MultiSuSiE improves multi-ancestry fine-mapping in All of Us whole-genome sequencing data

Jordan Rossen

PhD Candidate

Department of Epidemiology, Harvard School of Public Health

November 2, 2023

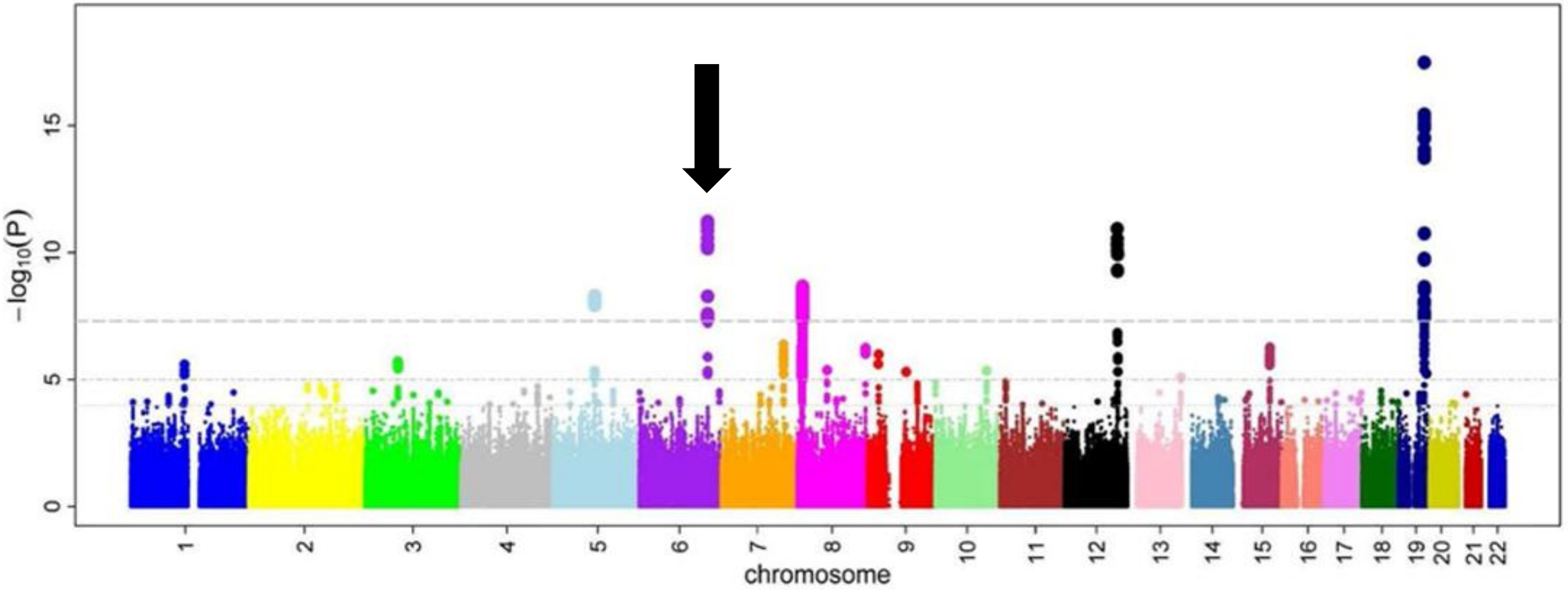
Outline

1. Background
2. Method and Simulations
3. Results on Real Traits
4. Conclusions and Future Directions

Outline

1. **Background**
2. Method and Simulations
3. Results on Real Traits
4. Conclusions and Future Directions

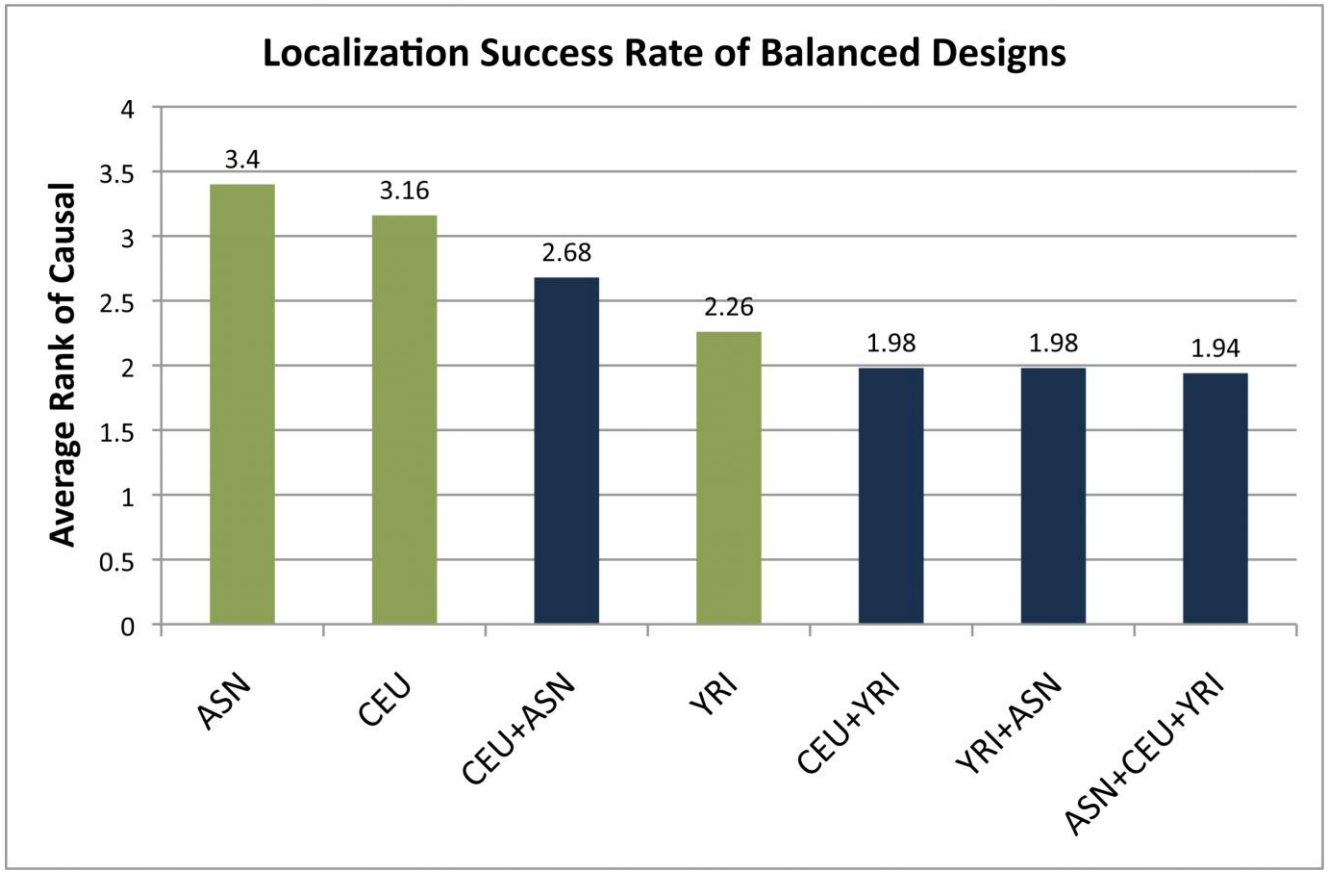
Fine-mapping methods identify causal mutations at GWAS loci



Manhattan plot from Ikram et al. 2010 PLoS Genet

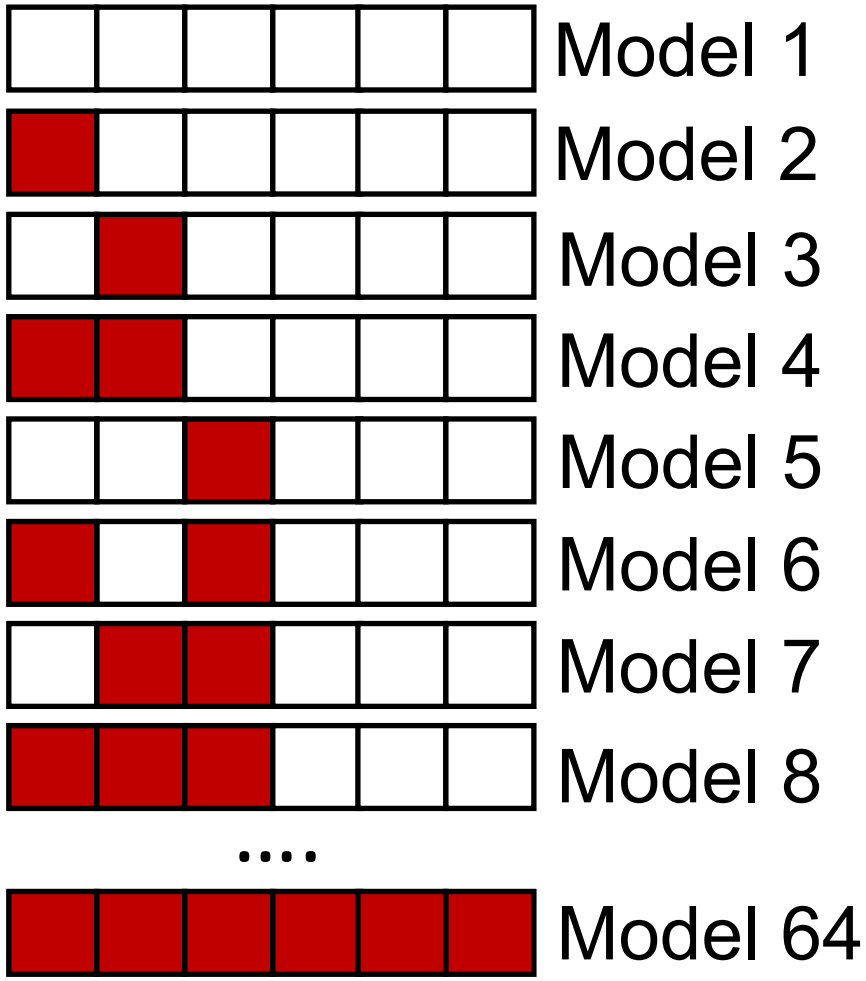
Integrating data from multiple ancestries improves fine-mapping

- Fine-mapping is hard because of linkage disequilibrium (LD)
- Integrating data from multiple population improves ability to detect causal SNPs



Zaitlen et al. 2010 Am. J. Hum. Genet.

Published multi-ancestry methods fail at multiple causal variant fine-mapping

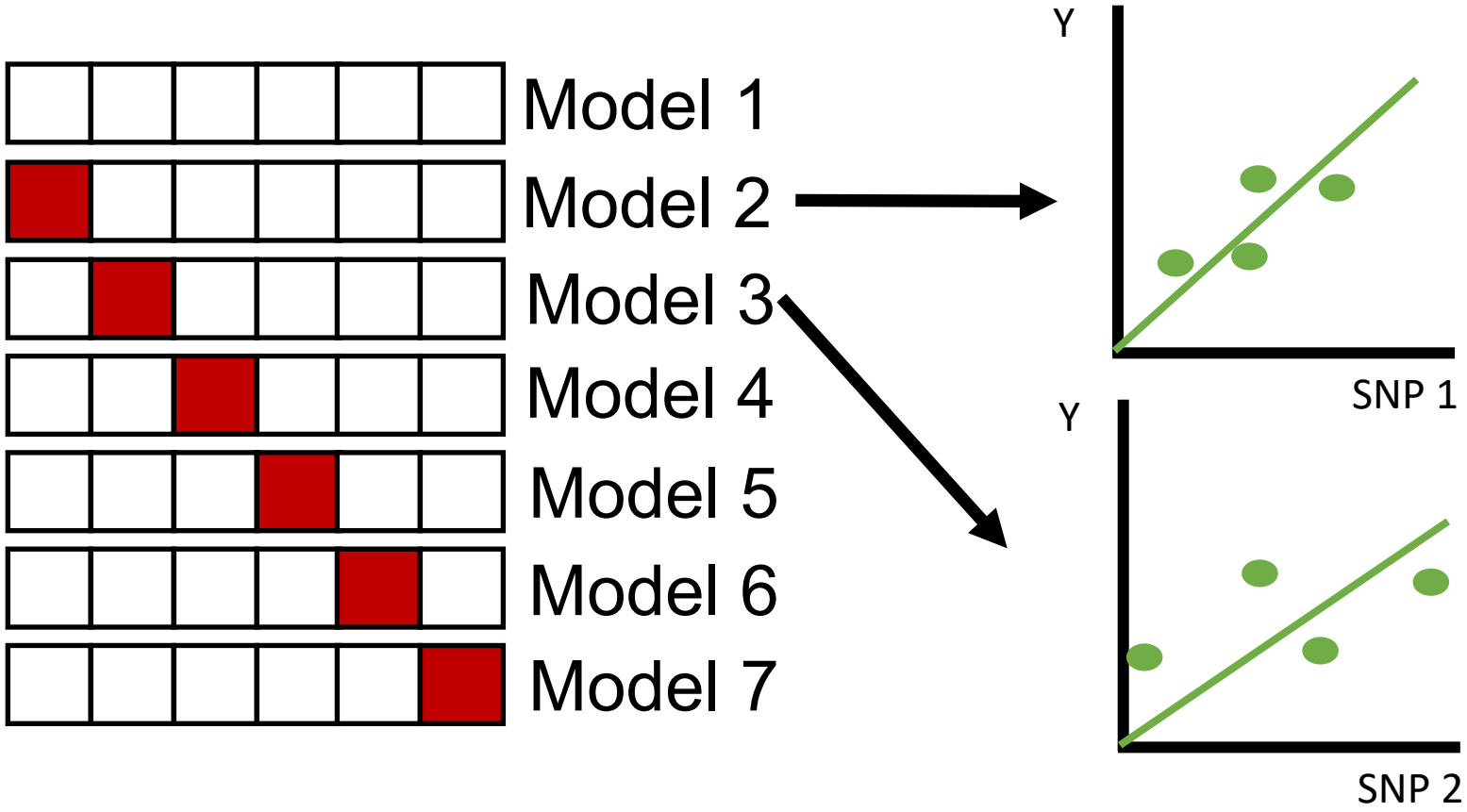


- Exhaustively evaluating all possible configurations of causal variants is very expensive for large numbers of SNPs, $O(2^P)$ models must be evaluated
- Published multi-ancestry methods use non-exhaustive search algorithms that often fail

Null SNP

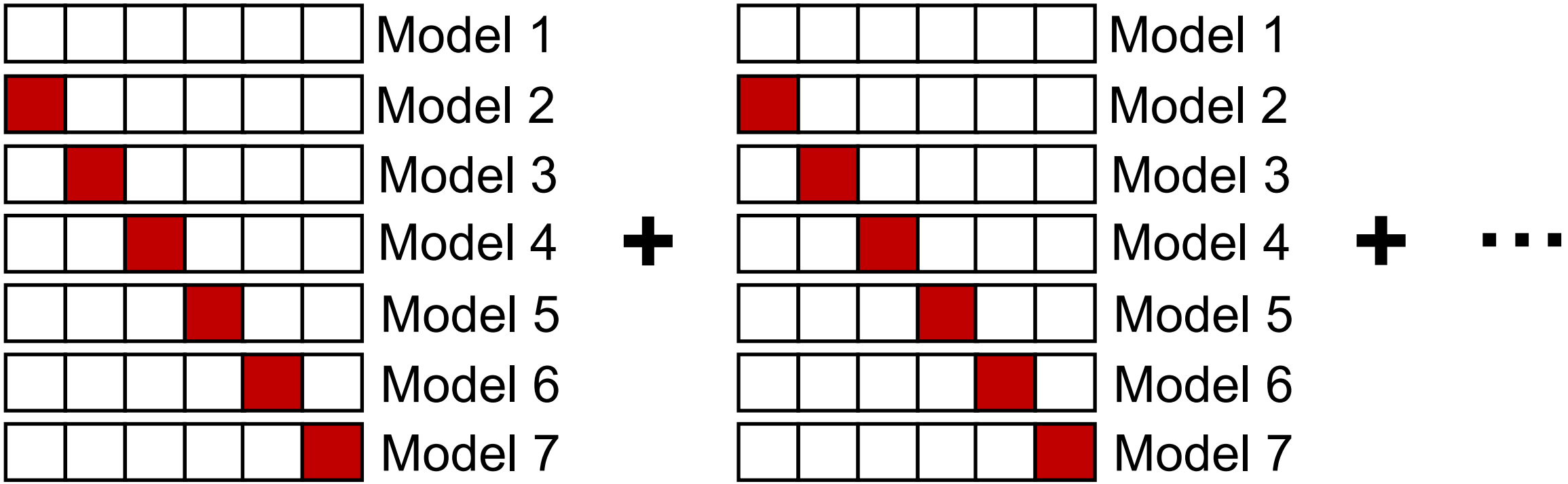
Causal SNP

SuSiE provides an elegant approach for efficient multiple-causal variant fine-mapping



Wang et al. 2020 J R Stat Soc Series B Stat Methodol

SuSiE provides an elegant approach for efficient multiple-causal variant fine-mapping



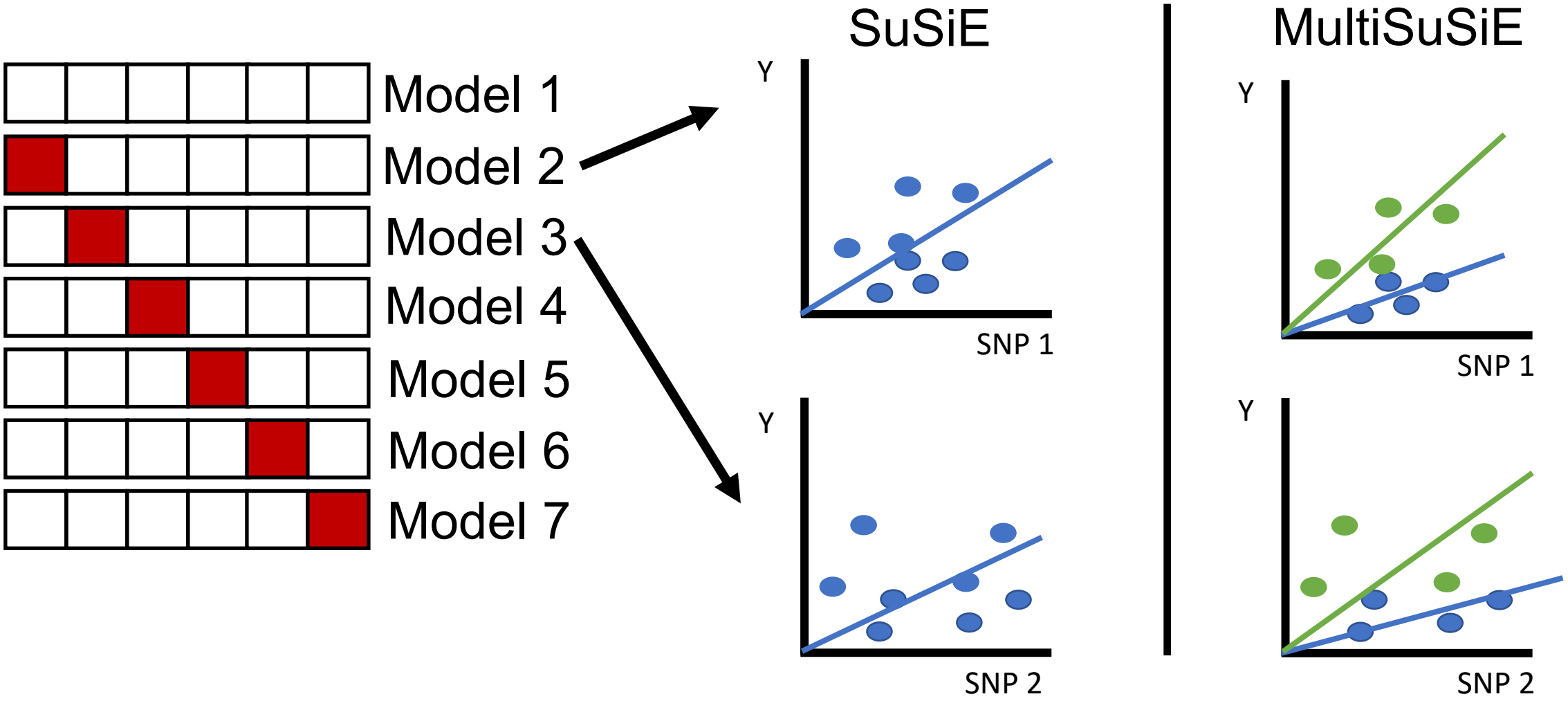
Null SNP Causal SNP

Wang et al. 2020 J R Stat Soc Series B Stat Methodol

Outline

1. Background
2. **Method and Simulations**
3. Results on Real Traits
4. Conclusions and Future Directions

MultiSuSiE extends SuSiE to allow for population-specific effect sizes



MultiSuSiE extends SuSiE to allow for population-specific effect sizes

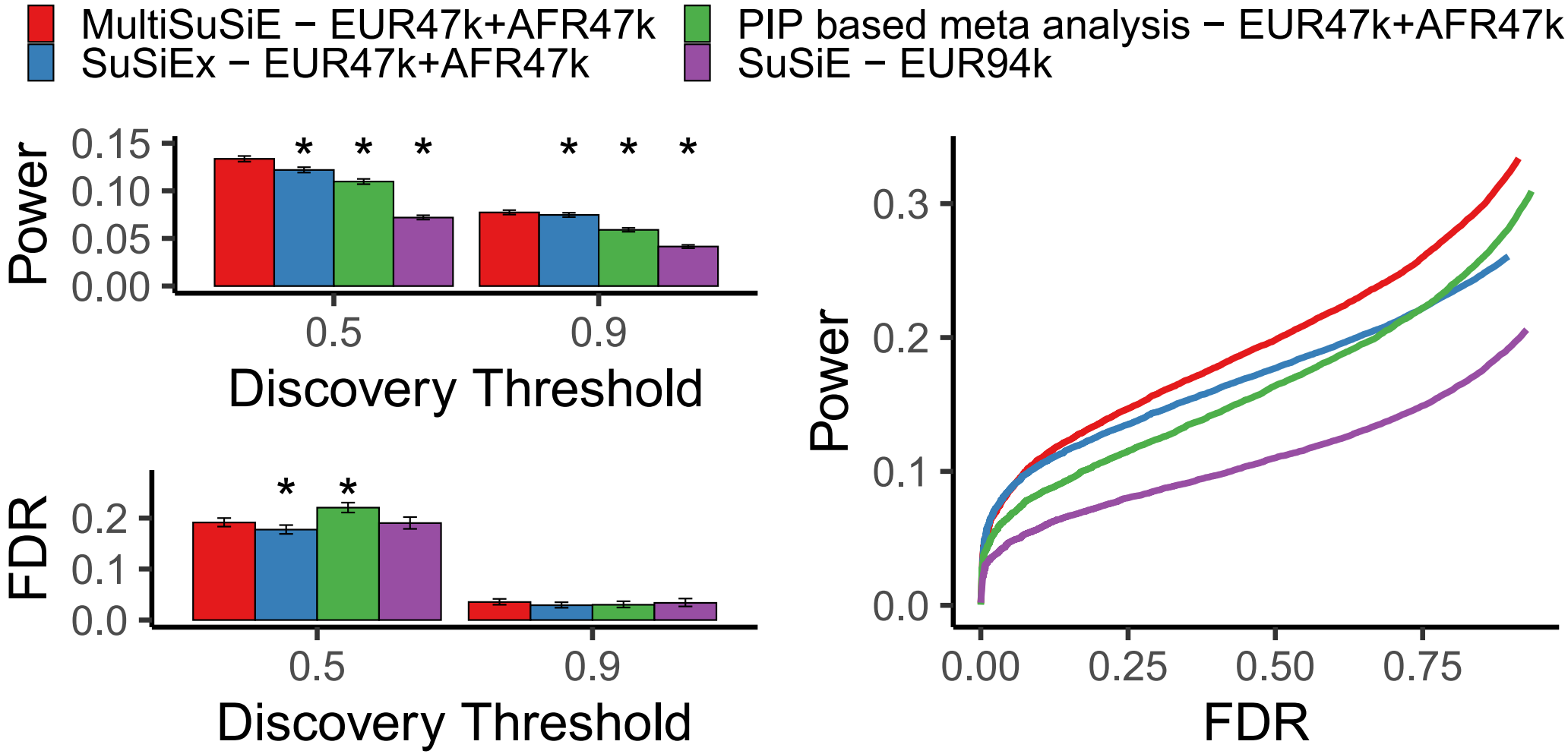
$$Y_k = \sum_{l=1}^{L} X_k \beta_{k,l} \gamma_l + \epsilon_k$$

$$\begin{bmatrix} \beta_{l,1} \\ \beta_{l,2} \end{bmatrix} \sim \mathrm{MVN}_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} v_{l,1}^2 & v_{l,1} v_{l,2} \rho \\ v_{l,1} v_{l,2} \rho & v_{l,2}^2 \end{bmatrix} \right)$$

$$\gamma_l \sim \mathrm{Categorical}_P(\pi)$$

- *k* indexes populations
- *l* indexes causal effects

MultiSuSiE outperforms alternative published methods in simulations

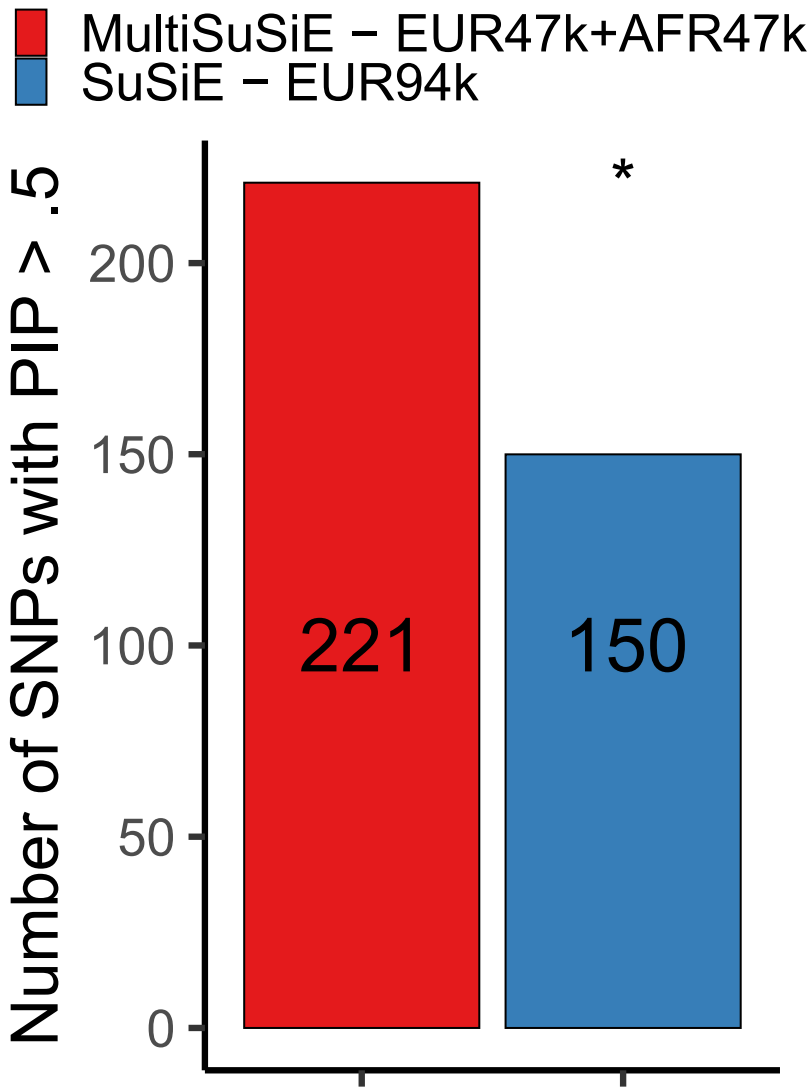


Outline

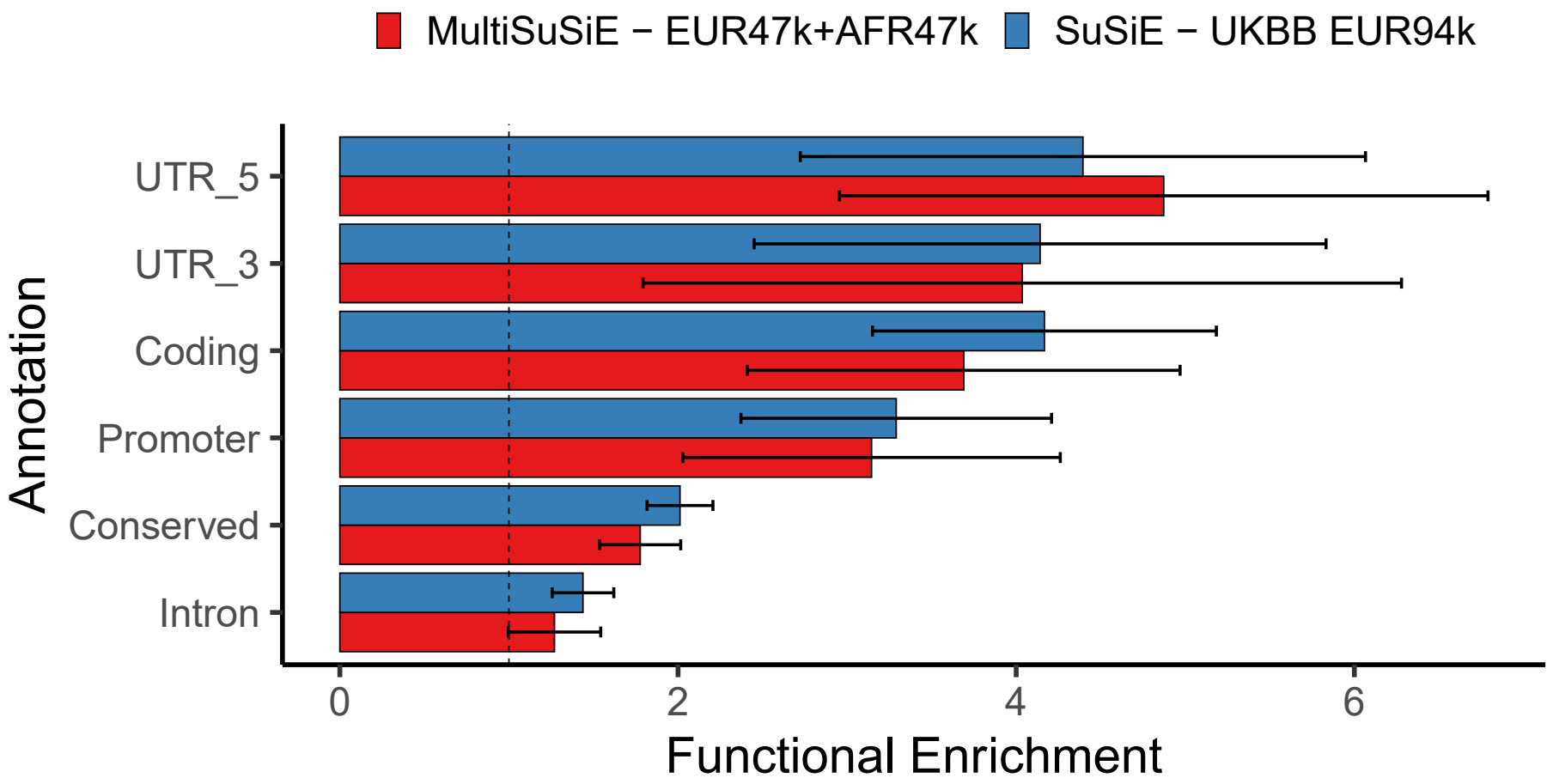
1. Background
2. Method and Simulations
3. **Results on Real Traits**
4. Conclusions and Future Directions

Multi-ancestry fine-mapping using MultiSuSiE outperforms single ancestry fine-mapping at matched sample sizes

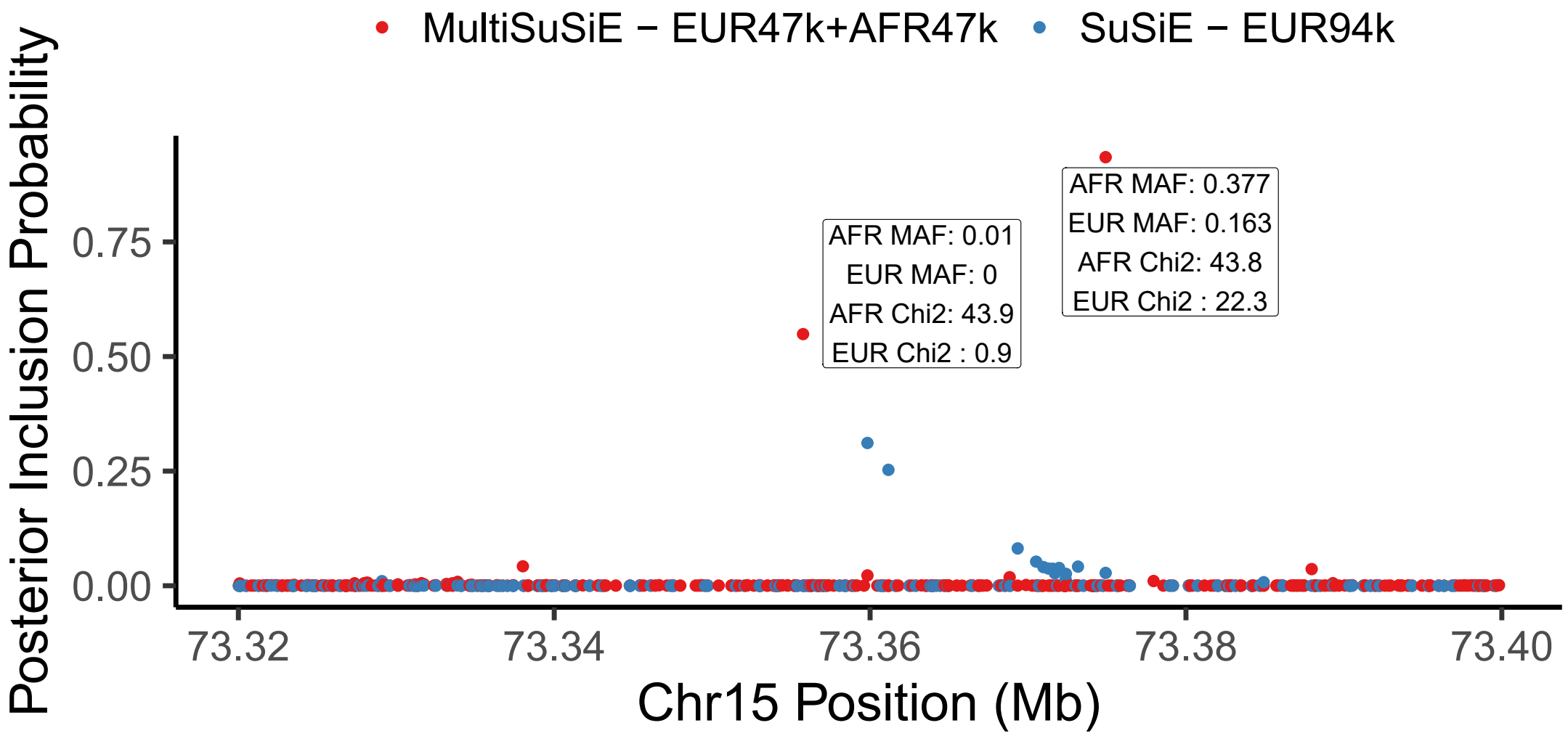
- We fine-mapped 5 anthropomorphic traits using All of Us whole-genome sequencing data
- MultiSuSiE uses 47,000 European-ancestry samples and 47,000 African-ancestry samples
- SuSiE uses 94,000 European-ancestry samples



Causal SNPs identified by MultiSuSiE have similar functional enrichment to those of SuSiE



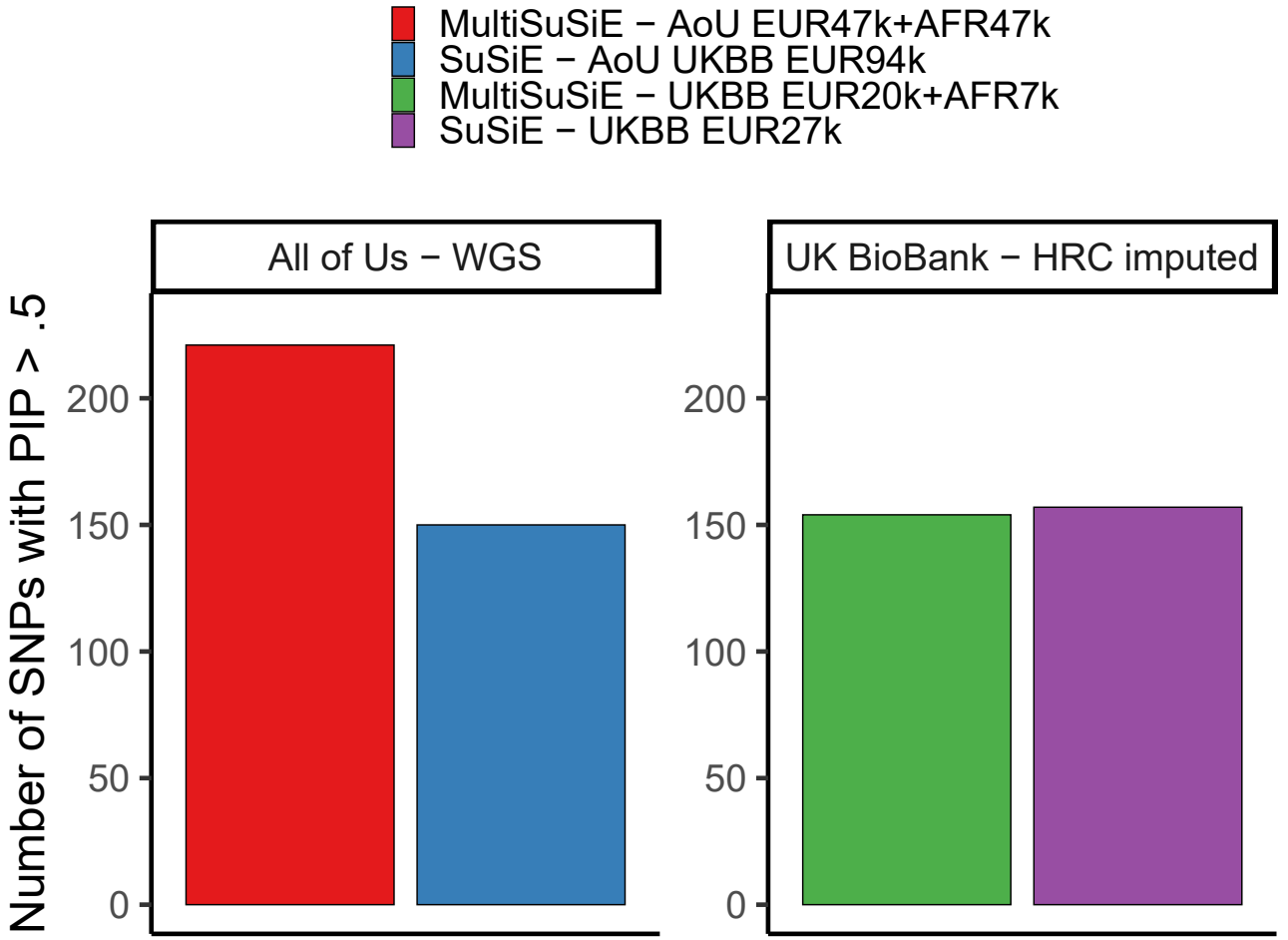
At the HCN4 locus, MultiSuSiE identifies two likely causal variants



Multi-ancestry fine-mapping does outperform single-ancestry using UKBioBank HRC imputed data

We fine-mapped:

- 5 anthropomorphic traits using All of Us whole-genome sequencing data
- 8 highly heritable traits using UKBioBank data



Outline

1. Background
2. Method and Simulations
3. Results on Real Traits
4. **Conclusions and Future Directions**

Conclusions and Future Directions

Conclusions:

- MultiSuSiE enables powerful multi-ancestry fine-mapping
- Multi-ancestry fine-mapping using whole-genome-sequencing data outperforms single-ancestry fine-mapping in simulations and in applications to real traits

Future Directions:

- Include EHR-derived traits with substantial missingness
- Develop recommendations for fine-mapping of binary traits