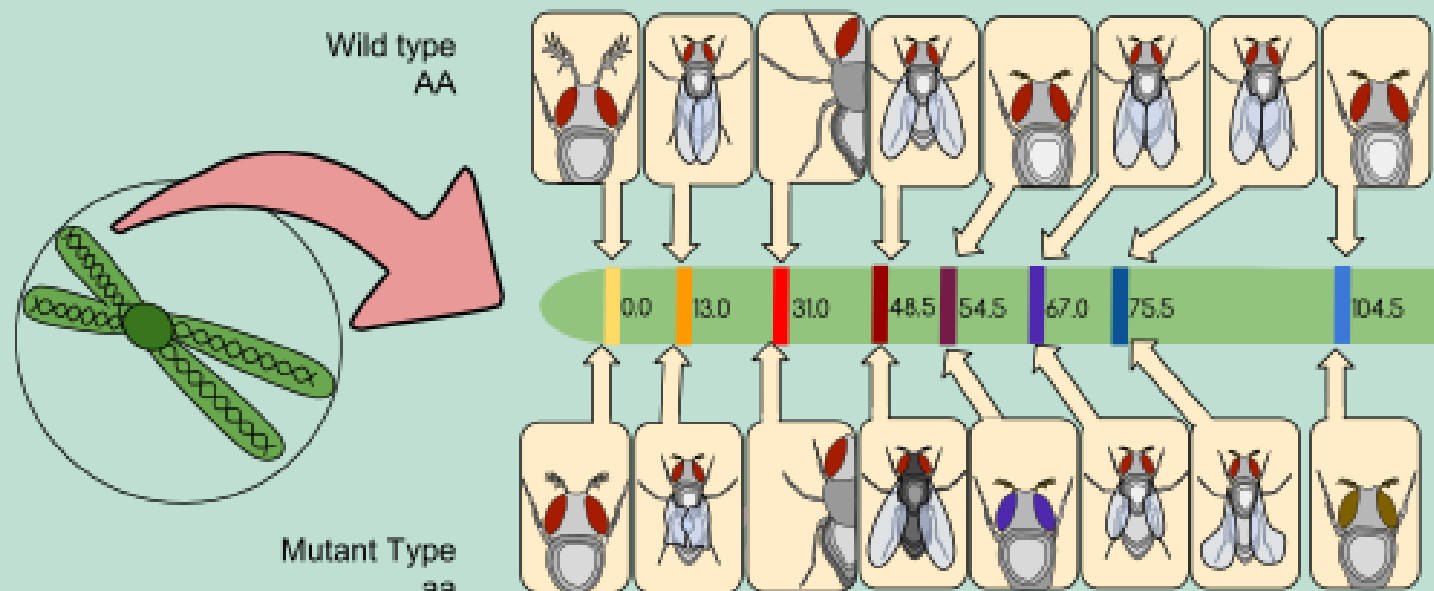


# RISK ESTIMATION FROM PEDIGREES

#3

PALLE DUUN ROHDE

palledr@hst.aau.dk



# 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 – 09:00 | Recap + Exercises E15 [Part III]             |
| 09:00 – 09:10 | Break  |
| 09:10 – 09:30 | Lecture 1 [ <i>Genetic risk assessment</i> ] |
| 09:30 – 10:00 | Group work                                   |
| 10:00 – 10:40 | Break + Exercises 1 [1-3]                    |
| 10:40 – 11:15 | Lecture 2 [ <i>Linkage</i> ]                 |
| 11:15 – 11:55 | Break + Exercises 2 [4-8]                    |
| 11:55 – 12:00 | Reflection                                   |

# OUTLINE

|                      |  |
|----------------------|--|
| <b>08:15 – 09:00</b> | <b>Recap + Exercises E15 [Part III]</b>      |
| <b>09:00 – 09:10</b> | Break  |
| <b>09:10 – 09:30</b> | Lecture 1 [ <i>Genetic risk assessment</i> ] |
| <b>09:30 – 10:00</b> | Group work                                   |
| <b>10:00 – 10:40</b> | Break + Exercises 1 [1-3]                    |
| <b>10:40 – 11:15</b> | Lecture 2 [ <i>Linkage</i> ]                 |
| <b>11:15 – 11:55</b> | Break + Exercises 2 [4-8]                    |
| <b>11:55 – 12:00</b> | 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]



# WHY IS POPULATION GENETICS IMPORTANT?

Population genomics tackles questions about genetic diversity

*0.08% of nucleotide base pair in human DNA vary among individuals*

*Humans and orangutans are ~97% similar*

## Why this little genetic diversity?

- Selection favour functionally different DNA alleles in different circumstances
- DNA variation is tolerated when the alleles of a gene are functionally equivalent

The **aim of population genomics** is to model the dynamics of **evolutionary change within and between populations**.

# THE FOUR FORCES

**Mutation** Copying errors during DNA replication, which introduce new alleles into the population

**Natural selection** differential transmission of alleles into the next generation due to the consequences of functional differences on an individual's survival and reproductive success

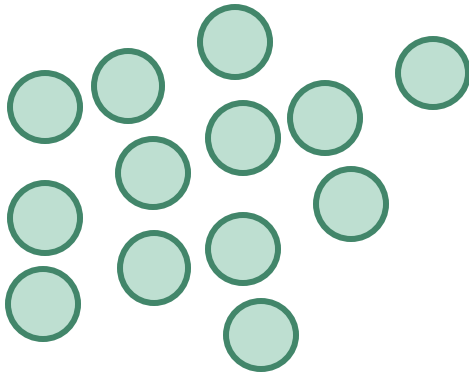
**Genetic drift** differential transmission of alleles into the next generation as a result of random sampling, and has the greatest potential impact in small populations

**Gene flow** spreads alleles from one population into another via migration, making them more genetically similar to each other, and countering genetic differentiation by drift



# 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<sub>AA</sub>**, **P<sub>Aa</sub>** and **P<sub>aa</sub>**.

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

# FREQUENCIES

|                    |            |            |            |          |
|--------------------|------------|------------|------------|----------|
| Genotype           | AA         | Aa         | aa         | $\Sigma$ |
| 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  $\rightarrow$  Allele frequencies

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

Allele frequencies  $\rightarrow$  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!

|         |       | Males |       |
|---------|-------|-------|-------|
|         |       | A (p) | a (q) |
| Females | A (p) | $p^2$ | pq    |
|         | a (q) | pq    | $q^2$ |

# 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 harmful alleles  
*diversity within populations* ↓  
*diversity between populations* ↓↑

## Non-random mating

do not change alleles, but change genotype frequencies

# OUTLINE

|               |  |
|---------------|--|
| 08:15 – 09:00 | Recap + Exercises E15 [Part III]             |
| 09:00 – 09:10 | Break  |
| 09:10 – 09:30 | Lecture 1 [ <i>Genetic risk assessment</i> ] |
| 09:30 – 10:00 | Group work                                   |
| 10:00 – 10:40 | Break + Exercises 1 [1-3]                    |
| 10:40 – 11:15 | Lecture 2 [ <i>Linkage</i> ]                 |
| 11:15 – 11:55 | Break + Exercises 2 [4-8]                    |
| 11:55 – 12:00 | Reflection                                   |

# OUTLINE

|                      |   |
|----------------------|---|
| 08:15 – 09:00        | Recap + Exercises E15 [Part III]                    |
| 09:00 – 09:10        | Break   |
| <b>09:10 – 09:30</b> | <b>Lecture 1</b> [ <i>Genetic risk assessment</i> ] |
| 09:30 – 10:00        | Group work  |
| 10:00 – 10:40        | Break + Exercises 1 [1-3]                           |
| 10:40 – 11:15        | Lecture 2 [ <i>Linkage</i> ]                        |
| 11:15 – 11:55        | Break + Exercises 2 [4-8]                           |
| 11:55 – 12:00        | Reflection  |



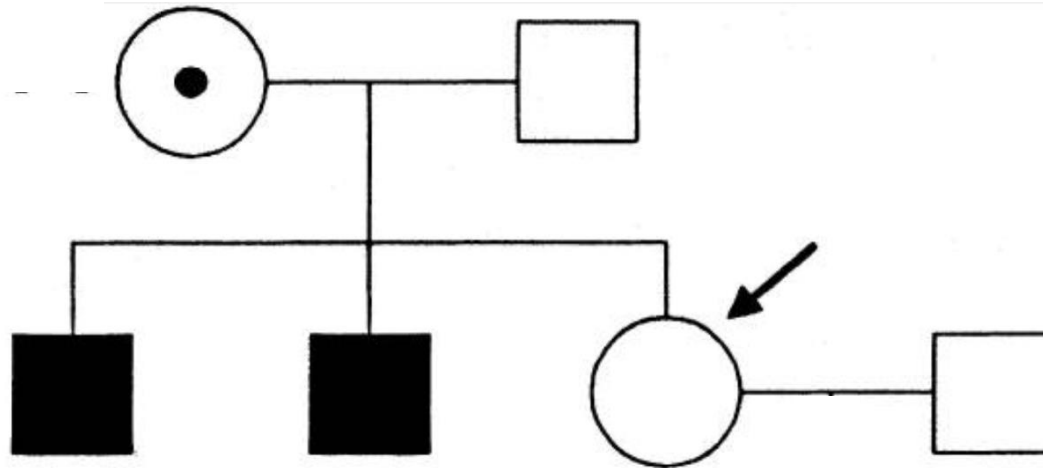
# MONOGENIC RISK ASSESSMENT



# MONOGENIC INHERITANCE



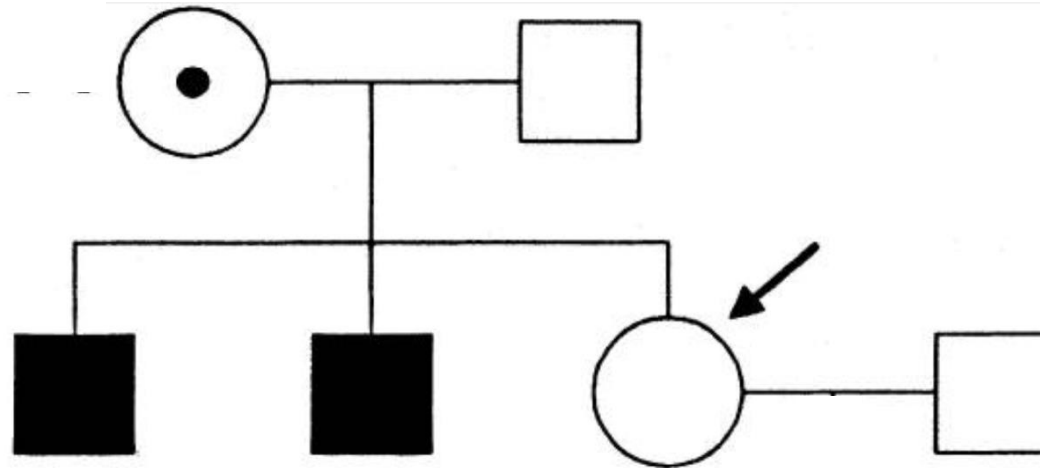
What type of inheritance is seen in the pedigree?



# MONOGENIC INHERITANCE



What type of inheritance is seen in the pedigree?

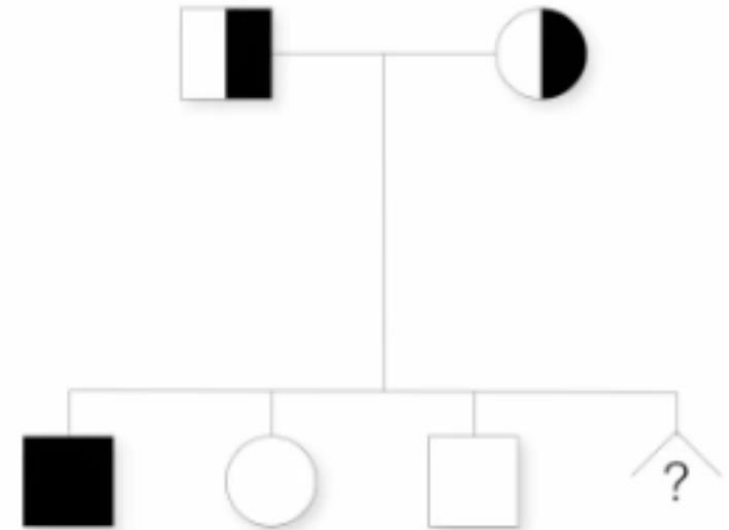


What is the probability that II.3 is a carrier?

# AUTOSOMAL RECESSIVE I

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

Their risk of getting a fourth affected child is =  $\frac{1}{4}$  [draw punnet square]



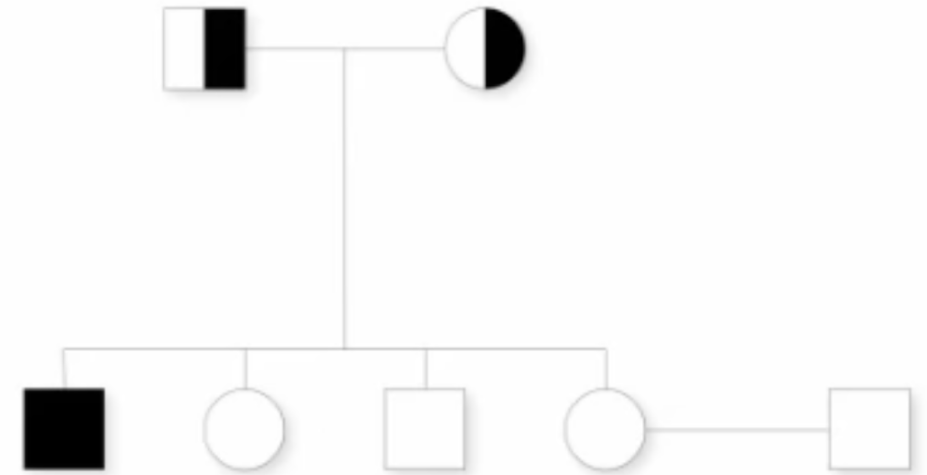
# AUTOSOMAL RECESSIVE II

The risk of II.4 being a carrier must be  $\frac{2}{3}$  [we know that she is not affected, thus she cannot be aa].

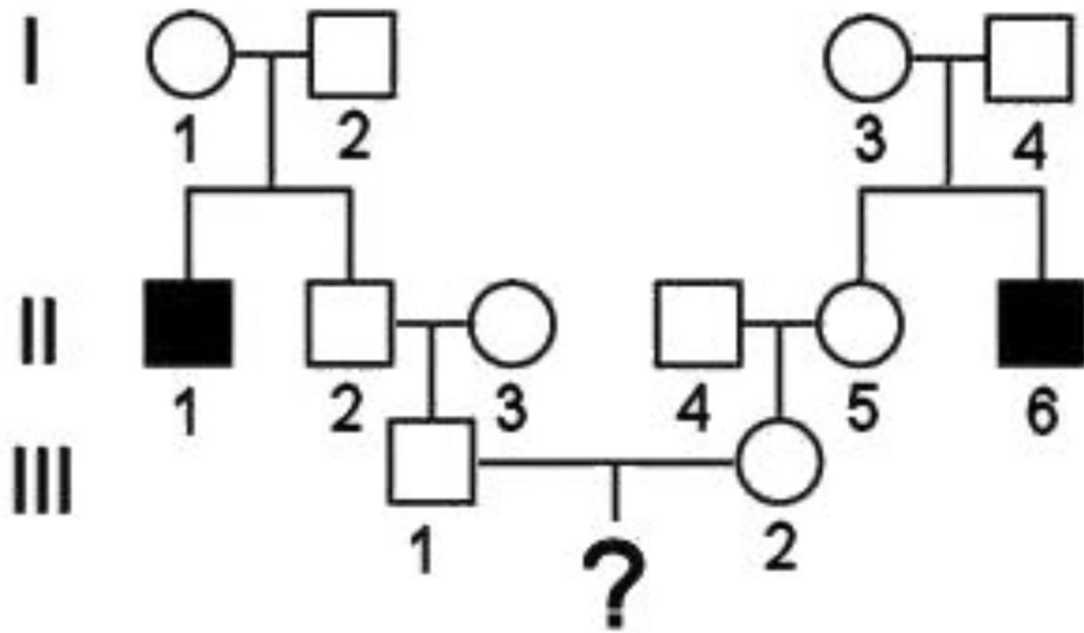
The risk of II.5 of being a carrier (given no family history of the disease) is the population risk.

For an AR the population frequency could be  $\frac{1}{25}$ .

The risk the couple will get an affected child is then:  
 $\frac{1}{2} * \frac{2}{3} * \frac{1}{25} * \frac{1}{2} = \frac{1}{150}$



# AUTOSOMAL RECESSIVE III

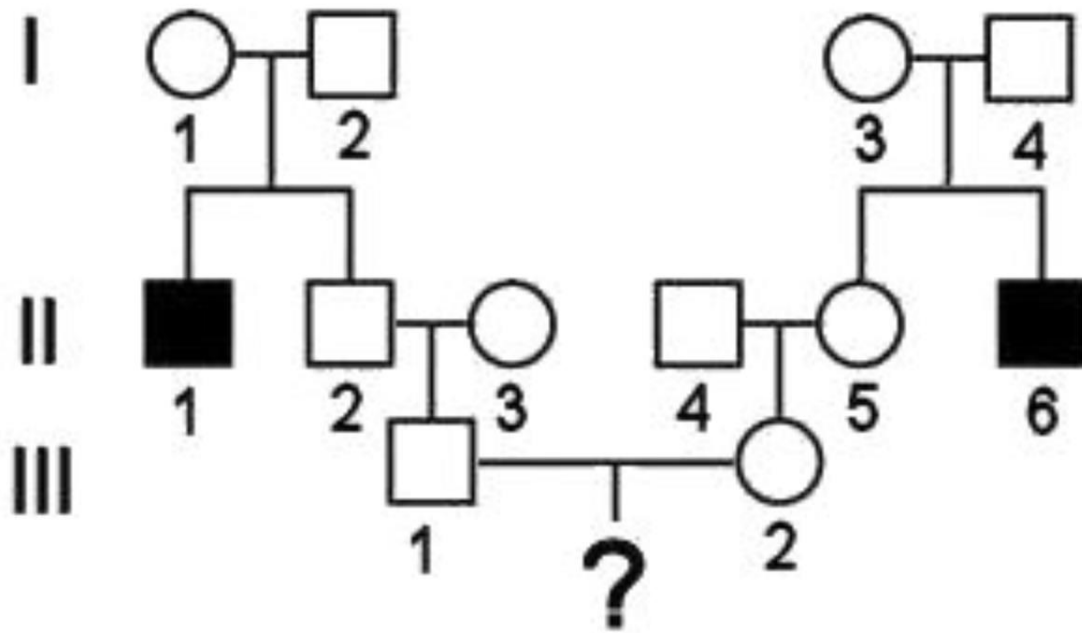


What is the **probability** that IV.1 is affected (**aa**)?

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 must 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 probability that he is **Aa** is the probability that II.2 is **Aa** and passes **a** to his son,  $\frac{1}{2} * \frac{2}{3} = \frac{1}{3}$
6. The probability that III.2 is **Aa** is  $\frac{1}{2} * \frac{2}{3} = \frac{1}{3}$
7. The probability that IV.1 is **aa**  $\frac{1}{4} * \frac{1}{3} * \frac{1}{3} = \frac{1}{36}$

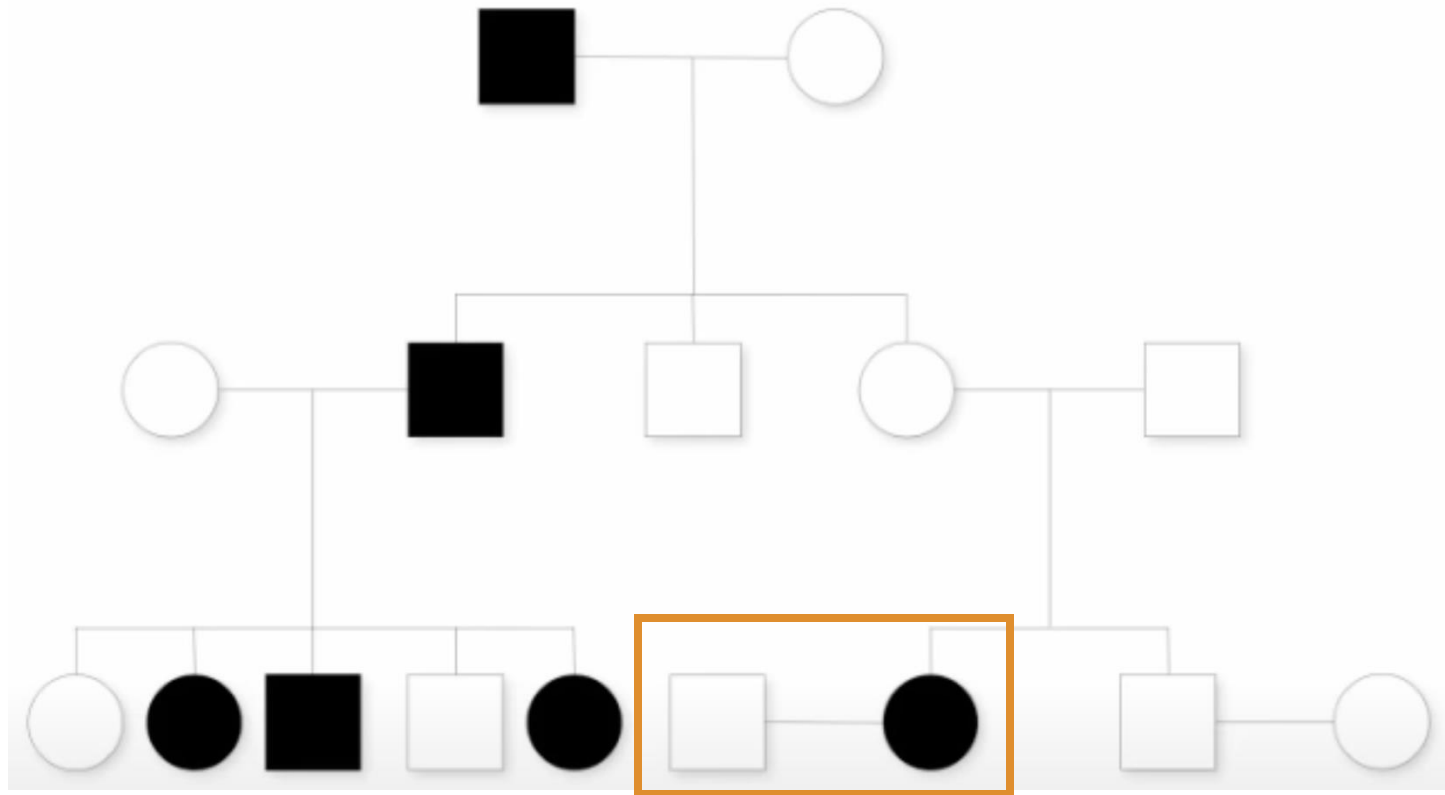
# AUTOSOMAL RECESSIVE IV

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



1. The probability that III.1 is **Aa**  $\frac{1}{2} * \frac{2}{3} = \frac{1}{3}$
2. The probability that III.2 is **Aa** is  $\frac{1}{2} * \frac{2}{3} = \frac{1}{3}$
3. The probability that IV.1 is **Aa**  $\frac{2}{4} * \frac{1}{3} * \frac{1}{3} = \frac{1}{18}$

# RISK WITH INCOMPLETE PENETRANCE

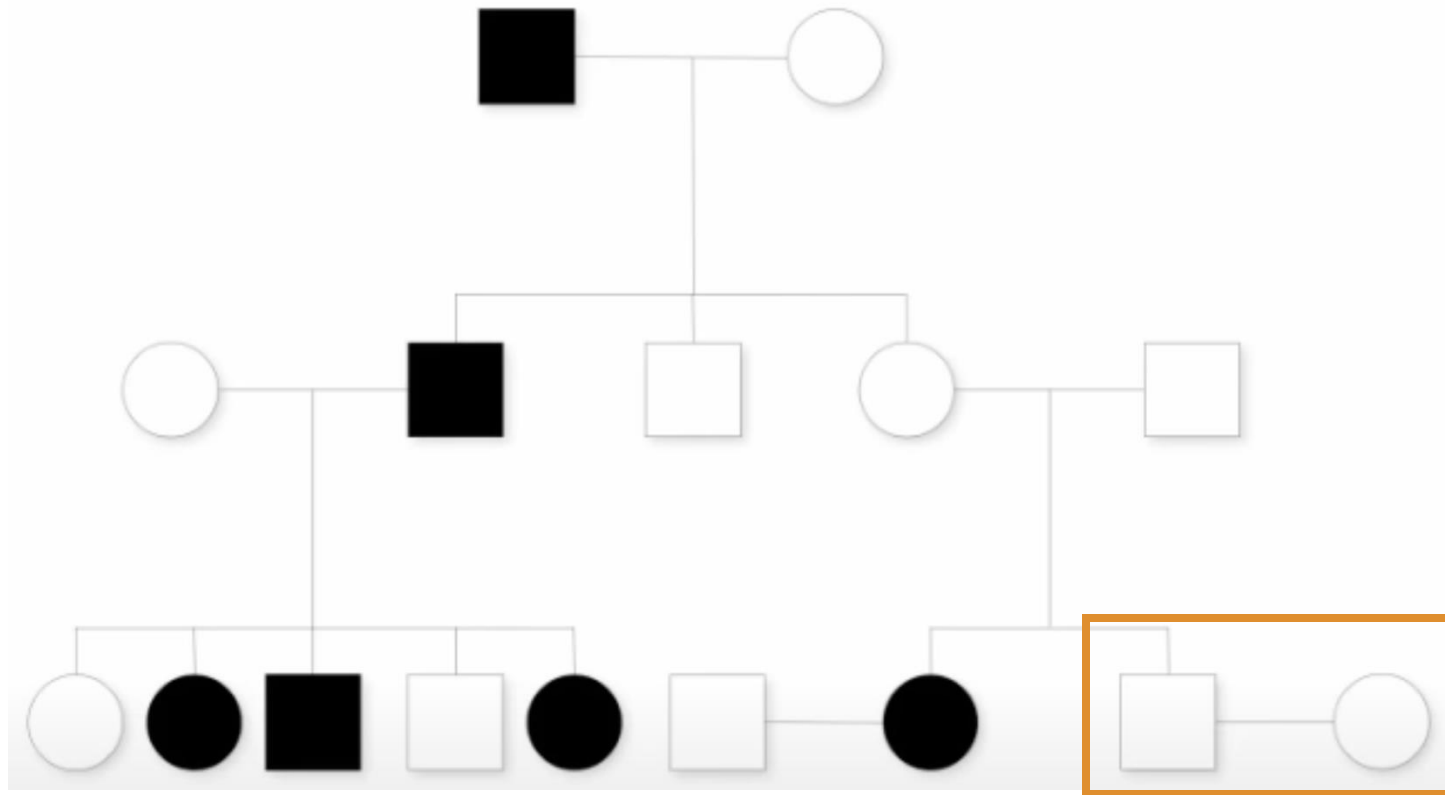


Assuming 80% penetrance  
→ 80% probability that an individual that inherits the mutation will show the phenotype

This couple's risk of getting an affected child:  
 $\frac{1}{2} * 0.8 = 0.4$



# RISK WITH INCOMPLETE PENETRANCE



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

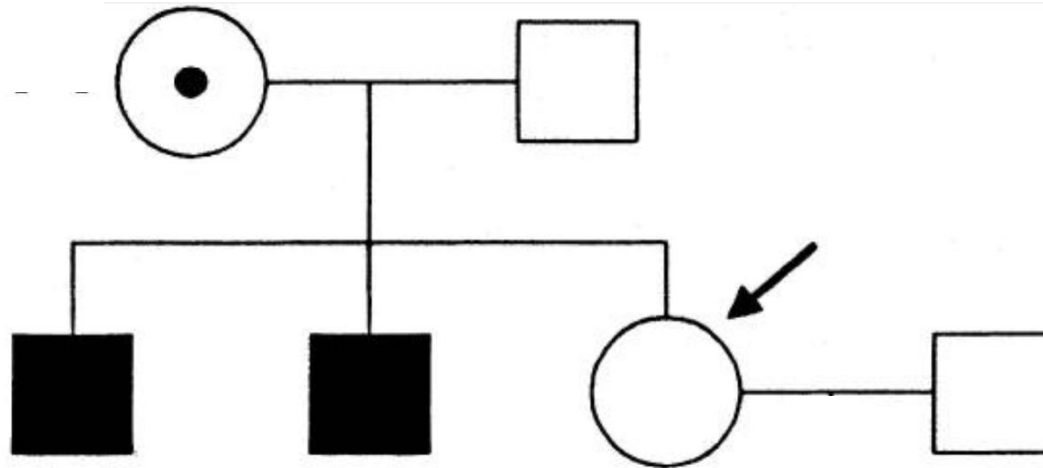
This couples risk of getting an affected child:  
 $\frac{1}{2} * 0.2 * \frac{1}{2} * 0.8 = 0.04$

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 probability that II.3 is a carrier?

# BAYES' THEOREM

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

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

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

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

# BAYES' THEOREM

## IN GENETICS

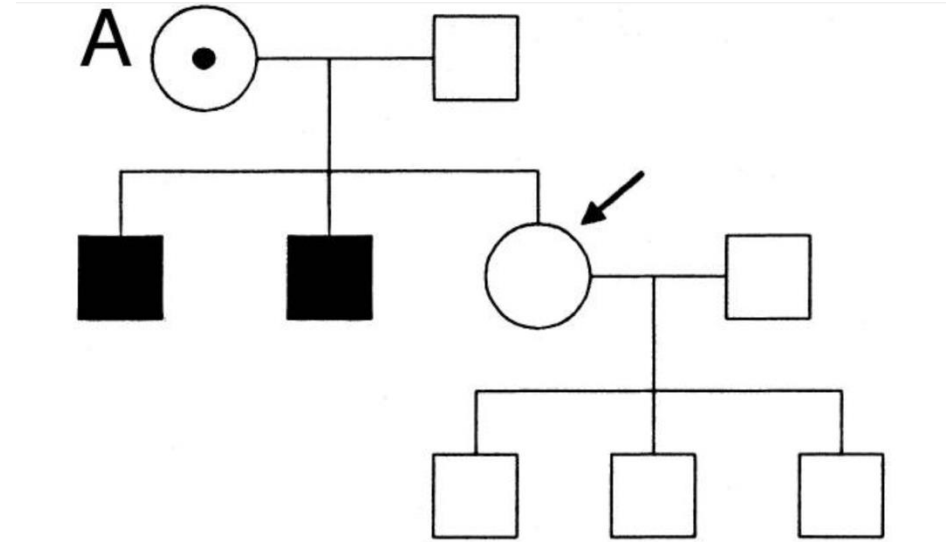
| Hypothesis              | H: Is a carrier                                      | H: Is not a carrier                                  |
|-------------------------|--|--|
| Prior probability       | $x_1$  | $x_2$  |
| Conditional probability | $y_1$  | $y_2$  |
| Joint probability       | $x_1 * y_1$  | $x_2 * y_2$  |
| Posterior probability   | $j.\text{prob1} / (j.\text{prob1} + j.\text{prob2})$ | $j.\text{prob2} / (j.\text{prob1} + j.\text{prob2})$ |

*Used when additional information becomes available.*

# BAYES' THEOREM

## IN GENETICS

What is the probability 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<br>(three normal sons) | $\frac{1}{2}^3 = 1/8$               | $1$                                |
| Joint probability                              | $1/16$                              | $1/2$                              |
| Posterior probability                          | $\frac{(1/16)}{(1/16 + 1/2)} = 1/9$ | $\frac{(1/2)}{(1/16 + 1/2)} = 8/9$ |

# OUTLINE

- 08:15 – 09:00    Recap + Exercises E15 [Part III]
- 09:00 – 09:10    Break
- 09:10 – 09:30    Lecture 1 [*Genetic risk assessment*]
- 09:30 – 10:00    **Group work**
- 10:00 – 10:40    Break + Exercises 1 [1-3]
- 10:40 – 11:15    Lecture 2 [*Linkage*]
- 11:15 – 11:55    Break + Exercises 2 [4-8]
- 11:55 – 12:00    Reflection





# Make 4 groups

Find Group-exercise on Github

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

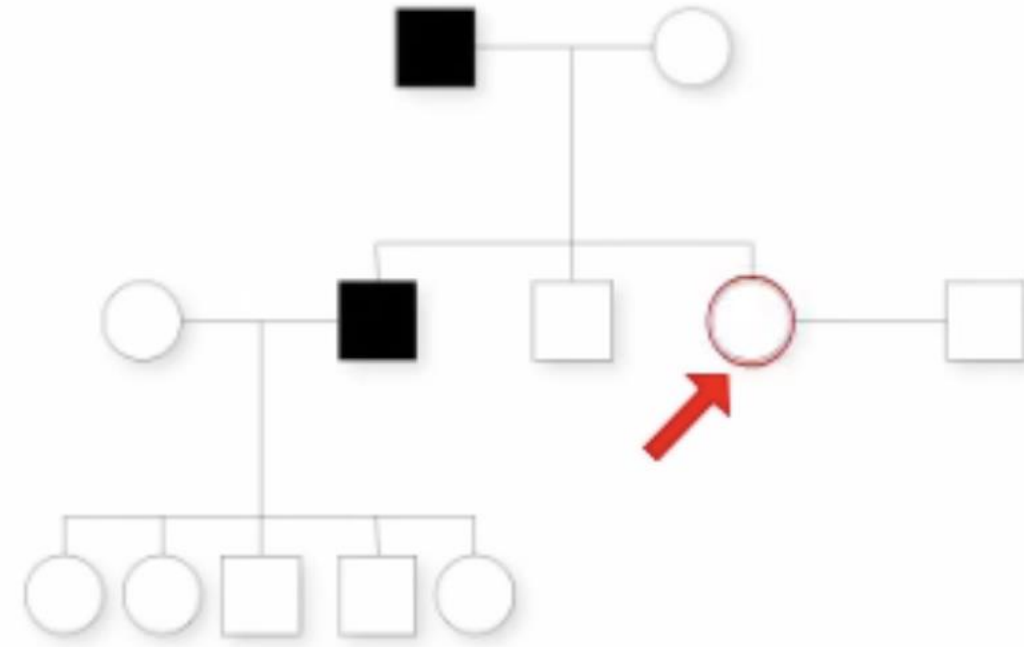
Group 2 and 4 works with '*Bayesian Analysis Using Genetic Test Results*'

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



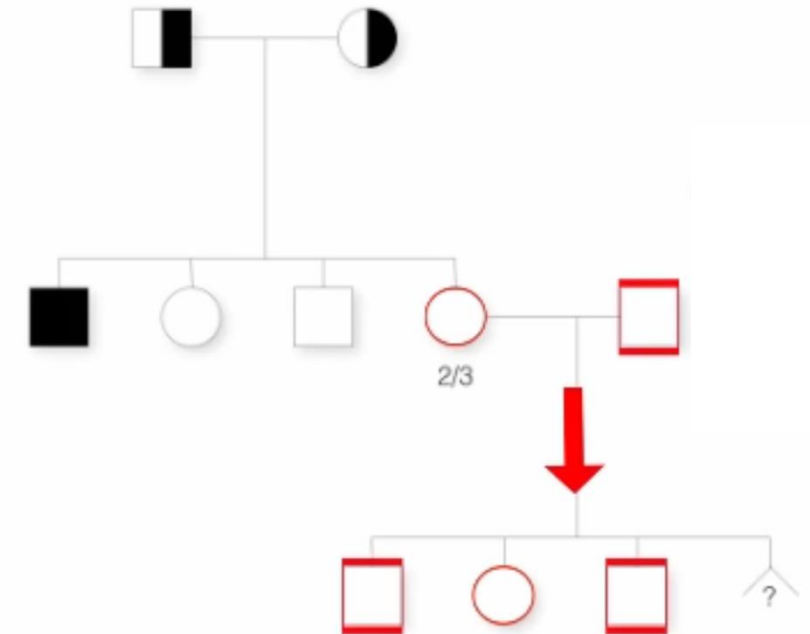
| Hypothesis   | H: Is a carrier                  | H: Is not a carrier             |
|--|----------------------------------|---------------------------------|
| Prior probability                                  | 0.5                              | 0.5                             |
| Conditional probability<br>(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<br>(three unaffected kids) | $3/4^3 = 0.42$              | 1                            |
| Joint probability                                  | 0.01                        | 0.974                        |
| Posterior probability                              | $\frac{0.01}{0.01 + 0.974}$ | $\frac{0.974}{0.01 + 0.974}$ |



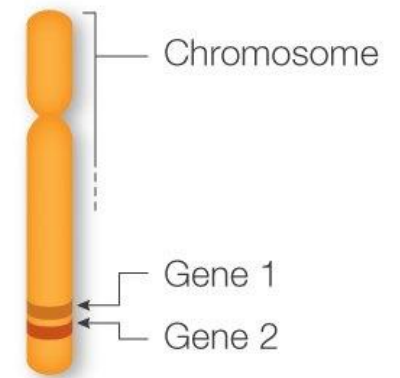
Probability that they are at risk (vs not at risk) and having three unaffected kids

# OUTLINE

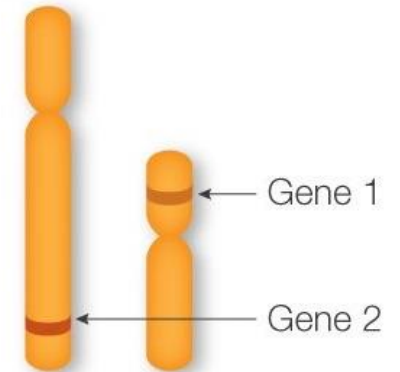
|                      |  |
|----------------------|--|
| 08:15 – 09:00        | Recap + Exercises E15 [Part III]             |
| 09:00 – 09:10        | Break  |
| 09:10 – 09:30        | Lecture 1 [ <i>Genetic risk assessment</i> ] |
| 09:30 – 10:00        | Group work                                   |
| <b>10:00 – 10:40</b> | Break + Exercises 1 [1-3]                    |
| 10:40 – 11:15        | Lecture 2 [ <i>Linkage</i> ]                 |
| 11:15 – 11:55        | Break + Exercises 2 [4-8]                    |
| 11:55 – 12:00        | Reflection                                   |

# LINKAGE

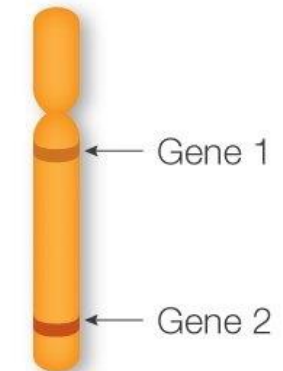
When alleles travel together



**Linked**

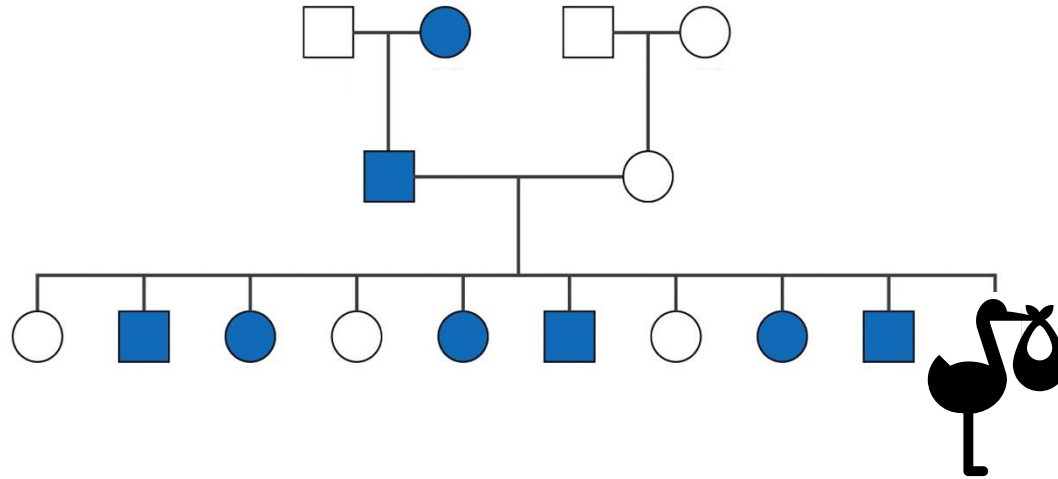


**Not Linked**



**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

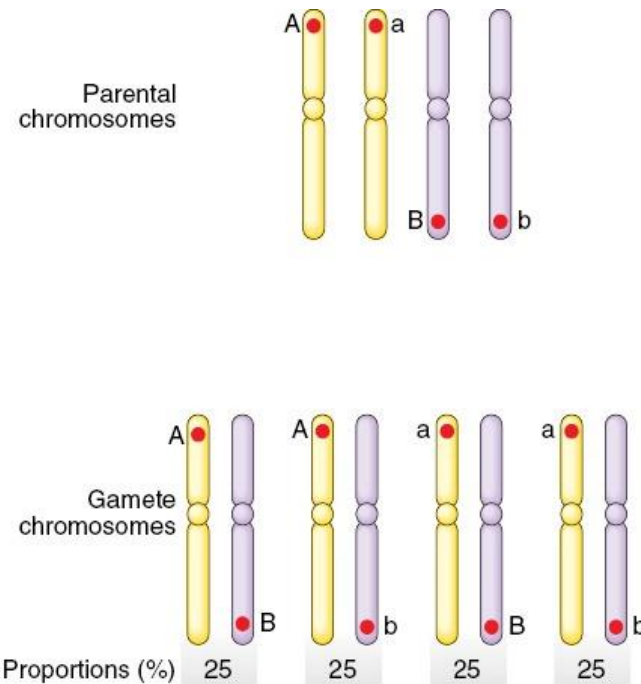


## Mendels 2. law

Alleles at different loci segregate independently during meiosis.

Only true for independent loci

|   | genloci                              |
|---|--------------------------------------|
|   | blomsterfarve frøfarve               |
| 1 |                                      |
| 2 |                                      |
| 3 |                                      |
| 4 | blomstens placering bælgens udseende |
| 5 | bælgfarve plantehøjde                |
| 6 |                                      |
| 7 | frøets udseende                      |



# INDEPENDENT ASSORTMENT



## Mendels 2. law

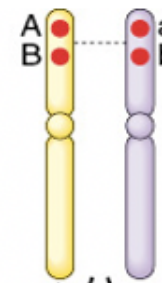
Alleles at different loci segregate independently during meiosis.

Only true for independent loci.

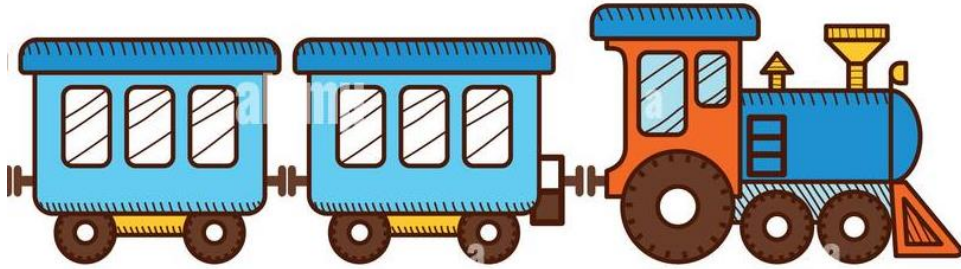
|   | genloci             |                  |
|---|---------------------|------------------|
|   | blomsterfarve       | frøfarve         |
| 1 |                     |                  |
| 2 |                     |                  |
| 3 |                     |                  |
| 4 | blomstens placering | bælgens udseende |
| 5 | bælgfarve           | plante højde     |
| 6 |                     |                  |
| 7 |                     | frøets udseende  |

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

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



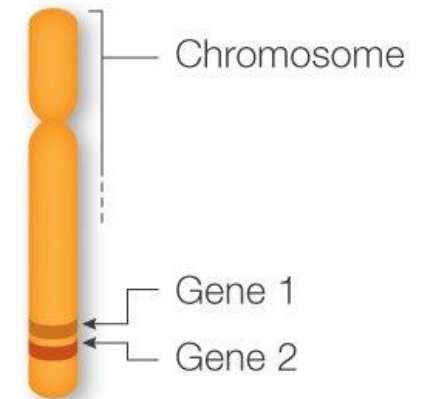
# LINKAGE



Linked train wagons

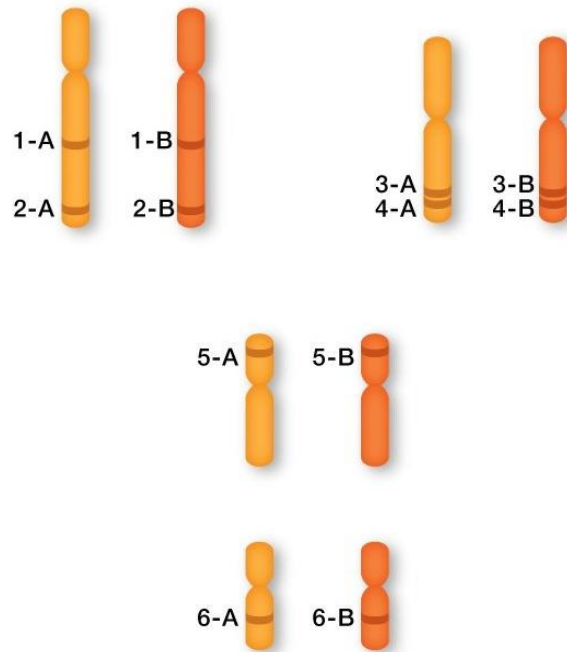


Linked prisoners



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?



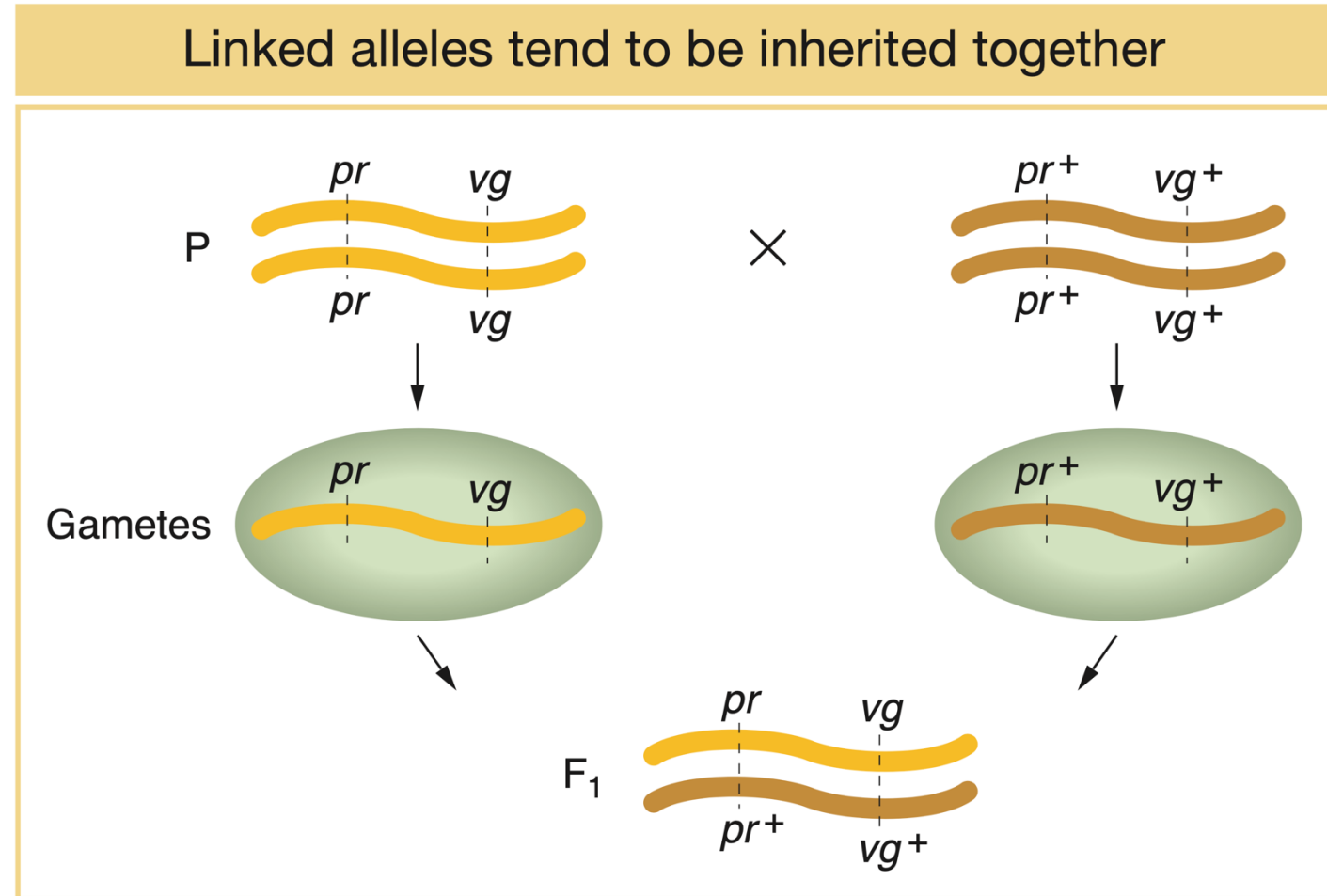
The **prisoners** are linked when they are seen together more often than by chance



# LINKED GENES

## DURING MEIOSIS

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



# LINKED GENES

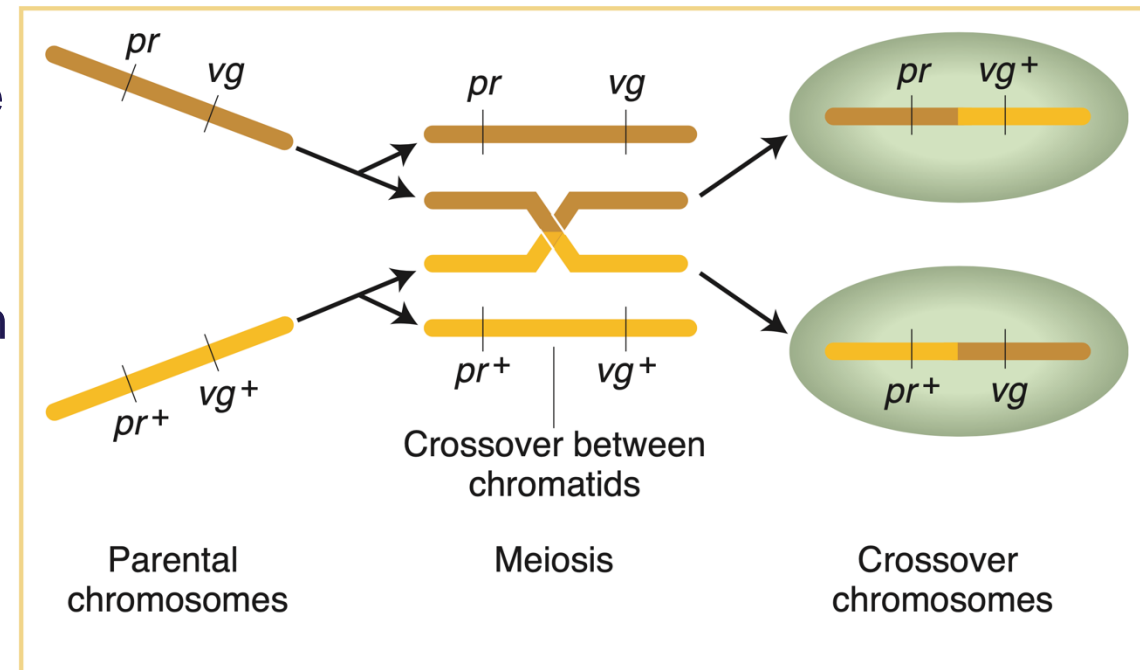
## DURING MEIOSIS

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).

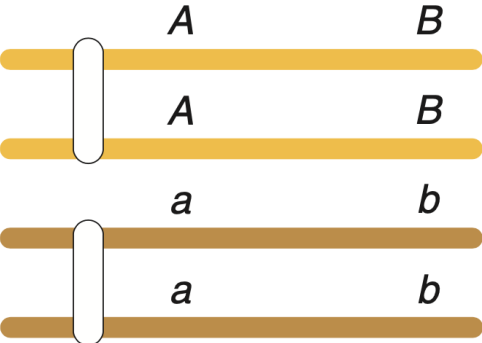
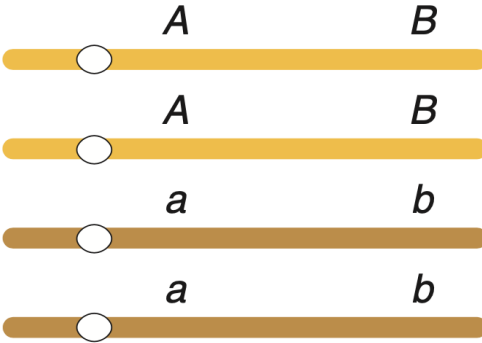
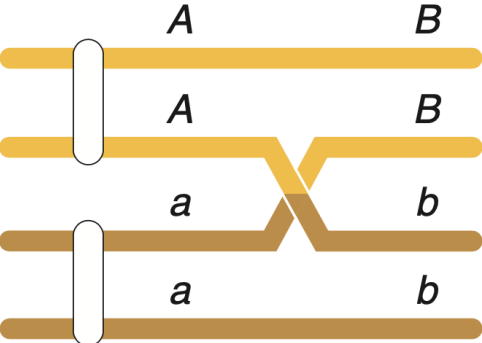
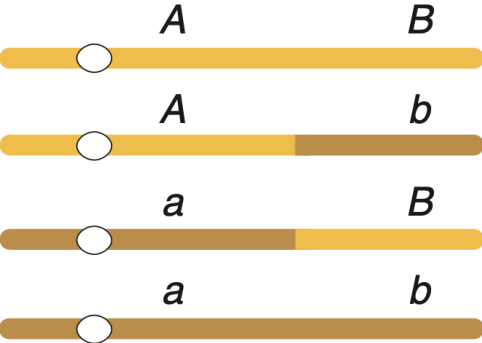
If crossover happens between 2 (or more) genes both parental and recombinant gametes are seen.

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



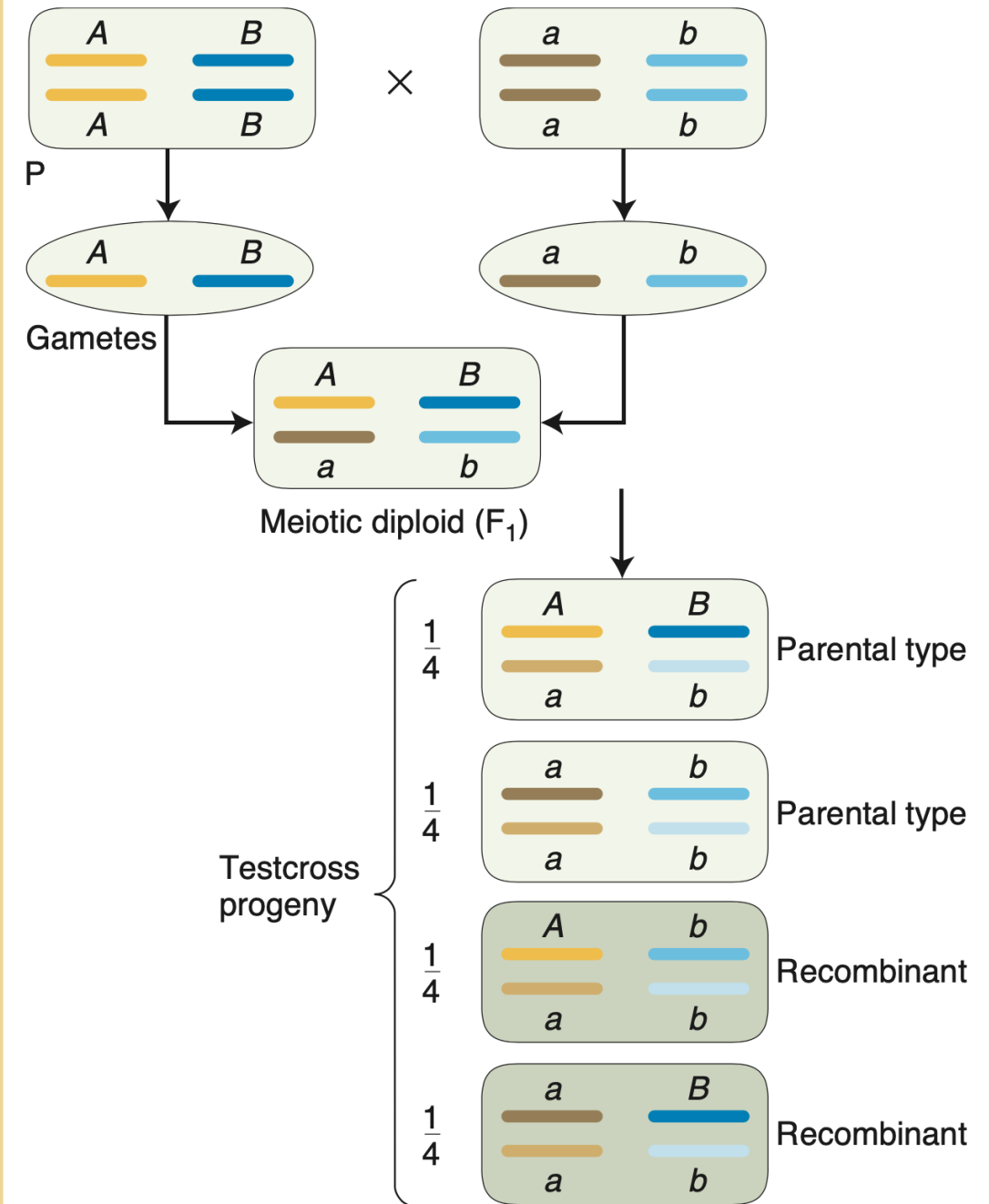
# CROSSTOVERS

## BETWEEN NON-SISTER CHROMATIDS

|   | Meiotic chromosomes   | Meiotic products   |  |
|---|---|--|--|
| Meioses with no crossover between the genes |   |   | Parental<br>Parental<br>Parental<br>Parental       |
| Meioses with a crossover between the genes  |  |  | Parental<br>Recombinant<br>Recombinant<br>Parental |

# FREE RECOMBINATION

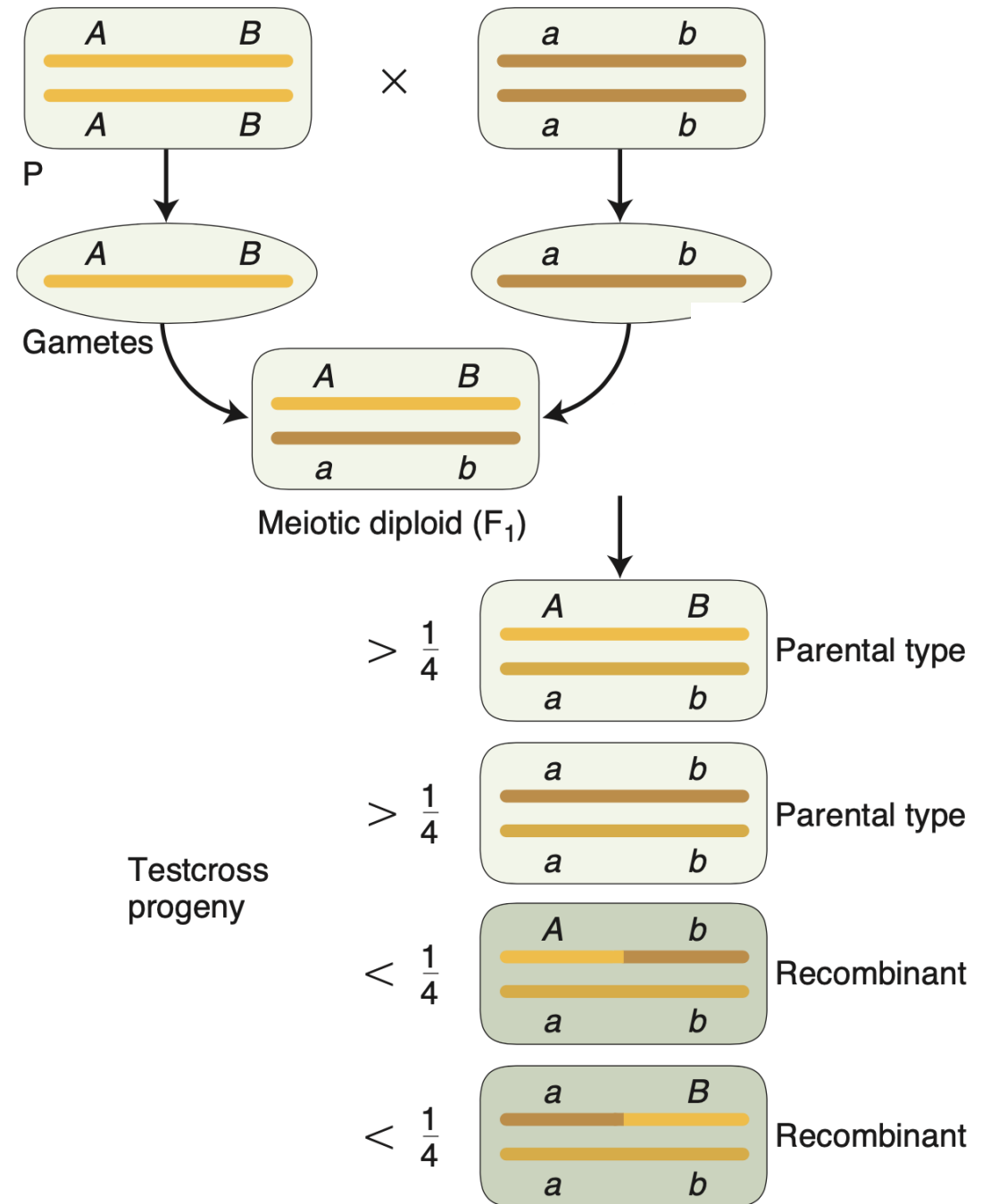
When genes are located on different chromosomes (= free recombination) equal amount of each gamete type is produced.



# 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.



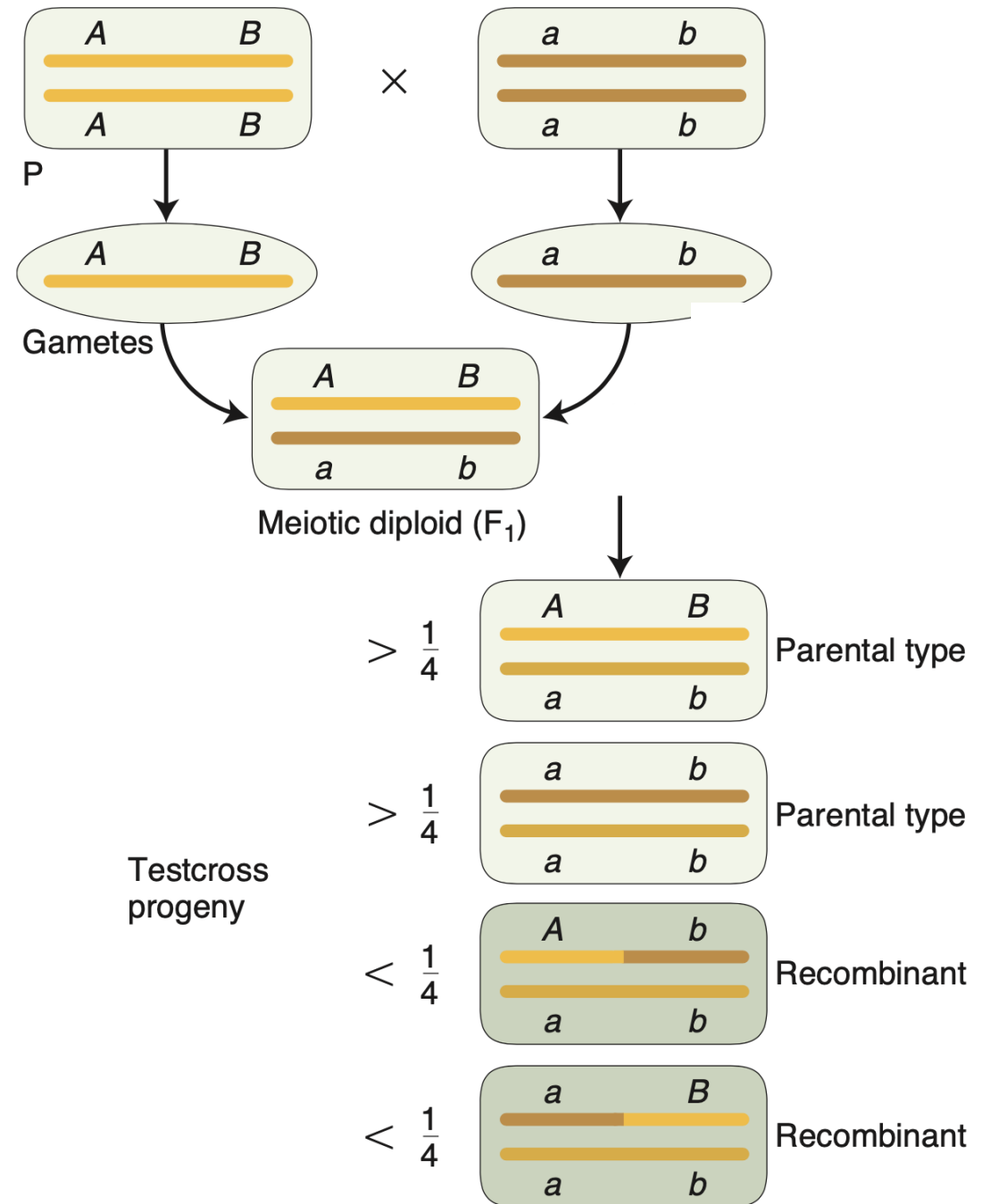
# LINKAGE

## 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 <sub>1</sub> | <i>A B</i> | 37 | parental     |
|                | <i>a b</i> | 43 |              |
|                | <i>A b</i> | 9  | recombinante |
|                | <i>a B</i> | 11 |              |

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

# 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)

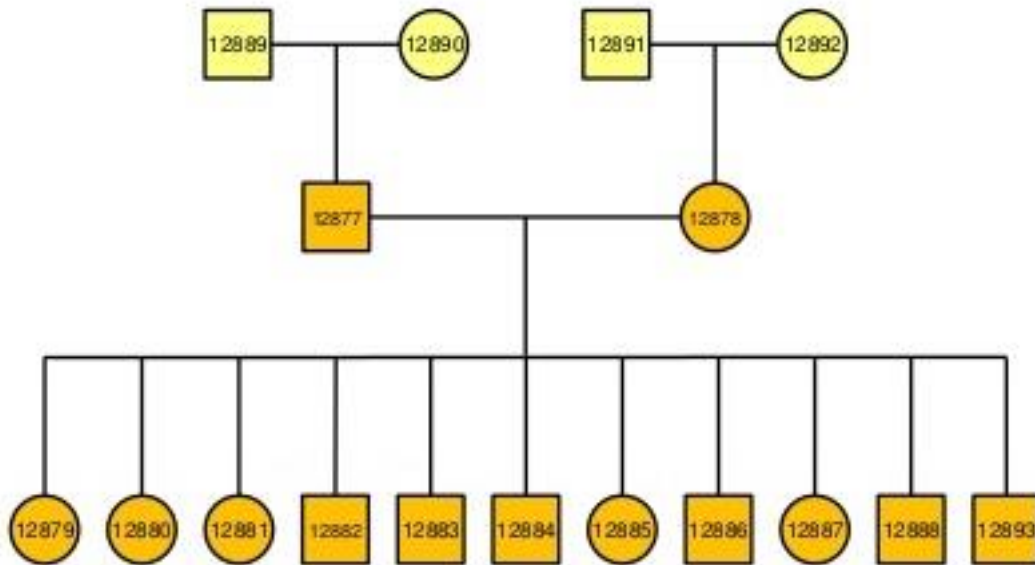
10 cM = 10% probability of crossover in each meiosis

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



scute bristles, *sc*  
white eyes, *w*

ruby eyes, *rb*

crossveinless wings, *cv*

singed bristles, *sn*

lozenge eyes, *lz*

vermilion eyes, *v*

sable body, *s*

scalloped wings, *sd*

*Bar* eyes, *B*

carnation eyes, *car*

little fly, *lf*



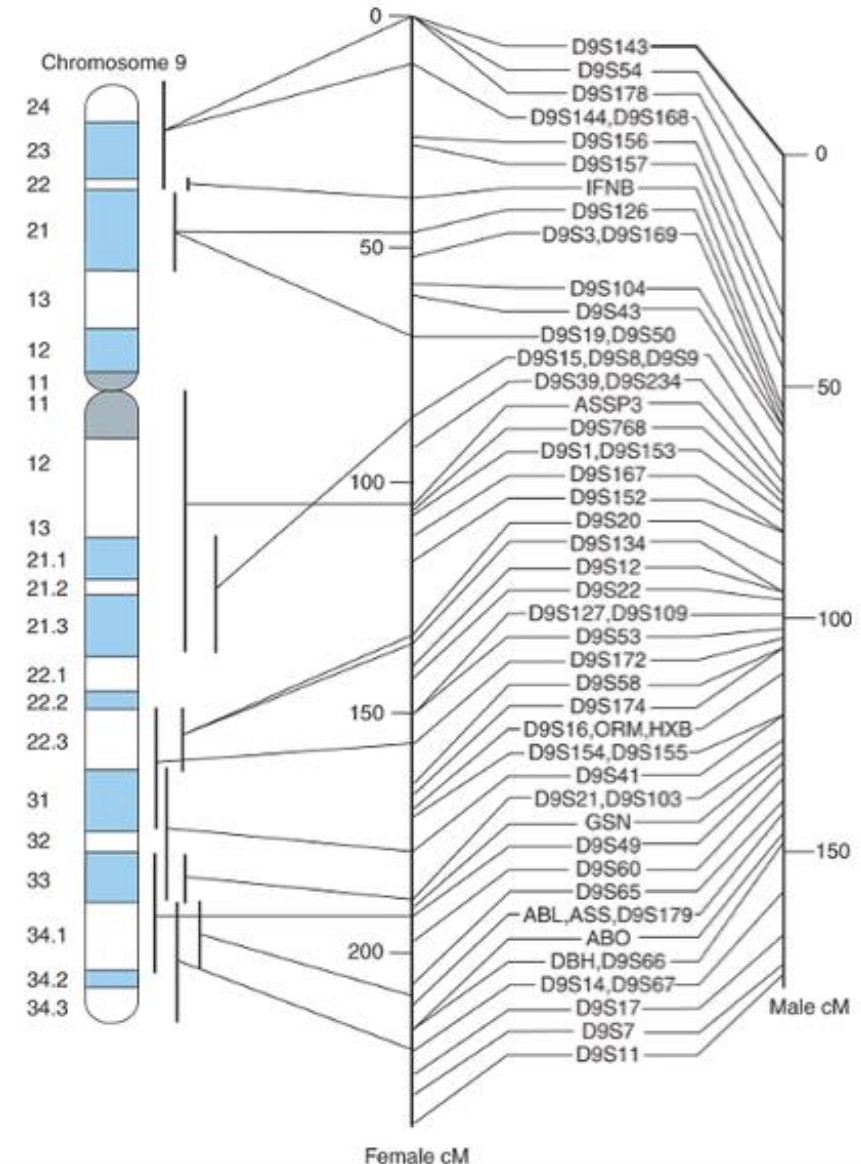
# THE MAP OF THE HUMAN GENOME

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)

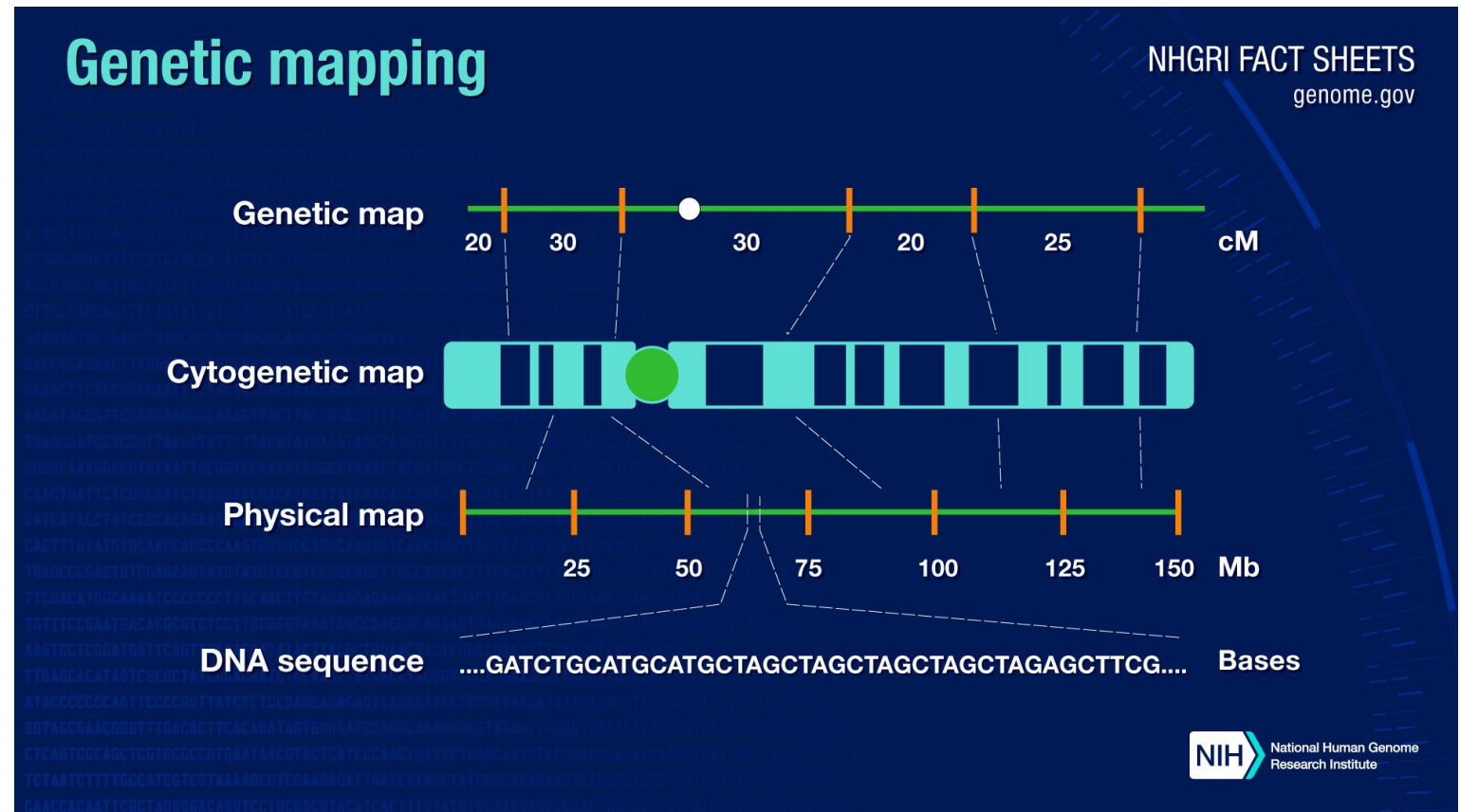


# TWO TYPES OF MAPS

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 – 09:00    Recap + Exercises E15 [Part III]
- 09:00 – 09:10    Break
- 09:10 – 09:30    Lecture 1 [*Genetic risk assessment*]
- 09:30 – 10:00    Group work
- 10:00 – 10:40    Break + Exercises 1 [1-3]
- 10:40 – 11:15    Lecture 2 [*Linkage*]
- 11:15 – 11:55    Break + Exercises 2 [4-8]
- 11:55 – 12:00    Reflection

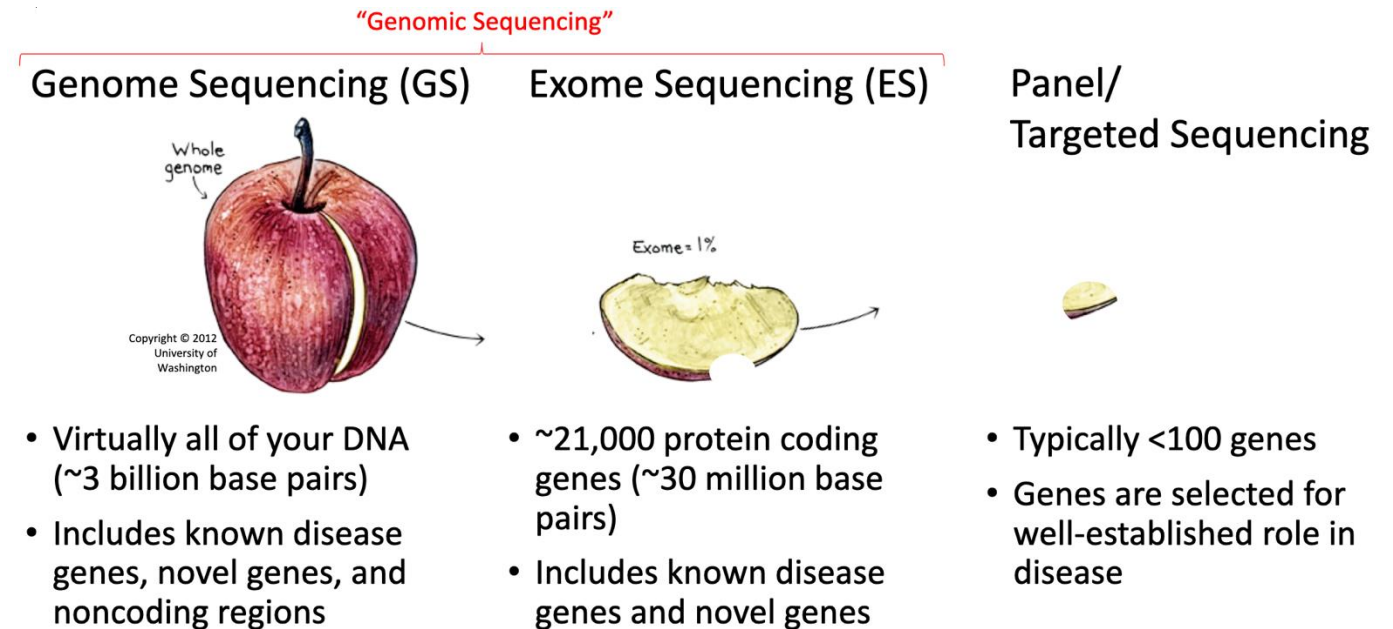


# OUTLINE

|                      |  |
|----------------------|--|
| 08:15 – 09:00        | Recap + Exercises E15 [Part III]             |
| 09:00 – 09:10        | Break  |
| 09:10 – 09:30        | Lecture 1 [ <i>Genetic risk assessment</i> ] |
| 09:30 – 10:00        | Group work                                   |
| 10:00 – 10:40        | Break + Exercises 1 [1-3]                    |
| 10:40 – 11:15        | Lecture 2 [ <i>Linkage</i> ]                 |
| 11:15 – 11:55        | Break + Exercises 2 [4-8]                    |
| <b>11:55 – 12:00</b> | <b>Reflection</b>                            |

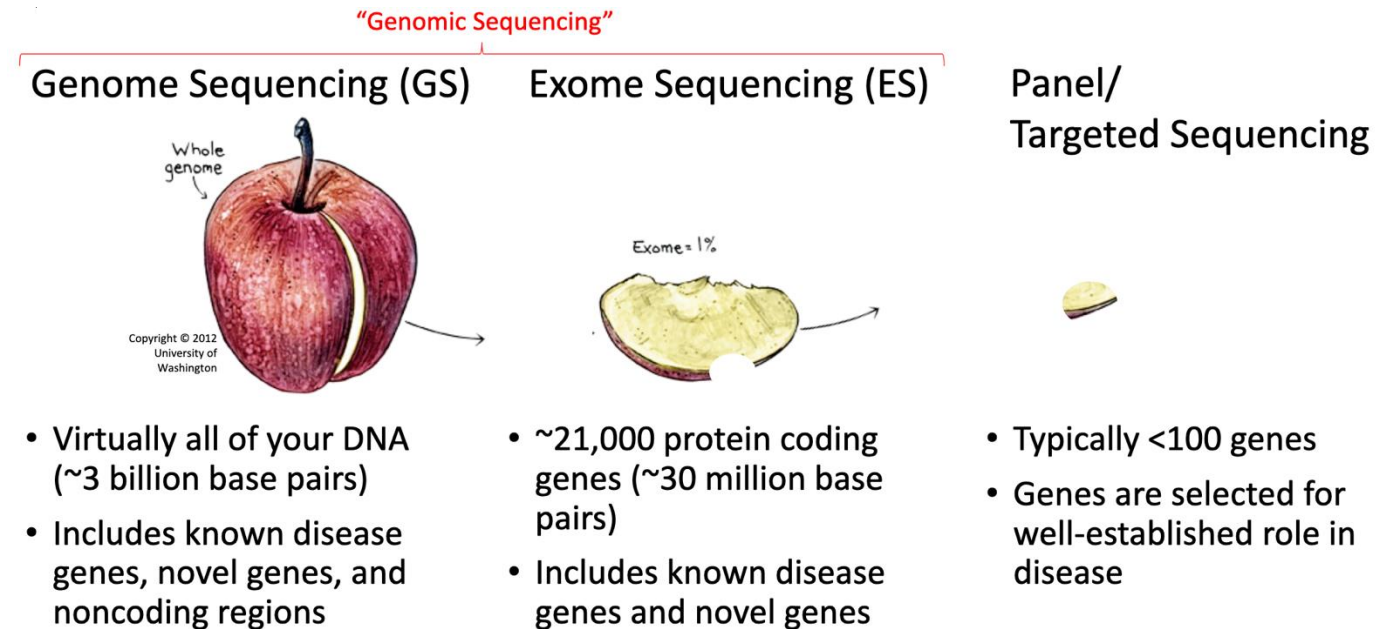
# CLINICAL APPLICATION OF NEXT-GENERATION SEQUENCING (NGS)

- ▶ 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



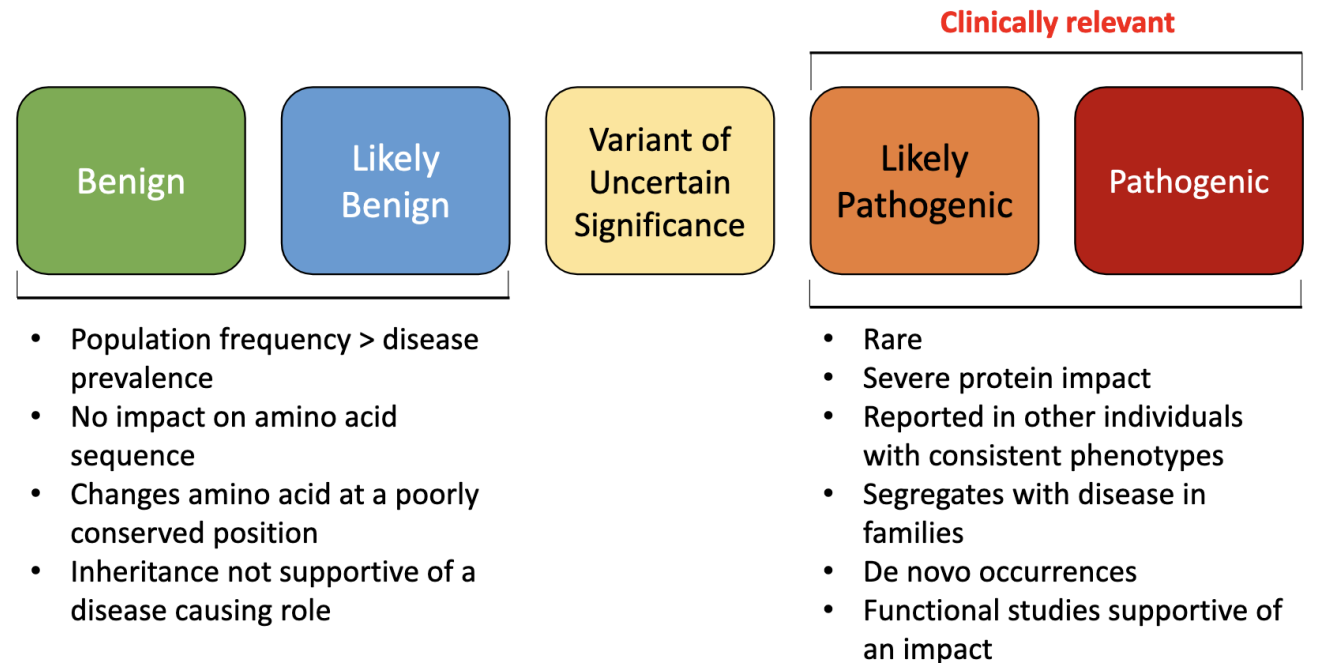
# CLINICAL APPLICATION OF NEXT-GENERATION SEQUENCING (NGS)

- ❖ 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



# 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)





# REFLECT TOGETHER 2 AND 2



- What will you remember from today?
- What did you find difficult?
- What do you need to follow-up on?
- Feedback / Questions to PDR?

# NEXT TIME

