Population-specific causal disease effect sizes in functionally important regions impacted by selection

ASHG 2019

October 17, 2019
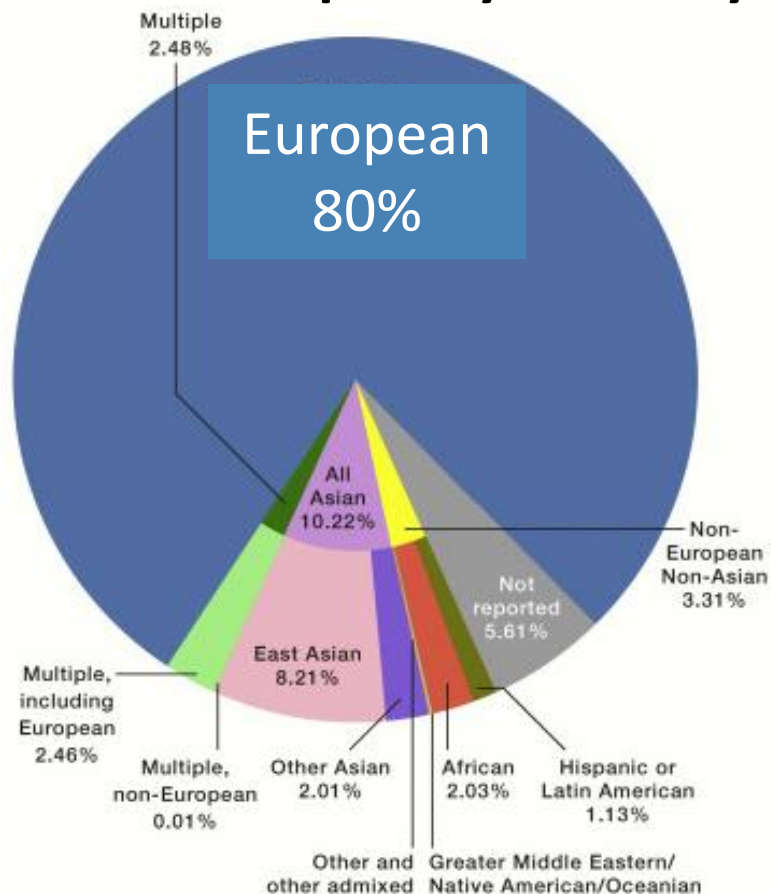
Huwenbo Shi

Alkes Price Group

Harvard University & Broad Institute

Most GWAS samples are still European

GWAS samples by ethnicity

Multiple 2.48%

European 80%

All Asian 10.22%

Non-European Non-Asian 3.31%

Not reported 5.61%

Multiple, including European 2.46%

East Asian 8.21%

Multiple, non-European 0.01%

Other Asian 2.01%

African 2.03%

Hispanic or Latin American 1.13%

Other and other admixed

Greater Middle Eastern/ Native American/Oceanian

non-European 20%

(Sirugo, Williams, Tishkoff 2019 *Cell*)
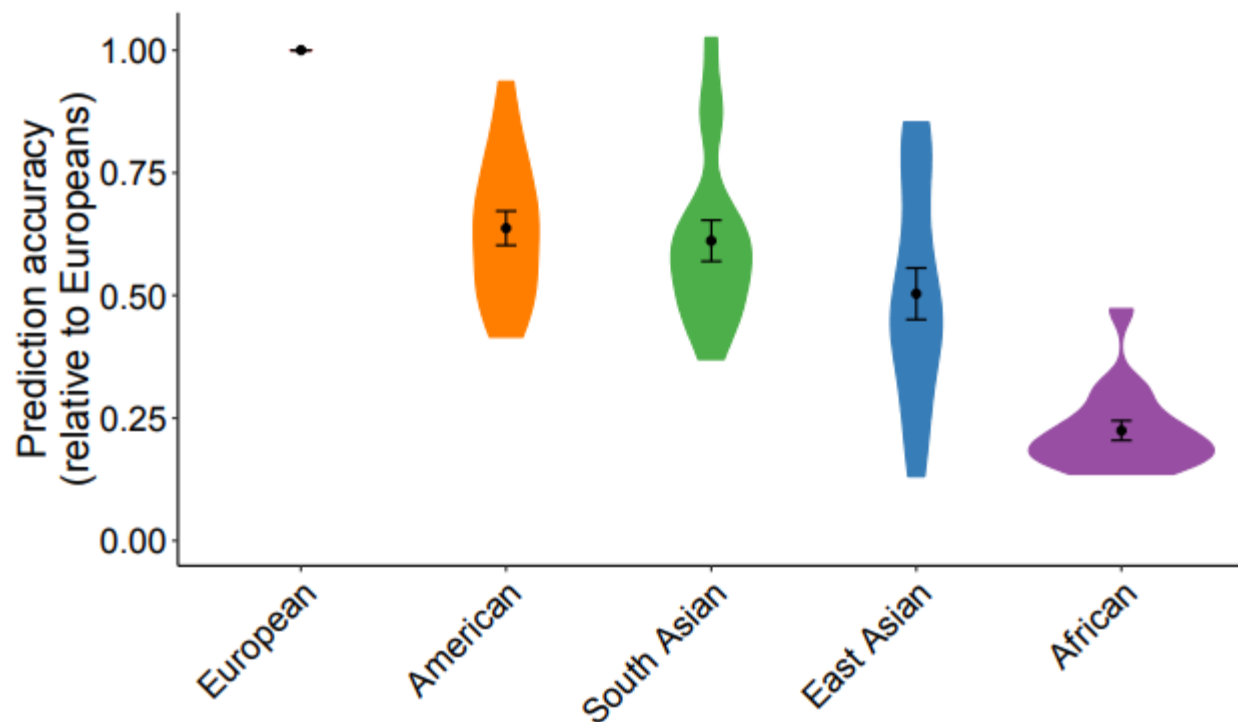
GWAS is becoming more diverse

Number of non-European GWAS samples

- 2009:   0.1 million

- 2016:   7.0 million

70x

(Popejoy & Fullerton 2016 *Nature*)

Trans-ethnic genetic risk prediction has limited accuracy

Relative to Europeans, trans-ethnic genetic risk prediction is less accurate for non-Europeans

(Martin et al. 2019 *Nat Genet*)

Factors impacting prediction accuracy

- Different LD patterns

- Population-specific variants

......

(Marquez-Luna et al. 2017 *Genet Epidemiol*)

- **Different causal effect sizes?**

Trans-ethnic genetic correlations are less than 1

- **Schizophrenia**: **0.6** (AFR American – EUR)
  (de Candia et al. 2013 *AJHG*)

- **Prostate cancer**: **0.9** (EAS – EUR)
  (Mancuso et al. 2016 *Nat Genet*) **0.6** (AFR American – EUR)

- **Rheumatoid arthritis**: **0.5** (EAS – EUR)
  (Brown et al. 2016 *AJHG;*
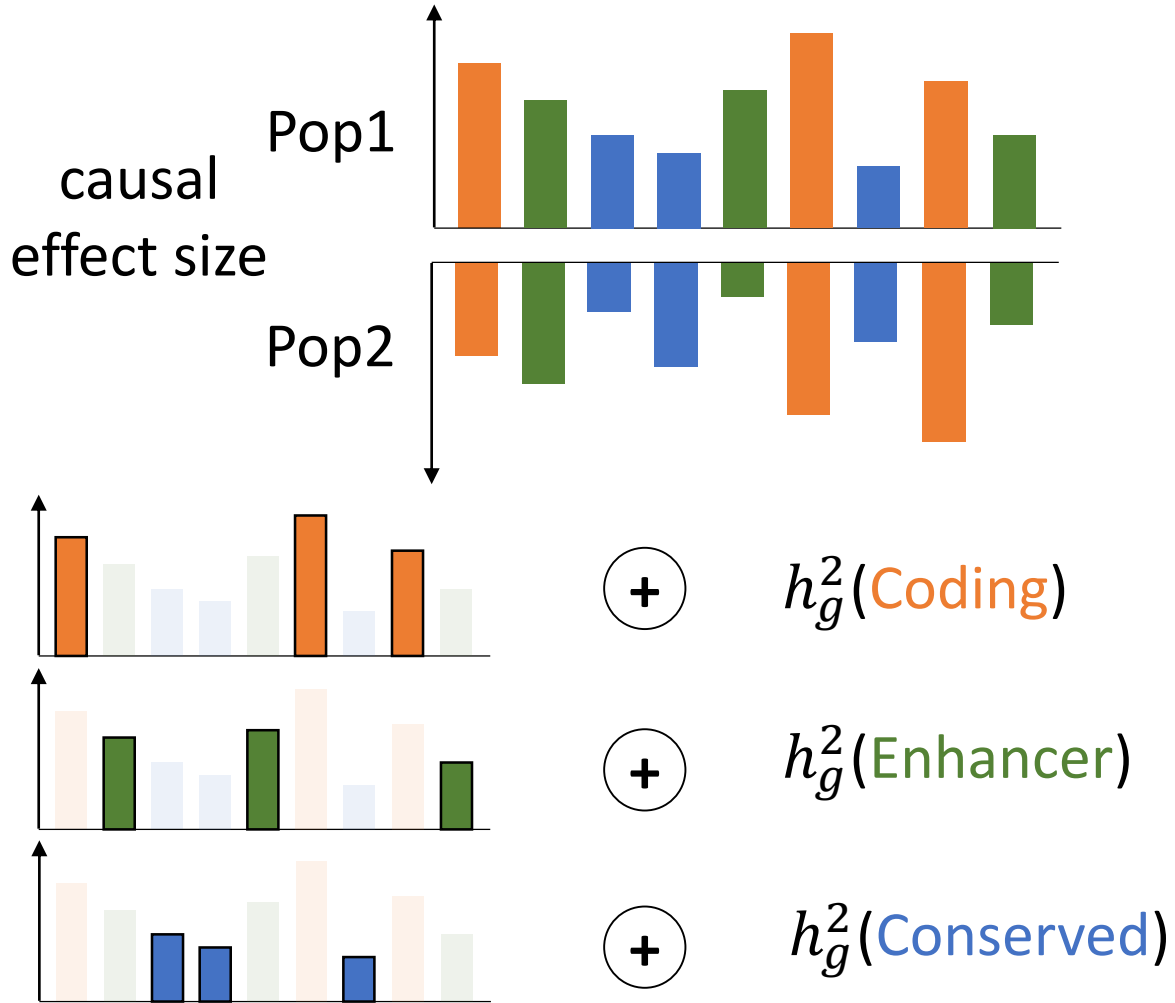 Galinsky et al. 2019 *Genet Epidemiol*)

- **Average across 22 traits:** **0.8** (EAS – EUR)
  (Martin et al. 2019 *Nat Genet*)

Why is trans-ethnic genetic correlation imperfect?

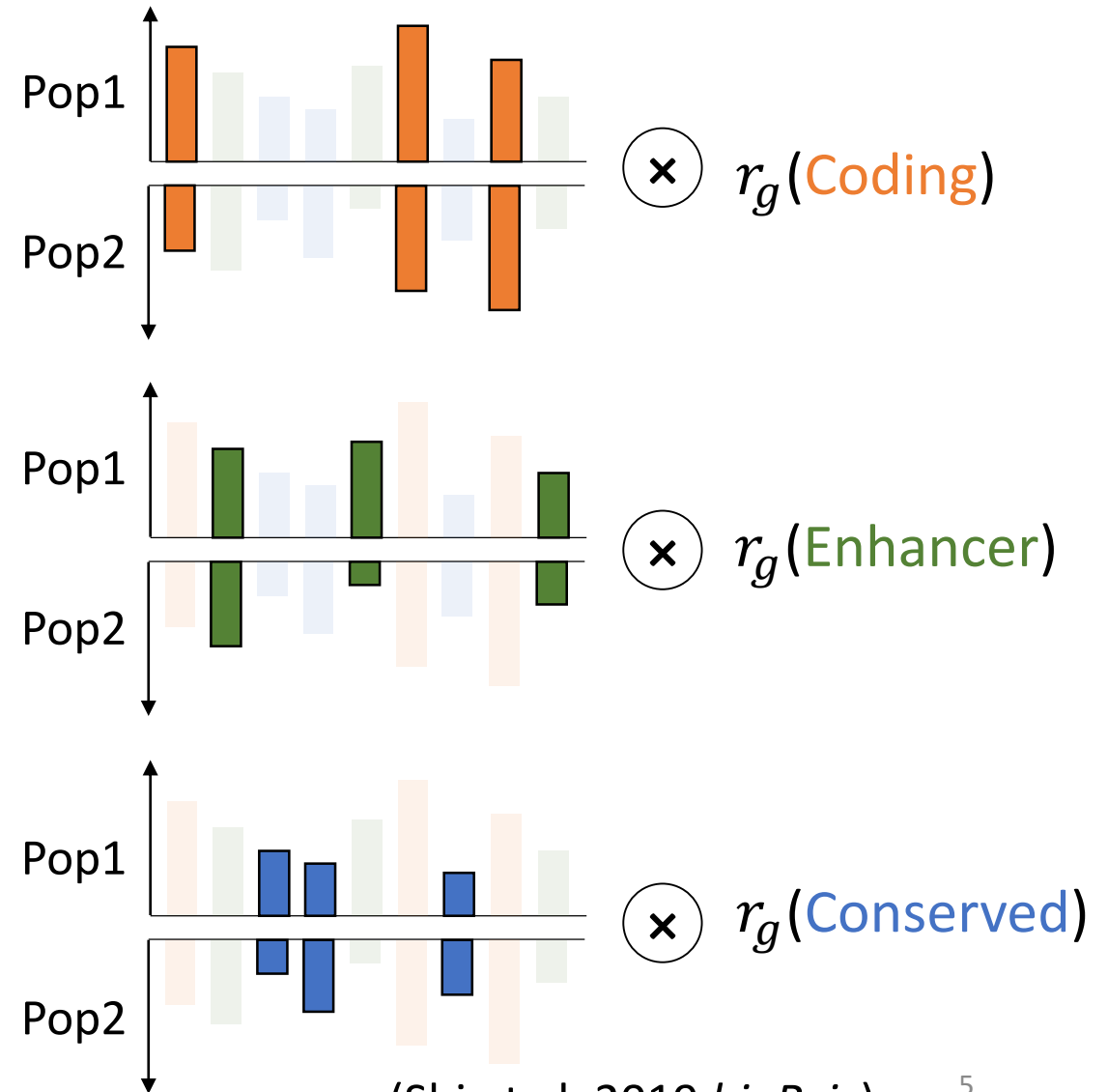# Stratifying trans-ethnic genetic correlation

Stratifying heritability

Stratifying trans-ethnic genetic correlation (r_g)

(Finucane et al. 2015 *Nat Genet*)

(Shi et al. 2019 *bioRxiv*)

Outline

- Method to stratify trans-ethnic genetic correlation (S-LDXR)

- Analysis of 30 complex diseases and traits

- Interpretation: potential role of positive selection

Outline

- **Method to stratify trans-ethnic genetic correlation (S-LDXR)**

- Analysis of 30 complex diseases and traits

- Interpretation: potential role of positive selection
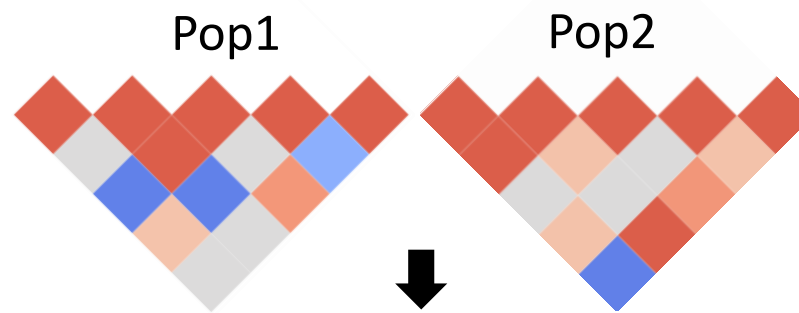
S-LDXR stratifies *squared* trans-ethnic genetic correlation from GWAS summary statistics
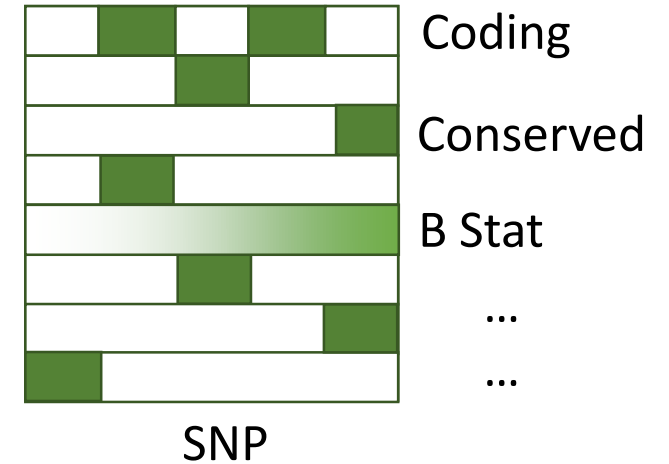
GWAS summary statistics
(in 2 populations)
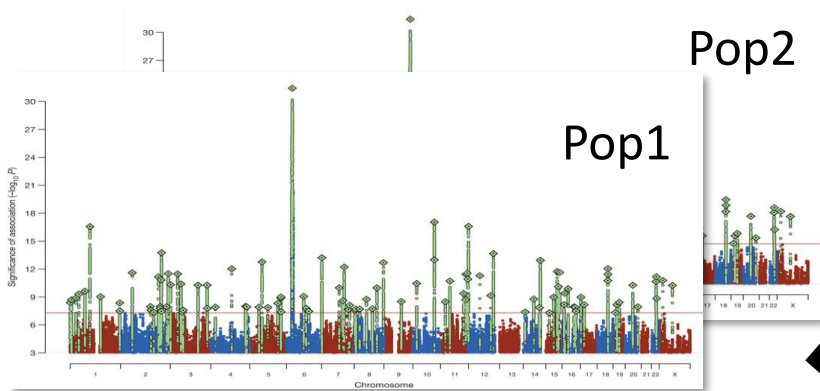
reference LD (1000 GP)
(in 2 populations)

Genomic annotations
(binary and continuous)

PGC 2014 *Nature*

Pop1   Pop2

Coding

Conserved

B Stat

...

...

SNP

**S-LDXR with baseline-LD-X model
(62 genomic annotations)**

S-LDXR models per-allele effect sizes, accounting for MAF differences, different from S-LDSC

use shrinkage to reduce noise

enrichment of trans-ethnic r_g^2 in annotation C

$$\lambda^2(C) = \frac{r_g^2(C)}{r_g^2}$$

(Finucane et al. 2015 *Nat Genet;* Gazal et al. 2017 *Nat Genet;* Gazal et al. 2019 *Nat Genet*)

Product of Z-scores is higher for SNPs tagging annotations enriched for trans-ethnic genetic covariance
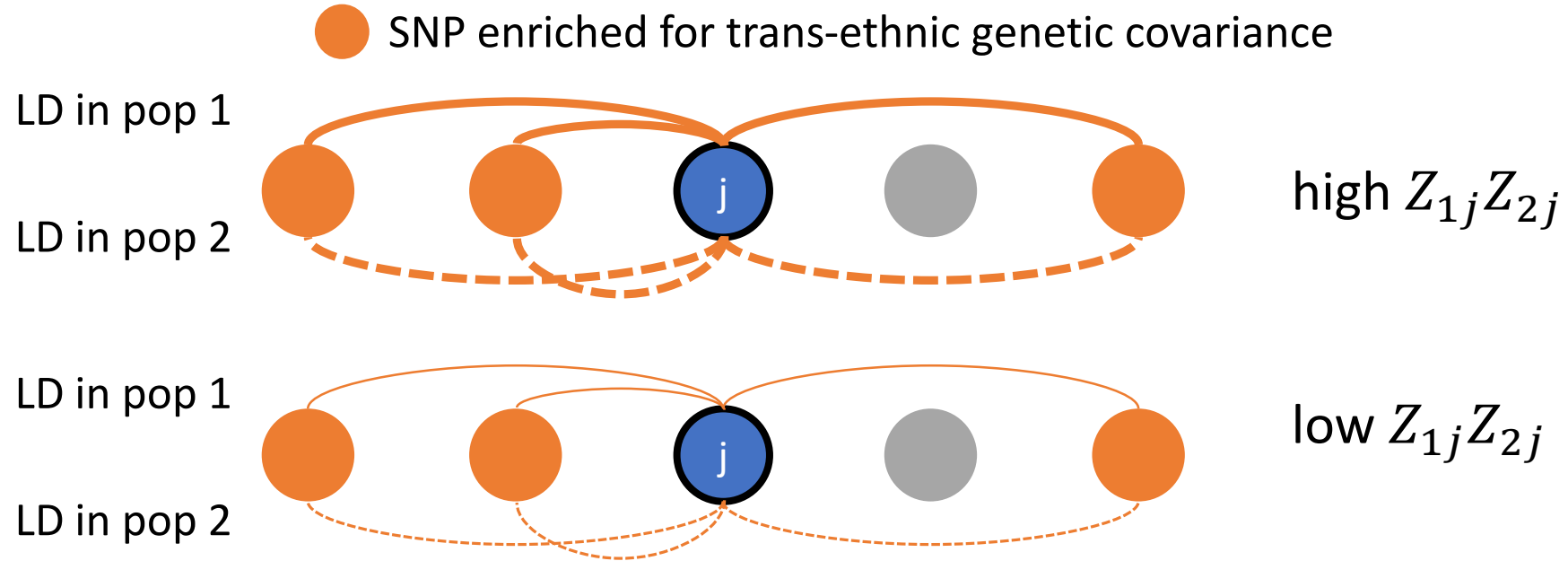
$$E\left[Z_{1j}Z_{2j}\right] = \sqrt{N_1 N_2} \sum_C l_{\times j}(C)\theta_C$$
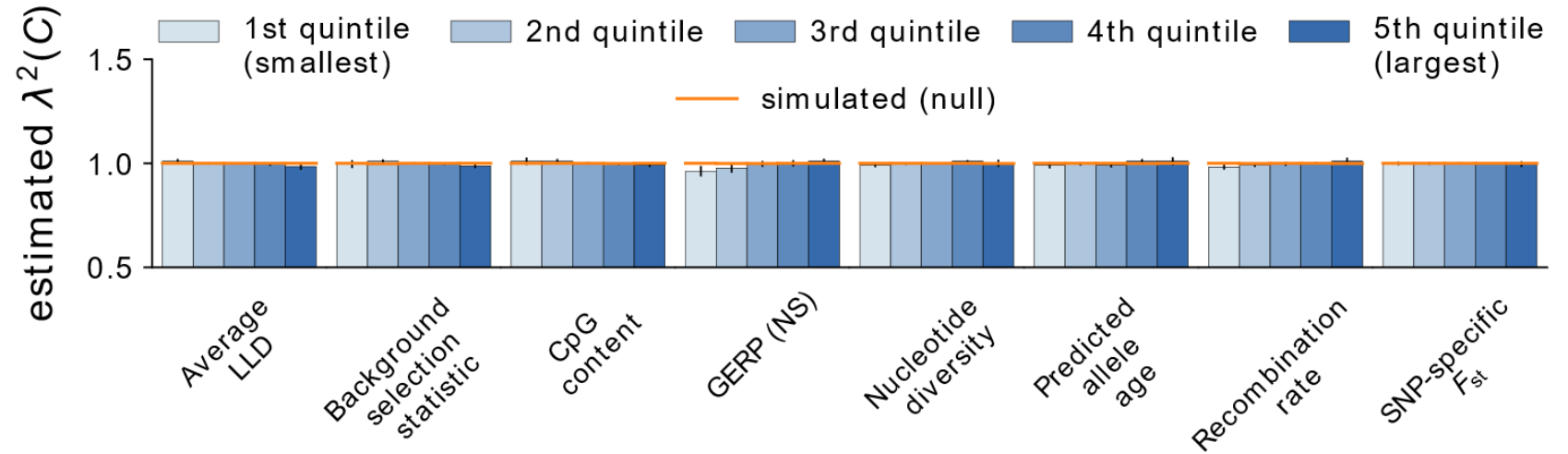
GWAS sample sizes

net contribution of annotation C

Z-score of SNP j in population 1 and 2

trans-ethnic LD score

(Finucane et al. 2015 *Nat Genet;* Gazal et al. 2017 *Nat Genet;* Gazal et al. 2019 *Nat Genet*)
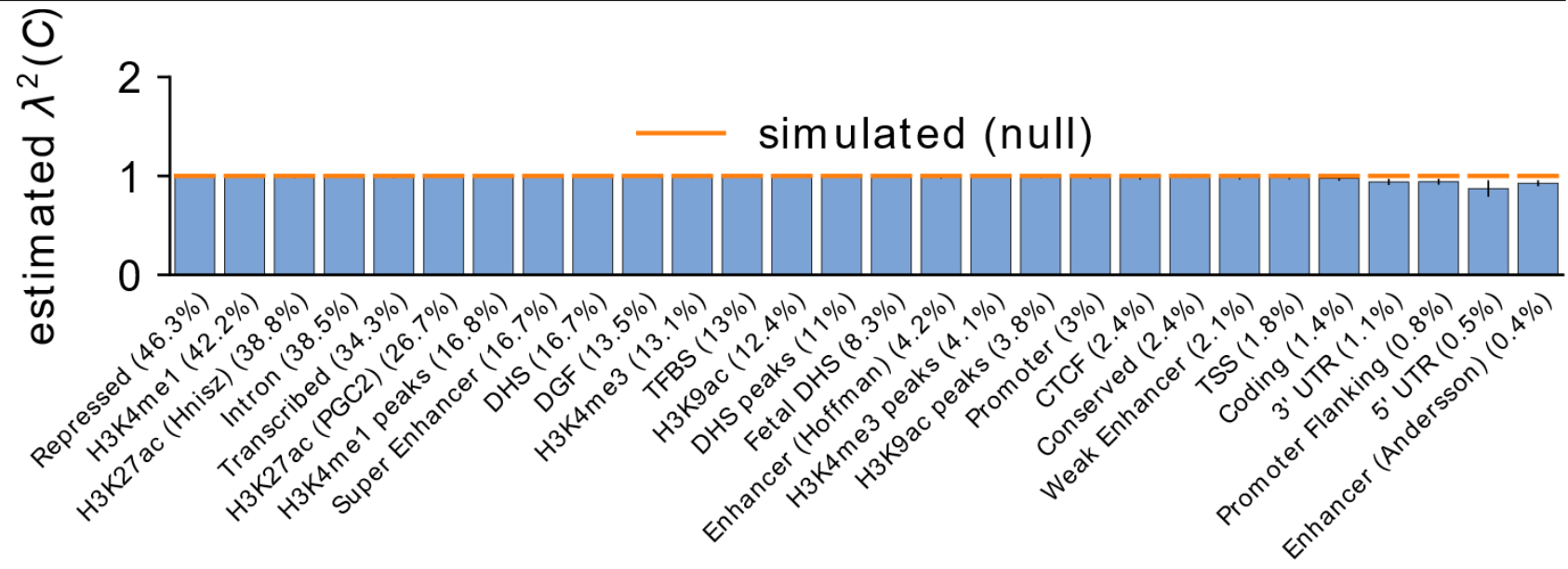
S-LDXR is unbiased in null simulations

simulated $\lambda^2(C) = 1$ for all annotations

$\lambda^2(C)$: enrichment of squared trans-ethnic genetic correlation in annotation C

- ## Quintiles of continuous-valued annotations
  (Gazal et al. 2017 *Nat Genet*)

- ## 28 binary functional annotations
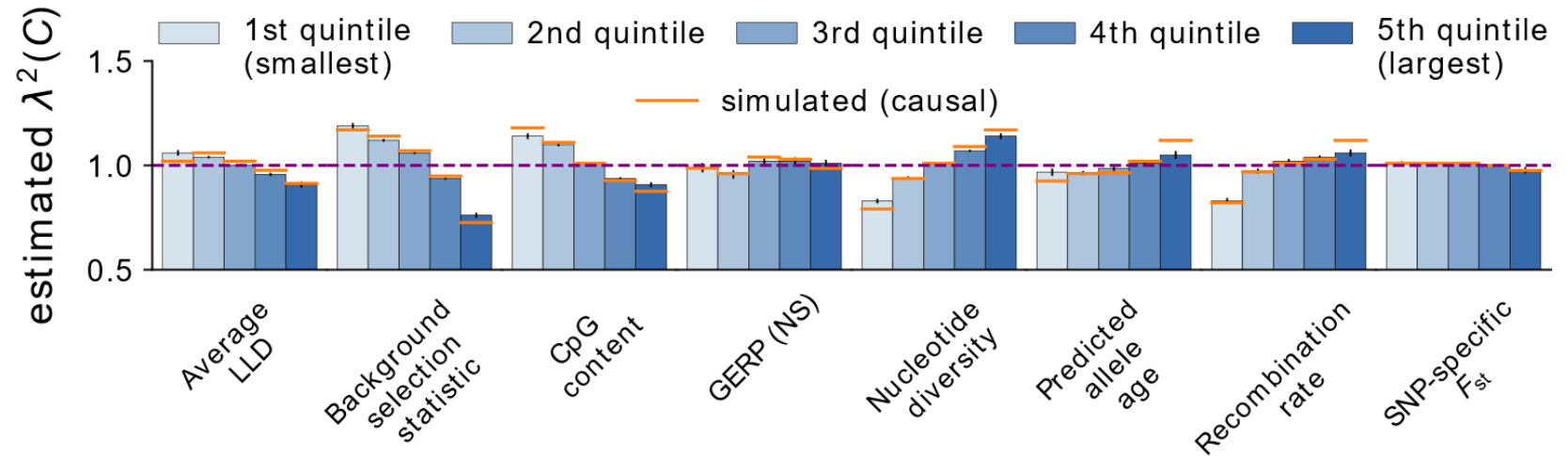 (Finucane et al. 2015 *Nat Genet*)

S-LDXR is conservative in causal simulations

simulated $\lambda^2(C)$ informed by real traits

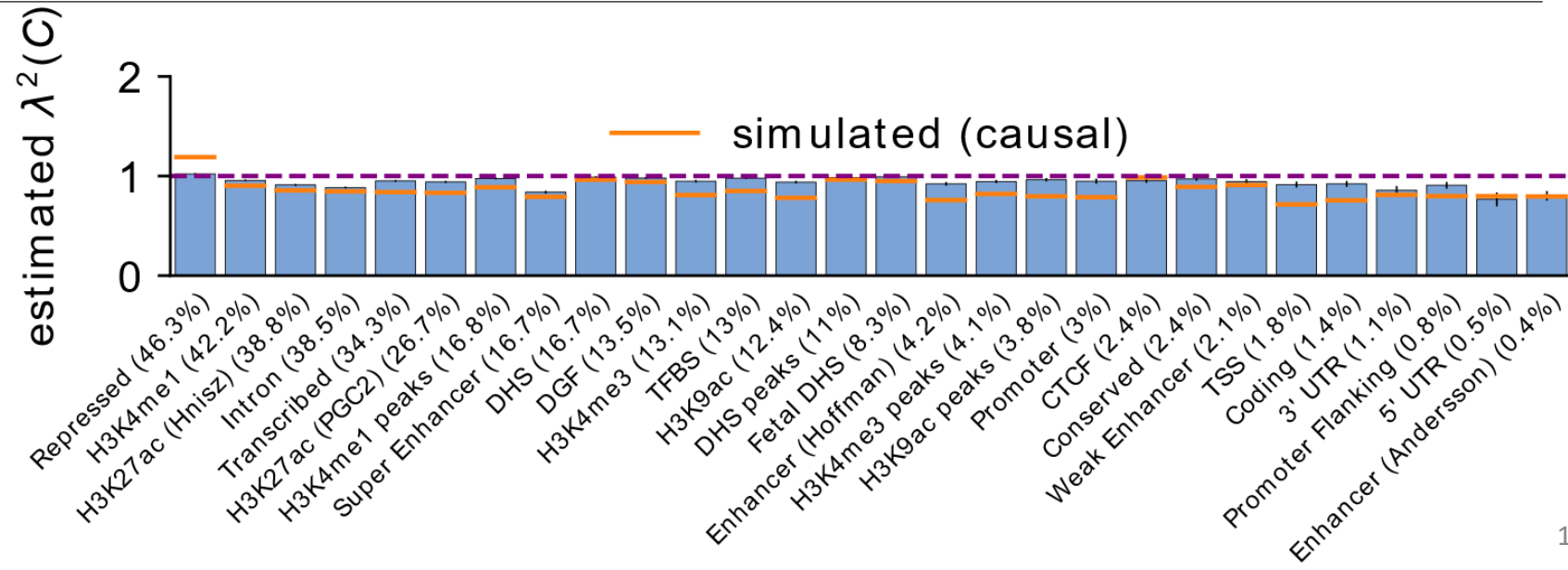$\lambda^2(C)$: enrichment of squared trans-ethnic genetic correlation in annotation C

- ## Quintiles of continuous-valued annotations
  (Gazal et al. 2017 *Nat Genet*)

  estimates are shrunk towards the null: $\lambda^2(C)=1$

- ## 28 binary functional annotations
  (Finucane et al. 2015 *Nat Genet*)

  estimates are shrunk towards the null: $\lambda^2(C)=1$

Outline

- Method to stratify trans-ethnic genetic correlation (S-LDXR)

- **Analysis of 30 complex diseases and traits**

- Interpretation: potential role of positive selection

Analysis of 30 complex diseases and traits

- We obtained GWAS Summary statistics from

**East Asian
(N ≈ 93K)**   Biobank Japan   (Kanai et al. 2018 *Nat Genet*)
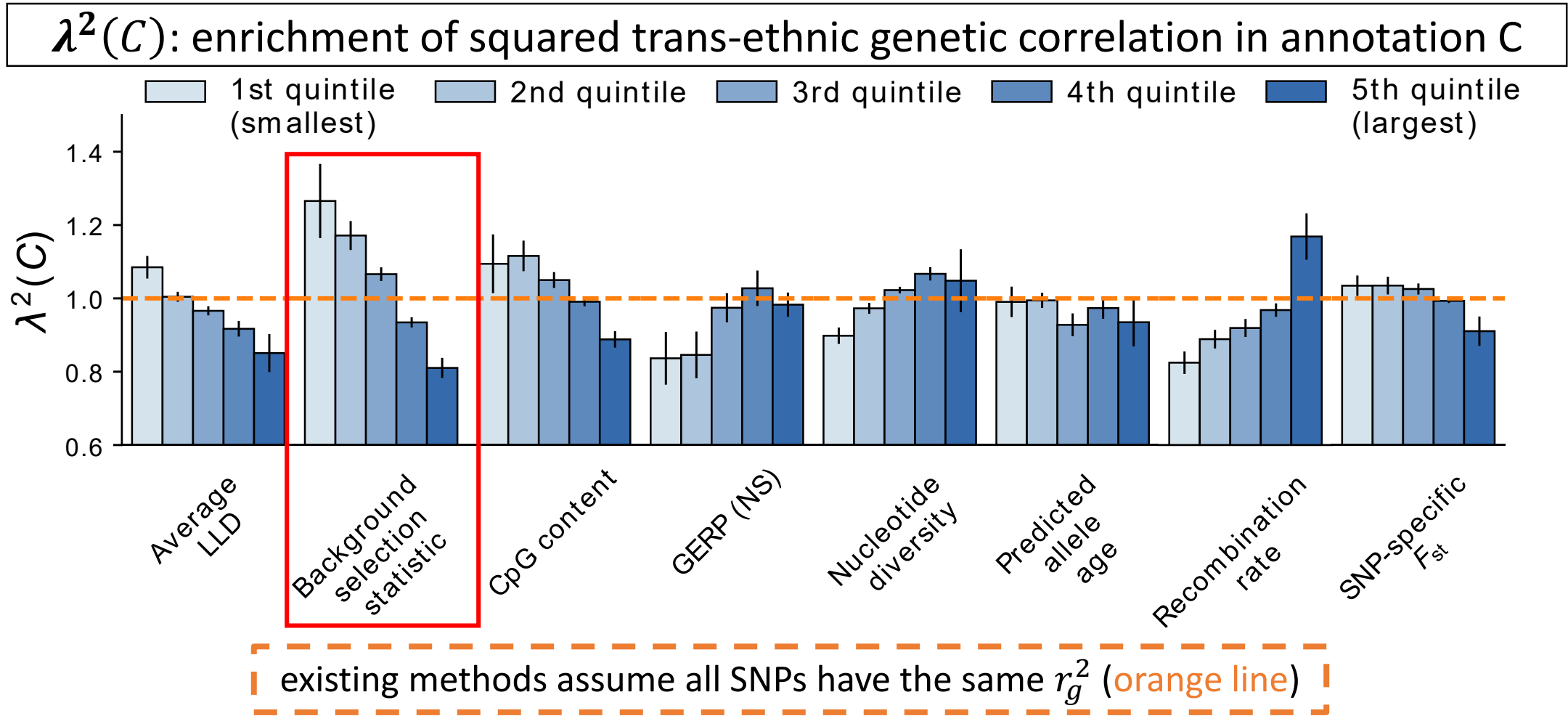
**European
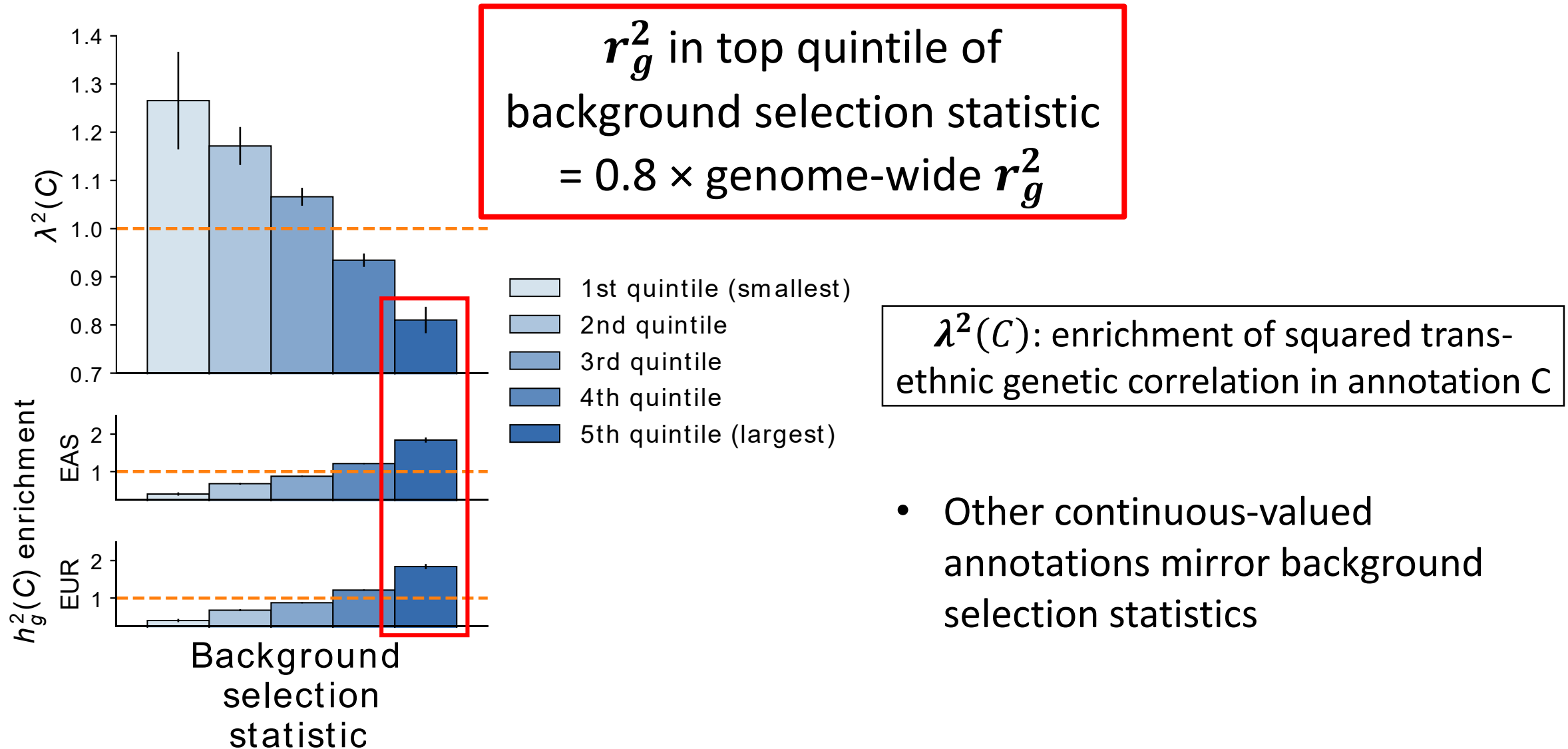(N ≈ 274K)**   UK Biobank   (Bycroft et al. 2018 *Nature*)

The method can be generalized to analyze other populations.

Trans-ethnic genetic correlation is depleted in top quintile of background selection statistic

- 8 continuous-valued annotations (Gazal et al. 2017 *Nat Genet*)

$\lambda^2(C)$: enrichment of squared trans-ethnic genetic correlation in annotation C

existing methods assume all SNPs have the same r_g^2 (orange line)

(Lee et al. 2012 *Bioinformatics*; Brown et al. 2016 *AJHG;* Galinsky et al. 2018 *Genet Epidemiol*)

Trans-ethnic genetic correlation is depleted in top quintile of background selection statistic

r_g^2 in top quintile of background selection statistic

= 0.8 × genome-wide r_g^2

$\lambda^2(C)$: enrichment of squared trans-ethnic genetic correlation in annotation C

- Other continuous-valued annotations mirror background selection statistics

Legend:
- 1st quintile (smallest)
- 2nd quintile
- 3rd quintile
- 4th quintile
- 5th quintile (largest)

(Gazal et al. 2017 *Nat Genet*)

Trans-ethnic genetic correlation is depleted in functionally important regions

$\lambda^2(C) < 1$ for most functional annotations

$\lambda^2(C)$: enrichment of squared trans-ethnic genetic correlation in annotation C
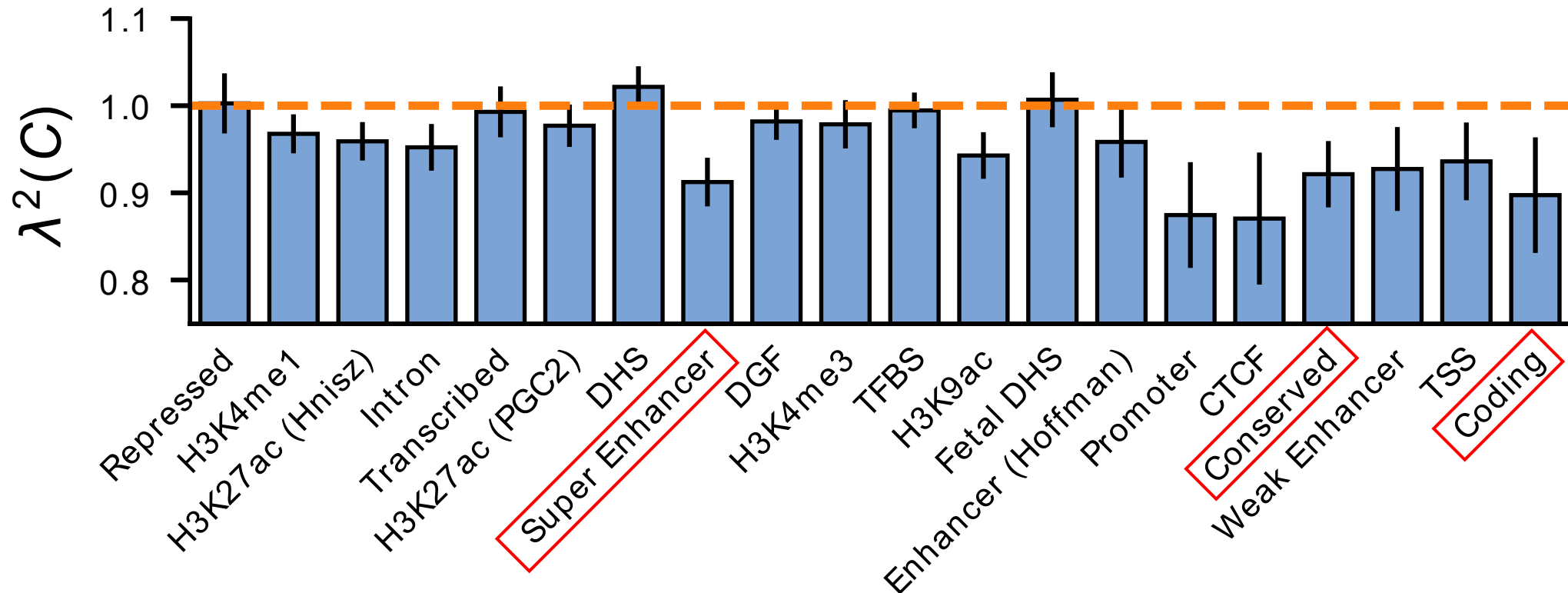
- Positively correlate with background selection statistic, enriched for heritability
  (Finucane et al. 2015 *Nat Genet*; Gazal et al. 2017 *Nat Genet*)

Trans-ethnic genetic correlation is most depleted near skin and immune genes and least depleted near brain genes

- Top 10% genes (± 100 kb) specifically expressed in 53 tissues (Finucane et al. 2018 *Nat Genet*)

$\lambda^2(C)$: enrichment of squared trans-ethnic genetic correlation in annotation C
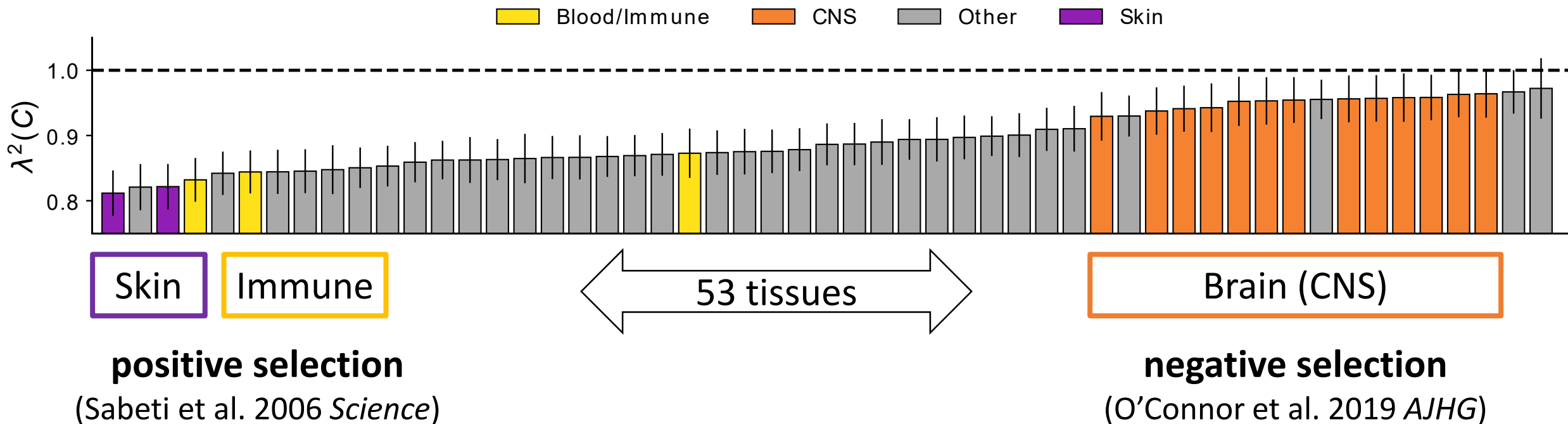
Outline

- Method to stratify trans-ethnic genetic correlation (S-LDXR)

- Analysis of 30 complex diseases and traits

- **Interpretation: potential role of positive selection**

Interpretation: potential role of positive selection

- Trans-ethnic genetic correlation is depleted in functionally important regions.

- This is likely due to stronger G x E. (Robinson et al. 2017 *Nat Genet*) Dominance and G x G are less likely. (Hill et al. 2008 *Plos Genet*; Zhu et al. 2015 *AJHG*)

- Results for immune and skin genes suggest G x E may be linked to positive selection.
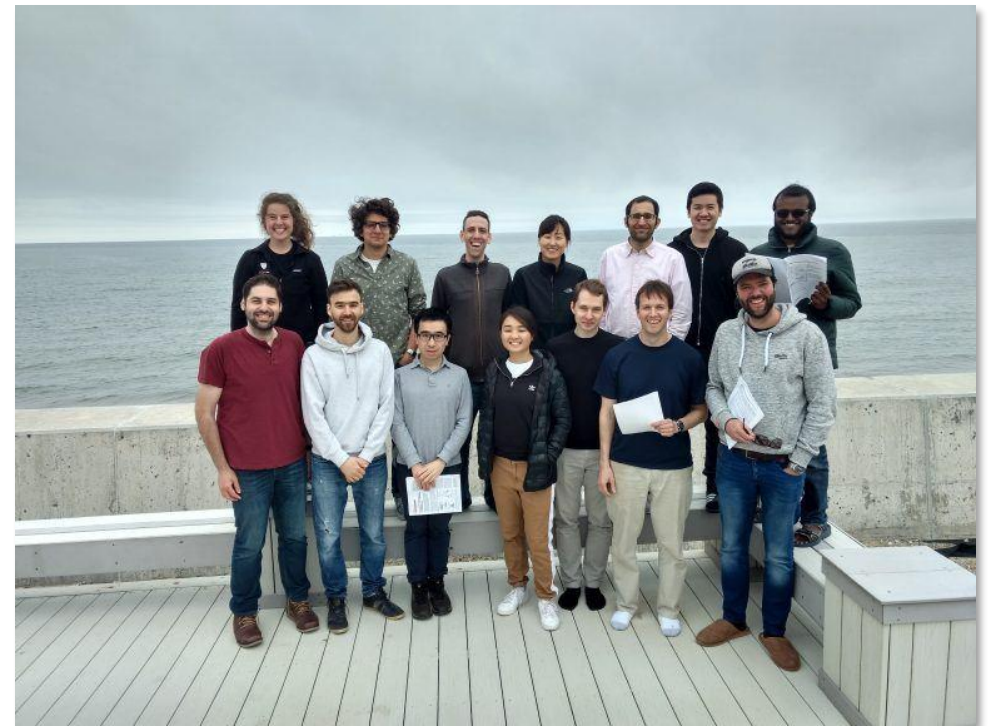
Acknowledgments

Also thanks to:
Luke O'Connor
Hilary Finucane
Daniel Kassler
Benjamin Neale
Raymond Walters
Alicia Martin
Nick Patterson
Swapan Mallick
Katherine Siewert
Brielin Brown
Bogdan Pasaniuc
Margaux Hujoel
Farhad Hormozdiari

Alkes Price
Steven Gazal
Masahiro Kanai
Evan Koch
Armin Schoech
Samuel Kim
Yang Luo
Tiffany Amariuta
Yukinori Okada
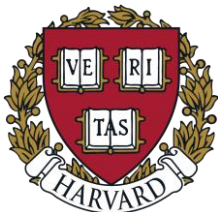Soumya Raychaudhuri
Shamil Sunyaev

Twitter:
@shihuwenbo

(Shi et al. 2019 *bioRxiv*)

(journal club retreat, May 2019)