Bioinformatics Virtual Core Journal Club

10 May 2023

Bioinformatics and Biostatistics Journal Club

The Bioinformatics and Biostatistics Journal Club is designed to promote learning and intellectual development within DCEG. This Journal Club offers a platform for us to stay current with recent research, explore new findings, and participate in thought-provoking discussions alongside our colleagues.

We invite you to contribute to this dynamic environment by recommending a paper that has captured your interest. Please submit your suggestions at the journal club GitHub repo discussion forum. Your participation is key to maintaining the richness of our discussions and the breadth of our learning. In addition to paper suggestions, we are always looking for volunteers who are willing to lead our discussions or present on topics of interest.

DCEG Bioinformatics and Biostatistics Journal Club Schedule

Meeting 23-05-10

- Title: Using Population Descriptors in Genetics and Genomics Research
- Authors: National Academies of Sciences, Engineering, and Medicine
- Link: https://nap.nationalacademies.org/catalog/26902/using-population-descriptors-in-genetics-and-genomics-research-a-new

Meeting 23-06-14

- Title: TBD - Looking for paper and volunteer!

Latest Posts

Announcements

We are thrilled to introduce our new monthly series, the Bioinformatics and Biostatistics Journal Club, designed to promote learning and intellectual development within DCEG. This Journal Club offers a platform for us to stay current with recent research, explore new findings, and participate in thought-provoking discussions alongside our colleagues.

Our first Journal Club meeting is scheduled for 11-12pm, May 10, 2023.

Read more

Published: May 3, 2023

https://nci-dceg.github.io/bioinformatics_journal_club/

Groundrules

- Be respectful and kind
- We're all here to learn—listen (don't interrupt)
- Remember: each of us only sees part of the elephant

- When in the large group, stay muted when not speaking
- Use "Raise Hand" function if you wish to ask a question live or…
- …type questions into comments section

"Science should benefit all"
"Do no harm"

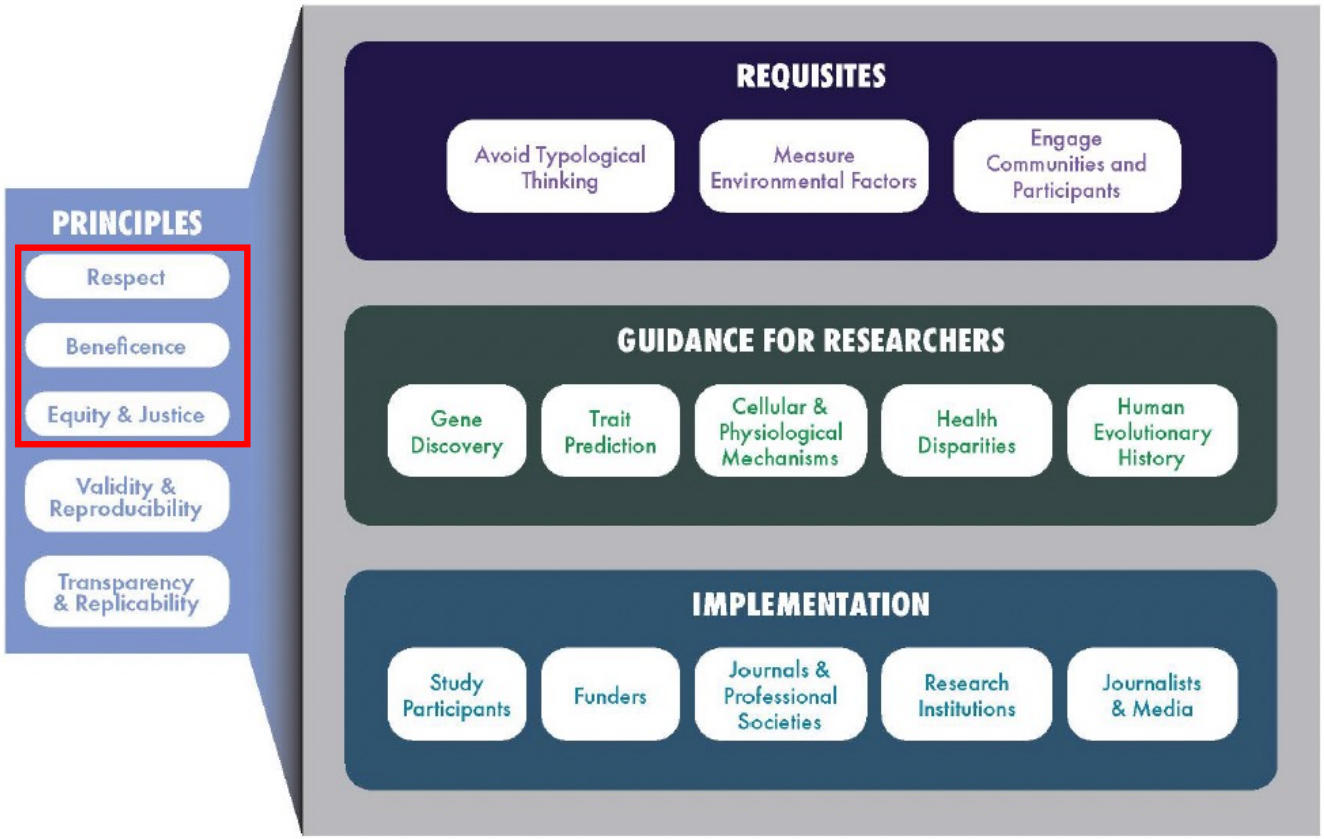


Figure S-1

Setting the agenda in research

Comment

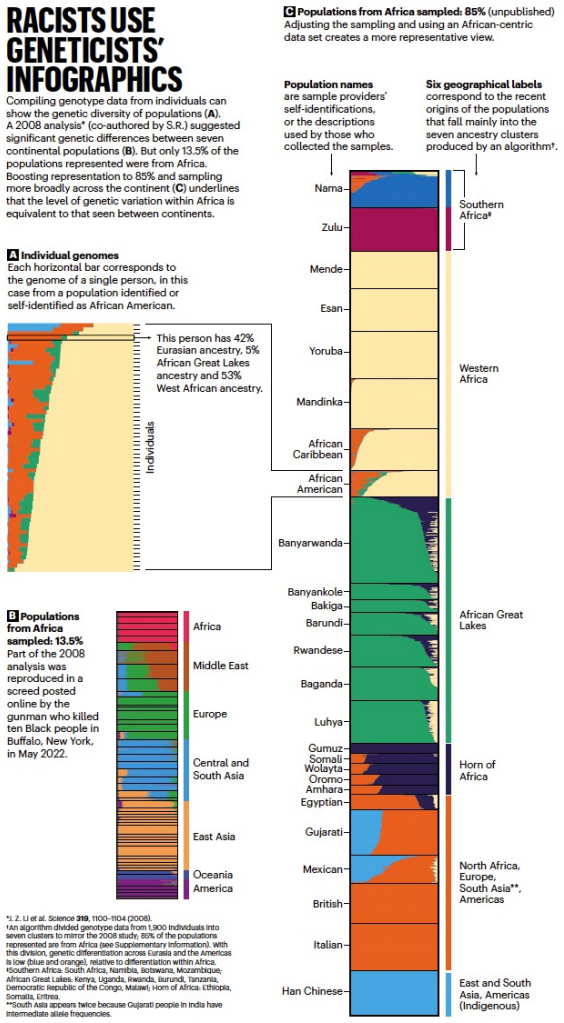
Racism in science

Counter the weaponization of genetics research by extremists

Jedidiah Carlson, Brenna M. Henn, Dana R. Al-Hindi & Sohini Ramachandran

KENNY HOLSTON/NEW YORK TIMES/REDUX/EYEVINE

A memorial to the ten Black people who were killed by a shooter outside a shop in Buffalo, New York, in May 2022.



Carlson (2022) *Nature*

Whether and how to use population descriptors depends on scientific goal

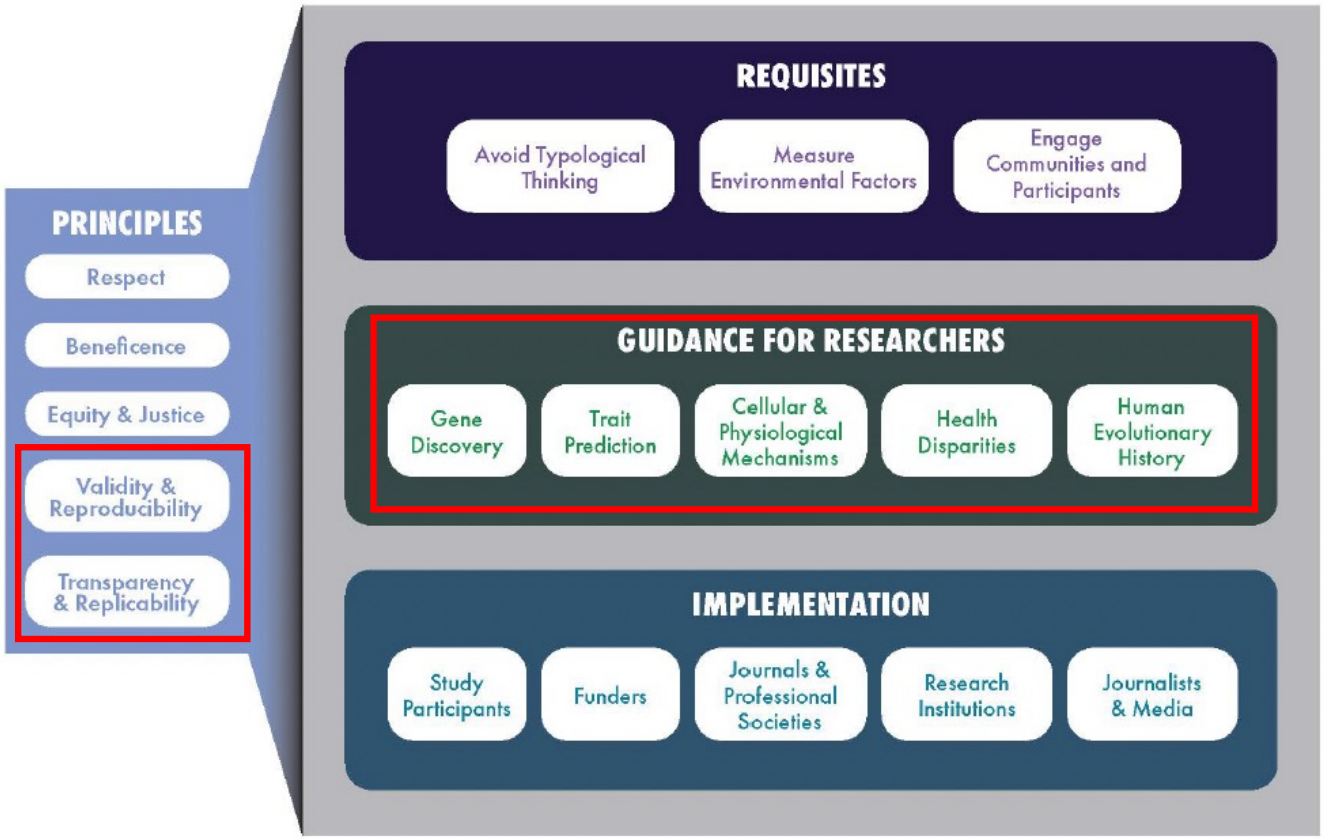


Figure S-1

1. Ancestry, genetic ancestry and genetic similarity

2. Examples from GWAS

LEGEND

+ Preferred population descriptor(s)	− Should not be used
? In some cases; refer to Ch. 5 text and the decision tree in Appendix D	E Descriptors could be used if appropriate proxies for environmental, not genetic, effects

GENOMICS STUDY TYPE	Race	Ethnicity/ Indigeneity	Geography	Genetic Ancestry	Genetic Similarity	Notes
1: Gene Discovery - Mendelian Traits	−	?	?	?	+	Similarity suffices as a genetic measure; at fine-scale, other variables may be useful
2: Trait Prediction - Mendelian Traits	−	E	E	?	+	No population descriptors may be necessary for analysis
3: Gene Discovery - Complex Traits	−	E	E	?	+	Similarity suffices as a genetic measure
4: Trait Prediction - Complex Traits	−	E	E	?	+	Similarity suffices as a genetic measure
5: Cellular and Physiological Mechanisms	−	E	E	−	?	No population descriptors may be necessary for analysis
6: Health Disparities with Genomic Data	E	E	E	?	+	Not all health disparities studies rely on descent-associated population groupings, so none may be necessary for analysis
7: Human Evolutionary History	−	?	+	+	+	Reconstructing genetic ancestry may be of central interest

Table S-1

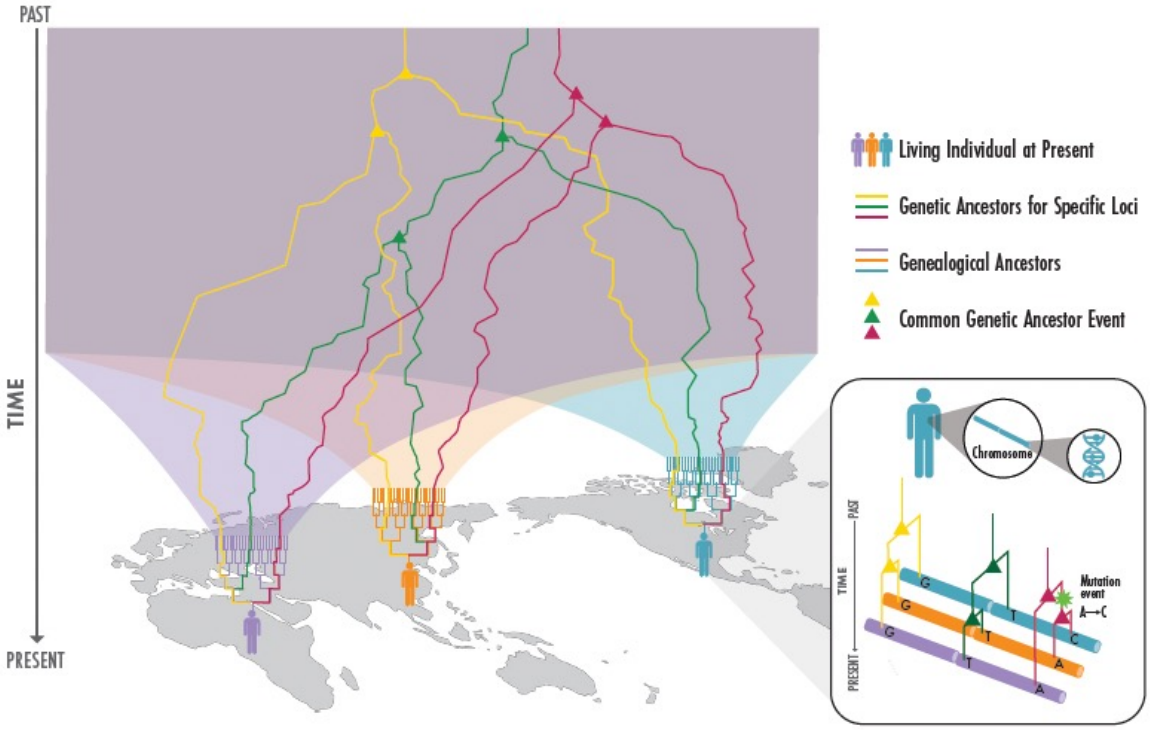
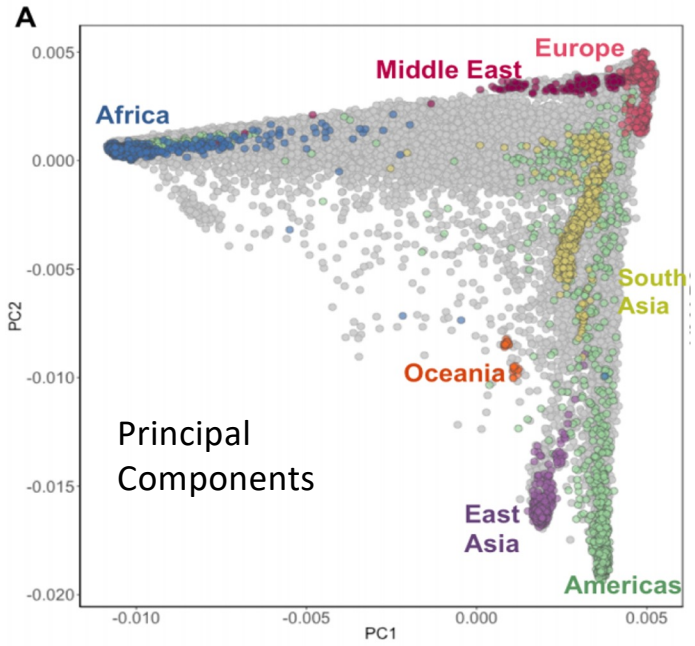


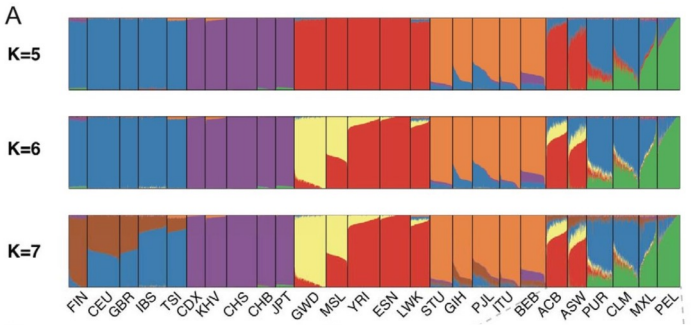
FIGURE 2-1 Visualization of genealogical vs. genetic ancestry. *Genealogical* (pedigree) *ancestors* over time are shown as family tree structures extending into the past (upward) for three present-day individuals sampled from different places around the world. Only a subset of genealogical ancestors contributes genetic material to the living individuals; these *genetic ancestors* and the lines of genetic transmission connecting them are indicated with bolded lines (yellow, green, red) for three exemplar genetic loci (see also inset). Moving back in time, all three individuals eventually share their genealogical ancestors and their genetic ancestors. As indicated, the overlap between family trees that indicate genealogical ancestors happens much more recently than the common ancestor events in the "gene trees" that describe the relationship of genetic ancestors. Who is most closely related genetically to whom varies along a genome, as shown in the inset panel. The inset also demonstrates how mutations that occurred in the past, during the transmission of DNA from one generation to the next, lead to observable changes in DNA today and result in greater or less observed genetic similarity between individuals.

genetic ancestors ≠ ancestors
genetic ancestry is locus- and time-scale specific
genetic ancestry cannot be directly observed

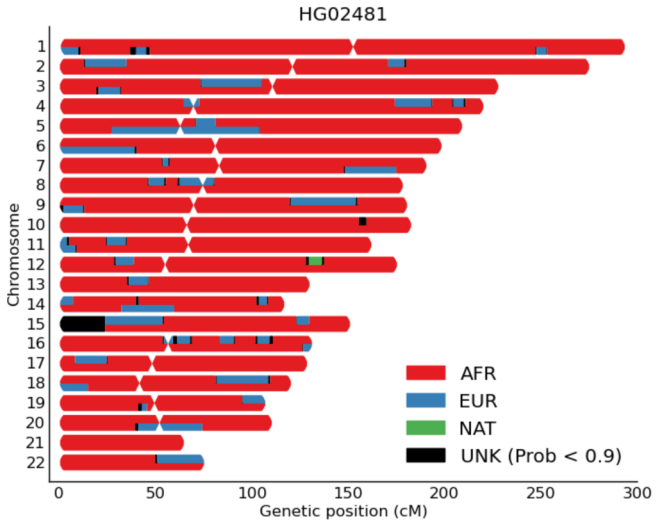
Figure 2-1

Genetic similarity





STRUCTURE



Local Ancestry*

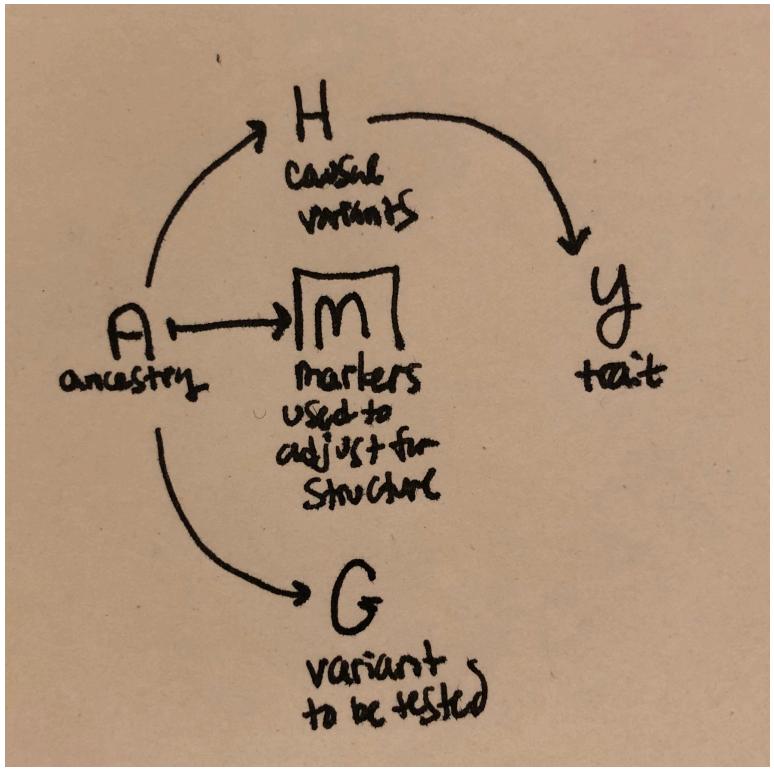
*not really ancestry

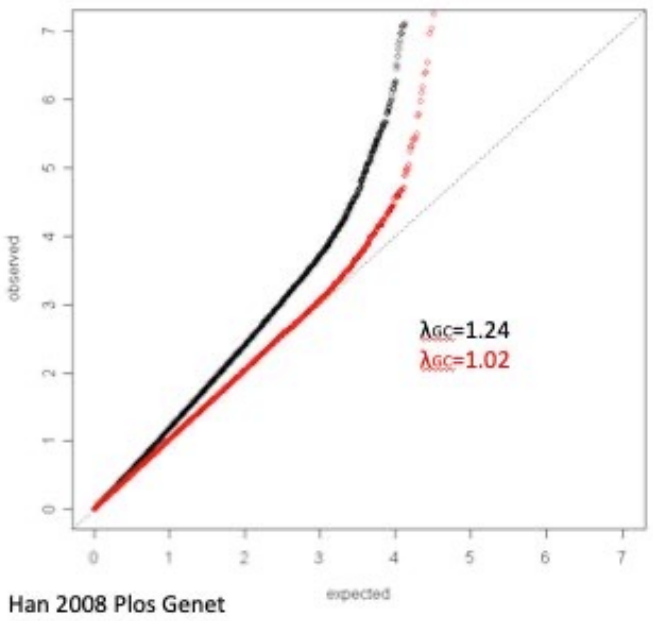
- Does not require subjects be placed in "buckets"
- Emphasizes continuous nature of genetic similarity
- Sample-specific

- Assumes each subject's genome can be portioned into segments drawn from discrete observed (or inferred) reference populations ("buckets")
- Sample- and reference-specific

Belbin (2021) Cell; Lewis (2022) Science; Martin (2017) Am J Hum Genet

Example 1: light-dark natural hair color among European-ancestry women



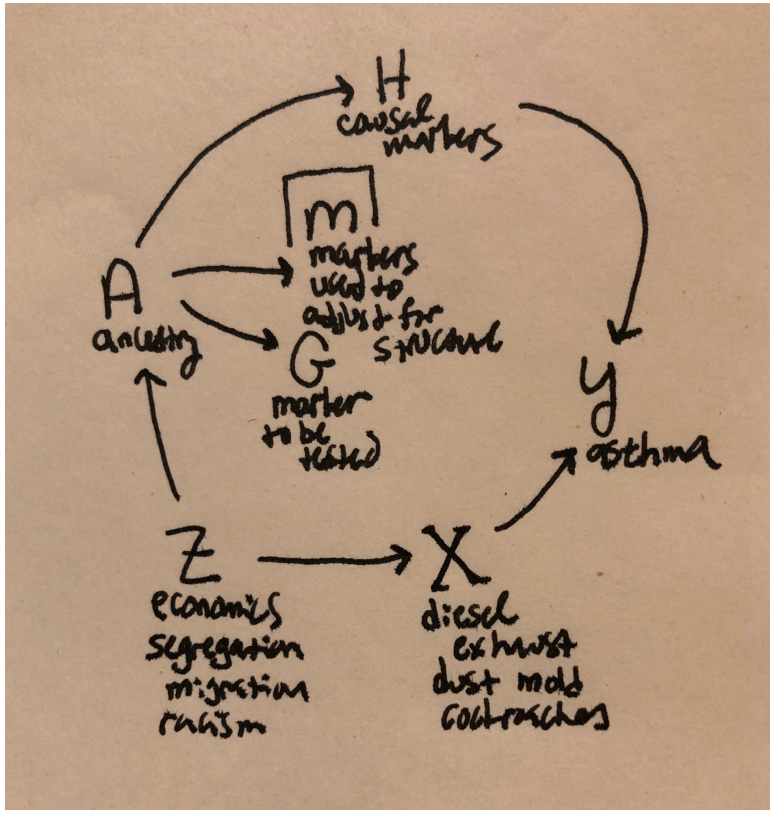


Han 2008 Plos Genet

Quantile-Quantile (QQ) plot for a GWAS of dark-light hair color in US European-ancestry subjects from the Nurses Health Study . The black points are the p-values from the unadjusted tests. The red points are from **principal-component adjusted tests**.

https://twitter.com/GENES_PK/status/1194969256645079041

Example 2: asthma among children in southern California



- Note: There is no arrow from ancestry to **X** (confounding due to **X** is *socially contingent*)
- If we could measure **X** exhaustively and accurately, that would provide better control of confounding
- Where would self-reported race and ethnicity go on this DAG? Could including it improve control of confounding?

https://twitter.com/GENES_PK/status/1194969256645079041

References

1: Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am J Hum Genet. 2017 Apr 6;100(4):635-649. doi: 10.1016/j.ajhg.2017.03.004. Epub 2017 Mar 30.

2: Belbin GM, Cullina S, Wenric S, Soper ER, Glicksberg BS, Torre D, Moscati A, Wojcik GL, Shemirani R, Beckmann ND, Cohain A, Sorokin EP, Park DS, Ambite JL, Ellis S, Auton A; CBIPM Genomics Team; Regeneron Genetics Center; Bottinger EP, Cho JH, Loos RJF, Abul-Husn NS, Zaitlen NA, Gignoux CR, Kenny EE. Toward a fine-scale population health monitoring system. Cell. 2021 Apr15;184(8):2068-2083.e11. doi: 10.1016/j.cell.2021.03.034. PMID: 33861964.

3: Lewis ACF, Molina SJ, Appelbaum PS, Dauda B, Di Rienzo A, Fuentes A, Fullerton SM, Garrison NA, Ghosh N, Hammonds EM, Jones DS, Kenny EE, Kraft P, Lee SS, Mauro M, Novembre J, Panofsky A, Sohail M, Neale BM, Allen DS. Getting genetic ancestry right for science and society. Science. 2022 Apr 15;376(6590):250-252. doi: 10.1126/science.abm7530. Epub 2022 Apr 14. PMID:35420968; PMCID: PMC10135340.

4: Howe CJ, Bailey ZD, Raifman JR, Jackson JW. Recommendations for Using Causal Diagrams to Study Racial Health Disparities. Am J Epidemiol. 2022 Nov 19;191(12):1981-1989. doi: 10.1093/aje/kwac140. PMID: 35916384; PMCID: PMC10144617.

5: Carlson J, Henn BM, Al-Hindi DR, Ramachandran S. Counter the weaponization of genetics research by extremists. Nature. 2022 Oct;610(7932):444-447. doi: 10.1038/d41586-022-03252-z. PMID: 36261568.