Comparing the connectome similarity of monozygotic twins

CS T680 – Computational Neuroscience

Abstract

The goal of this study is to explore to what degree monozygotic (genetically identical) twins have similarities in their connectomes. The connectomes of monozygotic twins are compared to each other and then to connectomes of non-twins, and it is discovered that the monozygotic twins have much higher connectomic similarity than that of the average of other nontwins, with an effect size of 1.56 and $p < 10^{-6}$. The same was observed with dizygotic twins although to a lesser degree, with an effect size of 0.647 and a p-value of 0.00065. No relationship was observed between monozygotic and dizygotic twin pair's connectome correlation with each other and sex or age. We assume that monozygotic twin brain networks are very similar because of their nearly identical genetic origin, while dizygotic twin brain networks are similar because of their familial relationship.

Introduction

Monozygotic twins, otherwise known as "identical" twins, are the product of a natural phenomenon that occurs when two human embryos are formed from a single zygote. The monozygotic twins born from this rare situation are physically similar and, with the exception of minuscule differences (Haque 2009), have an identical genome. This means that from birth these monozygotic twins have nearly identical organ structure, including their brain. We know that the structural connectivity pathways within the brain change as it learns new information, so monozygotic twins will not have identical structural connectivity in their brains especially later on in life. But how similar are their brains, as compared to other non-twin brains? This question is what this study attempts to answer.

The human brain has been an expanding field of study in the last several decades. Neuronal systems in human beings have been shown to be extremely complex and interconnected, full of billions of neurons. The field of brain imaging has enabled scientists to explore structural and functional connectivity at a deep scale using technologies such as magnetic resonance images (MRI) and other neuroimaging techniques. Connectomes are one way to study connectivity in the brain's neuronal network (Sporns 2005). They can allude to similarities and differences in the brain through structural or functional differences. Connectomes are focused, compact graphs derived from diffusion or functional MRI of the brain wherein their nodes represent regions of the brain, and their edges represent streamlines connecting regions of the brain.

The goal of this study will be to explore to what degree genetically identical twins have similarities in their connectomes. By comparing connectomes from monozygotic twins to each other and then to connectomes of non-twins, the relative contributions of genes and environment in shaping brain circuitry may be discovered. Later in the study we will also explore the degree that dizygotic twin connectomes are like each other as compared to monozygotic twins and nontwins. We will compare these connectomes using measures such as group differences.

Participants and Data Acquisition

The participants in the data that will be used included 117 pairs of genetically confirmed monozygotic twin pairs and 60 dizygotic same-sex twin pairs, as well as 619 non-twin patients (including 69 of the monozygotic twin's non-twin siblings and 48 of the dizygotic twins non-twin siblings)]. Monozygotic twins are assumed to be almost genetically identical whereas Dizygotic twins share about half of their DNA, which happens to be similar to the amount of DNA that non-twin siblings share.

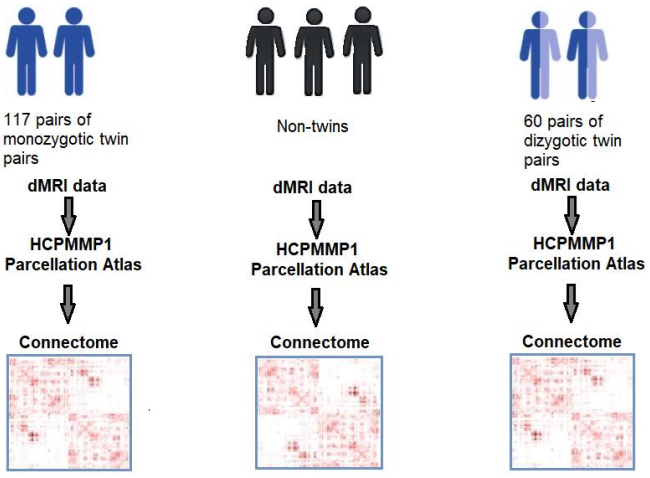
The data used in this study was obtained from a 1200 participant study done by the Washington University in Saint Louis, the University of Minnesota, and Oxford University as part of the Human Connectome Project (HCP) (Varga 2021). The cohort of monozygotic and dizygotic twin pairs from this HCP study are vital to this paper. The data was acquired on a customized Siemens 3T "Connectome Skyra" scanner at Washington University in St Louis, Missouri. Further details about the neuroimaging process can be found in Materials and Methods section of this paper.

Materials and Methods

Minimally processed DWI and structural data from the Human Connectome Project for 972 participants (age mean ± standard deviation: 28.7 ± 3.7, 522 females) were used, including a cohort of monozgotic and dizygotic twin pairs together with their non-twin siblings. Data was acquired on a customized Siemens 3T "Connectome Skyra" scanner at Washington University in St Louis, Missouri using a multi-shell protocol for the DWI (1.25 mm3 isotropic voxels, repetition time (TR) = 5520 ms, echo time (TE) = 89.5 ms, field-of-view (FOV) of 210 × 180 mm, 270 directions with b = 1000, 2000, 3000 s/mm2 (90 per b value), and 18 b = 0 volumes). Structural T1-weighted data were collected using 0.7 mm3 isotropic voxels, TR = 2400 ms, TE = 2.14 ms, FOV of 224 × 224 mm. Subject recruitment procedures and informed consent forms, including written informed consent to share de-identified data, were approved by the Washington University institutional review board.

In terms of pre-processing, normalization across diffusion acquisitions and correction for EPI susceptibility and signal outliers, eddy-current-induced distortions, slice dropouts, gradient-nonlinearities, and subject motion were performed. T1-weighted data was corrected for gradient and readout distortions. (Arnatkeviciute 2021)

Parcellation was done using the HCPMMP1 parcellation atlas which uses 360 regions (180 per brain hemisphere). The HCPMMP1 parcellation atlas determines nodes based on structural and functional connectivity information. 10 million streamlines were generated on a probabilistic basis using a dynamic seeding approach. Connection weights were quantified using both streamline count (number of streamlines connecting two regions, SC) and the mean fractional anisotropy (FA) of voxels traversed by streamlines connecting two regions. The accuracy of all tractography



methods is still an open challenge for the field, but data processing pipeline in the creation of this connectome was designed to limit errors from spurious streamlines and head motion.

The computed connectomes will be representative group-level connectivity matrices containing 12,924 unique connections between 360 brain regions defined by the HCPMMP1 atlas. They will be for monozygotic and dizygotic twin pairs as well as their non-twin siblings, which will be compared in the Results section of this paper. Comparisons will be made for monozygotic twin connectomes, dizygotic twin connectomes, and non-twin siblings connectomes (see figure 1).

We will be using group difference to determine to what degree monozygotic twins have similar structural connectivity as compared to nontwins. We will be using 117 monozygotic twin pairs (234 individual patients). A similarity matrix is first computed for all 234 monozygotic twin patients. This similarity matrix uses Pearson's correlation between flattened patient connectomes. This is shown in Figure 2 on the left. The nontwins for each twin pair in this case refer to all other 116 monozygotic twin pairs because they are unrelated to the relevant twin pair.

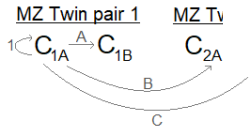


Figure 1

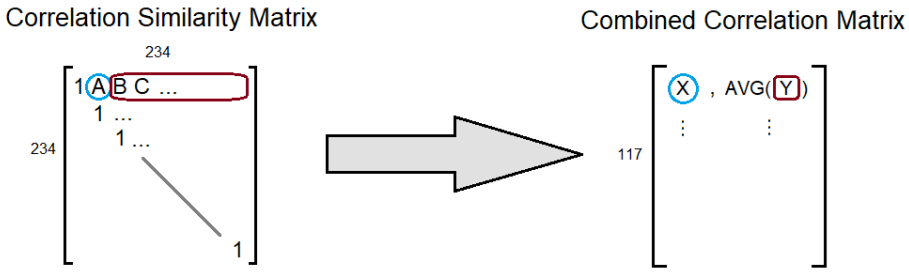


Figure 2

If twin A and twin B represent a monozygotic twin pair, and twin A represents the first row of the similarity matrix, then its correlation with twin B (represented by 'A' in Fig. 2) would be moved to a new matrix (denoted by 'X' in Fig. 2). Additionally, the average between twin A and all its nontwins ('B', 'C', and the rest of the row in Fig. 2) would be moved to the new matrix ('AVG(Y)' in Fig. 2). This new matrix will be called the combined correlation matrix and it contains these two correlation values per each monozygotic twin pair. Group difference will then be performed between both columns of this matrix.

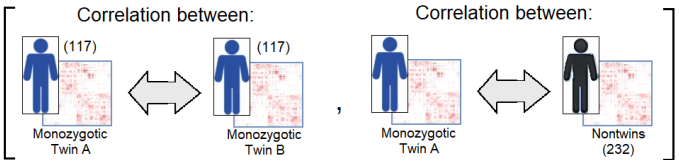


Figure 3 – Visual of the Combined Correlation Matrix

In choosing a parametric or non-parametric approach to calculating group difference, we need to discover the normality and equivalence of variance in both data distributions (columns of the combined correlation matrix). If the data in these proposed group differences in normally distributed, and the data distributions within each proposed group difference have equal variance, then we can use a parametric group difference. In this case, a student t-test for independent variables will be performed. This statistical analysis will ensure that any

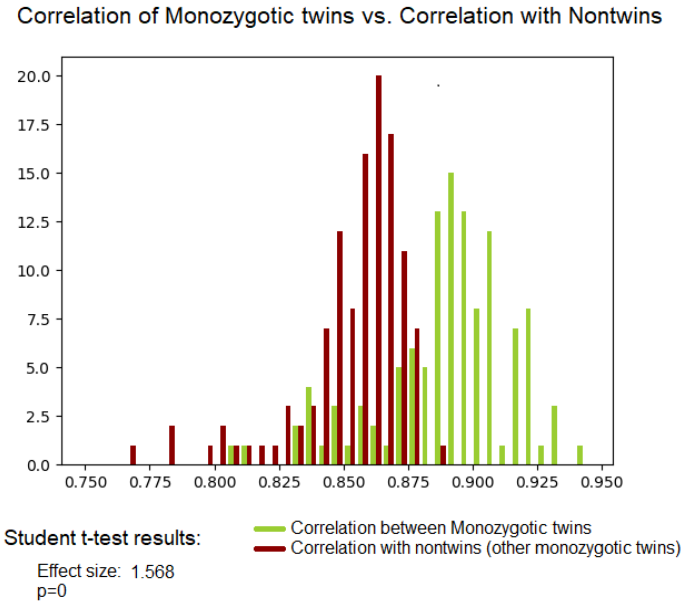
observed differences in structural connectivity between the two groups are real and not just a chance difference caused by the natural variation within the measurements.

Patient sex and age are data in addition to the connectome itself that we identify as relevant patient information. The potential relationship of age or sex of monozygotic twins with twin pair correlation will be determined.

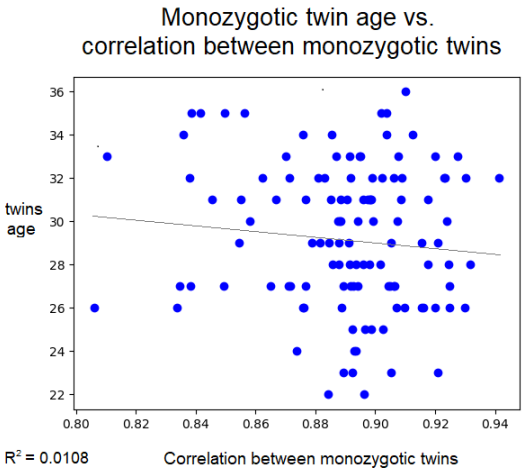
The above steps for computing similarity matrix, combined correlation matrix, and performing group difference will also be done for the have 60 dizygotic twin pairs (120 individual patients) as well as 119 of the 619 nontwin patients (including nontwin siblings to these dizygotic and monozygotic twin pairs). Two nontwin patients will be selected to act as twins so we can perform the same methods on our nontwin dataset. This gives us a useful control dataset, since dizygotic twin pairs effectively have sibling genetics which may affect the correlation of their connectomes.

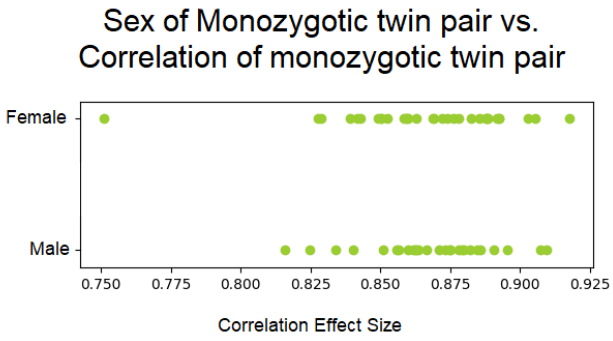
Results:

Computing the similarity matrix, combined correlation matrix, and then performing group difference on the two resultant datasets for monozygotic twins produces the following histogram. Because these datasets exhibit a normal distribution and have equal variance, a student-t test is used. The results are an effect size of 1.568 with a p-value of $p < 10^{-6}$ (virtually zero). As the statistical analysis shows, there is a definite strong difference between these two datasets.



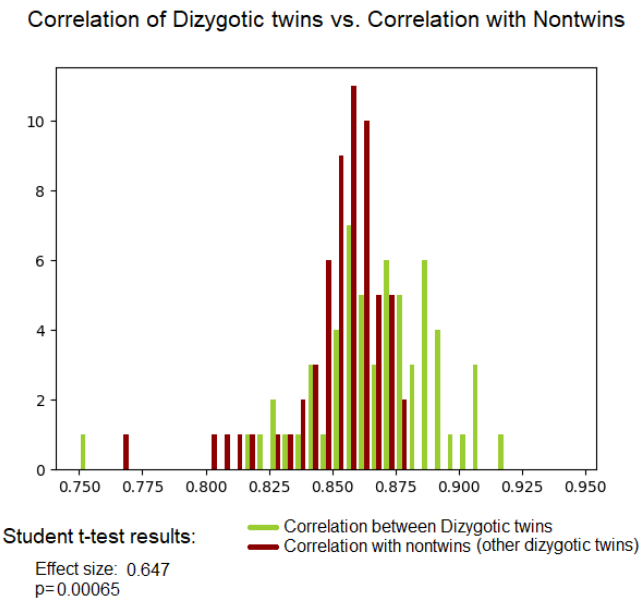
The relationship of the sex of monozygotic twins with twin pair correlation was also investigated, along with monozygotic twin pair age. The scatter plots conveying these relationships are shown in figures. There is no observable relationship that can be drawn from either the sex or age of monozygotic twins with twin pair correlation.

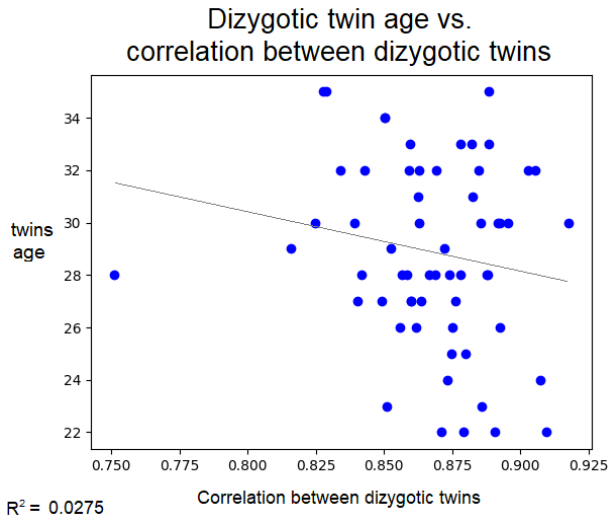


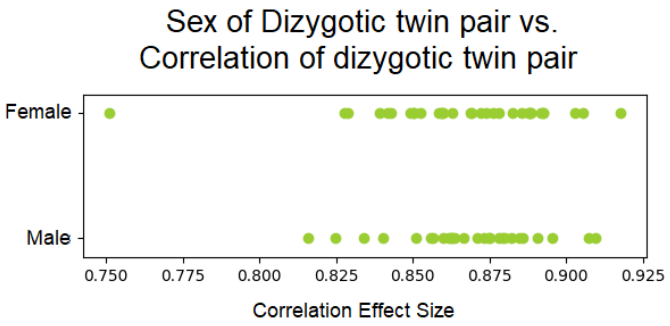


Computing the similarity matrix, combined correlation matrix, and then performing group difference on the two resultant datasets for dizygotic twins produces the following histogram. Like the monozygotic twins dataset, these datasets exhibit a normal distribution and have equal variance so a student-t test is used. The Student-t test returns an effect Size of 0.647 with a p-value of 0.00065. As the statistical analysis shows, there is a difference between these two datasets.

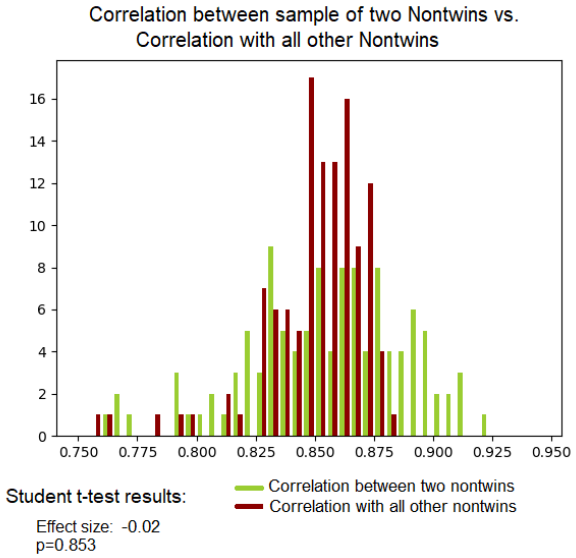
Just as we did with monozygotic twins, the relationship of the sex and age of dizygotic twins with twin pair correlation was also investigated. No observable relationship was drawn from either the sex or age of dizygotic twins with twin pair correlation.



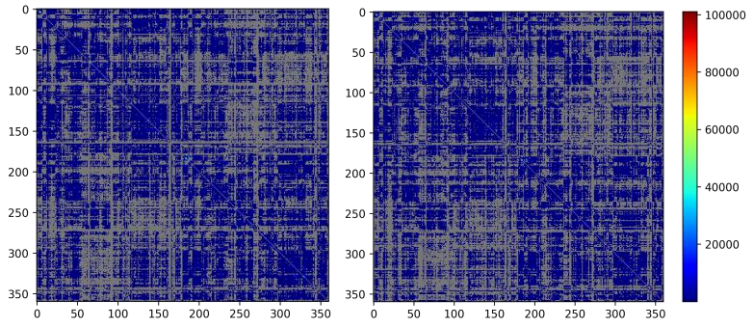




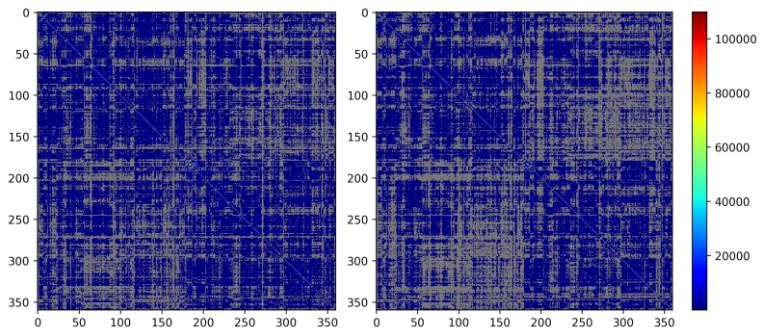
Lastly, the same process of computing a similarity matrix, combined correlation matrix, and then performing group difference was done using a sample of 119 nontwin patients. This was done by selecting two nontwin patients to act as twins so we can perform the same methods on our nontwin dataset, for the purpose of having a control dataset. We verified that the datasets have a normal distribution and have equal variance, and so a student-t test was used. The results are an effect size of -0.02 with a p-value of 0.853. As the statistical analysis shows, there is a not a statistically relevant relationship between these two datasets. This is as expected as the nontwin patients act as a control dataset.



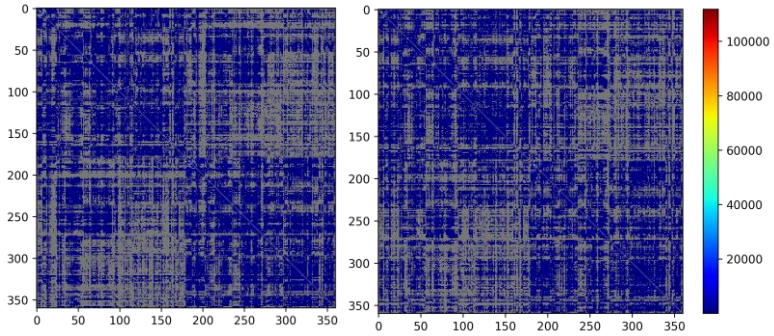
The heatmaps for a pair of monozygotic twins, dizygotic twins, and a nontwin connectomes are shown below:



Monozygotic twins



Dizygotic twins



Nontwins

Discussion:

Our results show that there is a strong difference between the correlations of monozygotic twin pairs with each other and the correlations of a monozygotic twin with unrelated monozygotic twins (effectively unrelated nontwins). Because this difference is statistically significant, it shows that monozygotic twin pairs have a higher connectomic similarity with each other than that of an unrelated nontwin. Or in other words, it appears true that monozygotic twins brains networks resemble each other more than nontwins. It is interesting to note that although monozygotic twin connectomes have around 90% correlation with each other, they are not 100% similar. This can be explained by the neuroplasticity of the brain. The structural connectivity of the brain can change due to a variety of factors such as different social and learning experiences, which monozygotic twins can have as well.

When exploring monozygotic twin connectome correlation with age, a possible hypothesis is that the connectomes of younger monozygotic twins would resemble each other more because their brains were identical when first formed and before they independently began learning anything. However, our results show a lack of a correlation with age. This may indicate that there is no change among the brains of twins with age. However, the monozygotic and dizygotic twin patients that participated in this study were over the age of 20. It may be that although there is no correlation in monozygotic twins over age 20, younger monozygotic twins closer to birth may show more correlation. More research would be required to explore these ideas further.

Exploring monozygotic twin connectome correlation with sex also showed a lack of correlation with the sex of patient.

The same statistical analysis that was performed on monozygotic twin pairs was performed on dizygotic twin pairs. Our results on the connectomic similarity of dizygotic twins yielded interesting results. Our expected result was that the dizygotic twins would have the same level of connectomic similarity as unrelated nontwins. However, it appears based on our results that dizygotic twins connectomes resemble each other more than unrelated nontwins. This observed result could be explained by dizygotic twins being, for all purposes, siblings. Dizygotic twins on average share half of their DNA, which is similar to non-twin siblings (Arnatkeviciute 2021). As siblings, dizygotic twins would have higher correlation than nontwins from different families.

Exploring dizygotic twin connectome correlation with both sex and age showed lack of correlation, similar to the results found with monozygotic twins.

Our results showed the absence of a significant difference when comparing the correlations of the unrelated nontwins with a randomly sampled pair. This is to be expected because this is our control dataset.

The nontwins control group we used in our study consists of a sample of 119 nontwins from the HCP dataset. This includes both nontwin siblings of twins from the dataset as well as unrelated nontwins. Although we discovered more similarity between dizygotic twins as compared to all nontwins, it may be of interest to determine how similar dizygotic twins are to their nontwin siblings. We were unable to obtain publicly available data on nontwin sibling identification for dizygotic twins within the HCP, which may be protected for patient privacy. This would be a good area to research more in the future

Another area where this study could be expanded is in calculated group difference between graph theory measures using the Brain Connectivity Toolbox. Although this was not performed in this study, we could calculate a matrix with several graph theory measures as the columns and the Twin A of the 117 monozygotic twins as the rows. We would calculate another matrix for Twin B in the monozygotic twins, and then would calculate group difference between each column of both matrices to find any graph theory measures that have significant similarities or differences between monozygotic twin pairs. We could repeat this with dizygotic twins as well. This would also be a good area to research more in the future.

There are several possible problems and topics worth discussing with the data used in this study. First, when comparing monozygotic twin correlation we compare monozygotic twin A to twin B, as well as twin A to all other monozygotic twins in our dataset. But we are treating these other twins as nontwins. It is good to remember that they do in fact each have their own twin, but because these twins from a different family they can thus be considered nontwins. This applies to our statistical analysis with dizygotic twins as well.

Secondly, during our study between monozygotic twin pair correlation and sex, we do not have monozygotic twins of different sexes to compare connectomes with. This is because all monozygotic twins in the HCP dataset are of the same sex. Whether having twin pairs of different sexes would show a relationship between the twin connectome correlation and the sex of the twins remains to be seen, however we would hypothesize that it would not change our results. This applies to correlation study between dizygotic twins and sex as well.

Lastly, our study between monozygotic twin pair correlation and age did not include age-progressed connectomes of the same twins because we did not have such data. Instead, our study used connectome data of various twins at different ages. We would be able to calculate more individual connectome results if we used age-progressed connectomes, which could yield interesting findings. We also don't have any data on monozygotic twins younger than twenty years old. We would hypothesize that having such data, especially connectome data on infants, may show more of a negative relationship where correlation between younger twin patients would be closer to 100%. Although our result showed that there may be no systematic change among brains of monozygotic twins with age, It could be the case that as the twins age grows closer to 20 years old, their connectomes have less correlation with each other because of the neuroplasticity of the brain. This would be a good area to research more in the future.

There are several relevant studies that should be mentioned in relation to familial relationships, heritability, and connectome similarity. The level of genetic similarity between patients affecting connectomic similarity was explored in a study that found each individual has a unique brain "fingerprint" that is unique to them. It found that identical twins only share about 12 percent of structural connectivity patterns and the brain's unique local connectome is sculpted over time, changing at an average rate of 13 percent every 100 days. Results specific to heritability from this study further supports the results found in this study. (Yeh 2016) Another study trained a machine learning model to predict whether a connectome represents a subjects sibling or an unrelated individual (Miranda-Dominguez 2018). Although not relevant to connectomes, another study was able to re-identify individuals (as well as their monozygotic twins) across time points using only behavioral factors included in the HCP using machine learning models. (Adolphs 2020)

Conclusion:

The goal of this study was to explore to what degree genetically identical twins have similarities in their connectomes. We compared the connectomes of monozygotic twins to each other and then to connectomes of non-twins, and discovered that the monozygotic twins have much higher connectomic similarity than that of the average of other nontwins, with an effect size of 1.56 and a p-value of $p < 10^{-6}$. The same was observed with dizygotic twins although to a lesser degree, with an effect size of 0.647 and a p-value of 0.00065. This supports some of the findings on studies done on hereditability in connectomics. No relationship was observed between monozygotic and dizygotic twin pair's connectome correlation with each other and sex or age. We assume that monozygotic twin brain networks are very similar because of their nearly identical genetic origin, while dizygotic twin brain networks are similar because of their familial relationship.

References:

Haque FN, Gottesman II, Wong AHC. 2009. Not really identical: Epigenetic differences in monozygotic twins and implications for twin studies in psychiatry. Am J Med Genet Part C Semin Med Genet 151C:136–141.

Sporns, O., Tononi, G. & Kötter, R. The human connectome: A structural description of the human brain. PLoS Comput. Biol. 1, e42 (2005).

Arnatkeviciute, A., Fulcher, B.D., Oldham, S. et al. Genetic influences on hub connectivity of the human connectome. Nat Commun 12, 4237 (2021). https://doi.org/10.1038/s41467-021-24306-2

Varga, B., Grolmusz, V. The braingraph.org database with more than 1000 robust human connectomes in five resolutions. Cogn Neurodyn 15, 915–919 (2021). https://doi.org/10.1007/s11571-021-09670-5

Yeh F-C, Vettel JM, Singh A, Poczos B, Grafton ST, Erickson KI, et al. (2016) Quantifying Differences and Similarities in Whole-Brain White Matter Architecture Using Local Connectome Fingerprints. PLoS Comput Biol 12(11): e1005203. https://doi.org/10.1371/journal.pcbi.1005203

Han Y, Adolphs R (2020) Estimating the heritability of psychological measures in the Human Connectome Project dataset. PLoS ONE 15(7): e0235860. https://doi.org/10.1371/journal.pone.0235860

Oscar Miranda-Dominguez, Eric Feczko, David S. Grayson, Hasse Walum, Joel T. Nigg, Damien A. Fair; Heritability of the human connectome: A connectotyping study. Network Neuroscience 2018; 02 (02): 175–199. doi: https://doi.org/10.1162/netn_a_00029