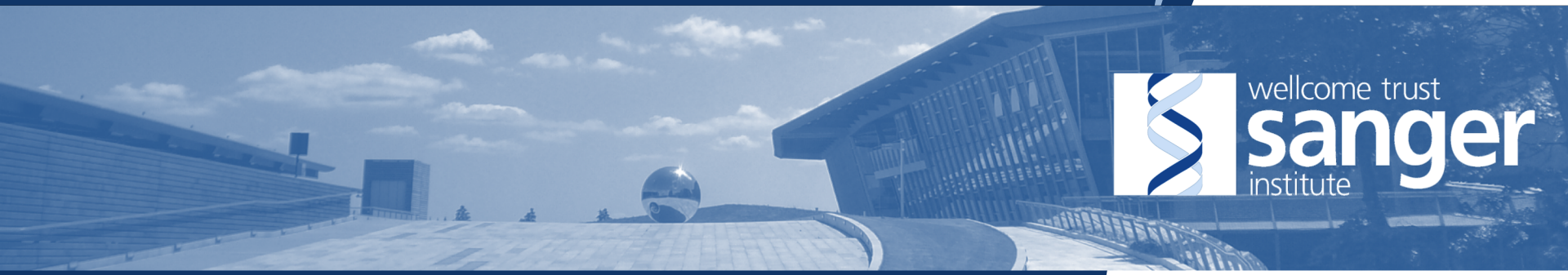
Pleiotropy in complex traits



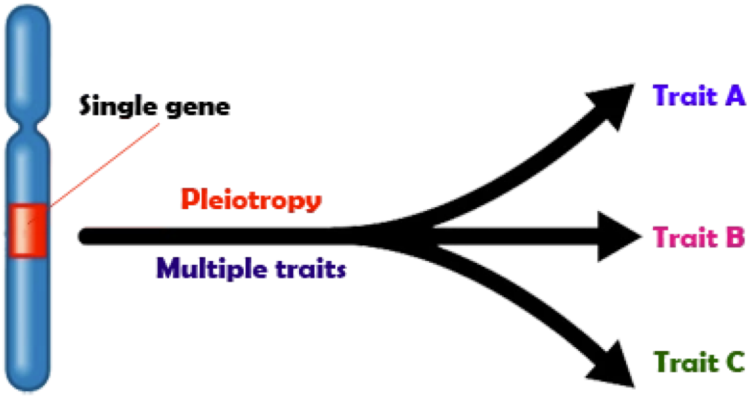
Sophie Hackinger

sh29@sanger.ac.uk

Volos Summer School of Human Genetics

Pleiotropy

- The phenomenon of one genetic variant or locus or gene affecting multiple traits
- In the context of GWAS, also referred to as cross-phenotype association/effect
- First described by Ludwig Plate in 1910



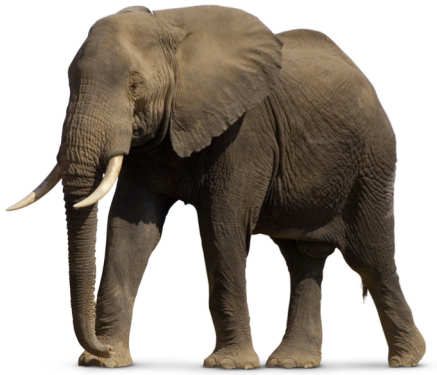
Ludwig Plate

Why bother?

Why bother?

- Increased power to detect associations

- Refine understanding of disease mechanisms
- Uncover common links between traits (not always obvious!)
- Publish in Nature Genetics



What data to use?

Biobank sample collections

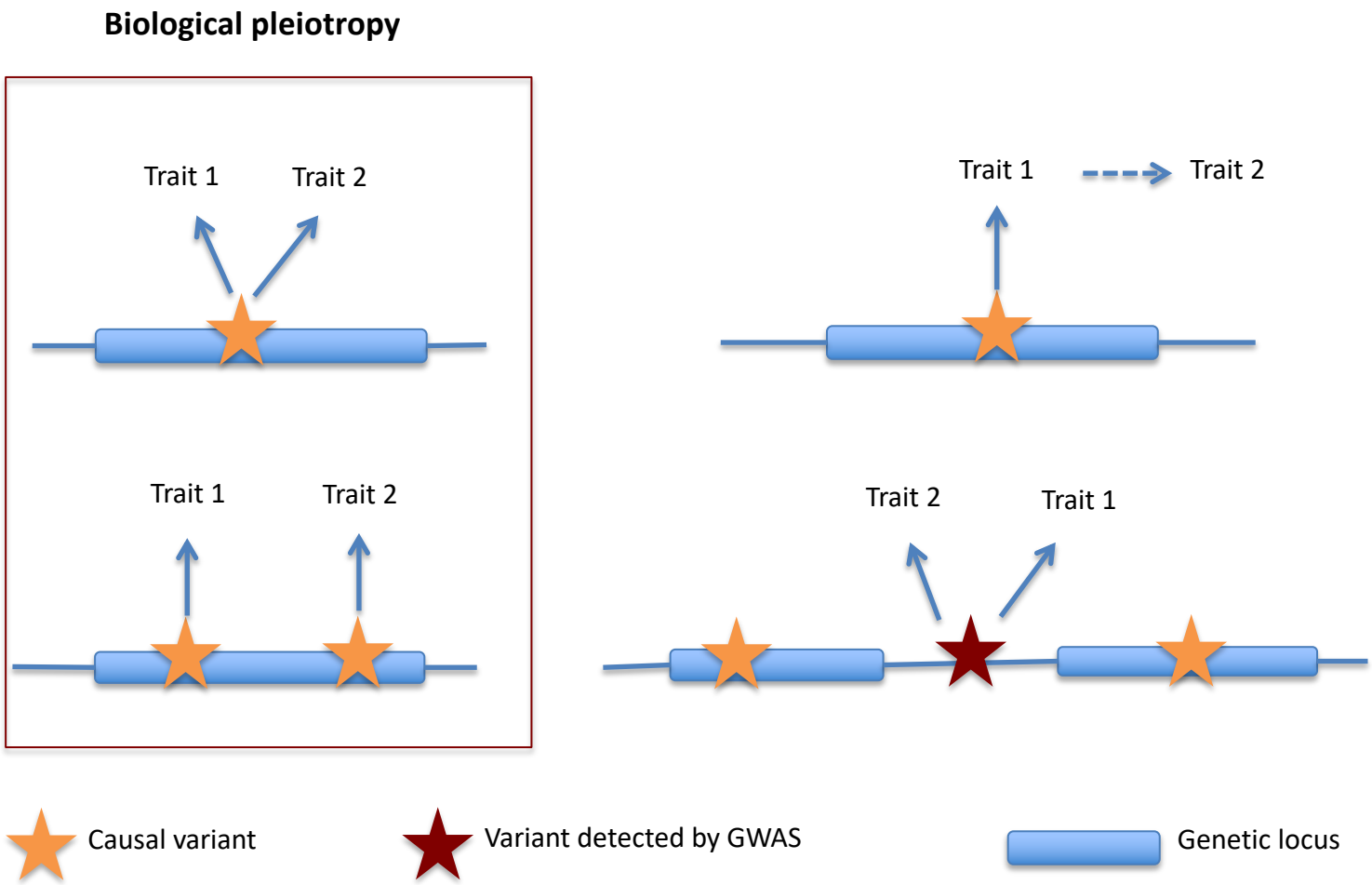
- Large number of phenotypes measured on same individuals
- Often linked to electronic health records
- Additional biological samples (e.g. urine, blood)
- Excellent resources for phenome-wide association studies (PheWAS)

biobank uk
Improving the health of future generations

THE PRECISION MEDICINE INITIATIVE

Many GWAS consortia of single traits have also made their summary data available.

Types of pleiotropy



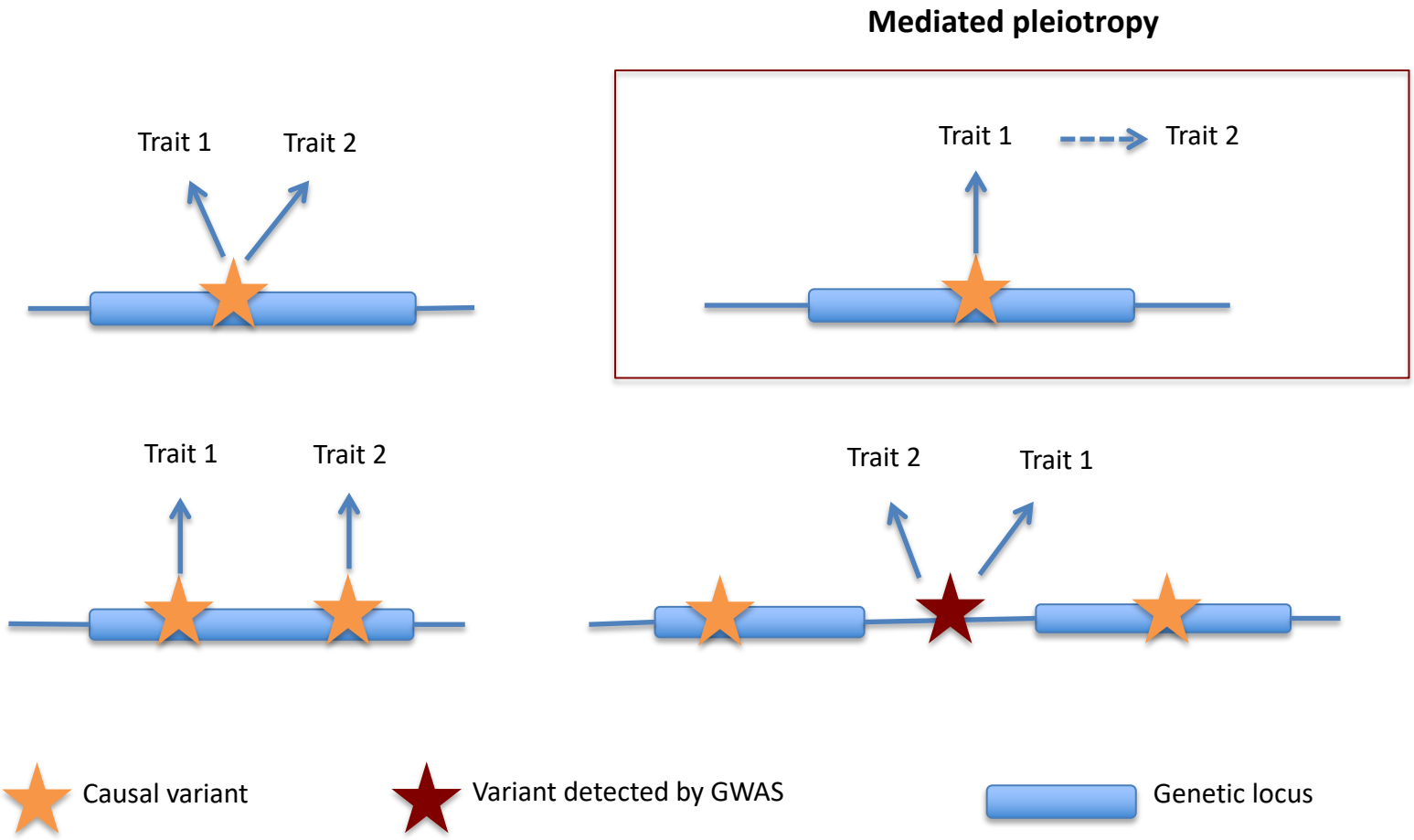
Solovieff et al. *Nat Rev Gen* (2013)

Biological pleiotropy

transcriptional modulator acting downstream of TGF-β

- Osteoarthritis
- Renal/cardiac fibrosis
- Bone density
- Colorectal cancer
- Coronary heart disease
- Autoimmune disorders

Types of pleiotropy



Solovieff et al. *Nat Rev Gen* (2013)

Mediated pleiotropy

EXTENDED REPORT

The effect of *FTO* variation on increased osteoarthritis risk is mediated through body mass index: a mendelian randomisation study

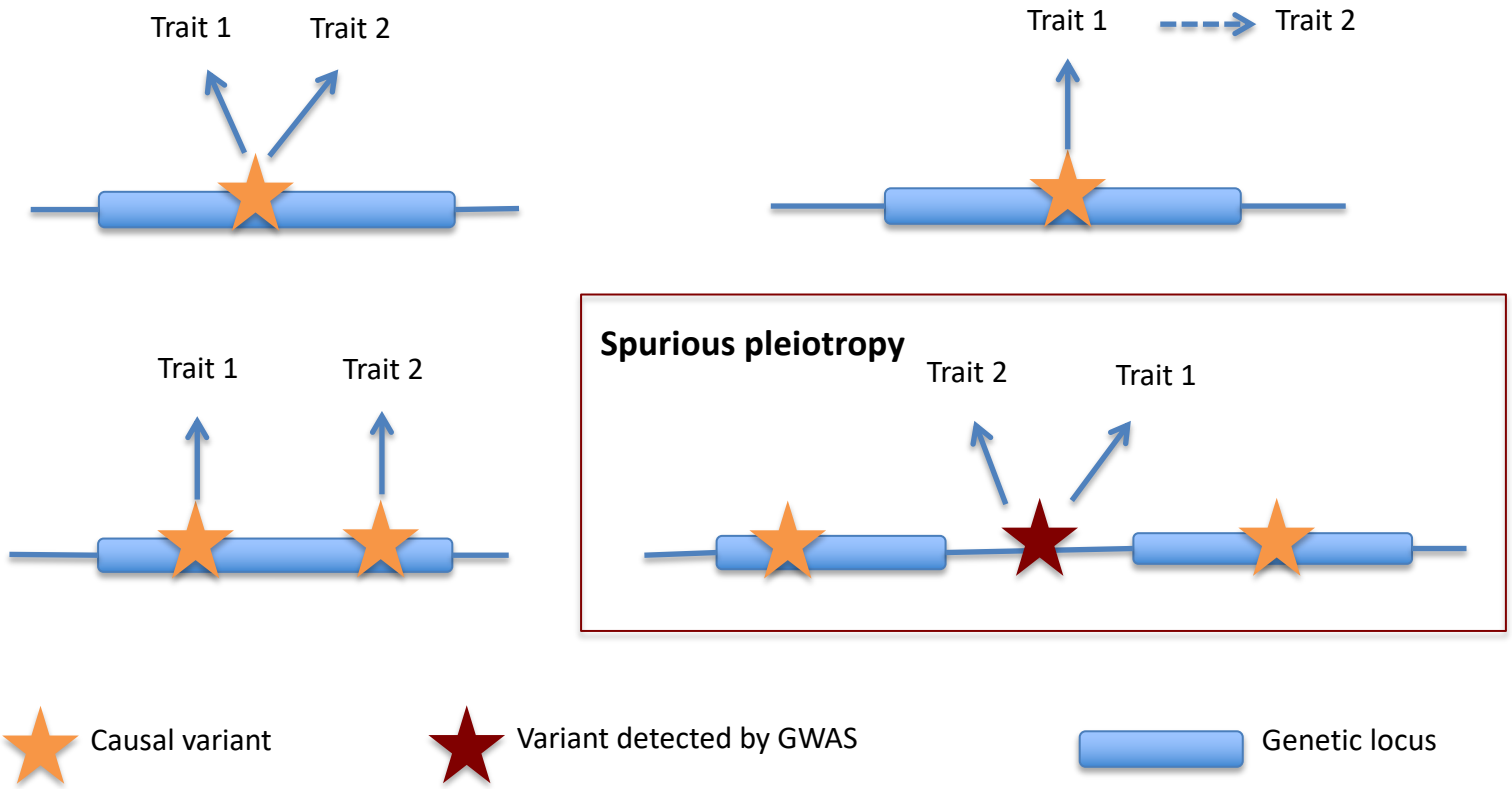
Kalliope Panoutsopoulou,[1] Sarah Metrustry,[2] Sally A Doherty,[3] Laura L Laslett,[4] Rose A Maciewicz,[5] Deborah J Hart,[2] Weiya Zhang,[3] Kenneth R Muir,[6,7] Margaret Wheeler,[3] Cyrus Cooper,[8,9] Tim D Spector,[2] Flavia M Cicuttini,[10] Graeme Jones,[4] Nigel K Arden,[8,9] Michael Doherty,[3] Eleftheria Zeggini,[1] Ana M Valdes,[2,3] arcOGEN Consortium

Common Variation in the *FTO* Gene Alters Diabetes-Related Metabolic Traits to the Extent Expected Given Its Effect on BMI

Rachel M. Freathy,[1] Nicholas J. Timpson,[2,3] Debbie A. Lawlor,[3,4] Anneli Pouta,[5] Yoav Ben-Shlomo,[4] Aimo Ruokonen,[5] Shah Ebrahim,[6] Beverley Shields,[1] Eleftheria Zeggini,[2] Michael N. Weedon,[1] Cecilia M. Lindgren,[2,7] Hana Lango,[1] David Melzer,[1] Luigi Ferrucci,[8] Giuseppe Paolisso,[9] Matthew J. Neville,[7] Fredrik Karpe,[7] Colin N.A. Palmer,[10] Andrew D. Morris,[10] Paul Elliott,[11] Marjo-Riitta Jarvelin,[5,11] George Davey Smith,[3,4] Mark I. McCarthy,[2,7] Andrew T. Hattersley,[1] and Timothy M. Frayling[1]

- *FTO* is associated with osteoarthritis and metabolic traits mainly through its BMI-increasing effect

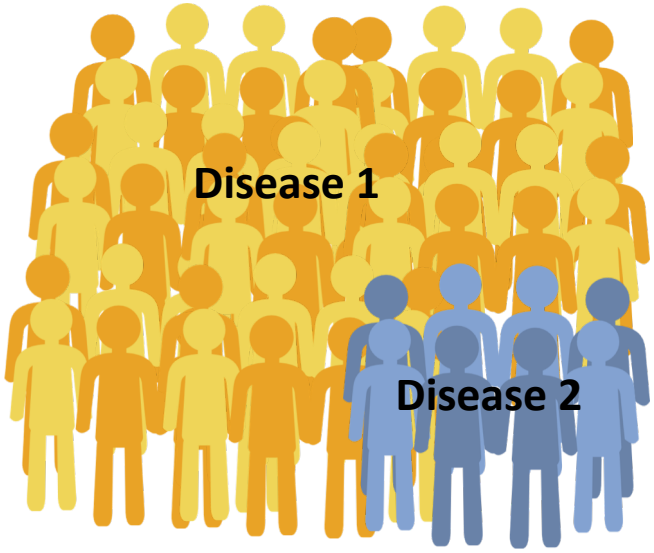
Types of pleiotropy



Solovieff et al. *Nat Rev Gen* (2013)

Spurious pleiotropy

- Ascertainment bias
  - Recruitment of people with one disease increases prevalence of second disease (biased subsample)



Spurious pleiotropy

- Disease Heterogeneity
  - A subgroup of patients with disease 1 are genetically closer to patients with disease 2

- Shared controls
- Strong LD

Analysis methods

Pleiotropy can be assessed at different levels:

- Genome-wide
  - E.g. genetic correlation, polygenic risk scores, multivariate models
- Regional
  - Gene- or locus-based, multivariate models
  - E.g. co-localisation analyses, burden tests
- Variant-level
  - E.g. meta-analysis, multivariate models

Methods to test for non-biological pleiotropy:

- Causal relationships
  - Mendelian randomisation
- Spurious pleiotropy
  - Currently only one tool for disease traits: BUHMBOX

Analysis Methods

There are two broad categories of statistical methods:

- Univariate
  - Generally only require summary statistics
  - Can combine data from different studies (sample overlap might be an issue!)
  - Computationally faster
- Multivariate
  - Require individual-level data
  - All traits measured on same individuals
  - More powerful
  - Computationally intensive

Things to consider

- Summary or raw data?
- Traits measured on same or different samples?
- Quantitative or binary traits?
- How many traits?
- Extent of trait correlation?

Univariate methods

Univariate statistics: one response variable (e.g. trait) modeled on one or more predictor variables (e.g. genotypes, principal components, age, etc.)

$$y_1 \sim x\beta_0 + c_1\beta_1 + c_2\beta_2 + c_3\beta_3 + \varepsilon$$

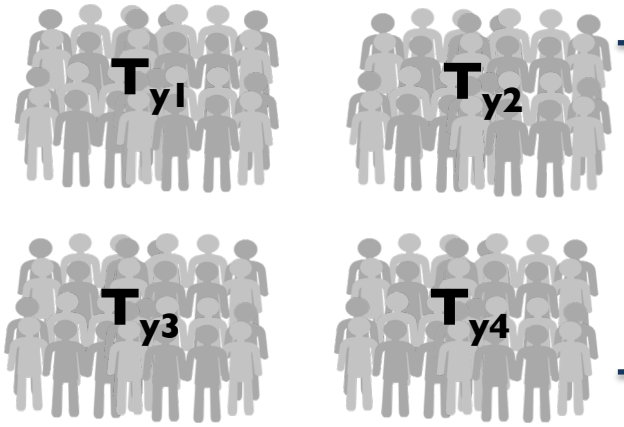
$y_{1n}=(y_{11}, y_{12}, y_3, y_4, y_5, y_6, \ldots, y_n)$ for n individuals

Univariate methods

Raw data

y_1 (e.g. height) → single-trait GWAS summary statistics (T_{y1})

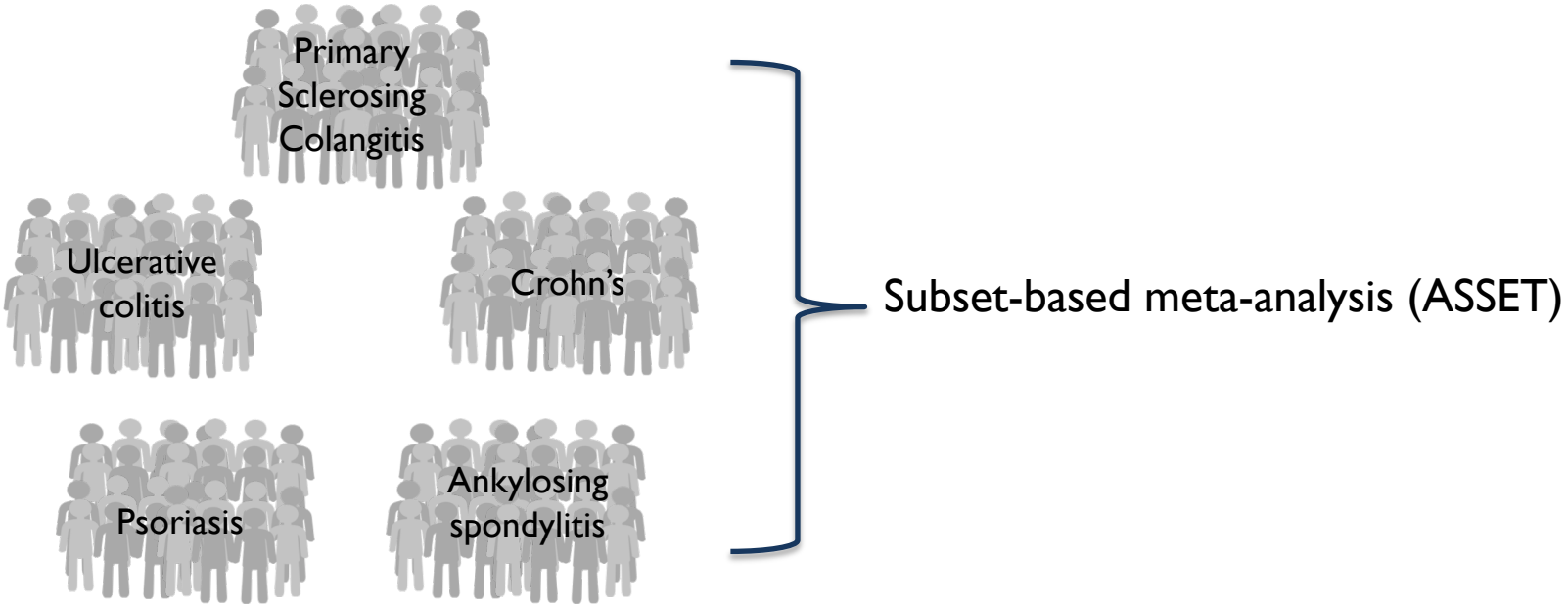
Summary data



Combine in meta-analysis or other approach

Univariate methods

Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci Ellinghaus at al. Nat Gen (2015)



- These 27 loci would have been missed by single-trait analysis
- Subsequent eQTL analysis and functional annotation showed biologically relevant consequences of identified variants

Multivariate methods

Multivariate statistics: several response variables (e.g. traits) modeled on one or more predictor variables (e.g. genotypes, principal components, age, etc.)

!!! multivariate regression ≠ multiple or multivariable regression !!!

$$\begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{mn} \end{bmatrix} \sim \mathrm{x}\beta_0 + \mathrm{c}_1\beta_1 + \mathrm{c}_2\beta_2 + \mathrm{c}_3\beta_3 + \varepsilon$$

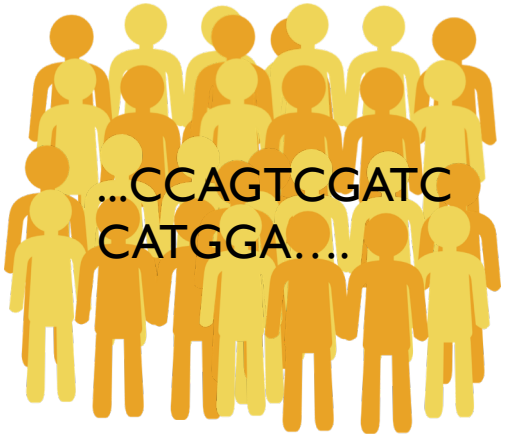
Matrix with m rows (traits) and n columns (individuals)

Each row contains a vector of trait measurements, as for univariate regression:

$y_{mn}=(y_{m1}, y_{m2}, y_{m3}, y_{m4}, y_{m5}, y_{m6}, \ldots, y_{mn})$ for n individuals

Multivariate methods

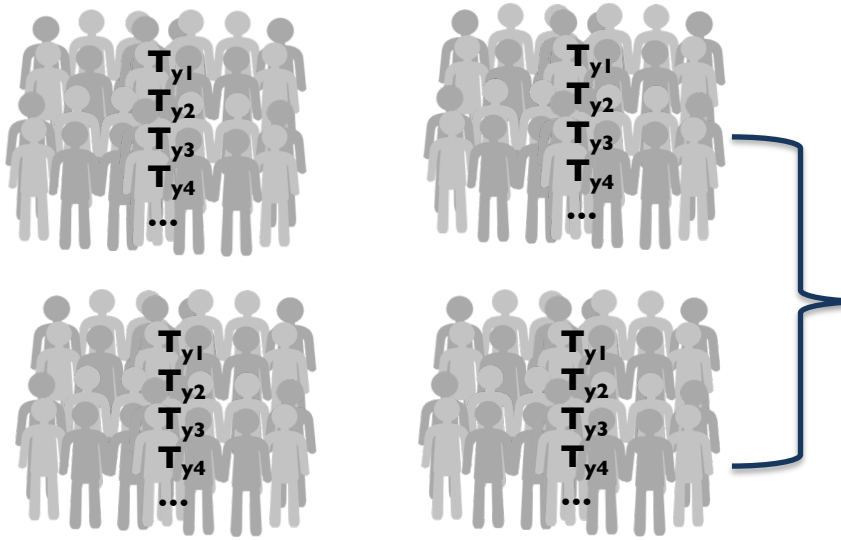
Raw data



$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \dots \end{bmatrix} = \begin{bmatrix} \text{Height} \\ \text{BMI} \\ \text{Blood pressure} \\ \text{Fasting glucose} \end{bmatrix}$$

Multi-trait GWAS summary statistic (T_y)

Summary data

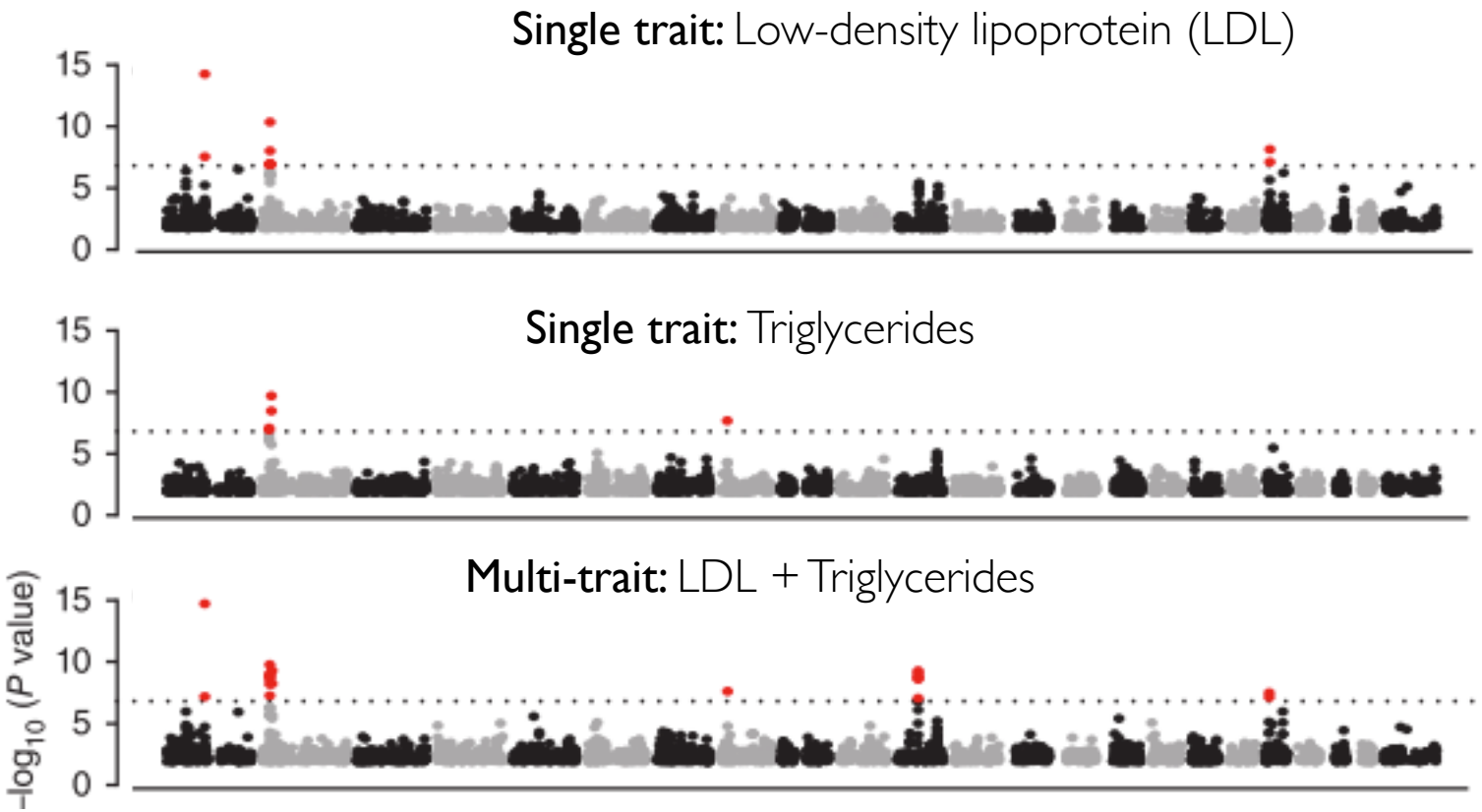


Multivariate methods

- Multivariate mixed models
- Dimension reduction techniques (PCA, CCA)
- And others…

Multivariate GWAS

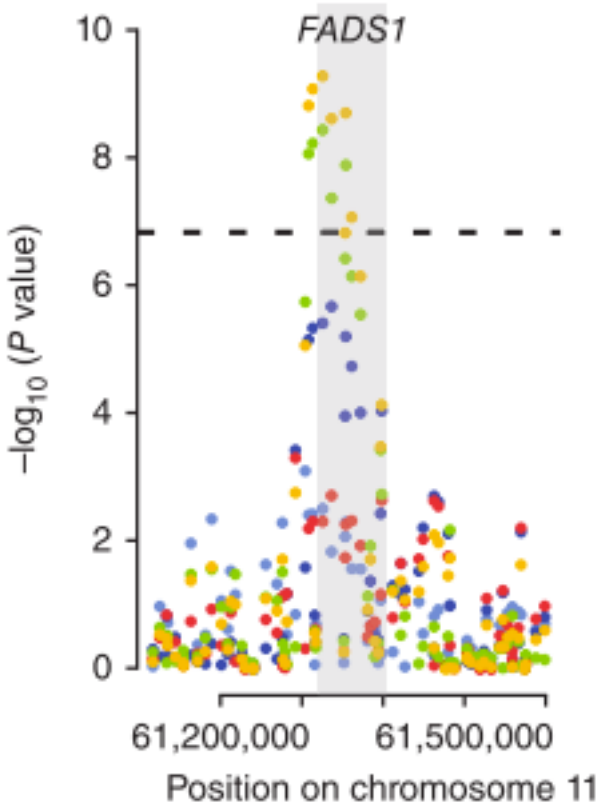
Mixed model analysis of blood metabolites



Korte et al. *Nat Gen* (2012)

Multivariate GWAS

Mixed model analysis of blood metabolites

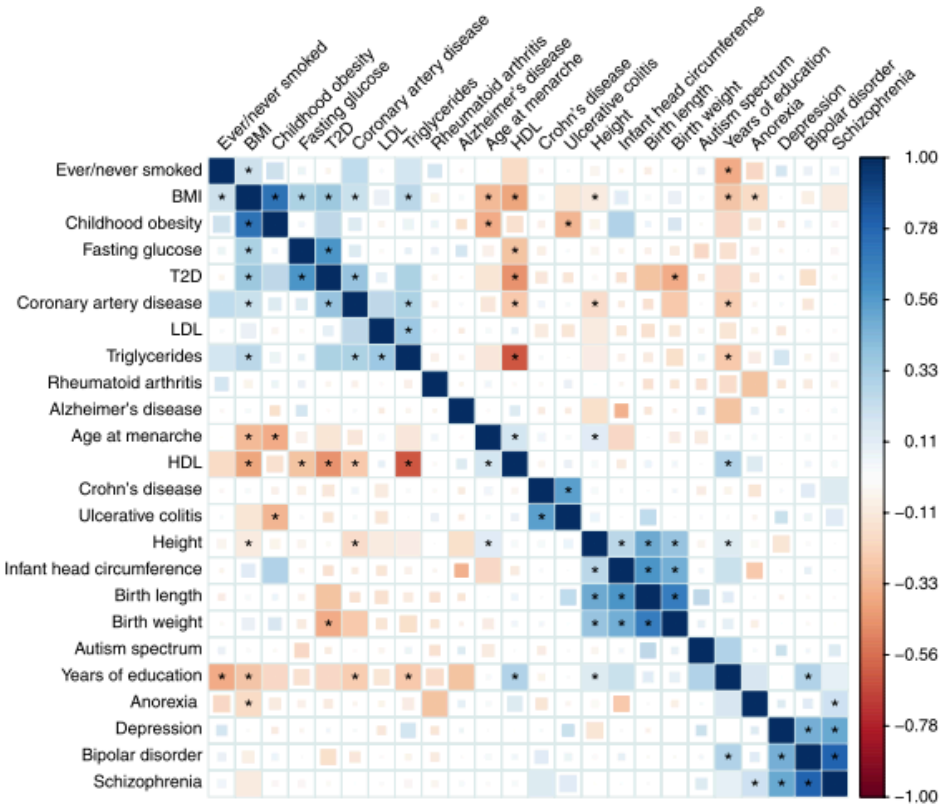


- *FADS1* signal was only detected by jointly analysing LDL and triglycerides

Korte et al. *Nat Gen* (2012)

Genetic Correlation

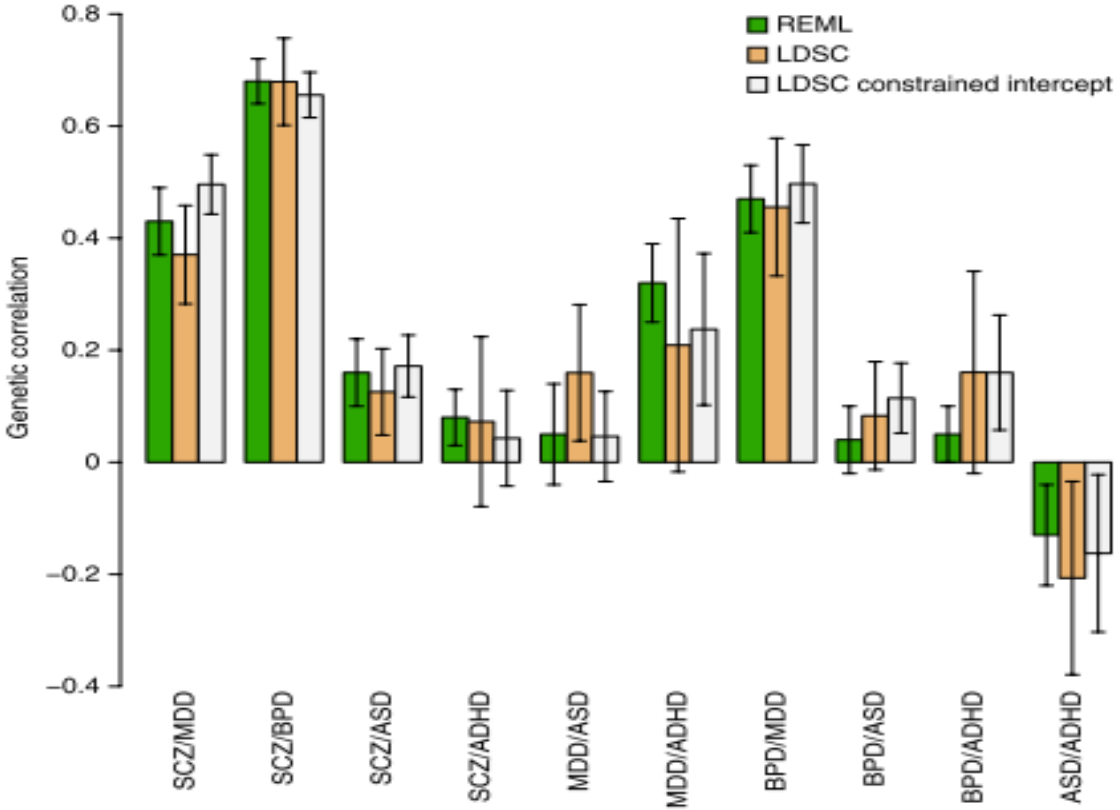
- The proportion of covariance between two phenotypes that is due to genetic factors
- Both univariate (LDSC) and multivariate (GCTA, BOLT-REML) methods exist



Bulik-Sullivan et al. *Nat Gen*, 2015

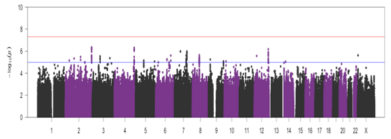
Genetic Correlation

- The proportion of covariance between two phenotypes that is due to genetic factors
- Both univariate (LDSC) and multivariate (GCTA, BOLT-REML) methods exist



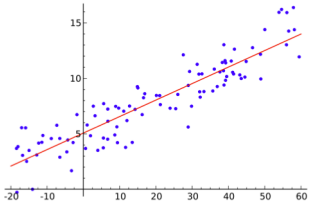
Bulik-Sullivan et al. *Nat Gen*, 2015

Polygenic Risk Scores (PRS)



Base data
(Summary statistics)

Target data (individual-level genotype): Construct risk scores using base phenotype effect estimates



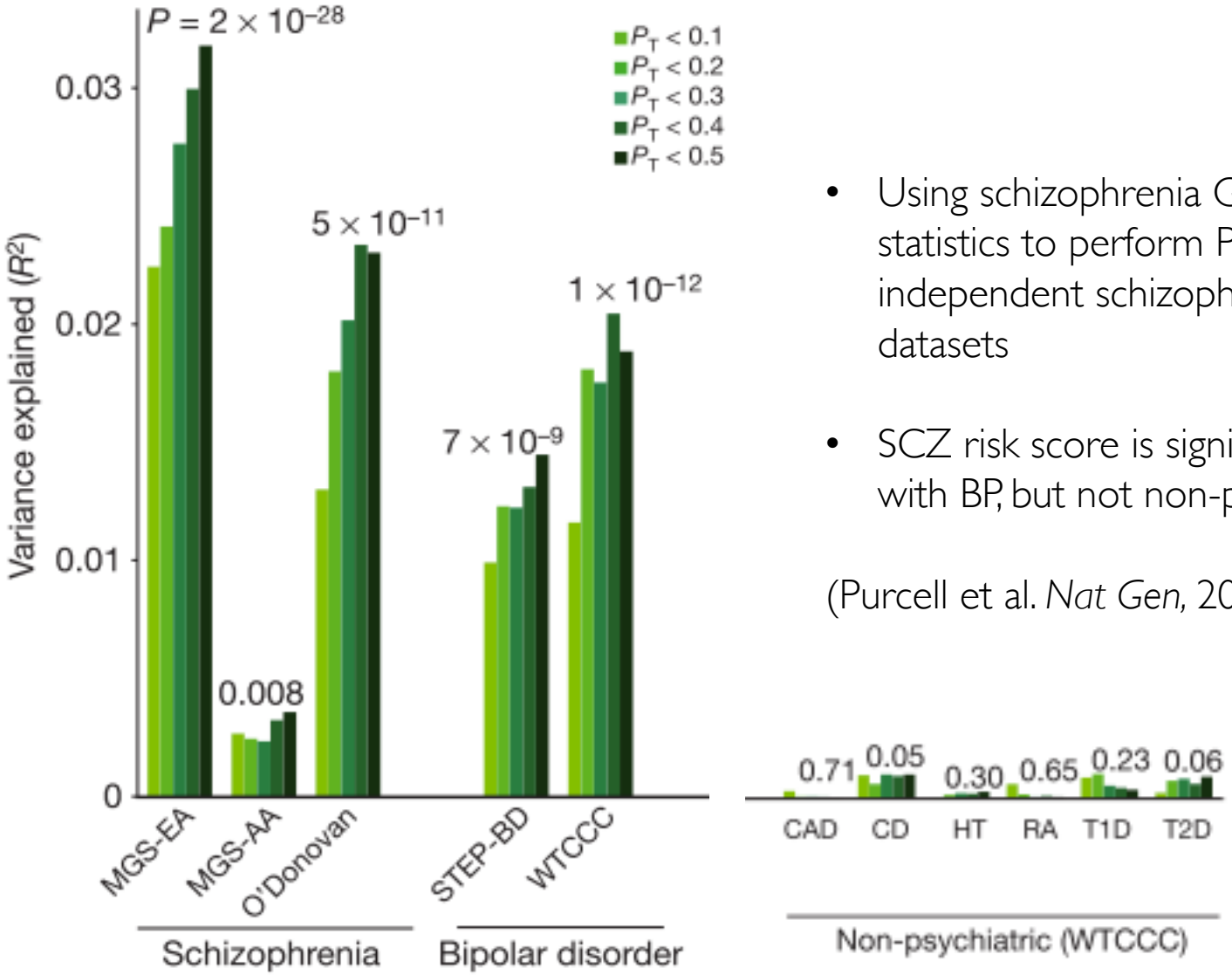
Regress target phenotype on scores to obtain (pseudo) R^2

- Base data: p-values and effect estimates (betas or odds ratio)
- Can use only established risk loci or all SNPs below a certain p-value threshold in base data
- Base and target phenotype can be the same or different

Risk score formula (EA=effect allele):

$$\frac{\left(\#\ of\ EA_{SNP_1} * log(OR)\right) + \left(\#\ of\ EA_{SNP_2} * log(OR)\right) + \ldots + \left(\#\ of\ EA_{SNP_K} * log(OR)\right)}{\#\ SNPs}$$

Polygenic Risk Scores

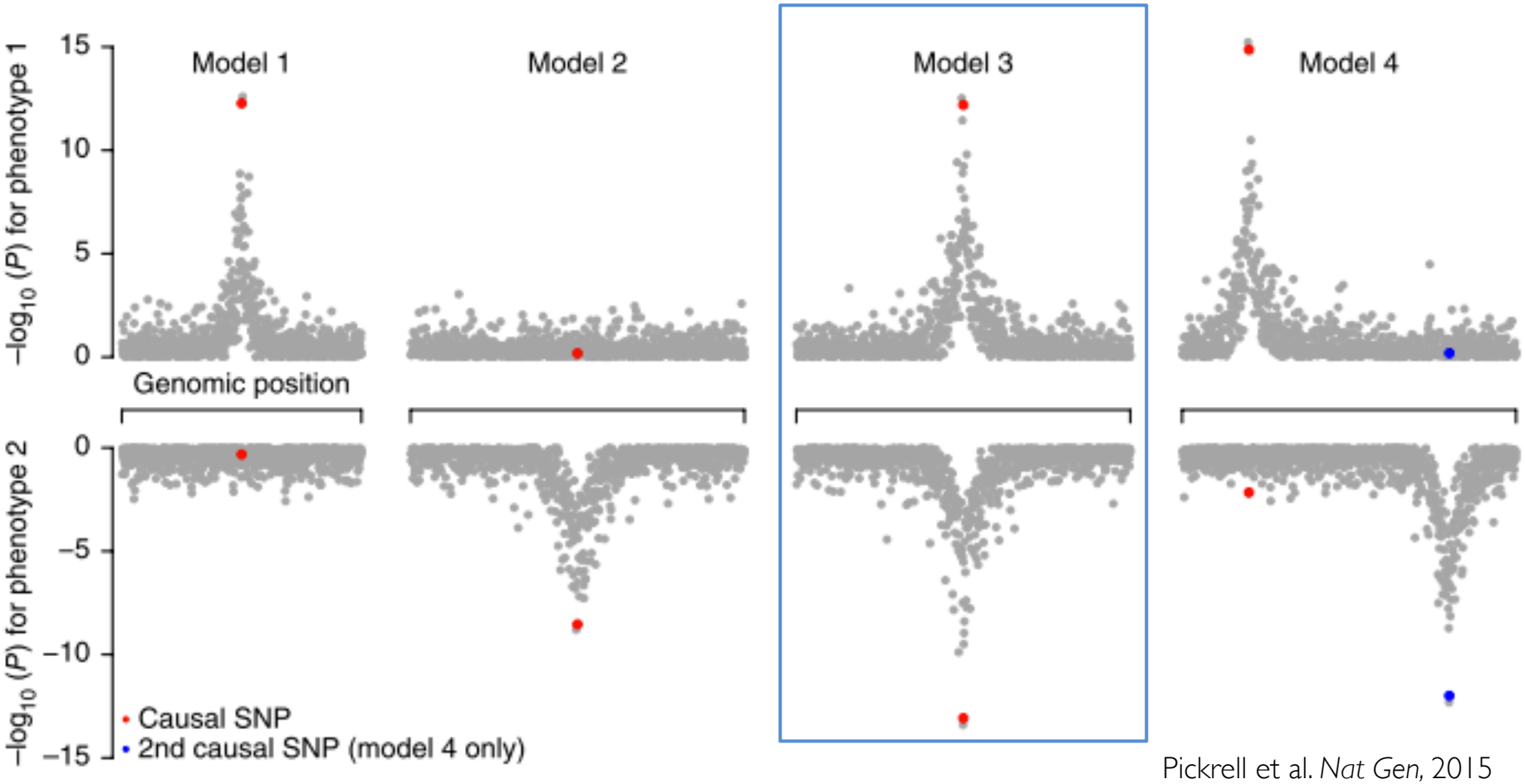


- Using schizophrenia GWAS summary statistics to perform PRS in independent schizophrenia and bipolar datasets
- SCZ risk score is significantly associated with BP, but not non-psychiatric traits

(Purcell et al. *Nat Gen*, 2009)

Co-localisation Analysis

- Originally used for eQTL and expression data
- What is the probability that a genomic region harbours signals for two traits?



Pickrell et al. *Nat Gen*, 2015

Follow-up and replication

- Interpretation of signals is not straightforward
  - Which subset of traits is associated?
  - Could the signal be driven by one trait?
  - What does it mean biologically?
- Replication more difficult than for single trait
  - Suitable independent datasets might not be available
  - Replicate in all or a subset of traits?

References

1. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet. 2013;14(7):483–95. (https://doi.org/10.1038/nrg3461)

2. Hodgkin J. Seven types of pleiotropy. Int J Dev Biol. 1998;505(3):501–5.

3. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460(7256):748–52. (https://doi.org/10.1038/nature08185)

4. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan L, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47(11):1236–41. (https://doi.org/10.1038/ng.3406)

5. Pickrell J, Berisa T, Segurel L, Tung JY, Hinds D. Detection and interpretation of shared genetic influences on 42 human traits [Internet]. Nature Genetics. 2015 May.

6. Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nat Genet. 2012;44(9):1066–71. (https://doi.org/10.1038/ng.2376)

7. Timpson NJ, Lawlor DA, Harbord RM, Gaunt TR, Day INM, Palmer LJ, Hattersley AT, Ebrahim S, Lowe GDO, Rumley A, et al. C-reactive protein and its role in metabolic syndrome: mendelian randomisation study. Lancet (London, England). 2005;366(9501):1954–9. (https://doi.org/10.1016/S0140-6736(05)67786-0)

APPENDIX

Univariate methods

Method	PMID	Year	Data	Trait #	Trait type	Implementation
CPMA	21852963	2011	P-values	>2	Any	R
ASSET	22560090	2012	Betas, SEs	≥2	Any	R
CPASSOC	25500260	2015	Z-scores	≥2	Any	R
MultiMeta	25908790	2015	Betas, SEs	≥2	Any	R
MTAG	NA	2017	Betas, SEs	≥2	Any	Python
cFDR	25658688	2015	P-values	2	Any	R
Bayesian overlap	26411566	2015	P-values	2	Any	NA
metaCCA	27153689	2016	Betas, SEs	≥2	Any	R
GPA	25393678	2014	P-values	2	Any	R
GPA-MDS	27868058	2016	P-values	≥2	Any	R
fastPAINTOR	27663501	2017	Z-scores	≥2	Any	C++
EPS	27153687	2016	P-values	2	Any	Matlab

Variant prioritisation and fine-mapping

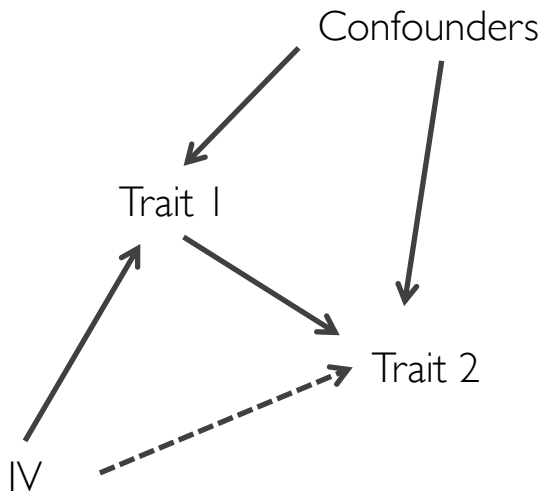
Multivariate methods

Method	PMID	Year	Data	Trait #	Trait type	Implementation
FBAT-PC	16646795	2004	Raw	≥2	Any	C
PCHAT	17922480	2008	Raw	≥2	Any	Fortran
mvPlink	19019849	2009	Raw	≥2	Any	C++
MTMM	22902788	2012	Raw	2	ND	R
GEMMA	24531419	2014	Raw	≥2	ND	C/C++
mvLMM	25724382	2015	Raw	≥2	ND	Python
GAMMA	27770036	2016	Raw	≥2	ND	R
B_EGEE	18924135	2009	Raw	2	Any	Fortran
PleioGRiP	22973300	2012	Raw	2	Binary	Java
mvBIMBAM	23861737	2013	Raw	≥2	ND	C/C++
Kendall's Tau	20711441	2010	Raw	≥2	Any	NA
MultiPhen	22567092	2012	Raw	≥2	Any	R
ATeMP	26479245	2015	Raw	≥2	Any	NA
BAMP	26493781	2015	Raw	≥2	Any	NA
TATES	23359524	2013	P-values	≥2	Any	R/Fortran
Extension to O'Briens	20583287	2010	Raw	≥2	Any	Upon request
Log-linear model	21849790	2011	Raw	≥2	Binary	NA
PET	25044106	2014	Raw	2	ND	R
FBAT-PC	16646795	2004	Raw	≥2	Any	C

ND=normally distributed

Mendelian randomisation (MR)

- Test for causal relationship between two traits using an instrumental variable (IV=genetic marker)
- MR estimate= $\frac{\beta_{IV-trait2}}{\beta_{IV-trait1}}$
- Other ways to obtain MR estimate, e.g. two-stage least squares analysis (2SLS)
- 3 key assumptions
  - IV is associated with trait 1
  - IV is not associated with confounding factors
  - IV is associated with trait 2 only via trait 1



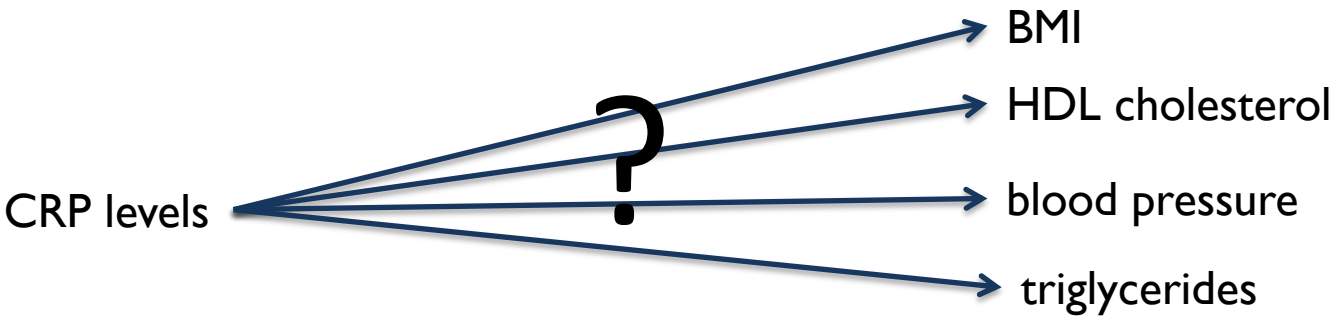
N.B. Unlike previous methods presented, MR assumes that there is <u>no</u> biological pleiotropy between the IV and the two traits analysed.

MR example

C-reactive protein and its role in metabolic syndrome: mendelian randomisation study

Lancet (2005)

Nicholas J Timpson, Debbie A Lawlor, Roger M Harbord, Tom R Gaunt, Ian N M Day, Lyle J Palmer, Andrew T Hattersley, Shah Ebrahim, Gordon D O Lowe, Ann Rumley, George Davey Smith



- 2SLS: Regress CRP levels on CRP SNPs, then regress second trait (e.g. BMI) on predicted values from this regression
- No association between IV CRP levels and metabolic syndrome traits