Considering factors affecting the connectome-based identification process: Comment on Waller et al.

Corey Horien [a,*], Stephanie Noble [a], Emily S. Finn [a,d], Xilin Shen [b], Dustin Scheinost [b,c], R. Todd Constable [a,b,e]

[a] Interdepartmental Neuroscience Program, Yale University School of Medicine, New Haven, CT, USA
[b] Department of Radiology and Biomedical Imaging, Yale University School of Medicine, New Haven, CT, USA
[c] The Child Study Center, Yale University School of Medicine, New Haven, CT, USA
[d] Section on Functional Imaging Methods, National Institute of Mental Health, Bethesda, MD, USA
[e] Department of Neurosurgery, Yale University School of Medicine, New Haven, CT, USA

ABSTRACT

A recent study by Waller and colleagues evaluated the reliability, specificity, and generalizability of using functional connectivity data to identify individuals from a group. The authors note they were able to replicate identification rates in a larger version of the original Human Connectome Project (HCP) dataset. However, they also report lower identification accuracies when using historical neuroimaging acquisitions with low spatial and temporal resolution. The authors suggest that their results indicate connectomes derived from historical imaging data may be similar across individuals, to the extent that this connectome-based approach may be inappropriate for precision psychiatry and the goal of drawing inferences based on subject-level data. Here we note that the authors did not take into account factors affecting data quality and hence identification rates, independent of whether a low spatiotemporal resolution acquisition or a high spatiotemporal resolution acquisition is used. Specifically, we show here that the amount of data collected per subject and in-scanner motion are the predominant factors influencing identification rates, not the spatiotemporal resolution of the acquisition. To do this, we investigated identification rates in the HCP dataset as a function of the amount of data and motion. Using a dataset from the Consortium for Reliability and Reproducibility (CoRR), we investigated the impact of multiband versus non-multiband imaging parameters; that is, high spatiotemporal resolution versus low spatiotemporal resolution acquisitions. We show scan length and motion affect identification, whereas the imaging protocol does not affect these rates. Our results suggest that motion and amount of data per subject are the primary factors impacting individual connectivity profiles, but that within these constraints, individual differences in the connectome are readily observable.

Introduction

A key goal of precision psychiatry is leveraging individual differences in neuroimaging data to generate predictive models related to behavior. As highlighted by Waller et al. (2017), finding reliable markers across datasets remains an important part of this process. As such, they investigate the generalizability of a previous method using functional connectivity fMRI data to identify individuals from a group ('connectome fingerprinting'; Finn et al., 2015). Waller et al. demonstrate that identification can be replicated in the same high spatiotemporal resolution dataset (i.e. acquired using multiband acquisition sequences), consistent with other work to replicate the method (Finn et al., 2017; Kaufmann et al., 2017; Vanderwal et al., 2017), though they note lower accuracies using a dataset acquired with lower spatiotemporal resolution (i.e. acquired using non-multiband acquisition sequences). In addition, the authors also show that the specificity of the identification procedure is lower when a within-subject correlation threshold is introduced into the

ID pipeline. Therefore, the authors argue that the identification method may not generalize to datasets with lower spatiotemporal resolution because individual features may only be detectable in data acquired with high spatiotemporal resolution. However, in their study, the authors did not take into account other factors affecting data quality and hence the identification process, namely scan duration and subject motion. Here we evaluate the impact of not only spatiotemporal resolution during image acquisition, but also other data quality factors on identification rates.

Methods

The HCP 900 subjects release (Van Essen et al., 2013) was used to investigate scan time and motion. Data were pre-processed and connectivity matrices were calculated as described elsewhere (Finn et al., 2015, 2017; Shen et al., 2017). All analyses were performed using the left-to-right (LR) phase encoding rest runs from days one and two. Of note, HCP TR = 720 ms. To study motion, subjects were separated into

low and high motion groups using a mean frame-frame displacement threshold of 0.1 mm averaged over the two sessions. Of the 819 subjects available with all data and day one and two LR rest scans, 603 were in the low motion group and 216 subjects were in the high motion group. To study the effect of scan time, we truncated time courses to correspond to the number of frames in 1, 2, …,14 min and calculated connectivity matrices from the shortened data. Because of the difference in sample sizes among the low and high motion groups, we repeatedly subsampled 216 subjects in the low motion group and performed identification 1000 times. The mean ID rate and 95% confidence intervals were therefore calculated from the subsampled data (Fig. 1A). To investigate when ID rates plateaued, we used the Levenberg-Marquardt nonlinear least squares algorithm to fit the following nonlinear regression model function: $IDrate = maxIDrate\left(1 - e^{-\frac{t}{x}}\right)$, where t = time, $IDrate$ = ID rate at time t, $maxIDrate$ = maximum ID rate determined by the model, and x = time required for the ID rate to reach approximately 63% of its maximum value. We defined plateauing of the ID rate to be the time points when the rate was 95 and 99% of the maximum ID rate.

In a separate analysis (Fig. 1B), we subsampled data (after low-pass filtering; approximate cutoff frequency of 0.12 Hz) from each of the 603 low motions subjects to simulate the effects of lower sampling frequencies (longer TR) versus total amount of scan time. For this analysis, we selected n frames from the duration of a subject's time course such that sampling every other frame produced 600 frames of the original 1200; sampling every 3rd frame resulted in 400 frames remaining, etc. Hence, these subsampled data still spanned the same overall acquisition time window. It should also be noted that this subsampled data has lower signal to noise ratio (SNR) than real data acquired at a longer TR because

of the additional T1 recovery that would occur with a longer TR. Connectivity matrices were subsequently calculated from the subsampled data. In addition, we performed a follow-up analysis using a similar strategy except that instead of removing every nth frame we averaged data from every n adjacent frames to again simulate a slower sampling frequency (Fig. 1B) and boost the SNR.

To study the effect of spatiotemporal resolution (Fig. 1C), we utilized a publically available test-retest dataset from the Nathan Kline Institute (NKI; http://fcon_1000.projects.nitrc.org/indi/CoRR/html/nki_1.html). This dataset contains individuals scanned with both multiband and non-multiband acquisition sequences, thus allowing us to investigate the impact of different pulse sequences on ID rates. Data acquisition parameters have been described previously (Liao et al., 2013). Briefly, three resting-state fMRI sequences were obtained for each of the 24 subjects: 1) multiband scan with TR = 645 ms; 2) multiband scan with TR = 1400 ms; and 3) and a non-multiband echo planar imaging (EPI) scan with TR = 2500 ms. One subject was excluded due to brain atrophy (subject 0021001); one subject was excluded due to excessive head motion (3795193; greater than 3° rotation); and we were unable to locate session 2 data for subject 6471972, leaving 21 subjects in the final analysis. We did not apply a further motion cutoff with these subjects due to the small sample size. The preprocessing steps have been previously described (Noble et al., 2017), except we performed skull-stripping using optiBET (Lutkenhoff et al., 2014). Though we did not perform slice-time correction on the multi-band data, we performed analyses on the TR = 2500 subjects with and without slice-time correction.

The identification procedure was carried out as described previously using Matlab code released by Finn et al. (2015) and utilized by Waller
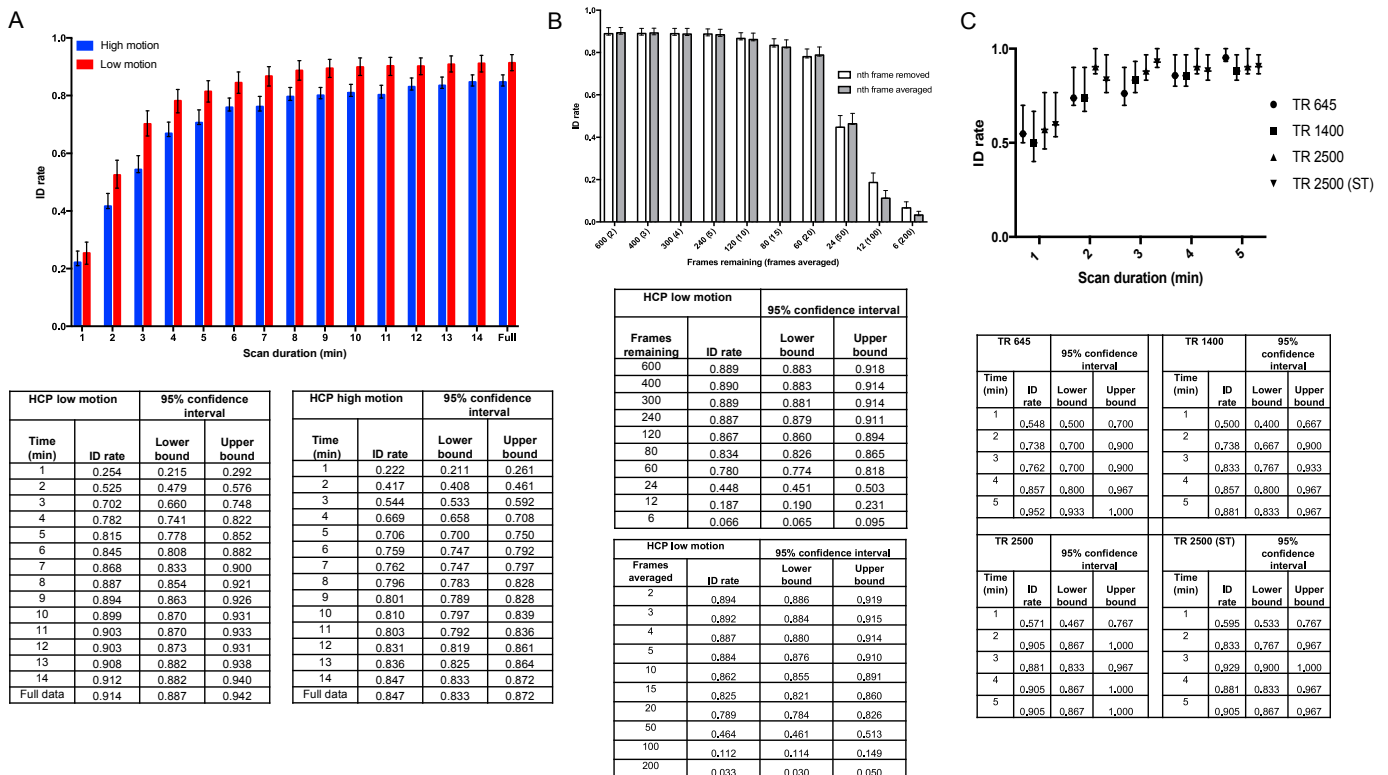

Fig. 1. The effect of scan duration, motion, and differences in spatiotemporal resolution on identification rates. (A) Top: Separating the HCP 900 subjects into groups based on motion and performing identification with increasing amounts of data. Identification rate for each group is indicated at each scan duration time. Note that the high and low motion groups have the same scan durations at a given time point on the x-axis. Both groups have equal sample sizes (n = 216). Red and blue bars represent the low and high motion groups, respectively. (B) Top: Simulating the effects of a lower TR in the HCP. Data from all 603 low motion subjects were subsampled (white bars); x-axis indicates the number of frames remaining. In a separate analysis every n adjacent frames were averaged (grey bars); x-axis indicates in parentheses the number of adjacent frames used to average. (C) Top: Identification rates achieved using multiband or non-multiband imaging parameters to assess the effect of spatiotemporal resolution. Multiband imaging was performed on groups labelled as TR 645 and TR 1400; TR 2500 was acquired via non-multiband imaging; TR 2500 (ST) indicates these subjects underwent slice-time correction. Identification rate achieved for each scanning protocol is indicated at each scan duration time. Error bars correspond to 95% confidence intervals. Note that in (A), (B), and (C), the lower part of each panel includes the actual ID rate obtained and the 95% confidence intervals.

and colleagues (https://www.nitrc.org/frs/?group_id=51). To directly compare our results to the main findings of Waller et al. we used the same subset of nodes from the frontoparietal and medial frontal networks to perform identification. To generate 95% confidence intervals, we calculated bootstrapped identification accuracies by subsampling approximately 70% of the subjects in each condition tested.

Results

Using only the edges derived from the frontoparietal and medial frontal networks, we found identification was affected by both motion and total scan time in the HCP 900 dataset. Given the differences in sample sizes of the datasets used in this study and that of the authors (BLP 85), ID rates should not be directly compared to those obtained by Waller et al. Rather our focus is on the relative importance of motion, scan time, and acquisition sequence.

We observed that identification rates were consistently higher for subjects in the low motion group compared to the high motion group ($P < .05$; Fig. 1A) with the only exception being when 1 min of data was used. Increasing scan duration, or the amount of data per subject, resulted in increasing rates of identification for both groups. Using nonlinear regression, we determined that ID rates reached 95% and 99% percent of maximum at approximately 6.85 and 10.3 min, respectively, for the low motion group, and at 8.7 and 14.4 min, respectively, for the high motion group, supporting the notion that increasing scan durations past 3.7 and 5.2 min (the scan times of 2/3 scans used in Waller et al.) results in higher ID rates for both groups, and that for high motion subjects, more data is needed to achieve a successful identification.

To further investigate the importance of total scan time, we subsampled the HCP 900 data from the 603 low motion subjects, in one case by taking every other volume, and in the other case by averaging adjacent frames, to simulate a lower TR. We found identification rates were relatively stable after both procedures (Fig. 1B) and did not begin to decrease until removing every 20th frame (60 frames remaining; $P < .05$) and averaging over every 50 frames ($P < .05$).

We next used the NKI dataset to evaluate the effect of spatiotemporal resolution on identification rates. Using the same framework as above, we found that identification rates were effectively the same across imaging conditions at a given time point ($P > .05$; Fig. 1C). Similar to the HCP results above, we observed that increasing the scan duration increased identification rates in the NKI dataset. Interestingly, it was scan duration, and not number of samples, that had the largest effect on identification rates. In other words, a shorter TR cannot compensate for a shorter acquisition time: in general, it is better to have fewer samples distributed across a longer temporal window than more samples acquired in quick succession (echoing results in Laumann et al., 2015; Airan et al., 2016; Noble et al., 2017).

General comments

The study by Waller et al. (2017) provides a valuable contribution to the goal of precision psychiatry by replicating the original connectome fingerprinting work, and we applaud their efforts to expand the method to other datasets and generalize the findings to larger groups of subjects. However, the authors point out that accuracy and specificity are expected to drop with lower quality datasets (in the authors' words, datasets of "standard quality," as opposed to "high quality" datasets like the HCP). While we agree that identification rates are expected to drop as sample size increases, higher identification accuracies can still be obtained by improving factors other than the particular EPI acquisition used. Here, we have demonstrated that scan duration and subject movement both affect identification rates, but were unable to find an effect of the particular EPI acquisition strategy. Given the known effect of motion on estimates of functional connectivity (Power et al., 2015; Satterthwaite et al., 2012), it is reasonable that high motion subjects might be harder to identify from scan to scan. Further, the findings with respect to EPI

protocol (i.e. spatiotemporal resolution) are generally consistent with the work of Airan et al. (2016), in which they applied a non-parametric measure to study the differentiation of subjects in the NKI dataset and observed no clear relationship regarding multiband versus non-multiband data. Taking all of these factors into account, we suggest that if Waller et al. (2017) had applied a similar motion threshold to the BLP 85 dataset and if they had longer scan durations per subject, they would have higher identification rates—even with their standard quality fMRI data.

In a more general sense, our results reinforce the importance of the amount of data collected per subject in detecting between-subject differences. In the original connectome fingerprinting article it was shown that identification rates increase with increasing data (Finn et al., 2015); follow-up work similarly demonstrated identification is affected by time (Finn et al., 2017). Other studies have also shown that longer acquisition times and more data are associated with increases in the test-retest reliability of functional connectivity measures (Birn et al., 2013; Mueller et al., 2015; Shah et al., 2016; Noble et al., 2017; Laumann et al., 2015) as well as estimates of individual differentiation (Airan et al., 2016).

We note that the authors showed that the specificity of the ID method could be low even when high accuracies are obtained. While the specificity findings are noteworthy and we appreciate their work, it is not clear to what extent specificity in the context of identification-based studies is necessary. Identifiability of the functional connectome by itself is interesting, yet identification provides no information related to behavioral/cognitive measures, disease course, etc. The goal is the development of connectome-based predictive models that the unique patterns of connectivity provide (Shen et al., 2017). Identification per se is not the primary objective, and thus identifying a patient with high confidence from scan to scan does not inform patient response to treatment or disease prognosis.

With this in mind, we reiterate an important aspect of our original work: capitalizing on individual variability in functional connectomes and using this to build meaningful models predicting some cognitive feature as opposed to only identification of participants. Prediction of individual features is a valuable objective for neuroimaging (Gabrieli et al., 2015) and our previous work (Finn et al., 2015; Rosenberg et al., 2017) supports the notion that it is possible to generate predictive models of behavior from neuroimaging data. While we agree with Waller et al. that further developments are needed to yield clinically available biomarkers based on individual connectomes, there are several emerging promising results suggesting the potential clinical utility (Drysdale et al., 2017) of this approach. In this development, attention to the impact of factors such as motion, scan duration, and acquisition parameters is a key part of the process. We maintain that establishing the link between individual connectomes and behavior is an important goal for precision psychiatry. We thank Waller et al. for extending the work of our identification method, and we look forward to further developments in using individual connectome data in precision psychiatry.

References

Airan, R.D., Vogelstein, J.T., Pillai, J.J., Caffo, B., Pekar, J.J., Sair, H.I., 2016. Factors affecting characterization and localization of interindividual differences in functional connectivity using MRI. Hum. Brain Mapp. 37, 1986–1997.

Birn, R.M., Molloy, E.K., Patriat, R., Parker, T., Meier, T.B., Kirk, G.R., Nair, V.A., Meyerand, M.E., Prabhakaran, V., 2013. The effect of scan length on the reliability of resting-state fMRI connectivity estimates. Neuroimage 83, 550–558.

Drysdale, A.T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., Fetcho, R.N., Zebley, B., Oathes, D.J., Etkin, A., Schatzberg, A.F., Sudheimer, K., Keller, J., Mayberg, H.S., Gunning, F.M., Alexopoulos, G.S., Fox, M.D., Pascual-Leone, A., Voss, H.U., Casey, B.J., Dubin, M.J., Liston, C., 2017. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat. Med. 23, 28–38.

Finn, E.S., Scheinost, D., Finn, D.M., Shen, X., Papademetris, X., Constable, R.T., 2017. Can brain state be manipulated to emphasize individual differences in functional connectivity? Neuroimage 160, 140–151.

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. Nat. Neurosci. 18, 1664–1671.

Gabrieli, J.D., Ghosh, S.S., Whitfield-Gabrieli, S., 2015. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. Neuron 85, 11–26.

Kaufmann, T., Alnaes, 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. Nat. Neurosci. 20, 513–515.

Laumann, T.O., Gordon, E.M., Adeyemo, B., Snyder, A.Z., Joo, S.J., Chen, M.Y., Gilmore, A.W., McDermott, K.B., Nelson, S.M., Dosenbach, N.U., Schlaggar, B.L., Mumford, J.A., Poldrack, R.A., Petersen, S.E., 2015. Functional system and areal organization of a highly sampled individual human brain. Neuron 87, 657–670.

Liao, X.H., Xia, M.R., Xu, T., Dai, Z.J., Cao, X.Y., Niu, H.J., Zuo, X.N., Zang, Y.F., He, Y., 2013. Functional brain hubs and their test-retest reliability: a multiband resting-state functional MRI study. Neuroimage 83, 969–982.

Lutkenhoff, E.S., Rosenberg, M., Chiang, J., Zhang, K., Pickard, J.D., Owen, A.M., Monti, M.M., 2014. Optimized brain extraction for pathological brains (optiBET). PLoS One 9, e115551.

Mueller, S., Wang, D., Fox, M.D., Pan, R., Lu, J., Li, K., Sun, W., Buckner, R.L., Liu, H., 2015. Reliability correction for functional connectivity: theory and implementation. Hum. Brain Mapp. 36, 4664–4680.

Noble, S., Spann, M.N., Tokoglu, F., Shen, X., Constable, R.T., Scheinost, D., 2017. Influences on the test-retest reliability of functional connectivity MRI and its relationship with behavioral utility. Cerebr. Cortex 1–15.

Power, J.D., Schlaggar, B.L., Petersen, S.E., 2015. Recent progress and outstanding issues in motion correction in resting state fMRI. Neuroimage 105, 536–551.

Rosenberg, M.D., Finn, E.S., Scheinost, D., Constable, R.T., Chun, M.M., 2017. Characterizing attention with predictive network models. Trends Cognit. Sci. 21, 290–302.

Satterthwaite, T.D., Wolf, D.H., Loughead, J., Ruparel, K., Elliott, M.A., Hakonarson, H., Gur, R.C., Gur, R.E., 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. Neuroimage 60, 623–632.

Shah, L.M., Cramer, J.A., Ferguson, M.A., Birn, R.M., Anderson, J.S., 2016. Reliability and reproducibility of individual differences in functional connectivity acquired during task and resting state. Bird. Behav. 6, e00456.

Shen, X., Finn, E.S., Scheinost, D., Rosenberg, M.D., Chun, M.M., Papademetris, X., Constable, R.T., 2017. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. Nat. Protoc. 12, 506–518.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.U.-M.H., 2013. The WU-Minn human connectome project: an overview. Neuroimage 80, 62–79.

Vanderwal, T., Eilbott, J., Finn, E.S., Craddock, R.C., Turnbull, A., Castellanos, F.X., 2017. Individual differences in functional connectivity during naturalistic viewing conditions. Neuroimage 157, 521–530.

Waller, L., Walter, H., Kruschwitz, J.D., Reuter, L., Muller, S., Erk, S., Veer, I.M., 2017. Evaluating the replicability, specificity, and generalizability of connectome fingerprints. Neuroimage 158, 371–377.