

# Heritability of Human Structural Connectomes

**Jaewon Chung**

(he/him) - [NeuroData lab](#)

*Johns Hopkins University - Biomedical Engineering*

✉ [j1c@jhu.edu](mailto:j1c@jhu.edu)

🐙 [@j1c \(Github\)](#)

🐦 [@j1c \(Twitter\)](#)



# What is heritability?

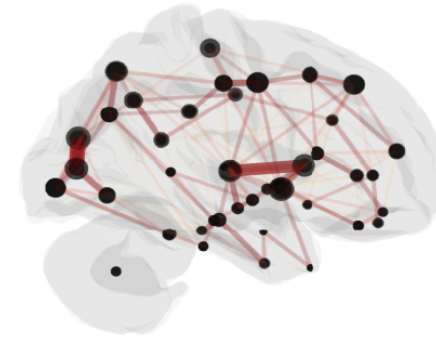
- Variations in phenotype caused by variations in genotype.
- Potentially discover relationships between diseases and genetics.

**Are the patterns of brain connectivity heritable?**

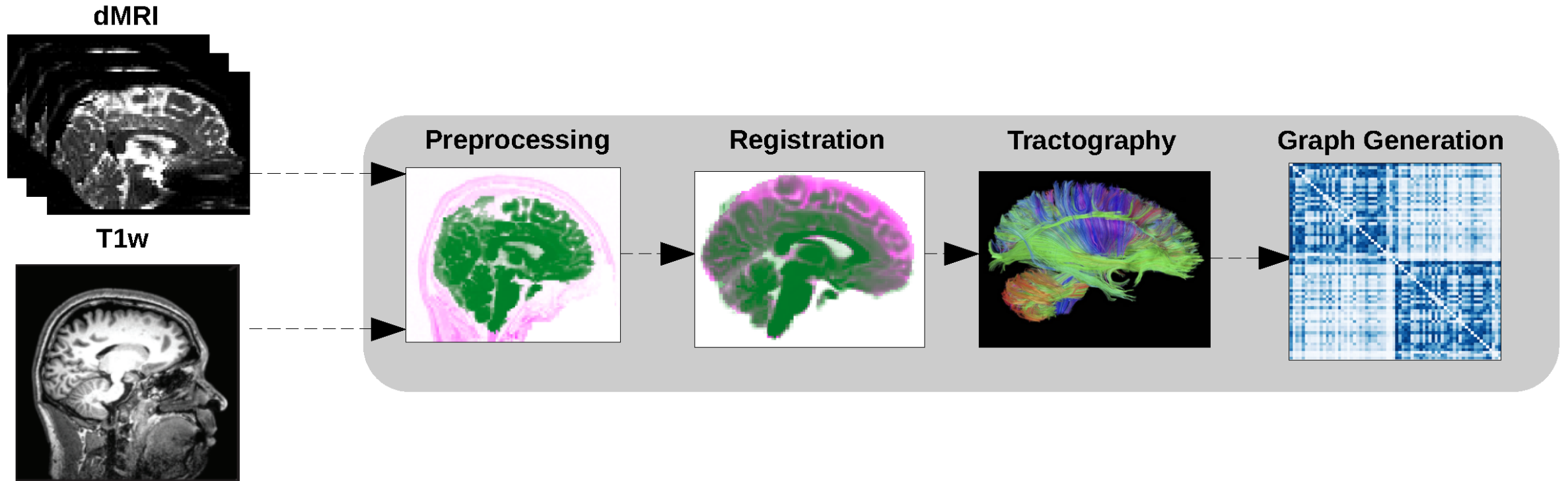
# Brain connectivity as connectomes

(aka networks or graphs)

- Vertex: region of the brain
- Edges: connectivity measure between a pair of vertices
- Structural connectomes: estimated # of neuronal fibers
- Undirected: neurons have no direction


$$\begin{bmatrix} 0 & 3 & 0 & 0 & 1 & 0 \\ 3 & 0 & 2 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 6 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 6 & 0 & 0 \end{bmatrix}$$

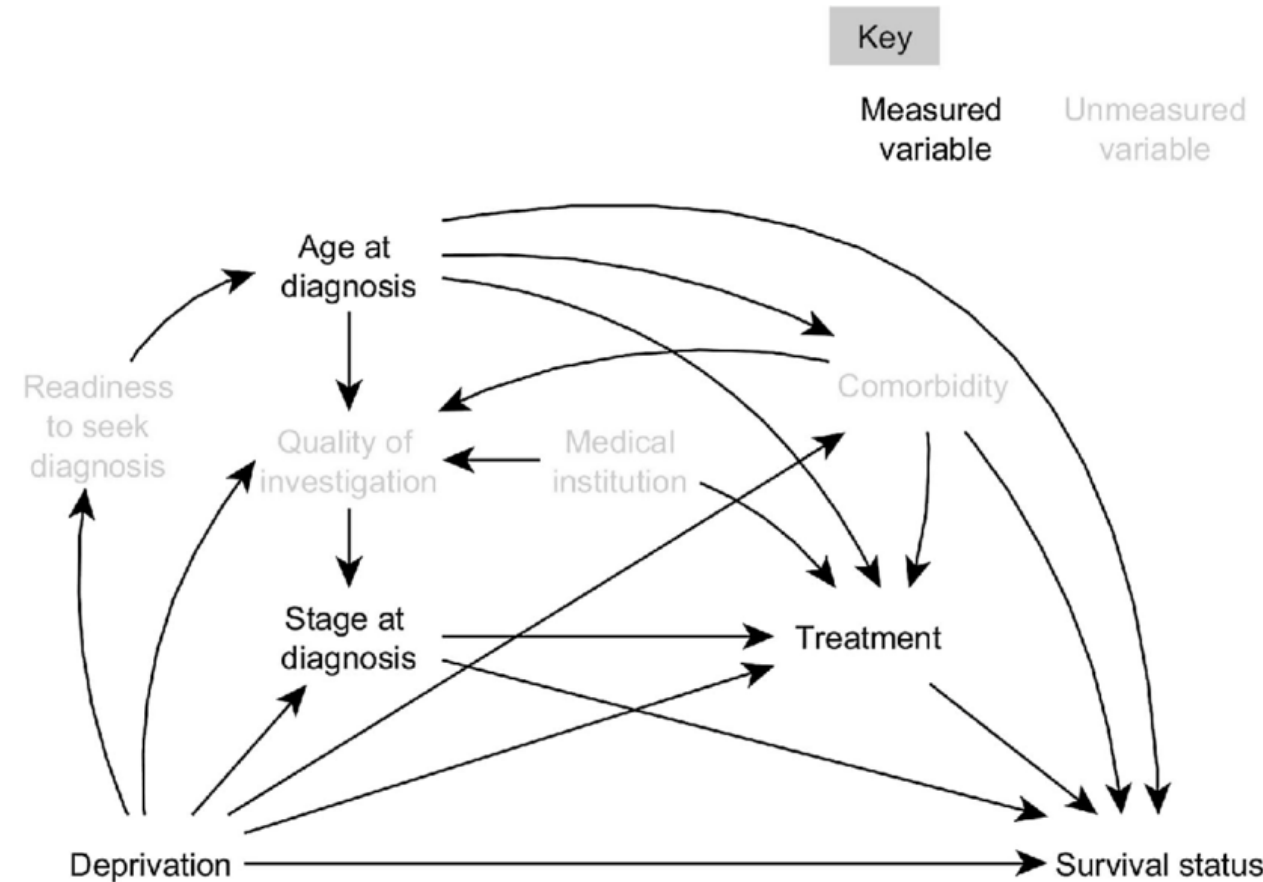
# How do we get structural connectomes?



**Quick detour to causal land**

# Directed acyclic graphs

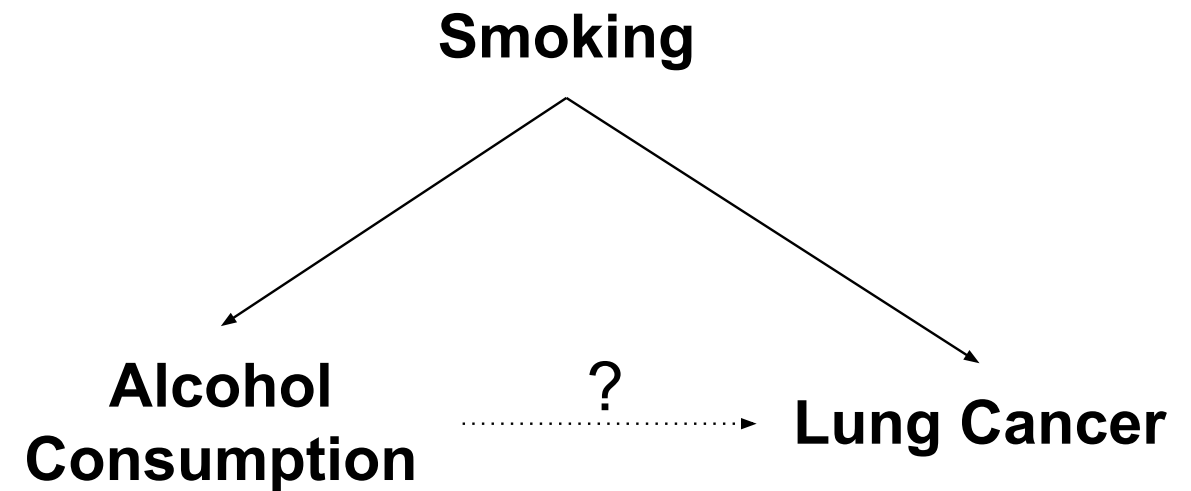
- Representation of causal relationships of variables
  - From domain knowledge



Li, Ruoran, Rhian Daniel, and Bernard Rachet. "How much do tumor stage and treatment explain socioeconomic inequalities in breast cancer survival? Applying causal mediation analysis to population-based data." *European journal of epidemiology* 31.6 (2016): 603-611.

# Confounding

- Common cause for both exposure and outcome
  - Exposure: alcohol consumption
  - Outcome: lung cancer
- Causes spurious association!
- Solution: "condition" confounding



# Toy example

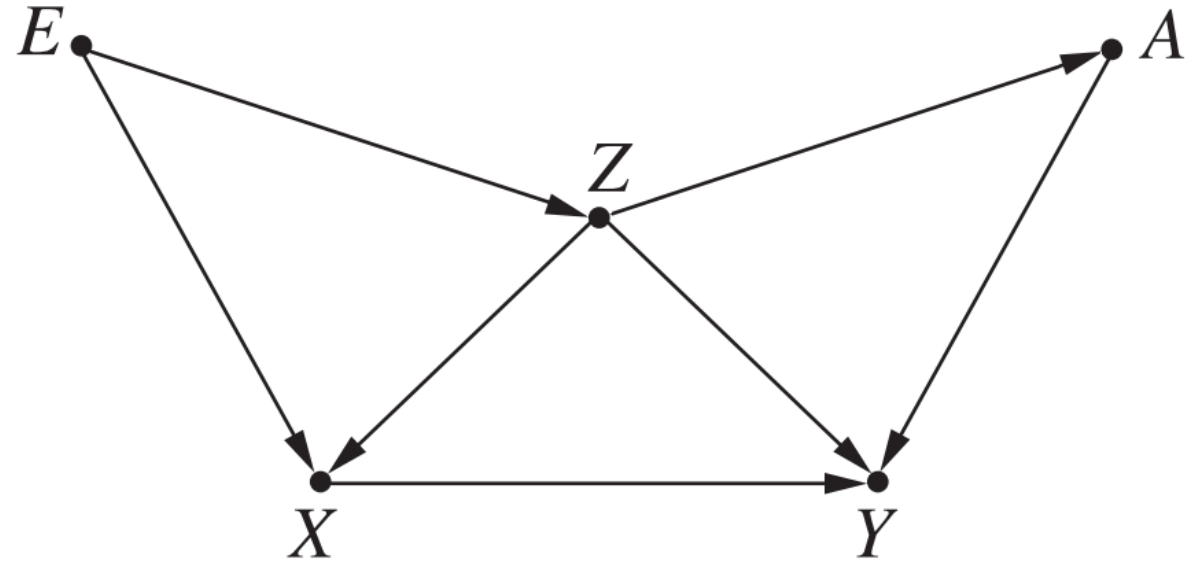
- Smoking =  $\mathcal{N}(5, 1)$
- Alcohol = Smoking +  $\mathcal{N}(0, 1)$
- Lung Cancer = Smoking +  $\mathcal{N}(0, 1)$
- n = 100

	Indep. Test	Conditional Indep. Test
Null Hypothesis	Alcohol $\perp$ Lung cancer	Alcohol $\perp$ Lung cancer   Smoking
Alternate Hypothesis	Alcohol $\not\perp$ Lung cancer	Alcohol $\not\perp$ Lung cancer   Smoking
Correlation ( $r$ )	$\approx 0.66$	$\approx 0.006$
p-value	$\approx 0$	$\approx 1$



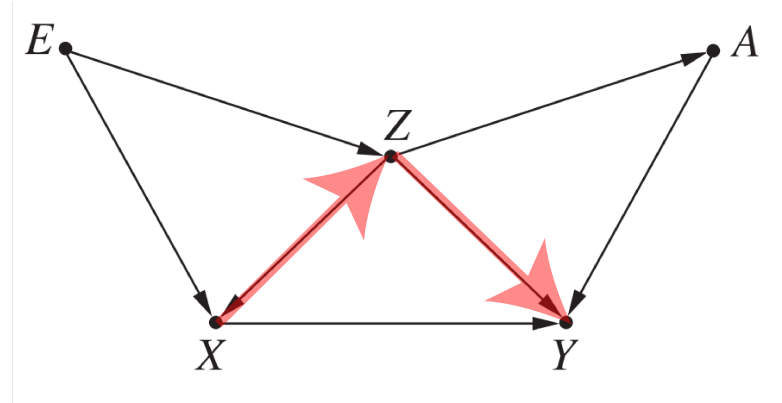
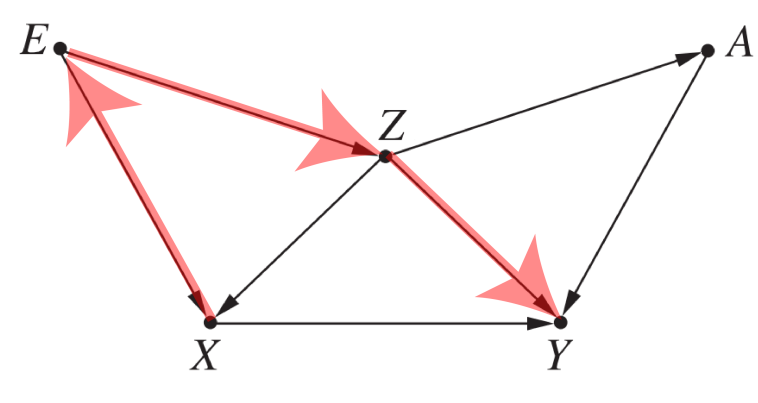
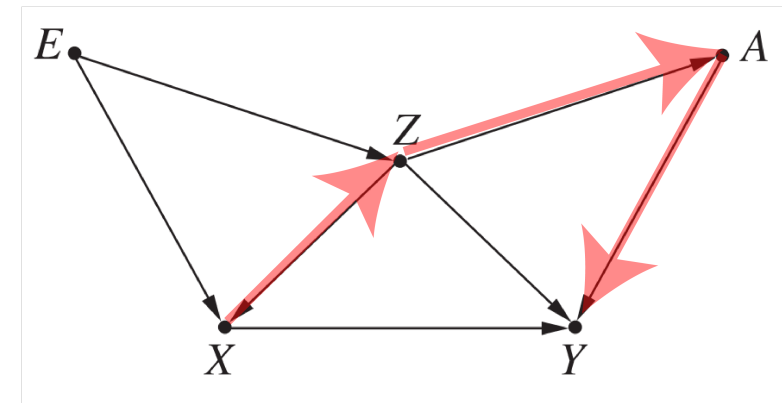
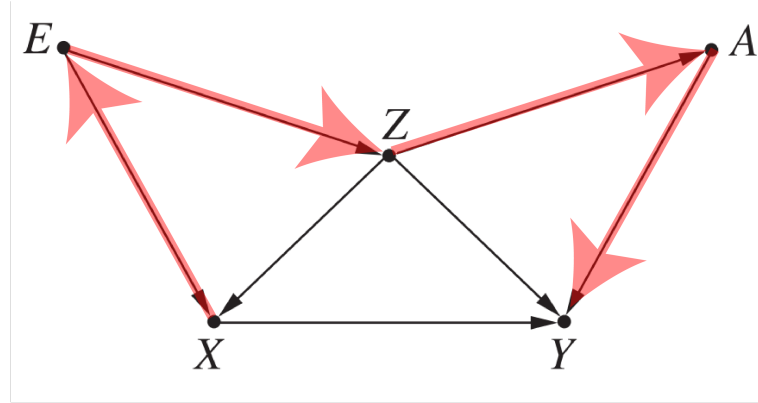
# Backdoor paths

- Non-causal path from exposure to outcome
- Flow backwards out of exposure (arrow pointing towards exposure)
- Confounders create backdoor paths
- Path is blocked if a variable in the path is conditioned on.



# Examples

- Four possible paths
- Sufficient condition sets:
  - $\{Z\}$
  - $\{Z, E\}$
  - $\{Z, A\}$
  - $\{Z, E, A\}$



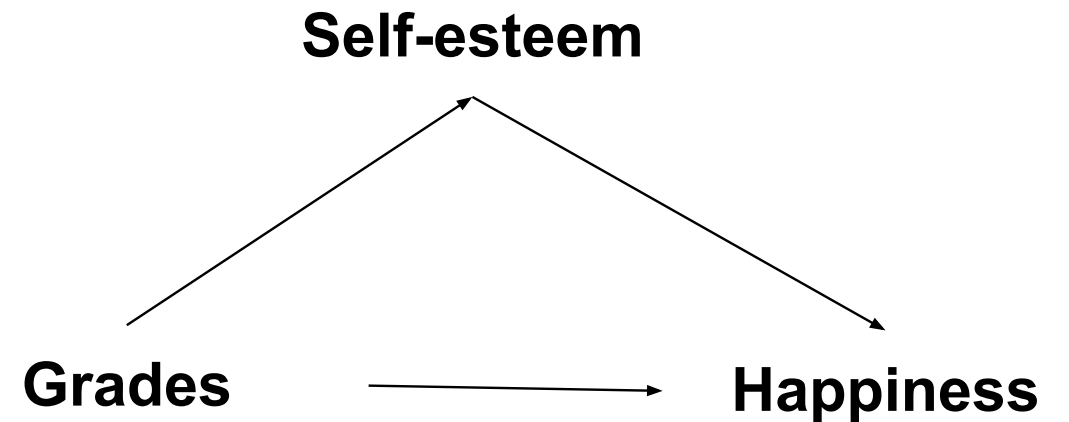
# Backdoor criterion

Causal effect exists between exposure ( $X$ ) and outcome ( $Y$ ) if:

1. No backdoor paths from  $X$  to  $Y$  exist
2. Measured covariates  $Z$  block all backdoor paths from  $X$  and  $Y$

# Mediation

- Lies on the causal path between exposure and outcome
  - Exposure: grades
  - Outcome: happiness
  - Mediator: self-esteem
- Total effect: effect from both exposure and mediators
- Direct effect: effect only from exposure



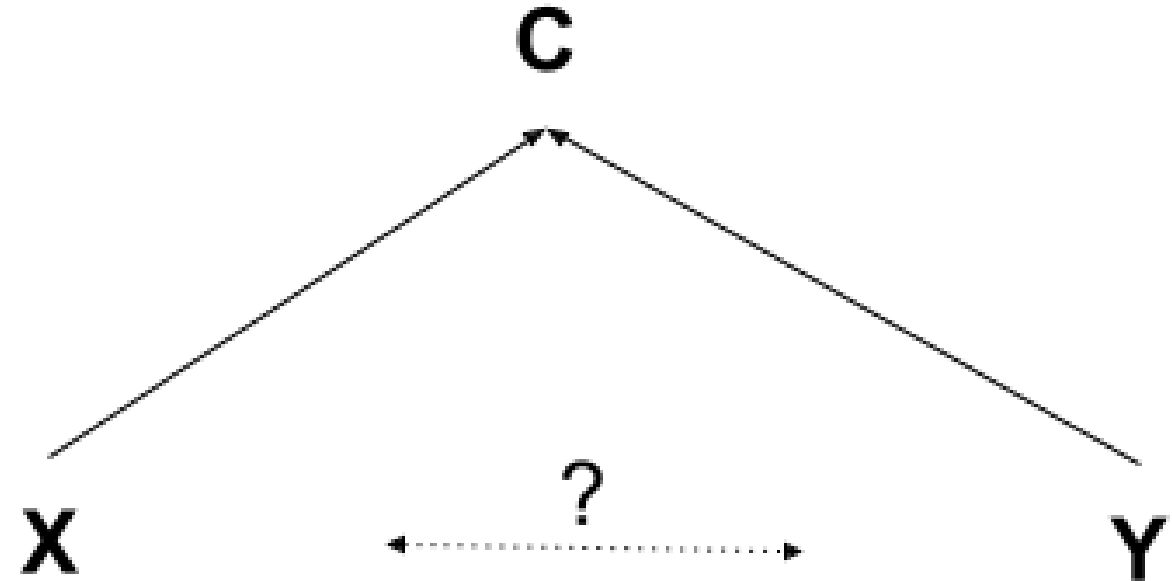
# Toy example

- Grades =  $\mathcal{N}(5, 1)$
- Self-esteem = Grades +  $\mathcal{N}(0, 1)$
- Happiness = Grades + Self-esteem +  $\mathcal{N}(0, 1)$
- n = 100

	Indep. Test	Conditional Indep. Test
Null Hypothesis	Grades $\perp$ Happiness	Grades $\perp$ Happiness   Self-esteem
Alternate Hypothesis	Grades $\not\perp$ Happiness	Grades $\not\perp$ Happiness   Self-esteem
Correlation ( $r$ )	$\approx 0.89$	$\approx 0.63$
p-value	$\approx 0$	$\approx 0$

# Collider

- A variable causally influenced by two or more variables
- Conditioning on collider does not block path
- Solution: also condition on a parent

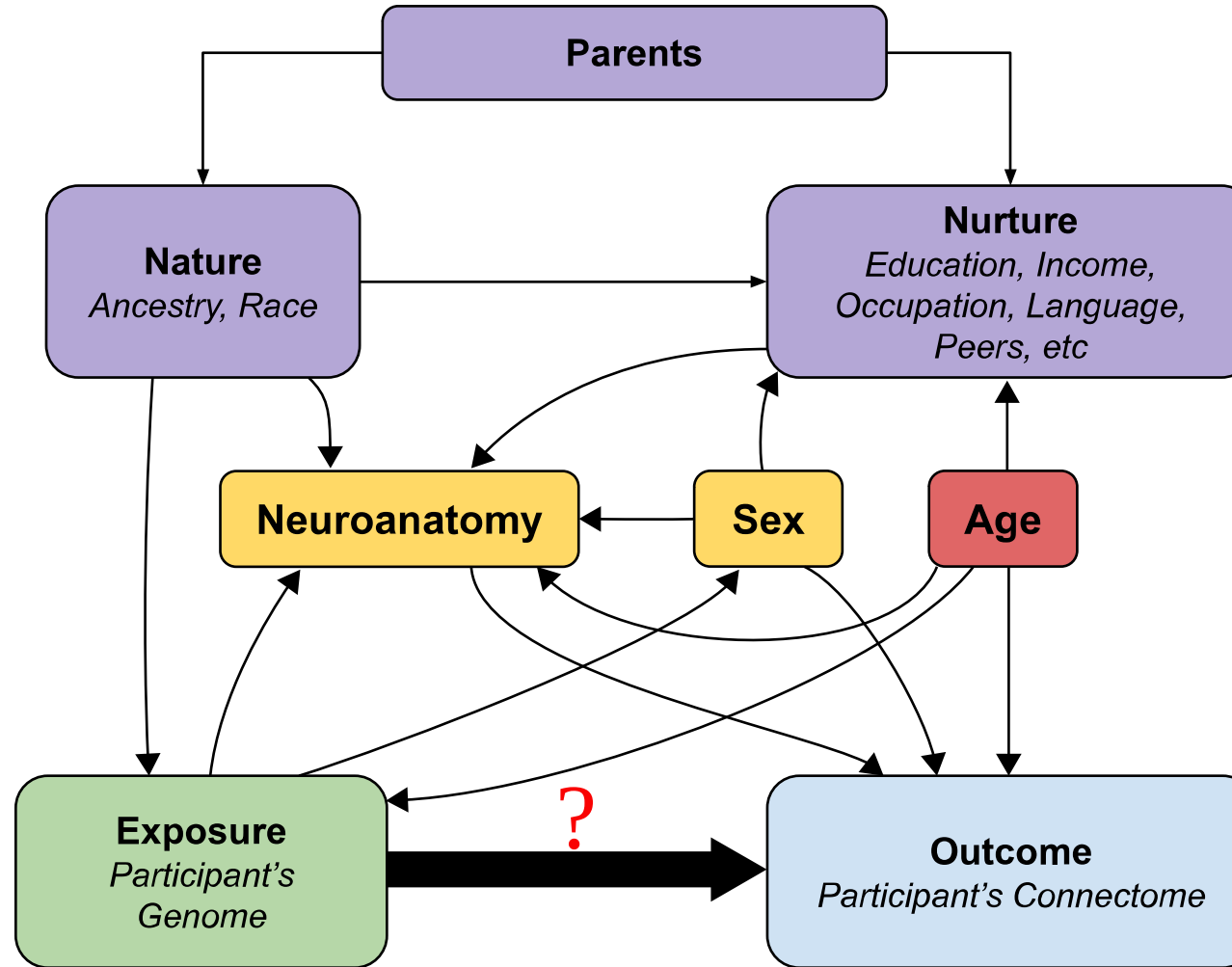


# Toy example

- $X = \mathcal{N}(5, 1)$
- $Y = \mathcal{N}(10, 1)$
- $C = X + Y + \mathcal{N}(0, 1)$

	Indep. Test	Conditional Indep. Test
Null Hypothesis	$X \perp\!\!\!\perp Y$	$X \perp\!\!\!\perp Y   C$
Alternate Hypothesis	$X \not\perp\!\!\!\perp Y$	$X \not\perp\!\!\!\perp Y   C$
Correlation ( $r$ )	$\approx 0.003$	$\approx -0.66$
p-value	$\approx 1$	$\approx 0$

# Heritability as causal problem





# Do genomes affect connectomes?

- Our hypothesis:

$$H_0 : F(\text{Connectome}|\text{Genome}) = F(\text{Connectome})$$

$$H_A : F(\text{Connectome}|\text{Genome}) \neq F(\text{Connectome})$$

- Alternatively:

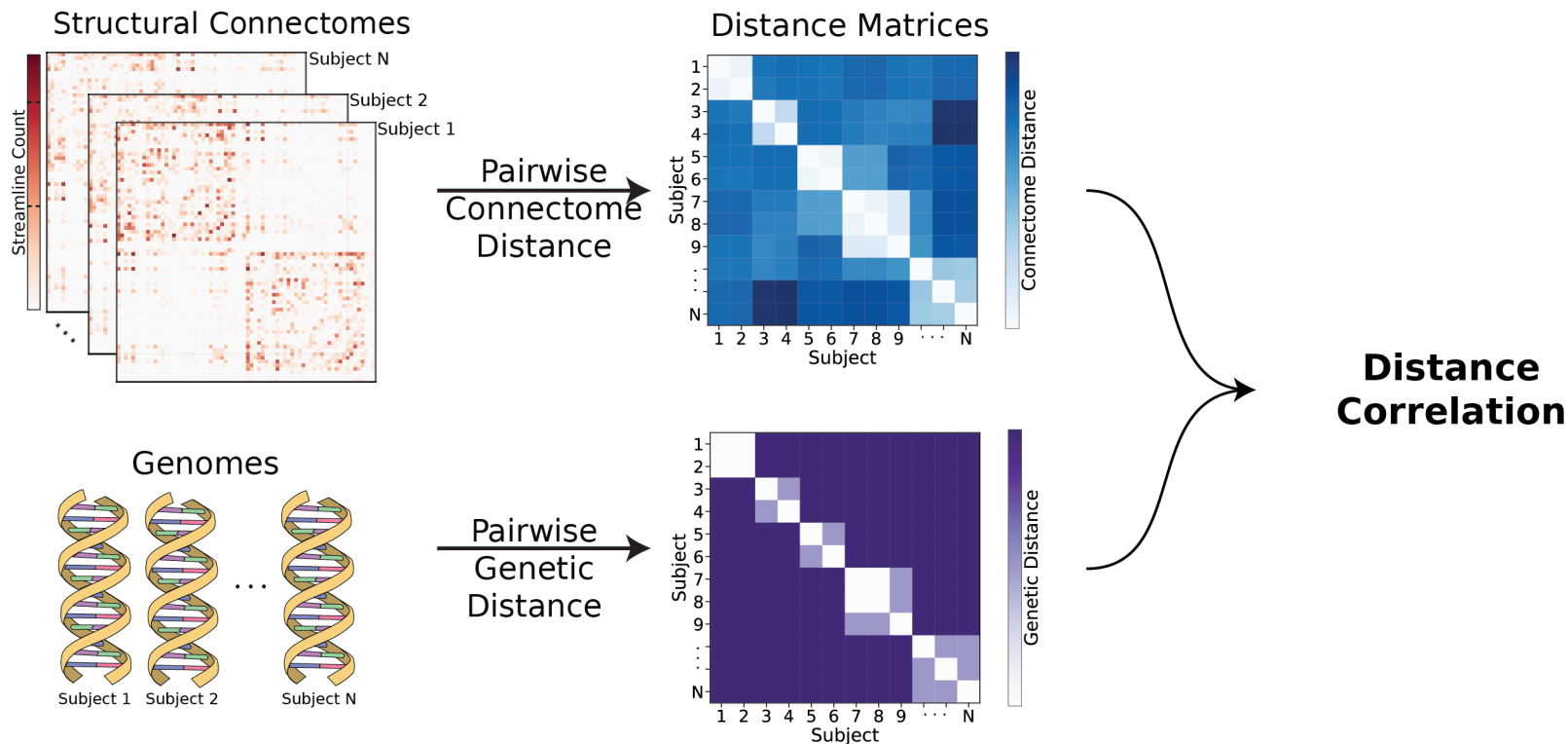
$$H_0 : F(\text{Connectome}, \text{Genome}) = F(\text{Connectome})F(\text{Genome})$$

$$H_A : F(\text{Connectome}, \text{Genome}) \neq F(\text{Connectome})F(\text{Genome})$$

- Known as independence testing
- Test statistic: *distance correlation* (*dcorr*)
- Implication if false: there exists an associational heritability.

# Distance correlation

- Measures dependence between two *multivariate* quantities.
  - For example: connectomes, genomes.
- Can detect nonlinear associations.
- Measures correlation between pairwise distances.



# Do genomes affect connectomes given covariates?

- Want to test:

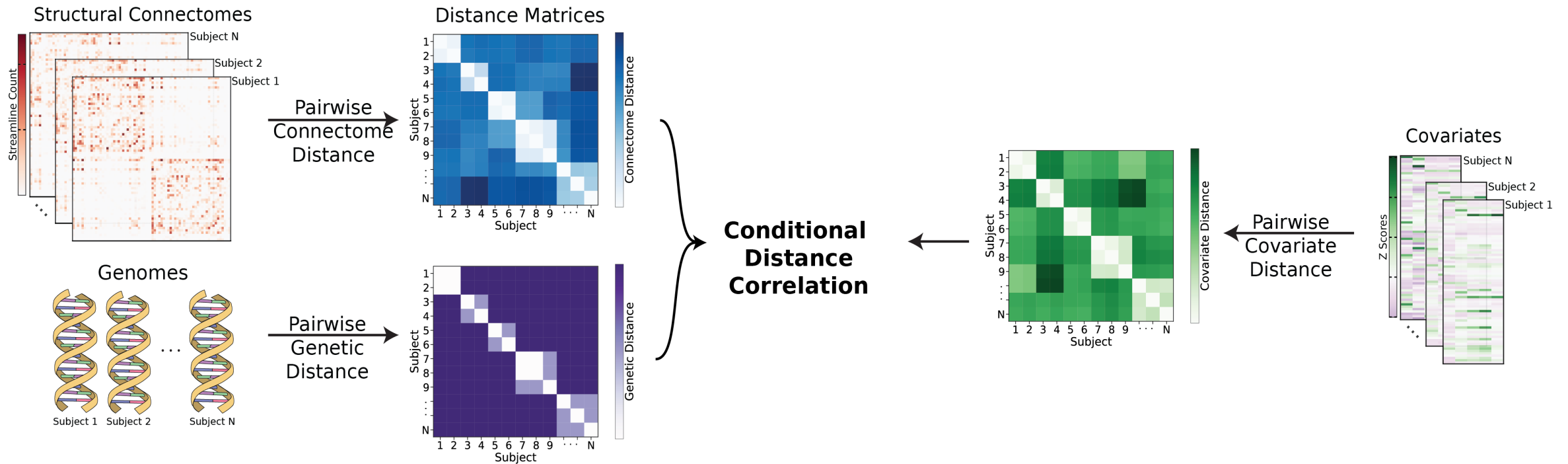
$$H_0 : F(\text{Conn.}, \text{Genome} | \text{Covariates}) = F(\text{Conn.} | \text{Covariates}) F(\text{Genome} | \text{Covariates})$$

$$H_A : F(\text{Conn.}, \text{Genome} | \text{Covariates}) \neq F(\text{Conn.} | \text{Covariates}) F(\text{Genome} | \text{Covariates})$$

- Known as conditional independence test
- Test statistic: Conditional distance correlation (cdcorr)
- Implication if false: there exists causal dependence of connectomes on genomes.

# Conditional distance correlation

- Augment distance correlation procedure with third distance matrix.



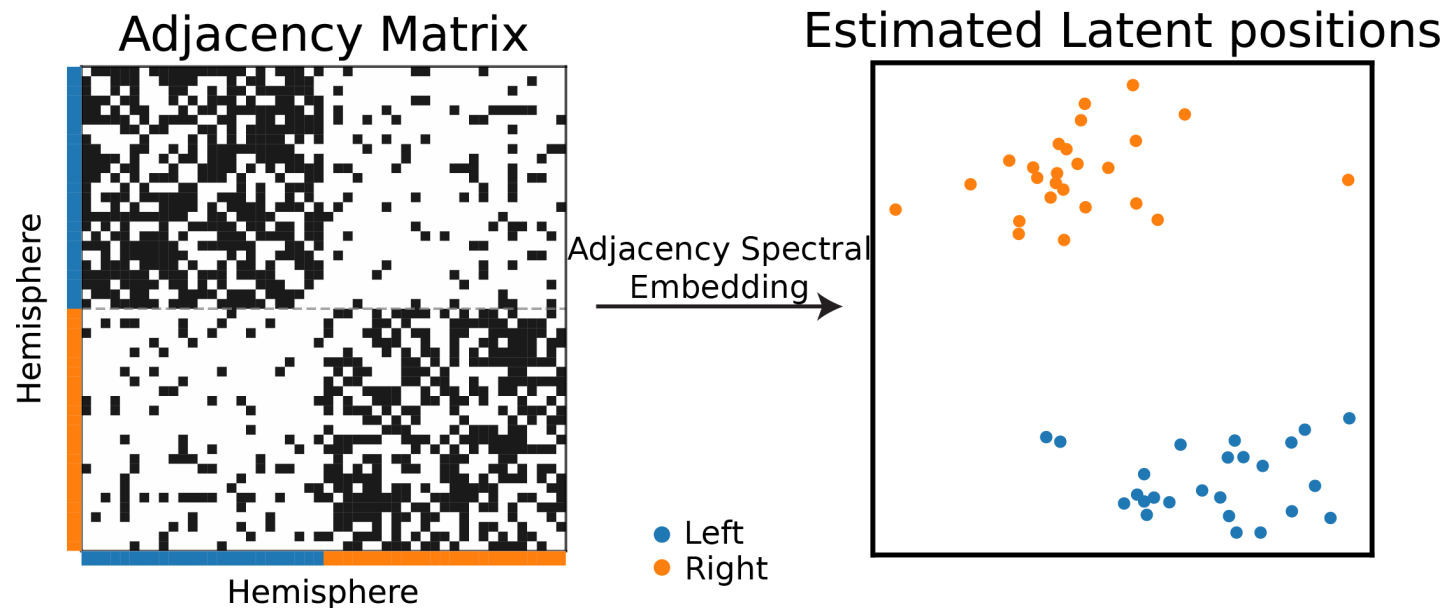
# How do we compare genomes?

- Neuroimaging twin studies do not sequence genomes.
- Coefficient of kinship ( $\phi_{ij}$ )
  - Probabilities of finding a particular gene at a particular location.
- $d(\text{Genome}_i, \text{Genome}_j) = 1 - 2\phi_{ij}$ .

Relationship	$\phi_{ij}$	$1 - 2\phi_{ij}$
Monozygotic	$\frac{1}{2}$	0
Dizygotic	$\frac{1}{4}$	$\frac{1}{2}$
Non-twin siblings	$\frac{1}{4}$	$\frac{1}{2}$
Unrelated	0	1

# How do we compare connectomes?

- Random dot product graph (RDPG)
  - Each vertex (region of interest) has a low  $d$  dimensional latent vector (position).
  - Estimate latent position matrix  $X$  via adjacency spectral embedding.



- $d(\text{Connectome}_k, \text{Connectome}_l) = ||X^{(k)} - X^{(l)} R||_F$

# Neuroanatomy (mediator), Age (confounder)

- Literature show:
  - neuroanatomy (e.g. brain volume) is highly heritable.
  - age affects genomes and potentially connectomes
- $d(\text{Covariates}_i, \text{Covariates}_j) = \|\text{Covariates}_i - \text{Covariates}_j\|_F$

# Human Connectome Project

- Brain scans from identical (monozygotic), fraternal (dizygotic), non-twin siblings.
- Regions defined using Glasser parcellation (180 regions).

Zygosit y	Monozygoti c	Dizygotic	Non-twin siblings
N	322	212	490
Sex	196 F, 126 M	125 F, 87 M	237 F, 253 M
Age (mean )	29.6 (3.3)	28.9 (3.4)	28.3 (3.9)
Age (range )	22-36	22-36	22-37

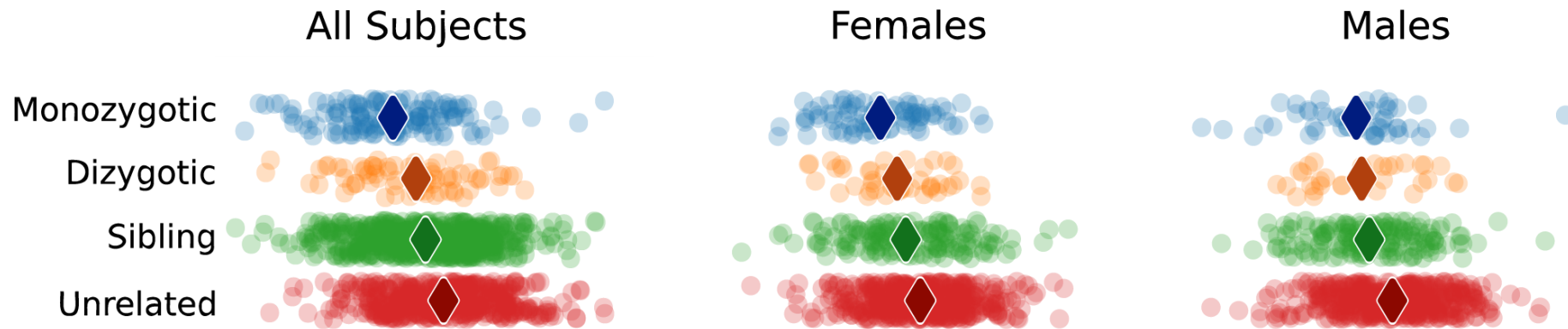
Van Essen, David C., et al., The WU-Minn human connectome project: an overview (2013)

Glasser, Matthew F., et al. "A multi-modal parcellation of human cerebral cortex." Nature (2016).



# Associational Test for Connectomic Heritability

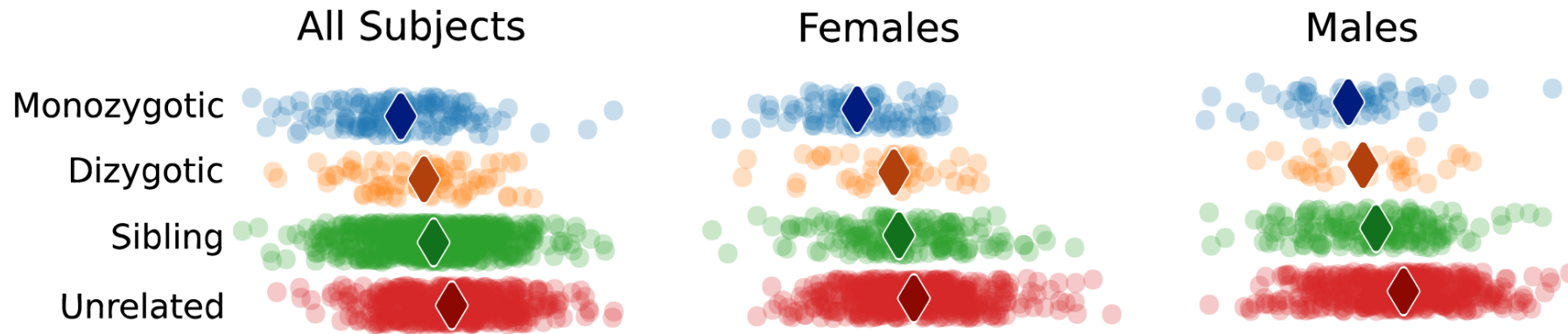
- $H_0 : F(\text{Connectome}, \text{Genome}) = F(\text{Connectome})F(\text{Genome})$   
 $H_A : F(\text{Connectome}, \text{Genome}) \neq F(\text{Connectome})F(\text{Genome})$



Sex	All	Females	Males
p-value	$< 1 \times 10^{-5}$	$< 1 \times 10^{-3}$	$< 1 \times 10^{-2}$

# Associational Test for Neuroanatomy

- $H_0 : F(\text{Neuroanatomy}, \text{Genome}) = F(\text{Neuroanatomy})F(\text{Genome})$   
 $H_A : F(\text{Neuroanatomy}, \text{Genome}) \neq F(\text{Neuroanatomy})F(\text{Genome})$



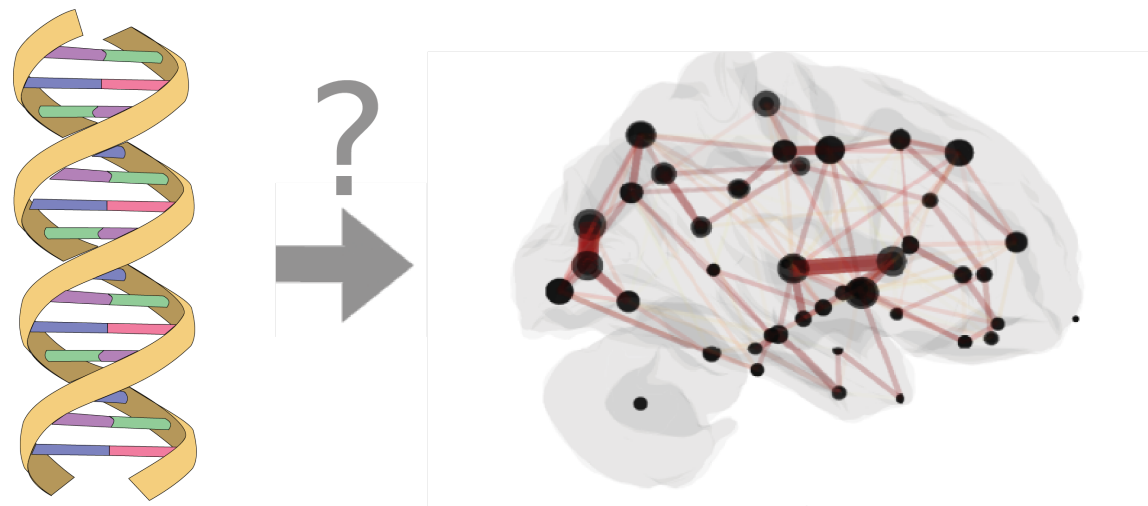
Sex	All	Females	Males
p-value	$< 1 \times 10^{-3}$	$< 1 \times 10^{-2}$	$< 1 \times 10^{-2}$

# Causal Test for Connectomic Heritability

- $H_0 : F(\text{Conn.}, \text{Genome} | \text{Covariates}) = F(\text{Conn.} | \text{Covariates}) F(\text{Genome} | \text{Covariates})$   
 $H_A : F(\text{Conn.}, \text{Genome} | \text{Covariates}) \neq F(\text{Conn.} | \text{Covariates}) F(\text{Genome} | \text{Covariates})$

Sex	All	Females	Males
p-value	$< 1 \times 10^{-2}$	$< 1 \times 10^{-2}$	$< 1 \times 10^{-2}$

# Summary



- Present a causal model for heritability of connectomes.
- Leveraged recent advances:
  - i. Statistical models for networks, allowing meaningful comparison of connectomes.
  - ii. Distance and conditional distance correlation as test statistic for causal analysis<sup>1</sup>.
- Connectomes are dependent on genome, suggesting heritability.

<sup>1</sup> Bridgeford, Eric W., et al. "Batch Effects are Causal Effects: Applications in Human Connectomics." (2021).

