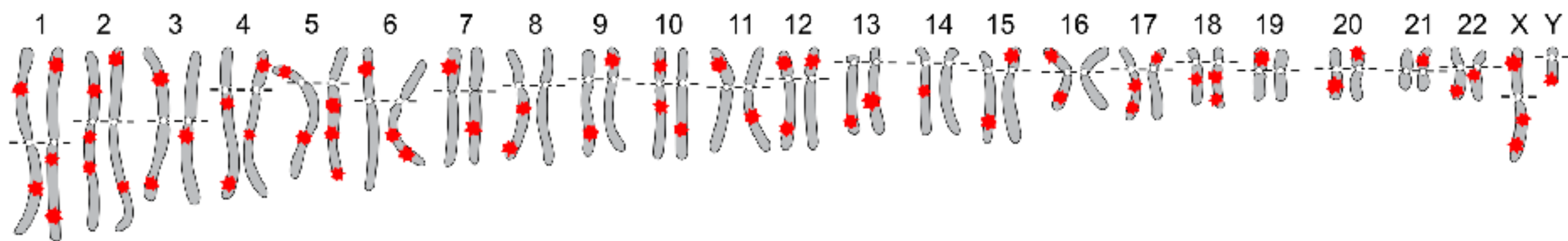


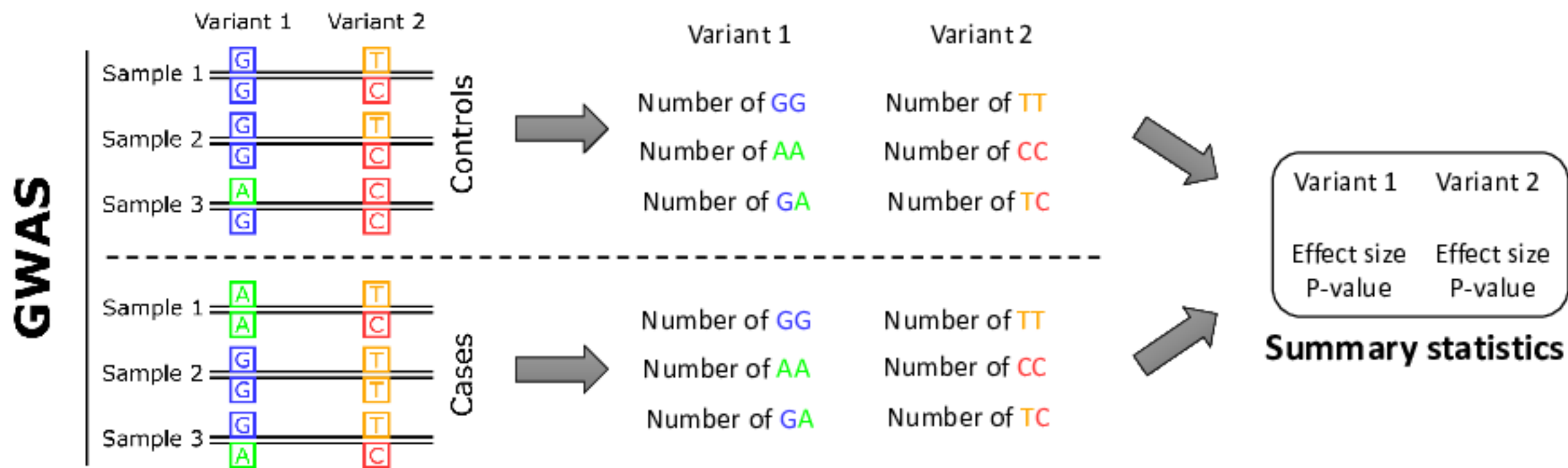
Beyond the standard GWAS

Advanced tools for post-GWAS analysis

Complex traits and GWAS



Complex traits -> multiple common variants with weak effects scattered throughout the genome -> GWAS



GWAS proved to be very successful

Genome-wide significant ($p \leq 5.0 \times 10^{-8}$)
SNP-trait associations, published in the
GWAS Catalog as of May 2018.

23,321 variants in total.



Digestive system disease	1003
Cardiovascular disease	701
Metabolic disease	229
Immune system disease	1250
Nervous system disease	1202
Liver enzyme measurement	154
Lipid or lipoprotein measurement	472
Inflammatory marker measurement	326
Hematological measurement	2249
Body measurement	1158
Cardiovascular measurement	679
Other measurement	5695
Response to drug	285
Biological process	1160
Cancer	1018
Other disease	1181
Other trait	4559

... however, there are a number of challenges

- Huge sample size is required to reach genome-wide significance for weak association signals scattered throughout genome
- True association signals are obscured by the LD structure, hindering understanding of causal underpinnings

... however, there are a number of challenges

- Huge sample size is required to reach genome-wide significance for weak association signals scattered throughout genome
 - True association signals are obscured by the LD structure, hindering understanding of causal underpinnings
-
- Is it possible to circumvent/mitigate these obstacles?
 - What can we do on top of the standard GWAS?

Two GWAS can be better than one

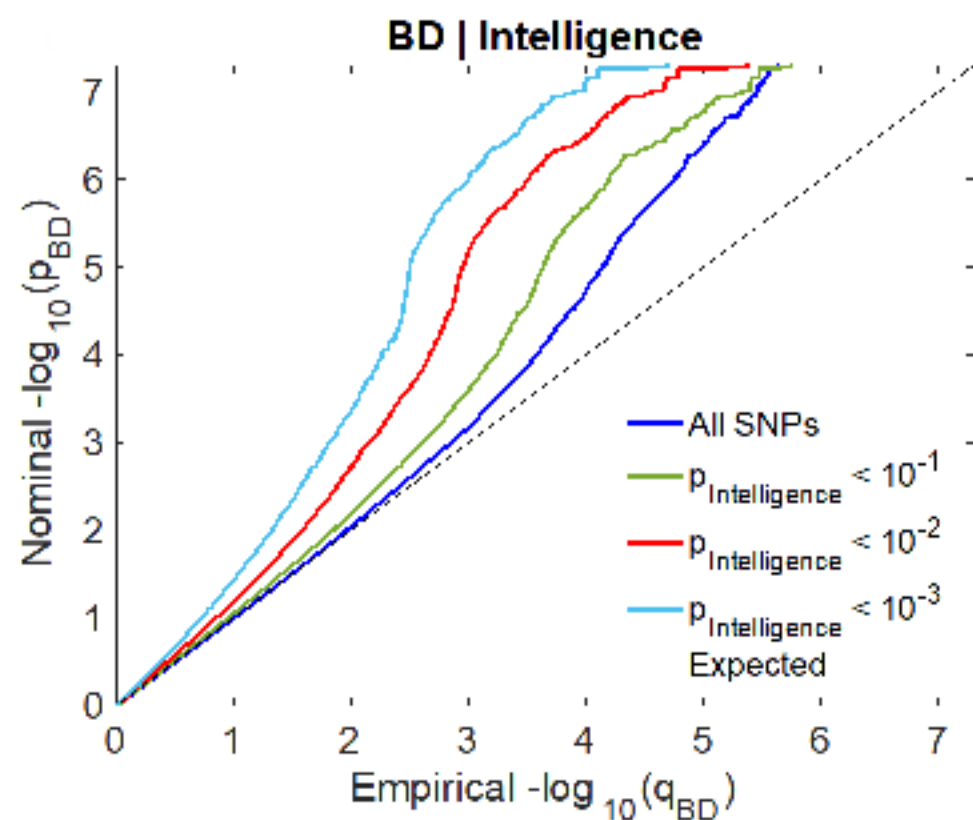
- Genetic overlap between different complex traits is ubiquitous
- Investigate common genetic basis of different phenotypes
- Leverage shared genetic underpinnings to boost power for genetic discovery

Conditional quantile-quantile (QQ) plots

- **Purpose**
 - Visual assessment of polygenic overlap between two phenotypes
- **Required input data**
 - GWAS summary statistics of two phenotypes
- **Key features**
 - A simple yet instructive way to get a general idea of whether two traits have a shared genetic background
 - Agnostic to direction of effects (can detect genetic overlap when genetic correlation is absent)

Conditional quantile-quantile (QQ) plots

- **Purpose**
 - Visual assessment of polygenic overlap between two phenotypes
- **Required input data**
 - GWAS summary statistics of two phenotypes
- **Key features**
 - A simple yet instructive way to get a general idea of whether two traits have a shared genetic background
 - Agnostic to direction of effects (can detect genetic overlap when genetic correlation is absent)



Smeland et al. (*in press*)

$$r_g = -0.05, p = 0.08$$

Stahl et al. (*bioRxiv*, 2018)

Conditional/Conjunctional False Discovery Rate

- **Purpose**
 - Boost identification of phenotype-associated loci
 - Identify loci shared between two different phenotypes
- **Required input data**
 - GWAS summary statistics of two phenotypes (non-overlapping samples)
- **Key features**
 - Exploits overlapping associations between different phenotypes
 - Prioritizes loci with elevated association in both phenotypes
 - Enhances loci discovery and improves replication rates of discovered risk variants
 - Makes no assumptions about distribution of association signals (model free)

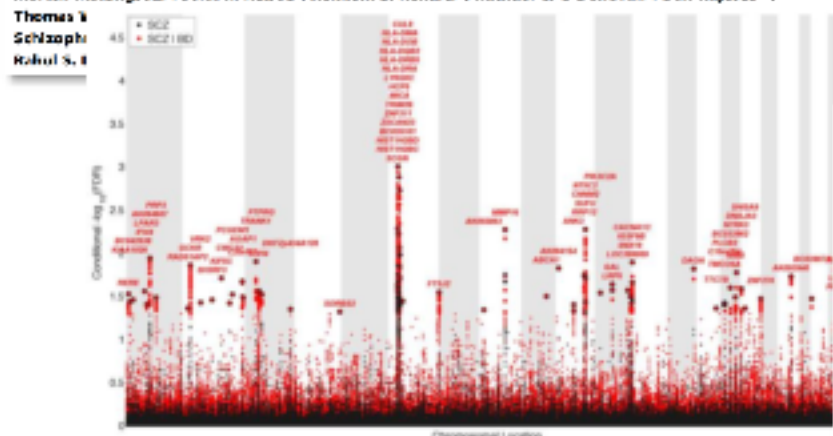
Conditional/Conjunctional False Discovery Rate

OPEN ACCESS freely available online

PLOS

Improved Detection of Common Variants Associated with Schizophrenia and Bipolar Disorder Using Pleiotropy-Informed Conditional False Discovery Rate

Olaf A. Andreassen^{1,2,3,4}, Wesley K. Thompson², Andrew J. Schork^{1,2,3}, Stephan Ripke⁷, Morten Mattingdal¹, John R. Kessler⁵, Kenneth S. Kendler⁶, Michael C. O'Donovan⁵, Dan Rujescu¹⁰, Thomas W. Schizophrenia¹, Rahul S. K.



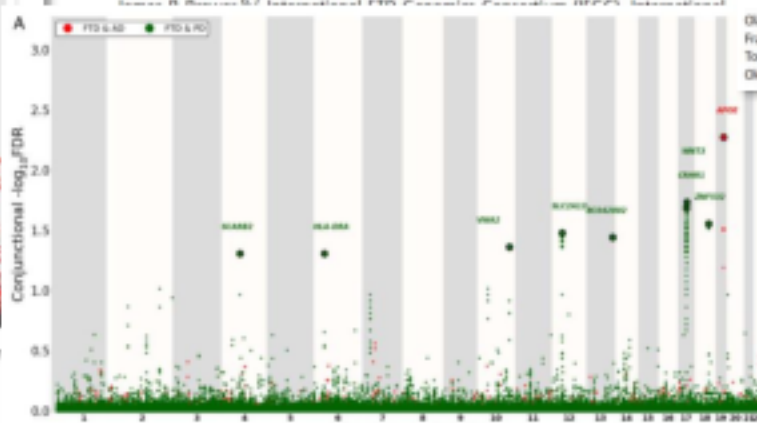
51 novel SCZ loci
11 novel BIP loci
14 loci shared between SCZ and BIP

Neurogenetics

RESEARCH PAPER

Genetic architecture of sporadic frontotemporal dementia and overlap with Alzheimer's and Parkinson's diseases

Raffaella Ferrari¹, Yunpeng Wang², Jana Vandrovcova^{1,3}, Sebastian Gueff^{1,3}, Aree Witteolar², Celeste M. Karch⁴, Andrew J. Schork⁵, Chun C. Fan⁵

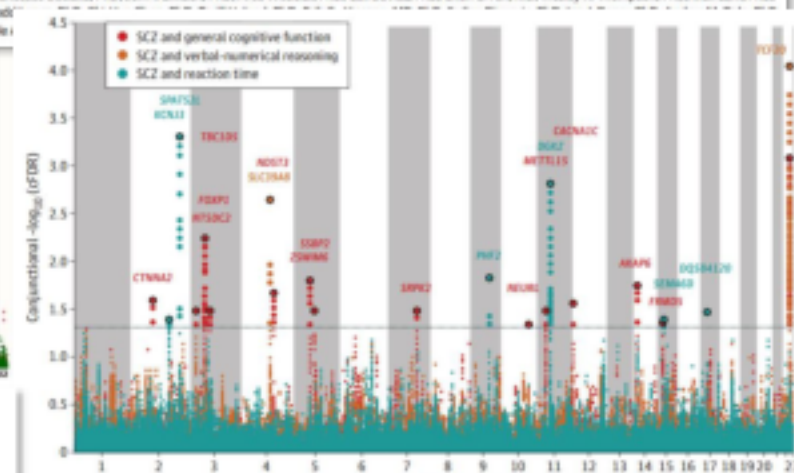


8 loci shared between FTD and PD
1 shared locus between FTD and AD
13 novel FTD loci

JAMA Psychiatry | Original Investigation

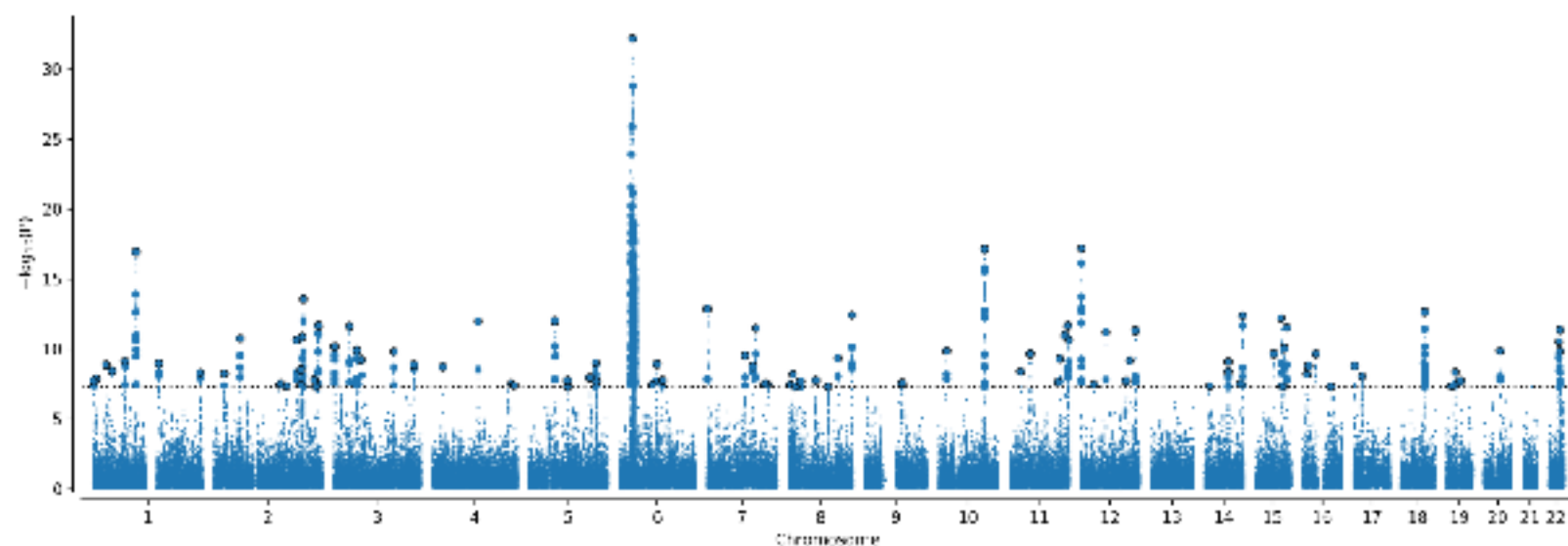
Identification of Genetic Loci Jointly Influencing Schizophrenia Risk and the Cognitive Traits of Verbal-Numerical Reasoning, Reaction Time, and General Cognitive Function

Olaf B. Smeland, MD, PhD; Oleksandr Frei, PhD; Karolina Kuoppi, PhD; W. David Hill, PhD; Wen Li, PhD; Yunpeng Wang, PhD; Florian Krull, PhD; Francesco Bettella, PhD; Jon A. Eidsen, PhD; Aree Witteolar, PhD; Gail Davies, PhD; Chun C. Fan, MD; Wesley K. Thompson, PhD; Max Lani, PhD; Ole A.



21 loci shared between SCZ and cognitive traits

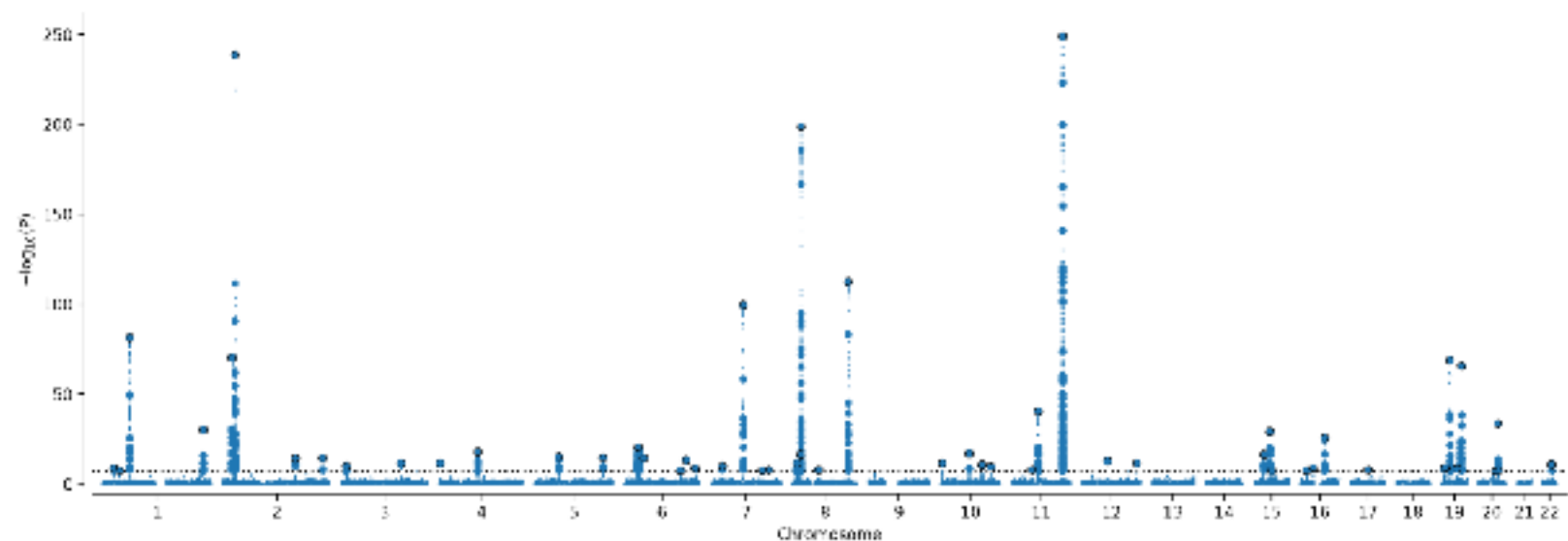
Different degrees of complexity



SCZ (PGC, 2014)

N cases = 35476

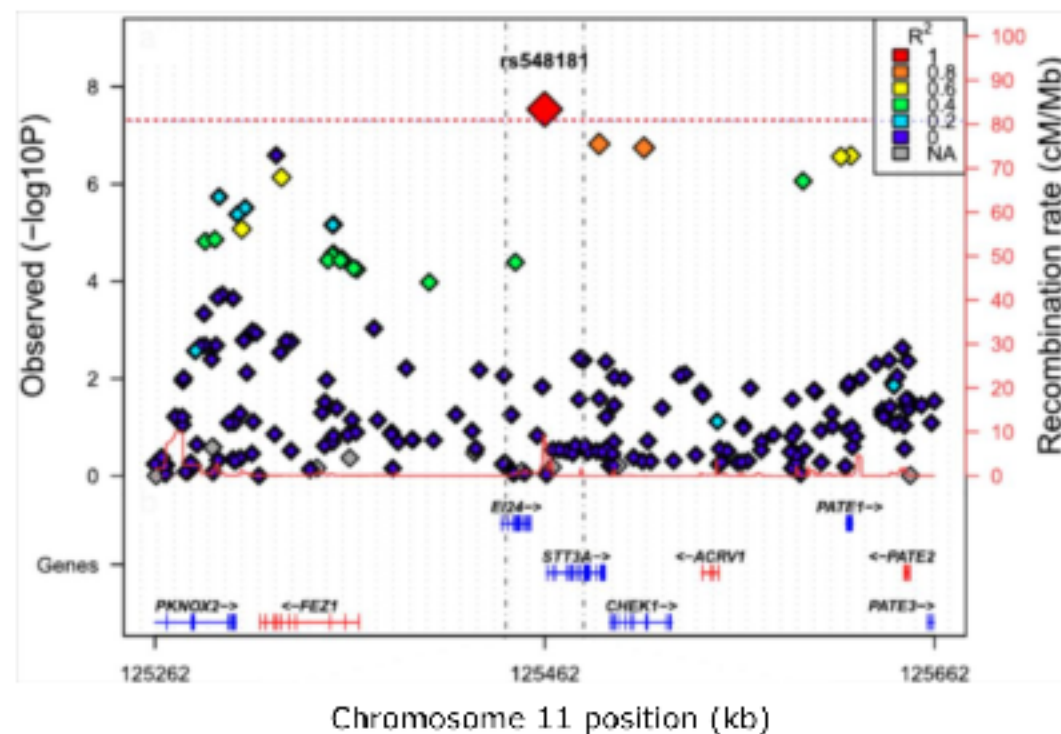
N controls = 46839



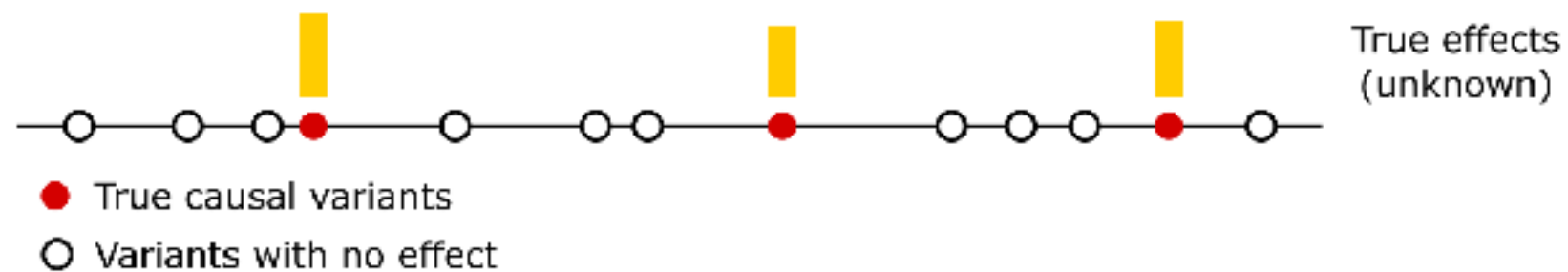
Triglycerides (LIPIDS , 2013)

N = 188577

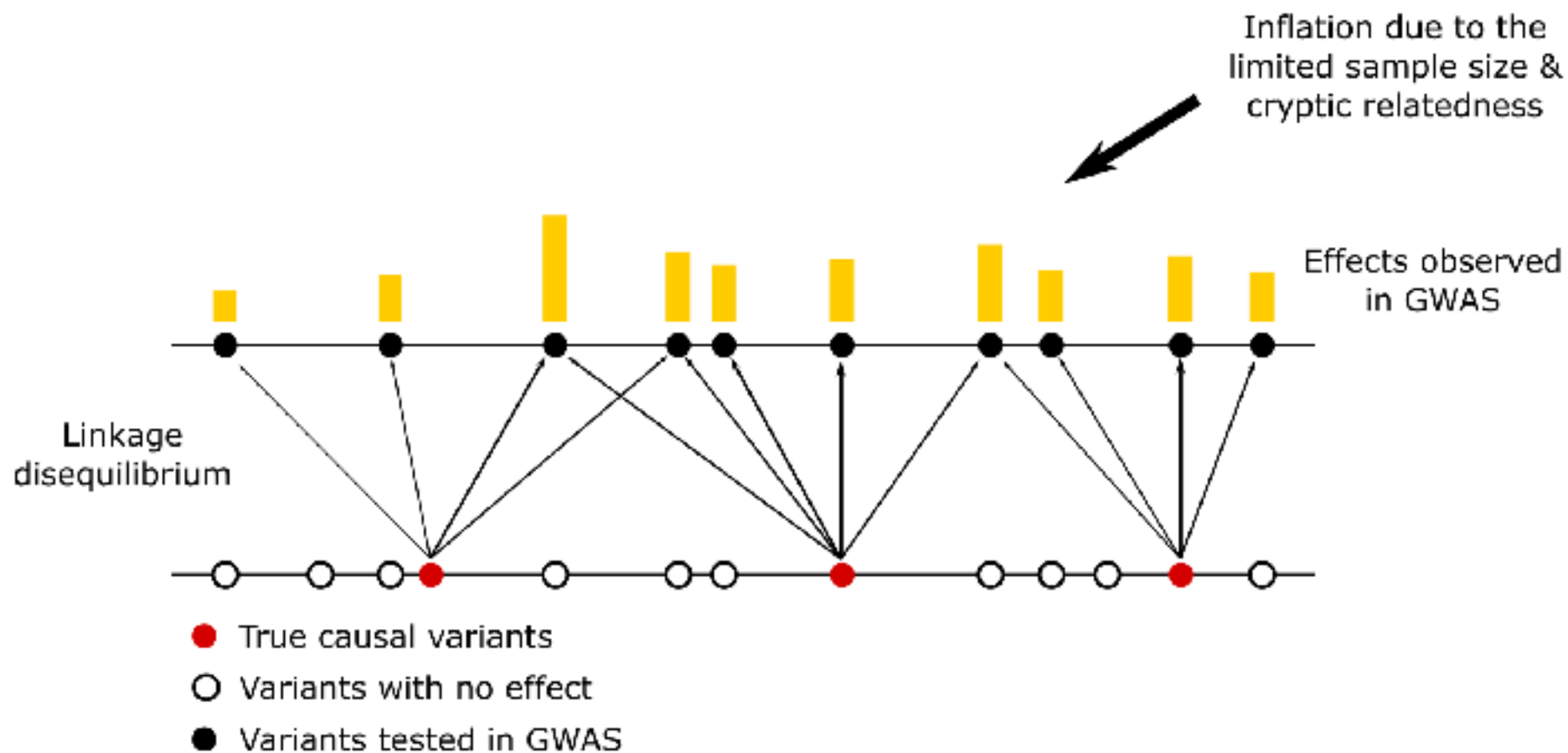
What we see in GWAS



What we want to see in GWAS



What happens here?



Causal Mixture Model (MiXeR)

- **Purpose**
 - Unravel LD structure
 - Investigate general properties of phenotype's genetic architecture:
 - estimate proportion of causal variants
 - estimate magnitude of effect sizes
 - assess genetic overlap between two phenotypes
- **Required input data**
 - GWAS summary statistics
- **Key features**
 - Effect sizes of variants are modeled as a mixture distribution of null (no effect) and non-null effects
 - Incorporates detailed LD structure of the genotype
 - Accounts for sample overlap and stratification

Frei et al. (*bioRxiv*, 2018)

Code is publicly on GitHub: <https://github.com/precimed/mixer>

Causal Mixture Model. Details

$$z|\pi, \sigma_{\beta}^2, \sigma_0^2 = \epsilon + \sqrt{N} \sum_{\substack{j \text{ in LD} \\ \text{with } i}} \sqrt{H_j} r_j \beta_j = \epsilon + \sum_{\substack{j \text{ in LD} \\ \text{with } i}} \xi_j$$

$$\beta_j = \begin{cases} 0 & , 1 - \pi \\ N(0, \sigma_{\beta,c}^2), & \pi \end{cases} \quad \begin{array}{l} \text{“true” effect of } j\text{-th LD neighbor} \end{array}$$

π - proportion of causal variants

σ_{β}^2 - phenotypic variance per variant

r_j - correlation j -th LD neighbor

N - number of genotyped individuals

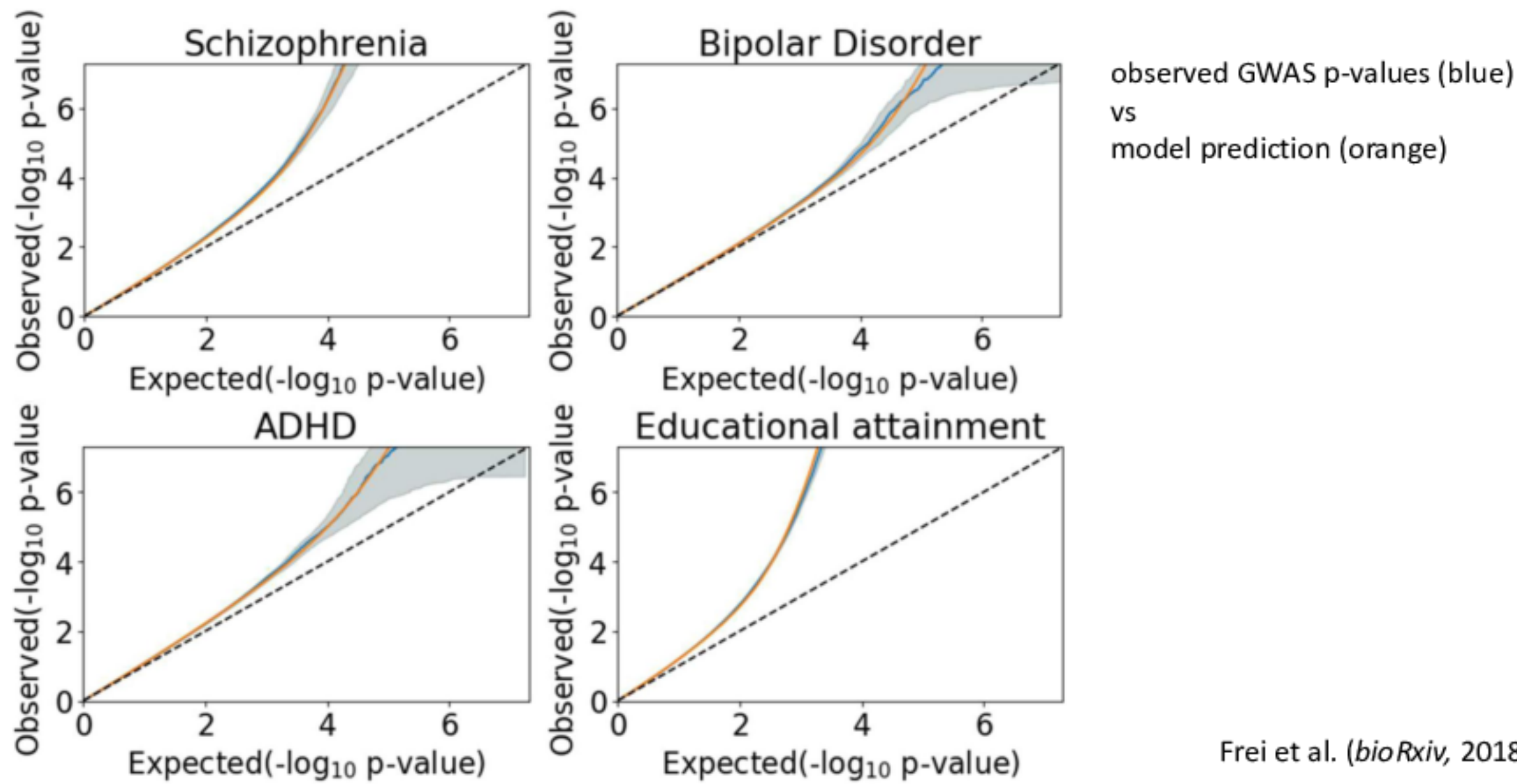
H_j - heterozygosity of j -th LD neighbor

$\epsilon \sim N(0, \sigma_0^2)$ - inflation

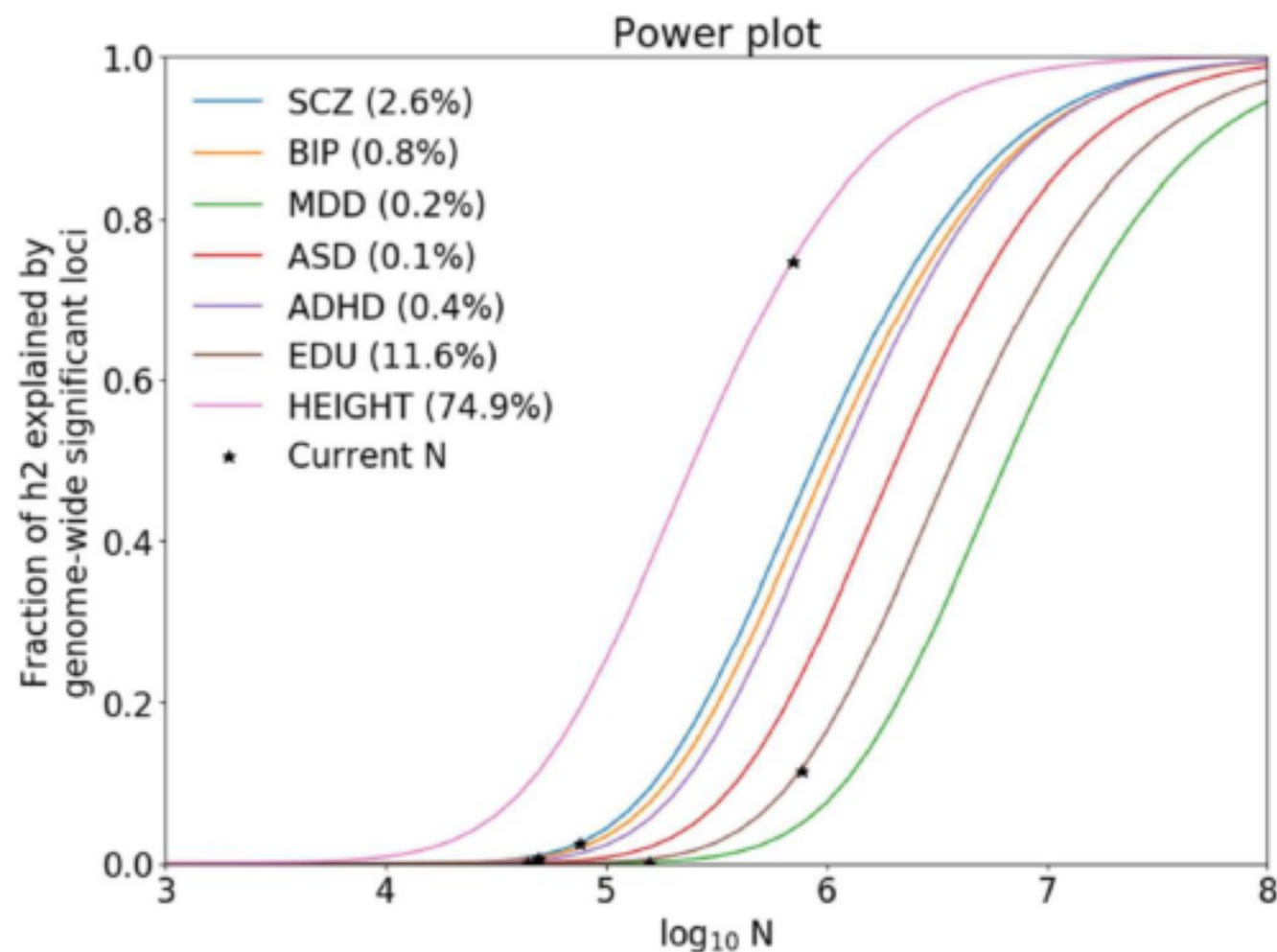
$$\xi_j = \begin{cases} 0 & , 1 - \pi \\ N(0, \sigma_{e,j}^2), & \pi \end{cases} \quad \begin{array}{l} \text{contribution of } j\text{-th LD neighbor to the observed effect, } \sigma_{e,j}^2 = N_i r_{ij}^2 H_j \sigma_{\beta}^2 \end{array}$$

$$pdf_z(z_0) \xrightarrow[\pi, \sigma_{\beta}^2, \sigma_0^2]{} max, \quad z_0 \text{ - z-score estimated in GWAS}$$

Causal Mixture Model. QQ plot reconstruction



Causal Mixture Model. GWAS power prediction

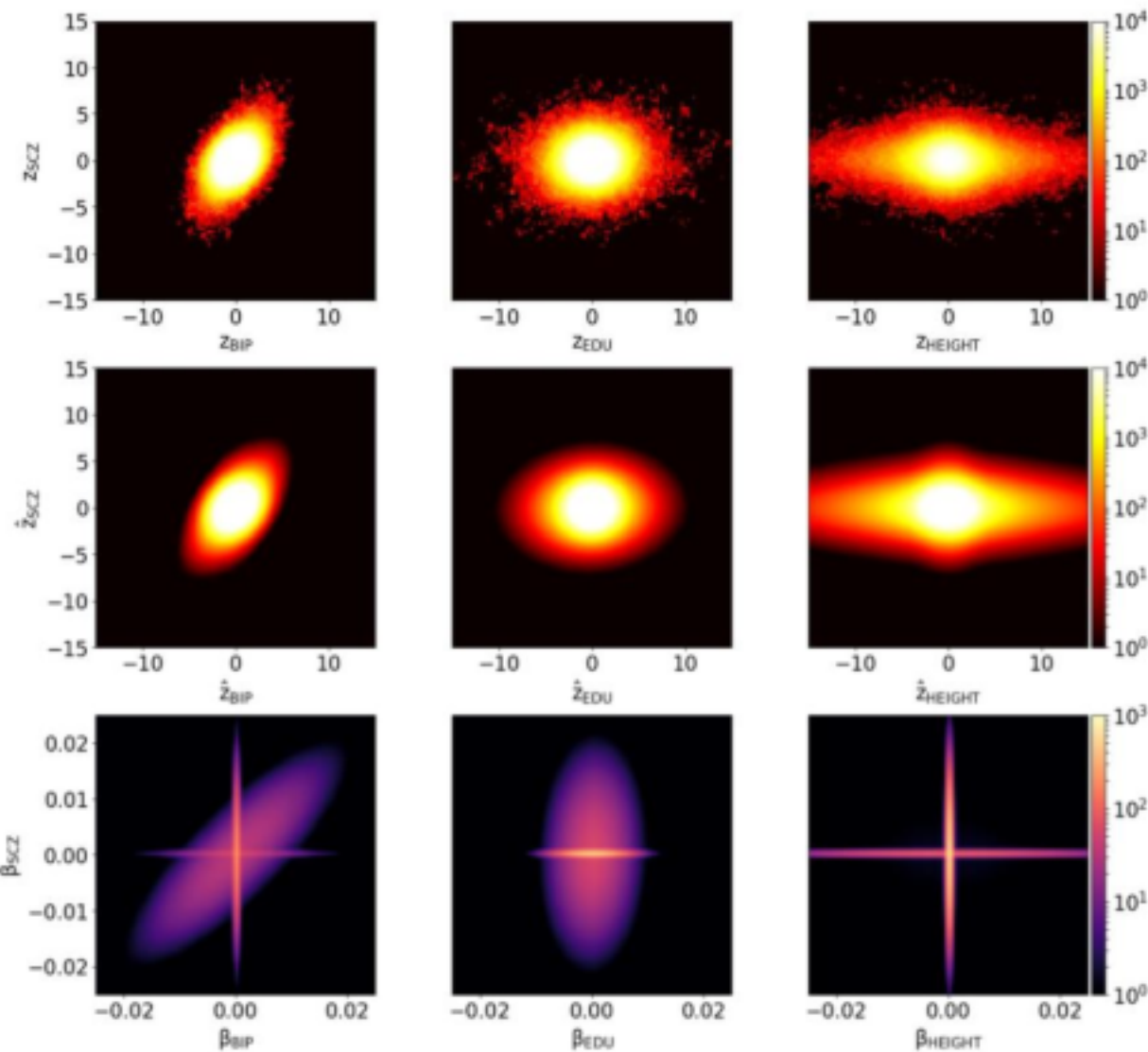


Proportion of SNP-heritability, captured by genome-wide significant SNPs, projected to the future GWAS sample size (N).

Values for current GWAS sample sizes are shown in parentheses.

Why not add a second trait here?

Bivariate MiXeR



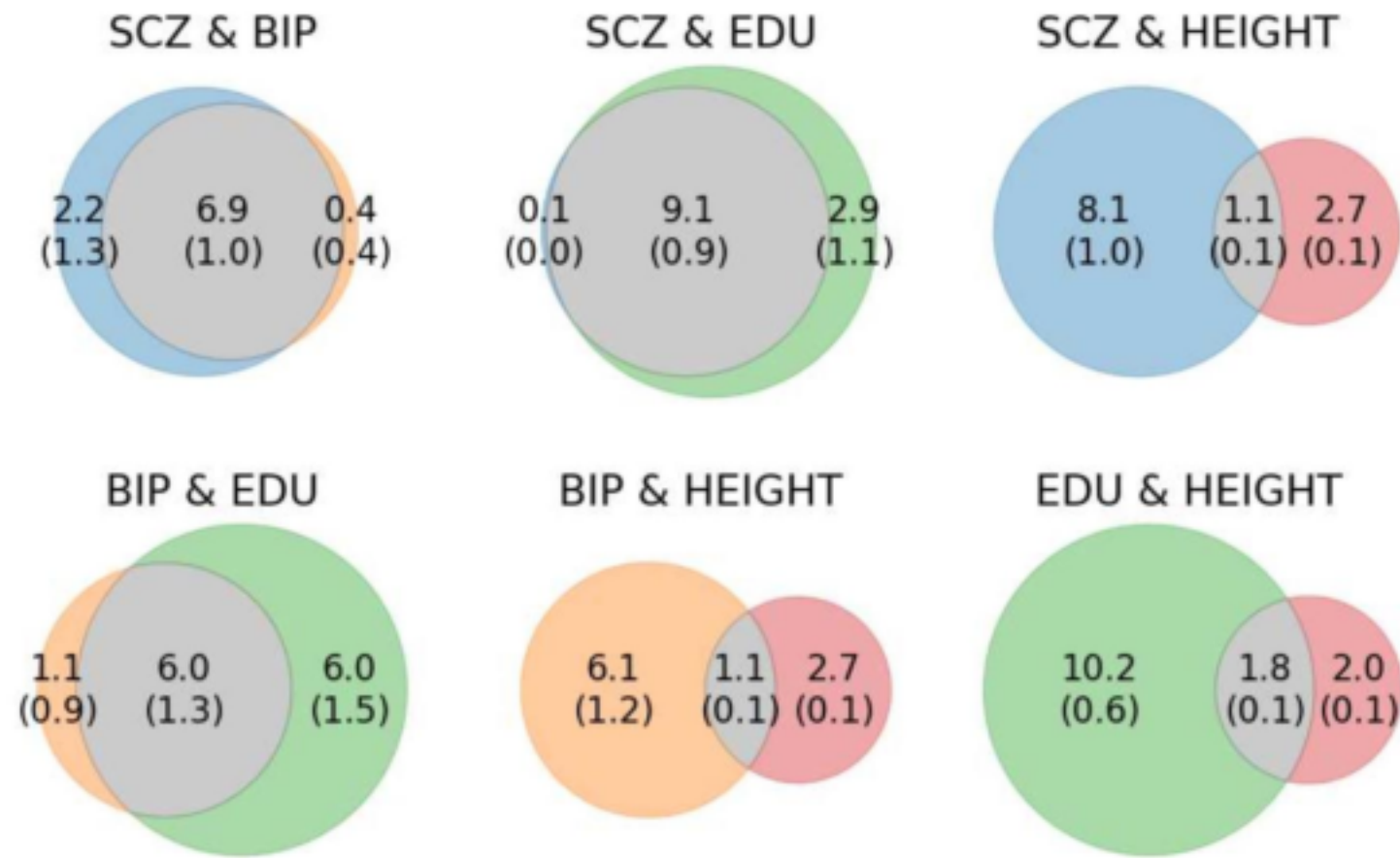
bivariate density of z scores observed in GWAS

model prediction at tag level

model prediction at causal level

Causal Mixture Model.

Polygenicity and genetic overlap



The number of causal variants per 1,000 variants, explaining 90% of SNP heritability in each phenotype (and corresponding standard errors).

The size of circles reflects the polygenicity.

Causal Mixture Model. Under development

- Incorporate information on functional annotations and estimate differential enrichment between different annotation categories
- Loci discovery
- Improved PRS