GENETICS OF COMPLEX DISEASES : METHODOLOGICAL TOOLS

STATISTICAL METHODS TO IDENTIFY GENETIC FACTORS IN HUMAN DISEASES

M1 Geniomhe

2024-2025

Valérie Chaudru
valerie.chaudru@univ-evry.fr

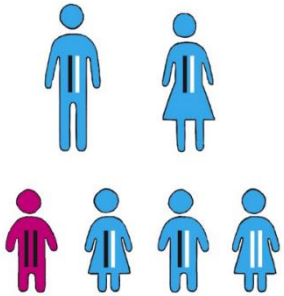
Elisabeth Petit-Teixeira

AIM OF THE COURSE

Human genome : ~3 billion base pairs (A, C, T, G)

=> Overall similarity across human genomes : 99.6%

=> Genetic diversity ≈0.4%



Epidemio-genetics methods aim to identify new genetic factors based on genetic diversity in complex human diseases

Scoring

50% CC (online Multiple Choice Questions) – end of the course

50% Exam – exam's week

RISK FACTORS IN HUMAN DISEASES

Genetic factors

- **Mendelian entity**
  - *mutation (very rare)*
  - *high disease-risk*

While rare in a population, carriers face high….

- **Susceptibility gene**
  - *Frequent variant*
  - *low disease-risk*

Due to their increased presence, they are more "controlled in the body", therefore less deadly

interactions

Environnemental factors

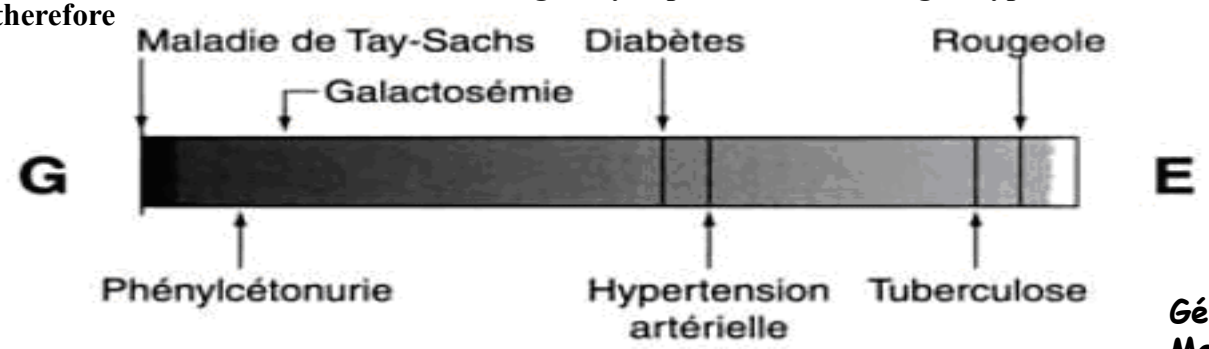
sun exposure, tobacco, diet …

Host factors

age, sex, nevi, skin color …

Race? Genetically induced moles??

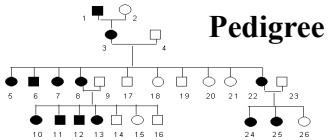
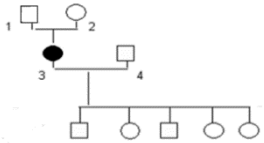
Consider the effect of environmental in multifactoral diseases as adding petrol to a fire; multifactoral diseases are on a spectrum, where the effect environmental factors have on the individual, greatly depend on the allelic genotype of the individual



Génétique médicale, Ed. Masson, 2004

➢ **Monogenic diseases / Multifactorial diseases**

MONOGENIC DISEASES vs MULTIFACTORIAL DISEASES

	Monogenic	Complex (multifactorial)
Familial aggregation: the occurence of diesease in multiple members of a family, across generations (inheritance)	 Pedigree	
Disease frequency	Mendelian entities -> One genetic factor =/= Only one mutation in the patient. Rare	Medium / High If high, environmental factors assume the function of mendelian entities (somewhat)
Familial aggregation	High *(hereditary diseases)*	Low *(except in 5-10% of cases)*
Nb of genes	1 Mendelian entity *(≠ mutations of the gene)*	Multiple *(interactions between genes)*
Penetrance	penetrance 0.8 < 1 High *(but incomplete)*	Low *(except if Mendelian entity)*
Environment effect	Low *(treatment aid)*	Medium / High *(interactions with genes)*

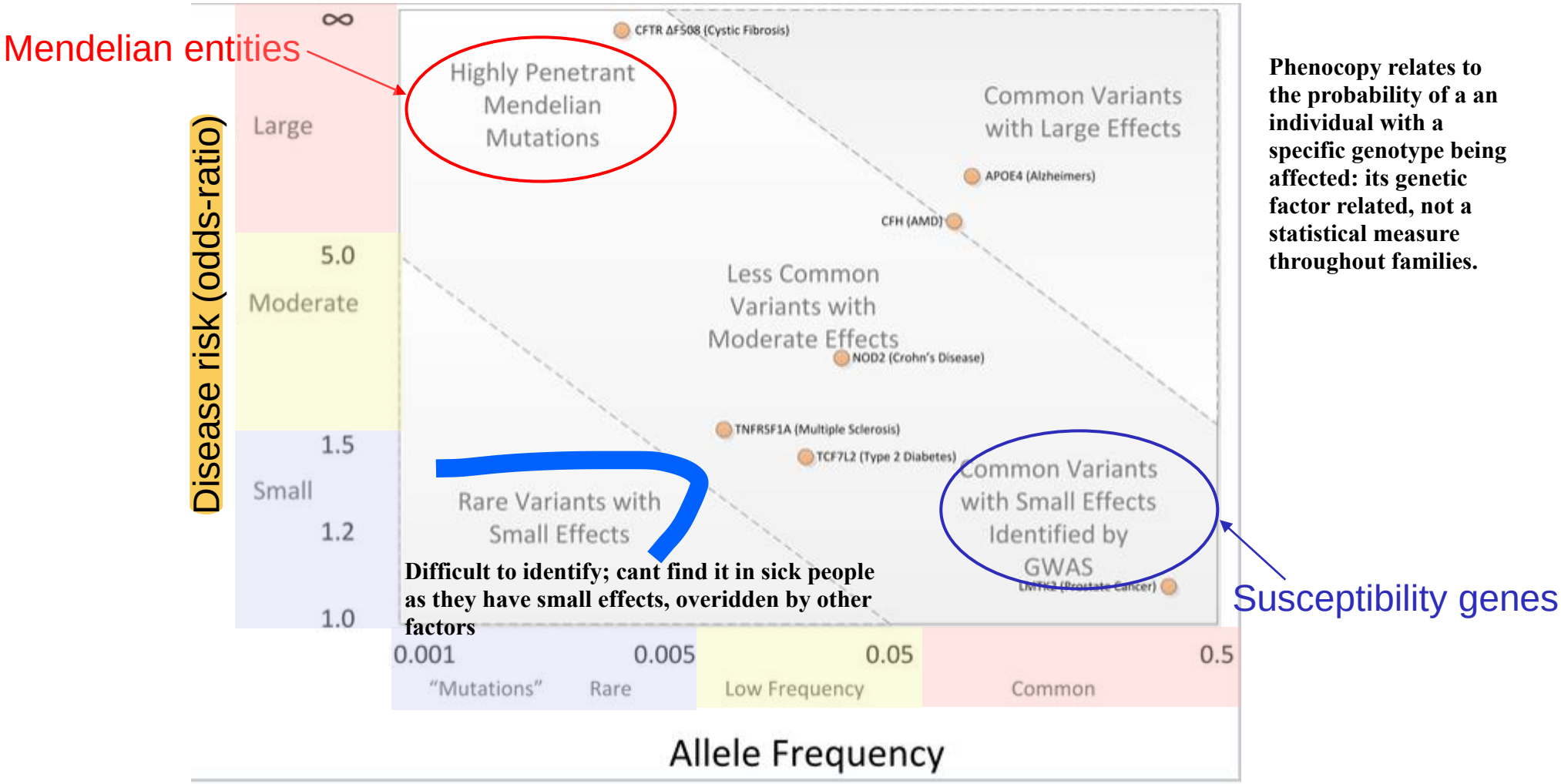
Definitions for a genetic risk factor :

If complete penetrance <=> P = 1

Compare observed cases of familial disease to decide penetrance of disease; you cannot look at random samples, only families

- *Penetrance = P([affected]/at-risk genotype)*
- *Phenocopy = P([affected]/non at-risk genotype)*

FOCUS ON THE SPECTRUM OF GENETIC FACTORS IN COMPLEX DISEASES



Phenocopy relates to the probability of a an individual with a specific genotype being affected: its genetic factor related, not a statistical measure throughout families.

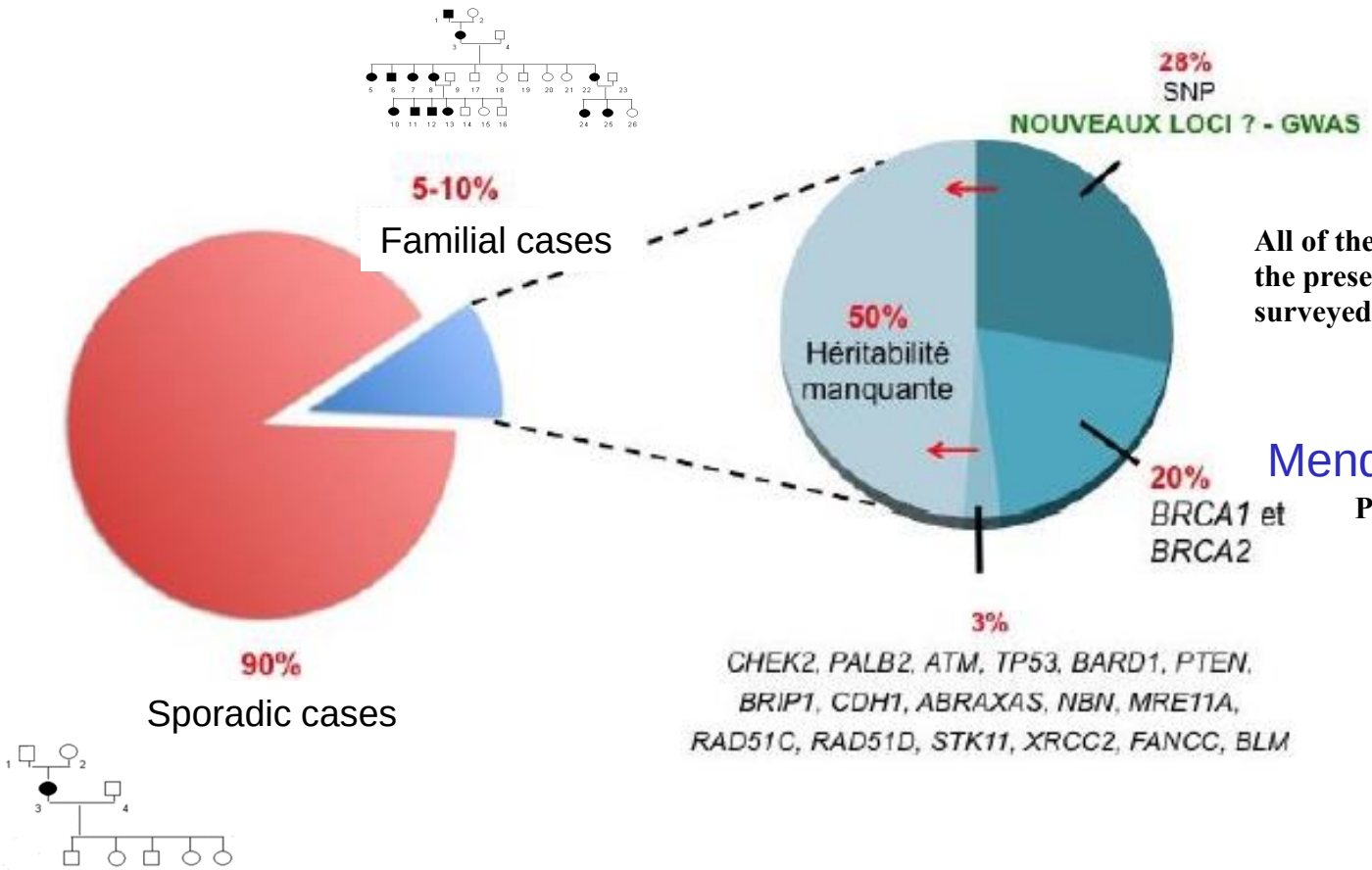
Bush et al., 2012.

EXAMPLE OF MONOGENIC DISEASE: PHENYLKETONURIA

- **DISEASE**
  - Inability to properly convert phenylalanine (one of the amino acids found in food) into tyrosine
    - *accumulation of toxic compounds for the brain and significant mental retardation from childhood*
  - Incidence : ≈ 1/16000

- **GENETIC FACTOR** Mendelian entity (Complete/Absolute) penetrance
  - Gene PAH (phenylalanine hydroxylase) - chrom.12 – recessive trans.
  - ≈ 500 mutations identified Confounding factors but not the "driver" mutations of the disease; are an after effect.
  - Mut. cause enzyme deficiency
  - High penetrance
    - *if dietary treatment at birth --> penetrance ≈ 0*

EXAMPLE OF A COMPLEX DISEASE: BREAST CANCER



Given alleles
AA
AB
BB

All of these calculations are performed to identify the presence of specific allelic genotypes within surveyed patients

Mendelian entities

Penetrance above 0.8

Missing heritability = part of the genetic component not yet identified

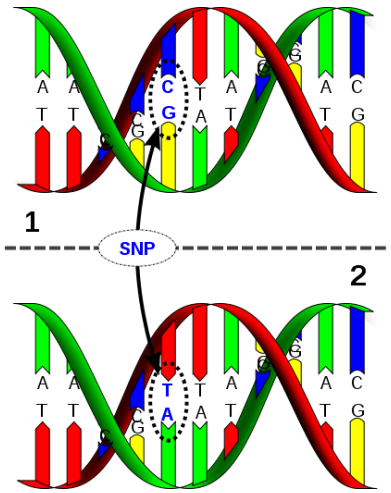
(While some women observe high penetration in their family, we do not know what the mendelian entity is to drive this inheritance in those extreme cases.)

=> Use of genetic variability between individuals and statistical methods to minimize missing heritability

GENETIC MARKERS FOR STUDYING GENETIC DIVERSITY

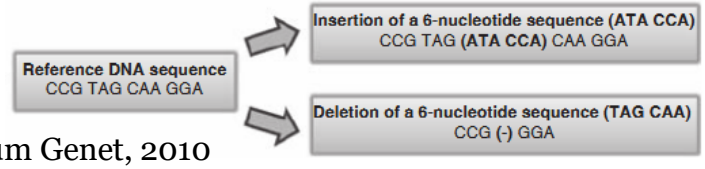
Most commonly used genetic marker for disease analysis is:

- **SNV = single nucleotide variant**:
  DNA sequence variation in which a single nucleotide — A, T, C or G — differs between members of the same species



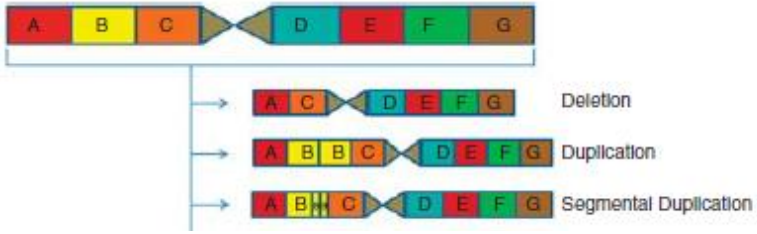
- **SNP = single nucleotide polymorphism**:
  SNV occurring commonly within a population (> 1%)

- **Small insertions/deletions (indels)**



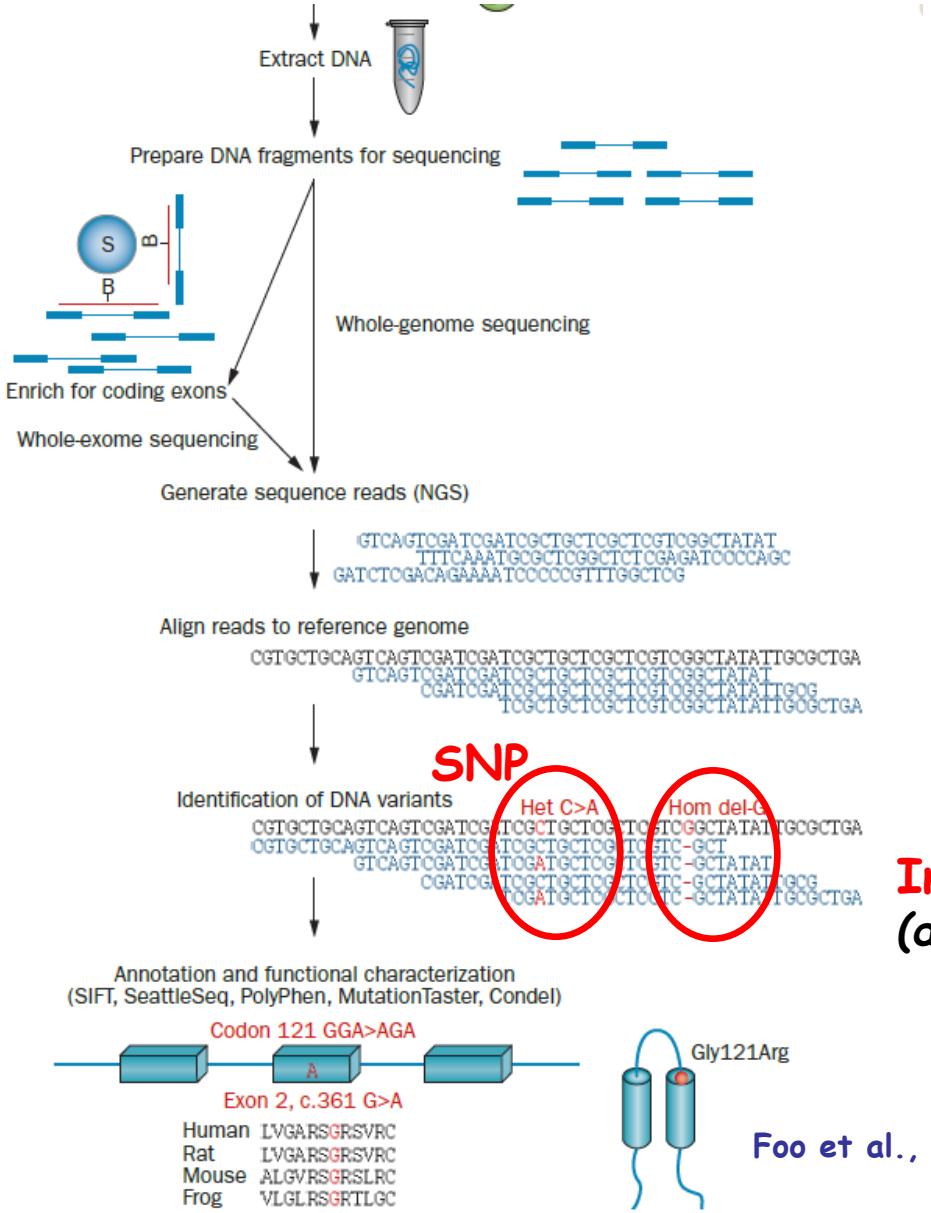
Ku et al., J Hum Genet, 2010

- **Large copy number variants (CNVs)**
  >1 kb



D'après Suhani H, et al., 2012

SEQUENCING AS TOOL TO IDENTIFY GENETIC MARKERS



1. Cutting of DNA in small fragments
2. Sequencing of the fragments = reads
3. Alignment
4. Variant calling

Indel :
(deletion of G)

5. Filtering

Foo et al., Nat Rev Neurol, 2012

HOW IS USED THE GENETIC MARKERS INFORMATION?

IT tools :

- Data storage
- Programs for :
  - * data management
  - * data analysis

Statistical analyses

+

Fonctional genomic

Genetic markers

Observe the effect of identified genetic markers on the function of vitality genes

Genes identification

Take for example counseling for newborns considering familial history of disease (monogenic diseases)

Therapeutics (genetic therapy)

Interactions G x E (pharmacogenetic)

Early diagnosis and prevention

+

Genetic counseling

Personalised medicine

WHAT DID I REMEMBER ?
(1)

1. Allez sur wooclap.com
2. Entrez le code d'événement dans le bandeau supérieur

Code d'événement
XELTEB

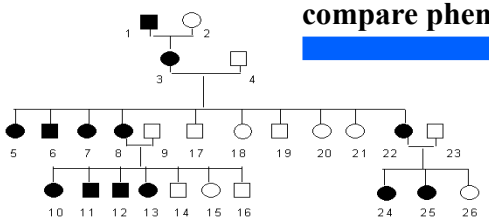
1. Envoyez @XELTEB au 06 44 60 96 62
2. Vous pouvez participer

Désactiver les réponses par SMS

HISTORY OF METHODS TO IDENTIFY GENETIC FACTORS (non exhaustive)

to calculate concordance: # of matching genetic factors between twins, over total genetic factors possible

Is there familial clustering ?
=> Recurrence risk



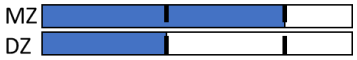
compare phenotypic concordance against control

Comparison of disease risk in relatives of cases vs risk in population

monozygotic twins by definition should have the same genetic components

Is there genetic component ?
=> Twin studies + others…

MZ: Monozygotic…..



E for environmental factors

Pheno. Concordance Rate (blue)

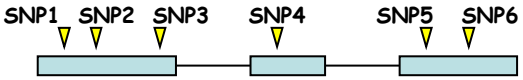
- If $PCR_{MZ}<1$ => Influence of E
- If $PCR_{DZ}< PCR_{MZ}$ => Influence of G

Localisation of the G factor ?
=> Linkage analysis

Disease gene close to which genetic marker?

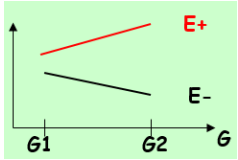
Not all alleles are susceptibility; increase disease risk!

Causal variant ? **See which genetic factors compound disease risk**
=> Association Analysis



f(allele1 SNP3) > in cases than in controls?

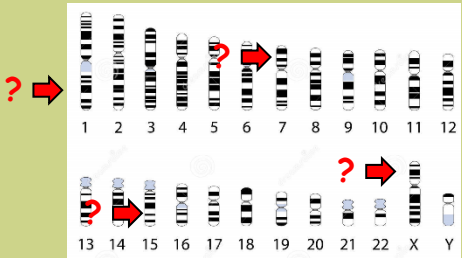
Search for interactions ?
=> GxG and/or GxE



Is genotypic risk on disease depends on E?

SOME EXAMPLES OF GENETIC ANALYSES

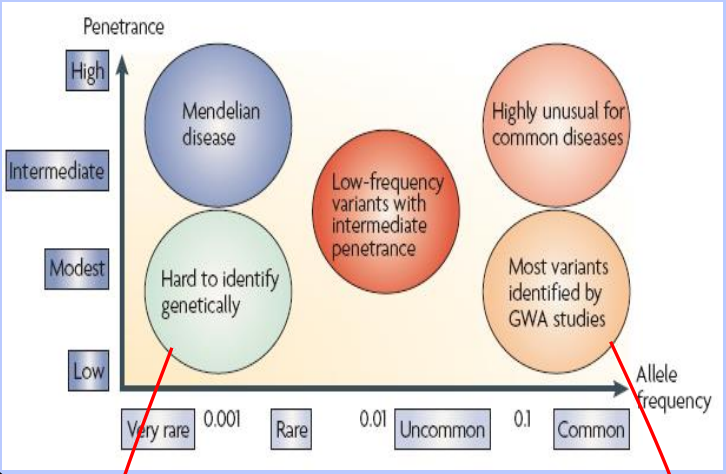
Linkage analyses



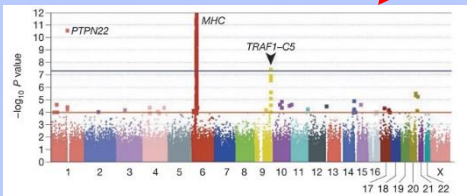
positional cloning or functional studies of candidate genes

Location of disease genes on the genome

Association analyses



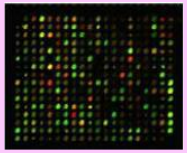
McCarthy et al. 2008

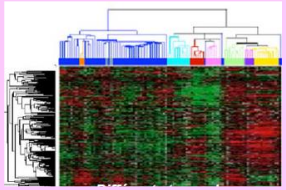


Search for causal variant

Identification of new genetic factor in a human disease

Transcriptomic analyses



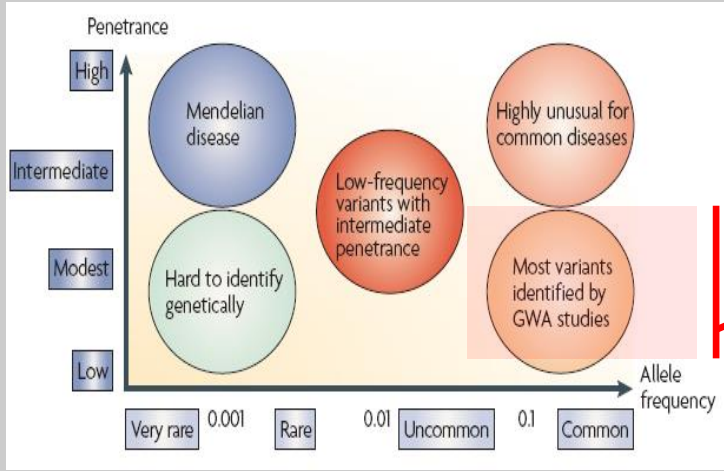


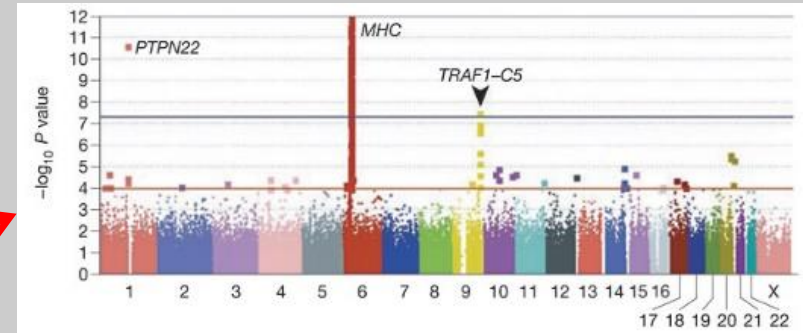
Identification of groups of genes (rows) over or under expressed in sample subgroups (columns)

PART 1

GENETIC ASSOCIATION ANALYSES

McCarthy et al. 2008





PREFERENTIALLY USED MARKERS : SNP

SNP = Single Nucleotide Polymorphism

=> very abundant and evenly distributed in the human genome ~ 1 SNP every 100-300 bp

=> Most SNPs are located in non-coding regions and have no direct impact on an individual's phenotype

=> Some SNPs may be in coding regions and have functional consequences

Example:

Sequence 1: ...CTGACTCCTG**A**GGAGAAG...
Sequence 2: ...CTGACTCCTG**T**GGAGAAG...

MAIN STRATEGIES

Positional cloning

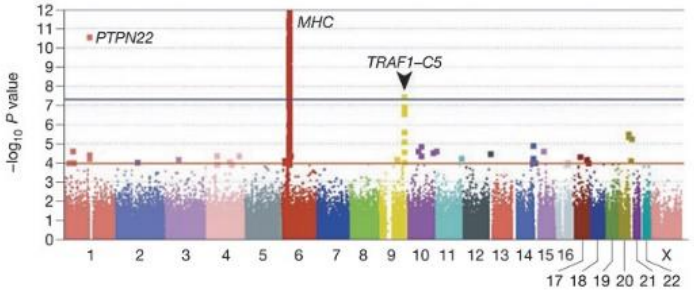
Linkage analyses

Refine the region by genotyping more markers (SNPs) in linkage region (or in candidate genes of the linkage region)

Association studies to search for the causal variant

Genome-Wide Association studies (GWAs)

Genotyping of 500,000 – 1 million SNPs covering the genome in thousands of cases / controls



Manhattan plot

- Discovery of new susceptibility genes without conducting linkage analyses

Samples	Methodology
Unrelated subjects	χ^2 or Regressive models
Nuclear families (parents + offsprings)	Transmission Disequilibrium Test (TDT)

ASSOCIATION ANALYSES

SEARCH FOR CAUSAL VARIANTS

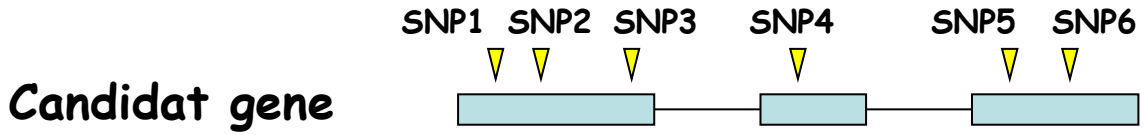
-

POPULATION-BASED

ASSOCIATION ANALYSES

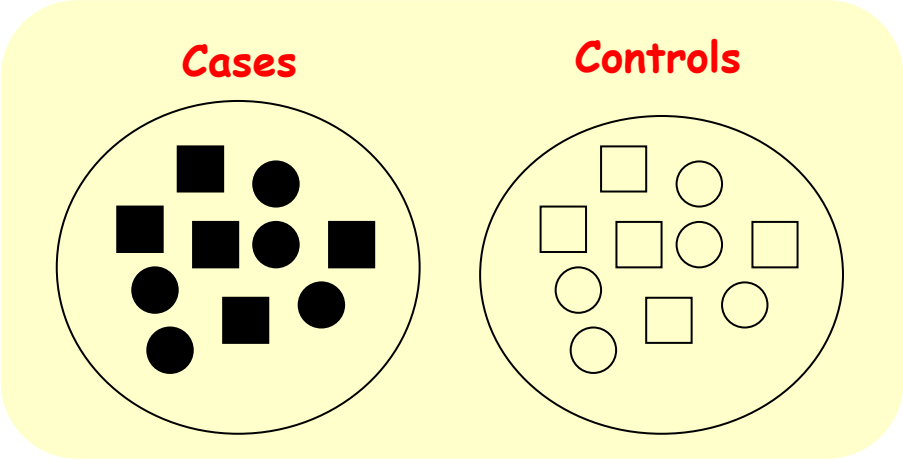
DESIGN OF CASES-CONTROLS ASSOCIATION STUDY

- **Aim : search for causal genetic variant**



- **Studied sample**

Cases (affecteds) and controls (unaffecteds) – all unrelated



For a given marker (SNP):

a) Are allele frequencies ≠ between cases and controls?

b) Are genotype frequencies ≠ between cases and controls?

For a set of marker (SNPs):

c) Are haplotype frequencies ≠ between cases and controls?

➢ *Quality controls on genotypes must be applied before analyses*

FEW WORDS ON QUALITY CONTROLS (1)

The main (non-exhaustive) steps for data quality control are :

- Deletion of SNP with % of missing genotypes (00 in the table) > threshold

  Ex: call rate < 95% => SNP with > 5% of missing genotypes => removed

- Deletion of individuals with % of missing genotypes (00 in the table) > threshold

  Ex: call rate < 95% => individuals with > 5% of missing genotypes => removed

	SNP1	SNP2	SNP3	SNP4	SNP5	…	SNPn'
Ind 1	11	11	12	00	22		11
Ind 2	00	00	22	00	00		00
Ind 3	11	22	12	12	12		22
Ind 4	22	12	22	00	11		11
…							
Ind n	12	12	22	00	12		11

DEL

DEL

1=ref
2=alt

FEW WORDS ON QUALITY CONTROLS (2)

The main (non-exhaustive) steps for data quality control are :

- **Deletion of genetic markers (SNP) « rare »**

  Ex: SNPs with MAF (Minor Allele Frequency) <5% removed

- **Deletion of genetic markers (SNP) which do not verify the Hardy-Weinberg (H-W) model in controls**

  Conformity test to the H-W model (chi-square test)

  H0: H-W verified ($f(11)=f(1)^2$; $f(12)=2*f(1)*f(2)$; $f(22)=f(2)^2$)
  H1 : H-W not verified

Genotypes :	11	12	22	Total
Observed counts	n1	n2	n3	N
Expected counts under H0	$N*f(1)^2$	$N*2*f(1)*f(2)$	$N*f(2)^2$	N

From obs. counts :

f(1) = (2*n1 + n2)/(2*N)

f(2) = (2*n3 + n2)/(2*N)

$$\chi^2(\text{nb categories} - 1 - (\text{nb all. freq} - 1)) = \Sigma_{cats}\,(\text{obs count} - \text{exp count})^2 / \text{exp count}$$

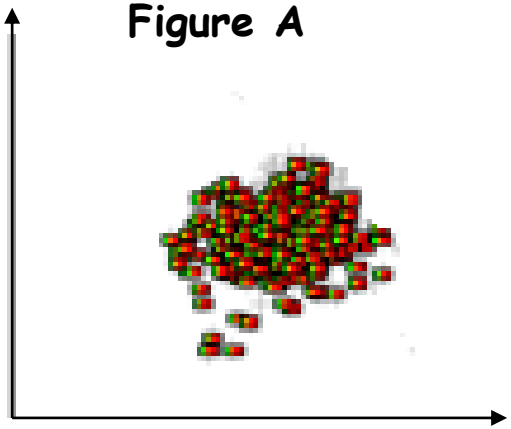
FEW WORDS ON QUALITY CONTROLS (3)

The main (non-exhaustive) steps for data quality control are :

- To check if cases and controls are genetically homogeneous for « neutral » SNPs

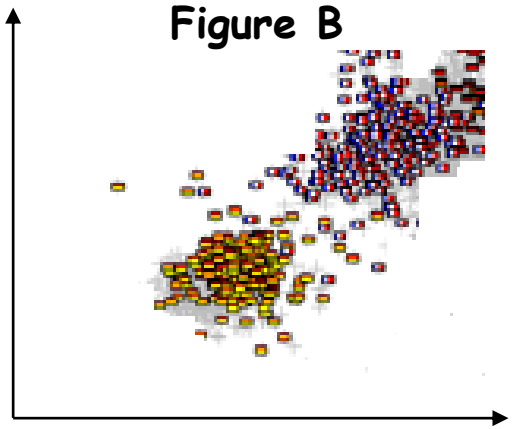
  Principal Componant Analysis (PCA)

Figure A



Cases (yellow dots) & controls (red dots) are pooled
=> homogeneity => good

Figure B



Cases (yellow dots) & controls (purple dots) are not pooled
=> heterogeneity => not good

POPULATION STRATIFICATION

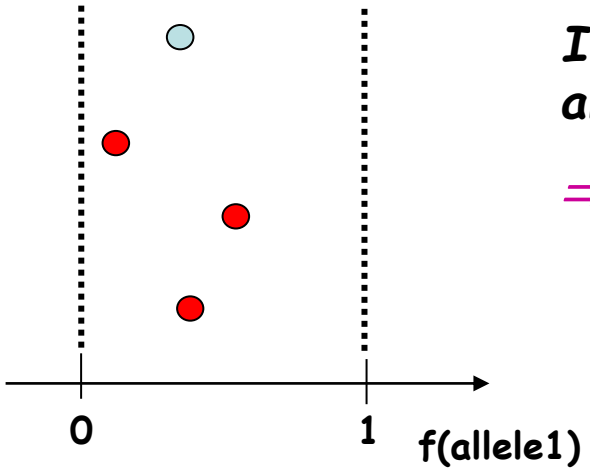
A population can be composed from ≠ sub-population having ≠ allele frequencies

Total Population

Sub-pop 1

Sub-pop 2

Sub-pop 3



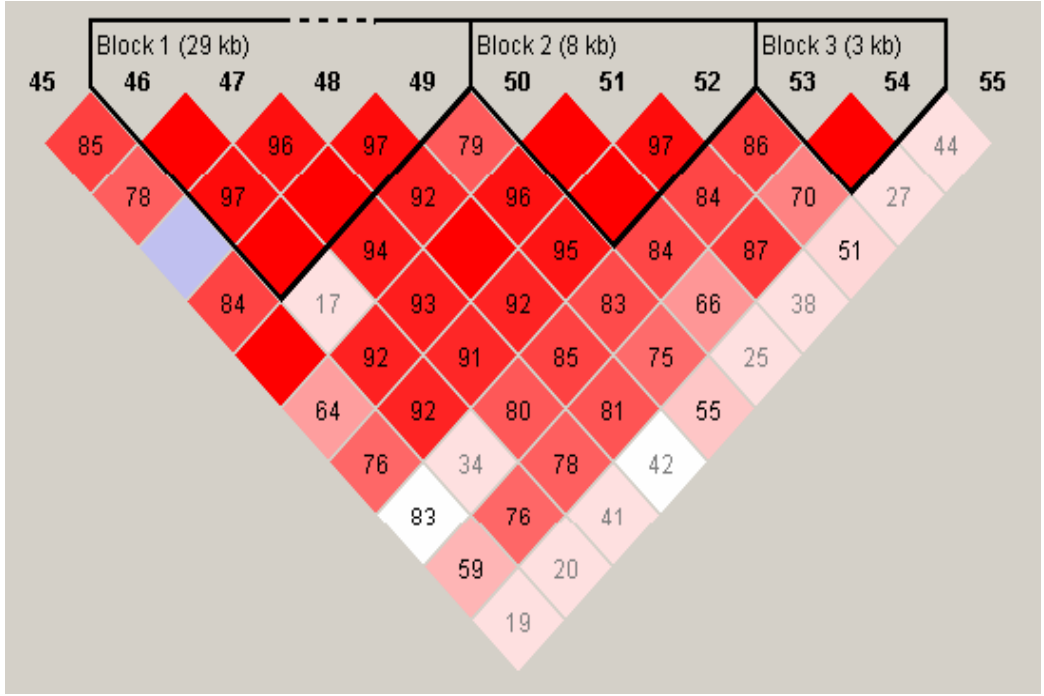
If cases are from sub-pop 2 and controls are from sub-pop 1

⇒ *an association will be found*

- Consequence of pop. stratification : **false positive association**
- Remedies to pop. stratification by matching cases and controls by ethnic background

Then use adapted test : Mantel-Haenszel or adjusted logistic regression by stratum…

USE OF LINKAGE DISEQUILIBRIUM



Measures that can be used:

- r^2*100
- $D'*100$

1. A priori to select a SNP (tagSNP) in a linkage disequilibrium block if a huge number of SNPs are analysed

2. A posteriori to search for the causal variant

LINKAGE DISEQUILIBRIUM (LD) DEFINITION

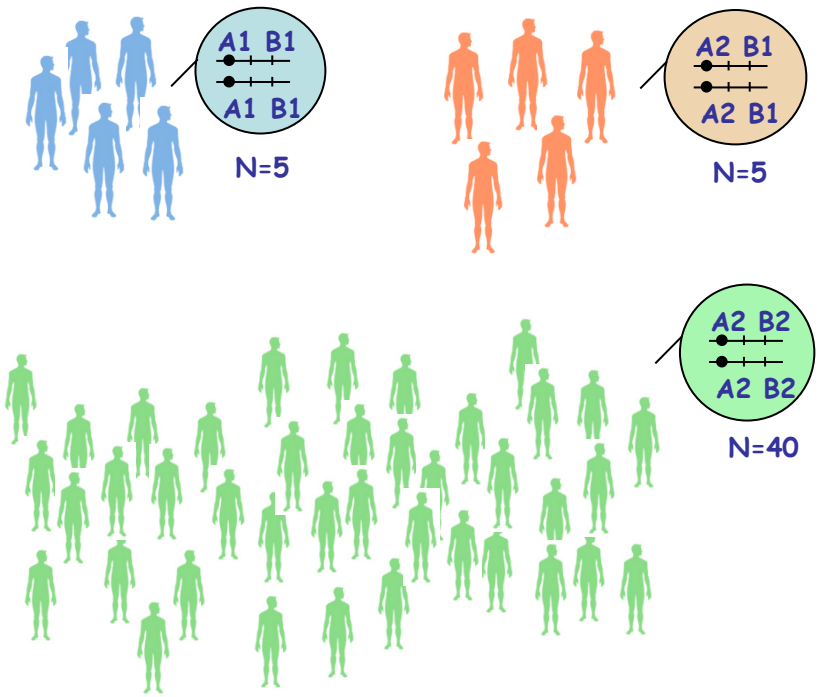
- Gametic disequilibrium : Preferential association between alleles taken at ≠ loci
- Linkage disequilibrium : Gametic disequilibrium that persists due to a genetic linkage (the 2 loci are close)
- Examples : - SNP A with 2 alleles : A1 (p1=0.1) et A2 (p2=0.9)
  - SNP B with 2 alleles : B1 (q1=0.3) et B2 (q2=0.7)

Equilibrium		Disequilibrium	
Random association of alleles		*Preferential association between alleles*	
Haplotypic freq.	Example	Haplotypic freq.	Example if A1 is always with B1
$f(A_1,B_1) = f(A_1)*f(B_1)$	0.1*0.3=0.03	$f(A_1,B_1) \neq f(A_1)*f(B_1)$	0.1
$f(A_1,B_2) = f(A_1)*f(B_2)$	0.1*0.7=0.07	$f(A_1,B_2) \neq f(A_1)*f(B_2)$	0
$f(A_2,B_1) = f(A_2)*f(B_1)$	0.9*0.3=0.27	$f(A_2,B_1) \neq f(A_2)*f(B_1)$	0.2
$f(A_2,B_2) = f(A_2)*f(B_2)$	0.9*0.7=0.63	$f(A_2,B_2) \neq f(A_2)*f(B_2)$	0.7

EXAMPLE : POPULATION IN LD FOR 2 SNP

If 2 linked loci studied : - A with 2 alleles : A1 (p1=0.1) et A2 (p2=0.9)
- B with 2 alleles : B1 (q1=0.2) et B2 (q2=0.8)

Population n°1 (n=50)



⇒ **Linkage disequilibrium**
(2 under-represented haplotypes)

Observed frequency of haplotypes:

- f(A1,B1) = (2*5) / 100 = 0,1
- f(A2, B1) = (2*5) / 100 = 0,1
- f(A2, B2) = (2*40) / 100 = 0,8
- f(A1, B2) = 0

Expected frequency if no LD:

- f(A1,B1) = f(A1) * f(B1) = 0,1*0,2 = 0,02
- f(A2, B1) = 0,9*0,2 = 0,18
- f(A2, B2) = 0,9*0,8 = 0,72
- f(A1, B2) = 0,1*0,8 = 0,08

EXAMPLE : POPULATION AT EQUILIBRIUM FOR 2 SNP

If 2 linked loci studied : - A with 2 alleles : A1 (p1=0.1) et A2 (p2=0.9)
- B with 2 alleles : B1 (q1=0.2) et B2 (q2=0.8)

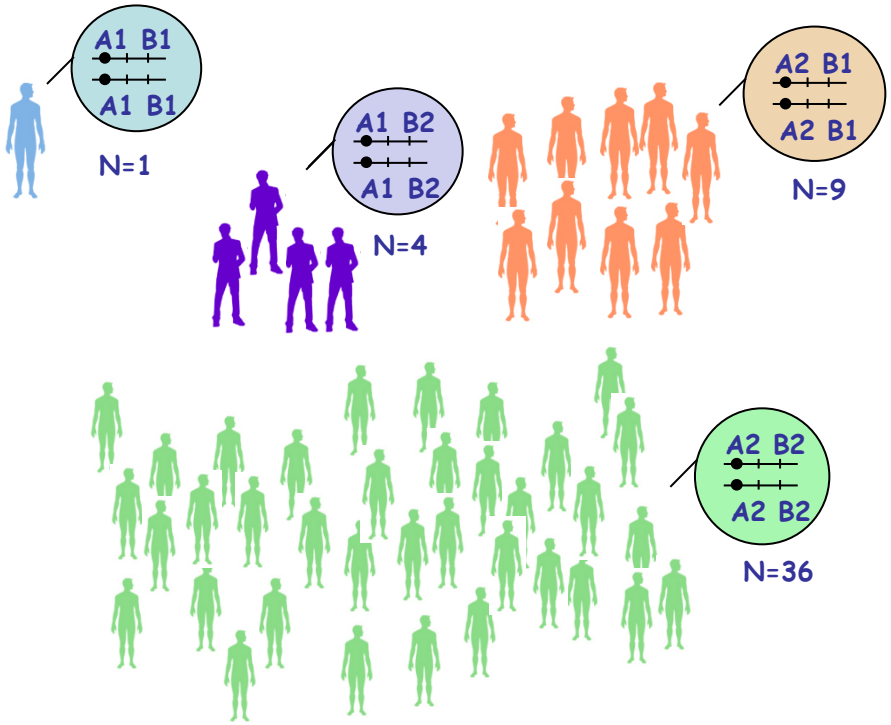
Observed frequency of haplotypes :

- f(A1,B1) = (2*1) / 100 = 0,02
- f(A2, B1) = (2*9) / 100 = 0,18
- f(A2, B2) = (2*36) / 100 = 0,72
- f(A1, B2) = (2*4) / 100 = 0,08

Expected frequency if no LD :

- f(A1,B1) = f(A1) * f(B1) = 0,1*0,2 = 0,02
- f(A2, B1) = 0,9*0,2 = 0,18
- f(A2, B2) = 0,9*0,8 = 0,72
- f(A1, B2) = 0,1*0,8 = 0,08

Population n°2 (n=50)



⇒ **Linkage equilibrium**
(random association between alleles)

MEASURES OF LINKAGE DISEQUILIBRIUM

Consider 2 bi-allelic genes : A with 2 alleles : A1 (p1) et A2 (p2)
B with 2 alleles : B1 (q1) et B2 (q2)

- Δ : $\Delta = f(A1, B1) - f(A1)\, f(B1)$

  $\Delta \in [-0.25, +0.25]$

- D' : $D' = \Delta / D_{max}$ (Lewontin, 1964)

  où $D_{max} = \begin{cases} \text{Min}\{p_1 q_2 , p_2 q_1\} & \text{si } \Delta > 0 \\ \text{Min}\{p_1 q_1 , p_2 q_2\} & \text{si } \Delta < 0 \end{cases}$

  $D' \in [-1, +1]$

- r^2 : $r^2 = \Delta^2 / (p_1\ p_2\ q_1\ q_2)$ (Hill et Robertson, 1968)

  $r^2 \in [0, +1]$

EXAMPLE OF CALCULATION OF LD MEASURES

- Allelic and haplotypic frequencies at the 2 loci studied

  A with 2 alleles : A1 (p1=0.1) et A2 (p2=0.9)
  B with 2 alleles : B1 (q1=0.3) et B2 (q2=0.7)

  f(A1, B1) = 0.05 f(A2, B1) = 0.25
  f(A1, B2) = 0.05 f(A2, B2) = 0.65

- Measures of LD
  - Δ_{A1B1} = f(A1, B1) - f(A1) * f(B1) = 0.05 – 0.1*0.3 = 0.02
  - $D'_{A1B1} = \Delta_{A1B1}$ / min {0.1*0.7; 0.9*0.3} = 0.02/0.07 = 0.29
  - $r^2 = \Delta^2_{A1B1}$ / (0.1*0.9*0.3*0.7)= 0.02^2/0.0189 = 0.02

- If equilibrium :
  - $\Delta = 0$
  - $D' = 0$
  - $r = 0$

High LD if r^2 (or D') > 0.8

WHAT DID I REMEMBER ?
(2)

1. Allez sur wooclap.com
2. Entrez le code d'événement dans le bandeau supérieur

Code d'événement
NHAUGF

1. Envoyez @NHAUGF au 06 44 60 96 62
2. Vous pouvez participer

Désactiver les réponses par SMS

TEST FOR ALLELIC ASSOCIATION

- Contingency table for a di-allelic locus
  - *Assume a multiplicative effect of allele*

	Cases	Controls	total
Allele 1	O11=a	O12=b	R1
Allele 2	O21=c	O22=d	R2
total	C1	C2	N

- Test for association
  - H0: No allelic association
    H1: Association
  - Homogeneous χ^2 (on 1-df)

$$\chi^2_{i,j} = \Sigma\ [(Oij - (Ri*Cj)/N)]^2 / ((Ri*Cj)/N)$$

df = (no row -1) * (no column -1)

- Allelic risk (2 vs 1) estimated by the odds ratio (OR)

OR = (c*b) / (a*d)

95% CI =exp [ln(OR) ± 1.96*√var(lnOR)]

$$Var(lnOR) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

EXAMPLE OF ALLELIC ASSOCIATION TEST

Search for an association between 1 SNP (A and B alleles) and a disease

- Observed count:

	A	B	total
Cas	175	265	440
Témoins	380	200	580
total	555	465	1020

- Allelic risk for B allele :

$$OR_{B/A} = (265 \times 380) / (175 \times 200) = 2,88$$

- Confidence Interval (CI):
  - Var(lnOR) = 1/175 + 1/265 + 1/380 + 1/200 = 0,017
  - CI for ln(OR) : ln(2,88) ± 1.96 √(0,017)

CI = [2,23-3,72]

- Homogeneity test:
  - $\chi^2 = 66,85$
  - Nb of degrees of freedom = (2-1) * (2-1) = 1

Ho rejected => association

TEST FOR GENOTYPIC ASSOCIATION

- Table for a di-allelic locus and a general genotypic model

	Cases	Controls	total	OR	95% CI
Genotype 11	O11=a	O12=b	R1	-	-
Genotype 12	O21=c	O22=d	R2	cb/ad	exp[ln(OR)±1.96√(1/a+1/b+1/c+1/d)]
Genotype 22	O31=e	O32=f	R3	eb/af	exp[ln(OR)±1.96√(1/a+1/b+1/e+1/f)]
total	C1	C2	N		

- *Test of genotypic association with homogeneity χ^2 (on 2-df)*
- *Genotype 11 is considered as the baseline*

- Other genotypic coding
  - Dominant model : pool 12+22 --> 2x2 table
  - Recessive model : pool 12+11 --> 2x2 table

Nb df	threshold	P-value
1	3,84	0,05
2	5,99	0,05
3	7,81	0,05

- *If the size of an expected category <5, use fisher exact test*
- *H0 rejected if p-value < 0,05*

EXAMPLE OF GENOTYPIC ASSOCIATION TEST

Obs. count	Cases	Controls	total	OR	95% CI
Genotype 11	72	84	156	-	-
Genotype 12	54	31	85	?	?
Genotype 22	46	13	59	?	?
total	172	128	300		

=> Genotypes 12 et 22 ↗ disease risk compared to 11 genotype

=> The risk associated to genotype 22 is x2 compared to the risk of genotype 12

- Homogeneity χ^2

Exp. count	Cases	Controls	total
Genotype 11	89.44	66.56	156
Genotype 12	48.73	36.27	85
Genotype 22	33.83	25.17	59
total	172	128	300

$\chi^2(2) = 19.57$
P-val = 0.000056

ANOTHER APPROACH : LOGISTIC REGRESSION

Dependant variable	Independant variable
Y=1 (cases)	X=1 (allele 2 of a SNP)
0 (controls)	0 (allele 1 of a SNP)

- Probability of being affected (or unaffected)

$$P = P(Y=1/X) = \exp\theta / (1 + \exp\theta)$$

$$1-P = P(Y=0/X) = 1 / (1 + \exp\theta)$$

where θ = Logit P
$= \alpha + \beta X$

- Allelic risk (2 vs 1) estimated by the odds ratio (OR)

$$OR = \exp(\beta)$$

$$95\%\ CI = \exp[\ \beta \pm 1.96 * \sqrt{var(\beta)}\ ]$$

- Likelihood ratio test ($L = \Pi_{i=1..n} P^{Yi} * (1-p)^{1-Yi}$; n=nb obs)

$$LR = -2 \ln [\ L(\beta=0) / L(\beta max)\ ] \sim \chi^2(1\ df)$$

LOGISTIC REGRESSION : EXAMPLE

Variable	Regressive coefficient	Standard error	Ln L
baseline	-5.3	1.134	-53.68
SNP	0.111	0.024	

↳ Ln L(β=0) = -63.14

- Likelihood ratio test

$$LR = -2 \ln [L(\beta=0) / L(\beta max)] = -2 \ln(L(\beta=0) - (-2) \ln(\beta max)$$
$$= -2 * (-63.14) - (-2) * (-53.68) = 18.92 \sim \chi^2(1)$$

=> P-val = $1.36*10^{-5}$

- OR et 95% CI

OR = exp(0.111) = 1.12

95% CI =exp [$\beta \pm 1.96*\sqrt{var(\beta)}$] = [1.07 ; 1.17]

- Test de Wald : $T = \beta / \sqrt{var(\beta)} = 0.111/0.024 = 4.61 \sim N(0,1)$

OTHER CODING SCHEME FOR THE INDEPENDANT VARIABLE

- General genotypic coding scheme

⇨ Introduction of 2 indicator variables

$$\text{Logit } P = \alpha + \beta_1 X_1 + \beta_2 X_2$$

Genotype	X1	X2
11	0	0
12	1	0
22	0	1

⇨ Baseline genotype

⇨ $OR_{12 \text{ vs } 11} = \exp(\beta 1)$

⇨ $OR_{22 \text{ vs } 11} = \exp(\beta 2)$

- Additive genotypic coding scheme

⇨ Multiplicative model => trend test

Genotype	X
11	0
12	1
22	2

⇨ $OR_{12 \text{ vs } 11} = \exp(\beta)$

⇨ $OR_{22 \text{ vs } 11} = \exp(2*\beta) = \exp(\beta)^2$

- Dominant or recessive coding scheme

	Dom	Rec
Genotype	X	X
11	0	0
12	1	0
22	1	1

MULTIVARIATE LOGISTIC REGRESSION

- Advantage of logistic regression

  To take into account other variables (genetic and/or environmental) to test association of a SNP with a disease

$$\text{Logit } P = \alpha + \beta_1 X_1 + \beta_2 X_2 \,..\, + \beta_\nu X_\nu$$

- Test of association for a SNP (X1)

  $H0 : \beta 1 = 0 \quad ; \quad H1 : \beta 1 \neq 0$

  $OR_{\text{adjusted for other variables}} = \exp(\beta_1)$

  $95\%\ CI = \exp\left[\ \beta_1 \pm 1.96 * \sqrt{var(\beta_1)}\ \right]$

- *Selection of the best model (combination of SNPs associated with the disease) – Example: stepwise procedure*

GENOME-WIDE ASSOCIATION ANALYSES

In the context of genome-wide association analyses (GWAs), since thousands of SNPs are analysed, the p-value must be corrected for multiple testing

- **If p-value=5%, expected number of false positive results:**

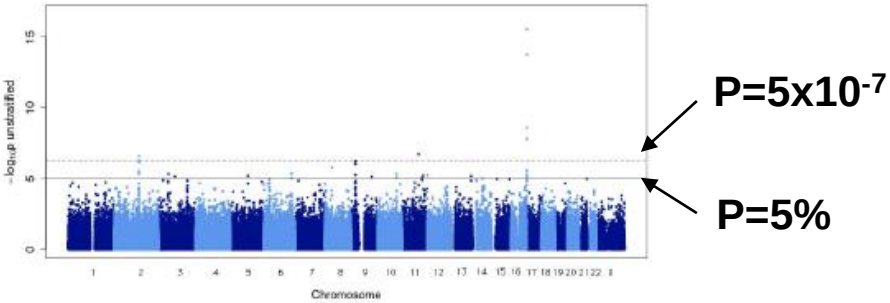
  => on 100 tests where H0 should be true, 5 false positives

  => on 100000 tests, 5000 false positives !

- **Different methods allow to correct p-value for multiple testing – the most simplest = Bonferroni correction**

  $p\text{-value}_{Corr}$ = 5% / nb of independants SNP tested

  Ex: if 100000 tests, $p\text{-value}_{Corr} = 0{,}05 / 100000 = 5\times10^{-7}$

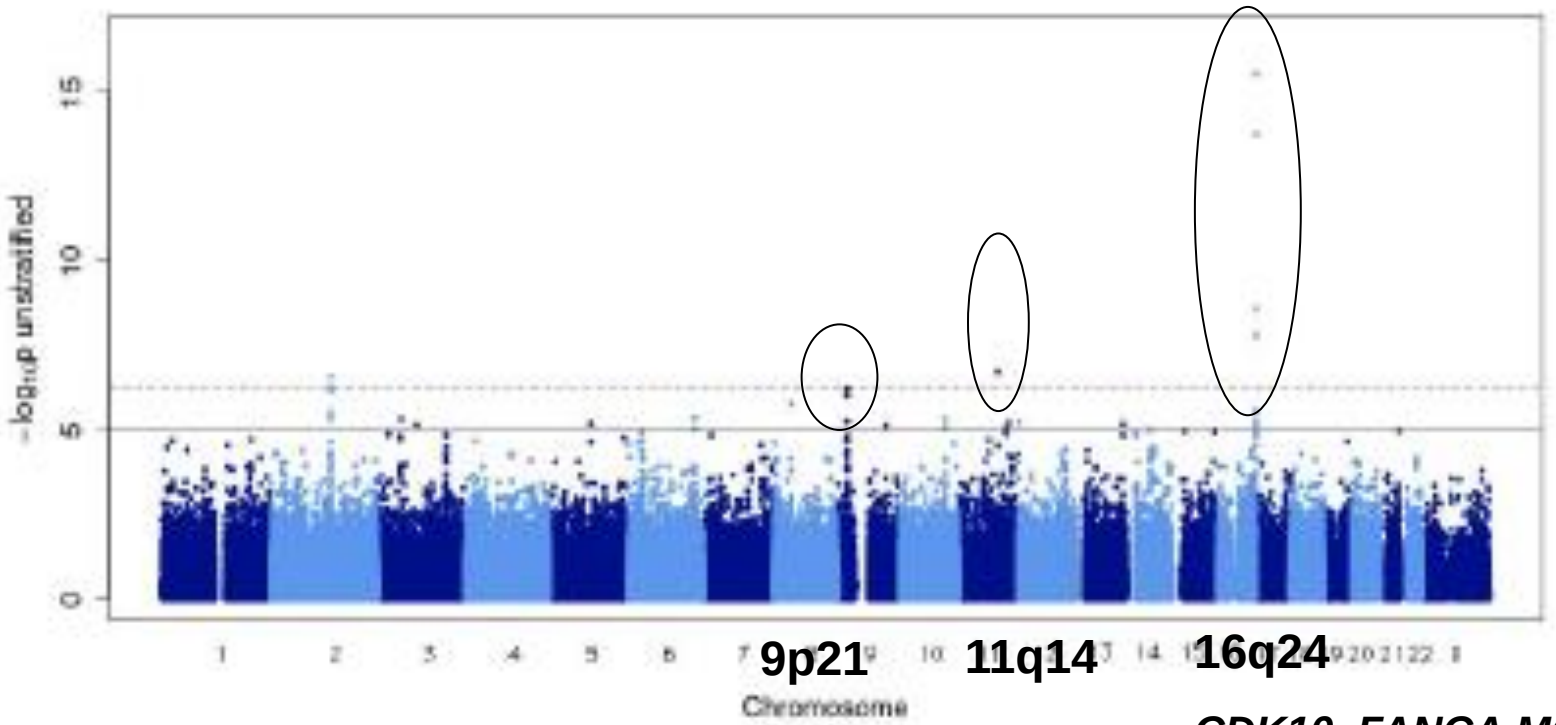


GENOME-WIDE ASSOCIATION STUDY FOR MELANOMA

(GenoMEL Consortium)

Bishop DT, Demenais F et al, Nat Genet, 2009

Corrected threshold for multiple testing consideration



$p<5x10^{-7}$

CDK10, FANCA,MC1R

GENOME-WIDE ASSOCIATION STUDY FOR OTHER CANCERS

Easton et al, Hum Mol Genet, 2008

Locus	Chromosome	SNP(s)	MAF[a]	Per allele OR[b]	*P*-value[c]
Breast cancer					
2q35	2	rs13387042	0.50	1.21	10^{-13}
MAP3K1	5	rs889312	0.28	1.13	7×10^{-20}
MRPS30	5	rs10941679	0.25	1.19	3×10^{-11}
ECHDC1, RNF146	6	rs2180341	0.27	1.41	3×10^{-8}
8q24	8	rs13281615	0.40	1.08	10^{-12}
FGFR2	10	rs2981582	0.38	1.26	2×10^{-76}
LSP1	11	rs3817198	0.30	1.07	3×10^{-9}
TNRC9, LOC643714	16	rs3803662	0.25	1.20	10^{-36}
Prostate cancer					
2p15	2	rs721048	0.19	1.15	8×10^{-9}
3p12	3	rs2660753	0.11	1.18	3×10^{-8}
6q25	6	rs9364554	0.29	1.17	6×10^{-10}
7q21	7	rs6465657	0.46	1.12	10^{-9}
JAZF1	7	rs10486567	0.77	1.12	10^{-7}
8q24	8	rs1447295, DG8S737	0.10	1.62	3×10^{-11}
8q24	8	rs6983267	0.50	1.26	9×10^{-13}
8q24	8	rs16901979, hapC	0.03	2.1	3×10^{-15}
HNF1B	17	rs4430796	0.49	1.24	10^{-11}
HNF1B	17	rs11649743	0.80	1.28	2×10^{-9}
17q	17	rs1859962	0.46	1.25	3×10^{-10}
MSMB	10	rs10993994	0.40	1.25	9×10^{-29}
CTBP2	10	rs4962416	0.27	1.17	3×10^{-8}
11q13	11	rs7931342	0.51	1.19	2×10^{-12}
KLK2/KLK3	19	rs2735839	0.85	1.20	2×10^{-18}
Xp11	X	rs5945619	0.36	1.19	2×10^{-9}
Colorectal cancer					
8q24	8	rs6983267	0.50	1.17	3×10^{-11}
SMAD7	18	rs4939827	0.53	1.15	10^{-12}
CRAC1	15	rs4779584	0.19	1.26	4×10^{-14}
EIF3H	8	rs16892766	0.07	1.25	3×10^{-18}
10p14	10	rs10795668	0.67	1.12	3×10^{-13}
11q23	11	rs3802842	0.29	1.10	6×10^{-10}
Lung cancer					
CHRNA3/CHRNA5	15	rs8034191	0.33	1.30	5×10^{-20}
Melanoma					
TYR	11	rs1126809 (R402Q)	0.30	1.21	10^{-7}
ASIP	20	rs1015362/ rs4911414 Haplotype	0.08	1.45	10^{-9}
	20	rs910873 rs1885120	0.09	1.75	10^{-15}

REASONS FOR A GENETIC ASSOCIATION

- The marker locus contains the causal (functional) variant

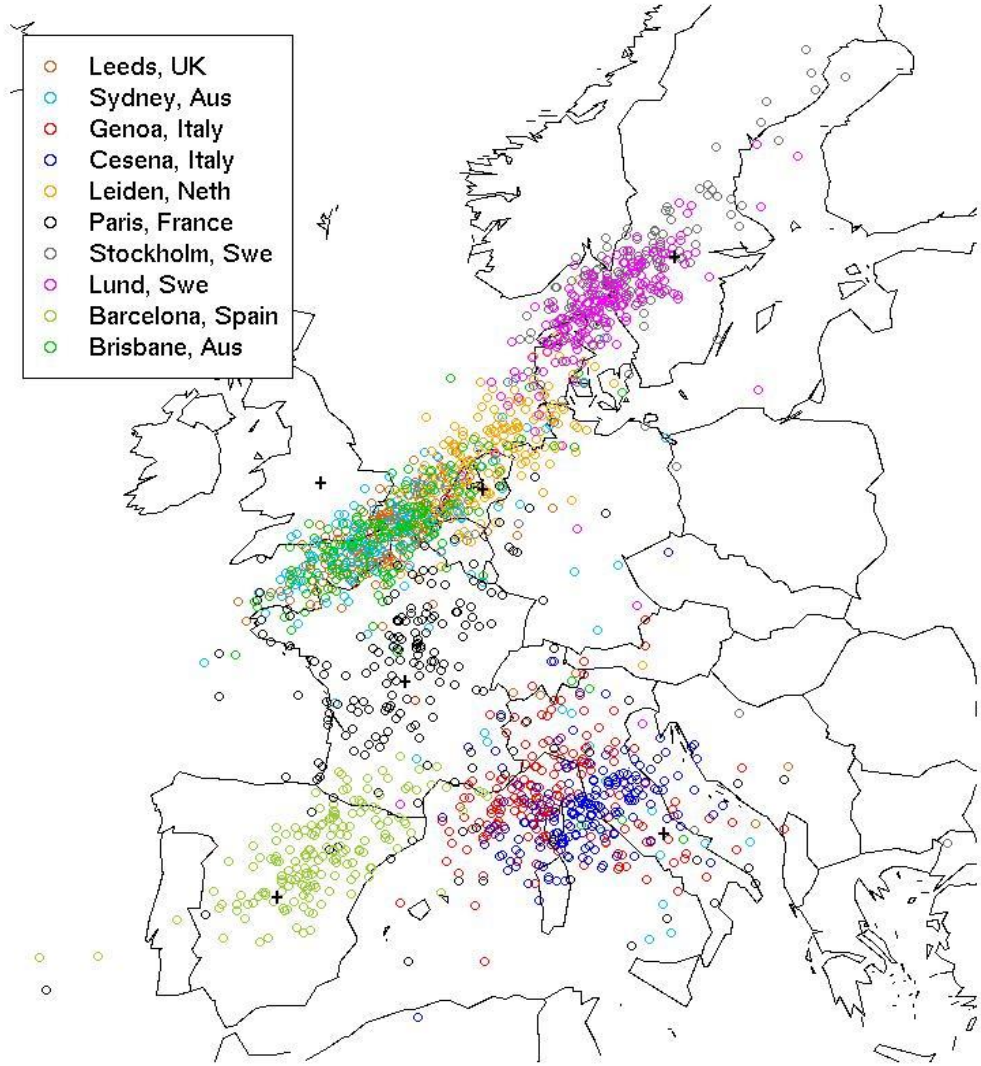
Causal variant --> Disease

- The marker locus is in linkage disequilibrium (LD) with the locus including the functional variant

Marker locus -LD-- Causal variant --> Disease

- The association is due to confounding by population stratification

EXAMPLE OF GENETIC DIVERSITY – PCA RESULTS



ASSOCIATION ANALYSES

SEARCH FOR CAUSAL VARIANTS

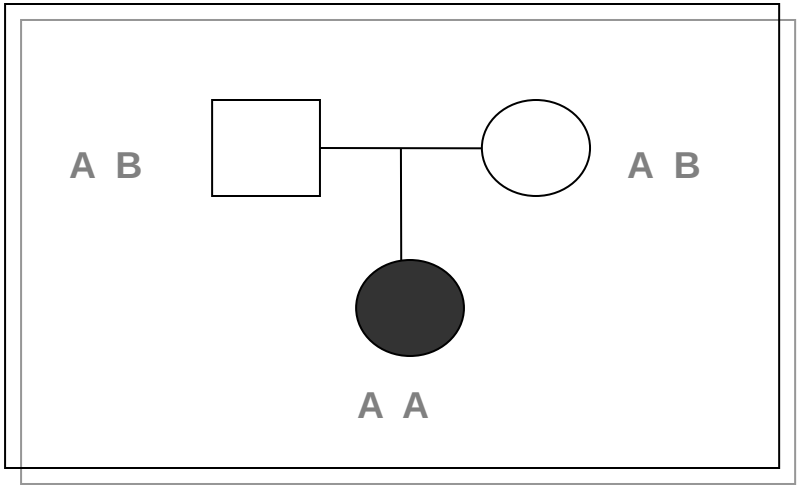
-

FAMILIAL

ASSOCIATION ANALYSES

Transmission Disequilibrium Test (TDT)

Trios: 2 parents + 1 affected offspring



Sample: N trios

Only <u>heterozygous</u> parents are informative

N1 = no times when A transmitted from an heterozygous to affected offspring
N2 = no times when B transmitted from an heterozygous to affected offspring

H0 : N1 = N2 = (N1+N2)/2 (No association or No linkage)
H1 : N1 ≠ N2

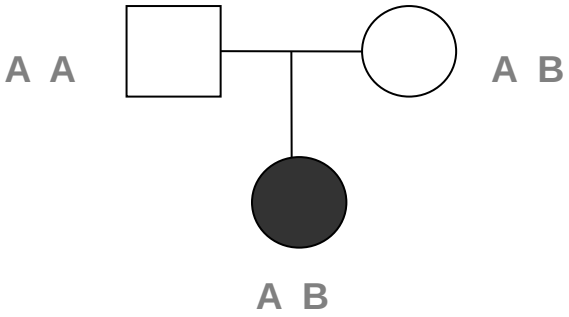
Test: $\chi 2 = (N1 - N2)^2/(N1 + N2)$ with 1 df

if significant test => Association & Linkage

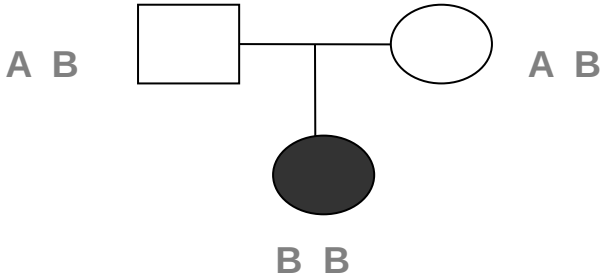
EXAMPLE

Given a sample of 100 trios of the following types

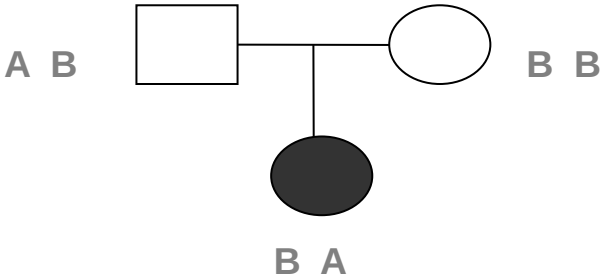
30 trios



50 trios



20 trios



N1 = nb A transmitted from an heterozygous to affected offspring = 20
N2 = nb B transmitted from an heterozygous to affected offspring = 130

H0: N1 = N2 = (N1+N2)/2 (No association or No linkage)
H1 : N1 ≠ N2 (association and linkage)

Test: $\chi 2$ (1 df) = $(130 - 20)^2 / 150 = 80{,}67 > 3.84$

=> Association & linkage

REASONS FOR A GENETIC ASSOCIATION

- The marker locus contains the causal (functional) variant

Causal variant --> Disease

- The marker locus is in linkage disequilibrium (LD) with the locus including the functional variant

Marker locus -LD-- Causal variant --> Disease

➢ *Advantage of family-based studies: avoids the problems of stratification that can occur with population-based studies*

➢ *Disadvantage of family-based studies: more difficult to collect*

WHAT DID I REMEMBER ?
(3)

1. Allez sur wooclap.com
2. Entrez le code d'événement dans le bandeau supérieur

Code d'événement
NJBZZS

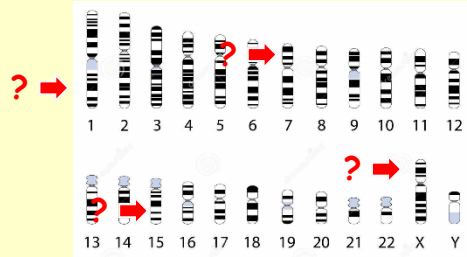
1. Envoyez @NJBZZS au 06 44 60 96 62
2. Vous pouvez participer

Désactiver les réponses par SMS

PART 2

GENETIC LINKAGE ANALYSES

Genetic linkage analyses



positional cloning or functional studies of candidate genes

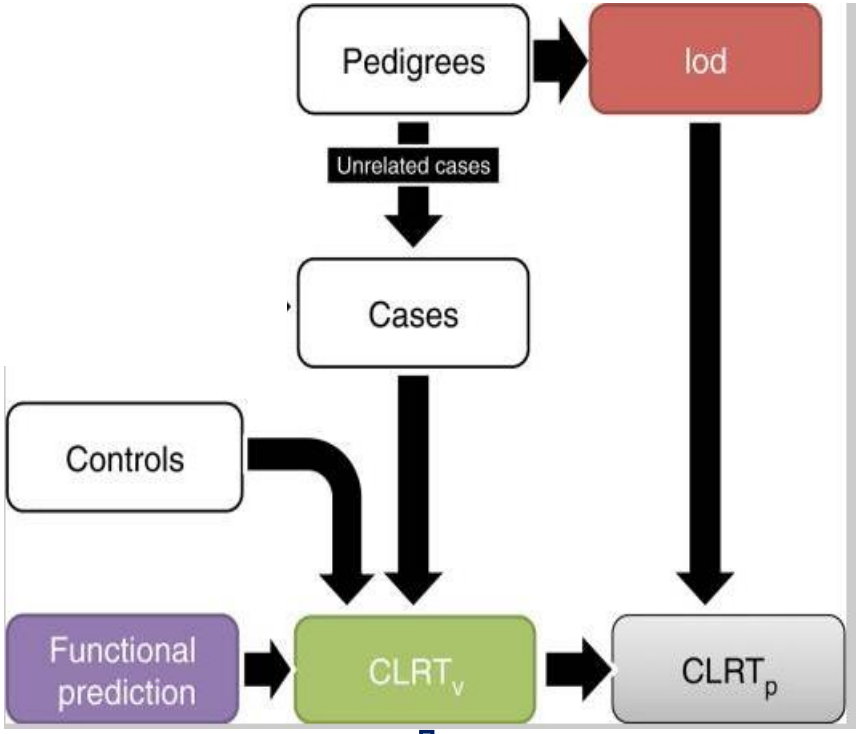
Location of disease genes on the genome

AIM OF A LINKAGE ANALYSIS

- At the beginning :

  Locating a disease gene on the genome using genetic markers (MK) whose location is known

- Today :
  - Combine linkage and association scores to identify new variants with familial data
  - Example : pVAAST (Hu et al., 2014)



=> Linkage score

⇩

Association score

GENERAL PRINCIPLE OF LINKAGE ANALYSES

- **Aim :**

  Locating a disease gene on the genome using genetic markers (MK) whose location is known

  - **To locate a Mendelian entity : lodscore method**
    - ➢ *Study of disease-marker co-transmission in families*

      + estimation of the recombination rate between the disease gene and the MKs

  - **To locate a susceptibility gene: sib-pairs method**

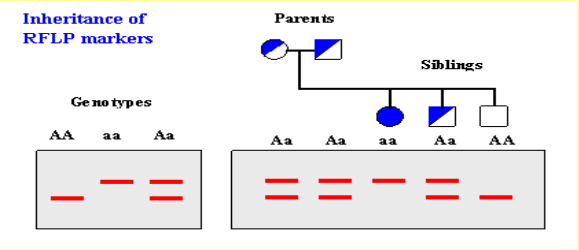
- **Analysed data:**
  - **Familial sample (nuclear/3 generations) identified by affected subject and with at least 2 children**
  - **Including : phenotypic information + MK genotypes**

PREFERENTIALLY USED MARKERS

- **RFLP (Restriction Fragment Length Polymorphism)**
  - The polymorphism is due to the length of the fragments
  - di-allelic

Ex: EcoRI restriction enzyme

AATTC GAATTC TAATTC



- **MICROSATELLITES**
  - Tandem repeat of short nucleotide motifs
  - Multi-allelics

Ex: (CA)n

➢ *The more polymorphic a marker, the more useful ("informative") it will be for estimating θ (=> MK most used= microsatellites)*

GENETIC LINKAGE ANALYSES

MENDELIAN ENTITY LOCATION

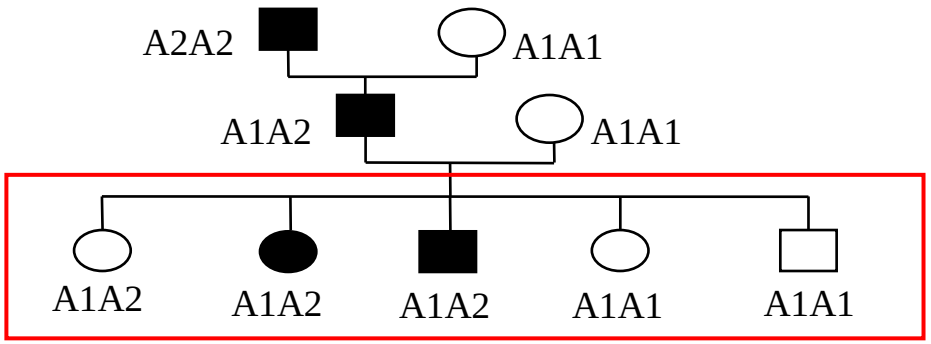
-

LODSCORE METHOD

RECOMBINATION RATE ESTIMATION

- Estimation of the recombination rate θ (fct of distance) :

  θ = nb of recombination events / total nb of events



θ calculated from gametes received by children

- Steps to estimate θ:

1) Establish the phenotype ⇔ genotype relationship to assign disease gene genotypes in families

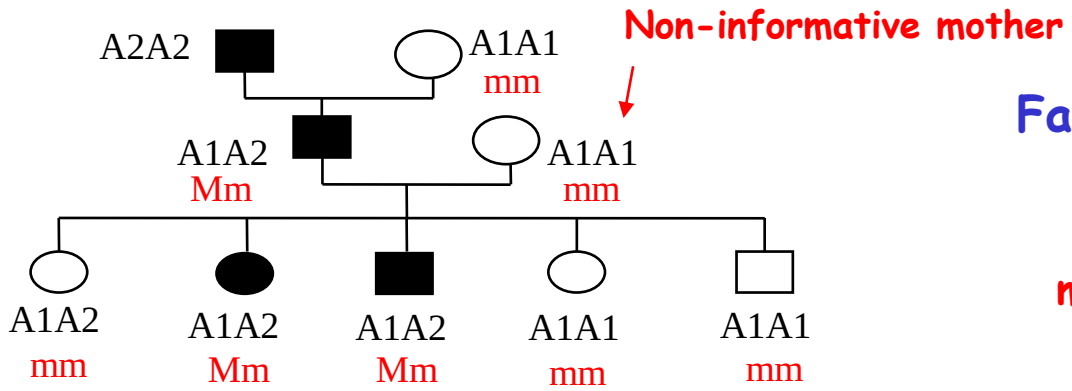
2) Determine the allelic phase in double-heterozygous parents
   (only informative subject to follow gene-MK segregation)

3) For each child, establish whether a parental or recombinant gamete was transmitted by the double-heterozygous parents

EXAMPLE n°1 : DIRECT ESTIMATE OF θ

- **Genetic model:**
  - **Autosomal dominant disease**
  - **di-allelic gene (M=deleterious; m=wild-type) – P(affected/MM)=1**
    P(affected/Mm)=1
    P(affected/mm)=0

- **Example n°1: Estimation of θ possible directly**



Father's allelic phase:

A1m / A2M

maternal γ paternal γ

A2m	A2M	A2M	A1m	A1m
R	P	P	P	P

γ from the father trans. to the children

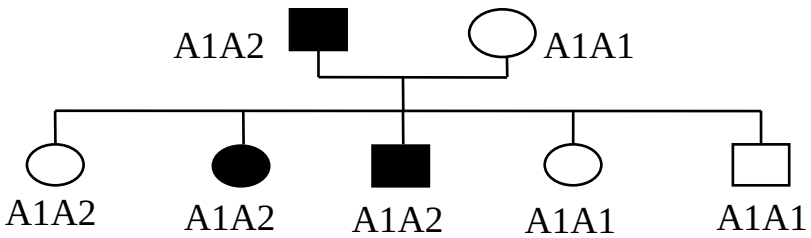
Types of γ : R=recombinant

P=parental

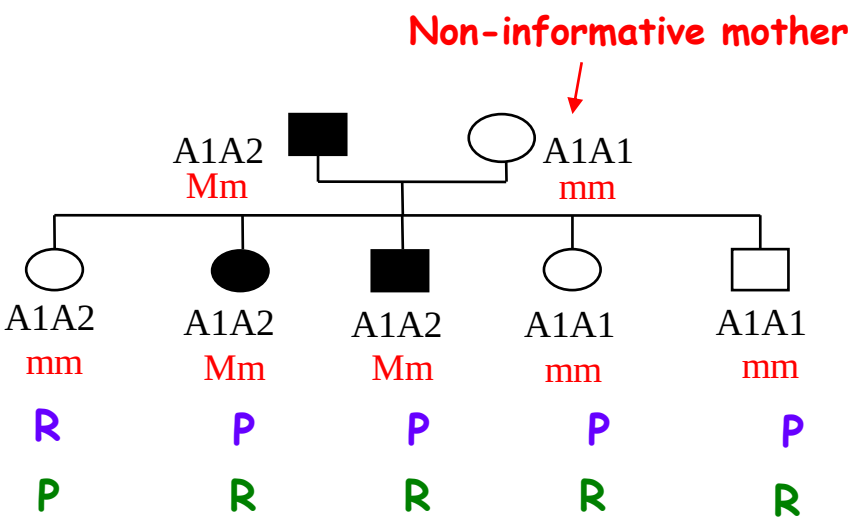
CONCLUSION : θ = 1/5 = 0,2

EXAMPLE n°2: ESTIMATION OF θ IMPOSSIBLE DIRECTLY

- Example n°2 : θ cannot be calculated directly



Genetic model identical to example n°1



Allelic phase of the father: indeterminate

Phase 1: A1m/A2M (proba ½)

Phase 2: A1M/A2m (proba ½)

Types of γ transmitted by father under phase 1

Types of γ transmitted by father under phase 2

➢ Statistical method to estimate θ ⇨ Lod-score test

LOD-SCORE TEST (Morton, 1955)

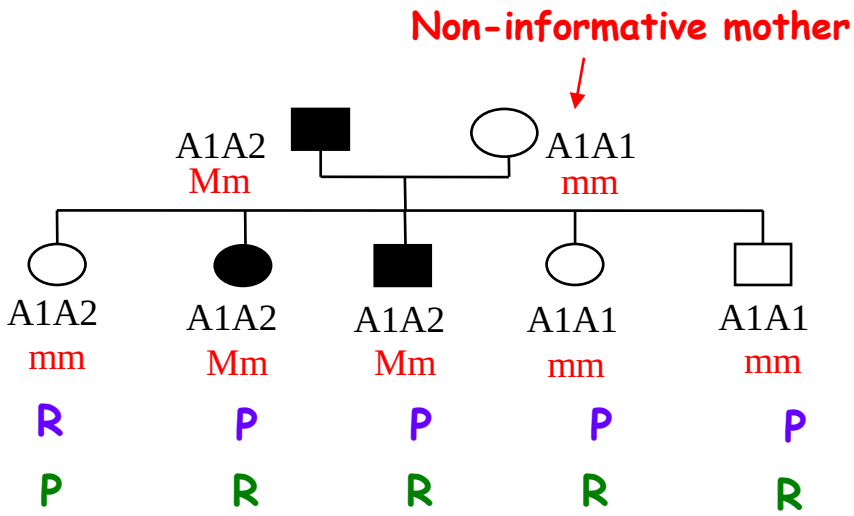
- **Aims**
  - Test if 2 loci are genetically linked or not
  - Estimate the recombination rate
- **Assumptions tested**
  - H0: Absence of genetic linkage (θ=0.5)
  - H1: Genetic linkage for estimated θ ($0 \leq \theta < 0,5$)
- **Calculation of the Lod-score for a family Fi and a given value of θ**

$$Z_i(\theta_1) = \log_{10} \frac{P(F_i / \theta_1)}{P(F_i / \theta=0,5)} = \log_{10} \frac{L_i(\theta_1)}{L_i(\theta=0,5)}$$

where $L_i(\theta_1) = \Sigma_{phases}$ P(phase) x $L_{i,phase}(\theta_1)$

and $L_{i,phase}(\theta_1)=\Pi_{nb\ enf}$ P(father gamete) x P(mother gamete)

EXAMPLE OF LIKELIHOOD CALCULATION



Allelic phase of the father: indeterminate

Phase 1: A1m/A2M (proba ½)

Phase 2: A1M/A2m (proba ½)

Types of γ transmitted by father under phase 1

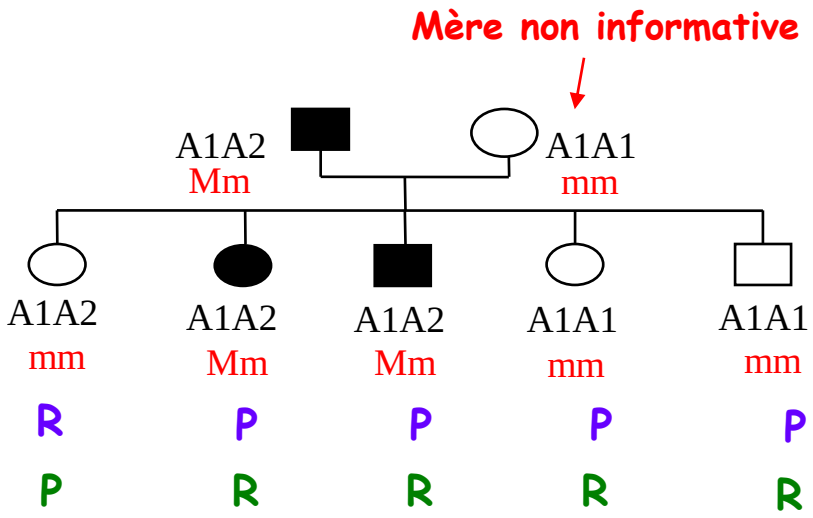
Types of γ transmitted by father under phase 2

- **Under phase 1:** $L_{i,phase1}(\theta)=(\theta/2)^1 \times [(1-\theta)/2]^4$
- **Under phase 2:** $L_{i,phase2}(\theta)=(\theta/2)^4 \times [(1-\theta)/2]^1$

➢ **Conclusion:** $L_i(\theta)= \frac{1}{2}(\theta/2)^1 \times [(1-\theta)/2]^4 + \frac{1}{2}(\theta/2)^4 \times [(1-\theta)/2]^1$

$$L_i(\theta)= (\tfrac{1}{2})^6 \times \theta \times (1-\theta) \times [(1-\theta)^3+\theta^3]$$

EXAMPLE OF LOD-SCORE CALCULATION



Allelic phase of the father: indeterminate

Phase 1: A1m/A2M (proba $\frac{1}{2}$)

Phase 2: A1M/A2m (proba $\frac{1}{2}$)

Types of γ transmitted by father under phase 1

Types of γ transmitted by father under phase 2

$$L_i(\theta) = (\tfrac{1}{2})^6 \times \theta \times (1-\theta) \times [(1-\theta)^3+\theta^3]$$

$$Z(\theta) = \log \left[\frac{(\tfrac{1}{2})^6 \times \theta \times (1-\theta) \times [(1-\theta)^3+\theta^3]}{(\tfrac{1}{2})^6 \times \tfrac{1}{2}^2 \times (\tfrac{1}{2})^2} \right]$$

$$Z(\theta) = \log [2^4 \times \theta \times (1-\theta) \times [(1-\theta)^3+\theta^3]]$$

THRESHOLDS TO CONCLUDE

- Calculation of the Lod-score in a sample of n families
  - $$Z(\theta) = \sum_{i=1}^{n} Z_i(\theta)$$
  - Find the value of θ that maximizes the Lod-Score (estimated θ_{max}) in the total sample

    + calculation of $Z(\theta)$ for given θ

- Thresholds
  - if $Z(\theta_{max}) \geq 3$ ➲ H0 rejected => linkage for θ_{max}
  - if $Z(\theta_i) \leq -2$ ➲ absence of genetic linkage for θ_i ($\theta_i \neq \theta_{max}$)
  - if $-2 < Z(\theta_{max}) < 3$ ➲ nothing can be concluded

    (families must be added to the analysed sample)

Propriétés statistiques

Morton a montré que pour de tels critères :

- **erreur de première espèce** α = proba (rejeter H0 / H0 vraie)

$$\alpha \leq 10^{-3}$$

- **erruer de seconde espèce** β = proba (rejeter H1 / H1 vraie)

$$\beta \leq 10^{-2}$$

- **fiabilité** ρ = proba (H0 vraie / on a conclu H1)

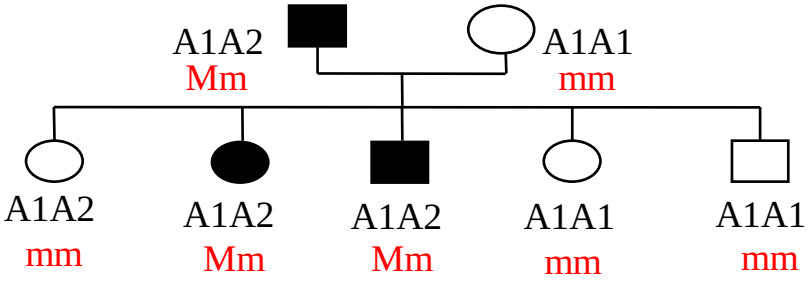
= probabilité que la liaison n'existe pas alors que $z(\theta 1)$ a dépassé 3

$$\rho \leq 0.05 \ \forall \ \theta 1$$

- **puissance** $P(\theta 1)$ = proba (rejeter H0 / θ est la vraie valeur)

$P(\theta) \geq 0.80$ $\forall \theta 1$ si la vraie valeur de $\theta < 0.10$

EXAMPLE OF INTERPRETATION OF A LOD-SCORE RESULT

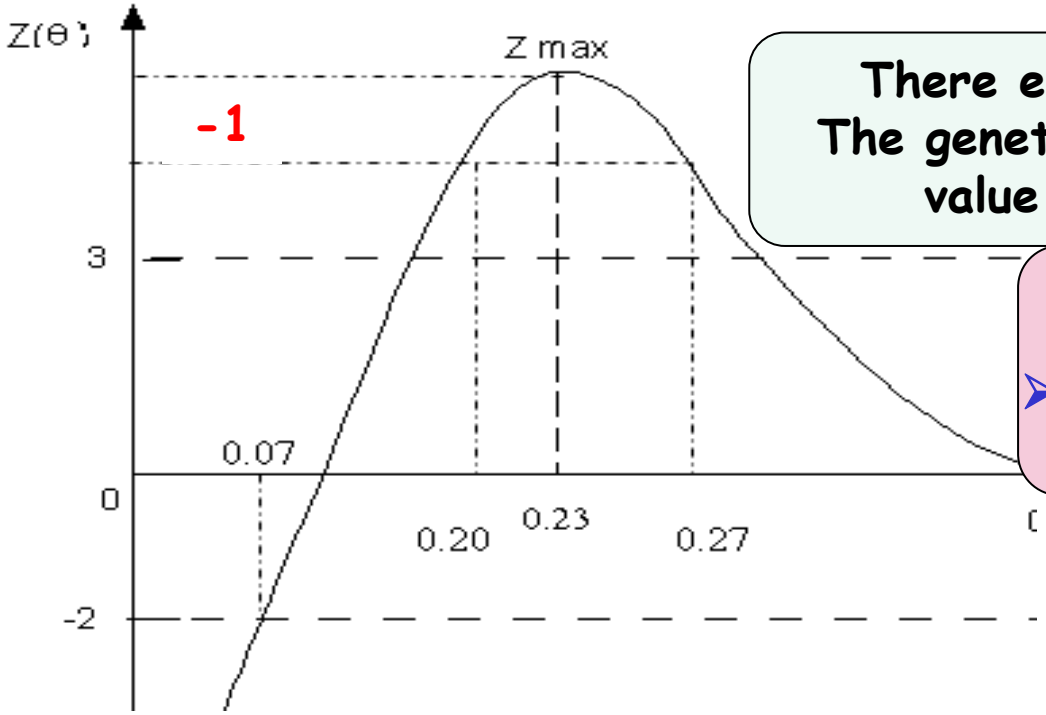


$$Z(\theta) = \log\left[2^4 \times \theta \times (1-\theta) \times \left[(1-\theta)^3+\theta^3\right]\right]$$

- **Estimation of θ maximizing the lod-score**
  - Performed using analysis software (LINKAGE)
  - θ_{max}= 0,212 et Z(0,212)=0,12

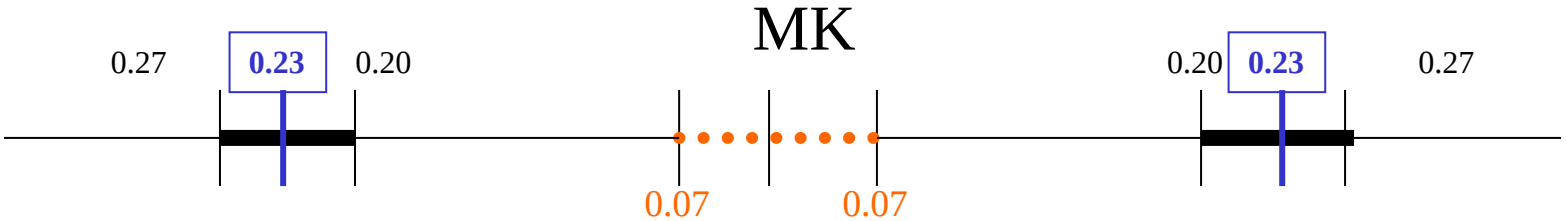
- **Conclusion**
  - $-2 < Z(\theta_{max}) < 3$ => nothing can be concluded
    - ➢ *families must be added to the sample*

CONFIDANCE INTERVAL



There exists a value of θ for which Z(θ)≥3
The genetic distance between the 2 loci is the value of θ which maximizes the Z fc°

Confidance Interval =
Possible regions around the maximum
- Values of θ such that the line Zmax-1 intersects the Lod-Score curve



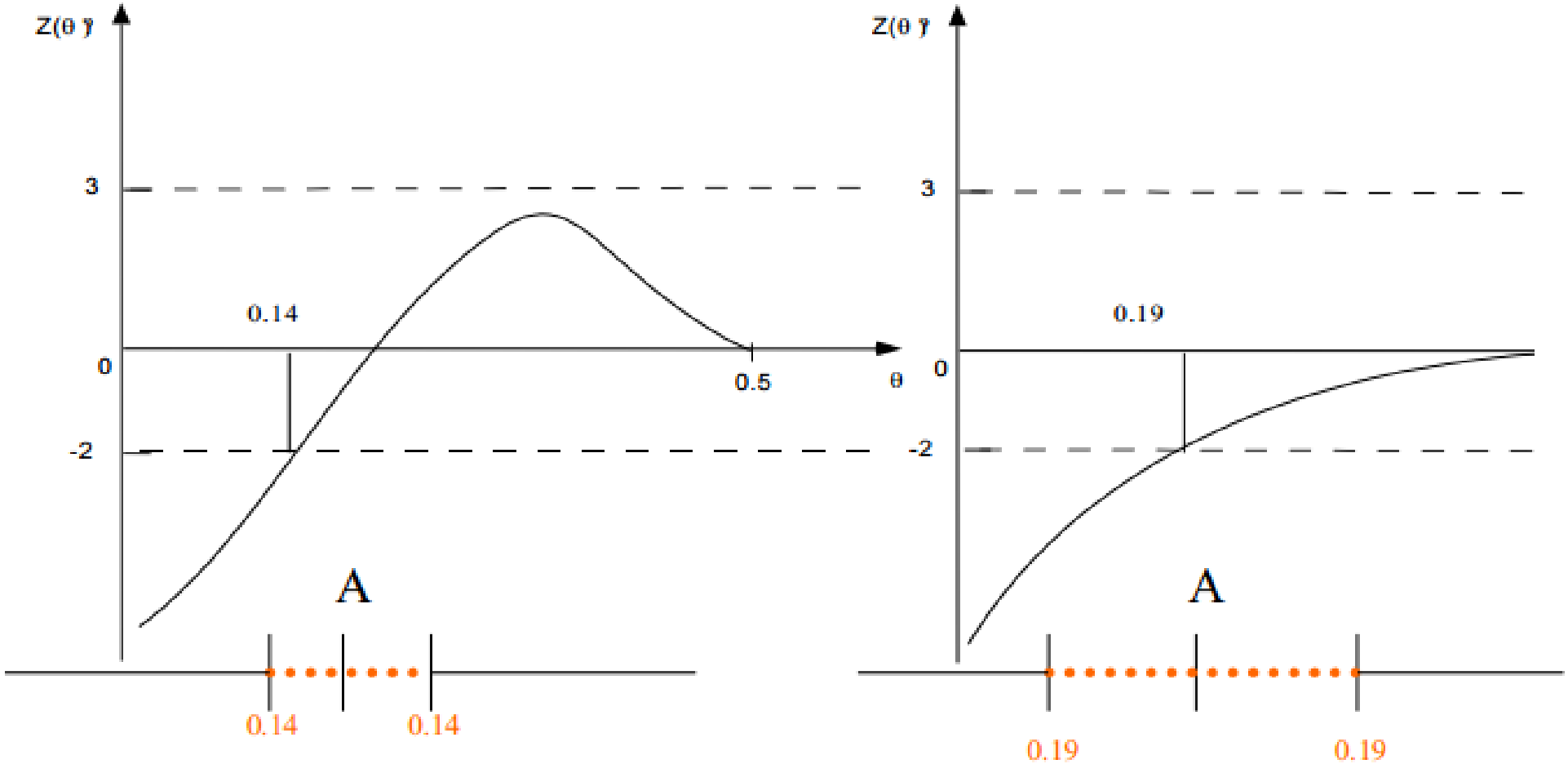
Impossible location of disease gene M

Most likely location of disease gene M

Possible location of the disease gene M

EXAMPLES OF LOD-SCORE CURVES

Genetic linkage is excluded for all $\theta \leq \theta_1$ if there is a value of θ_1 such that $Z(\theta_1) = -2$



Impossible location of disease gene M

RESULTS OF A LINKAGE ANALYSIS ON A FAMILIAL SAMPLE

θ:	0.0	0.1	0.2	0.3	0.4	0.5
1=2	-infini	21.71	18.00	13.14	7.21	0.00
1	-infini	2.24	2.28	1.82	1.06	0.00
2	-infini	4.15	3.48	2.55	1.41	0.00
3	6.02	5.11	4.08	2.92	1.58	0.00
4	6.02	5.11	4.08	2.92	1.58	0.00
5	6.02	5.11	4.08	2.92	1.58	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00

Z values in the total sample

Z values per family

(here 6 families)

- Z(0)=-infini --> at least 1 recombination is observed
- The disease locus is approximately located 10cM from the MK
- Families 3, 4 and 5 have the same lodscore
- Family 6 is not informative (L does not depend on θ)

  --> there is no double heterozygous parent

WHAT DID I REMEMBER ?
(4)

1. Allez sur wooclap.com
2. Entrez le code d'événement dans le bandeau supérieur

Code d'événement
QLRXDY

1. Envoyez @QLRXDY au 06 44 60 96 62
2. Vous pouvez participer

Désactiver les réponses par SMS

CONDITIONS OF APPLICATION AND DISADVANTAGES OF THE LOD-SCORE METHOD

- Lod-score method = model-dependent approach
  - The genetic model must be known
  - If genetic model unknown, the parameters *(allelic freq., penetrances, transmission mode)* can be estimated with recombination rate

    -> thresholds ?

- Impact of an error on used parameters
  - If genetic model of the disease not well known

    -> power loss, false exclusion possible
  - If error on marker allele frequency

    -> risk of false positives

ANALYSIS PROGRAMS

- Rockefeller website
  - allows to download genetic analysis software for free
  - http://linkage.rockefeller.edu/soft/

- Bi-point genetic linkage analysis programs
  - MLINK : calculates values of Z for ≠ values of θ
  - LODSCORE : calculates Z_{max} associated with θ_{max}

- Multi-point genetic linkage analysis programs
  - LINKMAP ou CMAP : Locate a disease locus in relation to a set of MKs *(established genetic map)*

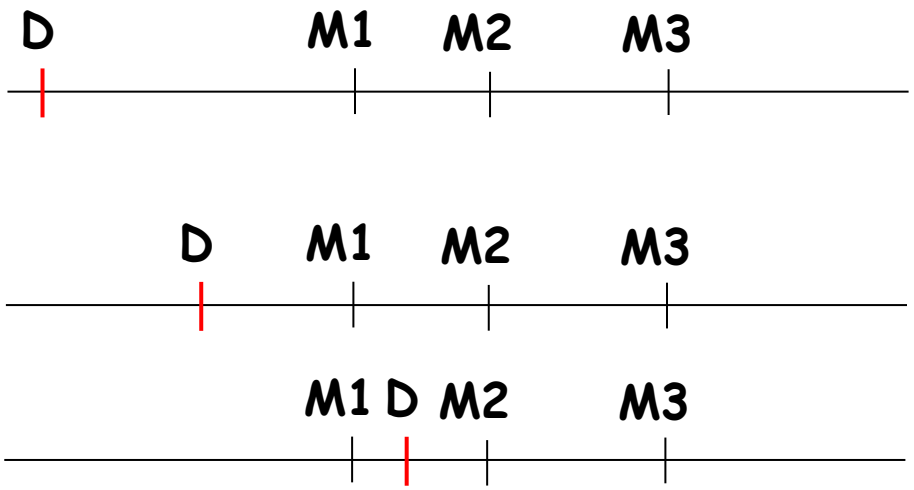
➢ *These programs can be used to establish genetic maps*

PRINCIPLE OF MULTIPOINT ANALYSIS

- **Idea :**

  From the genetic map established for a certain number of MK, we vary the position of the disease gene and we conclude with respect to the most likely location

- **Example :** use of a genetic map with 3 MK (M1, M2, M3)



Calculation of the likelihood L(I) considering D genetically independent

Calculation of the likelihood L(X1)

Calculation of the likelihood L(X2)

etc …

➢ Calculation of the multipoint lod-score at each position Xi :
log10 [L(Xi)/L(I)]

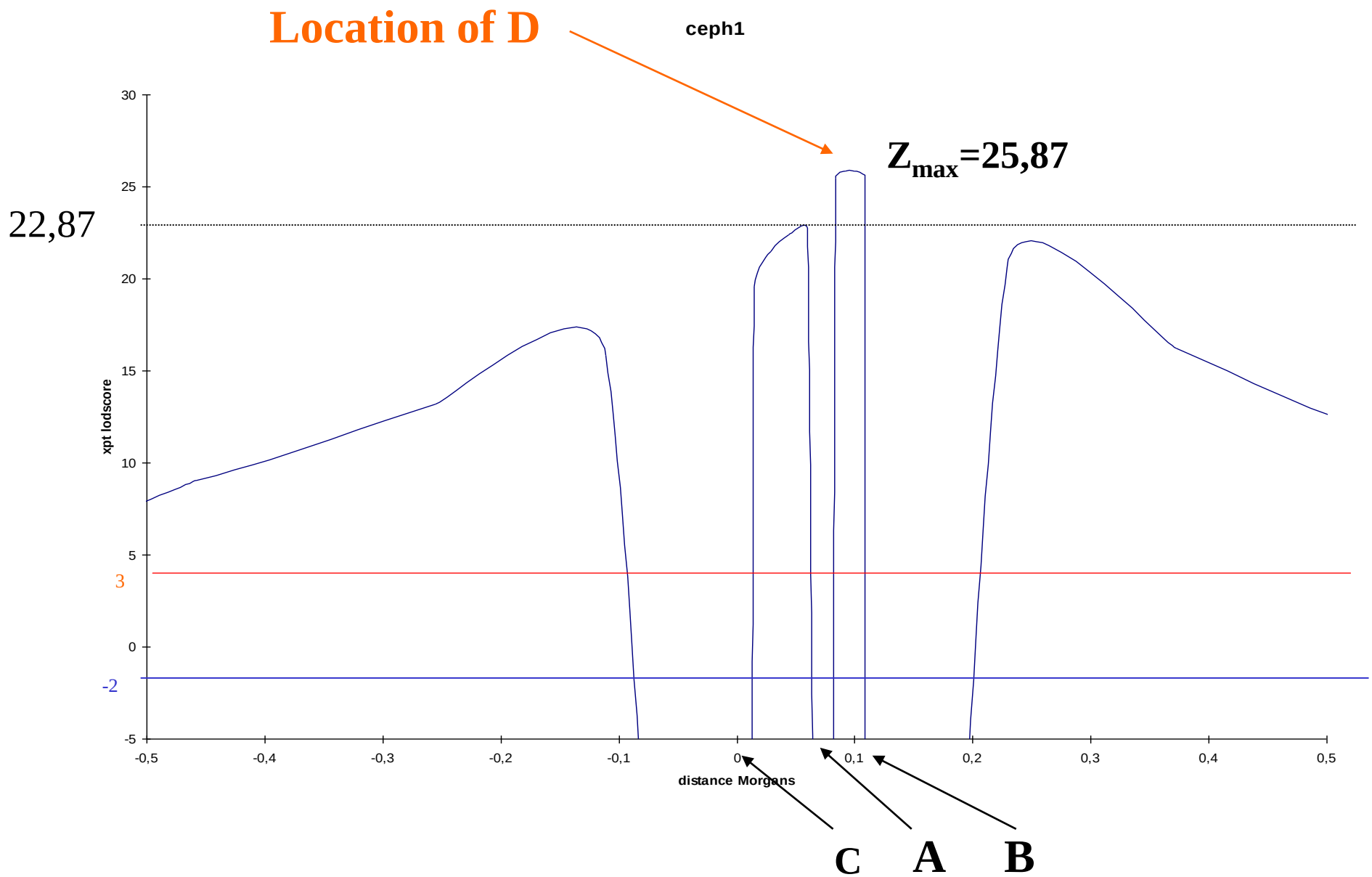
THRESHOLDS

- Genetic linkage
  - If for a given location the multipoint lod score is greater than 3 and included in the CI

    --> possible location of disease gene
  - Its probable location is the position for which the lod-score is maximum

- Linkage rejected
  - We can also reject regions for which the multipoint lod-score is less than or equal to -2

EXAMPLE OF MULTI-POINT CURVE



EXAMPLE OF USE OF LOD-SCORE TO LOCATE MENDELIAN ENTITIES

- **Breast cancer**
  - 1st gene localised in 17q21 *(Hall et al 1990)* → **BRCA1**
  - 2nd gene localised in 13q12 *(Wooster et al 1994)* → **BRCA2**
  - other genes…

- **Melanoma**
  - localised gene in 9p21*(Cannon-Albright et al 1992)* → **CDKN2A**
  - other genes : CDK4, 1p, others?

LINKAGE ANALYSES AND MULTIFACTORIAL DISEASES

- **Multifactorial diseases**

  Multifactorial human diseases most often result from complex interactions between genetic and environmental factors

- **Genetic model for multifactorial diseases**

  We cannot explain the transmission of the disease by a simple model (unless Mendelian entity (MG for Major Gene))

  - *the Lod-score method (model-dependent method) cannot be used to locate disease genes (except MG)*

  - *It is necessary to use model-independent methods*
    -> method of affected sib pairs (binary traits)

GENETIC LINKAGE ANALYSES

SUSCEPTIBILITY GENE LOCATION

-

AFFECTED SIB-PAIRS METHOD

AFFECTED SIB-PAIRS METHOD

- **Aim**

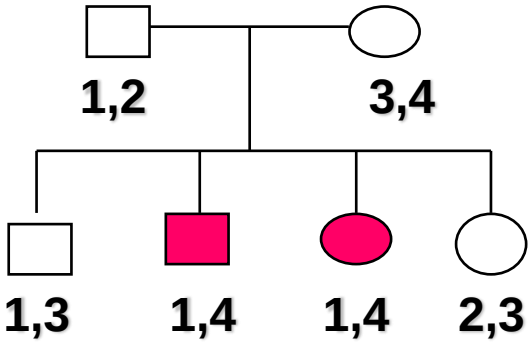
  Locate a gene involved in a disease without a priori knowledge of its genetic model

- **Principle**

  Related subjects (siblings) resembling each other phenotypically (affected) do they resemble each other for the genotypes of the marker?

  If yes ➔ genetic linkage

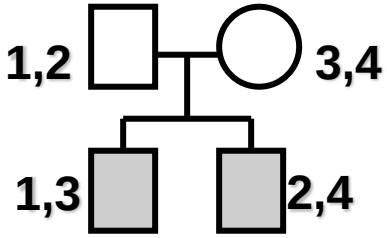
- **Sample**



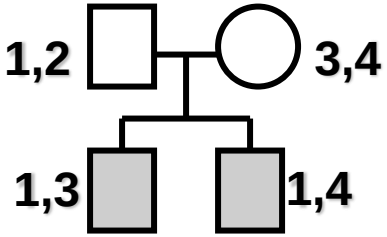
➲ Calculation of the IBD variable (IBD=Identity By Descent) for the 2 affected sibs

IBD: IDENTITY BY DESCENT

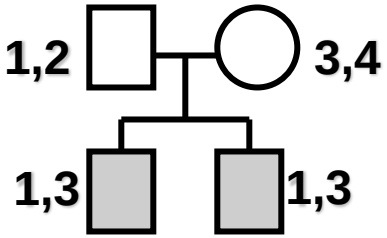
Two alleles present in two related individuals
are IBD (identical by descent)
if they both come from the same common ancestor



0 shared alleles
IBD=0

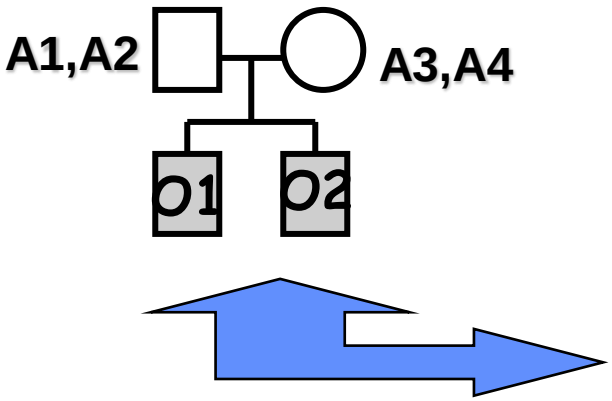


1 shared allele
IBD=1



2 shared alleles
IBD=2

EXPECTED PROPORTIONS OF IBD IF NO LINKAGE



The different possible genotypes for O1 and O2 are shown in the right table with the associated IBD status

G_{O2} \ G_{O1}	A1A3	A1A4	A2A3	A2A4
A1A3	2	1	1	0
A1A4	1	2	0	1
A2A3	1	0	2	1
A2A4	0	1	1	2

Note: whatever the genotypes of the parents, we always obtain the table above

IBD Status	0	1	2
Expected proportions	4/16=1/4	8/16=1/2	4/16=1/4

LINKAGE TEST BASED ON IBD STATUS OF AFFECTED SIBS (Penrose 1935)

	IBD		
	0	1	2
Expected counts (under H0)	N/4	N/2	N/4
Observed counts	n0	n1	n2

N : total number of sib pairs = n0+n1+n2
n0, n1, n2 : number of pairs sharing 0, 1 or 2 alleles

H0: no genetic linkage [P(IBD=0)=25%; P(IBD=1)=50%; P(IBD=2)=25%]

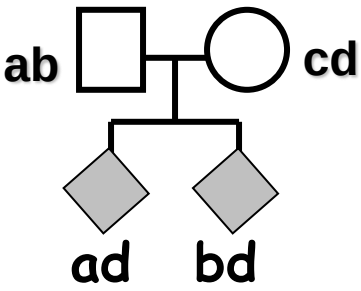
H1: genetic linkage [P(IBD=0)≠25%; P(IBD=1)≠50%; P(IBD=2)≠25%]

Test: $\chi^2 = \Sigma\ ((O-E)^2/E)$ **with df=2**

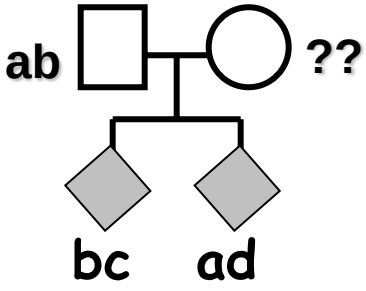
EXAMPLE OF CALCULATION OF THE LINKAGE TEST BASED ON THE IBD STATUS OF AFFECTED SIBS

Analysed sample : 100 affected sib-pairs / 1 marker with 4 alleles (a, b, c et d)

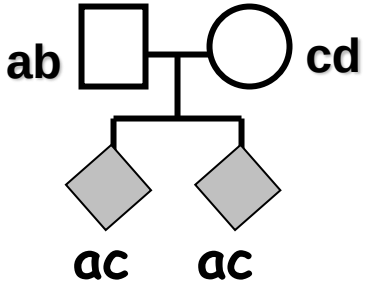
30 families



20 families



50 families



	IBD		
	0	1	2
Exp. counts	25	50	25
Obs. Counts	20	30	50

H0: no genetic linkage

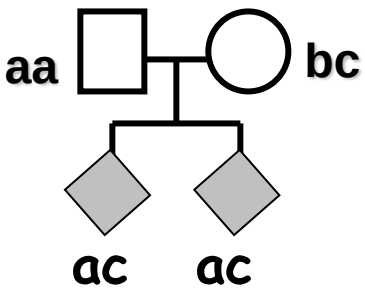
H1: genetic linkage

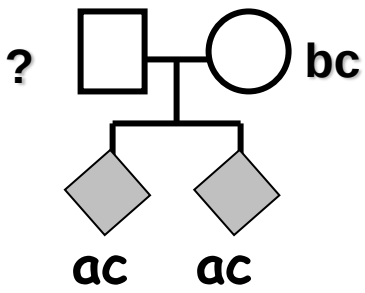
Test: χ^2 (2 df) = 34 > 5.99

➢ *H0 rejected --> genetic linkage*

UNCERTAINED IBD STATUS

Examples of families with uncertained IBD status :





IBD = 1 ou 2 ?

- ***Other methods if uncertained IBD status :***

  MLS (Maximum Likelihood Score) or NPL (Non Parametric Linkage)

ADVANTAGES-DISADVANTAGES
SIB-PAIRS METHOD VS LOD-SCORE METHOD

- **Advantages :**
  - conceptually simple
  - fast computing time
  - no specification of the genetic model
- **Disadvantages :**
  - less powerful methods
  - one cannot obtain an estimate of the recombination rate (therefore less precise)

WHAT DID I REMEMBER ?
(5)

1. Allez sur wooclap.com
2. Entrez le code d'événement dans le bandeau supérieur

Code d'événement
PKPEUY

1. Envoyez @PKPEUY au 06 44 60 96 62
2. Vous pouvez participer

Désactiver les réponses par SMS

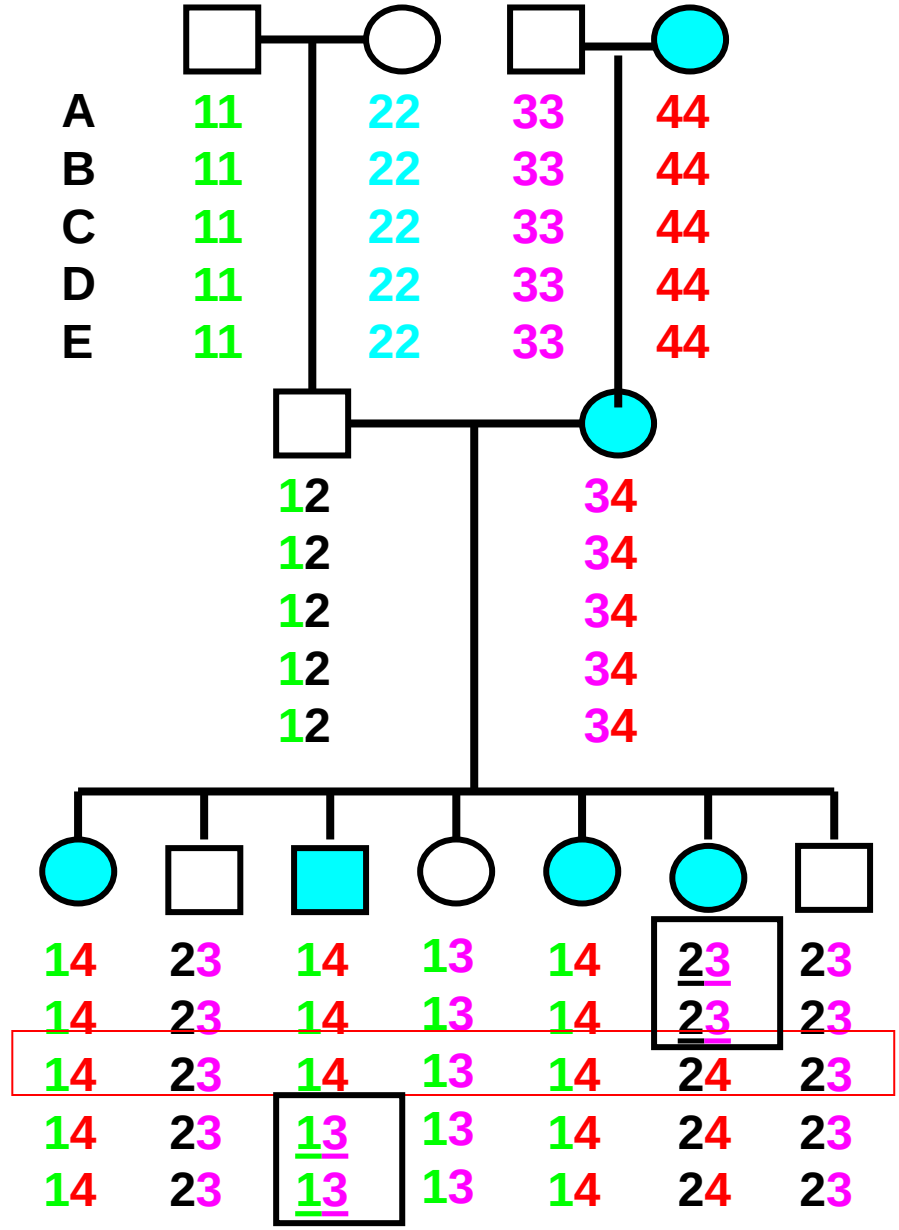
Use of haplotypes to localise disease gene

Autosomal dominant disease

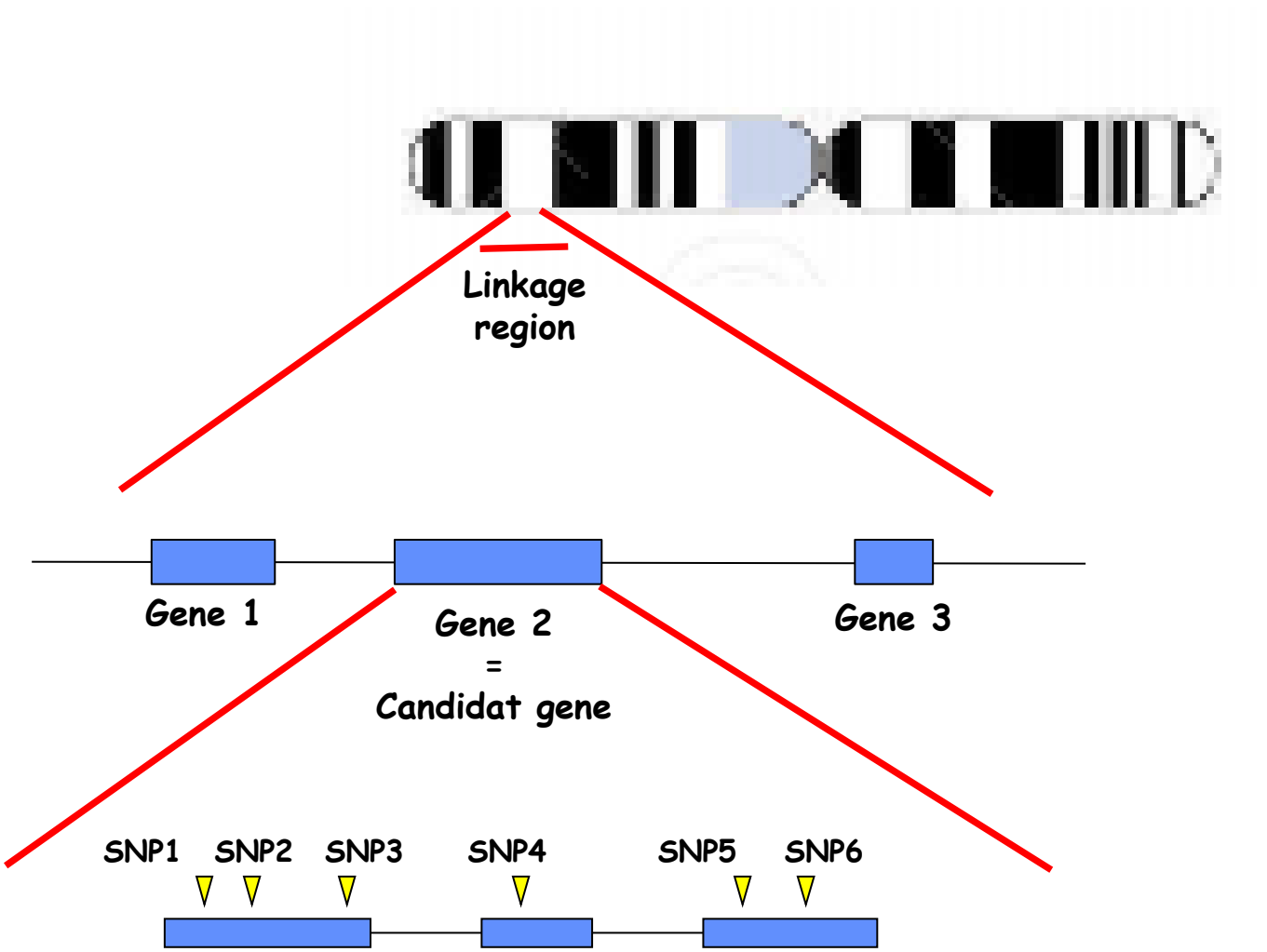
Search for haplotypes common to affecteds

III-3 & III-6 received recombinant gametes from their affected mother.

The disease gene is located between markers B and D.



AND AFTER …

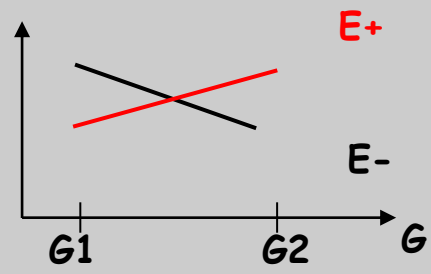


What is the causal variant?

PART 3

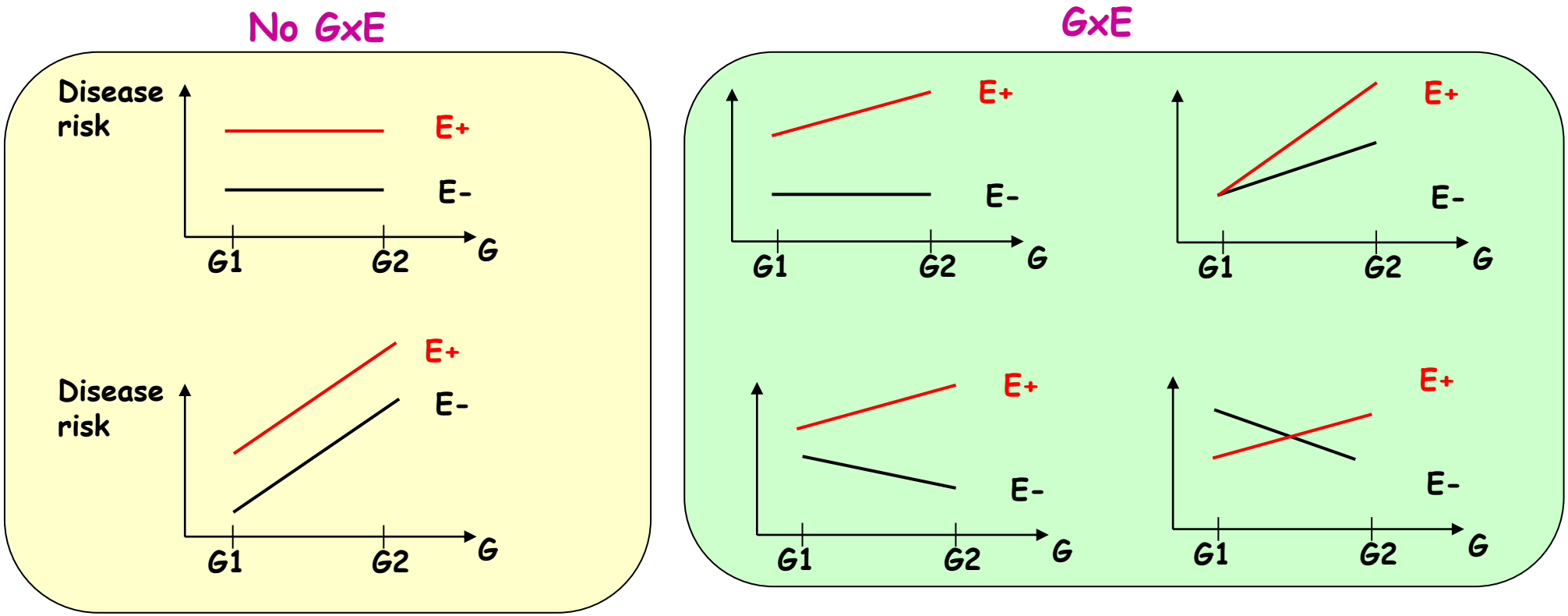
SEARCH FOR INTERACTIONS

Exemple : gene-environment interaction



INTERACTION GxE

G and E do not act independently on the development of the disease but interact (the response to E will depend on G)



INTERACTION TEST : Q-TEST

Stratum 1 : E+

	Cases	Controls	total
Allele 1	a1	b1	R11
Allele 2	c1	d1	R12
total	C11	C12	T1

Stratum 2 : E-

	Cases	Controls	total
Allele 1	a2	b2	R21
Allele 2	c2	d2	R22
total	C21	C22	T2

- OR in stratum i : $$\mathrm{OR}_i = \frac{a_i d_i}{b_i c_i}$$

- Homogeneity test for k strata (H0 : ORi= ; H1: ORi≠)

$$\chi^2(k-1) = \Sigma\, w_i * (\ln(OR_i) - Y)^2$$

For stratum i :

$w_i = 1/\mathrm{var}(\ln(OR_i))$

$Y = [\ \Sigma\, w_i * \ln(OR_i)\ ] / \Sigma w_i$

INTERACTION TEST : EXAMPLE

Stratum 1 : E+

	Cases	Controls	total
Allele 1	6	16	R11
Allele 2	48	322	R12
total	C11	C12	T1

$OR_1 = 2.52$
$Ln(OR_1) = 0.9225$
$w_1 = 3.95$

Stratum 2 : E-

	Cases	Controls	total
Allele 1	14	4	R21
Allele 2	25	56	R22
total	C21	C22	T2

$OR_2 = 7.84$
$Ln(OR_2) = 2.059$
$w_2 = 2.636$

- Homogeneity test for k=2 strata (H0: $OR_1=OR_2$; H1: $OR_1 \neq OR_2$)

$\chi^2(k-1) = \Sigma\, w_i*(\ln(OR_i)-Y)^2$

$\chi^2(1) = 2.04$

P-val = 0.15

For stratum i :

$w_i = 1/var(\ln(OR_i))$

$Y = [\ \Sigma\, w_i*\ln(OR_i)\ ] / \Sigma w_i$

INTERACTION TEST BY LOGISTIC REGRESSION (1)

$$P(\text{affected} / G,E) = P = \frac{\exp(\alpha + \beta_G G + \beta_E E + \beta_I GxE)}{(1 + \exp(\alpha + \beta_G G + \beta_E E + \beta_I GxE))}$$

$$\text{Logit } P = \log (P/1-P) = \alpha + \beta_G G + \beta_E E + \beta_I I$$

Coding of variables

$$G = \begin{cases} 1 \text{ if subjects } G+ \\ 0 \text{ if subjects } G- \end{cases} \qquad E = \begin{cases} 1 \text{ if subjects } E+ \\ 0 \text{ if subjects } E- \end{cases}$$

$$I = GxE = \begin{cases} 1 \text{ if } G=1 \ \& \ E=1 \\ 0 \text{ otherwise} \end{cases}$$

- *Other risk factors may be included in the logistics model*
- *To test a Gene-Gene interaction, I=GxG*

INTERACTION TEST BY LOGISTIC REGRESSION (2)

$$\text{Logit } P = \alpha + \beta_G G + \beta_E E + \beta_I I$$

- GxE interaction tests (H0 : $\beta_I=0$; H1 : $\beta_I \neq$)

  - Wald Test: $(\beta_I - 0)^2 \ / \ Var(\beta_I) \sim \chi^2(1 \text{ df})$

  - Likelihood Ratio Test : $-2\ln[\ L(\beta_I=0)/L(\beta_I)] \sim \chi^2(1 \text{ df})$

  ➢ *Sometimes, it is more powerful to test the effects of G and I together*

- Effect of G according to exposure to E:

  - Effect of G for E+ subjects : $OR_{GE} = \exp(\beta_G+\beta_I)$

  - Effect of G for E- subjects : $OR_G = \exp(\beta_G)$

  ➢ *If no interaction ($\beta_I \approx 0$), effect of G similar for E+ and E- subjects*

WHAT DID I REMEMBER ?
(6)

1. Allez sur wooclap.com
2. Entrez le code d'événement dans le bandeau supérieur

Code d'événement
GRMSHI

1. Envoyez @GRMSHI au 06 44 60 96 62
2. Vous pouvez participer

Désactiver les réponses par SMS