

Where are the bad guys?



Genome wide association studies

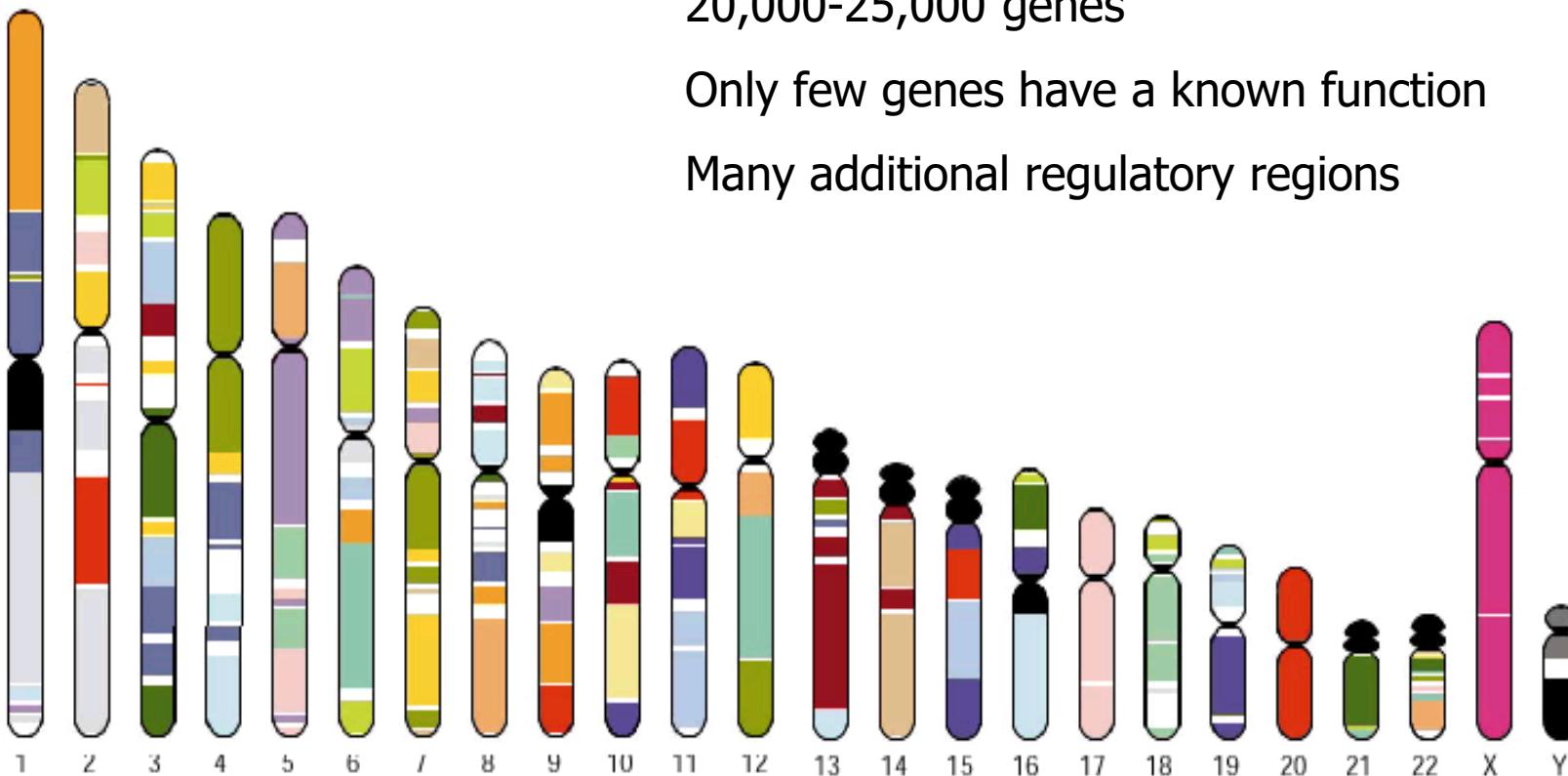
The human genome

~3,300,000,000 base pairs (~2 meter DNA in a single cell)

20,000-25,000 genes

Only few genes have a known function

Many additional regulatory regions

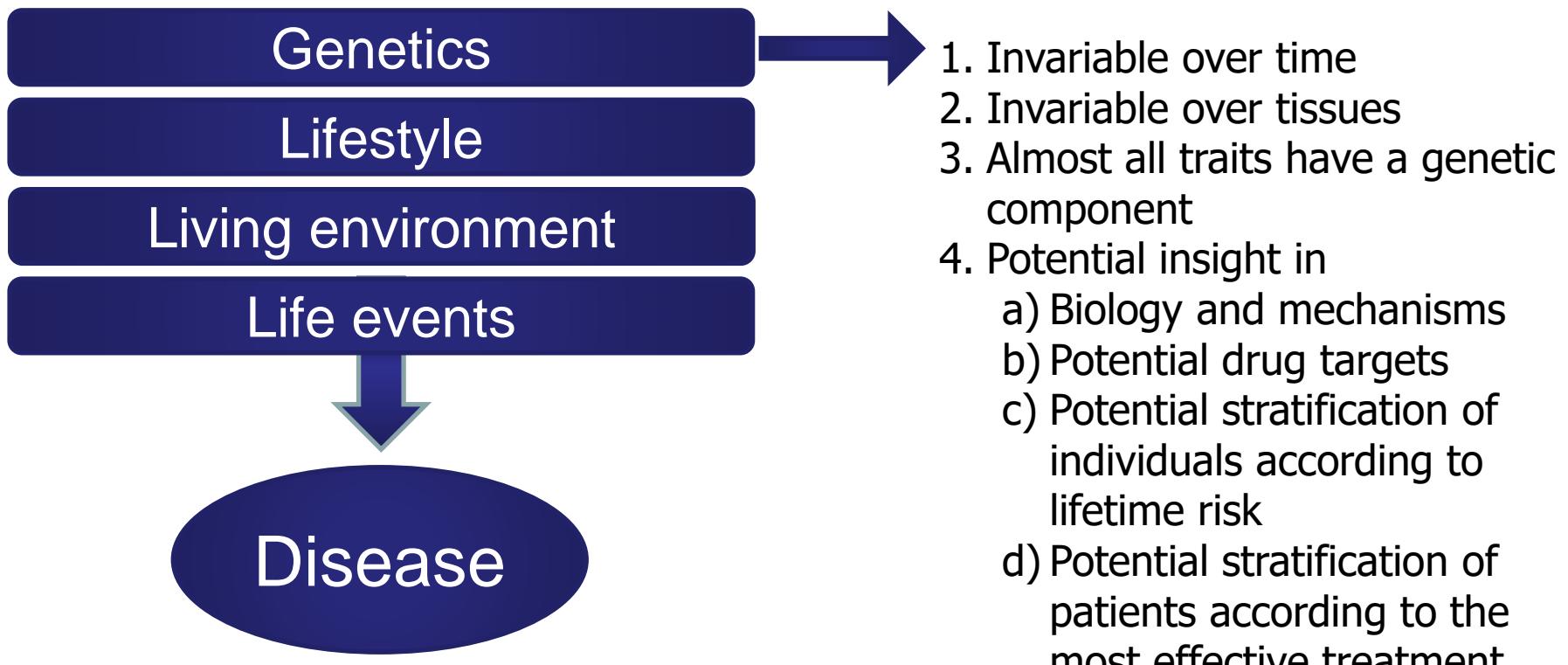


Learning goals

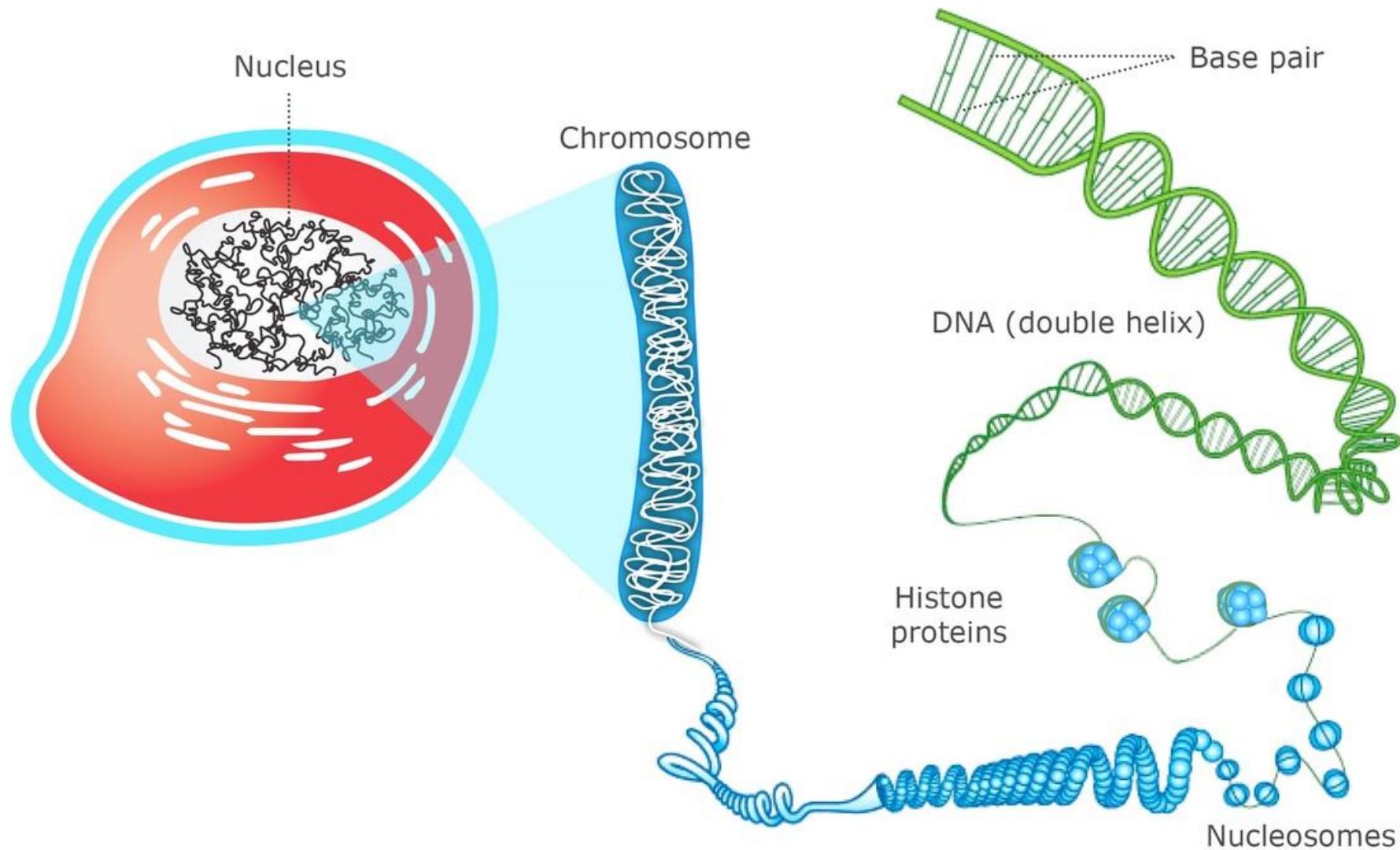
- After my introduction you are able to
 - Explain why genome wide association studies are being performed
 - List the prerequisites of a genome wide association study
 - Design a genome wide association study for an arbitrary trait



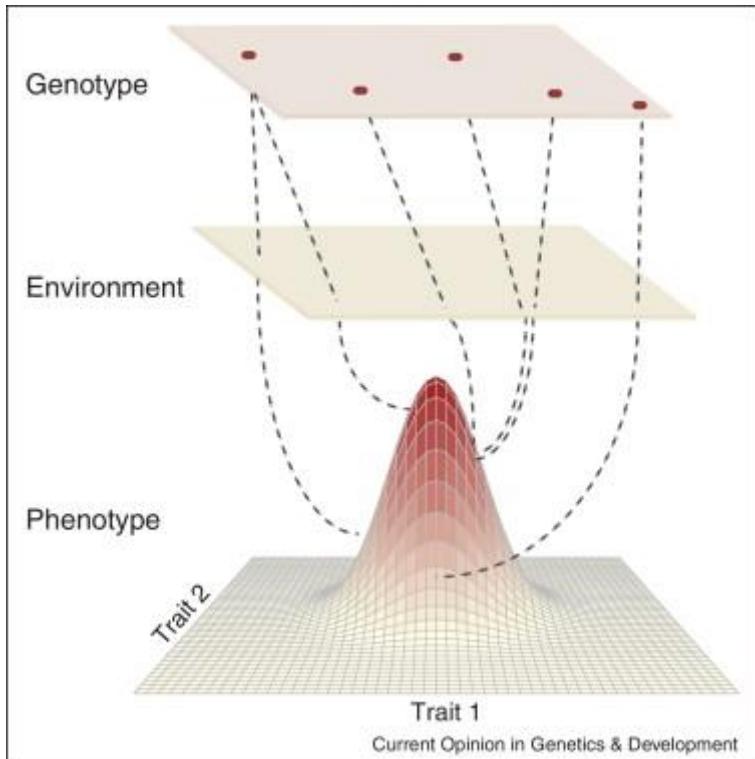
Disease mechanisms



Desoxyribo Nucleic Acid

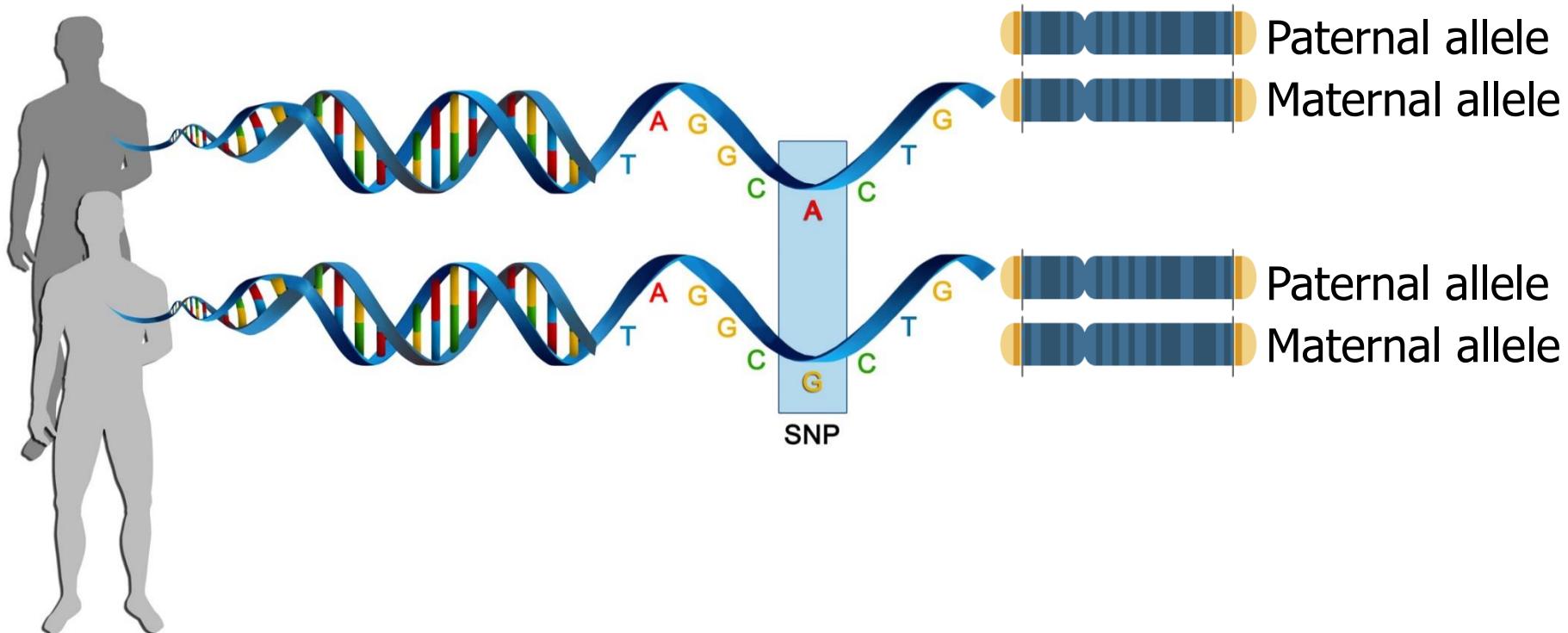


Heritability of a trait



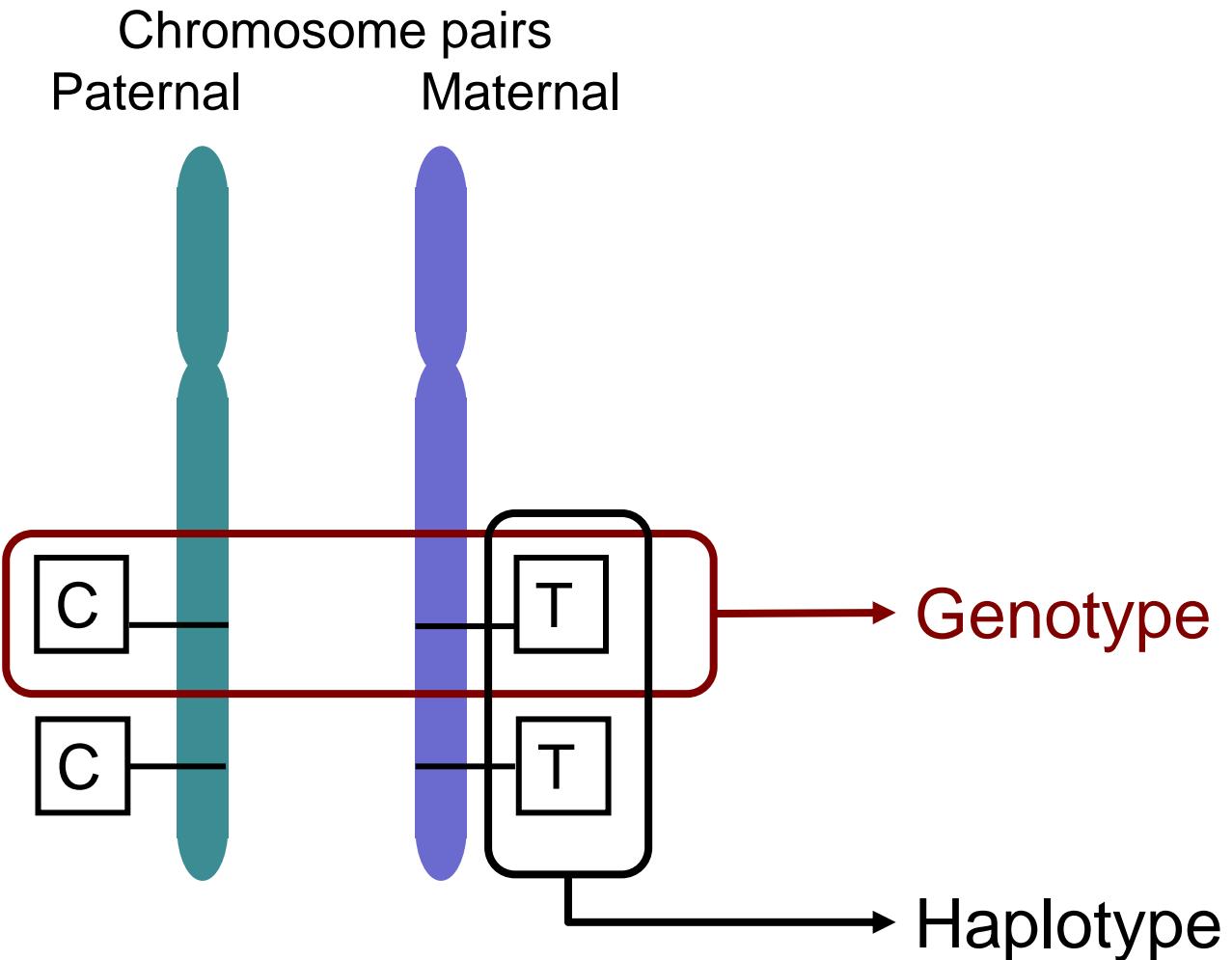
- How large is the contribution of the genotype in a trait?
- The size of the genetic component is not related with the amount of loci.

Single Nucleotide Polymorphism

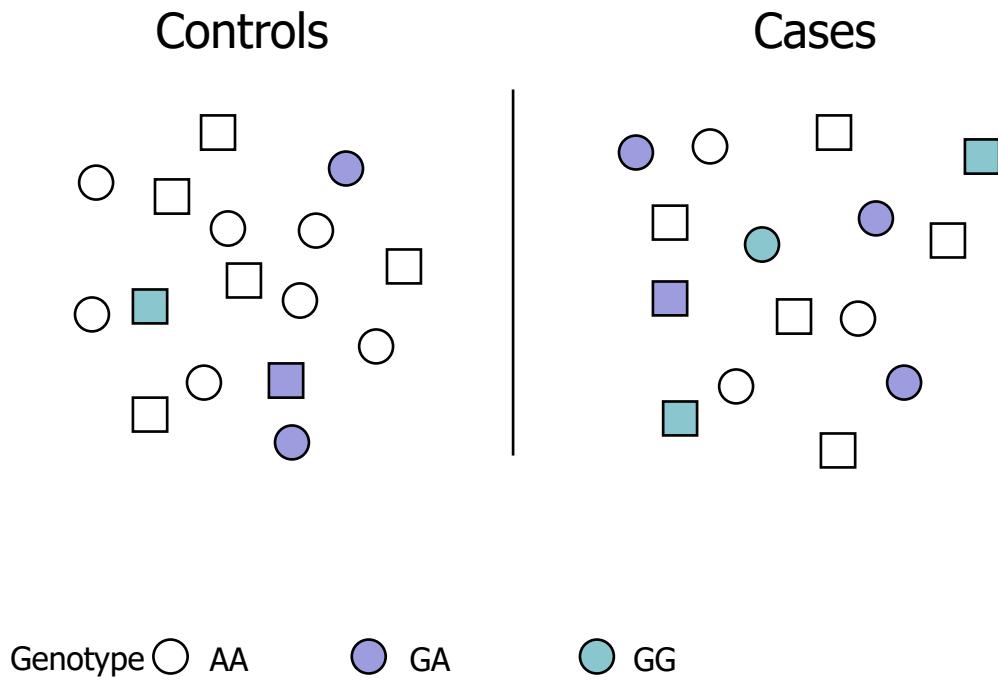


Genetic variation

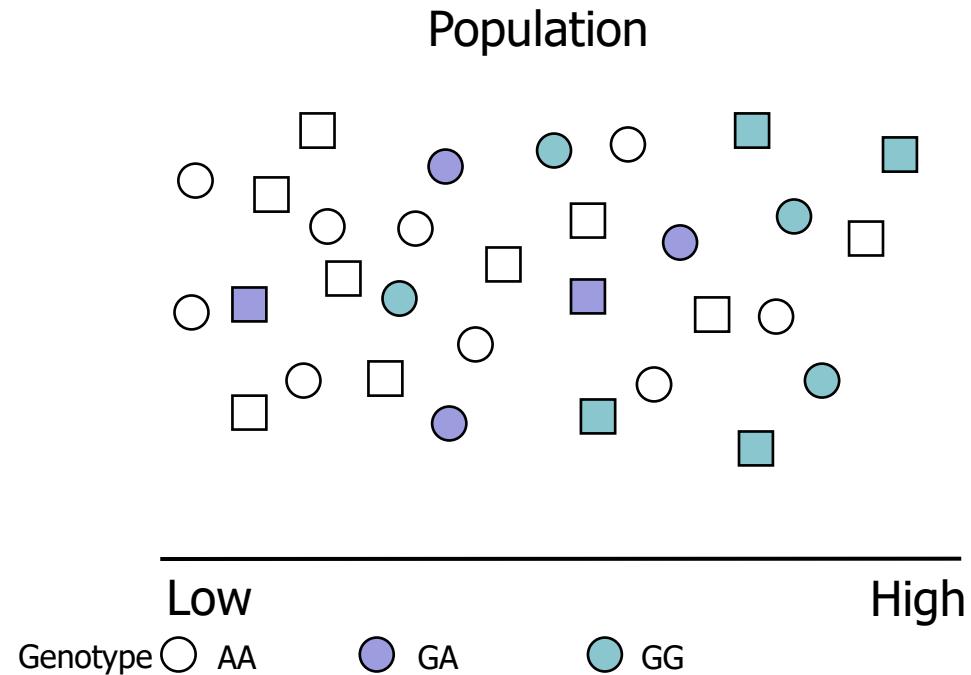
SNP rs429358 19:44,908,684
SNP rs7412 19:44,908,822



Case – Control study (qualitative trait)

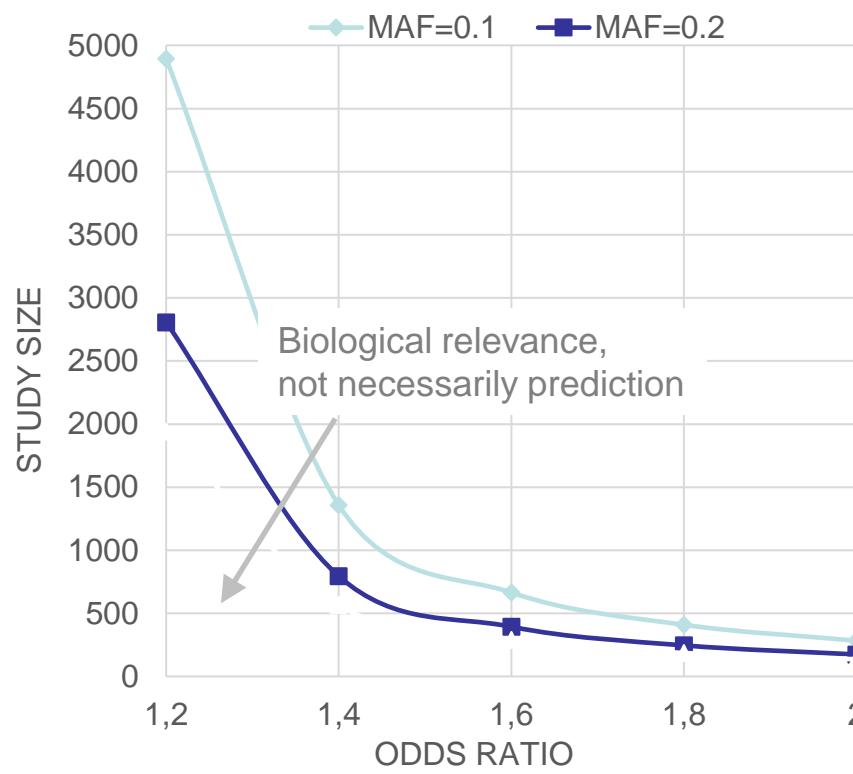


Biomarker study (quantitative trait)



Prerequisites for genetic association study

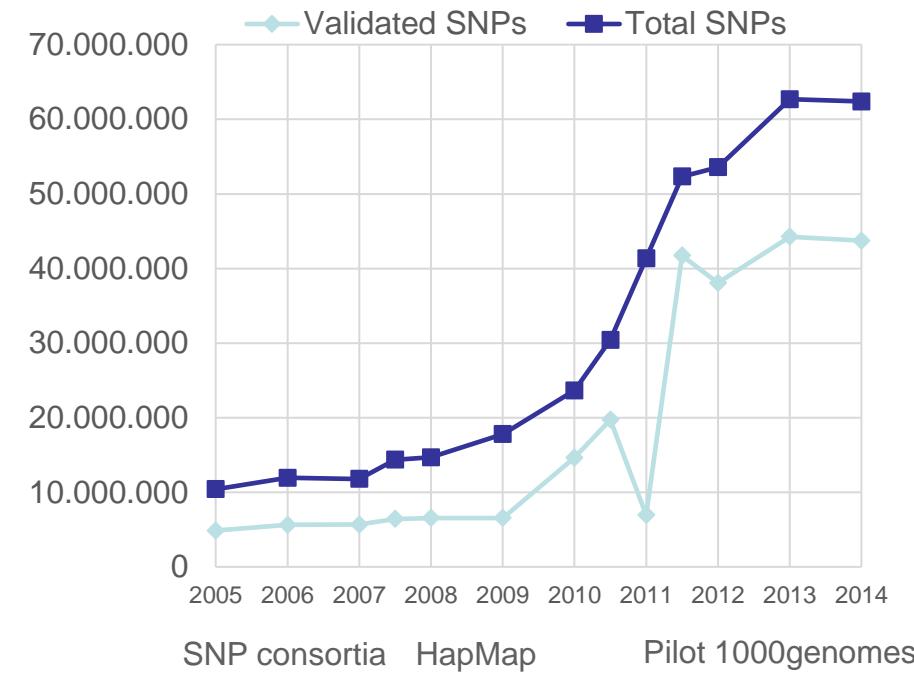
- Trait:
 - Genetic component
- Population
 - Sample size
- Genetic variation
 - Minor Allele Frequency



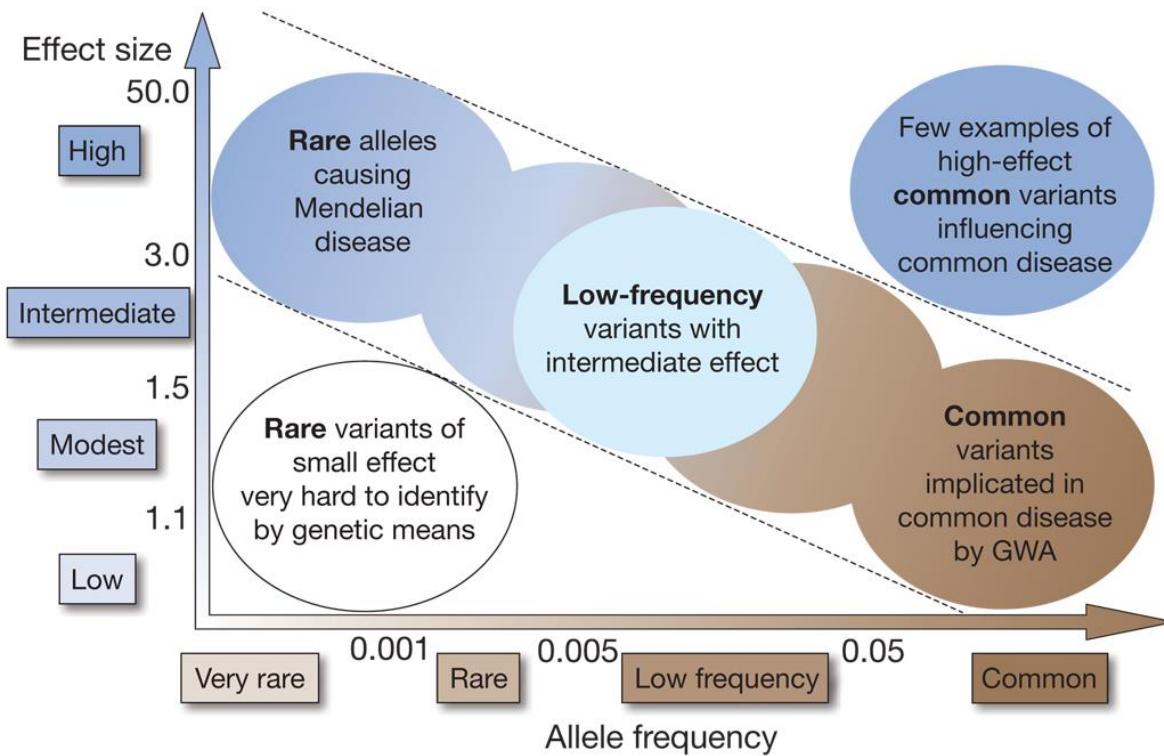
Study sizes assuming
Power=80%, P=0.05

Genome wide genetic association study

- Trait:
 - Genetic component
- Population
 - Sample size
- **Genetic variation**
 - Minor Allele Frequency
 - Genome wide genetic variation



Minor allele frequency and effect size

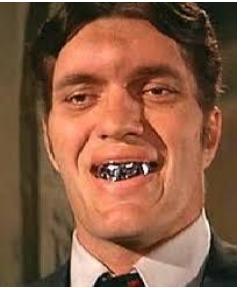


TA Manolio *et al.* *Nature* **461**, 747-753 (2009) doi:10.1038/nature08494

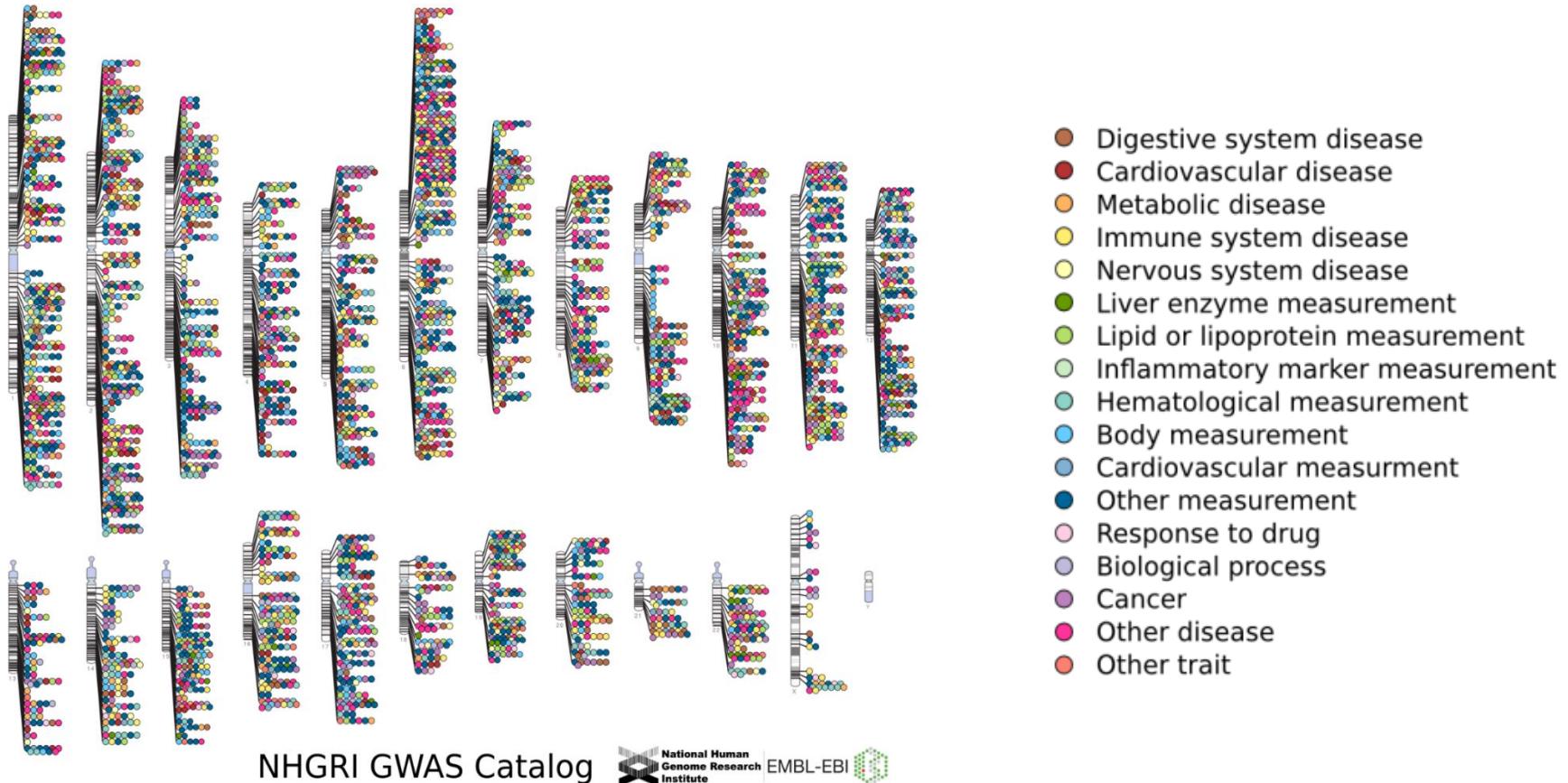
Prerequisites for GWAS

- Trait with an assumed/established genetic component
- Large population in which trait and genetic variation has been measured
 - Formation of large consortia
- Genetic variation
 - Common variants (MAF > 1%)
 - Localization of genetic variation
 - Technology
- Statistics and informatics

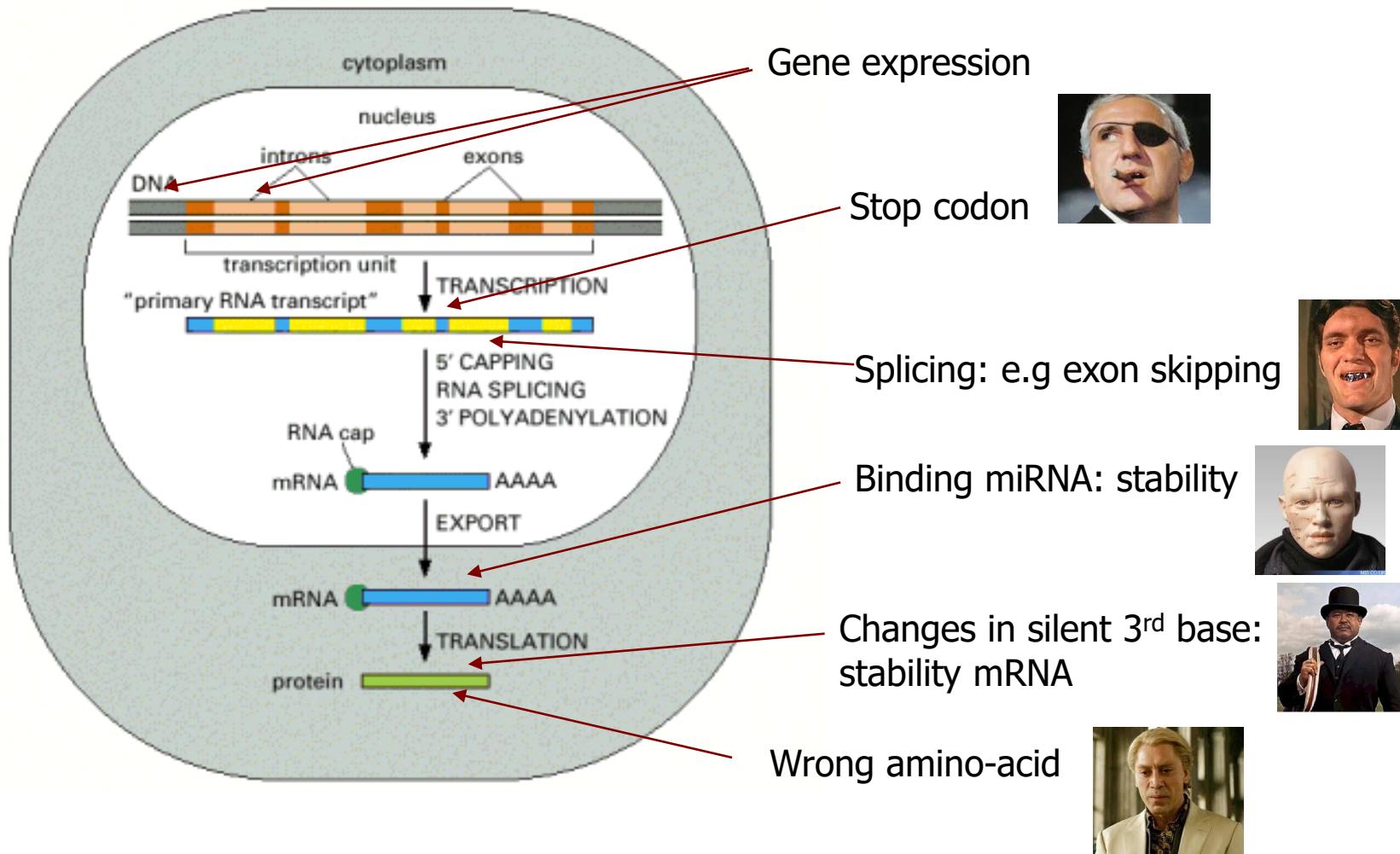
Some bad guys...



...have been identified



Biological effects of genetic variants



Prerequisites for GWAS

- Trait with an assumed/established genetic component
- Large population in which trait and genetic variation has been measured
 - Formation of large consortia
- Genetic variation
 - Common variants (MAF > 1%)
 - Localization of genetic variation
 - Technology
- Statistics and informatics

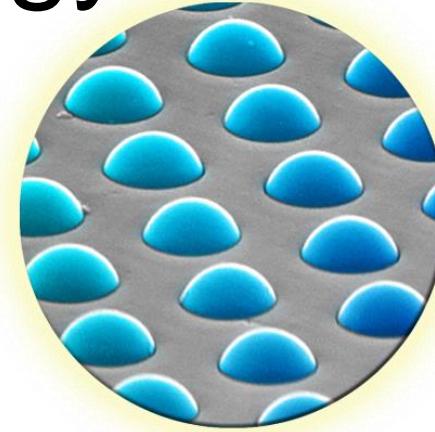
Learning goals

- After the second part you are able to
 - Interpret linkage disequilibrium
 - Explain the difference between D' and R^2
 - Understand the principle of genetic imputation

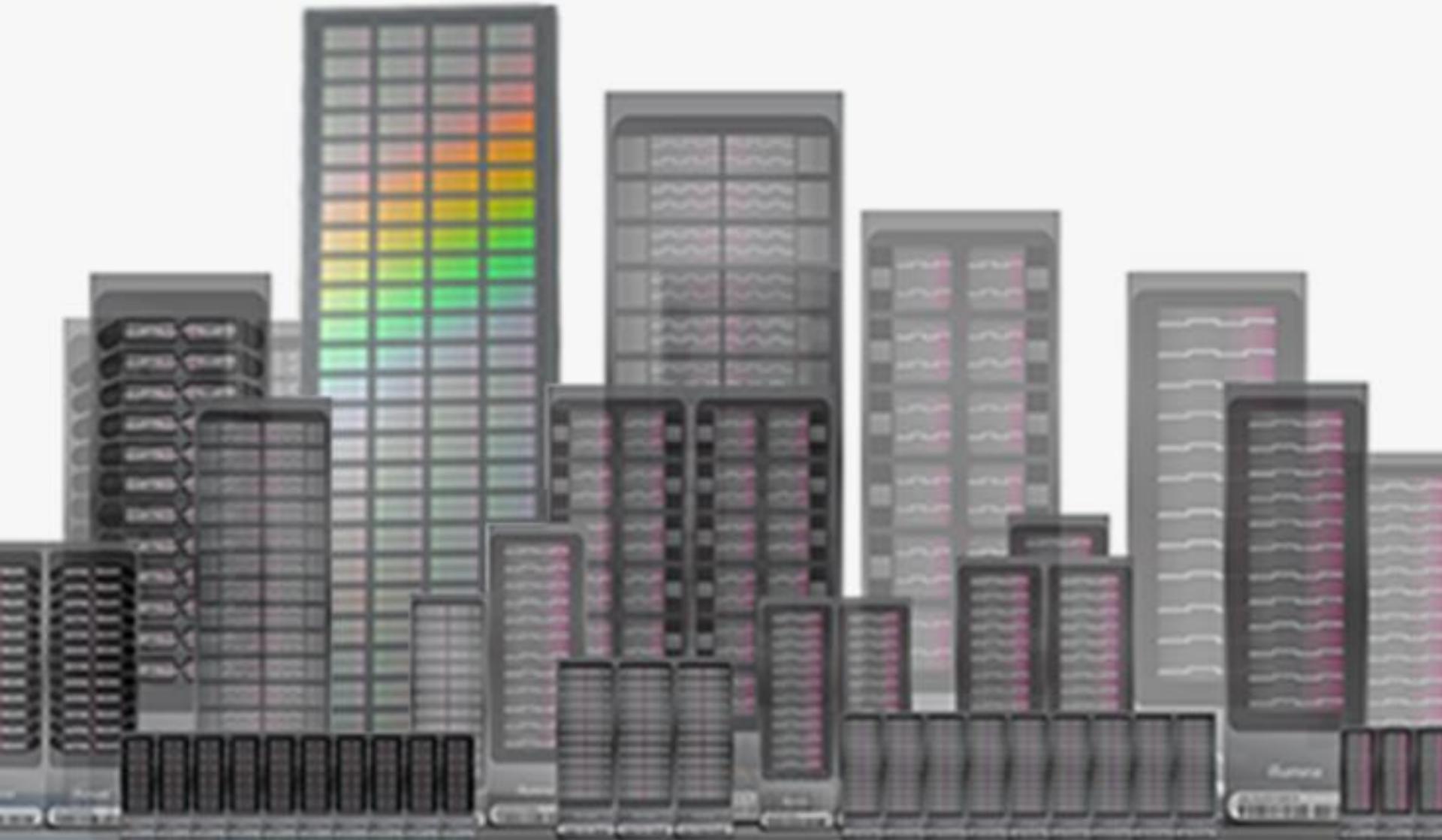


Genotyping technology

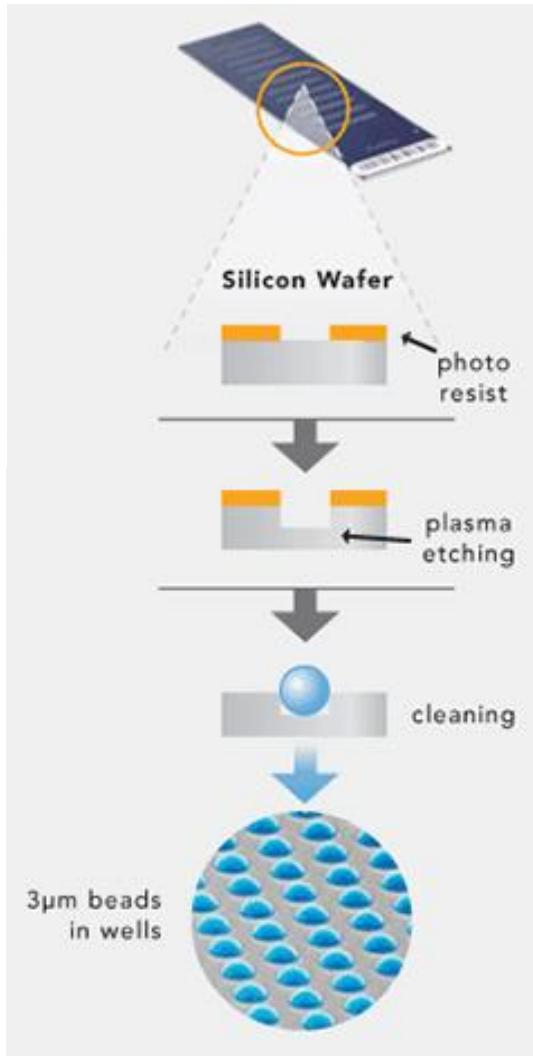
- Illumina Human OmniExpress
 - ~700k common SNPs
 - Copy Number Variations (CNV)
 - Very high data quality (call rate 99.84%)
 - Reasonable throughput (12 samples per chip)
 - Cost ~€ 200-300 per sample all in
- Other Illumina chips up to 2.5M SNPs



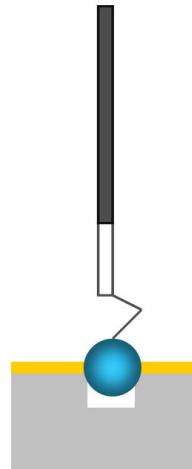
Genetic variation: Genotyping Illumina Infinium Beadchip technology



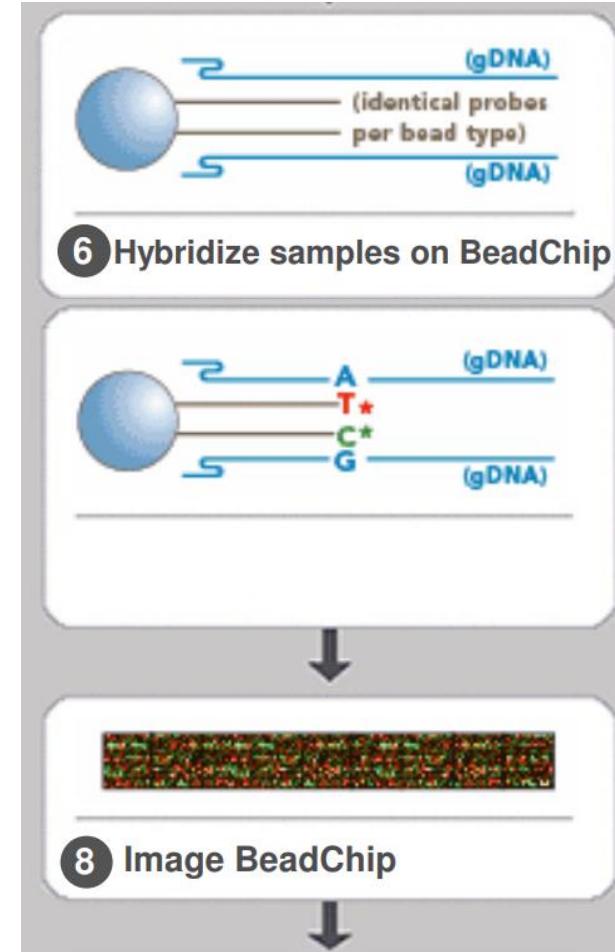
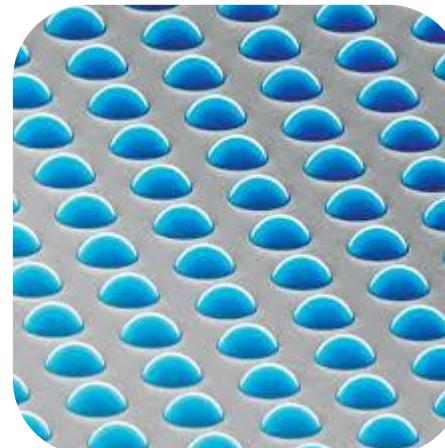
Genetic variation: Genotyping Illumina Infinium Beadchip technology



Specific Probe
(50 base nt)

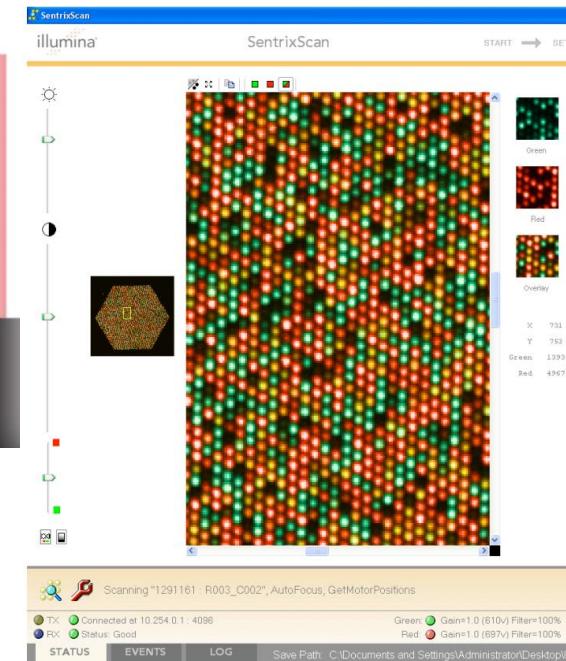
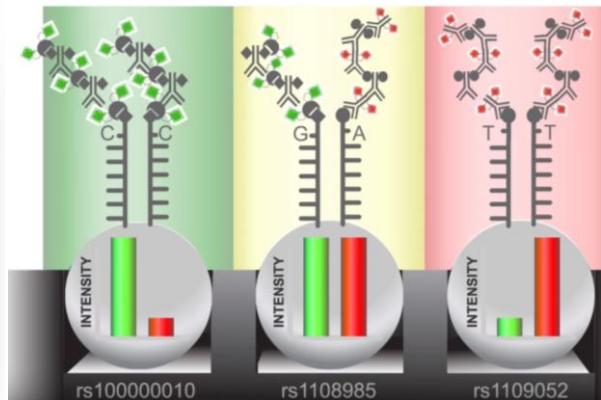
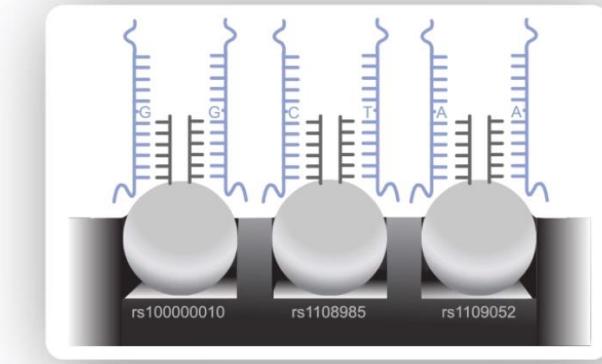


Bead Identifier
(30base nt)



Genetic variation: Genotyping Illumina Infinium Beadchip technology

Each probe binds to a complimentary sequence.



Indirect approach of association

- Test all common genetic variation...
- ...by genotyping a small subset of SNPs only (efficient!)
- Exploit ‘linkage disequilibrium’

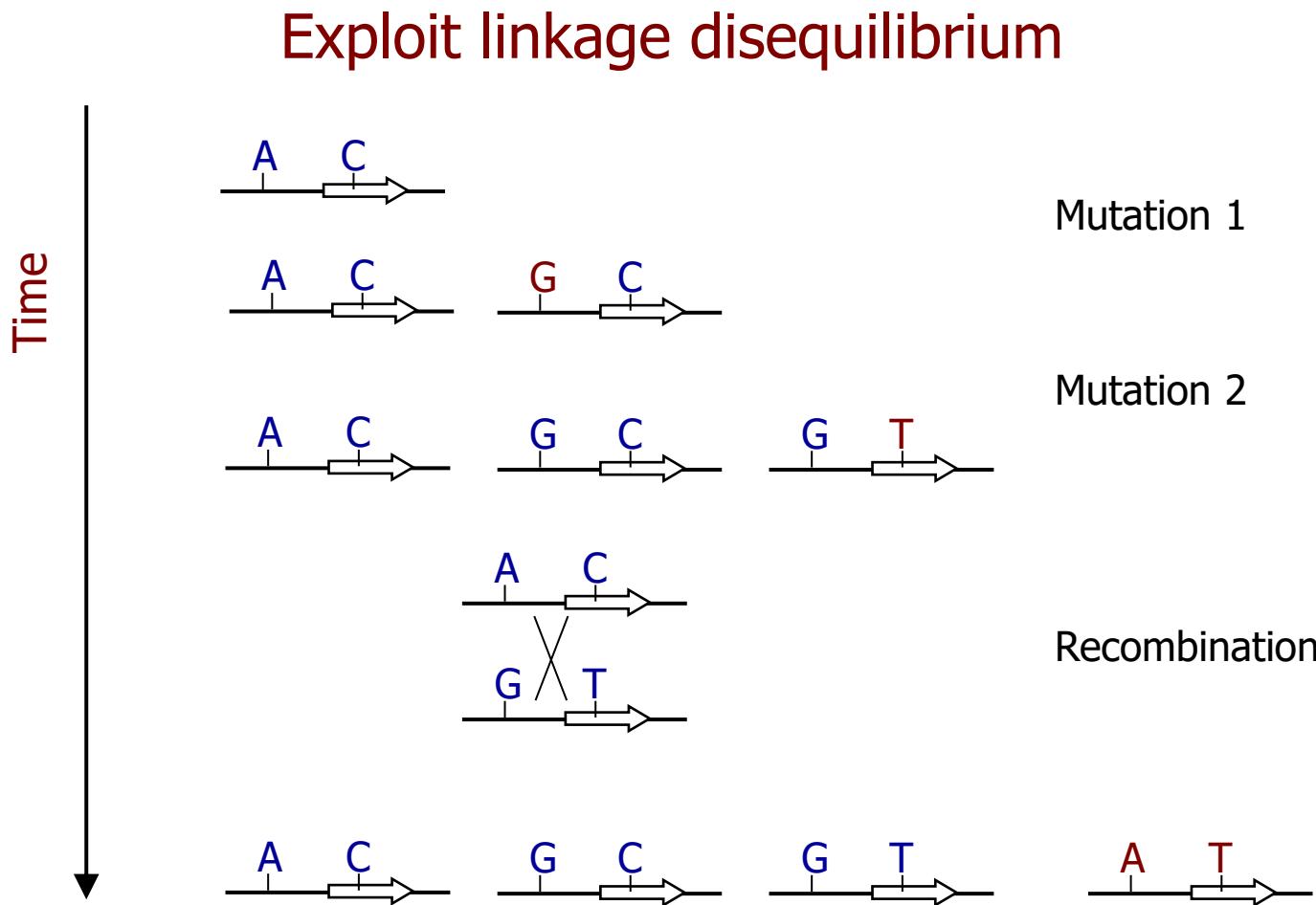


Indirect approach of association

- Two SNPs
 - 1. A/G with MAF=0.40
 - 2. C/T with MAF=0.20
- Expectation combinations
 - 1. A-C: $0.60 \times 0.80 = 0.48$
 - 2. A-T: $0.60 \times 0.20 = 0.12$
 - 3. G-C: $0.40 \times 0.80 = 0.32$
 - 4. G-T: $0.40 \times 0.20 = 0.08$
- Frequently this does not hold for close by SNPs:
→ *DISEQUILIBRIUM*

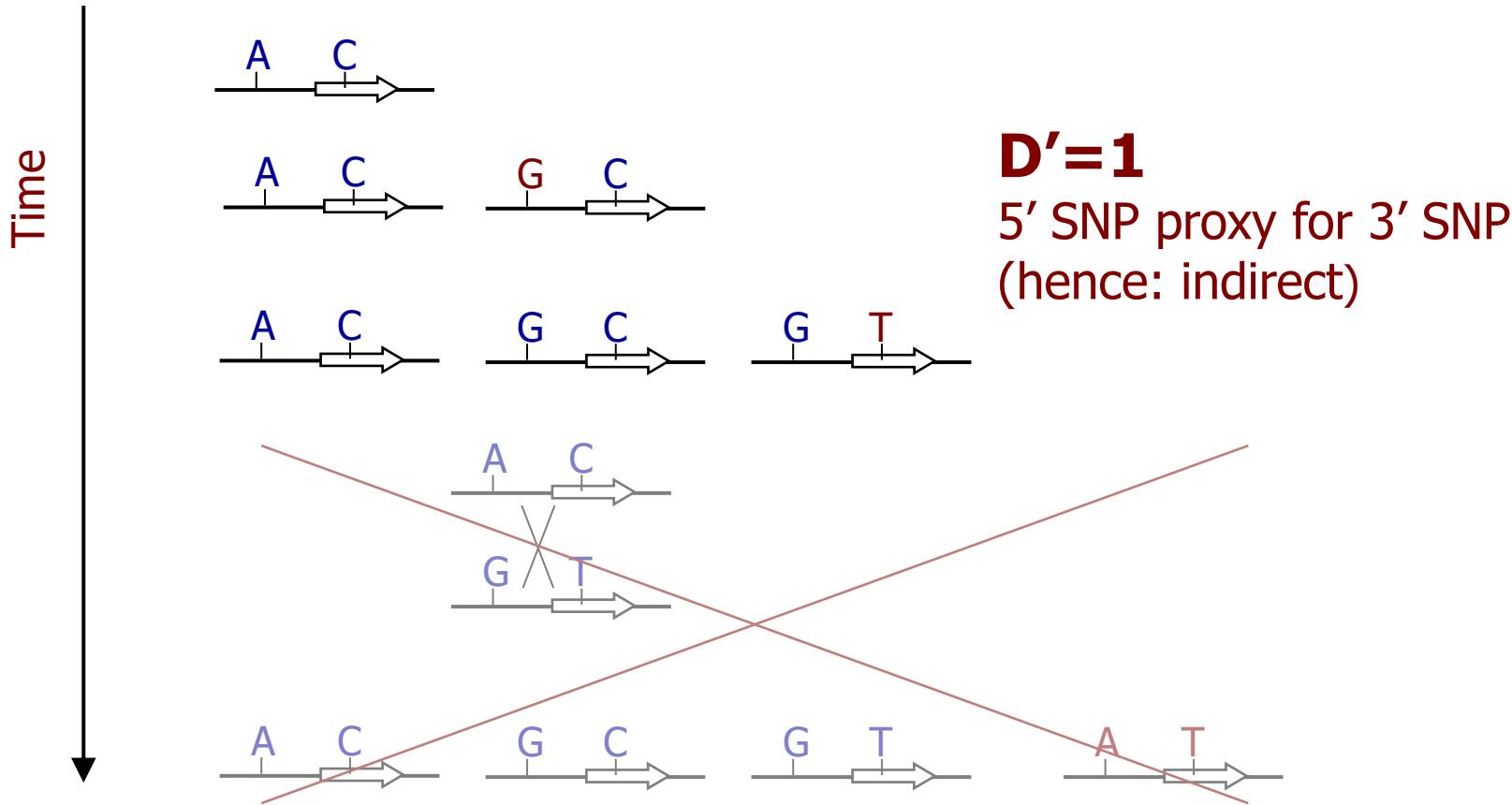


Indirect approach of association



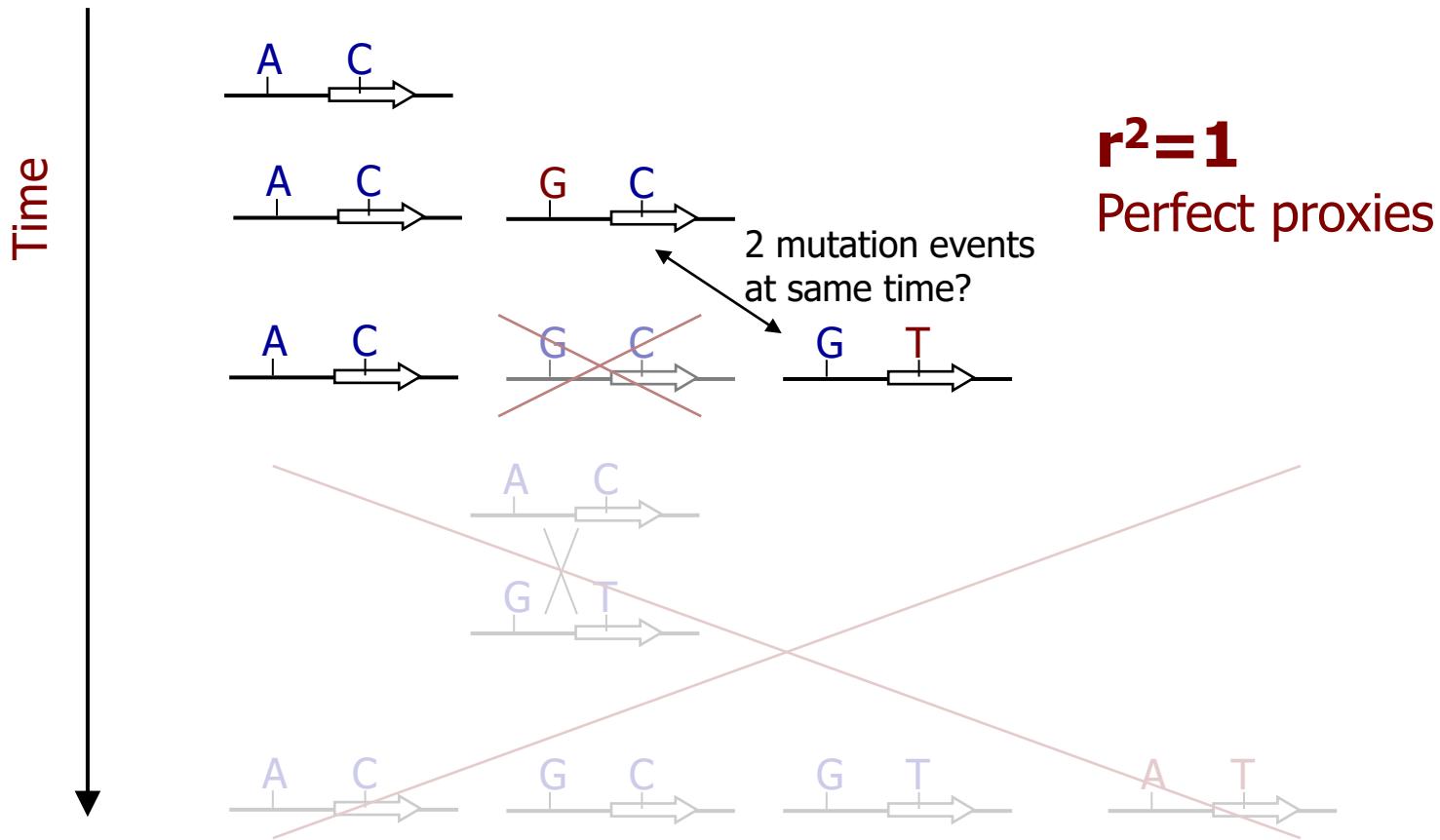
Indirect approach of association

If no recombination:



Indirect approach of association

If also equal allele frequencies:

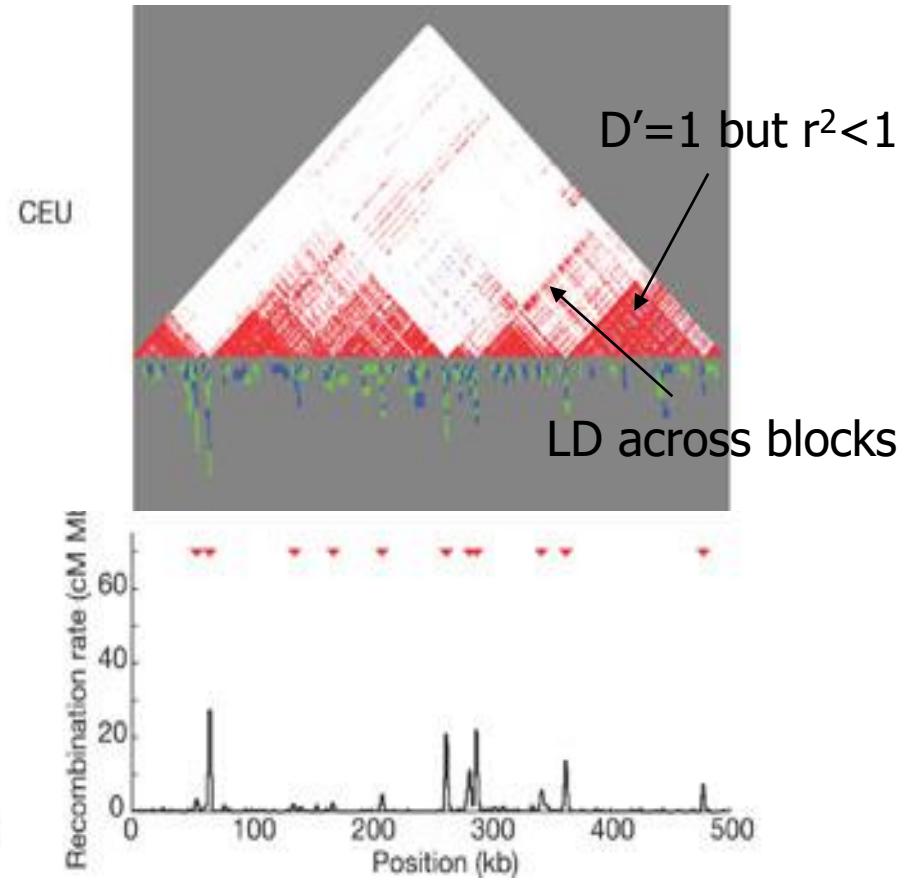
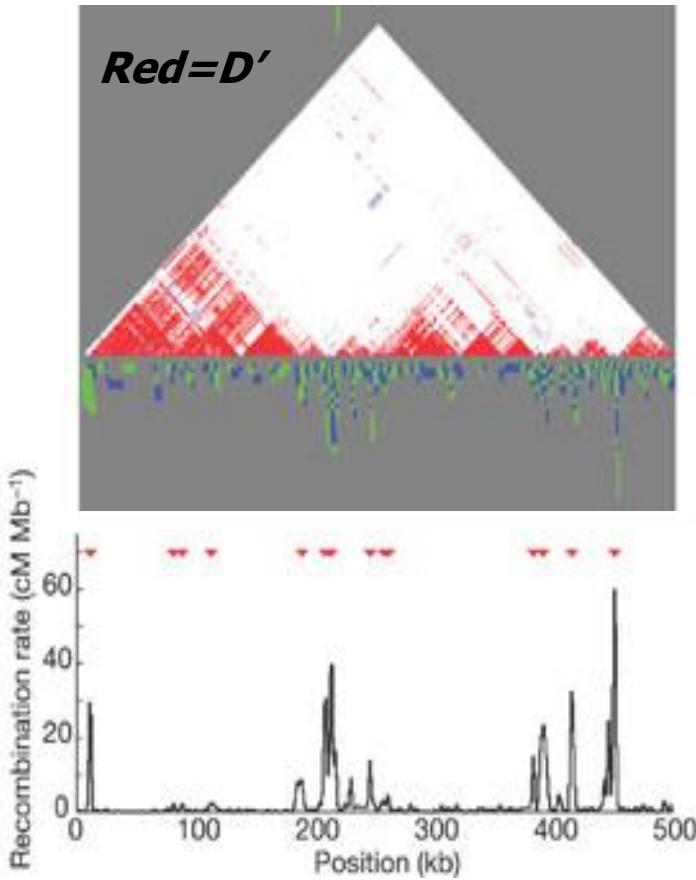


International HapMap project

- Samples HapMap phases 1+2
 - Yoruba, Nigeria (YRI):
 - n=90 (30 parent-offspring trios)
 - Ceph - Utah, USA (CEU):
 - n=90
 - Han Chinese, Beijing (CHB) + Japanese, Tokyo (JPT):
 - n=90
- Genotyping
 - 6,349,188 suspected SNPs assessed
 - 2,819,322 indeed polymorphic and MAF>0.05
 - No resequencing done but ENCODE regions (10 x 5Mb) sequenced as reference



Genetic variation is limited



Imputation Reference Panels

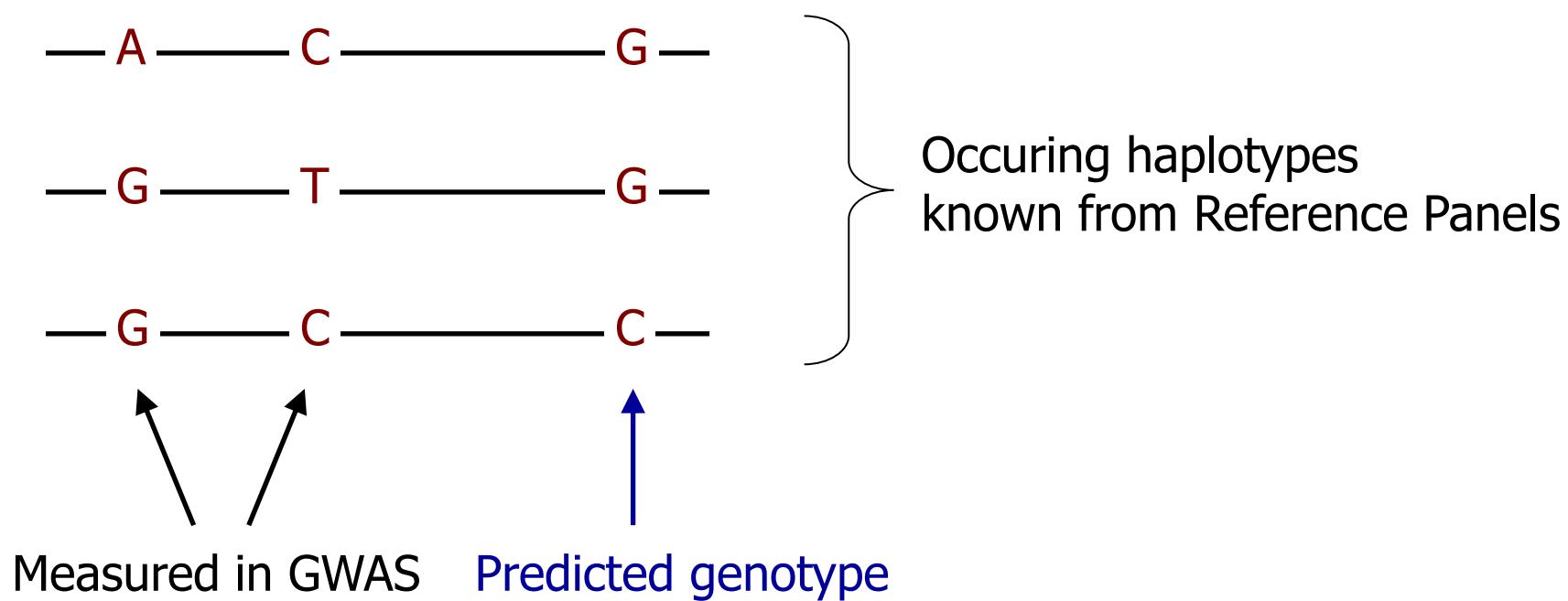
- International HapMap project (2007)
 - CEU (European), YRB (African), JPT/CHB (Asian)
 - 270 samples
 - ~2.5M SNPs
- 1000 Genomes Project (2010)
 - EUR (European), AFR (African), ASN/SAN (Asian), AMR (Americas)
 - 2,535 samples
 - ~30M SNPs + Indels
- The Haplotype Reference Consortium (2015)
 - Mainly EUR ancestry
 - 32,611 samples
 - 39.2 M SNPs



The Haplotype Reference Consortium

Genotype imputation

- Prediction of missing genotypes using LD



Genotype imputation

STAGE ONE

STUDY SAMPLE

.... A A A ...
.... G C A ...

REFERENCE HAPLOTYPES

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

STAGE TWO

STUDY SAMPLE

.... A A A ...
.... G C ... A ...

REFERENCE HAPLOTYPES

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

STAGE THREE

STUDY SAMPLE

c g a g A t c t c c c g A c c t c t g g
c g a a G c t c t t t C t t t c A t g g

REFERENCE HAPLOTYPES

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

Genotype imputation

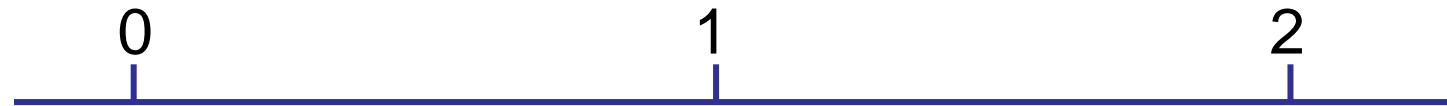
- Imputation software will provide probabilities per genotype
 - probability of the AA genotype = 0.5*
 - probability of the AB genotype = 0.3*
 - probability of the BB genotype = 0.2*
- Best guess genotype => AA
- Genotype dosages (using genotype probabilities)

AA

AB

BB

- Dosage =

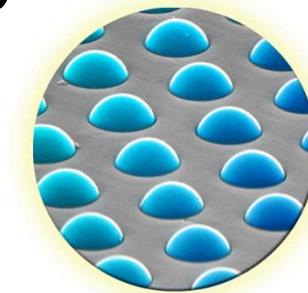


- $= (0 \times p_{AA}) + (1 \times p_{AB}) + (2 \times p_{BB}) \Rightarrow (0 \times 0.5) + (1 \times 0.3) + (2 \times 0.2) = 0.7$



Genetic variation: Genotyping and imputation

- Illumina Global Screening Array (GSA) array
 - ~640k common SNPs
 - Very high data quality (call rate 99.84%)
 - High throughput (24 samples per chip)
 - Cost ~€ 50 per sample all in
- => Haplotype Reference Consortium imputation up to 40 Million SNPs!



HELP!

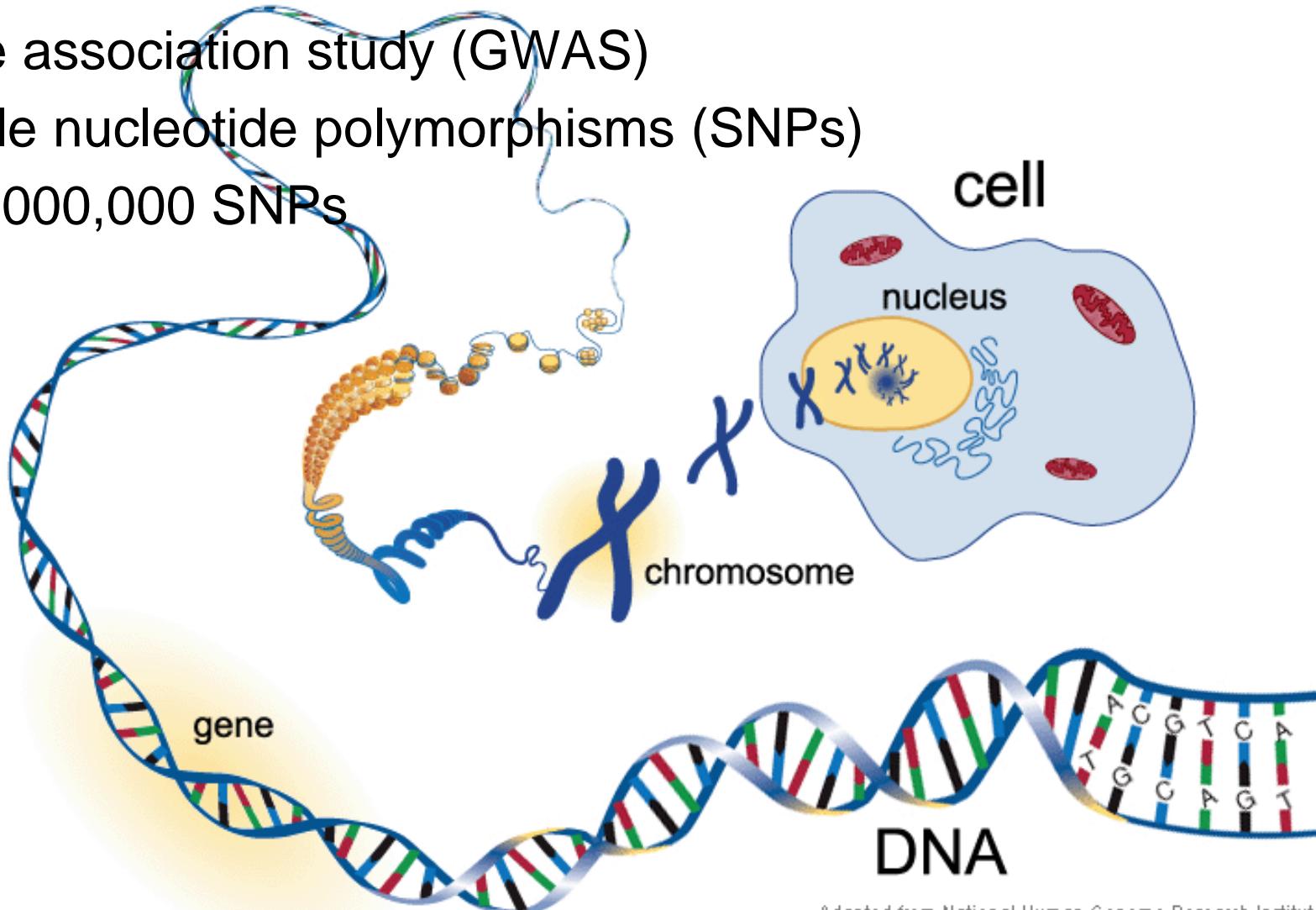
I generated 400 billion data points

- 40,000,000 SNPs
- 10,000 individuals



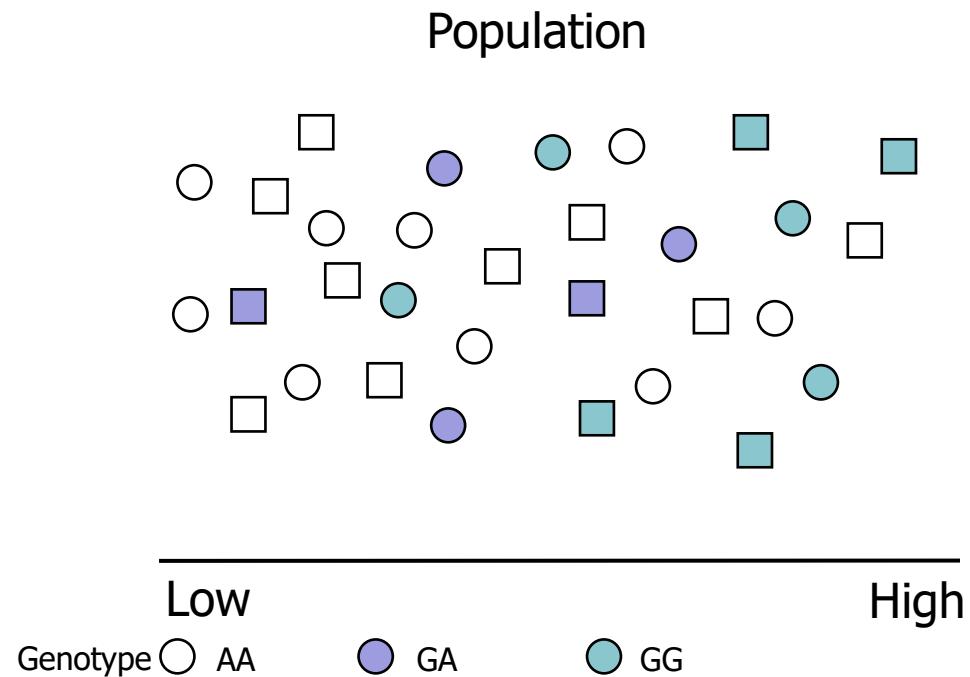
Practical

- Genome-wide association study (GWAS)
- Measure single nucleotide polymorphisms (SNPs)
- 300,000 – 40,000,000 SNPs



GWAS Practical

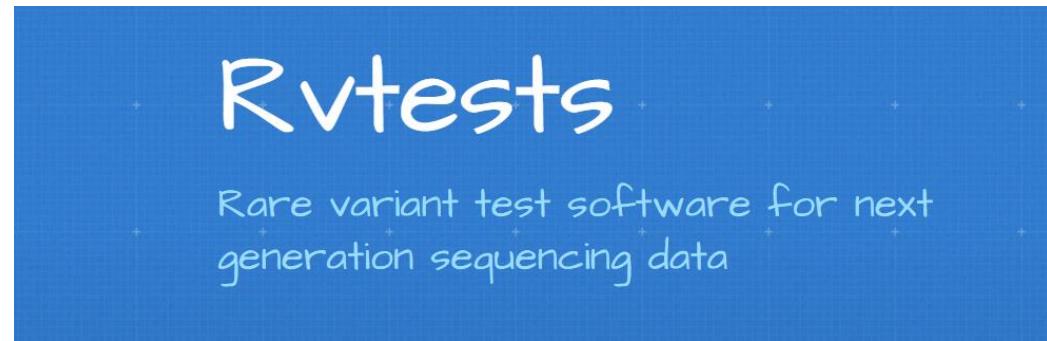
- Investigate whether genotypic variation is associated with different levels of a biomarker
- Successful approach for many complex diseases/traits



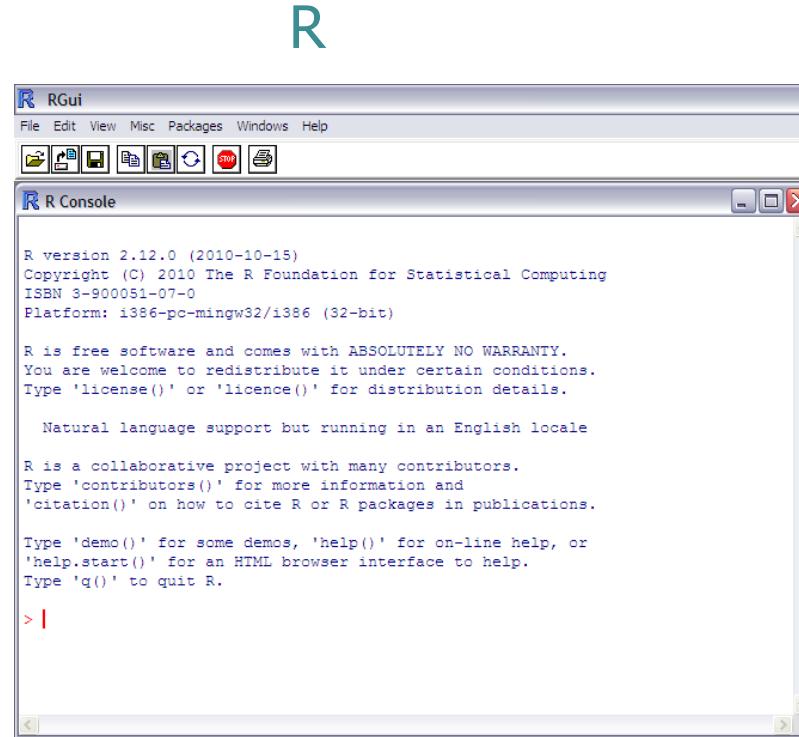
GWAS practical

- Genome-wide association study for the circulating levels of betaine.
- Low circulating levels of betaine have previously been associated with unfavorable cardiovascular risk
- Is there genetic variation associating with circulating Betaine levels?
- In Leiden Longevity Study participants we measured genetic variation and circulating Betaine levels.

Conquer your fear of The Blinking Cursor



1. Explore input files
2. Demo by Niels van den Berg
3. Explore and visualize results



BREAK

False Positive Findings

- **False positive** or spurious association = significant association signals **due to reasons other than true association with the trait under study**
- Possible sources for false positives:
 - Genotyping errors
 - Population stratification
 - Multiple testing

Hardy Weinberg Equilibrium

- HWE is used to check the **genotyping quality** in GWAS analyses
- Allele frequencies should correspond with genotype frequencies
- Example SNP rsXX has 2 alleles:
 - A with a frequency of 0.8
 - G with a frequency of 0.2

Hardy Weinberg Equilibrium

- A= 0.8 is represented by p
B = 0.2 is represented by q
 $\Rightarrow p+q=1$
- From allele frequencies the genotype frequencies can be calculated.

Hardy Weinberg Equilibrium

- Formula of Hardy-Weinberg Equilibrium:

$$AA=p^2$$

$$AG=2pq$$

$$GG=q^2$$

- In this case:

$$AA=0.8*0.8= 0.64$$

$$AG=2*0.8*0.2=0.32$$

$$GG=0.2*0.2=0.04$$

NB: under the assumption of random mating

Hardy Weinberg Equilibrium

- HWE test:

Genotypes	Observed N	Expected N	$(\text{obs-exp})^2/\text{exp}$
AA	60	64 ($p^2 * 100$)	0.25
AG	20	32 ($2pq * 100$)	4.5
GG	20	4 ($q^2 * 100$)	64
Total	100	$\chi^2 = 68.75$	
A frequency	0.8	Df=1	P<0.05
G frequency	0.2		

Population stratification

Systematic difference in allele frequencies between subpopulations in a population, possibly due to different ancestry.

- Can cause false positives if the trait values also differ between the subpopulations.



Population stratification: haircolor example

Sample 1 Americans: $\chi^2=0, p=1$			
	Blond hair		
	Yes	No	Total
Allele 1	320	320	640
Allele 2	80	80	160
Total	400	400	800

Sample 2 Swedish: $\chi^2=0, p=1$			
	Blond hair		
	Yes	No	Total
Allele 1	320	20	340
Allele 2	320	20	340
Total	640	40	680



There is a clear allele frequency difference between American (MAF=0.2) and Swedish (MAF=0.5) people

There is a clear difference between in proportion of "cases" and "controls" American (0.5 blonds) and Swedish (0.8 blonds) people

Population stratification: hair color example

Sample 1 Americans: $\chi^2=0, p=1$			
	Blond hair		
	Yes	No	Total
Allele 1	320	320	640
Allele 2	80	80	160
Total	400	400	800

Sample 2 Swedish: $\chi^2=0, p=1$			
	Blond hair		
	Yes	No	Total
Allele 1	320	20	340
Allele 2	320	20	340
Total	640	40	680



Sample 1 + 2 = Americans + Swedish:
 $\chi^2=34.2, p=4.9 \times 10^{-9}$

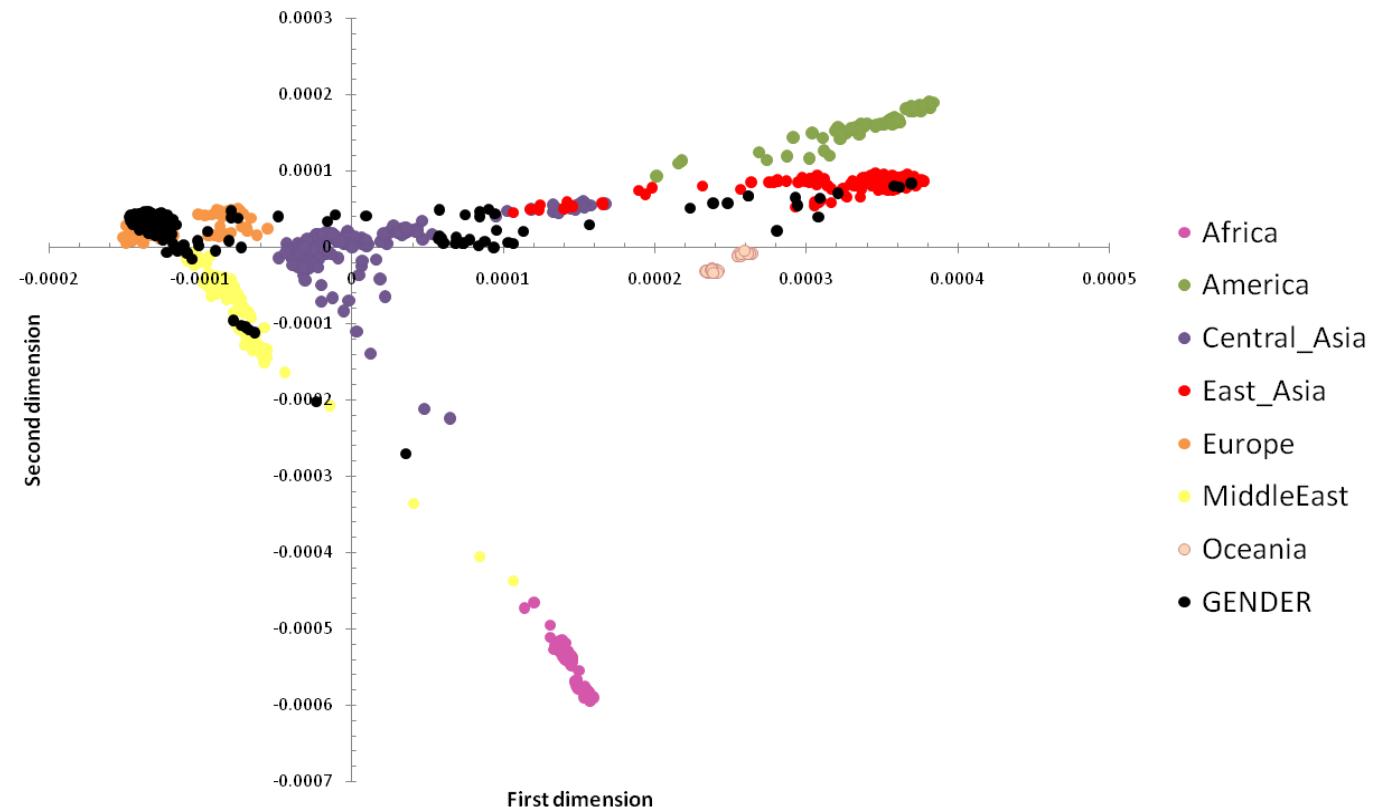
	Blond hair		
	Yes	No	Total
Allele 1	640	340	980
Allele 2	400	100	500
Total	1040	440	1480

MAF allele 2 =
 0.38 in blonds
 0.23 in non-blonds

=> association!

Prevent population stratification

- Detect heterogeneity in origin of participants by comparing genotyping results to HapMap data



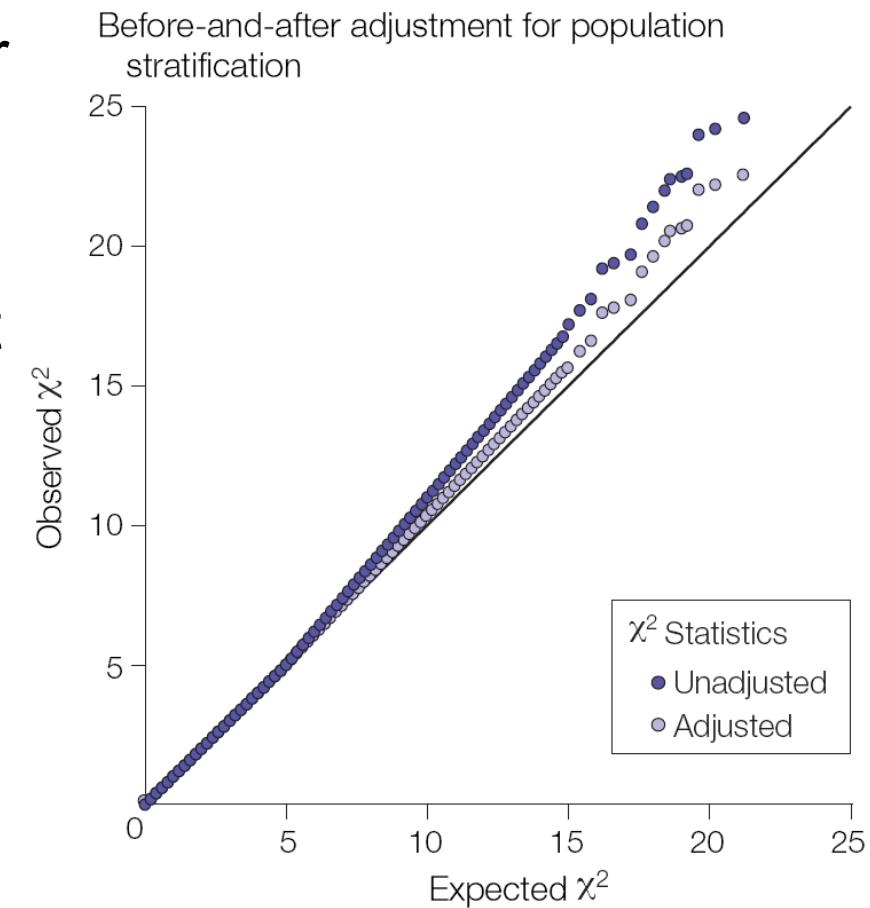
Dealing with population stratification

Ways to deal with population stratification:

- Genomic Control (GC)
- Principal Component Analysis

Genomic Control (GC)

- Population stratification can result in higher test statistics (= lower p -values)
- The genomic control method estimates the factor with which the test statistics are inflated due to the population structure $\rightarrow \lambda$
- Dividing the test statistics by λ cancels this effect out for all SNPs:
 - Unadjusted: $\lambda \chi^2$
 - Adjusted: χ^2

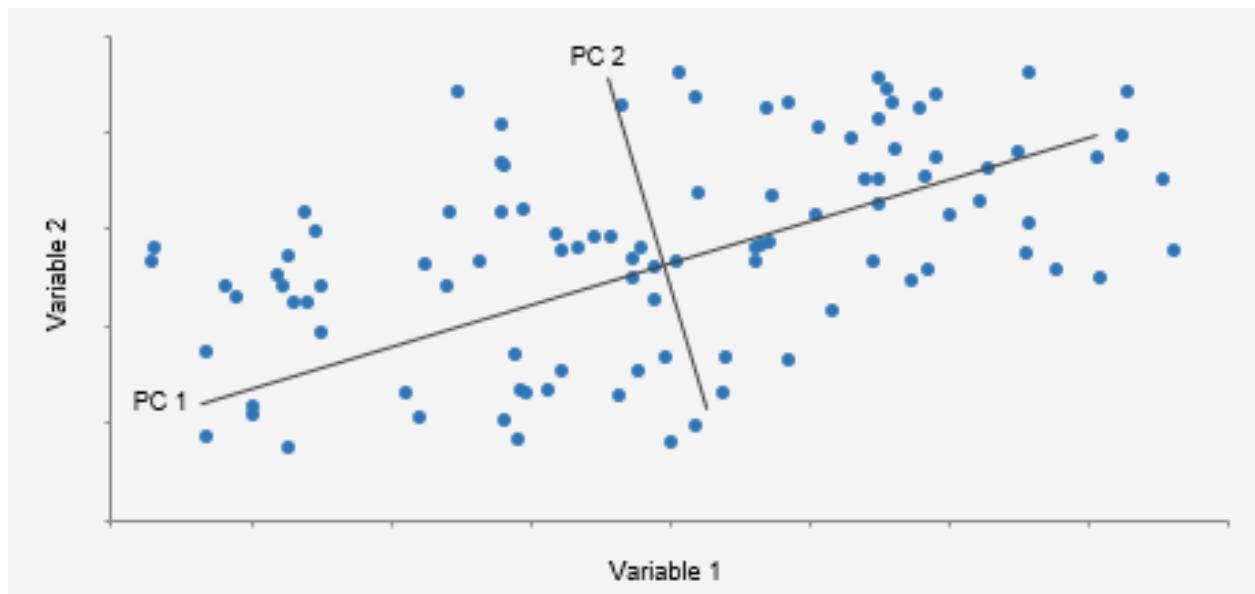


Genomic Control (GC)

- Genomic inflation factor λ is measured by dividing the **median** of the distribution of the chi-square statistics from the **actual tests** by the **median** of the chi-square distribution **under the null**.
- Then, GC applies its correction by dividing the actual association test chi-square statistic results by this λ , thus making these results appropriately more pessimistic.
- A genomic inflation factor λ close to unity reflects no evidence of inflation, while values up to 1.10 are generally considered acceptable for GWAS.

Principle Component Analysis (PCA)

- PCA reduces the dimensions into a few components that explain the main patterns
- The first PC is the mathematical combination of measurements that accounts for the largest amount of variability in the data.
- The second PC (uncorrelated with the first) accounts for the second largest amount of variability, etc.



Adjusting for the first PCs

- The values of the PCs can be used as covariates in subsequent association analysis to control for stratification

Take
—A—
Break

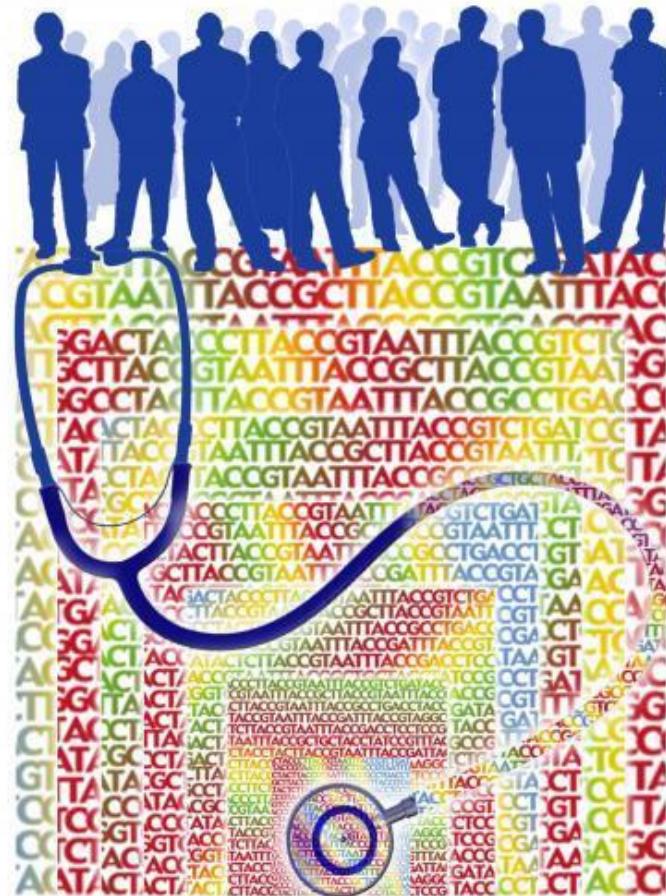
Learning goals

- After this third part you are able to
 - Explain the principles of a genetic association analysis
 - Apply adjustment for multiple testing in genome wide association studies
 - Understand the importance of large sample size and replication of results



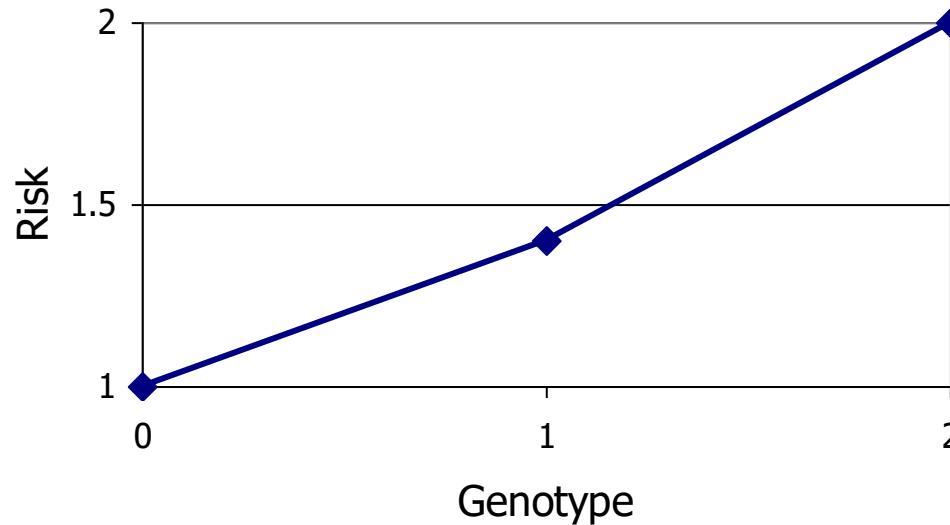
Statistical Analysis

- Keep it simple
 - Single SNP analysis
 - Easy to interpret
 - Minimal number of tests
 - All very basic statistical tests
(χ^2 or similar)



Statistical Analysis

- Cochrane-Armitage trend test
 - Same as linear-by-linear in SPSS
 - Additive effect: more risk alleles, more effect
 - Genotype coding: 0, 1, 2 (counting no. of rare alleles)
 - Plausible biological model
 - Robust against random fluctuations
 - Optimal power (df=1)



Regression analysis

$$\text{Trait} \sim \text{Constant} + \beta_1 \times \text{SNP} (+ \beta_2 \times \text{PC1} + \beta_3 \times \text{PC2} + \beta_n \times \text{PCn})$$

Where the dependent variable “Trait” is

- a) Quantitative trait => linear regression
- b) Case-Control status => logistic regression

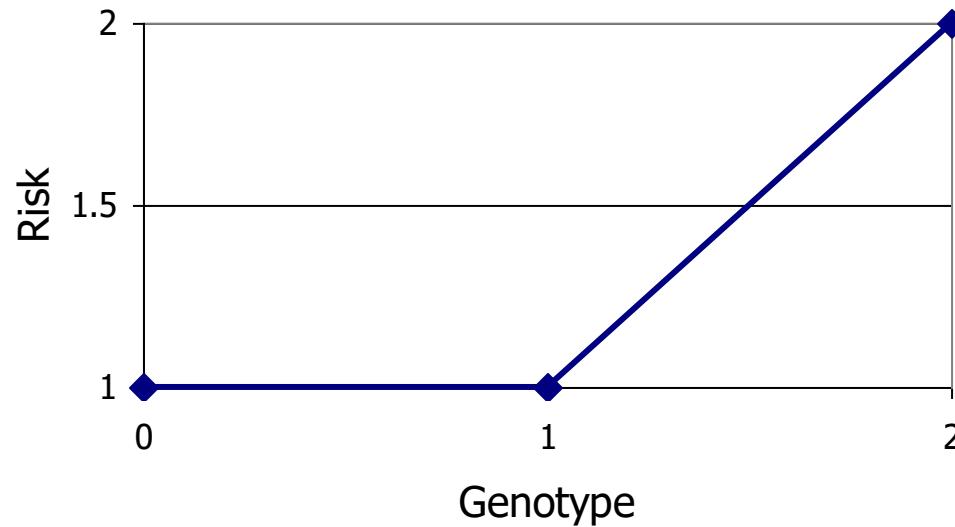
And where the independent variable “SNP” is coded as number of alternative alleles (0, 1 or 2) or in dosages (number between 0 and 2)

Cochrane Armitage Trend test $H_0: \beta_1 = 0$, No association

ASSOCIATION if β_1 is NOT EQUAL to 0

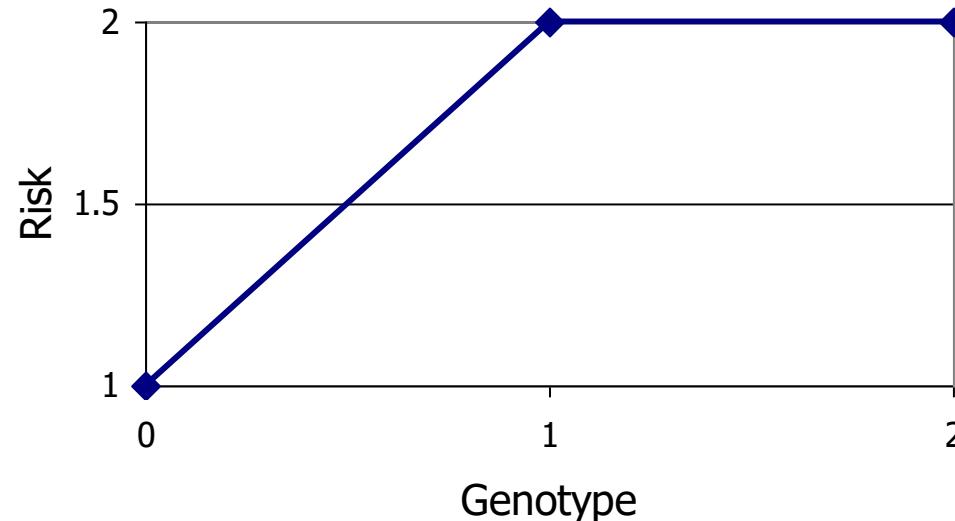
Alternatives

- Recessive test
 - Mendelian disease like Cystic Fibrosis
 - Assumes effect among rare homozygotes only

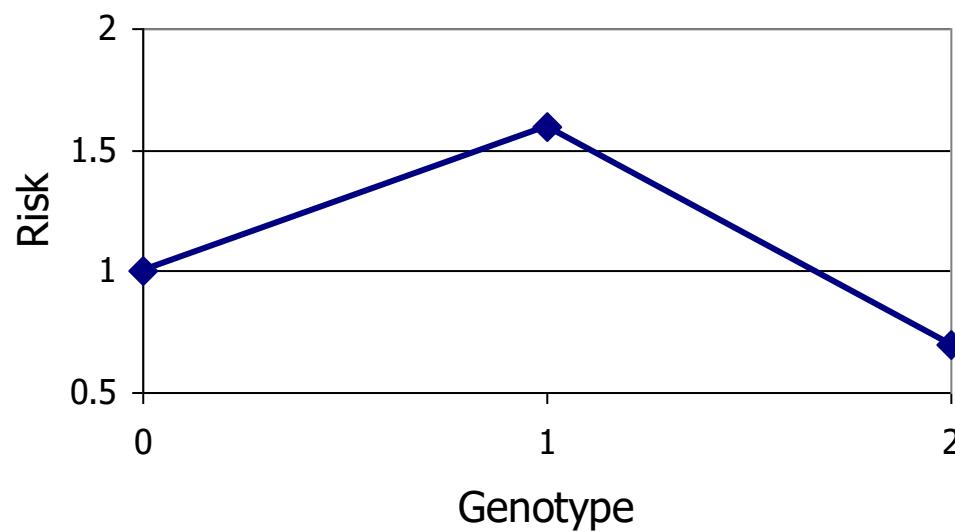
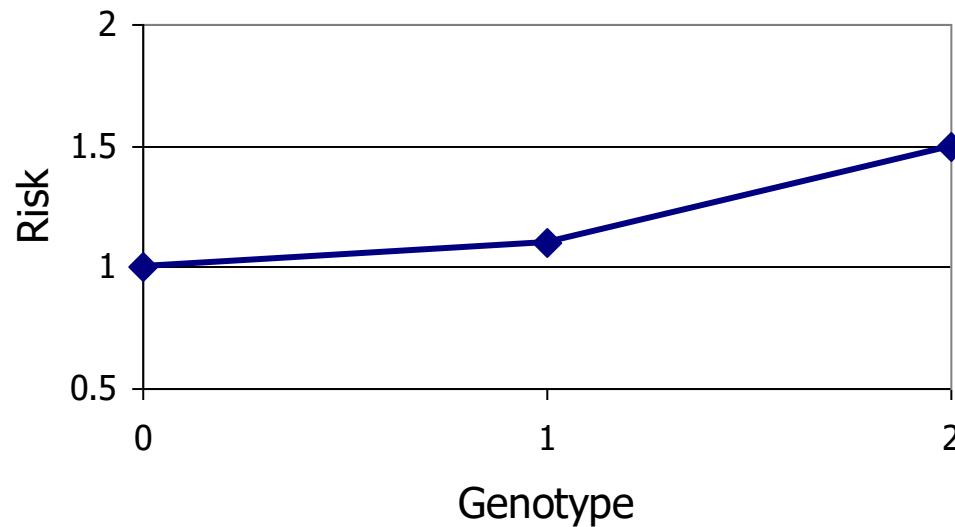


Alternatives

- Dominant test
 - Mendelian disease like Huntington’s Disease
 - A single rare allele is sufficient for disease trait
 - Common homozygotes and heterozygotes same effect
 - Often trend test has sufficient power

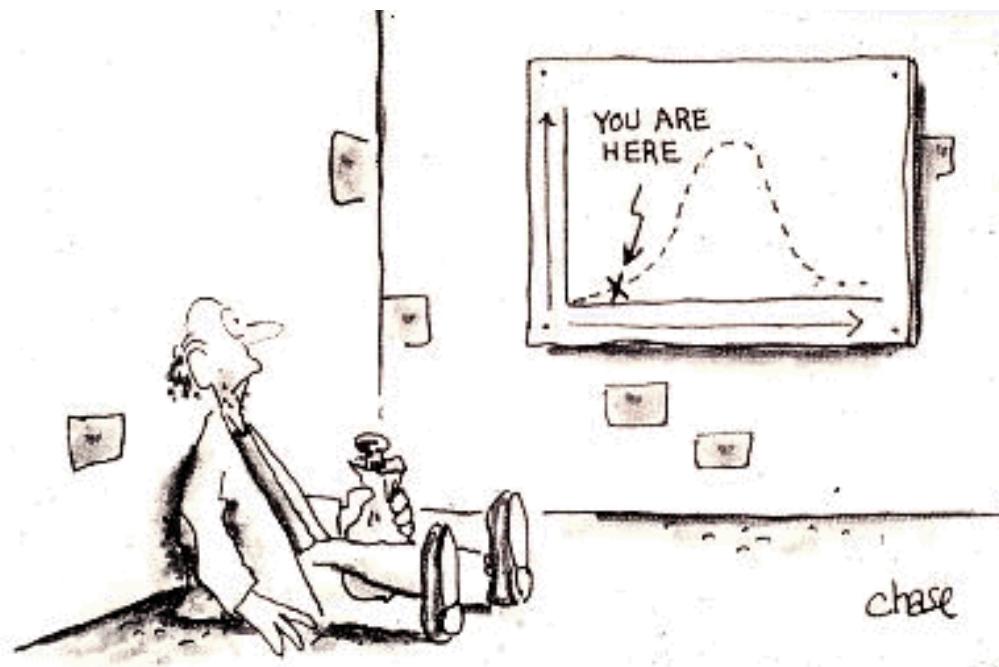


What if



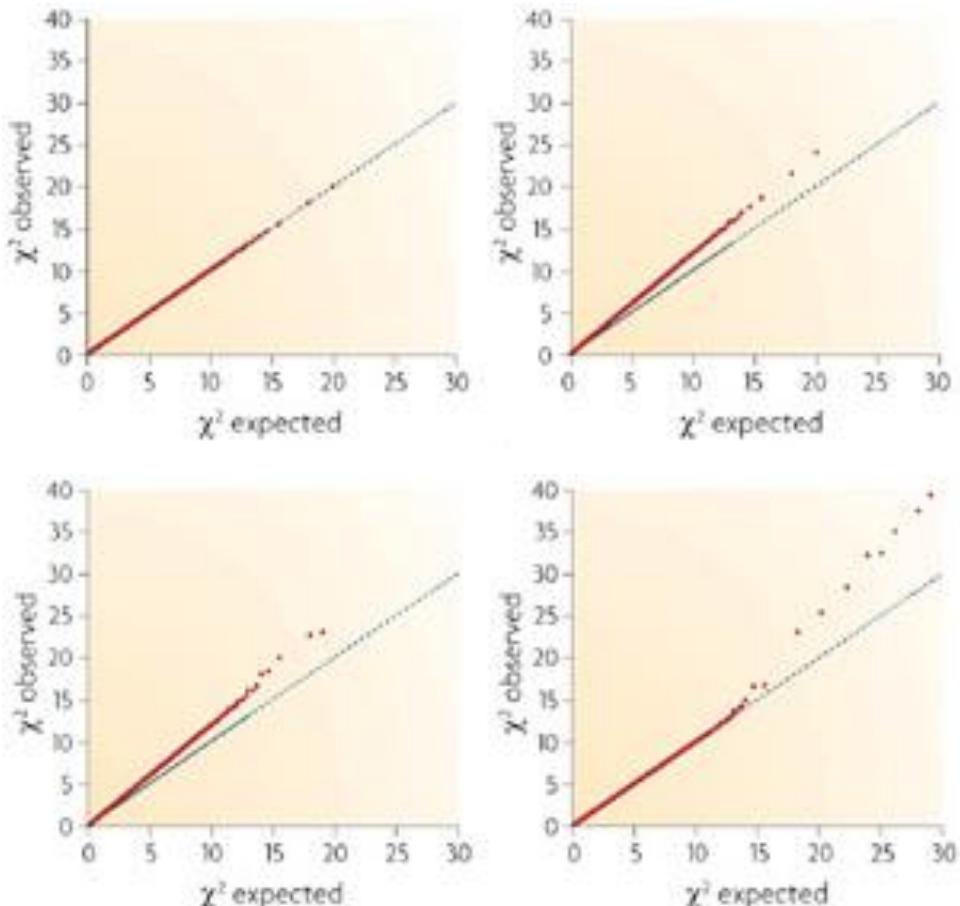
Significance in GWAs

- Single test: significance threshold = 0.05
- Multiple tests: significance threshold = $0.05 / \text{number of } \underline{\text{independent tests}}$ (Bonferroni)
- The whole genome is estimated to contain 1 million independent tests.
→ the p-value threshold to take into account should be $0.05/1.000.000 = 0.00000005 = 5 \times 10^{-8}$.



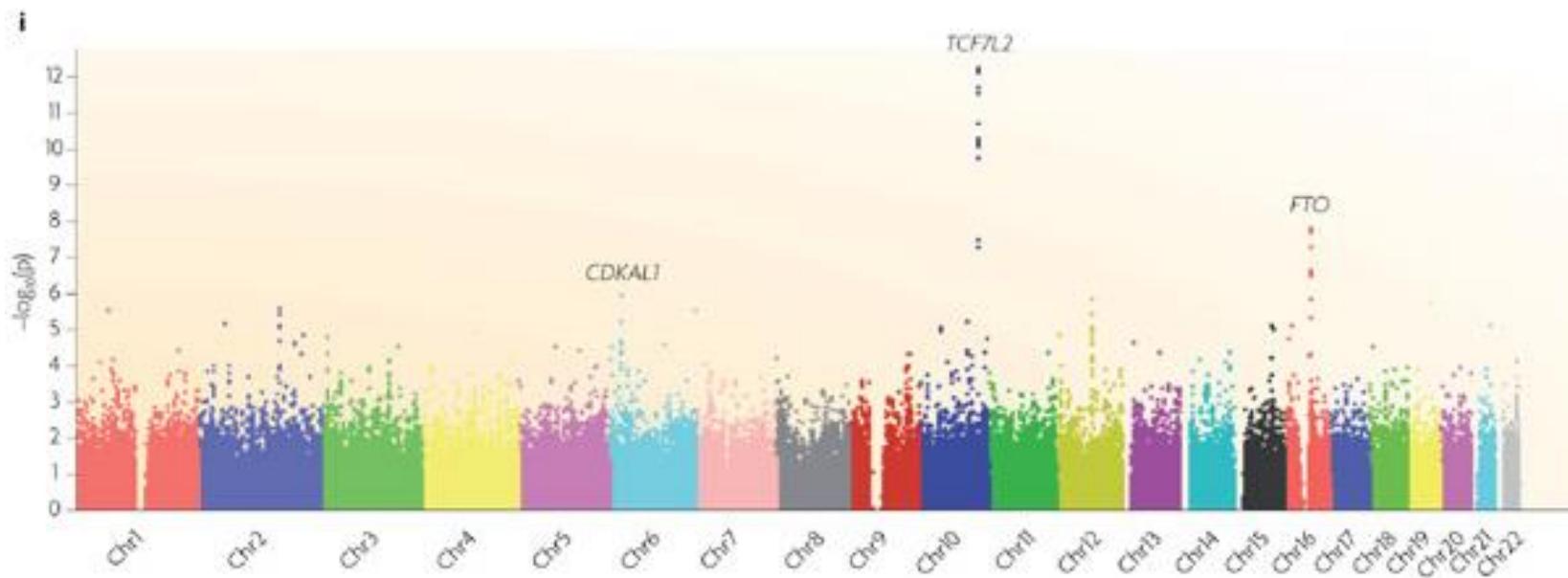
Bias and enrichment

- QQ plots allow to detect bias and enrichment for low p-values



Visualization

- Manhattan plots are standard way to display GWAS results



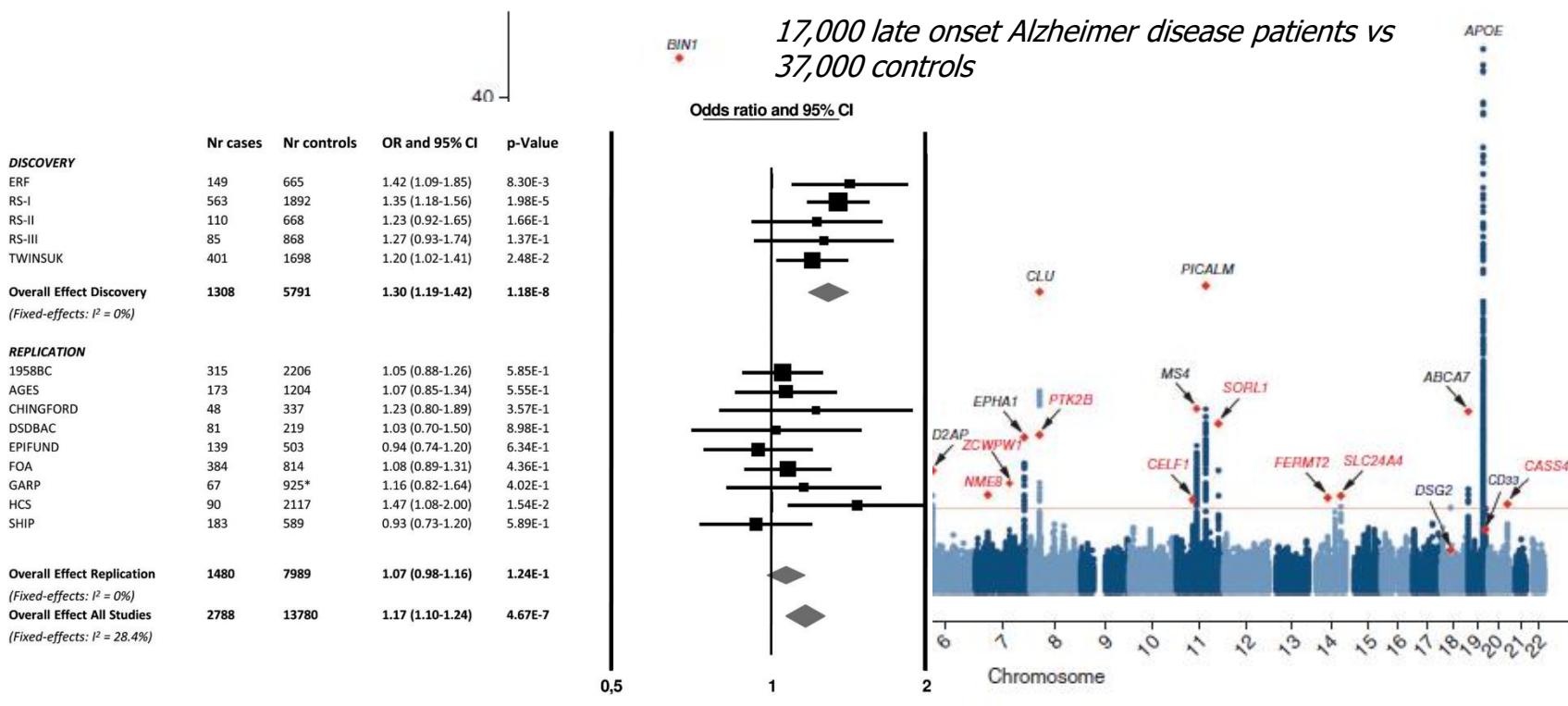


Bruno TASSAN

BREAK BREAK
BREAK BREAK
BREAK BREAK
BREAK BREAK

State of the art Meta-GWAS

- Combine multiple GWAS ($n > 50,000$)
- Follow-up in multiple cohorts ($n > 50,000$)
- Genome wide significant loci ($P < 5 \times 10^{-8}$)



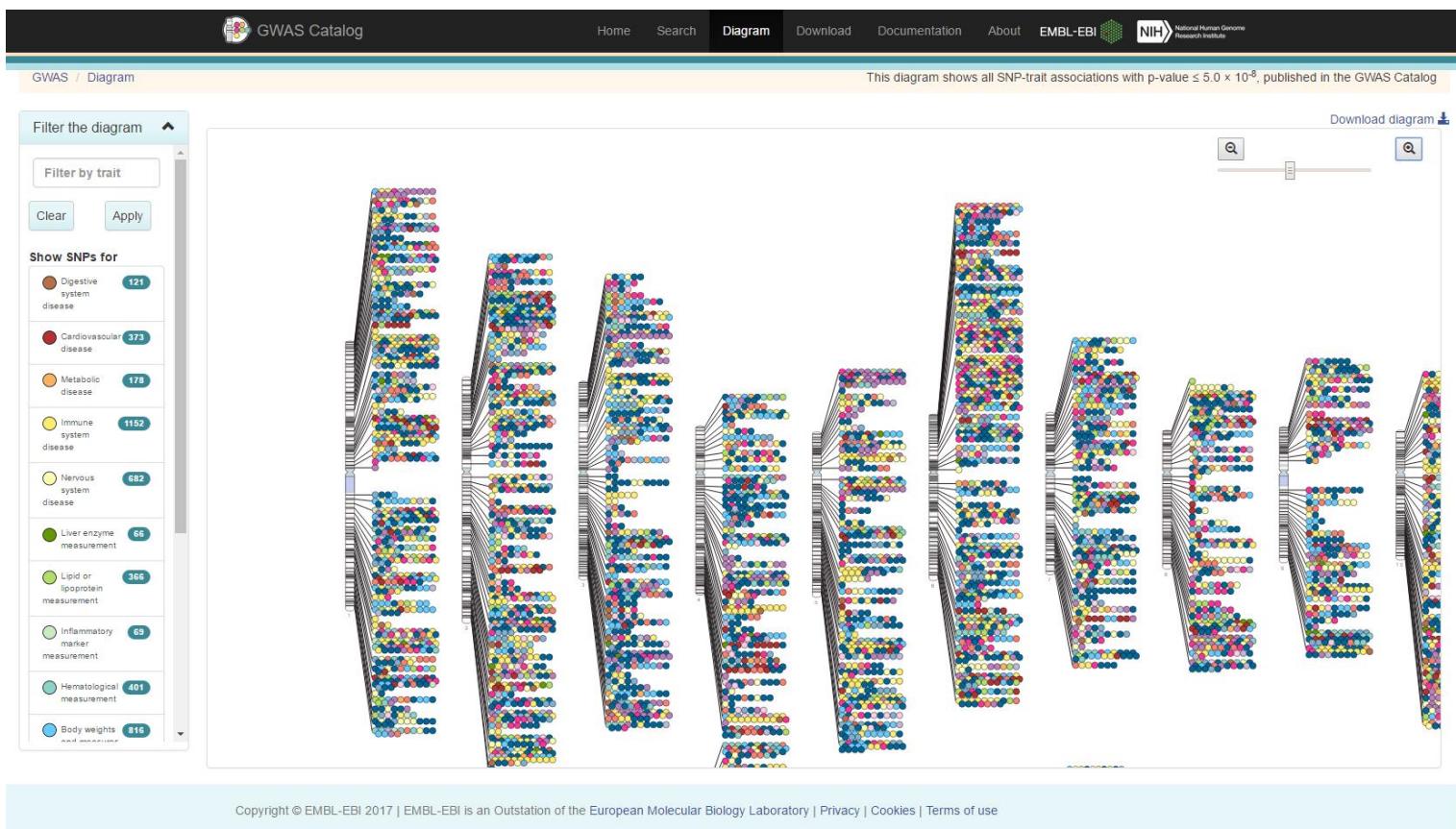
Replication

- Without replication no one will believe you
 - Same SNP
 - Same allele
 - Same phenotype
 - Same genetic model
- Considerations
 - Often ≥ 2 large replications required nowadays: collaboration is key
 - Replication in cohorts that do not have GWAS data available

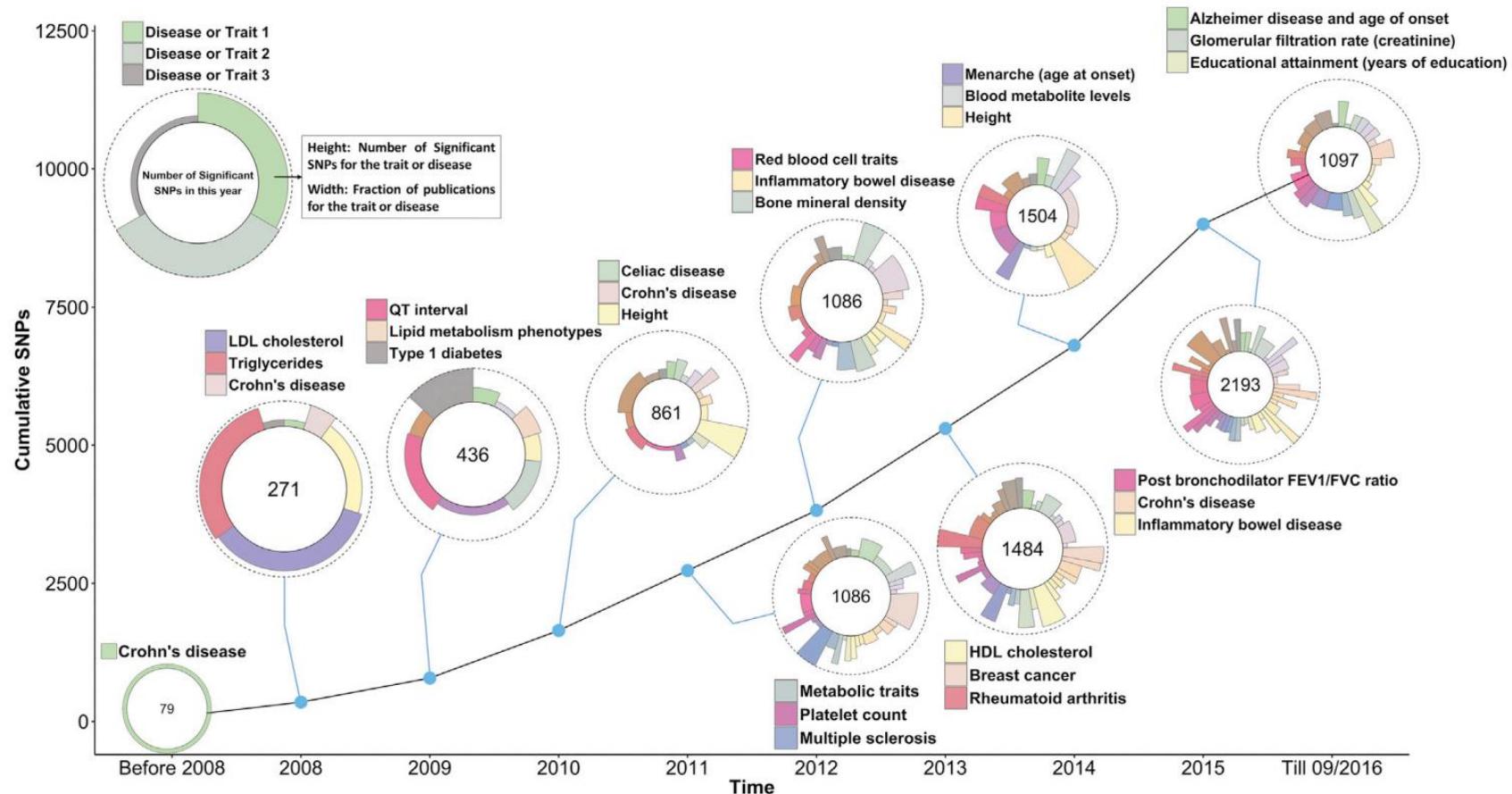


A Catalog of GWAS

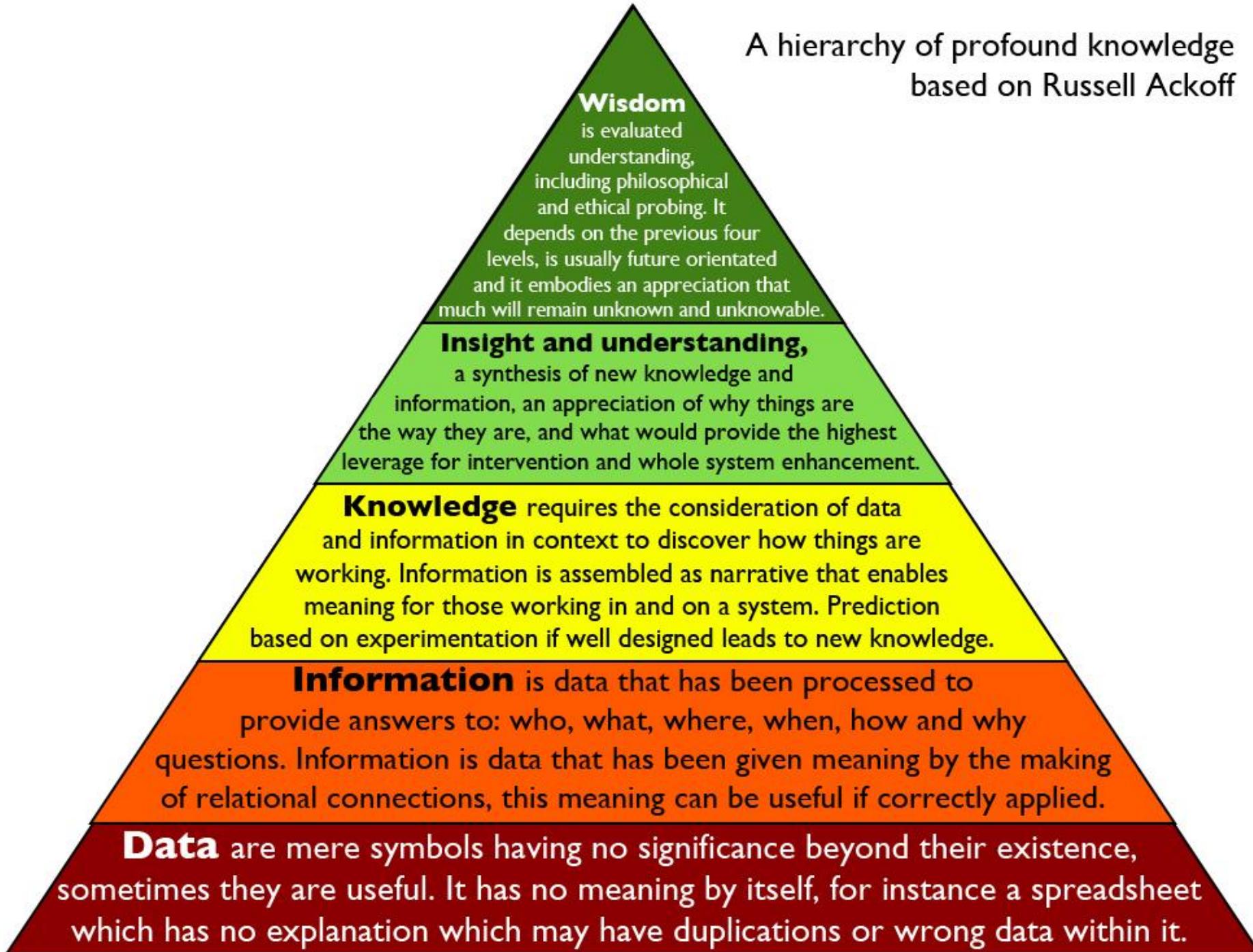
- Database of all GWAS results <https://www.ebi.ac.uk/gwas/>
- 11/08/2022: 6,096 publications and 434,351 unique SNP-Trait associations $P < 10^{-8}$



GWAS SNP-Trait discovery Timeline



A hierarchy of profound knowledge based on Russell Ackoff



Learning goals

- After this fourth part you are able to
 - Explain potential factors contributing to the missing heritability of traits
 - Understand that genetic loci associated with disease are no good predictors for disease
 - Provide an outlook how mechanisms underlying GWAS results may be investigated to establish causality



What we do not know after all those GWASes

- Missing heritability

Disease	Number of loci	Proportion of heritability explained
Age-related macular degeneration ⁷²	5	50%
Crohn's disease ²¹	32	20%
Systemic lupus erythematosus ⁷³	6	15%
Type 2 diabetes ⁷⁴	18	6%
HDL cholesterol ⁷⁵	7	5.2%
Height ¹⁵	40	5%
Early onset myocardial infarction ⁷⁶	9	2.8%
Fasting glucose ⁷⁷	4	1.5%

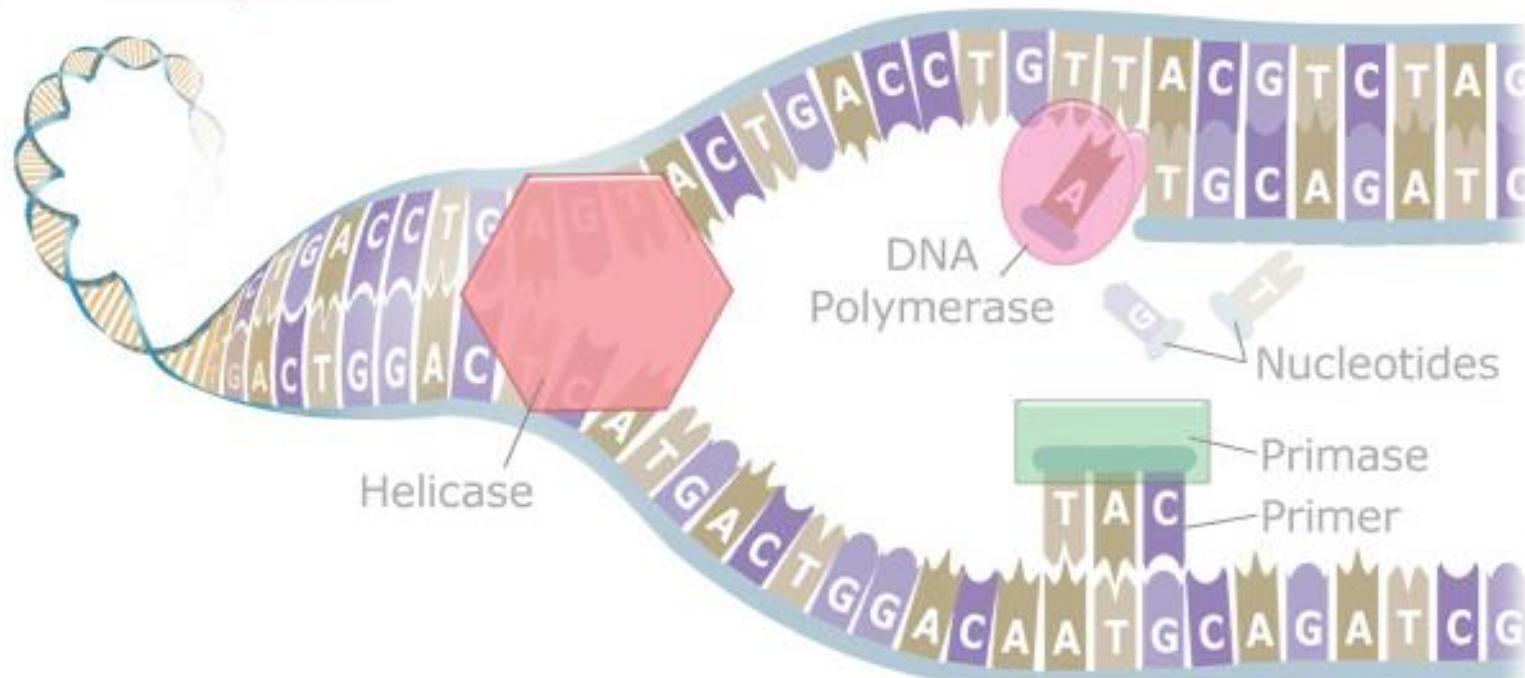
Incomplete detection genetic variation

- Factors
 - Common SNPs missed with (previous) arrays
 - Structural variation, copy number variants, in/dels
 - Much fewer known than SNPs



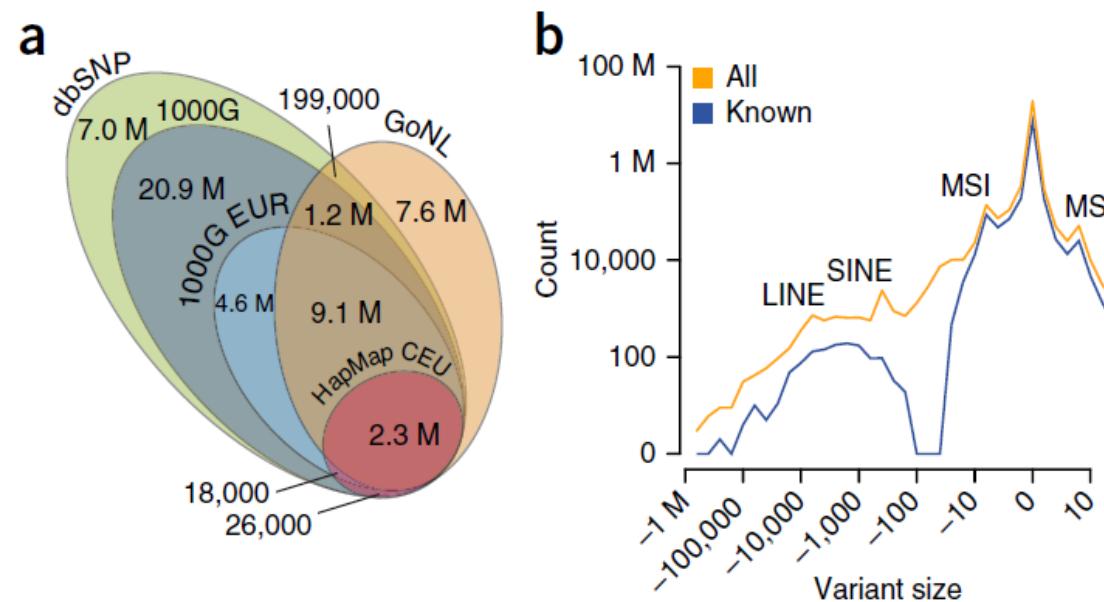
Incomplete detection genetic variation

- Factors
 - Common SNPs missed with (previous) arrays
 - Structural variation, copy number variants, in/dels
 - Much fewer known than SNPs
 - Rare variants: Sequencing



Current state-of-the-art

- Genome of The Netherlands (Go.NL)
 - 769 samples
 - ~20M SNPs
 - Insertion and deletions (Indels)



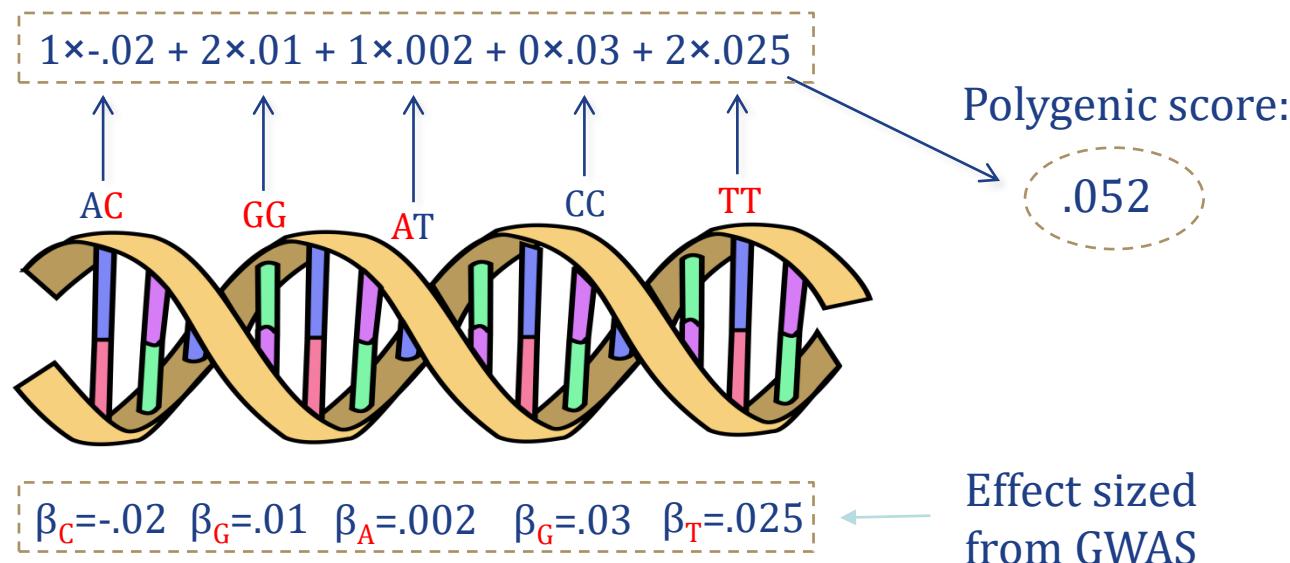
Prediction by single SNP

- Limited or no prediction despite biological insight

Region	Candidate gene(s)	Weight†	Reference‡	Risk allele	Risk allele frequency	Other allele	Coronary heart disease (total n=19790)	
							Pooled HR (95% CI)‡	p value
rs17465637	1q41	MIA3	1.14	15	C	0.75	A	0.99 (0.87-1.12) 0.854
rs11206510	1p32	PCSK9	1.15	15	T	0.84	C	0.94 (0.81-1.09) 0.431
rs646776	1p13	CELSR2-PSRC1-SORT1	1.19	15	T	0.79	C	0.96 (0.84-1.09) 0.512
rs6725887	2q33	WDR12	1.17	15	C	0.11	T	1.14 (0.96-1.35) 0.126
rs9818870	3q22	MRAS	1.15	16	T	0.10	C	0.88 (0.73-1.06) 0.174
rs3798220	6q26	LPA	1.68	18	C	0.01	T	2.07 (1.39-3.09) 3.8×10⁻⁹
rs9349379	6p24	PHACTR1	1.12	15	C	0.44	T	1.16 (1.04-1.29) 0.008
rs4977574	9p21	CDKN2A-CDKN2B	1.29	15	G	0.43	A	1.21 (1.08-1.34) 0.001
rs1746048	10q11	CXCL12	1.17	15	C	0.84	T	1.13 (0.97-1.33) 0.113
rs2259816	12q24	HNF1A	1.08	16	T	0.36	G	1.02 (0.91-1.14) 0.774
rs3184504	12q24	SH2B3	1.13	17	T	0.40	C	1.03 (0.92-1.15) 0.568
rs1122608	19p13	LDLR	1.15	15	G	0.79	T	1.00 (0.87-1.14) 0.988
rs9982601	21q22	SLC5A3-MRPS6-KCN2	1.20	15	T	0.14	C	1.29 (1.07-1.57) 0.009

Prediction by Polygenic Risk Score

- Polygenic Scores capture (part of) someone's genetic "risk" by summing all risk alleles weighted by the effect sizes estimated in a Genome-Wide Association Study (GWAS)



Prediction by Polygenic Risk Score

- By summing the collective effect sizes of many SNPs you can quantify part of the genetic “risk” in an independent dataset
- Polygenic Scores generally improve when adding SNPs that individually didn’t reach genome-wide significance

Prediction by Polygenic Risk Score

NATURE GENETICS

LETTERS

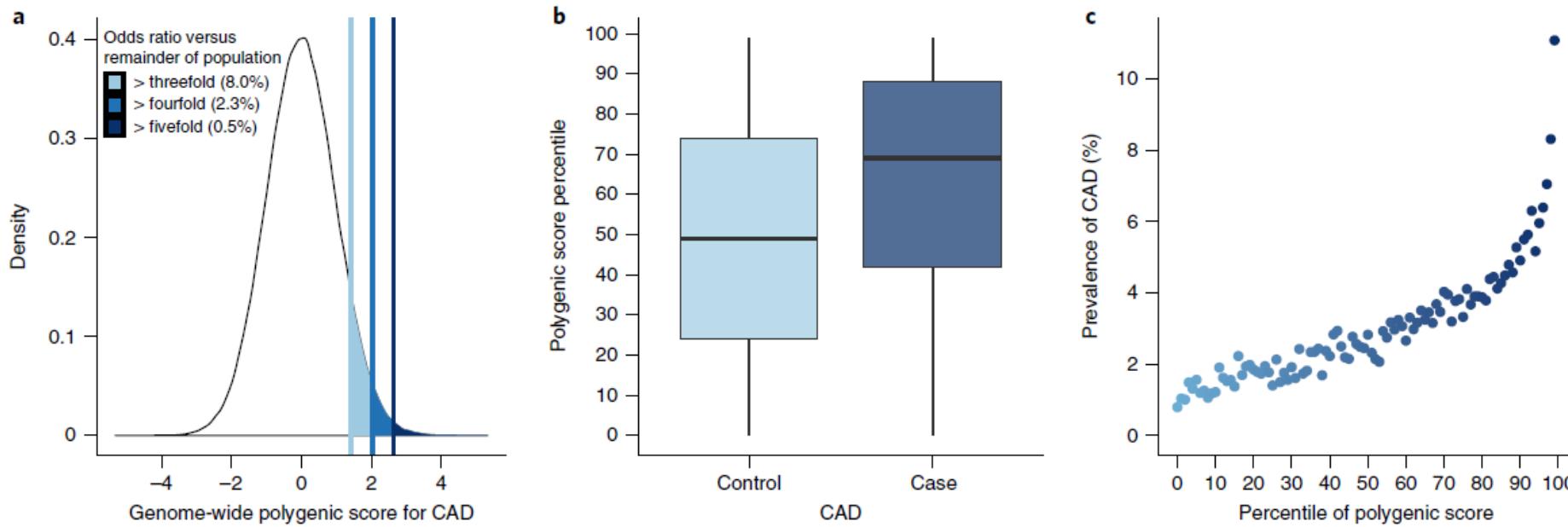
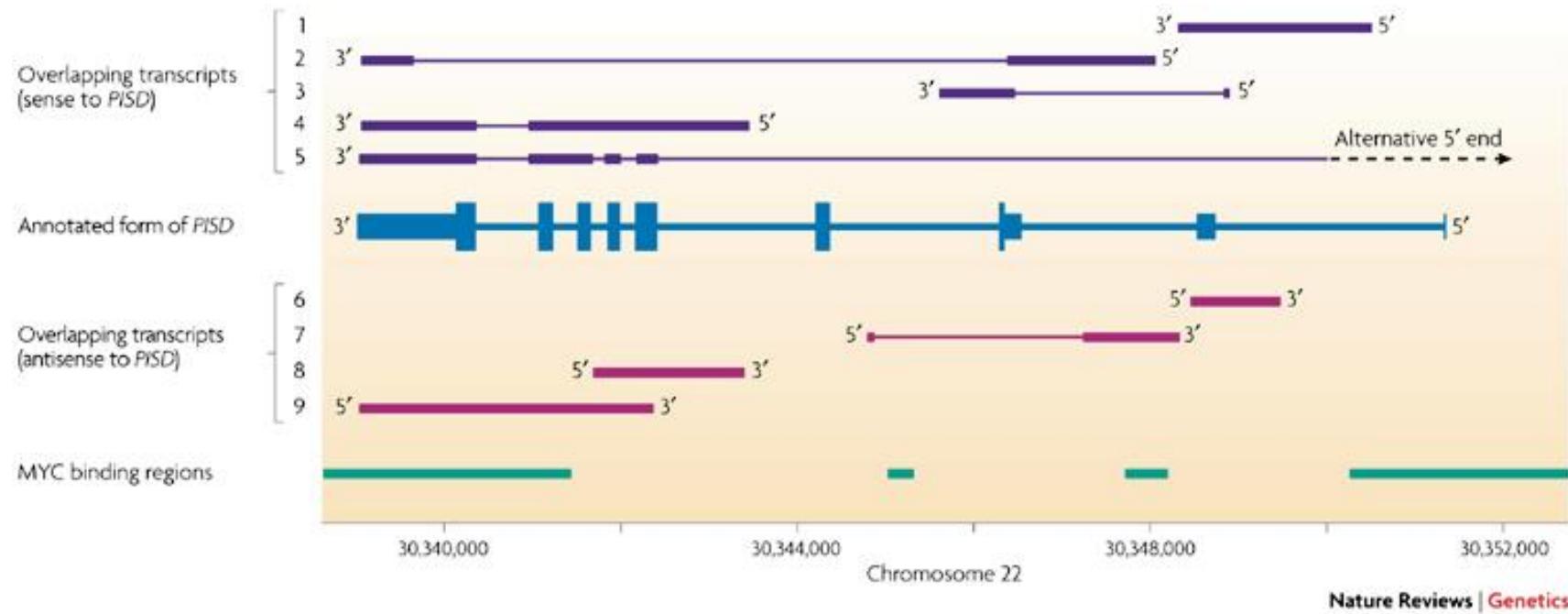


Fig. 2 | Risk for CAD according to GPS. **a**, Distribution of GPS_{CAD} in the UK Biobank testing dataset ($n = 288,978$). The x axis represents GPS_{CAD} , with values scaled to a mean of 0 and a standard deviation of 1 to facilitate interpretation. Shading reflects the proportion of the population with three-, four-, and fivefold increased risk versus the remainder of the population. The odds ratio was assessed in a logistic regression model adjusted for age, sex, genotyping array, and the first four principal components of ancestry. **b**, GPS_{CAD} percentile among CAD cases versus controls in the UK Biobank testing dataset. Within each boxplot, the horizontal lines reflect the median, the top and bottom of each box reflect the interquartile range, and the whiskers reflect the maximum and minimum values within each grouping. **c**, Prevalence of CAD according to 100 groups of the testing dataset binned according to the percentile of the GPS_{CAD} .

A gene is not what it used to be

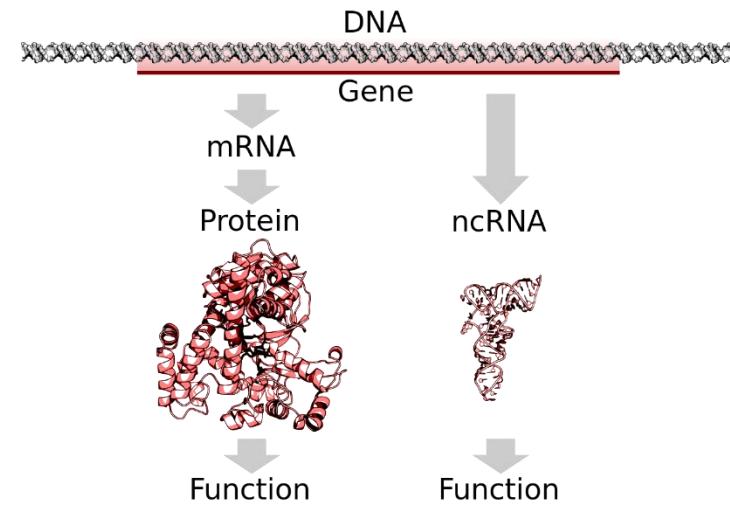
- Multiple use of same sequence
 - Phosphatidylserine decarboxylase (PSID)



Kapranov et al. Nat Rev Genet 2007 & ENCODE publications

A gene is not what it used to be

- Up to 90% of genomic DNA is transcribed
 - 1-2% encodes exons, ~15% exons+introns
- Alternative initiation of transcription: ~60%
 - Alternative TSS 10s-100s kb away
 - Encode: 90% of genes have unannotated exon/TSS
- Alternative splicing: 60%
- Transcripts with anti-sense counterpart: 60%
- Alternative polyadenylation
- Gene fusions
- Trans splicing
- RNA synthesis at enhancers (eRNA)
- Relevant transcription factor binding occurs anywhere



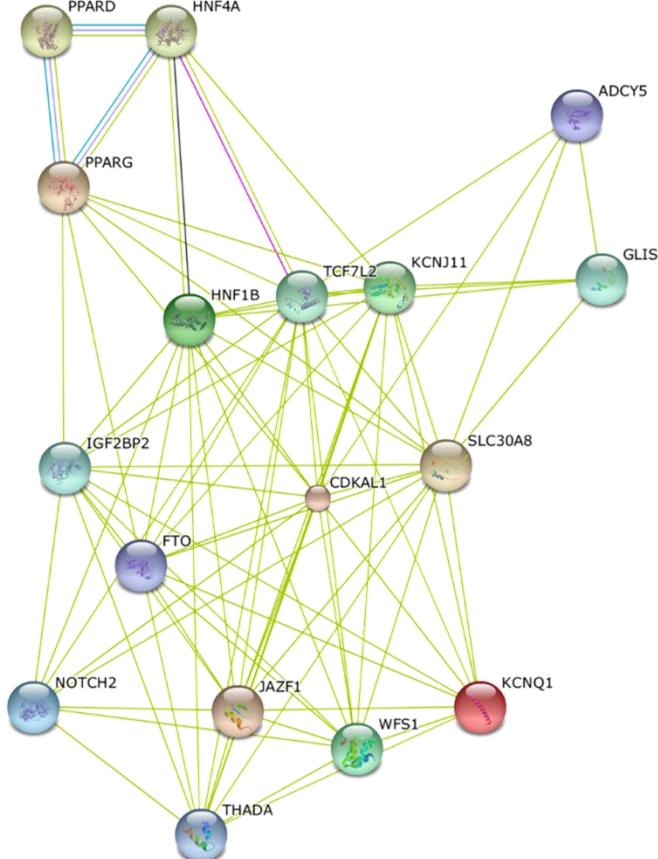
Underlying Mechanism

- Pathway analyses:
- DAVID
 - Database for Annotation, Visualization and Integrated Discovery
 - Gene Functional Classification Tool Based on Gene Ontology
- STRING
 - Search Tool for the Retrieval of Interacting Genes/Proteins
 - Known and predicted protein-protein interactions
- DAPPLE
 - Disease Association Protein-Protein Link Evaluator
 - Physical connectivity among proteins encoded for by genes according to protein-protein interactions reported in the literature.



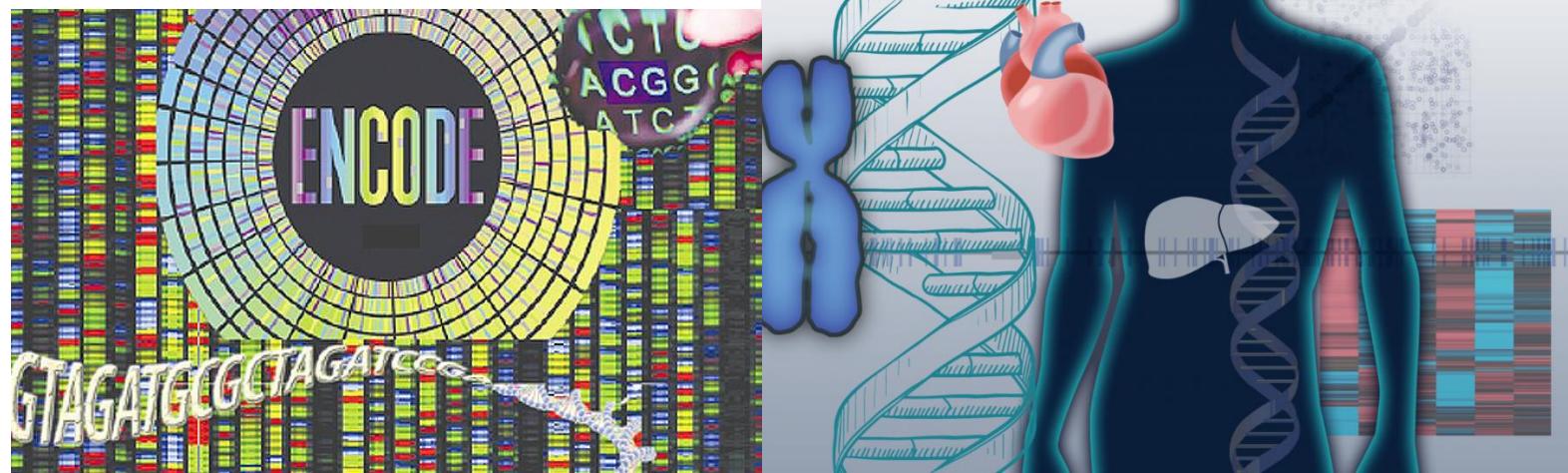
Underlying Mechanism

STRING



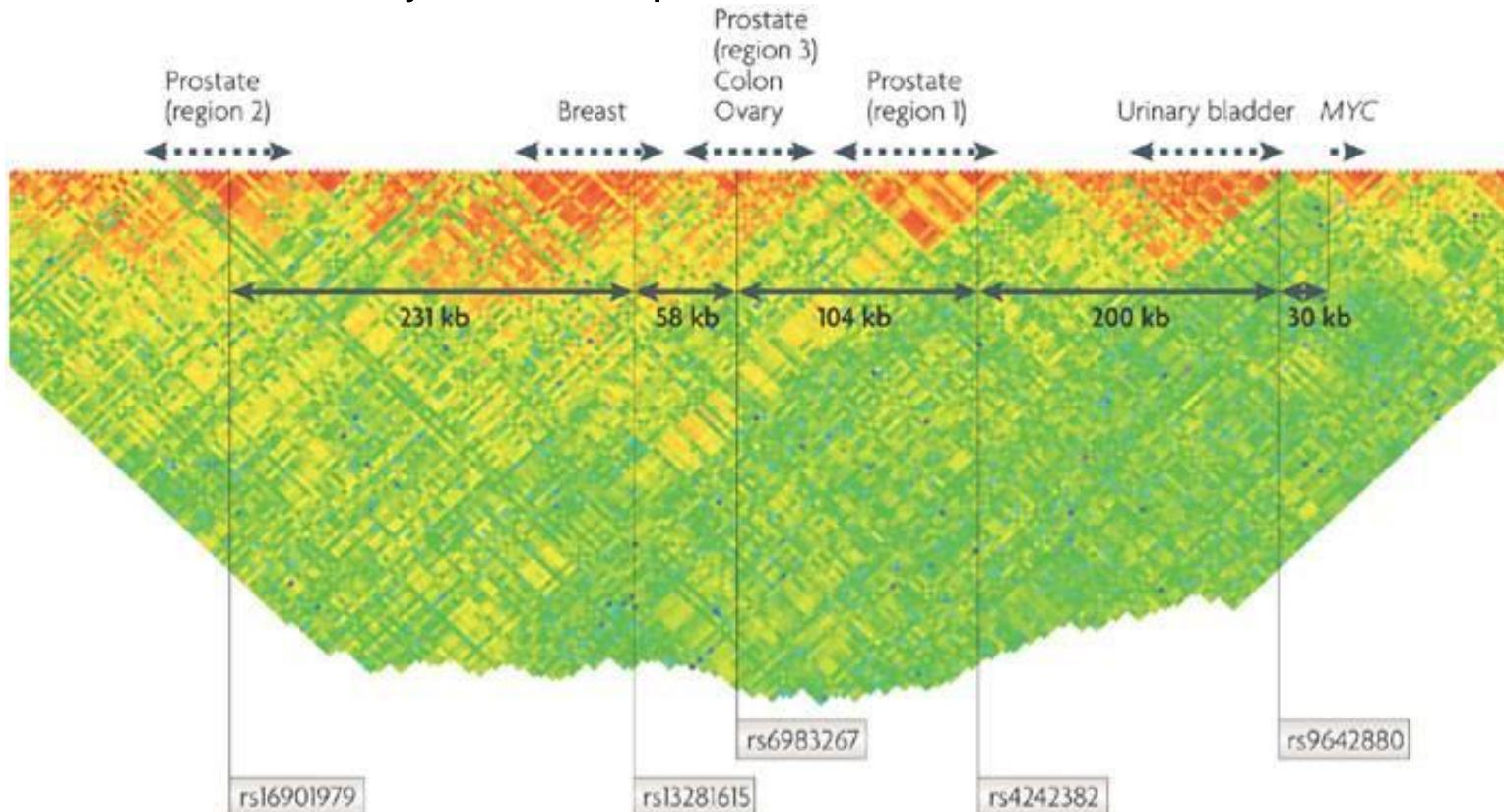
Causality

- Combine knowledge from public databases using UCSC genome browser
 - Databases with biological knowledge
 - ENCODE
 - Haploreg
 - Databases with functional knowledge
 - GTEx for eQTLs
 - Entrez Gene for gene functions



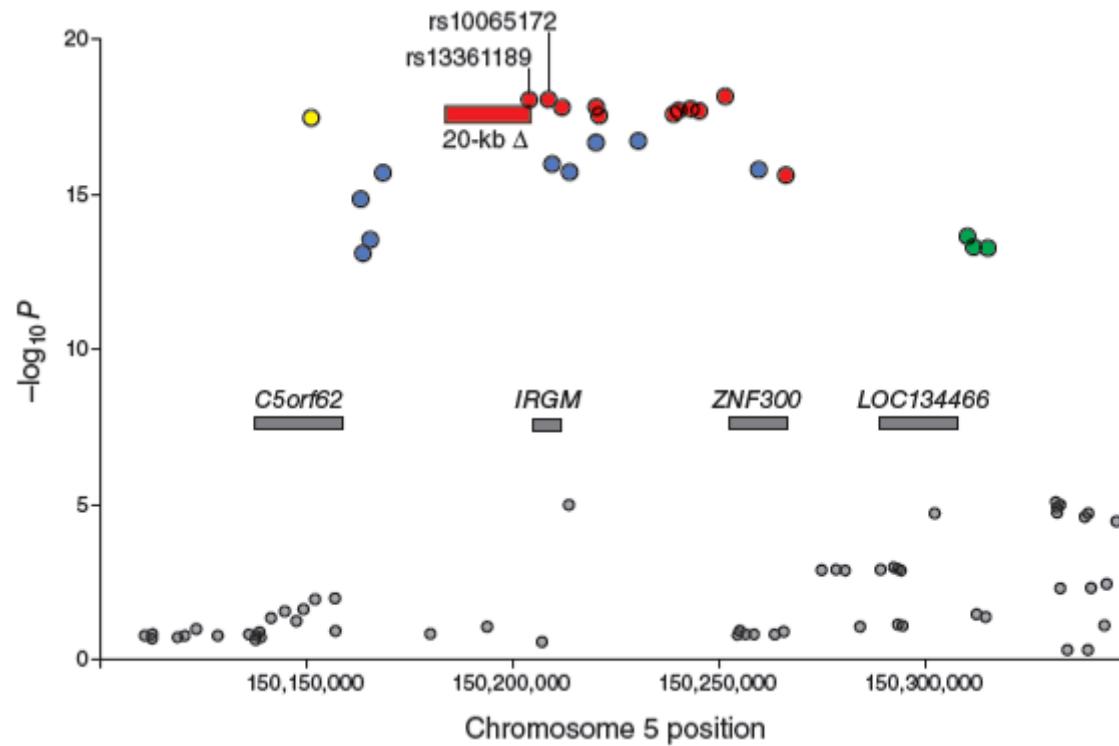
Truly novel findings

- ‘Gene deserts’
 - 8q24 region confers susceptibility to various cancers
 - Functional studies are very much required



Proven causality

- Rarely causality is proven or even plausible



Brest et al Nat Genet 2011. Locus associated with Crohn's Disease: a synonymous variant in the *IRGM* coding region alters a binding site for miR-196 and modulates *IRGM*-dependent autophagy.

Challenges in GWAS studies

- Weaknesses of designs: cross-sectional, broad phenotypes
- Complex effects: interactions, allele specific effects
- Incomplete detection of genetic variation



Complex effects

- Interaction
 - Interactions within biomolecular networks: rare combinations of common variants (epistasis – as routinely seen in yeast)
- Allele specific effects
 - Estimate: ~15% of type 2 diabetes heritability due to known variants involves parent-of-origin effects.

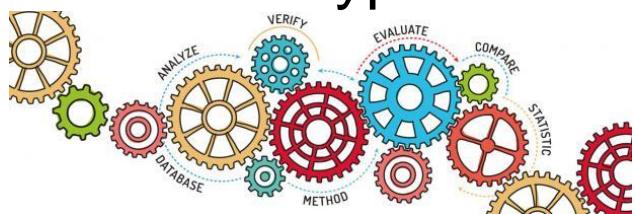
Table 1 | Parental-origin-specific analyses of disease-susceptibility variants

Disease, SNP [alleles]*	NCBI build36 position, N	M, F_{con}	Standard case-control test		Tests of association with parental origins						
			OR	P†	Paternal allele§		Maternal allele§		2-d.f. test		Paternal vs maternal (case only)
			OR	P	OR	P	OR	P	OR	P	n12:n21¶
T2D, rs2334499 [T/C]											
C11 1,653,425, 1,468 (discovery)	34,706,	0.412	1.11	0.017	1.41	4.3×10^{-9}	0.87	0.020	3.5×10^{-9}	437:276	7.0×10^{-9}
783 (replication)			1.02	0.71	1.23	0.0055	0.84	0.023	0.0018	222:157	8.0×10^{-4}
2,251 (combined)			1.08	0.034	1.35	4.7×10^{-10}	0.86	0.0020	5.7×10^{-11}	659:433	4.1×10^{-11}

Life after genome-wide studies



- Go big: meta-analysis
- Increase genetic detail (rare variants)
- Smarter clinical end-points
- Detailed intermediate phenotypes (biomarkers) and system approaches (vertical genomics)
- Pathway analyses
- Acquire biological knowledge from public databases
- Re-analysis publicly available data (e.g. interactions)
- Functional studies to prove causality (!)
- General: more hypothesis-driven, more depth



Upload answers to Brightspace,
TurnItIn

Filename: *Yourname_YourStudentnr_GWAS_2022*