Title:

Unique individual factors shape brain hubs organization of the human functional connectome

Authors:

Fernanda L. Ribeiro[1*], Walter H. L. Pinaya[1,2], João R. Sato[1], Claudinei E. Biazoli Jr.[1]

1 – Center of Mathematics, Computing and Cognition, Universidade Federal do ABC, Santo Andre, Brazil

2 – Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, 16 De Crespigny Park, SE5 8AF, United Kingdom

***Corresponding author:**

E-mail address: fernanda.lenita@ufabc.edu.br

Postal address:

Centre of Mathematics, Computing and Cognition, Universidade Federal do ABC

São Bernardo do Campo - Brazil, 09606-045

Abstract

Functional connectome fingerprints are partially heritable. Specifically, a higher heritability of high-order associative networks, which are greatly variable between individuals, has been proposed. However, such an interpretation relies on linear assumptions for the heritability inference and does not take into account the influence of the number of nodes on connectome fingerprinting analyses, since associative networks tend to include more brain regions. To overcome these limitations, we considered stable individual factors to estimate the relative genetic influence on whole-brain and eigenvector centrality-based networks of fixed size using the Human Connectome Project twin dataset. Individual identification accuracies were found to be higher in networks including more central nodes (94%) than the ones comprising peripheral nodes (22%). Nonetheless, the inferred additive genetic influence was more prominent in networks comprising peripheral nodes (50%) than hubs' networks (28%). Our results suggest that a brain hub's backbone is individually unique and less heritable than the connectivity profile of a subset of peripheral regions.

Keywords: Connectome fingerprint; Intersubject variability; Functional network; Brain hubs; Twin study

Introduction

In the past few years, fMRI research is living a paradigm shift, moving from population inferences to the study of individual differences[1,2]. Previous research has paved the way for the study of individual variability in functional connectivity patterns of the human brain[3–5]. In this context, resting-state fMRI (rs-fMRI) showed to be particularly powerful in determining underlying differences in the wiring patterns of functional connectome (FC) profiles. Indeed, connectome-based individual predictions achieved identification accuracies as high as 99% when comparing functional connectivity matrices[4]. Hence, the endeavor to identify and to characterize the individual functional connectivity architecture has been shown to have an imperative place in the study of individual differences.

Recent and mounting evidence suggests that FC profiles are stable at different time scales[6–9]. This characteristic posits the FC not only as an individual marker due to the comparably low intrasubject variability but also as a powerful discriminatory measure characterized by the high intersubject variability. Gratton et al. (2018)[8] showed that despite functional networks displaying common organization features at the group-level, the similarity between functional networks substantially increased at the individual level when evaluating the same subject in different tasks and sessions. This evidence supports the fact that individual stable patterns are crucial to explaining the intersubject variability of functional networks. Therefore, these findings suggest that sources of intersubject variability are stable over time, acting as individual signatures or ‘fingerprints’.

Seghier and Price (2018)[10] refer to the presence of distinct sources of intersubject variability that differ in their timescale. In the lower bound are sources of variability due

to mood states and context. The medium to long-term sources of intersubject variability include functional patterns built from the intimate interaction of an individual with the environment and inherent factors[10], respectively. Interestingly, functional networks show distinct levels of intersubject variability. Networks comprised of higher-order associative cortical areas seem to remarkably contribute to the FC distinctiveness[3–6,11], which, in turn, might be due to a higher intersubject[5,8] and lower intrasubject variability[12,13]. On the other hand, functional connectivity within networks that comprises primary sensory and motor regions was more variable across sessions of the same individual and less variable across individuals[5,8,12]. This might suggest that functional connections among brain regions within these high-order associative networks might be particularly determined by medium to long-term factors, such as genetics, in opposite to the primary sensorimotor networks.

Alternatively, the underlying organization supporting the coupling of these functional networks might also reveal the importance of different sources of intersubject variability[14]. Highly connected brain regions (hubs) are critical for the overall network organization and brain function, being important to link different functional networks in the brain[15]. Previous studies support that "brain hubs" are heavily connected among themselves and are thought to be the underlying structure for information transfer through different brain systems[16,17]. Interestingly, it is believed that this dense network continuously passes through developmental changes throughout adolescence[17]. Altogether, this suggests that the functional connectivity patterns underlying brain hubs function might comprehend high intersubject variability, and we questioned whether genetic and unique individual factors critically determine these patterns.

In fact, the intersubject variability in functional connectivity has been shown to be crucially influenced by genetic factors[6,18]. Miranda-Dominguez et al. (2018), using a

machine-learning algorithm and *a priori* defined functional networks, reported evidence for a heritable functional connectome, with a particularly higher contribution of high-order systems such as attentional and default mode networks. However, the extent to which genetic factors differentially influence the development of individual signatures in the FC remains elusive, for two main reasons. First, the underlying assumption of most models, which is that the phenotype is linearly affected by genetic and environmental factors, produce disparate estimates[6,18–20]. These estimates obscure the understanding of whether genetic factors are critical in determining intersubject variability, as most of them do not consider individual variability. Second, as resting-state networks have a different number of nodes and, consequently, of possible connections, differences in variability and identification accuracy between the networks cannot be directly interpreted as differences in heritability. Here, we sought to overcome these caveats by quantifying the relative effect of genetic factors on individual signatures in the FC by comparing pairs of whole-brain and eigenvector centrality-based (with fixed node number) networks, besides conventional functional networks.

In the first part of the study, we (1) reproduced the identification analysis introduced by Finn et al. (2015) to obtain the mean similarity value of FC profiles from pairs of subjects. Following, we (2) characterized the increase in similarity when comparing pairs of FC profiles with a different genetic relationship and (3) quantified the relative genetic factors influence on individual signatures in functional connectivity. In the second part of our study, we verified (4) whether the connections among the most influential nodes, as defined by eigenvector centrality, comprehended higher intersubject variability and (5) whether they were more susceptible to the influence of genetics. Thus, we hypothesized that (i) networks composed of the most influential

nodes (i.e. brain hubs) have specifically higher intersubject variability and (ii) are more susceptible to the influence of genetic factors.

Results

Part 1: Method reproduction and characterization of intersubject variability

Functional connectivity based identification analyses

Individual identification. Connectivity-based identification was performed comparing pairs of resting-state scans[4]. The individual identification was determined by computing the Pearson's correlation of each individual connectivity matrix from a 'target' set (n=380) with all the other connectivity matrices from the database (n=380), giving two sources of identification accuracy estimation (REST1xREST2 and REST2xREST1). Following that, the maximum correlation value among all comparisons between the target matrix and each of the FC matrices from the database should correspond to the correlation of the functional connectivity matrices of the same subject in different sessions. This process was repeated for all functional connectivity matrices within the 'target' set (Figure 1). The accuracy of the method was defined by the proportion of correct predicted subjects.

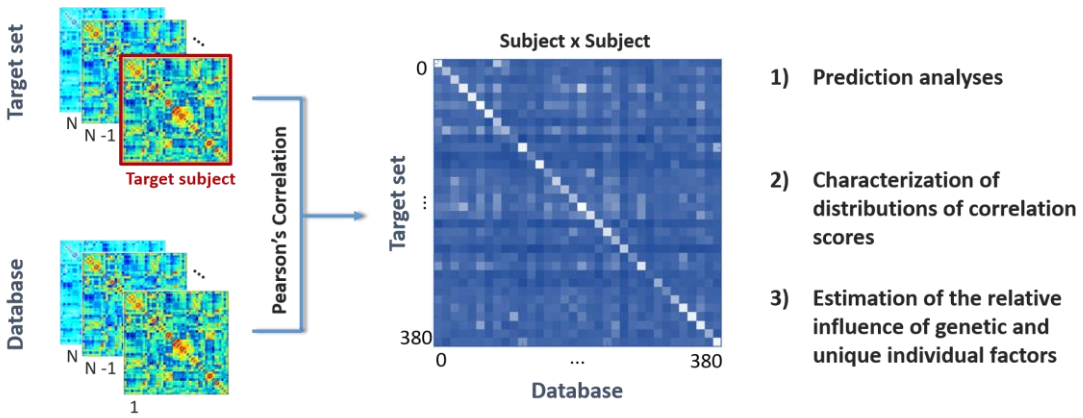

Figure 1 – Connectome-based identifications. For all analyses conducted at the first part of this study, we first grouped all the functional connectivity matrices from different sessions into two datasets, which could be either the 'target' set or the database. In the illustration, the 'target' set comprehends all functional connectivity matrices from the first resting-state session (REST1), while the database consists of the functional connectivity matrices from the

second resting-state session (REST2). Following that, we computed the Pearson's correlation of each individual connectivity matrix from a 'target' set with each connectivity matrix from the database. Therefore, each row within the subjectXsubject matrix contains the correlation scores between a target's FC (line) and each functional connectivity matrix of the database (column). With this, we were able to perform prediction analyses, to characterize the distribution of correlation scores and, finally, to estimate the relative influence of genetic and unique individual factors onto functional connectivity patterns.

Individual identifications were performed using the whole-brain connectivity matrices determined by two parcellation schemas: "Shen"[21] (268 nodes, 71,824 edges) and "Gordon"[22] (333 nodes, 110,889 edges). In the main transcript, the results reported were obtained with "Shen" parcels, but appropriate reference will be given to the supplementary material of similar results obtained with "Gordon" parcels. Thus, the resulting individual prediction accuracy was 97.8% (SD = 0.6%). This result is in agreement with previous reports[4,23], which further supports the power of the method for individual identification.

Then, we investigated the relevance of specific functional networks defined *a priori* for individual predictions by sectioning the whole-brain connectivity matrix into sub-matrices of single networks. From the 8 functional networks defined[4], the most successful networks were the medial frontal (93.9 ± 0.7%), and frontoparietal (91.8 ± 0.4%) networks. For the remaining results, refer to the Supplementary Table 1. Thus, the individual identification based on functional networks suggests that associative networks are particularly more discriminative. Indeed, the higher accuracy on the identification of an individual's functional connectivity profile indicates a higher intersubject and low intrasubject variability in those networks in opposite to sensorimotor networks. However, since the number of nodes varies across the *a priori* defined functional networks, differences in identification accuracy might not be directly comparable.

Twin identification. Previous reports indicate that the functional connectivity among higher-order associative brain regions greatly vary across individuals[5,8], even though they are comparably more stable within an individual across sessions[12,13]. Thus, we hypothesized that sources of high intersubject and low intrasubject variability in the functional connectome were governed by genetic factors. In order to test this hypothesis, we sought to determine whether the FC profiles of pairs of twins were more similar compared to pairs of unrelated individuals by using connectome-based predictions.

First, we evaluated monozygotic twin identification by computing the correlation coefficients of the functional connectivity matrices of monozygotic individuals (n=246) within the 'target' set with all matrices in the database (246x380=93,480 comparisons). Our prediction was based on the selection of the highest correlation coefficient (excluding the correlation coefficients corresponding to the same individual's FC) for each 'target' subject vs. database iteration. The mean prediction accuracy was 57.2% (SD = 3.0%). This result indicates that the idiosyncratic FC profiles might be genetically determined and they are sufficiently stable so one could identify monozygotic twins well above chance. Indeed, we have performed a permutation testing, by exchanging twin pairs' identities 1,000 times, such that for each identification iteration a new twin pair identity was assigned. The maximum identification accuracy found through these 1,000 permutations was 0.8%, indicating that the performance is significantly different from the chance level (p-value < 0.001).

Later on, we investigated the ability of specific functional networks in discriminating a twin pair from pairs of unrelated individuals (Figure 2). At this stage, the most successful functional networks were the subcortical-cerebellum (28.6 ± 1.7%) and medial frontal (21.1 ± 2.6%) networks. Noteworthy, the most successful functional

networks on twin identification were the ones that best performed on individual predictions.

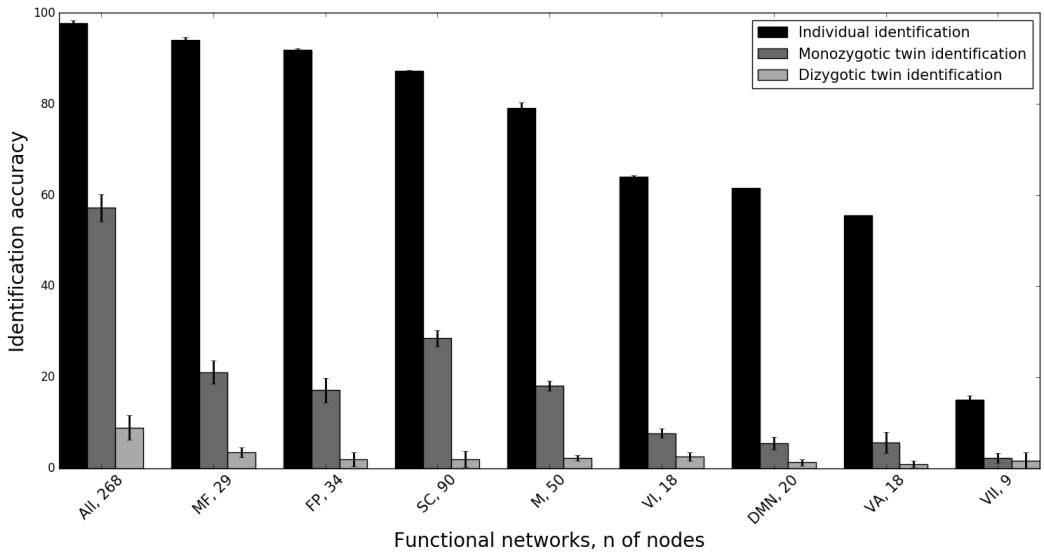

Figure 2 – Individual and twin identifications across pairs of functional connectivity matrices at the whole-brain and functional network levels. Identification accuracies for individual and twin prediction analyses considering the whole-brain (all) and each functional network connectivity patterns. Average identification for individual prediction was determined from two combinations of the database and 'target' sets (RESTXxRESTY, for X and Y = 1,2, X≠Y), while for twin prediction the mean identification was determined from four combinations (RESTXxRESTY, for X and Y = 1,2). Bar plots indicate average identification accuracy and error bars represent the standard deviation. All, whole-brain; MF, medial frontal; FP, frontoparietal; SC, subcortical-cerebellum; M, motor; VI, visual I; DMN, default mode network; VA, visual association; VII, visual II. We also present the number of nodes in each network.

Nonetheless, a substantial decrease in the successful twin identification rates was observed for functional networks when compared to the whole-brain connectome, and these results were particularly affected by the number of nodes within each network. The least successful functional networks on twin identification were the ones with the least number of nodes, while the networks with a larger number of nodes tended to present higher accuracies. This implies that the ability of *a priori* defined functional

networks to capture similarities in the FC profiles of monozygotic twins differentially relies on the amount of information provided (i.e., by the number of nodes).

Finally, we performed all the previous analyses for the dizygotic twin identification. At this time, we selected only the dizygotic individuals (n=134) within the 'target' set, giving 134x380=50,920 comparisons. For the whole-brain based identification, the mean prediction accuracy was 7.8% (SD = 2.3%). This abrupt change in twin identification accuracy indicates that the functional connectivity patterns of monozygotic twins are strictly more similar in comparison with dizygotic twins, which indicates the relevance of shared genetic background. We have also evaluated the individual functional networks on dizygotic twin identifications, and none of these functional networks were able to capture similarities in FC patterns between dizygotic twins (Figure 2).

Intra- and intersubject similarities in the functional connectome

In sequence, in order to characterize the intra- and intersubject (among unrelated individuals, monozygotic and dizygotic twin pairs) similarities for the whole-brain connectome and for each functional network, we summarized the correlation coefficients between all pairs of functional connectivity matrices (380x380=144,400 comparisons). We arranged the correlation coefficients in four groups according to their relationship: 1) same individual - SI (n=380); 2) monozygotic twins - MZ (n=246); 3) dizygotic twins - DZ (n=134) and 4) unrelated individuals - UN (n=143,640). The distributions of correlations across all these pairs at the whole-brain and each functional network level are illustrated in Figure 3 (Supplementary Figure 1).

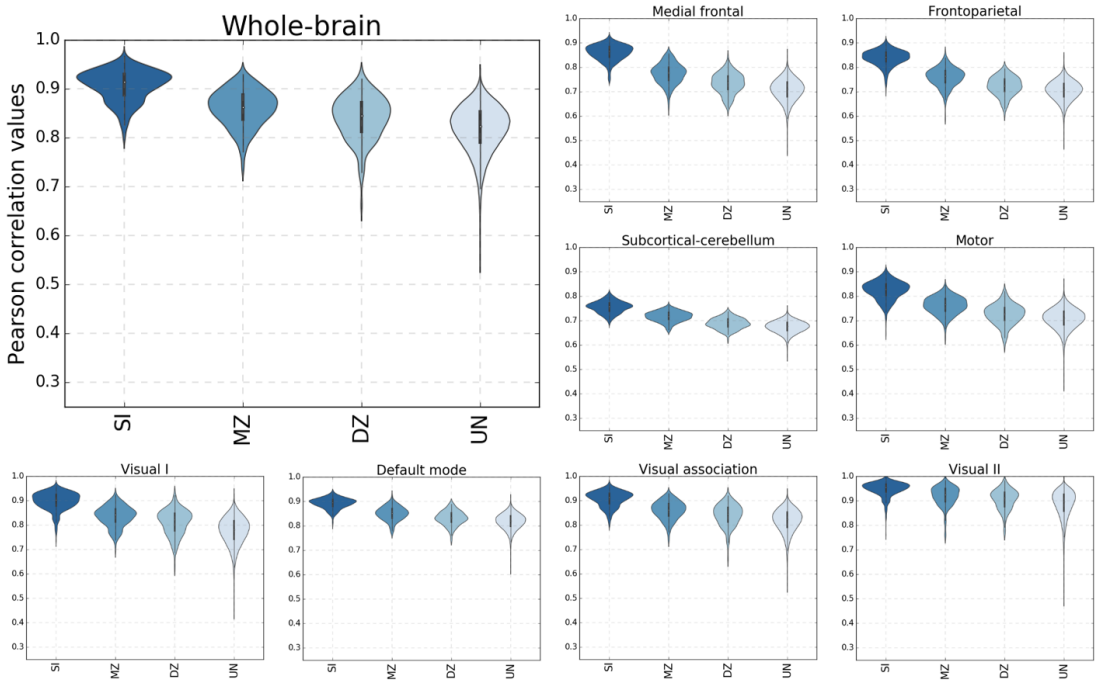

Figure 3 - Distribution of correlation coefficients between pairs of functional connectivity matrices for the whole-brain and individual functional networks. Pearson's correlation scores were determined from pairs of connectivity matrices (REST1*x*REST2) and they were grouped based on their relationship. Thus, violin plots show the distribution of the correlation scores determined from pairs of the same individual (SI), monozygotic twin (MZ), dizygotic twin (DZ) and unrelated subjects (UN) connectivity matrices. Connectivity matrices were determined for each functional network by selecting only the nodes assigned to it.

As one could expect, the mean of the distributions of correlation scores from the SI connectivity matrices is notably higher than the ones from the remaining distributions. This finding occurred not only at the whole-brain level but also at the functional network level, especially for medial frontal and frontoparietal functional networks. These patterns reflect the high accuracy on individual predictions, supporting the fact that the functional connectome might be an actual individual marker.

Additionally, we calculated the effect size of the difference among all groups using Cliff's delta[24]. For the whole-brain connectivity matrices, the effect size of the difference between the distributions of SI and UN correlation values was d=0.99. Similarly, the distributions of MZ correlation values are very unlike those distributions of UN correlations (d=0.89). At last, the distributions of DZ correlations are only

marginally different from the distributions of UN correlations (d=0.44), which is in agreement with the low accuracy found on dizygotic twins identification (Supplementary Figure 2). However, at the functional network level, the differences among MZ, DZ, and UN groups were considerably subtler.

Genetic factors influence stable individual signatures in functional connectivity profiles

Following, we sought to determine the portion of the increased similarity when we compared functional connectivity matrices from the same individual to the ones from unrelated individuals that was due to the influence of genetic factors. The increment in the mean similarity scores observed in SI in comparison to UN supports the presence of stable idiosyncratic patterns of functional connectivity. In addition, we could also observe higher similarity values within the MZ and DZ groups in comparison to unrelated individuals. Therefore, the increment of correlation scores within these groups might suggest that genetic and shared-environment factors might be important in determining unique FC profiles[25]. For this reason, we estimated the relative effect of those influences.

First, we assumed that the shared-environment and genetic factors that determined a higher similarity between DZ pairs in comparison to unrelated individuals are also present in MZ pairs[25]. Therefore, the further increment of correlation scores of MZ in comparison to DZ distributions might be a reasonable estimate of the relative influence of additional genetic factors on individual signatures in the FC profiles (Figure 4A). Thus, additional genetic factors explained 29% of the stable individual signatures at the whole-brain level (Figure 4B; Supplementary Figure 3). More specifically, among the best functional networks on individual identification, the medial frontal and frontoparietal networks had at least 26-27% of the increment of correlation scores in SI

group due to additional genetic factors. However, these networks were not the ones with the most substantial influence of shared genetics. Indeed, the subcortical-cerebellum (~35%) and motor (~31%) networks had the highest additional genetic factors' effect on individual patterns of their functional connectivity profiles (Figure 4B).

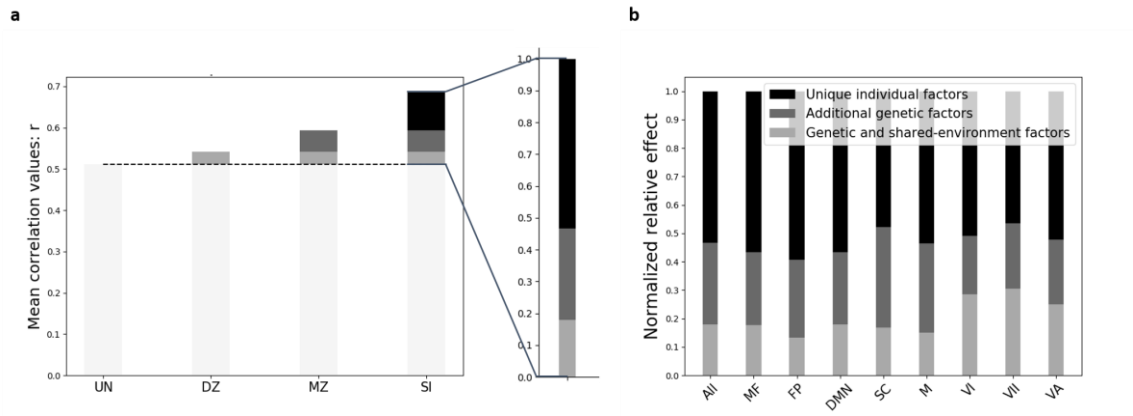

Figure 4 - The normalized relative effect of genetic and shared-environment factors on individual signatures in the functional connectome. (**a**) Representation of the average similarity, correlation scores of pairs of connectivity matrices, of each group: unrelated individuals – UN; dizygotic twins – DZ; monozygotic twins – MZ; and same individual – SI. The average similarity of UN was considered as the baseline. Thus, the increment of the average similarity of DZ over UN was considered to be due to genetic and shared-environment factors (light gray), the increment of the average similarity of MZ over DZ due to additional genetic factors (dark grey) and the increment of the average similarity of SI over MZ due to unique individual factors (black). These effects were normalized in relation to the overall increment of the similarity of SI over UN. (**b**) Normalized relative effects of genetic and shared-environment factors, additional genetic factors and unique individual factors on individual signatures in the whole-brain functional connectome and individual functional networks. All, whole-brain; MF, medial frontal; FP, frontoparietal; DMN, default mode network; SC, subcortical-cerebellum; M, motor; VI, visual I; VII, visual II; VA, visual association.

Part 2: Sources of intersubject variability

The individual and twin identifications seemed to be particularly affected by the number of nodes defining the functional connectivity matrices. The least successful functional networks were the ones with the least number of nodes, while the networks with a larger

number of nodes tended to present higher accuracies. This suggested that a more appropriate analysis was needed in order to compare different functional networks. Also, studies have been suggesting that the connectivity profile among the most central regions of the brain (nodes)[16] are the underlying structure responsible for transferring information across multiple functional systems (functional networks)[26]. Thus, we sought to study intersubject variability and the influence of genetic and unique individual factors on different organizational levels of brain networks (Figure 5).

Connections among the most central nodes comprehend the major source of intersubject variability and are less influenced by genetic factors

In order to test our hypotheses, we first determined the influence of each node in functional connectome profiles. To perform this, we computed the eigenvector centrality[27–29], which is a measure of the influence of a node in the network, of each node from the average functional connectivity matrix across subjects of the first resting-state session. Following that, we determined the minimum number of nodes necessary to reach at least 90% accuracy in individual identification (Figure 6; Supplementary Figure 4). For this, 60 nodes of the most central nodes were necessary to reach the performance (first panel in Figure 6A). On the other hand, 125 nodes of the least central nodes were necessary to reach the same performance (second panel in Figure 6A). The same analysis was done for monozygotic and dizygotic twin predictions, even though the maximum monozygotic identification rates were only achieved considering connectivity matrices determined by almost all nodes, identification rates for dizygotic twins were low at all levels.

To confirm the previous finding, we fixed the network size of 60 nodes and evaluated the change in identification accuracies when moving the network size window from the most to the least central nodes (step size=1 node). Here we see that as we move from

networks predominantly determined by the most central nodes to the ones determined by the least central nodes, individual and monozygotic twin identification accuracies decrease (first panel in Figure 6B). This result supports our hypothesis that connections among the most central nodes were actually the major source of intersubject variability, indicated by their power to discriminate among individuals.

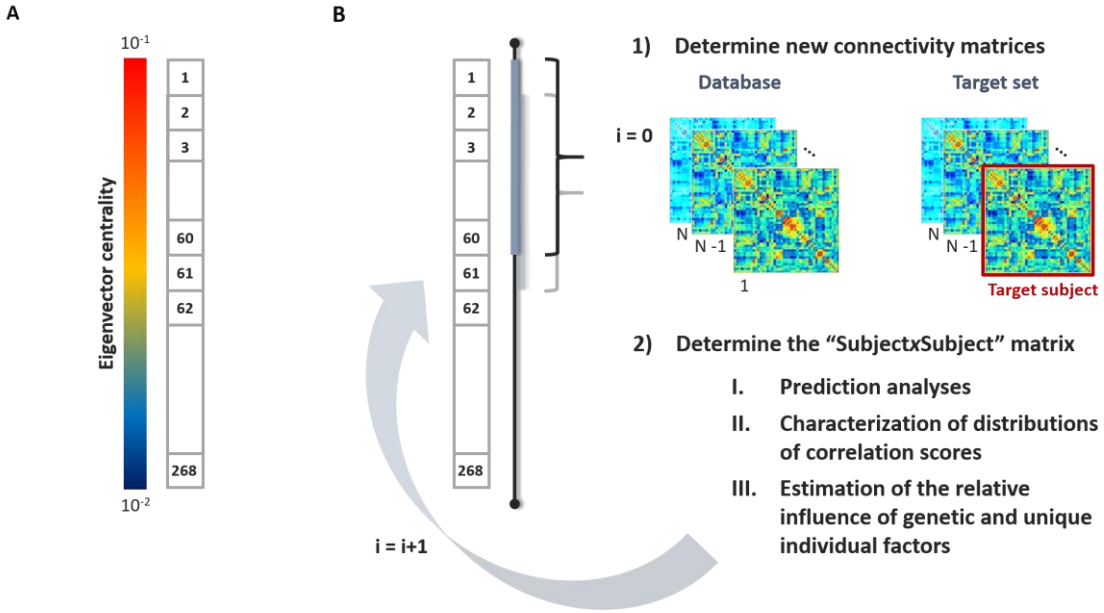

Figure 5 – Analyses based on new connectivity profiles determined by considering nodes centrality. First, the influence of different brain regions (nodes) was determined by computing their eigenvector centrality. Then, nodes were sorted based on their eigenvector centrality score, from the highest (red color in the gradient bar) to the lowest (blue color in the gradient bar). Following that, in an iterative process, new connectivity matrices were determined with the number of nodes fixed at 60, and all previous analyses were conducted. The fixed size window (60 nodes) moved from the most to the least central nodes with the step size equal to 1 node. Thus, in the first iteration, the connectivity matrices were determined by the 60 most central nodes. In the second iteration, the window moved and nodes selected ranged from the 2nd to the 61rst position on eigenvector centrality score. This process was repeated until the last iteration, where the connectivity matrices were determined by the 60 least central nodes. The total number of iteration was 208, which is defined by the total number of nodes *minus* the window size.

Finally, we estimated the additional genetic factors' effects on these fixed size networks (second panel in Figure 6B). Although the networks that were predominantly determined by the most central nodes were the best on monozygotic twin

identifications, they were actually the ones with the least contribution of additional genetic factors. Indeed, as the network size window moved from the most to the least central nodes, the additional genetic factors' effect on the functional connectivity patterns increased. Thus, the functional connectivity patterns among the least influential nodes might be particularly determined by genetic factors, while the connectivity patterns among the most influential nodes are more susceptible to unique individual factors.

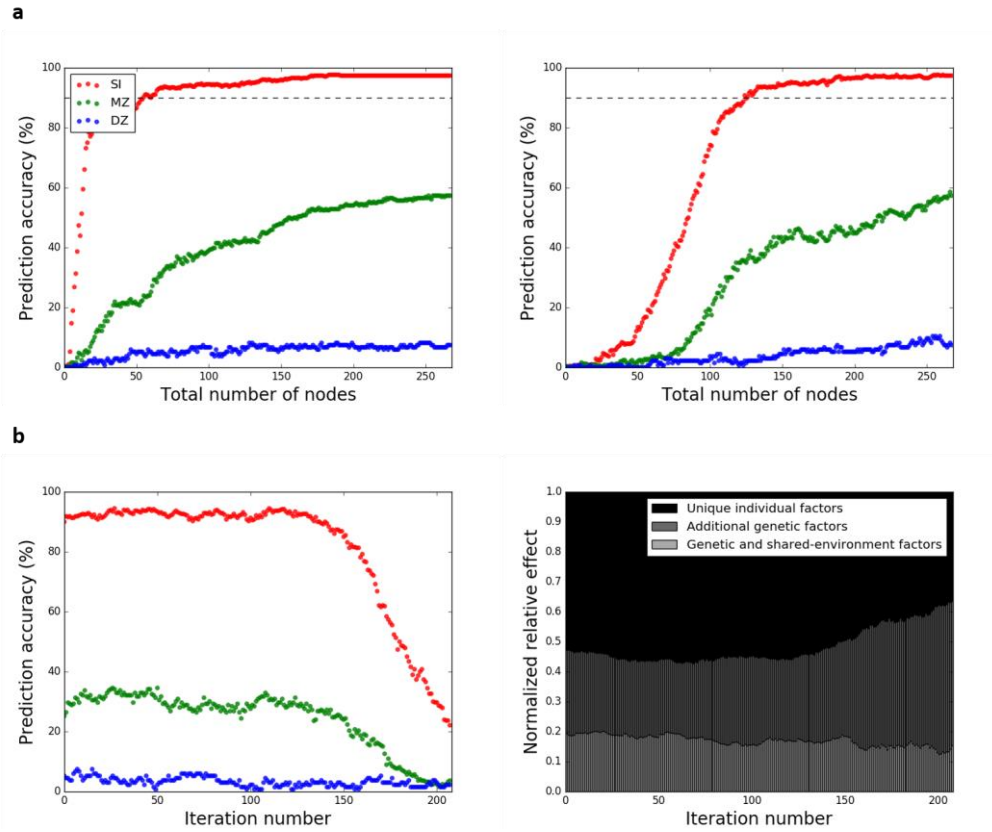

Figure 6 - Influence of nodes centrality on predictions and genetic factors' effects on patterns of functional connectivity. A) The first (from left to right) and second panel indicate changes on identification accuracies as the number of nodes determining functional connectivity matrices increased by adding nodes based on their eigenvector centrality, from nodes with the highest score to the least (first panel) and the other way round (second panel). Individual identification accuracy plot is in red, while monozygotic twin and dizygotic identification accuracies were plotted in green and blue, respectively. B) The number of nodes determining functional networks was fixed at 60. In the first panel, plots indicate the change on identification accuracies by moving from networks predominantly determined by the most central nodes to the ones determined by the least central nodes. The second panel gives the

estimates of unique individual factors, additional genetic factors, and genetic and shared-environment factors in individual signatures within the functional connectivity patterns of these networks.

Discussion

Genetic factors partially determine the individual architecture of functional connectivity profiles. As expected, we observed that the functional connectivity profiles of pairs of twins were more similar than of pairs of unrelated individuals. We also confirmed that different functional networks have distinct discriminatory power in connectome fingerprinting analyses[4,5,7,8,12,13]. However, these findings might not be due to valid individual differences[30], but to differences in the number of nodes defining functional networks[31]. Nonetheless, we found that the best functional networks on twin identifications were not the most heritable, in agreement with previous report[6]. Finally, identification analyses based on different organizational levels of brain networks suggest that the functional connectivity among brain hubs might comprehend a higher intersubject variability, which is particularly due to unique individual factors.

The genetic resemblance between twin pairs determines similarities in FC profiles

Functional connectivity matrices from pairs of genetically related individuals (monozygotic/dizygotic twins) were more similar than the ones from unrelated individuals. Also, the higher accuracy in the determination of monozygotic twin pairs in comparison to dizygotic twins suggests that this increment in the shared genetic background directly increases similarities in functional connectivity of genetically related individuals. Therefore, these findings indicate that individual stable patterns in the functional connectome might be determined by genetic factors.

Furthermore, the most distinctive functional networks[4,11] were also able to distinguish twins from unrelated individuals. Among the networks comprising higher-order associative cortical areas, within-network connectivity patterns were more similar among monozygotic twin pairs in comparison to pairs of unrelated individuals. In agreement, Miranda-Dominguez and colleagues (2018)[6] showed that the identification of genetically related individuals using a machine learning algorithm was mostly driven by data from high-order associative systems, including the frontoparietal, dorsal attention, ventral attention, and default mode networks.

Connectome-based identifications are limited by the number of nodes defining connectivity matrices

Connectome-based individual and twin identifications depend on the number of nodes defining connectivity matrices, and thus comparisons across different functional networks would likely lead to spurious results[32]. Here, we found that the least successful functional networks in identification analyses were the ones with the least number of nodes, while the networks with a larger number of nodes tended to give higher identification rates. Furthermore, the visual II network (smallest functional network defined) showed distributions of high correlation scores among pairs of connectivity matrices (Figure 3). Thus, these findings indicate that the different performances of a priori defined functional networks might rather be due to differences in the amount of information (edges) available to capture the similarity between pairs of connectivity matrices.

Genetic factors influence individual signatures in the functional connectome

In sequence, we estimated the relative contribution of genetic factors on individual signatures. Recently, Gratton et al. (2018) showed that functional networks were

particularly similar at the group-level comparisons, showing common organizational features. However, the similarity between functional networks substantially increased at the individual level[8], which supports that sources of intersubject variability are stable and are crucial to explaining the variability of functional networks between individuals. Our findings, therefore, sum to the current literature, as we have demonstrated that these stable individual signatures in brain functional connectivity are partly determined by genetic factors.

Additionally, we found that the functional networks that were comparably more susceptible to genetic factors were not the ones that best performed on twin identifications, in agreement with previous report[6]. By exploring the distribution of correlation scores between pairs of functional connectivity matrices, it became clear that even though the similarity among monozygotic twins' connectivity matrices was higher across the associative networks, the portion of the increment in similarity in the individual-level in comparison to the group level due to genetic factors was lower when compared to sensorimotor networks. Indeed, for sensory and motor networks, the subtle increase in the individual-level similarity was largely explained by genetic factors. Therefore, we found that although individual signatures showed to be highly dependent on genetic factors in many functional networks, a higher ability of a specific functional network to differentiate siblings from unrelated individuals does not directly determine a higher contribution of genetic factors in their functional connectivity arrangement.

However, considering the dependence on the number of nodes, direct comparisons between *a priori* defined functional networks might be misleading. Additionally, the connectivity profile among the most central regions of the brain (hubs)[16] constitutes the underlying structure responsible for transferring information across multiple functional systems (functional networks). Thus, we sought to study intersubject variability and the

influence of genetic and unique individual factors on different organizational levels of brain networks, with fixed numbers of nodes (Figure 5).

Hence, we sought to investigate the intersubject variability at different organizational levels of brain networks by considering nodes' eigenvector centrality. We hypothesized that the functional connections among brain hubs were the major source of intersubject variability. We confirmed that not only the number of nodes in the network is crucial for this identification framework, but also the relative functional "influence" of a set of nodes. In fact, by fixating the number of nodes defining functional networks, we were able to demonstrate that individual identification accuracy decreases when we move from networks mostly determined by the most central to the ones determined by the least central nodes. Previous findings indicated that regions that were predictors of individual differences in cognitive and behavioral domains comprehended high functional connectivity variability[5]. This might support the idea that the functional connectivity among brain hubs underpin human cognition, particularly due to their possible role in integrating information from different functional systems[26,29] and their high intersubject variability found here.

Additionally, we found that genetic factors influence on the stable individual signatures in the functional connectome was more prominent among connections between the least influential nodes, which were less variable across subjects. In fact, a lower portion of these individual signatures in the connectivity patterns among the most central nodes was explained by genetic factors. This finding is interesting as it suggests that intersubject variability, evolving from the unique interaction of the individual with the environment, is most prominent in a functional backbone of hub regions. In a recursive process, individual-environment interactions throughout development shape the functional wiring of such a hub network, which, in turn, might result in the complex and

unique human behaviors. Moreover, the FC architecture of a subset of less influential regions is highly genetically determined and presumably less permeable to environmental influences.

Parcellation schema

The individual and twin identification analyses resulted in high prediction accuracy using both parcellation schemas, "Shen" and "Gordon". Notably, the results obtained using "Shen" parcellation schema is about the same of previous studies[4,23], even though we had a more homogenous sample. At the network-level, associative networks were particularly better in discriminating between individuals. This result further supports a recent finding[8] showing that associative networks accommodate higher intersubject variability in comparison to sensorimotor networks. Despite that, we observed that the default mode network (DMN) defined by the parcellation schemas used in this work performed differently in the identification analyses. For “Gordon” schema, the DMN figured among the most distinctive networks, similarly to other associative networks. However, this pattern was not observed using "Shen" schema, in which the defined DMN figured among the worst functional network on individual predictions, similarly to other sensorimotor networks. Although for “Shen” schema the DMN figured among the networks with the least amount of nodes, for “Gordon” schema the DMN comprises the highest amount of nodes. Thus, these results suggest that the number of nodes is a confounding variable in individual identification analyses and it should be considered while comparing different functional networks. Furthermore, we observed a considerable difference in prediction accuracies among the functional networks defined by both schemas. Critically, the results concerning the intersubject variability and the

influence of genetic and unique individual factors on different organizational levels of brain networks were replicated with both parcellation schemas.

Limitations

The two data-driven parcellation schemas used were developed based on the average connectivity matrix across subjects (second parcellation scheme presented in the supplementary material). Thus, inherent difficulties in drawing conclusions at the individual-level arise. In fact, we first observed that differences in parcellation schemas resulted in differences in identification accuracy at the functional network level. The different definition of functional networks for each parcellation schema resulted in consistent differences across identification analyses, particularly for the default mode network. Therefore, a personalized parcellation[35] in the context of connectome-based predictions should be applied in future studies. Nonetheless, to study the intersubject variability of the functional connectivity patterns among nodes with different levels of centrality, we must first determine nodes centrality based on their common organization across subjects, and then evaluate the variability that emerges by looking at individual subjects. Thus, a personalized parcellation in this context would result in different and non-overlapping brain regions, adding another level of difficulty on drawing conclusions from the data.

Conclusions

Here, we sought to determine the genetic and shared-environment contribution on individual signatures in whole-brain connectome, pre-defined networks, and centrality-based functional connectivity patterns. Our results indicate that direct comparison among *a priori* defined functional networks, at least based on the framework adopted

here, is limited due to their irrevocable differences in the number of nodes assigned to them. Thus, we propose to overcome this limitation by studying the role of the eigenvector centrality-based networks on individual and twin prediction. Although the functional connectivity patterns among the hubs presented a high intersubject variability when compared to the ones associated with the least central nodes, they were less susceptible to genetic influences. These findings have important implications for the study of individual differences, suggesting that intersubject variability is not constrained to specific functional networks, but might be better captured by functional connectivity patterns among the most central nodes. This can be particularly relevant to better understand the relationship between unique functional connectome and cognitive and behavioral features.

Methods

Database and subject information

In this study, we used the dataset from the "1200 subjects data release" of the Human Connectome Project - HCP[36]. We restricted our analysis to monozygotic (MZ) and dizygotic (DZ) individuals. So, we initially selected all MZ and DZ individuals from the original sample. From this subsample, we excluded the subjects who did not have both resting-state fMRI sessions (ICA-FIX versions) available, and who did not have the twin within the group. Therefore, our final sample size was n=380. Table 1 summarizes the demographic data.

Table 1 - Subject demographic data.

	Monozygotic (n=246)	Dizygotic (n=134)
Age, y		
Mean ± SD	29.4 ± 3.3	29.1 ± 3.5
Range (min-max)	22 – 36	22 – 35
Sex, n(%)		
Female	144 (58.5)	78 (58.2)
Male	102 (41.5)	56 (41.8)

Data acquisition

The acquisition protocol has been previously described[36]. In summary, functional and structural data were acquired in a 3T Siemens Skyra scanner using a 32-channel head coil. Resting-state data were collected in two separated sessions (REST1 and REST2) in different days, each session containing two runs of 15 minutes. In this protocol, subjects had to keep their eyes open with a relaxed fixation on a projected bright cross-hair in a dark background. Each run within a session is distinguished by the oblique axial acquisition, of which one run used phase encoding in a right-to-left (RL) direction and the other used phase encoding in a left-to-right (LR) direction.

Data pre-processing

Pre-processing pipeline

For this study, we used the spatial and temporal pre-processed rs-fMRI timecourses[37,38], which have undergone the steps of artifact removal, motion correction, and registration to standard space. Furthermore, we applied additional pre-processing steps by using the CONN toolbox (v.17.f)[39], which included: structural segmentation, functional outlier detection (intermediate setting: 5 for z-score scan-to-scan global signal changes and 0.9

mm for scan-to-scan head-motion composite changes), and functional smoothing. Following that, a component-based noise correction method (CompCor)[40] and a temporal band-pass filtering (preserving frequencies between 0.01 and 0.10Hz) were applied. For spatial smoothing, a Gaussian with the full width at half maximum (FWHM) equal to 6mm was used. We also included subject movement parameters as first-level covariates to regress out their linear components from the BOLD time series.

Parcellations and functional networks

Timecourses were calculated as the mean signal within the regions of interest (ROIs) defined by different parcellation schemas used: “Gordon”[22] and “Shen”[21]. Both "Gordon" and "Shen" schemas are data-driven parcellation schemas. The first defines 333 ROIs clustered in 12 functional networks (Supplementary Table 1), in addition to 47 ROIs not assigned to any specific network. The latter defines 268 ROIs clustered in 8 networks (Supplementary Table 1).

Functional connectivity matrices

Finally, for the two resting-state sessions, data from both the left-right (LR) and right-left (RL) phase-encoding runs were used to calculate the connectivity matrices. To obtain the connectivity matrices, ROI-to-ROI bivariate correlation connectivity measures were computed for all ROIs defined by both parcellation methods, obtaining two symmetric connectivity matrices for each session for each subject.

Individual identification

The identification analysis was based on previous work[4] with few alterations. Initially, two databases were created containing the functional connectivity matrices for each session (REST1 and REST2). The individual identification was determined by computing the Pearson’s correlation of each individual connectivity matrix from one

database with all the other connectivity matrices from the second database (RESTXxRESTY, for X and Y = 1,2, X≠Y). For a pair of functional connectivity matrices linearly transformed in a column vector (vectorization), T_i and D_n, where T_i is the connectivity matrix of a target subject i, and D_n is the connectivity matrix of a subject (n=1, ..., 380) from the other database, the Pearson's correlation coefficient r is:

$$r_{i,N} = \frac{\sum_{j=1}^{e}(T_{i_j}-\bar{T}_i)(D_{N_j}-\bar{D}_N)}{\sqrt{\sum_{j=1}^{e}(T_{i_j}-\bar{T}_i)^2}\sqrt{\sum_{j=1}^{e}(D_{N_j}-\bar{D}_N)^2}} \qquad \text{Equation 1}$$

where *e* is the number of edges. In order to predict the identity of the target subject, the maximal Pearson's correlation coefficient was selected (Figure 1). Additionally, we also investigated the contribution of single networks to identification accuracy by sub-sectioning the functional connectivity matrices into sub-matrices of single networks. To perform this, we selected only connection within a specified network. Then, we calculated the Pearson's correlation coefficients, similarly to the previous approach. Results are reported as mean ± SD.

Twin identification

The twin pair identification algorithm was based on the previous individual identification analysis. At this stage, we removed the correlations corresponding to the same individual in different databases (sessions). Therefore, if the chosen maximum correlation value in each identification iteration belonged to the target subject's twin, the prediction was considered correct. Monozygotic and dizygotic twins were analyzed separately, and all conditions (RESTXxRESTY, for X and Y = 1,2) were tested. Results are reported as mean ± SD.

Statistical significance assessment

To assess the statistical significance of twin identification analyses, we performed a permutation testing. To assure the independence of the dataset, we permuted the twin pairs' identities, such that for each row of the SubjectxSubject matrix (Figure 1) a new twin pair identity was assigned, which corresponded to a different twin pair identity in the dataset. The permutation process was repeated 1,000 times.

Effect size

The distribution of correlation scores between pairs of connectivity matrices (i.e. correlation among the vectorized form of the connectivity matrices) were determined by grouping these scores based on familial relationship: 1) same individual - SI; 2) monozygotic twins - MZ; 3) dizygotic twins - DZ and 4) unrelated individuals - UN. Following that, the effect size of the differences between the distributions of correlation values was measured through the calculation of Cliff's delta. This a non-parametric effect size measure based on all pairwise differences[24], which gives how often values from one distribution are larger than the ones from a second distribution (Equation 2).

$$Delta\ (d) = \frac{Sum(x_1 > x_2) - Sum(x_1 < x_2)}{n_1 n_2} \qquad \textit{Equation 2}$$

Therefore, the number of times that values from one group are higher than the ones from a second group is calculated for all possible combinations of values between the two groups (n_1n_2, where n_1 and n_2 are the number of values within the distribution 1 and 2, respectively). The final Cliff's delta value is the difference between the previous calculations divided by all possible combinations. Thus, a positive and high value of d ($d_{maximum}$ = 1) mean that values within distribution 1 are mostly higher than the ones

within distribution 2, a negative and high absolute value of d ($d_{minimum}$ = -1) means the other way round, that values within distribution 1 are mostly lower than the ones within distribution 2, and d = 0 means that the distribution 1 and 2 are equal.

Genetic and shared-environment effects estimates based on the level of similarity

Genetic and shared-environment effects were estimated by determining their relative effects in the increment of the mean of the distribution of correlation values when comparing the same individual connectivity matrices in opposite to pairs of unrelated individuals. We built upon Gratton et al (2018) stability analysis, and the average similarity was examined for the whole-brain and functional network matrices that were from each group previously described (SI, DZ, MZ, and UN). Thus, the increment of the average similarity of each group (SI, MZ, and DZ) in comparison to the group of unrelated individuals was interpreted as due to:

1) Genetic and shared-environment factors: increment of the average similarity of pairs of DZ connectivity matrices over UN (Equation 3);

2) Additional genetic factors: increment of the average similarity of pairs of MZ connectivity matrices over DZ (Equation 4);

3) Unique individual factors: increment of the average similarity of pairs of SI connectivity matrices over MZ (Equation 5).

These effects were contrasted by determining their normalized relative effect on the increment of the average similarity among SI's pairs of connectivity matrices in comparison to unrelated individuals. In this procedure, the average similarity of unrelated individuals was considered the baseline. Thus, each of the effects was determined relative to the increment of the average similarity of SI group. The mathematical formulation of the procedure is as follows:

$$Genetic\ and\ shared - environment\ factors = \frac{\overline{DZ} - \overline{UN}}{\overline{SI} - \overline{UN}} \qquad Equation\ 3$$

$$Additional\ genetic\ factors = \frac{\overline{MZ} - \overline{DZ}}{\overline{SI} - \overline{UN}} \qquad Equation\ 4$$

$$Unique\ individual\ factors = 1 - \frac{\overline{MZ} - \overline{UN}}{\overline{SI} - \overline{UN}} \qquad Equation\ 5$$

Eigenvector centrality

For the subsequent analyses, we determined the influence of each node in functional connectome profiles by computing the eigenvector centrality. Thus, we first determined the average functional connectivity matrix across subjects in the first resting-state session (REST1) for each parcellation schema. To determine the eigenvector centrality, which is a metric of the influence of a node not only considering its degree of centrality (i.e the number of links) but also the importance of the nodes linked to it, we took the absolute values of each edge of the average functional connectivity matrix. Following that, the eigenvector centrality was determined by using the function *bct.eigenvector_centrality_und()* from the brain connectivity toolbox previously developed[41] and ported to python. In summary, the eigenvector centrality (x) of a node i (x_i) is defined as the i-th component of the eigenvector determined by the largest eigenvalue of the adjacency matrix, i.e. the functional connectivity matrix.

Effects of nodes centrality on identification rates

Finally, to evaluate the effect of the influence of the nodes on identification rates, new connectivity matrices were determined by increasing the number of nodes from both the most central to the least central nodes and the other way round. By doing so, we selected the minimum number of nodes such that connectome-based identification accuracies would reach at least 90%, which would be the fixed network size for the following analyses. In sequence, we evaluated the effect of the influence of the nodes and their connectivity patterns by shifting from networks predominantly determined by

the most influential nodes to the ones predominantly determined by the least influential nodes. In this procedure, the fixed network size was used as a window, which moved from the most to the least central nodes by one node step in each iteration. Individual and twin identification were computed for each iteration along with the average correlation scores of the pairs of connectivity matrices within the SI, MZ, DZ and UN groups, with which we determined genetic and shared-environment, additional genetic and unique individual factors influence in each network.

Code availability

Python scripts were written to perform all analyses described here. All scripts are available at https://github.com/felenitaribeiro/connectome-fingerprint-analyses.

Acknowledgments

This work was supported by the Universidade Federal do ABC (UFABC) and Coordination of Improvement of Higher Education Personnel (CAPES). We are grateful to the Human Connectome Project for their effort to increase the access to data of high quality.

Author contributions

F.L.R. and C.E.B. conceptualized the study. F.L.R. designed and performed the analyses with support from W.H.L.P. F.L.R. wrote the original draft. All authors revised and edited the manuscript. J.R.S., W.H.L.P., and C.E.B provided support and guidance with data interpretation.

Competing Interests

The authors declare no competing interests.

References

1. Seghier, M. L. & Price, C. J. Interpreting and Utilising Intersubject Variability in Brain Function. *Trends Cogn. Sci.* **xx,** 1–14 (2018).
2. Dubois, J. & Adolphs, R. Building a science of individual differences from fMRI. *Trends Cogn. Sci.* **20,** 425–443 (2016).
3. Miranda-Dominguez, O. *et al.* Connectotyping: Model based fingerprinting of the functional connectome. *PLoS One* **9,** (2014).
4. Finn, E. S. *et al.* Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18,** 1664–1671 (2015).
5. Mueller, S. *et al.* Individual Variability in Functional Connectivity Architecture of the Human Brain. *Neuron* **77,** 586–595 (2013).
6. Miranda-Dominguez, O. *et al.* Heritability of the human connectome: a connectotyping study. *Netw. Neurosci.* **2,** 175–199 (2018).
7. Sato, J. R., White, T. P. & Biazoli, C. E. Commentary: A test-retest dataset for assessing long-term reliability of brain morphology and resting-state brain activity. *Front. Neurosci.* **11,** 1–4 (2017).
8. Gratton, C. *et al.* Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. *Neuron* 439–452 (2018). doi:10.1016/j.neuron.2018.03.035
9. Pannunzi, M. *et al.* Resting-state fMRI correlations: From link-wise unreliability to whole brain stability. *Neuroimage* **157,** 250–262 (2017).
10. Seghier, M. L. & Price, C. J. Interpreting and Utilising Intersubject Variability in Brain Function. *Trends Cogn. Sci.* **22,** 517–530 (2018).
11. Kaufmann, T. *et al.* Delayed stabilization and individualization in connectome development are related to psychiatric disorders. *Nat. Neurosci.* **20,** 513–515 (2017).
12. Laumann, T. O. *et al.* Functional System and Areal Organization of a Highly Sampled Individual Human Brain. *Neuron* **87,** 658–671 (2015).
13. Poldrack, R. A. *et al.* Long-term neural and physiological phenotyping of a single human. *Nat. Commun.* **6,** (2015).
14. Grayson, D. S. & Fair, D. A. Development of large-scale functional networks from birth to adulthood: A guide to the neuroimaging literature. *Neuroimage* **160,** 15–31 (2017).

15. Power, J. D., Schlaggar, B. L., Lessov-Schlaggar, C. N. & Petersen, S. E. Evidence for hubs in human functional brain networks. *Neuron* **79,** 798–813 (2013).

16. van den Heuvel, M. P. & Sporns, O. Rich-Club Organization of the Human Connectome. *J. Neurosci.* **31,** 15775–15786 (2011).

17. Grayson, D. S. *et al.* Structural and functional rich club organization of the brain in children and adults. *PLoS One* **9,** 1–13 (2014).

18. Yang, Z. *et al.* Genetic and Environmental Contributions to Functional Connectivity Architecture of the Human Brain. *Cereb. Cortex* **26,** 2341–2352 (2016).

19. Colclough, G. L. *et al.* The heritability of multi-modal connectivity in human brain activity. *Elife* **6,** 1–19 (2017).

20. Ge, T., Holmes, A. J., Buckner, R. L., Smoller, J. W. & Sabuncu, M. R. Heritability analysis with repeat measurements and its application to resting-state functional connectivity. *Proc. Natl. Acad. Sci.* **114,** 5521–5526 (2017).

21. Shen, X., Tokoglu, F., Papademetris, X. & Constable, R. T. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage* **82,** 403–415 (2013).

22. Gordon, E. M. *et al.* Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cereb. Cortex* **26,** 288–303 (2014).

23. Waller, L. *et al.* Evaluating the replicability, specificity, and generalizability of connectome fingerprints. *Neuroimage* **158,** 371–377 (2017).

24. Cliff, N. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychol. Bull.* **114,** 494–509 (1993).

25. Thompson, P., Cannon, T. D. & Toga, A. W. Mapping genetic influences on human brain structure. *Ann. Med.* **34,** 523–536 (2002).

26. van den Heuvel, M. P. & Sporns, O. Network hubs in the human brain. *Trends Cogn. Sci.* **17,** 683–696 (2013).

27. Lohmann, G. *et al.* Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain. *PLoS One* **5,** (2010).

28. Zuo, X. N. *et al.* Network centrality in the human functional connectome. *Cereb. Cortex* **22,** 1862–1875 (2012).

29. Hwang, K., Hallquist, M. N. & Luna, B. The development of hub architecture in the human functional brain network. *Cereb. Cortex* **23,** 2380–2393 (2013).

30. Ren, Y., Nguyen, V. T., Guo, L. & Guo, C. C. Inter-subject Functional Correlation Reveal a Hierarchical Organization of Extrinsic and Intrinsic Systems in the Brain. *Sci. Rep.* **7,** 10876 (2017).

31. Byrge, L. & Kennedy, D. P. High-accuracy individual identification using a “thin slice” of the functional connectome. *Netw. Neurosci.* 1–48 (2018).

32. van Wijk, B. C. M., Stam, C. J. & Daffertshofer, A. Comparing brain networks of

different size and connectivity density using graph theory. *PLoS One* **5,** (2010).

33. Singh, K. D. & Fawcett, I. P. Transient and linearly graded deactivation of the human default-mode network by a visual detection task. *Neuroimage* **41,** 100–112 (2008).

34. Power, J. D. *et al.* Functional Network Organization of the Human Brain. *Neuron* **72,** 665–6678 (2011).

35. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nat. Publ. Gr.* **536,** 171–178 (2016).

36. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: An overview. *Neuroimage* **80,** 62–79 (2013).

37. Smith, S. M. *et al.* Resting-state fMRI in the Human Connectome Project. *Neuroimage* **80,** 144–168 (2013).

38. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80,** 105–124 (2013).

39. Whitfield-Gabrieli, S. & Nieto-Castanon, A. *Conn* : A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks. *Brain Connect.* **2,** 125–141 (2012).

40. Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37,** 90–101 (2007).

41. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* **52,** 1059–1069 (2010).