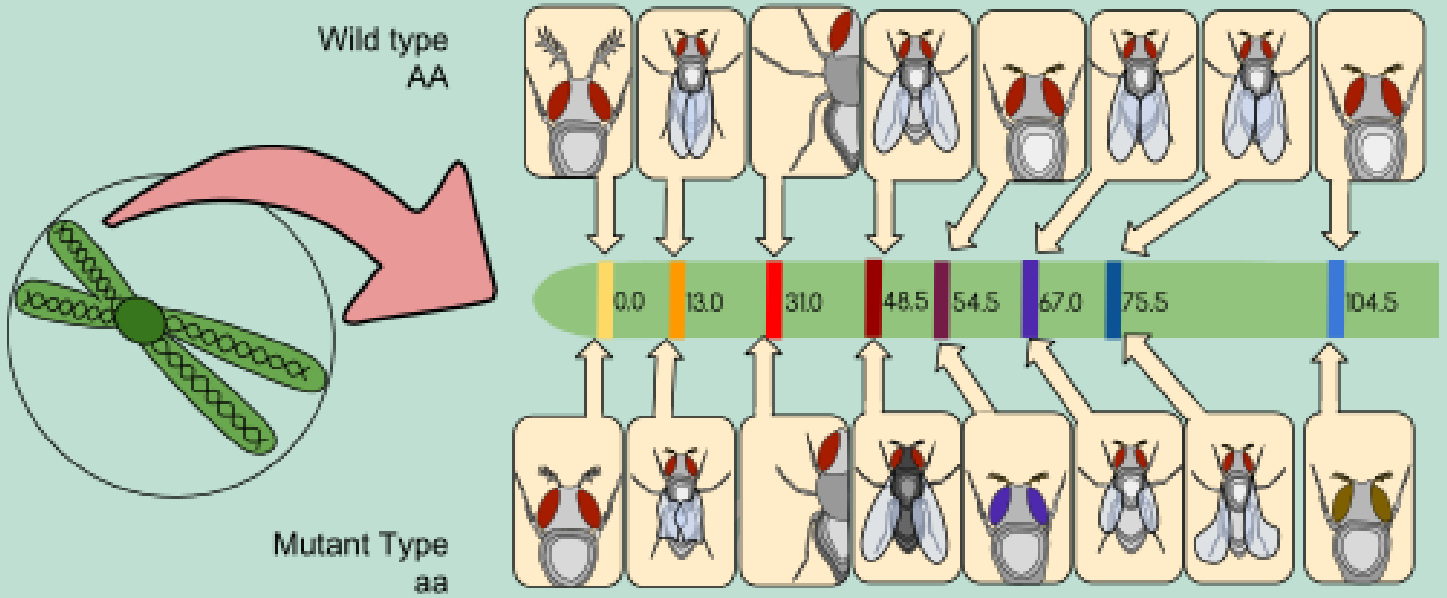
RISK ESTIMATION FROM PEDIGREES

#3

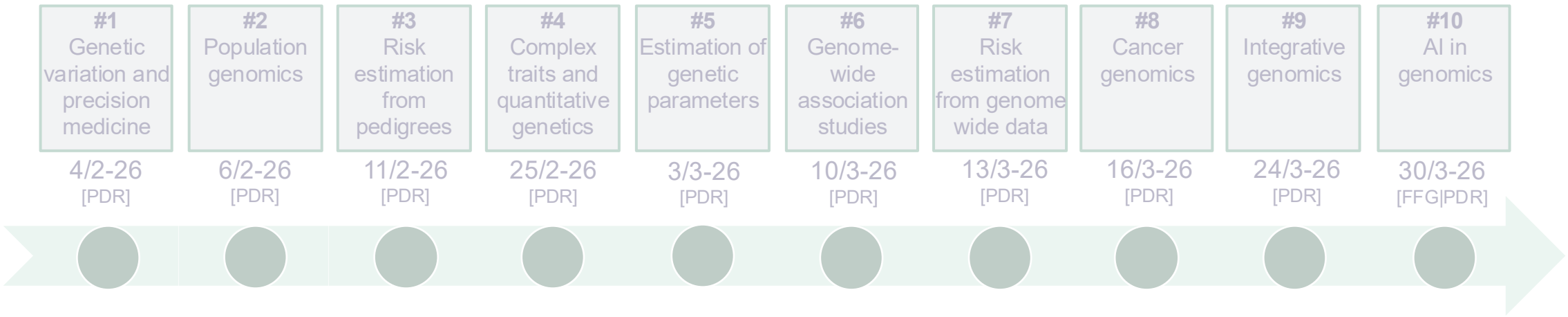
PALLE DUUN ROHDE

palledr@hst.aau.dk



AALBORG
UNIVERSITY

LETS GET STARTED



LINKAGE AND GENETIC TESTING

Today we will talk about

- Risk calculations [Bayes theorem]
- Linkage
- Molecular tools for diagnosis;
  - direct vs indirect test

Moved to session 4

OUTLINE

08:15 – 09:15	Recap + Exercises E15 [Part III]
09:15 – 09:30	Break
09:30 – 09:50	Lecture 1 [*Genetic risk assessment]*
09:50 – 10:30	Group work
10:30 – 11:10	Break + Exercises I [E1-E3]
11:10 – 11:40	Lecture 2 [*Linkage*]
11:40 – 11:55	Exercises II [E4-E8]
11:55 – 12:00	Reflection

OUTLINE

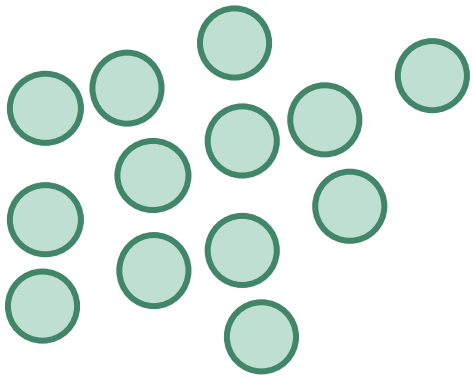
08:15 – 09:15	Recap + Exercises E15 [Part III]
09:15 – 09:30	Break
09:30 – 09:50	Lecture 1 [*Genetic risk assessment]*
09:50 – 10:30	Group work
10:30 – 11:10	Break + Exercises I [E1-E3]
11:10 – 11:40	Lecture 2 [*Linkage*]
11:40 – 11:55	Exercises II [E4-E8]
11:55 – 12:00	Reflection

SHORT RECAP FROM LAST

- Population genomics
  - The study of the distribution of hereditary variation across time and space in species and populations [Bugge, F. 2008]

GENETIC VARIATION

IN A SINGLE LOCUS



A random sample of individuals of whom we know the genotype of in a single locus

Co-dominant (i.e., we can observe both alleles in heterozygote individuals).

The population is polymorph in one autosomal locus with the alleles **A** and **a**, and three genotypes, **AA**, **Aa** and **aa**.

The frequencies of the alleles are denoted **p** and **q**, and the frequency of the genotypes are P_{AA}, P_{Aa} and P_{aa}.

Note! There is a difference between $\hat{\mathbf{p}}$ and $\mathbf{p}$. The hat ($\widehat{\ }$) indicates that it is an estimate ($\hat{\mathbf{p}}$) over the true parameter ($\mathbf{p}$). For simplicity we ignore $\widehat{\ }$.

FREQUENCIES

Genotype	AA	Aa	aa	Σ
Count	n_{AA}	n_{Aa}	n_{aa}	N
Genotype frequency	n_{AA}/N	n_{Aa}/N	n_{aa}/N	1

Allele frequency of A: $p = (2 \times n_{AA} + n_{Aa})/2 \times N$

Allele frequency of a: $q = (2 \times n_{aa} + n_{Aa})/2 \times N$

We are counting the alleles

Check! $p + q = 1$

All alleles are counted

HARDY-WEINBERG LAW

So far, we have computed allele frequencies by counting genotypes

Genotype frequencies → Allele frequencies

Under certain conditions, we can compute genotype frequencies in the next generation

Allele frequencies→ Genotype frequencies

However, that requires some assumptions.

THE NEUTRAL POPULATION

- Random mating
- No selection
- No genetic drift (infinite population size)
- No migration
- No mutation

Hardy-Weinberg principal describes the relationship allele- and genotype frequencies in the neutral population

HARDY-WEINBERG EQUILIBRIUM

After one generation under HW assumptions the genotype frequencies will be in equilibrium:

Genotype	AA	Aa	aa
Frequency	p^2	$2pq$	q^2

Allele frequencies do not change!

THE NEUTRAL POPULATION?

The **constancy of allele frequencies** from generation to generation only holds under the assumptions of HW-law.

- Random mating
- No selection
- No genetic drift (infinite population size)
- No migration
- No mutation

Does the neutral population exists

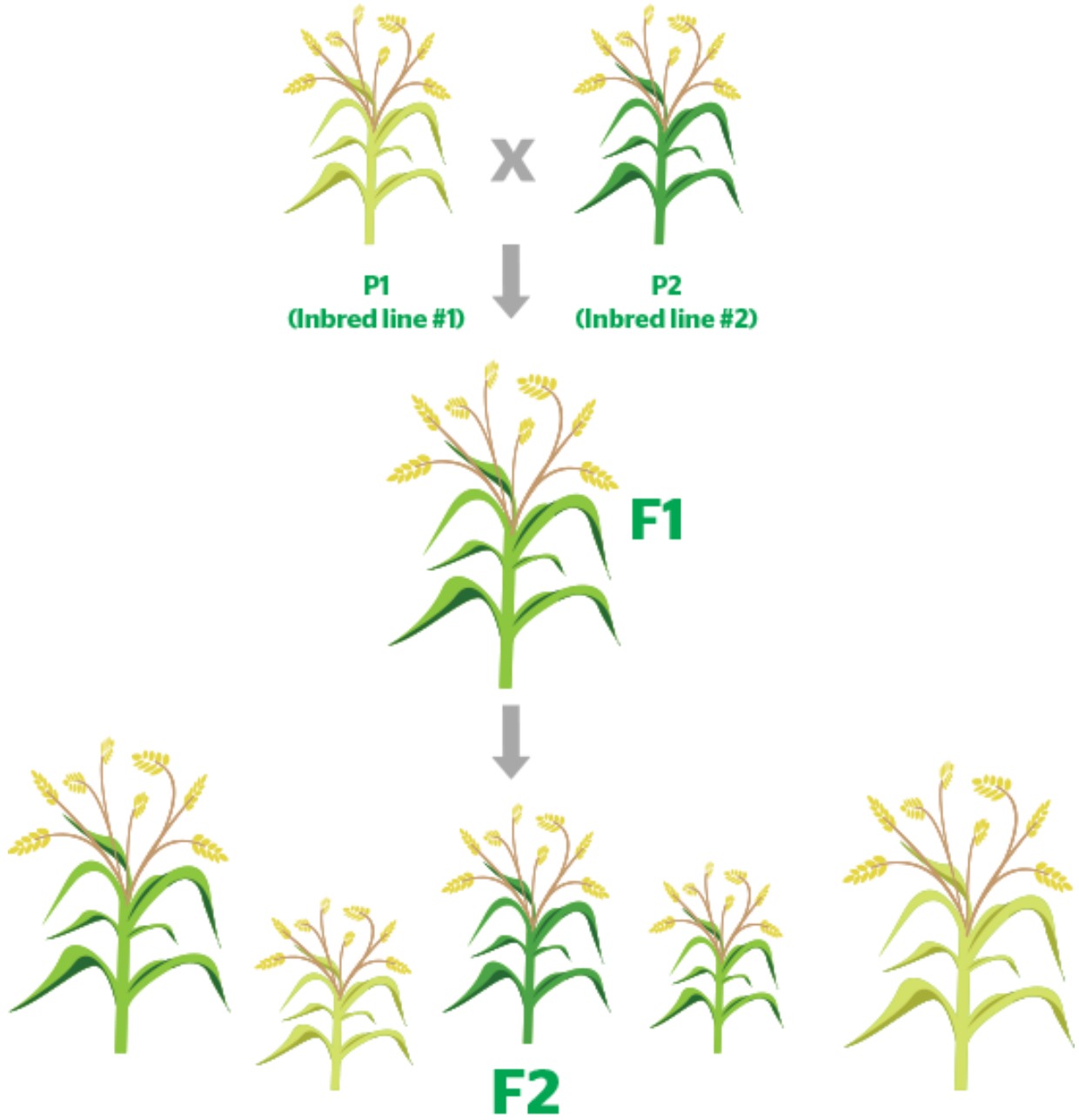
?

THE NEUTRAL POPULATION

- **Random mating**
- No selection
- No genetic drift (infinite population size)
- No migration
- No mutation

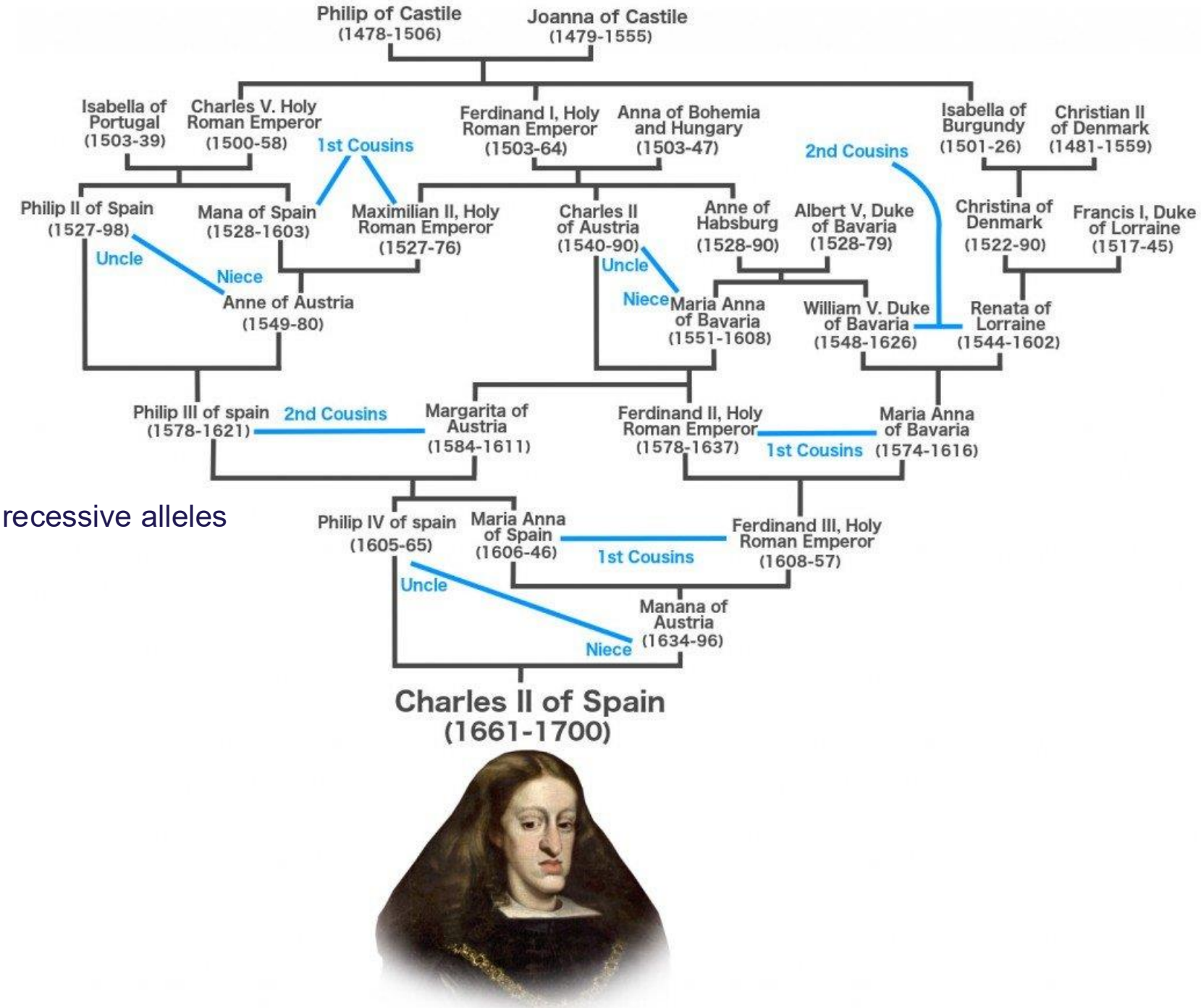
INBREEDING

- Mating between relatives
  - Heterosis | Hybrid vigor



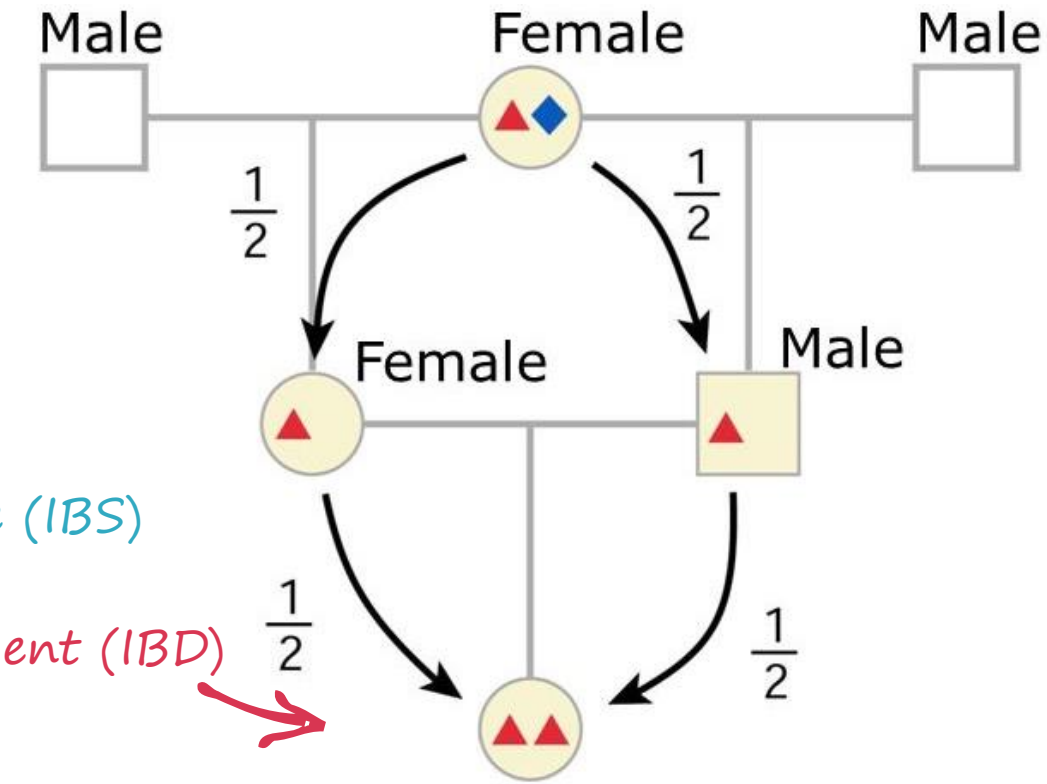
INBREEDING

- Mating between relatives
  - Heterosis | Hybrid vigor
  - Inbreeding depression
    - Accumulation of deleterious recessive alleles



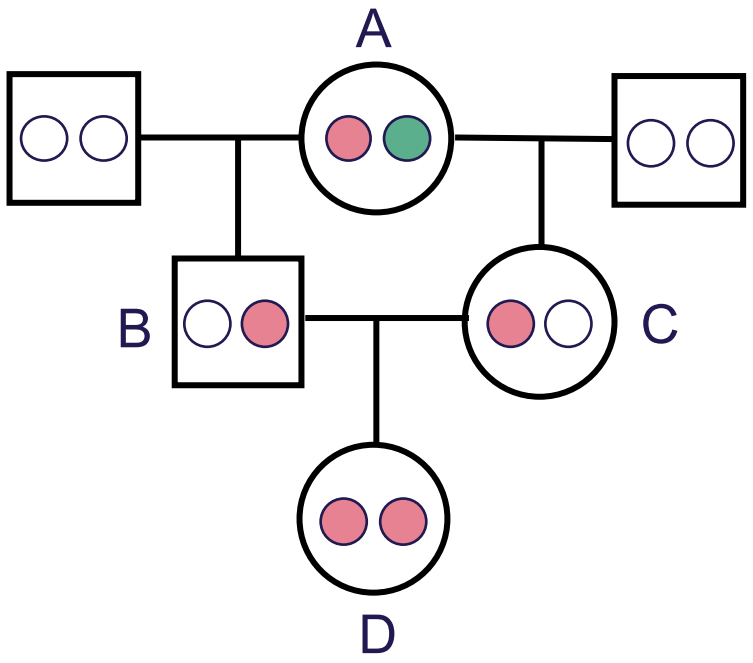
THE INBREEDING COEFFICIENT

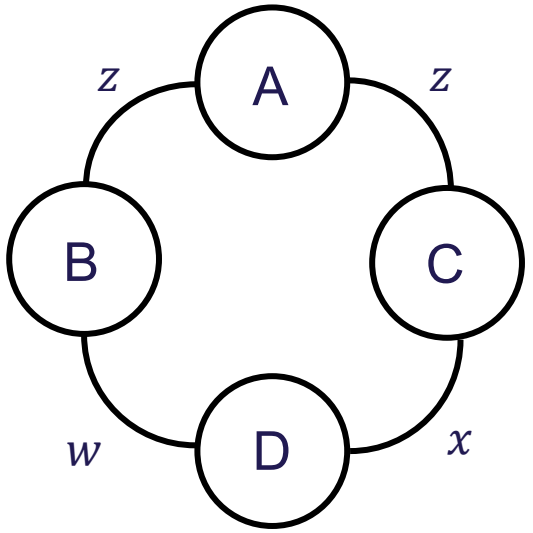
The inbreeding coefficient (F) is the probability that two alleles in an individual trace back to the same copy in a common ancestor.



Identical by state (IBS)
VS
Identical by descent (IBD)

THE INBREEDING COEFFICIENT





Follow the transmission of alleles.

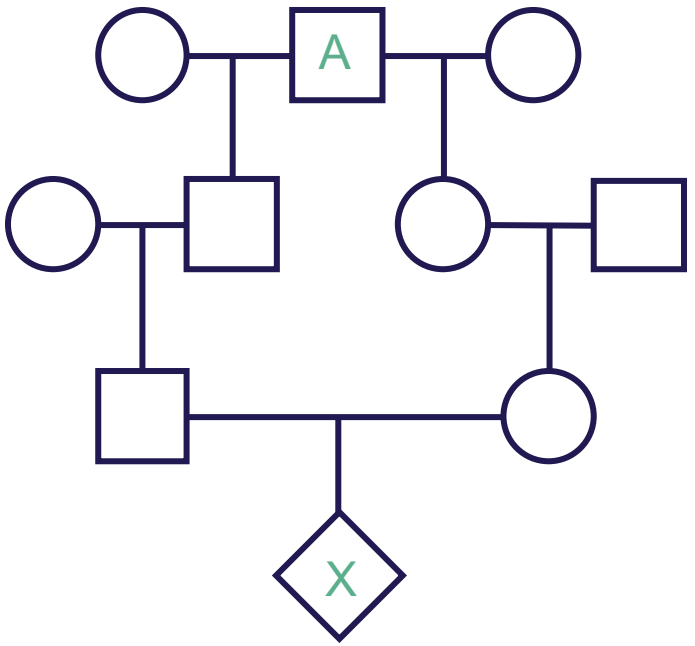
$$F_D = \left(\frac{1}{2}\right)^n (1 + F_A)$$

where n is the number of individuals in the loop without the individual we are computed F for.

$$F_D = \left(\frac{1}{2}\right)^3 (1 + F_A)$$

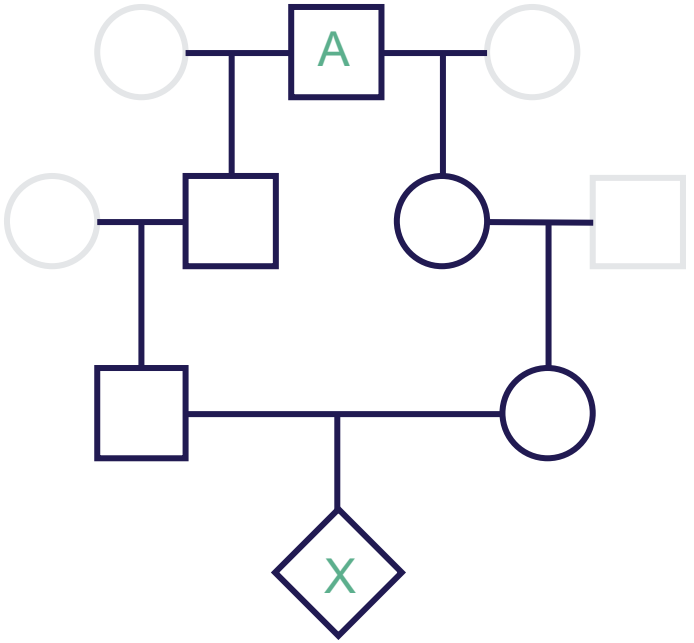
YOUR TURN

What is the inbreeding coefficient for individual X assuming individual A is not inbred ($F_A = 0$)?

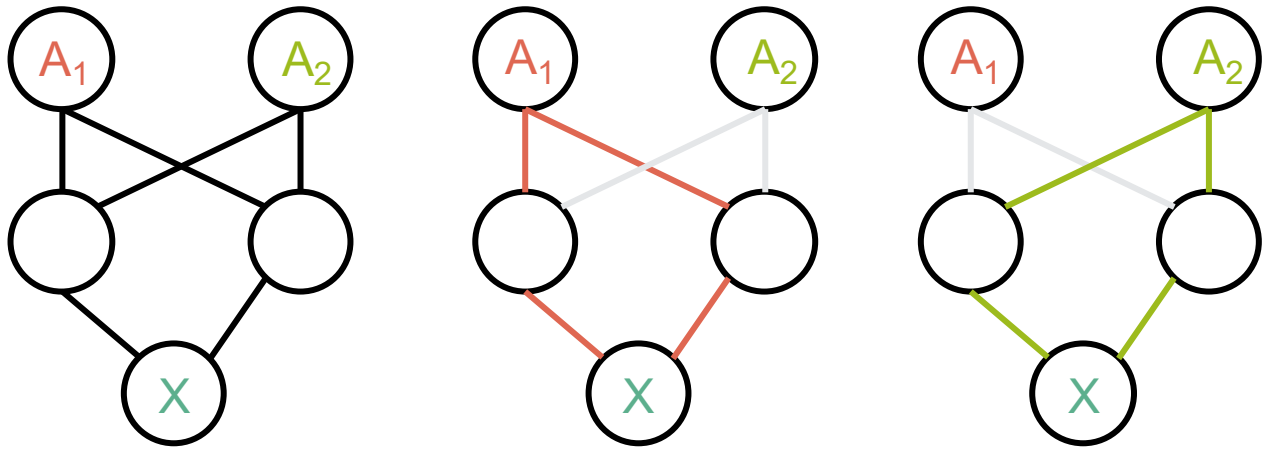


YOUR TURN

What is the inbreeding coefficient for individual X assuming individual A is not inbred ($F_A = 0$)?



WHEN THERE ARE MULTIPLE ANCESTORS



Follow the transmission of alleles over multiple loops.

$$F_X = \sum_{loops} \left(\frac{1}{2}\right)^n (1 + F_A)$$

INBREEDING

CHANGES GENOTYPE FREQUENCIES

If the population is in HW proportions

Genotype	AA	Aa	aa
Frequency	p^2	$2pq$	q^2

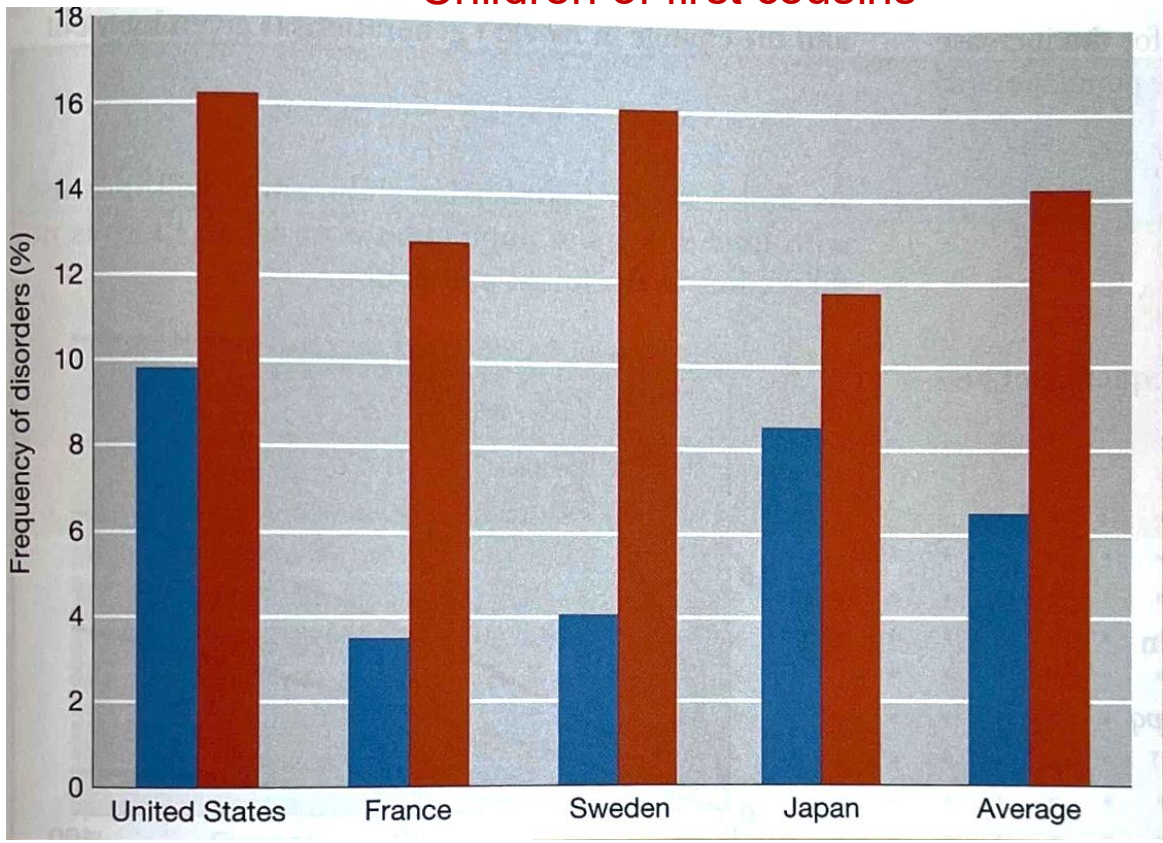
If there is inbreeding

Genotype	AA	Aa	aa
Frequency	p^2+pqF	$2pq-2pqF$	q^2+pqF

Results in excess in homozygotes

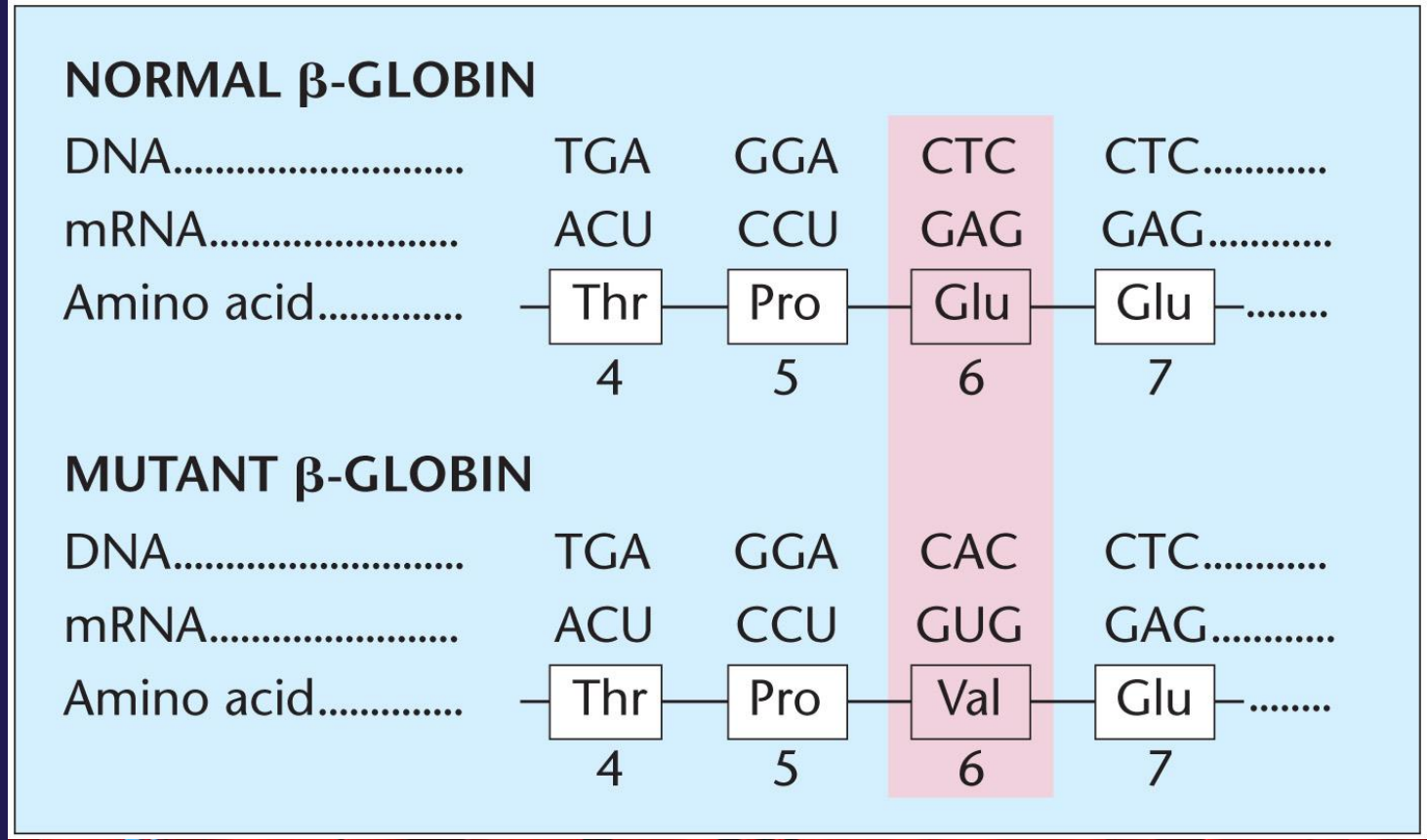
Children of unrelated parents

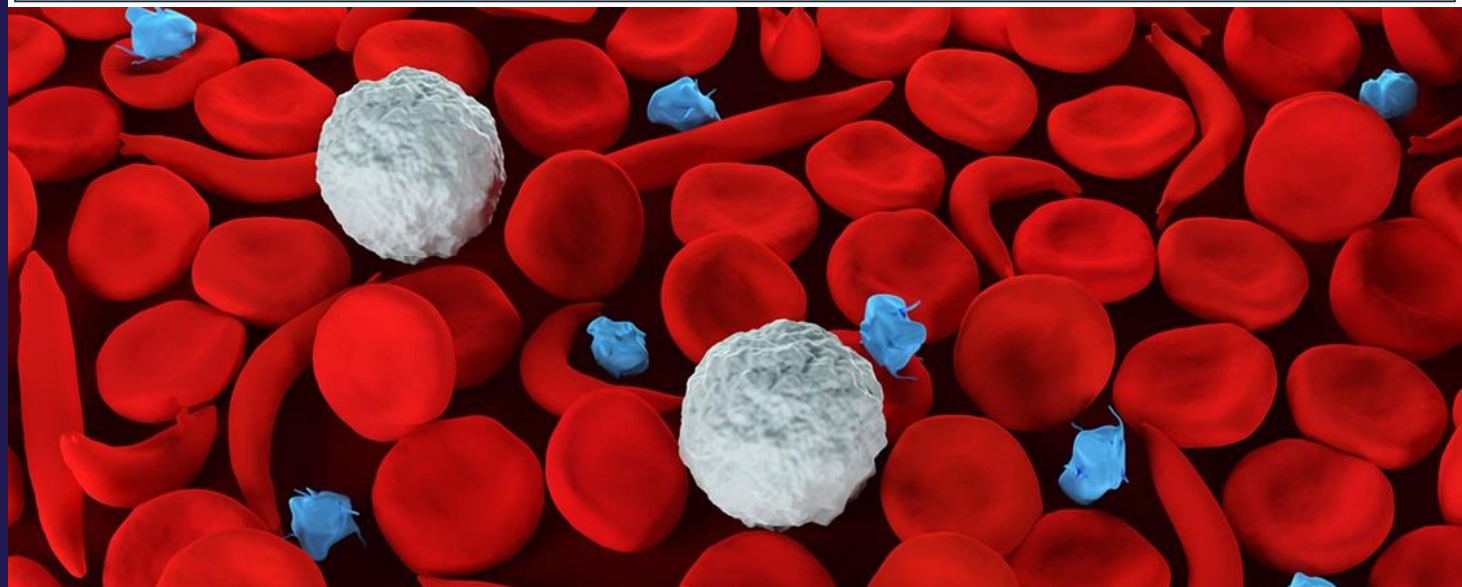
Children of first cousins



THE NEUTRAL POPULATION

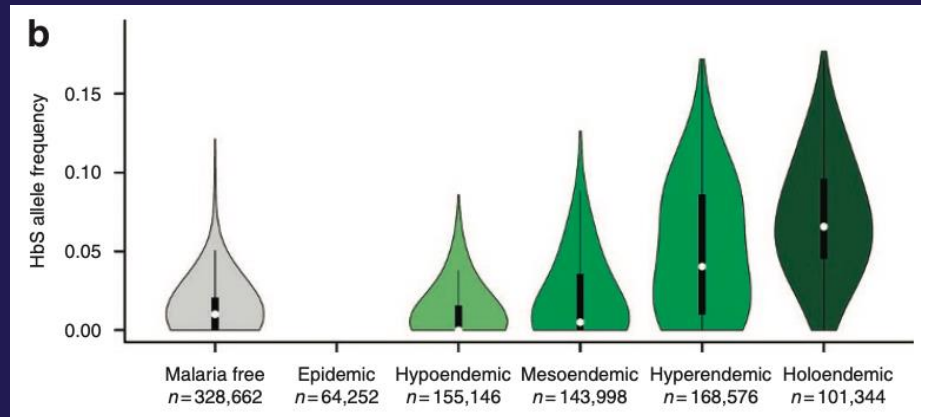
- Random mating
- No selection
- No genetic drift (infinite population size)
- No migration
- **No mutation**

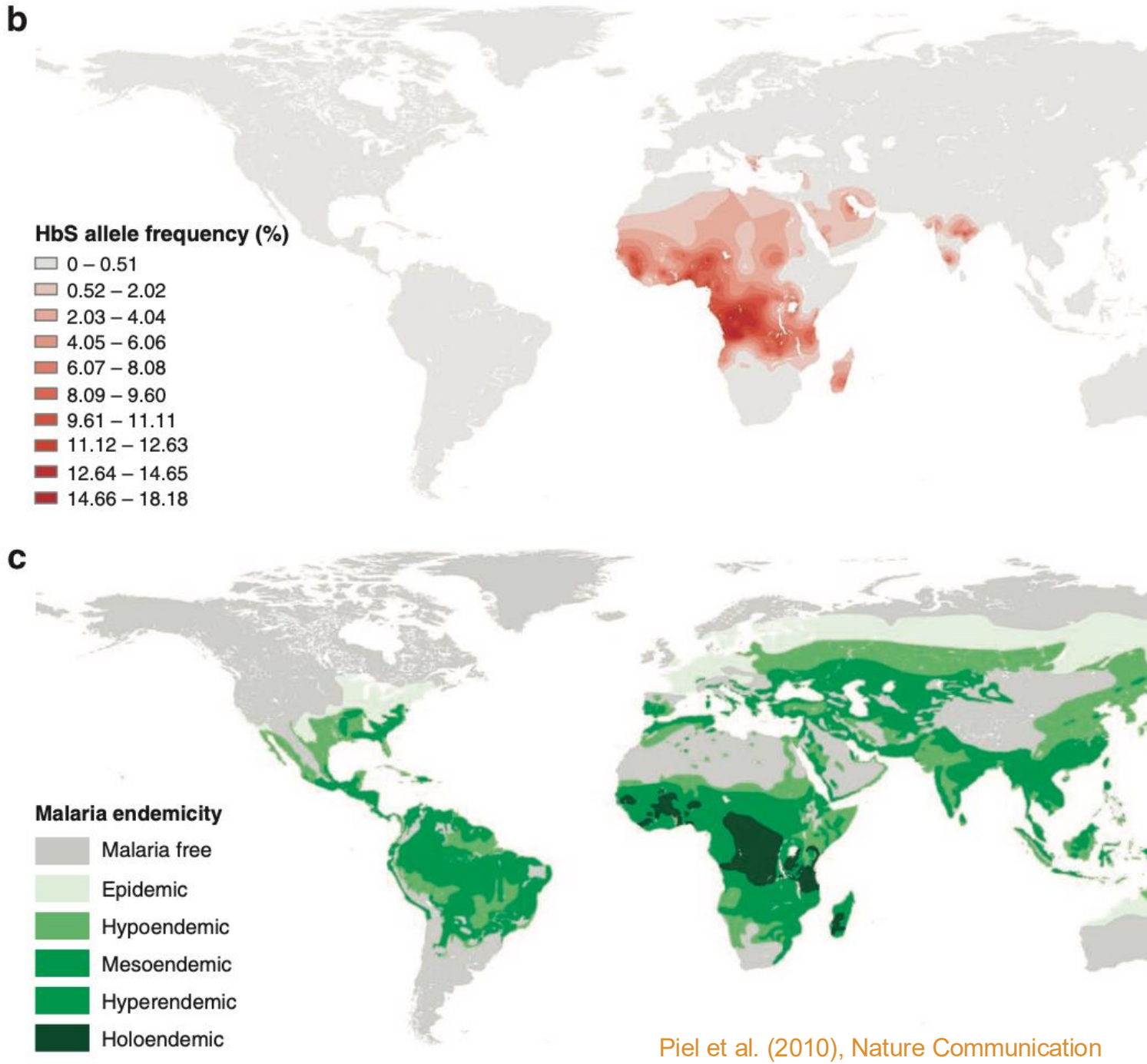




THE NEUTRAL POPULATION

- Random mating
- **No selection**
- No genetic drift (infinite population size)
- No migration
- No mutation





Piel et al. (2010), Nature Communication

MUTATION AND SELECTION $a^+ \xrightarrow{\mu} a$

Number wildtype alleles in a population of *2N* is $2Np$, which with the raten μ mutates to harmfull allele.

In the next generation the proportion of new harmfull alleles are: $\Delta q_\mu = 2Np\mu$

Recessive harmfull

$$\Delta q_\mu = 2Nsq^2$$

$$q = \sqrt{\frac{\mu}{s}}$$

Genotype	a^+a^+	a^+a	**aa**
Fitness	*1*	*1*	*1-s*

Dominant harmfull

$$\Delta q_\mu = Ns2pq + 2Nsq^2$$

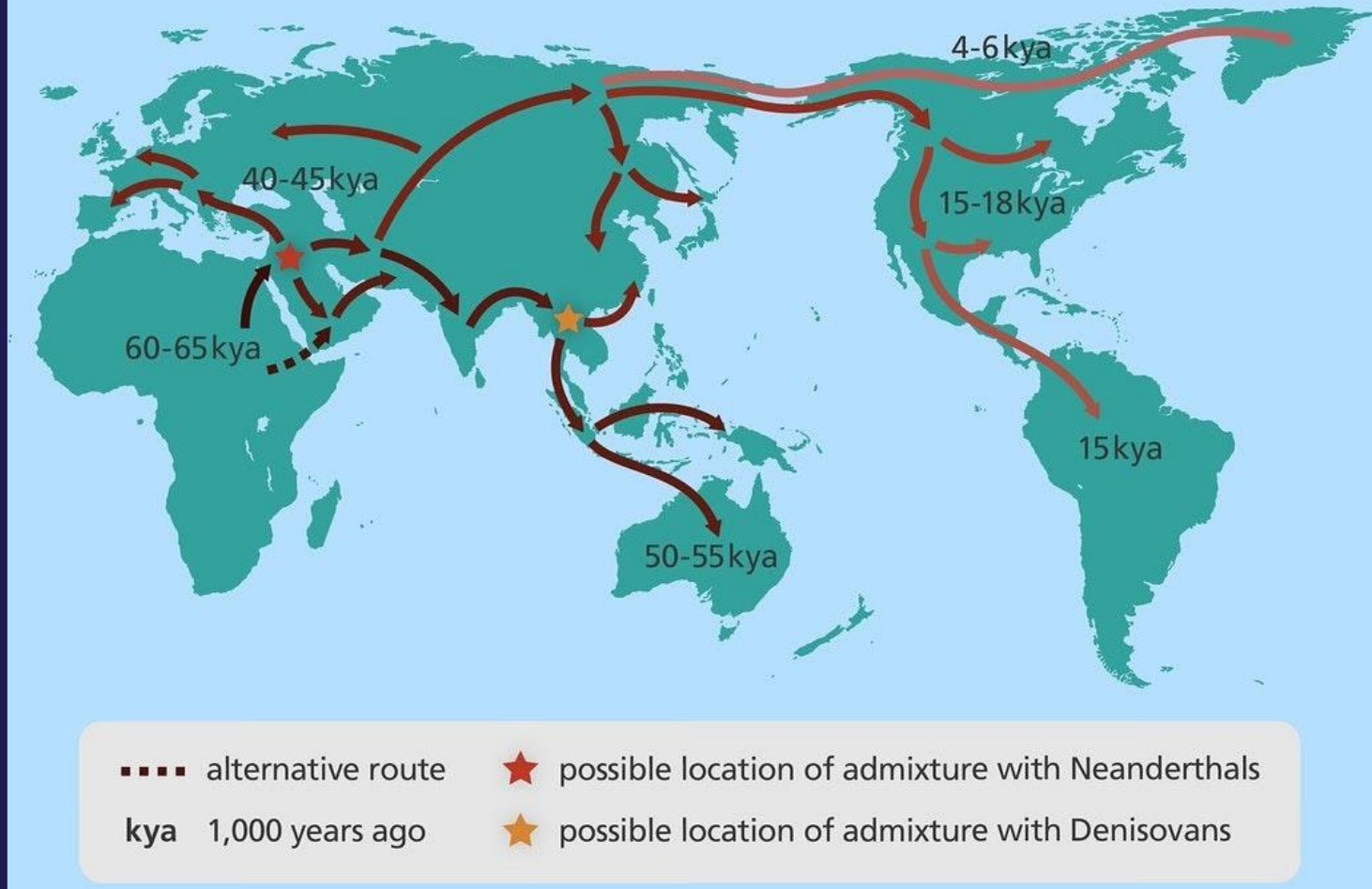
$$q = \frac{\mu}{s}$$

Genotype	a^+a^+	a^+a	**aa**
Fitness	*1*	*1-s*	*1-s*

THE NEUTRAL POPULATION

- Random mating
- No selection
- No genetic drift (infinite population size)
- **No migration**
- No mutation

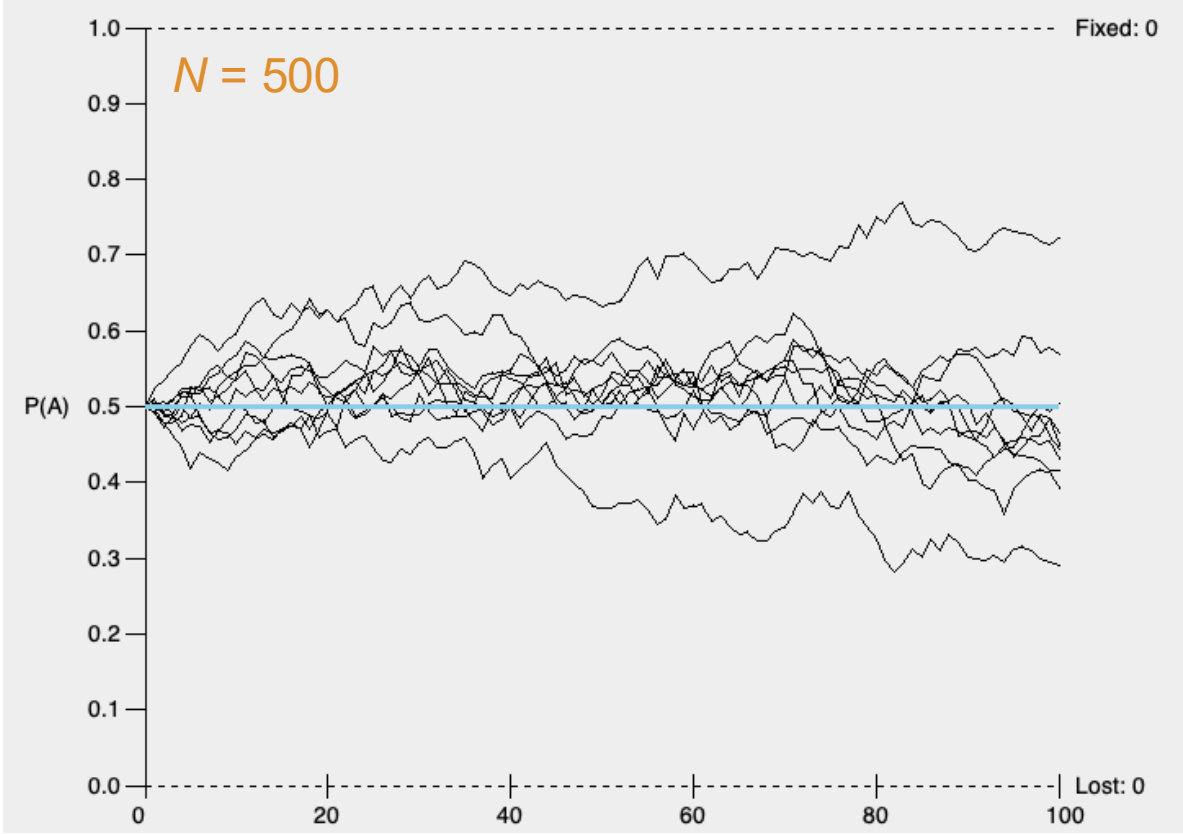
$$q_1 = mq_m + (1 - m)q_1$$



THE NEUTRAL POPULATION

- Random mating
- No selection
- **No genetic drift** (infinite population size)
- No migration
- No mutation

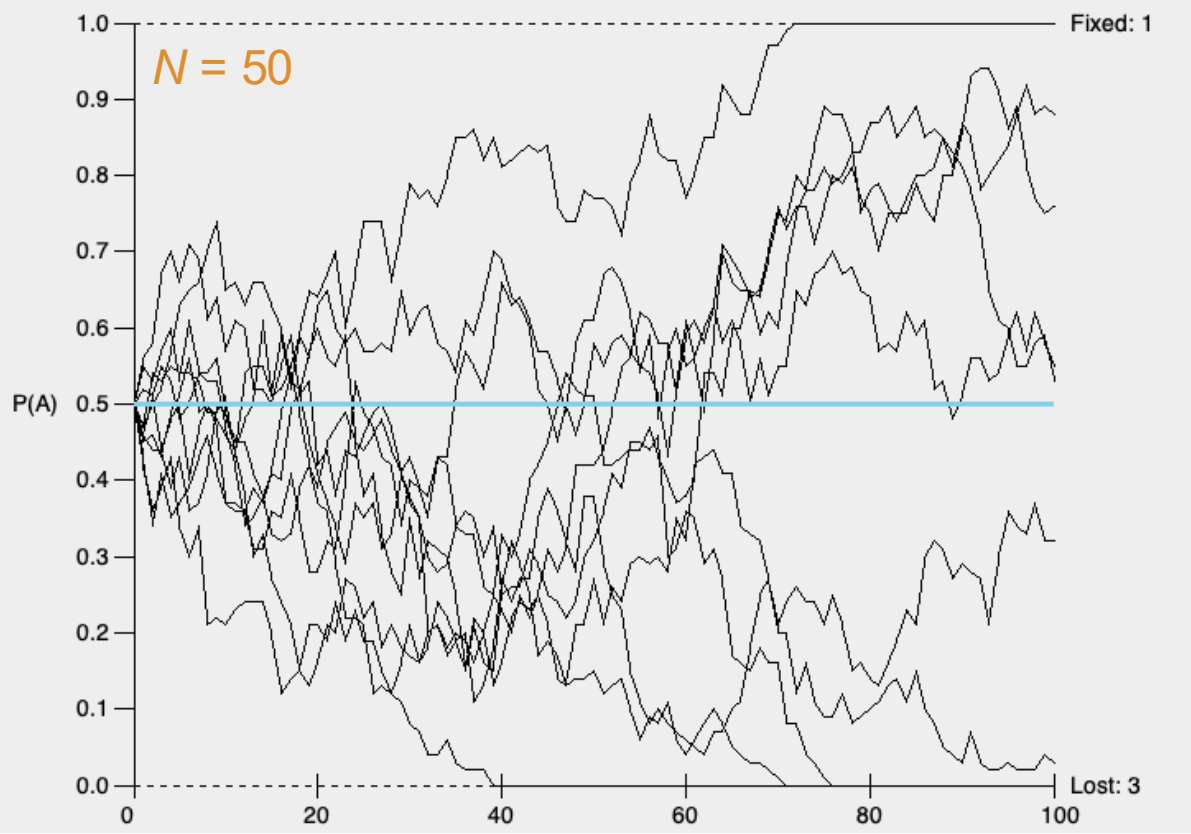
Genetic drift is changes in allele frequencies between generations due to sampling error



THE NEUTRAL POPULATION

- Random mating
- No selection
- **No genetic drift** (infinite population size)
- No migration
- No mutation

Genetic drift is changes in allele frequencies between generations due to sampling error



GENETIC DRIFT AND INBREEDING

Genetic drift entails loci in a sub-population becomes fixed, thus, the degree of homozygosity increases (thus, *F* increase).

The probability of selecting two gametes carrying the same allele is *1/(2N).*

The degree of inbreeding increase with time

$$F_t = 1 - (1 - \frac{1}{2N})^t$$

The rate of loss of heterozygosity (H) per generation

$H_t = (1 - \frac{1}{2N})^t H_0$, the rate depend on N

If there is inbreeding

Genotype	AA	Aa	aa
Frequency	p^2+pqF	2pq-2pqF	q^2+pqF

Results in excess in homozygotes

MODULATION OF FREQUENCIES

Mutation
introduces new alleles
diversity within populations ↑

Migration
introduces new alleles
diversity within populations ↑
diversity between populations ↓

Genetic drift
loss of alleles
diversity within populations ↓
diversity between populations ↑

Selection
removes harmfull alles
diversity within populations ↓
diversity between populations ↓↑

Non-random mating
do not change alleles, but change genotype frequencies

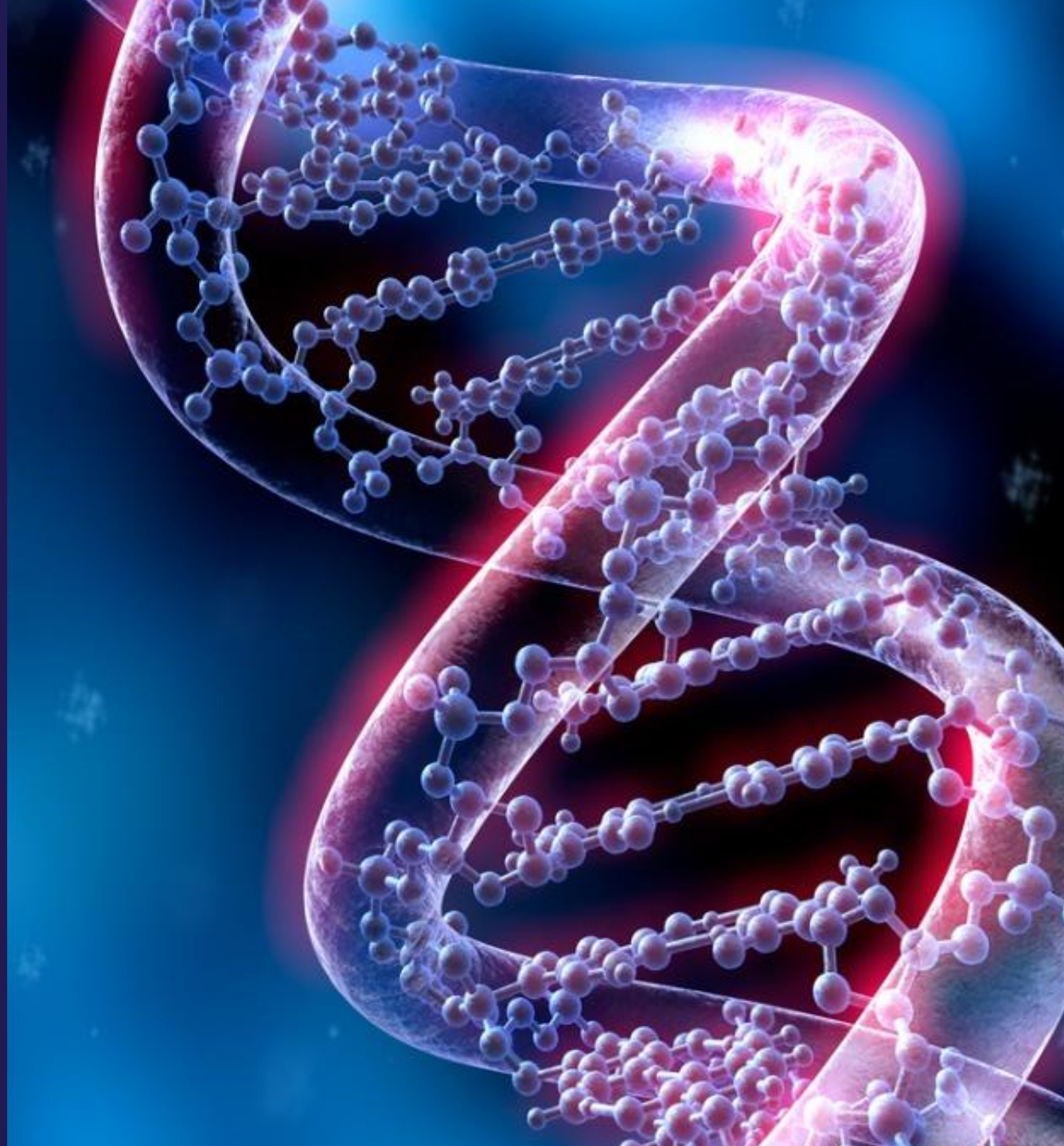
OUTLINE

08:15 – 09:15	Recap + Exercises E15 [Part III]
09:15 – 09:30	Break
09:30 – 09:50	Lecture 1 [*Genetic risk assessment]*
09:50 – 10:30	Group work
10:30 – 11:10	Break + Exercises I [E1-E3]
11:10 – 11:40	Lecture 2 [*Linkage*]
11:40 – 11:55	Exercises II [E4-E8]
11:55 – 12:00	Reflection

OUTLINE

08:15 – 09:15	Recap + Exercises E15 [Part III]
09:15 – 09:30	Break
09:30 – 09:50	Lecture 1 [*Genetic risk assessment]*
09:50 – 10:30	Group work
10:30 – 11:10	Break + Exercises I [E1-E3]
11:10 – 11:40	Lecture 2 [*Linkage*]
11:40 – 11:55	Exercises II [E4-E8]
11:55 – 12:00	Reflection

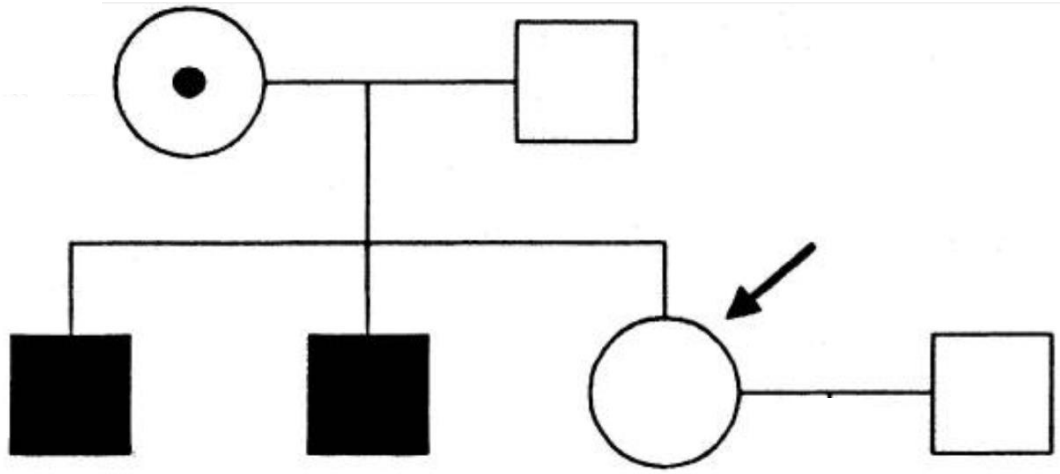
MONOGENIC RISK ASSESSMENT



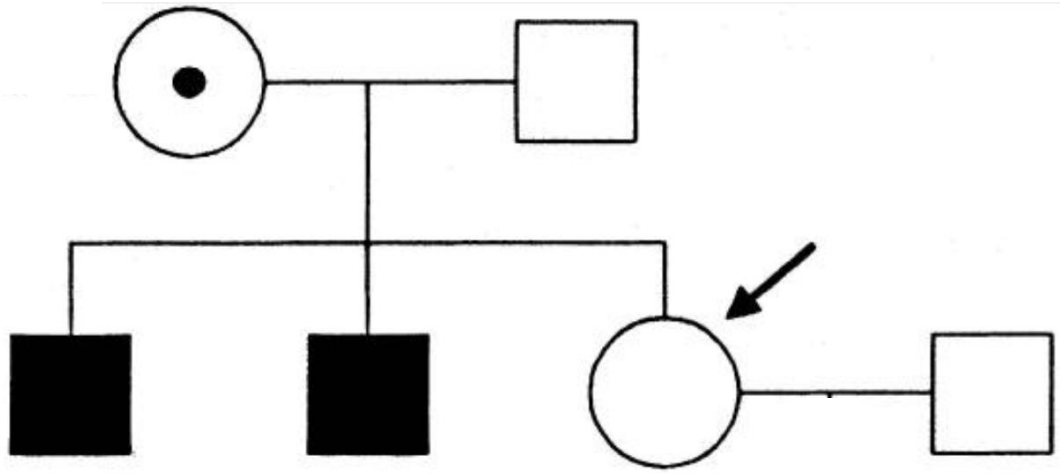
AALBORG UNIVERSITY

MONOGENIC INHERITANCE

What type of inheritance is seen in the pedigree?



MONOGENIC INHERITANCE



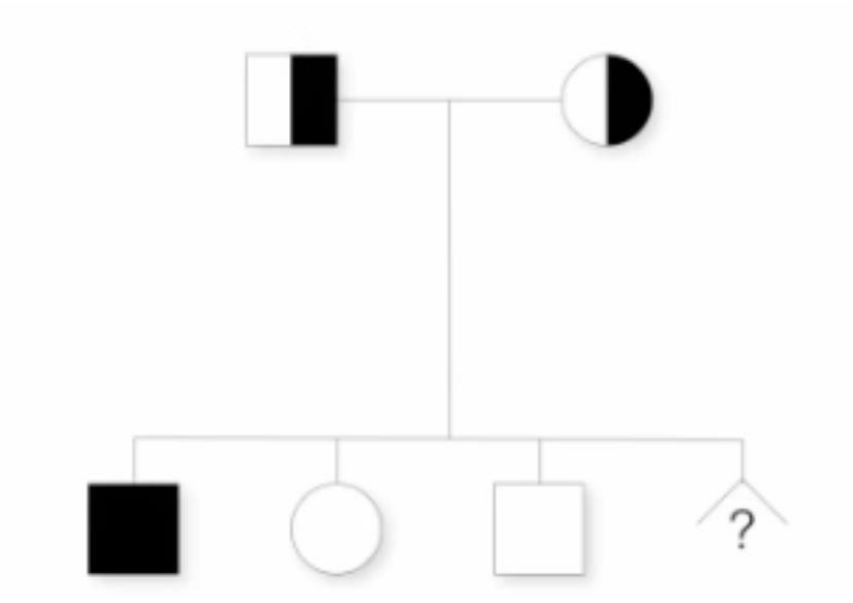
What type of inheritance is seen in the pedigree?

What is the probabiity that II.3 is a carrier?

AUTOSOMAL RECESSIVE I

Both parents must be carriers (Aa) to get an affected child.

Their risk of getting a fourth affected child is = ¼ [draw punnet square]



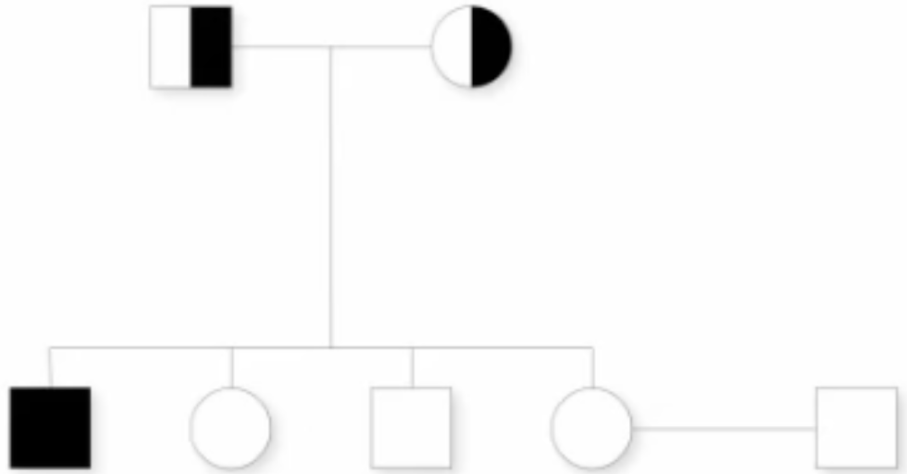
AUTOSOMAL RECESSIVE II

The risk of II.4 being a carrier must be 2/3 [we know that she is not affected, thus se cannot be aa].

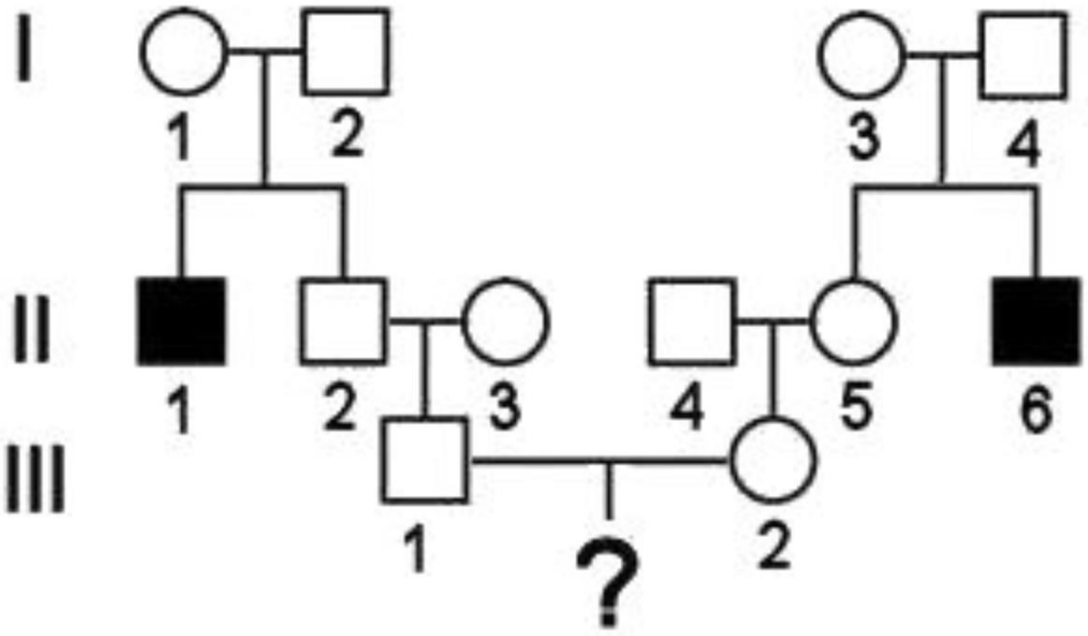
The risk of II.5 of being a carrier (given no family history of the disease) is the population risk.

For an AR the population frequency could be 1/25.

The risk the couple will get an affected child is then:
½*2/3 *1/25*1/2= 1/150



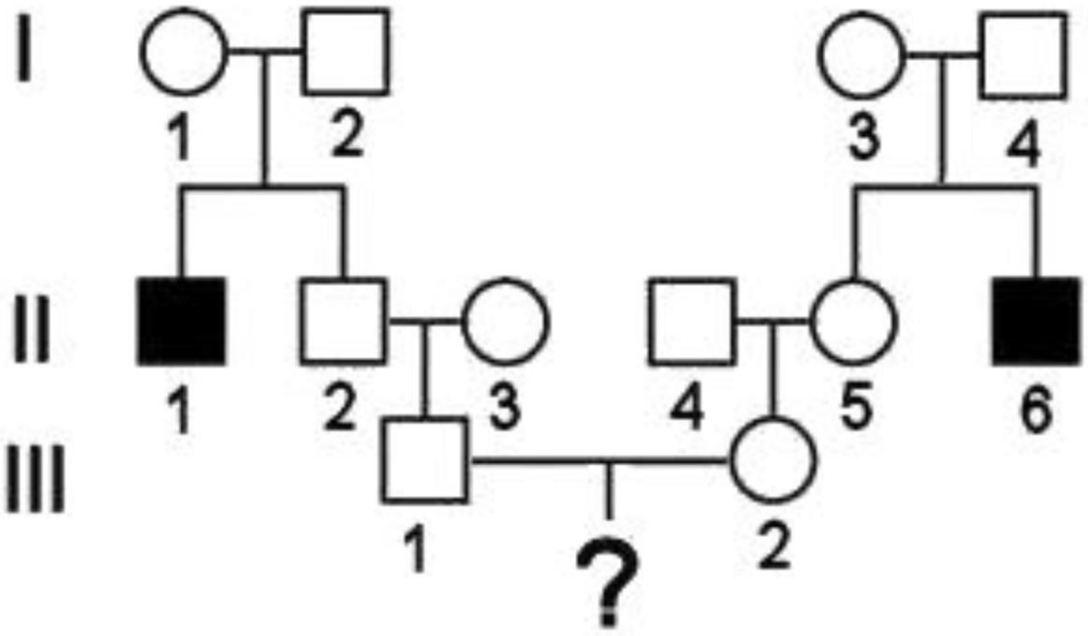
AUTOSOMAL RECESSIVE III



What is the probability that IV.1 is affected (aa)?

1. IV.1 must inherit an a-allele from III.1 and III.2
2. II.1 has the genotype aa, thus I.1 and I.2 most both have the genotype Aa.
3. II.2 has the dominant phenotype, thus he must have at least one A. The probability that the other is a, is 2/3 (he is not affected).
4. II.3 is from outside the family, thus we assume she is AA.
5. III.1 has the dominant phenotype (A-). The probaility that he is Aa is the probability that II.2 is Aa and passes a to his son, ½ * 2/3 = 1/3
6. The probability that III.2 is Aa is ½ * 2/3 = 1/3
7. The probability that IV.1 is aa ¼ * 1/3 * 1/3 = 1/36

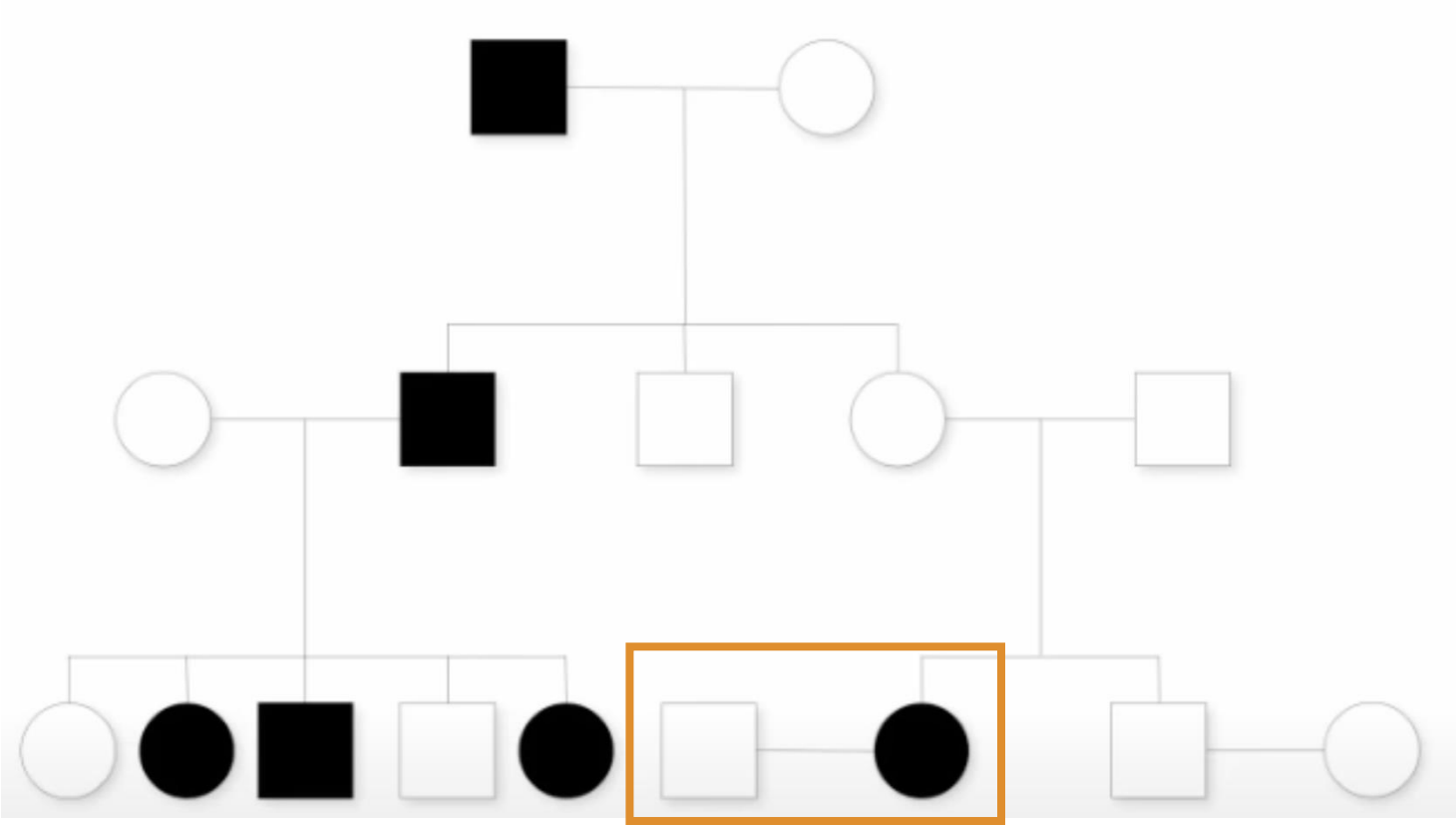
AUTOSOMAL RECESSIVE IV



What is the probability that IV.1 is a carrier (Aa)?

1. The probaility that III.1 is Aa ½ * 2/3 = 1/3
2. The probability that III.2 is Aa is ½ * 2/3 = 1/3
3. The probability that IV.1 is Aa 2/4 *1/3 *1/3 = 1/18

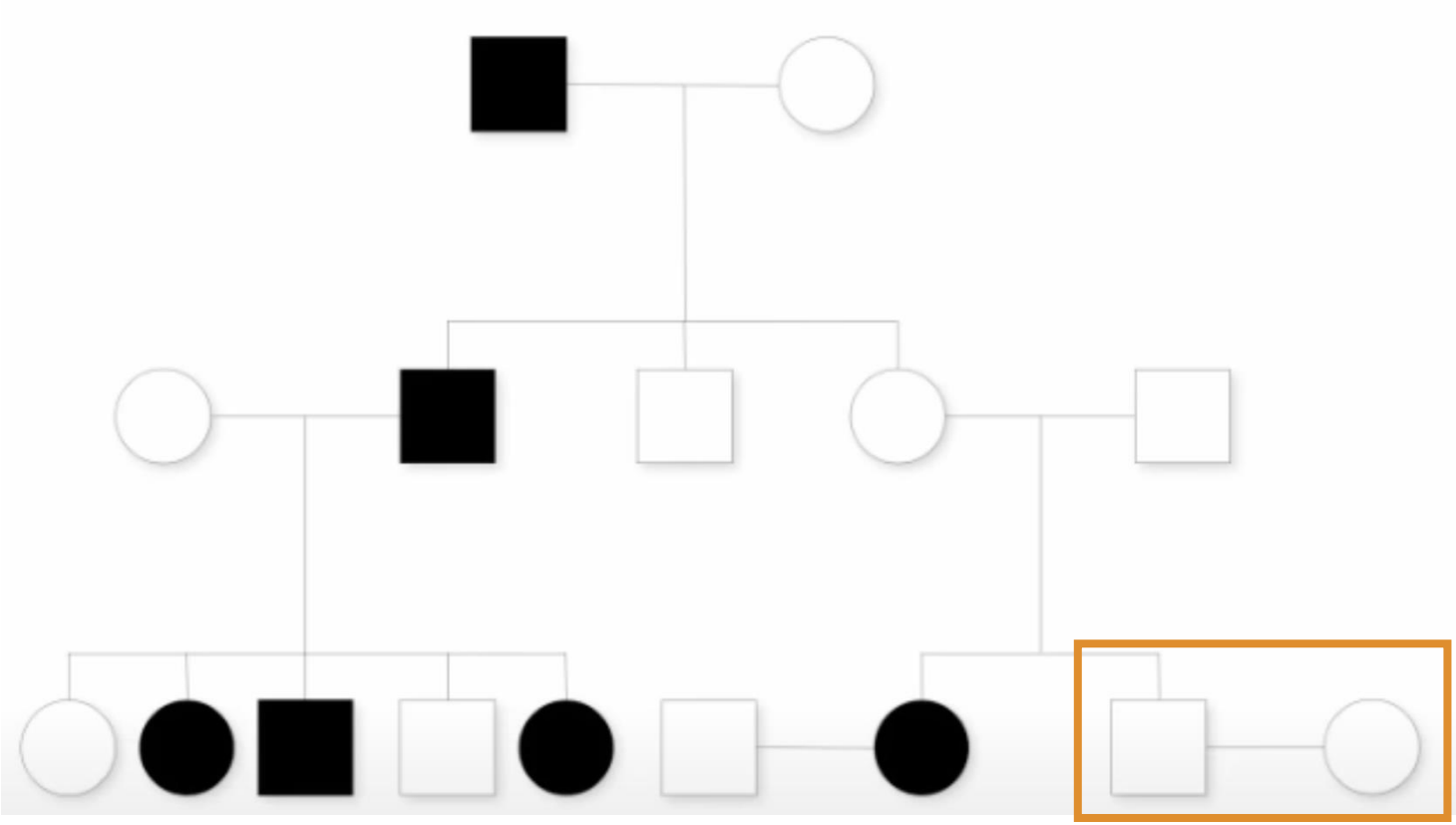
RISK WITH INCOMPLETE PENETRANCE



Assuming 80% penetrance
→ *80% probability that an individual that inherrent the mutation will show the phenotype*

This couples risk of getting an affected child:
½ * 0.8 = 0.4

RISK WITH INCOMPLETE PENETRANCE

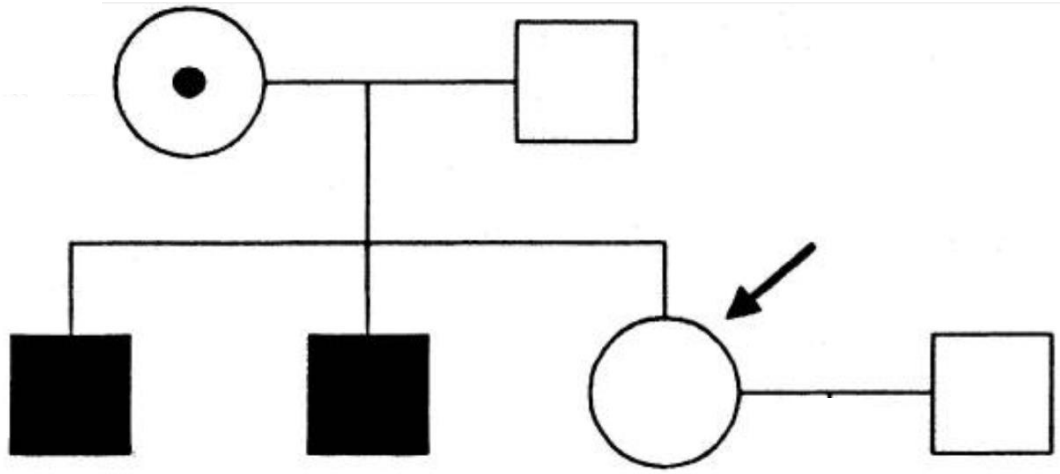


Assuming 80% penetrance
→ *80% probability that an individual that inherrent the mutation will show the phenotype*

This couples risk of getting an affected child:
½*0.2*1/2*0.8=0.04

Fathers risk of being a carrier
Childs risk of being affected

MONOGENIC INHERITANCE
WITH ADDITIONAL INFORMATION



What type of inheritance is seen in the pedigree?

What is the probabiity that II.3 is a carrier?

ıat is the probabiity that II.3 is a carrier now?

BAYES' THEOREM

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

The probability of B given that A is true.
→ Likelihood of A given a fixed B

The probabilities of observing A and B, respectively without any conditions.
→ prior probability

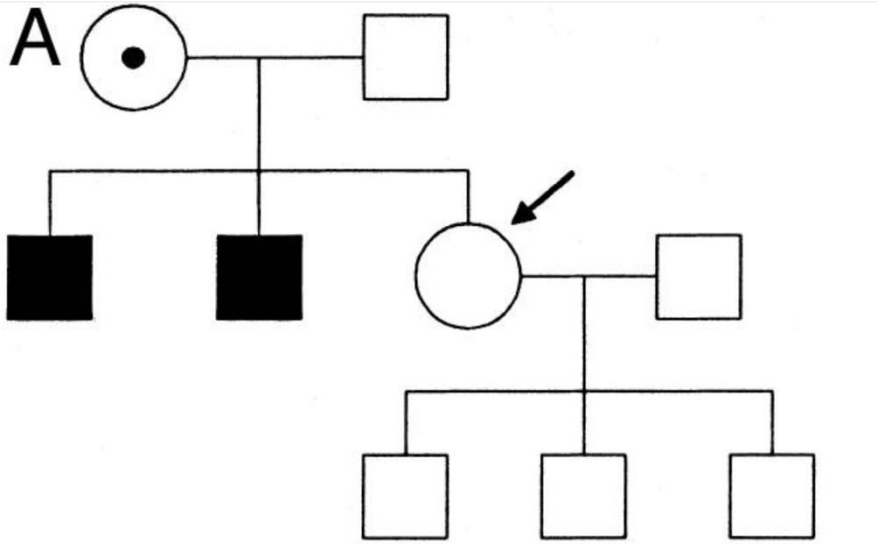
The probability of A given that B is true.
→ The posterior probability of A given B

BAYES' THEOREM
IN GENETICS

Hypothesis	H: Is a carrier	H: Is not a carrier
Prior probability	x1	x2
Conditional probability	y1	y2
Joint probability	x1*y1	X2*y2
Posterior probability	j.prob1 / (j.prob1+j.prob2)	j.prob2 / (j.prob1+j.prob2)

Used when additional information becomes available.

BAYES' THEOREM
IN GENETICS



What is the probabiity that II.3 is a carrier now?

Hypothesis	H: Is a carrier	H: Is not a carrier
Prior probability	1/2	1/2
Conditional probability (three normal sons)	½^3=1/8	1
Joint probability	1/16	1/2
Posterior probability	(1/16)/(1/16+1/2) =1/9	(1/2)/(1/16+1/2) =8/9

OUTLINE

08:15 – 09:15 Recap + Exercises E15 [Part III]

09:15 – 09:30 Break

09:30 – 09:50 Lecture 1 [*Genetic risk assessment]*

09:50 – 10:30 Group work

10:30 – 11:10 Break + Exercises I [E1-E3]

11:10 – 11:40 Lecture 2 [*Linkage*]

11:40 – 11:55 Exercises II [E4-E8]

11:55 – 12:00 Reflection

Make 4 groups

Find Group-exercise on Github

Group 1 and 3 works with '*Bayesian Analysis Using Pedigree Information*'

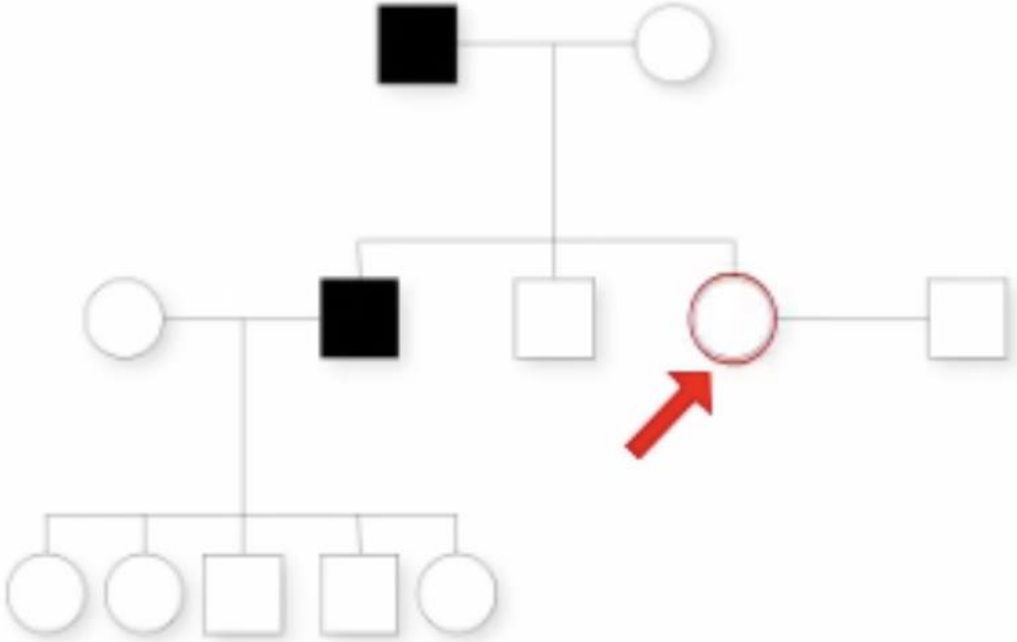
Group 2 and 4 works with '*Bayesian Analysis Using Genetic Test Results*'

25 min to read and understand your example

15 min to explain example to new group

__Assuming__ that at the age of 30, 70% of individuals with the mutation will display the phenotype

AGE-DEPENDENT PENETRANCE

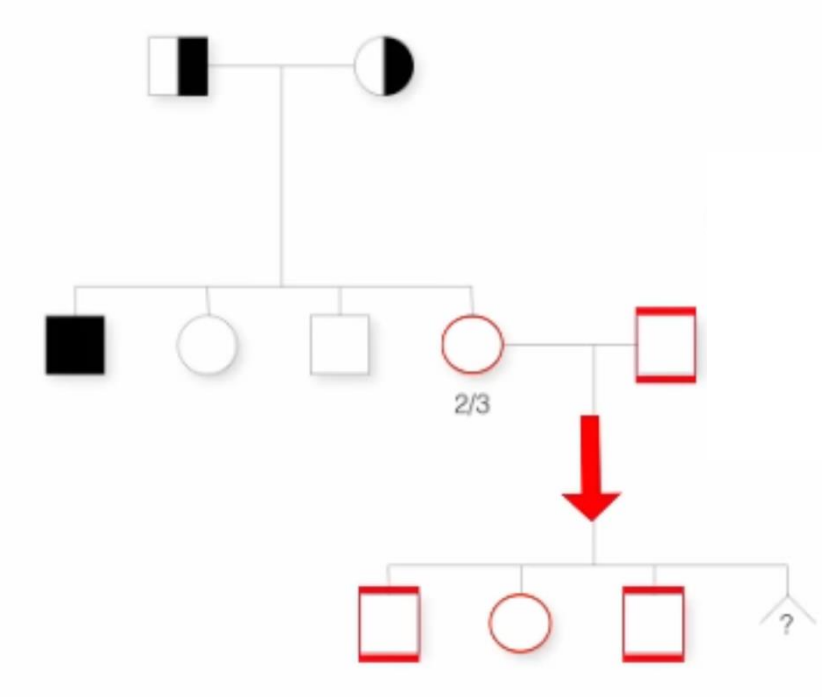


Hypothesis	H: Is a carrier	H: Is not a carrier
Prior probability	0.5	0.5
Conditional probability (unaffected at age 30)	0.7	1
Joint probability	0.35	0.5
Posterior probability	$\frac{0.35}{0.35+0.5} = 0.41$	$\frac{0.5}{0.35+0.5} = 0.59$

AUTOSOMAL RECESSIVE

New evidence, the couple has three unaffected children.

Hypothesis	Couple at risk	Couple not at risk
Prior probability	2/3*1/25=0.026	1-0.026=0.974
Conditional probability (three unaffected kids)	¾^3=0.42	1
Joint probability	0.01	0.974
Posterior probability	$\frac{0.01}{0.01 + 0.974}$	$\frac{0.974}{0.01 + 0.974}$



Probability that they are at risk (vs not at risk) and having three unaffected kids

OUTLINE

08:15 – 09:15 Recap + Exercises E15 [Part III]

09:15 – 09:30 Break

09:30 – 09:50 Lecture 1 [*Genetic risk assessment]*

09:50 – 10:30 Group work

10:30 – 11:10 Break + Exercises I [E1-E3]

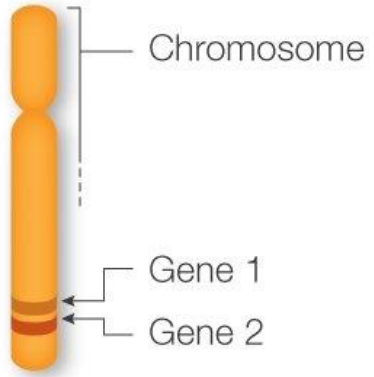
11:10 – 11:40 Lecture 2 [*Linkage*]

11:40 – 11:55 Exercises II [E4-E8]

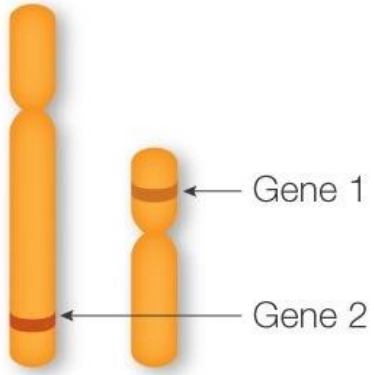
11:55 – 12:00 Reflection

LINKAGE

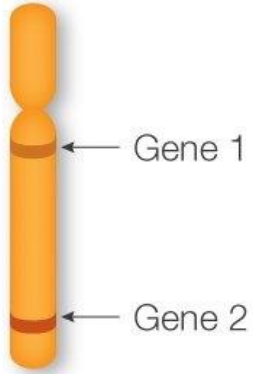
When alleles travel together



Linked

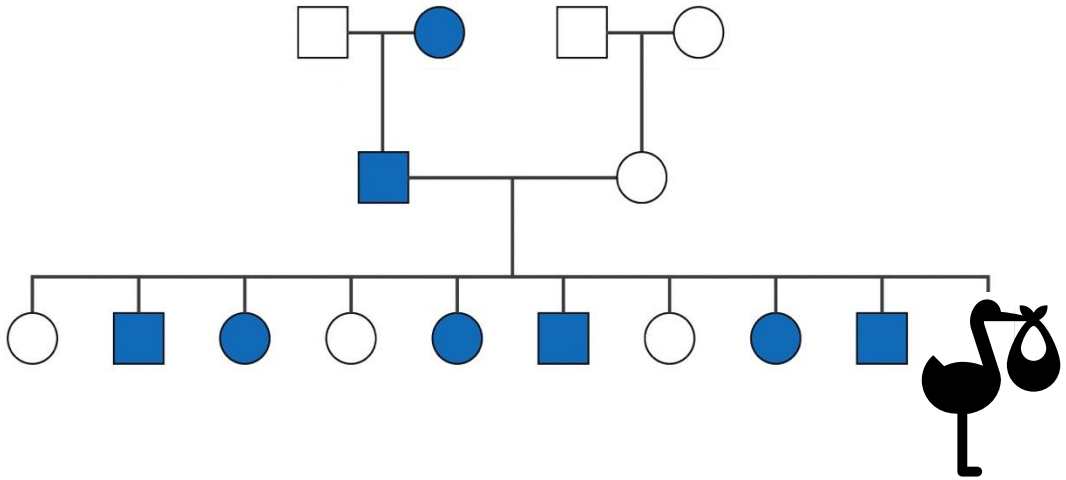


Not Linked



Not Linked

OVERALL WE AIM TO



We need to understand how variants segregate in families first

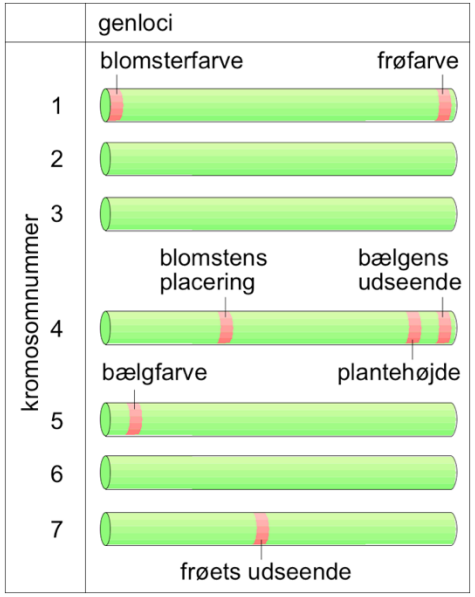
- Carrier status / prenatal testing
- Prognosis
- Guided treatment
- Genetic counselling - *you can help even without knowing the mutation*

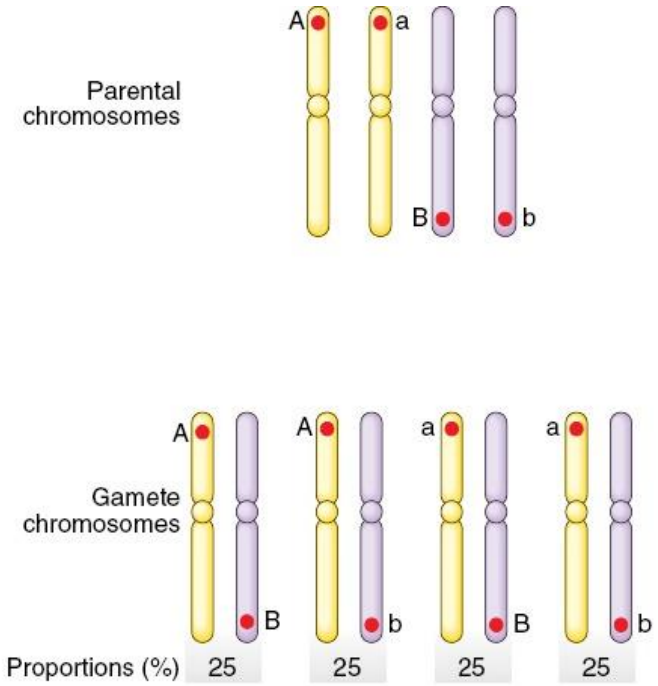
INDEPENDENT ASSORTMENT

Mendels 2. law

Alleles at different loci segregate independently during meiosis.

Only true for independent loci



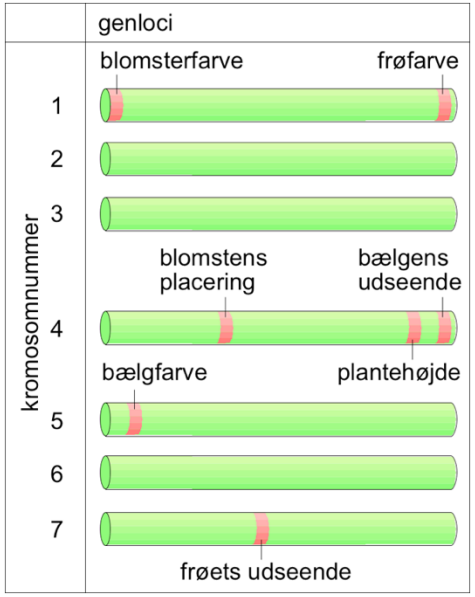


INDEPENDENT ASSORTMENT

Mendels 2. law

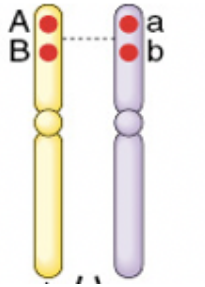
Alleles at different loci segregate independently during meiosis.

Only true for independent loci.

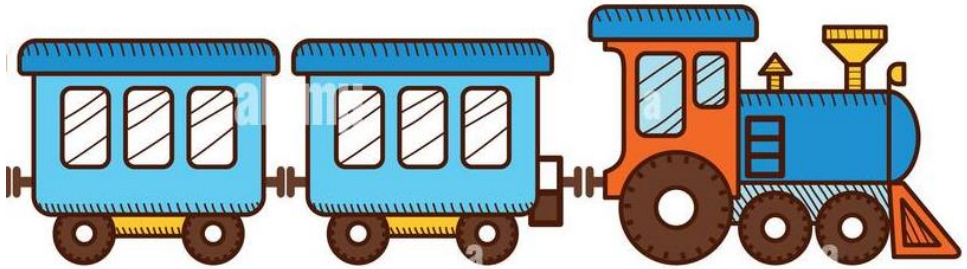


If – *in contrast* – loci are close, alleles do no longer segregate independently.

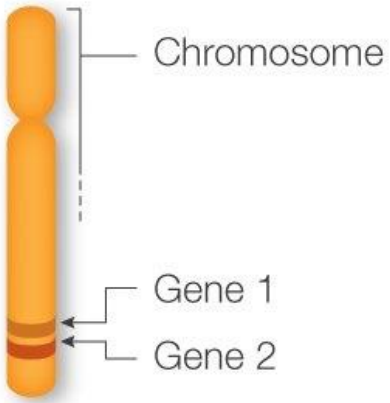
When this happens – we say the loci are ***linked***



LINKAGE

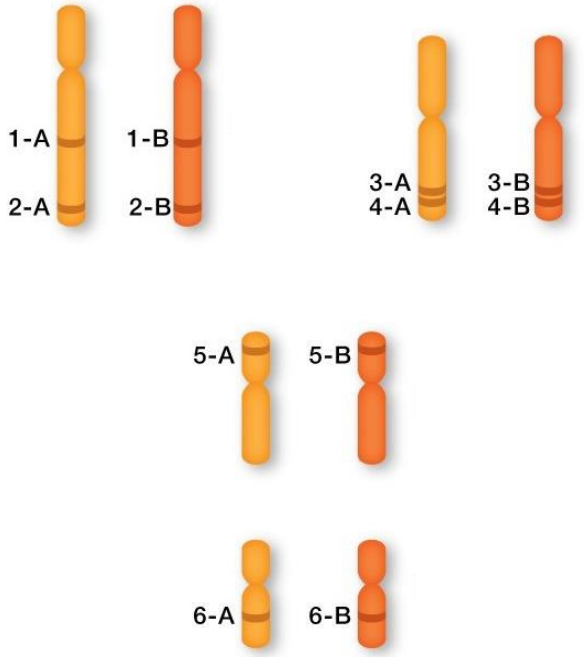


Linked train wagons



Linked loci
(Physical proximity)

LINKED LOCI



Two **loci** are linked when the **alleles** segregate together more often than by chance

Your turn

Linked or unlinked?

Gene 1 and Gene 2

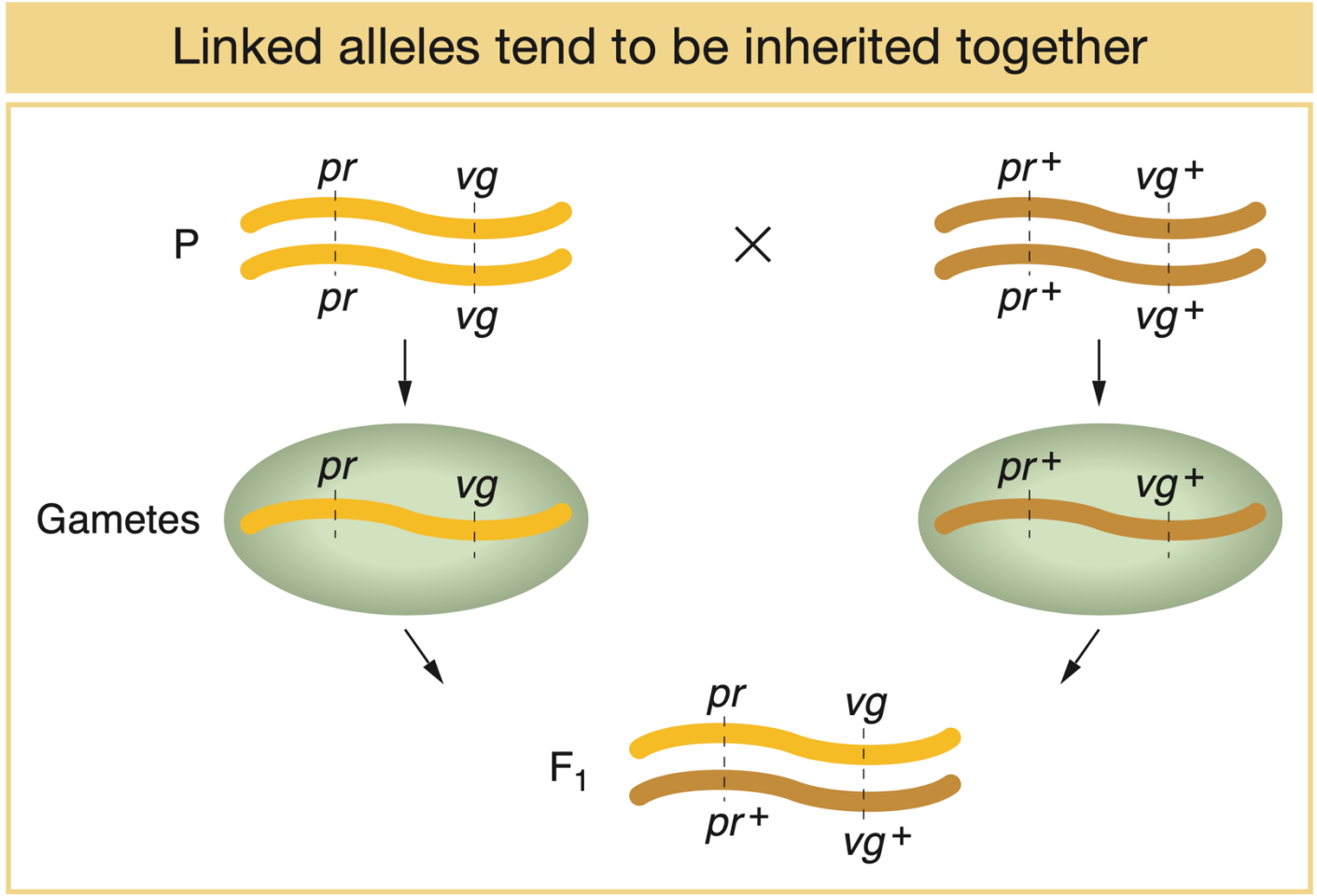
Gene 3 and Gene 4

Gene 5 and Gene 6

What about the other combinations?

LINKED GENES
DURING MEIOSIS

HAPLOTYPE = haploid genotype combination of genetic information on a single chromosome.



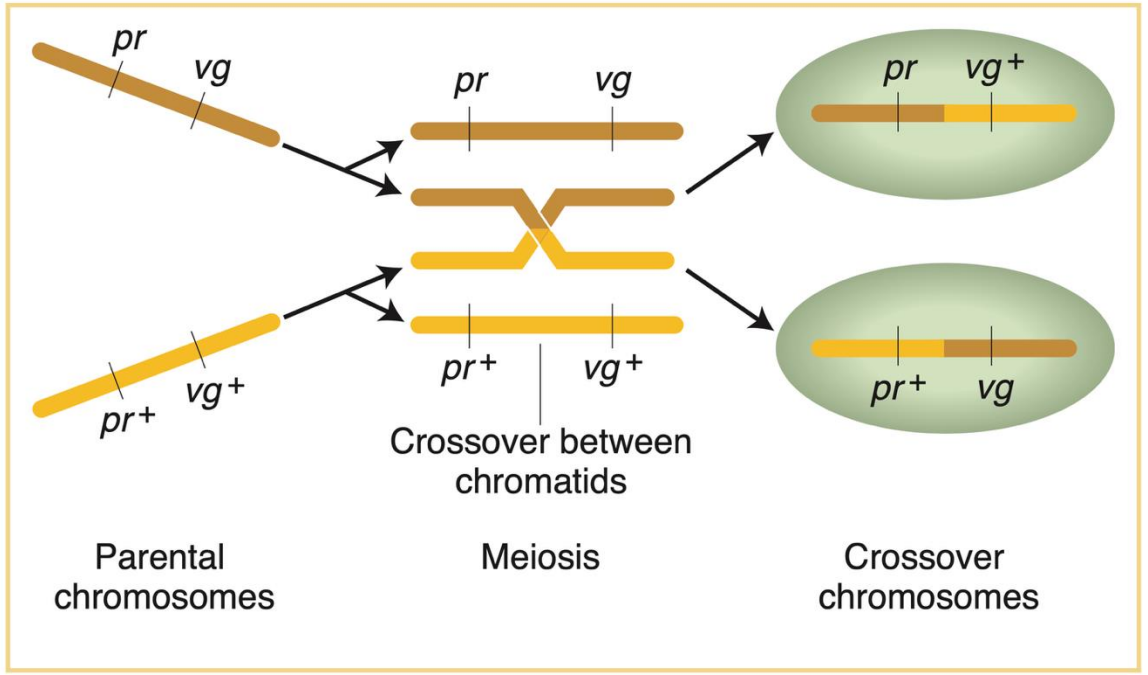
LINKED GENES
DURING MEIOSIS

$$\frac{pr^{+} \quad vg^{+}}{pr \quad vg}$$

Genes segregate independently if they are on different chromosomes, but can be linked if they are on the same chromosome

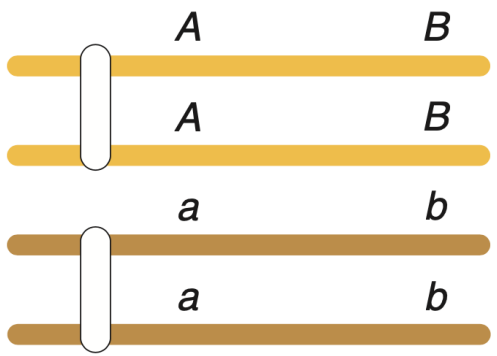
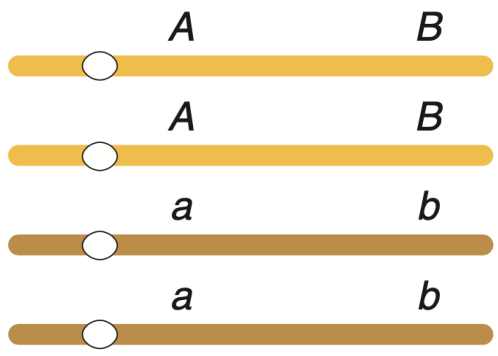
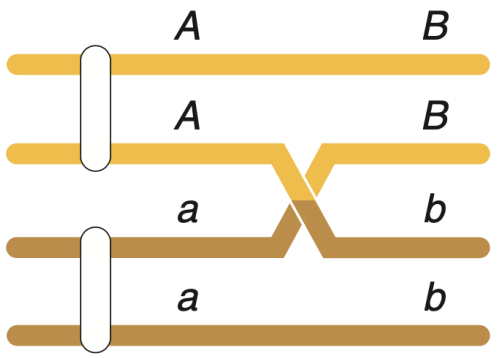
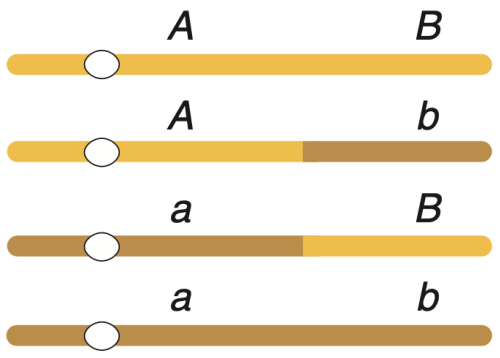
At complete linkage only parental gametes are seen (non-crossover; NCO).

If crossover happens between 2 (or more) genes both parental and recombinant gametes are seen.



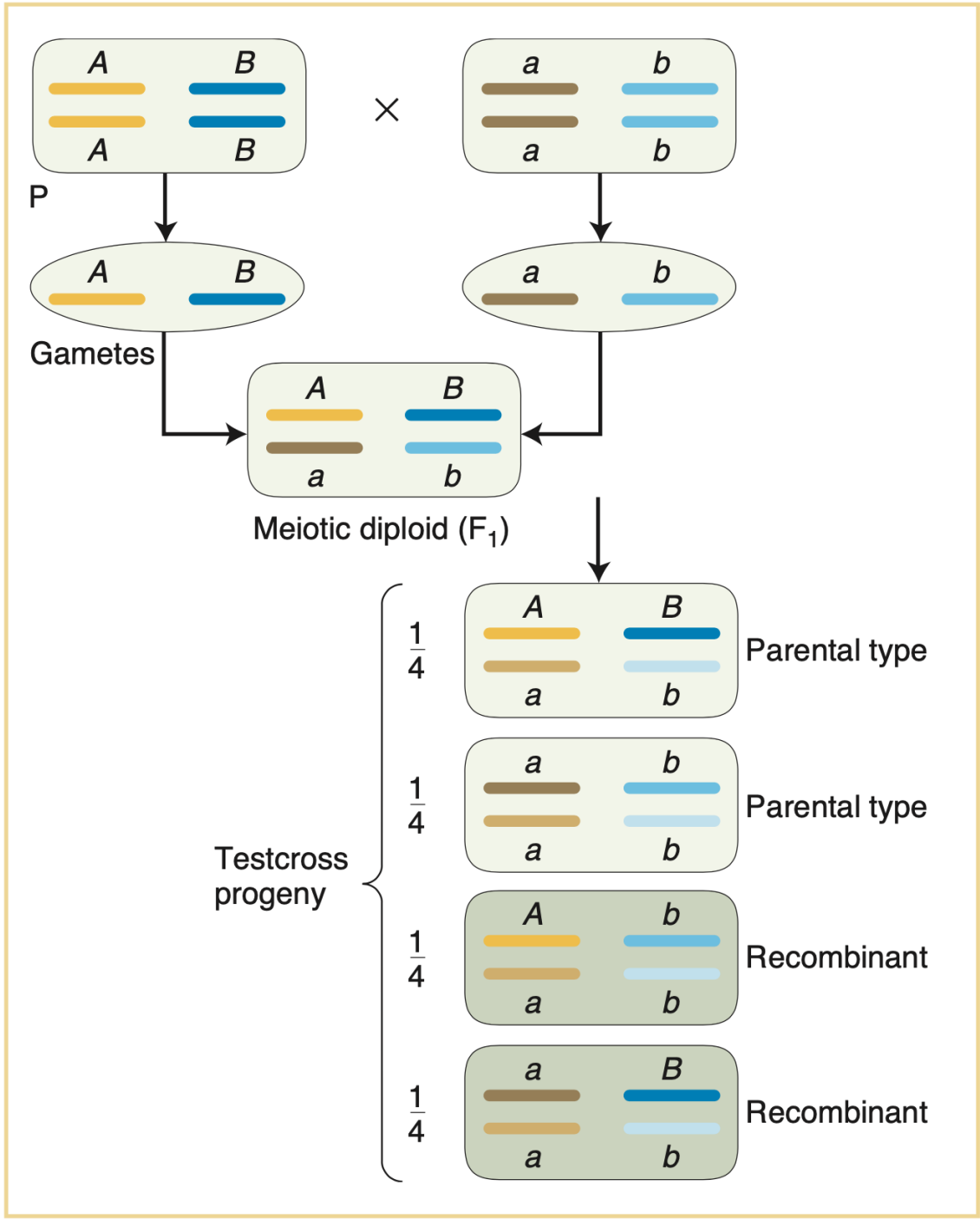
CROSSOVERS

BETWEEN NON-SISTER CHROMATIDES

	Meiotic chromosomes	Meiotic products	
Meioses with no crossover between the genes			Parental Parental Parental Parental
Meioses with a crossover between the genes			Parental Recombinant Recombinant Parental

FREE RECOMBINATION

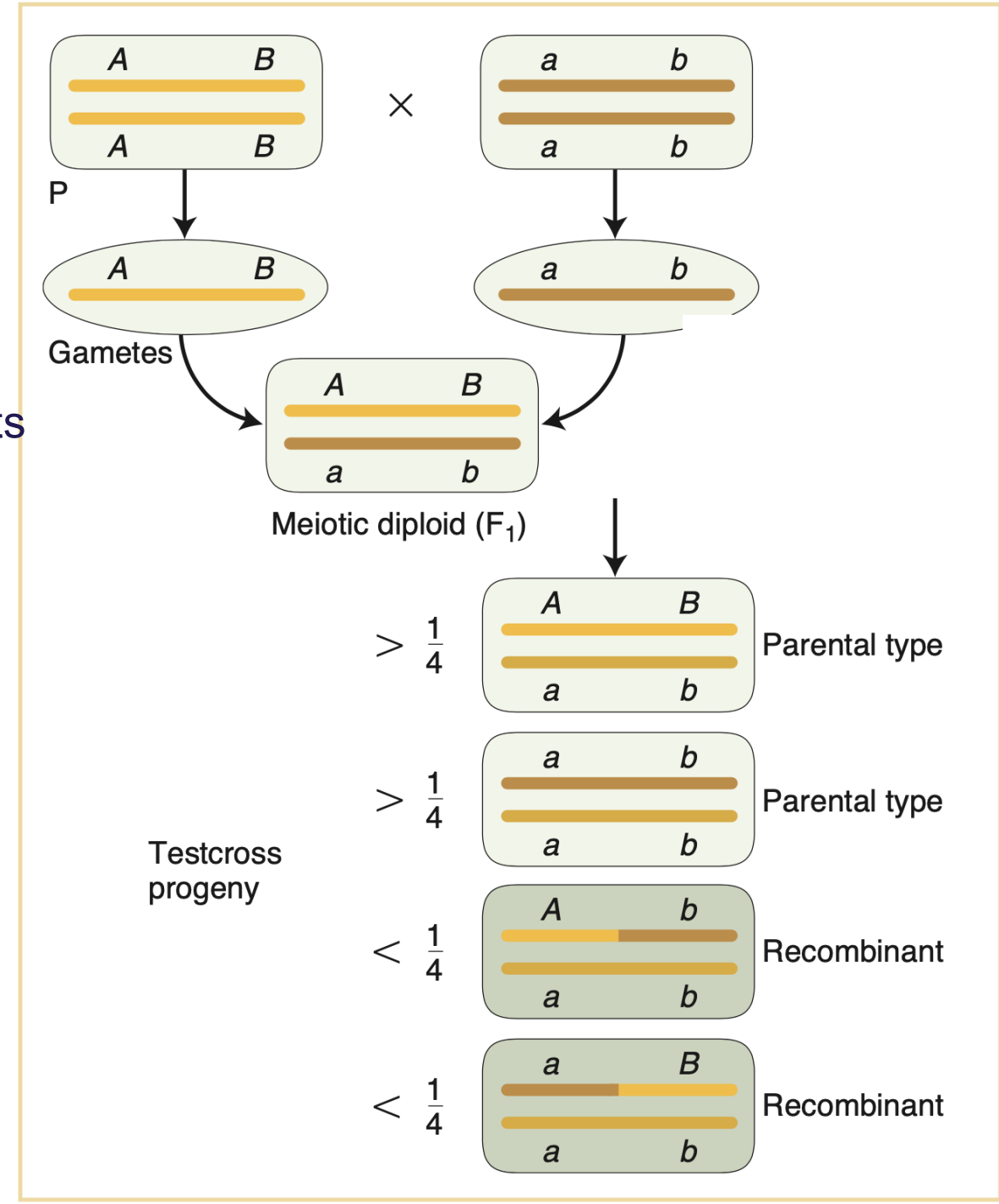
When genes are located on different chromosomes (= free recombination) equal amount of each gamete type is produced.



LINKAGE
RECOMBINATION

Maximum 50% of the gametes can be recombinants

If 50% of the gametes are recombinants then there will be two parental and two recombinant gametes.

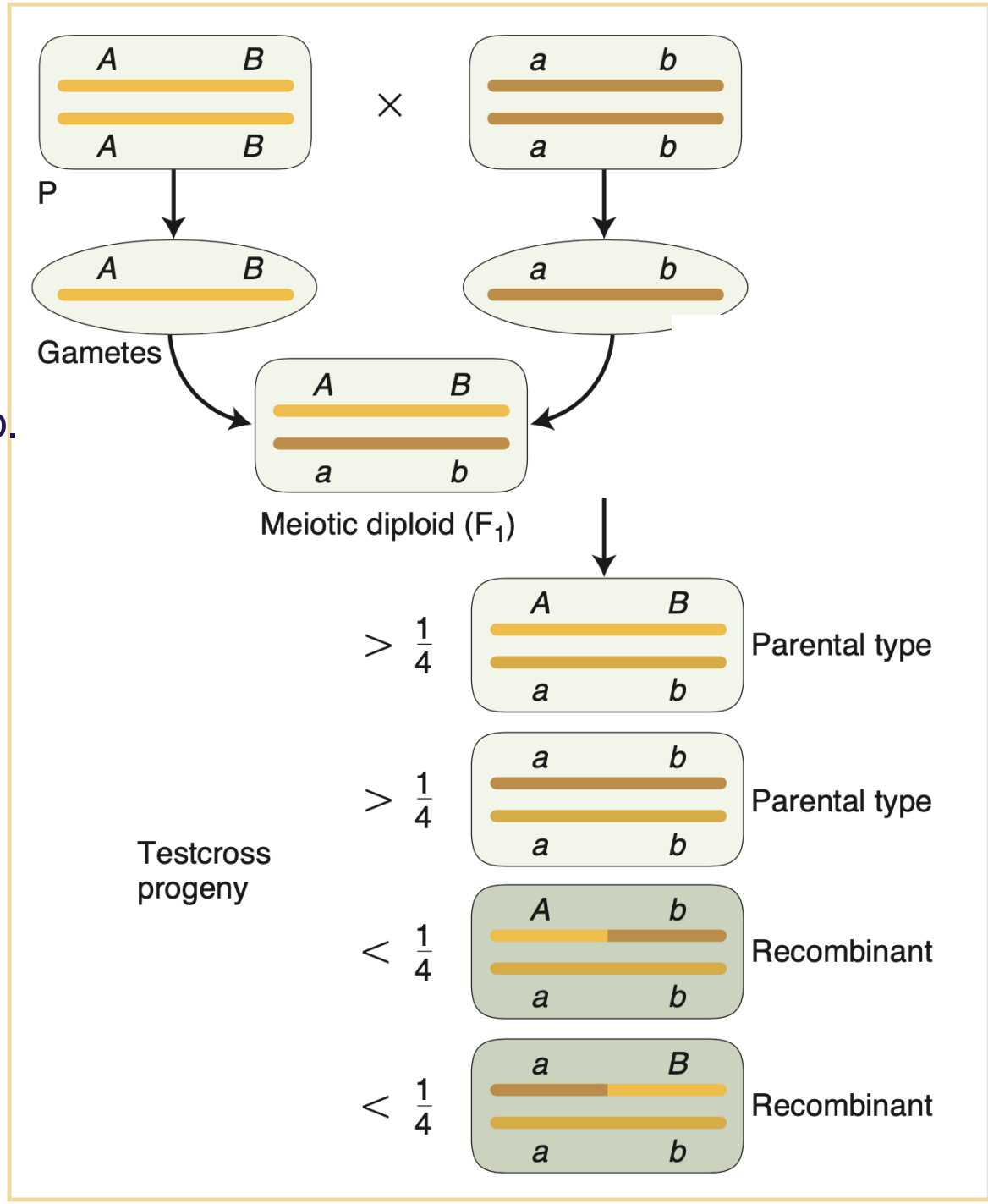


LINKAGE
RECOMBINATION

Genes on the same chromosome is a linkage group.

The proportion of recombinant gametes depend on the distance between the two genes.

Short distance – small probability for crossovers.



RECOMBINATION FREQUENCY

A MEASURE OF DISTANCE

One map unit (mu) is defined as 1% recombination between two genes (RF=0.01).
Map unit is also known as centimorgan (cM) [named after Thomas Hunt Morgan]

Genetic distance (cM) = (number of recombinant chromosomes / total chromosomes) x 100

F_1

$A\ B$	37	parental
$a\ b$	43	
$A\ b$	9	recombinante
$a\ B$	11	

Disrtance between $A - B$ $= \frac{9+11}{100} = 0.2; 20\text{cM}$

GENETIC DISTANCE

- Distance between two loci (two markers) is measured as a probability (cM)
- Two loci could be
  - Two neutral loci; Locus1 and Locus2
  - One neutral and one disease-causing locus; Locus1 and a disease locus

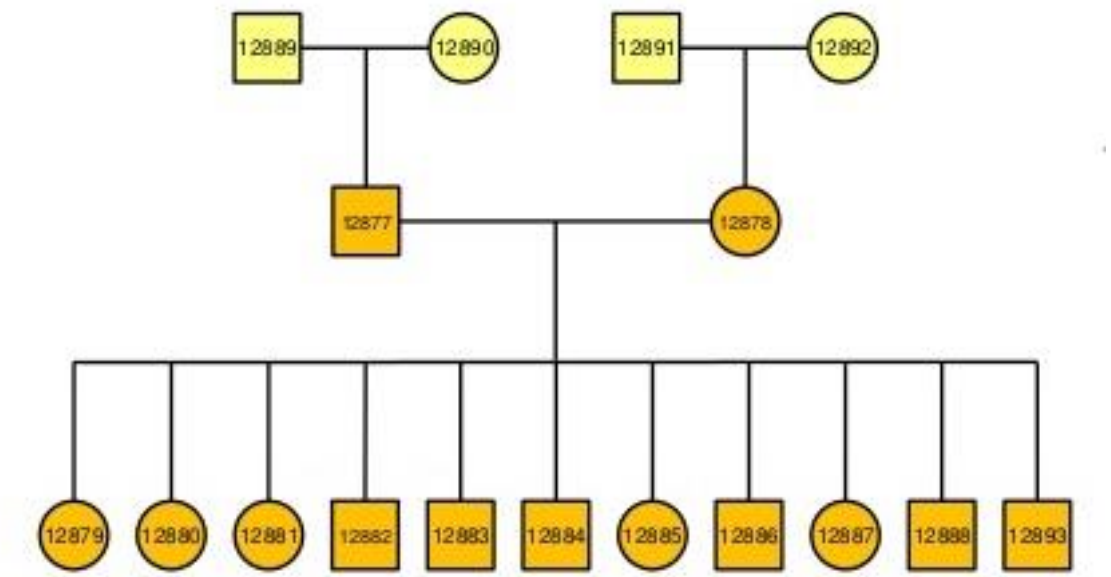
0 cM= no recombination (loci completely linked; always segregate together)

10 cM = 10% probability of crossover in each meiosis

50 cM = 50% probability (the two loci are completely unlinked, as if loci were on different chromosomes)

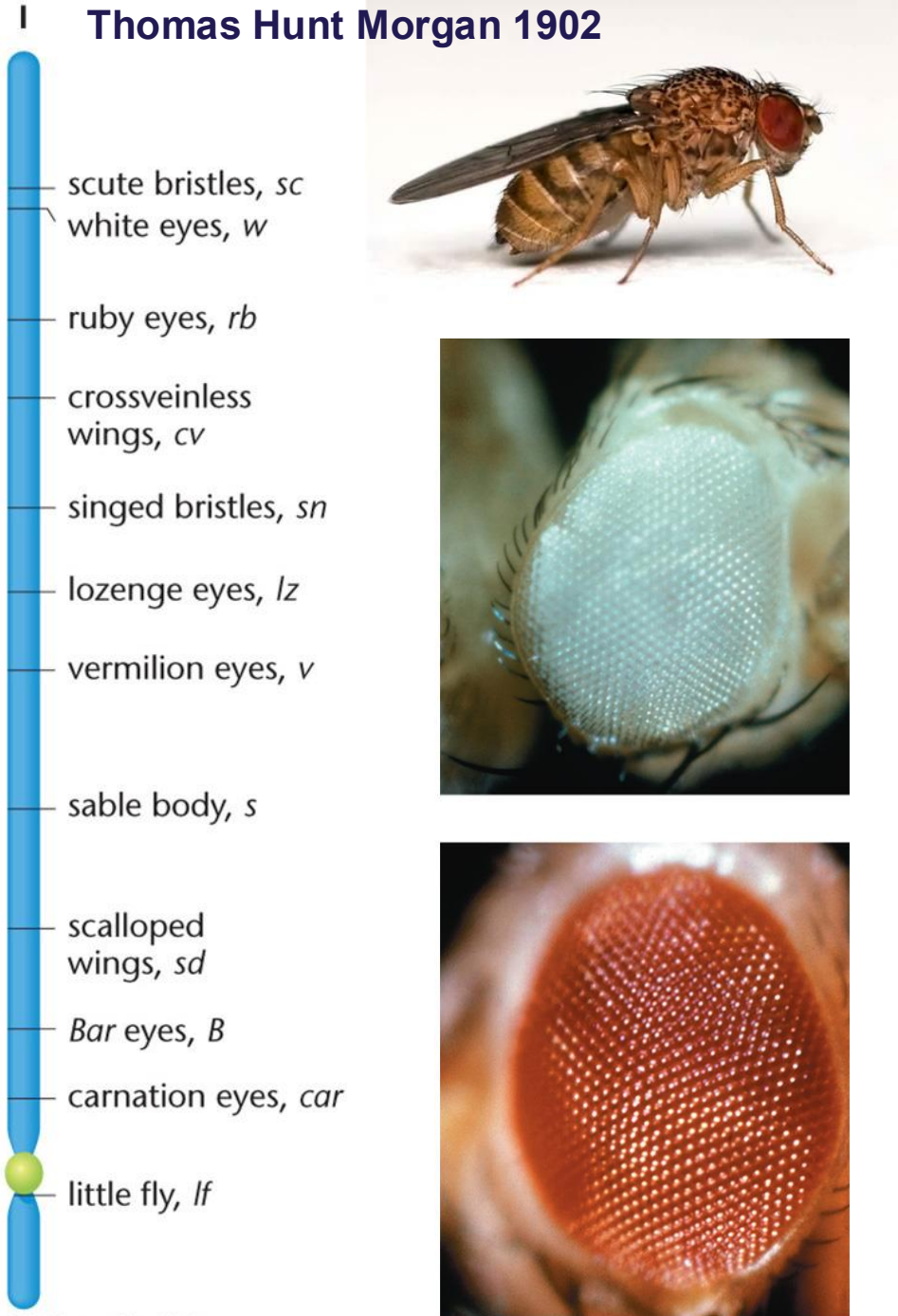
MANY MEIOSIS' ARE NEEDED TO BUILD A MAP

Large families (CEPH families, had many children)
Pick any two loci. Count parental and recombinant haplotypes.



Took systematically all loci – one pair at a time
Build a map

Thomas Hunt Morgan 1902



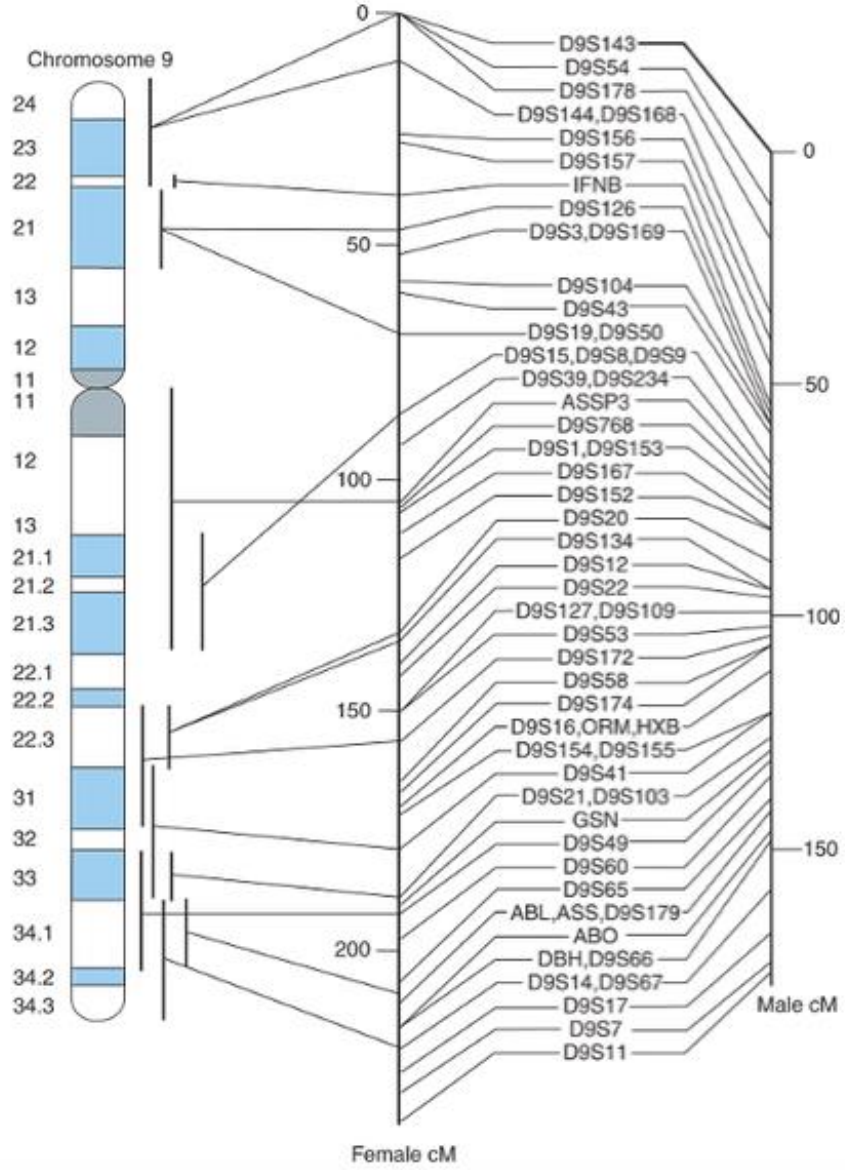
THE MAP OF THE HUMAN GENOME

All loci got a position in the human genetic map.

However, it turned out, that the female map was longer than the male map.

Why?!

- difference in amount of meiosis between sexes;
- higher recombination frequency among females (increase distance)

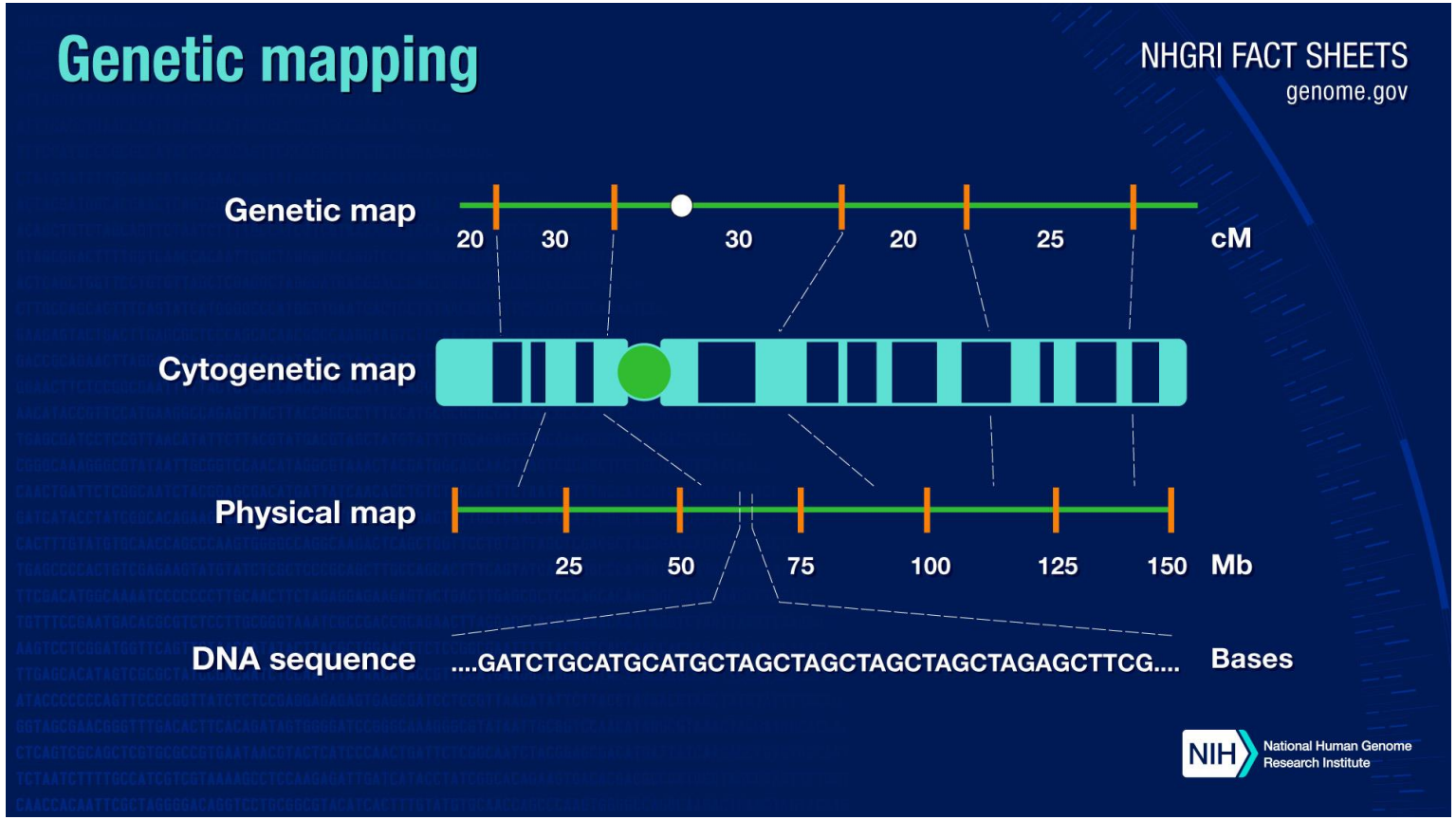


TWO TYPES OF MAPS

With genome sequencing of the human genome a physical map was generated.

Genes appear closer together when there is low recombination frequency between two genes.

Genes appear farther apart when there is high recombination frequency between two genes.



OUTLINE

08:15 – 09:15 Recap + Exercises E15 [Part III]

09:15 – 09:30 Break

09:30 – 09:50 Lecture 1 [*Genetic risk assessment]*

09:50 – 10:30 Group work

10:30 – 11:10 Break + Exercises I [E1-E3]

11:10 – 11:40 Lecture 2 [*Linkage*]

11:40 – 11:55 Exercises II [E4-E8]

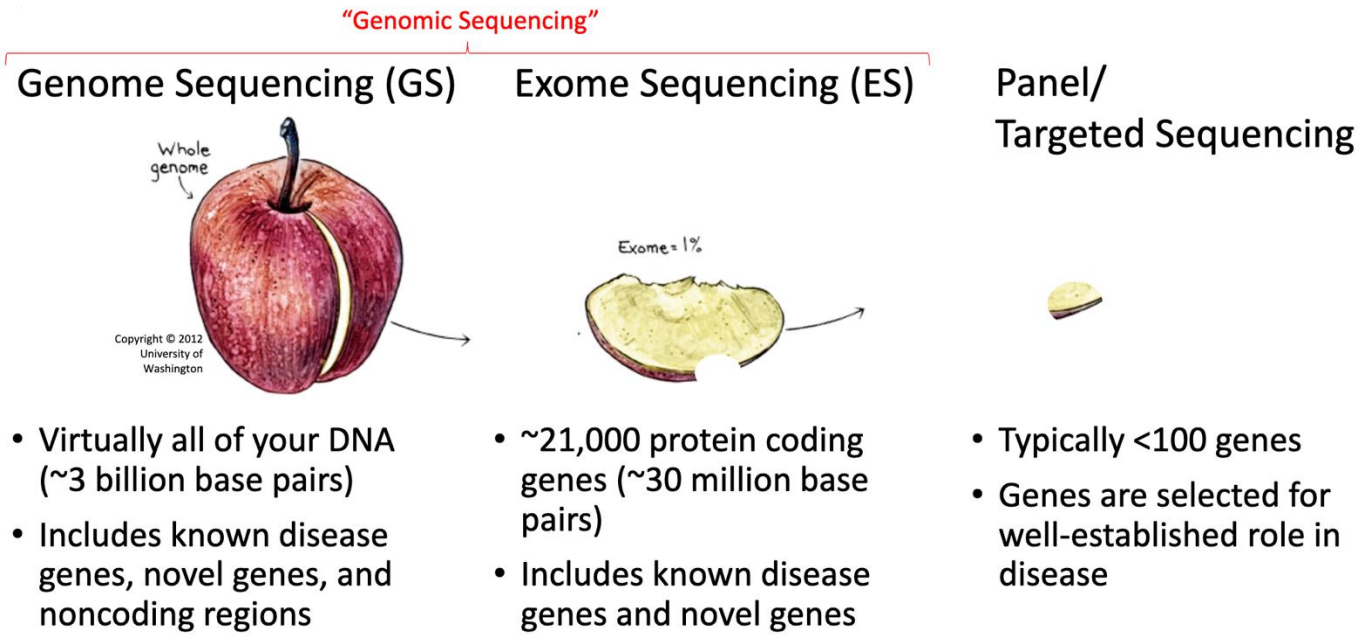
11:55 – 12:00 Reflection

OUTLINE

08:15 – 09:15	Recap + Exercises E15 [Part III]
09:15 – 09:30	Break
09:30 – 09:50	Lecture 1 [*Genetic risk assessment]*
09:50 – 10:30	Group work
10:30 – 11:10	Break + Exercises I [E1-E3]
11:10 – 11:40	Lecture 2 [*Linkage*]
11:40 – 11:55	Exercises II [E4-E8]
11:55 – 12:00	Reflection

CLINICAL APPLICATION OF NEXT-GENERATION SEQUENCING (NGS)

- A general disadvantage is that often multiple different assays are needed – NGS solves this
  - e.g., in the case of genetic heterogeneity
- NGS can be used on disorders that have variable penetrance

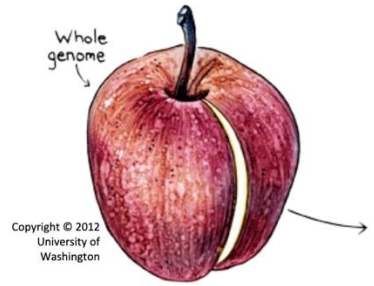
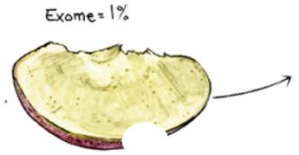


CLINICAL APPLICATION OF NEXT-GENERATION SEQUENCING (NGS)

- In principle all variants detected
- Relatively low cost (3000 kr)
- Things that took 20 years can now be done in few days

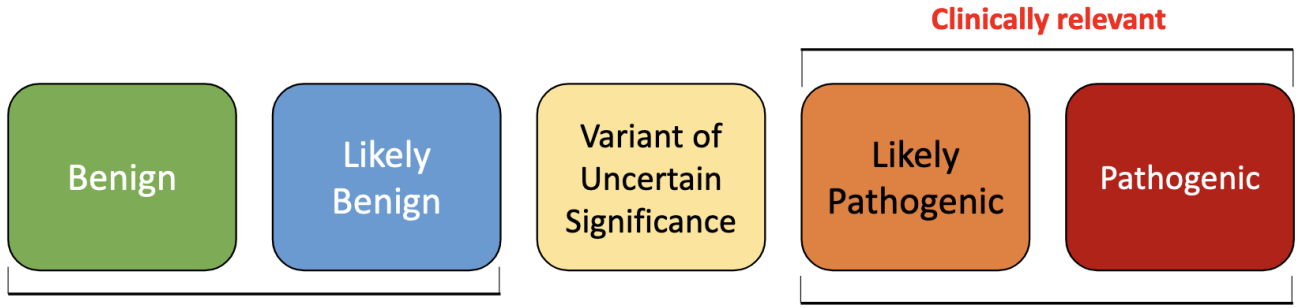
- Produces lots of data
- Can be hard to find the real pathogenic variant, if not seen before
- Mutations, that we were not looking for (e.g., *BRCA1* mutation) – incidental finding

"Genomic Sequencing"

Genome Sequencing (GS)	Exome Sequencing (ES)	Panel/ Targeted Sequencing
		
• Virtually all of your DNA (~3 billion base pairs) • Includes known disease genes, novel genes, and noncoding regions	• ~21,000 protein coding genes (~30 million base pairs) • Includes known disease genes and novel genes	• Typically <100 genes • Genes are selected for well-established role in disease

VARIANTS ON INTEREST?

- ACMG guidelines [American College of Medical Genetics and Genomics]
  - Put all variants into any of these categories by looking at the variants impact on the protein (missense, synonymous, nonsens)

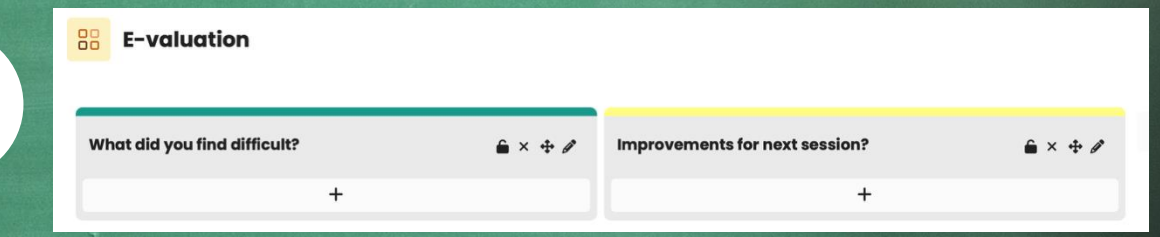


- Population frequency > disease prevalence
- No impact on amino acid sequence
- Changes amino acid at a poorly conserved position
- Inheritance not supportive of a disease causing role

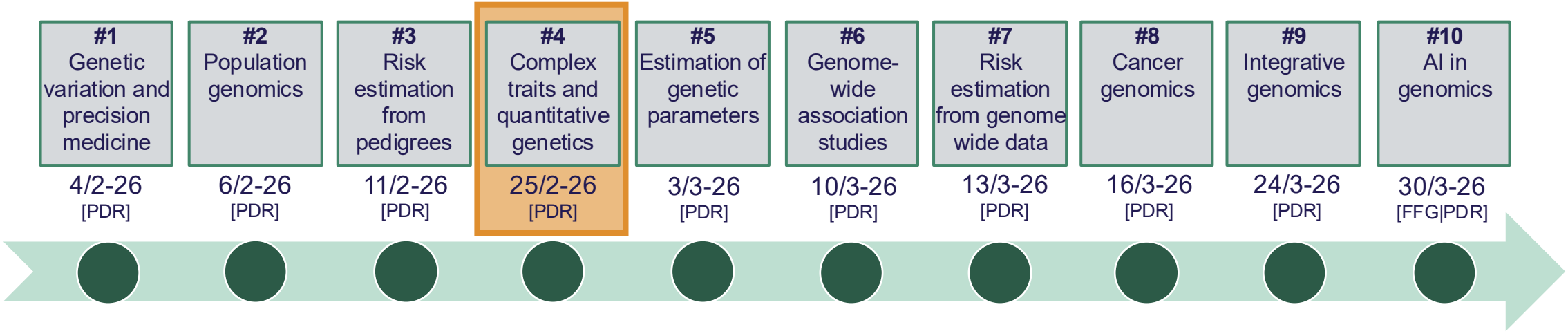
- Rare
- Severe protein impact
- Reported in other individuals with consistent phenotypes
- Segregates with disease in families
- De novo occurrences
- Functional studies supportive of an impact

REFLECT TOGETHER 2 AND 2

- What will you remember from today?
- What do you need to follow-up on?



NEXT TIME



- Molecular tools for diagnosis;
  - direct vs indirect test

Moved to session 4