Heritability of Human Structural Connectomes

Jaewon Chung

(he/him) - NeuroData lab

Johns Hopkins University - Biomedical Engineering

j1c@jhu.edu

@j1c (Github)

@j1c (Twitter)

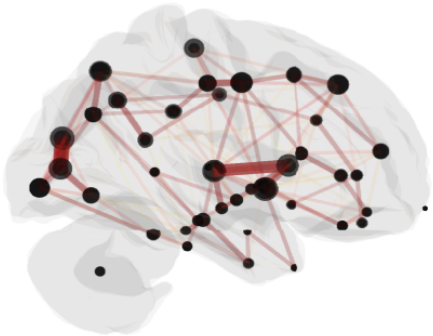
What is heritability?

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

Are the brain connectivity patterns heritable?

Brain connectivity as connectomes

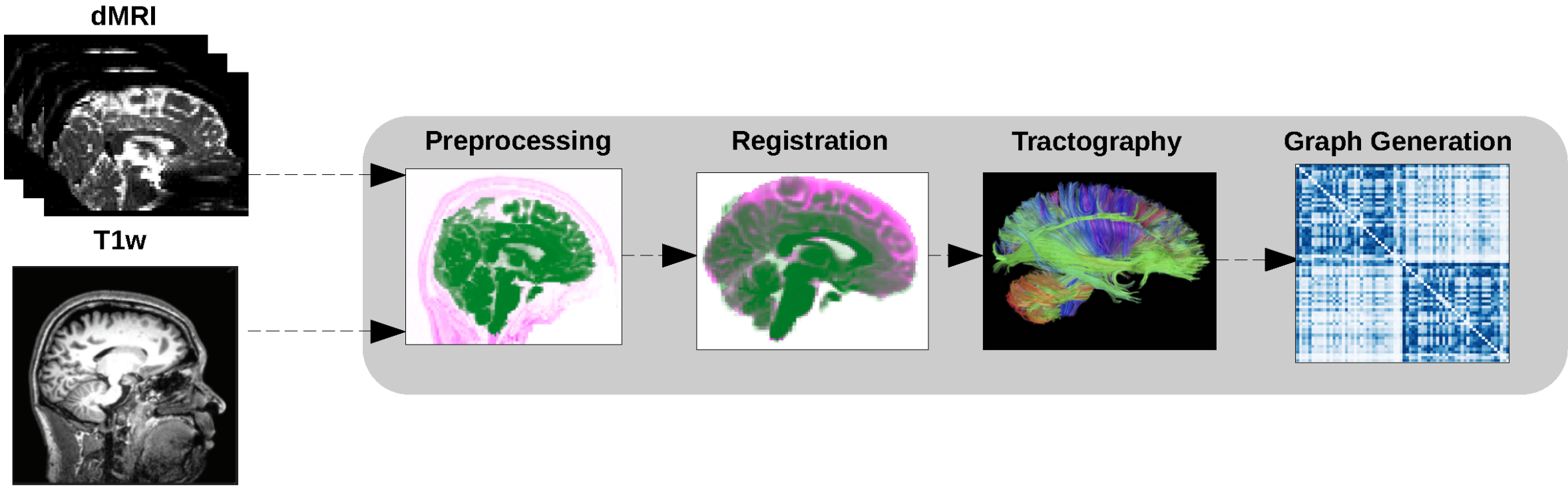
- Vertex: region of interest
- Edges: connectivity measure between a pair of vertices
- Diffusion MRI: # of estimated neuronal fibers
- Undirected: neurons have no direction



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

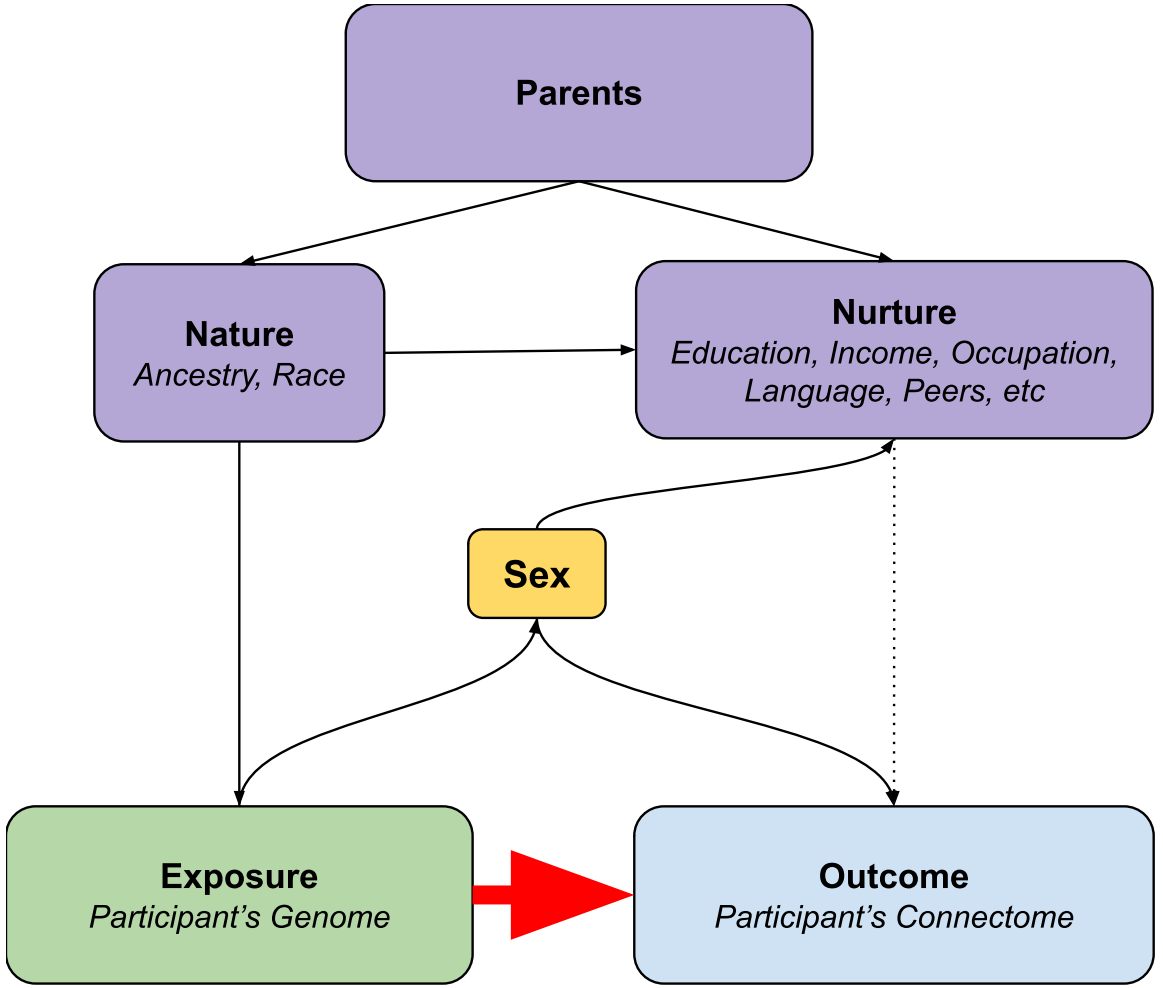
Image from Gu, Zijin, et al. "Heritability and interindividual variability of regional structure-function coupling." (2021)

How do we get structural connectomes?



Heritability as causal problem

- Directed acyclic graph

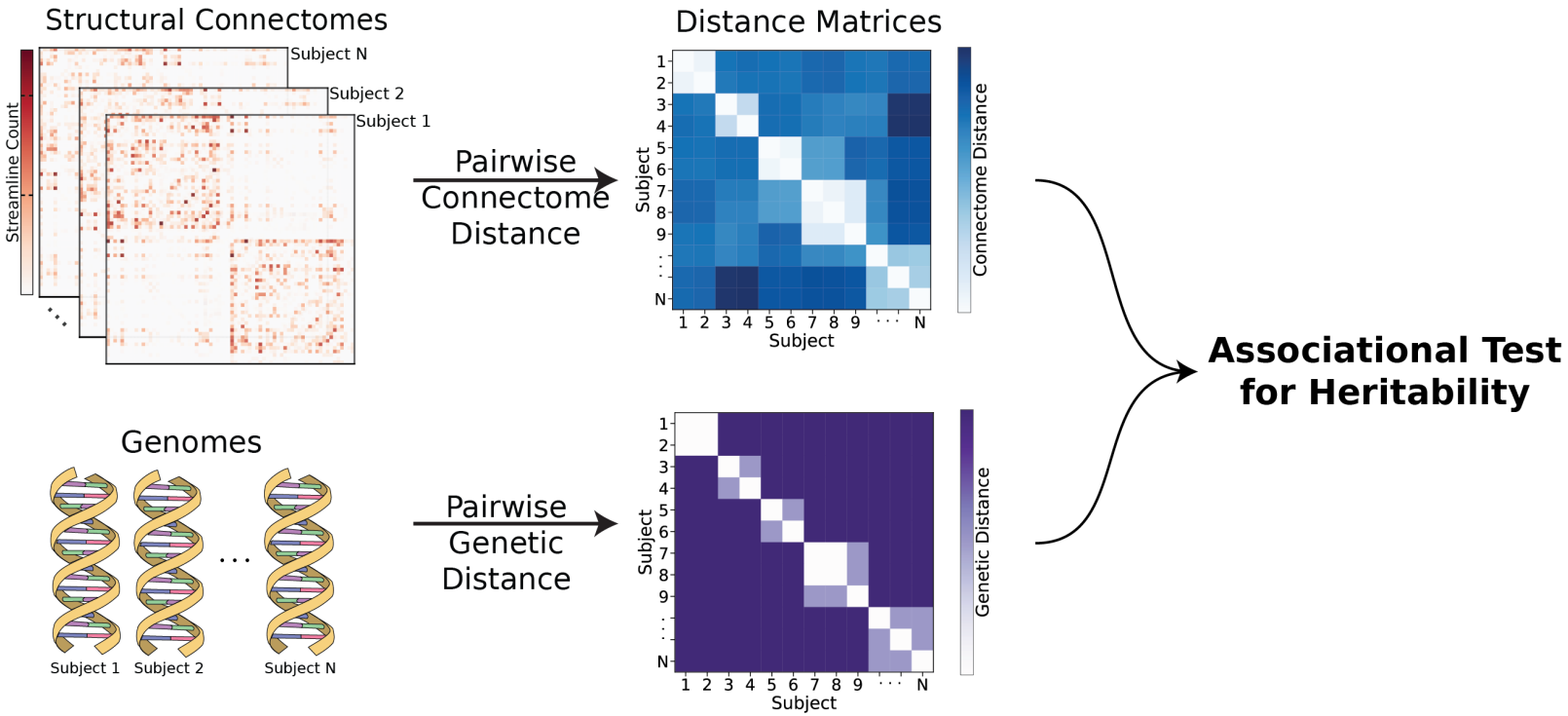


Do genomes affect connectomes?

- 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 true: there exists an associational heritability.

What is distance correlation?

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



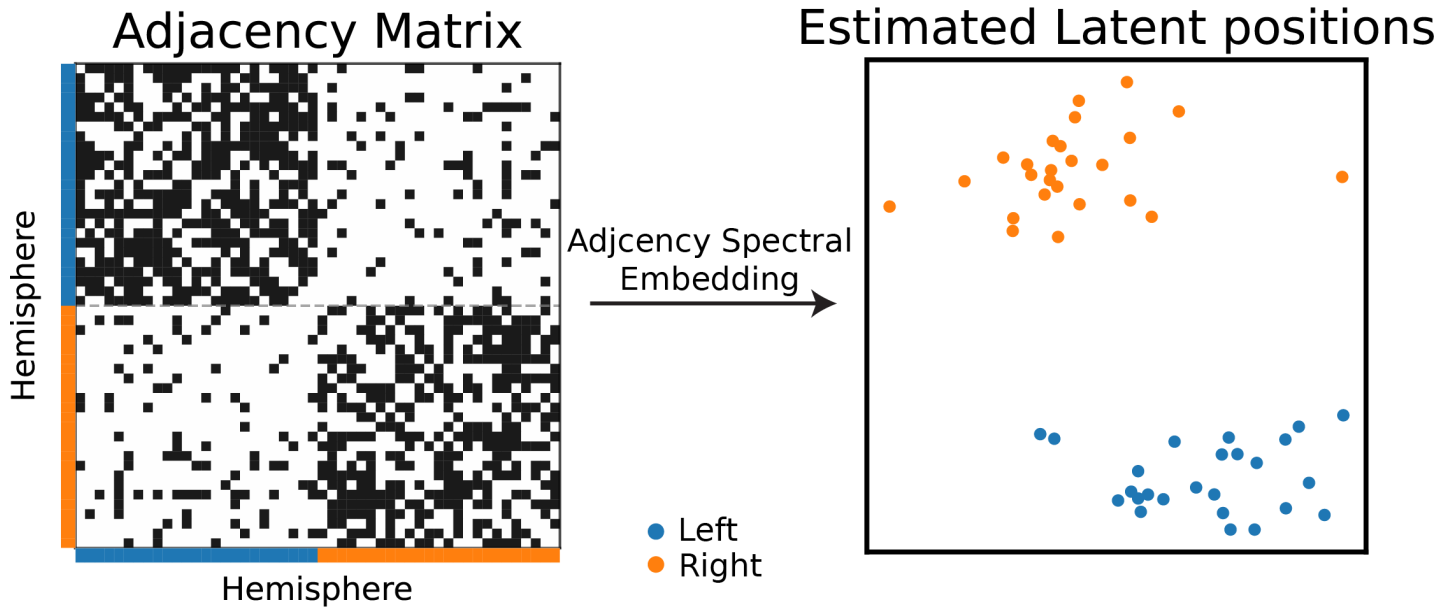
How to compare genomes?

- Typical twin studies do not sequence genomes.
- Coefficient of kinship (ϕ_{ij})
  - Probabilities of finding particular genes as identical among subjects.
- d(Genome$_i$, Genome$_j$) = 1 - 2ϕ_{ij}.

Relationship	ϕ_{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 to compare connectomes?

- Random dot product graph (RDPG)
  - Each vertex (region of interest) has a low d dimensional latent vector.
  - $P[i \rightarrow j] = \langle x_i, x_j \rangle$
- Latent vectors =
- d(Connectome$_k$, Connectome$_l$) = $||X^{(k)} - X^{(l)}R||_F$



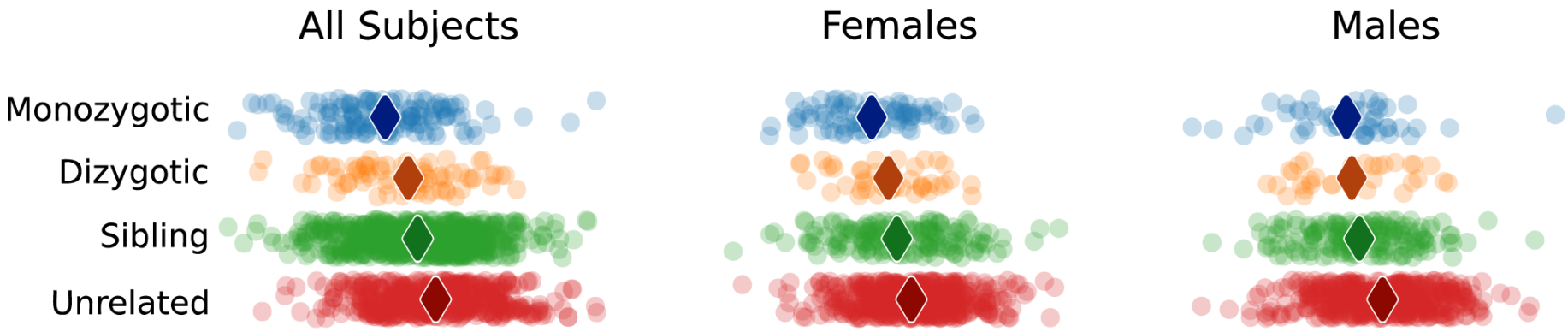
Human Connectome Project

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

Zygosity	Monozygotic	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).

Genome and connectomes are dependent

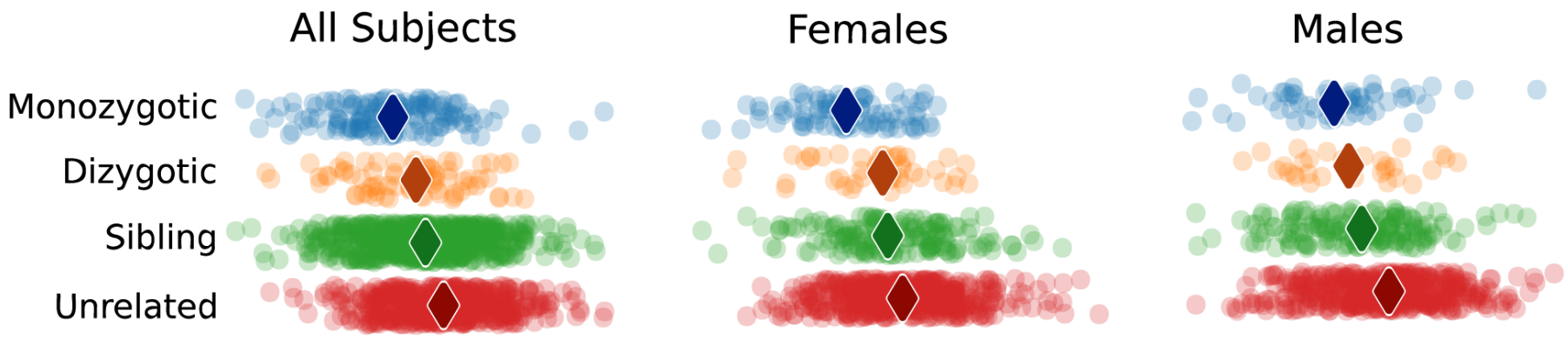


Sex	All	Females	Males
p-value			

Neuroanatomy (effect mediator)

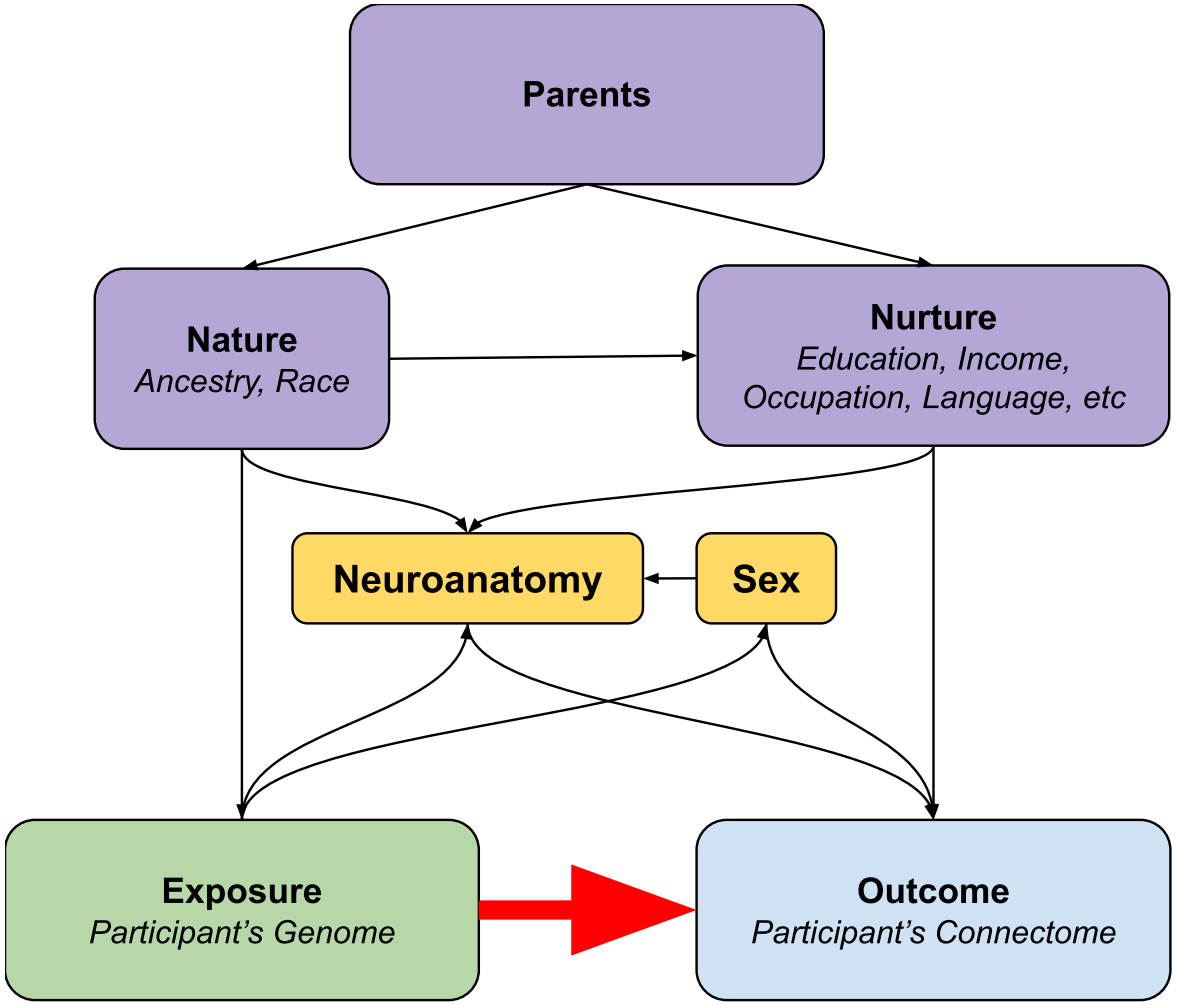
- Literature show neuroanatomy (e.g. brain volume) is highly heritable.
- Want to test:
  $H_0 : F(\text{Neuroanatomy}, \text{Genome}) = F(\text{Neuroanatomy})F(\text{Genome})$
  $H_A : F(\text{Neuroanatomy}, \text{Genome}) \neq F(\text{Neuroanatomy})F(\text{Genome})$
- Implication if true: causal model should include neuroanatomy.

Genome and neuroanatomy are dependent



Sex	All	Females	Males
p-value			

DAG including interactions of neuroanatomy



Do genomes affect connectomes given neuroanatomy?

- Want a conditional independence test!

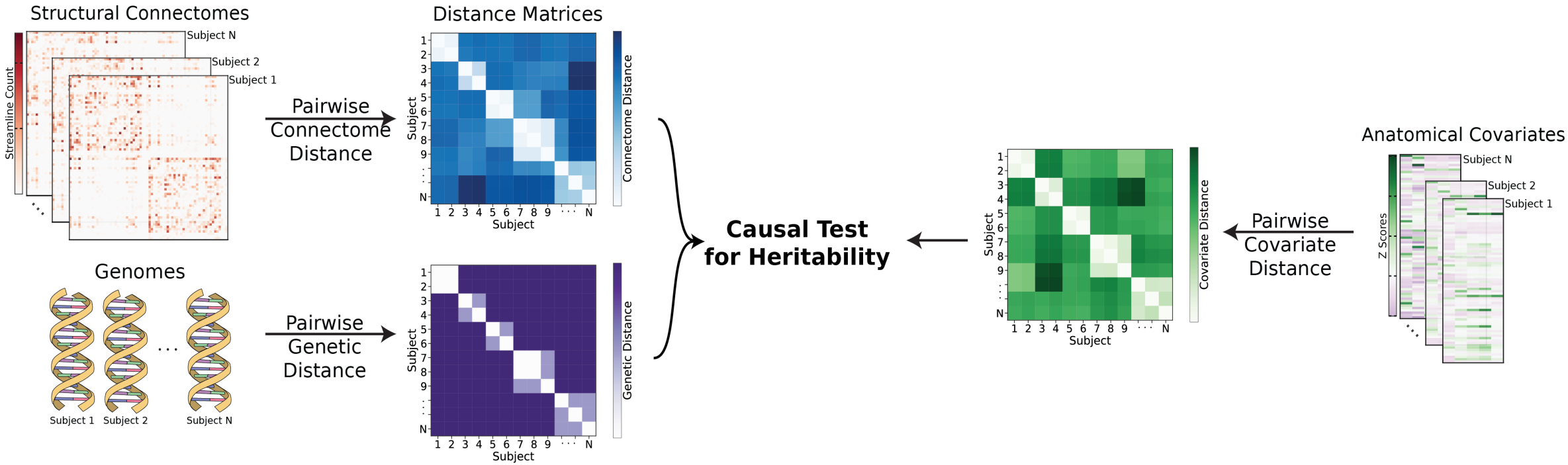
  $H_0 : F(\text{Conn.}, \text{Genome}|\text{Neuro.}) = F(\text{Conn.}|\text{Neuro.})F(\text{Genome}|\text{Neuro.})$

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

- Test statistic: Conditional distance correlation (cdcorr)
- Implication if true: there exists causal dependence of connectomes on genomes.

What is conditional distance correlation?

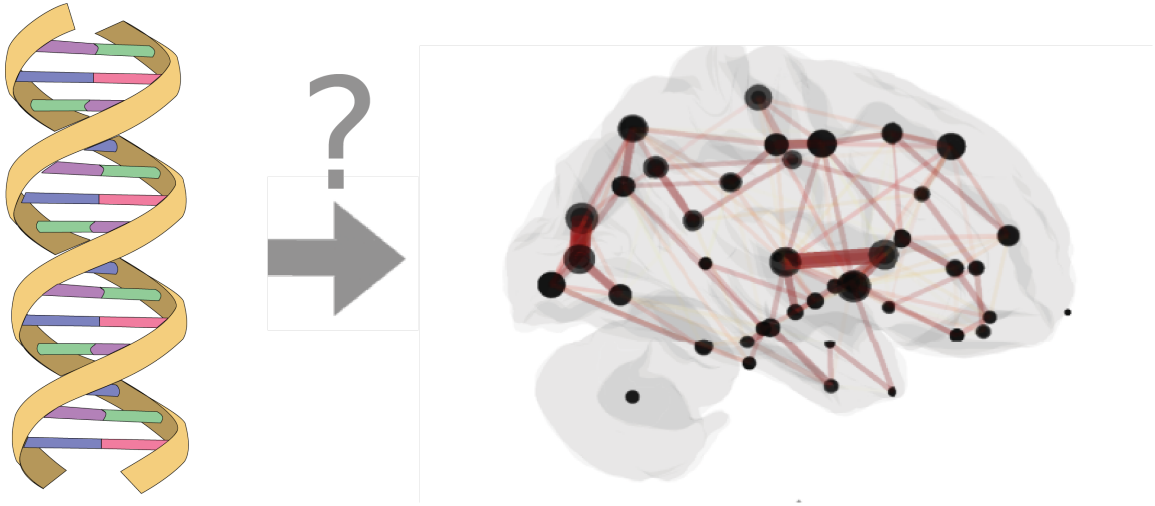
- Augment distance correlation procedure with third distance matrix.
- d(Neuroanatomy$_i$, Neuroanatomy$_j$) = ||Neuroanatomy$_i$ - Neuroanatomy$_j$||$_F$



Connectomes are still dependent on genome

Sex	All	Females	Males
p-value			

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[1].
- Connectomes are dependent on genome, suggesting heritability.

[1] Bridgeford, Eric W., et al. "Batch Effects are Causal Effects: Applications in Human Connectomics." (2021).

Acknowledgements

Team

Mike Powell

Eric Bridgeford

Carey Priebe

Joshua Vogelstein

Additional slides

Causal model

- X denote exposure, Y denote outcome, W denote measured covariates, Z denote unmeasured covariates
- Want to estimate the effect of different exposures on the outcome, which is quantified using the backdoor formula if W and Z close all backdoor paths.

$$f_{w,z}(y|x) = \int_{\mathcal{W}\times\mathcal{Z}} f(y|x,w,z)f(w,z)\mathrm{d}(w,z)$$

- Above integrates over *all* measured and unmeasured covariates.

$$f(y|x) = \int_{\mathcal{W}\times\mathcal{Z}} f(y|x,w,z)f(w,z|x)(w,z)$$

- Averages the true outcome distribution over the *conditional* distribution of the measured and unmeasured covariates.

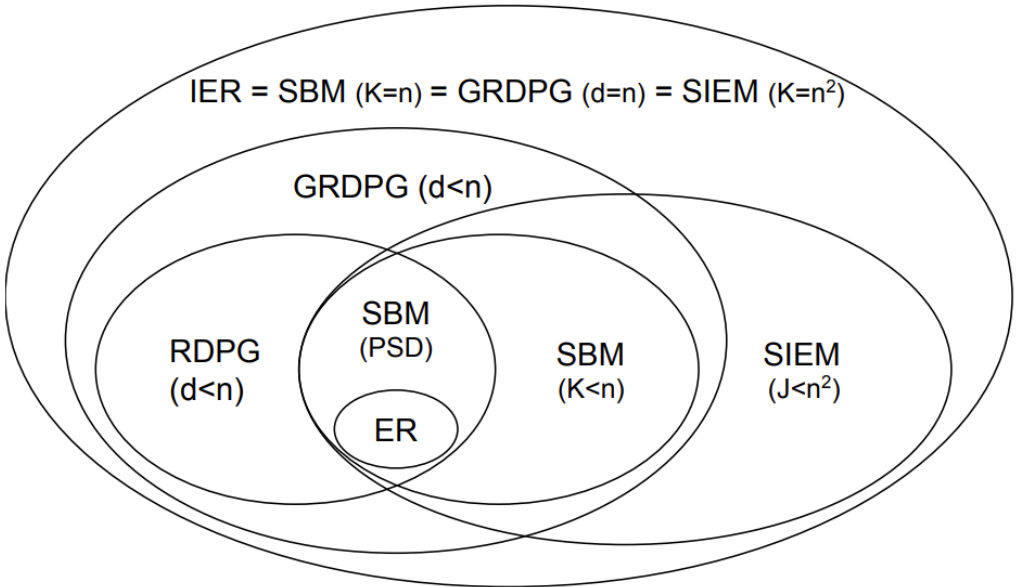
Causal model (cont.)

- We observe the triples (x_i, y_i, w_i) for $i \in [n]$.
- Only be able to estimate the functions of (X, Y, W)
- The corresponding hypothesis test is:

$$H_0 : f(y|x, w) = f(y|w) \quad \text{vs} \quad H_A : f(y|x, w) \neq f(y|w).$$

Shortcomings - Network model

- Problems with connectome estimation.
  - Inability to determine the precise origin/termination of connections in the cortex.
    - -> false negatives
  - Crossing fibers
    - -> false positives
- RDPG can only represent subset of independent edge networks.



Shortcomings - Model assumptions

- No interaction between genome and environment
- No epistatsis
  - Effect of one gene is dependent on another
  - Ex: black hair and baldness
- No dominance effects
- Strong assumptions in genetic distances

What are environmental effects?

- Shared
  - Common experiences of siblings living in the same household.
    - household income, the family’s living situation, the dynamics between the parents, food consumed
- Non-shared
  - Everything else
  - Epigenetics
  - Luck
  - schools, peers

Random dot product graphs

- Adjacency spectral embedding
- representation of the vertices of the graphs into d dimensions via its singular value decomposition, given by $A = USU^\top$ where $U \in \mathbb{R}^{n \times n}$ is the orthogonal matrix of eigenvectors and $S \in \mathbb{R}^{n \times n}$ is a diagonal matrix containing the eigenvalues of A ordered by magnitude.
- $ASE(A) = \hat{X} = \hat{U}\hat{S}^{1/2}$ where $\hat{U} \in \mathbb{R}^{n \times d}$ contains the first d columns of U, which correspond to the largest eigenvectors, and $\hat{S} \in \mathbb{R}^{d \times d}$ is the submatrix of $ S $ corresponding to the d largest eigenvalues in magnitude.