Connectome Fingerprinting with Twins & with Different Parcellations

Team members: Roger Xu, Emily Liao

Introduction

Finn et al. (2015) showed that subject identification by functional connectivity matrix similarity can be highly successful, especially using resting state scans (92.9% and 94.4%). Other studies support identifiability in additional contexts including different stages of life (St-Onge et al., 2015), in dynamic characteristics of FC (Liu et al., 2018), and in relation to cognitive and clinical characteristics (Kaufmann et al., 2017; Lopez et al., 2022). In siblings and twins specifically, fingerprinting analyses have found that related individuals were still distinguishable, but more closely correlated than with non-related individuals (Miranda-Dominguez et al., 2018); increased genetic similarity also correlated to better classification accuracy of siblings and twins (Demeter et al., 2020). Based on such literature, we hypothesized that we would see similar success in subject identification and also signs of sibling similarity (either misidentification or second-closest identification) in our project using monozygotic (MZ) twin data from the Human Connectome Project Young Adult dataset. A subset of data from this initiative was previously included in the Finn et al. and the Miranda-Dominguez et al. studies when data collection was still ongoing, but we extend their analyses to a larger group with monozygotic twins specifically and 7T ultra high resolution scans.

Higher feature similarity observed across monozygotic twins can strongly suggest underlying genetic influences (Jansen et al., 2015). While the genetic and environmental influences on the individuality observed in fMRI data is still an open area of study, selecting a twin subset as our main focus gives us a potential starting point for possible future comparison studies where subjects are not genetically identical. We also thought it might help us elucidate questions of preprocessing especially with regard to parcellation. One of the challenges of integrating genomic and neuroimaging datasets is that there are many approaches to identifying regions. By repeating our identification analysis with 3 parcellation images (268-node atlas from Shen and the 200 and 400 node atlases from Yan et al. 2023) using the same subjects, we sought to clarify whether interindividual differences and sibling similarities might be hidden or enhanced by one of these parcellation choices.

Methods

We accessed preprocessed, FIX-denoised HCP 7T data (1.6mm slice thickness) from 15 minute 7T fMRI resting scans collected by the Connectome Coordination Facility, selecting for a total of 100 subjects who had a monozygotic twin in the same dataset (50 pairs). We also accessed parcellation images for the Shen 268-region atlas used by Finn et al. and the 200- and 400-region atlases by Yan/Kong (Yan et al., 2023). Documentation and visualization in FSLeyes were used to confirm that both the HCP data and parcellation images were in MNI152 space. When the images and HCP data had different voxel sizes, MATLAB imresize3 with nearest neighbor interpolation was used to match the parcellation images to HCP data dimensions. Visualization in FSLeyes was used to confirm their alignments (Fig. 1, from milestone).

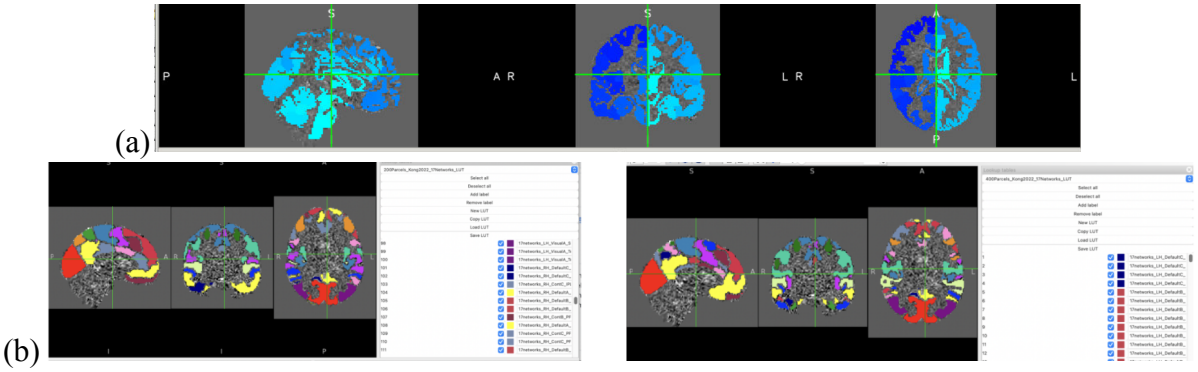

Fig. 1: Overlay of (a) Shen and (b) Yan parcellation on HCP subject 1 in 7T MZ twin subset.

After the space and sizing were matched, each parcellation image was applied to HCP data by looping through region numbers and averaging time series of voxels within each region. Functional connectivity matrices were then calculated by Pearson correlation between every pair of regions. In total, we have a matrix per resting state scan stored for each subject and parcellation scheme (4 scans x 100 subjects x 3 schemes). An example is shown in Fig. 2. NaN matrices were used for missing scans.

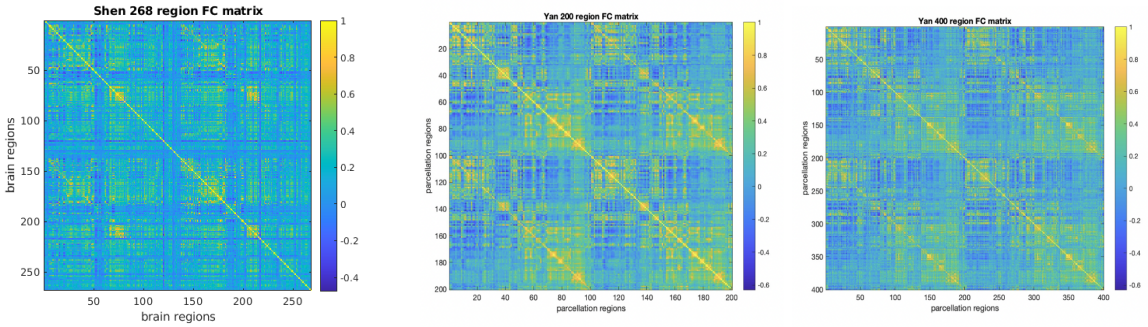

Fig. 2: Resting state scan 1 FC matrices (one per parcellation) for HCP subject 1 of our 7T MZ twin subset.*

We next flattened all matrices to vectors, and computed Pearson correlation across all possible pairs of target and database vectors from different scans. We have not currently controlled for whether the scans occurred on the same day (scans 1&2, 3&4) or with the same phase encoding procedure (scans 1&3 PA, scans 2&4 AP). Correlation outputs were sorted for each subject to identify their best match (expected to be self), as well as second and third closest matches. This information was used to create a table recording whether the subject was correctly identified, misidentified as their twin, had their twin as their second closest match if they were correctly identified, and shared their third closest match with their twin if both they and their twin were identified. For each of these categories, mean and standard deviations across all the scan combinations were calculated and presented in the Results section. For a basic test of significance, paired t-tests between parcellations were evaluated and differences were annotated on the plots using modified code from MATLAB toolbox sigstar (Campbell, 2018) with default

cutoffs of *p<=0.05, ** p<=0.01, *** p<=0.001. Since many of our comparisons were significant, other/stricter tests could be explored in future analyses.

A similar procedure was also repeated using a separate connectivity matrix for each of Yeo's 7 networks. Yan/Kong parcellation images had network labels available by lookup table. For the Shen parcellation, Yeo network labels were assigned in the same approach used by Finn et al. to evaluate the Freesurfer atlas: for each Shen node the number of voxels belonging to each Yeo network was counted. The network with the most voxels became the primary label, and a second label was allowed if the second highest exceeded 30% of the total voxels in the node. Nodes for which the target network was either the primary or secondary label were included when defining our connectivity matrix for a network.

Results

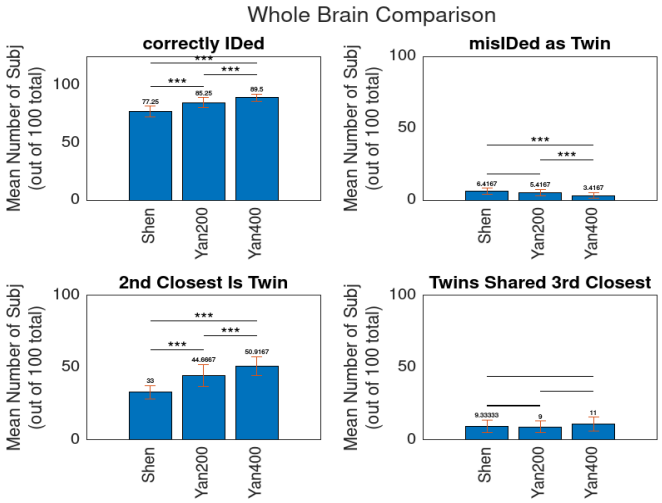

Fig. 3. Whole brain identification results across the 3 atlases. Significance thresholds were as follows for paired t-test comparisons between atlases: *p<=0.05, ** p<=0.01, *** p<=0.001.

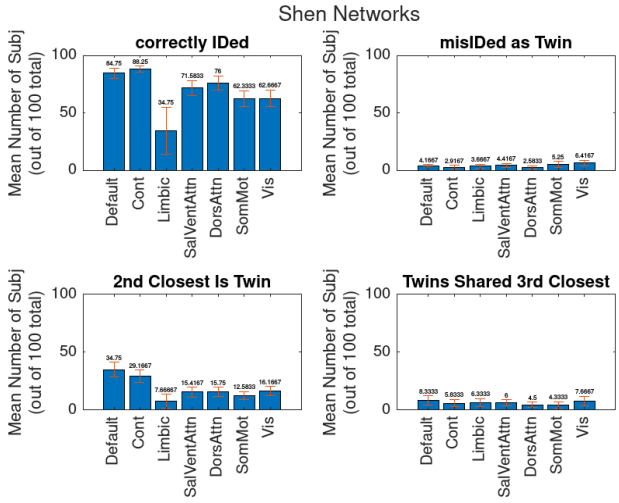

Fig. 4 continued on next page

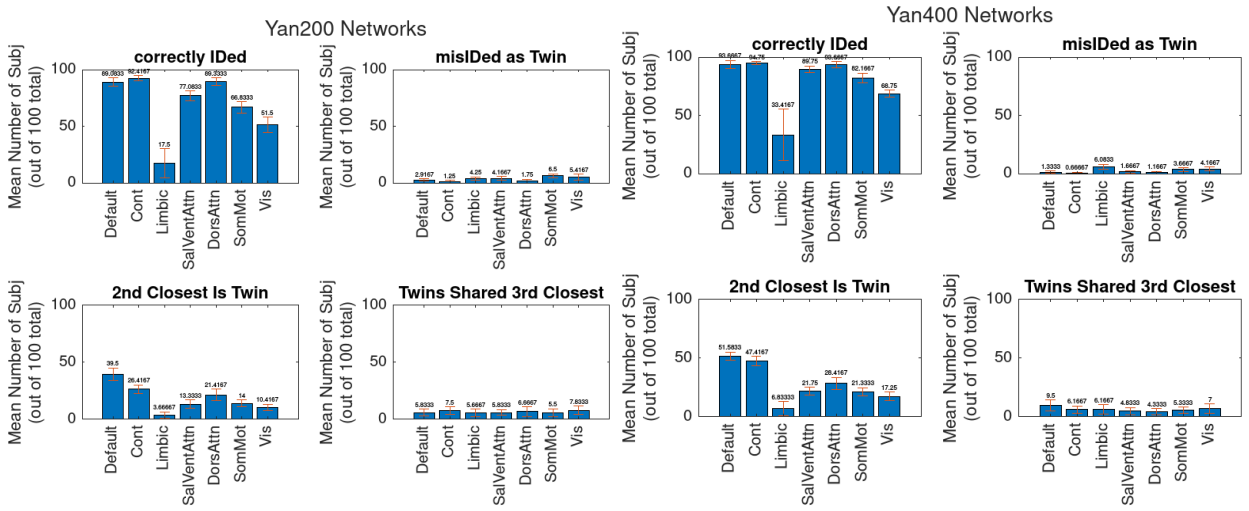

Fig. 4: For each of 7 networks in each atlas, the mean number of subjects (out of 100) across all database-target scan combos that were correctly identified, misidentified as their twin, had their twin as their second closest match if they were correctly identified, and shared their third closest match with their twin if both they and twin were identified.

Discussion and Conclusions

Our whole brain results support the background literature that indicate individual identification from resting state connectivity profiles is generally successful, with our observed rates ranging from 77 to 89% of subjects identified correctly on average. Although these rates are lower than those published by Finn et al. (92.9% and 94.4%), our identification task may have also been harder due to our selection of participants who had a monozygotic twin in the dataset. A small number of subjects get misidentified as their twin under all parcellations, although on average the Yan 400 region parcellation had the fewest instances of this and also the highest correct identification rate, followed by the Yan 200 parcellation. This trend in parcellation scheme performance was also observed in the success of identifying a subjects' twin based on finding the subject with the second closest correlation. While we did expect that the number of regions would play a role in the accuracy of identification (higher node numbers might help capture unique local activity), the effects of parcellation also do not appear to be entirely driven by node number alone because the Yan 200 atlas outperformed the 268-region Shen atlas in self and twin identification. In future work, we could confirm what we observe here by repeating the analyses with other parcellation images available from the Yan set (100, 300, 500, etc.). Also, many other parcellation schemes are available. Both the atlases we have used here were made based on functional analyses, with consideration towards network organization, but it would be interesting to compare with an atlas derived from multimodal or structural features. This type of comparative analyses to clarify which patterns are brought out by which pre-processing could be informative for future parcellation choices in our own research projects, especially in cases where biological variables are expected to play a key role (e.g. linking fMRI data with genetics).

Aside from investigating the effects of parcellation, we also conducted our network analysis to see if we might observe certain networks driving the identification success. Across all parcels, top performers included the default, control, and attention networks. There is some overlap between these areas with the Finn et al. frontoparietal network areas, and the same

networks are also highlighted in findings by Demeter et al. (2020) which showed that connections in the DMN and fronto-parietal (control) network are the most important features for the classification task of subject identification. These networks, especially the default and control, were also best at identifying subjects' twins. In contrast, the limbic network was a poor identifier across all parcellations. Low performance of limbic networks could potentially be partly biological, based on its association with more basic/survival functions expected to be common across individuals. At the same time, one potential confound we are still looking to explore is the effect of fMRI data quality, evaluated by tSNR. While we have begun this process, the 4D nature of fMRI data and our different parcellations do make choosing and implementing SNR evaluation a little more challenging than in some other signal processing domains. Thus, this part of the project is still in progress. We hope such analyses can help to later clarify the relationship between cognitive/basic functions and individual variation observed in resting brain scans.

Contributions

Both team members contributed significantly to this project. Our initial parcellation and functional connectivity matrix calculation functions were prepared by Emily and then reviewed together before being used to apply the Shen parcellation (Emily) and Yan/Kong parcellations (Roger). The assignment of network labels to the parcellation nodes of each atlas was first completed by Roger, then scaled to all subjects/scans by Emily. Both team members had practice with computing functional connectivity matrices, either for the whole brain or within networks. The code for the identification analyses and charts was provided by Emily. Roger and Emily worked on learning about SNR in fMRI analyses together (selecting tSNR based on resources including Welvaert & Rosseel, 2013). Roger has provided some initial code to prepare the tSNR calculations for the networks which will need work from both team members. Our current plan is to first find a mean time course for each network (average all region time courses belonging to the network at each point in time), then compute the mean signal over time and the mean standard deviation over time. We can seek advice to confirm these steps as well as possible future directions.

Code

https://drive.google.com/drive/folders/1lELGQWhSk8hFQ0S0k1Jylad-R3m-eeeS?usp=sharing

- *networks_fc_and_snr.mlx* (should be run first): calculate FC matrices and tSNR for each network. The outputs are saved and passed to functions in id_report.mlx
- *id_report.mlx*: calculate whole brain FC matrices (when toggle variable is not set to load pre-saved results). Also runs identification report given the whole brain and network fc matrices
- *sigstar.m*: code from MATLAB file exchange by Rob Campbell. Annotates identification rate bar charts with brackets/stars. Slightly modified for visualization on our bar plots.
- Files needed to run the code are found in ece8395_files, along with HCP 7T dataset scans located on Accre at /data1/datasets/hcp/<subjId>

References

Campbell, R. (2018). *Sigstar - File Exchange - MATLAB Central*. [Matlab code]. https://www.mathworks.com/matlabcentral/fileexchange/39696-raacampbell-sigstar

Connectome Coordination Facility. (n.d.). *Connectome—HCP 7T Imaging Protocol Overview*. https://www.humanconnectome.org/hcp-protocols-ya-7t-imaging

Demeter, D. V., Engelhardt, L. E., Mallett, R., Gordon, E. M., Nugiel, T., Harden, K. P., Tucker-Drob, E. M., Lewis-Peacock, J. A., & Church, J. A. (2020). Functional Connectivity Fingerprints at Rest Are Similar across Youths and Adults and Vary with Genetic Similarity. *iScience*, *23*(1), 100801. https://doi.org/10.1016/j.isci.2019.100801

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., & Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, *18*(11), Article 11. https://doi.org/10.1038/nn.4135

Jansen, A. G., Mous, S. E., White, T., Posthuma, D., & Polderman, T. J. C. (2015). What twin studies tell us about the heritability of brain development, morphology, and function: A review. *Neuropsychology Review*, *25*(1), 27–46. https://doi.org/10.1007/s11065-015-9278-9

Kaufmann, T., Alnæs, D., Doan, N. T., Brandt, C. L., Andreassen, O. A., & Westlye, L. T. (2017). Delayed stabilization and individualization in connectome development are related to psychiatric disorders. *Nature Neuroscience*, *20*(4), Article 4. https://doi.org/10.1038/nn.4511

Kong, R., Yang, Q., Gordon, E., Xue, A., Yan, X., Orban, C., Zuo, X.-N., Spreng, N., Ge, T., Holmes, A., Eickhoff, S., & Yeo, B. T. T. (2021). Individual-Specific Areal-Level Parcellations Improve Functional Connectivity Prediction of Behavior. *Cerebral Cortex*, *31*(10), 4477–4500. https://doi.org/10.1093/cercor/bhab101

Liu, J., Liao, X., Xia, M., & He, Y. (2018). Chronnectome fingerprinting: Identifying individuals and predicting higher cognitive functions using dynamic brain connectivity patterns. *Human Brain Mapping*, *39*(2), 902–915. https://doi.org/10.1002/hbm.23890

Miranda-Dominguez, O., Feczko, E., Grayson, D. S., Walum, H., Nigg, J. T., & Fair, D. A. (2018). Heritability of the human connectome: A connectotyping study. *Network Neuroscience*, *2*(2), 175–199. Scopus. https://doi.org/10.1162/netn_a_00029

St-Onge, F., Javanray, M., Pichet Binette, A., Strikwerda-Brown, C., Remz, J., Spreng, R. N., Shafiei, G., Misic, B., Vachon-Presseau, É., & Villeneuve, S. (2023). Functional connectome fingerprinting across the lifespan. *Network Neuroscience*, *7*(3), 1206–1227. https://doi.org/10.1162/netn_a_00320

Lopez, E., Minino, R., Liparoti, M., Polverino, A., Romano, A., De Micco, R., Lucidi, F., Tessitore, A., Amico, E., Sorrentino, G., Jirsa, V., & Sorrentino, P. (2023). Fading of brain network fingerprint in Parkinson's disease predicts motor clinical impairment. *Human Brain Mapping*, *44*(3), 1239–1250. https://doi.org/10.1002/hbm.26156

Welvaert M, Rosseel Y. On the definition of signal-to-noise ratio and contrast-to-noise ratio for FMRI data. *PLoS One*. 2013;8(11):e77089. Published 2013 Nov 6. doi:10.1371/journal.pone.0077089

Yan, X., Kong, R., Xue, A., Yang, Q., Orban, C., An, L., Holmes, A. J., Qian, X., Chen, J., Zuo, X.-N., Zhou, J. H., Fortier, M. V., Tan, A. P., Gluckman, P., Chong, Y. S., Meaney, M. J., Bzdok, D., Eickhoff, S. B., & Yeo, B. T. T. (2023). Homotopic local-global parcellation of the human cerebral cortex from resting-state functional connectivity. NeuroImage, 273, 120010. https://doi.org/10.1016/j.neuroimage.2023.120010