CityU
香港城市大學
City University of Hong Kong

GIFT: Conditional TWAS for fine-mapping candidate causal genes

Speaker: Yuekai Li

Major: Biostatistics

CONTENTS

1. Background

2. Challenges

3. Methods

4. Results

01

Background

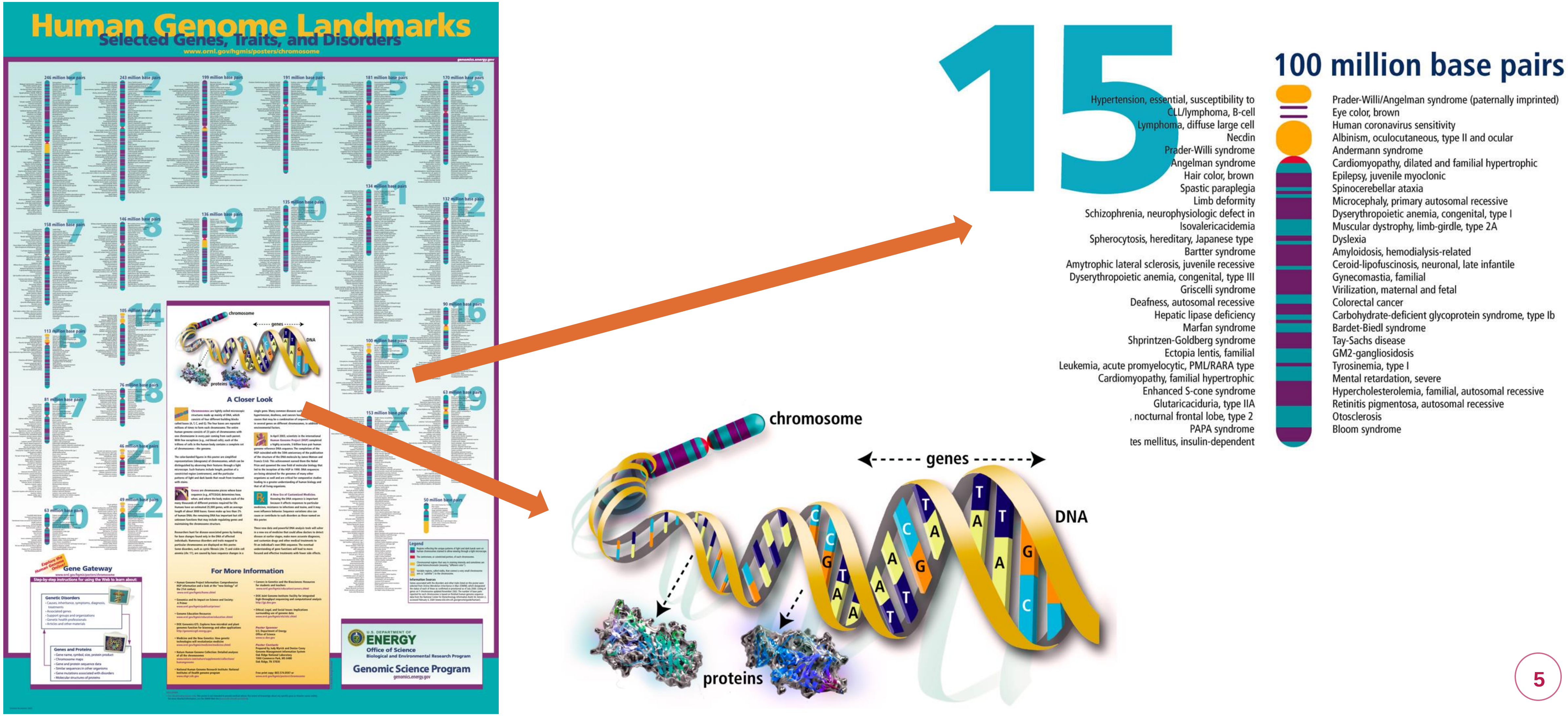
Outline

01 The relationship between traits, genes, SNPs

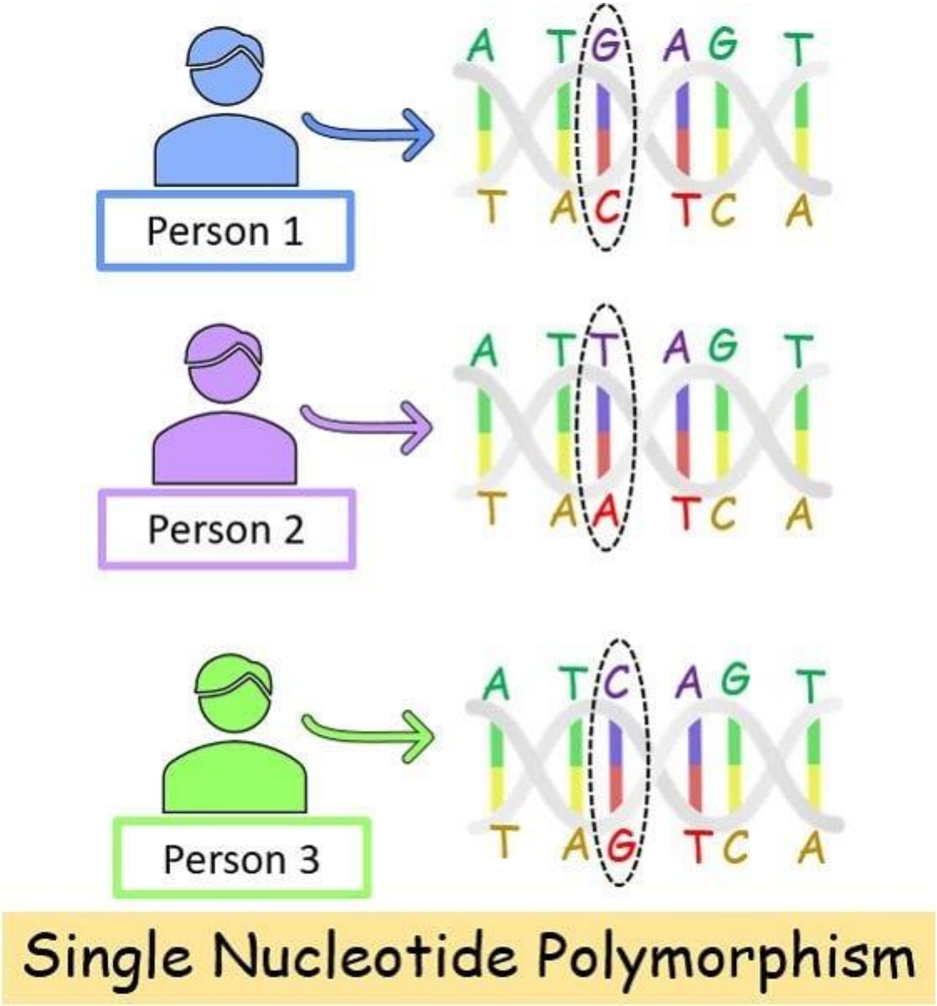
02 What is GWAS

03 The motivation of TWAS

■ The relationship between traits, genes, SNPs


Human Genome Landmarks
Selected Genes, Traits, and Disorders
www.ornl.gov/hgmis/posters/chromosome
A Closer Look
For More Information
Gene Gateway
U.S. DEPARTMENT OF ENERGY
Office of Science
Biological and Environmental Research Program
Genomic Science Program
genomics.energy.gov
15
100 million base pairs
Hypertension, essential, susceptibility to
CLL/lymphoma, B-cell
Lymphoma, diffuse large cell
Necdin
Prader-Willi syndrome
Angelman syndrome
Hair color, brown
Spastic paraplegia
Limb deformity
Schizophrenia, neurophysiologic defect in
Isovalericacidemia
Spherocytosis, hereditary, Japanese type
Bartter syndrome
Amytrophic lateral sclerosis, juvenile recessive
Dyserythropoietic anemia, congenital, type III
Griscelli syndrome
Deafness, autosomal recessive
Hepatic lipase deficiency
Marfan syndrome
Shprintzen-Goldberg syndrome
Ectopia lentis, familial
Leukemia, acute promyelocytic, PML/RARA type
Cardiomyopathy, familial hypertrophic
Enhanced S-cone syndrome
Glutaricaciduria, type IIA
, nocturnal frontal lobe, type 2
PAPA syndrome
tes mellitus, insulin-dependent
Prader-Willi/Angelman syndrome (paternally imprinted)
Eye color, brown
Human coronavirus sensitivity
Albinism, oculocutaneous, type II and ocular
Andermann syndrome
Cardiomyopathy, dilated and familial hypertrophic
Epilepsy, juvenile myoclonic
Spinocerebellar ataxia
Microcephaly, primary autosomal recessive
Dyserythropoietic anemia, congenital, type I
Muscular dystrophy, limb-girdle, type 2A
Dyslexia
Amyloidosis, hemodialysis-related
Ceroid-lipofuscinosis, neuronal, late infantile
Gynecomastia, familial
Virilization, maternal and fetal
Colorectal cancer
Carbohydrate-deficient glycoprotein syndrome, type Ib
Bardet-Biedl syndrome
Tay-Sachs disease
GM2-gangliosidosis
Tyrosinemia, type I
Mental retardation, severe
Hypercholesterolemia, familial, autosomal recessive
Retinitis pigmentosa, autosomal recessive
Otosclerosis
Bloom syndrome
chromosome
genes
DNA
proteins
T A T G A C C A A G T A T C G T A A C C

The relationship between traits, genes, SNPs



- SNP is the replacement of a single base pair in the DNA sequence.

- SNP is the most common type of genetic variation.

- More than 600 million SNPs have been identified across the human genome in the world's population.

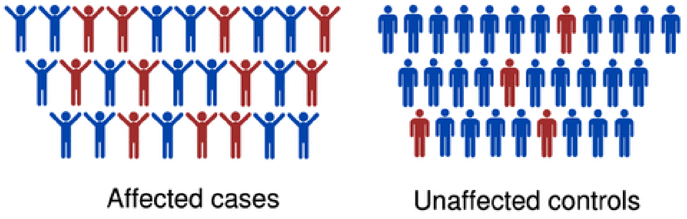
What is GWAS

Sequence Variation

ATGCCAGTGTTTCAAGATGCTTGGCCAGCTGGACGAGGGCGATGAC
ATGCCAGTGTTTCAAGATGTTTGGCCAGCTGGACGAGGGCGATGAC

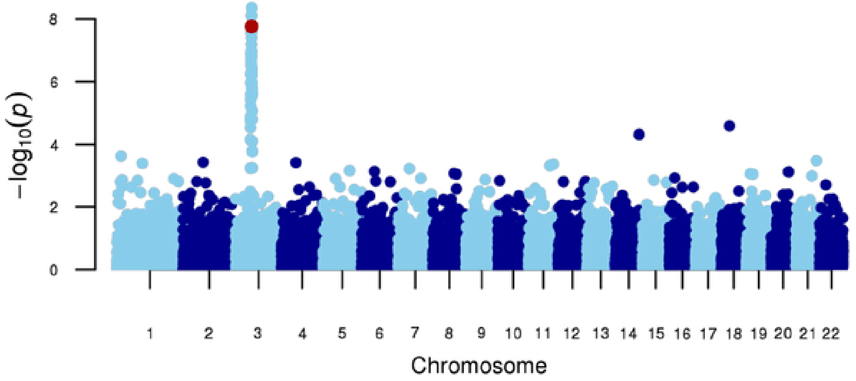
↓

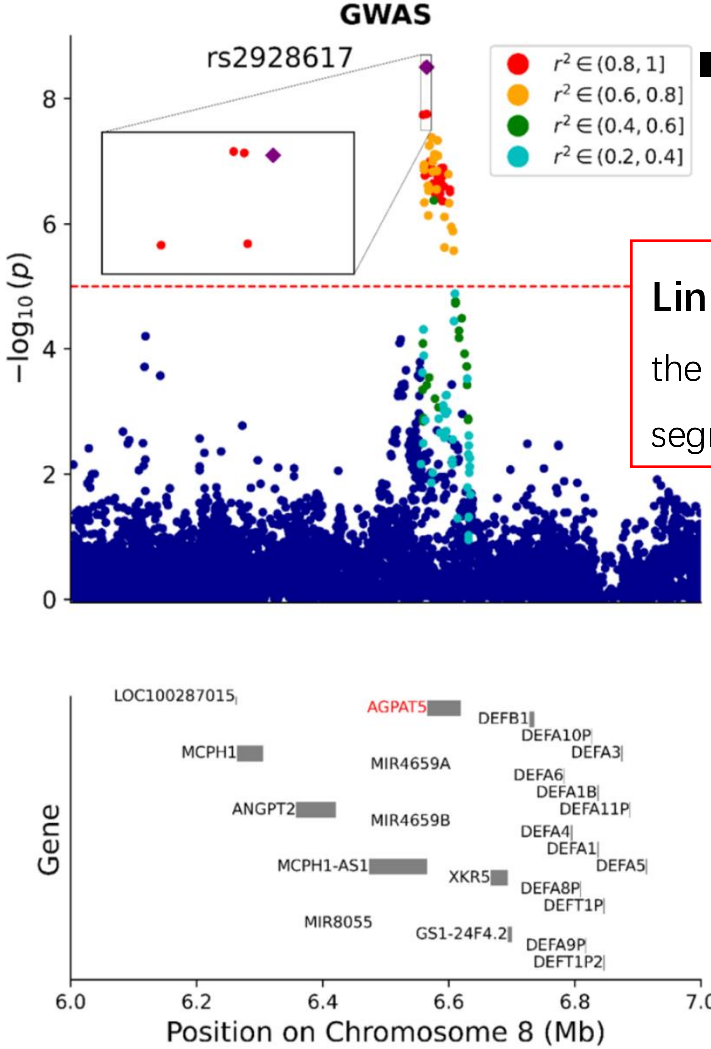
Disease



↓

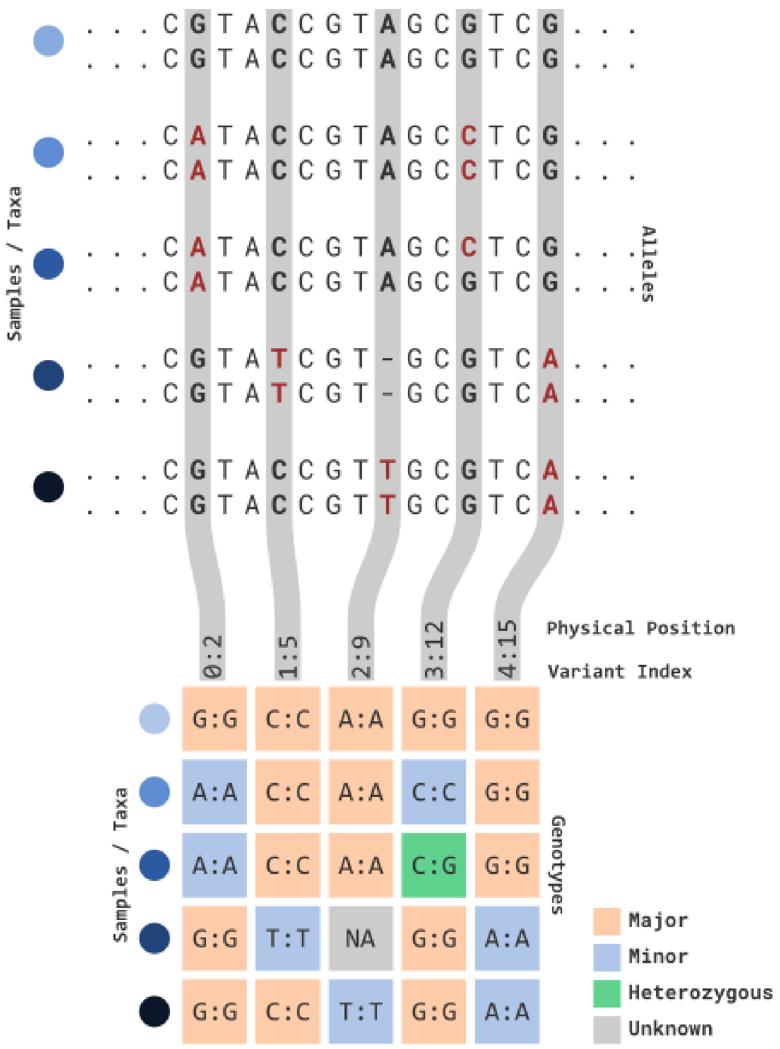
GWAS





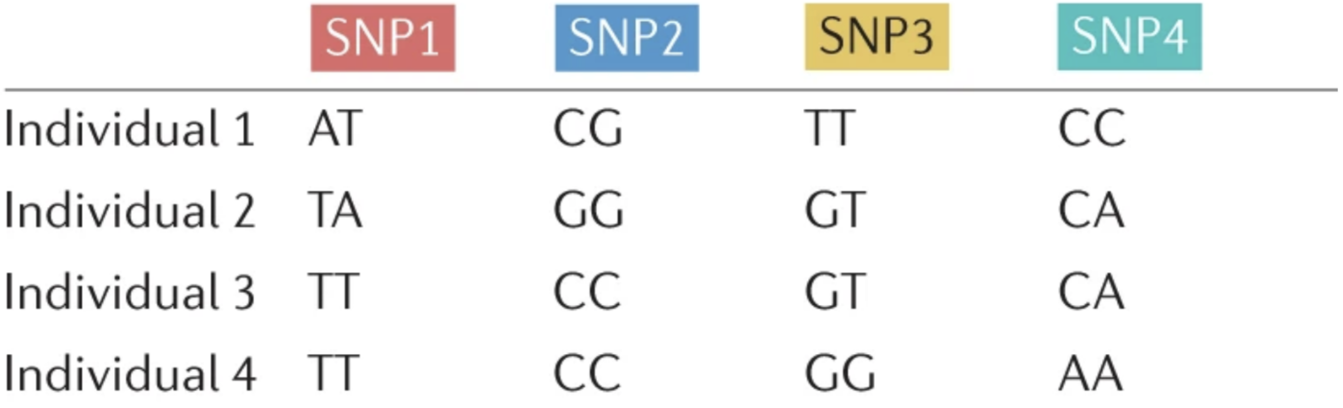
Linkage disequilibrium (LD) : the non-independent segregation of genetic variants.

What is GWAS


. . . C G T A C C G T A G C G T C G . . .
. . . C G T A C C G T A G C G T C G . . .
. . . C A T A C C G T A G C C T C G . . .
. . . C A T A C C G T A G C C T C G . . .
. . . C A T A C C G T A G C C T C G . . .
. . . C A T A C C G T A G C G T C G . . .
. . . C G T A T C G T - G C G T C A . . .
. . . C G T A T C G T - G C G T C A . . .
. . . C G T A C C G T T G C G T C A . . .
. . . C G T A C C G T T G C G T C A . . .
Samples / Taxa
Alleles
0:2
1:5
2:9
3:12
4:15
Physical Position
Variant Index
Samples / Taxa
Genotypes
Major
Minor
Heterozygous
Unknown

0:2	1:5	2:9	3:12	4:15
G:G	C:C	A:A	G:G	G:G
A:A	C:C	A:A	C:C	G:G
A:A	C:C	A:A	C:G	G:G
G:G	T:T	NA	G:G	A:A
G:G	C:C	T:T	G:G	A:A

Genotype data:



	SNP1	SNP2	SNP3	SNP4
Individual 1	AT	CG	TT	CC
Individual 2	TA	GG	GT	CA
Individual 3	TT	CC	GT	CA
Individual 4	TT	CC	GG	AA

Major=2 Heterozygous=1 Minor=0

For example, if we assume A is the major allele,

then A:A=2, A:C/C:A=1, CC=0

What is GWAS

Linear regression models for GWAS can be written as follows:

Fixed effect

$$\boldsymbol{Y} \sim \boldsymbol{W}\boldsymbol{\alpha} + \boldsymbol{X}_s \boxed{\boldsymbol{\beta}_s} + g + e \quad (1)$$

Its p-value measures the strength of the association between SNPs and trait.

$$g \sim N(0, \sigma_{\mathrm{A}}^2 \boldsymbol{\psi}) \quad (2)$$

$$e \sim N(0, \sigma_e^2 \boldsymbol{I}) \quad (3)$$

$\boldsymbol{Y}$: the phenotype value

$\boldsymbol{W}$: the vector of covariates including an intercept term

$\boldsymbol{\alpha}$: the corresponding vector of effect

$\boldsymbol{X}_s$: the genotype value for the genetic variant s

$\boldsymbol{\beta}_s$: the corresponding fixed effect

g : the random effect that captures the polygenic effect of other SNPs

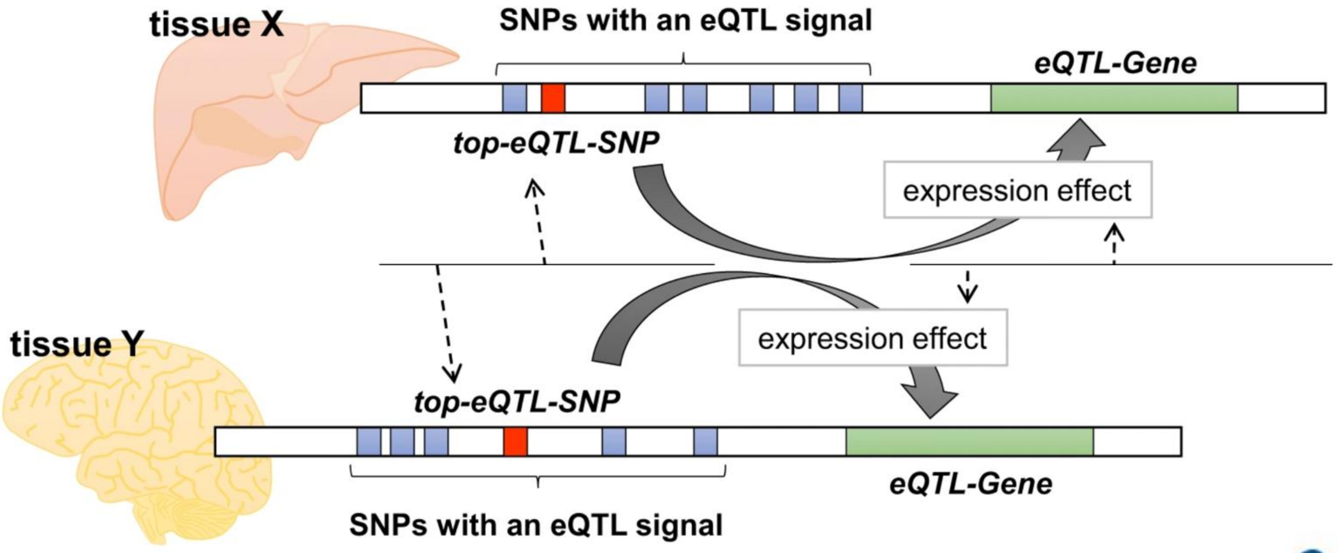
e : the random effect of residual errors

σ_A^2 : the additive genetic variation of the phenotype

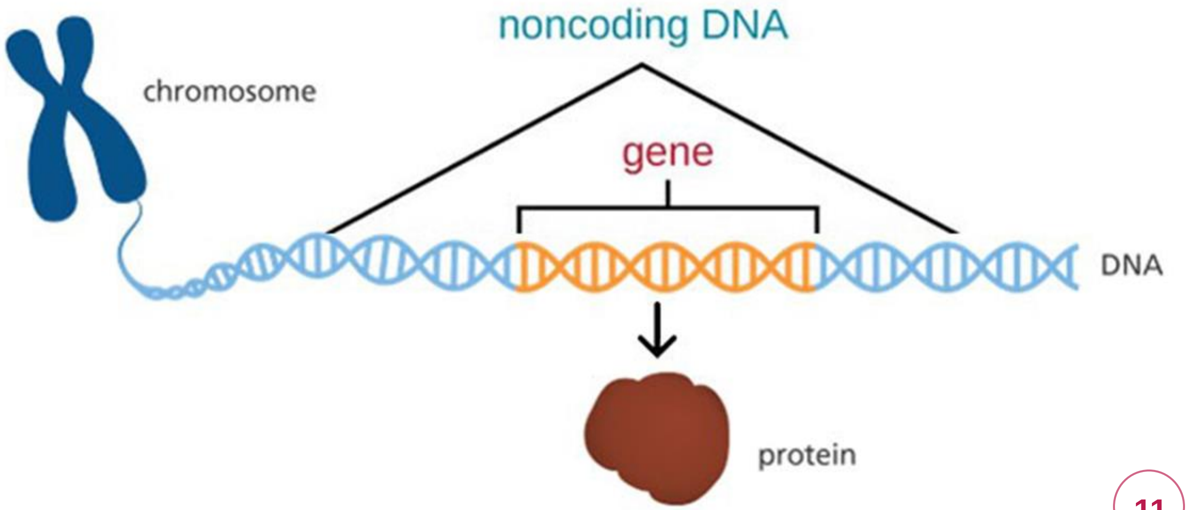
$\boldsymbol{\psi}$: the standard genetic relationship matrix

σ_e^2 : residual variance

The motivation of TWAS



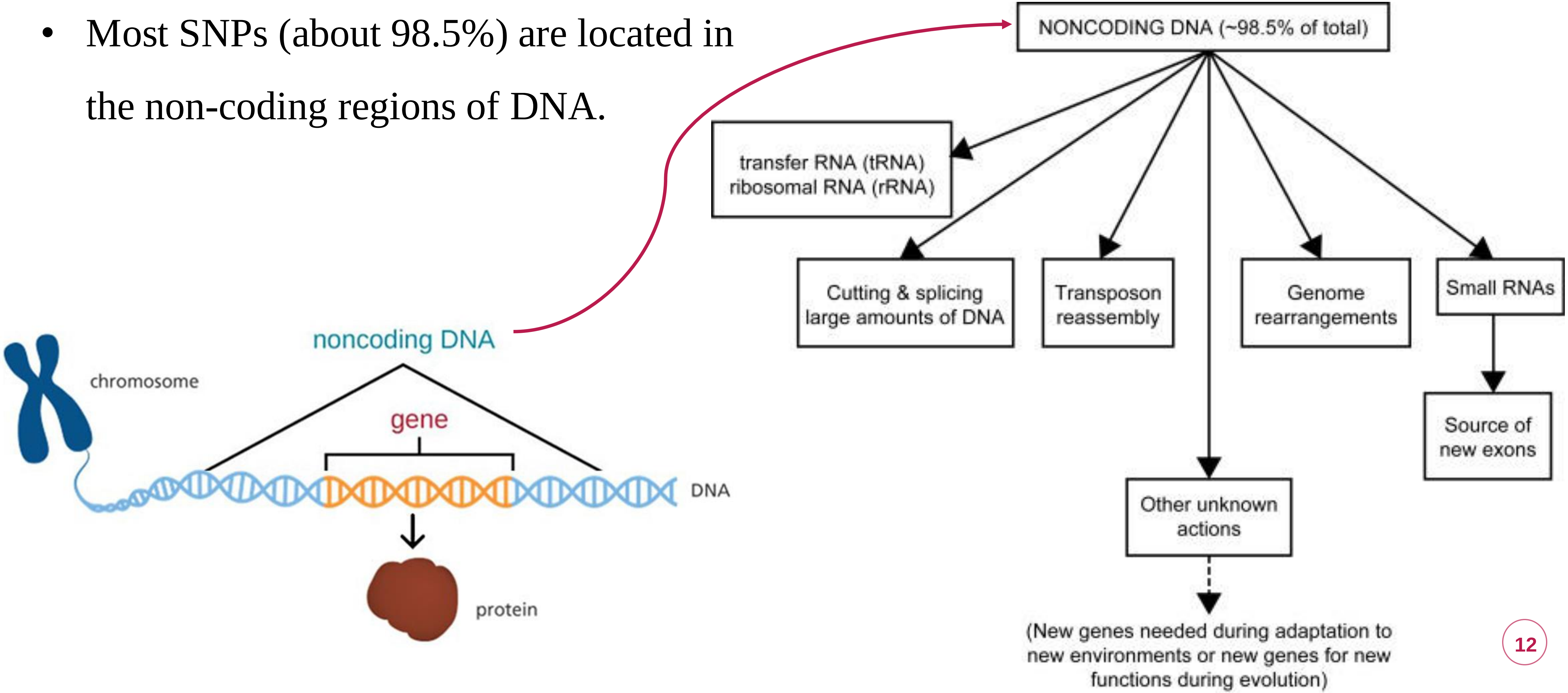
- SNPs in different tissues have different regulatory effects on gene expression.



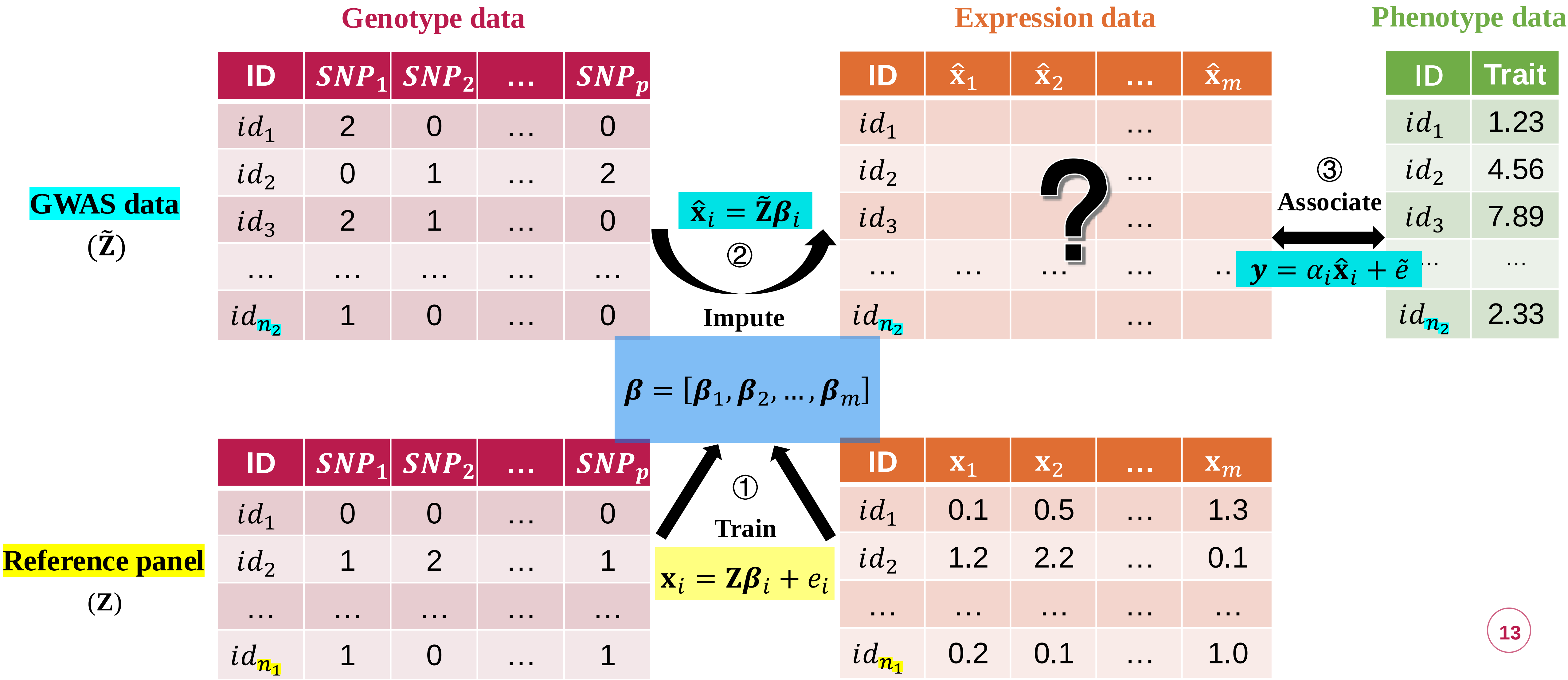
- Most SNPs (about 98.5%) are located in the non-coding regions of DNA.

The motivation of TWAS

- Most SNPs (about 98.5%) are located in the non-coding regions of DNA.


noncoding DNA
chromosome
gene
DNA
protein
NONCODING DNA (~98.5% of total)
transfer RNA (tRNA)
ribosomal RNA (rRNA)
Cutting & splicing large amounts of DNA
Transposon reassembly
Genome rearrangements
Small RNAs
Source of new exons
Other unknown actions
(New genes needed during adaptation to new environments or new genes for new functions during evolution)

The motivation of TWAS



GWAS data

($\tilde{\mathbf{Z}}$)

Genotype data

ID	SNP_1	SNP_2	…	SNP_p
id_1	2	0	…	0
id_2	0	1	…	2
id_3	2	1	…	0
…	…	…	…	…
id_{n_2}	1	0	…	0

$\hat{\mathbf{x}}_i = \tilde{\mathbf{Z}}\boldsymbol{\beta}_i$

②

Impute

Expression data

ID	$\hat{\mathbf{x}}_1$	$\hat{\mathbf{x}}_2$	…	$\hat{\mathbf{x}}_m$
id_1			…	
id_2		?	…	
id_3			…	
…	…	…	…	…
id_{n_2}			…	

③

Associate

$y = \alpha_i\hat{\mathbf{x}}_i + \tilde{e}$

Phenotype data

ID	Trait
id_1	1.23
id_2	4.56
id_3	7.89
…	…
id_{n_2}	2.33

$\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_m]$

①

Train

$\mathbf{x}_i = \mathbf{Z}\boldsymbol{\beta}_i + e_i$

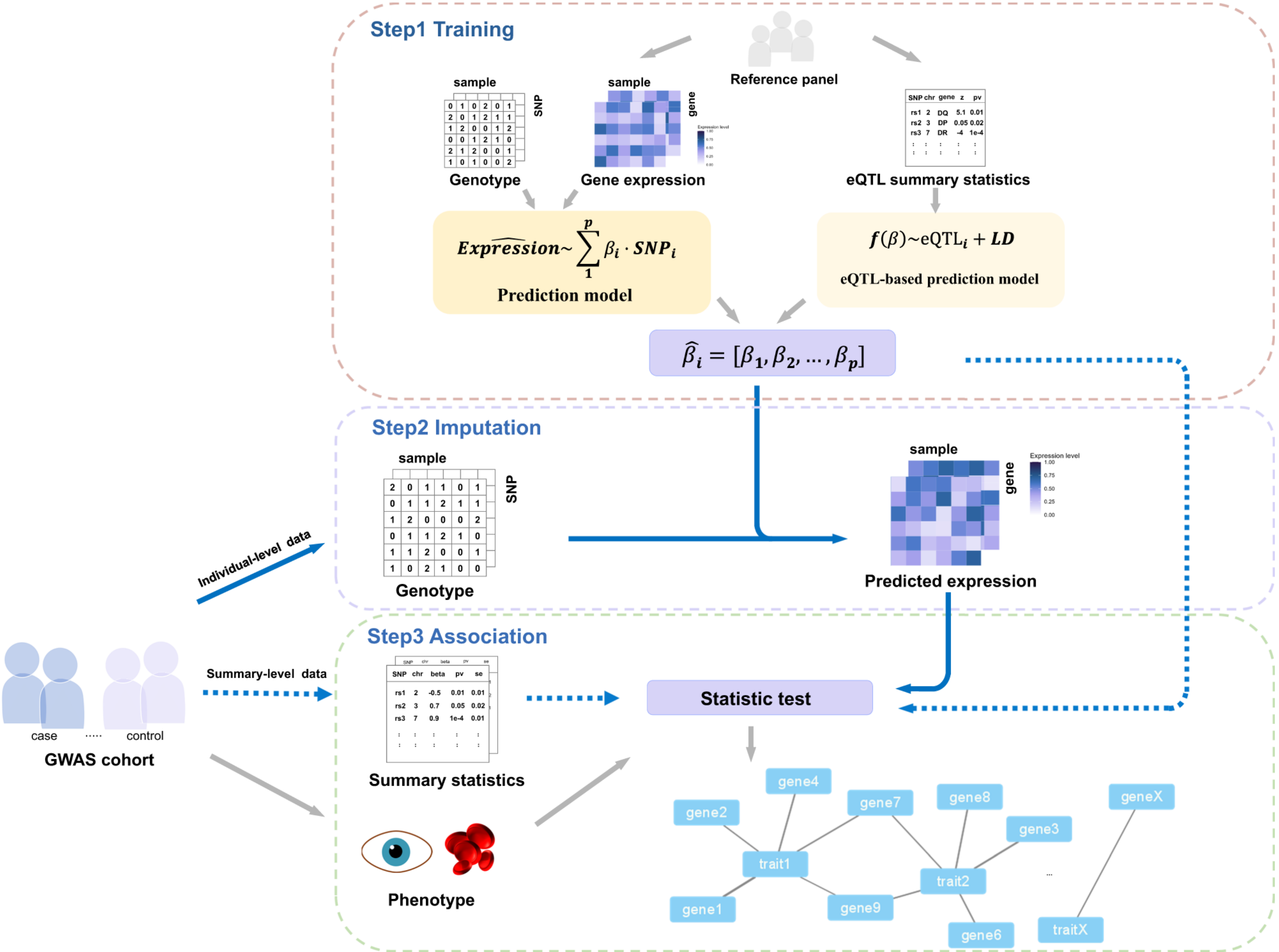
Reference panel

($\mathbf{Z}$)

ID	SNP_1	SNP_2	…	SNP_p
id_1	0	0	…	0
id_2	1	2	…	1
…	…	…	…	…
id_{n_1}	1	0	…	1

ID	$\mathbf{x}_1$	$\mathbf{x}_2$	…	$\mathbf{x}_m$
id_1	0.1	0.5	…	1.3
id_2	1.2	2.2	…	0.1
…	…	…	…	…
id_{n_1}	0.2	0.1	…	1.0

The motivation of TWAS



$$\widehat{Expression} \sim \sum_{1}^{p} \beta_i \cdot SNP_i$$

$$f(\beta) \sim \text{eQTL}_i + LD$$

$$\widehat{\beta_i} = [\beta_1, \beta_2, \dots, \beta_p]$$

0	1	0	2	0	1
2	0	1	2	1	1
1	2	0	0	1	2
0	0	1	2	1	0
2	1	2	0	0	1
1	0	1	0	0	2

SNP	chr	gene	z	pv
rs1	2	DQ	5.1	0.01
rs2	3	DP	0.05	0.02
rs3	7	DR	-4	1e-4
:	:	:	:	:
:	:	:	:	:

2	0	1	1	0	1
0	1	1	2	1	1
1	2	0	0	0	2
0	1	1	2	1	0
1	1	2	0	0	1
1	0	2	1	0	0

SNP	chr	beta	pv	se
rs1	2	-0.5	0.01	0.01
rs2	3	0.7	0.05	0.02
rs3	7	0.9	1e-4	0.01
:	:	:	:	:
:	:	:	:	:

1. **Training stage**: Estimate regulatory effect sizes of multiple SNPs on the gene expression level from a small reference panel with genotype and expression data.

2. **Imputation stage**: Obtain the predicted gene expression of GWAS individuals.

3. **Association stage**: Implement hypothesis tests between predicted gene expression and the target trait

02

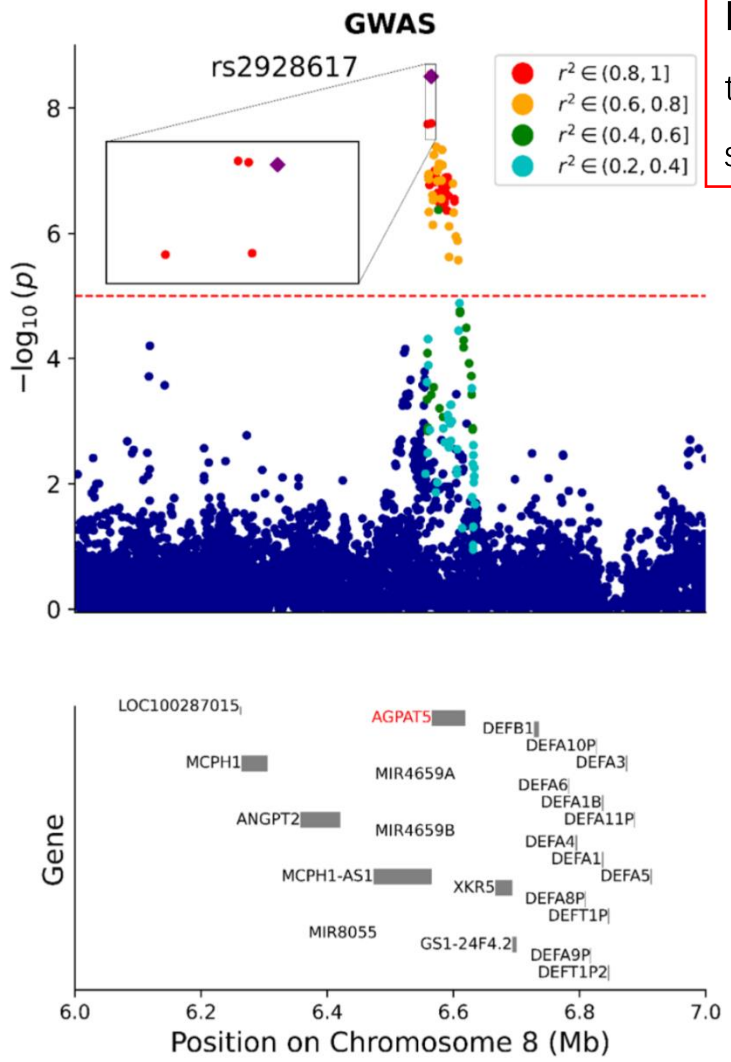
Challenges

Outline

☑ LD and expression correlation lead to confounding

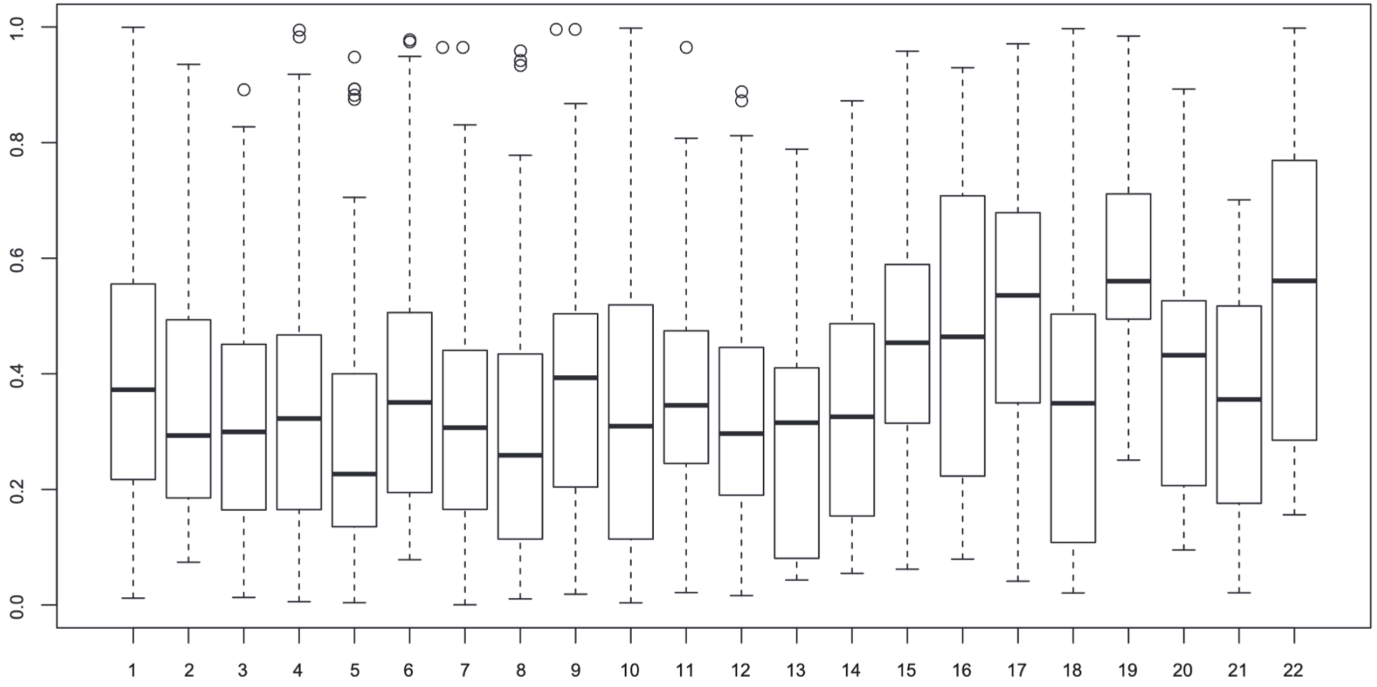
☑ Two-Step Inference Procedure lead to power loss

LD and expression correlation lead to confounding



Linkage disequilibrium (LD) : the non-independent segregation of genetic variants.

Boxplot displays the maximum of the absolute value of **expression correlation** estimates in each region across the 22 chromosomes

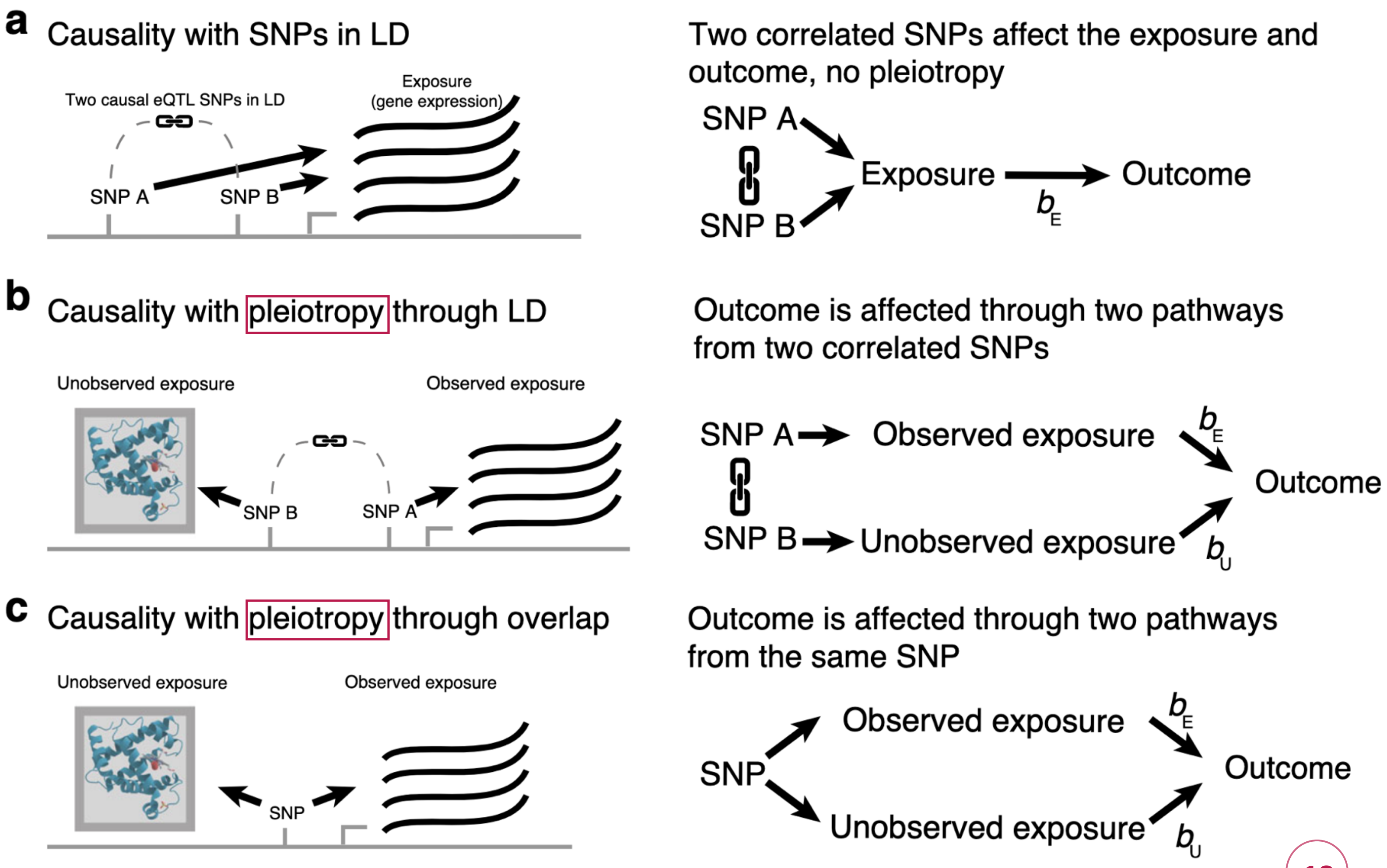


■ LD and expression correlation lead to confounding

There are no unobserved exposures.

Such situations rarely occur.

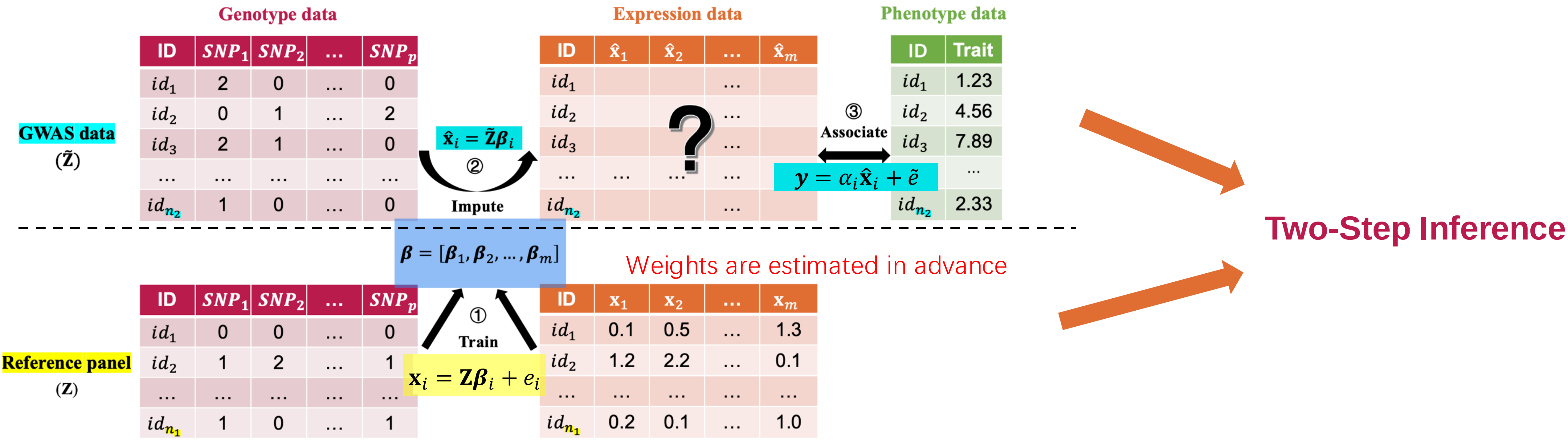
Confounding


a
Causality with SNPs in LD
Two causal eQTL SNPs in LD
Exposure
(gene expression)
SNP A
SNP B
Two correlated SNPs affect the exposure and outcome, no pleiotropy
SNP A
SNP B
Exposure
Outcome
b_E
b
Causality with pleiotropy through LD
Unobserved exposure
Observed exposure
SNP B
SNP A
Outcome is affected through two pathways from two correlated SNPs
SNP A
Observed exposure
b_E
Outcome
SNP B
Unobserved exposure
b_U
c
Causality with pleiotropy through overlap
Unobserved exposure
Observed exposure
SNP
Outcome is affected through two pathways from the same SNP
Observed exposure
b_E
SNP
Outcome
Unobserved exposure
b_U

Pleiotropy:

A genetic variant affects the outcome through a pathway that does not involve the risk factor of interest.

Two-Step Inference Procedure lead to power loss



Genotype data

Expression data

Phenotype data

GWAS data
($\tilde{\mathbf{Z}}$)

ID	SNP_1	SNP_2	…	SNP_p
id_1	2	0	…	0
id_2	0	1	…	2
id_3	2	1	…	0
…	…	…	…	…
id_{n_2}	1	0	…	0

$\hat{\mathbf{x}}_i = \tilde{\mathbf{Z}}\boldsymbol{\beta}_i$

② Impute

ID	$\hat{\mathbf{x}}_1$	$\hat{\mathbf{x}}_2$	…	$\hat{\mathbf{x}}_m$
id_1			…	
id_2		?	…	
id_3			…	
…	…	…	…	
id_{n_2}			…	

③ Associate

$\boldsymbol{y} = \alpha_i \hat{\mathbf{x}}_i + \tilde{e}$

ID	Trait
id_1	1.23
id_2	4.56
id_3	7.89
	…
id_{n_2}	2.33

$\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_m]$

Weights are estimated in advance

Reference panel
($\mathbf{Z}$)

ID	SNP_1	SNP_2	…	SNP_p
id_1	0	0	…	0
id_2	1	2	…	1
…	…	…	…	…
id_{n_1}	1	0	…	1

① Train

$\mathbf{x}_i = \mathbf{Z}\boldsymbol{\beta}_i + e_i$

ID	$\mathbf{x}_1$	$\mathbf{x}_2$	…	$\mathbf{x}_m$
id_1	0.1	0.5	…	1.3
id_2	1.2	2.2	…	0.1
…	…	…	…	…
id_{n_1}	0.2	0.1	…	1.0

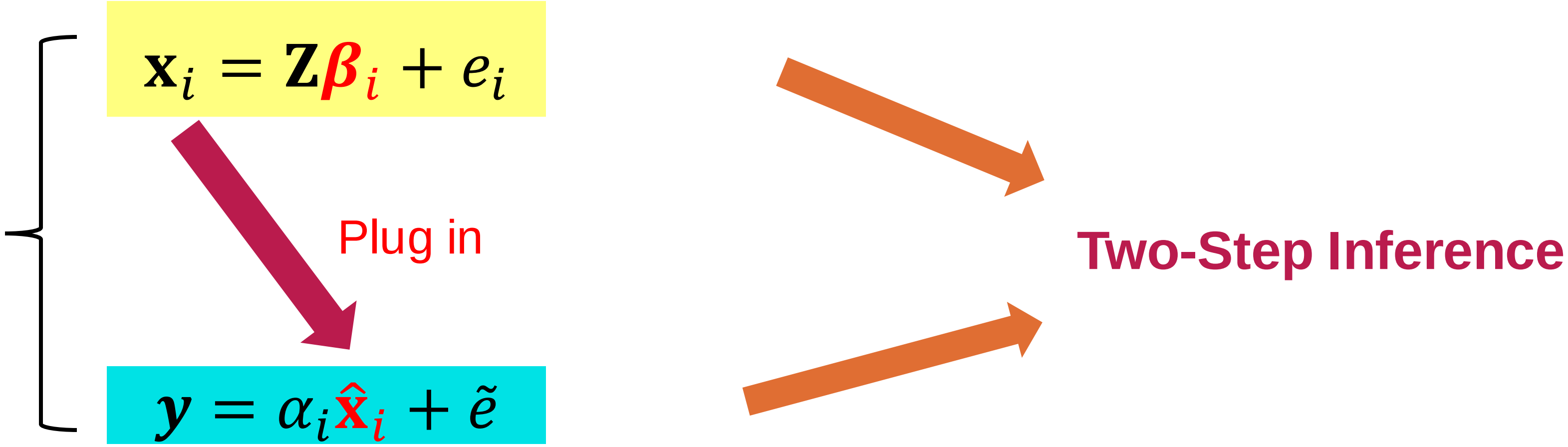
Two-Step Inference

Most of the existing marginal TWAS methods consist of two separate analytical steps.

The characteristic of these methods is that they estimate the weights $\boldsymbol{\beta}$ from the reference panel in advance. TWAS (or TWAS fine-mapping) is performed given weights $\boldsymbol{\beta}$.

Two-Step Inference Procedure lead to power loss

The point estimation of $\boldsymbol{\beta}_i$ has more uncertainty.



$$\mathbf{x}_i = \mathbf{Z}\boldsymbol{\beta}_i + e_i$$

Plug in

$$\boldsymbol{y} = \alpha_i \hat{\mathbf{x}}_i + \tilde{e}$$

Two-Step Inference

Most of the existing marginal TWAS methods consist of two separate analytical steps.

The characteristic of these methods is that they estimate the weights $\boldsymbol{\beta}$ from the reference panel in advance. TWAS (or TWAS fine-mapping) is performed given weights $\boldsymbol{\beta}$.

03

Methods

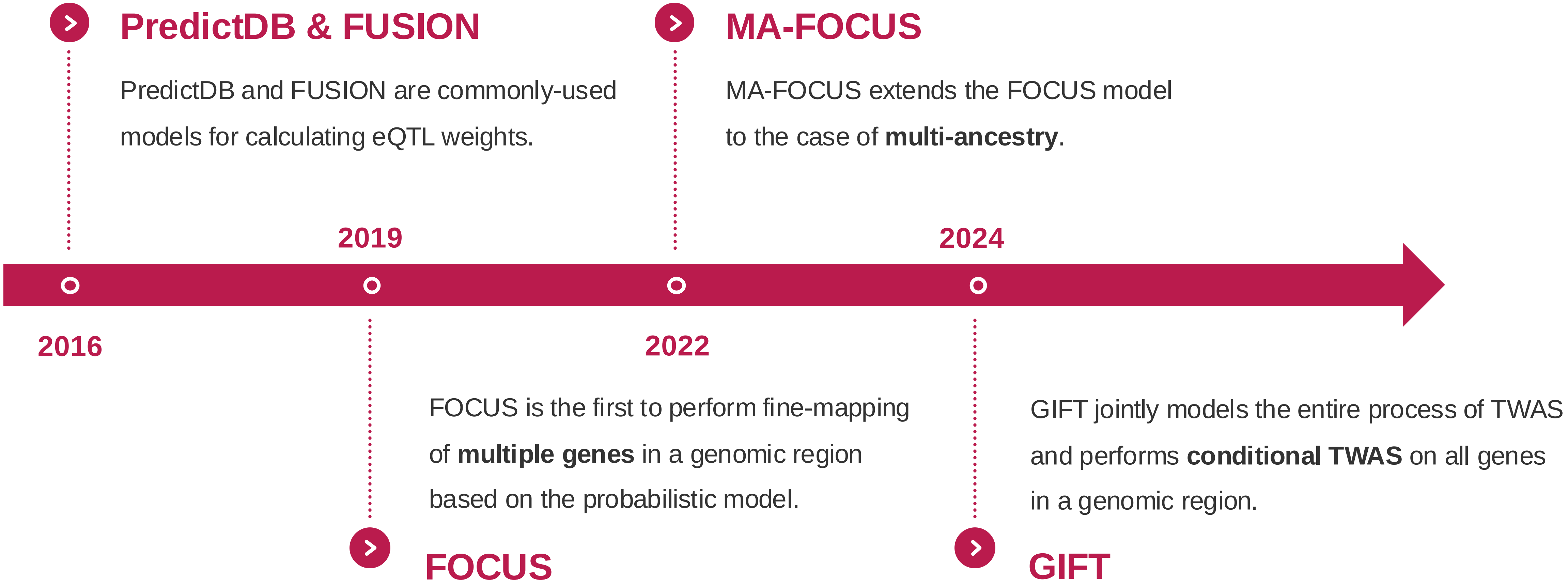
Outline

01 TWAS method timeline

02 The model of GIFT

03 The advantages of GIFT

TWAS method timeline



2016

PredictDB & FUSION

PredictDB and FUSION are commonly-used models for calculating eQTL weights.

2019

FOCUS

FOCUS is the first to perform fine-mapping of **multiple genes** in a genomic region based on the probabilistic model.

2022

MA-FOCUS

MA-FOCUS extends the FOCUS model to the case of **multi-ancestry**.

2024

GIFT

GIFT jointly models the entire process of TWAS and performs **conditional TWAS** on all genes in a genomic region.

The model of GIFT (individual-level)

GIFT (Gene-based Integrative Fine-mapping through conditional TWAS),

jointly models all *k* genes residing in the focal region and carries out TWAS conditional analysis:

$$\begin{cases} \mathbf{x}_i = Z_i \boldsymbol{\beta}_i + \mathbf{e}_i, i = 1, \cdots, k & (1) \\ \mathbf{y} = \sum_{i=1}^{k} \alpha_i (\tilde{Z}_i \boldsymbol{\beta}_i) + \tilde{\mathbf{e}} & (2) \end{cases}$$

$\mathbf{x}_i \in \mathbb{R}^{n_1 \times 1}$: expression vector for the *i*-th gene

$\mathbf{y} \in \mathbb{R}^{n_2 \times 1}$: phenotype vector

$\mathbf{Z}_i \in \mathbb{R}^{n_1 \times p_i}$: genotype matrix in the reference panel for the *i*-th gene

$\widetilde{\boldsymbol{Z}}_i \in \mathbb{R}^{n_2 \times p_i}$: genotype matrix in the GWAS data for the *i*-th gene

$\boldsymbol{\beta}_i \in \mathbb{R}^{p_i \times 1}$: eQTL random effects on the *i*-th gene expression

$\alpha_i \in \mathbb{R}$: effects of predicted expression for the *i*-th gene

Assumed that **y**, $\mathbf{x}_i$ and each column of $\mathbf{Z}_i$ and $\widetilde{\boldsymbol{Z}}_i$ have all been standardized to have a **mean of zero** and standard **deviation of 1**.

The model of GIFT (individual-level)

GIFT (Gene-based Integrative Fine-mapping through conditional TWAS),

jointly models all *k* genes residing in the focal region and carries out TWAS conditional analysis:

$$\begin{cases} \mathbf{x}_i = Z_i \boldsymbol{\beta}_i + \mathbf{e}_i, i = 1, \cdots, k & (1) \\ \mathbf{y} = \sum_{i=1}^{k} \alpha_i (\tilde{Z}_i \boldsymbol{\beta}_i) + \tilde{\mathbf{e}} & (2) \end{cases}$$

Due to $p_i > n_1$

$\boldsymbol{\beta}_i \sim N(\mathbf{0}, \sigma_{\beta_i}^2 \cdot \boldsymbol{I}_{p_i})$

Using the posterior distribution of $\boldsymbol{\beta}_i$ for eQTL effect instead of the point estimate $\widehat{\boldsymbol{\beta}}_i$

$\mathbf{x}_i \in \mathbb{R}^{n_1 \times 1}$: expression vector for the *i*-th gene

$\mathbf{y} \in \mathbb{R}^{n_2 \times 1}$: phenotype vector

$\mathbf{Z}_i \in \mathbb{R}^{n_1 \times p_i}$: genotype matrix in the reference panel for the *i*-th gene

$\tilde{\boldsymbol{Z}}_{\boldsymbol{i}} \in \mathbb{R}^{n_2 \times p_i}$: genotype matrix in the GWAS data for the *i*-th gene

$\boldsymbol{\beta}_i \in \mathbb{R}^{p_i \times 1}$: eQTL random effects on the *i*-th gene expression

$\alpha_i \in \mathbb{R}$: effects of predicted expression for the *i*-th gene

The model of GIFT (individual-level)

GIFT (Gene-based Integrative Fine-mapping through conditional TWAS),

jointly models all *k* genes residing in the focal region and carries out TWAS conditional analysis:

$$\begin{cases} \mathbf{x}_i = Z_i \boldsymbol{\beta}_i + \mathbf{e}_i, i = 1, \cdots, k & (1) \\ \mathbf{y} = \sum_{i=1}^{k} \alpha_i (\tilde{Z}_i \boldsymbol{\beta}_i) + \tilde{\mathbf{e}} & (2) \end{cases}$$

$\mathbf{e}_i \in \mathbb{R}^{n_1 \times 1}$: residual errors for the *i*-th gene,

where $\boxed{\left(\mathbf{e}_{l,1}, \mathbf{e}_{l,2}, \dots, \mathbf{e}_{l,k}\right)^T \sim N_k(0, \Omega)}$ for the same individual l

GIFT takes the correlation of gene expressions into account.

$\tilde{\mathbf{e}} \in \mathbb{R}^{n_2 \times 1}$: residual error with each element *i.i.d.* from the same normal distribution $N(0,\ \sigma_y^2)$

The model of GIFT (summary-level)

GIFT can also be extended to perform inference using summary statistics only.

The corresponding model for summary statistics are:

$$\begin{cases} \widehat{\boldsymbol{\beta}}^*_{\boldsymbol{x}_i} = \boldsymbol{\Sigma}_{1i}\boldsymbol{\beta}_i + \boldsymbol{e}_{\boldsymbol{x}_i}, i = 1, \cdots, k \\ \widehat{\boldsymbol{\beta}}^*_{\boldsymbol{y}} = \boldsymbol{\Sigma}_2\left(\alpha_1\boldsymbol{\beta}_1^T, \cdots, \alpha_k\boldsymbol{\beta}_k^T\right)^T + \boldsymbol{e}_{\boldsymbol{y}} \end{cases}$$

$\widehat{\boldsymbol{\beta}}^*_{\boldsymbol{x}_i} \in \mathbb{R}^{1\times p_i}$: the estimates for the marginal SNP effects on the *i*-th gene expression

$\widehat{\boldsymbol{\beta}}^*_{\boldsymbol{y}} \in \mathbb{R}^{1\times p}$: the estimates for the marginal SNP effects on the trait

$\boldsymbol{\Sigma}_{1i} \in \mathbb{R}^{p_i\times p_i}$: correlation matrix of all cis-SNPs for the *i*-th gene in the reference panel

$\boldsymbol{\Sigma}_2 \in \mathbb{R}^{p\times p}$: correlation matrix of all cis-SNPs for all the genes in the focal region in the GWAS data.

$\boldsymbol{\beta}_i \in \mathbb{R}^{p_i\times 1}$: eQTL effects on the *i*-th gene experssion

$\alpha_i \in \mathbb{R}$: effects of predicted expression for the *i*-th gene

The advantages of GIFT

01

Conditional TWAS analysis

GIFT performs TWAS fine-mapping conditional on the effects of the other genes to avoid confounding.

02

Joint likelihood inference framework

The joint inference framework accounts for the uncertainty in the SNP effect-size estimates on gene expression and the uncertainty in the predicted expression.

03

PX-EM algorithm

GIFT introduces the auxiliary parameter λ through the parameter expansion method to significantly improve the convergence speed.

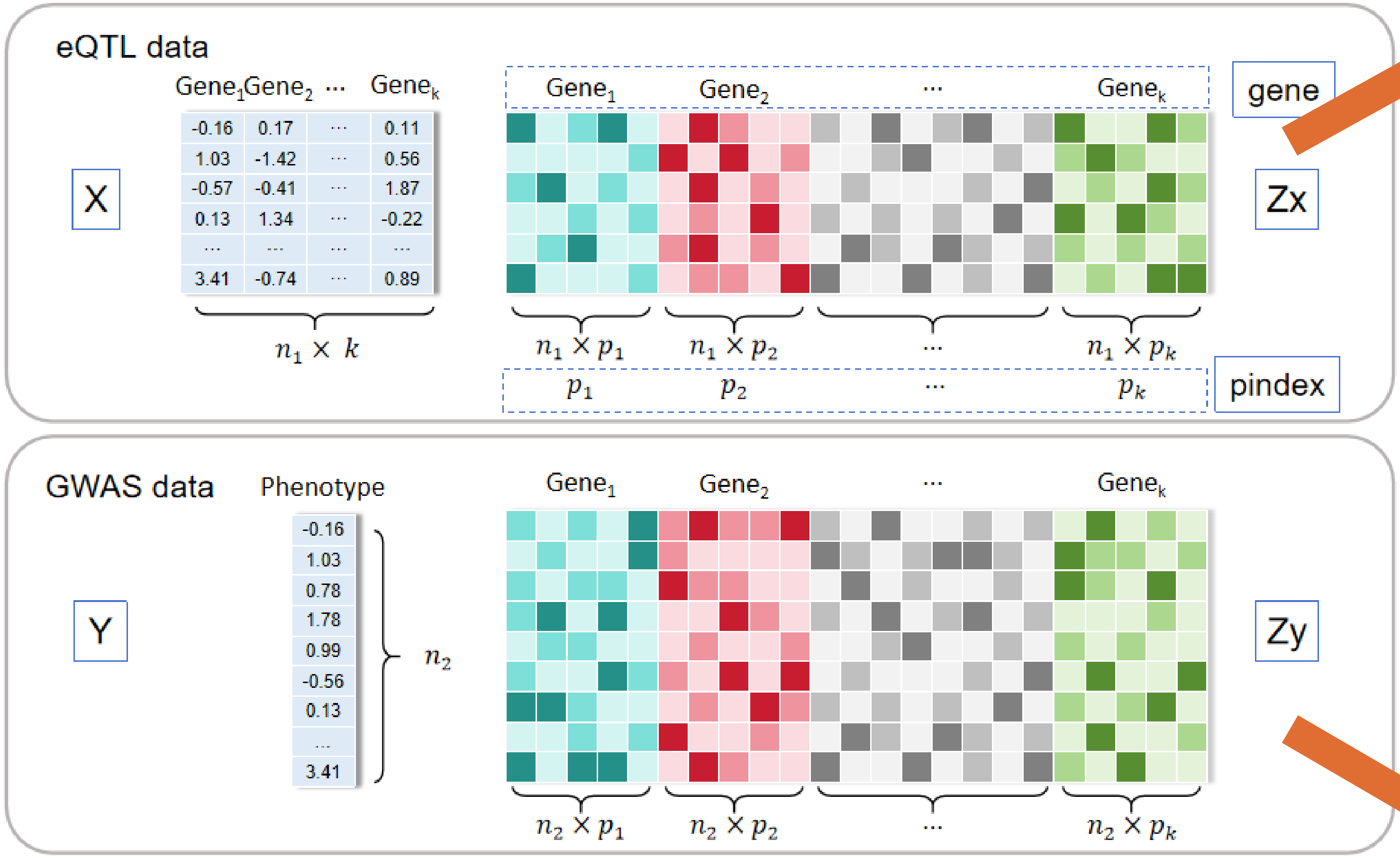
04

Results

Outline

- ☑ Data input
- ☑ GIFT produces calibrated P values under the null simulations
- ☑ GIFT is powerful under a range of alternative simulations
- ☑ Real-data applications

Data input

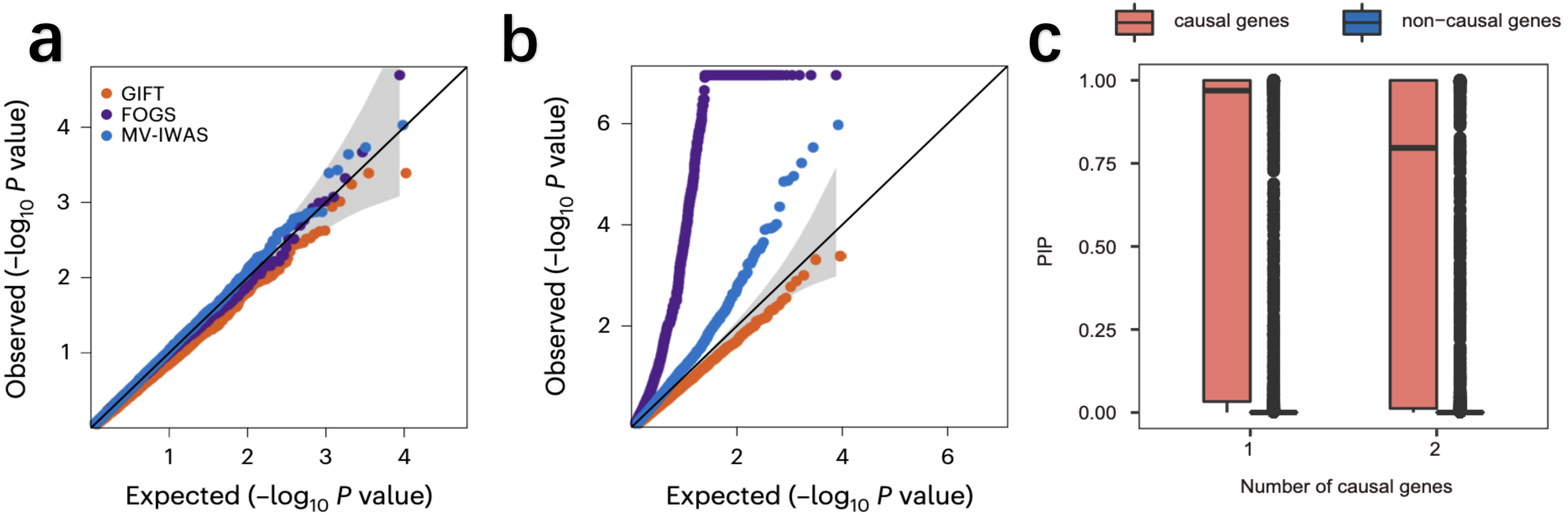


Data from GEUVADIS: n_1=465

	Median	Mean	Min	Max
Genes per region (k)	8	11.75	1	100
Cis-SNPs per gene (p_i)	199	229	1	6879
Cis-SNPs per region (Σp_i)	2,242	2751	2	16,655

Data from UKBiobank: n_2=487,298

GIFT produces calibrated P values under the null simulations



All genes in the region are null

One gene in the region is casual

a, b: Quantile–quantile plots of −log10(P values) from the three frequentist methods,

which are both displayed for the **non-causal genes**.

c: Boxplot from FOCUS displays the PIPs from causal genes and non-causal genes.

*All of the simulation results are from 1,000 random region.

GIFT is powerful under a range of alternative simulations

When there is one causal gene in the region and it explains 1% of phenotypic variance, the FDR and power under the recommended thresholds from different methods as follows:

The 3 frequentist methods (GIFT, FOGS and MV-IWAS) are based on **Bonferroni's adjusted P-value threshold (0.05/m)** and the Bayesian method (FOCUS) is based on 90% credible sets.

Methods	FDR	power
GIFT	↓ 0%	46.8%
FOCUS	42.1%	↑ 70.6%
FOGS	39.6%	49.6%
MV-IWAS	0.5%	56.2%

Bonferroni correction:

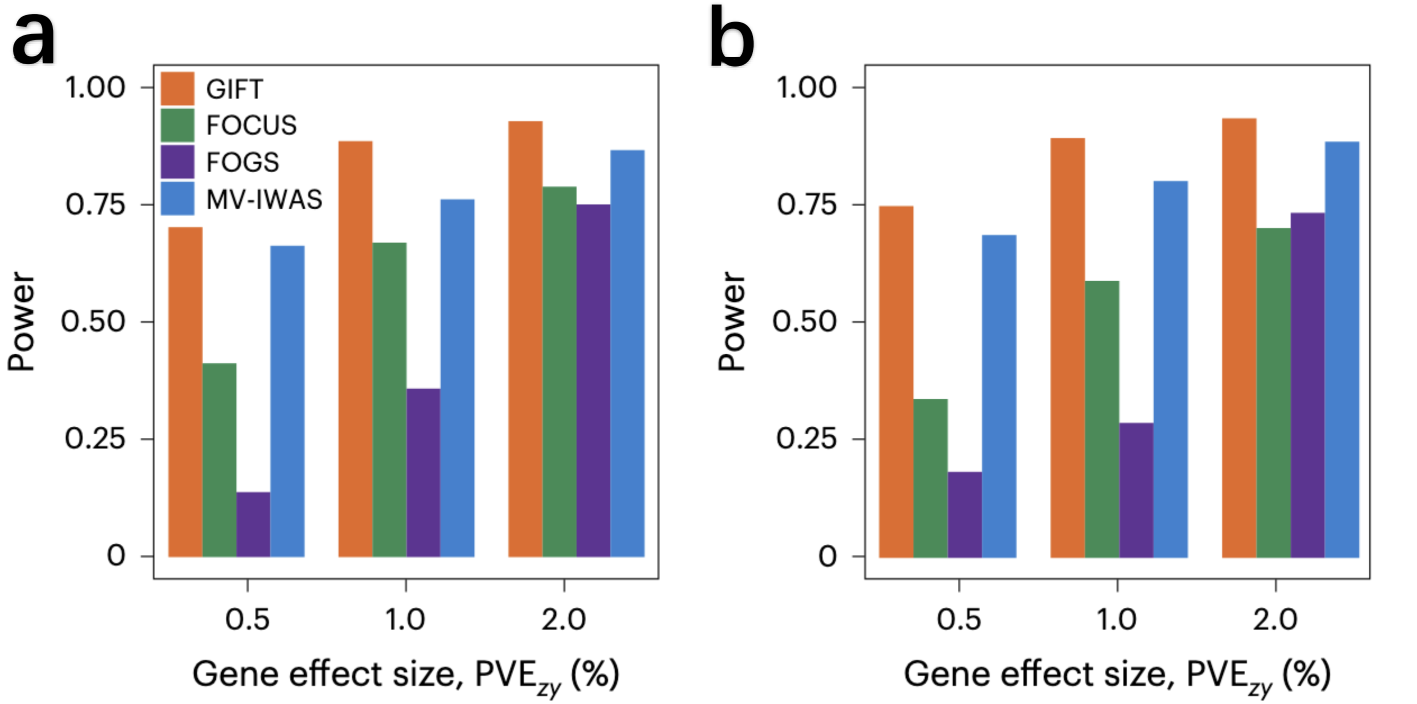
The most common way to control the familywise error rate. You will find the critical value (alpha) for an individual test by dividing the familywise error rate (usually 0.05) by the number of tests.

*All of the simulation results are from 1,000 random region.

GIFT is powerful under a range of alternative simulations

As the threshold for GIFT corresponds to a much lower FDR than the other three methods, such a threshold naturally leads to a lower power for GIFT.

To allow for fair, we further computed power **based on a true FDR of 0.05**:



One gene in the region is casual Two genes in the region are casual

a, b: Power comparisons for different methods based on a **true FDR** of 0.05.

*All of the simulation results are from 1,000 random region.

GIFT is powerful under a range of alternative simulations

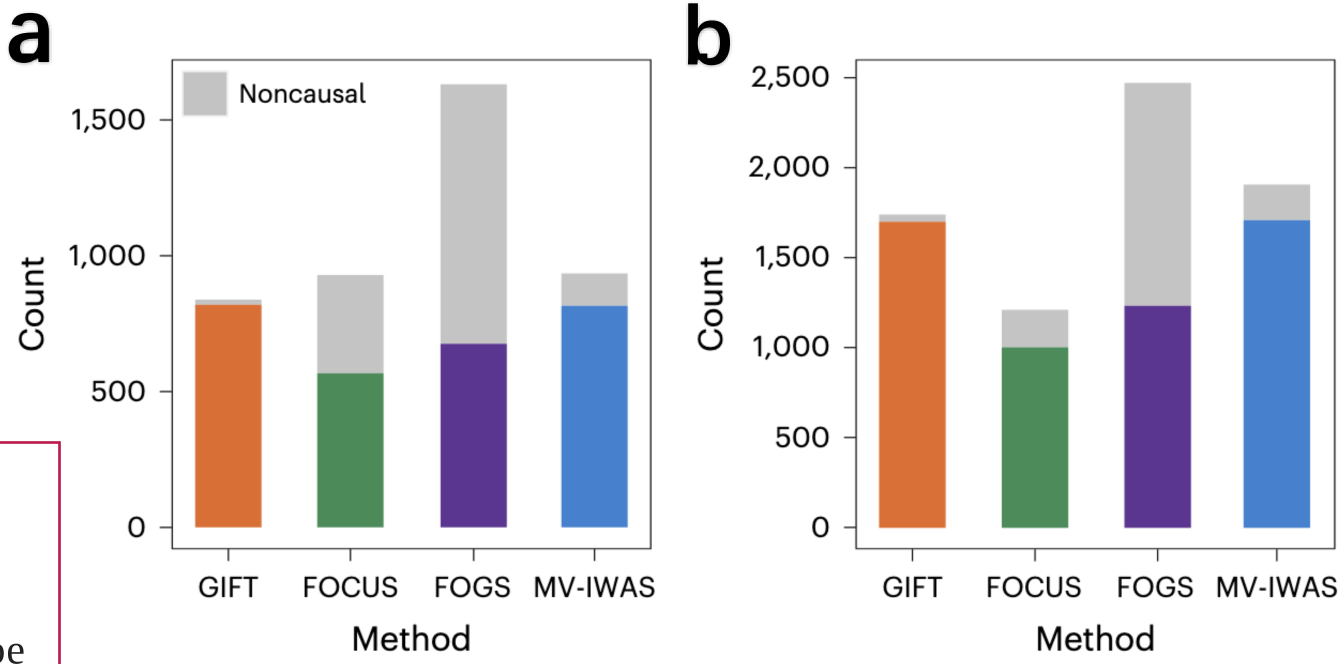
As the true FDR is known only in simulations but unknown for any real dataset, we also used P-value to compared power **based on the estimated FDR of 0.05**:

Benjamini–Hochberg method:

Order the m hypothesis by ascending p-values, where P_i is the p-value at the *i*-th position with the associated hypothesis H_i . Let *k* be the largest *i* for which:

$$P_i = \frac{i}{m} q$$

Reject hypotheses *i* = 1, 2, 3,..., *k*. The Benjamini–Hochberg method has been proven to control the FDR for all tests at a level of q



One gene in the region is casual

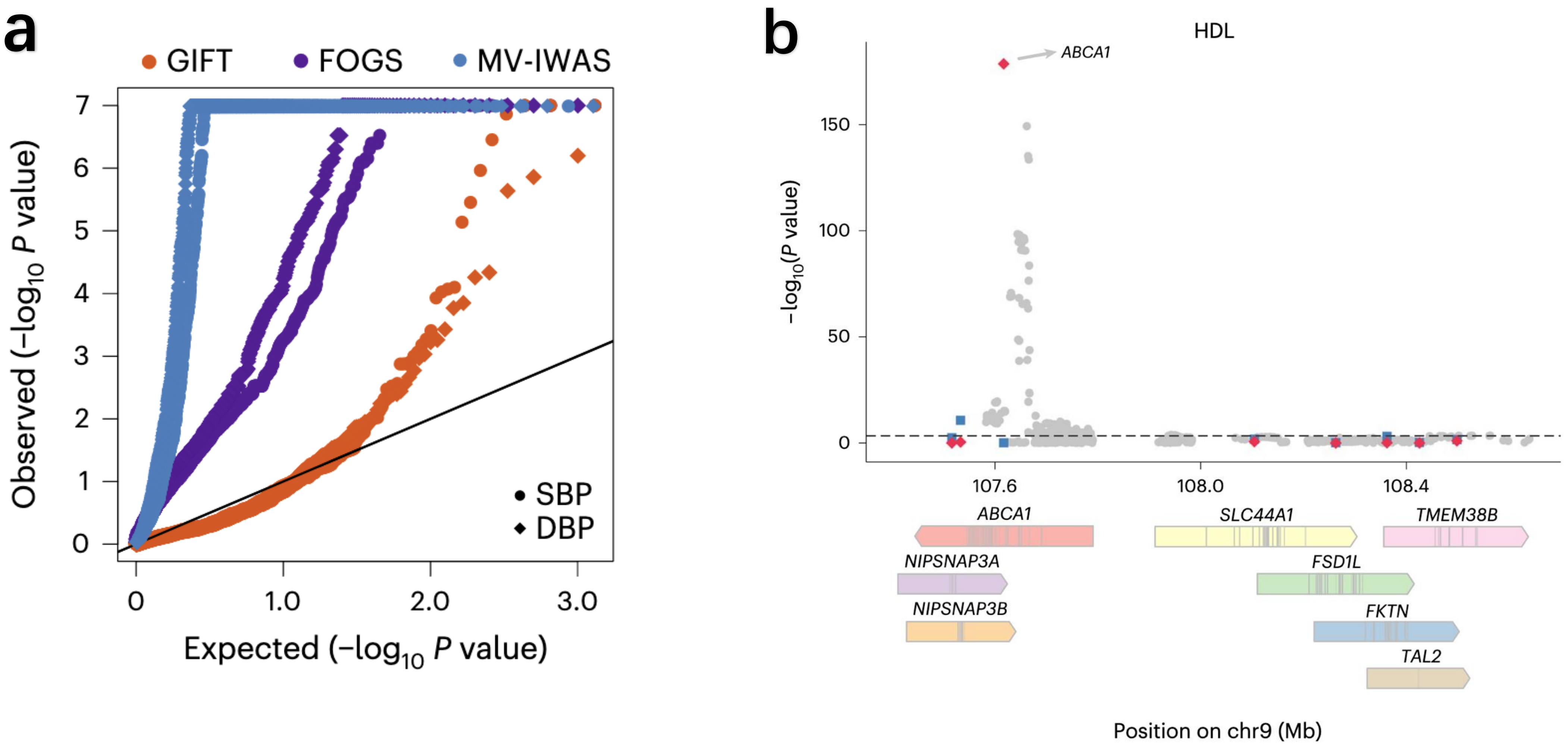
Two genes in the region are casual

a, b: Number of genes identified by different methods based on a **estimated FDR** of q=0.05.

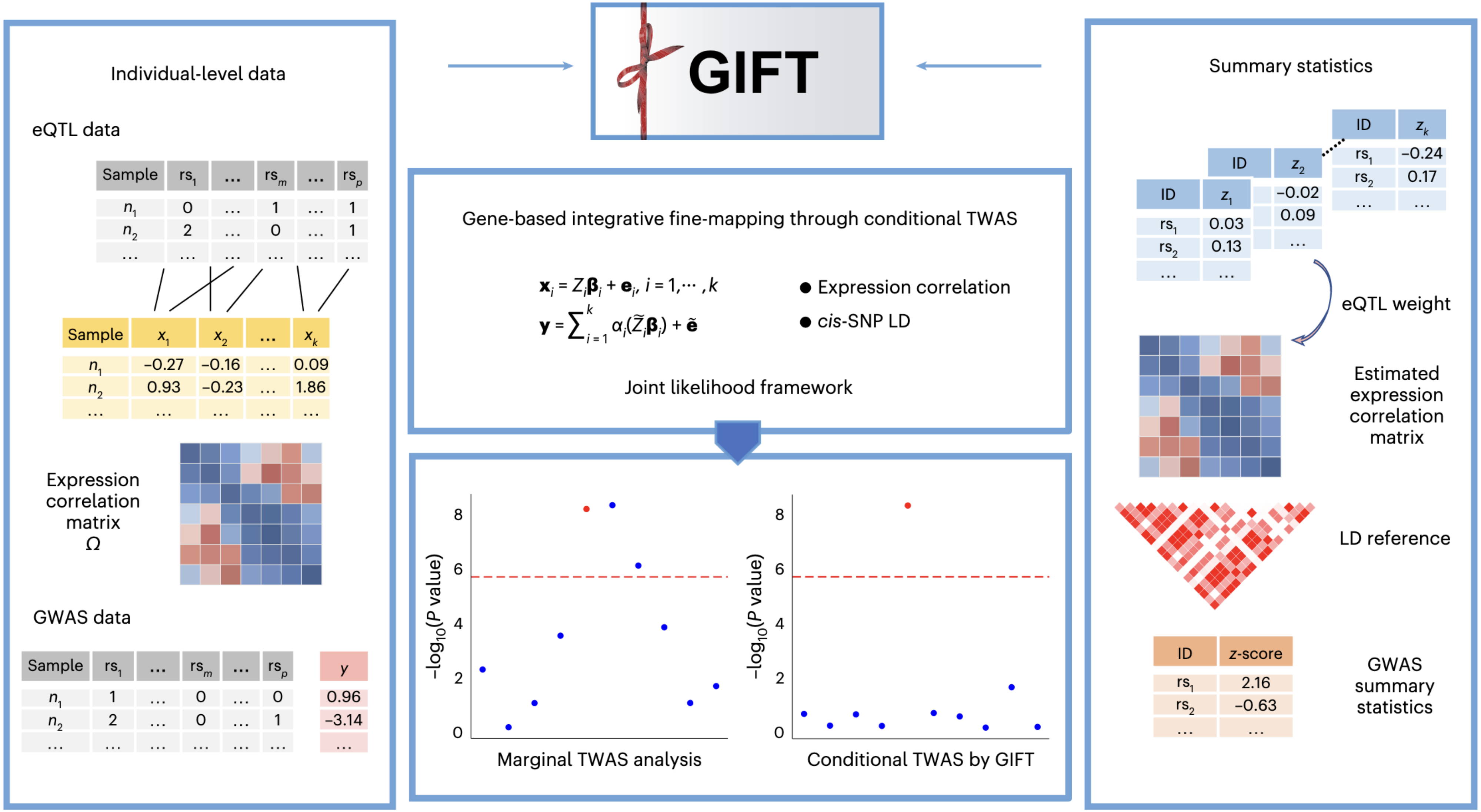
Colors represent the number of detected causal genes.

*All of the simulation results are from 1,000 random region.

Real-data applications



Discussion



GIFT

Individual-level data

eQTL data

Sample	rs_1	…	rs_m	…	rs_p
n_1	0	…	1	…	1
n_2	2	…	0	…	1
…	…	…	…	…	…

Sample	x_1	x_2	…	x_k
n_1	−0.27	−0.16	…	0.09
n_2	0.93	−0.23	…	1.86
…	…	…	…	…

Expression correlation matrix Ω

GWAS data

Sample	rs_1	…	rs_m	…	rs_p	y
n_1	1	…	0	…	0	0.96
n_2	2	…	0	…	1	−3.14
…	…	…	…	…	…	…

Gene-based integrative fine-mapping through conditional TWAS

$$\mathbf{x}_i = Z_i\boldsymbol{\beta}_i + \mathbf{e}_i,\ i = 1,\cdots,k$$

$$\mathbf{y} = \sum_{i=1}^{k}\alpha_i(\tilde{Z}_i\boldsymbol{\beta}_i) + \tilde{\mathbf{e}}$$

- Expression correlation
- *cis*-SNP LD

Joint likelihood framework

Summary statistics

ID	z_1
rs_1	0.03
rs_2	0.13
…	…

ID	z_2
	−0.02
	0.09
	…

ID	z_k
rs_1	−0.24
rs_2	0.17
…	…

eQTL weight

Estimated expression correlation matrix

LD reference

ID	*z*-score
rs_1	2.16
rs_2	−0.63
…	…

GWAS summary statistics

香港城市大學
City University of Hong Kong

THANK YOU

楊建文學術樓
Yeung Kin Man Academic Building
CityU
香港城市大學
City University of Hong Kong
楊建文學術樓
Yeung Kin Man Academic Building

Speaker: Yuekai Li

Major: Biostatistics