Disclosure Slide

Financial Disclosure for:
Omer Weissbrod

I have nothing to disclose
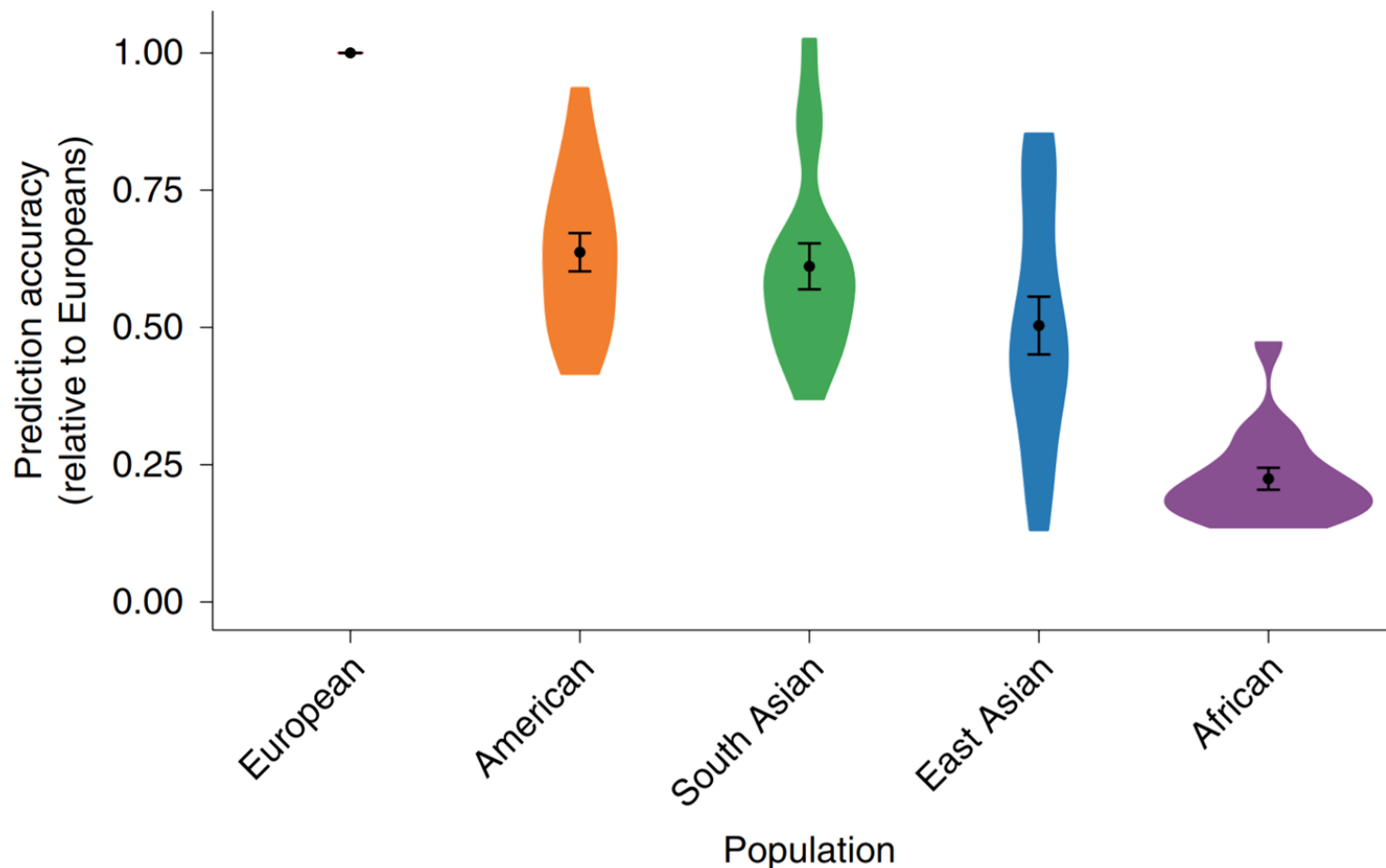
Leveraging fine-mapping
and non-European training data
to improve trans-ethnic polygenic risk scores

Omer Weissbrod

Alkes Price Group

Harvard School of Public Health

Polygenic risk scores lose accuracy in non-European target populations

Outline

➢ Introduction:
Why do polygenic risk scores lose accuracy across populations?

➢ Methods

➢ Results on real traits

Outline

➢ **Introduction:**
Why do polygenic risk scores lose accuracy across populations?

➢ Methods

➢ Results on real traits

Why do polygenic risk scores lose accuracy across populations?

1. LD differences

2. Allele frequency differences

Marquez-Luna *et al.* 2017 Genet Epidem.
Duncan et al. 2019 Nat Commun.

Martin *et al.* 2019 Nat Genet
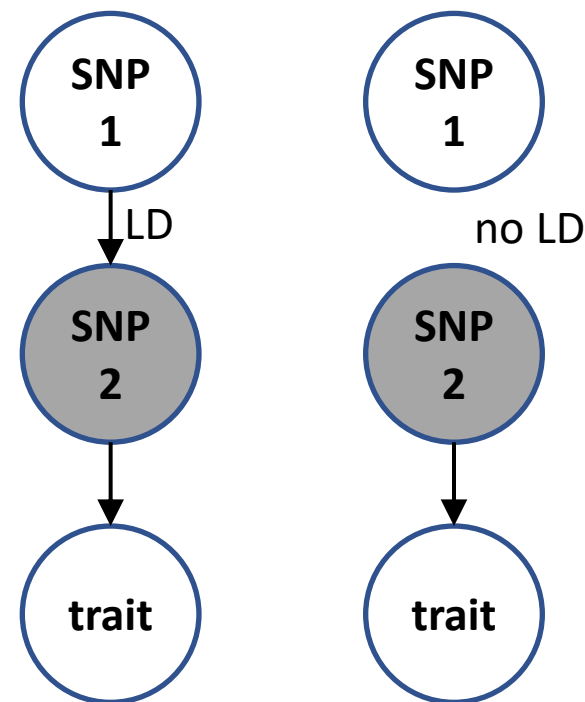Wang *et al.* 2020 Nat Commun.

Why do polygenic risk scores lose accuracy across populations?

1. **LD differences**
   (when using non-causal SNPs to predict)

2. Allele frequency differences

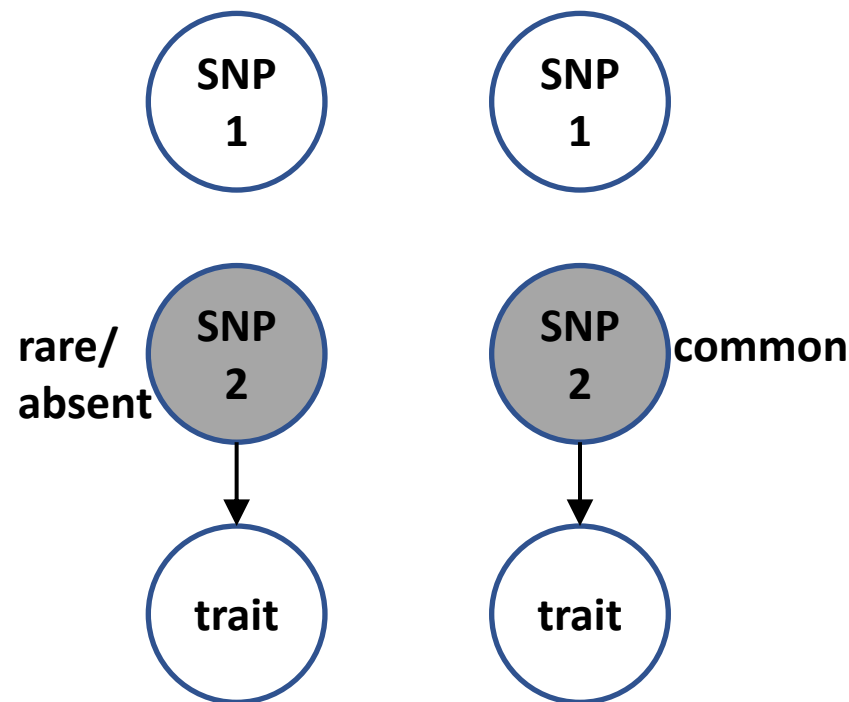**European pop.   Non-European pop.**

Why do polygenic risk scores lose accuracy across populations?

1. LD differences
   (when using non-causal SNPs to predict)

2. **Allele frequency differences**
   (even when using causal SNPs)

European pop.   Non-European pop.

SNP 1   SNP 1

rare/absent   SNP 2   SNP 2   common

trait   trait

Two strategies to mitigate loss of polygenic risk score accuracy

1. **LD differences**
(when using non-causal SNPs to predict)

**Predict using
<u>causal</u> SNPs
(fine-mapping)**

2. **Allele frequency differences**
(even when using causal SNPs)

**Combine data
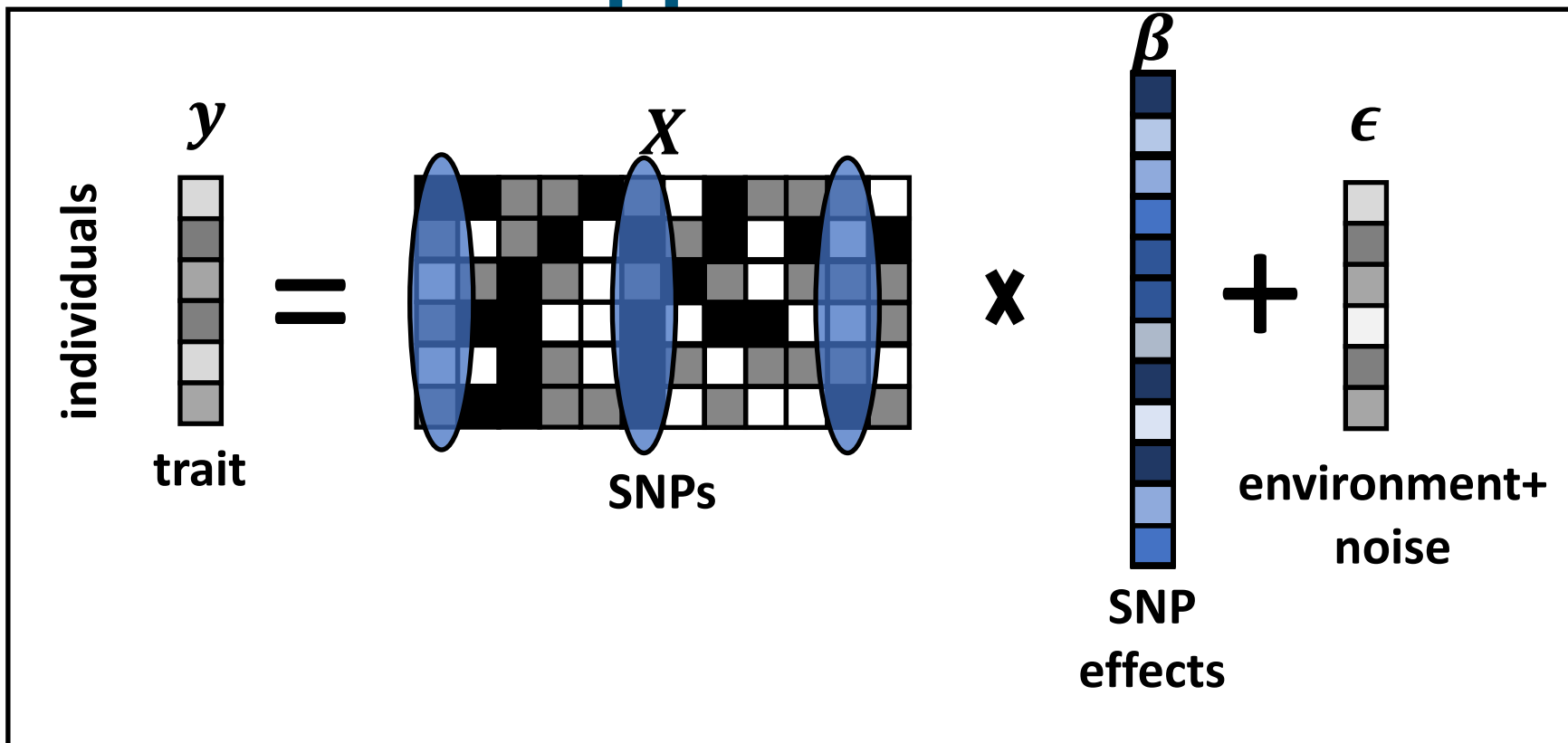from Europeans
and non-Europeans**

Outline

➤ Introduction:
Why do polygenic risk scores lose accuracy across populations?

➤ **Methods**

➤ Results on real traits

Fine-mapping is closely related to PRS applied to all SNPs

Fine-mapping:
- Estimate effect sizes for **all** SNPs
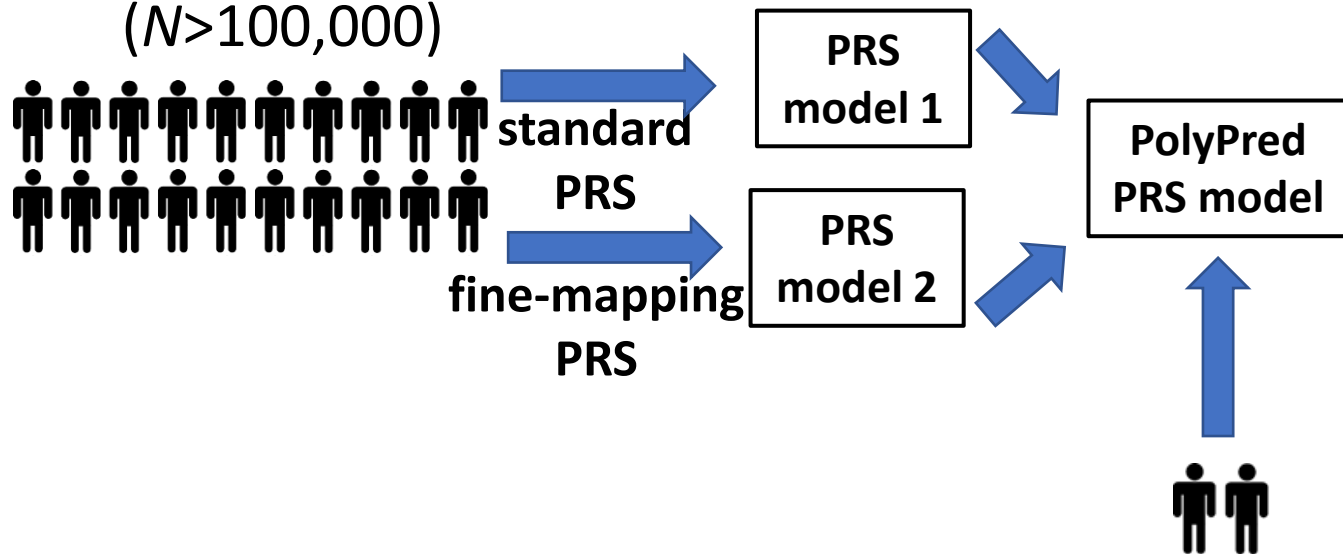- β_i represents a **causal** effect

PRS:
- Estimate effect sizes for **a subset** of SNPs
- β_i represents a **causal+tagging** effect

PolyPred combines a standard PRS with a fine-mapping PRS (accounts for LD differences)
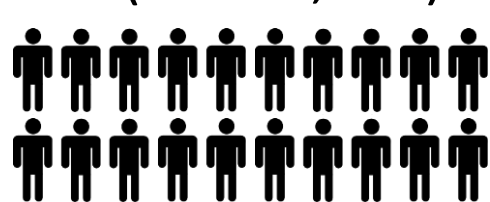
large European sample
(*N*>100,000)

standard PRS → PRS model 1

fine-mapping PRS → PRS model 2

PRS model 1 → PolyPred PRS model

PRS model 2 → PolyPred PRS model

Small training sample
from **target cohort** (N>100) → PolyPred PRS model

PolyPred+ extends PolyPred to include a non-European PRS (if available) (accounts for LD, MAF, effect size differences)

large European sample
(*N*>100,000)

standard PRS

fine-mapping PRS

PRS model 1
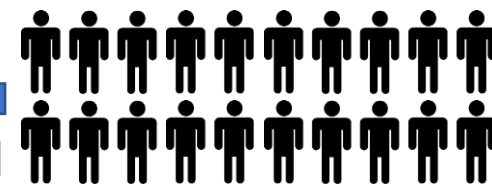
PRS model 2

PolyPred+ PRS model

PRS model 3

standard PRS

large non-European sample (*N*>100,000)

Small training sample
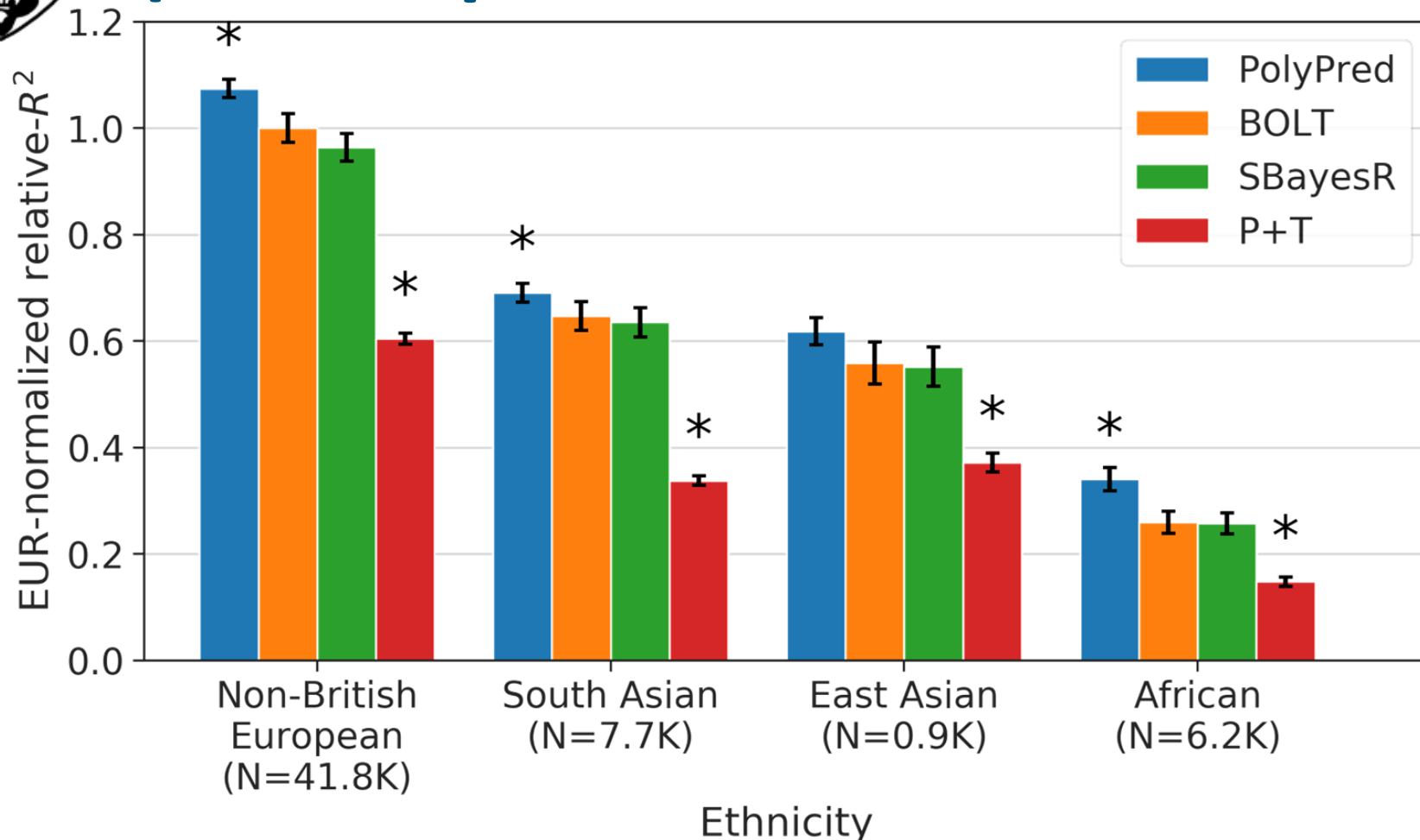from **target cohort** (N>100)

Outline

➢ Introduction:
   Why do polygenic risk scores lose accuracy across populations?

➢ Methods

➢ **Results on real traits**

PolyPred significantly improves PRS accuracy in the UK Biobank (32% improvement in Africans vs BOLT)

traits:
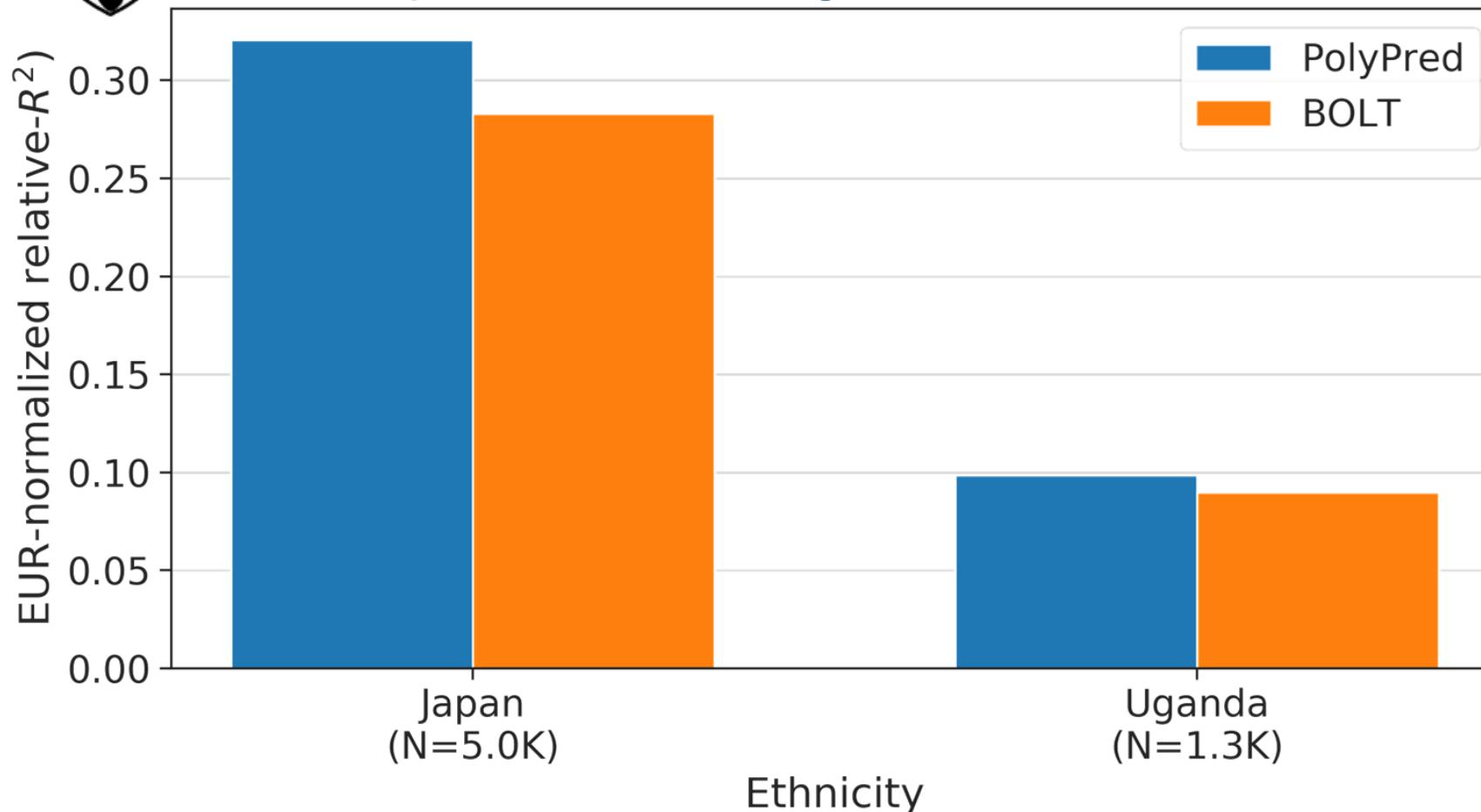7 independent complex traits

training data:
N=325K UKB British

* P<0.05 (diff vs BOLT)

conservative block-jackknife s.e.

BOLT: BOLT-LMM (Loh *et al.* 2015a Nat Genet, 2018)
SBayesR (Lloyd-Jones *et al.* 2019 Nat Commun)

PolyPred significantly improves PRS accuracy in Biobank Japan (13.4% improvement vs BOLT)

training data: *N*=325K UKB British

Large drop in **absolute** accuracy compared to within-UKB PRS

Biobank Japan: Nagai *et al.* 2017 J. Epidemiol
Uganda-APCDR: Asiki *et al.* 2013 Int. J. Epidemiol

PolyPred+ significantly improves PRS accuracy in UK Biobank East Asians (24% improvement vs BOLT)
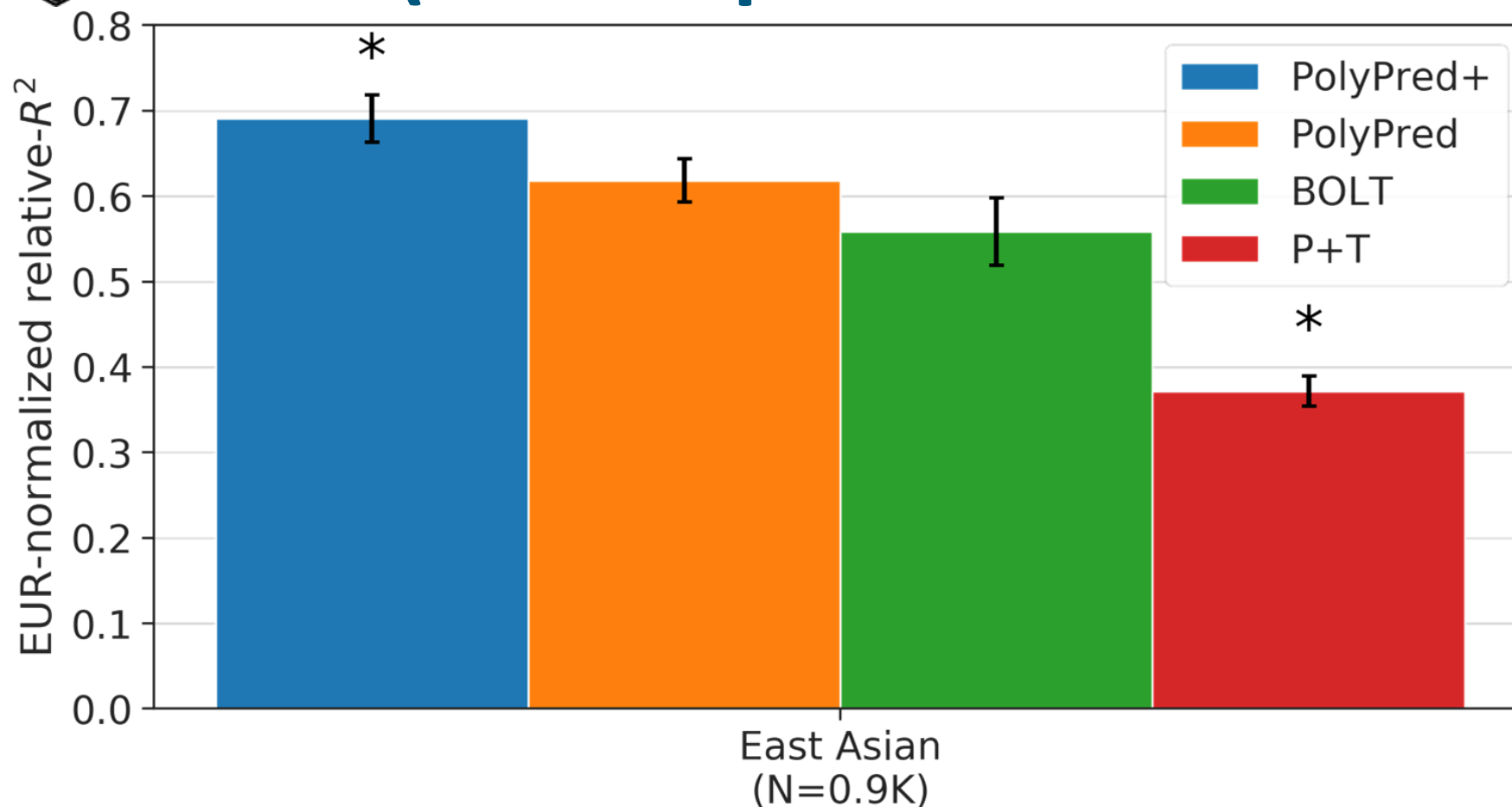
training data:
N=325K UKB British +
N=124K BBJ Japanese

Conclusions

- **PolyPred** leverages **fine-mapping** to improve trans-ethnic PRS
  (32% improvement vs BOLT in UKB Africans,
   11% improvement vs. BOLT in UKB East Asians)

- **PolyPred+** leverages fine-mapping and **non-European data**
  (24% improvement vs BOLT in UKB East Asians)

Acknowledgements

Alkes Price
Huwenbo Shi
Steven Gazal
Wouter Peyrot
Price Group

Masahiro Kanai
Alicia Martin
Hilary Finucane
Amit Khera
Yukinori Okada