



Testing 15 billion QTL associations: fine-mapping and multivariate methods

Gao Wang¹, Yuxin Zou², Wei Wang² and Matthew Stephens^{1,2}

¹Department of Human Genetics and ²Department of Statistics, The University of Chicago, Chicago, IL 60637, USA

Motivation

Association testing of many effects on many conditions

- **Fine-mapping** of “causal” variants
 - ▷ A challenging task due to correlated observed effects
 - ▷ Typically using Bayesian Variable Selection Regression
- **Multi-phenotype association** for shared genetic factors
 - ▷ Effects can be similar or different cross conditions
 - ▷ Learn and exploit patterns of sharing to boost power

Current success and limitations

- Two new methods developed for these two problems
- A fundamental problem remains: *pleiotropy or linkage?*

Fine-mapping Wang et. al. 2018

Sum of Single Effect (SuSiE) regression [1]

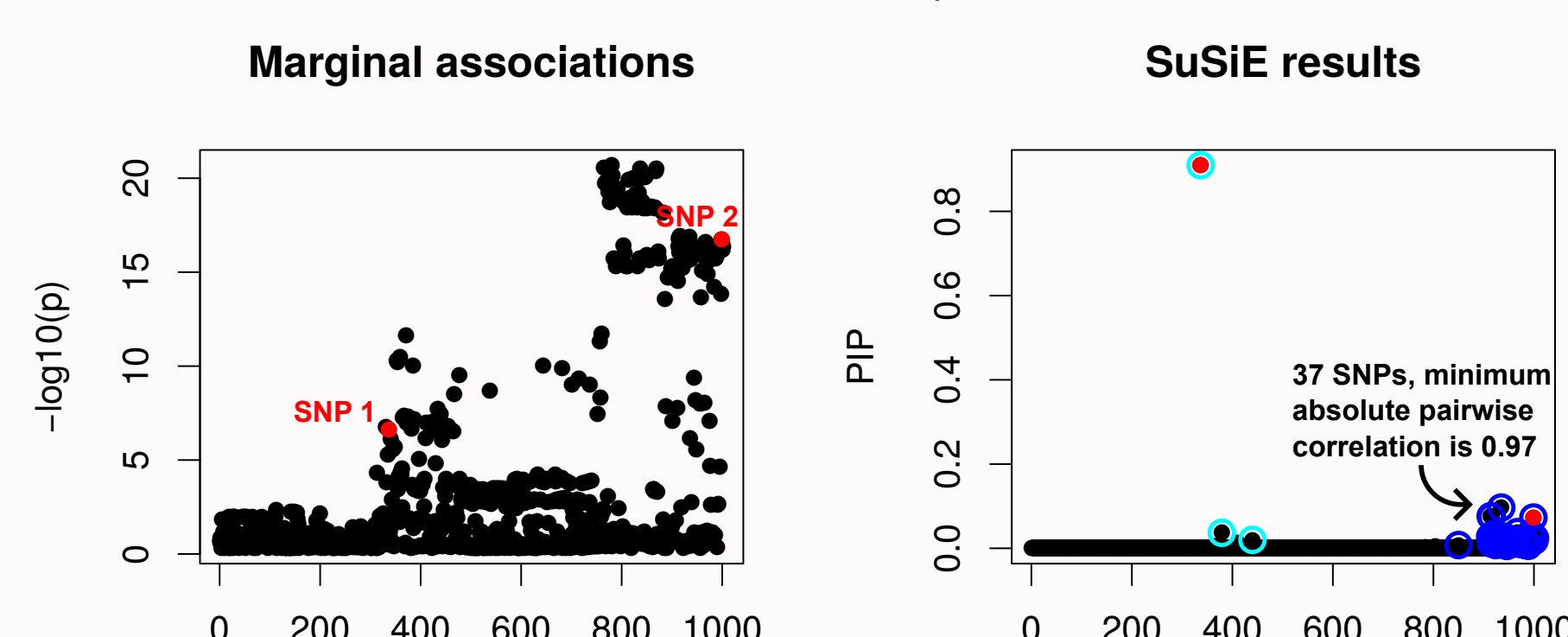
$$y = Xb + e, e \sim N(0, \sigma^2 I_n)$$

$$b = \sum_{l=1}^L b_l, b_l = \gamma_l \beta_l$$

$$\gamma_l \sim \text{Mult}(1, \pi), \beta_l \sim N(0, \sigma_{\beta_l}^2)$$

Posterior of effects is approximated by

$$q(b_1, \dots, b_L) = \prod_l q_l(b_l)$$



- The companion algorithm is fast (complexity $\mathcal{O}(np)$)
- Summary statistics model extension available

Multivariate analysis Urbat et. al. 2018

Multivariate Adaptive SHrinkage (MASH) model [2]

$$\hat{\beta} \sim N_R(\beta, S)$$

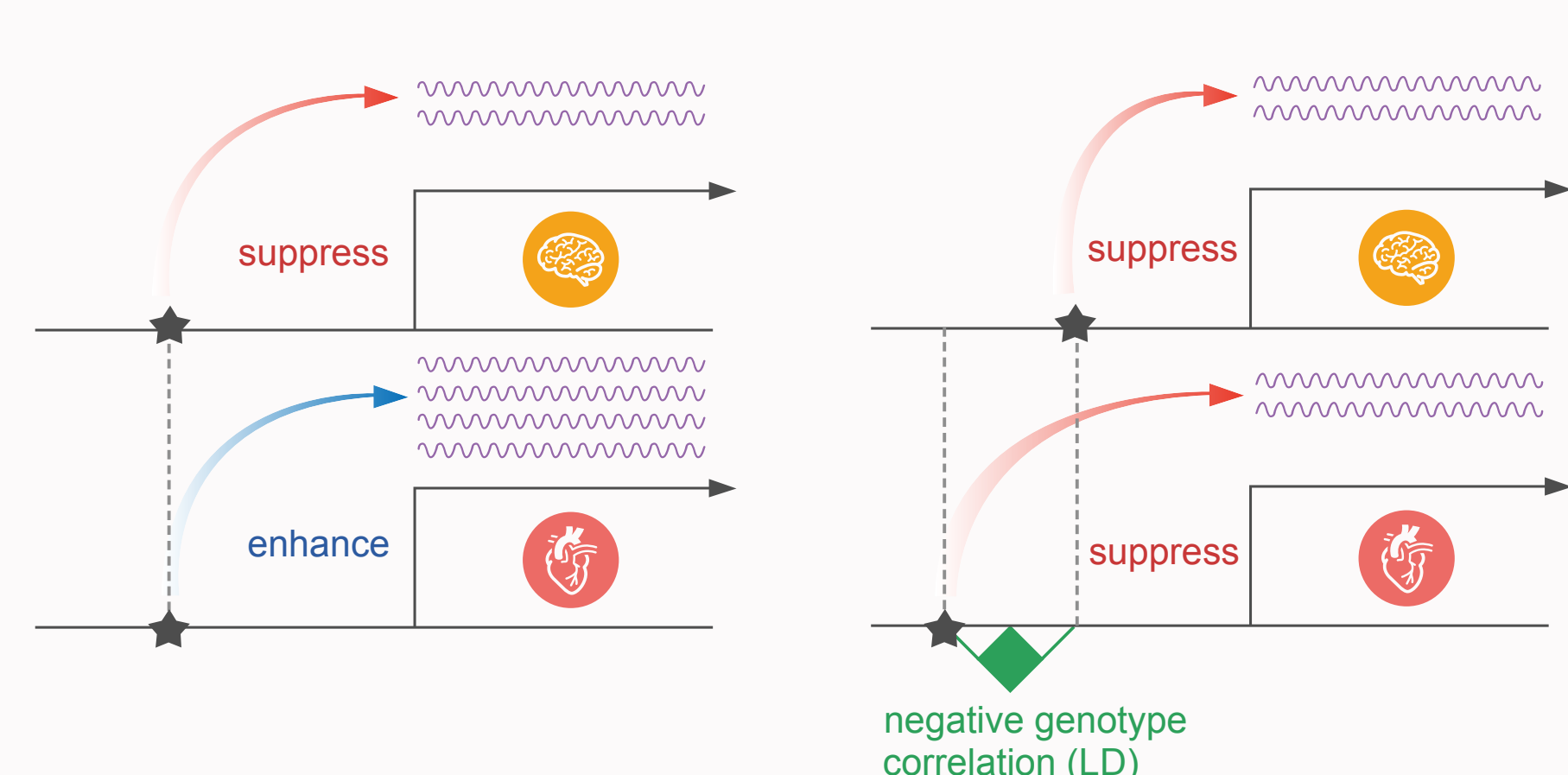
$$\beta \sim \sum_k \pi_k N_R(0, U_k)$$

U_k : what to share; π_k : how much to share



A fundamental problem

MASH model ignores correlations in $\hat{\beta}$



- Shared yet opposite effects observed between conditions
- Induced by two condition specific effects in LD
- Motivates a joint fine-mapping and multivariate method

Multivariate fine-mapping and adaptive shrinkage (M&M)

$$Y = XB + E$$

$$E \sim N_{N \times R}(0, S \otimes I)$$

$$B = \sum_{l=1}^L B_l$$

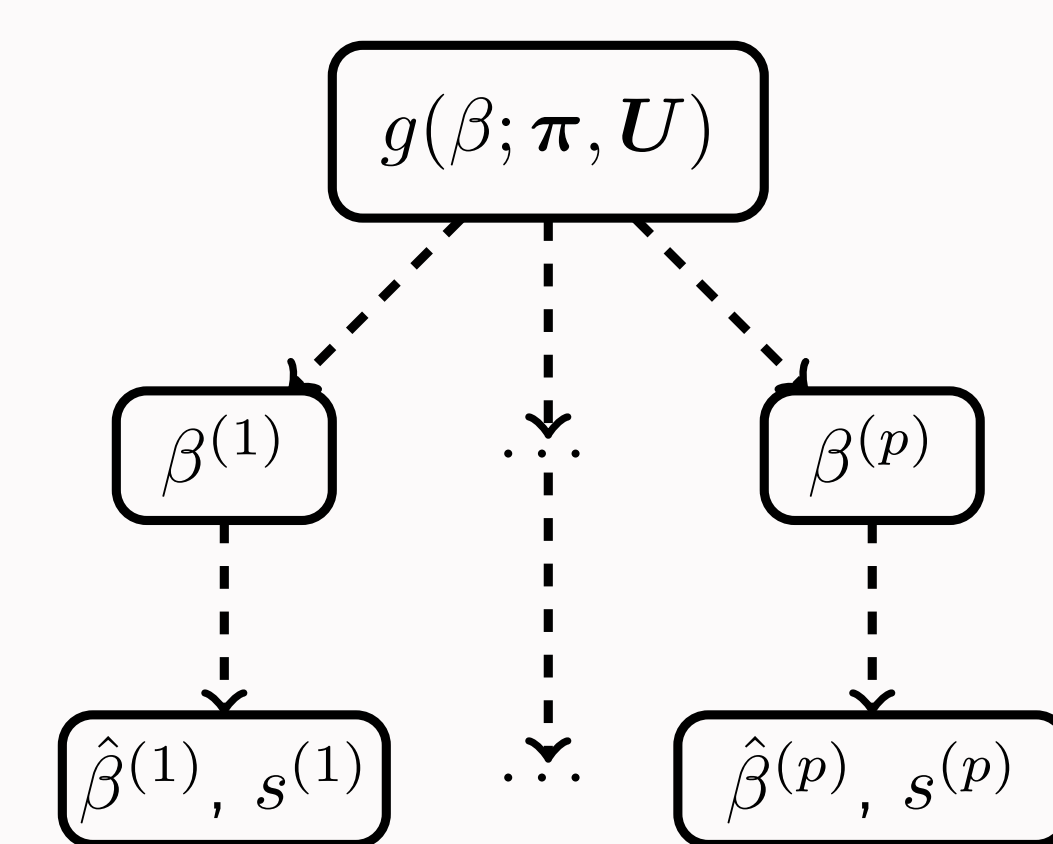
$$B_l = \gamma_l b_l^\top$$

$$\gamma_l \sim \text{Mult}(1, \alpha)$$

$$b_l \sim \sum_k \pi_k N_R(0, U_k)$$

An Empirical Bayes approach to 15 billion association tests

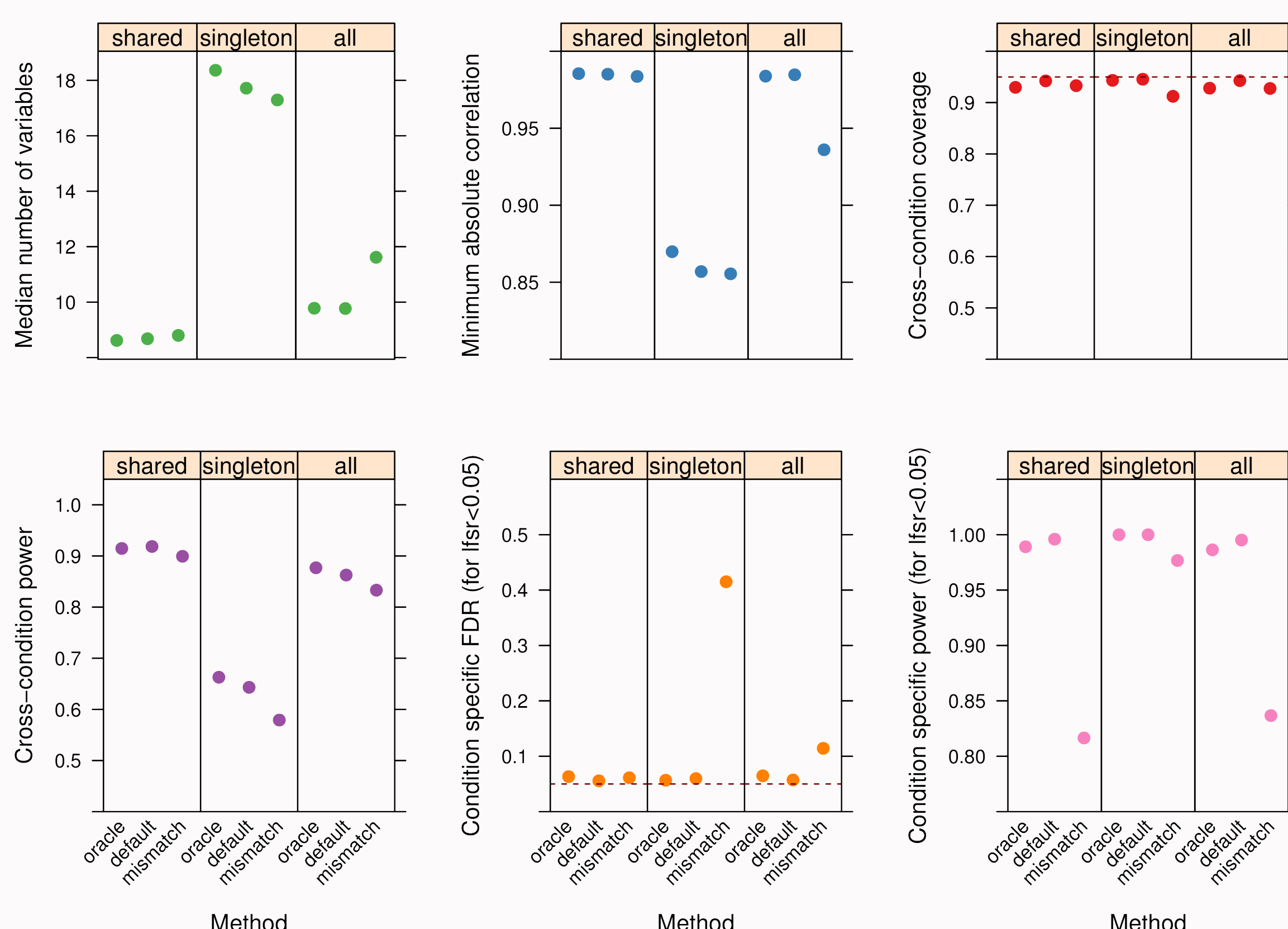
1. Obtain summary statistics per effect per condition
2. Estimate residual covariance S
3. Learn possible patterns of effect sharing U_k [3]
4. Estimate mixture prior weights π_k [2]
5. \leftrightarrow Fit multivariate SuSiE model with MASH prior



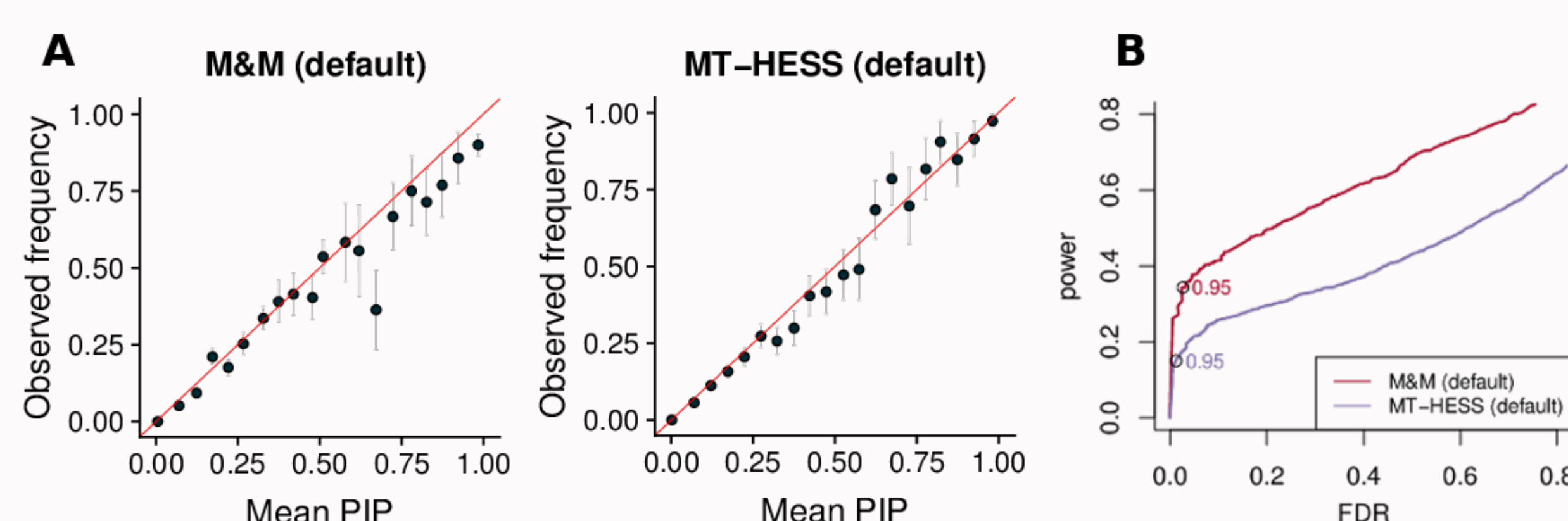
Generative model illustration

- Results from Step 5 can be used to refine Steps 2 — 4, until converged.

Numerical studies and comparisons with MT-HESS



Condition specific posterior inclusion probabilities (PIP)

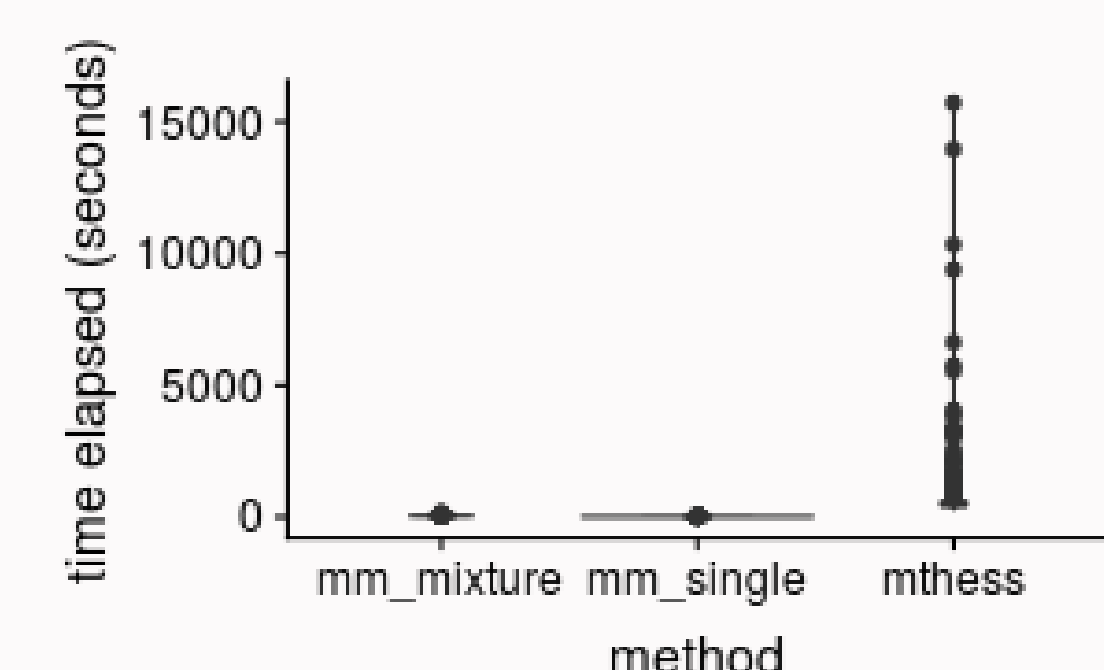


Benefits of adaptive shrinkage prior

- **Oracle**: prior under which data is simulated (shared, singleton or combined)
- **Default**: M&M default mixture prior (default to uniform weights)
- **Mismatch**: when choice of prior mismatch the simulated data, eg, use “shared” prior to analyze singleton scenarios, or vice versa.

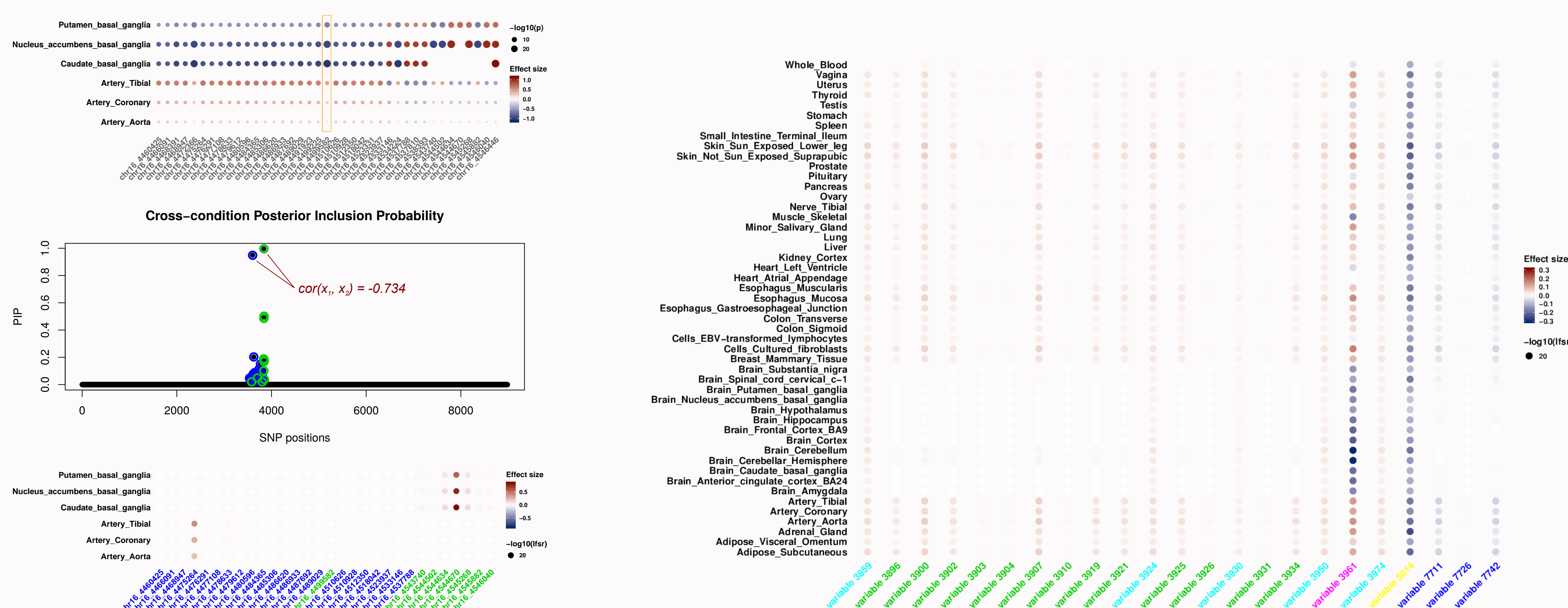
Computation speed (unit: second)

| | Walltime | M&M | MT-HESS |
|--------|----------|--------|---------|
| Median | | 41.95 | 499.97 |
| Min. | | 25.18 | 376.67 |
| Max. | | 181.12 | >20hrs |



M&M is written in R with C++ extensions.
MT-HESS (version 1.99) is in C++.

GTEx [4] multi-tissue eQTL fine-mapping: potentials and challenges



References

- [1] Gao Wang, Abhishek Sarkar, Peter Carbonetto, and Matthew Stephens. A simple new approach to variable selection in regression, with application to genetic fine-mapping. *bioRxiv*, December 2018.
- [2] Sarah M. Urbat, Gao Wang, Peter Carbonetto, and Matthew Stephens. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nature Genetics*, 51(1):187–195, November 2018.
- [3] Wei Wang and Matthew Stephens. Empirical bayes matrix factorization. *arXiv*, 2018.
- [4] The GTEx Consortium. The GTEx consortium atlas of genetic regulatory effects across human tissues. *bioRxiv*, October 2019.

Gao Wang
wangow@gmail.com
Matthew Stephens
mstephens@uchicago.edu

M&M R package
<http://github.com/gaow/mmbr>

