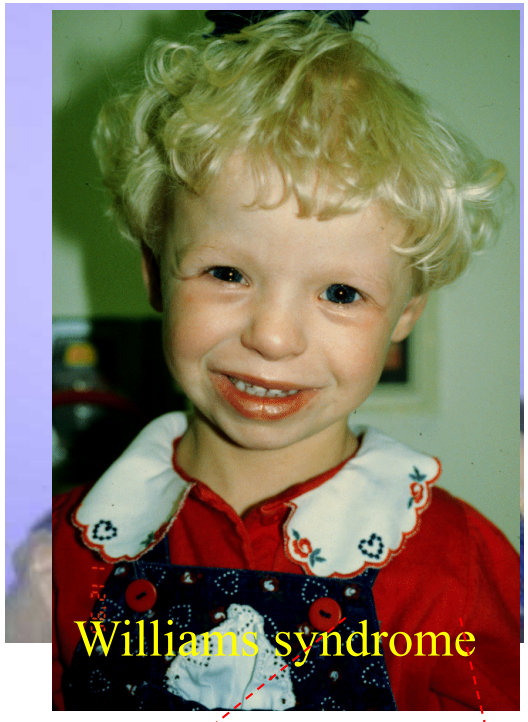
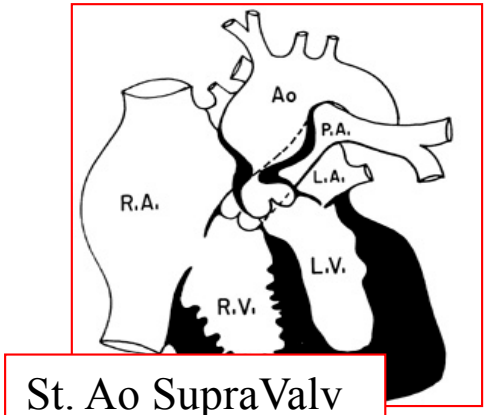
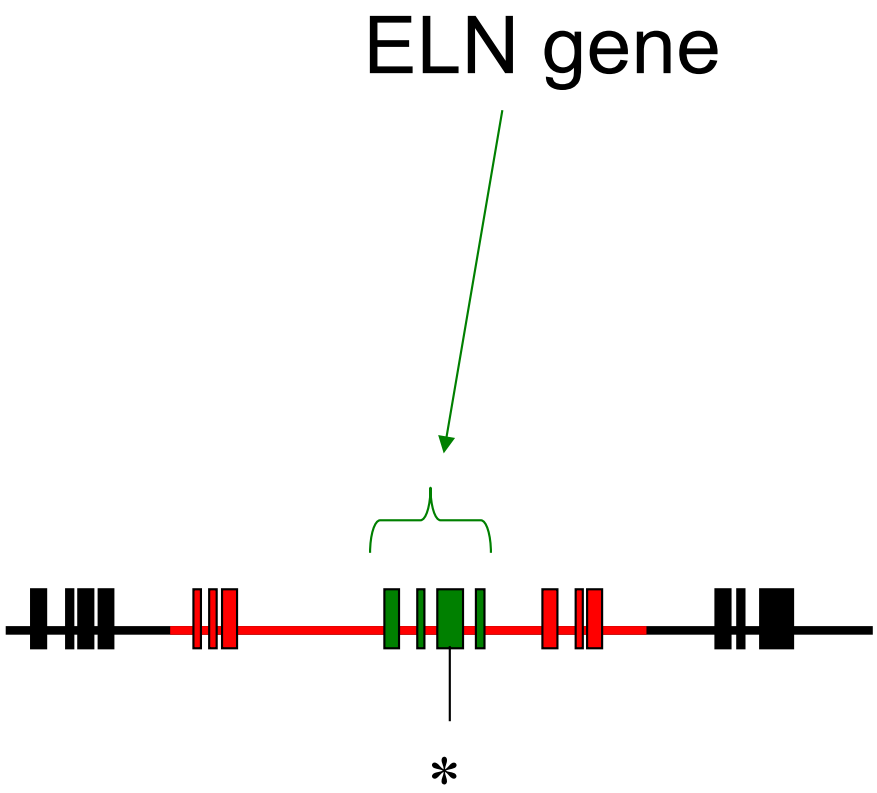
Williams syndrome

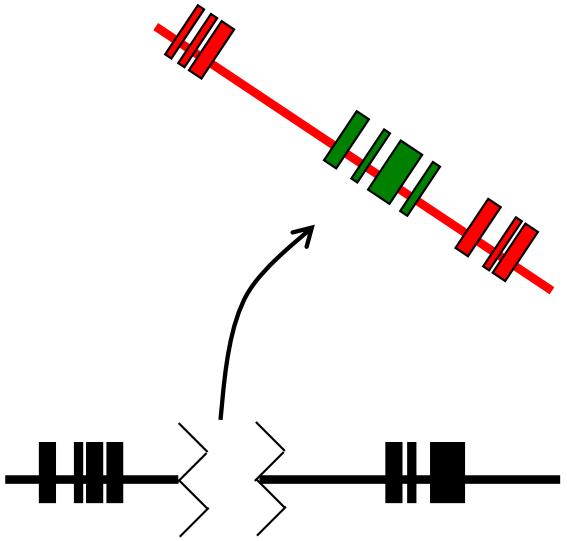




St. Ao SupraValv

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





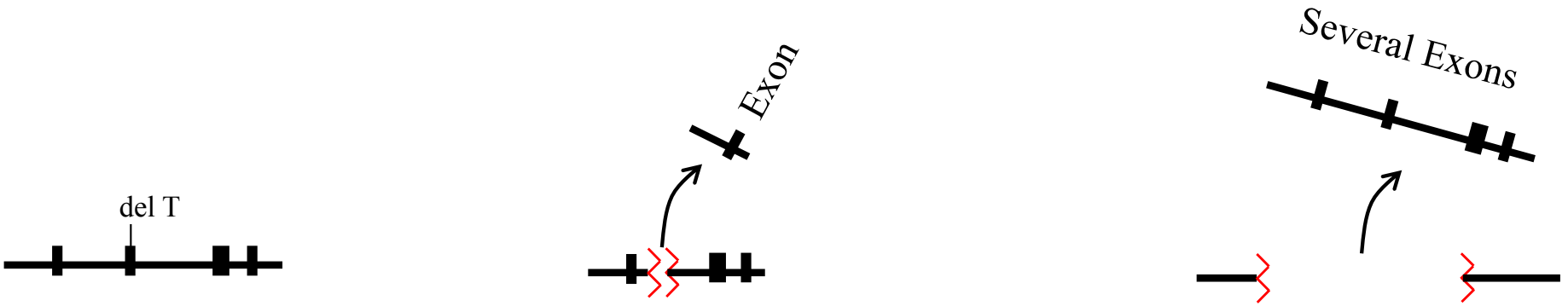
Point mutation
=> St Ao , isolated
AD transmission

Null mutation

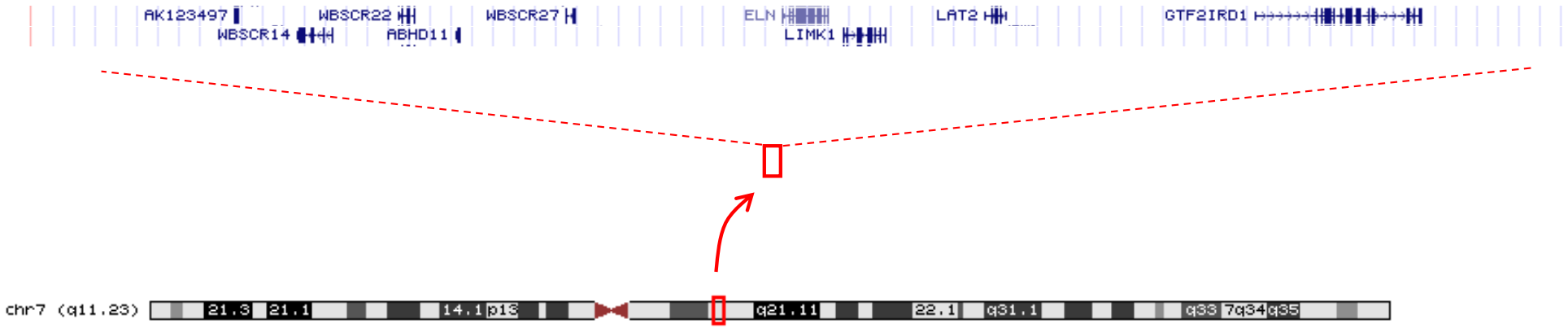
Deletion ELN + **contiguous genes**
=> St Ao + malfo + MR
+ dysmorphism

confirms haploinsufficiency

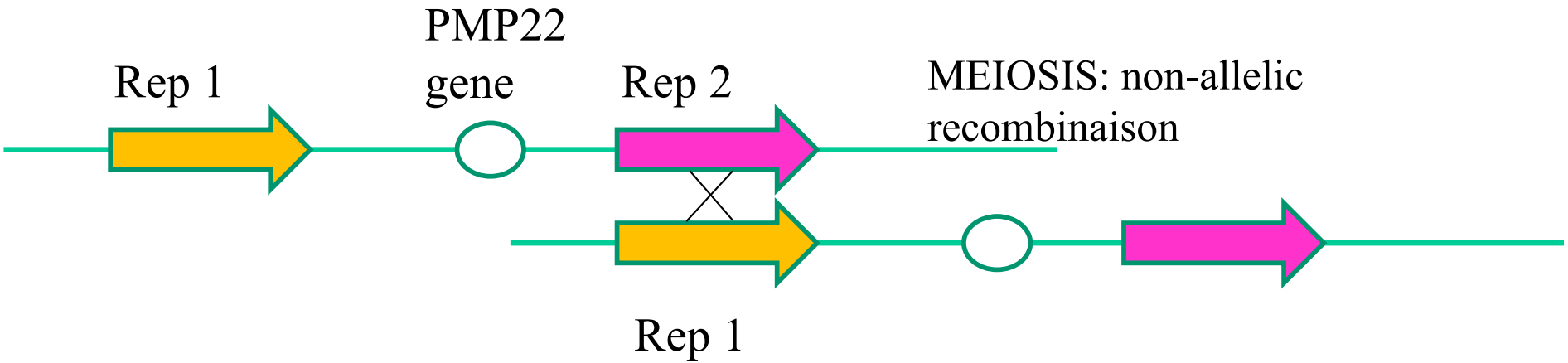
Variable size of mutations

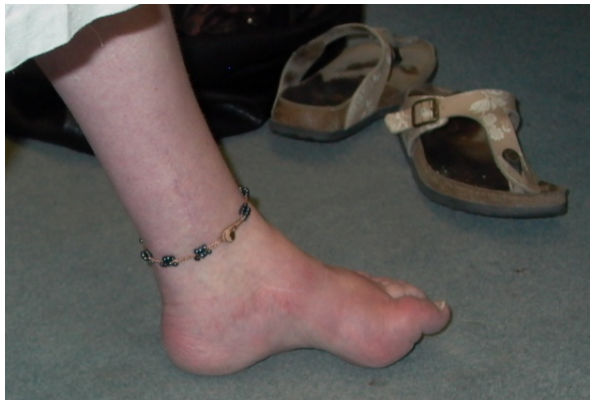


Several genes



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





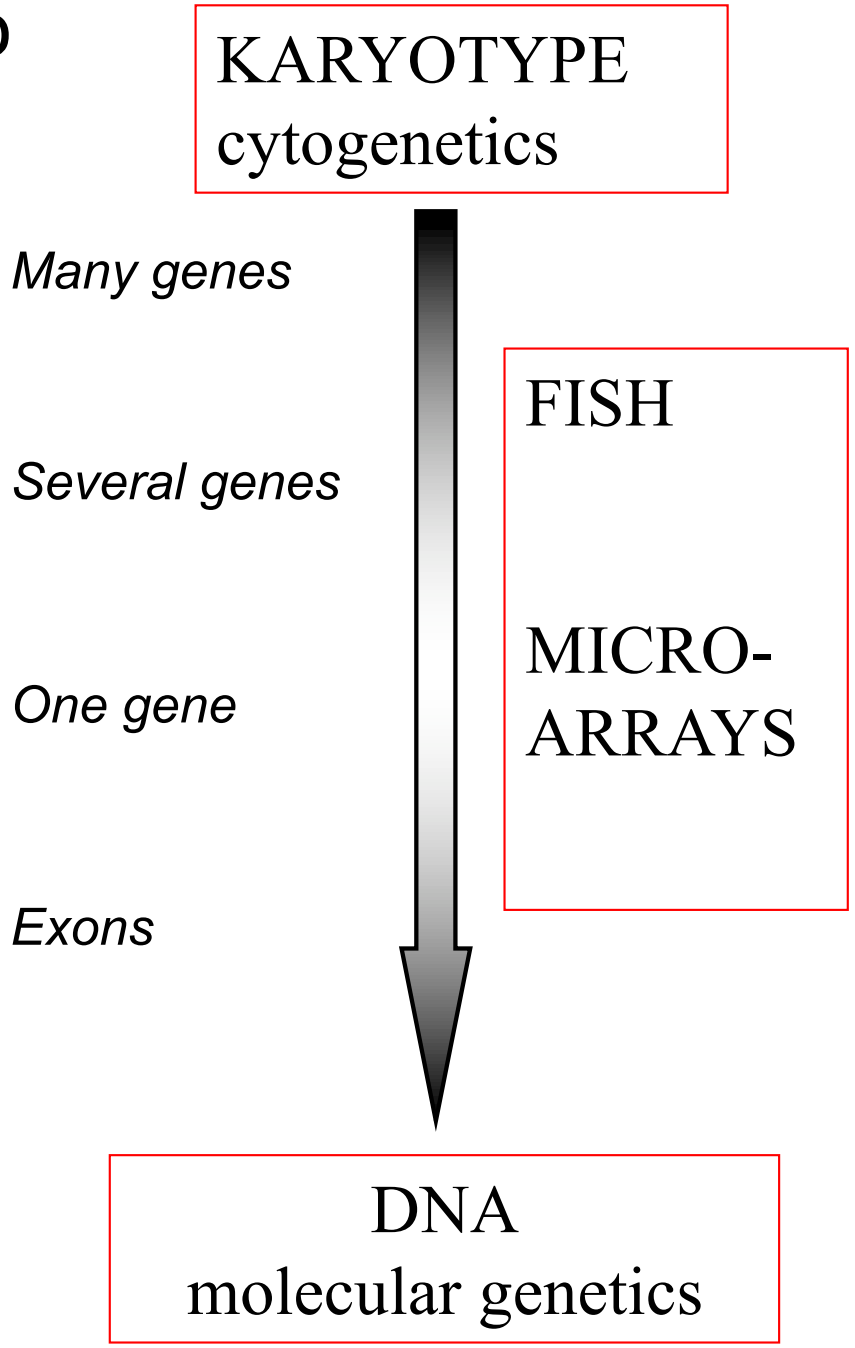
- PMP22 => péripheral myelin protein
- Duplication => 3 doses
- Causes polyneuropathy

✓ 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

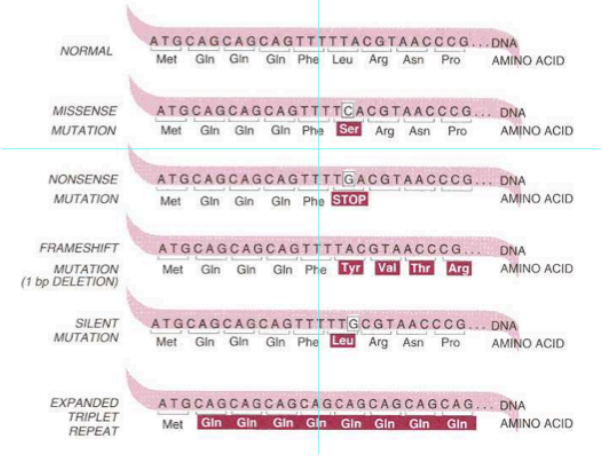
- Many syndromes MCA+MR
- Most hereditary diseases



Various mutations currently need different **methods**

Point mutation

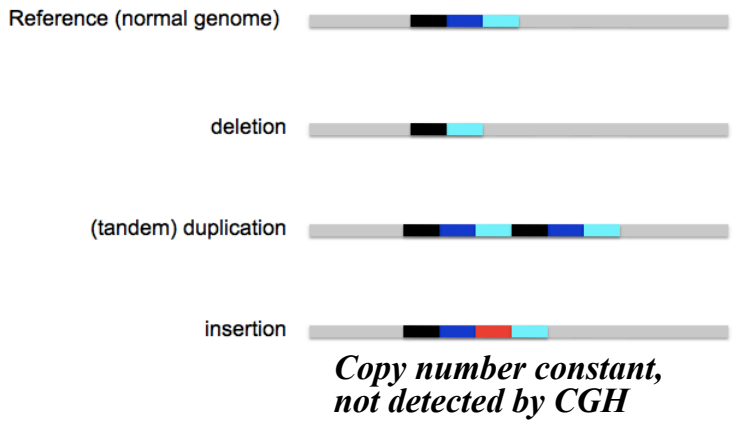
Point mutations (CNVs)
SNVs and small in / dels



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)

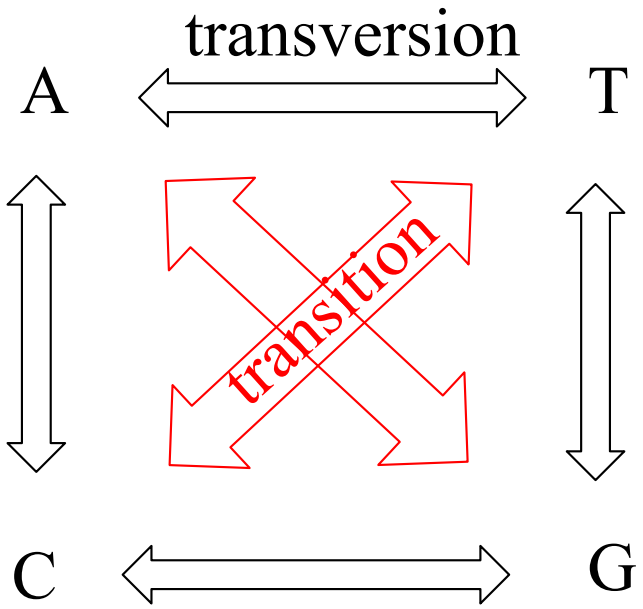
In the future, all types of mutation will be amenable to whole genome sequencing

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

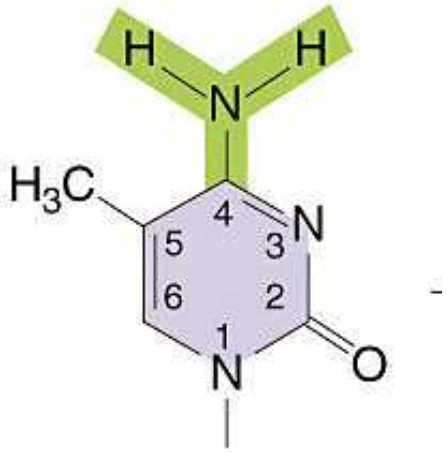
- Most Genome mutations and Chromosome mutation affect survival and/or fertility => fitness ~0

Single-nucleotide point mutations



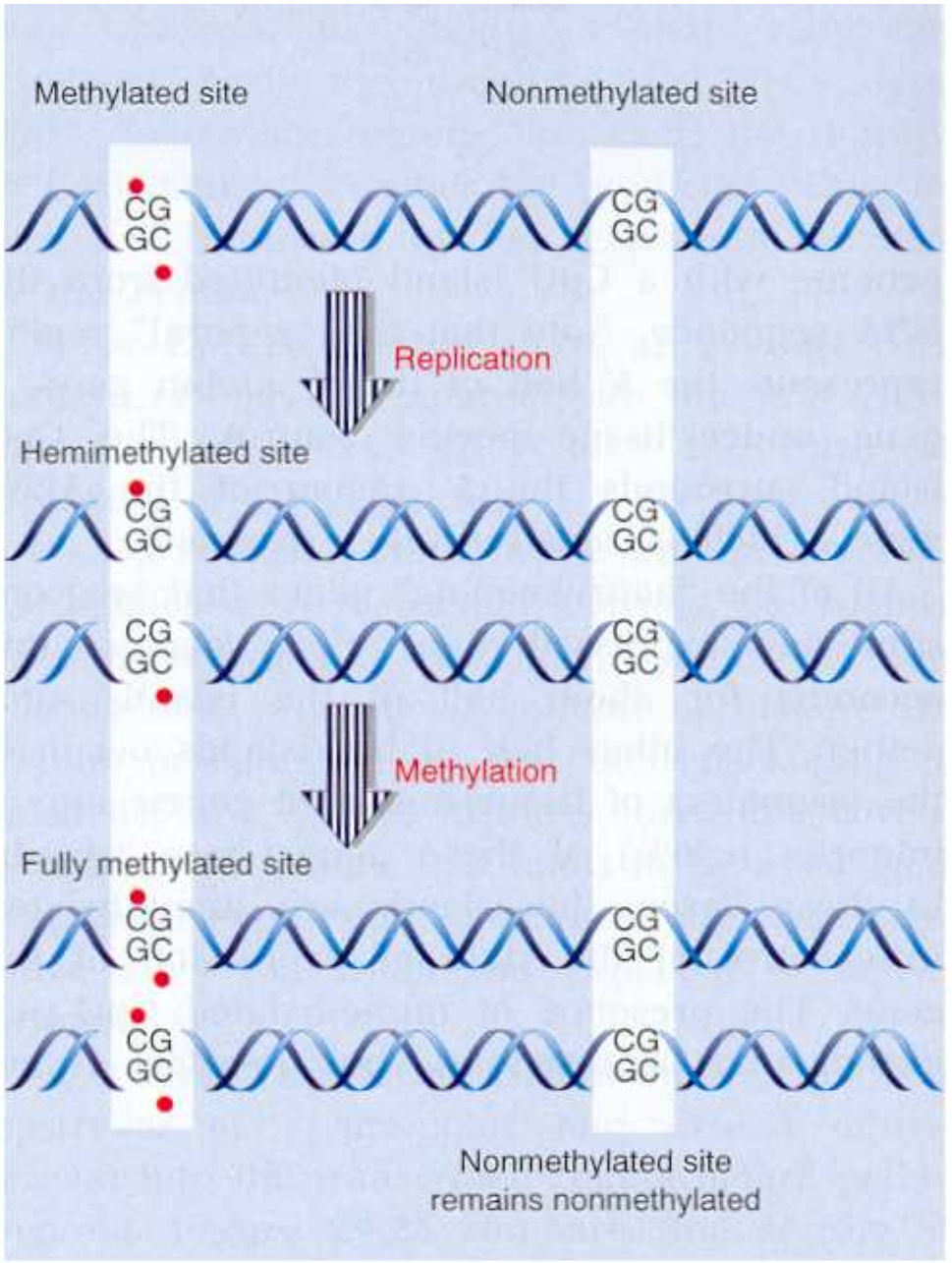
- Transition: purine to purine
- Transversion: pu / py
- Expect 2 x more transversions
- In fact transitions are more frequent
- Most frequent is C>T (G>A) in 5’ CpG 3’ dinucleotide

DNA (hemi-) methylation

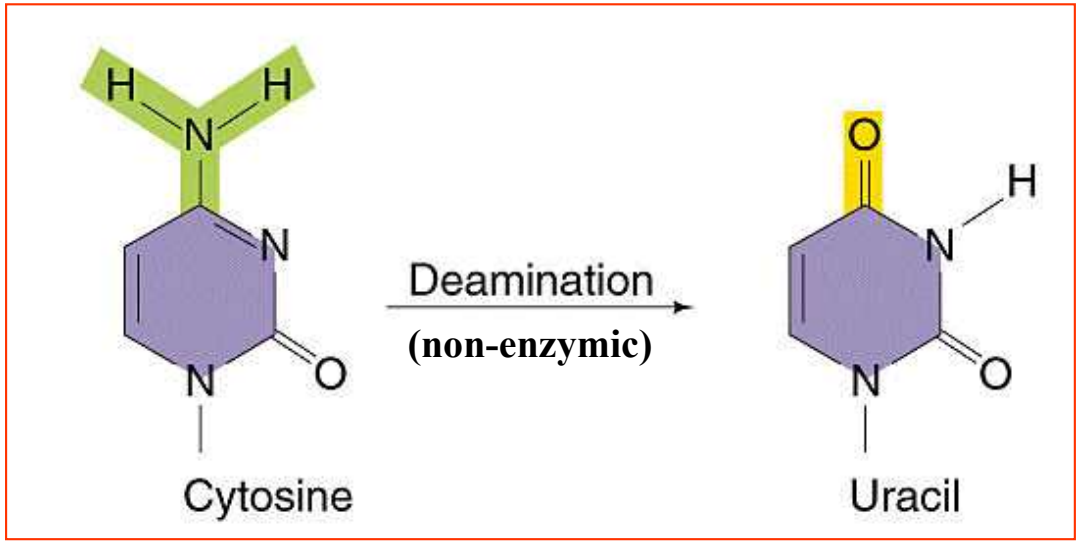


5-Methylcytosine

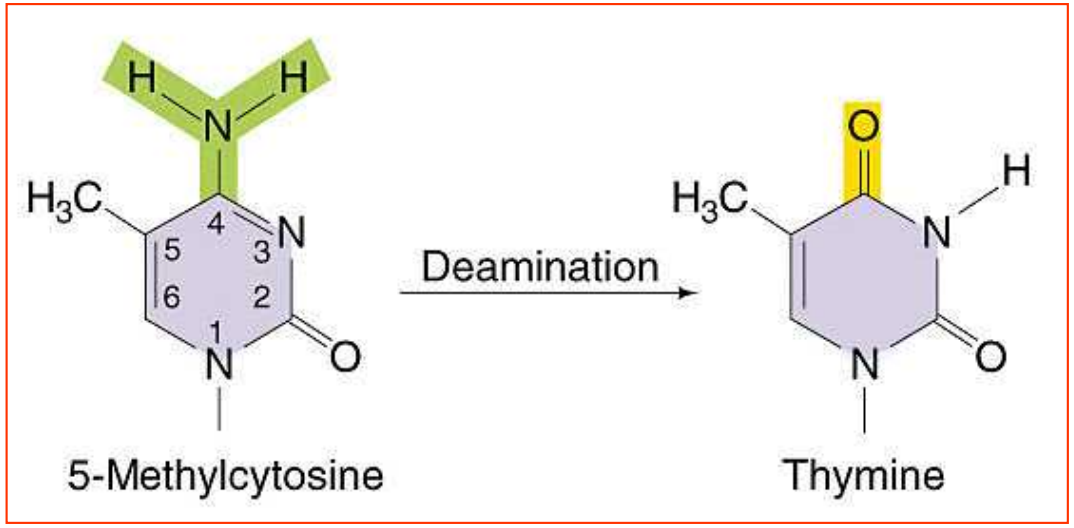
- ^{met}C = 5^{th} base of DNA
- Methylated promoters inactive (usually)
- Implication in gene silencing; imprinting.
- Epigenetic heredity



^{met}Cytosine occasional deamination



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



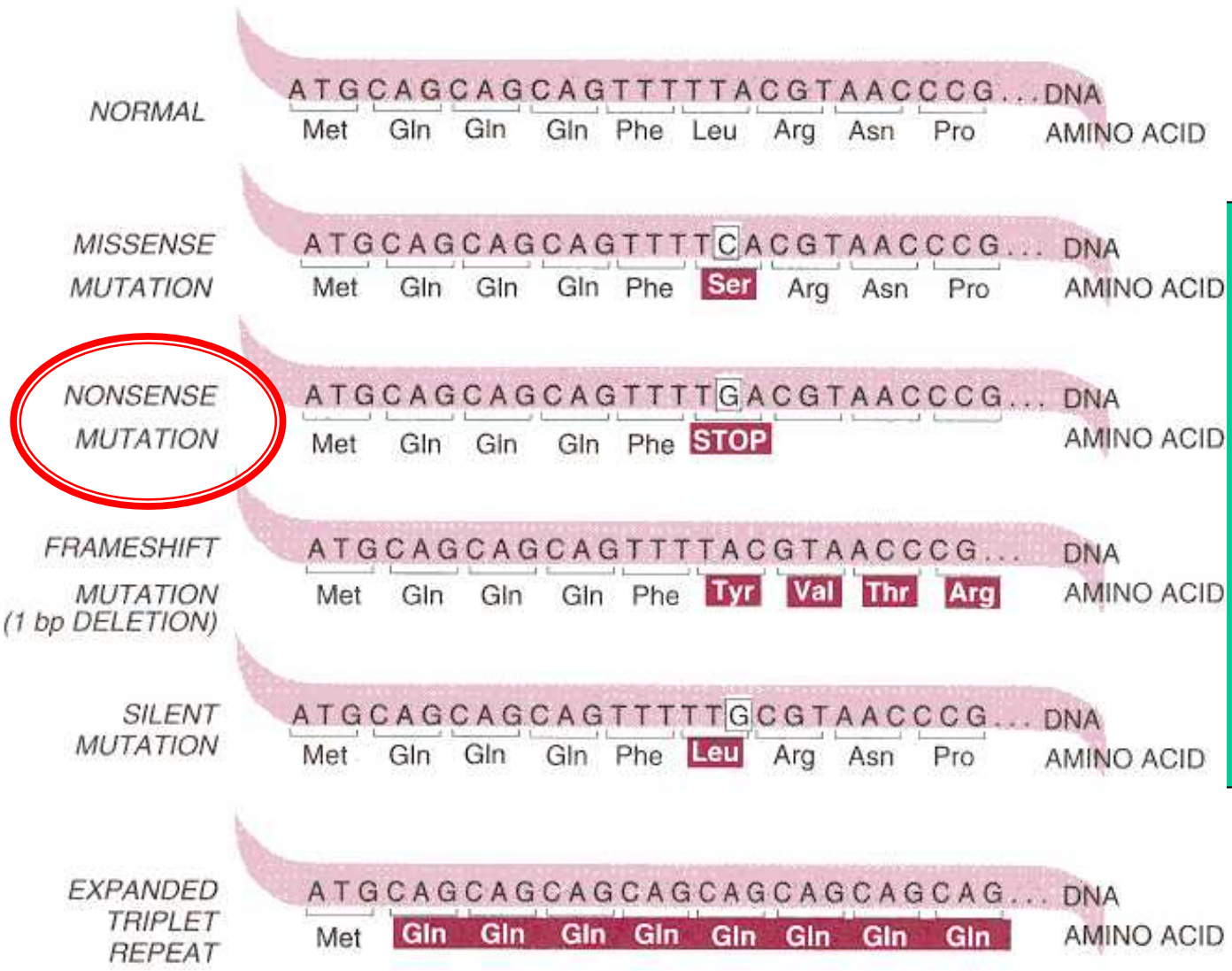
- Thymine not recognized as abnormal
  => mutation remains

MUTATIONS IN CODING SEQUENCES

Point mutations

Affect 1 or a few base pairs

MUTATIONS IN CODING SEQUENCES

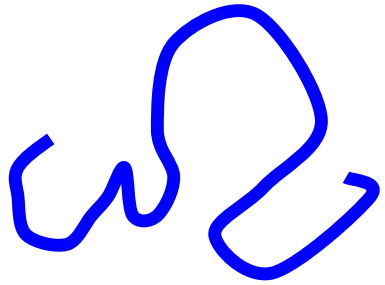


Point mutations

Affect 1 or a few base pairs

Nonsense mutation generally causes null allele

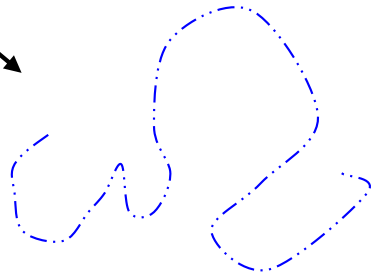
Wt allele > mRNA > normal protein:



Nonsense Mutation > Premature Termination Codon (PTC)

OR

Truncated protein product



Unstable mRNA > decay

Usually complete loss of function

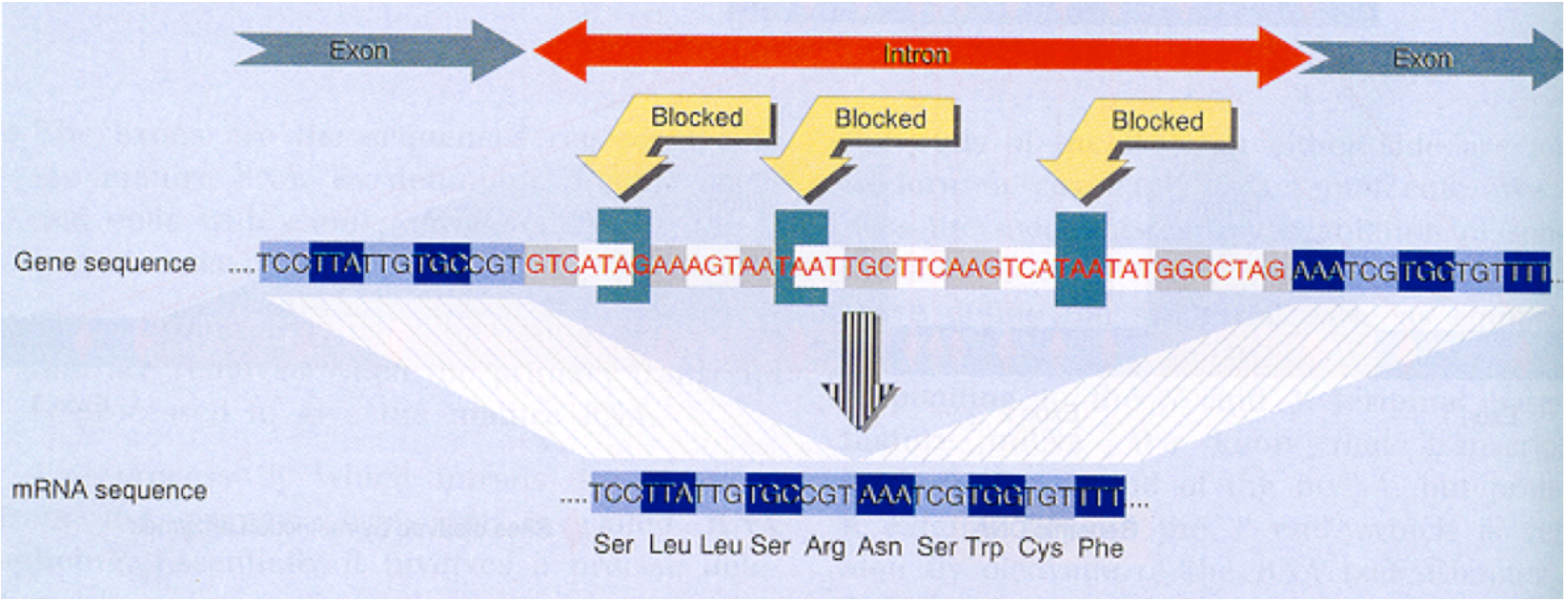
always complete loss of funtion

MUTATIONS IN CODING SEQUENCES

Point mutations

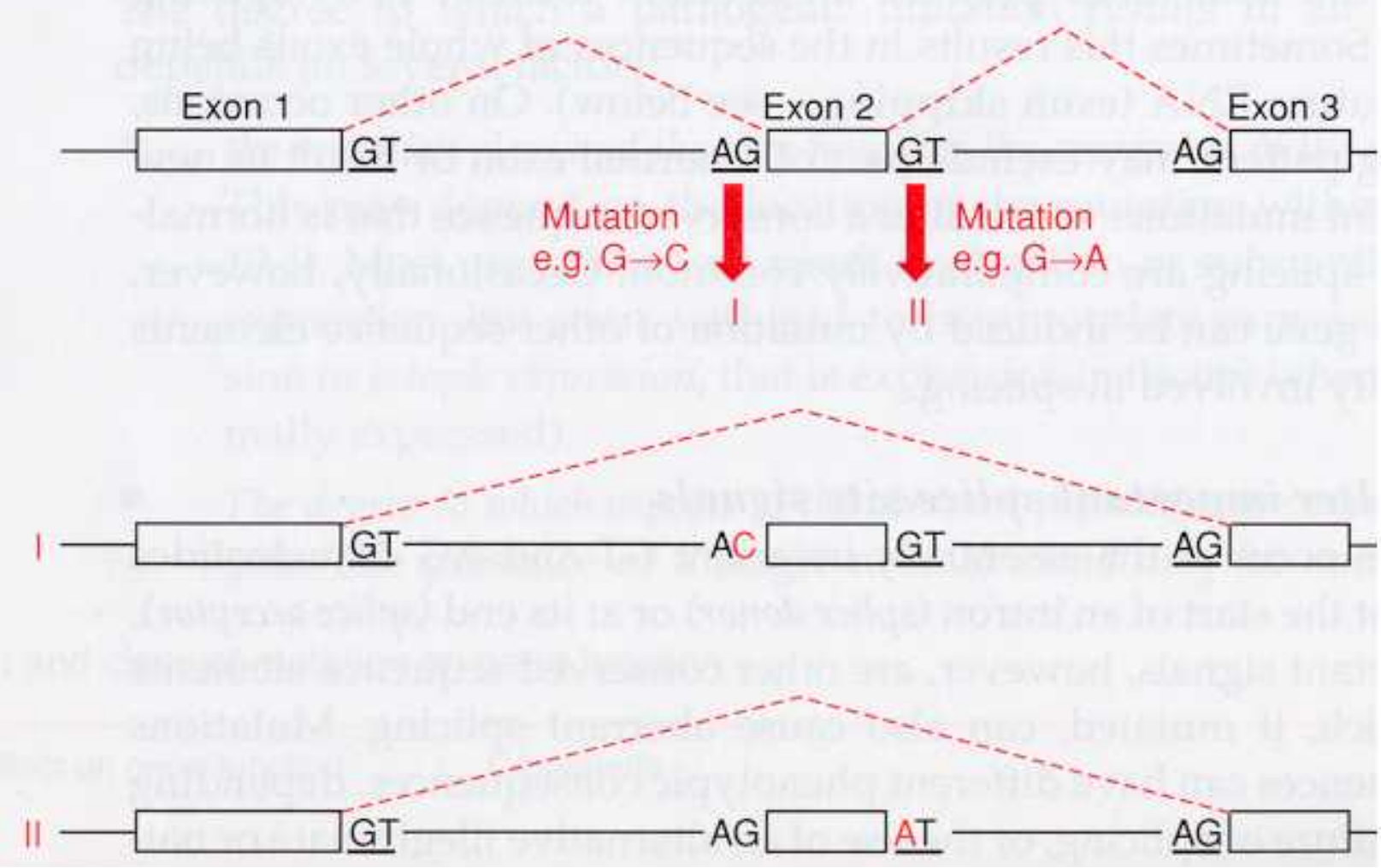
Affect 1 or a few base pairs

Frameshift mutations and premature termination codons (PTCs)



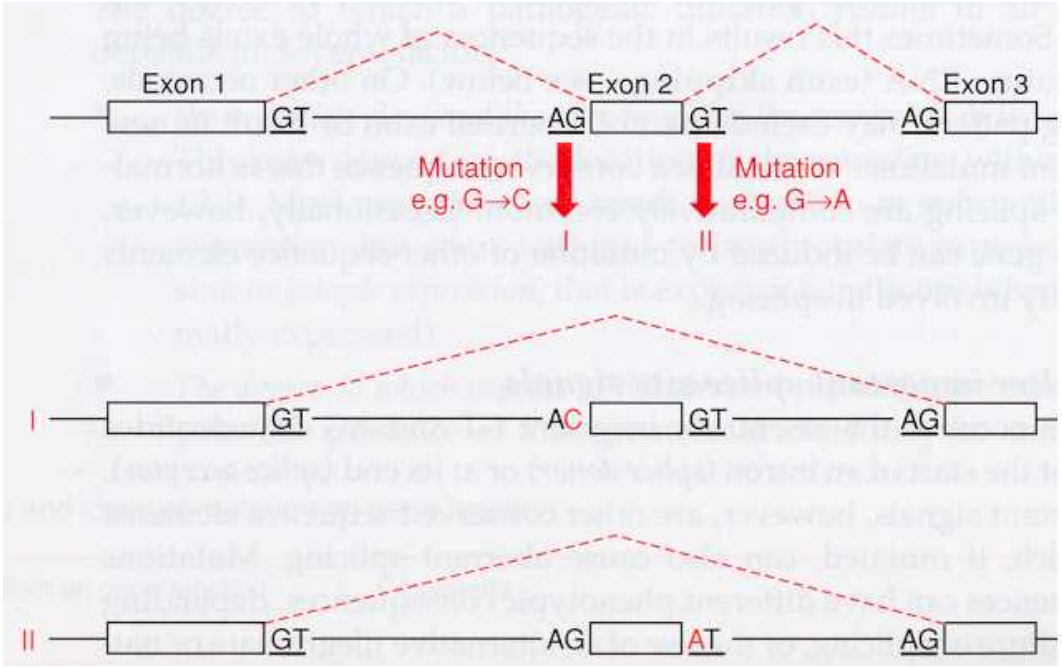
- Normal reading frame is open (ORF): no premature stop codon
- The two other phases (normally unread) contain many stop codond (ex: TAG, TAA)
- hence, a mutation that shifts the triplet frame of codon reading (frameshift) has 2 consequences
    1. Changes all downstream AA
    2. Then premature termination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

Splice-site Mutation

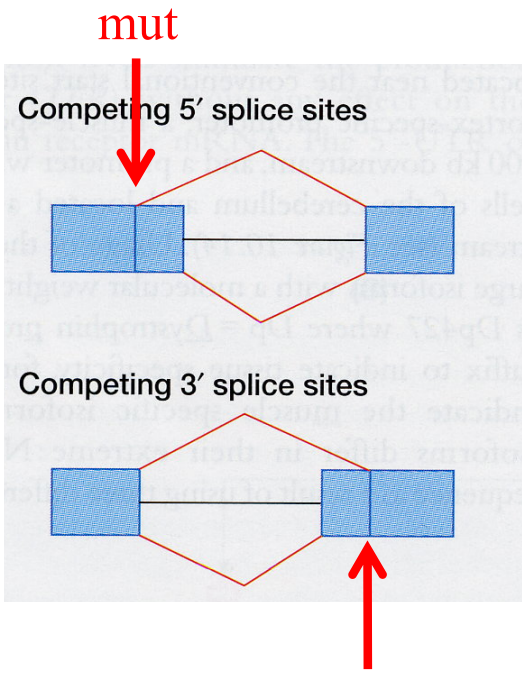


exon skipping

Splice Mutations



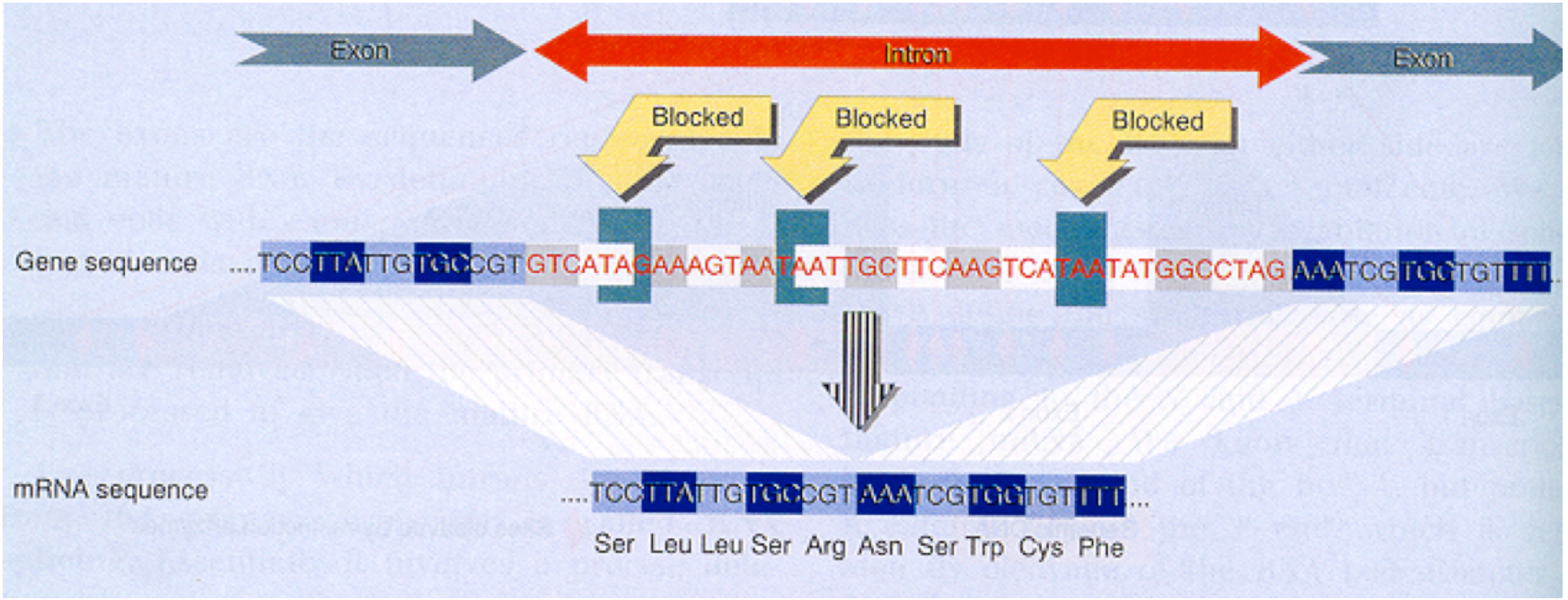
exon skipping



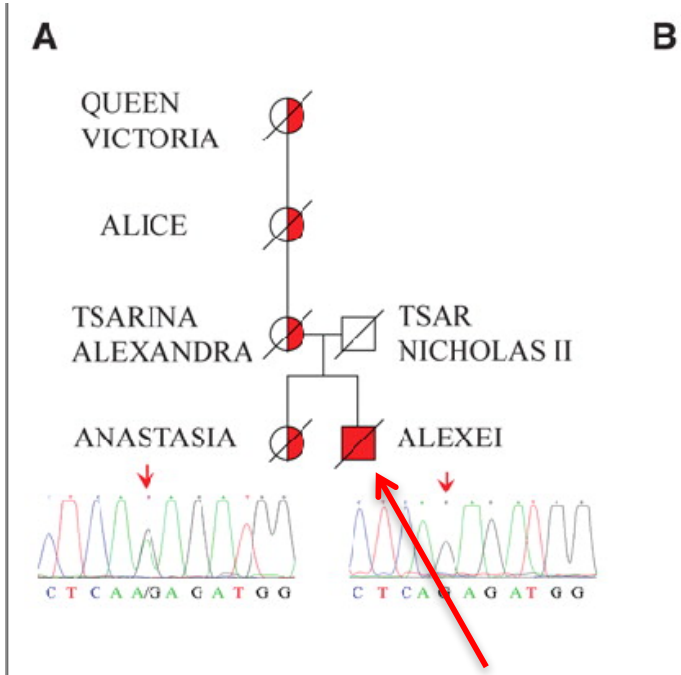
Retention of a piece of intron

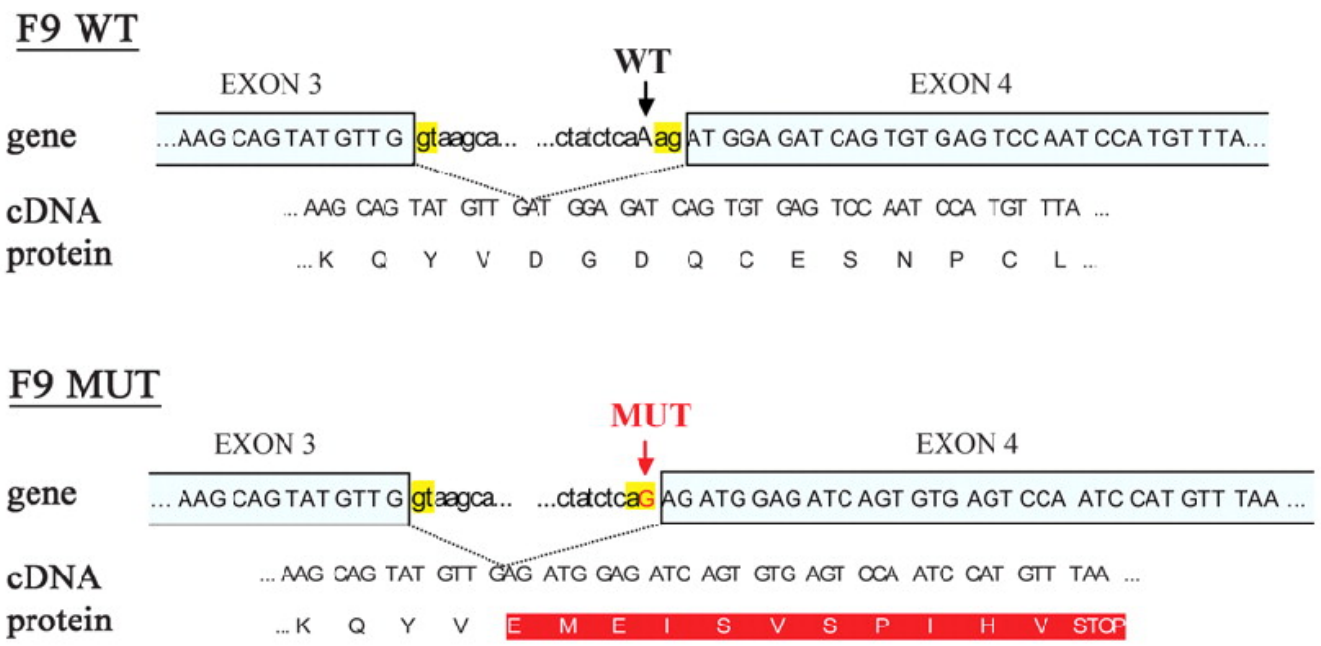
Splice Mutations

- *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. *Ajout / perte d'un morceau de protéine*
    2. *Modification des AA en aval de l'ajout / perte*
    3. *Codon stop en aval de tout cela*
- *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)

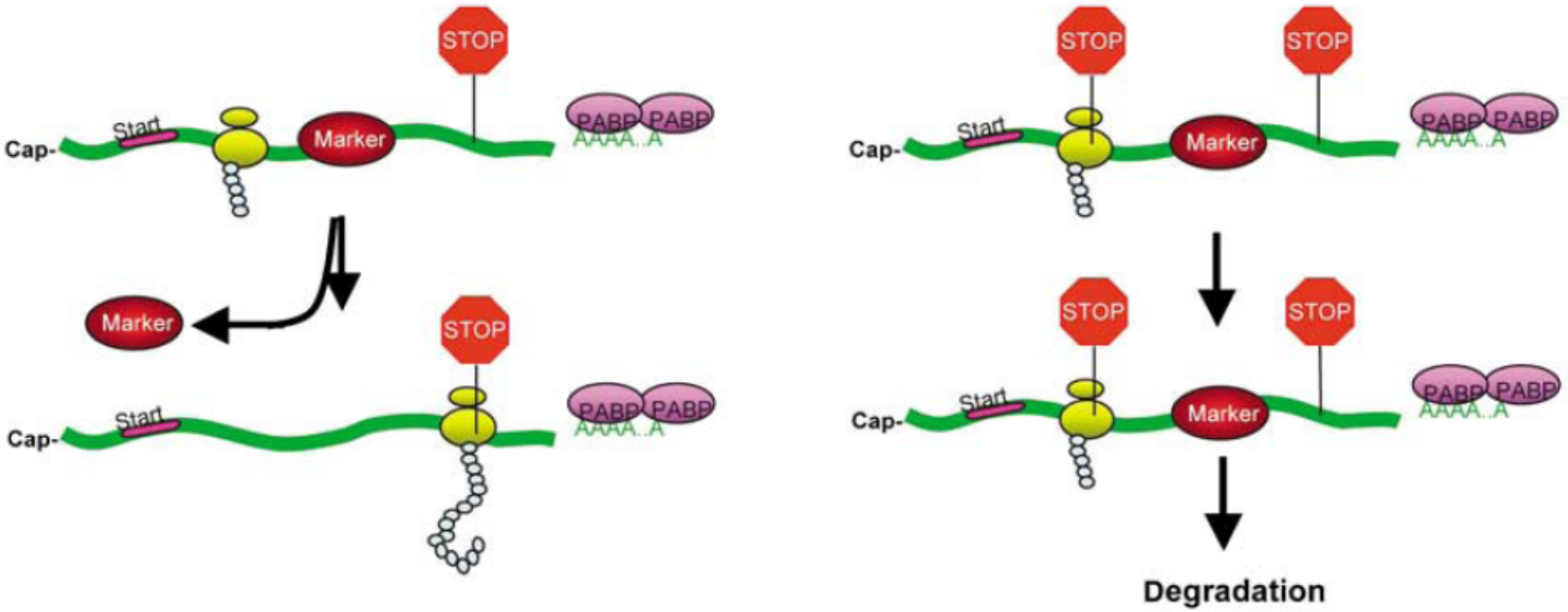
- Truncates the reading frame
- Several types of mutations
  - nonsense
  - Indel (small insertion/small deletion) (not multiple of 3)
  - Splice mutations

- Often causes nonsense-mediated mRNA decay

Nonsense-mediated decay of mRNA.

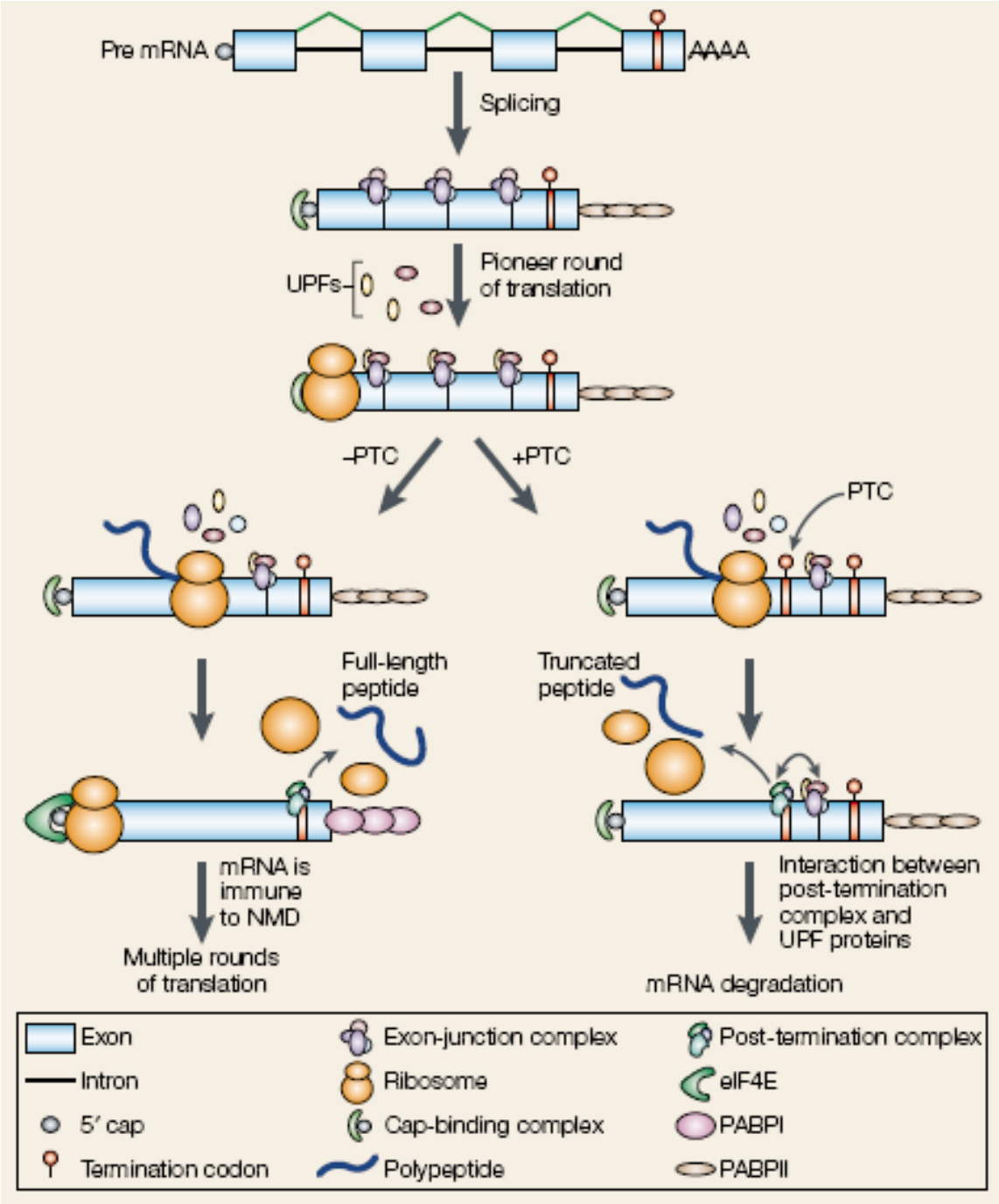
NORMAL

PREMATURE TERMINATION CODON



Shyu et al 2008 EMBO J

Marker for degradation = Exon Junction Complex

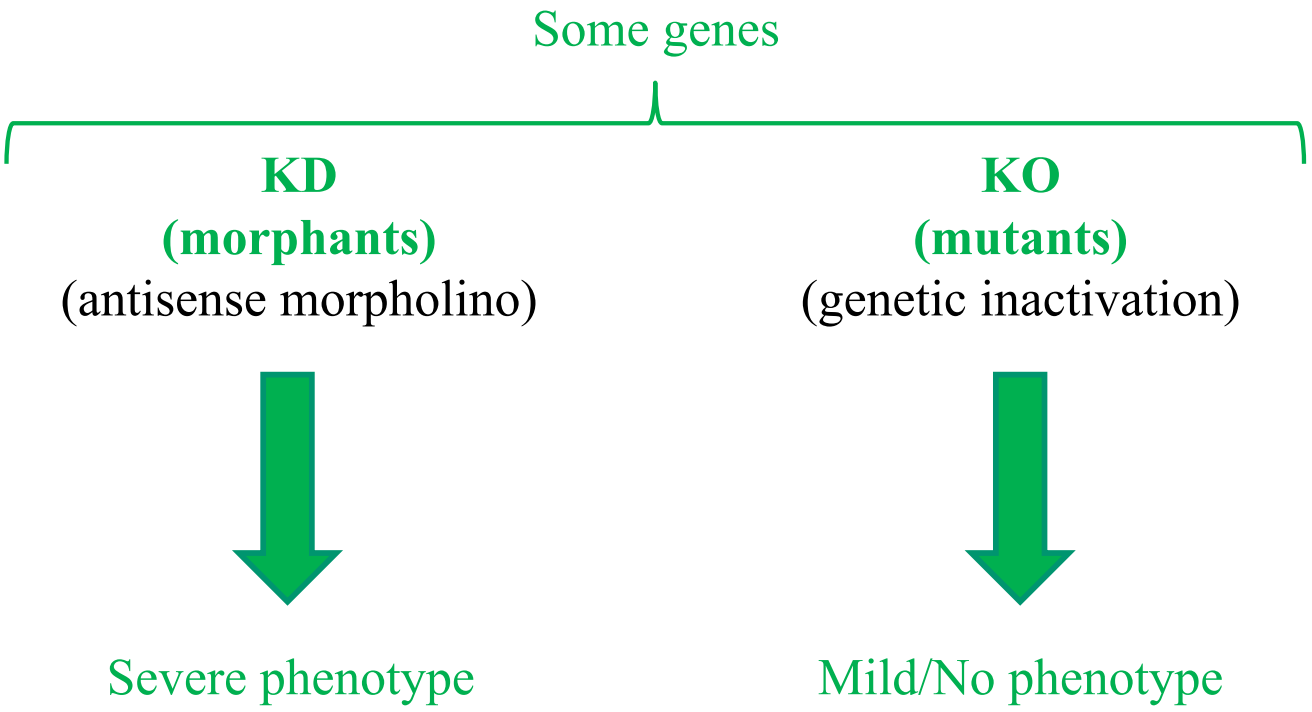


Cartegni et al 2002 Nat Rev Genet

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)



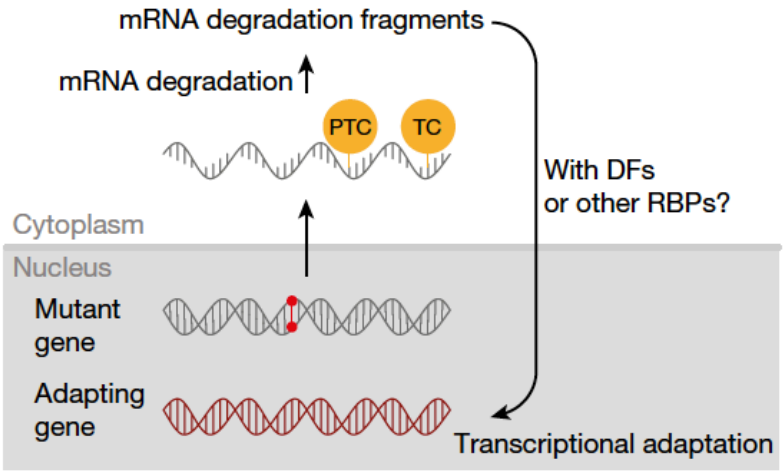
- ❑ Off target effect of morpholino
- ❑ Toxicity of excipient
- ❑ Genetic compensation response (transcriptional regulation)

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

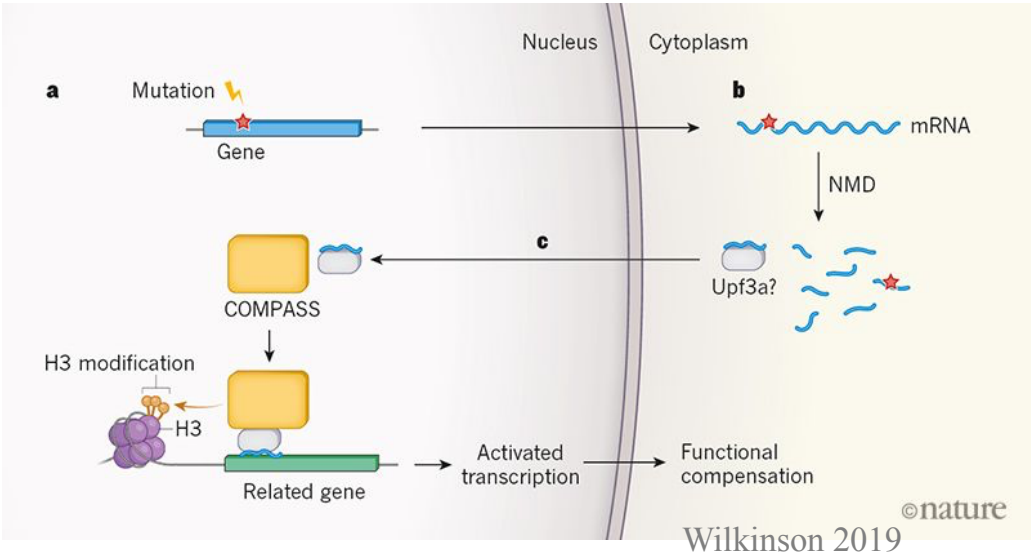
El Brolosy et al. 2019 Nature ; Ma et al. 2019 Nature

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

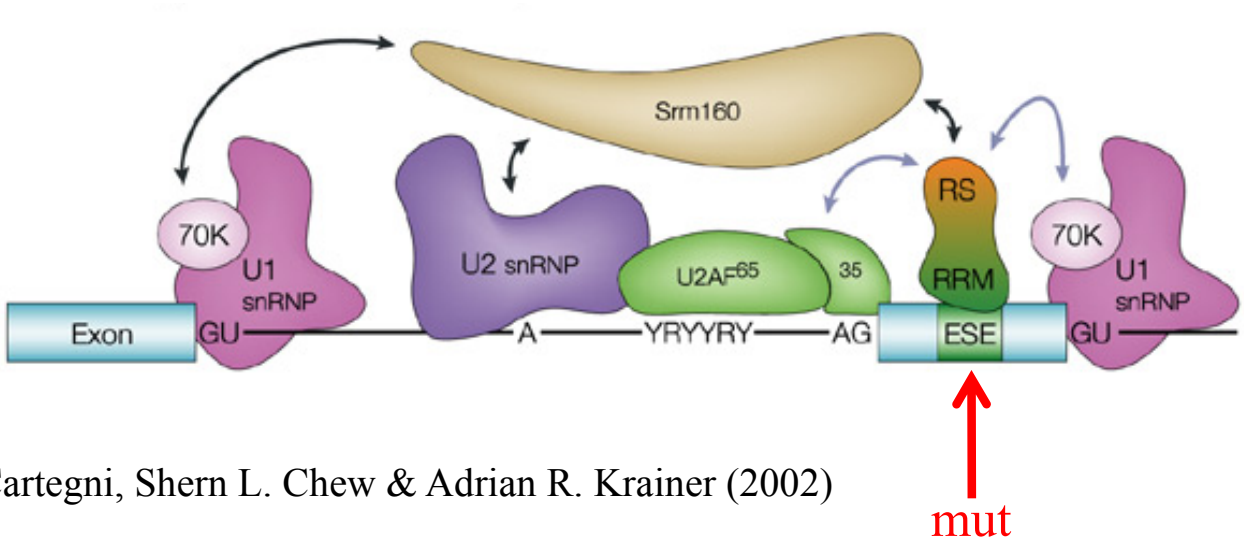
MUTATIONS IN CODING SEQUENCES

Point mutations

Affect 1 or a few base pairs

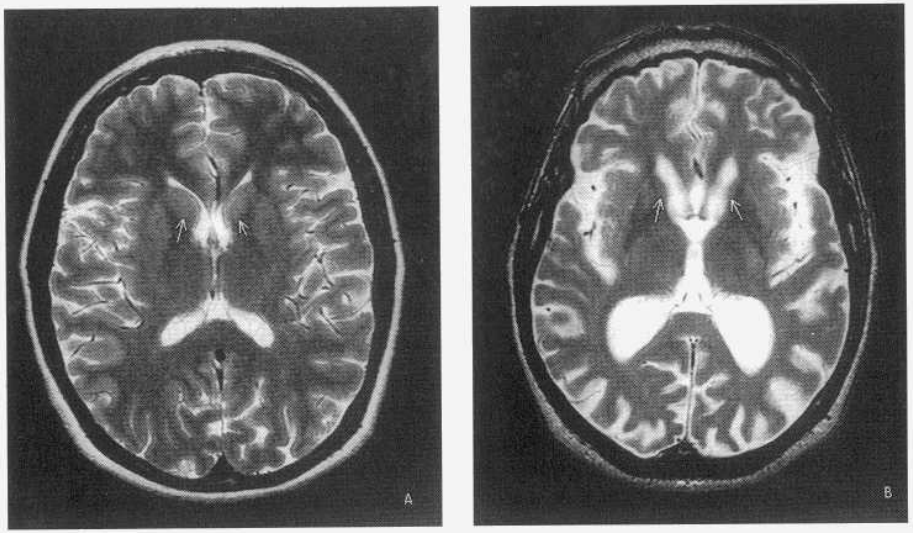
Mutations that do not change an AA

- Often 3rd base ofcodons (the genetic code is « degenerate »)
- IN PRINCIPLE 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



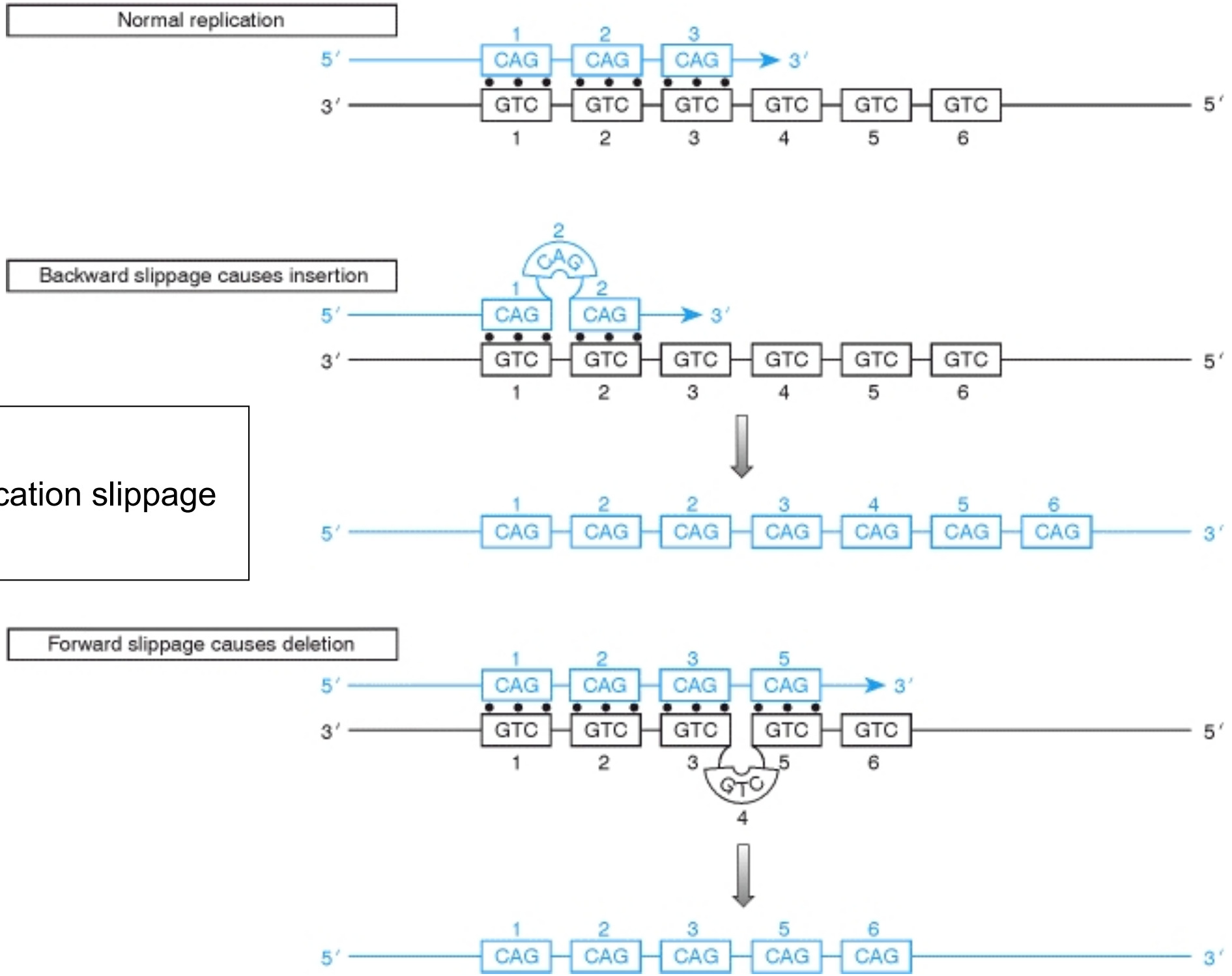
(a) Normal volunteer
(Courtesy of Dr M. Lowry, Hull, UK.)

(b) Huntington's disease

Huntington Disease

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

Replication slippage



MUTATION in SINGLE GENE (gene mutation)

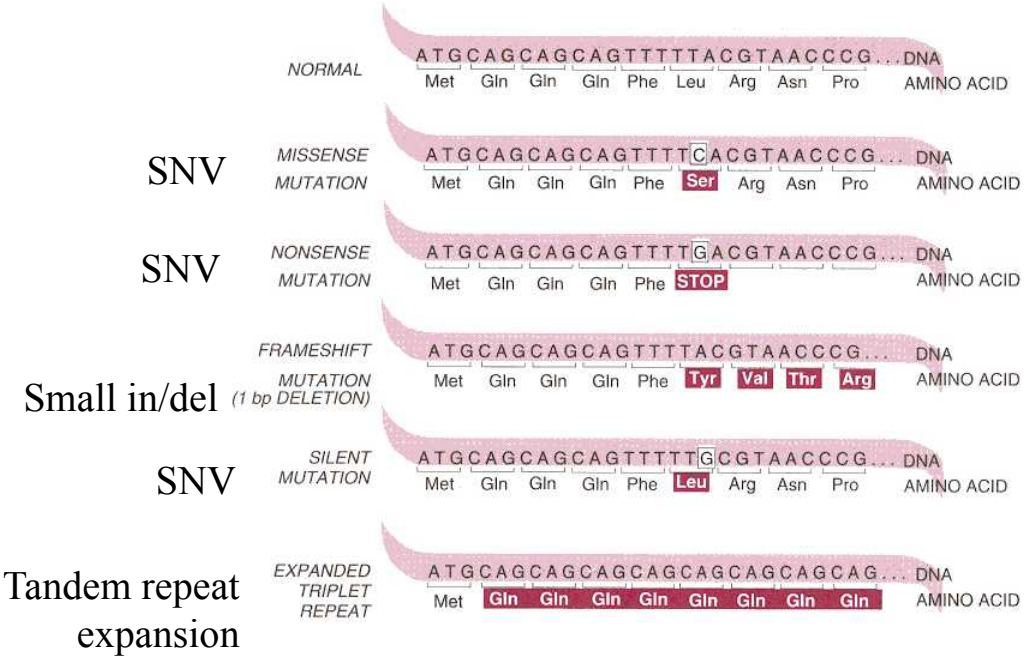
MUTATION IN NONCODING SEQUENCE
Promoter, enhancer, silencer

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

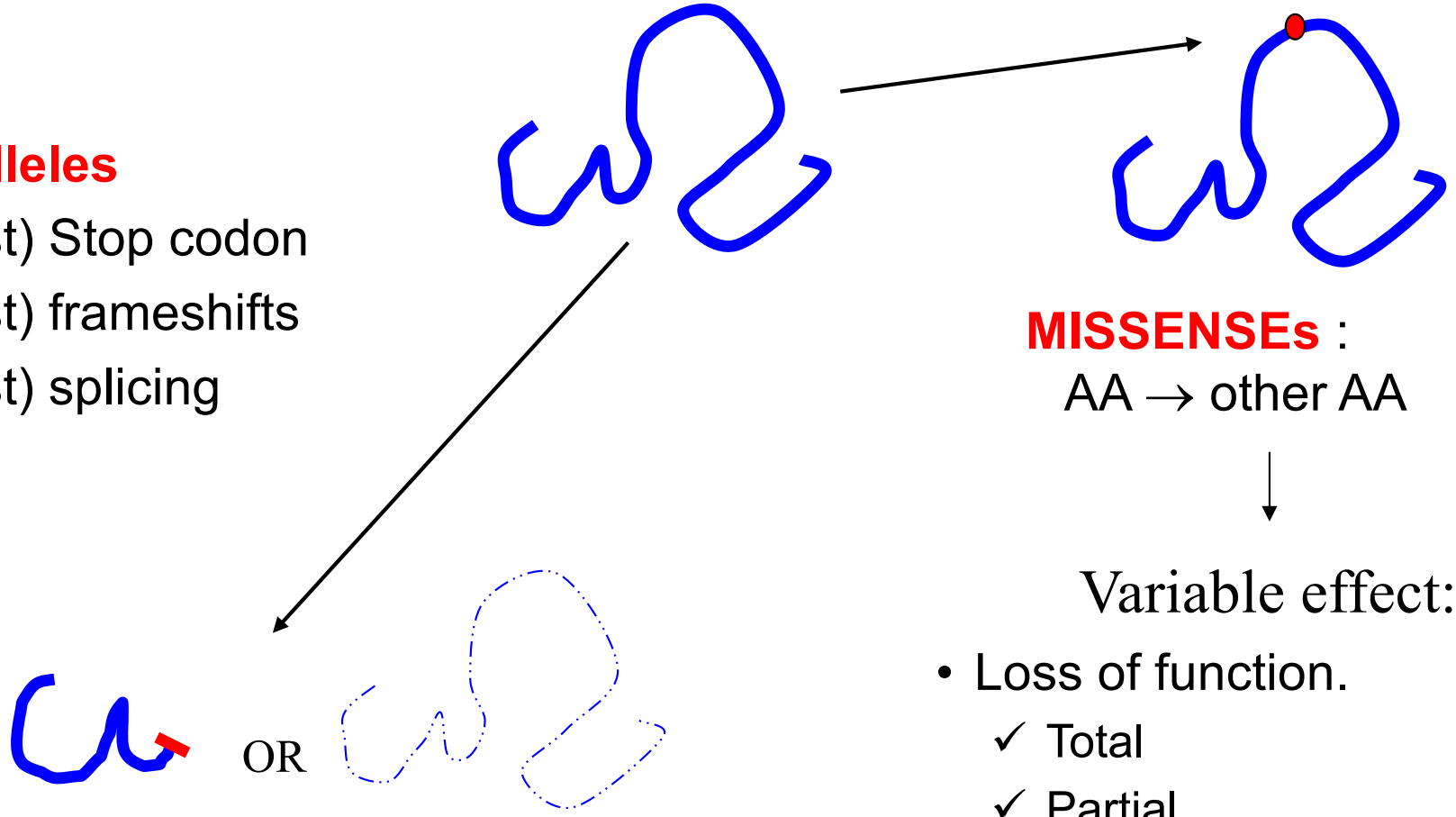
Point mutations

Affect 1 or a few base pairs

Functional effect of coding mutations

Null alleles

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



Loss of function (null allele)

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

MISSENSEs :

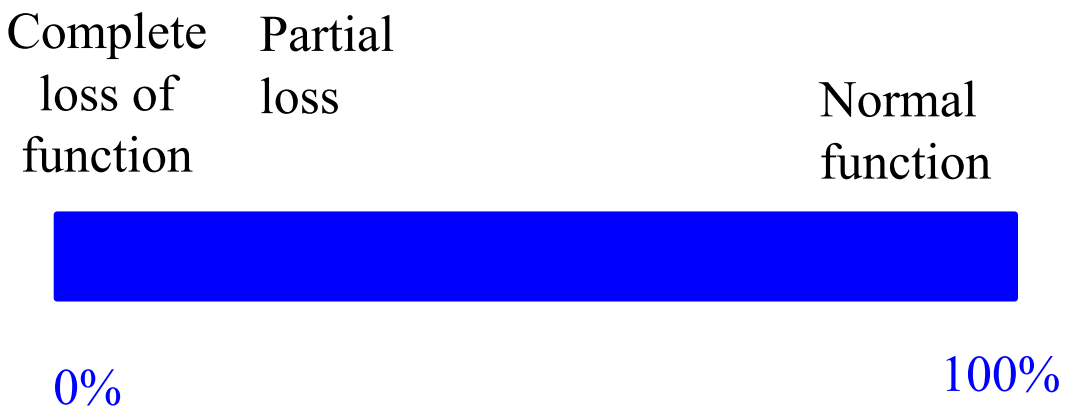
AA → other AA

Variable effect:

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

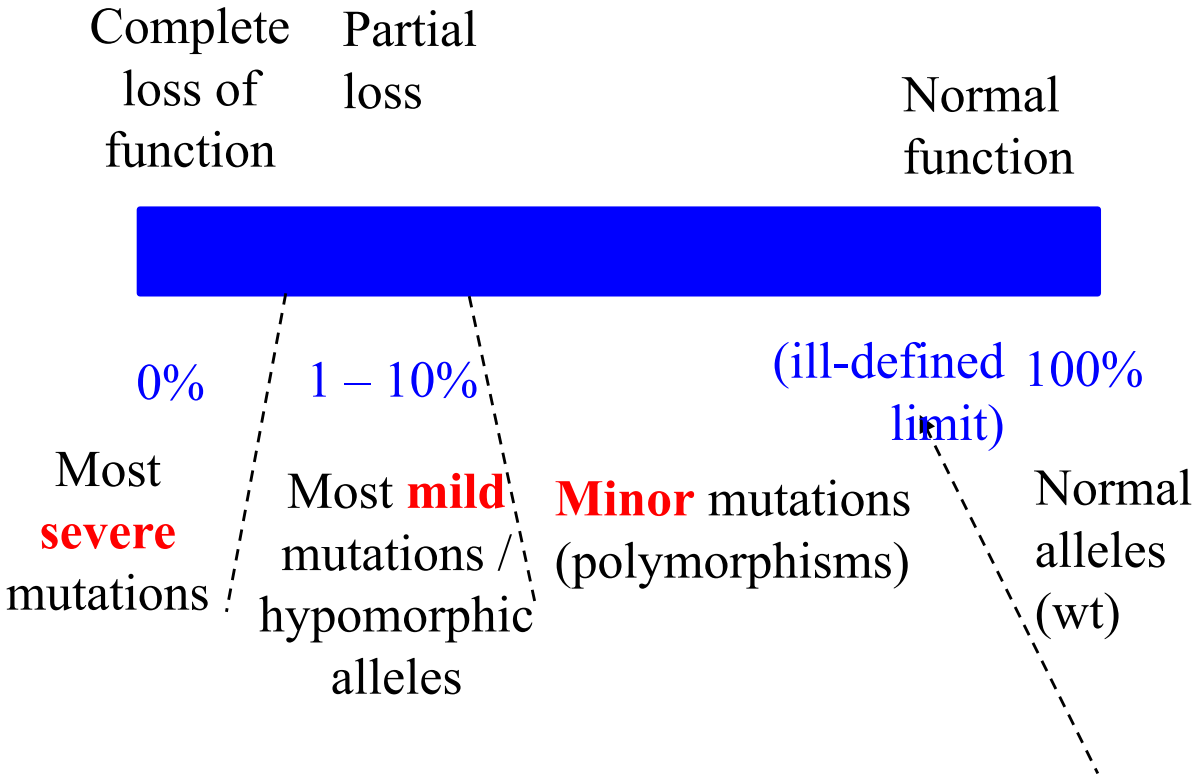
Functional effet, **loss of function** type

- Quantitative effect, with continuum



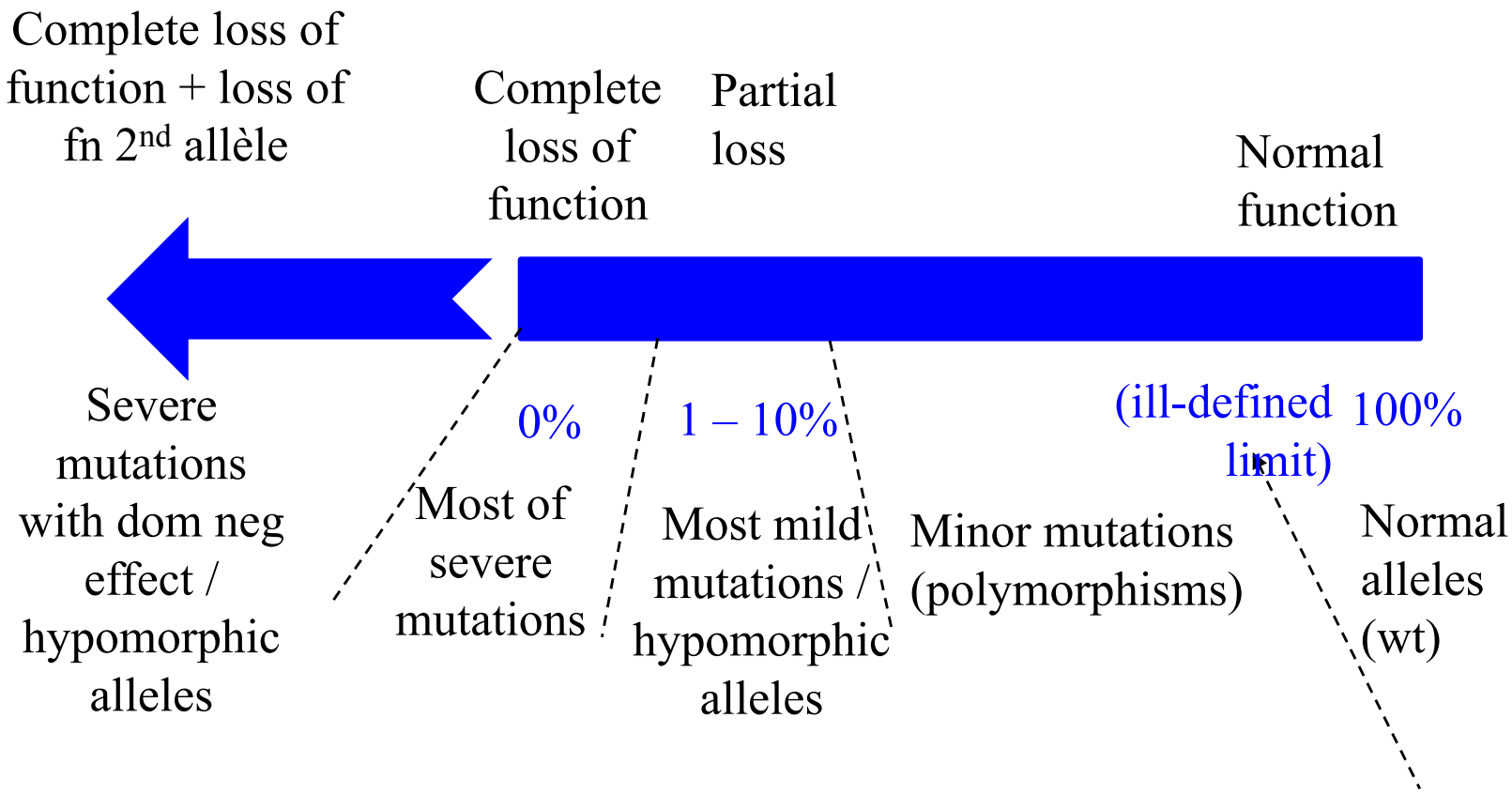
Functional effet, **loss of function** type

- 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

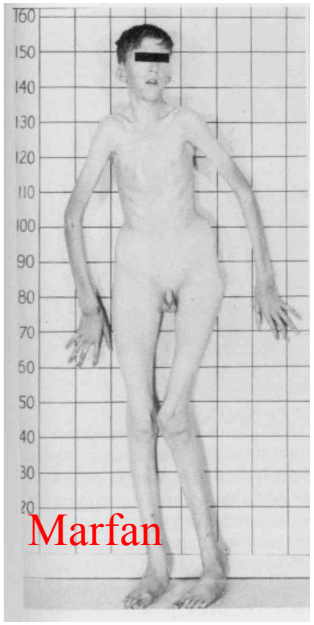
Partial loss

Normal activity

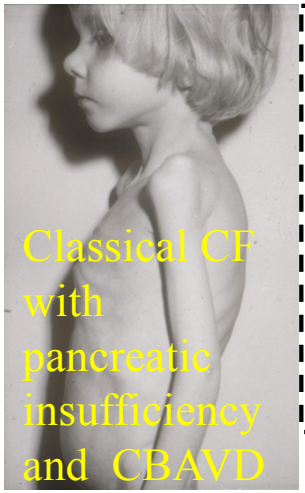
Gain of function:
- Dose effect
- Gain toxic effect

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

Antimorphs:
- Missenses (most)



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



Mild mutations
Missenses,
Or others

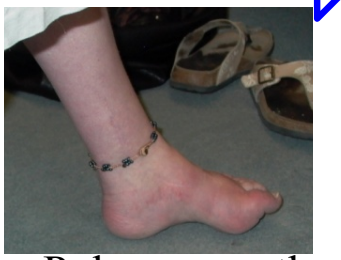
Hypomorphic alleles

CF without pancreatic insufficiency,

Or
Isolated CBAVD

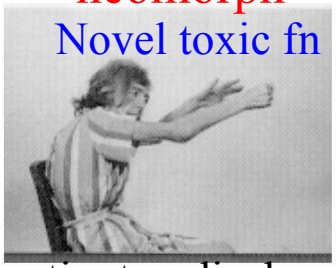
Wild type alleles (wt) and Neutral polymorphisms :
- Intronic
- 3rd base codons, …

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

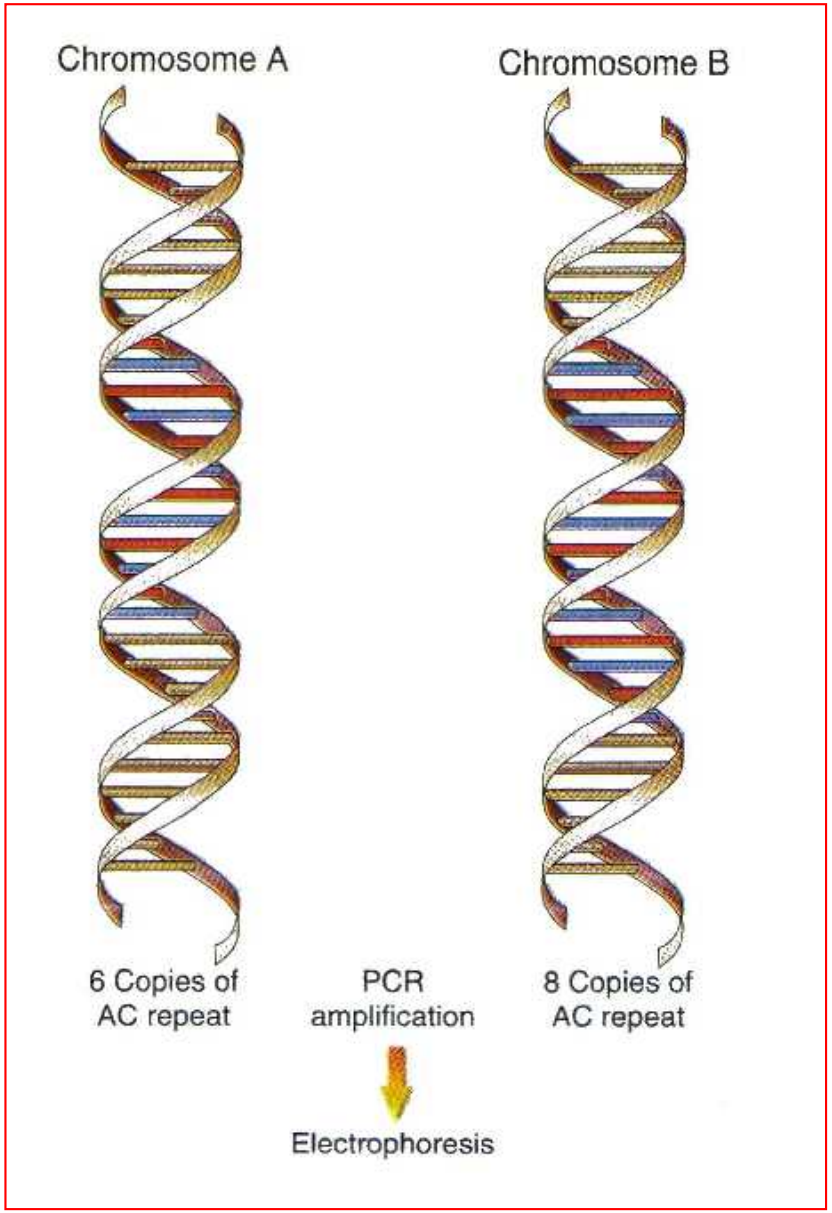
- Daltonism mutation
- HFE*C282Y
- CFTR*DF508
  (! According to definition, DF508 is a human polymorphism !)

Polymorphic markers

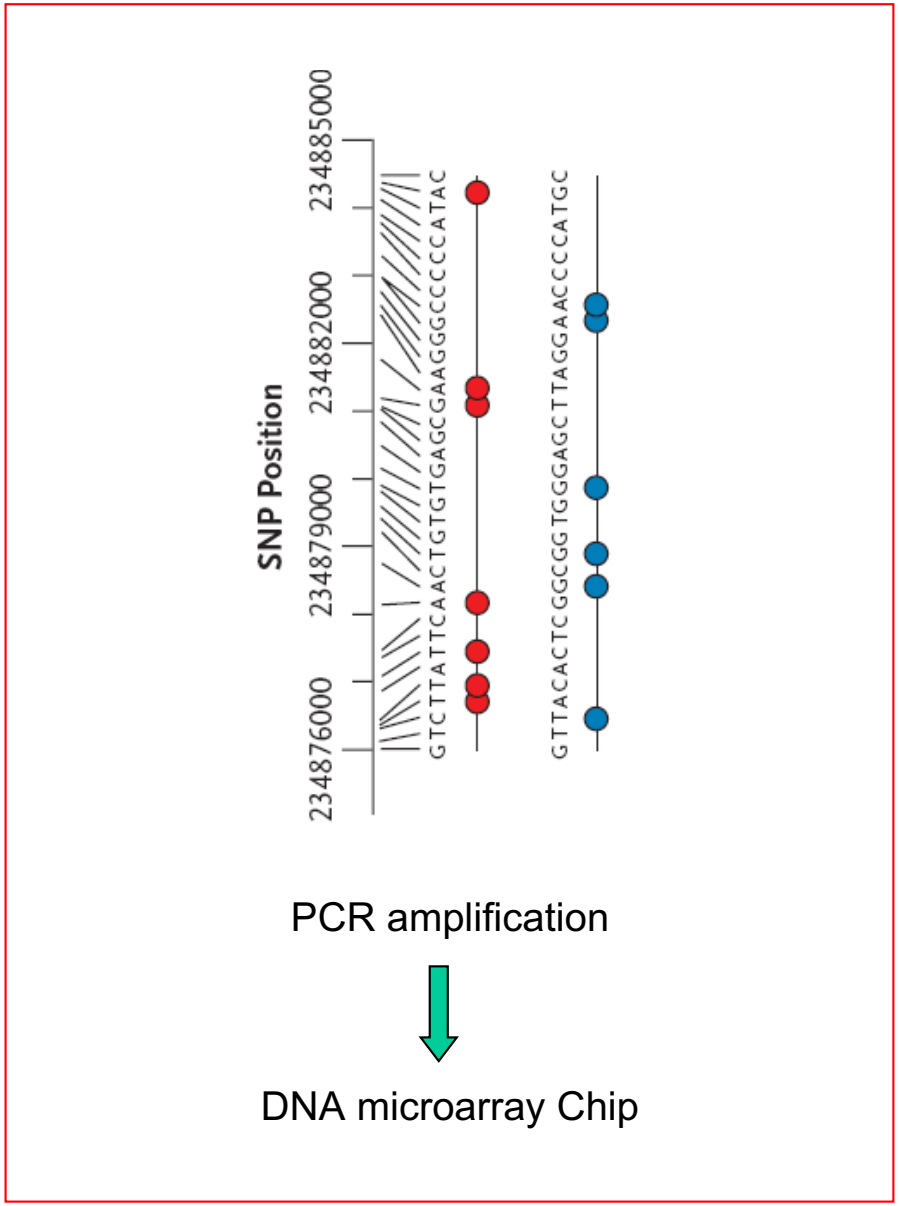
- Neutral polymorphisms, frequent in population (typically MAF >.05)
  - Minisatellites (obsolete)
  - Microsatellites (= short tandem repeats)
  - SNPs
  - others
- May serve as markers of chromosomal segment
  - Linkage studies, in families
  - Association studies, in populations (Gwas)

Genotyping polymorphic markers

Microsatellite



SNPs



Single Nucleotide Polymorphism (SNP)

…5’ AATCGAGG 3’…
…3’ TTAGCTCC 5’…

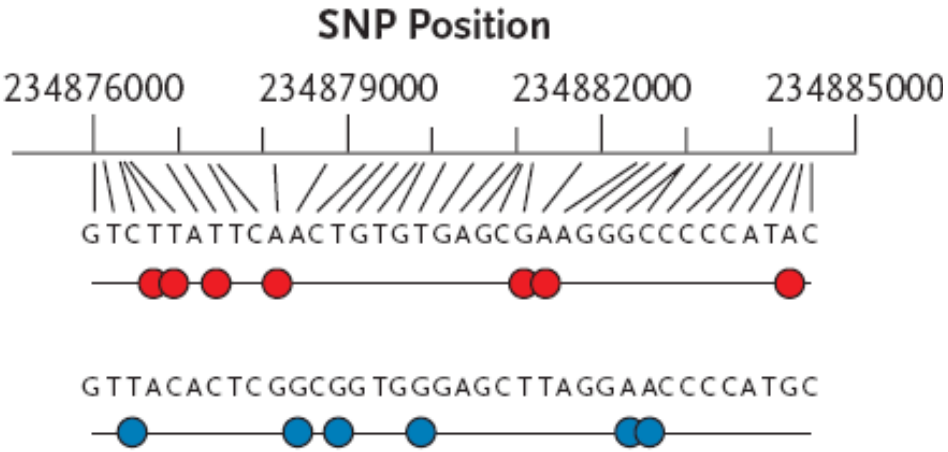
Allele 1 (allele C), frequency = p

…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

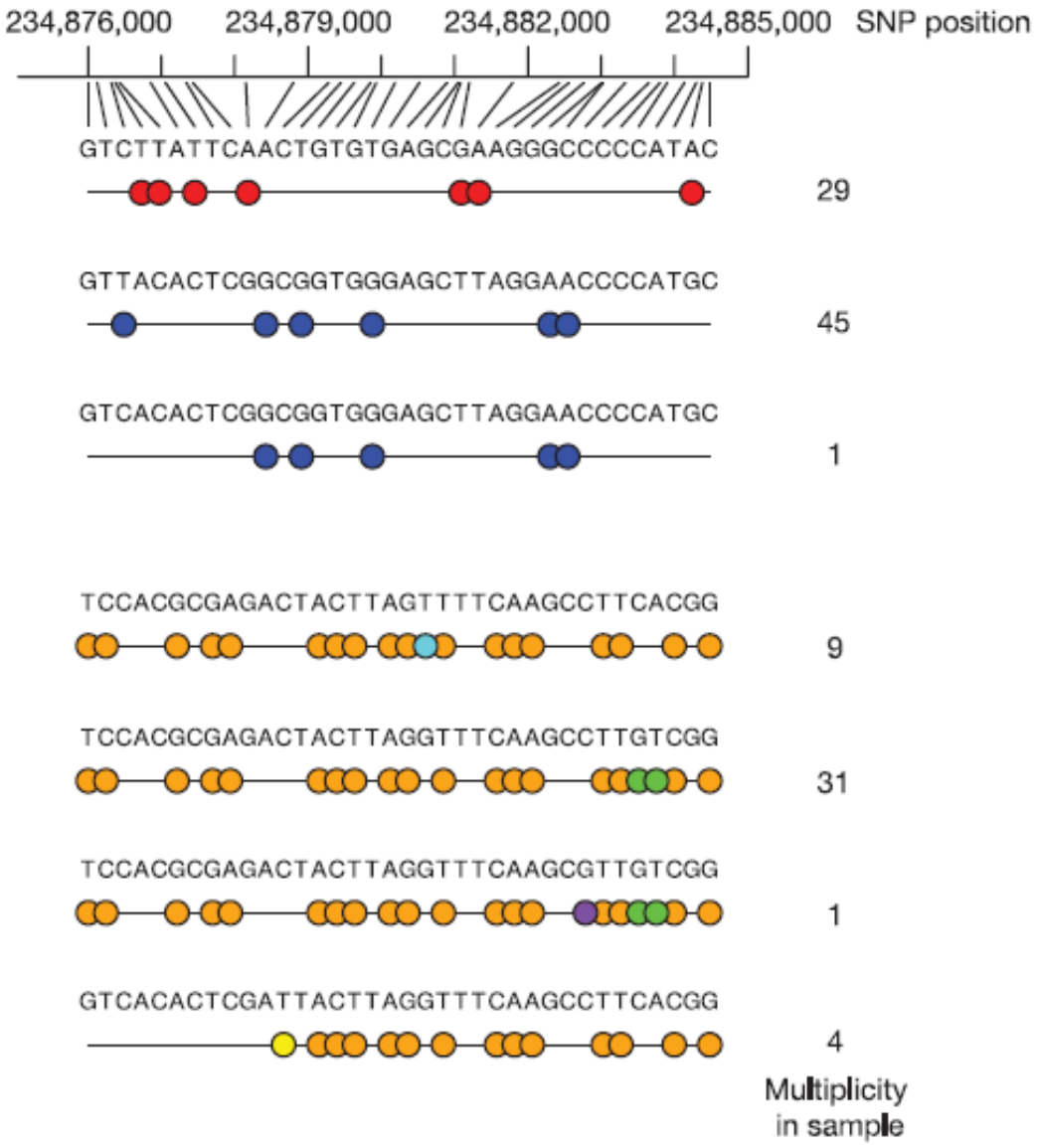


- millions of SNPs in genome

- Many CNPs (copy number polymorphisms)

Haplotype = sequence of alleles on a short piece of chromosome

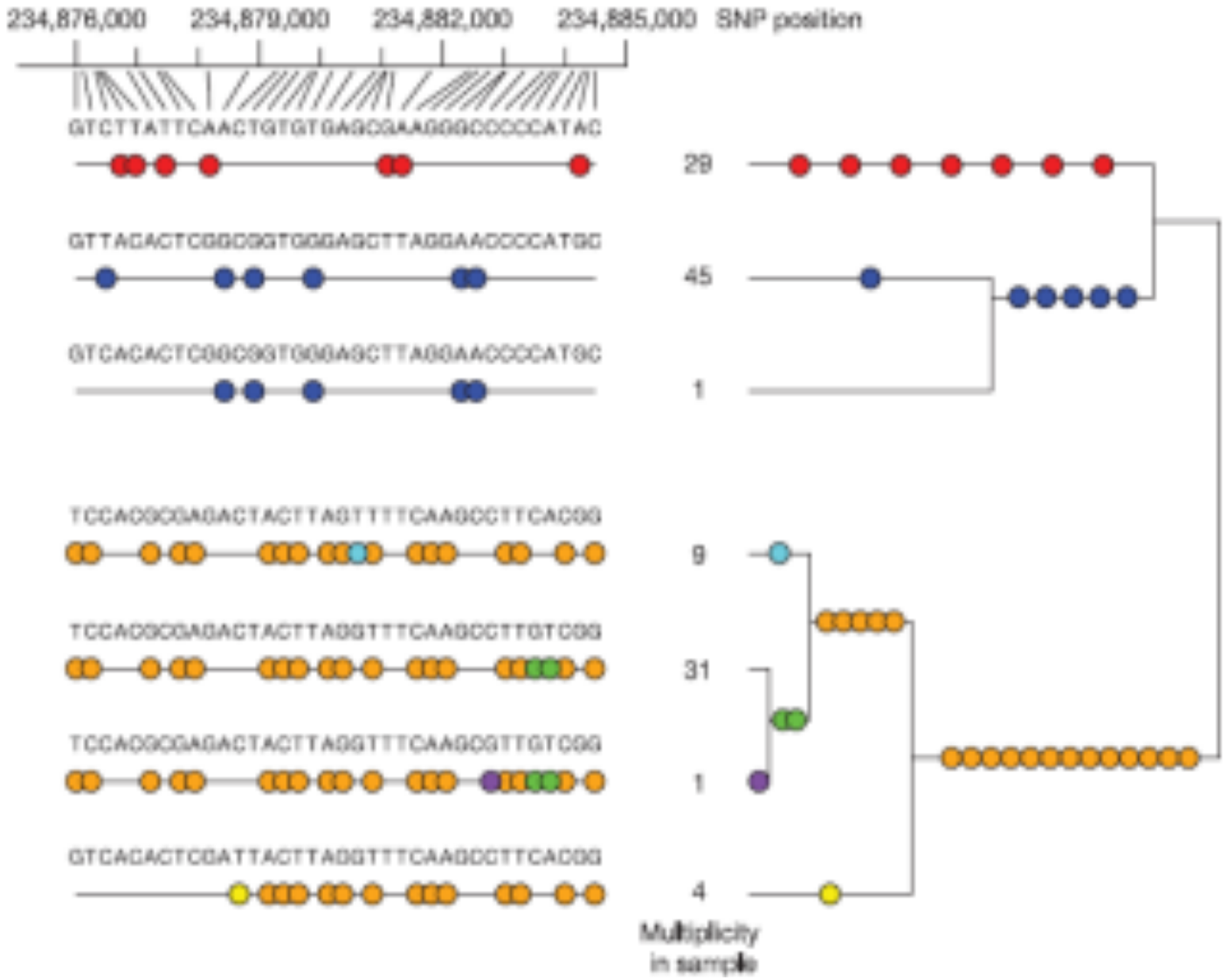
Limited number of haplotypes at 10kb loci

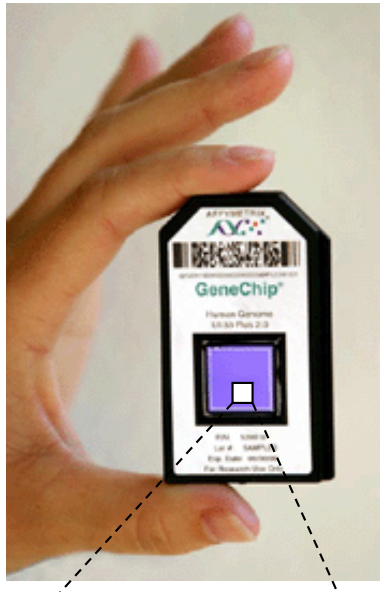


A haplotype map of the human genome

The International HapMap Consortium*

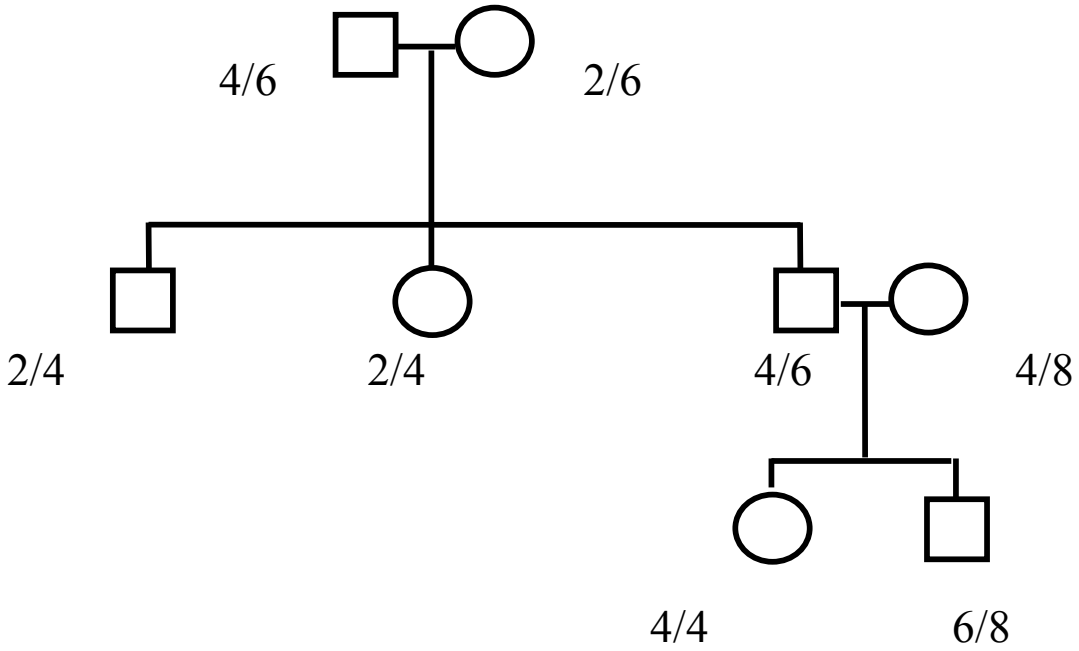
NATURE|Vol 437|27 October 2005



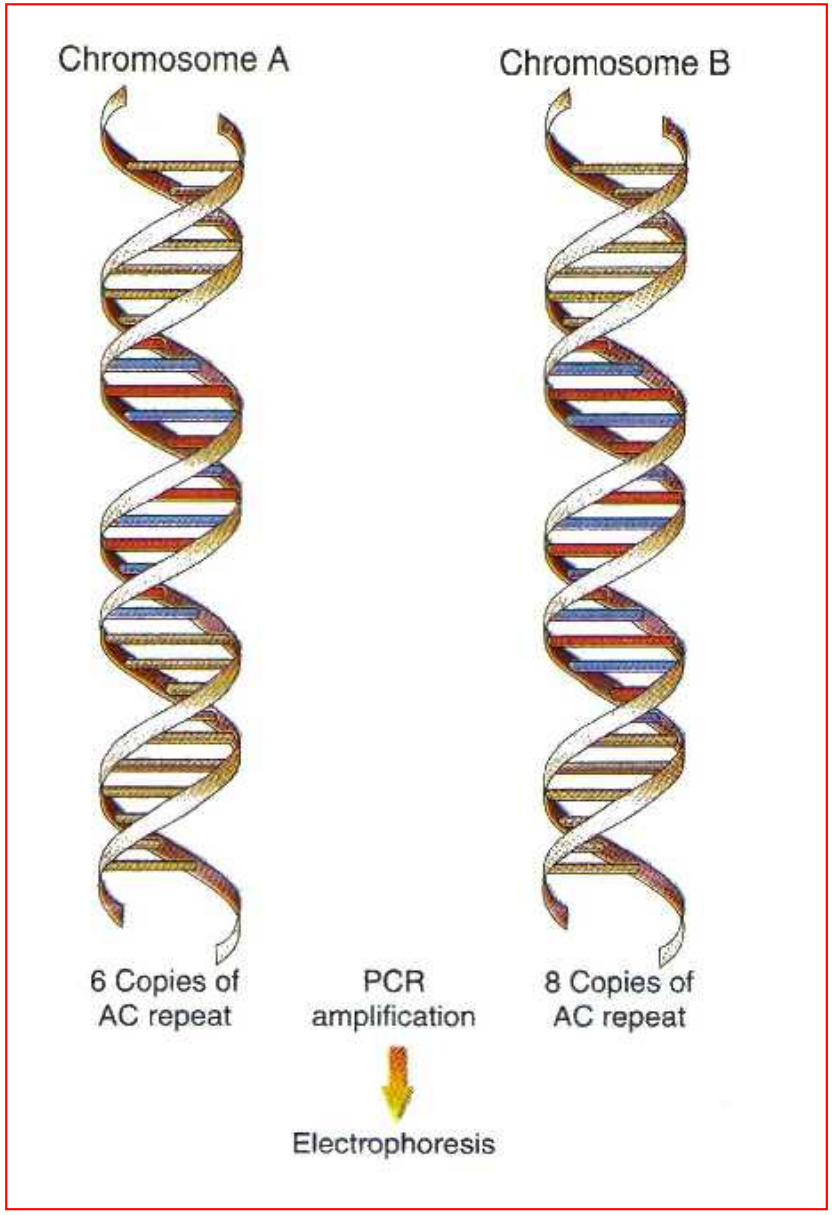


1,000,000 SNPs
DNA microarray chip

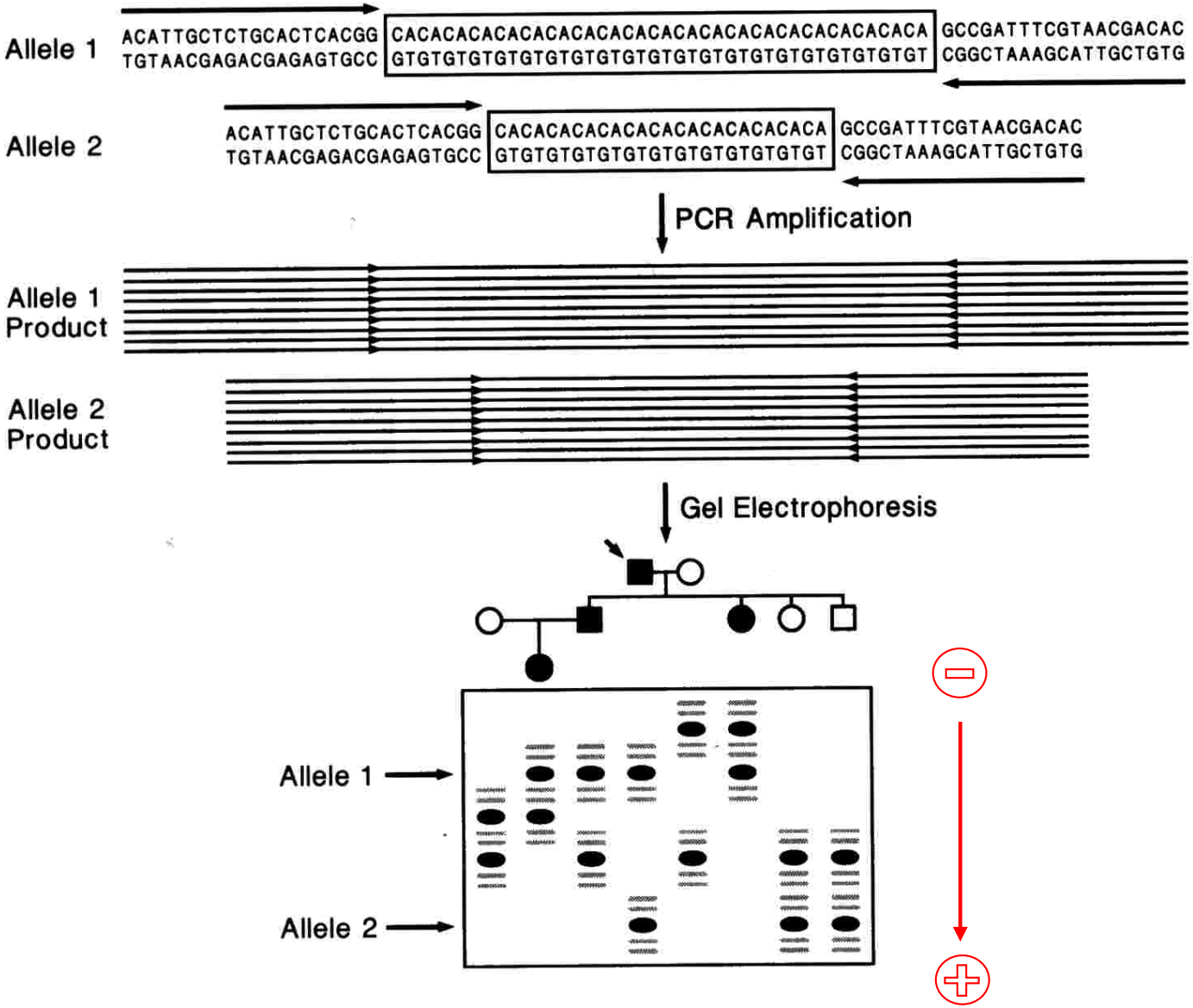
Microsatellite polymorphisms



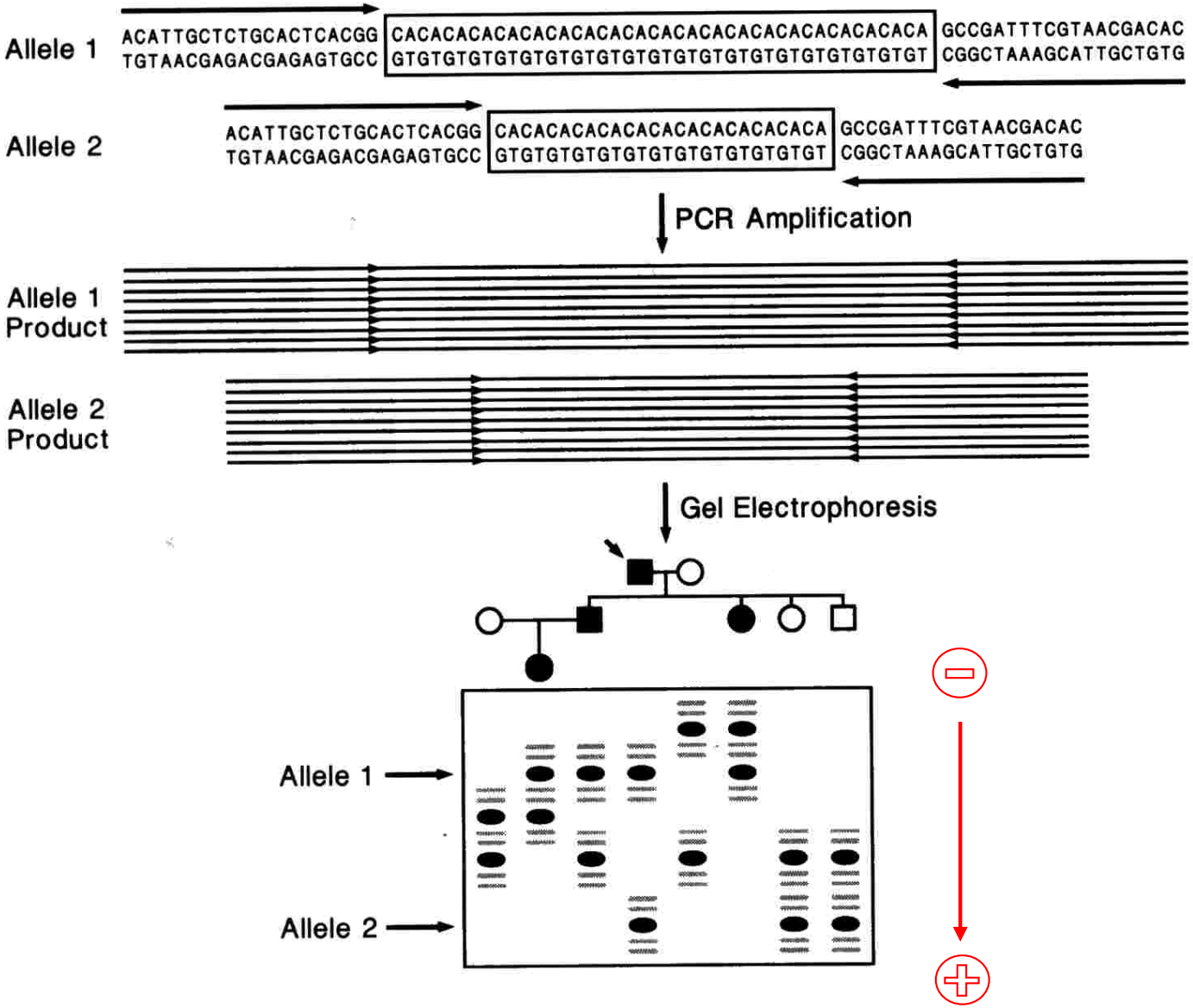
- One microsatellite locus
- Many alleles

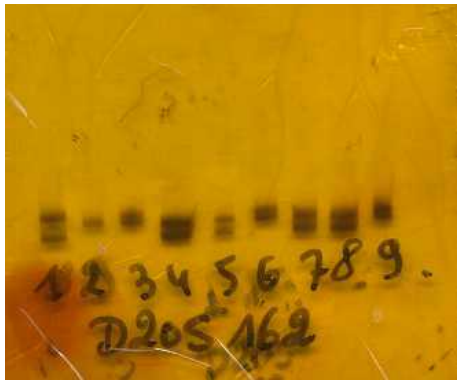


Genotyping a microsatellite

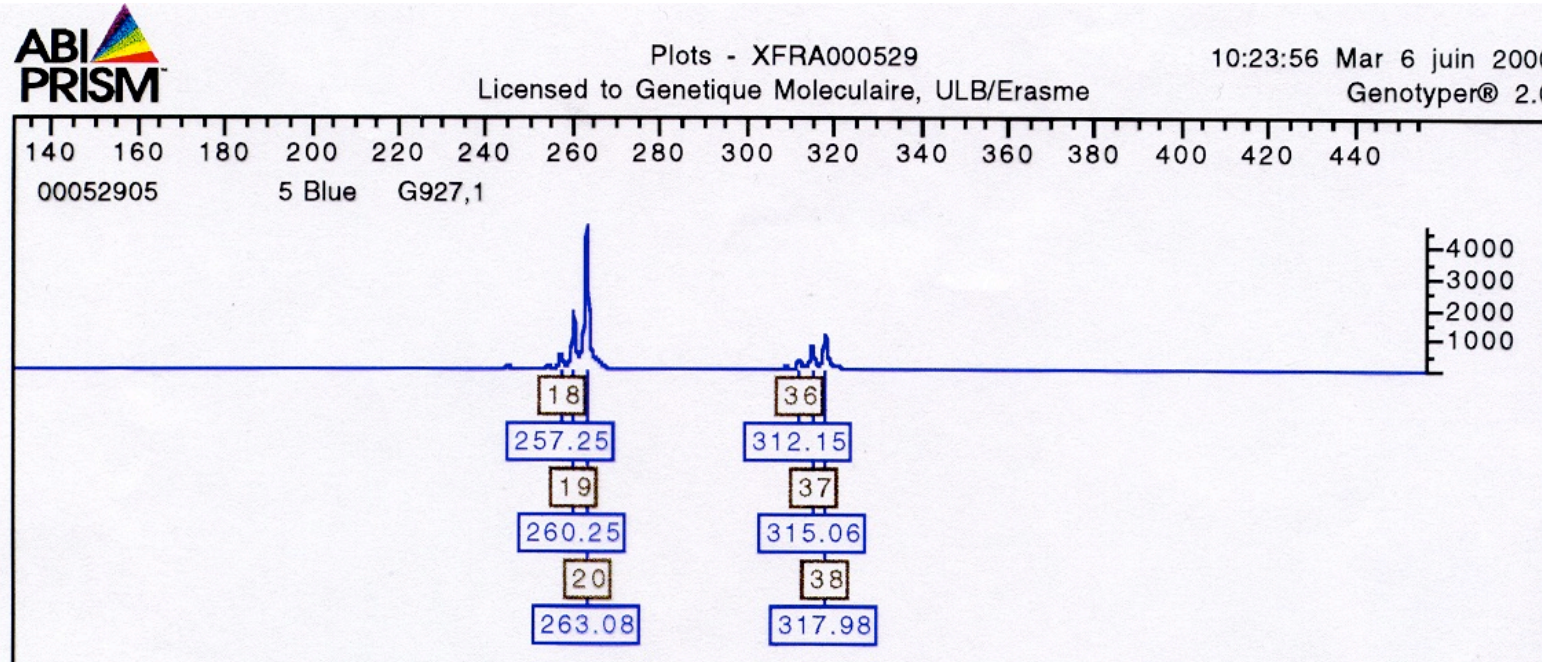


Genotyping a microsatellite



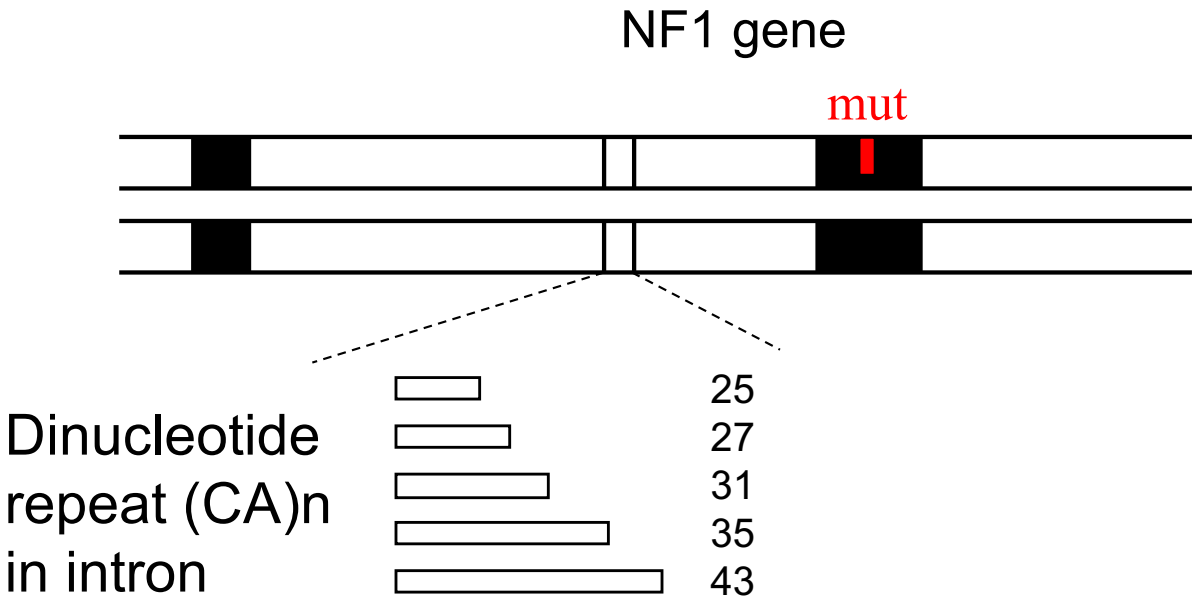


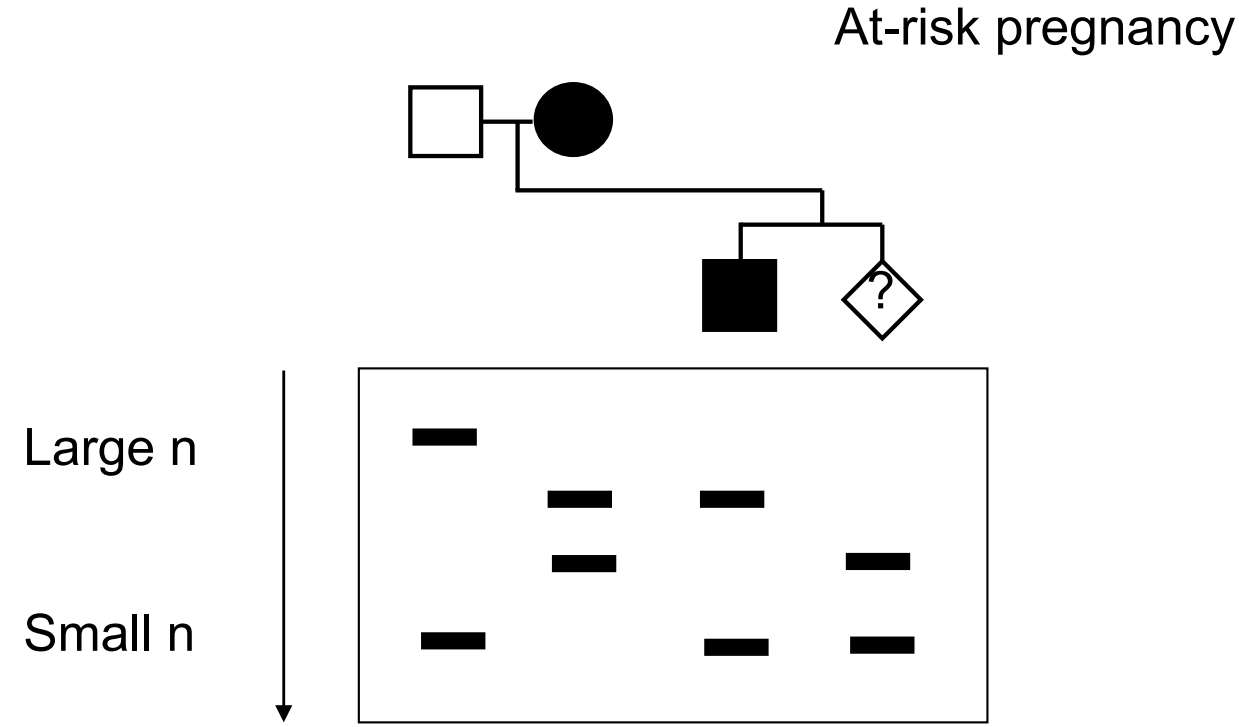
Capillary Electrophoresis, laser read-out



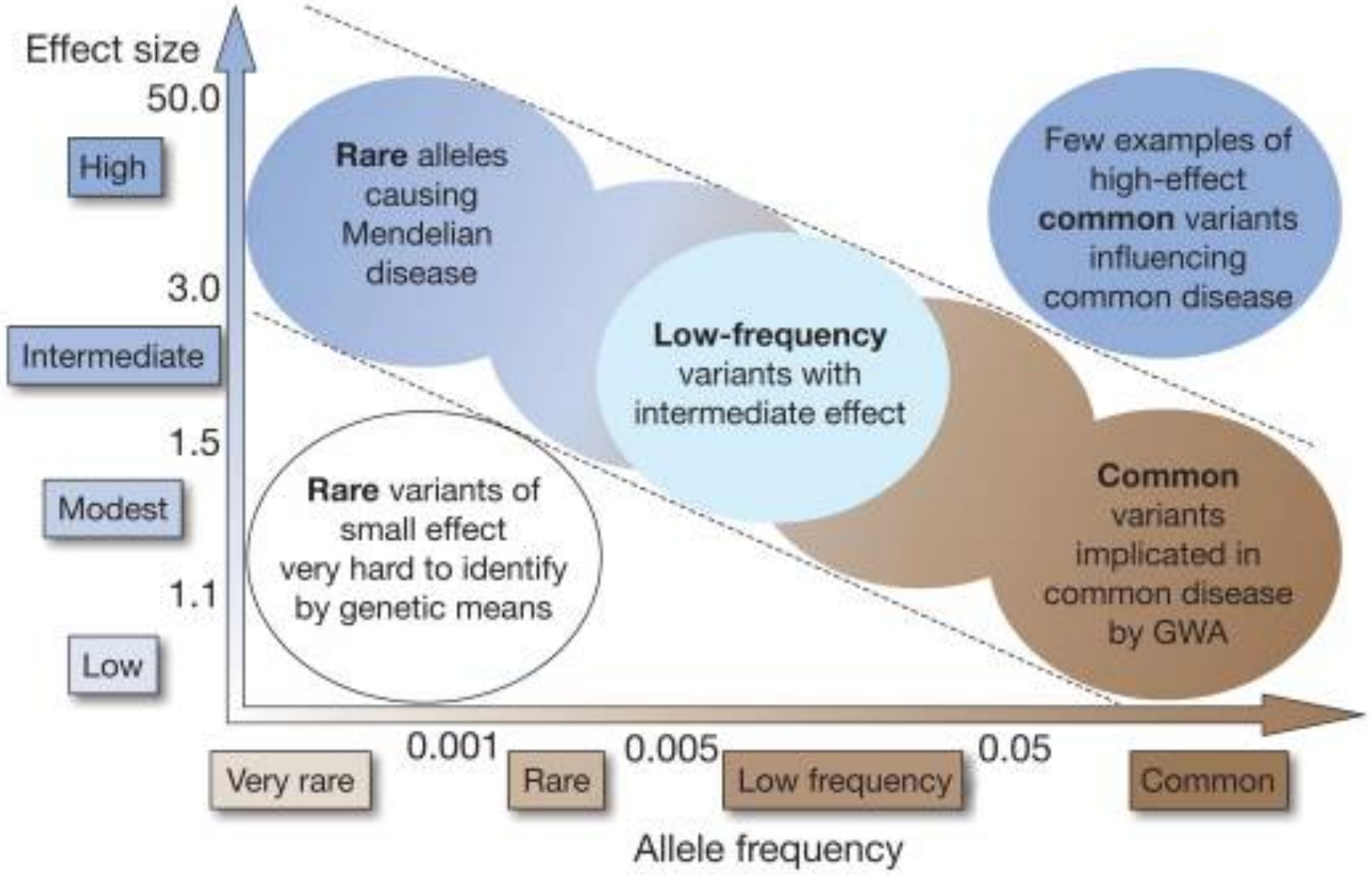
2 alleles of a microsatellite

Linkage study in family using a microsatellite marker



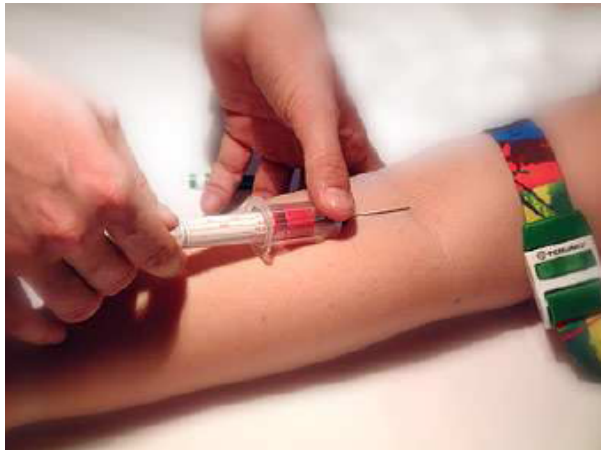


Rare genetic variants



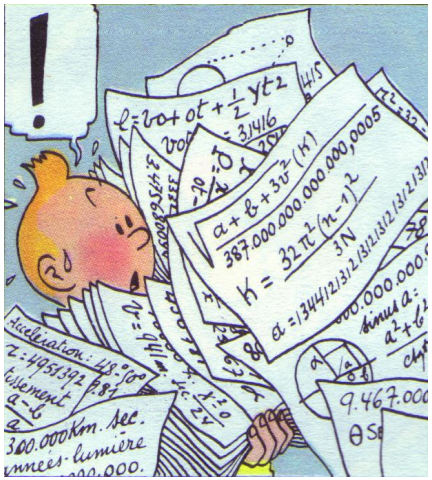
Manolio TA et al. Nature 2009: 8; 747-53

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

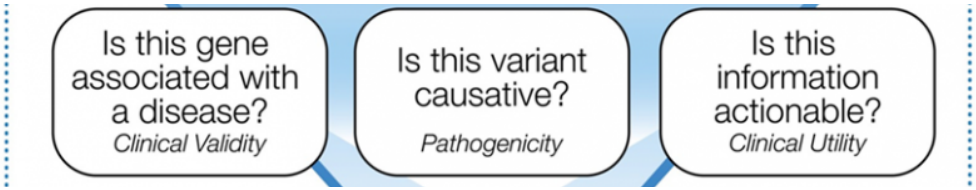
?

Intersting variant in Rotatin gene.

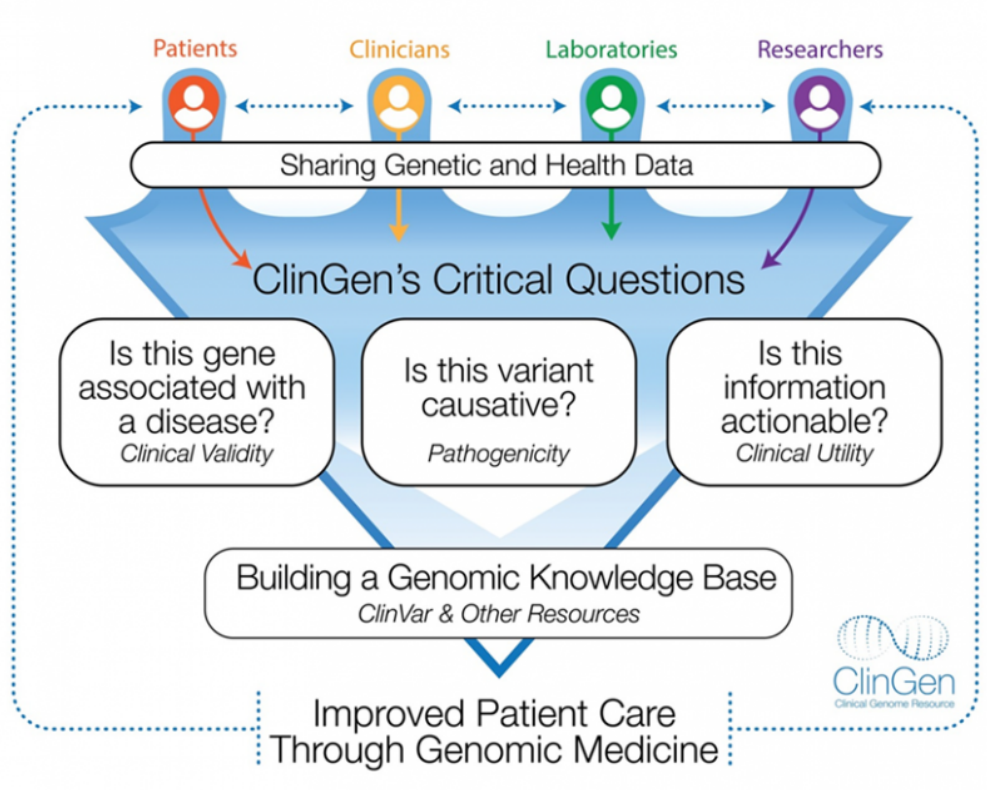
What if only one family affected?

Human population = saturation mutagenesis population?

Areas of uncertainty



Areas of uncertainty



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)

VUS

Areas of uncertainty :

1. Causes phenotype, or not involved ?

2. Complete or incomplete penetrance ?

VUS : how tell if pathogenic or benign ?

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

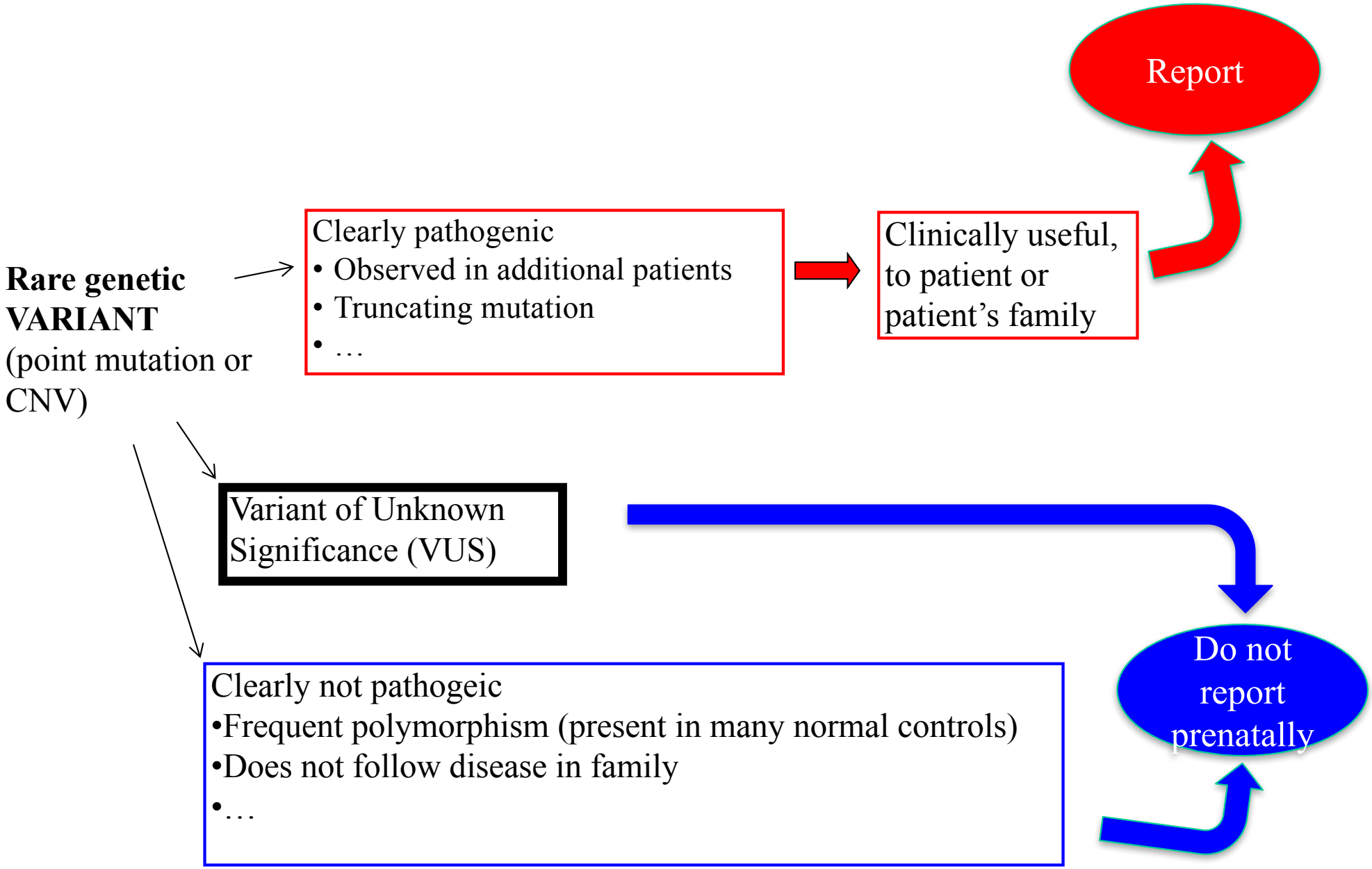
- Population data: test many controls
  - Family
  - Local controls
  - Regional
  - National
  - Worldwide

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 variants (from sequencing / from CGH arrays)



Findings out of scope of initial phenotype

Genome-wide analyses will show variants beyond initial question
= incidental findings

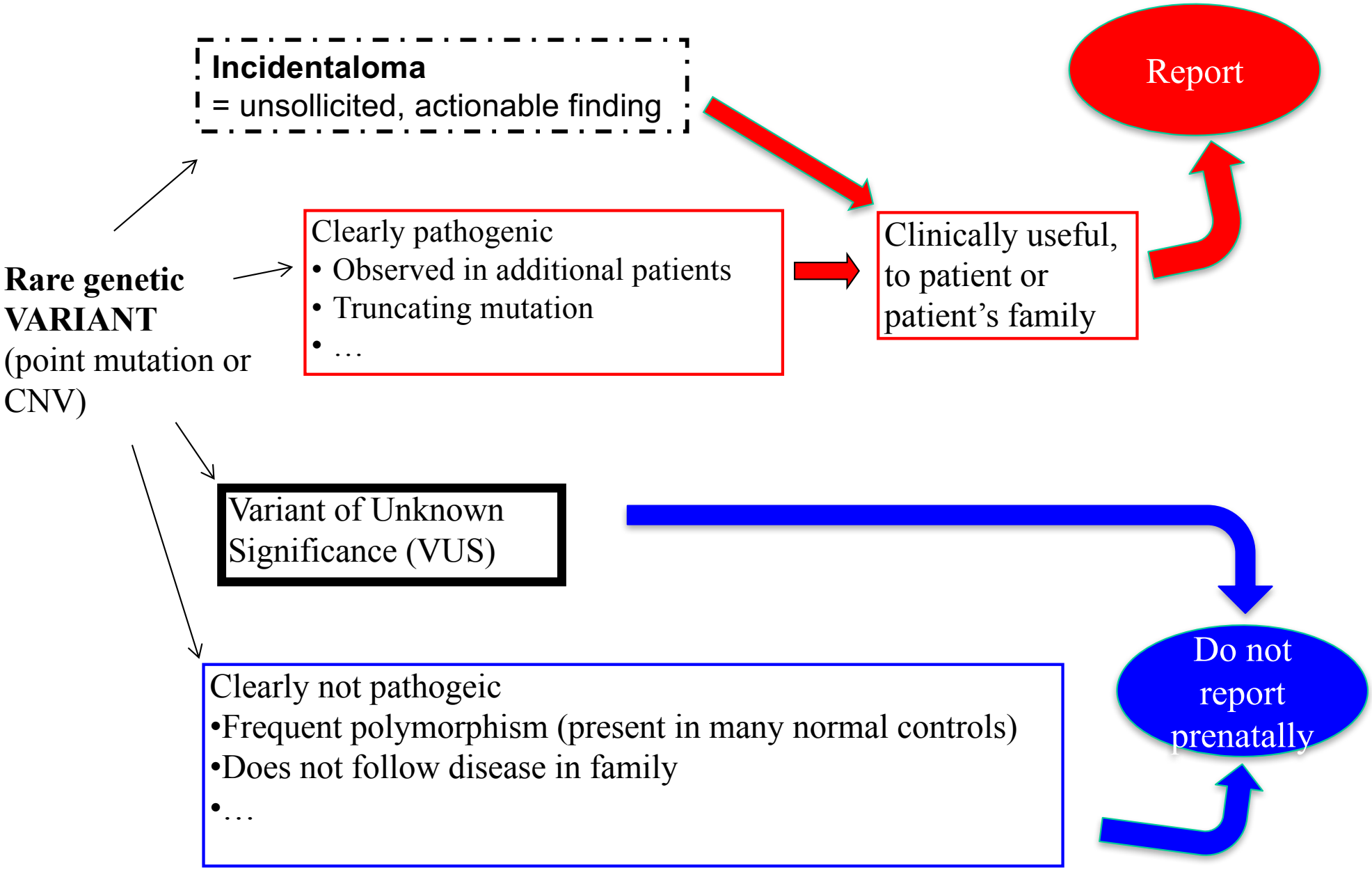
•Ex: child tested for ID => CGH array shows BRCA1 locus deletion (causes breast and ovarian cancers in adults)

•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

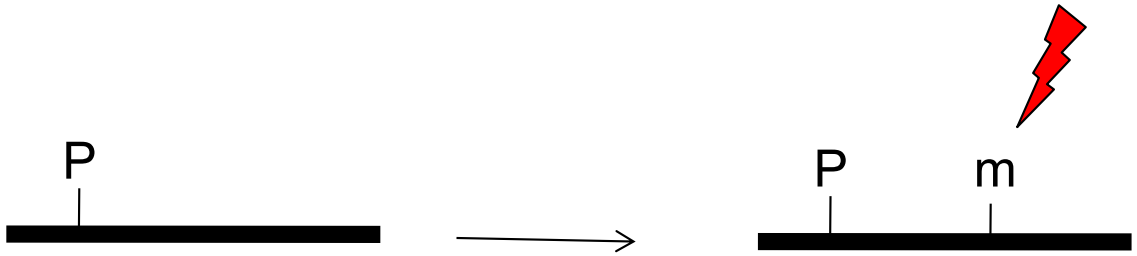
- Unsollicited finding
- Actionable or not
- If actionable, inform patient and offer genetic counseling (patient and family)
- Opt-out procedure (discuss in pre-test genetic counseling)

Secondary variant

- Actionable change
- In predefined, international consensus set of genes (~150 genes in 2015)
- 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

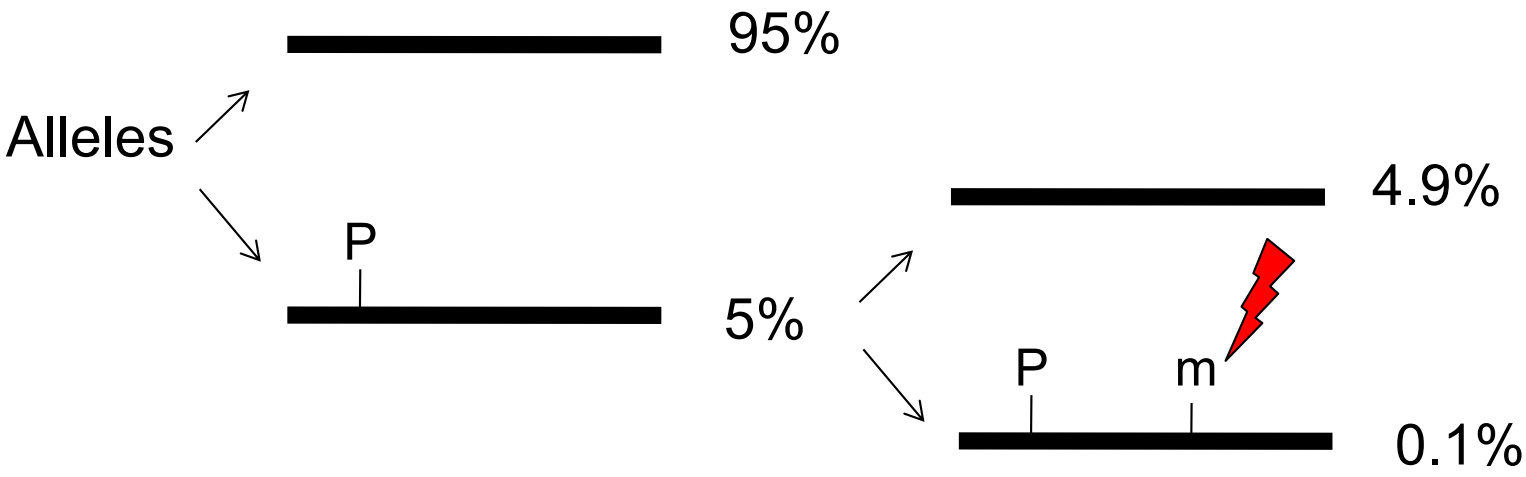
COMPLEX ALLELES

Polymorphism and mutation may coexist on same allele



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

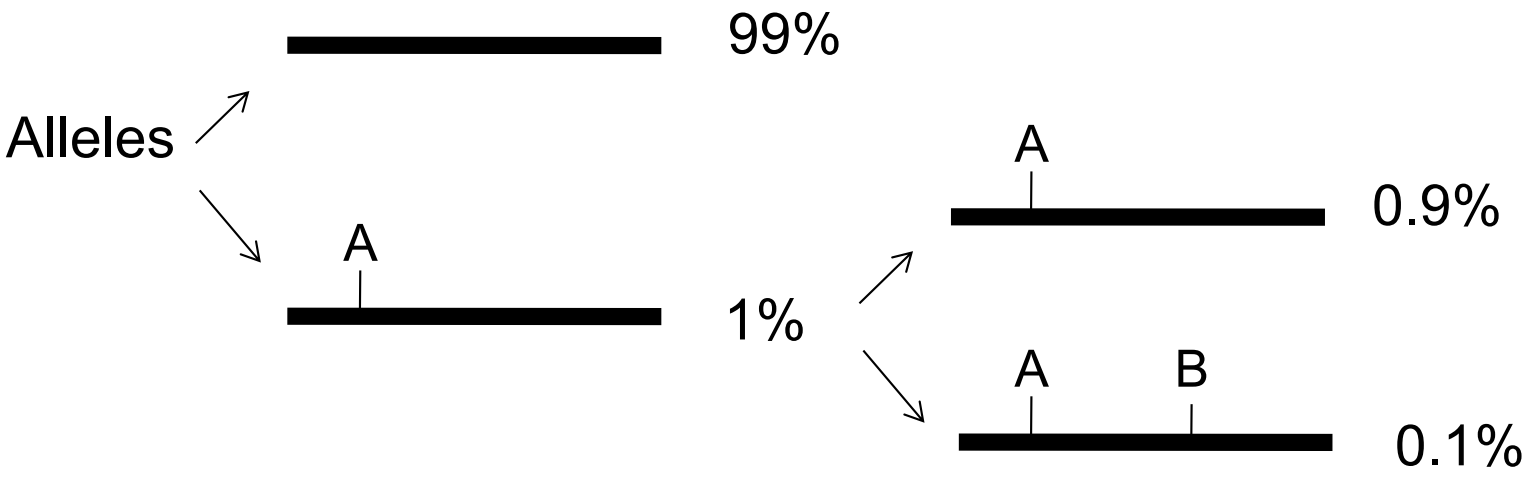
Polymorphism and mutation may coexist on same allele



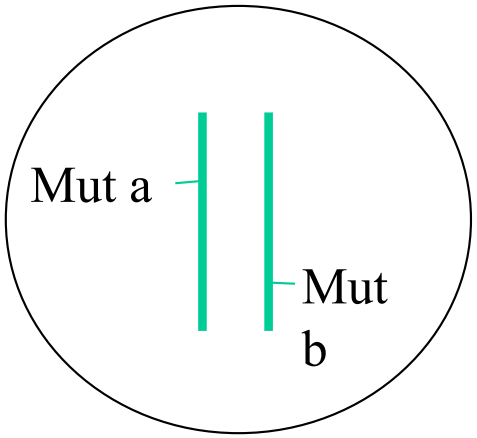
- Mutated alleles are rare : 0.1% in this example
- P is known and frequent, hence no problem in interpretating m as a possible disease causing mutation.

- If P was rare, it might be hard to tell which of the 2 rare variants, P and m, are disease-causins: « complex allele »

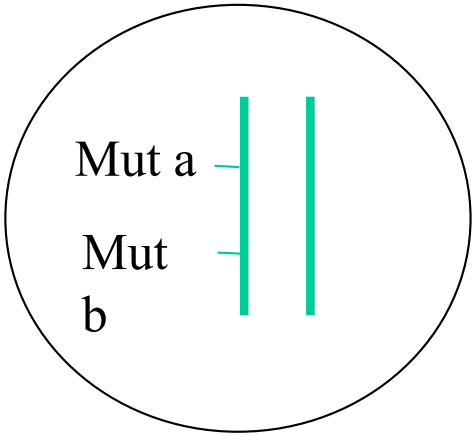
Complex alleles with 2 rare variants



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



Mut a and Mut b
< both parents

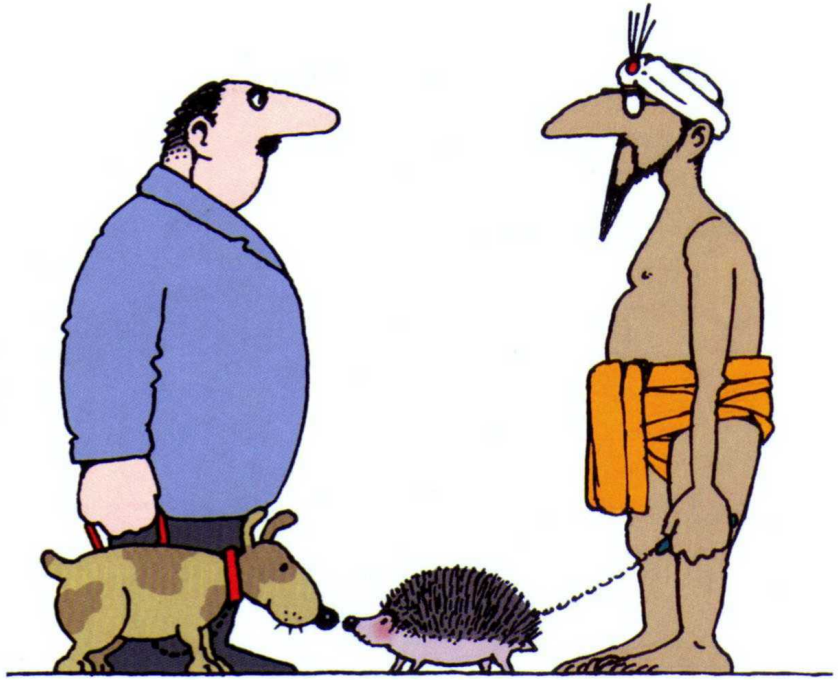


Mut a and Mut b
< same parent

- 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) !

Ex: complex allele of the CFTR gene

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)

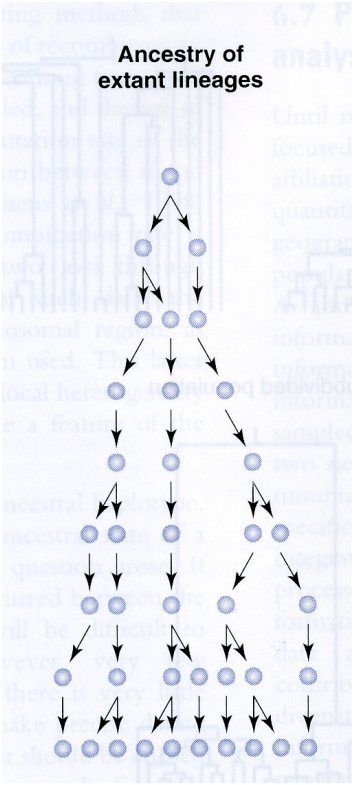
- BUT : There are no races in the human species

Ethnic prevalence of ancestral mutations

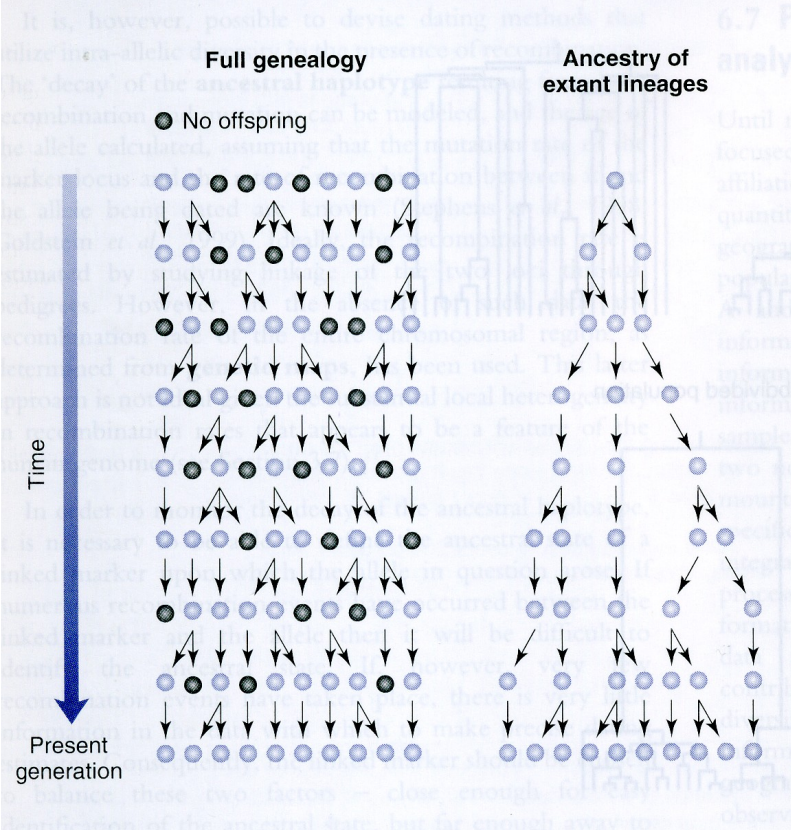
- Race = group of individuals defined by common biological characteristics.
  No race in human species.
  Human groups mix and depart constantly.

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

Most Recent Common Ancestor

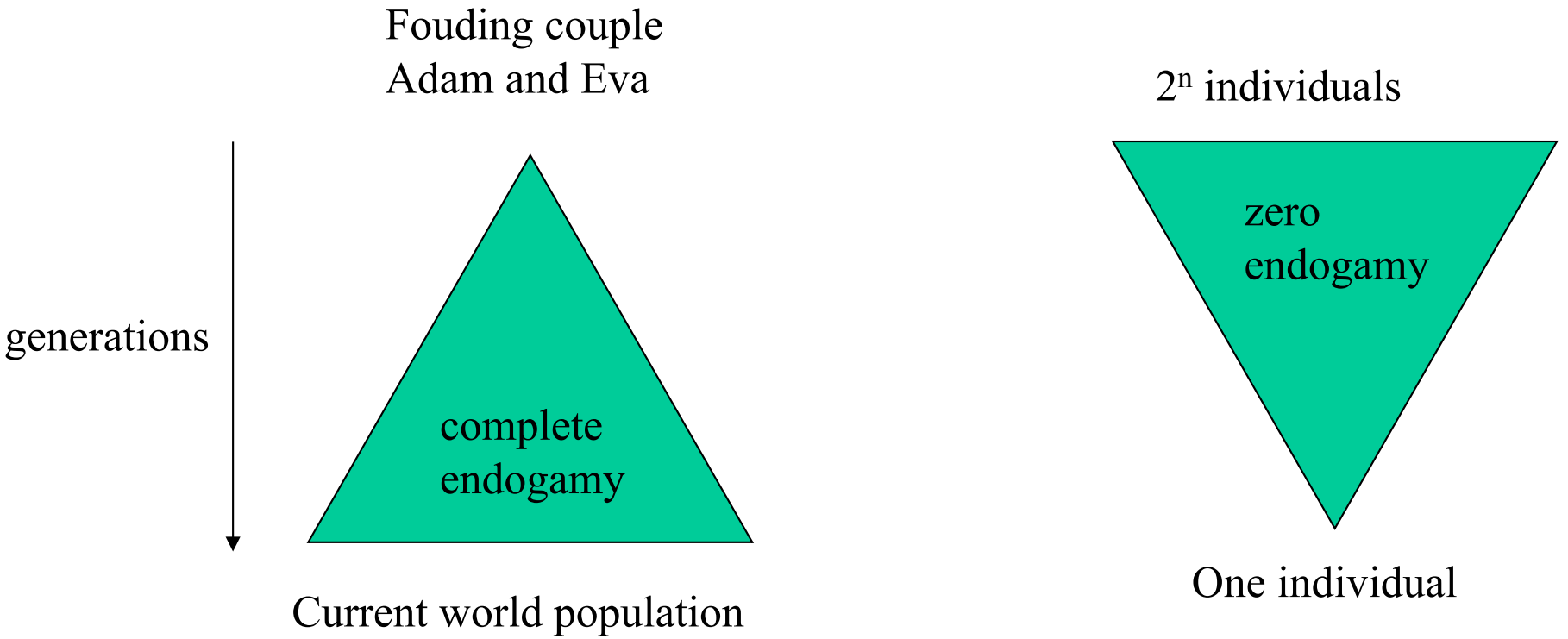


Most Recent Common Ancestor



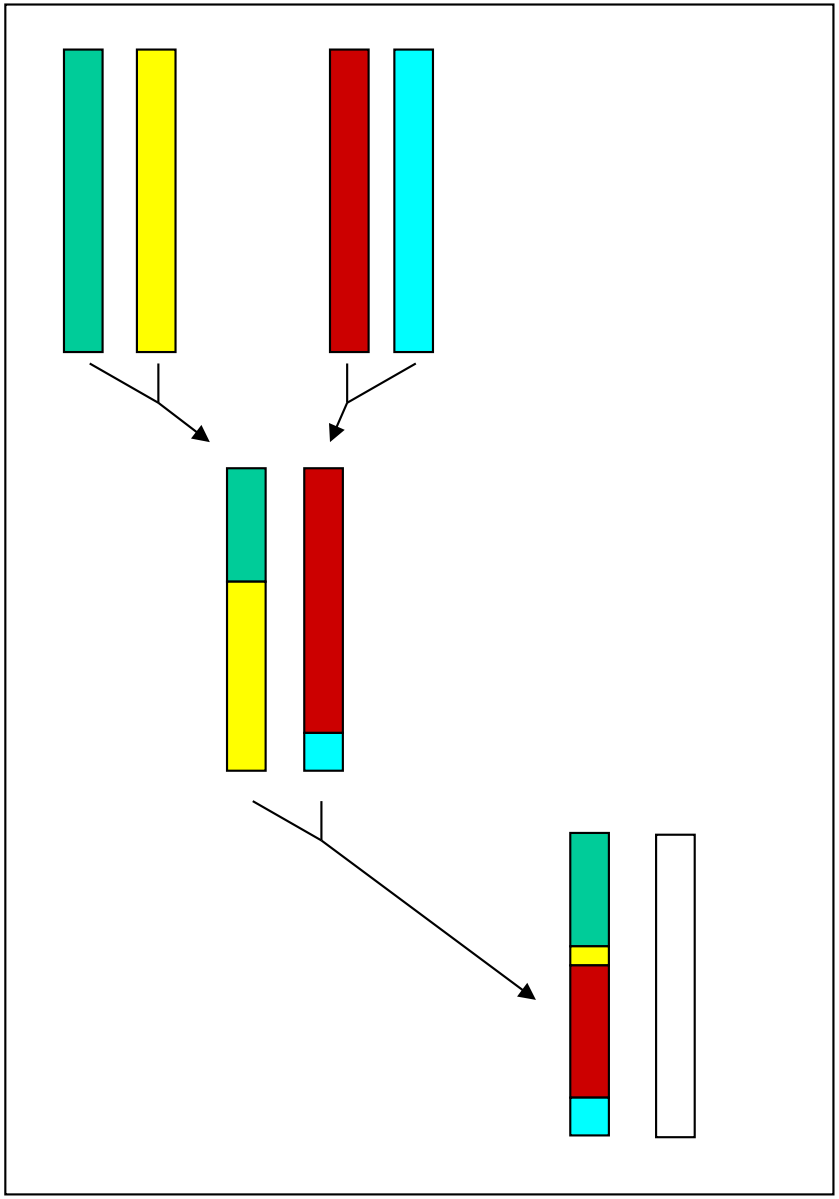
Population of constant size over generations

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

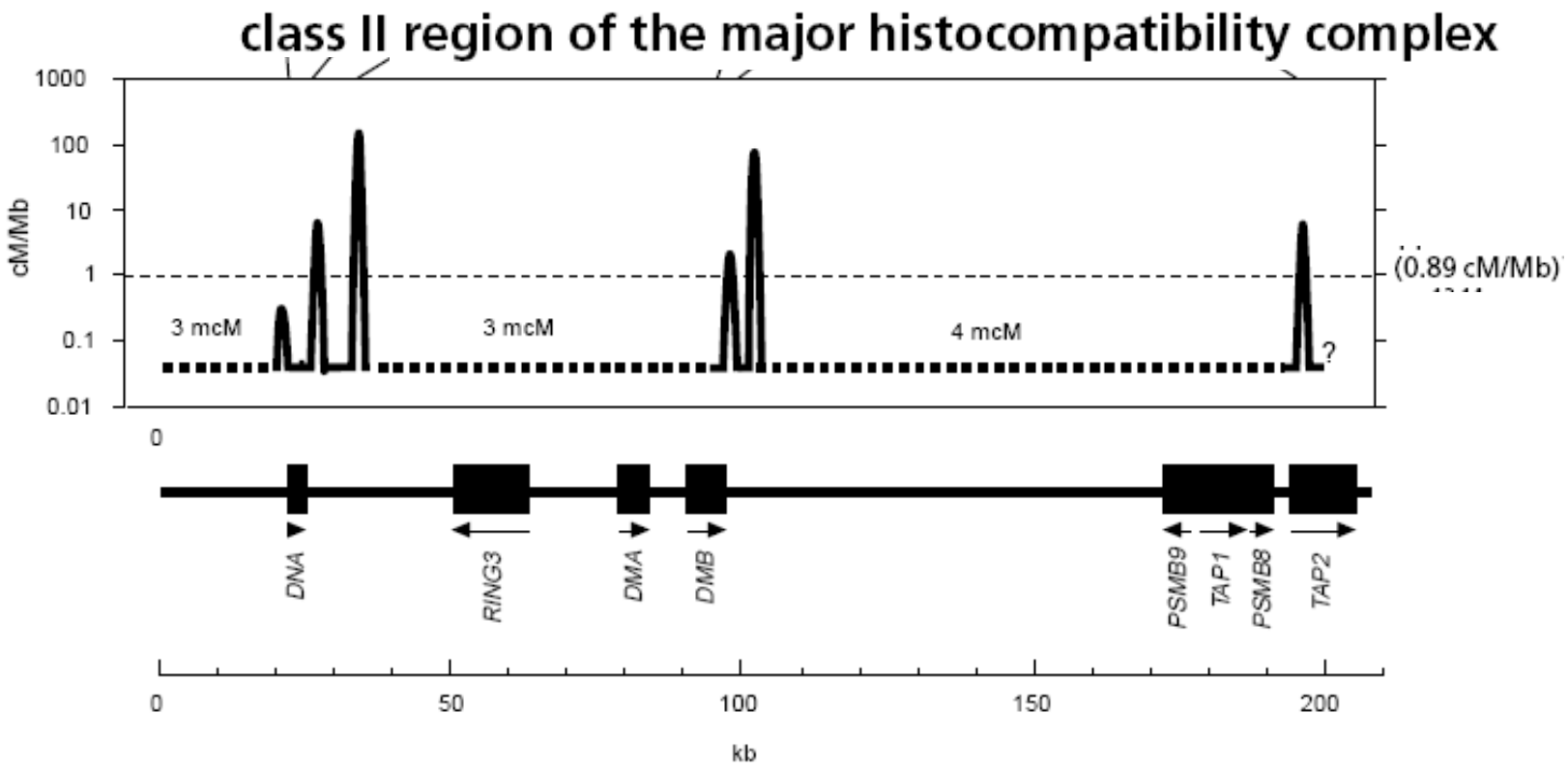
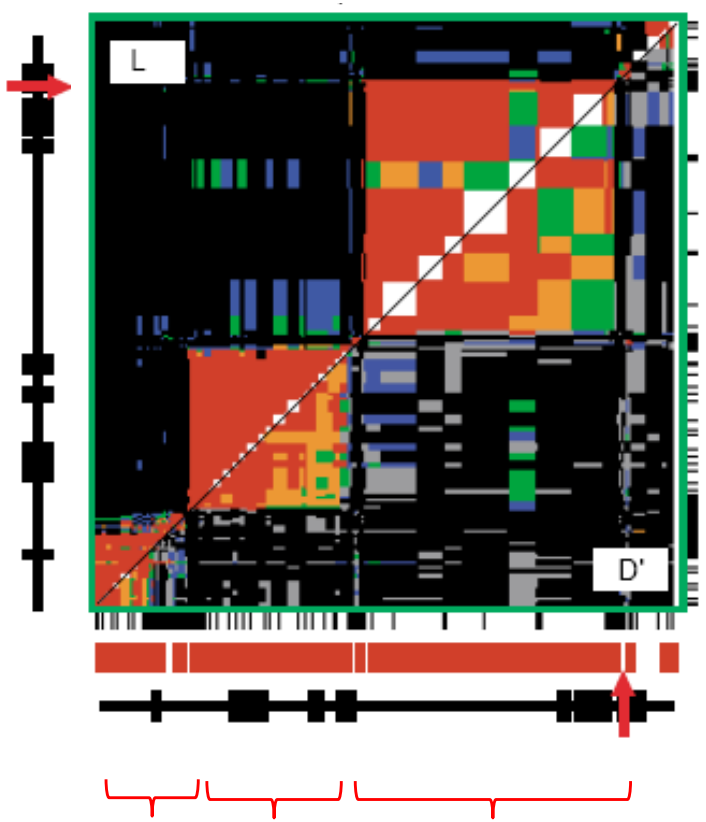


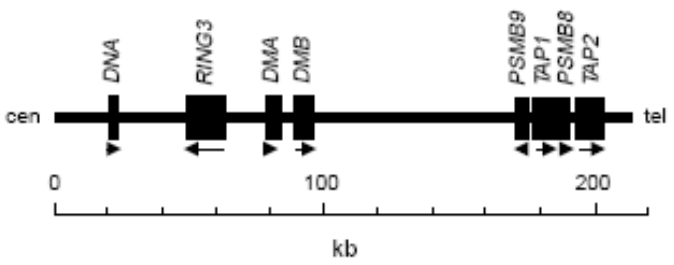
Fig. 5 Sperm crossover activity in the class II region of the MHC. The number of men tested and the total number of sperm crossovers mapped are given for each hot spot, together with approximate hot-spot center coordinate in the consensus sequence of the human MHC[10]. The width of each hot spot, within 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].

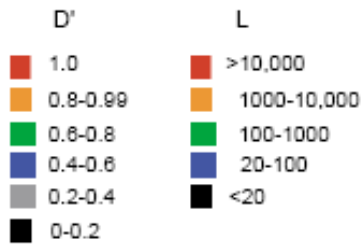
Jeffreys et al. **nature genetics • volume 29 • october 2001**

Intensely punctate meiotic recombination In MHC

Haplotype blocks





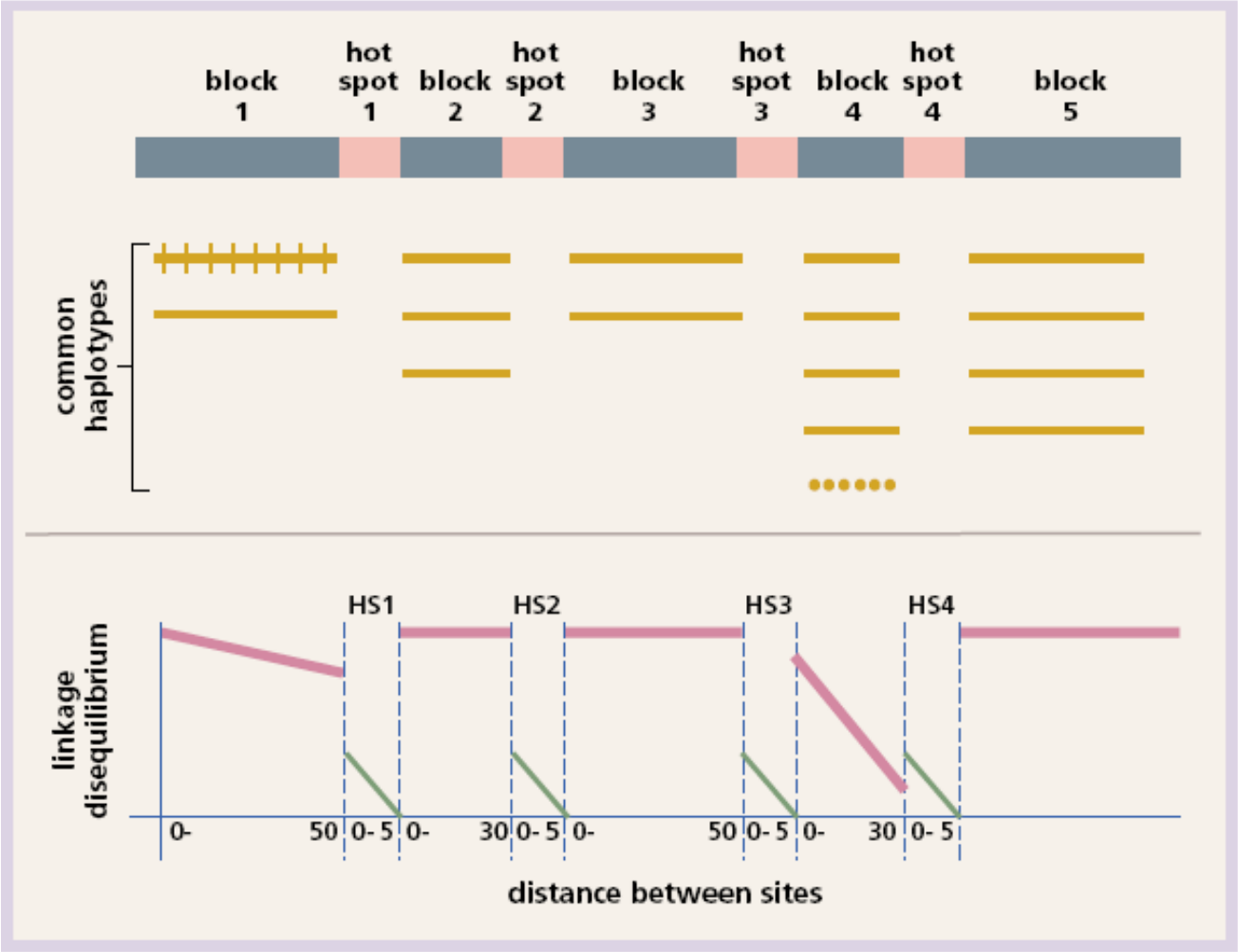


D' and L are measures of LD

Crossover hotspot in *TAP2* gene (known)

Jeffreys et al. 2001

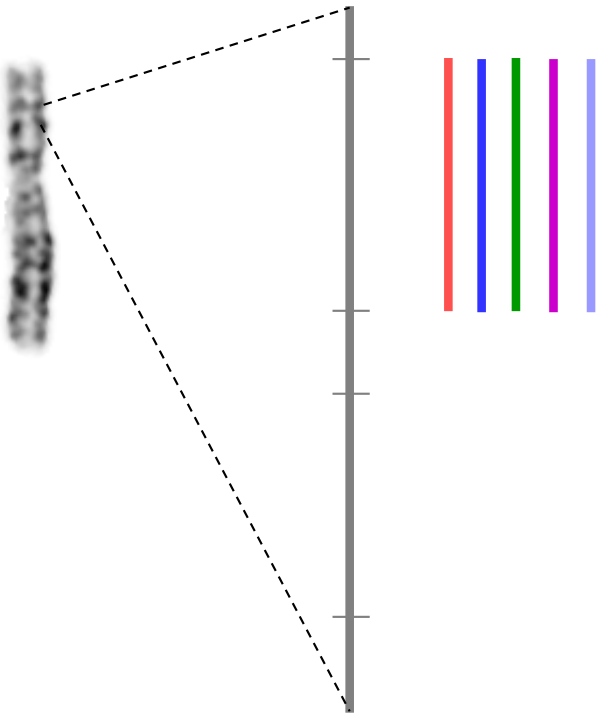
Haplotype block <=> absolute LD



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.

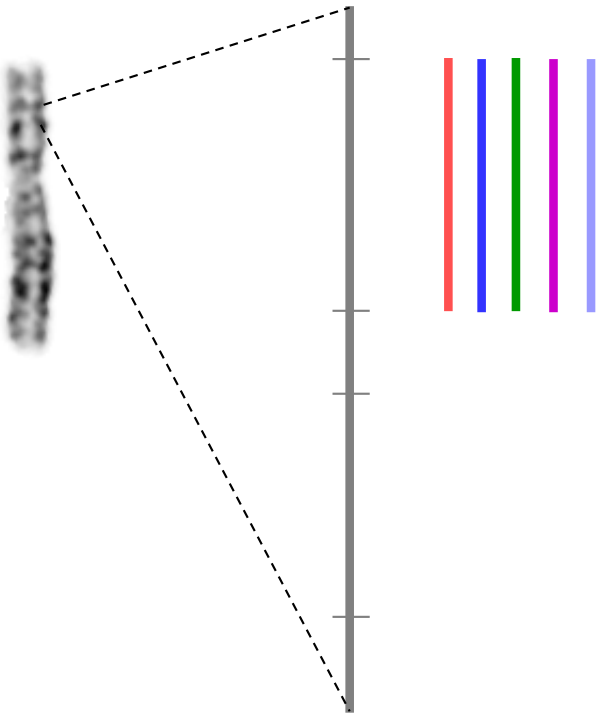
Islands of Linkage Disequilibrium. Goldstein DB 2001 Nat Genet

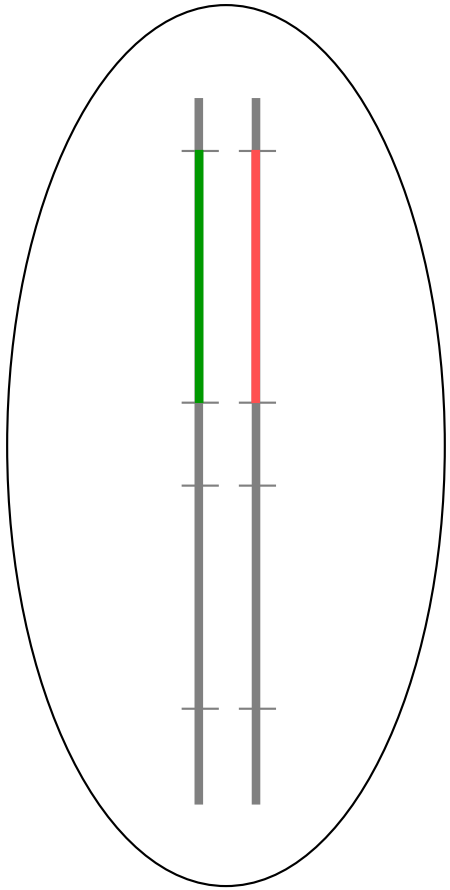
Haplotype blocks



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

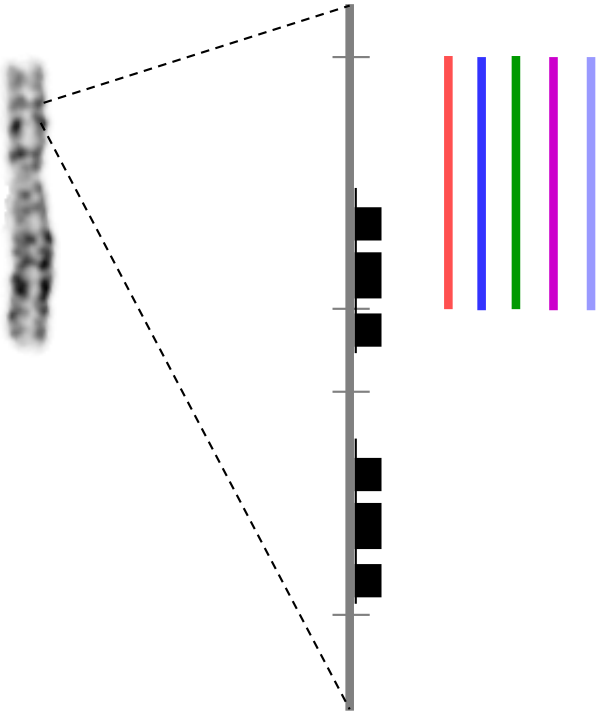
Haplotype blocks

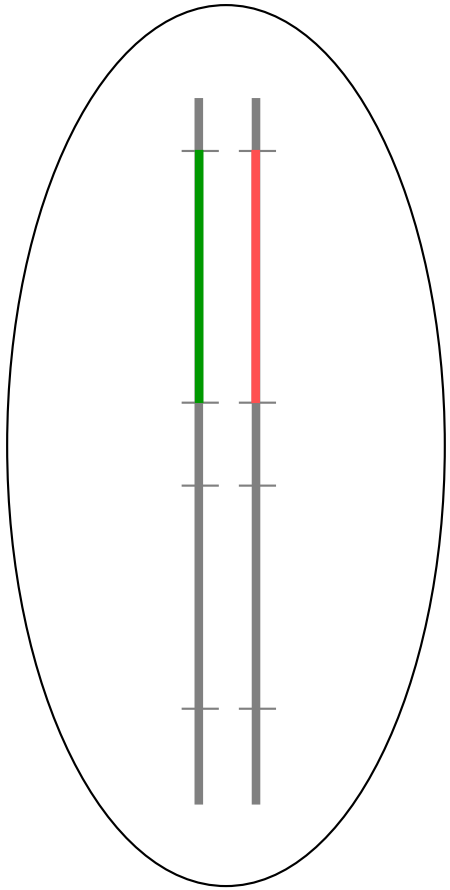




- Whole genome = 100,000 blocks, with a few haplotypes in population.
- Everyone has 100,000 x 2 haplotypes

Haplotype blocks

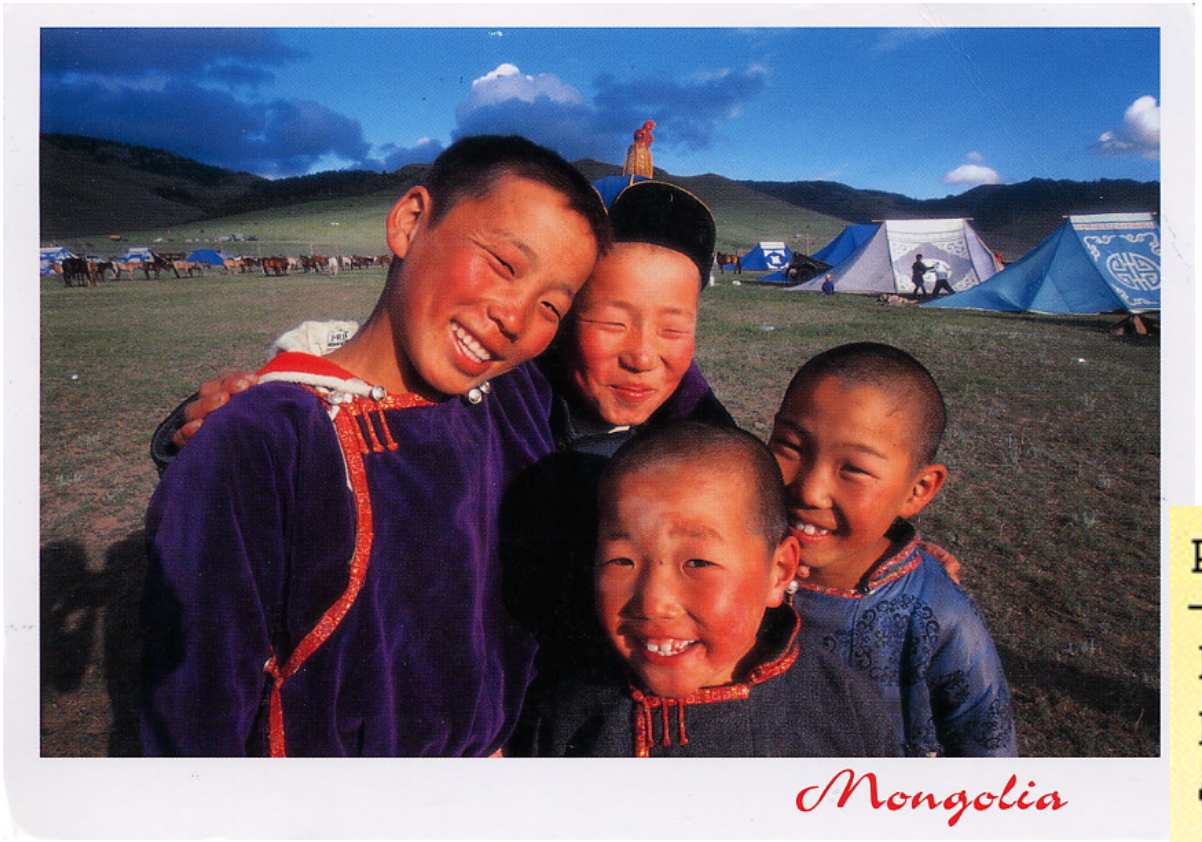




- Whole genome = 100,000 blocks, with a few haplotypes in population.
- Everyone has 100,000 x 2 haplotypes
- Genes may overlap blocks

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

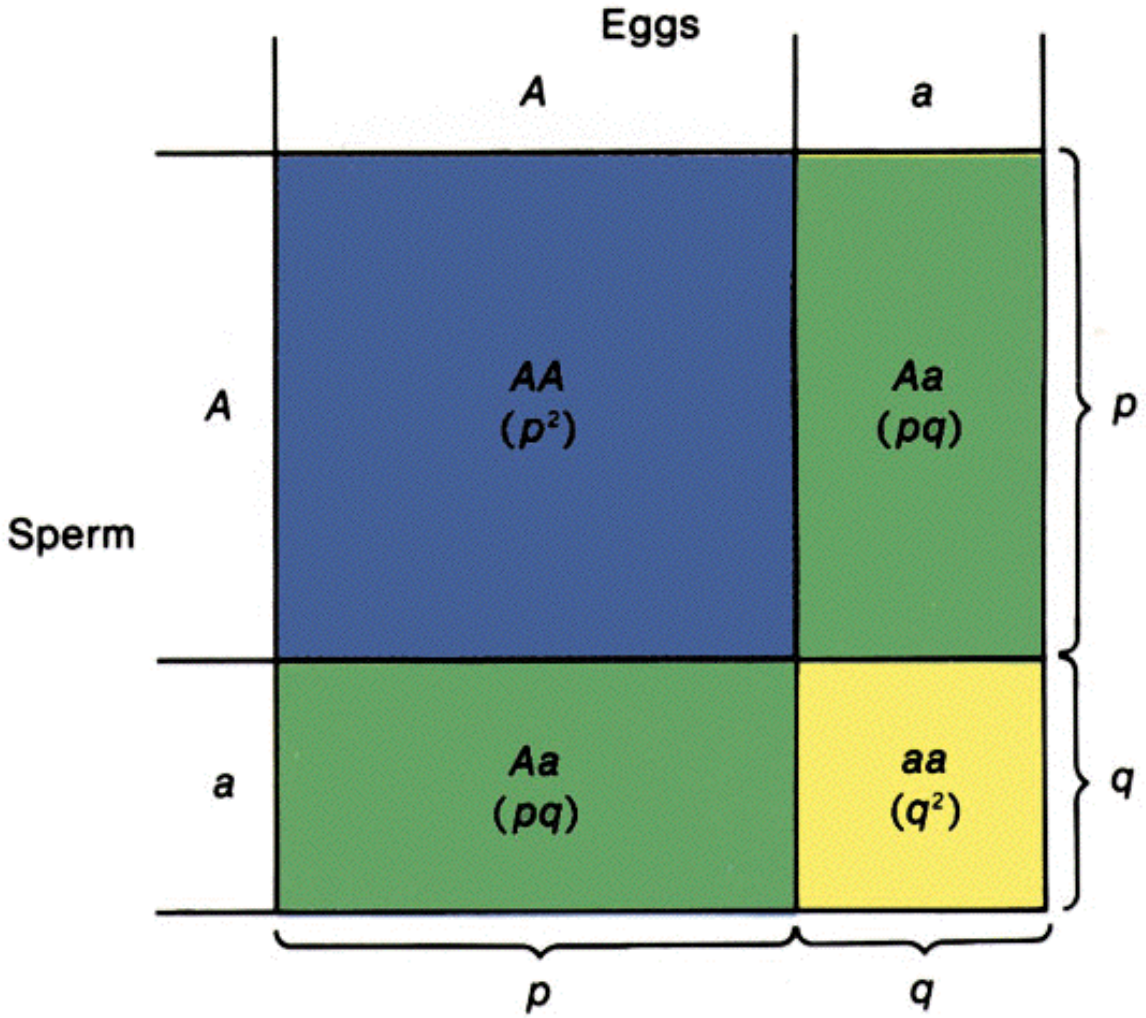
- 8000 individuals are AA
- 2000 individuals are aa

=> is this population in HW equilibrium?

! Phenotypes are not considered here !

A is not dominant and a is not recessive

Punnett square for population (vs family)



Equilibrium reached in 1 generation

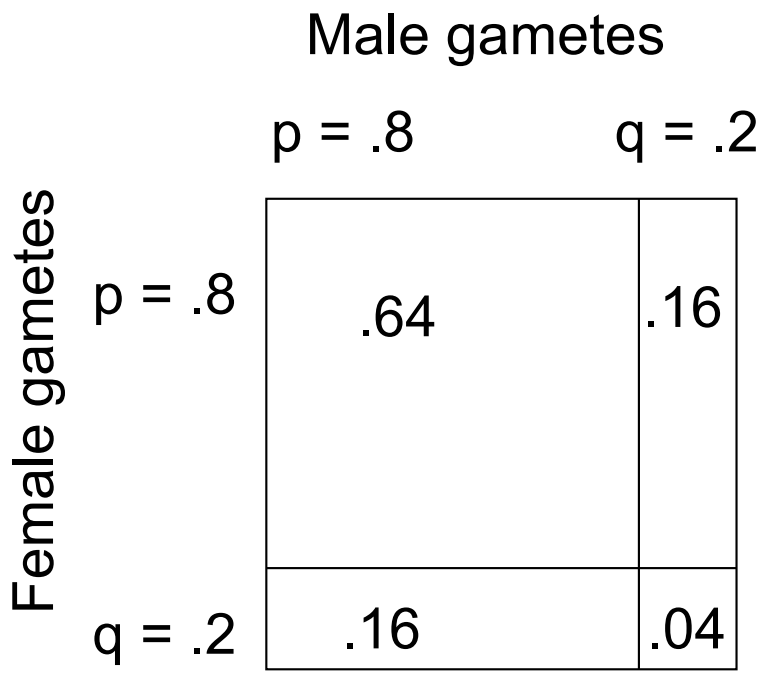
- N=10000; 8000 are AA, 2000 are aa, 0 are htz : NOT in equlibrium
- Alleles: 16000 A, 4000 a (p=16000/20000=0.8, q=0.2)

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*



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

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

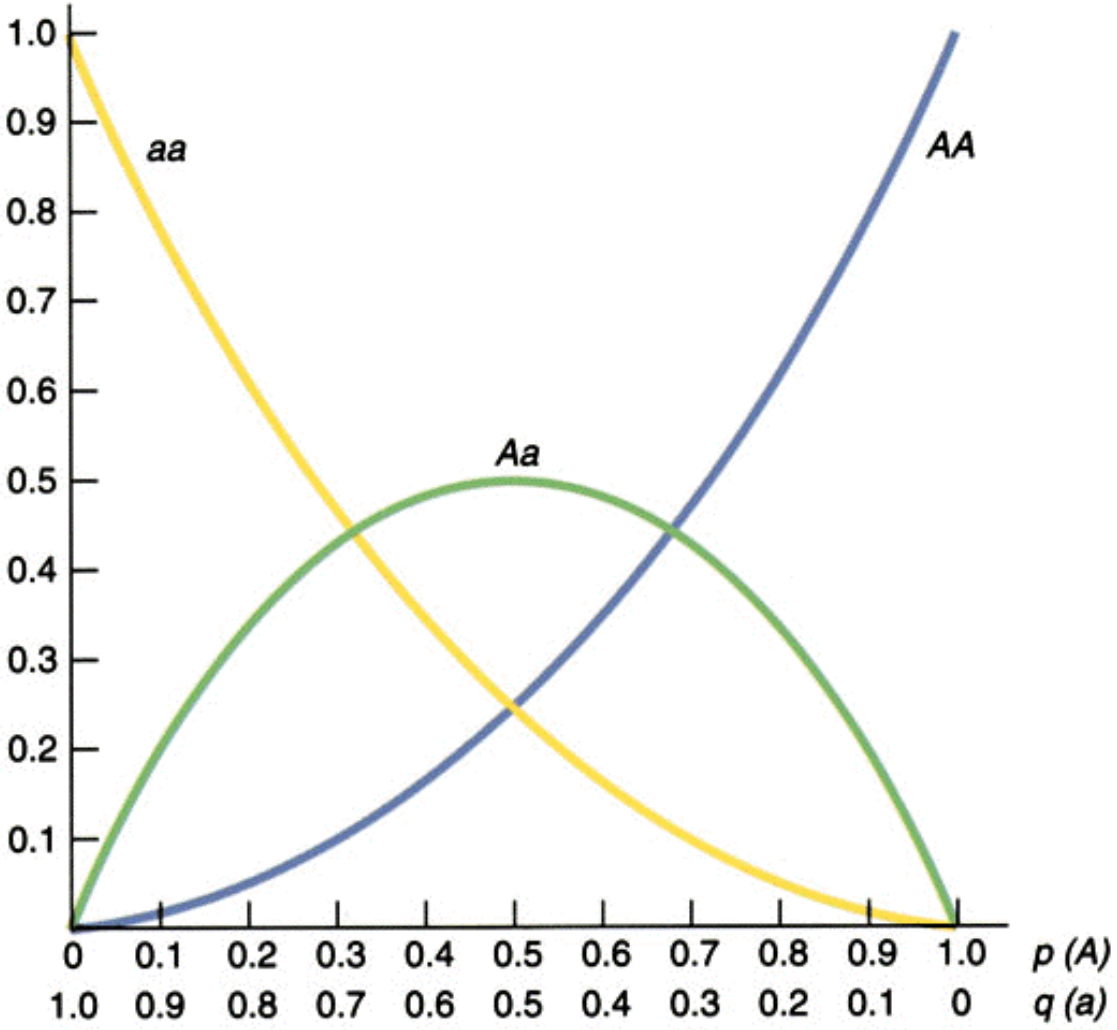
Equilibrium reached in 1 generation

- 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**
- Next generation: Punnett square, N stable
  => 6400 AA, 3200 Aa, 400 aa

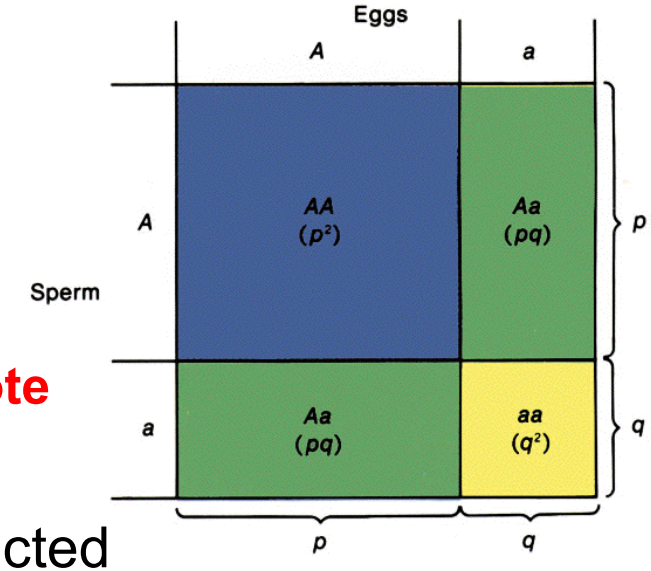
Male gametes

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

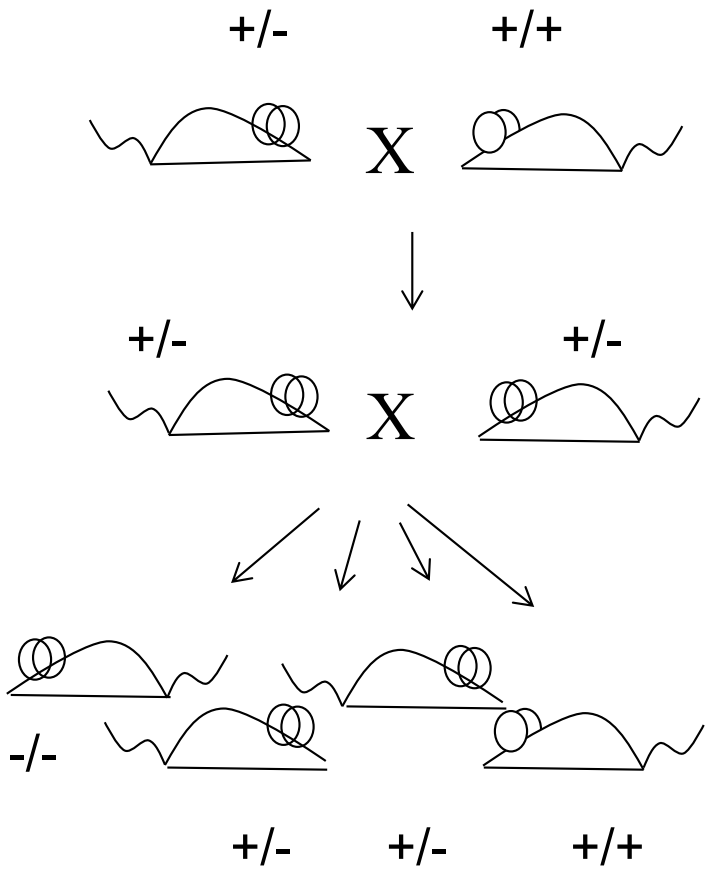
ΔCCR5: 2 x 7 + 134 = 148 => 148 / 1576 = 0.094

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

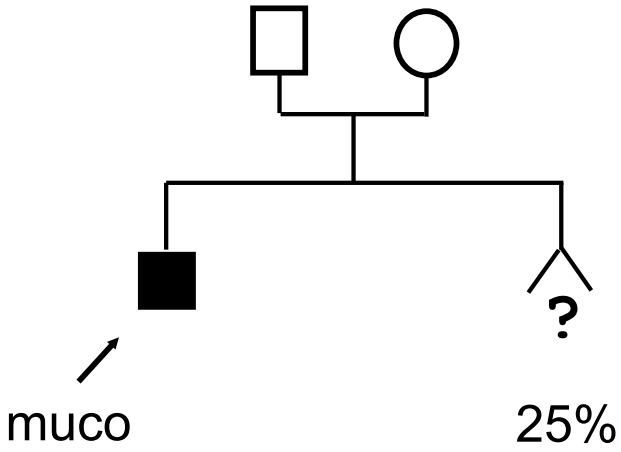


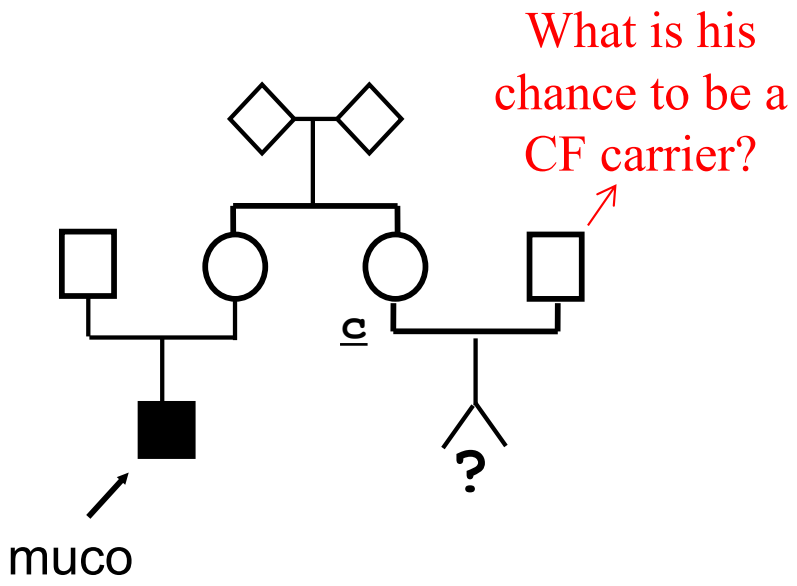
The observed distribution is not expected, a load of -/- homozygotes are missing => suggests embryonic lethality in homozygous -/- KO.

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

Cystic fibrosis affects 1 newborn in 2500

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



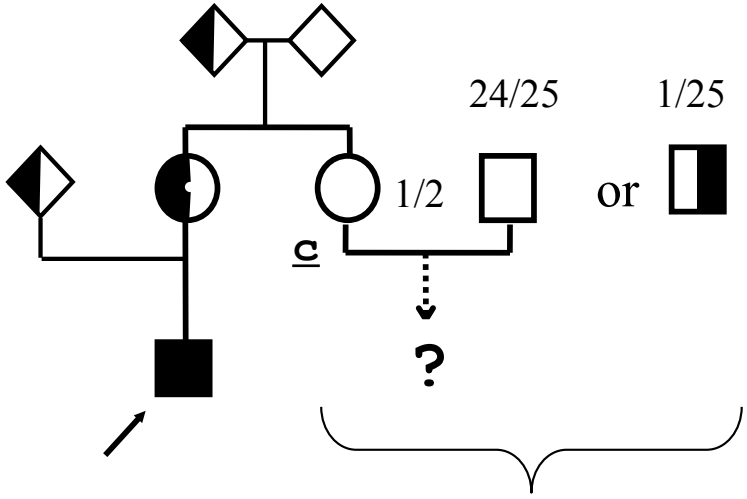


HW in AR disease: eg CF

- 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)



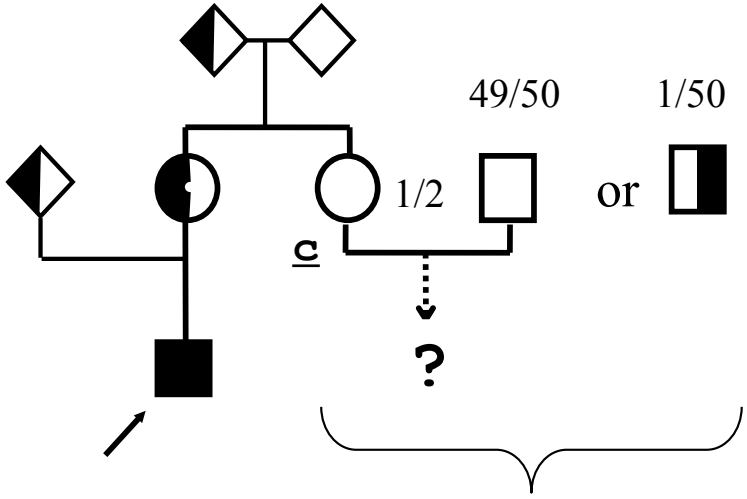
Offspring risk
(a priori) = 1/200

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

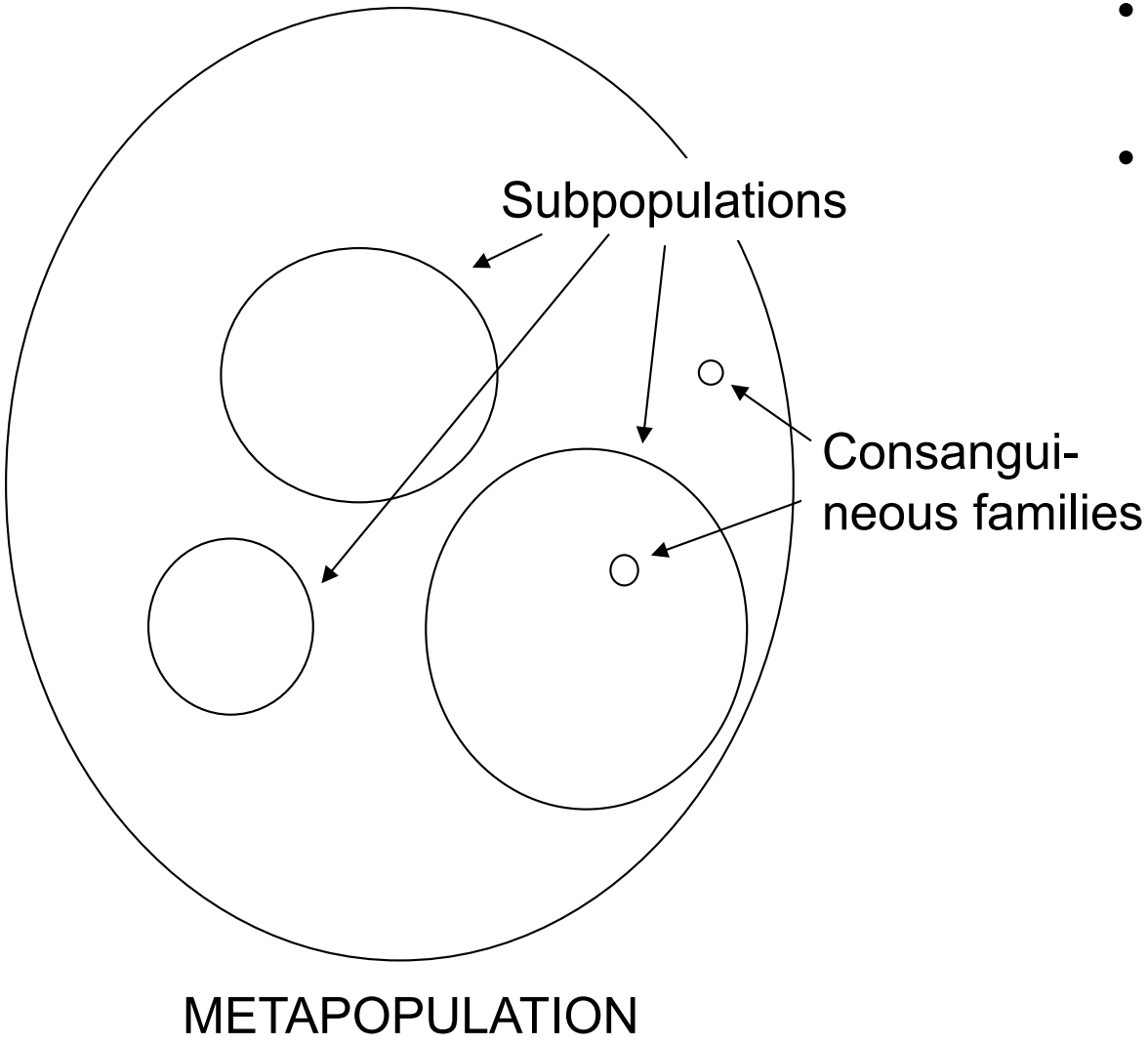
Check: 2% carriers
=> 1/50 x 1/50 x 1/4 affected
= 1/10,000 affected

(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

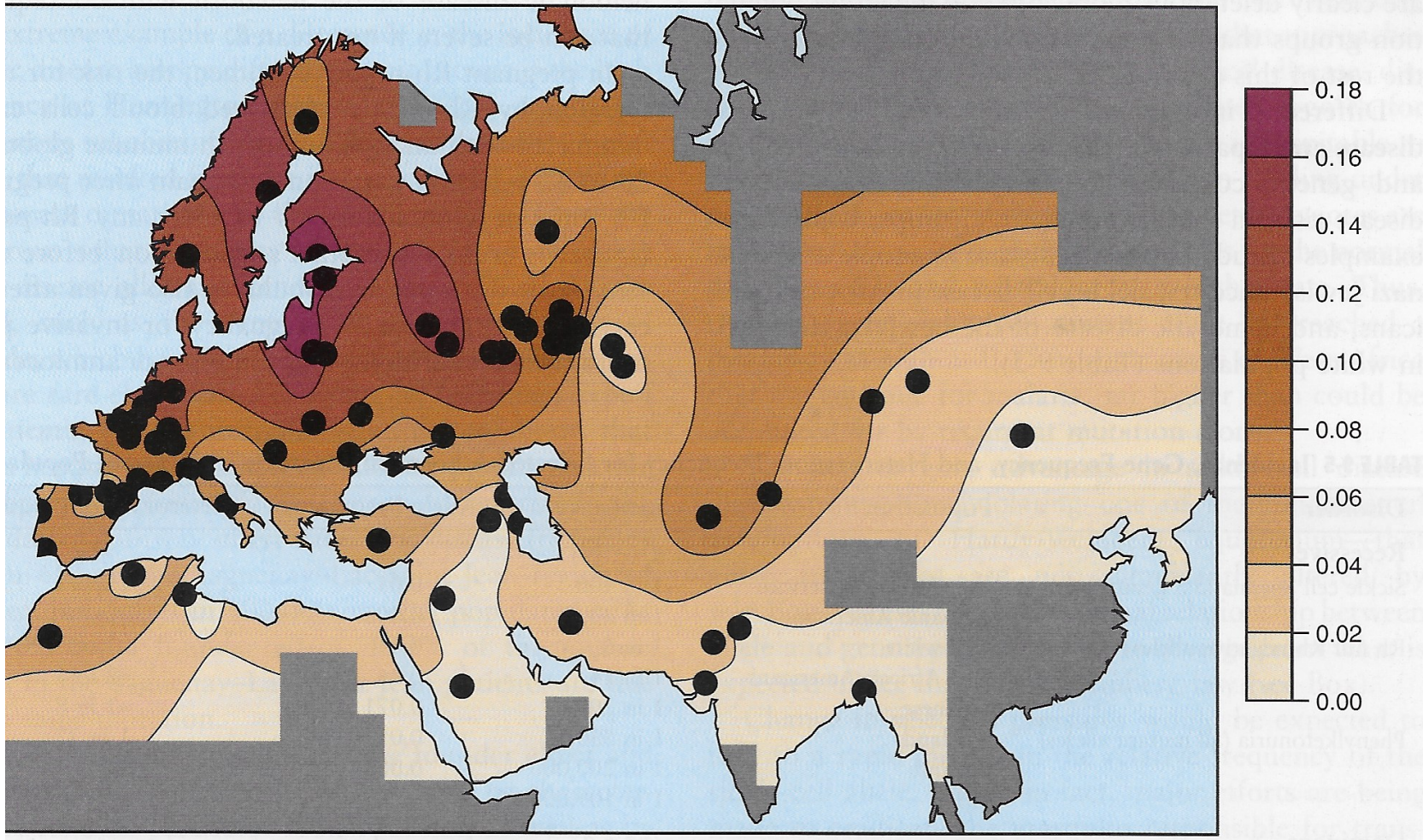


Figure 9-1 **The frequency of Δ*CCR5* alleles in various geographical regions of Europe, the Middle East, and the Indian subcontinent.** The various allele frequencies are shown with color coding provided on the right. *Black dots* indicate the locations where allele frequencies were sampled; the rest 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

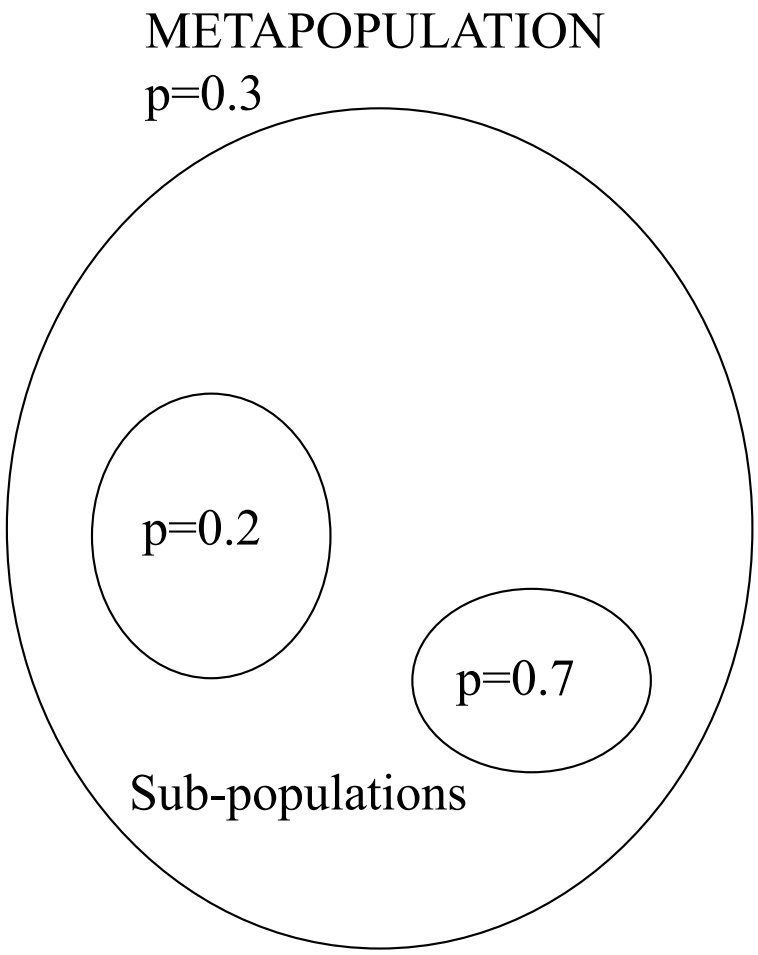
Assortative matings

= 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

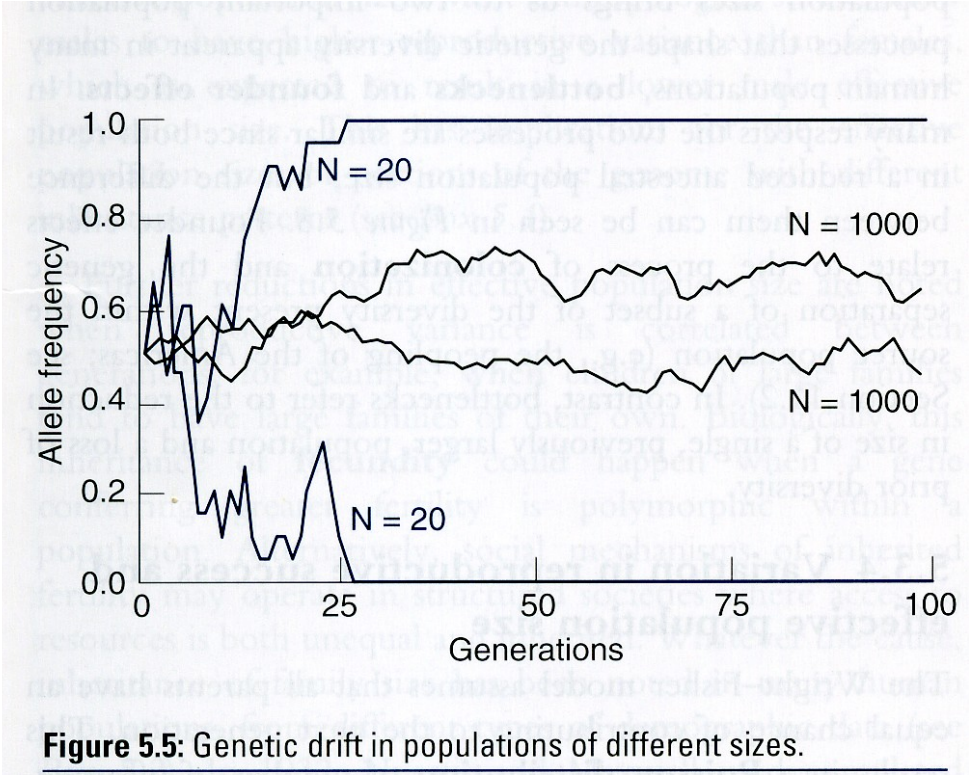
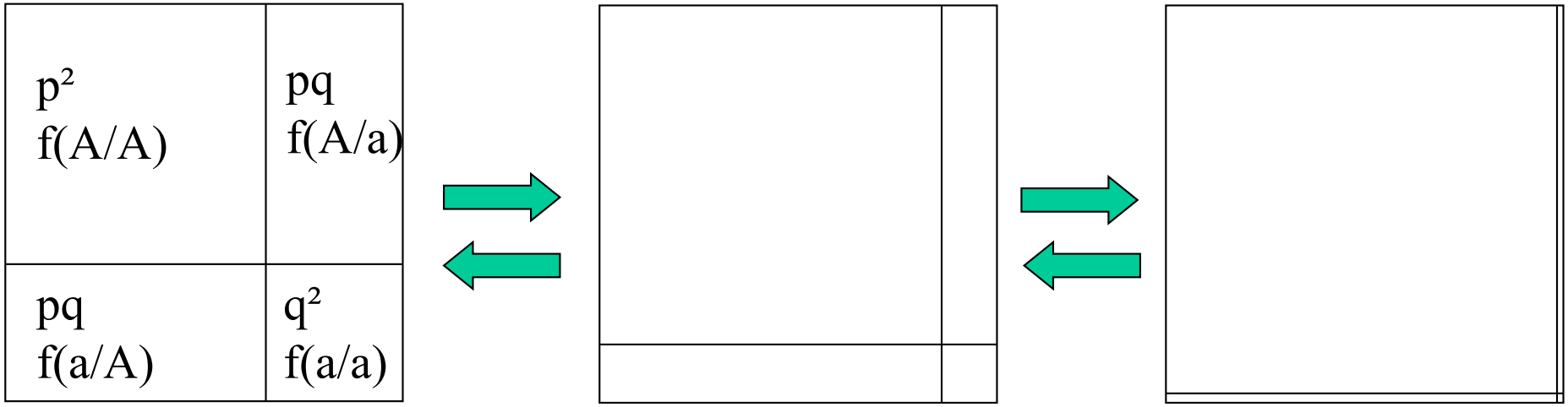


Figure 5.5: Genetic drift in populations of different sizes.

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

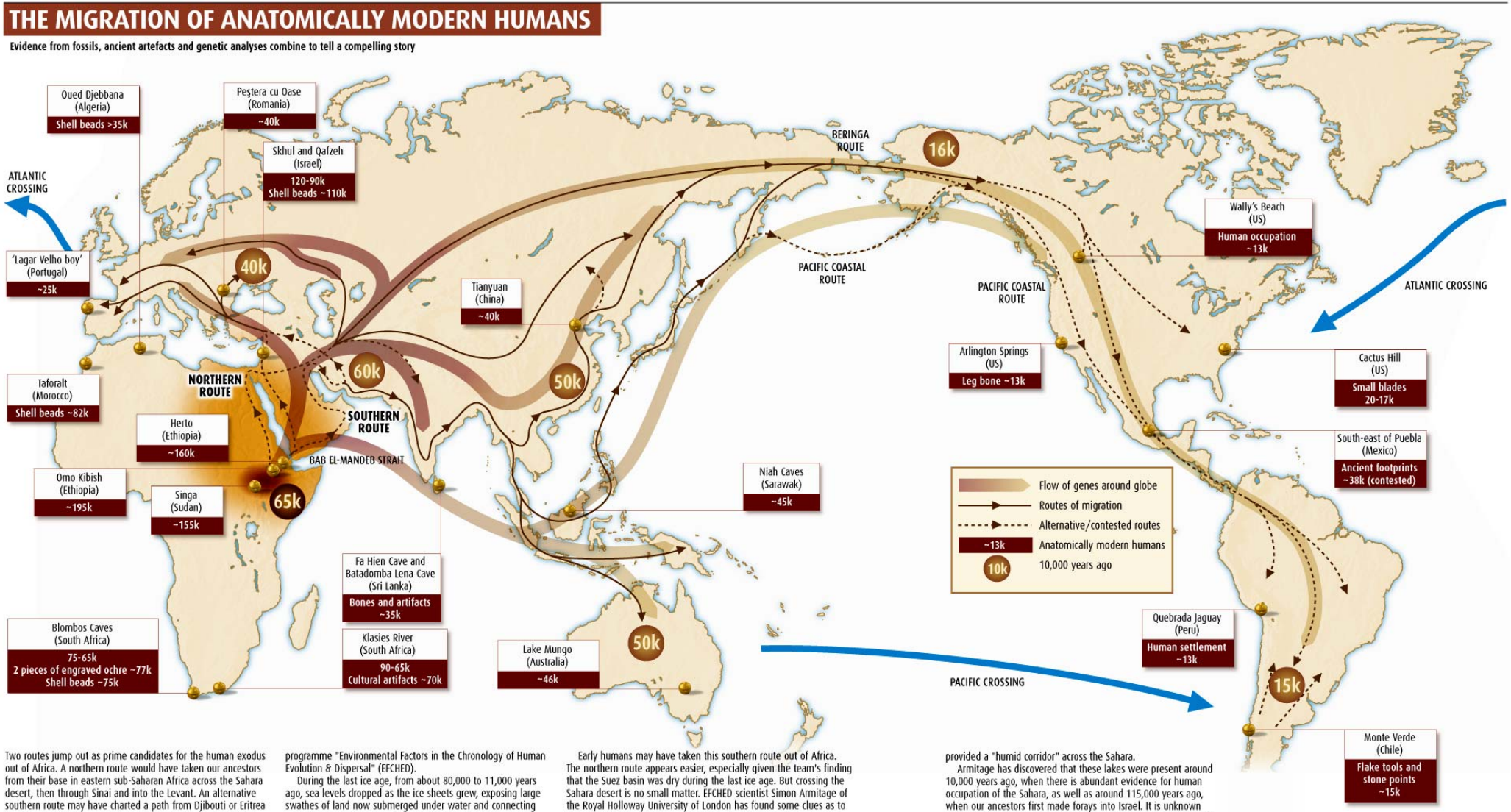
N=20 => one allele gets FIXED

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).

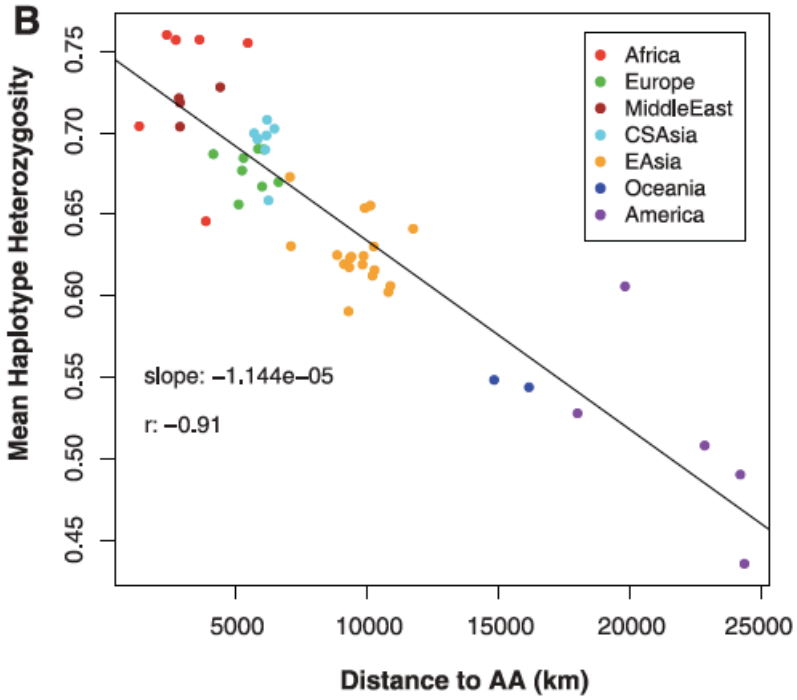
During the last ice age, from about 80,000 to 11,000 years ago, sea levels dropped as the ice sheets grew, exposing large swathes of land now submerged under water and connecting regions now separated by the sea. By reconstructing ancient shorelines, the EFCHED team found that the Bab el-Mandeb strait, 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

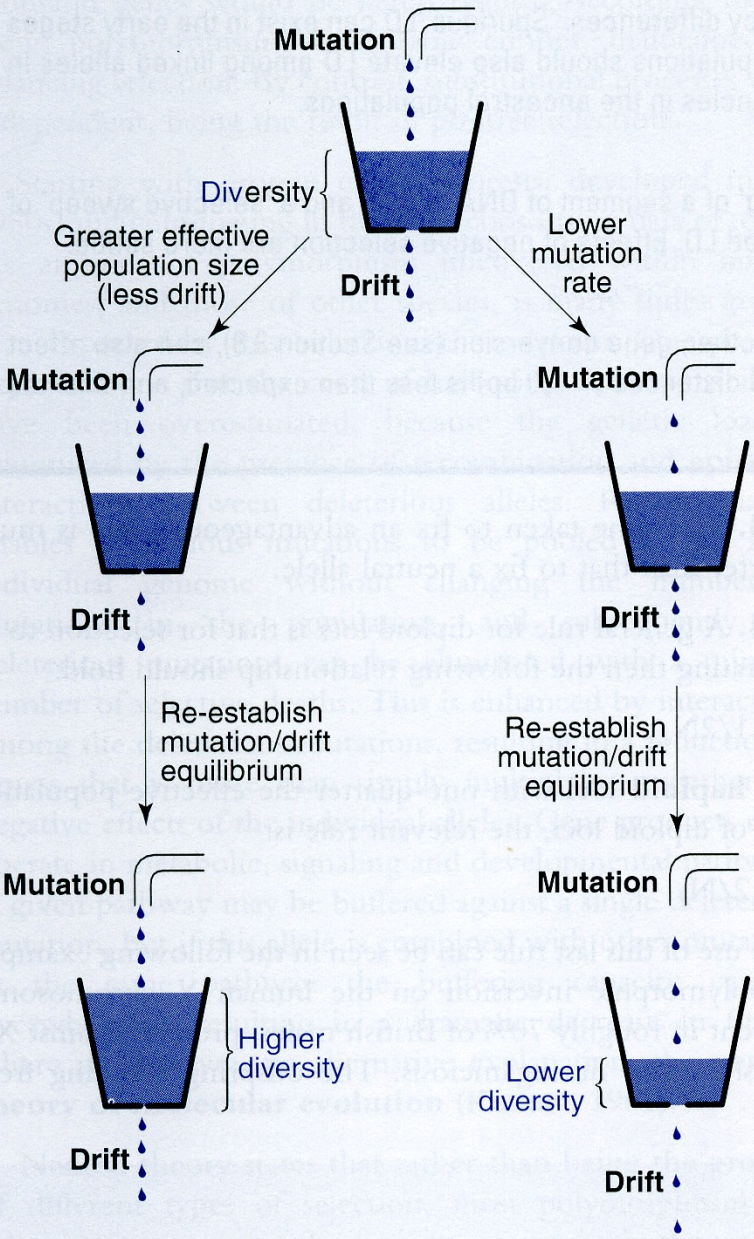


Figure 5.13: A metaphorical depiction of the relationship between mutation rate, drift and diversity.

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

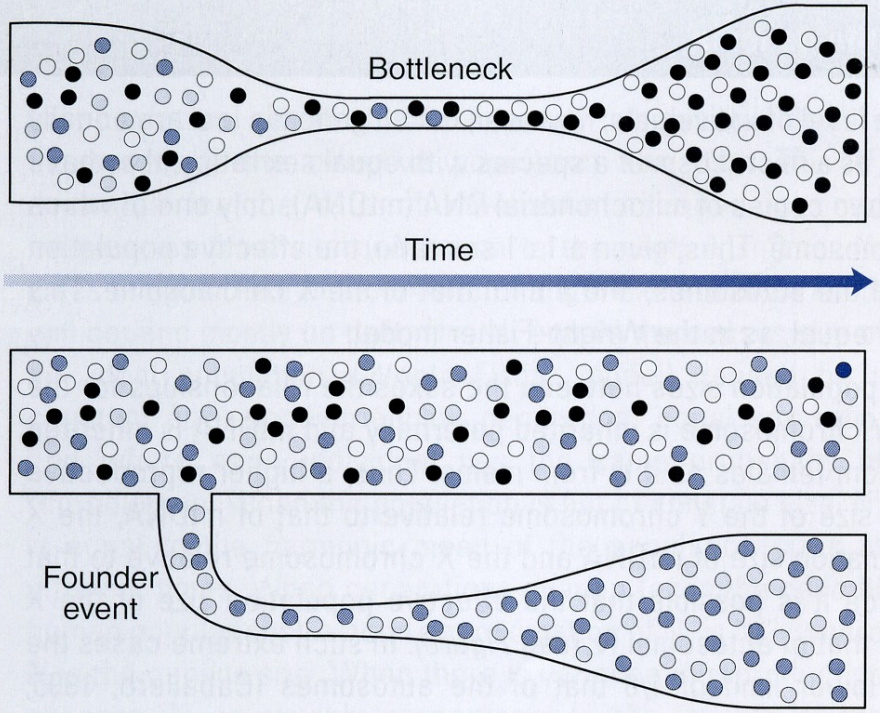


Figure 5.8: Bottlenecks and founder events.

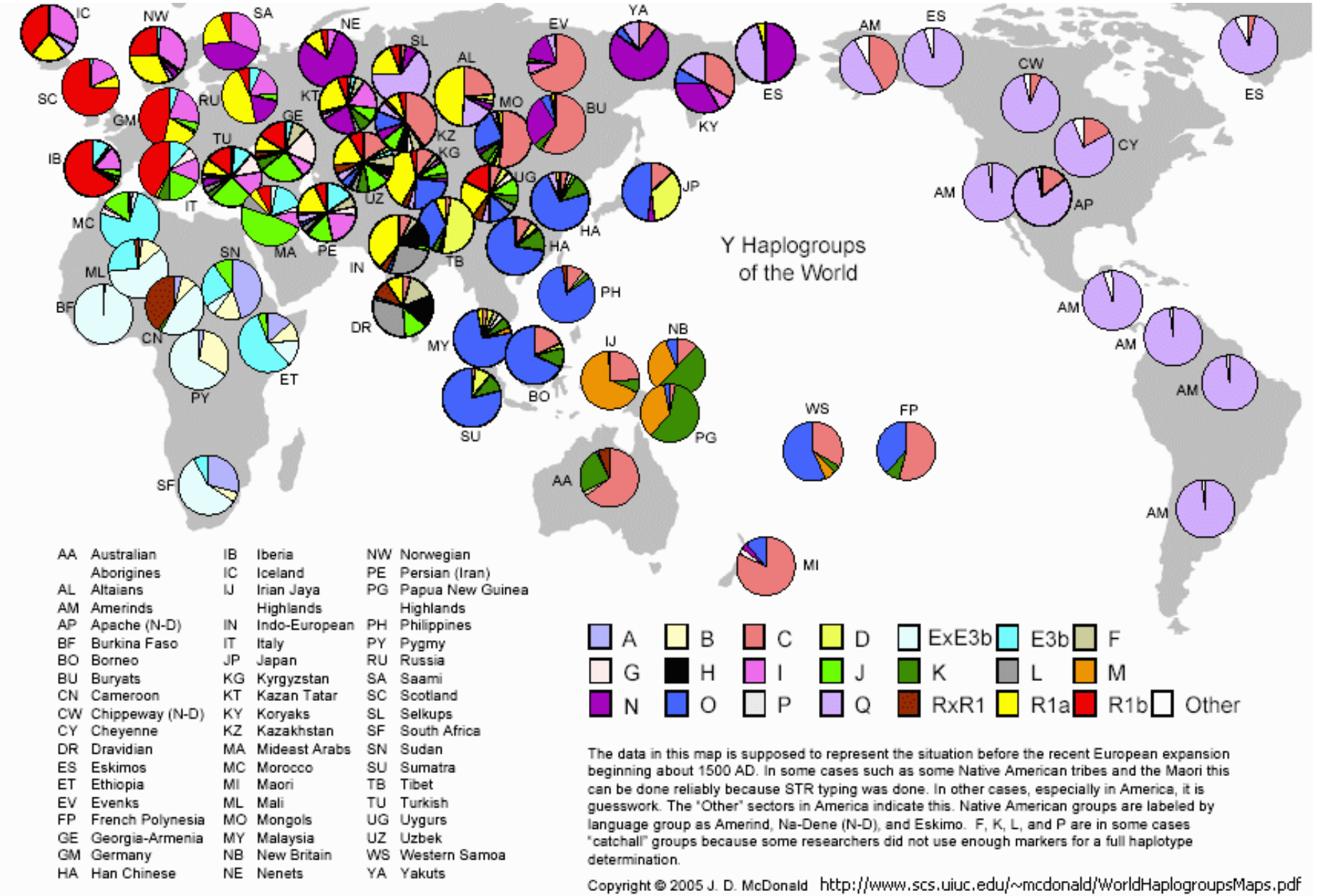
Circles of different colors represent different alleles. Both bottlenecks and founder events result in a loss of allelic diversity.

Male and female reproductive variance

Eg, in Aka pygmies (Hewlett BS, 1988)

- Male progeny mean = 6.34 variance = 8.64
- Female progeny mean = 6.34 variance = 5.20

Y-haplogroups



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

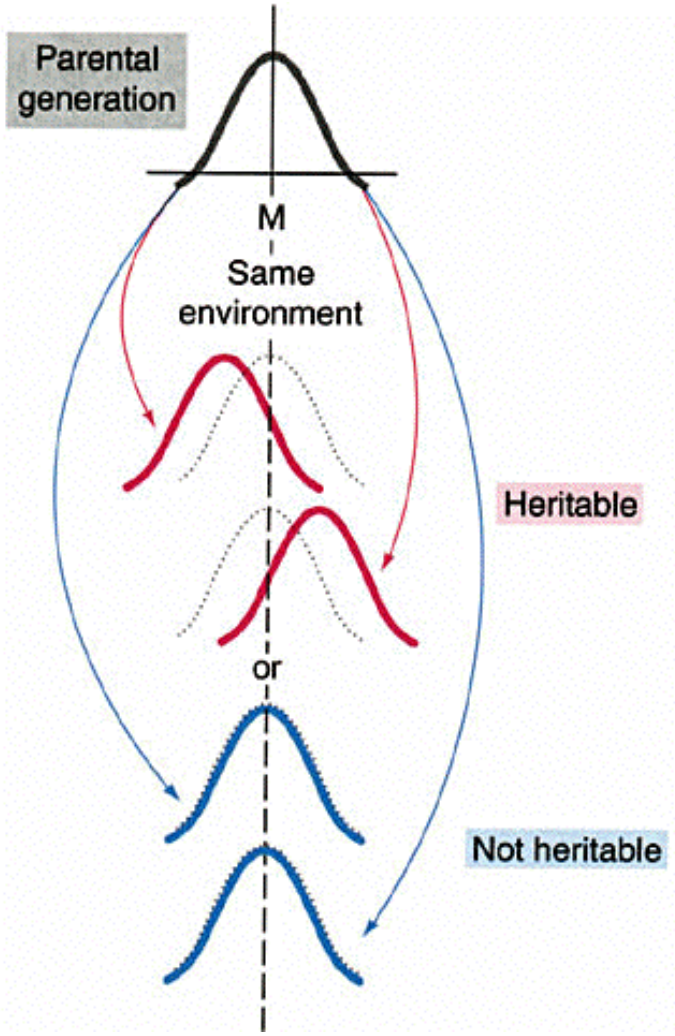
Artificial selection (empirical)

- 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

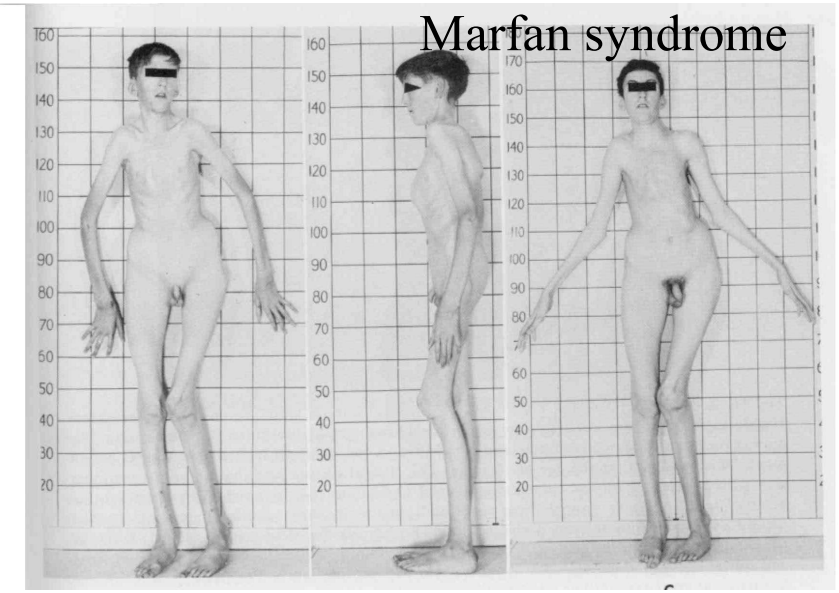
- Natural or artificial
- Differential survival and reproduction of individuals

- **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)

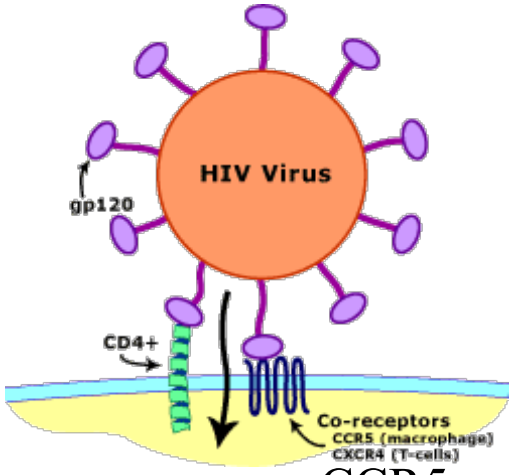


Positive selection of mutation CCR5 * delta32

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 -

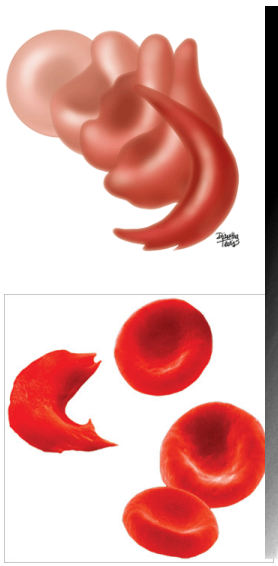
- **NEGATIVE SELECTION** : reduced fitness
  = purifying selection

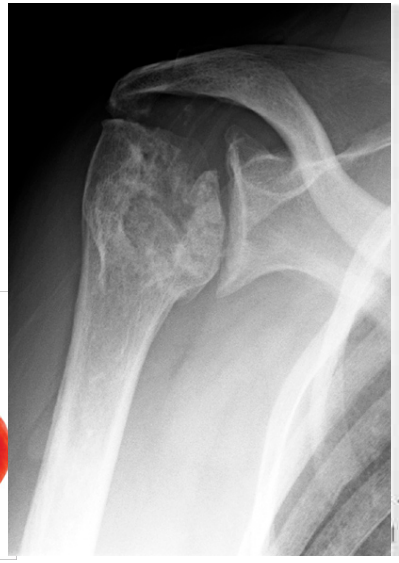
- **POSITIVE SELECTION** : increased fitness
  = adaptive selection

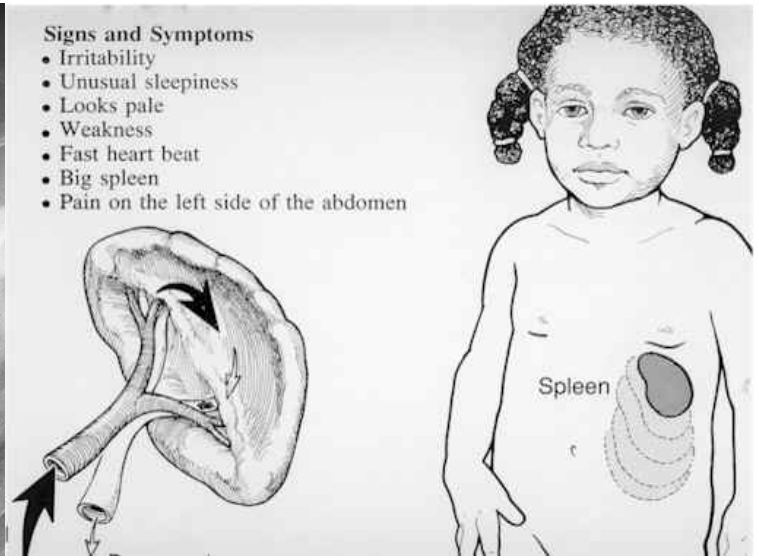
- **BALANCED SELECTION** : htz performs best 

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

Balanced Selection : ex: hereditary anemia



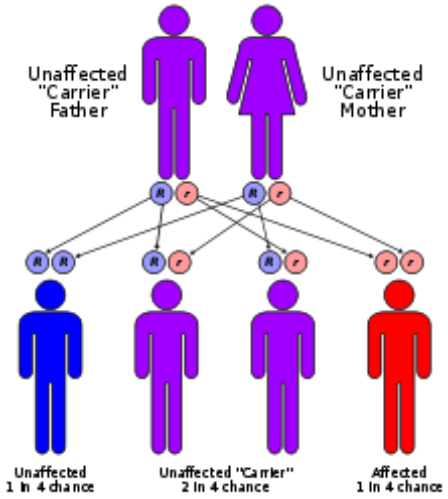




Drépanocytose (Sickle cell disease)

Thalassemia

Autosomal recessive, monogenic

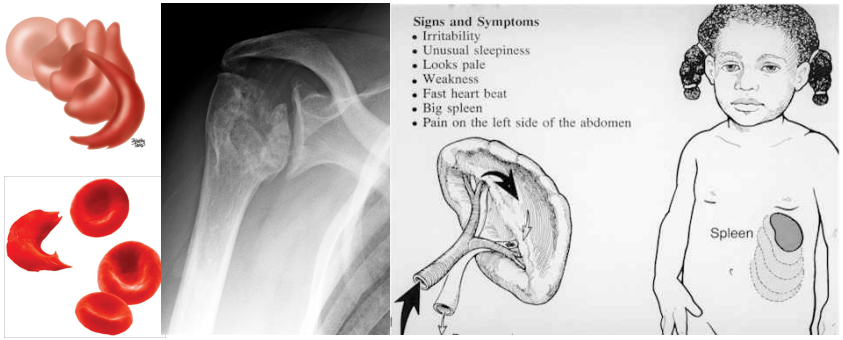


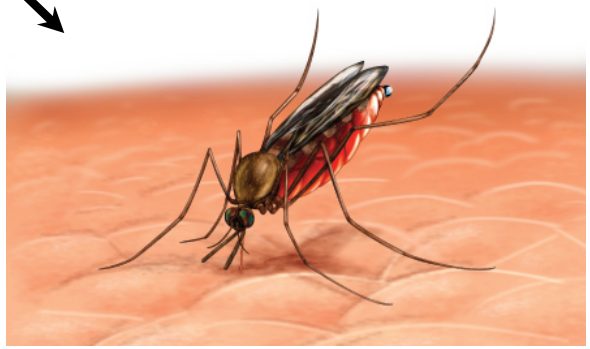
=> 10-20% carriers in some populations

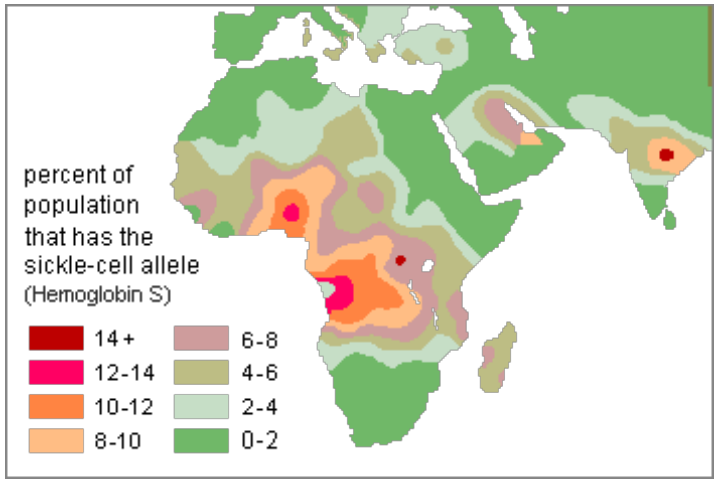
although mutated alleles disappear as patients die !

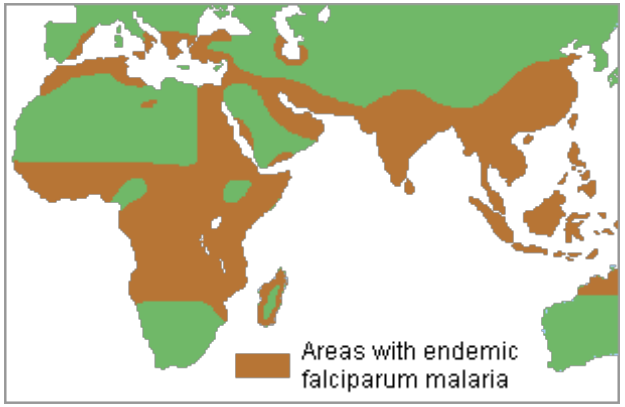
Sickle cell (and thalassemia)

follow malaria

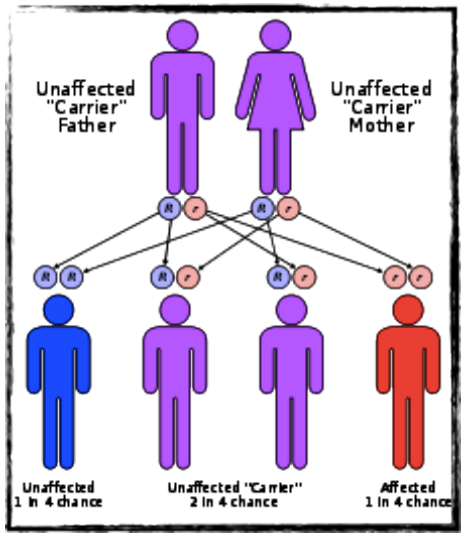






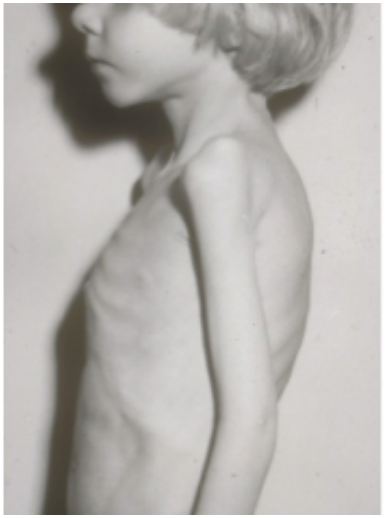


balanced polymorphisms



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

Cystic Fibrosis
(mucoviscidose)



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

Thalassemia
(beta 0)

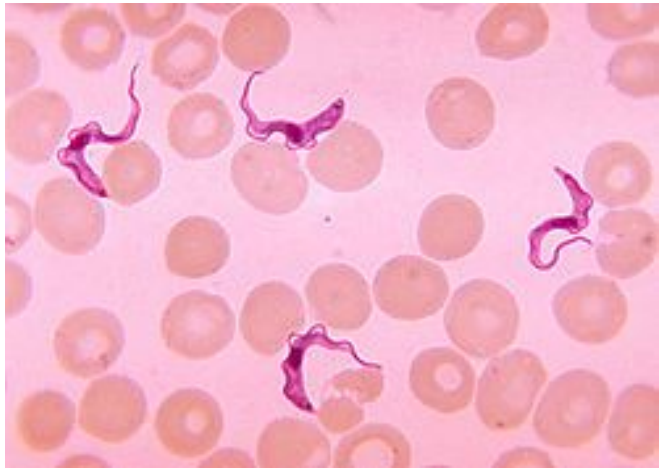
Severe disease in hmz => These genetic changes can not get fixed in population

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



TseTse fly

Balanced selection

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

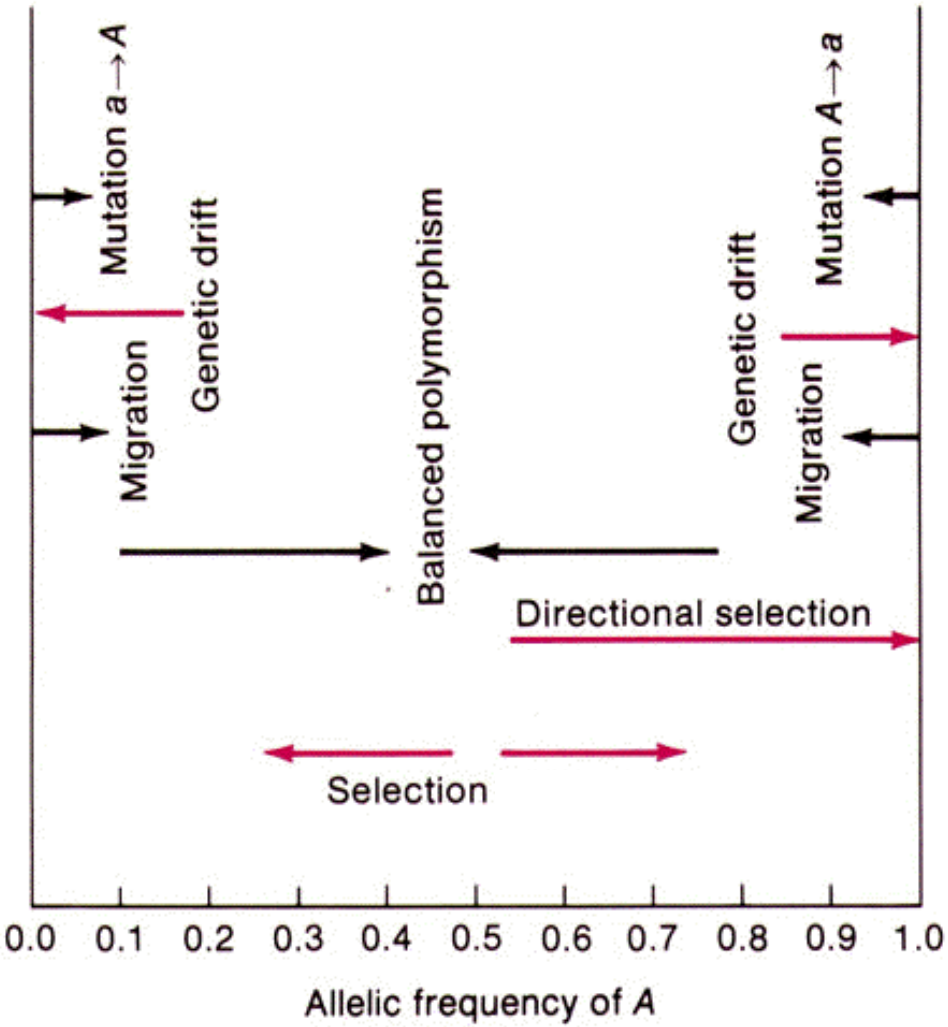
=> Such genetic changes can not get fixed in population.. or everyone would be affected with CF, Sickle Cell, …

Balanced selection in oligogenic / multigenic/ complex disorders

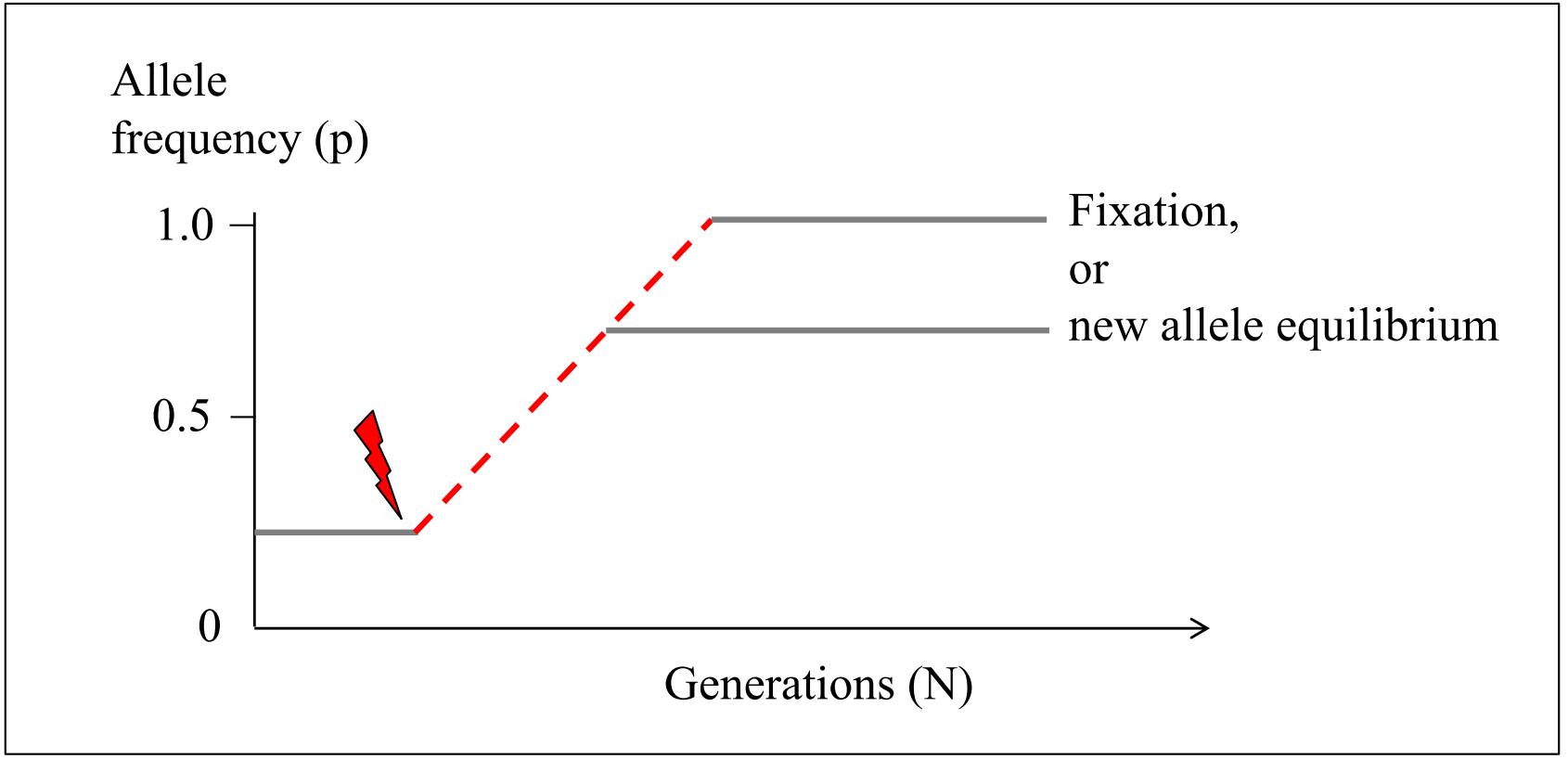
- TLR4
  - Septic shock < > ischemic cardiopathy
- HFE
  - Iron deficiency < > iron overload
- FV Leiden
  - Fewer hemorrhages < > thrombophilia

Microevolution

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



Positive selection (adaptive mutation)

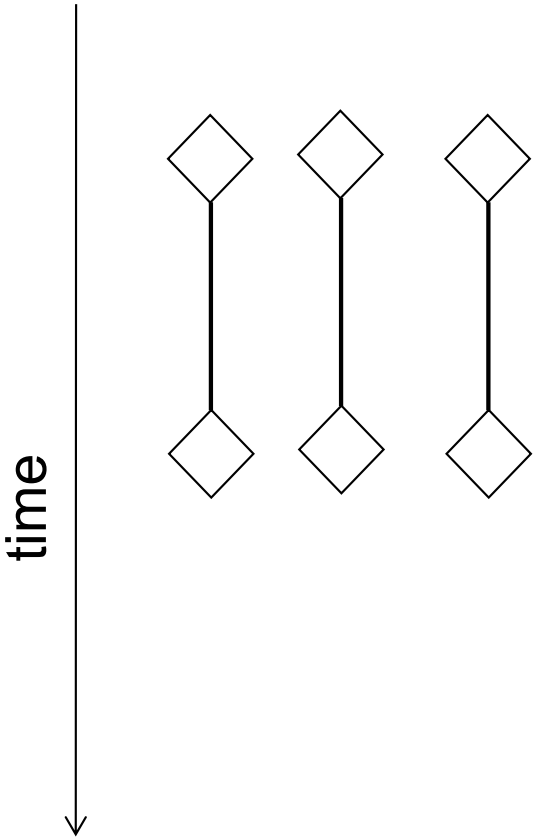


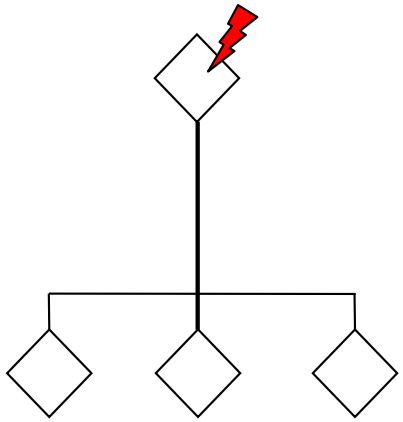
Δp/ΔN ≡ Selection = differential number of offspring, untill fixation, or untill new equilibrium reached

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

Mutation with selective advantage

Population with constant size



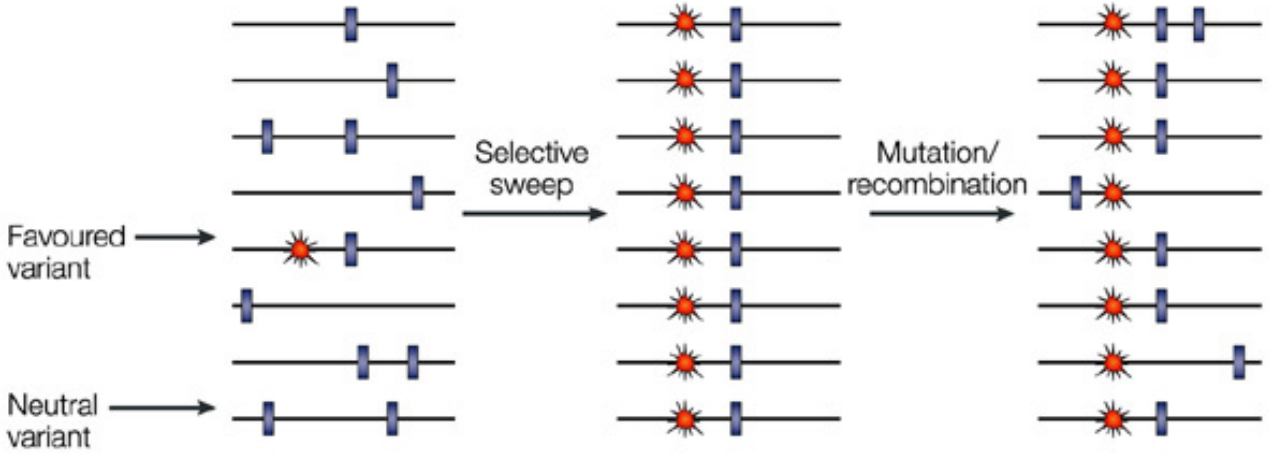


Because of selective advantage, fitness > 1

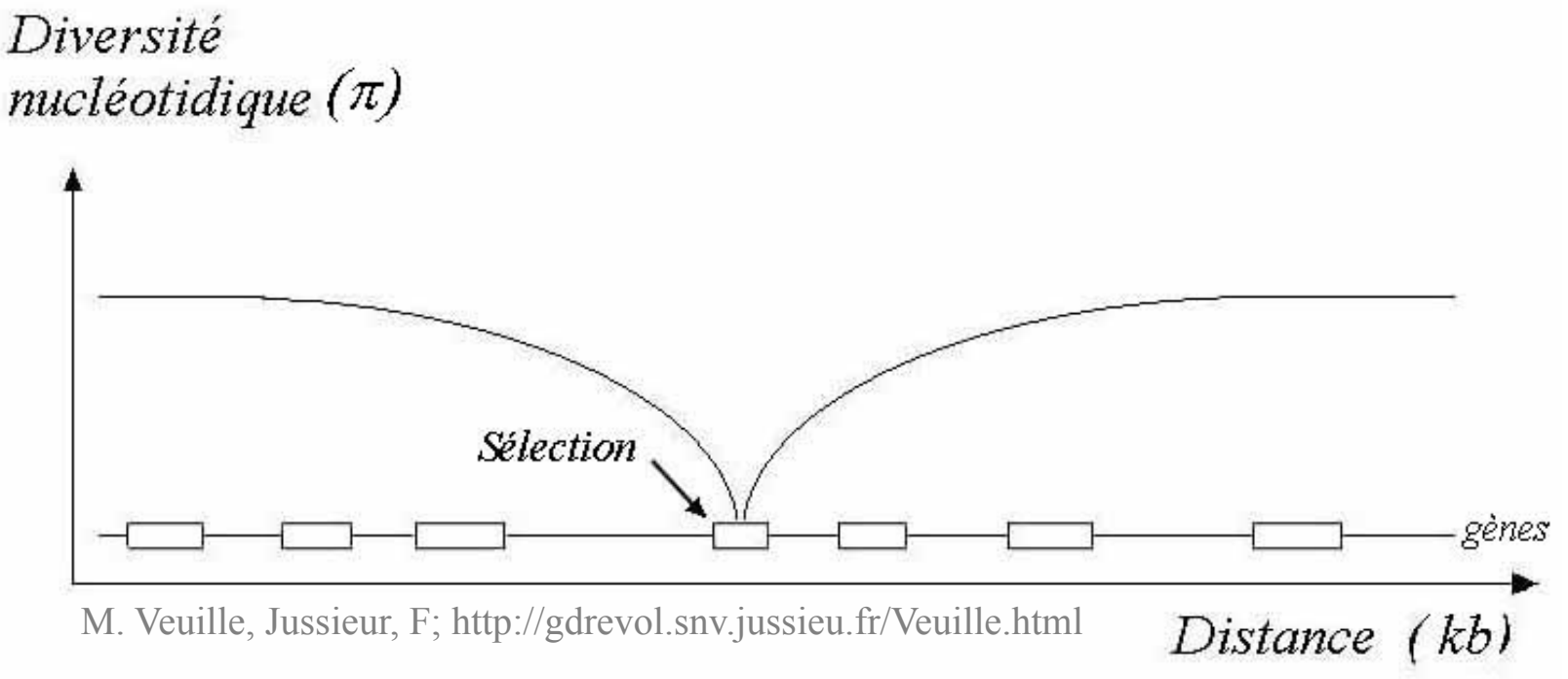
Over time, mutation settles, surrounded by large haplotype in LD (with constant size of population)

Haplotype structure analysis => detects recent positive selection events

Selective sweep

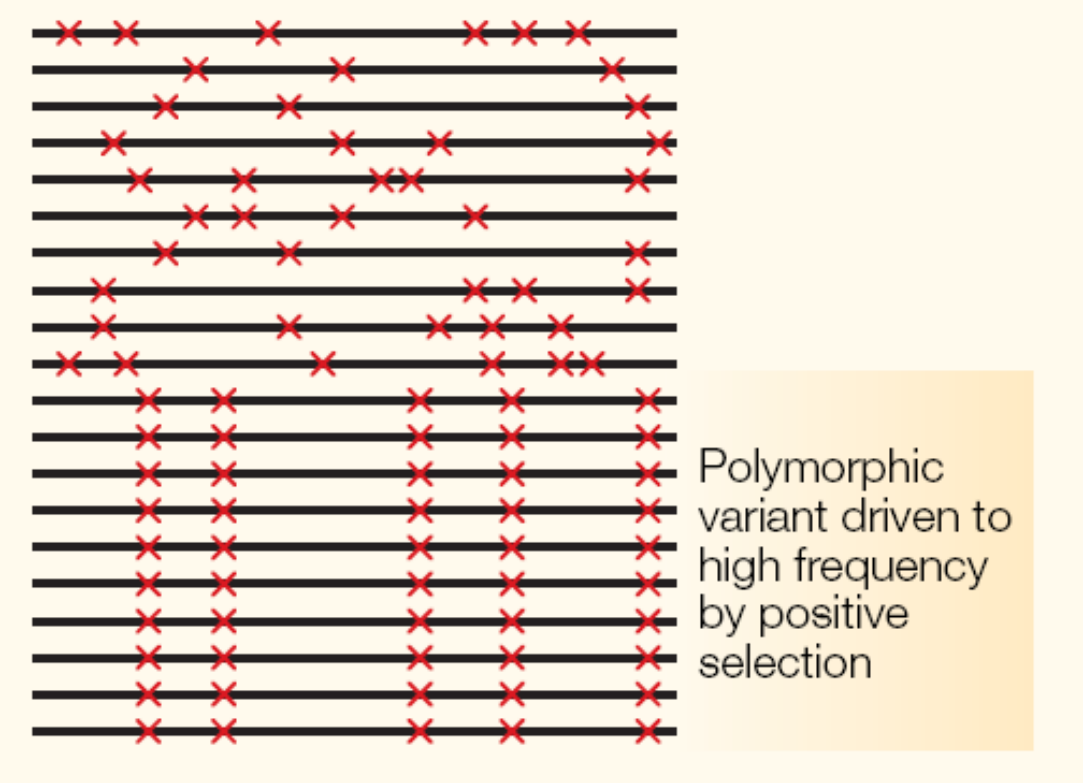


Boffelli D et al. 2004



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

Selective sweep



Gilbert et al. 2004

Selective sweep in ongoing evolution

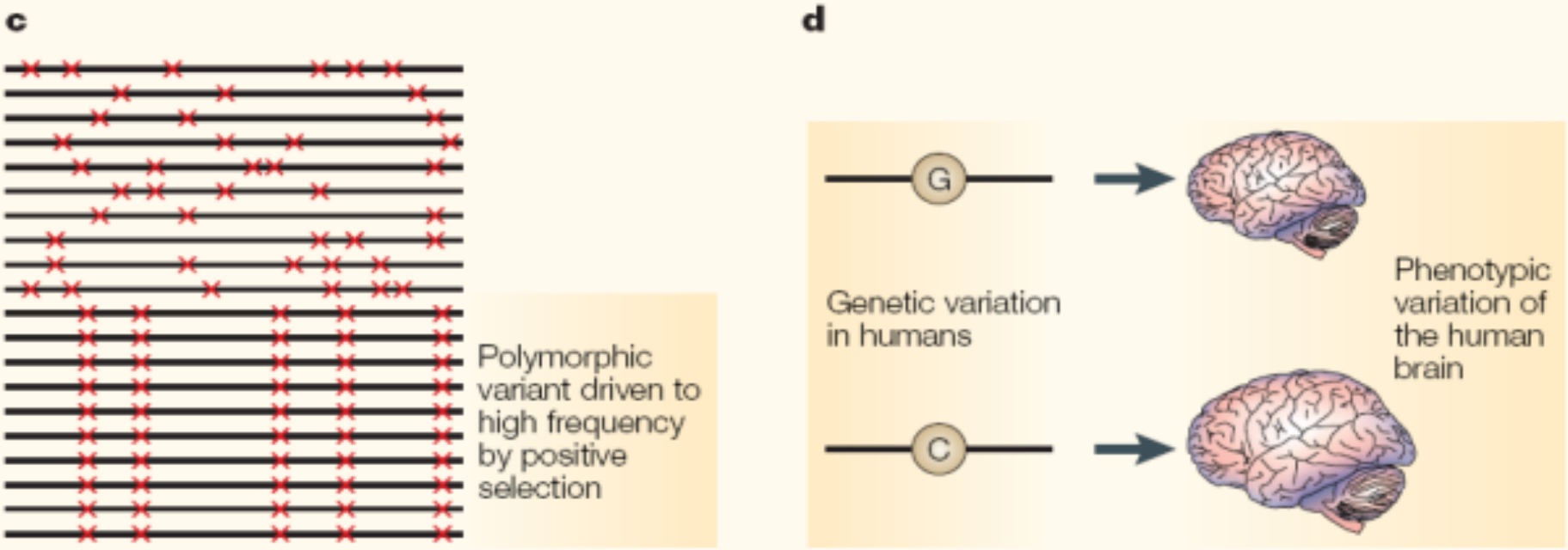


Figure 4 | **A methodological template for investigating the genetic basis of human brain evolution. a** | Large-scale comparisons of brain-related genes across four strategically selected species that include the human, Old World monkey, rat and mouse. These comparisons can reveal broad genome-wide 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

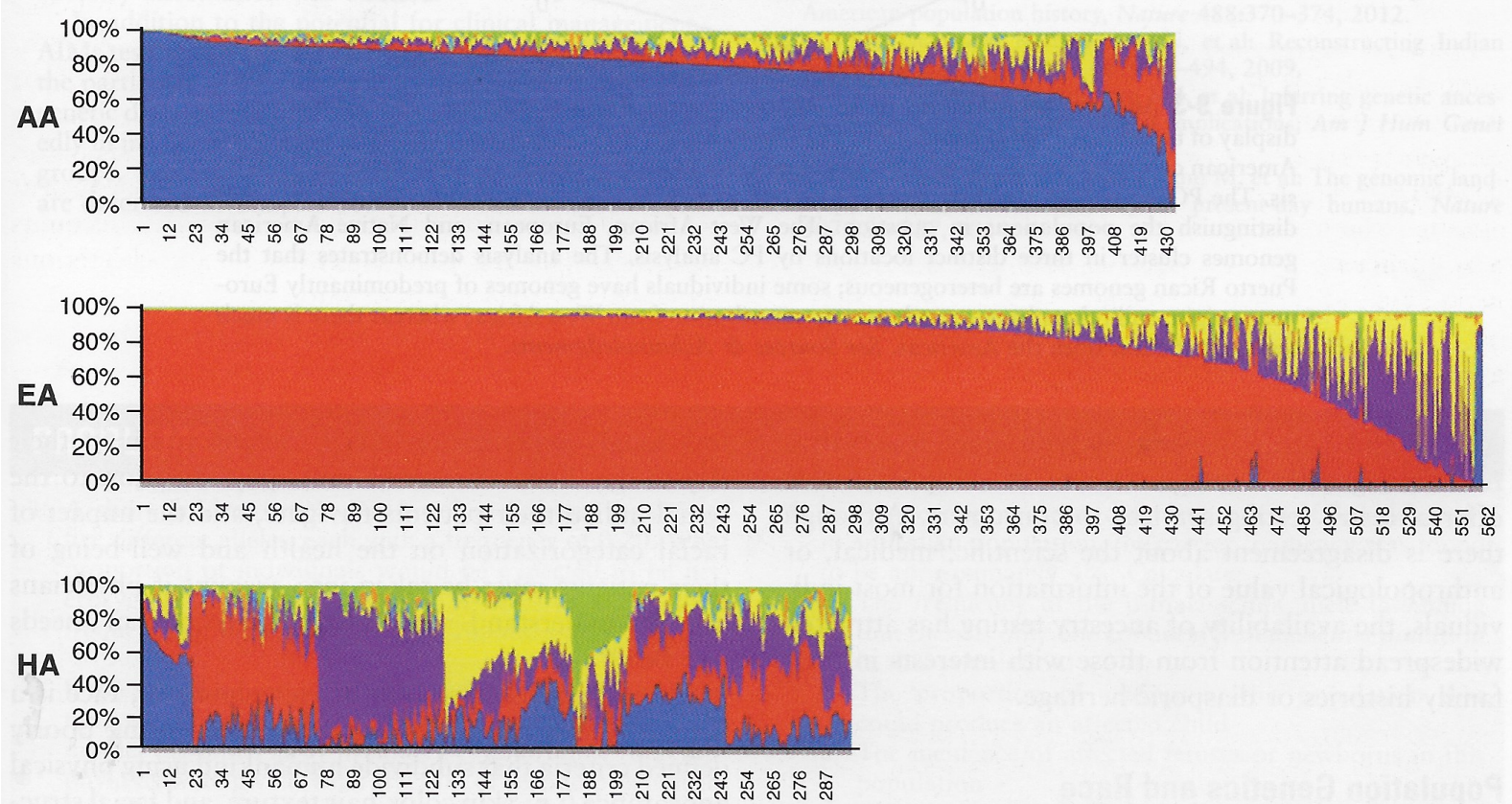
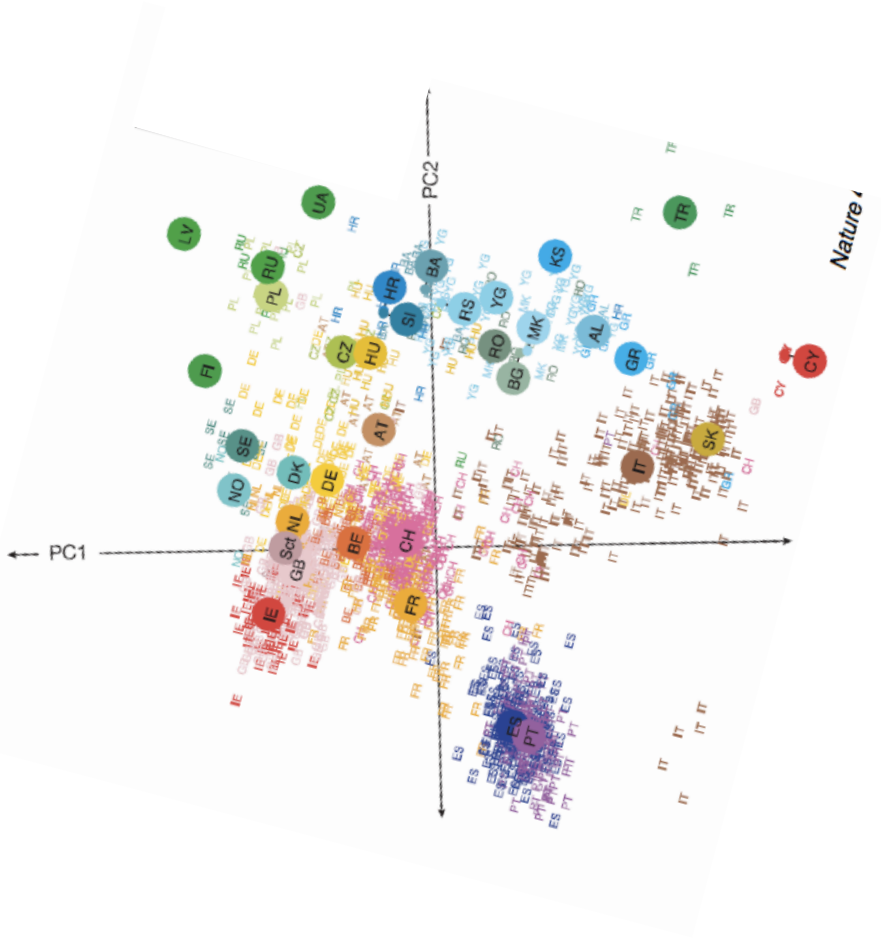


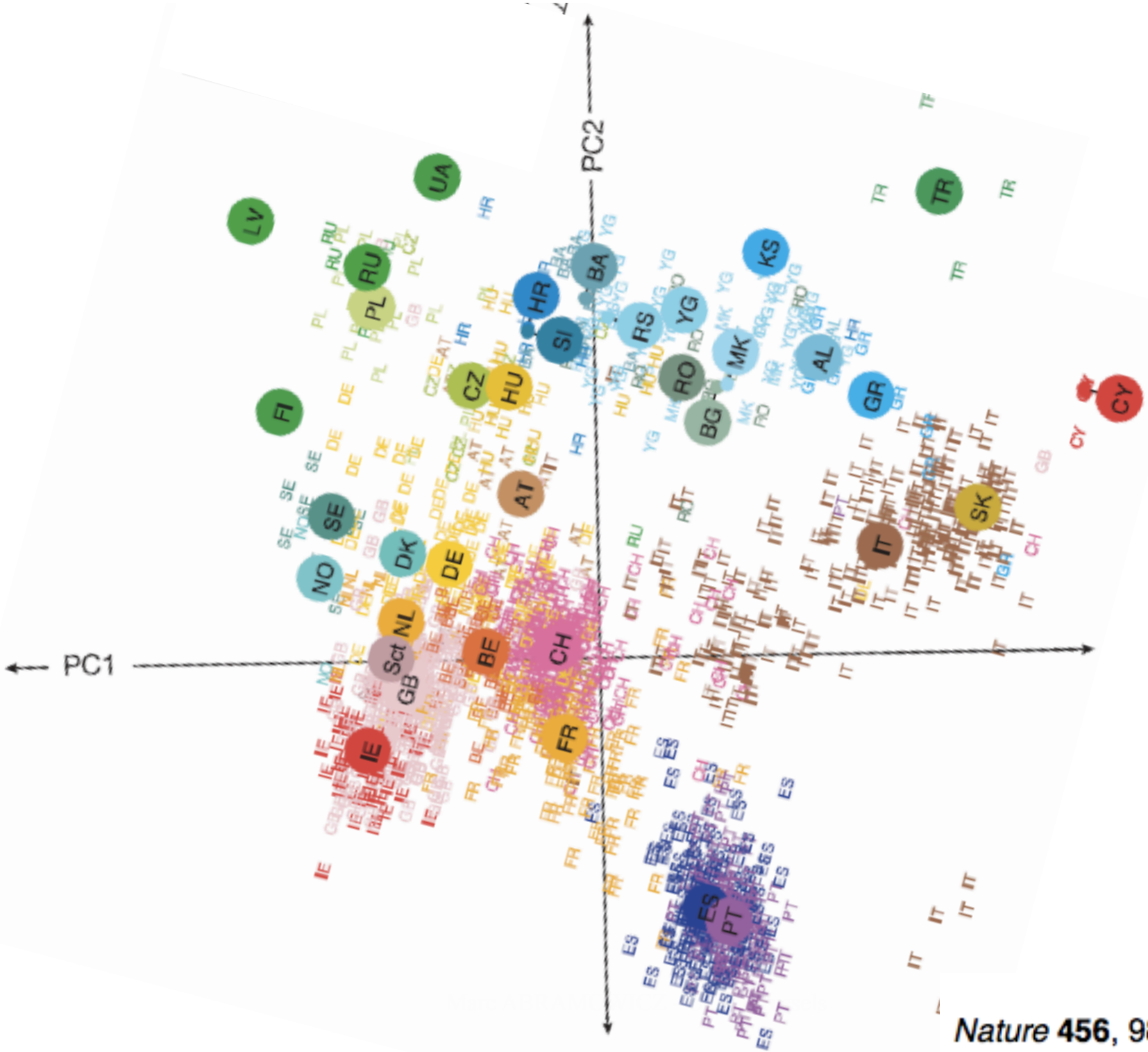
Figure 9-2 **Mixed ancestry of a group of Americans who self-identify as African American (AA), European American (EA), and Hispanic American (HA) using ancestry informative markers.** Each vertical line represents one individual ((totaling hundreds, as shown by the numbers), and subjects are displayed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

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

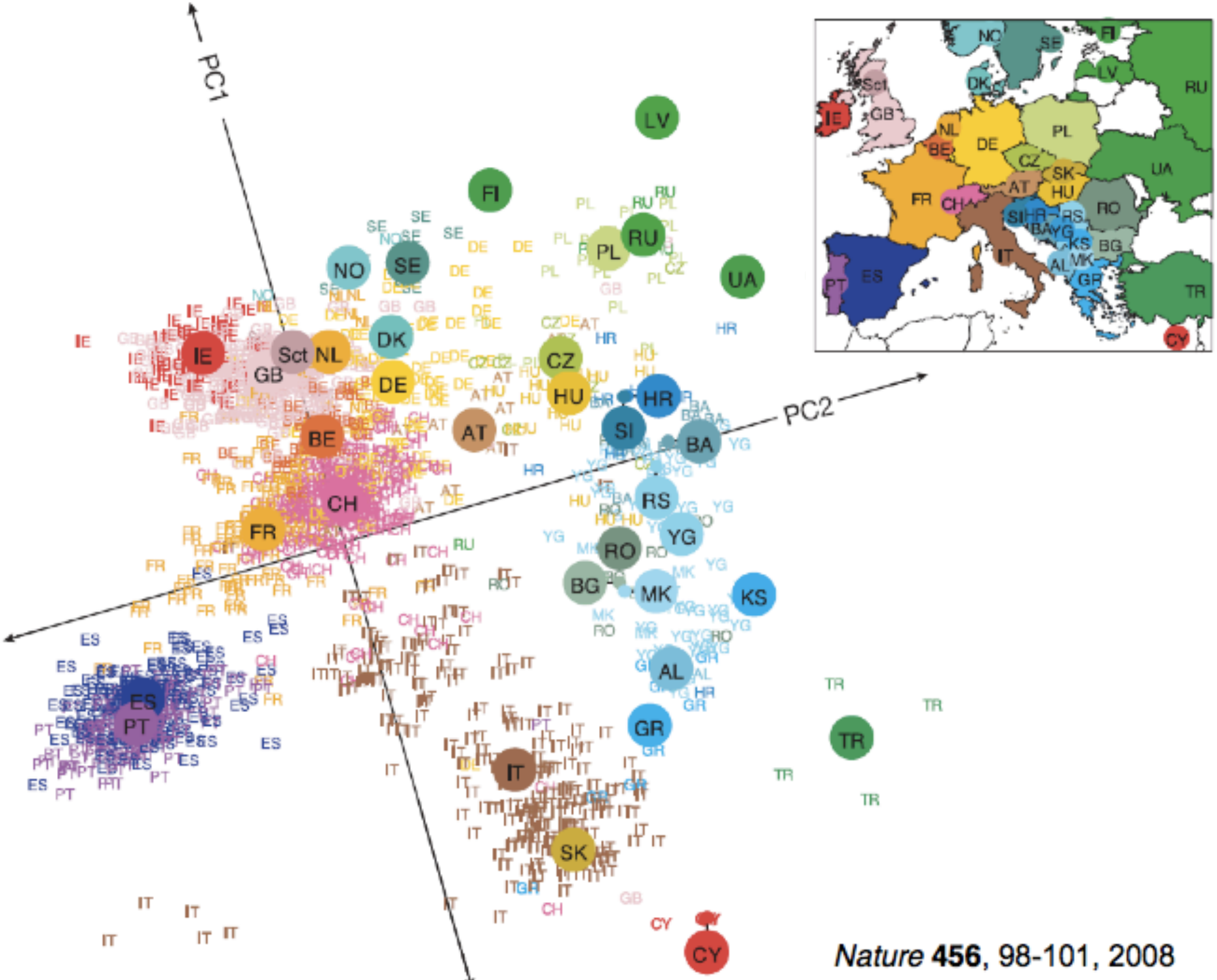
Principal Component Analysis

- 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