Local Ancestry and Population Structure in Biobanks - Methods and Implications

Dan Lawson, School of Mathematics

External IEU Faculty

Work with Yaoling Yang, School of Maths,

Aimee Hanson, IEU

Important contributions from Richard Durbin, Astrid Iversen, Will Barrie, Simon Myers, and more

Is population genetics relevant to CHD?

KeAi
CHINESE ROOTS
GLOBAL IMPACT

Gynecology and Obstetrics Clinical Medicine

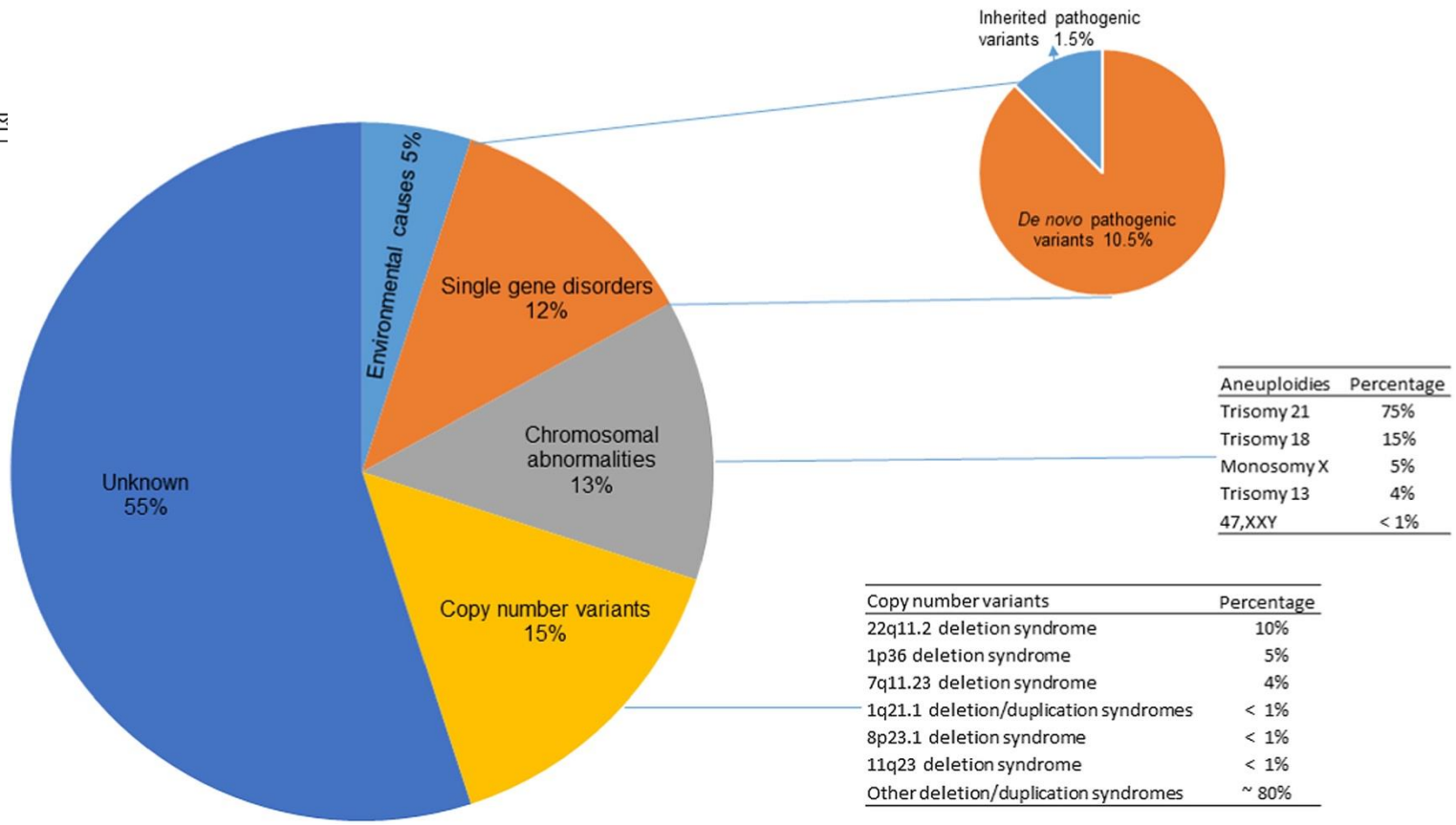
Volume 2, Issue 3, September 2022, Pages 109-123

Review Article

Genetic detection of congenital heart disease

Sumathi I. Rachamadugu [a b 1], Kristen A. Miller [c 1], Ina H. Lee [d], Ying S. Zou [b e f]

- … about 400 human genes are involved in causing or contributing to structural CHD,[6,52,53] < 200 genes are currently recognized as being definitively associated,[54] …
- some known CHD genes display an inheritance pattern characterized by incomplete penetrance



Aneuploidies	Percentage
Trisomy 21	75%
Trisomy 18	15%
Monosomy X	5%
Trisomy 13	4%
47,XXY	< 1%

Copy number variants	Percentage
22q11.2 deletion syndrome	10%
1p36 deletion syndrome	5%
7q11.23 deletion syndrome	4%
1q21.1 deletion/duplication syndromes	< 1%
8p23.1 deletion syndrome	< 1%
11q23 deletion syndrome	< 1%
Other deletion/duplication syndromes	~ 80%

CAD GWAS

- …largest GWAS for angiographically derived coronary atherosclerosis …15 loci of genome-wide significance that robustly overlap with established loci for clinical CAD
- markedly reduced transferability of existing PRSs to Black individuals…
- …no region has convincingly reached genome-wide significance in Black or Hispanic populations…

nature medicine

Explore content · About the journal · Publish with us

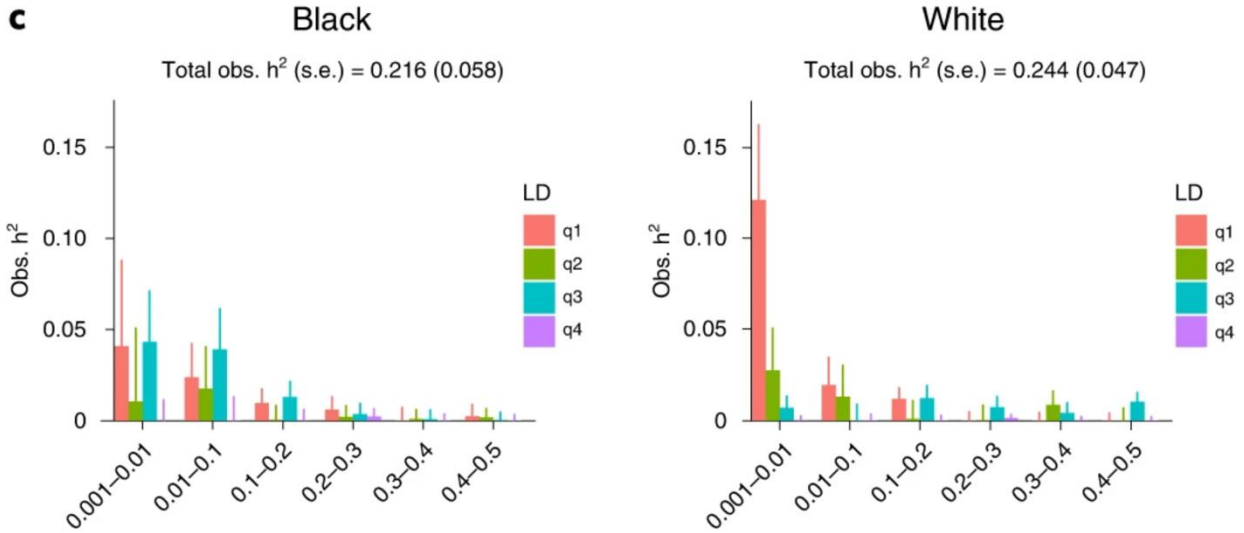
nature › nature medicine › articles › article

Article | Published: 01 August 2022

Large-scale genome-wide association study of coronary artery disease in genetically diverse populations

Catherine Tcheandjieu, Xiang Zhu, Austin T. Hilliard, Shoa L. Clarke, Valerio Napolioni, Shining Ma, Kyung Min Lee, Huaying Fang, Fei Chen, Yingchang Lu, Noah L. Tsao, Sridharan Raghavan, Satoshi Koyama, Bryan R. Gorman, Marijana Vujkovic, Derek Klarin, Michael G. Levin, Nasa Sinnott-Armstrong, Genevieve L. Wojcik, Mary E. Plomondon, Thomas M. Maddox, Stephen W. Waldo, Alexander G. Bick, Saiju Pyarajan, Regeneron Genetics Center, CARDIoGRAMplusC4D Consortium, Biobank Japan, Million Veteran Program, … Themistocles L. Assimes

+ Show authors

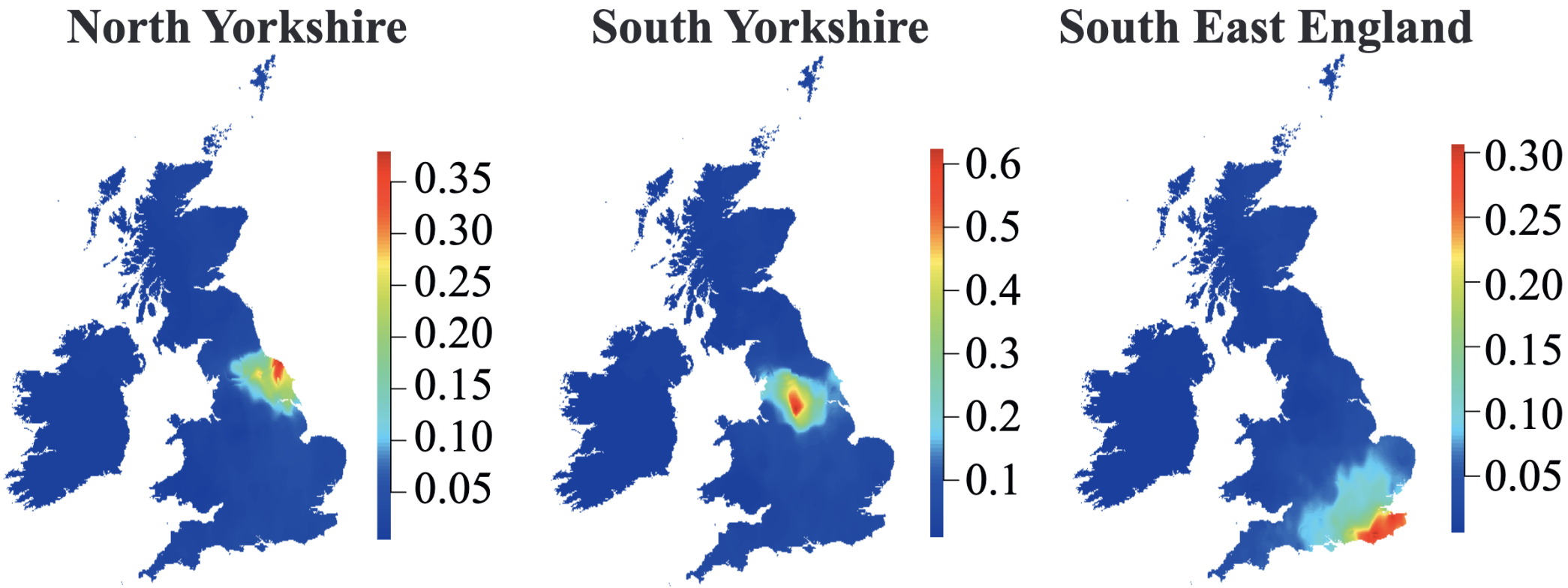


What can population genetics tell us?

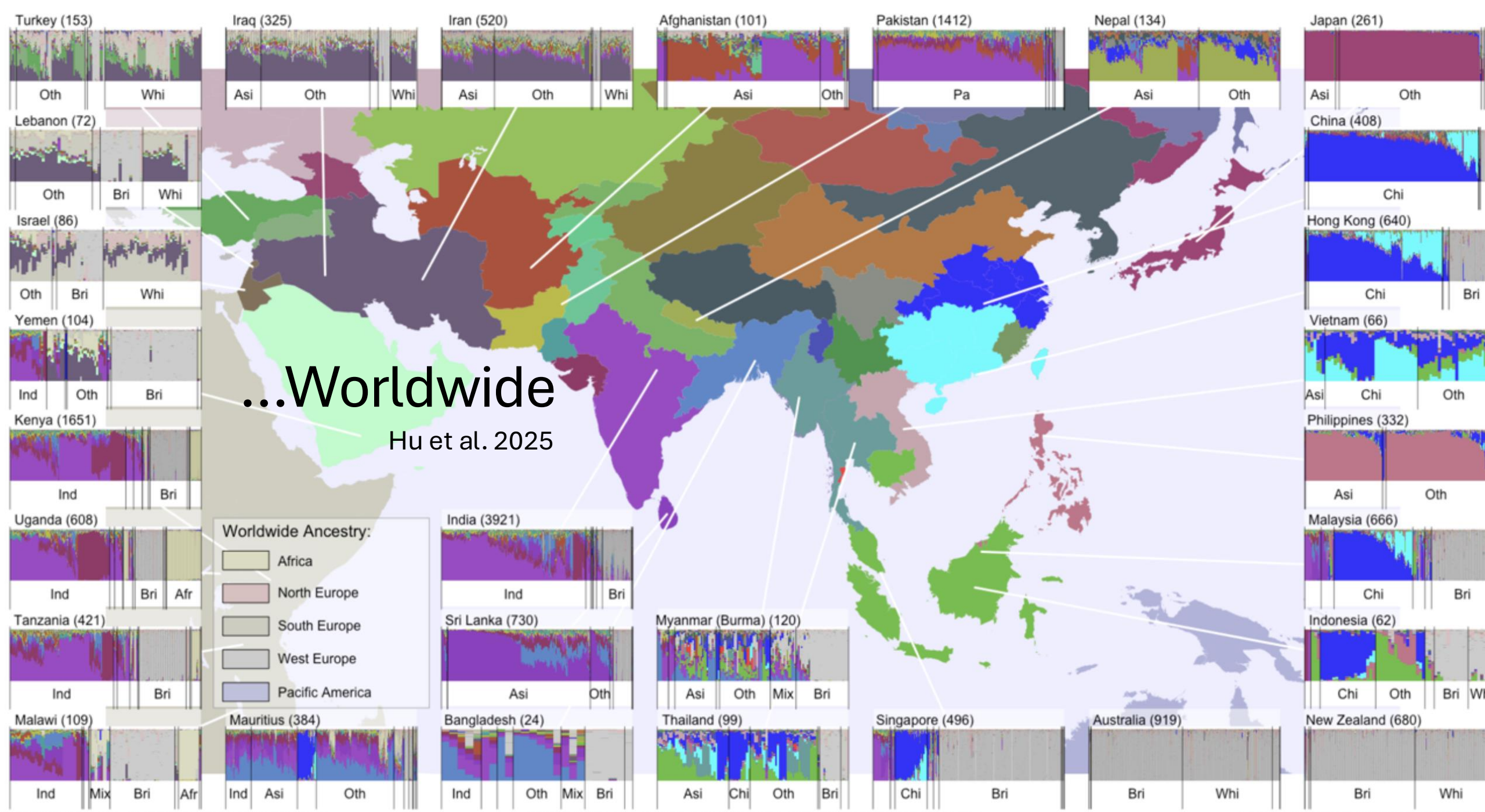
Consequences of ancestry?

- Population Structure describes with which population an individual shares ancestry
- We are here learning the ancestry of particular segments of DNA
- We have new software to do this very quickly called SparsePainter [Yang et al 2024], which makes fine-scale ancestry possible [Hu et al. 2025]
- [Yang et al. 2024] also shows that PBWTpaint is much more accurate than PCA correction, and is laptop-fast to process the whole UK Biobank
- We made the selection metrics (ancestry segments too short, or unusual at a locus) from [Barrie et al. Nature 2024] deployable at scale
- We see massive recent selection on immune genes, and consequently autoimmune disease
- Interested in talking to people that work on related problems!

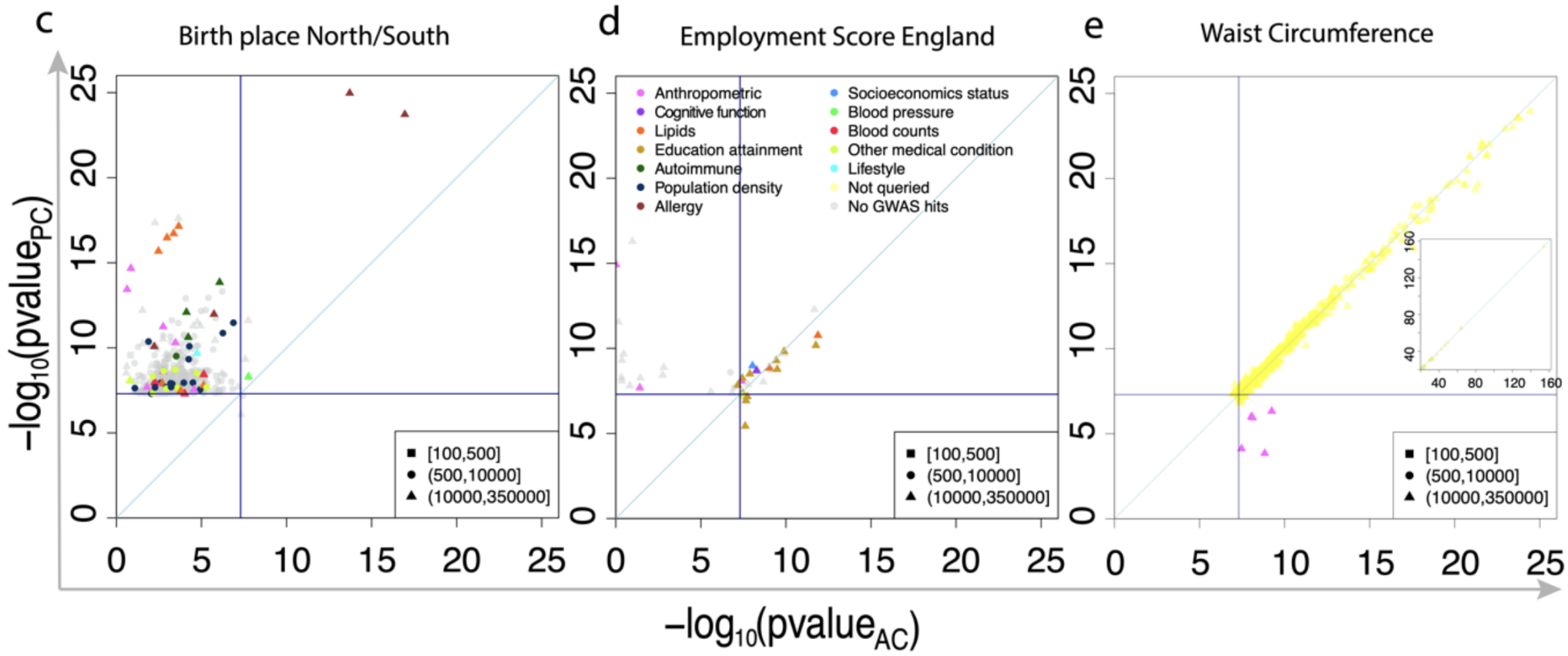
Fine-scale ancestry decomposition is now “easy”



Hu et al. 2025


Turkey (153)
Oth
Whi
Iraq (325)
Asi
Oth
Whi
Iran (520)
Asi
Oth
Whi
Afghanistan (101)
Asi
Oth
Pakistan (1412)
Pa
Nepal (134)
Asi
Oth
Japan (261)
Asi
Oth
Lebanon (72)
Oth
Bri
Whi
China (408)
Chi
Israel (86)
Oth
Bri
Whi
Hong Kong (640)
Chi
Bri
Yemen (104)
Ind
Oth
Bri
Vietnam (66)
Asi
Chi
Oth
...Worldwide
Hu et al. 2025
Kenya (1651)
Ind
Bri
Philippines (332)
Asi
Oth
Uganda (608)
Ind
Bri
Afr
India (3921)
Ind
Bri
Malaysia (666)
Chi
Bri
Worldwide Ancestry:
Africa
North Europe
South Europe
West Europe
Pacific America
Tanzania (421)
Ind
Bri
Sri Lanka (730)
Asi
Oth
Myanmar (Burma) (120)
Asi
Oth
Mix
Bri
Indonesia (62)
Chi
Oth
Bri
Malawi (109)
Ind
Mix
Bri
Afr
Mauritius (384)
Ind
Asi
Oth
Bangladesh (24)
Ind
Oth
Mix
Bri
Thailand (99)
Asi
Chi
Oth
Bri
Singapore (496)
Chi
Bri
Australia (919)
Bri
Whi
New Zealand (680)
Bri
Whi
Ethnicity Key: Europe: Bri=British, Whi=White, Ire=Irish
Asia: Chi:Chinese, Asi:Asian, Pa:Pakistani, Ind:Indian, Mix:Mixed, Oth:Other
African: Afr:African, Blk:Black, Car:Caribb

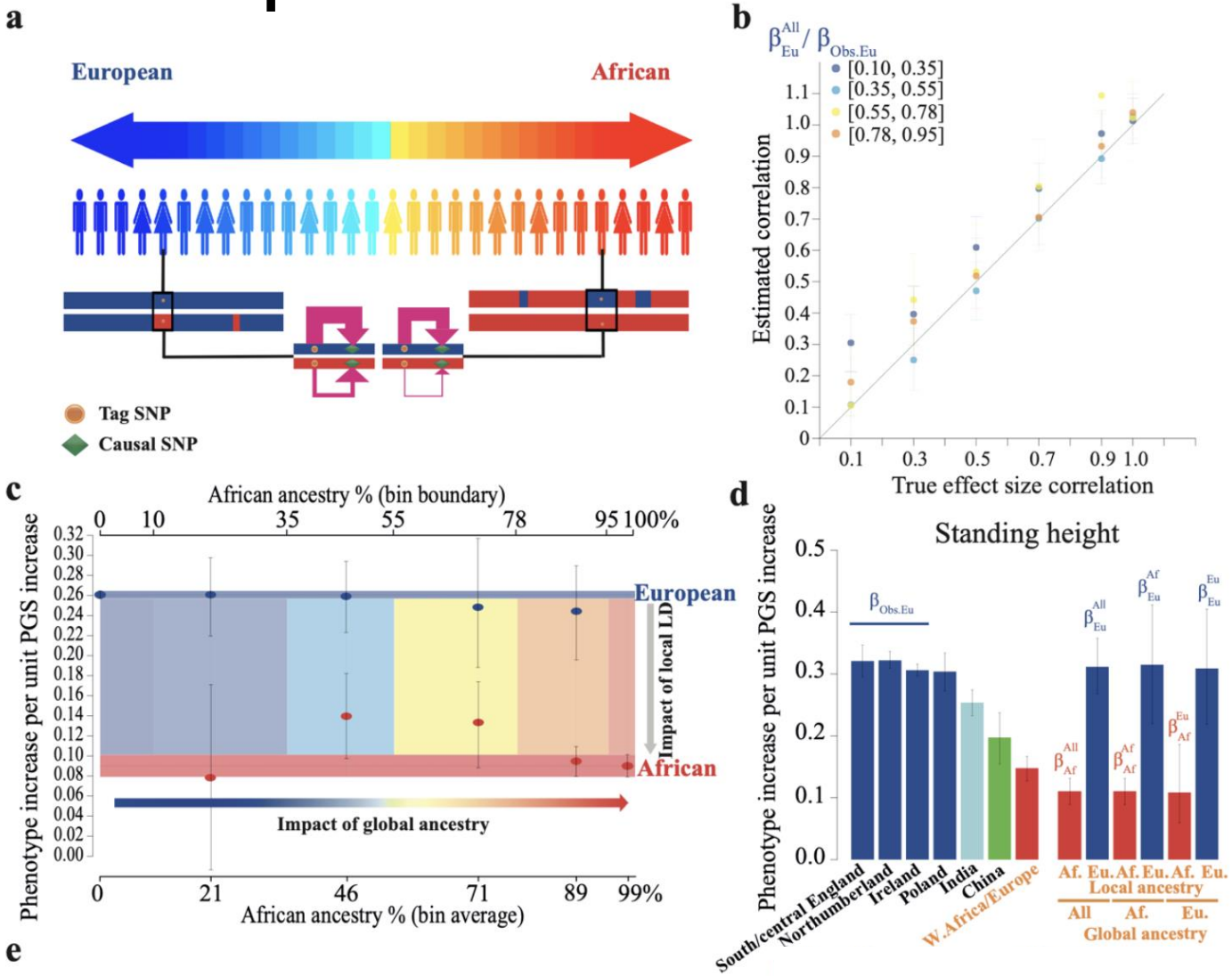
It is useful for GWAS*



* But the difference for “biological” traits is negligible…

Hu et al. 2025

We find that polygenic scores only decay due to linkage disequilibrium*



(*for all traits we examined, see paper)

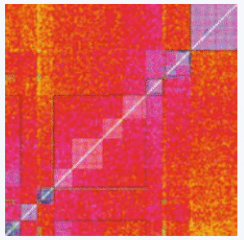
Hu et al. 2025

And we can do it computationally efficiently

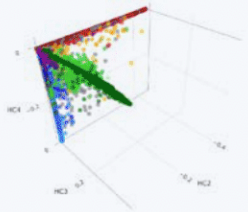
PBWTpaint
Unsupervised Learning

Fine Scale Structure
All-vs-All

Clustering

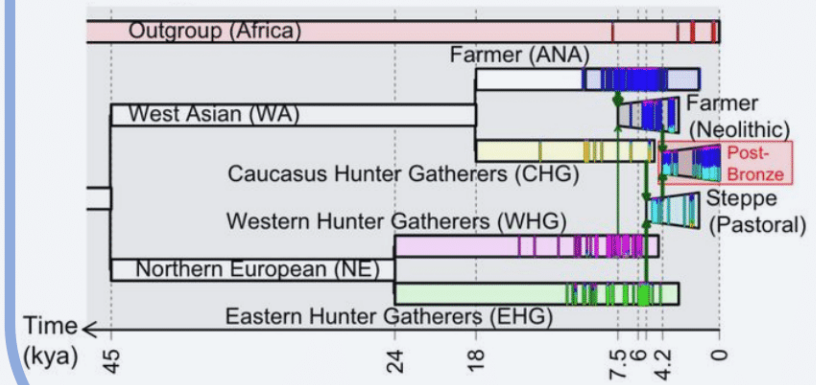


PCA



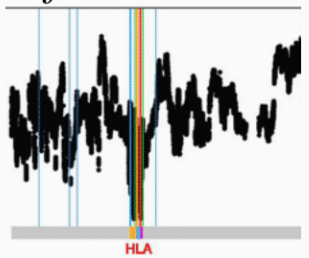
SparsePainter
Supervised Learning

Admixture Modelling
Reference-vs-Reference
+ Target vs Reference

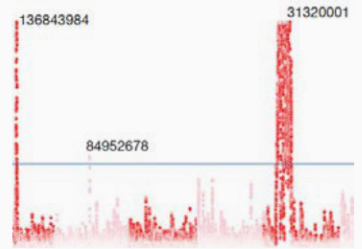


Local Ancestry
Target-vs-Reference

Ancestry LD Score



Ancestry Anomaly Score



Genome-wide properties, including:

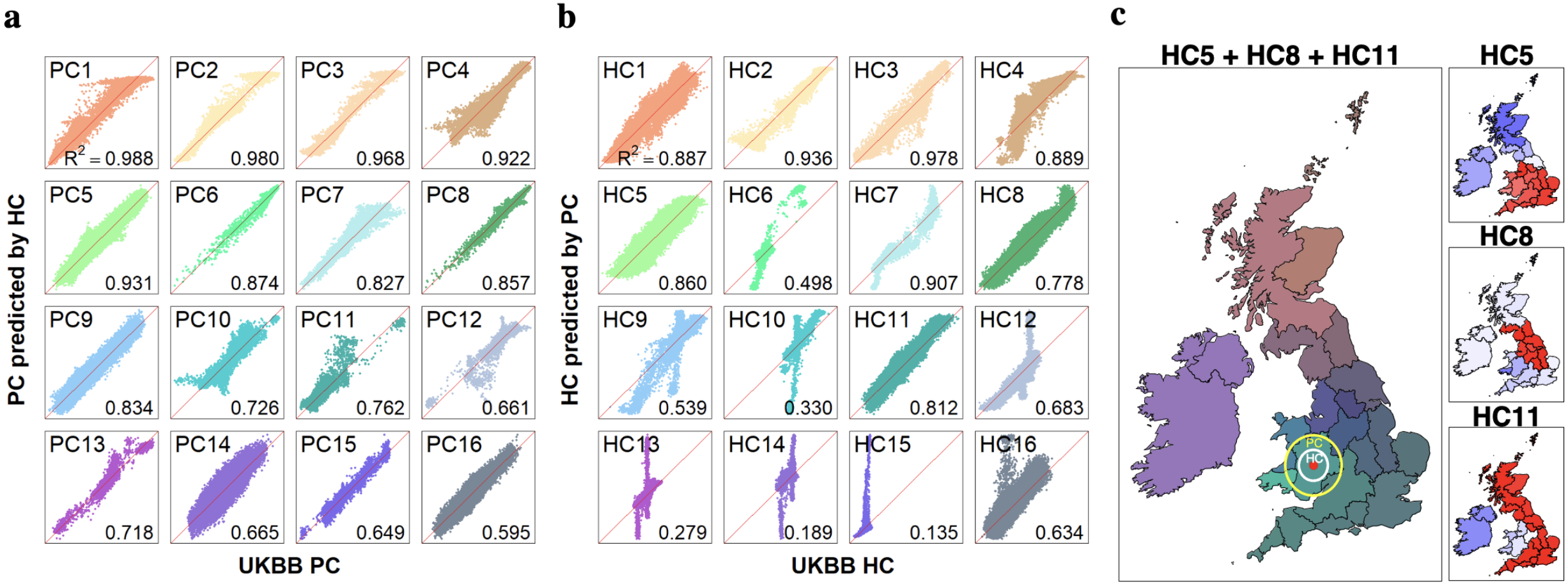
Population History	GWAS
	Polygenic Scores

Genetically localized, including:

Ancestral GWAS	Selection
Introgression	Trait evaluation

Yang et al. 2024

PBWTs Haplotypes Components (HCs) define population structure better than PCs

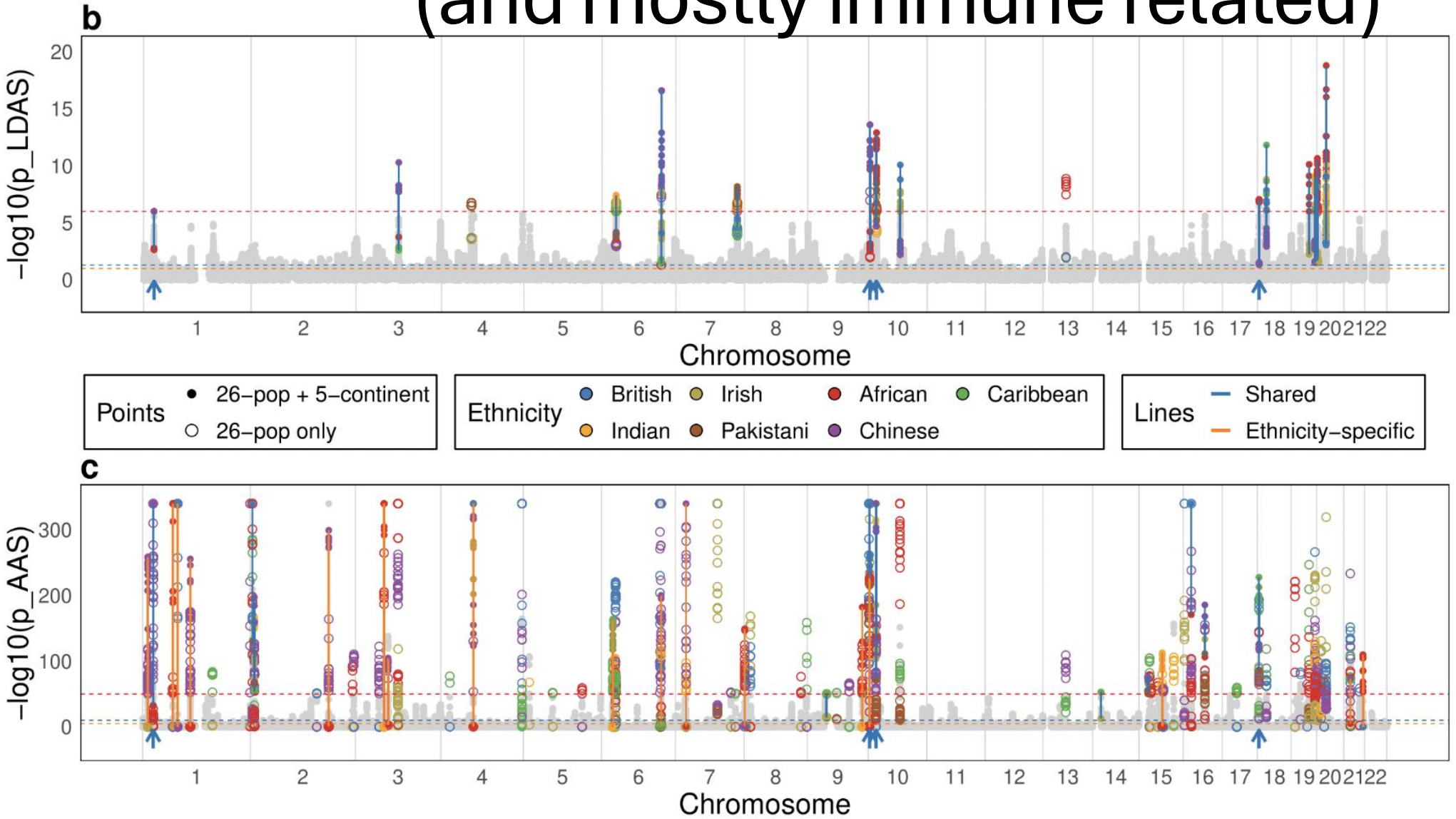


Yang et al. 2024

Very recent balancing selection is detectable (and mostly immune related)

Ancestry segments are "too short"

Ancestry at locus different to genome-wide



Yang et al. 2024 (see also Barrie et al Nature 2024)

Rare variant analysis and sibling design

nature genetics

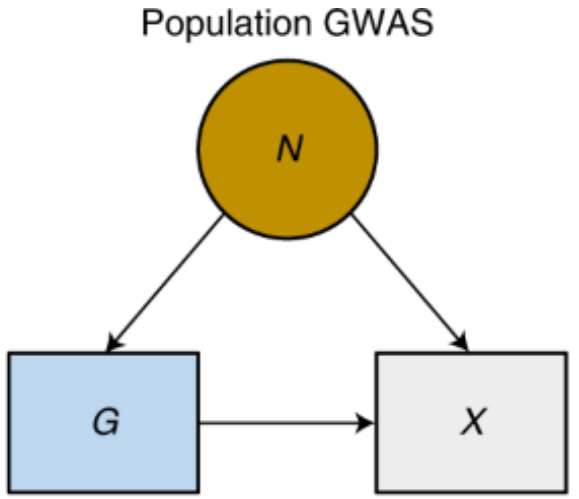
Explore content ⌄ About the journal ⌄ Publish with us ⌄

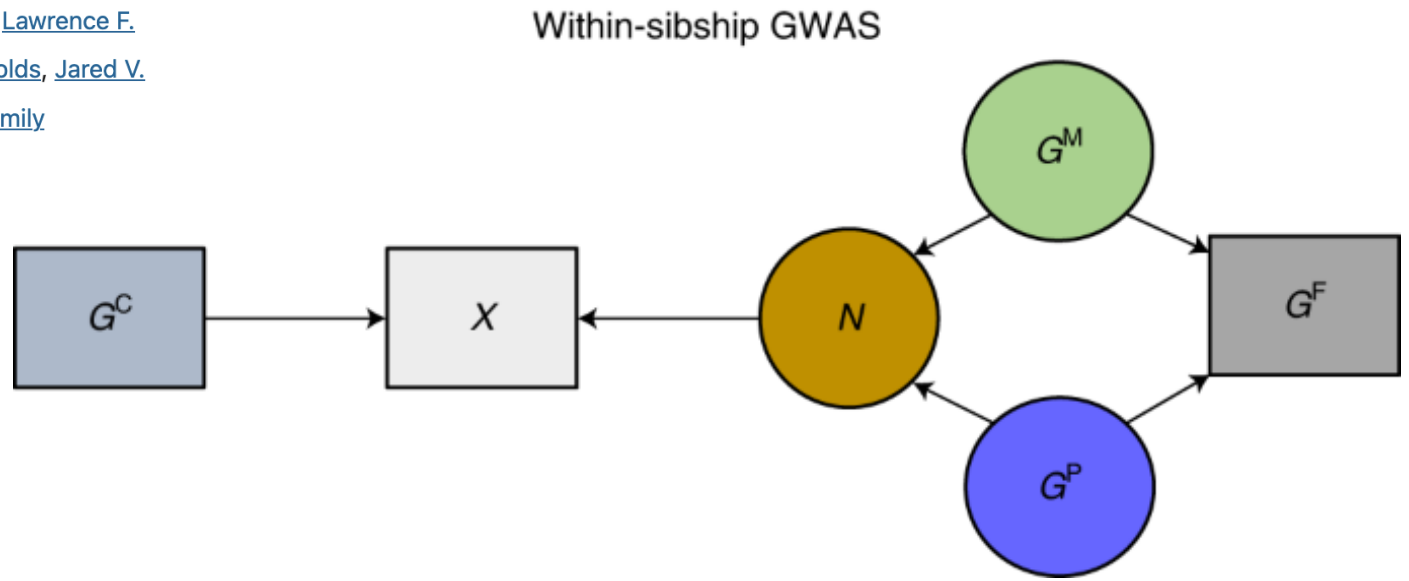
nature › nature genetics › articles › article

Article | Open access | Published: 09 May 2022

Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects

Laurence J. Howe ✉, Michel G. Nivard, Tim T. Morris, Ailin F. Hansen, Humaira Rasheed, Yoonsu Cho, Geetha Chittoor, Rafael Ahlskog, Penelope A. Lind, Teemu Palviainen, Matthijs D. van der Zee, Rosa Cheesman, Massimo Mangino, Yunzhang Wang, Shuai Li, Lucija Klaric, Scott M. Ratliff, Lawrence F. Bielak, Marianne Nygaard, Alexandros Giannelis, Emily A. Willoughby, Chandra A. Reynolds, Jared V. Balbona, Ole A. Andreassen, Social Science Genetic Association Consortium, Within Family Consortium, … Neil M. Davies ✉ + Show authors



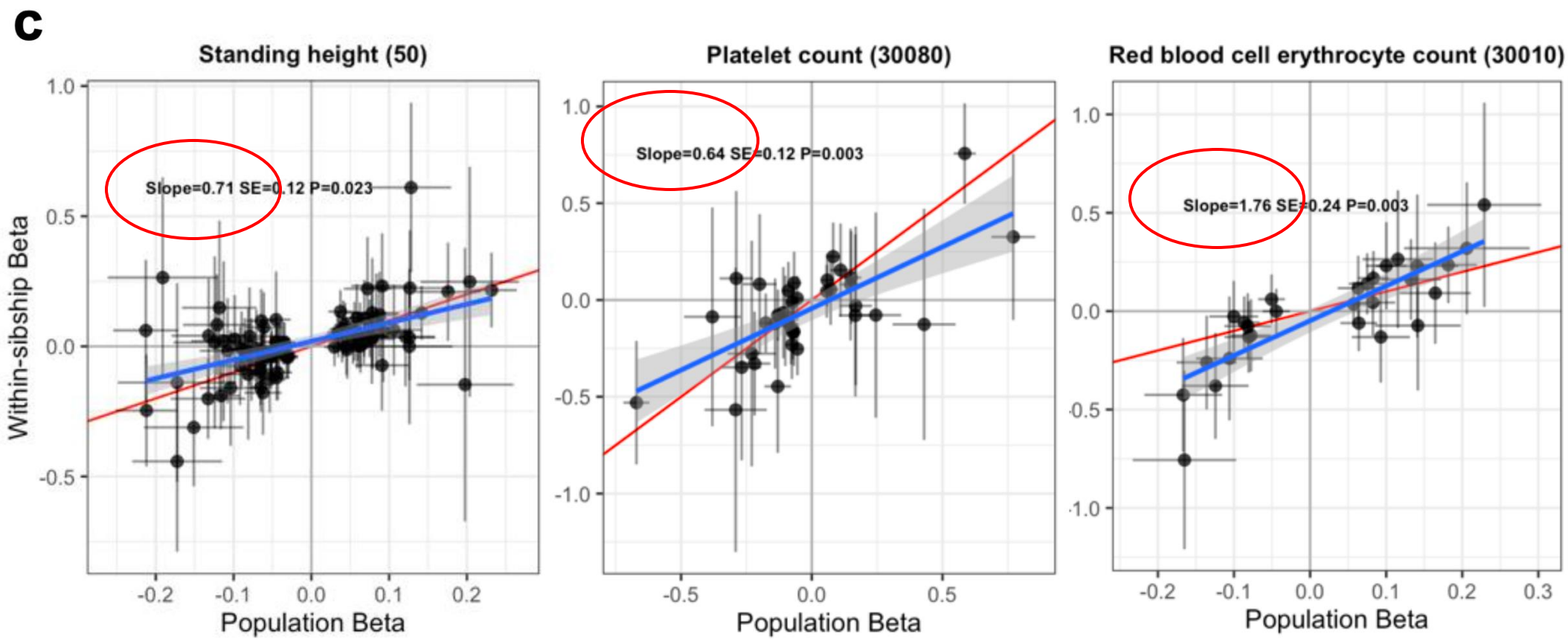


Do sibling estimates replicate in rare SNPs?

Evaluating Confounding in Rare Variant Genome Wide Association Studies

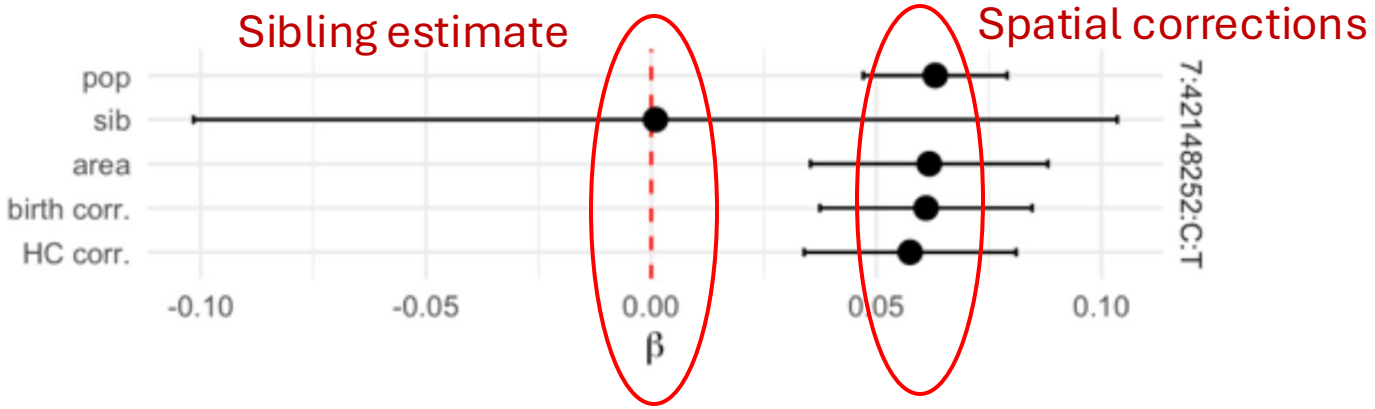
Aimee L. Hanson[1], Gareth J. Griffith[1], Si Fang[1], Neil M. Davies[1], George Davey Smith[1], Daniel J. Lawson[1], Gibran Hemani[1]

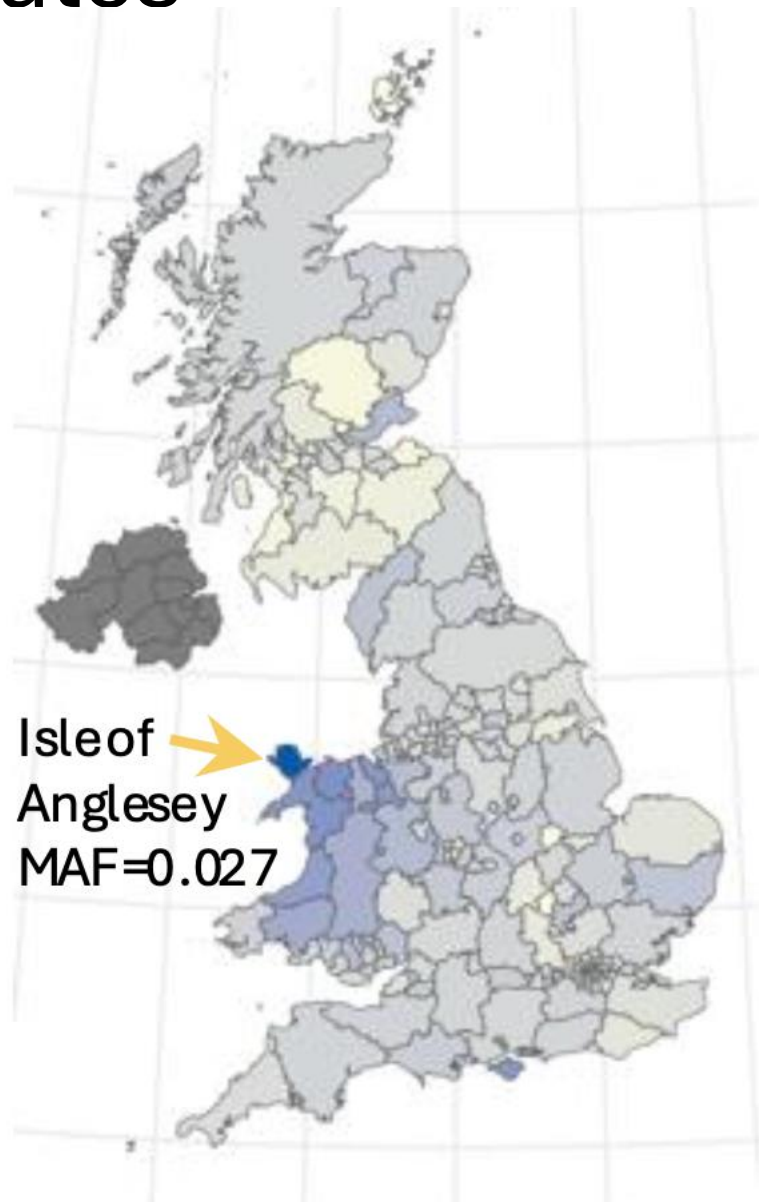
… not really…



Spatial confounding does not explain bias in rare variant height effect estimates

- 116 height associated rare variants from Backman et al. (MAF < 1%, P<5x10-8).
- Many are strongly associated with geography
- Sadly, correcting for geography* doesn't correct the estimate





* With GRMs, haplotypes, rare variant PCs, levels for region, etc

In conclusion

- Genetic risk varies by population
  - We need local ancestry methods to properly attribute cause to loci
- We need more diverse studies to:
  - Identify risk variants for everyone
  - Better identify causal variants
- Some special diseases have been selected recently and population structure is important for them
- However, population structure is not the big confounder:
  - Family, assortative mating, cultural practice, socio-economics…
  - All matter much more and only sibling studies work there
- Conclusion: Large diverse sibling studies please!

References

[1] Yaoling Yang, Richard Durbin, Astrid K. N. Iversen, Daniel J. Lawson, *Sparse haplotype-based fine-scale local ancestry inference at scale reveals recent selection on immune responses*. medRxiv 2024.03.13.24304206.

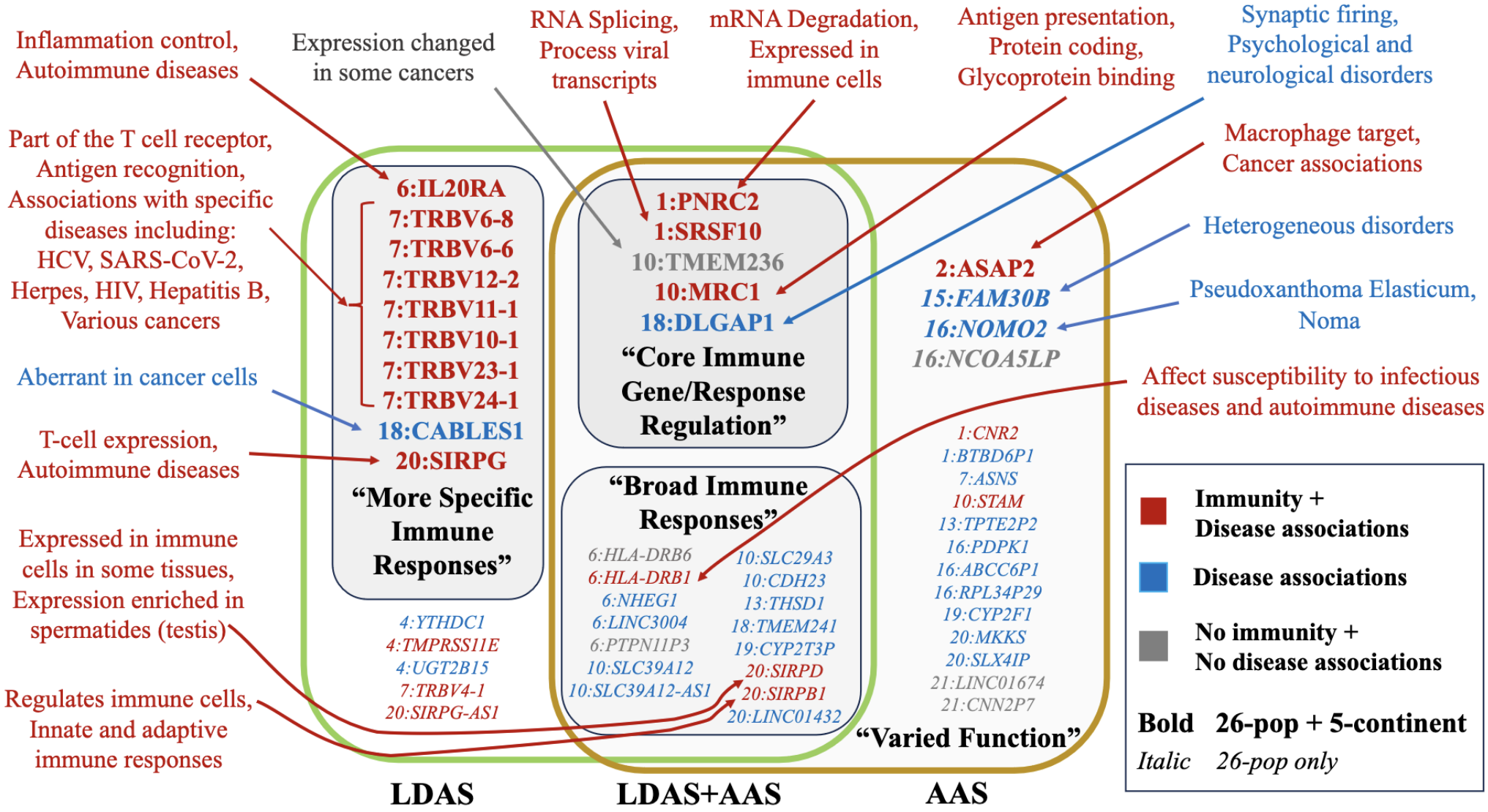
Software: https://github.com/YaolingYang/SparsePainter

[2] Sile Hu, Lino A. F. Ferreira, Sinan Shi, Garrett Hellenthal, Jonathan Marchini, Daniel J. Lawson, Simon R. Myers, *Leveraging fine-scale population structure reveals conservation in genetic effect sizes between human populations across a range of human phenotypes*. bioRxiv 2023.08.08.552281.

[3] Barrie, W., Yang, Y., Irving-Pease, E.K. *et al*. *Elevated genetic risk for multiple sclerosis emerged in steppe pastoralist populations*. *Nature* **625**, 321–328 (2024). https://doi.org/10.1038/s41586-023-06618-z

With special thanks to **Santi Rodriguez** for HaploType Regression with eXtra flexibility (HTRX)

Very recent selection is immune related



Yang et al. 2024 (see also Barrie et al Nature 2024)