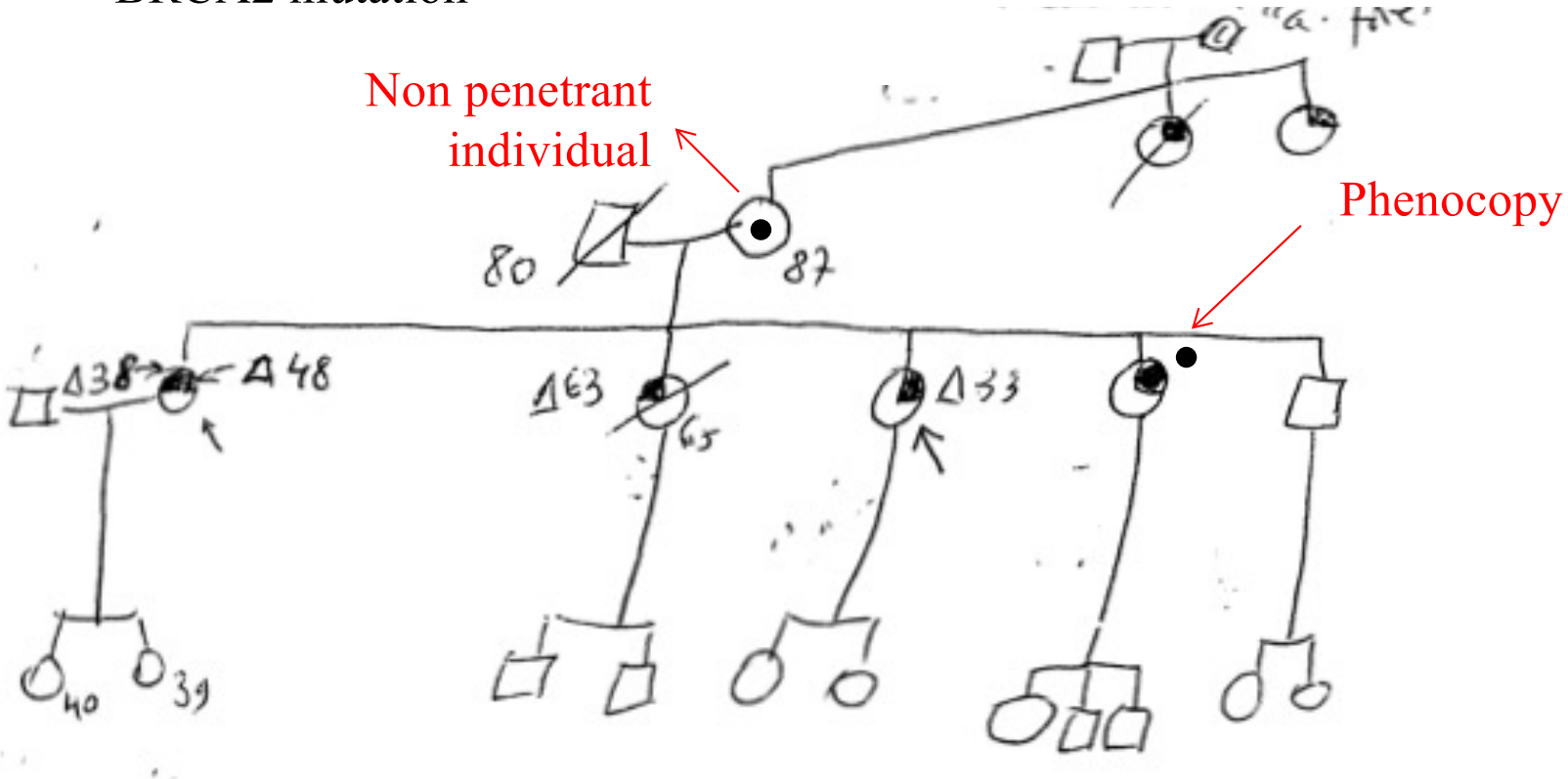
B

Struewing et al. 1997, NEJM 336: 1401-8.



Incomplete penetrance

BRCA2 mutation



Male-limited precocious puberty
is a sex-limited AD dis. expressed only in males

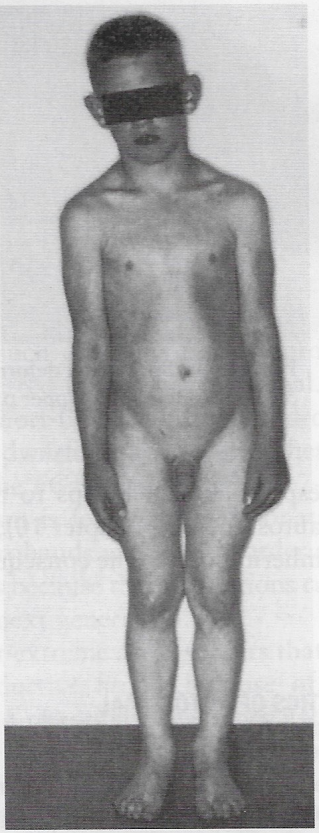


Figure 7-7 **Male-limited precocious puberty, a sex-limited autosomal dominant disorder expressed exclusively in males.** This child, at 4.75 years, is 120 cm in height (above the 97th percentile for his age). Note the muscle bulk and precocious development of the external genitalia. Epiphyseal fusion occurs at an early age, and affected persons are relatively short as adults.

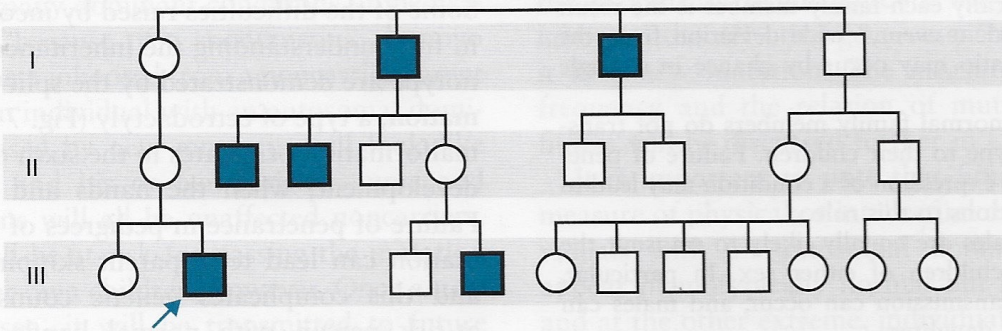
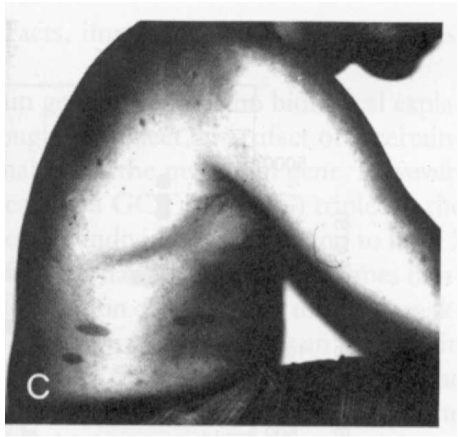


Figure 7-8 **Part of a large pedigree of male-limited precocious puberty in the family of the child shown in Figure 7-7.** This autosomal dominant disorder can be transmitted by affected males or by unaffected carrier females. Male-to-male transmission shows that inheritance is autosomal, not X-linked. Transmission of the trait through carrier females shows that inheritance cannot be Y-linked. *Arrow* indicates proband.

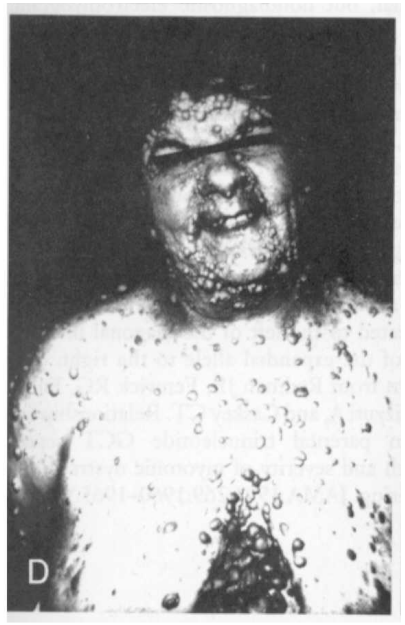
Variable expressivity

Mechanisms :

- Genetic
  - Mutated locus, 2nd allele
  - Modifyer gene(s)
  - Dynamic Mutations (rare)
- Epigenetic
- Environmental
- Stochastic

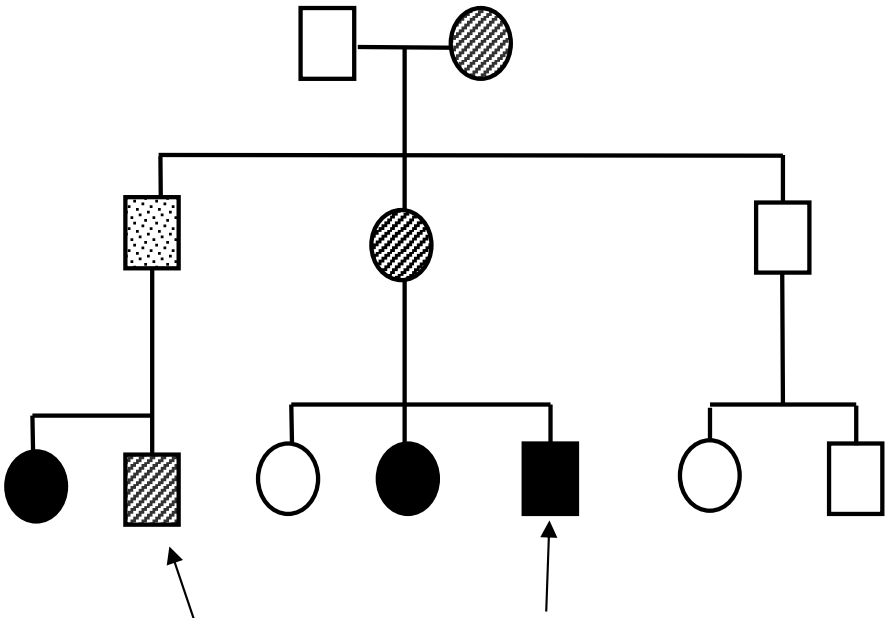


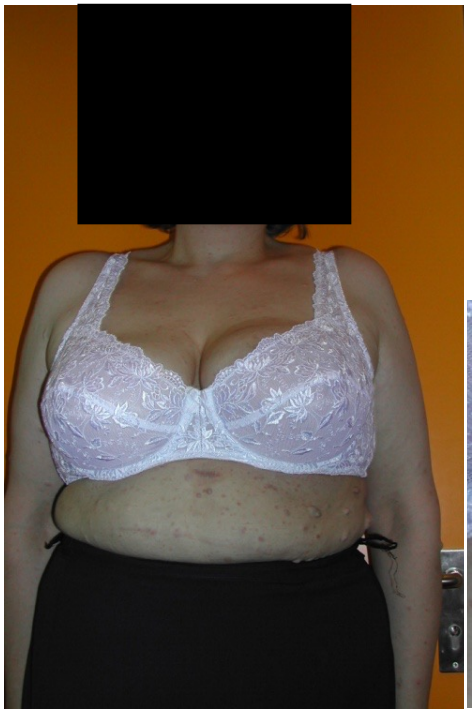
NF1

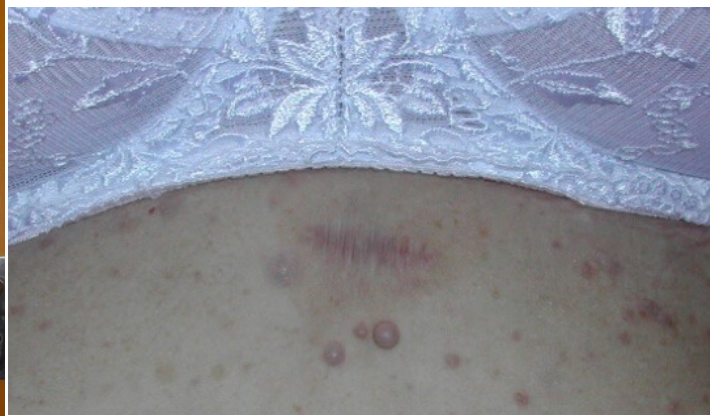


Variable expressivity, intrafamilial

- Same family, same mutation
- Hence, mere detection of mutation (eg prenatally) does not predict severity
- Especially if loss-of-fn mutation
- ex: NF1







Location of spots: truly random

Huntington : anticipation

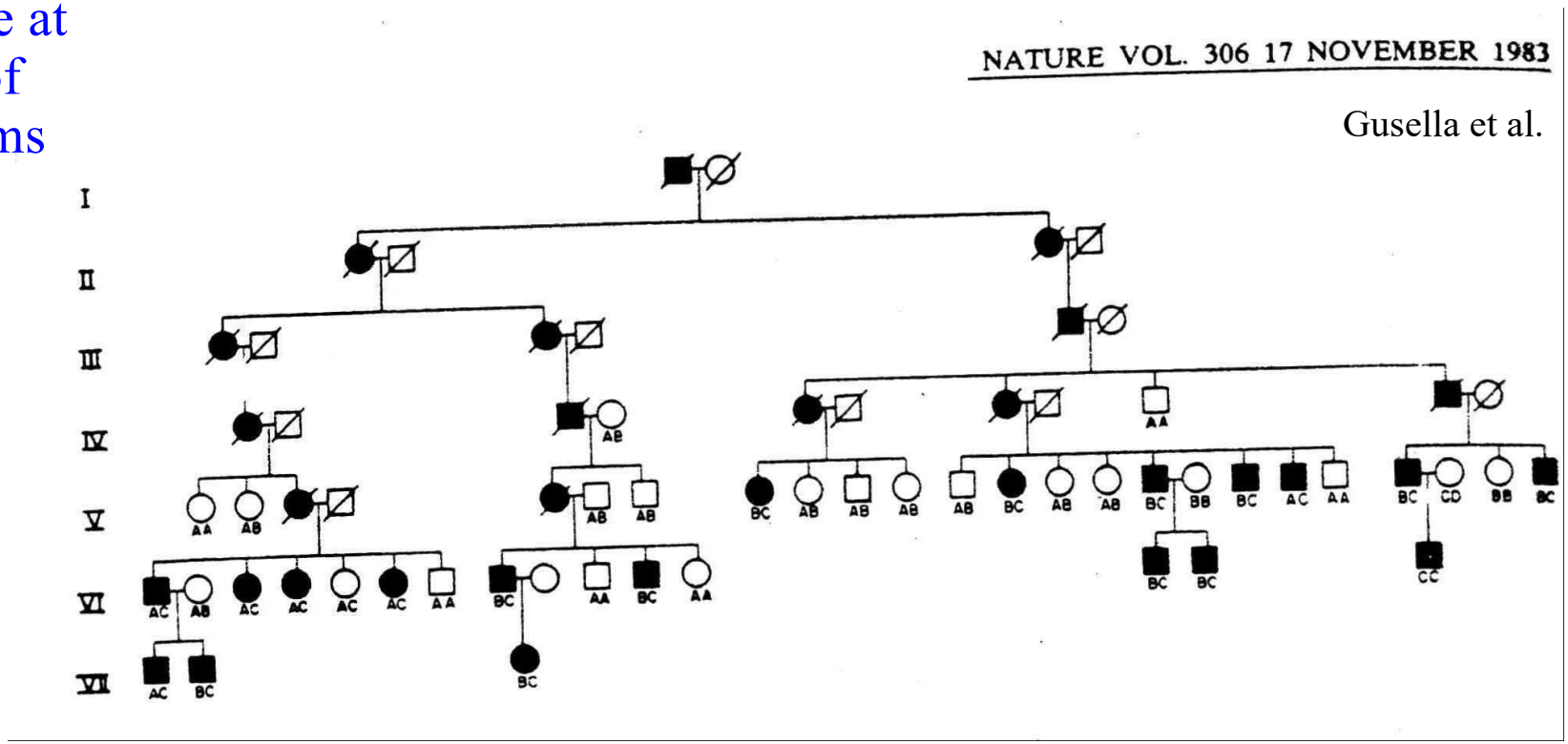
Mean age at onset of symptoms

70

60

50

40



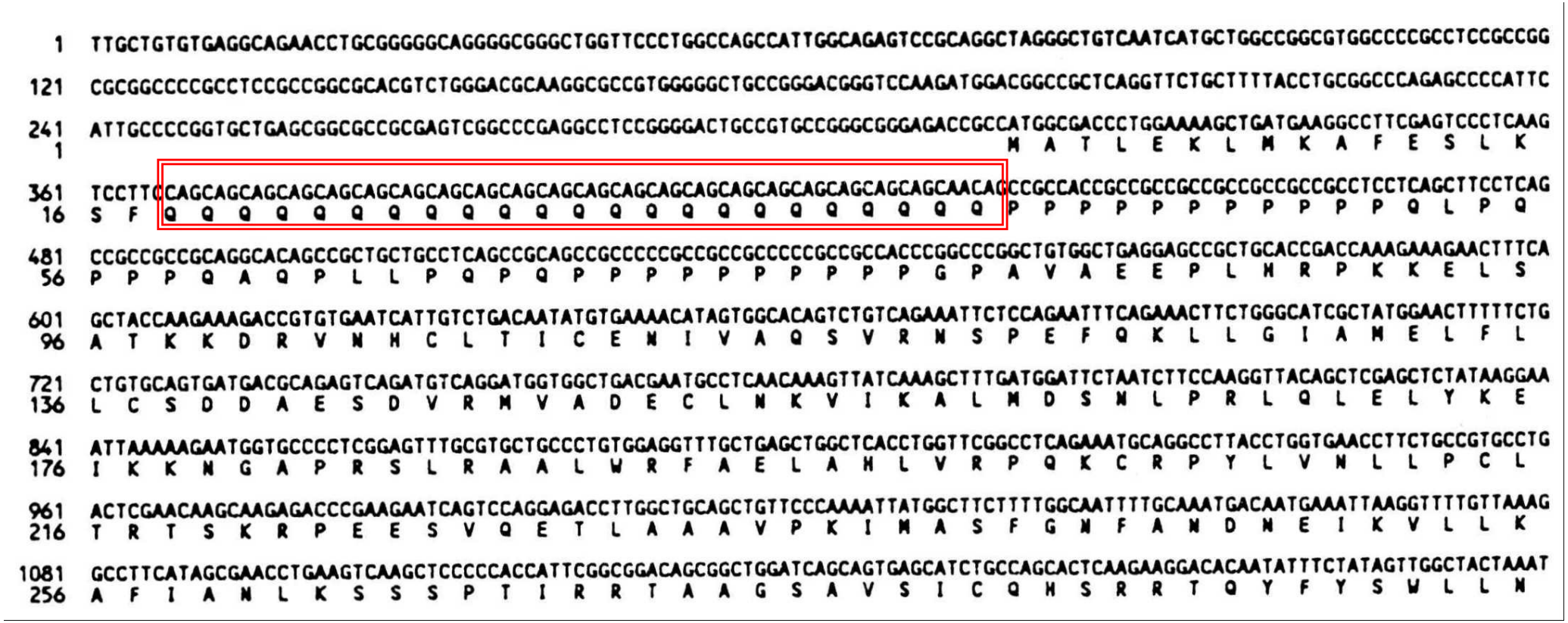
Anticipation = clinical observation (phenotype)

Molecular correlate : progressive expansion of triplets with generations

Near-mendelian inheritance

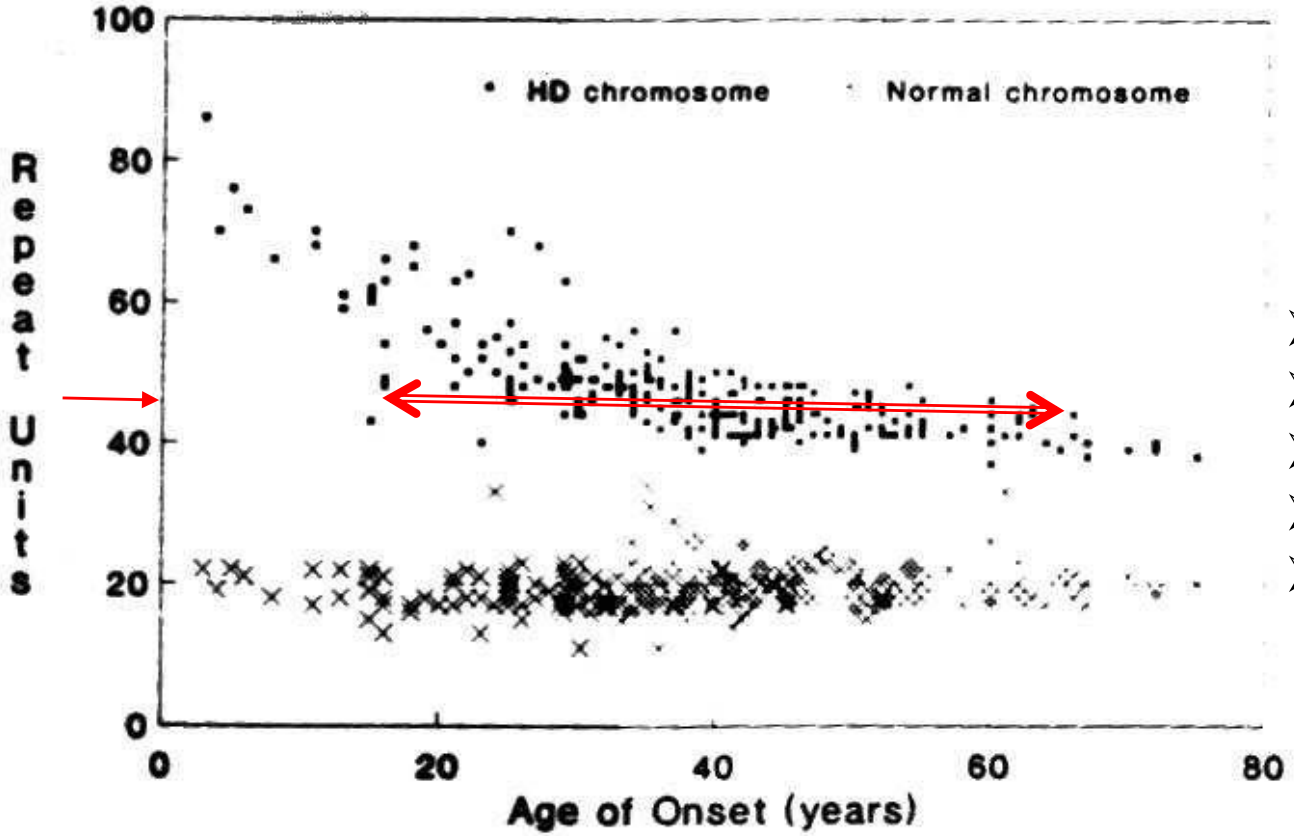
- Penetrance — all-or-nothing presence of phenotype (on/off button)
  - Complete penetrance
  - Incomplete penetrance
- Expressivity — quantitative (volume button)
  - Mild, moderate, severe expressivity
- Phenocopy — acquired, non-inherited mimic
- Anticipation — increased expressivity over generations

Huntington Disease(HD)



- (CAG)n, coding for (Gln)n, N-term part
- n is polymorphic, < 30 , stable, in general population
- n est > 40 in HD chromosomes . And unstable when transmitted, especially through male.
- Larger n causes more severe disease. Explains 2/3 of variability.

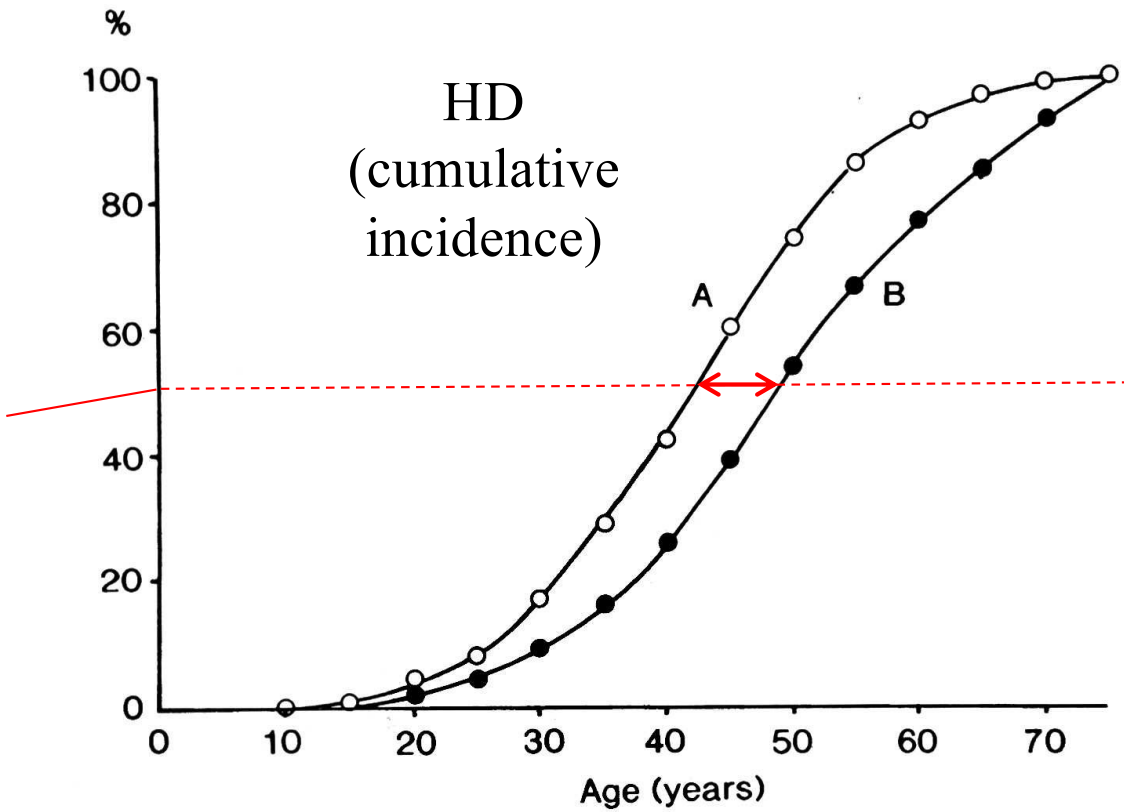
HD: dynamic mutation



- Juvenile HD alleles (>50)
- Classical HD alleles (>36)
- Low penetrance alleles
- Unstable alleles
- Normal alleles

- n ↑ ⇒ age at onset ↓.
- Statistical only.
  No reliable individual predictions.
- Anticipation parallels ↑ n over generations.

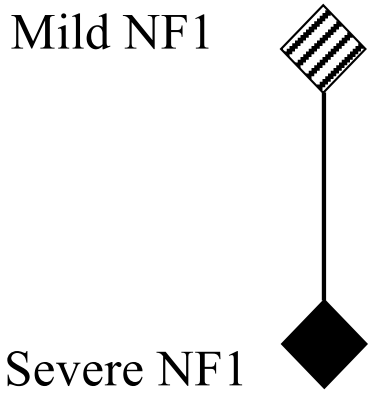
Ascertainment bias (not linked to anticipation)



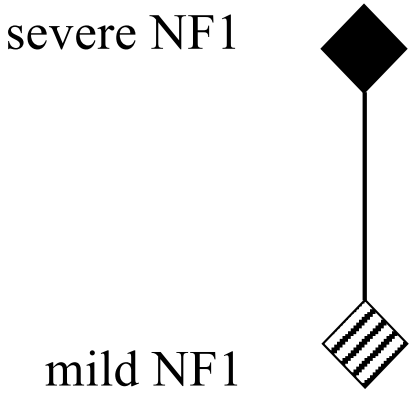
Median age at onset of symptoms

- A: retrospective: patients who consulted because of symptoms
  → biased for increased severity, earlier onset
- B: prospective: mutation carriers.
  → Includes those who would not have consulted.

Variable expressivity interpreted as anticipation



May be reported as possible anticipation



Not reported as possible anticipation

Less frequent (reduced fitness)

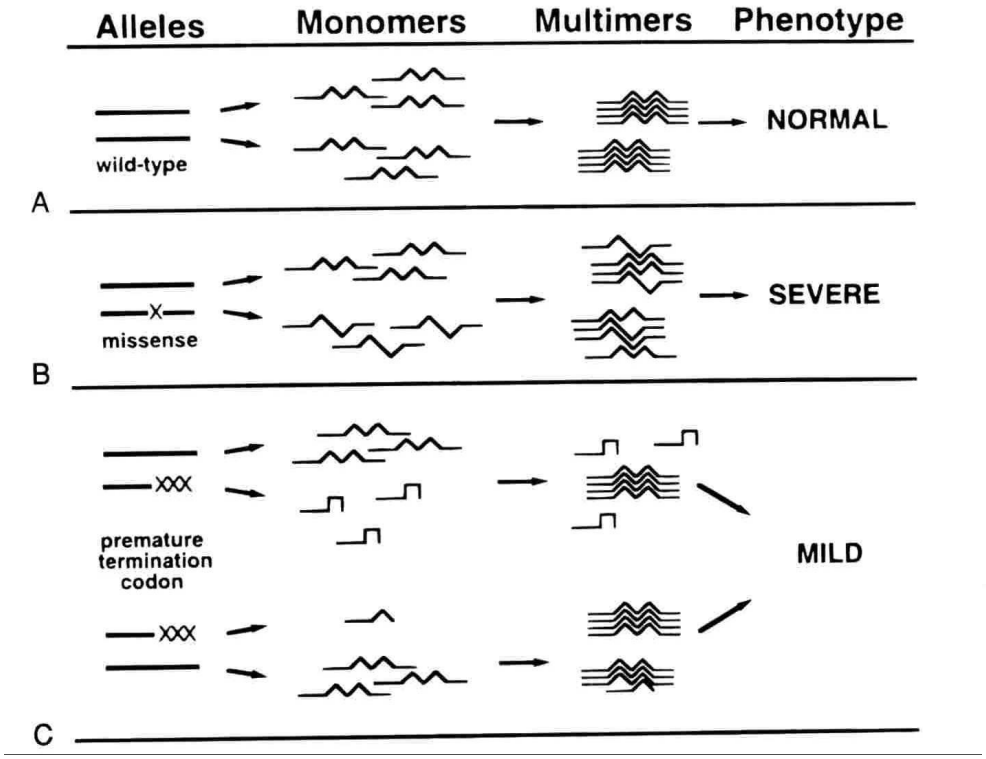
Some mecanism for dominance

Most genes have robust functional reserve:

>10% gene activity (e.g. enzymic) enough for normal fn
==> why phenotype in htz?

- Haploinsufficiency (>50% not enough) → Acute intermittent Porphyria
- Gain of toxic fn → Huntington
- Dominant negative effet (multimer) → Marfan, THR,
- Somatic mutation of 2nd allele frequent → Cancers héréditaires
- Dose effect (triplication) → Charcot Marie Tooth
- Ectopic expression → Corticoids-remediable HTN

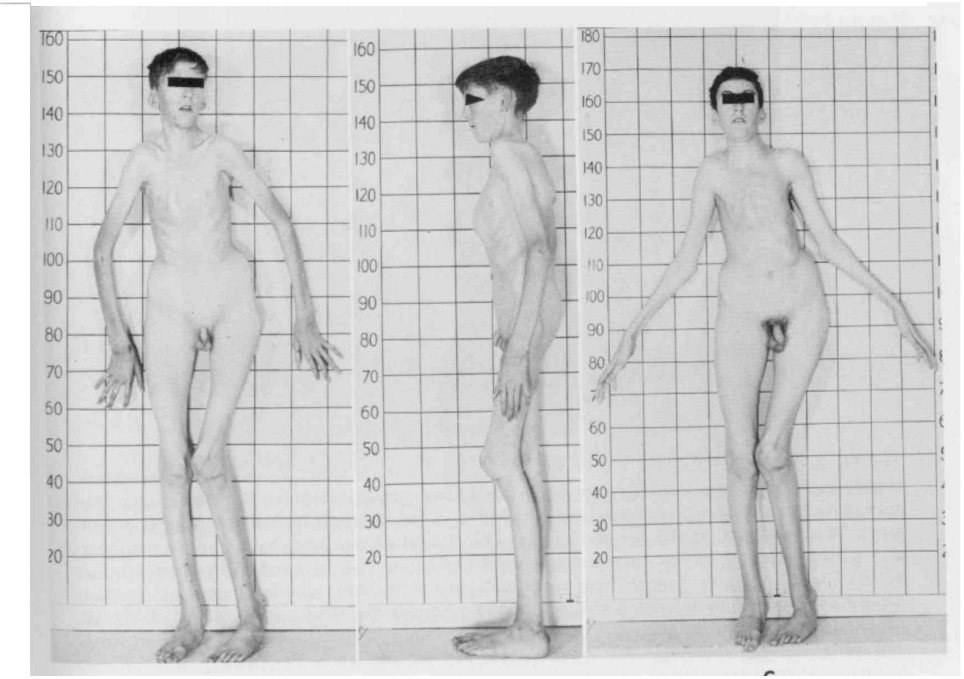
Dominant negative (antimorphic) alleles



- Mutated allele loses fn AND interferes with wt allele
- >1 subunit (dimers, multimers): 1 mutation hampers whole structure

- Usually causse more severe phenotype than null mutation
  - Ostogenesis imperfecta
  - Marfan

Marfan Syndrome (FBN1 gene)



Mutations affect Fitness

- Natural selection favours or hampers chances to transmit gene
  - Survival, up to reproducing age
  - Find a mate (sexual selection)
  - Be fertile
  - Raise children to reach reproducing age
  - …
- Positive selection (adaptive change)
- Negative selection (purifying selection)
  - Fitness = (# offspring) / (mean # offspring in population)

ex: f=.95

after 10 generations: $.95^{10} = .60$

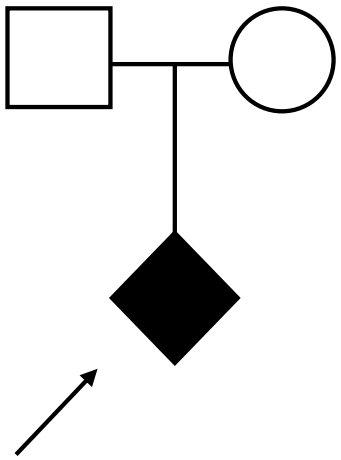
after 20 generations: $.95^{20} = .36$

after 100 generations: $.95^{100} = .0060$

Neomutations (de novo mutations)

- Sporadic. No ethnic preponderance
- Cause a fraction of AD cases, disease-specific
- Fraction reflects effect of disease on *fitness* (f)

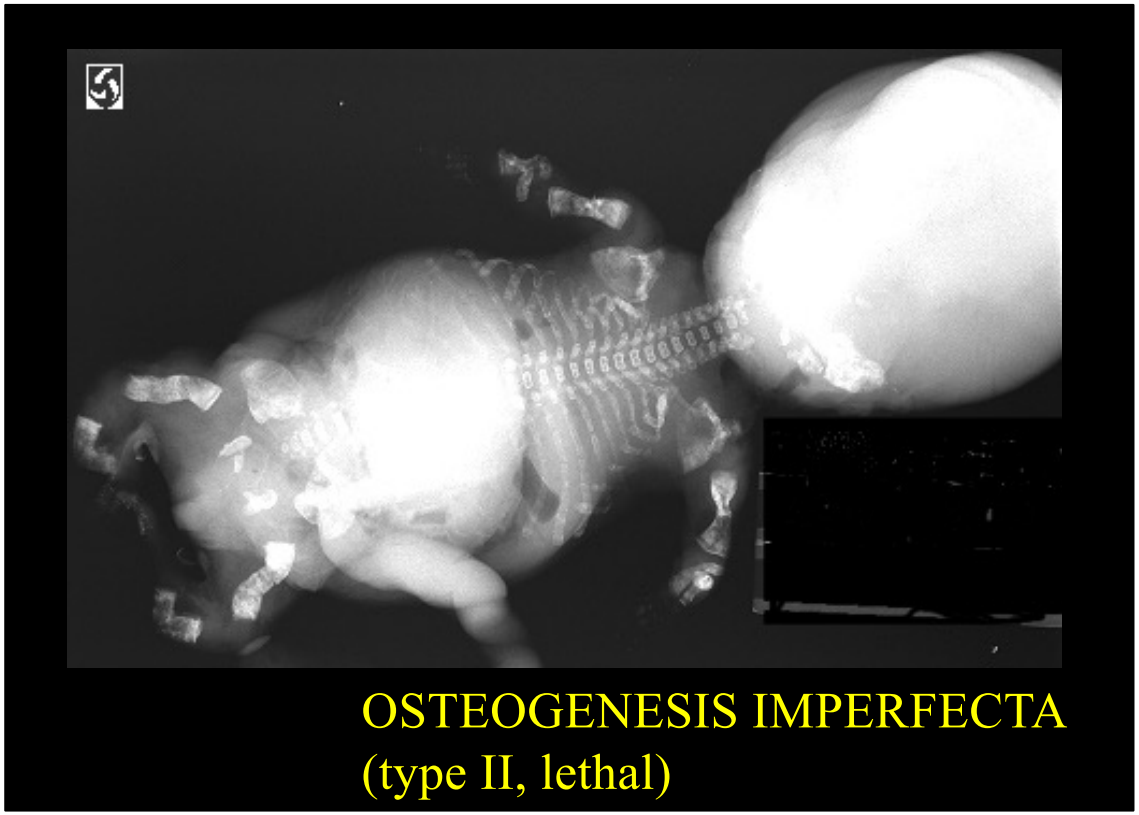
  f = (No offspring of individual) / (mean No offspring in population)

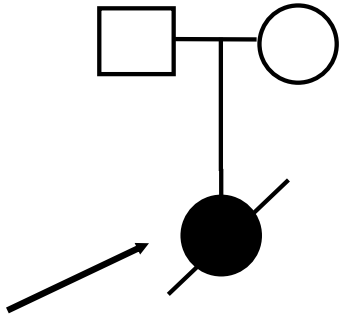


disease	*% neomutations*
Huntington Chorea	*< 1%*
Fam Adenom Polyposis	*10-25%*
Polykystosis	*25%*
NF1	*50%*
Tuberous sclerosis	*80%*
Achondroplasia	*90%*
Lethal OI	~100%

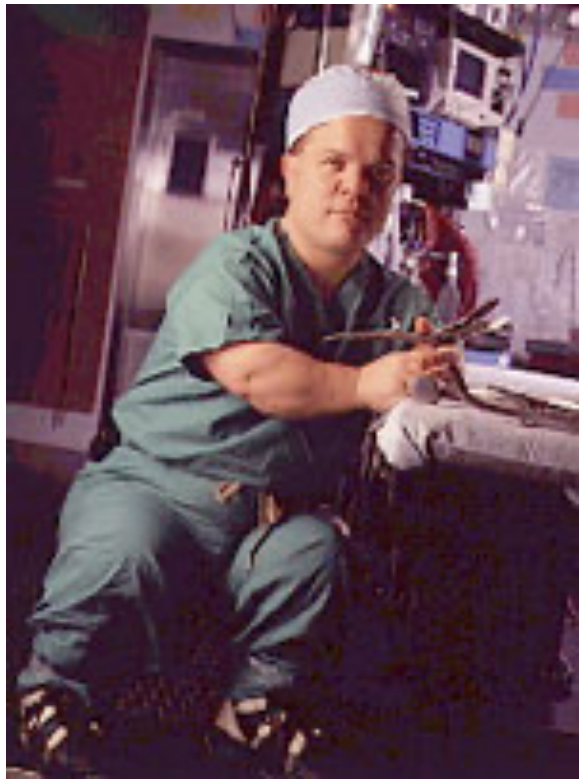
New mutation lethal OI

- Procollagen gene, hts mutation, lethal phenotype (fitness = 0)
- New mutations only



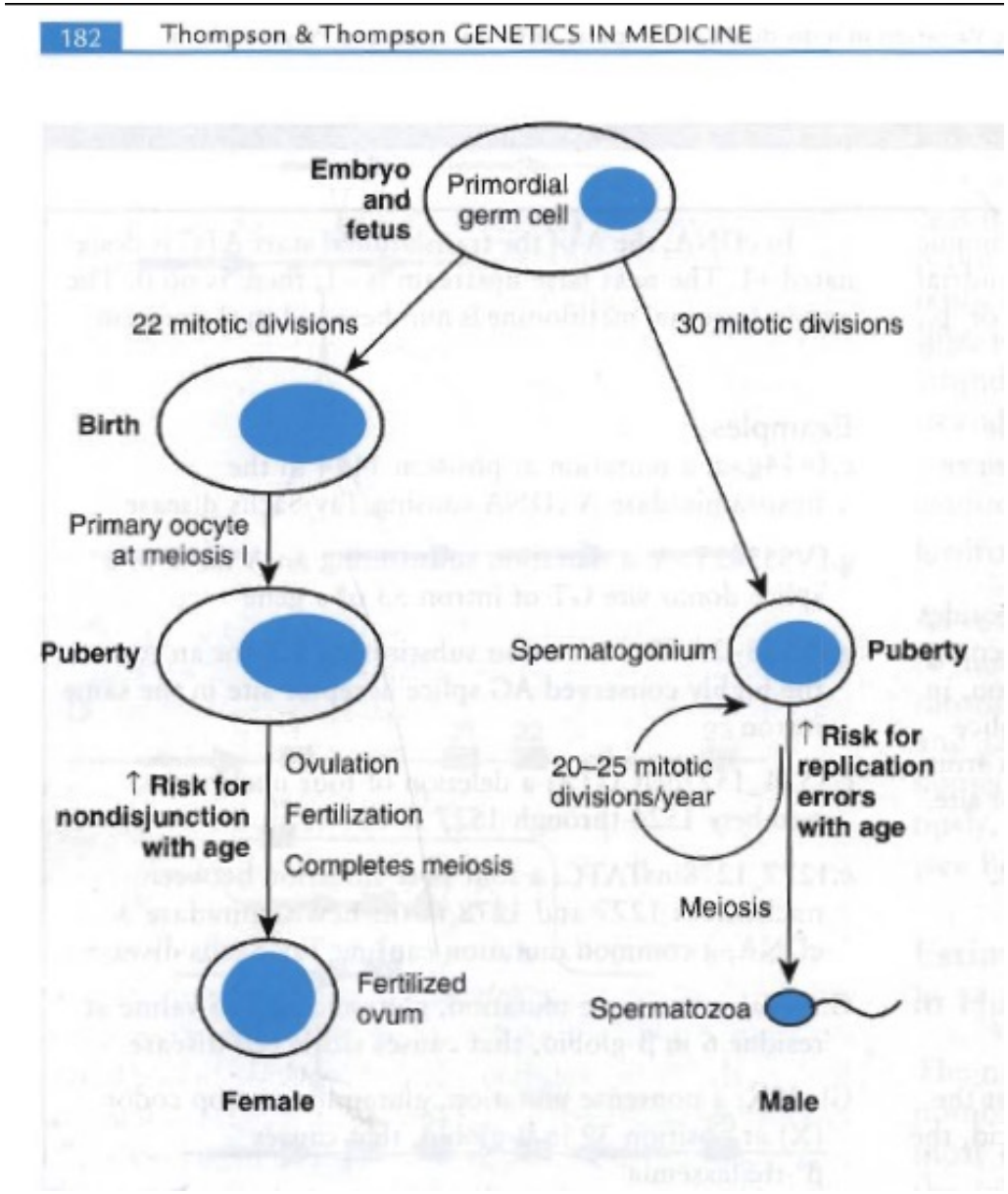


Achondroplasia



- FGFR3 gene
- Neomutations => no LD with close markers in different subjects

New mutations are more frequent in male germ-line + paternal age effect



DNA replication:
mutation rate 10^{-10}
2 x 3.10^9 bp/cell
10^{11} cells

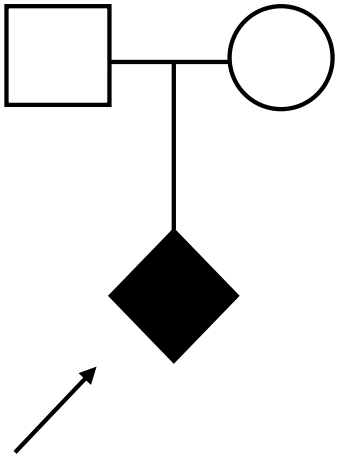
This is true for point mutations

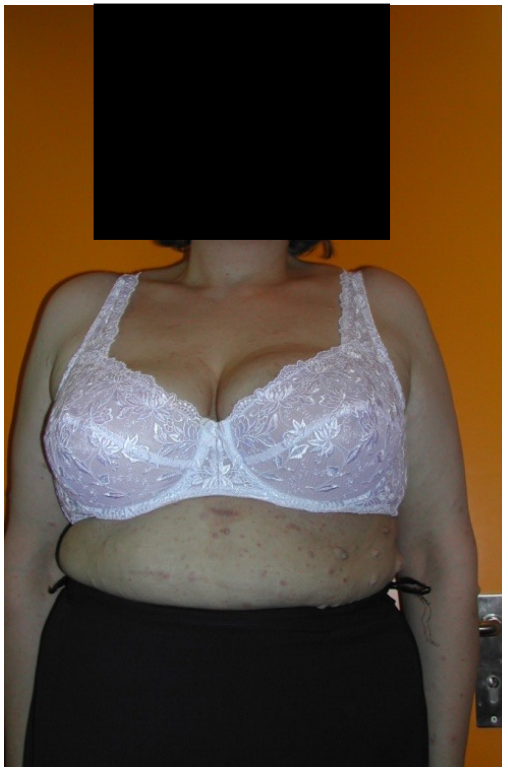
Large deletions de novo are more frequent in female gametogenesis

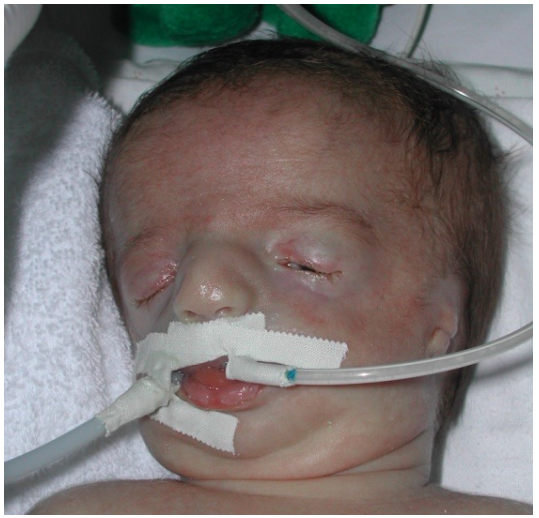
New mutations in AD disease

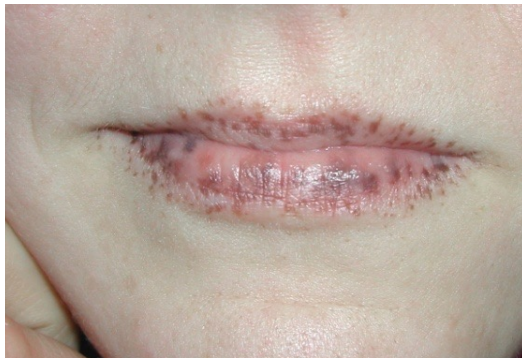
- f = fitness
  μ = mutation rate / generation
  q = allele frequency

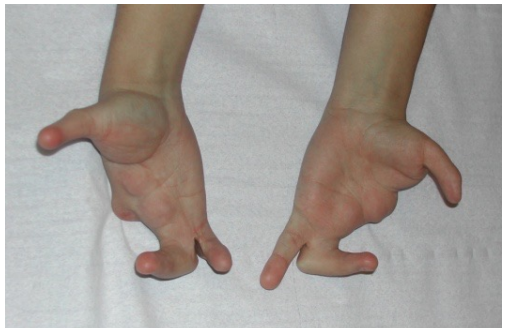
- μ = (1-f)q
  f = 0 → q = μ

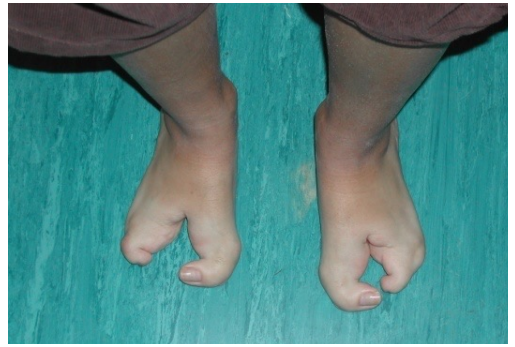




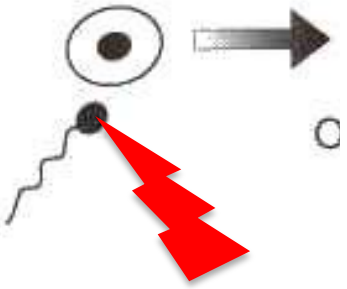








Neomutation in germline

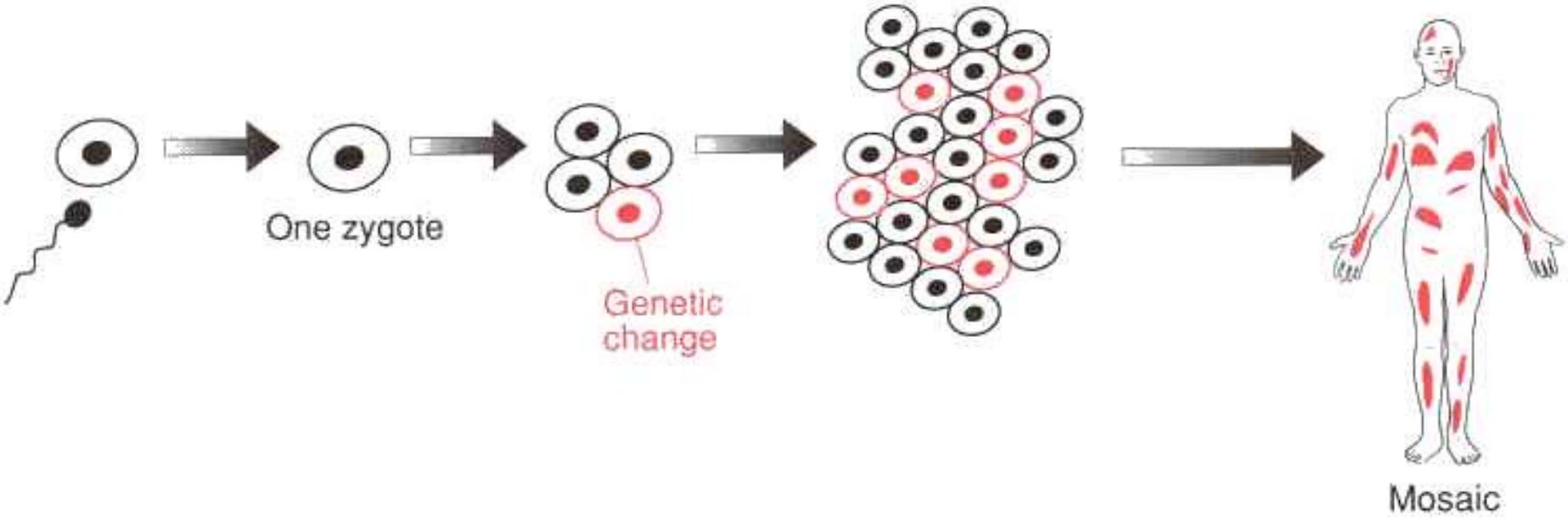


- Affects one allele
- In one gamete
- heterozygous
- May be lethal in utero
- Or asymptomatic
- Or in between: phenotype in heterozygous carrier subject

Typically

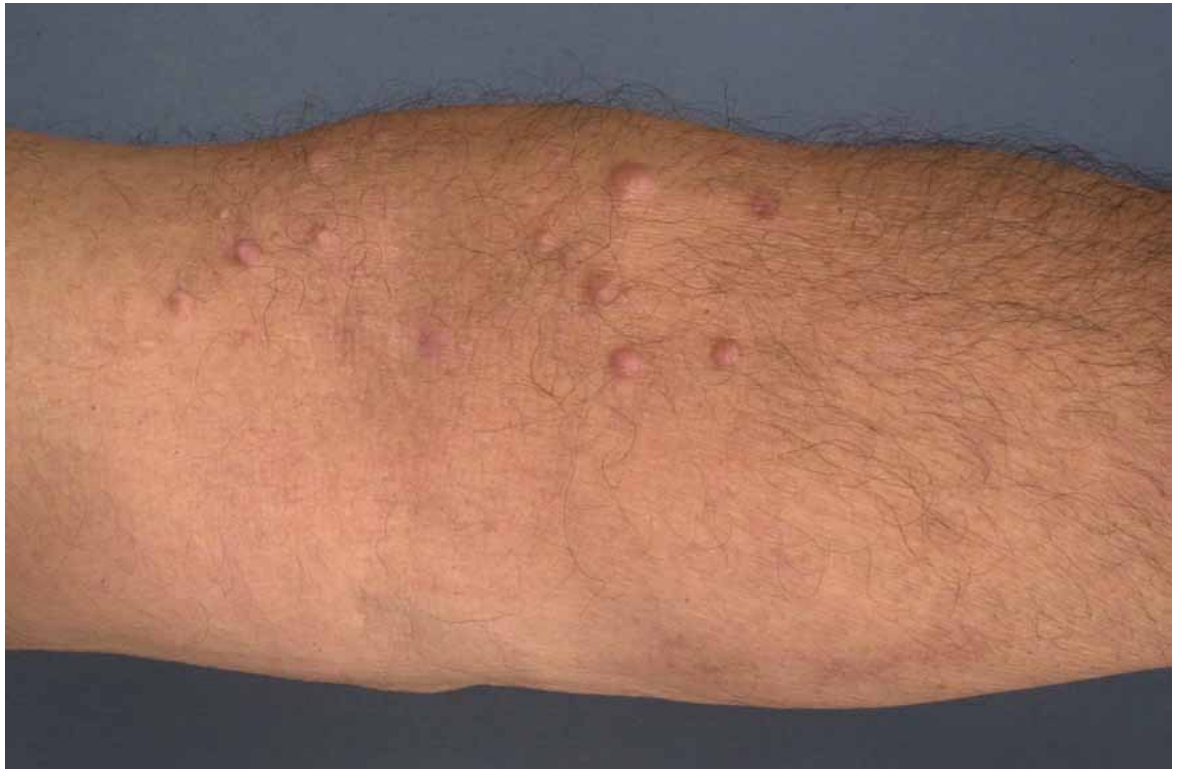
- 30-60 new point mut in newborn
- Of which 1 or 2 is in a coding sequence

Mosaic



- De novo mutation in one postmitotic cell during development
- Heterozygous
- Phenotype if mutation produces dominant effect in mutated cells

Somatic mosaic: segmental NF1

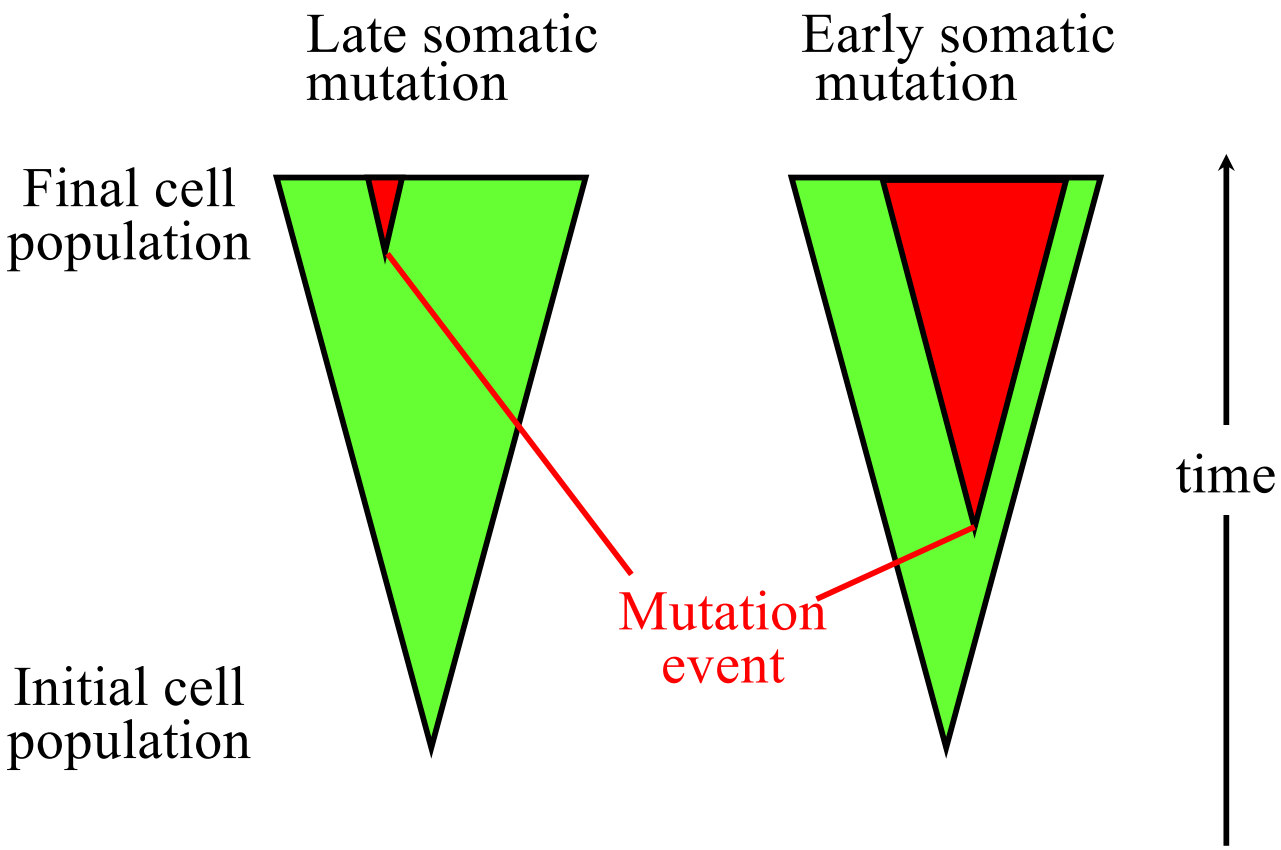


FC Victor Dermatology Online Journal 11 (4): 20

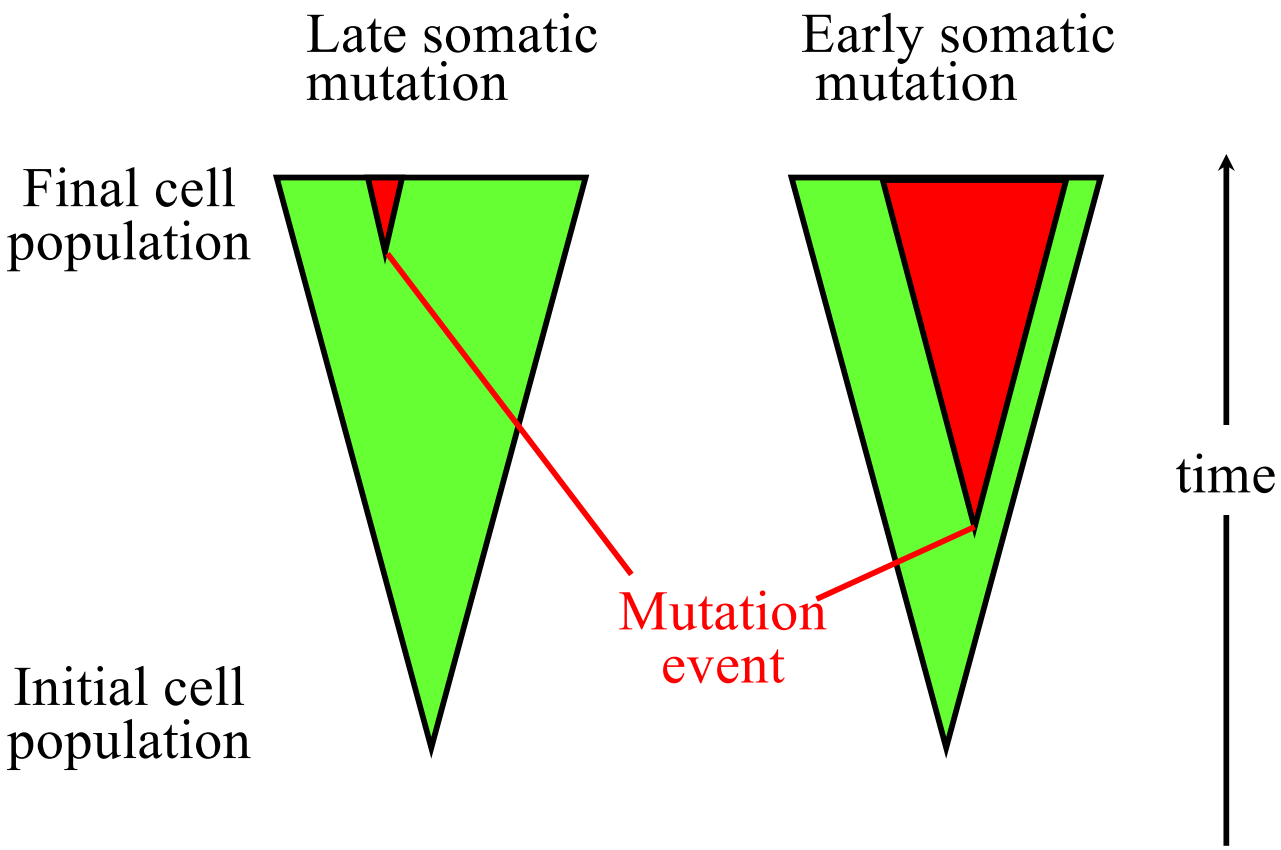
NF1 gene mutation, in a population of patient’s cells

Sporadic

Somatic mutation in embryo => Mosaic



Somatic mutation in embryo => Mosaic

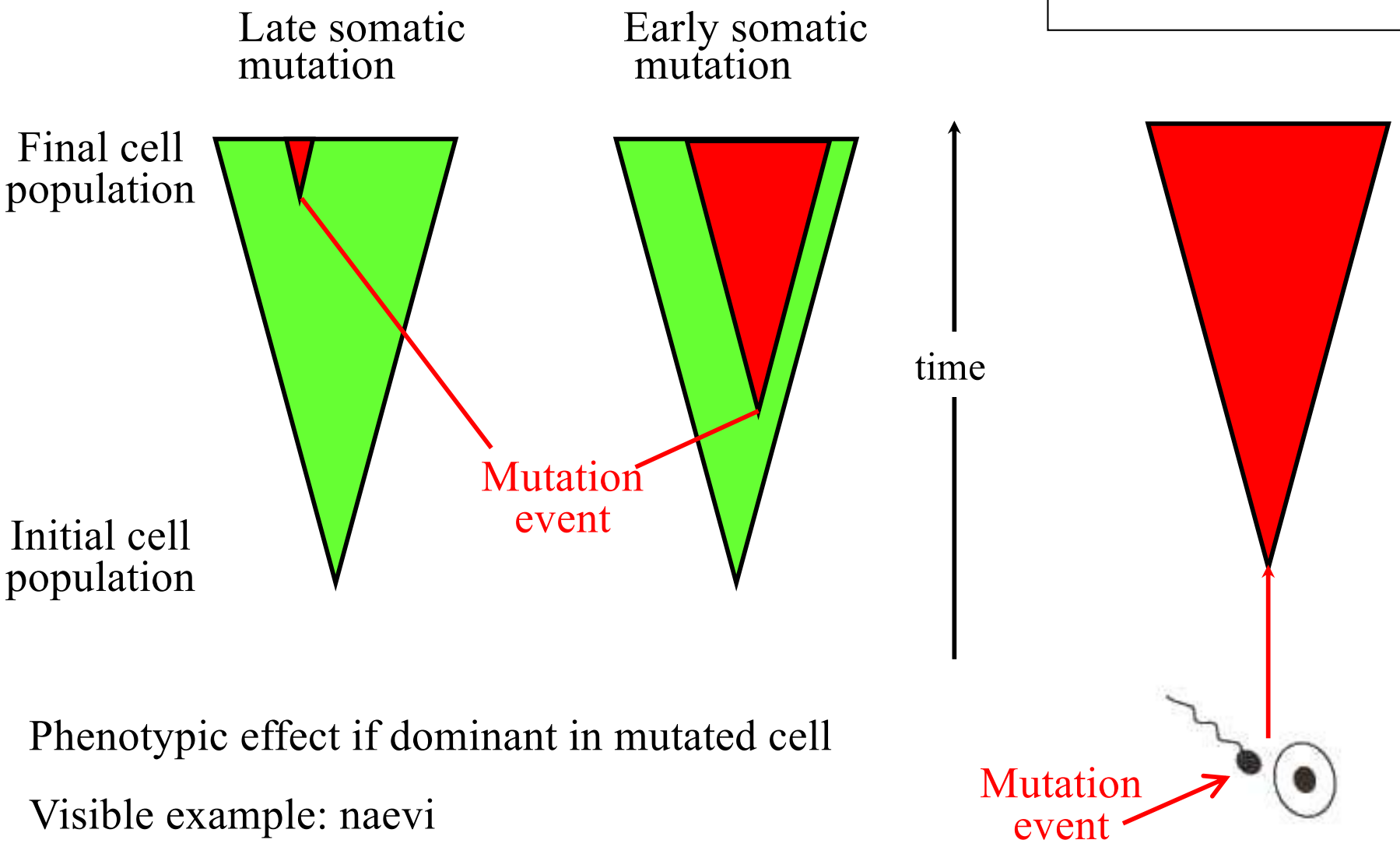


Phenotypic effect if dominant in mutated cell

Visible example: naevi

Somatic mutation in embryo => Mosaic

mutation in germline

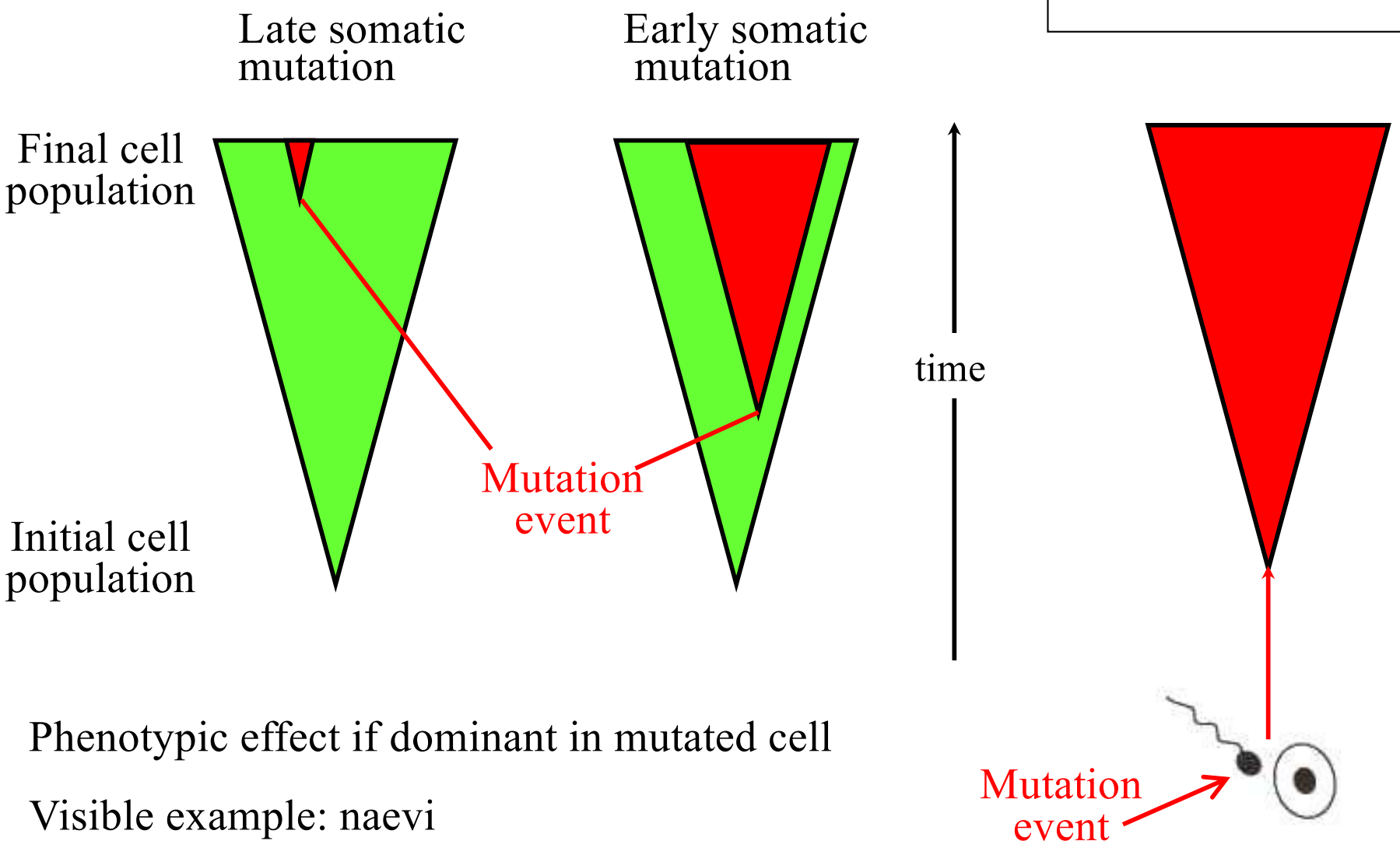


Phenotypic effect if dominant in mutated cell

Visible example: naevi

Somatic mutation in embryo => Mosaic

mutation in germline

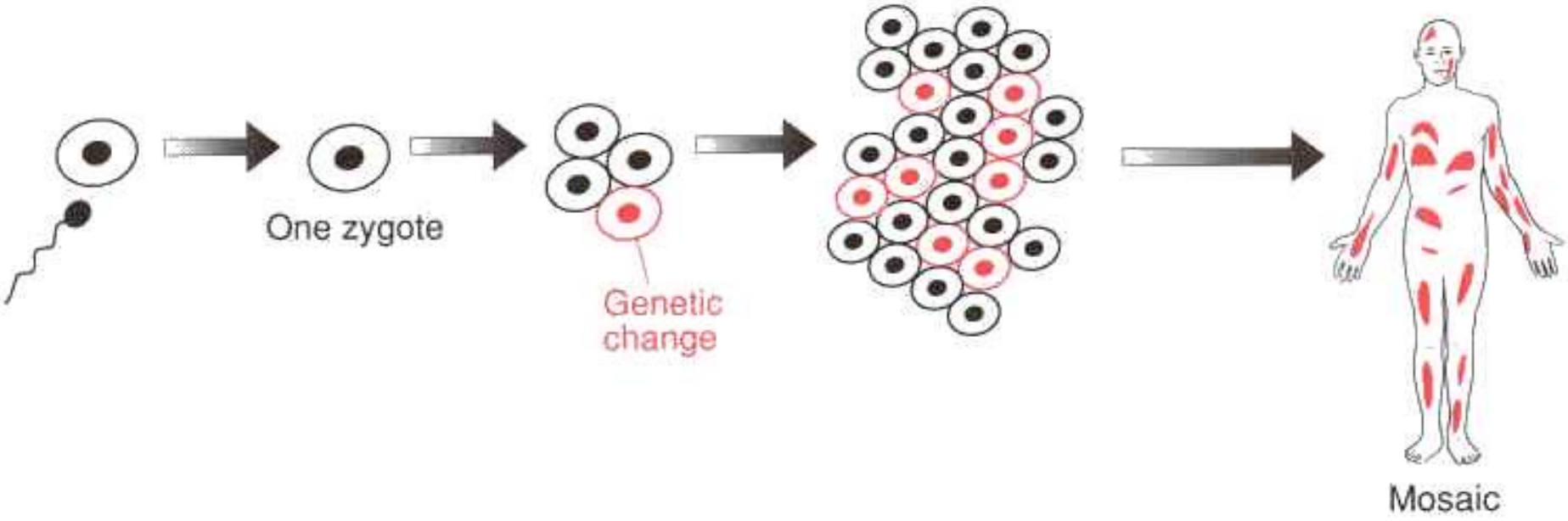


Phenotypic effect if dominant in mutated cell

Visible example: naevi

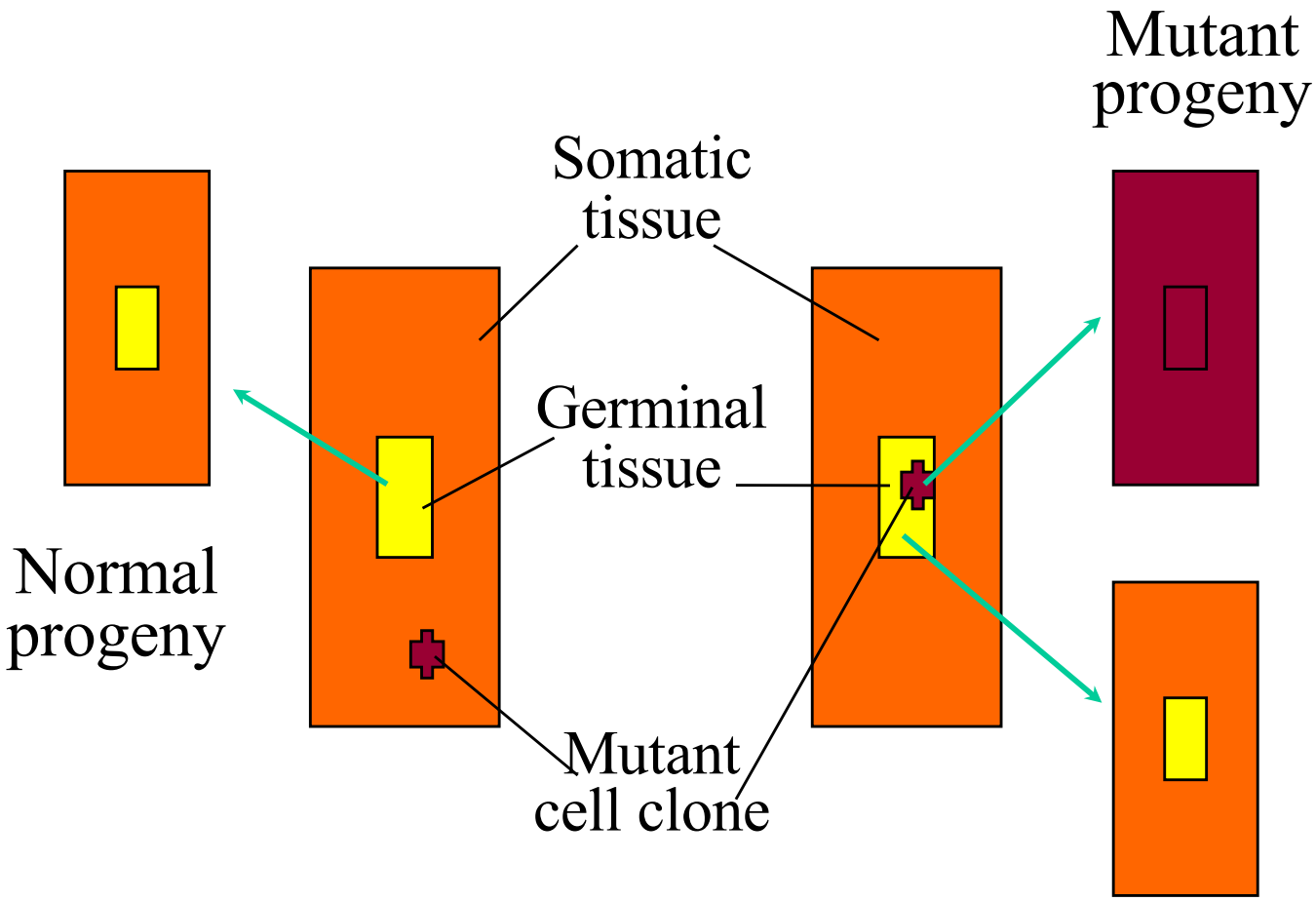
Proportion of mutated allele in blood: 0 – 50%

Mosaic



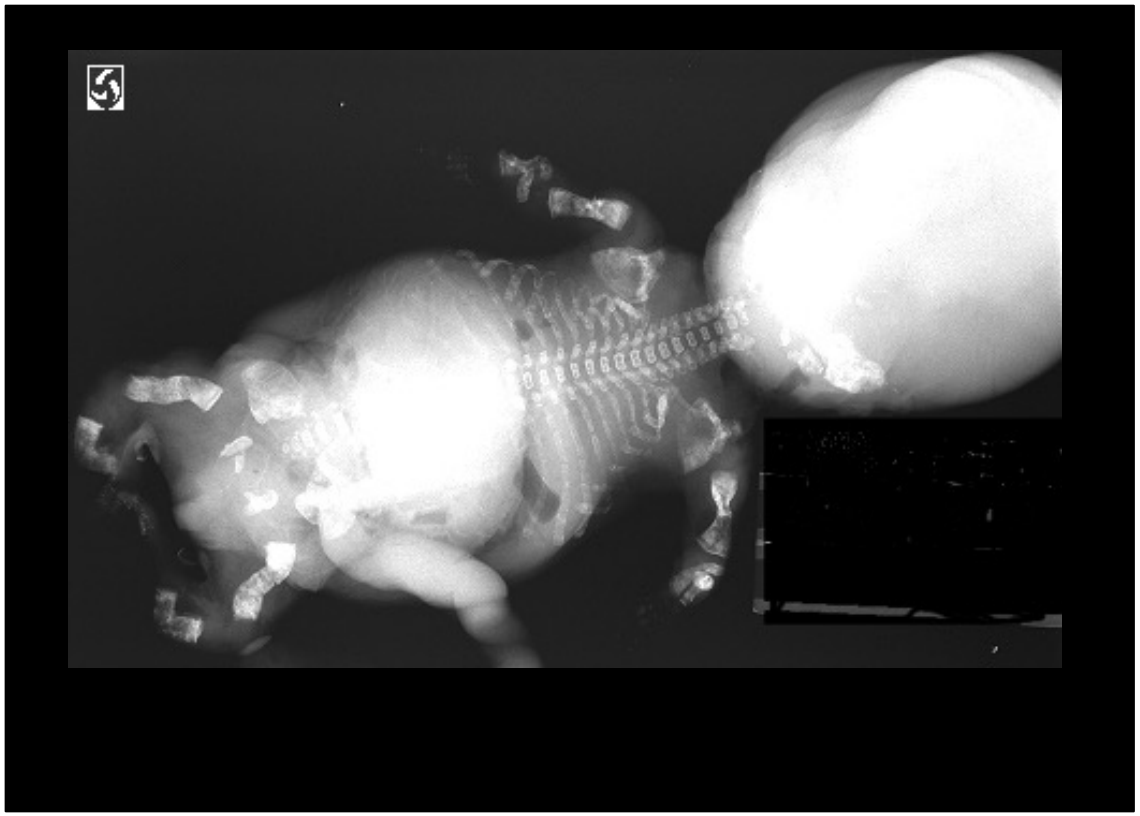
- De novo mutation in one postmitotic cell during development
- Heterozygous
- Phenotype if mutation produces dominant effect in mutated cells
- Proportion of mutation in individual = 0 – 50%

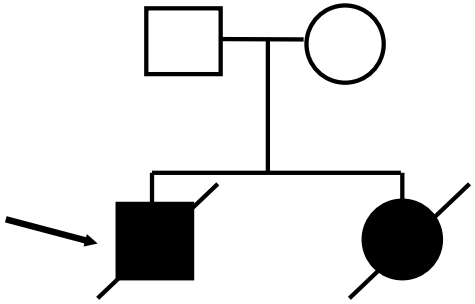
Somatic and/or **germ-line mosaic**



Germ-line mosaic in lethal OI

- Procollagen gene, hts mutation, lethal phenotype (fitness = 0)
- New mutations only
- 2 affected children here because germ-line mosaic in father (no mutation in bone cells)





OI

Germ-line mosaics

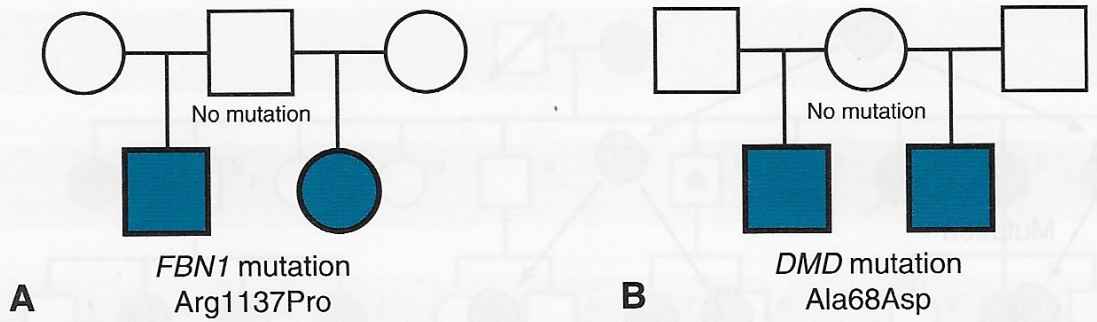


Figure 7-18 **Pedigrees demonstrating two affected siblings with the autosomal dominant disorder Marfan syndrome (Family A) and the X-linked condition Becker muscular dystrophy (Family B).** In Family A, the affected children have the same point mutation inherited from their father, who is unaffected and does not carry the mutation in DNA from examined somatic tissues. He must have been a mosaic for the *FBN1* mutation in his germline. In Family B, the affected children have the same point mutation inherited from their mother who is unaffected and does not carry the mutation in DNA from examined somatic tissues. She must have been a mosaic for the *DMD* mutation in her germline.

Parent-of-origin effect

See part about Genomic Imprinting

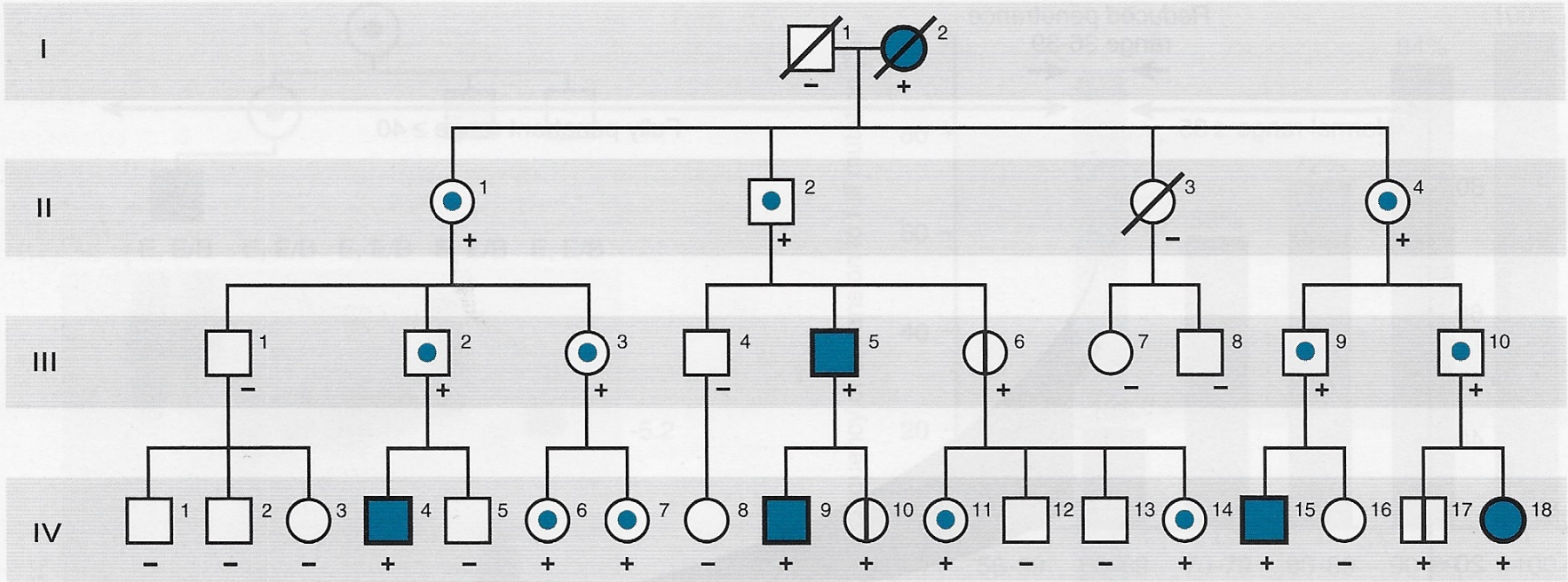


Figure 7-19 **Pedigree of a family with paraganglioma syndrome 1 caused by a mutation in the *SDHD* gene.** Individuals II-1, II-2, II-4, III-2, III-3, III-9, III-10, IV-6, IV-7, IV-11, and IV-14 each inherited the mutation from their mothers but are unaffected. However, when the males in this group pass on the mutation, those children can be affected. In addition to the imprinting, the family also demonstrates the effect of reduced and age-dependent penetrance in the children (III-6, IV-10, IV-17) of heterozygous fathers. The + and – symbols refer to the presence or absence of the *SDHD* mutation in this family.

Gene interactions

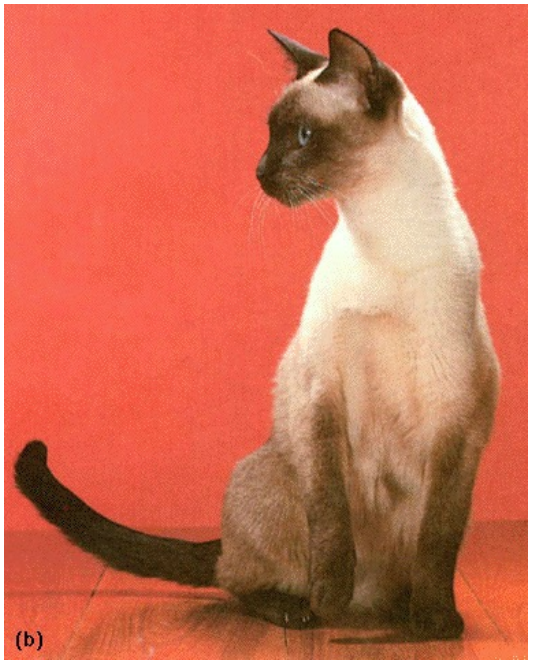
Ordinary cat

Siamese cat

(Himalayan mouse, Himalayan rabbit)

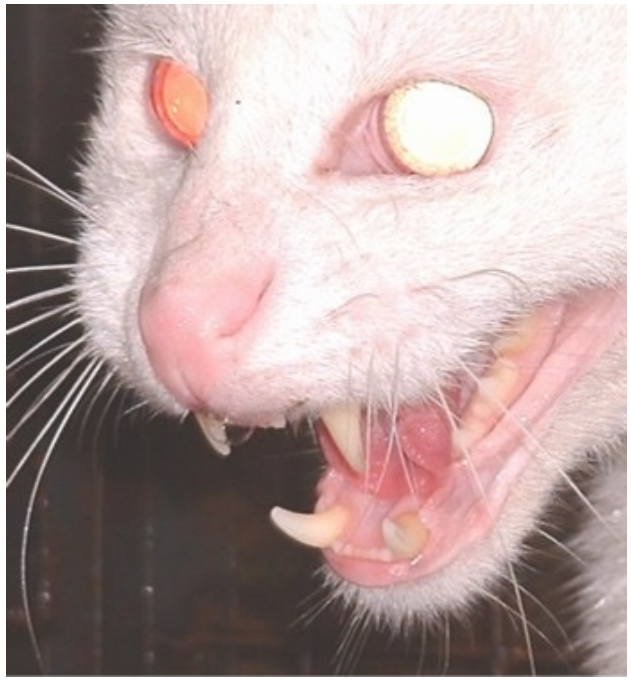
t° -sensitive allele in C gene of colour deposition. Recessive.

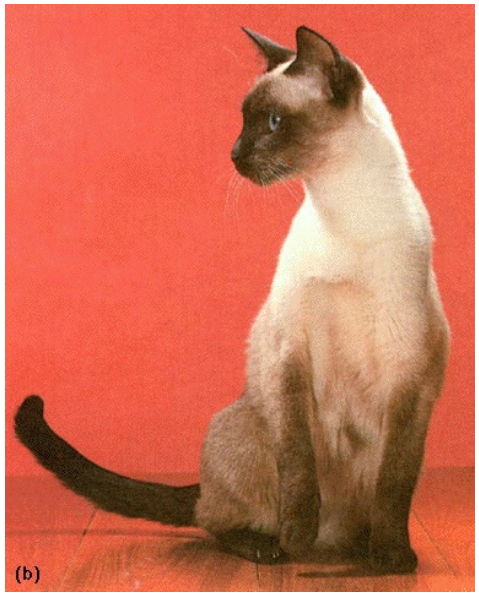
C^h/C^h prevents colour deposition in warmer parts of the body



Albino cat

No pigment produced
(tyrosinase -/-)





EPISTASIS

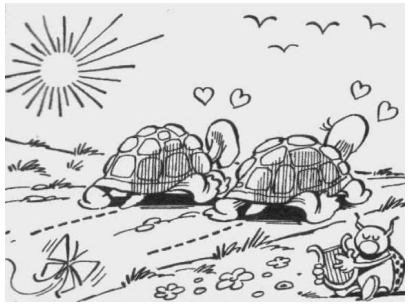
epistatic gene (Albino)
masks the effect of
another gene (Siamese)

Siamese albino

Non-Siamese albino

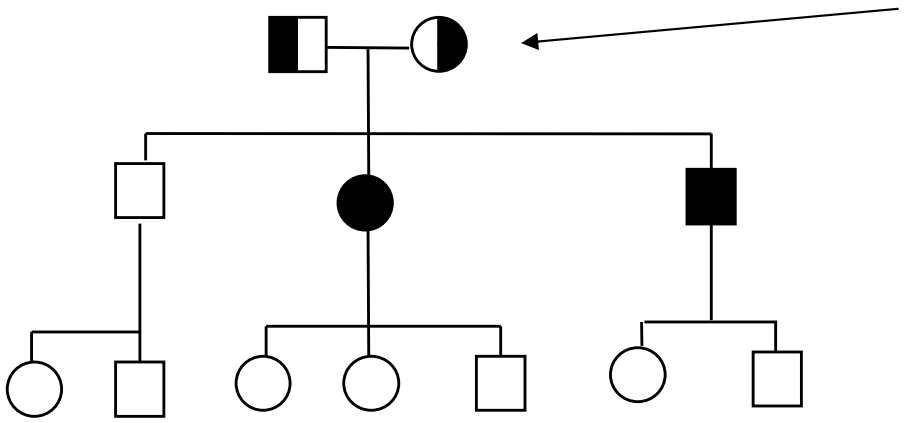
Patterns of single-gene inheritance

AUTOSOMAL RECESSIVE



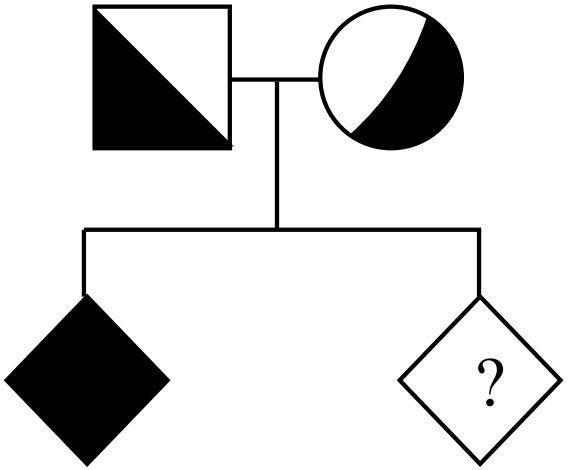
AR phenotype: horizontal transmission

Genotype: bi-allelic mutations, one gene, autosome

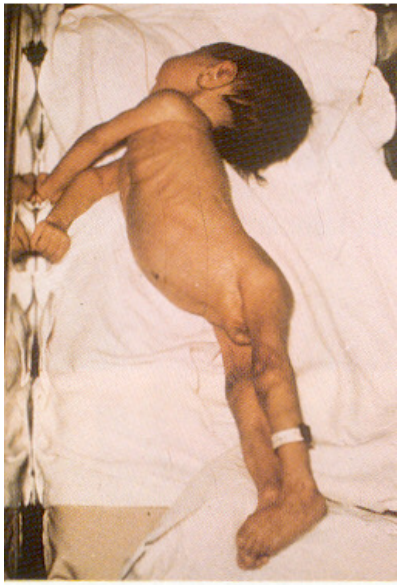


- Heterozygote, healthy carrier
- We all are healthy carriers of hts mutation which, when hms, cause severe disease
  - ~1 mutation compatible with post-natal life
  - 2-3 mutations causing miscarriage / non-implantation (?)

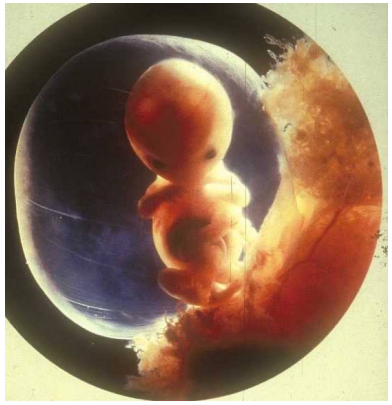
Krabbe disease (AR)



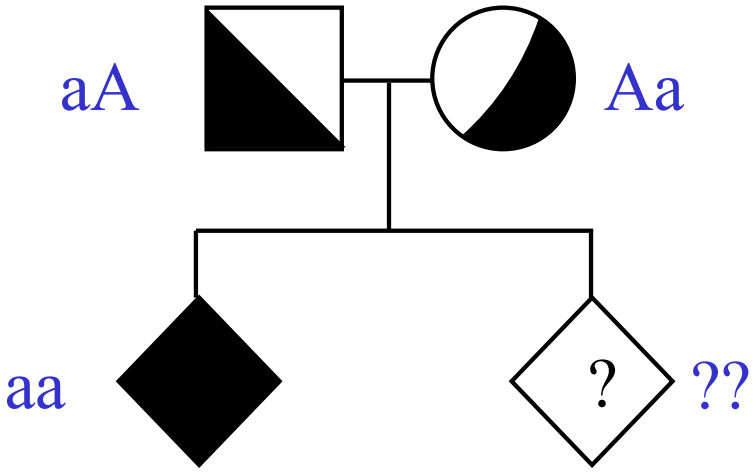
Affected child



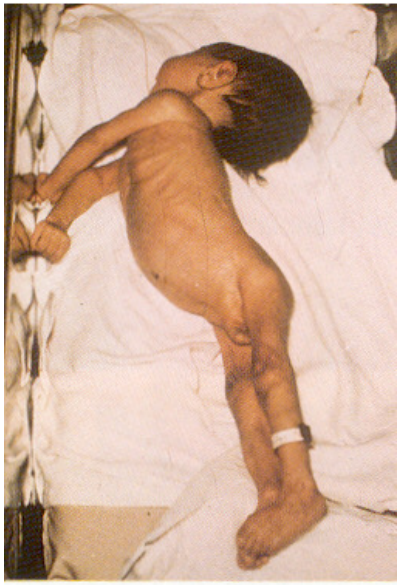
New pregnancy
risk = 25 %



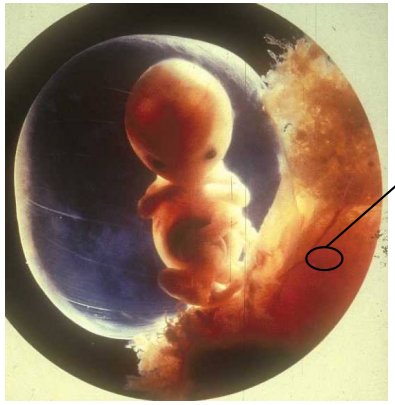
Krabbe disease (AR)



Affected child



New pregnancy
risk = 25 %



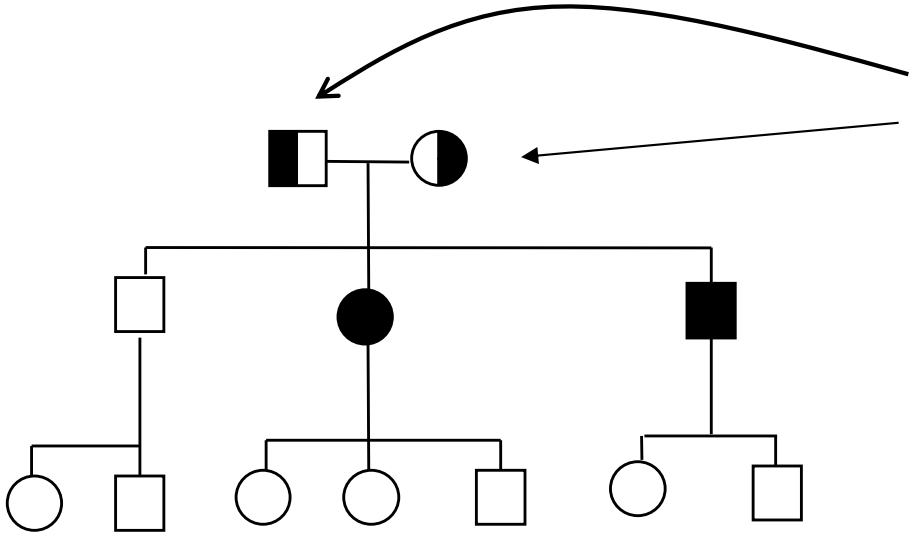
→ Prenatal diagnosis
(CVS, 10 wks.)

PUNNET'S SQUARE

Female gametes \ Male gametes	A	a
A	AA	Aa
a	aA	aa

AR phenotype: **bi-allelic mutations**,

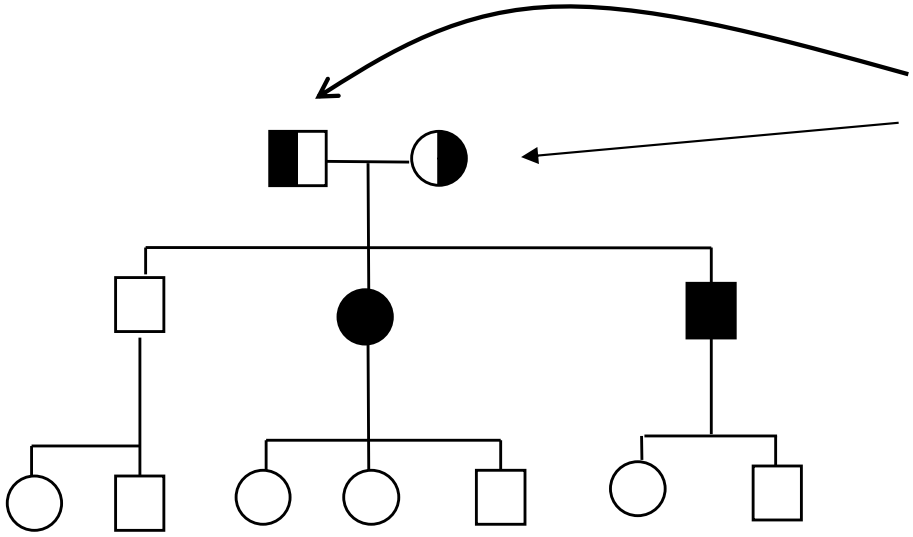
loss of function



- Same parental mutation
  => hmz
- Different parental mutations
  => compound htz

AR phenotype: **bi-allelic mutations**,

loss of function

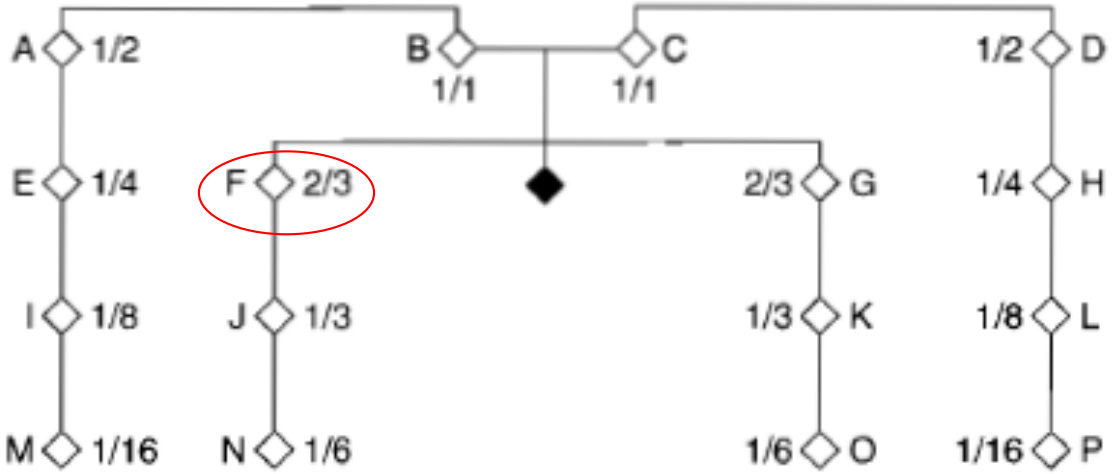


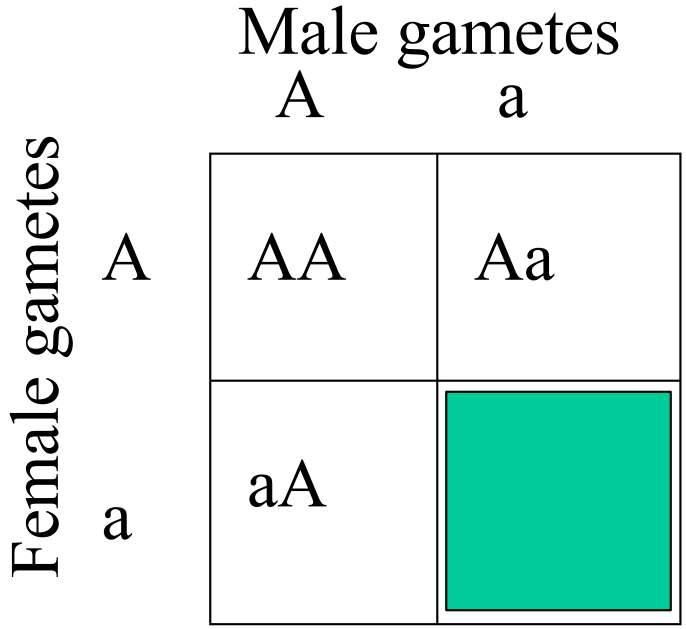
- Same parental mutation
  => hmz
- Different parental mutations
  => compound htz

Loss of function: variable

- Complete (severe mutation)
- Partial (mild mutation)
- Minimal (minor mutation, polymorphism)

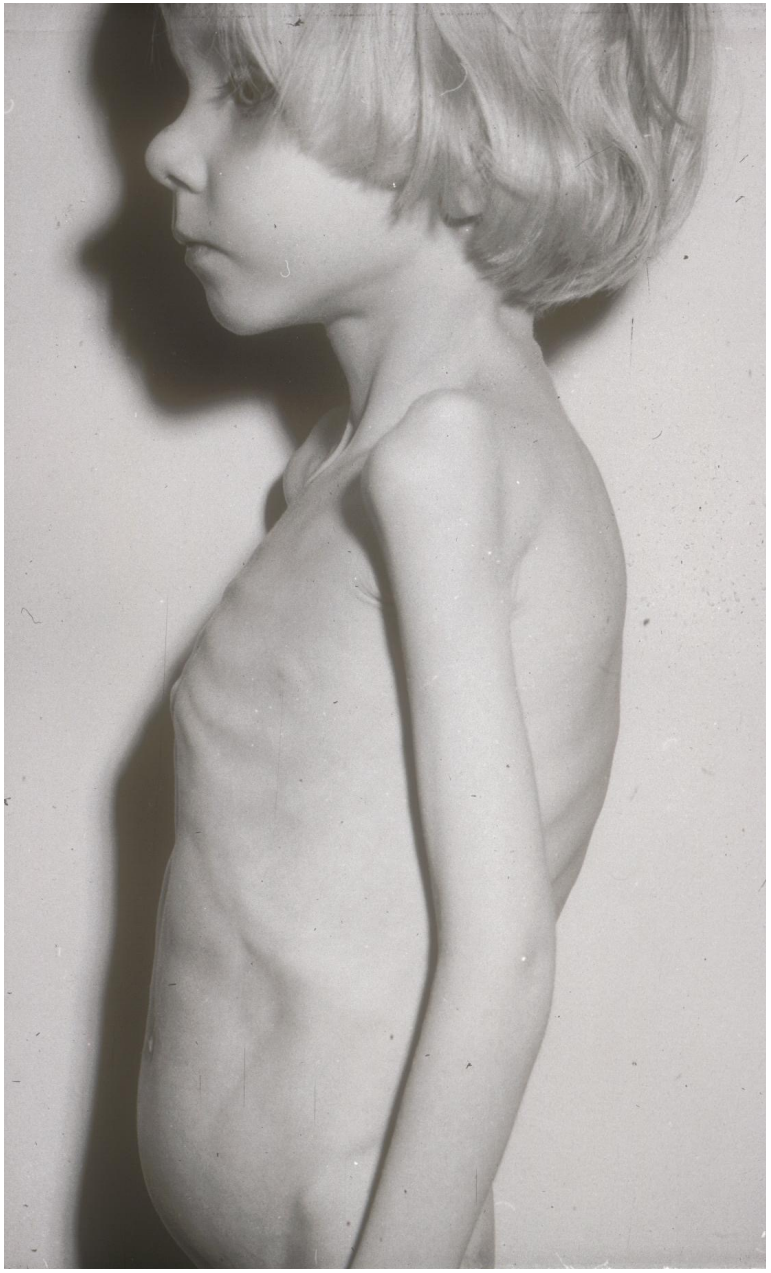
Probability of being a carrier, AR disease





PS Harper, Practical Genetic Counselling, 6th ed

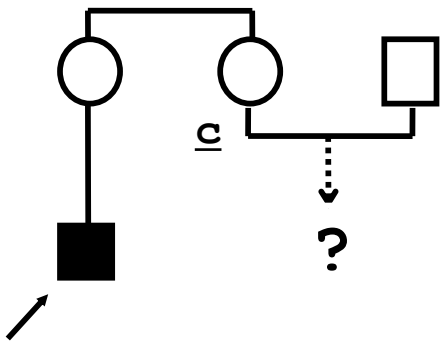
Cystic Fibrosis, CF



- Syndrome :
  - COPD
  - Pancreatic insufficiency
  - Na et Cl elevated in sweat
- Vas deferens agenesis (CBAVD)
- No MR

- 1/2500 children affected at birth
  1/25 healthy carriers

Genetic counselling in CF

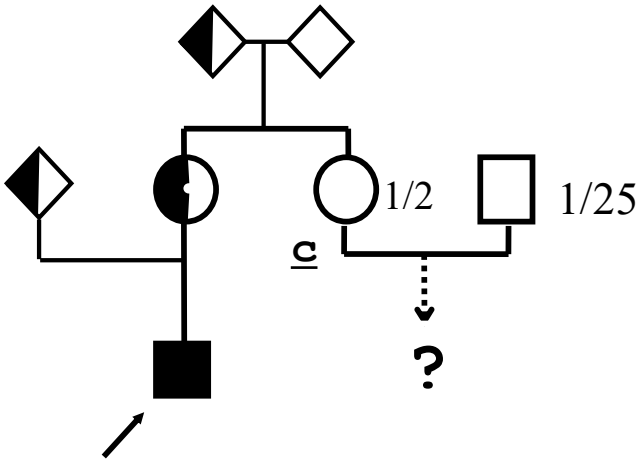


Cystic Fibrosis (Northern EU) 1/2500

- Carriers in general population (1/25)
- Risk = 1/2 x 1/25 x 1/4 = 1/200

.

Genetic counselling in CF

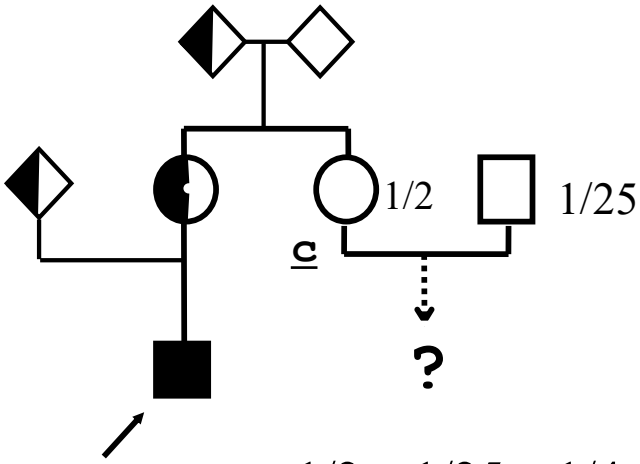


Cystic Fibrosis (Northern EU) 1/2500

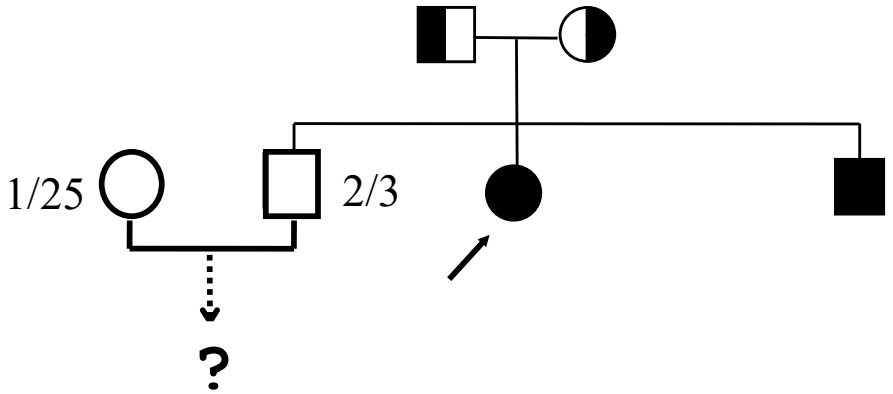
- Carriers in general population (1/25)
- Risk = 1/2 x 1/25 x 1/4 = 1/200

.

Genetic counselling in CF



1/2 x 1/25 x 1/4
= 1/200 = 0.05



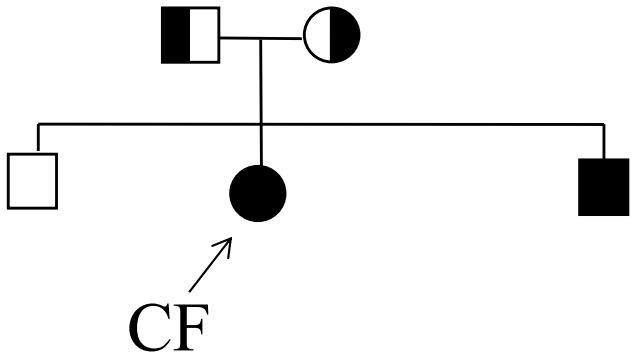
2/3 x 1/25 x 1/4
= 2/300 = 0.067

Cystic Fibrosis (Northern EU) 1/2500 affected at birth

– Carriers in general population 1/25

Genotype – phenotype correlation

- Pancreatic sufficiency is concordant in sibs
  - Depends on mutation

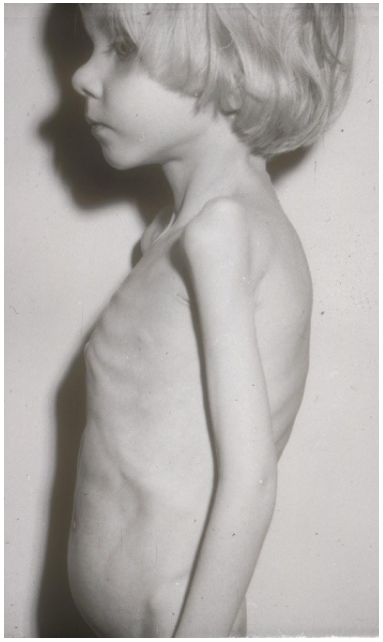


- Lung disease severity is less concordant
  - Depends on mutation and on environment (Pseudomonas infection, …)

Variable expressivity, AR disease (CF)

DF508 W1282X

DF508 / W1282X

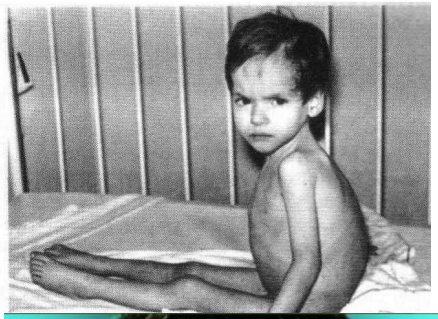


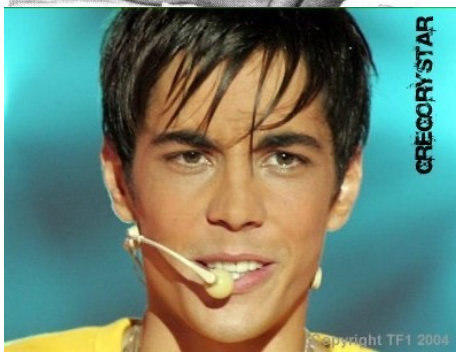
No sign in hts parents

Homoz severe mutation

DF508 DF508

DF508/DF508





Modifyer genes (TGFb1,…)

Severe + mild mutation

DF508 A445E

DF508/A445E

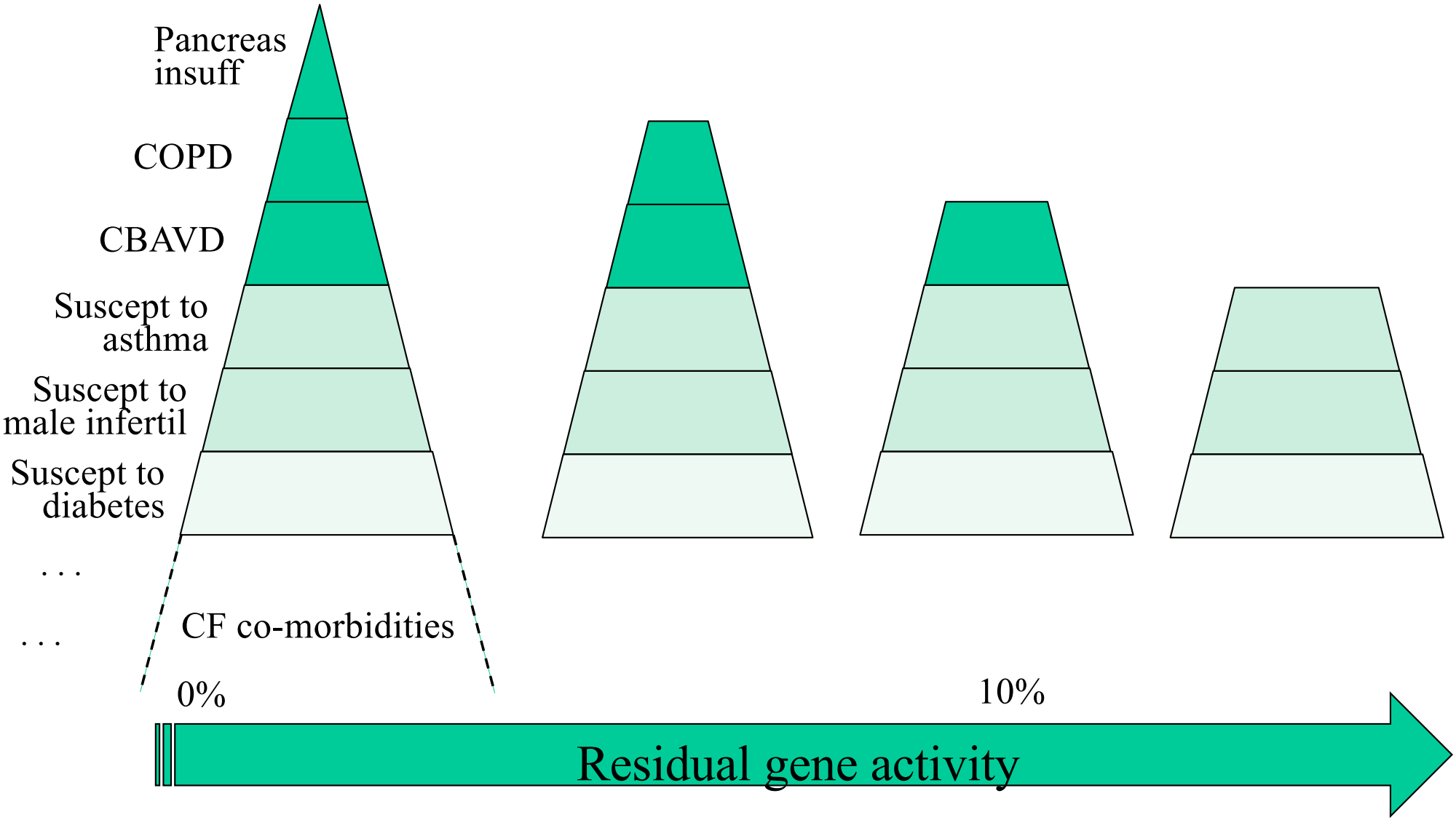
Pancreas sufficient
Lung variable

Severe + minor mutation

DF508 hypomorphic allele

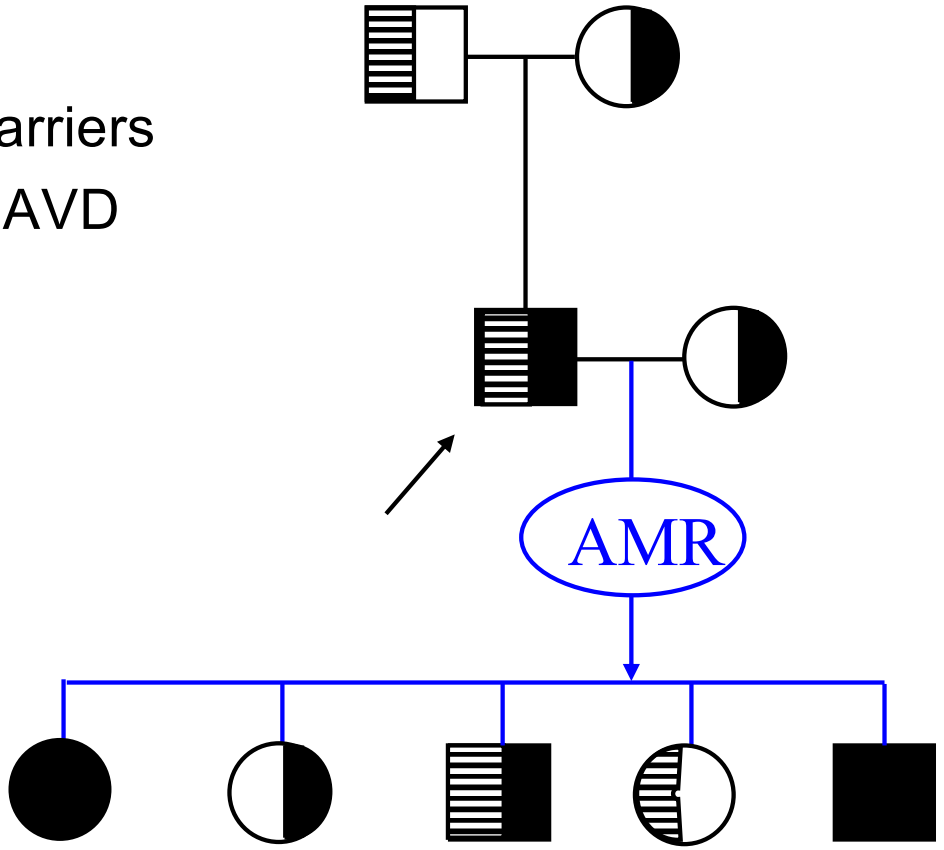
Healthy… or almost (CBAVD)

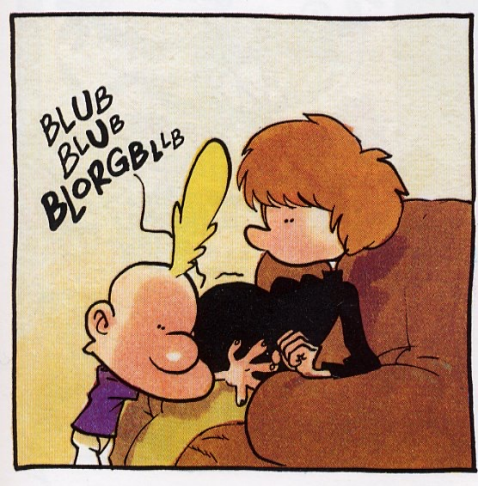
Increased gene activity removes phenotypic features



CBAVD, AMR and CF

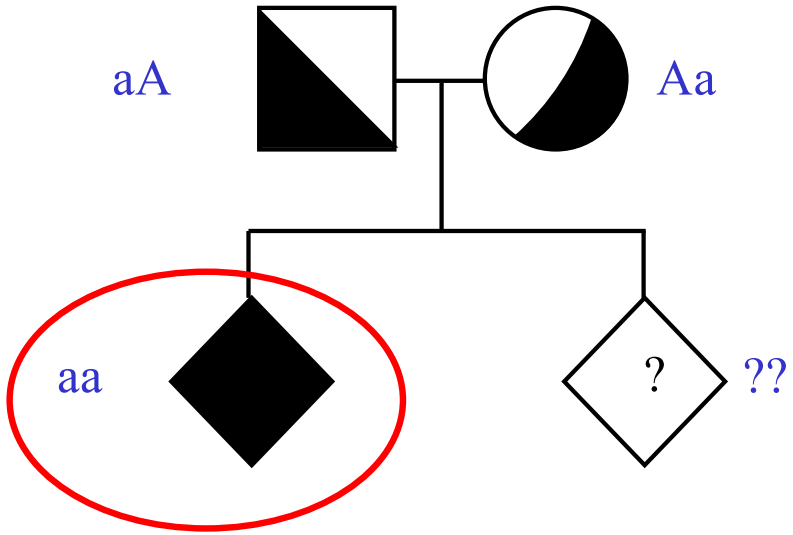
- Assisted Medical Reproduction
  for men with CBAVD
- 4% of partner women are CF carriers
- If carrier, offspring at risk of CBAVD
  and of CF



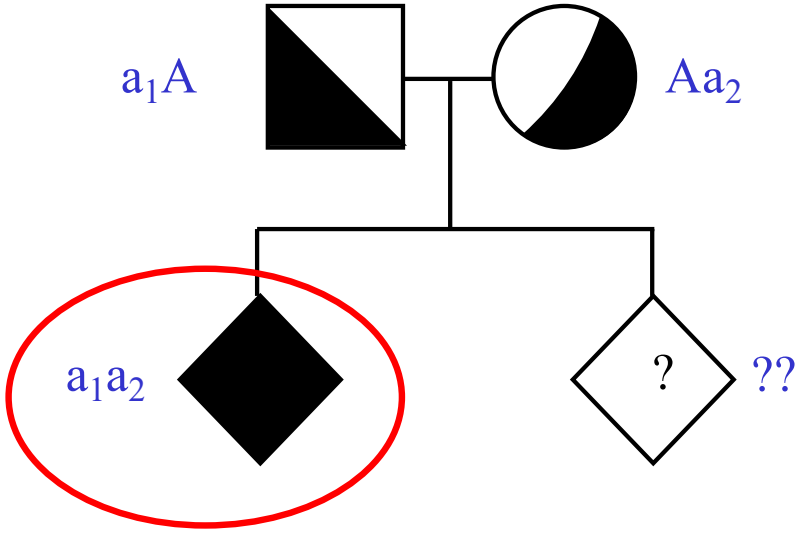


bi-allelic mutations

- homozygote



- compound heterozygote



Heterogeneity of mutations

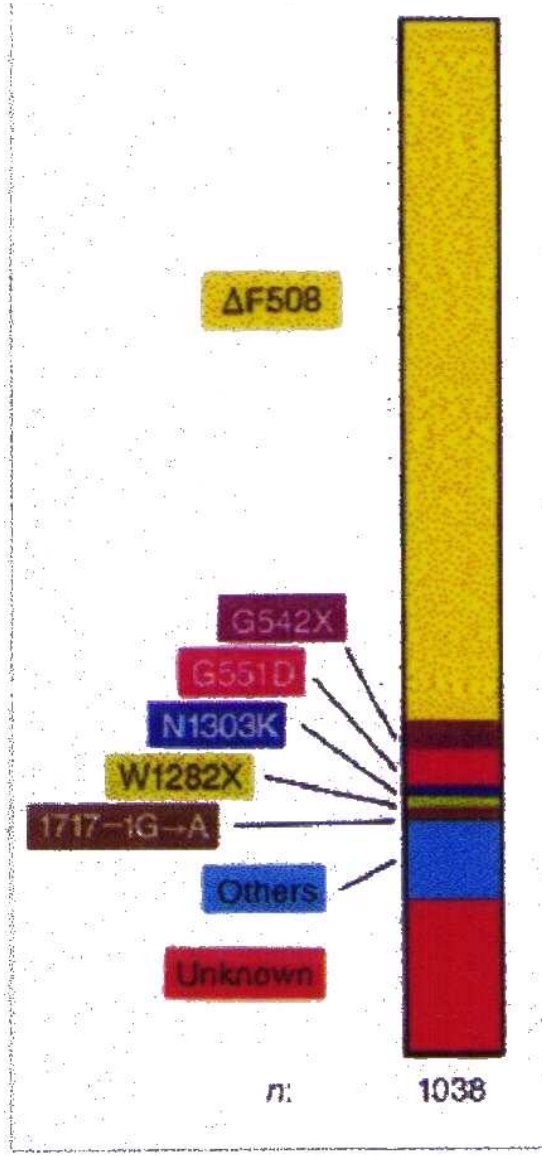
- Practical nomenclature A / a of alleles conceals a great diversity of DNA sequences : $a = a_1 + a_2 + a_3 + \ldots + a_n$

- Ex: CF (CFTR gene): mutations with complete loss of function
  - Many missense mutations G551D, N1303K
  - Many nonsense mutations W1282X, G542X
  - Many indels DF508
  - Many splicing mutations 1717-1 G>A
  - Many chromosomal mutations

All severe mutations = allele « a » in A,a nomenclature

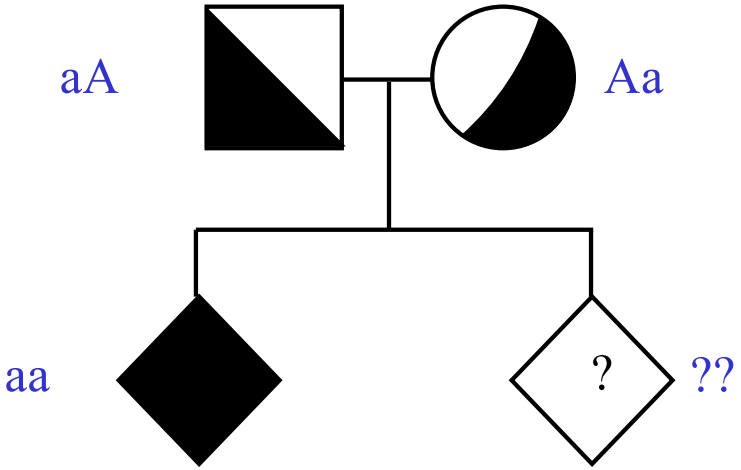
All are silent in htz, and cause CF when hmz or compound htz

Allelic heterogeneity

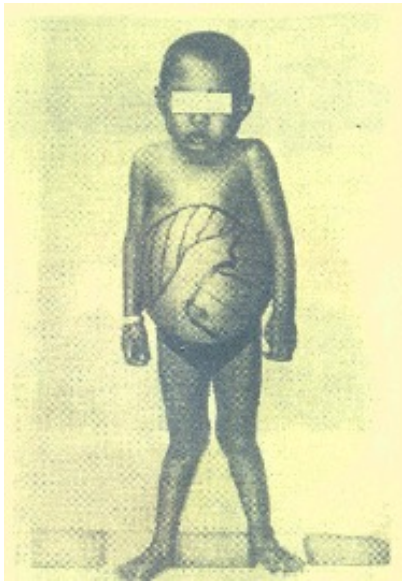


- CF: one gene, one locus
- Many mutations
- Ethnic prevalences

b-thalassemia (AR)



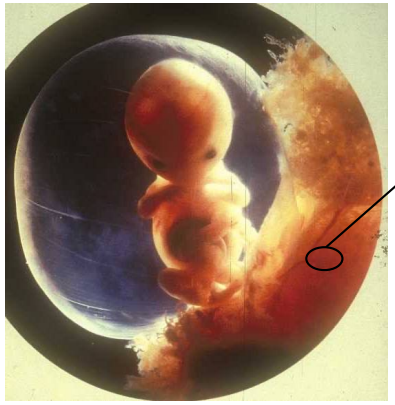
Affected child



PUNNET'S SQUARE

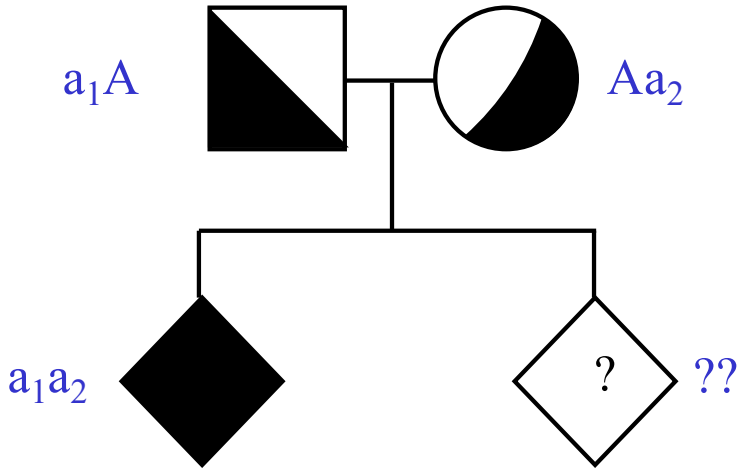
Female gametes \ Male gametes	A	a
A	AA	Aa
a	aA	aa

New pregnancy
risk = 25 %

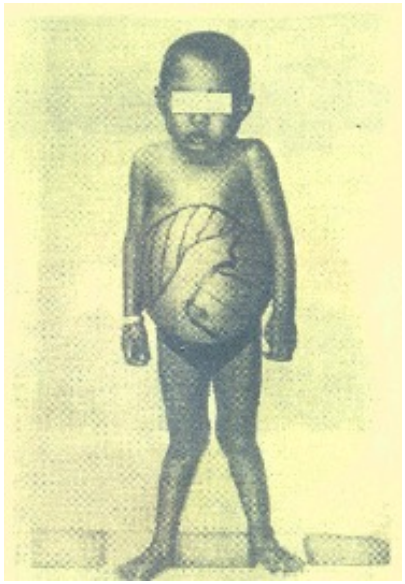


→ Prenatal diagnosis
(CVS, 10 wks.)

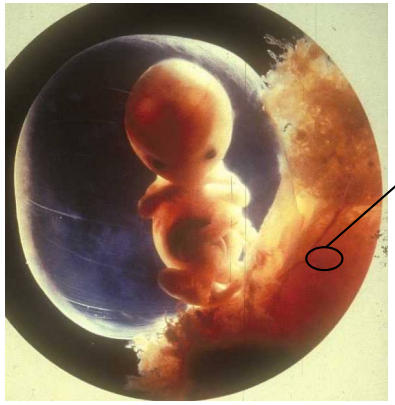
b-thalassemia (AR)



Affected child



New pregnancy
risk = 25 %



→ Prenatal diagnosis
(CVS, 10 wks.)

PUNNET'S SQUARE

Female gametes \ Male gametes	A	a_1
A	AA	Aa_1
a_2	a_2A	a_1a_2

AR transmission = approximation

- CF is truly recessive (?)
- HbS carriers have some signs if hypoxia
- Beta-thalassemia carriers have microcytosis (sometimes low Hb)
- Some carriers of Wilson disease have low Ceruloplasmin
  - Because their 2nd allele is hypomorphic?

Why are some AR mutations frequent ?

1. **FOUNDER EFFECT**
   - Founder event
   - Population bottleneck

2. **OVERDOMINANCE**
   htz has selective advantage

   ex: HbS and malaria

Both (1) and (2) tend to be population-specific

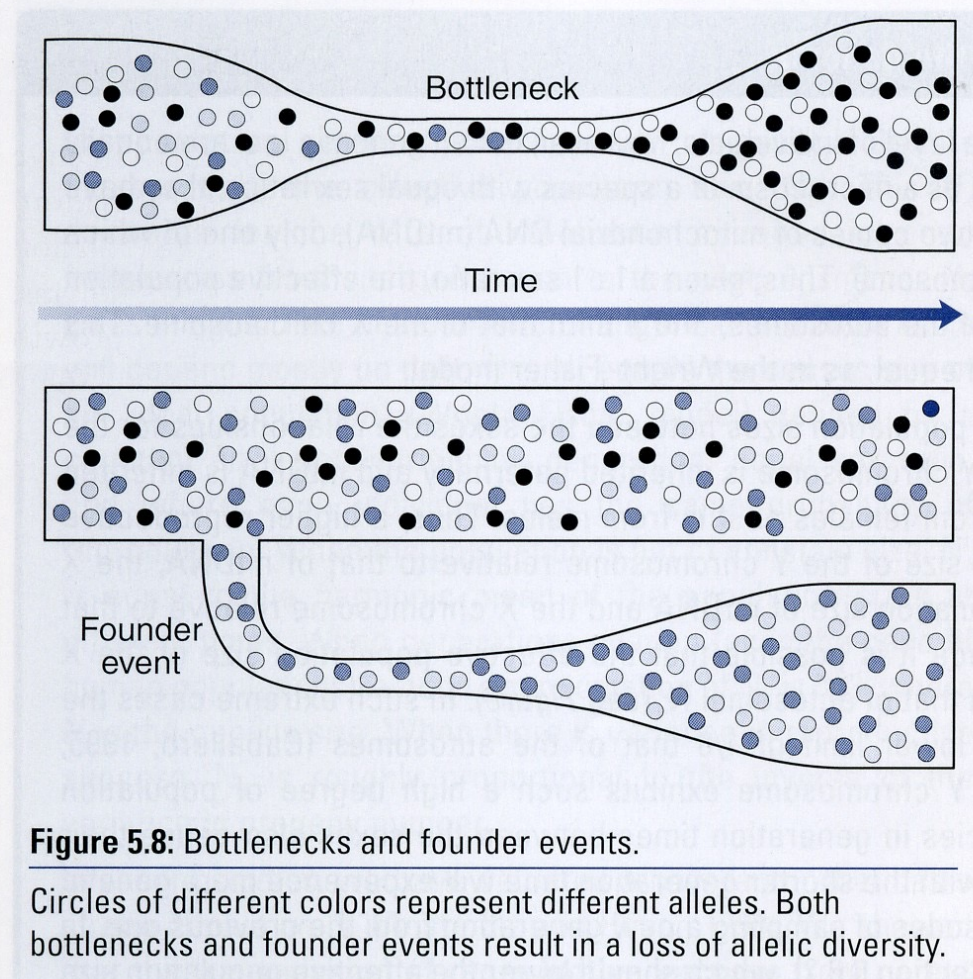


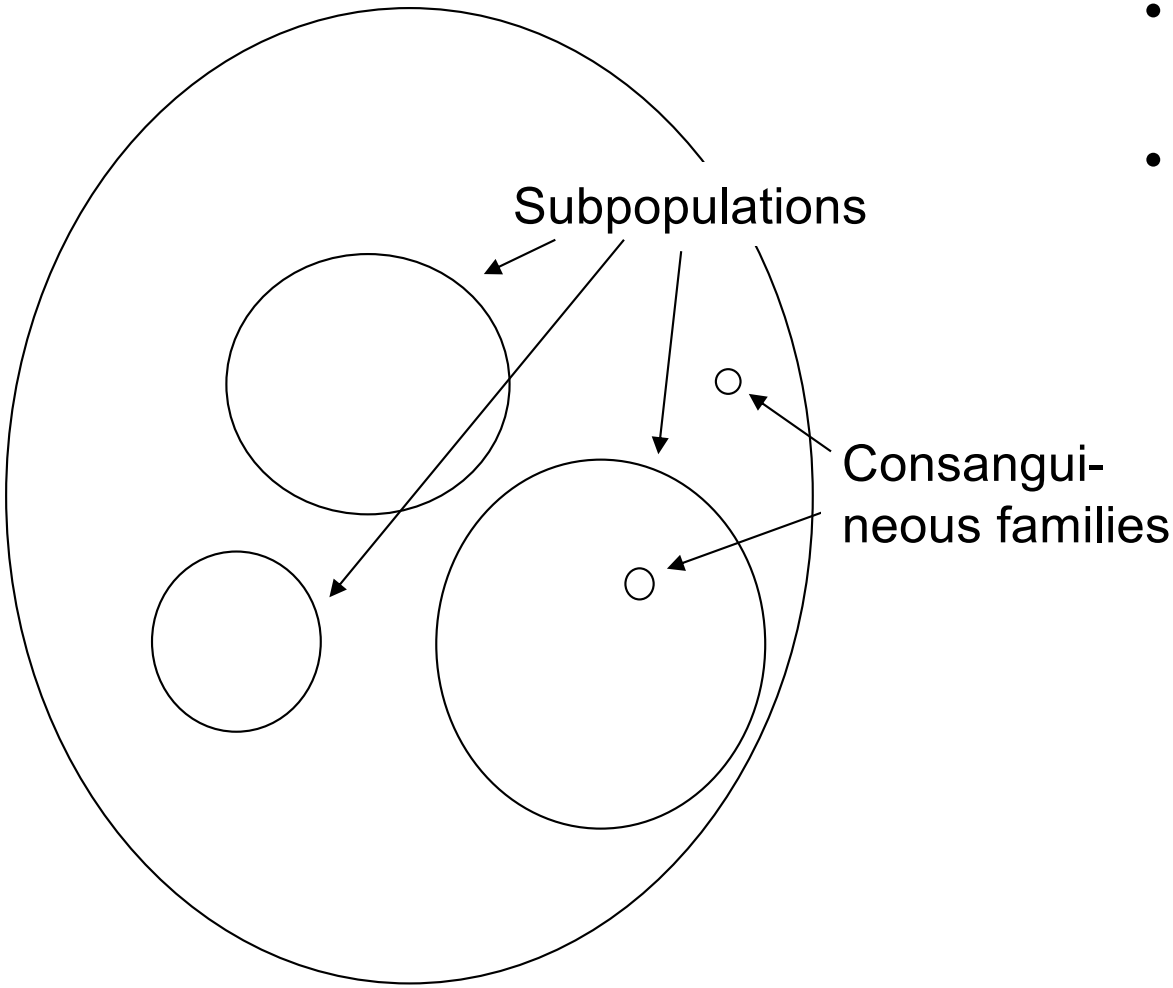
Figure 5.8: Bottlenecks and founder events.

Circles of different colors represent different alleles. Both bottlenecks and founder events result in a loss of allelic diversity.

AR phenotypes tend to be ethnic = population specific

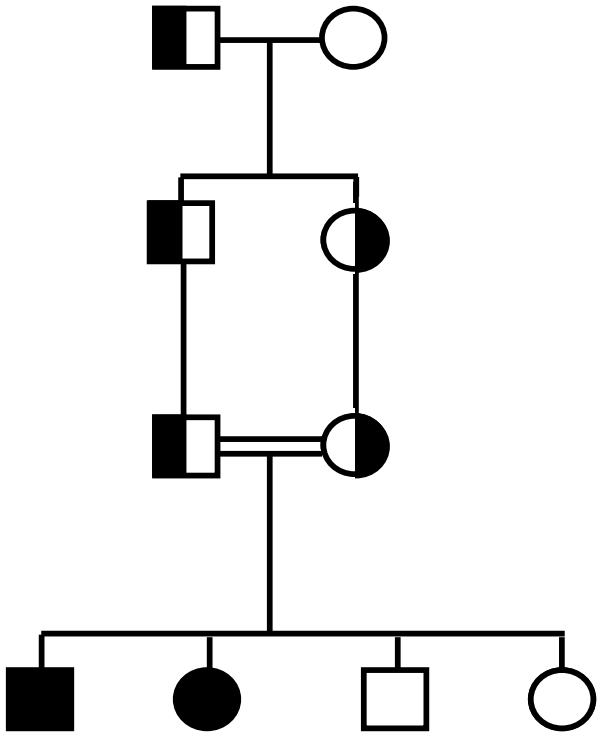
- Founder effects
- Population bottlenecks
- Overdominance < population-specific selection (HbS < malaria)

Subpopulations in Metapopulation



- Cross-fertile individuals (species)
- Subpopulations isolated by
  - Geography
  - Language
  - Religion
  - …
  - Inbreeding
  - Consanguinity

Consanguinity and AR disease



- One ancestral mutation
- Mutation is identical-by-descent (**IBD**) in first cousins
- Affected offspring
  = true homozygote
  = homozygote by descent
  = « **autozygote** »
- Rarer mutation <=> more cases due to consanguinity

Burden of consanguinity:
3% excess disease in offspring of 1st cousins

- We all carry one recessive disease [non embryonic-lethal].

- hence :

population risk of handicap :	3%
additional risk from consanguinity :	3%
total risk:	6%

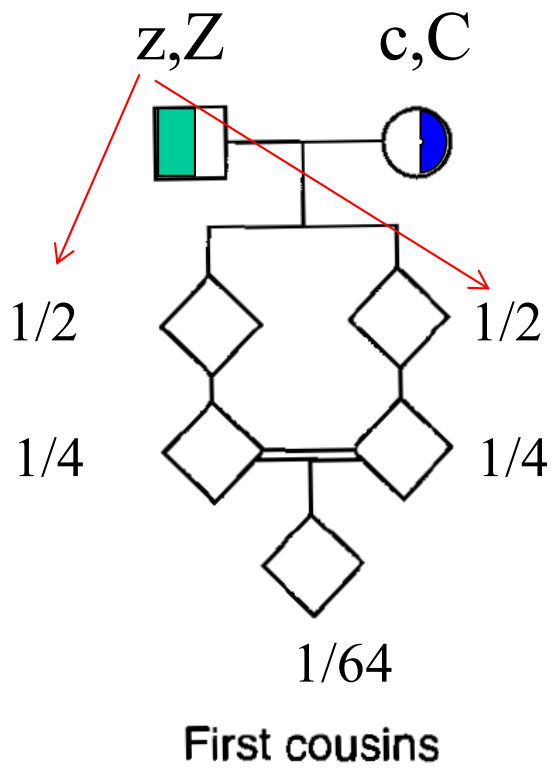
Homozygosity by descent in first cousins

- Assume everyone carries 1! severe recessive disease
  - Grandfather carries Zellweger disease
  - Grandmother carries CF

- Proba hmz (z,z) = 1/64

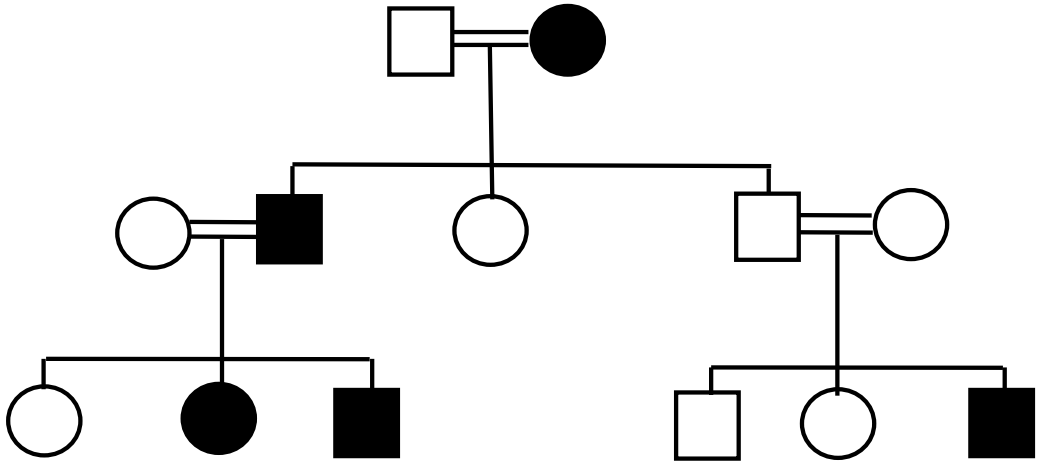
- Proba hmz (c,c) = 1/64

=> excess risk = 1/32
  - fits with observed 3%



First cousins

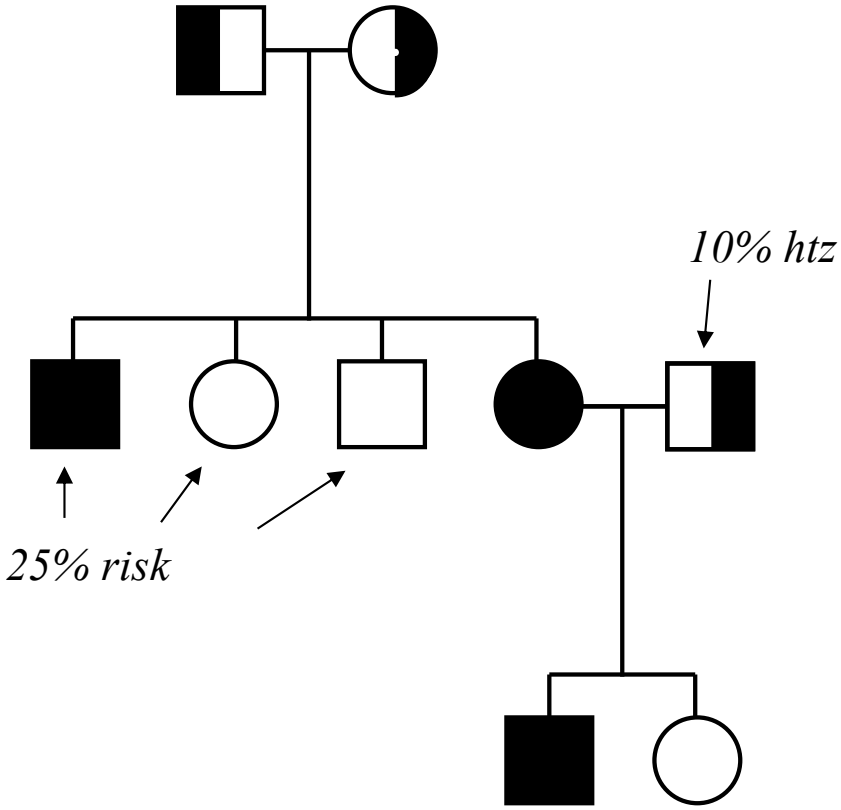
Pseudodominance in rare AR trait



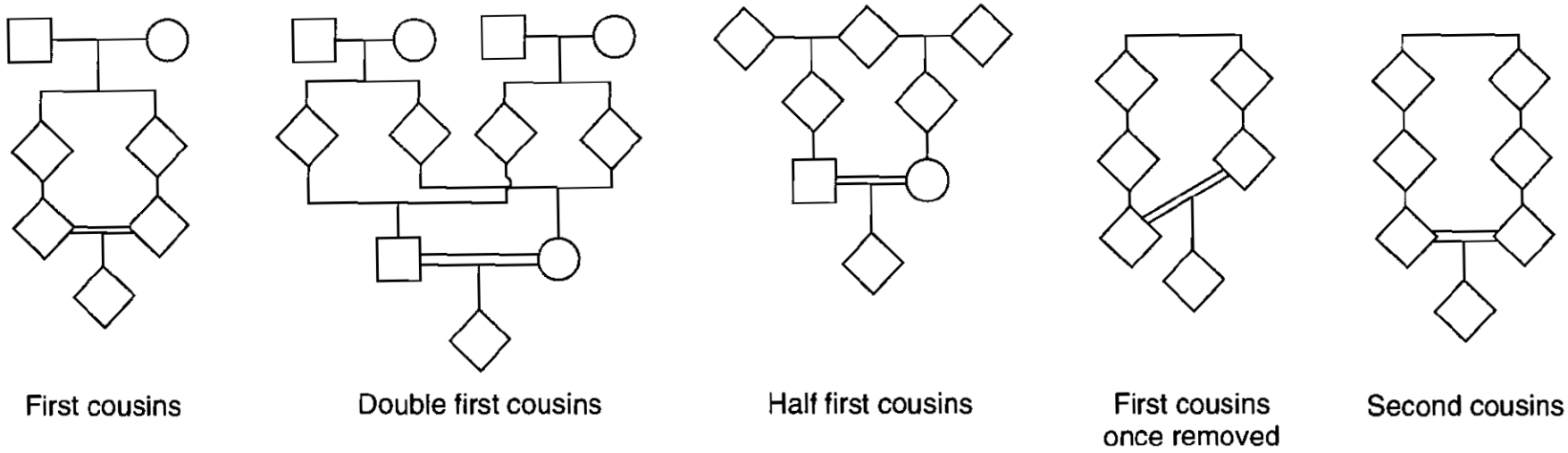
Multi-generation consanguinity => AR disease may seem AD.

Pseudodominance in frequent AR trait: hemochromatosis

- 10% (!) carriers in Western Europe
- Low penetrance of full-blown disease



Consanguinity
= having (relatively) close common ancestors

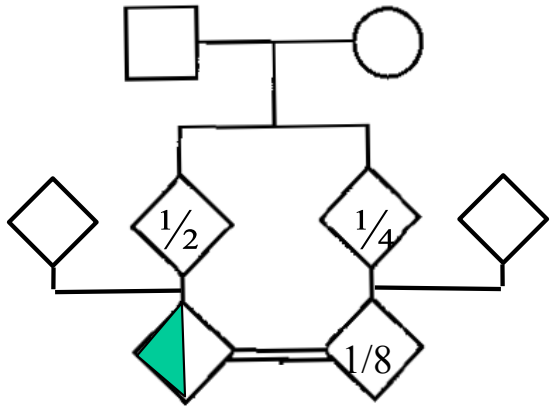


- ✓ « germans » = siblings = brothers and sisters
- ✓ First cousins = offspring of siblings (cousins issus de germains)

First cousins share 1/8 genome

FIRST COUSINS:

- Probability of finding green allele in consanguinity loop

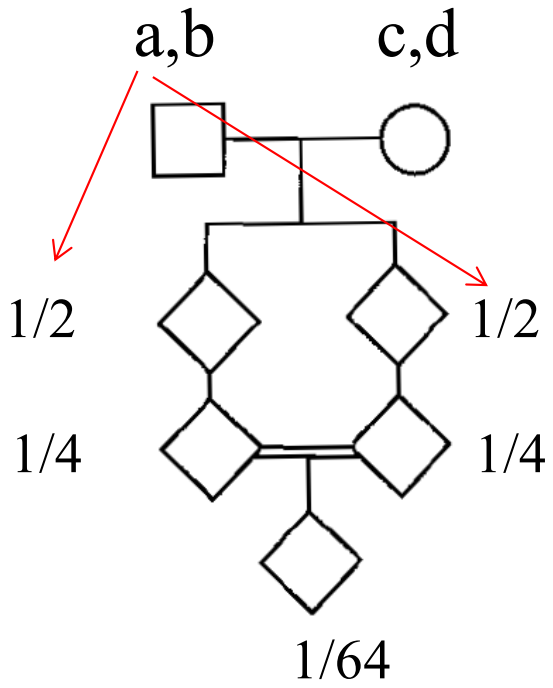


First cousins

F = coefficient of inbreeding
= p(hmz by descent)

FIRST COUSINS:

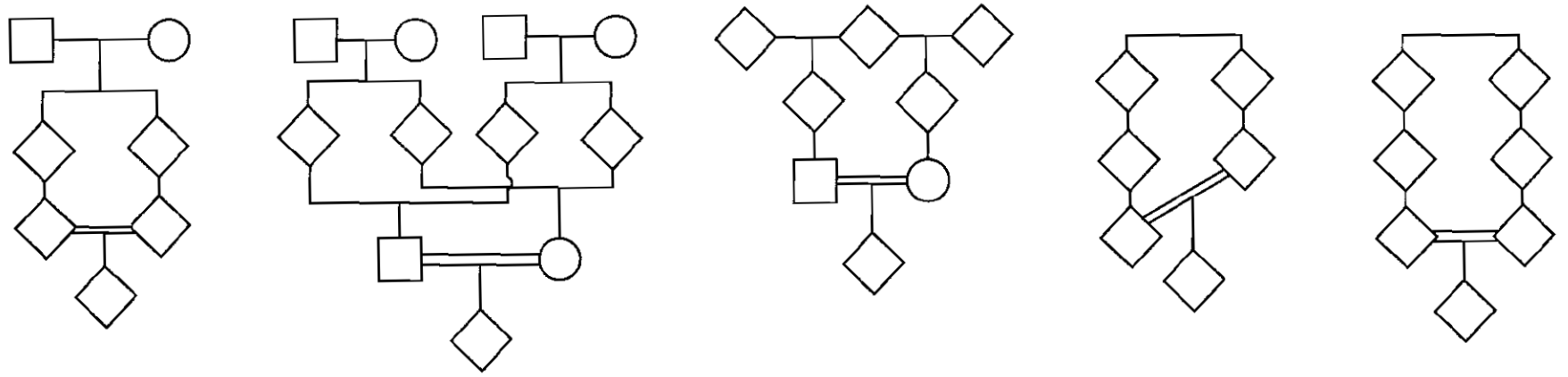
- 4 alleles in common ancestors
- Proba of *one* ancestral allele (a) hmz = 1/64
- Proba *any* of the 4 ancestral alleles hmz = 1/16 = F



First cousins

4 x 1/64 = **1/16**

Coefficient of inbreeding (F)

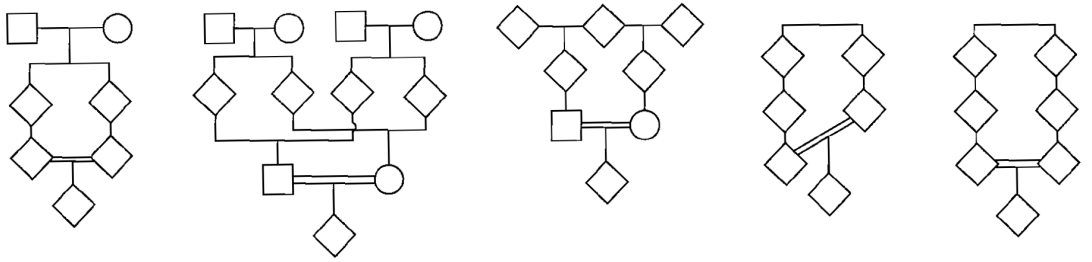


Type	Degree of Relationship	Proportion of Genes in Common	Coefficient of Inbreeding of Child (F)
Monozygotic twins	NA	1	NA
Parent-child	1st	1/2	1/4
Brother-sister (including dizygotic twins)	1st	1/2	1/4
Brother–half sister	2nd	1/4	1/8
Uncle-niece or aunt-nephew	2nd	1/4	1/8
Half uncle–niece	3rd	1/8	1/16
First cousins	3rd	1/8	1/16
Double first cousins	2nd	1/4	1/8
Half first cousins	4th	1/16	1/32
First cousins once removed	4th	1/16	1/32
Second cousins	5th	1/32	1/64

Coefficients of inbreeding for the offspring of a number of consanguineous matings. If a person is inbred through more than one line of descent, the separate coefficients are summed to find his or her total coefficient of inbreeding. NA, not applicable

PS Harper, Practical Genetic Counselling, 6th ed

Coefficient of inbreeding (F)



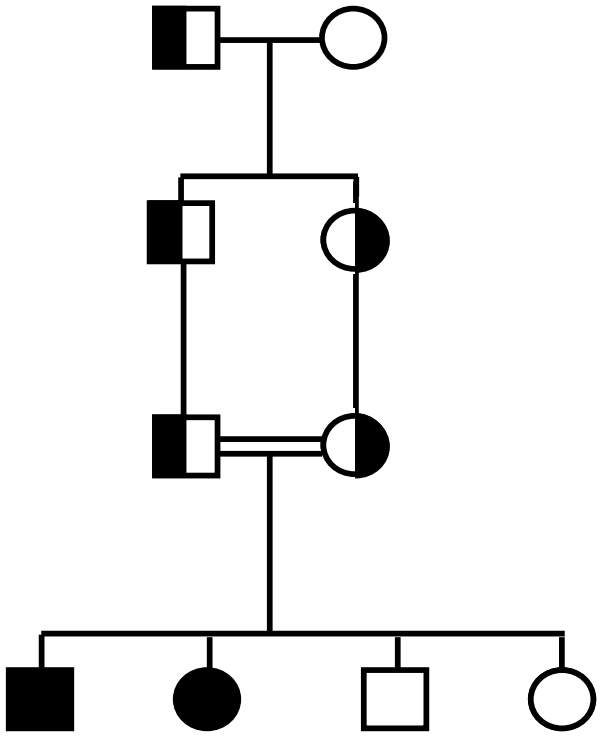
First cousins | Double first cousins | Half first cousins | First cousins once removed | Second cousins

Type	Degree of Relationship	Proportion of Genes in Common	Coefficient of Inbreeding of Child (F)
Monozygotic twins	NA	1	NA
Parent-child	1st	1/2	1/4
Brother-sister (including dizygotic twins)	1st	1/2	1/4
Brother–half sister	2nd	1/4	1/8
Uncle-niece or aunt-nephew	2nd	1/4	1/8
Half uncle–niece	3rd	1/8	1/16
First cousins	3rd	1/8	1/16
Double first cousins	2nd	1/4	1/8
Half first cousins	4th	1/16	1/32
First cousins once removed	4th	1/16	1/32
Second cousins	5th	1/32	1/64

Coefficients of inbreeding for the offspring of a number of consanguineous matings. If a person is inbred through more than one line of descent, the separate coefficients are summed to find his or her total coefficient of inbreeding. NA, not applicable

- F = 1/16 in offspring of first-cousin parents
- F = probability of homozygosity at any given locus
  true homozygosity, identity-by-descent, autozygosity

Consanguinity and AR disease



- One ancestral mutation
- Identical-by-descent (IBD) in first cousins
- Affected offspring = true homozygote = « autozygote »
- Rarer mutation <=> more cases due to consanguinity

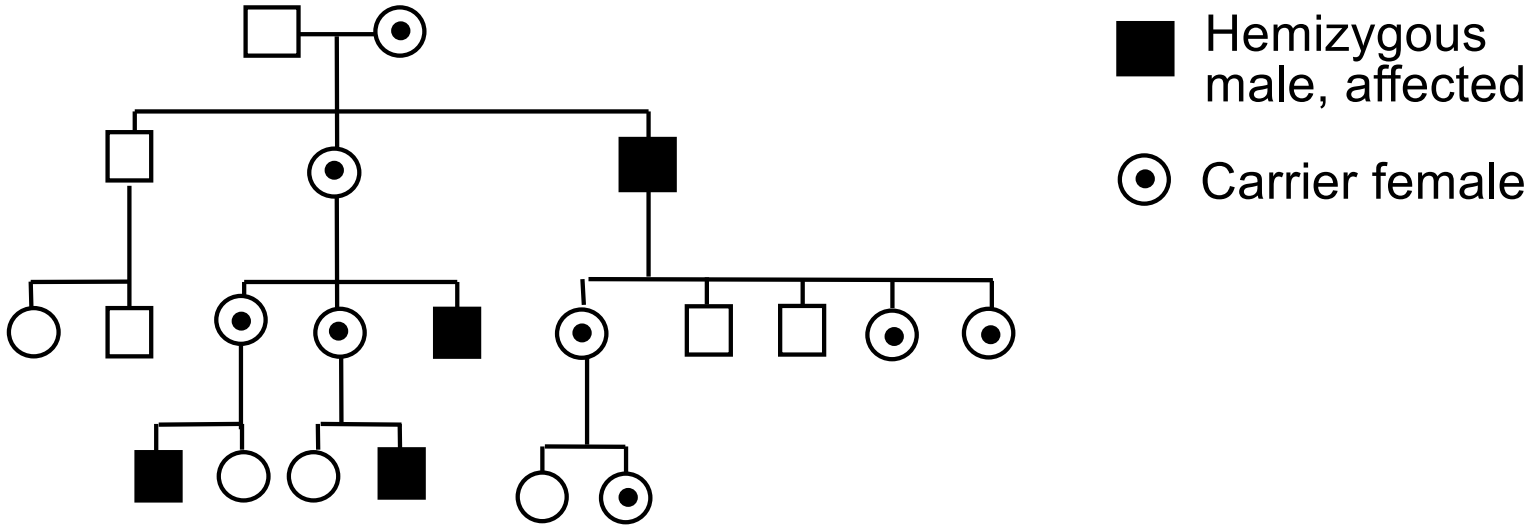
q << F => autozygosity likely to cause the disease

Patterns of single-gene inheritance

X-LINKED DISORDERS

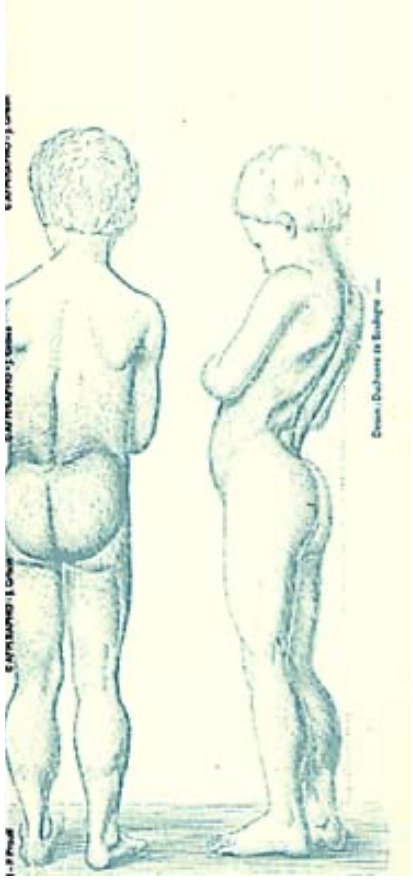
X-linked recessive phenotype: oblique transmission

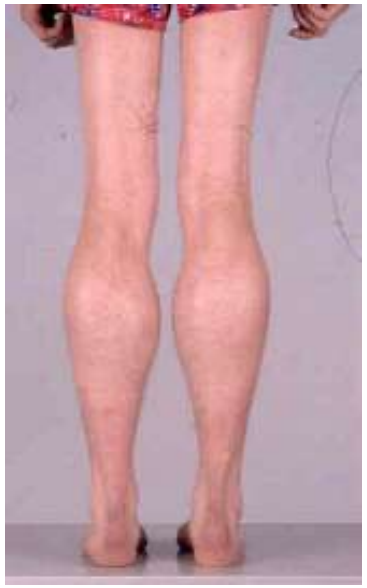
genotype: hemizygous mutation in male



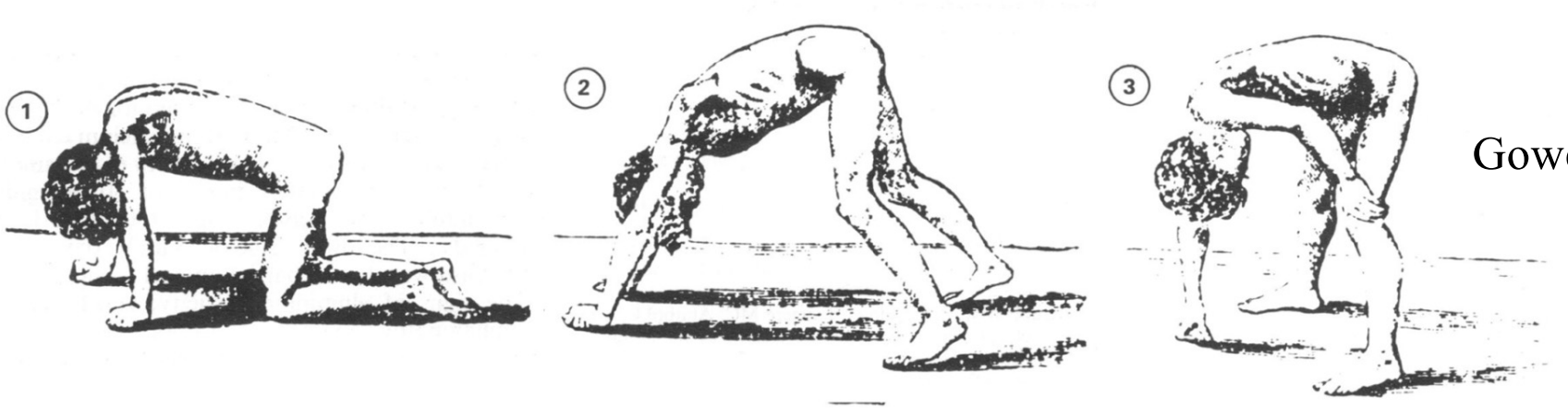
- Ex: hemophilia A, B; Duchenne Muscular Dystrophy (DMD)...
- All daughters of affected males are carriers .
- All sons of affected male are unaffected (Y chromosome).
- No male-to-male transmission.

Duchenne Muscular Dystrophy (DMD)



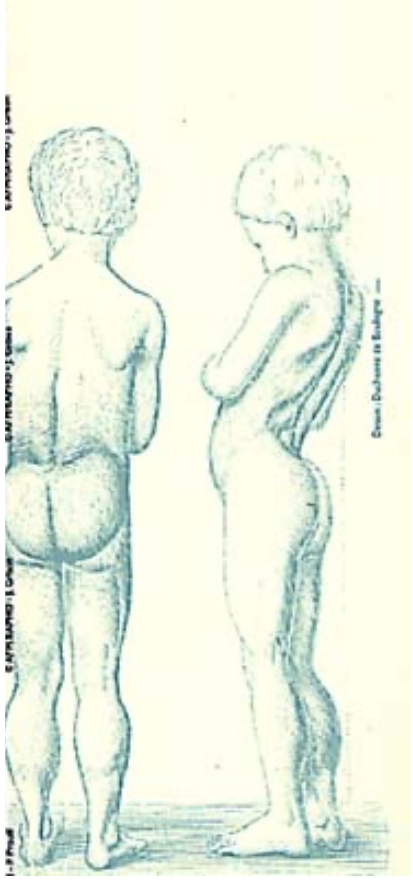


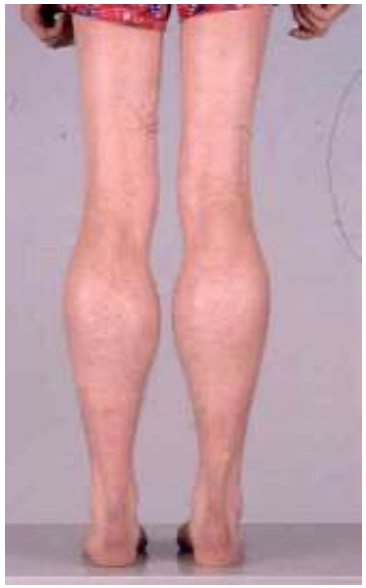
- 1/10,000 birth
- Boys, almost all patients
- Progressive decay of muscular fibres

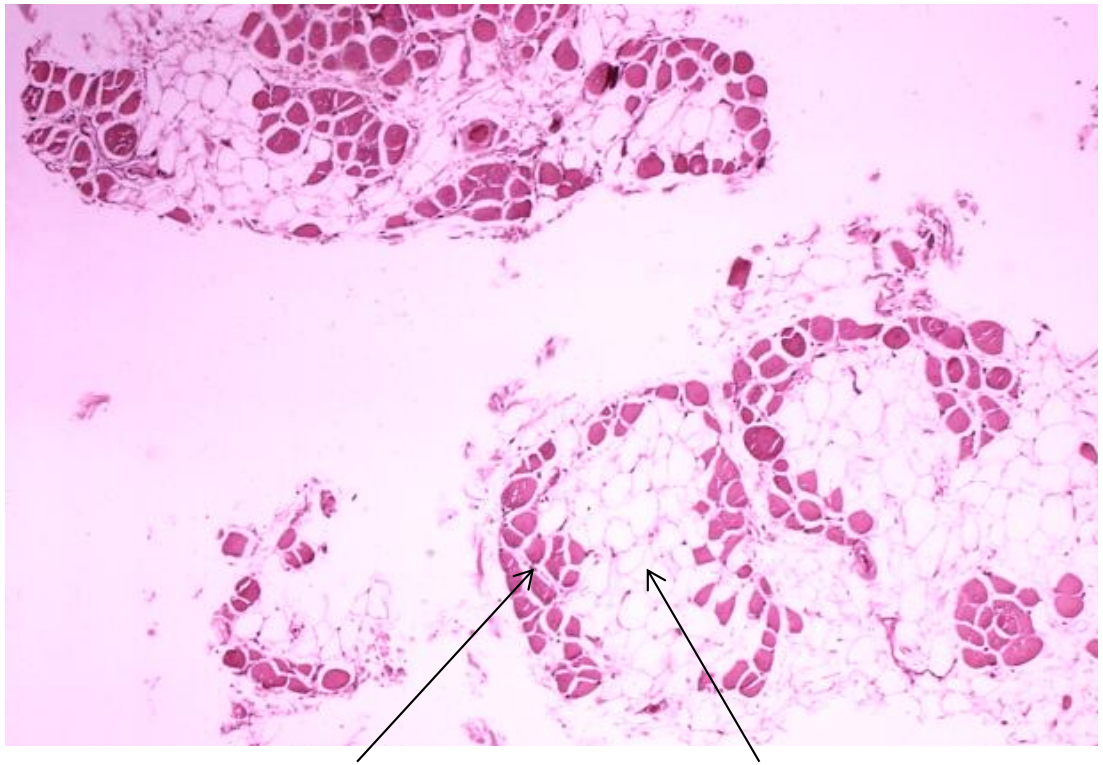


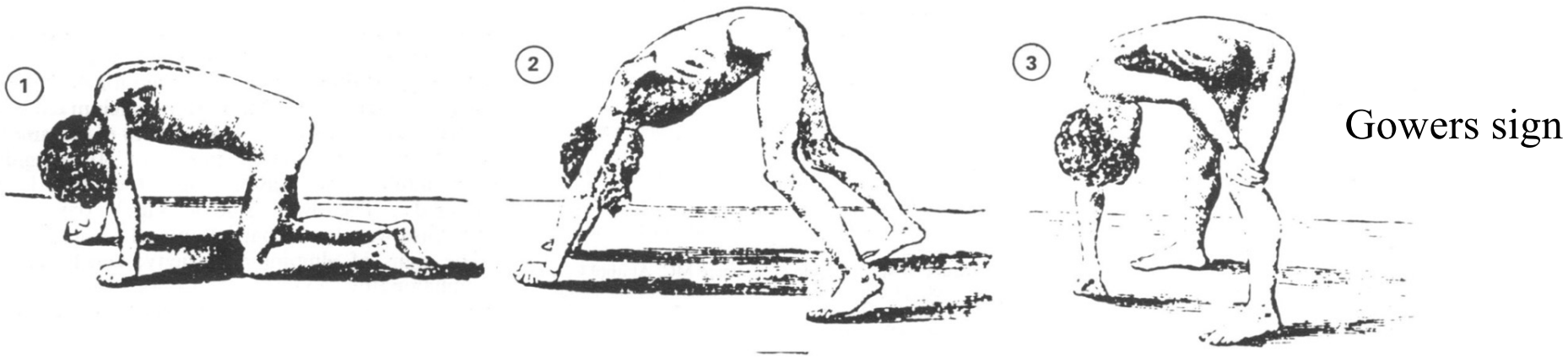
Gowers sign

Duchenne Muscular Dystrophy (DMD)



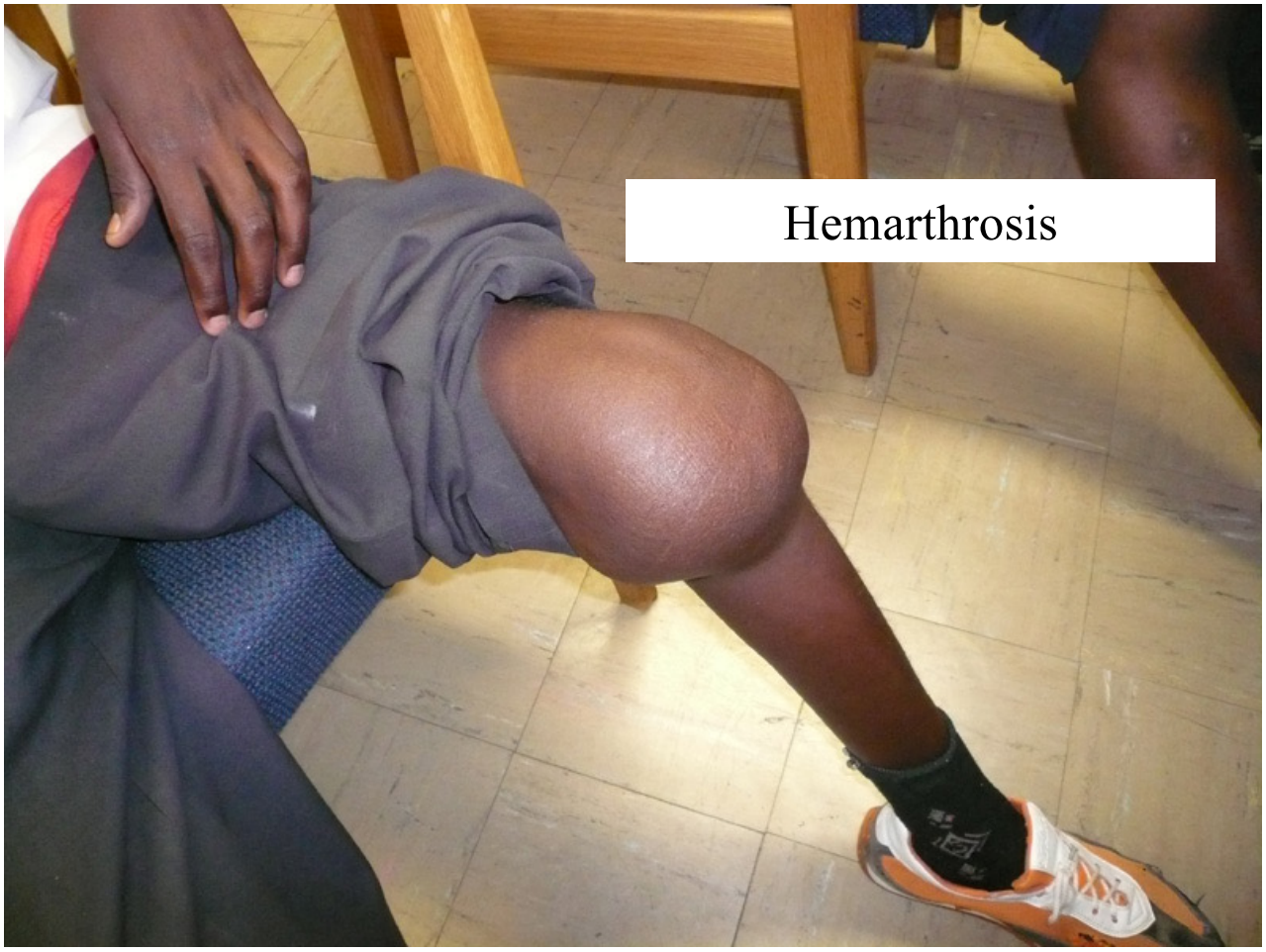


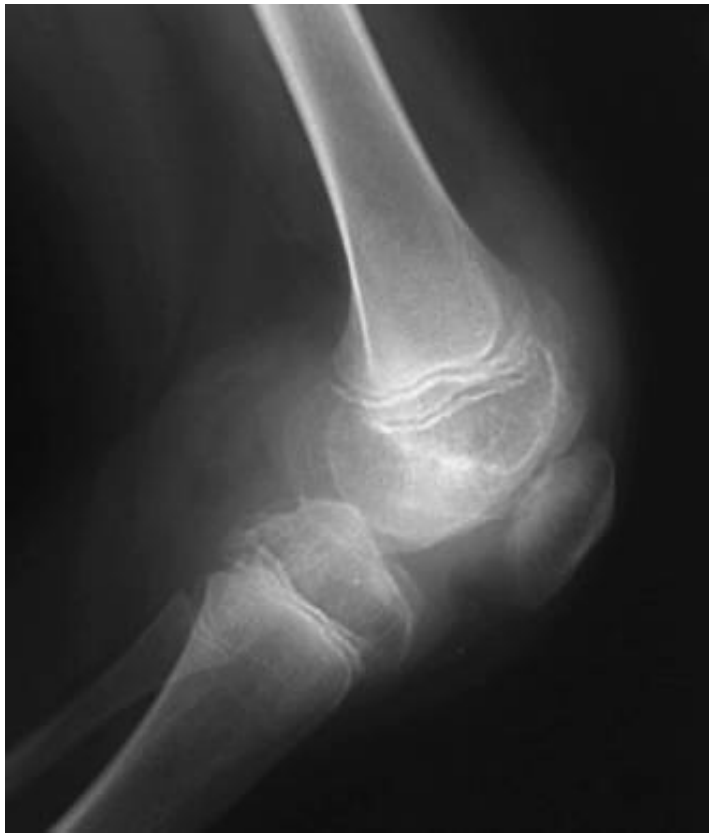




Gowers sign

Hemophilia





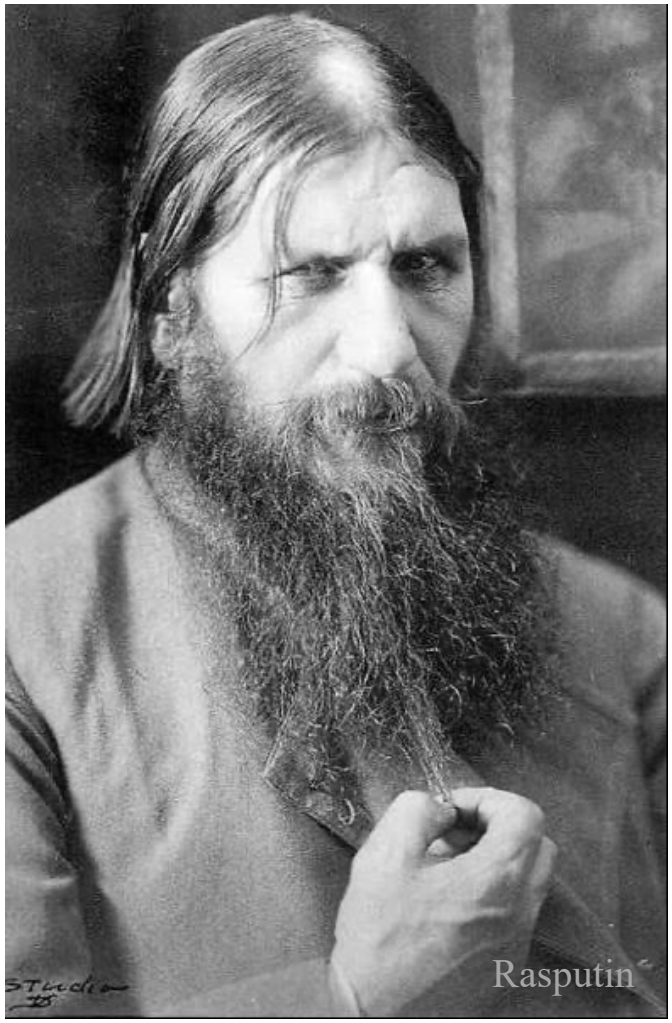
Rx knee

Inherited defect of coagulation. Rare: 1/10.000

Hemophilia

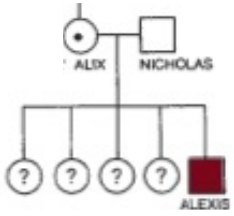
Inherited defect of coagulation. Rare: 1/10.000

Hemophilia



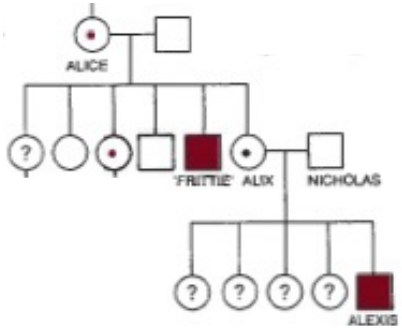
Inherited defect of coagulation. Rare: 1/10.000

Hemophilia



Inherited defect of coagulation. Rare: 1/10.000

Hemophilia



Inherited defect of coagulation. Rare: 1/10.000

Hemophilia

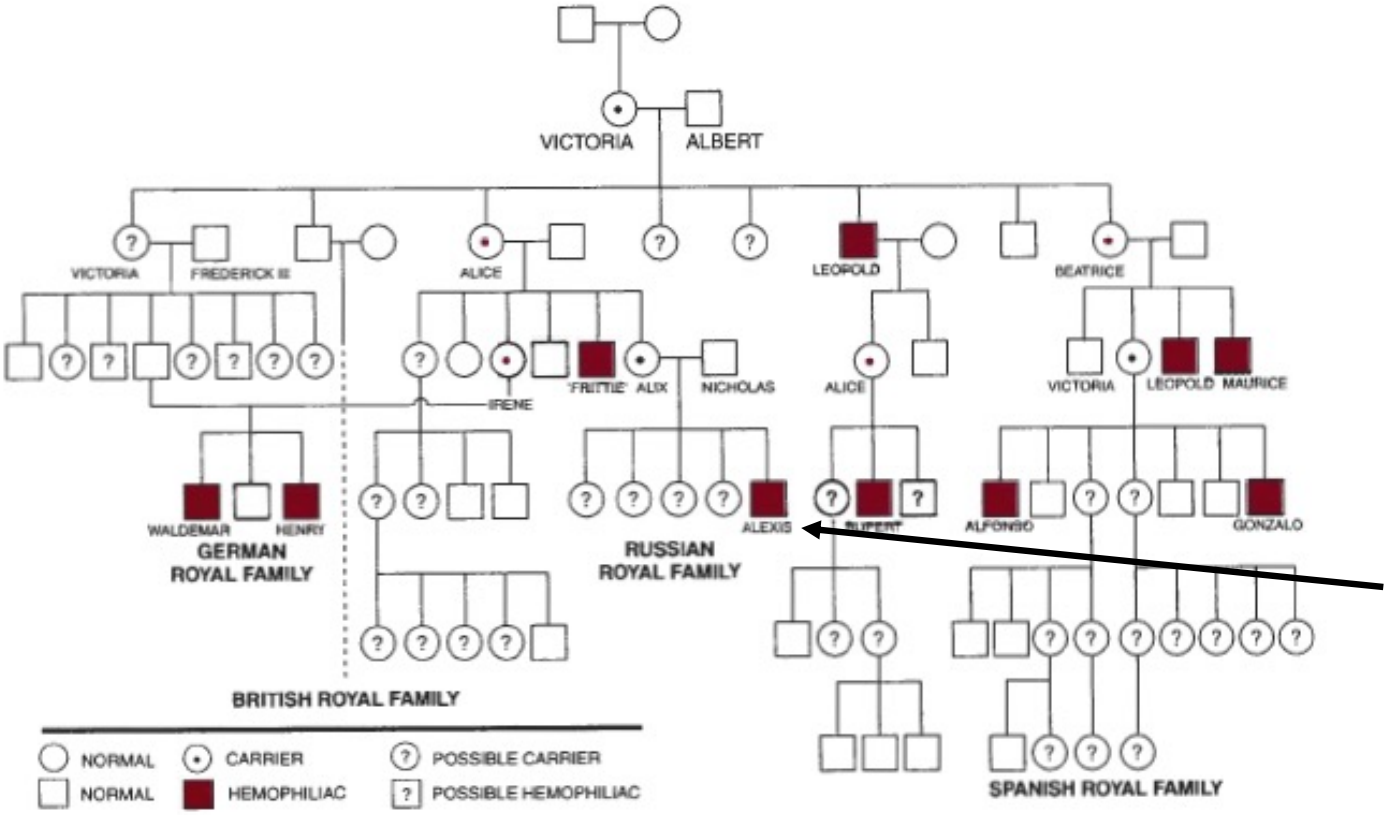


Figure 7.16. **Queen Victoria's pedigree.** Though the X-linked recessive inheritance pat-

- Inherited defect of coagulation. Rare: 1/10.000
- This family shows that the disease must be genetic and monofactorial from mutation on X chromosome

Hemophilia

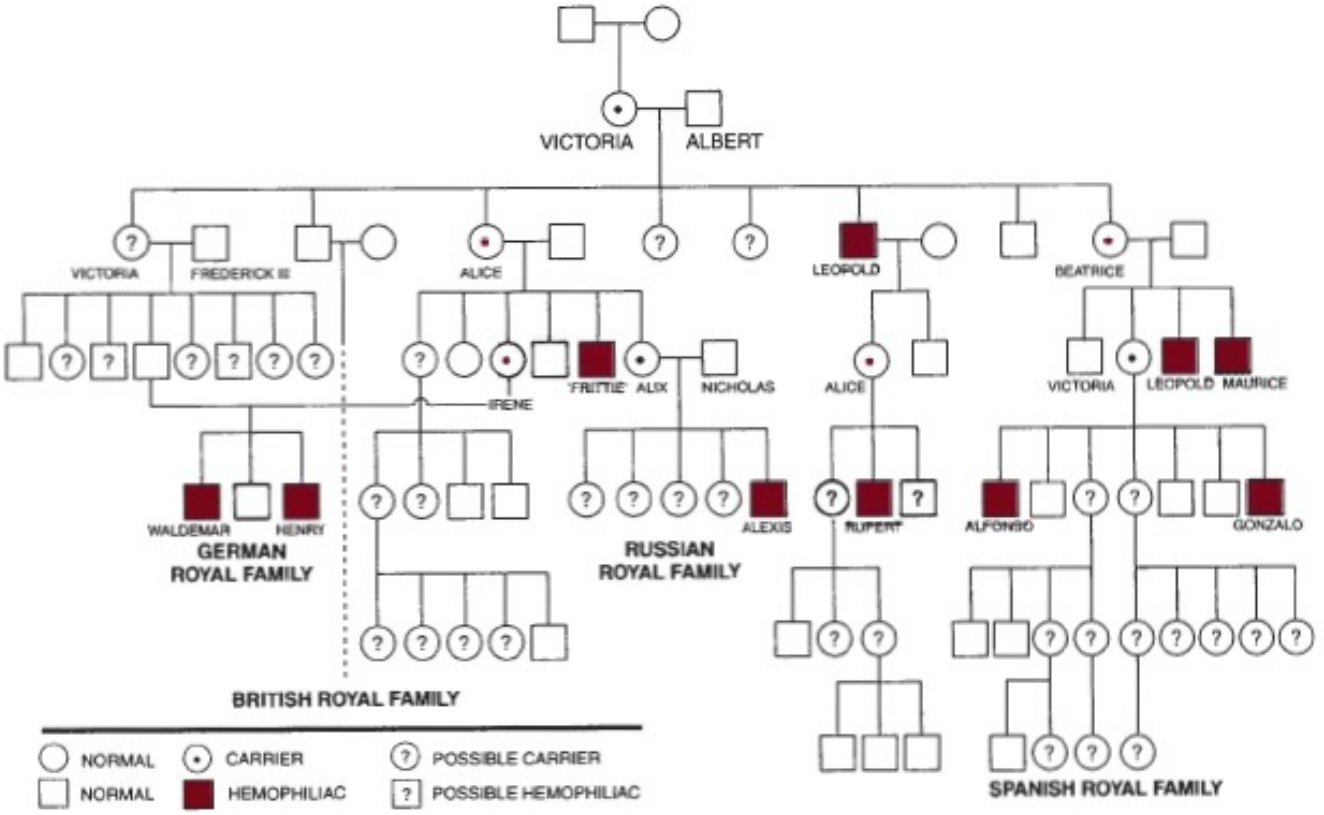
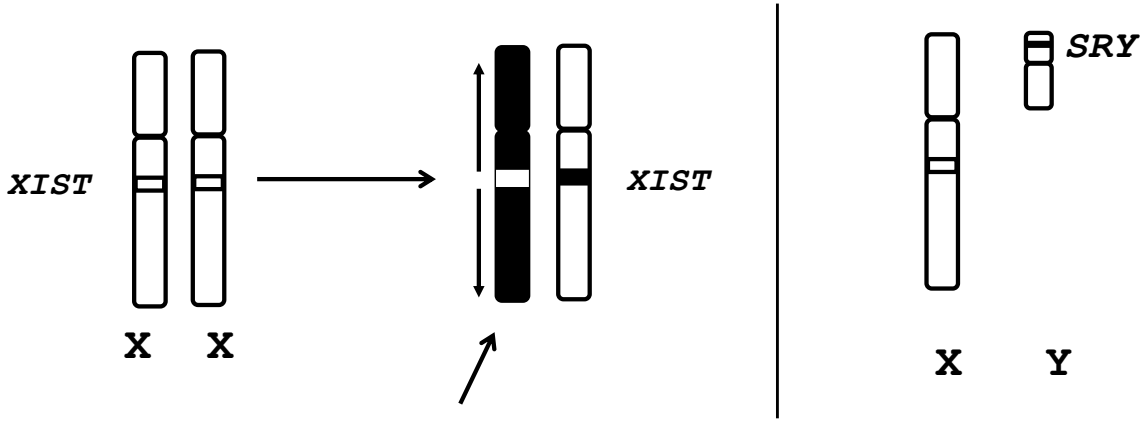


Figure 7.16. Queen Victoria's pedigree. Though the X-linked recessive inheritance pat-

- High penetrance, severe monogenic disease
- Women asymptomatic

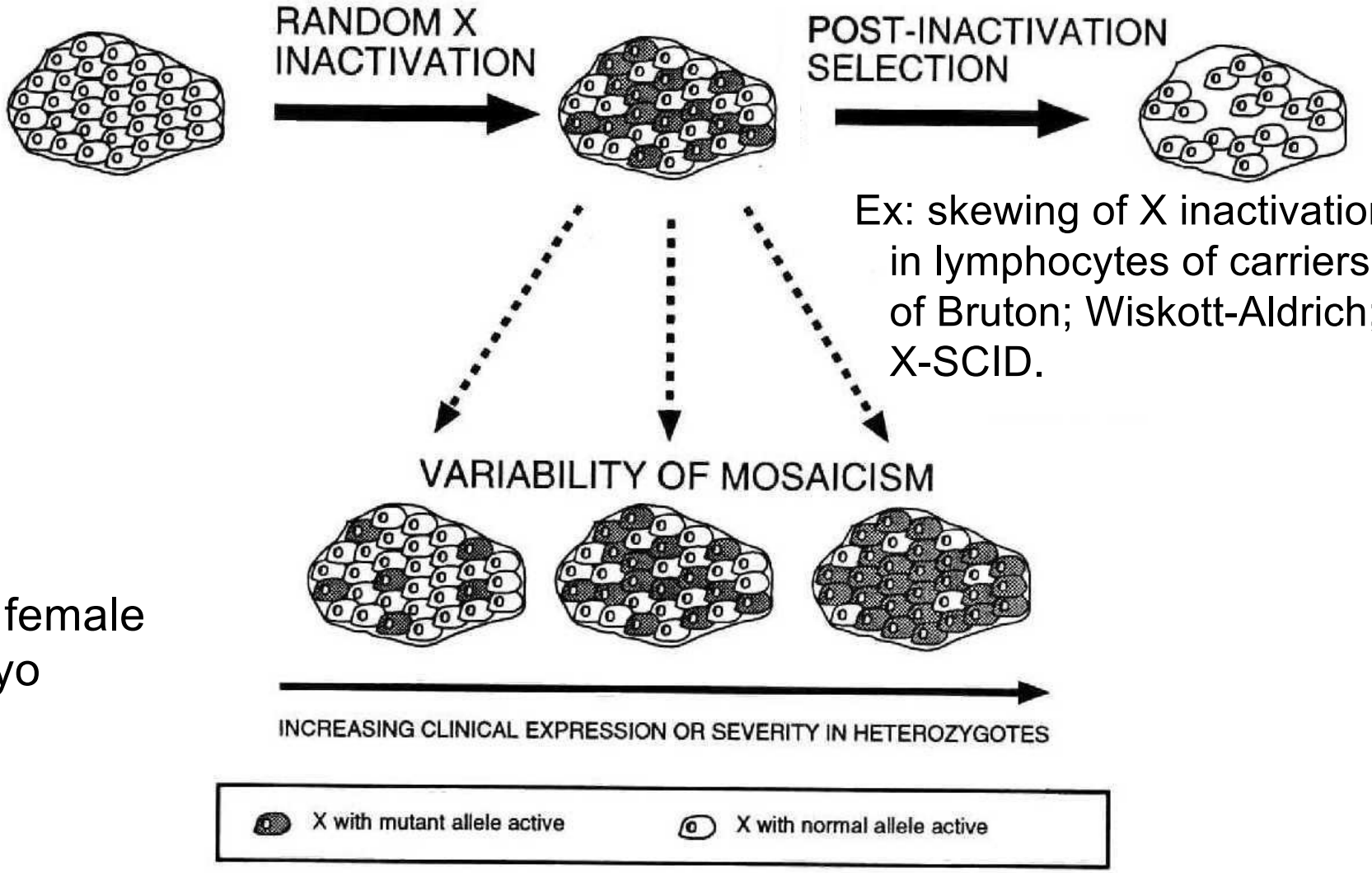
Women are mosaic for 2 cell populations



XIST gene active
on inactive chr.X

XX Embryo < 100 cells => random inactivation of one X chromos
(Lyonisation)

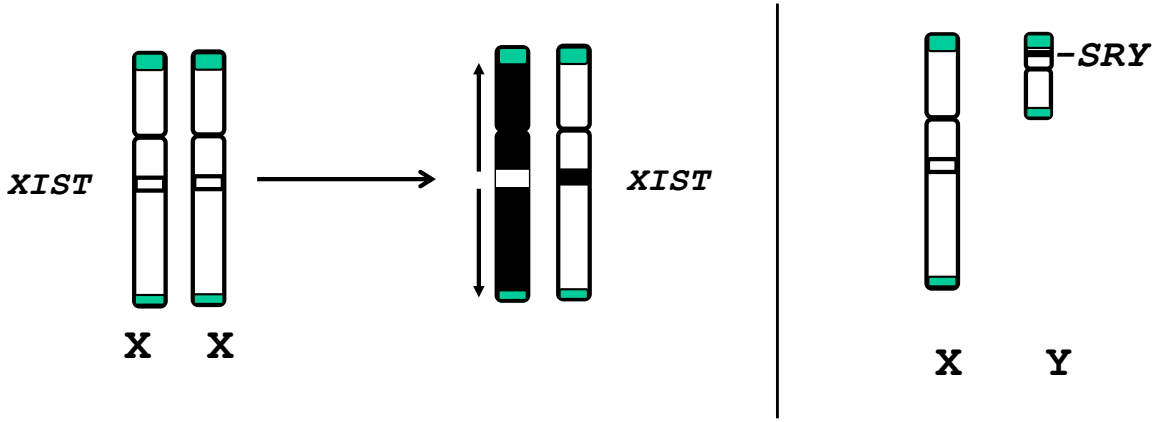
X chromosome inactivation in female somatic cells. Early. Random.



Ex: skewing of X inactivation in lymphocytes of carriers of Bruton; Wiskott-Aldrich; X-SCID.

Early female embryo

Pseudo-autosomal regions

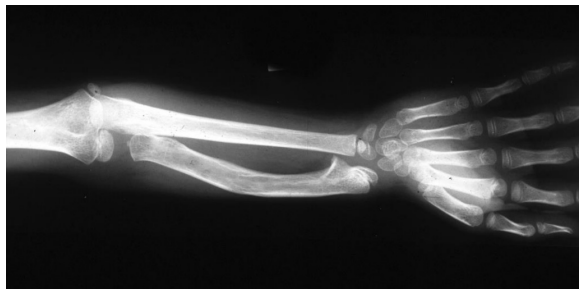


- On X and Y
- Escape X-inactivation
- Recombine by crossing-over

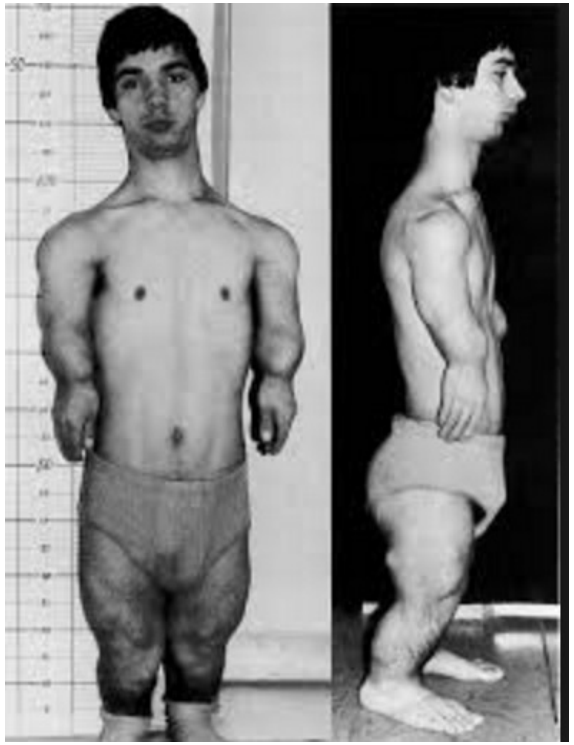
- AD (or AR) heredity, not X-linked
  - eg, SHOX mutations
- SRY very close to Y-PAR
  - Too centromeric cr-ov produces XX males

Pseudo-autosomal inheritance

- A few phenotypes
- Eg: SHOX gene – linked dyschondrosteosis
- Male to male transmission, male and female affected… if dominant: AUTOSOMAL DOMINANT pattern



- Or AUTOSOMAL RECESSIVE: Langer mesomelic dysplasia



Baxova et al. 2008 AJMG

- « 23rd chromosome »

AD inheritance of SHOX-associated phenotype

On the X and Y chromosome, so inheritance is not X-linked

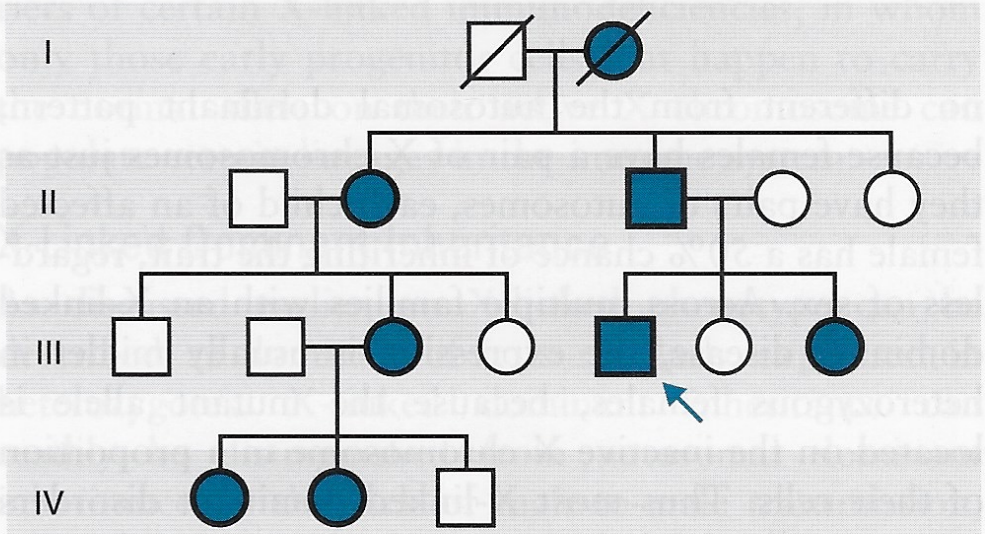
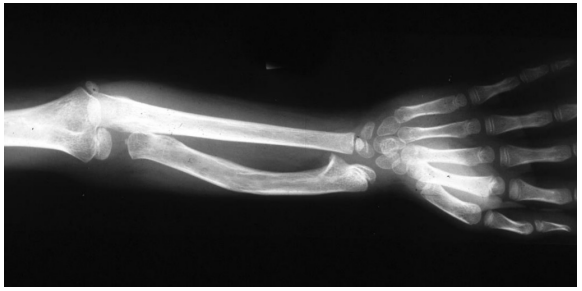
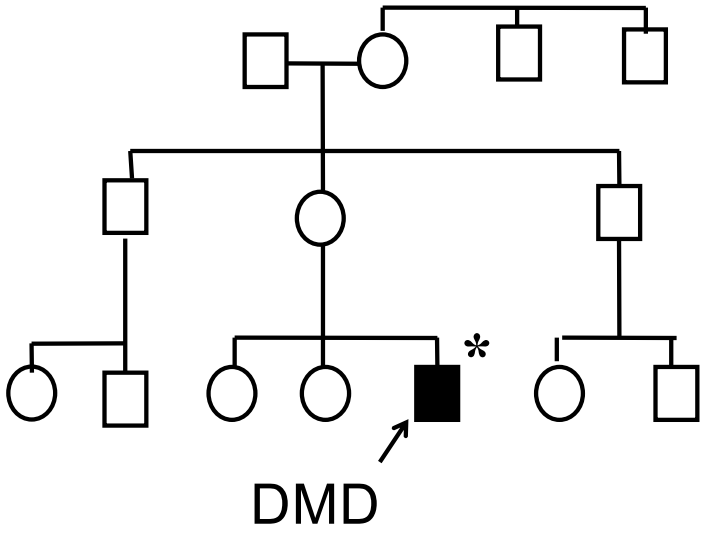


Figure 7-16 Pedigree showing inheritance of dyschondrosteosis due to mutations in *SHOX*, a pseudoautosomal gene on the X and Y chromosomes. The *arrow* shows a male who inherited the trait on his Y chromosome from his father. His father, however, inherited the trait on his X chromosome from his mother. *See Sources & Acknowledgments.*



X-linked gene new mutation: in patient OR mother

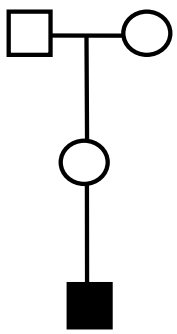


New mutation:

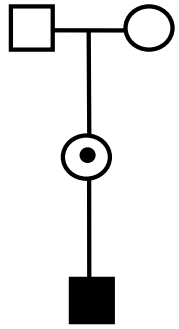
1. Either in patient
   - Mother's egg cell
2. Or in patient's mother
   - Grandmaternal egg cell
   - Grandpaternal sperm

---

1. Risk in further boy very small

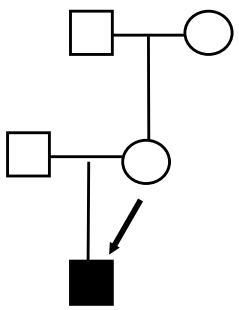


2. Risk in further boy = 50%

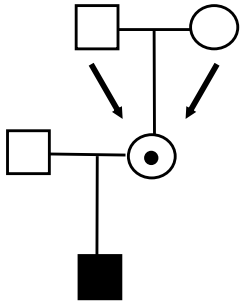


In-between (1) and (2):
maternal germ-line mosaic

Many new X-linked gene mutations arise in maternal Grandfather



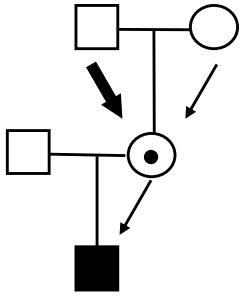
1/3 of cases



2/3 cases

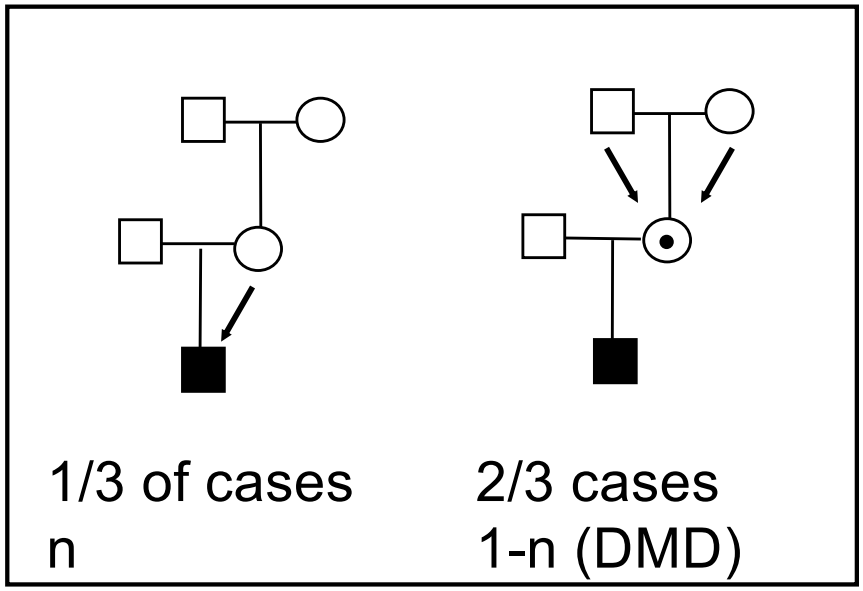
- 66% carrier mothers rate applies to DMD
  - High contribution of large genomic deletions
- Higher (>85%) for most other Xgenes
  - High contribution of testicle-derived point mutations

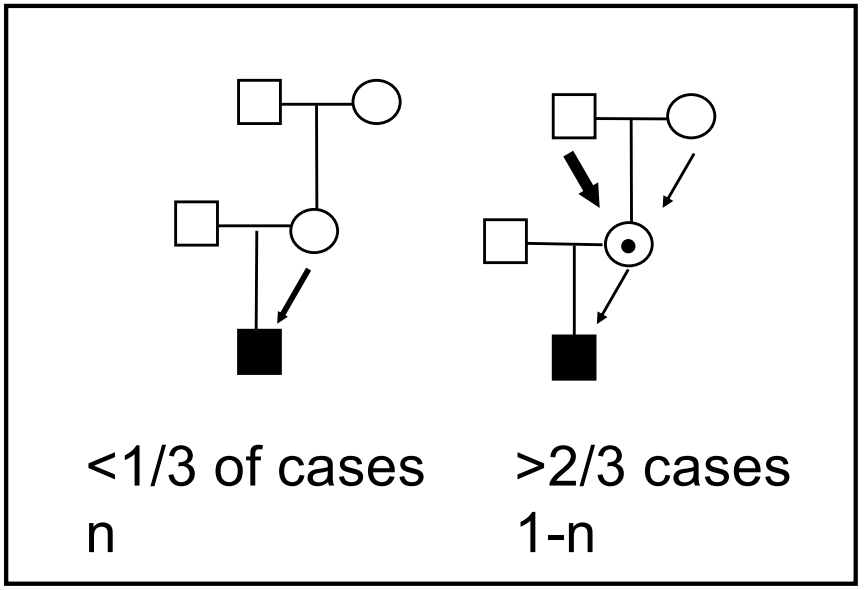
- Grand-paternal age effect



NB: father not involved (Y chromosome)

Many new X-linked gene mutations arise in maternal Grandfather





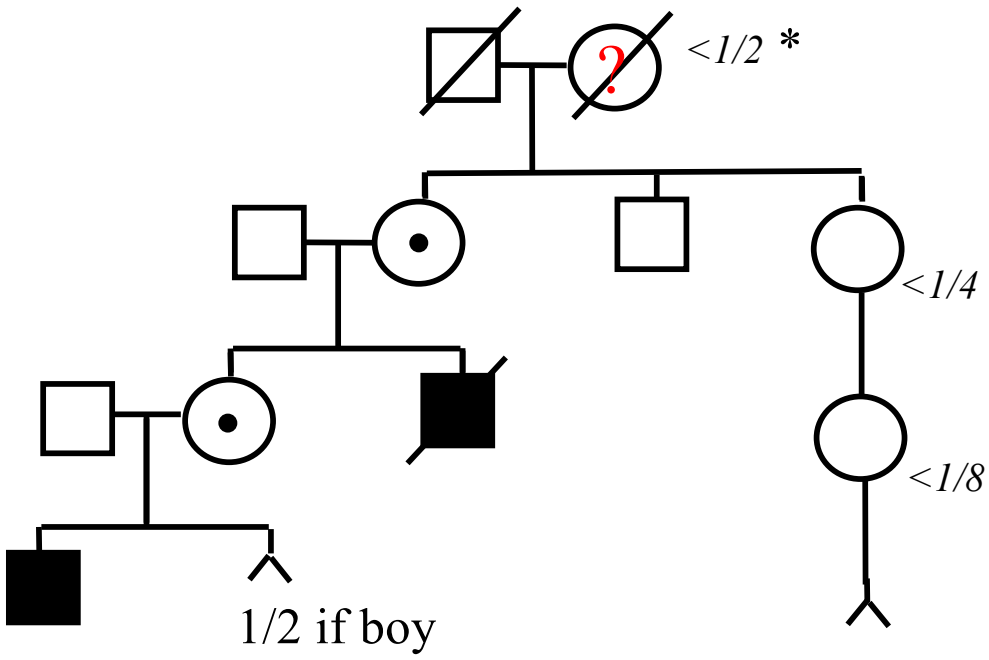
n = frequency of new mutation in affected boy, not in mother

n = (1-f) / (k + 2) where k = male mutation rate / female mutation rate ; f = fitness

- DMD: equal rate in both sexes: k = 1, and f = 0
  => n = 1/3
- Most X genes: k > 1 (k range: 2 – 10) => n < 1/3
- If f = 1 (and patient number constant): no new mutation

Risk in mother of carrier woman?

Ex: DMD



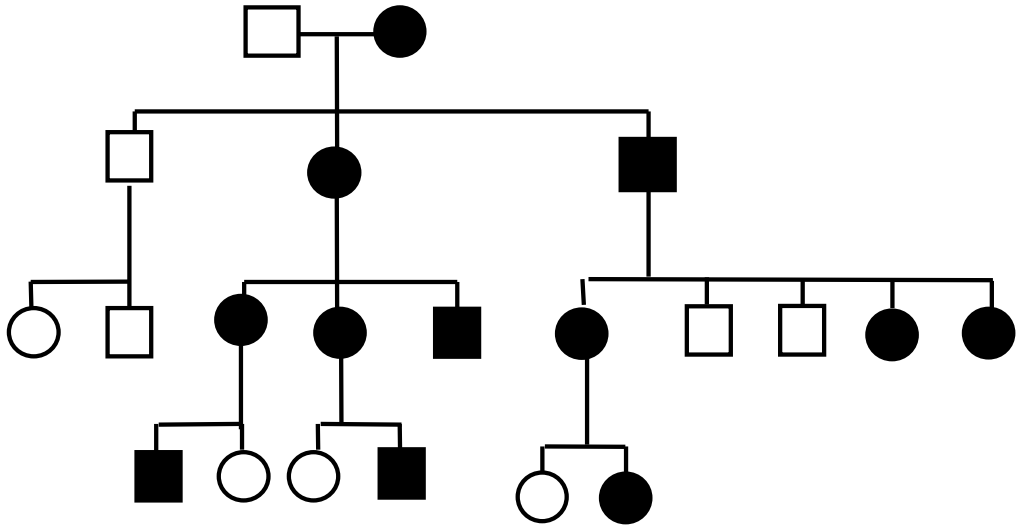
* proportion of carrier women's mothers who are carrier themselves is maximum ½

If male = female mutation rate (k=1) and f=0, maximum risk = 1/2

Risk <1/2 if k>1

In previous generations of women, risk is halved going up the pedigree (as it is halved going down the pedigree)

X-linked dominant phenotype

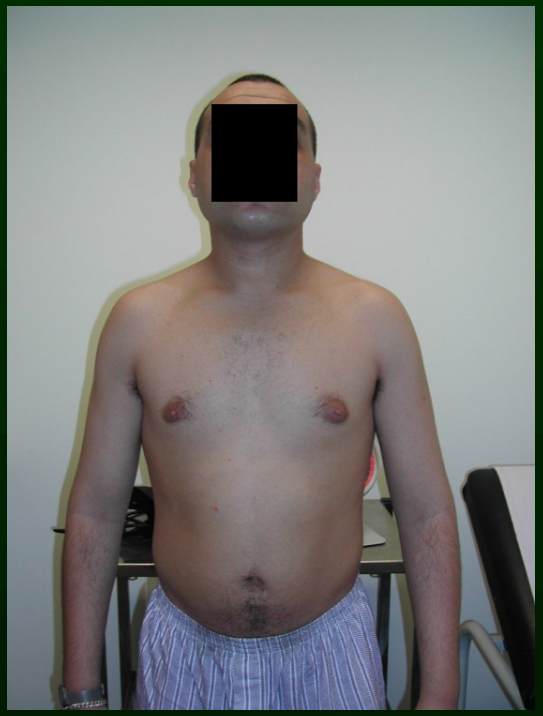


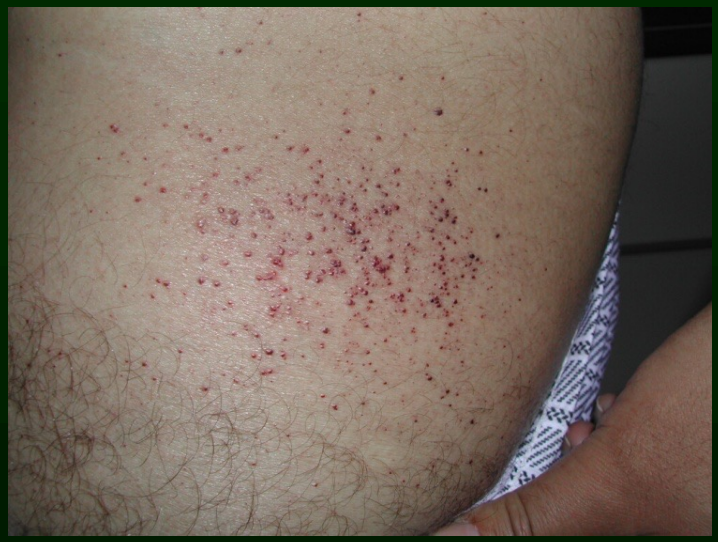
Hemizygous male, affected

Heterozygous female, affected

- Females have 2 chromosome X
- F/M ratio = 2/1

Fabry

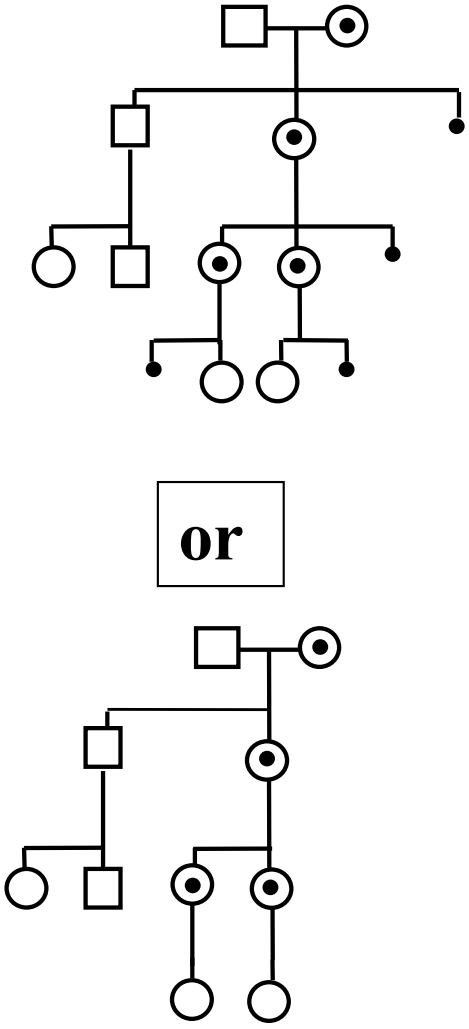




- Renal failure, cardiophaty, stroke.
- X gene, but most carrier females become sick

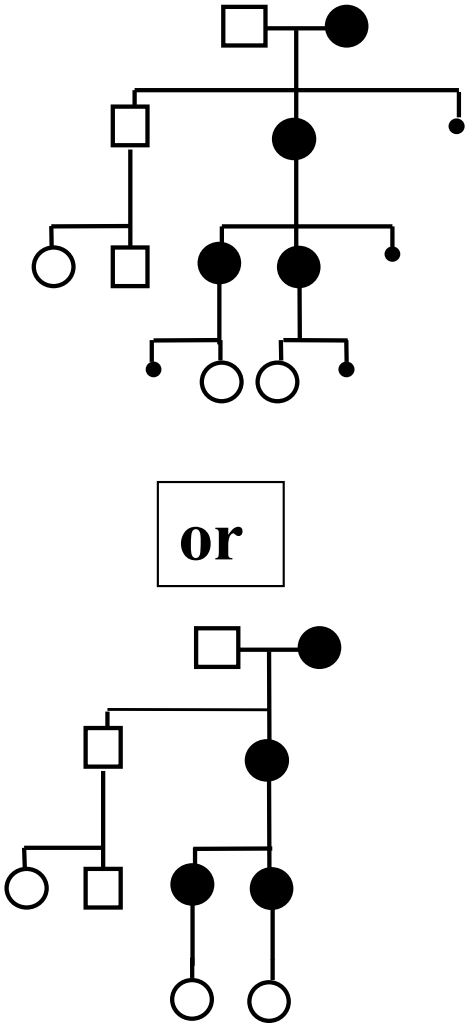
X-linked recessive, lethal in hemizygous male

- 50% of expected male births



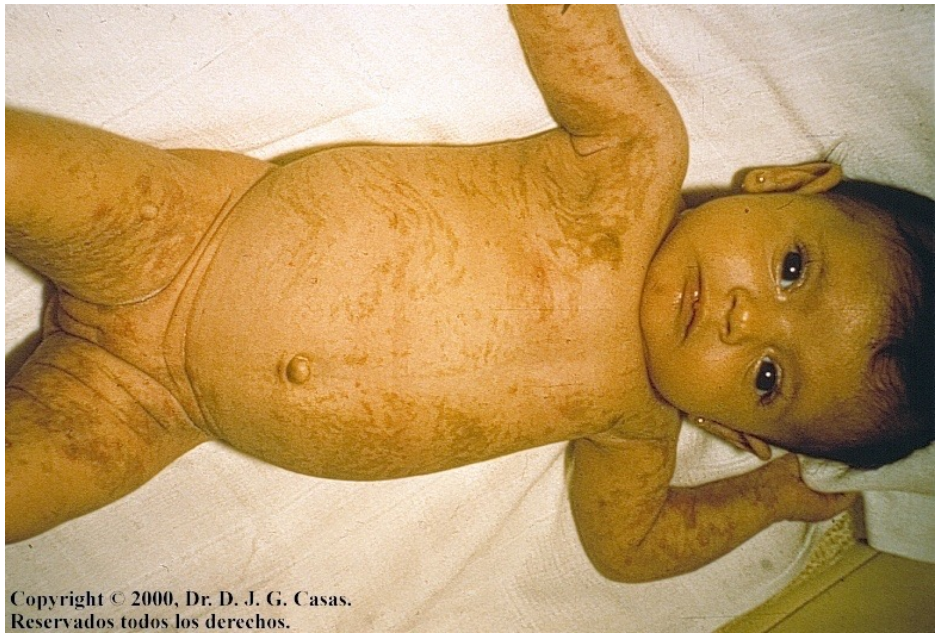
X-linked dominant, lethal in hemizygous male

- 50% of expected male births
- Affected females only

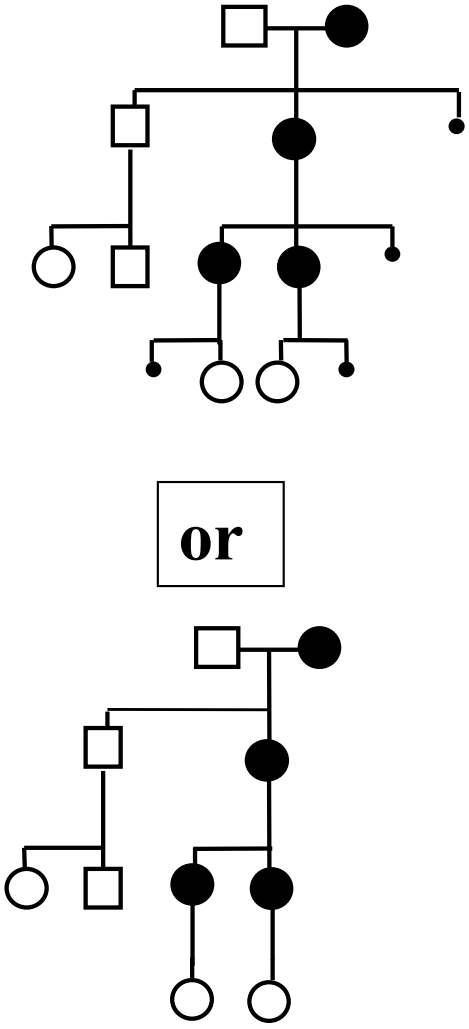


X-linked dominant, lethal in hemizygous male

- 50% of expected male births
- Affected females only
- Ex: Incontinentia pigmentii

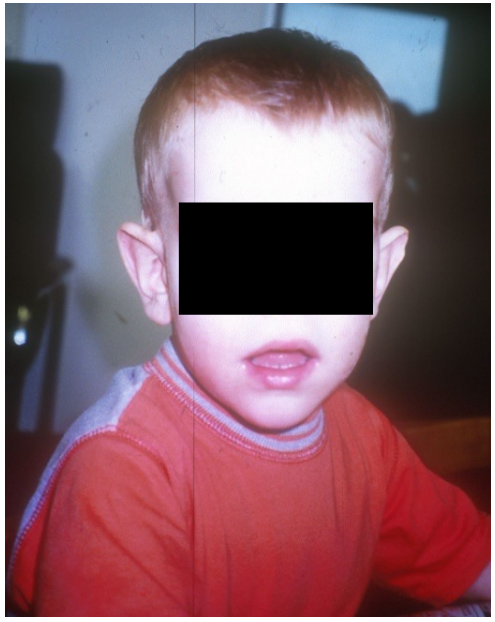


- Ex: Rett syndrome
  - Fitness ~0 => neomut only (except germ-line mosaics)

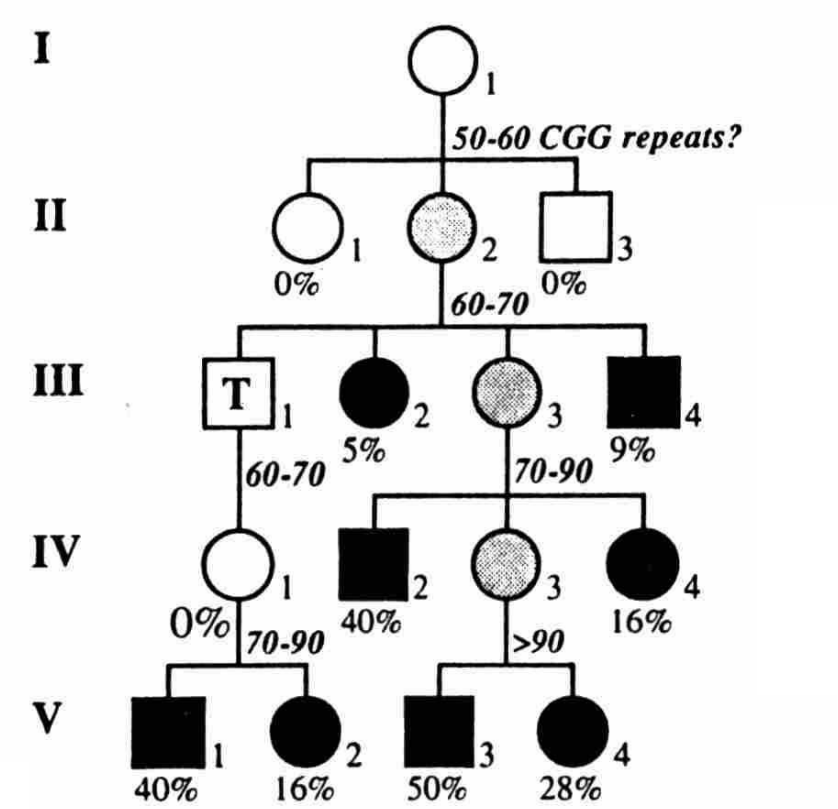


Fragile X syndrome

- 1st cause of hereditary MR ?
- X-linked, complex: « semi-dominant with incomplete penetrance »
  - Women often affected
  - Normal transmitting males
  - MR risk depends on position in pedigree



Fra-X: complex X-linked inheritance

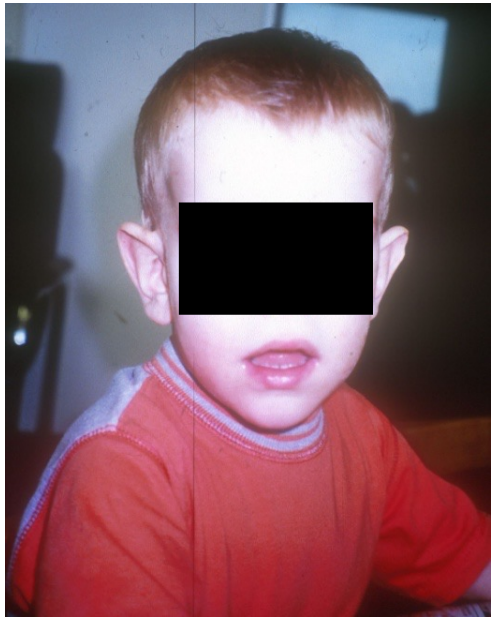


Empiric Risk of Mental Retardation Varies with Pedigree Position

Fu et al, 1991, Cell.

An example pedigree is given with data from Sherman et al. (1985) showing percent risk (below each individual) of mental retardation based on pedigree position from studies of fragile X families.

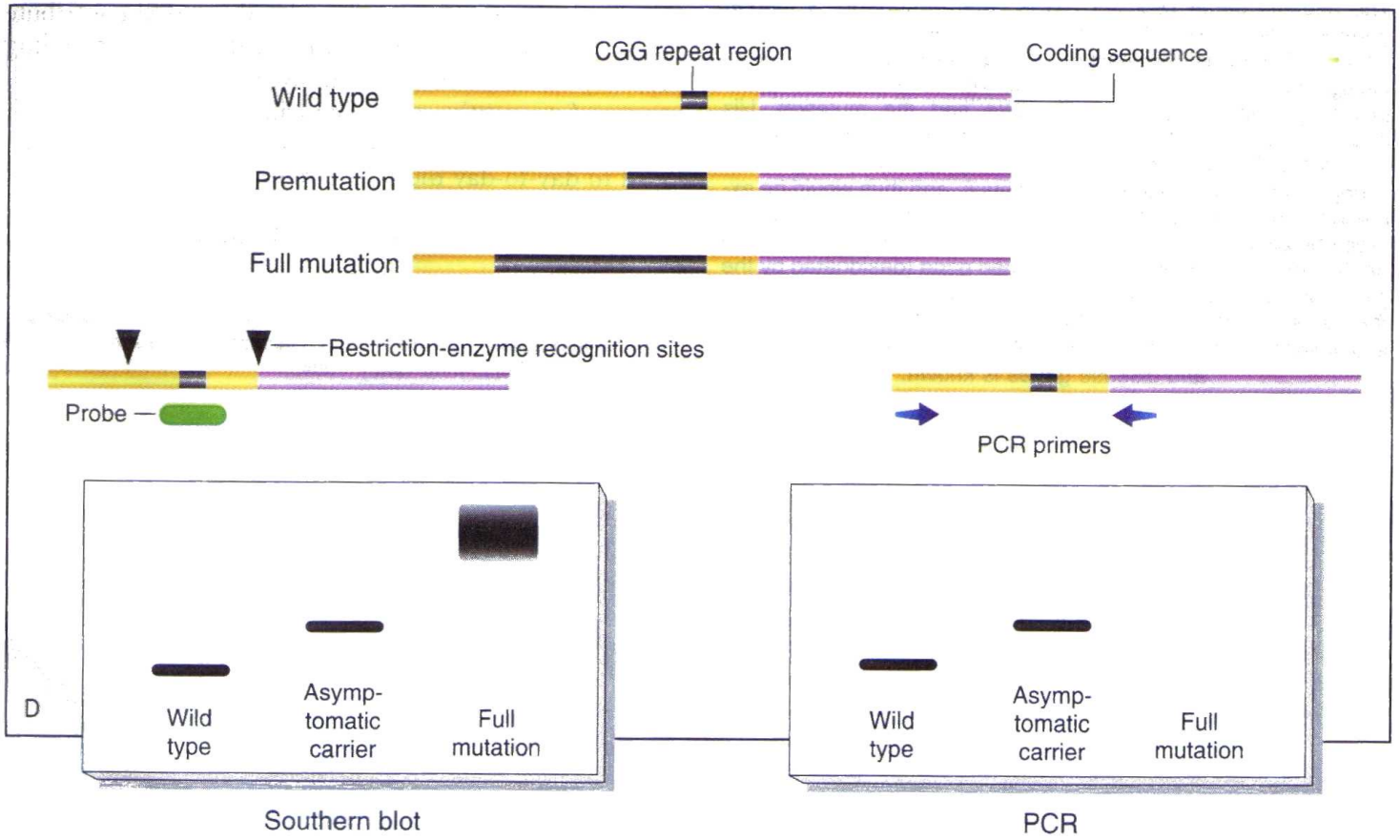
- Women often affected
- Normal transmitting males
- MR risk depends on position in pedigree



Dynamic mutation in Fra-X

- 3 types of alleles
  - Normal n<50
  - Premutated : unstable in femal transmission : n = 50-200
  - Full mutated : loss of function, semi-dominant (?): n usually >300

DNA analysis in Fra-X males

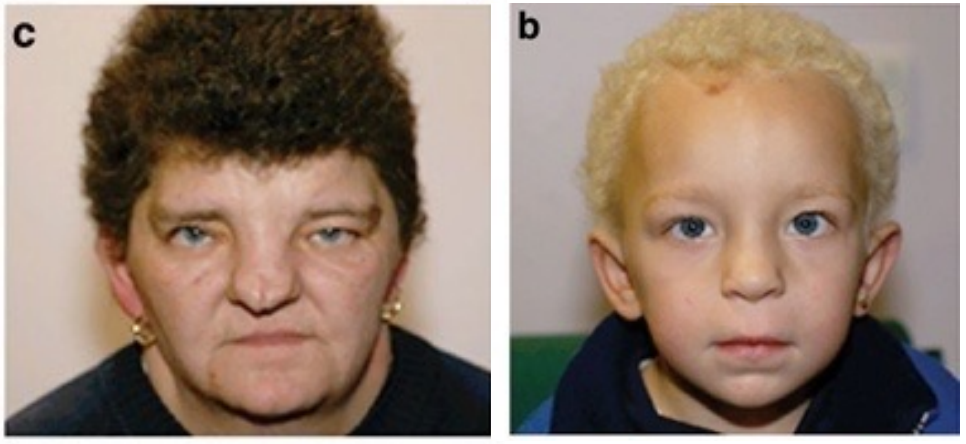


(CGG)n

- Premutation: non-methylated DNA, functional gene.
  May expand further via female meiosis.
  Expansion risk increases with size ÷ taille (n)
- Full Mutation: methylated DNA, loss of gene fn, MR

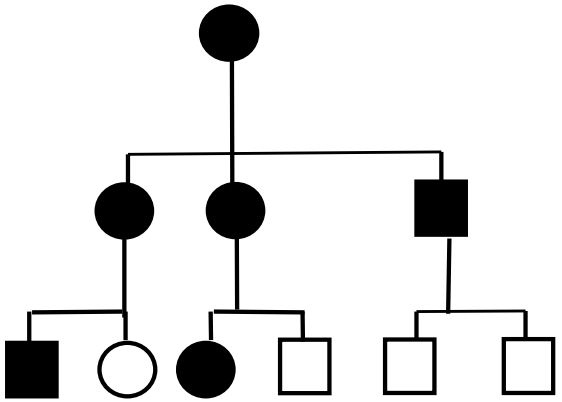
X-linked more severe in females

- Paradoxical
- Cranio-fronto-nasal dysplasia
  – *EFNB1* gene (Xq12)



Mother son

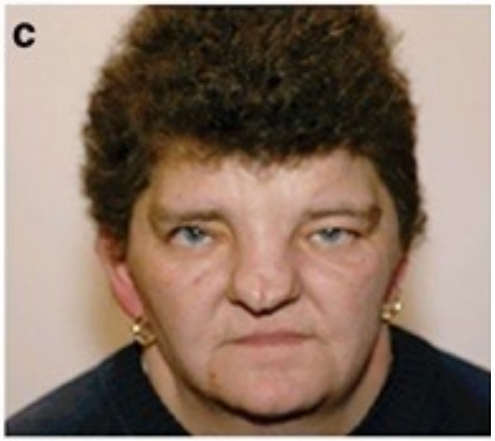
Wilkie&Coll 2006 EJHG



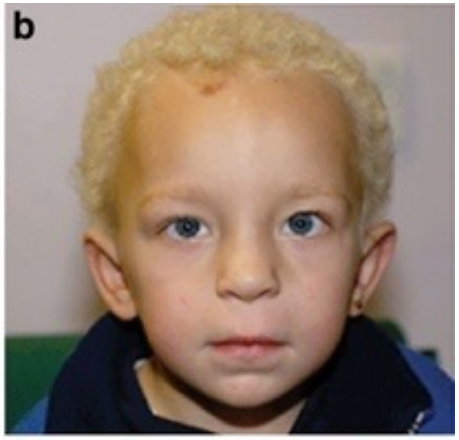
X-dom inheritance

X-linked more severe in females

- Paradoxical
- Cranio-fronto-nasal dysplasia
  - *EFNB1* gene (Xq12)
    cell autonomous
  - CELLULAR INTERFERENCE
    model

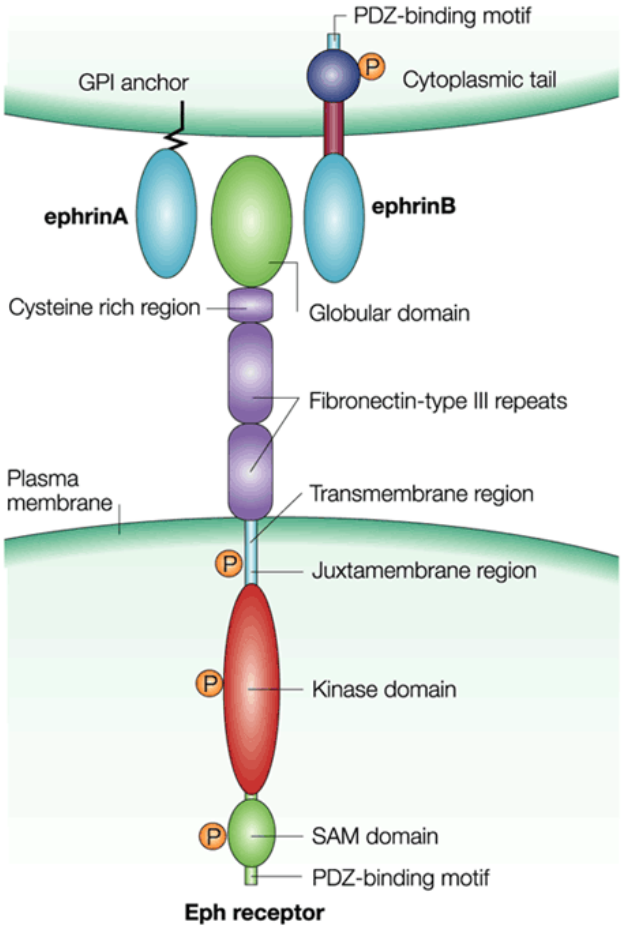


Mother



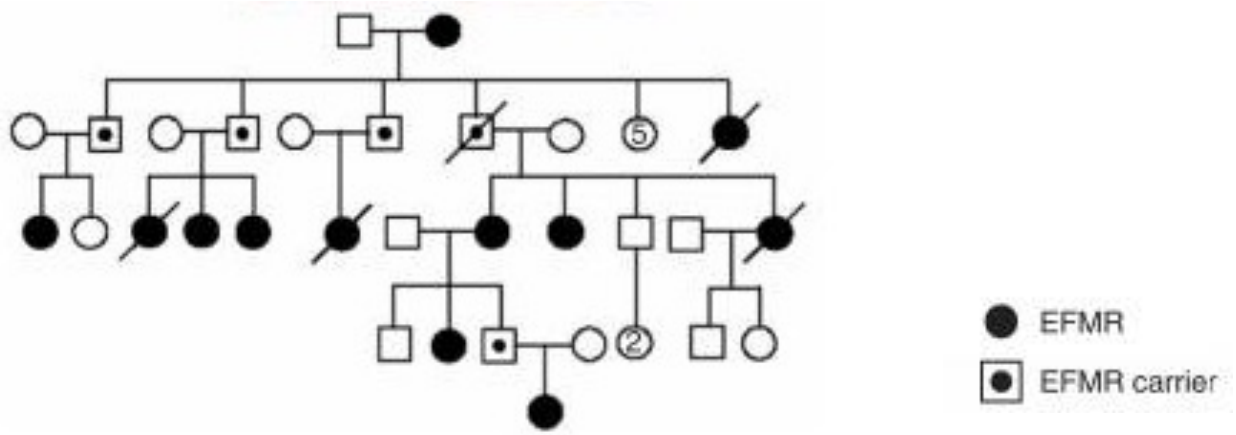
son

Wilkie&Coll 2006 EJHG



Nature Reviews | Molecular Cell Biology

EPILEPSY, FEMALE-RESTRICTED, WITH MENTAL RETARDATION; EFMR

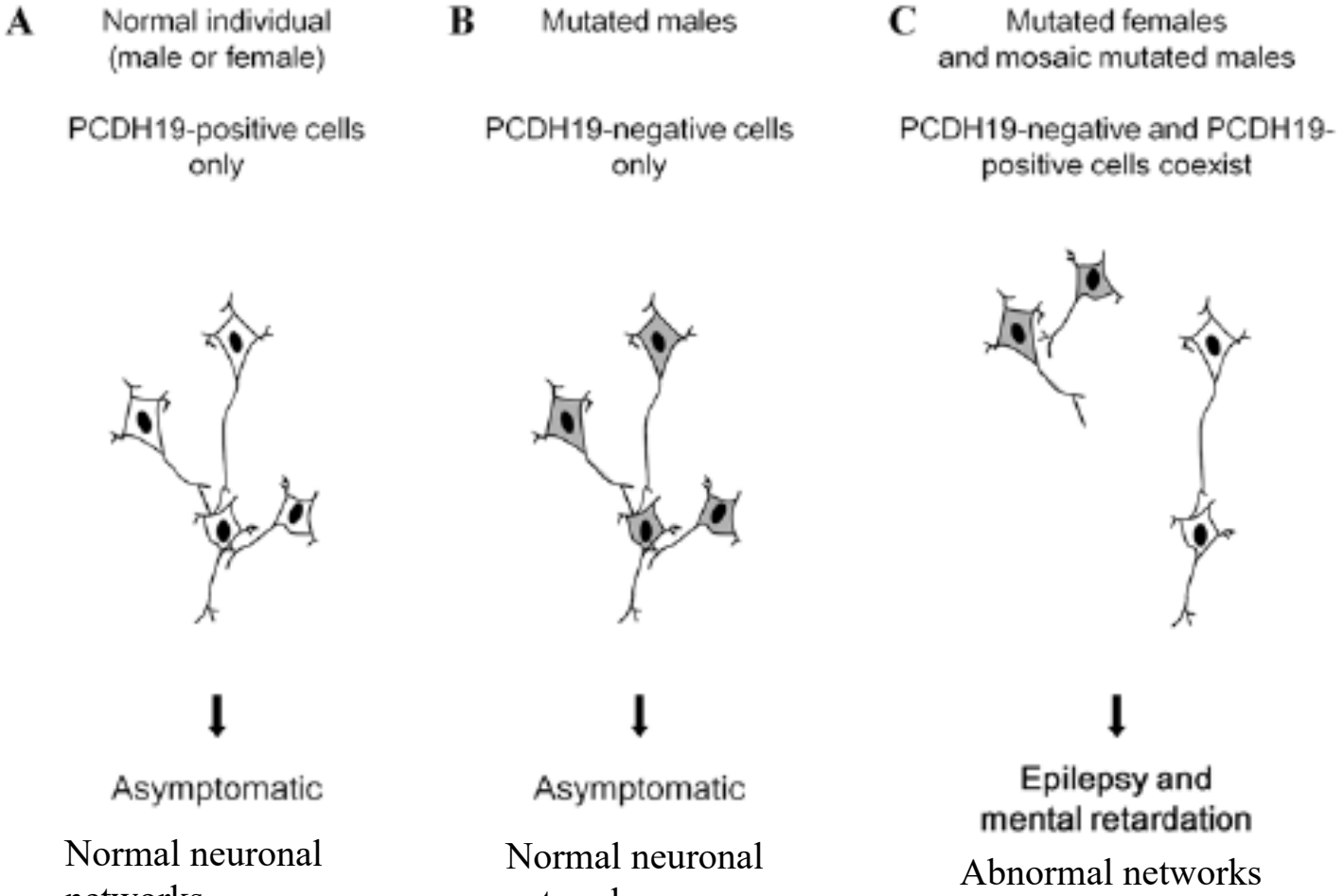


Dibbens LM et al. Nat Genet 2008
PCDH19 gene mutation

Cellular interference in another X-linked disorder

All patients are female, except 1 male who is a mosaic !

PCDH19 Mutations in Dravet-Like Syndrome
Depienne et al 2009



X-linked, cell autonomous lethal

PLP gene mutations

- X-linked disorder severe in affected boys => no sign in carrier mothers
- X-linked disorder milder in affected boys => carrier mothers are symptomatic

  paradoxical

X-linked, cell autonomous lethal

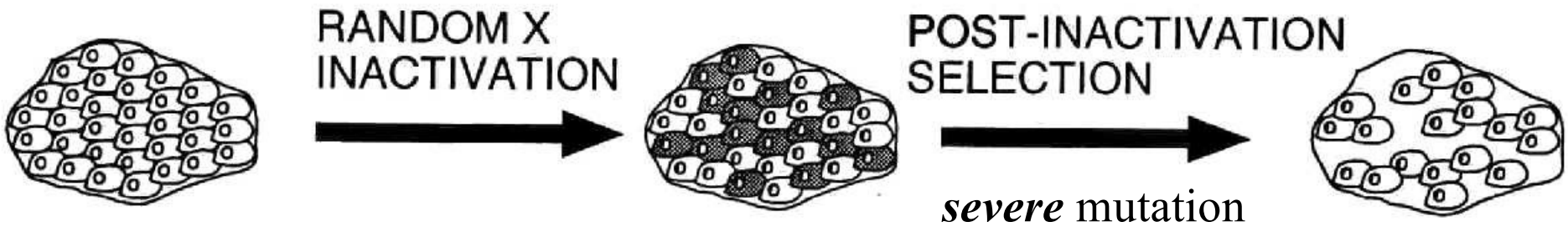
PLP gene mutations

- X-linked disorder severe in affected boys => no sign in carrier mothers
- X-linked disorder milder in affected boys => carrier mothers are symptomatic

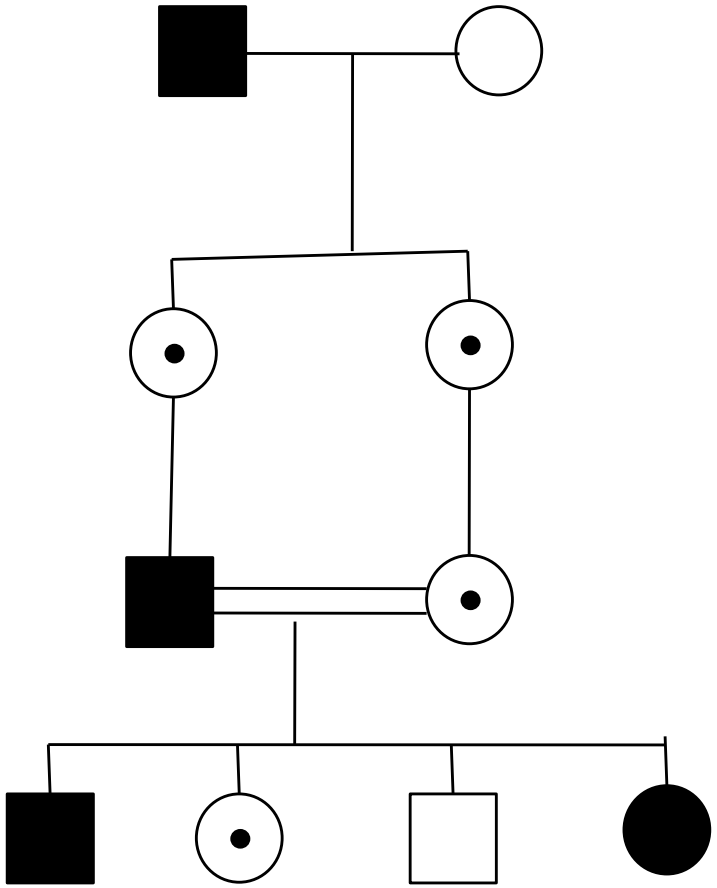
paradoxical

Mechanistic model:

- severe mutation in carrier => cell die after lyonisation healthy cell repopulate
- Mild mutation in carrier => all cells survive lyonisation, mutated cell degenerate in adult



X-linked in consanguineous family



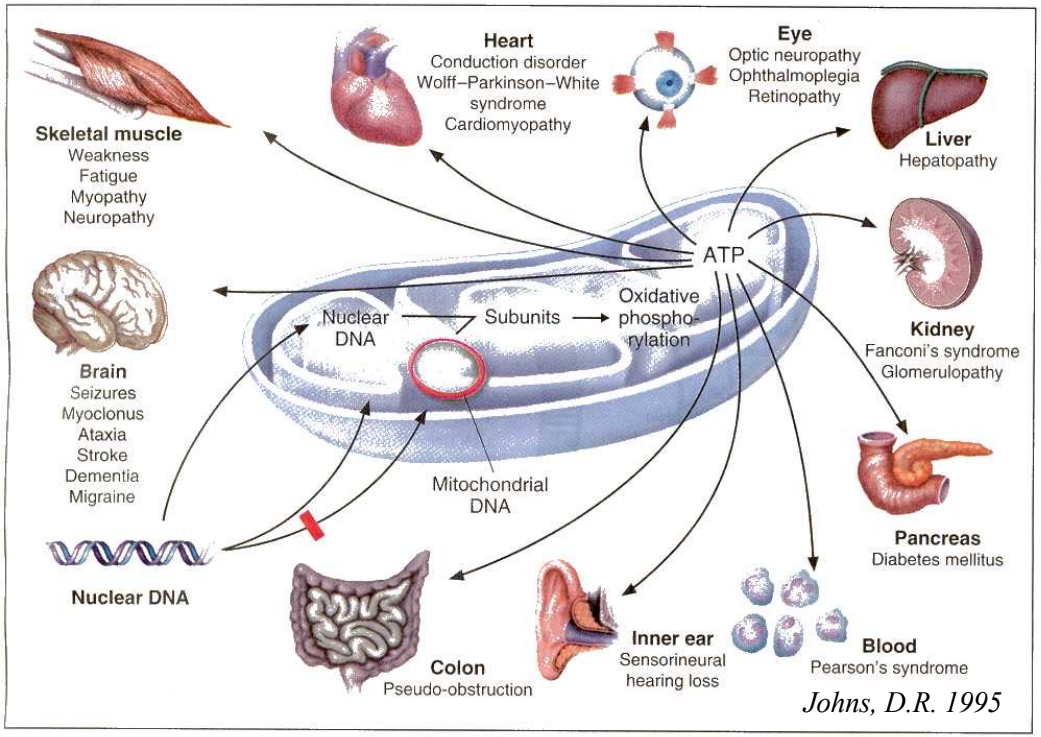
- Affected females
- Male-to-male transmission

- Also if frequent X-linked trait red-green colour-blindness

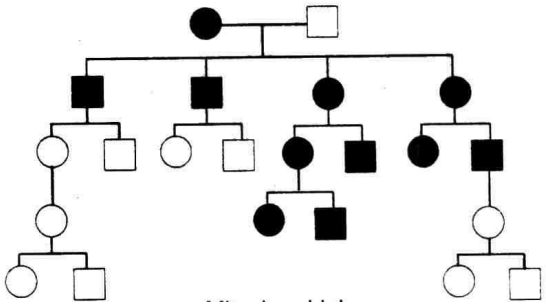
Patterns of single-gene inheritance

MITOCHONDRIAL DISORDERS

Maternal transmission
mtDNA mutation

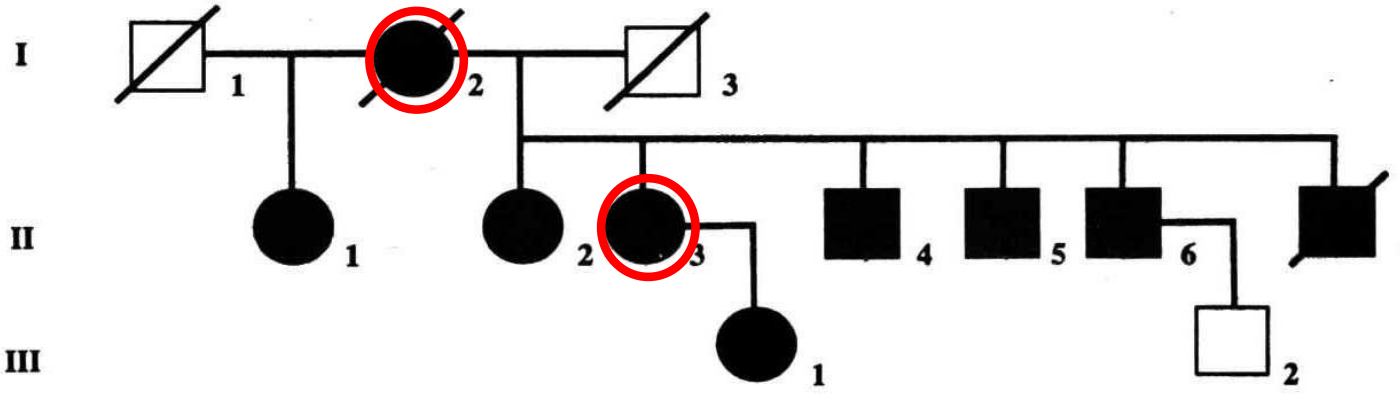


- 100 – 1000 mitoch/cell
- Few genomes/mitoch
- mtDNA: 16 kb
- Encode some components of mitochondrial respiratory chain complexes
- Variable phenotypes
  - Mostly neuro-muscular
- MATERNAL INHERITANCE
- Disease or susceptibility
  - Hearing loss ← aminosides



- Other components <= nucl. genes
- Respiratory chain defects: >1/10,000

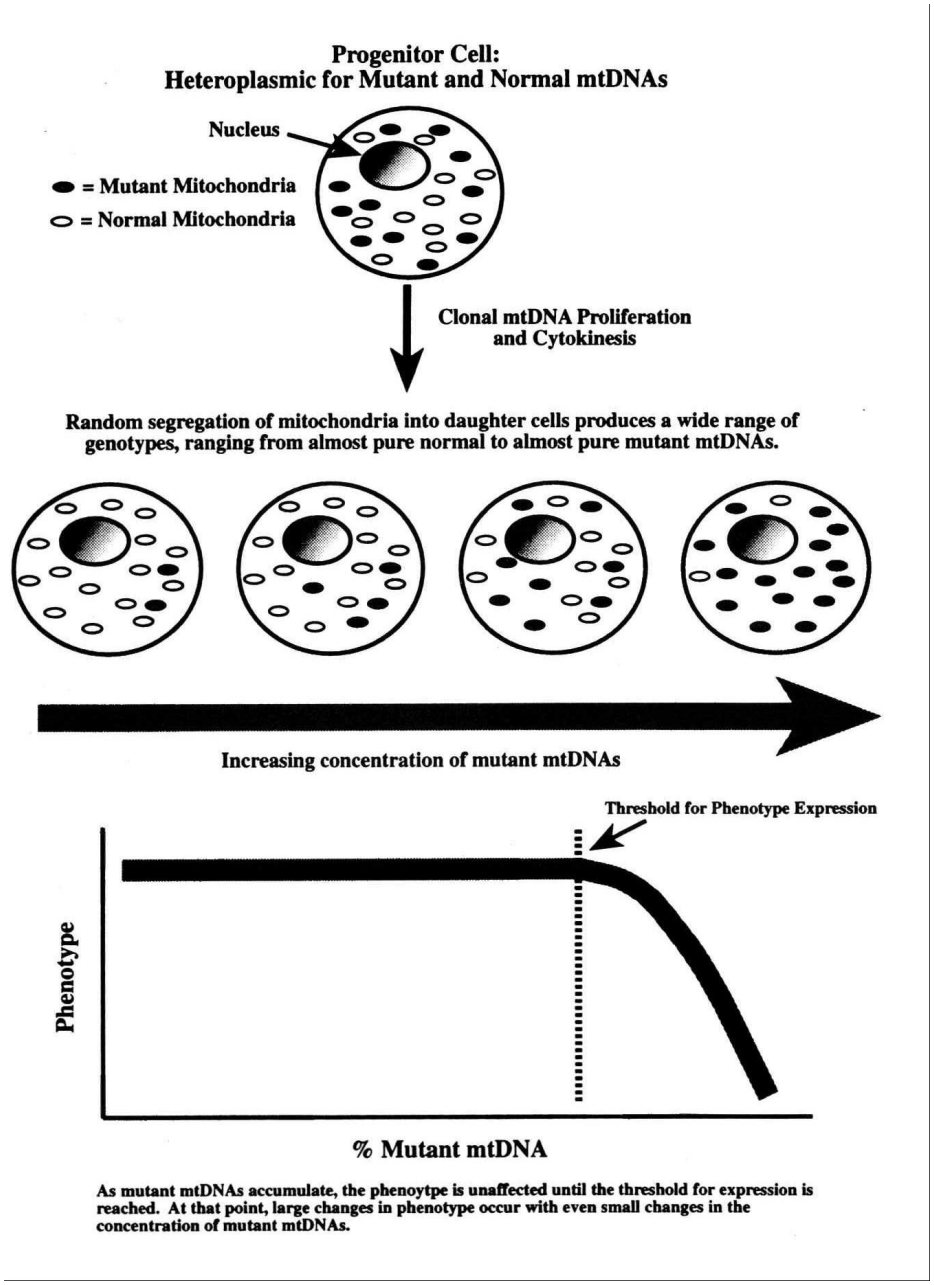
Variable expressivity in mt disease



	II-1	I-2	II-2	II-3	III-1	II-4	II-5	II-6	III-2	II-7
Age	53	-	51	49	25	46	45	-	29	-
%Deletion	39	ND	42	44	44	38	43	ND	0	ND
Insulin Dependent Diabetes Mellitus	+	+	+	+	+[1]	+	+	+	-	+
Diabetic Ketoacidosis	-	-	-	+	-	+	+	-	-	-
Deafness	+	+	+	+	+	+	+	+	-	+
Stroke	-	+	-	+	-	-	-	-	-	-
Ophthalmoplegia	-	-	-	-	-	-	-	-	-	-
Ptosis	-	-	-	-	-	-	-	-	-	-
Mitochondrial Myopathy	ND	ND	ND	-	ND	-	-	ND	ND	ND
OXPHOS Biochemistry	ND	ND	ND	+	ND	+	+	ND	ND	ND

[1] Diabetes mellitus has recently developed and the patient is currently managed with oral hypoglycemics. + = abnormal; - = normal.

Mitochondrial encephalomyopathy with lactic acidosis and stroke-like episodes

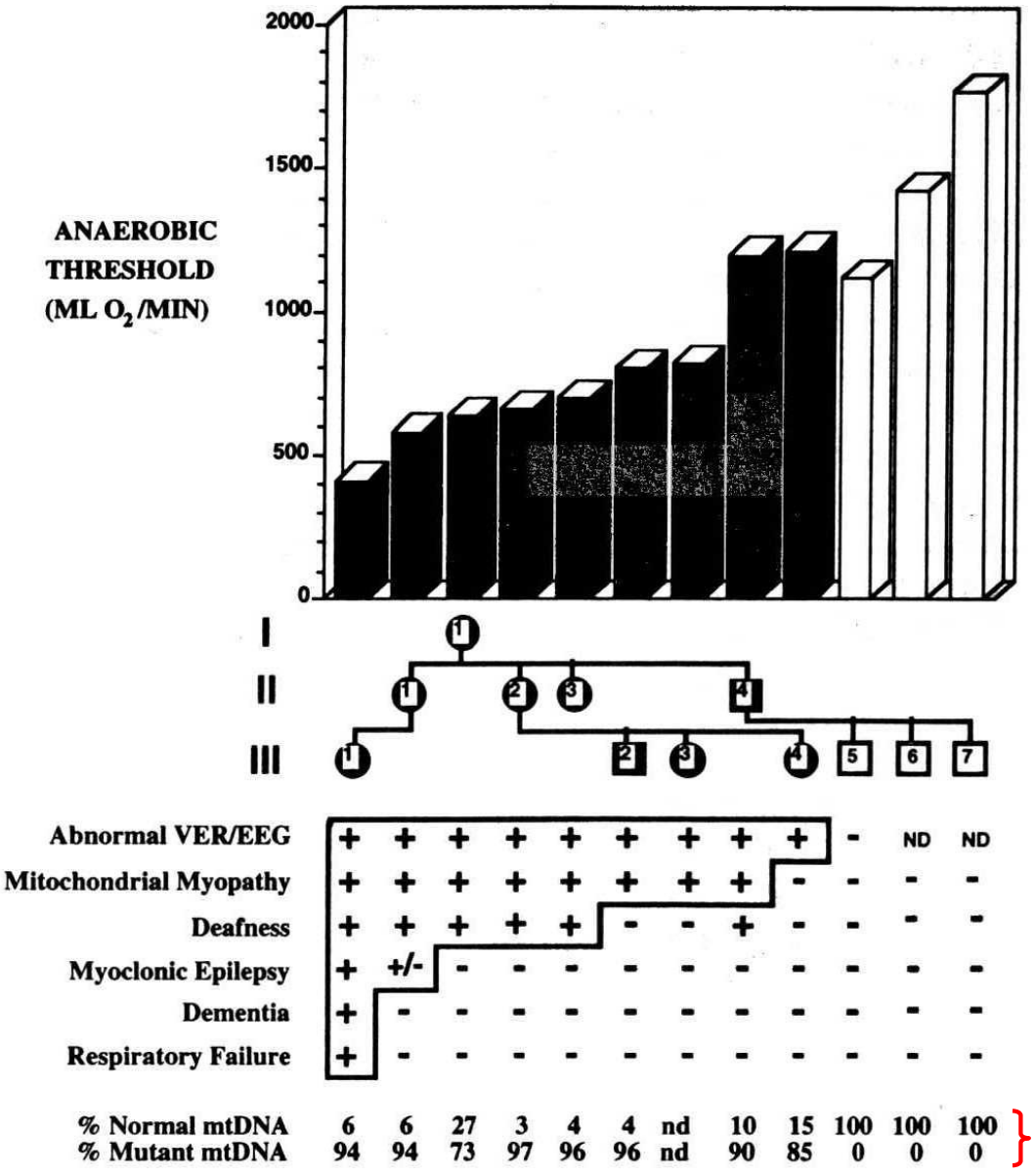


heteroplasmy

- Normal and mutated mtDNA mixed in one cell:
  « intracellular mosaic »

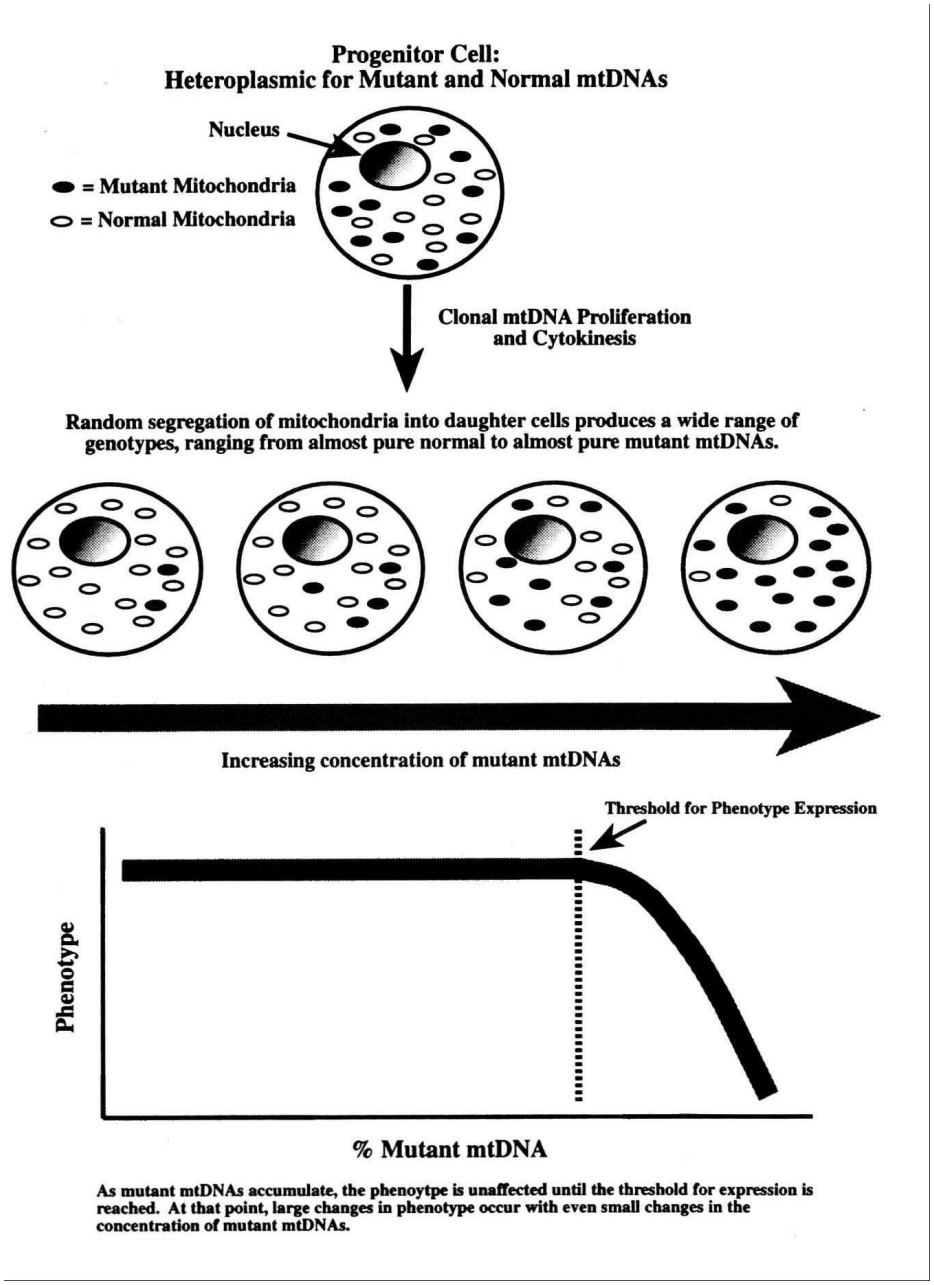
- TRESHOLD EFFECT

Heteroplasmy



Variable penetrance of clinical manifestations due to replicative segregation in MERRF.

- Normal and mutated mtDNA mixed in one cell: « intracellular mosaic »
- TRESHOLD EFFECT



Mitotic segregation

- Random distribution of mitoch to daughter cells
- Modifies heteroplasmy => modifies clinical course

Respiratory chain disorders

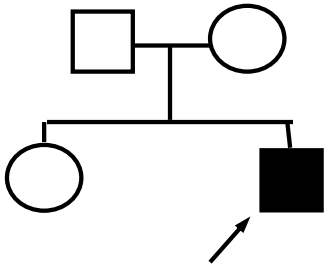
mtDNA

- Maternal inheritance
- Heteroplasmia, treshold effect
- Mitotic segregation

Nuclear DNA

- AD
- AR
- X

In small families, it may be difficult to identify mode of heredity

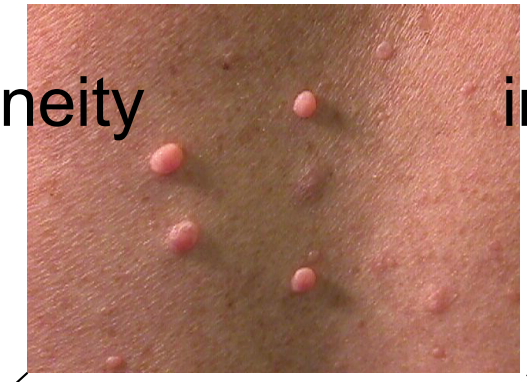


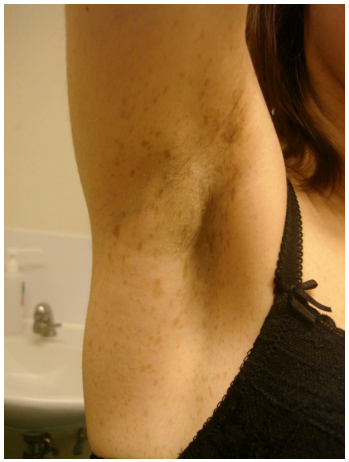
- Recessive?
- Dominant with low expressivity in affected parent ?
  — *(forme fruste)* ?
- Dominant with incomplete penetrance in carrier parent?
- Dominant, neomutation ?
- X chromosome-linked ?
- Multigenic/ Multifactorial ?
- Mitochondrial ?
- Non genetic ?

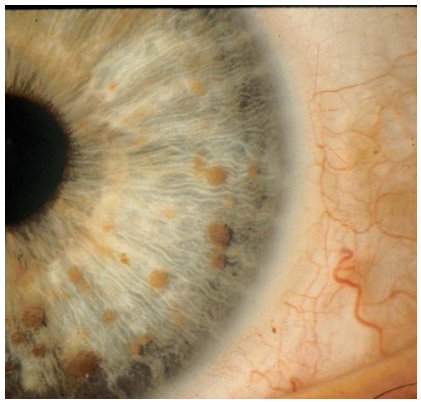
Locus heterogeneity

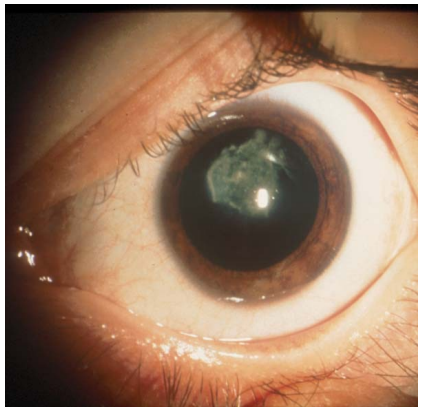
- MEN1 and MEN2
- NF1 and NF2; NF1-Like
- BRCA1 and BRCA2
- SCA1, SCA2, SCA3, SCA6, ….
- … and many disorders : identical phenotype, several loci

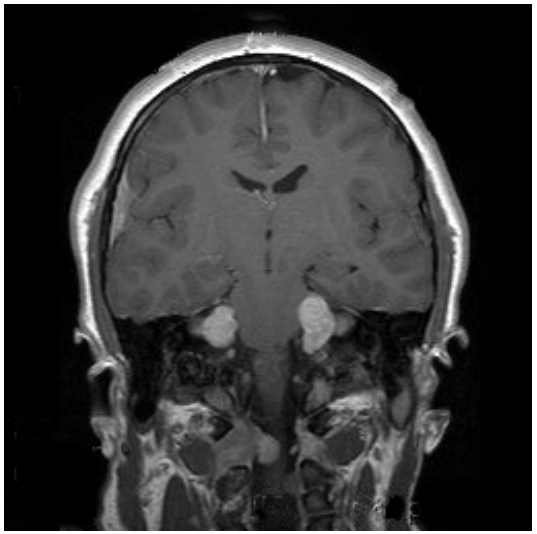
Genetic heterogeneity in Neurofibromatosis









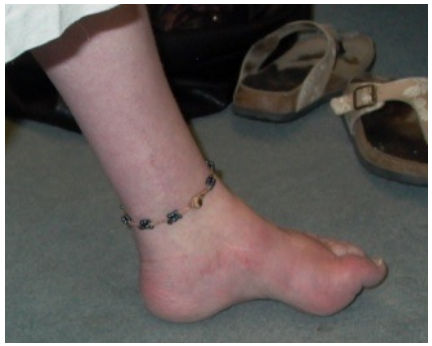


<u>NF1 (Von Recklinghausen) 1/3000</u>
Peripheral NF; >6 café-au-lait spots;
iris hamartomas (Lisch nodules);
freckling;
f/u: optic tract gliomas; PNST
NF1 gene, #17

<u>NF2 : 1/25000</u>
CentralNF; few café-au-lait spots;
VIII schwannomas;
f/u: meningiomas

NF2 gene, #22

Genetic heterogeneity in Charcot-Marie-Tooth disease = HSMN, Hereditary Sensory-Motor Neuropathy



CMT1 demyelinating

CMT1a PMP22 gene

CMT1b

CMT1c

CMT1n

…..

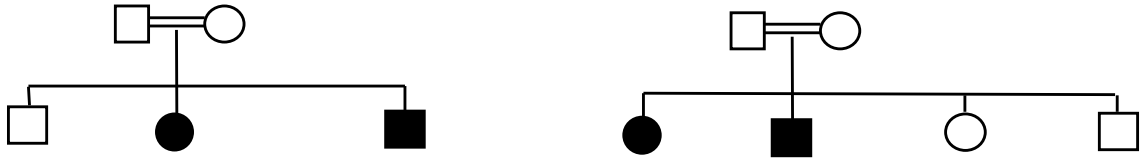
CMT2 axonal

CMT2n

…..

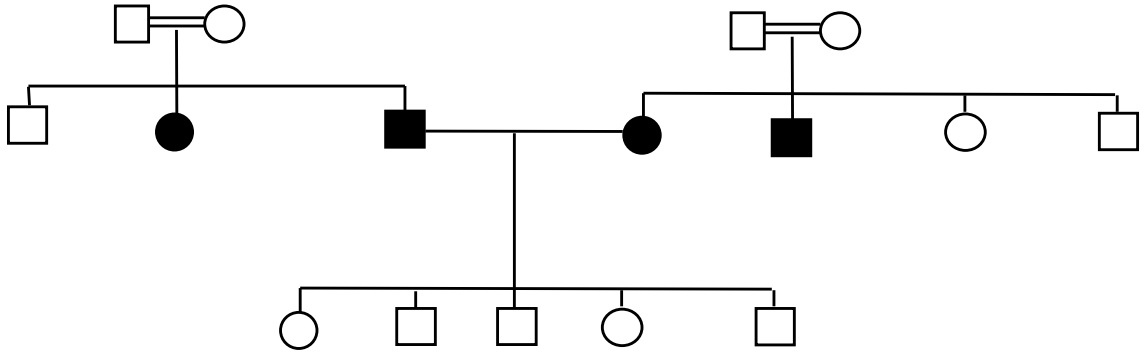
- CMT1 is heterogeneous beyond our ability to distinguish sub-phenotypes (>< NF1 and NF2)
- CMT2 also

Locus heterogeneity, AR disease



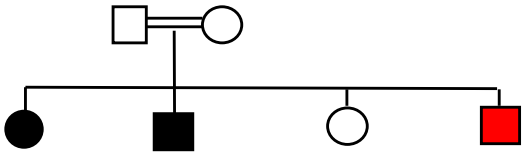
ex: deafness (> 100 db) prelingual onset

Locus heterogeneity, AR disease



ex: deafness (> 100 db) prelingual onset

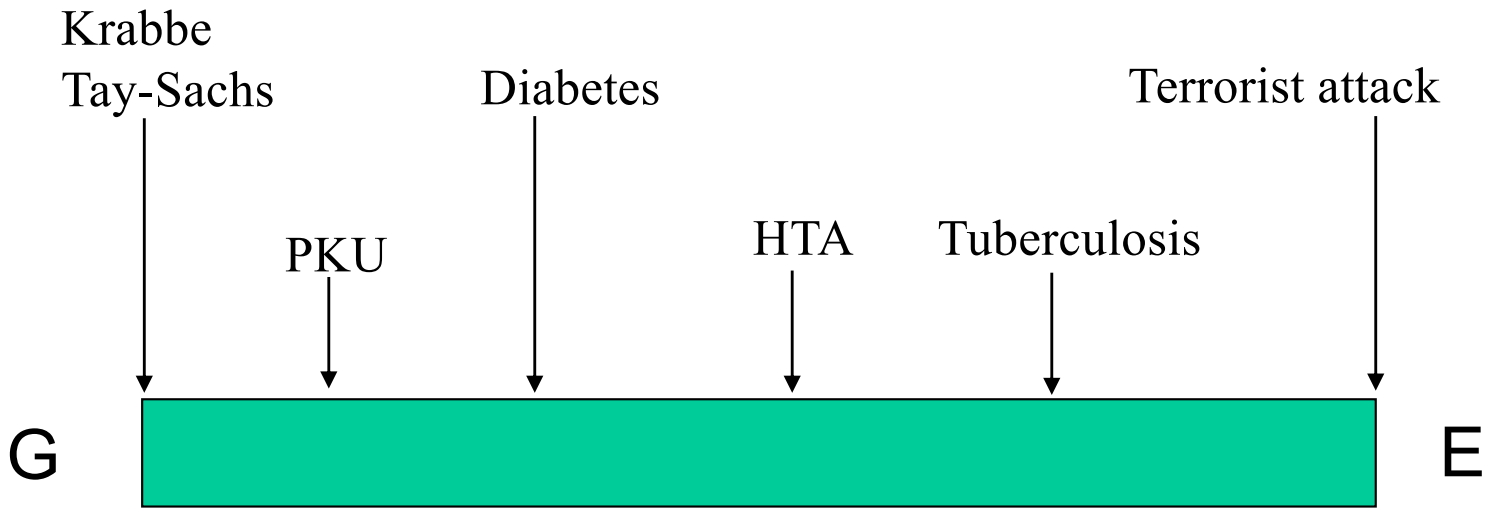
Phenocopy



Deafness, AR

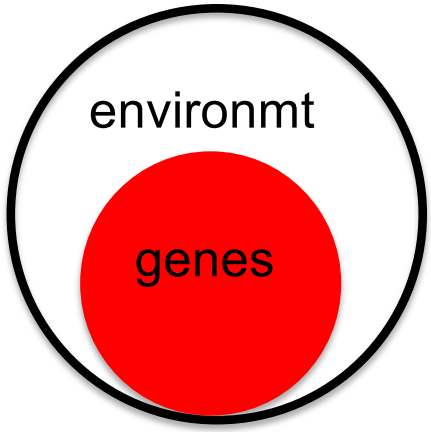
Deafness < fetal rubella

Gene – Environment interactions



Multifactorial characters

- Genetic contribution to phenotype is **COMPLEX** because it is due to many genes
- Genetic contribution to phenotype is globally **LIMITED** :
  global penetrance of the genome is incomplete

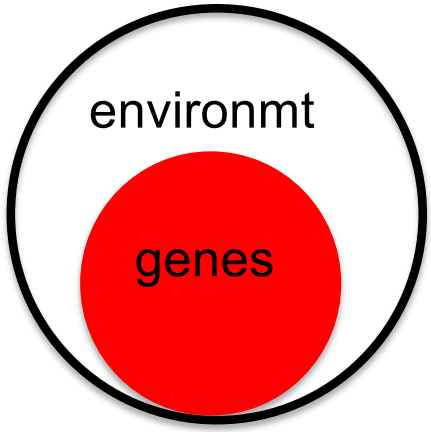


Multifactorial, complex

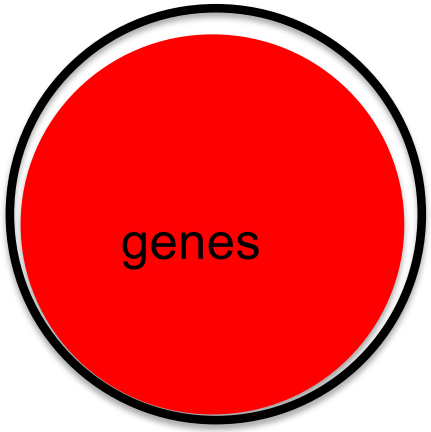
Mendelian, simple

Multifactorial characters

- Genetic contribution to phenotype is **COMPLEX** because it is due to many genes
- Genetic contribution to phenotype is globally **LIMITED** : ***globale penetrance of the genome*** is incomplete



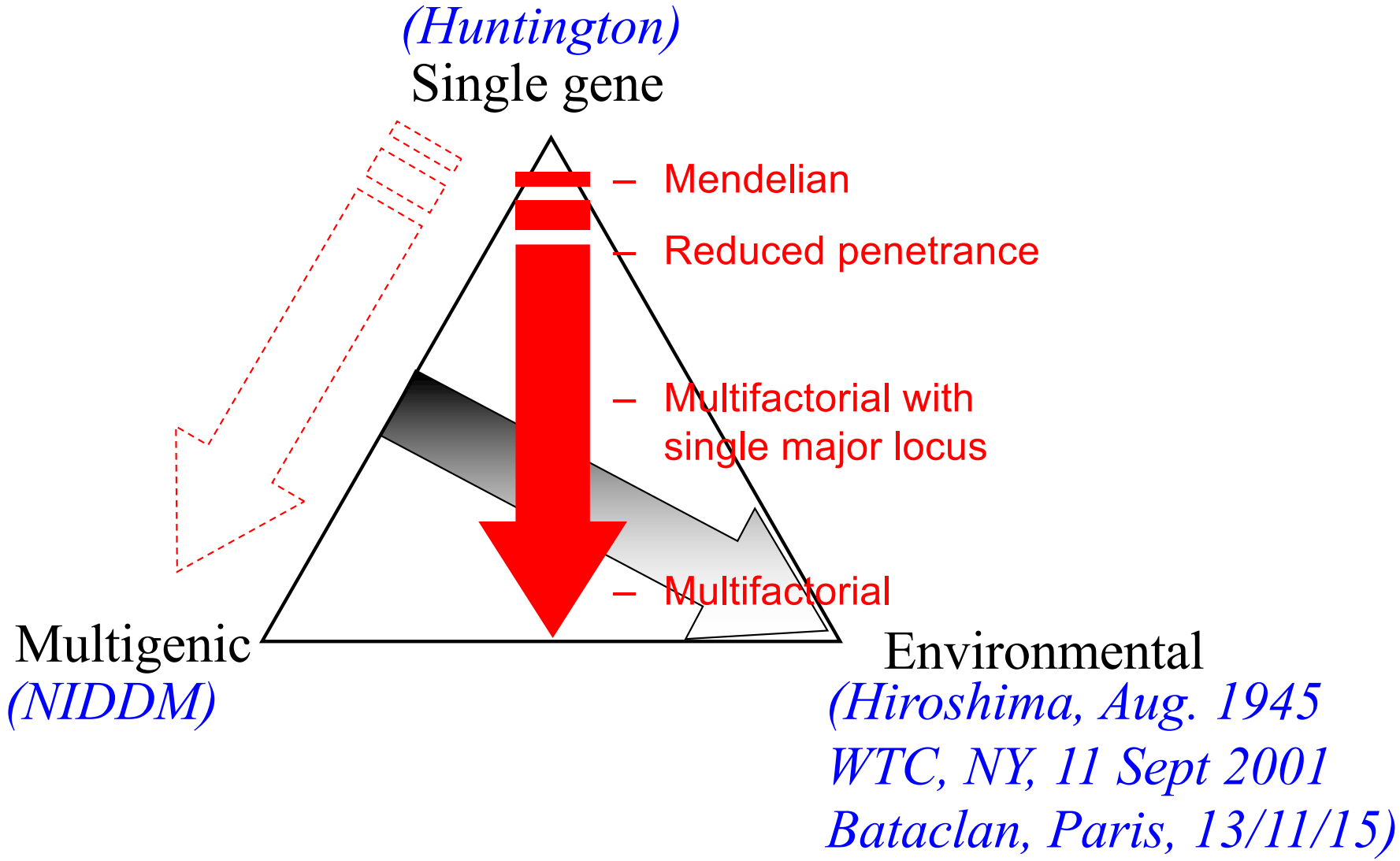
Multifactorial, complex



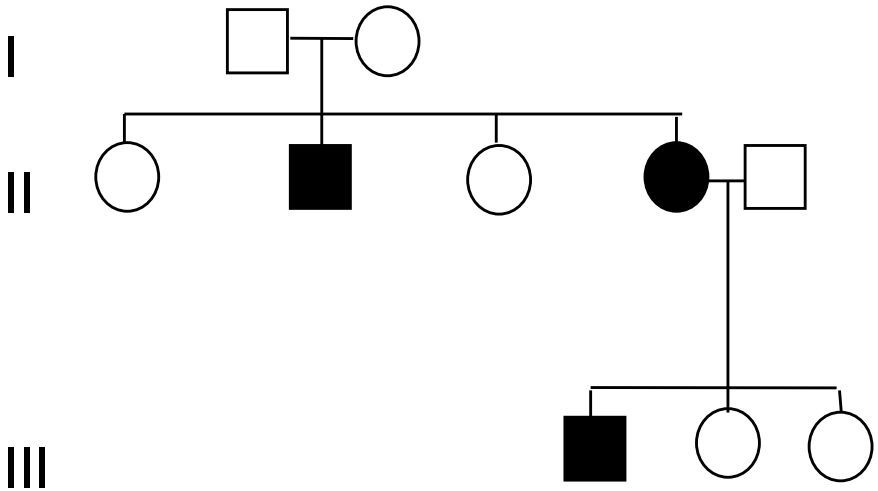
Multigenic, pure (exceptional)

Mendelian, simple

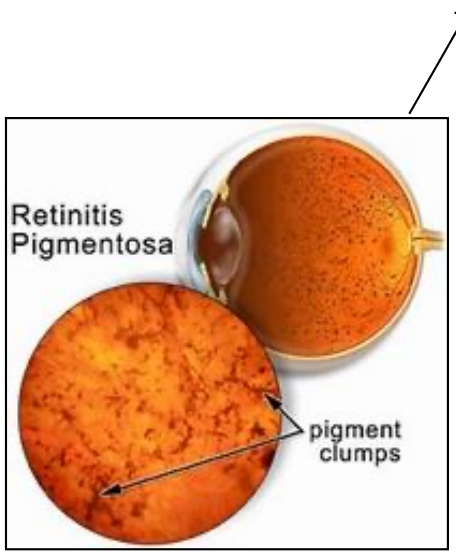
Gene-Gene and Gene-Environment interactions

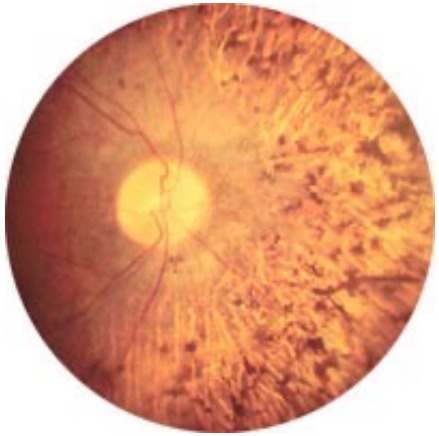


Complex inheritance in a case of pigmentary retinitis



Non mendelian: dominant?
Recessive? X-linked? …

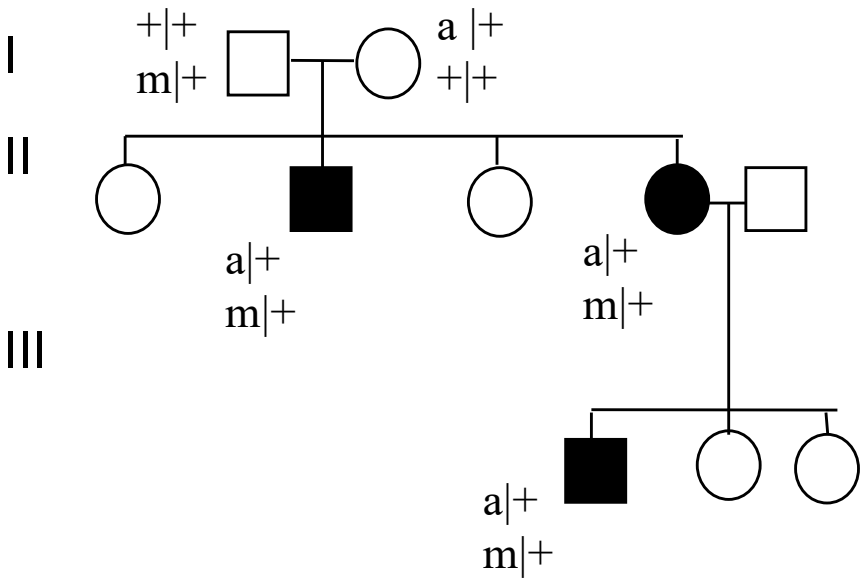




Tunnel vision

Digenic inheritance of a retinitis pigmentosa

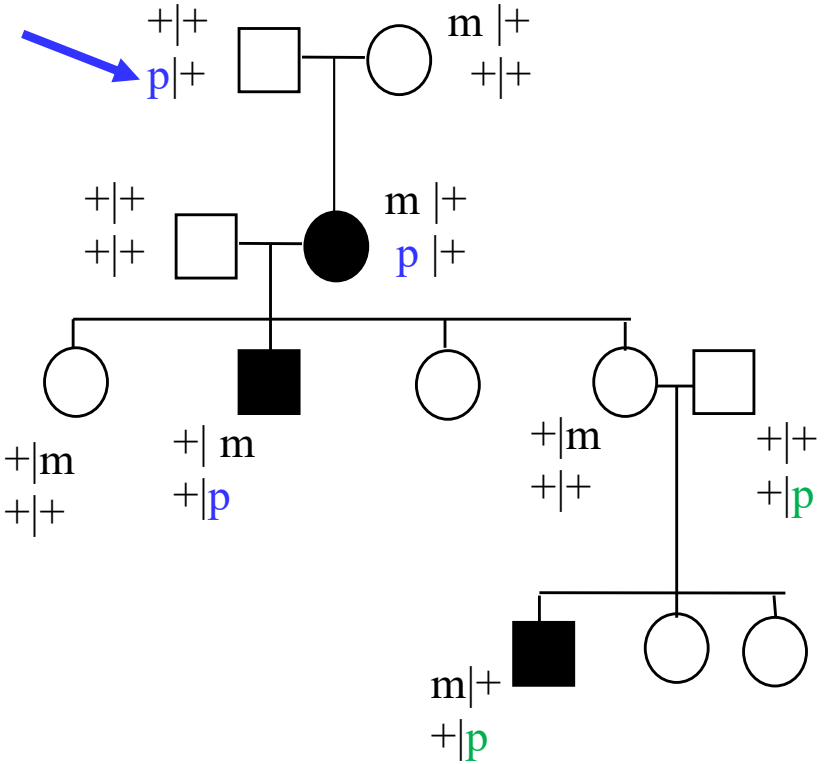
ex: mutations a and m
not genetically linked
(+ = wt allele)



Ex: Retinitis pigmentosa from mutations at 2 non-linked loci : peripherin/*RDS* and *ROM1*. (Kajiwara et al. Science 1994)

- Disease if **double heterozygote** (htz at 2 ≠ loci). Complete penetrance then.
  - Parents are simple heteroz and unaffected
  - 1/4 of children will be affected in génération II ; resembles AR
- Over next generations (génération III etc...) :
  - ¼ offspring will be affected: very different from AR
- a et m are 2 rare mutations, independently inherited

Digenic inheritance, with 1 rare mutation rare (m), 1 frequent polym (p)



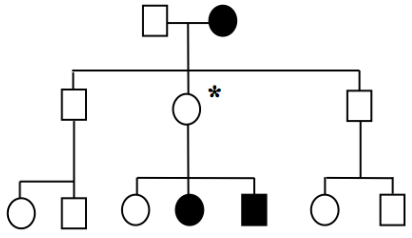
- AD with penetrance <1
- Penetrance = 1 if p present in everyone

Re-entry *of p in the pédigree*

Fqcy of allele m=1/10,000
Fqcy of allele p= 1/10

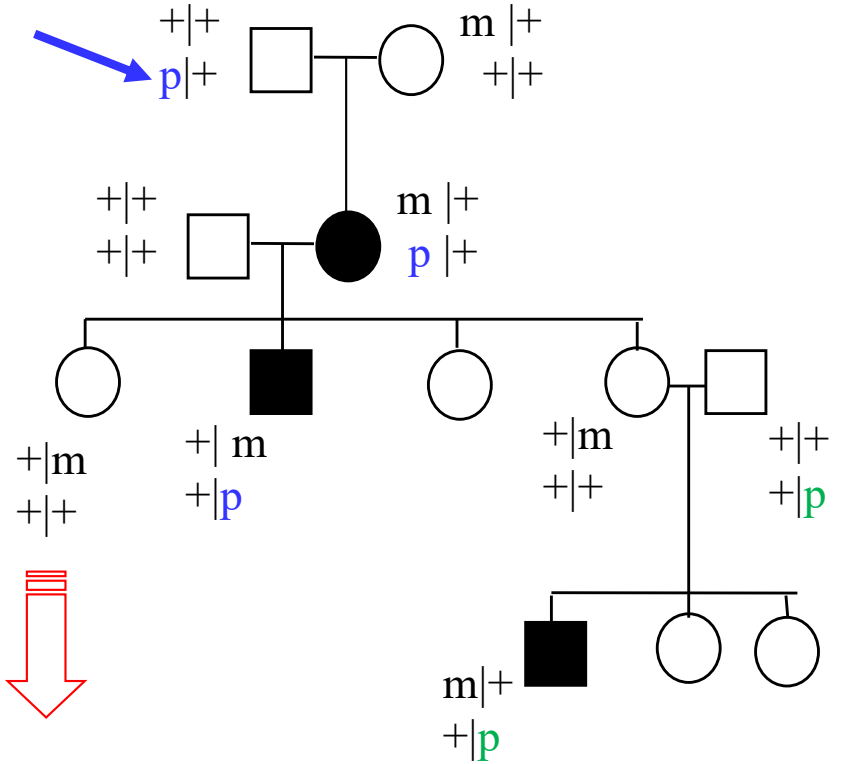
Incomplete penetrance

- **PENETRANCE**: % of mutation carriers who express phenotype



* « non-penetrant » subject :

Digenic inheritance, with 1 rare mutation rare (m), 1 frequent polym (p)



- AD with penetrance <1
- Pénétrance = 1 if p present in everyone

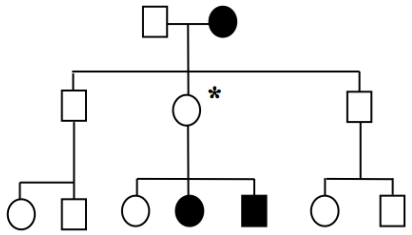
Re-entry *of p in the pédigree*

Migrates into island where all are p|p
=> AD , following transmission of m

Fqcy of allele m =1/10,000
Fqcy of allele p= 1

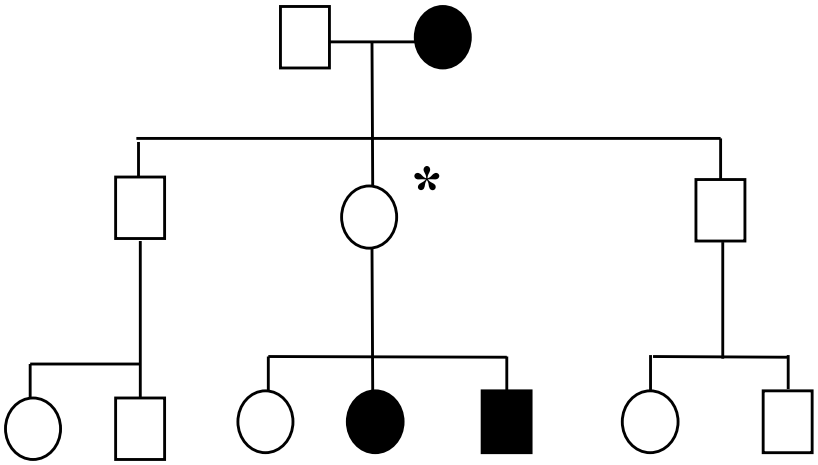
Incomplete penetrance

- **PENETRANCE**: % of mutation carriers who express phenotype



* « non-penetrant » subject :

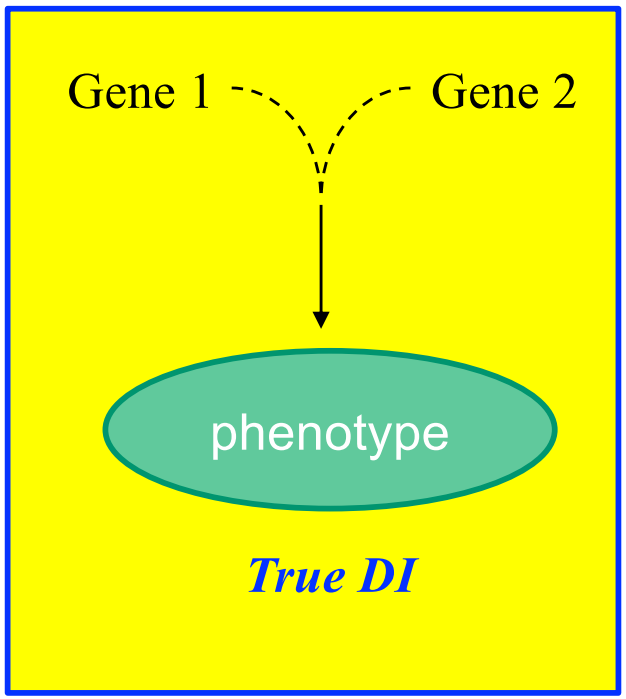
MAJOR GENE: necessary but not sufficient
Modifyer gene: modifies penetrance/expressivity



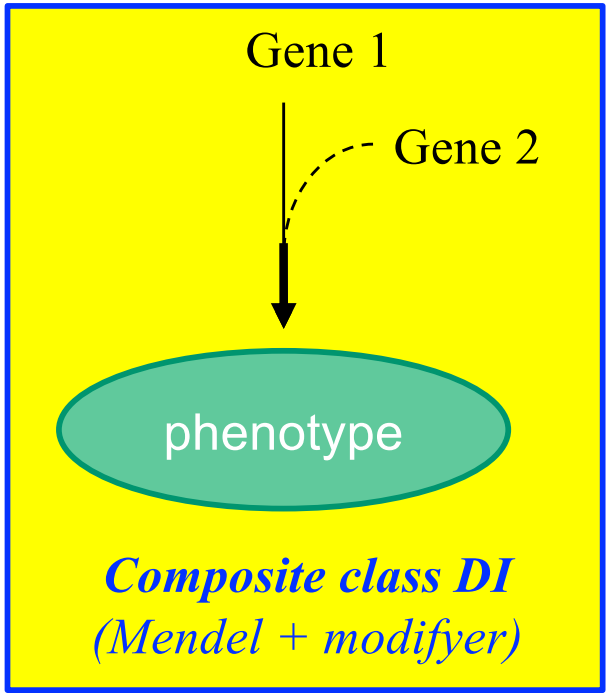
Incomplete penetrance:

* "non-penetrant » individual

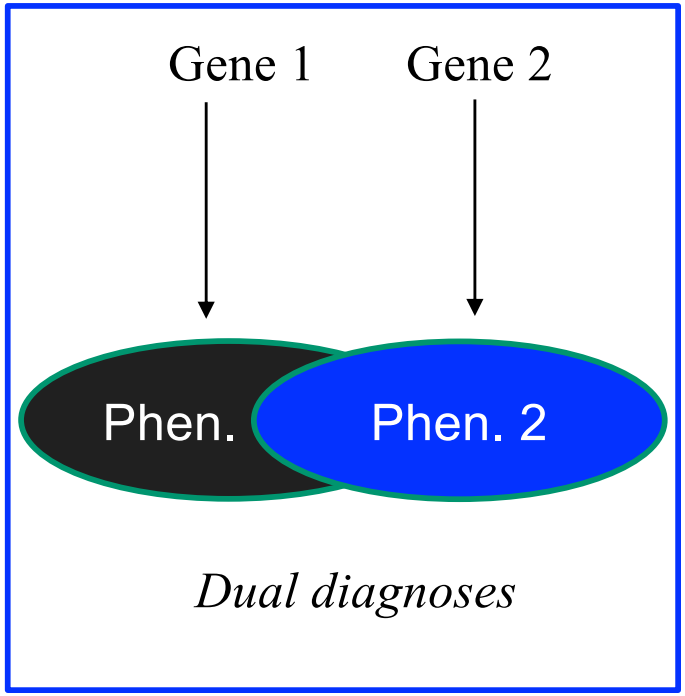
Digenic inheritance (bilocus)



Retinitis (Kajiwara et al. 1994); FSHD type 2 (Lemmers et al 2012); Midline Craniosynostosis (Timberlake et al., 2016) ;



Bardet-Biedl syndrome (Katsanis et al. 2001); Cystic Fibrosis (Dorfman et al. 2008);



RP + OA
CF + A1AT ddef

Functional interaction for synergy

- Protein – protein
- Protein – DNA
- Shared pathway

Clinical presentation

- Locus heterogeneity
- Reduced penetrance
- Variable expressivity

Lupski 2012; Schaffer 2013; Posey et al. 2017; Papadimitriou et al. 2019

Human genetic diversity:

MUTATION and POLYMORPHISM

Mutations and polymorphisms

- MINOR MUTATIONS = Polymorphisms

  Little or no functional effect on phenotype

  (Little or no penetrance)

- MILD MUTATIONS

- MAJOR (SEVERE) MUTATIONS High penetrance

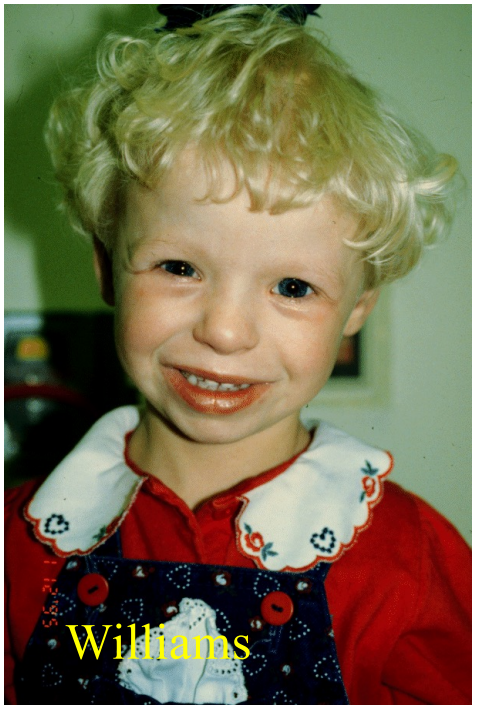
ALL are genetic VARIANTS = not wild type

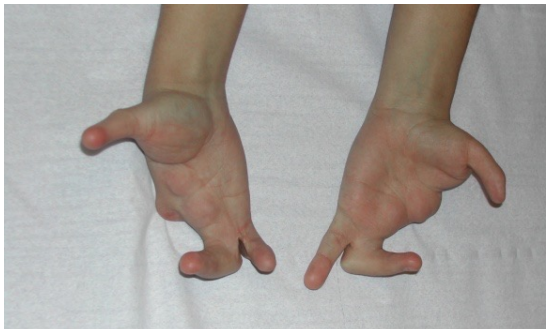
Mutations : tentative classification

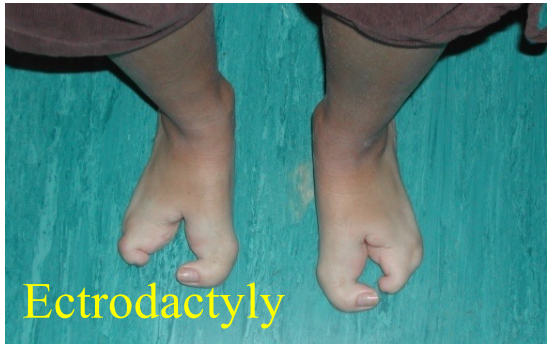
Class	Mechanism	Frequency	Examples
Genome mutation	Chromosome missegregation	>10% meioses	T21, other aneuploidies
Chromosome mutation	Chromosome rearrangement	1/1000 meioses	Microdeletion, translocation
Gene mutation	Base pair mutation	Varies with loci ~10^{-6} /locus /generation	Point mutations, indels

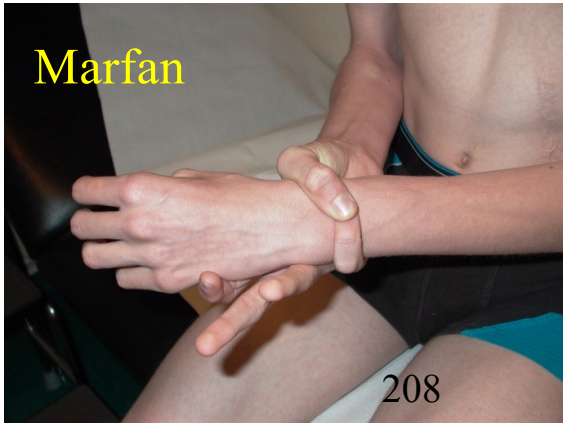
Most Genome mutations and Chromosome mutation affect survival and/or fertility => fitness ~0

Mutant phenotypes



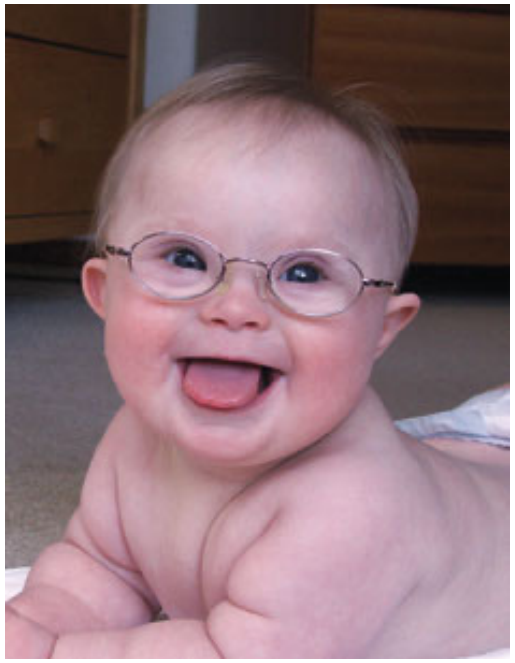




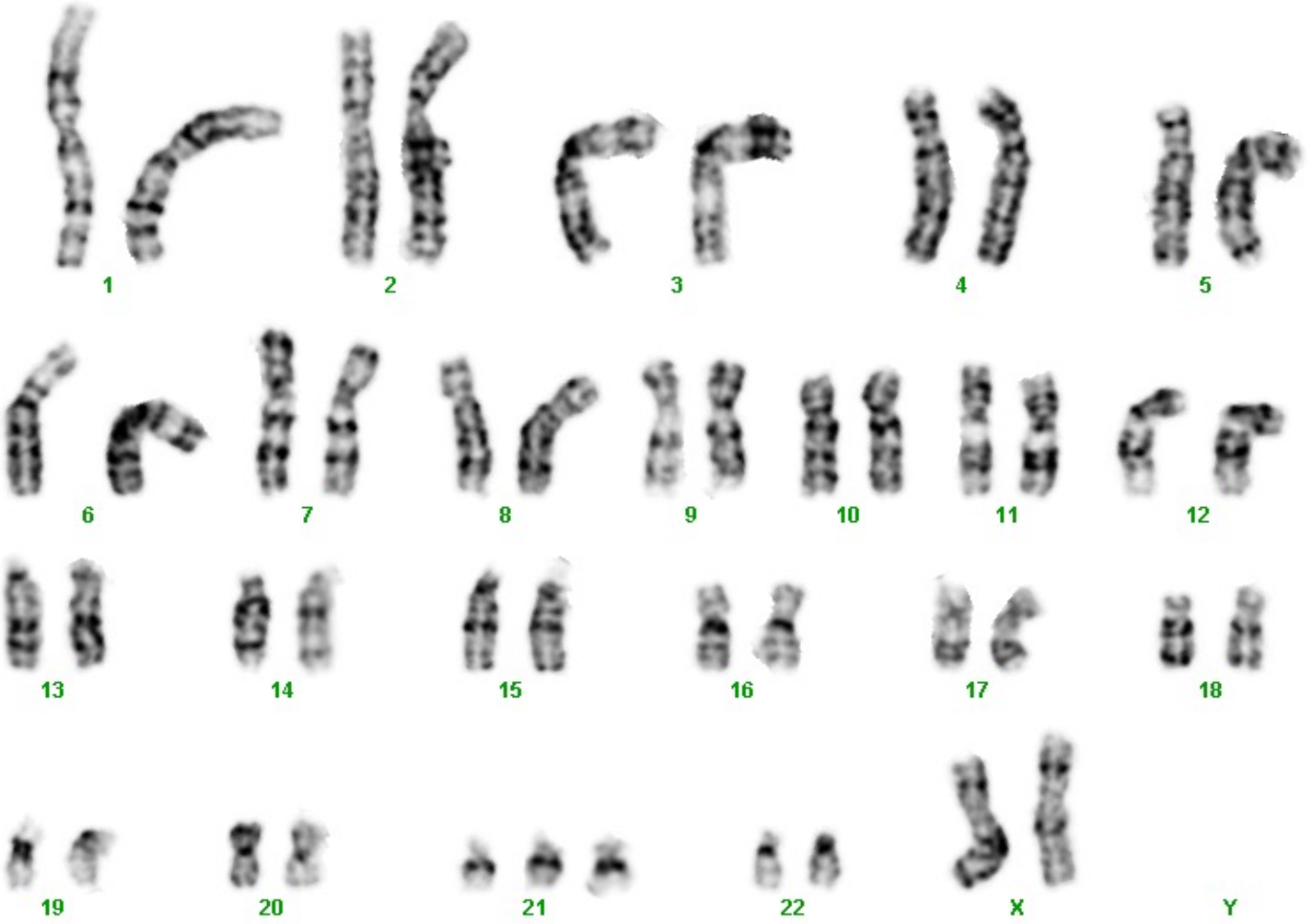


Genome mutation (T21) are *de novo* mutations

PHENOTYPE
= Down syndrome



GENOTYPE = T21



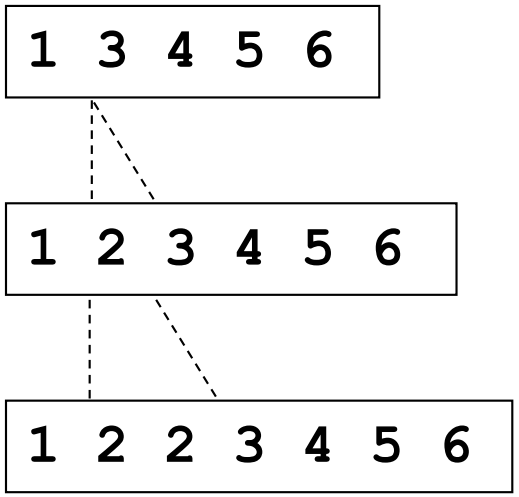
Mutations

Class	Mechanism	Frequency	Examples
Genome mutation	Chromosome missegregation	>10% meioses	T21, other aneuploidies
Chromosome mutation	Chromosome rearrangement	1/1000 meioses	Microdeletion, translocation
Gene mutation	Base pair mutation	Varies with loci ~10^{-6} /locus /generation	Point mutations, indels

- Most Genome mutations and Chromosome mutation affect survival and/or fertility => fitness ~0

Chromosome mutations (chromosomal rearrangements; structural anomalies)

INTRACHROMOSOMAL:



INTERSTITIAL DELETION

NORMAL

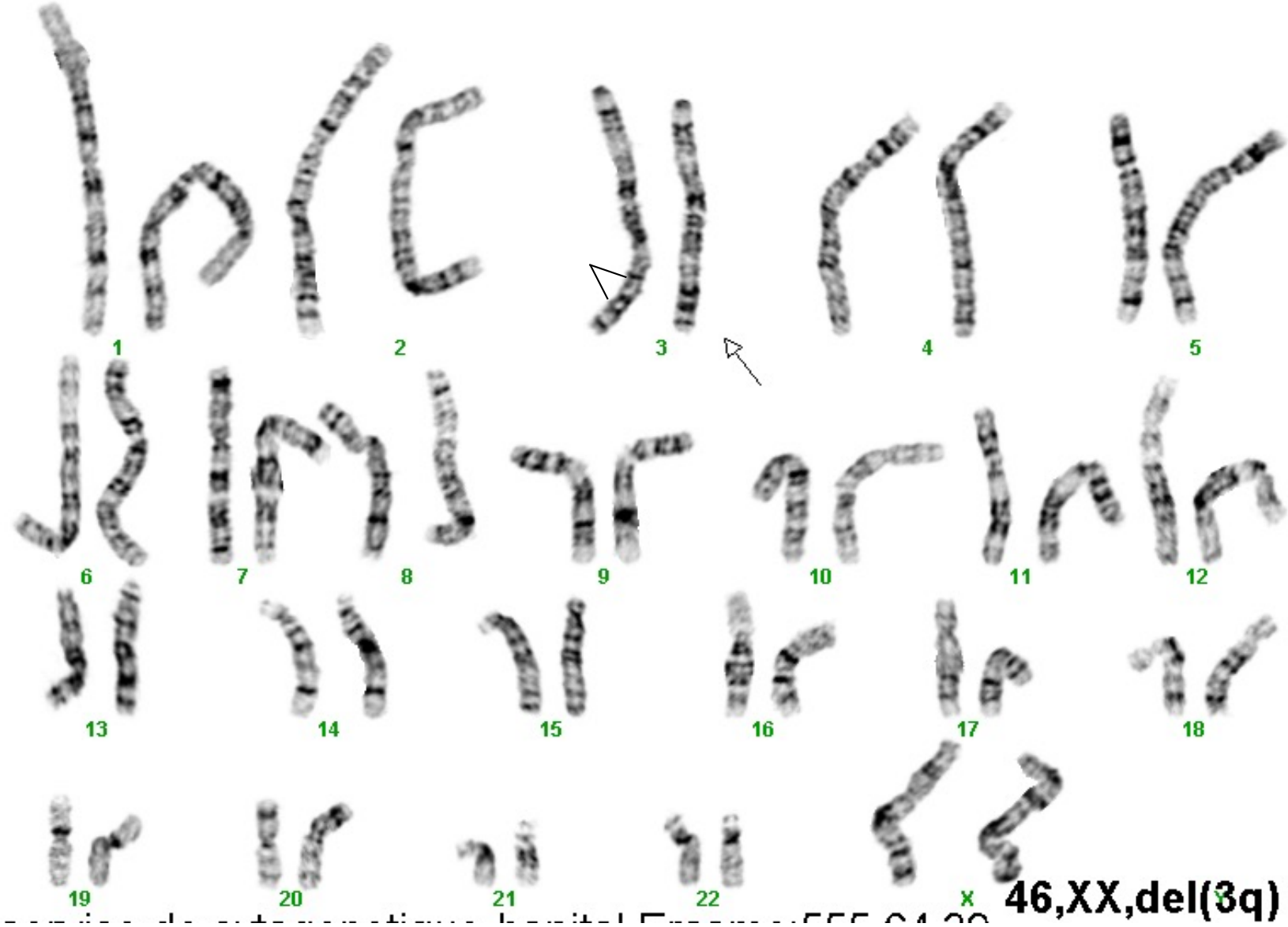
INTERSTITIAL DUPLICATION

=> Partial monosomies and trisomies (or tetrasomies…)
may be compatible with live birth

INTERCHROMOSOMAL

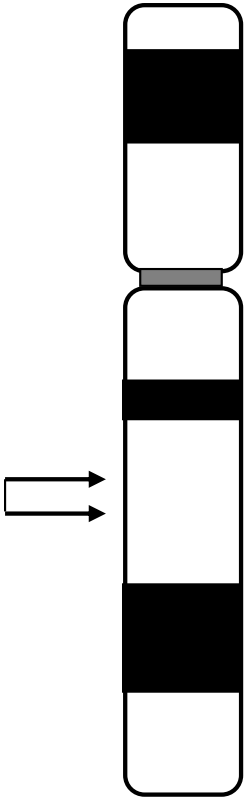
- translocations
- Inversions : pericentric, paracentric

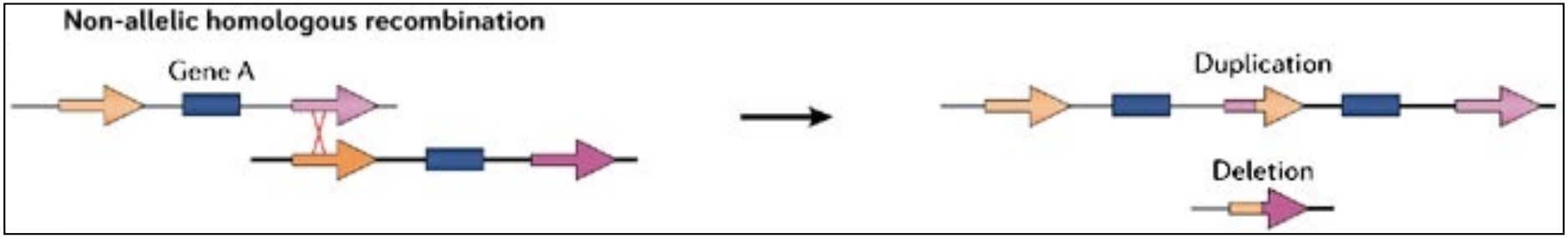
Interstitial deletion, long arm, # 3



Sub–microscopic chromosomal interstitial del (microdeletions)

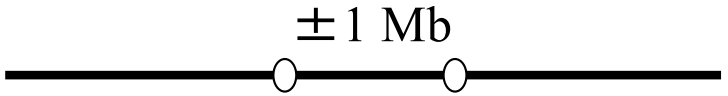
- Unseen on standard Karyotype
- May cause MCA+ID
  - Partial monosomy
  - Phenotype if involves haplo-insufficient genes
- Recurrent Microdeletions => known syndromes
- Single gene or contiguous genes
- Same for microduplication.



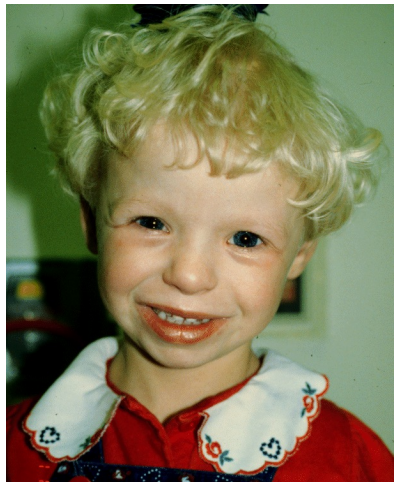


Recurrent microdeletions => Known syndromes

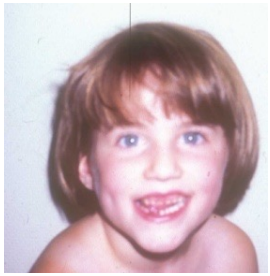
- Breakpoint hot-spots
  (interspersed repeats)
  – 1 meiose / 10,000



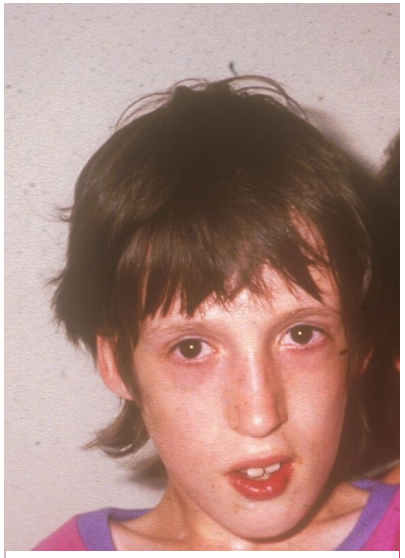
- Known Syndromes
  – **Williams** syndrome (7q)
  – **DiGeorge** syndrome (22q)
  – Prader-Willi / **Angleman**
  (15q, imprinting)
  – …et autres…



St. Ao, ↑Ca^{++}…



Ataxia, laughter,…



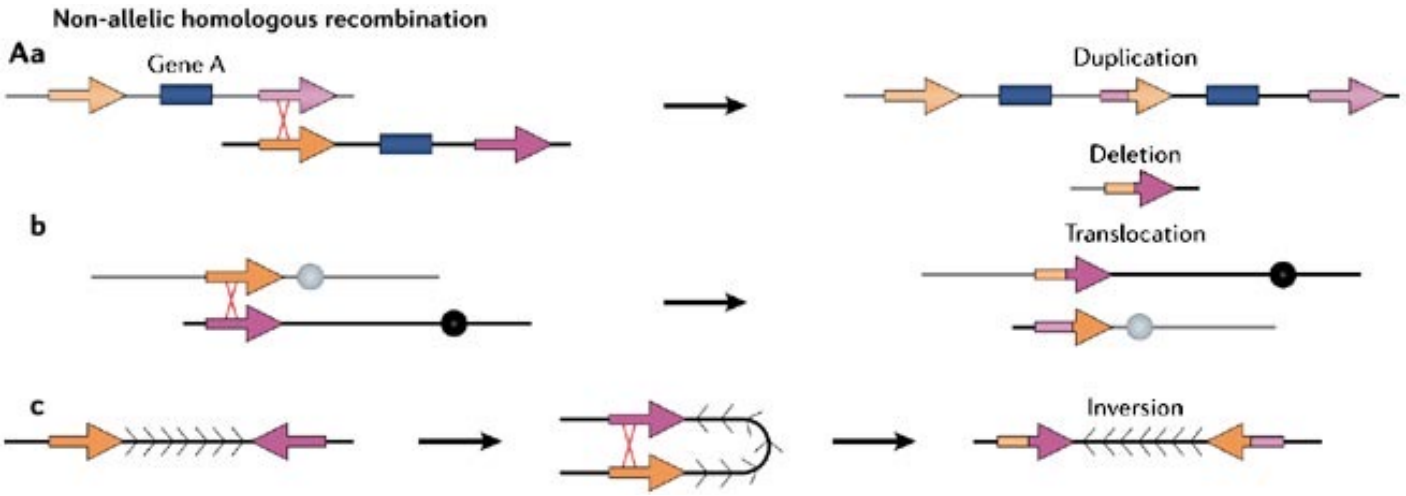
Lymphopenia T,
hypoPTH,…

Non Allelic Homologous Recombination (NAHR)

eg, Deletion PMP22 => HNPP
Duplication PMP22 => CMT1a

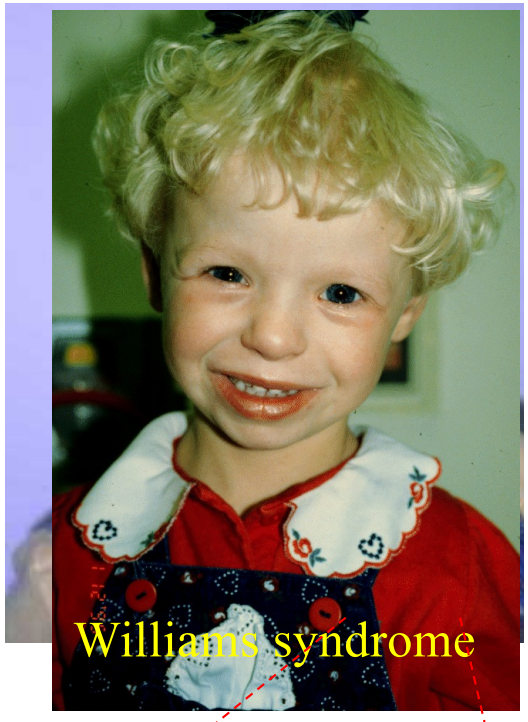
Bailey & Eichler

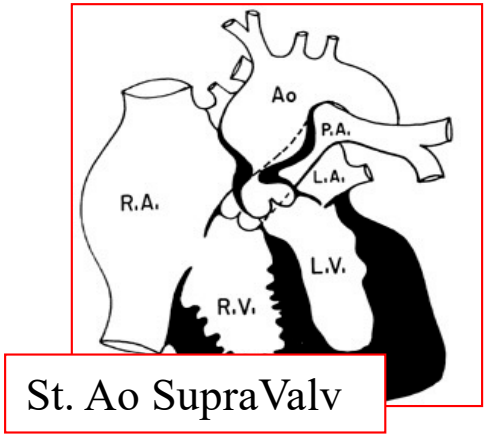
NAHR



Bailey & Eichler

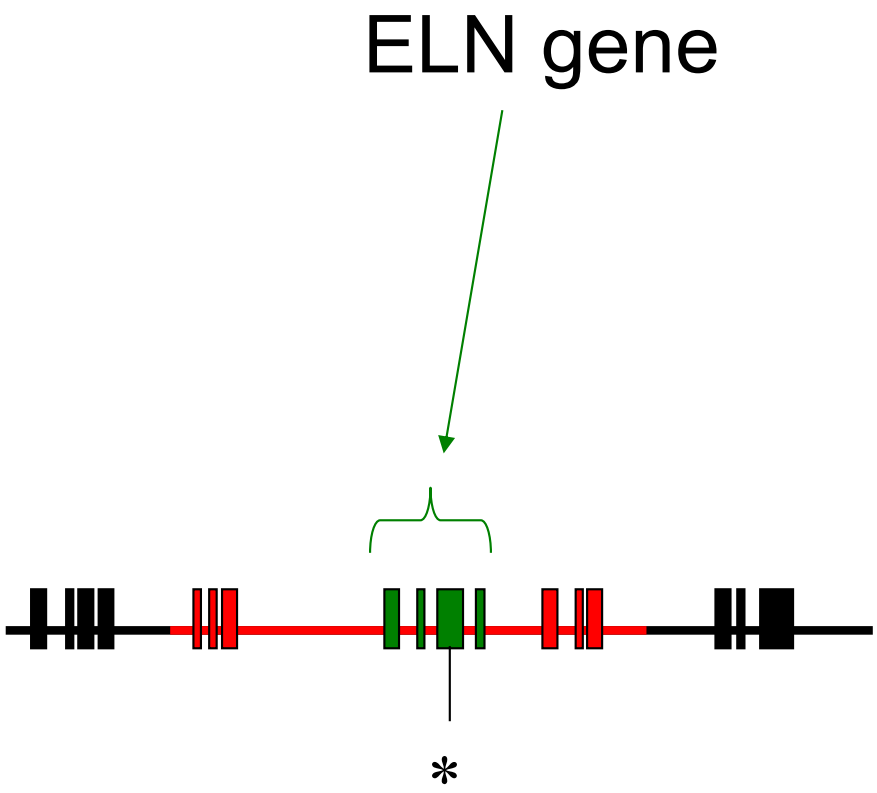
Williams syndrome

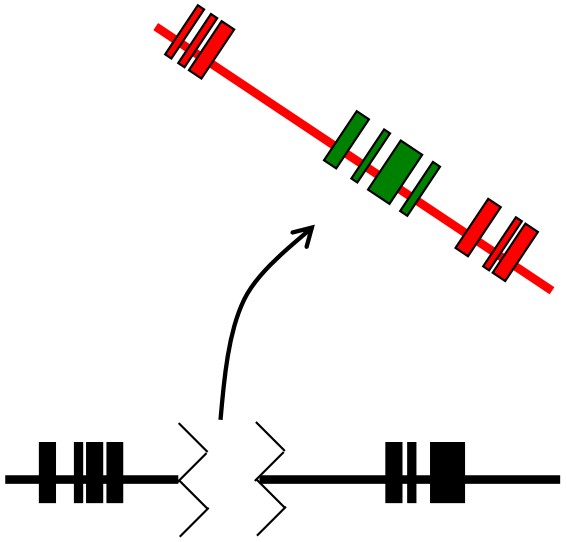




St. Ao SupraValv

- Microdeletion
  elastin gene + contiguous genes
- Supravalvular Ao stenosis
  - Isolated form(non syndromic) of SupraValv Ao St is also described; familial; AD; inactivating mutations of the elastin gene
- Dysmorphism
- ID, talkative, music gifted, friendly
  - Gene mechanism = ?
- Hypersensitivity to VitD
  - > early diagnosis allows prevention of hyperCa^{++}
- Microdeletion most often de novo (neomutation); rarely familial, AD.





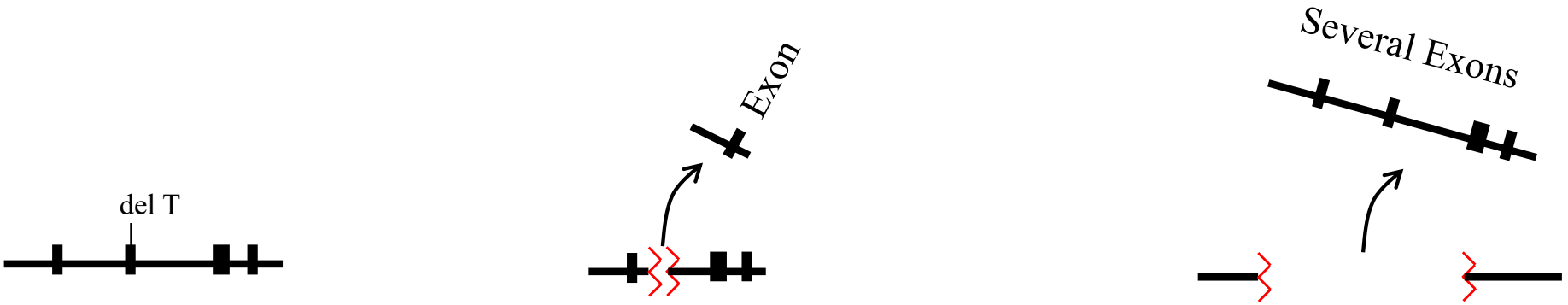
Point mutation
=> St Ao , isolated
AD transmission

Null mutation

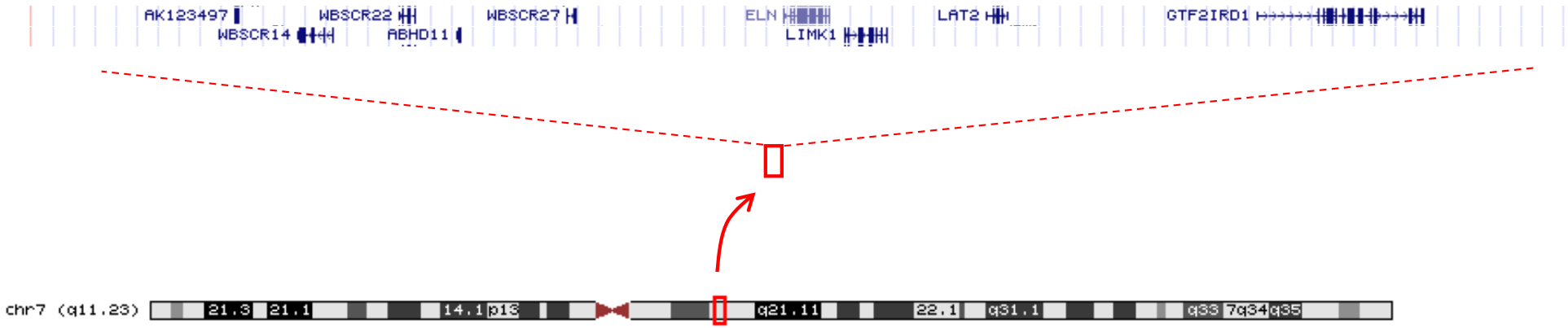
Deletion ELN + **contiguous genes**
=> St Ao + malfo + MR
+ dysmorphism

confirms haploinsufficiency

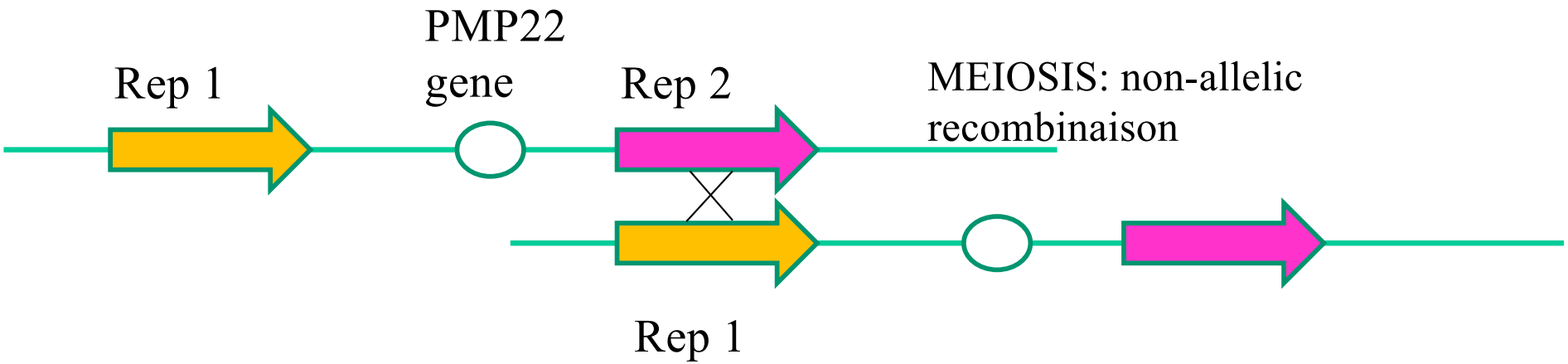
Variable size of mutations

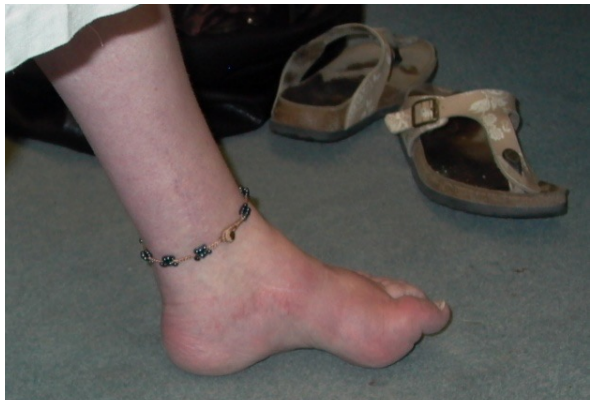


Several genes



Recurrent interstitial microduplication (1/10000 meioses), involving a single gene (PMP22, chr.17)





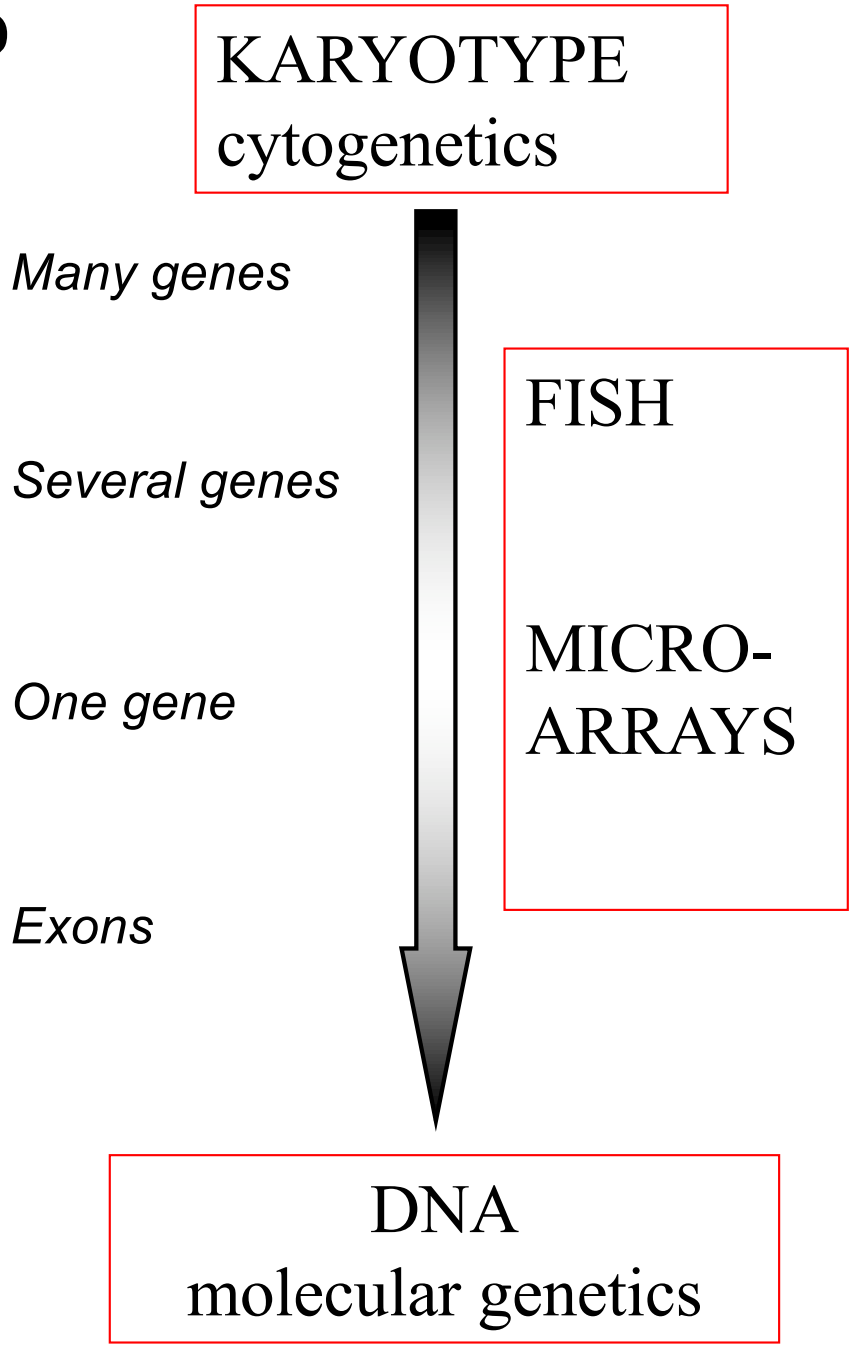
- PMP22 => péripheral myelin protein
- Duplication => 3 doses
- Causes polyneuropathy

✓ Structurally : chromosomal mutation, microduplication
✓ Fonctionally: single gene involved => simple, mendelian inheritance, AD

Williams is also AD !! But syndromic, and low fitness, >95% neomutations

From chromosome to nucleotide

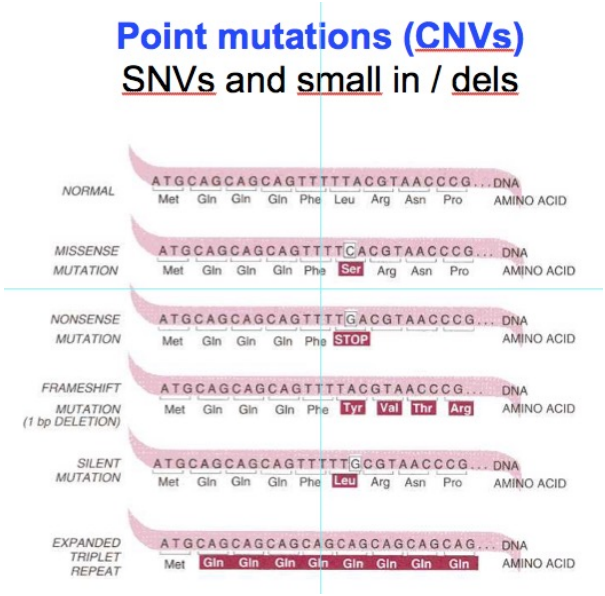
- Many syndromes MCA+MR
- Most hereditary diseases



Various mutations currently need different **methods**

Point mutation

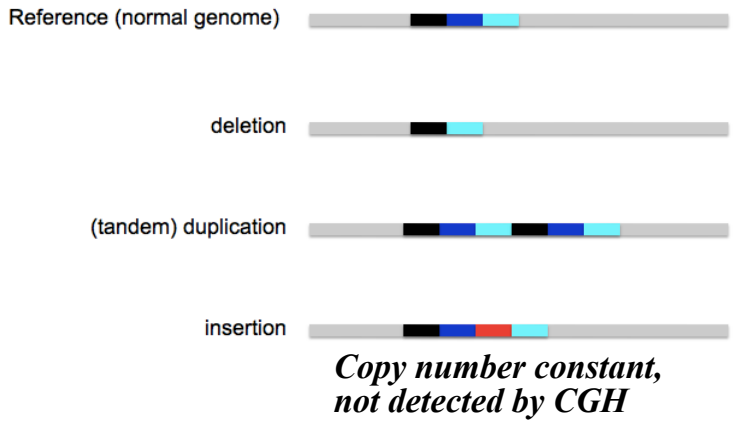
Point mutations (CNVs)
SNVs and small in / dels



DNA sequencing

Chromosomal mutation

Copy Number Variants (CNVs)
>1kb ; some large enough for cytogenetics
and other chr. rearrangements



Copy number constant, not detected by CGH

CGH array or SNP array

Genome mutation

Aneuploidies

47,XX or XY, +21 (T21)
47,XX or XY, +18 (T18)
47,XX or XY, +13 (T13)
47, XXX ; 47, XXY; 47,XYY
45,X

chromosomal translocations

Karyotype (cytogenetics)

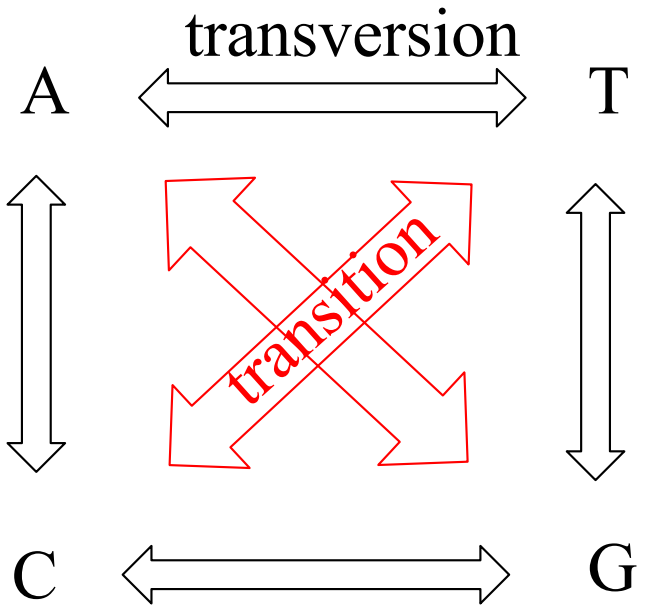
In the future, all types of mutation will be amenable to whole genome sequencing

Mutations

Class	Mechanism	Frequency	Examples
Genome mutation	Chromosome missegregation	>10% meioses	T21, other aneuploidies
Chromosome mutation	Chromosome rearrangement	1/1000 meioses	Microdeletion, translocation
Gene mutation	Base pair mutation	Varies with loci ~10^{-6} /locus /generation	Point mutations, indels

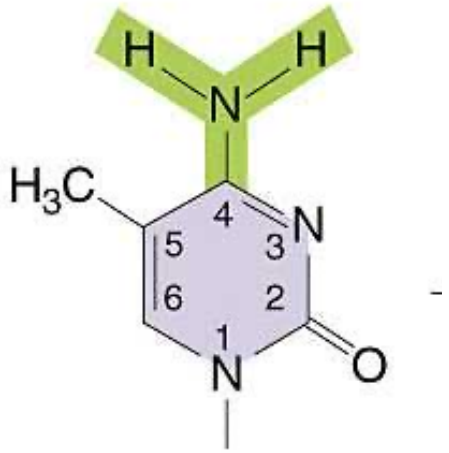
- Most Genome mutations and Chromosome mutation affect survival and/or fertility => fitness ~0

Single-nucleotide point mutations



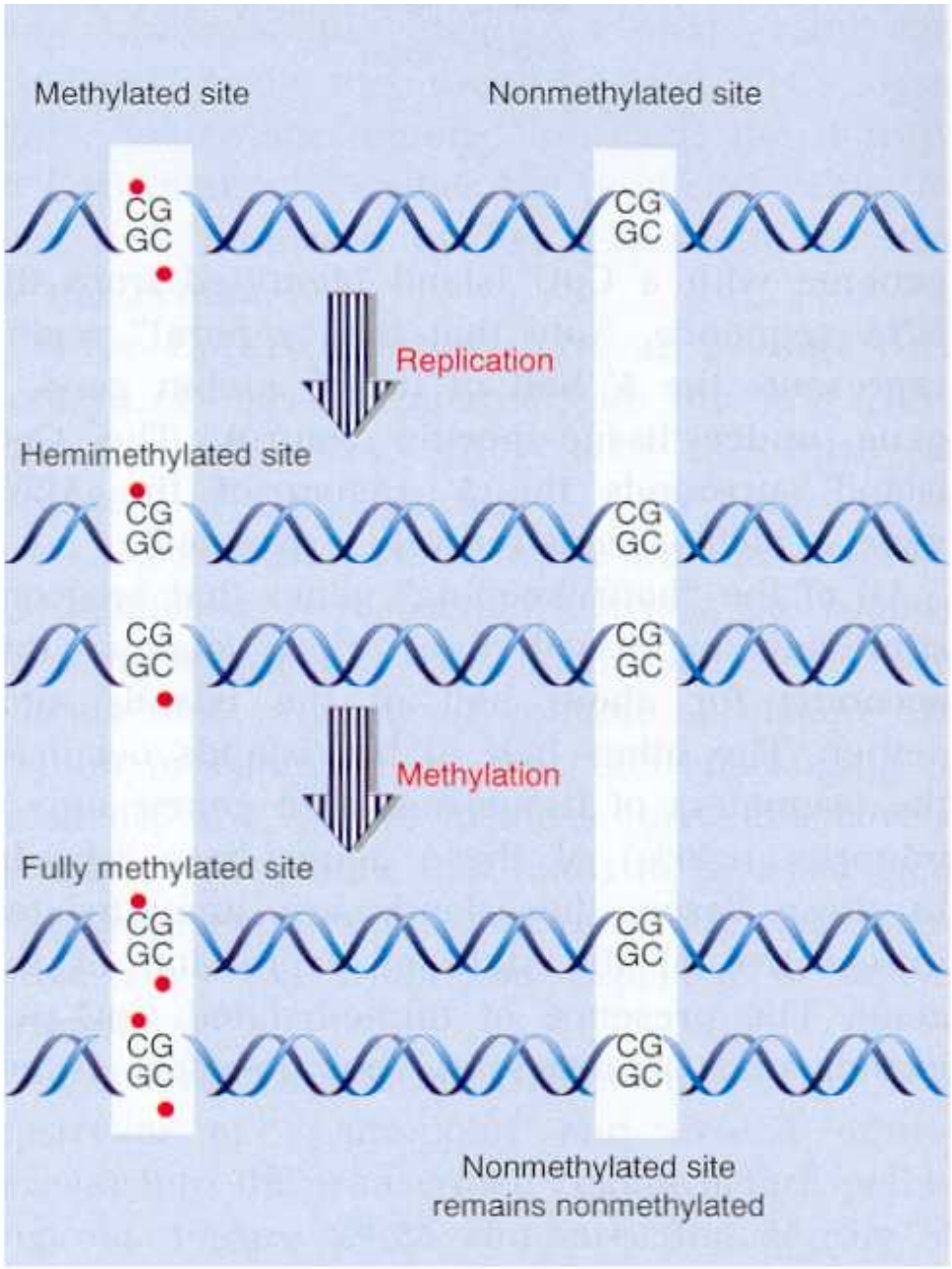
- Transition: purine to purine
- Transversion: pu / py
- Expect 2 x more transversions
- In fact transitions are more frequent
- Most frequent is C>T (G>A) in 5’ CpG 3’ dinucleotide

DNA (hemi-) methylation

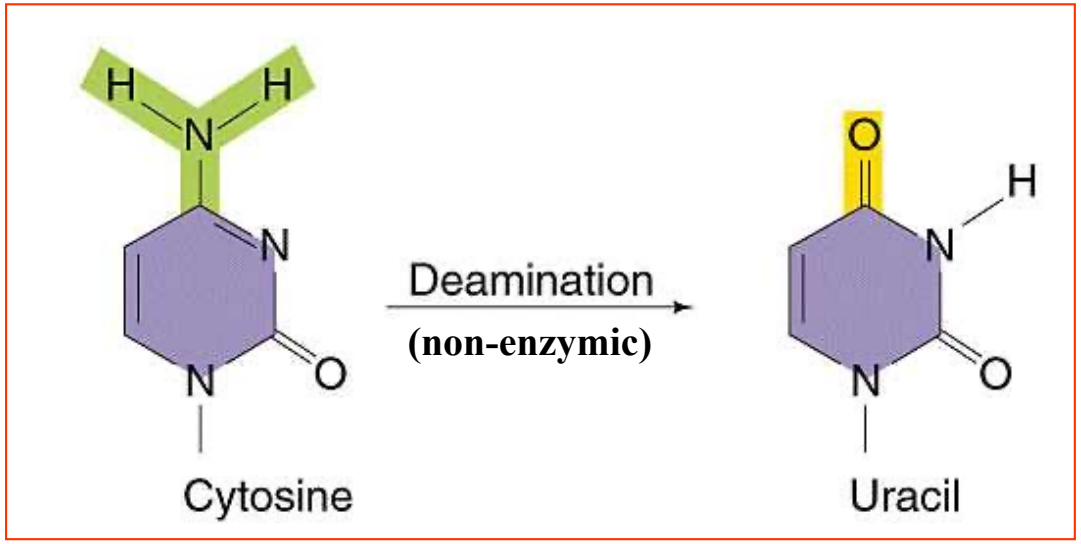


5-Methylcytosine

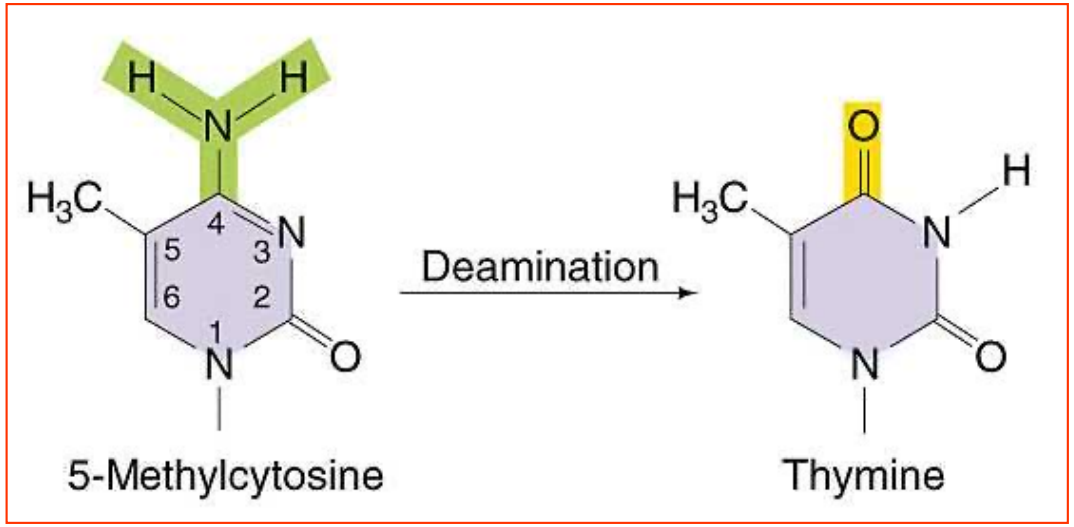
- ^{met}C = 5^{th} base of DNA
- Methylated promoters inactive (usually)
- Implication in gene silencing; imprinting.
- Epigenetic heredity



^{met}Cytosine occasional deamination



- In DNA, uracil recognized as abnormal
  => mutation corrected



- Thymine not recognized as abnormal
  => mutation remains

MUTATIONS IN CODING SEQUENCES

Point mutations

Affect 1 or a few base pairs

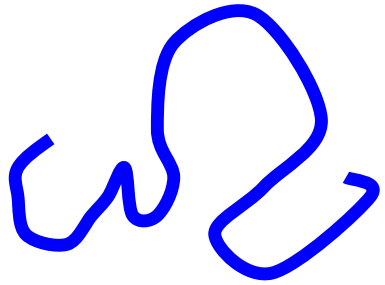
MUTATIONS IN CODING SEQUENCES

Point mutations

Affect 1 or a few base pairs

Nonsense mutation generally causes null allele

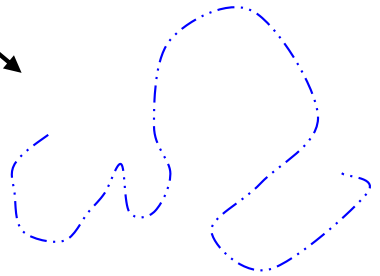
Wt allele > mRNA > normal protein:



Nonsense Mutation > Premature Termination Codon (PTC)

OR

Truncated protein product



Unstable mRNA > decay

Usually complete loss of function

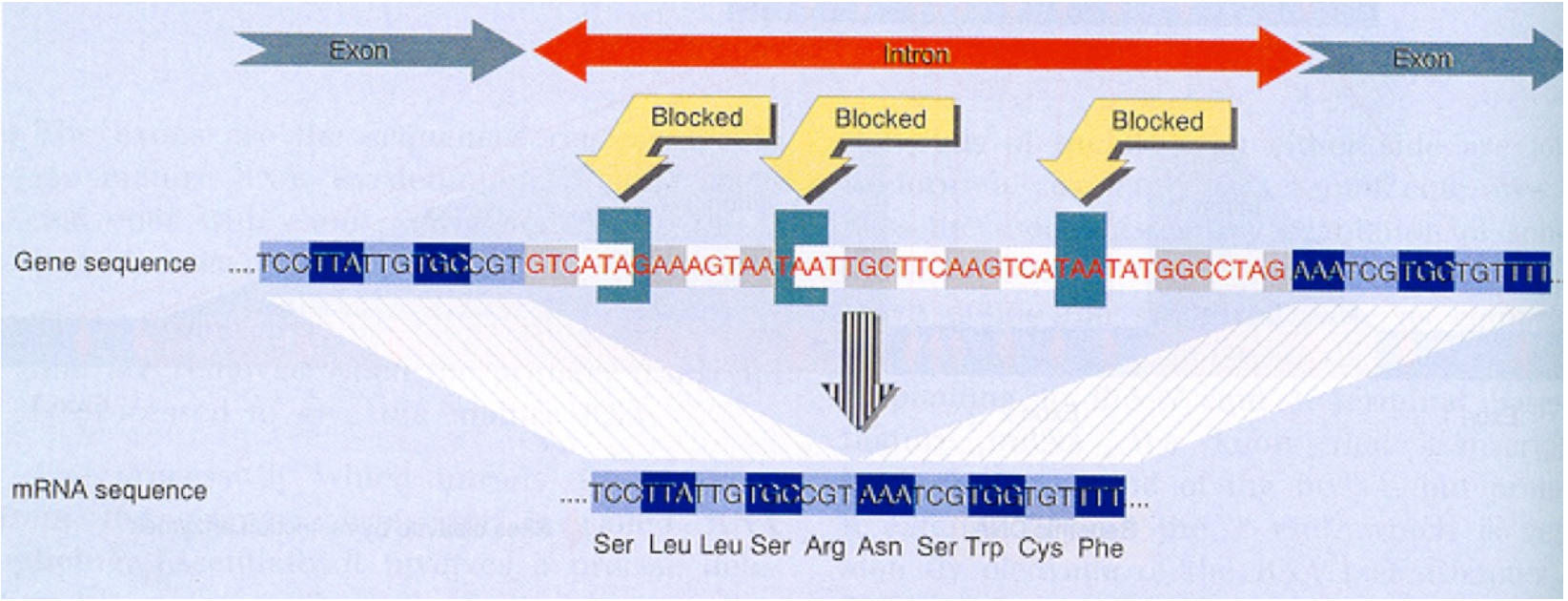
always complete loss of funtion

MUTATIONS IN CODING SEQUENCES

Point mutations

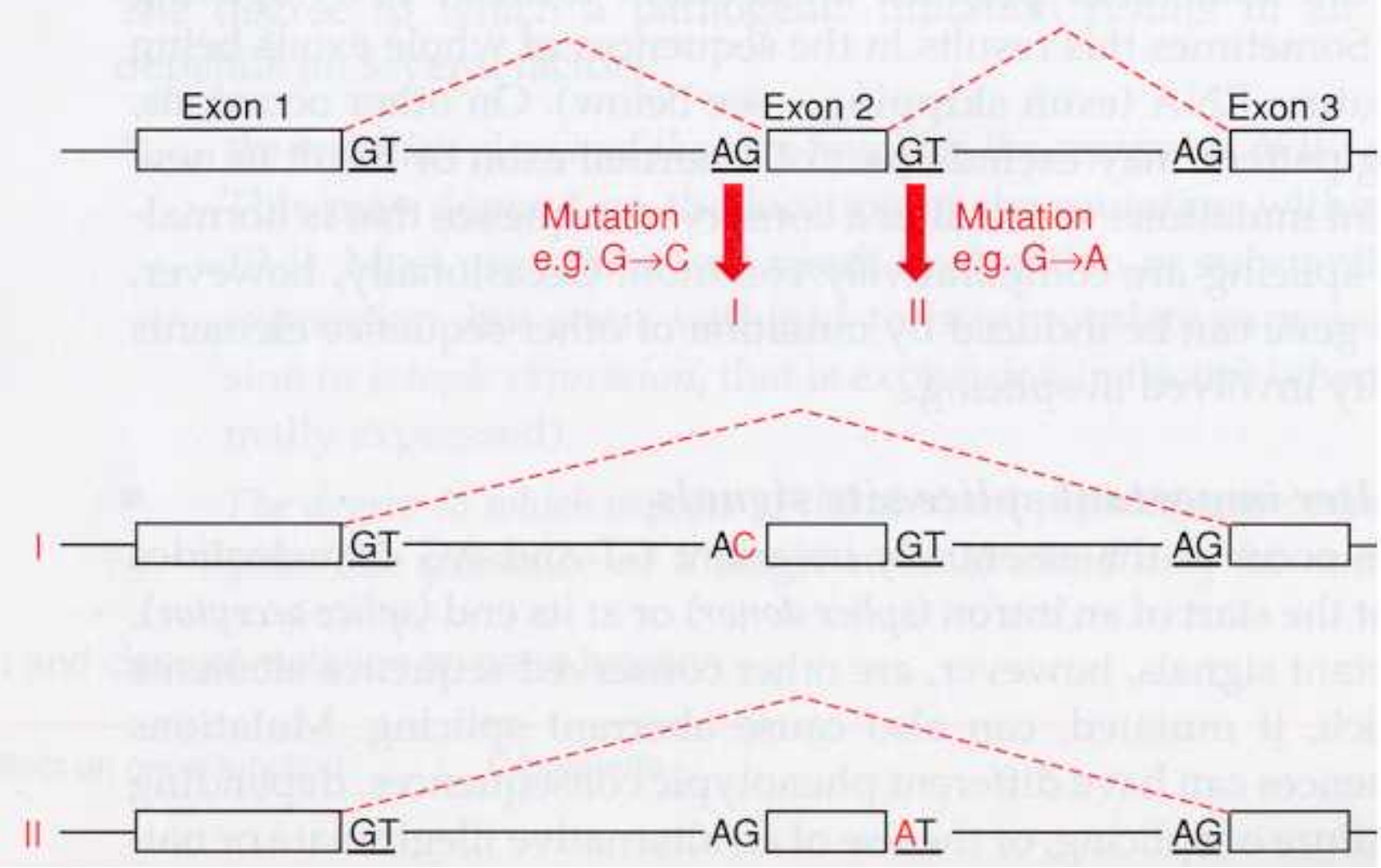
Affect 1 or a few base pairs

Frameshift mutations and premature termination codons (PTCs)



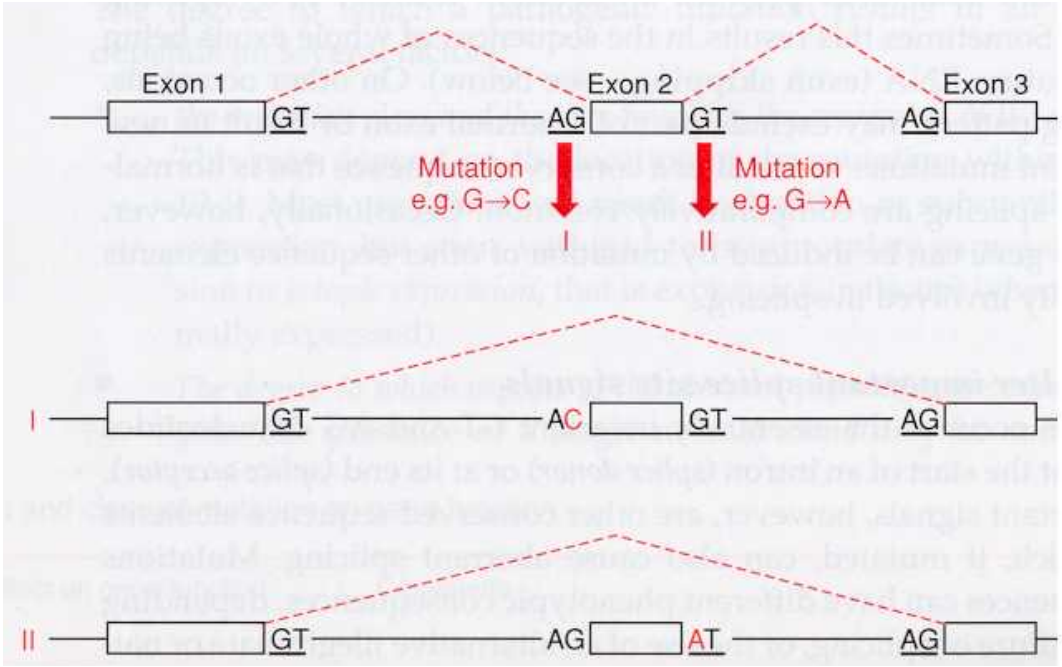
- Normal reading frame is open (ORF): no premature stop codon
- The two other phases (normally unread) contain many stop codond (ex: TAG, TAA)
- hence, a mutation that shifts the triplet frame of codon reading (frameshift) has 2 consequences
    1. Changes all downstream AA
    2. Then premature termination codon (= truncation)

Splice-site Mutation

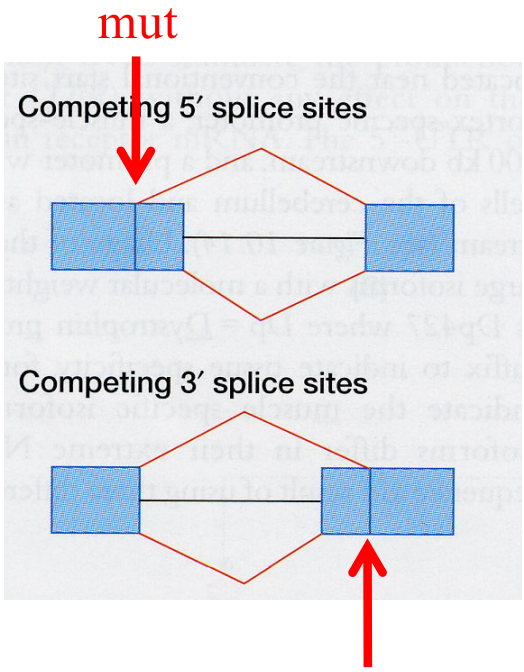


exon skipping

Splice Mutations



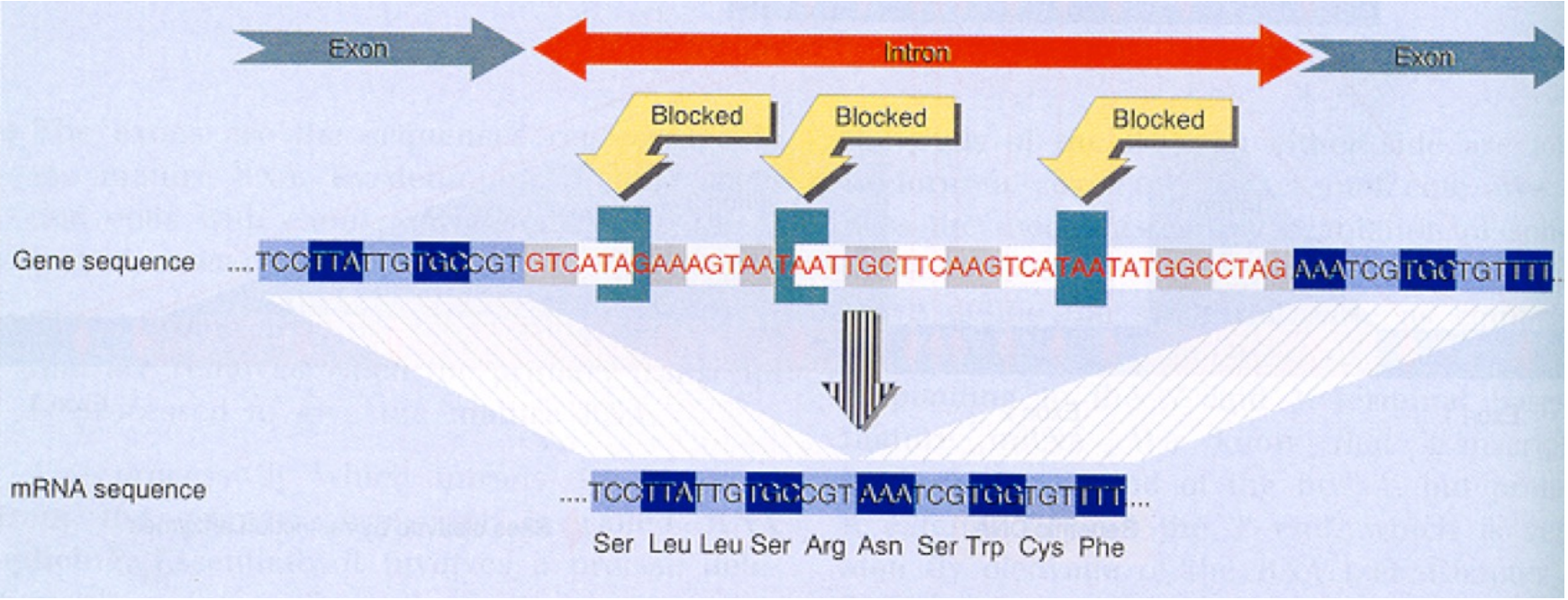
exon skipping



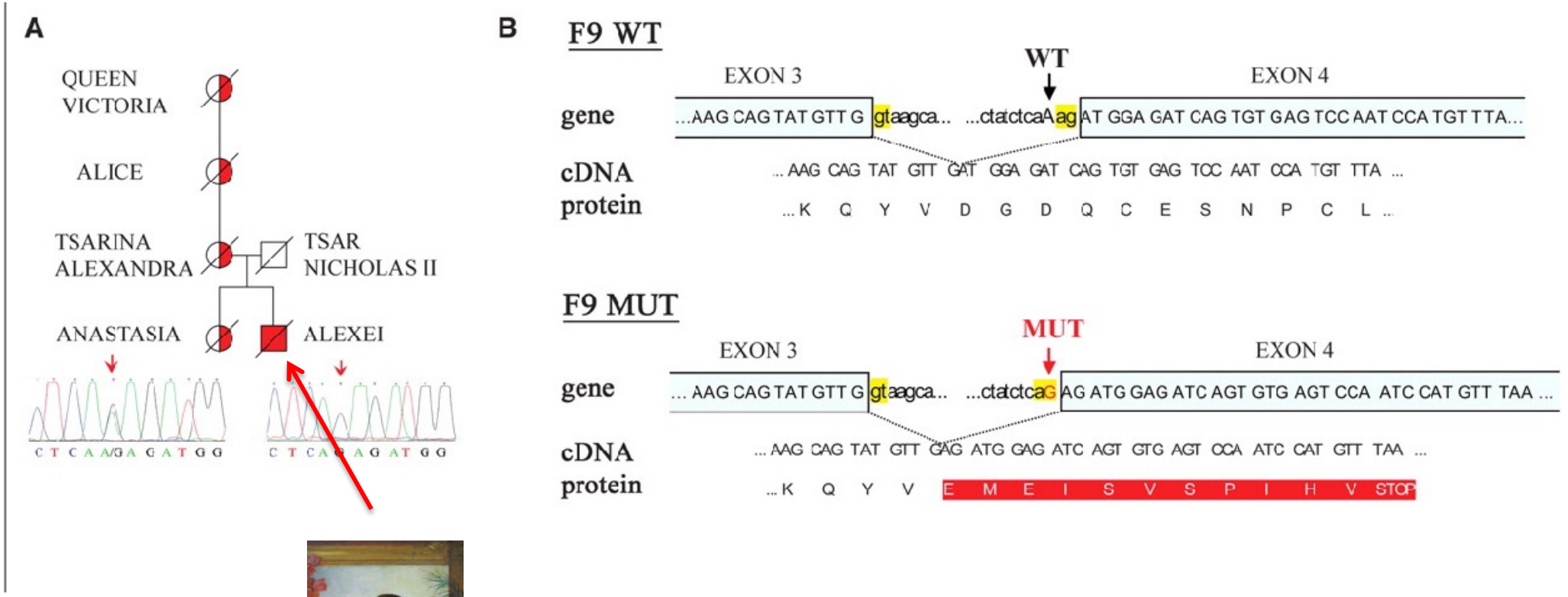
Retention of a piece of intron

Splicing Mutations

- *Added/lost exon may be IN PHASE exon = contains 3n nucleotides*
- *If added/lost exon is OUT OF PHASE, splicing mutation will add frameshfit to insertion/loss of protein fragment:*
    1. *gain/ loss of protein portion*
    2. *Modification of AA downstream of the gain/loss*
    3. *Stop codon stop downstream of all this*
- *Idem with retention of intronic sequence in mRNA, because introns contain numerous would-be stop codons in all reading frames*



The Royal Disease explained



A>G mutation creates a new splice acceptor site

=> exonisation of 2 nucleotides, AG

=> production of a truncated factor IX protein

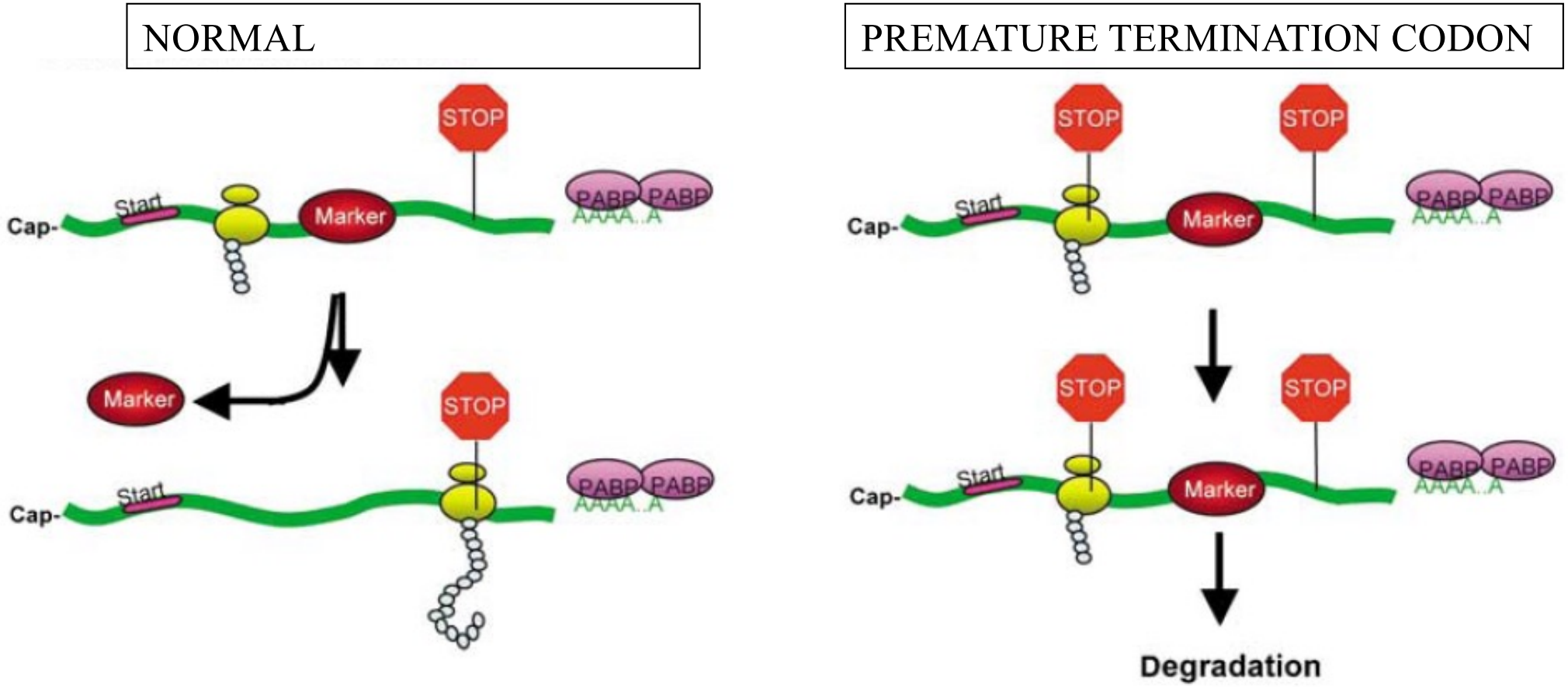
Rogaev et al. *Science* 6 November 2009: Vol. 326. no. 5954, p. 817

Premature Termination Codon (PTC)

- Truncates the reading frame
- Several types of mutations
  - nonsense
  - Indel (small insertion/small deletion) (not multiple of 3)
  - Splice mutations

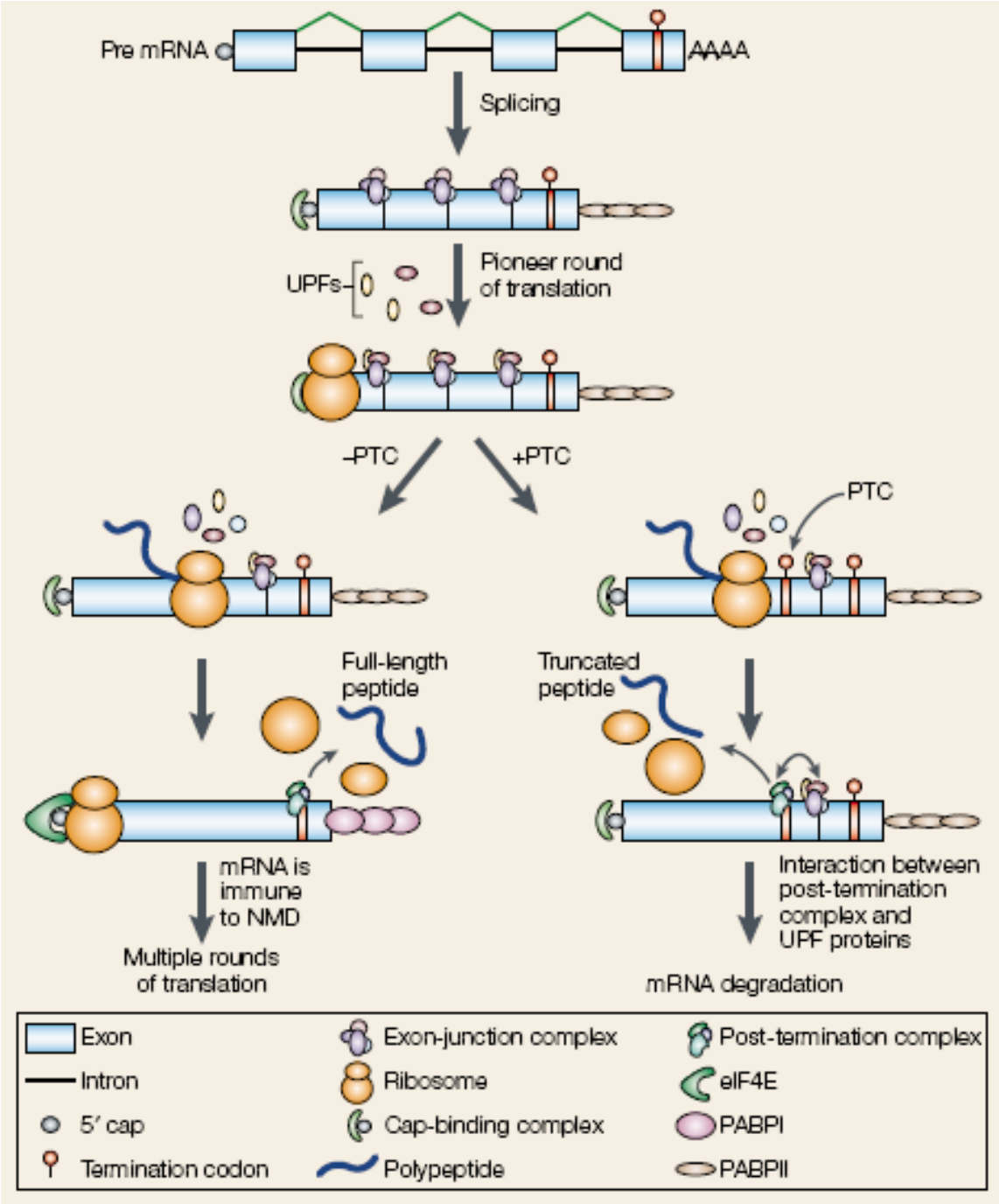
- Often causes nonsense-mediated mRNA decay

Nonsense-mediated decay of mRNA.



Shyu et al 2008 EMBO J

Marker for degradation = Exon Junction Complex

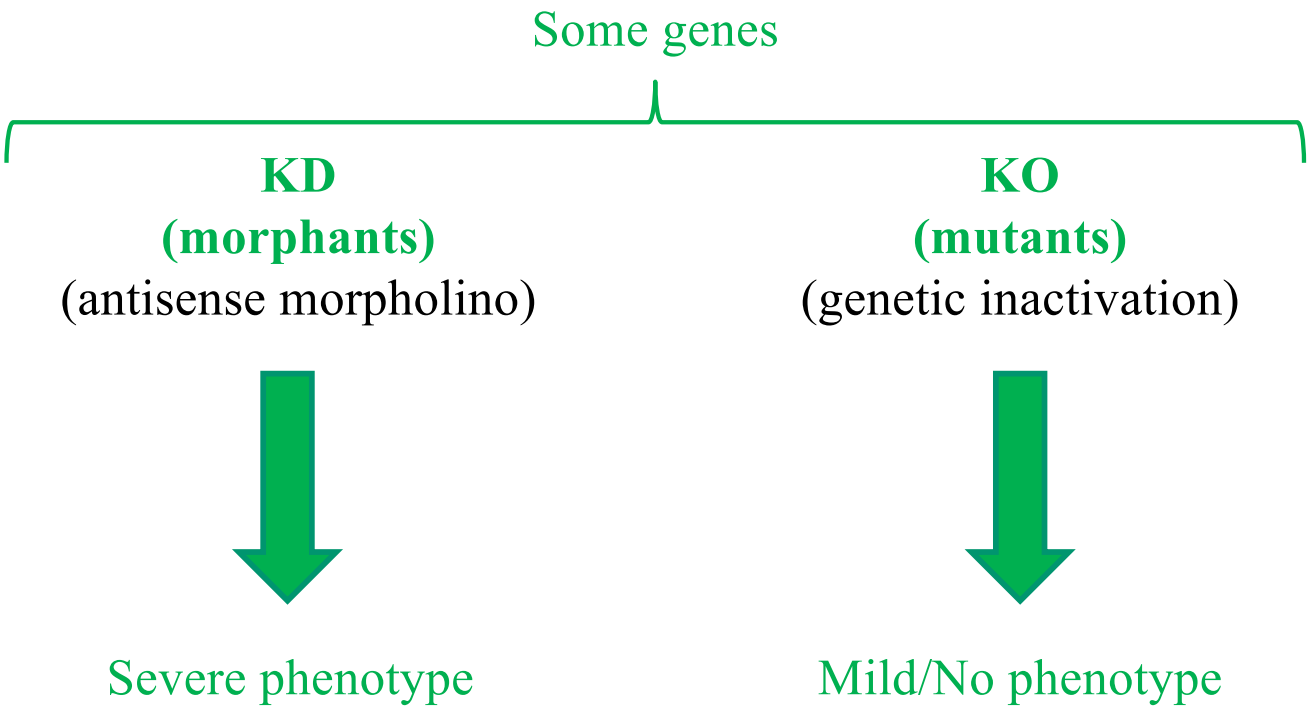


Cartegni et al 2002 Nat Rev Genet

Exon Junction complex

- Appears during mRNA maturation
- Displaced during first round of ribosomal read; special round (in nucleus?)
- If remains on cytoplasmic mRNA, induces its degradation

Paradox:
some PTCs produce milder phenotypes in model organisms
(ZF ; mouse)



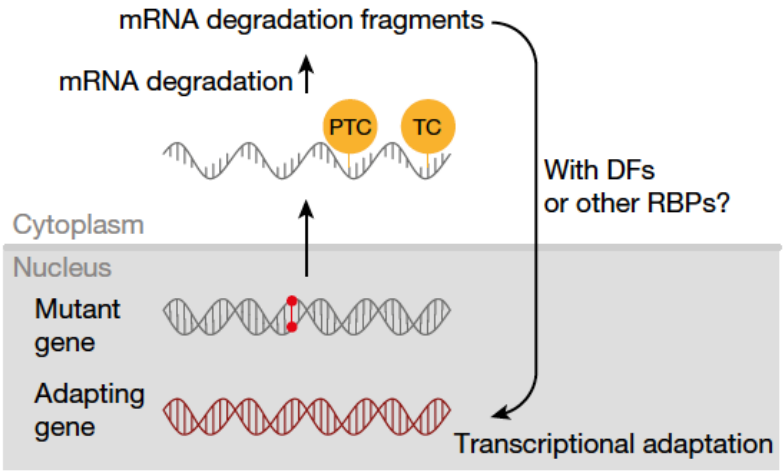
- ❑ Off target effect of morpholino
- ❑ Toxicity of excipient
- ❑ Genetic compensation response (transcriptional regulation)

GENETIC COMPENSATION triggered by NMD

- Transcriptional adaptation
- Correlates with mutant mRNA degradation
- Favours genes that exhibit sequence similarity with the mutated gene's mRNA
- Via Upf3a and COMPASS components

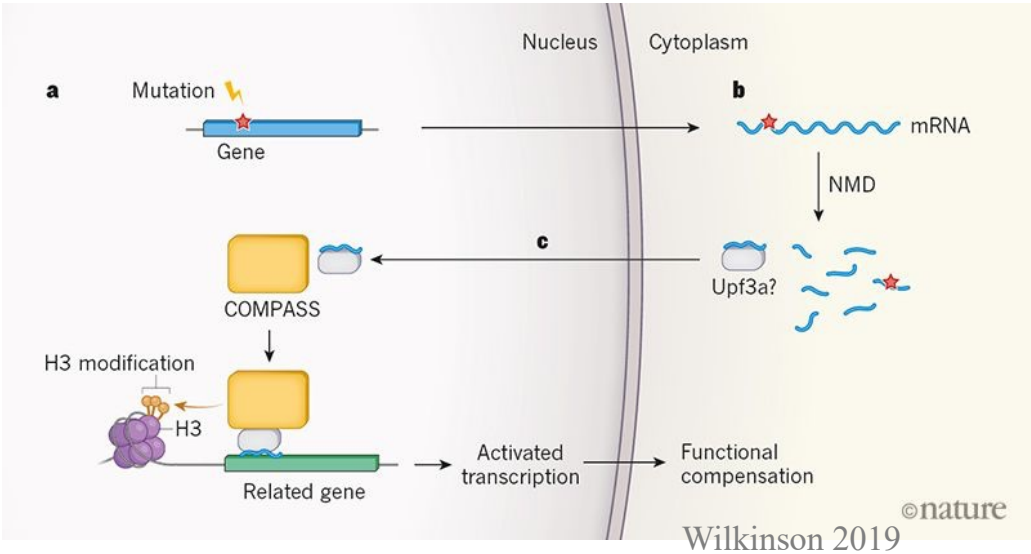
El Brolosy et al. 2019 Nature ; Ma et al. 2019 Nature

Genetic compensation triggered by NMD



DF, decay factors
RBPs, RNA-binding proteins

El-Brolosy et al. 2019



Wilkinson 2019

Gene mutations that truncate the encoded protein can trigger the expression of related genes. The discovery of this compensatory response changes how we think about genetic studies in humans and model organisms. ***Nonsense-induced transcriptional compensation***

Implications

- Missense may be more severe than nonsense even without dom neg
- Interindividual variability in transcriptional adaptation may explain variable phenotype in haploinsufficiency with PTC
  - Including upregulation of the wt allele
- Phenotype of up-regulated genes = ?
- ZF KD may be better model than ZF KO
- Some up-regulated paralogues = modifyer genes > Therap targets?
- RNAseq data may eventually help interpreting mutation effects

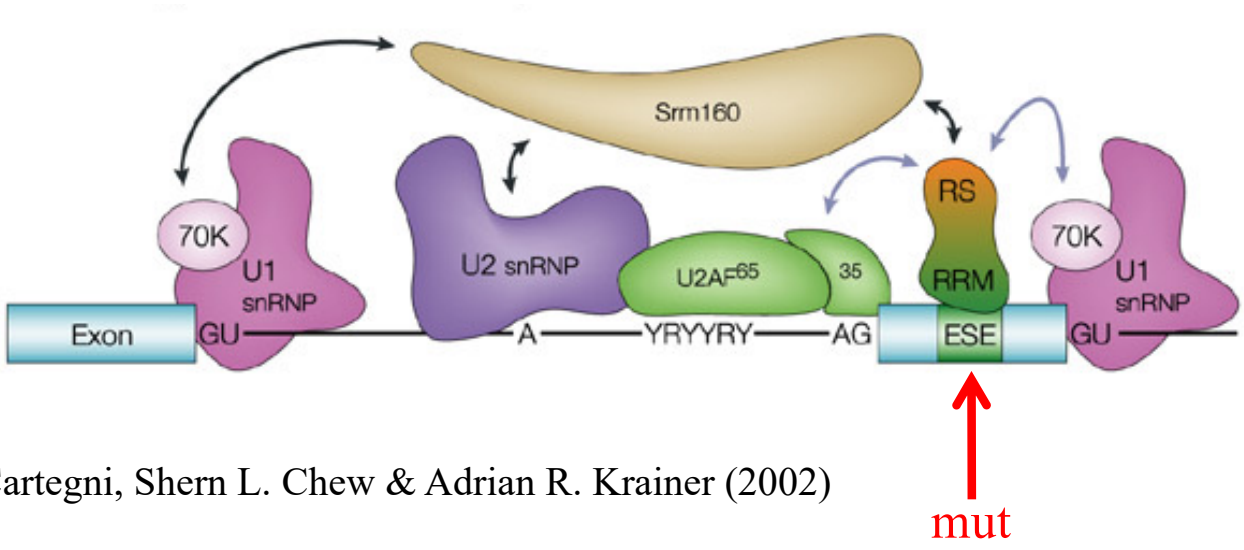
MUTATIONS IN CODING SEQUENCES

Point mutations

Affect 1 or a few base pairs

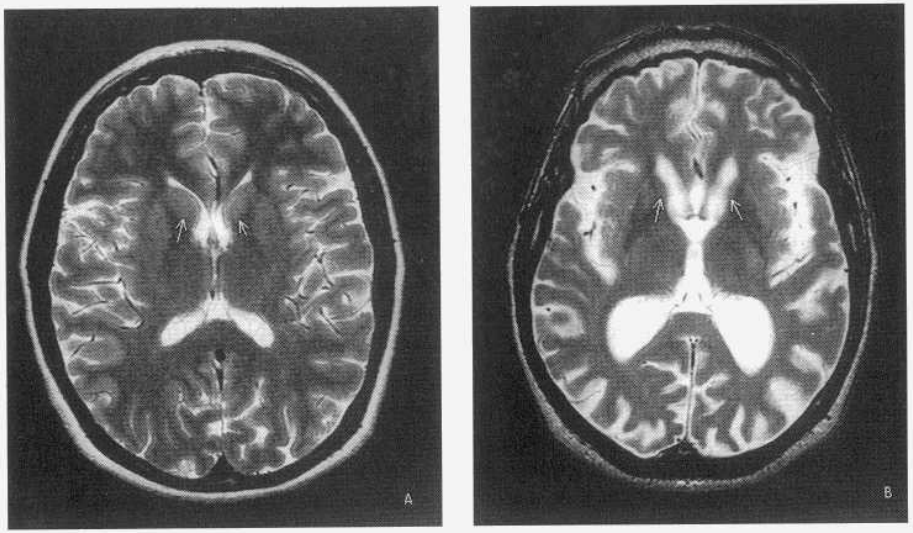
Mutations that do not change an AA

- Often 3rd base of codons (the genetic code is « degenerate »)
- USUALLY no effect on gene function because no effect on protein structure
- But not always: if mutation affects an Exon Splicing Enhancer, can have major functional effect independent of polypeptide sequence



Luca Cartegni, Shern L. Chew & Adrian R. Krainer (2002)

Nucleotide triplet Expansion



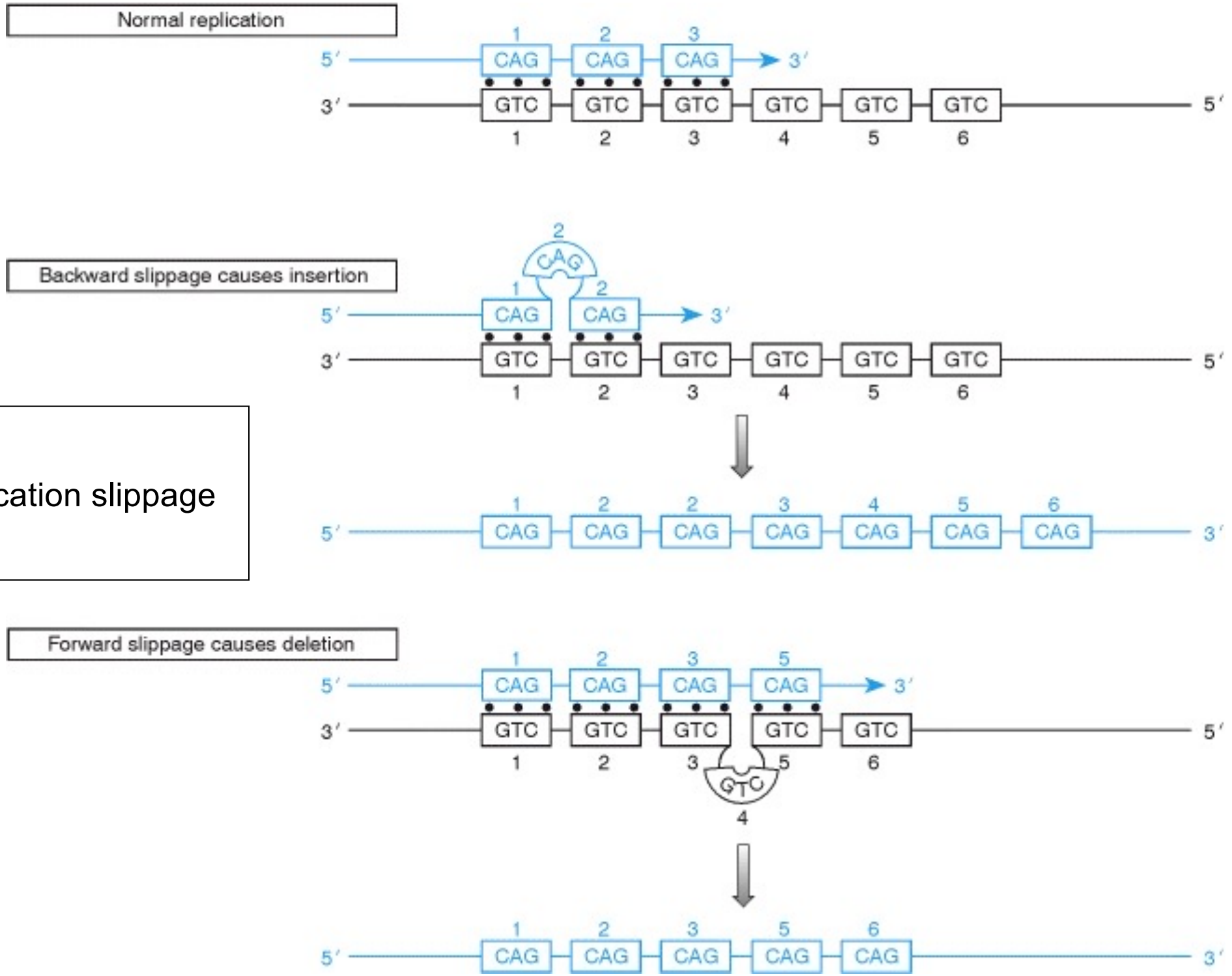
(a) Normal volunteer
(Courtesy of Dr M. Lowry, Hull, UK.)

(b) Huntington's disease

Huntington Disease

- Degeneration of striatal neurons (caudate nuclei)
- ↓ GABA

Replication slippage



MUTATION in SINGLE GENE (gene mutation)

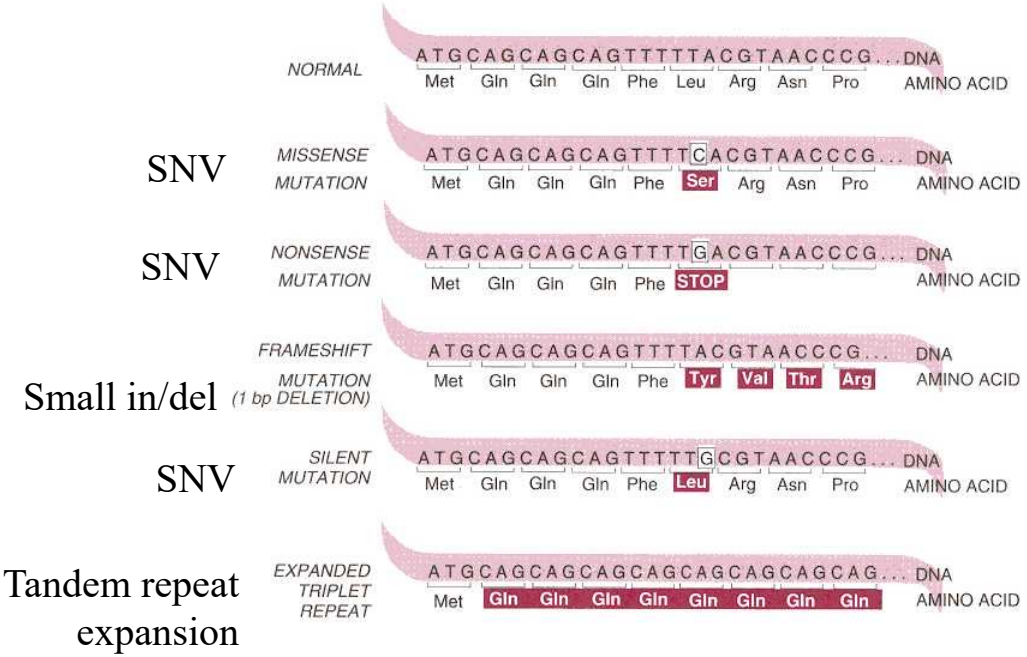
MUTATION IN NONCODING SEQUENCE
Promoter, enhancer, silencer, intron

MUT. IN CODING SEQUENCE

POINT MUTATION
- Small in/del (ex: DF508 du CFTR)
- Repeat expansion (ex: Huntington)
- Single Nucleotide Variant (SNV)

DELETION OR DUPLICATION of EXONS
(large intragenic in/del)

Large in/del = small CNV



Large indel gene mutation : **deletion or duplication of multiple exons**

Consider a gene with exons A, B, C, D, E.

Breakpoints in two introns: /

== A =/== B == C ==/== D ==== E==

NAHR during meiosis yields two gametes with mutations in this gene:

Large indel gene mutation : **deletion or duplication of multiple exons**

Consider a gene with exons A, B, C, D, E.

Breakpoints in two introns: /

== A =/== B == C ==/== D ==== E==

NAHR during meiosis yields two gametes with mutations in this gene:

- Interstital deletion (exons del, intragenic deletion):

  == A === D ==== E==

  => mRNA: ADE

Is this deletion IN FRAME?
= is nb of nucleotides 3n?
If≠ 3n, frameshift, causing complete LOF

- Interstitial duplication :

  == A === B == C ==== B == C ==== D ==== E==

=> mRNA: ABCBCDE

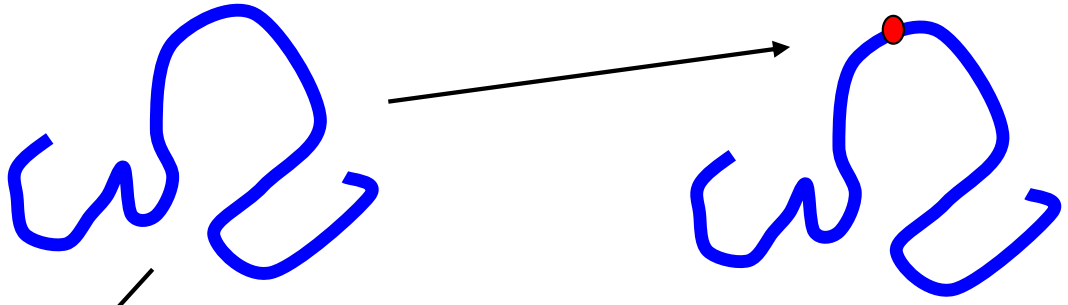
Is this duplication IN FRAME?
= is nb of nucleotides 3n?
If≠ 3n, frameshift, causing complete LOF

MUTATIONS IN CODING SEQUENCES

Point mutations

Affect 1 or a few base pairs

Functional effect of coding mutations

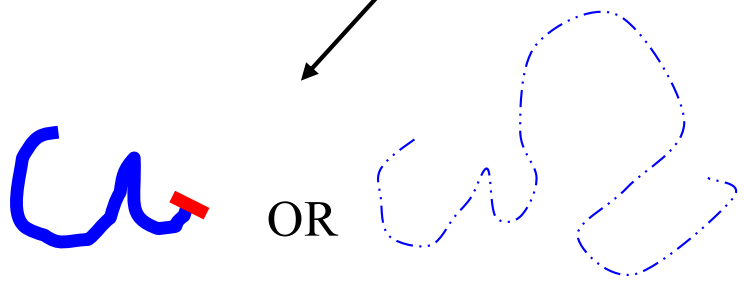


Null alleles

- (most) Stop codon
- (most) frameshifts
- (most) splicing

MISSENSEs :

AA → other AA



Loss of function (null allele)

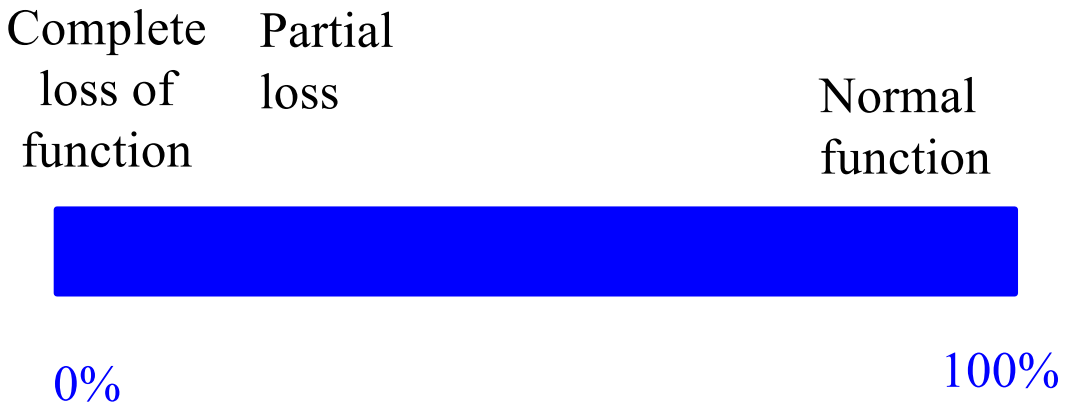
Occasionally, trunctated product still has function: antimorph, neomorph or hypomorph

Variable effect:

- Loss of function.
  - ✓ Total
  - ✓ Partial
  - ✓ Total + 2nd allele
- Gain of function.
- Variant normal/polym.

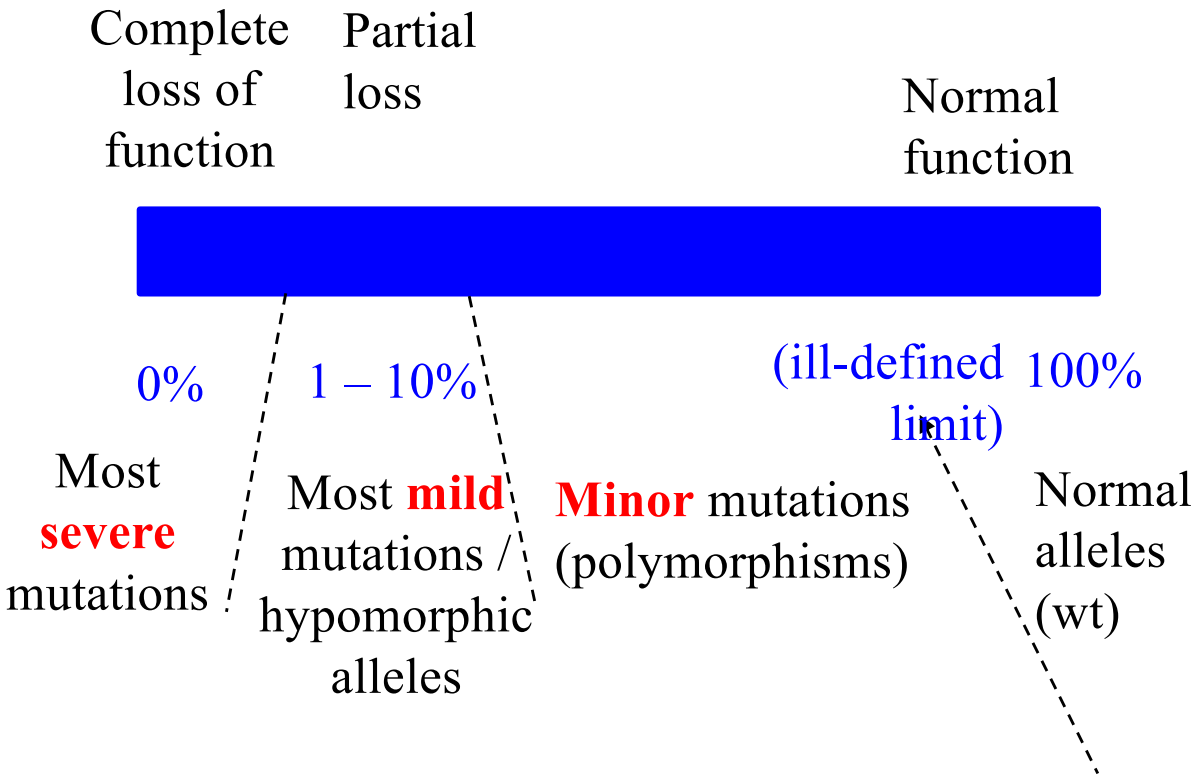
Functional effet, **loss of function** type

- Quantitative effect, with continuum



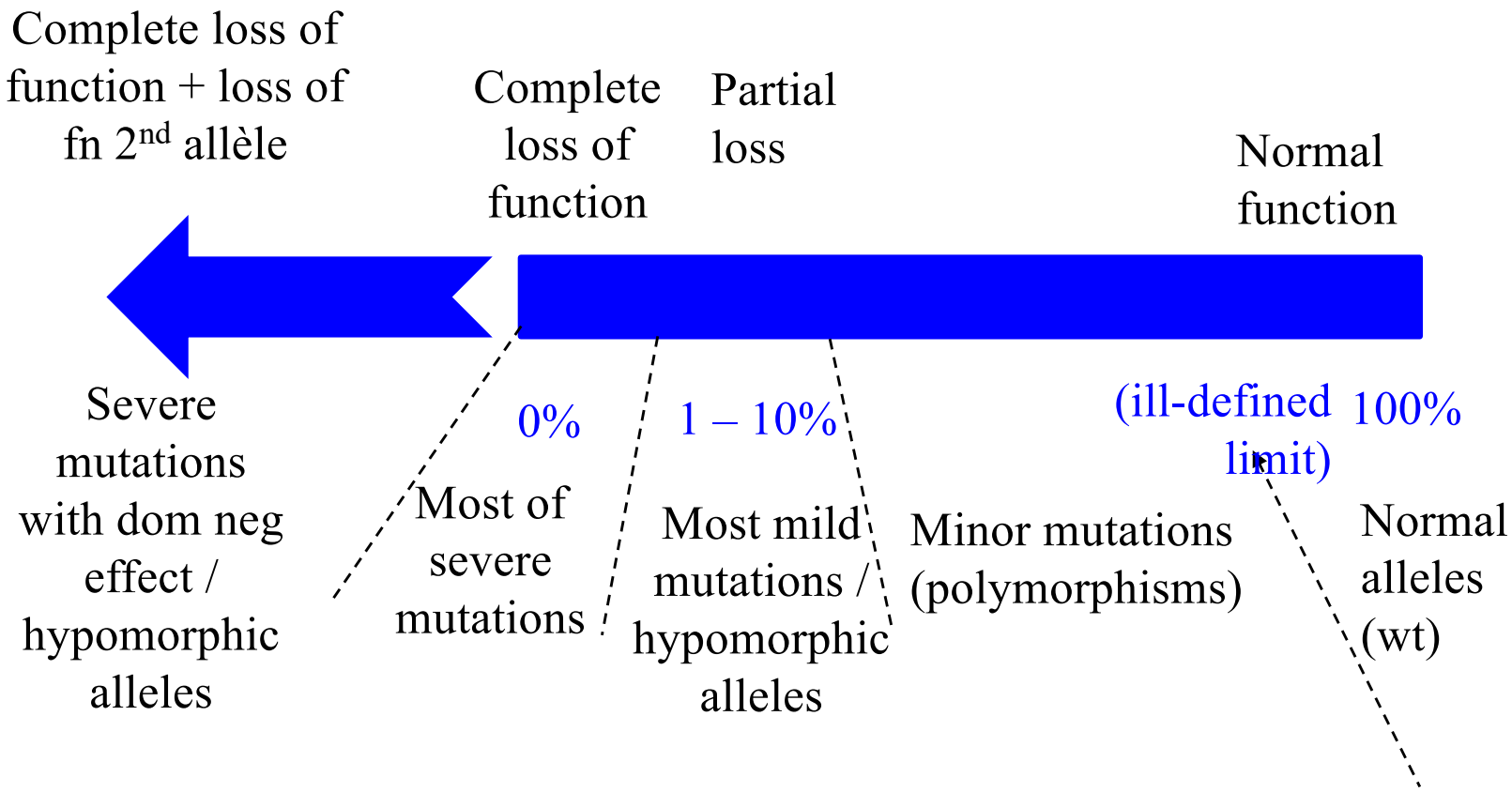
Functional effet, **loss of function** type

- Quantitative effect, with continuum



Functional effet, **loss of function** type

- Quantitative effect, with continuum



Mutations causing gain or loss of function

Complete Loss + loss fn 2nd allele

Complete loss

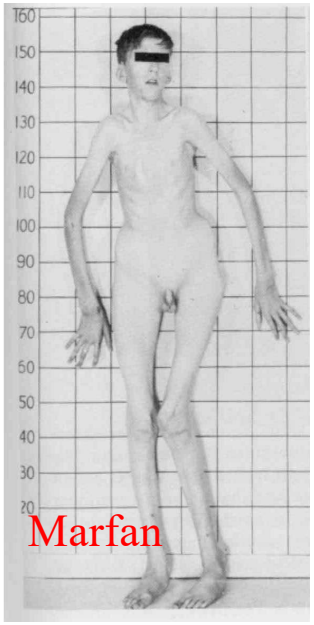
Partial loss

Normal activity

Gain of function:
- Dose effect
- Gain toxic effect

-90% -50% 0% 1-10% 100% 150%

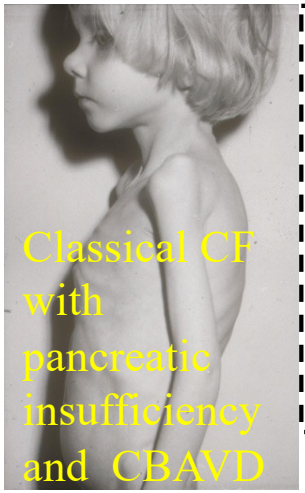
Antimorphs:
- Missenses (most)



Null alleles:
- Nonsenses
- Missenses
- Frameshift
- Locus deletion

Mild mutations
Missenses, Or others

Hypomorphic alleles

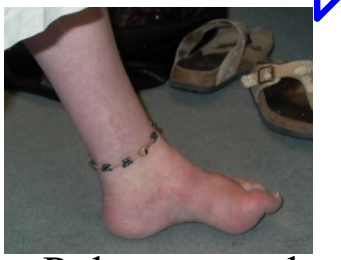


CF without pancreatic insufficiency,

Or
Isolated CBAVD

Wild type alleles (wt) and Neutral polymorphisms :
- Intronic
- 3rd base codons, …

Or not completely neutral, and contribute to disease susceptibility



Polyneuropathy < duplication PMP22 allele

neomorph
Novel toxic fn

Huntington dis. by expansion of (CAG)n > (Gln)n

Mutations and polymorphisms

- MINOR MUTATIONS = Polymorphisms

- MILD MUTATIONS

- MAJOR (SEVERE) MUTATIONS High penetrance

all are genetic VARIANTS = not wild type

By definition, Polymorphism if allele frequency ≥ 0.01

- Consider a locus with 2 alleles: A and B
- With frequences = p and q
- If p>q, q = minor allele frequency (MAF)

- **POLYMORPHISM if q ≥ 0.01**

  if q <0.01, « rare genetic variant »

Polymorphism: allele fqcy ≥ 0.01

- A, B, O blood group
- HLA B27
- Many, many other coding changes
- Many, many non-coding SNPs
- Many, many CNPs

- Daltonism mutation
- HFE*C282Y
- CFTR*DF508
  (! According to definition, DF508 is a human polymorphism !)

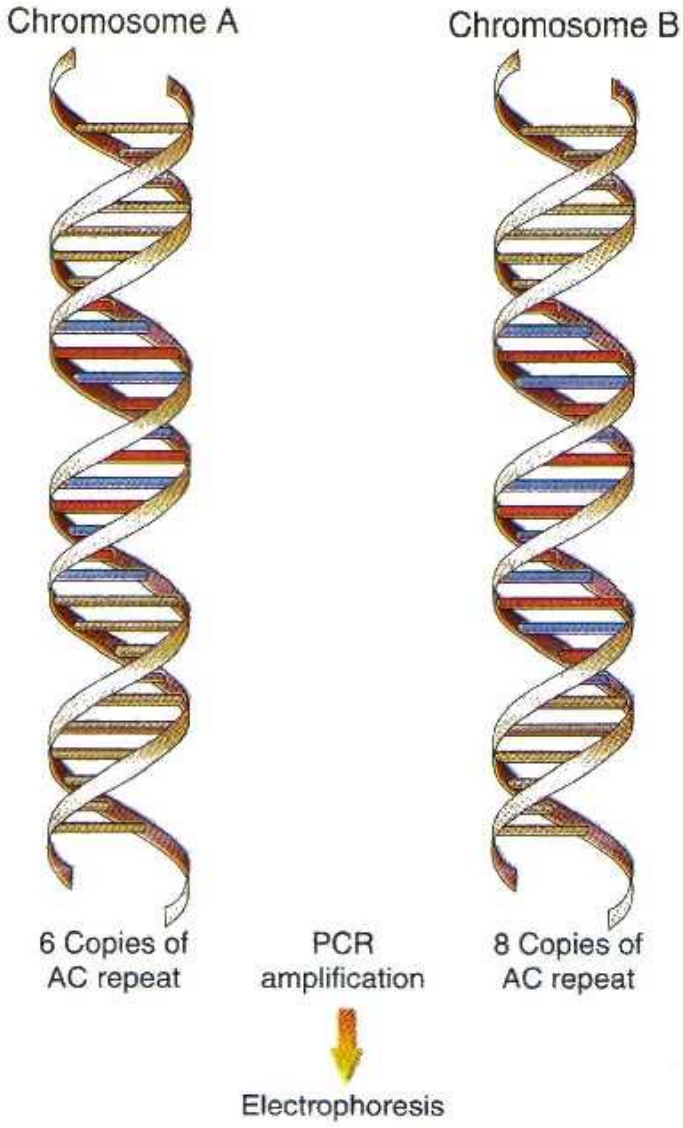
Polymorphic markers

- Neutral polymorphisms, frequent in population (typically MAF >.05)
  - Minisatellites (obsolete)
  - Microsatellites (= short tandem repeats)
  - SNPs
  - others

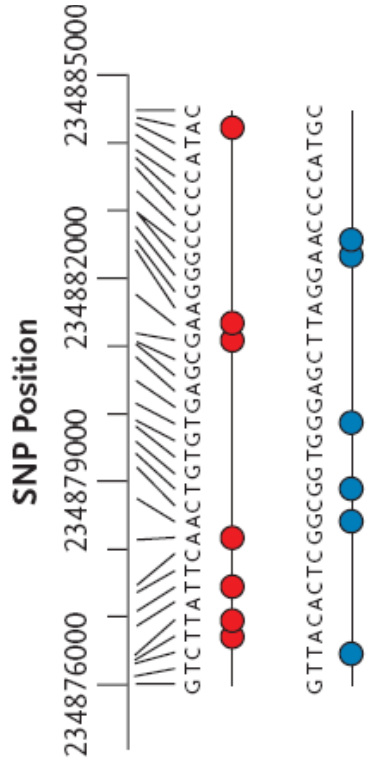
- May serve as markers of chromosomal segment
  - Linkage studies, in families
  - Association studies, in populations (Gwas)

Genotyping polymorphic markers

Microsatellite



SNPs



PCR amplification

DNA microarray Chip

Single Nucleotide Polymorphism (SNP)

…5’ AATCGAGG 3’…
…3’ TTAGCTCC 5’…

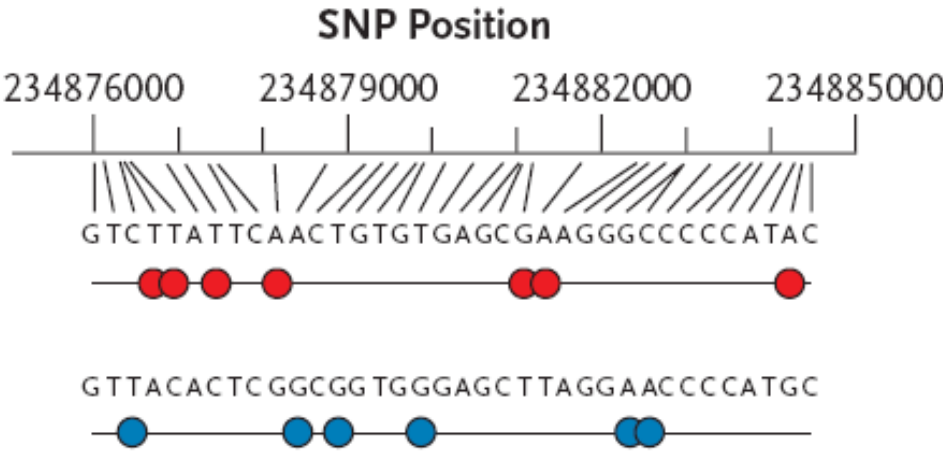
Allele 1 (allele C), frequency = p

…5’ AATTGAGG 3’…
…3’ TTAACTCC 5’…

Allèle 2 (allele T), frequency = q

Single Nucleotide Polymorphisms

- Ex: 10,000 bp (#2)
- Coding or non-coding
- 2 haplotypes shown

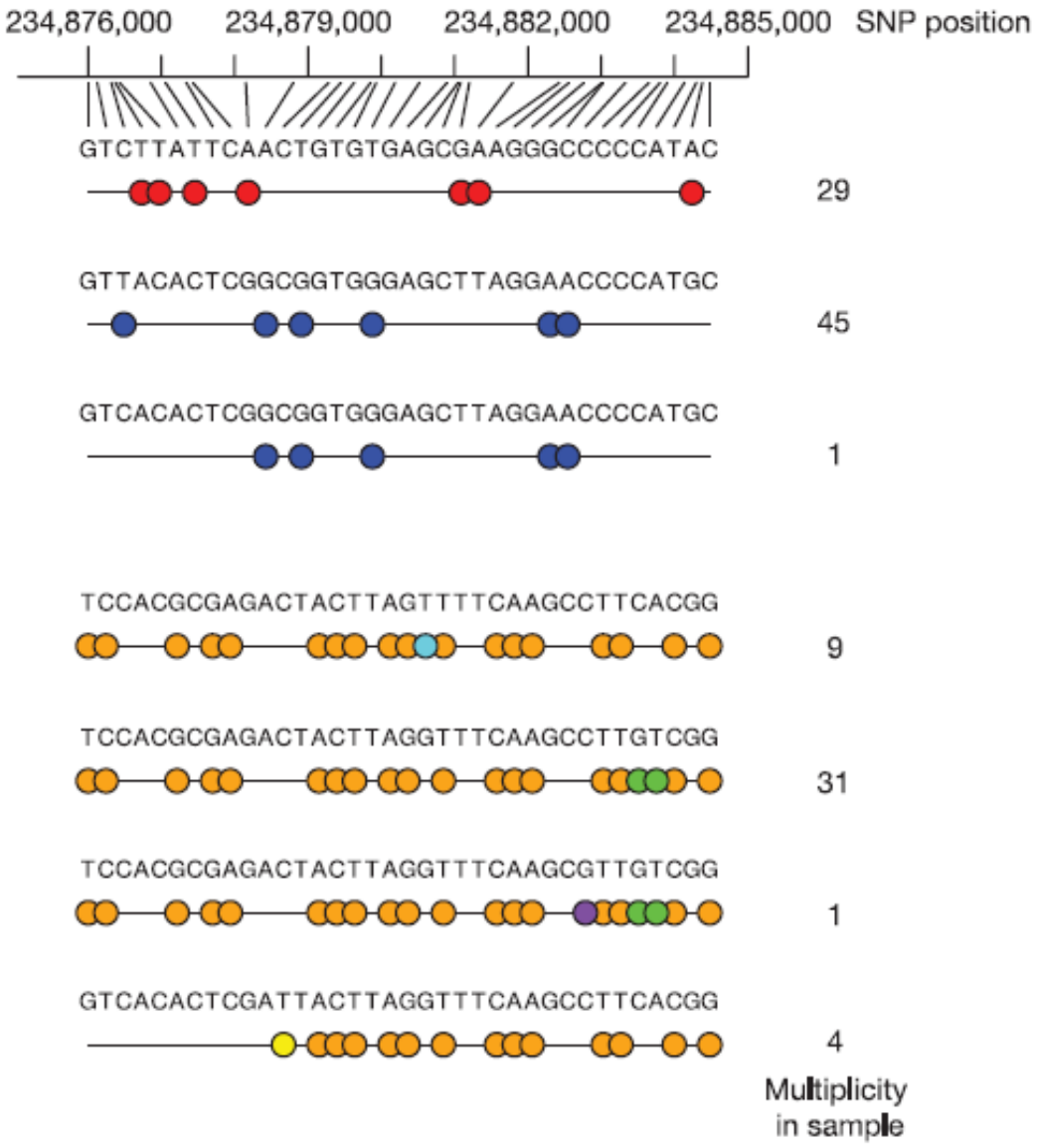


- millions of SNPs in genome

- Many CNPs (copy number polymorphisms)

Haplotype = sequence of alleles on a short piece of chromosome

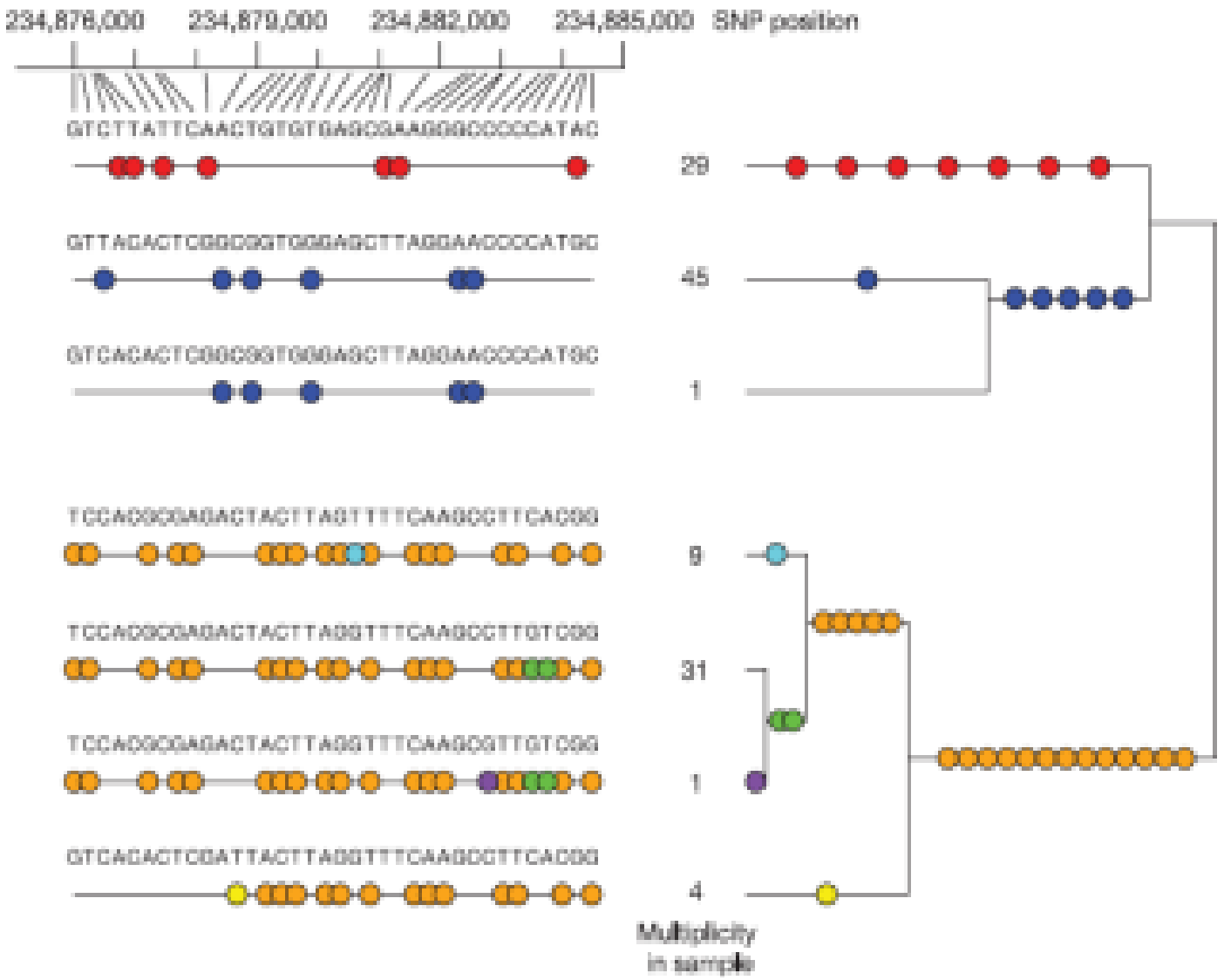
Limited number of haplotypes at 10kb loci



A haplotype map of the human genome

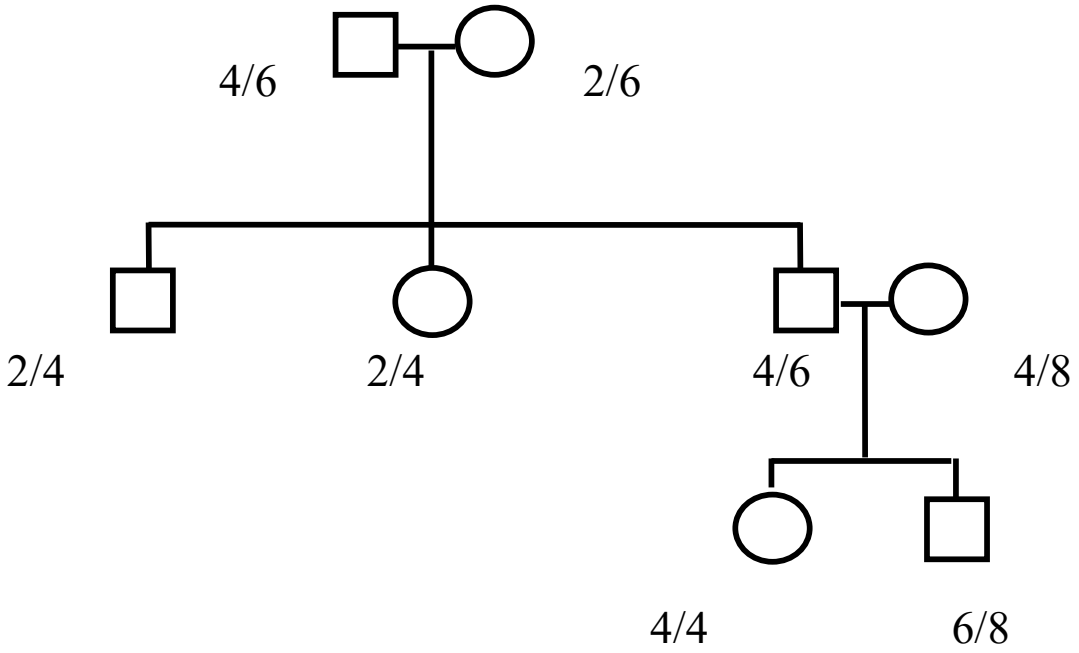
The International HapMap Consortium*

NATURE|Vol 437|27 October 2005

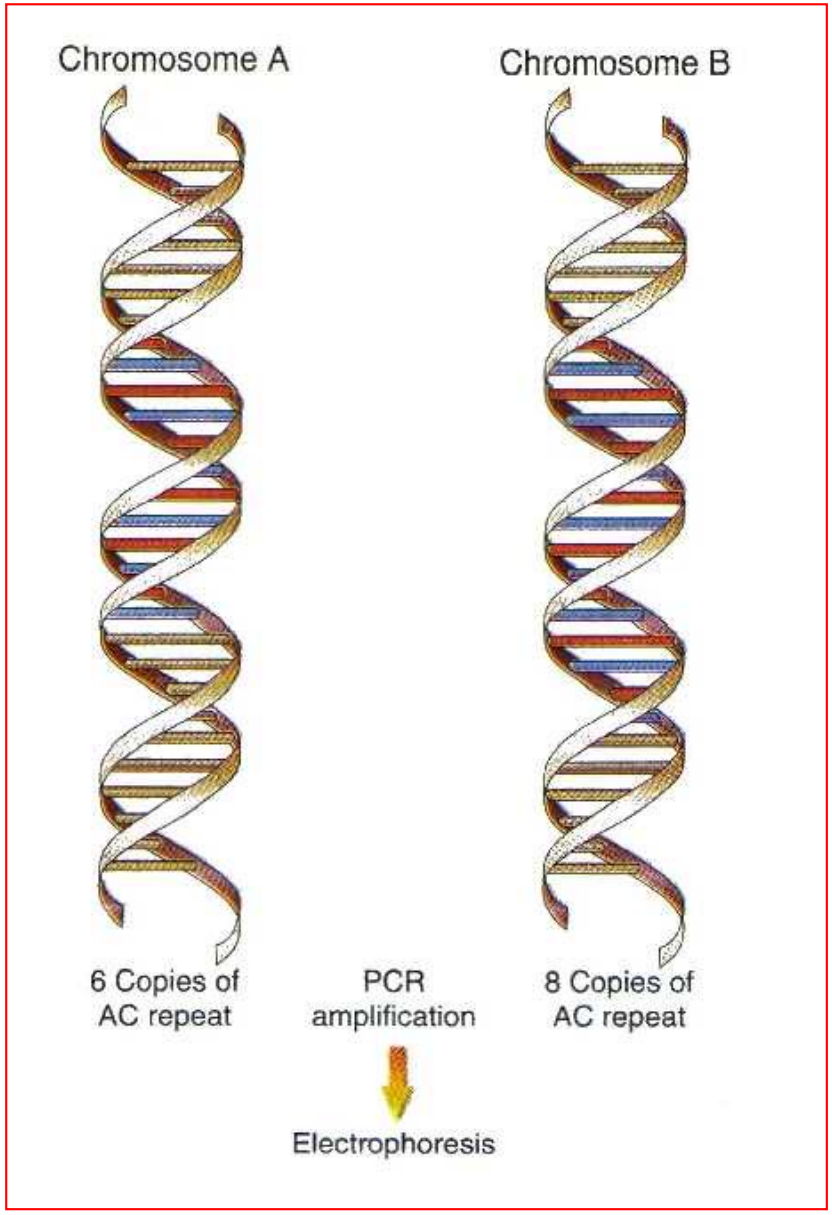


1,000,000 SNPs
DNA microarray chip

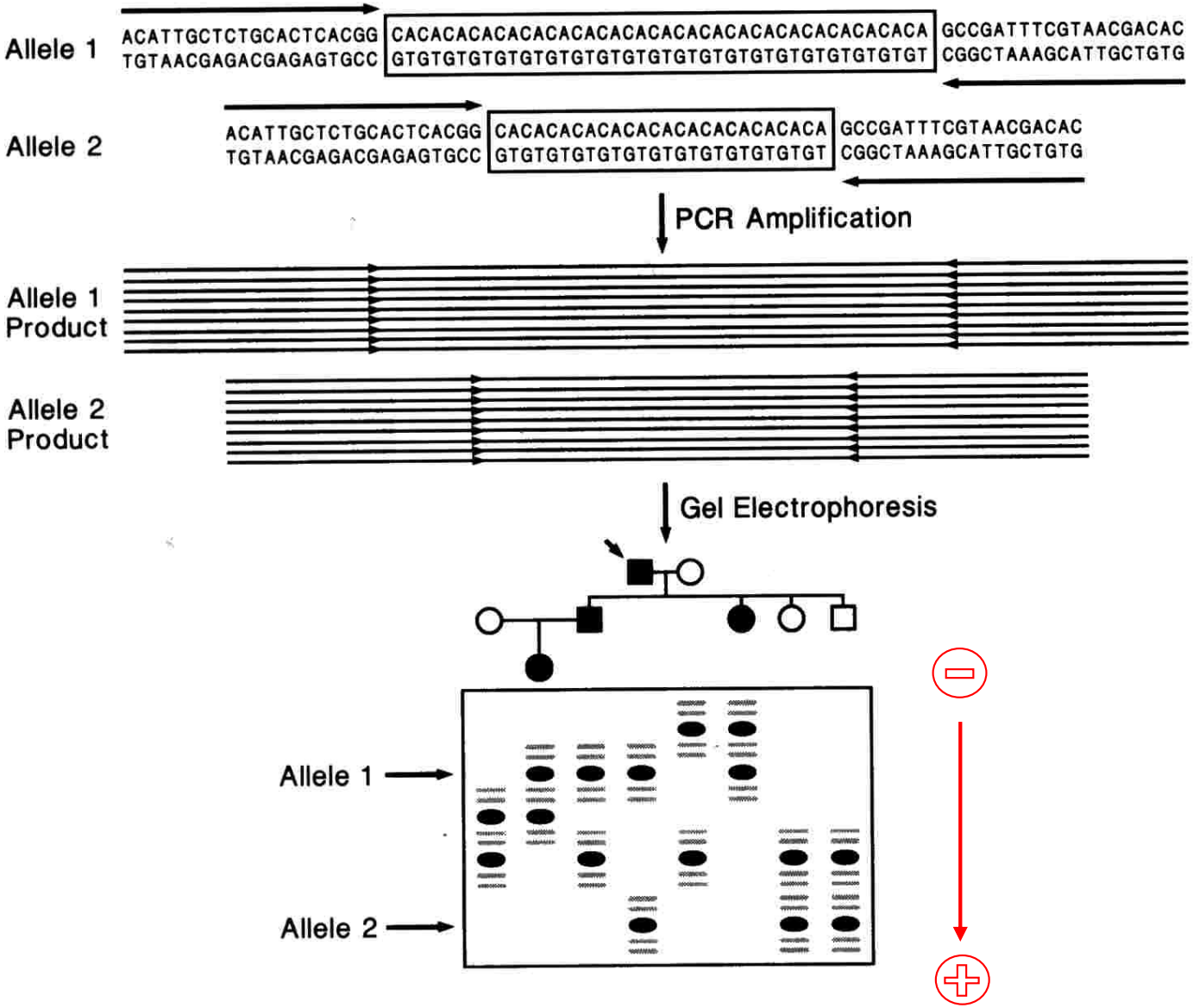
Microsatellite polymorphisms



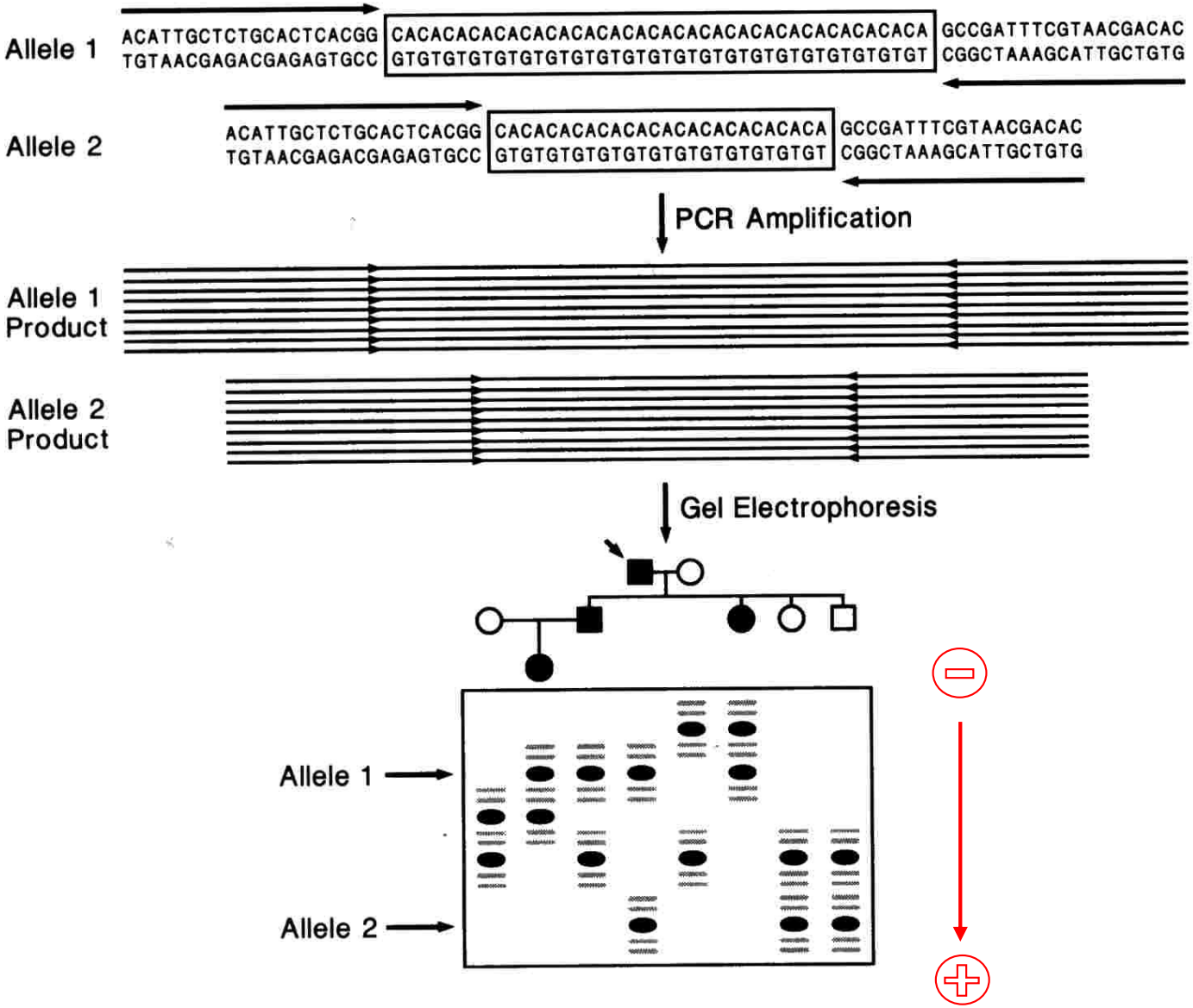
- One microsatellite locus
- Many alleles

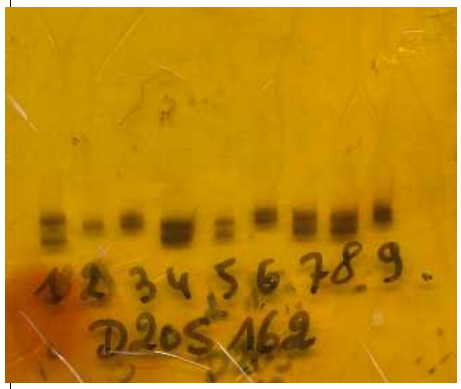


Genotyping a microsatellite

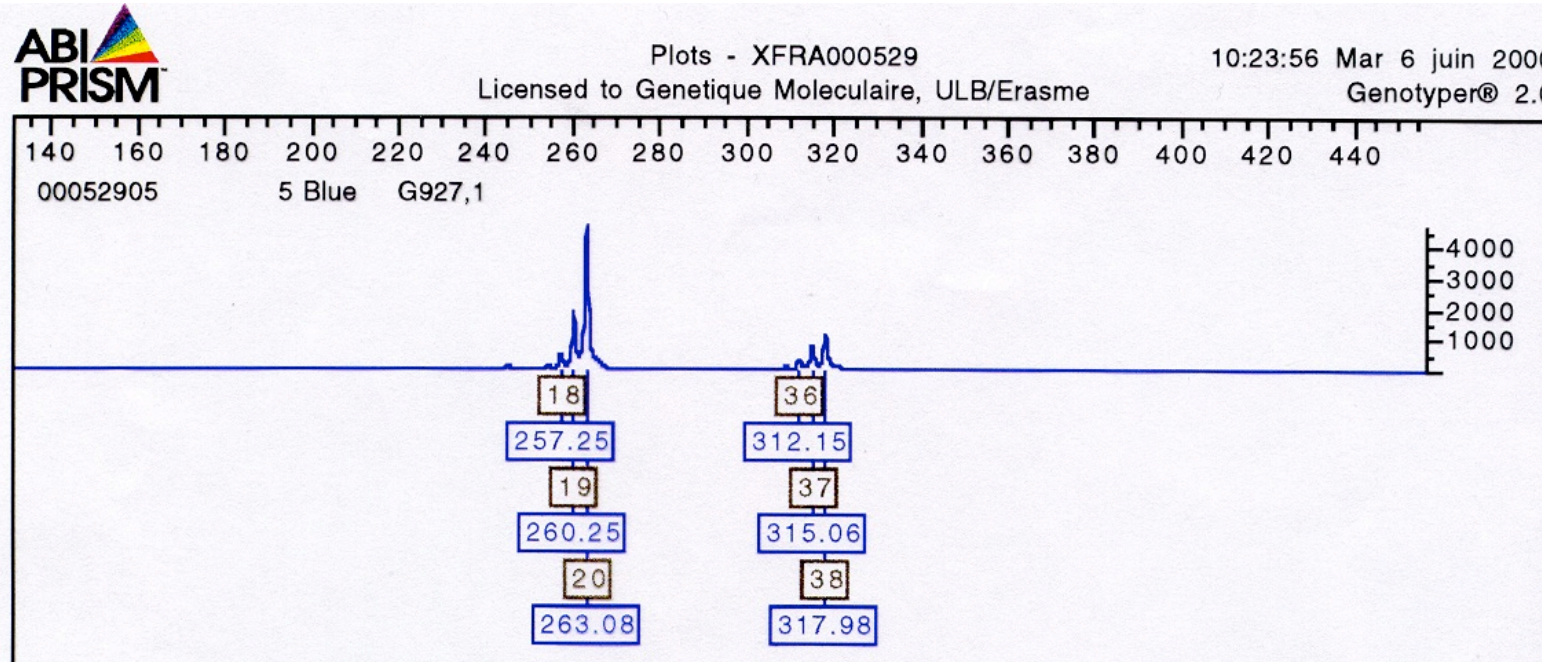


Genotyping a microsatellite

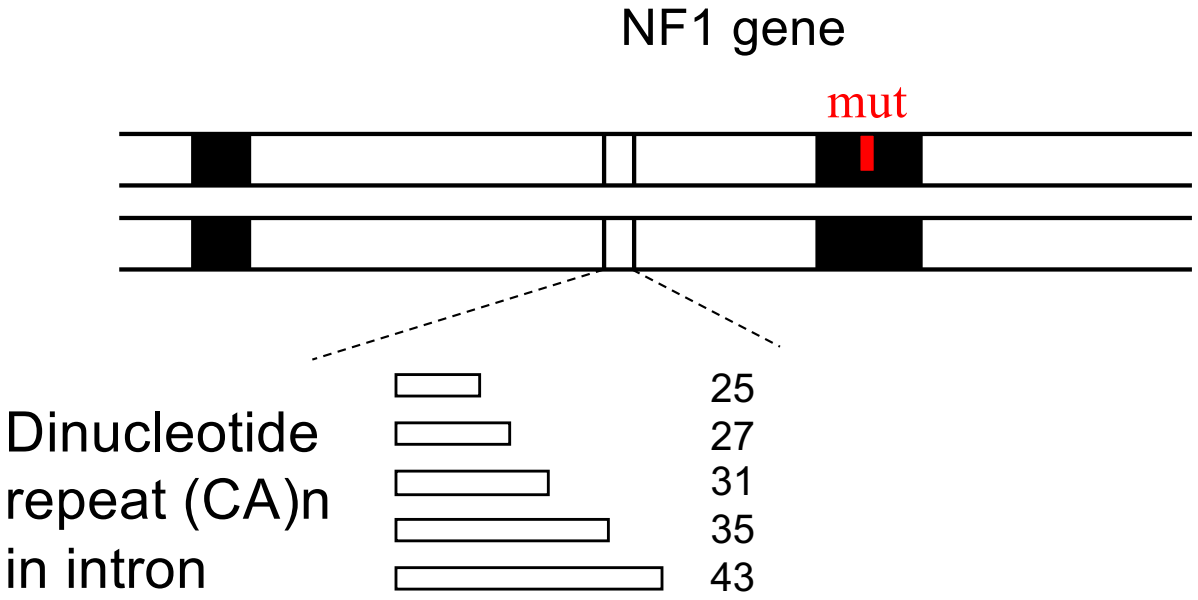




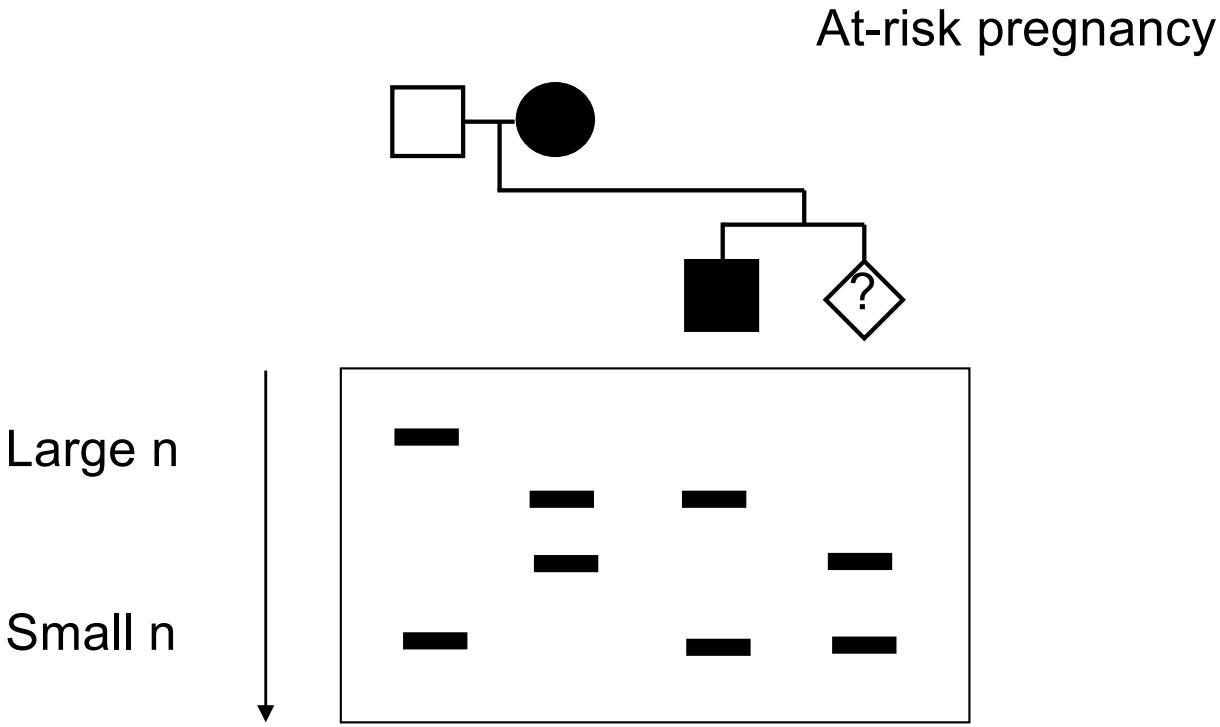
Capillary Electrophoresis, laser read-out



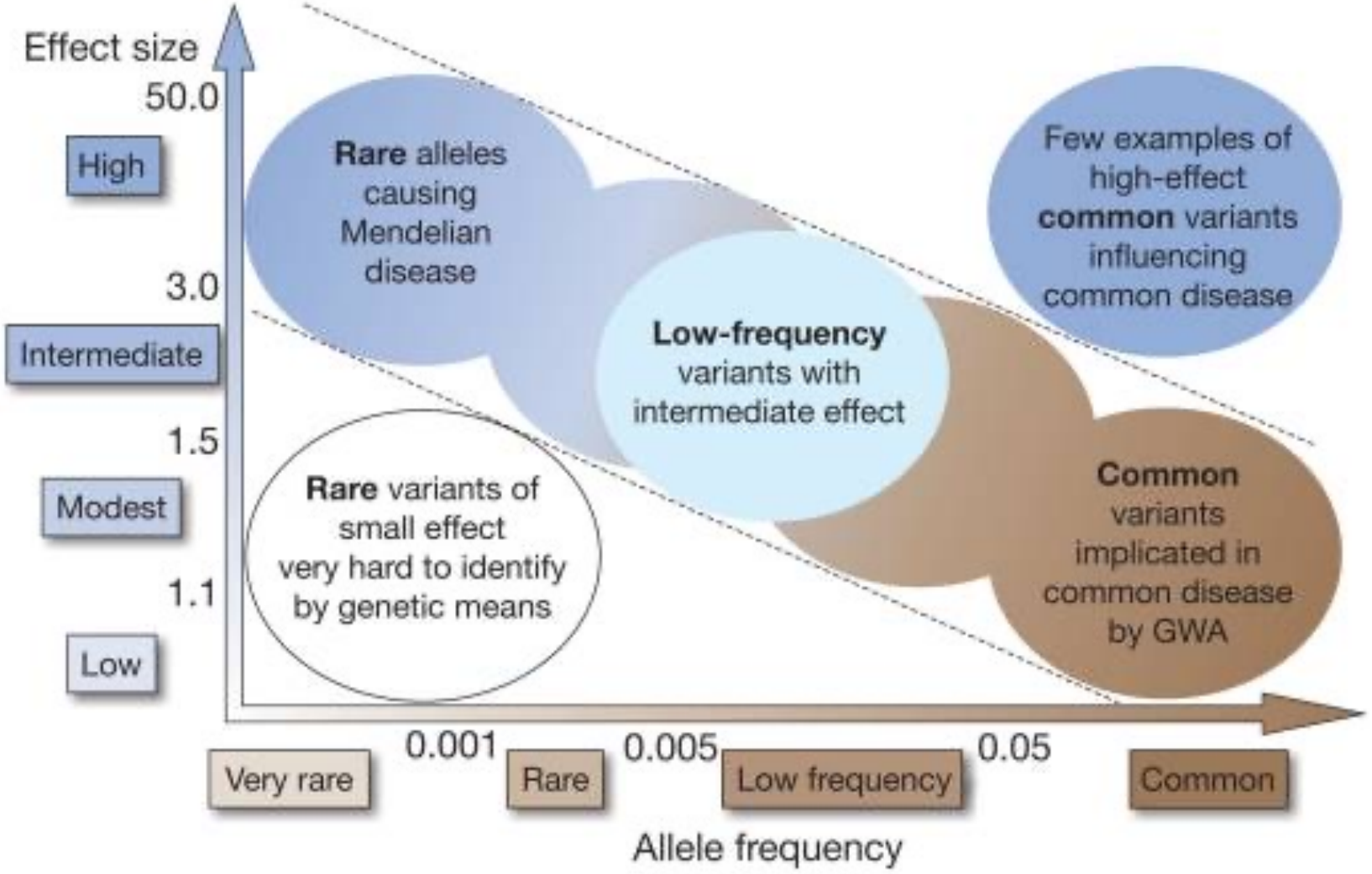
2 alleles of a microsatellite



Linkage study in family using a microsatellite marker

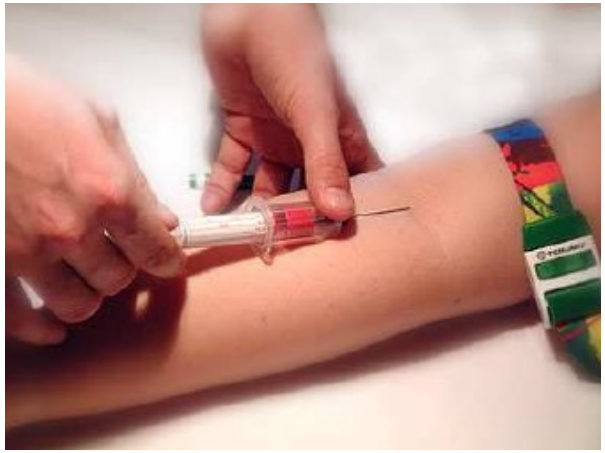


Rare genetic variants



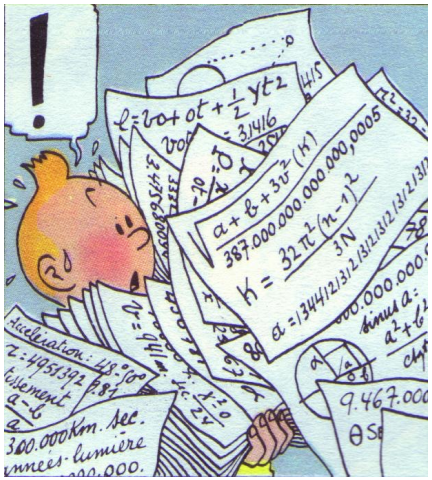
Manolio TA et al. Nature 2009: 8; 747-53

DNA sequencing goes faster than interpretation



Blood sampling
> DNA extraction

DNA sequencing



DNA sequence analysis

N of 1

Novel mutation

novel genetic aqriant

(never observed before)

is it disease-causing

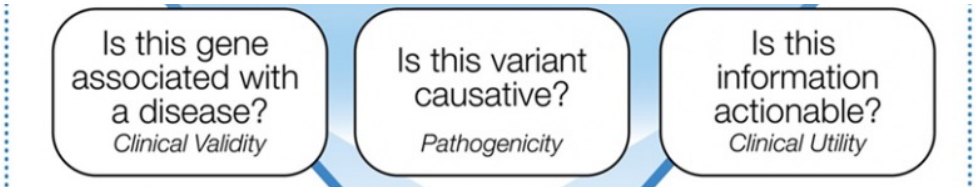
?

Intersting variant in Rotatin gene.

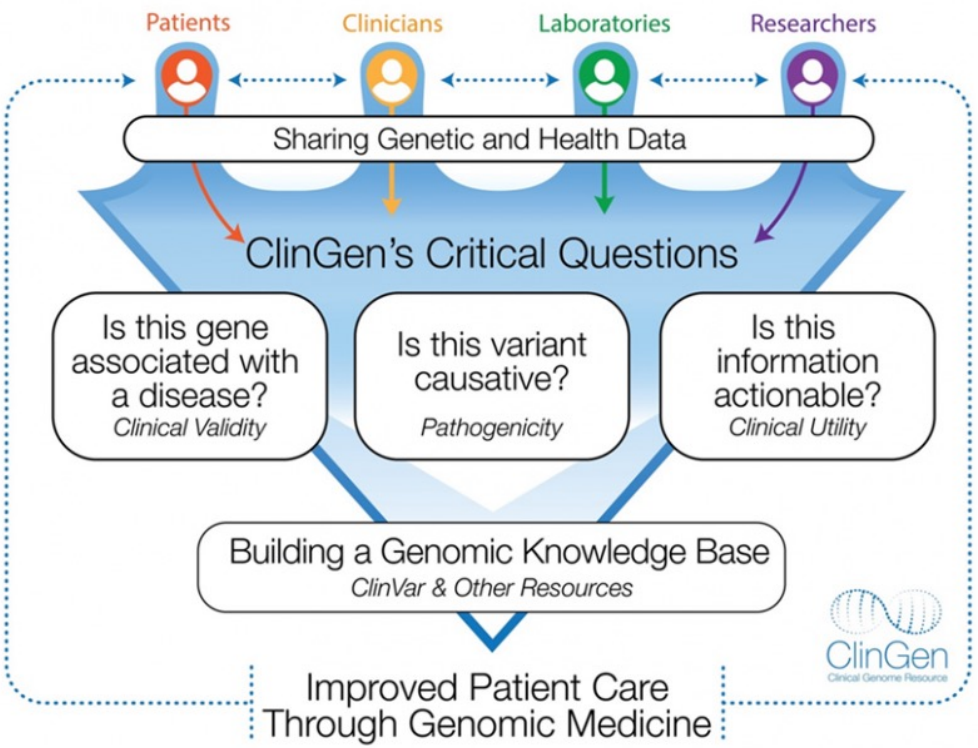
What if only one family affected?

Human population = saturation mutagenesis population?

Areas of uncertainty



Areas of uncertainty



Two areas of uncertainty

1. Does the **gene** cause the disease ?
   – Eg: BRIP1 does ***not*** increase breast cancer risk

   => gene retirement

? GUS ?

2. Does the patient's **variant** alter the gene function ?

? VUS ?

Focused gene panel analysis

Ex: Cystic Kidney Disease

https://panelapp.genomicsengland.co.uk/panels/283/

71 genes, 28 green (GenomicsEngland, UK)

Evidence for gene causality

- **Suggestive or Minimal evidence**
- **Some/Good evidence**
- **High confidence**

GENETIC **VARIANTS**

VARIANT	**Frequency**	**Penetrance (fonctional effect)**
Mutation	Rare	High
Polymorphism	Frequent	Low or none

Polymorphism = frequent genetic variant (MAF >.01 in population)

GENETIC **VARIANTS**

VARIANT	**Frequency**	**Penetrance (fonctional effect)**
Mutation	Rare	High
VUS	Rare	? ?
Polymorphism	Frequent	Low or none
« Rare polymorphism »	Rare	Low or none

VUS = variant of uncertain significance : currently impossible to tell if high penetrance (phenotype-causing, mutation) or low/null penetrance (« rare polymorphism)

VUS classification will require epidemiology of mutation and/or functional data (bioinformatics, machine learning approach)

SNVs and SNPs ; CNVs and CNPs

- Genetic variant affecting one (or few) bp (**SNV**) < sequencing
  - Point mutation
    (ex: point mutation in SCN1A causing Dravet syndrome)
  - Polymorphism : **SNP**
  - VUS

- Copy number variant (**CNV**) < CGH array
  - Mutation
    (ex: chromosomal interstitial deletion causing Williams syndrome)
  - Polymorphism: **CNP**
  - VUS

5 classes of genetic variants from CGH array or sequencing

1. Benign (polymorphism)
2. Probably benign
3. VUS
4. Probably pathogenic
5. Pathogenic (mutation)

5 classes of genetic **variants**

1. Benign (polymorphism)
2. Likely Benign (<10%)
3. **VUS**
4. Likely Pathogenic (>90%)
5. Pathogenic (mutation)

Diagnostic

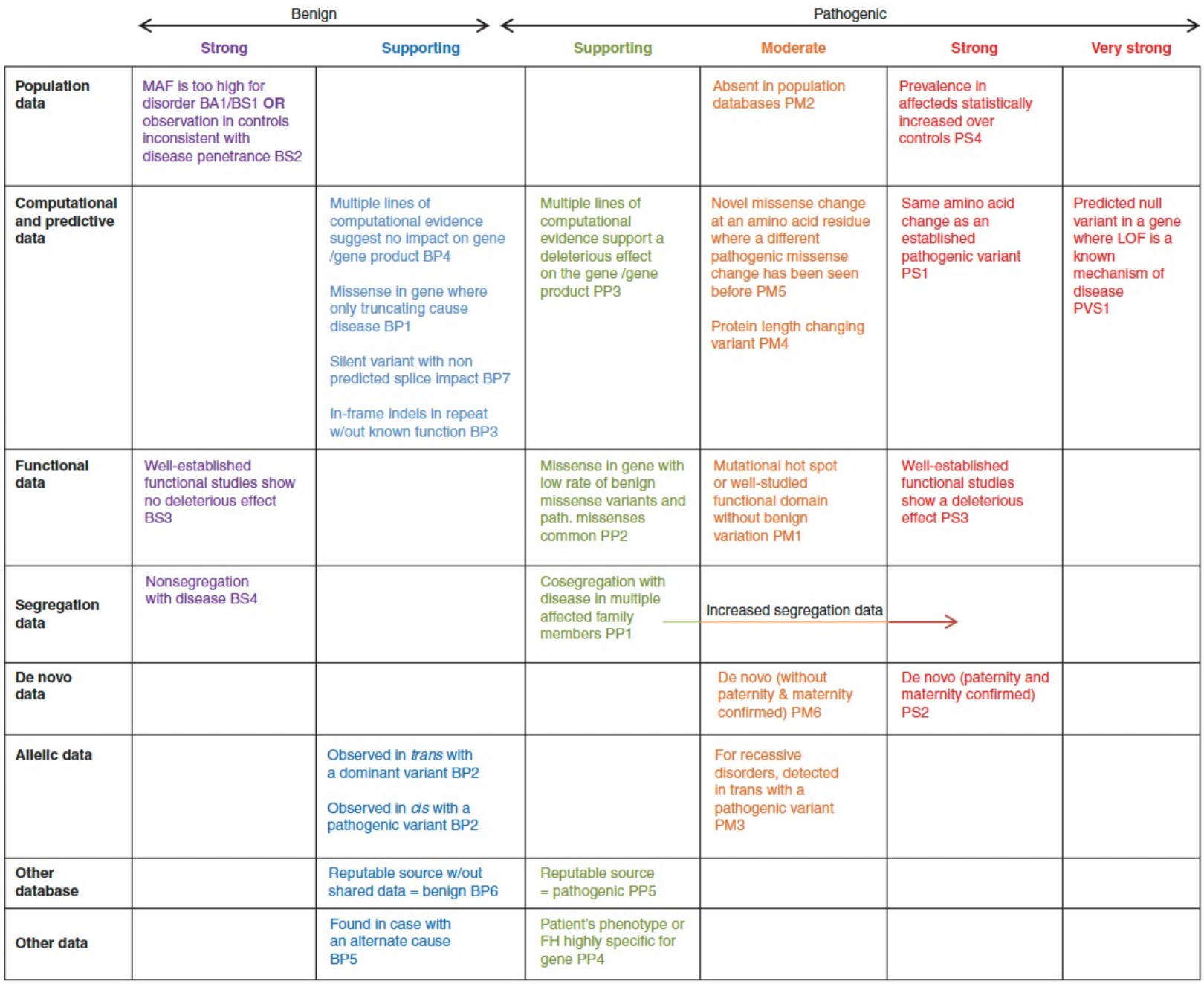
Variants of Uncertain Significance

90% VUS are benign

(false positive results)

Likely = 90%

Richards et al. 2015



	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Figure 1 Evidence framework. This chart organizes each of the criteria by the type of evidence as well as the strength of the criteria for a benign (left side)

Other variant classification systems are coming of age, too

Table 5 Rules for combining criteria to classify sequence variants

Pathogenic	(i) 1 Very strong (PVS1) *AND* (a) ≥1 Strong (PS1–PS4) *OR* (b) ≥2 Moderate (PM1–PM6) *OR* (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) *OR* (d) ≥2 Supporting (PP1–PP5) (ii) ≥2 Strong (PS1–PS4) *OR* (iii) 1 Strong (PS1–PS4) *AND* (a)≥3 Moderate (PM1–PM6) *OR* (b)2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR* (c)1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) *AND* 1 moderate (PM1–PM6) *OR* (ii) 1 Strong (PS1–PS4) *AND* 1–2 moderate (PM1–PM6) *OR* (iii) 1 Strong (PS1–PS4) *AND* ≥2 supporting (PP1–PP5) *OR* (iv) ≥3 Moderate (PM1–PM6) *OR* (v) 2 Moderate (PM1–PM6) *AND* ≥2 supporting (PP1–PP5) *OR* (vi) 1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Benign	(i) 1 Stand-alone (BA1) *OR* (ii) ≥2 Strong (BS1–BS4)
Likely benign	(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) *OR* (ii) ≥2 Supporting (BP1–BP7)
Uncertain significance	(i) Other criteria shown above are not met OR (ii) the criteria for benign and pathogenic are contradictory

VUS : how tell if pathogenic or benign ?

- Functional data
  - In silico : bioinformatics, machine learning
  - Experimental : beyond scope of clinical diagnosis !

- Population data: test many controls
  - Family
  - Local controls
  - Regional
  - National
  - Worldwide

Review

Implementation of genomic arrays in prenatal diagnosis: The Belgian approach to meet the challenges

Olivier Vanakker[a], Catheline Vilain[d], Katrien Janssens[b], Nathalie Van der Aa[b], Guillaume Smits[d], Claude Bandelier[h], Bettina Blaumeiser[b], Saskia Bulk[g], Jean-Hubert Caberg[g], Anne De Leener[d], Marjan De Rademaeker[c], Thomy de Ravel[f], Julie Desir[e], Anne Destree[e], Annelies Dheedene[a], Stéphane Gaillez[g], Bernard Grisart[e], Ann-Cécile Hellin[g], Sandra Janssens[a], Kathelijn Keymolen[c], Björn Menten[a], Bruno Pichon[d], Marie Ravoet[h], Nicole Revencu[h], Sonia Rombout[e], Catherine Staessens[c], Ann Van Den Bogaert[c], Kris Van Den Bogaert[f],

Genetic analysis

Preliminary report

gene | variant | Population db | bio-informatics | Patients db

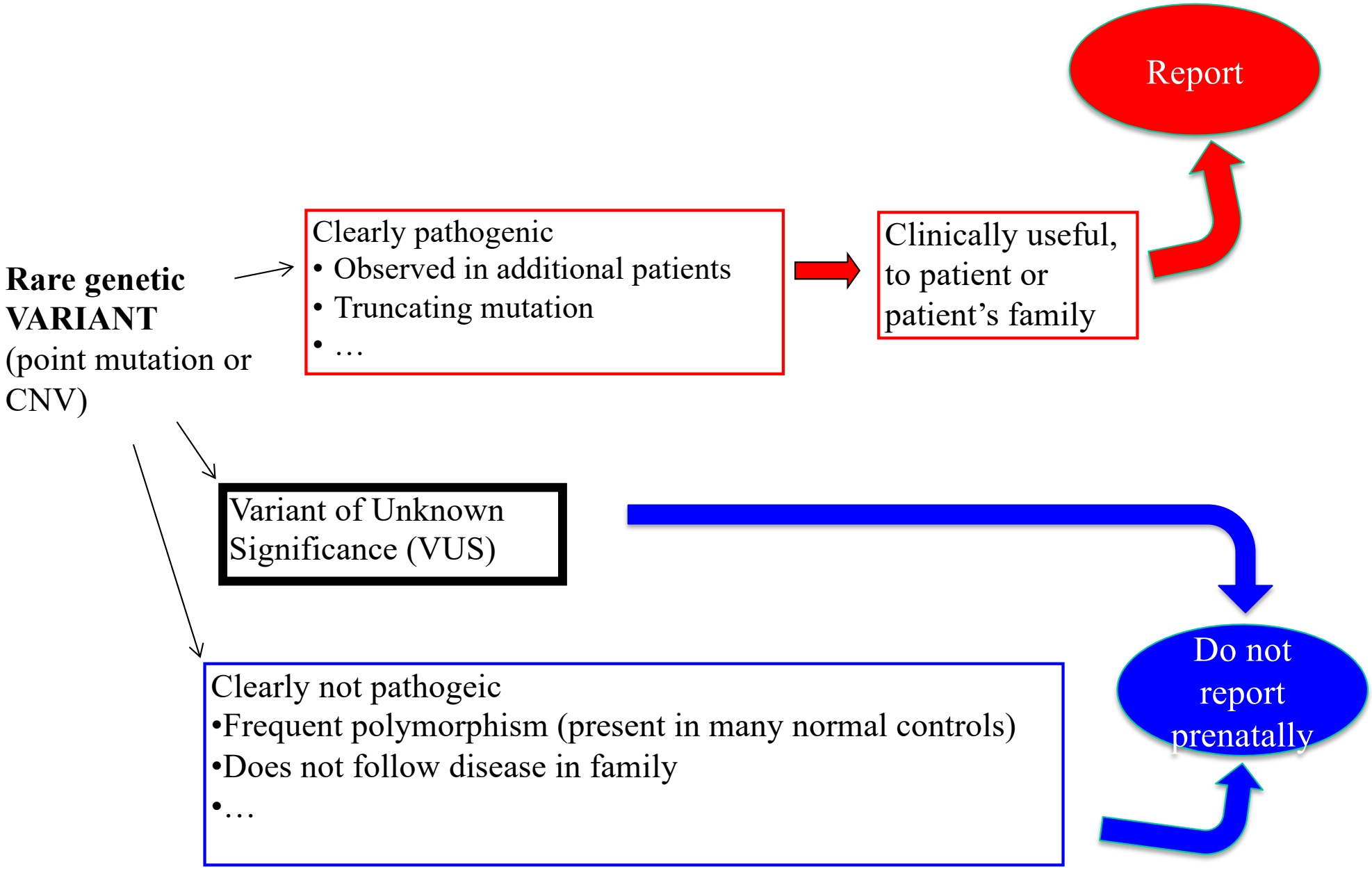
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	J	T	M	Gene	Al	Mutation	Zy	Balan	Mol	CGEN	gnomAD	A	hd	SIF	PP		CA	p/	C	MP	dbscSN	spliceAI	pL	ClinVar	OMIM
11				ATM	ns	NM_000051:exon29:c.4258C>T:p.Leu1420Phe	het	30-23	AD/AR/SMu	69/5124	0,0185555	3103	21	0,07	0,06	D	16	0,64					0	B(8)_LB(3)_VUS(1)	{Breast cancer, susceptibility to}
13				POLRMT	ns	NM_005035:exon10:c.2572C>T:p.Arg858Trp	het	34-50	.	6/5102	0,00308261	67	2		0,67	D	26	0,24		0,95			0	.	.
15				DNA2	ns	NM_001080449:exon1:c.68C>T:p.Ala23Val	het	37-26	AR/AD	6/5110	0,00137079	207	0	0,31	0	B	6	0,29					0	LB(1)_VUS(1)	?Seckel syndrome 8 (AR)/Progres
16	hom			POLG	ns	NM_001126131:exon7:c.1399G>A:p.Ala467Thr	hom	1-70	AD/AR	7/5104	0,000983048	143	0	0	1	D	31	0,65					0	P	Mitochondrial DNA depletion sy
17				MSTO1	sp	NM_018116:exon9:c.966+5G>C	het	36-16	AD/AR	1/5090	0	0	0			.				0,71	1	0,7	0,04	.	Myopathy, mitochondrial, and a
18				PITRM1	fs_ins	NM_001242307:exon24:c.2715dupT:p.Ala906fXaa	het	36-35	.	1/5110	0	0	0			.			97				0	.	.
20																									
21				Coniferix: vide																					
22																									

POLG Mutation

Valproate contra-indicated

cgem@hcuge.ch

Genetic variants (from sequencing / from CGH arrays)



Findings out of scope of initial phenotype

Genome-wide analyses may show variants beyond initial question
= **incidental findings** – unsollicited

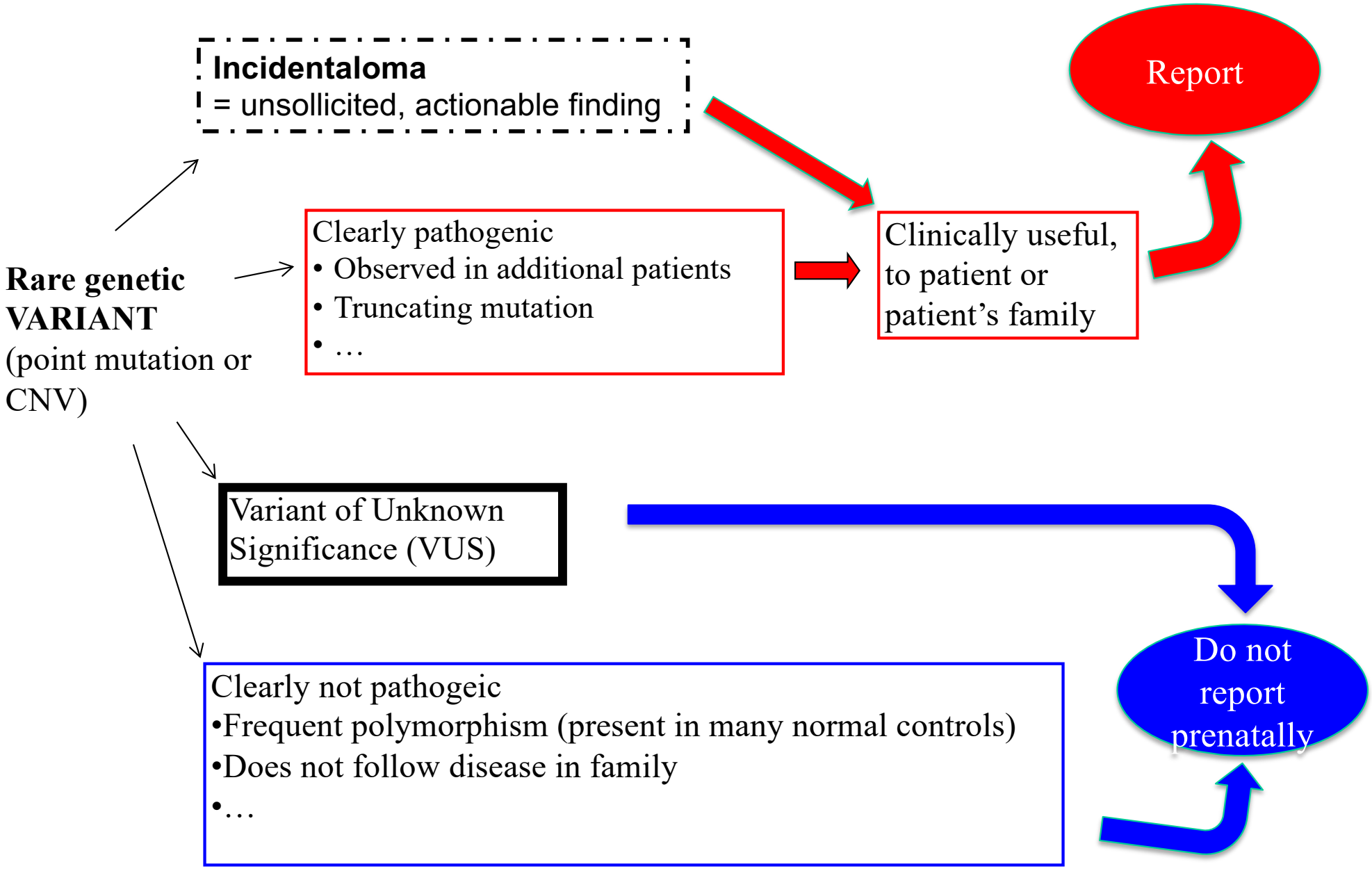
- Ex: child tested for ID => CGH array shows BRCA1 locus deletion (causes breast and ovarian cancers in adults)
- Ex: child tested for ID => exome shows ApoE4 mutation (causes marginal increase in Alzheimer risk)

=> Attitude ?

➢ Consider **actionable** vs non-actionable variant

➢ Opt-in / **opt-out** choice for patient: pretest genetic counseling

Genetic variants (from sequencing / from CGH arrays



Unsollicited and sollicited findings

Incidentaloma

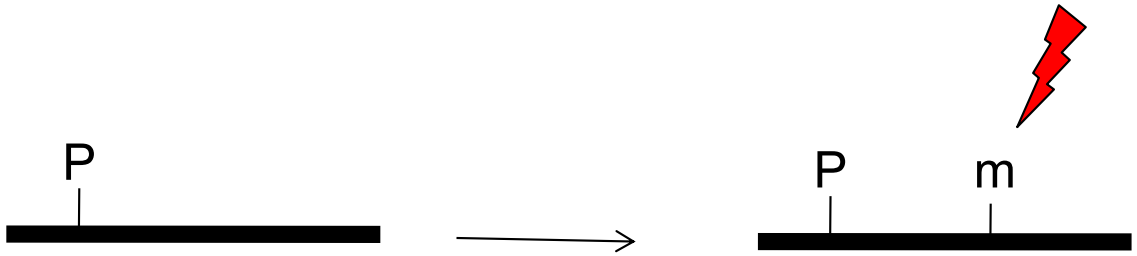
- Unsollicited finding
- Actionable or not
- If actionable, inform patient and offer genetic counseling (patient and family)
- Opt-out procedure (discuss in pre-test genetic counseling)

Secondary variant

- Actionable change
- Sought for, in predefined, international consensus set of genes (~75 genes in 2023)
- In the future, obligation to complete diagnostic-grade analysis of these genes, in any exome/genome sequenced
- Opt-out choice (pre-test counseling)
- Post-test counseling, patient and family, if positive

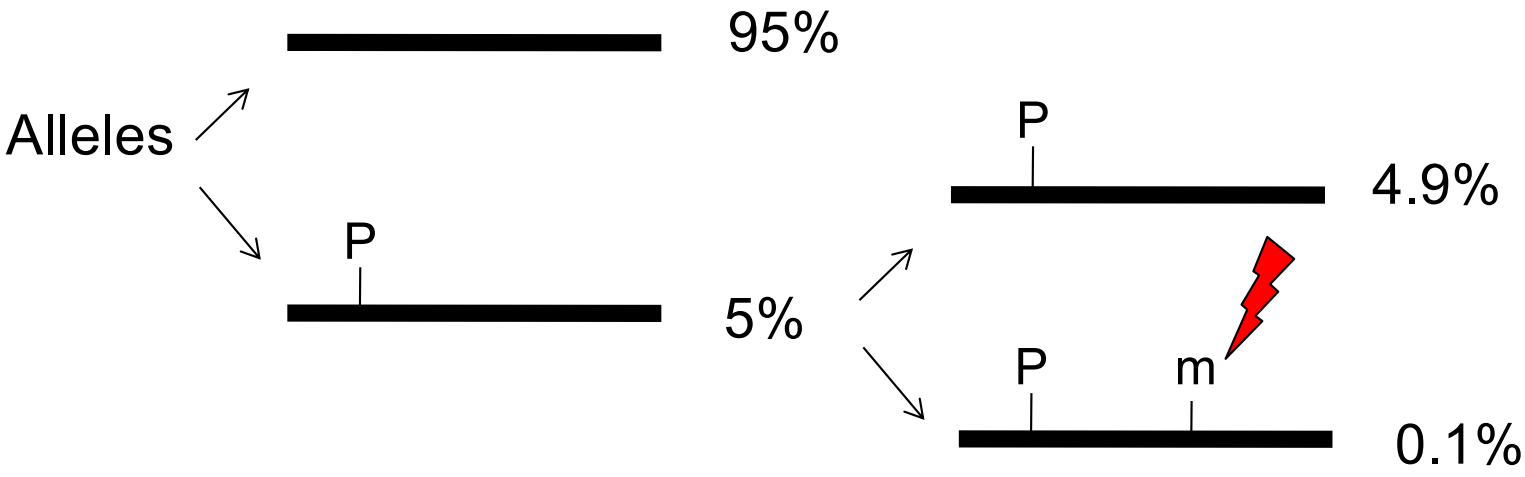
COMPLEX ALLELES

Polymorphism and mutation may coexist on same allele



- Here, a mutation (m) appeared on an allele that already carried a polymorphism (P)

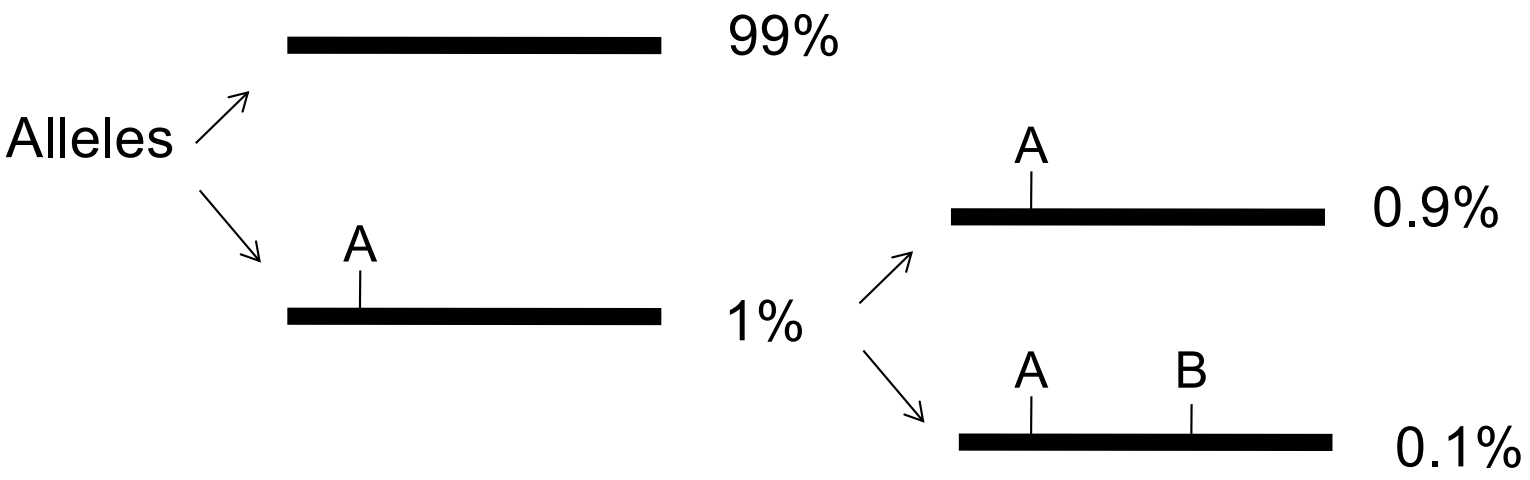
Polymorphism and mutation may coexist on same allele



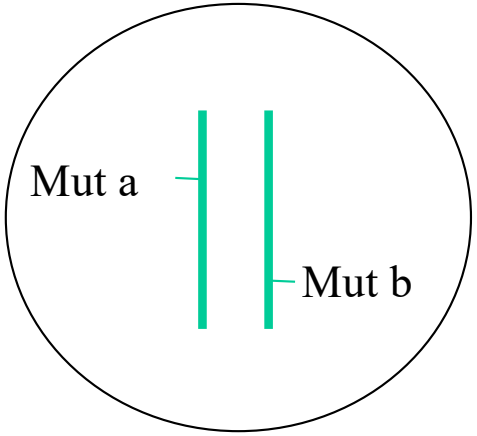
- Mutated alleles are rare : 0.1% in this example
- P is known and frequent, hence no problem in interpreting m as a possible disease causing mutation.

- If P was rare, it might be hard to tell which of the 2 rare variants, P and m, is disease-causing: « complex allele » (next slide)

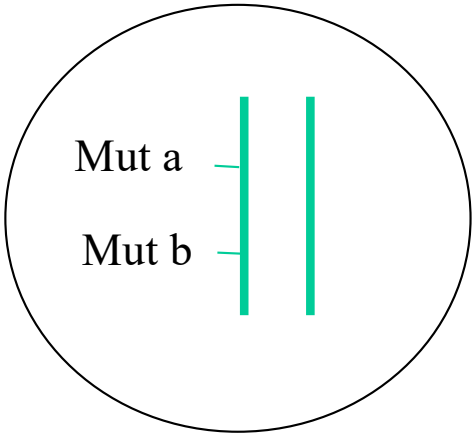
Complex alleles with 2 rare variants



2 rare variants may lie on same allele (in cis)



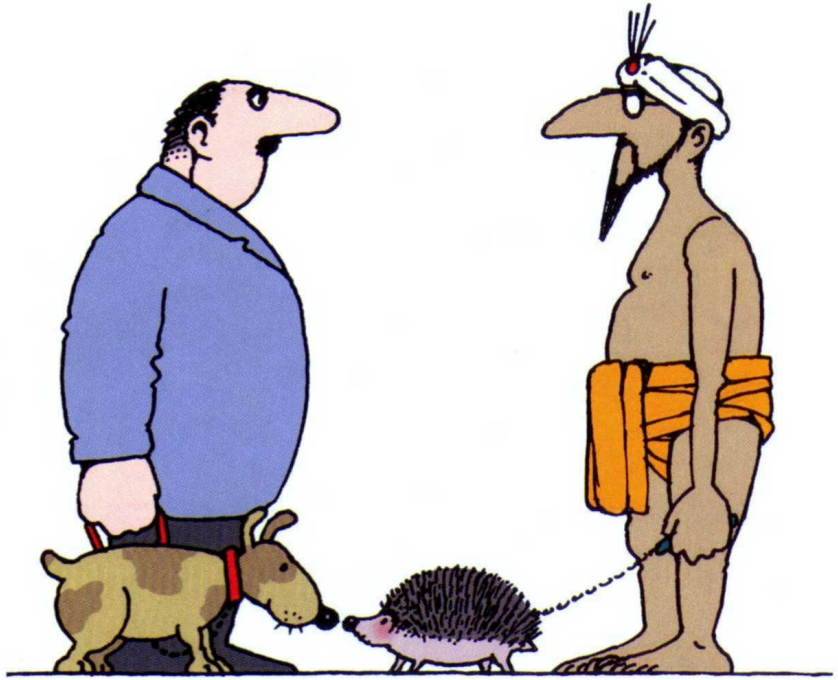
Mut a and Mut b
< both parents



Mut a and Mut b
< same parent

- In Autosomal Recessive disease, make sure Mut a et Mut b are biallelic = in trans (left panel)
- ! If mut a et b are in cis, the mutation of 2nd allele remains unidentified (right panel) !

GENETIC VARIATION IN POPULATIONS



Populations are very polymorphic

- Individuals are all different
- genetic (and epigenetic) polymorphism
- Reveal our differences
  - Identity
  - Family links
  - Historical, geopolitical links
  - On-going evolution, adaptive changes

Human populations

- No races, but
- Sub-populations (« ethnic groups »)
- **Common ancestors**, close or distant, between all humans

CFTR*DF508 ; HBB*sickle; ethnic groups

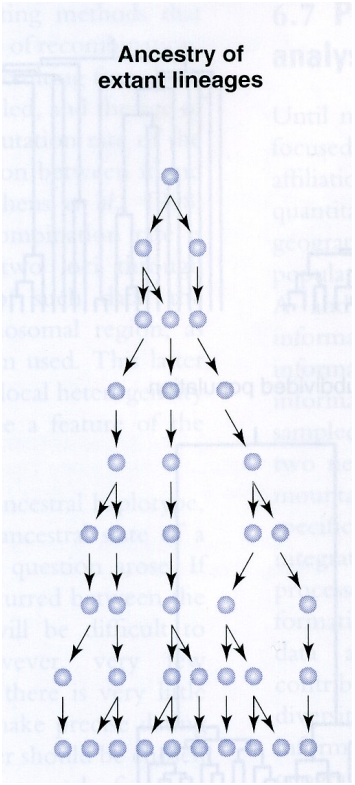
- CF more frequent in the Northern populations (3% carry DF508)
- Sickle cell more frequent in Central Africa (10% carry drepano)

Ethnic prevalence of ancestral mutations

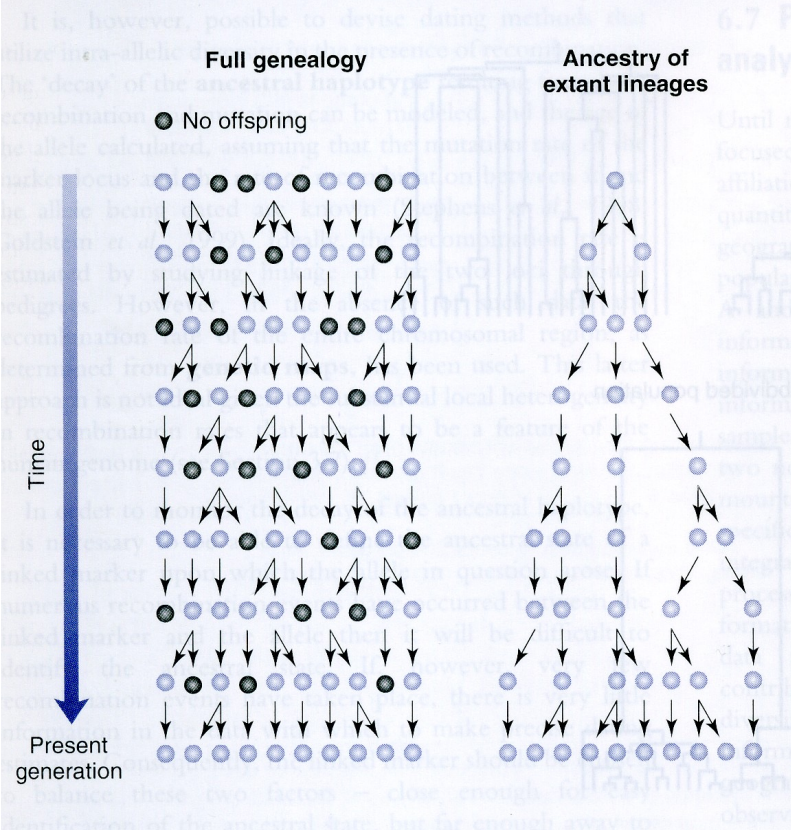
- Race = group of individuals defined by common biological characteristics, different from other group (mice).
  No race in human species.
  Human groups mix and depart constantly.

- **Ethny**: human group caracterized by biological ancestrality and/or by common language, religion, culture…
  Ill-defined borders.

Most Recent Common Ancestor

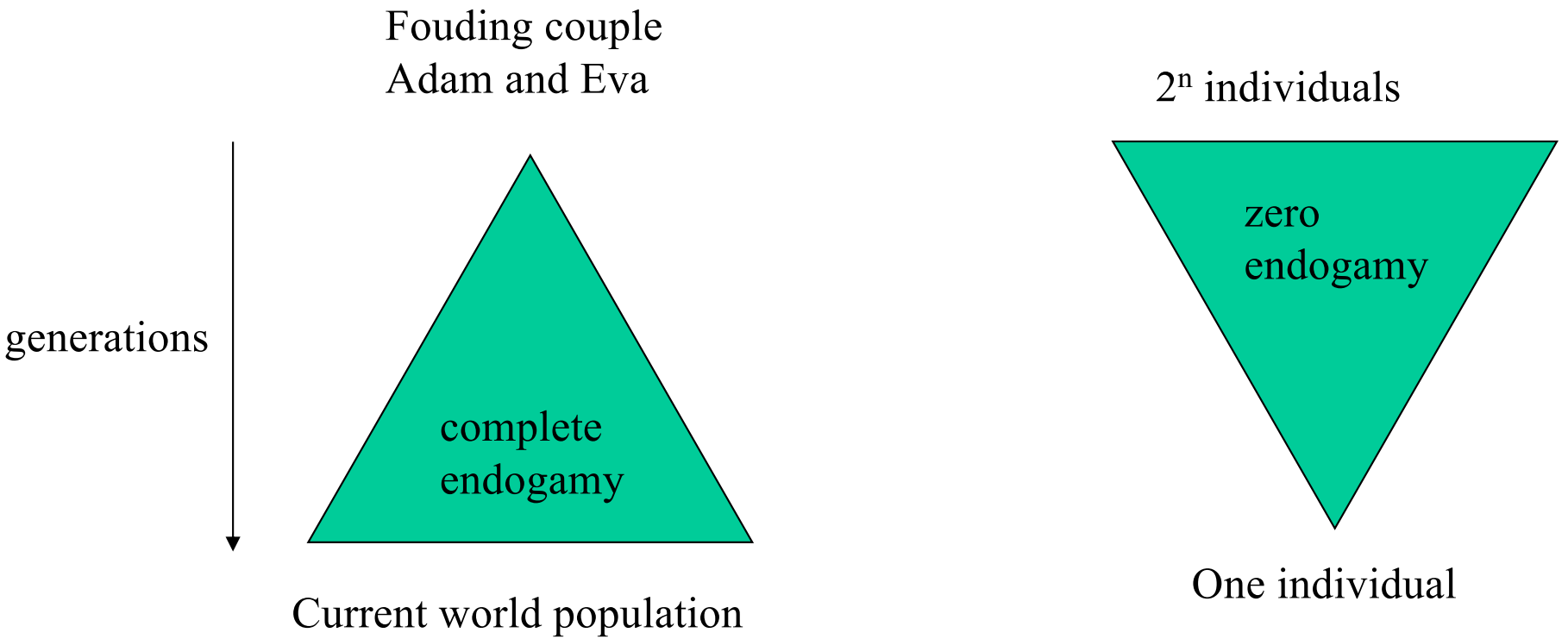


Most Recent Common Ancestor



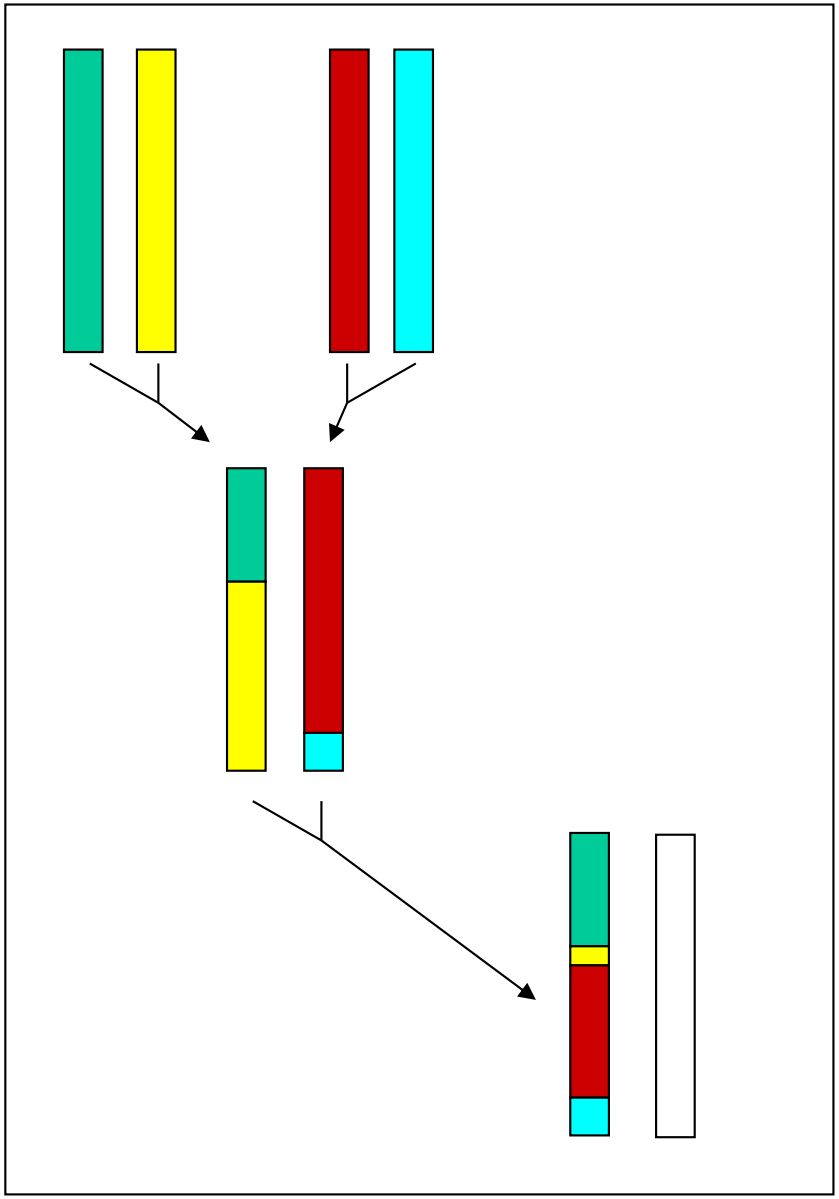
Population of constant size over generations

genealogies



- Reality = mixture of both
- In a constant-sized population, every 2 individuals are related through a paternal and a maternal MOST RECENT COMMON ANCESTOR

The origin of genetic diversity



1. MUTATION : diversity by change

   pieces of homologuous chromosomes differ

2. MEIOSIS : diversity by assembly (crossing-overs)

   pieces are re-shuffled

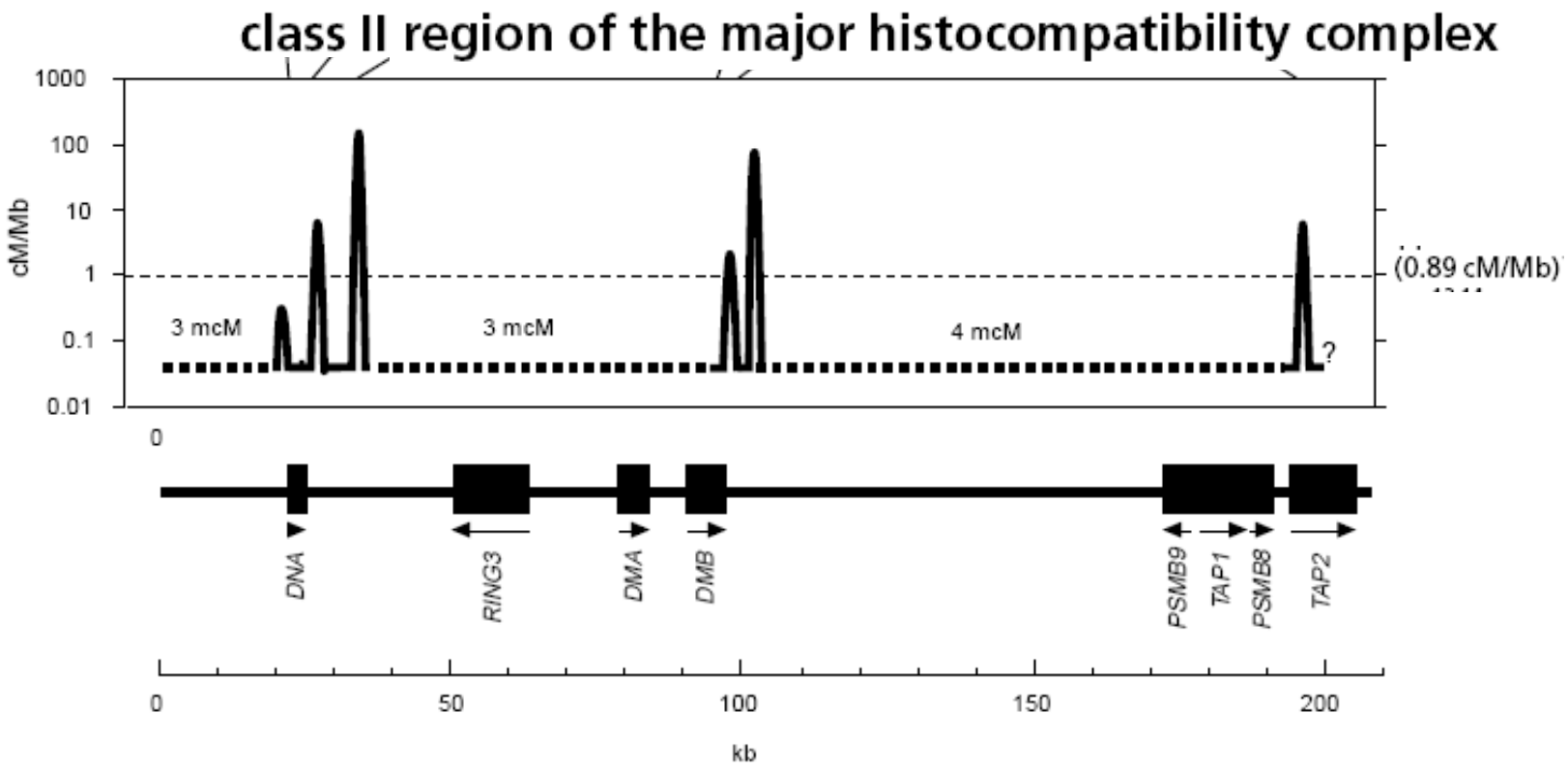
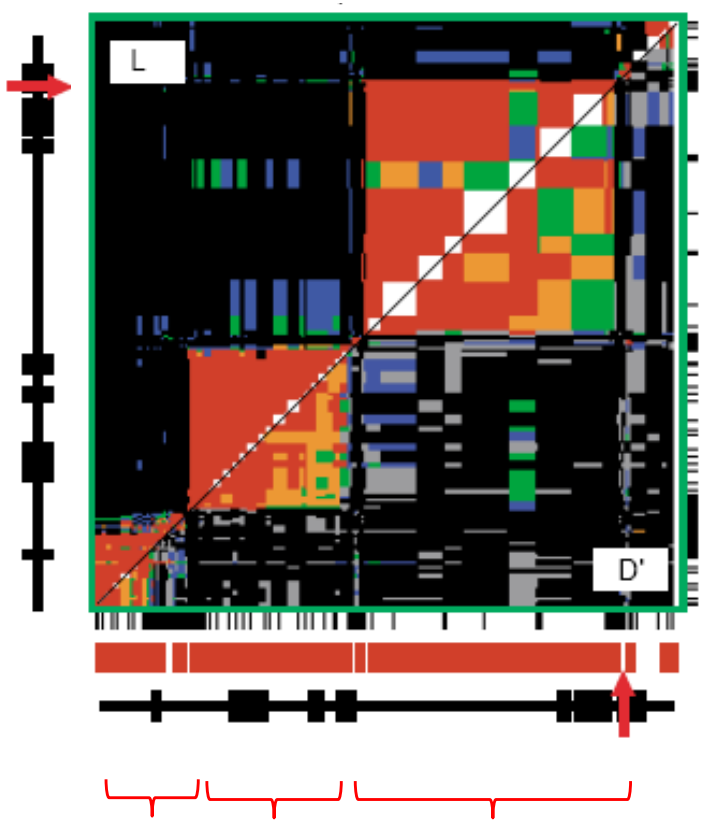


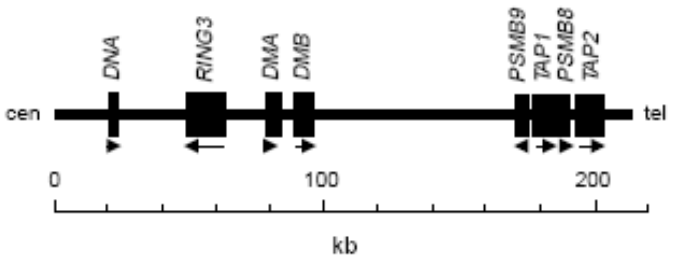
Fig. 5 Sperm crossover activity in the class II region of the MHC. The number of men tested and the total number of sperm crossovers mapped are given for each hot spot, together with approximate hot-spot center coordinate in the consensus sequence of the human MHC[10]. The width of each hot spot, within which 95% of crossovers occur, was determined by normal-distribution fitting (Fig. 3). The mean male linkage map distance contributed by each hot spot, plus range seen in the different men tested, was determined from the observed hot spot crossover frequency per sperm and is given in millicentimorgans (mcM, $cM \times 10^{-3}$); only the hot spot *DNA* 2 shows significant variation in activity between tested men. Inter–hot spot distances were estimated from data in Fig. 4. The background recombination rate of 0.04 cM/Mb is very approximate and should be treated with caution. The mean rate of male meiotic recombination in the human genome (0.89 cM/Mb)[16] is shown as a thin dashed line. *TAP2* and minisatellite MS32 estimates were from data published elsewhere[12,14].

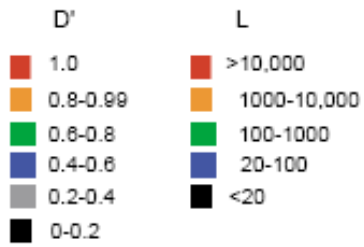
Jeffreys et al. **nature genetics • volume 29 • october 2001**

Intensely punctate meiotic recombination In MHC

Haplotype blocks





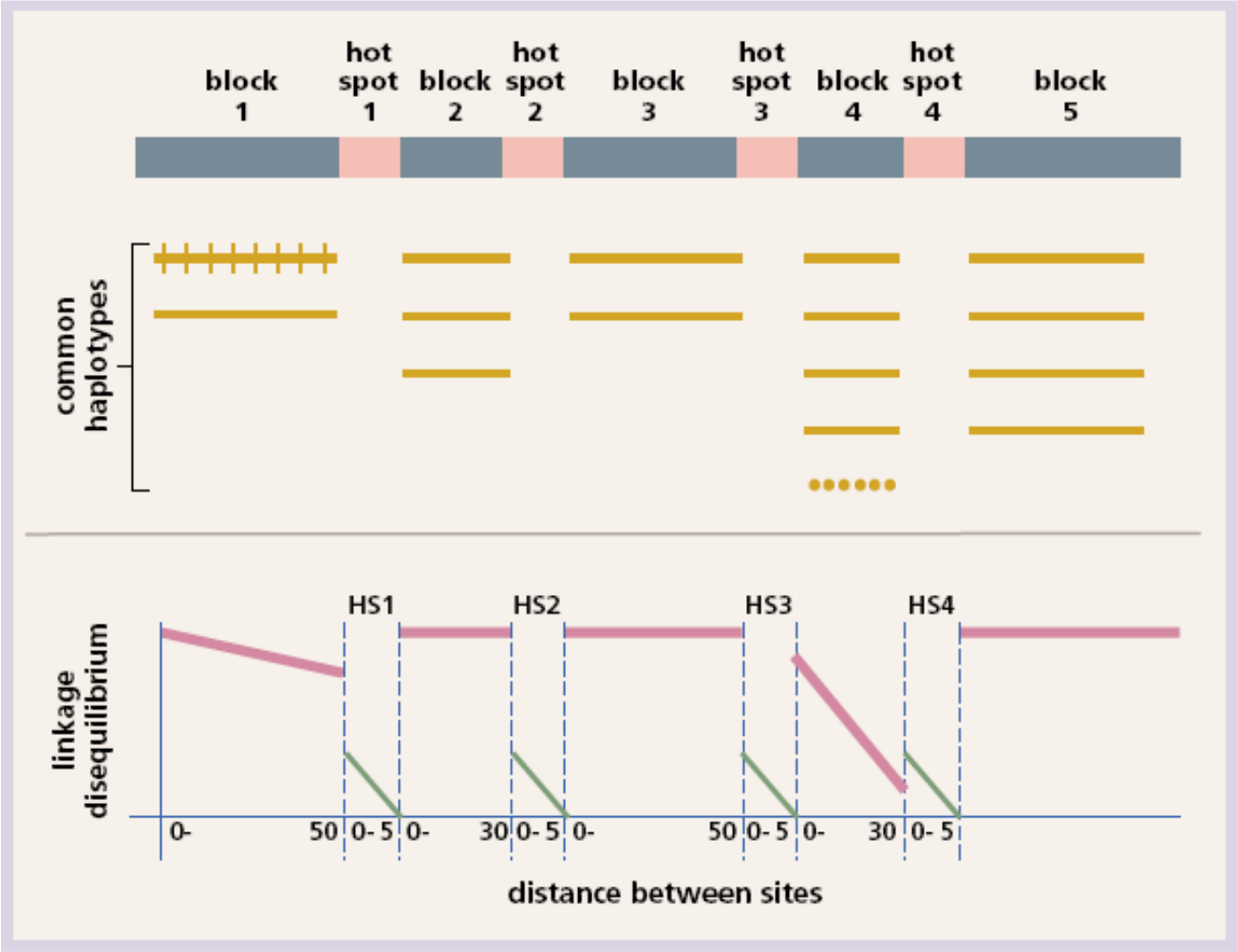


D' and L are measures of LD

Crossover hotspot in *TAP2* gene (known)

Jeffreys et al. 2001

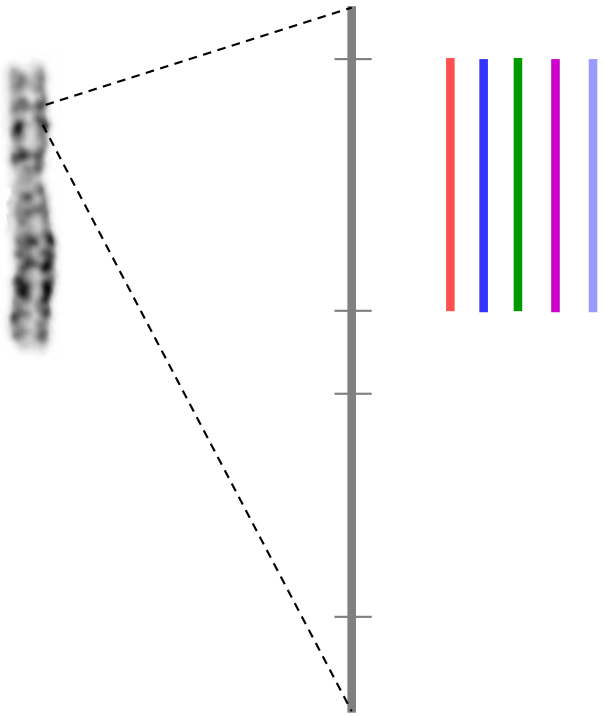
Haplotype block <=> absolute LD



The lowdown on LD. Idealized representation of block-like structure of linkage disequilibrium, with regions of low haplotype diversity separated by recombinational hot spots. Lines below the blocks represent examples of the number of common haplotypes that might be present for such blocks. SNPs distinguishing the two common haplotypes in block 1 are represented by short vertical lines. The graphs plot (idealized) LD as a function of distance, averaged across pairs of sites, either for sites within a given block or within a hot spot. The plots show that within a block LD decays only gradually with distance, or not at all. Within hot-spot areas, however, LD falls away much more rapidly with distance. If no LD-generating event, such as a bottleneck, has recently occurred in the population, then there may have been enough recombination across the hot spots that the haplotypes in adjacent blocks are randomly associated. Similarly, with sufficient time, or in blocks with higher within-block recombination rates, LD may be substantially reduced for distant sites within a block, as represented here in block 4. Note that for block 1, any of the SNPs indicated would be sufficient to represent the majority of the haplotypic variation within this block. If haplotype 1 were shown to increase the risk of a condition relative to haplotype 2, however, it would be impossible to determine from association data which of the SNPs distinguishing haplotypes 1 and 2 was the biological cause of the increased risk.

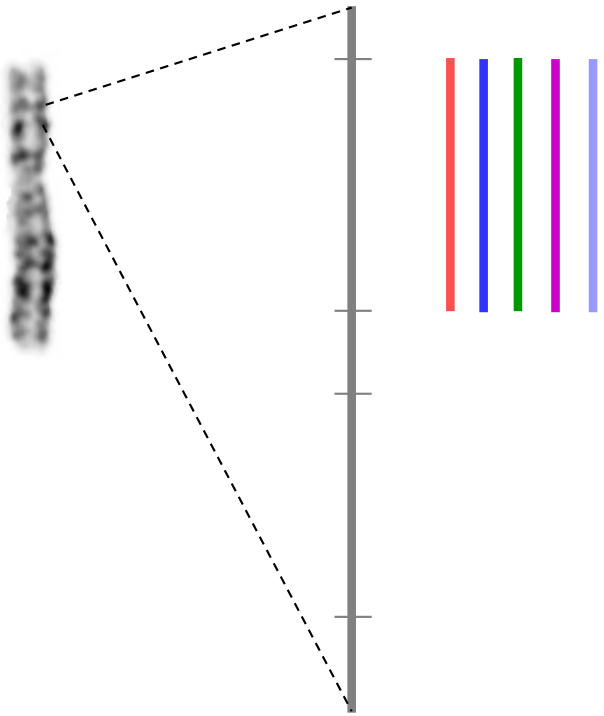
Islands of Linkage Disequilibrium. Goldstein DB 2001 Nat Genet

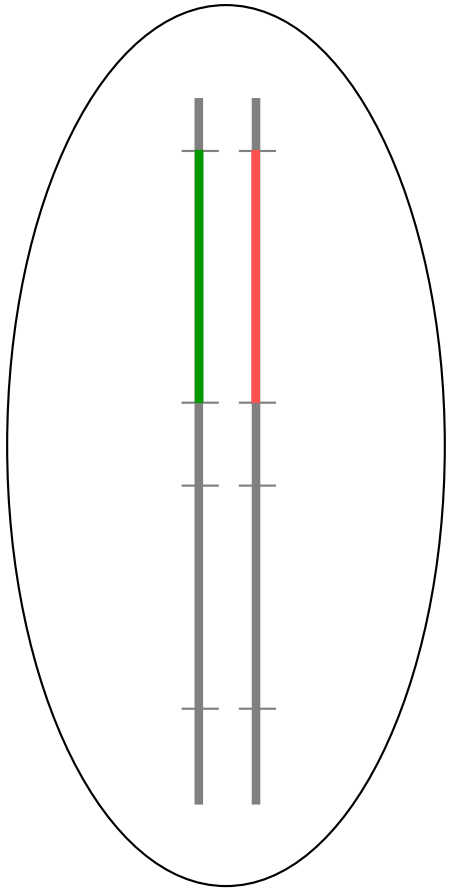
Haplotype blocks



- Whole genome = 100,000 blocks,
  with a few haplotypes in population.

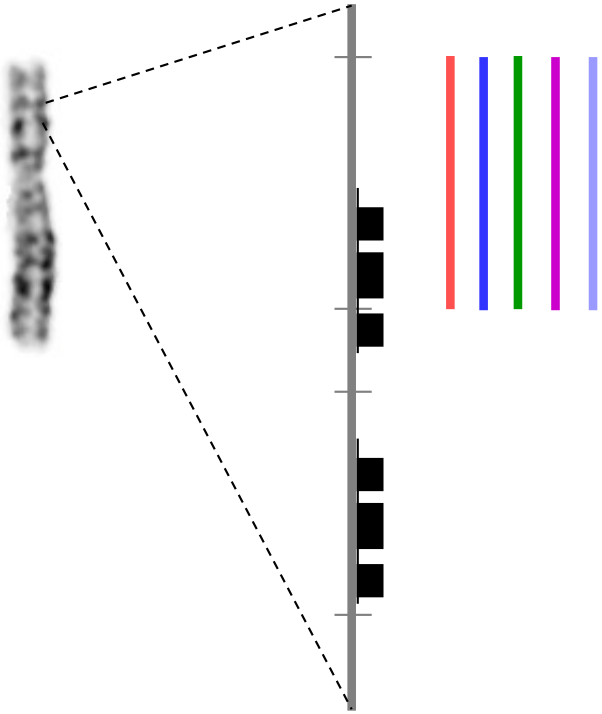
Haplotype blocks

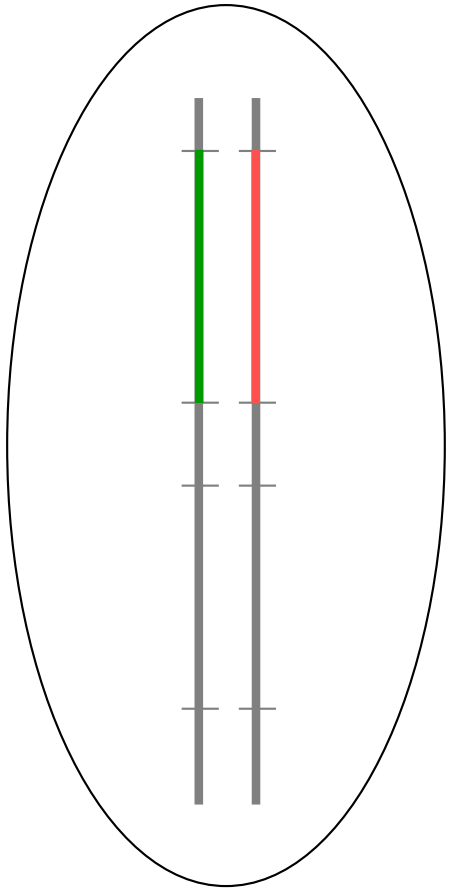




- Whole genome = 100,000 blocks, with a few haplotypes in population.
- Everyone has 100,000 x 2 haplotypes

Haplotype blocks

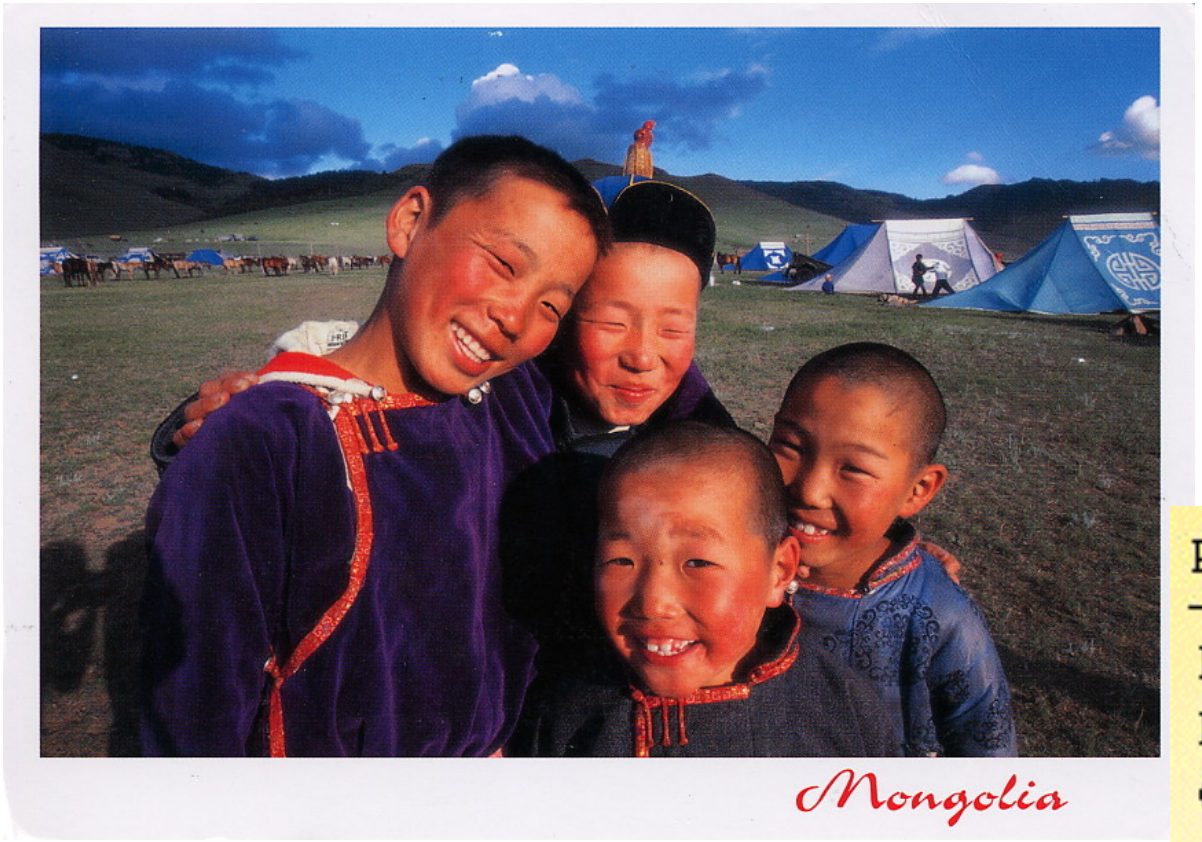




- Whole genome = 100,000 blocks, with a few haplotypes in population.
- Everyone has 100,000 x 2 haplotypes
- Genes may overlap blocks

Mongolia

MN blood group



Population	*MM*	*MN*	*NN*
Eskimo	0.835	0.156	0.009
Egyptian	0.278	0.489	0.233
Chinese	0.332	0.486	0.182
Australian aborigine	0.024	0.304	0.672

Hardy – Weinberg law

- The frequency of the three genotypes AA, Aa and aa is given by the tems of the binomial expansion of $(p+q)^2$
  $= p^2 + 2pq + q^2$
- And does not change over generations
- Under certain conditions :
  - Random matings
  - No mutation
  - No selection
  - No drift
  - No migration in or out
  - Equal generations
  - Stable population

Hardy Weinberg

Consider dimorphic locus : 2 alleles, A or a

Population, N = 10000

- 8000 individuals are AA
- 2000 individuals are aa

=> is this population in HW equilibrium?

! Phenotypes are not considered here !

A is not dominant and a is not recessive

Hardy Weinberg

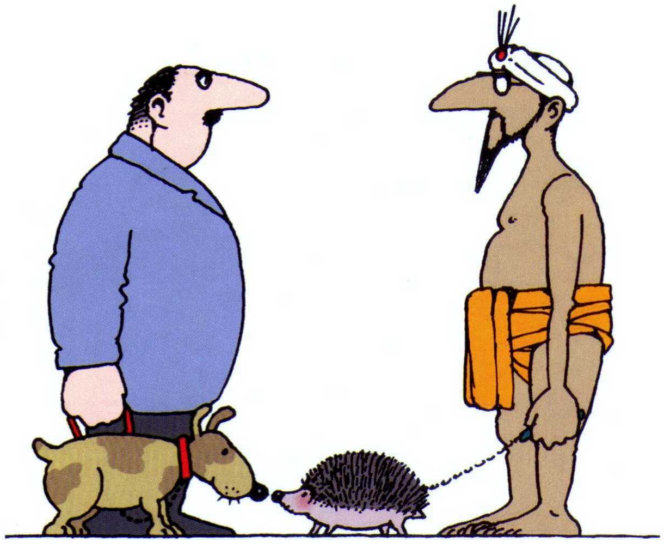
Consider dimorphic locus : 2 alleles, A or a

Population, N = 10000

- 8000 individuals are AA
- 2000 individuals are aa

=> is this population in HW equilibrium?

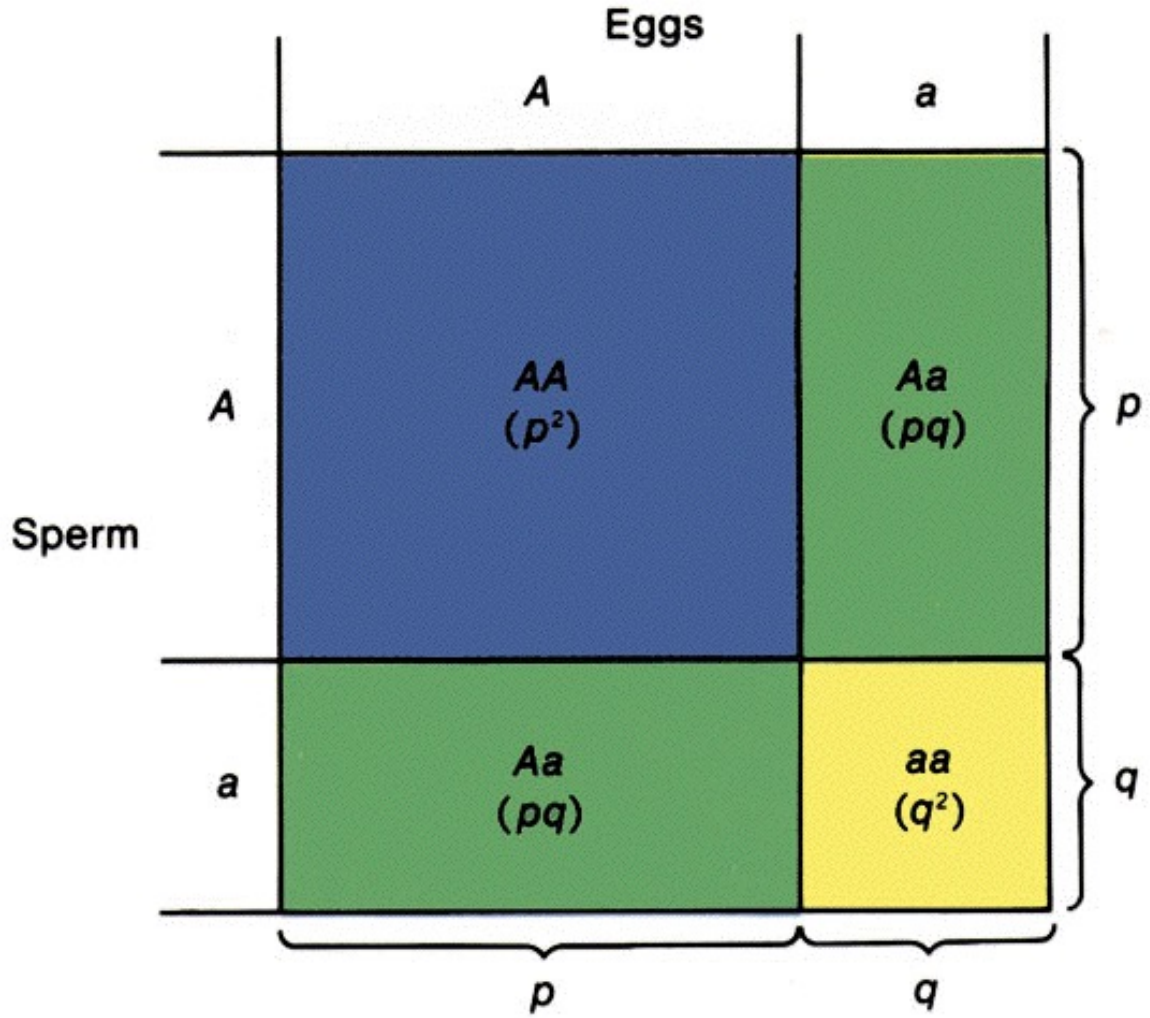
Of course not ! Heterozygotes are lacking
This is a mixture of 2 independent populations



! Phenotypes are not considered here !

A is not dominant and a is not recessive

Punnett square for population (vs family)



Equilibrium reached in 1 generation

- N=10000; 8000 are AA, 2000 are aa, 0 are htz : NOT in equlibrium
- Alleles: 16000 A, 4000 a (p=16000/20000=0.8, q=0.2)

1. *Measure allele frequencies*
2. *Compute expected frequencies of the genotypes*

Equilibrium reached in 1 generation

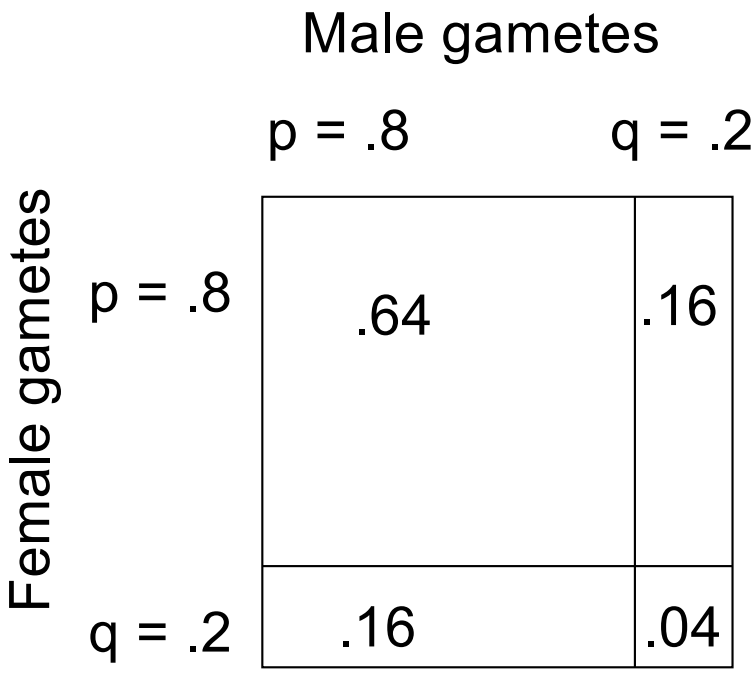
- N=10000; 8000 are AA, 2000 are aa, 0 are htz : NOT in equlibrium
- Alleles: 16000 A, 4000 a (p=16000/20000=0.8, q=0.2)
- Gametes: 16000 A, 4000 a
- Next generation: Punnett square: => 6400 AA, 3200 Aa, 400 aa

1. *Measure allele frequencies*
2. *Compute expected frequencies of the genotypes*

		Male gametes	
		p = .8	q = .2
Female gametes	p = .8	.64	.16
	q = .2	.16	.04

Equilibrium reached in 1 generation

- N=10000; 8000 are AA, 2000 are aa,
  0 are htz : NOT in equlibrium
- Alleles: 16000 A, 4000 a
  (p=16000/20000=0.8, q=0.2)
- Gametes: 16000 A, 4000 a
- Next generation: Punnett square, N stable
  => 6400 AA, 3200 Aa, 400 aa
- Alleles:
  12800 A + 3200 A = 16000 A
  800 a + 3200 a = 4000 a
  (p=0.8, q=0.2)
- Gametes: 16000 A, 4000 a



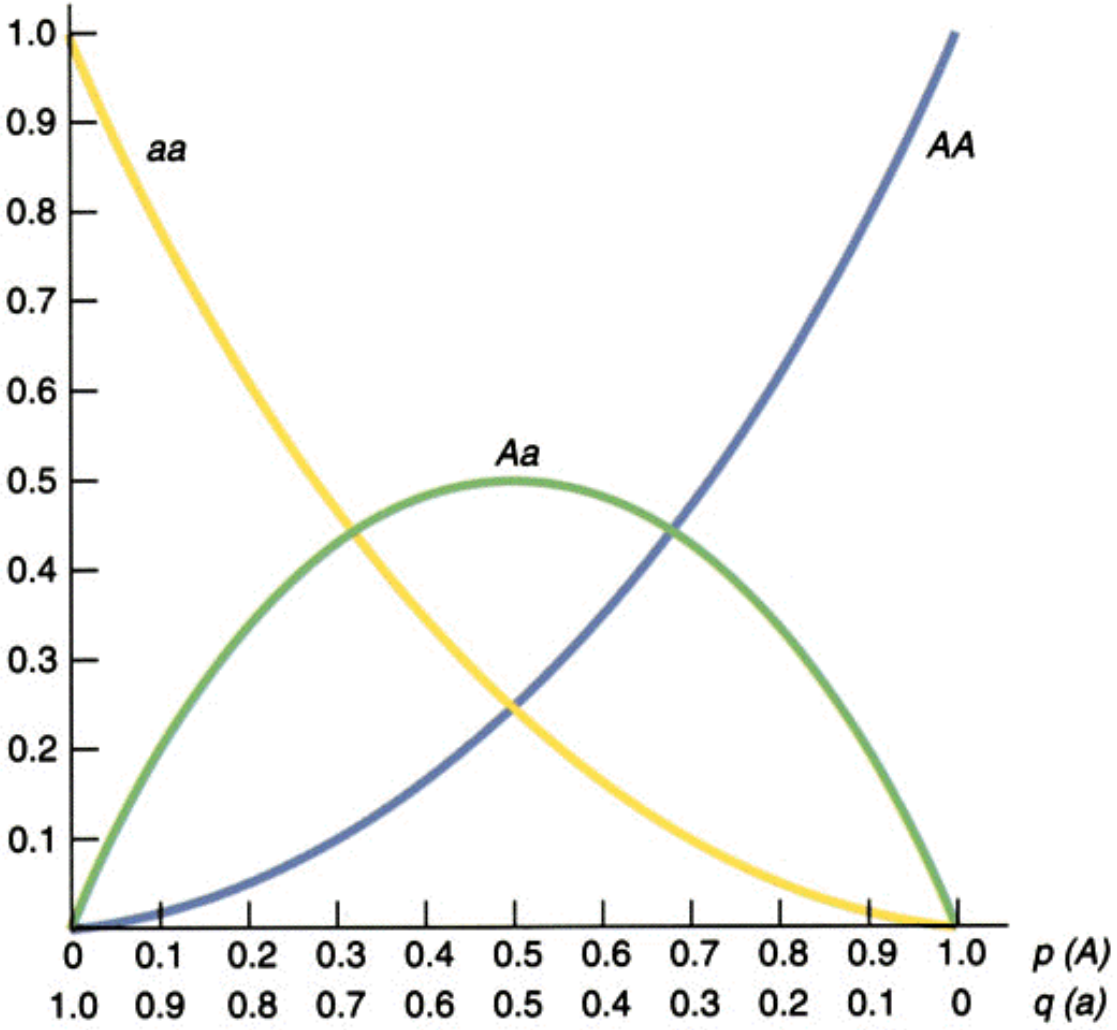
Equilibrium reached in 1 generation

- N=10000; 8000 are AA, 2000 are aa,
  0 are htz : NOT in equlibrium
- Alleles: 16000 A, 4000 a
  (p=16000/20000=0.8, q=0.2)
- **Gametes: 16000 A, 4000 a**
- Next generation: Punnett square, N stable
  => 6400 AA, 3200 Aa, 400 aa
- Alleles:
  12800 A + 3200 A = 16000 A
  800 a + 3200 a = 4000 a
  (p=0.8, q=0.2)
- **Gametes: 16000 A, 4000 a**
- Next generation: Punnett square, N stable
  => 6400 AA, 3200 Aa, 400 aa

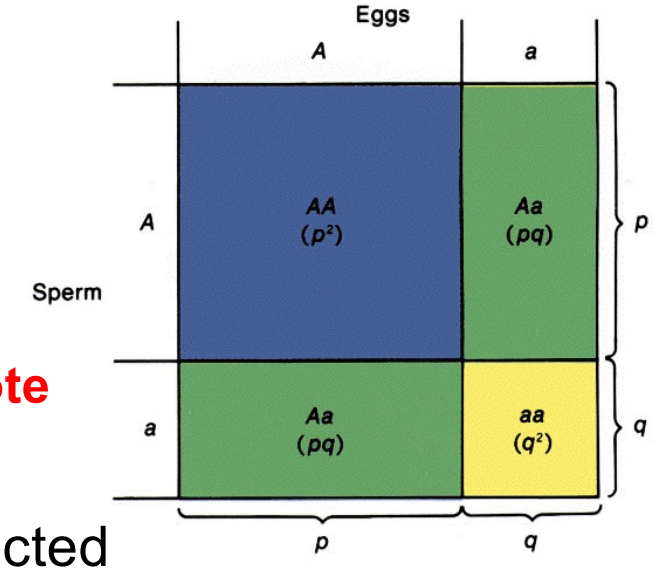
Male gametes

Female gametes	p = .8	q = .2
p = .8	.64	.16
q = .2	.16	.04

Allele proportions at equilibrium



HWE



- ✓ Measure frequency of **ALLELES**
  - ➢ **2 alleles per homozygote + 1 per heterozygote**
- ✓ Knowing allele frequencies, compute the expected frequencies of the **GENOTYPES**
  - ➢ **$p^2 + 2pq + q^2$**
- ✓ Check if observed frequencies of genotypes match the expected frequencies
  - – If yes, alleles are at HW equilibrium
  - – If not => find why they aren't

CCR5 alleles in one population

Genotype	indiv
CCR5/CCR5	647
CCR5/ΔCCR5	134
ΔCCR5/ΔCCR5	7
Total individuals:	788

Total alleles = 2 x 788 = 1576

CCR5 alleles in one population

Genotype	indiv	Genotype Frequencies
CCR5/CCR5	647	647 / 788 = .821
CCR5/ΔCCR5	134	134 / 788 = .170
ΔCCR5/ΔCCR5	7	7 / 788 = .009
Total individuals:	788	1.000

Total alleles = 2 x 788 = 1576

CCR5 alleles in one population

Genotype	indiv	Genotype Frequencies
CCR5/CCR5	647	647 / 788 = .821
CCR5/ΔCCR5	134	134 / 788 = .170
ΔCCR5/ΔCCR5	7	7 / 788 = .009
Total individuals:	788	1.000

Total alleles = 2 x 788 = 1576

Allele frequency:

CCR5: 2 x 647 + 1 x 134 = 1428 => 1428 / 1576 = 0.906

ΔCCR5: 2 x 7 + 134 = 148 => 148 / 1576 = 0.094

Are the genotypes in Hardy – Weinberg equilibrium?

$.906^2 = .821$; $.094^2 = 0.009$; 2 x .906 x .094 = 170 yes

CCR5 alleles in one population

Genotype	indiv	Genotype Frequencies
CCR5/CCR5	647	647 / 788 = .821
CCR5/ΔCCR5	134	134 / 788 = .170
ΔCCR5/ΔCCR5	7	7 / 788 = .009
Total individuals:	788	1.000

Total alleles = 2 x 788 = 1576

Allele frequencies:

CCR5: 2 x 647 + 1 x 134 = 1428 => 1428 / 1576 = 0.906

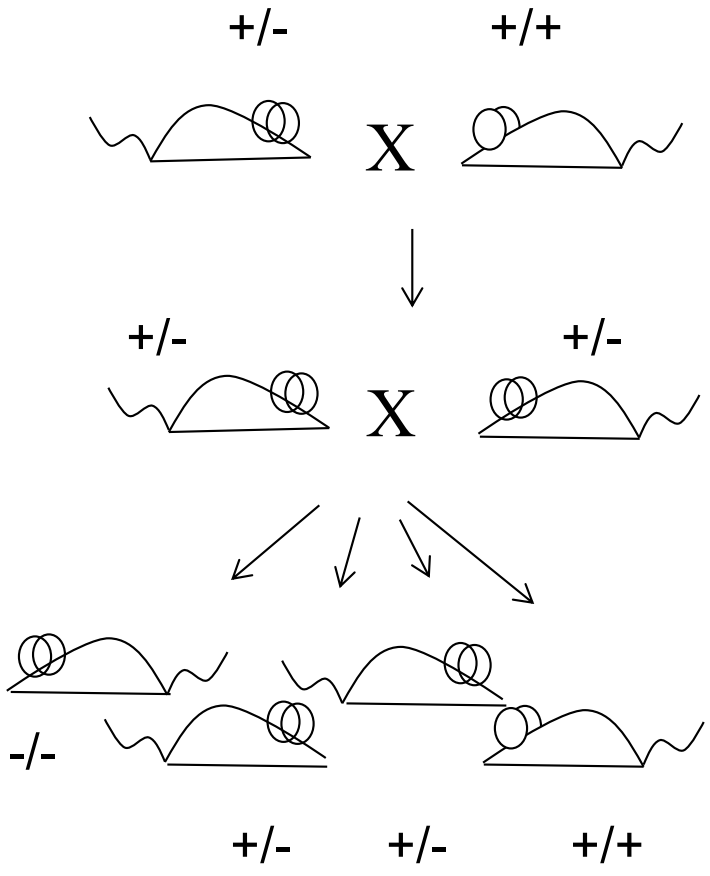
ΔCCR5: 2 x 7 + 134 = 148 => 148 / 1576 = 0.094

Are the genotypes in Hardy – Weinberg equilibrium?

$.906^2 = .821$; $.094^2 = 0.09$; 2 x .906 x .094 = 170

(seems almost too exactly right to be true!!)

You make a strain of KO mice for a transcription factor and count the number of homozygotes and heterozygotes in F2

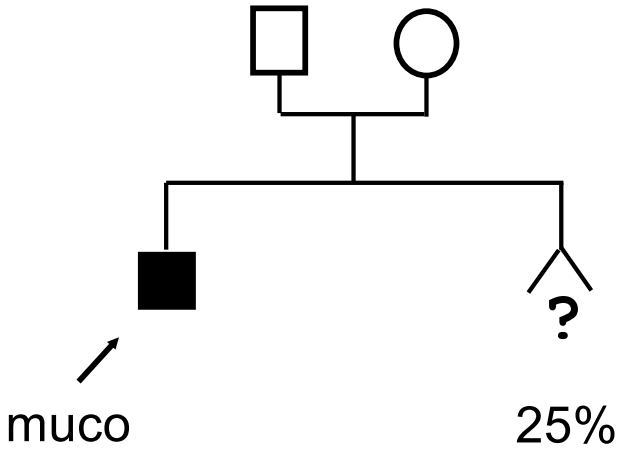


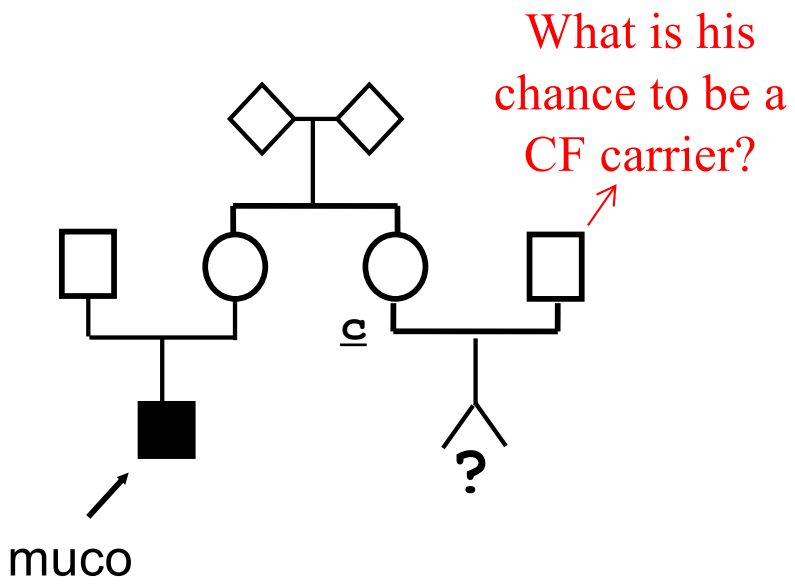
The observed distribution is not expected, a load of -/- homozygotes are missing => suggests embryonic lethality in homozygous -/- KO.

+/+	+/-	-/-
32%	64%	3%

Cystic fibrosis affects 1 newborn in 2500

=> what is the risk of CF in the 2 following future children?



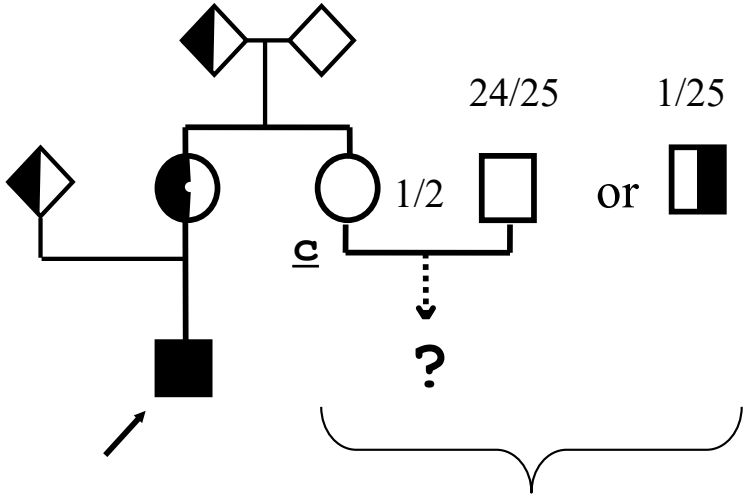


HW in AR disease: eg CF

- CF (aa) 1/2,500
- q^2 = 1/2,500
- q=0.02, p=.98
- htz (Aa) = 2pq = 0.04 = 1/25

Check: 4% carriers
=> 1/25 x 1/25 x 1/4 affected
= 1/2,500 affected newborns

(Selection acts on very few individuals (1/2,500) => discard)



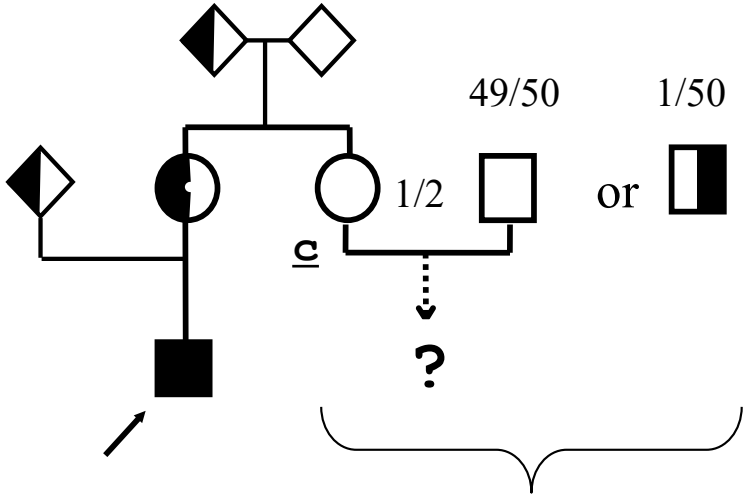
Offspring risk
(a priori) = 1/200

HW in AR disease: eg PKU

- PKU (aa) 1/10,000
- q^2 = 1/10,000
- q=0.01, p=.99
- htz (Aa) = 2pq = 0.02 = 1/50

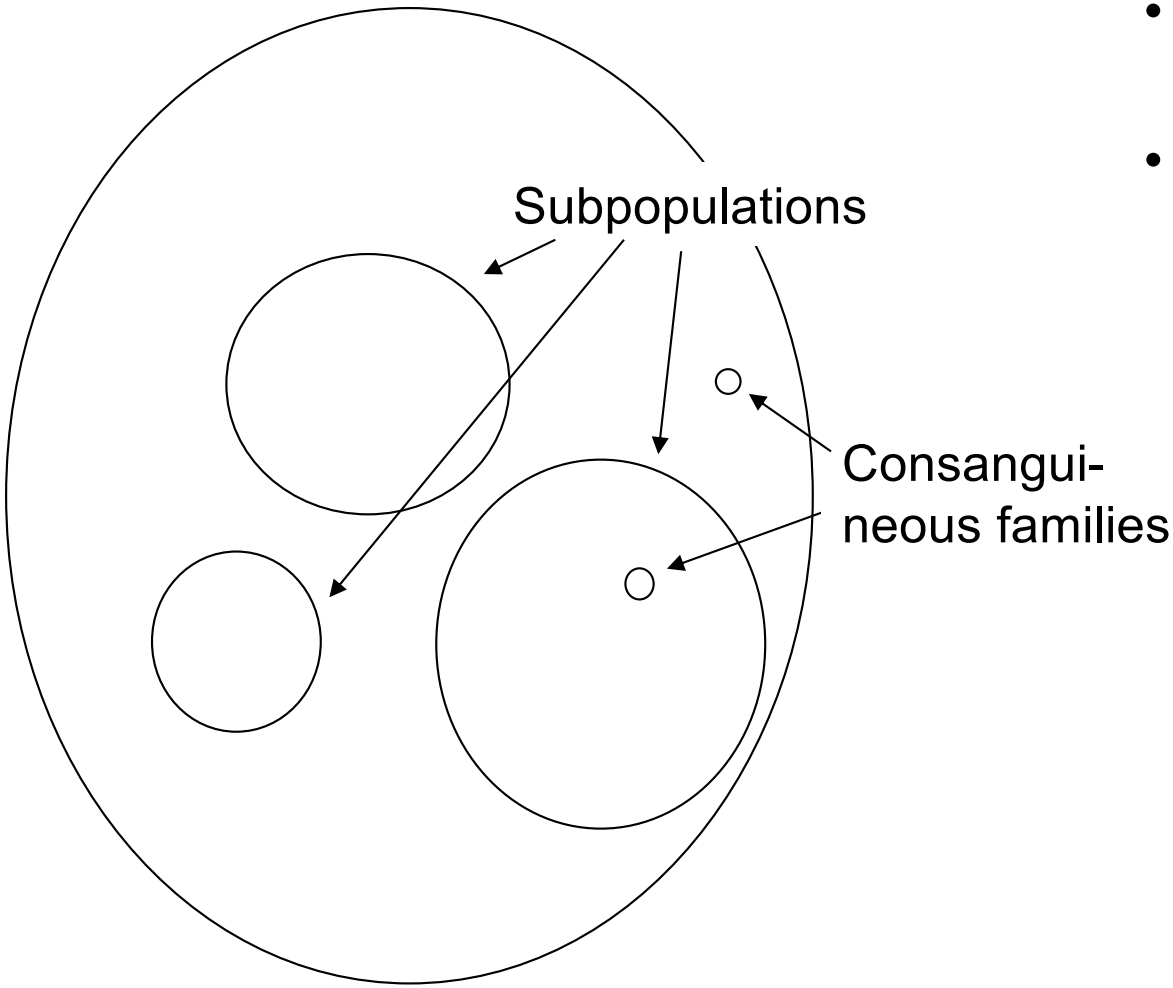
Check: 2% carriers
=> 1/50 x 1/50 x 1/4 affected
= 1/10,000 affected

(Selection acts on very few individuals (1/10,000) => discard)



Offspring risk (a priori) = 1/400

Genetics in families,
Genetics in populations



- Cross-fertile individuals (species)
- Subpopulations isolated by
  - Geography
  - Language
  - Religion
  - …
  - Inbreeding
  - Consanguinity

Allele frequencies vary in different populations

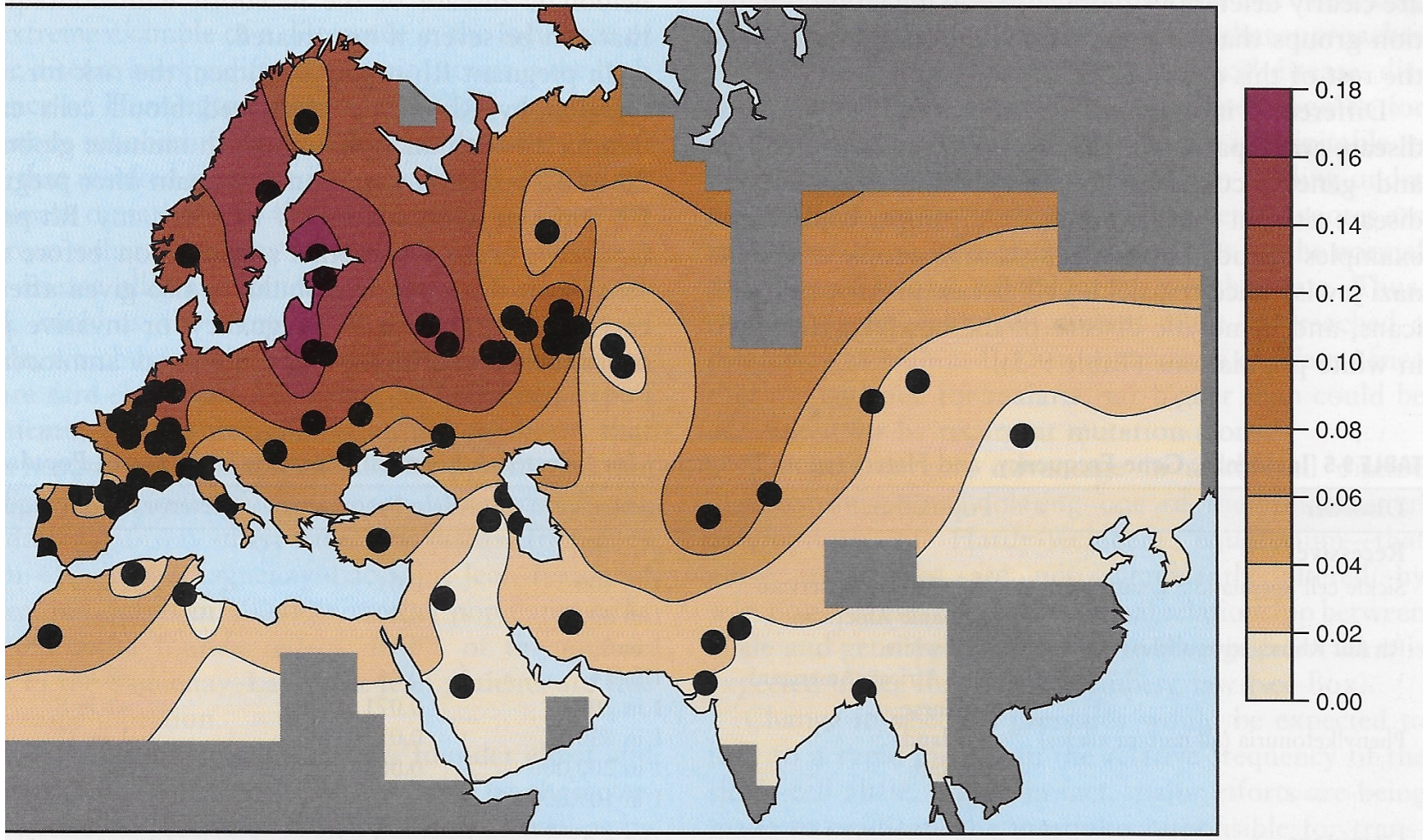


Figure 9-1 **The frequency of ΔCCR5 alleles in various geographical regions of Europe, the Middle East, and the Indian subcontinent.** The various allele frequencies are shown with color coding provided on the right. *Black dots* indicate the locations where allele frequencies were sampled; the rest of the frequencies were then interpolated in the regions between where direct sampling was done. *Gray areas* are regions where there were insufficient data to estimate allele frequencies. *See Sources & Acknowledgments.*

Alleles in stable populations are at H-W equilibrium

Table 26-10 **Comparison between Observed Frequencies of Genotypes for the MN Blood Group Locus and the Frequencies Expected from Random Mating**

	Observed			Expected		
Population	*MM*	*MN*	*NN*	*MM*	*MN*	*NN*
Eskimo	0.835	0.156	0.009	0.834	0.159	0.008
Egyptian	0.278	0.489	0.233	0.274	0.499	0.228
Chinese	0.332	0.486	0.182	0.331	0.488	0.181
Australian aborigine	0.024	0.304	0.672	0.031	0.290	0.679

NOTE: The expected frequencies are computed according to the Hardy-Weinberg equilibrium, using the values of p and q computed from the observed frequencies.

Genes in population

DISTORTIONS TO H-W EQUIL

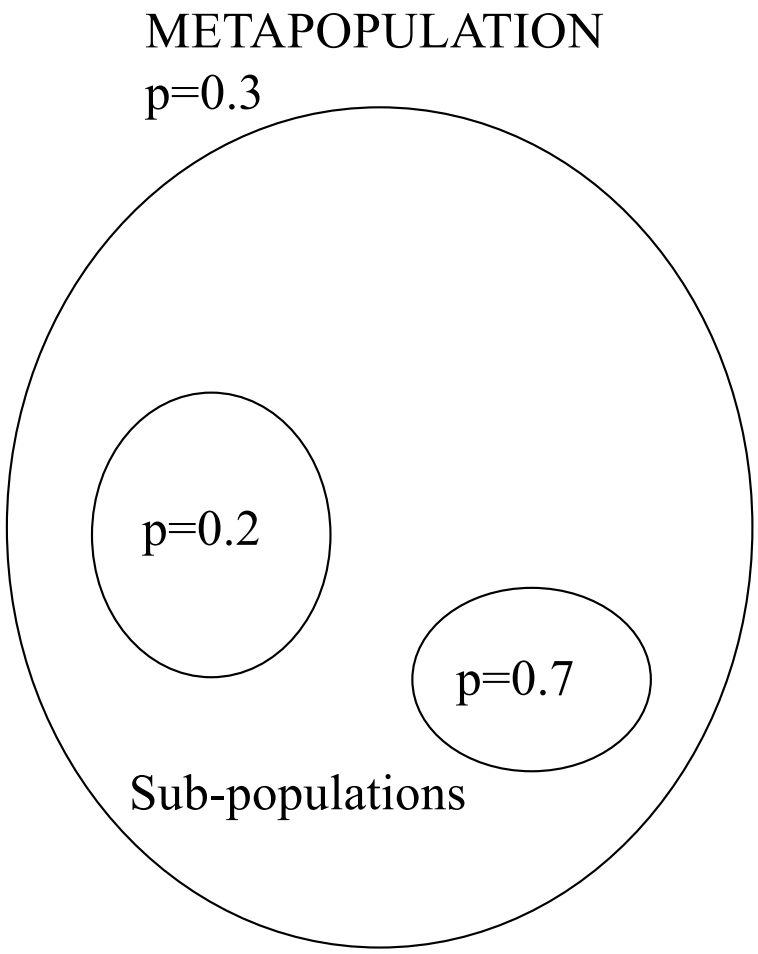
Assortative matings

= if you chose your mate non-randomly

- Height; deafness; …
- Consanguinity

- Geography
- Language
- Religion

STRATIFICATION of population



Sub-populations have their own H-W equilibrium

Table 26-1 Frequencies of Genotypes for Alleles at MN Blood Group Locus in Various Human Populations

Population	Genotype			Allele frequencies	
	MM	*MN*	*NN*	*p*(*M*)	*q*(*N*)
Eskimo	0.835	0.156	0.009	0.913	0.087
Australian aborigine	0.024	0.304	0.672	0.176	0.824
Egyptian	0.278	0.489	0.233	0.523	0.477
German	0.297	0.507	0.196	0.550	0.450
Chinese	0.332	0.486	0.182	0.575	0.425
Nigerian	0.301	0.495	0.204	0.548	0.452

SOURCE: W. C. Boyd, *Genetics and the Races of Man*. D. C. Heath, 1950.

HW equilibria are not additive

- Consider 2 populations in HW equilibrium at one locus
- Sample them and pool the samples
- The resulting pool is NOT at HW equilibrium
  - Stratification of the metapopulation

- If the 2 populations actually mix and mate randomly, equilibrium will be reached, at the next generation

Random genetic drift

- No population is infinitely large
- Each generation is a sample of the previous one
- Stochastic variation in allele frequency between generations

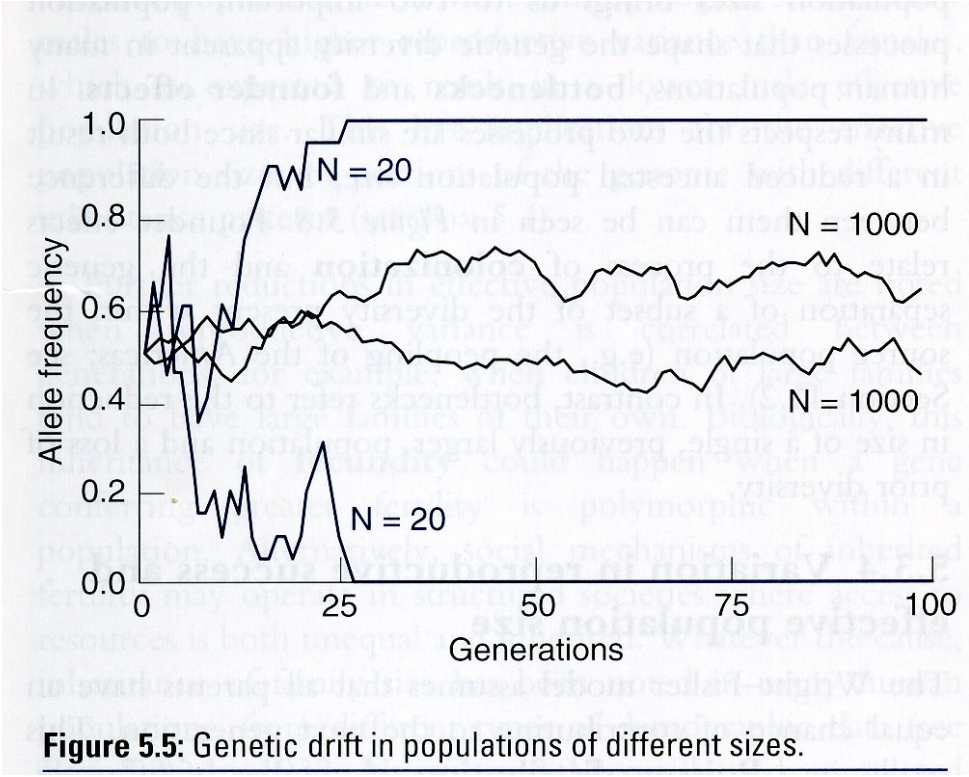
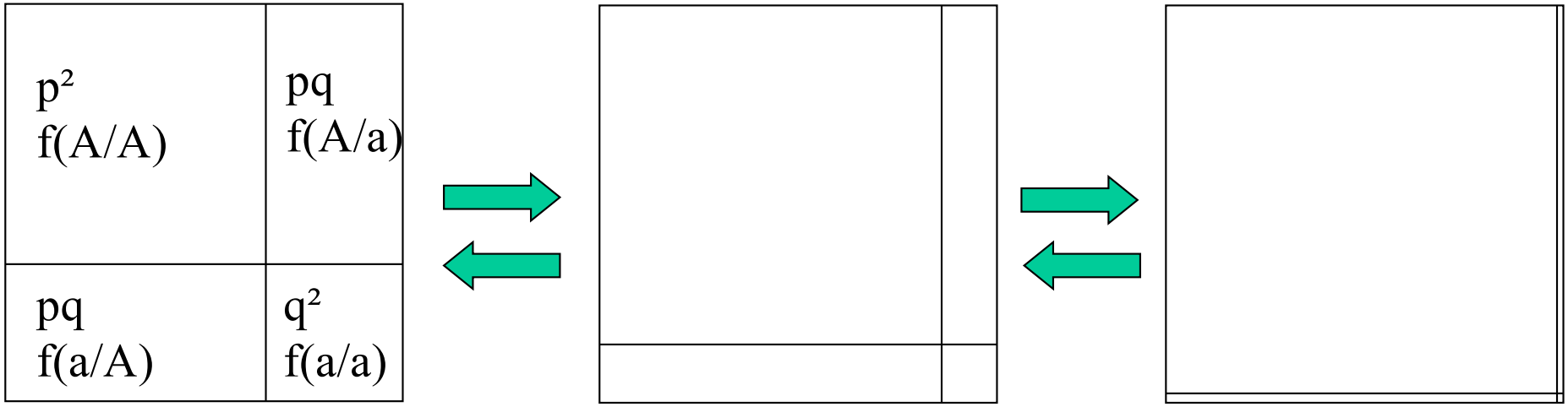


Figure 5.5: Genetic drift in populations of different sizes.

Ex: p=0.5, N=20 (simulation over 100 generations)

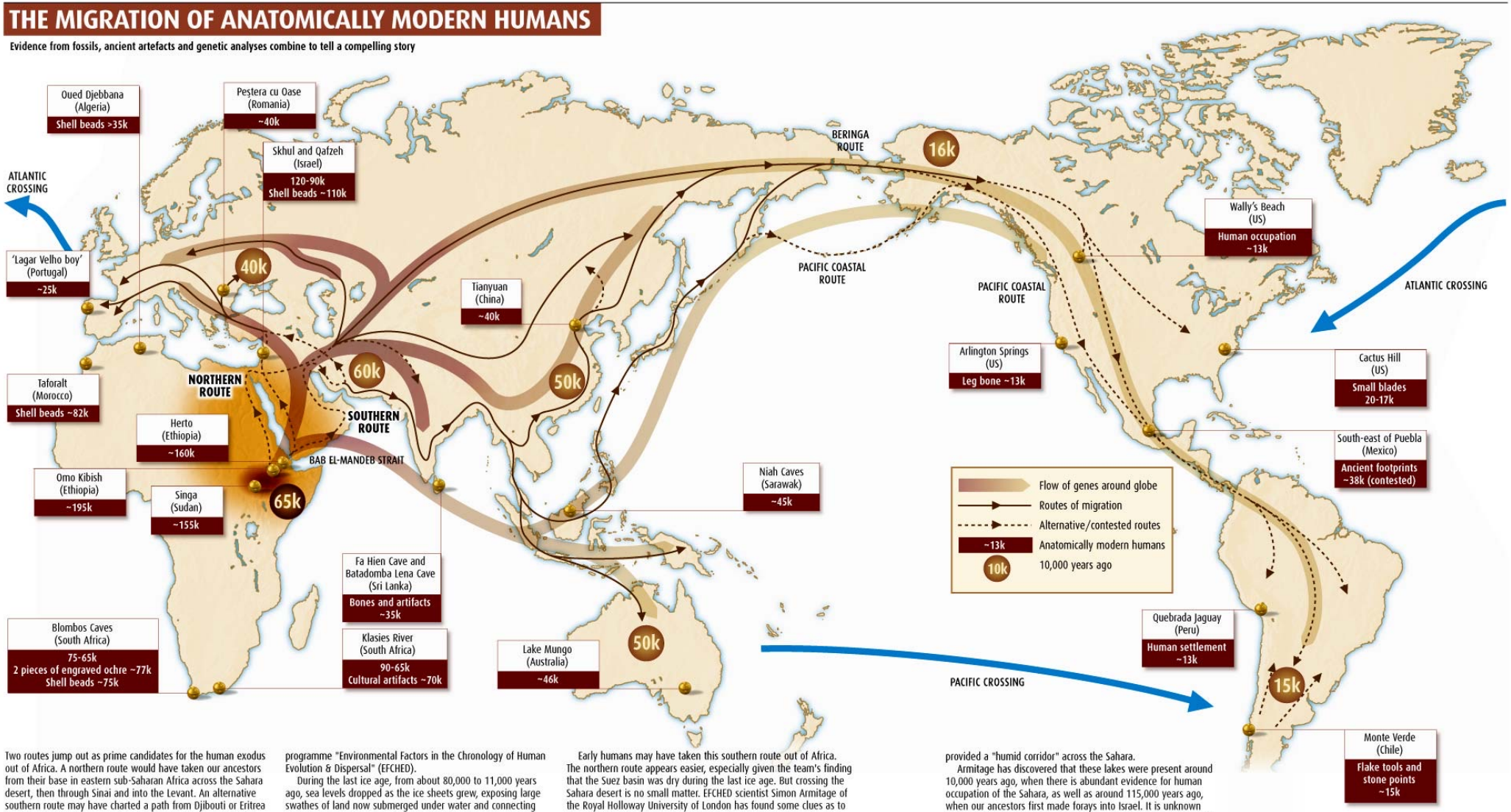
N=20 => one allele gets FIXED

Genetic drift and allele fixation



- Random variation of p and q, over a generation
- In small population
- Once q = 0, q remains 0
- Allele FIXATION: p=1

Out of Africa model:



Two routes jump out as prime candidates for the human exodus out of Africa. A northern route would have taken our ancestors from their base in eastern sub-Saharan Africa across the Sahara desert, then through Sinai and into the Levant. An alternative southern route may have charted a path from Djibouti or Eritrea in the Horn of Africa across the Bab el-Mandeb strait and into Yemen and around the Arabian peninsula. The plausibility of these two routes as gateways out of Africa has been studied as part of the UK's Natural Environment Research Council's programme "Environmental Factors in the Chronology of Human Evolution & Dispersal" (EFCHED).

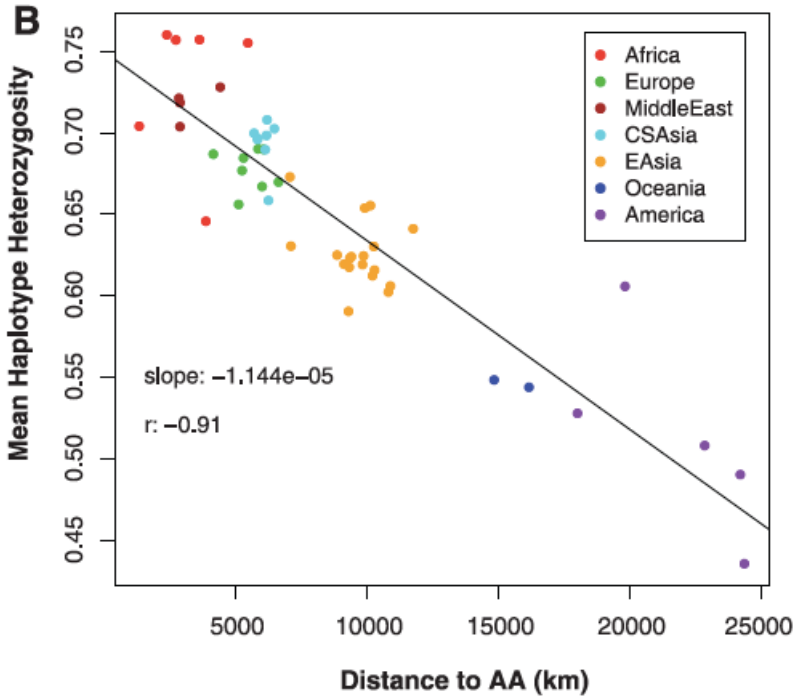
During the last ice age, from about 80,000 to 11,000 years ago, sea levels dropped as the ice sheets grew, exposing large swathes of land now submerged under water and connecting regions now separated by the sea. By reconstructing ancient shorelines, the EFCHED team found that the Bab el-Mandeb strait, now around 30 kilometres wide and one of the world's busiest shipping lanes, was then a narrow, shallow channel.

Early humans may have taken this southern route out of Africa. The northern route appears easier, especially given the team's finding that the Suez basin was dry during the last ice age. But crossing the Sahara desert is no small matter. EFCHED scientist Simon Armitage of the Royal Holloway University of London has found some clues as to how this might have been possible. During the past 150,000 years, North Africa has experienced abrupt switches between dry, arid conditions and a humid climate. During the longer wetter periods huge lakes existed in both Chad and Libya, which would have provided a "humid corridor" across the Sahara.

Armitage has discovered that these lakes were present around 10,000 years ago, when there is abundant evidence for human occupation of the Sahara, as well as around 115,000 years ago, when our ancestors first made forays into Israel. It is unknown whether another humid corridor appeared between about 65,000 and 50,000 years ago, the most likely time frame for the human exodus. Moreover, accumulating evidence is pointing to the southern route as the most likely jumping-off point.

Out of Africa, progressive drift

- Observe >100k SNP polymorphisms
- Measure variability (= measure heterozygosity) in various populations
- Plot variability as a fn of distance from Ethiopia capital, Addis Ababa (AA)



Li et al. Science 2008

Mutations and drift, Bottlenecks and founder effects

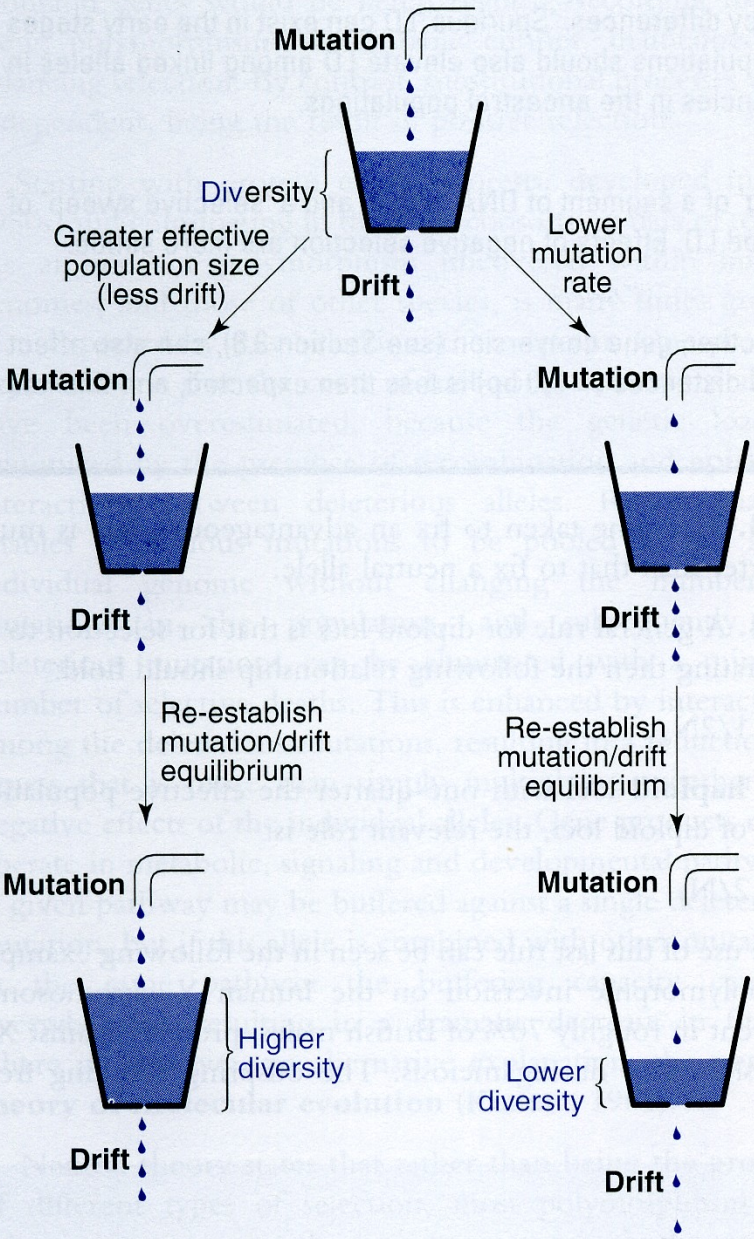


Figure 5.13: A metaphorical depiction of the relationship between mutation rate, drift and diversity.

A change in either the mutation rate or effective population size changes the diversity at mutation-drift equilibrium – see text.

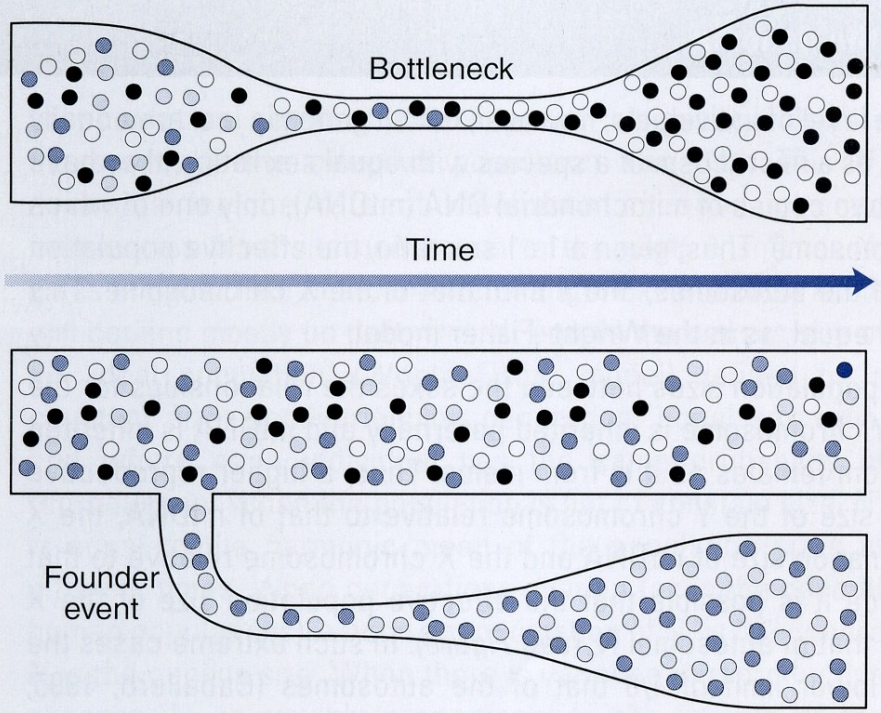


Figure 5.8: Bottlenecks and founder events.

Circles of different colors represent different alleles. Both bottlenecks and founder events result in a loss of allelic diversity.

Hardy – Weinberg law

- The frequency of the three genotypes AA, Aa and aa is given by the tems of the binomial expansion of $(p+q)^2$
  $= p^2 + 2pq + q^2$
- And does not change over generations
- Under certain conditions :
  - Random matings
  - No mutation
  - No selection 
  - No drift
  - No migration in or out
  - Equal generations
  - Stable population

Selection

- All individuals in one generation differ qualitatively from one another
- Differential rates of survival and reproduction (fitness)
  - Natural selection (environment)
  - Artifical selection (plant or animal breeders)
- **If variability is (partly) inherited**, this results in the evolution of the population (microevolution)
- Via change in allele frequencies

Natural selection

Ex: Cystic Fibrosis (CF) affects 1/2500 individual at birth (incidence measured at birth)
Patients are normal at birth
progressive disease in children and young adults
life expectancy = 39 yrs)

- A birth, H-W equilibrium :
  1/25 heteroz <=>1/2500 affected

AA≃.96 ; Aa = .04 ; aa=1/2500

- À 75 yrs, H-W equilibrium not observed :
  1/25 heteroz <=> 0 affected (all are dead)

AA≃.96 ; Aa = .04 ; aa=0

Selection: + or -

- **NEGATIVE SELECTION** : reduced fitness
  = purifying selection

- **POSITIVE SELECTION** : increased fitness
  = adaptive selection

- **BALANCED SELECTION** : htz performs best

- **NO SELECTION** : for most mutations
  neutral evolution

Fitness

- Survival into reproductive age
- Success in mating: *sexual selection*
- Ability to fertilize: *gamete selection*

  fertility, meiotic drive
- Number of progeny: *fecundity*

Selection

- All individuals in one generation qualitatively different from one another
- Differential rates of survival and reproduction (fitness)
  - Natural selection (environment)
  - **Artifical selection (plant or animal breeders)**
- If variability is (partly) inherited, this results in the evolution of the population (microevolution)
- Via change in allele frequencies

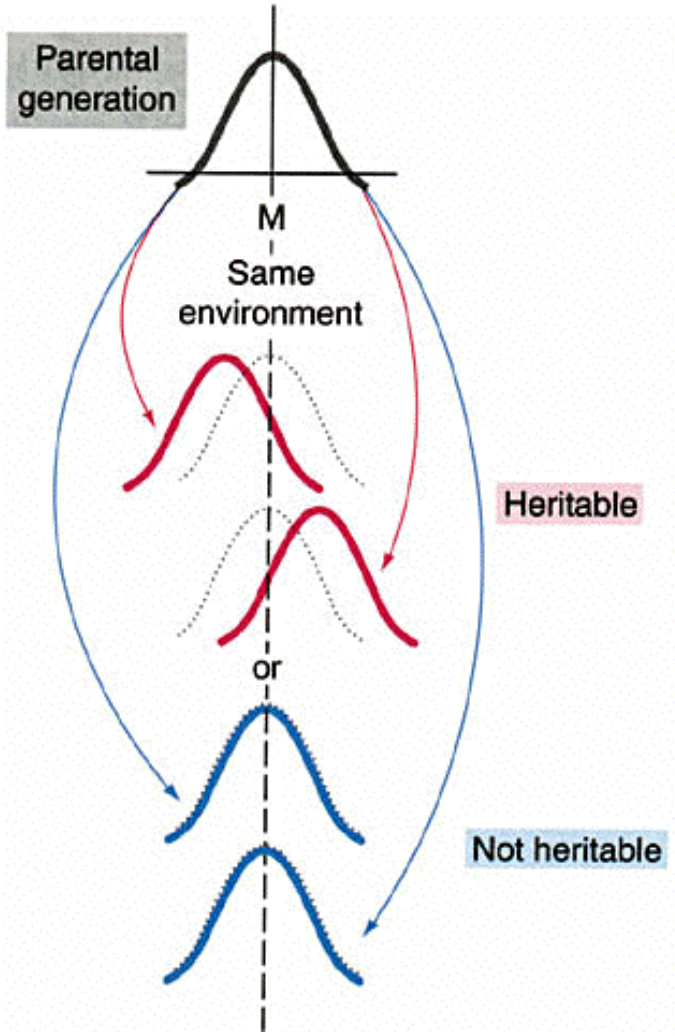
Artificial selection (empirical)

- Since 10,000 yrs in agriculture
- Since 10,000 yrs in farming

Works only on (partially) inherited characters

Regression to the mean indicates non-heritability of variation

- Cross individuals from the extremes of the distribution
- If variation not inherited (= environment effect only): => crosses from both extremes will produce same distribution = regression to the mean
- If variation (partly) inherited (= genetic effect present) => distribution different in two groups



SELECTION

- Natural or artificial
- Differential survival and reproduction of individuals

➢ **NEGATIVE**
purifying selection

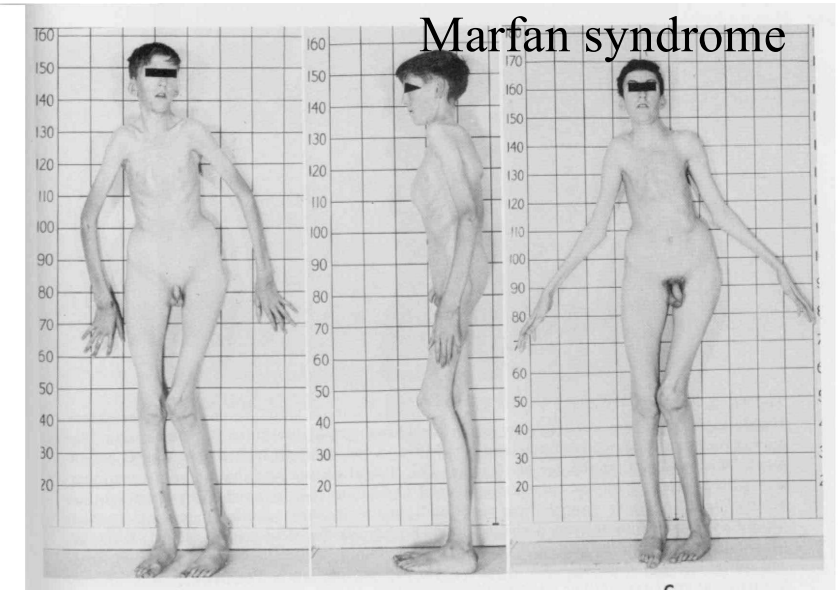
➢ **POSITIVE**
adaptive selection

➢ **BALANCING**

- Most changes are not selected for or against
NEUTRAL evolution

Negative selection

- Most mutations that cause dominant disease
- Because these patients have fewer children (fitness <1)

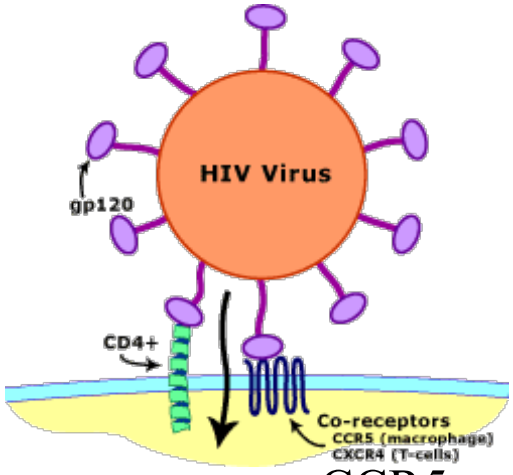


Positive selection of mutation CCR5 * delta32

Mutation delta32 inactivates CCR5 , a co-recepteur for HIV virus

Mutation does not seem to cause any problem per se

SELECTION of this mutation by
- Plague (14th century)
- Smallpox (Variola)
- AIDS



CCR5

=> This mutation will get fixed (settle) in population if selective pressure maintained (?)

Selection: + or -

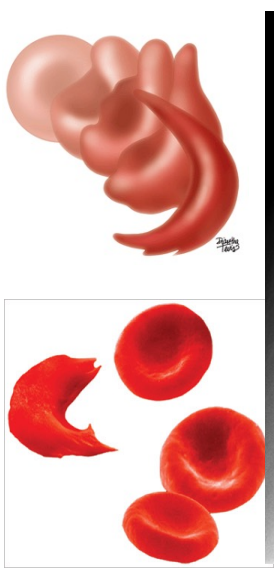
- **NEGATIVE SELECTION** : reduced fitness
  = purifying selection

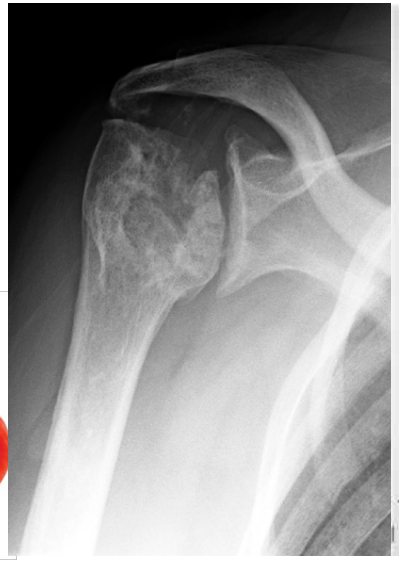
- **POSITIVE SELECTION** : increased fitness
  = adaptive selection

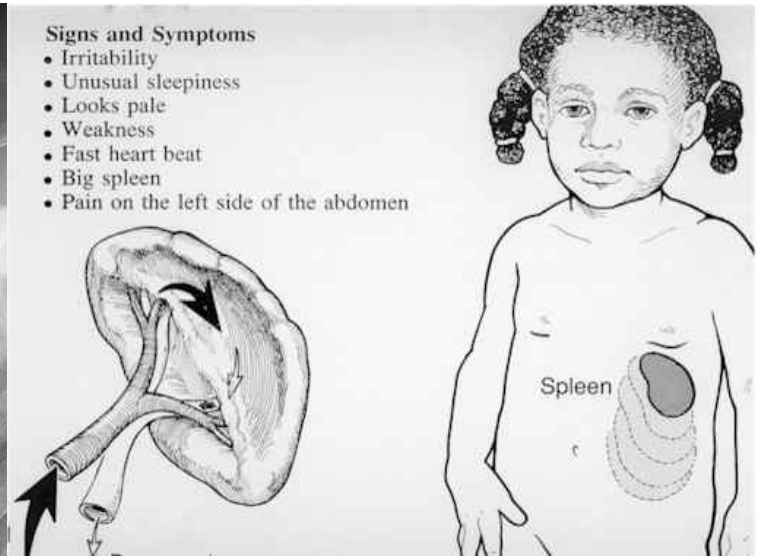
- **BALANCED SELECTION** : htz performs best 

- **NO SELECTION** : for most mutations
  neutral evolution

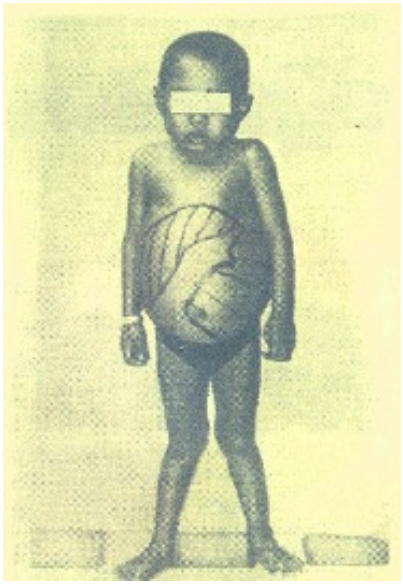
Balanced Selection : ex: hereditary anemia





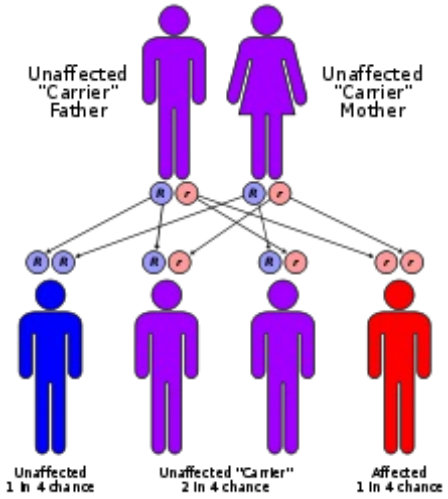


Drépanocytose (Sickle cell disease)



Thalassemia

Autosomal recessive, monogenic

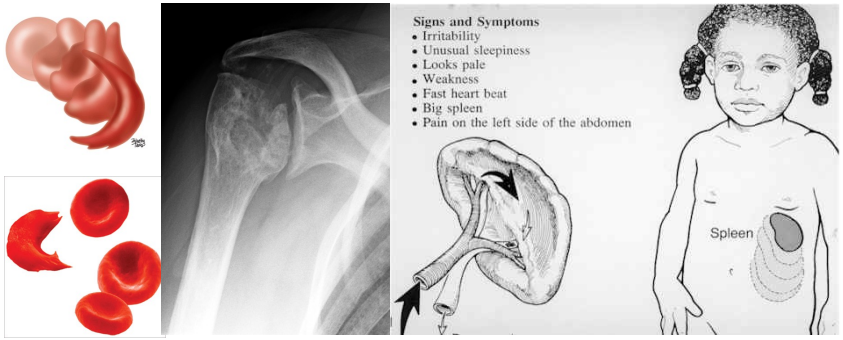


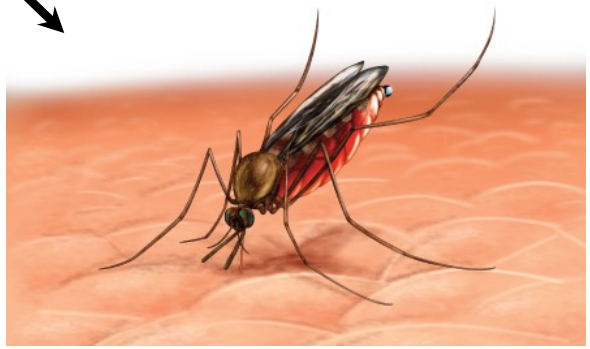
=> 10-20% carriers in some populations

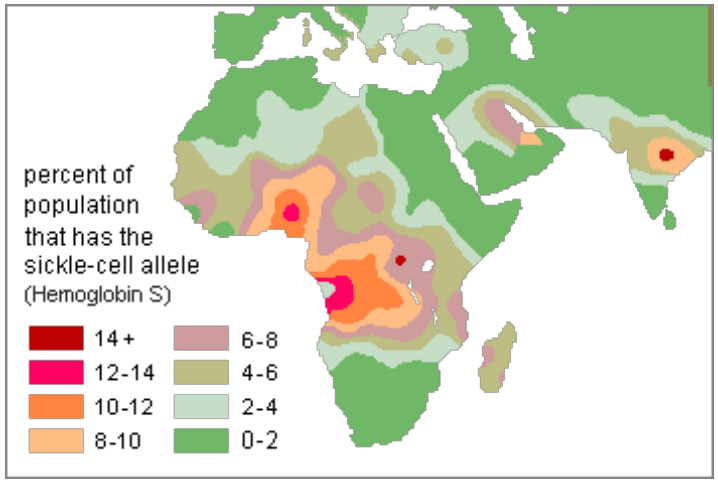
although mutated alleles disappear as patients die !

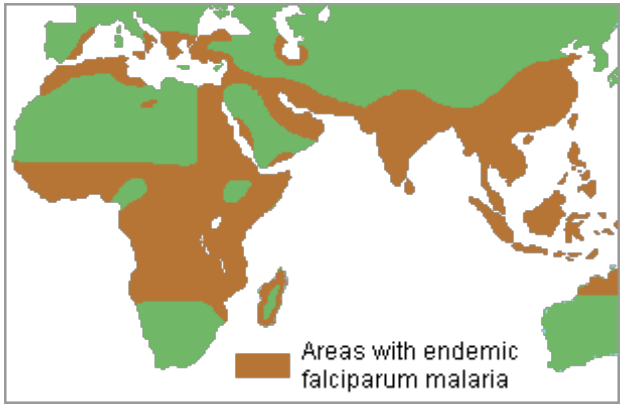
Sickle cell (and thalassemia)

follow malaria

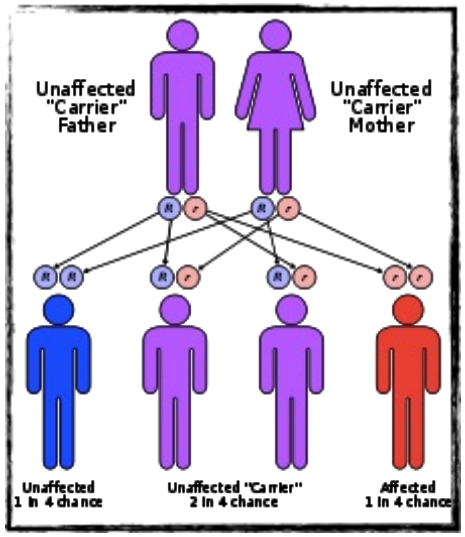






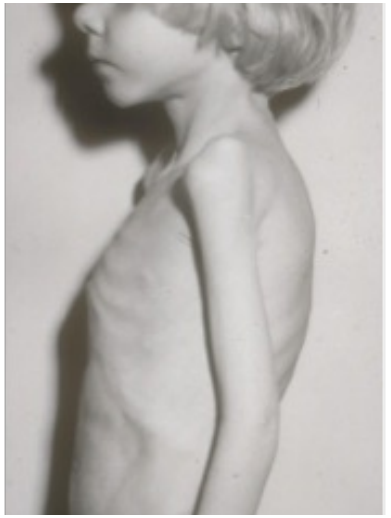


balanced polymorphisms



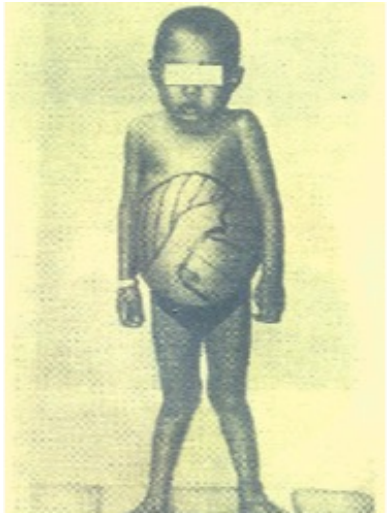
- CFTR mutations
  - 1/25 (4%) carriers
    1/2500 affected

Cystic Fibrosis
(mucoviscidose)



- HbBeta * null
  - 1/10 (10%) carriers
    1/400 affected

Thalassemia
(beta 0)



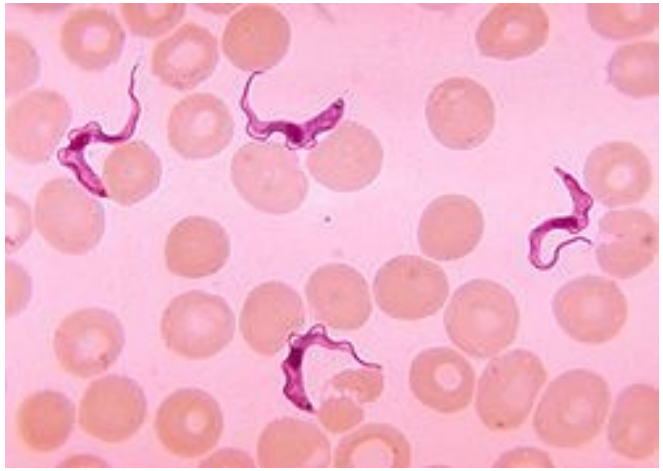
Severe disease in hmz => These genetic changes can not get fixed in population

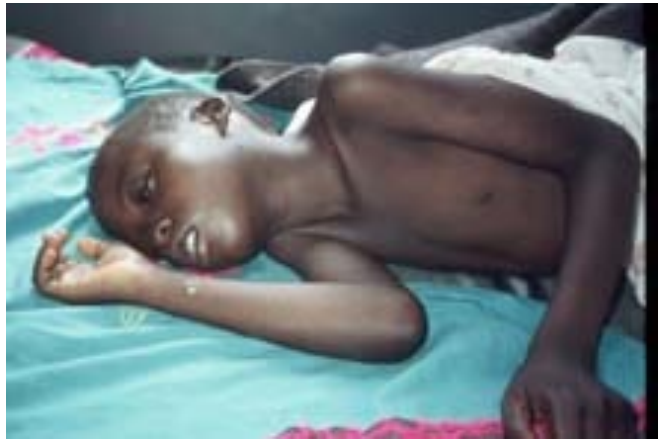
CFTR*DF508

- Is a mutation, causing disease with 100% penetrance (if biallelic)
- Is a polymorphism as q = 1.5%
- Balanced selection (overdominance)

Balanced selection of APOL1 mutation in Africa

- resistance to trypanosoma infection (sleeping disease) in heterozygotes
- Nephrosis in homozygotes (focal segmental glomerulosclerosis)
    - African American have higher rates of renal disease than European Americans





Balanced selection

- HbS mutation
  - Malaria < > Sickle-cell anemia
- CFTR mutations
  - Infant diarrhea (?) < > CF

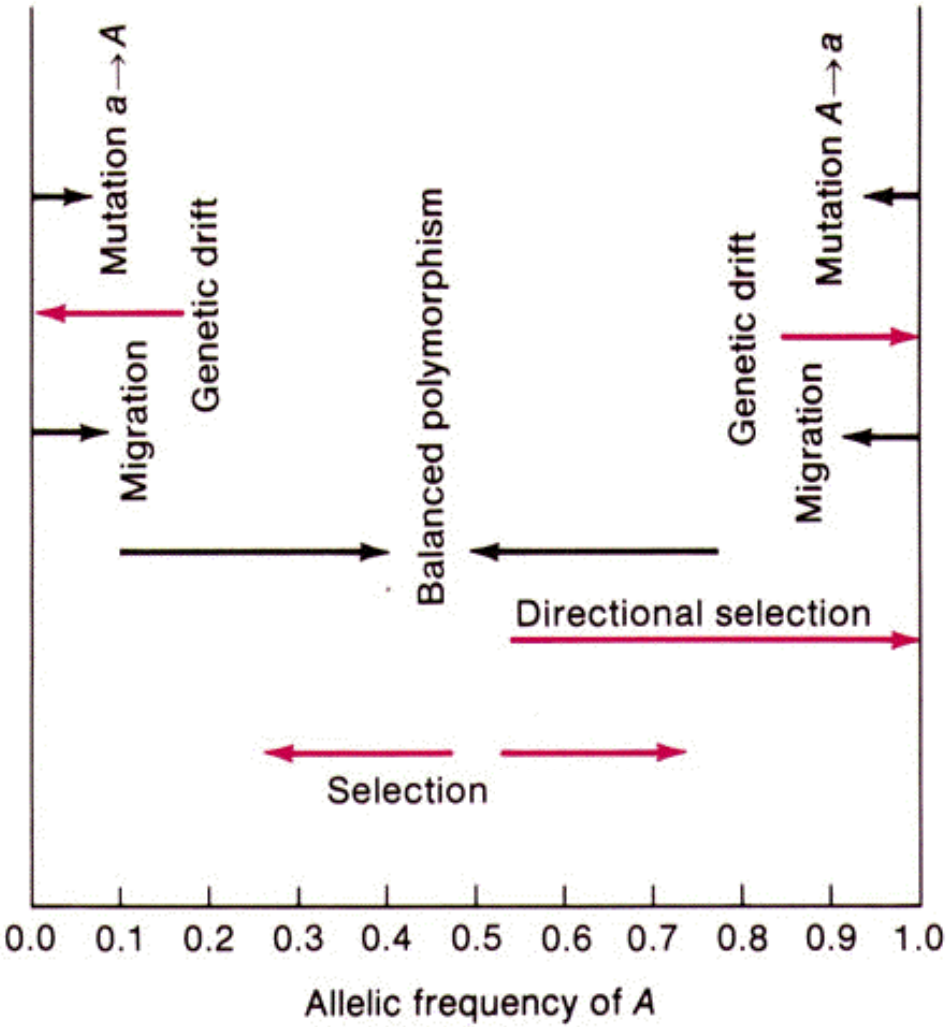
=> Such genetic changes can not get fixed in population.. or everyone would be affected with CF, Sickle Cell, …

Balanced selection in oligogenic / multigenic/ complex disorders

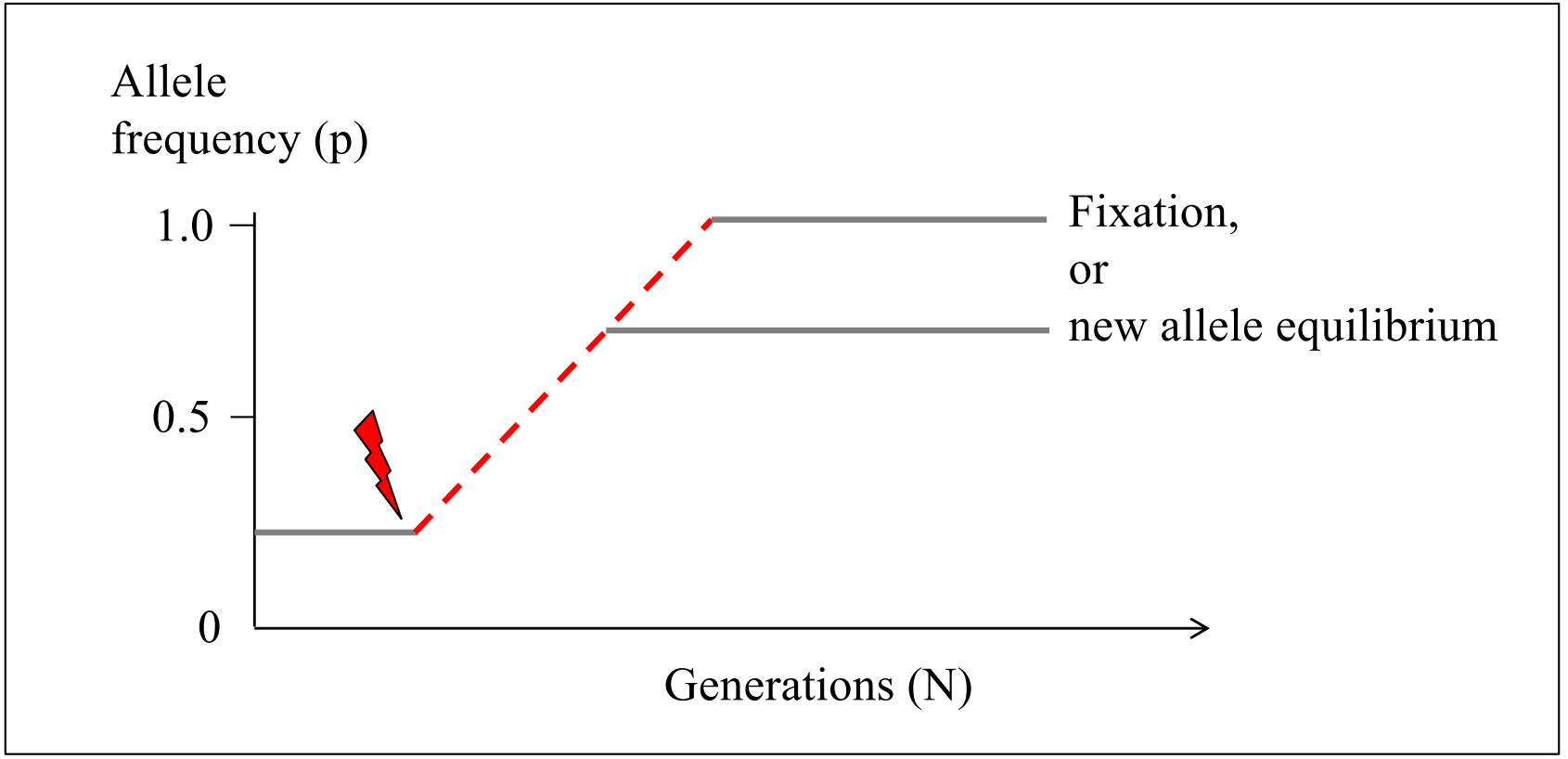
- TLR4
  - Septic shock < > ischemic cardiopathy
- HFE
  - Iron deficiency < > iron overload
- FV Leiden
  - Fewer hemorrhages < > thrombophilia

Microevolution

- Changes in allele frequency in population
- Cross-fertile individuals



Positive selection (adaptive mutation)

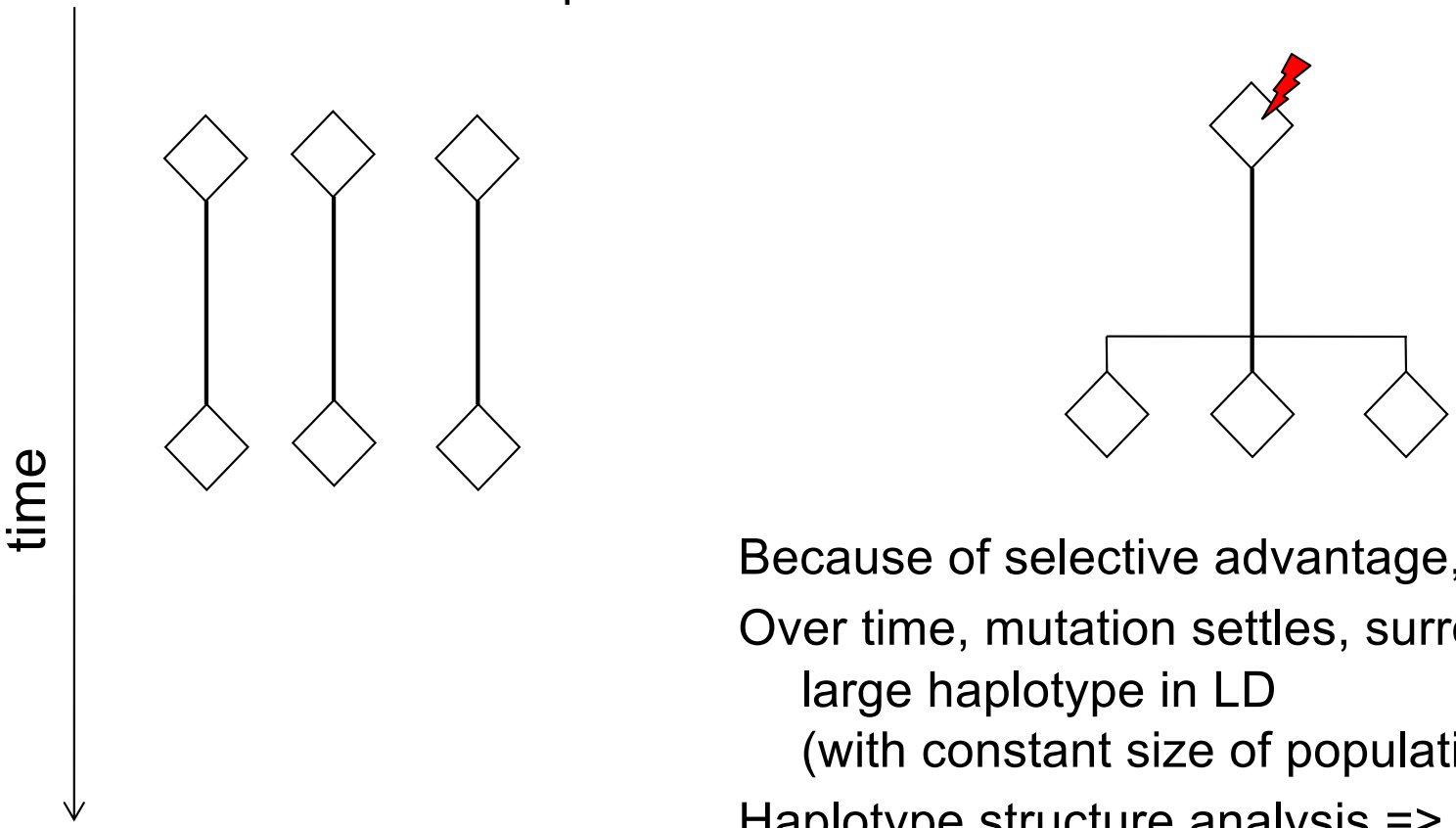


Δp/ΔN ≡ Selection = differential number of offspring, untill fixation, or untill new equilibrium reached

=> What happens to haplotype around adaptive mutation during selection ?

Mutation with selective advantage

Population with constant size

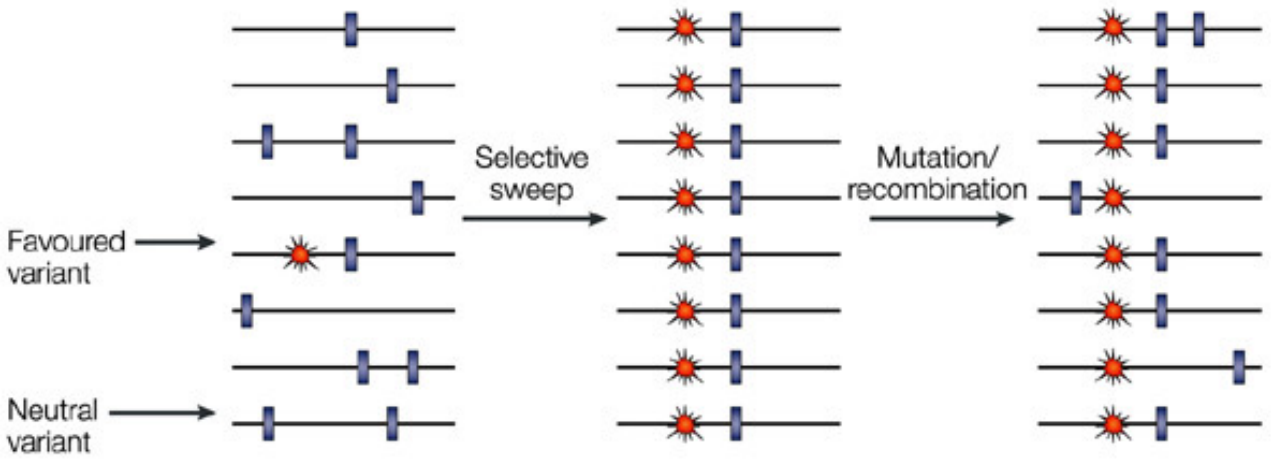


Because of selective advantage, fitness > 1

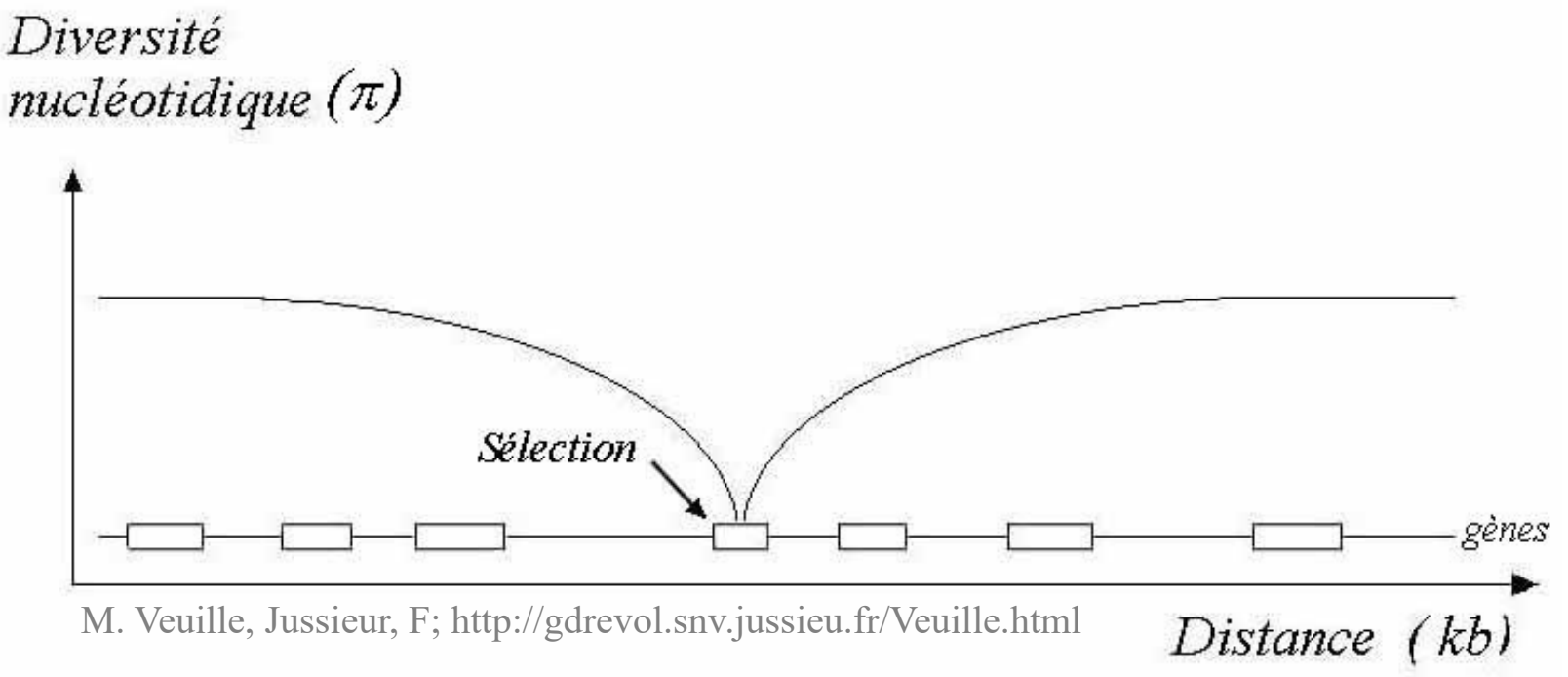
Over time, mutation settles, surrounded by large haplotype in LD (with constant size of population)

Haplotype structure analysis => detects recent positive selection events

Selective sweep

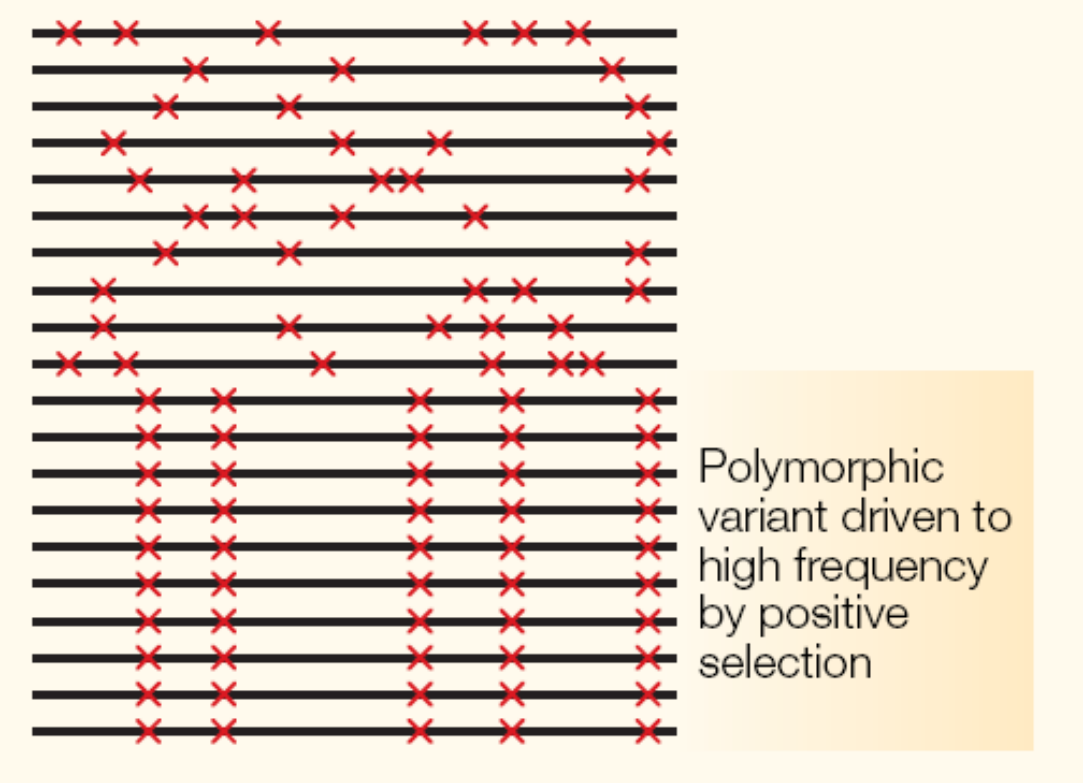


Boffelli D et al. 2004



M. Veuille, Jussieur, F; http://gdrevol.snv.jussieu.fr/Veuille.html

Selective sweep



Gilbert et al. 2004

Selective sweep in ongoing evolution

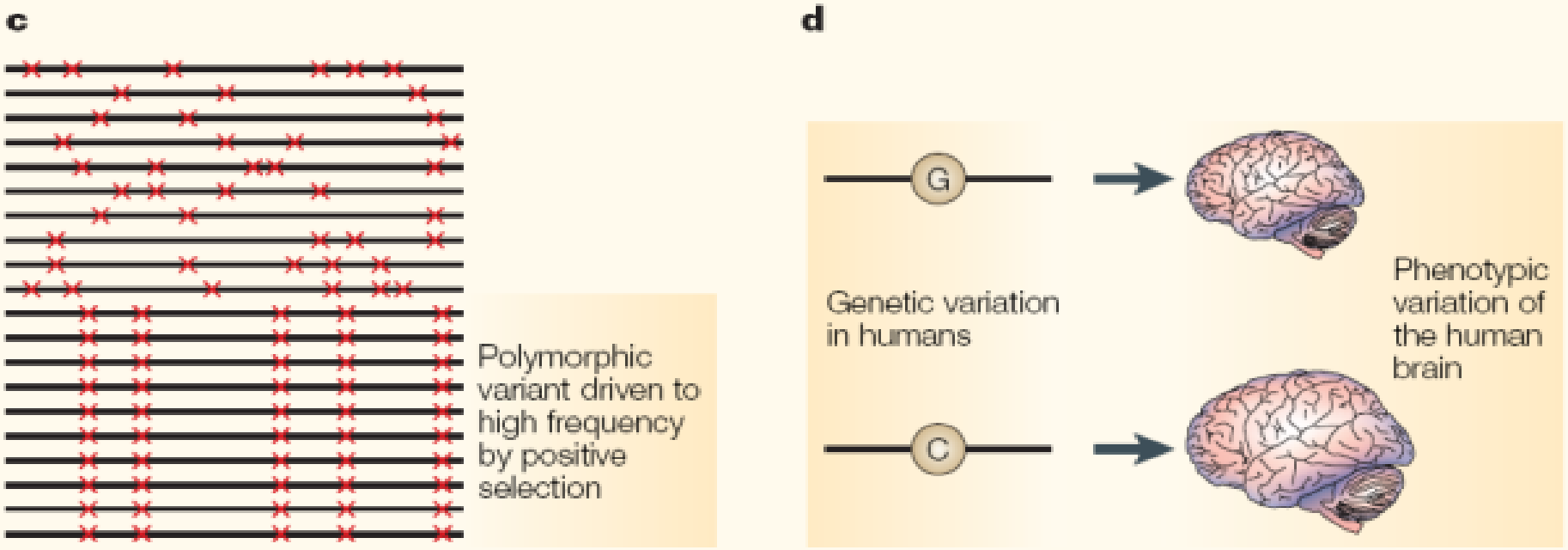


Figure 4 | **A methodological template for investigating the genetic basis of human brain evolution. a** | Large-scale comparisons of brain-related genes across four strategically selected species that include the human, Old World monkey, rat and mouse. These comparisons can reveal broad genome-wide trends and uncover specific genes of interest (for example, genes with significantly higher rates of evolution in primates than rodents). **b** | Analysis of interesting genes identified through (**a**) in a wider range of species. This analysis allows a more detailed evolutionary investigation of individual genes to address questions such as whether the evolution of these genes is specifically accelerated in the lineage leading to humans compared with that in other primate and non-primate taxa. **c** | Polymorphism studies of interesting genes in humans. Each line represents a copy of a locus under investigation and each cross represents a mutational polymorphism. **d** | Correlating polymorphisms in humans with variations in brain phenotype (such as brain size). The phylogenetic relationships and evolutionary timescales depicted in (**a**) and (**b**) are based on data from REFS 114–118.

How recognize human populations ?

Ancestry Informative Markers

= alleles with widely different frequencies among populations originating in different parts of the world

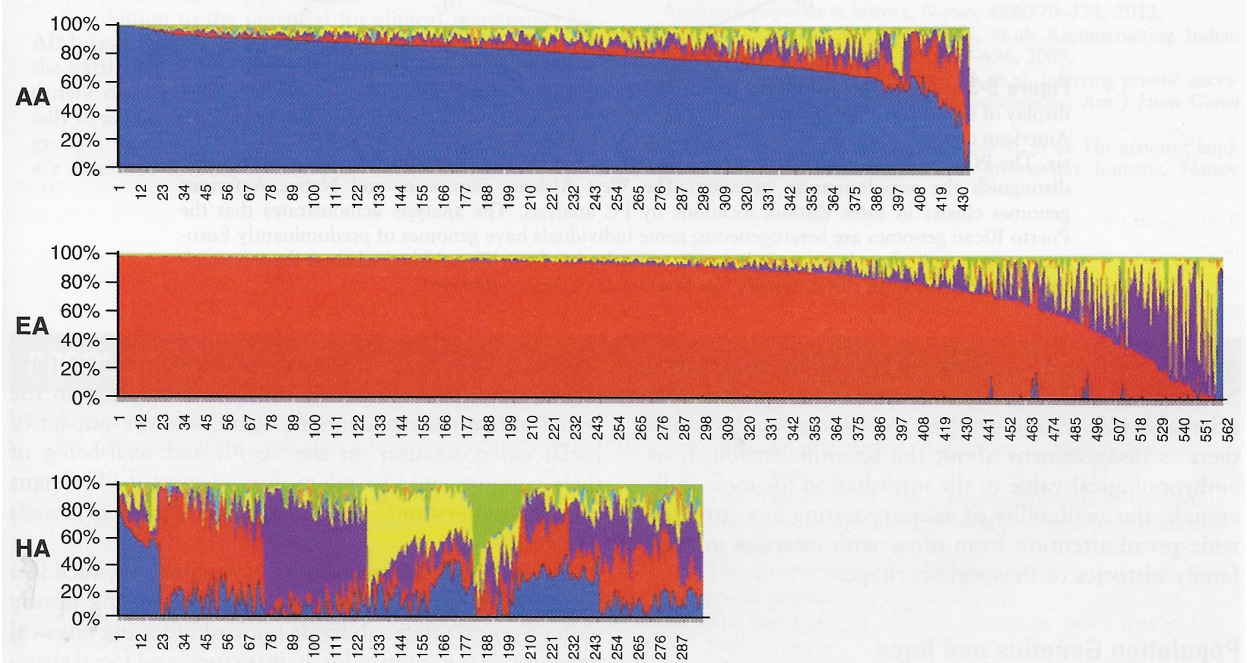


Figure 9-2 **Mixed ancestry of a group of Americans who self-identify as African American (AA), European American (EA), and Hispanic American (HA) using ancestry informative markers.** Each vertical line represents one individual ((totaling hundreds, as shown by the numbers), and subjects are displayed according to the predominant ancestry contribution to their genomes. Different colors indicate origin from a different geographical origin, as inferred from AIMs, as follows:

Actually markers of geographical origin. Snapshot.

Geographical distance is tighltly associated with ancestry

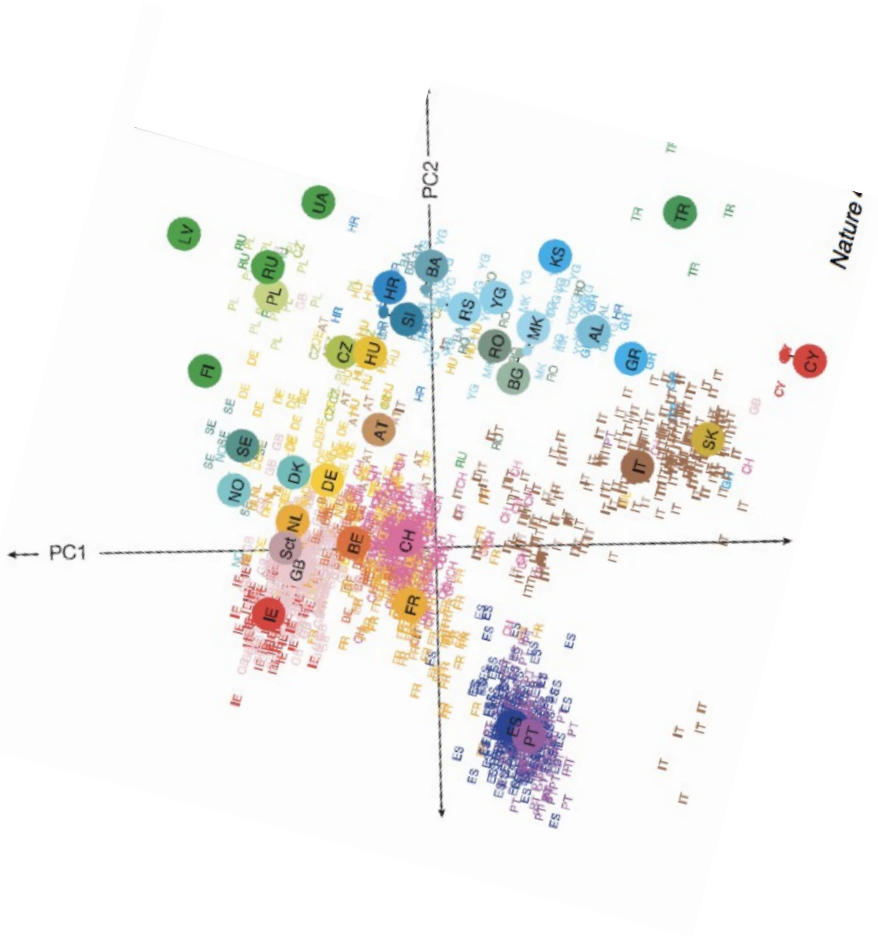
Allows for probabilistic interpretation of ancestry in an individual

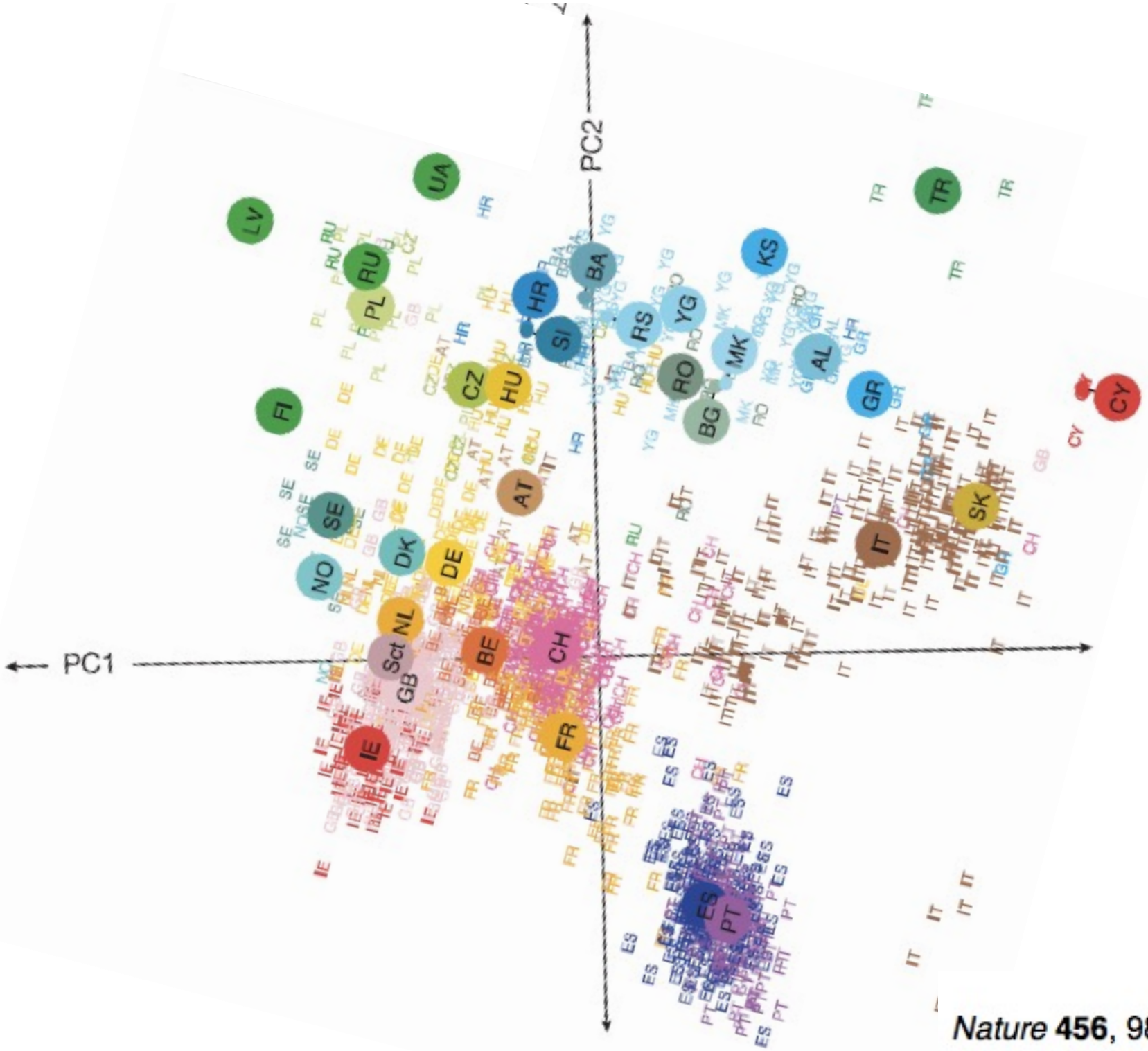
Indicates need for sub-population-specific Gwas

Tell me where I come from: genetic polymorphism reflect geographical origin

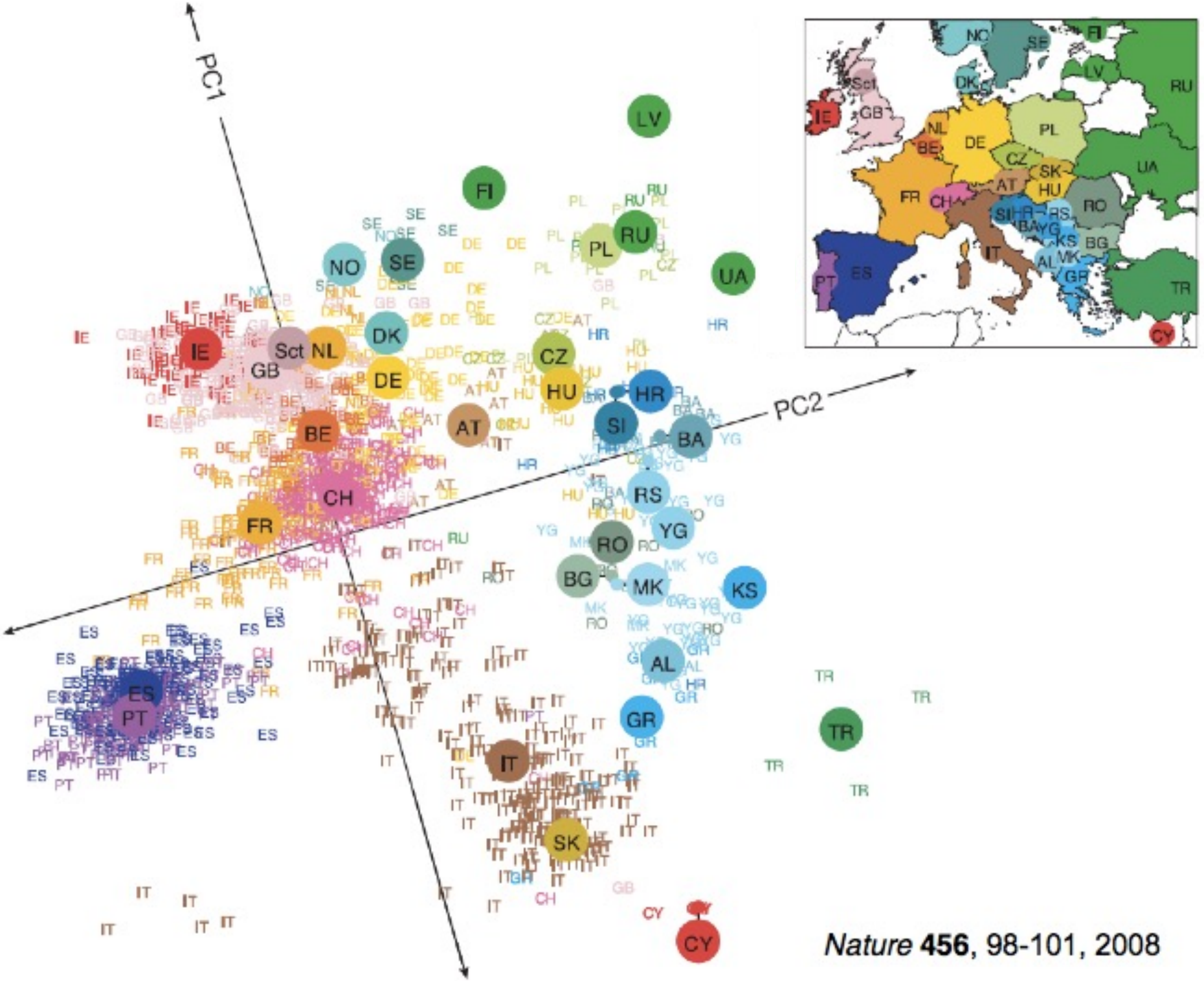
Principal Component Analysis

- Start from 250k SNP polymorphsims
- Multivariate analysis
- Generate 2 (or 3) graphical representations of distances between populations
  = 2 (or 3) eigenvectors





Nature **456**, 98-101, 2008



Nature **456**, 98-101, 2008

The private sector owns a lot of personal genomic and phenotypic data

- ✓ Please fill in the evaluation form ;-)

  and hand it to the teacher of the last session

Thank you