Putting Pipeline Implementation-related Variation Into Perspective For Functional Connectomics

Xinhui Li[1], Lei Ai[1], Steve Giavasis[1], Hecheng Jin[1], Jon Clucas[1], Alexandre Franco[2,1], Eric Feczko[3], Joshua Vogelstein[4],
Cameron Craddock[5], Ting Xu[1], Oscar Esteban[6], Russell Poldrack[7], Damien Fair[3], Theodore Satterthwaite[8], Michael Milham[1,2]

[1]Child Mind Institute, New York, NY, [2]Center for Biomedical Imaging and Neuromodulation, Nathan Kline Institute, Orangeburg, NY, [3]University of Minnesota, Minneapolis, MN, [4]Johns Hopkins University, Baltimore, MD,
[5]The University of Texas at Austin, Austin, TX, [6]University Hospital of Lausanne and University of Lausanne, Lausanne, Switzerland, [7]Stanford University, San Francisco, CA, [8]University of Pennsylvania, Philadelphia, PA

Introduction

A basic set of preprocessing steps have emerged in the functional connectomics literature and are now widely implemented across pipeline packages. However, marked variation exists in how specific algorithms implement these steps, as well as how pipelines link them together. Alarmingly, recent studies have found that when analyzing the same data, differences in implementation are a substantial and undesirable source of variation (e.g., Bowring et al., 2019).

Here, inspired by the growing consensus that increasing the fMRI data quantity per participant can overcome suboptimal test-retest reliability, we examine the impact of data quantity on cross-pipeline reliability. To accomplish this goal, we leverage the Configurable Pipeline for the Analysis of Connectomes (CPAC) to enable generation of harmonized configurations for each of three minimal processing pipelines in a common framework, which can then be used to analyze the origins of differences among pipelines.

Methods

Dataset. The HNU Sample (N=30) from the Consortium for Reliability and Reproducibility (CoRR), including ten 10-min single-band resting state fMRI per subject (TR=2000ms).

Pipelines. Our cross-pipeline comparisons focused on the impact of differing minimal preprocessing pipelines on functional connectivity estimates. To accomplish this, we extended the range of configurable options in CPAC pipeline to include those necessary to replicate the key minimal preprocessing steps in each of the 3 established preprocessing pipelines (fMRIPrep, the ABCD Pipeline, and the Connectome Computation System) (see Table 1 for comparison of pipeline steps used to guide harmonization). Prior to cross-pipeline comparisons, we first validated each CPAC harmonized configuration with the package it was based on (target cross-pipeline reliability: > 0.9).

Evaluation Metrics. Reproducibility was quantified using a multifaceted assessment strategy, including: 1) matrix-wise Pearson spatial correlation, 2) edge-wise intraclass correlation coefficient (ICC), 3) image-wise intraclass correlation coefficient (I2C2), and 4) discriminability (matrix-level index of reliability).

Table 1. Summarizing minimal processing pipeline differences

Comparing Minimal Processing Pipelines
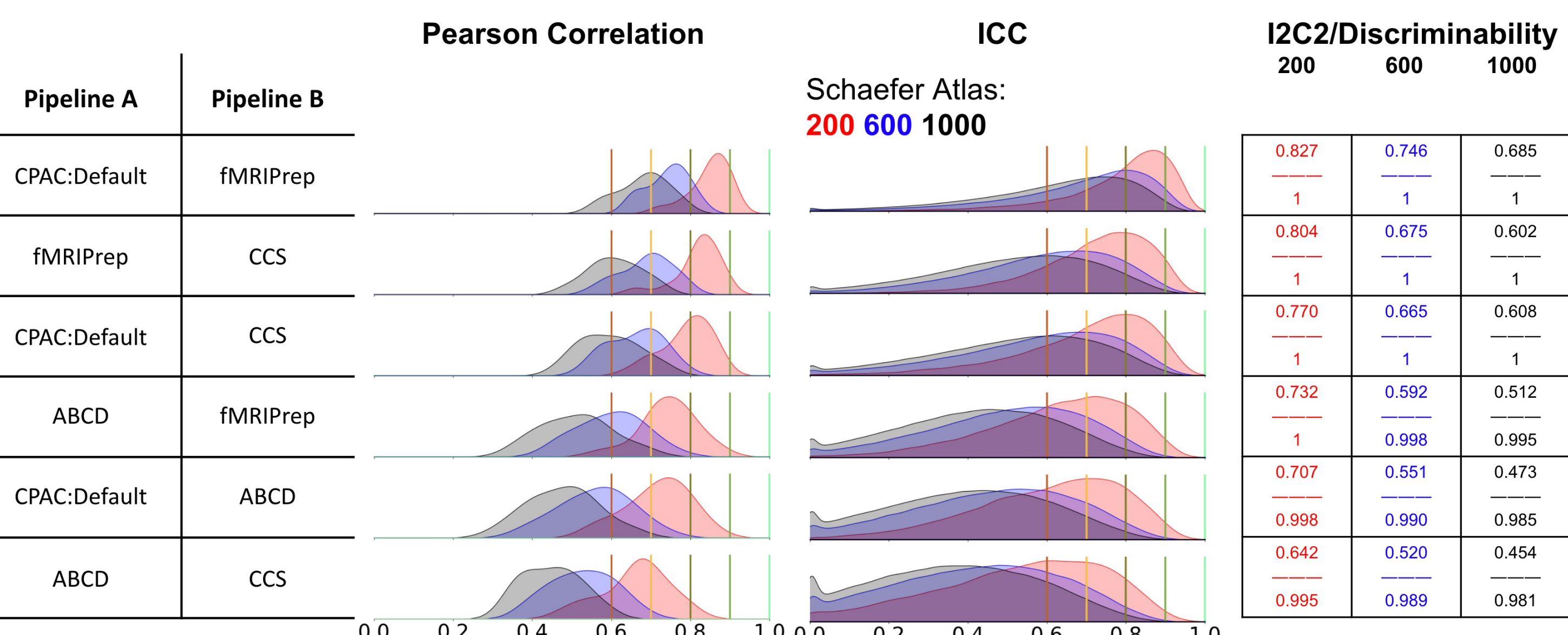

Figure 1. Comparing minimal preprocessing across a sample of fMRI pipeline packages (CPAC:Default, fMRIPrep, ABCD and CCS).

Evaluating Harmonized Pipelines

Figure 2. Evaluating cross-pipeline reliability (matrix-wise, edge-wise) for functional connectivity matrices generated using CPAC harmonized pipelines against the originals (fMRIPrep, ABCD, CCS).
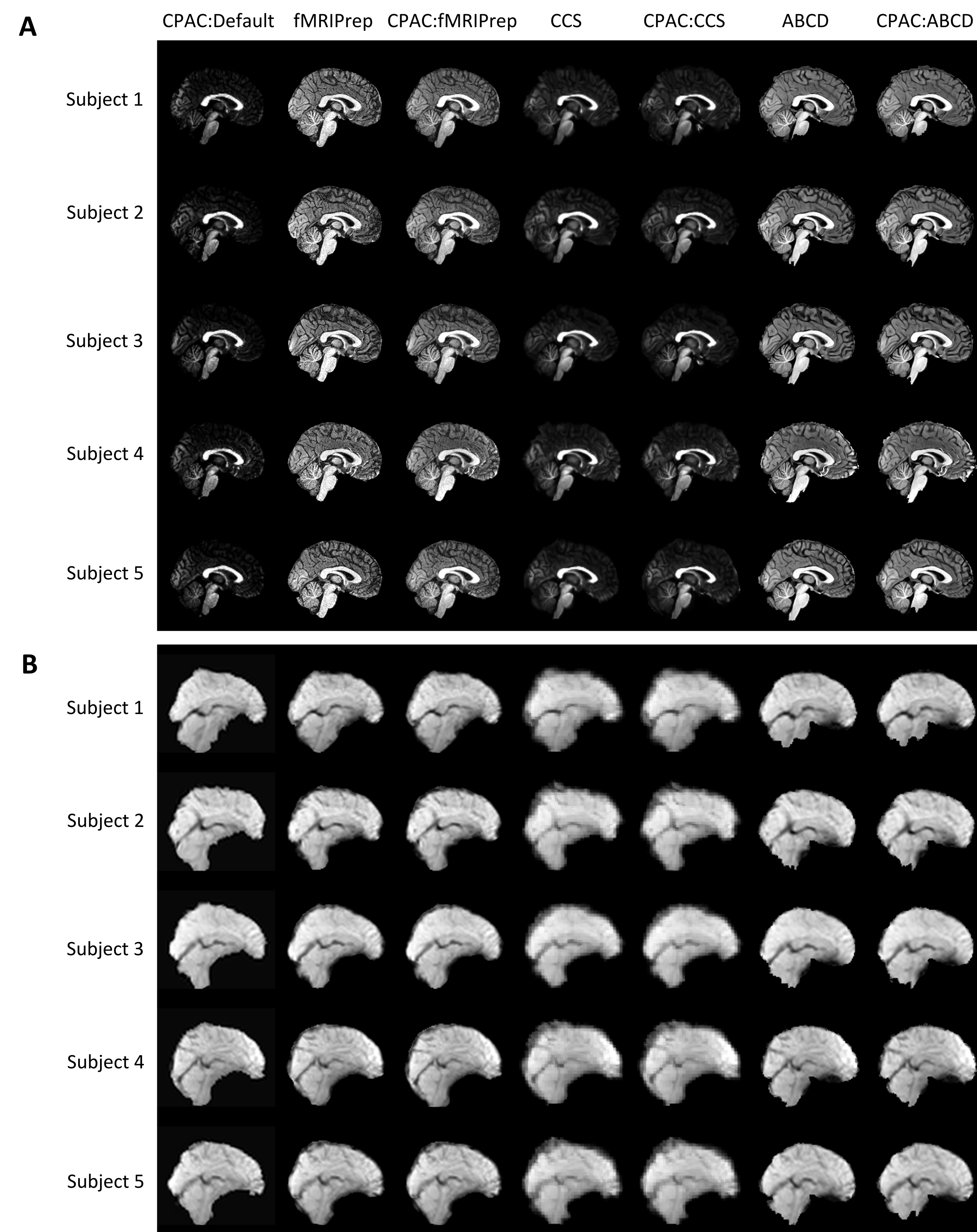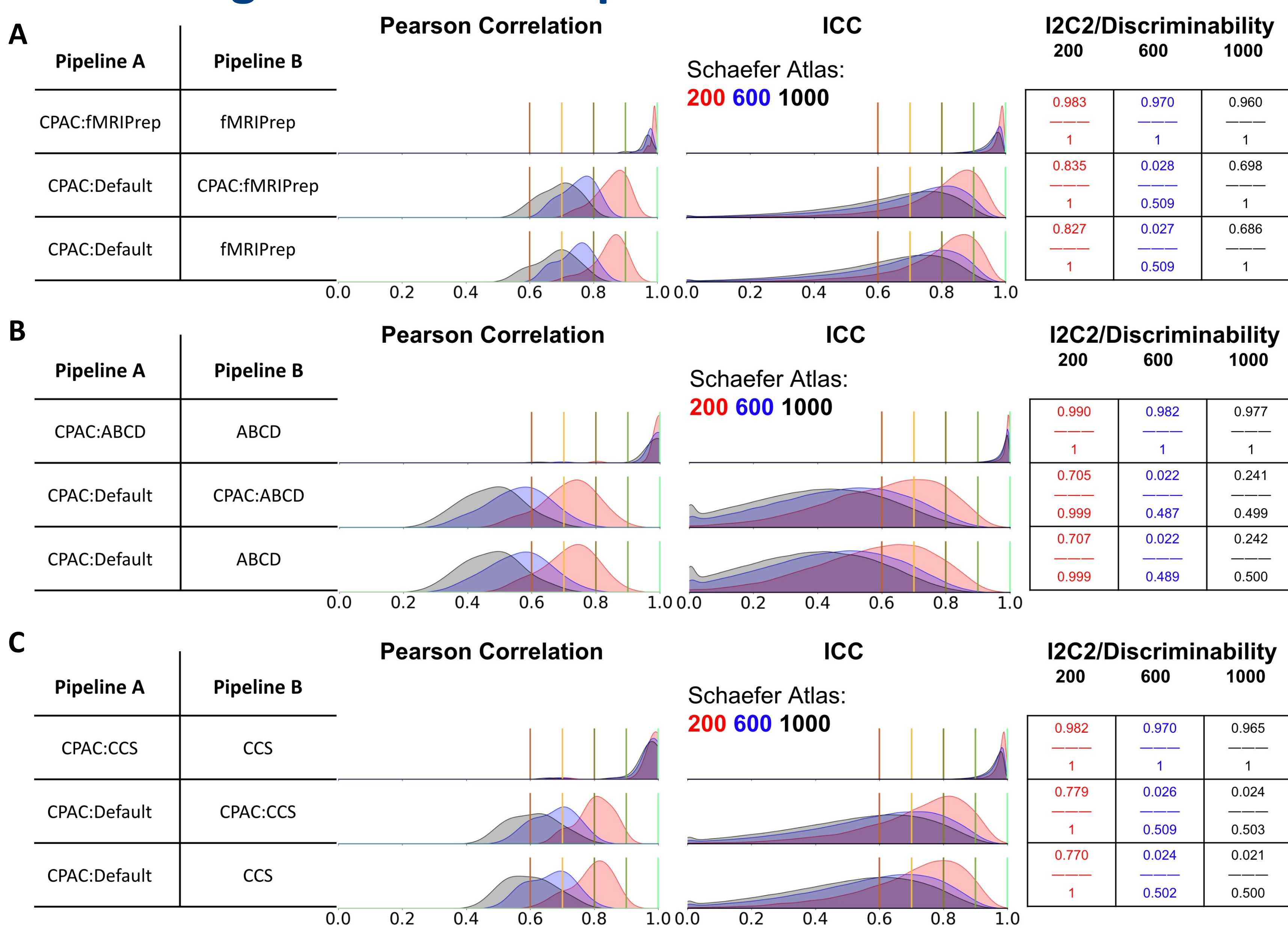
Visualizing Minimal Processing Outputs

Figure 3. Visualizing harmonized minimal preprocessing outputs.
A. Anatomical images in template space. B. Functional images in template space.

Evaluating Scan Duration Impact

CROSS-PIPELINE RELIABILITY FOR FUNCTIONAL CONNECTIVITY MATRICES AS FUNCTION OF DATA SOURCE AND QUANTITY
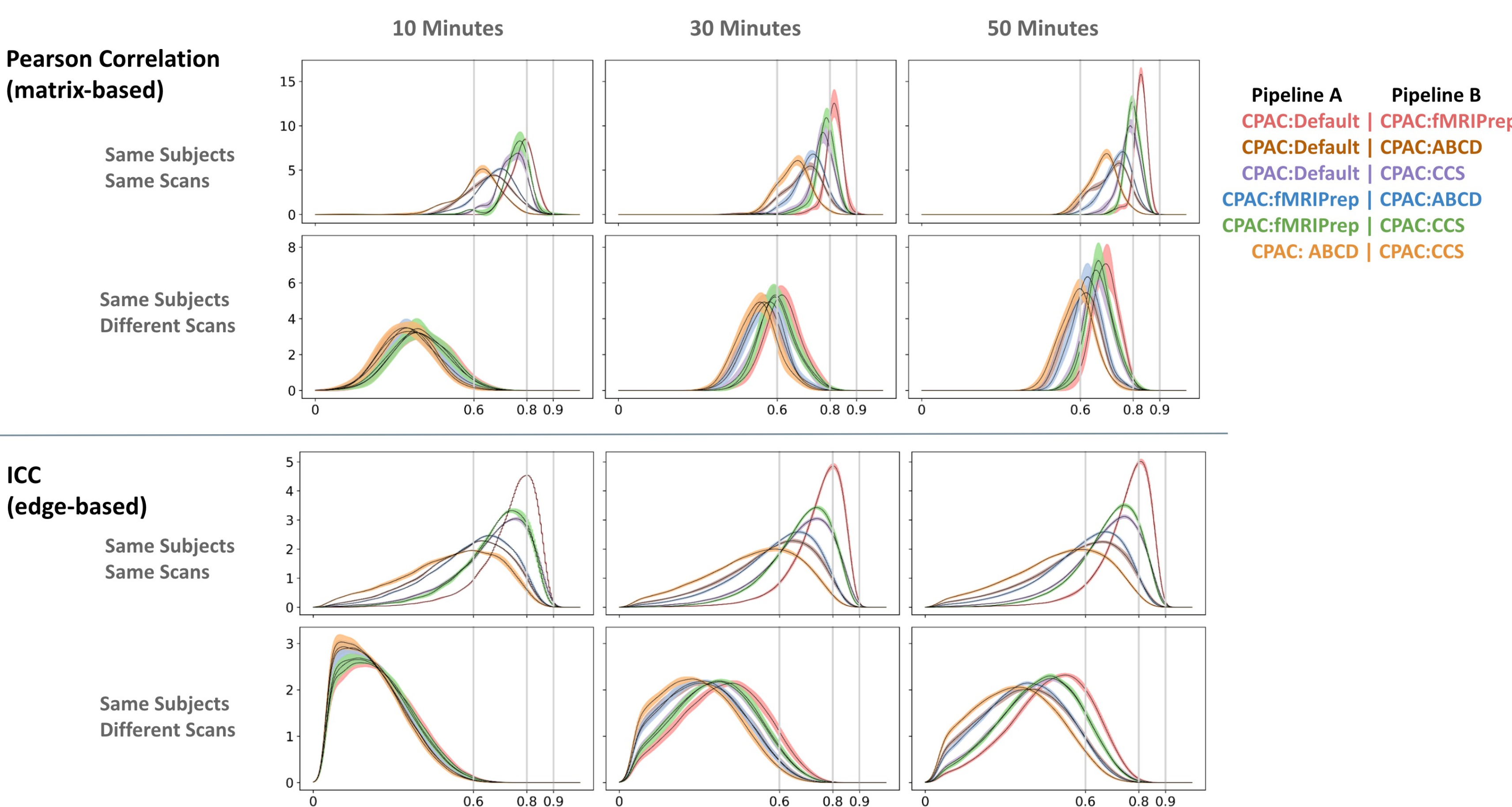

Figure 4. Evaluating the scan duration impact. We varied the scan duration (10/30/50 min) by concatenating the data from different sessions and measured the cross-pipeline reliability. Results using a 200 parcellation unit atlas are reported here (findings are consistent at 600 and 1000).

Conclusion

It is common and justifiable to compare the impact of different processing steps on the same datasets from the same subjects. The present work clearly demonstrates that such comparisons can give skewed perspectives. As robust as our findings were regarding alarmingly suboptimal cross-pipeline reliability were when applied to the same data, from the same subjects, they were dwarfed by session-related variation when shorter scan durations (e.g., 10 minutes) were included. Our findings emphasize the need for both – more dense sampling per subject, and an increased focus on harmonizing pipeline implementation - to improve reproducibility across studies.

References

Bowring A, Maumet C, Nichols TE. Exploring the impact of analysis software on task fMRI results. Hum Brain Mapp. 2019;40(11):3362-3384. doi:10.1002/hbm.24603

Craddock C, Sikka S, Cheung B, et al. Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac). Front Neuroinform. 2013;42:10-3389. https://www.frontiersin.org/10.3389/conf.fninf.2013.09.00042/event_abstract

Esteban O, Markiewicz CJ, Blair RW, et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat Methods. 2019;16(1):111-116. doi:10.1038/s41592-018-0235-4

Satterthwaite T, et al. XCP Imaging Pipeline — xcpEngine 1.0 documentation. Accessed February 1, 2021. https://xcpengine.readthedocs.io/

Sturgeon, D., Perrone, A., Earl, E., & Snider, K. DCAN_Labs/abcd-hcp-pipeline. doi: 10.5281/zenodo.2587209

Xu T, Yang Z, Jiang L, Xing XX, Zuo XN. A Connectome Computation System for discovery science of brain. Sci Bull. 2015; 60(1):86-95.

Shou H, Eloyan A, Lee S, et al. Quantifying the reliability of image replication studies: the image intraclass correlation coefficient (I2C2). Cogn Affect Behav Neurosci. 2013;13(4):714-724. doi:10.3758/s13415-013-0196-0