eurac
research

Running GWAS

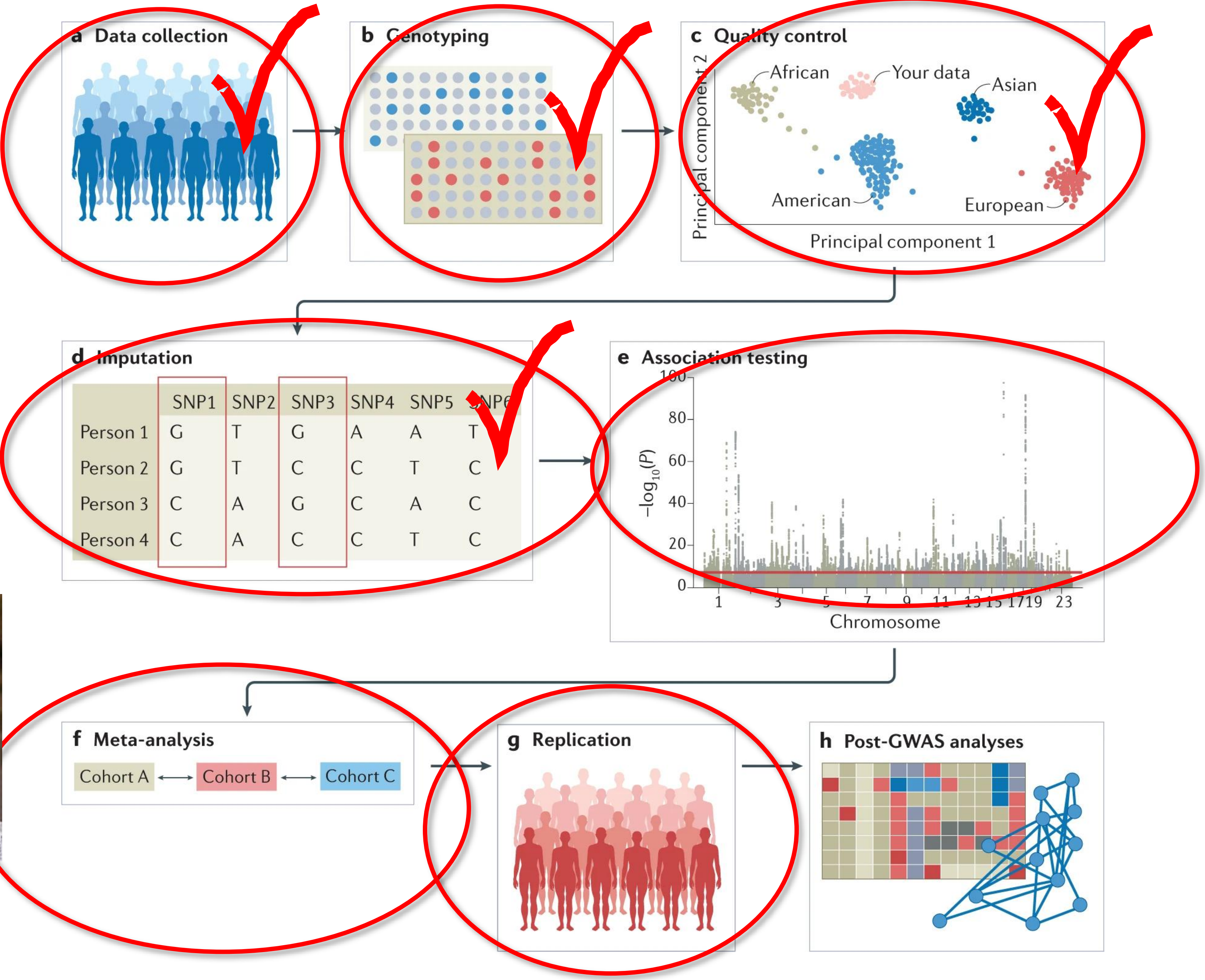
Cristian Pattaro

Eurac Research, Institute for Biomedicine, Bolzano, Italy

Johannesburg, 28 January 2026

cristian.pattaro@eurac.edu

@cpattaro.bsky.social

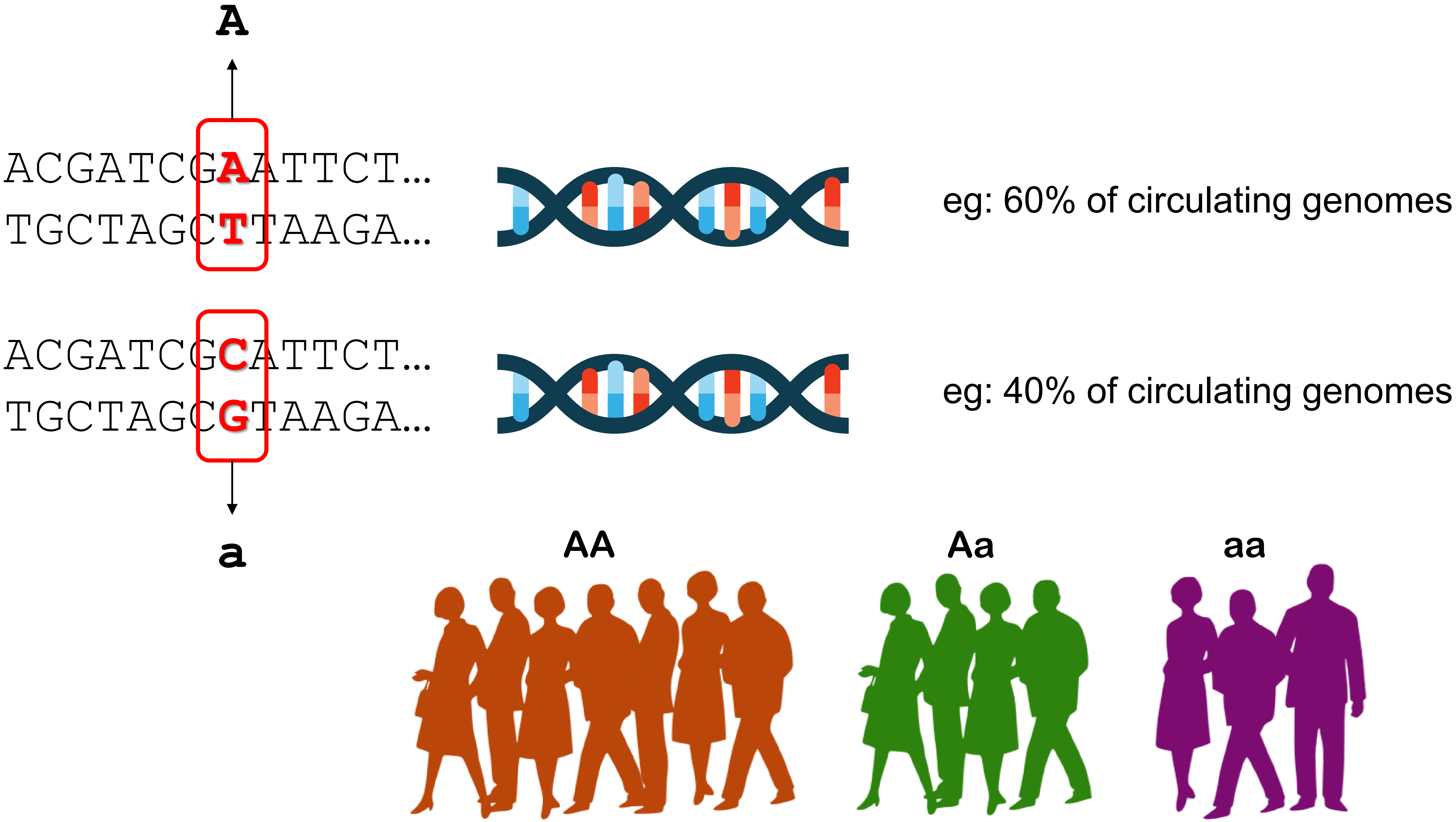


	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6
Person 1	G	T	G	A	A	T
Person 2	G	T	C	C	T	C
Person 3	C	A	G	C	A	C
Person 4	C	A	C	C	T	C

Andrew Morris
U Manchester

Uffelmann et al, Nat Rev Methods Primers 2021; 1:59

SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs)



A

ACGATCGAATTCT…
TGCTAGCTTAAGA…

eg: 60% of circulating genomes

ACGATCGCATTCT…
TGCTAGCGTAAGA…

eg: 40% of circulating genomes

a

AA

Aa

aa

Consider a **sample of N study participants**

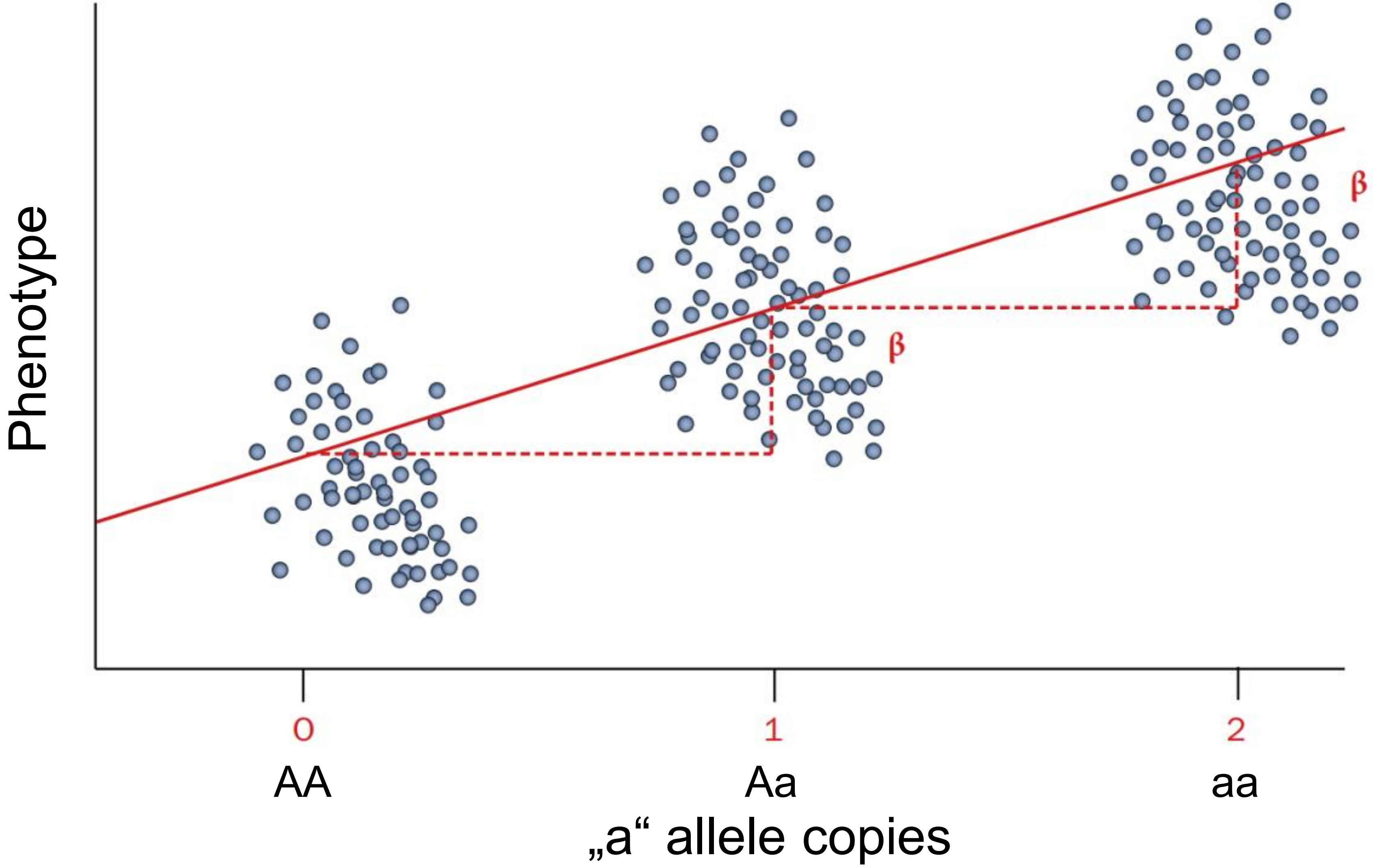
In each participant

- we measured a **quantitative trait or diagnosed a disease of interest**
- we **genotyped a few million SNPs**

- is any SNP associated with the trait of interest?

Association between SNP and quantitative trait

linear model, assuming an additive genetic model



O’Seaghdha & Fox, Nat Rev Nephrol (2012)

$\boldsymbol{P} \gg \boldsymbol{N}$ problem

→ fit one linear regression model per SNP:

$$\boldsymbol{y} = \beta_0 + \beta_i SNP_i + \boldsymbol{\varepsilon}$$

$$i = 1..some\ millions$$

$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{I})$

$H_0: \{\beta_i = 0\}$ vs $H_1: \{\beta_i \neq 0\}$

$\alpha = 5 \times 10^{-8}$

Often, the phenotype is **standardized** to mean = 0 & SD = 1 ➔

$$\boldsymbol{y} = \beta_i SNP_i + \boldsymbol{\varepsilon}$$

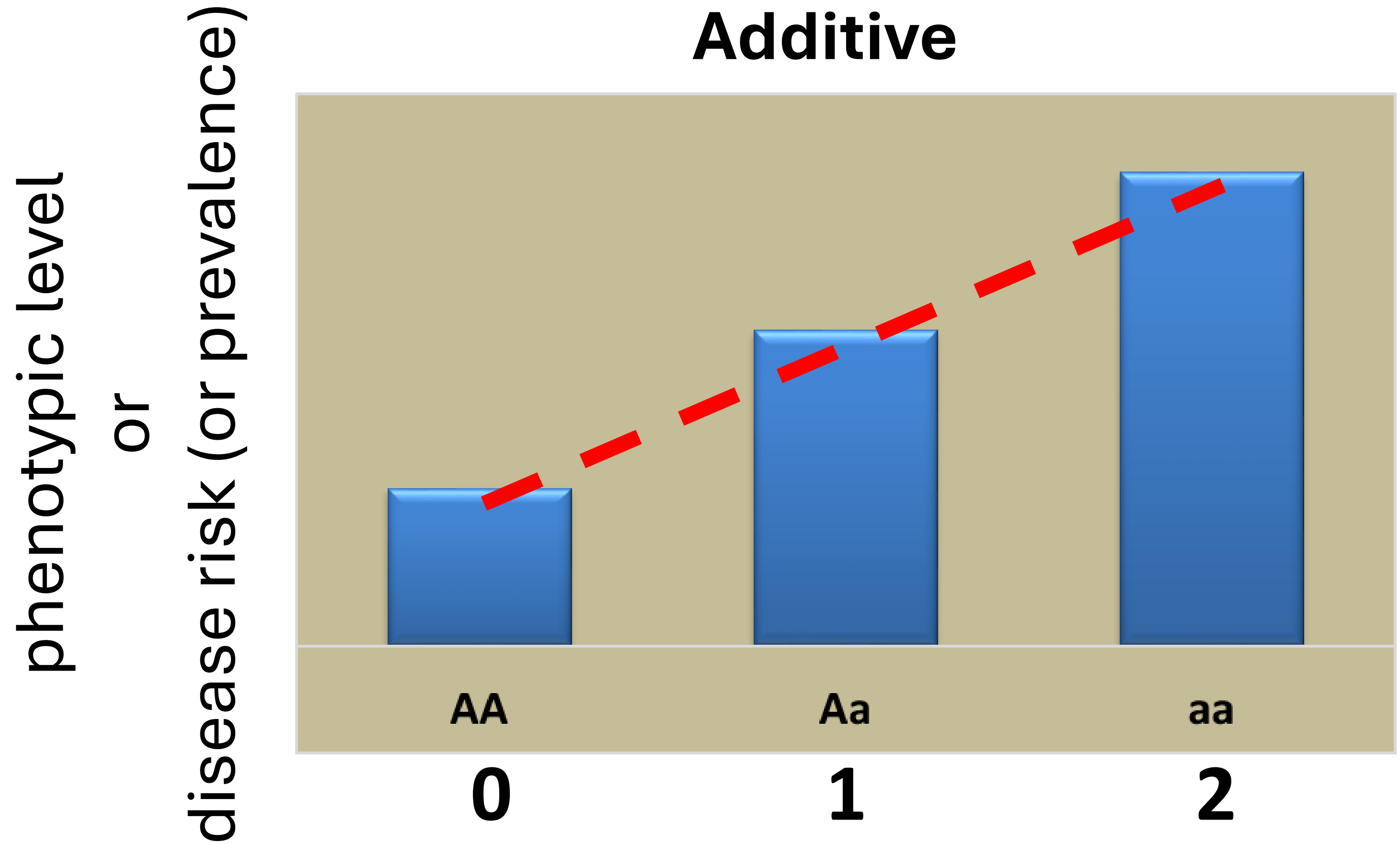
$$i = 1..some\ millions$$

$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{I})$

$H_0: \{\beta_i = 0\}$ vs $H_1: \{\beta_i \neq 0\}$

$\alpha = 5 \times 10^{-8}$

Genetic models

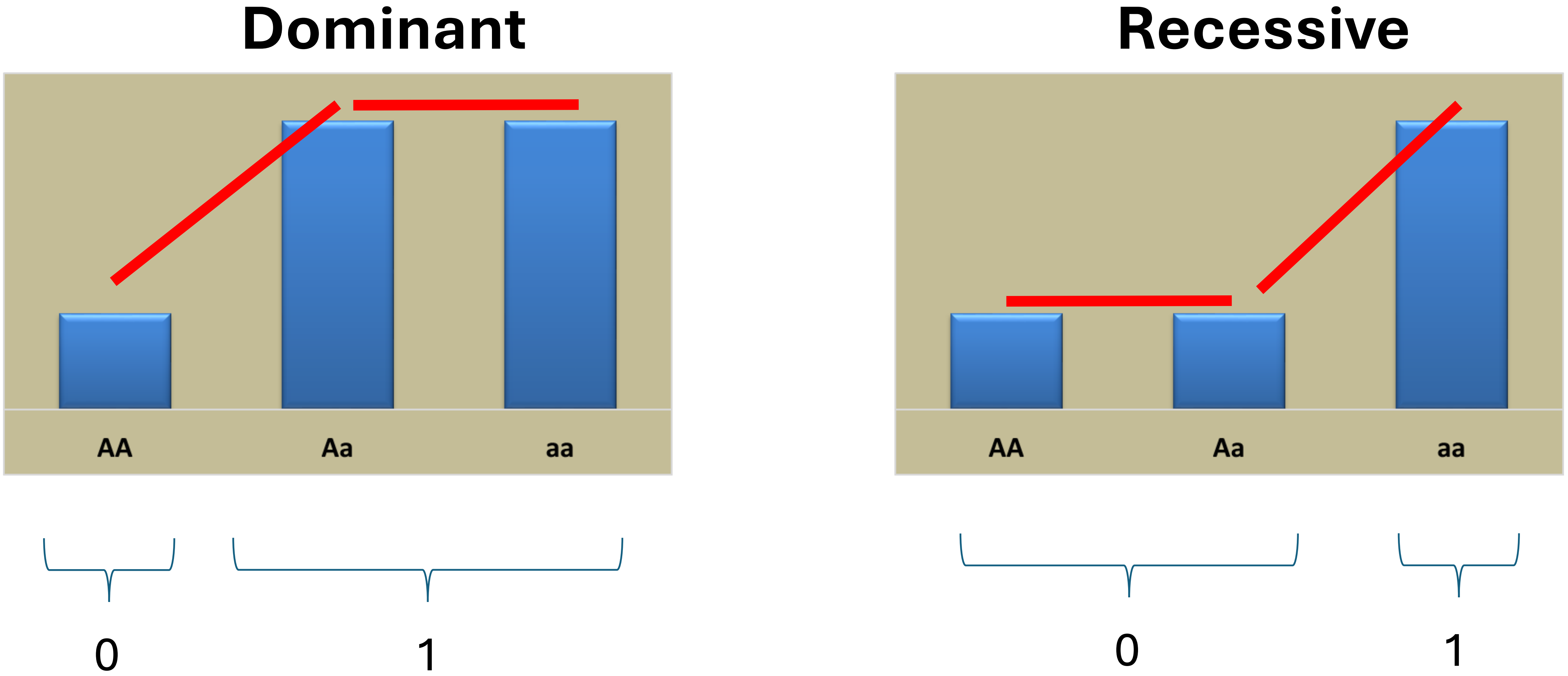


The phenotypic level* is increased proportionally

for each copy of the risk allele (dose effect)

*Either levels of a quantitative phenotpye (biomarker) or risk of a disease

Genetic models

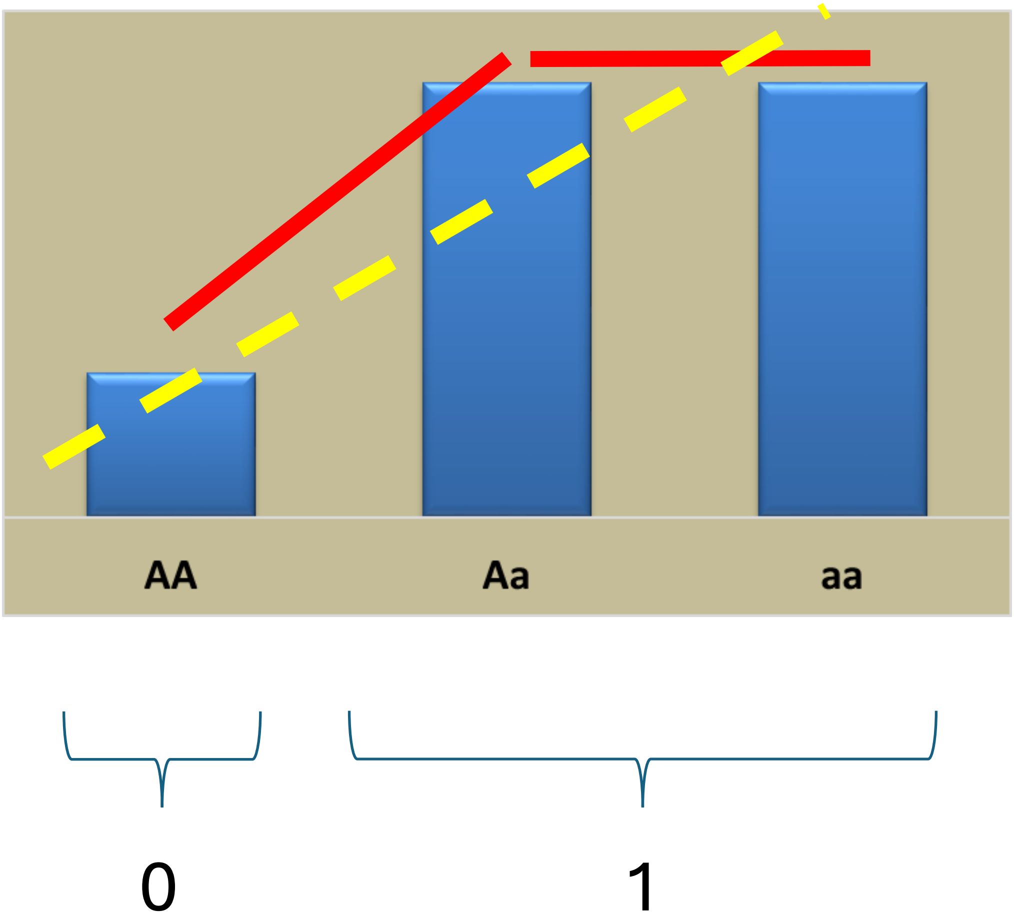


Dominant

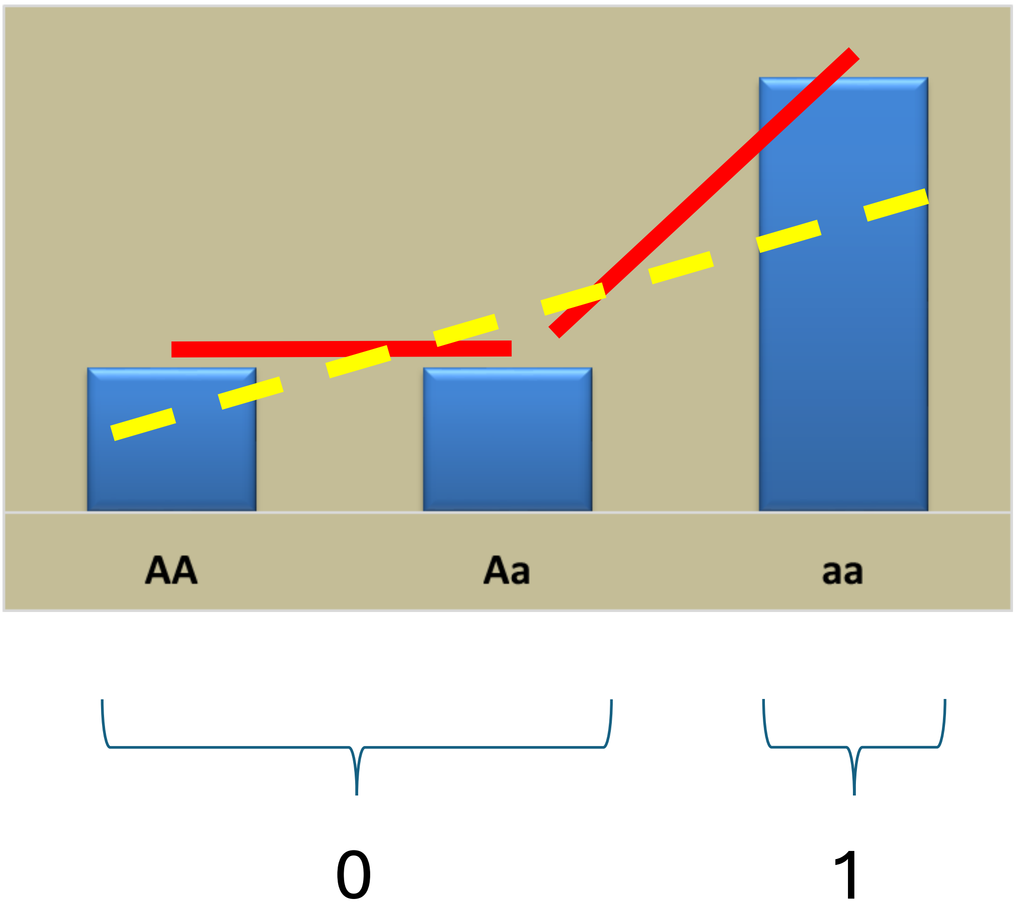
Recessive

Genetic models

Dominant



Recessive



Binary traits → logistic regression model

$$logit(\boldsymbol{p}) = \beta_0 + \beta_i SNP_i + \boldsymbol{\varepsilon}$$

$$i = 1..some\ millions$$

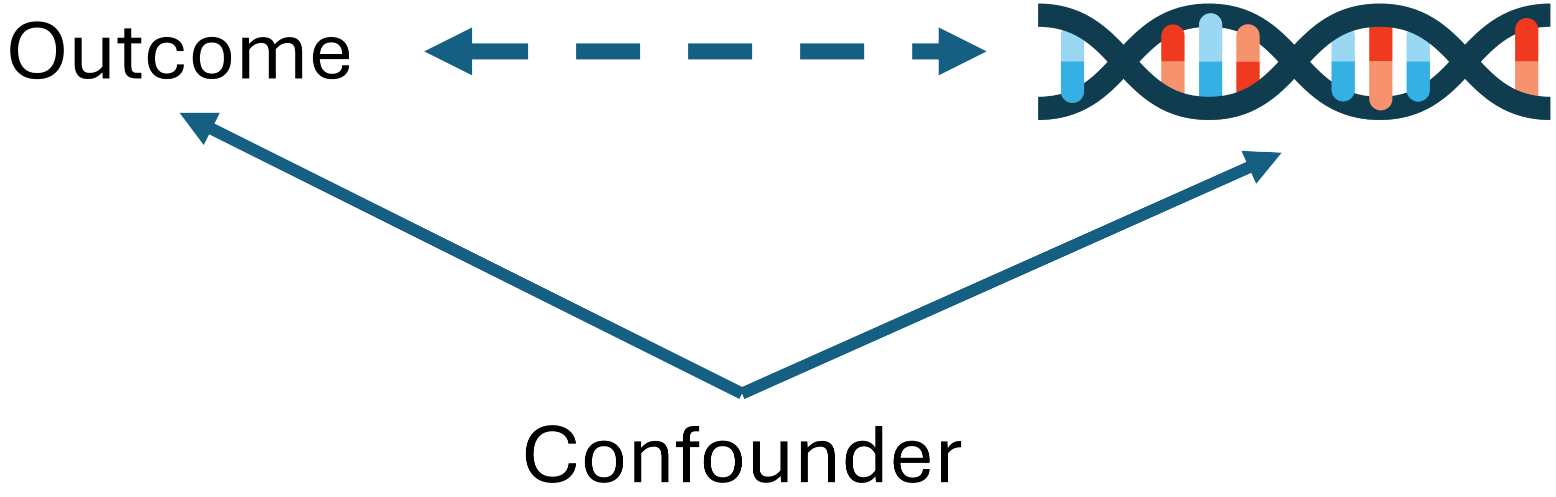
$$p = \Pr(disease)$$

$\boldsymbol{\varepsilon} \sim binomial$

$H_0: \{\beta_i = 0\}$ vs $H_1: \{\beta_i \neq 0\}$

$\alpha = 5 \times 10^{-8}$

Confounding



Outcome

Confounder

recruitment center

genotyping batch (batch effect)

between-individual relatedness

population stratification...

Population stratification = presence of more than one genotypic group hidden within the study sample.

Population stratification happens if both the two following conditions are realized:

1. Subgroups have different genotype frequency
2. Subgroups have different disease prevalence (*when studying quantitative phenotypes there will always be a difference*!)

When paired up with different disease prevalence, population stratification can cause spurious associations = false findings = false positive results.

[**see appendix**]

Linear mixed regression model

$$\boldsymbol{y} = \beta\, \boldsymbol{SNP} \quad + \sum_{j=1}^{K} \gamma_j \boldsymbol{x}_j + \boldsymbol{g} + \boldsymbol{e}$$

$$V = Cov(\boldsymbol{y}) = Cov(\boldsymbol{g}) + Cov(\boldsymbol{e}) = \sigma_1^2 K + \sigma_2^2 I$$

K = relatedness matrix, estimated from genetic data (pairwise coefficients of „distance“ between individuals), usually K$_{LOCO}$

$$H_0 \colon \beta = 0 \;\; vs \;\; H_1 \colon \beta \neq 0$$

$$\chi^2_{LMM} = \frac{(SNP'V^{-1}y)^2}{(SNP'V^{-1}SNP)}$$

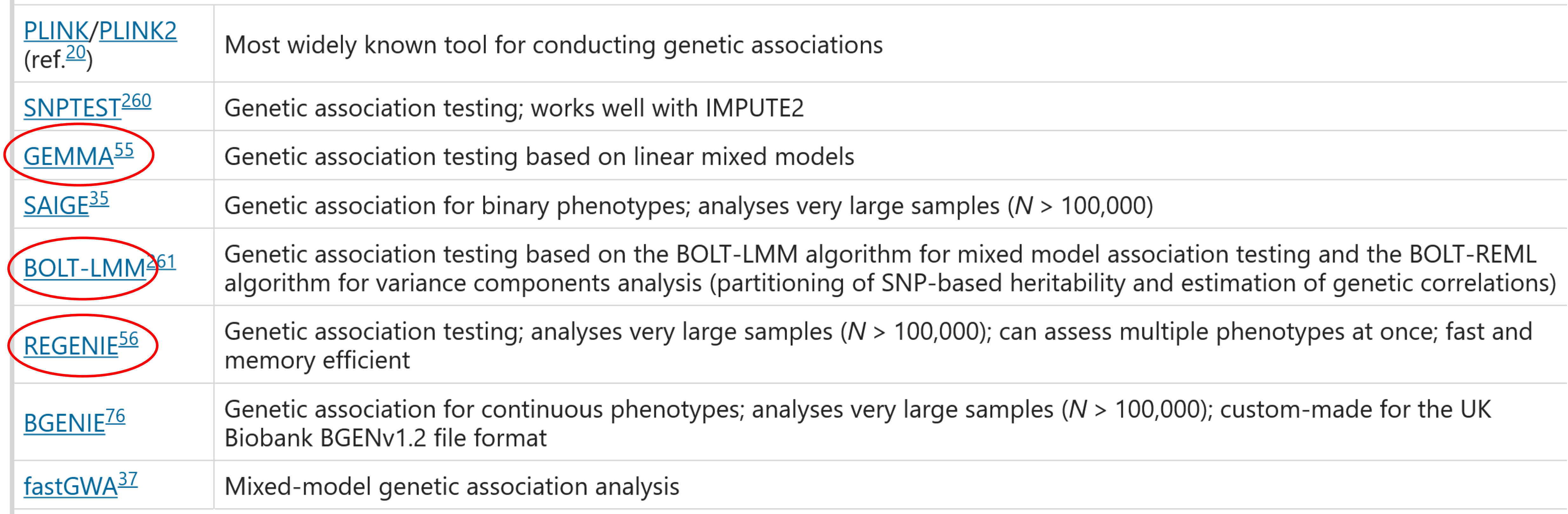
Loh et al, Nat Genet 2015; 47(3): 284-290

Adjustments

Relatedness is always there → correct for it, even when individuals are apparently unrelated

Covariates: only technical covariates; do not adjust for variables that could be in the causal pathway unless there is a special reason

SOFTWARE



PLINK/PLINK2 (ref.[20])	Most widely known tool for conducting genetic associations
SNPTEST[260]	Genetic association testing; works well with IMPUTE2
GEMMA[55]	Genetic association testing based on linear mixed models
SAIGE[35]	Genetic association for binary phenotypes; analyses very large samples (*N* > 100,000)
BOLT-LMM[261]	Genetic association testing based on the BOLT-LMM algorithm for mixed model association testing and the BOLT-REML algorithm for variance components analysis (partitioning of SNP-based heritability and estimation of genetic correlations)
REGENIE[56]	Genetic association testing; analyses very large samples (*N* > 100,000); can assess multiple phenotypes at once; fast and memory efficient
BGENIE[76]	Genetic association for continuous phenotypes; analyses very large samples (*N* > 100,000); custom-made for the UK Biobank BGENv1.2 file format
fastGWA[37]	Mixed-model genetic association analysis

Uffelmann et al, Nat Rev Methods Primers 2021; 1:59

References

Loh et al. **Efficient Bayesian mixed model analysis increases association power in large cohorts.** Nat Genet. 2015 Feb 2;47(3):284–290. https://pmc.ncbi.nlm.nih.gov/articles/PMC4342297/

Mbatchou et al, **Computationally efficient whole-genome regression for quantitative and binary traits**. Nat Genet . 2021 Jul;53(7):1097-1103. doi: 10.1038/s41588-021-00870-7.
https://pubmed.ncbi.nlm.nih.gov/34017140/ , https://rgcgithub.github.io/regenie/

Results

Allele on which the effect is calculated

Effect estimate: change of the phenotype level for each copy of the „effect allele“

Standard error of *b*

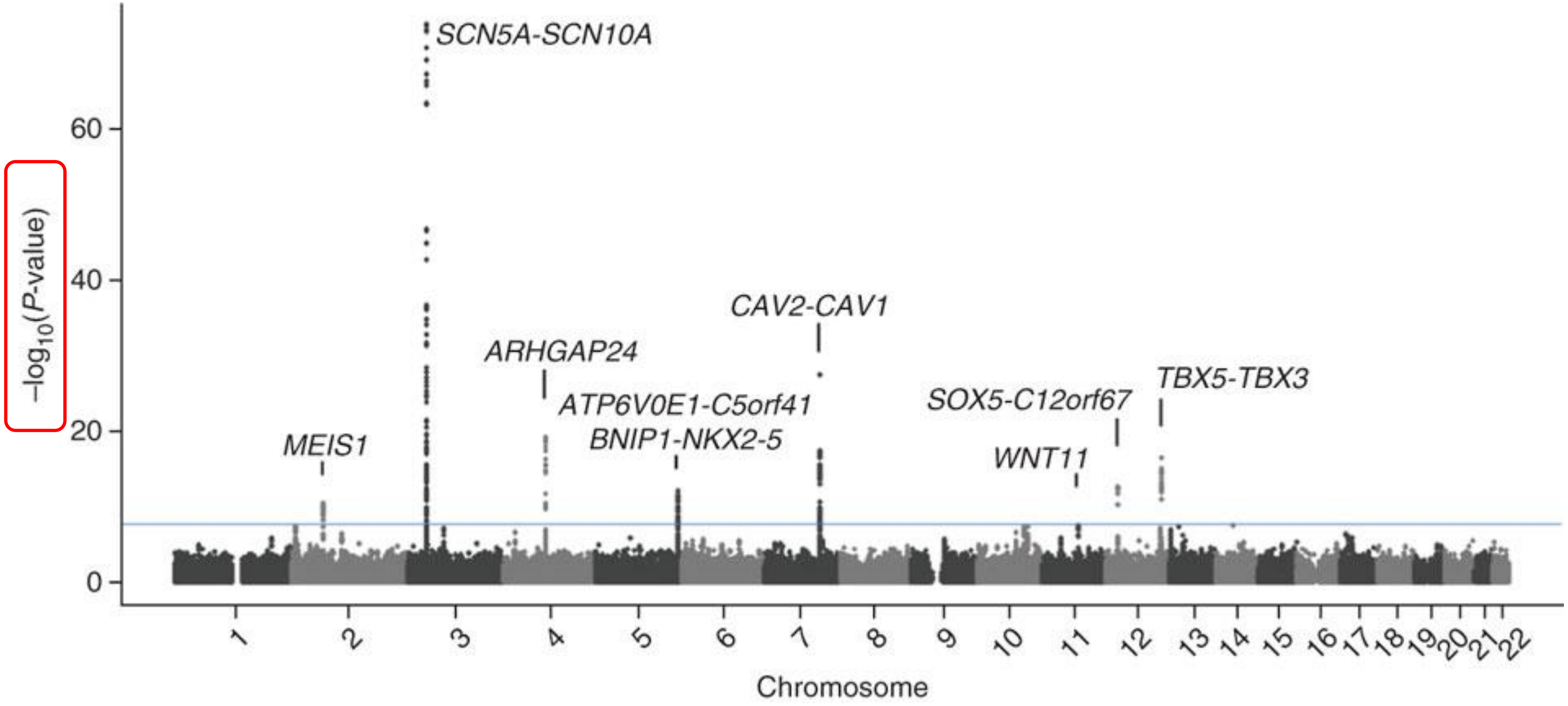
P-value: evidence of association

#	SNP ID	Chr	Pos	Effect Allele	Other Allele	Effect Allele Frequency	b	SE(b)	P-value
							OR	SE(OR)	
1	rs…	…	…	…	…		…	…	…
2	rs…	…	…	…	…		…	…	…
3	rs…	…	…	…	…		…	…	…
		…	…	…	…		…	…	…
		…	…	…	…		…	…	…
M	rs…	…	…	…	…		…	…	…

Example: Genome-wide scan of total cholesterol levels

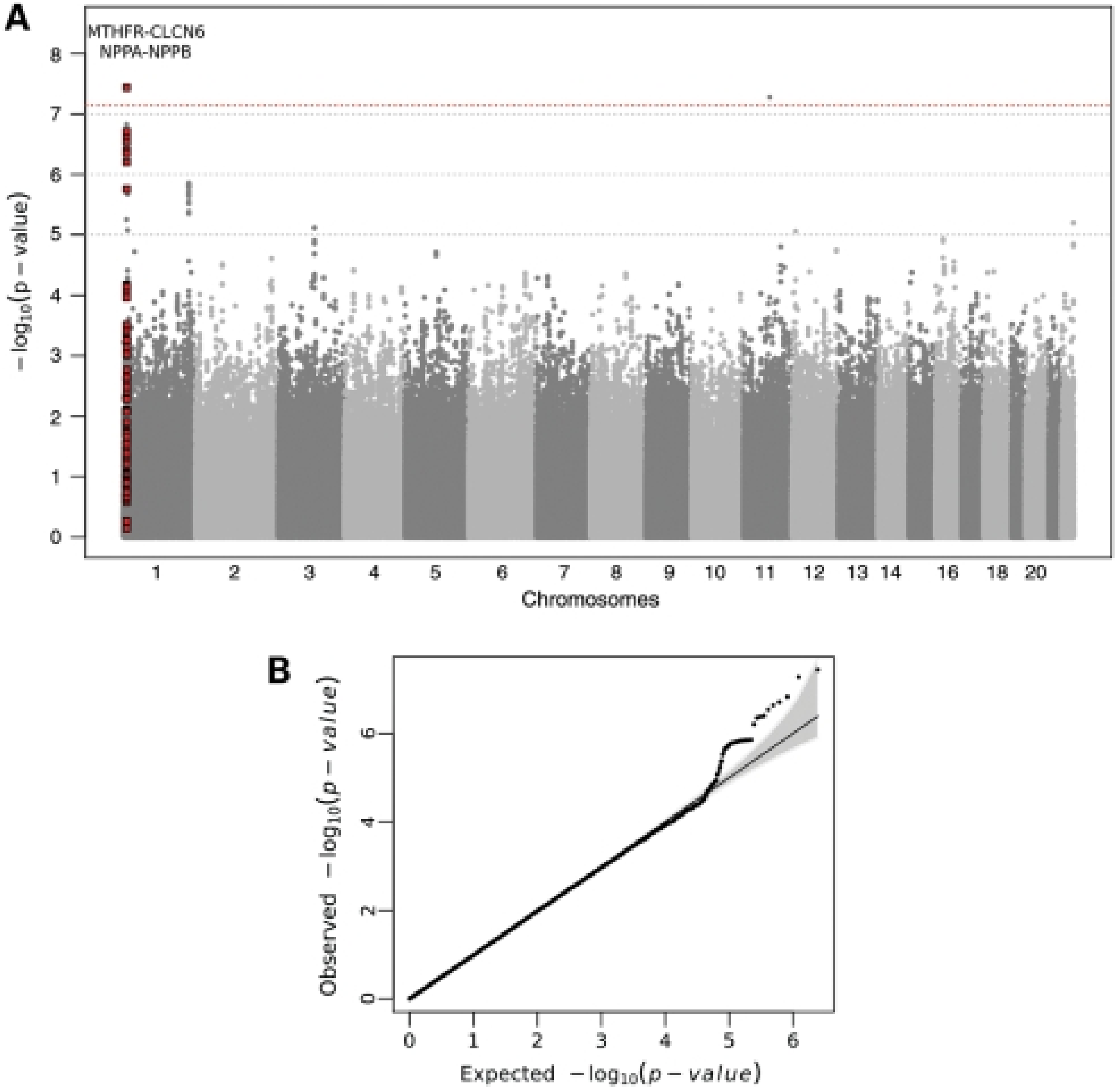
#	SNP ID	Chr	Pos	Eff. All.	Other All.	Eff. All. Freq.	b	SE(b)	P-value
1	rs1	1	56,023	T	A	0.40	0.101	0.600	0.8663211
2	rs2	1	70,231	G	C	0.23	-3.302	5.302	0.5334266
3	rs3	1	75,444	G	A	0.05	1.432	1.500	0.3397463
	…	…	…	…	…		…	…	…
		…	…	…	…		…	…	…
10,000,000	rs120137103	22	…	C	T	0.11	2.512	8.230	0.760195

Manhattan Plot



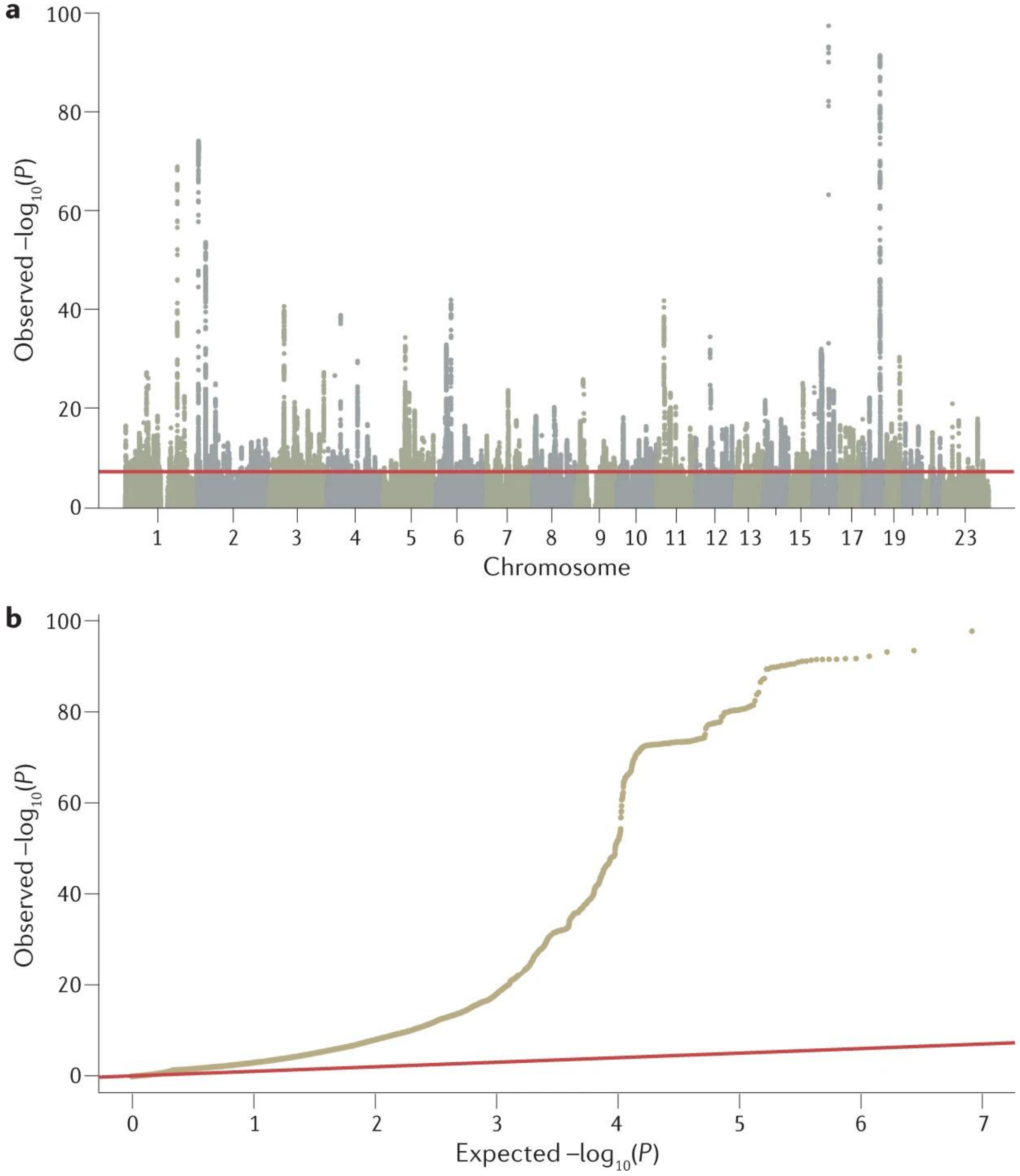
Pfeufer. Nature Genetics, 42, 153–159(2010)

Small sample size



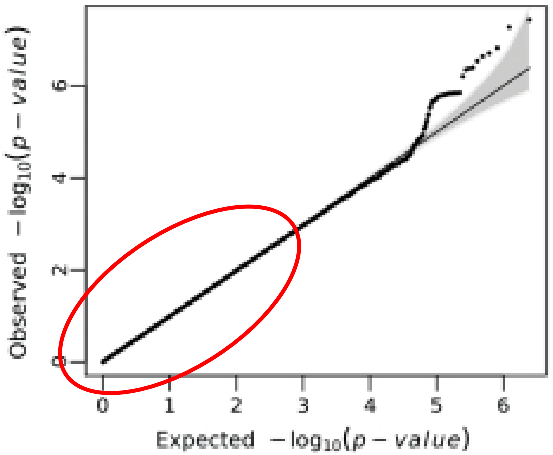
Del Greco et al, Hum Mol Genet 2011; 20:1660-71

Large sample size



Uffelmann et al, Nat Rev Methods Primers 2021; 1:59

Estimating the genomic control inflation factor, λ_{GC}



Under the null hypothesis of no association,

if N is large,

$$t_i = \frac{\beta_i}{SE(\beta_i)} \sim N(0,1)$$

$$t_i^2 \sim \chi_1^2$$

$$\lambda_{GC} = \frac{median(t_1^2, \ldots.t_S^2)}{\chi_1^2(0.5)} = \begin{cases} > 1\ inflation \\ 1, no\ inflation \\ < 1, deflation \end{cases}$$

$$S = no.of\ SNPs$$

Genomic control for association studies. Devlin B, Roeder K. Biometrics. 1999 Dec;55(4):997-1004.

LD-score regression

Alternative (better) way to assess inflation

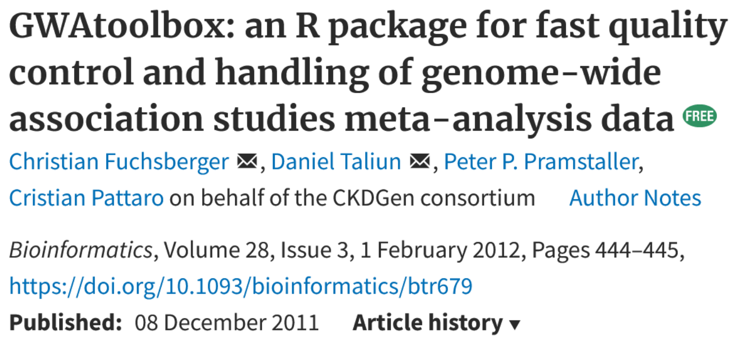
If a trait is truly polygenic, SNPs with more neighbours (higher LD scores) should, on average, show stronger associations than SNPs with fewer neighbors.

By regressing the χ2 test statistics from GWAS vs LD Scores (LD Score regression), the **intercept** minus one estimates the mean contribution of confounding bias to the inflation in the test statistics.

Bulik-Sullivan et al, Nat Genet 2015

Other checks

- Does λ_{GC} depends on imputation quality (IQ) or minor allele frequency (MAF)? → re-calculate after filtering for IQ and/or MAF
- Is the distribution of p-values ~ Uniform?
- Is the distribution of b/SE(b) ~ symmetric?

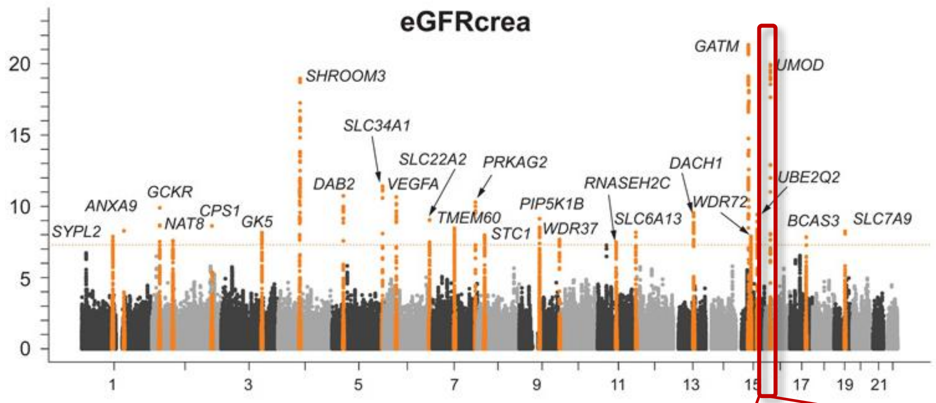


GWAtoolbox: an R package for fast quality control and handling of genome-wide association studies meta-analysis data FREE

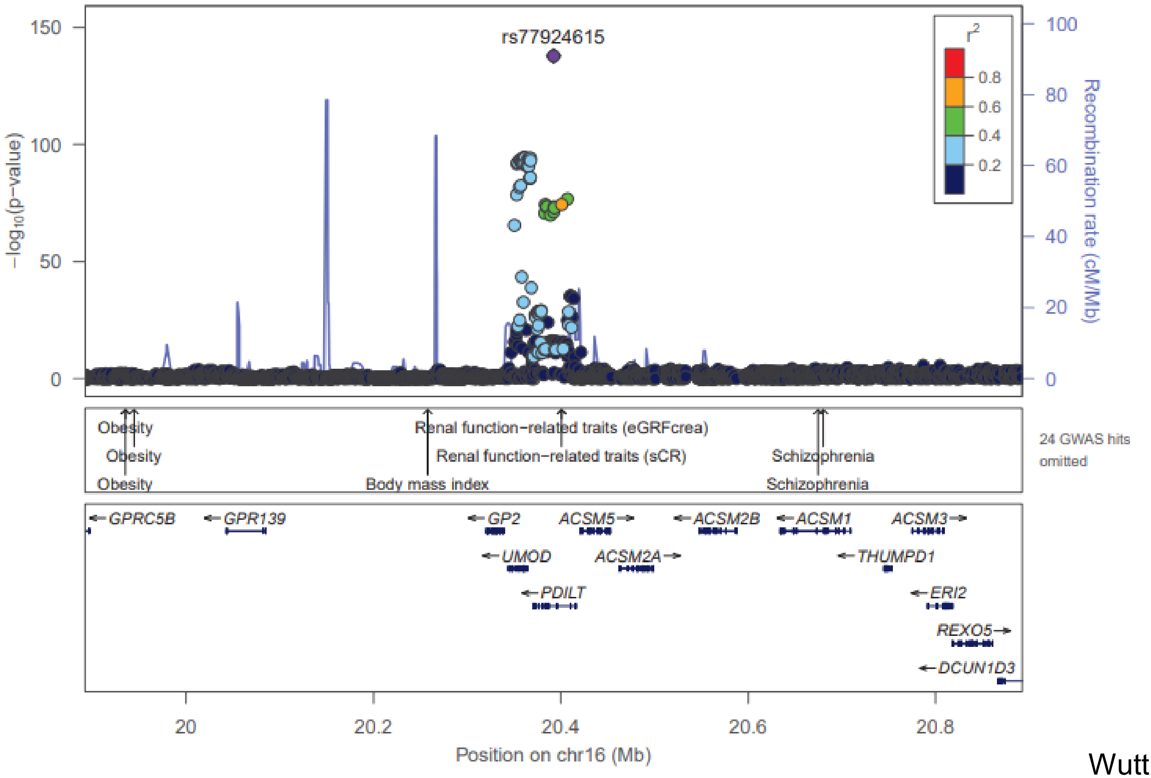
Christian Fuchsberger ✉, Daniel Taliun ✉, Peter P. Pramstaller, Cristian Pattaro on behalf of the CKDGen consortium Author Notes

Bioinformatics, Volume 28, Issue 3, 1 February 2012, Pages 444–445, https://doi.org/10.1093/bioinformatics/btr679

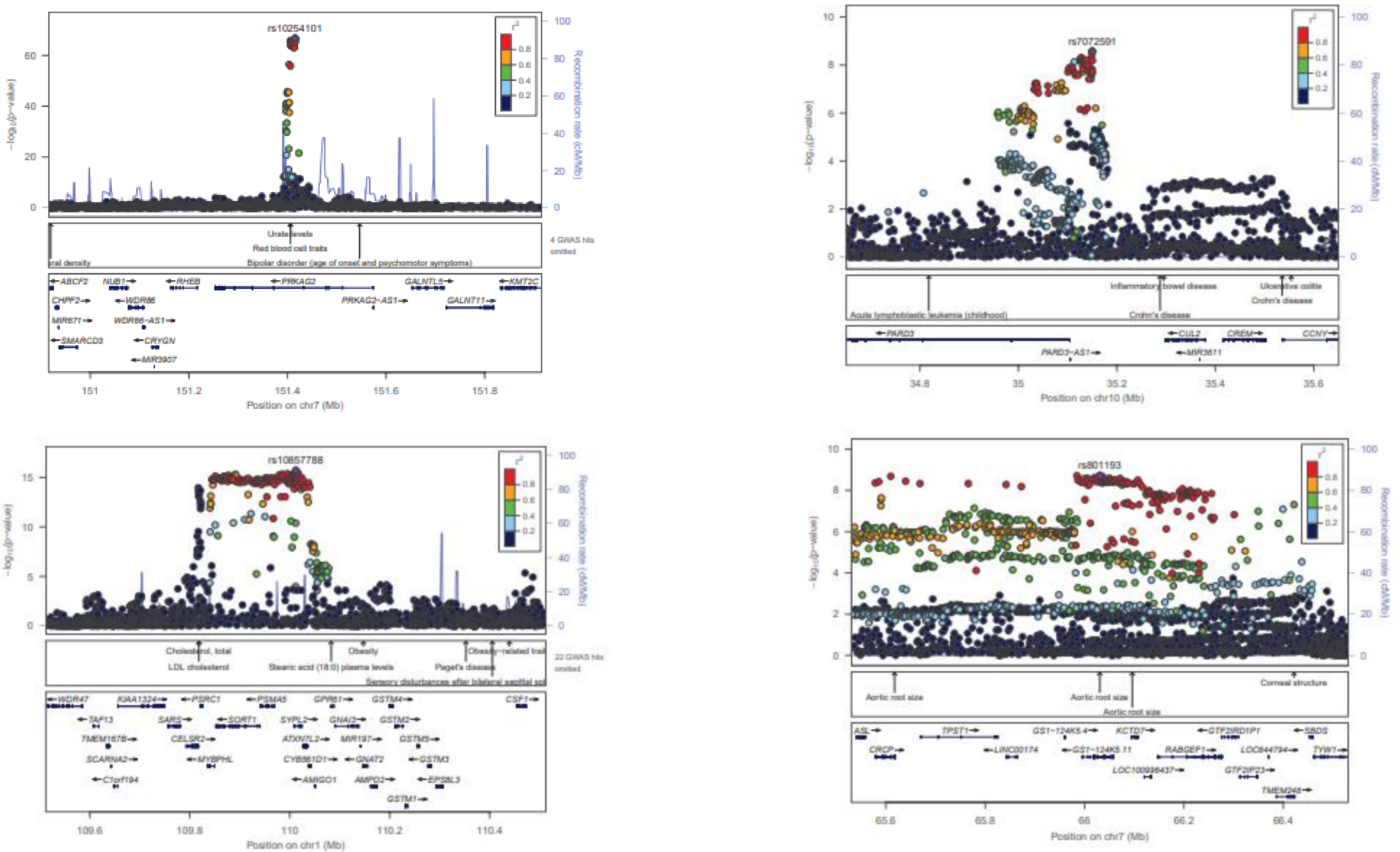
Published: 08 December 2011 **Article history** ▾



Köttgen et al. *Nat Genet* 2010



Wuttke et al. *Nat Genet* 2019



Wuttke et al, Nat Genet 2019

Meta-analysis and replication

Genome-wide association studies have introduced a different way to look at meta-analysis

Because of privacy issues, studies do not share genetic data (typically)

Thus, meta-analysis is used as a tool to increase the sample size and the power of a study

Rather than a method to assess the average effect

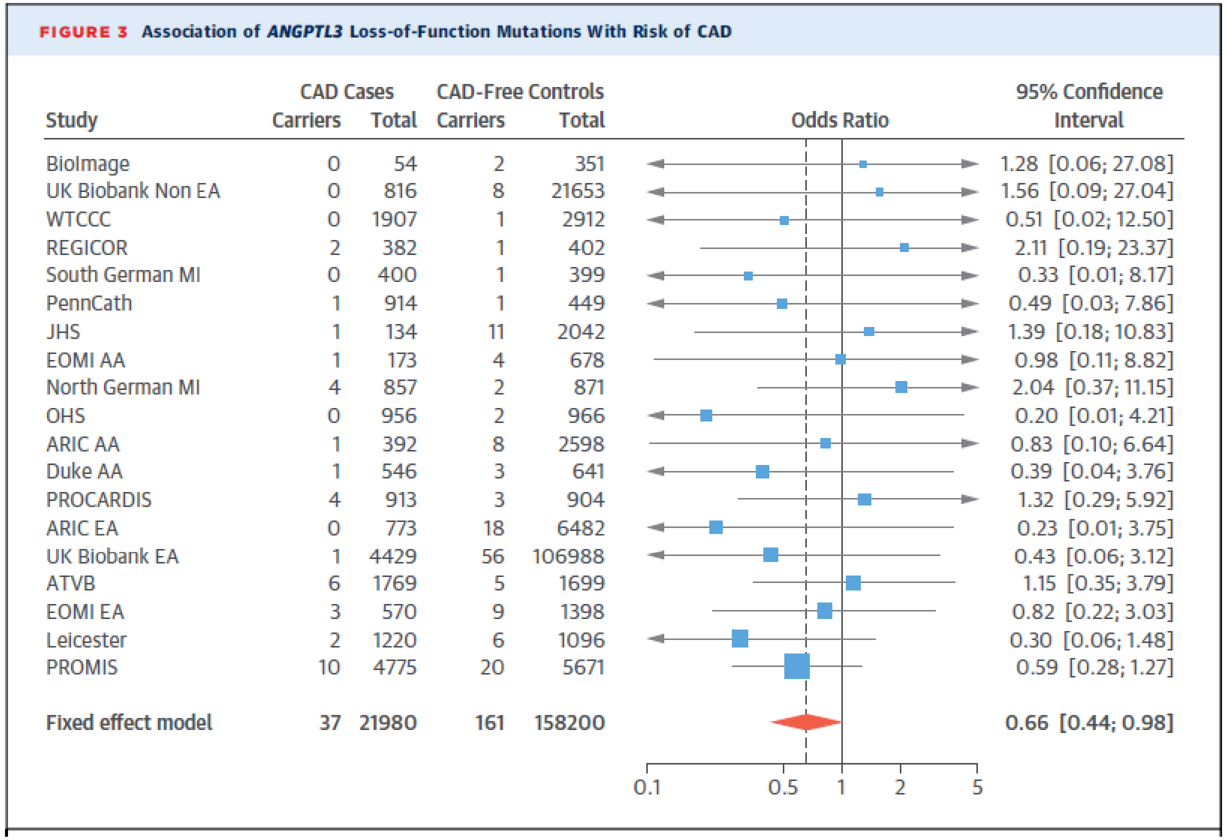


FIGURE 3 Association of *ANGPTL3* Loss-of-Function Mutations With Risk of CAD

	CAD Cases		CAD-Free Controls			95% Confidence
Study	Carriers	Total	Carriers	Total	Odds Ratio	Interval
BioImage	0	54	2	351		1.28 [0.06; 27.08]
UK Biobank Non EA	0	816	8	21653		1.56 [0.09; 27.04]
WTCCC	0	1907	1	2912		0.51 [0.02; 12.50]
REGICOR	2	382	1	402		2.11 [0.19; 23.37]
South German MI	0	400	1	399		0.33 [0.01; 8.17]
PennCath	1	914	1	449		0.49 [0.03; 7.86]
JHS	1	134	11	2042		1.39 [0.18; 10.83]
EOMI AA	1	173	4	678		0.98 [0.11; 8.82]
North German MI	4	857	2	871		2.04 [0.37; 11.15]
OHS	0	956	2	966		0.20 [0.01; 4.21]
ARIC AA	1	392	8	2598		0.83 [0.10; 6.64]
Duke AA	1	546	3	641		0.39 [0.04; 3.76]
PROCARDIS	4	913	3	904		1.32 [0.29; 5.92]
ARIC EA	0	773	18	6482		0.23 [0.01; 3.75]
UK Biobank EA	1	4429	56	106988		0.43 [0.06; 3.12]
ATVB	6	1769	5	1699		1.15 [0.35; 3.79]
EOMI EA	3	570	9	1398		0.82 [0.22; 3.03]
Leicester	2	1220	6	1096		0.30 [0.06; 1.48]
PROMIS	10	4775	20	5671		0.59 [0.28; 1.27]
Fixed effect model	**37**	**21980**	**161**	**158200**		**0.66 [0.44; 0.98]**

Stitziel et al, J Am Coll Cardiol 2017

GWAS are typically conducted out in the context of a **consortium**

No. of studies: 2-3… 120-150…

No. of pooled samples: 1000 … >5 M

Crucial steps for GWAS meta-analyses:

- Centralized analysis plan, distributed to all partners
- Analysis plan must be tested before
- Post-phenotype preparation and post-GWAS QC should be centralized

FIXED-EFFECTS META-ANALYSIS BASED ON INVERSE-VARIANCE WEIGHTING

STUDY-1	*STUDY-2*	*STUDY-K*
$b_{1,\boldsymbol{1}}$, $SE(b_{1,\boldsymbol{1}})$	$b_{1,\boldsymbol{2}}$, $SE(b_{1,\boldsymbol{2}})$	$b_{1,\boldsymbol{K}}$, $SE(b_{1,\boldsymbol{K}})$
$b_{2,\boldsymbol{1}}$, $SE(b_{2,\boldsymbol{1}})$	$b_{2,\boldsymbol{2}}$, $SE(b_{2,\boldsymbol{2}})$	$b_{2,\boldsymbol{K}}$, $SE(b_{2,\boldsymbol{K}})$
…	…	…
…	…	…
$b_{2500000,\boldsymbol{1}}$, $SE(b_{2500000,\boldsymbol{1}})$	$b_{2500000,\boldsymbol{2}}$, $SE(b_{2500000,\boldsymbol{2}})$	$b_{2500000,\boldsymbol{K}}$, $SE(b_{2500000,\boldsymbol{K}})$

$$b = \frac{\dfrac{b_1}{SE(b_1)^2} + \dfrac{b_2}{SE(b_2)^2} + .. + \dfrac{b_K}{SE(b_K)^2}}{\dfrac{1}{SE(b_1)^2} + \dfrac{1}{SE(b_2)^2} + .. + \dfrac{1}{SE(b_K)^2}}$$

FOREST PLOT: the best way to visualize a meta-analysis

A tool to assess homogeneity of the SNP-trait association between different studies

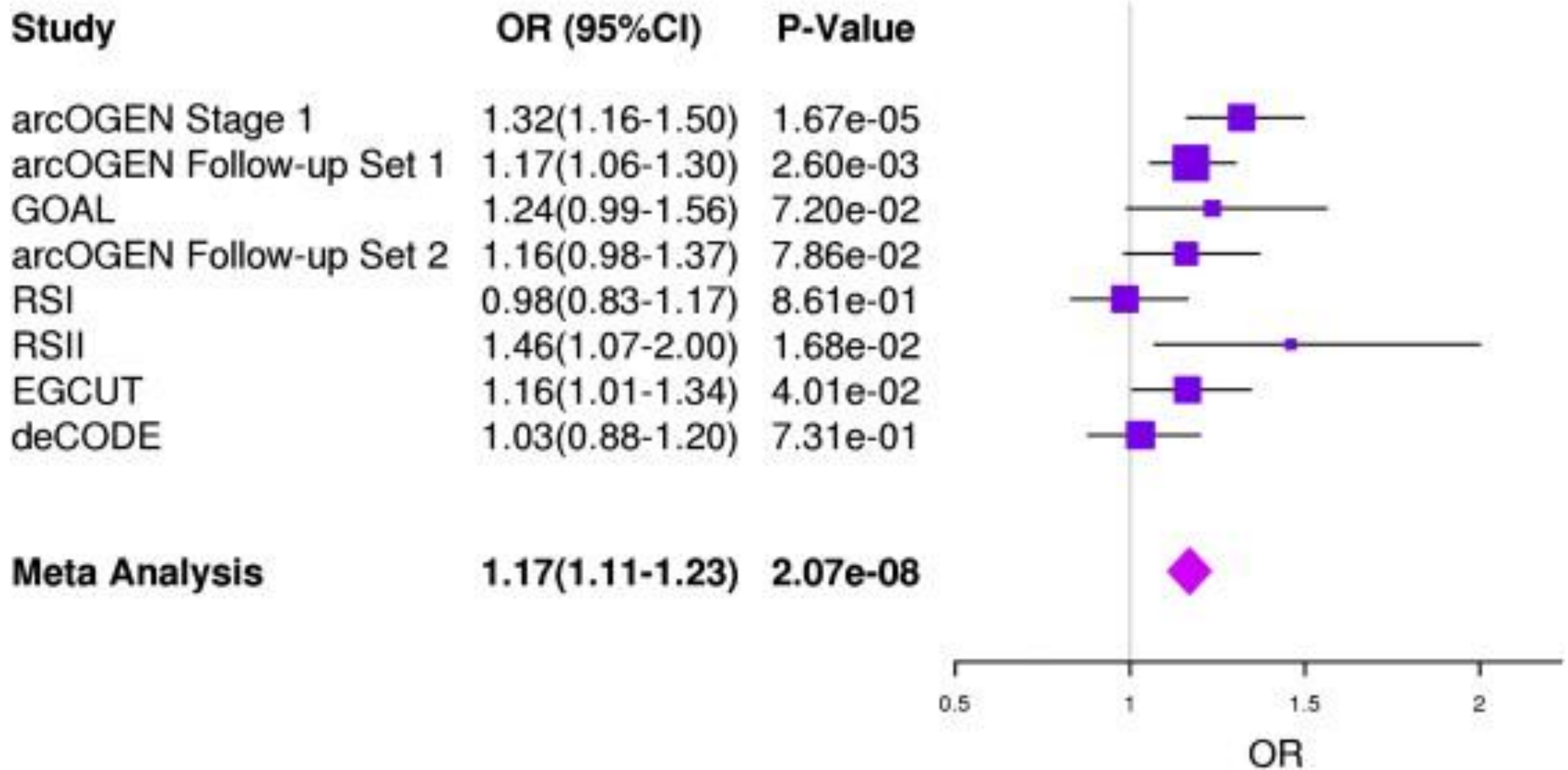
AJHG

Volume 89, Issue 3, 9 September 2011, Pages 446–450

Report

A Variant in *MCF2L* Is Associated with Osteoarthritis

Association between a genetic variant in MCF2L and osteoarthritis

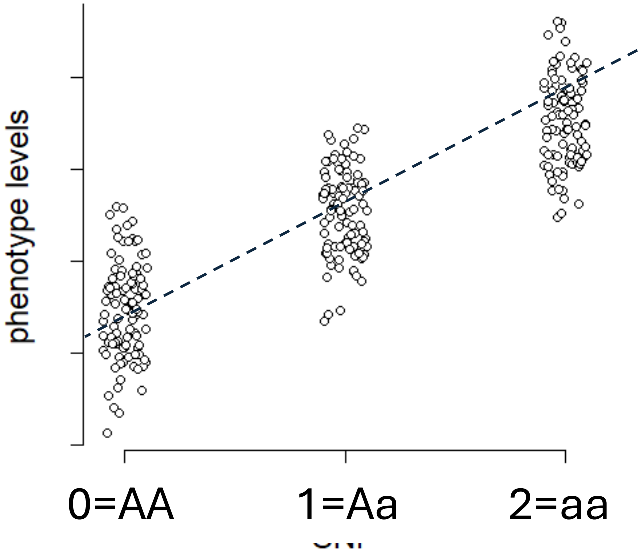


Study	OR (95%CI)	P-Value
arcOGEN Stage 1	1.32(1.16-1.50)	1.67e-05
arcOGEN Follow-up Set 1	1.17(1.06-1.30)	2.60e-03
GOAL	1.24(0.99-1.56)	7.20e-02
arcOGEN Follow-up Set 2	1.16(0.98-1.37)	7.86e-02
RSI	0.98(0.83-1.17)	8.61e-01
RSII	1.46(1.07-2.00)	1.68e-02
EGCUT	1.16(1.01-1.34)	4.01e-02
deCODE	1.03(0.88-1.20)	7.31e-01
Meta Analysis	**1.17(1.11-1.23)**	**2.07e-08**

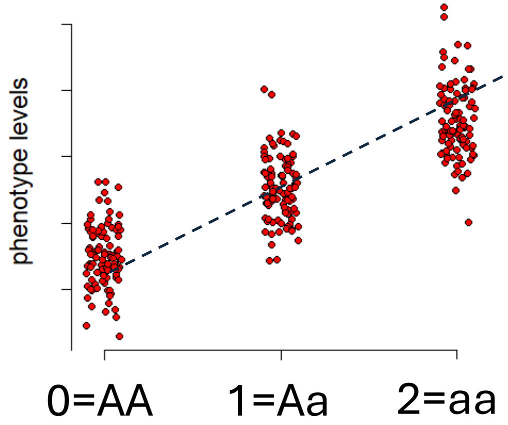
REPLICATION

1. From the discovery GWAS, identify the significant hits (take the most associated SNP for each locus)
2. Find a **similar** study, which **must have the same phenotype** (**!!!**), with **adequate sample size [power calculation given the minor allele frequency is recommended]**, and ask if they can analyze your SNPs in their sample.
3. Verify if the **same allele** at each SNP is associated with the trait in the same direction (you can use a 1-sided test, with significance level of 0.05 / number of SNPs being tested)

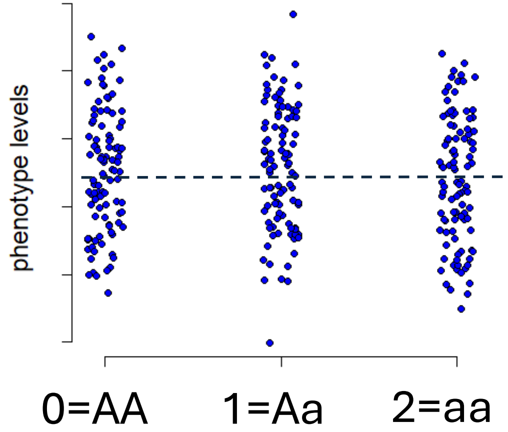
Discovery study



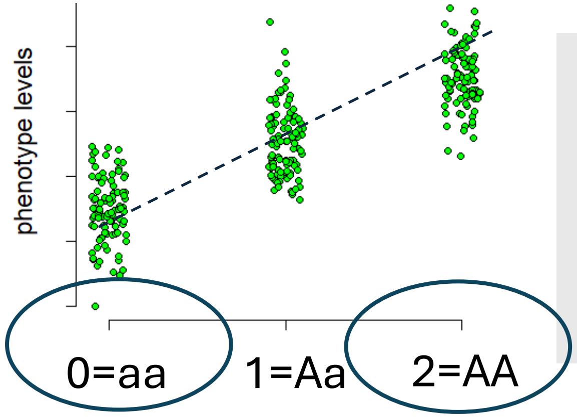
Replicatio n study



Replication! effect in the same direction



No replication: no effect

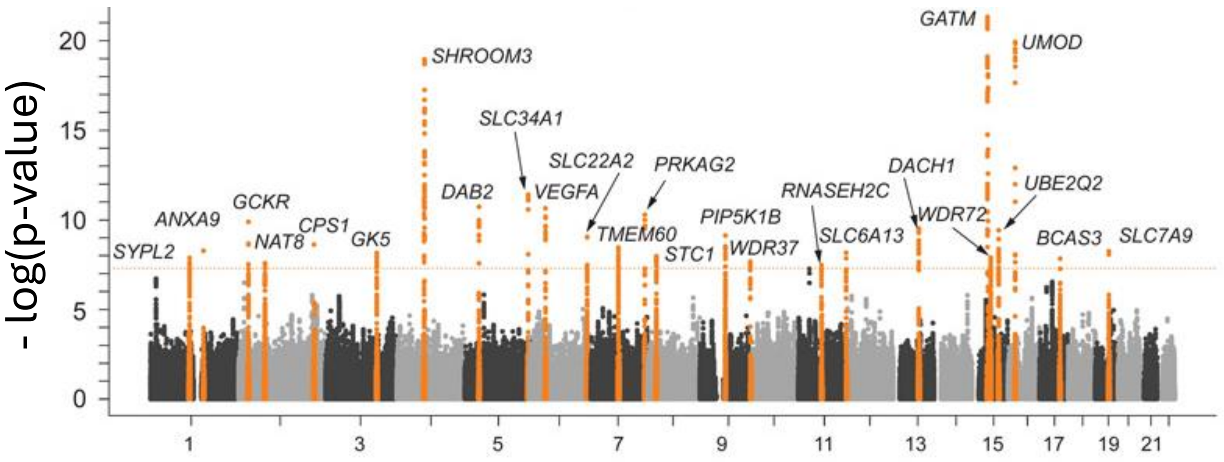


No replication: effect in the opposite direction (allele swapping)

EXAMPLES

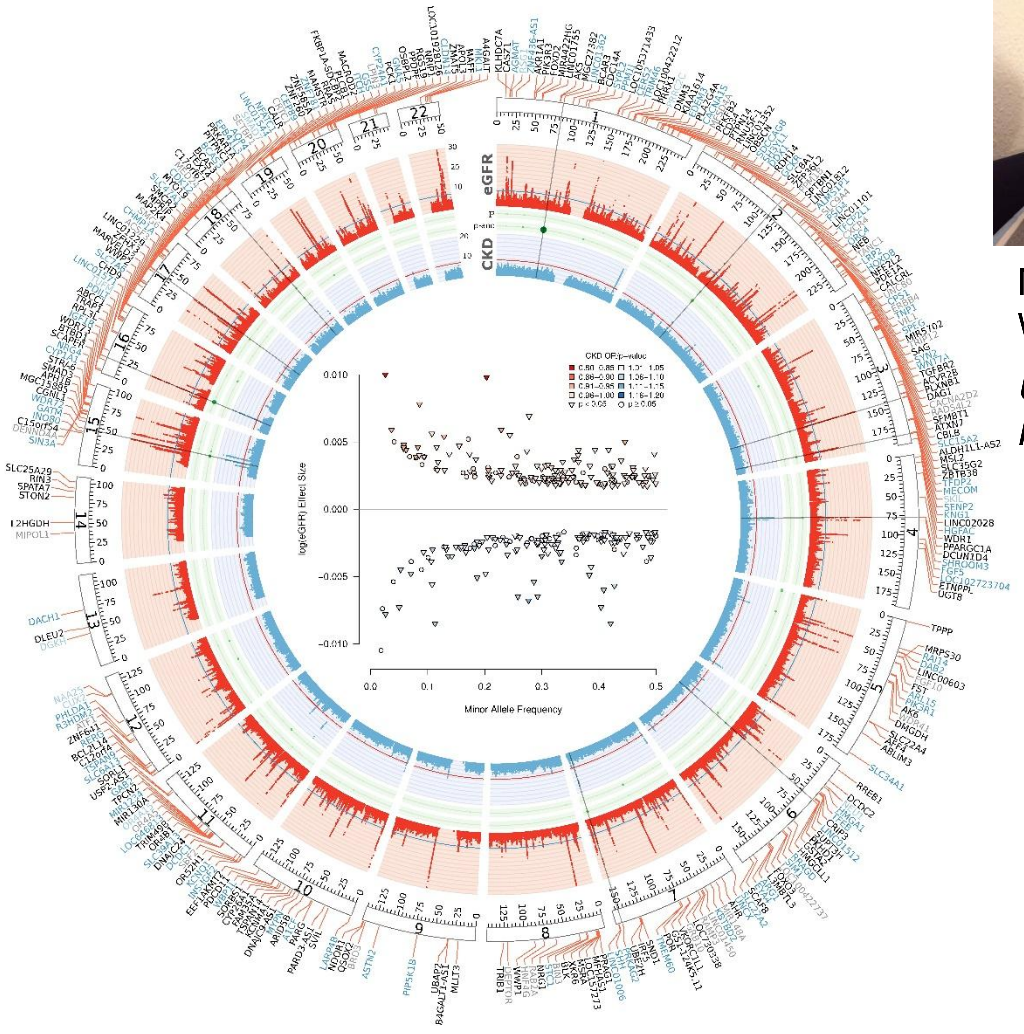
CKD Gen CONSORTIUM

Genetic loci for eGFRcrea



- Outcome: estimated glomerular filtration rate
- N = **67,093** (discovery) + 22,982 (Replication)
- **2.5 Mio** SNPs
- 24 loci associated with kidney function

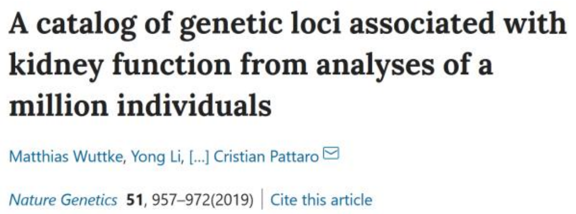
Köttgen et al. Nat Genet 2010



Matthias Wuttke

University of Freiburg

- N = **765,348** (discovery) + 280,722 (Replication)
- **8.2 Mio** SNPs
- 246 loci associated

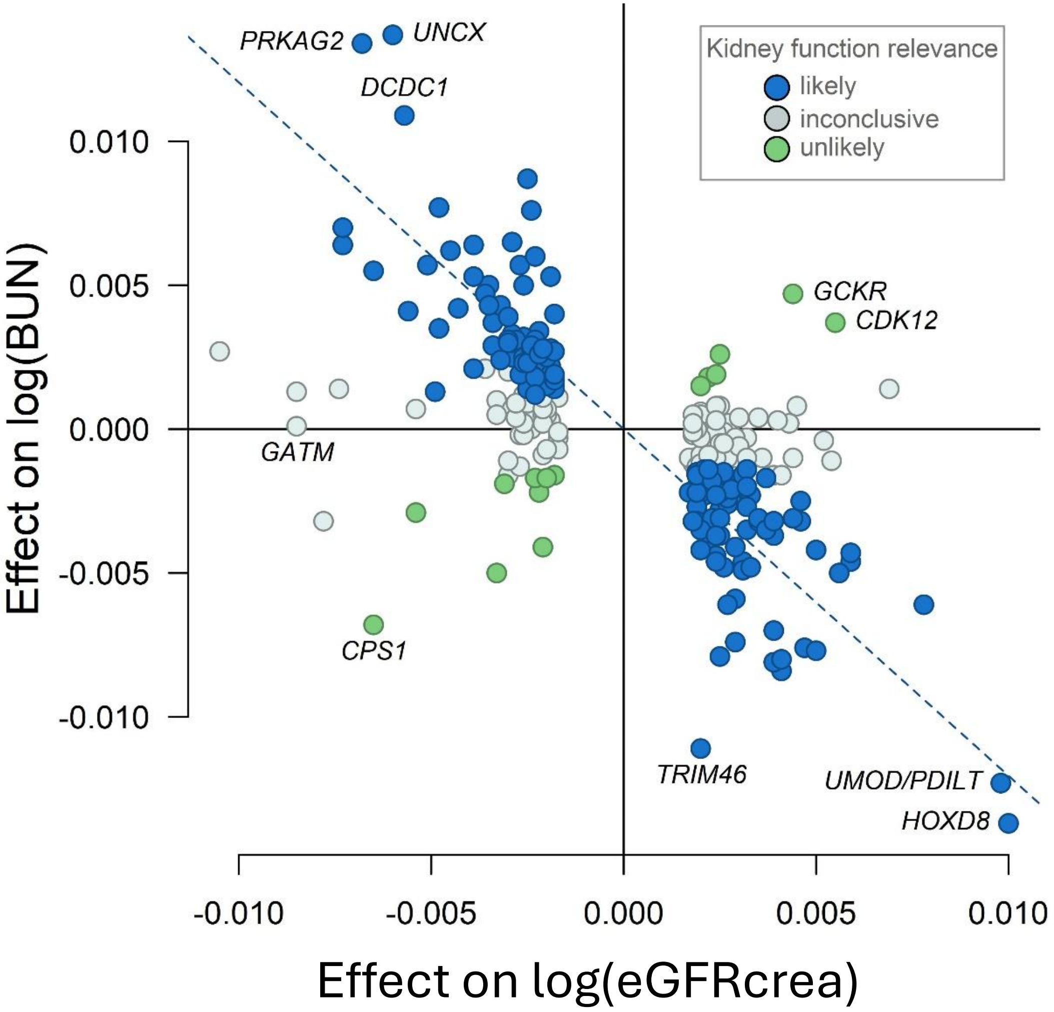


A catalog of genetic loci associated with kidney function from analyses of a million individuals

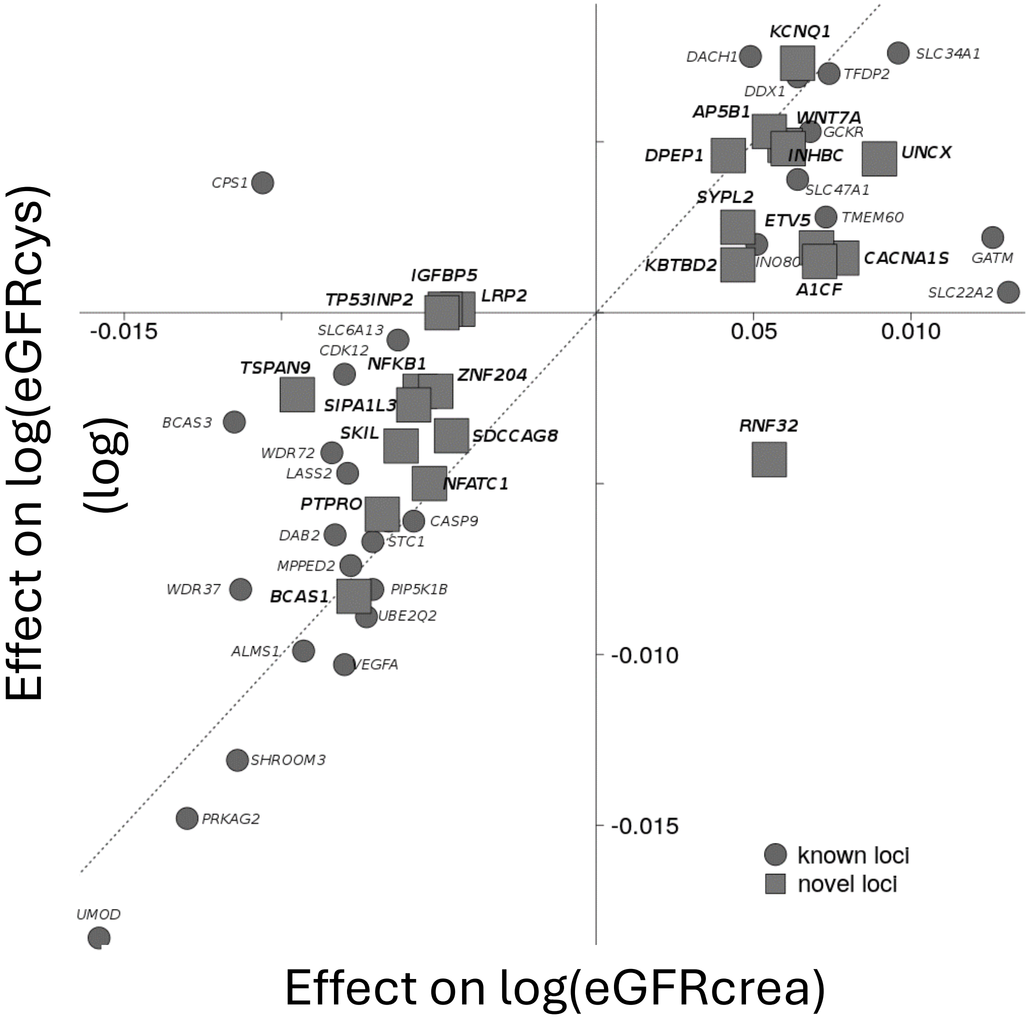
Matthias Wuttke, Yong Li, [...] Cristian Pattaro

Nature Genetics **51**, 957–972(2019) | Cite this article

Agreement between alternative markers

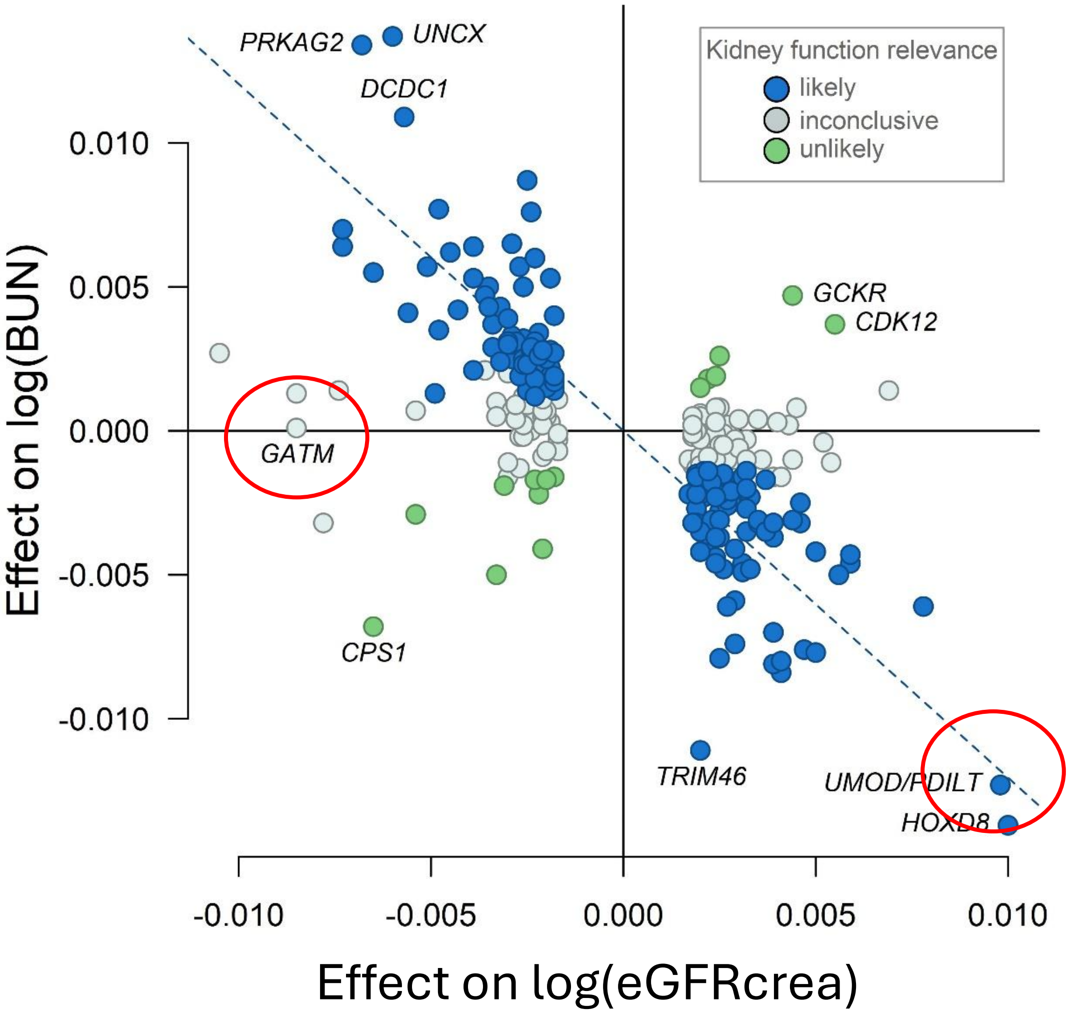


Köttgen & Pattaro. *Kidney Int* 2020

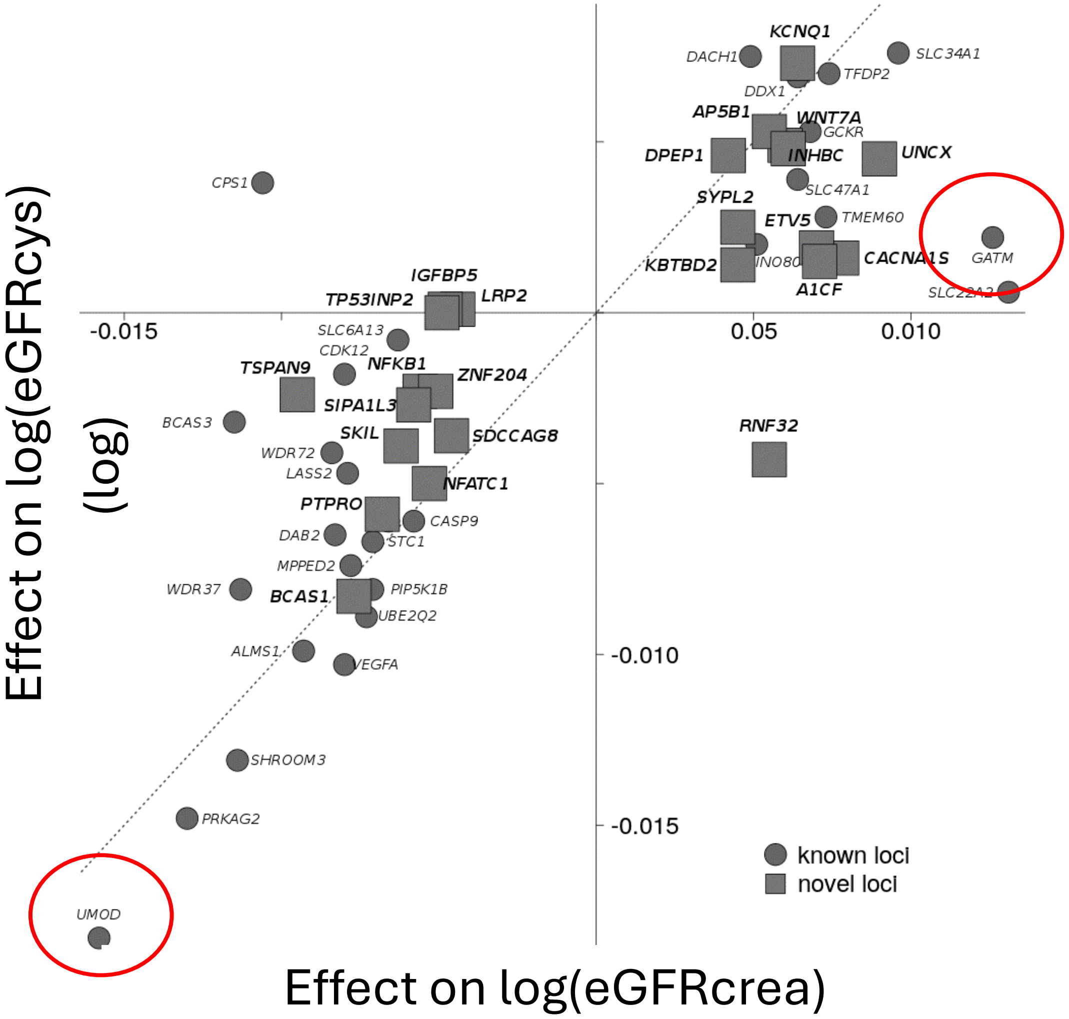


Pattaro et al *Nat Commun* 2016

Agreement between alternative markers



Köttgen & Pattaro. *Kidney Int* 2020



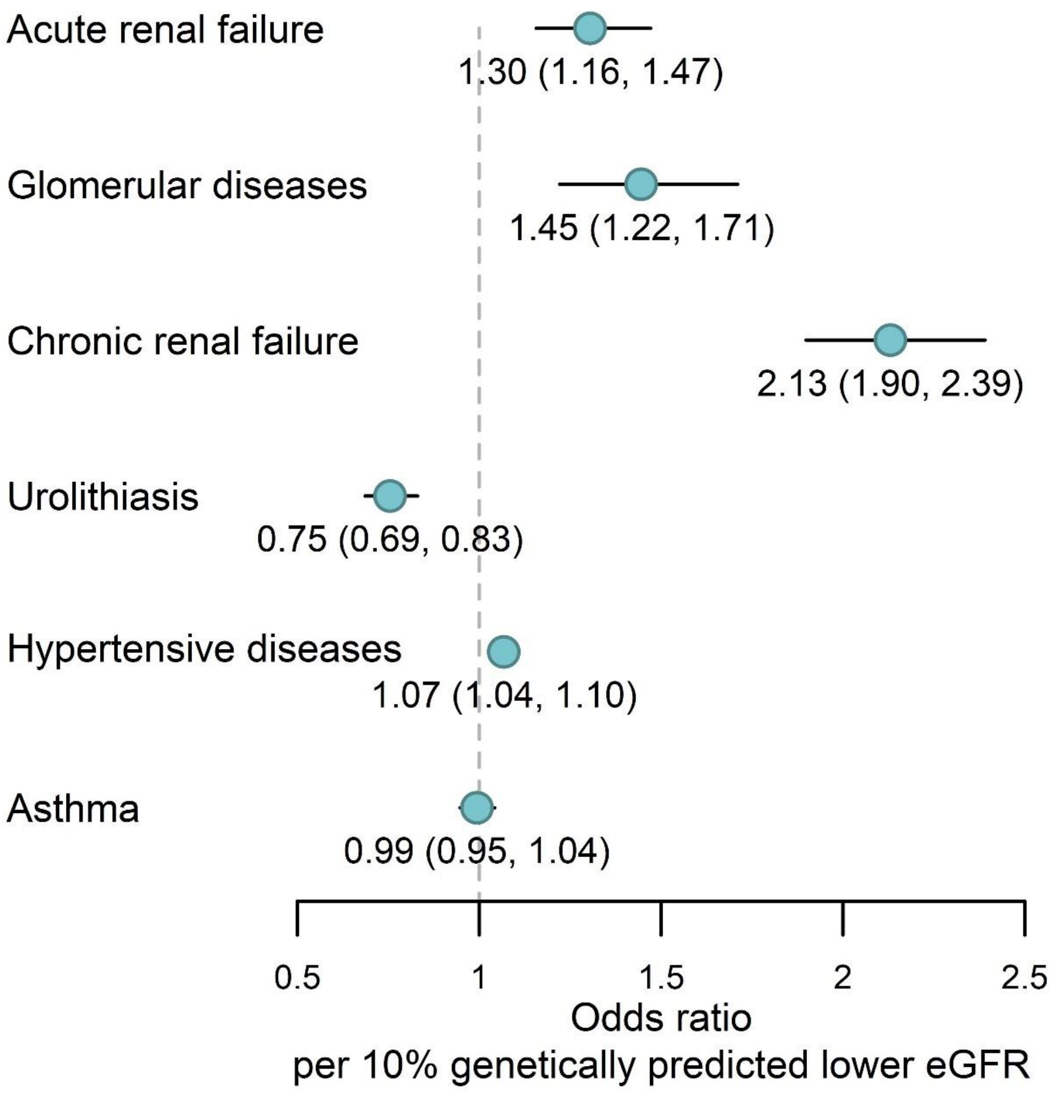
Pattaro et al *Nat Commun* 2016

Polygenic score for eGFR & BUN

applied to ICD10 kidney-related outcomes from the UK Biobank (N=452,264)

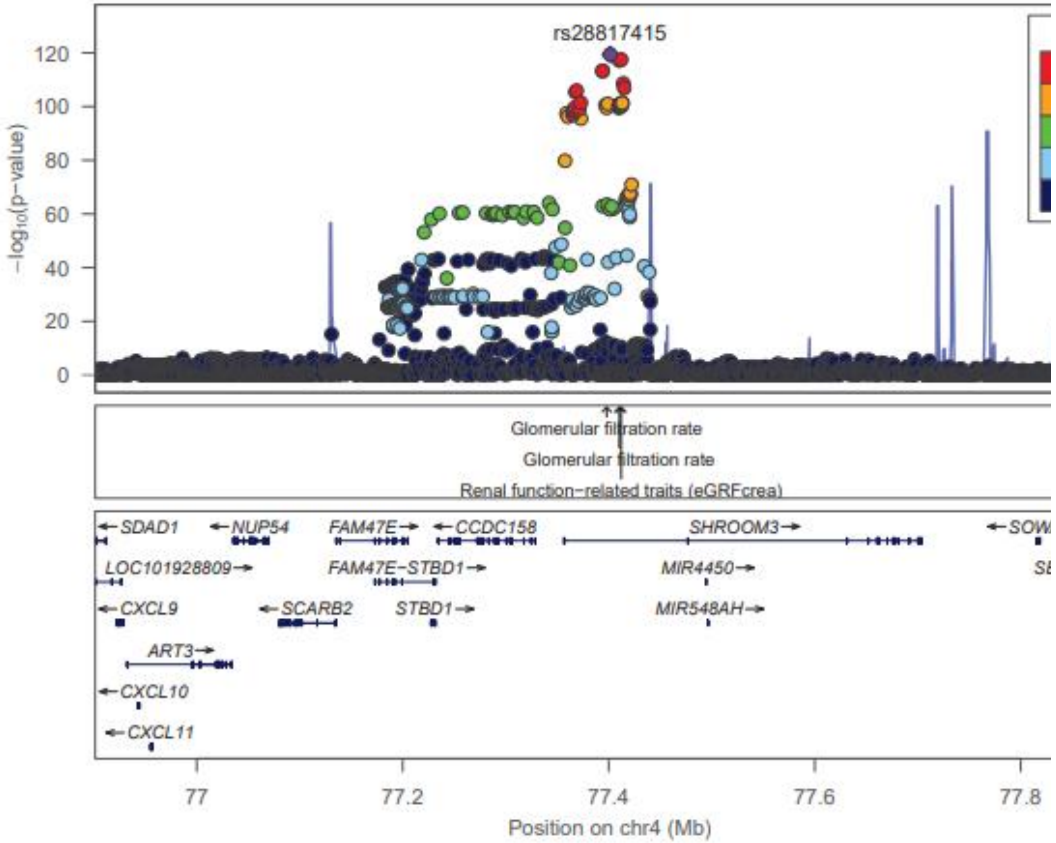
PGS based on **147 SNPs associated with eGFRcrea <u>and</u> with BUN** in a direction-consistent manner

better than PGS based on 246 SNPs associated with eGFRcrea, unfiltered for BUN support

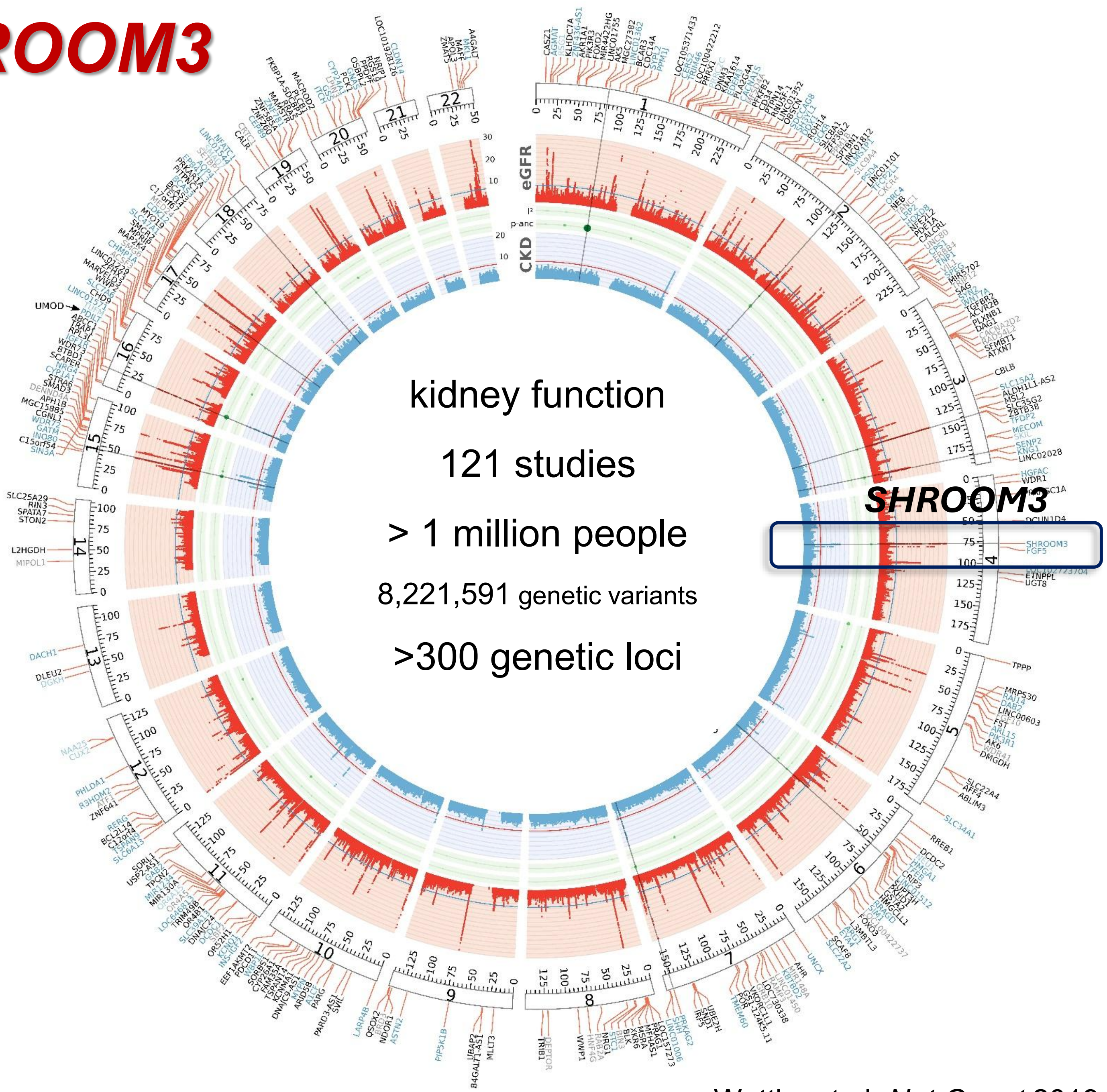


Wuttke et al. *Nat Genet* 2019

From GWAS to function: SHROOM3



actin-binding protein involved in cell shape, neural tube formation, and epithelial morphogenesis



SHROOM3

kidney function

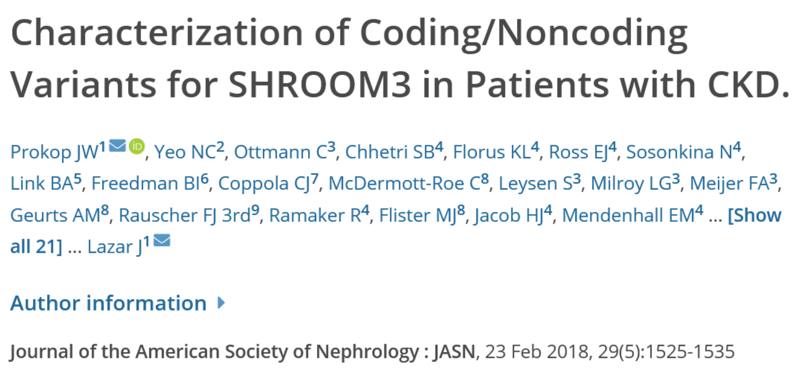
121 studies

> 1 million people

8,221,591 genetic variants

>300 genetic loci

Wuttke et al. *Nat Genet* 2019

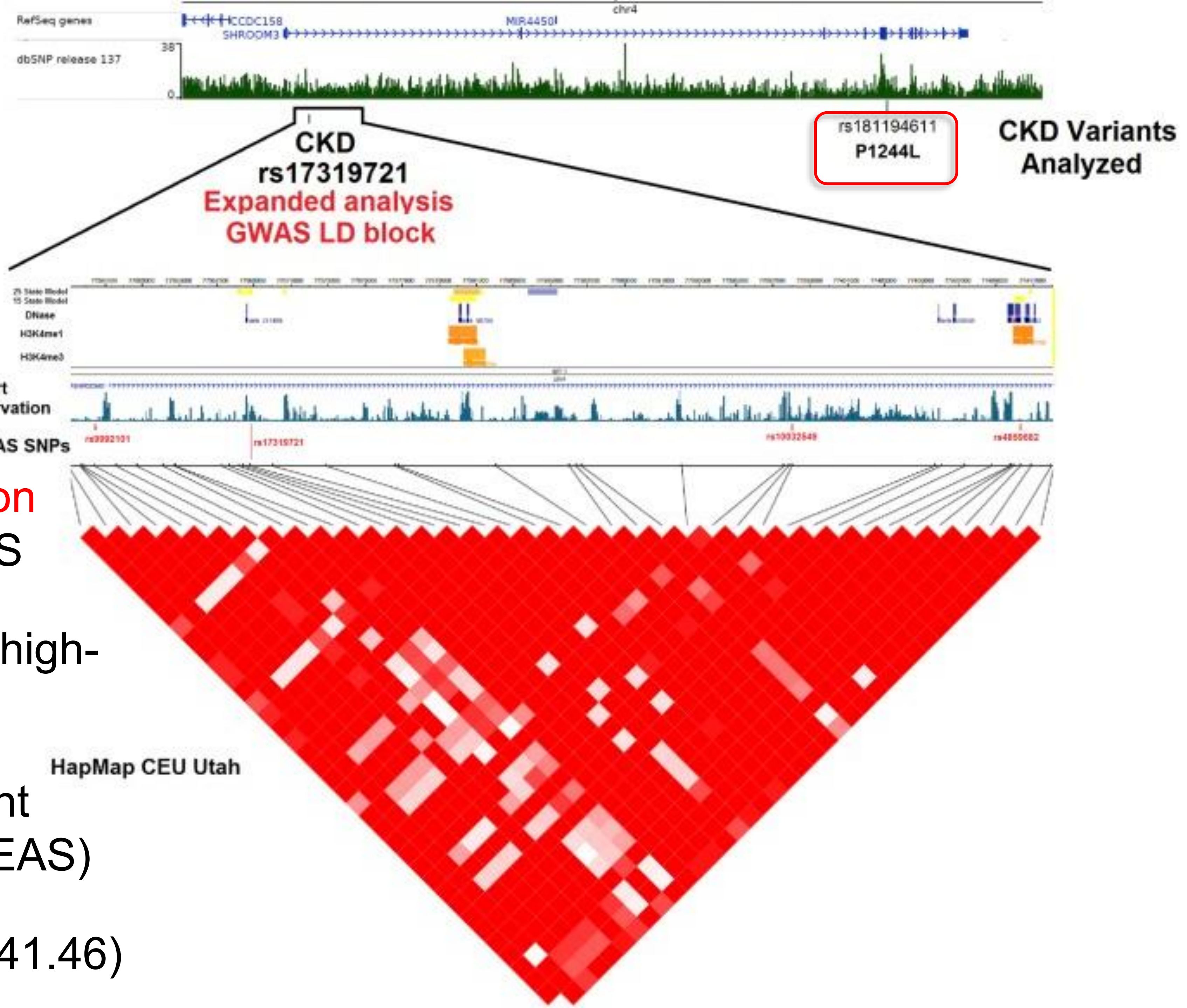


Characterization of Coding/Noncoding Variants for SHROOM3 in Patients with CKD.

Prokop JW[1], Yeo NC[2], Ottmann C[3], Chhetri SB[4], Florus KL[4], Ross EJ[4], Sosonkina N[4], Link BA[5], Freedman BI[6], Coppola CJ[7], McDermott-Roe C[8], Leysen S[3], Milroy LG[3], Meijer FA[3], Geurts AM[8], Rauscher FJ 3rd[9], Ramaker R[4], Flister MJ[8], Jacob HJ[4], Mendenhall EM[4] ... [Show all 21] ... Lazar J[1]

Author information ▸

Journal of the American Society of Nephrology : JASN, 23 Feb 2018, 29(5):1525-1535



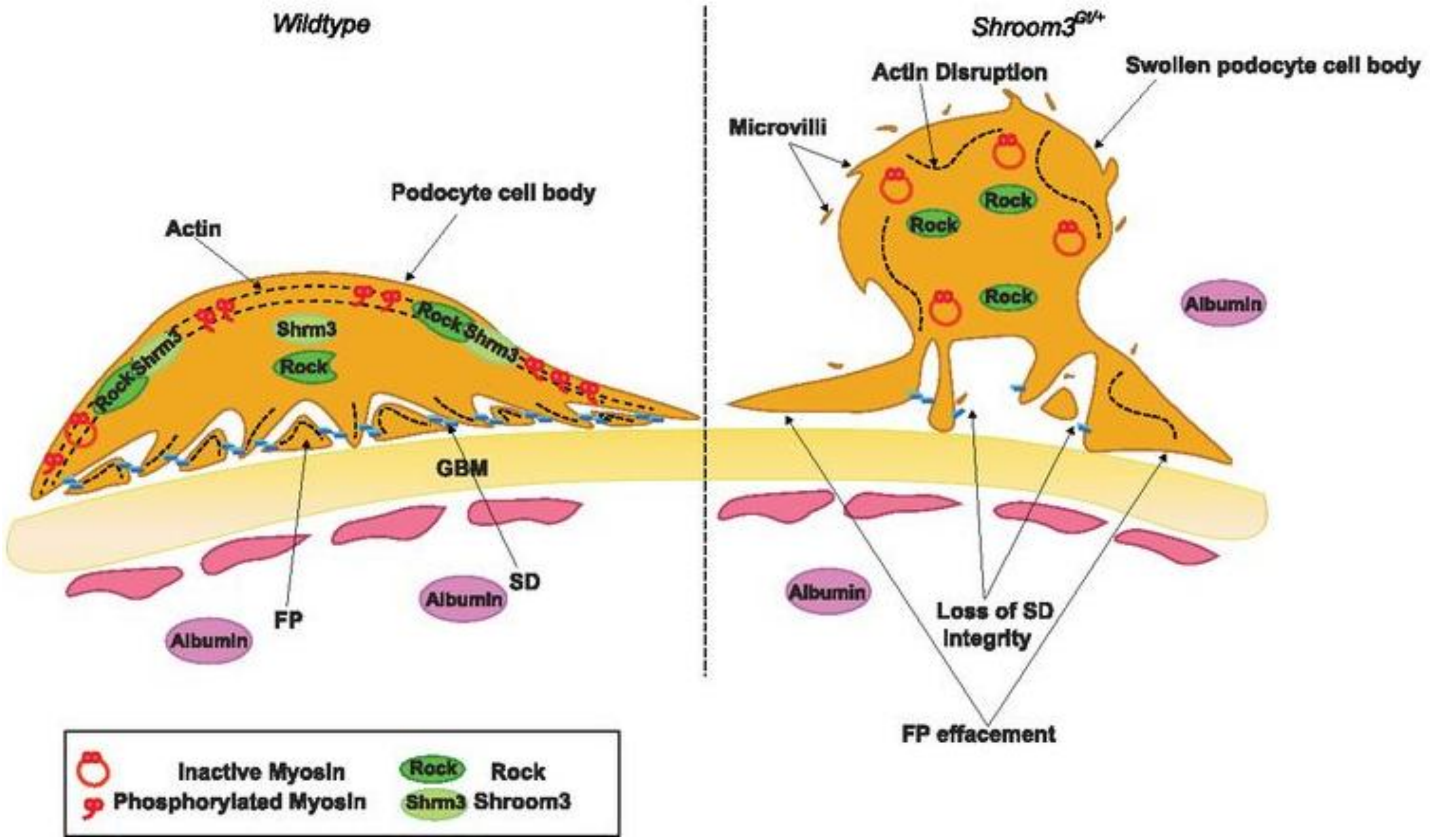
- In *SHROOM3*, the common variant identified by GWAS
- …is in LD with 35 nearby high-effects variants
- …and a rare coding variant P1244L (MAF=0.0027 in EAS)
- OR for CKD = 7.95 (1.53-41.46)

In Fawn Hooded Hypertensive rats, missense variants within Shroom3 **affects normal maintenance of kidney glomerular filtration**

Yeo NC et al (2015). *Genome Res* **25**: 57–65

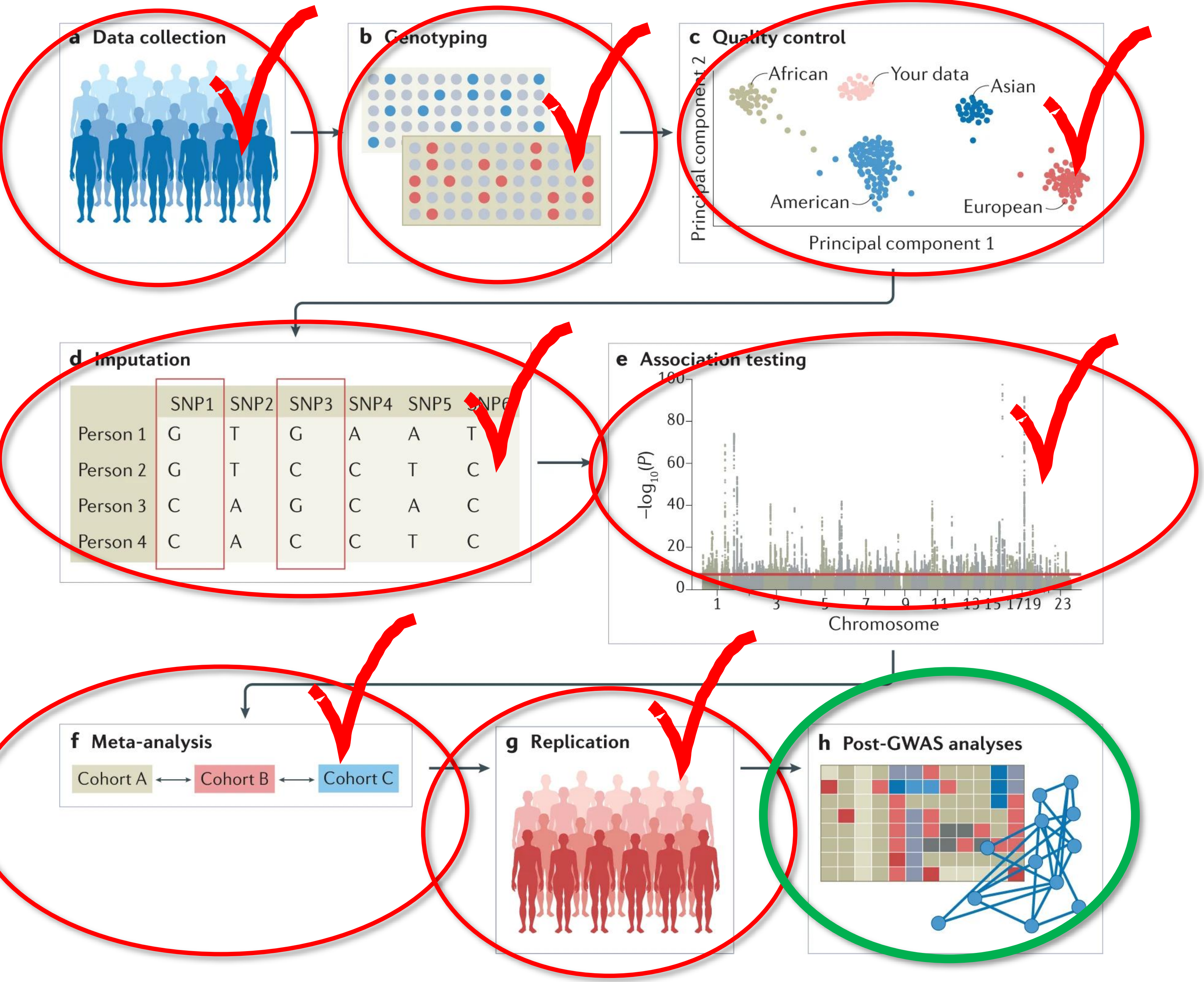
In mice, genetic deletion of Shroom3 affects **glomerular function** and **maintenance of proper podocyte morphology**, with alterations of apically distributed actin.

Khalili H, et al (2016) *JASN* **27**: 2965–73



Take home messages

1. GWAS assess association with any trait over imputed SNPs
2. GWAS is the first step of genomic characterization
3. Quality controls are essential
4. While GWAS bring interesting results, digging into causal mechanisms requires further downstream analyses

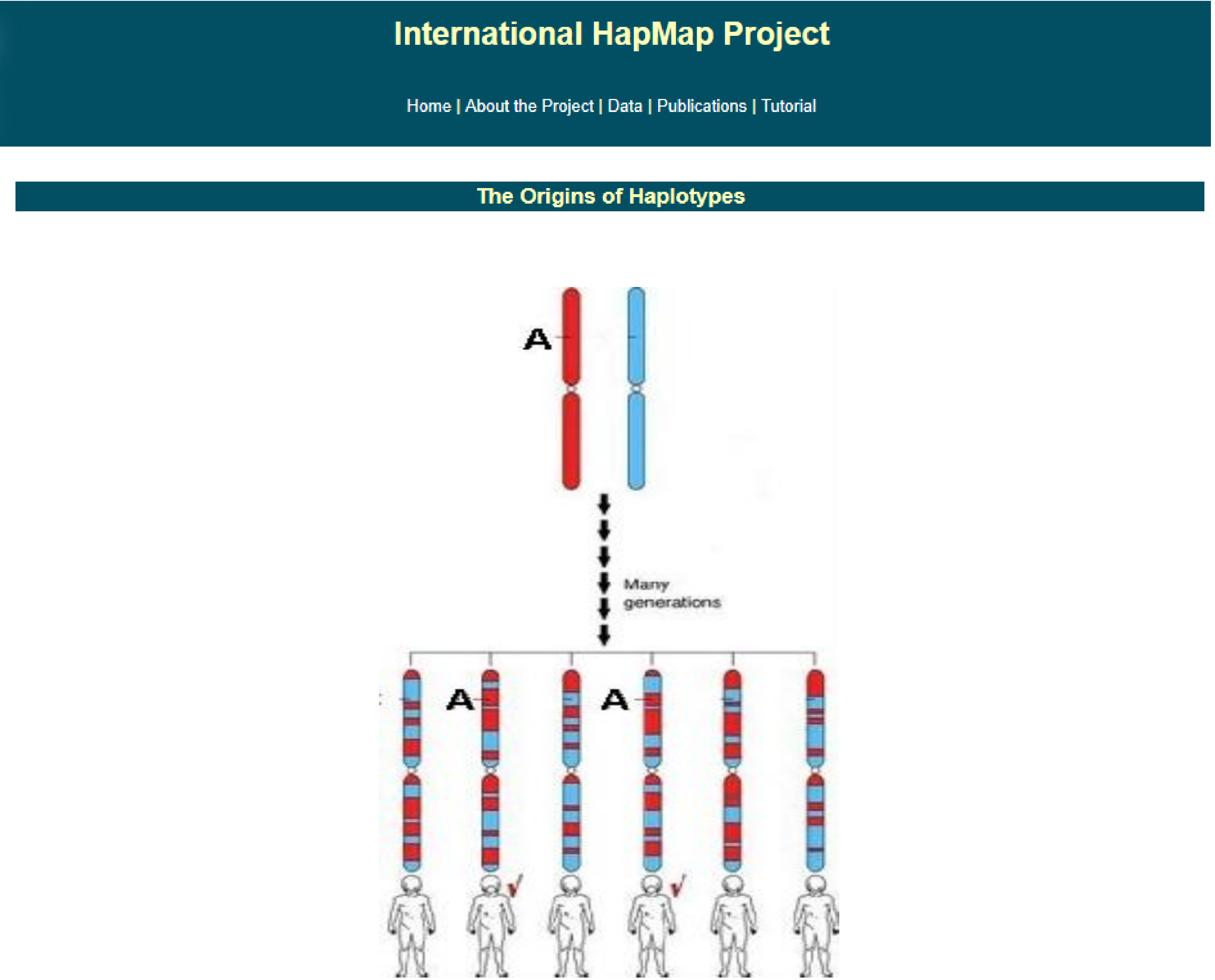


	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6
Person 1	G	T	G	A	A	T
Person 2	G	T	C	C	T	C
Person 3	C	A	G	C	A	C
Person 4	C	A	C	C	T	C

Andrew Morris
U Manchester

Uffelmann et al, Nat Rev Methods Primers 2021; 1:59

BACKUP

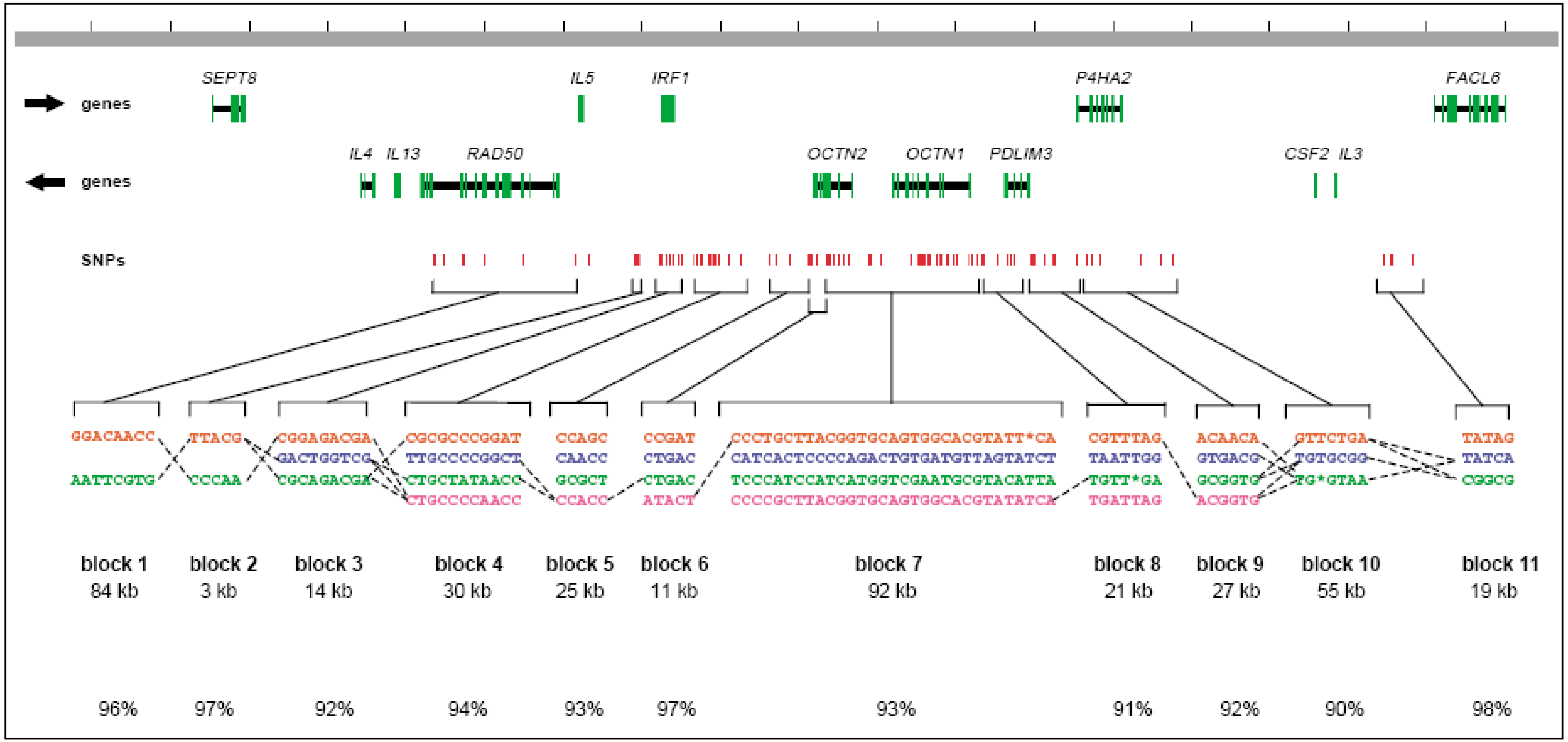


International HapMap Project

The Origins of Haplotypes

Taken from https://hapmap.ncbi.nlm.nih.gov/originhaplotype.html.en (not accessible anymore)

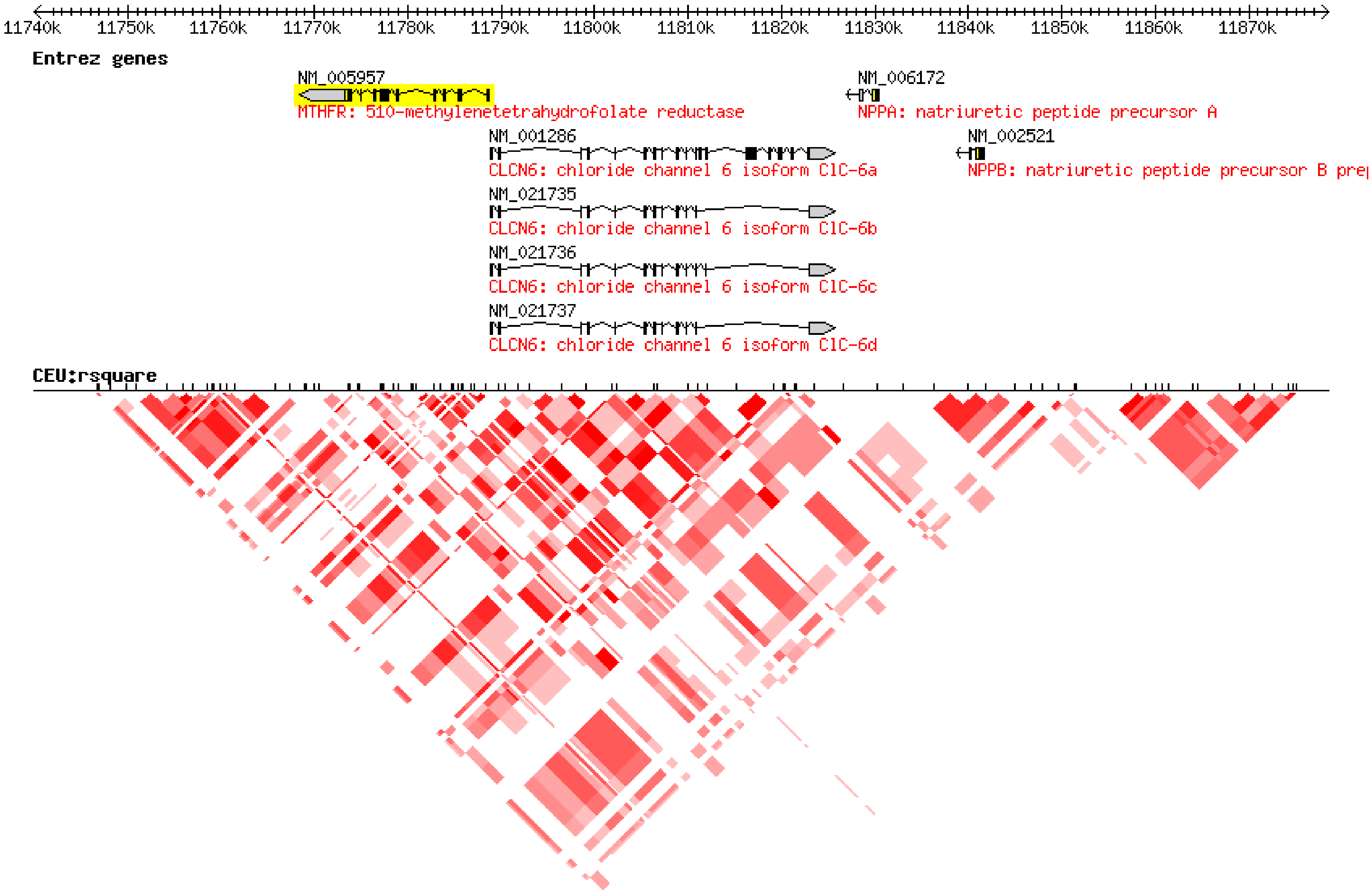
Haplotypes: block-like distribution on the genome



Daly et al., Nature Genetics 29, 2001

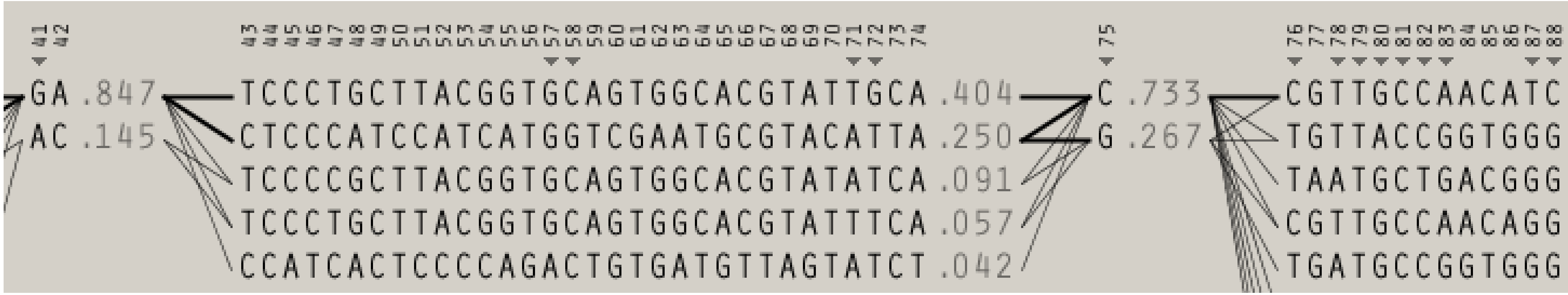
When typing large numbers of SNPs within small genomic regions, it is commonly found that there is rather **little haplotype diversity.**

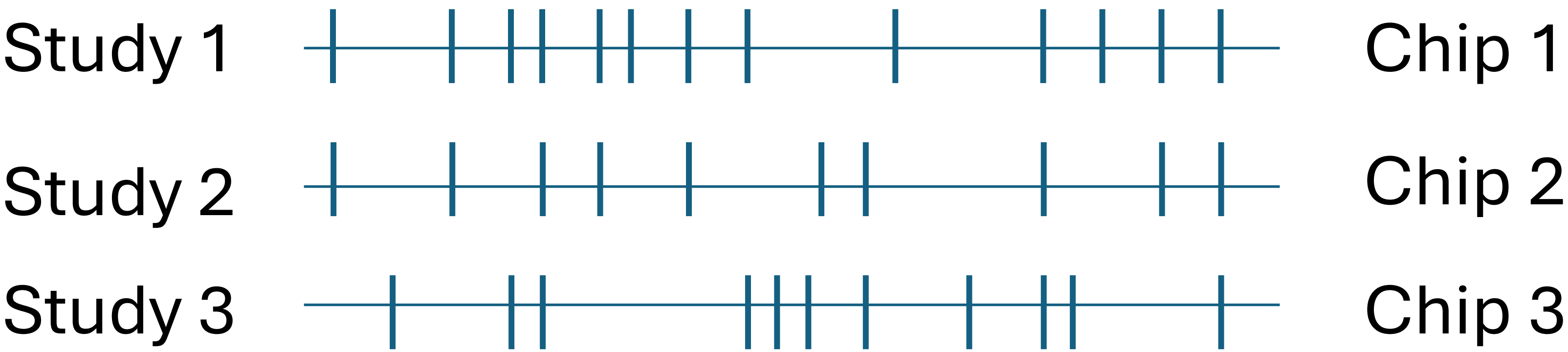
The observed haplotypes fall into rather few major groups with only minor differences between haplotypes within groups. Haplotype diversity within the region can be captured by a much smaller subset of variants



11740k
11750k
11760k
11770k
11780k
11790k
11800k
11810k
11820k
11830k
11840k
11850k
11860k
11870k
Entrez genes
NM_005957
MTHFR: 510-methylenetetrahydrofolate reductase
NM_006172
NPPA: natriuretic peptide precursor A
NM_001286
CLCN6: chloride channel 6 isoform ClC-6a
NM_002521
NPPB: natriuretic peptide precursor B pre
NM_021735
CLCN6: chloride channel 6 isoform ClC-6b
NM_021736
CLCN6: chloride channel 6 isoform ClC-6c
NM_021737
CLCN6: chloride channel 6 isoform ClC-6d
CEU:rsquare

Genotype imputation





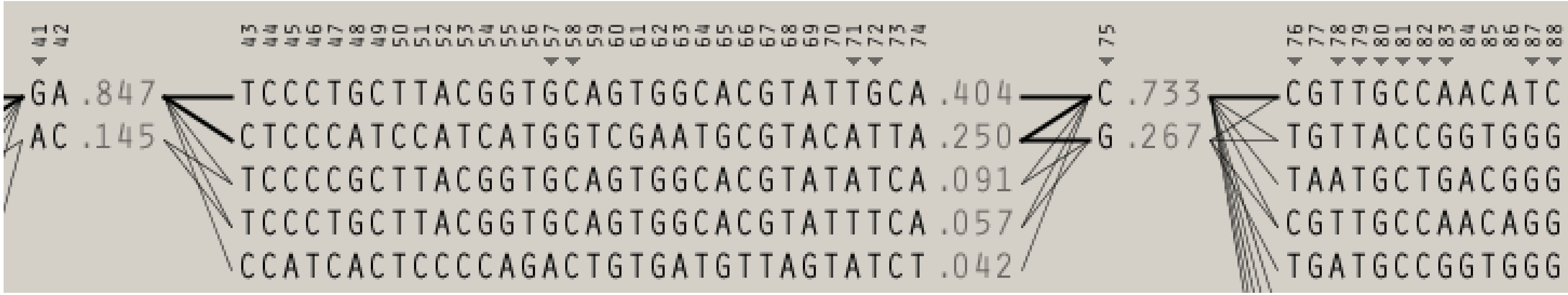
Study 1 Chip 1

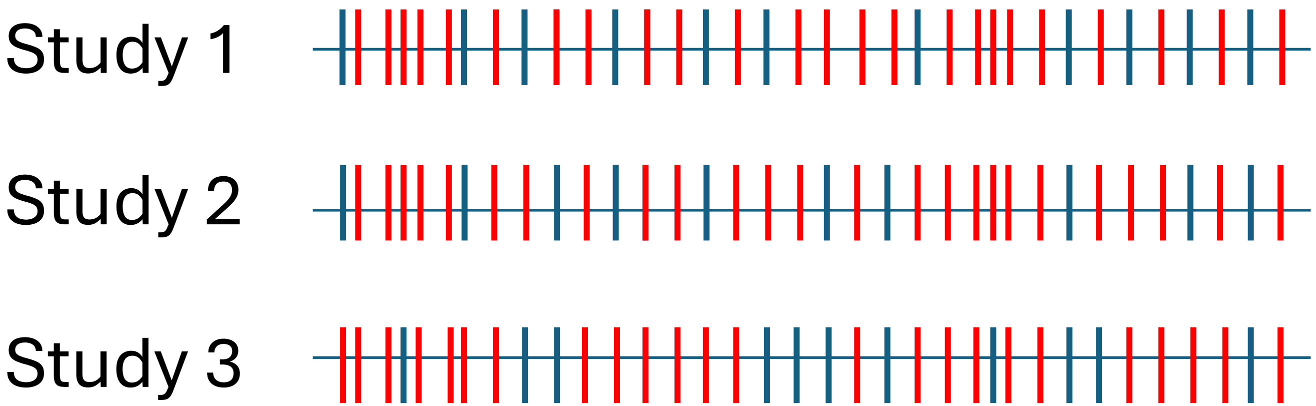
Study 2 Chip 2

Study 3 Chip 3

Suppose to have 3 studies using 3 different genotyping arrays: different no. of markers, different locations

Genotype imputation





By means of genotype imputation we can derived unobserved markers in a probabilistic way and so producing a perfect alignmet between SNP arrays at different studies

Genotype imputation

- Based on complex probabilistic methods
- Produces very reliable estimates of the true genotypes for most of the common SNPs (MAF > 0.5%)
- Has enable to expand SNP-chips with only ~300,000-1 Million SNPs to 10 Million imputed SNPs, when using the 1000 Genomes as reference panel
- Limited value for rare variants (MAF < 0.5%)

RISK OF POPULATION STRATIFICATION

Population no. 1

$p = 0.8, q = 0.2$

$p(D+) = 0.03$

genotypes are distributed according to HWE

	AA	Aa	aa	
D-	1862	931	116	2910
D+	58	29	4	90
	1920	960	120	3000

Chi-square test (2 df) → p-value = 1
no association

Population no. 2

$p = 0.6, q = 0.4$

$p(D+) = 0.08$

genotypes are distributed according to HWE

	AA	Aa	aa	
D-	994	1325	442	2760
D+	86	115	38	240
	1080	1440	480	3000

Chi-square test (2 df) → p-value = 1
no association

The two populations differ by

- SNP minor allele frequency
- Disease prevalence

If we mix such two groups, we create population stratification and the risk of spurious results

Population 1 + 2

$p = 0.7$, $q = 0.3$ $p(D+) = 330/6000$

$= 5.5\%$

	AA	Aa	aa	
D-	2856	2256	558	5670
D+	144	144	42	330
	3000	2400	600	6000

Population 1 + 2

p = 0.7, q = 0.3

p(D+) = 300/6000 =5.5%

	AA	Aa	aa	
D-	2856	2256	558	5670
D+	144	144	42	330
	3000	2400	600	6000

4.8% **6%** **7%**

Chi-squared = 6.58 (2 df), P-value = 0.037

significant association

Causes of population stratification:

1. Admixture of different ethnic groups or families
2. Batch effects → different genotype quality
3. Different genotyping platforms → different genotype quality

Batch effects

Differential genotype allocation by plate

Plate 1

Plate 2

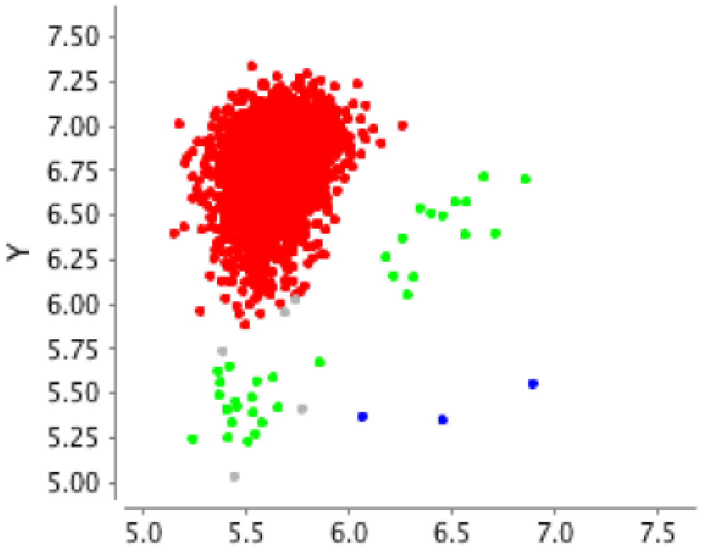
Lab conditions 1 ≠ Lab conditions 2

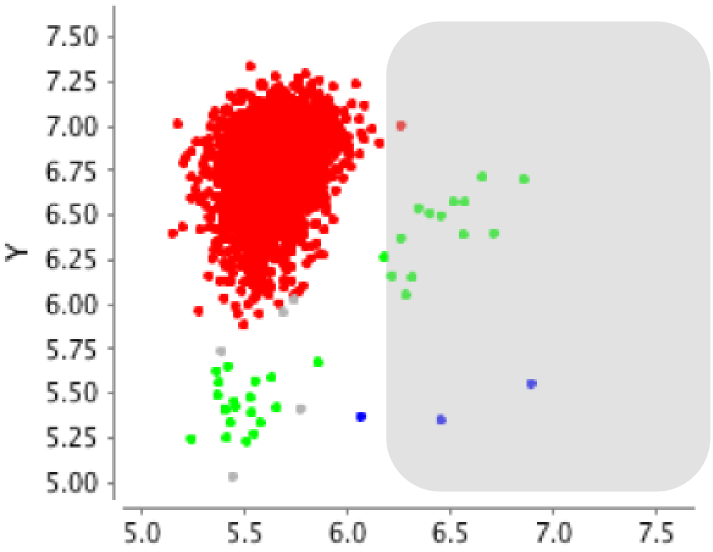
Batch effects

Plate 1

Plate 2

Lab conditions 1 ≠ Lab conditions 2



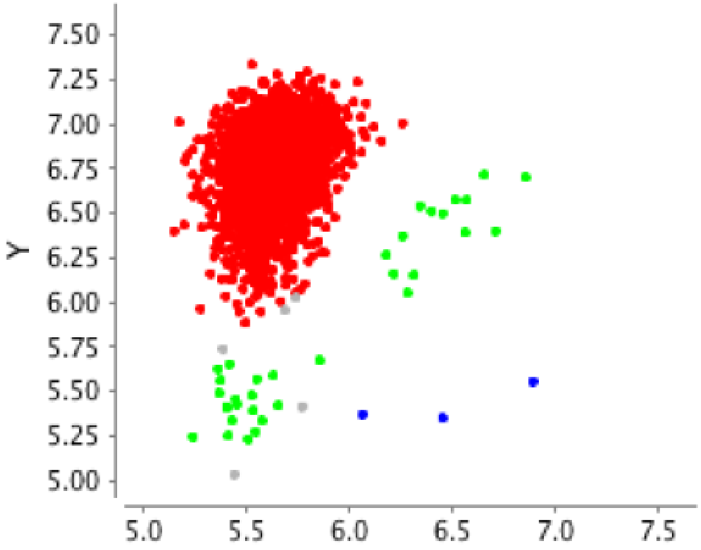


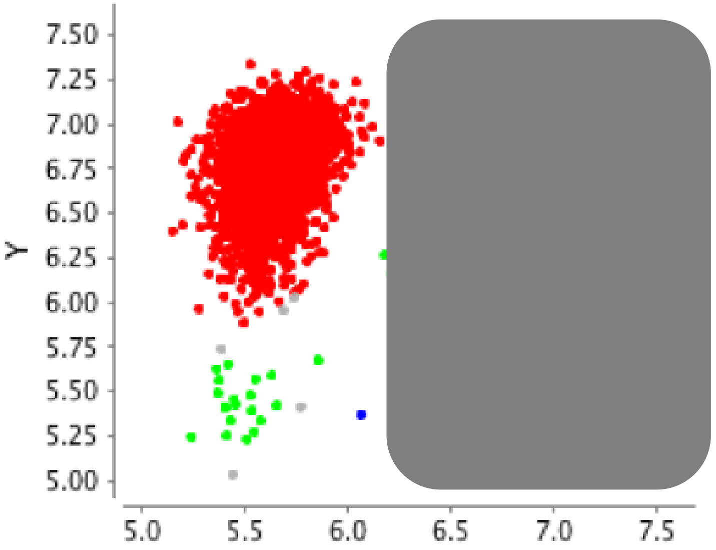
Batch effects

Plate 1

Plate 2

Lab conditions 1 ≠ Lab conditions 2





Consequence: the same SNPs has different genotype frequency in the two plates

Batch effects

Differential case-control (phenotype) allocation by plate

Plate 1

Plate 2

Proportion of affected and non affected individuals 1 ≠ Proportion of affected and non affected individuals 2

If

- Genotype allocation differs by plate
- Phenotype allocation differs by plate

high risk of false results

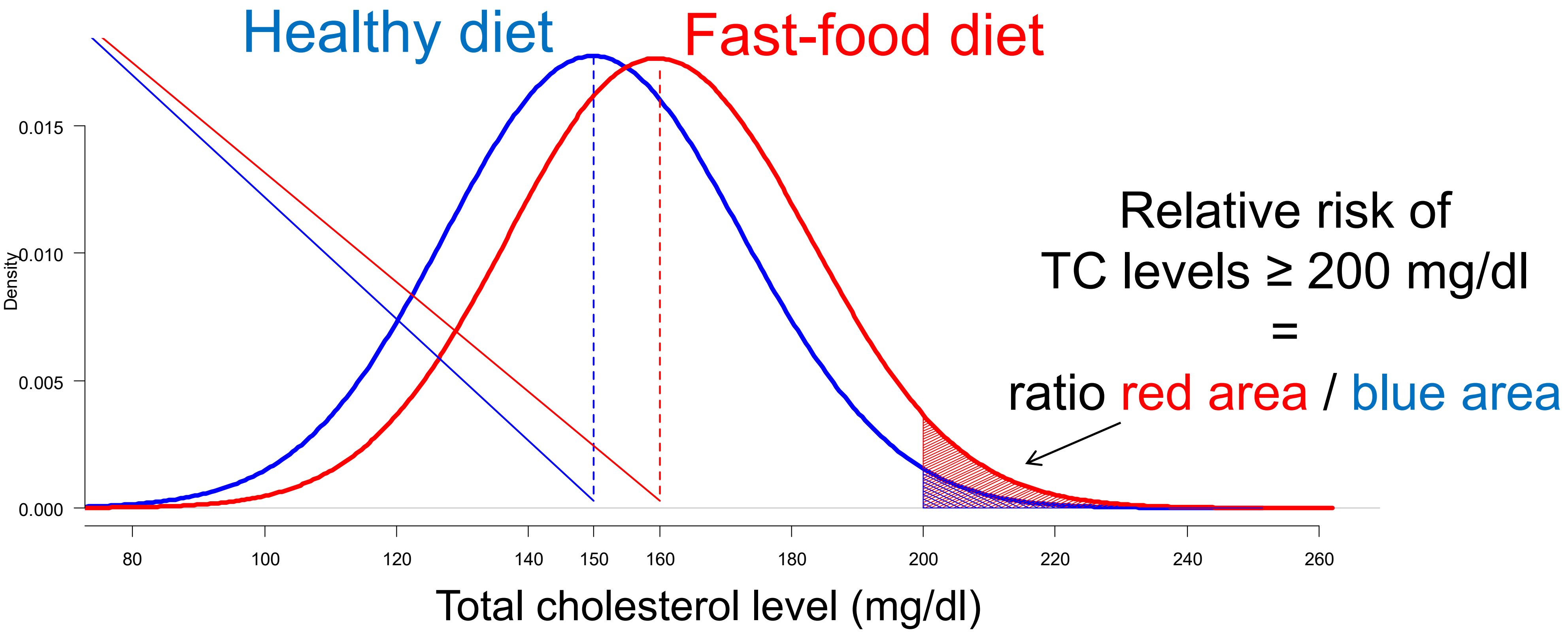
Batch effects due to different genotyping platforms

Array 1

Array 2

Genotype frequency between the two chips might be different due to differential genotyping quality/error, implying different call rate or HWE results. Issues are more sever for SNPs with very low MAF

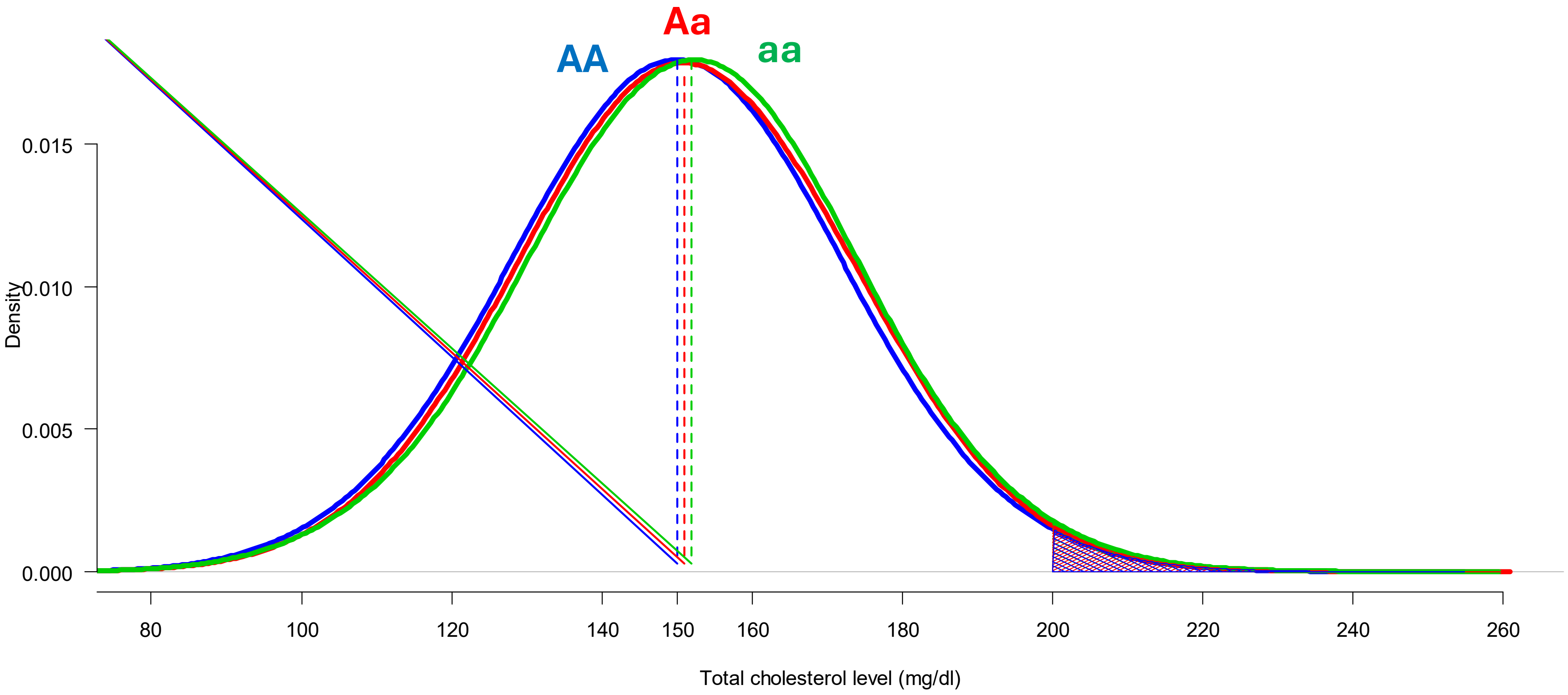
A non-genetic risk factor increases the risk of disease



In clinical epidemiology, we are used to think that the presence of a risk factor corresponds to a substantially large difference of the mean phenotypic levels.

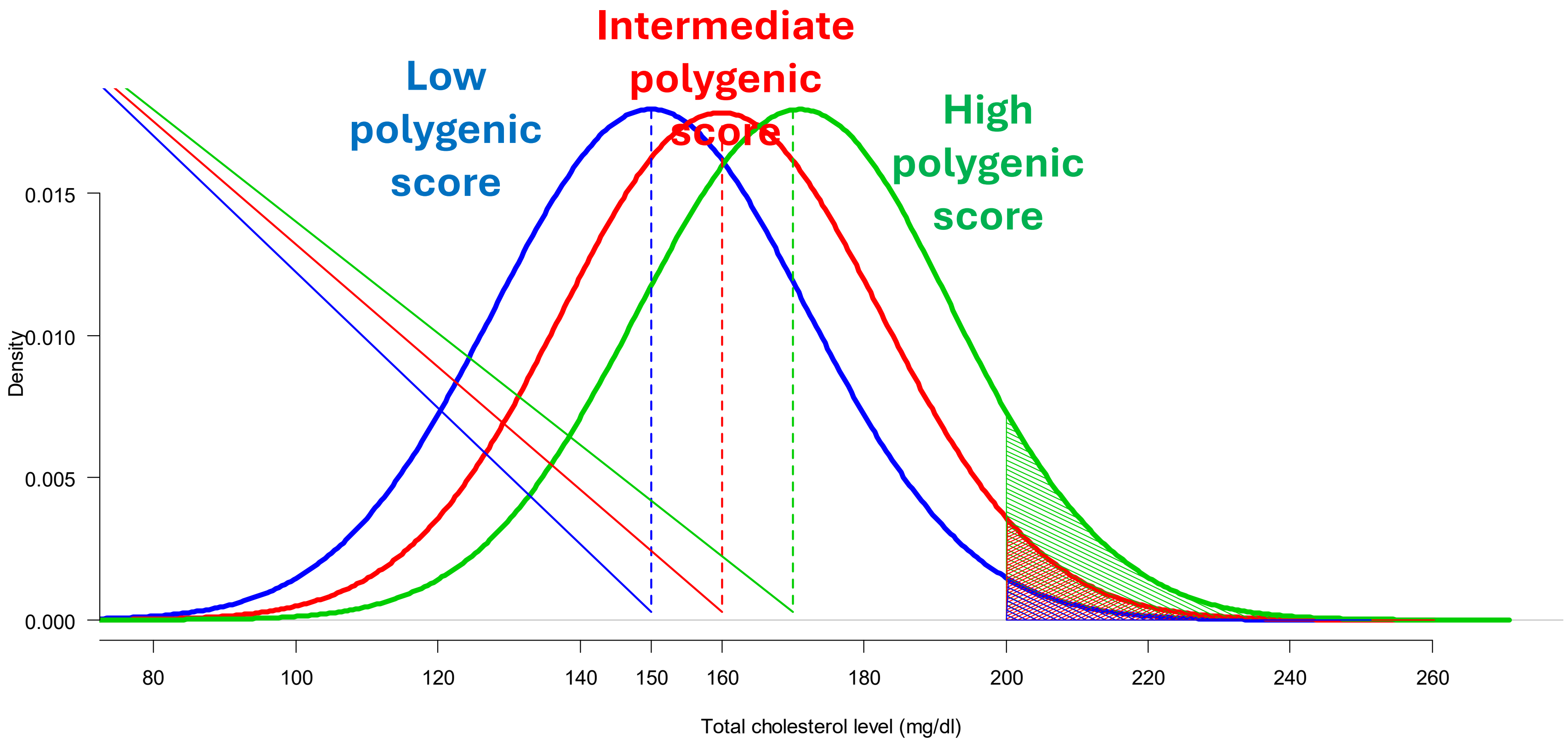
In the example, the mean cholesterol level would be 10 mg/dl larger in the „fast-food“ diet group compared to the other group.

A genetic risk factor increases the risk of disease



Associaton between a SNP and TC levels [***common variant effect on a complex trait***], assuming an additive genetic model

Sum of genetic risk factors increases the risk of disease



Polygenic effect on a complex disease or trait (TC levels)

On the genetic basis of predominatly environmental diseases

- ✓ Primary prevention, aimed at removing environmental exposures, is certainly the most effective way to tackle complex diseases (public health)
- ✓ Measuring the **genetic background** helps
  - o identify more precise biomarkers
  - o identify molecular targets → developing new or more effective treatments
  - o stratification of individual susceptibility (risk)
  - o identify cases of direct genetic origin