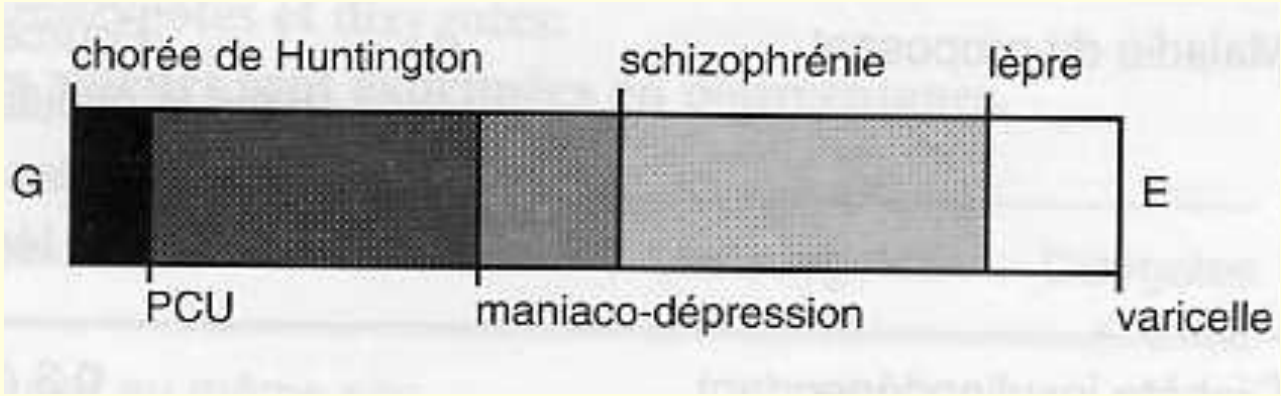
# Common disorders (multifactorial)

Individual variations (polymorphisms) in some genes can predispose (or protect) for a disease, in interaction with environmental factors (food, smoking, infection, stress…)

| | | |
|---|---|---|
| Metabolic diseases | type 1 diabetes | T1DM |
| | type 2 diabetes | T2DM |
| Neurodegenerative dis. | Multiple sclerosis | MS |
| | Parkinson disease | PD |
| | Alzheimer disease | AD |
| Psychiatric diseases | Schizophrenia | |
| Vascular disease | Arterial hypertension | |
| Joint diseases | Rheumatoïd arthritis | RA |
| Infectious diseases ! | Predisposition to leprosy | |

Environmental and inherited factors

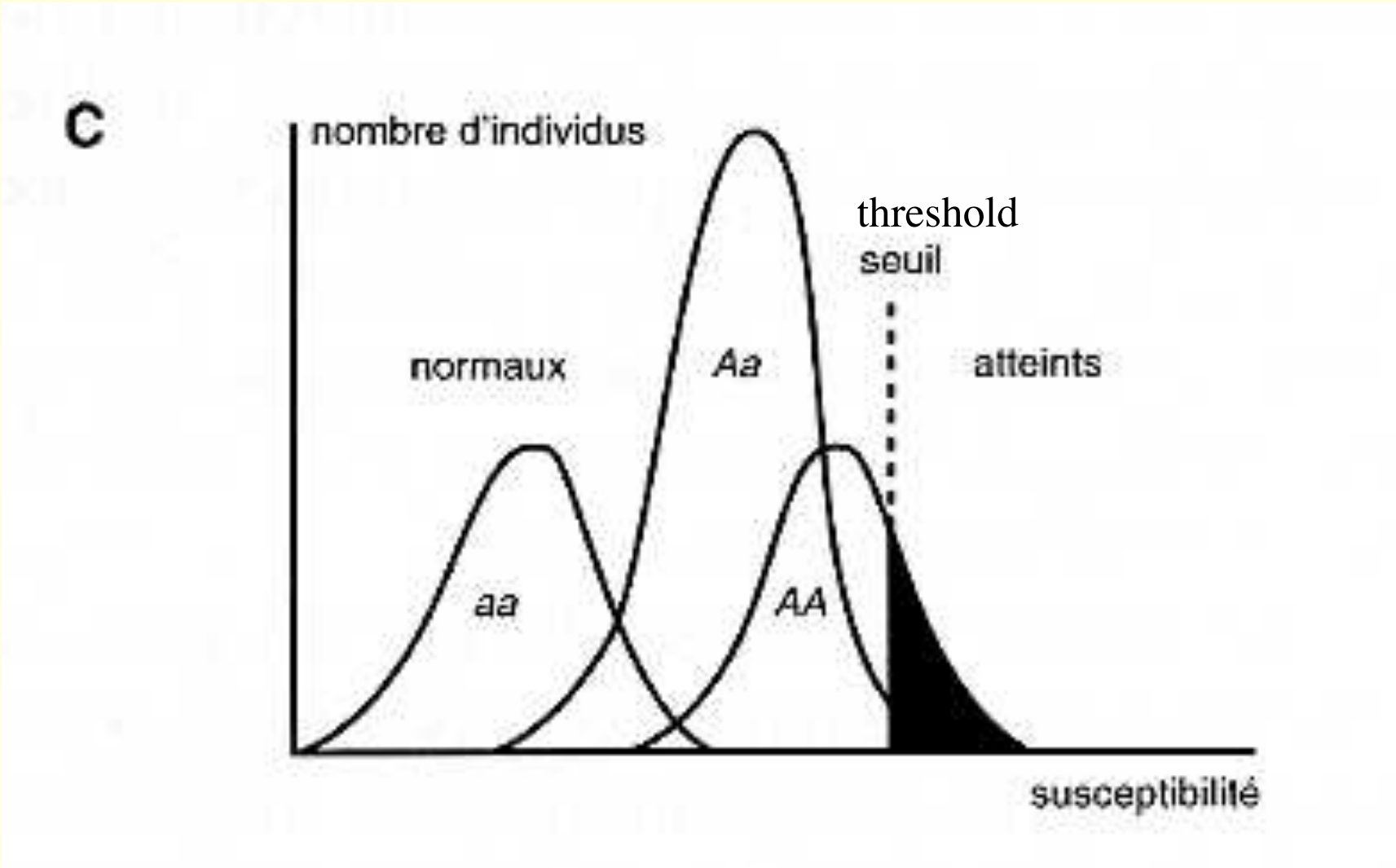
Genetics




Environment

**Principes de génétique humaine**
Josué Feingold, Marc Fellous, Michel Solignac
Hermann éditeurs

# Risk variation acording to genotype




Feingold et al

# Familial recurrence in some common disorders

| Maladie du proposant | Pourcentage dans la population générale (ou chez les apparentés des témoins) | Pourcentage chez les apparentés du propositus au premier degré | Risque relatif |
|---|---|---|---|
| Diabète insulinodépendant | 0,3 à 0,4 | 3 à 8 | 10 à 20 |
| Diabète non insulinodépendant | 2 à 4 | 10 à 15 | 4 |
| Asthme | 3,8 | 9,2 | 2,5 |
| Sclérose en plaques (MS) | 0,02 à 0,06 | 0,5 à 1 | 19 |
| Fente labiale (bec-de-lièvre) | 0,1 | 2 à 5 | 35 |
| Schizophrénie | 1 | 4 à 10 | 7 |
| Hypertension artérielle | 10 | 20 à 30 | 3 |

Relative Risk (RR) = Odds Ratio (OD) = $\lambda s$

Feingold et al

How to identify genes and alleles causing this increase of Relative Risk?

## 1. Family studies (linkage analysis)

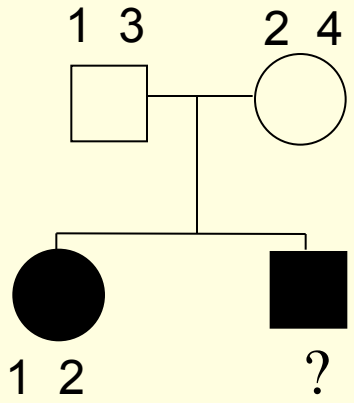
- Large« monogenic » families : parametric methods
- Small families : non-parametric methods :

  Sib-pair studies

## 2. Association studies

- Case-control studies :
  comparison of allelic frequencies or genotype frequencies (aa/aA/AA) (testing for recessive, dominant or additive effect)
- In the general population or in « isolates » (Iceland, Finland, Ashkenazy Jews, remote villages [mountains]…) with possible founder effects

# 1. Sib-pair method




*a priori* likelihood (for informative parents = IBD)

| | |
|---|---|
| 2 shared alleles (1 2) | 0.25 |
| 1 shared allele (1 4 ou 2 3) | 0.50 |
| 0 shared allele (3 4) | 0.25 |

IBD: Identity by descent;

IBS: Identity by state (if parents not genotyped or non informative)

# Sib-pair method

| | Allèles partagés | | | Excès de concordance | $\chi^2_2$ ; p | |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | | 100 paires | 200 paires |
| Expected | 25 % | 50 % | 25 % | | | |
| Observé | 20 % | 50 % | 30 % | 10 % | 1,0 ; p = 0,60 | 2,0 ;p = 0,36 |
| | 15 % | 55 % | 30 % | 15 % | 3,2 ; p = 0,20 | 6,4 ; p = 0,04 |
| | 10 % | 60 % | 30 % | 20 % | 7,8 ; p = 0,02 | 15,6 ; p = 0,0004 |

$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

Determine the p value from $\chi^2$ distribution

Feingold et al

# Susceptibility locus to leprosy

« Genome Wide » study Mira et al. Nature Genetics, 2003

86 families 205 patients (90 paucibacillary, 115 multibacillary)

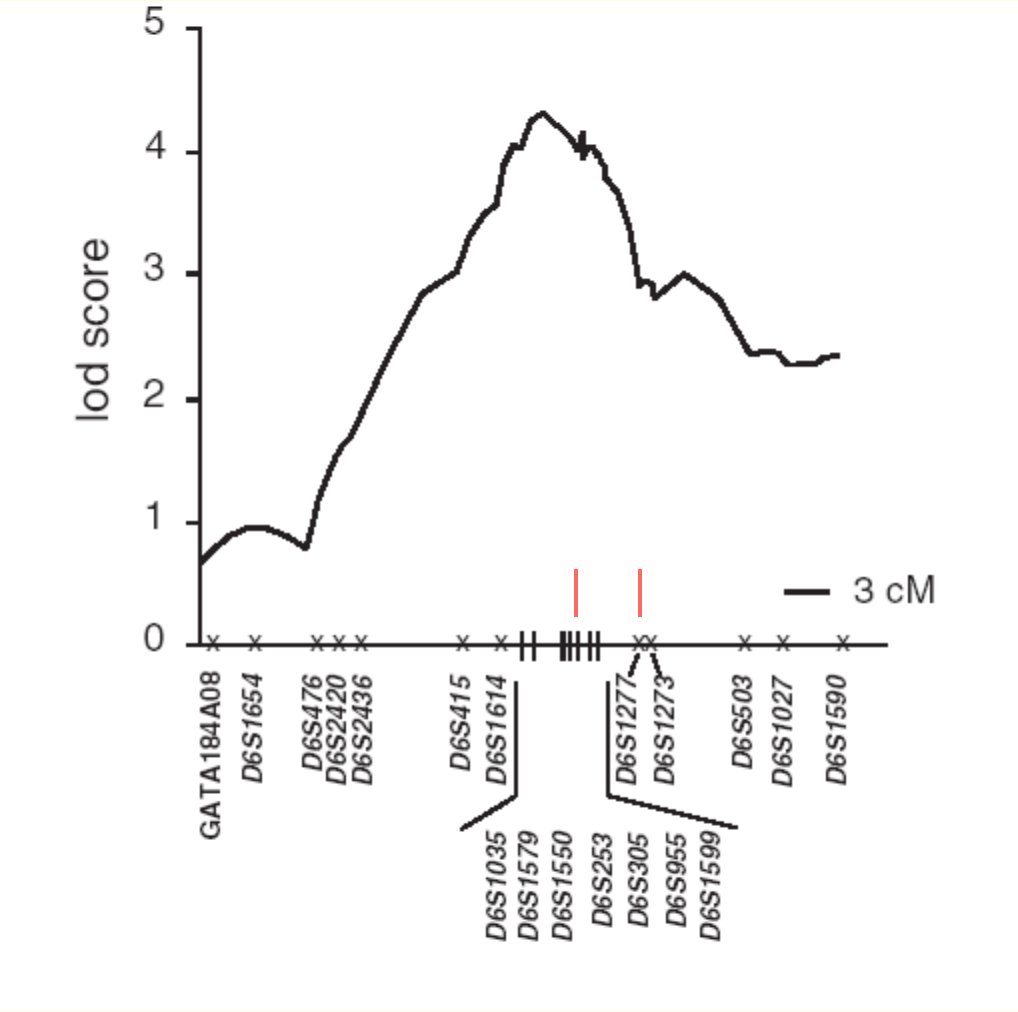
**Table 2 • Chromosome regions selected for linkage fine mapping**

| Cytogenetic location | Microsatellite name | Map position | Two-point lod score | Two-point Info content | Multipoint lod score | Multipoint Info content |
|---|---|---|---|---|---|---|
| 1p32 | D1S200 | 82.41 | 1.20 | 0.69 | 1.18 | 0.86 |
| 1q42.3 | D1S235 | 254.64 | 2.62 | 0.31 | 2.13 | 0.61 |
| 2q36.3 | D2S1363 | 227 | 1.01 | 0.55 | 1.00 | 0.76 |
| 3p21 | D3S2409 | 70.61 | 1.31 | 0.37 | 1.55 | 0.64 |
| | D3S1766 | 78.64 | 0.63 | 0.55 | 1.67 | 0.71 |
| 5p14 | D5S807 | 19.02 | 1.04 | 0.65 | 1.35 | 0.86 |
| | D5S817 | 22.88 | 0.70 | 0.49 | 1.44 | 0.79 |
| 5q11.2 | D5S2500 | 69.23 | 1.91 | 0.70 | 1.27 | 0.83 |
| 6p21 | D6S1051 | 51 | 2.43 | 0.66 | 2.39 | 0.77 |
| | D6S1017 | 63.28 | 0.61 | 0.57 | 1.25 | 0.77 |
| 6q25–q27 | D6S305 | 166 | 3.27 | 0.75 | 3.88 | 0.84 |
| | D6S1277 | 173.31 | 2.82 | 0.66 | 2.66 | 0.85 |
| | D6S503 | 184.51 | 1.41 | 0.57 | 2.33 | 0.86 |
| | D6S1027 | 187.23 | 1.32 | 0.65 | 2.23 | 0.84 |
| 7p15 | D7S1802 | 33.09 | 1.66 | 0.60 | 1.15 | 0.81 |
| 10p13 | D10S674 | 41.79 | 0.21 | 0.64 | 0.08 | 0.77 |
| | D10S1423 | 46.23 | 0.09 | 0.68 | 0.22 | 0.86 |
| 13q22.1 | D13S788 | 45.55 | 0.21 | 0.56 | 1.68 | 0.68 |
| | D13S800 | 55.31 | 1.19 | 0.70 | 1.18 | 0.85 |
| 20p12.3 | D20S473 | 9.53 | 0.67 | 0.44 | 1.13 | 0.59 |

Linkage of microsatellites on 11 cytogenetic locations with multipoint MLB lod score >1 and of markers on chromosome 10p13 with leprosy *per se*. Map positions are according to the Marshfield map.

# Susceptibility locus in leprosy

20 microsatellites in the 6q25-27 region




Mira et al. Nature Genetics, 2003

# 2. Association studies

- Case-control studies
- Comparison of allelic frequencies or genotype frequencies (aa/aA/AA) (testing for recessive, dominant or additive effect)
- Hypothesis of a founder effect

## Relative risk and attributable risk :

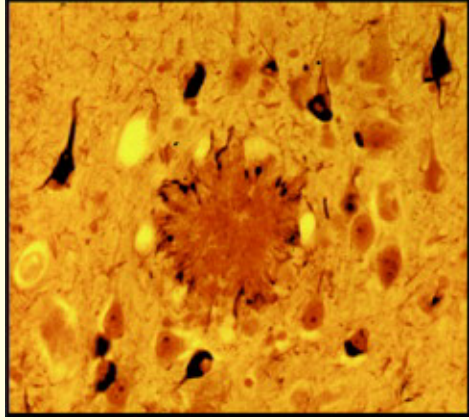
- Risk = p(m)
- Relative risk RR = p(m)f+/p(m)f- (f+; f-: present or absent risk factor)
- Attributable risk defines, at the population level, the proportion of cases imputable to the risk factor AR= f(RR-1)/1+ f(RR-1)  f=frequency of carriers

# Association studies

- Candidate at-risk allele
- Candidate gene(s)
- Candidate region
- Whole Genome Association Studies (GWAS) : possible since 2005…

- Candidate allele :

  (the allele is directly predisposing to the disease)



Amyloïde plaques

Alzheimer disease (multifactorial form)

Predisposing gene : ApoE

Allèles :

| | | | |
|---|---|---|---|
| Apo ε4 | arg | arg | 17 % |
| Apo ε3 | cys | arg | 73 % |
| Apo ε2 | cys | cys | 10 % |

Genotype frequencies

| | | | |
|---|---|---|---|
| ε4 homozygote (aa) | : | 0,17 x 0,17 | ≈ 0,03 (3 %) |
| ε4 heterozygote (aA) | : | 2 x 0,17 x 0,83 | ≈ 0,28 (28 %) |
| ε2 or ε3 (no ε4) (AA) | : | 0,83 x 0,83 | ≈ 0,69 (69 %) |

# Relative risk of ε4 allele for Alzheimer disease

| genotype | patients % | controls % | Relative Risk RR | Attributable Risk AR (%) |
|---|---|---|---|---|
| ε4 ε4 or ε3 ε4 | 58 | 31 | x 3.1 | 39 |
| ε4 ε4 | 14 | 3 | x 7.7 | |
| ε3 ε4 | 44 | 28 | x 2.6 | |
| ε2 or ε3 (no ε4) | 42 | 69 | | |

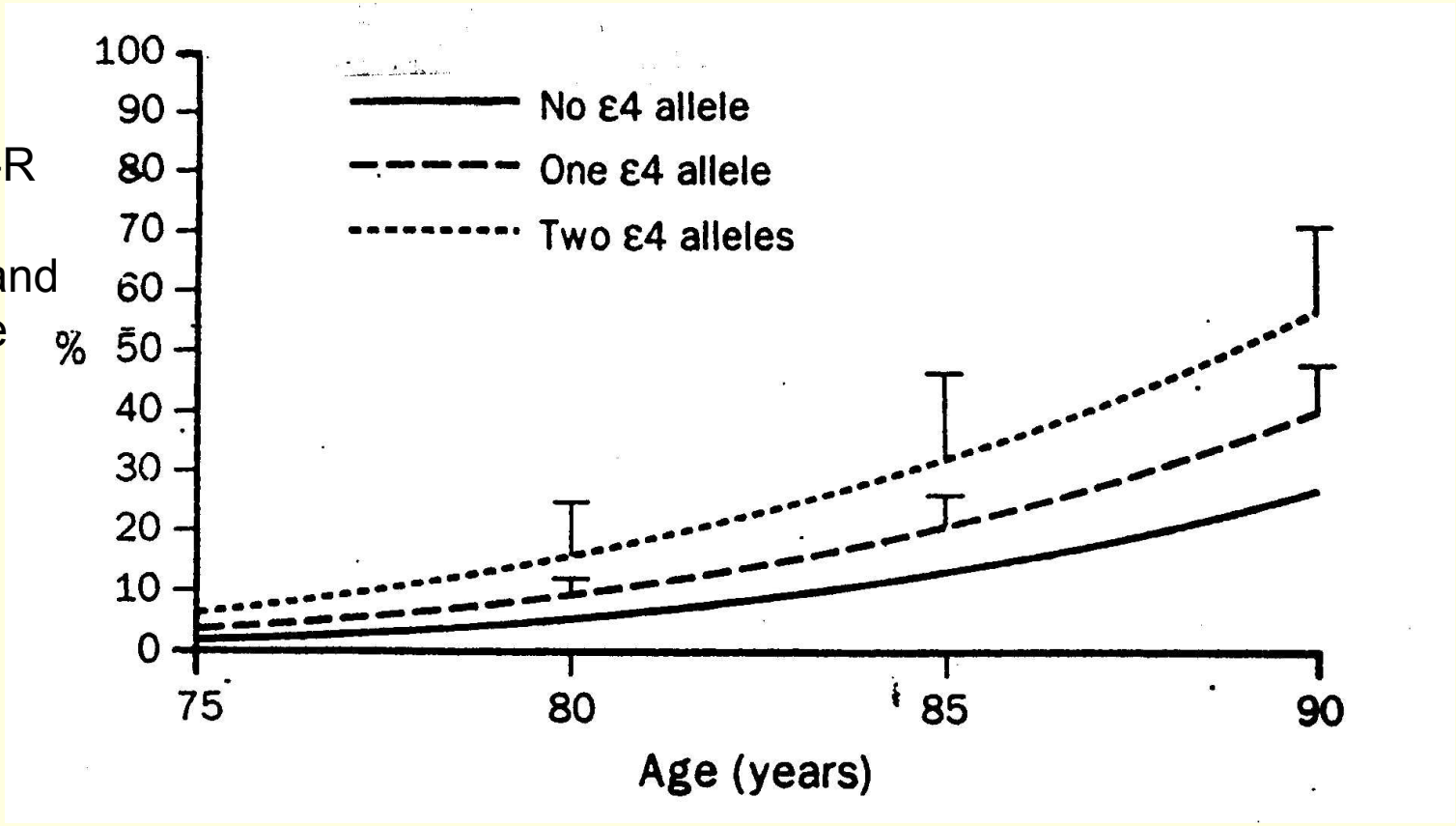
Incidence of the disease at 85 years ≈ 10%

$$RR_{homoz} = \frac{\frac{0.1 \times 0.14}{0.03}}{\frac{0.1 \times 0.42}{0.69}} = 7.7 \qquad RR_{heteroz} = \frac{\frac{0.1 \times 0.44}{0.28}}{\frac{0.1 \times 0.42}{0.69}} = 2.6$$

Relative risk of ε4 allele for Alzheimer disease:

| genotypes | (freq.) | RR | risk at 85 years |
|---|---|---|---|
| ε4 ε4 | (3 %) | 7.7 x | 46 % |
| ε3 ε4 | (28 %) | 2.6 x | 16 % |
| ε2 ou ε3 | (69 %) | 1 x | 6 % |

Estimation of the prevalence of dementia DSM-III-R (Alzheimer) according to age and ApoE e4 genotype

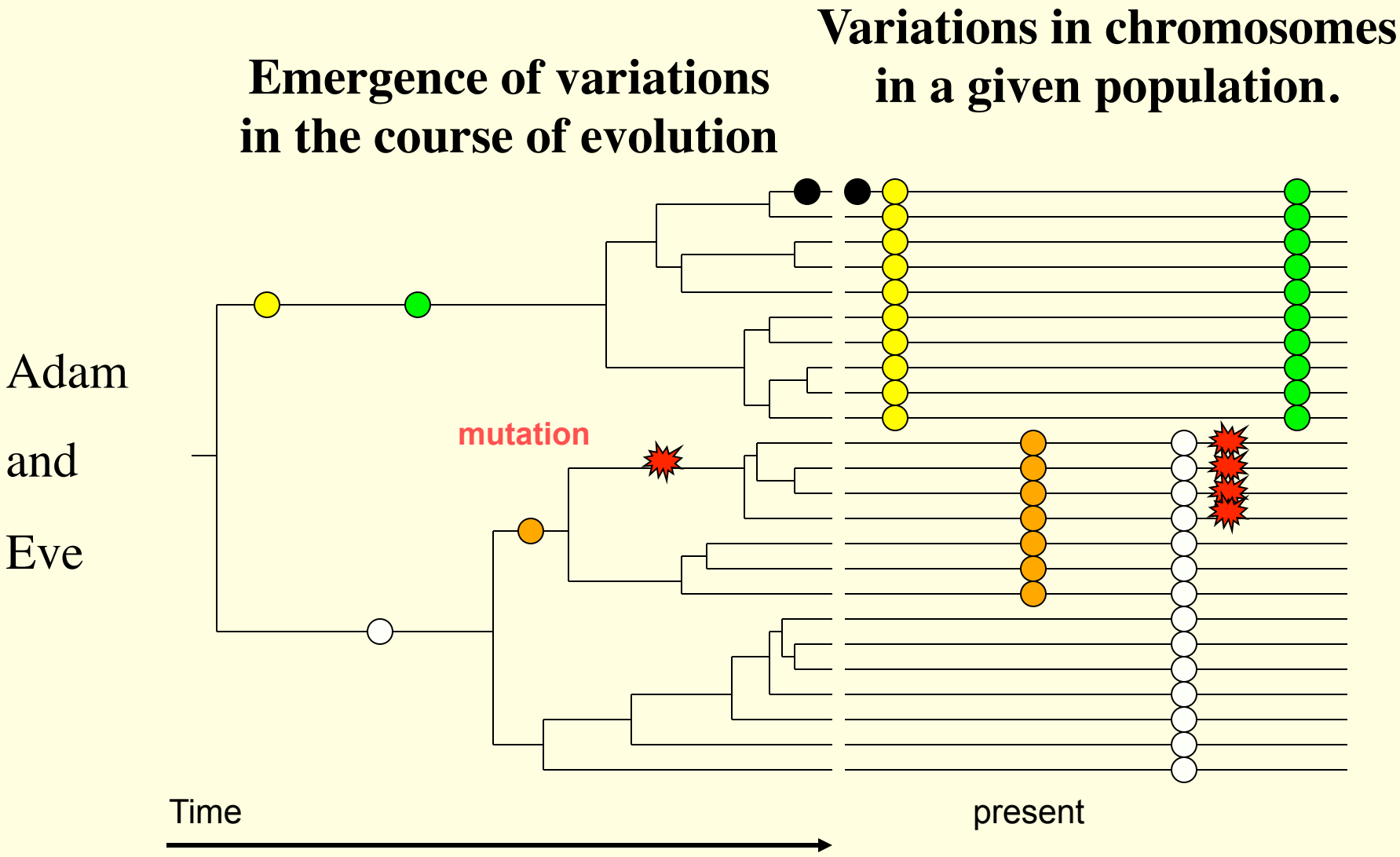


- Candidate gene but no candidate allele

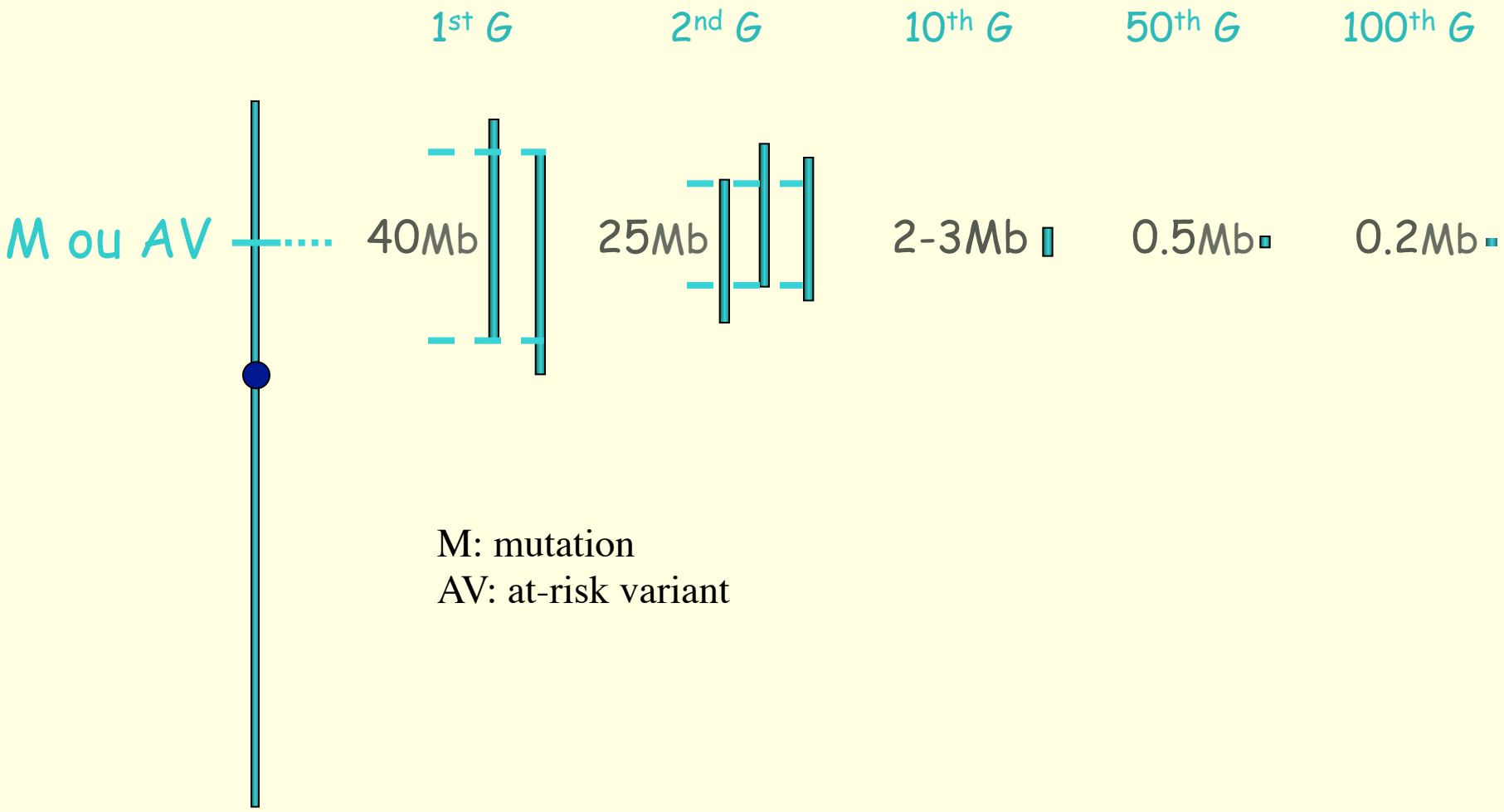
  Predisposing allele is in linkage disequilibrium with flanking polymorphic markers

  Requires the sequencing of the entire at-risk haplotype, possibly also in different populations

# Linkage disequilibrium and haplotype



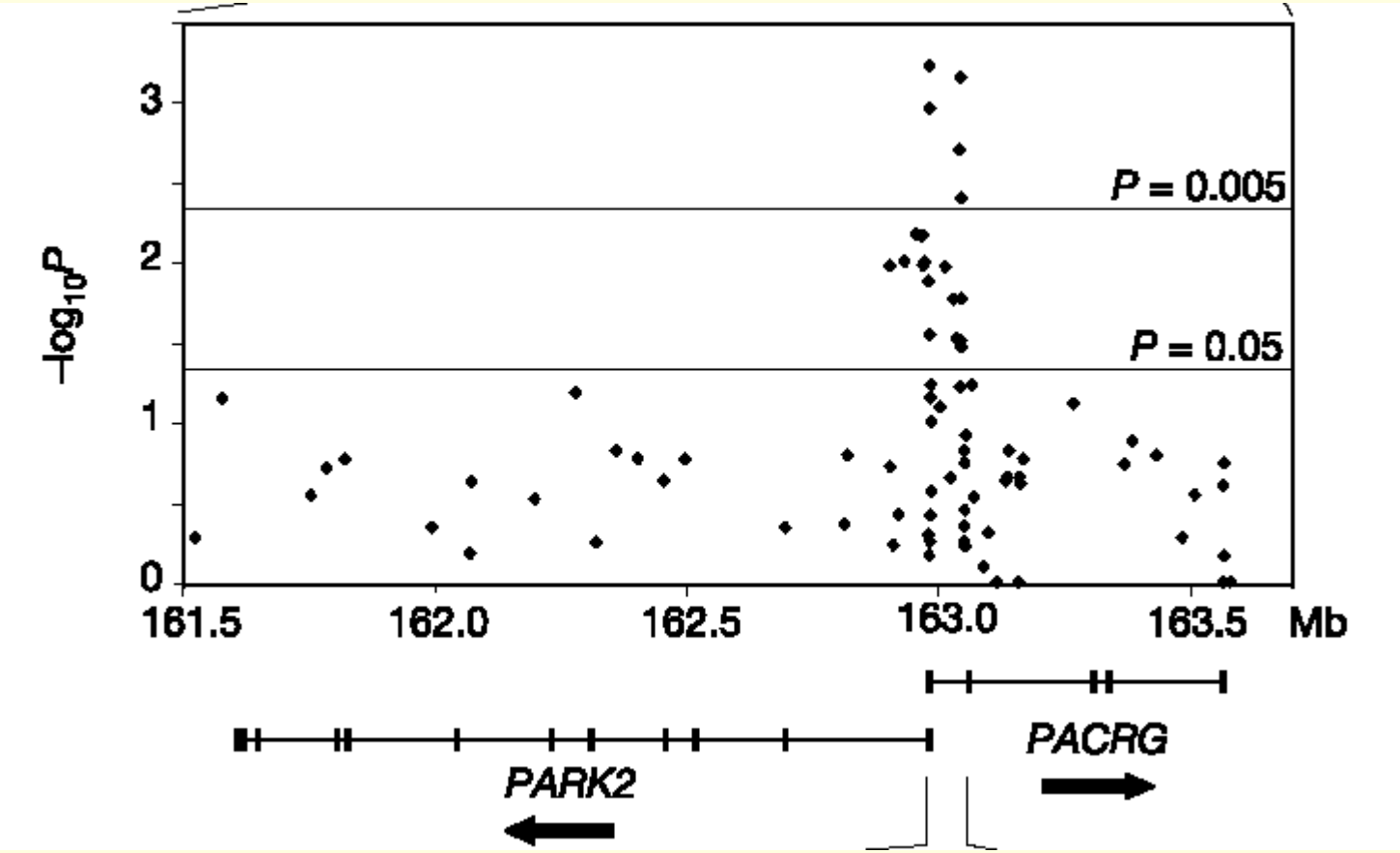
# Evolution of the length of the shared region in patients, around a mutation or a risk variant




M: mutation
AV: at-risk variant

- Candidate region (based on linkage)

Association between leprosy
and 81 SNPs in the candidate region (2 Mb)




Mira et al, Nature, 2004

Recessive protective C-T haplotype and not dominant at-risk haplotypes :

Table 1 **Association with leprosy of genotypes defined from the three PARK2_e01(−2599)–rs1040079 haplotypes**

| PARK2(−2599)–rs1040079 genotypes | | Frequency | Odds ratio | 95% confidence interval | *P* |
|---|---|---|---|---|---|
| C–T | C–T | 0.11 | 1.00 | | |
| T–T | C–T | 0.32 | 3.15 | 1.29–7.71 | 0.012 |
| T–T | T–T | 0.26 | 3.68 | 1.41–9.62 | 0.008 |
| T–C | C–T | 0.10 | 5.35 | 1.92–14.89 | 0.001 |
| T–C | T–T | 0.17 | 5.72 | 2.10–15.59 | 0.0007 |
| T–C | T–C | 0.04 | 4.75 | 1.32–17.18 | 0.017 |
| Accounting for dominance effect | | | | | |
| C–T | C–T | 0.11 | 1.00 | | |
| T–T | C–T/T–T | 0.58 | 3.23 | 1.34–7.82 | 0.009 |
| T–C | T–C/T–T/C–T | 0.31 | 5.28 | 2.06–13.55 | 0.0005 |

Frequencies have been estimated from non-transmitted parental haplotypes. Odds ratios were computed by means of conditional logistic regression. The overall comparison showed a significant association with leprosy, accounting ($P = 0.0002$) or not ($P = 0.006$) for dominance effect. Likelihood ratio tests showed that the genotypic model including three risk levels denoted as 'accounting for dominance effect' (that is, T–T haplotype is dominant over C–T, and T–C haplotype is dominant over both T–T and C–T) was not rejected against the full genotypic model with six risk levels ($\chi^2 = 0.54$, 3 degrees of freedom, not significant), whereas the genotypic model with only two risk levels (T–T and T–C haplotypes have the same dominant effect over C–T) is rejected against the genotypic model with three risk levels ($\chi^2 = 4.21$, 1 degree of freedom, $P = 0.04$). This indicates that the best-fitting genotypic model was the one denoted in the table as 'accounting for dominance effect'.

Mira et al, Nature, 2004

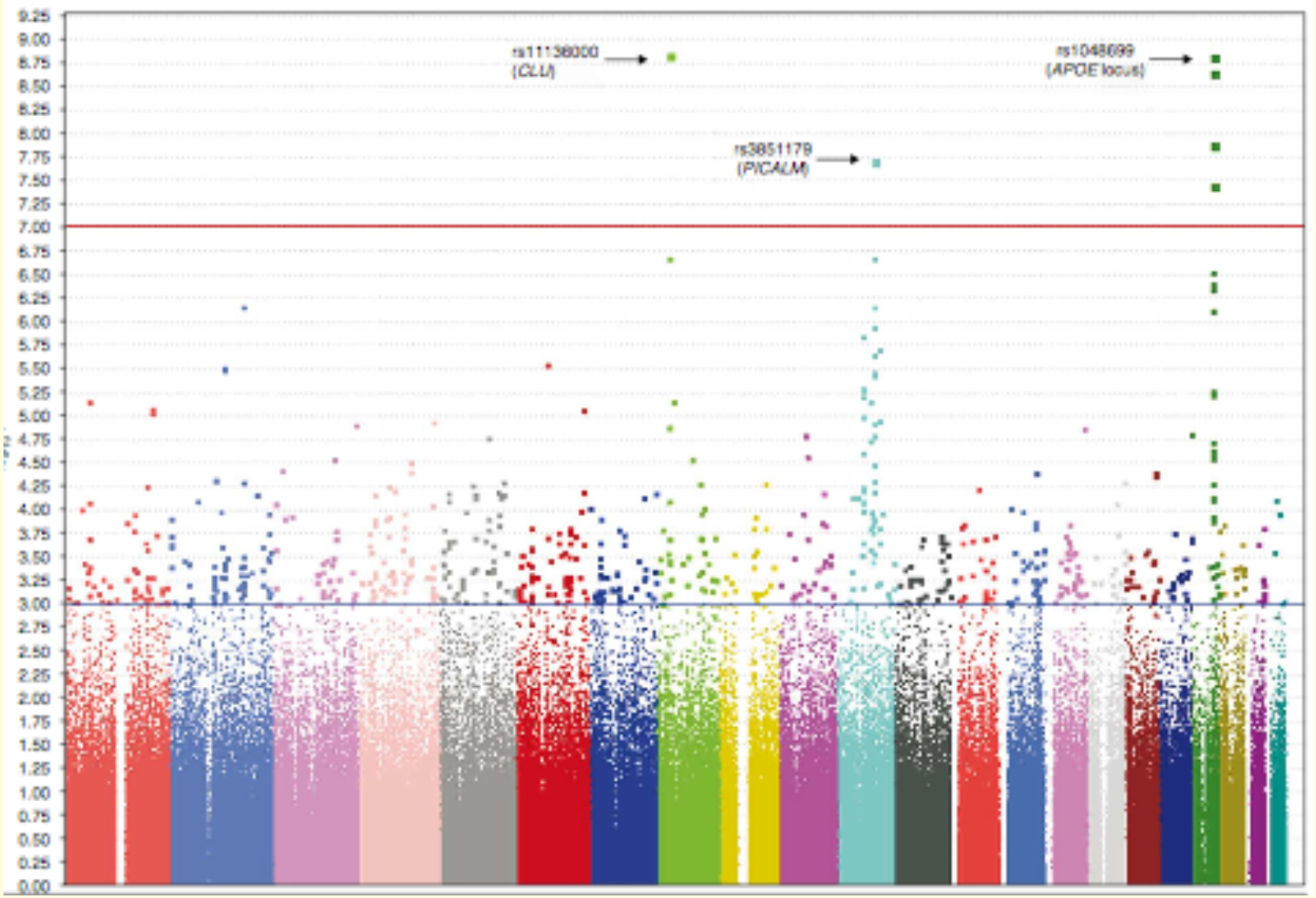
Is there additional genes and alleles causing increased relative risk for Alzheimer disease?

# Whole genome association study (GWAS)

Polymorphisms in linkage disequilibrium with a nearby allele directly causing the increased susceptibility

# GWAS 2009

Alzheimer Disease, Harold et al Nat. Genet. 2009



# GWAS 2009

Alzheimer disease, Harold et al Nat. Genet. 2009

SNPs showing genome-wide significant association with AD in stage 1 of GWAS

| SNP | Chr | Closest RefSeq Gene | Location Relative to Gene | MAF | GWAS: 3941 cases 7848 controls P-value (two-tailed) | GWAS OR (95% CI) | Extension: 2023 cases 2340 controls P-value (one-tailed) | Extension OR (95% CI) | Combined: 5964 cases 10188 controls P-value (two-tailed) | Combined OR (95% CI) |
|---|---|---|---|---|---|---|---|---|---|---|
| rs2075650 * | 19 | *TOMM40* | Intron | 0.15 | $1.8\times10^{-157}$ | 2.53 (2.37-2.71) | | | | |
| rs157580 | 19 | *TOMM40* | Intron | 0.39 | $9.6\times10^{-54}$ | 0.63 (0.59-0.66) | | | | |
| rs6859 | 19 | *PVRL2* | 3′ UTR | 0.43 | $6.9\times10^{-41}$ | 1.46 (1.38-1.54) | | | | |
| rs8106922 | 19 | *TOMM40* | Intron | 0.40 | $5.4\times10^{-39}$ | 0.68 (0.64-0.72) | | | | |
| rs405509 | 19 | *APOE* | 5′ | 0.52 | $4.9\times10^{-37}$ | 0.70 (0.66-0.74) | | | | |
| **rs11136000** | **8** | ***CLU*** | **Intron** | **0.40** | $\mathbf{1.4\times10^{-9}}$ | **0.84 (0.79-0.89)** | **0.017** | **0.91 (0.83-0.99)** | $\mathbf{8.5\times10^{-10}}$ | **0.86 (0.82-0.90)** |
| **rs3851179** | **11** | ***PICALM*** | **5′** | **0.37** | $\mathbf{1.9\times10^{-8}}$ | **0.85 (0.80-0.90)** | **0.014** | **0.90 (0.82-0.99)** | $\mathbf{1.3–10^{-9}}$ | **0.86 (0.82-0.90)** |

# Association studies: interpretation of statistically significant results)

- **False positive** (type 1 error) (significance of the $p \leq 0{,}01\ldots$), major problem of multiple testing (needs Bonferroni correction or other tests) and publication bias of positive results : requires replication studies

# Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease

Kirk E. Lohmueller[1,2], Celeste L. Pearce[3], Malcolm Pike[3], Eric S. Lander[1,4] & Joel N. Hirschhorn[1,5,6]

**Table 2 • Pooled odds ratios for follow-up studies of 25 associations**

| Associated gene, phenotype | Number of studies | Number of studies removed to achieve homogeneity | Fixed effects OR (95% c.i.) |
|---|---|---|---|
| Associations for which meta-analysis replicates original report | | | |
| *ABCC8*, exon 22, type 2 diabetes[a] | 4 | 1 | 2.28 (1.27–4.10) |
| *COL1A1*, osteoporotic fracture[a] | 12 | 0 | 1.59 (1.36–1.86) |
| *CTLA4*, type 1 diabetes[a] | 20 | 6 | 1.27 (1.17–1.37) |
| *DRD3*, schizophrenia[a] | 48 | 2 | 1.12 (1.02–1.23) |
| *GSTM1*, head/neck cancer[a] | 25 | 1 | 1.20 (1.09–1.33) |
| *HTR2A*, schizophrenia[a] | 28 | 0 | 1.07 (1.01–1.14) |
| *PPARG*, type 2 diabetes[a] | 14 | 1 | 1.22 (1.08–1.37) |
| *SLC2A1*, type 2 diabetes[a] | 3 | 1 | 1.76 (1.35–2.31) |
| Associations for which meta-analysis does not provide replication for original report | | | |
| *ABCC8*, intron 24, type 2 diabetes | 9 | 2 | 1.02 (0.92–1.14) |
| *ADD1*, hypertension | 18 | 5 | 1.09 (0.98–1.20) |
| *APOE*, schizophrenia | 12 | 0 | 1.0 (0.85–1.18) |
| *BLMH*, Alzheimer disease | 5 | 0 | 0.92 (0.67–1.26) |
| *COMT*, bipolar disorder | 12 | 0 | 1.08 (0.95–1.23) |
| *COMT*, schizophrenia | 9 | 0 | 0.99 (0.87–1.12) |
| *DRD2*, schizophrenia | 8 | 1 | 0.84 (0.70–1.01) |
| *SLC2A2*, type 2 diabetes | 3 | 0 | 0.96 (0.63–1.45) |
| *GSTM1*, breast cancer | 15 | 0 | 1.07 (0.98–1.18) |
| *GYS1*, type 2 diabetes[b] | 3 | 0 | 0.56 (0.35–0.90) |
| *INSR* RFLP, type 2 diabetes | 4 | 0 | 1.32 (0.90–1.94) |
| *INSR* Met985, type 2 diabetes | 4 | 0 | 0.89 (0.49–1.64) |
| *KCNJ11*, type 2 diabetes | 6 | 0 | 1.09 (0.95–1.24) |
| *NTF3*, schizophrenia | 7 | 0 | 1.01 (0.80–1.29) |
| *PON1*, CAD | 14 | 3 | 1.06 (0.98–1.14) |
| *SERPINE1*, MI | 13 | 0 | 1.04 (0.95–1.15) |
| *TPH*, bipolar disorder | 5 | 0 | 1.12 (0.98–1.28) |

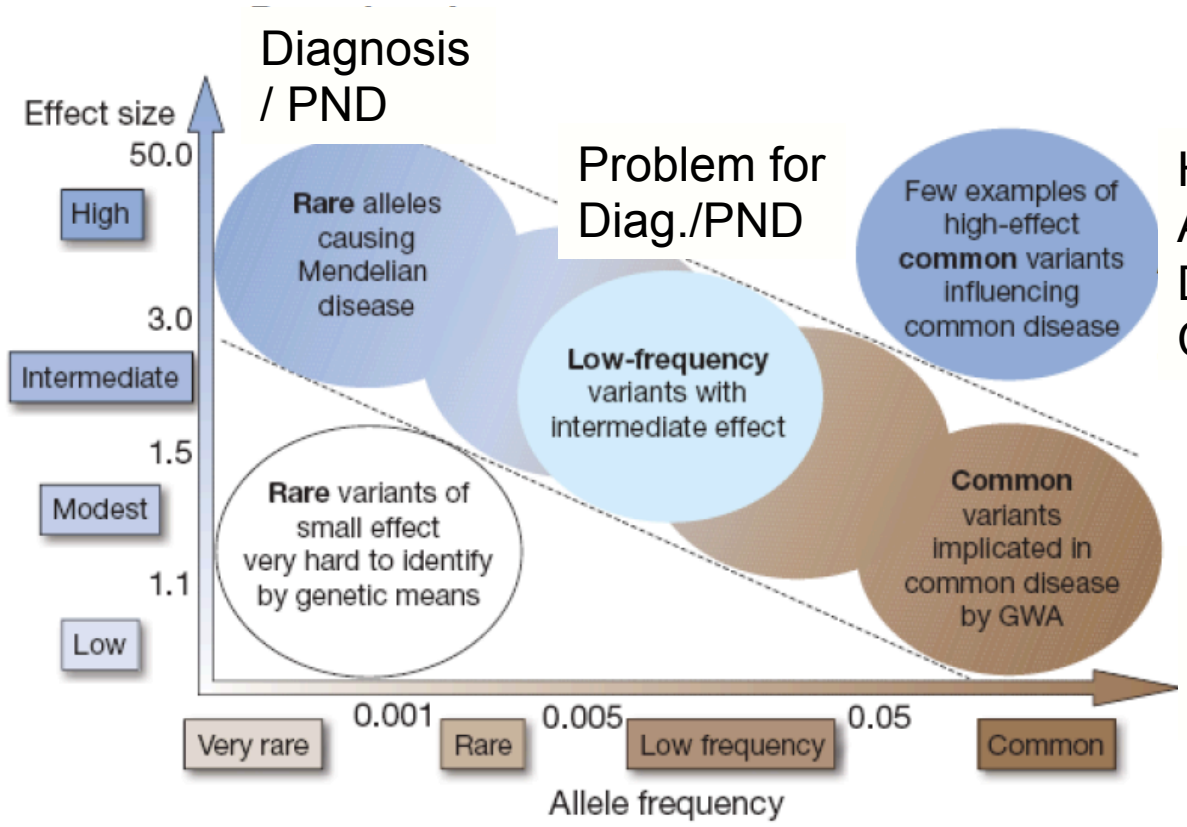
[a]Statistically significant in fixed and random effects models. [b]Statistically significant in fixed effects model only; opposite direction from original report. OR, estimated pooled odds ratio; CAD, coronary artery disease; MI, myocardial infarction.

Nature Genetics, 2003

# Rares variants, common polymorphisms

- « Whole genome association »strategies are based on one hypothesis: common disease, frequent genetic risk factor in the population (common disease/common variant); in these cases, RR is seldom above 2. But monogenic disease studies suggests that individually rare variants can have high effect on the disease risk.
- Variant fr 1/10, RR2: attributable risk = 9%
- Variant fr 1/1000, RR10: AR = 1%
- In the 2nd case: the variant is important for the individual, but little contributive for the population

# Genetic variants and et common and rare diseases



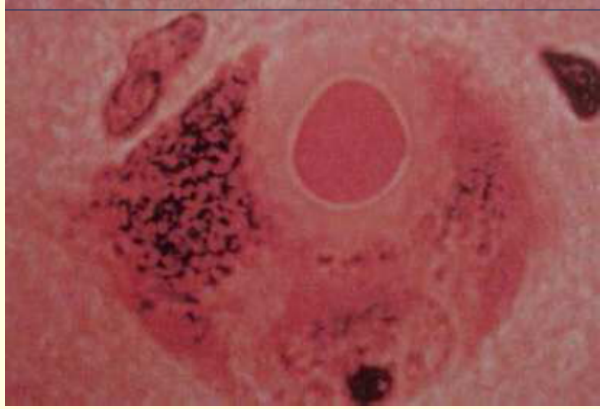

Problem for Diag./PND

HLA
ApoE & Alzheimer
DMLA
Chron

No diagnostic application

**Figure 1 | Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio).** Most emphasis and interest lies in identifying associations with characteristics shown within diagonal dotted lines. Adapted from ref. 42.

# Rare variants, common diseases



Levy bodies in substantia nigra

- An emerging concept
- Parkinson disease – multifactorial form
- glucocerebrosidase (GBA)

(homozygous mutations —> Gaucher disease)

Gan-Or et al. Neurology 2008 : moderate (N370S, R496H) and severe mutations (84insGG, IVS2+1, V394L, D409H, L444P, RecTL) loss of function mutations.

- Relative risk heterozygous GBA mutations

| genotype | patients % | controls % | Relative risk RR | Attributable risk RA (%) |
|---|---|---|---|---|
| Heterozygous mutations | 17.9 | 6.3 | x 3.2 | 12% |
| No mutation | 82.1 | 93.7 | | |

$$RR_{heteroz} = \frac{\frac{0.03 \times 0.179}{0.063}}{\frac{0.03 \times 0.821}{0.937}} = 3.2$$

Confirmed in Neumann et al. Brain 2009

# CONCLUSIONS :

Identifications of rare variants in common diseases

—> next generation sequencing

requires human genome sequence for 1 000 $…

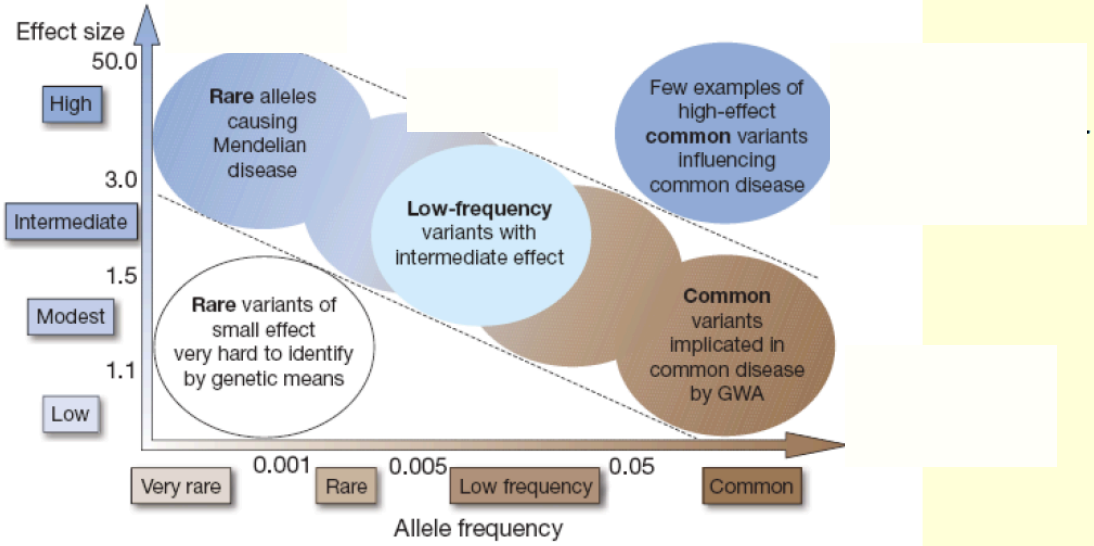



**Figure 1 | Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio).** Most emphasis and interest lies in identifying associations with characteristics shown within diagonal dotted lines. Adapted from ref. 42.

Progress in interpretation, and in diagnostic applications will be much more slow, due to the complexity of the risk factors