Fine-mapping causal tissues and genes at disease-associated loci

Benjamin Strober

Price lab, Harvard School of Public Health

ASHG 2023 annual meeting

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

@bennystrobes

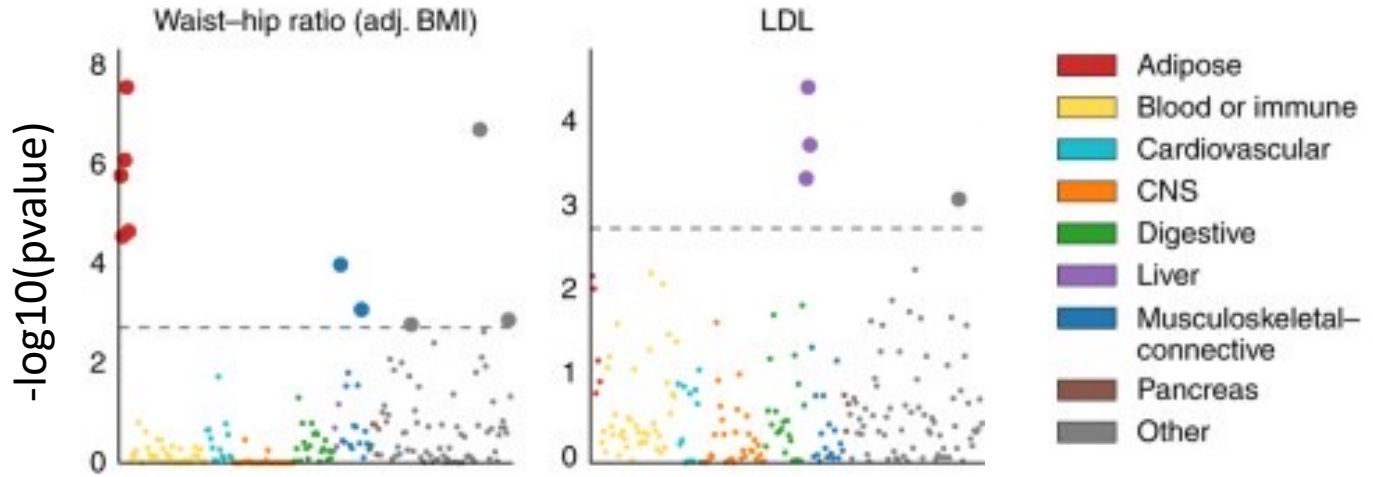
Outline

1. Motivation for fine-mapping causal tissues and genes at disease loci
2. Methods
3. Simulations
4. Results: Tissue-gene fine-mapping for 45 UK Biobank diseases and complex traits with 38 GTEx tissues

Outline

1. Motivation for fine-mapping causal tissues and genes at disease loci
2. Methods
3. Simulations
4. Results: Tissue-gene fine-mapping for 45 UK Biobank diseases and complex traits with 38 GTEx tissues

Genetic predisposition for complex disease/trait acts through disease-relevant tissues/cell-types

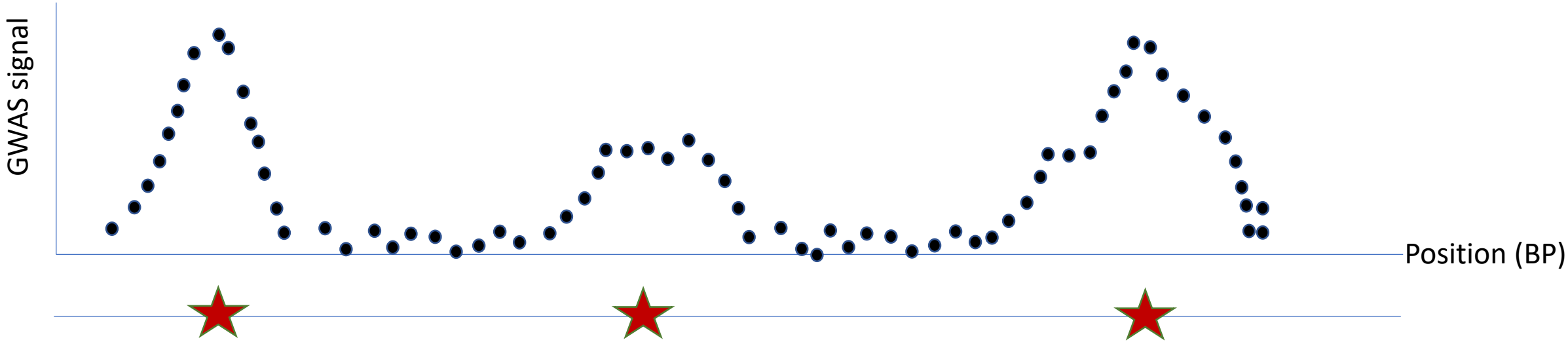


Evidence from genome-wide analyses:

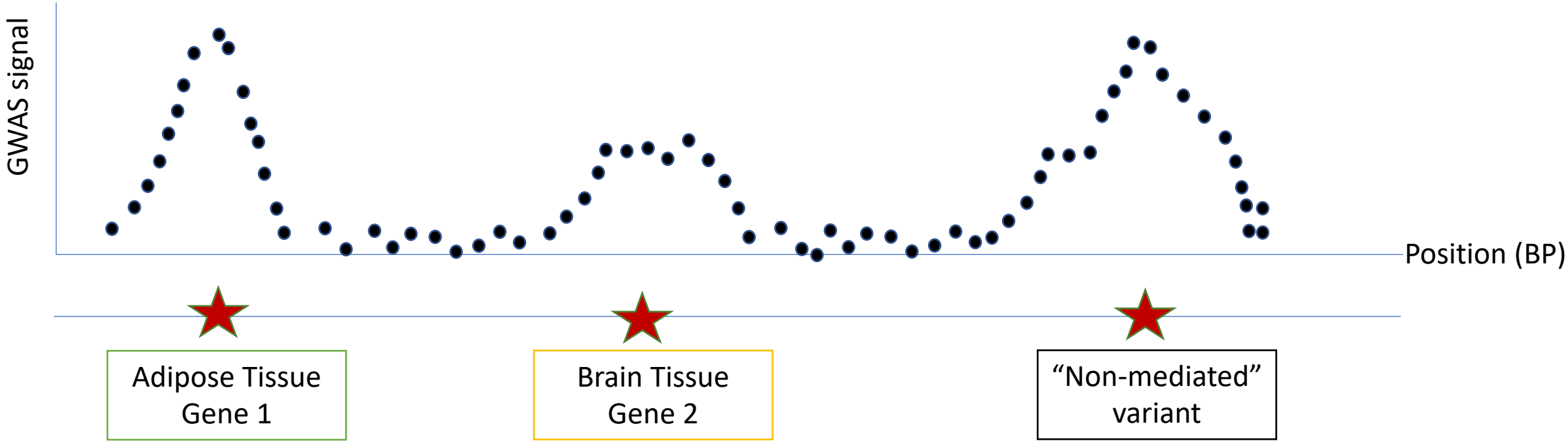
- **Finucane et al, Nature Genetics 2018:** Trait heritability enrichment around specifically expressed genes of specific GTEx tissues and cell types
- **Ongen et al, Nature Genetics, 2017:** Assess enrichment of tissue eQTLs colocalizing with GWAS signal
- **Amariuta et al, Nature Genetics 2023:** Trait heritability explained by genetically predicted tissue expression

Different loci from a single disease can be mediated by different tissues

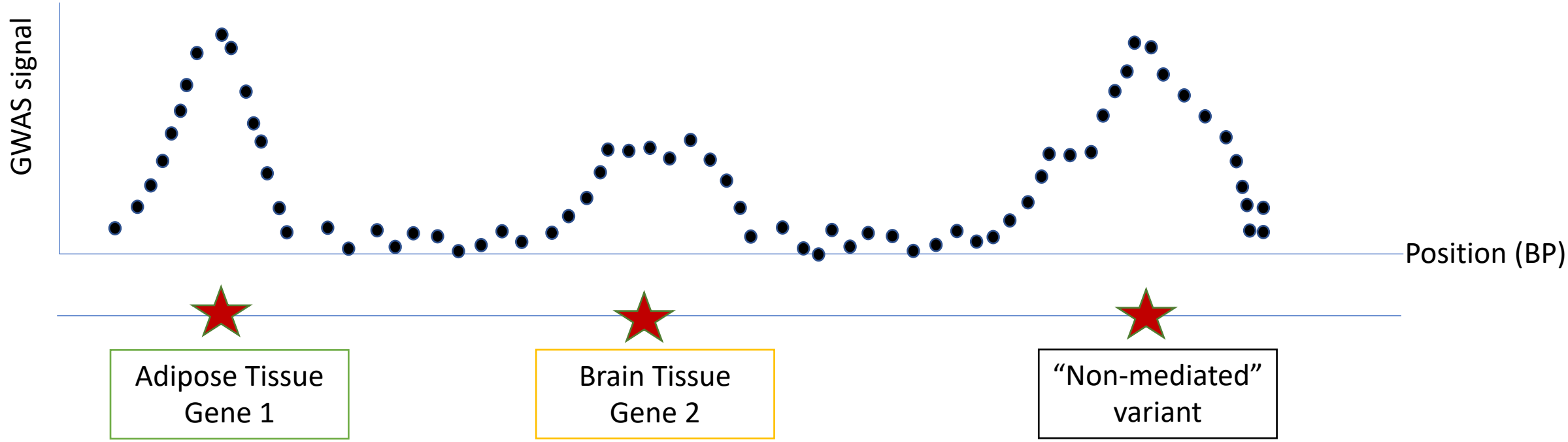
Project goal: Fine-map the causal tissue and gene at each disease signal



Project goal: Fine-map the causal tissue and gene at each disease signal



Project goal: Fine-map the causal tissue and gene at each disease signal

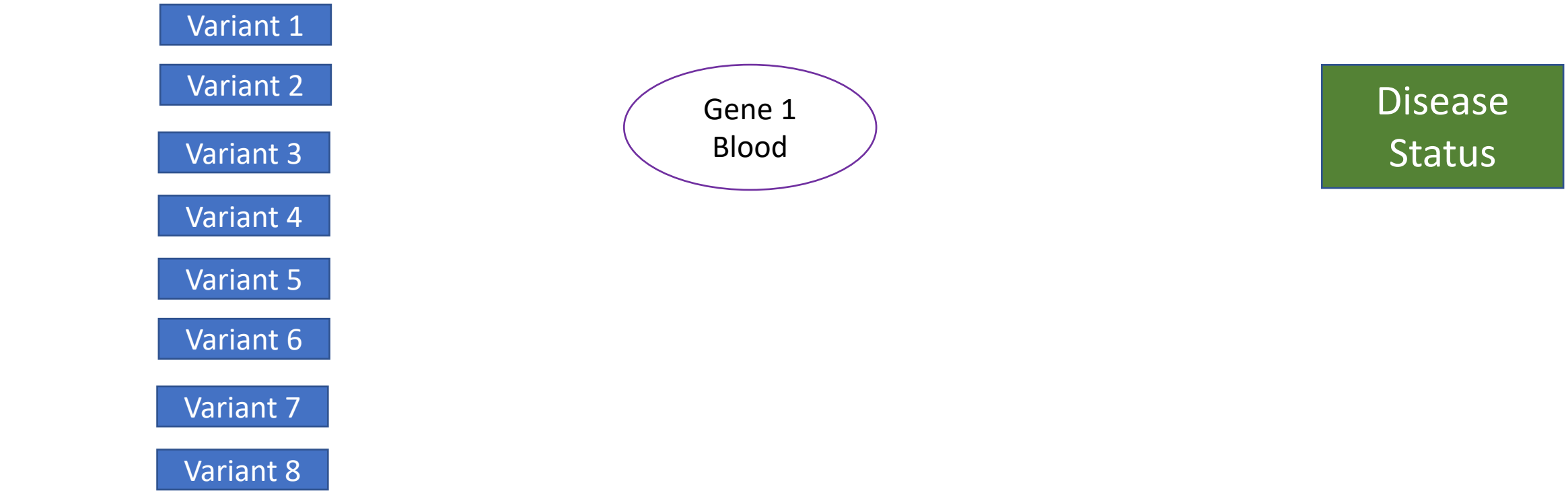


- Develop statistical model, TGFM (tissue-gene fine-mapping), to fine-map the causal tissue and gene at each genetic disease signal
  - Calculate causal probabilities for each (gene, tissue) pair and causal probabilities for each non-mediated variant
  - Based on integration of GWAS data and multi-tissue/cell type eQTL data

Outline

1. Motivation for fine-mapping causal tissues and genes at disease loci
2. Methods
3. Simulations
4. Results: Tissue-gene fine-mapping for 45 UK Biobank diseases and complex traits with 38 GTEx tissues

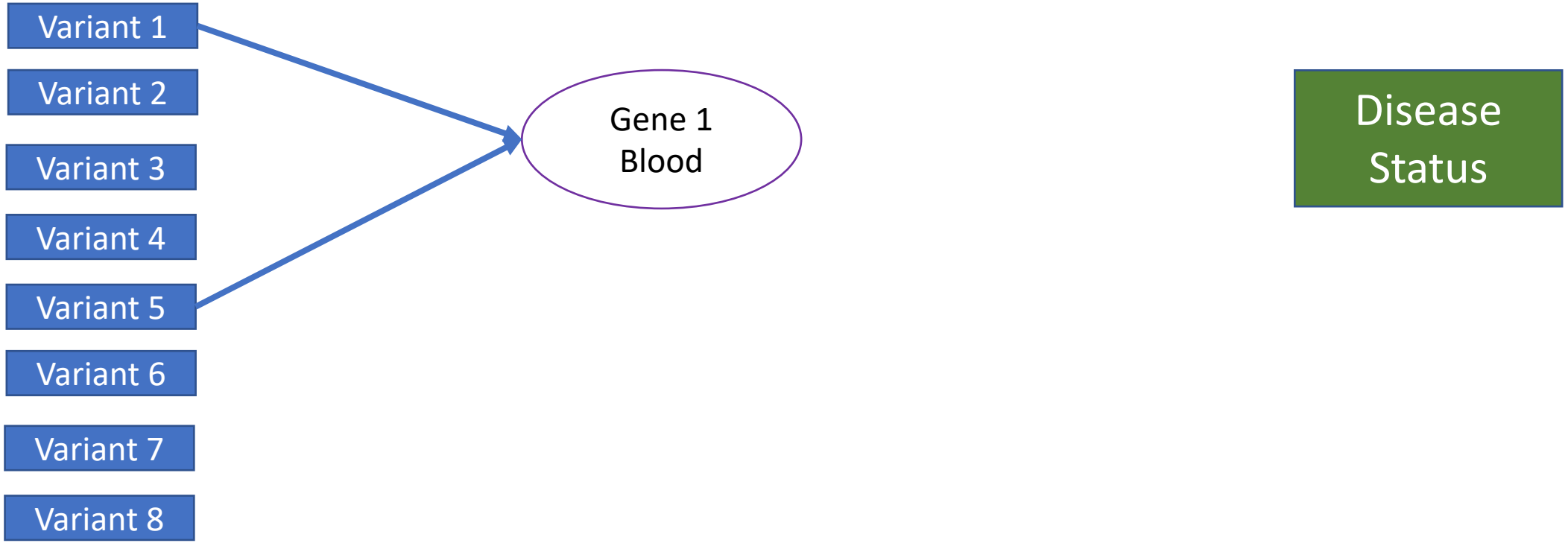
TWAS calculates marginal association between predicted genetic gene expression and disease



Gamazon et al. Nat. Genet. 2015
Gusev et al. Nat. Genet. 2016
Wainberg et al. Nat Genet. 2019

TWAS calculates marginal association between predicted genetic expression and disease

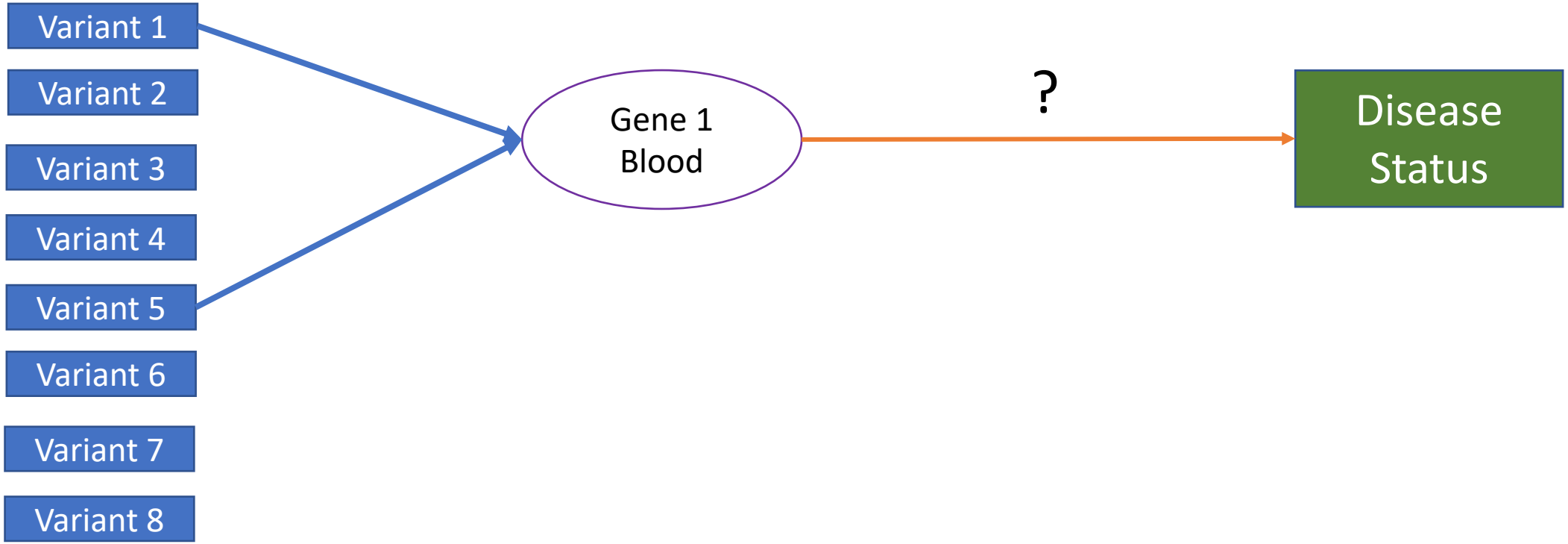
Step 1: Estimate predicted genetic gene expression using eQTL data



Gamazon et al. Nat. Genet. 2015
Gusev et al. Nat. Genet. 2016
Wainberg et al. Nat Genet. 2019

TWAS calculates marginal association between predicted genetic expression and disease

Step 2: Associate predicted genetic gene expression with disease



Gamazon et al. Nat. Genet. 2015
Gusev et al. Nat. Genet. 2016
Wainberg et al. Nat Genet. 2019

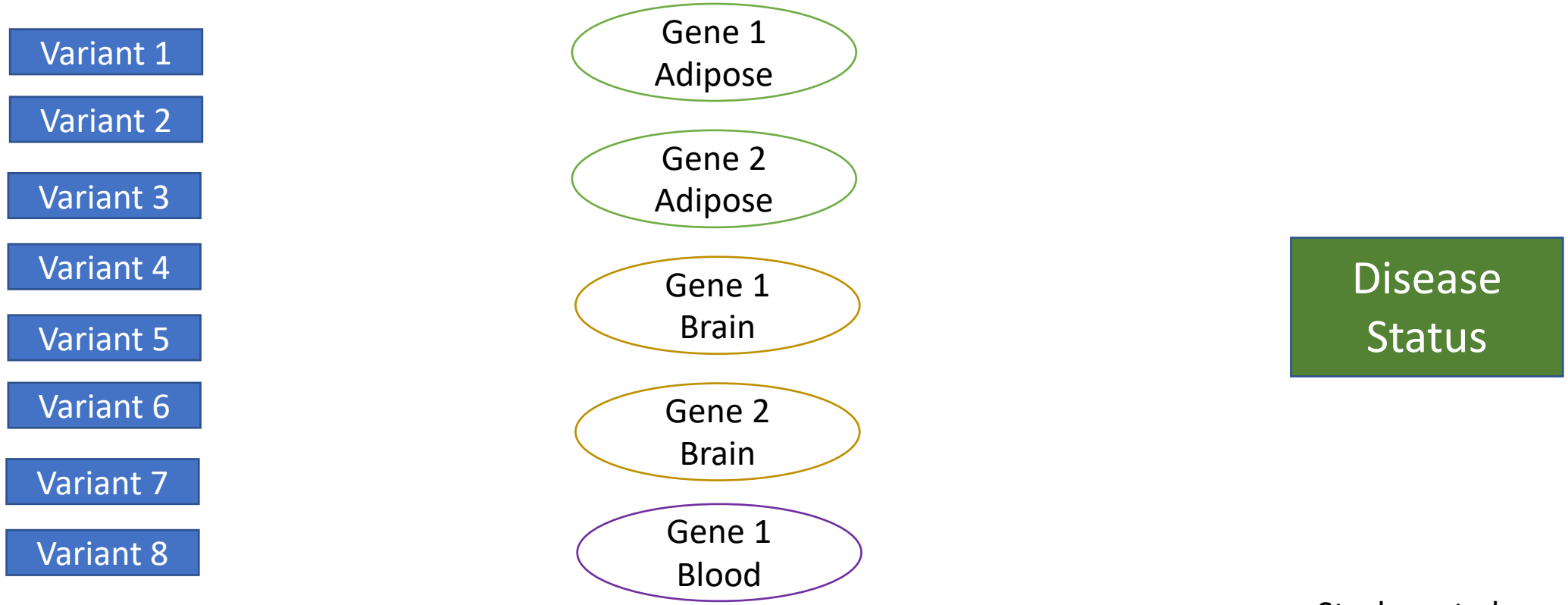
TWAS cannot distinguish causal from tagging gene-tissue pairs

- Analogous to variant LD obscuring causal from tagging variants using marginal GWAS associations
- A significant TWAS association for a gene-tissue pair can result from:

1. The gene-tissue pair having a causal effect on the disease
2. The gene-tissue pair does not have a causal effect on disease, but is correlated with another causal gene-tissue pair or causal non-mediated variant

Mancuso et al. Nat. Genet. 2019
Yao et al Nat. Genet 2020
Amariuta et al. Nat. Genet. 2023
Zhao et al. BioRxiv 2022

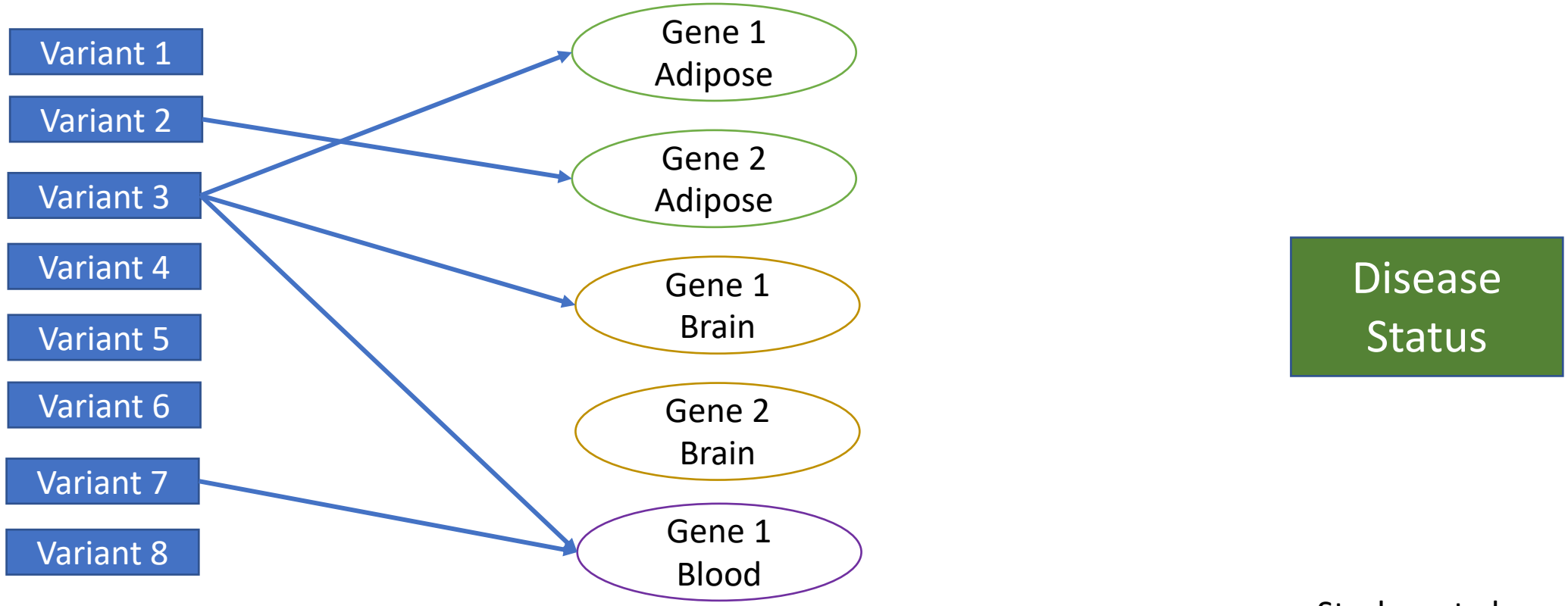
TGFM (tissue gene fine-mapping) distinguishes causal from tagging gene-tissue pairs



Strober et al. medRxiv 2023

TGFM (tissue gene fine-mapping) distinguishes causal from tagging gene-tissue pairs

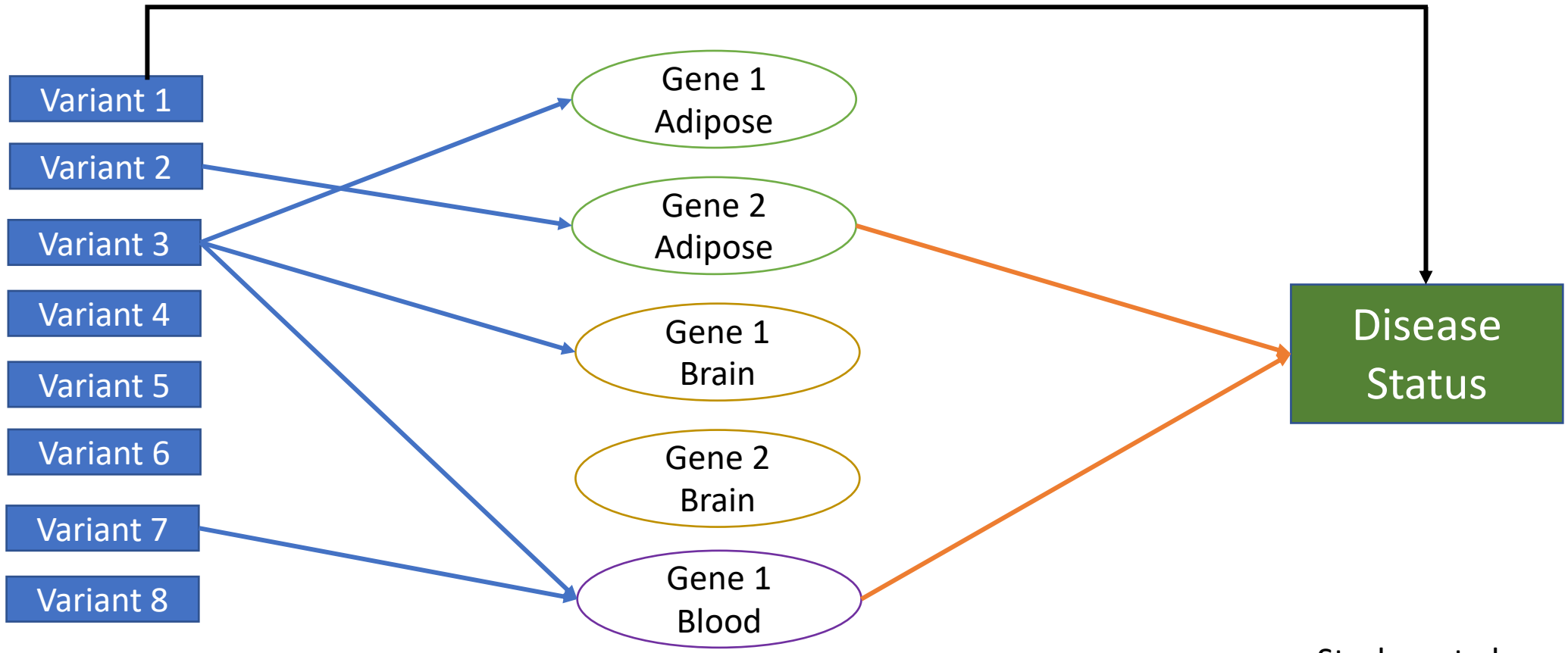
Step 1: Estimate predicted genetic gene expression for each gene-tissue pair using eQTL data



Strober et al. medRxiv 2023

TGFM (tissue gene fine-mapping) distinguishes causal from tagging gene-tissue pairs

Step 2: Identify gene-tissue pairs with non-zero effects on disease while accounting for other gene-tissue pairs and non-mediated genetic effects



Strober et al. medRxiv 2023

TGFM performs joint statistical fine-mapping of eQTL effects and non-mediated effects on disease

Generative Model likelihood

$$Y = G\,\vec{\beta} + \sum_{g}\sum_{t} G\,\overrightarrow{\delta^{gt}}\,\alpha^{gt} + \epsilon$$

Non-mediated genetic effects

Gene-tissue mediated genetic effects

TGFM performs joint statistical fine-mapping of eQTL effects and non-mediated effects on disease

Generative Model likelihood

$$Y = G\,\vec{\beta} + \sum_{g}\sum_{t} G\,\overrightarrow{\delta^{gt}}\,\alpha^{gt} + \epsilon$$

$$\left[\vec{\beta}, \vec{\alpha}\right] = ?$$

Non-mediated genetic effects

Gene-tissue mediated genetic effects

TGFM performs joint statistical fine-mapping of eQTL effects and non-mediated effects on disease

Generative Model likelihood

$$Y = G\,\vec{\beta} + \sum_{g}\sum_{t} G\,\overrightarrow{\delta^{gt}}\,\alpha^{gt} + \epsilon$$

Non-mediated genetic effects

Gene-tissue mediated genetic effects

SuSiE fine-mapping prior

$$[\vec{\beta}, \vec{\alpha}] = \sum_{l} \gamma_l d_l$$

$$\gamma_l \sim Categorical(\pi)$$

$$d_l \sim N(0, \sigma_l^2)$$

Wang et al. J R Stat Soc Series B 2020

TGFM performs joint statistical fine-mapping of eQTL effects and non-mediated effects on disease

Generative Model likelihood

$$Y = G\,\vec{\beta} + \sum_{g}\sum_{t} G\,\overrightarrow{\delta^{gt}}\,\alpha^{gt} + \epsilon$$

Non-mediated genetic effects

Gene-tissue mediated genetic effects

SuSiE fine-mapping prior

$$[\vec{\beta}, \vec{\alpha}] = \sum_{l} \gamma_l d_l$$

$$\gamma_l \sim Categorical(\pi)$$

$$d_l \sim N(0, \sigma_l^2)$$

- **Posterior Inclusion Probabilities (PIPs):** Probability each genetic element (non-mediated variant or gene-tissue pair) has non-zero effect on the disease/trait

TGFM performs joint statistical fine-mapping of eQTL effects and non-mediated effects on disease

Generative Model likelihood

$$Y = G\,\vec{\beta} + \sum_{g}\sum_{t} G\,\overrightarrow{\delta^{gt}}\,\alpha^{gt} + \epsilon$$

Non-mediated genetic effects

Gene-tissue mediated genetic effects

SuSiE fine-mapping prior

$$[\vec{\beta}, \vec{\alpha}] = \sum_{l} \gamma_l d_l$$

$$\gamma_l \sim Categorical(\pi)$$

$$d_l \sim N(0, \sigma_l^2)$$

TGFM allow for multiple (l) causal genetic elements in a genomic region

TGFM performs joint statistical fine-mapping of eQTL effects and non-mediated effects on disease

Generative Model likelihood

$$Y = G\,\vec{\beta} + \sum_{g}\sum_{t} G\,\overrightarrow{\delta^{gt}}\,\alpha^{gt} + \epsilon$$

Non-mediated genetic effects

Gene-tissue mediated genetic effects

SuSiE fine-mapping prior

$$[\vec{\beta}, \vec{\alpha}] = \sum_{l} \gamma_l d_l$$

$$\gamma_l \sim Categorical(\pi)$$

$$d_l \sim N(0, \sigma_l^2)$$

TGFM improves fine-mapping power by leveraging data from across the genome to learn tissue-specific prior probability (π) of each genetic element being causal

TGFM performs joint statistical fine-mapping of eQTL effects and non-mediated effects on disease

Generative Model likelihood

$$Y = G\,\vec{\beta} + \sum_{g}\sum_{t} G\,\overrightarrow{\delta^{gt}}\,\alpha^{gt} + \epsilon$$

Non-mediated genetic effects

Gene-tissue mediated genetic effects

SuSiE fine-mapping prior

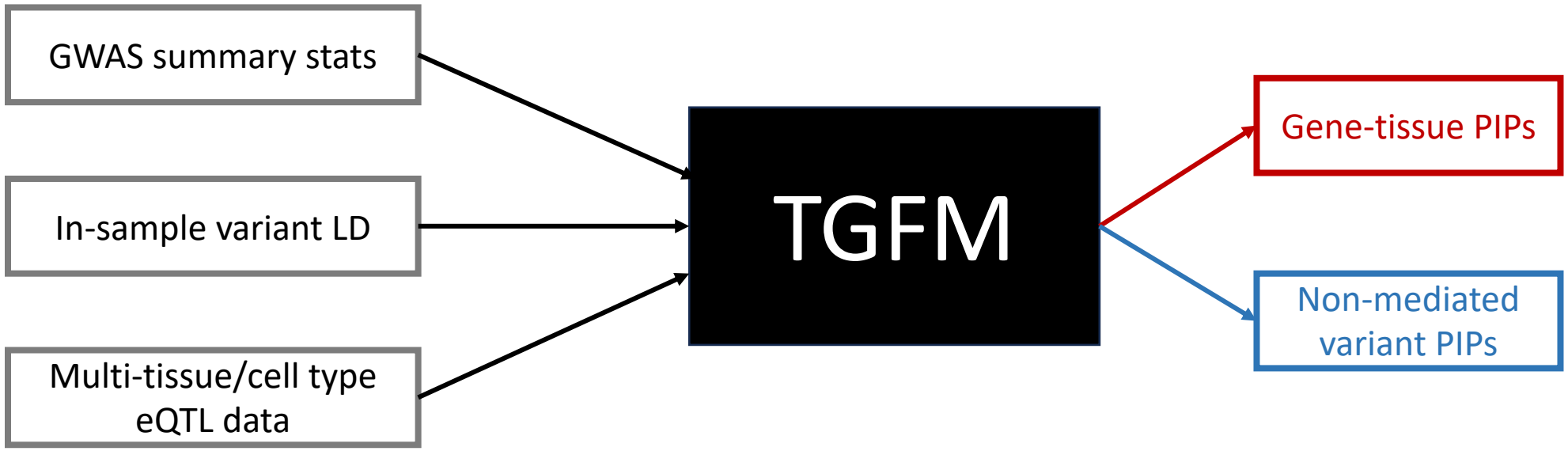
$$[\vec{\beta}, \vec{\alpha}] = \sum_{l} \gamma_l d_l$$

$$\gamma_l \sim Categorical(\pi)$$

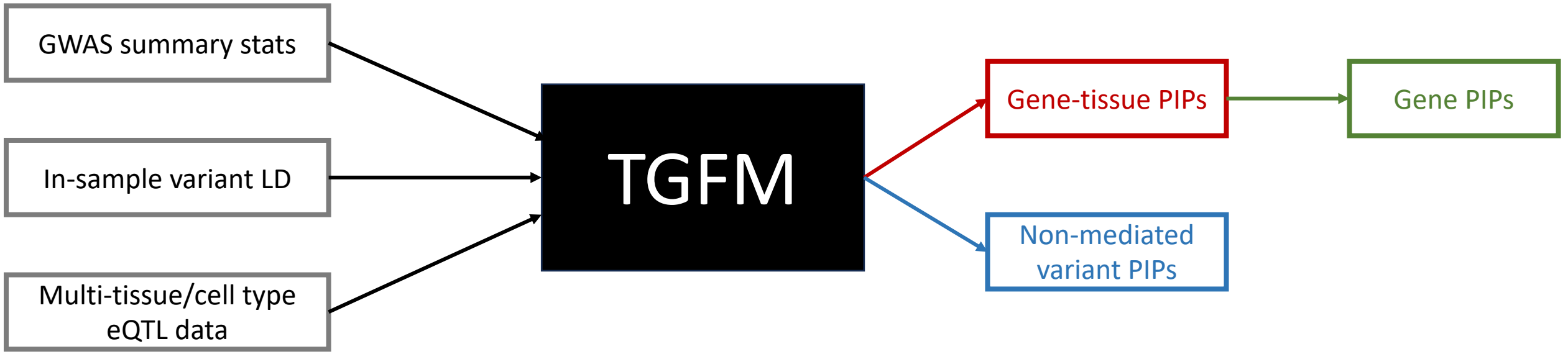
$$d_l \sim N(0, \sigma_l^2)$$

TGFM improves fine-mapping calibration using a sampling approach to model uncertainty in predicted causal eQTL effect sizes (δ^{gt})

TGFM input data and output statistics



TGFM input data and output statistics

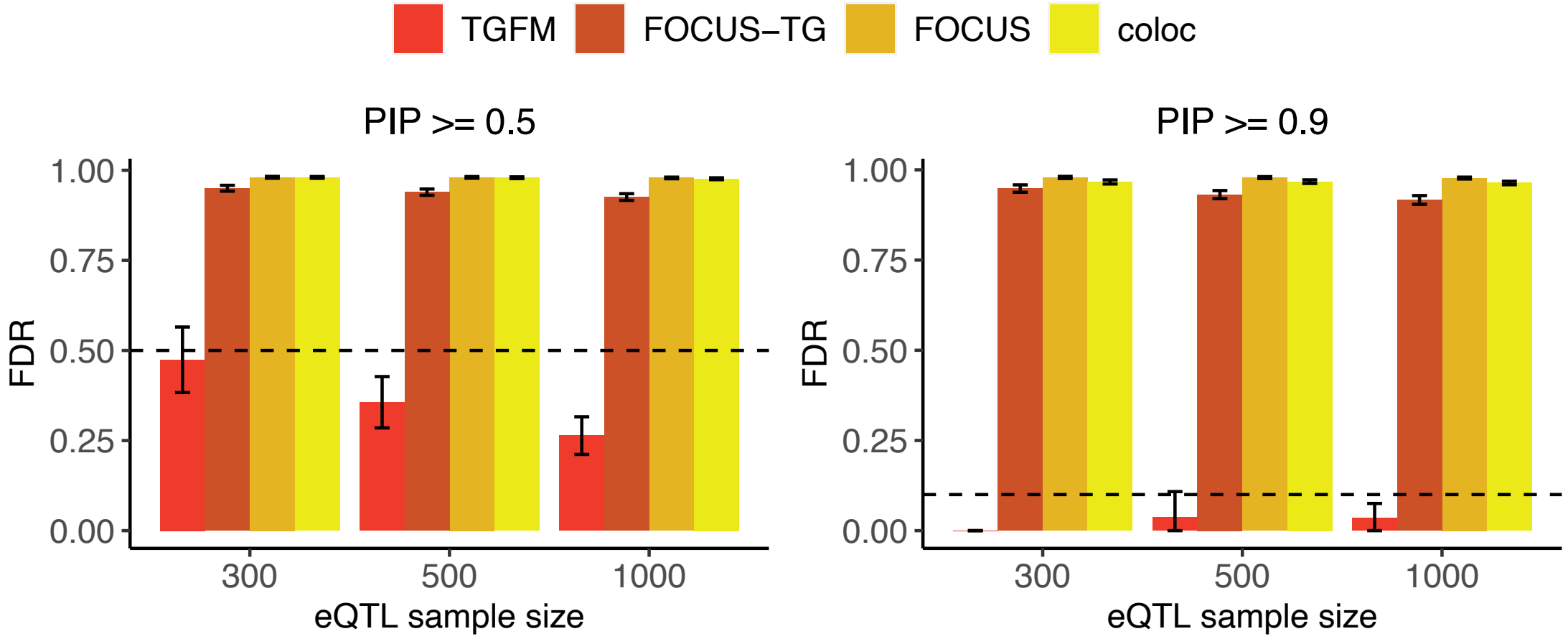


*We refer to variants, (genetic) genes, and (genetic) gene-tissue pairs as **genetic elements***

Outline

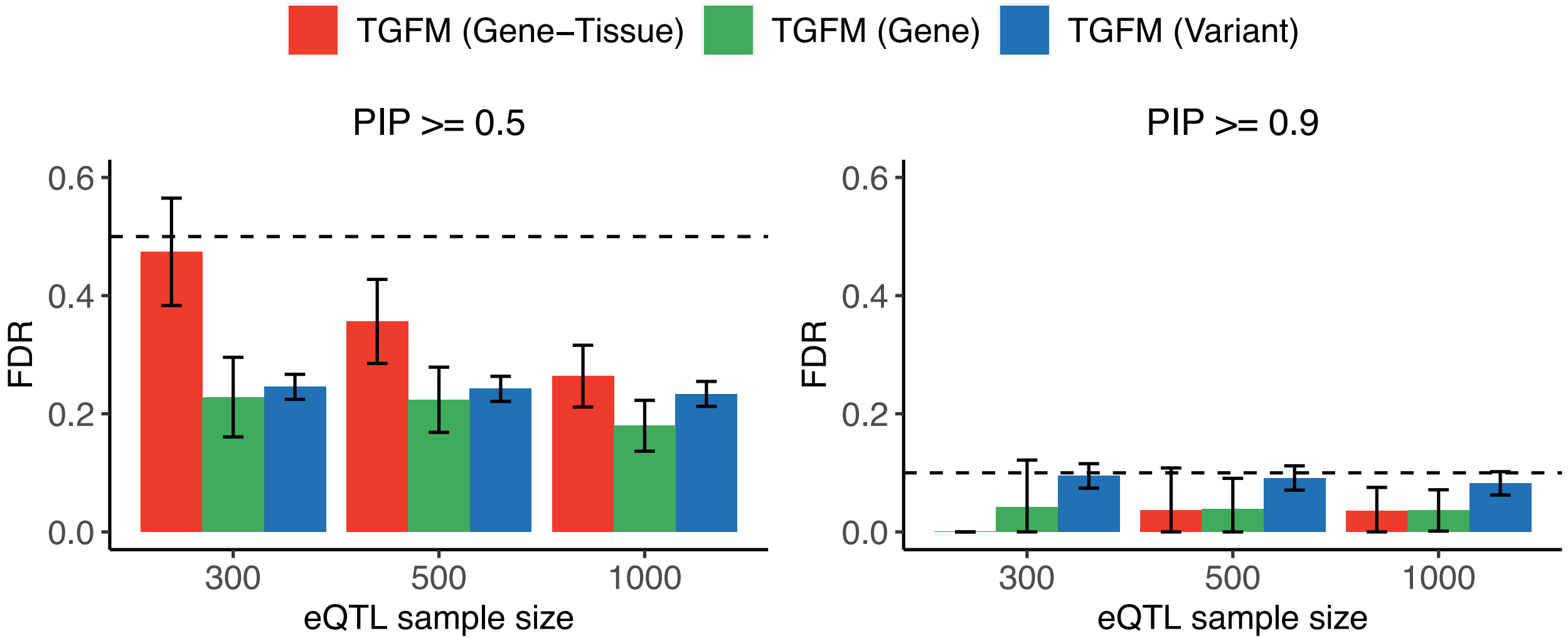
1. Motivation for fine-mapping causal tissues and genes at disease loci
2. Methods
3. Simulations
4. Results: Tissue-gene fine-mapping for 45 UK Biobank diseases and complex traits with 38 GTEx tissues

TGFM gene-tissue pair fine-mapping FDR is well-calibrated in simulations



FOCUS: Mancuso et al. Nat. Genet. 2019
coloc: Giambartolomei et al. Plos Genet. 2014

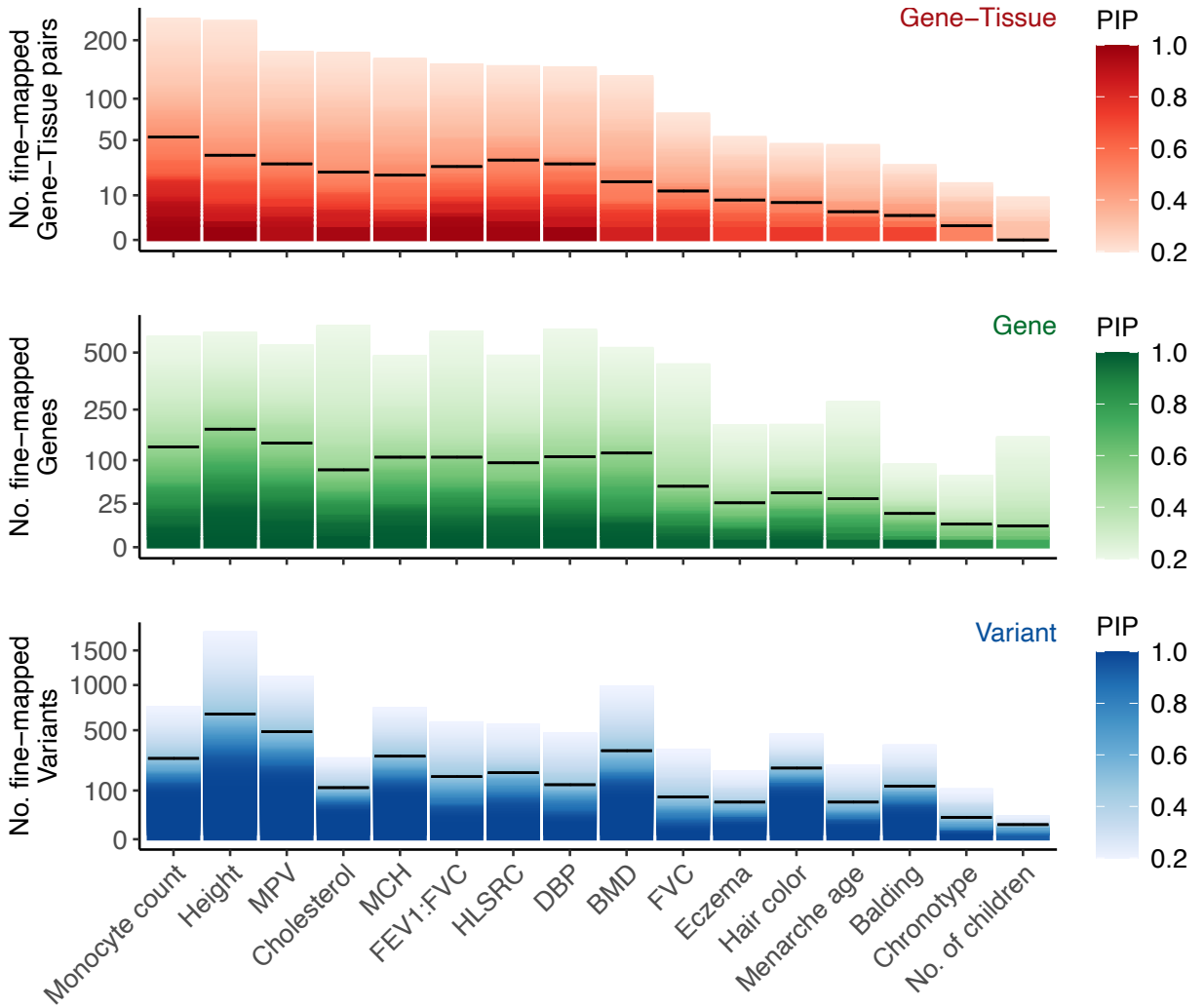
TGFM well calibrated to fine-map multiple classes of genetic elements in simulations



Outline

1. Motivation for fine-mapping causal tissues and genes at disease loci
2. Methods
3. Simulations
4. Results: Tissue-gene fine-mapping for 45 UK Biobank diseases and complex traits with 38 GTEx tissues

TGFM fine-maps genetic elements underlying 45 UK Biobank diseases and traits using 38 GTEx tissues

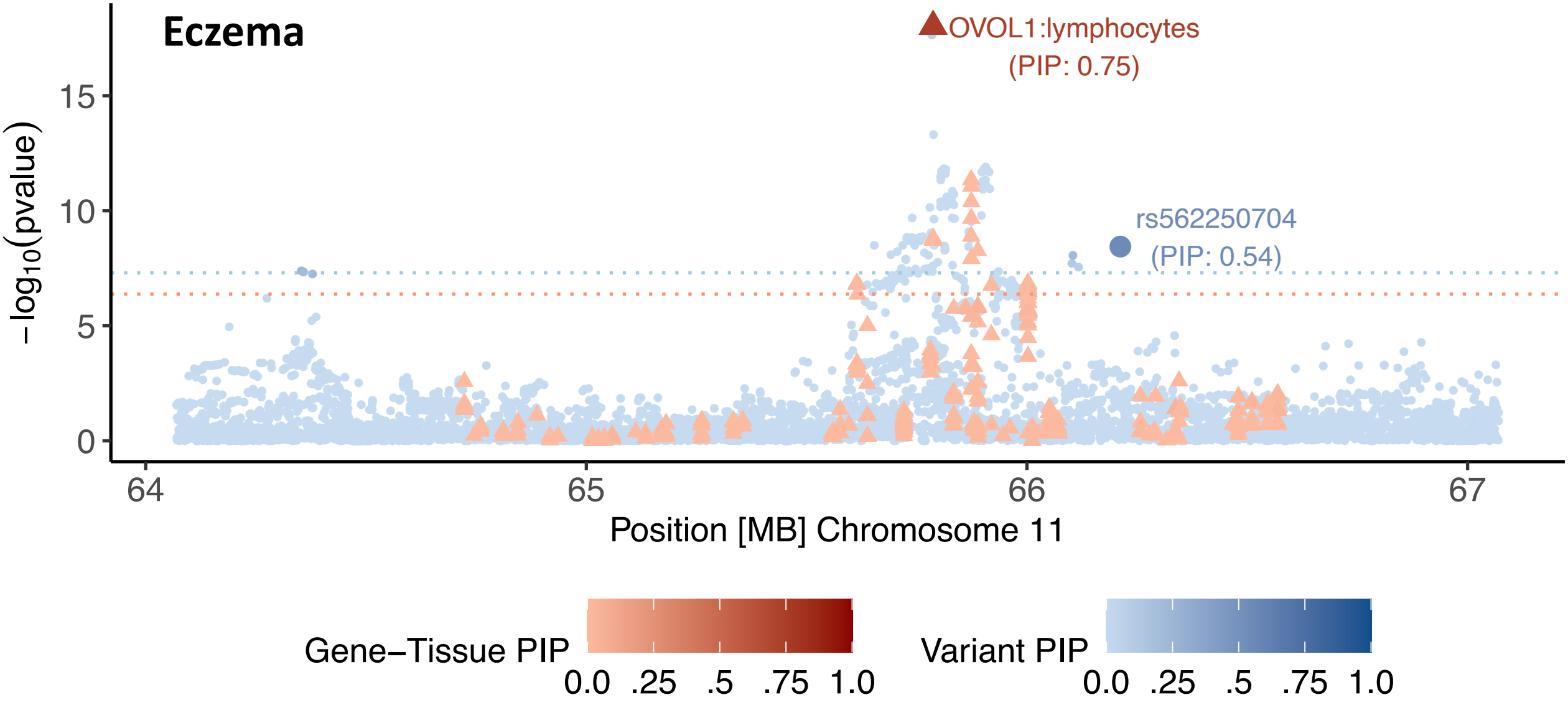


16 gene-tissue pairs/trait at PIP > 0.5

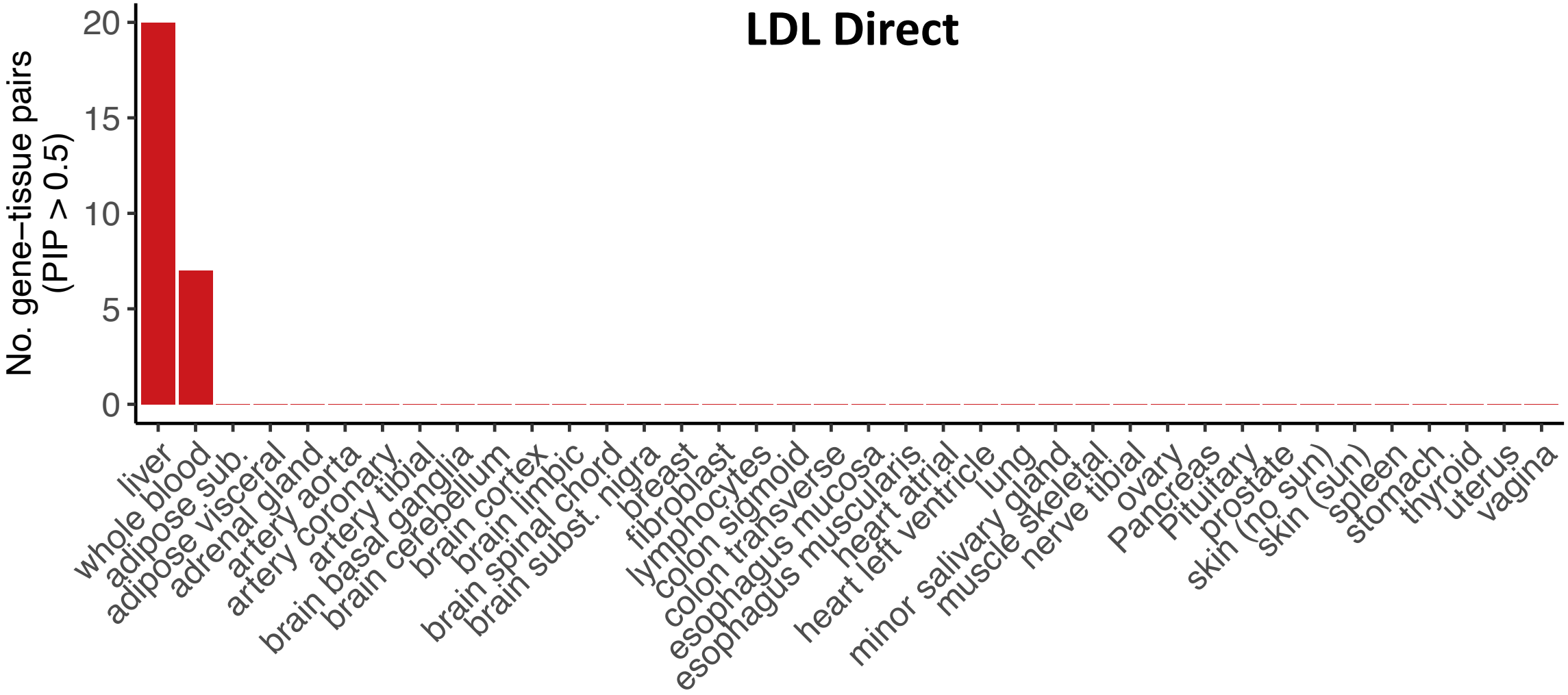
62 genes/trait at PIP > 0.5

131 variants/trait at PIP > 0.5

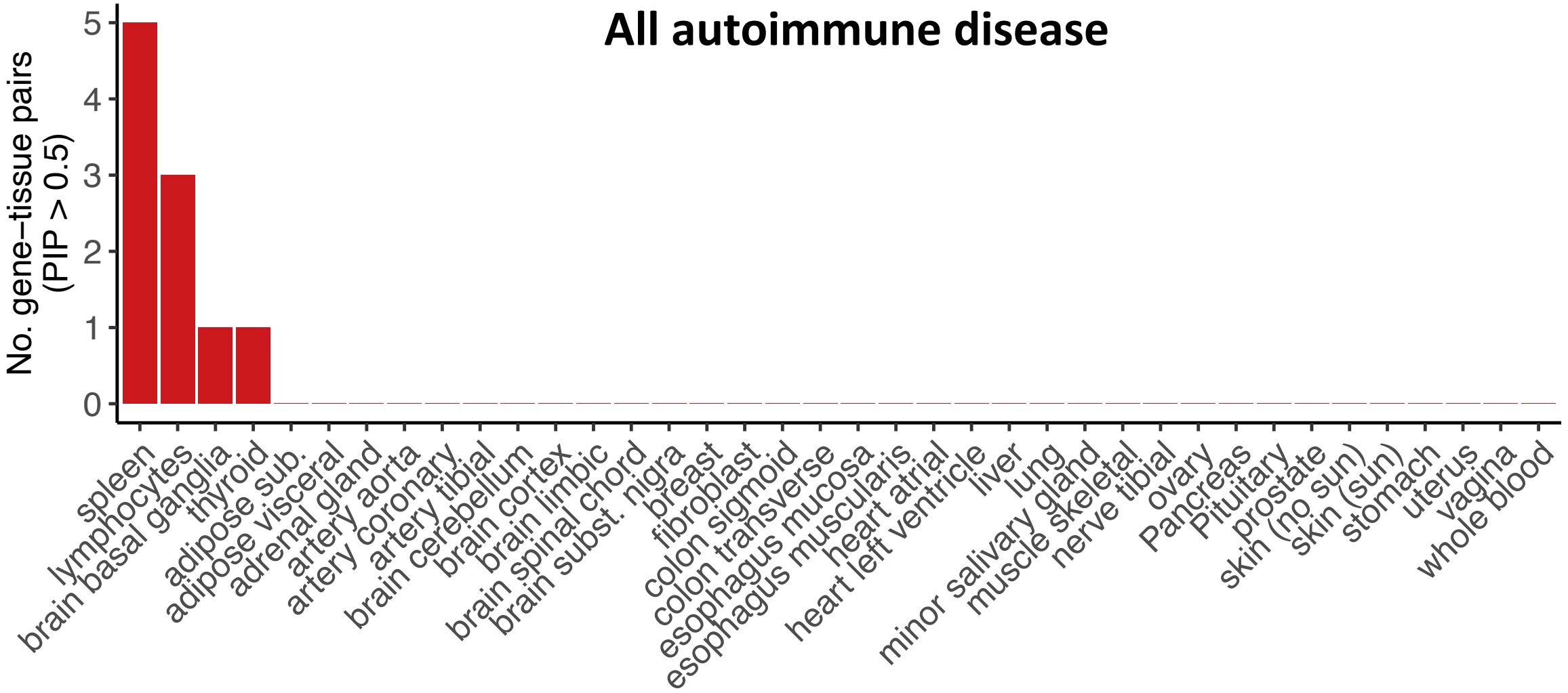
TGFM applied to UK Biobank diseases/traits identifies disease genes and their tissue of action



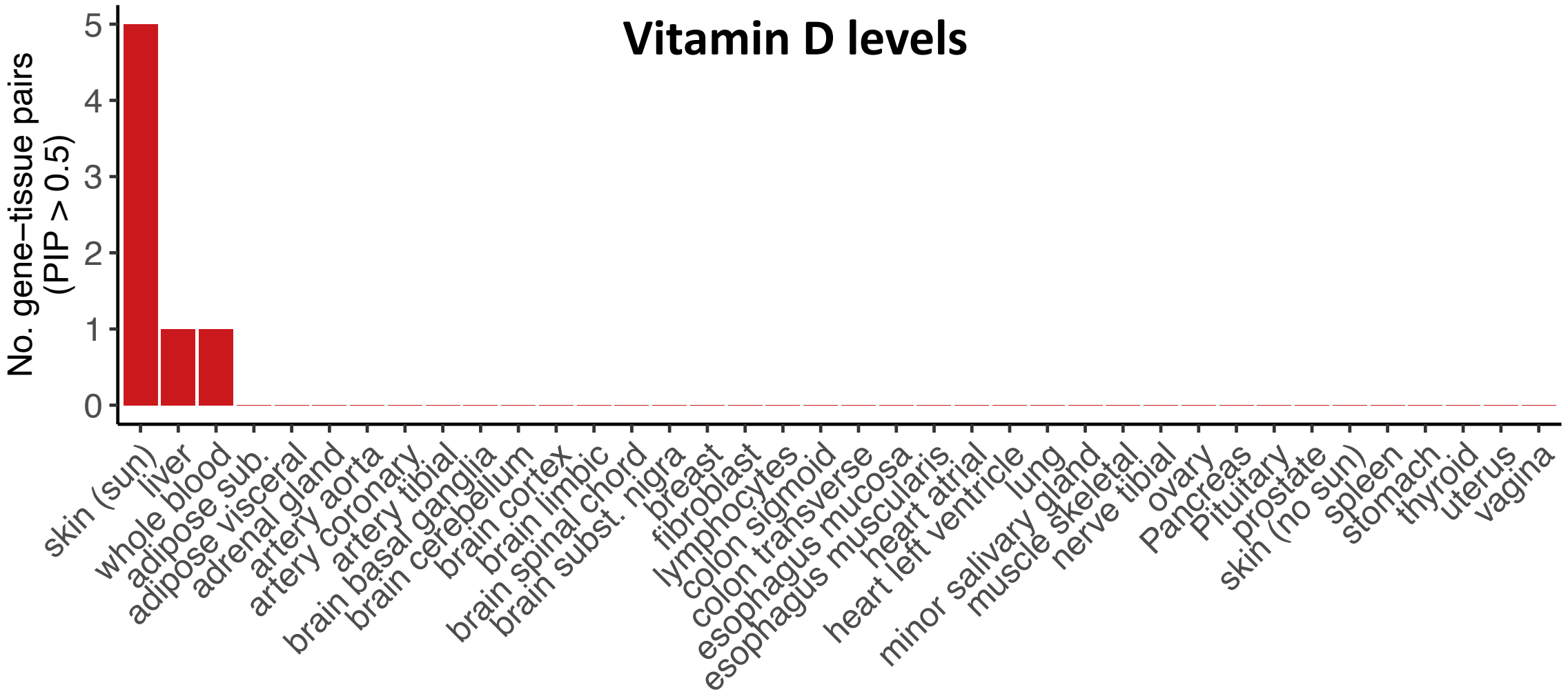
TGFM implicated gene-tissue-disease triplets are concentrated in disease-critical tissues



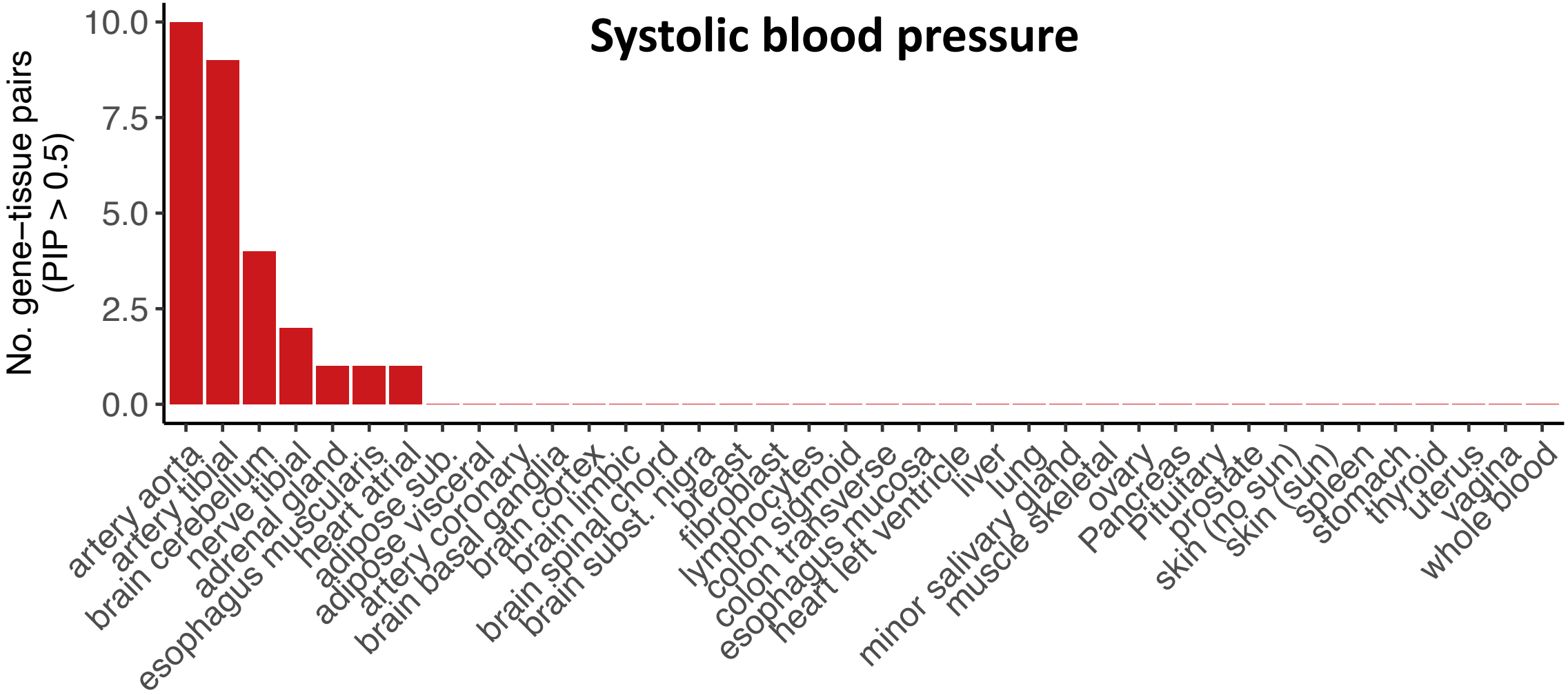
TGFM implicated gene-tissue-disease triplets are concentrated in disease-critical tissues



TGFM implicated gene-tissue-disease triplets are concentrated in disease-critical tissues

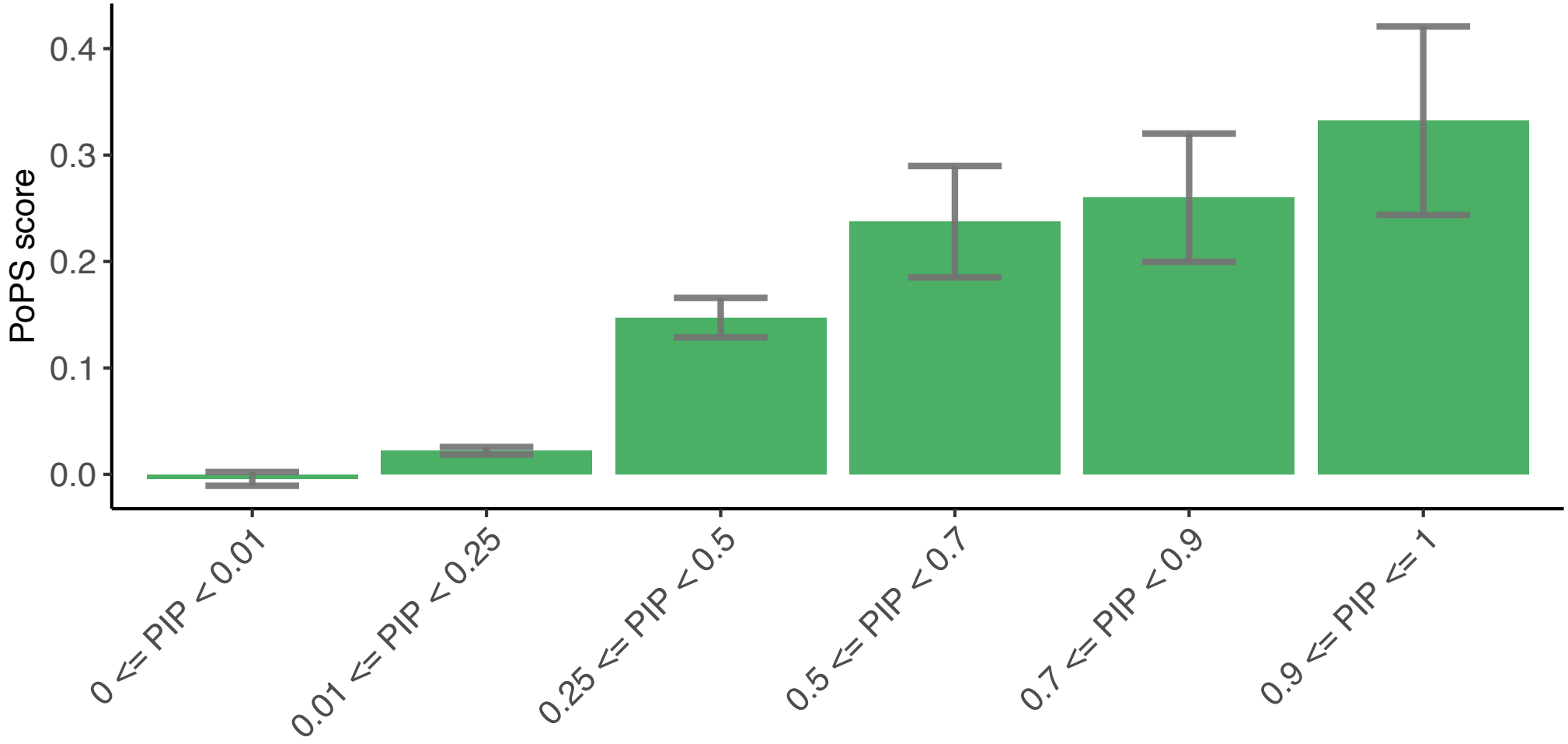


TGFM implicated gene-tissue-disease triplets are concentrated in disease-critical tissues



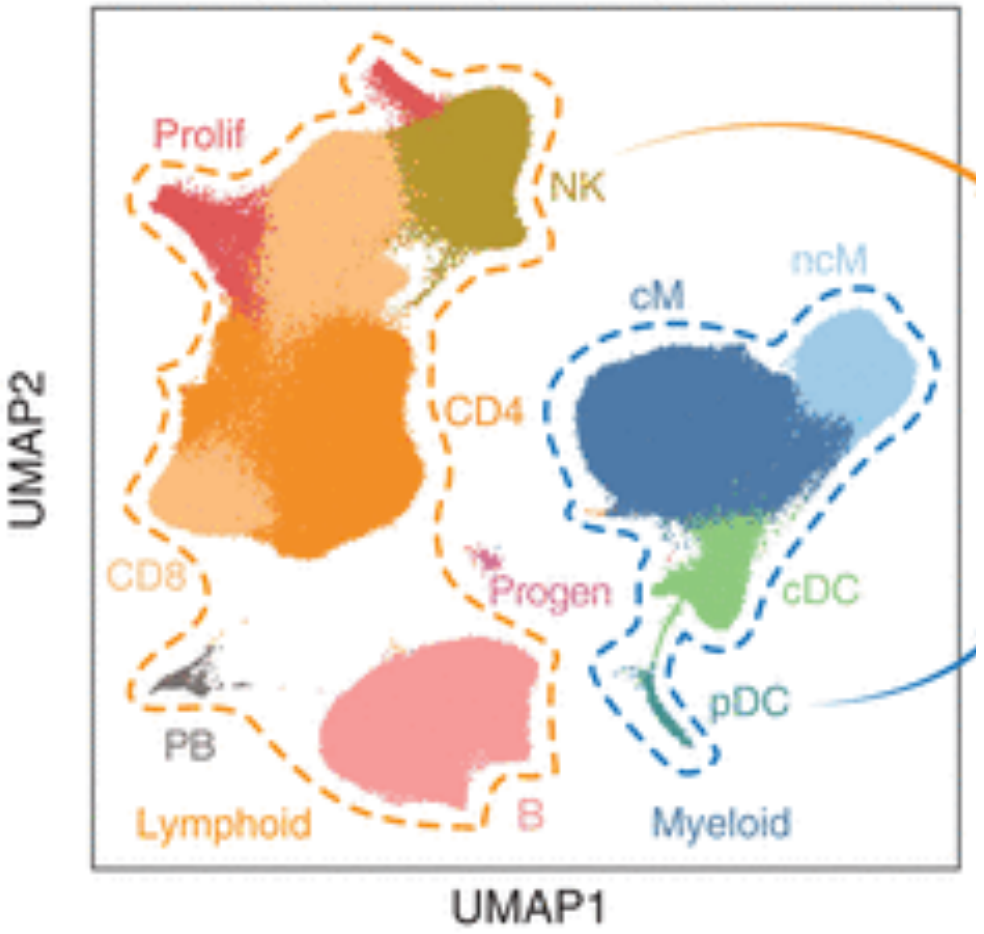
High TGFM PIP genes are strongly enriched within high PoPS genes

- PoPS (Weeks et al. Nat. Genet 2023) is trait-specific gene score based many gene features, such as cell-type specific expression

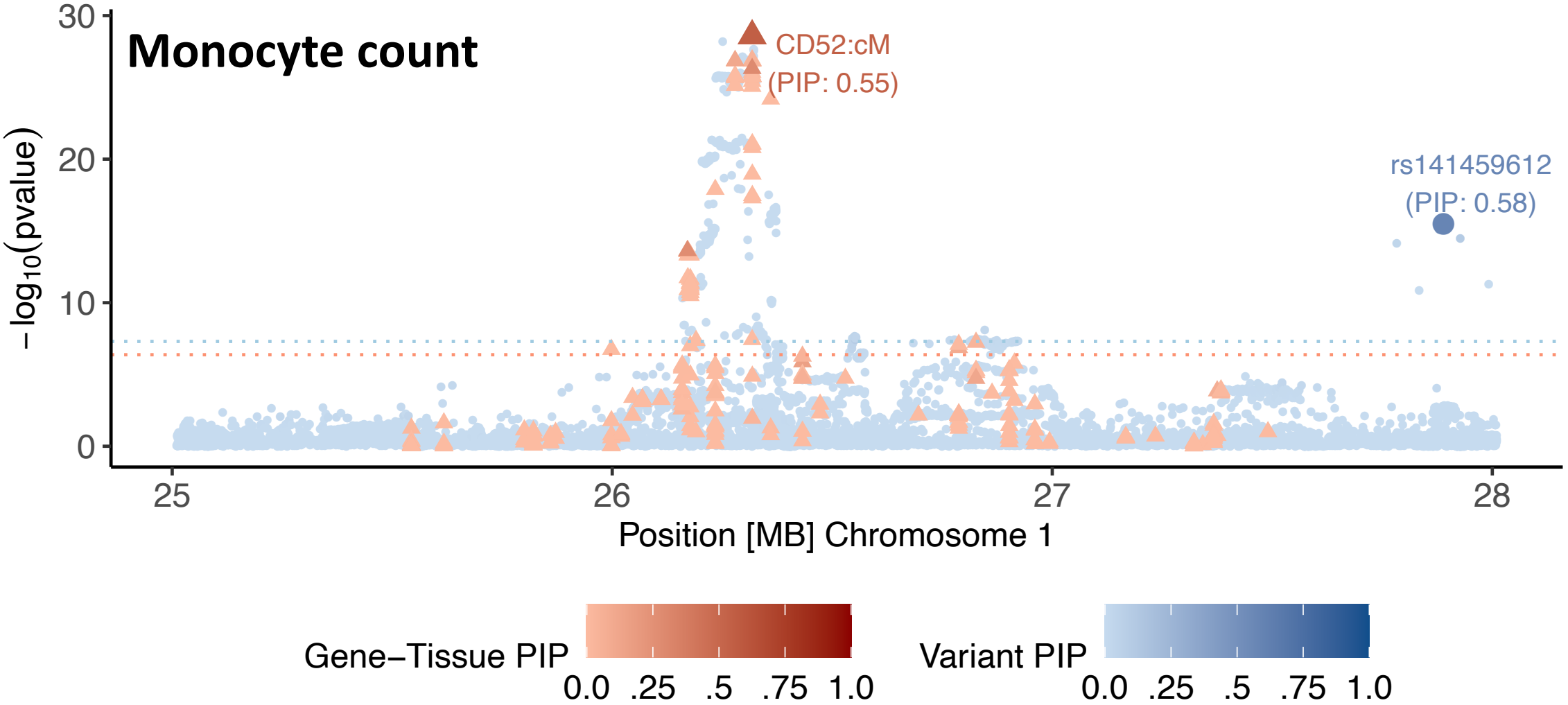


Application of TGFM to fine-grained PBMC cell types (Perez et al. Science 2022)

- Generate pseudobulk in each pre-defined cell type
- 9 cell types:
  - B
  - NK: Natural Killer cells
  - Prolif: Proliferation cells
  - T4
  - T8
  - cDC: classical Dendritic cells
  - pDC: Plasmacytoid dentric cells
  - cM: classical Monocytes
  - ncM: non-classical Monocytes

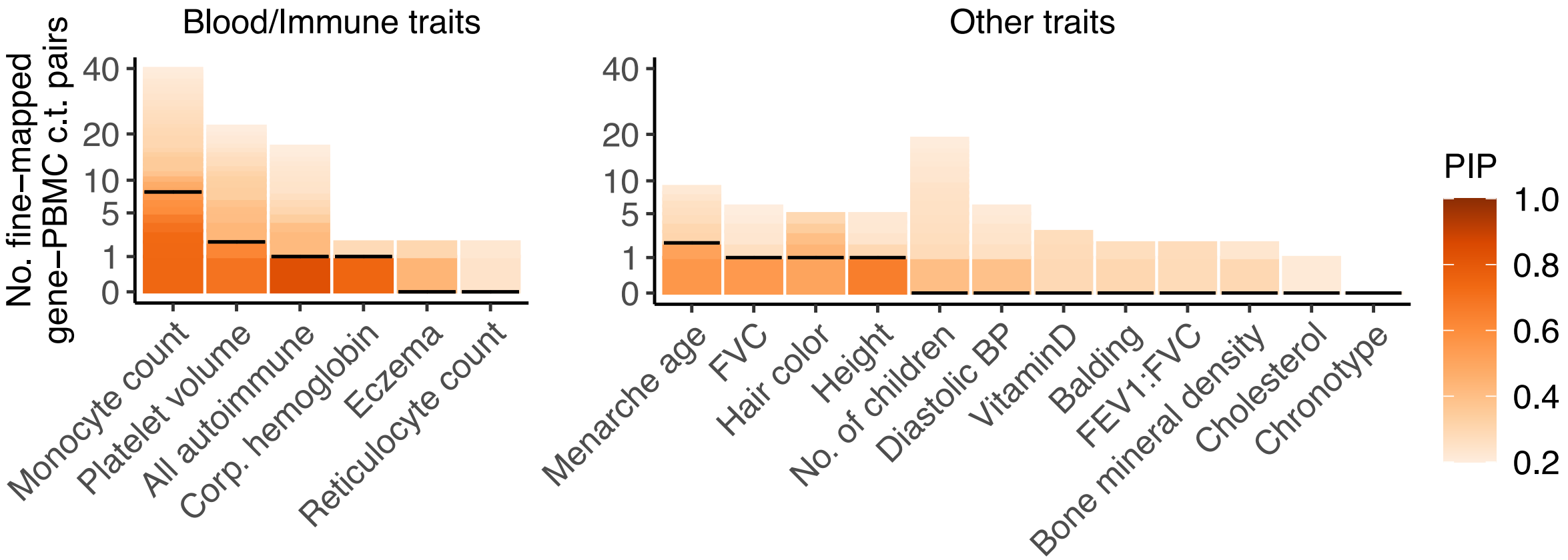


Fine-grained PBMC cell type help resolve the biological mechanism underlying disease loci



TGFM identified 30 addition gene-PBMC cell type pairs at PIP > 0.5

- Primarily (23 of 30) for autoimmune and blood cell traits



Conclusions

- Develop new statistical method (TGFM) to jointly fine-map the causal gene and tissue at disease loci
- TGFM generates well-calibrated PIPs for gene-tissue pairs, genes, and non-mediated variants in simulations
- High PIP gene-tissue pairs often originate in disease relevant tissues and high PIP genes are enriched in independent gene sets
- TGFM applied to fine-grained cell types in single-cell eQTL data can help resolve the biological mechanism underlying disease loci

Acknowledgements

- Alkes Price
- Martin Zhang
- Tiffany Amariuta
- Jordan Rossen

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH