

# Add Health Genetic Data and Genome- Wide Association Study

Brandt Levitt Ph.D.

[addhealth\\_genetics@unc.edu](mailto:addhealth_genetics@unc.edu)



Credit : Lauren Solomon, Broad Communications

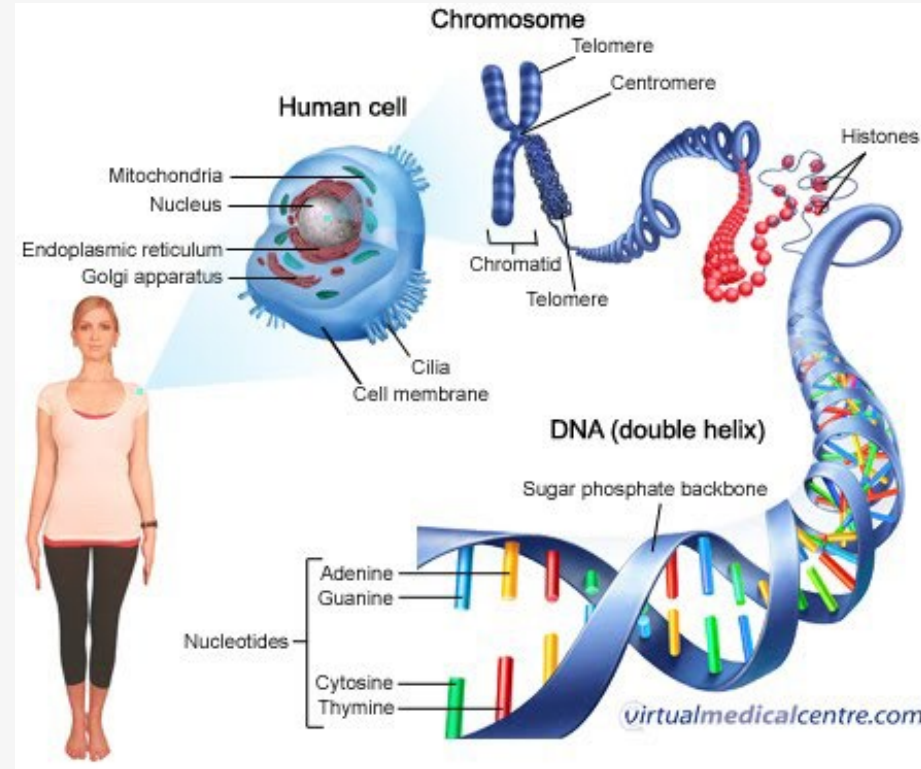
# Outline

- GWAS background
- Rationale
- Conducting a GWAS
- Recent examples
- AddHealth genotyping data
- Potential uses
- Resources

# Outline

- **GWAS background**
- Rationale
- Conducting a GWAS
- Recent examples
- AddHealth genotyping data
- Potential uses
- Resources

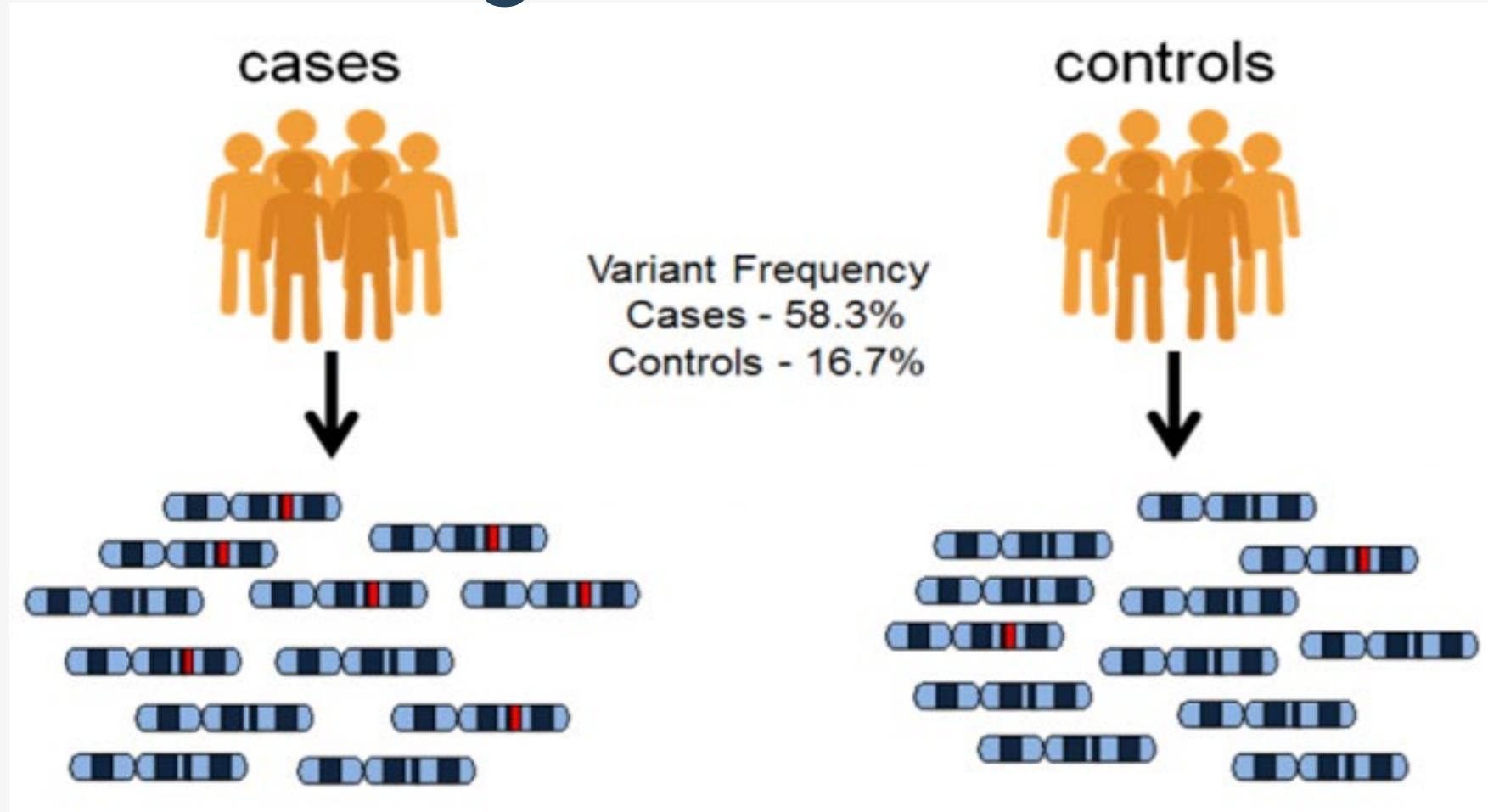
# Background - DNA



# Background - SNPs

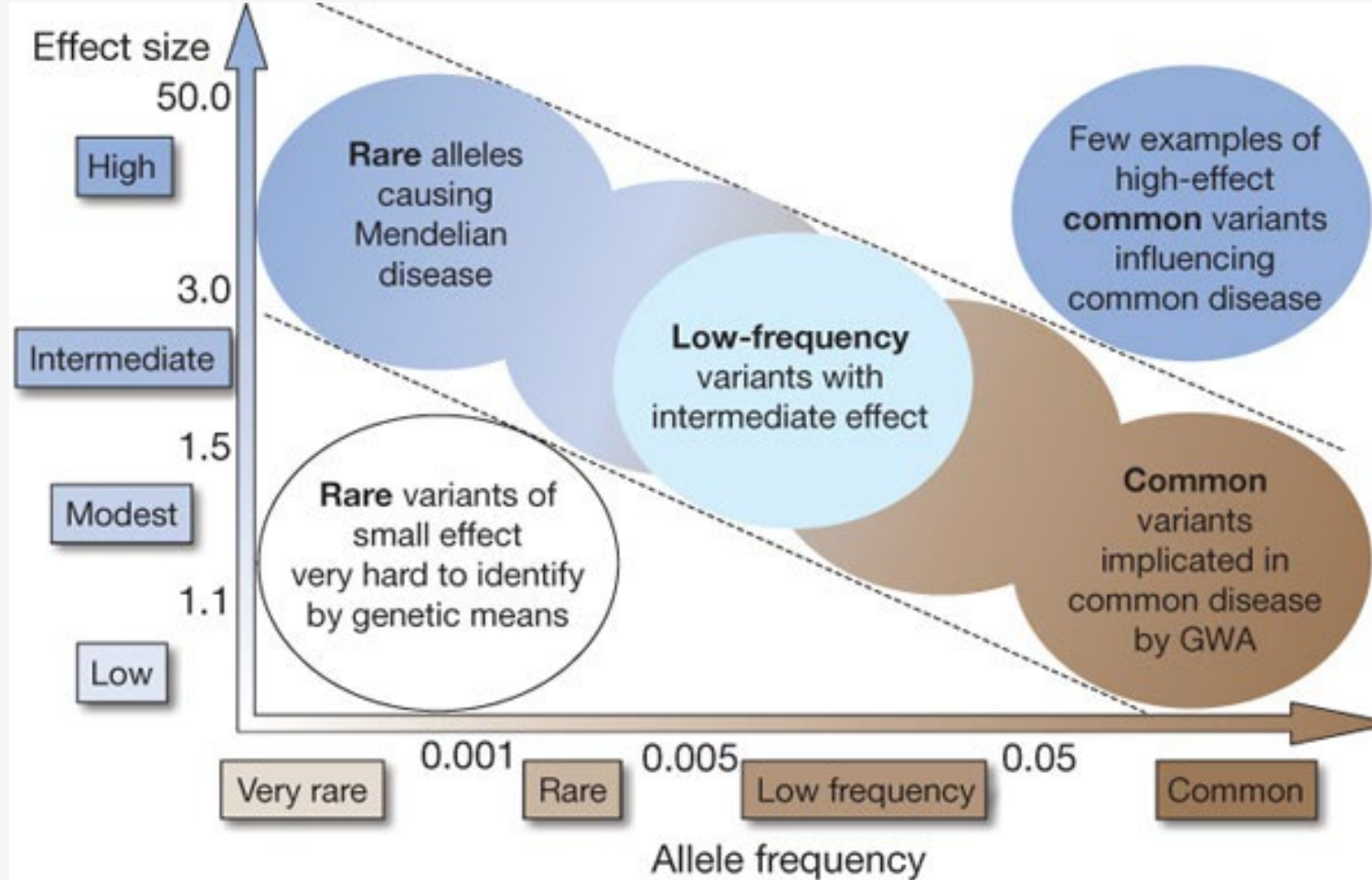
Chromosome 19	44911184...44911194.....44911204	
Consensus	...CGATATTCCATCGAATGTC...	Chr19:44911194C
Individual 1a	...CGATATTCCATCGAATGTC...	Chr19:44911194C>T
Individual 1b	...CGATATTCCATCGAATGTC...	Chr19:44911194C
Individual 1		Chr19:44911194 T/C
Individual 2a	...CGATATTCCATCGAATGTC...	Chr19:44911194C
Individual 2b	...CGATATTCCATCGAATGTC...	Chr19:44911194C
Individual 2		Chr19:44911194 C/C
Individual 3a	...CGATATTCCATCGAATGTC...	Chr19:44911194C>T
Individual 3b	...CGATATTCCATCGAATGTC...	Chr19:44911194C>T
Individual 3		Chr19:44911194 T/T
Individual 4a	...CGATATTCCATCGAATGTC...	Chr19:44911194C
Individual 4b	...CGATATTCCATCGAATGTC...	Chr19:44911194C
Individual 4		Chr19:44911194 C/C
Individual 5a	...CGATATTCCATCGAATGTC...	Chr19:44911194C
Individual 5b	...CGATATTCCATCGAATGTC...	Chr19:44911194C
Individual 5		Chr19:44911194 C/C

# Background - GWAS

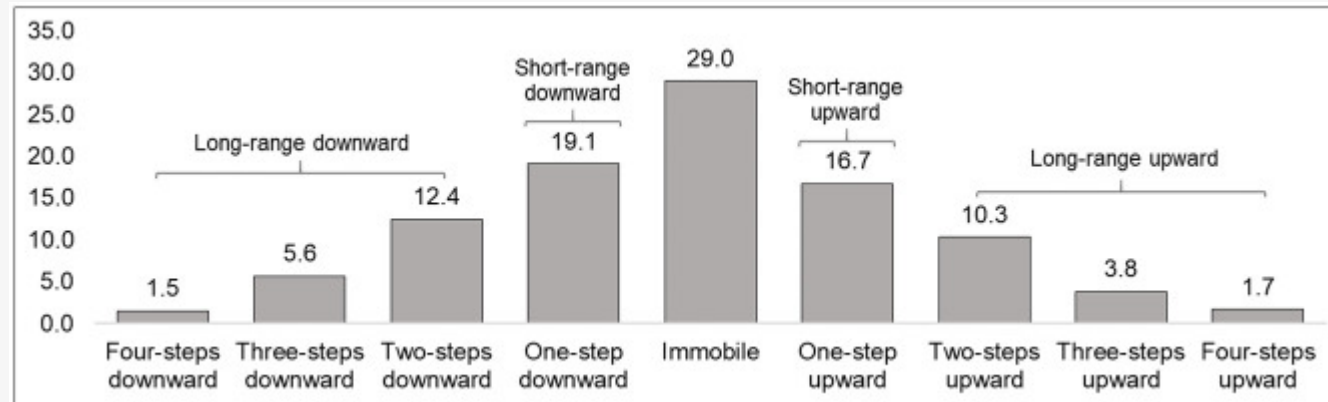




# Background – SNP Frequency and Effect



# Background – Sample Size



Gugushvili A, Bulczak G, Zelinska O, Koltai J. Socioeconomic position, social mobility, and health selection effects on allostatic load in the United States. PLoS One. 2021 Aug 4;16(8):





# Background – Sample Size

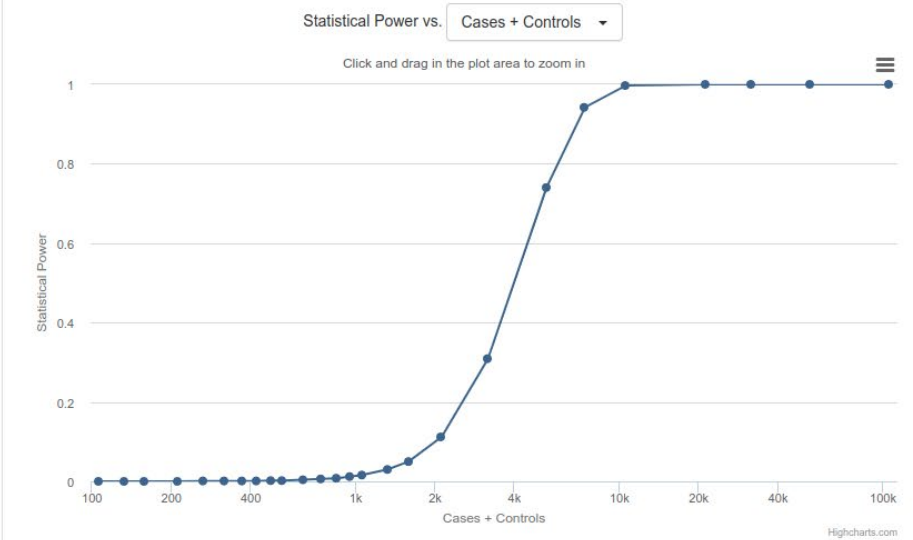
## About GAS Power Calculator

This Genetic Association Study (GAS) Power Calculator is a simple interface that can be used to compute statistical power for large one-stage genetic association studies. The underlying method is derived from the [CaTS](#) power calculator for two-stage association studies (2006).

## Inputs

<b>Sample Size</b>		Cases/Controls = 0.058
Cases	<input type="range"/>	<input type="text" value="550"/>
Controls	<input type="range"/>	<input type="text" value="9450"/>
<b>Study Design</b>		
Significance Level	<input type="range"/>	<input type="text" value="0.0000100"/>
Disease Model	<input type="text" value="Multiplicative"/>	
Prevalence	<input type="range"/>	<input type="text" value="0.0550"/>
Disease Allele Frequency	<input type="range"/>	<input type="text" value="0.5000"/>
Genotype Relative Risk	<input type="range"/>	<input type="text" value="1.5000"/>

## Graph

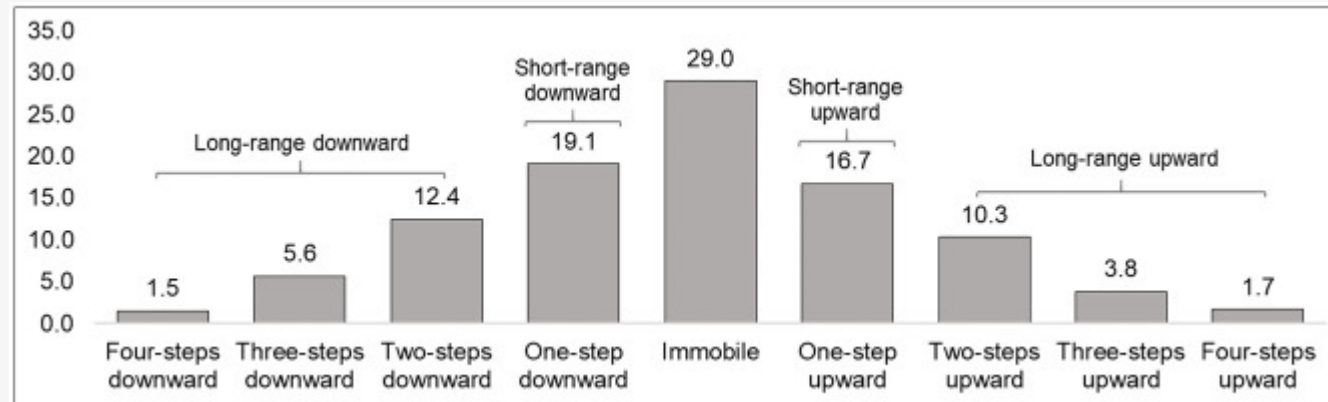


## Results

Expected power for a one-stage study

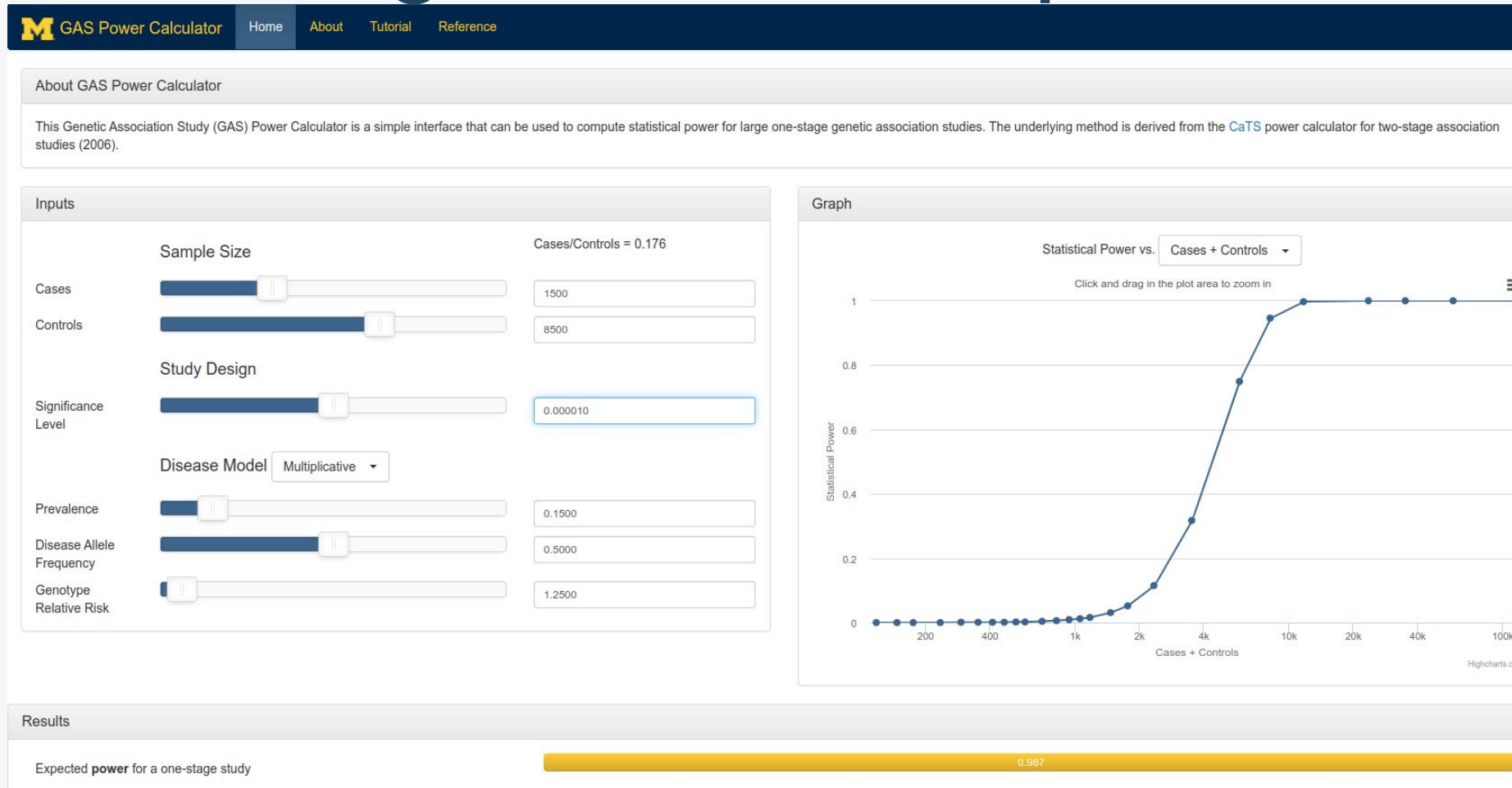
0.994

# Background – Sample Size



Gugushvili A, Bulczak G, Zelinska O, Koltai J. Socioeconomic position, social mobility, and health selection effects on allostatic load in the United States. PLoS One. 2021 Aug 4;16(8):

# Background – Sample Size



# Outline

- GWAS background
- **Rationale**
- Conducting a GWAS
- Recent examples
- AddHealth genotyping data
- Potential uses
- Resources

# Rationale

## Intellectual

- biological contribution to outcomes
- evolutionary past
- family history

## Biomedical

- drug targets
- biomarkers
- disease subtyping

## Public Health

- at risk populations
- preventative practices
- genetic counseling
- Prenatal/newborn screening



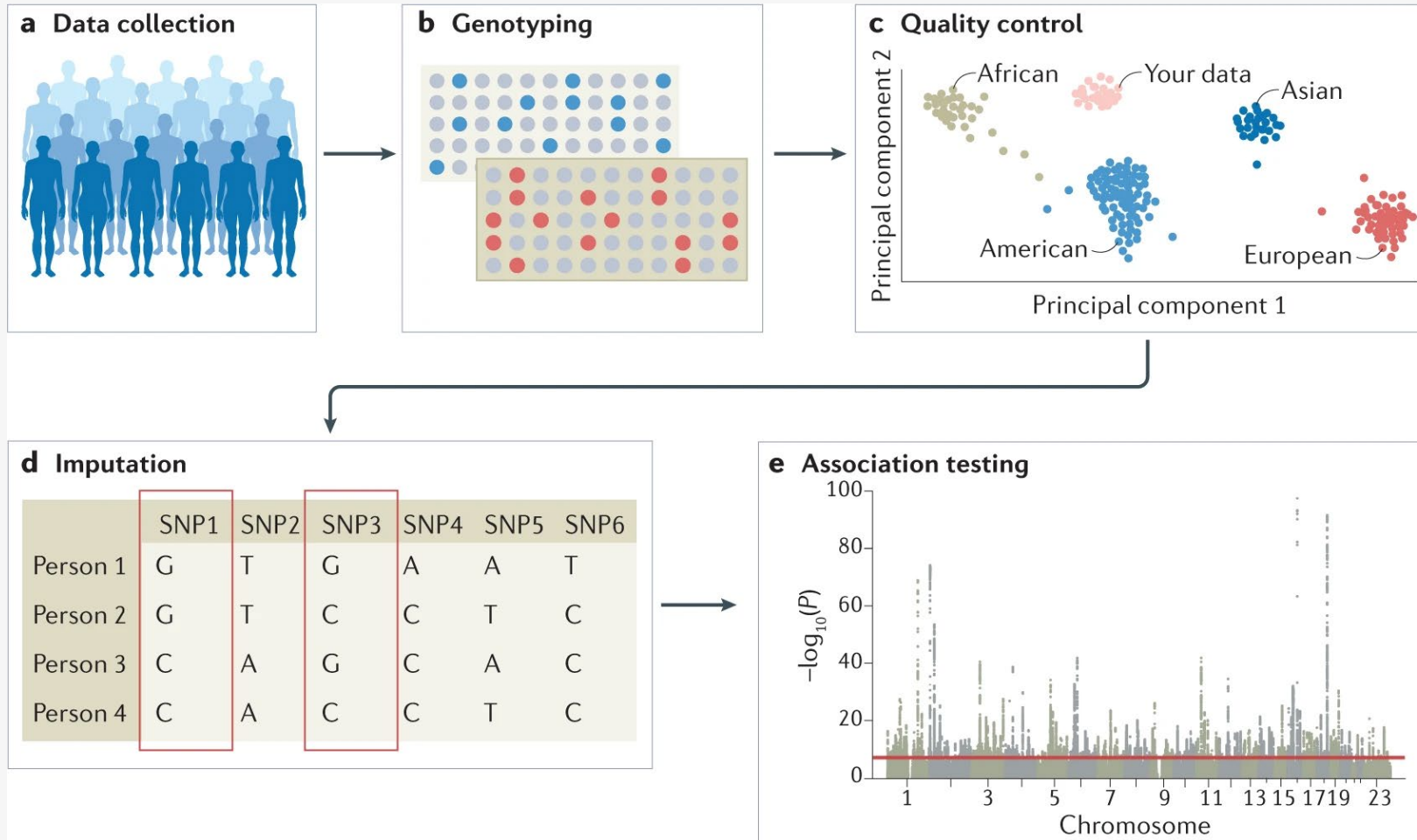
# Rationale

- What **proportion of outcome** is due to subjects' genes?
- **Which populations** are most likely to develop outcome?
- Are there **mitigating biological factors** that limit the connection between my exposure and outcome of interest?
- Does my outcome of interest **share genetic predisposition** to other outcomes?
- Does the **environment** of my subjects conspire with genetic elements in an unpredictable way?

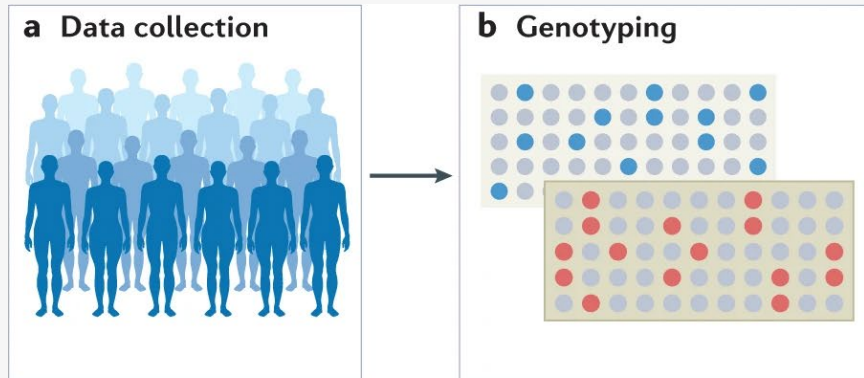
# Outline

- GWAS background
- Rationale
- **Conducting a GWAS**
- Recent examples
- AddHealth genotyping data
- Potential uses
- Resources

# Conducting a GWAS - Outline



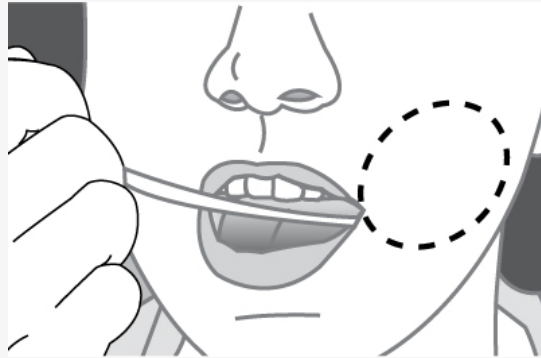
# Conducting a GWAS - Outline



# Conducting a GWAS – Data Collection



Participants  
Consented



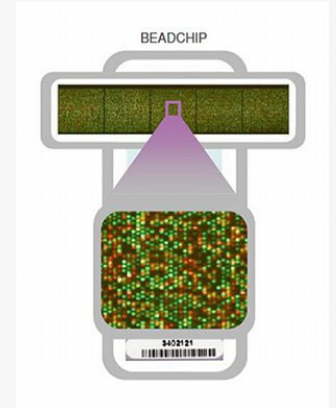
Saliva  
Collection



DNA  
Extraction

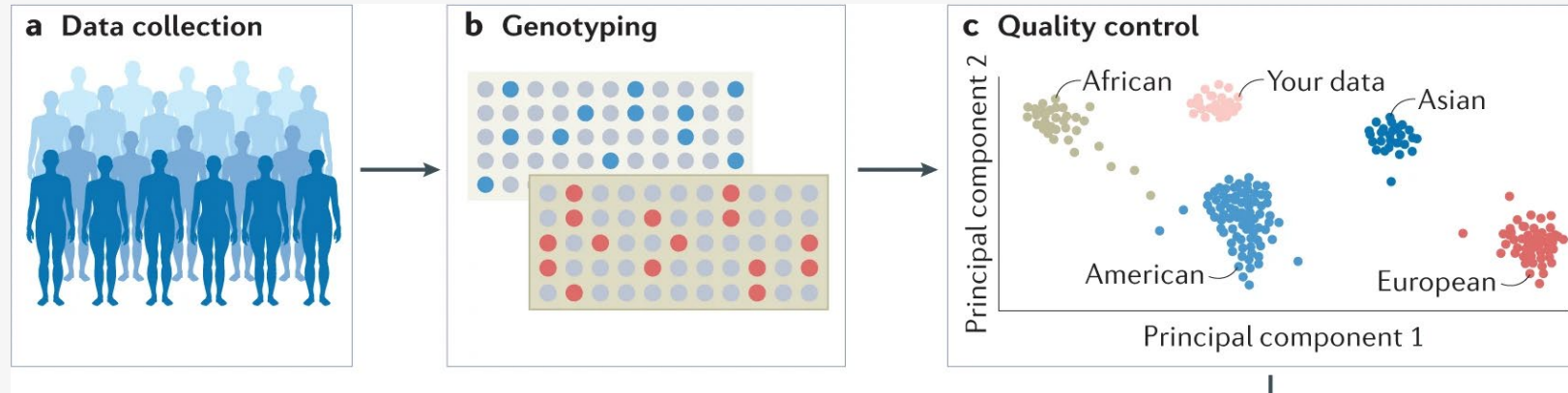


Genotyping  
Chip



Data  
Imaging

# Conducting a GWAS - Outline





# Conducting a GWAS – Sample QC

- Remove participants with **call rate <90%** (n=373)
- Remove participants with **poor duplicate concordance** (n=14)
- Remove participants with **sex discordance** or unclear sex chromosomes (n=177)
- Remove participants with **identifier abnormalities** (n=3)
- **Yielded** n=11554 participants passing quality control

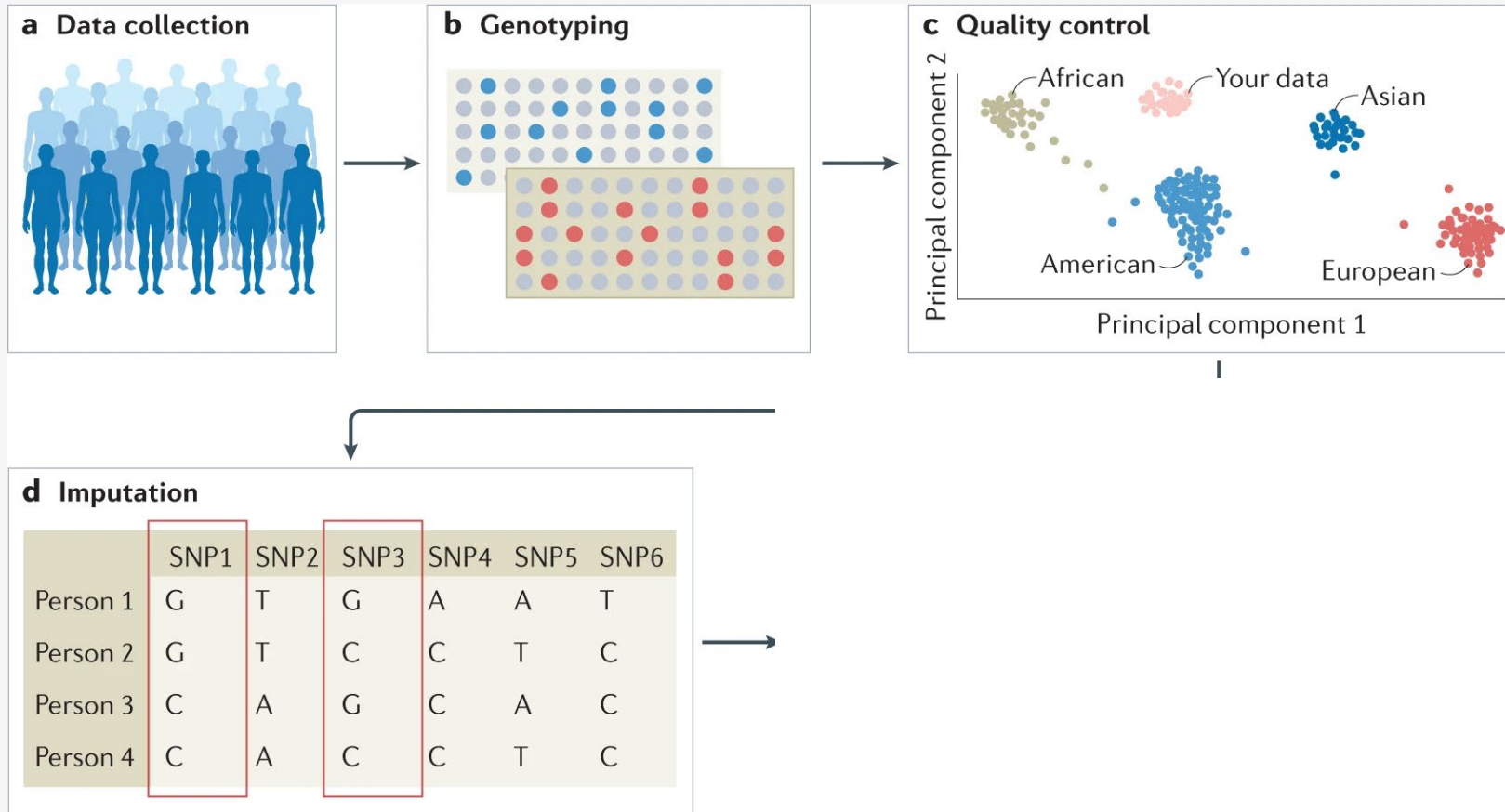
# Conducting a GWAS – Sample QC

- Variants on **Omni1** = 1,058,111 / **Omni2.5** = 2,369,541
- Variants removed for **failure to map** (n=142,695)
- Variants removed for **triallelic** status (n=1,985)
- Variants removed for **low call rate** (n=20,008 / 56,182)
- Variants removed for **rarity** (n=33,333 / 571,200)
- Variants removed for **HWE** (n=7953 / 4,607)
- Variants removed for **duplicate discordance**
- **Yielded** n=609,130 variants

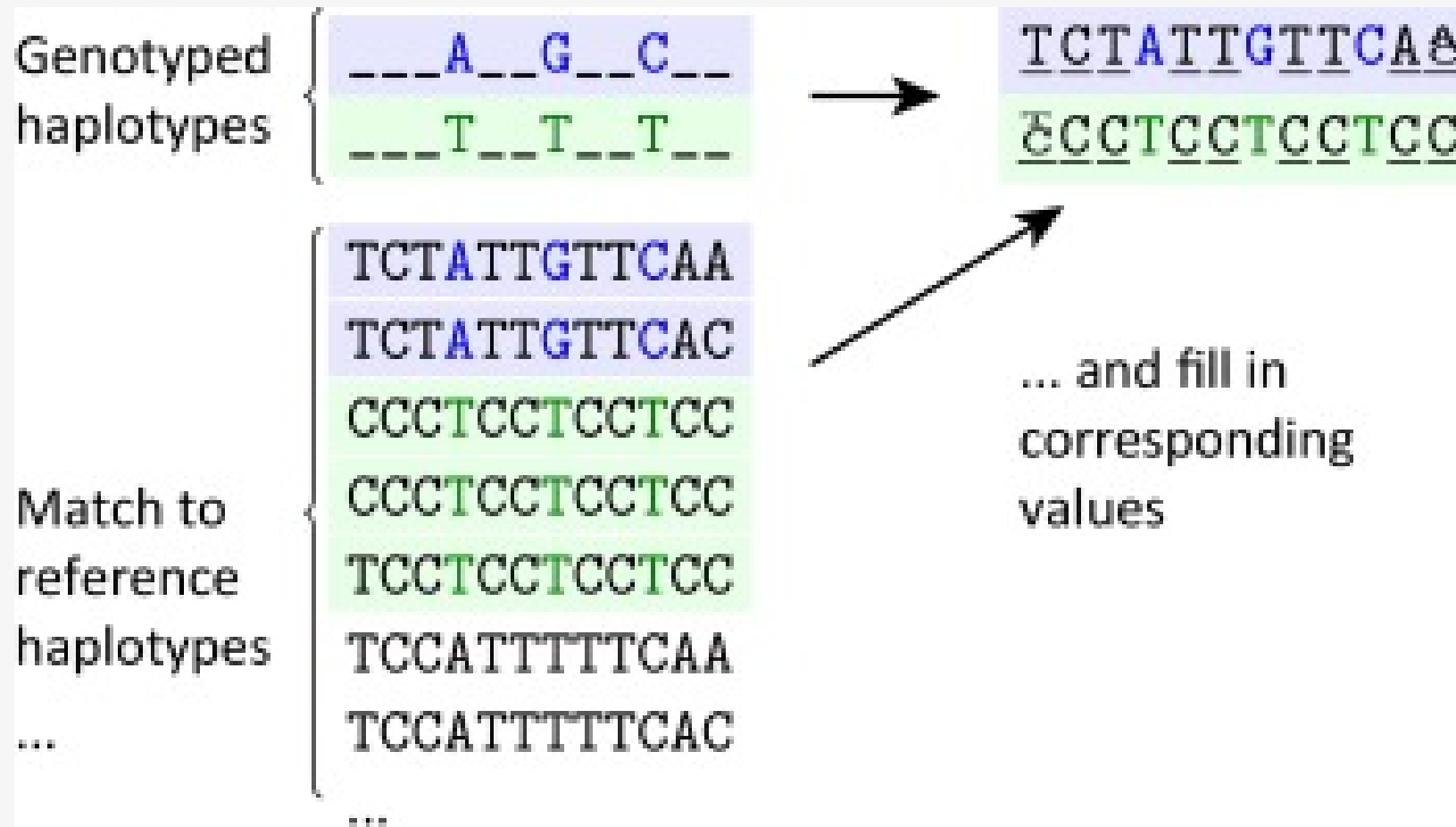
# Conducting a GWAS - Ancestry



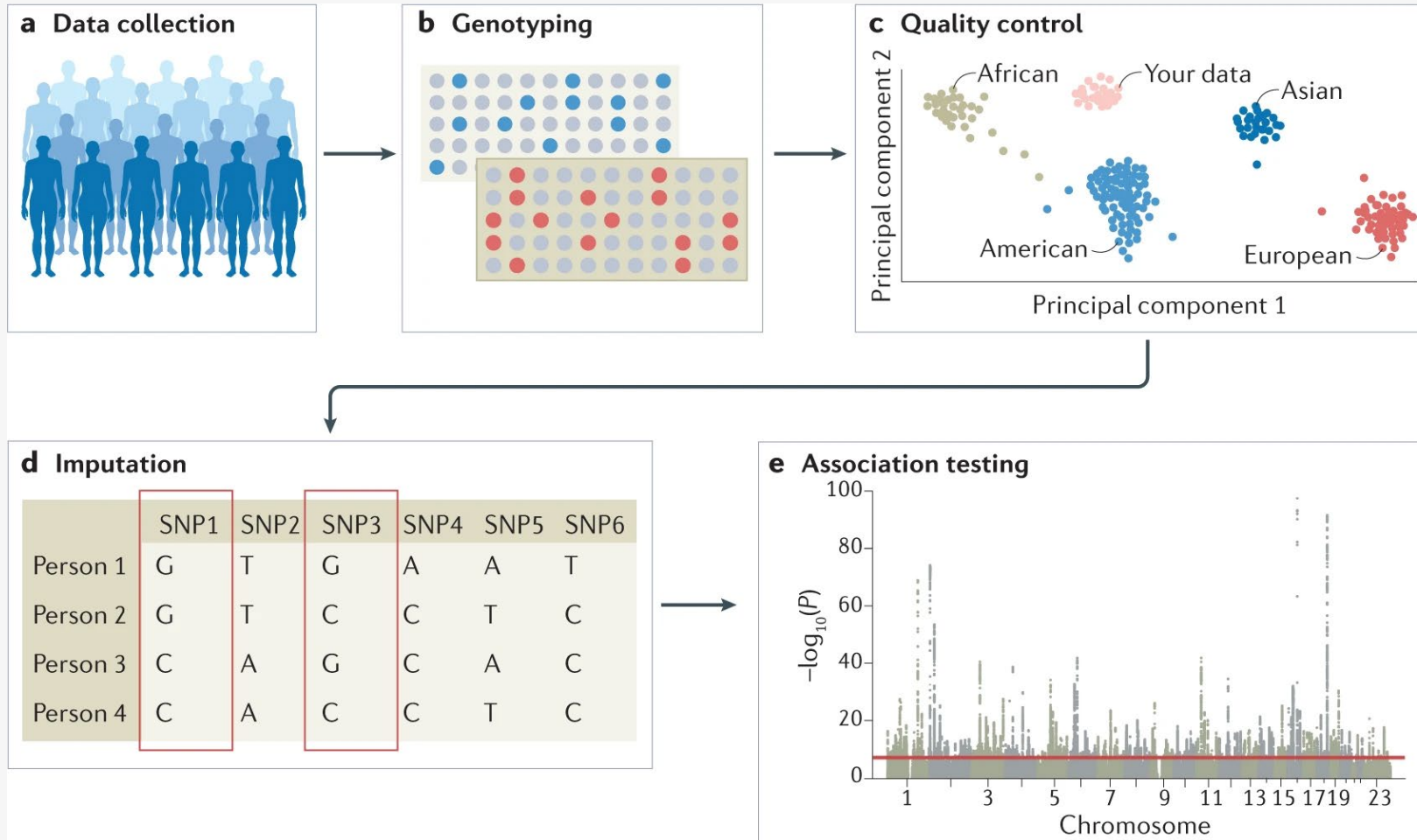
# Conducting a GWAS - Outline



# Conducting a GWAS - Imputation



# Conducting a GWAS - Outline



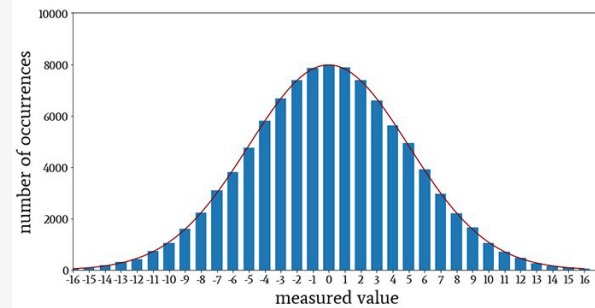


# Conducting a GWAS - Outcome

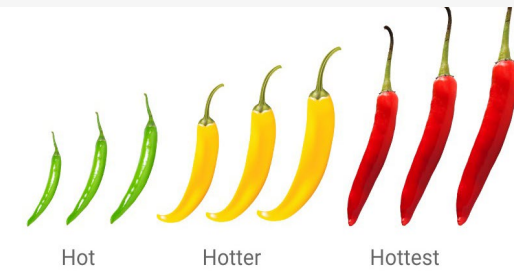
## Dichotomous



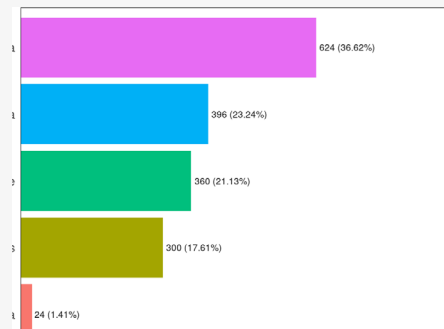
## Quantitative



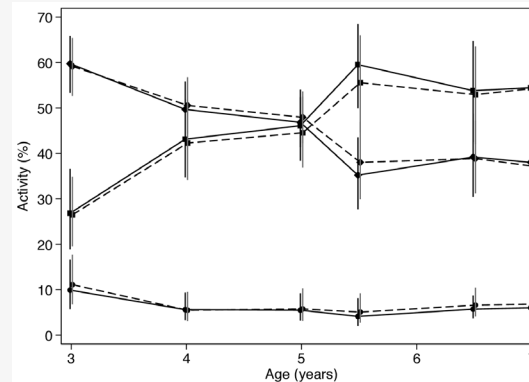
## Ordinal



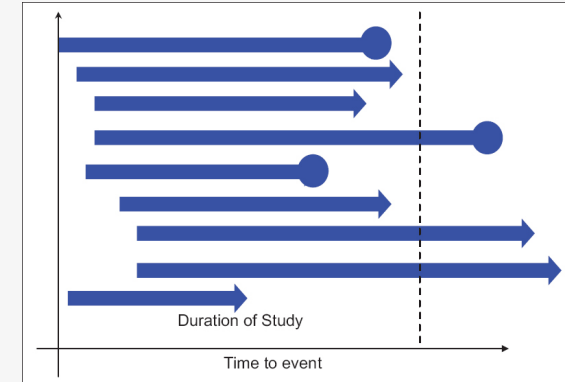
## Categorical



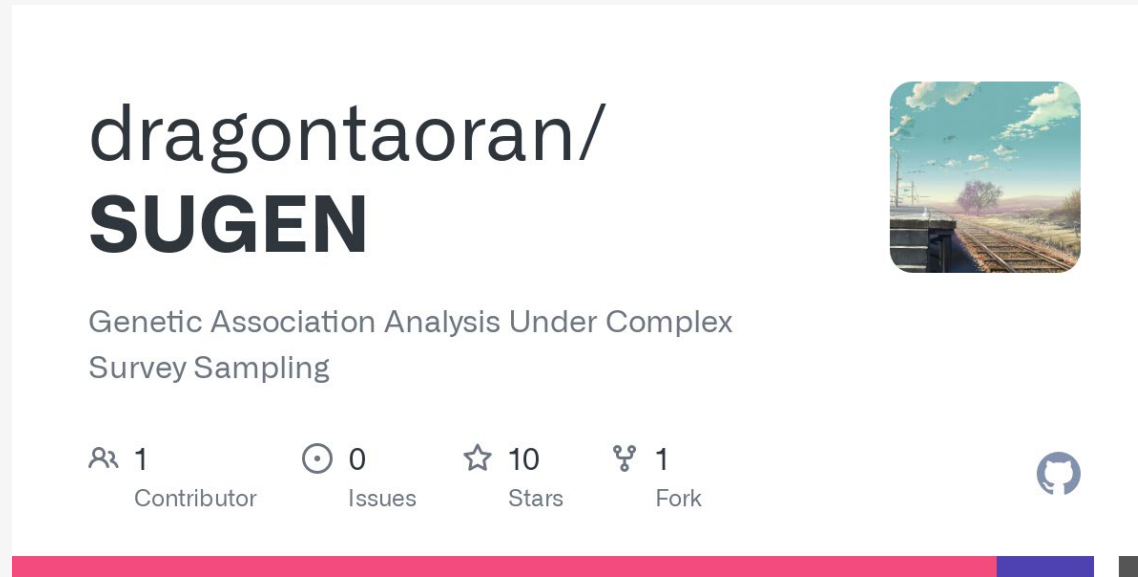
## Longitudinal



## Time to Event



# Conducting a GWAS – Finding Associations



- Allows for family correction
- Allows for heteroscedastic variance flags
- Accommodates GWAS, GWEIS

# Conducting a GWAS – Results

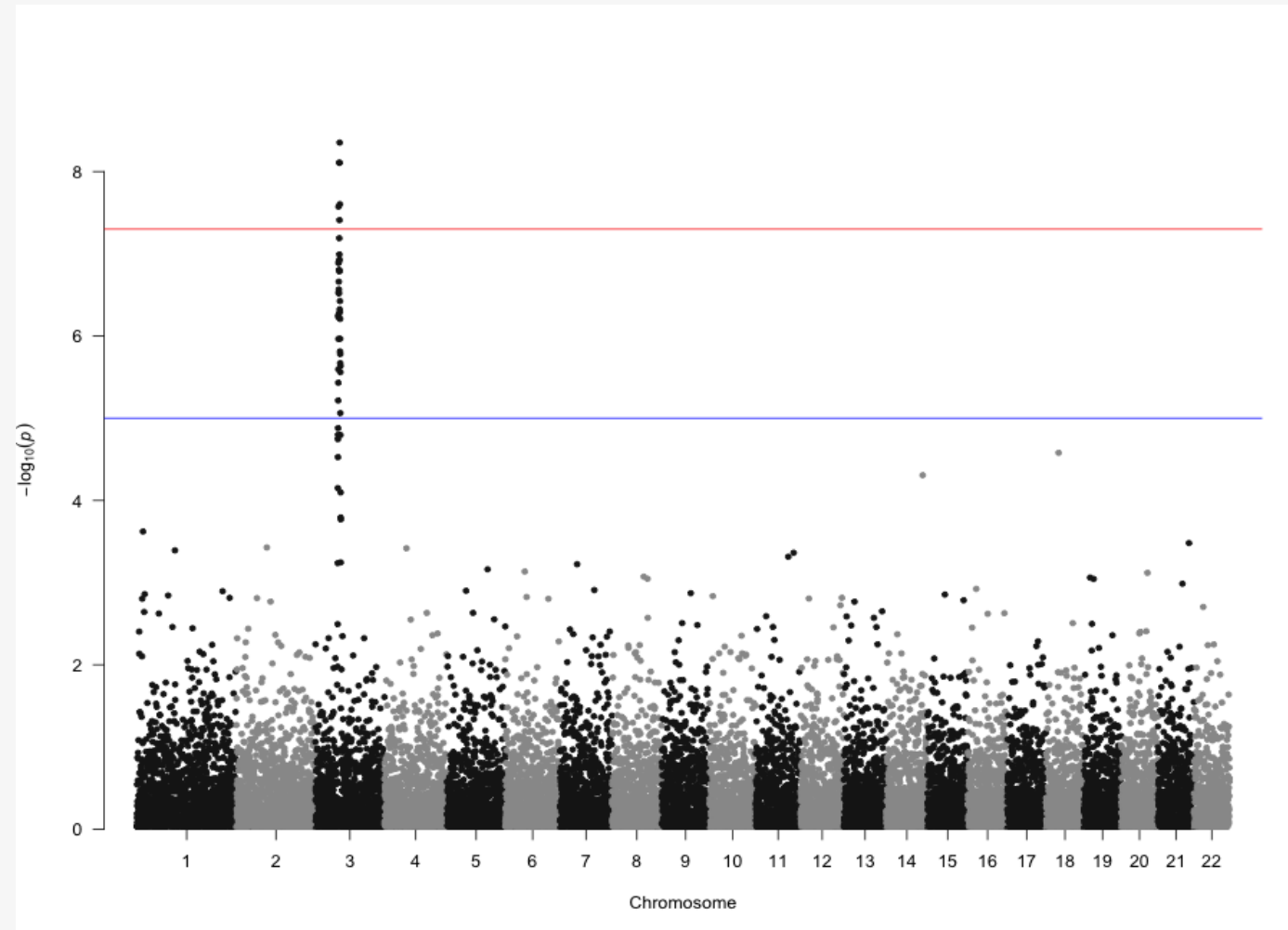
CHROM	POS	VCF_ID	REF	ALT	ALT_AF	N_INFORMATIVE	BETA	SE	PVALUE
22	49262800	22:49262800	T	C	0.27	9939	0.34	0.087	9.9e-05
22	23266118	22:23266118	A	G	0.50	9939	-0.14	0.038	1.4e-04
22	50460076	22:50460076	G	T	0.10	9939	0.25	0.066	2.0e-04
22	23268052	22:23268052	A	G	0.50	9939	-0.14	0.039	2.7e-04
22	19846611	22:19846611	CT	C	0.10	9939	-0.34	0.095	2.8e-04
22	32729773	22:32729773	C	T	0.57	9939	-0.14	0.040	3.4e-04

- Results filtered for
- Minor allele frequency
- Imputation quality
- Effective population size

# Conducting a GWAS – Interpretation

- **Number** of significant hits (n)
- Multiple hits in a single gene (**clustering** of hits)
- Degree of **significance** (p value)
- **Strength** of association (beta)
- Inclusion of hits in **gene set**
- **Functional effect** of mutation

# Conducting a GWAS – Visualization



# Outline

- GWAS background
- Rationale
- Conducting a GWAS
- **Recent examples**
- AddHealth genotyping data
- Potential uses
- Resources

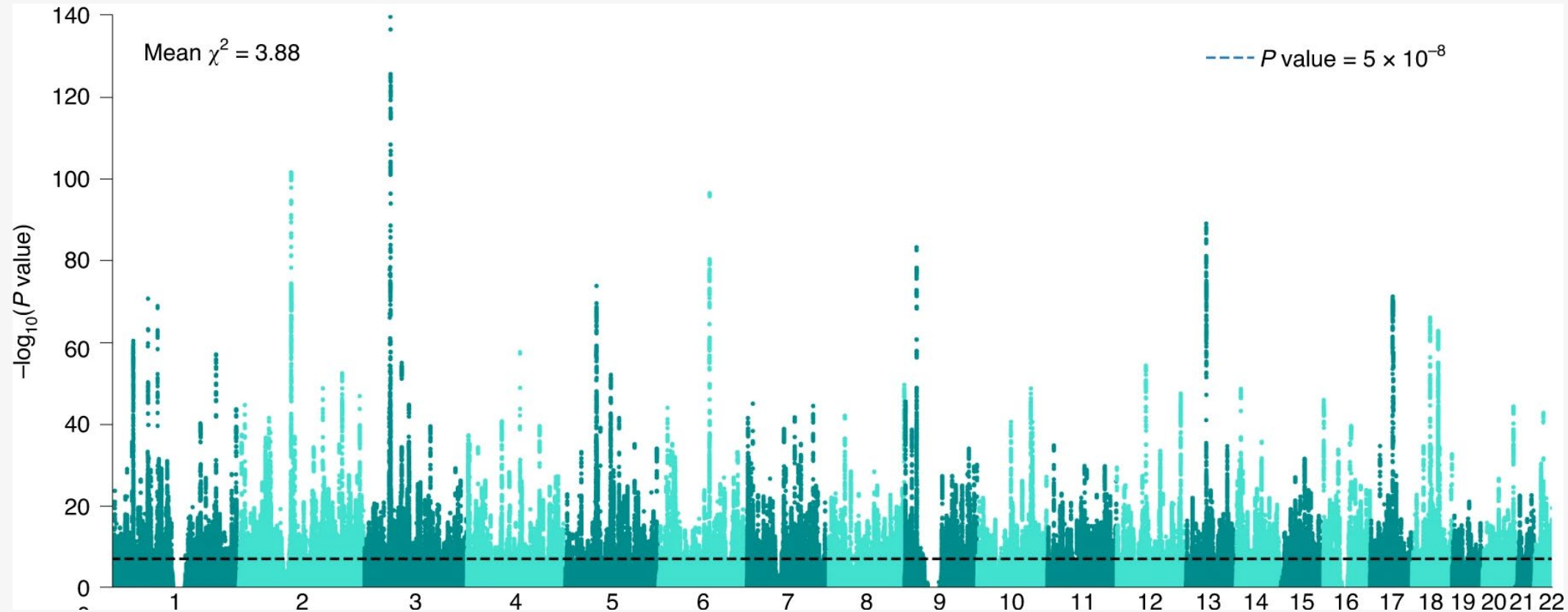


# Recent Examples – Social Isolation



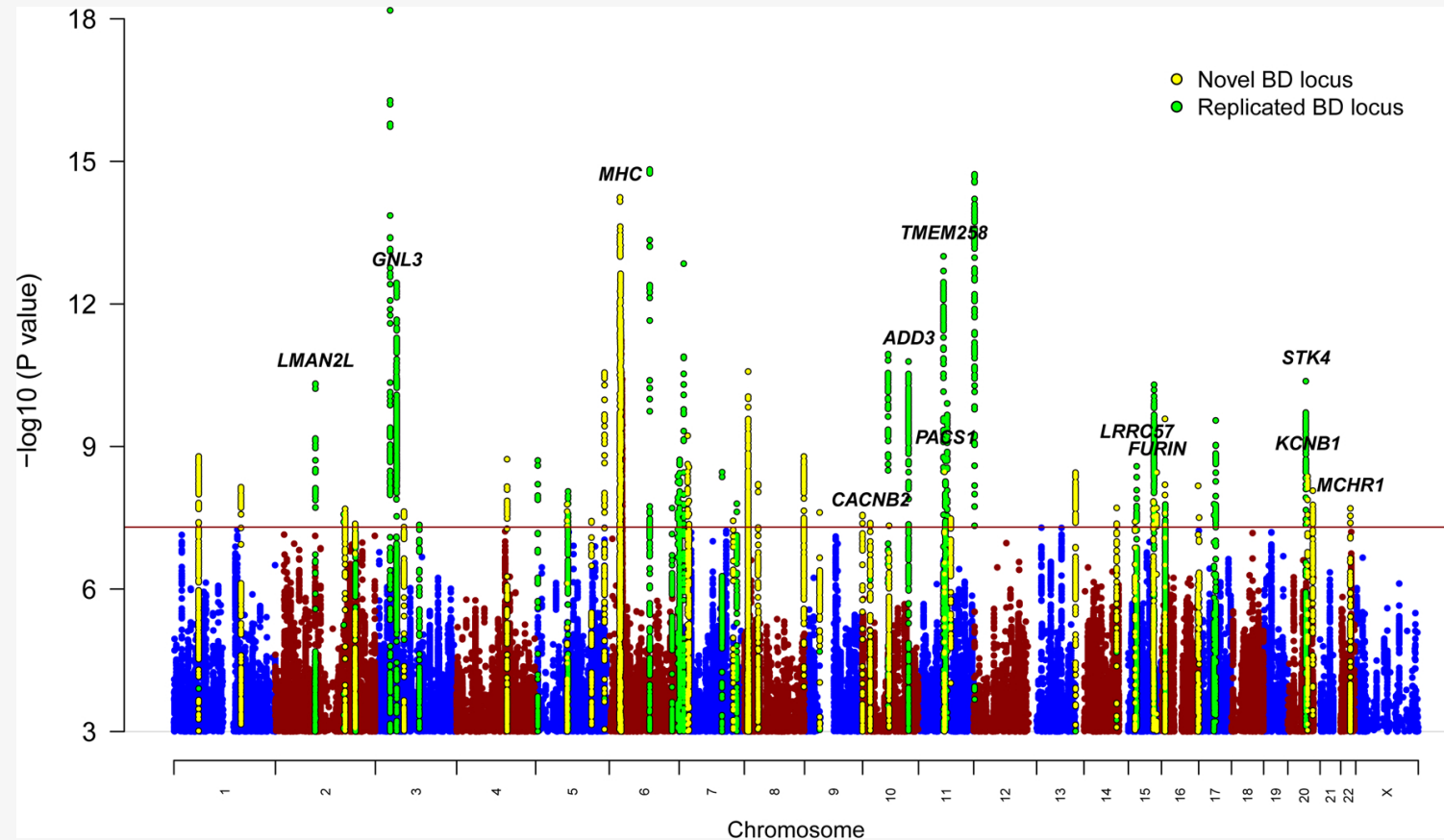
Day, F.R., Ong, K.K. & Perry, J.R.B.  
Elucidating the genetic basis of social  
interaction and isolation. *Nat  
Commun* 9, 2457 (2018).

# Recent Examples – Educational Attainment



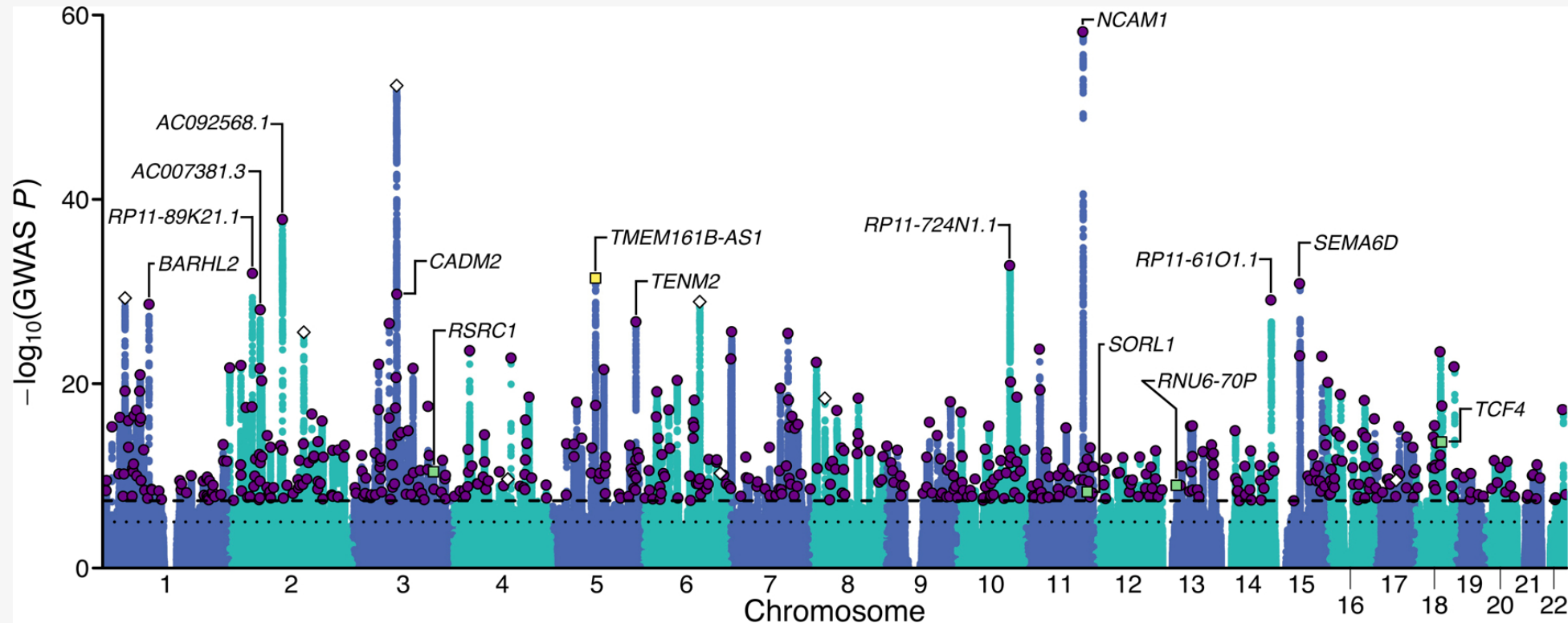
Okbay, A., Wu, Y., Wang, N. et al. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat Genet* 54, 437–449 (2022).

# Recent Examples – Bipolar Disorder



Mullins N, Forstner AJ, et al., Andreassen OA. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet. 2021 Jun;53(6):817-829.

# Recent Examples – Behavioral Regulation

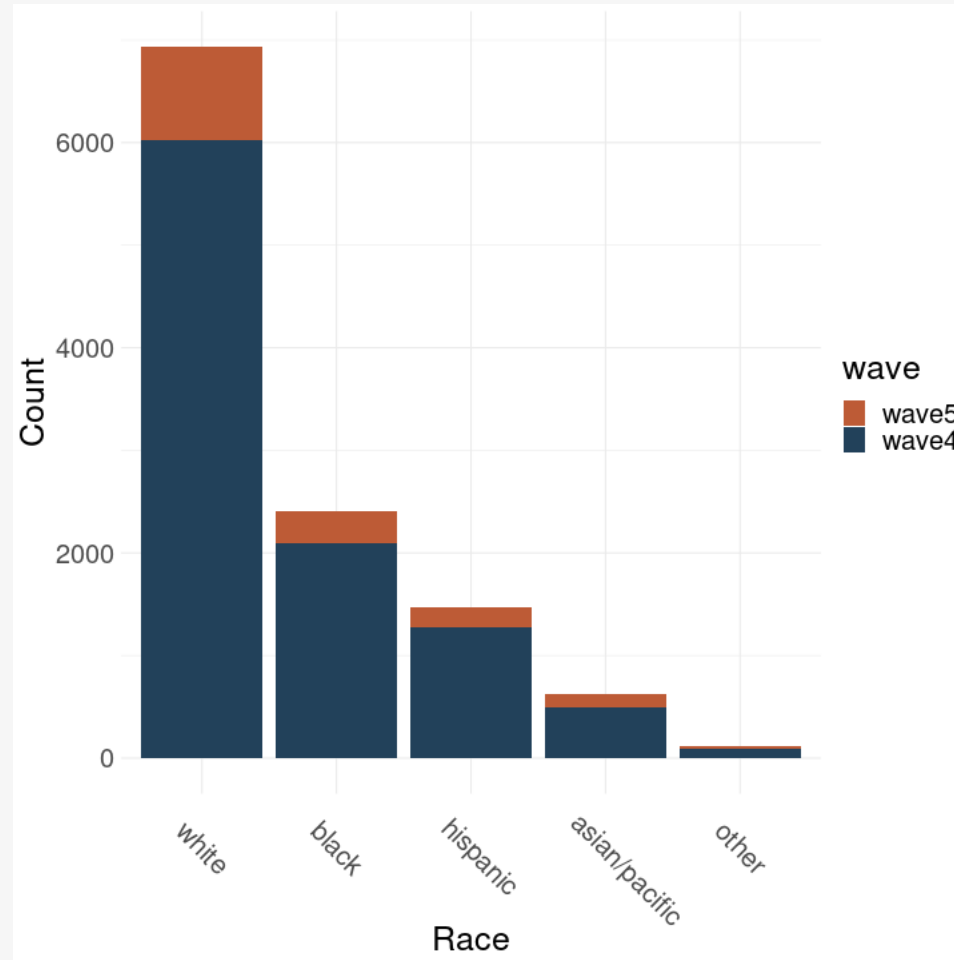


Karlsson Linnér R,...Dick DM. Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. Nat Neurosci. 2021 Oct;24(10):1367-1376.

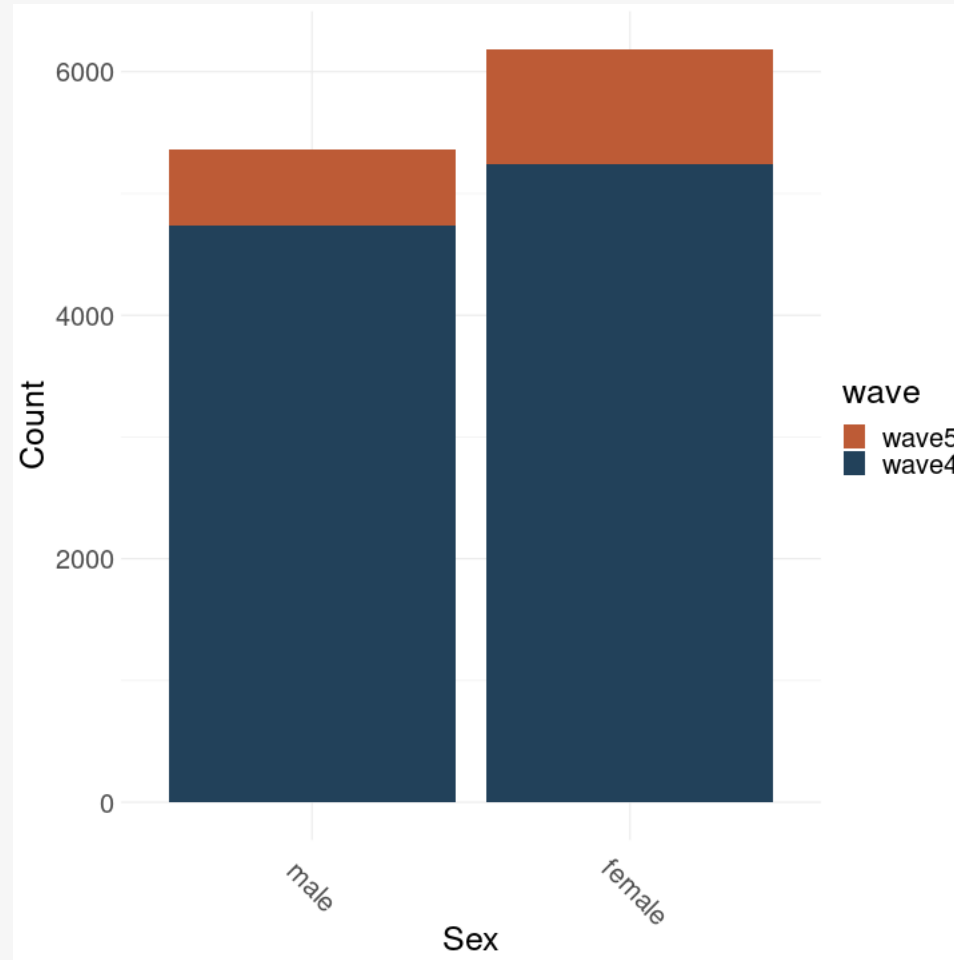
# Outline

- GWAS background
- Rationale
- Conducting a GWAS
- Recent examples
- **AddHealth genotyping data**
- Potential uses
- Resources

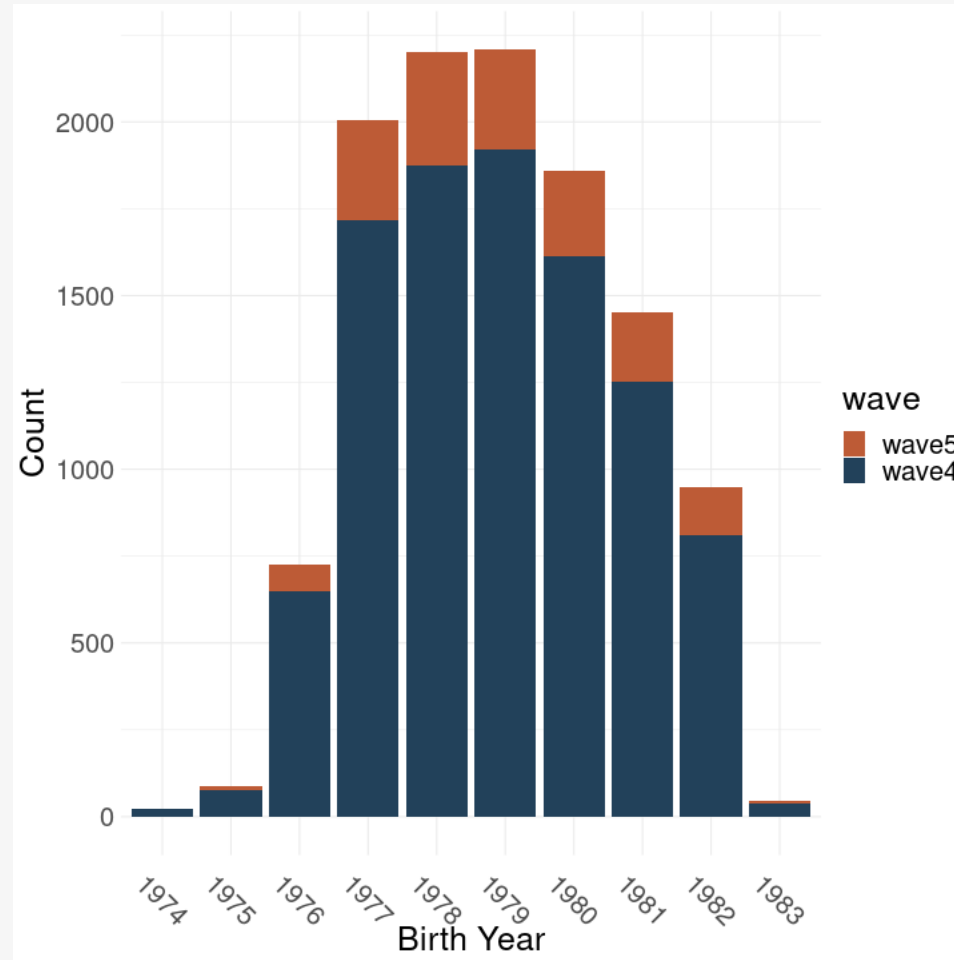
# AddHealth Genotyping Data - Race



# AddHealth Genotyping Data - Sex

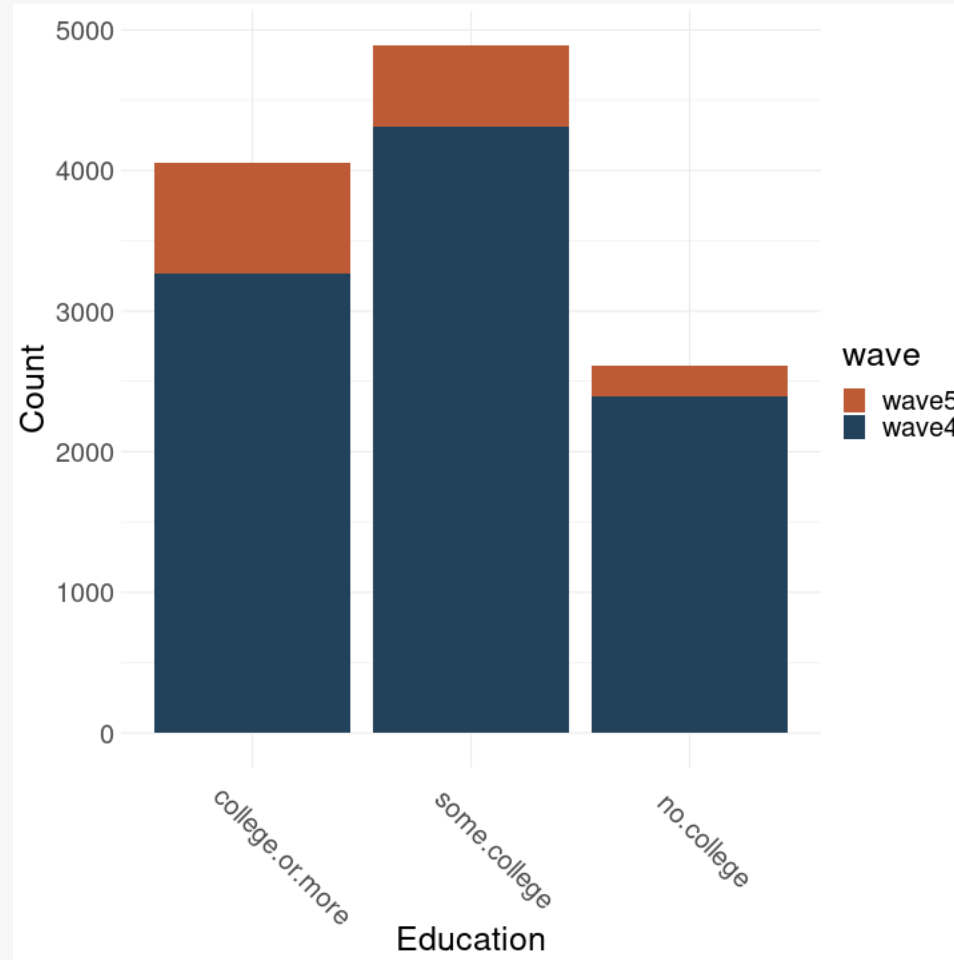


# AddHealth Genotyping Data – Birth Year

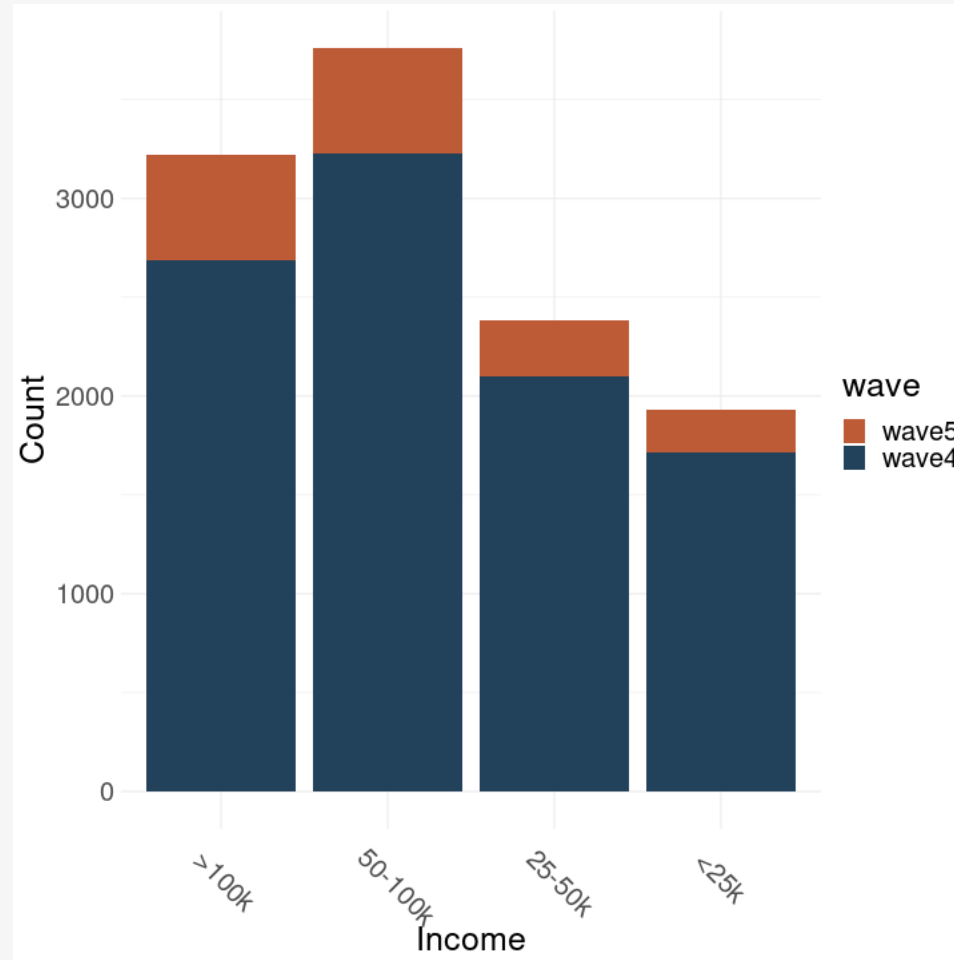




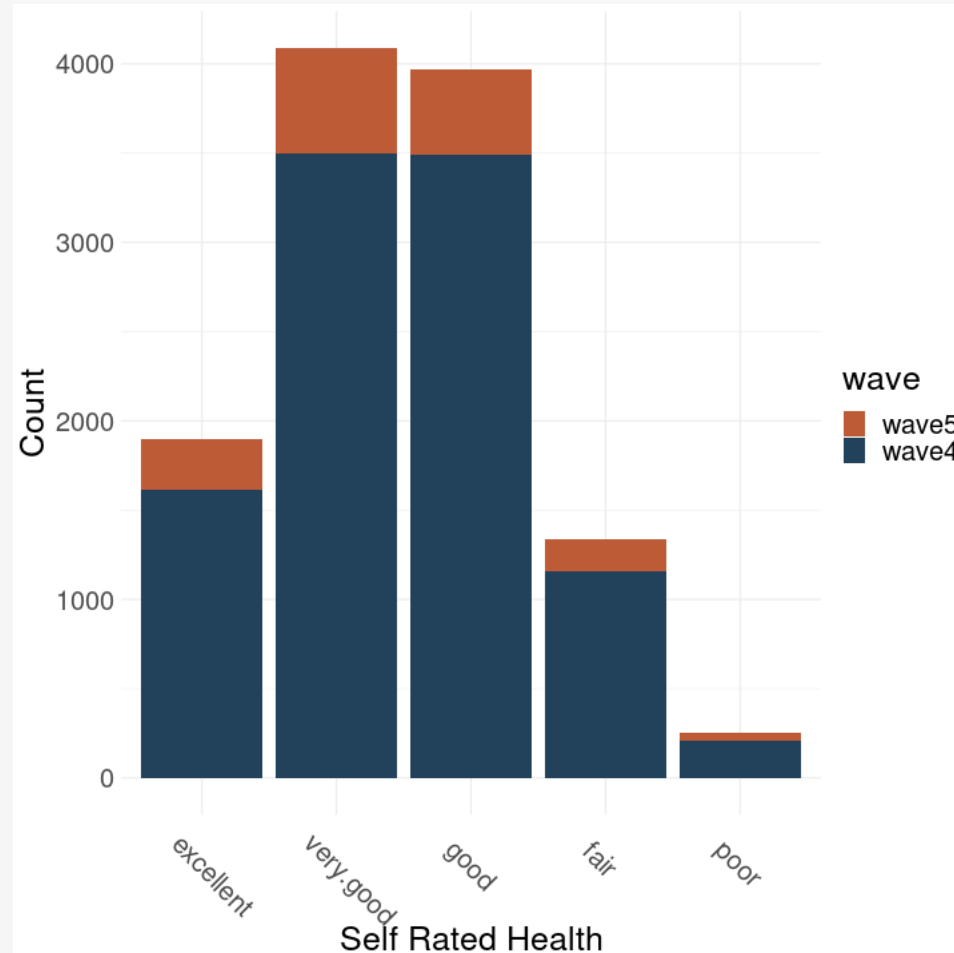
# AddHealth Genotyping Data – Educational Attainment



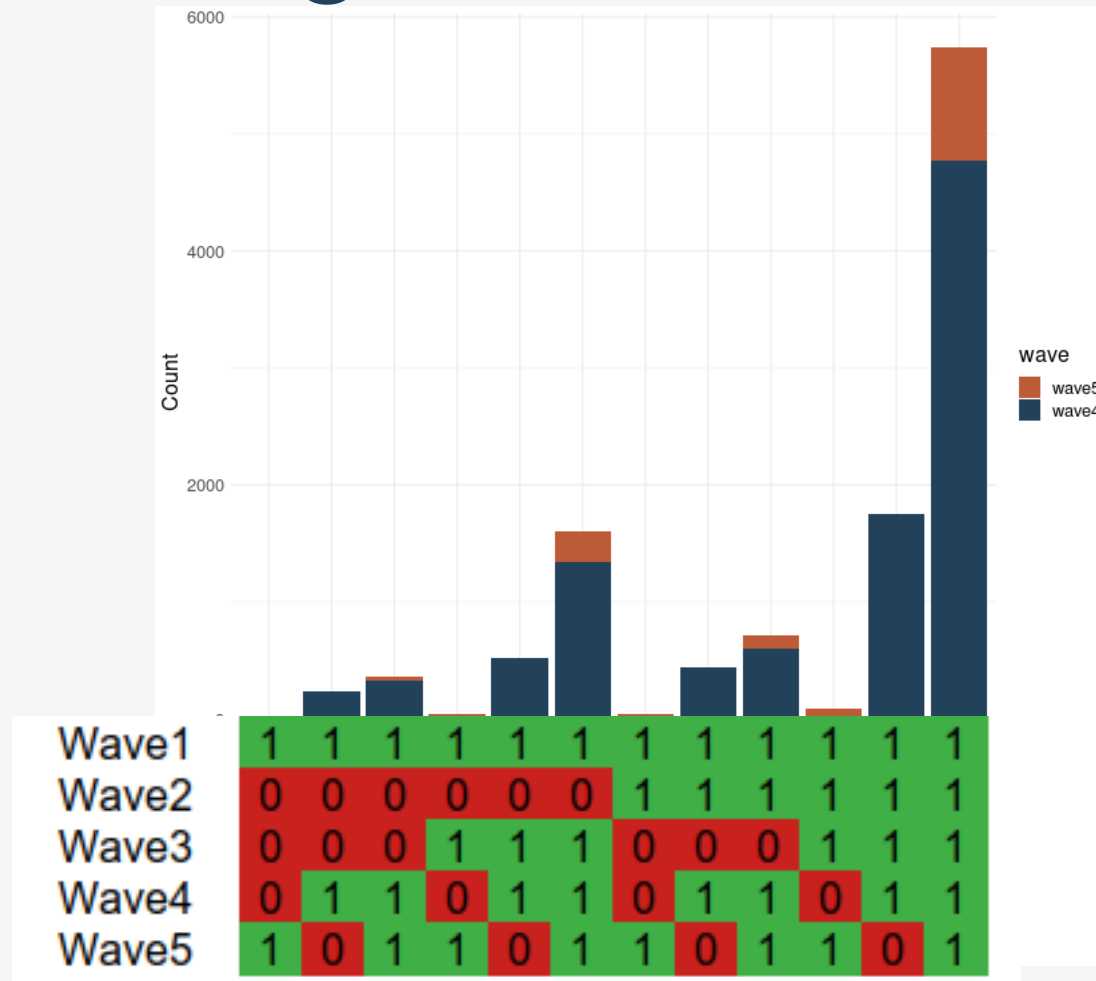
# AddHealth Genotyping Data - Income



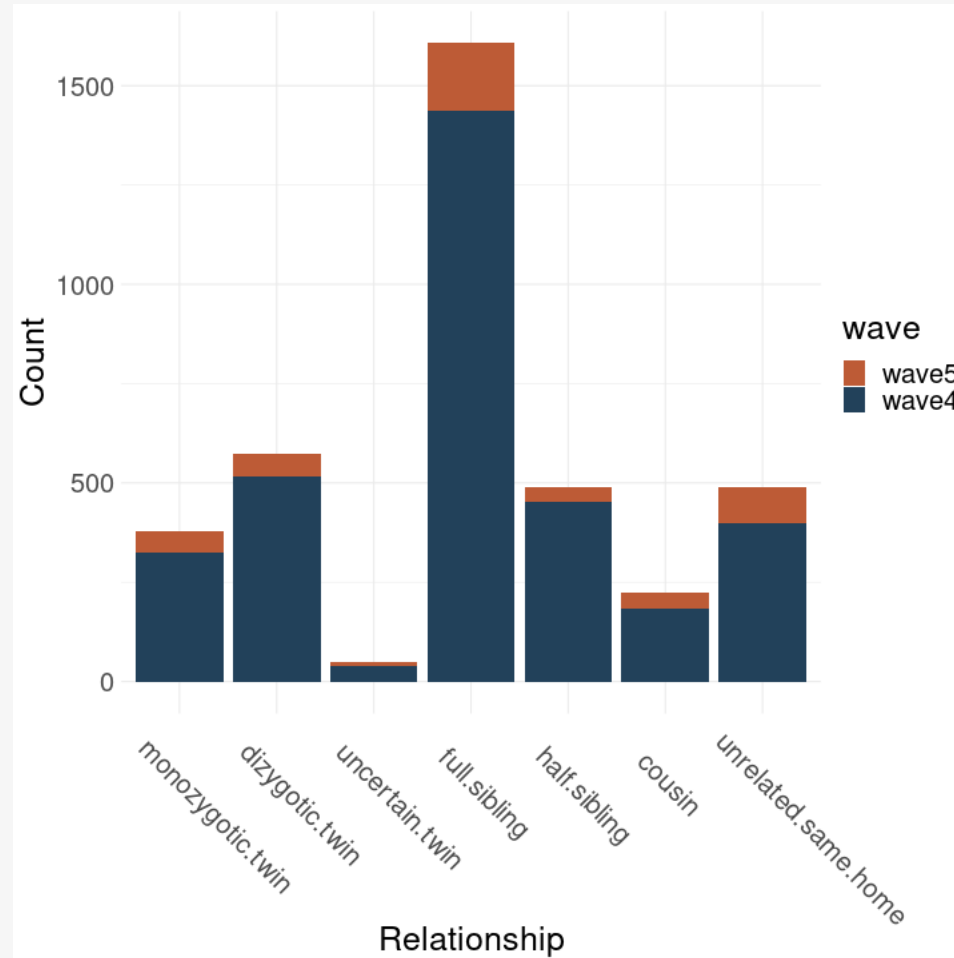
# AddHealth Genotyping Data – Self Rated Health



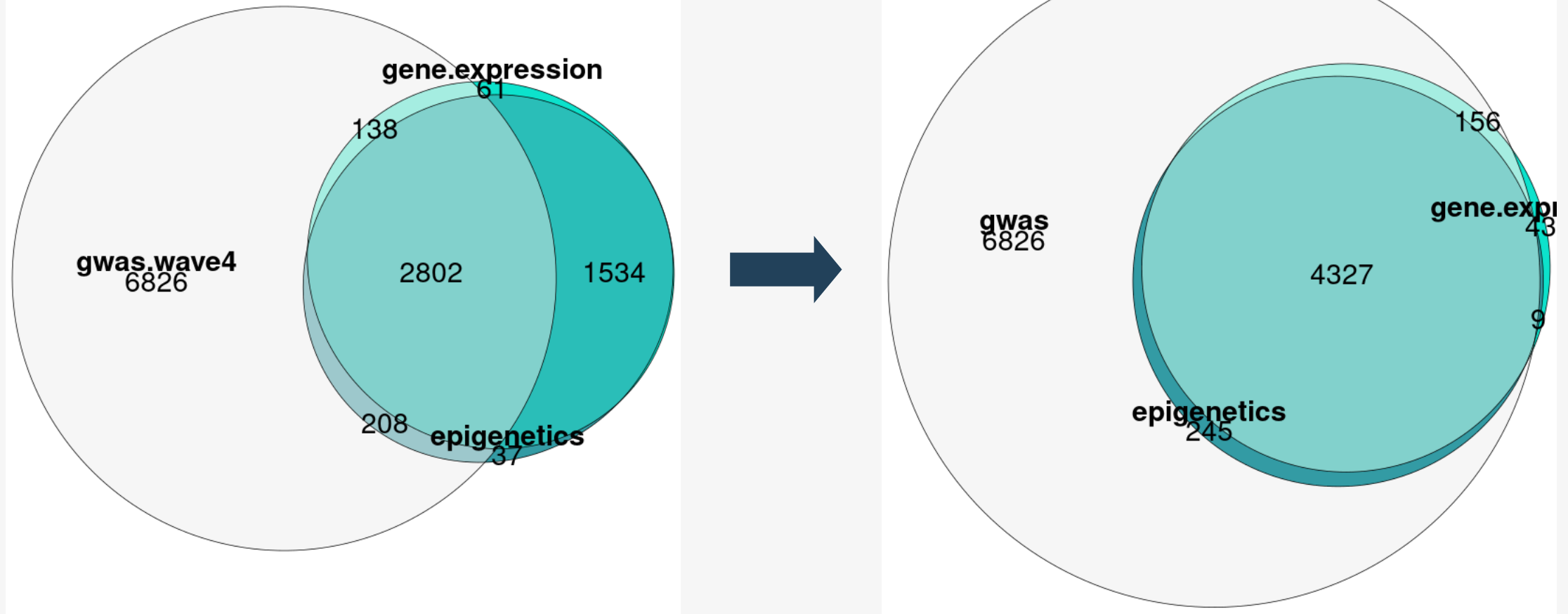
# AddHealth Genotyping Data – Longitudinal Pattern



# AddHealth Genotyping Data – Nested Pairs



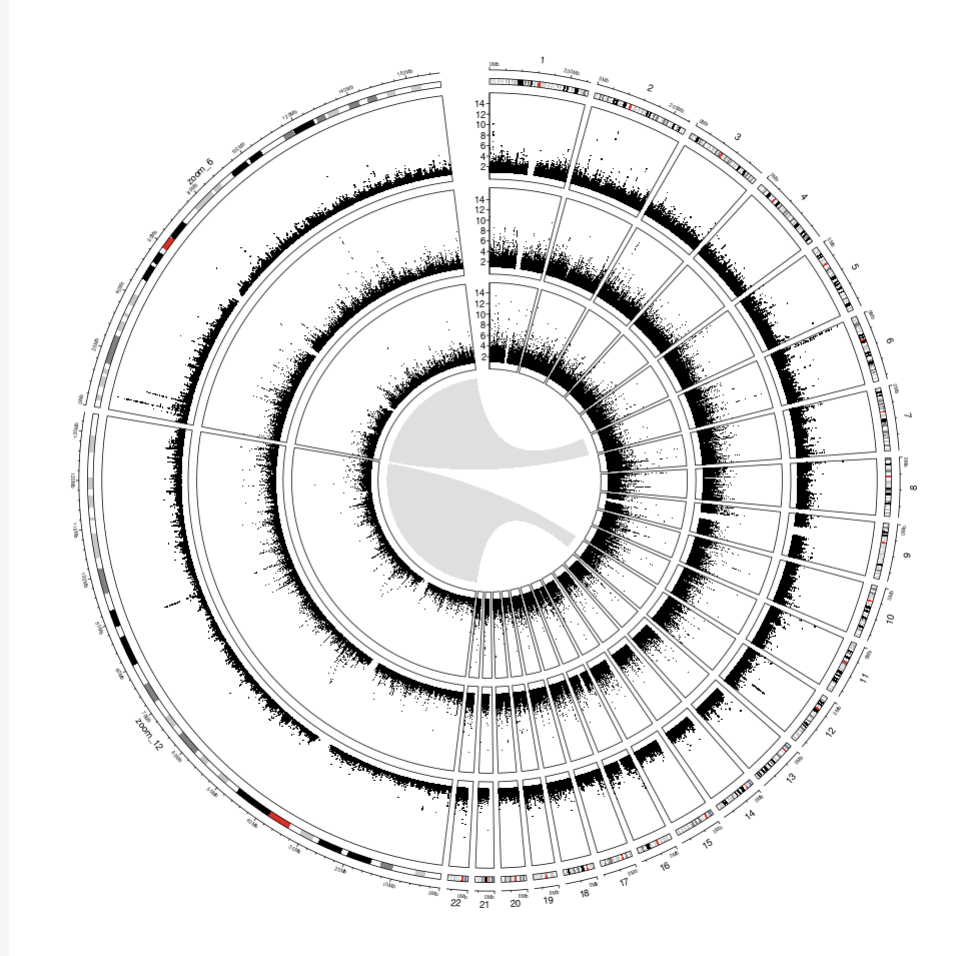
# AddHealth Genotyping Data – Overlap With Other Data



# Outline

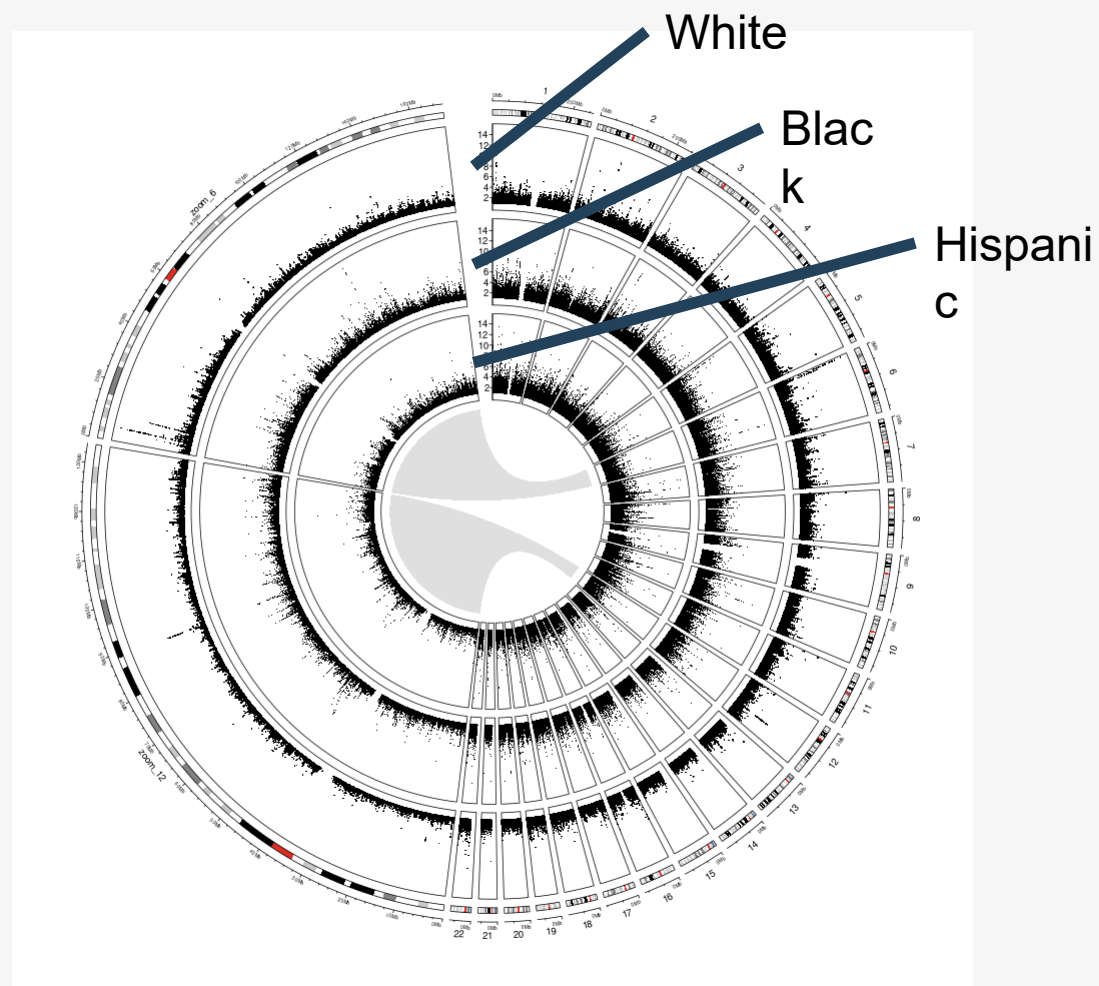
- GWAS background
- Rationale
- Conducting a GWAS
- Recent examples
- AddHealth genotyping data
- **Potential uses**
- Resources

# Trans-ancestry Analysis

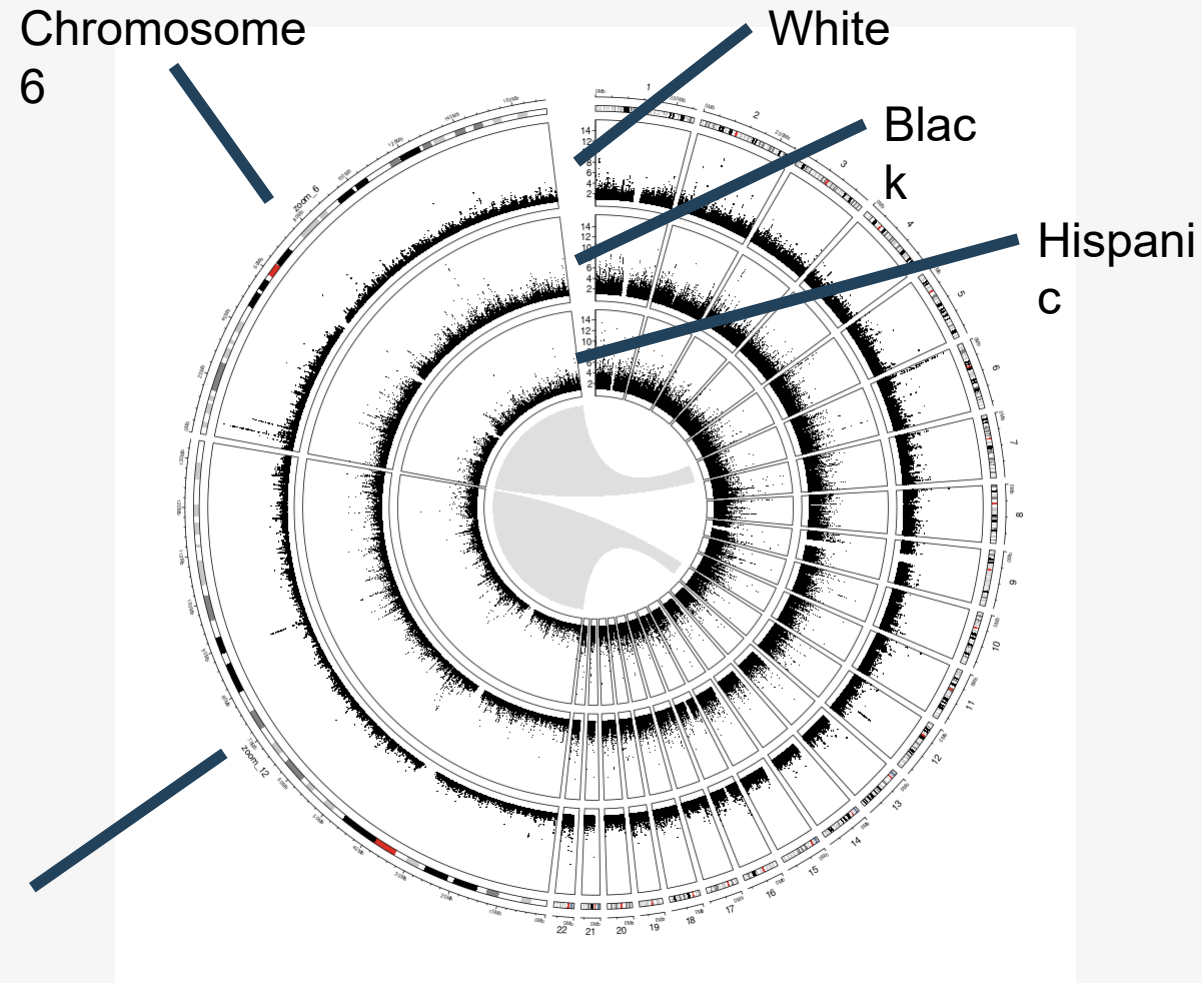




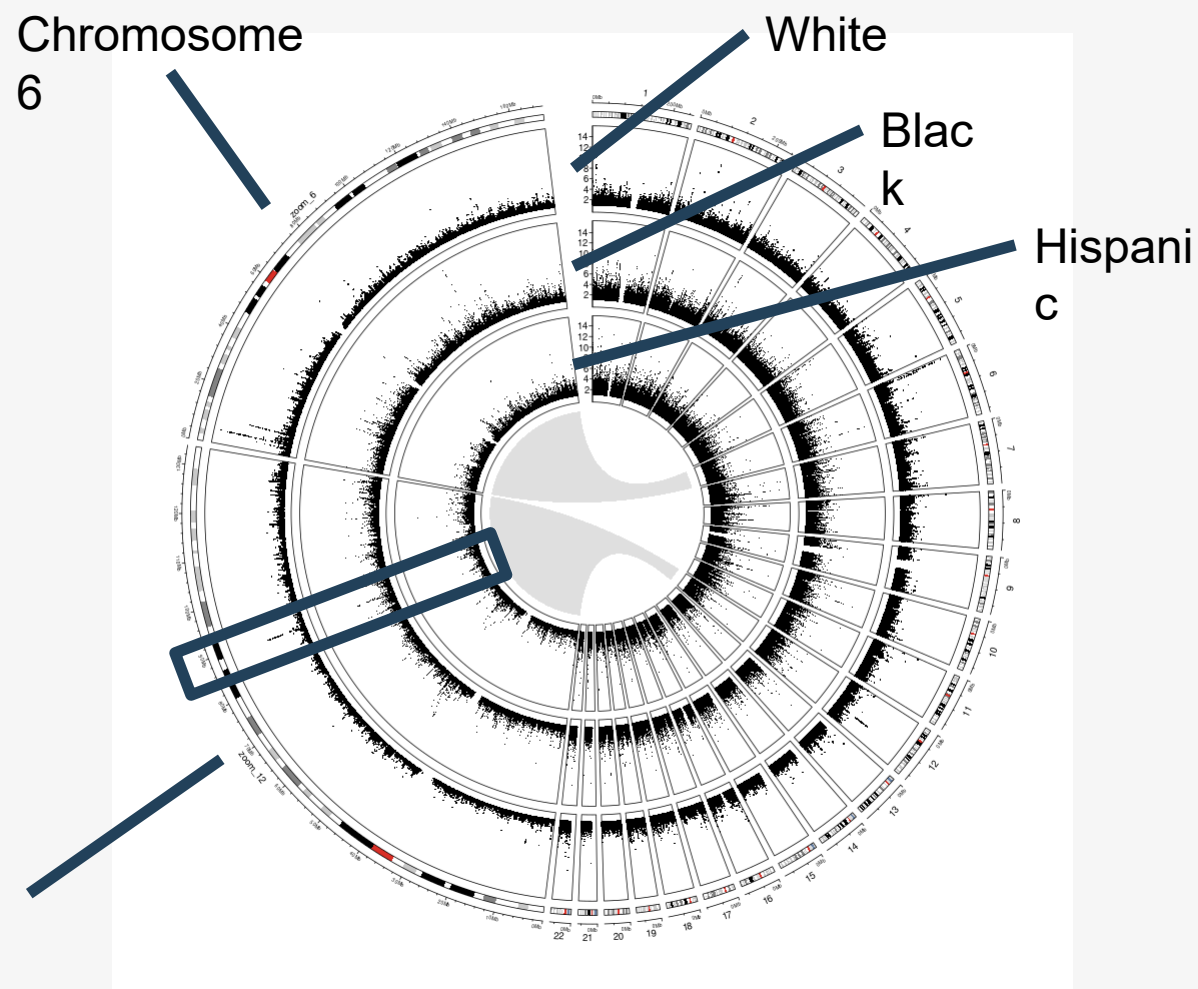
# Trans-ancestry Analysis



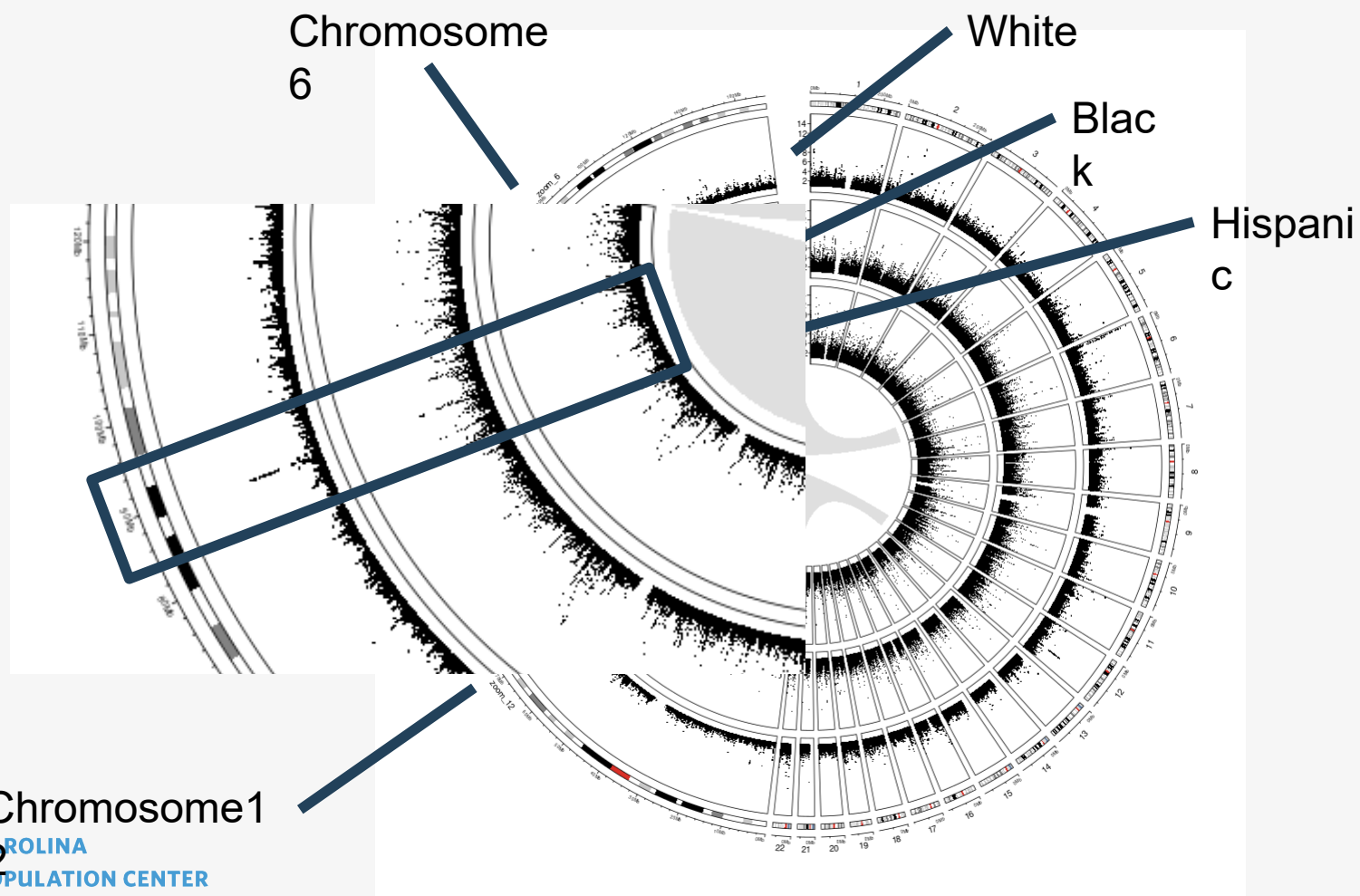
# Trans-ancestry Analysis



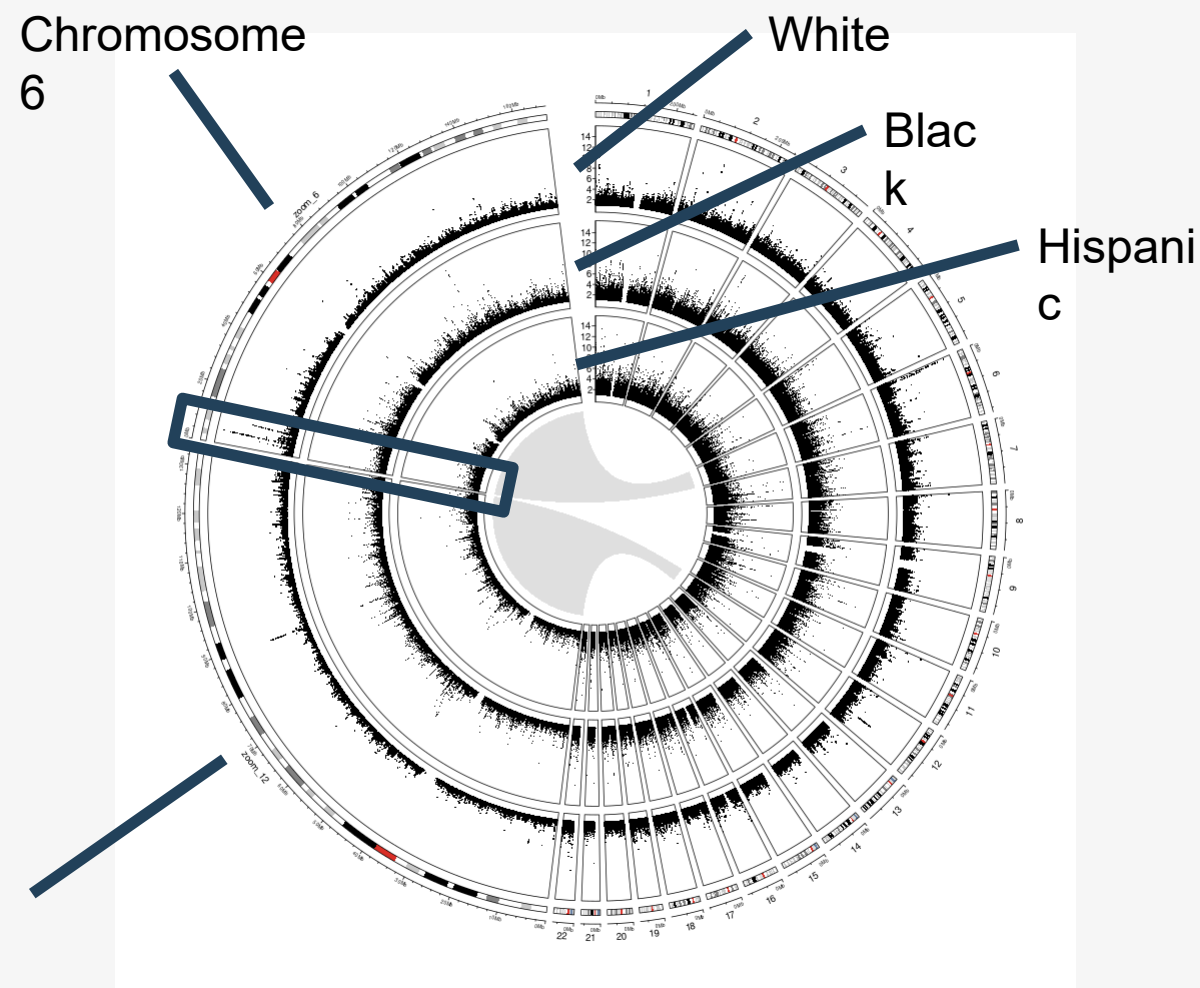
# Trans-ancestry Analysis



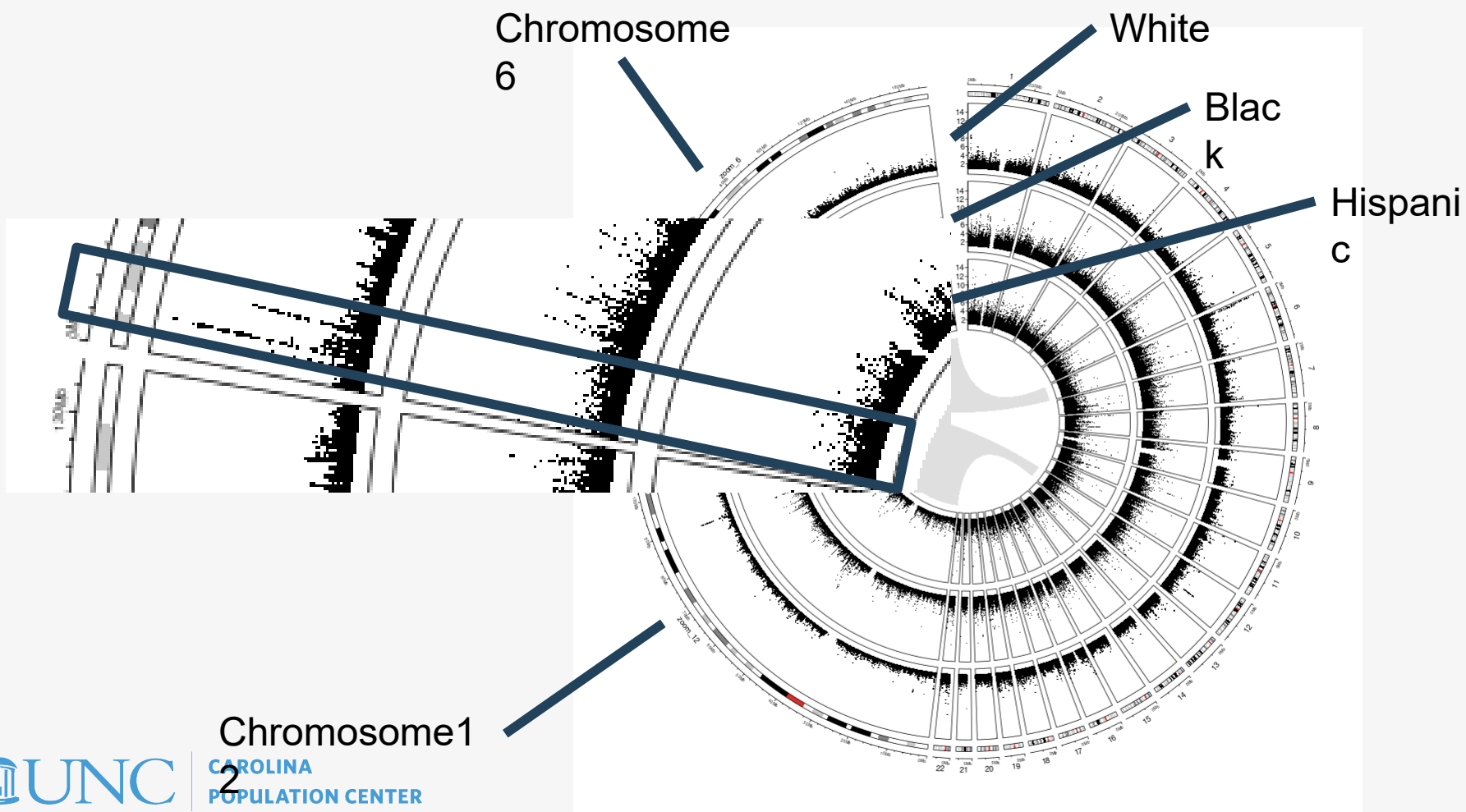
# Trans-ancestry Analysis



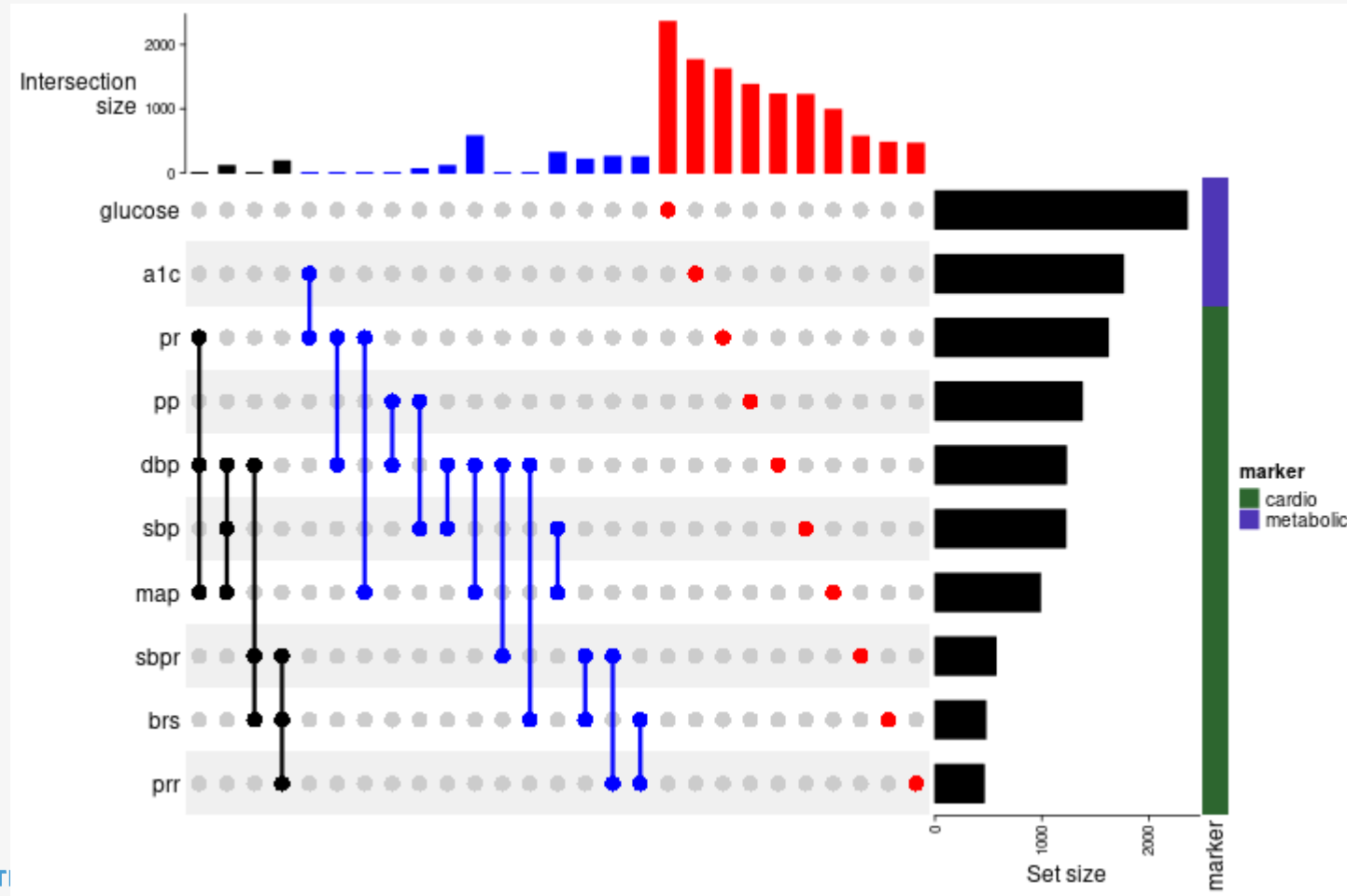
# Trans-ancestry Analysis



# Trans-ancestry Analysis

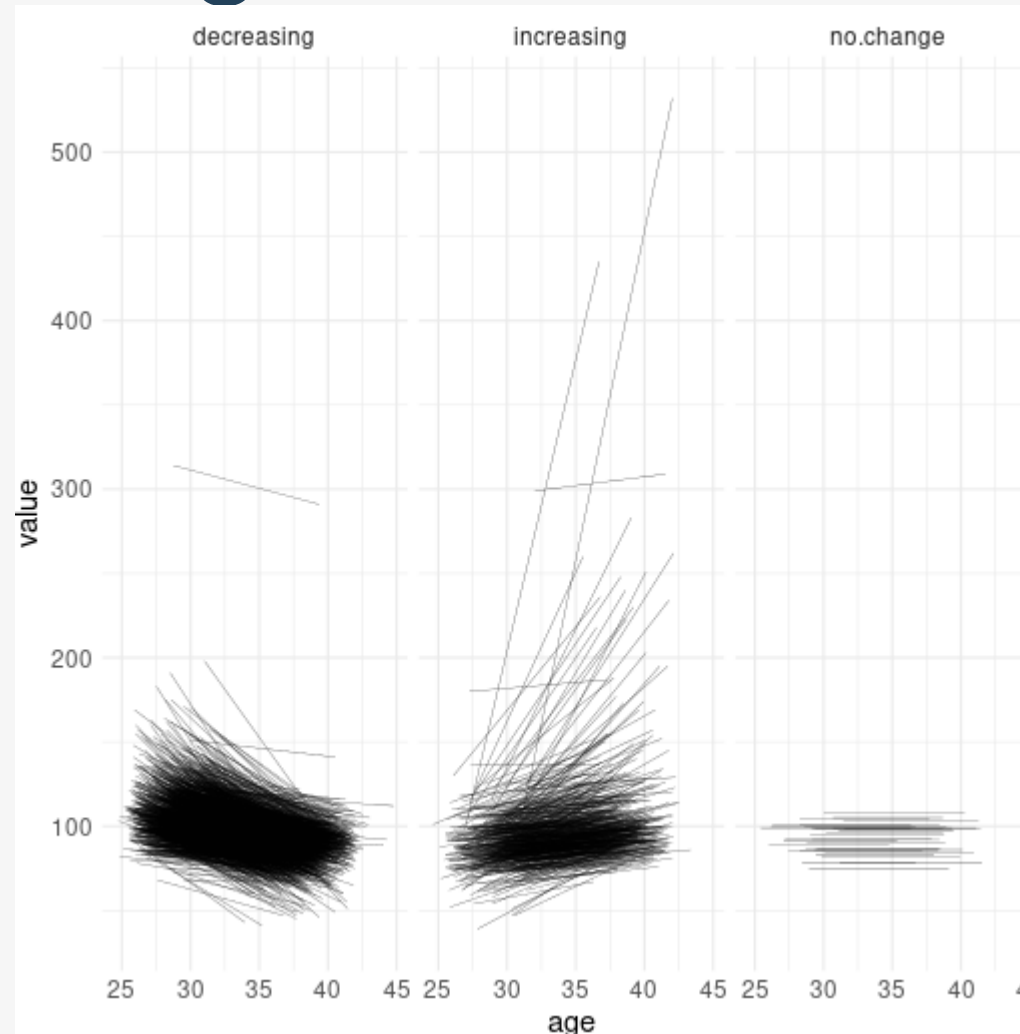


# Shared Genetic Variation





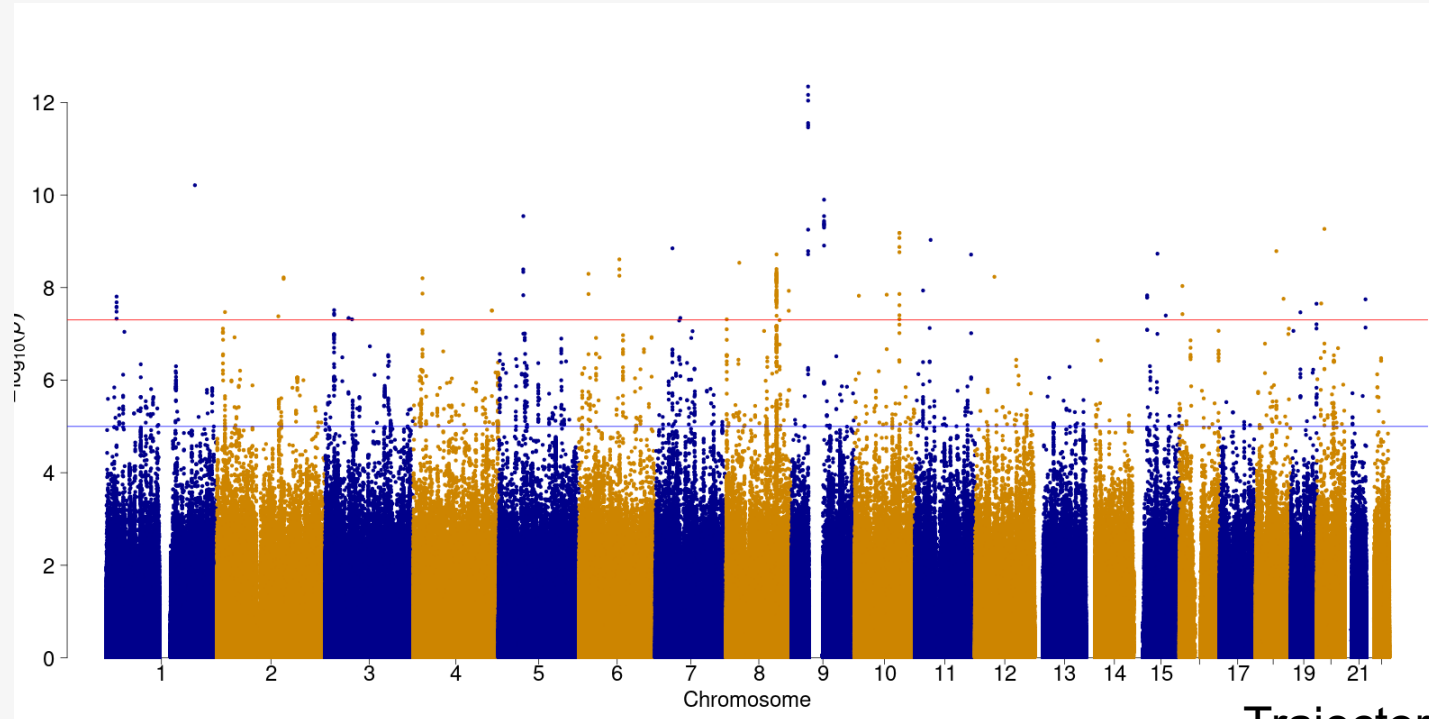
# Longitudinal Phenotypes



Trajectory of blood glucose  
during adolescence



# Longitudinal Phenotypes



Trajectory of blood glucose  
during adolescence

# Outline

- GWAS background
- Rationale
- Conducting a GWAS
- Recent examples
- AddHealth genotyping data
- Potential uses
- **Resources**

# Resources

[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001367.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001367.v1.p1)

<https://addhealth.cpc.unc.edu/about/omics>

[addhealth\\_genetics@unc.edu](mailto:addhealth_genetics@unc.edu)