# Weak and Unstable Prediction of Personality from the Structural Connectome

Amelie Rauland[1,2], Kyesam Jung[2,3], Theodore D. Satterthwaite[4,5,6], Matthew Cieslak[4,5], Kathrin Reetz[7,8], Simon B. Eickhoff[2,3], Oleksandr V. Popovych[2,3]

[1] Department of Psychiatry, Psychotherapy and Psychosomatics, RWTH Aachen, Aachen, Germany
[2] Institute of Neuroscience and Medicine, Brain and Behaviour (INM-7), Research Centre Jülich, Jülich, Germany
[3] Institute for Systems Neuroscience, Medical Faculty, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany
[4] Lifespan Informatics and Neuroimaging Center, University of Pennsylvania Perelman School of Medicine, Philadelphia Pennsylvania, United States
[5] Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia Pennsylvania, United States
[6] Penn-CHOP Lifespan Brain Institute, Philadelphia Pennsylvania, United States
[7] Department of Neurology, RWTH Aachen University, Aachen, Germany
[8] JARA-BRAIN Institute of Molecular Neuroscience and Neuroimaging, Forschungszentrum Jülich GmbH and RWTH Aachen University, Aachen, Germany

Keywords: structural connectome, diffusion weighted MRI, big five personality traits, machine learning prediction analysis, cognition, individual differences

## Abstract

Personality neuroscience aims to discover links between personality traits and features of the brain. Previous neuroimaging studies have investigated the connection between the brain structure, microstructural properties of brain tissue or the functional connectivity (FC) and these personality traits. Analyses relating personality to diffusion weighted MRI measures were limited to investigating the voxel-wise or tract-wise association of microstructural properties with trait scores. We expanded past work in two ways by focusing on the entire structural connectome (SC) and by predicting personality trait scores instead of performing a statistical correlation analysis in order to assess an out-of-sample performance and determine whether there is an individually predictive relationship between the brain SC and the big five personality traits. Prediction of personality from the SC is however not yet as established as prediction of behavior from the FC, and sparse studies in this field so far delivered rather heterogeneous results. We therefore further dedicated our study to investigate if and how different pipeline settings influence prediction performance. In a sample of 426 unrelated subjects with high quality MRI acquisitions from the Human Connectome Project, we analyzed 19 different brain parcellations, three SC weightings, three groups of subjects and four feature classes for the prediction of the five personality traits using a ridge regression. From the large number of evaluated pipelines, only very few lead to promising results of prediction accuracy $r > 0.2$, while the vast majority lead to a small prediction accuracy centered around zero. A markedly better prediction was observed for a cognition target confirming the chosen methods for SC calculation and prediction and indicating limitations of the personality trait scores and their relation to the SC. We therefore report, that for methods evaluated here, the SC cannot predict personality trait scores. Overall, we found that all considered pipeline conditions influence the predictive performance of both cognition and personality trait scores and should be chosen carefully according to the problem at hand. The strongest differences were found

for the trait openness and the SC weighting by number of streamlines which outperformed the other traits and weighting respectively.

# 1 Introduction

Personality strongly affects interindividual differences of human behavior as well as personal strengths and vulnerabilities influencing all aspects of life [1]. Personality neuroscience therefore aims to explore the neurobiological basis of personality. One of the most generally recognized and comprehensive models of personality is the Five-Factor Model (FFM), also known as the big five personality traits [2]. It describes five major dimensions of personality as openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. Scores for these dimensions are acquired from subjects by completion of self-report inventories.
From the neurobiological side, diffusion-weighted magnetic resonance imaging (dwMRI) measures the directional diffusion of water molecules in the brain [3] and enables both the *in-vivo* mapping of white matter (WM) tracts using tractography algorithms [4] and the estimation of microstructural properties of the WM [5]. It therefore offers a diverse and novel insight into the human brain in addition to structural and functional MRI and opens up further options to investigate neural correlates of personality accordingly. Here, we aim to investigate if there is an individual predictive relationship between the structural connectome (SC) of the brain calculated from dwMRI acquisitions and the big five personality trait scores.

Apart from prior research investigating neural correlates of personality extracted from structural or functional MRI (e.g. [6], [7]), there has been previous work relating structural properties of the brain WM extracted from dwMRIs with the FFM personality traits. The majority of these studies used a technique called tract-based spatial statistics (TBSS) [8]. TBSS is a voxel-wise analysis of multi-subject diffusion data to relate different microstructural properties of the WM, which can be calculated using e.g. diffusion tensor imaging (DTI), to a certain target – here, personality trait scores. Most commonly, fractional anisotropy (FA) and mean diffusivity (MD) were estimated for the analysis [9-16]. FA describes the fraction of the diffusion tensor that can be assigned to directionally dependent diffusion, and MD describes the average magnitude of diffusion in all directions [5]. However, results in these studies were quite heterogeneous. Leshem et al. [9], Jung et al. [10] and Wilhelms et al. [11] solely investigated the relationship with one trait (extraversion (Leshem), openness (Jung) and neuroticism (Wilhelms)) in different samples, and all reported statistically significant correlations for clusters of voxels and the respective personality trait. Other studies investigated all of the big five personality traits and reported statistically significant correlations only for a subset of traits. For example, microstructural properties of several WM voxel clusters were found to correlate only with openness, neuroticism and agreeableness [12], neuroticism [13], conscientiousness [14], agreeableness and conscientiousness [15], or with all traits except openness [16]. Furthermore, a recent study of Avinun et al. [17] reported null-findings for the relationship between mean FA across the brain and personality for all of the big five personality traits.
Alongside the voxel-wise analysis, other approaches applied special tractography algorithms to track and select major and known WM tracts for each subject (see Ref. [18] for the specific

algorithm used in the presented approaches) and calculate the average dwMRI microstructural properties along these tracts. The evaluated track properties can then be related to personality. While McIntosh et al. [19] investigated the relationship between the mean FA of 12 WM tracts and the traits extraversion and neuroticism, Lewis et al. [20] studied how the FA of 10 WM tracts related with openness, agreeableness, and conscientiousness. Both studies found statistically significant correlations with at least one tract for each of the considered personality traits. However, in another study Privado et al. [21] only found a connection between openness and the FA of some of the extracted tracts. The reported results on the relationships between dwMRI data and personality traits thus remain very sparse and heterogeneous also for the tract-wise analyses.
The tract-based analysis of dwMRI data can be extended to the whole brain, where the entire structural connectome (SC) can be calculated for each subject based on the whole-brain tractography (WBT) and a predefined parcellation of the gray matter (GM) [22]. In the first approximation, the SC is represented by an adjacency matrix, where each entry indicates whether any two regions of interest (ROIs) of the GM parcellation are connected with each other or not. The connectivity can then be defined in many different ways by weighting the SC adjacency matrix by using, for example, the number of streamlines (NOS) between two ROIs. Another SC weighting can also be the mean of any microstructural property, e.g., MD or FA of all voxels that the streamlines connecting two ROIs pass through. So far, only a few studies have related graph metrics of local and global efficiencies and clustering coefficient of the SC to personality traits of extraversion, neuroticism and openness to experience ([23], [24]).

Overall, the available results on the relationships between dwMRI data and personality traits are rather fragmentary, heterogeneous and controversial, which calls for additional and more systematic investigations of this problem for sufficiently large subject cohorts and high-quality neuroimaging data. All of the above studies have in common that they correlated dwMRI measures with personality trait scores and therefore do not offer any insight into the generalizability of their findings. On the other hand, predicting trait scores using machine learning (ML) algorithms in a cross-validation (CV) setting can, in contrast to statistical correlation analysis, give an indication on the out-of-sample performance and how generalizable certain findings and connections are ([25], [26]). Further, the relations between the SC itself and all five personality traits have so far not been thoroughly investigated and still await a systematic investigation. However, compared to the functional connectome (FC), the SC has been studied less as a predictor of individual phenotypes in general. In the process of calculating the SC from the diffusion MRI acquisition and predicting phenotypes from (the features of) the SC, there are many parameters that can each be adjusted in different ways. It has not yet been thoroughly investigated how the choice of brain parcellation, connectome weighting, feature selection of prediction and CV algorithm may influence the prediction of one or another personality trait for one or another subject group, e.g., males or females. While studies looking into these parameters did find differences between distinct settings, there are no clear best practices and parameter choices as there is a complex interplay between the different settings ([27], [28], [29]).

In this work, we therefore aimed to extend available results on the relationship between personality traits and dwMRI in two ways by making use of the entire SC instead of using a voxel-wise or tract-based analysis and by applying a predictive CV method instead of a statistical correlation analysis for an assessment on the out-of-sample performance. Considering that the SC has been studied less as a predictor of behavior than the FC, we

further dedicated our study to evaluate a large number of different pipelines for the prediction of personality trait scores. We varied the brain atlas, the SC weighting, the subject group and the feature class as well as feature selection process in order to unfold potential connections between the structural connectivity of the brain and personality and determine how they relate to different settings of the prediction process.

# 2 Materials and Methods

In this study we investigated whether and how personality trait scores can be predicted from the structural connectome (SC) calculated from diffusion-weighted MRI (dwMRI) data. To address this question, we considered a large number of pipeline conditions including different data processing and weightings, subject samples, feature classes and feature selections within classes as well as prediction targets. We considered many different pipeline conditions given that there are no established best practices yet for setting these conditions when predicting from the SC. Additionally, previous studies relating dwMRI measures to personality using voxel-wise or tract-based analysis showed heterogeneous results (see Section 1) and evaluating many different pipeline conditions enabled us to provide a broader picture on if and how personality and the SC are related. Overall, we compared the prediction results for 19 different cortical parcellations, three SC weightings, four feature classes - three of them with additional conditions of feature selection - and three subject groups to predict five different personality traits. All these selections resulted in a total of 3,420 (19 x 3 x 4 x 5 x 3) different prediction pipelines that were repeatedly evaluated and compared for all considered cases of feature selection. An overview of the considered settings is schematically illustrated in Fig. 1.

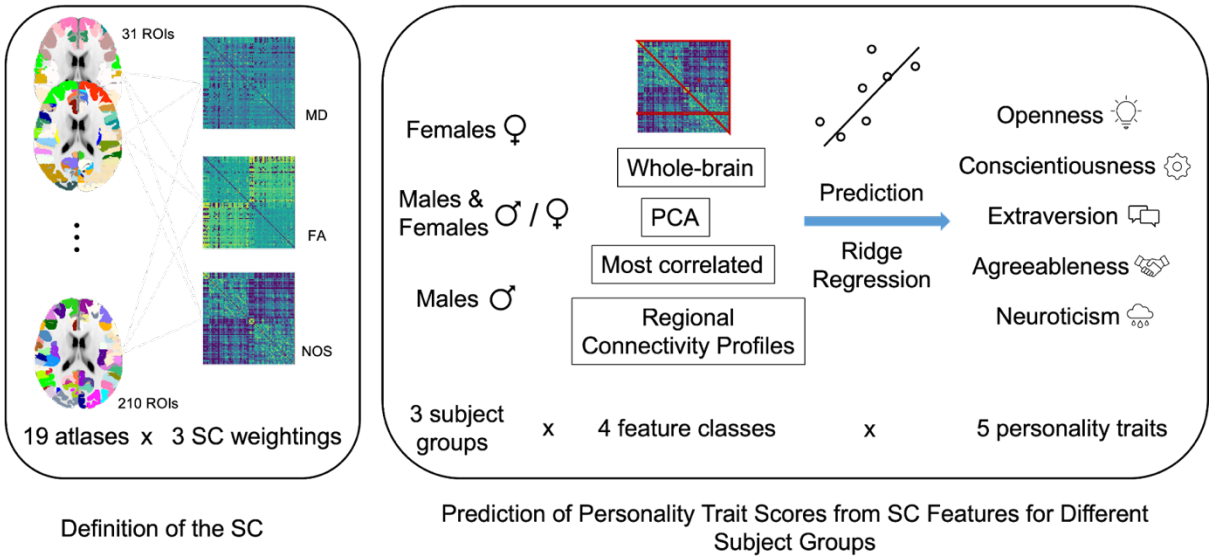



*Fig. 1* **Overview of different options that were evaluated for predicting personality trait scores from the structural connectome (SC).** For every subject, 19 parcellations were used to calculate SC from dwMRI for three different SC weightings given by the mean diffusivity (MD), fractional anisotropy (FA) and number of streamlines (NOS). The predictions of 5 personality traits were separately performed by ridge regression for 3 subject groups and 4 feature classes that included several conditions of feature selection like number of features or a given region of used parcellation (see text for details).

We used the 19 different state-of-the-art cortical parcellations as described in Ref. [30]. The parcellations were selected in a way to balance between parcellations derived from functional data and parcellations derived from structural data and provide a large variety of granularities from 31 to 210 regions of interest (ROI).

For the connectome weighting, we considered the number of streamlines (NOS) connecting any two brain regions, calculated from the whole-brain tractography for a given parcellation [4]. We further employed the mean diffusivity (MD) and the fractional anisotropy (FA) [5] as connectome weighting. MD and FA are microstructural measures describing diffusion

properties of the underlying tissue, where MD describes the average magnitude of diffusion in all directions, and FA describes the degree of directionality of the diffusion. A known limitation of streamline tractography is that the density of reconstructed connections is not necessarily closely reflecting the density of the underlying WM fibers as the number of reconstructed streamlines depends, e.g., on the distance between two ROIs, their size and the shape / curvature of the streamlines [31]. To correct for this, the spherical-deconvolution-informed filtering of tractograms (SIFT2) algorithm was developed to obtain a more quantitative measure for the connection density by calculating weights for each streamline based on the fiber orientation distribution (FOD) [32]. Therefore, in addition to the NOS-weighted SCs, we also investigated NOS-weighted SCs after the SIFT2 refinement and compared these two weightings for the larger 'males and females' subject group shown in Fig. 1. All other pipeline conditions (parcellation, feature classes and personality traits) were varied as for the case of NOS weighting.

Further, we considered three subject groups, only females, only males and a mixed-sex subject group. Such a setup of subject groups was chosen to evaluate if there is an effect of sex on the personality prediction by SC. This consideration was motivated by previous findings, where an influence of sex on brain-personality correlations has been shown for GM [7] volume and the resting-state functional connectivity [33].

For prediction, features need to be selected from the SC. The most strait forward form feature class considers the entire SC (whole-brain). Depending on the granularity of the parcellation scheme, the number of features in this feature class in most cases exceeded the number of subjects in our dataset. This may lead to problems with overfitting and therefore motivated us to consider three additional feature classes that selected / calculated a much smaller number of features from the SC (details in Section 2.4).

## 2.1 MRI Data and Subject Groups

The present study considered structural and diffusion MRI data from 560 healthy subjects from the Human Connectome Project (HCP) S1200 release dataset [34]. The dataset includes diffusion imaging data for 972 individuals. However, there are many siblings and twins within the dataset, and we here decided to only use groups of unrelated subjects to not obscure results by possible family effects. The local ethics committee of the HCP WU-Minn Consortium gave its approval for the study and written informed consent was obtained from all subjects.

We created three subject groups of unrelated subjects. The first group consisted of n=426 unrelated subjects (213 females, age 22-36 years, mean 28.5±3.7). This corresponds to the highest number of unrelated subjects with diffusion data and personality trait scores available preserving an equal ratio of males and females. The second group consisted of n=278 (age 22-36 years, mean 29.3±3.7) unrelated males, and the third group was made up of n=278 (age 22-37 years, mean 28.0±3.8) unrelated females. There are overlaps between the mixed-sex group and the all-male and all-female group, respectively. In total, considering the overlap between groups, data from 560 subjects was considered.

All data was acquired with a 3T Siemens PRISMA scanner. The T1-weighted (T1w) images were acquired at 0.7mm isotropic resolution, the diffusion weighted images (DWI) at 1.25mm isotropic resolution with 90 gradient directions at each of three shells with b-values of 1000, 2000 and 3000 s/mm$^2$ and a total of 18 b=0 scans. The high-quality diffusion acquisition from a total of 270 directions on three different shells, enables a more accurate estimation of the

diffusion tensor and the microstructural properties derived from it, as well as the application of advanced methods for the estimation of fiber orientation distribution functions (fODFs) as a basis for tractography. For both T1w and DWI data, we used the minimally pre-processed images provided with the HCP dataset described in detail in Ref. [35].

## 2.2 SC Calculation

For each subject we calculated 76 SCs based on the 19 different parcellations and the three different connectome weightings based on NOS, MD and FA as well as on the SIFT2-refined weighting (Fig. 1). All SCs were based on the same whole-brain tractography (WBT) for each respective subject. To calculate the WBT and extract the SC matrices, we used the pre-processed T1w and DWI data provided with the HCP dataset and an in-house pipeline optimized for parallel processing on high-performance computational clusters. The preprocessing included intensity normalization across runs, *topup* and *eddy* corrections and gradient nonlinear correction. A detailed description of the HCP pre-processed data can be found in Ref. [35].

The pre-processed images were used for co-registration between the T1w and DWI spaces by FSL [36], as well as for the estimation of linear and non-linear transformation matrices from the standard MNI space to the native T1w space and vice versa. GM, WM, cortical/subcortical, cerebellar, and cerebrospinal fluid masks were generated in the DWI space as part of the registration process. Further, MD and FA images were calculated from the DWIs for each subject using the *dwi2tensor* and *tensor2metric* functionalities of MRtrix3 [37]. After registration, the WBT was calculated. This part of the pipeline only used MRtrix3 functions. Shell- and matter-specific response functions were estimated using the *dwi2response dhollander* algorithm implemented in MRtrix3. Fiber orientation distribution functions (fODF) were then calculated using the multi-shell-multi-tissue constrained spherical deconvolution (MSMT-CSD) [38] in each voxel of the DWIs. Subsequently, the WBT was calculated using an anatomically constrained probabilistic fiber tracking algorithm with second-order integration. Anatomically constrained tracking algorithms discard streamlines ending in WM or CSF to obtain a more realistic set of streamlines [39]. The WBT density was set to 10M streamlines, and other tracking parameters were set as follows based on the recommended default values in MRtrix3: step size = 0.625 mm, angle = 45 degrees, minimal length = 2.5 mm, maximal length = 250 mm, FOD amplitude for terminating tract = 0.06, maximum attempts per seed = 1000, maximum number of sampling trials = 1000, and downsampling = 3.

For all atlases, only the cortical ROIs were selected. The brain atlas images were sampled in the volumetric MNI152 nonlinear 6th generation standard space included in the FSL software package [36]. To calculate the SC matrix from the WBT for a given parcellation, the atlas images were transformed to the native diffusion space for each subject using the pre-calculated transformation matrices. Then the labeled voxels within the GM mask were selected as seed and target regions to define streamlines connecting any two regions in the parcellation. For the SC weighted by streamline count (NOS-weighted), the number of streamlines connecting any two regions was entered in the corresponding cells of the connectivity matrix. For the SC weighted by MD and FA, additional steps were necessary. In the first step, the mean FA / MD per streamline was determined by taking the mean over all voxels that a respective streamline passes through. Then, to determine the final weights of the SC matrix, the mean FA and MD values were averaged across all streamlines connecting

any two regions of a given parcellation. For the SIFT2-weighted connectomes, a weight was first determined for each streamline using the *tcksift2* function in MRtrix3 [37] such that the density of reconstructed streamlines more accurately represents the density of underlying fibers**.** These weights were subsequently summed up across all respective streamlines between any two regions resulting in a SIFT2-weighted connectome. For all SC matrices the diagonals containing self-connections were set to zero.

## 2.3 Personality Trait Scores

One of the most generally recognized and comprehensive models of personality is the Five-Factor Model (FFM) also known as the big five personality traits [2]. It describes five major dimensions of personality as openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. To obtain a score for each of the big five personality traits for each subject, all selected subjects filled out the English version of the 60-item version of the Neuroticism/Extraversion/Openness Five Factor Inventory (NEO-FFI) by McCrae and Costa [40] as part of the Penn Computerized Cognitive Battery [41]. There were 12 items for each of the five traits and all items were answered on a Likert scale ranging from 0 (strong disagreement) to 4 (strong agreement). The total score for each trait was obtained by adding up the responses from the separate items. Items keyed inversely were reversed before computation so that the total score for each trait was in the range [0, 48]. Additional information on the trait scores, such as their distributions and correlations of the traits between each other can be found in Supplementary Fig. S1.

## 2.4 Feature Classes

Four feature classes were considered and derived from the SC for prediction of the personality trait scores (Fig. 1).

1. *Whole-Brain*: The first feature class was the entire upper triangle of the SC matrices (without diagonal). As the SC is symmetric, this included information from all structural connections considered in the respective parcellation.
2. *Most correlated (corr) edges*: The second class of features included a certain number *k* of SC edges that were most correlated with the personality traits across subjects. The selection of the most correlated edges was performed for the training set only and then applied to the test set to prevent data leakage.
3. *Principal component analysis (PCA):* The third feature class was composed of *k* first principal components (PC) as determined by a PCA of the SC matrices explaining most of the variance of the SC across subjects. To prevent data leakage, the PCA was fit on the training set only.
4. *Regional connectivity profiles (RCP)*: This feature class corresponded to separate rows of the SC reflecting the connectivity profile of a given brain region to the rest of the brain. For this case, every single row of the SC was evaluated as features.

Depending on the granularity of the parcellation scheme, the number of features in the whole-brain class in most cases exceeded the number of subjects in our dataset, which may lead to overfitting. This motivated us to consider the other feature classes above that contained a much smaller number of features, and that we compared with each other with respect to their prediction performance. For both the PCA and corr feature class, 19 different cases of feature selection were evaluated corresponding to different number of features *k* included in the

analysis: $k \in \{1, 2, 3, 4, 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, \#ROIs\}$. The latter number #ROIs corresponded to the number of brain regions of a given parcellation according to its granularity and was added to the feature selection cases of the corr- and PCA-classes in order to obtain the same number of features used in the RCP class. The RCP class also evaluated different cases of feature selection as each row of the SC was evaluated for the prediction. The number of different feature selection cases of the RCP feature class therefore depended on the number of ROIs of the considered parcellation.

## 2.5 Prediction

Taking into account the findings from literature concerning the impact of used prediction algorithms on the prediction performance ([28], [42], [27]) (see Discussion) and the computational burden of the different prediction models, we considered a simple Ridge regression model for the prediction of personality trait scores from SC features. Ridge regression is a linear model that penalizes the regression coefficients based on their l2-norm and is defined as in Eq. 1.

$$argmin_{\beta} \sum_{i} (y_i - \beta' x_i + b)^2 + \alpha \sum_{k=1}^{K} \beta_k^2 \qquad \textit{Eq. 1}$$

A nested 5-fold cross validation (CV) was employed to tune the hyperparameter alpha of the ridge regression in the inner loop and assess the prediction performance on unseen subject samples in the outer loop. Alpha determines the strength of the l2-regularization, and the following values were evaluated in the inner loop of the nested CV: α in [0.001, 0.01, 1, 10, 50, 100, 500, 1000, 5000, 10000]. The prediction performance was measured in terms of Pearson's correlation between the predicted and empirical personality trait scores of the subjects in the test sets of the outer 5-fold CV loop. Then all five correlation values of the five different test sets were averaged such that a single prediction result was reported for a given CV procedure. The entire prediction procedure was repeated 100 times for different random splits of the data, which resulted in a distribution of prediction results (correlations) for each constellation of other conditions mentioned above. The 100 different random splits of the data were consistent across pipelines.

The SC features used for prediction were normalized before being used as input to the ridge regression using a global min-max normalization. The global maximum and minimum feature values from the training set were selected to scale all features to the range [0,1] as described in Eq. 2:

$$x_{norm} = \frac{x - train_{min}}{train_{max} - train_{min}}, \qquad \textit{Eq. 2}$$

where x is the feature used for training and prediction, and $train_{min}$ and $train_{max}$ are the minimum and maximum feature values calculated of the training set. For the PCA feature class, the normalization was applied to the SC before fitting the PCA on the data. For the other methods - whole-brain, most correlated edges, and RCPs - the features were first extracted from the SC, and only the selected features were then normalized and used to determine the minimum and maximum of the training set. As the distribution of the NOS was very skewed and stretched towards large NOS values, a $\log_{10}$-function was applied to scale all edges with values > 0 before the normalization was used to obtain a more symmetric and condensed NOS distribution without affecting small values too much.

## 2.6 Comparison of Feature Classes

To compare the different feature selection methods (feature classes), for any specific pipeline of atlas – SC weighting – subject group – feature class – feature selection – target trait, we selected the best result from all evaluated feature selection scenarios. For example, for the pipeline MIST31 – NOS – mixed sex – PCA – openness, we ran the pipeline for 19 different numbers of PCs. To compare this pipeline to the same pipeline with another feature class, we chose the number of PCs that led to the best prediction results, i.e., the largest mean prediction accuracy of the test set averaged over 100 random data splits used for CV procedure. The same was applied to the feature classes of the most correlated SC edges and the RCPs, where we chose the best prediction results with respect to the number of correlated edges and appropriate brain region, respectively. Using this method, for each feature class, we only had one distribution of test correlations from 100 random data splits that represented the obtained upper boundary of prediction accuracy for this feature class. This selection process did not have to be applied for the whole-brain feature class as there was only one option (whole SC) for the feature selection.

As the best number of PCs, best number of most correlated features and best RCP were chosen based on the prediction performance on the test set, these values represented an idealized performance estimate for each of the feature classes and might be overfit to this specific data set. For this reason, we performed additional calculations using another approach for the best feature selection. In this case these variables were independently optimized in the inner CV loop of the nested CV, together with the regularization parameter alpha of the ridge regression. This resulted in only one distribution of prediction accuracy values calculated for the test set by the CV procedure for each specific pipeline with fixed atlas, SC weighting, subject sample, feature class and prediction target. These prediction results can then be compared between the pipelines, for example, between feature classes. We only ran such a feature optimization on the mixed-sex subject group.

## 2.7 Prediction Brain Maps

For the RCP feature class, we can assign a prediction performance to every brain region of a given parcellation. Indeed, each case of the feature selection from the RCP feature class corresponds to selecting and using one row of the SC matrix for prediction, i.e., the connectivity profile of one specific ROI (parcellated brain region) to the rest of the brain. Then the obtained prediction accuracy was assigned to all voxels of the selected region. For each parcellation we iterated through all ROIs and created the whole-brain prediction map of that parcellation. Such a mapping procedure was repeated for all considered parcellations, and the obtained prediction maps were averaged across parcellations. This was done separately for all traits, SC weightings and subject groups, which resulted in a total of 45 brain overlay maps reflecting the contribution of one or another group of voxels to the prediction of personality traits.

## 2.8 Prediction of Cognition

To obtain context to the results for the personality prediction, we additionally predicted a cognition variable provided with the HCP S1200 dataset. We chose the age-adjusted total composite score from the NIH Toolbox Cognitive Function Battery [41], which includes scores

from picture vocabulary, reading, flanker, dimensional change card sort, picture sequence memory, list sorting and pattern comparison tests and comprises both fluid and crystallized cognition measures. For these experiments we only used the larger mixed-sex subject group. There was no significant correlation of the cognition score with age or sex across the considered subjects ($r_{Cognition\text{-}Sex}$ = -0.014, $r_{Cognition\text{-}Age}$= -0.055).

# 3 Results

We applied a Ridge regression model to predict personality trait scores from features of the SC evaluating many different pipelines to 1) investigate the predictive relationship between personality and the structural connections of the brain and 2) the influence that different settings in the prediction pipeline have on the prediction accuracy.
Below we first present a few examples of prediction results for varying the feature class and feature selection before giving an overview of the results from all different pipelines. Then we discuss the prediction brain maps illustrating how the connections from groups of voxels participate in prediction and, finally, evaluate and illustrate the influence of different pipeline parameters on prediction accuracy.

## 3.1 Prediction Performance

### 3.1.1 Personality Trait Prediction

The prediction performance of personality traits from the DWI data is illustrated in Fig. 2 for a few examples of the feature class and feature selection. The shown results were obtained using the Desikan-Killiany atlas [43] with 70 ROIs, NOS weighting of the SC, the mixed sex subject group and the trait openness as target variable (see Methods). Overall, we can see that the prediction accuracy as given by the correlation between predicted and empirical openness scores does not on average reach values higher than ~0.19 for this set up and for all feature classes and feature selections considered. In particular, for the feature class *PCA* and varying the different number of PCs (Fig. 2A), a larger number of PCs (~20) must be used to obtain a positive prediction accuracy, which then stays approximately constant at ~0.1 when adding further PCs. For another feature class referred to as "*corr*" of the most correlated features (see Methods), the prediction accuracy reached plateaus of ~0.15 at around 40 of the most correlated SC edges (Fig. 2B). However, when considering the *whole-brain* feature class (Fig. 2C), which used all SC edges leading to 2415 features in total including the edges of lower correlation, the accuracy was reduced to about ~0.1, which indicates that the plateau seen for the *corr* feature class (Fig. 2B) is expected to decrease again when more SC edges are added as features into the prediction process. Finally, the best accuracy of the considered setup was observed for the *RCP* feature class, when the connectivity profiles of individual brain regions to the rest of the brain were selected as features (Fig. 2D). For the RCP feature selection, the prediction accuracy depends substantially on the selected RCP, which can be observed by the strongly varied correlation values across brain regions. Nevertheless, the RCP features that may include SC edges with high and low correlations with the target score as compared with the *corr* feature class, have the potential to outperform the latter and other feature classes, when an appropriate RCP was chosen. An in-depth comparison of the different feature classes follows in Section 3.2.4.
A second example is illustrated in Supplementary Fig. S2. The presented results were obtained using the Shen atlas [44] with 79 ROIs, the NOS weighting of the SC, the mixed sex subject group and the trait neuroticism. Compared to the first example, the parcellation and the trait were exchanged for another condition. Here, we see overall lower prediction accuracies. For the whole-brain, corr and PCA feature class, only negative correlations were obtained (Supplementary Fig. S2A-C). Solely for the RCP feature class some few RCPs led to positive correlations (Fig. S2D). The proportion of RCPs leading to positive values as well

as the maximum achieved correlation lie below that found for the RCP feature class in the first example (Fig. 3D).

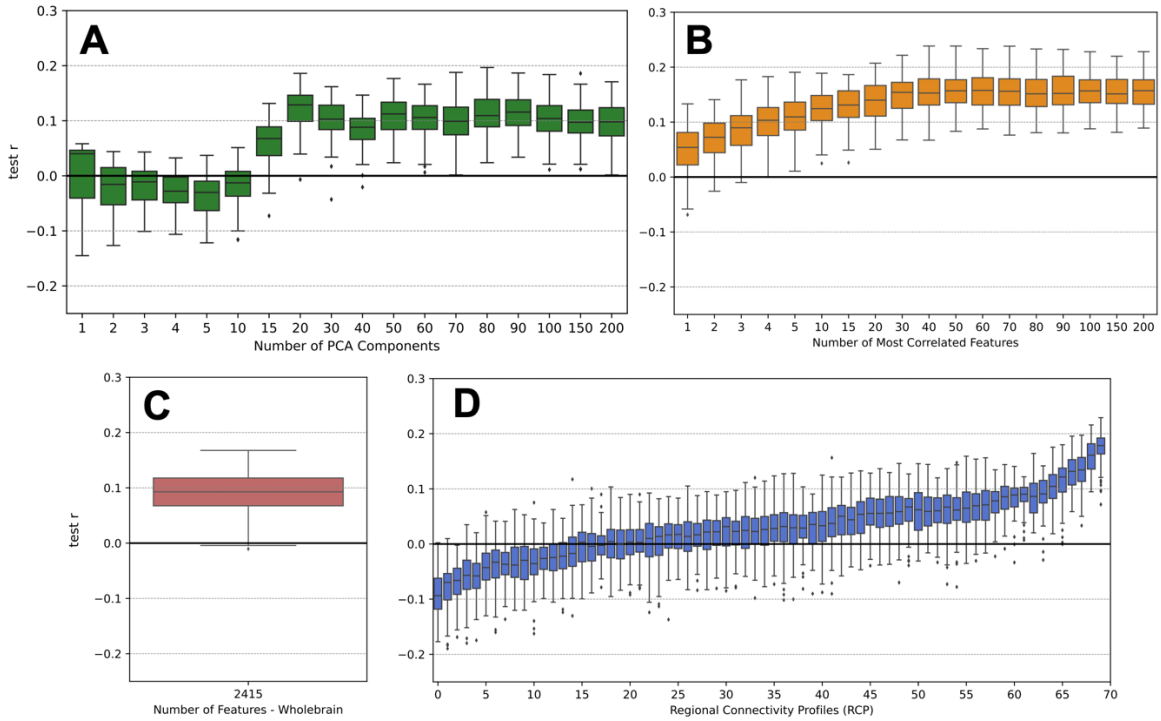



*Fig. 2* **Examples of prediction results for all four feature classes.** The calculations were performed by the pipeline with the Desikan-Killiany atlas (70 ROIs), NOS weighting, mixed-sex subject group, and the trait openness. The box plots show the distributions of the prediction accuracy as given by Pearson's correlation between the predicted and empirical personality scores obtained for the test sets over 100 random splits of the data of the 5-fold cross-validation. The prediction results are depicted for **A** the PCA feature class with different numbers of PCA components, **B** the *corr* feature class with different numbers of the most correlated features, **C** the whole-brain feature class and **D** the Regional Connectivity Profile (*RCP)* feature class for different brain regions, i.e., rows of the SC matrix. RCPs are sorted by the mean prediction accuracy. The respective conditions of the feature selection are indicated on the horizontal axes with the non-linear scaling in plots A and B.

### 3.1.2 Distribution of all Prediction Results

In this study we performed a large number of prediction calculations for 3,420 different pipelines reflecting combinatorial combinations of 19 atlases x 3 SC weightings x 3 subject groups x 4 feature classes x 5 target traits (see Methods). Every pipeline was executed several times using different conditions for the feature selection of a given feature class such as different numbers of PCs (*PCA* class) and the most correlated edges (*corr* class), and different brain regions (*RCP* class), see Fig. 2. By this, the total number of prediction results accumulated to 123,615, which were obtained by averaging over 100 random splits of the subjects into 5-fold CV. These average results therefore required a 100-fold amount of individual prediction runs consisting of a nested 5-fold CV loop used to tune the hyperparameter of the ridge regression and calculate the prediction accuracy on the test sets.

To get an overview over the prediction performance of all evaluated pipelines, we illustrate the distribution of the average prediction accuracies from all conducted calculations in Fig. 3A. With only a few exceptions, the prediction results over all different settings yielded low correlations for the prediction of personality trait scores from the SC features. The mean of the overall distribution of the average prediction accuracies was around zero at r ≈ -0.003. Some very few prediction pipelines led to promising prediction accuracies of $r > 0.2$ (Fig. 3A,

enlargement), which are in the range of values reported in the literature for prediction of personality traits from FC [6], [33], [45].

The grand collection of all prediction results in Fig. 3A can be split into subgroups to reveal if one or another case of the pipeline conditions may lead to different prediction performance than the others. For example, we may compare the prediction results for different personality traits and split the overall result distribution into 5 subgroups of specific traits, while the other pipeline conditions vary within every subgroup. The distribution histograms reflecting the prediction accuracy for the five personality traits are illustrated in the left plot in Fig. 3B. Here one can observe that, despite the overall low prediction performance, the highest prediction accuracies were achieved for the trait openness. The mean correlations of the different distributions all still remain close to zero with the highest mean correlation for the openness trait at r = 0.012, shifted slightly to the right. The effect size in terms of Cohen's d was small to medium when comparing the result distribution of openness to the other traits ($d_{O\text{-}A}$ = 0.28 (Cohen's d between the distributions for the openness (O) and agreeableness (A) trait), $d_{O\text{-}C}$ = 0.26, $d_{O\text{-}E}$ = 0.15, $d_{O\text{-}N}$ = 0.40). No notable differences in predictability could be observed among the other traits, as it is difficult to compare the overall very low effect sizes found for the other traits other than $d_{E\text{-}N}$ = 0.28. Furthermore, we observed a much larger proportion of the best prediction results with $r > 0.2$ for the openness than for the other traits (Fig. 3B right). We also found that especially good results can be reported for this trait in the female only subject group, but also in the mixed sex subject group which can be seen in Fig. 7.

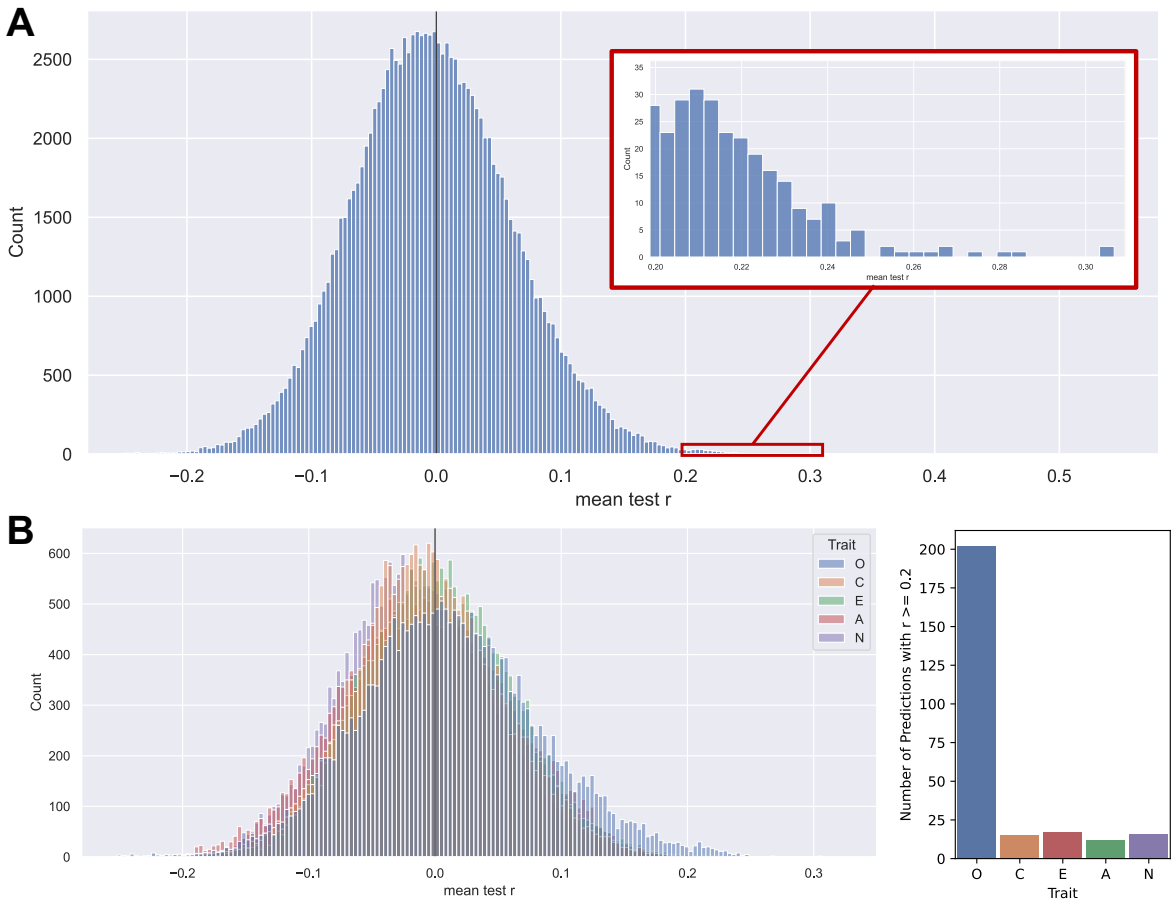



*Fig. 3* ***Overview of all results on personality prediction by structural connectome. A*** *Histogram showing the overall distribution of the prediction accuracy given by the mean correlation r between empirical and predicted personality traits obtained for the test sets and averaged across 100 random subject splits for cross-validation. The prediction results are collected from all different pipelines with varying parcellation, SC weighting, subject group, personality trait, and feature class as well as all conditions of the feature selection. The latter includes different*

*numbers of PCA components (19 options), different numbers of the most correlated SC edges (19 options) and all RCPs (number of options depended on the granularity of the parcellation). The histogram in total contains 123,526 average prediction accuracies. The selected and enlarged area of the histogram shows the cases of r > 0.2.* ***B*** *Left: Five overlaid histograms of the prediction results from plot (A) separated by the five different personality traits as indicated in the legend. Right: Number of occurrences of the different personality traits in the pipelines leading to prediction accuracies r > 0.2, marked by the red box in A. The traits are indicated on the horizontal axis. The trait abbreviations are as follows O: Openness, C: Conscientiousness, E: Extraversion, A: Agreeableness, N: Neuroticism.*

### 3.1.2 Cognition Prediction

To provide an additional context for the prediction results of personality traits illustrated in Fig. 3, we also considered a cognition score from the HCP dataset as another prediction target. We therefore calculated the prediction accuracy for this score for the mixed-sex subject group, and varied brain parcellation, SC weighting and feature class. The obtained correlation distribution for the whole-brain feature class is depicted in Fig. 4A together with that of all five personality traits for the same conditions of the prediction pipeline. We chose the whole-brain feature class since this doesn't have the additional variable of the feature selection, i.e., different RCPs or different numbers of PCA and corr features, which facilitates the comparison. For completeness, Fig. 4B shows prediction accuracies for all feature classes and their evaluated feature selection options for the prediction of cognition. One can see that predicting cognition led to slightly higher values of the prediction accuracies than predicting personality trait scores did (Fig. 4A). Interestingly, the mean of the accuracy distribution for cognition prediction lies at about $r$ = 0.15, whereas the prediction accuracy for personality traits is distributed around $r$ = 0.02 for the pipelines considered in Fig. 4A. The difference between the two distributions shows a very high effect size (Cohen's d) of 2.06. Furthermore, the highest obtained correlation for cognition was $r$ = 0.28, while it was $r$ = 0.21 for openness, $r$ = 0.23 for conscientiousness, $r$ = 0.19 for extraversion, $r$ = 0.20 for agreeableness, and $r$ = 0.21 for neuroticism. This demonstrates that the used prediction pipeline and DWI data can in fact lead to reasonable prediction results for cognition, where the accuracy distribution was clearly shifted to positive values (Fig. 4A). This was however not the case for predicting personality traits (Fig. 3 and Fig. 4A), where the overall and the best-case correlations were smaller than those for cognition.

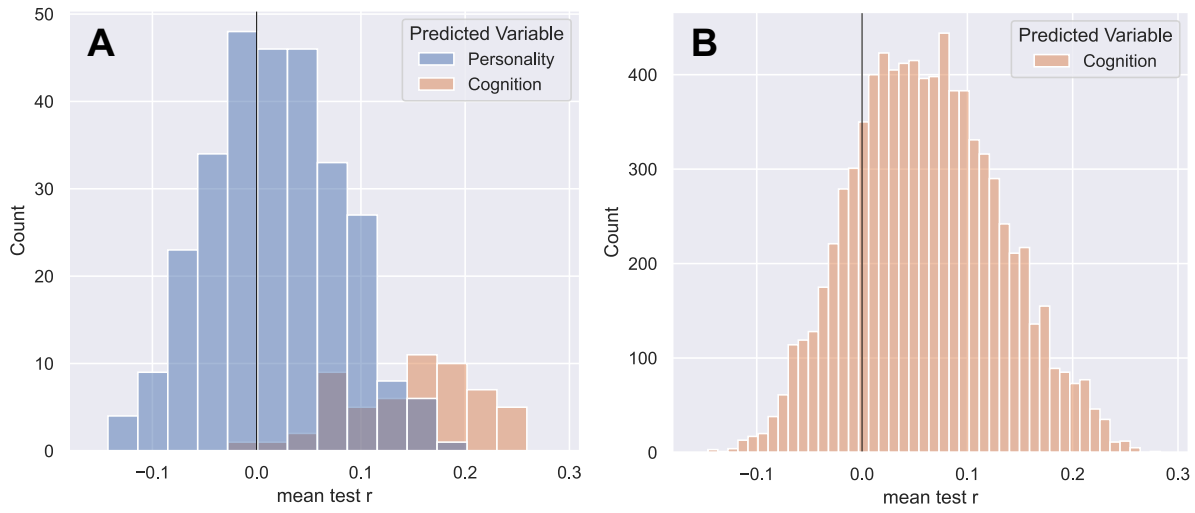



*Fig. 4* **Results for the prediction of cognition. A** Comparing the prediction results for personality traits and cognition. The distributions of the prediction accuracy (correlation) for cognition and all five personality traits are shown as indicated in the legend. In both cases the same prediction pipeline and DWI data were used of the mixed

sex group (n=426), the entire SC (whole-brain feature class) and varied parcellations and connectome weightings. **B** The prediction accuracy of all evaluated prediction pipelines for cognition. Compared to (A) this distribution includes results for all four feature classes (instead of only the whole-brain feature class) and all of the evaluated feature selection options.

## 3.2 Prediction Brain Maps

In addition to evaluating the prediction accuracy, we are interested in exploring which brain regions actively participated in personality prediction. We address this question for the prediction results obtained from the pipelines using the RCP feature class. For this, we calculated the prediction brain maps of the mean prediction accuracy averaged over all considered parcellations as described in Section 2.7 in Methods.

Fig. 5 shows exemplary maps for different traits and different SC weightings. Voxels colored in red show that, across all parcellations, this voxel belonged to a ROI which's RCP led to on average a positive test set correlation. The 'stronger' this red color is, the higher the positive test set correlation. Voxels colored in blue indicate that, on average across all parcellations, these voxels belonged to ROIs which's connections to all other parts of the brain led to a negative test set correlation. I.e. no generalizable connection between the RCP and the personality trait score was found. The stronger the blue color, the more negative the test set correlation was. We do want to clarify that one cannot say that a stronger negative average correlation corresponds to a 'worse' prediction than a slightly negative average correlation. Correlations $r <= 0$ indicate an unsuccessful prediction but it is unclear how negative prediction accuracies relate to each other.

We found that predicting the five different target traits was associated with different prediction brain maps as illustrated in Fig. 5A for the NOS weighting and mixed sex subject group. In particular, the map for openness stands out as it indicates that, especially for the left hemisphere, RCPs from almost all regions of the brain actively participated in the prediction of this trait and delivered positive prediction accuracy across different parcellations. This may be compared with the conscientiousness map exhibiting a uniformly low average correlation (low color intensity in Fig. 5A, middle), which makes it difficult to select any brain areas clearly associated with this trait. On the other hand, the well-pronounced landscapes of high and low average prediction accuracies in the maps for extraversion, agreeableness and neuroticism can help to distinguish brain regions whose connections are important for predicting the respective trait (Fig. 5A).

The discussed prediction brain maps can also be calculated for the three SC weightings as illustrated in Fig. 5B for the mixed sex subject group and the trait neuroticism. Here we also observe that different SC weightings can be associated with different brain maps. In particular, the NOS weights were accompanied by a distinct landscape of the brain map as compared to those of MA and FA weightings (Fig. 5B). For the NOS weighting, one can clearly distinguish the brain areas positively contributing to the personality prediction from those with a little or even negative impact on the prediction performance (Fig. 5B, red and blue domains). The other two SC weightings based on FA and MD led to less contrasted prediction brain maps that are rather similar to each other, but distinct from the NOS case, which we consider in more detail later below.

It must be noted that the maximal mean correlations observed in the maps are low at $r = 0.12$ when comparing traits and $r = 0.08$ when comparing weightings. Nevertheless, the obtained brain maps can be compared to baseline maps of 'random prediction' (Supplementary Fig.

S3), where we additionally ran permutation tests for the RCP feature class. There, we again performed 100 repetitions of a nested 5-fold CV for the same pipelines, but the target values were randomly shuffled between subjects. The baseline maps for the different traits show a mean correlation across parcellations close to zero for the entire brain (Supplementary Fig. S3). Therefore, it can be reported that distinct brain maps arise for the different prediction targets and connectome weightings marking RCPs which lead to positive or negative prediction accuracies across all 19 parcellations. This shows that it depends on the prediction target and the considered weighting which RCPs can predict the personality target more or less accurately, i.e. which brain areas and their connections might be more or less related with the trait of interest.

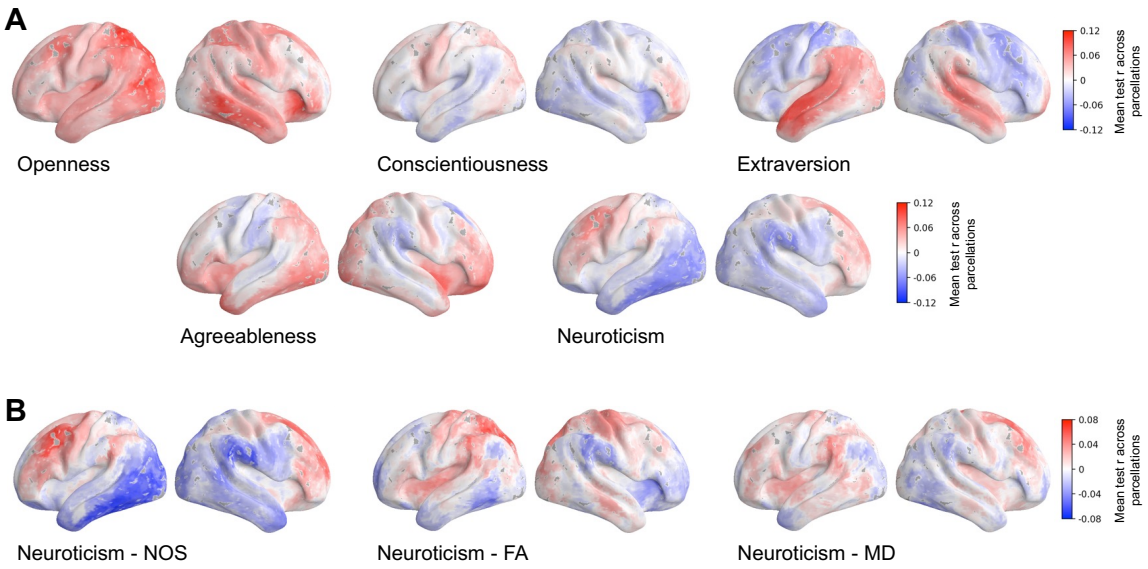



*Fig. 5* **Prediction brain maps associated with the performance of personality prediction from DWI data.** The prediction accuracy (correlation) of the RCP feature class was assigned to all voxels of the respective brain regions and averaged over all considered parcellations for fixed other conditions of the prediction pipeline (see Methods). The brain maps are illustrated for **A** the five different personality traits indicated in the plots as prediction targets obtained for the mixed sex subject group and the NOS SC weighting, and **B** the three different SC weightings indicated in the plots obtained for the mixed sex subject group and the personality trait neuroticism. The average correlation is depicted by colors with the scaling indicated in the color bars. Left and right hemispheres are shown from the lateral view. Abbreviations are NOS: the number of streamlines, MD: the mean diffusivity, and FA: the fractional anisotropy. The visualization of the maps was created using the neuromaps toolbox [46] including the volume-to-surface transformations as proposed and defined in [47], [48].

## 3.3 Influence of Different Pipeline Settings

### 3.3.1 Parcellation

We observed no clear trend towards a superiority of certain brain parcellations for personality prediction. There seems to be a slight tendency towards an advantage of a finer granularity (more brain regions) within the same parcellation scheme, for example, when the best prediction results were selected from the RCP feature class (Supplementary Fig. S4). A similar trend can be observed when considering pipelines leading to correlations $r > 0.2$ for the prediction of cognition. Here, parcellations with higher granularities more often lead such results than coarser parcellations (Supplementary Fig. S5B). When considering all different pipelines, it however depends on the combination of the individual settings of the pipeline conditions which brain parcellation performs best.

3.3.2 Connectome Weighting

We also did not clearly observe an overall best connectome weighting for the prediction of personality traits. The prediction performance of the weighting depends on the other settings, such as the parcellation scheme, the trait or the feature class. For predicting cognition, we can however see a quite clear difference between the different SC weightings. NOS performs better than FA which in turn performs better than MD (Supplementary Fig. S6). Furthermore, among the best prediction results, characterized by an average prediction accuracy of $r > 0.2$, NOS is the weighting occurring most frequently for both cognition and personality traits. In particular, ~69.5% and 76% of the best pipelines leading to prediction accuracies of $r > 0.2$ used NOS as SC weighting for the prediction of cognition and personality traits, respectively (Fig. 6A).

The properties of SC weightings and their impact on the prediction results can also be compared based on the prediction brain maps calculated for the RCP feature class by averaging across all considered parcellations (Fig. 5B). We calculated the correlations between any of these two maps, which was repeated for all combinations of different personality traits and subject groups to obtain distributions of the correlations (Fig. 6, right). One can see a high similarity (correlation) between the cortical maps from the MD and FA weighting, whereas the brain maps of both microstructural DWI weightings exhibit distinct patterns as compared with that of the NOS weighting, which is reflected by a low correlation of NOS maps with both MD and FA maps. Considering that the meaning of more or less negative correlations is unclear but is given a meaning by correlating the maps of different weightings, we further correlated the maps after setting all voxels with $r < 0$ to zero and only considering voxels with $r > 0$ for all three weightings. While these configurations might in turn have other drawbacks, the stronger similarity between maps of MD and FA weighting and the more distinct pattern of the NOS map were confirmed for all three setups. This also holds for individual parcellations for both personality traits and cognition (Supplementary Fig. S7B and C). Analogous similarity patterns can also be observed when calculating the correlation between the differently weighted normalized SCs for all subjects selected for the study (Supplementary Fig. S7A). Correlations between the FA- and MD-weighted SCs are close to 1 whereas correlations between NOS-weighted connectomes and the others are significantly lower for all parcellations except for the parcellation with the coarsest granularity (MIST 31) (Supplementary Fig. S7A).

We also compared the prediction results obtained for the NOS-weighted connectomes with those based on the SIFT2-weighted version, which is supposed to more accurately represent the density of structural connections [32]. Nevertheless, we observed only a very marginal improvement of prediction results when using the SIFT2-weighted connectomes for personality prediction performed for the mixed-sex subject group and varying conditions of the feature class, parcellation and all five personality traits (Supplementary Fig. S8). While the difference between the two distributions of the prediction accuracy is statistically significant with $p << 0.001$ (two-sided t-test), the effect size in terms of Cohen's d values is very small at $d = 0.072$.

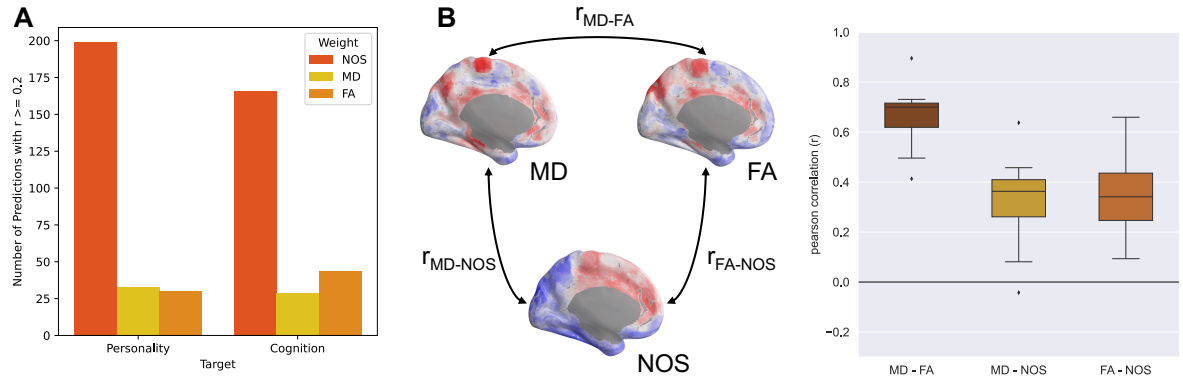



*Fig. 6* **Impact of connectome weighting on prediction of personality and cognition. A** Fraction of the predictions results with the prediction accuracy r > 0.2 broken down by the different connectome weightings as indicated in the legend for both the prediction of personality and cognition as denoted on the horizontal axis. **B** Comparison of the prediction brain maps obtained for different connectome weightings and the RCP feature class, see Fig. 5B. The maps for different weightings were compared by correlation across voxels for all five personality traits and all three subject groups, so that the boxes in the graph on the right each represent a distribution of 15 different correlation values. Abbreviations are NOS: the number of streamlines, MD: the mean diffusivity, and FA: the fractional anisotropy. The example maps on the left show maps for the trait neuroticism and the mixed-sex subject group.

### *3.3.2* Subject Group

In this study three different subject groups of males, females and mixed-sex subjects were considered as one of the pipeline conditions. We observed some differences in terms of prediction accuracies between these subject groups, which were however not consistent across different pipeline settings such as different SC weightings and different traits. Fig. 7 shows prediction accuracies from the pipelines using the whole-brain feature class for the different subject groups across the five different traits for the NOS weighting (Fig. 7A), the MD weighting (Fig. 7B) and the FA weighting (Fig. 7C). For the NOS weighting the female subject group performs a lot better than the mixed-sex and male group for the openness trait. This is however not the case for openness using the MD or FA weighting. For most other traits, there is no improvement when separating the subjects by sex or even a deterioration such as for the trait agreeableness for all three SC weightings. As the prediction accuracies and their differences were overall low, it is little conclusive whether the differences between subject groups arise from distinct brain-personality relationships in males and females or simply from the fact that these groups were composed of different subjects independently of subject sex. Further investigations would be necessary to ensure there is a difference if a more stable SC-personality connection could be established.

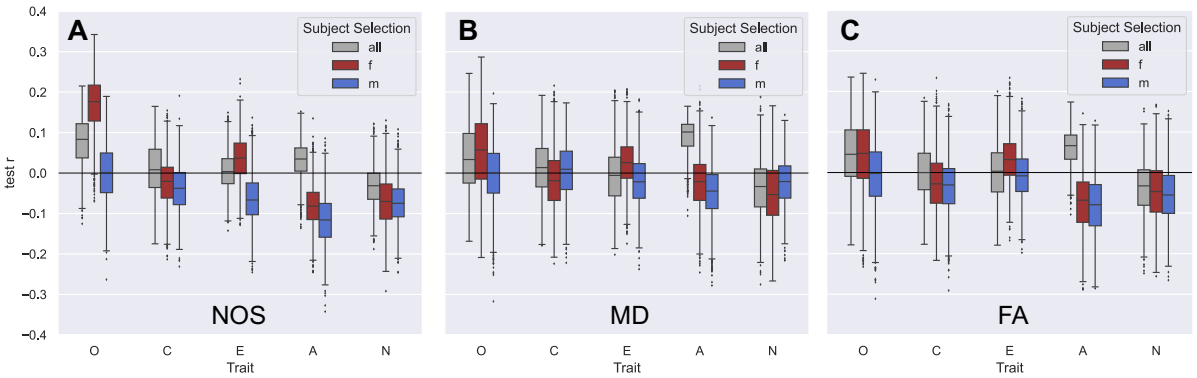



*Fig. 7* **Impact of different subject groups on the prediction results.** The plots include results from the whole-brain feature class and all different parcellations for **A** the NOS weighting, **B** the MD weighting and **C** the FA

weighting. In each plot the results are separated by trait as indicated on the x-axis and the different subject groups are indicated in the legend - all: mixed-sex, f: females, m: males.

### 3.3.4 Feature Classes

An important pipeline condition that can influence the prediction performance is the employed feature class. The prediction results obtained for different feature classes were compared in this study in two different ways, see Section 2.7 in Methods. According to the first approach, we calculated the prediction accuracies on the test set for all considered cases of pipeline conditions and feature selections. Then the best prediction results were selected for each of the feature classes across the feature selection instances, i.e. the best number of most correlated features, the best RCP and the best number of PCs. In Fig. 2A, B and D, for example only the distribution with the highest mean value is chosen for each feature class respectively such that they can be compared with each other and the distribution of Fig. 2C for the whole-brain feature class. This procedure was repeated for each combination of other pipeline settings. In this comparison, the *RCP* feature class outperformed the other feature classes in almost all prediction pipelines including all personality traits, see Fig. 8A. In addition to the mixed-sex subject group illustrated in Fig. 8A, we verified and confirmed this conclusion for the female only and male only subject groups. Furthermore, we also found that the majority (over 50%) of the results with prediction accuracies $r > 0.2$ collected over all cases and conditions of the pipelines were obtained from the *RCP* feature class. Interestingly, a different picture was observed for the prediction of cognition, where almost 60% of all best results were obtained from the *PCA* feature class. This is visualized in the Sankey plots in the Supplementary Fig. S5 which shows the distribution of promising results ($r > 0.2$) over the different pipeline conditions.

Overall, for all different pipeline conditions, there were only a few RCPs that lead to the promising results (Fig. 2D) meaning that only a few connectivity profiles in the SC have a stronger relationship with the trait of interest, while the other connectivity profiles in the SC have little association with the personality scores. Furthermore, the predictive RCPs differ between different SC weightings and between the different personality traits as can be seen in Fig. 5.

The above comparison, obtained by selecting the best prediction accuracy out of many others after the CV loops, i.e. based on the test set correlation, may serve as an idealistic upper boundary of prediction performance. As a more sophisticated approach, we ran an additional prediction analysis, where the issue of the optimal feature selection was resolved within the inner CV loop, see Methods. In more detail, we included the selection of the optimal RCP, number of PCs and number of most correlated features as an additional hyperparameter to be optimized together with the regularization hyperparameter of the ridge regression in the inner loop of the 5-fold nested CV. The obtained results are illustrated in Fig. 8B for the mixed-sex subject group, where we observe a drop of the prediction accuracies for almost all illustrated cases, especially for the RCP feature class. With this comparison approach, there is no clear difference between the prediction performances of different feature classes (Fig. 8B).

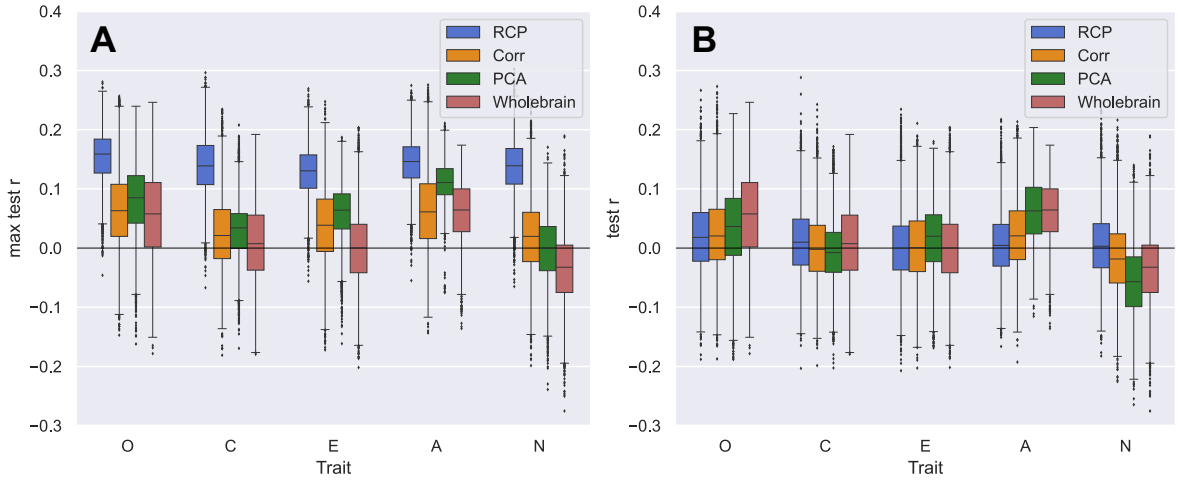



*Fig. 8* ***Comparing the performance of different feature classes for personality prediction from DWI data.*** ***A*** *Comparing the best prediction results for different feature classes as indicated in the legend and the five traits indicated on the horizontal axis (O – Openness, C – Conscientiousness, E – Extraversion, A – Agreeableness, N – Neuroticism). Boxplots depict the distributions of the prediction accuracies calculated for 100 random data splits for CV loops and collected over all considered parcellations, all SC weights and the mixed-sex subject group. For every pipeline condition (fixed trait, weighting, parcellation and subject group) the best result was selected across all feature selection instances of the respective feature class (RPC, number of PCs and most correlated SC edges).* ***B*** *Data presentation as in plot A. Here, the optimal feature selection instances for the different feature classes were however determined as a hyperparameter in the inner loop of the nested CV (see text for details).*

# 4 Discussion

## 4.1 Personality Trait Prediction

In the current study, we predicted the big five personality trait scores from the SC derived from dwMRI data of unrelated subjects from the HCP young adult dataset. We systematically evaluated the effect of different pipeline settings on the prediction performance including 19 cortical brain parcellations, three SC weightings, three subject groups and four feature classes. For the largest of the three subject groups (mixed-sex), we ran additional analysis for the prediction of a cognition target and for an extension of the NOS weighting – the SIFT2 weighted connectome.

Overall, we find only a few cases of promising prediction performance (correlation $r > 0.2$) which are similar to values reported in the literature for personality prediction from the resting state FC ([6], [33], [45], [49]). Most results, however, are centered around a correlation of zero with the mean at $r \approx -0.003$ (Fig. 3A) indicating no generalizable predictive relationship between personality traits and SC for the pipelines and data evaluated here. Considering these already very low correlations we did not further investigate the effect of potential confounds apart from sex. The effect of sex was investigated independently by considering both a mixed-sex group and subject groups separated by sex. Nevertheless, it cannot be ruled out that other confounds such as in scanner head movement [50] or age (despite the small range from 22-35 years) inflated the results. This would however further strengthen the already observed null-findings.

While earlier studies relating dwMRI features to personality focused on TBSS and correlation analysis, the inconsistent results obtained in different analyses well support our findings. Nevertheless, for a more direct comparison we also conducted a TBSS analysis for the data considered here. The exact method and results can be found in the supplementary. Overall, our findings for TBSS are in line with our prediction results: For most traits, there are no significant TBSS-based correlation results, and the few significant findings have low correlations with $|r| < 0.21$. Prior research based on TBSS that found significant correlations between voxel clusters and personality, in many cases reported correlations $r \leq 0.3$. Only one approach by Xu and Potenza [12] found very strong relationships between WM microstructural properties and personality with r's up to 0.75 using the TBSS analysis on a small sample size of N = 56. [12]. A study by Ooi et al. [51] predicted different aspects of behavior as determined by a factor analysis from many behavioral scores from the SC with different weighting. The factors of *emotion* and *dissatisfaction* in the HCP dataset, as well as the *personality* factor in the ABCD dataset [52], were predicted with correlations $r < 0.1$ for all weightings; this again is in line with the low correlations we obtained for the prediction of personality traits. Although predictions from the FC showed better results than our findings for the SC, the correlation in these cases was still limited to values below $r = 0.29$ ([6], [45], [49]). The only exception is the study by Nostro et al. [33] that found correlations up to $r = 0.42$ for predicting traits from certain functional brain networks. The approach by Dubois et al. [45] provided prediction accuracies for different setups and also reported negative / unsuccessful prediction accuracies for some setups for all traits except openness. This work therefore found the trait openness to be predicted best and most reliably from the FC, just as we did here for the SC (Fig. 3B). Considering results for brain structural measures, a review and meta-analysis by Chen et al.

[53] systematically evaluated research relating personality traits to brain structure, and their meta-analysis showed no replicable results for gray matter volume (GMV), cortical thickness and surface area.
All in all, despite some promising results for certain modalities, the prediction of the big five personality traits from brain imaging data remains to be a challenging task. This is underlined by the results of our study, showing that no generalizable and clear predictive relationship between the SC and the big five personality trait scores can be found so far.

#### 4.1.1 Comparison with the Prediction of Cognition

To further investigate why no strong relationship between personality and the structural connections of the brain could be found, we additionally predicted a cognition score from the SC for the mixed sex group as a comparison. Here, we found improved performance both in terms of the best performing pipelines and the average prediction accuracy compared to the prediction of personality (Fig. 4). The improved, but still limited prediction of the cognition target indicates that the results for the personality prediction might be influenced by both inherent limitations of the standard SC as well as reliability issues of the self-reported evaluation of personality scores themselves, which has been shown to limit brain-behavior predictions [54]. Additionally, the big five personality traits represent a model of personality, and it remains unclear how close they come to actual true factors of personality and how these could ideally be measured. We therefore expect that further advances in both dwMRI / SC extraction and reliable estimations of the personality trait scores are necessary to find a potential connection between our personality and the structural connections of our brain. In addition, large sample sizes, enabling the use of more complex, non-linear models such as graph neural networks [55] might offer new opportunities to relate structural brain connectivity to personality.
Related literature on prediction of cognition from SC or FC shows diverse results, some of them in line with what we find here. Research by Ooi et al. [51] and Yeung et al. [28] find similar correlation values for the prediction of cognition from the SC. Work by Feng et al. [27] also show a large range of possible prediction accuracies for different pipelines, the maximally reached values are however higher than in our findings (r up to 0.50). Reported accuracies for the prediction from the FC are mostly higher compared to those from the SC ([56], [51], [57], Zhu), the range of reported correlations however still strongly varies in r = [0, 0.77]. In contrast an approach by Litwińczuk et al. finds target for which the SC performs better than the FC for prediction (executive function, language) [58]. These results show the improved prediction of cognition targets compared to personality and the general applicability of the SC for cognition prediction across literature. The diverse results again highlight how different data, pipelines and targets, as well as the presentation of results, can influence the prediction and complicate comparability of results.

#### 4.1.2 Prediction Brain Maps

Using the results from the *RCP* feature class pipelines, we generated cortical maps of average ‘local’ prediction accuracy across all 19 parcellations. These maps showed distinct pictures of which RCPs contributed positively to the personality trait prediction for the five different traits. They also indicated a unique role of the different SC weightings to the prediction. Considering the overall very low prediction accuracies, we may not thoroughly interpret the RCPs leading to promising results for the different personality traits. We however highlight the potential of such maps to identify regions whose connections to other parts of the brain may have a strong link to the trait / target of interest. We expect that the connections of local voxel clusters that

belong to RCPs leading to superior prediction results for many different parcellations are more strongly linked to the target. Such maps might therefore help to determine connections or regions of interest for further studies focusing on a connection between the brain SC and behavior.

#### 4.1.3 Prediction Model

In this study we applied a ridge regression model to prediction. There is different evidence concerning the effect of the prediction algorithm on the effect size of the prediction. Both Yeung et al. [28] and Schulz et al. [42] reported comparable results between classical linear ML models such as ridge regression or elastic net and non-linear deep learning models for predicting different behavioral phenotypes from the SC and FC respectively. Further, a work by Cui et al. [59] compared six regression algorithms for the prediction of behavioral variables from features of the rsFC. Of the six evaluated algorithms, four, including ridge regression, presented a similarly strong performance outperforming the other two algorithms. Of the four superior algorithms, ridge regression had the lowest computational burden. On the other hand, Feng et al. [27] found advanced prediction accuracy using a DL model over classical ML models for the prediction of age and cognition targets from the SC while still observing a comparably strong signal for prediction using the classical ML models. The simple linear ridge regression model was chosen based on these findings and the computational burden of the different prediction models. The comparably low computational load of the ridge regression enabled the comparison of many different other pipeline parameters as we did in this study. In addition, findings from literature comparing different algorithms suggest that the low prediction accuracy found here for prediction of personality are most likely not due to the selected prediction algorithm.

## 4.2 Influence of Different Pipeline Conditions

Despite the observed overall low prediction accuracy we still identified differences in prediction results between distinct pipeline settings.

#### 4.2.1 Brain Parcellations

There was no clear best brain parcellation for personality prediction. While there was a slight tendency towards prediction improvement with higher granularity within the same parcellation scheme, differences were not large enough to give a clear recommendation. It is also important to add that while we investigated a large number of brain parcellations, the highest granularity we used was 210 ROIs and there are significantly finer brain parcellations available (e.g. Schaefer parcellation with up 1000 ROIs [60]). Considering the inherent limitations of diffusion tractography [61] and the lower resolution of diffusion MRI compared to other MRI modalities, the tendency towards an improvement with higher granularity might saturate or reverse at some point when continuously increasing granularity possibly leading to an optimal granularity, at least for given recording and data preprocessing settings.

Investigations by Feng et al. [27] for prediction of cognition from the SC showed improved performance of the Human Brainnetome Atlas with 246 ROIs [62] compared to the Automated Anatomic Labeling with 90 ROIs [63] across two datasets and several cognition traits. Both of these parcellations were also evaluated in this work and the improved performance of the Brainnetome atlas was confirmed for the prediction of cognition for all three SC weightings. This however didn’t hold for most configurations when predicting personality traits, showing

an influence of the target. This is in line with research conducted by Zhang et al. [64] for predicting different behavioral measures from the SC. They found improvement of prediction performance for a higher granularity parcellation only for some of the traits. Dhamala et al. [56] predict cognition from both SC and FC for a parcellation with 86 ROIs (combination of the Desikan-Killiany atlas [43] and additional subcortical structures [65]) and an in-house parcellation with 439 ROIs. While the prediction performance improves with a higher granularity for the FC, this does not hold for the SC for which performance doesn't improve or even deteriorates. This might affirm our assumption that at a certain point higher granularity does not further improve prediction accuracy for the SC.

#### 4.2.2 SC Weightings

For predicting personality, there was no global best SC weighting. However, when only considering the pipelines leading to prediction accuracies $r > 0.2$, the NOS weighting was used in a vast majority of cases (76%, Fig. 6A). For the cognition target one could see a clear ranking of SC weightings (NOS > FA > MD) (Supplementary Fig.S5). This is consistent with the study of Yeung et al. [28]. They also found that generally the best results for prediction of sex, age, general cognition, and mental health from the SC were achieved for the NOS weighting. For predicting the general cognition factor with the BrainNetCNN, they also found a similar ranking of the SC weightings: $r_{MD} = 0.138$, $r_{FA} = 0.168$ and $r_{NOS} = 0.201$). Also studies by Liu et al. [66] and Zhang et al. [64] predicting continuous behavioral phenotypes from SCs, found improved or equal performance of the NOS weighting compared to the microstructural weightings MD and FA. A study by Ooi et al. [51] on the other hand found no consistent best weighting when predicting several behavioral targets in two large scale datasets, which demonstrates the difficulty of comparing different weightings when other pipeline conditions are set differently between studies.

Considering that for different weightings, different RCPs are predictive of a certain personality trait (Fig. 5B), it might be valuable to combine different weightings for the prediction. The higher similarity between the two maps of MD and FA compared to the more distinct pattern for the NOS map (Fig. 6B) is most likely due to the fact that MD and FA are both calculated from the eigenvalues of the diffusion tensor and are related by definition [5], while the NOS weighting represents a completely different form of weighting. Taking this into consideration it might be worthwhile to only consider one DTI-based microstructural weighting and combine it with the NOS weighting to enhance information capacity and improve prediction performance.

#### 4.2.3 SIFT2 Refinement

We furthermore investigated applying the SIFT2 algorithm [32] to the NOS-weighted connectomes to compare predictive performance for the mixed-sex subject group and found a very slight improvement of the prediction accuracy when applying the SIFT2 weighting (Supplementary Fig. S8). The additional post-processing, i.e. weight calculation for each reconstructed streamline, that is necessary for the SIFT2 weighted SC, increases the time that is necessary to construct the SC from the dwMRI data by 28-48% per subject [32]. It therefore depends on available resources if additional refinement, which only showed very slight improvement for the prediction compared to the NOS weighting, should be applied. One possible explanation for the observed slight improvement after applying the SIFT2 weighting is that the NOS- and SIFT2-weighted SCs are still very strongly correlated. Even though the mean relative change across all subjects can reach values of up to 67% for certain connections, the mean Pearson's correlation across all subjects and parcellations between NOS- and SIFT2-weighted connectome lies at $r = 0.99$ (Supplementary Fig. S9). This means

that despite quite high relative changes for some connections between SIFT2 and NOS, the connectivity patterns seem to mostly be retained for the two weightings.

#### 4.2.4 Subject Groups

A review and meta-analysis by Chen et al. [53] investigating research on brain structure - personality associations found that for each trait, there was at least one study that found sex differences in the association between personality and regional gray matter volume (GMV). However, for all the traits there was more research that examined the role of sex on the personality – brain structure relationship and found no influence of sex showing that there are overall mixed findings with the majority not showing any sex-depended associations.

Here, we did find differences in performance between the different groups (mixed-sex, only male and inly female) (Fig. 7). However, there was no clear pattern and, considering the overall low correlations and relatively small sample size, it is hard to claim if there was an influence of sex on the relationship between brain structure and personality or if differences only arrived due to a different group of subjects independent of their sex. This would require additional investigations.

Considering the FC, not many studies investigated the role of sex on the prediction or relationship with personality traits. A study by Nostro et al. [33] found that additional functional networks predicted extraversion, neuroticism and openness when separating subjects by sex in comparison to a mixed sex subject group. Specifically, they e.g. found that openness was significantly better predicted in women compared with men for the reward network. This is similar to one of our findings that openness is the only trait for which we see an improvement for the female group when separating subjects by sex (Fig. 7). This however only hold for the NOS weighting.

#### 4.2.5 Feature Classes

Among all considered feature classes, the *RCP* class led to the most promising prediction performances for the personality traits (Fig. 8A). It is however not clear how well these RCP results generalize to other datasets. In particular, the beneficial impact of the RCP feature class on the prediction performance disappeared, when the feature selection process was optimized under the nested CV loop (Fig. 8B). This might however be connected with a relatively small sample size considered in this study, and further investigations of this issue are necessary. In general, this finding highlights the importance of out of sample prediction and proper assessment of the generalizability of predictive pipelines. When evaluating many different pipelines, it needs to be ensured that the best performing pipelines generalize well to new data. In addition, giving an overview over the results from all evaluated pipelines gives the reader a chance to assess how difficult it is to predict the target in general.

As mentioned above for the connectome weightings, only a few RCPs led to relatively high positive correlations between predicted and empirical personality trait scores (Fig. 2D). This may indicate that the relationship between brain structural connections and personality, to the extent that could be found here, is localized in certain connections. Single RCPs would lead to better prediction results than using the entire SC for almost all pipelines (Fig. 8Fig. 8A). In contrast to the personality traits, the *PCA* feature class led to the best results for the prediction of cognition. Here, compared to the personality traits, it also worked better to use the whole brain feature class (Supplementary Fig. S5B). Interestingly, for both the cognition and personality targets the *corr* feature class did not show a particularly strong performance (Supplementary Fig. S5). In comparable literature for prediction of behavior from the FC and

SC this type of feature selection is used quite commonly through the application of the connectome based predictive modeling [67]. Despite promising prediction performance of these approaches (see e.g. [6], [68], [69]), considering the results here, it might be worthwhile to investigate other feature classes as well.

## 4.3 Importance of Result Presentation

Publication bias led to positive and significant findings being reported and published more frequently than null-findings or non-significant results in the past [70], [71]. An overly optimistic picture might therefore be established in the literature for certain prediction investigations when several analyses were performed and only significant or the best results were reported to enhance chances of the work being published. Presenting all the obtained prediction accuracies gives the reader a chance to assess the prediction performance that can be expected when randomly selecting a pipeline from the set of evaluated conditions. Considering the entire spectrum of the obtained results can further give a good impression on how easy or difficult it is to predict a certain target from the features of interest in general. The best results achieved by some of the pipelines evaluated here, are in line with other results found in the literature. However, not presenting the vast amount of unsuccessful predictions, illustrated in Fig. 3A, would leave the reader with an unrealistically good impression for the prediction of personality from the SC. Our results further show that despite using a CV prediction method to assess an out-of-sample performance, overfitting, in this case to the test set, can still occur. Fig. 8 clearly shows that selecting the best RCP on an independent set in the inner loop of the nested CV leads to significantly worse results than selecting it post-hoc based on the calculated test set accuracies. Here, this presentation of overfitting confirms and strengthens the assumptions of null-findings for the SC-personality relationship. This also demonstrates the importance of verifying selected ‘good’ pipeline conditions for prediction either by having them determined in the inner loop of the CV or by applying the successful pipelines to completely unseen data.

Independent of the exact targets and features used for prediction, our results show how strongly cherry-picking results obtained from several analyses might influence the perceived prediction performance and further highlight the importance of reporting all performed analyses and testing for potential overfitting.

# 5 Summary

In the present study, we systematically evaluated different pipeline conditions for the prediction of the big five personality trait scores from the SC. Despite the large variability in pipelines, we found no linear generalizable relationship between personality trait scores and the structural connectome derived from dwMRI data. Only a few pipelines could be found that lead to promising results while still of low correlations between empirical and predicted scores. Considering the vast majority of unsuccessful predictions and the comparably small sample sizes, we do not expect the more successful setups of the prediction pipelines to generalize well to new unseen data. Comparisons with the results for prediction of a cognition target indicate that there might be limitations considering the reliability of the personality target as well as the reliability of current tractography and SC reconstruction algorithms. We expect that larger advances in the reliability of both trait scores and dwMRI / SC calculation as well as larger samples will be necessary to uncover a potential connection between the structural connection of the brain and personality traits. Overall, it became apparent that all different design choices we evaluated in the process of calculating the SC and predicting a behavioral target from it influenced the prediction outcome. Apart from comparing obtained prediction accuracies for different combinations of pipeline conditions, we calculated and compared prediction brain maps of local prediction accuracy. We found distinct prediction brain maps for different personality trait scores and highlighted the potential of such maps to determine regions whose connections to the rest of the brain resemble good features for predicting a certain target trait. Distinct maps could also be found for different connectome weightings. While the maps of the two microstructural weightings were quite similar, the map of the NOS weighting showed a distinct pattern. Even though the NOS weighting outperformed microstructural properties as SC weighting for both the cognition and personality variables, it might be beneficial to combine the NOS weighting with a microstructural weighting for enhanced information capacity and possibly better prediction.

# Author Contributions

**Amelie Rauland:** conceptualization, data curation, formal analysis, investigation, software, visualization, writing – original draft preparation. **Kyesam Jung:** data curation, software, writing – review and editing. **Theodore D. Satterthwaite:** conceptualization, supervision, writing – review and editing. **Matthew Cieslak:** conceptualization, writing – review and editing. **Kathrin Reetz:** supervision, writing – review and editing. **Simon B. Eickhoff:** conceptualization, resources, supervision, writing – review and editing. **Oleksandr V. Popovych:** conceptualization, project administration, resources, supervision, writing – original draft preparation

# Acknowledgements

This work was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) – grant number 269953372/GRK2150 (International Research Training Group 2150) and by the Portfolio Theme Supercomputing and Modeling for the Human Brain by the Helmholtz association, the Human Brain Project and the European

Union’s Horizon 2020 Research and Innovation Programme under the Grant Agreements 945539 (HBP SGA3) and 826421 (VirtualBrainCloud). SBE acknowledges funding by the Deutsche Forschungsgemeinschaft (SPP 2041, SFB 1451, IRTG 2150).

Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

The authors gratefully acknowledge the computing time granted through JARA on the supercomputer JURECA at Forschungszentrum Jülich.

# Conflict of Interest

The authors declare no potential conflict of interest.

# Data Availability Statement

The anonymized data from the Human Connectome Project used for this analysis is publicly available from ConnectomeDB.
The structural connectivity pipeline (https://github.com/ameliecr/SC-calculation-pipeline) as well as the code used for prediction (https://github.com/ameliecr/SCandBigFive) are available on GitHub.

[1] C. G. DeYoung, J. B. Hirsh, M. S. Shane, X. Papademetris, N. Rajeevan, and J. R. Gray, “Testing Predictions From Personality Neuroscience,” *Psychological Science*, Apr. 2010, doi: 10.1177/0956797610370159.
[2] R. McCrae and P. Costa, “The five factor model of personality: Theoretical Perspective,” Jan. 1996.
[3] S. Mori and P. B. Barker, “Diffusion magnetic resonance imaging: Its principle and applications,” *The Anatomical Record*, vol. 257, no. 3, pp. 102–109, Jun. 1999, doi: 10.1002/(SICI)1097-0185(19990615)257:3<102::AID-AR7>3.0.CO;2-6.
[4] B. Jeurissen, M. Descoteaux, S. Mori, and A. Leemans, “Diffusion MRI fiber tractography of the brain,” *NMR in Biomedicine*, vol. 32, no. 4, p. e3785, Apr. 2019, doi: 10.1002/nbm.3785.
[5] D. L. Bihan *et al.*, “Diffusion tensor imaging: Concepts and applications,” *Journal of Magnetic Resonance Imaging*, vol. 13, no. 4, pp. 534–546, Apr. 2001, doi: 10.1002/jmri.1076.
[6] H. Cai, J. Zhu, and Y. Yu, “Robust prediction of individual personality from brain functional connectome,” *Social Cognitive and Affective Neuroscience*, vol. 15, no. 3, pp. 359–369, May 2020, doi: 10.1093/scan/nsaa044.
[7] A. D. Nostro, V. I. Müller, A. T. Reid, and S. B. Eickhoff, “Correlations Between Personality and Brain Structure: A Crucial Role of Gender,” *Cerebral Cortex*, vol. 27, no. 7, pp. 3698–3712, Jul. 2017, doi: 10.1093/cercor/bhw191.
[8] S. M. Smith *et al.*, “Acquisition and voxelwise analysis of multi-subject diffusion data with Tract-Based Spatial Statistics,” *Nat Protoc*, vol. 2, no. 3, Art. no. 3, Mar. 2007, doi: 10.1038/nprot.2007.45.
[9] R. Leshem *et al.*, “Inward versus reward: white matter pathways in extraversion,” *Personality Neuroscience*, vol. 2, p. e6, Jan. 2019, doi: 10.1017/pen.2019.6.
[10] R. E. Jung, R. Grazioplene, A. Caprihan, R. S. Chavez, and R. J. Haier, “White Matter Integrity, Creativity, and Psychopathology: Disentangling Constructs with Diffusion Tensor Imaging,” *PLOS ONE*, vol. 5, no. 3, p. e9818, Mar. 2010, doi: 10.1371/journal.pone.0009818.
[11] D. A. Wilhelms, “MR-Morphometrie und subklinischer Neurotizismus,” 2021, Accessed: Dec. 15, 2023. [Online]. Available: https://archiv.ub.uni-marburg.de/diss/z2021/0201
[12] J. X. Marc N. Potenza, “White matter integrity and five-factor personality measures in healthy adults,” *NeuroImage*, vol. 59, no. 1, pp. 800–807, Jan. 2012, doi: 10.1016/j.neuroimage.2011.07.040.
[13] A. Bjørnebekk, A. M. Fjell, K. B. Walhovd, H. Grydeland, S. Torgersen, and L. T. Westlye, “Neuronal correlates of the five factor model (FFM) of human personality: Multimodal imaging in a large healthy sample,” *NeuroImage*, vol. 65, pp. 194–208, Jan. 2013, doi: 10.1016/j.neuroimage.2012.10.009.
[14] T. Booth *et al.*, “Personality, health, and brain integrity: The Lothian Birth Cohort Study 1936,” *Health Psychology*, vol. 33, no. 12, pp. 1477–1486, 2014, doi: 10.1037/hea0000012.
[15] C. Rodriguez *et al.*, “Structural Correlates of Personality Dimensions in Healthy Aging and MCI,” *Frontiers in Psychology*, vol. 9, 2019, Accessed: Jan. 30, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02652
[16] H. Sanjari Moghaddam *et al.*, “Microstructural white matter alterations and personality traits: A diffusion MRI study,” *Journal of Research in Personality*, vol. 88, p. 104010, Oct. 2020, doi: 10.1016/j.jrp.2020.104010.
[17] R. Avinun, S. Israel, A. R. Knodt, and A. R. Hariri, “Little evidence for associations between the Big Five personality traits and variability in brain gray or white matter,” *NeuroImage*, vol. 220, p. 117092, Oct. 2020, doi: 10.1016/j.neuroimage.2020.117092.
[18] J. D. Clayden, A. J. Storkey, and M. E. Bastin, “A Probabilistic Model-Based Approach to Consistent White Matter Tract Segmentation,” *IEEE Transactions on Medical Imaging*, vol. 26, no. 11, pp. 1555–1561, Nov. 2007, doi: 10.1109/TMI.2007.905826.

[19] A. M. McIntosh *et al.*, “Neuroticism, depressive symptoms and white-matter integrity in the Lothian Birth Cohort 1936,” *Psychological Medicine*, vol. 43, no. 6, pp. 1197–1206, Jun. 2013, doi: 10.1017/S003329171200150X.
[20] G. J. Lewis *et al.*, “Trait conscientiousness and the personality meta-trait stability are associated with regional white matter microstructure,” *Social Cognitive and Affective Neuroscience*, vol. 11, no. 8, pp. 1255–1261, Aug. 2016, doi: 10.1093/scan/nsw037.
[21] J. Privado, F. J. Román, C. Saénz-Urturi, M. Burgaleta, and R. Colom, “Gray and white matter correlates of the Big Five personality traits,” *Neuroscience*, vol. 349, pp. 174–184, May 2017, doi: 10.1016/j.neuroscience.2017.02.039.
[22] S. N. Sotiropoulos and A. Zalesky, “Building connectomes using diffusion MRI: why, how and but,” *NMR in Biomedicine*, vol. 32, no. 4, p. e3752, 2019, doi: 10.1002/nbm.3752.
[23] Y. Pang *et al.*, “Extraversion and Neuroticism Related to Topological Efficiency in White Matter Network: An Exploratory Study Using Diffusion Tensor Imaging Tractography,” *Brain Topogr*, vol. 32, no. 1, pp. 87–96, Jan. 2019, doi: 10.1007/s10548-018-0665-4.
[24] N. Talaei and A. Ghaderi, “Integration of structural brain networks is related to openness to experience: A diffusion MRI study with CSD-based tractography,” *Frontiers in Neuroscience*, vol. 16, 2022, Accessed: Dec. 15, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2022.1040799
[25] T. Yarkoni and J. Westfall, “Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning,” *Perspect Psychol Sci*, vol. 12, no. 6, pp. 1100–1122, Nov. 2017, doi: 10.1177/1745691617693393.
[26] J. Dubois and R. Adolphs, “Building a Science of Individual Differences from fMRI,” *Trends in Cognitive Sciences*, vol. 20, no. 6, pp. 425–443, Jun. 2016, doi: 10.1016/j.tics.2016.03.014.
[27] G. Feng *et al.*, “Methodological evaluation of individual cognitive prediction based on the brain white matter structural connectome,” *Human Brain Mapping*, vol. 43, no. 12, pp. 3775–3791, 2022, doi: 10.1002/hbm.25883.
[28] H. W. Yeung *et al.*, “Predicting sex, age, general cognition and mental health with machine learning on brain structural connectomes,” *Human Brain Mapping*, vol. 44, no. 5, pp. 1913–1933, 2023, doi: 10.1002/hbm.26182.
[29] H. He *et al.*, “Model and Predict Age and Sex in Healthy Subjects Using Brain White Matter Features: A Deep Learning Approach,” in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, Mar. 2022, pp. 1–5. doi: 10.1109/ISBI52829.2022.9761684.
[30] J. W. M. Domhof, K. Jung, S. B. Eickhoff, and O. V. Popovych, “Parcellation-based structural and resting-state functional brain connectomes of a healthy cohort (v1.1).” EBRAINS, 2022. doi: 10.25493/NVS8-XS5.
[31] D. K. Jones, T. R. Knösche, and R. Turner, “White matter integrity, fiber count, and other fallacies: The do’s and don’ts of diffusion MRI,” *NeuroImage*, vol. 73, pp. 239–254, Jun. 2013, doi: 10.1016/j.neuroimage.2012.06.081.
[32] R. E. Smith, J.-D. Tournier, F. Calamante, and A. Connelly, “SIFT2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography,” *NeuroImage*, vol. 119, pp. 338–351, Oct. 2015, doi: 10.1016/j.neuroimage.2015.06.092.
[33] 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.
[34] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, and K. Ugurbil, “The WU-Minn Human Connectome Project: An overview,” *NeuroImage*, vol. 80, pp. 62–79, Oct. 2013, doi: 10.1016/j.neuroimage.2013.05.041.
[35] M. F. Glasser *et al.*, “The minimal preprocessing pipelines for the Human Connectome Project,” *NeuroImage*, vol. 80, pp. 105–124, Oct. 2013, doi: 10.1016/j.neuroimage.2013.04.127.

[36] S. M. Smith *et al.*, “Advances in functional and structural MR image analysis and implementation as FSL,” *NeuroImage*, vol. 23, pp. S208–S219, Jan. 2004, doi: 10.1016/j.neuroimage.2004.07.051.
[37] J.-D. Tournier *et al.*, “MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation,” *NeuroImage*, vol. 202, p. 116137, Nov. 2019, doi: 10.1016/j.neuroimage.2019.116137.
[38] B. Jeurissen, J.-D. Tournier, T. Dhollander, A. Connelly, and J. Sijbers, “Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data,” *NeuroImage*, vol. 103, pp. 411–426, Dec. 2014, doi: 10.1016/j.neuroimage.2014.07.061.
[39] R. E. Smith, J.-D. Tournier, F. Calamante, and A. Connelly, “Anatomically-constrained tractography: Improved diffusion MRI streamlines tractography through effective use of anatomical information,” *NeuroImage*, vol. 62, no. 3, pp. 1924–1938, Sep. 2012, doi: 10.1016/j.neuroimage.2012.06.005.
[40] R. R. McCrae and P. T. Costa, “A contemplated revision of the NEO Five-Factor Inventory,” *Personality and Individual Differences*, vol. 36, no. 3, pp. 587–596, Feb. 2004, doi: 10.1016/S0191-8869(03)00118-1.
[41] R. C. Gur *et al.*, “A cognitive neuroscience based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation,” *J Neurosci Methods*, vol. 187, no. 2, pp. 254–262, Mar. 2010, doi: 10.1016/j.jneumeth.2009.11.017.
[42] 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.
[43] R. S. Desikan *et al.*, “An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest,” *NeuroImage*, vol. 31, no. 3, pp. 968–980, Jul. 2006, doi: 10.1016/j.neuroimage.2006.01.021.
[44] X. Shen, F. Tokoglu, X. Papademetris, and R. T. Constable, “Groupwise whole-brain parcellation from resting-state fMRI data for network node identification,” *NeuroImage*, vol. 82, pp. 403–415, Nov. 2013, doi: 10.1016/j.neuroimage.2013.05.081.
[45] J. Dubois, P. Galdi, Y. Han, L. K. Paul, and R. Adolphs, “Resting-State Functional Brain Connectivity Best Predicts the Personality Dimension of Openness to Experience,” *Personality Neuroscience*, vol. 1, p. e6, Jul. 2018, doi: 10.1017/pen.2018.8.
[46] R. D. Markello *et al.*, “neuromaps: structural and functional interpretation of brain maps,” *Nat Methods*, vol. 19, no. 11, pp. 1472–1479, Nov. 2022, doi: 10.1038/s41592-022-01625-w.
[47] R. L. Buckner, F. M. Krienen, A. Castellanos, J. C. Diaz, and B. T. T. Yeo, “The organization of the human cerebellum estimated by intrinsic functional connectivity,” *Journal of Neurophysiology*, vol. 106, no. 5, pp. 2322–2345, Nov. 2011, doi: 10.1152/jn.00339.2011.
[48] J. Wu *et al.*, “Accurate nonlinear mapping between MNI volumetric and FreeSurfer surface coordinate systems,” *Human Brain Mapping*, vol. 39, no. 9, pp. 3793–3808, 2018, doi: 10.1002/hbm.24213.
[49] W.-T. Hsu, M. D. Rosenberg, D. Scheinost, R. T. Constable, and M. M. Chun, “Resting-state functional connectivity predicts neuroticism and extraversion in novel individuals,” *Soc Cogn Affect Neurosci*, vol. 13, no. 2, pp. 224–232, Feb. 2018, doi: 10.1093/scan/nsy002.
[50] G. L. Baum *et al.*, “The impact of in-scanner head motion on structural connectivity derived from diffusion MRI,” *NeuroImage*, vol. 173, pp. 275–286, Jun. 2018, doi: 10.1016/j.neuroimage.2018.02.041.
[51] 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.

[52] H. Garavan *et al.*, “Recruiting the ABCD sample: Design considerations and procedures,” *Developmental Cognitive Neuroscience*, vol. 32, pp. 16–22, Aug. 2018, doi: 10.1016/j.dcn.2018.04.004.
[53] Y.-W. Chen and T. Canli, “‘Nothing to see here’: No structural brain differences as a function of the Big Five personality traits from a systematic review and meta-analysis,” *Personal Neurosci*, vol. 5, p. e8, Aug. 2022, doi: 10.1017/pen.2021.5.
[54] M. Gell *et al.*, “The Burden of Reliability: How Measurement Noise Limits Brain-Behaviour Predictions.” bioRxiv, p. 2023.02.09.527898, Feb. 10, 2023. doi: 10.1101/2023.02.09.527898.
[55] A. Bessadok, M. A. Mahjoub, and I. Rekik, “Graph Neural Networks in Network Neuroscience,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 5833–5848, May 2023, doi: 10.1109/TPAMI.2022.3209686.
[56] 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,” *Human Brain Mapping*, vol. 42, no. 10, pp. 3102–3118, 2021, doi: 10.1002/hbm.25420.
[57] R. Jiang *et al.*, “Gender Differences in Connectome-based Predictions of Individualized Intelligence Quotient and Sub-domain Scores,” *Cereb Cortex*, vol. 30, no. 3, pp. 888–900, Mar. 2020, doi: 10.1093/cercor/bhz134.
[58] M. C. Litwińczuk, N. Muhlert, L. Cloutman, N. Trujillo-Barreto, and A. Woollams, “Combination of structural and functional connectivity explains unique variation in specific domains of cognitive function,” *NeuroImage*, vol. 262, p. 119531, Nov. 2022, doi: 10.1016/j.neuroimage.2022.119531.
[59] 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.
[60] A. Schaefer *et al.*, “Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI,” *Cerebral Cortex*, vol. 28, no. 9, pp. 3095–3114, Sep. 2018, doi: 10.1093/cercor/bhx179.
[61] C.-H. Yeh, D. K. Jones, X. Liang, M. Descoteaux, and A. Connelly, “Mapping Structural Connectivity Using Diffusion MRI: Challenges and Opportunities,” *Journal of Magnetic Resonance Imaging*, vol. 53, no. 6, pp. 1666–1682, 2021, doi: 10.1002/jmri.27188.
[62] L. Fan *et al.*, “The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture,” *Cereb Cortex*, vol. 26, no. 8, pp. 3508–3526, Aug. 2016, doi: 10.1093/cercor/bhw157.
[63] N. Tzourio-Mazoyer *et al.*, “Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain,” *Neuroimage*, vol. 15, no. 1, pp. 273–289, Jan. 2002, doi: 10.1006/nimg.2001.0978.
[64] Z. Zhang, G. I. Allen, H. Zhu, and D. Dunson, “Tensor network factorizations: Relationships between brain structural connectomes and traits,” *NeuroImage*, vol. 197, pp. 330–343, Aug. 2019, doi: 10.1016/j.neuroimage.2019.04.027.
[65] B. Fischl *et al.*, “Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain,” *Neuron*, vol. 33, no. 3, pp. 341–355, Jan. 2002, doi: 10.1016/s0896-6273(02)00569-x.
[66] W. Liu *et al.*, “Fiber Tract Shape Measures Inform Prediction of Non-Imaging Phenotypes.” arXiv, May 20, 2023. doi: 10.48550/arXiv.2303.09124.
[67] X. Shen *et al.*, “Using connectome-based predictive modeling to predict individual behavior from brain connectivity,” *Nat Protoc*, vol. 12, no. 3, Art. no. 3, Mar. 2017, doi: 10.1038/nprot.2016.178.
[68] R. Jiang *et al.*, “Connectome-based individualized prediction of temperament trait scores,” *NeuroImage*, vol. 183, pp. 366–374, Dec. 2018, doi: 10.1016/j.neuroimage.2018.08.038.

[69] C. Yoo, S. Park, and M. J. Kim, “Structural connectome-based prediction of trait anxiety,” *Brain Imaging and Behavior*, vol. 16, no. 6, pp. 2467–2476, Dec. 2022, doi: 10.1007/s11682-022-00700-2.
[70] P. J. Easterbrook, R. Gopalan, J. A. Berlin, and D. R. Matthews, “Publication bias in clinical research,” *The Lancet*, vol. 337, no. 8746, pp. 867–872, Apr. 1991, doi: 10.1016/0140-6736(91)90201-Y.
[71] K. Dickersin, “Publication Bias: Recognizing the Problem, Understanding Its Origins and Scope, and Preventing Harm,” in *Publication Bias in Meta-Analysis*, John Wiley & Sons, Ltd, 2005, pp. 9–33. doi: 10.1002/0470870168.ch2.

# Supplementary Materials

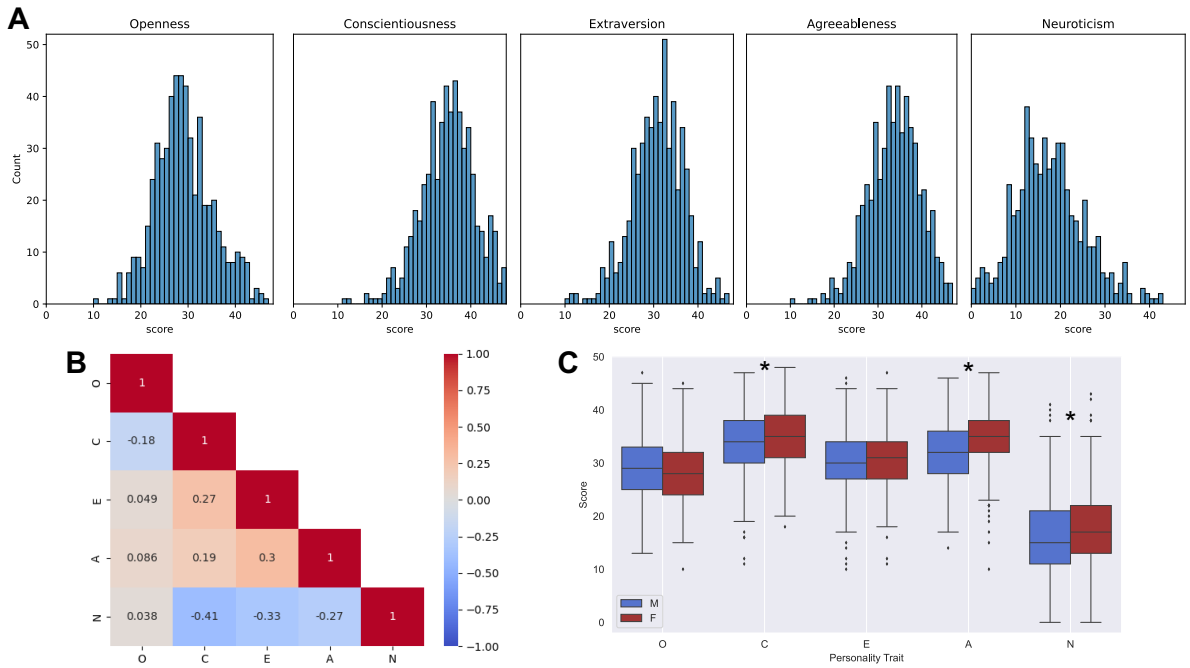



*Figure S1* **Supplementary information on the personality trait scores.** Only scores of subjects used in the analysis were used for the preparation of the figures. **A** The distribution of the big five personality trait scores as obtained through the NEO-FFI questionnaire. **B** The correlation between the different trait scores (O - Openness, C - Conscientiousness, E - Extraversion, A - Agreeableness, N - Neuroticism) **C** Comparison of trait distributions between males and females. For traits marked with an asterisk, the difference between the distributions of male and female trait scores was statistically significant with $p < 0.05$.

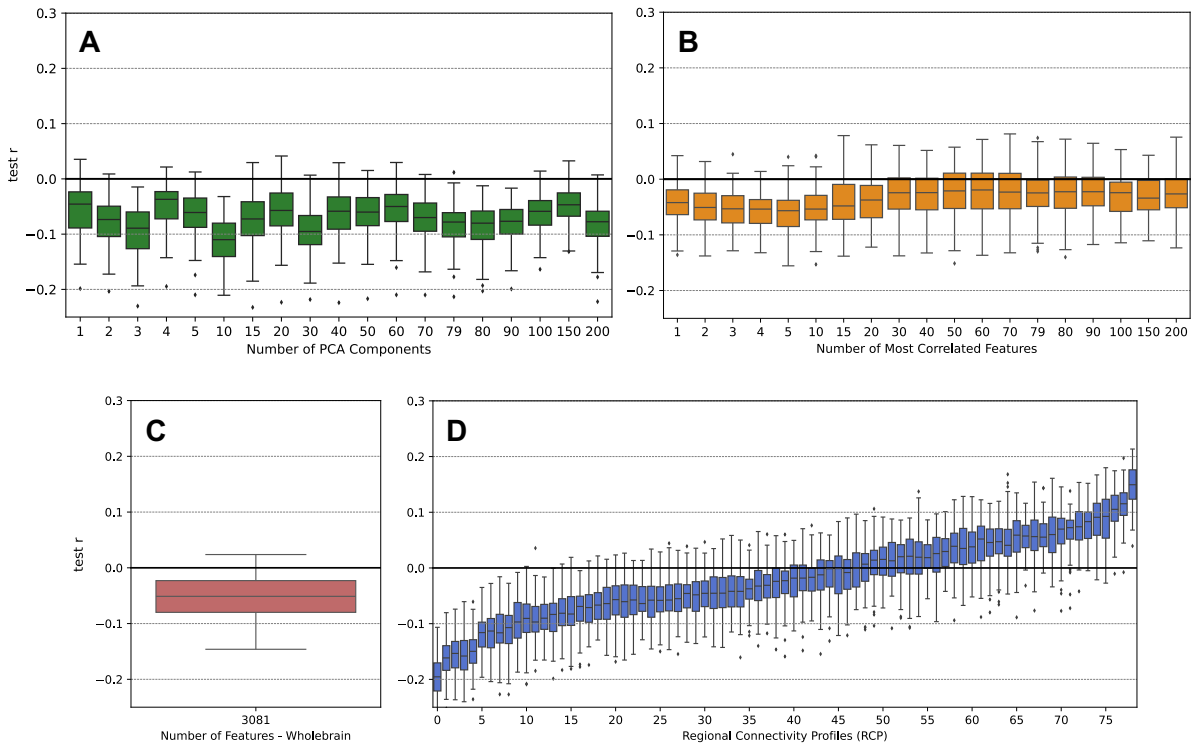



*Fig. S2* **Examples of prediction results for all four feature classes.** The calculations were performed by the pipeline with the Shen atlas (79 ROIs), NOS weighting, mixed-sex subject group, and the trait neuroticism. The box plots show the distributions of the prediction accuracy as given by Pearson's correlation between the predicted and empirical personality scores obtained for the test sets over 100 random splits of the data of the 5-fold cross-validation. The prediction results are depicted for **A** the PCA feature class with different numbers of PCA

components, **B** the *corr* feature class with different numbers of the most correlated features, **C** the whole-brain feature class and **D** the Regional Connectivity Profile (*RCP)* feature class for different brain regions, i.e., rows of the SC matrix. RCPs are sorted by the mean prediction accuracy. The respective conditions of the feature selection are indicated on the horizontal axes with the non-linear scaling in plots A and B.

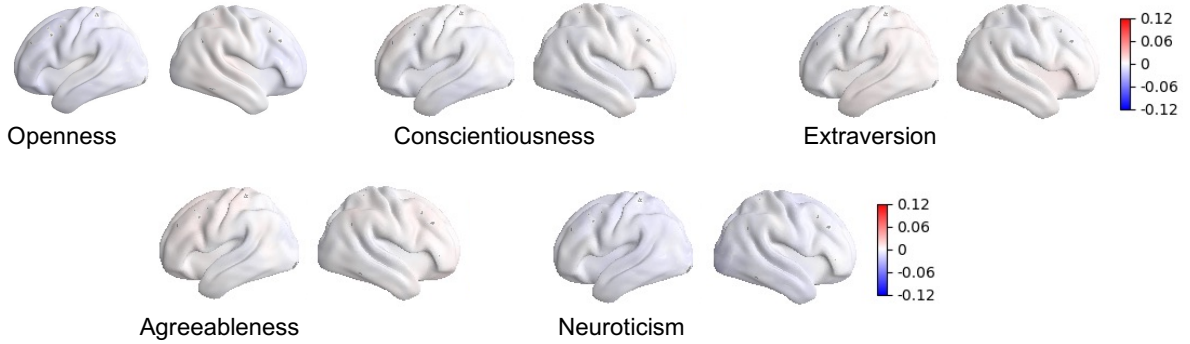



*Figure S3* **Prediction brain maps for permutation test predictions.** The prediction accuracy (correlation) of the RCP feature class was assigned to all voxels of the respective brain regions and averaged over all considered parcellations for fixed other conditions of the prediction pipeline (see Methods). The brain maps are illustrated for the five different personality traits indicated in the plots as prediction targets obtained for the mixed sex subject group and the NOS SC weighting for randomly permuted target values. The visualization of the maps was created using the neuromaps toolbox [1] including the volume-to-surface transformations as proposed and defined in [2], [3].

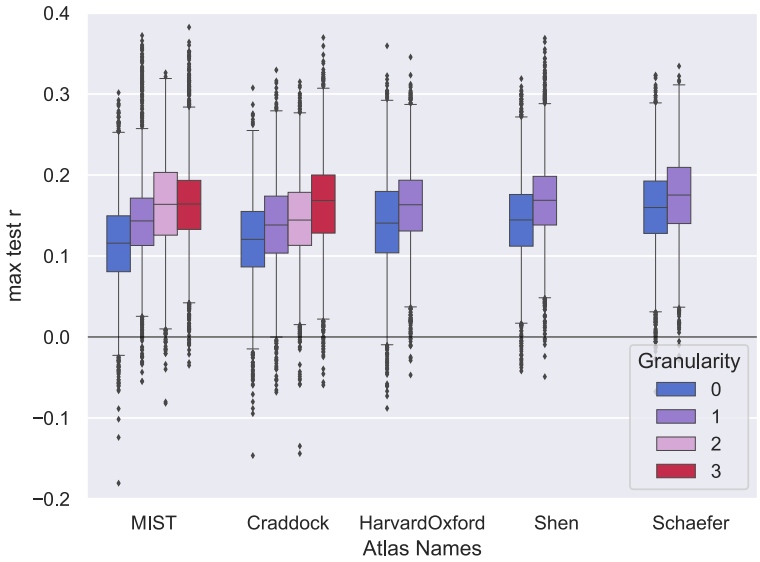



*Figure S4* **Comparison of brain parcellation granularity.** The figure compares the distributions of the maximum prediction correlations from pipelines using the RCP feature class for brain parcellations with the same parcellation scheme at different granularities. The distributions contain results from all personality traits and all connectome weightings.

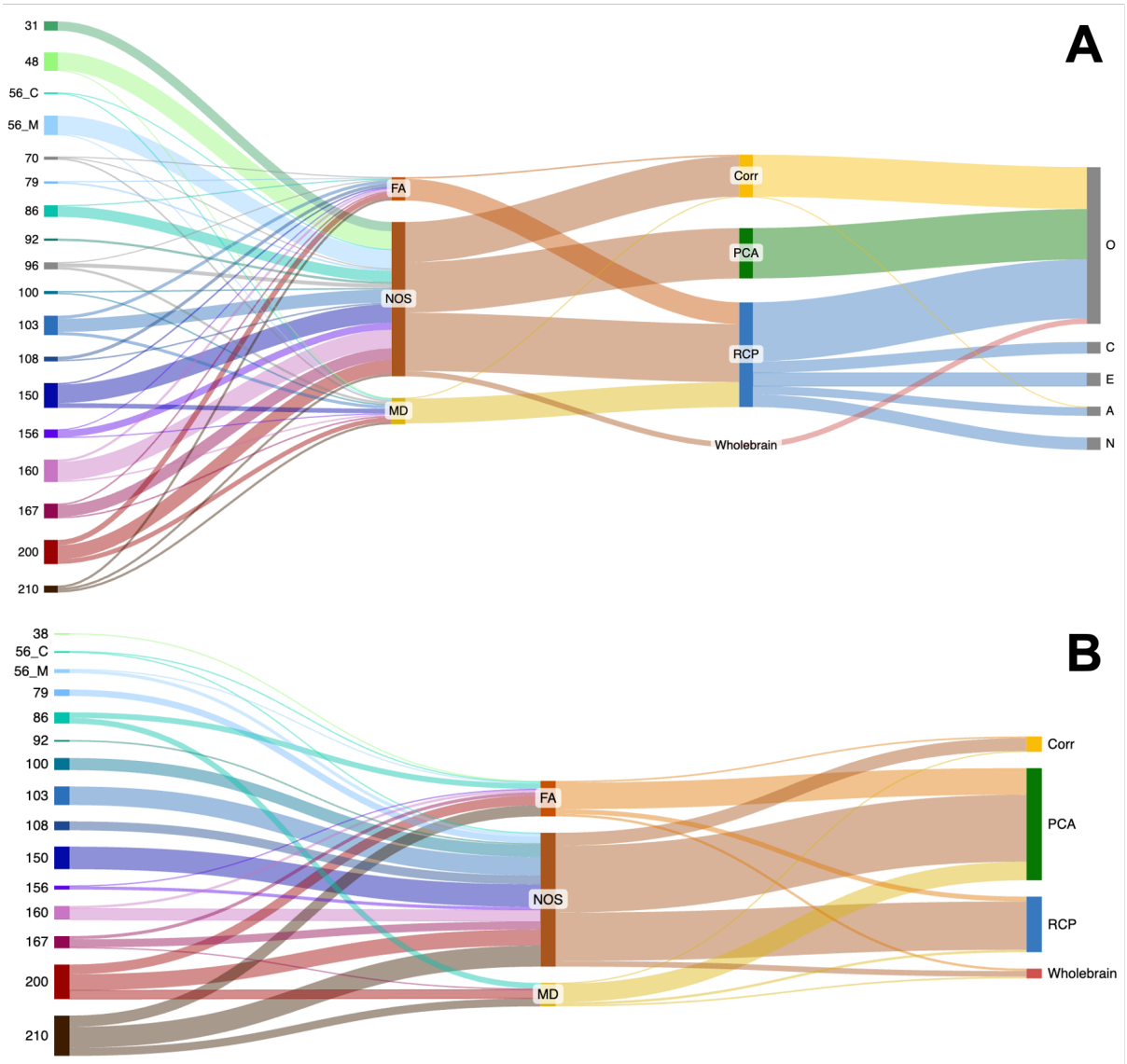



*Figure S5* **Sankey graphs of the best pipelines leading to test set correlations r > 0.2.** On the left are the different parcellation schemes, followed by the three different SC weightings and then the feature selection methods and the different feature selections / representations. The larger a node, the more often it was part of a pipeline leading to a test set correlation >= 0.2. Diagram created using SankeyMATIC. **A** For personality trait scores (O: Openness, C: Conscientiousness, E: Extraversion, A: Agreeableness, N: Neuroticism). **B** For the cognition target.

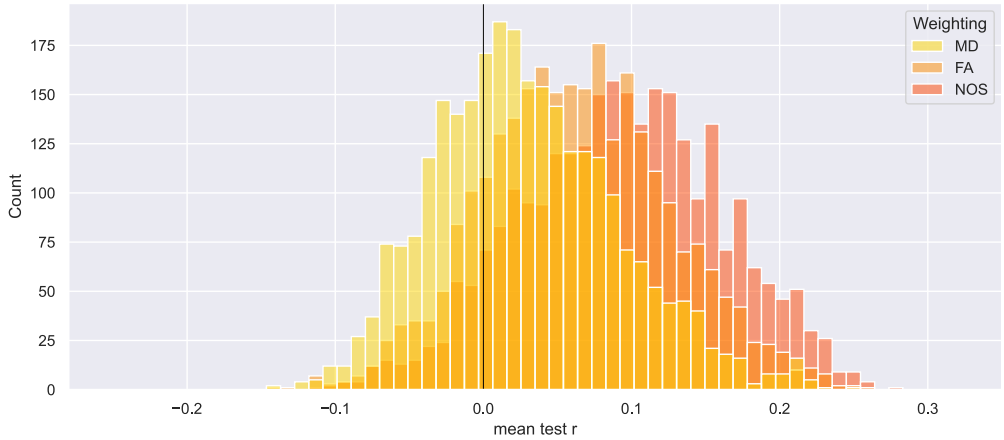
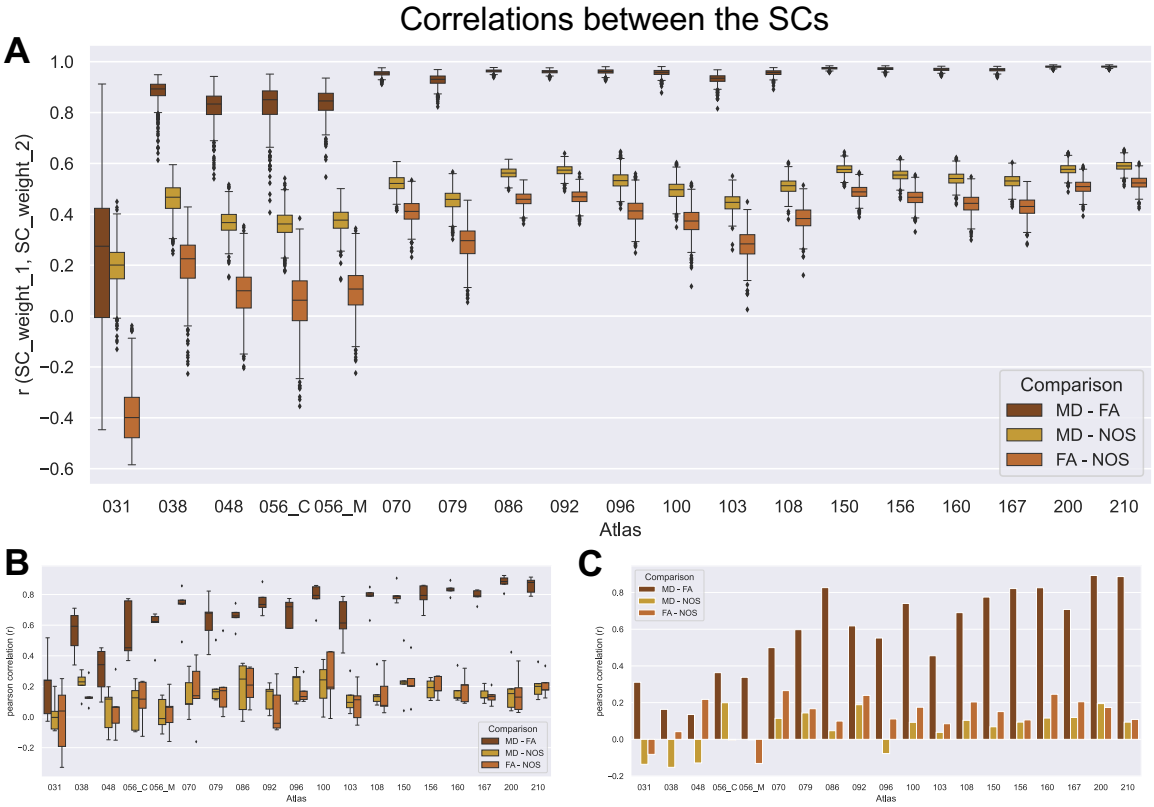



*Figure S6* Distribution of mean test set correlations for the prediction of cognition separated by connectome

weighting. Effect size between the distributions was calculated as Cohen's d: $d_{MD\text{-}FA}$ = 0.55, $d_{FA\text{-}NOS}$ = 0.46, $d_{MD\text{-}NOS}$ = 0.99.




Correlations between test r maps personality

Correlations between test r maps cognition

*Figure S7* **A** Pearson correlation r between the SCs with different weightings for every individual cortical brain atlas. The correlation was calculated for each subject included in the study (n=560), leading to a distribution of correlation values. **B** Pearson correlation r between the test r maps for different connectome weightings obtained from the RCP feature representation predictions of personality trait scores for individual cortical brain parcellations. The distributions each contain 15 values (three different subject groups x five personality trait scores). **C** Pearson correlation r between the test r maps obtained from the RCP feature representation predictions of a cognition score for individual cortical brain parcellations. Here, there is only one correlation value and not a distribution of correlations as these predictions were only performed for the mixed sex subject group and one composite cognition score.

## TBSS Analysis

In addition to our prediction analysis, we ran a TBSS analysis for our data to obtain results which can directly be compared to published findings. The analysis was conducted for all three subject groups (mixed-sex, male only and female only) for the fractional anisotropy (FA) images. We used the FA images calculated with MRtrix3 [4] for the structural connectome (SC) construction and then used the implementation of the TBSS analysis provided by FSL [5]. All FA images were aligned to a standard space by applying a non-linear registration. Subsequently the mean across all aligned FA images was calculated to extract the mean FA skeleton from it. The skeleton was thresholded at a value of 0.2 and only voxels that were part of the thresholded skeleton mask were used in the following statistical analysis. A detailed description of the algorithm up to this point can be found in Ref. [6].

Voxel-wise cross-subject statistical analysis between FA and the different personality traits were performed with the *randomize* tool in FSL [5] which implemented a general linear model in conjunction with, here, 10,000 Monte Carlo simulations. Threshold-free cluster enhancement (TFCE) [7] was applied for final voxel-wise inference. All personality trait scores

were demeaned before the analysis and both positive and negative associations between trait scores and FA values were investigated.

Statistically significant findings are listed in Table 1. For each subject group, only one respective trait was found to significant correlate with a certain voxel clusters. For the mixed-sex subject group openness was significantly positively correlated with the FA values in one cluster of 45 voxels, for the female only subject group neuroticism was significantly negatively correlated with the FA in a cluster of 31 voxels, and for the male only subject group a cluster of 12 voxels was significantly correlated with conscientiousness trait scores. The table further gives information on the mean and max absolute Pearson's correlation across voxels in the cluster as well as the MNI coordinates of the center of gravity (COG) of the cluster. For all other traits in each of the subject groups, there were no significant findings. The previous literature relating personality to measures extracted from dwMRIs via TBSS only investigated mixed-sex subject groups. None of these approaches found only significant results for the openness trait as we did here for the mixed sex subject group: Several approaches found no significant results for openness ([8], [9], [10], [11]) and another approach found significant results for openness but also for agreeableness and neuroticism [12]. Jung et al. [13] found significant findings only for openness. However, they did not investigate any of the other traits. Overall, our findings only comprise comparably small clusters of voxels with low maximum absolute correlations for a few different traits for different subject groups. This therefore goes well in line with our analyses demonstrating weak and unstable prediction of personality traits from SC.

| | | | | | MNI152 coordinates of COG | | |
|---|---|---|---|---|---|---|---|
| Subject Group | Trait | Cluster Size | Mean r | Max \|r\| | X | Y | Z |
| All | O | 45 | 0.12 | 0.18 | -43 | -60 | -5 |
| F | N | 31 | -0.08 | \|-0.21\| | 44 | -59 | -7 |
| M | C | 12 | 0.14 | 0.14 | 37 | 17 | 37 |

*Table 1* **Significant results for the TBSS analysis.** Results are separated by subject groups (all – mixed-sex, F – females only, M – males only). Given is the trait for which significant results were found (O – Openness, N – Neuroticism, C – Conscientiousness), the cluster size in number of voxels, the mean Pearson's correlation between the FA values and demeaned personality trait scores across all voxels in the cluster, the maximum Pearson's correlation of all voxels in the cluster and the MNI coordinates of the center of gravity of the cluster.


[1] R. D. Markello *et al.*, "neuromaps: structural and functional interpretation of brain maps," *Nat Methods*, vol. 19, no. 11, pp. 1472–1479, Nov. 2022, doi: 10.1038/s41592-022-01625-w.

[2] R. L. Buckner, F. M. Krienen, A. Castellanos, J. C. Diaz, and B. T. T. Yeo, "The organization of the human cerebellum estimated by intrinsic functional connectivity," *Journal of Neurophysiology*, vol. 106, no. 5, pp. 2322–2345, Nov. 2011, doi: 10.1152/jn.00339.2011.

[3] J. Wu *et al.*, "Accurate nonlinear mapping between MNI volumetric and FreeSurfer surface coordinate systems," *Human Brain Mapping*, vol. 39, no. 9, pp. 3793–3808, 2018, doi: 10.1002/hbm.24213.

[4] J.-D. Tournier *et al.*, "MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation," *NeuroImage*, vol. 202, p. 116137, Nov. 2019, doi: 10.1016/j.neuroimage.2019.116137.

[5] S. M. Smith *et al.*, "Advances in functional and structural MR image analysis and implementation as FSL," *NeuroImage*, vol. 23, pp. S208–S219, Jan. 2004, doi: 10.1016/j.neuroimage.2004.07.051.

[6] S. M. Smith *et al.*, “Acquisition and voxelwise analysis of multi-subject diffusion data with Tract-Based Spatial Statistics,” *Nat Protoc*, vol. 2, no. 3, Art. no. 3, Mar. 2007, doi: 10.1038/nprot.2007.45.
[7] S. M. Smith and T. E. Nichols, “Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference,” *NeuroImage*, vol. 44, no. 1, pp. 83–98, Jan. 2009, doi: 10.1016/j.neuroimage.2008.03.061.
[8] R. Avinun, S. Israel, A. R. Knodt, and A. R. Hariri, “Little evidence for associations between the Big Five personality traits and variability in brain gray or white matter,” *NeuroImage*, vol. 220, p. 117092, Oct. 2020, doi: 10.1016/j.neuroimage.2020.117092.
[9] T. Booth *et al.*, “Personality, health, and brain integrity: The Lothian Birth Cohort Study 1936,” *Health Psychology*, vol. 33, no. 12, pp. 1477–1486, 2014, doi: 10.1037/hea0000012.
[10] C. Rodriguez *et al.*, “Structural Correlates of Personality Dimensions in Healthy Aging and MCI,” *Frontiers in Psychology*, vol. 9, 2019, Accessed: Jan. 30, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02652
[11] H. Sanjari Moghaddam *et al.*, “Microstructural white matter alterations and personality traits: A diffusion MRI study,” *Journal of Research in Personality*, vol. 88, p. 104010, Oct. 2020, doi: 10.1016/j.jrp.2020.104010.
[12] J. X. Marc N. Potenza, “White matter integrity and five-factor personality measures in healthy adults,” *NeuroImage*, vol. 59, no. 1, pp. 800–807, Jan. 2012, doi: 10.1016/j.neuroimage.2011.07.040.
[13] R. E. Jung, R. Grazioplene, A. Caprihan, R. S. Chavez, and R. J. Haier, “White Matter Integrity, Creativity, and Psychopathology: Disentangling Constructs with Diffusion Tensor Imaging,” *PLOS ONE*, vol. 5, no. 3, p. e9818, Mar. 2010, doi: 10.1371/journal.pone.0009818.