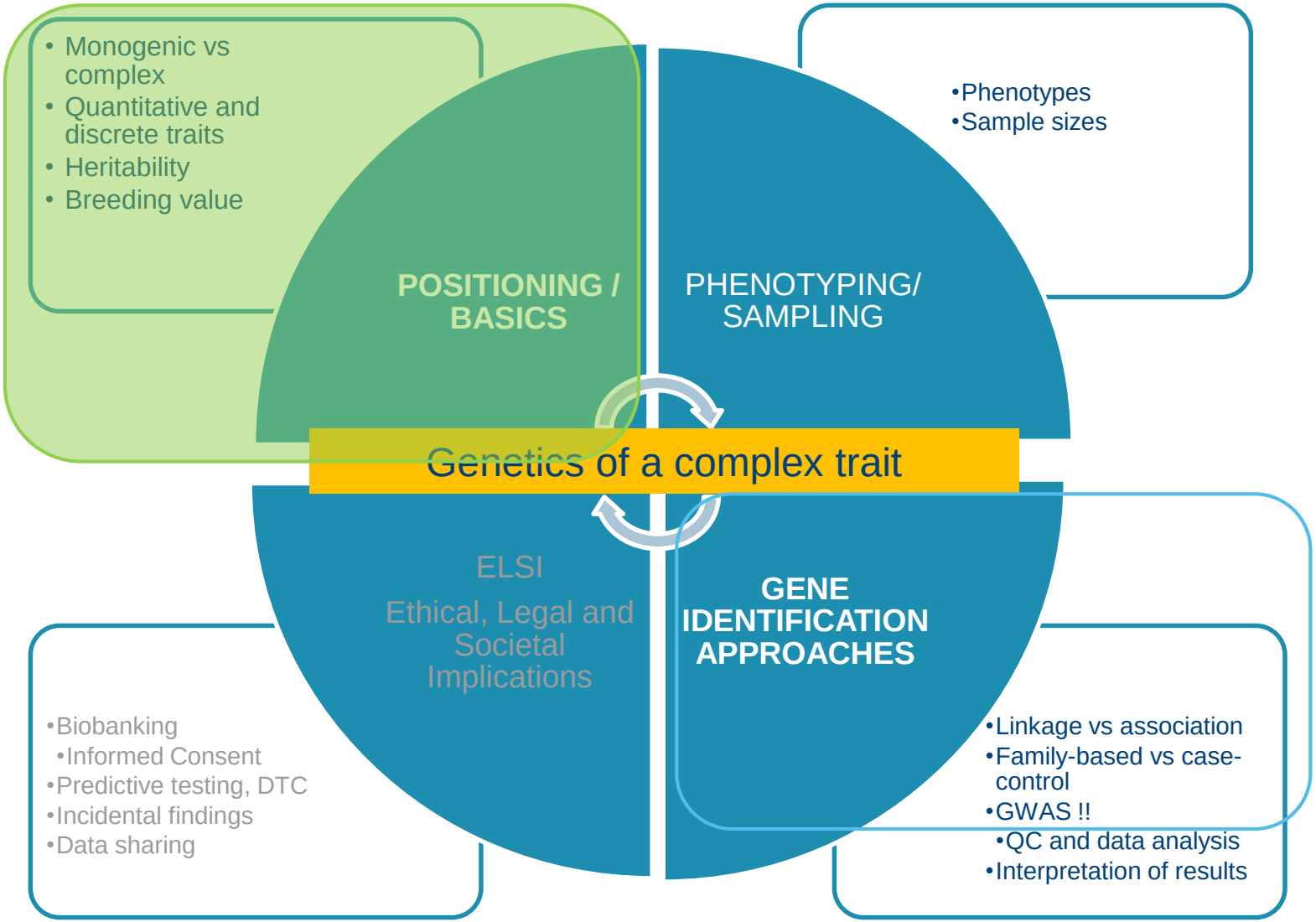
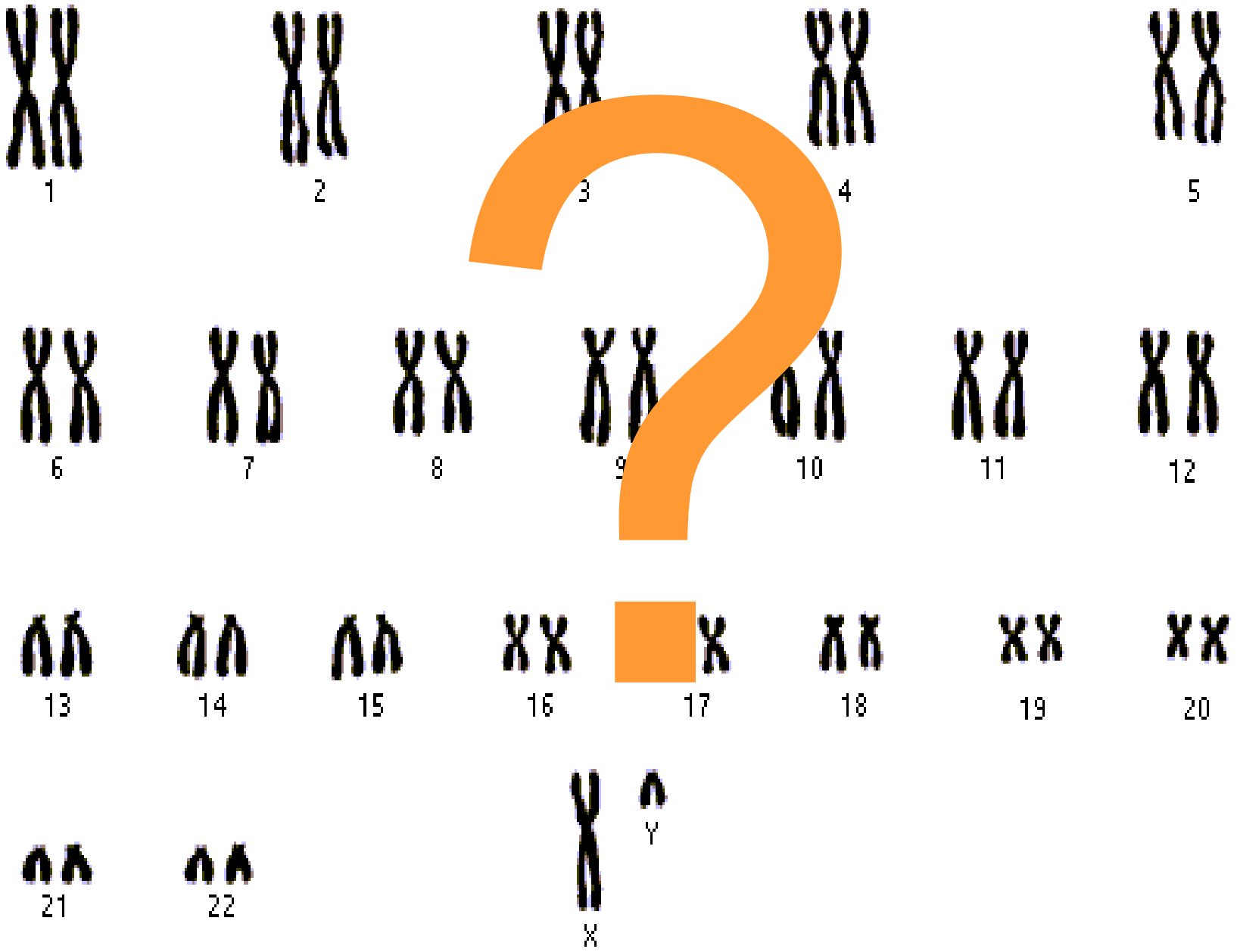
Quantitative genetics

2. Gene identification approaches

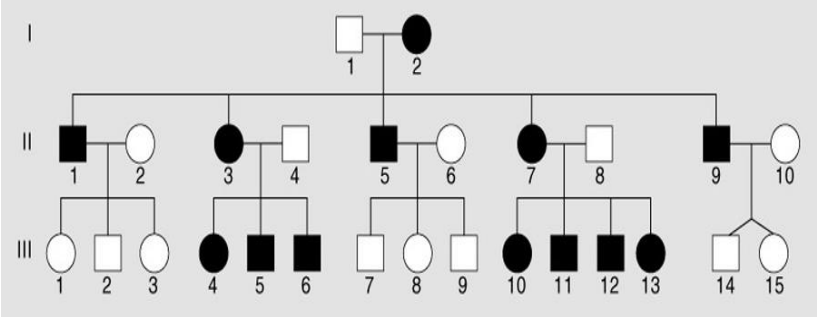
KU LEUVEN


• Monogenic vs complex
• Quantitative and discrete traits
• Heritability
• Breeding value
POSITIONING / BASICS
•Phenotypes
•Sample sizes
PHENOTYPING/ SAMPLING
Genetics of a complex trait
ELSI
Ethical, Legal and Societal Implications
GENE IDENTIFICATION APPROACHES
•Biobanking
•Informed Consent
•Predictive testing, DTC
•Incidental findings
•Data sharing
•Linkage vs association
•Family-based vs case-control
•GWAS !!
•QC and data analysis
•Interpretation of results


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y

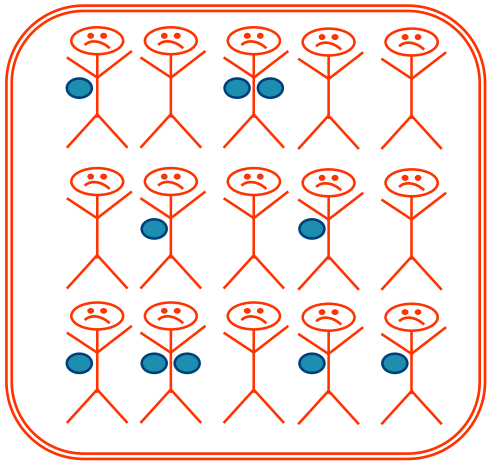
Linkage versus association

- Linkage studies

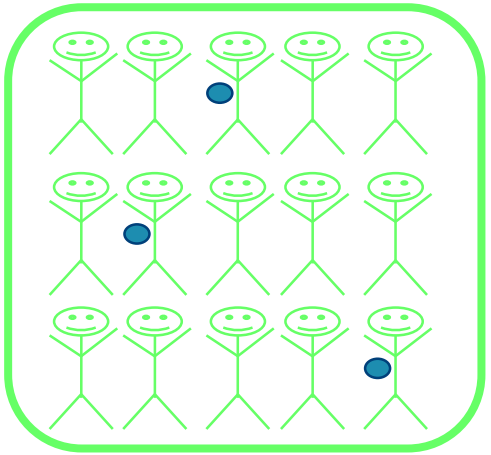


- Association studies

cases



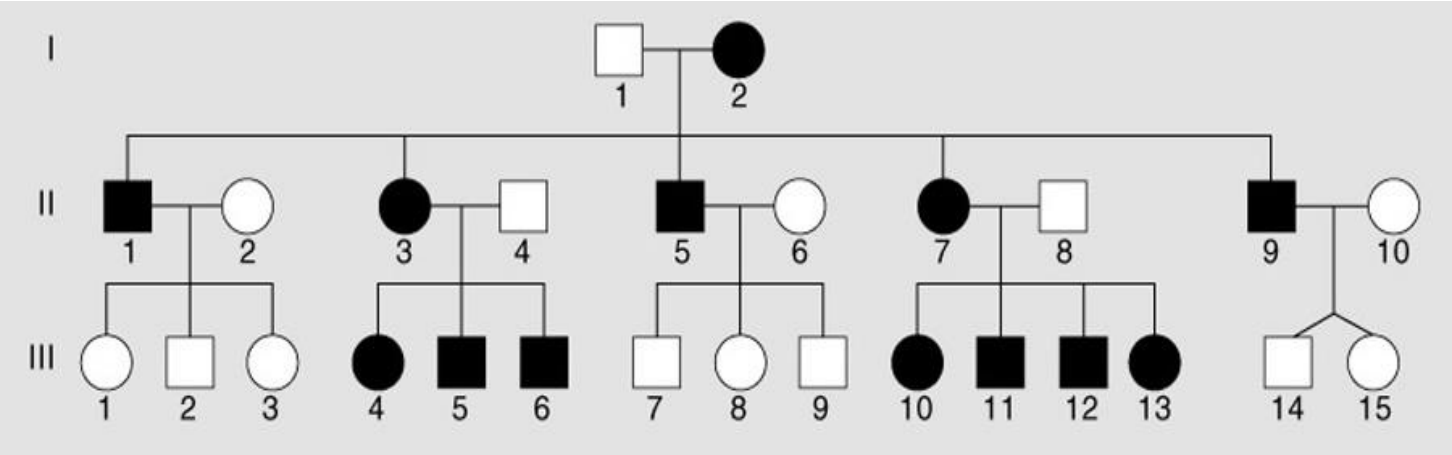
controls



KU LEUVEN

Linkage analysis

- co-segregation of genetic marker(s) and disease in a family

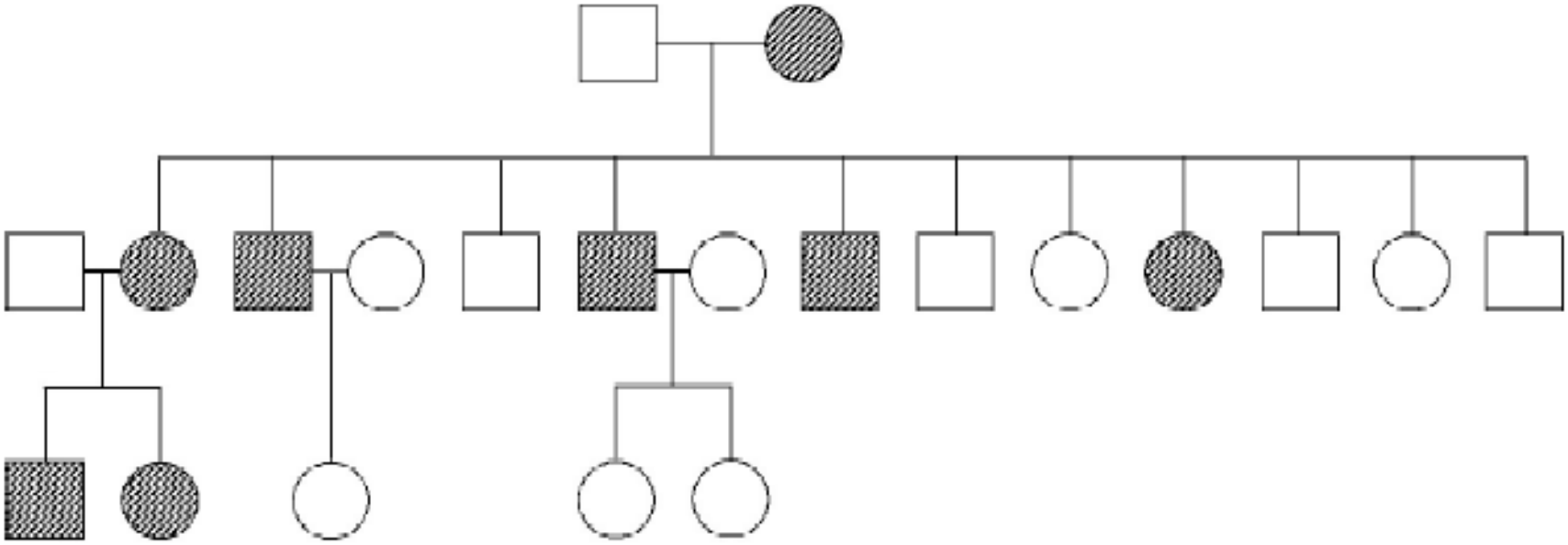


- Goal:
  - identify chromosomal region with risk gene
  - map traits of interest to particular chromosomal region

- What is linkage?
- How to quantify?
- How to use it to identify chromosomal region(s) with disease gene(s)?

Example – nail patella syndrome (NPS)

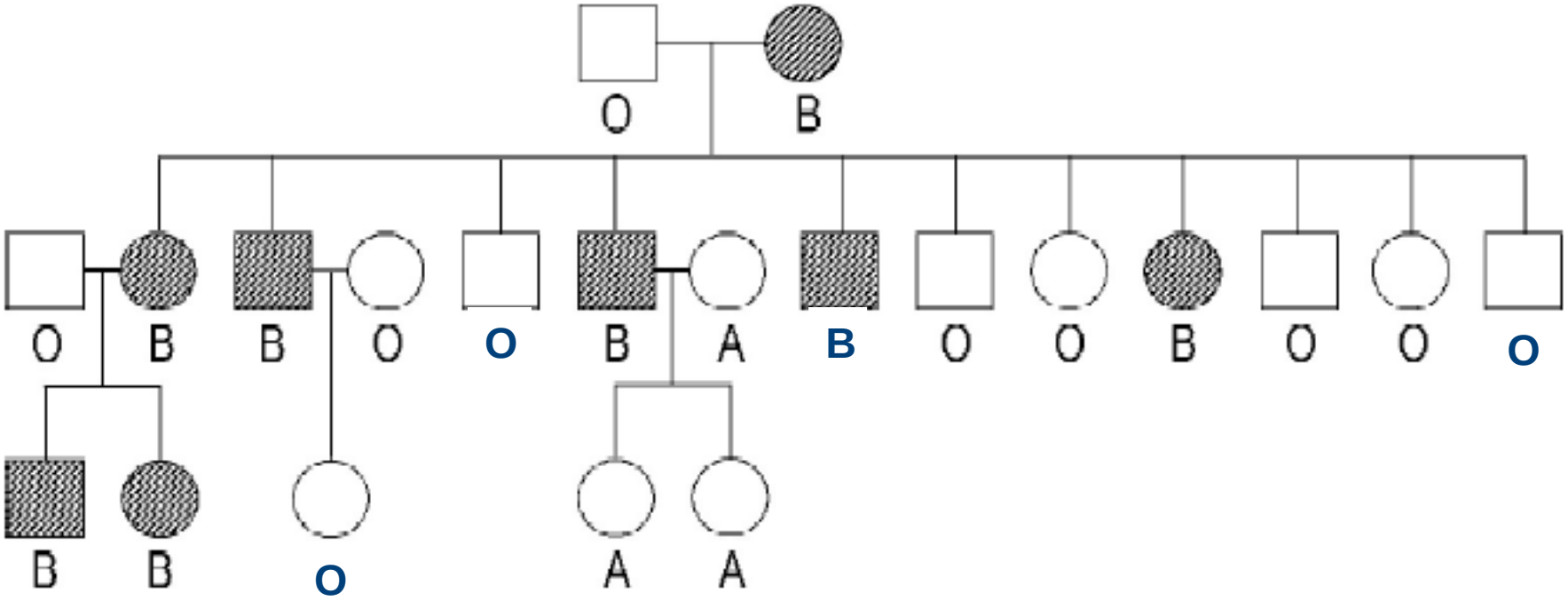
Example – Nail patella syndrome



Autosomal dominant

Example – Nail patella syndrome

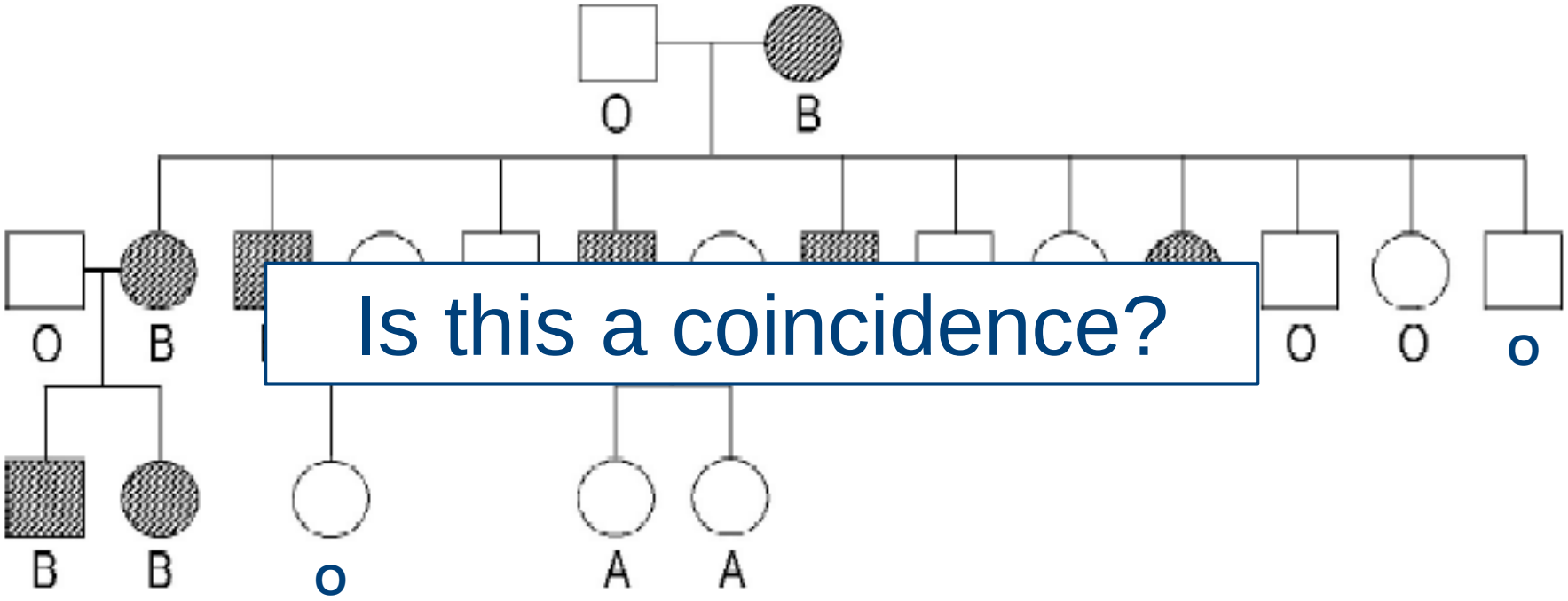
ABO-blood type



All individuals with NPS in this family have the same blood type

Example – Nail patella syndrome

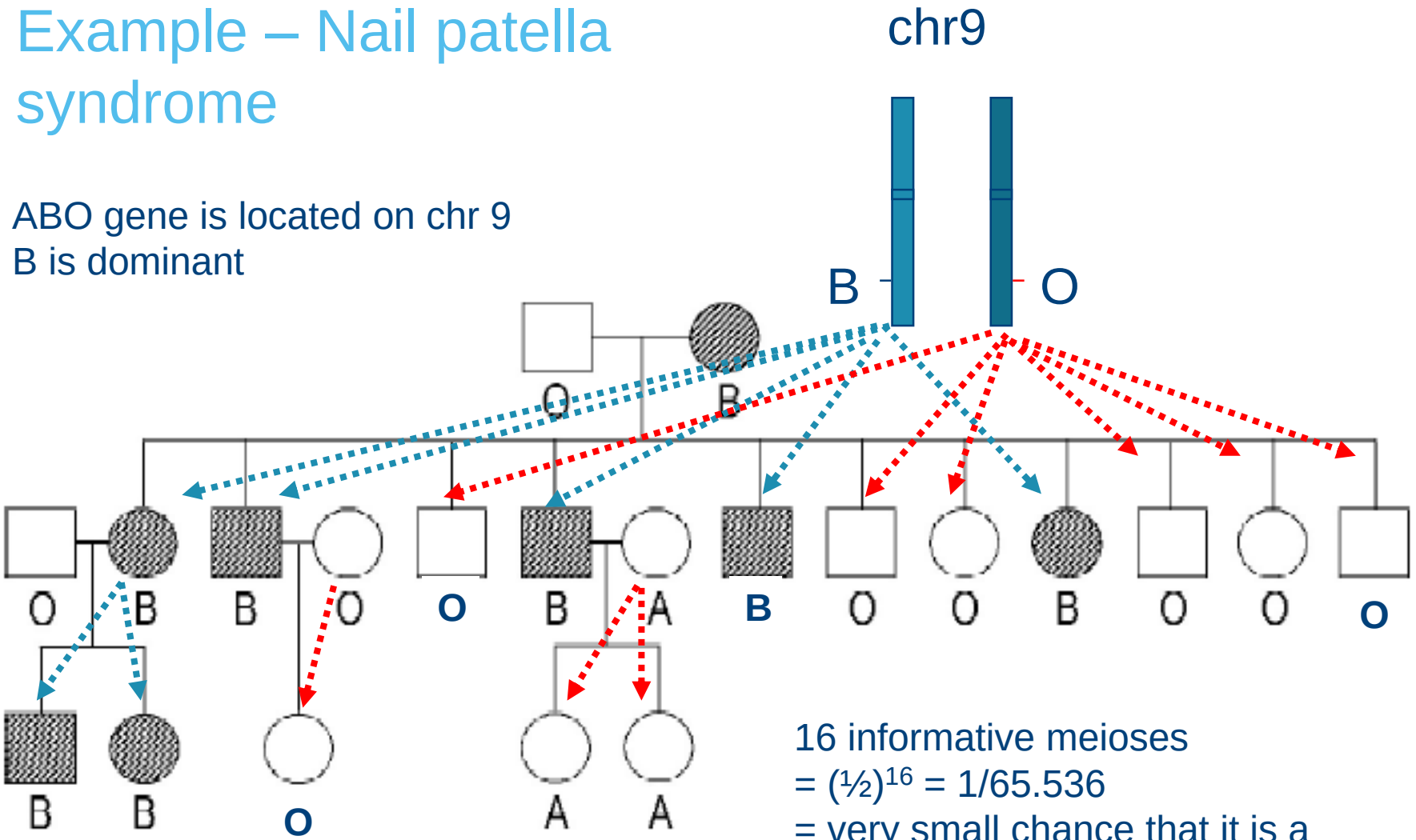
ABO-blood type



All individuals with NPS in this family have the same blood type

Example – Nail patella syndrome

- ABO gene is located on chr 9
- B is dominant

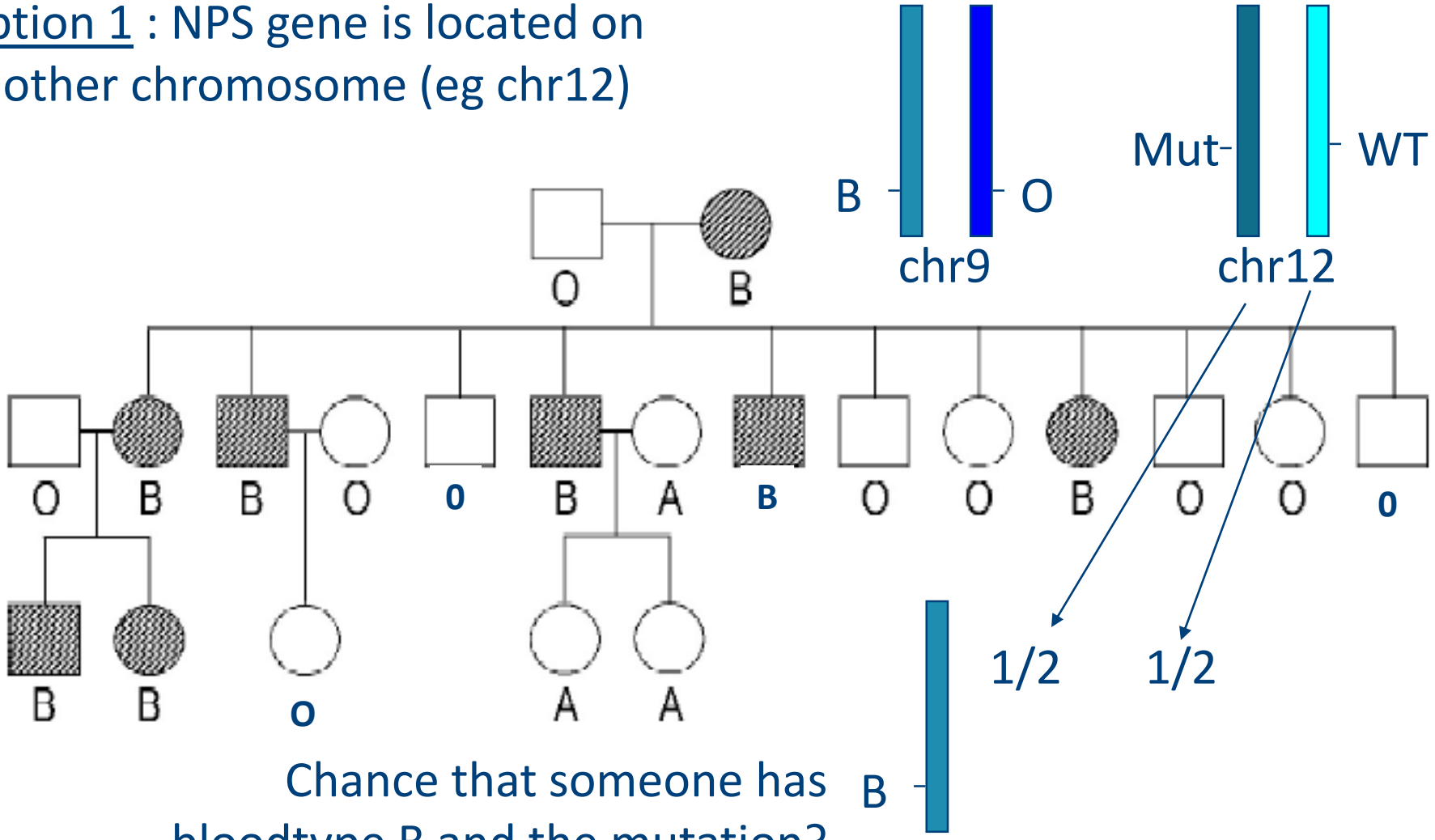


16 informative meioses
= $(½)^{16}$ = 1/65.536
= very small chance that it is a coincidence that NPS is always co-inherited with blood type

KU LEUVEN

Where is the NPS gene located?

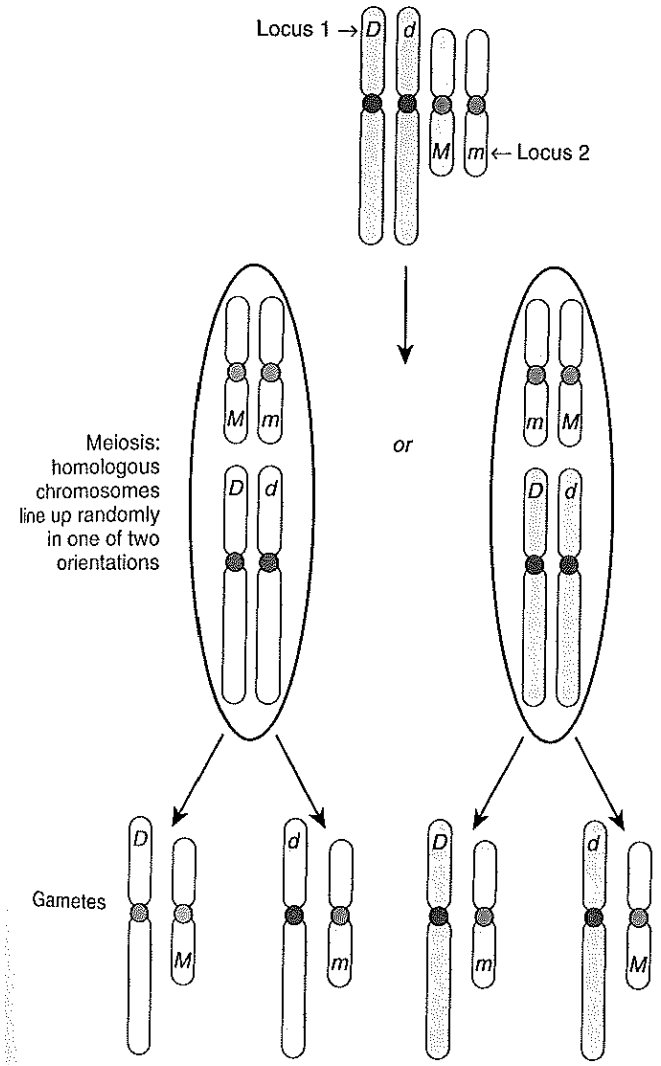
Option 1 : NPS gene is located on another chromosome (eg chr12)



Chance that someone has bloodtype B and the mutation?

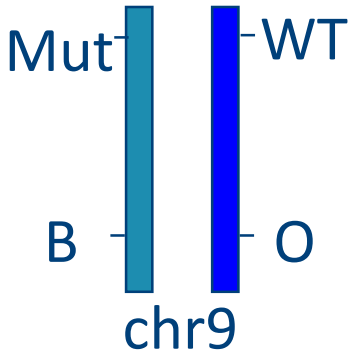
Law of independent assortment (Mendel)

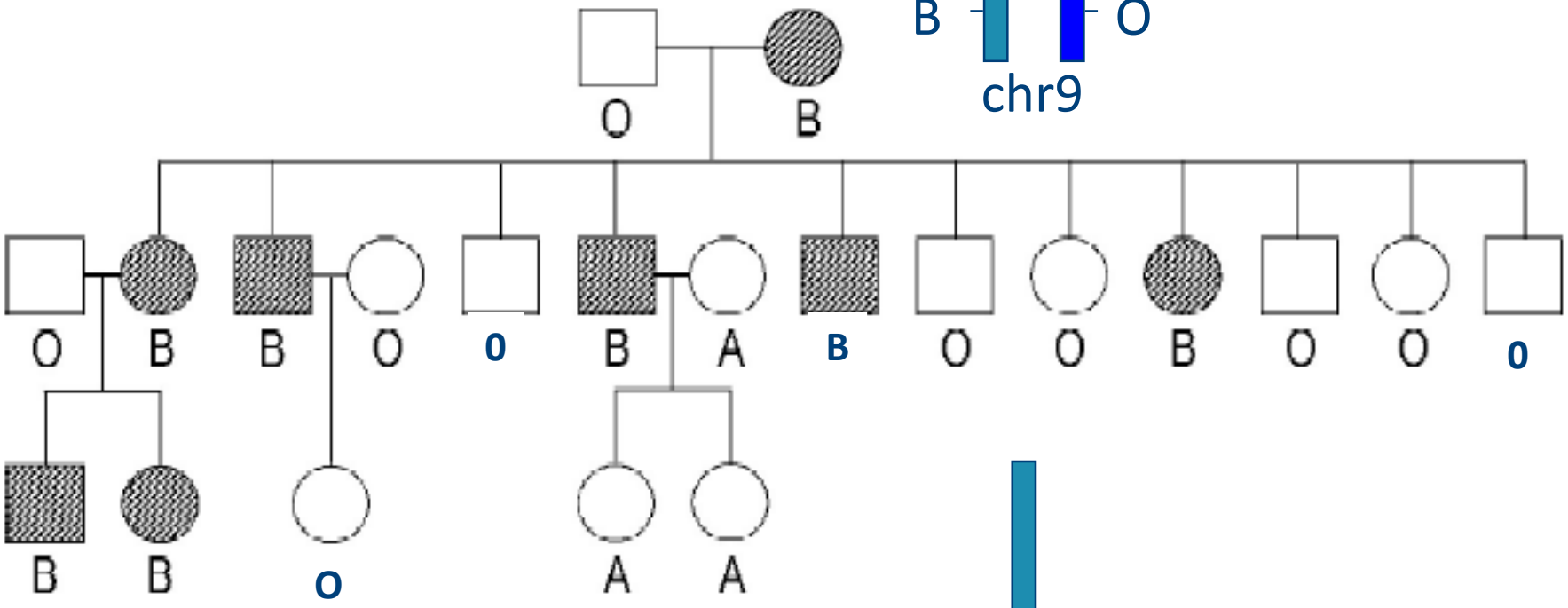
Alleles at loci on different chromosomes assort independently



Where is the NPS gene located?

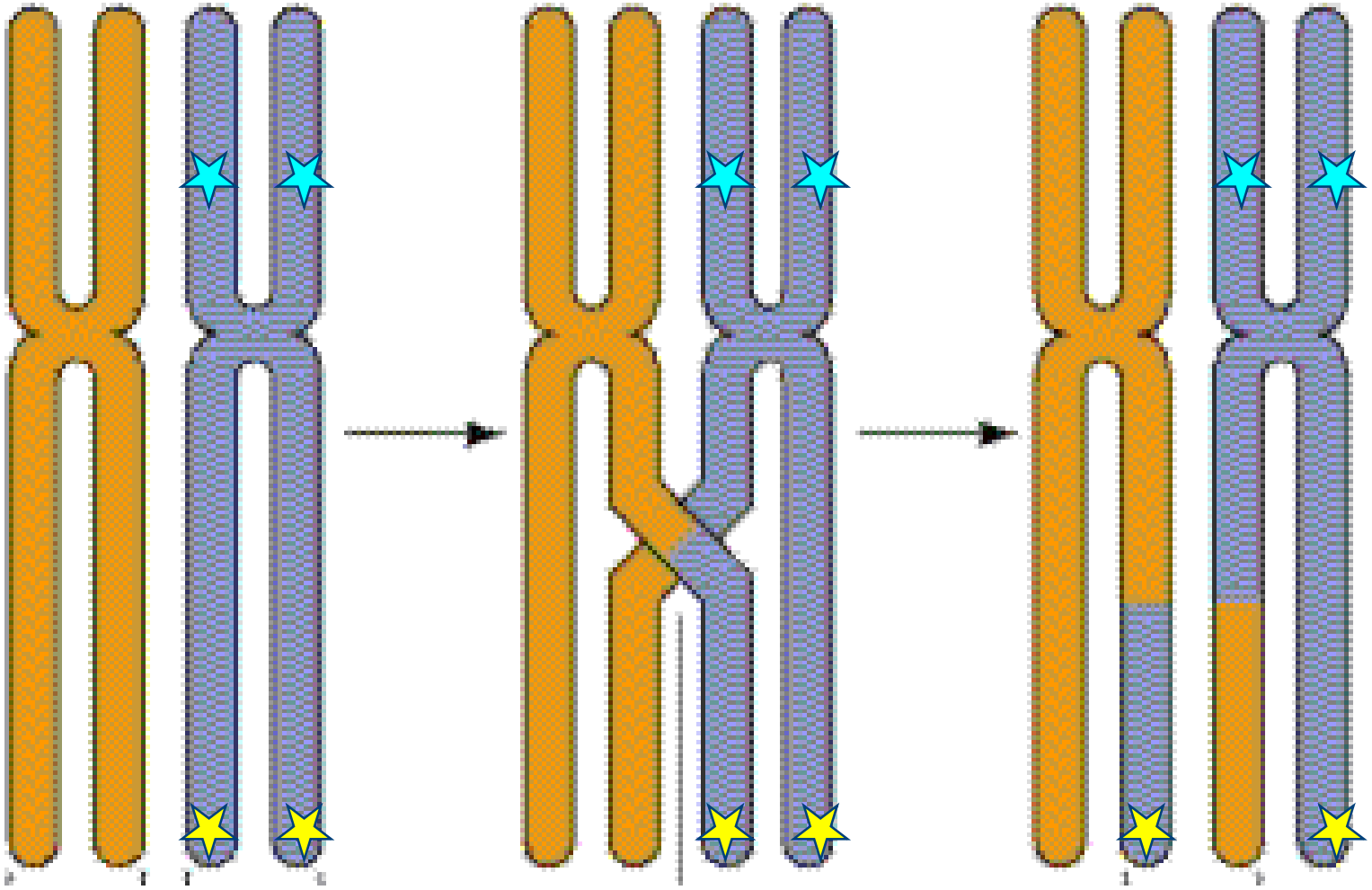
Option 1 : NPS gene is located on same chromosome, far from ABO gene





Chance that someone has bloodtype B and the mutation?

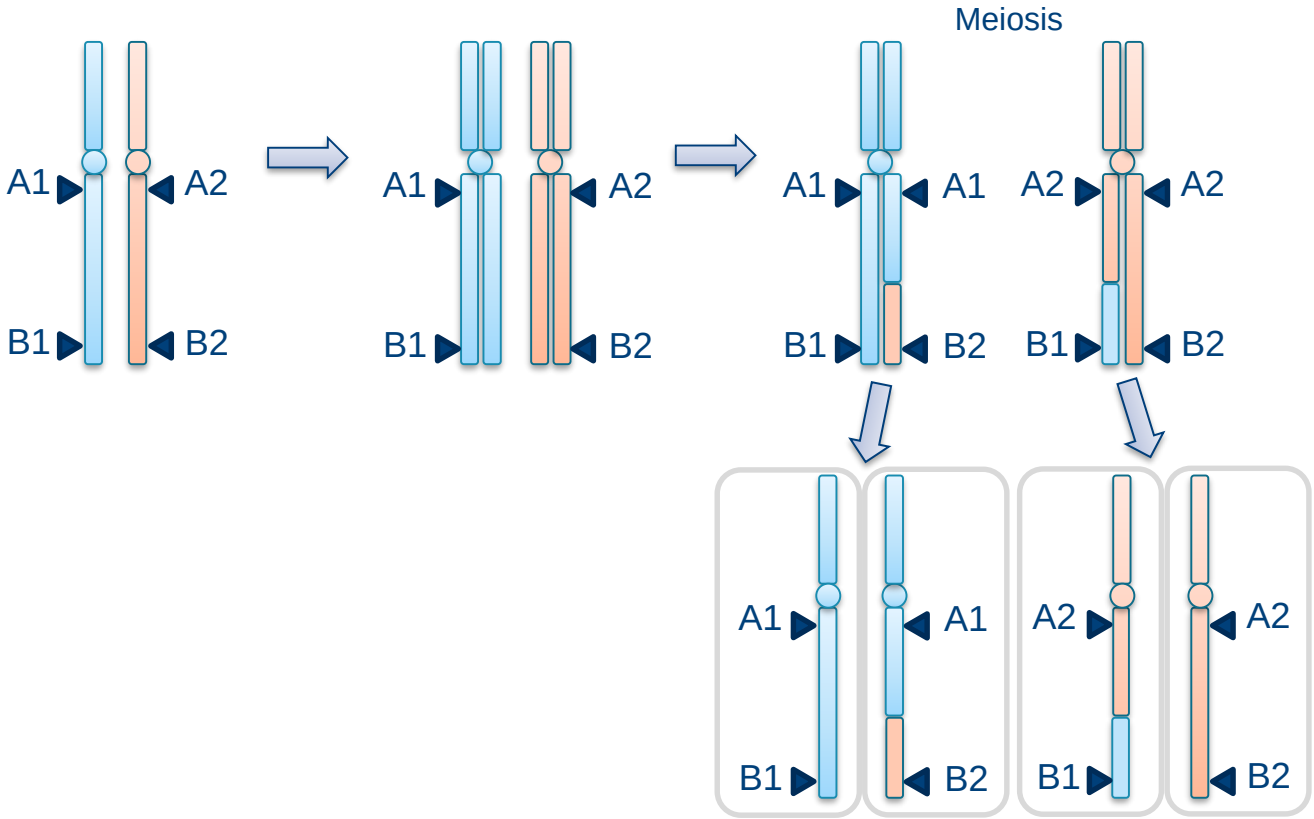
50%



CROSSING-OVER
=> are not co-inherited

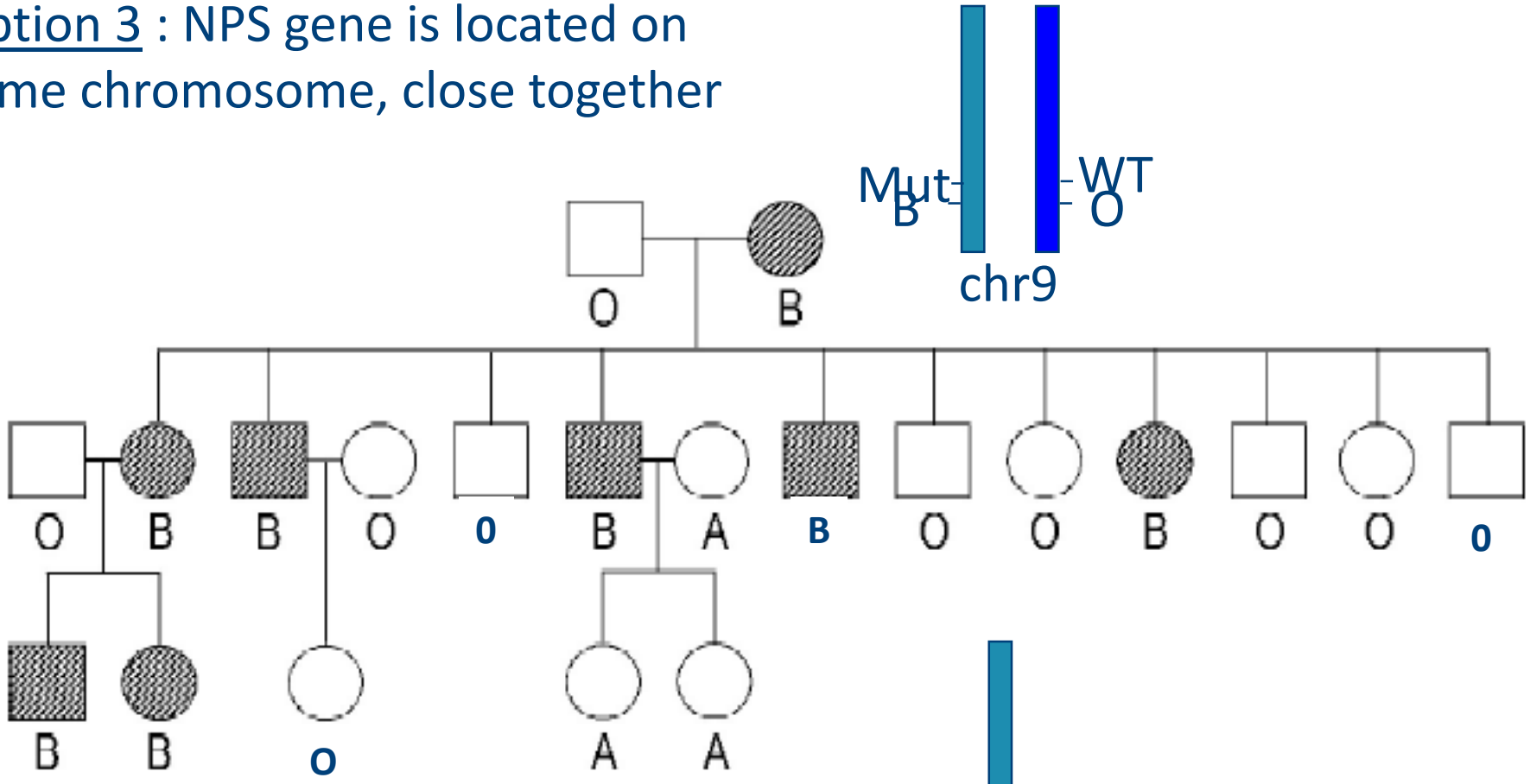
KU LEUVEN

Recombination during meiosis



Where is the NPS gene located?

Option 3 : NPS gene is located on same chromosome, close together

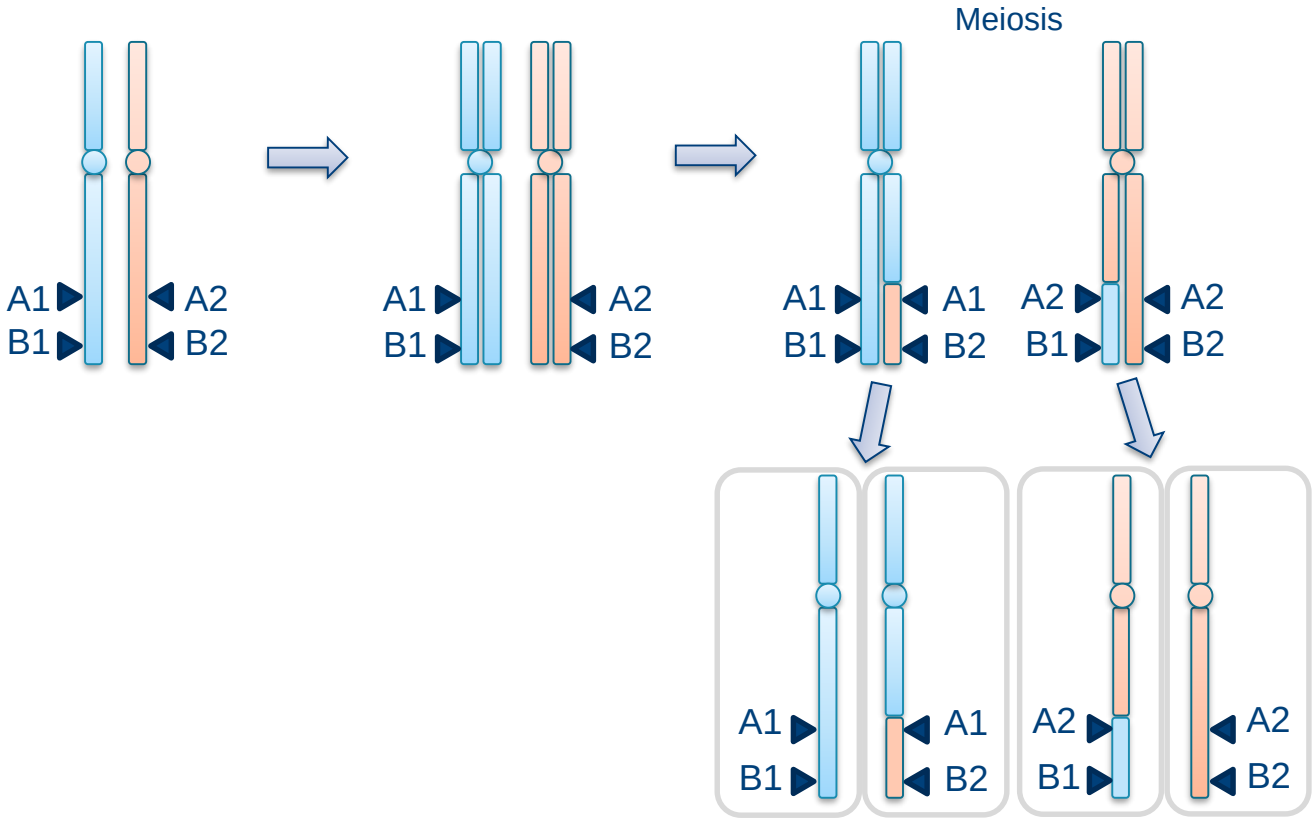


Chance that someone has bloodtype B and the mutation?

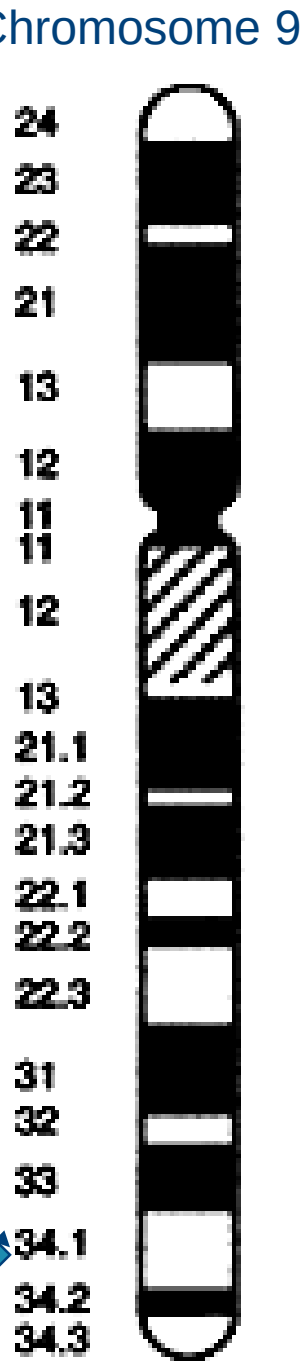
> 50%

KU LEUVEN

Recombination during meiosis



Where is the NPS gene located?



ABO blood type gene

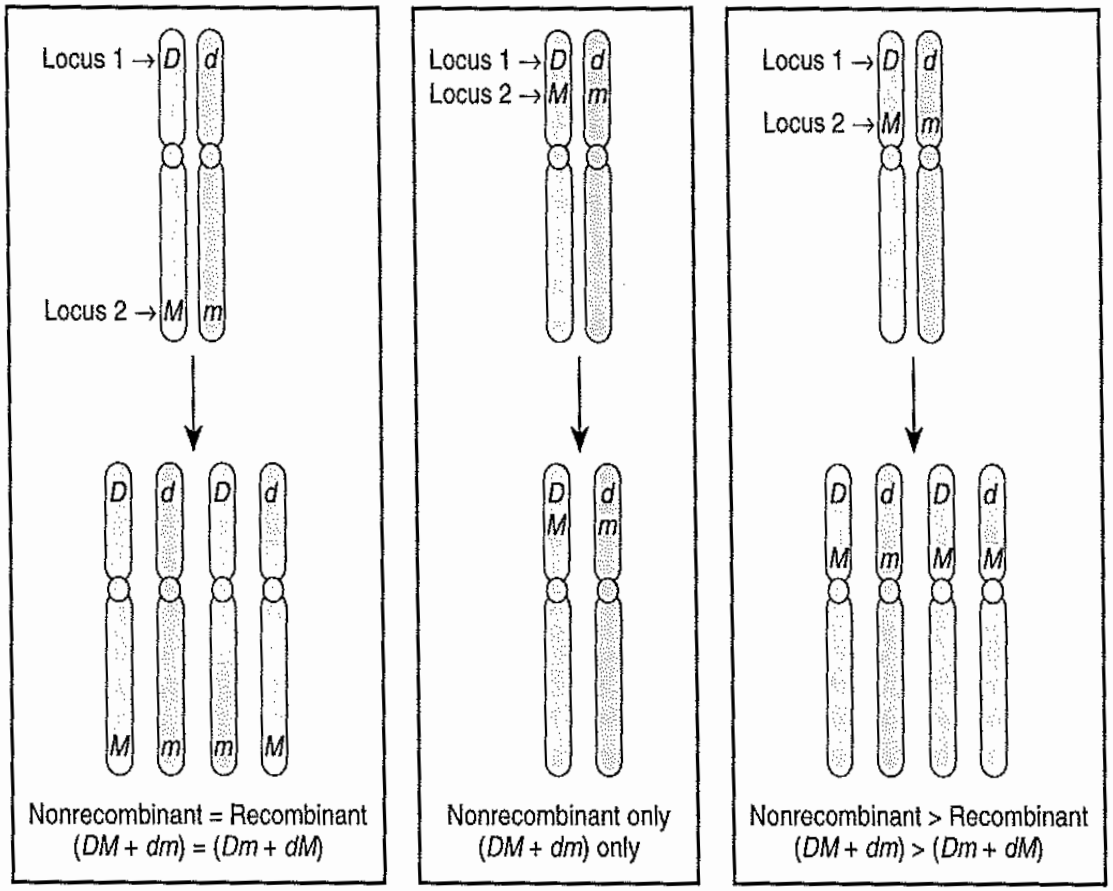
Nail Patella Syndrome gene

Linkage

- When genes are found on different chromosomes or far apart on the same chromosome, they assort independently = **unlinked**
- When genes are close together on the same chromosome, the alleles, or gene versions, will be inherited as a unit more frequently than not = **linked**

- How can we measure this?

Recombination during meiosis



50% 0% 0%<θ<50% Recombination fraction

The recombination fraction (θ) between two loci is a measure for the distance between these two loci

KU LEUVEN

Linkage

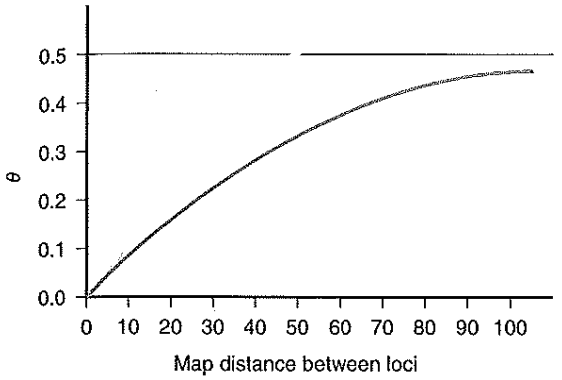
- When genes are found on different chromosomes or far apart on the same chromosome, they assort independently = **unlinked**
- When genes are close together on the same chromosome, the alleles, or gene versions, will be inherited as a unit more frequently than not = **linked**

- How can we measure this?
  - use data from genetic crosses (‘meiotic mapping’) to calculate the **recombination fraction (θ)**
  - If you do this for many gene/marker pairs, we can make **linkage maps** that show the relative distances of the genes on the chromosome

Recombination fraction θ

$$\theta = \frac{number\ of\ recombined\ gametes\ (r)}{number\ of\ gametes\ transmitted\ (n)}$$

Parent	Gametes	θ
	50% non-rec and 50% rec	0.5
	90% non-rec and 10% rec	0.1
	99% non-rec and 1% rec	0.01
	100% non-rec	0



Genetic distance

- Chromosomal position: physical vs. genetic
  - physical: base pair position (bp)
  - genetic: recombination fraction (Morgan)

- Mapping function

Kosambi (1943): $W = \frac{1}{4}\ln\left(\frac{1+2\theta}{1-2\theta}\right)$

- Globally: 1cM = 1Mb = 1% recombination
  - recombination deserts: 0.3 cM/Mb
  - recombination jungles: 3 cM/Mb

Recombination fraction θ

$$\theta = \frac{number\ of\ recombined\ gametes\ (r)}{number\ of\ gametes\ transmitted\ (n)}$$

- Recombination fraction (θ) as a measure for linkage
- $0 \leq \theta < \frac{1}{2}$?

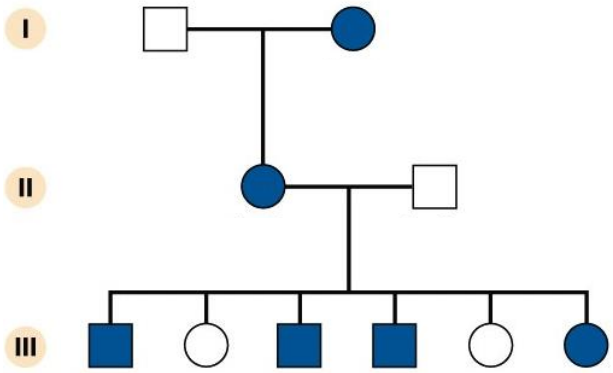
50% chance of recombination when completely unlinked

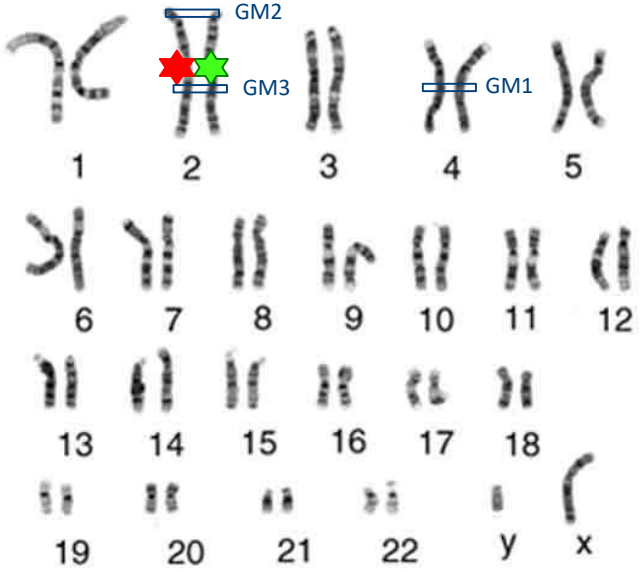
- What is linkage?
- How to quantify?
- **How to use it to identify chromosomal region(s) with disease gene(s)?**

KU LEUVEN

Linkage analysis

- Linkage analysis is a method that is used to decide if two loci, or a genetic marker (GM) and a disease gene, are linked :
  - Ascertain whether the recombination fraction θ between the two loci deviates significantly from 0.5





KU LEUVEN

Linkage analysis – how?

1. Collect familie(s)
2. Choose a genetic marker (eg microsatellite, SNP) for linkage analysis
3. Genotype this marker in all individuals of the familie(s)
4. Identify informative meioses (R vs NR → between heterozygous GM and heterozygous disease locus)
   - Based on the phenotype you know the disease locus genotype (affected vs unaffected)
   - Is phase known?
5. Determine whether there is linkage between the GM locus and the disease locus
   - Determine recombination fraction θ

$$log_{10}\left[\frac{Likelihood\ of\ linkage\ (\theta)}{Likelihood\ that\ loci\ are\ unlinked\ \left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{L\ (\theta)}{L\left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{(1-\theta)^{NR} * \theta^{R}}{0.5^{(R+NR)}}\right]$$ = max

   - Is this recombination significantly different from 0.5?

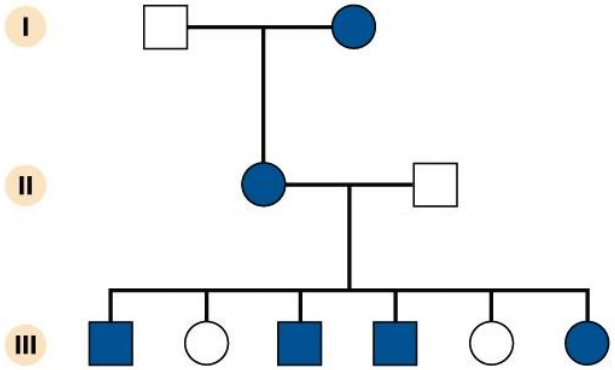
     (= null hypothesis; no linkage between GM and disease locus)

$$log_{10}\left[\frac{Likelihood\ of\ linkage\ (\theta)}{Likelihood\ that\ loci\ are\ unlinked\ \left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{L\ (\theta)}{L\left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{(1-\theta)^{NR} * \theta^{R}}{0.5^{(R+NR)}}\right]$$ ≥ 3 ?

6. Do we need more informative meioses / families?

KU LEUVEN

Collect families



KU LEUVEN

Linkage analysis – how?

1. Collect familie(s)
2. Choose a genetic marker (eg microsatellite, SNP) for linkage analysis
3. Genotype this marker in all individuals of the familie(s)
4. Identify informative meioses (R vs NR → between heterozygous GM and heterozygous disease locus)
   - Based on the phenotype you know the disease locus genotype (affected vs unaffected)
   - Is phase known?
5. Determine whether there is linkage between the GM locus and the disease locus
   - Determine recombination fraction θ

$$log_{10}\left[\frac{Likelihood\ of\ linkage\ (\theta)}{Likelihood\ that\ loci\ are\ unlinked\ \left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{L\ (\theta)}{L\left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{(1-\theta)^{NR} * \theta^{R}}{0.5^{(R+NR)}}\right]$$ = max

   - Is this recombination significantly different from 0.5?

     (= null hypothesis; no linkage between GM and disease locus)

$$log_{10}\left[\frac{Likelihood\ of\ linkage\ (\theta)}{Likelihood\ that\ loci\ are\ unlinked\ \left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{L\ (\theta)}{L\left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{(1-\theta)^{NR} * \theta^{R}}{0.5^{(R+NR)}}\right]$$ ≥ 3 ?

6. Do we need more informative meioses / families?

KU LEUVEN

Linkage analysis: genetic markers

- Requirements
  - polymorphic
  - informative = high mean heterozygosity

- Types: RFLPs, microsatellites & SNPs

- Advantages
  - easy to detect
  - relatively inexpensive to test easily available material

Linkage analysis – how?

1. Collect familie(s)
2. Choose a genetic marker (eg microsatellite, SNP) for linkage analysis
3. Genotype this marker in all individuals of the familie(s)
4. Identify informative meioses (R vs NR → between heterozygous GM and heterozygous disease locus)
   - Based on the phenotype you know the disease locus genotype (affected vs unaffected)
   - Is phase known?
5. Determine whether there is linkage between the GM locus and the disease locus
   - Determine recombination fraction θ

$$log_{10}\left[\frac{Likelihood\ of\ linkage\ (\theta)}{Likelihood\ that\ loci\ are\ unlinked\ \left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{L\ (\theta)}{L\left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{(1-\theta)^{NR} * \theta^{R}}{0.5^{(R+NR)}}\right]$$ = max

   - Is this recombination significantly different from 0.5?

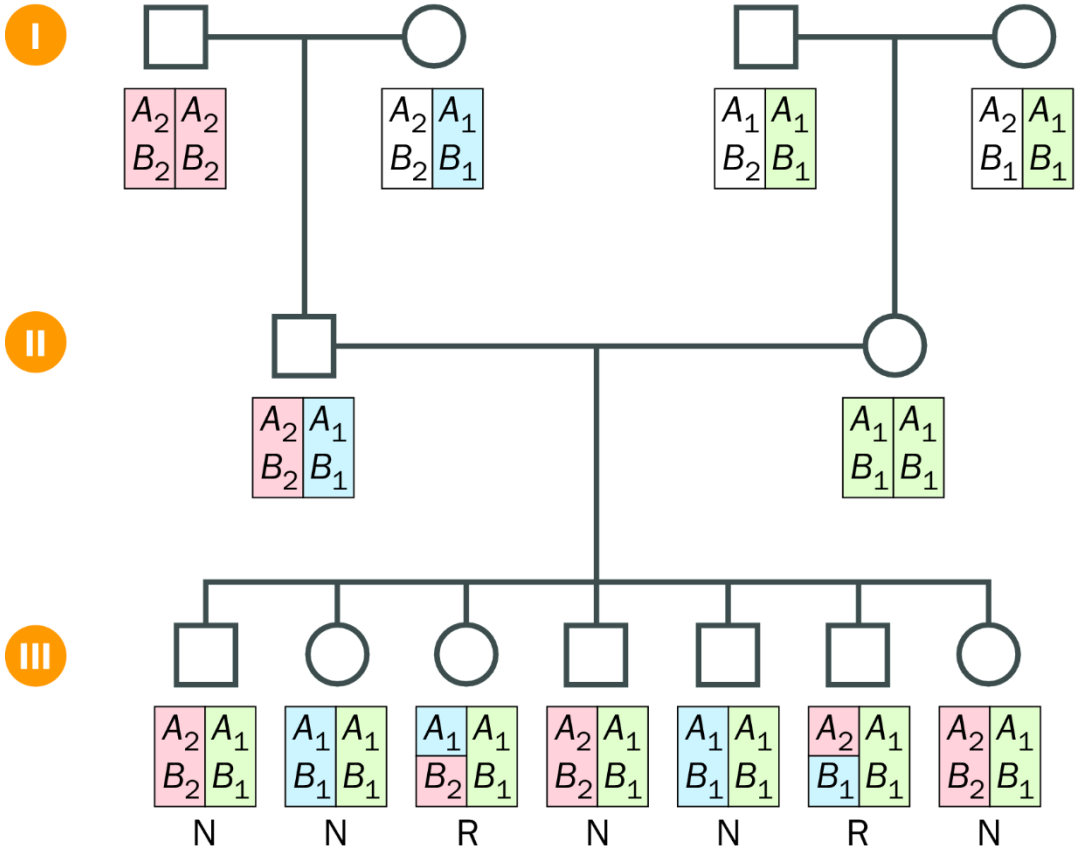
     (= null hypothesis; no linkage between GM and disease locus)

$$log_{10}\left[\frac{Likelihood\ of\ linkage\ (\theta)}{Likelihood\ that\ loci\ are\ unlinked\ \left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{L\ (\theta)}{L\left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{(1-\theta)^{NR} * \theta^{R}}{0.5^{(R+NR)}}\right]$$ ≥ 3 ?

6. Do we need more informative meioses / families?

KU LEUVEN

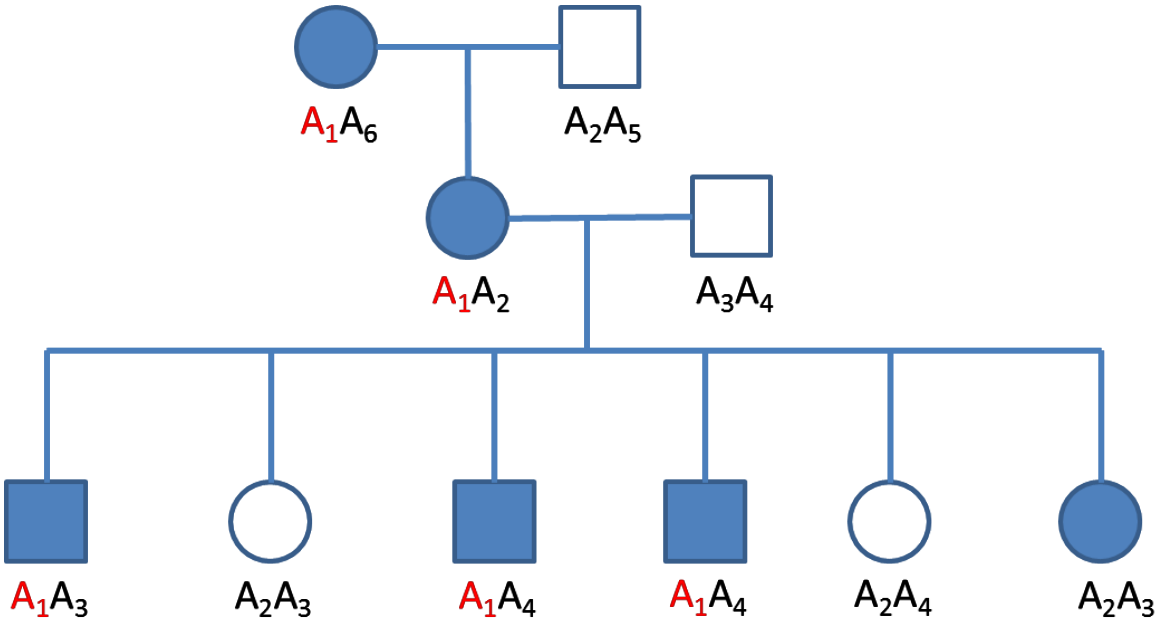
Recombinants vs non-recombinants



KU LEUVEN

Recombinants vs non-recombinants

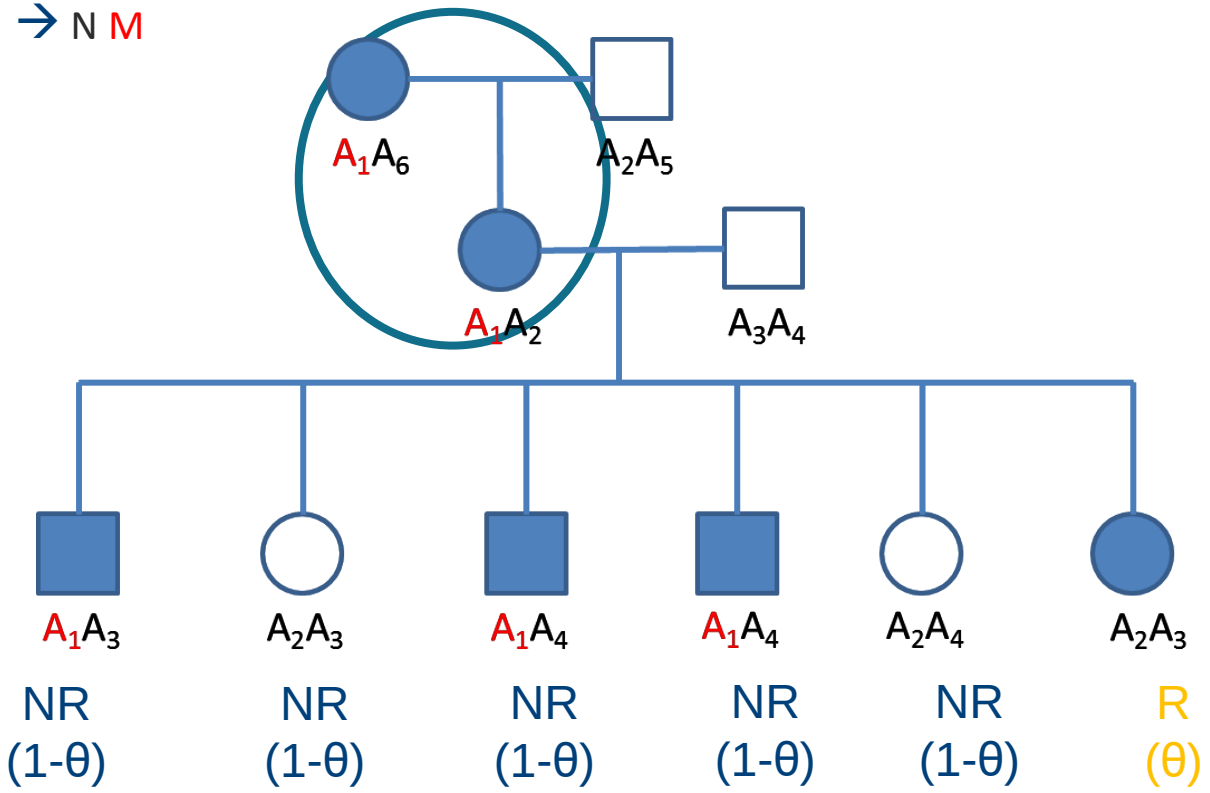
Phase known?



Based on Strachan & Read (2004)

Recombinants vs non-recombinants

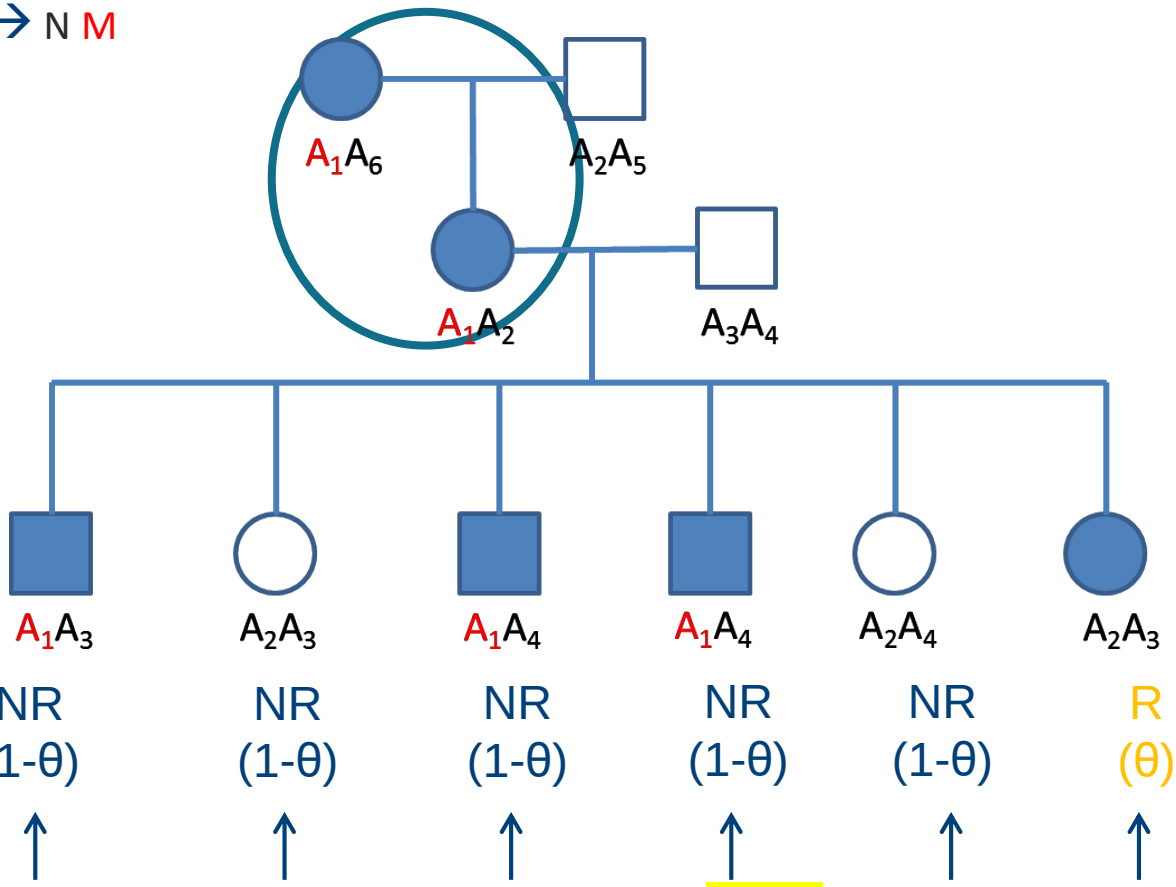
Phase known → N M



Based on Strachan & Read (2004)

Recombinants vs non-recombinants

Phase known → N M



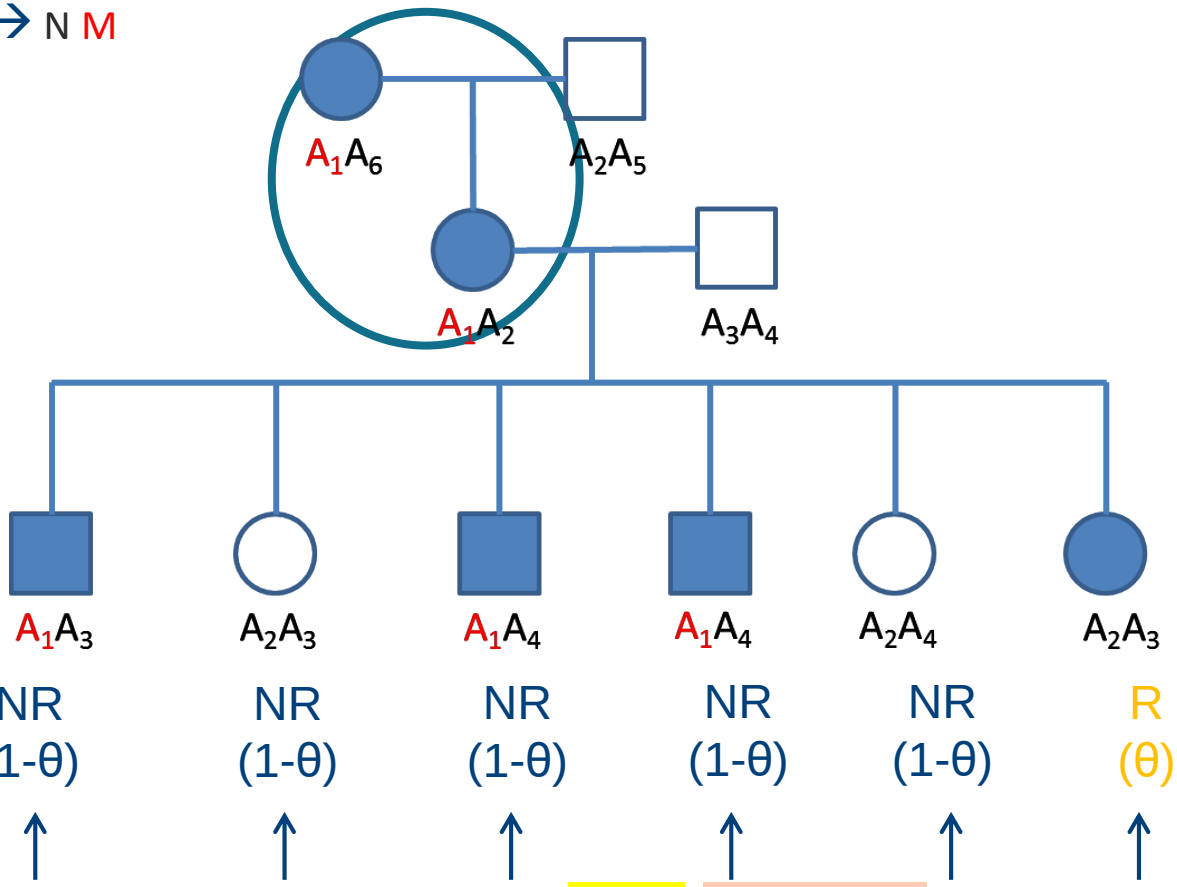
Chance for this observation when marker locus is linked with disease locus via θ?

$$L(\theta) = (1 - \theta)^5 * \theta^1$$

Based on Strachan & Read (2004)

Recombinants vs non-recombinants

Phase known → N M



Chance for this observation when marker locus is linked/NOT linked with disease locus via θ?

$L(\theta) = (1 - \theta)^5 * \theta^1$

$(1 - 0.5)^5 * 0.5 = 0.5^6$

KU LEUVEN

Based on Strachan & Read (2004)

Linkage analysis – LOD score

- $Odds = \frac{Likelihood\ of\ linkage\ (\theta)}{Likelihood\ that\ loci\ are\ unlinked\ \left(\theta=\frac{1}{2}\right)} = \frac{L\ (\theta)}{L\left(\theta=\frac{1}{2}\right)} = \frac{(1-\theta)^5 * \theta^1}{0.5^6}$

- $Logarithm\ of\ odds = LOD\ score\ (Z) = log_{10}\left[\frac{L\ (\theta)}{L\left(\theta=\frac{1}{2}\right)}\right]$

KU LEUVEN

Linkage analysis – LOD score

- $Odds = \frac{Likelihood\ of\ linkage\ (\theta)}{Likelihood\ that\ loci\ are\ unlinked\ \left(\theta=\frac{1}{2}\right)} = \frac{L\ (\theta)}{L\left(\theta=\frac{1}{2}\right)} = \frac{(1-\theta)^5 * \theta^1}{0.5^6}$

>1

What is most likely θ?

- $Logarithm\ of\ odds = LOD\ score\ (Z) = log_{10}\left[\frac{L\ (\theta)}{L\left(\theta=\frac{1}{2}\right)}\right]$

>0

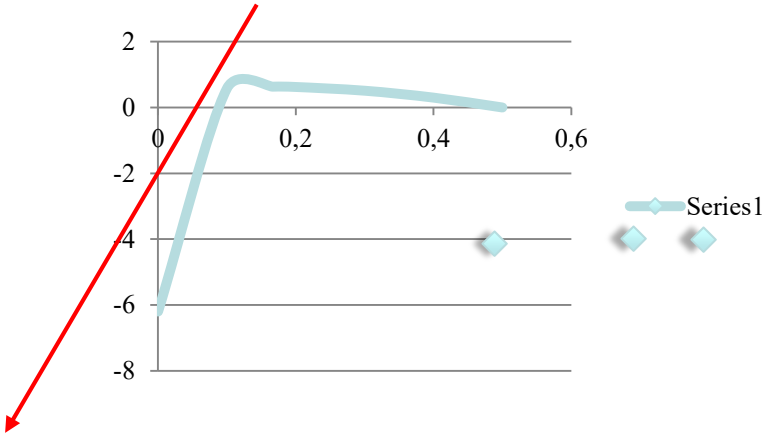
Most likely θ ?

$$\log_{10}\left[\frac{\Theta^{R} \times (1-\Theta)^{NR}}{0.5^{(R+NR)}}\right]$$

R = # rec. = 1

NR = # non-rec. = 5

teta	0.00000001	0.1	0.16666667	0.2	0.3	0.4	0.5
lod score	-6.19382	0.57739252	0.63212249	0.6226599	0.50879143	0.29899622	0



Is this significantly different from 0.5 ?

KU LEUVEN

Linkage analysis – LOD score

- $Odds = \frac{Likelihood\ of\ linkage\ (\theta)}{Likelihood\ that\ loci\ are\ unlinked\ \left(\theta=\frac{1}{2}\right)} = \frac{L\ (\theta)}{L\left(\theta=\frac{1}{2}\right)} = \frac{(1-\theta)^5 * \theta^1}{0.5^6}$

  >1

- $Logarithm\ of\ odds = LOD\ score\ (Z) = log_{10}\left[\frac{L\ (\theta)}{L\left(\theta=\frac{1}{2}\right)}\right]$

  >0 → ≥3 ?

KU LEUVEN

Z=3.0 threshold for accepting linkage with a 5% chance of a type1 error (falsely rejecting the null hypothesis)

Hypothesis	**Loci are linked** (recombination fraction = θ)	**Loci are not linked** (recombination fraction = 0.5)
Prior probability	1/50	49/50
Conditional likelihood: 1000:1 odds of linkage [lod score $Z(\theta) = 3.0$]	1000	1
Joint probability (prior × conditional)	20	~1

Box 14.3 Human Molecular Genetics, 4ed. (© Garland Science)

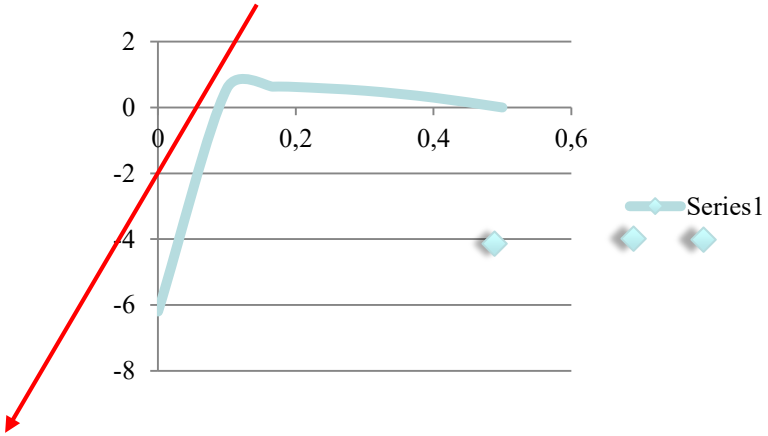
Most likely θ ?

$$\log_{10}\left[\frac{\Theta^{R} \times (1-\Theta)^{NR}}{0.5^{(R+NR)}}\right]$$

R = # rec. = 1

NR = # non-rec. = 5

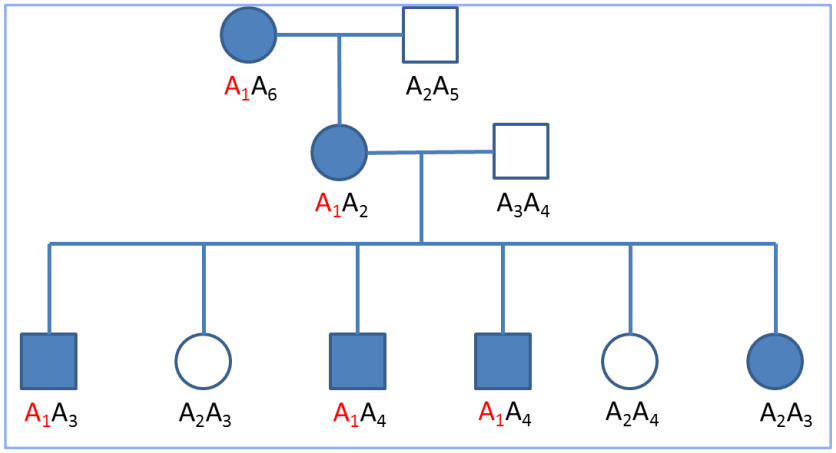
teta	0.00000001	0.1	0.16666667	0.2	0.3	0.4	0.5
lod score	-6.19382	0.57739252	0.63212249	0.6226599	0.50879143	0.29899622	0

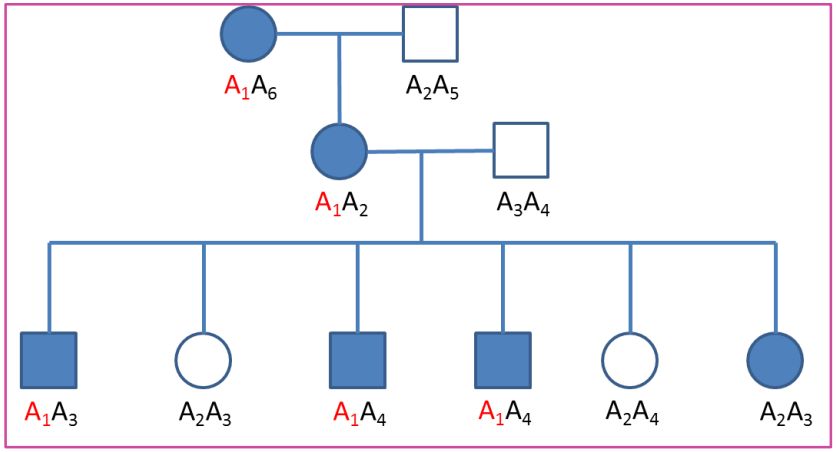


Is this significantly different from 0.5 ? → ≥3 ?

KU LEUVEN

More families…





$$\log_{10}\left[\frac{\Theta^{(R1+R2)} \times (1-\Theta)^{(NR1+NR2)}}{0.5^{(R1+NR1)+(R2+NR2)}}\right]$$

Family 1
R1 = # rec.
NR1 = # non-rec.

Family 2
R2 = # rec.
NR2 = # non-rec.

$$\log_{10}\left[\frac{\Theta^{R1} \times (1-\Theta)^{NR1}}{0.5^{(R1+NR1)}}\right]$$

$$+ \log_{10}\left[\frac{\Theta^{R2} \times (1-\Theta)^{NR2}}{0.5^{(R2+NR2)}}\right]$$

SUM ≥ 3 ?

KU LEUVEN

Potential LOD score curves

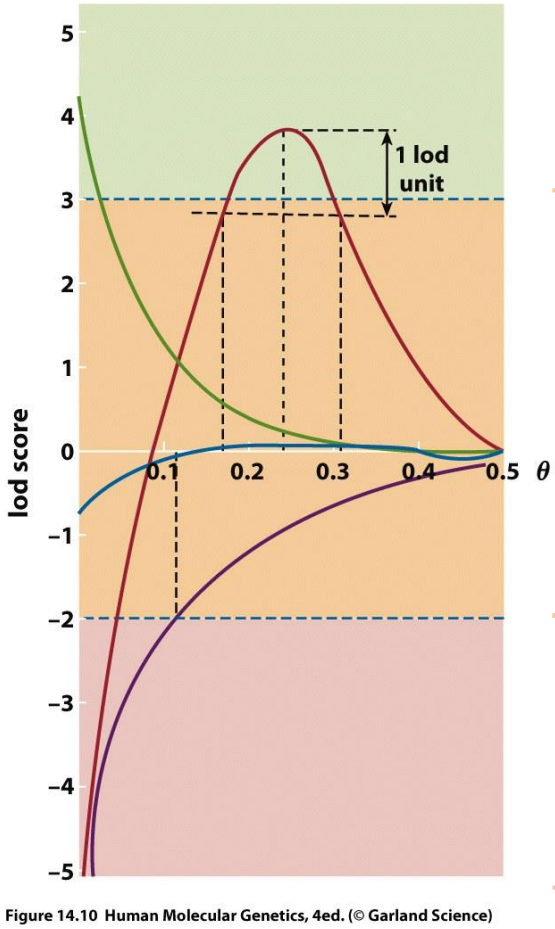
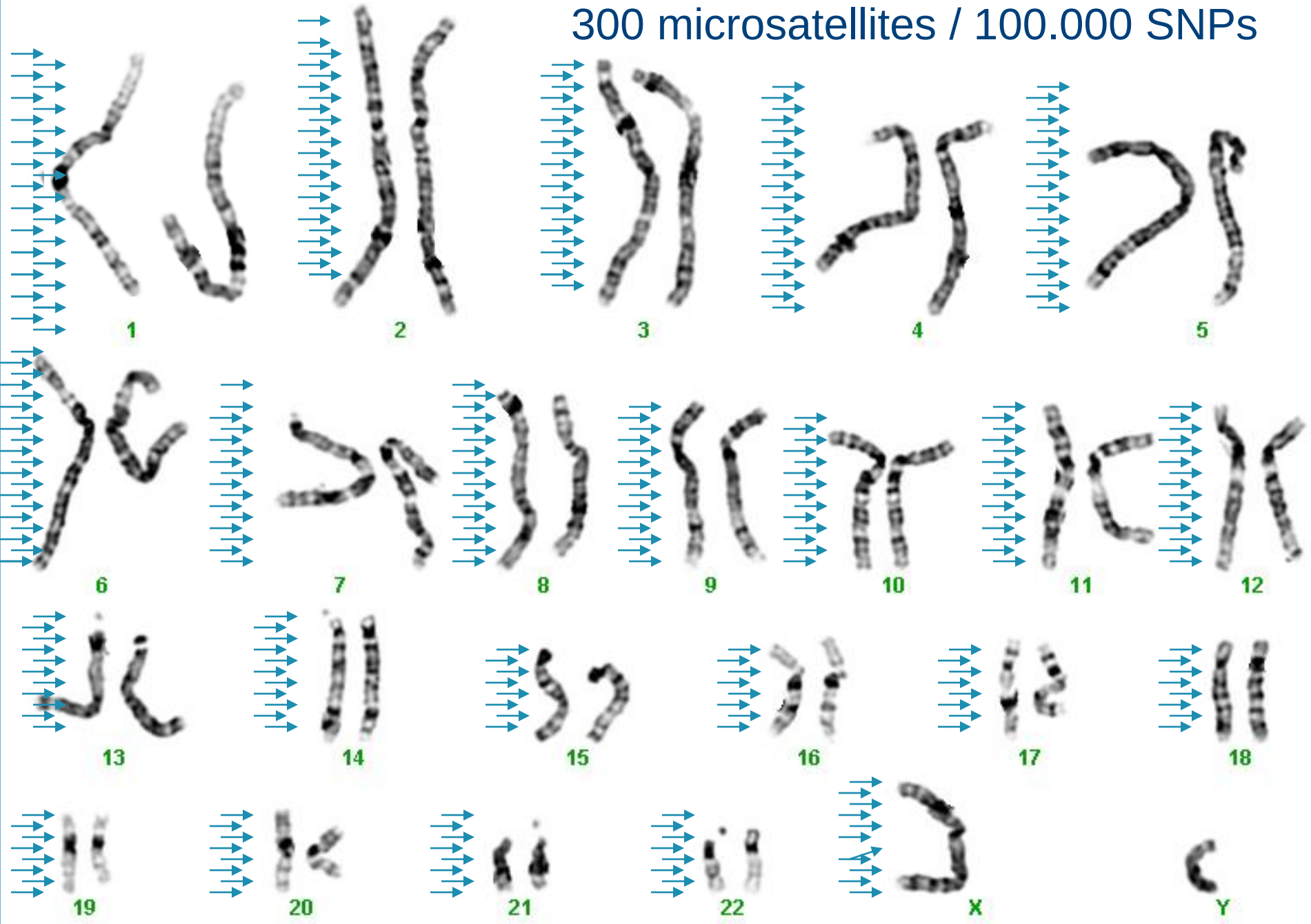


Figure 14.10 Human Molecular Genetics, 4ed. (© Garland Science)

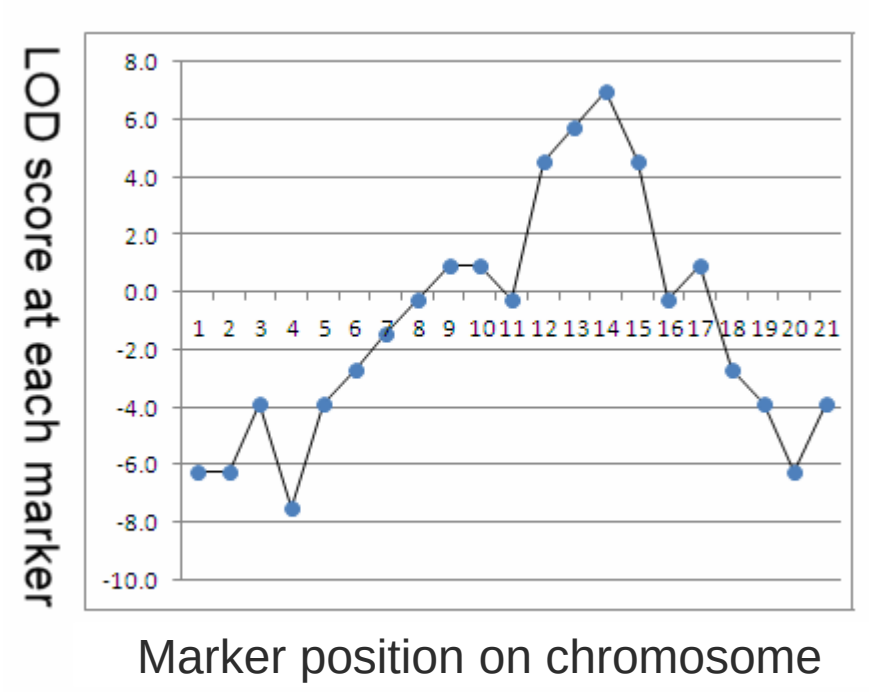
inconclusive

Exclusion mapping

300 microsatellites / 100.000 SNPs

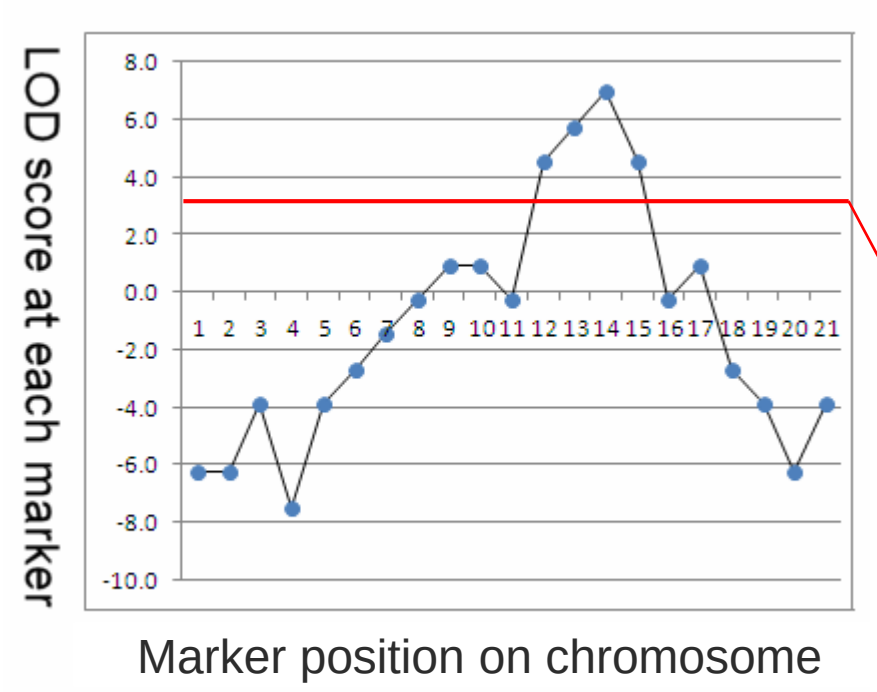


LOD score - visualization



- θ between genetic markers & disease locus
  - LOD score plotted against chromosomal positions
  - highest peak = most likely locus

LOD score - visualization



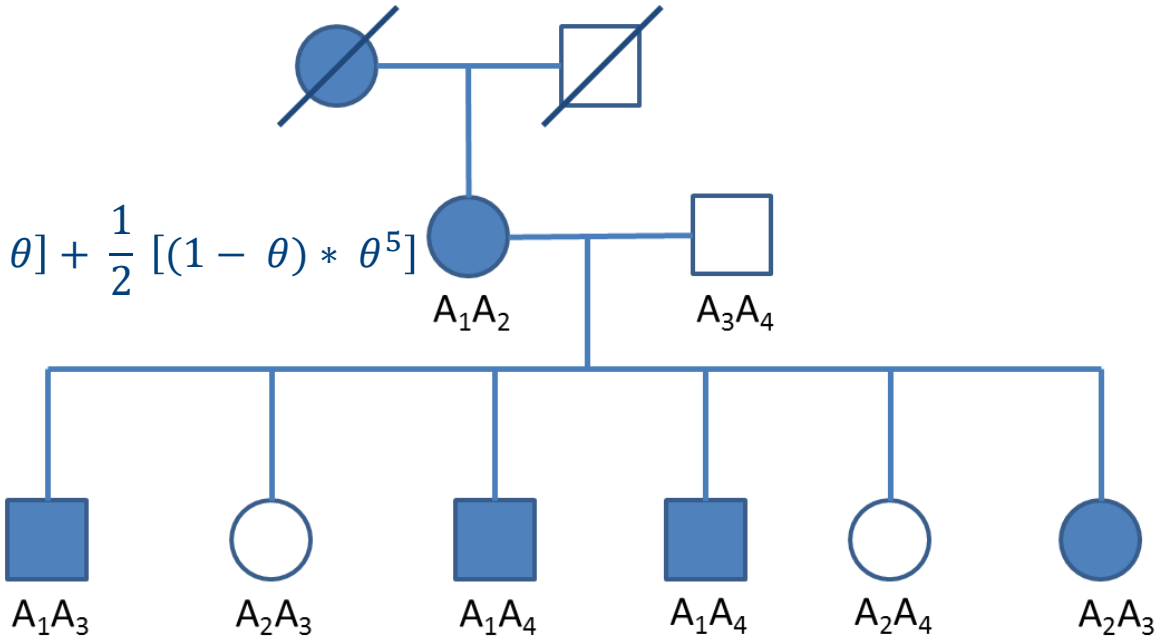
- θ between genetic markers & disease locus
  - LOD score plotted against chromosomal positions
  - highest peak = most likely locus
- LOD score = 3
  - ➢ 1 / 10^3 chance that the observed concordance is due to chance

KU LEUVEN

Recombinants vs non-recombinants

Phase unknown

$$L(\theta) = \frac{1}{2}\left[(1-\theta)^5 * \theta\right] + \frac{1}{2}\left[(1-\theta) * \theta^5\right]$$



Linkage analysis – how?

1. Collect familie(s)
2. Choose a genetic marker (eg microsatellite, SNP) for linkage analysis
3. Genotype this marker in all individuals of the familie(s)
4. Identify informative meioses (R vs NR → between heterozygous GM and heterozygous disease locus)
   - Based on the phenotype you know the disease locus genotype (affected vs unaffected)
   - Is phase known?
5. Determine whether there is linkage between the GM locus and the disease locus
   - Determine recombination fraction θ

$$log_{10}\left[\frac{Likelihood\ of\ linkage\ (\theta)}{Likelihood\ that\ loci\ are\ unlinked\ \left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{L\ (\theta)}{L\left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{(1-\theta)^{NR} * \theta^{R}}{0.5^{(R+NR)}}\right]$$

= max

   - Is this recombination significantly different from 0.5?

     (= null hypothesis; no linkage between GM and disease locus)

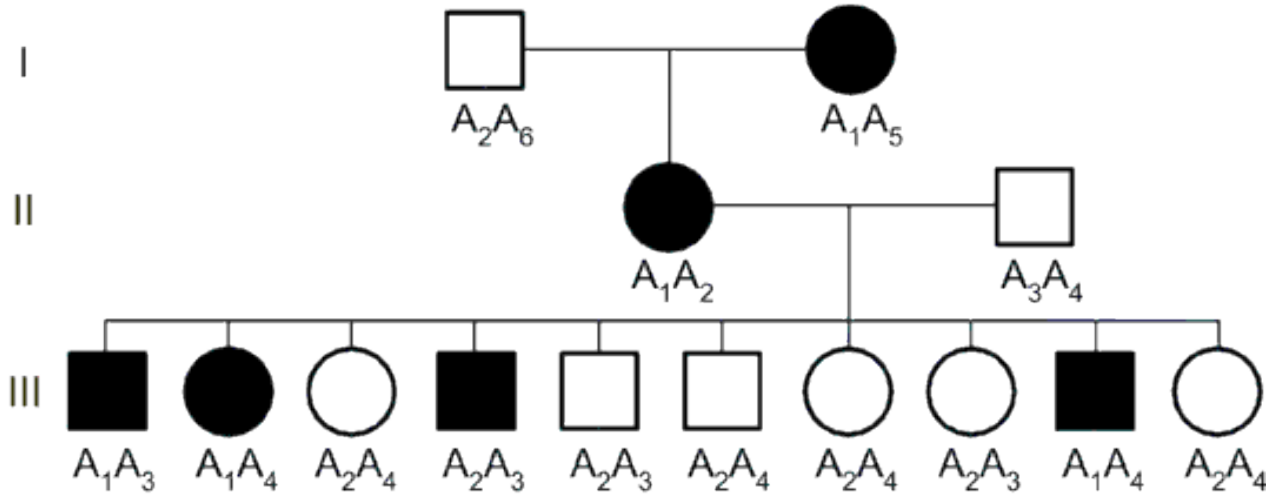
$$log_{10}\left[\frac{Likelihood\ of\ linkage\ (\theta)}{Likelihood\ that\ loci\ are\ unlinked\ \left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{L\ (\theta)}{L\left(\theta=\frac{1}{2}\right)}\right] = log_{10}\left[\frac{(1-\theta)^{NR} * \theta^{R}}{0.5^{(R+NR)}}\right]$$

≥ 3 ?

6. Do we need more informative meioses / families?

KU LEUVEN

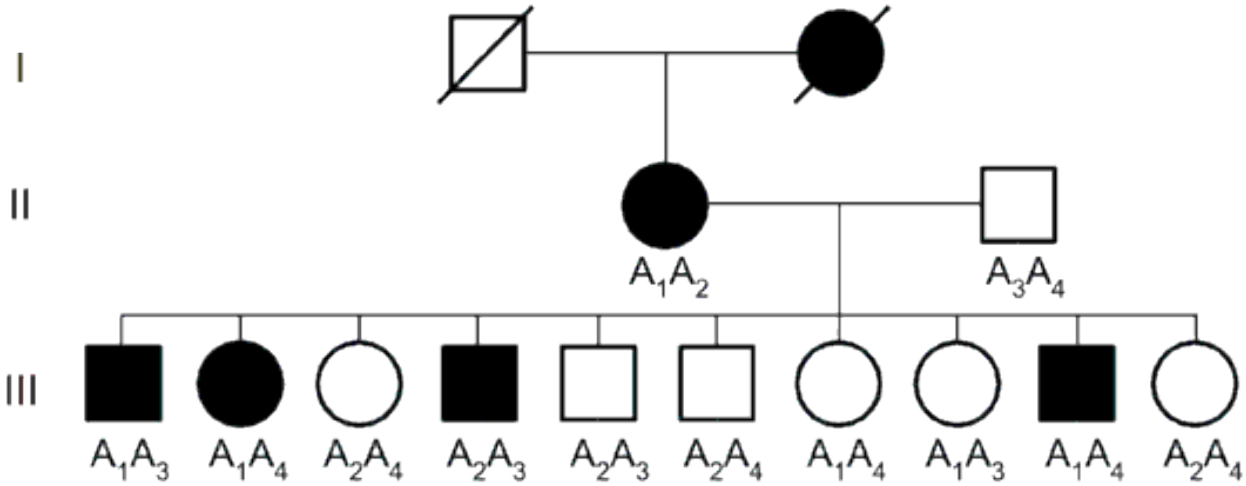
Linkage analysis – homework → assignment 2



θ = ?

LOD = ?

Linkage analysis – homework → assignment 2



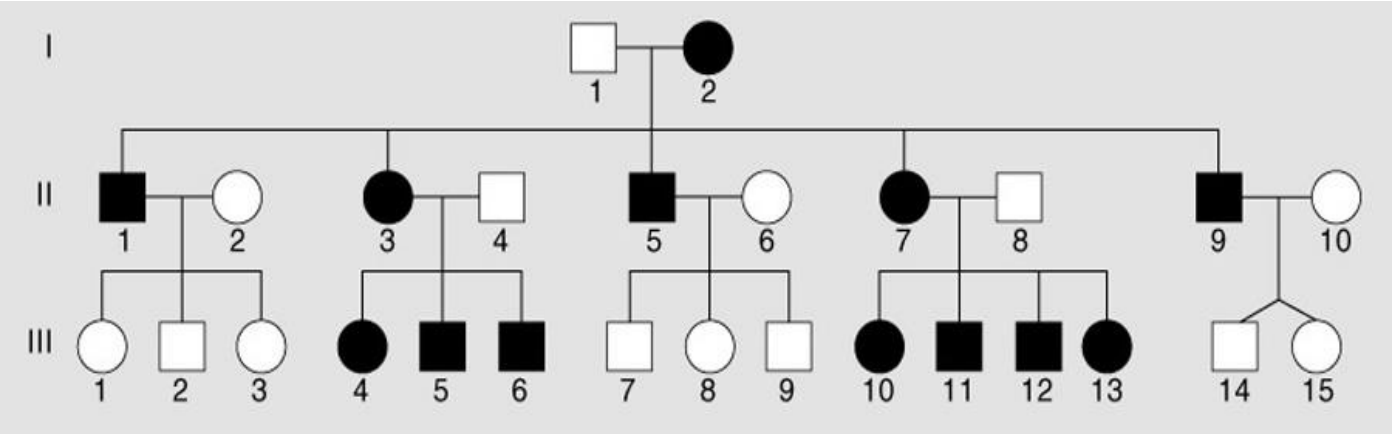
θ = ?

LOD = ?

KU LEUVEN

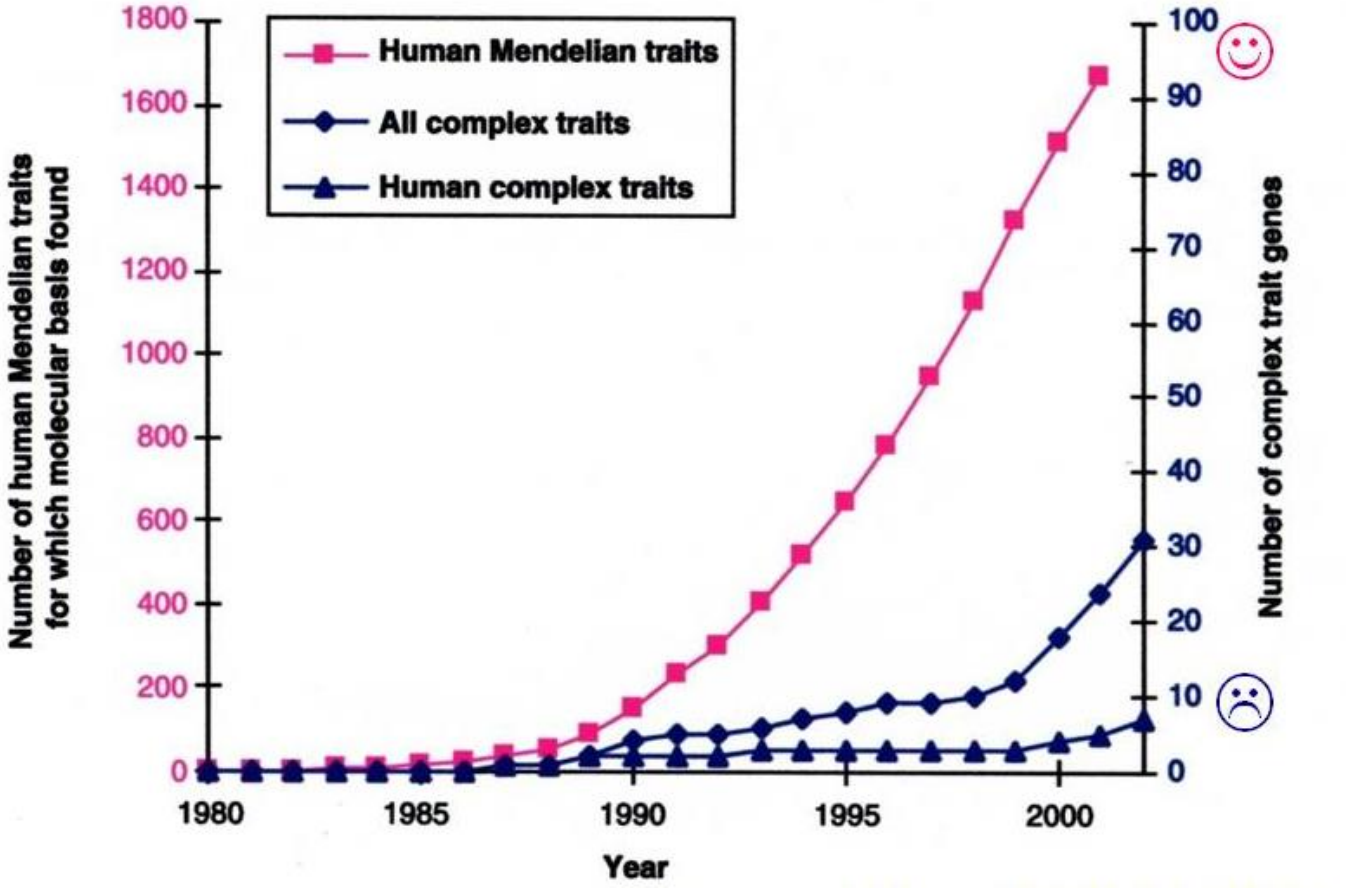
Linkage analysis

- co-segregation of genetic marker(s) and disease in a family



- Goal: identify chromosomal region with risk gene
- Limits :
  - good for detection of highly penetrant variant (ie extremely likely to cause the disease when present)
  - Hard to get big enough pedigrees with enough informative individuals

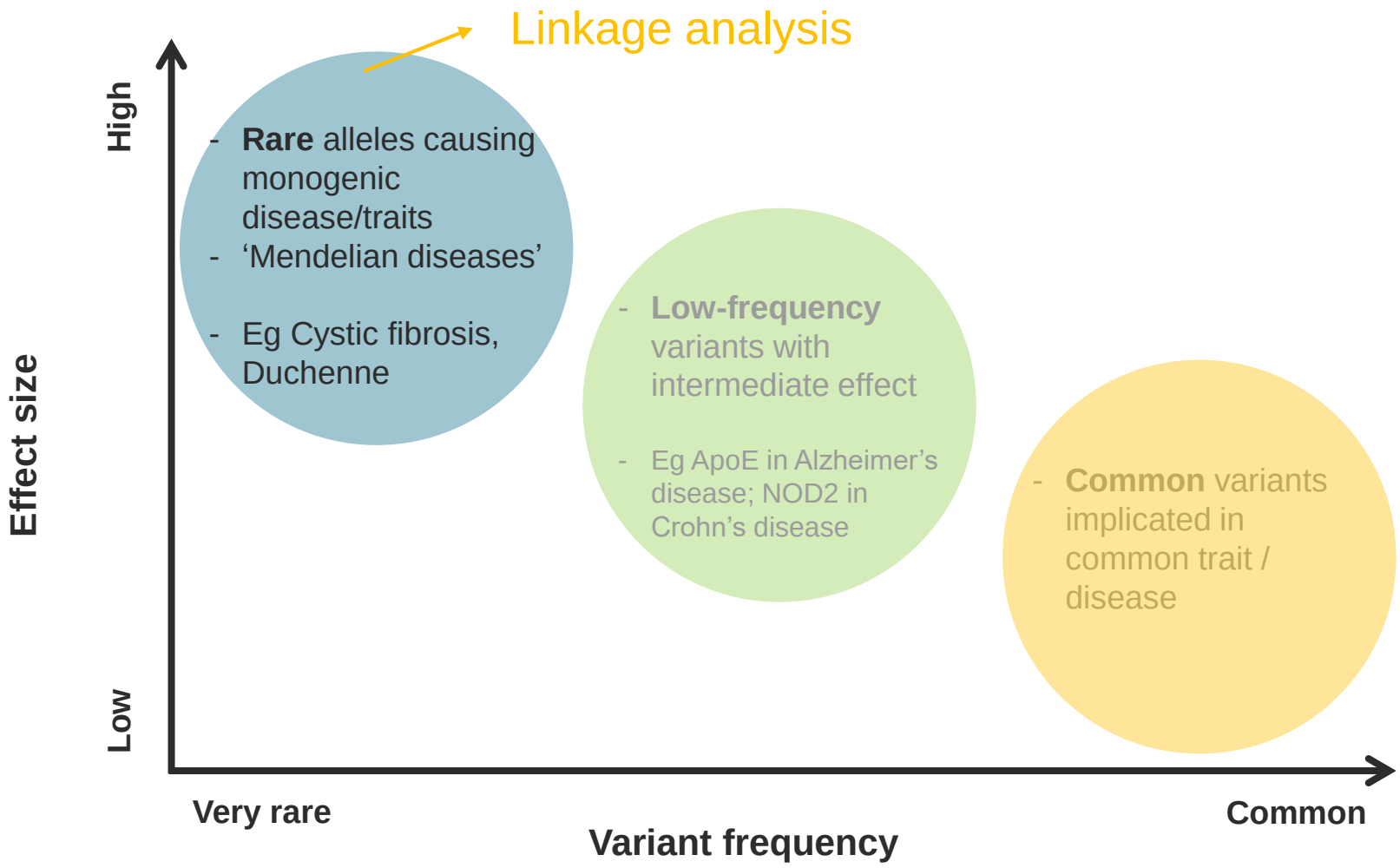
Genetic successes in Mendelian >< complex disorders



Glazier et al, Science 298: 2345-9,2002

KU LEUVEN

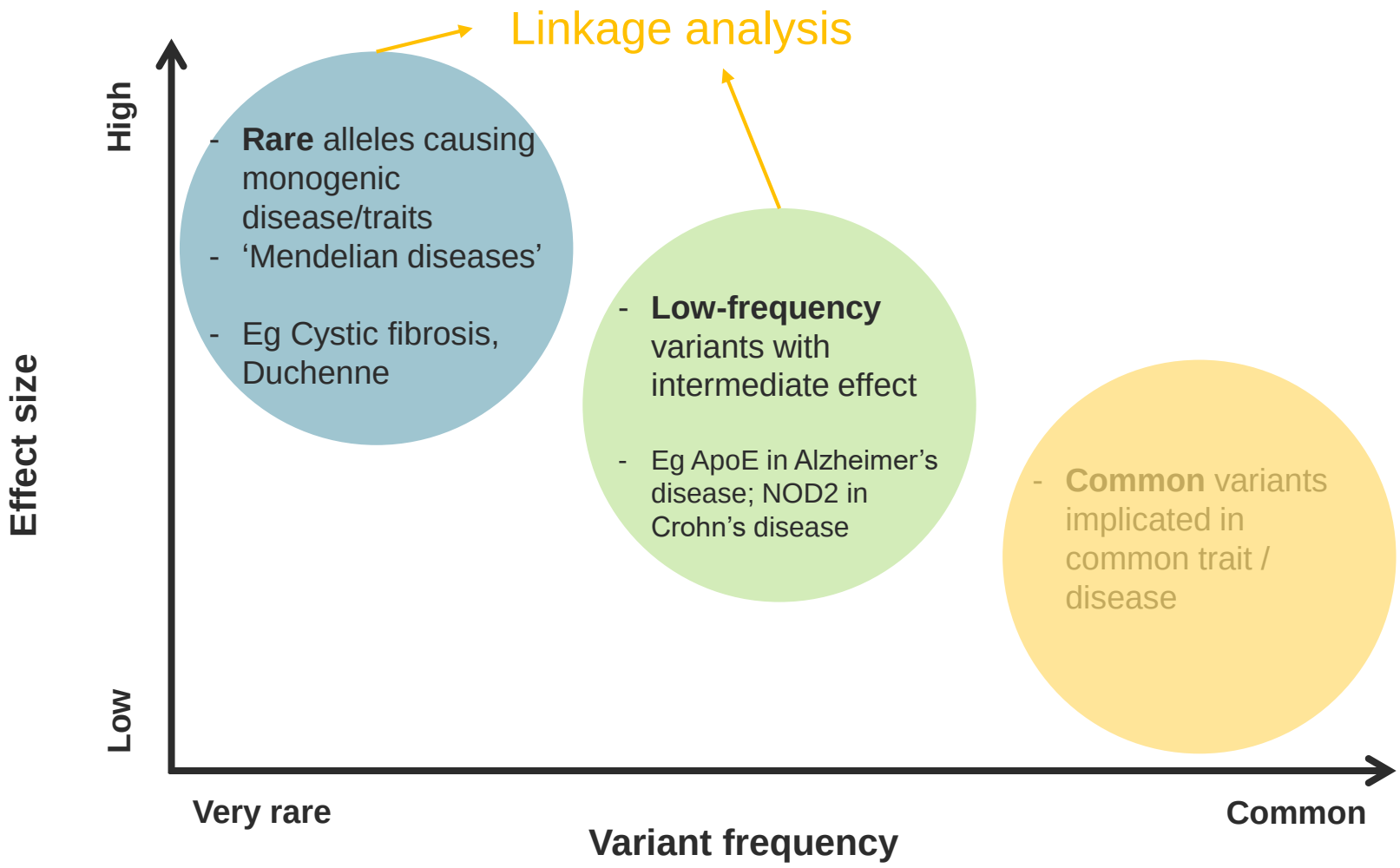
The genetic spectrum



Adapted from Manolio TA et al, Nature Reviews, 2009

KU LEUVEN

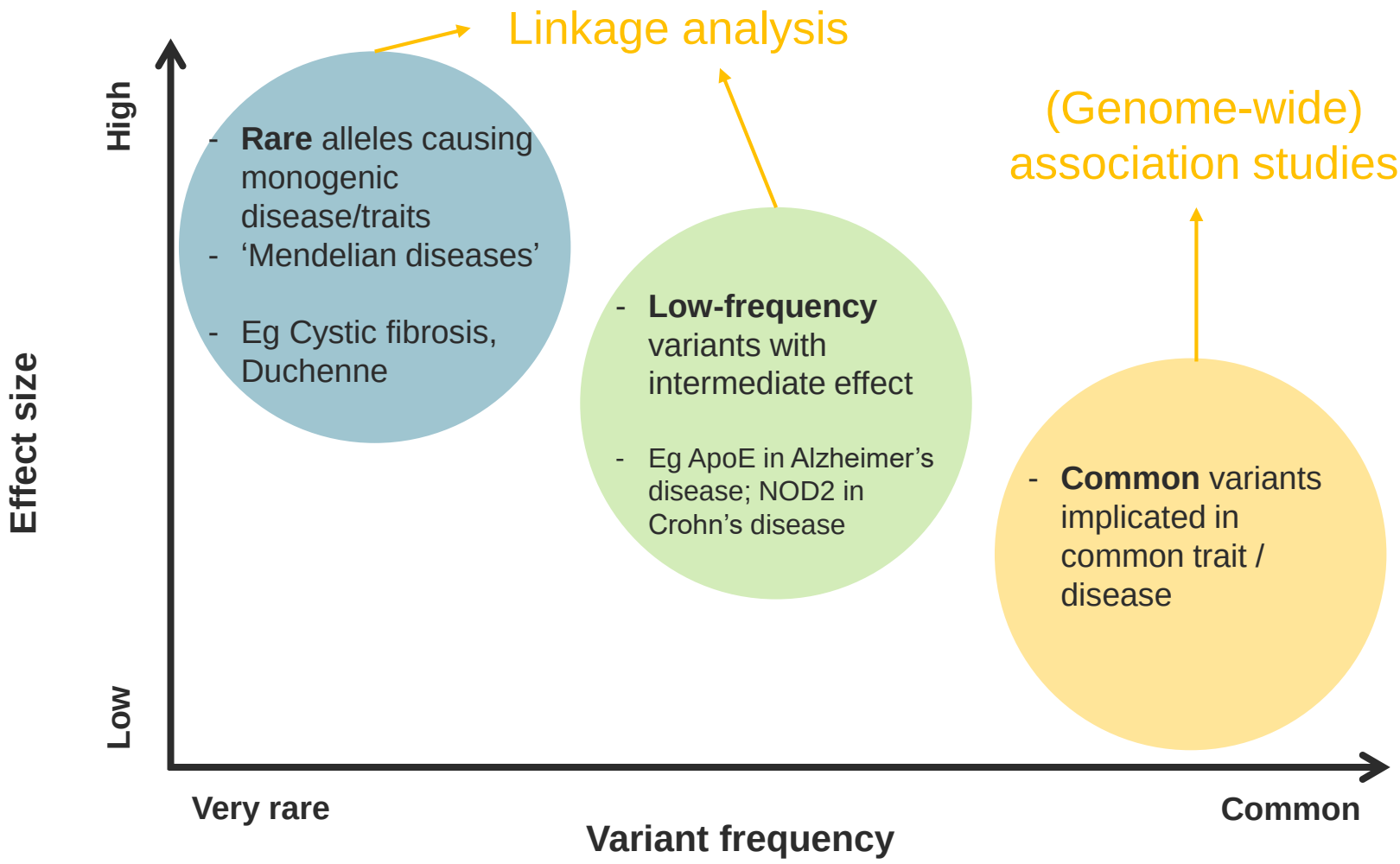
The genetic spectrum



Adapted from Manolio TA et al, Nature Reviews, 2009

KU LEUVEN

The genetic spectrum



Adapted from Manolio TA et al, Nature Reviews, 2009

KU LEUVEN