

Determining four confounding factors in individual cognitive traits prediction with functional connectivity: an exploratory study

Pujie Feng^{1,†}, Rongtao Jiang^{2,†}, Lijiang Wei^{1,3}, Vince D. Calhoun⁴, Bin Jing^{1,*}, Haiyun Li^{1,*}, Jing Sui^{1,3,4,*}

¹School of Biomedical Engineering, Capital Medical University, Xitoutiao No. 10, Youanmenwai Street, Fengtai District, 100069 Beijing, China,

²Department of Radiology and Biomedical Imaging, Yale University, New Haven, 300 Cedar Street, New Haven, 06510 CT, United States,

³State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, No. 19, Xijiekou Outer Street, Haidian District, 100875 Beijing, China,

⁴Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, Emory University, 55 Park Pl NE, Atlanta, 30303, GA, United States

*Corresponding authors: Bin Jing, School of Biomedical Engineering, Capital Medical University, Xitoutiao No. 10, Youanmenwai Street, Fengtai District, 100069 Beijing, China, Email: bjing@ccmu.edu.cn; Haiyun Li, School of Biomedical Engineering, Capital Medical University, Xitoutiao No. 10, Youanmenwai Street, Fengtai District, 100069 Beijing, China, Email: haiyunli@ccmu.edu.cn; Jing Sui, State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, No. 19, Xijiekou Outer Street, Haidian District, 100875 Beijing, China, Email: jsui@bnu.edu.cn

[†]Rongtao Jiang and Pujie Feng contributed equally to the work.

Resting-state functional connectivity (RSFC) has been widely adopted for individualized trait prediction. However, multiple confounding factors may impact the predicted brain-behavior relationships. In this study, we investigated the impact of 4 confounding factors including time series length, functional connectivity (FC) type, brain parcellation choice, and variance of the predicted target. The data from Human Connectome Project including 1,206 healthy subjects were employed, with 3 cognitive traits including fluid intelligence, working memory, and picture vocabulary ability as the prediction targets. We compared the prediction performance under different settings of these 4 factors using partial least square regression. Results demonstrated appropriate time series length (300 time points) and brain parcellation (independent component analysis, ICA100/200) can achieve better prediction performance without too much time consumption. FC calculated by Pearson, Spearman, and Partial correlation achieves higher accuracy and lower time cost than mutual information and coherence. Cognitive traits with larger variance among subjects can be better predicted due to the well elaboration of individual variability. In addition, the beneficial effects of increasing scan duration to prediction partially arise from the improved test-retest reliability of RSFC. Taken together, the study highlights the importance of determining these factors in RSFC-based prediction, which can facilitate standardization of RSFC-based prediction pipelines going forward.

Key words: cognitive trait; confounding factors; functional connectivity; ICA; individualized prediction.

Introduction

Recent years have seen increased interest in predicting cognitive traits using multimodal neuroimaging data (Rosenberg et al. 2016; Sui et al. 2020), especially as some large open-source dataset become increasingly available such as Human Connectome Project (HCP) and UK Biobank. Resting-state functional magnetic resonance imaging (rs-fMRI) is one the most popularly used imaging techniques in brain-behavior prediction. Resting-state functional connectivity (RSFC) is calculated using a variety of different methods. Due to the capability in detecting spatially remote but functionally coherent information, RSFC has been widely used in understanding the neurobiology of cognition and brain diseases of human brain (Jung et al. 2010; Cribben et al. 2012; Finn et al. 2015; Meskaldji et al. 2016; Jiang et al. 2020a). Numerous studies have revealed great potential of RSFC in revealing individual difference in cognitive and behavioral phenotypes (Power et al. 2011; Sui et al. 2020). However, an unsolved problem in these predictive

studies is that prediction accuracy is usually not high, which may limit their further usage in clinical diagnosis.

Because of this, it is essential to investigate the influence of different confounding factors in RSFC-based brain-behavior prediction studies. Lots of work has been done to examine the role of heterogeneous parameter settings, including global signal regression (Li et al. 2019), motion artifact correction (Nielsen et al. 2019), feature selection (Wei et al. 2020), sample size (Zaixu and Gaolang 2018), and prediction algorithms. Nevertheless, there are still additional confounding factors that may affect the prediction performance such as length of time series, brain parcellation approach, calculation of functional connectivity (FC), and variance of the prediction target (Van Dijk et al. 2010; Anderson et al. 2011; Jiang et al. 2020b). The first 3 factors determine the generation of RSFC used for the prediction, and the last one concerns the distribution of the predicted target. Though a few studies have performed some initial analysis of these factors (Au et al. 2015;

Elliott et al. 2019; Jiang et al. 2020b), their effects on the RSFC-based cognition prediction are still not clear.

Motivated by this, the current study aims to investigate the different impact of the 4 above-mentioned confounding factors: (i) time series length, (ii) functional connectivity calculation method, (iii) coarse to fine brain parcellation with different number of independent components (ICs; Meng et al. 2021), and (iv) the variance of the predicted target metrics. For this, we analyzed the large HCP ($n = 1,206$) dataset. Three representative cognitive traits like fluid intelligence (FI), working memory (WM), and picture vocabulary test (PVT) were used as the predicted targets, since they play important roles in higher-order cognition and can be linked with human intrinsic functional network connectivity (Sui et al. 2018, 2020; Uddin 2021). Prediction models were built by a widely used linear regression model, partial least squares regression (PLSR; Foodeh et al. 2020), since PLSR can capture the brain-behavior relationship without additional feature selection and can directly transform the input features into regression models while controlling other confounding factors. PLSR has been successfully used for prediction of multiple cognitive traits such as episodic memory and sustained attention (Meskaldji et al. 2016; Yoo et al. 2018; Chen et al. 2019; Fong et al. 2019; Jiang et al. 2020b).

Specifically, we compared the prediction performance of each of the 3 cognitive traits by testing 12 different time series lengths, 6 brain parcellations by independent component analysis (ICA; Calhoun et al. 2001; Calhoun and Adali 2012) with IC number ranging from 15 to 300, and 5 methods to calculate FCs. The variance of 3 cognitive scores was also controlled through sampling subjects for each prediction target, and the prediction accuracy was further compared under different variance of cognitive traits. Such an investigation may help us better understand the role of multiple confounding factors played in RSFC-based cognition prediction, which may enhance the interpretability and guide us to choose the most appropriate parameters in the individualized brain-behavior prediction in future studies.

Materials and methods

Participants

We used 1,206 healthy subjects (age 22–37, 550 males and 656 females) released from the WU-HCP (1,200), in which 2 scanning sessions of rs-fMRI data, REST1 and REST2, were collected with high spatial-temporal resolution during 2 successive days. Both sessions of data were scanned at right-to-left (run1) and left-to-right (run2) phase encoding direction, and the functional images comprised 1,200 volumes in each run. Both sessions were available for a total of 1,003 subjects with complete rs-fMRI runs (a total of 4,800 time points). In the study, REST1-run1 scan data were used to investigate the different impact of 4 confounding factors, whereas REST1-run1 and REST2-run1 data were used as the test-retest reliability dataset. For more details of inclusion

and exclusion criteria on the HCP dataset, please see Van Essen et al. (2013).

MRI acquisition

In the HCP S1200 dataset, high-resolution resting-state fMRI images were acquired using a customized Siemens Skyra 3-T scanner with a 32-channel head coil. Rs-fMRI data were collected with a multiband pulse sequence with the following scanning parameters: time repetition (TR) = 720 ms, time echo = 33.1 ms, field of view = 208×180 mm², flip angle = 52°, and voxel size = 2.0-mm isotropic cube (Smith et al. 2013).

fMRI preprocessing

The extensively preprocessed PTN (Parcellation+ Timeseries+ Netmats) data were employed in the study. Details of preprocessing steps can refer to Smith et al. (2013). In short, each run of rs-fMRI data went through a minimal preprocessed pipeline and ICA + FIX to eliminate the latent artifacts (Glasser et al. 2013). After brain region parcellation by group ICA with different number of ICs, the corresponding time courses/series of each ICs were also obtained. Correlations of time courses between any 2 ICs were then calculated by different methods (e.g. Pearson and Spearman) and transferred into Z-scores, resulting in a symmetric function network connectivity matrix for every subject. After removing diagonal elements, we extracted the upper triangle elements of the FC matrix as the candidate features for prediction.

Prediction with PLSR

In this study, we adopted PLSR to predict the cognitive traits. PLSR detects the relationship between 2 sets of variables by firstly projecting them onto a low-dimensional subspace to generate representative latent components and then detecting the optimal linear expression of the new variable with the prediction target (Guo and Mu 2011; Krishnan et al. 2011). This is similar to the combination of principal component analysis and canonical component analysis and has been successfully used for predicting various cognitive traits such as episodic memory and sustained attention (Meskaldji et al. 2016; Yoo et al. 2018; Chen et al. 2019; Fong et al. 2019; Jiang et al. 2020b).

Let $X_{n \times N}$ denote the feature matrix, where n is the number of samples, and N is the number of features, and let $Y_{n \times M}$ denote the label matrix, where M is the number of labels; then PLSR decomposes X and Y into the following form:

$$X_{n \times N} = U_{n \times d}(P_{N \times d})^T + E_{n \times N} \quad (1)$$

$$Y_{n \times M} = V_{n \times d}(Q_{M \times d})^T + F_{n \times M} \quad (2)$$

where U and V are matrices of the d extracted score vectors (latent variables), and P and Q represent matrices of weighting, i.e. the coefficient of the regression of X to U (or Y to V). E , F , and F^* are the residual errors. Partial least

squares decompose X and Y to obtain the maximized covariance between U and V . Based on X , Y , U , and V , an explicit $N \times M$ matrix B that satisfies the following linear relationship can be obtained:

$$Y_{n \times M} = X_{n \times N} (B_{N \times M}) + F_{n \times M}^* \quad (3)$$

For each cognitive trait, the prediction process was implemented within a nested 10-fold cross validation. The model performance was evaluated by the correlational values (R) between the predicted and observed cognitive scores. An inner 10-fold cross validation was used to determine the optimal parameters (latent component “ d ”) for PLSR, which ranged from 1 to 10, then parameters resulting in the highest prediction accuracy were determined as the optimal parameter; this is for 1-fold. Since the dataset was randomly divided into 10-folds, the whole prediction accuracy was obtained after looping all 10-folds. Note that the performances might depend on different sample partitions, thus the above procedure was repeated 100 times and the final prediction results were reported as the average value of 100 repetitions (1,000 loops).

Cognitive measure

The HCP dataset includes a battery of cognitive tests, and 3 representative cognitive abilities were chosen as the prediction targets in the study, including WM capacity and FI and PVT. The details of each cognitive score were described in HCP official document [https://www.humanconnectome.org/storage/app/media/documentation/s1200/HCP_S1200_Release_Reference_Manual.pdf]. The first 2 tests measured by National Institutes of Health (NIH) Toolbox were normalized to age-adjusted scores. After excluding the subjects with missing fMRI image data or cognitive measures, the number of healthy subjects who were retained for predicting cognitive traits of FI, WM, and picture vocabulary are 992, 996, and 988, respectively.

The analysis pipeline for the confounding factor test

Figure 1 displayed the whole testing pipeline for the 4 confounding factors, namely, functional connectivity type, variance of cognitive score, length of time series (time points) and number of ICs used in group ICA for brain parcellation. In particular, when predicting each of the 3 cognitive measures (WM, FI, and PVT), we compared the prediction accuracy by testing 5 methods to calculate FCs, 12 different time series lengths and 6 number of ICs for brain decomposition respectively. In addition, the variance of 3 cognitive scores was also controlled through sampling subjects. More details are provided below.

FC Type

FC is typically calculated by Pearson correlation on resting-state fMRI signal between paired brain regions.

However, multiple alternative methods have been reported to result in FC types displaying dissimilar sensitivity to motion artifact, test-retest reliability, and fingerprinting accuracy (Mahadevan et al. 2021). To verify whether different FC types may impact on the predictive performance of cognition, 5 representative FC types including Pearson, Spearman, partial correlation, mutual information (time domain), and coherence were used to calculate FC. Specifically, partial correlation is a measure of the linear correlation between 2 time series after regressing out the time series of all other nodes in the network. Mutual information is a statistical measure of the shared information between 2 time series, which could be defined through its Shannon entropy (Shannon 1948). Coherence is a measure of the cross-correlation between 2 signals in the frequency domain, and the average coherence in the frequency range [0.009 and 0.08 Hz] was used to define connectivity (Mahadevan et al. 2021). For these FC types, we used an open-source MATLAB toolbox (Zhou et al. 2009) for functional connectivity calculation. Besides the prediction accuracy, the spatial similarity of the features from different FC types were also compared (Tian and Zalesky 2021), which was calculated by ICC (1,1) on the feature weights (β coefficients) and finally reported with the average value in 100 times model training.

Variance of predicted target

For every cognitive ability, a fixed number of sample subsets were randomly chosen from the whole cohort to generate new dataset with different variance of cognitive score. In the study, the number of selected subsets was constantly set at 600 (~50%), and in order to make the variance of each repetition as different as possible, pre-defined score range thresholds were respectively used to ensure all selected subsets to be within this range.

Time series length

Although HCP records long rs-fMRI time series, most other datasets only collect 5–6-min rs-fMRI time series. There is large inconsistency in time series length among different datasets, so FC was constructed with 12 different time series lengths ranging from 100 to 1,200 with 100 time points as interspace (i.e. 100, 200, 300, ..., 1,200 from run1). All other factors in the prediction model were kept consistent in order to make sure the sole role of time series length.

Number of ICs

Group ICA has been found to be more suitable and adaptive for defining intrinsic functional networks (ICN, brain nodes) than anatomy atlases (Dadi et al. 2019), hence, 6 number of ICs (15/25/50/100/200/300 ICs) were adopted to provide as a coarse to fine brain parcellation, resulting in FCs with remarkably different dimensions. All other settings were identical when examining brain parcellation effects.

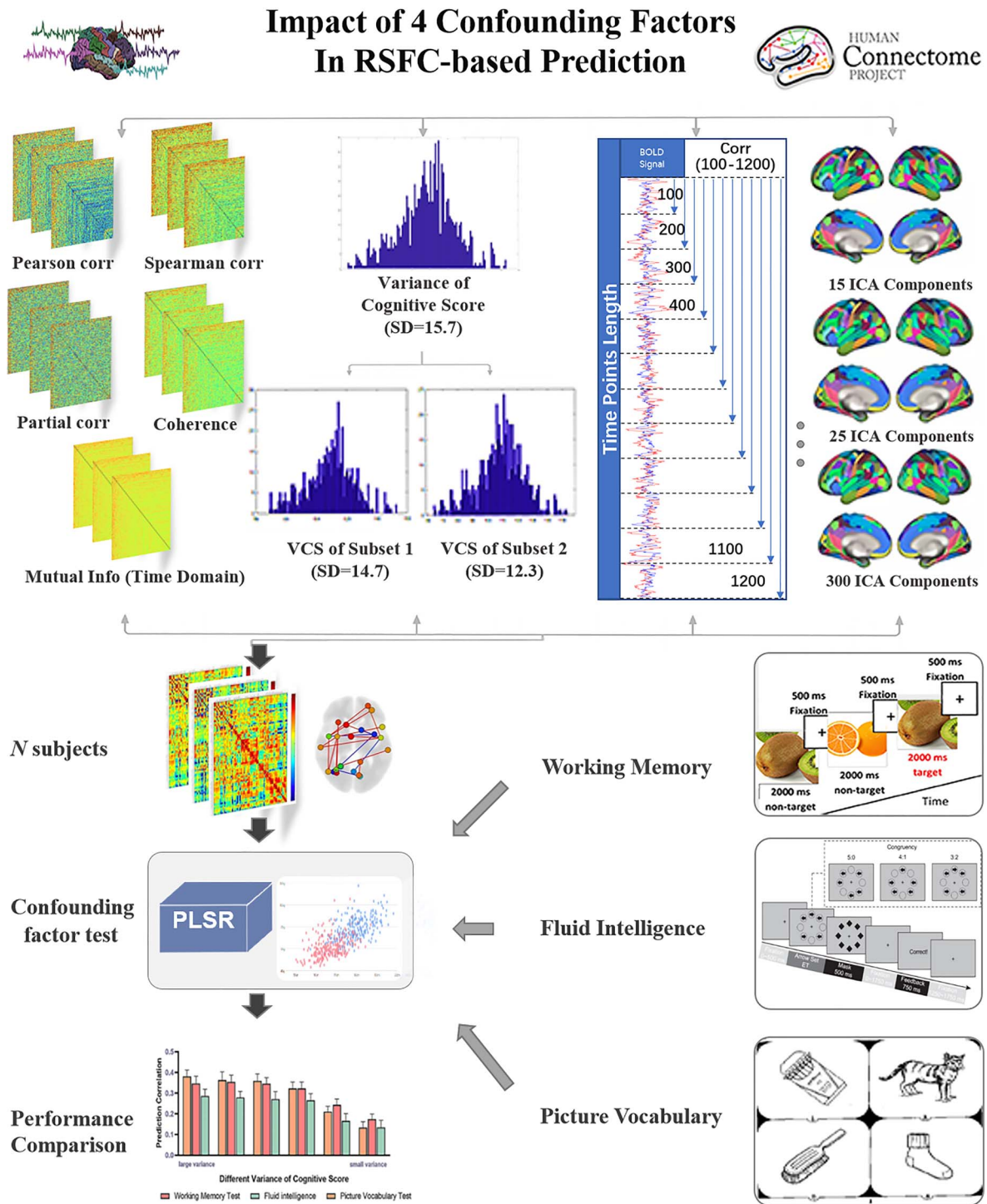


Fig. 1. Pipeline to testing the impact of 4 confounding factors.

Test-retest reliability

Beyond the prediction performance, test-retest reliability is also a crucial metric influencing the feasibility and rational explanation of the brain-behavior prediction. Therefore, we further tested intraclass correlation coefficient (ICC; Noble et al. 2021) under different length of time series, and then investigate the relationship between ICC and the prediction accuracy, as well as the contribution of FC with different reliability to the prediction.

ICC is defined in the same manner as in Paek et al. (2019) with different time series length:

$$ICC = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \quad (4)$$

where σ_b is the between-subject variation and σ_w is the within-subject variation. The value of ICC ranges from 0 to 1, and the larger the ICC value, the higher the reliability is.

To reduce the computational burden, we selected IC number=200 as the FC node number, i.e. calculated ICC using 19,900 FCs, and classified all ICCs into 4 categories as reported in (Zhang et al. 2018; Elliott et al. 2019): $ICC < 0.39$, poor reliability; $0.4 < ICC < 0.59$, moderate reliability; $0.6 < ICC < 0.74$, good reliability, and $0.75 < ICC < 1$, excellent reliability. Next, the FCs with $ICC < 0.4$ were discarded, and the remaining FC with good reliability served as the input features to the prediction framework. If the prediction accuracy did not change much compared with using all FCs, it means the FC features with good reliability play the key role in capturing stable individual differences. In addition, only FCs with the low reliability were also used as the input feature for performance comparison. Note that in order to exclude the influence of the feature number on the prediction performance, the whole FCs were randomly selected as the same number of FC with $ICC > 0.4$ for prediction and repeated this procedure 50 times.

In addition, we further explored whether the test-retest reliability of FC was influenced by head motion, sample size, gender, and age. For head motion, we used the relative RMS (root-mean-square) movement, provided by the HCP, as the primary measure of motion, and the correlation between the test-retest reliability of mean RMS and FC was evaluated for different time series length. For sample size, gender and age, we grouped the whole dataset into respective subgroups using the following criteria: age (22–25, 26–30, and 31+), gender (male and female) and sample size ($n = 100, 200, 400$, and 600) by random sampling, and compared whether there was significant difference in ICC between subgroups.

Computational cost

Computational cost is another noticeable concern especially for big data. In our prediction, estimation of the latent parameter “ d ” in PLSR and the 100 times repetitions of cross validation were the main source of computational cost. The running time was recorded and compared for every time series length, brain parcellation, and functional connectivity type. To reduce the computational cost and to facilitate comparison with other studies, Pearson correlation is used in prediction models when evaluating time series length and brain parcellation. All procedures were carried out with 2 multi-core CPUs (Intel(R) Xeon(R) E5–2630 v4 @2.20GHz, 10 cores, 20 threads) with 128 GB memory.

Results

Prediction accuracies under different parameters

Prediction performance while controlling for different confounding factors is shown in Fig. 2. Figure 2a compared the prediction accuracy measured by Pearson correlation using FCs generated from 6 different number of ICs and with 12 length of time series. It is clear that finer brain parcellation (more ICs) yield better prediction performance in general, which may influence the prediction

more than the time length. For example, in WM prediction, the performance from 300 components with 100 time points ($r = 0.260 \pm 0.010$) is still significantly higher ($P < 1.0 \times 10^{-4}$) than that from 15 components with 1,200 time points ($r = 0.223 \pm 0.006$). But higher IC number also comes with additional computational cost (Fig. 4a). And increase of time points also improved prediction accuracy (e.g. when $IC = 300$, WM: r from 0.282 ± 0.010 to 0.373 ± 0.008 ; PVT: r from 0.356 ± 0.007 to 0.426 ± 0.006 ; FI: r from 0.260 ± 0.009 to 0.307 ± 0.008 ; $P < 0.05$ FDR corrected) and tended to be stable after certain time points. However, longer fMRI time lengths may not necessarily ensure better prediction accuracy, e.g. for FI, the inflection point is time = 300, which may be influenced by the intrinsic dynamics of fMRI signal (Calhoun et al. 2014).

Figure 2b illustrated the prediction performances under different variances of cognitive score (VCS) when $IC = 200$ Time point = 1,200. As shown, for the same predicted target, subjects with the mid or higher VCS can better reflect the individual variability, thus leading to better estimation accuracy accordingly. In addition, we also tried to change VCS of the prediction target by min-max normalization and Z-score transformation, but the results did not show any improvement in prediction, indicating that simply enlarging the variance brings no effect to the prediction. Although among the 5 different FC types, Pearson, Spearman, and partial correlation performed similarly or slightly better than mutual information and coherence, as shown in Fig. 2c, where $IC = 200$ time point = 1,200. Thus, we choose Pearson correlation when calculating the computational cost. The result of feature spatial similarity (Fig. 2d) illustrated that the feature weights of Spearman correlation and Pearson correlation were very similar ($ICC = 0.901$), whereas the similarity of partial correlation and Pearson correlation was relatively poor ($ICC = 0.190$). In addition, Mutual information and coherence have almost no similarity with other 3 FC types.

Influence of test-retest reliability on prediction accuracy

Figure 3 illustrate the test-retest reliability on prediction accuracy with different time points via ICC. According to Fig. 3a, when the fMRI scan time increase, the percentage of FC with poor reliability decrease and a large portion of them turn into good reliability. Further calculation in Fig. 3b revealed that with the increase of time series length, the mean ICC of the FC significantly improved.

Specifically, we observed very high correlation ($r = 0.989$) between the prediction accuracy and the mean ICC under different scan duration (time points) in Fig. 3c, and the mean value of $ICC > 0.4$ corresponds to the robust prediction accuracy at Pearson correlation $r > 0.35$. Moreover, we further studied the contribution of FC with different reliability to the prediction by using work memory as an example. As shown in Fig. 3d, after removing the FC with $ICC < 0.4$, the prediction accuracy would slightly increase compared with using all FCs (in most

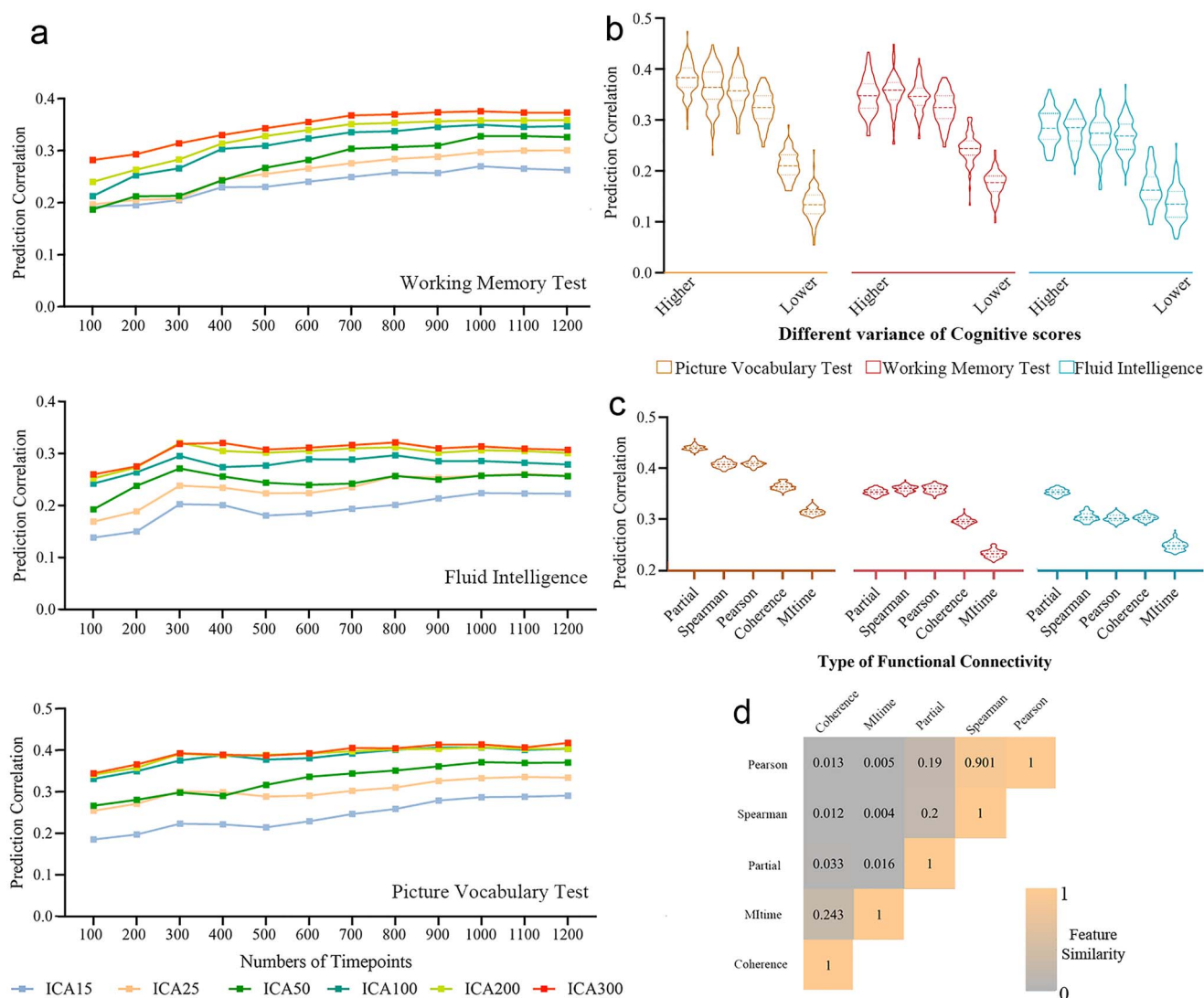


Fig. 2. Prediction performance comparison under 4 confounding factors a) with 12 different length of time series and 6 IC numbers, b) under different VCS, c) with 5 methods to calculate FC types, and d) Feature spatial similarity between different FC types.

cases, $r > 0.33$) because the non-reliable information was thrown off; Instead, deleting FC with ICC > 0.4 will however worsen the prediction accuracy significantly ($P < 0.01$, FDR corrected), the performance maintained at low level $r < 0.27$. In contrast, the prediction performance of random feature selection was between the results with high-reliability FC and those with low-reliability FC. In Fig. 3e, sample size, age and gender didn't obviously affect the test-retest reliability. However, there was a strong correlation between the test-retest reliability of mean RMS and FC ($r = 0.851$).

As for the computational cost, Fig. 4a indicated that difference of time series length did not affect the time consumption remarkably; instead, more ICs for fine brain parcellation would significantly increase the time cost, because more ICs result in nearly square growth of FC numbers, very fragmented brain regions, and consequently, the dozens and hundreds of time cost. Considering the performance in Figs. 2a and 4a, IC num=100/200 could be the optimal choice between accuracy and time cost. Furthermore, among 5 FC types,

mutual information and coherence were more time-consuming than the other 3 correlation types that act similarly (Fig. 4b). Different VCS did not affect the time consumption.

Discussion

In this study, we aimed to discover the optimal settings and influence of 4 confounding factors in FC-based cognition prediction, including length of time series, method to calculate functional connectivity, brain parcellation choice by ICA, and variance of the predicted target. We compared the prediction performance of FI, WM, and picture vocabulary ability by applying PLSR on RSFC from 1206 subjects of HCP data. Results demonstrated that an appropriate time series length at ~ 300 time points and a relatively finer brain parcellation (ICA100) could achieve better prediction performance without too much time consumption. FC calculated by Pearson, Spearman, and partial correlation can achieve higher estimation accuracy and lower time cost than mutual information

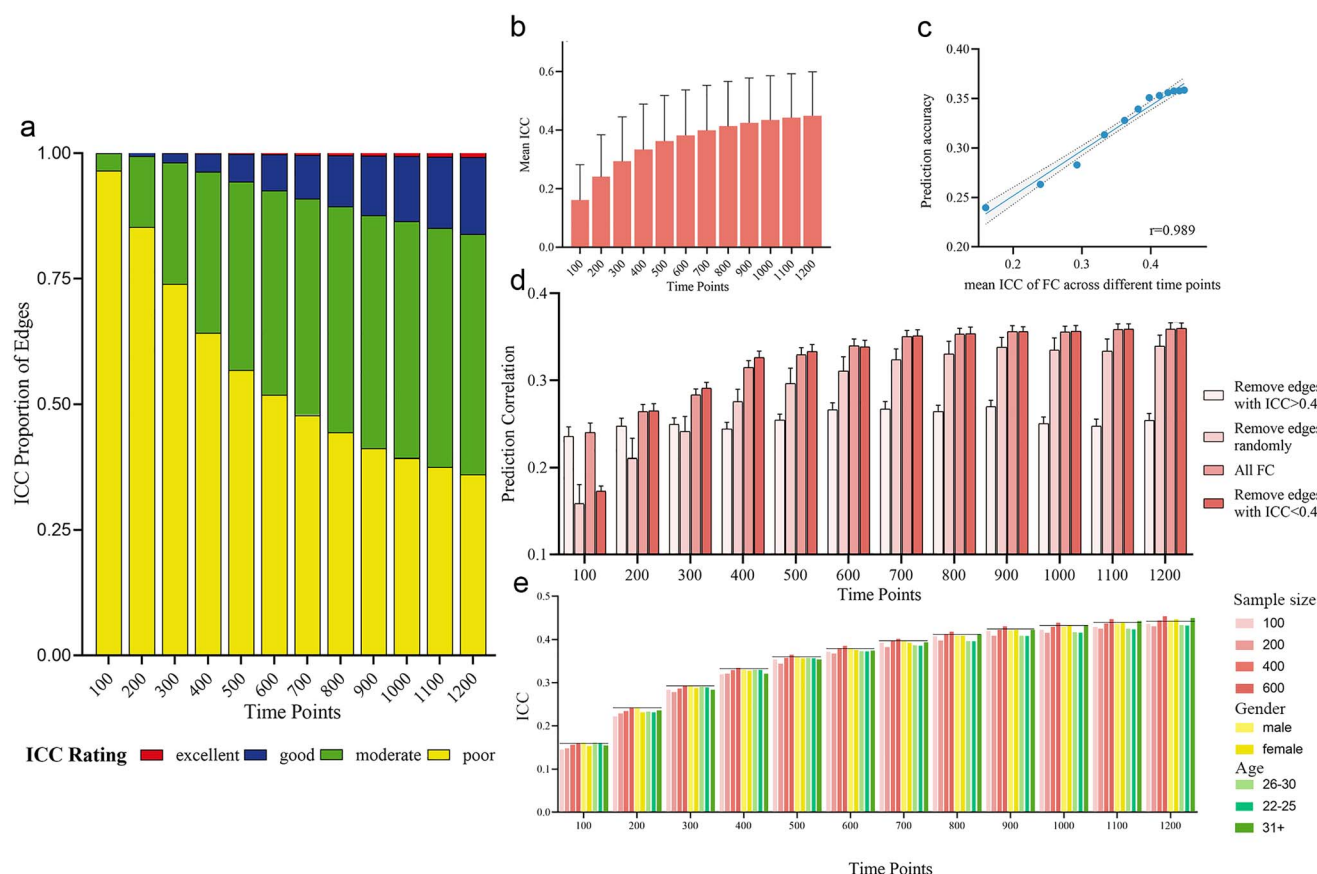


Fig. 3. Influence of test-retest reliability with different time points. a) ICC rating percentage in all edges. b) The average ICC of all FC increases with the time points numbers. c) The mean ICC is strongly correlated with the prediction accuracy. d) Shows that removing the high-reliability edges will lead to a significant loss of prediction performance, whereas retaining the high-reliability edges and removing the low-reliability edges will not cause significant changes. In addition, when randomly selecting the same number of edges as the high-reliability edge, the average prediction performance is between the above 2 conditions. e) Sample size, age, and gender have little impact on the test-retest reliability. The horizontal line stands for the ICC value in corresponding whole dataset.

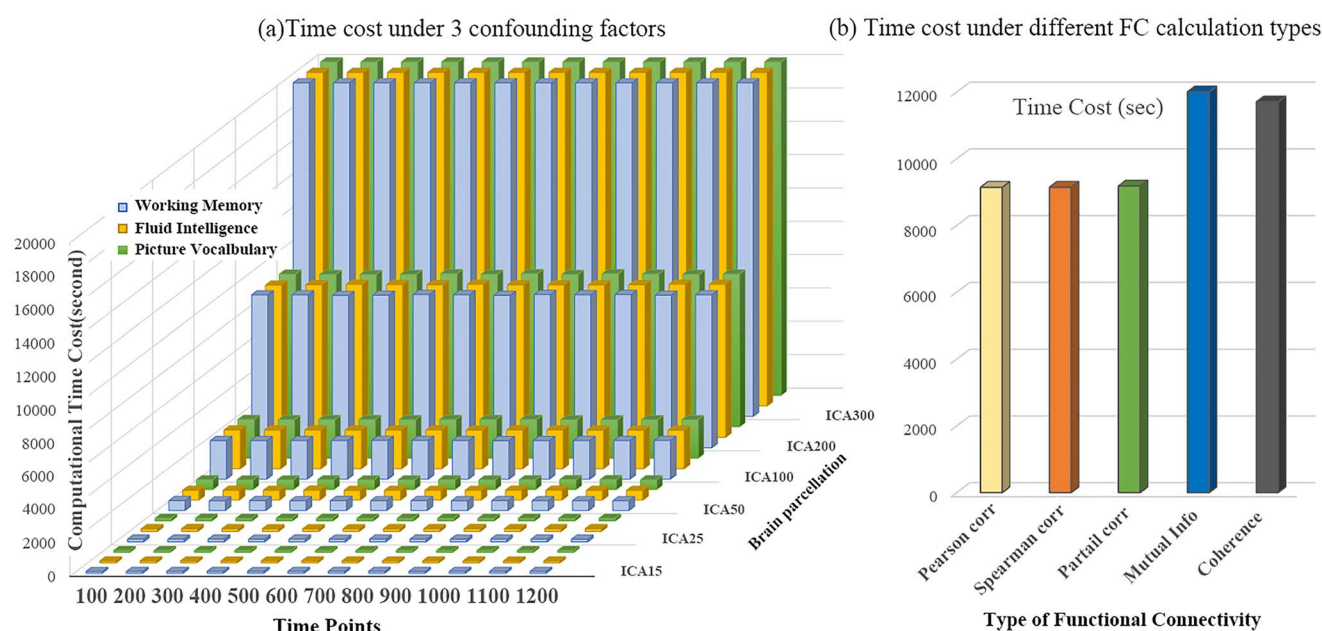


Fig. 4. The mean computational time cost under different factor settings.

and coherence. In addition, cognitive traits with larger variance among subjects can be better predicted due to the well elaboration of individual variability. This work on optimal parameter exploration can be very helpful for the neuroimaging-based individual trait prediction.

For the test-retest reliability, both the percentage of ICC ratings and the mean ICC of the whole brain were gradually improved when the time series length increased (Fig. 3a and b), suggesting that longer time series could yield better reliability and more reliable FC features. This in turn may better capture individual differences (Van Dijk et al. 2010; Anderson et al. 2011; Jiang et al. 2020b; Fig. 3c) and improve prediction of cognition traits (Finn et al. 2015). In light of this, rs-fMRI data with time points < 200 may generate FCs with poor test-retest reliability, which is thus not recommended in prediction and may be mitigated by the spatially constrained ICA (Du et al. 2020). More importantly, our results suggest removing FC with low reliability ($ICC < 0.4$) may benefit on the prediction performance while reducing the feature dimension (computation load) simultaneously, by contrast, deleting FC with high reliability ($ICC > 0.4$) can degrade the prediction accuracy significantly (Fig. 3d). In parallel, removing FC randomly showed reduced prediction performance than using original whole FCs, and there is a strong positive correlation between ICC and prediction accuracy. Taken together, these results suggest that FCs with high ICC are the main contributors to the prediction of cognitive ability. Coincidentally, Finn et al. also demonstrated an improved identification accuracy with longer time courses when using FC as brain fingerprint (Finn et al. 2015). Moreover, Van De Ville et al. reported longer fMRI sessions improved the identification for human brain fingerprint (Van De Ville et al. 2021), and demonstrated the optimal fingerprints at a time scale of 200 s (~280 time points). In addition, the most important regions contributing to individual fingerprint were mainly located at “higher-order” cognitive systems. Our study reported consistent conclusions and further proved that FCs with higher reliability were the main contributors to individual cognitive prediction. Overall, this finding suggests that the beneficial effects of increasing scan duration to prediction partially stem from the improved test-retest reliability of individual functional connectome features.

It is not unexpected that finer brain parcellation yield better prediction performance than the coarser one, given the fact that fine-parcellations provide much more information at higher spatial resolution. Note that the prediction performances increased slowly with the refinement of brain parcellation, especially when number IC > 100, whereas the computational cost increased substantially as well (IC num = 200 or 300). One potential reason may be that the 300-ICA parcellation may result in more noise components and overlapped fragmented brain regions, which may be not helpful but even harmful to predictive performance (Dadi et al. 2019). Furthermore, predictive models based on ICA-300

parcellation have an increased likelihood of overfitting, as the feature dimension was much larger (44,850 vs 19,900 vs. 4,950). Taken together, ICA100 or ICA200 parcellation are better choice for individual cognitive prediction, which are in a similar range for the number of brain regions included in most brain atlases. In contrast, coarse brain parcellation with IC number < 50 are not recommended in cognition prediction; as the prediction accuracy was relatively low for even the longer time series.

With regard to the FC type, generally speaking, different FC types show different tolerance to the physiological noise like head motion (Mahadevan et al. 2021), which may affect the delineating of individual difference in FC. We recommend the commonly used methods to calculate FC, such as Pearson correlation, Spearman correlation, and Partial correlation, they are simple yet effective compared with complex models such as mutual information or coherence. The potential reasons may lie in: (i) The definition of mutual information itself is sensitive to unstable physiological noise during the scanning, thereafter weakening its performance in prediction. (ii) Coherence is a type of FC constructed in the frequency domain, although it is not sensitive to the head motion, it displays not high test-retest reliability, subject identity and system identifiability (Mahadevan et al. 2021), thereafter limiting its performance in prediction. Although all of FCs generated by Pearson, Spearman, and Partial correlation displayed good prediction performances, the adopted feature sets in Partial correlation were largely different from the other 2, implying a potential way to integrate these features together to improve the prediction performance.

Finally, variance of predicted target also affects the prediction performance, and larger variance in cognitive abilities is able to elaborate individual variability more dedicatedly than binary scores or nearly constant values slightly, thus may yield relatively better predictive accuracy as shown in Fig. 2c. Although PLSR model does not need additional feature selection step, it first compresses the original high-dimensional data like PCA based on the data variance, thereafter leading to more effective feature reduction representation. Specifically, the ceiling effect has been reported in previous WM study (Peterburs et al. 2019), when we deleted some extremely high scored participants in the WM prediction, the accuracy slightly improved (Fig. 2b). Therefore, for cognitive scores with possible ceiling effect, to remove the outliers can be helpful and essential for prediction.

Implications for future studies

The study found all 4 factors have significant impacts on the predictive performance, which may have beneficial implications for future prediction studies. In terms of time series length, we recommend the scanning time with 200–800 time points. On one hand, the prediction performance will be greatly reduced when the time length was < 200. On the other hand, the prediction

performance does not improve significantly for time series longer than 800 (Jiang et al., 2020b). Longer scan time could place high burden on participants to finish the scanning, especially those with psychiatric disorder and children. In addition, considering the prediction performance, computational consumption and model interpretability, ICA 200 components is suggested as an optimal brain parcellation dimensionality. For FC type, we recommend FC calculated by simple full or partial correlation because of their prediction performance and computational cost. It may be not necessary to use complex models like mutual information or Coherence for prediction, at least in current settings. At last, variance of predicted cognitive scores should also be focused. Every cognitive measure displays unique data distribution, and ensuring relatively large variance may be very beneficial to the prediction.

Limitation and future directions

Several limitations should be mentioned in this study. (i) We only evaluated the PLSR model for prediction, and there are many other commonly-used machine learning models in the field, which could be evaluated in future studies. (ii) Considering the huge computational burden, only 3 representative cognitive measures were examined. (iii) The study was performed on HCP dataset with TR=0.72 s, however, different TRs may affect the time series length conclusion. (iv) Although the HCP dataset is large, its sample size is still much lower than the feature dimension, which can be alleviated using other open-source datasets like UK biobank and Adolescent Brain Cognitive Development (ABCD).

Conclusion

In summary, this study explored the effects of 4 factors in RSFC-based prediction for cognitive scores, and found that RSFC generated with longer time series length and larger cognitive scores variance could yield better prediction performances. Moreover, we propose appropriate selection of FC type and brain parcellation may also be beneficial to the prediction accuracy. Furthermore, this study emphasizes the influence of these potential factors on revealing the relationship between behavior and brain connectivity, which may help construct standardized RSFC-based prediction pipeline in future. Taken together, the study highlights the importance to select the optimal confounding factors in RSFC-based prediction when revealing brain-behavior relationship, which may guide the parameter selection and facilitate the standardization of RSFC-based prediction pipeline in future.

Authors' contribution

PF performed the data analysis and wrote the draft. RJ provided help in data analysis, designed the experiments and revised the paper. LW performed some data analysis.

BJ conceived and designed the experiments, and revised the paper. HL and VDC revised the draft. JS supervised the whole study and revised the paper. All authors have read and approved the manuscript.

Data and code availability statement

The data were available from request to the Human Connectome Project and WUMinn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil). Code for RSFC-based prediction analyses is freely available at the GitHub repository (<https://github.com/fengpujie/Four-Confounding-Factors>).

Funding

This research was supported by the National Natural Science Foundation of China (nos. 82022035 and 61773380 to JS); Beijing Municipal Commission of Education (no. KM202010025025 to BJ); Beijing Municipal Natural Science Foundation (no. L192044 to HL), Project of "Practical Training Plan" (College Students Scientific Training Plan Project of Beijing Colleges and Universities High-level Talents Cross-Cultivation to PF), the National Science Foundation (no. 2112455) and the National Institutes of Health (nos. R01MH117107 and R01MH118695 to VDC).

Conflict of interest: None declared.

References

- Anderson JS, Ferguson MA, Lopez-Larson M, Yurgelun-Todd D. Reproducibility of single-subject functional connectivity measurements. *AJNR Am J Neuroradiol*. 2011;32:548–555.
- Au J, Sheehan E, Tsai N, Duncan GJ, Buschkuhl M, Jaeggi SM. Improving fluid intelligence with training on working memory: a meta-analysis. *Psychon Bull Rev*. 2015;22:366–377.
- Calhoun VD, Adali T. Multisubject independent component analysis of fMRI: a decade of intrinsic networks, default mode, and neurodiagnostic discovery. *IEEE Rev Biomed Eng*. 2012;5:60–73.
- Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A method for making group inferences from functional MRI data using independent component analysis. *Hum Brain Mapp*. 2001;14:140–151.
- Calhoun VD, Miller R, Pearlson G, Adali T. The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron*. 2014;84:262–274.
- Chen C, Cao X, Tian L. Partial least squares regression performs well in MRI-based individualized estimations. *Front Neurosci*. 2019;13:1282.
- Cribben I, Haraldsdottir R, Atlas LY, Wager TD, Lindquist MA. Dynamic connectivity regression: determining state-related changes in brain connectivity. *NeuroImage*. 2012;61:907–920.
- Dadi K, Rahim M, Abraham A, Chyzyhyk D, Milham M, Thirion B, Varoquaux G. Benchmarking functional connectome-based predictive models for resting-state fMRI. *NeuroImage*. 2019;192:115–134.
- Du Y, Fu Z, Sui J, Gao S, Xing Y, Lin D, Salman M, Abrol A, Rahaman MA, Chen J, et al. NeuroMark: an automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders. *Neuroimage Clin*. 2020;28:102375.

- Elliott ML, Knodt AR, Cooke M, Kim MJ, Melzer TR, Keenan R, Ireland D, Ramrakha S, Poulton R, Caspi A, et al. General functional connectivity: shared features of resting-state and task fMRI drive reliable and heritable individual differences in functional brain networks. *NeuroImage*. 2019;189:516–532.
- Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat Neurosci*. 2015;18:1664–1671.
- Fong AHC, Yoo K, Rosenberg MD, Zhang S, Li CR, Scheinost D, Constable RT, Chun MM. Dynamic functional connectivity during task performance and rest predicts individual differences in attention across studies. *NeuroImage*. 2019;188:14–25.
- Foodeh R, Ebadollahi S, Daliri MR. Regularized partial Least Square regression for continuous decoding in brain-computer interfaces. *Neuroinformatics*. 2020;18:465–477.
- Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, et al. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*. 2013;80:105–124.
- Guo G, Mu G, editors. *Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression*. Vol. 2011. CVPR; 2011. pp. 657–664.
- Jiang R, Calhoun VD, Fan L, Zuo N, Jung R, Qi S, Lin D, Li J, Zhuo C, Song M, et al. Gender differences in connectome-based predictions of individualized intelligence quotient and sub-domain scores. *Cereb Cortex*. 2020a;30:888–900.
- Jiang R, Zuo N, Ford JM, Qi S, Zhi D, Zhuo C, Xu Y, Fu Z, Bustillo J, Turner JA, et al. Task-induced brain connectivity promotes the detection of individual differences in brain-behavior relationships. *NeuroImage*. 2020b;207:116370.
- Jung YC, Ku J, Namkoong K, Lee W, Kim SI, Kim JJ. Human orbitofrontal-striatum functional connectivity modulates behavioral persistence. *Neuroreport*. 2010;21:502–506.
- Krishnan A, Williams LJ, McIntosh AR, Abdi H. Partial least squares (PLS) methods for neuroimaging: a tutorial and review. *NeuroImage*. 2011;56:455–475.
- Li J, Kong R, Liegeois R, Orban C, Tan Y, Sun N, Holmes AJ, Sabuncu MR, Ge T, Yeo BTT. Global signal regression strengthens association between resting-state functional connectivity and behavior. *NeuroImage*. 2019;196:126–141.
- Mahadevan AS, Tooley UA, Bertolero MA, Mackey AP, Bassett DS. Evaluating the sensitivity of functional connectivity measures to motion artifact in resting-state fMRI data. *NeuroImage*. 2021;241:118408.
- Meng X, Iraj A, Fu Z, Kochunov P, Belger A, Ford J, McEwen S, Mathalon DH, Mueller BA, Pearlson G, et al. Multimodal order independent component analysis: a data-driven method for evaluating brain functional network connectivity within and between multiple spatial scales. *Brain Connect*. 2021.
- Meskaldji DE, Preti MG, Bolton TA, Montandon ML, Rodriguez C, Morgenthaler S, Giannakopoulos P, Haller S, Van De Ville D. Prediction of long-term memory scores in MCI based on resting-state fMRI. *Neuroimage Clin*. 2016;12:785–795.
- Nielsen AN, Greene DJ, Gratton C, Dosenbach NUF, Petersen SE, Schlaggar BL. Evaluating the prediction of brain maturity from functional connectivity after motion Artifact Denoising. *Cereb Cortex*. 2019;29:2455–2469.
- Noble S, Scheinost D, Constable RT. A guide to the measurement and interpretation of fMRI test-retest reliability. *Curr Opin Behav Sci*. 2021;40:27–32.
- Paek EJ, Murray LL, Newman SD, Kim DJ. Test-retest reliability in an fMRI study of naming in dementia. *Brain Lang*. 2019;191:31–45.
- Peterburs J, Blevins LC, Sheu YS, Desmond JE. Cerebellar contributions to sequence prediction in verbal working memory. *Brain Struct Funct*. 2019;224:485–499.
- Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, et al. Functional network organization of the human brain. *Neuron*. 2011;72:665–678.
- Rosenberg MD, Finn ES, Scheinost D, Papademetris X, Shen X, Constable RT, Chun MM. A neuromarker of sustained attention from whole-brain functional connectivity. *Nat Neurosci*. 2016;19:165–171.
- Shannon CE. A mathematical theory of communication. *Bell Syst Tech J*. 1948;27.
- Smith SM, Beckmann CF, Andersson J, Auerbach EJ, Bijsterbosch J, Douaud G, Duff E, Feinberg DA, Griffanti L, Harms MP, et al. Resting-state fMRI in the human connectome project. *NeuroImage*. 2013;80:144–168.
- Sui J, Qi S, van Erp TGM, Bustillo J, Jiang R, Lin D, Turner JA, Damaraju E, Mayer AR, Cui Y, et al. Multimodal neuromarkers in schizophrenia via cognition-guided MRI fusion. *Nat Commun*. 2018;9:3028.
- Sui J, Jiang R, Bustillo J, Calhoun V. Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: methods and promises. *Biol Psychiatry*. 2020;88:818–828.
- Tian Y, Zalesky A. Machine learning prediction of cognition from functional connectivity: are feature weights reliable? *NeuroImage*. 2021;245:118648.
- Uddin LQ. Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. *Nat Rev Neurosci*. 2021;22:167–179.
- Van De Ville D, Farouj Y, Preti MG, Liégeois R, Amico E. When makes you unique: temporality of the human brain fingerprint. *Sci Adv*. 2021;7:eabj0751.
- Van Dijk KR, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *J Neurophysiol*. 2010;103:297–321.
- Wei L, Jing B, Li H. Bootstrapping promotes the RSFC-behavior associations: an application of individual cognitive traits prediction. *Hum Brain Mapp*. 2020;41:2302–2316.
- Yoo K, Rosenberg MD, Hsu WT, Zhang S, Li CR, Scheinost D, Constable RT, Chun MM. Connectome-based predictive modeling of attention: comparing different functional connectivity features and prediction methods across datasets. *NeuroImage*. 2018;167:11–22.
- Zaixu C, Gaolang G. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *NeuroImage*. 2018;178:622–637.
- Zhang C, Baum SA, Adduru VR, Biswal BB, Michael AM. Test-retest reliability of dynamic functional connectivity in resting state fMRI. *NeuroImage*. 2018;183:907–918.
- Zhou D, Thompson WK, Siegle G. MATLAB toolbox for functional connectivity. *NeuroImage*. 2009;47:1590–1607.