



JOHNS HOPKINS  
UNIVERSITY

# multi-ancestry Polygenic Risk scOres based on enSemble of PEnalized Regression models (PROSPER)

---

**Jingning Zhang\***, Jianan Zhan, Jin Jin, Cheng Ma, Ruzhang Zhao, Jared O'Connell,

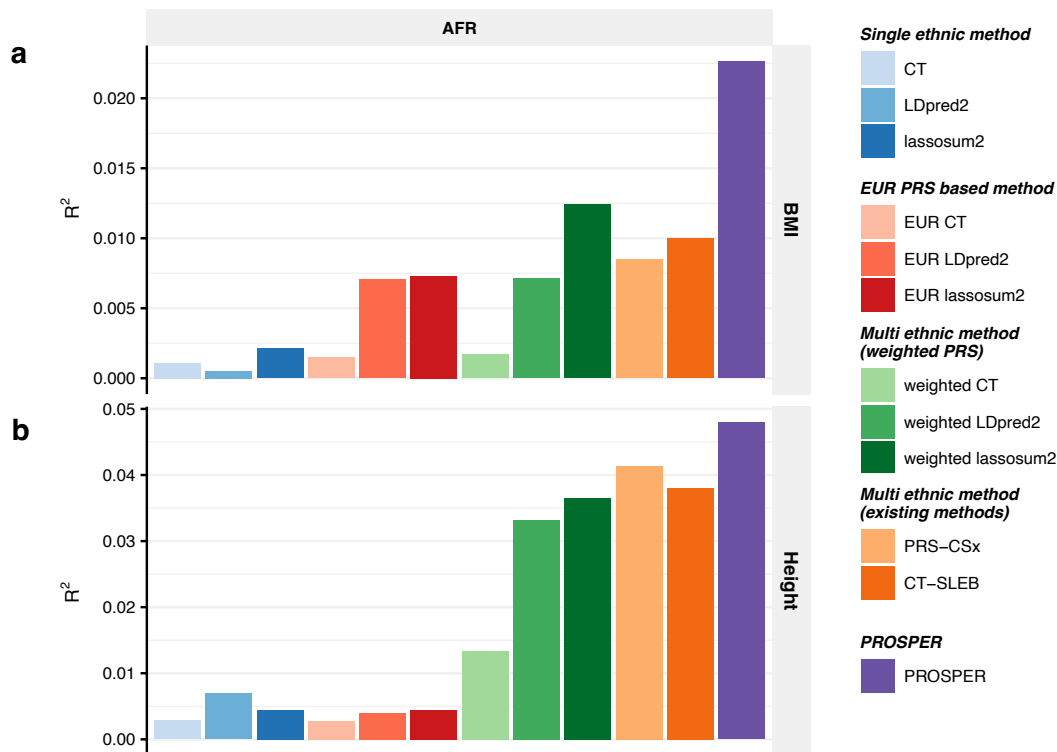
Yunxuan Jiang, 23andMe Research Team, Bertram L. Koelsch, Haoyu Zhang, **Nilanjan Chatterjee\***

\*Correspondence to: **Jingning Zhang (jzhan218@jhu.edu)** and **Nilanjan Chatterjee (nilanjan@jhu.edu)**

# The proposed method -- PROSPER

Details of the proposed model will be released  
right after the manuscript being submitted to journal

# Results on data from All of US (AoU)



► Sample size

► EUR (N = 48K)

► AFR (N = 22K)

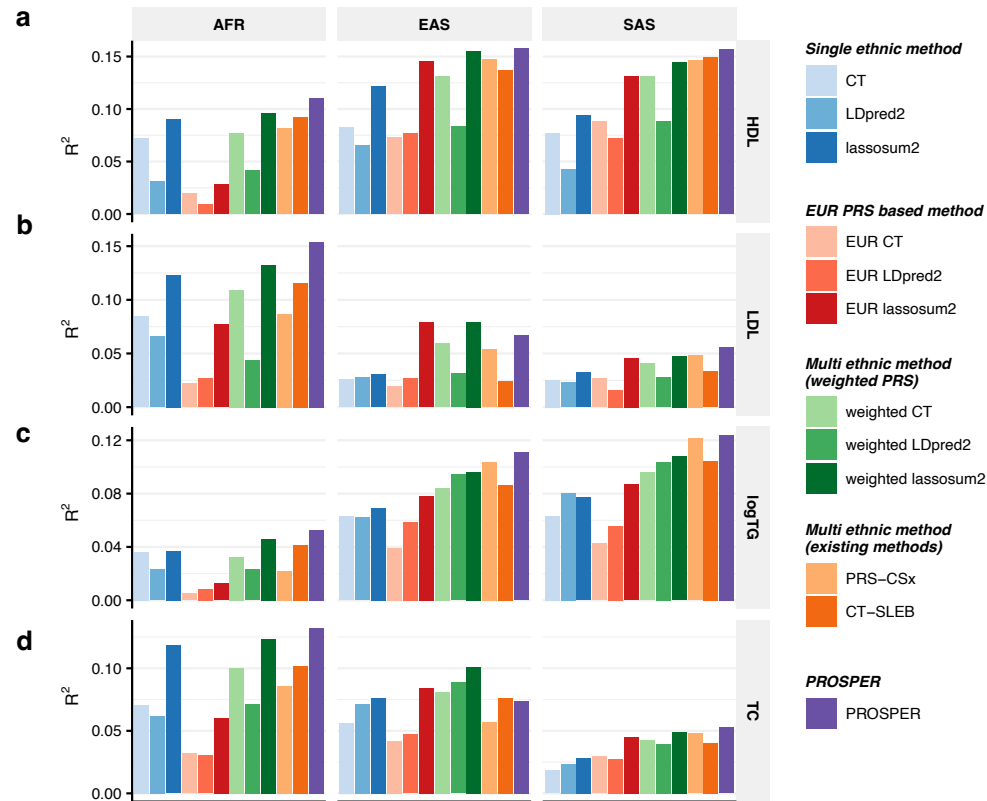
► Relative improvement in  $R^2$  of PROSPER over alternative methods (average across traits and ancestries)

► 91.3% over PRS-CSx

► 76.5% over CT-SLEB

► 56.8% over weighted lassosum2

# Results on data from Global Lipids Genetics Consortium (GLGC)



## Sample size

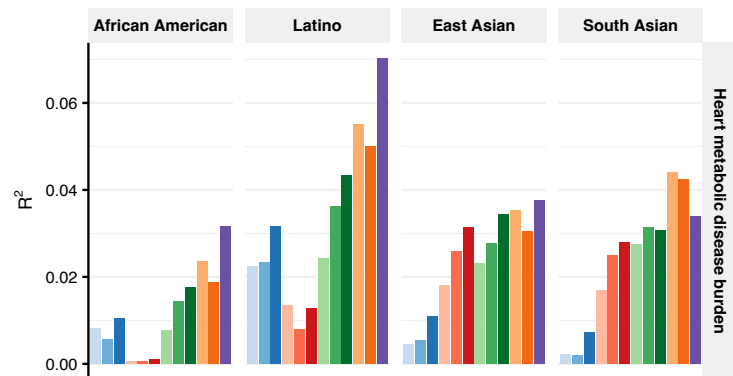
- EUR (N = 931K)
- AFR (N = 93K)
- EAS (N = 146K)
- SAS (N = 34K)

## Relative improvement in $R^2$ of PROSPER over alternative methods (average across traits and ancestries)

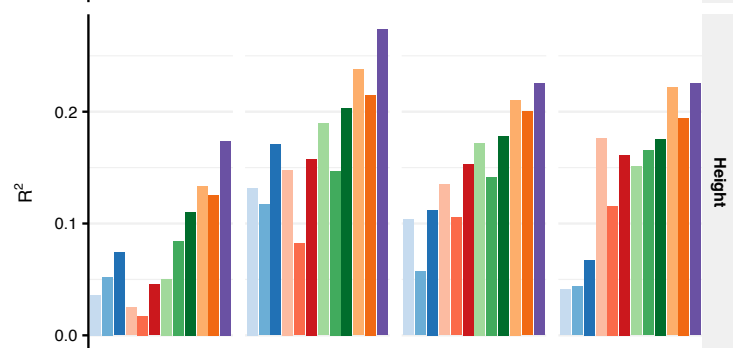
- 32.5% over PRS-CSx
- 36.9% over CT-SLEB
- 11.0% over weighted lassosum2 in AFR
- 11.2% over weighted lassosum2 in SAS

# Results on data from 23andMe Inc. (23andMe) continuous traits

a

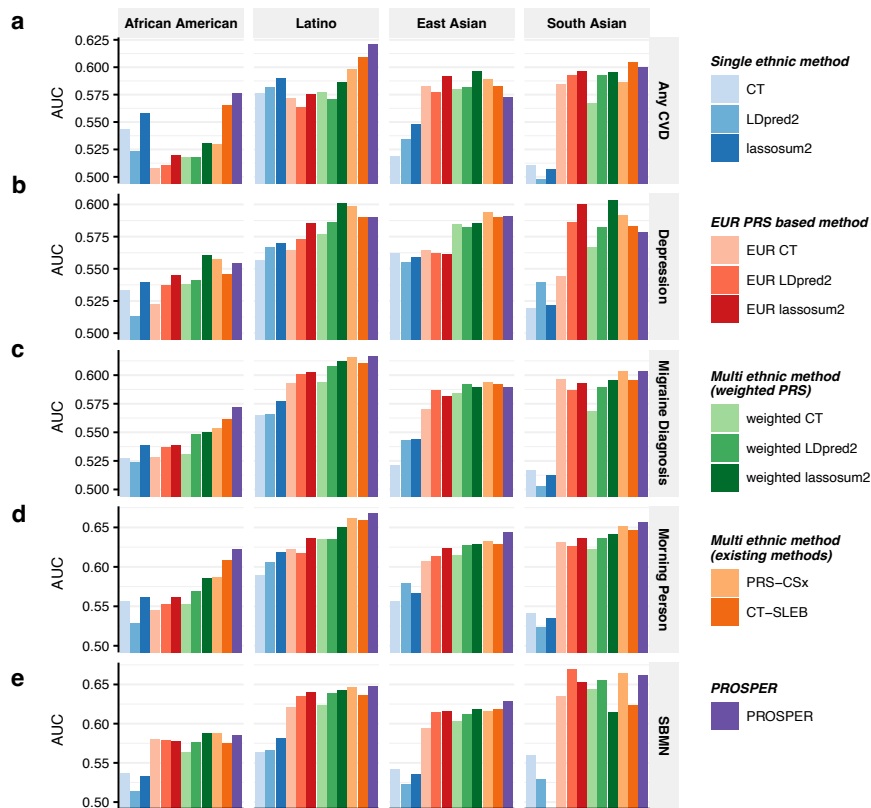


b



- ▶ Sample size
  - ▶ EUR ( $N \approx 2500K$ )
  - ▶ African American ( $N \approx 135K$ )
  - ▶ Latino ( $N \approx 400K$ )
  - ▶ EAS ( $N \approx 115K$ )
  - ▶ SAS ( $N \approx 30K$ )
- ▶ Relative improvement in  $R^2$  of PROSPER over alternative methods (average across traits and ancestries)
  - ▶ 12.5% over PRS-CSx
  - ▶ 25.7% over CT-SLEB

# Results on data from 23andMe Inc. (23andMe) binary traits



- ▶ Sample size
  - ▶ EUR ( $N \approx 2370K$ )
  - ▶ African American ( $N \approx 110K$ )
  - ▶ Latino ( $N \approx 400K$ )
  - ▶ EAS ( $N \approx 86K$ )
  - ▶ SAS ( $N \approx 24K$ )
- ▶ Relative improvement in AUC of PROSPER over alternative methods (average across traits and ancestries)
  - ▶ 1.1% over PRS-CSx
  - ▶ 1.3% over CT-SLEB

# Runtime and memory usage

- ▶ Chromosome 22
- ▶ two ancestries: AFR, and EUR  
five ancestries: AFR, AMR, EAS, EUR, and SAS
- ▶ Runtime and memory usage for model training part

Method	Computational time (minutes)	Memory (Gb)
PROSPER (two ancestries)	3.0	2.24
PROSPER (five ancestries)	6.8	2.35
PRS-CSx (two ancestries)	111.1	0.78
PRS-CSx (five ancestries)	595.8	0.84

# Conclusion

- ▶ PROSPER has substantial improvement over alternative methods and can deal with traits with a variety of genetic architectures
- ▶ PROSPER is not only more powerful for complex traits **with high polygenicity**, but also **robust to biomarker traits with large-effect loci**. In addition, PROSPER developed based on penalized regression is **an order of magnitude faster** compared to alternative Bayesian methods.