Is resting state fMRI better than individual characteristics at predicting cognition?

Amir Omidvarnia[1,2*], Leonard Sasse[1,2], Daouia I. Larabi[1,2], Federico Raimondo[1,2], Felix Hoffstaedter[1,2], Jan Kasper[1,2], Juergen Dukart[1,2], Marvin Petersen[3], Bastian Cheng[3], Götz Thomalla[3], Simon B. Eickhoff[1,2], Kaustubh R. Patil[1,2]

[1]Institute of Neuroscience and Medicine, Brain & Behavior (INM-7), Research Center Jülich, Wilhelm-Johnen-Straße, Jülich, 52428, Germany
[2]Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Moorenstr. 5, Düsseldorf, 40225, Germany
[3]Klinik und Poliklinik für Neurologie, Kopf- und Neurozentrum, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

*Corresponding author(s). Email(s): a.omidvarnia@fz-juelich.de

Abstract

Changes in spontaneous brain activity at rest provide rich information about behavior and cognition. The mathematical properties of resting-state functional magnetic resonance imaging (rsfMRI) are a depiction of brain function and are frequently used to predict cognitive phenotypes. Individual characteristics such as age, gender, and total intracranial volume (TIV) play an important role in predictive modeling of rsfMRI (for example, as "confounders" in many cases). It is unclear, however, to what extent rsfMRI carries independent information from the individual characteristics that is able to predict cognitive phenotypes. Here, we used kernel ridge regression modeling to thoroughly examine the predictability of four cognitive phenotypes in 20,000 healthy UK Biobank subjects. We extracted common rsfMRI features of functional brain connectivity (FC) and temporal complexity (TC). We assessed the ability of these features to predict outcomes in the presence and absence of age, gender, and TIV. Additionally, we assessed the predictiveness of age, gender, and TIV only. We find TC and FC features to perform comparably with regard to predicting cognitive phenotypes. As compared to rsfMRI features, individual characteristics provide systematically better predictions with smaller sample sizes and, to some extent, in larger cohorts. It is also consistent across different levels of inherent temporal noise in rsfMRI. Our results suggest that when the objective is to perform cognitive predictions as opposed to understanding the relationship between brain and behavior, individual characteristics outperform rsfMRI features.

Keywords: functional MRI, temporal complexity, functional connectivity, cognitive phenotype, fluid intelligence, memory, entropy, hurst exponent, confound removal, kernel ridge regression

Resting-state functional magnetic resonance imaging (rsfMRI) is a widely-used technique for studying human brain function [1]–[3]. Functional brain connectivity (FC) is an important aspect of rsfMRI defined as the statistical dependence between cortical and subcortical areas during periods of rest or low cognitive demand [4]. A common application of rsfMRI is for the prediction of cognitive performance [5]–[9], and clinical phenotypes [5], [10]–[13] [14]–[20]. It is usually accomplished by extracting various features from rsfMRI, such as widely used FC measures, and using them for predictive modeling. This approach has been boosted by modern MRI scanners with powerful magnetic field strengths, large databases, high-performance computing systems, computational software packages, and improved machine learning techniques [21]. However, the field has struggled to advance to real-world applications due to systematic challenges such as modest prediction accuracy in large populations ($N_{subject}$ > 2000) [5], [6], [22], [23] and biased findings in studies with small sample sizes [23]. This makes it difficult to use the current predictive models of cognitive phenotypes for generating individual-level outputs in real-world applications. An effective improvement could be to consider features beyond prevalent FC measures.

Functional brain connectivity and large-scale nonlinear interactions between brain regions are tightly related to each other [24], [25]. Nonlinear brain dynamics help to process information from the neuronal to the observable hemodynamic levels in an adaptive way [26]. The nonlinearity of functional connections at different spatial scales gives rise to a temporally complex (TC) behavior in the hemodynamic response of the brain across time, as measured by fMRI [27], [28]. In fact, if neural connections do not fluctuate in a balanced way throughout time, emotions, learning, and memory cannot be managed by the brain effectively [29]. This is in line with the brain criticality hypothesis, which postulates that neural networks and, by extension, many aspects of brain function, self-organize into a critical state or a transition between ordered and disordered states [30]. There is evidence that the TC of rsfMRI and cognitive phenotypes are correlated [31]–[34], making it promising for brain-behavior predictions [31], [32], [35][36]–[38][36], [38]–[40]. Having said that, the relationship between these rsfMRI features and “noise” across different brain areas is not entirely understood yet. In the context of cognitive phenotypic prediction, it is a significant challenge to define what the “signal of interest” is given the influence of numerous factors, such as internal brain states, external stimuli, the impact of thermal noise inside the MRI scanner, the inevitable impact of head motion, and the persistence of physiological changes such as heartbeat and respiration [41].

In the current rsfMRI-based prediction pipelines, individual characteristics such as age, gender, and total intracranial volume (TIV) are frequently treated as "confounds" and removed from the rsfMRI features or from the prediction targets before analysis [42]. It stems from the idea that any information other than that directly related to brain activity should be discarded because it may prevent us from determining the neuronal origin of the predictive signal [43]. However, a previous study has shown that features based on individual characteristics might be better at diagnosing brain abnormalities than those based on fMRI [44]. Nevertheless, it is unclear how much of the shared information between rsfMRI and cognitive phenotypes is lost as a result of removing individual characteristics.

To examine these questions, we developed four cognitive prediction scenarios using rsfMRI features based on the role of individual characteristics. We sought to evaluate the relevance of treating commonly used individual characteristics (age, gender, and TIV) as "confounds" for cognitive phenotypic prediction in contrast to the TC and FC features of rsfMRI. We used a large subset of the UK Biobank ($N_{subject}$ = 20,000) that included high-resolution rsfMRI and four cognitive phenotypes: fluid intelligence, processing speed, visual memory, and numerical memory [45]. We used nine brain region-wise rsfMRI features to predict the cognitive phenotypes. The rsfMRI features covered five prominent characteristics of FC (*fractional amplitude of low-frequency fluctuations* or *fALFF* [46], *local correlation* or *LCOR* [47], *global correlation* or *GCOR* [48], *Eigenvector centrality* or *EC* [49], and *weighted clustering coefficient* or *wCC* [49]) as well as four TC metrics (*Hurst exponent* or *HE* [50], *Weighted permutation entropy* or *wPE* [51], *Range entropy* or *RangeEn* [52], and *Multiscale entropy* or *MSE* [53]). We then entered these features into four predictive modeling pipelines, considering different roles for age, gender, and TIV as individual characteristics. These analysis pipelines performed predictive modeling using: (*i*) rsfMRI features without removing individual characteristics, (*ii*) rsfMRI features with the removal of individual characteristics (i.e., treating them as confounds), (*iii*) a combination of rsfMRI features and individual characteristics, and (*iv*) individual characteristics only (i.e., age, gender, and TIV). We also investigated the impact of brain region-wise tSNR on the performance of predictive models using rsfMRI features. We wanted to examine to what extent brain regions with low tSNR contribute to better predictions.

We found that the TC and FC characteristics of rsfMRI had comparable predictive capacity. The removal of age, gender, and TIV from the features or targets resulted in reduced performance. In most cases, the sole use of these individual characteristics yielded the best prediction accuracy. The accuracy of the predictions improved marginally when using a combination of rsfMRI

features and individual characteristics. In line with previous works, our results show that age and gender could be predicted much more accurately than cognitive phenotypes. Our findings demonstrate that rsfMRI features are not necessarily better than individual characteristics at predicting cognitive phenotypes, even when the sample size is increased to large numbers.

Results

Quantifying rsfMRI complex dynamics and cognitive phenotypes We used preprocessed rsfMRI data from 20,000 unrelated UK Biobank participants for this study [54] (see Methods). We calculated four TC measures and five ROI-wise FC-derived measures (see Methods). As prediction targets, we chose the four most reliable cognitive phenotypes in the UK Biobank database, measuring fluid intelligence, processing speed, visual memory, and numerical memory [55]. See Table S1 in the Supplementary Materials for the full list of features and targets.

We used kernel ridge regression with l2-norm regularization for predictive modeling, a widely used prediction method [5], [56], [57], to assess rsfMRI markers' cognitive phenotypic prediction ability. Model performance was measured through cross-validation using the Pearson correlation between the real and predicted targets (kernel ridge regression) or the balanced accuracy for gender classification (ridge classification). Model hyper-parameterization was done using nested cross-validation. Individual characteristics, i.e., age, gender, and TIV, were addressed through four scenarios, outlined in Figures 1-B.1 to B.4 (see also Methods).

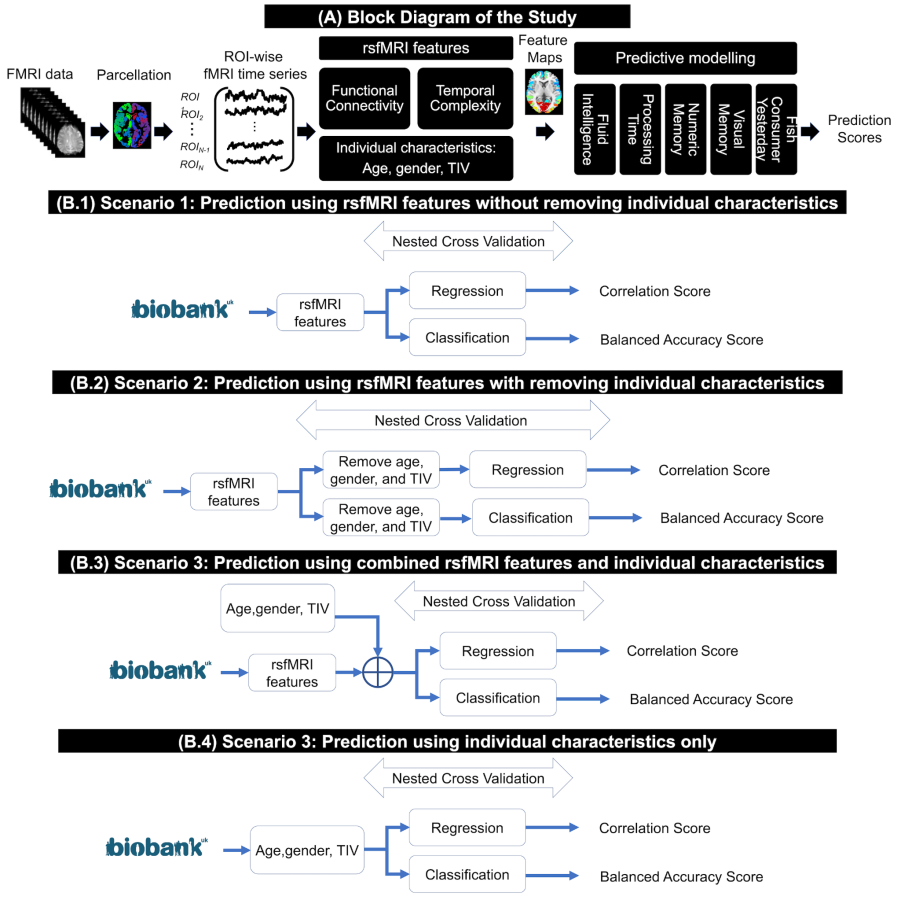

Figure 1: (A): main block diagram of this study, including the rsfMRI features and the prediction targets from the UK Biobank. (B) Four analysis scenarios based on the role of individual characteristics, i.e., age, gender, and total intracranial volume (TIV), in cognitive phenotypic prediction.

Larger sample sizes increase accuracy but eventually reach a plateau. First, we examined whether increasing the sample size could improve the prediction accuracy of cognitive phenotypes in all four scenarios (Figure 1). As illustrated in Figure 2, increasing the number of subjects improved accuracy most of the time, but the performance curves reached a plateau when using approximately more than 2,000 participants.

As a sanity check, we tested all the predictive modeling scenarios using fish consumption (the day prior to fMRI) as a target presumably unrelated to the rsfMRI features. The performance for all sample sizes, rsfMRI features,

individual characteristics, and their combinations remained at chance level (Figure 2).

Temporal complexity and FC features show comparable predictive capacities. Next, we investigated how TC and FC measures compare in cognitive phenotype prediction across different sample sizes. The average performance of ridge regression models suggested that certain features, specifically *fALFF*, *LCOR*, *wPE*, and $RangeEn_B$, performed better than others in all contexts, regardless of the target. Both types of features were situated at the lower and upper bands of prediction accuracy. The correlation between actual and predicted targets remained below 0.35 even at the maximum sample size. Voxel-based local brain activity measures of *fALFF* and *LCOR* showed the highest predictive capacity among the FC measures. Among TC measures, *wPE* and $RangeEn_B$ resulted in the highest accuracy, comparable to *fALFF* and *LCOR*.

Even with a sample size of 20,000 individuals, not all cognitive phenotypes could be predicted with equal accuracy (Figure 2). Processing speed was predicted with the highest correlation coefficient of up to 0.35, followed by fluid intelligence with 0.25 when using *fALFF* together with the individual characteristics. For all combinations, predictions for visual memory and numeric memory scores were less accurate with a correlation accuracy of less than 0.2. The prediction accuracy of processing speed was again higher than that of the other three cognitive phenotypes when using only the individual characteristics (age, gender, and TIV) (Scenario 4, see Figure 1). However, as shown in the black colored curves of Figure 2, the predictability of fluid intelligence, visual memory, and numerical memory scores was close to each other. Worthy to note that in all cases, removing the individual characteristics from cognitive phenotypes worsened the predictive performance.

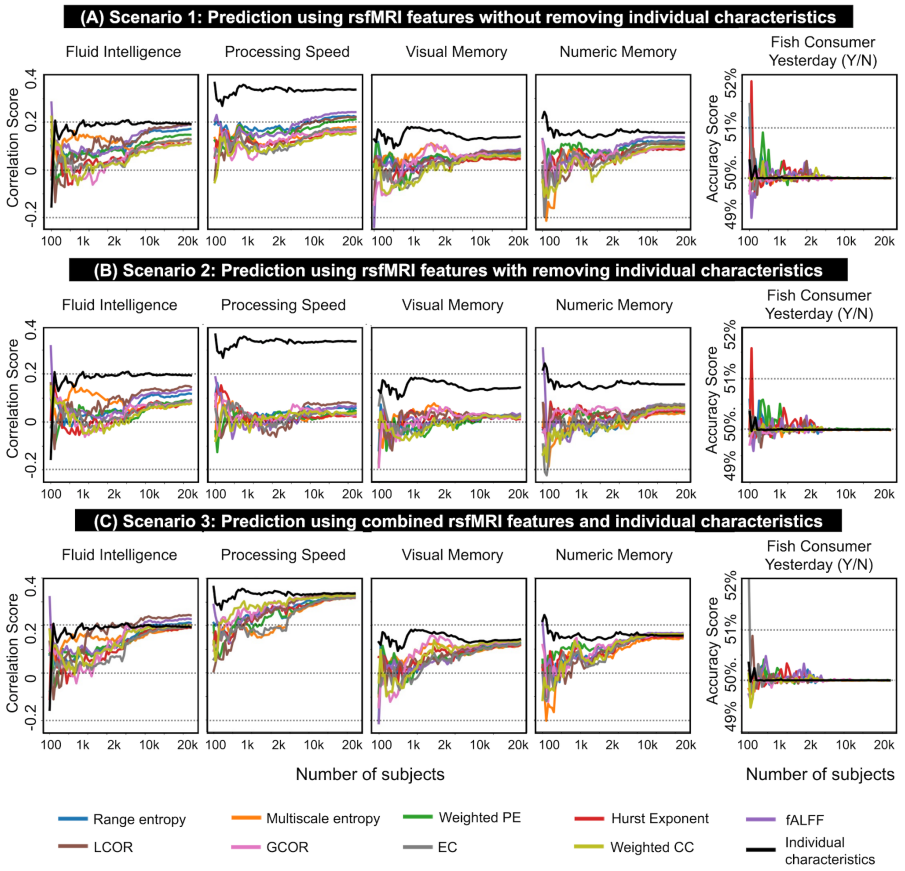

Figure 2: Prediction accuracy scores associated with nine rsfMRI features and five prediction targets using scenarios 1–3 of this study (see also Figures 1-B.1–B.3 and Methods). The prediction accuracies of individual characteristics only (Scenario 4 in Figure 1-B.4) have been plotted in bold black on all panels. Prediction accuracies of the fluid intelligence, processing speed, visual memory, and numeric memory scores are computed as the Pearson correlation between the actual values and predicted values through kernel ridge regression modeling. The prediction accuracy of Fish consumer yesterday is computed as the balanced accuracy through ridge binary classification. Each rsfMRI feature is illustrated in a distinct color and listed in the figure legend. In each figure panel, the x-axis represents the population size in the analysis, and the y-axis shows the prediction accuracy. The predictive modeling of each pair of features and targets is repeated for different sample sizes in the UK Biobank, ranging from $N_{subject}$ = 100 to $N_{subject}$ = 20,000. The population sizes from 100 to 2000 were increased with a 50-step increment (see the light orange shadow in the figure panels) and from 2000 to 20,000 with a 500-step increment (see the light blue shadow in the figure panels). As the panels show, the prediction accuracy improved with increasing sample size and the number of suprathreshold ROIs (i.e., higher accuracy toward the upper right corner of the color-coded maps). See Supplementary Figure S1 for the boxplot representation of these results.

Age, gender, and TIV result in higher accuracy than rsfMRI features. Next, we tested how age, gender, and TIV predict cognitive performance when used

as sole input features and without any rsfMRI data involved. To this end, we used these individual characteristics as input to the kernel ridge regression to predict cognitive phenotypes (Figure 1-D.1). As shown in Figure 2 and Supplementary Figure S1, this approach resulted in the highest correlation between actual and predicted targets across all sample sizes, outperforming all scenarios where rsfMRI features were utilized (Figure 1-D.1 to D.3). When individual characteristics served as input features, the sample size required to reach the plateau was substantially lower (lower than 500 subjects; see Figure 2). In other words, the ability of individual characteristics to predict cognitive phenotypes from a small sample size was better than the ability of rsfMRI features to predict the same targets, even when a larger sample size was used.

Given that the individual characteristics outperformed rsfMRI features in predicting cognitive phenotypes, the next logical step was to combine the TC and FC features with individual characteristics and see if it improved the prediction accuracy. For all rsfMRI features, this scenario produced the highest prediction accuracy of the first three analysis scenarios using rsfMRI features (Figure 2C, see also Supplementary Figure S1). This result shows that the inclusion of individual characteristics such as age, gender, and TIV may improve the performance of rsfMRI features, in particular for large sample sizes. The distinction between combined rsfMRI features and individual characteristics (Scenario 3) and rsfMRI features only (scenarios 1 and 2) was more pronounced when predicting processing speed in comparison to the other three cognitive phenotypes. Additionally, when the rsfMRI features were combined with individual characteristics and with larger sample sizes, the prediction accuracy was more similar (Figure 2, Scenario 1 versus Scenario 3).

The temporal signal-to-noise ratio plays no major role. We then asked if background noise in rsfMRI data affects prediction performance. To this end, we investigated whether excluding brain regions with high noise levels would increase prediction accuracy. We used a group-level tSNR map to threshold the rsfMRI feature maps (see Methods). Figure 3 illustrates the prediction accuracies for fluid intelligence using rsfMRI features (Scenario 1), rsfMRI features after removing individual characteristics (Scenario 2), and when combining rsfMRI features with individual characteristics (Scenario 3), following stepwise thresholding on the tSNR maps from 0% (no threshold, corresponding to results in Figure 2) to 60% with 5% increments. Prediction accuracies improved with increasing sample size and the number of suprathreshold ROIs. The results were similar for the other cognitive phenotypes (see Supplementary Figures S3–S5). Prediction accuracy for fish consumption remained at chance-level for all tSNR thresholds (Supplementary Figure S6).

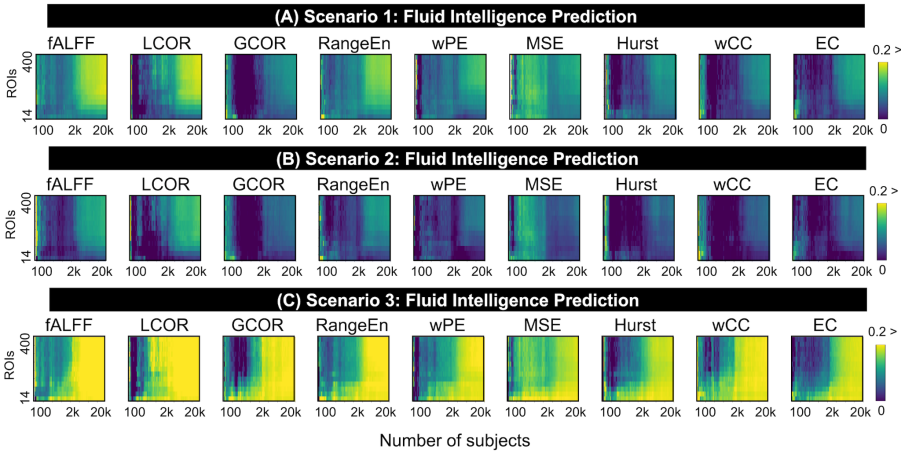

Figure 3: Pearson correlation accuracy associated with kernel ridge regression modeling of *fluid intelligence* using nine rsfMRI features and after tSNR thresholding from 0% (no threshold) to 65%. In each figure panel, the accuracy values are color-coded. Additionally, the x-axis represents the population size in the analysis, and the y-axis shows the number of suprathreshold ROIs after tSNR thresholding. The predictive modeling of each pair of features and targets is repeated for different sample sizes in the UK Biobank, ranging from $N_{subject}$ = 100 to $N_{subject}$ = 20,000. The population sizes from 100 to 2000 were increased with a 50-step increment, and from 2000 to 20,000 with a 500-step increment.

Age and gender are easier to predict than cognitive phenotypes. We investigated the capability of the rsfMRI features to predict individual characteristics. Compared to the prediction of four cognitive phenotypes, the prediction accuracy of all rsfMRI features was higher for both age and gender prediction (Figure 4). Two TC features, *wPE* and $RangeEn_B$, performed best at large sample sizes, as well as *fALFF* and *LCOR* (FC), with correlation coefficients of up to 0.5. This accuracy was considerably better than the prediction accuracy of cognitive phenotypes, which was typically less than 0.25 (see Figure 2). This result was noticeably different when individual characteristics were used as features for predictive modeling (gender and TIV for age prediction, and age and TIV for gender prediction). Gender could be classified using age and TIV with 88% accuracy. However, the individual characteristics did not perform well in age prediction, with a Pearson correlation of 0.2 between actual and predicted values.

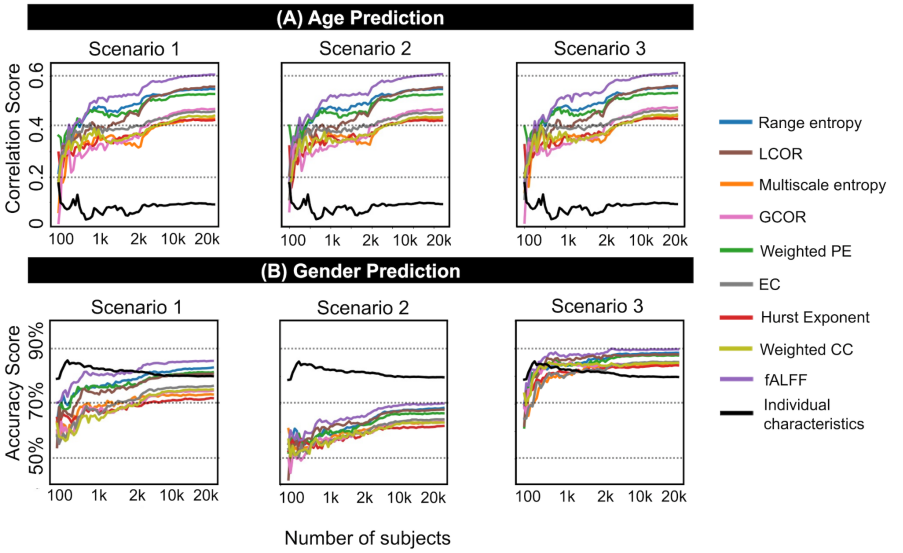

Figure 4: Prediction accuracy scores associated with nine rsfMRI features and age and gender as targets using scenarios 1–3 of this study (see also Figures 1-B.1–B.3 and Methods). The prediction accuracies of individual characteristics only (Scenario 4 in Figure 1-B.4) have been plotted in bold black on all panels. For age prediction, we considered gender and TIV as confounds, while for gender prediction, we considered age and TIV as confounds. Age prediction accuracies are computed as the Pearson correlation between the actual values and predicted values through kernel ridge regression modeling. Gender prediction accuracies are computed as the balanced accuracy through ridge binary classification. Each rsfMRI feature is illustrated in a distinct color and listed in the figure legend. In each figure panel, the x-axis represents the population size in the analysis, and the y-axis shows the prediction accuracy. The predictive modeling of each pair of features and targets is repeated for different sample sizes in the UK Biobank, ranging from $N_{subject}$ = 100 to $N_{subject}$ = 20,000. The population sizes from 100 to 2000 were increased with a 50-step increment (see the light orange shadow in the figure panels) and from 2000 to 20,000 with a 500-step increment (see the light blue shadow in the figure panels). See Supplementary Figure S2 for the boxplot representation of these results.

Similar individual patterns across rsfMRI features. We looked into how much information rsfMRI features with comparative prediction capacity share with one another. Our findings show that some rsfMRI features have comparable predictive capacity, despite their mathematical definitions and interpretations being quite different. For instance, *fALFF* and *wPE* were frequently among the most predictive features across three analysis scenarios, despite describing different aspects of rsfMRI. To check how well different rsfMRI features match with each other, we quantified the similarity between them using the *identification accuracy* score (see Methods). A number of rsfMRI feature pairs showed a high level of match (Figure 5). The pairs *wCC-EC*, *wPE-RangeEn*$_B$, *fALFF-LCOR*, and *MSE-HE* were among the most highly matched. The identification accuracy changed when individual characteristics were removed or when rsfMRI features were added to individual characteristics

(Figure 5, panels B, C, and D). Importantly, identification accuracy decreased as the number of subjects increased. This was in contrast to the increase in prediction accuracy (Figure 2).

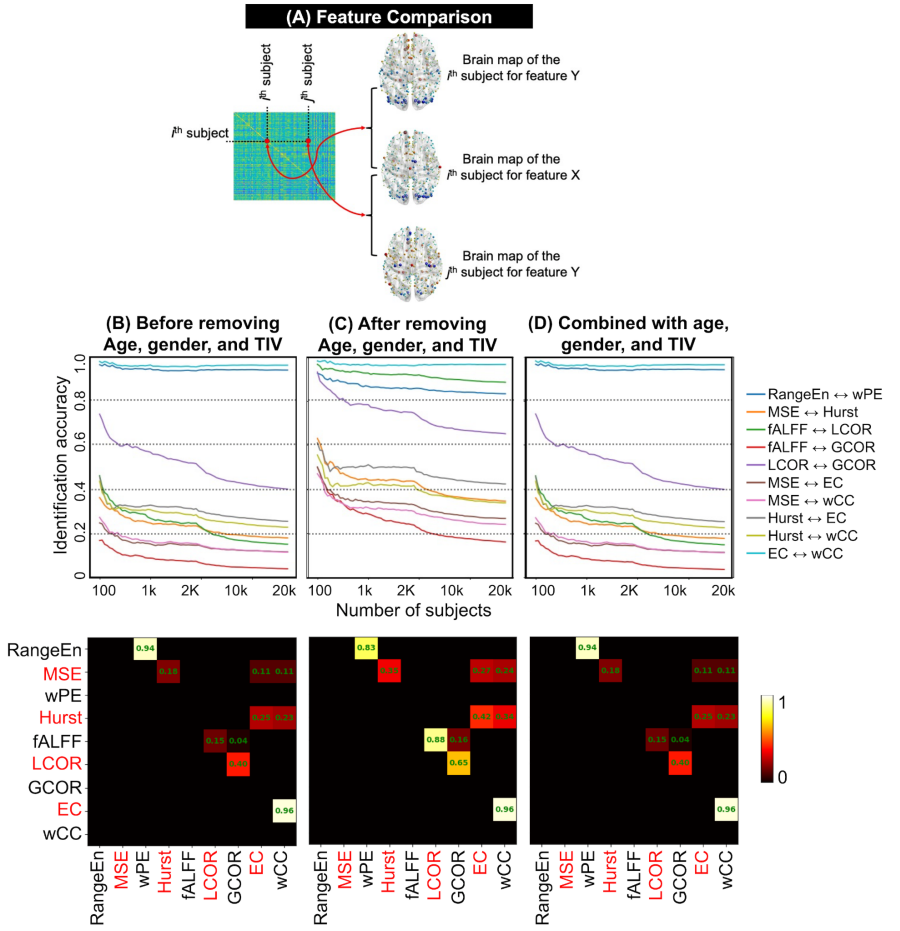

Figure 5: The process and results of rsfMRI feature comparison. (A) A schematic example of comparing two rsfMRI features *X* and *Y* from the same subject in a sample. This comparison leads to the computation of an identification accuracy score (see Methods). (B-D) Identification accuracy patterns of 10 rsfMRI feature pairs with above zero matching are associated with three analysis scenarios of this study (see Figure 1 as well as Methods). Each pair in the middle row panels has been depicted in a distinct color, and all pairs are listed in the figure legend. In each figure panel, the x-axis represents the population size in the analysis, and the y-axis shows the identification accuracy. The identification analyses are repeated for different sample sizes in the UK Biobank, ranging from $N_{subject}$ = 100 to $N_{subject}$ = 20,000. The population sizes from 100 to 2000 were increased with a 50-step increment (see the light orange shadow in the figure panels) and from 2000 to 20,000 with a 500-step increment (see the light blue shadow in the figure panels). The color-coded matrices in the row illustrate the identification accuracy of rsfMRI feature pairs for $N_{subject}$ = 20,000.

Discussion

A primary goal of neuroscience research is to investigate the relationship between complex brain dynamics and individual differences in behavior [58]. Spontaneous fluctuations in blood oxygenation level-dependent (BOLD) changes measured by fMRI have been shown to exhibit complex dynamics in the time domain referred to as TC [60], [61]. The interactions between BOLD changes across brain areas, also known as FC, provides useful perspectives on brain activity at a large scale. It has been demonstrated that these functional interactions are crucial for accomplishing tasks and are related to cognitive phenotypes [62].

The goal of this study was to investigate the use of TC/FC properties of spontaneous BOLD changes to predict cognitive phenotypes and to learn more about the effects of individual characteristics. To this end, we looked at how four cognitive phenotypes, fluid intelligence, processing speed, and memory characteristics, can be predicted by various aspects of TC/FC in rsfMRI. We first demonstrated that, despite having different mathematical definitions, the TC and FC metrics of rsfMRI lead to comparable performance across a wide range of sample sizes [45]. It naturally follows that a more thorough examination and a wider variety of rsfMRI features are required for deriving a more concrete conclusion.

Comparing MRI modalities for cognitive prediction has been the subject of several recent studies [63]–[65]. However, most of these studies have utilized the same datasets, primarily the widely used Human Connectome Project (HCP) database [66], with a medium sample size of fewer than 1500 participants. Cognitive prediction studies that use the UK Biobank database have a lot more participants than HCP, but as far as the authors know, only a few studies with large fMRI datasets have been done so far [22], [56], [57]. This makes extrapolating the findings of low sample size studies to larger sample sizes challenging [5]. Additionally, it may result in the problem of *dataset decay*, which means that repeated statistical tests on the same dataset in different studies may result in a rise in false positives [67]. Studies on reproducible brain-wide associations have also been established to require the involvement of thousands of participants [23]. To take these issues into account, we used a sizable portion of the UK Biobank database [45], varied the population size from 50 to 20,000 subjects, and examined the effects of sample size scaling on the predictive capacity of rsfMRI TC and FC-derived measures for cognitive phenotypic prediction. Altering the sample size indeed has a significant impact on the predictive accuracy of rsfMRI features (Figures 2 and 4). Previous cognitive prediction studies using the UK Biobank and other rsfMRI databases also found that the accuracy of kernel ridge regression modeling reaches a plateau by increasing the sample size [56], [68]. This finding implies that

prediction accuracy derived from small populations may not be reliable and can lead to large variance in cross-validation [69]. The comparative predictive ability of the TC and FC features of rsfMRI (Figures 2 and 3) suggests that the complex dynamics of rsfMRI and spontaneous functional interaction between brain regions carry overlapping information about behavior and cognition rather than being two distinct facets of brain function [62]. Such a relationship has also been seen in the characteristics of neurological conditions such as epilepsy, which has been described as both a disorder of functional networks in the brain and an abnormality of its dynamics at the same time [70].

Depending on the analysis workflows of this study, we either used rsfMRI features, individual characteristics, or both as input features for the kernel ridge regression models with l2-norm regularization. Interestingly, the prediction accuracy using individual characteristics was higher than that of all rsfMRI features (Figure 2). This result is consistent with a previous study in the ADHD-200 Global Competition, which found that when using logistic classification for ADHD diagnostics, individual characteristic data (site of data collection, age, gender, handedness, performance IQ, verbal IQ, and full scale IQ) performed better than a variety of fMRI features [44]. One way to reconcile these findings is to retain age, gender, and anatomical brain properties in rsfMRI features when predicting cognitive phenotypes and diagnosing diseases rather than removing them as is currently common practice in the literature [71]. If supported by additional research and if prediction accuracy itself is the goal (as opposed to identifying the neuronal origins of the predictive signal), then it may also suggest that using demographic and structural MRI data rather than fMRI recordings is a simpler and more practical method for cognitive phenotypic prediction in humans. The three individual characteristics performed better than the combined feature vectors of size 403 and the 400 rsfMRI feature vectors. A possible technical reason for this is that the number of samples needed to estimate a model with a given level of accuracy rises exponentially with the number of features, i.e. *the curse of dimensionality* [72]. However, given our large sample size, this is unlikely to be a concern.

Previous studies have shown that some cognitive phenotypes can be predicted better than others using neuroimaging data [5]. This is supported by our prediction results (Figure 2 and Supplementary Figure S1), which show that regardless of the rsfMRI features used and the sample size, the processing speed measure was usually predicted better than the visual and numerical memory scores. A recent review of human fluid intelligence prediction using neuroimaging data has reported an average Pearson correlation of 0.15 with a $CI_{95\%}$ of [0.13, 0.17] across the fMRI literature [9]. This is confirmed by our fluid intelligence prediction results with a maximum correlation score of 0.23 using combined *LCOR* and individual characteristics and at very high sample sizes

(20,000 - see Figure 2). Contrary to cognitive phenotypes, age and gender were easier to predict using both TC and FC features, as shown by a comparison between the prediction accuracy curves (Figures 2 and 4 as well as the boxplots in Supplementary Figures S1 and S2). However, gender prediction using age and TIV was better than age prediction using gender and TIV (see the black-color curves in Figure 4). All four analysis workflows passed a sanity check using the chance-level prediction of yesterday's fish consumption (Figure 2 and Supplementary Figure S1).

In the analysis of biological systems, "noise" is a term used to describe arbitrary or irregular disturbances that hinder or prevent the transmission of true, biologically plausible information to the recorded biosignals. This issue is particularly apparent in fMRI features defined by bandpass filtering of fMRI time series, such as *fALFF* [46]. This challenge is not limited to the rsfMRI literature, but it also includes new conceptualizations of macroscopic and microscopic brain functional organization [74]. In fact, *what was once "noise" is now "signal," as* has frequently occurred in the history of biosignal analysis research [74]. The temporal signal-to-noise ratio (tSNR) is a metric for comparing the strength of an interest signal to the amount of background noise in the time domain [75]. The tSNR analysis results of our study (Figure 4, Supplementary Figures S3-S6) indicate that even the rsfMRI features of brain regions with a high level of noise, which are typically found in deeper areas of the brain and close to sinus cavities, still contain predictive information about cognition. It is corroborated by our observation that using more brain regions, even when their tSNRs are not very high, leads to higher accuracy. To retain the same number of brain regions across individuals, we used the group-mean tSNR map of the full sample with 20,000 UK Biobank individuals to threshold subject-specific rsfMRI feature brain maps. This is because tSNR brain maps do not always agree on the same brain regions across participants. We believe that the information in this group-mean tSNR map from a very large sample is so compressed and dimensionally reduced that any influence of data leakage would be minimal.

Our results show a relatively inverse association between identification accuracy and prediction accuracy across different sample sizes. While adding more subjects improved cognitive phenotypic prediction accuracy (Figure 2), doing so reduced identification accuracy (Figure 5). This shows that the identification problem becomes harder as the sample size grows because there are more chances of obtaining a match with another subject than with the self. On the other hand, the prediction problem becomes relatively easier for larger populations because more information is available for learning. Additionally, it is consistent with a recent study that found a similar dichotomy between the neural identity facets that best predict a person's behavior and cognition and

those that best distinguish them from other people [64]. Figure 5 also suggests that the nine investigated rsfMRI features can be categorized into various matched pairs. The matches between *wPE* and $RangeEn_B$, *HE* and *MSE*, *fALFF* and *LCOR*, and *wCC* and *EC*, are most notable. The similar prediction performance of these feature pairs can be partly explained by the measurements yielding similar individual-level patterns (as shown by the high identification accuracy), even though they are conceptually different. We have previously demonstrated spatial overlap between the distribution of sample entropy, the *MSE*'s fundamental unit, and the *HE* across cortical brain areas [52]. The identification accuracy of rsfMRI gets more reasonable at time scales greater than 200 sec [76]. In particular, the reliability of FC connections (as determined by intraclass correlation) is at its peak for longer time scales of almost 400 sec and above [76]. Although we did not examine how unique the rsfMRI features were in an FC fingerprinting sense [10], [76], we think that the length of the rsfMRI datasets in the UK Biobank (490 sec) is long enough to analyze identification using rsfMRI features.

In conclusion, rsfMRI TC measures show some potential for cognitive, phenotypic, and individual characteristic prediction. This capacity is comparable to the widely used rsfMRI FC features. However, the individual characteristics showed a higher capacity for cognitive prediction than the rsfMRI features. Having said that, there are several limitations that should be taken into account when interpreting the findings of this study. First, we only used one predictive algorithm, kernel ridge regression. Although ridge regression has been shown to be effective in similar applications in the literature, other models might capture different information. Second, even though many more variables, such as handedness and genetic factors, could influence the rsfMRI features, we only considered three individual characteristics in our predictive modeling. Third, we attempted to address the challenge of accurately quantifying cognitive phenotypes using some of the most reliable cognitive phenotypes available in the UK Biobank. Despite this, these quantitative scores might still be noisy and subject to oversimplification. Unfortunately, standardized normed scores that account for demographic factors such as age and gender are not available in the UK Biobank database. Fourth, we only included cortical areas in our analyses. Using subcortical and cerebellar areas may give a more complete picture. Taken together, our findings could aid future research in creating accurate, individualized predictive models for clinical, behavioral, and cognitive measurements.

Methods

Data and preprocessing. We used the rsfMRI data of 20,000 unrelated UK Biobank (UKB) participants after excluding subjects with mental and cognitive disorders (ICD10, category F), diseases of the nervous system (ICD10, category

G), and cerebrovascular diseases (ICD10, categories I60 to 69). Data management of the UKB datasets was performed using DataLad [77] on JURECA, a pre-exascale modular supercomputer operated by the Jülich Supercomputing Center at the Forschungszentrum Jülich, Germany. The duration of each rsfMRI scan was 6 minutes (490 time points), with a repetition time (*TR*) of 0.735 sec, an echo time (*TE*) of 39 msec, a voxel size of 2.4×2.4×2.4 mm, and a field of view of 88×88×64. The following procedure was performed on the rsfMRI datasets as part of a pipeline developed by the UK Biobank [54]: grand-mean intensity normalization of the entire 4D fMRI dataset by a single multiplicative factor; highpass temporal filtering using Gaussian-weighted least-squares straight line fitting with σ = 50 sec; echo planar imaging unwarping; gradient distortion correction unwarping; and structured artifact removal through independent component analysis (ICA), followed by an ICA-based X-noiseifier (ICA-FIX) [78]–[80]. No spatial or temporal smoothing was applied to the fMRI volumes. The preprocessed data files, referred to as "*filtered_func_data_clean.nii"*, were normalized to the MNI space using FSL's *applywarp* function with spline interpolation and parcellated using the Schaefer brain atlas into 400 regions of interest (ROIs) [81]. Since we needed a continuous fMRI time series for the extraction of TC features, we did not apply motion scrubbing. Finally, we considered age, gender, and TIV as individual characteristics in the analyses and incorporated them into four analysis scenarios illustrated in Figure 1. The TIV of each subject was extracted after brain extraction from the T1 image using the Computational Anatomy Toolbox (CAT12) for SPM [82].

Four cognitive phenotypes were selected as the predictive targets among the most reliable UK Biobank cognitive phenotypes, including fluid intelligence (data field 20016), processing speed (data field 20023), numeric memory (data field 20240), and visual memory (data field 399) [55]. Additionally, an unrelated binary target (fish consumption yesterday - data field 103140) was used as a sanity check of the rsfMRI features in the predictive modeling scenarios.

Temporal complexity features. *HE* [50] is used to determine whether a time series contains a *long-memory process*. It quantifies three different types of trends: (*i*) values between 0.5 and 1, indicating that the time series is complex and has long-range dependence; (*ii*) values less than 0.5, indicating that the time series is random and has short-range dependence; or (*iii*) a value close to 0.5, indicating that the time series is a random walk with no memory of the past. HE has been shown to be stable and reproducible across different fMRI datasets [83]. In this study, we estimated *HE* using the rescaled range analysis technique [50]. The *wPE* [51] is a modified version of permutation entropy [84], a robust metric of TC that captures order relations between time points in a signal and generates an ordinal pattern probability distribution using an

embedding dimension m and a time delay τ, where the former is the length of the patterns and the latter is a lag parameter denoting the number of time points to shift throughout the time series. Unlike its predecessor, *wPE* assigns a weight to each time point to take amplitude information into account. In this study, we used the parameters $m = 4$ and $\tau = 1$ and normalized the *wPE* values by dividing them by $log_2(m!)$ in order to get the numbers between 0 and 1. *RangeEn* offers two versions ($RangeEn_A$ and $RangeEn_B$) as modifications to approximative entropy [85] and sample entropy [86], respectively. A property of $RangeEn_B$ is that regardless of the nature of the signal dynamics, it always reaches 0 at its tolerance value of $r = 1$ [52]. In light of this, one can obtain a complete trajectory of signal dynamics in the r-domain using this measure. Therefore, we extracted it in this study from ROIwise rsfMRI and reduced its dimensionality by computing the area under each curve along the r-axis ($m = 2$). We have recently shown that range entropy is robust to variations in signal length [52], making it a viable option for relatively short-length time series such as rsfMRI. *MSE* is an extension of sample entropy that provides insights into the complexity of rsfMRI fluctuations over a range of time scales [53]. The measure returns a trajectory of sample entropy values across the time scales 1 to τ_{max}. We have already shown that *MSE* may be linked to higher-order cognition [36]. In this study, we chose the parameters $m = 2$, $r = 0.5$, and $\tau_{max} = 10$ for *MSE*. We then reduced its dimensionality by taking the area under their curves and dividing by τ_{max}.

Functional connectivity features. We computed the FC measures of rsfMRI at two spatial scales: (*i*) at the ROI level averaged time series (*EC*, *wCC*) and (*ii*) first at the voxel level, then averaged within the ROIs (*fALFF*, *LCOR*, *GCOR*). For the ROIwise measures, we characterized the FCs between every pair of voxels or ROIs in each rsfMRI dataset and extracted the connections using Pearson correlation between fMRI time series [49]. *GCOR* serves as a representative of brain-wide correlation properties and a voxel-level representation of node centrality [48]. *LCOR* measures voxel-level local coherence. It is defined as the average of the correlation coefficients between a voxel and its immediate surroundings (a Gaussian kernel with FWHM of 25 mm) [47]. Similar to *GCOR*, *LCOR* takes both the strength and sign of functional connections into consideration. The *fALFF* quantifies the contribution of low frequency fluctuations to the total frequency range within a given frequency band (here, 0.008-0.09 Hz [46]). While *GCOR and LCOR* assess the strength of interregional and local cooperation by measuring the temporal similarity between voxels, *fALFF* evaluates the amplitude of regional neuronal activity. For each subject, the voxelwise *GCOR*, *LCOR*, and *fALFF* brain maps were parcellated into 400 ROIs using a brain atlas [81]. *EC* is an ROI-based measure in our study that indicates the impact of an ROI on the functional brain network [49]. The eigenvector centrality of the i^{th} ROI corresponds to the i^{th} element in the eigenvector corresponding to the largest eigenvalue of the ROIwise functional connectome. *The wCC* quantifies how much the ROIs in the brain network

functionally cluster together. This metric is calculated as the ratio of all triangles in which the i^{th} ROI participates to all triangles that, theoretically, could be formed given the degree of the i^{th} ROI's involvement in the brain's functional network [49]. We chose to exclude the vectorized upper/lower triangle of the FC matrix from the FC feature sets (i.e., feature vectors of length 79,800), because it would greatly increase the degrees of freedom in the predictive models and would increase the risk of overfitting. Also, it would lead to a very high computational load for the analysis of large sample sizes. The list of rsfMRI features in this study is summarized in Table S1 in the Supplementary Materials.

tSNR analysis. We calculated the temporal signal-to-noise ratio (tSNR) for each brain region as the ratio between the mean and standard deviation of the rsfMRI time series at each ROI [75]. This led to a tSNR brain map for each participant, which we normalized over ROIs and later averaged across the entire UK Biobank population ($N_{subject}$ = 20,000). We used the group-average map for thresholding to exclude the *noisiest* brain regions at multiple tSNR levels. The thresholding levels were varied from no threshold (i.e., preserving all ROIs for the prediction) to 65%, resulting in 14 suprathreshold ROIs. See Figure 3 as well as Figures S1 to S4.

Predictive modeling. Following previous studies in the field [5], [56], [57], we chose to use kernel ridge regression with l2-norm regularization and classification for predictive modeling. As illustrated in Figures 1-B.1 to B.4, we designed four analysis scenarios based on the role of individual characteristics in predictive modeling. We trained 78 kernel ridge regression models for each cognitive phenotype on a wide range of UK Biobank subjects from $N_{subject}$ = 100 to $N_{subject}$ = 2000 with a 50-step increment and from $N_{subject}$ = 2000 to $N_{subject}$ = 20,000 with a 500-step increment and each tSNR level, resulting in a total number of 36504 models (9 features × 4 targets × 78 population sizes × 13 tSNR levels). We also trained 78 ridge binary classifiers using each rsfMRI feature to predict fish consumption yesterday (total number of models: 9126). In all cases, we estimated the best model hyperparameter λ of the ridge regression/classification over the following values: [0.001, 0.01, 0.1, 1, 5, 10, 100, 1000, 10000, 100000] through grid search. For the evaluation of prediction accuracy, we performed five repeats of 5-fold nested cross-validation using the scikit-learn [87] and Julearn (https://juaml.github.io/julearn/main/index.html) libraries in Python. For evaluation of the regression models, we computed Pearson's correlation coefficient between the actual targets and the model's predictions. To evaluate the binary classifications, we used balanced accuracy, taking into account any imbalance between the two classes. We repeated the predictive modeling for five targets (four cognitive phenotypes as well as fish consumption yesterday) and nine rsfMRI features at a range of sample sizes varying from 100 to 20,000.

At each sample size, we randomly sampled the data to contain an equal number of males and females. We developed four predictive modeling scenarios based on the role of personal information or the individual characteristics (age, gender, and TIV) in our study, as illustrated in Figure 1. These scenarios included (*B.1*) prediction using rsfMRI features before removing individual characteristics, (*B.2*) prediction using rsfMRI features after treating individual characteristics as confounds and removing them, (*B.3*) prediction using combined rsfMRI features and individual characteristics, and (*B.4*) prediction using individual characteristics only. Individual characteristics were regressed out at the target level for regression modeling [5], [7] and at the feature level for the classification analyses [42], [71] using linear regression. Confound removal was performed in a cross-validation consistent manner to avoid data leakage [42].

Feature comparison via identification analysis. We adapted the concept of *identification accuracy* from the functional connectome fingerprinting literature [10], [88] and applied it to comparing different rsfMRI features of the same subject across a population. In this context, "identification" refers to the process of identifying a rsfMRI feature vector (brain map) *X* having the highest spatial correlation with *Y,* one of the other eight rsfMRI feature maps across the entire population. In this study, identification accuracy was defined as the proportion of correctly identified individuals based on matching their two rsfMRI features. The score ranges between 0 and 1, with higher values indicating a better match. See Figure 5-A for a schematic example of comparing two rsfMRI features across a given sample.

Acknowledgments

This work was supported by the Schwerpunktprogramm (SPP2041), project number 454012190, EI 816 28-1 (Machine-learning on Brain Connectomics: Individual Prediction of Cognitive Functioning in Health and Cerebral Small Vessel Disease), and the Helmholtz Portfolio Theme “Supercomputing and Modelling for the Human Brain”. SBE acknowledges funding by the European Union’s Horizon 2020 Research and Innovation Program (grant agreements 945539 (HBP SGA3) and 826421 (VBC)) and the Deutsche Forschungsgemeinschaft (DFG, SFB 1451 & IRTG 2150). All data analyses for this study were performed on the JURECA pre-exascale modular supercomputer operated by the Jülich supercomputing center at Forschungszentrum Jülich, Germany.

References

[1] B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde, “Functional connectivity in the motor cortex of resting human brain using echo-planar mri,” *Magn. Reson. Med.*, vol. 34, no. 4, pp. 537–541, Oct. 1995, doi: 10.1002/mrm.1910340409.

[2] M. D. Fox and M. E. Raichle, "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging," *Nat. Rev. Neurosci.*, vol. 8, no. 9, Art. no. 9, Sep. 2007, doi: 10.1038/nrn2201.

[3] B. B. Biswal, "Resting state fMRI: a personal history," *NeuroImage*, vol. 62, no. 2, pp. 938–944, Aug. 2012, doi: 10.1016/j.neuroimage.2012.01.090.

[4] B. B. Biswal, J. V. Kylen, and J. S. Hyde, "Simultaneous assessment of flow and BOLD signals in resting-state functional connectivity maps," *NMR Biomed.*, vol. 10, no. 4–5, pp. 165–170, 1997, doi: 10.1002/(SICI)1099-1492(199706/08)10:4/5<165::AID-NBM454>3.0.CO;2-7.

[5] L. Q. R. Ooi *et al.*, "Comparison of individualized behavioral predictions across anatomical, diffusion and functional connectivity MRI," *NeuroImage*, vol. 263, p. 119636, Nov. 2022, doi: 10.1016/j.neuroimage.2022.119636.

[6] C. Sripada *et al.*, "Prediction of neurocognition in youth from resting state fMRI," *Mol. Psychiatry*, vol. 25, no. 12, Art. no. 12, Dec. 2020, doi: 10.1038/s41380-019-0481-6.

[7] J. Chen *et al.*, "Shared and unique brain network features predict cognitive, personality, and mental health scores in the ABCD study," *Nat. Commun.*, vol. 13, no. 1, Art. no. 1, Apr. 2022, doi: 10.1038/s41467-022-29766-8.

[8] R. Kong *et al.*, "Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion," *Cereb. Cortex*, vol. 29, no. 6, pp. 2533–2551, Jun. 2019, doi: 10.1093/cercor/bhy123.

[9] B. H. Vieira, G. S. P. Pamplona, K. Fachinello, A. K. Silva, M. P. Foss, and C. E. G. Salmon, "On the prediction of human intelligence from neuroimaging: A systematic review of methods and reporting," *Intelligence*, vol. 93, p. 101654, Jul. 2022, doi: 10.1016/j.intell.2022.101654.

[10] E. S. Finn *et al.*, "Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity," *Nat. Neurosci.*, vol. 18, no. 11, Art. no. 11, Nov. 2015, doi: 10.1038/nn.4135.

[11] S. M. Smith *et al.*, "A positive-negative mode of population covariation links brain connectivity, demographics and behavior," *Nat. Neurosci.*, vol. 18, no. 11, pp. 1565–1567, Nov. 2015, doi: 10.1038/nn.4125.

[12] S. Weis, K. R. Patil, F. Hoffstaedter, A. Nostro, B. T. T. Yeo, and S. B. Eickhoff, "Sex Classification by Resting State Brain Connectivity," *Cereb. Cortex*, vol. 30, no. 2, pp. 824–835, Mar. 2020, doi: 10.1093/cercor/bhz129.

[13] R. N. Pläschke *et al.*, "Age differences in predicting working memory performance from network-based functional connectivity," *Cortex J. Devoted Study Nerv. Syst. Behav.*, vol. 132, pp. 441–459, Nov. 2020, doi: 10.1016/j.cortex.2020.08.012.

[14] S. H. Hojjati, A. Ebrahimzadeh, and A. Babajani-Feremi, "Identification of the Early Stage of Alzheimer's Disease Using Structural MRI and Resting-State fMRI," *Front. Neurol.*, vol. 10, 2019, Accessed: Oct. 14, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fneur.2019.00904

[15] J. R. Younce *et al.*, "Resting-State Functional Connectivity Predicts STN DBS Clinical Response," *Mov. Disord.*, vol. 36, no. 3, pp. 662–671, 2021, doi: 10.1002/mds.28376.

[16] H. Lin *et al.*, "Brain connectivity markers in advanced Parkinson's disease for predicting mild cognitive impairment," *Eur. Radiol.*, vol. 31, no. 12, pp. 9324–9334, Dec. 2021, doi: 10.1007/s00330-021-08086-3.

[17] M. Krishnamurthy *et al.*, "Resting-state functional MRI for motor cortex mapping in childhood-onset focal epilepsy," *J. Neuroimaging*, vol. n/a, no. n/a, doi: 10.1111/jon.13030.

[18] M. Kowalczyk, A. Omidvarnia, T. Dhollander, and G. Jackson, "Dynamic analysis of

fMRI activation during epileptic spikes can help identify the seizure origin," *Epilepsia*, vol. To appear, 2020.

[19] S. Moguilner *et al.*, "Weighted Symbolic Dependence Metric (wSDM) for fMRI resting-state connectivity: A multicentric validation for frontotemporal dementia," *Sci. Rep.*, vol. 8, no. 1, Art. no. 1, Jul. 2018, doi: 10.1038/s41598-018-29538-9.

[20] E. H. I. Claeys, T. Mantingh, M. Morrens, N. Yalin, and P. R. A. Stokes, "Resting-state fMRI in depressive and (hypo)manic mood states in bipolar disorders: A systematic review," *Prog. Neuropsychopharmacol. Biol. Psychiatry*, vol. 113, p. 110465, Mar. 2022, doi: 10.1016/j.pnpbp.2021.110465.

[21] M. Khosla, K. Jamison, G. H. Ngo, A. Kuceyeski, and M. R. Sabuncu, "Machine learning in resting-state fMRI analysis," *Magn. Reson. Imaging*, vol. 64, pp. 101–121, Dec. 2019, doi: 10.1016/j.mri.2019.05.031.

[22] T. He *et al.*, "Meta-matching as a simple framework to translate phenotypic predictive models from big to small data," *Nat. Neurosci.*, vol. 25, no. 6, Art. no. 6, Jun. 2022, doi: 10.1038/s41593-022-01059-9.

[23] S. Marek *et al.*, "Reproducible brain-wide association studies require thousands of individuals," *Nature*, vol. 603, no. 7902, Art. no. 7902, Mar. 2022, doi: 10.1038/s41586-022-04492-9.

[24] G. Deco and M. Corbetta, "The Dynamical Balance of the Brain at Rest," *The Neuroscientist*, vol. 17, no. 1, pp. 107–123, Feb. 2011, doi: 10.1177/1073858409354384.

[25] K. J. Friston, "Functional and effective connectivity: a review," *Brain Connect.*, vol. 1, no. 1, pp. 13–36, 2011, doi: 10.1089/brain.2011.0008.

[26] D. R. Chialvo, "Emergent complex neural dynamics," *Nat. Phys.*, vol. 6, no. 10, Art. no. 10, Oct. 2010, doi: 10.1038/nphys1803.

[27] F. Kj, J. O, R. G, and T. R, "Nonlinear event-related responses in fMRI," *Magn. Reson. Med.*, vol. 39, no. 1, Jan. 1998, doi: 10.1002/mrm.1910390109.

[28] K. J. Friston, A. Mechelli, R. Turner, and C. J. Price, "Nonlinear Responses in fMRI: The Balloon Model, Volterra Kernels, and Other Hemodynamics," *NeuroImage*, vol. 12, no. 4, pp. 466–477, Oct. 2000, doi: 10.1006/nimg.2000.0630.

[29] L. Waschke, N. A. Kloosterman, J. Obleser, and D. D. Garrett, "Behavior needs neural variability," *Neuron*, vol. 109, no. 5, pp. 751–766, Mar. 2021, doi: 10.1016/j.neuron.2021.01.023.

[30] V. Zimmern, "Why Brain Criticality Is Clinically Relevant: A Scoping Review," *Front. Neural Circuits*, vol. 14, p. 54, Aug. 2020, doi: 10.3389/fncir.2020.00054.

[31] Z. Wang, Y. Li, A. R. Childress, and J. A. Detre, "Brain Entropy Mapping Using fMRI," *PLoS ONE*, vol. 9, no. 3, p. e89948, Mar. 2014, doi: 10.1371/journal.pone.0089948.

[32] N. W. Churchill *et al.*, "The suppression of scale-free fMRI brain dynamics across three different sources of effort: aging, task novelty and task difficulty," *Sci. Rep.*, vol. 6, no. 1, Art. no. 1, Aug. 2016, doi: 10.1038/srep30895.

[33] S. Gao, G. Mishne, and D. Scheinost, "Nonlinear manifold learning in functional magnetic resonance imaging uncovers a low-dimensional space of brain dynamics," *Hum. Brain Mapp.*, vol. 42, no. 14, pp. 4510–4524, 2021, doi: 10.1002/hbm.25561.

[34] G. N. Saxe, D. Calderone, and L. J. Morales, "Brain entropy and human intelligence: A resting-state fMRI study," *PLOS ONE*, vol. 13, no. 2, p. e0191582, Feb. 2018, doi: 10.1371/journal.pone.0191582.

[35] M. O. Sokunbi *et al.*, "Nonlinear Complexity Analysis of Brain fMRI Signals in Schizophrenia," *PLoS ONE*, vol. 9, no. 5, May 2014, doi:

10.1371/journal.pone.0095146.

[36] A. Omidvarnia, A. Zalesky, S. Mansour L, D. Van De Ville, G. D. Jackson, and M. Pedersen, "Temporal complexity of fMRI is reproducible and correlates with higher order cognition," *NeuroImage*, vol. 230, p. 117760, Apr. 2021, doi: 10.1016/j.neuroimage.2021.117760.

[37] M. Pedersen, A. Omidvarnia, E. K. Curwood, J. M. Walz, G. Rayner, and G. D. Jackson, "The dynamics of functional connectivity in neocortical focal epilepsy," *NeuroImage Clin.*, vol. 15, pp. 209–214, 2017, doi: 10.1016/j.nicl.2017.04.005.

[38] M. Nezafati, H. Temmar, and S. D. Keilholz, "Functional MRI Signal Complexity Analysis Using Sample Entropy," *Front. Neurosci.*, vol. 14, 2020, Accessed: Oct. 14, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2020.00700

[39] I. M. McDonough and K. Nashiro, "Network complexity as a measure of information processing across resting-state networks: evidence from the Human Connectome Project," *Front. Hum. Neurosci.*, vol. 8, Jun. 2014, doi: 10.3389/fnhum.2014.00409.

[40] A. D. Nostro *et al.*, "Predicting personality from network-based resting-state functional connectivity," *Brain Struct. Funct.*, vol. 223, no. 6, pp. 2699–2719, Jul. 2018, doi: 10.1007/s00429-018-1651-z.

[41] J. D. Power, A. Mitra, T. O. Laumann, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, "Methods to detect, characterize, and remove motion artifact in resting state fMRI," *NeuroImage*, vol. 84, p. 10.1016/j.neuroimage.2013.08.048, Jan. 2014, doi: 10.1016/j.neuroimage.2013.08.048.

[42] S. More, S. B. Eickhoff, J. Caspers, and K. R. Patil, "Confound Removal and Normalization in Practice: A Neuroimaging Based Sex Prediction Case Study," in *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track*, Cham, 2021, pp. 3–18. doi: 10.1007/978-3-030-67670-4_1.

[43] F. Alfaro-Almagro *et al.*, "Confound modelling in UK Biobank brain imaging," *NeuroImage*, vol. 224, p. 117002, Jan. 2021, doi: 10.1016/j.neuroimage.2020.117002.

[44] M. Brown *et al.*, "ADHD-200 Global Competition: diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements," *Front. Syst. Neurosci.*, vol. 6, 2012, Accessed: Oct. 22, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnsys.2012.00069

[45] K. L. Miller *et al.*, "Multimodal population brain imaging in the UK Biobank prospective epidemiological study," *Nat. Neurosci.*, vol. 19, no. 11, pp. 1523–1536, Nov. 2016, doi: 10.1038/nn.4393.

[46] Q.-H. Zou *et al.*, "An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: Fractional ALFF," *J. Neurosci. Methods*, vol. 172, no. 1, pp. 137–141, Jul. 2008, doi: 10.1016/j.jneumeth.2008.04.012.

[47] G. Deshpande, S. LaConte, S. Peltier, and X. Hu, "Integrated local correlation: a new measure of local coherence in fMRI data," *Hum. Brain Mapp.*, vol. 30, no. 1, pp. 13–23, Jan. 2009, doi: 10.1002/hbm.20482.

[48] Z. S. Saad *et al.*, "Correcting brain-wide correlation differences in resting-state FMRI," *Brain Connect.*, vol. 3, no. 4, pp. 339–352, 2013, doi: 10.1089/brain.2013.0156.

[49] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, Sep. 2010, doi: 10.1016/j.neuroimage.2009.10.003.

[50] H. E. Hurst, "Long-Term Storage Capacity of Reservoirs," *Trans. Am. Soc. Civ. Eng.*,

vol. 116, no. 1, pp. 770–799, 1951.

[51] B. Fadlallah, B. Chen, A. Keil, and J. Príncipe, "Weighted-permutation entropy: a complexity measure for time series incorporating amplitude information," *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, vol. 87, no. 2, p. 022911, Feb. 2013, doi: 10.1103/PhysRevE.87.022911.

[52] A. Omidvarnia, M. Mesbah, M. Pedersen, and G. Jackson, "Range Entropy: A Bridge between Signal Complexity and Self-Similarity," *Entropy*, vol. 20, no. 12, p. 962, Dec. 2018, doi: 10.3390/e20120962.

[53] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, no. 6, p. 068102, Aug. 2002, doi: 10.1103/PhysRevLett.89.068102.

[54] F. Alfaro-Almagro *et al.*, "Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank," *NeuroImage*, vol. 166, pp. 400–424, Feb. 2018, doi: 10.1016/j.neuroimage.2017.10.034.

[55] C. Fawns-Ritchie and I. J. Deary, "Reliability and validity of the UK Biobank cognitive tests," *PLoS ONE*, vol. 15, no. 4, p. e0231627, Apr. 2020, doi: 10.1371/journal.pone.0231627.

[56] T. He *et al.*, "Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics," *NeuroImage*, vol. 206, p. 116276, Feb. 2020, doi: 10.1016/j.neuroimage.2019.116276.

[57] M.-A. Schulz *et al.*, "Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets," *Nat. Commun.*, vol. 11, no. 1, Art. no. 1, Aug. 2020, doi: 10.1038/s41467-020-18037-z.

[58] D. S. Bassett and M. S. Gazzaniga, "Understanding complexity in the human brain," *Trends Cogn. Sci.*, vol. 15, no. 5, pp. 200–209, May 2011, doi: 10.1016/j.tics.2011.03.006.

[59] Z. J. Lau, T. Pham, S. H. A. Chen, and D. Makowski, "Brain entropy, fractal dimensions and predictability: A review of complexity measures for EEG in healthy and neuropsychiatric populations," *Eur. J. Neurosci.*, vol. 56, no. 7, pp. 5047–5069, 2022, doi: 10.1111/ejn.15800.

[60] P. CIUCIU, G. Varoquaux, P. Abry, S. Sadaghiani, and A. Kleinschmidt, "Scale-free and multifractal properties of fMRI signals during rest and task," *Front. Physiol.*, vol. 3, 2012, Accessed: Aug. 01, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fphys.2012.00186

[61] B. J. He, "Scale-Free Properties of the Functional Magnetic Resonance Imaging Signal during Rest and Task," *J. Neurosci.*, vol. 31, no. 39, pp. 13786–13795, Sep. 2011, doi: 10.1523/JNEUROSCI.2111-11.2011.

[62] R. Liégeois *et al.*, "Resting brain dynamics at different timescales capture distinct aspects of human behavior," *Nat. Commun.*, vol. 10, no. 1, Art. no. 1, May 2019, doi: 10.1038/s41467-019-10317-7.

[63] E. Dhamala, K. W. Jamison, A. Jaywant, S. Dennis, and A. Kuceyeski, "Distinct functional and structural connections predict crystallised and fluid cognition in healthy adults," *Hum. Brain Mapp.*, vol. 42, no. 10, pp. 3102–3118, 2021, doi: 10.1002/hbm.25420.

[64] S. Mansour L, Y. Tian, B. T. T. Yeo, V. Cropley, and A. Zalesky, "High-resolution connectomic fingerprints: Mapping neural identity and behavior," *NeuroImage*, vol. 229, p. 117695, Apr. 2021, doi: 10.1016/j.neuroimage.2020.117695.

[65] J. Rasero, A. I. Sentis, F.-C. Yeh, and T. Verstynen, "Integrating across neuroimaging modalities boosts prediction accuracy of cognitive ability," *PLOS Comput. Biol.*, vol. 17, no. 3, p. e1008347, Mar. 2021, doi: 10.1371/journal.pcbi.1008347.

[66] D. C. Van Essen *et al.*, "The Human Connectome Project: a data acquisition perspective," *NeuroImage*, vol. 62, no. 4, pp. 2222–2231, Oct. 2012, doi: 10.1016/j.neuroimage.2012.02.018.

[67] W. H. Thompson, J. Wright, P. G. Bissett, and R. A. Poldrack, "Dataset decay and the problem of sequential analyses on open datasets," *eLife*, vol. 9, p. e53498, May 2020, doi: 10.7554/eLife.53498.

[68] Z. Cui and G. Gong, "The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features," *NeuroImage*, vol. 178, pp. 622–637, Sep. 2018, doi: 10.1016/j.neuroimage.2018.06.001.

[69] G. Varoquaux, "Cross-validation failure: Small sample sizes lead to large error bars," *NeuroImage*, vol. 180, no. Pt A, pp. 68–77, Oct. 2018, doi: 10.1016/j.neuroimage.2017.06.061.

[70] A. T. Berg *et al.*, "Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009," *Epilepsia*, vol. 51, no. 4, pp. 676–685, Apr. 2010, doi: 10.1111/j.1528-1167.2010.02522.x.

[71] D. Chyzhyk, G. Varoquaux, M. Milham, and B. Thirion, "How to remove or control confounds in predictive models, with applications to brain biomarkers," *GigaScience*, vol. 11, p. giac014, Mar. 2022, doi: 10.1093/gigascience/giac014.

[72] M. Verleysen and D. François, "The Curse of Dimensionality in Data Mining and Time Series Prediction," in *Computational Intelligence and Bioinspired Systems*, Berlin, Heidelberg, 2005, pp. 758–770. doi: 10.1007/11494669_93.

[73] L. Wiersch *et al.*, "Accurate sex prediction of cisgender and transgender individuals without brain size bias." bioRxiv, p. 2022.07.26.499576, Jul. 28, 2022. doi: 10.1101/2022.07.26.499576.

[74] L. Q. Uddin, "Bring the Noise: Reconceptualizing Spontaneous Neural Activity," *Trends Cogn. Sci.*, vol. 24, no. 9, pp. 734–746, Sep. 2020, doi: 10.1016/j.tics.2020.06.003.

[75] K. Murphy, J. Bodurka, and P. A. Bandettini, "How long to scan? The relationship between fMRI temporal signal to noise and necessary scan duration," *NeuroImage*, vol. 34, no. 2, pp. 565–574, Jan. 2007, doi: 10.1016/j.neuroimage.2006.09.032.

[76] D. Van De Ville, Y. Farouj, M. G. Preti, R. Liégeois, and E. Amico, "When makes you unique: Temporality of the human brain fingerprint," *Sci. Adv.*, vol. 7, no. 42, p. eabj0751, Oct. 2021, doi: 10.1126/sciadv.abj0751.

[77] Y. O. Halchenko *et al.*, "DataLad: distributed system for joint management of code, data, and their relationship," *J. Open Source Softw.*, vol. 6, no. 63, p. 3262, Jul. 2021, doi: 10.21105/joss.03262.

[78] C. F. Beckmann and S. M. Smith, "Probabilistic independent component analysis for functional magnetic resonance imaging," *IEEE Trans. Med. Imaging*, vol. 23, no. 2, pp. 137–152, Feb. 2004, doi: 10.1109/TMI.2003.822821.

[79] G. Salimi-Khorshidi, G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti, and S. M. Smith, "Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers," *NeuroImage*, vol. 90, pp. 449–468, Apr. 2014, doi: 10.1016/j.neuroimage.2013.11.046.

[80] L. Griffanti *et al.*, "ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging," *NeuroImage*, vol. 95, pp. 232–247, Jul. 2014, doi: 10.1016/j.neuroimage.2014.03.034.

[81] A. Schaefer *et al.*, "Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI," *Cereb. Cortex N. Y. N 1991*, vol. 28, no. 9,

pp. 3095–3114, 01 2018, doi: 10.1093/cercor/bhx179.

[82] G. C, D. R, T. Pm, K. F, L. E, and undefined, "CAT – A Computational Anatomy Toolbox for the Analysis of Structural MRI Data," Jun. 2022, doi: 10.1101/2022.06.11.495736.

[83] O. L. Campbell and A. M. Weber, "Monofractal analysis of functional magnetic resonance imaging: An introductory review," *Hum. Brain Mapp.*, vol. 43, no. 8, pp. 2693–2706, 2022, doi: 10.1002/hbm.25801.

[84] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, no. 17, p. 174102, Apr. 2002, doi: 10.1103/PhysRevLett.88.174102.

[85] S. M. Pincus, "Approximate entropy as a measure of system complexity.," *Proc. Natl. Acad. Sci.*, vol. 88, no. 6, pp. 2297–2301, Mar. 1991, doi: 10.1073/pnas.88.6.2297.

[86] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Am. J. Physiol. Heart Circ. Physiol.*, vol. 278, no. 6, pp. H2039-2049, Jun. 2000, doi: 10.1152/ajpheart.2000.278.6.H2039.

[87] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python." arXiv, Jun. 05, 2018. doi: 10.48550/arXiv.1201.0490.

[88] E. Amico and J. Goñi, "The quest for identifiability in human functional connectomes," *Sci. Rep.*, vol. 8, no. 1, Art. no. 1, May 2018, doi: 10.1038/s41598-018-25089-1.

Supplementary Materials

Table S1 RsfMRI features and the prediction targets in this study

RsfMRI Feature name	**Category**	**Spatial resolution**
Weighted permutation entropy (*wPE*) [51]	Temporal complexity	ROIwise
Range entropy ($RangeEn_B$) [52]	Temporal complexity	ROIwise
Multiscale entropy (*MSE*) [53]	Temporal complexity	ROIwise
Hurst exponent (*HE*) [50]	Temporal complexity	ROIwise
Eigenvector centrality (*EC*) [49]	Functional connectivity	ROIwise
Weighted clustering coefficient (*wCC*) [49]	Functional connectivity	ROIwise
Fractional amplitude of low-frequency fluctuations (*fALFF*) [46]	Functional connectivity	Voxelwise
Local correlation (*LCOR*) [47]	Functional connectivity	Voxelwise
Global correlation (*GCOR*) [48]	Functional connectivity	Voxelwise
Prediction target	**UK Biobank Data field**	
Fluid intelligence	20016	
Processing time	20023	
Visual memory	399	
Numeric memory	20240	
Fish consumer yesterday	103140	

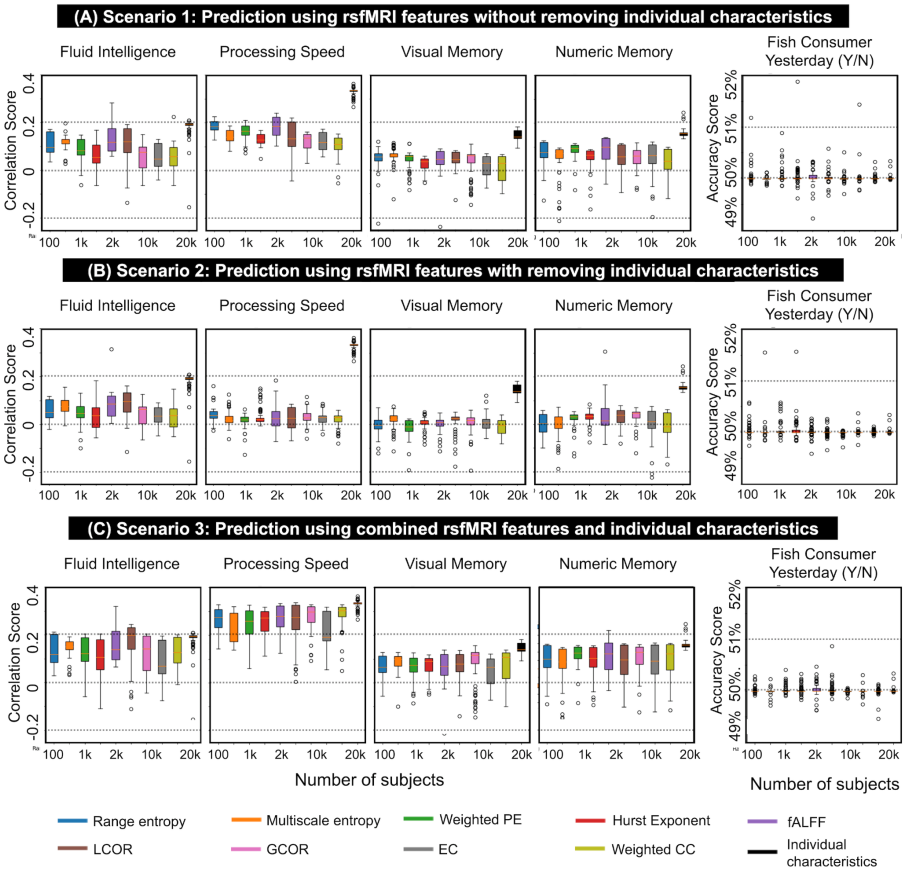

Figure S1: Prediction accuracy scores associated with nine rsfMRI features and five prediction targets using scenarios 1–3 of this study (see also Figures 1-B.1–B.3 and Methods). Prediction accuracies of the fluid intelligence, processing speed, visual memory, and numeric memory scores are computed as the Pearson correlation between the actual values and predicted values through kernel ridge regression modeling. The prediction accuracy of Fish consumer yesterday is computed as the balanced accuracy through ridge binary classification. Each rsfMRI feature is illustrated in a distinct color and listed in the figure legend. In each figure panel, the box has a line at the median and spans the complete range of sample sizes (from 100 to 20,000 participants), extending from the lower to upper quartile values of the prediction accuracies. The whiskers extend outside the box to display the data's range. The population sizes from 100 to 2000 were increased with a 50-step increment and from 2000 to 20,000 with a 500-step increment. See Figure 2 for the representation of prediction accuracies over the range of sample sizes.

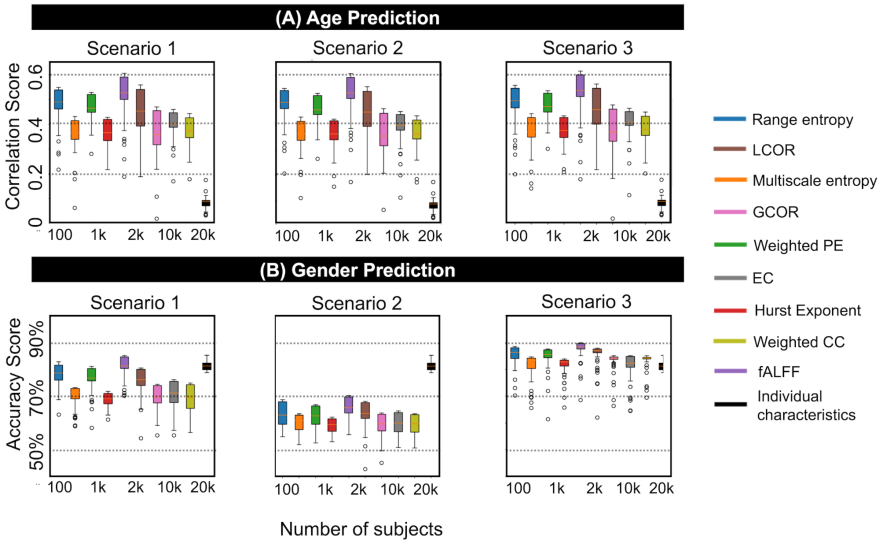

Figure S2: Prediction accuracy scores associated with nine rsfMRI features and age and gender as targets using scenarios 1–3 of this study (see also Figures 1-B.1–B.3 and Methods). For age prediction, we considered gender and TIV as confounds, while for gender prediction, we considered age and TIV as confounds. Age prediction accuracies are computed as the Pearson correlation between the actual values and predicted values through kernel ridge regression modeling. Gender prediction accuracies are computed as the balanced accuracy through ridge binary classification. Each rsfMRI feature is illustrated in a distinct color and listed in the figure legend. In each figure panel, the box has a line at the median and spans the complete range of sample sizes (from 100 to 20,000 participants), extending from the lower to upper quartile values of the prediction accuracies. The whiskers extend outside the box to display the data's range. The population sizes from 100 to 2000 were increased with a 50-step increment and from 2000 to 20,000 with a 500-step increment. See Figure 4 for the representation of prediction accuracies over the range of sample sizes.

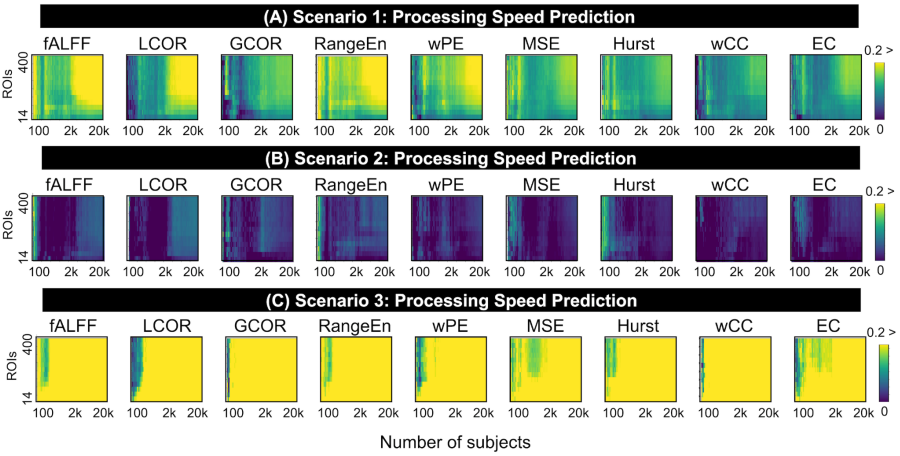

Figure S3: Pearson correlations associated with kernel ridge regression modeling of the *processing speed* score using nine rsfMRI features and after tSNR thresholding from 0% (no threshold) to 65%. In each figure panel, the accuracy values are color-coded. Additionally, the x-axis represents the population size in the analysis, and the y-axis shows the number of suprathreshold ROIs after tSNR thresholding. The predictive modeling of each pair of features and targets is repeated for different sample sizes in the UK Biobank, ranging from $N_{subject}$ = 100 to $N_{subject}$ = 20,000. The population sizes from 100 to 2000 were increased with a 50-step increment, and from 2000 to 20,000 with a 500-step increment.

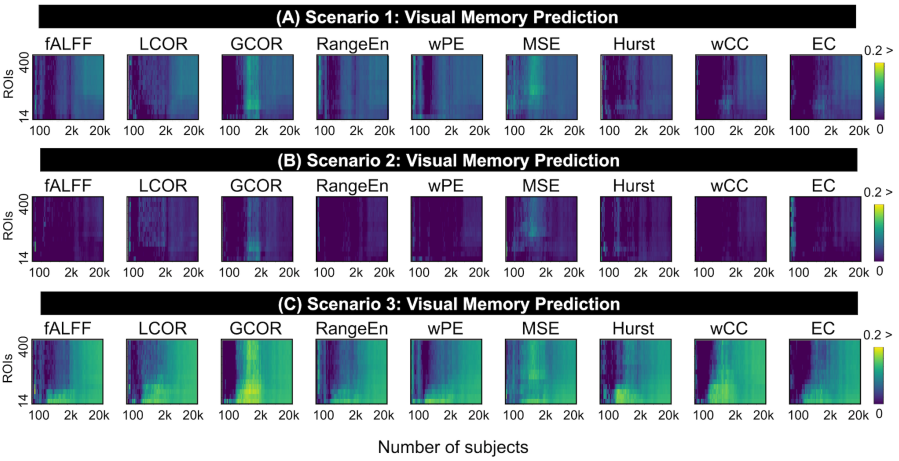

Figure S4: Pearson correlation accuracies associated with kernel ridge regression modeling of the *visual memory* score using nine rsfMRI features and after tSNR thresholding from 0% (no threshold) to 65%. In each figure panel, the accuracy values are color-coded. Additionally, the x-axis represents the population size in the analysis, and the y-axis shows the number of suprathreshold ROIs after tSNR thresholding. The predictive modeling of each pair of features and targets is repeated for different sample sizes in the UK Biobank ranging from $N_{subject}$ = 100 to $N_{subject}$ =

20,000. The population sizes from 100 to 2000 were increased with a 50-step increment, and from 2000 to 20,000 with a 500-step increment.

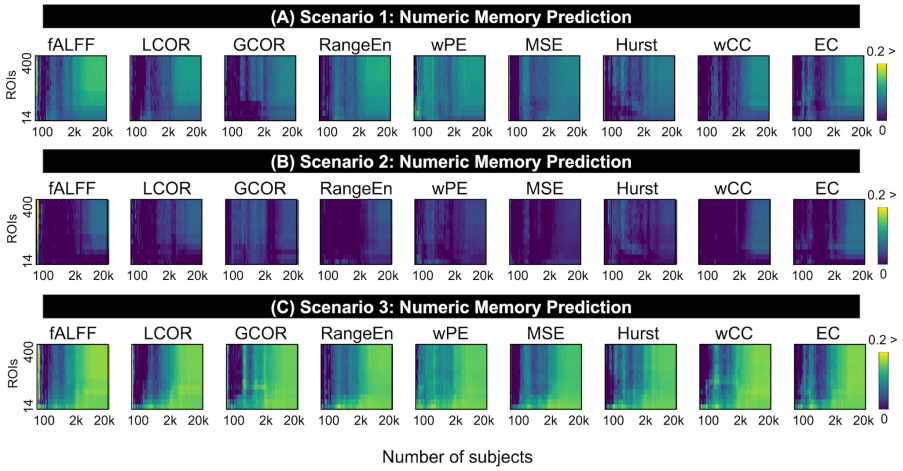

Figure S5: Pearson correlation accuracies associated with kernel ridge regression modeling of the *numeric memory* score using nine rsfMRI features and after tSNR thresholding from 0% (no threshold) to 65%. In each figure panel, the accuracy values are color-coded. Additionally, the x-axis represents the population size in the analysis, and the y-axis shows the number of suprathreshold ROIs after tSNR thresholding. The predictive modeling of each pair of features and targets is repeated for different sample sizes in the UK Biobank, ranging from $N_{subject}$ = 100 to $N_{subject}$ = 20,000. The population sizes from 100 to 2000 were increased with a 50-step increment, and from 2000 to 20,000 with a 500-step increment.

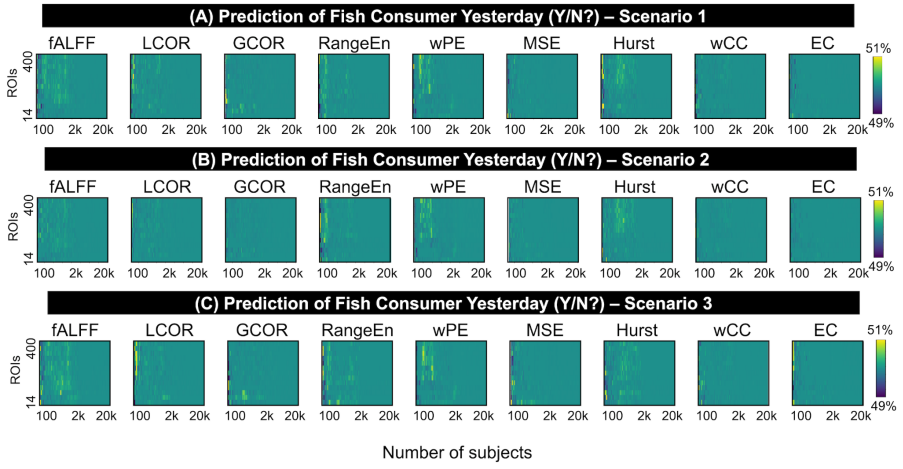

Figure S6: Balanced accuracy associated with ridge binary classification of *fish consumption yesterday* using nine rsfMRI features and after tSNR thresholding from 0% (no threshold) to 65%.

In each figure panel, the accuracy values are color-coded. Additionally, the x-axis represents the population size in the analysis, and the y-axis shows the number of suprathreshold ROIs after tSNR thresholding. The predictive modeling of each pair of features and targets is repeated for different sample sizes in the UK Biobank, ranging from $N_{subject}$ = 100 to $N_{subject}$ = 20,000. The population sizes from 100 to 2000 were increased with a 50-step increment, and from 2000 to 20,000 with a 500-step increment.