

Università degli Studi di Salerno



Dottorato di Ricerca in Informatica e Ingegneria  
dell'Informazione  
Ciclo 30 - a.a. 2016/2017

TESI DI DOTTORATO / PH.D. THESIS

Building functional neuromarkers from  
resting state fMRI to describe  
physiopathological traits

**Supervisor**

Prof. Roberto TAGLIAFERRI

**Candidate**

Paola GALDI

**Ph.D. Program Director**

Prof. Pasquale CHIACCHIO

Dipartimento di Ingegneria dell'Informazione  
ed Elettrica e Matematica Applicata  
Dipartimento di Informatica

A Michele,  
per avermi s[uo]pportato in questi anni.

# Acknowledgements

No scientist is an island. All the people I met during these years have contributed with their insights to my work and to my growth as a researcher. I would like to thank my supervisor, Prof. Roberto Tagliaferri, that has been my guide since my Bachelor's years and has introduced me to the wonders of artificial intelligence in one of his (long) digressions during a Computer Architecture class. Thanks to Prof. Fabrizio Esposito, who transmitted me his enthusiasm for neuroimaging data analysis, and to Prof. Michele La Rocca, who has trained me in the rigorous art of statistical validation. Thanks to Prof. Giancarlo Raiconi, a living reminder to study more Maths, and to Prof. Dario Greco, whose inspiring talks would encourage anybody to do science. Thanks to Ciro Donalek, George Djorgovski, Ashish Mahabal and all the staff of the Center for Data Driven Discovery, that hosted me during my first exciting experience at Caltech. Thanks to Julien Dubois, that has supervised my work in the last year and has provided me constant guidance and inspiration. Thanks to Prof. Ralph Adolphs, Lynn Paul and all the brilliant scientists I met at Adolphs Lab, who taught me that doing good science requires being a good person. A special thanks goes to Angela Serra, Michele Fratello and Luca Puglia, who as senior PhD students have helped me through my days at NeuRoNe Lab, and have lightened up my nights as cherished friends. Finally, thanks to my parents and my sisters, for their (somewhat biased) beliefs about my ability to succeed.

# Abstract

The overarching goal of this work has been that of devising novel methods for building functional neuromarkers from resting-state fMRI data to describe healthy and pathological human behaviour. Observing spontaneous fluctuations of the BOLD signal, resting-state fMRI allows to have an insight into the functional organisation of the brain and to detect functional networks that are consistent across subjects. Studying how patterns of functional connectivity vary both in healthy subjects and in subjects affected by a neurodegenerative disease is a way to shed light on the physiological and pathological mechanisms governing our nervous system.

The first part of this thesis is devoted to the description of fully data-driven feature extraction techniques based on clustering aimed at supporting the diagnosis of neurodegenerative diseases (e.g., amyotrophic lateral sclerosis and Parkinson's disease). The high-dimensional nature of resting state fMRI data implies the need of suitable feature selection techniques. Traditional univariate techniques are fast and straightforward to interpret, but are unable to unveil relationships among multiple features. For this reason, this work presents a methodology based on consensus clustering, a particular approach to the clustering problem that consists in combining different partitions of the same data set to produce more stable solutions. One of the objectives of fMRI data analysis is to determine regions that show an abnormal activity with respect to a healthy brain and this is often attained with comparative statistical models applied to single voxels or brain parcels within one or several functional networks. Here, stochastic rank aggregation is applied to identify brain regions that exhibit a coherent behaviour in groups of subjects affected by the same disorder. The proposed methodology was validated on real data and the results are consistent with previous literature, thus indicating that this approach might be suitable to support early diagnosis of neurodegenerative diseases.

The second part of the thesis is dedicated to the study of the link between brain functional connectivity and individual differences in intelligence and per-

sonality. Functional connectivity networks are built by parcelling the brain into regions according to some criterion, so that regions become the nodes of the network, and by adding a connection between two nodes if the time series of the involved brain regions are correlated. One of the challenge of functional connectivity estimation is that it can be greatly affected by motion artifacts and nowadays there is still no universal agreement on what is the best strategy for noise and motion artifact removal. For this reason, during this project a preprocessing pipeline was developed implementing the most used denoising steps found in literature, to study how different approaches affect the final outcome of the analysis. After preprocessing, starting from the connectivity network of each subject, regression models were built using the edges as features to predict IQ scores. The results have shown the presence of a correlation, although weak, between the strength of the connections and the intelligence score, but also how this value can be greatly affected by the presence of motion artefacts and experimental choices. The same framework was applied also to predict personality traits, to investigate the neural correlates of psychometric models currently adopted in personality psychology.

# Contents

<b>Acknowledgements</b>	<b>1</b>
<b>Abstract</b>	<b>2</b>
<b>List of original articles</b>	<b>5</b>
<b>1 Introduction</b>	<b>7</b>
<b>2 Feature extraction from rfMRI</b>	<b>11</b>
2.1 Background . . . . .	11
2.2 Methods . . . . .	13
2.2.1 Consensus clustering . . . . .	13
2.2.2 Consensus-based feature filtering . . . . .	14
2.2.3 Anatomical cluster filtering . . . . .	15
2.3 Data description . . . . .	15
2.4 Experimental setup . . . . .	16
2.5 Results and discussion . . . . .	18
2.6 Conclusion . . . . .	25
<b>3 Building neuromarkers from rfMRI</b>	<b>27</b>
3.1 Background . . . . .	28
3.2 Data description . . . . .	28
3.3 Overview of the methodology . . . . .	29
3.3.1 Brain parcellation . . . . .	29
3.3.2 Stochastic rank aggregation . . . . .	30
3.4 Results . . . . .	31
3.5 Discussion . . . . .	37
3.6 Conclusion . . . . .	39

---

<b>4</b>	<b>Predicting individual differences</b>	<b>41</b>
4.1	Denoising of rfMRI data . . . . .	42
4.1.1	Description of denoising pipelines . . . . .	43
4.2	Data description . . . . .	45
4.3	Intelligence prediction from rfMRI . . . . .	46
4.3.1	Methods . . . . .	47
4.3.2	Results . . . . .	48
4.4	Resting-state fMRI and personality . . . . .	49
4.4.1	Methods . . . . .	51
4.4.2	Results . . . . .	55
4.5	Conclusion . . . . .	56
<b>5</b>	<b>Conclusion</b>	<b>67</b>
	<b>Bibliography</b>	<b>70</b>

# List of original articles

Part of the work described in this thesis appears in the following articles:

**P. Galdi**, M. Fratello, F. Trojsi, A. Russo, G. Tedeschi, R. Tagliaferri, & F. Esposito. (2017). Consensus-based feature extraction in rs-fMRI data analysis. *Soft Computing*, 1-11.

**P. Galdi**, M. Fratello, F. Trojsi, A. Russo, G. Tedeschi, R. Tagliaferri, & F. Esposito. (2017). Stochastic rank aggregation for the identification of functional neuromarkers. *Manuscript submitted*.

J. Dubois, **P. Galdi**, Y. Han, L. K. Paul, R. Adolphs. (2017) Predicting personality traits from resting-state fMRI. *Manuscript submitted*.

# Chapter 1

## Introduction

Functional magnetic resonance imaging (fMRI) is a neuroimaging technique based on magnetic resonance used to describe brain activity over time. It measures neural activity indirectly, exploiting the blood-oxygen-level dependent (BOLD) effect, first discovered by Seiji Ogawa in the early 1990s, that can be summarized as follows: after a neural activation, the oxygen consumption causes an increase of the blood flow in the activation locus that in turn causes an increase of the intensity of the measured signal.

The introduction of fMRI allowed for non-invasive in vivo observation of brain function with an unprecedented spatial resolution, thus dramatically changing the landscape of brain science. One of the main research goal of this field is to analyse the human brain network, often referred to as human connectome, in order to understand its anatomical and functional organisation. Gaining such an understanding is fundamental to support early diagnoses of neurological disorders and to improve the treatments of these pathologies. Furthermore, connectome-based predictive models can be used to study the links between measured brain activity and individual differences in demographics, behaviour and psychological profile.

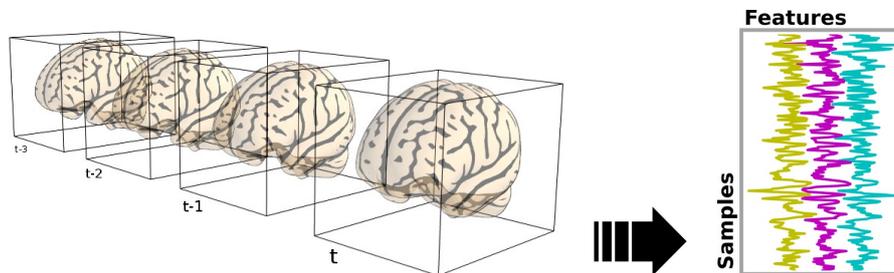


Figure 1.1: Representation of fMRI data in 2D form.

---

fMRI data consist of 3D volumetric images of the brain acquired over time, that can be represented in matrix form, with a time series associated to each voxel (volumetric pixel), as depicted in figure 1.1. Traditionally, fMRI has been applied to observe brain activity during the performance of a task or in conjunction with the administration of a stimulus (e.g., visual or auditory stimuli). However, already in 1995, in the seminal work by Biswal and colleagues, spatially correlated fluctuations of the BOLD signal were observed in the brain at rest, i.e. in the absence of a task. The common variance of signals from different regions of the brain has been interpreted as an indicator of synchronous neural activity, demonstrating that these regions are functionally connected. Since then, a large body of literature has been devoted to the study of resting-state fMRI (rfMRI), to investigate the organisation of the brain as a functional network.

The central theme of this thesis is the design of methods for building functional neuromarkers from rfMRI data. Identifying brain correlates of health-related outcomes, such as diagnostic categories or cognitive performance, could bring further insight in neuropathology and in the assessment of healthy brain function. One of the challenges of defining brain markers, as opposed to other biomarkers based on relatively stable trait indicators, is to cope with the considerable variability in brain structure, brain function and cognition during the lifetime of the same individual. Ideally, a neuromarker should be stable across age, not significantly affected by brain states, sensitive when compared with normative controls and specific across disorders (Gordon, 2007). Moreover, features that can be easily understood and interpreted by humans tend to be more useful, since it is easier to assess their neuroscientific plausibility. Early neuroimaging studies were aimed at understanding which functions were encoded in isolated regions of interest, but it is now commonly believed that features of brain disorders are more likely to be encoded in neural systems encompassing networks of many regions (Woo et al., 2017), therefore models that integrate the contributions of multiple regions, as those based on brain functional networks, have a better chance of obtaining more accurate neuromarkers. Similarly, studies investigating the neural mechanisms underlying higher cognitive abilities in healthy subjects discovered that human intelligence, rather than arising from specialized regions, is linked with global properties of brain networks (Barbey, 2017).

In this work, we consider two of the main approaches used to model functional brain networks. The first is based on independent component analysis (ICA), a blind source separation method that assumes that the observed signal  $X$  is the result of the contributions of different independent sources, the components:  $X = MC$ , where each row of  $C$  is a component and  $M$  is a mixing matrix representing the contribution of each component to the observed sig-

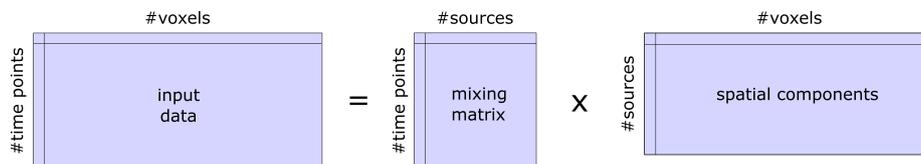


Figure 1.2: ICA on fMRI data: the input matrix consists of time series associated to brain voxels; the mixing matrix contains, for each time point, the relative contribution of each independent component to the global signal; the component matrix indicates the contribution of the single voxels to the components.

nal for each time point. The goal of ICA is to determine an unmixing matrix  $W$  to separate signal components:  $C = WX$ . With fMRI data, every component identified by ICA consists of a set of voxel values with an associated time course (figure 1.2). These values are a measure of the voxel contribution to the specific component. We can distinguish between spatial and temporal ICA depending on whether we require the spatial pattern or the time courses to be independent, but the most commonly used is the spatial ICA, and this is also due to the fact that there are many more voxels than time points. When applied to rfMRI data, ICA is used to identify resting state networks (RSNs), i.e. functionally connected networks of regions that are spatially independent (Beckmann et al., 2005; Damoiseaux et al., 2006; Tedeschi and Esposito, 2012). The most commonly reported RSNs are six: the default-mode network (DMN), the visual network, the fronto-parietal network, the sensori-motor network, the auditory network and the self-referential network. The DMN is by far the most studied RSN. It is involved in attention-related processes and it appears to be deactivated when the brain is involved in cognitive tasks while active when the subject is awake and at rest, therefore it is considered a cognitive baseline for a subject (Tedeschi and Esposito, 2012).

A second approach builds functional connectivity networks first defining network nodes as spatial coordinates or as regions of interest derived from a brain parcellation; a time series is then identified for each node, e.g. as the average time series across all voxels in a parcel; finally, the strength of the connections between nodes is estimated from data, commonly as the correlation of each time series with every other time series (figure 1.3).

While in the latter approach the dimensionality of the problem is determined by the number of network nodes, ICA performs a whole brain analysis, and a feature selection step is often necessary to allow for the application of advanced analysis techniques. In chapter 2, a framework is presented for the extraction of features from the DMN maps based on clustering techniques, as opposed to traditional univariate feature selection methods. Clustering is the task of grouping similar objects together. In this context, it is used to identify groups of voxels showing a common behaviour, shifting the perspective from a

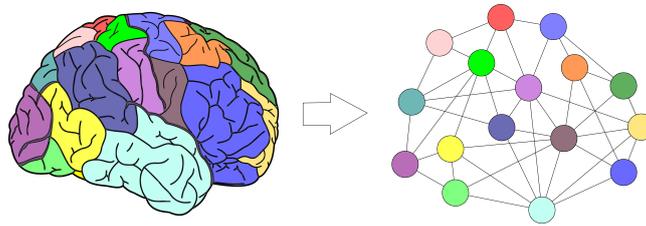


Figure 1.3: Network modelling applied to the investigation of brain connectivity: nodes represent brain regions and an edge between two regions indicates that there is a functional connection between the measured brain activity in the two regions.

mere geometric representation to the level of brain parcels. This has the dual advantage of making the system more robust to noise and to divide the brain into functional modules that are easier to interpret. Consensus clustering, a particular approach to the clustering problem that consists in combining different partitions of the same data set to produce more robust solutions, is introduced as a technique for obtaining stable subsets of features. In chapter 3, a method based on stochastic rank aggregation is proposed for the identification of neuromarkers of neurodegenerative diseases that uses as input clustering derived features. Briefly, this approach consists in using the data of each subject to create a ranking of the most activated brain regions; all the rankings of a given condition are then integrated to derive a subset of informative parcels.

Chapter 4 explores the possibility of training learning models on individual functional connectivity networks to predict cognitive abilities and personality traits in healthy subjects. Since functional connectivity is known to be greatly affected by motion artefacts (Burgess et al., 2016; Siegel et al., 2016), the main denoising strategies adopted in the field are discussed. The experimentation was based on a large cohort of subjects made available by the Human Connectome project, an open data initiative that provides access to a great number of fMRI acquisitions in order to promote research on the human brain connectivity.

# Chapter 2

## Feature extraction from rfMRI

Over the past decades a great effort has been made by researchers in order to shed light on the physiological and pathological mechanisms governing our nervous system. Recent advances in neuroimaging have allowed the collection of a considerable amount of data describing living human brains. However, the high-dimensional nature of fMRI data implies the need of suitable feature selection techniques. Traditional univariate techniques are fast and straightforward to interpret, but are unable to unveil relationships among multiple features. The aim of this chapter is to discuss the applicability of clustering based techniques to the problem of feature extraction in resting state fMRI data analysis. More specifically, a methodology is presented based on consensus clustering, a particular approach to the clustering problem that consists in combining different partitions of the same data set in a final solution. The approach was validated on a real-world data set, deriving from multiple clinical studies on Parkinson's disease and amyotrophic lateral sclerosis. Specifically, it was applied to default-mode network maps extracted from each subject with ICA. Results show that the adoption of consensus-based techniques can indeed lead to an improvement of the results, not only in terms of feature discriminability, but also from the point of view of interpretability.

### 2.1 Background

After preprocessing fMRI data, a feature selection step is often necessary before performing further analyses. The reasons for reducing data dimensionality are several: prohibitive computational requirements; metrics that lose meaning when applied in a high dimensional space; the relatively higher number of samples required to train a learning model.

Traditional univariate techniques, such as t-test or analysis of variance (A-

NOVA), suffer from many drawbacks. First, false positives and the multiple comparisons problem should be taken into account (Eklund et al., 2016), since the number of features in a whole-brain analysis is in the order of several thousands and above. Second, univariate approaches are unable to unveil relationships among features and this is a serious limitation in this domain. In fact, voxels are features deriving from a geometrical representation of the brain that does not reflect the actual organization of neurons, therefore it is most unlikely that a single voxel could alone explain an underlying neurobiological phenomenon.

Since the ultimate goal of fMRI studies is to understand which voxels covary and whether the activity of some voxel influences the activity of another voxel, then one possible solution is to employ multivariate statistical tests (e.g. multivariate ANOVA); but, since each independent variable takes up a degree of freedom, these approaches are problematic when working with a high number of features. Alternative techniques employ anatomical information to build regions of interest, but they imply the need for strong a priori assumptions, or are based on searchlight analysis, a method for combining the information deriving from all voxels lying into a sphere centred on a given voxel (the ‘searchlight’). The latter approach has, however, some limitations that could lead to the wrong interpretation of a cluster as informative, or to the inability to detect truly informative voxels (see Etzel et al. (2013) for a detailed discussion of these aspects).

In exploratory data analysis, clustering is the task of segmenting data in groups in order to highlight underlying relationships among data objects. Typically, a clustering algorithm works by building a partition in which a given measure of intra-group similarity between items is maximized. When applying clustering to fMRI data, the goal is to obtain a compressed representation of the original data set. Voxels are partitioned into groups using correlation as a similarity measure, in order to cluster together voxels that show a similar behaviour; then, a representative feature is selected from each cluster and the extracted features become the input for further analysis. On the one hand this can be considered as a dimensionality reduction step, aimed to decrease the number of variables and thus facilitating the application of more sophisticated statistical models. On the other hand, aggregating the information described by several voxels allows to work at a higher level of abstraction, that of brain regions, and this has multiple advantages: a) it captures the modular organization of the brain; b) it aids the generalization of the results since clusters of voxels are built across subjects; c) features are easier to interpret because they can be reliably mapped to brain regions. Compared to other dimensionality reduction techniques (e.g. PCA), clustering produces features that can be put into bijective correspondence with voxels, retaining information useful for visualization and interpretation. Working with small regions instead of single voxels

guarantees a higher reliability of the results, especially in group studies, where even if every acquired brain is transformed in order to fit in a standard common space, it might be troublesome to assign a single voxel to the same anatomical area across subjects with a high degree of confidence.

Consensus clustering is a methodology used to combine multiple clustering solutions to obtain more accurate and stable results (Kuncheva et al., 2006), since a single clustering run could be affected by noise and prone to overfitting the data. Moreover, many clustering algorithms are sensitive to the choice of initialization parameters. The idea is borrowed from classifier ensembles, where many weak classifiers are combined together to improve the overall accuracy of the model. In this context, consensus clustering was applied not only in the attempt to get better quality partitions of the data, but also to achieve a higher reliability in the results and to devise a robust method for the extraction of stable set of features.

## 2.2 Methods

In this section the basic concepts of consensus clustering are presented together with a methodology for extracting stable subsets of features inspired by this technique. Then, a domain specific evaluation measure for cluster quality is introduced.

### 2.2.1 Consensus clustering

Consensus clustering is a particular approach to the clustering problem that consists in aggregating different partitions of the same data set into a final solution for the purpose of improving the quality of individual data clusterings. Intuitively, results obtained with this methodology are more reliable because they are supported by many base solutions and the final clusters are more stable, since they represent the consensus among multiple different partitions; furthermore, this technique is more robust to noise and overfitting than a simple execution of a clustering algorithm (Monti et al., 2003; Hornik, 2005; Nguyen and Caruana, 2007). A wide literature exists on this subject (Strehl and Ghosh, 2003; Topchy et al., 2005; Kuncheva et al., 2006; Nguyen and Caruana, 2007; Ghaemi et al., 2009), but the algorithms differ mainly in two aspects: the generation of multiple clustering solutions and the consensus function used to quantify the concordance among partitions. Different base clusterings can be generated in several ways: varying the number of clusters in the partitions; using different clustering algorithms or, where applicable, using a random initialization of the same algorithm (e.g. K-means); perturbing the original data set by means of subsampling (Monti et al., 2003) or random projections (Fern and Brodley, 2003; Bertoni and Valentini, 2006). The most simple and common way to measure consensus is

to build a pairwise similarity matrix (the consensus or co-association matrix) counting how many times two objects are assigned to the same cluster across partitions. More formally, given a data set of  $N$  points and  $P$  data partitions, let

$$M^{(p)}(i, j) = \begin{cases} 1 & \text{if items } i \text{ and } j \text{ belong to the same cluster} \\ 0 & \text{otherwise.} \end{cases} \quad (2.1)$$

be an  $N \times N$  connectivity matrix corresponding to partition  $p \in \{1, \dots, P\}$  (see figure 2.1 for an illustrative example). We can build a consensus matrix in the following way:

$$\mathcal{M}(i, j) = \frac{\sum_p M^{(p)}(i, j)}{P}. \quad (2.2)$$

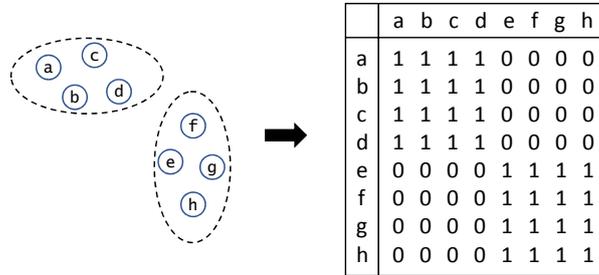


Figure 2.1: An example of connectivity matrix. The dotted lines represent two clusters.

The final consensus clustering can then be obtained using the matrix  $\mathcal{M}$  as input to a hierarchical clustering algorithm.

### 2.2.2 Consensus-based feature filtering

The information contained in a consensus matrix can give an insight into the stability of a feature set. If we consider only data points that are clustered often together, we can obtain a subset of more stable features, since many of the base solutions agree on their collocation in a partition. This idea inspired the approach called consensus-based feature filtering. Considering the consensus matrix as a pair-wise similarity matrix, we choose two thresholds  $\nu$  and  $\sigma$  and select only the items that have a similarity greater than  $\sigma$  with at least  $\nu$  other objects. In more detail:

- $\sigma$  is a real value between 0 and 1, and indicates how often two features have to be clustered together to be considered a *stable pair*.
- $\nu$  is an integer number: a feature is selected if it is included in a number of *stable pairs* greater than  $\nu$ .

The choice for the values of the thresholds is problem-specific. Threshold  $\sigma$  is related to the variability in the base solutions: for instance, if the various partitions are the results of the application of the same clustering algorithm with varying initialization, a stricter value for the threshold would be appropriate since solutions are expected to share a certain amount of information; on the contrary, when combining clusterings resulting from different algorithms, solutions are likely to be more diverse and therefore a more tolerant threshold would be preferable. Threshold  $\nu$  is related to the average size of clusters and so indirectly to the number of clusters in the partition: the bigger the original clusters are, the higher the threshold should be.

### 2.2.3 Anatomical cluster filtering

In correlation based clustering, voxels are grouped together if they behave coherently across subjects, ignoring their spatial position. Nevertheless, in this specific domain, a desirable property for a cluster is to be formed by a sufficient number of contiguous voxels in order to be easily associated with an anatomical area and to be a posteriori validated.

In order to assess the quality of the obtained clusters, an evaluation measure based on the degree of clustering (Formisano et al., 2002) is adopted. For each cluster, this score is defined as follows:

$$DoC = \begin{cases} 0 & \text{if cluster size is smaller than } \theta \\ \frac{N_c}{N_{tot}} & \text{otherwise.} \end{cases} \quad (2.3)$$

where  $\theta$  is a user-defined threshold for the minimum size of a cluster,  $N_c$  is the number of contiguous voxels in a cluster and  $N_{tot}$  is the total number of voxels in a cluster. For instance, if a cluster is too small then its score will be 0; if a sufficiently big cluster is constituted by contiguous voxels its score will be 1.

This measure can be used to filter out clusters that do not meet the desired properties.

## 2.3 Data description

Data come from a cohort of 121 subjects, with age ranging from 38 to 82 years (mean age  $63.87 \pm 8.2$ ). Specifically, they include 41 patients (20 women) from a clinical study on amyotrophic lateral sclerosis (Tedeschi et al. 2012); 37 patients (14 women) from a clinical study on Parkinson’s Disease (Tessitore et al. 2012a; Tessitore et al. 2012b; Esposito et al. 2013; Amboni et al. 2015) and 43 control subjects (23 women) from the same clinical studies. A 3T scanner equipped with an 8-channel parallel head coil (General Electric Healthcare, Milwaukee, Wisconsin) was used for the acquisition of MRI images. A sequence of 240

volume was acquired using gradient-echo T2\*-weighted MR imaging (TR = 1508ms, axial slices = 29, matrix =  $64 \times 64$ , field of view = 256mm, thickness = 4mm, inter-slice gap = 0mm). Subjects were asked to stay awake and motionless and to keep their eyes closed during the scans. To register and normalize fMRI images, high resolution T1-weighted sagittal images were acquired in the same session (GE sequence IR-FSPGR, TR = 6988ms, TI = 1100ms, TE = 3.9ms, flip angle = 10, voxel size =  $1\text{mm} \times 1\text{mm} \times 1.2\text{mm}$ ). Data pre-processing was performed with BrainVoyager QX (Brain Innovation BV, Maastricht, the Netherlands) including slice timing correction, 3D rigid body motion correction and high-pass temporal filtering. For each subject, 40 independent components (ICs) were extracted with the fastICA algorithm (Hyvarinen 1999), accounting for more than 99.9% of the total variance. The number of ICs corresponds to one sixth of the number of time points (following the example of Greicius et al. 2007). Each component identifies a whole brain spatial map, which assigns to each voxel a weight representing the voxel’s contribution to the signal of the component. Among the extracted spatial maps, the one associated with the DMN was selected as the one with the highest goodness of fit (GoF) with a DMN mask from a previous study on the same MRI scanner with the same protocol and pre-processing (Esposito et al. 2010), where the GoF is computed as the mean IC value inside the mask minus the mean IC value outside the mask (Greicius et al. 2004; Greicius et al. 2007). To avoid ICA sign ambiguity, each component sign was adjusted to have all GoF values as positive.

## 2.4 Experimental setup

After the preprocessing of fMRI data (as described in section 2.3), ICA was applied to each subject data to extract the independent component corresponding to the DMN, that was selected as the component with the best fit with a template of the DMN from a previous study of Esposito et al. (2010).

In order to speed up the tests, data were subsampled from 3 mm to 4.5 mm voxels, reducing the number of features from  $\sim 50\,000$  to  $\sim 13\,000$ .

We obtained an  $v \times n$  matrix  $X$ , where  $v$  is the number of voxels and  $n$  is the number of subjects, and entry  $X(i, j)$  indicates the contribution of voxel  $i$  to the DMN component of subject  $j$ .

Two clustering algorithms were chosen to generate the initial partitions of  $X$ :

1. K-means clustering, a standard method in the cluster analysis literature, which tries to find a partition of the data set that minimizes the distortion error, i.e. the distance between cluster points and centroids. The main drawback of this approach is that the output might be affected by the presence of local minima in the error function.

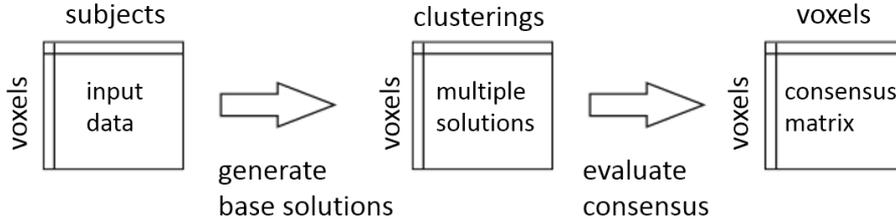


Figure 2.2: The steps for building a consensus matrix.

2. DBSCAN (Ester et al., 1996), a density based clustering algorithm that identifies clusters starting from an estimate of the density distribution of points, and was chosen for its ability to detect noise in the form of outliers that lie in low-density regions.

Figure 2.2 shows the steps performed to obtain a consensus matrix starting from matrix  $X$ . Multiple base clusterings of the voxels were performed by 100 runs of the K-means algorithm with a random initialization of the initial centroids or by 100 runs of DBSCAN algorithm shuffling the input data, since this method is sensitive to the ordering of data points. In both cases, one minus the Pearson correlation coefficient was used as a dissimilarity measure. The number of clusters in the partitions was set to 500 for K-means and the parameters of DBSCAN were tuned in order to obtain clusters of  $\sim 30$  voxels, which is a reasonable size to map a cluster to an anatomical region. A consensus matrix was built as described in section 2.2.1 and a final clustering was obtained by means of hierarchical clustering with average linkage. The resulting clusters were filtered using the degree of clustering (section 2.2.3) to select only those with a score higher than 0.5 and a minimum size of 10 voxels. The consensus based voxel filtering (section 2.2.2) was then applied to select only voxels that are often clustered together, with parameters  $\sigma \in [0.4, 0.7]$  and  $\nu \in [3, 5]$ . These ranges of values were chosen to guarantee a high degree of stability in the final solution while preserving cluster structure. Specifically, the upper bounds were chosen experimentally, since with stricter thresholds too few voxels were selected to constitute a meaningful solution.

Features were extracted from clusters in two ways:

- selecting the medoid of each cluster, i.e. the voxel which has on average the higher correlation with the other voxels forming the cluster;
- computing the geometric mean of the cluster voxels.

To assess the quality of the selected features, they were used as input in a three-class classification problem (i.e. to distinguish healthy controls, subjects affected by amyotrophic lateral sclerosis, and subjects diagnosed with Parkin-

---

<i>all</i>	all the features were used
<i>mean</i>	means of clusters used as features
<i>mean+DoC</i>	means of clusters used as features with DoC filtering
<i>mds</i>	medoids of clusters used as features
<i>mds+DoC</i>	medoids of clusters used as features and DoC filtering
<i>mean+filter</i>	means of clusters used as features with both consensus filtering
<i>mean+filter+DoC</i>	means of clusters used as features with both consensus and DoC filtering
<i>mds+filter</i>	medoids of clusters used as features with both consensus filtering
<i>mds+filter+DoC</i>	medoids of clusters used as features with both consensus and DoC filtering
<i>RF</i>	Random Forest
<i>SVM</i>	Support Vector Machine
$\langle algorithm \rangle_C$	clustering obtained with $\langle algorithm \rangle$ and consensus
$\langle algorithm \rangle_S$	clustering obtained with a simple run of $\langle algorithm \rangle$

---

Table 2.1: Abbreviations used to the identify the different tests.

son’s disease). Random Forest (RF) and Support Vector Machines (SVM) with linear kernel were chosen as classifiers because these models are known to cope well with high dimensional data. In order to evaluate the generalization abilities of the models, data were split into a training and a test set (70-30%). Since the three classes were slightly unbalanced (see section 2.3), proportions were preserved when training and test folds were generated. All the clusterings were performed on the training set. The obtained partitions were then applied on the test set to extract the features. Since the consensus-based voxel filtering depends on two tuning parameters, stratified 5-fold cross validation was performed on the training set to select them in a grid search fashion; namely, different combinations of the two parameters were compared using the cross-validated score (table 2.2 reports the selected parameters for each test). While Random Forest uses a built-in procedure for feature selection, recursive feature elimination (RFE) was applied in combination with SVM to select a subset of the features. In this case, a nested cross-validation was implemented: an outer loop with a stratified 5-fold cross validation for the consensus filter parameters and an inner loop with Leave-one-out cross validation for the parameters of SVM and RFE.

## 2.5 Results and discussion

We compared the obtained classification accuracies using all the features, the medoids and the means of the clusters generated with a simple run of K-means or DBSCAN and with consensus clustering. We also compared the results achieved with or without the application of the DoC and the consensus filters. Please

	K-means <sub>C</sub>		DBSCAN <sub>C</sub>	
	$\sigma$	$\nu$	$\sigma$	$\nu$
<b>Random Forest</b>				
<i>mean+filter</i>	0.5	3	0.7	3
<i>mean+filter+DoC</i>	0.5	3	0.4	3
<i>mds+filter</i>	0.5	3	0.4	4
<i>mds+filter+DoC</i>	0.7	5	0.4	4
<b>SVM</b>				
<i>mean+filter</i>	0.6	4	0.6	3
<i>mean+filter+DoC</i>	0.5	4	0.5	3
<i>mds+filter</i>	0.6	4	0.6	4
<i>mds+filter+DoC</i>	0.4	3	0.6	5

Table 2.2: Parameters of consensus filter, selected with cross validation.

#clusters	K-means <sub>C</sub>	K-means <sub>S</sub>	DBSCAN <sub>C</sub>	DBSCAN <sub>S</sub>
<i>mean/mds</i>	500	500	500	622
<i>DoC</i>	434	486	275	305
<b>Random Forest</b>				
<i>mean+filter</i>	453	-	61	-
<i>mean+filter+DoC</i>	405	-	10	-
<i>mds+filter</i>	453	-	304	-
<i>mds+filter+DoC</i>	187	-	154	-
<b>SVM</b>				
<i>mean+filter</i>	445	-	140	-
<i>mean+filter+DoC</i>	389	-	101	-
<i>mds+filter</i>	445	-	105	-
<i>mds+filter+DoC</i>	422	-	21	-

Table 2.3: Number of clusters in each test. The DoC and the consensus filtering preserve much more clusters when the base clusterings were generated with K-means rather than with DBSCAN, suggesting that the former algorithm is able to detect more compact clusters.

<b>RF accuracies</b>	K-means <sub>C</sub>	K-means <sub>S</sub>	DBSCAN <sub>C</sub>	DBSCAN <sub>S</sub>
<i>all</i>	74%	74%	74%	74%
<i>mean</i>	76%	76%	74%	71%
<i>mean+DoC</i>	79%	<b>79%</b>	<b>85%</b>	68%
<i>mds</i>	71%	65%	76%	<b>82%</b>
<i>mds+DoC</i>	76%	65%	74%	74%
<i>mean+filter</i>	<b>82%</b>	-	76%	-
<i>mean+filter+DoC</i>	<b>82%</b>	-	76%	-
<i>mds+filter</i>	76%	-	71%	-
<i>mds+filter+DoC</i>	76%	-	65%	-

Table 2.4: Classification accuracies obtained with Random Forest classifier (best results in bold).

<b>SVM accuracies</b>	K-means <sub>C</sub>	K-means <sub>S</sub>	DBSCAN <sub>C</sub>	DBSCAN <sub>S</sub>
<i>all</i>	<b>65%</b>	<b>65%</b>	65%	65%
<i>mean</i>	62%	<b>65%</b>	<b>68%</b>	47%
<i>mean+DoC</i>	56%	59%	56%	<b>68%</b>
<i>mds</i>	53%	53%	56%	44%
<i>mds+DoC</i>	41%	59%	53%	53%
<i>mean+filter</i>	56%	-	44%	-
<i>mean+filter+DoC</i>	56%	-	47%	-
<i>mds+filter</i>	50%	-	53%	-
<i>mds+filter+DoC</i>	59%	-	50%	-

Table 2.5: Classification accuracies obtained with SVM classifier in the various tests (best results in bold).

refer to table 2.1 for an explanation of the abbreviations used to identify the various tests. Table 2.3 reports the number of clusters in each test. Tables 2.4 and 2.5 show the classification accuracies obtained with Random Forest and SVM respectively. Figures 2.3 and 2.4 show bar plots of the results grouped by clustering method and by classifier. For better readability, results relative to tests based on the medoids of clusters were omitted, since in all but one case they were outperformed by the ones based on the means of clusters. This might suggest that the geometric mean is a more effective method for extracting features compared to the selection of the cluster medoids. In the following paragraphs, a brief comment is provided for each combination of clustering algorithm and classifier.

**K-means Consensus and RF** In this test, the best result has been achieved with the means of the clusters used as features in combination with the consensus based filtering. Even applying the DoC filter, the result is unchanged, suggesting that the clusters that form the solution are made of a sufficient amount of

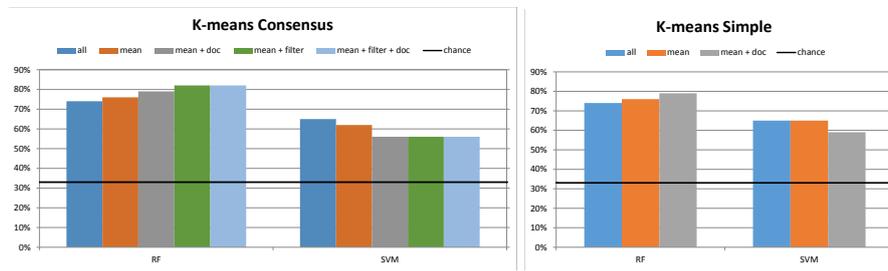


Figure 2.3: Classification accuracies obtained with Random Forest and SVM with features extracted with K-means Consensus (left) and K-means Simple (right). The black line indicates chance level.

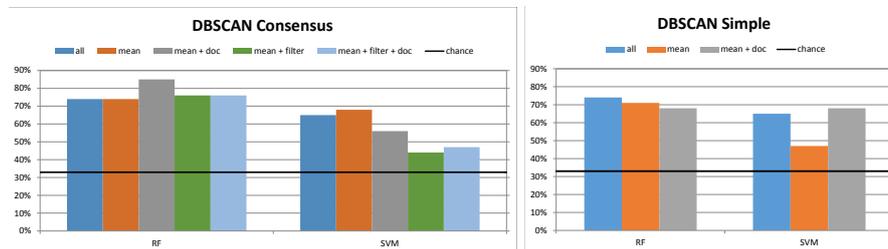


Figure 2.4: Classification accuracies obtained with Random Forest and SVM with features extracted with DBSCAN Consensus (left) and DBSCAN Simple (right). The black line indicates chance level.

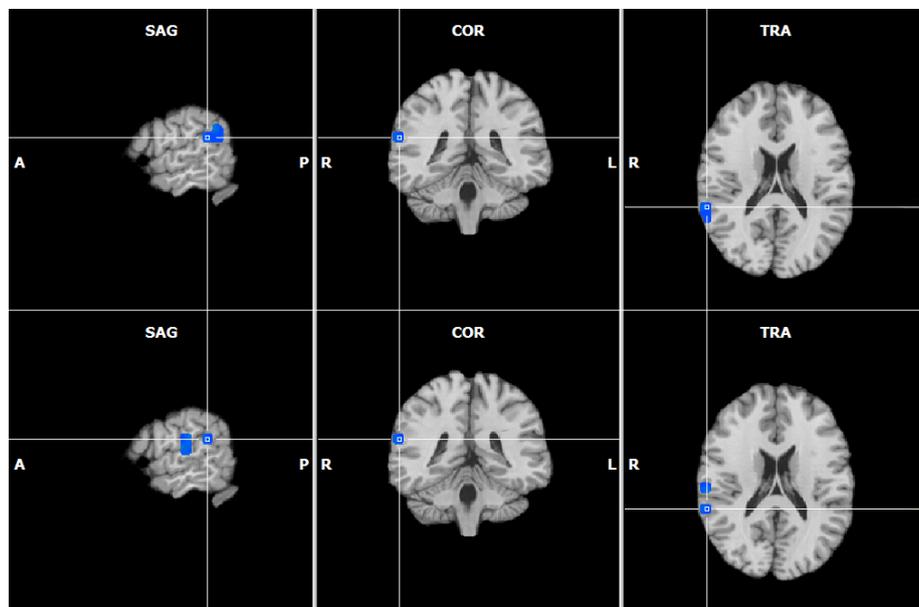


Figure 2.5: Comparison between the two most similar clusters generated with K-means Consensus and both DoC and consensus filtering (top row) and simple K-means with DoC filtering (bottom row).

contiguous voxels.

**K-means Consensus and SVM** The best accuracy is obtained when using all the features, but this result is not helpful from the point of view of the interpretability. No one of the techniques improves the standard result, however the best score obtained with SVM is lower than the average score obtained with RF.

**K-means Simple and RF** The best result is attained with the means of the clusters used as features in combinations with the DoC filtering, but the score is lower than the best score obtained with K-means and consensus-based filtering.

**K-means Simple and SVM** The results of this test show a similar trend to the one with K-means Consensus, except for an improvement in the run with medoids and DoC filters.

**DBSCAN Consensus and RF** Here the best result is achieved with the means of the clusters as features and the DoC filtering, followed closely by the run with medoids and means with consensus filtering (both with and without DoC filtering).

**DBSCAN Consensus and SVM** The best result is the one with the means of clusters as features, but also in this case the score is lower than the average score achieved with RF.

**DBSCAN Simple and RF** The best result is obtained in the run with the medoids of clusters as features, but we can observe how after the application of the DoC filter there is a significant decrease in accuracy, indicating that the selected clusters do not satisfy the minimum quality requirements.

**DBSCAN Simple and SVM** Similarly to the other tests with SVM, the best score (achieved in the run with cluster means as features and DoC filtering), is lower than the scores obtained with RF.

In reference to table 2.3, where the number of clusters of each test is reported, it is important to note that, when using the consensus-based voxel filtering, the subsequent application of the DoC filtering plays a relevant role. In fact, the filtering could break cluster structure, leading to sparse or too small groups of voxels.

We can observe that in general the DoC and the consensus filtering preserve much more clusters when the base clusterings were generated with K-means

rather than with DBSCAN, suggesting that the former algorithm is able to detect more compact clusters and indeed, while the clusters built by K-means are homogeneous in size, DBSCAN tends to find few big clusters and many small sized groups of voxels that are then filtered out.

For what concerns the adopted classifiers, Random Forest outperformed SVM in every test. This is probably due to the ability of Random Forest of handling non-linearity, but the choice of a linear kernel for SVM was obliged considering the high dimensionality of data compared to the number of training samples available. Considering then the test based on K-means and Random Forest, we can see an improvement in the classification accuracy both in combination with consensus clustering and with the consensus-based feature filtering. The application of the DoC filtering removes a little number of clusters without affecting the accuracy, indicating that the resulting clusters were already compact and sufficiently big.

To give an idea of the qualitative difference between clusters built with our consensus approach and with a simple clustering algorithm, in figure 2.5 we compare two clusters taken from the best solutions yielded by K-means Consensus and K-means Simple respectively. The clusters were chosen as the two most similar with respect to Jaccard index, defined as

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}. \quad (2.4)$$

As we can see, the cluster generated with consensus is more compact, while the other has a more sparse structure, therefore in the former case it is easier to map the cluster to an anatomical region (in this case Brodmann area 40) and this helps to increase the interpretability of the results from a neurobiological perspective. Indeed, in Mohammadi et al. (2009) the DMN showed less activation in patients affected by ALS in left and right inferior parietal cortex, and in Tessitore et al. (2012b) patients with Parkinson's disease showed decreased functional connectivity of the bilateral inferior parietal cortex. In order to observe the effect of the consensus-based voxel filtering, figure 2.6 shows some of the clusters from the best solution before and after the application of the filter. The voxels that are removed by the filter (in white in the figure) lie in border regions of the clusters or are isolated voxels, thus the remaining voxels constitute core regions of the initial clusters, indicating that the use of the proposed method has the desired effect of selecting stable subsets of the original features. Since one of the main goal of fMRI data analysis is that of detecting relevant regions of the brain related to physiological and pathological mechanisms, we selected the top 10% most discriminative clusters according to Random Forest's feature importance score (see figure 2.7) and made a comparison with literature. We found out that more than a half of the top clusters fall in regions that have

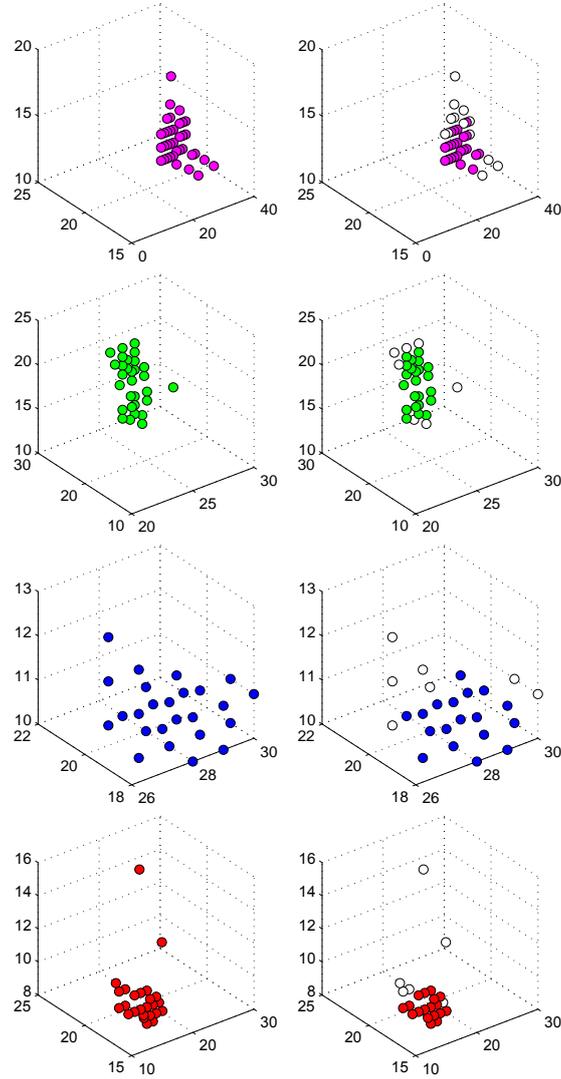


Figure 2.6: The effect of the consensus-based voxel filtering on clusters. On the left column some of the clusters generated with consensus. On the right, the same clusters after the consensus filtering (white points are filtered out).

some relevance with respect to the diseases object of the study. Specifically, these regions were associated with patterns of enhanced activation or deactivation in patients when compared to controls, and this would explain why the classifier model selected the features associated with this set of regions. In more detail, we found clusters in the following areas:

- the *prefrontal cortex*, that was found to be deactivated in Parkinson's (Am-boni et al., 2015; van Eimeren et al., 2009; Gorges et al., 2013; Tessitore et al., 2012a) and ALS patients (Trojsi et al., 2015; Agosta et al., 2013);

- the *angular gyrus*, that was found to be deactivated in Parkinson’s patients with freezing of gait (Tessitore et al., 2012a);
- the *inferior parietal cortex*, which was associated with decreased functional connectivity in Parkinson’s patients (Tessitore et al., 2012b; Amboni et al., 2015) and enhanced connectivity in ALS patients. (Agosta et al., 2013);
- the *precuneus*, that showed enhanced connectivity both in Parkinson’s (van Eimeren et al., 2009) and in ALS patients (Agosta et al., 2013).

## 2.6 Conclusion

In this chapter, a framework based on consensus clustering for feature extraction from rfMRI data was presented. The proposed approach was tested on a real world data set in a three-class classification task. Firstly, results show that clustering based techniques constitute a promising alternative to univariate feature selection methods. In fact, not only they led to higher accuracy scores, but also allowed to reliably map features to anatomical regions for a posteriori validation. Secondly, it was demonstrated how further improvements can be achieved with a consensus approach and how the information contained in a consensus matrix can be exploited to extract a stable subset of features.

As one would expect, the final consensus solution is affected by the quality of the base clusterings. Indeed, the application of a domain-specific score (DoC) to filter out clusters that did not meet the given quality requirements showed that in general solutions composed by more compact clusters (in this case the ones based on K-means) attained the best results also in term of accuracy. Additionally, comparing the solutions obtained with K-means with the ones generated with the proposed approach, we observed how consensus clusters exhibit a tight structure of contiguous voxels, a desirable property in this application. It should be noted that the advantages of the adoption of consensus clustering in this domain rely not only in quantifiable enhancements of the results, but also in qualitative aspects such as a higher reliability and interpretability of the results and a higher robustness of the method to noise and overfitting.

The framework presented here is meant to be used as an exploratory tool to derive meaningful features without relying on prior assumptions on what the regions of interest should be. Since the general approach is data-driven, the model should be tuned to adapt to the data set under study, but once a consensus partition is computed and the parameters of the consensus filter are set, it is trivial to extract features from previously unseen data samples. The extracted features can then become the input of further analyses. In the next chapter we will discuss how these consensus-based features can be used in

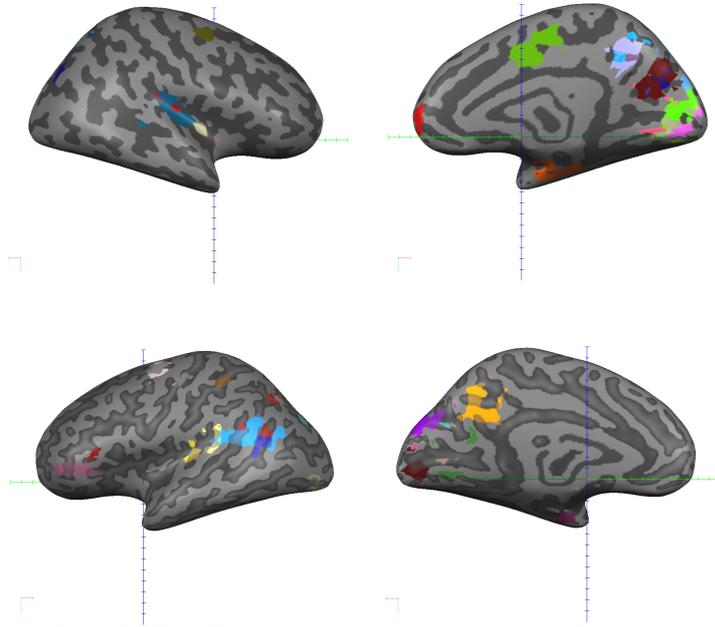


Figure 2.7: The highlighted areas correspond to the top 10% more discriminative clusters according to Random Forest on an inflated representation of the cortex (right hemisphere on top row and left hemisphere on bottom row).

combination with a stochastic rank aggregation method to automatically select a subset of relevant features.

# Chapter 3

## Building neuromarkers from rfMRI

The main challenge in analysing fMRI data from extended samples of subjects ( $N > 100$ ) is to extract as much relevant information as possible from big amounts of noisy data. When studying neurodegenerative diseases with resting-state fMRI, one of the objectives is to determine regions with abnormal background activity with respect to a healthy brain and this is often attained with comparative statistical models applied to single voxels or brain parcels within one or several functional networks. This chapter presents a novel approach based on clustering and stochastic rank aggregation methods to identify functional neuromarkers starting from single-subject activation maps. This framework was tested on a real-world data set consisting of individual ICA-derived default-mode network (DMN) maps from resting-state fMRI scans of subjects affected by neurodegenerative diseases (amyotrophic lateral sclerosis and Parkinson's disease) and of healthy controls.

Clustering has been previously applied in fMRI data analysis to extract patterns from raw time series (Goutte, 1999) or from second level features extracted from data (Goutte et al., 2001), and in group level analyses (Thirion et al., 2006; van den Heuvel et al., 2008). In combination with single-subject independent component analysis (ICA) (Hyvärinen and Oja, 2000), clustering has been also used to identify the most similar ICA components within a single group of subjects (Esposito et al., 2005), but the novelty of the proposed methodology is that it uses clustering in combination with rank aggregation to identify parcels that exhibit a coherent behaviour in groups of subjects affected by the same disorder. This approach being fully data-driven, there is no need for strong assumptions about data distribution as in parametric models, and it can therefore be equally applied to, e. g., ICA maps derived from resting-state fMRI data

or conventional activation maps derived from a general linear model analysis of activation time-courses. Nor it is necessary to specify in advance regions of interest, since a small subset of informative regions automatically emerges from the analysis.

## 3.1 Background

The recent advances in fMRI technology have made available high quality data characterized by ever higher resolution images and shorter repetition times. This translated into an explosion of data dimensionality, thus generating the need of analysis techniques able to cope with the increased complexity of the problem. To investigate how information is represented in the brain of healthy subjects, and how neurodegenerative diseases affect the physiological mechanisms underlying such a representation, a wide variety of statistical and computational methods have been applied to extract meaningful patterns of neural activity from fMRI data.

In population-level analysis, brain voxels are usually analysed in isolation with traditional univariate techniques such as t-test or ANOVA, and we discussed some drawbacks of this approach in the chapter 2. While this might not constitute an issue in single-subject study, in a cohort study inter-subject variability hinders the generalizability of the results: in fact, a voxel found to be significant on a given subject may not be significant on a different subject, or even fall in a different brain region. When conducting population studies, the goal is to analyse group-specific behaviour starting from the product of a first-level analysis consisting in single-subject activation maps. Due to the low signal-to-noise ratio of single-subject images, and to within and between subject variability, the measurable effects might be small or masked by noise. For this reason, the proposed approach is based on similarity within groups, as opposed to traditional comparative approaches that focus on searching discriminative patterns.

## 3.2 Data description

Data presented in this chapter come from multiple clinical studies on amyotrophic lateral sclerosis (Tedeschi et al., 2012; Trojsi et al., 2015) and Parkinson's disease (Amboni et al., 2015; Esposito et al., 2013; Tessitore et al., 2012a,b) and consist of 115 rs-fMRI acquisitions from 3 classes of subjects: 37 controls, 41 ALS and 37 PD. MRI images were acquired on a 3T scanner equipped with an 8-channel parallel head coil (General Electric Healthcare, Milwaukee, Wisconsin). Subjects were asked to rest motionless and with eyes closed during the acquisition. In the same session, high-resolution structural images were acquired

to provide spatial reference for registration and normalization of the functional data.

Data preprocessing was performed with BrainVoyager QX software (Brain Innovation BV, Maastricht, the Netherlands) and consisted of slice timing correction, 3D rigid body motion correction and temporal high-pass filtering. Finally, functional data were coregistered to structural data and then transformed to the Talairach standard space through a 12-parameter affine transformation. After registration to structural images, functional images were normalized to fit the Talairach standard space using a 12-parameter affine transformation and resampled to an isometric 3mm grid covering the entire Talairach box. Finally, all volumes were visually inspected to assess the impact of geometric distortion on the final images, which was judged to be negligible for a whole-brain analysis.

### 3.3 Overview of the methodology

Following the pre-processing and the extraction of the DMN from each subject's data (as described in section 3.2), the input data consist of a matrix  $N \times V$  where  $N$  is the number of subject and  $V$  is the number of voxels, and each entry  $(i, j)$  represents the contribution of voxel  $j$  to the DMN map of subject  $i$ . First, voxels are partitioned into parcels using one of the two methodologies detailed in the following section, and a representative feature (the median) is selected for each group. A ranking is then computed for each subject by sorting in descending order the extracted features. Finally, rankings are aggregated by class of subjects through stochastic rank aggregation; the goal is obtaining a subset of brain regions that share a common behaviour throughout classes.

#### 3.3.1 Brain parcellation

As discussed previously, working on brain parcels instead of single voxels is convenient for many reasons. First of all, brain activity is likely to span over multiple voxels. Therefore, the aggregation of several voxels in a single agglomerated feature may reduce redundancy and improve signal-to-noise ratio, and this could in turn increase the prediction accuracies of learning models. To validate our method for the selection of relevant regions of the brain based on stochastic rank aggregation, we adopted two different approaches to brain parcellation: one based on anatomical information and the other based on the data-driven clustering technique presented in chapter 2.

Anatomical parcellations are derived from an atlas that defines brain regions on a template image. The one used in this work is the Automated Anatomical Labelling (AAL) atlas (Tzourio-Mazoyer et al., 2002) that consists of 90 parcels. Since this approach is data-independent, it allows for an objective comparison of the results of different models.

For the clustering based parcellation, we adopted k-means clustering and Pearson’s correlation coefficient as measure of similarity between voxels across subjects. Moreover, to obtain more stable and reliable sets of features, clustering solutions were enhanced through consensus techniques (as in chapter 2). The final solution consists of 405 brain parcels. Once voxels are segmented in parcels, a representative feature can be extracted from each group to obtain a compressed representation of the input data.

The main drawback of atlas-based approaches is that they are prone to errors in the segmentation of functional regions and this might translate into a decreased sensitivity of the models. Data-driven approaches are, by contrast, more sensitive to noise and since they are not based on a priori defined anatomical regions they require a further step to map features onto an atlas to allow a biological interpretation.

### 3.3.2 Stochastic rank aggregation

*TopKLists* (Schimek et al., 2012, 2015) is a stochastic rank aggregation method that, starting from an ensemble of rankings of a set of items, outputs a new ranking of a subset of the same objects. It works by estimating the rank position  $k$  beyond which the concordance among the input rankings degenerates into noise. Once  $k$  has been computed, all rankings are cut at position  $k$ , thus selecting the top  $k$  elements of each ordered list (hence the name). Finally, all the sublists of length  $k$  are aggregated through a cross-entropy Monte Carlo method (Lin and Ding, 2009).

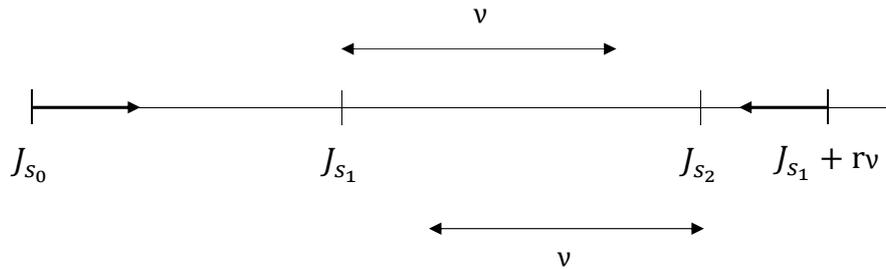


Figure 3.1: The first iteration starts at position  $J_{s_0}$  (the first entry of the indicator vector) and stops in position  $J_{s_1}$ , that is the first position where  $\hat{p}_i^+ - \frac{1}{2} \leq z_\nu$ . The second iteration starts at position  $J_{s_1} + r\nu$  and ends in position  $J_{s_2}$ , the first position where  $\hat{p}_i^- - \frac{1}{2} > z_\nu$ .

The estimation algorithm can be summarized as follows. For each pair of ordered lists an indicator vector  $I$  is built, where  $I_i = 1$  if the item ranked  $i$  by the first assessor is ranked no more than  $\delta$  positions away from rank  $i$  by the

second assessor, and  $I_i = 0$  otherwise. The underlying assumption is that the variables  $I_1, \dots, I_N$  are independent Bernoulli random variables, with  $p_i \geq \frac{1}{2}$  for  $i < i_0$  and  $p_i = \frac{1}{2}$  for  $i \geq i_0$ , where  $p_i$  is the probability that  $I_i = 1$  and  $i_0$  is the rank position where the consensus of the two lists breaks down and noise takes over. For each  $I_i$  an estimate of  $i_0$  is computed: starting from the first position of  $I_i$ , the algorithm updates the current estimate of the position  $i_0$  along the vector by alternating two steps. Even steps start  $r\nu$  positions to the right of the point where the previous odd step ended, while odd steps start  $r\nu$  positions to the left of where the previous even step ended. At step  $s_j$ , a sample of size  $\nu$  is extracted consisting of elements  $I_i$  with  $i$  comprised among the first  $\nu$  indices to the right of  $J_{s_{j-1}} - r\nu$ , if  $j$  is odd, or to the left of  $J_{s_{j-1}} + r\nu$ , if  $j$  is even (where  $J_{s_{j-1}}$  is the point where the previous step ended). See figure 3.1 for a schematisation of these steps. For this sample, an estimate of consensus probability  $p_i$  is computed as the sample mean, i.e. for even steps  $p_i^- = \frac{1}{\nu} \sum_{l=j}^{j+\nu-1} I_l$  and for odd steps  $p_i^+ = \frac{1}{\nu} \sum_{l=j}^{j+\nu-1} I_l$ . In even steps, we move to the left by unitary steps until we reach the first point where  $p_i^- - \frac{1}{2} > z_\nu$ . In odd steps, we move to the right as long as the inequality  $p_i^+ - \frac{1}{2} > z_\nu$  holds. The threshold  $z_\nu$  is defined as

$$z_\nu = (C\nu^{-1} \log \nu)^{1/2}$$

with  $C > \frac{1}{4}$ , to control for moderate deviations (Hall and Schimek, 2012). The algorithm terminates when one of the following stop conditions is met: a) the algorithm enters a loop between two adjacent stages; b) for some  $j$ ,  $J_{s_{2j+1}} \leq J_{s_{2j-1}}$ ; c)  $J_{s_{2j}} - r\nu \leq 1$ .

In this work, *TopKLists* has been applied to combine the rankings of brain parcels expressed by each subject of a class. Specifically, for each subject the medians of the brain regions were computed and sorted in descending order, then all rankings of a class were aggregated in a single list, containing a subset of regions that were ranked similarly across subjects of the same class.

### 3.4 Results

Tables 3.1 and 3.2 report the regions selected by TopKLists for each class with the relative rankings for anatomical parcels and clusters, respectively (the number of voxels and the Talairach coordinates of the geometric mean is reported for each region). The number of anatomical areas selected per class (30 for controls, 27 for ALS, 26 for PD) was higher than the number of selected clusters (8 for controls, 9 for ALS, 8 for PD) albeit the size of functional clusters was smaller than the size of anatomical areas.

Figure 3.2 represents with Venn's diagrams the overlap of regions and clusters shared among classes. There are more anatomical areas shared among all

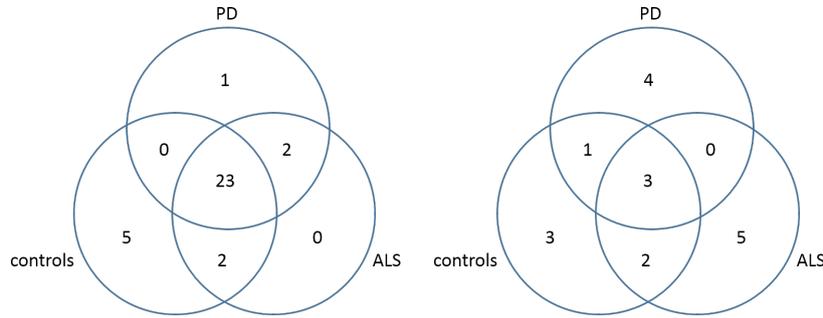


Figure 3.2: Venn diagram of the selected anatomical areas (left) and of selected clusters (right). Most of the anatomical areas are shared among the three classes, while there are more class-specific regions when using clusters. For the ALS group, there are 5 class-specific parcels in the clustering-based ranking while there are none in the anatomical based one.

three classes than class-specific anatomical areas; conversely, there are more class-specific than shared clusters. For the ALS group, there are 5 class-specific parcels in the clustering-based ranking while there are none in the anatomical-based one.

Figures 3.3 and 3.4 show a comparison between the anatomical and the clustering based solutions in corresponding regions, on an MRI image and on an inflated representation of the cortex. Cluster 265 and AAL area 25 (corresponding to the left medial orbitofrontal cortex) are both in the first or the second rank in all three classes (figure 3.3). Both parcels identify approximately the same region, but the functional cluster is smaller and has a substantial overlap with Brodmann area 10. Based on their respective rankings, AAL region 86 (right middle temporal gyrus) can be compared with functional clusters 380, 382 and 383. In this case, the AAL parcel is present in the top rankings of all classes, while clusters 380 and 382 are specific for the control class, and cluster 382 is specific to the ALS class.

Figures 3.5 to 3.8 represent with a box-plot the distribution of the median of each of the selected regions across subjects of the same class (for anatomical regions in figures 3.5 to 3.7 and for clusters in figure 3.8), with the parcels ordered according to the rankings. In most cases, class specific regions (black box-plots) exhibit less variability than others. In the anatomical-based solution for the control class (figure 3.5), the five class-specific regions occupy the highest ranks and exhibit narrow distributions; all of the remaining regions are shared among all classes except for two that are shared only with the ALS class and are present in the first half of the ranking. The only class-specific region for PD occupies the first rank, followed by the two regions shared with the ALS class (figure 3.7). The remaining regions are shared with all classes and show a high variability across subjects (figure 3.5). For what concerns the anatomical

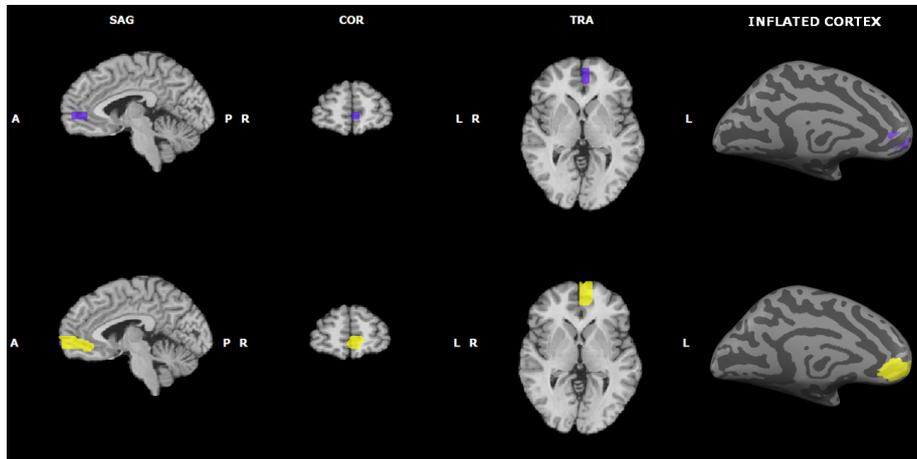


Figure 3.3: Comparison between anatomical areas and clusters in corresponding regions superimposed on a standard MRI image and on an inflated cortex. The top row shows cluster 265 while the bottom row shows AAL ROI 25.

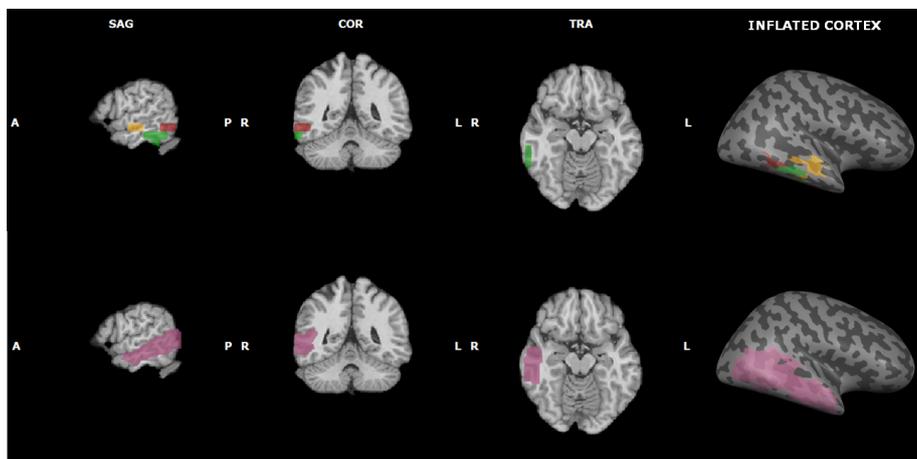


Figure 3.4: Comparison between anatomical areas and clusters in corresponding regions superimposed on a standard MRI image and on an inflated cortex. The top row shows clusters 380 (in orange), 382 (in green) and 383 (in red). The bottom row shows AAL ROI 86.

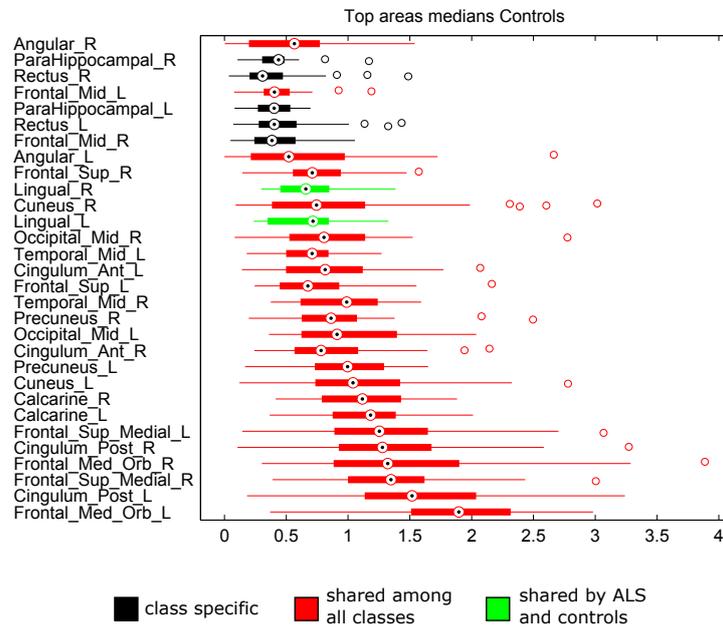


Figure 3.5: Box plots of the distribution of the medians of each of the selected anatomical areas across subjects of the control class. The five class-specific regions (in black) occupy the highest ranks and exhibit narrow distributions.

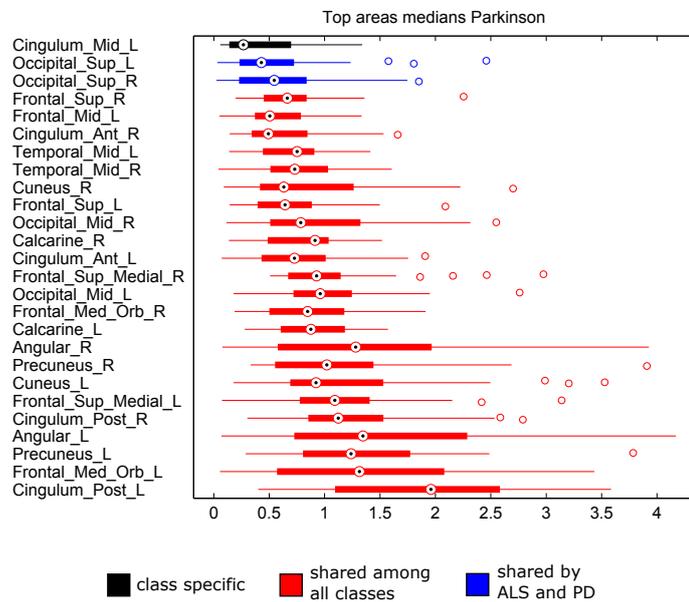


Figure 3.6: Box plots of the distribution of the medians of each of the selected anatomical areas across subjects of the PD class. The only class-specific region (in black) occupies the first rank, followed by the two regions shared with the ALS class (in blue).

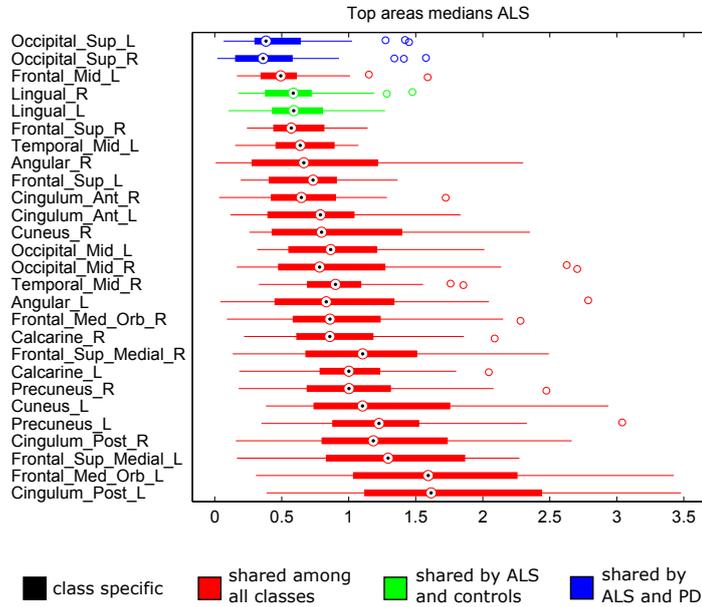


Figure 3.7: Box plots of the distribution of the medians of each of the selected anatomical areas across subjects of the ALS class. There are no class-specific regions, but the ones shared with PD (in blue) and controls (in green) are listed in the top ranks with narrower distributions.

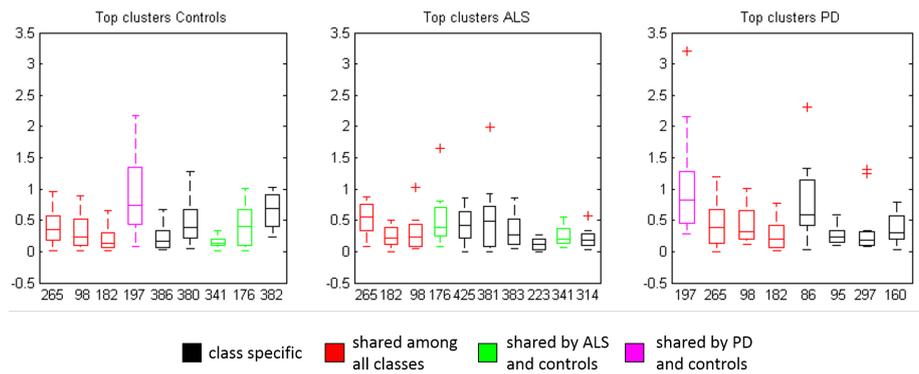


Figure 3.8: Box plots of the distribution of the medians of each of the selected clusters across subjects of the same class. In general the parcels show a reduced variability compared to the anatomical regions, but this is not unexpected since the clusters are smaller and the selected parcels are fewer.

areas relative to the ALS class (figure 3.7), there are no class-specific regions, but the ones shared with PD and controls, respectively, are listed in the top ranks with narrow distributions. Also in this case the regions shared among all classes exhibit a high variability. Considering now the clustering-based solutions (figure 3.8), we can see that in general the parcels show a reduced variability compared to the anatomical regions, but this is not unexpected since the clusters are smaller and the selected parcels are fewer. Another difference with the anatomical approach is that class-specific parcels occupy the second half of the ranking while the clusters shared by all classes (265, 98 and 182) are in the top ranks.

AAL ROI	AAL #	CTRL(30)	ALS(27)	PD(26)	voxel #
Frontal_Sup_L	3	15	19	17	1402
Frontal_Sup_R	4	22	22	23	1513
Frontal_Mid_L	7	27	25	22	1765
Frontal_Mid_R	8	24	-	-	1866
Frontal_Sup_Medial_L	23	6	3	6	1099
Frontal_Sup_Medial_R	24	3	9	13	853
Frontal_Med_Orb_L	25	1	2	2	300
Frontal_Med_Orb_R	26	4	11	11	367
Rectus_L	27	25	-	-	374
Rectus_R	28	28	-	-	352
Cingulum_Ant_L	31	16	17	14	590
Cingulum_Ant_R	32	11	18	21	537
Cingulum_Mid_L	33	-	-	26	674
Cingulum_Post_L	35	2	1	1	205
Cingulum_Post_R	36	5	4	5	152
ParaHippocampal_L	39	26	-	-	410
ParaHippocampal_R	40	29	-	-	471
Calcarine_L	43	7	8	10	866
Calcarine_R	44	8	10	15	743
Cuneus_L	45	9	6	7	579
Cuneus_R	46	20	16	18	577
Lingual_L	47	19	23	-	868
Lingual_R	48	21	24	-	940
Occipital_Sup_L	49	-	27	25	554
Occipital_Sup_R	50	-	26	24	573
Occipital_Mid_L	51	12	15	12	1222
Occipital_Mid_R	52	18	14	16	789
Angular_L	65	23	12	4	461
Angular_R	66	30	20	9	679
Precuneus_L	67	10	5	3	1308
Precuneus_R	68	13	7	8	1257
Temporal_Mid_L	85	17	21	20	1604
Temporal_Mid_R	86	14	13	19	1542

Table 3.1: **Ranking per class (AAL)** The first two columns report the name and the number of the AAL ROI, for reference. Columns CTRL(30), ALS(27) and PD(26) report the ranking for each parcel for controls, ALS and PD patients, respectively. The number between brackets indicates the number of top ranking regions selected per each class. The last column reports the number of voxels for each parcel.

Cluster	CTRL(8)	ALS(9)	PD(8)	BA	AAL Area	# voxels
86	-	-	5	40, 41	-	27
95	-	-	6	40	-	51
98	2	3	3	10	25, 27	81
160	-	-	8	40	-	71
182	3	2	4	WM	-	27
197	4	-	1	WM	-	61
223	-	7	-	23, 24	31	81
265	1	1	2	10, 32	25	54
297	-	-	7	39	65, 85	54
314	-	9	-	7, 18	67,45	54
341	7	8	-	WM	-	27
380	6	-	-	21	86	101
381	-	5	-	22	-	17
382	8	-	-	20, 37	86	98
383	-	6	-	37	86	72
386	5	-	-	22, 47	-	78
425	-	4	-	8	3, 7	26

Table 3.2: **Ranking per class (clusters)** The first column reports numeric identifiers of clusters, for reference. Columns CTRL(8), ALS(9) and PD(8) report the ranking for each cluster for controls, ALS and PD patients, respectively. The number between brackets indicates the number of top ranking clusters selected per each class. The fourth column indicates in which Brodmann area (BA) each cluster falls. The column AAL AREA indicates the corresponding parcel in the anatomical based solution; when more than one region is reported, the clusters lie on the boundary of parcels. The last column reports the number of voxels for each parcel.

### 3.5 Discussion

A novel framework based on clustering and stochastic rank aggregation has been evaluated using DMN maps from resting-state fMRI scans of ALS and PD patients and of healthy controls. As an alternative to clustering, a purely anatomical definition of brain parcels to extract regional DMN features for ranking, was also considered. While in the clustering-based analyses about 2% of the clusters (8 or 9 out of 405) were selected in the final rankings, in the case of anatomical defined areas up to one third of the parcels (30 out of 90) are part of the solution. One reason for this might be that anatomical areas are bigger and fewer compared to clusters and since the median is used as representative feature this might flatten the differences across subjects resulting in more conforming rankings. This would also explain why more anatomical parcels are shared in the rankings across classes. Indeed, observing the Venn diagrams in figure 3.2, most of the anatomical areas are shared among the three classes, while there are more class-specific regions when using clusters. It is interesting to observe that the anatomical parcels cover most of the DMN, that is the spatial component that was extracted by ICA, meaning that although not very selective, the approach based on the anatomical parcellation led to the subset of regions that contribute to this network. As mentioned above, clusters are

smaller than anatomical parcels and might therefore unveil differences at a finer granularity. Indeed, the example of figure 3.4, where a region included in the anatomical solutions of all three classes is compared to two clusters that are specific for the control group and one cluster that is specific for the ALS class, suggests that working with smaller regions might bring to light differences that are not evident at a higher level, because they are average out when considering a larger region. Nevertheless, most of the clusters in the solution are included in one of the anatomical parcels (see last column of Table 3.2), often with a similar ranking between the two approaches, demonstrating the consistency of the results. The fact that the clustering-based approach results in more localised regions, and that the features extracted from these regions exhibit a more stable behaviour across subjects of the same group, makes this method potentially better suited than the anatomical-based one when searching for a set of regional features characteristic of a specific neurological condition, i.e. what is usually called a neuromarker. If we study in detail the clusters selected for each class, we can observe that clusters 86, 95 and 160, that are class specific for PD, all lie in Brodmann area 40, in the inferior parietal cortex, that has been shown to be relevant for this disease in previous fMRI studies: in Tessitore et al. (2012b) a decreased functional connectivity of the bilateral inferior parietal cortex in the DMN was observed in patients with PD; the results of Amboni et al. (2015) suggest that a functional disconnection of the frontoparietal network could be associated with mild cognitive impairment in PD. Another class specific cluster for PD is cluster 297 in Brodmann area 39 in the medial temporal lobe, whose relation with PD has been investigated in Gorges et al. (2013) and Tessitore et al. (2012b), where a decreased connectivity in the DMN was observed between this area and the posterior cingulate cortex, and with the prefrontal cortex, respectively. Clusters 98 and 265, that are high-ranking in all three classes, lie in Brodmann area 10 in the prefrontal cortex, a region related to both ALS and PD: a weaker connectivity of the prefrontal region was observed in ALS patients in the DMN (Mohammadi et al., 2009; Agosta et al., 2013) and in the salience network (Trojsi et al., 2015); a deactivation of the medial prefrontal cortex was measured in PD patients in the DMN (van Eimeren et al., 2009; Gorges et al., 2013) and in the fronto-parietal network (Amboni et al., 2015); Tessitore et al. (2012a) showed that PD patients with freezing of gait present reduced functional connectivity within the executive-attention network in the middle frontal gyrus. Considering the five clusters that are specific to the ALS class, we observe that clusters 381 and 383 are located in the temporal lobe; cluster 314 is included in Brodmann areas 18 (in the occipital lobe) and 7, in the left precuneus, that has been previously associated to ALS in the study by Agosta et al. (2013), where this region exhibited enhanced connectivity in the DMN; cluster 223 covers Brodmann areas 23 and 24, corresponding to the

posterior and anterior cingulate cortex, mentioned in Tedeschi et al. (2012), where the DMN showed a disease-by-age interaction in the posterior cingulate cortex, and in Mohammadi et al. (2009) where the DMN showed less activation in ALS patients compared to controls; cluster 425 lies in Brodmann area 8 on the middle frontal gyrus, that has been associated with cognitive deficit in ALS patients in a PET study by Wicks et al. (2008), while in Terada et al. (2016) the gray matter volume measured within the right middle frontal gyrus in ALS patients was significantly lower than in healthy controls. If we observe the clusters that are class specific for control subjects, two of them (clusters 380 and 382) are on the middle temporal gyrus and one (cluster 386) on the fusiform gyrus. Both these regions are mentioned in studies on cortical thickness that investigated healthy aging as opposed to neurodegenerative disorders: in a work by Convit et al. (2000), the volumes of the fusiform gyrus and the middle (and inferior) temporal gyrus are shown to predict decline to Alzheimer's disease (AD) in non-demented elderly; while in a work of Hänggi et al. (2011), the volume of the right middle temporal gyrus revealed promising diagnostic values to distinguish AD from mild cognitive impairment. Another study (Huettel et al., 2001) investigated the aging-related changes of the haemodynamic response in regions surrounding the fusiform gyrus. Finally, three clusters (182, 197 and 341) fall in regions of white matter and might therefore be resulting from noise.

### 3.6 Conclusion

This chapter presented a data-driven methodology to detect regions that show a common behaviour in a class of subjects. Looking for commonalities instead of differences between groups is advantageous because the actual differences might be masked by noise of different origins. This approach takes into account both inter-subject variability and noise by excluding from the analysis brain parcels whose patterns of activation are incoherent across subjects of the same diagnostic group, while retaining regions for which a sufficient degree of consensus exist.

The proposed framework combines unsupervised clustering, consensus techniques and stochastic rank aggregation methods to automatically identify a small subset of regions of interest, without requiring prior hypotheses. However, the same framework can be applied with parcels derived from pre-existing brain atlases. In this work, it was applied on DMN maps derived from resting-state fMRI acquisitions of healthy controls, PD and ALS patients, and results are consistent with previous literature, thus indicating that this approach might be suitable to support early diagnosis of neurodegenerative diseases.

In future work, it would be interesting to apply this approach in task fMRI, where it is easier to detect activation loci and more prior knowledge is available

to define regions of interest (e.g., consider the specialization of motor cortex). This would allow to observe how the rankings of regions vary between healthy subjects and patients in the performance of specific tasks.

# Chapter 4

## Predicting individual differences

In most fMRI studies, investigators resort to averaging data across subjects, both for increasing signal-to-noise ratio (SNR) and to perform group comparisons, e.g., in clinical studies. However, the recent advances in MRI technologies, leading to higher field strength and shorter acquisition times, substantially improved SNR, thus allowing for investigations at single subject level (Dubois and Adolphs, 2016). Every brain is unique in its structural and functional organisation, and this intrinsic variability could be used to explain why people differ in their attitudes and abilities. Brain can be considered as the proximal cause of all behaviour, hence it is reasonable to search for the neural bases of behaviour in the brain. Correlations from resting-state fMRI have recently been used to predict a number of individual differences in cognitive abilities. In these studies, an estimate of the resting state functional connectivity (FC) network is derived for each subject in the following way: network nodes are defined, as spatial coordinates or as regions of interest; a time series is identified for each node; the strength of the connections between nodes is estimated, e.g., as the correlation of each time series with every other time series (Smith et al., 2013). Then, a (multivariate or univariate) statistic is derived from the network and used in combination with confound variables to predict an individual measure. However, motion artefacts can greatly influence the estimation of functional connectivity and consequently introduce bias when estimating its relationship with individual measures (Siegel et al., 2016). In this chapter, starting from the replication of a previous work that successfully predicted IQ in a cohort of subjects of the HCP dataset, we compare the effect of different denoising strategies on the outcome of the prediction. The same approach is then used to investigate whether functional connectivity can predict personality traits.

## 4.1 Denoising of rfMRI data

Contrary to task fMRI, that is more robust to artefactual influences because it relies on a predefined temporal model, resting-state data need to be preprocessed beyond minimal preprocessing, since the correlations between voxels' time series might be corrupted by artefacts spanning multiple voxels (Smith et al., 2013).

Subject motion and physiological fluctuations are two of the principal noise sources that are targeted by denoising techniques. Temporal filtering is usually applied to remove fluctuations outside the frequency range of interest for resting-state data (between  $\approx 0.01$  Hz and  $\approx 0.1$  Hz), since signal components related to cardiac pulsations and respiration have relatively high frequencies ( $\approx 1$  Hz and  $\approx 0.3$  Hz, respectively). Nevertheless, if scan repetition time is not short enough to avoid undersampling of the physiological fluctuations, these will appear as aliased frequencies in the band of interest (Caballero-Gaudes and Reynolds, 2017). Since the signal related to neural activity is expected to be present mainly in gray matter voxels, nuisance regressors can be derived from regions like white matter and ventricles where the signal is likely to be dominated by physiological noise (Liu, 2016). Similarly, global signal regression removes the global mean of the signal computed across all brain voxels, following the assumption that processes captured globally across the brain cannot be related to neural activity but are linked to confounding factors (Power et al., 2014), although it has been argued that in doing so also meaningful components are removed as well (Caballero-Gaudes and Reynolds, 2017). Motion during an acquisition has three main consequences: first, the tissue composition within a given voxel might change, especially at tissue interfaces, thus causing a change in the BOLD signal; second, movement alters the homogeneity of the magnetic field inducing local distortions; third, when movement occurs across successive slices, timing between excitations will change generating spin history effects (Murphy et al., 2013). To reduce the effect of head movement, volume registration is performed by aligning each scan to a reference volume through a rigid body transformation described by 3 translational and 3 rotational parameters. However, to remove any residual variance due to motion-related signal changes, the time series of realignment parameters are often used as nuisance regressors. Some models include also temporal derivatives and squared regressors of the realignment parameters to account for spin history effects (Friston et al., 1996). Censoring (or scrubbing) of high motion volumes has been proposed as an additional step to contrast changes in image intensity due to motion that cannot be explained by the realignment parameters (Power et al., 2012; Siegel et al., 2014), but, as pointed out in Carp (2013), censored points should be interpolated before performing temporal filtering to prevent motion artifacts in censored volumes from affecting the signal of retained frames. Alternatively, scan nulling regressors, modeled as a unit impulse function, can be added to the model (Sat-

terthwaite et al., 2013). To account for residual unmodeled noise, data-driven methods based on principal component analysis (Behzadi et al., 2007) and independent component analysis (Salimi-Khorshidi et al., 2014) have been proposed to detect noise-related components to be removed from the signal.

There is currently no consensus on the best denoising pipeline for resting-state fMRI data (Caballero-Gaudes and Reynolds, 2017; Siegel et al., 2016; Ciric et al., 2017; Murphy and Fox, 2017). The following section illustrates 14 different pipelines representative of the most common approaches found in literature. To conduct the analyses described in this chapter using a common framework, all pipelines were implemented in Python using well-established open source libraries for scientific computing, including SciPy, Numpy, NiLearn, NiPype and Scikit-learn.

### 4.1.1 Description of denoising pipelines

In the following, details are provided on the denoising steps constituting each pipeline. Each strategy is given a code name that will be used from now on to identify it. Figure 4.1 is a graphical summary of the pipelines, that shows that denoising steps can be grouped in seven categories, that are: scrubbing, normalization, detrending, tissue regression, global signal regression, motion regression and temporal filtering. Figure 4.2 highlights the order in which every step is performed and allows to observe that there are some steps that are very common among different strategies (demeaning, polynomial detrending, WM and CSF signal regression), while for others there is less consensus (e.g., global signal regression).

**Finn** This pipeline is based on the work of Finn et al. (2015) and consists of seven consecutive steps: first, the signal at each voxel is z-score normalized; then, using tissue masks computed with FSL, temporal drifts from CSF and white matter are removed with third degree Legendre polynomial regressors; subsequently, the mean signals of CSF and white matter are computed and regressed from gray matter voxels; translational and rotational realignment parameters and their derivatives are used as explanatory variables in motion regression; to perform temporal filtering, signals are low-pass filtered with a Gaussian kernel; as a last step, the temporal drift from gray matter signal is removed using again third degree Legendre polynomial regressors. As we can observe in figure 4.2, this pipeline comprises some denoising steps that are uncommon in literature. In addition, while other strategies tend to include all confounds in a single regression step, in this pipeline regression is performed in three consecutive steps.

The following nine pipelines are described in Ciric et al. (2017). If not stated

otherwise, all pipelines have three initial steps in common: the demeaning of each voxel's time series, the removal of linear and quadratic trends, and temporal filtering with a first order Butterworth filter with a passband between 0.01 and 0.08 Hz. The pipelines differ from each other for the adopted confound regression strategy.

**Ciric1** This pipeline regresses out from whole brain signal two confounds computed as the mean signal in WM and the mean signal in CSF.

**Ciric2** In this pipeline, the only explanatory variables used in confound regression are the 6 translational and rotational realignment parameters.

**Ciric3** This pipeline uses translational and rotational realignment parameters and their derivatives together with WM, CSF and whole brain mean signals in confound regression.

**Ciric4** This pipeline is an extension of model *Ciric2* that includes 6 motion parameters, 6 temporal derivatives, 6 quadratic terms, and 6 quadratic expansions of the derivatives of motion estimates for a total 24 regressors.

**Ciric5** This is an extension of model *Ciric3* that uses its 9 regressors plus their derivatives, quadratic terms, and squares of derivatives, for a total of 36 regressors.

**Ciric7** This pipeline uses the same 36 regressors as in *Ciric5*. Additionally, for each volume with an RMS (root-mean-squared) displacement that exceeded 0.25 mm, a regressor is added consisting of a unit impulse function with a value of 1 at the time point to be censored and 0 elsewhere (spike regression).

**Ciric9** This pipeline uses aCompCor (Muschelli et al., 2014), a method based on PCA, to derive 5 principal components each from the WM and CSF, that are used as regressors in addition to motion estimates and their temporal derivatives.

**Ciric13** This pipeline uses ICA-AROMA (Pruim et al., 2015), an ICA-based procedure for removal of motion-related variance from BOLD data, together with mean WM and CSF regressors.

**Ciric14** This pipeline combines ICA-AROMA with global signal regression.

The following two pipelines are described in Siegel et al. (2016). As a first step, data go through the HCP ICA-FIX denoising pipeline based on independent component analysis (Salimi-Khorshidi et al., 2014), that regresses out the signal of components classified as noise.

**SiegelA** After ICA-FIX, voxel signals are demeaned and detrended with a first degree polynomial.

**SiegelB** In addition to ICA-FIX, voxel demeaning and detrending as in model *SiegelA*, PCA-based method CompCor (Behzadi et al., 2007) is applied to derive CSF and white matter regressors that are used together with gray matter and whole brain mean signals in a single regression step. Also in this case as in *Ciric7*, regressors are added to remove the contributions of censored time points, identified as volumes with a frame-wise displacement greater than 0.25 mm and a variance of differentiated signal (DVARs) greater than 105% of the run median DVARs. Finally, a first order Butterworth band-pass filter between 0.01 and 0.08 Hz is used for temporal filtering.

The remaining two pipelines are described in Gordon et al. (2016).

**Gordon1** After voxel demeaning and detrending, confound regression is performed combining WM, CSF and whole brain signals with motion regressors derived by Volterra expansion (Friston et al. 1996)

**Gordon2** The only difference with *Gordon1* is that in this case temporal filtering is implemented with a discrete cosine transform and it is performed together with confound regression.

When performing temporal filtering after scrubbing, to prevent motion artifacts in censored volumes from affecting the signal of retained frames, censored time points were replaced with linear interpolation.

## 4.2 Data description

Data come from the public repository of the Human Connectome Project (HCP) (Van Essen et al., 2013). The HCP provides MRI data and extensive behavioural assessment from almost 1200 subjects. Acquisition parameters and minimal preprocessing of the resting-state fMRI data is described in the original publication (Glasser et al., 2013). Briefly, each subject underwent two sessions of resting-state fMRI on separate days (REST1 and REST2), each session with two separate 15 minute acquisitions generating 1200 volumes (customized Siemens

Skyra 3 Tesla MRI scanner, TR = 720 ms, TE = 33 ms, flip angle =  $52^\circ$ , voxel size = 2 mm isotropic, 72 slices, matrix = 104 x 90, FOV = 208 mm x 180 mm, multiband acceleration factor = 8). The two runs acquired on the same day differed in the phase encoding direction, left-right and right-left (which leads to differential signal intensity especially in ventral temporal and frontal structures). The HCP data were downloaded in their minimally preprocessed form, i.e. after motion correction, B0 distortion correction, co-registration to T1-weighted images and normalization to MNI space. The experiments described in section 4.3 are based on the Q2 release of HCP (136 subjects), while in section 4.4 all analyses are run on the full dataset of 1200 subjects.

### 4.3 Intelligence prediction from rfMRI

Individual differences in intelligence can be measured with psychometric tests assessing reasoning ability, processing speed, executive function, memory and spatial ability. However, it is believed that all these aspects are influenced by a common factor (Deary et al., 2010), denominated general intelligence (or *g* factor). This general factor can be assessed with cognitive tasks as the Raven's progressive matrices test, in which the subject is presented with a series of patterns, one of which is incomplete, and the task is to identify the missing tile, as showed in figure 4.3.

The challenge for cognitive neuroscience is to understand the neural structures and mechanisms underlying such a multifaceted ability. Early studies have linked intelligence to brain size, proportion of gray matter and cortical thickness (Brancucci, 2012), but one emerging theory is that intelligence is correlated to neural processing efficiency (Deary et al., 2010; Neubauer and Fink, 2009). For example, in Kievit et al. (2016), the authors built a hierarchical model to explain the interaction between intelligence, processing speed and white matter organisation and reported that individual differences in white matter anatomy predicted individual differences in processing speed, which in turn predicted over 58% of the variance in intelligence scores. Using diffusion tensor tractography to map white matter fibers, Li et al. (2009) provided evidence that subjects with higher IQ show a higher global efficiency of the brain anatomical network.

Other studies used graph analysis to assess the efficiency of functional networks built from rfMRI data, and significant correlation were found between IQ and functional efficiency, supporting the theory that more intelligent people tend to use less brain resources for completing a task (i.e., they tend to process information more efficiently). In van den Heuvel et al. (2009), it was found that IQ was negatively correlated with characteristic path length of functional connectivity networks, a measure of the average number of connections that have to be crossed to travel from each node to every other node in the network

that provides information about the level of global communication efficiency. In Cole et al. (2012), a lateral prefrontal cortex region’s activity was found to predict performance in a high control demand working memory task and also to exhibit high global connectivity, suggesting it may constitute a functional hub involved in control processes central to cognitive ability. In Song et al. (2008), the strengths of functional connections in the frontal and parietal lobes were found to be predictive for the variability of IQ scores across subjects.

It has been recently reported in Finn et al. (2015) that a measure of IQ based on the raw score on a 24-item version of the Raven’s Progressive Matrices can be predicted from individual whole brain functional connectivity networks. The authors of this work claim that they were able to predict IQ with a correlation of 0.5 between predicted and actual scores. Yet, it has been observed that data quality plays a relevant role in the estimation of FC, in that motion related artefacts can introduce bias if not properly accounted for Siegel et al. (2016): specifically, FC estimates are inflated for high motion time points and high motion participants (Burgess et al., 2016). More recently, Geerligs et al. (2017) suggested that head motion might constitute a biological trait per se and recommended against regressing motion out across participants, since it might have a dramatic impact on connectivity differences between individuals.

To understand to what degree preprocessing can affect predictions, in the following the performances of each of the denoising strategy described in the previous section are compared with respect to the task of predicting IQ from FC using the same data, parcellation and predictive model as in Finn et al. (2015). This model was chosen as a reference because it was shown to generalize to other datasets (Rosenberg et al., 2016; Yoo et al., 2018; Jangraw et al., 2018). Furthermore, to investigate whether these predictions are completely free of any residual motion information in the data, a predictive model was built for motion itself: if such a model is not able to predict motion it is likely that preprocessing succeeded in disrupting motion contribution; otherwise, it might be the case that motion carries all the information. Indeed, as shown in figure 4.4 for the HCP dataset, IQ and motion are correlated, and this relationship holds even after discarding acquisitions of high movers.

### 4.3.1 Methods

Subject scans with a root-mean-squared (RMS) frame-to-frame head motion estimate exceeding 0.14 mm were removed from the analysis, leaving a sample of 115 subjects in REST1 session and a sample of 114 subjects in REST2 session. After performing one of the preprocessing pipelines presented in section 4.1.1, a volumetric parcellation of the brain into 268 nodes (Shen et al., 2013) was applied to each subject’s data and the average time series was computed for each parcel. Functional connectivity between a pair of regions was estimated as

the Pearson’s correlations coefficient between the two corresponding time series. FC matrices were averaged across runs acquired with left-right and right-left phase encoding in each session. Then a univariate regression model was built where the dependent variable is the score to be predicted and the explanatory variable is a scalar value that summarize the FC network strength (i.e., the sum of edge weights). A simple filtering approach was used to select features that are correlated with the behavioural score on the training set. Two distinct models are built using edges of the network that are positively and negatively correlated with the score, respectively, with a  $p$ -value less than 0.01. In the following we refer to positive and negative models to distinguish between the two approaches. The prediction ability of the models is assessed using a leave-one-out cross-validation scheme. The IQ score is the 24-item version of the Raven’s Progressive Matrices (PMAT24\_A\_CR) as in Finn et al. (2015), while motion is measured using the mean RMS frame-to-frame displacement.

### 4.3.2 Results

In the original work by Finn et al. (2015) the authors reported only the result for the positive model for IQ on session REST1 ( $\rho = 0.5$ ). Figure 4.5 highlights the results for the positive models for both IQ and motion. Additionally, the results on session REST2 are reported to assess the stability of the predictions across sessions held at different times. The correlations between actual and predicted scores for all experiments are showed in tables 4.3.2 and 4.3.2.

The best performing pipeline in predicting IQ was *Ciric7*, ranking first in both sessions, followed closely by *Finn* strategy, ranked second in REST1 and third in REST2. In session REST1, eight pipelines out of 14 yielded significant correlations, and ten in session REST2, but just seven of them have significant results in both sessions, with an average correlation of 0.30 in session REST1 and of 0.29 in session REST2.

The trend in the IQ predictions across denoising strategies is fairly stable between the two sessions, while we can observe more variability comparing RMS predictions in session REST1 and in session REST2 (figure 4.5), even if there was no significant difference between the average RMS measure across subjects between the two sessions (0.0850 vs. 0.0853). In REST1 the motion predictions seem to follow IQ predictions, except for pipelines *SiegelB*, *Gordon1* and *Gordon2* that scored low on IQ predictions and very high on motion prediction, while in REST2 motion predictions achieved in general lower correlations, with the same exception for pipelines *SiegelB*, *Gordon1* and *Gordon2* that seem to predict motion very well. This might indicate that, after all, the selected features are better predictors for IQ than for motion.

Yet, the fact that in many cases motion could be predicted from FC suggests that the preprocessing did not succeed in isolating and removing the effects

of head motion. It is interesting to note that the negative models performed sensibly worse on motion predictions (table 4.3.2): correlations are either not significant or negative, meaning that the models predict in the wrong direction. This is in line with the results of Burgess et al. (2016), that reported that motion artefacts tend to inflate functional connectivity estimates, hence the contribution of motion should be found in the FC edges positively correlated with RMS score.

One encouraging datum is that when comparing the functional connections used to predict IQ and motion, there were very few overlapping edges (less than 10 on average) and this might be an indicator that FC contains information related to intelligence that is not resulting from the contribution of motion. However, in none of the predictive models the selected connections exhibited regional specificity. While this contrasts with previous results on IQ where edges predictive of intelligence were found in fronto-parietal regions (Song et al., 2008; Cole et al., 2012; Finn et al., 2015), it is in accord with the theory of globally distributed effects of head motion (Burgess et al., 2016).

In conclusion, there seems to be a relationship between the strength of individual functional connectivity networks and the IQ score, and this association emerges even when varying consistently the preprocessing strategy. Yet, the role played by head motion needs to be taken into account and the variability of the results means that preprocessing can greatly affect results. Some of these aspects will be further investigated in the following section, where the IQ experiments are repeated on a larger sample for selected pipelines and a multivariate model for regression is examined as well.

## 4.4 Resting-state fMRI and personality

Personality traits are relatively stable dispositions that influence behavioural, social, and emotional expression. According to the the Big Five model (McCrae and Costa, 1986), the major features of personality can be described by five factors, namely openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism, that can be described as follows:

- **Openness to experience** concerns an individual's willingness to try new things and the ability to think outside the box. It is also sometimes called intellect or imagination.
- **Conscientiousness** can be described as the tendency to control impulses and act in socially acceptable ways, to be organized and show self-discipline, act dutifully, aim for achievement, and prefer planned rather than spontaneous behaviour.

- **Extraversion** is the tendency to seek out opportunities for social interaction, to feel comfortable with others, and prone to action rather than contemplation.
- **Agreeableness** concerns how well people get along with others. It is the tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others.
- **Neuroticism** concerns anxiety, sadness, worry, and low self-esteem. It describes the tendency to experience unpleasant emotions easily, such as anger, depression, and vulnerability.

Since personality traits are stable over time, one might expect that brain measures that are similarly stable over time (e.g., structural, connectional and neurochemical measures) are the most promising candidates for predicting such traits. Resting-state data do not explicitly engage cognitive processes that are thought to be related to personality traits. However, they are used to study correlated self-generated activity between brain areas while a subject is at rest, and these correlations are thought to reflect stable aspects of brain organization. Several studies report an association between Neuroticism and Extraversion and properties of brain connectivity networks. In Gao et al. (2013), in a sample of 71 subjects, Extraversion was correlated with the normalized clustering coefficient of individual networks (a measure of modularity), while Neuroticism was associated with high betweenness centrality of nodes situated in the right precentral gyrus, right caudate nucleus, right olfactory cortex, and bilateral amygdala. In Aghajani et al. (2014), a seed-based correlation analysis focused on the amygdala was performed on a sample of 50 subjects with the following results: higher Neuroticism was observed in presence of increased amygdala connectivity with the precuneus, and decreased amygdala connectivity with the temporal poles, insula, and superior temporal gyrus; higher Extraversion scores were associated with increased amygdala connectivity with the putamen, temporal pole, insula, and several regions of the occipital cortex. Lei et al. (2013) investigated the association between personality and the scale-free dynamics of default mode network and found that Extraversion correlated with the Hurst exponent, a measure of long memory of temporal dynamics, in a sample of 20 subjects. More recently, Pang et al. (2016) applied Granger causality analysis to the resting state functional networks of 70 subjects and reported a positive correlation between Extraversion and the influence from the right inferior occipital gyrus (IOG) to the left amygdala, and from the bilateral IOG to the right amygdala, while Neuroticism was found to be associated with an increased influence from right amygdala to right middle frontal gyrus and a decreased influence from right precuneus to right amygdala.

In a study of the interaction of personality traits with cingulate functionality in

the prediction of stressor-evoked cardiovascular reactivity (Sheu et al., 2011), Agreeableness was correlated with posterior and perigenual cingulate connectivity in a sample of 39 subjects; additionally, cingulate cortex connectivity was found to mediate the covariation between Agreeableness and blood pressure reactivity.

In Adelstein et al. (2011), the Big Five traits are used to predict patterns of functional connectivity in a sample of 39 subjects, and each dimension predicted connectivity in non-overlapping regions responsible for cognitive and affective processing, although selected connections were inconsistent across subjects.

The association between the Big Five traits and the connectivity of the default mode network was investigated in Beaty et al. (2016) using structural equation modeling. After controlling for age, gender and intelligence (measured with the Intelligence Structure Battery), Openness was the only significant predictor of the global efficiency of the default mode network, while Extraversion, Agreeableness, and Conscientiousness had marginal effects.

Beyond the five-factor model, other personality dimensions have been studied using resting state fMRI, as harm avoidance (Baeken et al., 2014), procrastination (Wu et al., 2016), and creativity (Takeuchi et al., 2012; Beaty et al., 2014; Jiao et al., 2017). In Deris et al. (2017), a study on the correlation of functional connectivity of the amygdala with the Affective Neuroscience Personality Scales reported an association between connection in the basolateral amygdala and the SADNESS scale.

However, previous studies suffer from several methodological limitations. First of all, essentially all studies to date have been severely underpowered due to small sample sizes (Button et al., 2013; Schönbrodt and Perugini, 2013; Yarkoni, 2009). Second, most studies have failed to use a predictive or replication framework, making their generalizability unclear. To overcome these issues, in this work a predictive framework, together with a built-in replication, was applied to a large, homogeneous resting-state fMRI dataset.

#### 4.4.1 Methods

**Personality factors** The 60 item version of the Costa and McCrae Neuroticism Extraversion and Openness Five Factor Inventory (NEO-FFI) was administered to HCP subjects. The NEO-FFI is a self-report questionnaire, the abbreviated version of the 240-item Neuroticism Extraversion and Openness Personality Inventory Revised (NEO-PI-R). For each item, participants reported their level of agreement on a 5-point Likert scale, from strongly disagree to strongly agree. As the Big Five were correlated with each another in the selected sample of subjects, factor analysis (with varimax rotation) was performed on the 60 item scores to extract two superordinate factors (DeYoung, 2006) variably referred to in the literature as alpha (or socialization or stability) and beta (or personal

growth or plasticity).

**Subject selection** The total number of subjects in the 1200-subject release of the HCP dataset is  $N=1206$ . The following criteria were applied to select subjects for the analyses.

1. Subjects must have completed all resting-state fMRI scans, as well as the Raven’s matrices intelligence test, the NEO-FFI and the Mini Mental Status Exam (208 subjects excluded).
2. One subject with a score MMSE below 26 was excluded, since it could indicate severe cognitive impairment or delirium.
3. Subjects with high motion were excluded. Specifically, subjects with a root-mean-squared frame-to-frame head motion estimate exceeding 0.15 mm in any of the 4 resting-state runs (72 subjects excluded).
4. As an attempt to identify subjects that did not answer the self-report NEO-FFI in a consistent manner, a robust outlier detection method was used (Leys et al., 2018) in the five-dimensional space spanned by the five personality factors. As per recommendations in Leys et al. (2018), a robust Mahalanobis distance was used based on the Minimum Covariance Determinant (with a breakdown point of 0.25), and a threshold based on the chi-square value with 5 degrees of freedom for quantile 99.9 (58 subjects excluded).

The final sample consisted of 867 subjects, including 402 males (age range 22-36).

**Removal of potential confound** Gender, age, IQ, brain size and multi-band reconstruction algorithm (which changed in the third quarter of HCP data collection) were regressed from each of the personality scores to remove their confounding effects. The multiple linear regression used for removing the variance shared with confounds was performed on training data only (in each cross-validation fold during the prediction analysis), and then the fitted weights were applied to both the training and test data. This is critical to avoid any leakage of information, however negligible, from the test data into the training data.

**Resting-state data preprocessing** Three different pipelines (*Finn*, *Ciric7* and *SiegelB*, described in section 4.1.1) were selected and compared as representative alternatives of the implemented denoising strategies.

**Inter-subject alignment and parcellation** The most common approach to align subjects is to warp individual brains to a common volumetric template, typically MNI152. However, it is possible to exploit functional information alongside anatomical information, as in the multimodal surface matching (MSM) framework (Robinson et al., 2014). MSM-All aligned data, in which inter-subject registration uses individual cortical folding, myelin map, and resting-state fMRI correlation data, are available for download from the HCP database. Two different approaches were compared: one based on the classical volumetric alignment with a volumetric parcellation of the brain into 268 nodes (Shen et al., 2013; Finn et al., 2015), and one based on MSM-All data and a multi-modal parcellation that was specifically derived from these data (Glasser et al., 2016).

**Functional connectivity estimation** Following the same approach described in section 4.3, time series were extracted for each node of the network by averaging data across voxels within each brain parcel, and a connectivity matrix was built computing pairwise Pearson’s correlations between time series. FC matrices were averaged across runs acquired with left-right and right-left phase encoding in each session, i.e. two FC matrices were derived per subject, one for REST1 and one for REST2.

**Test-retest comparisons** In the following, three metrics are described that were adopted to compare the FC matrices produced by the three different denoising strategies.

- **Identification Success Rate (ISR):** identification of subject  $S$  is successful if, out of all subjects’ FC matrices derived from REST2, subject  $S$ ’s is the most highly correlated with subject  $S$ ’s FC matrix from REST1. The ISR gives an estimate of the reliability and specificity of the entire FC matrix at the individual subject level, and is influenced both by within-subject test-retest reliability as well as by discriminability among all subjects in the sample.
- **Similarity across sessions:** it is desirable to have stable similarities (and differences) between all subjects across repeated testing sessions. Following an approach introduced in Geerligs et al. (2015a), the pairwise similarity between subjects was computed separately for sessions REST1 and REST2, constructing a  $N_{\text{subjects}} \times N_{\text{subjects}}$  matrix for each session. The two matrices were then compared using Pearson’s correlation coefficient.
- **Behavioural utility:** for each entry of the connectivity matrix, the correlation between the entry and a trait across subjects was computed and

matrix was built representing the relationship of each entry to this trait, separately for session REST1 and session REST2. These two matrices were then compared using Pearson’s correlation coefficient as a similarity measure. The more edges reliably correlate with the stable trait, the higher the correlation between session REST1 and session REST2 matrices (Geerligs et al., 2015a).

**Predictive models** The choice of the predictive model is limited by the curse of dimensionality: although the HCP dataset is one of the largest available to date, the number of subjects (in this case 867) is one order of magnitude smaller than the number of features (the edges of the connectivity network), hence too complex models, e.g., non linear models, are likely to overfit the data. For this reason, two linear models were chosen, of which one is univariate and one multivariate. In both cases, to alleviate the problem of high dimensionality, a filter method was applied to select only features correlated with the score to be predicted (on the training set). The univariate model is the same described in section 4.3.1, since it was able to reproduce a previous finding (Finn et al., 2015). This method has the advantage of being extremely fast to compute, but it is limited in that it condenses all the information contained in the connectivity network in a single measure, it does not account for any interactions between edges and it arbitrarily builds two separate models (one for positively correlated edges and one for negatively correlated edges), but does not offer a way to integrate them. The multivariate model is the elastic net, a regularized regression method that linearly combines L1- (lasso) and L2- (ridge) penalties to shrink some of the regressor coefficients toward zero, thus retaining just a subset of features. The lasso model performs continuous shrinkage and automatic variable selection simultaneously, but in the presence of a group of highly correlated features, it tends to arbitrarily select one feature from the group. With high-dimensional data and few examples, the ridge model has been shown to outperform lasso; yet it cannot produce a sparse model since all the predictors are retained. Combining the two approaches, elastic net is able to do variable selection and coefficient shrinkage while retaining groups of correlated variables. 3-fold nested cross-validation was used to choose the alpha parameter that weighs the penalty term.

**Statistical assessment of predictions** The HCP dataset includes data of twins subjects and their non-twin siblings. To avoid any bias resulting from the inclusion of genetically related subjects in the training and in the test sets, all analyses were validated using a leave-one-family-out cross-validation scheme. The performances of the predictive models were measured computing the correlation between predicted and actual scores ( $\rho$ ). To assign statistical significance to the results, permutation testing would be preferable to parametric tests, since

in a cross-validation scheme the folds are not independent from each other. However, to reduce the computational demand, permutation tests were performed only for the analyses that passed the parametric significance threshold, since in general these results tend to be more optimistic. The null distribution was computed performing the whole analysis on 1000 random permutations of the scores between subjects, while keeping everything else unchanged.

#### 4.4.2 Results

**Test-retest reliability** Results regarding test-retest reliability were consistent across denoising pipelines. All three strategies scored high ( $> 87\%$ ) on the Identification Success Rate (ISR), with the best result attained by pipeline *Ciric7* (Figure 4.6a). Intra-session pairwise similarities between subjects were similar for all strategies and also in this case model *Ciric7* reported the highest score (Figure 4.6b). The best score for behavioural utility was obtained with pipeline *Finn* (Figure 4.6c). We can observe that Neuroticism has lower scores in terms of behavioural utility (4.6c), indicating that it is unlikely that connectivity matrices contain information about this specific personality trait. In general, analyses carried out with MSM-All alignment outperformed those based on MNI alignment.

**IQ prediction** The IQ prediction experiment presented in section 4.3 was repeated on the full HCP dataset, but in this case removing the confounding factors prior to performing regression and using a leave-family-out cross-validation scheme (4.4.1). In all experiments, positive correlations were found between predicted and actual scores (Figure 4.7a). The mean effect size across all 12 alternative analyses was  $\rho=0.147$  for session REST1, and  $\rho=0.123$  for session REST2 (see Table 4.3). Comparing the results obtained with pipeline *Finn* and the alignment in MNI space with those presented in section 4.3, we can observe that on the extended sample the prediction is barely significant in REST1 (at parametric  $p<0.05$ ) and not significant in REST2. It might be that in the previous analysis some variance from confounds was used in the predictions. Results are consistently better in the experiments based on the MSM-All alignments (4.3). In figure 4.7b, the null distributions of the performance scores for the univariate and multivariate models are reported for pipeline *Finn*. As mentioned above, the parametric statistics underestimate the confidence intervals, overestimating significance. Considering non-parametric statistics, the IQ prediction with *Finn* strategy and MNI alignment does not yield to significant results using the univariate model (permutation  $p=0.063$  vs. parametric  $p=0.005$ ) as in the previous experiments. Results are however significant when using the multivariate model and MSM-All alignment (permutation  $p=0.001$ ), showing how the choice of the model can affect statistical thresholds.

**Prediction of personality factors** In general, the results of prediction were quite poor (see Table 4.3), with the highest correlation between predicted and actual score equal to 0.265 with a permutation  $p$ -value of 0.003 (while the parametric  $p$ -value is  $2.46 \times 10^{-9}$ ), attained with pipeline *Finn*, MSM-All alignment and the multivariate model for prediction. Some experiments produced negative correlations, hinting at the unreliability of the models. However, we can observe some trends in the results. As mentioned above, experiments based on the MSM-All alignment and parcellation tend to score higher and results in the REST2 session are better than results in the REST1 session. It is interesting to note that the REST2 session was held closer in time to the personality assessment. Elastic net yielded overall better predictions than the univariate model. No significant results were obtained for Conscientiousness, Agreeableness, and Neuroticism, as well as the superordinate factor  $\alpha$  derived primarily from these traits. By contrast, Openness and Extraversion, as well as their corresponding superordinate factor  $\beta$  showed considerably better predictability (mean correlations of  $\rho=0.11$ ,  $\rho=0.08$  and  $\rho=0.12$ , respectively). Finally, the variability across the 12 different experimental settings demonstrates how sensitive the task of predicting individual differences is with respect to preprocessing and modelling choices.

## 4.5 Conclusion

In this chapter, we investigated the potential of using individual functional connectivity networks to predict intelligence and several personality factors. Comparing different denoising strategies and using data collected in two distinct fMRI sessions, we observed that the estimated FC networks were stable across time, a fundamental prerequisite when trying to predict individual traits that are expected to be stable in time as well. After replicating a previous finding on the prediction of IQ, thus confirming that this framework is indeed able to predict individual differences, the same approach was extended to personality factors. A total of 24 analyses were run, choosing among 2 resting sessions, 3 preprocessing pipelines, 2 predictive models and 2 alignment and parcellation schemes. Results showed that the best performances were attained with the MSM-All multimodal parcellation scheme and elastic net as predictive model. The only personality traits that could be predicted (although with rather low scores, approaching 0.2 in the best cases) were Extraversion and Openness and the superordinate factor  $\beta$ . In particular the  $\beta$  factor was the most reliably predicted of all, with an effect size similar to the prediction of IQ (mean  $\rho=.12$ ).

However significant in a large enough population of subjects like the one considered here, a correlation of 0.2 corresponds to just 4% of explained variance. The first conclusion that we can draw from this datum is that previous stud-

ies on personality, that were based on smaller samples and had lower statistical power, could not detect an actual effect. Additionally, predictions were unstable with respect to changes in preprocessing or across sessions. It seems therefore unlikely that a connectome-based model could explain the causal mechanisms bringing personality into being, but this does not imply that functional predictors cannot be used as markers of personality, intended not as neural correlates of individual traits, but as features that a computational model can learn to map to personality factors, in a supervised fashion. To explore this possibility, the first step would certainly be that of obtaining larger samples. More complex, non linear model could reach higher accuracies, but are inapplicable given the current sample sizes. The latter problem might be however alleviated adopting advanced dimensionality reductions methods or more sophisticated tools derived from network analysis. Further advances would also be possible with longitudinal studies, that monitoring the evolution of the brain over time would allow to account for the stability and variability of individual traits.

To effectively assess whether these findings reflect the inability to predict personality from functional connectivity data or rather the unsuitability of the chosen models, further replications and a more rigorous statistical evaluation of the results based on permutation testing would be needed, together with the examination of additional preprocessing alternatives. One concluding consideration is that the adopted psycho-metric tests might reflect just socio-cultural aspects and not biological individual differences, therefore a joint goal for psychology, personality and cognitive neuroscience could be that of defining a new taxonomy for personality and intelligence that better correlates with neurobiological substrates.

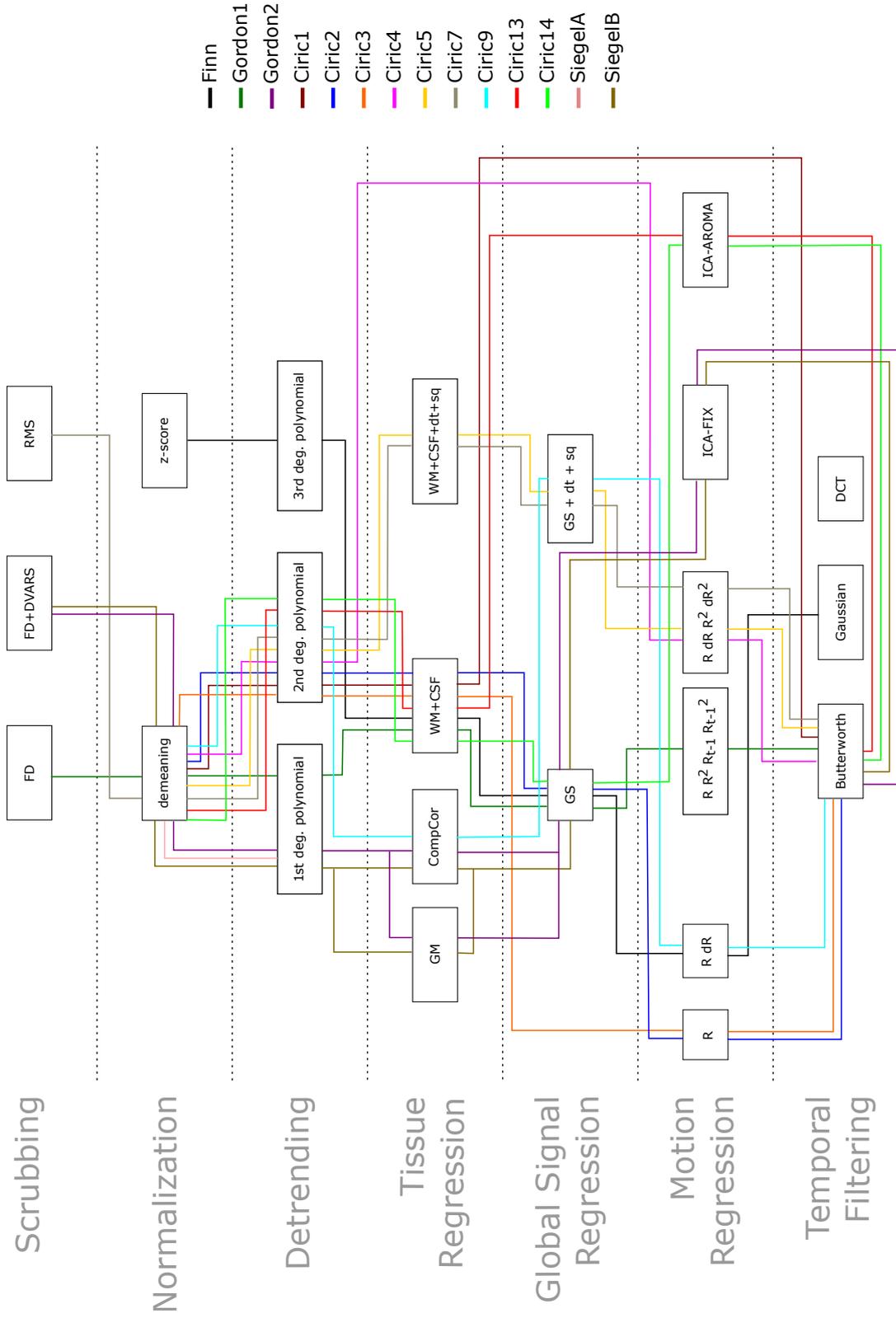


Figure 4.1: Pipeline flow chart.

	demean	z-score	poly detrend 1	poly detrend 2	poly detrend 3	Butterworth filter	Gaussian filter	DCT	R	R dR	R dR R2 dR2	R R2 R-1 (R-1)2	ICA-AROMA	ICA-FIX	CompCor	GM	WM	CSF	WMCSF+dt+sq	Scrubbing FD	Scrubbing DVARS	Scrubbing RMS	GS	GS+dt+sq
Finn	1			2		5				4							3	3					6	
Circ1	1		2	3													4	4						
Circ2	1		2	3				4																
Circ3	1		2	3				4									4	4					4	
Circ4	1		2	3						4														
Circ5	1		2	3						4									4					4
Circ7	2		3	4						5									5			1		5
Circ9	1		2	3					4						4									4
Circ13	1		2	3								4					4	4						
Circ14	1		2	3									4				4	4					4	
SiegelA	1	2																						
SiegelB	2	3			5								4	4	4					1	1		4	
Gordon1	2	3			5						4						4	4		1			4	
Gordon2	2	3			5	4					4						4	4		1			4	

Figure 4.2: Pipeline flow chart.

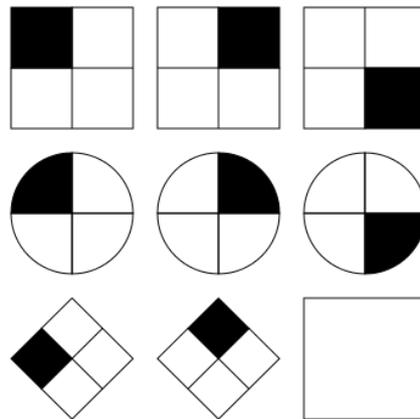
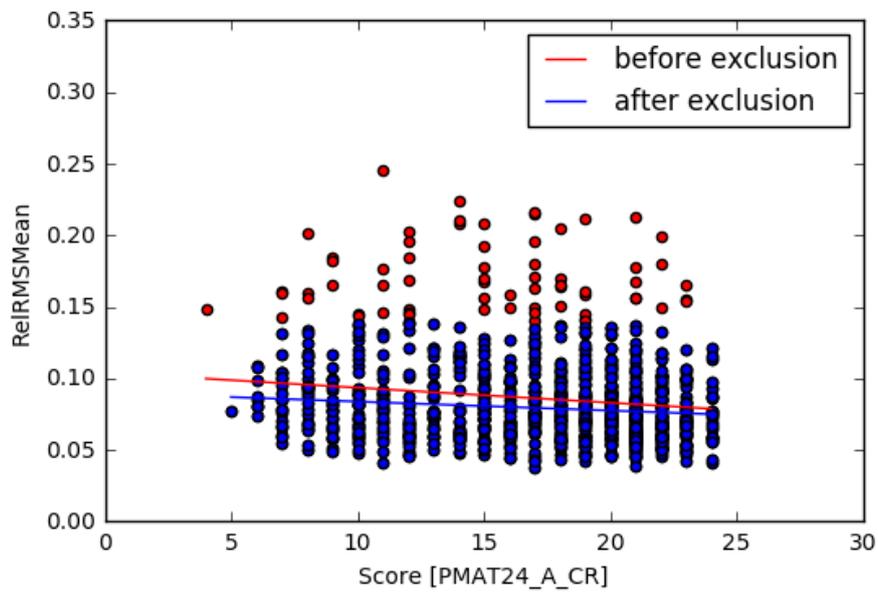


Figure 4.3: An item of the Raven's progressive matrices test.



With all subjects:  $\text{corr}(\text{IQ}, \text{motion}) = -0.147$  ( $p = 0.000$ )  
After discarding high movers:  $\text{corr}(\text{IQ}, \text{motion}) = -0.134$  ( $p = 0.000$ )  
Keeping 888 subjects [420 M]

Figure 4.4: Correlation between IQ score (PMAT24\_A\_CR) and motion (RMS) in the HCP dataset, before and after the exclusion of high motion subjects.

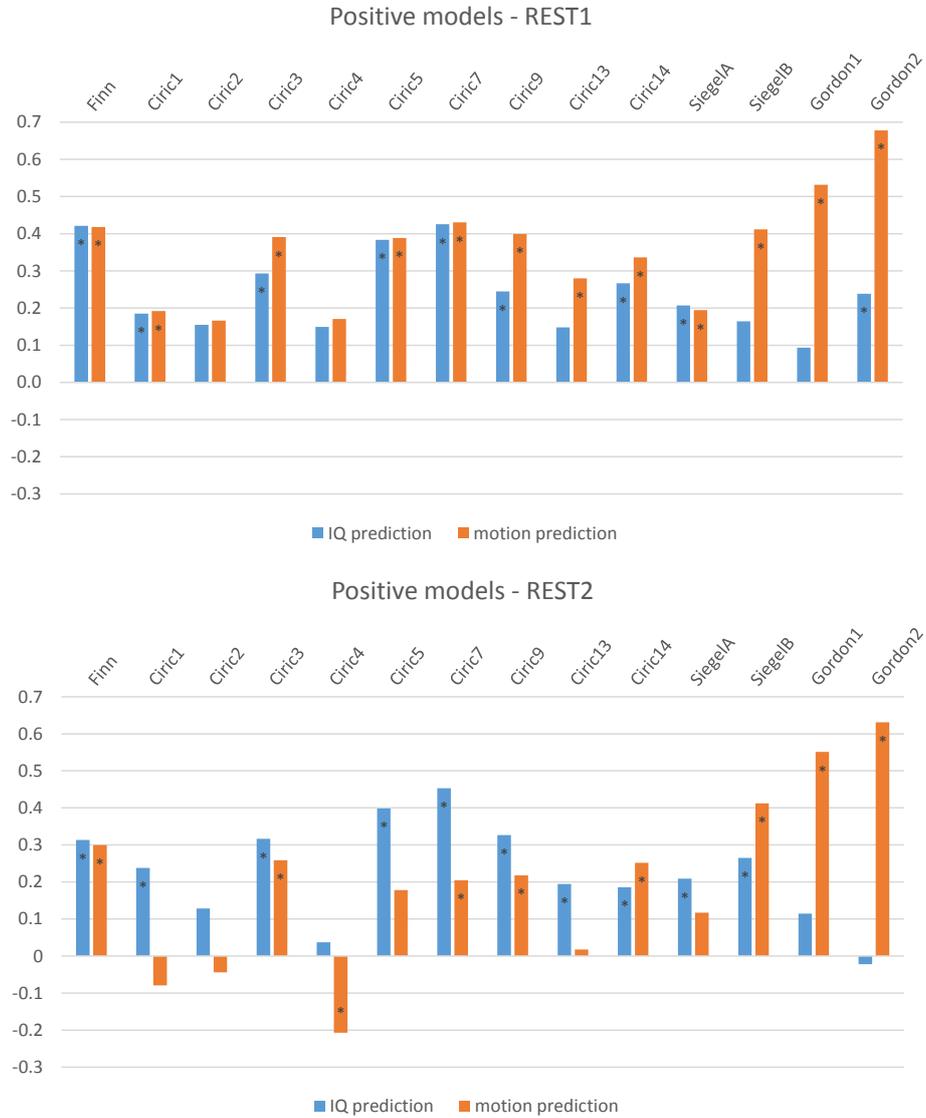


Figure 4.5: Prediction results obtained with positive models for session REST1 (top) and REST2 (bottom). The bars represent the correlation between predicted and actual scores. Correlations with a  $p$ -value  $< 0.05$  are marked with a star.

<b>IQ REST1</b>	$\rho^+$	$p\text{-value}^+$	$\rho^-$	$p\text{-value}^-$
Finn	<b>0.4213</b>	<b>2.75E-06</b>	<b>0.3013</b>	<b>1.07E-03</b>
Ciric1	<b>0.1852</b>	<b>4.76E-02</b>	<b>0.2378</b>	<b>1.05E-02</b>
Ciric2	0.1550	9.82E-02	0.0277	7.69E-01
Ciric3	<b>0.2933</b>	<b>1.46E-03</b>	<b>0.3928</b>	<b>1.41E-05</b>
Ciric4	0.1493	1.11E-01	0.1770	5.85E-02
Ciric5	<b>0.3835</b>	<b>2.33E-05</b>	<b>0.3872</b>	<b>1.91E-05</b>
Ciric7	<b>0.4253</b>	<b>2.15E-06</b>	<b>0.3890</b>	<b>1.73E-05</b>
Ciric9	<b>0.2445</b>	<b>8.44E-03</b>	<b>0.3596</b>	<b>7.91E-05</b>
Ciric13	0.1480	1.15E-01	0.1190	2.05E-01
Ciric14	<b>0.2666</b>	<b>3.97E-03</b>	<b>0.2361</b>	<b>1.11E-02</b>
SiegelA	<b>0.2072</b>	<b>2.63E-02</b>	-0.0318	7.35E-01
SiegelB	0.1647	7.87E-02	0.1441	1.25E-01
Gordon1	0.0934	3.23E-01	0.0054	9.55E-01
Gordon2	<b>0.2383</b>	<b>1.07E-02</b>	-0.0835	3.77E-01
<b>IQ REST2</b>	$\rho^+$	$p\text{-value}^+$	$\rho^-$	$p\text{-value}^-$
Finn	<b>0.3136</b>	<b>6.79E-04</b>	<b>0.3197</b>	<b>5.26E-04</b>
Ciric1	<b>0.2378</b>	<b>1.08E-02</b>	<b>0.2549</b>	<b>6.20E-03</b>
Ciric2	0.1286	1.73E-01	0.1601	8.89E-02
Ciric3	<b>0.3164</b>	<b>6.05E-04</b>	<b>0.3848</b>	<b>2.36E-05</b>
Ciric4	0.0373	6.94E-01	0.1667	7.62E-02
Ciric5	<b>0.3985</b>	<b>1.13E-05</b>	<b>0.4289</b>	<b>1.92E-06</b>
Ciric6	<b>0.4531</b>	<b>4.14E-07</b>	<b>0.3888</b>	<b>1.91E-05</b>
Ciric9	<b>0.3267</b>	<b>3.89E-04</b>	<b>0.2776</b>	<b>2.79E-03</b>
Ciric13	<b>0.1944</b>	<b>3.82E-02</b>	0.1514	1.08E-01
Ciric14	<b>0.1856</b>	<b>4.81E-02</b>	<b>0.2852</b>	<b>2.10E-03</b>
SiegelA	<b>0.2087</b>	<b>2.59E-02</b>	0.1086	2.50E-01
SiegelB	<b>0.2652</b>	<b>4.35E-03</b>	<b>0.2906</b>	<b>1.71E-03</b>
Gordon1	0.1145	2.29E-01	<b>-0.2322</b>	<b>1.37E-02</b>
Gordon2	-0.0221	8.17E-01	0.0069	9.42E-01

Table 4.1: IQ prediction results for session REST1 (top) and REST2 (bottom). The values indicate the correlations between predicted and actual scores ( $\rho$ ) and the associated  $p$ -values for positive (+) and negative (-) models. Correlations with a  $p$ -value  $< 0.05$  are in bold.

RMS REST1	$\rho^+$	$p\text{-value}^+$	$\rho^-$	$p\text{-value}^-$
Finn	<b>0.418051</b>	<b>3.34E-06</b>	-0.02169	0.817996
Ciric1	<b>0.191935</b>	<b>0.039885</b>	-0.1279	0.173149
Ciric2	0.166155	0.075947	-0.10114	0.282158
Ciric3	<b>0.390901</b>	<b>1.57E-05</b>	-0.02137	0.820668
Ciric4	0.170949	0.067749	-0.07583	0.420558
Ciric5	<b>0.3886</b>	<b>1.77E-05</b>	-0.01756	0.852275
Ciric6	<b>0.430703</b>	<b>1.55E-06</b>	-0.04332	0.645698
Ciric9	<b>0.399074</b>	<b>9.97E-06</b>	0.030367	0.747325
Ciric13	<b>0.279706</b>	<b>0.002467</b>	-0.06066	0.5196
Ciric14	<b>0.336586</b>	<b>0.000235</b>	-0.0024	0.979665
SiegelA	<b>0.194479</b>	<b>0.037278</b>	<b>-0.21659</b>	<b>0.020075</b>
SiegelB	<b>0.412015</b>	<b>4.76E-06</b>	<b>-0.24096</b>	<b>0.009484</b>
Gordon1	<b>0.531467</b>	<b>1.17E-09</b>	-0.08242	0.383349
Gordon2	<b>0.67818</b>	<b>1.14E-16</b>	-0.02027	0.83047
RMS REST2	$\rho^+$	$p\text{-value}^+$	$\rho^-$	$p\text{-value}^-$
Finn	<b>0.2996</b>	<b>1.20E-03</b>	-0.0978	3.01E-01
Ciric1	-0.0790	4.04E-01	-0.1843	4.97E-02
Ciric2	-0.0439	6.43E-01	-0.0122	8.98E-01
Ciric3	<b>0.2586</b>	<b>5.47E-03</b>	-0.0135	8.87E-01
Ciric4	<b>-0.2070</b>	<b>2.71E-02</b>	0.0113	9.05E-01
Ciric5	0.1781	5.80E-02	-0.0889	3.47E-01
Ciric6	<b>0.2045</b>	<b>2.90E-02</b>	-0.0433	6.47E-01
Ciric9	<b>0.2178</b>	<b>1.99E-02</b>	<b>-0.2196</b>	<b>1.89E-02</b>
Ciric13	0.0173	8.55E-01	-0.0597	5.28E-01
Ciric14	<b>0.2517</b>	<b>6.91E-03</b>	-0.0240	8.00E-01
SiegelA	0.1170	2.15E-01	-0.1487	1.14E-01
SiegelB	<b>0.4120</b>	<b>5.24E-06</b>	<b>-0.2283</b>	<b>1.46E-02</b>
Gordon1	<b>0.5513</b>	<b>3.01E-10</b>	0.0876	3.59E-01
Gordon2	<b>0.6309</b>	<b>8.91E-14</b>	-0.0348	7.16E-01

Table 4.2: Motion prediction results for session REST1 (top) and REST2 (bottom). The values indicate the correlations between predicted and actual scores ( $\rho$ ) and the associated  $p$ -values for positive (+) and negative (-) models. Correlations with a  $p$ -value  $< 0.05$  are in bold.

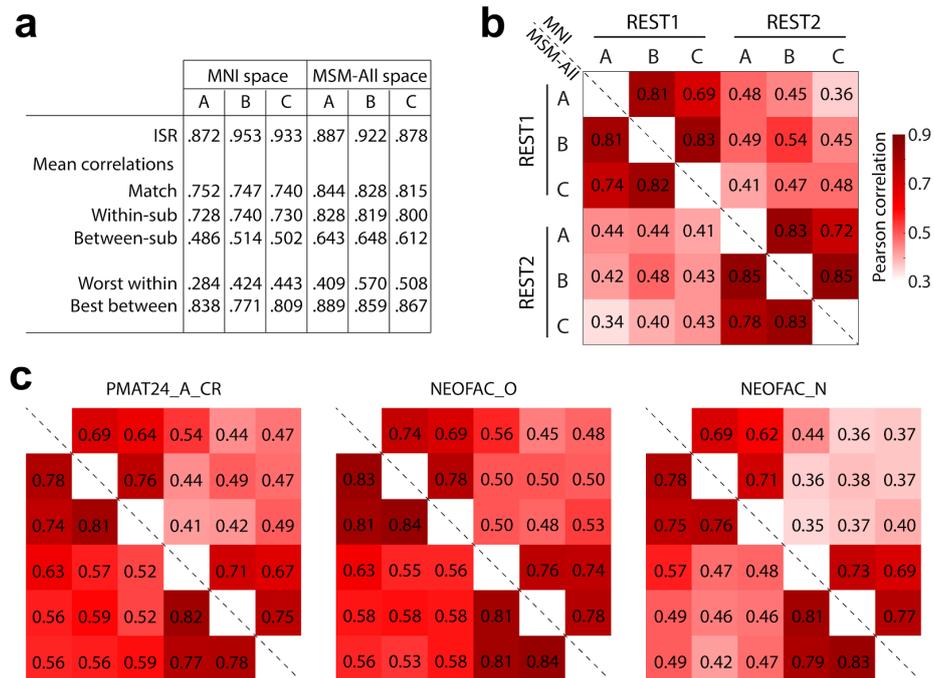


Figure 4.6: Test-retest comparisons between alignment (in MNI and in MSM-All spaces) and denoising strategies (A, B and C refer to *Finn*, *Ciric7* and *SiegelB* models, respectively). **a.** Identification success rate and other statistics. **b.** Test-retest of the pairwise similarities (based on Pearson's correlation) between all subjects **c.** Test-retest reliability of behavioral utility, quantified as the pattern of correlations between each edge and a behavioral score of interest.

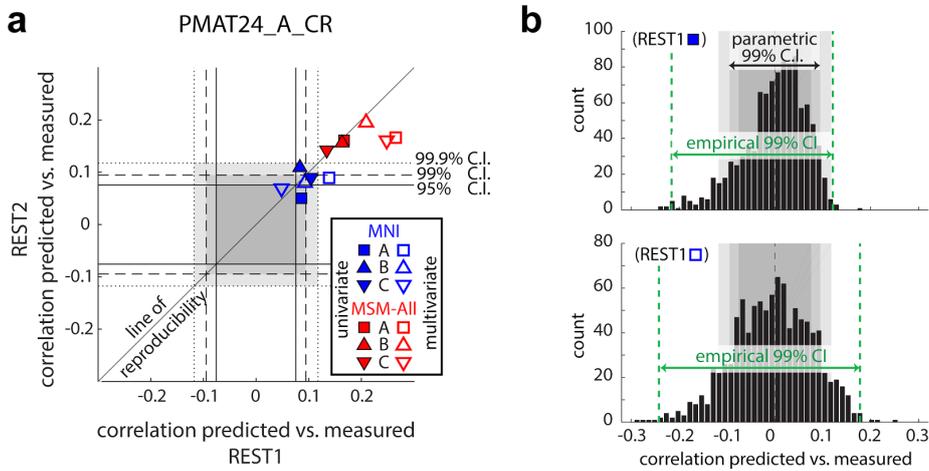


Figure 4.7: Prediction results for de-confounded IQ (PMAT24\_A\_CR). **a.** All predictions are assessed using the correlation between the predicted and the actual scores. The prediction score in REST2 is plotted against the prediction score in REST1, to assess test-retest reliability of the results. Parametric confidence intervals (95%, 99% and 99.9%) for the null hypothesis are shown as shades of gray. **b.** The distribution of prediction scores under the null hypothesis is shown (black histogram). Note that the empirical 99% confidence interval is wider than the parametric CI used in **a**, and features a heavy tail on the left side for the univariate model.

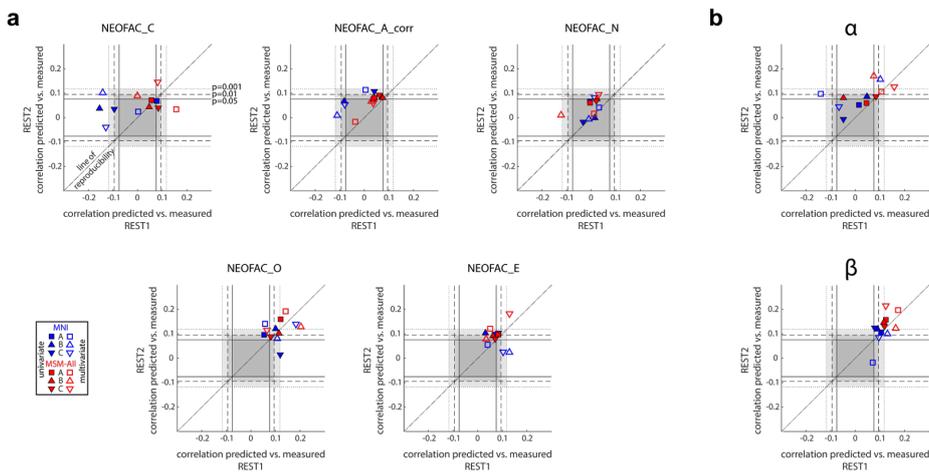


Figure 4.8: Prediction results for the Big Five and superordinate personality factors. **a.** Prediction results for each of the Big Five. **b.** Prediction results for superordinate factors  $\alpha$  and  $\beta$ .

Table 4.3: Pearson's correlation coefficient between predicted and actual scores for every experiment.

	IQ		O		C		E		A		N		$\alpha$		$\beta$		
	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	
<b>MNI</b>																	
Finn univariate	0.086	0.050	0.053	0.096	0.078	0.068	0.097	0.035	0.075	0.065	0.015	0.052	0.105	0.106			
Finn multivariate	0.139	0.089	0.057	0.140	0.002	0.024	0.040	0.055	0.005	0.114	0.032	0.040	-0.140	0.097	0.071	-0.018	
Citic7 univariate	0.083	0.109	0.100	0.120	-0.154	0.037	0.031	0.103	-0.082	0.067	0.015	-0.003	0.047	0.086	0.090	0.118	
Citic7 multivariate	0.093	0.080	0.107	0.080	-0.141	0.102	0.129	0.024	-0.111	0.009	-0.009	-0.008	0.101	0.156	0.130	0.099	
SiegeIB univariate	0.105	0.090	0.119	0.015	-0.094	0.035	0.084	0.103	0.040	0.108	-0.033	-0.019	-0.048	-0.007	0.078	0.124	
SiegeIB multivariate	0.048	0.069	0.182	0.139	-0.128	-0.039	0.103	0.026	-0.079	0.054	0.015	0.081	-0.066	0.044	0.095	0.085	
<b>MSM-All</b>																	
Finn univariate	0.167	0.161	0.120	0.159	0.057	0.071	0.083	0.097	0.041	0.081	-0.004	0.059	0.045	0.059	0.124	0.157	
Finn multivariate	0.265	0.166	0.141	0.192	0.157	0.034	0.051	0.121	-0.038	-0.017	0.011	0.015	0.106	0.106	0.173	0.197	
Citic7 univariate	0.164	0.156	0.114	0.101	0.047	0.043	0.062	0.091	0.074	0.081	0.023	0.075	-0.048	0.081	0.115	0.146	
Citic7 multivariate	0.209	0.195	0.203	0.128	-0.001	0.089	0.034	0.079	0.028	0.066	-0.122	0.009	0.075	0.170	0.164	0.122	
SiegeIB univariate	0.134	0.142	0.081	0.088	0.084	0.040	0.071	0.080	0.061	0.093	0.021	0.068	0.083	0.086	0.118	0.130	
SiegeIB multivariate	0.248	0.161	0.065	0.114	0.082	0.145	0.129	0.182	0.036	0.059	0.031	0.093	0.158	0.126	0.124	0.215	
average univariate	0.091	0.083	0.091	0.077	-0.057	0.047	0.061	0.101	-0.002	0.083	-0.008	0.014	0.005	0.044	0.091	0.116	
average multivariate	0.093	0.079	0.116	0.120	-0.089	0.029	0.091	0.035	-0.062	0.059	0.013	0.038	-0.035	0.099	0.099	0.055	
average univariate	0.155	0.153	0.105	0.116	0.063	0.051	0.072	0.089	0.059	0.085	0.013	0.067	0.027	0.075	0.119	0.144	
average multivariate	0.241	0.174	0.137	0.145	0.080	0.090	0.071	0.128	0.009	0.036	-0.027	0.039	0.113	0.134	0.154	0.178	
Overall average	0.147	0.123	0.113	0.115	-0.001	0.054	0.074	0.089	0.001	0.066	-0.002	0.040	0.027	0.088	0.116	0.124	
	0.134		0.113		0.027		0.081		0.033		0.019		0.058		0.120		

# Chapter 5

## Conclusion

We can think of our brains as computational machines that receive inputs from the surrounding environment and produce behaviour as an output. In the past decades, neuroscience has tried to quantitatively analyse neural activity to understand how the information processing taking place in our brains translates into physiological and pathological behaviour. Despite the daunting complexity of this challenge, great progress has been made in this field, e.g. in mapping functions to brain regions. Yet, a complete understanding of brain disorders is still lacking.

Every person is unique because every individual is the result of a unique point of view on the world. However, to be able to diagnose a neurodegenerative disease, it is necessary to identify anomalies that go beyond inter-subject variability and in fMRI data analysis this task is hindered by the presence of multiple sources of noise. The overarching goal of this work has been that of devising novel methods to derive neuromarkers built upon brain functional networks, taking advantage of machine learning and computational intelligence approaches. The need for these advanced methods is justified by the additional difficulties posed by resting state data: while in task fMRI a big portion of noise can be removed by averaging multiple acquisitions, in absence of a defined task it is crucial to isolate the signal of interest from background noise. Aggregating the information coming from multiple voxels, clustering techniques increase the signal-to-noise ratio and hence facilitate the task of extracting neurophysiological information. On the other hand, ever more studies on brain functional organisation support the theory of a modular community structure, and clustering allows to identify functional modules of co-varying voxels. We have also discussed how consensus clustering techniques can be used in population studies to extract meaningful features able to increase discriminability but also interpretability of the results in a classification task. Additionally, a new approach

was presented that uses stochastic rank aggregation methods to identify brain regions that exhibit a coherent behaviour in subjects of the same diagnostic group, as opposed to traditional methods based on comparative statistics that look at differences between groups. The proposed framework was applied on a resting state activation map, but potentially it can be extended also to task fMRI, to study specific aspects of behaviour.

A commonly raised objection to diagnostic markers derived from neuroimaging studies is that they only recapitulate existing findings. However, neuromarkers should not be considered just as diagnostic tools, but primarily as a means to study the neurobiological basis of disorders. Nevertheless, neuromarkers can support complex diagnostic tasks as patient subtyping, early detection, risk assessment and predicting conversion to full-scale disorders (Woo et al., 2017).

With more and more quality data being available, nowadays it is possible to investigate the variability of the functional organisation of the brain at the individual level. In this work, we discussed the possibility of predicting behavioural traits like fluid intelligence and personality factors from functional connectivity networks, and, to the best of my knowledge, this was the first time that this problem was investigated in a systematic way in a sample of this extension. Despite some negative findings, results showed that a measurable effect is present, especially in the case of intelligence. Indeed, in the IQ experiments, most of the results were significant, independently from the chosen preprocessing pipeline and regression model: this is in accord with the principle of inter-rater reliability (Dubois and Adolphs, 2016), that states that if all the different analyses choices are reasonable, they should all yield a similar result. Results were also stable when replicated on the data of the second session. Furthermore, preliminary experiments have shown that better predictability can be achieved when using factor analysis to derive a general factor of intelligence from multiple cognitive tasks, instead of relying on a single test like Raven's progressive matrices.

The main contribution of this work has been that of reporting the challenges for prospective investigators approaching this field, first among all the need for more data. It has been demonstrated that common sample sizes of less than 50 subjects have extremely low statistical power, and consequently are unable to represent actual effects, leading to a higher variability in subsequent replications (Yarkoni, 2009). Another fundamental aspect to consider is the adoption of predictive frameworks, that provide a means to assess the generalizability of the results on previously unseen data. In the near future, thanks to many data collection initiatives, like UK Biobank, larger datasets will be accessible, that will allow to replicate the presented experiments on independent data and to have more conclusive results, but also to adopt more sophisticated methods. We have seen how the application of multivariate models and of parcellations derived from multi-modal data led to better results. It is probable that the application

of non-linear models could yield to further improvements. For example, deep learning is one of the most trending areas in machine learning research, with outstanding results in several applications, but to achieve its full potential in this field samples sizes in the order of tens of thousands would be needed, while currently available datasets comprise hundreds of samples in the best cases. Future investigations aimed at understanding which is the best model for a given individual trait could give us a minimal representation of a trait in terms of brain function, and this in turn could constitute a basis to derive new psychometric tools that are better supported by neural correlates.

# Bibliography

- Jonathan S. Adelstein, Zarrar Shehzad, Maarten Mennes, Colin G. DeYoung, Xi Nian Zuo, Clare Kelly, Daniel S. Margulies, Aaron Bloomfield, Jeremy R. Gray, F. Xavier Castellanos, and Michael P. Milham. Personality is reflected in the brain's intrinsic functional architecture. *PLoS ONE*, 6(11), 2011. ISSN 19326203. doi: 10.1371/journal.pone.0027633.
- Moji Aghajani, Ilya M. Veer, Marie-José van Tol, André Aleman, Mark A. van Buchem, Dick J. Veltman, Serge A. R. B. Rombouts, and Nic J. van der Wee. Neuroticism and extraversion are associated with amygdala resting-state functional connectivity. *Cognitive, Affective, & Behavioral Neuroscience*, 14(2):836–848, 2014. ISSN 1530-7026. doi: 10.3758/s13415-013-0224-0. URL <http://link.springer.com/10.3758/s13415-013-0224-0>.
- Federica Agosta, Elisa Canu, Paola Valsasina, Nilo Riva, Alessandro Prelle, Giancarlo Comi, and Massimo Filippi. Divergent brain network connectivity in amyotrophic lateral sclerosis. *Neurobiology of aging*, 34(2):419–27, feb 2013. ISSN 1558-1497. doi: 10.1016/j.neurobiolaging.2012.04.015.
- Marianna Amboni, Alessandro Tessitore, Fabrizio Esposito, Gabriella Santangelo, Marina Picillo, Carmine Vitale, Alfonso Giordano, Roberto Erro, Rosa de Micco, Daniele Corbo, et al. Resting-state functional connectivity associated with mild cognitive impairment in parkinson's disease. *Journal of neurology*, 262(2):425–434, 2015.
- Chris Baeken, Daniele Marinazzo, Peter Van Schuerbeek, Guo Rong Wu, Johan De Mey, Robert Luybaert, and Rudi De Raedt. Left and right amygdala - Medial frontal cortical functional connectivity is differentially modulated by harm avoidance. *PLoS ONE*, 9(4), 2014. ISSN 19326203. doi: 10.1371/journal.pone.0095740.
- Aron K. Barbey. Network Neuroscience Theory of Human Intelligence. *Trends in Cognitive Sciences*, 0(0), nov 2017. ISSN 13646613. doi: 10.1016/j.

- tics.2017.10.001. URL <http://linkinghub.elsevier.com/retrieve/pii/S1364661317302218>.
- Roger E. Beaty, Mathias Benedek, Robin W. Wilkins, Emanuel Jauk, Andreas Fink, Paul J. Silvia, Donald A. Hodges, Karl Koschutnig, and Aljoscha C. Neubauer. Creativity and the default network: A functional connectivity analysis of the creative brain at rest. *Neuropsychologia*, 64:92–98, 2014. ISSN 18733514. doi: 10.1016/j.neuropsychologia.2014.09.019.
- Roger E. Beaty, Scott Barry Kaufman, Mathias Benedek, Rex E. Jung, Yoed N. Kenett, Emanuel Jauk, Aljoscha C. Neubauer, and Paul J. Silvia. Personality and complex brain networks: The role of openness to experience in default network efficiency. *Human Brain Mapping*, 37(2):773–779, 2016. ISSN 10970193. doi: 10.1002/hbm.23065.
- Christian F Beckmann, Marilena DeLuca, Joseph T Devlin, and Stephen M Smith. Investigations into resting-state connectivity using independent component analysis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 360(1457):1001–13, may 2005. ISSN 0962-8436. doi: 10.1098/rstb.2005.1634. URL <http://www.ncbi.nlm.nih.gov/pubmed/16087444><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1854918>.
- Yashar Behzadi, Khaled Restom, Joy Liau, and Thomas T. Liu. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, 37(1):90–101, 2007. ISSN 10538119. doi: 10.1016/j.neuroimage.2007.04.042.
- Joe Beitchman. *Biology of Personality and Individual Differences*, volume 18. Canadian Academy of Child and Adolescent Psychiatry, 2009.
- Alberto Bertoni and Giorgio Valentini. Ensembles based on random projections to improve the accuracy of clustering algorithms. In *Neural nets*, pages 31–37. Springer, 2006.
- Bharat Biswal, F Zerrin Yetkin, Victor M Haughton, and James S Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4):537–541, 1995.
- Astrid Bjørnebekk, Anders M. Fjell, Kristine B. Walhovd, Håkon Grydeland, Svann Torgersen, and Lars T. Westlye. Neuronal correlates of the five factor model (FFM) of human personality: Multimodal imaging in a large healthy sample. *NeuroImage*, 65:194–208, 2013. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.10.009.
- Alfredo Brancucci. Neural correlates of cognitive ability, 2012. ISSN 03604012.

- Gregory C. Burgess, Sridhar Kandala, Dan Nolan, Timothy O. Laumann, Jonathan D. Power, Babatunde Adeyemo, Michael P. Harms, Steven E. Petersen, and Deanna M. Barch. Evaluation of Denoising Strategies to Address Motion-Related Artifacts in Resting-State Functional Magnetic Resonance Imaging Data from the Human Connectome Project. *Brain Connectivity*, 6(9):669–680, nov 2016. ISSN 2158-0014. doi: 10.1089/brain.2016.0435. URL <http://online.liebertpub.com/doi/10.1089/brain.2016.0435>.
- Katherine S. Button, John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson, and Marcus R. Munafò. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5):365–376, 2013. ISSN 1471-003X. doi: 10.1038/nrn3475. URL <http://www.nature.com/doi/10.1038/nrn3475>.
- César Caballero-Gaudes and Richard C. Reynolds. Methods for cleaning the BOLD fMRI signal. *NeuroImage*, 154:128–149, 2017. ISSN 10959572. doi: 10.1016/j.neuroimage.2016.12.018.
- Joshua Carp. Optimizing the order of operations for movement scrubbing: Comment on power et al. *Neuroimage*, 76:436–438, 2013.
- Rastko Ciric, Daniel H. Wolf, Jonathan D. Power, David R. Roalf, Graham L. Baum, Kosha Ruparel, Russell T. Shinohara, Mark A. Elliott, Simon B. Eickhoff, Christos Davatzikos, Ruben C. Gur, Raquel E. Gur, Danielle S. Bassett, and Theodore D. Satterthwaite. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*, 154:174–187, 2017. ISSN 10959572. doi: 10.1016/j.neuroimage.2017.03.020.
- M. W. Cole, T. Yarkoni, G. Repovs, A. Anticevic, and T. S. Braver. Global Connectivity of Prefrontal Cortex Predicts Cognitive Control and Intelligence. *Journal of Neuroscience*, 32(26):8988–8999, 2012. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.0536-12.2012. URL <http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.0536-12.2012>.
- A Convit, J. De Asis, M. J. De Leon, C Y Tarshish, S De Santi, and H Rusinek. Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer’s disease. *Neurobiology of Aging*, 21(1):19–26, 2000. ISSN 01974580. doi: 10.1016/S0197-4580(99)00107-4. URL <http://www.ncbi.nlm.nih.gov/pubmed/10794844>.
- J S Damoiseaux, S A R B Rombouts, F Barkhof, P Scheltens, C J Stam, S M Smith, and C F Beckmann. Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences of the United States of America*, 103(37):13848–53, sep 2006. ISSN 0027-8424. doi: 10.1073/pnas.

0601417103. URL <http://www.ncbi.nlm.nih.gov/pubmed/16945915><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1564249>.
- Ian J. Deary, Lars Penke, and Wendy Johnson. The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, 11(MARch), 2010. ISSN 1471-003X. doi: 10.1038/nrn2793. URL <http://www.nature.com/doifinder/10.1038/nrn2793>.
- Nadja Deris, Christian Montag, Martin Reuter, Bernd Weber, and Sebastian Markett. Functional connectivity in the resting brain as biological correlate of the Affective Neuroscience Personality Scales. *NeuroImage*, 147:423–431, 2017. ISSN 10959572. doi: 10.1016/j.neuroimage.2016.11.063.
- Colin G. DeYoung. Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology*, 91(6):1138–1151, 2006. ISSN 1939-1315. doi: 10.1037/0022-3514.91.6.1138. URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.91.6.1138>.
- Julien Dubois and Ralph Adolphs. Building a Science of Individual Differences from fMRI, 2016. ISSN 1879307X.
- Anders Eklund, Thomas E. Nichols, and Hans Knutsson. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, 2016. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1602413113.
- Fabrizio Esposito, Tommaso Scarabino, Aapo Hyvarinen, Johan Himberg, Elia Formisano, Silvia Comani, Gioacchino Tedeschi, Rainer Goebel, Erich Seifritz, and Francesco Di Salle. Independent component analysis of fMRI group studies by self-organizing clustering. *NeuroImage*, 25(1):193–205, mar 2005. ISSN 10538119. doi: 10.1016/j.neuroimage.2004.10.042. URL <http://www.sciencedirect.com/science/article/pii/S1053811904006603>.
- Fabrizio Esposito, Giuseppe Pignataro, Gianfranco Di Renzo, Alessandra Spinali, Antonella Paccone, Gioacchino Tedeschi, and Lucio Annunziato. Alcohol increases spontaneous bold signal fluctuations in the visual network. *Neuroimage*, 53(2):534–543, 2010.
- Fabrizio Esposito, Alessandro Tessitore, Alfonso Giordano, Rosita De Micco, Antonella Paccone, Rentà Conforti, Giuseppe Pignataro, Lucio Annunziato, and Gioacchino Tedeschi. Rhythm-specific modulation of the sensorimotor network in drug-naive patients with parkinson’s disease by levodopa. *Brain*, 136(3):710–725, 2013.
- Martin Ester, Hans P Kriegel, Jorg Sander, and Xiaowei Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.

- Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pages 226–231, dec 1996. ISSN 09758887. doi: 10.1.1.71.1980. URL <https://www.osti.gov/scitech/biblio/421283https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf>.
- Joset A Etzel, Jeffrey M Zacks, and Todd S Braver. Searchlight analysis: promise, pitfalls, and potential. *Neuroimage*, 78:261–269, 2013.
- Xiaoli Zhang Fern and Carla E Brodley. Random projection for high dimensional data clustering: A cluster ensemble approach. In *ICML*, volume 3, pages 186–193, 2003.
- Emily S Finn, Xilin Shen, Dustin Scheinost, Monica D Rosenberg, Jessica Huang, Marvin M Chun, Xenophon Papademetris, and R Todd Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(11):1664–1671, 2015. ISSN 1097-6256. doi: 10.1038/nn.4135. URL <http://www.nature.com/doifinder/10.1038/nn.4135>.
- Elia Formisano, Fabrizio Esposito, Nikolaus Kriegeskorte, Gioacchino Tedeschi, Francesco Di Salle, and Rainer Goebel. Spatial independent component analysis of functional magnetic resonance imaging time-series: characterization of the cortical components. *Neurocomputing*, 49(1):241–254, 2002.
- Elia Formisano, Giancarlo Valente, Fabrizio Esposito, Federico de Martino, and Rainer Goebel. Using ICA for the analysis of fMRI data. *Simultaneous EEG and fMRI: Recording, Analysis, and Application*, page 135, 2010.
- Karl J Friston, Steven Williams, Robert Howard, Richard SJ Frackowiak, and Robert Turner. Movement-related effects in fmri time-series. *Magnetic resonance in medicine*, 35(3):346–355, 1996.
- Qing Gao, Qiang Xu, Xujun Duan, Wei Liao, Jurong Ding, Zhiqiang Zhang, Yuan Li, Guangming Lu, and Huaifu Chen. Extraversion and neuroticism relate to topological properties of resting-state brain networks. *Frontiers in Human Neuroscience*, 7, 2013. ISSN 1662-5161. doi: 10.3389/fnhum.2013.00257. URL <http://journal.frontiersin.org/article/10.3389/fnhum.2013.00257/abstract>.
- L. Geerligs, M. Rubinov, Cam-CAN, and R. N. Henson. State and Trait Components of Functional Connectivity: Individual Differences Vary with Mental State. *Journal of Neuroscience*, 35(41):13949–13961, 2015a. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.1324-15.2015. URL <http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.1324-15.2015>.

- Linda Geerligs, Remco J. Renken, Emi Saliasi, Natasha M. Maurits, and Monicque M. Lorist. A Brain-Wide Study of Age-Related Changes in Functional Connectivity. *Cerebral Cortex*, 25(7):1987–1999, 2015b. ISSN 14602199. doi: 10.1093/cercor/bhu012.
- Linda Geerligs, Kamen A Tsvetanov, and Richard N Henson. Challenges in measuring individual differences in functional connectivity using fmri: The case of healthy aging. *Human Brain Mapping*, 2017.
- Reza Ghaemi, Md Nasir Sulaiman, Hamidah Ibrahim, Norwati Mustapha, et al. A survey: clustering ensembles techniques. *World Academy of Science, Engineering and Technology*, 50:636–645, 2009.
- Matthew F. Glasser, Stamatios N. Sotiropoulos, J. Anthony Wilson, Timothy S. Coalson, Bruce Fischl, Jesper L. Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R. Polimeni, David C. Van Essen, and Mark Jenkinson. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80:105–124, 2013. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.04.127.
- Matthew F. Glasser, Timothy S. Coalson, Emma C. Robinson, Carl D. Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F. Beckmann, Mark Jenkinson, Stephen M. Smith, and David C. Van Essen. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615): 171–178, 2016. ISSN 0028-0836. doi: 10.1038/nature18933. URL <http://www.nature.com/doifinder/10.1038/nature18933>.
- Evan M. Gordon, Timothy O. Laumann, Babatunde Adeyemo, Jeremy F. Huckins, William M. Kelley, and Steven E. Petersen. Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cerebral Cortex*, 26(1):288–303, jan 2016. ISSN 1047-3211. doi: 10.1093/cercor/bhu239. URL <http://www.cercor.oxfordjournals.org/lookup/doi/10.1093/cercor/bhu239>.
- Evian Gordon. Integrating genomics and neuromarkers for the era of brain-related personalized medicine. *Personalized Medicine*, 4(2):201–215, may 2007. ISSN 1741-0541. doi: 10.2217/17410541.4.2.201. URL <http://www.futuremedicine.com/doi/10.2217/17410541.4.2.201>.
- Martin Gorges, Hans-Peter Müller, Dorothée Lulé, Albert C Ludolph, Elmar H Pinkhardt, and Jan Kassubek. Functional connectivity within the default mode network is associated with saccadic accuracy in Parkinson’s disease: a resting-state FMRI and videooculographic study. *Brain connectivity*, 3(3): 265–72, jan 2013. ISSN 2158-0022. doi: 10.1089/brain.2013.0146.

- Krzysztof Gorgolewski, Christopher D. Burns, Cindee Madison, Dav Clark, Yaroslav O. Halchenko, Michael L. Waskom, and Satrajit S. Ghosh. Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Frontiers in Neuroinformatics*, 5, 2011. ISSN 1662-5196. doi: 10.3389/fninf.2011.00013. URL <http://journal.frontiersin.org/article/10.3389/fninf.2011.00013/abstract>.
- C Goutte. On clustering fMRI time series. *Neuroimage*, 9(3), 1999. ISSN 10538119. URL <http://www.forskningsdatabasen.dk/en/catalog/2282408503>.
- Cyril Goutte, Lars Kai Hansen, Matthew G. Liptrot, and Egill Rostrup. Feature-space clustering for fMRI meta-analysis. *Human Brain Mapping*, 13(3):165–183, jul 2001. ISSN 1065-9471. doi: 10.1002/hbm.1031. URL <http://doi.wiley.com/10.1002/hbm.1031>.
- Michael D Greicius, Gaurav Srivastava, Allan L Reiss, and Vinod Menon. Default-mode network activity distinguishes Alzheimer’s disease from healthy aging: evidence from functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, 101(13):4637–42, 2004. ISSN 0027-8424. doi: 10.1073/pnas.0308627101.
- Michael D. Greicius, Benjamin H. Flores, Vinod Menon, Gary H. Glover, Hugh B. Solvason, Heather Kenna, Allan L. Reiss, and Alan F. Schatzberg. Resting-State Functional Connectivity in Major Depression: Abnormally Increased Contributions from Subgenual Cingulate Cortex and Thalamus. *Biological Psychiatry*, 62(5):429–437, 2007. ISSN 00063223. doi: 10.1016/j.biopsych.2006.09.020.
- Peter Hall and Michael G. Schimek. Moderate-Deviation-Based Inference for Random Degeneration in Paired Rank Lists. *Journal of the American Statistical Association*, 107(498):661–672, jun 2012. ISSN 0162-1459. doi: 10.1080/01621459.2012.682539. URL <http://www.tandfonline.com/doi/abs/10.1080/01621459.2012.682539>.
- Jürgen Hänggi, Johannes Streffer, Lutz Jäncke, and Christoph Hock. Volumes of lateral temporal and parietal structures distinguish between healthy aging, mild cognitive impairment, and Alzheimer’s disease. *Journal of Alzheimer’s disease : JAD*, 26(4):719–34, 2011. ISSN 1875-8908. doi: 10.3233/JAD-2011-101260. URL <http://www.ncbi.nlm.nih.gov/pubmed/21709375>.
- Kurt Hornik. Cluster ensembles. In *Classification The Ubiquitous Challenge*, pages 65–72. Springer, 2005.

- Scott A. Huettel, Jeffrey D. Singerman, and Gregory McCarthy. The Effects of Aging upon the Hemodynamic Response Measured by Functional MRI. *NeuroImage*, 13(1):161–175, jan 2001. ISSN 10538119. doi: 10.1006/nimg.2000.0675. URL <http://www.sciencedirect.com/science/article/pii/S105381190090675X><http://linkinghub.elsevier.com/retrieve/pii/S105381190090675X>.
- Aapo Hyvarinen. Fast and Robust Fixed-Point Algorithm for Independent Component Analysis. *IEEE Trans. Neur. Net.*, 10(3):626–634, 1999.
- Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4):411–430, 2000.
- Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent component analysis*, volume 46. John Wiley & Sons, 2001.
- David C. Jangraw, Javier Gonzalez-Castillo, Daniel A. Handwerker, Merge Ghane, Monica D. Rosenberg, Pujja Panwar, and Peter A. Bandettini. A functional connectivity-based neuromarker of sustained attention generalizes to predict recall in a reading task. *NeuroImage*, 166:99–109, feb 2018. ISSN 1053-8119. doi: 10.1016/J.NEUROIMAGE.2017.10.019. URL <http://www.sciencedirect.com/science/article/pii/S105381191730839X?via%3Dihub>.
- Bingqing Jiao, Delong Zhang, Aiyang Liang, Bishan Liang, Zengjian Wang, Junchao Li, Yuxuan Cai, Mengxia Gao, Zhenni Gao, Song Chang, Ruiwang Huang, and Ming Liu. Association between resting-state brain network topological organization and creative ability: Evidence from a multiple linear regression model. *Biological Psychology*, 129:165–177, 2017. ISSN 18736246. doi: 10.1016/j.biopsycho.2017.09.003.
- Mitzy Kennis, Arthur R. Rademaker, and Elbert Geuze. Neural correlates of personality: An integrative review, 2013. ISSN 01497634.
- Rogier A. Kievit, Simon W. Davis, John Griffiths, Marta M. Correia, CamCAN, and Richard N. Henson. A watershed model of individual differences in fluid intelligence. *Neuropsychologia*, 91:186–198, 2016. ISSN 00283932. doi: 10.1016/j.neuropsychologia.2016.08.008.
- Nikolaus Kriegeskorte, Rainer Goebel, and Peter Bandettini. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 103(10):3863–3868, 2006.
- Ludmila I Kuncheva, Stefan Todorov Hadjitodorov, and Ludmila P Todorova. Experimental comparison of cluster ensemble methods. In *2006 9th International Conference on Information Fusion*, pages 1–7. IEEE, 2006.

- Xu Lei, Zhiying Zhao, and Hong Chen. Extraversion is encoded by scale-free dynamics of default mode network. *NeuroImage*, 74:52–57, 2013. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.02.020.
- Christophe Leys, Olivier Klein, Yves Dominicy, and Christophe Ley. Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. *Journal of Experimental Social Psychology*, 74:150–156, 2018. ISSN 10960465. doi: 10.1016/j.jesp.2017.09.011.
- Yonghui Li, Yong Liu, Jun Li, Wen Qin, Kuncheng Li, Chunshui Yu, and Tianzi Jiang. Brain anatomical network and intelligence. *PLoS Computational Biology*, 5(5):e1000395, may 2009. ISSN 1553734X. doi: 10.1371/journal.pcbi.1000395. URL <http://dx.plos.org/10.1371/journal.pcbi.1000395>.
- Shili Lin and Jie Ding. Integration of Ranked Lists via Cross Entropy Monte Carlo with Applications to mRNA and microRNA Studies. *Biometrics*, 65(1):9–18, mar 2009. ISSN 0006341X. doi: 10.1111/j.1541-0420.2008.01044.x. URL <http://www.ncbi.nlm.nih.gov/pubmed/18479487><http://doi.wiley.com/10.1111/j.1541-0420.2008.01044.x>.
- Thomas T. Liu. Noise contributions to the fMRI signal: An overview. *NeuroImage*, 143:141–151, 2016. ISSN 10959572. doi: 10.1016/j.neuroimage.2016.09.008.
- Robert R. McCrae and Paul T. Costa. Clinical assessment can benefit from recent advances in personality psychology. *American Psychologist*, 41(9):1001–1003, 1986. ISSN 1935-990X. doi: 10.1037/0003-066X.41.9.1001. URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.41.9.1001>.
- Robert R. McCrae and Paul T. Costa. A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, 36(3):587–596, 2004. ISSN 01918869. doi: 10.1016/S0191-8869(03)00118-1.
- Bahram Mohammadi, Katja Kollewe, Amir Samii, Klaus Krampfl, Reinhard Dengler, and Thomas F Münte. Changes of resting state brain networks in amyotrophic lateral sclerosis. *Experimental neurology*, 217(1):147–53, may 2009. ISSN 1090-2430. doi: 10.1016/j.expneurol.2009.01.025.
- Stefano Monti, Pablo Tamayo, Jill Mesirov, and Todd Golub. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Machine Learning*, 52(1-2):91–118, jul 2003. ISSN 0885-6125, 1573-0565. doi: 10.1023/A:1023949509487.
- Kevin Murphy and Michael D. Fox. Towards a consensus regarding global signal regression for resting state functional connectivity MRI. *NeuroImage*, 154:169–173, 2017. ISSN 10959572. doi: 10.1016/j.neuroimage.2016.11.052.

- Kevin Murphy, Rasmus M. Birn, and Peter A. Bandettini. Resting-state fMRI confounds and cleanup. *NeuroImage*, 80:349–359, 2013. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.04.001.
- John Muschelli, Mary Beth Nebel, Brian S. Caffo, Anita D. Barber, James J. Pekar, and Stewart H. Mostofsky. Reduction of motion-related artifacts in resting state fMRI using aCompCor. *NeuroImage*, 96:22–35, 2014. ISSN 10959572. doi: 10.1016/j.neuroimage.2014.03.028.
- Aljoseha C. Neubauer and Andreas Fink. Intelligence and neural efficiency, 2009. ISSN 01497634.
- Nam Nguyen and Rich Caruana. Consensus clusterings. *Proceedings - IEEE International Conference on Data Mining, ICDM*, pages 607–612, 2007.
- Stephanie Noble, Marisa N. Spann, Fuyuze Tokoglu, Xilin Shen, R. Todd Constable, and Dustin Scheinost. Influences on the Test-Retest Reliability of Functional Connectivity MRI and its Relationship with Behavioral Utility. *Cerebral Cortex*, pages 1–15, 2017. ISSN 1047-3211. doi: 10.1093/cercor/bhx230. URL <http://academic.oup.com/cercor/article/doi/10.1093/cercor/bhx230/4139668/Influences-on-the-TestRetest-Reliability-of>.
- Seiji Ogawa, Tso-Ming Lee, Alan R Kay, and David W Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872, 1990.
- Kazufumi Omura, R Todd Constable, and Turhan Canli. Amygdala gray matter concentration is associated with extraversion and neuroticism. *Neuroreport*, 16(17):1905–1908, 2005. ISSN 0959-4965. doi: 10.1097/01.wnr.0000186596.64458.76.
- Yajing Pang, Qian Cui, Yifeng Wang, Yuyan Chen, Xiaona Wang, Shaoqiang Han, Zhiqiang Zhang, Guangming Lu, and Huafu Chen. Extraversion and neuroticism related to the resting-state effective connectivity of amygdala. *Scientific Reports*, 6(1):35484, 2016. ISSN 2045-2322. doi: 10.1038/srep35484. URL <http://www.nature.com/articles/srep35484>.
- Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.
- Jonathan D. Power, Kelly A. Barnes, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. Spurious but systematic correlations in functional

- connectivity MRI networks arise from subject motion. *NeuroImage*, 59(3): 2142–2154, 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2011.10.018.
- Jonathan D. Power, Anish Mitra, Timothy O. Laumann, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, 84: 320–341, 2014. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.08.048.
- Raimon H R Pruim, Maarten Mennes, Daan van Rooij, Alberto Llera, Jan K. Buitelaar, and Christian F. Beckmann. ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage*, 112, 2015. ISSN 10959572. doi: 10.1016/j.neuroimage.2015.02.064.
- Emma C. Robinson, Saad Jbabdi, Matthew F. Glasser, Jesper Andersson, Gregory C. Burgess, Michael P. Harms, Stephen M. Smith, David C. Van Essen, and Mark Jenkinson. MSM: A new flexible framework for multimodal surface matching. *NeuroImage*, 100:414–426, 2014. ISSN 10959572. doi: 10.1016/j.neuroimage.2014.05.069.
- M. D. Rosenberg, S. Zhang, W.-T. Hsu, D. Scheinost, E. S. Finn, X. Shen, R. T. Constable, C.-S. R. Li, and M. M. Chun. Methylphenidate Modulates Functional Network Connectivity to Enhance Attention. *Journal of Neuroscience*, 36(37):9547–9557, sep 2016. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.1746-16.2016. URL <http://www.ncbi.nlm.nih.gov/pubmed/27629707><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5039242><http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.1746-16.2016>.
- Gholamreza Salimi-Khorshidi, Gwenaëlle Douaud, Christian F. Beckmann, Matthew F. Glasser, Ludovica Griffanti, and Stephen M. Smith. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90:449–468, 2014. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.11.046.
- Theodore D. Satterthwaite, Mark A. Elliott, Raphael T. Gerraty, Kosha Ruparel, James Loughhead, Monica E. Calkins, Simon B. Eickhoff, Hakon Hakonarson, Ruben C. Gur, Raquel E. Gur, and Daniel H. Wolf. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, 64(1):240–256, 2013. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.08.052.
- Michael G. Schimek, Alena Myšičková, and Eva Budinská. An Inference and Integration Approach for the Consolidation of Ranked Lists. *Communications in Statistics - Simulation and Computation*, 41(7):1152–1166,

- aug 2012. ISSN 0361-0918. doi: 10.1080/03610918.2012.625843. URL <http://www.tandfonline.com/doi/abs/10.1080/03610918.2012.625843>.
- Michael G Schimek, Eva Budinská, Karl G Kugler, Vendula Švendová, Jie Ding, and Shili Lin. TopKLists: a comprehensive R package for statistical inference, stochastic aggregation, and visualization of multiple omics ranked lists. *Statistical applications in genetics and molecular biology*, 14(3):311–6, jun 2015. ISSN 1544-6115. doi: 10.1515/sagmb-2014-0093. URL <http://www.ncbi.nlm.nih.gov/pubmed/25968440>.
- Felix D. Schönbrodt and Marco Perugini. At what sample size do correlations stabilize? *Journal of Research in Personality*, 47(5):609–612, 2013. ISSN 00926566. doi: 10.1016/j.jrp.2013.05.009.
- X. Shen, F. Tokoglu, X. Papademetris, and R. T. Constable. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage*, 82:403–415, 2013. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.05.081.
- Lei K. Sheu, John P. Ryan, and Peter J. Gianaros. Resting state functional connectivity within the cingulate cortex jointly predicts agreeableness and stressor-evoked cardiovascular reactivity. *NeuroImage*, 55(1):363–370, 2011. ISSN 10538119. doi: 10.1016/j.neuroimage.2010.11.064.
- Joshua S Siegel, Jonathan D Power, Joseph W Dubis, Alecia C Vogel, Jessica A Church, Bradley L Schlaggar, and Steven E Petersen. Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. *Human brain mapping*, 35(5):1981–1996, 2014.
- Joshua S. Siegel, Anish Mitra, Timothy O. Laumann, Benjamin A. Seitzman, Marcus Raichle, Maurizio Corbetta, and Abraham Z. Snyder. Data Quality Influences Observed Links Between Functional Connectivity and Behavior. *Cerebral Cortex*, 2016. ISSN 1047-3211. doi: 10.1093/cercor/bhw253. URL <http://cercor.oxfordjournals.org/cgi/doi/10.1093/cercor/bhw253>.
- Stephen M. Smith, Diego Vidaurre, Christian F. Beckmann, Matthew F. Glasser, Mark Jenkinson, Karla L. Miller, Thomas E. Nichols, Emma C. Robinson, Gholamreza Salimi-Khorshidi, Mark W. Woolrich, Deanna M. Barch, Kamil UÇşurbil, and David C. Van Essen. Functional connectomics from resting-state fMRI, 2013. ISSN 13646613.
- Ming Song, Yuan Zhou, Jun Li, Yong Liu, Lixia Tian, Chunshui Yu, and Tianzi Jiang. Brain spontaneous functional connectivity and intelligence. *NeuroImage*, 41(3):1168–1176, 2008. ISSN 10538119. doi: 10.1016/j.neuroimage.2008.02.036.

- Olaf Sporns. The human connectome: Origins and challenges. *NeuroImage*, 80: 53–61, 2013. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.03.023.
- Alexander Strehl and Joydeep Ghosh. Cluster Ensembles - a Knowledge Reuse Framework for Combining Multiple Partitions. *J. Mach. Learn. Res.*, 3:583–617, mar 2003. ISSN 1532-4435. doi: 10.1162/153244303321897735.
- Hikaru Takeuchi, Yasuyuki Taki, Hiroshi Hashizume, Yuko Sassa, Tomomi Nagase, Rui Nouchi, and Ryuta Kawashima. The association between resting functional connectivity and creativity. *Cerebral Cortex*, 22(12):2921–2929, 2012. ISSN 10473211. doi: 10.1093/cercor/bhr371.
- Gioacchino Tedeschi and Fabrizio Esposito. *Neuronal networks observed with resting state functional magnetic resonance imaging in clinical populations*. INTECH Open Access Publisher, 2012.
- Gioacchino Tedeschi, Francesca Trojsi, Alessandro Tessitore, Daniele Corbo, Anna Sagnelli, Antonella Paccone, Alessandro D’Ambrosio, Giovanni Piccirillo, Mario Cirillo, Sossio Cirillo, Maria Rosaria Monsurrò, and Fabrizio Esposito. Interaction between aging and neurodegeneration in amyotrophic lateral sclerosis. *Neurobiology of aging*, 33(5):886–98, may 2012. ISSN 1558-1497. doi: 10.1016/j.neurobiolaging.2010.07.011.
- Tatsuhiko Terada, Tomokazu Obi, Jun Miyata, Manabu Kubota, Miho Yoshizumi, Toshiya Murai, Kinya Yamazaki, and Kouichi Mizoguchi. Correlation of frontal atrophy with behavioral changes in amyotrophic lateral sclerosis. *Neurology and Clinical Neuroscience*, 4(3):85–92, may 2016. ISSN 20494173. doi: 10.1111/ncn3.12046. URL <http://doi.wiley.com/10.1111/ncn3.12046>.
- Alessandro Tessitore, Marianna Amboni, Fabrizio Esposito, Antonio Russo, Marina Picillo, Laura Marcuccio, Maria Teresa Pellecchia, Carmine Vitale, Mario Cirillo, Gioacchino Tedeschi, et al. Resting-state brain connectivity in patients with parkinson’s disease and freezing of gait. *Parkinsonism & related disorders*, 18(6):781–787, 2012a.
- Alessandro Tessitore, Fabrizio Esposito, Carmine Vitale, Gabriella Santangelo, Marianna Amboni, Antonio Russo, Daniele Corbo, Giovanni Cirillo, Paolo Barone, and Gioacchino Tedeschi. Default-mode network connectivity in cognitively unimpaired patients with Parkinson disease. *Neurology*, 79(23):2226–32, dec 2012b. ISSN 1526-632X. doi: 10.1212/WNL.0b013e31827689d6.
- Bertrand Thirion, Guillaume Flandin, Philippe Pinel, Alexis Roche, Philippe Ciuciu, and Jean-Baptiste Poline. Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. *Human Brain*

*Mapping*, 27(8):678–693, aug 2006. ISSN 1065-9471. doi: 10.1002/hbm.20210. URL <http://doi.wiley.com/10.1002/hbm.20210>.

Alexander Topchy, Anil K Jain, and William Punch. Clustering ensembles: models of consensus and weak partitions. *IEEE transactions on pattern analysis and machine intelligence*, 27(12):1866–81, dec 2005. ISSN 0162-8828. doi: 10.1109/TPAMI.2005.237.

Francesca Trojsi, Fabrizio Esposito, Manuela de Stefano, Daniela Buonanno, Francesca L Conforti, Daniele Corbo, Giovanni Piccirillo, Mario Cirillo, Maria Rosaria Monsurrò, Patrizia Montella, et al. Functional overlap and divergence between als and bvftd. *Neurobiology of aging*, 36(1):413–423, 2015.

Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

Martijn van den Heuvel, Rene Mandl, and Hilleke Hulshoff Pol. Normalized Cut Group Clustering of Resting-State fMRI Data. *PLOS ONE*, 3(4):e2001, 2008. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0002001. URL <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0002001>.

Martijn P van den Heuvel, Cornelis J Stam, René S Kahn, and Hilleke E Hulshoff Pol. Efficiency of functional brain networks and intellectual performance. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(23):7619–7624, 2009. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.1443-09.2009.

Stéfan Van Der Walt, S. Chris Colbert, and Gaël Varoquaux. The NumPy array: A structure for efficient numerical computation. *Computing in Science and Engineering*, 13(2):22–30, 2011. ISSN 15219615. doi: 10.1109/MCSE.2011.37.

Thilo van Eimeren, Oury Monchi, Benedicte Ballanger, and Antonio P Strafella. Dysfunction of the default mode network in Parkinson disease: a functional magnetic resonance imaging study. *Archives of neurology*, 66(7):877–83, jul 2009. ISSN 1538-3687. doi: 10.1001/archneurol.2009.97.

David C. Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy E.J. Behrens, Essa Yacoub, and Kamil Ugurbil. The WU-Minn Human Connectome Project: An overview. *NeuroImage*, 80:62–79, 2013. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.05.041.

- Paul Wicks, Martin R. Turner, Sharon Abrahams, Alexander Hammers, David J. Brooks, P. Nigel Leigh, and Laura H. Goldstein. Neuronal loss associated with cognitive performance in amyotrophic lateral sclerosis: An (<sup>11</sup>C)-flumazenil PET study. *Amyotrophic Lateral Sclerosis*, 9(1):43–49, jan 2008. ISSN 1748-2968. doi: 10.1080/17482960701737716. URL <http://www.ncbi.nlm.nih.gov/pubmed/18270878><http://www.tandfonline.com/doi/full/10.1080/17482960701737716>.
- Choong-Wan Woo, Luke J Chang, Martin A Lindquist, and Tor D Wager. Building better biomarkers: brain models in translational neuroimaging. *Nature Neuroscience*, 20(3):365–377, feb 2017. ISSN 1097-6256. doi: 10.1038/nn.4478. URL <http://www.nature.com/doi/10.1038/nn.4478>.
- Christopher I. Wright, Danielle Williams, Eric Feczko, Lisa Feldman Barrett, Bradford C. Dickerson, Carl E. Schwartz, and Michelle M. Wedig. Neuroanatomical correlates of extraversion and neuroticism. *Cerebral Cortex*, 16(12):1809–1819, 2006. ISSN 10473211. doi: 10.1093/cercor/bhj118.
- Yan Wu, Ling Li, Binke Yuan, and Xuehong Tian. Individual differences in resting-state functional connectivity predict procrastination. *Personality and Individual Differences*, 95:62–67, 2016. ISSN 01918869. doi: 10.1016/j.paid.2016.02.016.
- Tal Yarkoni. Big Correlations in Little Studies: Inflated fMRI Correlations Reflect Low Statistical Power – Commentary on Vul et al. (2009). *Perspectives on Psychological Science*, 4(3):294–298, 2009. ISSN 17456924. doi: 10.1111/j.1745-6924.2009.01127.x.
- Kwangsun Yoo, Monica D. Rosenberg, Wei-Ting Hsu, Sheng Zhang, Chiang-Shan R. Li, Dustin Scheinost, R. Todd Constable, and Marvin M. Chun. Connectome-based predictive modeling of attention: Comparing different functional connectivity features and prediction methods across datasets. *NeuroImage*, 167:11–22, feb 2018. ISSN 1053-8119. doi: 10.1016/J.NEUROIMAGE.2017.11.010. URL <http://www.sciencedirect.com/science/article/pii/S1053811917309175>.
- Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 67(2):301–320, 2005. ISSN 13697412. doi: 10.1111/j.1467-9868.2005.00503.x.