

Re-evaluate ultrarare variants' contribution to *cis*-heritability of human gene expression

COS 551 / QCB 455 // Final Project

Yushi Tang

Lewis-Sigler Institute for Integrative Genomics, Princeton University

Dec 3, 2019

How to measure heritability (narrow-sense h^2)?

- h^2 : The proportion of phenotypic variance that is due to **additive** genetic effects

$$\mathbf{Y}_{N \times 1} = \mathbf{X}_{N \times c} \mathbf{B}_{c \times 1} + \mathbf{u}_{N \times 1} + \boldsymbol{\varepsilon}_{N \times 1} \quad (\text{eq. 1})$$

• \mathbf{Y} := phenotypes, \mathbf{X} := covariates, \mathbf{B} := effect sizes

• \mathbf{XB} := fixed effects, \mathbf{u} := random effects, $\boldsymbol{\varepsilon}$:= environment effects

nature genetics

Article | Published: 02 September 2019

Ultrarare variants drive substantial *cis* heritability of human gene expression

Ryan D. Hernandez , Lawrence H. Uricchio, Kevin Hartman, Chun Ye, Andrew Dahl & Noah Zaitlen 

Nature Genetics 51, 1349–1355(2019) | Cite this article

7201 Accesses | 2 Citations | 139 Altmetric | Metrics

$$h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2) \quad (\text{eq. 2})$$

• $\mathbf{K}_{N \times N}$:= pedigree relationship matrix

• $\mathbf{u} \xrightarrow{d} \mathcal{N}(0, \sigma_g^2 \mathbf{K})$ | the residual variance due to genetic effects

• $\boldsymbol{\varepsilon} \xrightarrow{d} \mathcal{N}(0, \sigma_e^2 \mathbf{I})$ | the residual variance due to environment effects

• $\mathbf{V} = \mathbb{V}(\mathbf{u} + \boldsymbol{\varepsilon}) = \sigma_g^2 \mathbf{K} + \sigma_e^2 \mathbf{I}$ | total residual variance

• $\mathbf{V} = \mathbf{V}/(\sigma_g^2 + \sigma_e^2) = h^2 \mathbf{K} + (1 - h^2) \mathbf{I}$ | normalized variance

• Y_i := phenotype for sample i after normalization on $\mathbf{Y}_{N \times 1}$

$$\mathbb{E}(Y_i Y_j | K_{ij}) = h^2 K_{ij}, \quad i \neq j \quad (\text{eq. 3})$$

H-E (Haseman-Elston) regression:

$$h^2 = \frac{\sum_{i \neq j} K_{ij} Y_i Y_j}{\sum_{i \neq j} K_{ij}^2} \quad (\text{eq. 4})$$

Haseman & Elston 1972 *Behavior Genetics* | Yang et al. 2010 *Nat Genet*
Visscher et al. 2008 *Nat Rev Genet* | Zaitlen and Kraft 2012 *Hum Genet*

More about h^2 and relationship matrix K

- **H-E regression:**
$$h^2 = \frac{\sum_{i \neq j} K_{ij} Y_i Y_j}{\sum_{i \neq j} K_{ii}^2} \quad (\text{eq. 4})$$
- **Scenario I: related individuals**
 - K measures the proportion of the genome that individuals inherited from a recent common ancestor i.e. identical-by-descent (IBD)
- **Scenario II: unrelated individuals**
 - Replace K with A to measure the proportion of the genome that individuals share the same genotype i.e. identical-by-state (IBS)
 - $$A_{ij} = \frac{1}{M} \sum_k \frac{(g_{ki} - 2p_k)(g_{kj} - 2p_k)}{2p_k(1-p_k)}$$
 - Now we are actually estimate heritability explained by genotyped SNPs (h_g^2)
- $$h_g^2 = \frac{\sum_{i \neq j} A_{ij} Y_i Y_j}{\sum_{i \neq j} A_{ii}^2} \quad (\text{eq. 5})$$
- **Decomposing genetic effects from genotyped SNPs**
 - $\mathbf{u} = \mathbf{G}_{N \times M} \boldsymbol{\beta}_{M \times 1}$
 - $$Y_i = \sum_{k=1}^c X_{ik} b_k + \sum_{j=1}^M g_{ij} \beta_j + \varepsilon_i \quad (\text{eq. 6})$$
- **Narrow-sense heritability --> partitioning into SNP bins**
 - $$h_k^2 = \sigma_k^2 / (\sigma_g^2 + \sigma_e^2)$$
 - σ_k^2 := the genetic variance contributed by SNPs in the k^{th} partition
- **Idea: will partitioning method improve the performance of heritability estimation?**

Haseman & Elston 1972 *Behavior Genetics* | Yang et al. 2010 *Nat Genet*
Powell et al. 2010 *Nat Rev Genet* | Speed & Balding 2015 *Nat Rev Genet*
Hernandez et al. 2019 *Nat Genet*

Partitioning alleles alleviates the problem of over-estimation for h^2

- **Study cohort: GEUVADIS**
 - 375 individuals with 4 European ancestries
 - 360 unrelated samples
 - Overlap with 1KGP
- **Partitioning method**
 - Minor allele freq. (MAF)
 - 20 bins of SNPs resulted in the smallest bias
- **Bias direction**
 - Upward-bias for high true h^2
 - Downward-bias for small one

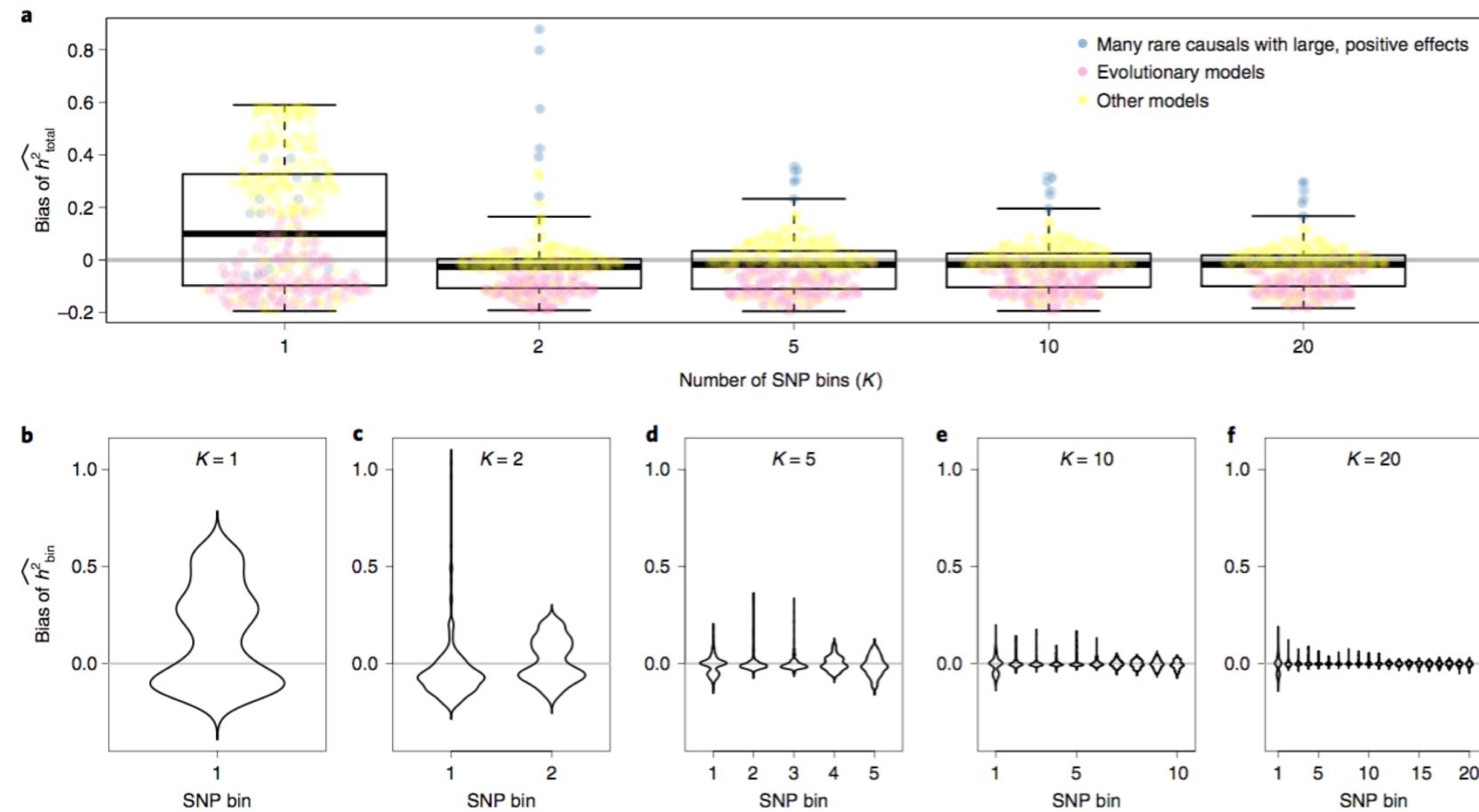
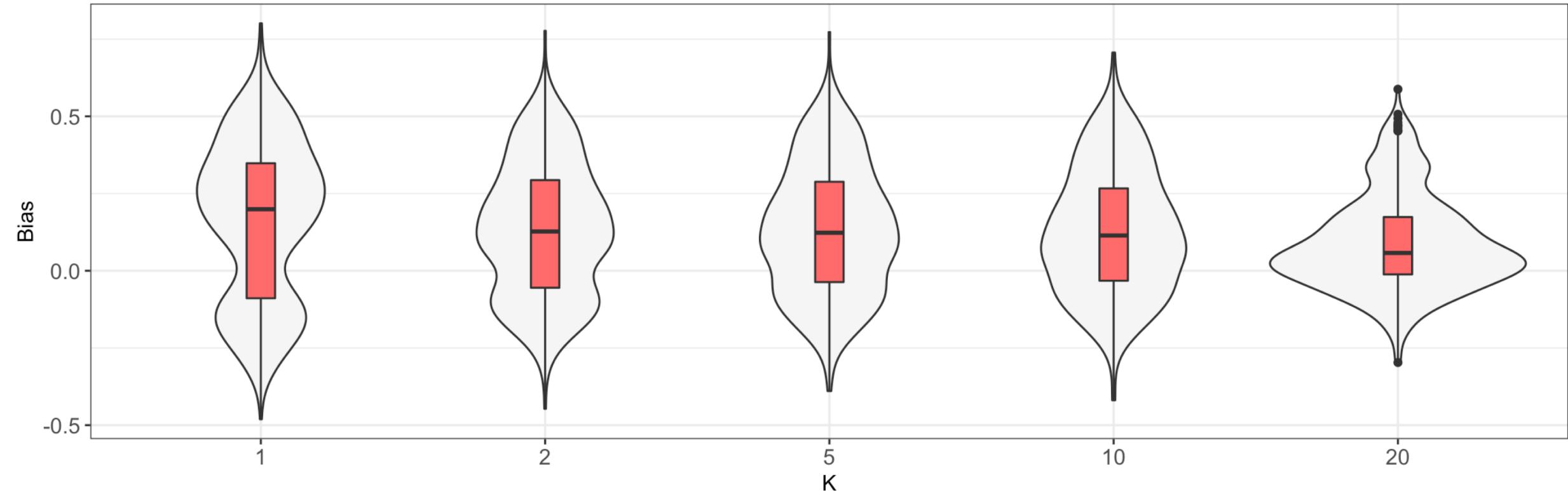


Fig. 1 | Simulation results. Across a broad range of parameters, the accuracy of heritability inference improves as the number of SNP bins (partitioned by MAF) increases. **a**, Mean bias of total heritability (inferred-true) for different numbers of SNP bins (K), where each point represents the mean of 500 simulations for different parameters, and box plot summarizing the bias distribution across all parameters (indicating the median, upper and lower quartile and twice the interquartile range). **b-f**, Distribution of average bias across simulated parameters for each SNP bin, showing that both mean and variance of the bias decrease as K increases ($n=500$ simulations in each plot).

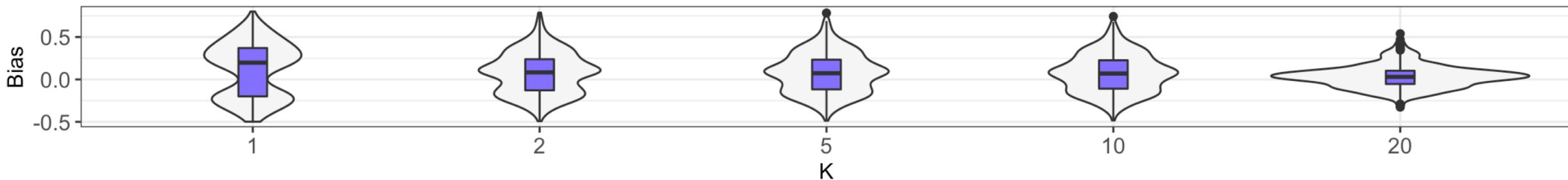
Partitioning method works for CEU cohort in HapMap 3

Simulation results for bias estimation for CEU (N = 112)

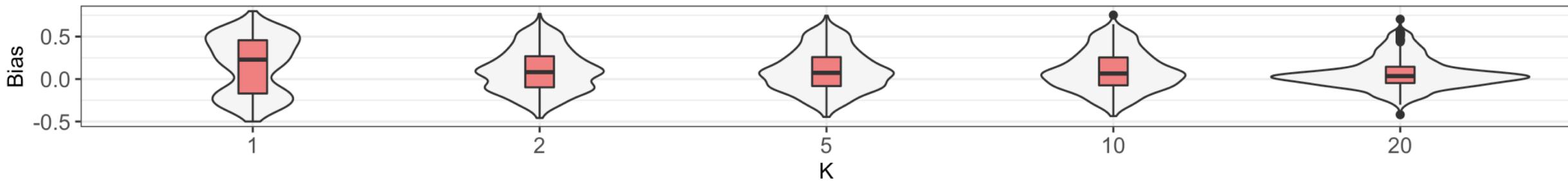


Partitioning method works for several other cohorts in HapMap 3

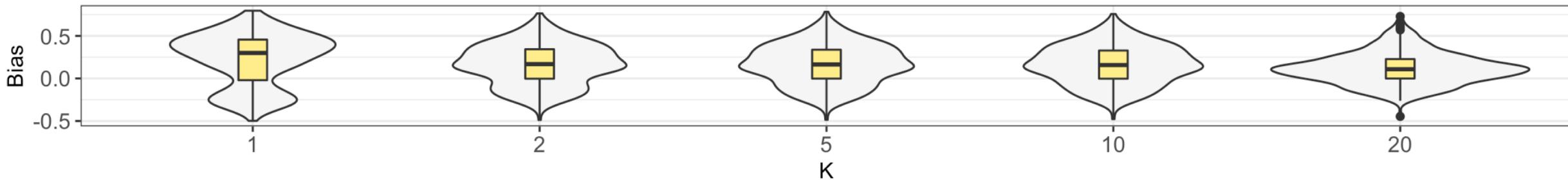
Simulation results for bias estimation for CHB ($N = 84$)



Simulation results for bias estimation for JPT ($N = 86$)

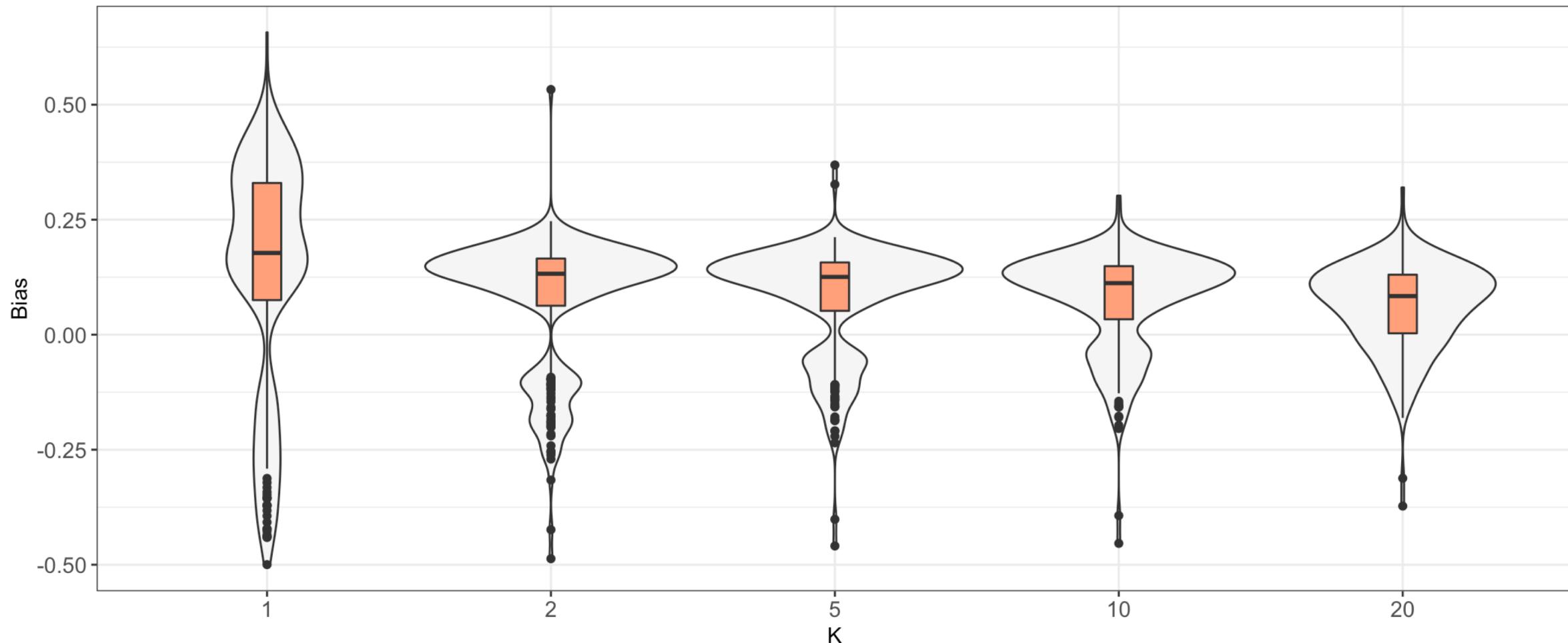


Simulation results for bias estimation for YRI ($N = 113$)



Partitioning method is not working well for cohort with complex population structure

Simulation results for bias estimation for HapMap3 ($N = 988$)



Rare variants dominate the genetic architecture of human gene expression?

- Singletons represent ~25% of the total heritability, dominating the estimates from other MAF bins
- ~90% singleton heritability is contributed by variants with extremely low population MAF inferred from gnomAD

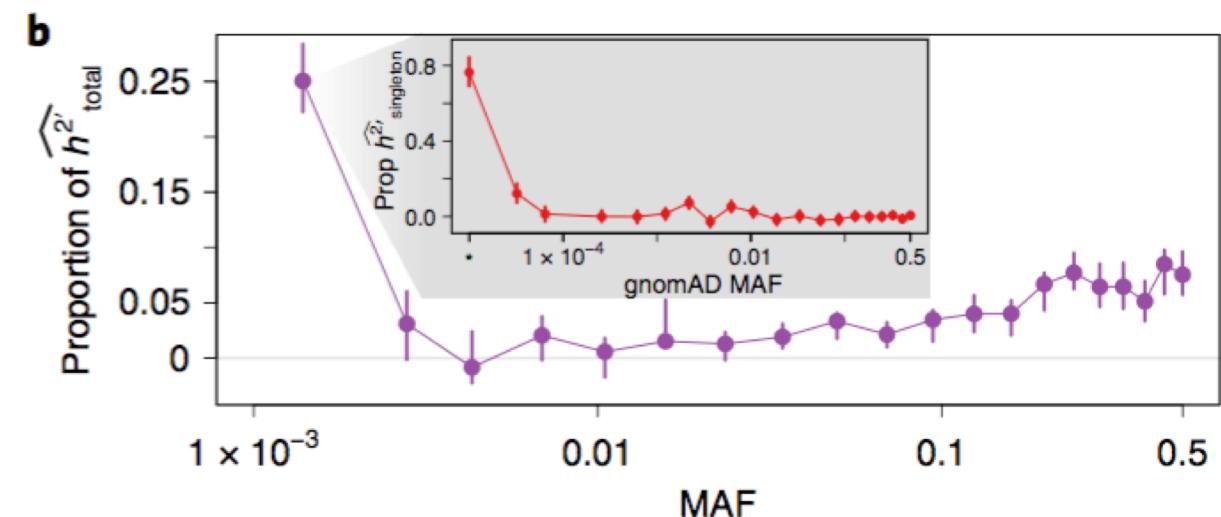
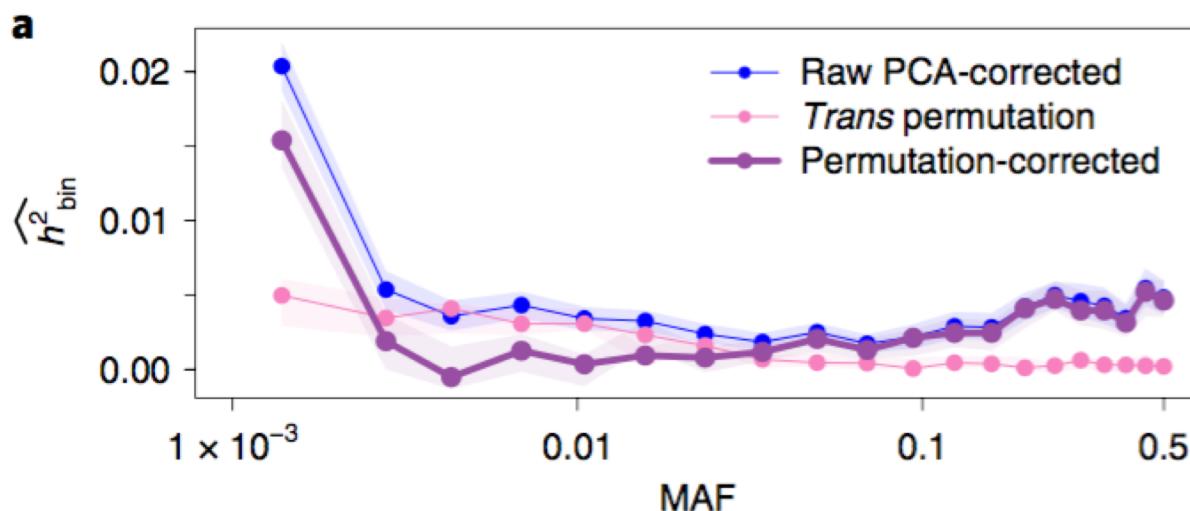


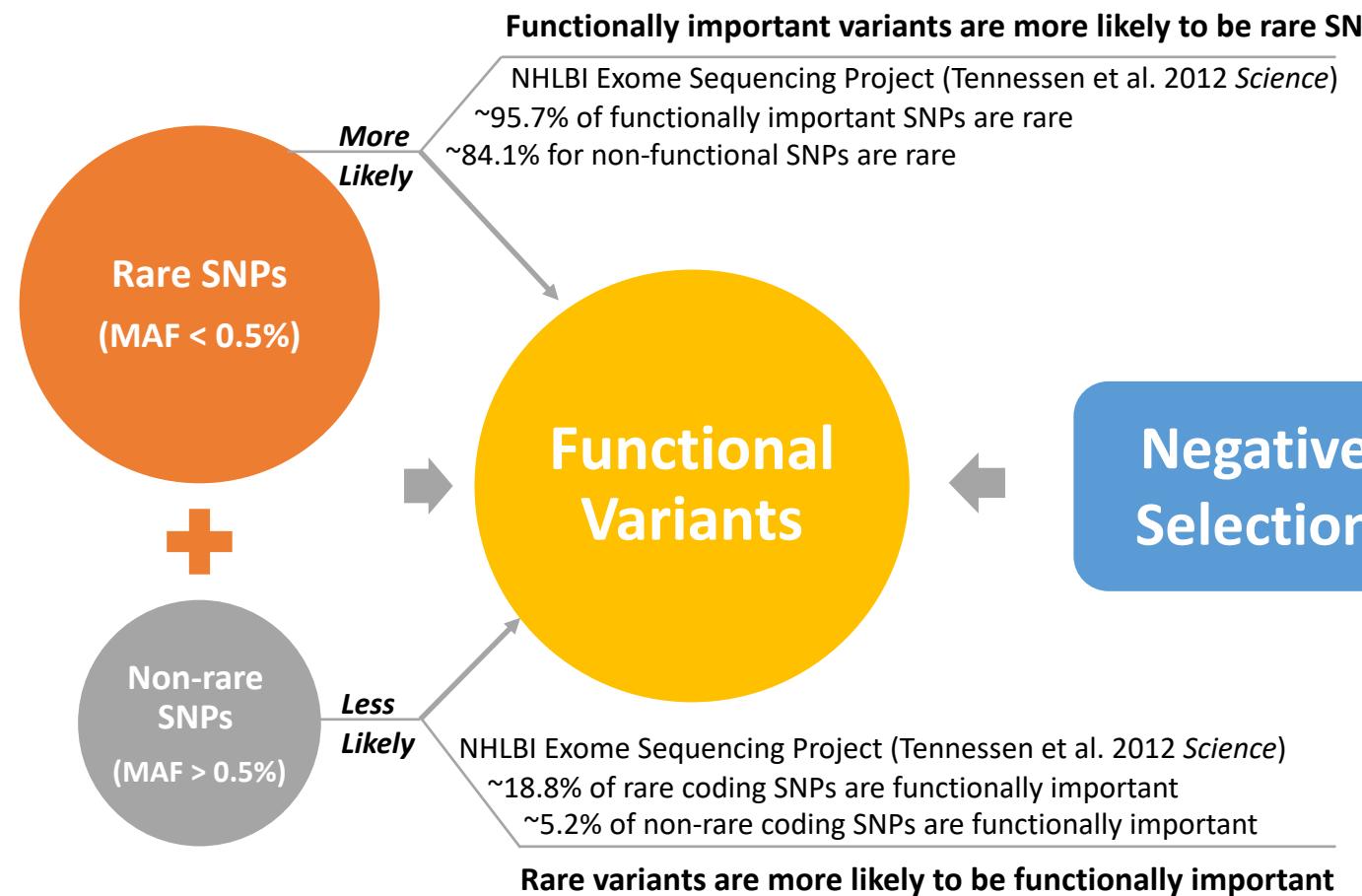
Fig. 2 | Partitioning heritability. **a**, The average h^2 estimated across genes in each SNPs bin. Highest h^2 was detected in the first bin. **b**, Proportion of h^2 explained by each MAF bin. Rare variants dominate the estimates from other MAF bins.

Hernandez et al. 2019 *Nat Genet*

What drives the ultrarare variant heritability? Negative selection...

- **Negative selection:** the negative pressure on allele frequencies of mutations that reduce fitness

- Effects: pushing down the MAF of functional SNPs --> rare variants



- Model the power of negative selection

- β : effect size of causal SNPs

- $h^2 := \beta^2 p_{maf} (1 - p_{maf})^2$

- NULL: Naturally

- $\mathbb{V}(\beta) \rightarrow_p [p_{maf}(1 - p_{maf})]^0$

- $\mathbb{E}(h^2) \rightarrow_p [p_{maf}(1 - p_{maf})]^{1+0}$

- Alternative: Selection

- $\mathbb{V}(\beta) \rightarrow_p [p_{maf}(1 - p_{maf})]^\alpha$

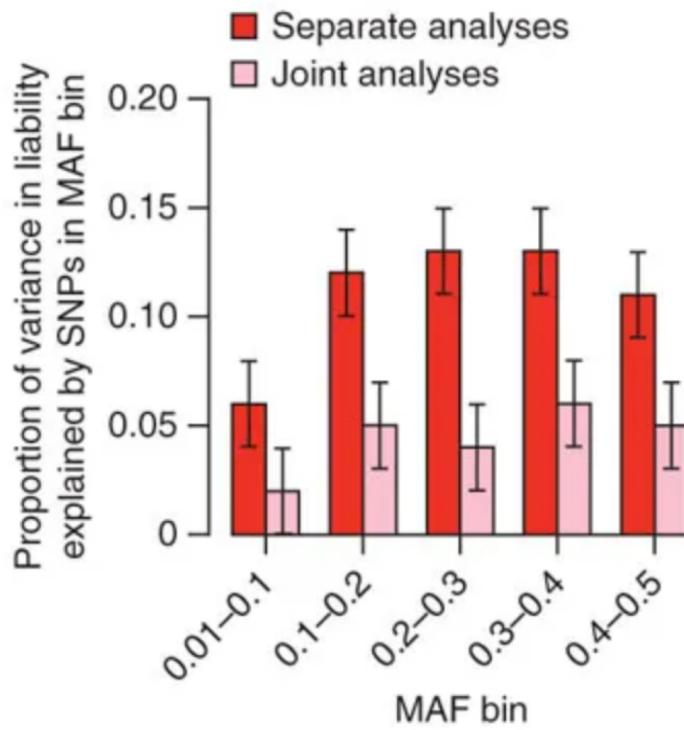
- $\mathbb{E}(h^2) \rightarrow_p [p_{maf}(1 - p_{maf})]^{1+\alpha}$

- $\alpha < 0$

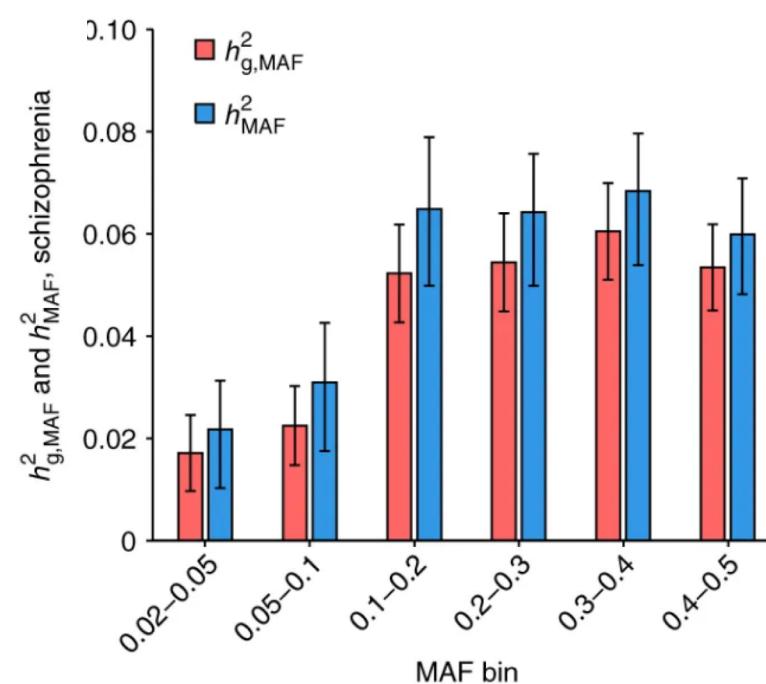
Schoech et al. 2019 *Nat Commun* | Speed et al. 2012 *Am J Hum Genet*
Eyre-Walker 2010 *PNAS*

How much heritability could be explained by rare/ultrarare variants might depend on negative selection

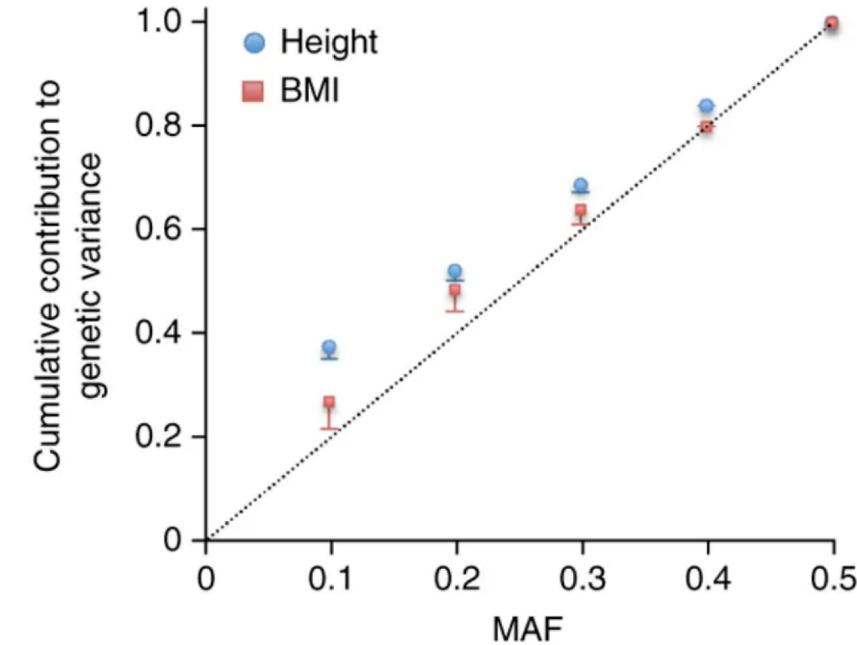
- Contribution of rare/ultrarare SNPs also varies across different traits



Schizophrenia (9,087 cases / 12,171 controls)
Lee et al. 2012 *Nat Genet*



Schizophrenia (22,177 cases / 27,629 controls)
Loh et al. 2012 *Nat Genet*



BMI and Height (44,126 samples)
Yang et al. 2015 *Nat Genet*

FIN

Thanks for your attention!

Yushi Tang

Lewis-Sigler Institute for Integrative Genomics, Princeton University

Dec 3, 2019