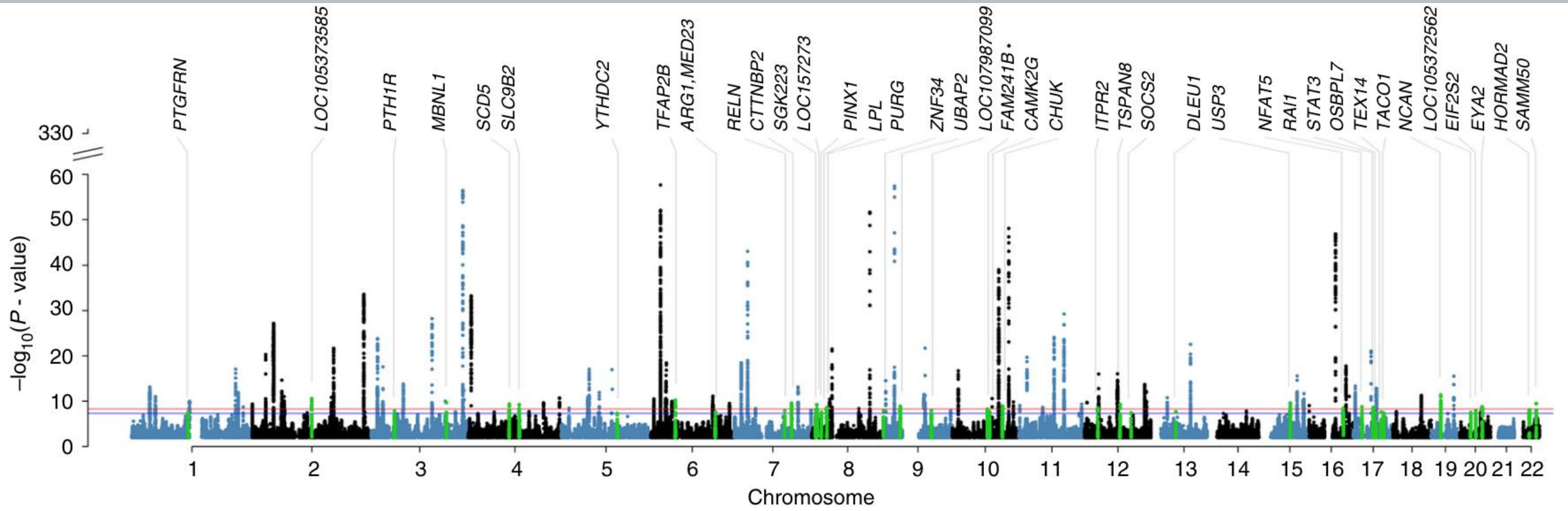
Genome-wide Association Studies (GWAS)

#6

PALLE DUUN ROHDE

palledr@hst.aau.dk



LETS GET STARTED

AGENDA

08:15 – 08:30	Recap [*Complex triats and genetic parameters*]
08:30 – 09:00	Group presentations from last
09:00 – 09:15	Break
09:15 – 09:45	Lecture 1 [*Genetic associations + GWAS part 1*]
09:45 – 10:15	Exercise 1 + 2
10:15 – 10:30	Break
10:30 – 11:00	Lecture 2 [*GWAS part 2*]
11:00 – 11:55	Group work
11:55 – 12:00	Evaluation at Moodle

AGENDA

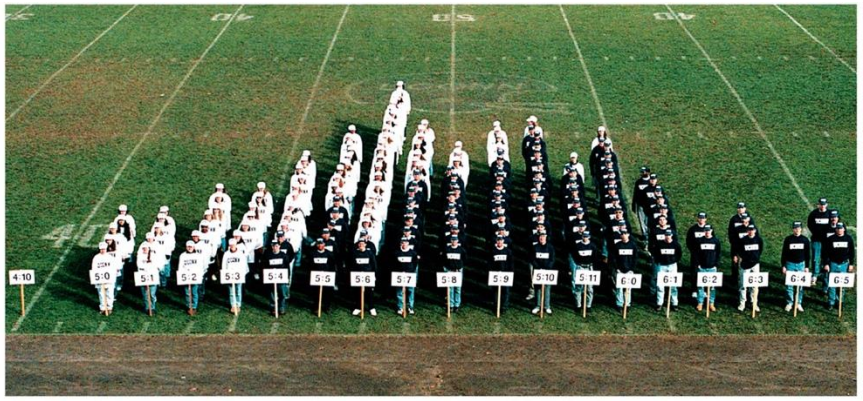
08:15 – 08:30	Recap [*Complex triats and genetic parameters*]
08:30 – 09:00	Group presentations from last
09:00 – 09:15	Break
09:15 – 09:45	Lecture 1 [*Genetic associations + GWAS part 1*]
09:45 – 10:15	Exercise 1 + 2
10:15 – 10:30	Break
10:30 – 11:00	Lecture 2 [*GWAS part 2*]
11:00 – 11:55	Group work
11:55 – 12:00	Evaluation at Moodle

MULTIFACTORIAL INHERITANCE

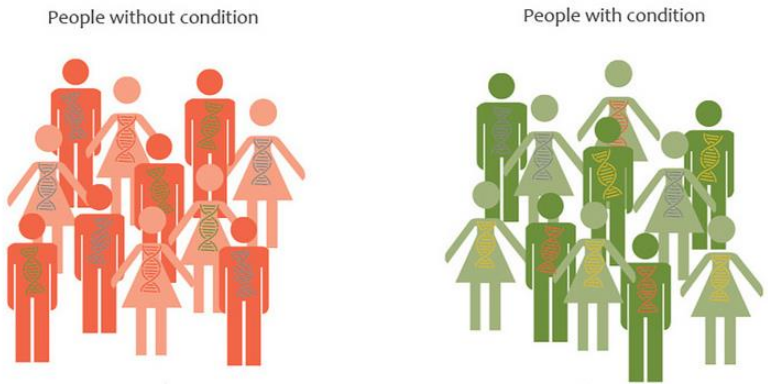
- Monogenic (single gene/variant)
- Polygenic (multiple gene variants at the same time)
- Multifactorial (multiple gene variants + environment variation)

MULTIFACTORIAL TRAITS

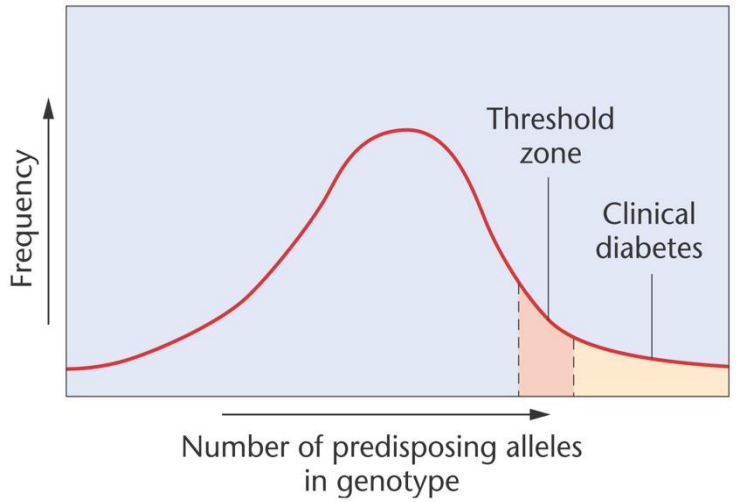
Continous variation



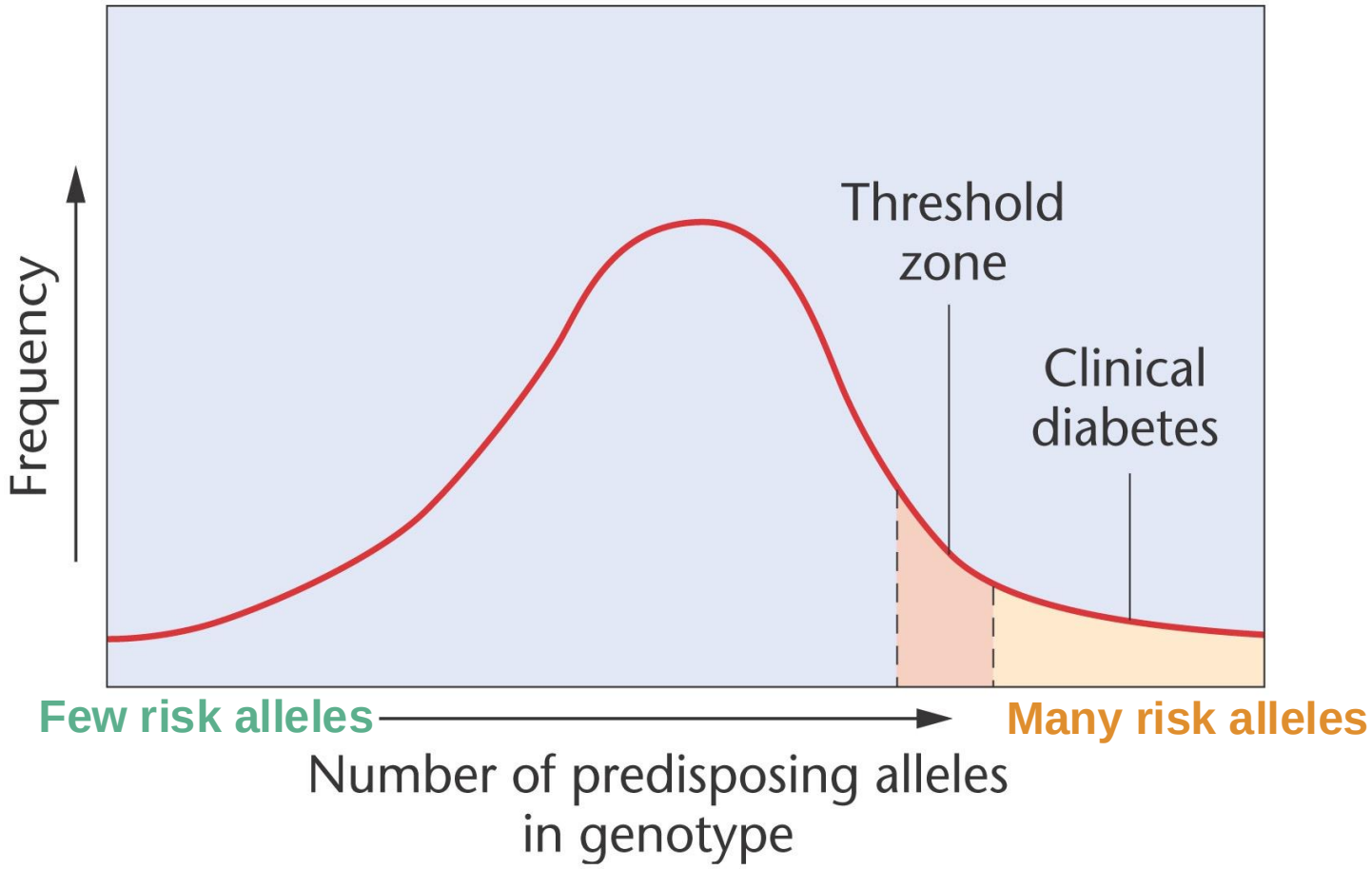
Dichotomous outcome



Categorical variation



LIABILITY (THRESHOLD) MODEL



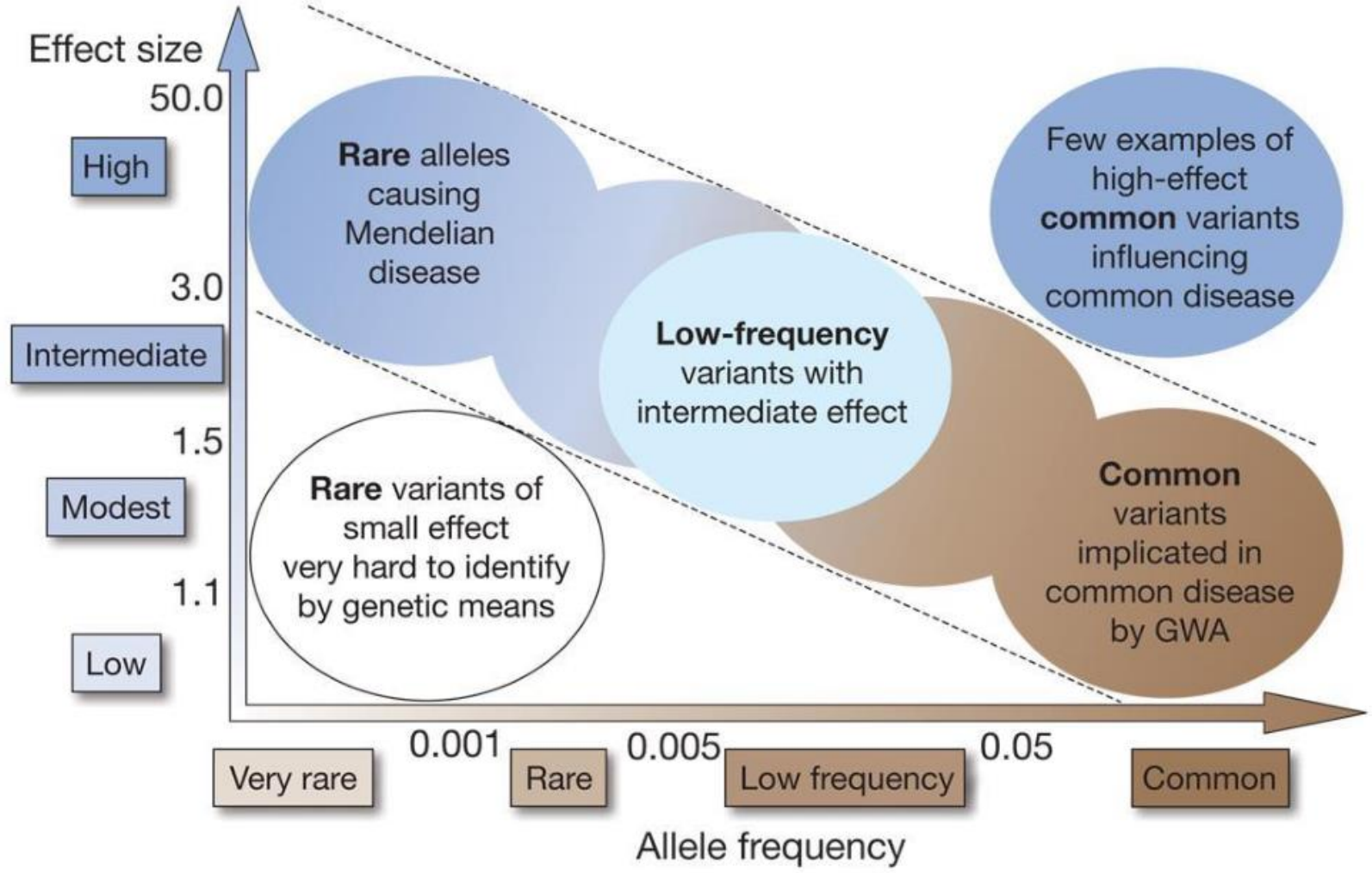
Liability model

Only individuals with a liability over a certain threshold will become affected

The **sum** of many genetic variants with **small effect/risk**.

Each locus follow Mendelian inheritance pattern, although the trait does not

GENETIC VARIATION



Multifactorial traits are caused by the sum of MANY variants exherting small effects

MULTIFACTORIAL TRAITS

External effects that modulates the phenotypic value.

… or things that we cannot explain

Multifactorial traits = polygenic effect + environmental effect

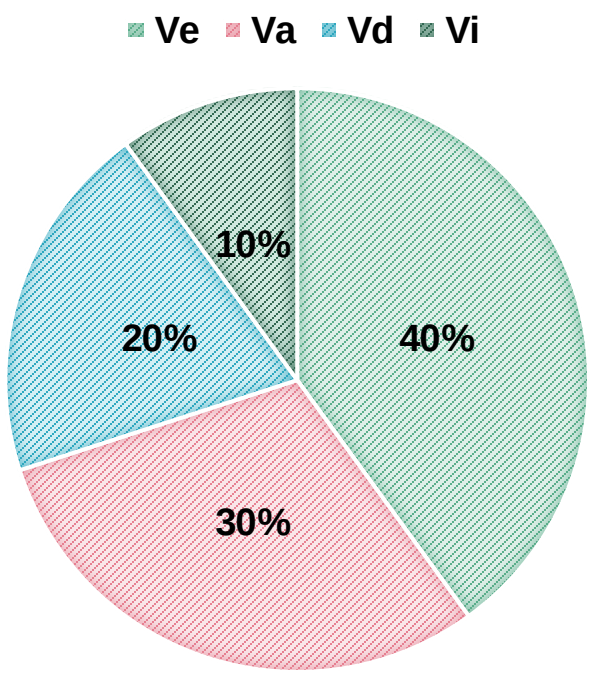
many genes/alleles

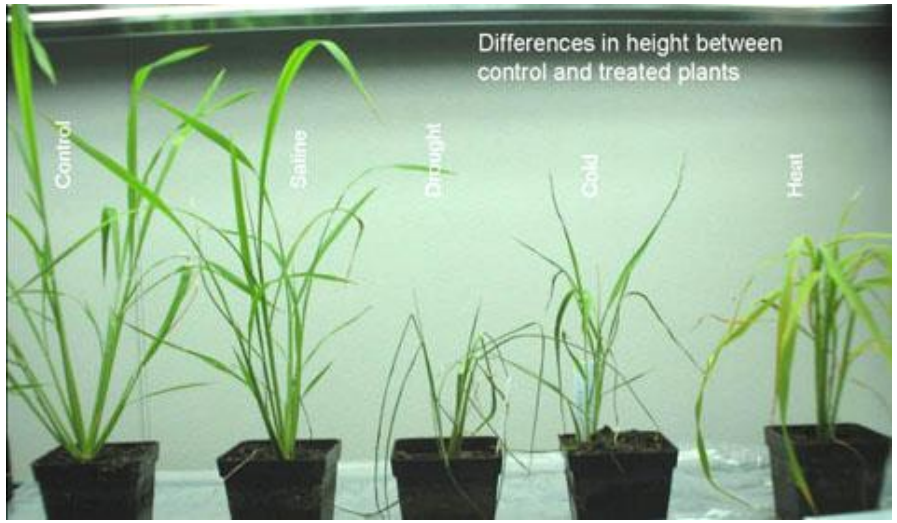
$V_P = V_G + V_E$

$V_P = V_a + V_d + V_i + V_e$

$H^2 = V_G / (V_G + V_E)$

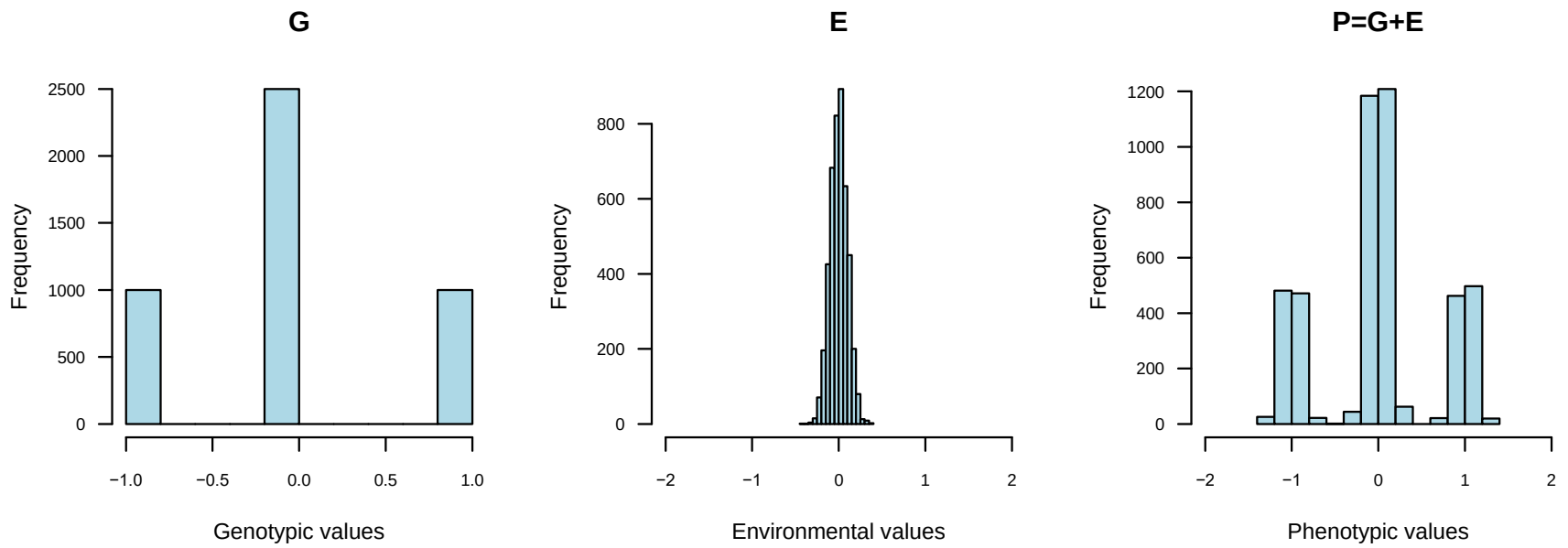
$h^2 = V_a / (V_a + V_E)$





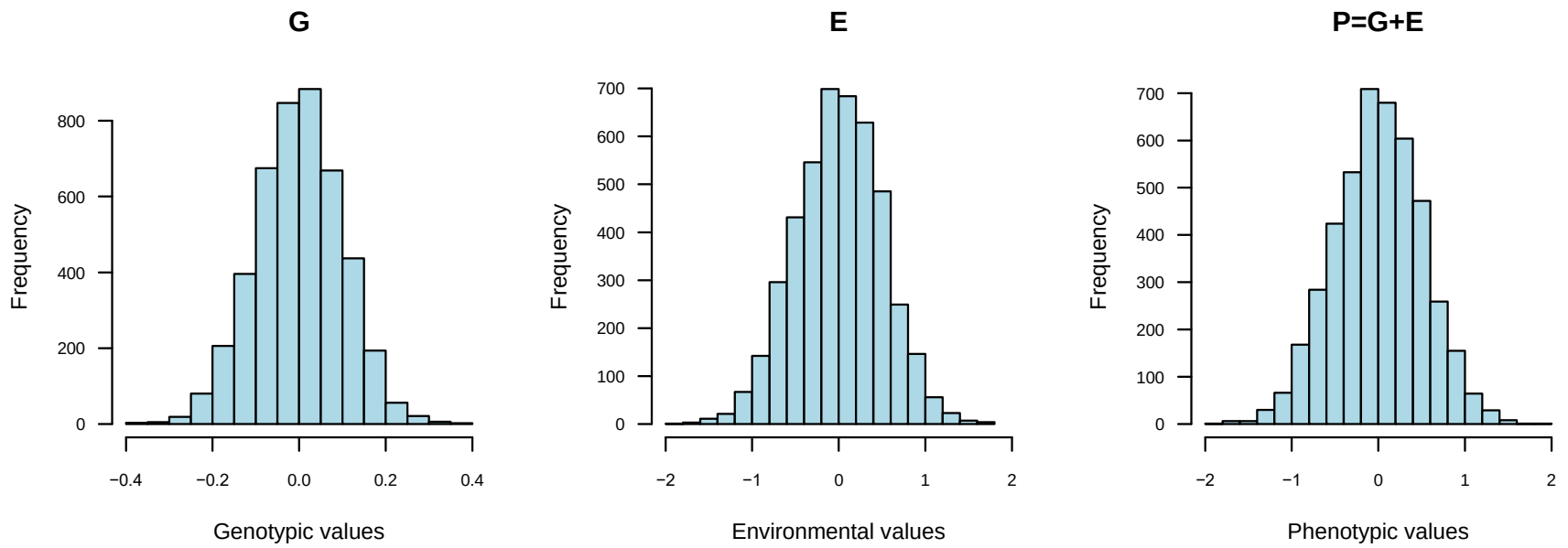
MULTIFACTORIAL TRAITS

Environmental variance (non-genetic factors) blurs phenotypic classes



MULTIFACTORIAL TRAITS

Genotypic and environmental variance creates infinite many phenotypic classes



BROAD-SENSE HERITABILITY

Broad-sense **heritability (H^2) describes the proportion of the phenotypic variance that is explained by genetic difference variation**.

$$V_P = V_G + V_E$$

$$H^2 = \frac{V_G}{V_P} = \frac{V_G}{V_G + V_E}$$

H^2 can take values between 0 and 1:

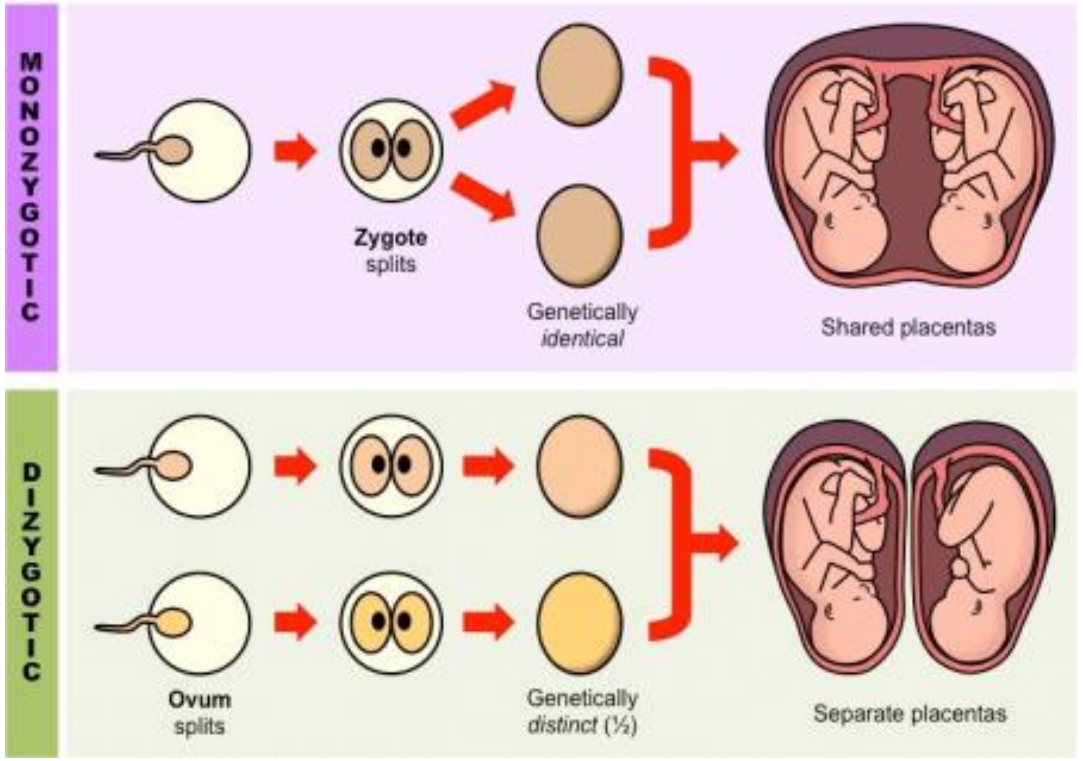
$H^2 = 0$→ all variation is due to environmental variation

$H^2 = 1$→ all variation is due to genetic variation

USING TWINS TO ESTIMATE H^2

Monozygotic twins (MZ): Difference in V_P is V_E

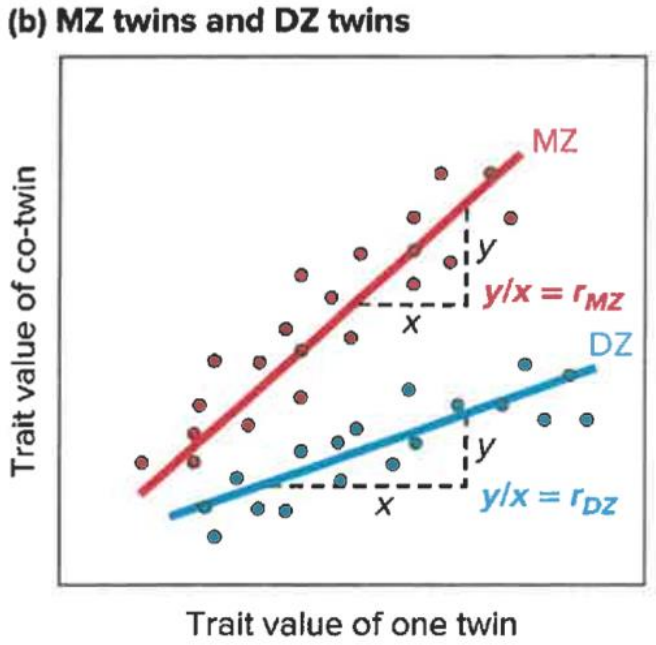
Dizygotic twins (DZ): Difference in V_P is $V_G + V_E$



Compare MZ [V_e] and DZ [$V_G + V_E$] twins – the difference is V_G

USING TWINS TO ESTIMATE H^2

QUANTITATIVE COMPLEX TRAITS



Because DZ share 50% of the genetic material

$$H^2 = 2(r_{MZ} - r_{DZ})$$

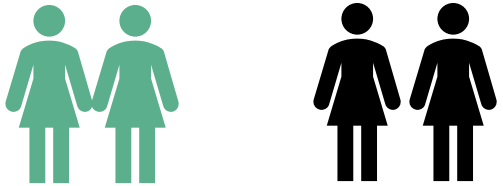
All phenotypic variation between MZ is environment, whereas all phenotypic variation between DZ is both

Thus, the difference is genetic variation

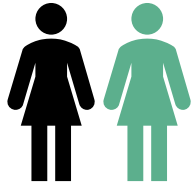
USING TWINS TO ESTIMATE H^2

DICHOTOMOUS COMPLEX TRAITS

Concordance rate (C) = the frequency with which the other twin has the trait



Concordant pair: same phenotype



Discordant pair: only one in the pair has the trait

$$H^2 = \frac{C_{MZ} - C_{DZ}}{1 - C_{DZ}}$$

Important

There has to be a difference between C_{MZ} and C_{DZ} for a trait to be under genetic influence

The larger ratio $\frac{C_{MZ}}{C_{DZ}}$ the higher H^2

If $C_{MZ} < 1$ environmental exposures affect the trait

GROUP WORK

THE HERITABILITY OF HUMAN DISEASE

PART 1

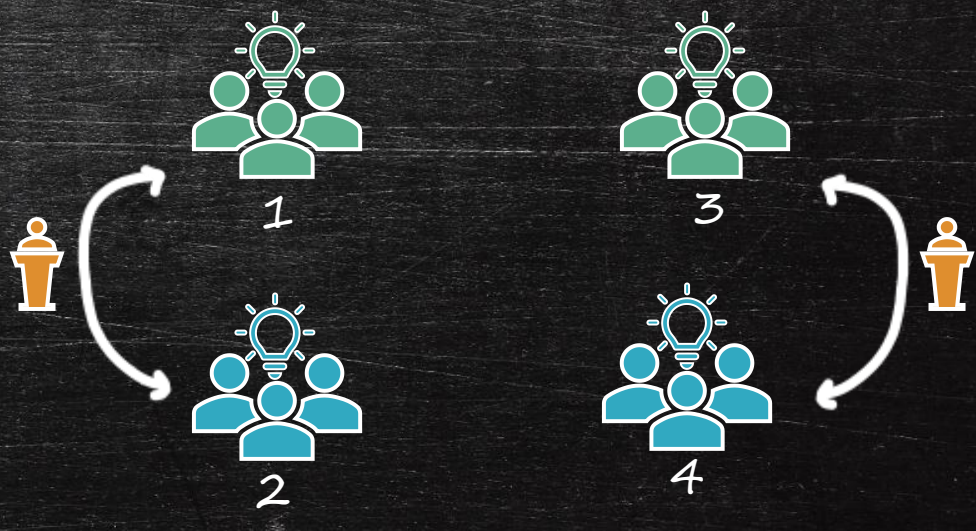
1) Make 4 groups & prepare a 5-7 min presentation
   - ❑ Group 1 & 3 works with section 'Estimating heritability' p141-144
   - ❑ Group 2 & 4 works with section 'Biased heritability' p144-148

PART 2 – *next time (17/3)*

- ❑ Group 1 present to group 2 and *vise versa*
- ❑ Group 3 present to group 4 and *vise verse*

What did you find difficult?

Artiklen er på et meget højt og svært niveau, som kan være ret svær at forstå - selvom man har læst den flere gange :(

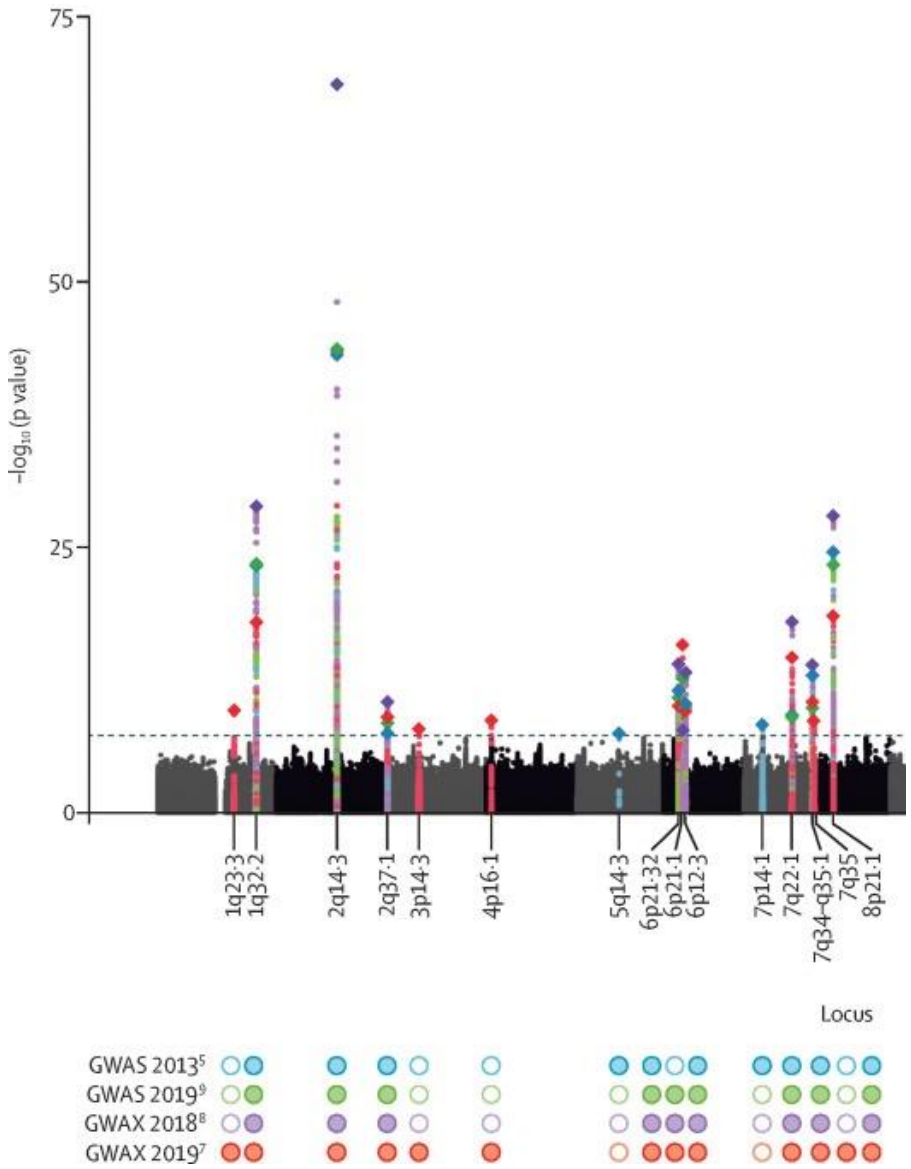



BREAK

AGENDA

08:15 – 08:30	Recap [*Complex triats and genetic parameters*]
08:30 – 09:00	Group presentations from last
09:00 – 09:15	Break
09:15 – 09:45	Lecture 1 [*Genetic associations + GWAS part 1*]
09:45 – 10:15	Exercise 1 + 2
10:15 – 10:30	Break
10:30 – 11:00	Lecture 2 [*GWAS part 2*]
11:00 – 11:55	Group work
11:55 – 12:00	Evaluation at Moodle

GENETIC ASSOCIATION



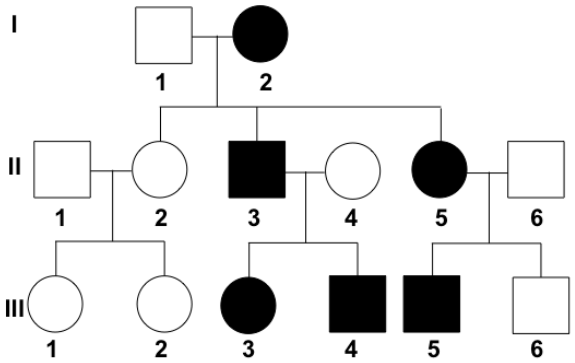
INHERITANCE PATTERN

OF MULTIFACTORIAL TRAITS

The exact inheritance pattern depends on

- The number of risk genes/alleles involved
- The effect size distribution of the genetic risk variants
- The interaction among genetic variants
- The interaction with environmental exposures

Monogenic disorders



Mutation

Each genetic variant is both necessary and sufficient

Linkage analysis
(pedigree)

Common complex disorders

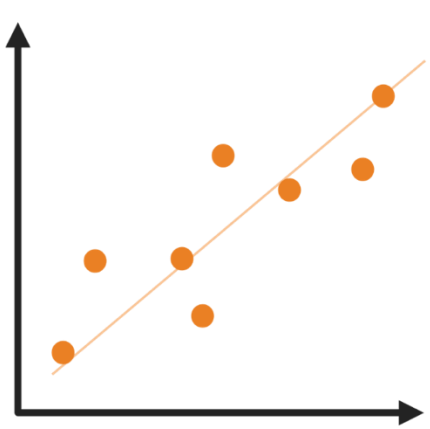
Each genetic variant is neither necessary nor sufficient

Association study
(unrelated population)

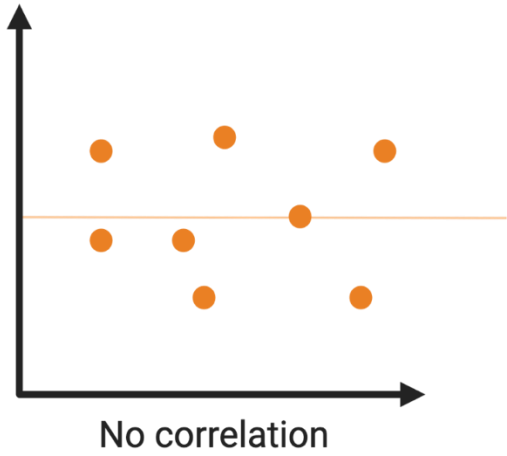
? What does association mean

ASSOCIATION

An association defines a relationship between two entity objects based on common attributes.



High positive correlation



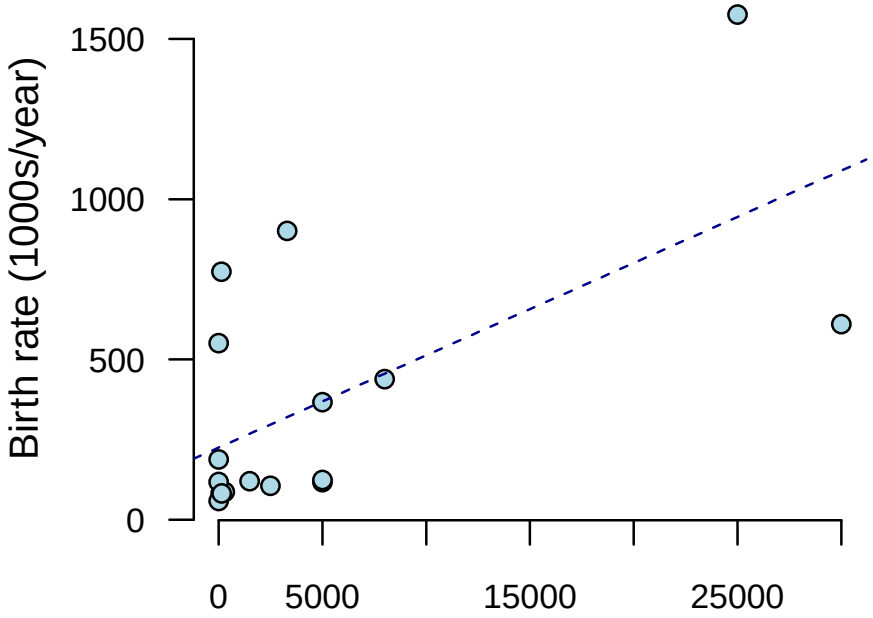
No correlation

Is there an association between exposure and outcome?

ASSOCIATION *NOT CAUSATION*

Country	Area (km^2)	Storks (pairs)	Humans (10^6)	Birth rate (10^3/yr)
Albania	28,750	100	3.2	83
Austria	83,860	300	7.6	87
Belgium	30,520	1	9.9	118
Bulgaria	111,000	5000	9.0	117
Denmark	43,100	9	5.1	59
France	544,000	140	56	774
Germany	357,000	3300	78	901
Greece	132,000	2500	10	106
Holland	41,900	4	15	188
Hungary	93,000	5000	11	124
Italy	301,280	5	57	551
Poland	312,680	30,000	38	610
Portugal	92,390	1500	10	120
Romania	237,500	5000	23	367
Spain	504,750	8000	39	439
Switzerland	41,290	150	6.7	82
Turkey	779,450	25,000	56	1576

Table 1. Geographic, human and stork data for 17 European countries




```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.250e+02  9.356e+01   2.405   0.0295 *
Storks      2.879e-02  9.402e-03   3.063   0.0079 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Matthews, R. (2000), Storks Deliver Babies (p= 0.008). Teaching Statistics, 22: 36-38.

ASSOCIATION

Rare variants may have a strong effect on outcome (large OR)

Common variants have small effect on outcome (OR=1.1-1.8)

OR>1
- -allele is risk allele
- -allele is seen more often among cases

OR<1
- -allele is protective
- -allele is seen more often among controls

Is there an association between genetic variant and disease status?

Outcome

Exposure	Case	Control	Odds
Carry variant (A –)	a	b	a/b
Don't carry variant (aa)	c	d	c/d

Odds ratio (OR) is the ratio of the odds of disease among the exposed to the odds of disease among the unexposed

$$OR = \frac{a/b}{c/d} = \frac{a \times d}{c \times b}$$

ASSOCIATIONS

ALLELIC VS GENOTYPIC

We can count alleles OR genotypes. Example;

Association of rs6983267 on 8q24 with colorectal cancer [C/T, allele C is the risk allele]

	CC	CT	TT
Cases	250	375	150
Controls	460	940	500

GENOTYPIC ASSOCIATION

Note, these ORs are relative to TT (the lowest-risk genotype)

$$OR_{TT} = \frac{odds(disease|TT)}{odds(disease|TT)} = 1$$

$$OR_{CT} = \frac{odds(disease|CT)}{odds(disease|TT)} = \frac{375 \times 500}{940 \times 150} = 1.33$$

$$OR_{CC} = \frac{odds(disease|CC)}{odds(disease|TT)} = \frac{250 \times 500}{460 \times 150} = 1.81$$

ASSOCIATIONS

ALLELIC VS GENOTYPIC

We can count alleles OR genotypes. Example;

Association of rs6983267 on 8q24 with colorectal cancer [C/T, allele C is the risk allele]

	C	T
Cases	875	675
Controls	1860	1940

ALLELIC ASSOCIATION

Cases
C alleles = 2 * 250 CC + 375 CT = 875
T alleles = 2 * 150 TT + 375 CT = 675

Controls
C alleles = 2 * 460 CC + 940 CT = 1860
T alleles = 2 * 500 TT + 940 CT = 1940

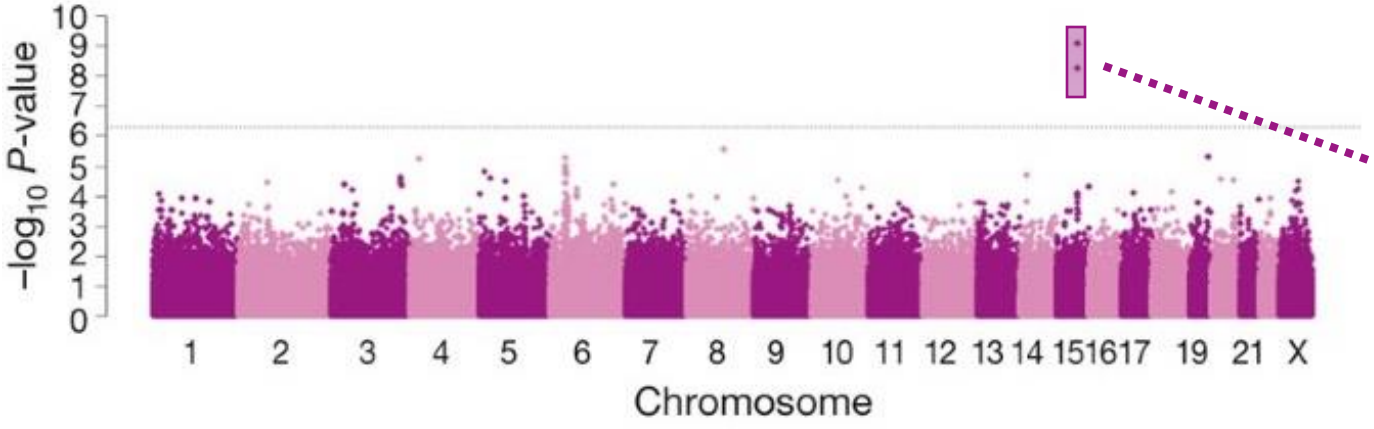
$$OR_C = \frac{odds(disease|C)}{odds(disease|T)} = \frac{875 \times 1940}{1860 \times 675} = 1.35$$

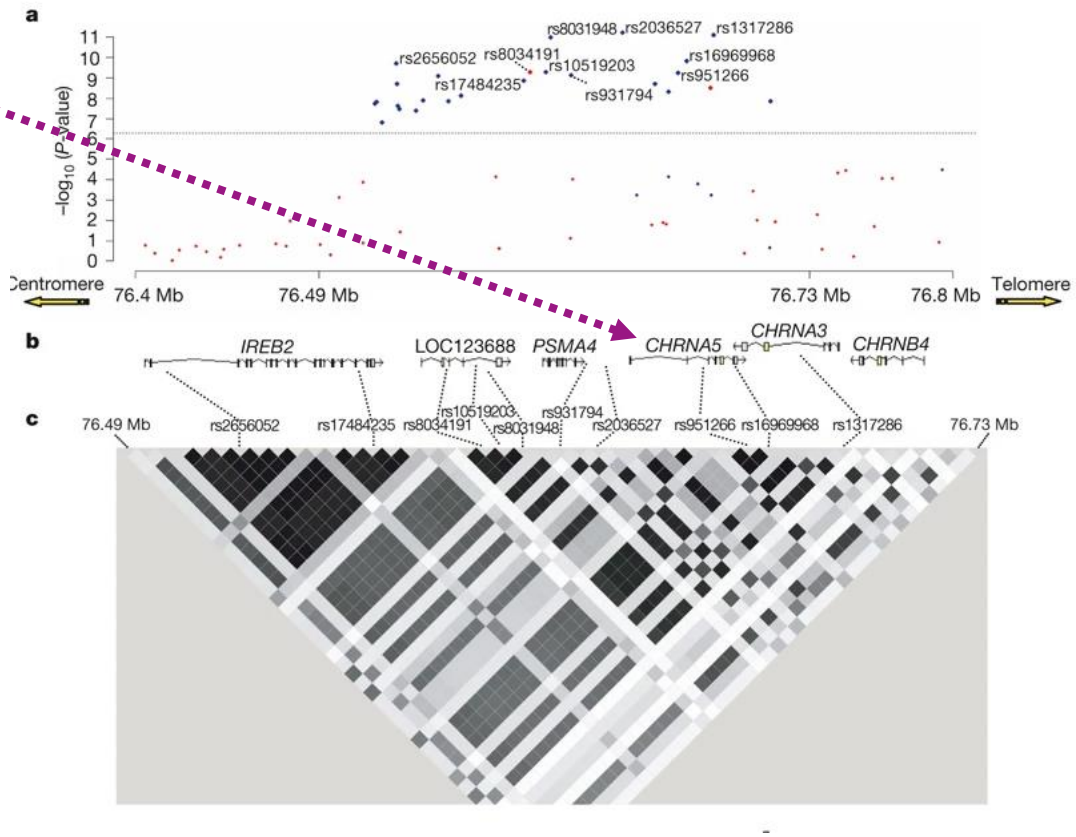
GENETIC RISK FOR LUNG CANCER?

Vol 452|3 April 2008|doi:10.1038/nature06885 nature

LETTERS

A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25





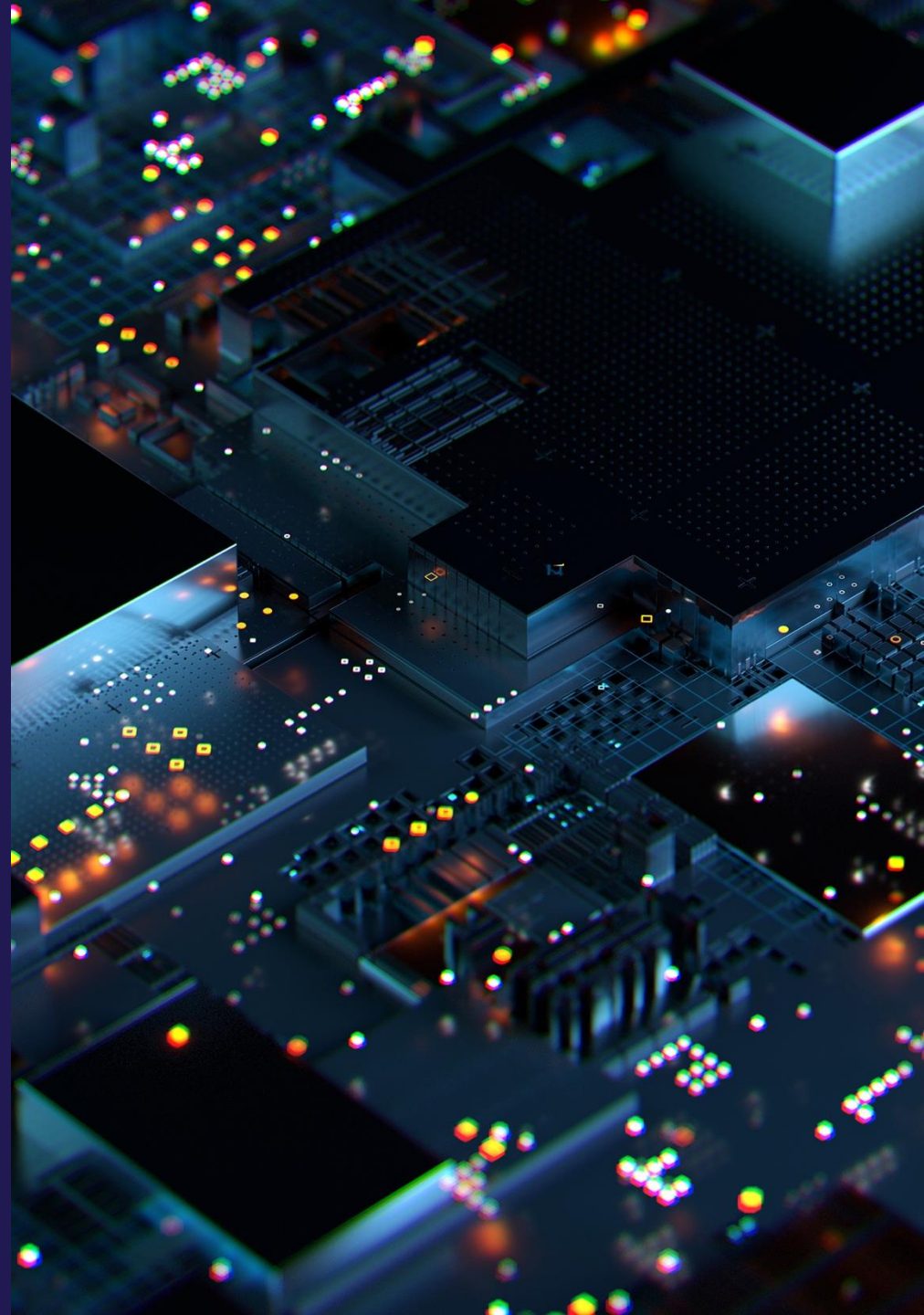
A genetic studie of lung cancer (LC, 1989 cases and 2625 controls) found a nicotine receptor (*CHRNA3/5*) to be associated with the risk of developing lung cancer.

Does that then mean that *CHRNA3/5* is a risk loci for LC?

Risk loci for nicotine addition → addicted to smoking → increase LC risk

GENOME-WIDE ASSOCIATION ANALYSIS (GWAS) – PART 1

- What is a GWAS
- LD



AALBORG UNIVERSITY

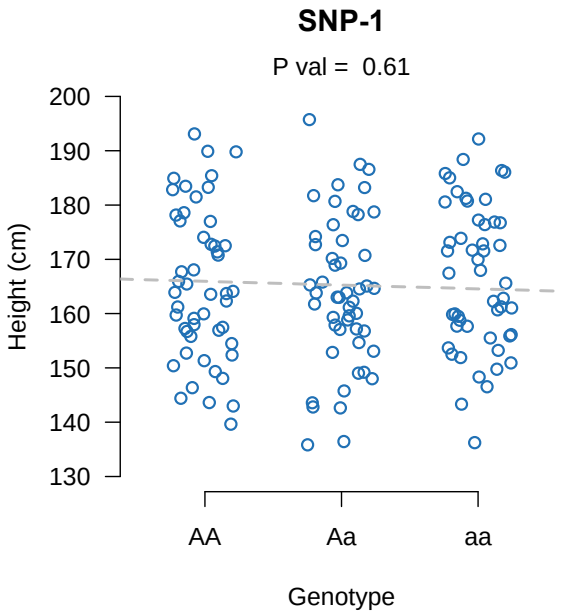
GENOME-WIDE ASSOCIATION STUDY (GWAS)

A systematic analysis of all common genetic variants without a priori hypothesis.

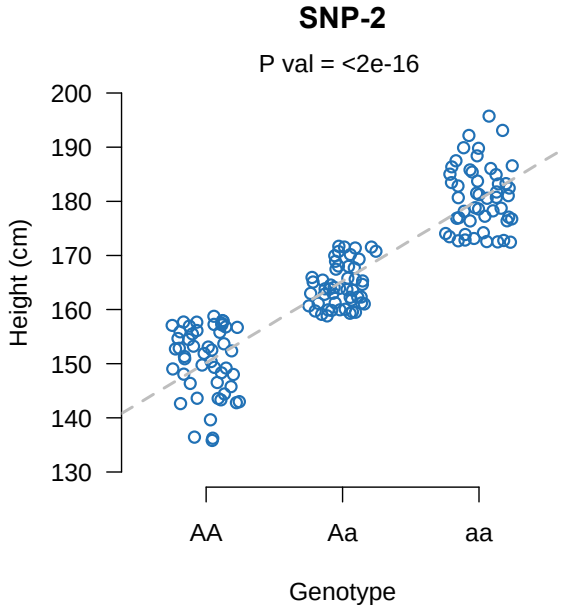
The aim is to identify risk variants for complex traits.

Testing one variant at a time

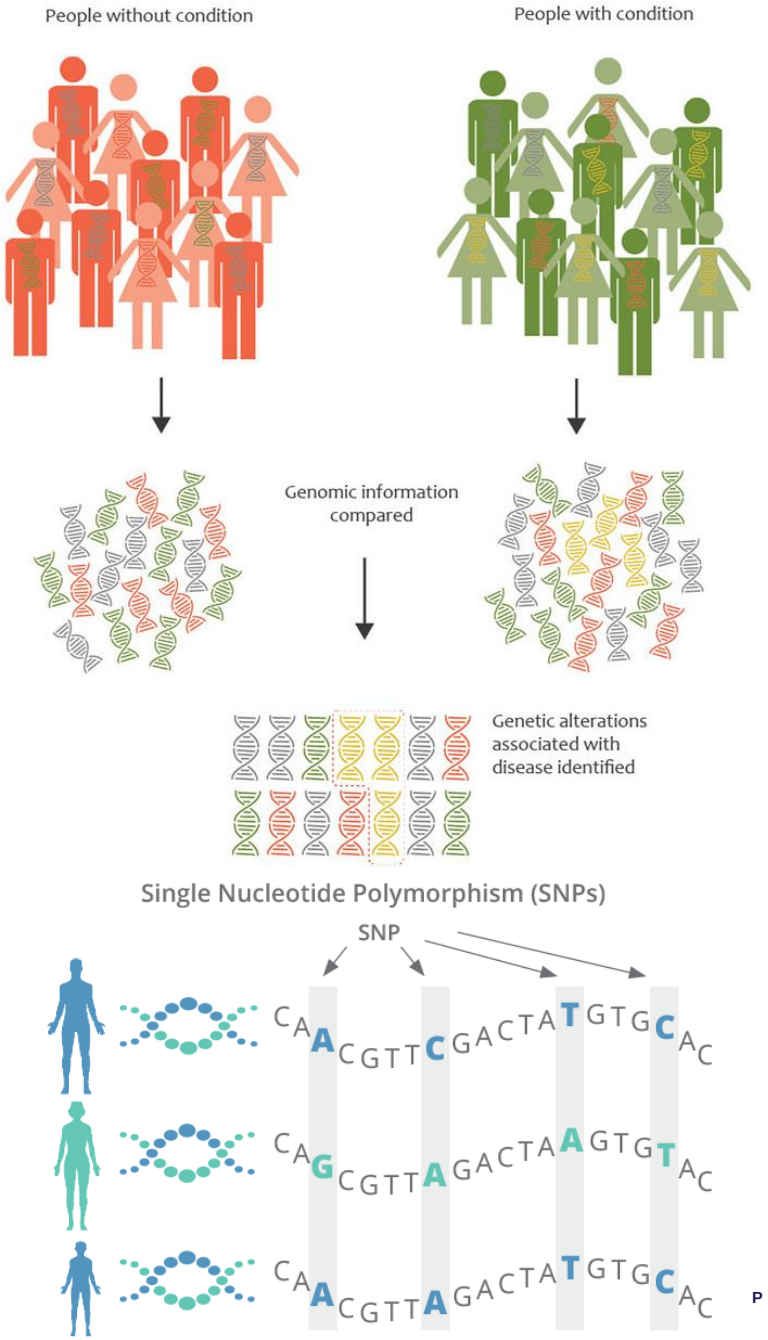
ASSOCIATION TESTING



No association

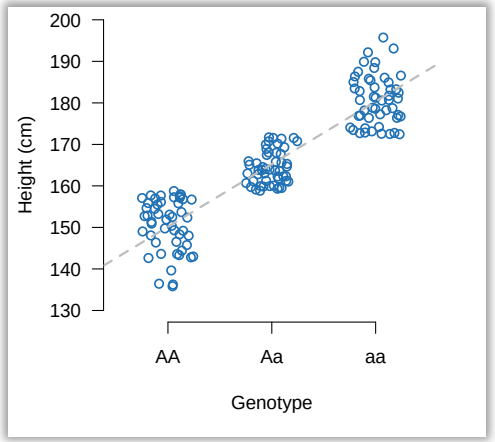


Association



THE BIOMETRIC MODEL

WHY THIS MODEL?



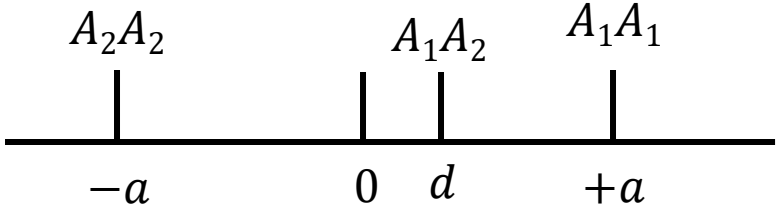
Consider a single locus with two alleles; A_1 and A_2, with the frequencies p and q.

Under HWE, the genotype frequencies are p^2, $2pq$, and q^2.

Under the biometric model, the genotypic value of A_1A_1 is a, and $-a$ for A_2A_2.

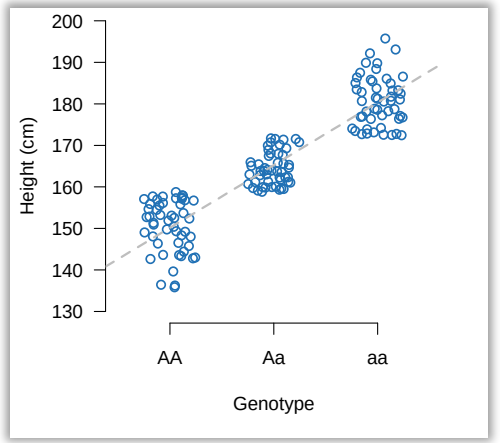
The genotype calue for A_1A_2 is d.

In this model, every single genetic variant with non-zero effect on a phenotype contributes to the population mean ($P = G + E$).



THE BIOMETRIC MODEL

WHY THIS MODEL?

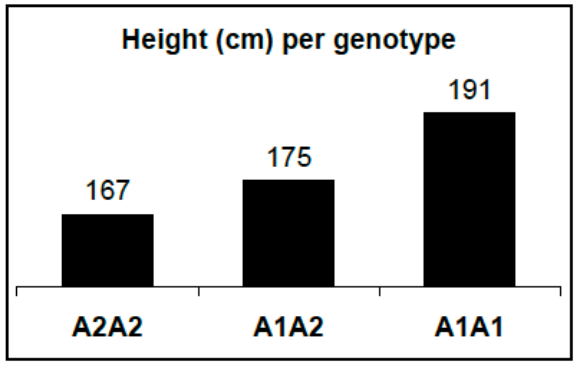


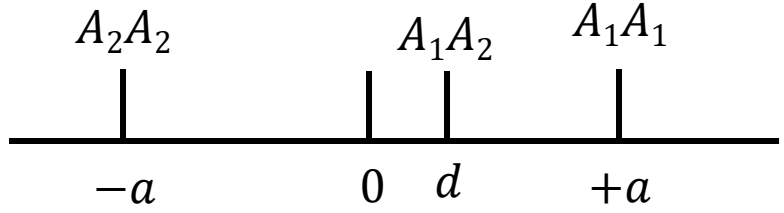
The mean effect (on the phenotype) is a function of the genotype effects ($-a$, a, and d) weighted by allele frequencies:

$$(ap^2) + (d2pq) + (-aq^2) = \mathrm{a(p - q) + d2pq}$$

Polygenic traits are influenced by many genetic variants; thus, assuming additive and independent effects, the population mean is:

$$\mu = \sum_{i=1}^{m} a_i(p_i - q_i) + \sum_{i=1}^{m} d_i 2p_i q_i$$



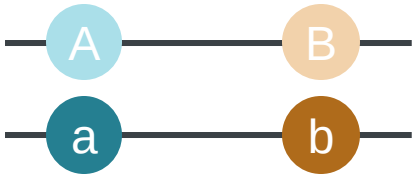


HOW MANY SITES IN THE GENOME?

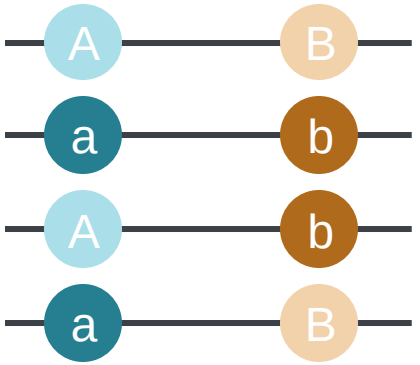
Should we test ALL 3,000,000,000 nucleotides within the genome?

No necessary, because of an old fried – Linkage Disequilibrium (LD)

WHAT IS LD?



Which gamtes can be produced?



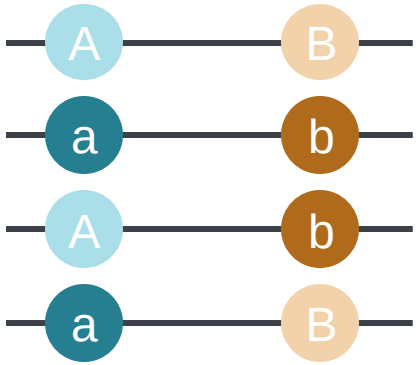
What are the frequencies of the alleles?

P(A), P(a)
P(B), P(b)

What are the frequencies of the haplotypes?

P(AB), P(Ab)
P(aB), P(ab)

WHAT IS LD?



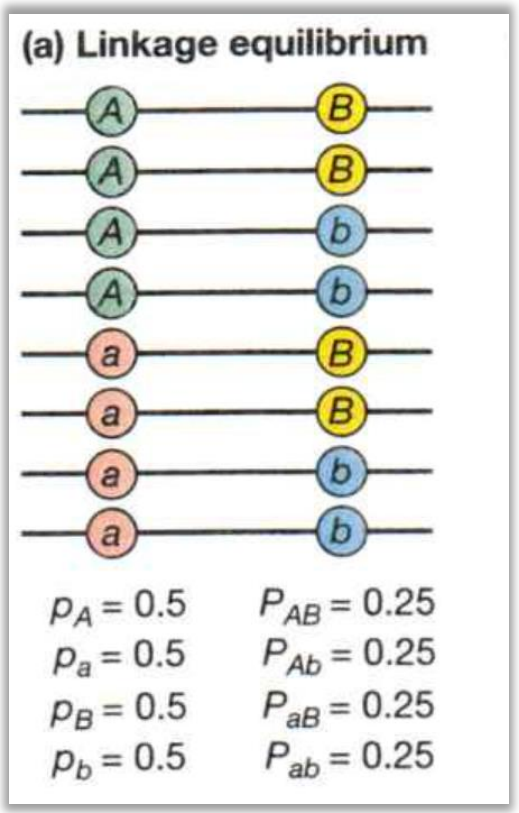
If there is random relationship among alleles at the two loci then the frequency of the haplotypes will be the product of the frequencies of the two alleles:

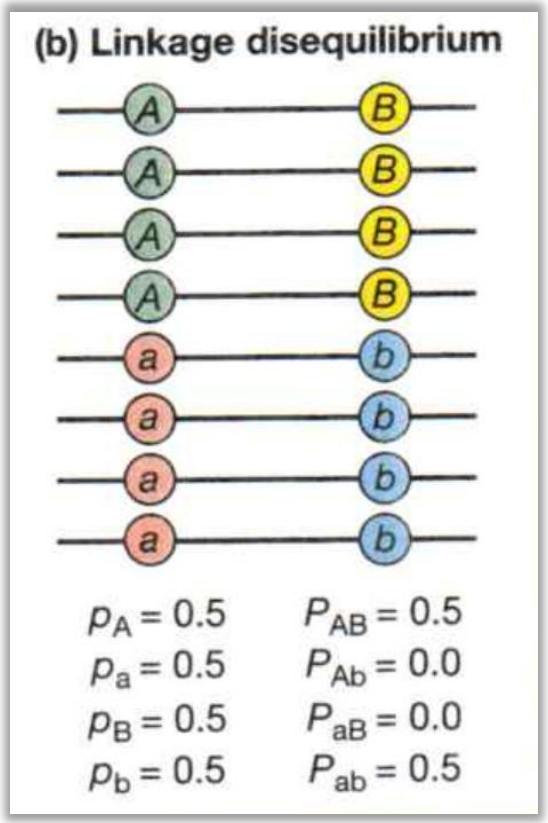
$P(AB)=P(A)xP(B)$

$P(Ab)=P(A)xP(b)$

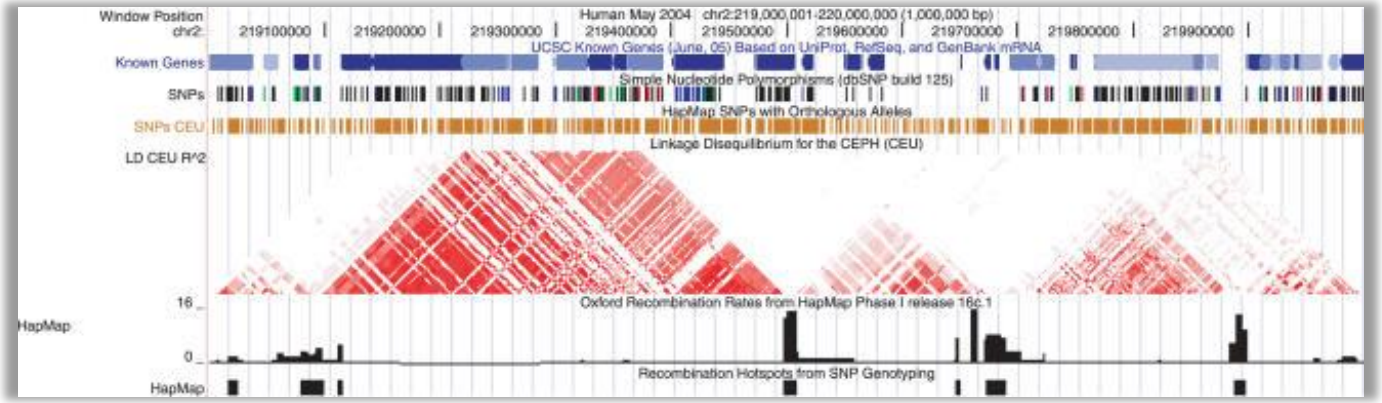
$P(aB)=P(a)xP(B)$

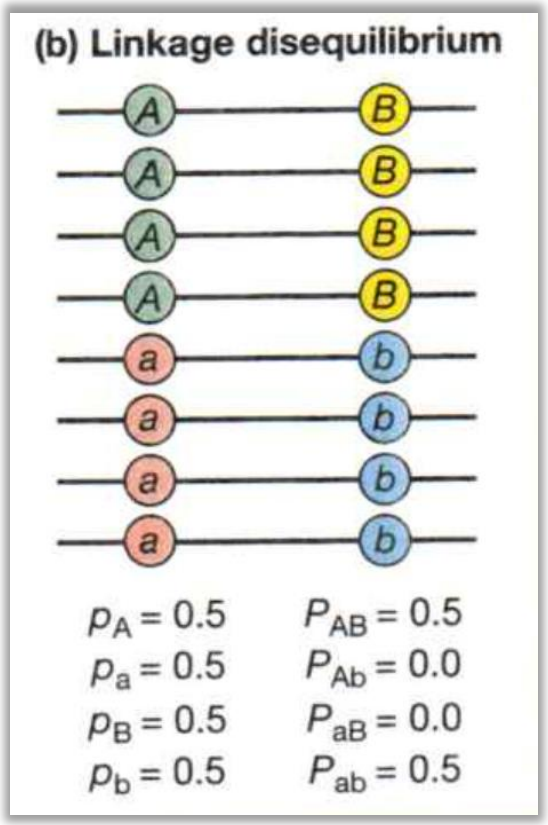
$P(ab)=P(a)xP(b)$





WHAT IS LD?



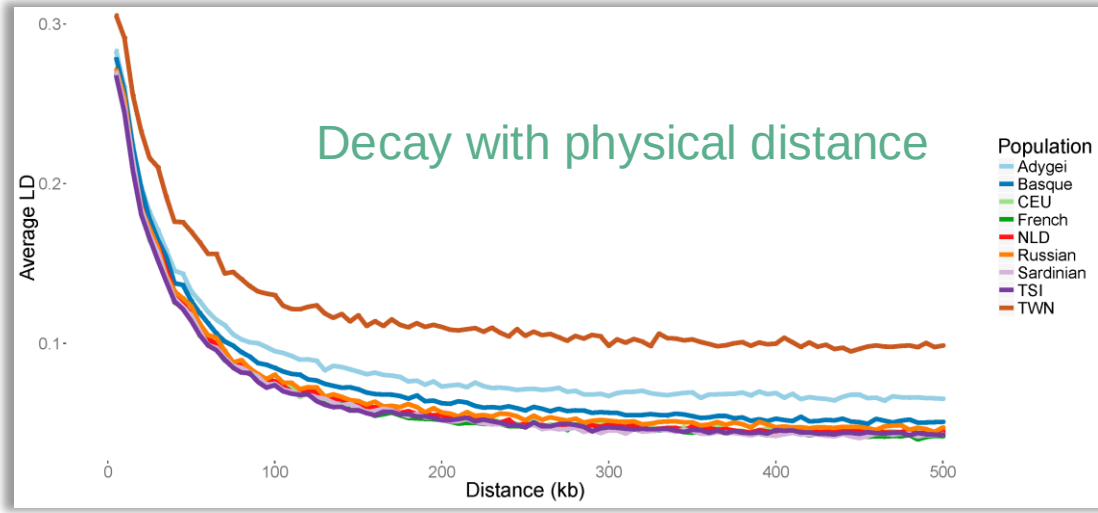


When the association between alleles at two loci is **non-random** they are said to be in **linkage disequilibrium**

The degree of LD can be measure in several ways – the simples one is:

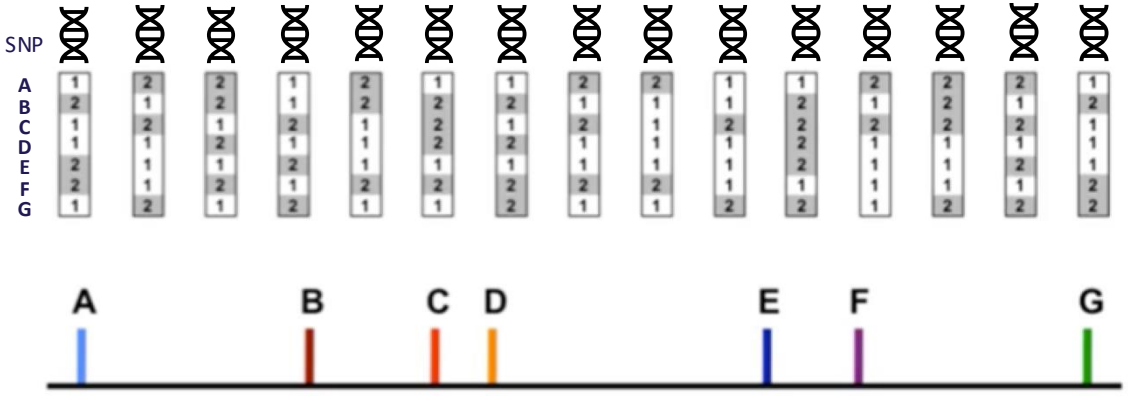
$$D = P_{AB} - P_A P_B$$

If D=0, no LD, if D>0 LD

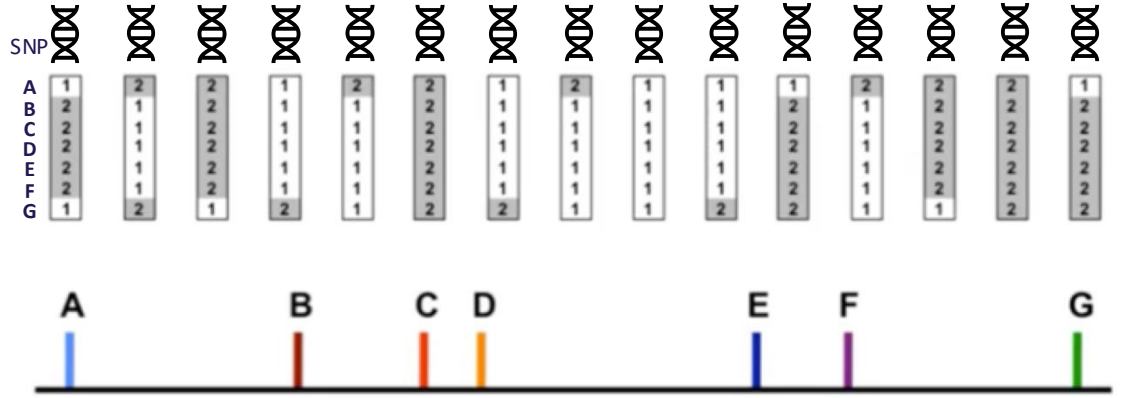


LD AND GENE MAPPING

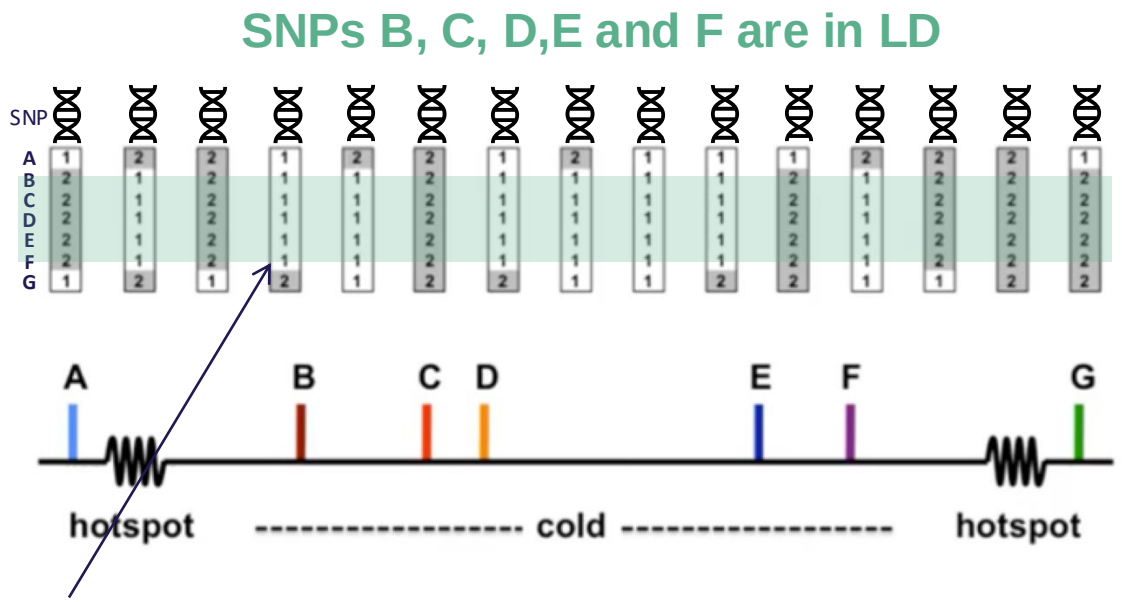
Linkage equilibrium – ***random association***



Linkage disequilibrium – *non-random association*



LD AND GENE MAPPING



If you have allele 1 here, I know what you are at the remaing sites in this haploblok

HOW MANY SITES IN THE GENOME?

Should we test ALL 3,000,000,000 nucleotides within the genome?

No necessary, because of an old fried – Linkage Disequilibrium (LD)

Typically, we test 5,000,000 – 10,000,000 SNPs.

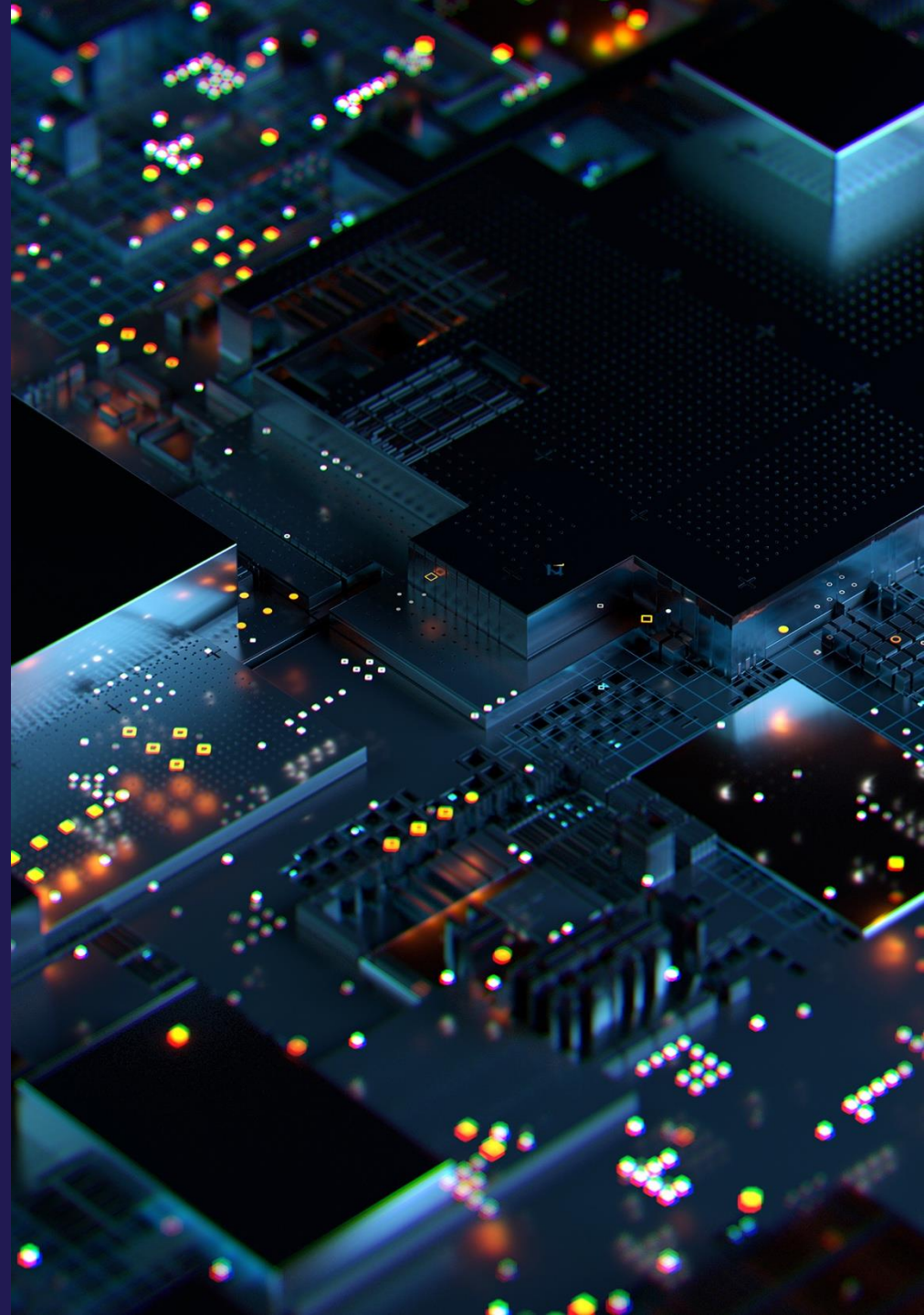
AGENDA

08:15 – 08:30	Recap [*Complex triats and genetic parameters*]
08:30 – 09:00	Group presentations from last
09:00 – 09:15	Break
09:15 – 09:45	Lecture 1 [*Genetic associations + GWAS part 1*]
09:45 – 10:15	Exercise 1 + 2
10:15 – 10:30	Break
10:30 – 11:00	Lecture 2 [*GWAS part 2*]
11:00 – 11:55	Group work
11:55 – 12:00	Evaluation at Moodle


BREAK

GENOME-WIDE ASSOCIATION ANALYSIS (GWAS) – PART 2

- What is a GWAS
- LD
- GWAS by steps



AALBORG UNIVERSITY

GWAS BY STEPS

GWAS BY STEPS

1. Select trait/disease

SELECT PHENOTYPE

Do we know anything about the phenotype already?

Is it heritable?

Do we know whether it is monogenic or polygenic?

Is it a common or rare disorder?

GWAS BY STEPS

2. Extract genetic variants

1. Select trait/disease

GENOTYPING

Because of LD you do not have to analyse all 3,000,000,000 variants in the genome.

Typically, we genotype ½ - 1 million variants

Because of LD we can impute (“guess” what variants are next to the genotyped variant) up til 10 million common genetic variants.

IMPUTATION

USING HAPLOTYPES

The true haplotypes

A	T	C
G	C	A

This individual has inherited a chromosome with alleles A-T-C from one parent, and G-C-A from the other parent

We observe only the genotypes

A/G T/C C/A

Genotype data does not carry information about the haplotypes.

We do not know whether A at SNP1 is coming from the same parent as T or C at SNP2

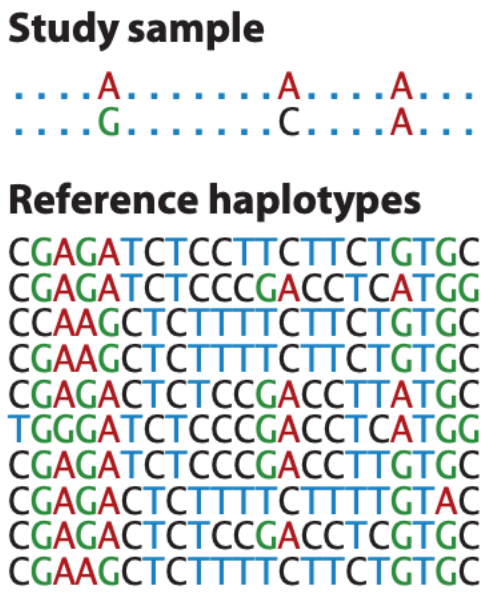
Different haplotypes

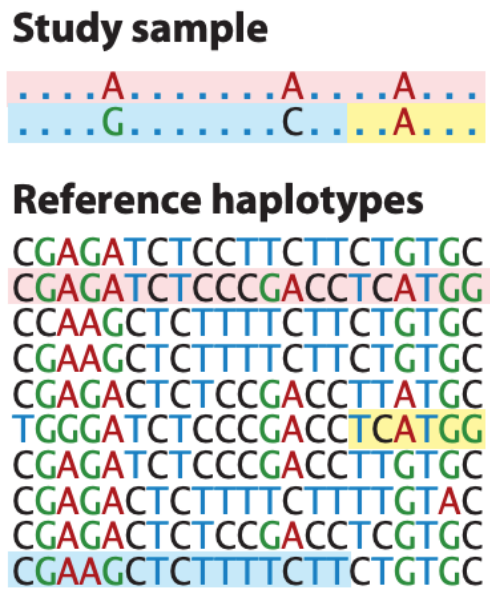
A	C	A
A	C	C
A	T	A
A	C	A
G	C	A
G	C	C
G	T	A
G	T	C

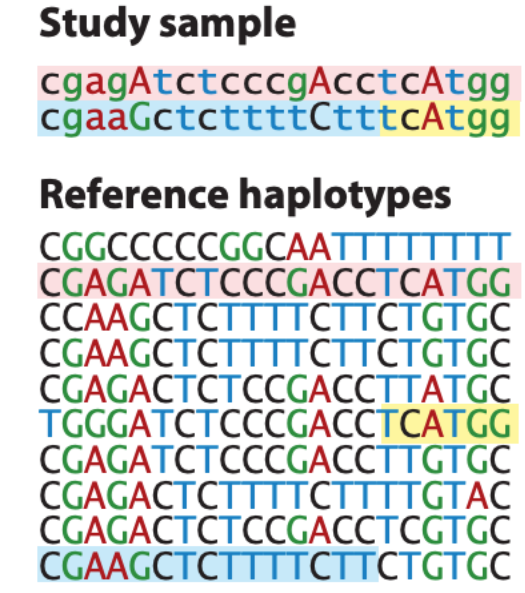
Phasing = estimate the most likely haplotypes

IMPUTATION

Genotypes







From a sequencing study

GWAS BY STEPS

3. GWAS

2. Extract genetic variants

1. Select trait/disease

STATISTICAL ANALYSIS

For each genetic variant [genotyping chip or after imputation] measure the degree of association between SNP and disease/trait

Fischers exact test

Exposure	Outcome: Case	Outcome: Control
Carry variant (A–)	a	b
Don't carry variant (aa)	c	d

$$OR = \frac{a \times d}{c \times d}$$

Linear regression

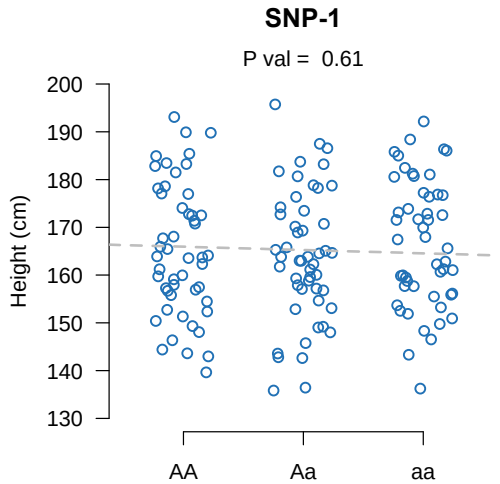
Allow us to account for confunding effects like, sex, age and ancestry.

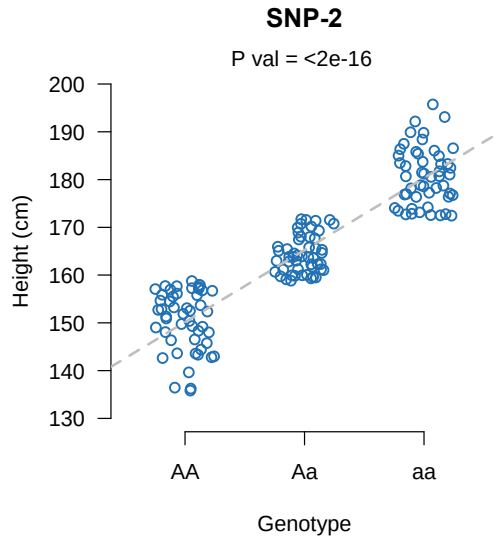
If trait follows Gaussian distribution:

- Linear regression
- Linear Mixed Model

If trait is dichotomous:

- Linear model with non-linear transformation
  [logistic regression with logit(x)]



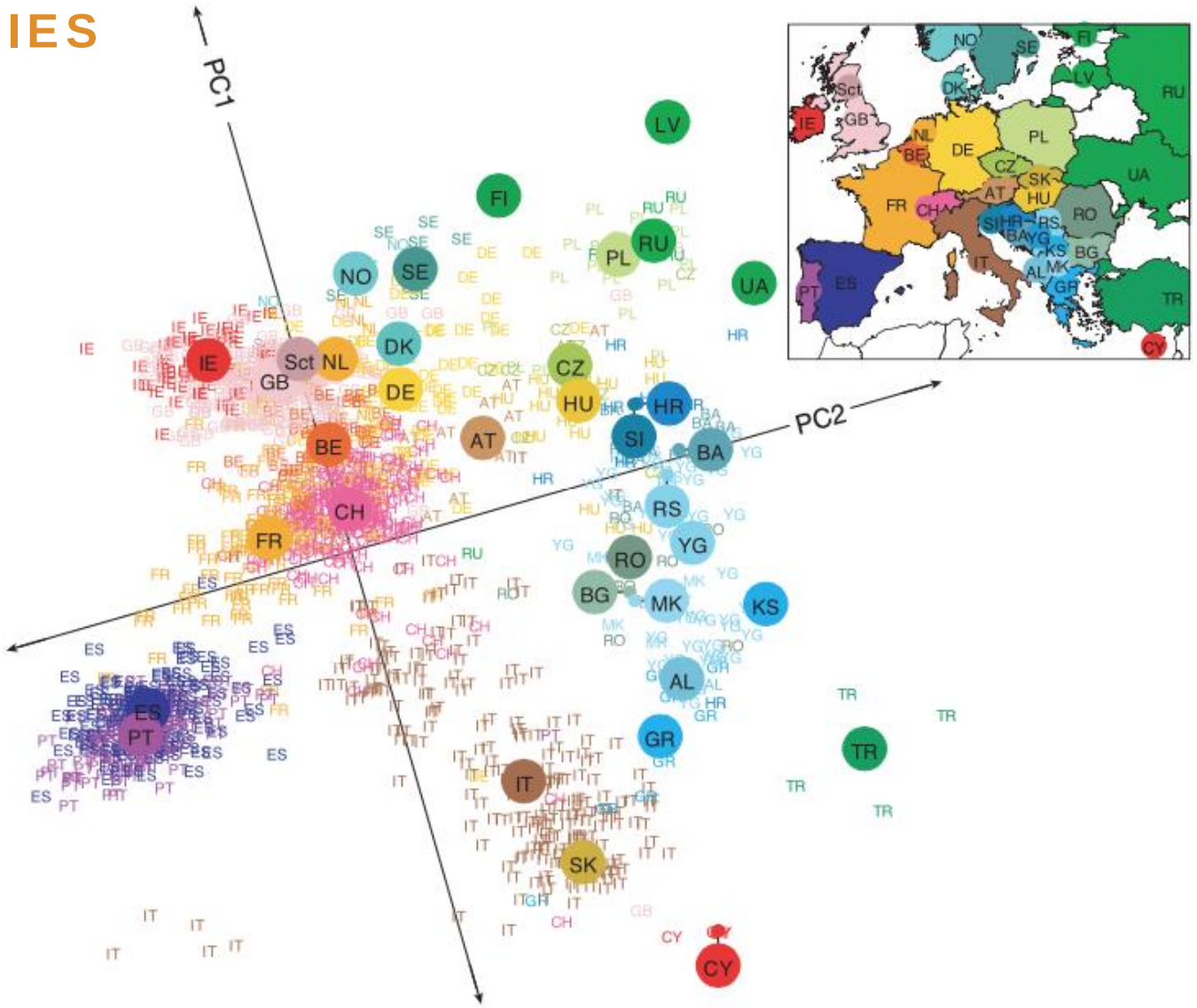


ANCESTRY

AFFECTS ALLELE FREQUENCIES

They surveyed genetic variation in a sample of **3,192 European** individuals which were genotyped at 500,568 loci using the Affymetrix 500K single nucleotide polymorphism (SNP) chip.

"*Our main result holds even when we relax nearly all of these stringency criteria, we focus our analyses on genotype data from 197,146 loci in 1,387 individuals (Supplementary Table 2), for whom we have high confidence of individual origins.*"



AALBORG UNIVERSITY

doi:10.1038/nature07331

Computerome HPC
DTU

TOP?

not really...

then how?

... and specilized software designed to handle large genetic data sets

GenomeDK HPC
Aarhus University

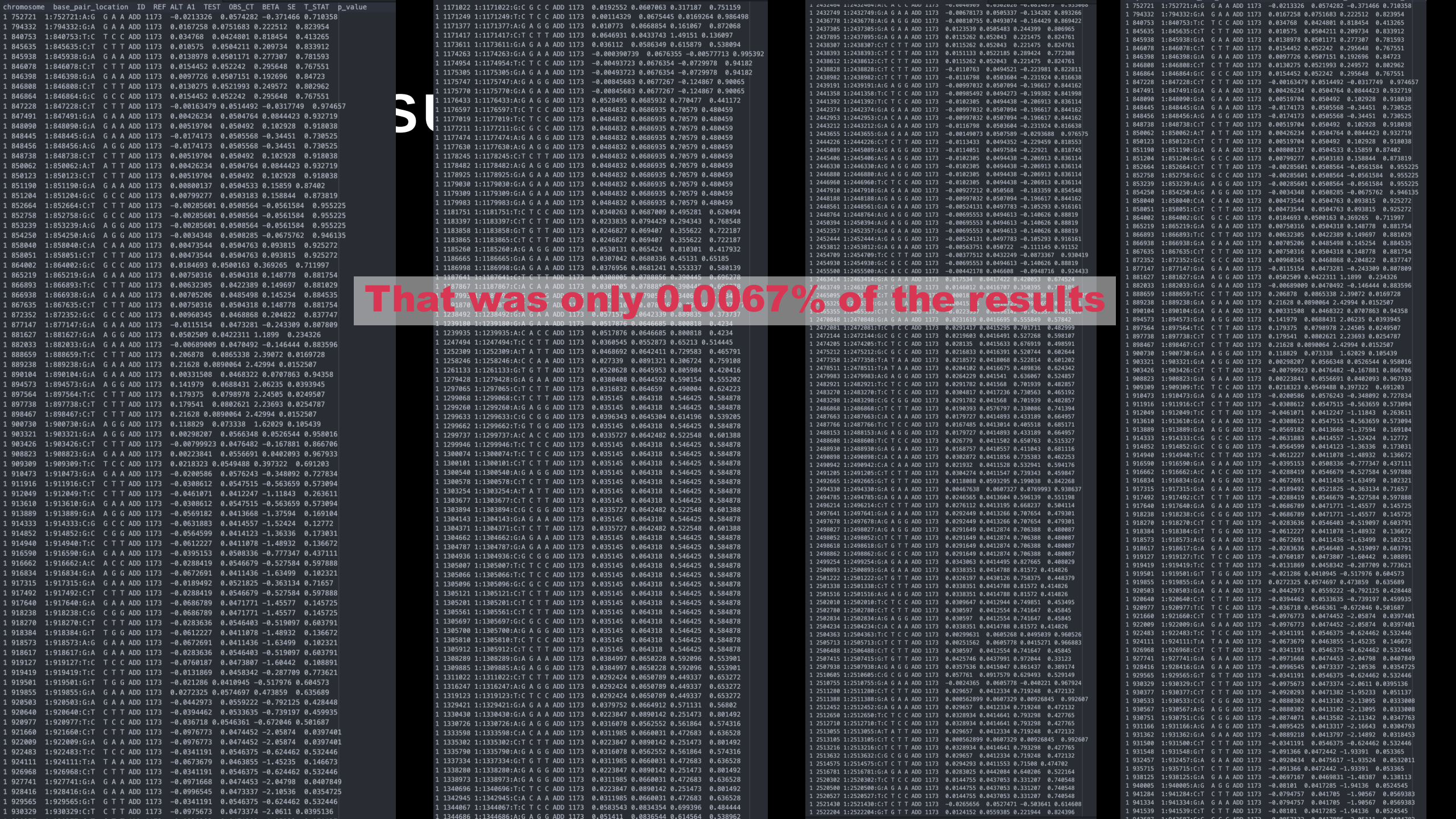
GWAS BY STEPS

4. Summaries 10M linear regressions

3. GWAS

2. Extract genetic variants

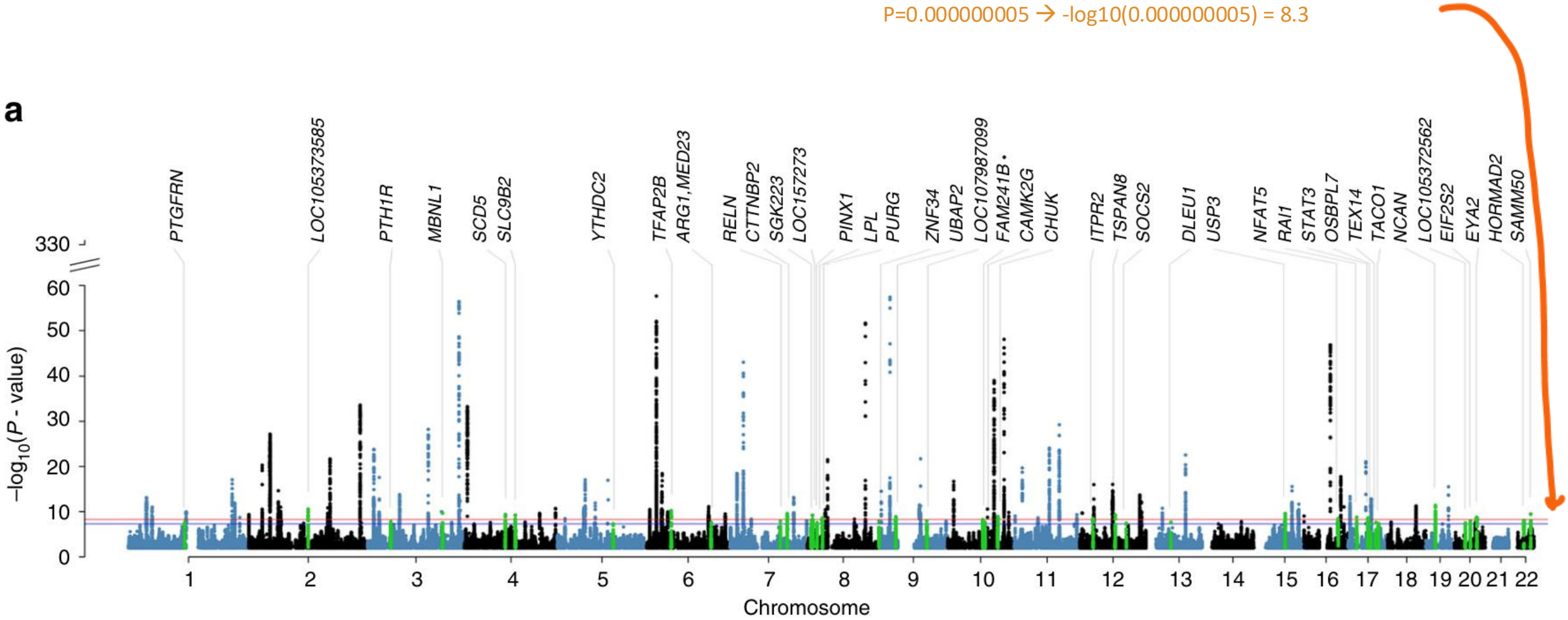
1. Select trait/disease


That was only 0.0067% of the results

MANHATTAN PLOT

For each variant, plot the –log10(P-value) as function of chromosomal position.

P=0.05 → -log10(0.05) = 1.3
P=0.001 → -log10(0.001) = 3
P=0.000000005 → -log10(0.000000005) = 8.3



HYPOTHESIS TESTING

	Null hypothesis (H_0) is true	Null hypothesis (H_0) is false
Reject null hypothesis (H_0)	Type I error α **False positive**	Correct outcome **True positive**
Accecpt null hypothesis (H_0)	Correct outcome **True negative**	Type II error β **False negative**

The probability (*P*) of making a type I error is denoted by α; we reject the null hypothesis if the inferred *P* value is less than the significance level (α=0.05). I.e., the probability of rejecting the null hypothesis when should be accepted.

Why multiple testing correction? If we test 500,000 SNPs, then by chance we expect 25.000 SNPs to be significant (if α=0.05) → i.e., **25.000 false-positive associations**.

One solution is to correct for number of tests performed; Bonferroni correction;

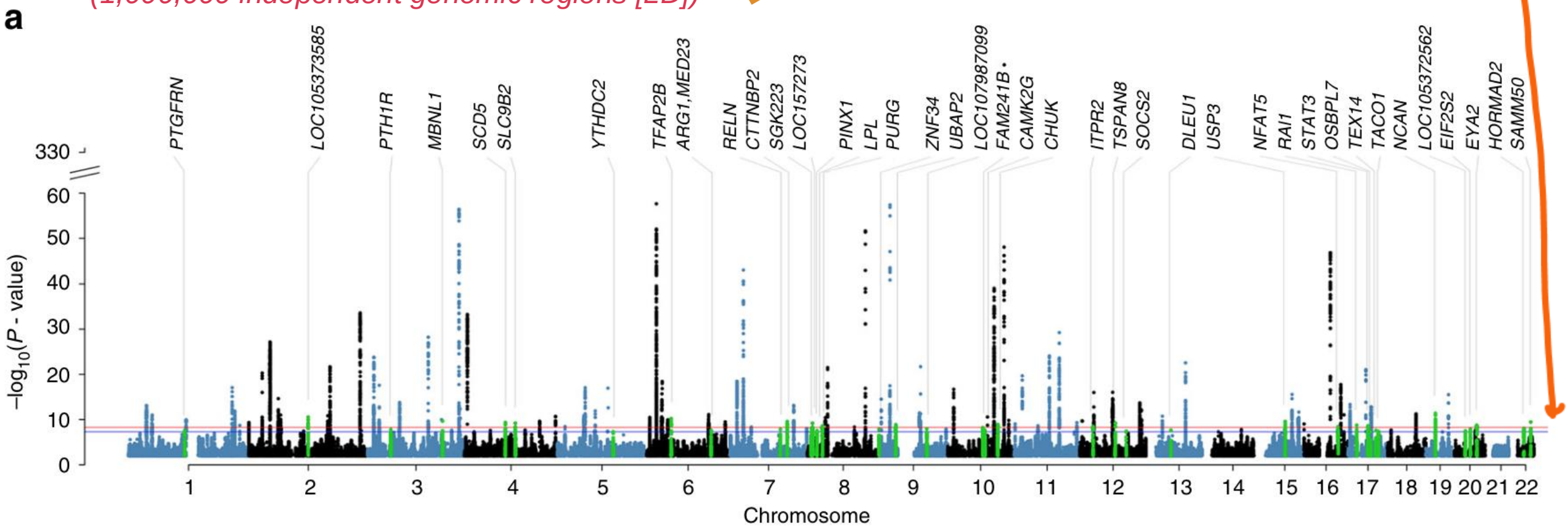
Corrected *P*-value = $P \times n_{tests} \leq 0.05$ OR $\frac{\alpha}{n_{tests}}$ =**new significance threshold**

MANHATTAN PLOT

For each variant, plot the –log10(P-value) as function of chromosomal position.

P=0.05 → -log10(0.05) = 1.3
P=0.001 → -log10(0.001) = 3
P=0.000000005 → -log10(0.000000005) = 8.3

Genome-wide significance level?
-adjust for no. of independent statistical tests (1,000,000 independent genomic regions [LD])

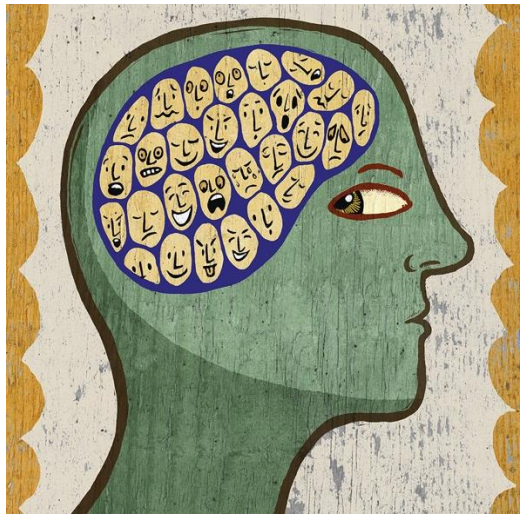


POWER IS EVERYTHING IN GWAS

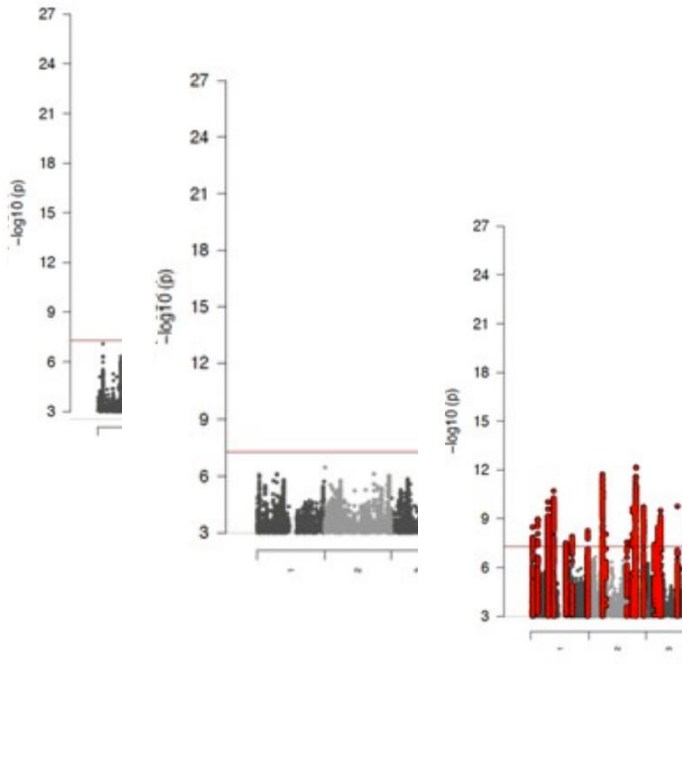
the probability of detecting an effect, if there is a true effect present to detect

depends among otherthings on sample size

Schizophrenia



Has high heritabilitet (h^2= 0.85)
The population prevalence is 1%
Emerge in late teens
Molecular aetiology is unknown



110.000 people
108 genes !

320.000 people
287 genes !

GWAS BY STEPS

5. Find the causal variant

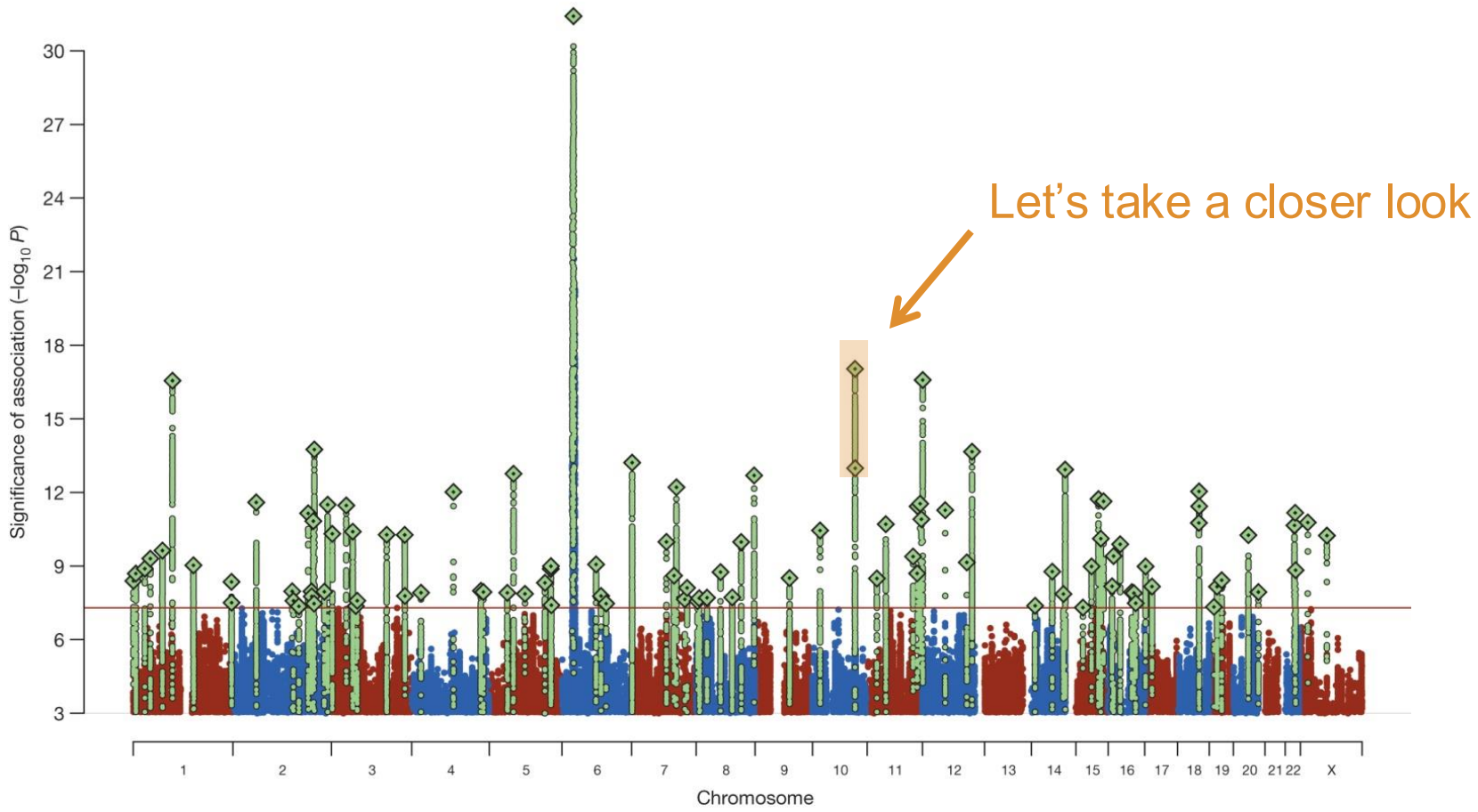
4. Summaries 10M linear regressions

3. GWAS

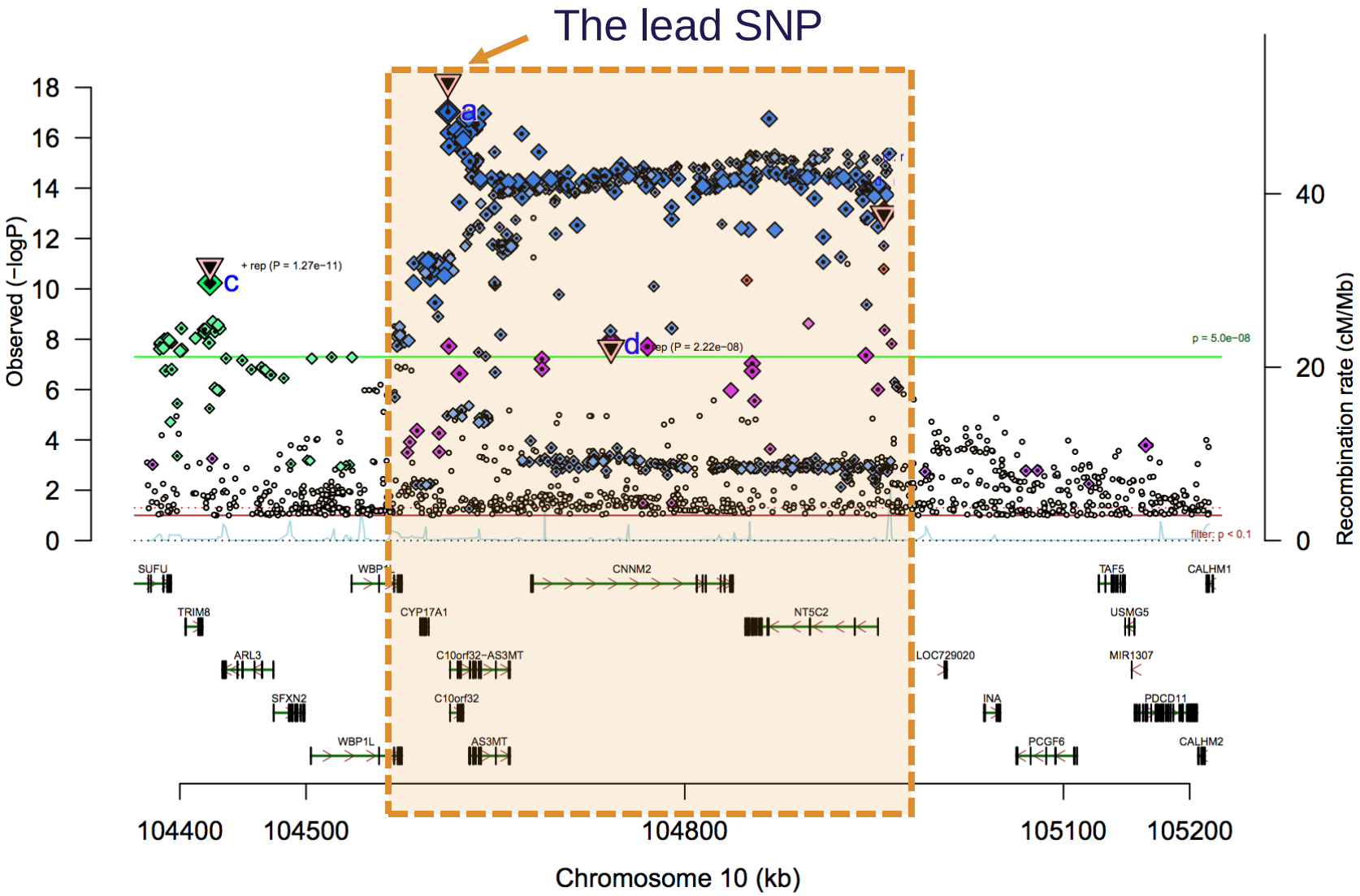
2. Extract genetic variants

1. Select trait/disease

Biological insights from 108 schizophrenia-associated genetic loci



Nature **511**, 421–427 (24 July 2014) | doi:10.1038/nature13595



All the “blue” variants are in LD – which gene is associated with SCZ?

Statistical “Fine mapping” fitting multiple variants at a time [adjust of LD]

FINE MAPPING

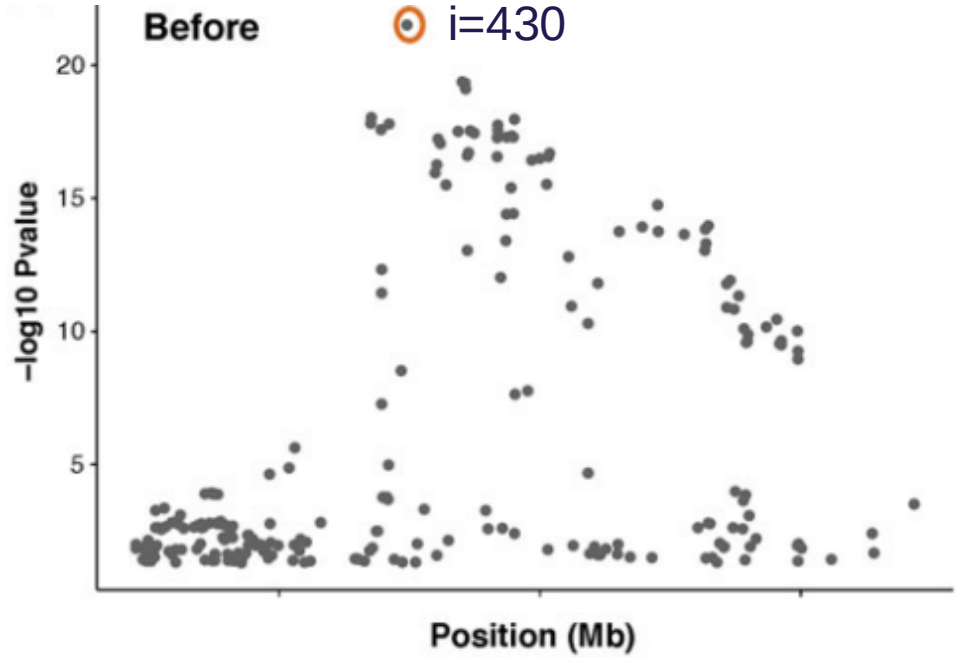
Statistical "Fine mapping" fitting multiple variants at a time [adjust of LD]

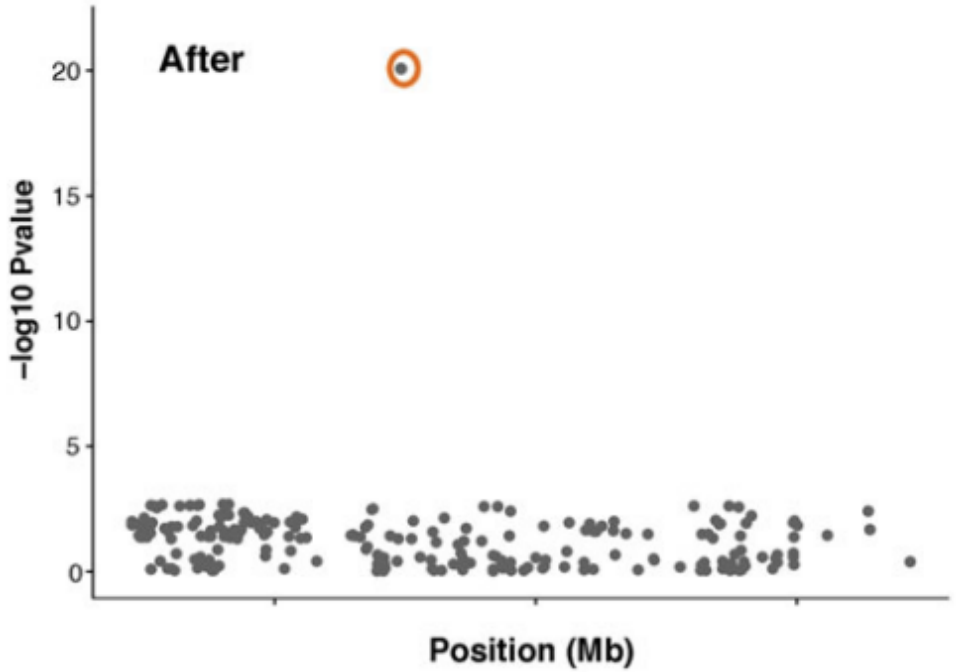
$$y = X_i\beta_i + g + e$$

$$y = X_{i=430}\beta_{i=430} + g + e$$

Condition on index variant

$$y = X_{i=430}\beta_{i=430} + X_{i=431}\beta_{i=431} + g + e$$
$$\vdots$$
$$\vdots$$





GWAS BY STEPS

6. Meta analysis

5. Find the causal variant

4. Summaries 10M linear regressions

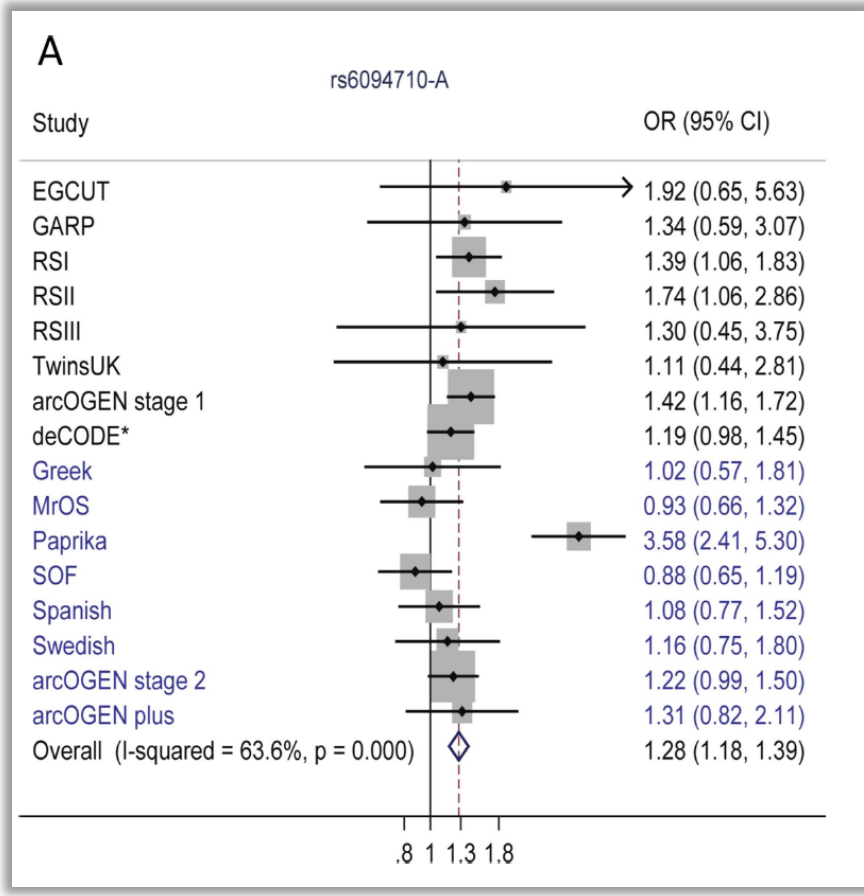
3. GWAS

2. Extract genetic variants

1. Select trait/disease

META ANALYSIS

- Meta-analysis is a set of methods that allows the quantitative combination of data from multiple studies
- Meta-analysis of GWA datasets can increase the power to detect association signals by increasing sample size and by examining more variants throughout the genome than each dataset alone
- Assume two independent estimates $\hat{x}_1 = 1.0$ and $\hat{x}_2 = 2.0$, and that the precision of the first estimate is twice that of the second (precision = $1/SE^2$)
- $\hat{x}_{meta} = \frac{(2\hat{x}_1 + \hat{x}_2)}{2+1} = \frac{2+2}{3} = 1.33$



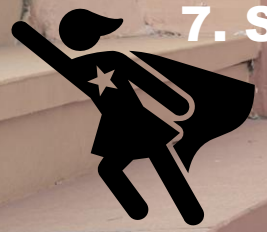
doi:10.1136/annrheumdis-2012-203114

$$\hat{\beta}_{l,F} = \frac{w_{1l}\hat{\beta}_{1l} + \ldots + w_{Kl}\hat{\beta}_{Kl}}{w_{1l} + \ldots + w_{Kl}} \qquad \text{studies } 1,\ldots,K$$

$$\mathrm{SE}_{l,F} = (w_{1l} + \ldots + w_{Kl})^{-\frac{1}{2}}, \qquad \text{where the weight}$$

$$w_{kl} = \frac{1}{\mathrm{SE}_{kl}^2} \text{ is the inverse-variance of study } k.$$

GWAS BY STEPS



7. Spurious associations

6. Meta analysis

5. Find the causal variant

4. Summaries 10M linear regressions

3. GWAS

2. Extract genetic variants

1. Select trait/disease

SPURIOUS ASSOCIATIONS

NEWS & VIEWS

Beware the chopsticks gene

Once upon a time, an ethnogeneticist decided to figure out why some people eat with chopsticks and others do not. His experiment was simple. He rounded up several hundred students from a local university, asked them how often they used chopsticks, then collected buccal DNA samples and mapped them for a series of anonymous and candidate genes.

Molecular Psychiatry (2000) 5, 11–13

Cases

Controls

SPURIOUS ASSOCIATIONS

NEWS & VIEWS

Beware the chopsticks gene

The results were astounding. One of the markers, located right in the middle of a region previously linked to several behavioral traits, showed a huge correlation to chopstick use, enough to account for nearly half of the observed variance. When the experiment was repeated with students from a different university, precisely the same marker lit up. Eureka! The delighted scientist popped a bottle of champagne and quickly submitted an article to *Molecular Psychiatry* heralding the discovery of the 'successful-use-of-selected-hand-instruments gene' (SUSHI).

Are there any problems?

Cases
p=0.50

Controls
p=0.01

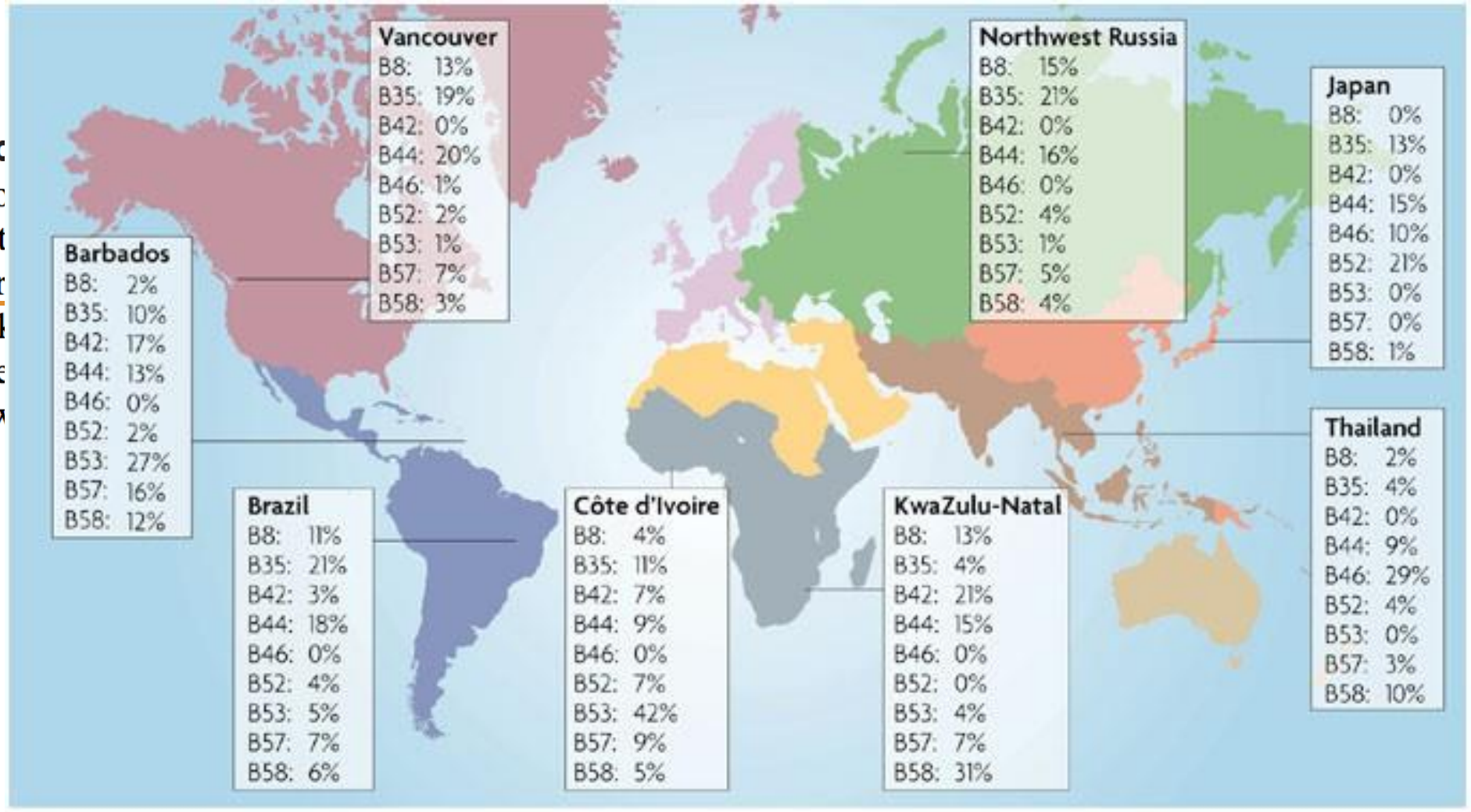
SUSHI gene

SPURIOUS ASSOCIATIONS

NEWS & VIEWS

Beware the chopstic

It took another 2 years tc
antigen gene that has not
to have different allele fr
course differ in chopstick
reasons. Even though the
readily replicated, they w



HLA class I diversity is illustrated by the prevalence of nine HLA-B molecules

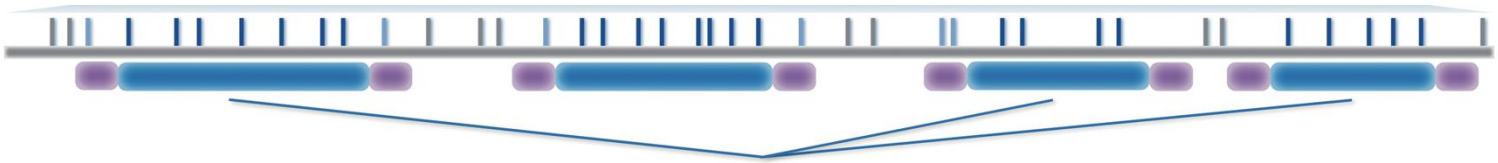
Nature Reviews | Immunology

GWAS BY STEPS

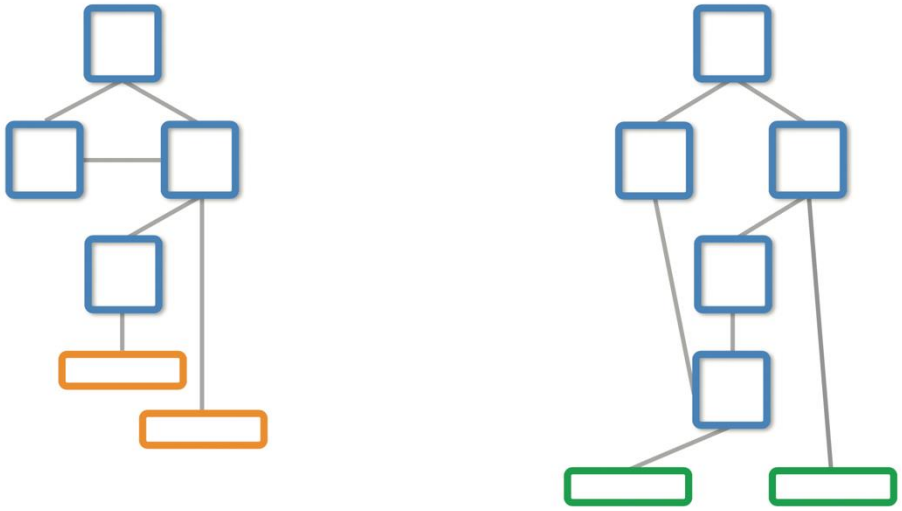
8. Post hoc analyses
7. Spurious associations
6. Meta analysis
5. Find the causal variant
4. Summaries 10M linear regressions
3. GWAS
2. Extract genetic variants
1. Select trait/disease

GWAS POST HOC ANALYSES

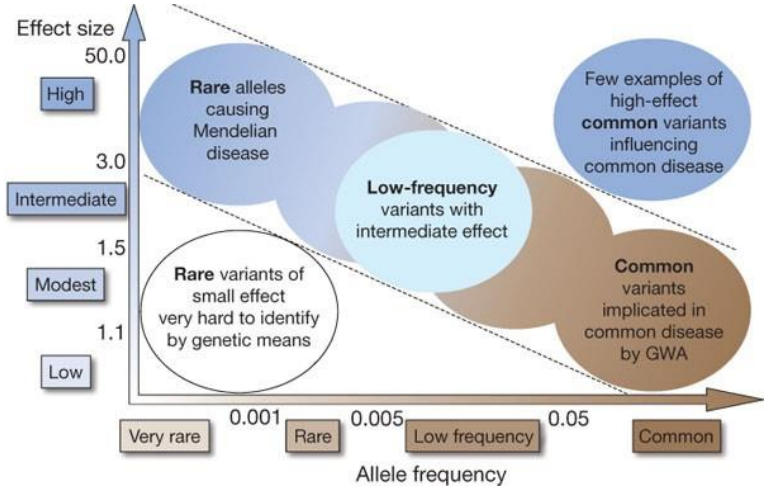
GENE SET ENRICHMENT



Genes (+/- regulatory sequences)
connected in biological pathways and networks



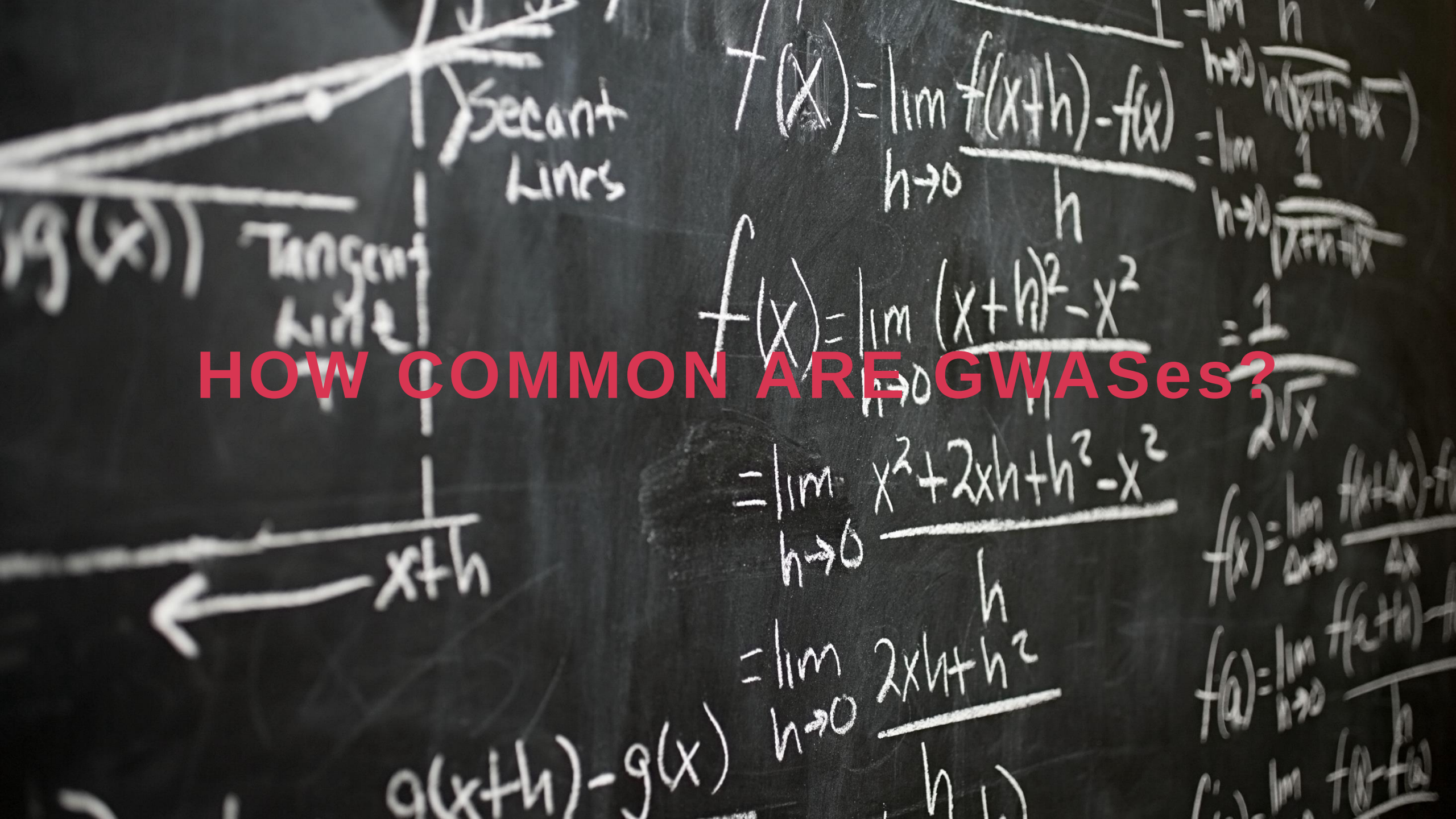
or other newtorks such as
protein interactions or metabolites



Looking for variants with small effects

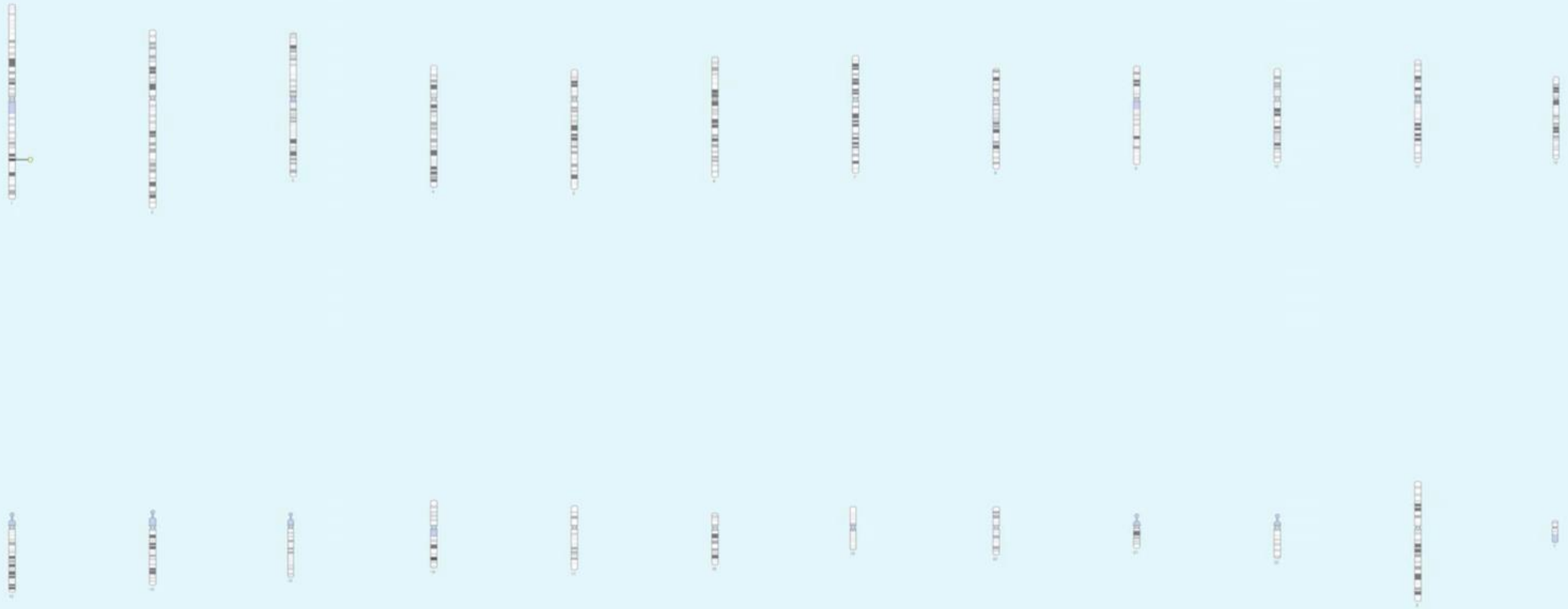
In a GWAS we go through all SNPs one by one.

Gene enrichment analyses we examine whether a group of SNPs (within a biological entity) display a more extreme association signal than by chance.

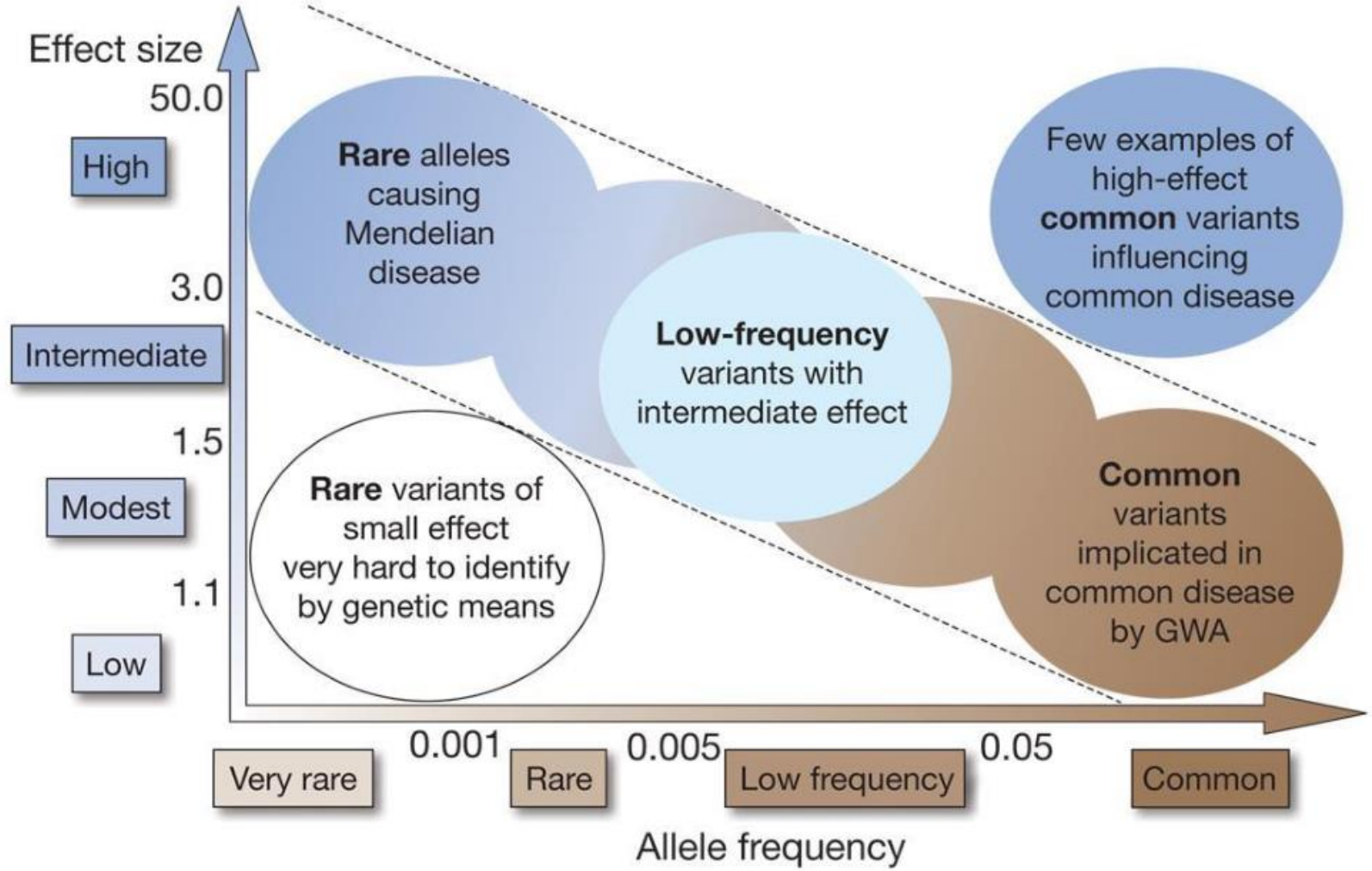


HOW COMMON ARE GWASes?

2006 Jan

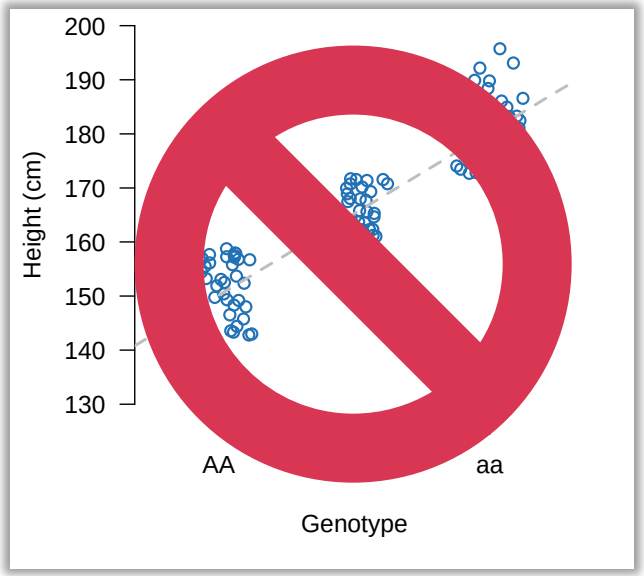


www.ebi.ac.uk/gwas

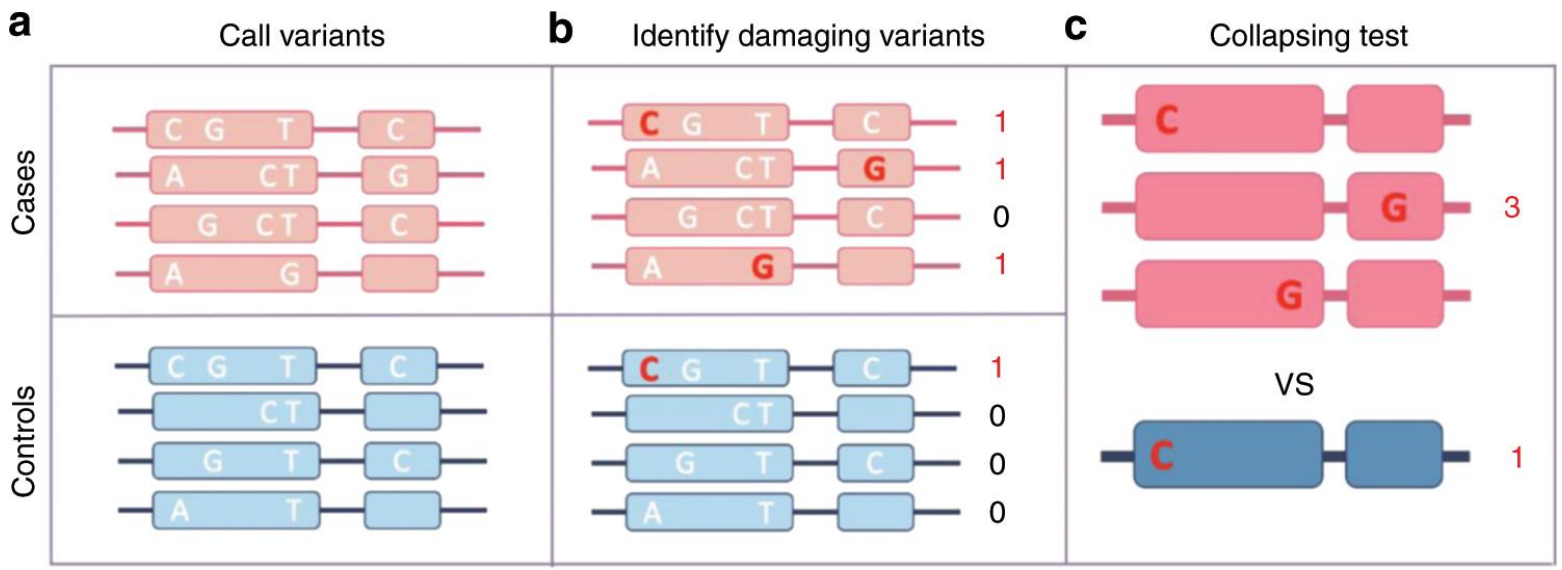


What about rare genetic variants for common diseases?

RARE VARIANT TESTING



If variation at the locus is rare association testing is not possible (no aa or few Aa individuals exists)

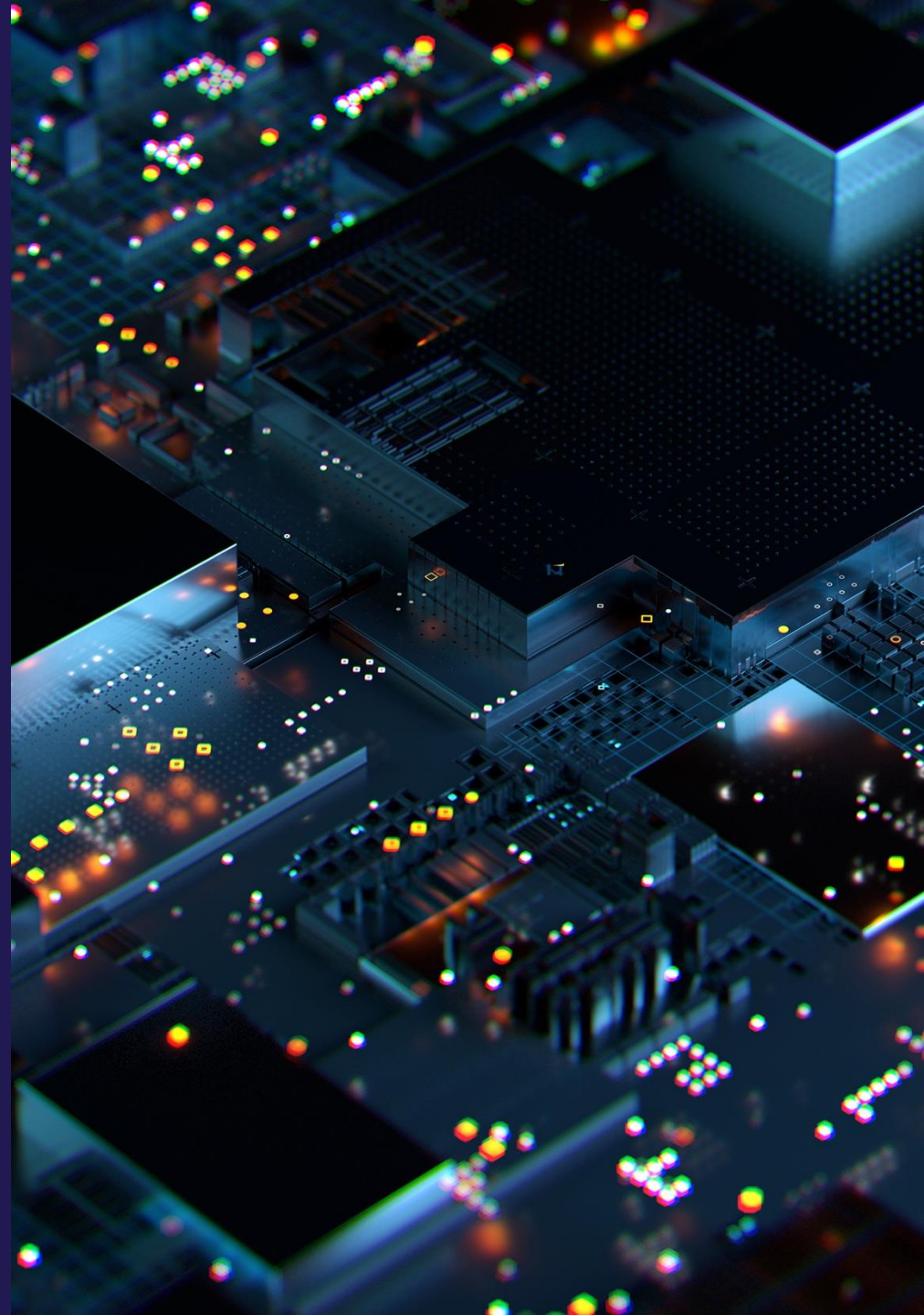


Test the “burden” of all rare variants within a gene.

Often variants are group by functional consequence (missense, pLOF etc)

GENOME-WIDE ASSOCIATION ANALYSIS (GWAS)

- What is a GWAS
- LD
- GWAS by steps



GROUP WORK

GWAS AND FUTURE DIRECTION

1) Go into your ‘complex traits’ group [45 min]
   - ❑ Discuss what did you learn from the GWAS you selected
   - ❑ Read pages 3-10 in the ICDA white-paper
   - ❑ Discuss how the reccomdations could be important for your trait
2) Plenum discuss [10 min]

YOUR OPPINION MATTERS

MOODLE EVALUATION