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

Martin et al. 2019 Nat Genet
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
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)
Two strategies to mitigate loss of polygenic risk score accuracy

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

   Predict using causal 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
- $\beta_i$ represents a **causal** effect

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

\[
\begin{align*}
    y &= X\beta + \epsilon \\
    \text{individuals} & \quad \text{trait} \\
    \text{SNPs} & \quad \text{SNP effects} \\
    \text{environment+ noise} &
\end{align*}
\]
PolyPred combines a standard PRS with a fine-mapping PRS (accounts for LD differences)

- **Standard PRS**: BOLT-LMM (Loh et al. 2015a Nat Genet, 2018) and SBayesR (Lloyd-Jones et al. 2019 Nat Commun)
- **Fine-mapping PRS**: PolyFun + SuSiE (Weissbrod et al. accepted in principle Nat Genet)

Diagram:
- Large European sample ($N > 100,000$)
- Standard PRS model 1
- Fine-mapping PRS
- PolyPred PRS model
- Small training sample from target cohort ($N > 100$)
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 \((N>100)\)
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)
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