Evaluation and applications of network modeling methods on a fine scale in the human visual cortex and in mild cognitive impairment

Debra Ann Dawson

Department of Neurology and Neurosurgery

Faculty of Medicine

McGill University, Montreal

January 2021

A thesis submitted to McGill University in partial fulfillment of the requirements of the Doctoral degree in the Integrated Program of Neuroscience.

© Debra Ann Dawson, 2021

Table of Contents

Abstract English		*4*
Abstract French		*6*
Acknowledgements		*8*
Contribution to original knowledge		*10*
Contribution of Authors		*12*
Thesis Introduction and Literature Review		*15*
Introduction to Manuscript 1		*32*
Manuscript 1:	*"Partial-correlation based retinotopically organized resting-state functional connectivity within and between areas of the visual cortex reflects more than cortical distance"*	*34*
Introduction to Manuscript 2 and the Corresponding Data Manuscript		*78*
Manuscript 2:	*"Local Principal Component regression, partial correlation and inverse covariance identify direct connections in an unknown, fine-scale, densely connected functional network"*	*81*
Introduction to Manuscript 3		*152*
Manuscript 3:	*"High positive predictive value in a subset of individuals using a multi-step machine learning classification procedure for identification of progression from mild cognitive impairment to dementia"*	*155*
Thesis Discussion		*233*
Conclusion/Summary		*251*
References		*253*
Appendices		
Appendix A: Supplementary Material from Manuscript 1.		*266*
Appendix B: Details regarding the differences between Part 2 of the thesis and Dawson et al., NeuroImage 2013		*274*

Appendix C: Supplementary Material from Manuscript 2.	*278*
Appendix D: A Manuscript. Human Connectome Project resting state fMRI data organized into 60 fine-scaled retinotopically organized regions in cortical areas V1, V2 and V3. bioRxiv. Epub: June 6, 2018. https://doi.org/10.1101/340174, Submitted to Data In Brief	333
Appendix E: A comment on computer precision.	*344*
Appendix F: ICOV pursued at different regularization parameters.	*345*
Appendix G: Supplementary Material from Manuscript 3.	*348*
Appendix H: Manuscript Specific Acknowledgement.	*369*

Abstract: English

Resting-State functional MRI (RS-fMRI) is often utilized to characterize the functional relationships between brain regions, these relationships being termed functional connectivity (FC). While whole-brain and network-specific FC have been widely explored, there is still much to be discovered, particularly with respect to FC on a fine scale. In the first part of this thesis, the fine-scaled RS-FC of the human visual cortex was explored in detail. The RS-FC was found to be retinotopically organized and partial correlation-based FC was found to reflect more than simply the distance between regions. In fact, the connectivity described by partial correlation in the visual cortex appeared to roughly follow that of the direct anatomical connectivity known to exist based on the thoroughly studied monkey visual cortex.

It has been theorized that RS-FC may be driven by direct anatomical connections. There are, however, a number of computational approaches one can use to describe FC and these do not all result in the same pattern of predicted connectivity. Thus, in the second part of the thesis, 16 network models were tested in terms of their abilities to predict direct connectivity in the densely connected early visual cortex, in the context of no prior knowledge of the system of study. The evaluation of the model predictions was based on previous non-human primate tracer-injection studies. Multivariate models, partial correlation and Inverse Covariance, were successfully able to predict direct connectivity. In a novel approach, a "global signal" of the local network (Local Principal Component Regression) was regressed from network time-courses prior to computing the FC between regions. With this regression, bivariate models, correlation and Kappa, were also highly successful in predicting direct connectivity. Importantly, with a larger sample size, network predictions converged toward the inferred ground truth using common statistical significance thresholding, whereas they did not when using connection density-based thresholding.

Finally, having identified computational methods that appeared promising for extracting direct connectivity from RS-fMRI data with relatively fine-scaled brain regions, we sought to test the usefulness of this approach in a more challenging context: the classification of individuals with mild cognitive impairment (MCI) using a whole-brain, data-driven approach.

MCI is often a precursor state for dementia, thus, if subjects who will progress can be identified while they are only mildly impaired, more timely and targeted dementia treatment development and implementation can be pursued. A novel positive predictive value (PPV)-optimized multi-classifier machine learning pipeline was developed that could classify a subset of the subjects who did not progress to dementia with high PPV and high specificity. Three subtypes of MCI individuals who progressed to dementia and two subtypes of individuals who did not progress were characterized. MRI-based features (gray matter volume, cortical thickness, and correlation-based FC) involving occipital regions, sensorimotor regions, and medial parietal regions were most often implicated in the differences between MCI subtypes.

We conclude that network modeling methods applied to RS-fMRI in the context of fine-scaled brain regions are useful to characterize retinotopically organized connectivity in the human visual cortex, can extract direct connectivity reliably in a densely connected local network, and contribute to the precise classification of a subset of stable MCI subjects.

Abstract: French

L'IRM fonctionnelle à l'état de repos (ER-IRMf) est souvent utilisée pour caractériser les relations fonctionnelles entre les régions cérébrales, ces relations étant appelées connectivité fonctionnelle (CF). La CF du cerveau entier et de réseaux spécifiques a été bien explorée, mais il reste encore beaucoup à découvrir en particulier en ce qui concerne la CF à une échelle fine. Dans la première partie de cette thèse, nous avons exploré en détail l'ER-CF à échelle fine du cortex visuel humain. Nous avons constaté que l'ER-CF est organisé de manière rétinotopique et que le CF basé sur une corrélation partielle reflète plus que simplement la distance entre les régions. En fait, la connectivité décrite par la corrélation partielle dans le cortex visuel semblait suivre à peu près celle de la connectivité anatomique directe connue sur la base du cortex visuel du singe largement étudié.

Il a été théorisé que l'ER-CF est entraîné par des connexions anatomiques directes. Il existe, cependant, nombreux approches informatiques que l'on peut utiliser pour calculer la CF et celles-ci ne décrivent pas toutes les mêmes réseaux de connectivité. Ainsi, dans la deuxième partie de la thèse, nous avons testé 16 modèles de CF en termes de leurs capacités à prédire la connectivité anatomique directe dans le cortex visuel dans le cadre d'une absence de connaissance préalable du système d'étude. Nous avons constaté que les modèles multivariés de corrélation partielle et Covariance Inversée réussissaient dans cette tâche ainsi que la corrélation et Kappa lorsque notre nouvelle approche pour supprimer le signal global a été appliquée (régression de composante principale locale). Nous avons également constaté qu'en utilisant le seuil de signification statistique commun, les prédictions du réseau convergeaient vers la vérité avec une taille d'échantillon plus grande, alors que lors de l'utilisation du seuillage basé sur la densité de connexion, ce n'était pas le cas.

Enfin, après avoir identifié des méthodes de calcul qui paraissaient prometteuses pour extraire la connectivité directe des données ER-IRMf avec des régions cérébrales relativement fines, nous avons cherché à tester l'utilité de cette approche dans un contexte plus difficile: la classification des individus ayant une déficience cognitive légère (DCL) en utilisant une approche du cerveau entier et axée sur les données. Le DCL est souvent un état précurseur de la

démence. Par conséquent, si les sujets qui progresseront et qui ne progresseront pas peuvent être identifiés quand ils ne sont que légèrement déficients, le développement et la mise en œuvre d'un traitement de la démence plus rapide et ciblé peuvent être poursuivis. Nous avons développé une nouvelle approche d'apprentissage machine multi-classificateurs optimisé par la valeur prédictive positive (VPP) qui a pu classer un sous-ensemble de sujets qui n'ont pas progressé vers la démence avec un VPP élevé et une spécificité élevée. Nous avons également caractérisé 3 sous-types d'individus DCL qui ont progressé vers la démence, et 2 sous-types d'individus qui n'ont pas progressé. Nous avons constaté que les mesures basées sur l'IRM (volume de matière grise, épaisseur corticale et CF basée sur la corrélation) impliquant les régions occipitales, les régions sensorimotrices, et les régions pariétales médiales étaient les plus souvent impliquées dans les différences entre les sous-types de DCL.

Nous concluons que les méthodes de modélisation de réseau appliquées à l'ER-IRMf dans le contexte de régions cérébrales à échelle fine sont utiles pour caractériser la connectivité organisée de manière rétinotopique dans le cortex visuel humain, peuvent extraire la connectivité directe de manière fiable dans un réseau local densément connecté, et contribuer à une classification précise d'un sous-ensemble de sujets stables DCL.

Acknowledgements

I would like to take this opportunity to thank my supervisor, Amir Shmuel, for his guidance and support over the 3 years and 3 months of full-time work we have spent together on this PhD. Your continued support while I finalize my thesis since returning to medicine has also been invaluable. I have learned a lot from you and will take those lessons with me as a move on in my career.

My PhD advisory committee has also been invaluable in guiding me through my journey to becoming a capable scientist. Thank you, Jean Gotman, Pierre Bellec, and David Buckeridge.

I would also like to thank the lab members (past and present) who have helped me along the way, in particular with navigating technical issues (!), including Kelvin Mok, Sethu Boopathy Jegathambal, Ricardo Pizarro, and Fadi Ayad.

There has also been a number of undergraduate students who have contributed to my projects as well as provided me with the opportunity to mentor, so I thank you for your enthusiasm and work: Wenzong Xia, Tianzhu Fu, Ziqi Hoa, Samuel Gregoire, and Narendra Shivaraman.

I have been greatly inspired by the McGill Imaging Neuroscience community since I first began in my graduate studies at McGill in 2010, so I thank its members, particularly those based at the Montreal Neurological Institute.

I would like to thank the McGill Undergraduate Medical Education (UGME) Program as well the MD-PhD Program's director, Mark Eisenberg, and the program's advisory committee for taking a chance on me and giving me the wonderful opportunity to study medicine as well as pursue my passion for science. The experiences I have accumulated with the UGME and the MD-PhD program have shaped my professional goals and have inspired me beyond what I could have imagined at the start of my journey.

To the MD-PhD students, thank you for your support and encouragement through the many years we have navigated this path together. I wish the best of luck to all of you on your respective journeys.

Last but not least, I would very much like to thank my friends and family for their unwavering support and encouragement along the way. You guys are the best: Mom, Dad, Jen, Neal, Louise, Kelvin, Alex, Menglan, and Lucas.

Contribution to Original Knowledge

Summary of contributions and value of the work

Throughout these studies, **a focus is on evaluating and proposing methodological approaches** which we optimize for how closely they can reflect reality (Manuscripts 1, 2, and 3).

The original methods we have developed include (1) Local Principal Component regression (LPC), and (2) a Positive Predictive Value (PPV)-optimized multi-step machine learning classification (Manuscripts 2 and 3, respectively).

Better methods support meaningful interpretations of experimental measurements.

Specifically, our work is contributing to the **identification of an accurate non-invasive measure of fine-scaled direct functional connectivity**, something as of yet not achievable. (Manuscripts 1 and 2)

We also contribute to better understanding the important features in **distinguishing individuals with Mild Cognitive Impairment** (MCI) who will progress to dementia from those who will remain stable in the MCI state. This will allow for better prognostication and eventually will lead to the development of better treatments through clarification of mechanisms underlying this pathological state. (Manuscript 3)

Finally, **our set of machine learning analyses optimizing PPV** can contribute to future identification of individuals for whom we can make a confident prognosis, allowing to optimize care and cohort definitions from an earlier time frame. (Manuscript 3)

Below we explain the original contributions of knowledge of each of the manuscripts.

1. *"Partial-correlation based retinotopically organized resting-state functional connectivity within and between areas of the visual cortex reflects more than cortical distance"*

- Our paper is the first to present comprehensive fine-scaled FC network consisting of V1, V2, and V3 analyzed by correlation and partial correlation. Earlier studies reported correlation within V1 and retinotopically sensitive connectivity between V2 and V4.
- Our paper is the first to show that in the network consisting of V1, V2, and V3, partial correlation-based functional connectivity (FC) reflects more than distance effects.
- We present the first thorough discussion of the possible mechanisms of correlation between distant eccentricity zones in V1.

2. *"Local Principal Component regression, partial correlation, and inverse covariance identify direct connections in an unknown, fine-scale, densely connected functional network"*

- We present the first evaluation of network modeling methods on a fine scale, in the context of a densely connected network of which one has no prior knowledge.
- Our work is the first to show that with a large sample size and statistical thresholding, specific modeling methods' network predictions converge to the real direct connectivity.
- Our work is the first to show that statistical thresholding is superior to density-based thresholding in the context of no prior knowledge of the system of study.
- We present a novel method of regressing the "global signal" within a local network, which improves network modeling methods' performances in term of predicting direct connections.

3. *"High positive predictive value in a subset of individuals using a multi-step machine learning classification procedure for identification of progression from mild cognitive impairment to dementia"*

- We have proposed a novel multi-classifier machine learning pipeline which optimizes class-wise positive predictive value in order to precisely classify a subset of the population of interest.
- Our work is the first to classify, with high precision, a subset of individuals with mild cognitive impairment (MCI).
- We present the first findings that FC feature-based classifiers are more contributory than structural feature-based classifiers to a set of multiple classifiers optimizing the positive predictive value of a subtype of MCI individuals
- Our work is the first to use the Human Connectome Project's cortical parcellation in a whole brain classification study of MCI individuals.
- Our work is the first to use L1-regularized Inverse Covariance (ICOV) as a FC measure in a whole brain classification study of MCI individuals.

Contribution of Authors

Manuscript 1: "*Partial-correlation based retinotopically organized resting-state functional connectivity within and between areas of the visual cortex reflects more than cortical distance"*

Debra Ann Dawson[1,2], Jack Lam[1,2], Lindsay B. Lewis[1,3], Felix Carbonell[1,2], Janine D. Mendola[1,3] and Amir Shmuel[1,2,4]

[1]McConnell Brain Imaging Centre, Montreal Neurological Institute

[2]Department of Neurology and Neurosurgery

[3]McGill Vision Research, Department of Ophthalmology

[4]Departments of Physiology and Biomedical Engineering

McGill University, Montréal, Québec, Canada

This paper is published in Brain Connectivity 6(1), pages 57-75 (2016).

DAD: Primary author. Concept development, methods development, writing, analyses, figure generation, interpretation of results.
JL: Preprocessing pipeline generation and Data preprocessing.
LBL: Data acquisition.
FC: Selection of statistical methods.
JDM: Data acquisition, early concept development.
AS: Principal Investigator. Concept development, data acquisition, methods refinement, interpretation of results, text editing.

| Contribution of Authors

Manuscript 2: *"Local Principal Component regression, partial correlation and inverse covariance identify direct connections in an unknown, fine-scale, densely connected functional network"*

Debra Ann Dawson[1,2], Jack Lam[1,2], and Amir Shmuel[1,2,3]

[1]McConnell Brain Imaging Centre, Montreal Neurological Institute

[2]Department of Neurology and Neurosurgery

[3]Departments of Physiology and Biomedical Engineering
McGill University, Montréal, Québec, Canada

This paper has been resubmitted for publication to NeuroImage following revisions.

DAD: Primary author. Concept development, methods development, writing, analyses, figure generation, interpretation of results.
JL: Preprocessing pipeline generation and data preprocessing.
AS: Principal Investigator. Concept development, data acquisition, methods refinement, interpretation of results, text editing.

| Contribution of Authors

Manuscript 3: *"High positive predictive value in a subset of individuals using a multi-step machine learning classification procedure for identification of progression from mild cognitive impairment to dementia"*

Debra Ann Dawson [1, 2], Ziqi Hao [5], Wenzong Xia [4], Tianzhu Fu [4], Kelvin Mok [1,3], Li Lin [6], Ying Han [6], Pierre Bellec [7], Amir Shmuel [1, 2, 3]

[1]McConnell Brain Imaging Centre, Montreal Neurological Institute

[2]Department of Neurology and Neurosurgery

[3]Department of Biomedical Engineering

[4]Departement of Electrical and Computer Engineering
McGill University, Montréal, Québec, Canada

[5]Electronic Information Engineering, Sichuan University, Chengdu, Sichuan, China

[6]Xuanwu Hospital of Capital Medical University, Beijing, Hebei, China

[7]Centre de recherche de l'institut de gériatrie de Montréal, Université de Montréal, Montréal, Québec

This paper is in preparation for publication.

DAD: Primary author. Concept development, methods development, preprocessing pipeline development and data preprocessing, analyses, figure generation, interpretation of results, writing.
ZH: Preprocessing pipeline refinement and data preprocessing.
WX: Python analysis code development, methods refinement.
TF: Python analysis code development, methods refinement.
KM: Methods development and refinement.
LL: Data acquisition.
LY: Data acquisition.
PB: Concept development and methods refinement.
AS: Principal Investigator. Concept development, methods refinement, interpretation of results, text editing.

Thesis Introduction and Literature Review

General Introduction

Neuroimaging has grown immensely over recent years as imaging modalities such as Magnetic Resonance Imaging (MRI) have now reached excellent spatial resolution, and different data acquisition paradigms make the technology more widely applicable. MRI scanning is an ideal way to explore the brain given its non-invasiveness and the absence of a need for contrast agents or radiation. Structural information, such as cortical thickness or gray matter volume, can be readily computed from structural T1 MRI scans. Functional MRI (fMRI) scans allow the interpreter to observe dynamic relationships between brain areas, such interactions being termed "functional connectivity" (FC). FC was originally defined as "temporal correlations between spatially remote neurophysiological events" (Friston, 1994), but is now more broadly measured via correlation or alternative interactions scores, including wavelet coherence, Granger causality, Bayesian graph estimators, or partial correlation (Smith et al., 2011; Dawson et al., 2013).

fMRI is frequently used as an imaging modality in neuroscience research. This modality allows us to study the brain in its many states. Since the discovery of the default mode network and the realization that the brain is very active even when our minds are "at rest", the resting state (RS) of the brain is being increasingly studied (resting state: i.e. when neurons in the brain fire spontaneously in the absence of a stimulus or task). It has been theorized that these resting state fluctuations are shaped by the underlying anatomical connectivity of the brain: neurons firing spontaneously during rest in order to reinforce important connections needed to fulfill the brain's various duties (van Dijk, 2010). Connectivity analyses of fMRI data in the resting state are numerous and contribute to our understanding of both the RS-fMRI signal and the functional connectivity of the brain. Although many studies have been conducted, there is still much to be learned about the best ways to extract relevant connectivity information from the signal, the fine-scaled functional connectivity within the healthy human brain, and differences in connectivity that may be observed in pathological states.

These are the main topics that will be presented throughout the thesis. The thesis is manuscript based with three stand-alone manuscripts coming together to comprise this larger

body of work. The format of the presentation will be that of a general introduction and literature review of topics relevant to all three studies, as well as a brief overview of the experimental approach and results from each individual study. Following this introduction, the three manuscripts will be presented, each with their own more tailored introduction and literature review, and of course detailed methods, results, and discussion. The manuscripts will be separated by short connecting sections reminding the reader how one study led to the next. Finally, after all three manuscripts are presented, a common discussion will explore some of the overarching ideas from across the three manuscripts, and from the thesis as a whole. Now, we delve further into our literature review, elucidating the concepts and motivations behind the studies to come.

More on Resting-State fMRI

A question one might ask is: Why focus on the resting-state as opposed to task or other states? Firstly, for task-based fMRI, if we wanted to cover the whole brain, we'd need to know specific tasks that can activate all parts of the brain. We don't know this; thus the extent of the brain we can access with this approach is limited. Of note, task networks are identifiable in the RS (Smith et al., 2009), suggesting that the RS may allow a comprehensive study of multiple, possibly all task states of the brain. Secondly, a sleep state might also have been considered. Studies show that in early sleep (Larson-Prior et al., 2009) and even during deep non-REM sleep (Tagliazucchi et al., 2013) the spatial extent of resting-state networks is preserved. Thus, one can view sleep as an alternative option to RS. Thirdly, the spatial extent of resting-state networks has been shown to be preserved even under anesthesia and in patients with disorders of consciousness (Boly et al., 2008). In other words, RS networks are expected to be preserved in all healthy brain states, suggesting that any state could have been used to study these networks. That being said, there is also the consideration of the strength of these networks emerging amidst all other fluctuations captured in the fMRI signal. Items that *do change* between each of these brain states are the strengths of activation of the brain regions and also the coherence of fluctuations with other regions. In task-based analyses, the networks outside of the task of focus are often overwhelmed and obscured by the strong signal from the present task (Smith et al., 2009; Raemaekers et al., 2014). In sleep, the intensities in regions responsible for cognition, attention, and awareness are significantly decreased (Tagliazucchi et al., 2013). The resting-state

reflects a middle ground between these extremes that is also very clinically practical. RS has the benefit of being very convenient as there are basic instructions for participants, and no need for additional efforts associated with setting up various stimuli and tasks. This makes RS-fMRI especially attractive for imaging patients, that might not be able to perform tasks. The fact that RS networks are preserved in all states could be interpreted as support of the hypothesis that they capture the relationships forged by direct anatomical linkages, as these are expected to remain constant independent of the activity state of the brain at any given moment.

Fine-scale networks

Throughout the coming work, we will often refer to the idea of "fine-scale" brain regions or connectivity. We will take a moment to clarify what is meant by this. The definition or use of this phrase in the context of brain networks/parcellations/atlases is admittedly vague. Researchers will use the term "fine-scale" to refer to anything from entire visual areas, to 500-2000 regions of interest (ROIs) or to voxel-based analysis (Smith et al. 2019, Liu et al., 2017). In our case, we will use this terminology to refer to the segmentation of cortical regions beyond their basic classical size. For example, in evaluating the visual cortex, we look at a sub-visual-area ROI size. In looking at a whole-brain analysis, we choose a parcellation that captures, in many instances, ROIs on a fine-grained-cortical-area scale. In the upcoming studies we present, we will discuss the nature of the "fine-scale" networks we study, but the above will be the loose definition with which we will work. Certainly, our scale of ROIs will be significantly smaller than those found in traditional atlases such as the Automatic Anatomical Labeling atlas (AAL), the Talairach Atlas using the Broadman Areas, etc., which usually contain under 100 ROIs across the whole brain. Studying finer-scaled ROIs is a valuable next step in FC analyses, given that large-scaled analyses do not allow for an understanding of local connectivity within brain areas, and may, in fact, misrepresent or conceal the details underlying the observable longer-range connectivity between brain areas (Smith, 2012).

Fucntional Connectivity in the Visual Cortex

There is still much to be learned about the functional relationships within the brain, and fine-scaled functional connectivity in even the most well understood sensory cortices is something that can still benefit from more detailed investigations. Among the primary sensory

cortices, the visual cortex is understood best, based on studies in primates and humans (the monkey visual cortex is known to be highly comparable to human visual cortex; e.g. Sereno et al., 1995; Engel et al., 1997; Dumoulin and Wandell, 2008). This makes it a good candidate for a fine-scaled study as we will have a fairly comprehensive context within which to work while we explore the RS-FC. Recent fine-scaled FC analyses in the visual cortex demonstrated that FC between area V1 and V3 is specific to voxels that respond to similar stimulus eccentricities (Heinzle et al. , 2011). Similarly, Gravel et al. (2014) showed that connective field maps between V1 and V2 or V3 are visuotopically organized, and Raemaekers et al. (2014) found greater FC in V1, V2 and V3 between like topographic locations, contralateral functional homologues, and iso-eccentric locations within the same visual area. From these studies, we are inspired to seek out a more comprehensive picture, across V1, V2, and V3, of the fine-scaled FC within the visual cortex. Additionally, we seek to further understand the source of the observed FC relationships through an analysis of FC versus distance.

Our results from Manuscript 1 include significant within visual quadrant Correlation and Partial correlation for all connections within a visual area. In general, Correlation and Partial correlation values were found to decrease with increasing eccentricity separation. Between-visual area within quadrant Partial correlation was found to span to same eccentricity regions and adjacent eccentricity regions. Similarly, pairs of ROIs with similar eccentricity in two different areas showed higher correlation or Partial correlation than pairs of ROIs distant in eccentricity. In terms of our exploration of the relationship of FC vs. distance along the cortex, correlation-based FC vs. cortical distance was approximately linear, whereas Partial correlation showed a more complex relationship to distance, indicating that it is reflecting more than distance effects. We conclude that resting-state functional connectivities within and between lower visual areas are retinotopically organized.

Some Specific Models of FC Explained

As evident in the results, our approach to investigating the fine-scale FC in the visual cortex is in utilizing two very widely used, easily interpretable, and broadly effective models: Pearson's correlation and Partial correlation. In fact, these two models hold a prominent role in all three of the manuscripts contained within the thesis. We will therefore explain them and go over their strengths and weaknesses. Correlation is a simple method that utilizes the covariance

between two vectors, representing, in our case, FC between two time-courses. It is equivalent to the covariance between the two time-courses divided by the product of the standard deviations of each time-course (Marrelec et al., 2005). This allows us to quantify the similarities between these two time-courses, i.e. how often their fluctuations across time correspond. The maximal value is 1 if the time-courses are perfectly linear with each other, and the minimal would be -1 if the fluctuations in the time-courses are perfectly linear but opposite. The midpoint between these, 0, would represent a 0 covariance between the two ROIs, or in other words, no relationship between the fluctuations of their time-courses. Correlation is a model of FC requiring minimal computational resources which also has the benefit of being very sensitive (Smith, 2012; Schmittmann et al., 2015). That being said, it is not very specific to direct connections between ROIs of a network (Marrelec et al., 2006; Friston, 2011; Smith 2012). For example, if A and B are connected, and B and C are connected, but A and C are not, most often correlation will identify a relationship between A and C as there is indeed some commonalities to the fluctuations in their time-courses as influenced by their independent relationships with B (see Figure 1). Some of the common inputs into network ROIs can be accounted for by specific regression of confounds such as motion, physiologic noise, or even a global signal (Marrelec et al., 2005; Smith et al., 2011). However, given that these are broad regressions applied to all ROIs or voxels, the finer-scaled, within-network common influences are still not filtered out (Schmittmann et al., 2015).

Partial correlation is based on the same concepts as Pearson's correlation, however it also strives to take into account information from other regions of the network via a regression procedure. It can be computed as the Pearson's correlation between the residual time-courses of two network ROIs after having regressed out the effects of all other ROIs in the network from each of the two time-courses. The idea of doing these regressions is that the residual will contain only non-redundant information, i.e. information that could not be explained by connections with other network ROIs. Thus, when the partial correlation between two ROIs is non-zero, one can interpret that the connection is a direct linkage (Marrelec et al., 2006; Dawson et al., 2013; Liu and Luo, 2015).

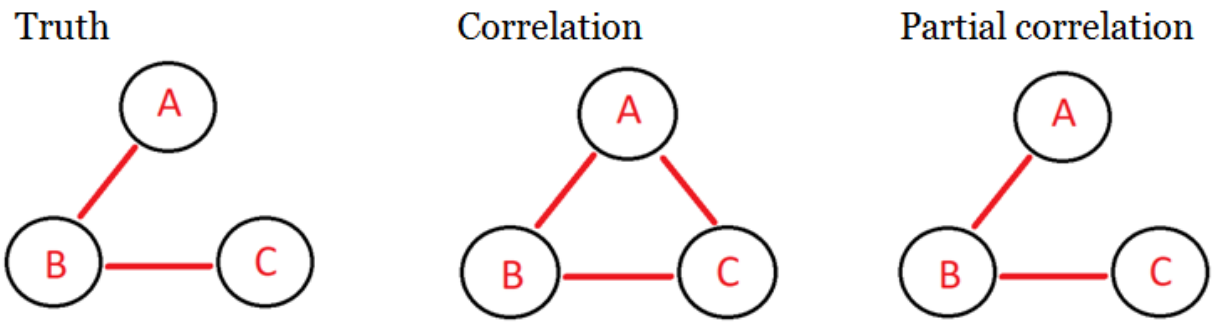


Figure 1. Connectivity in a simple network as predicted by correlation and partial correlation.

A major limitation that one faces with the use of partial correlation is the mathematical necessity that the number of observations (i.e., the number of time-points within a time-course) must be greater than the number of ROIs in the network (Whittaker, 1990; Edward, 2000; Peng at al., 2009; Schmittmann et al., 2015). The full set of regressions is not possible in the absence of this requirement. Thus for best results, the number of observations or length of time-courses in a partial correlation analysis should be long relative to the number of ROIs in the network (Smith et al., 2011; Dawson et al, 2013; Smith, 2012). This is not always feasible or ideal as MRI scanner time is expensive, and subjects may tire in their task, limiting the possible length of time-courses. The number of ROIs in a network (i.e. the network complexity) can easily be very large, the simplest example of this being a voxel-wise analysis (Biswal et al., 1995). If a full-brain voxel-wise analysis is done, the number of ROIs in the network is on the order of tens of thousands (Tong et al., 2016). Importantly, most fMRI time-courses are under 10 minutes in length, or at the most approximately 1 hour, with a repetition time (TR) typically over 1 second (Glover et al., 2011; McDowell et al., 2019). Even at 1 hour and a 1s TR, this would leave us with only 3600 time-points in a time-course: a value much smaller than the number of regions that we have in a whole-brain voxel-wise analysis. In order to overcome the problem of mismatched network complexity and time-course length, researchers have a few options. Firstly, they can either focus on a smaller portion of the brain or cluster voxels into ROIs in order to limit the number of regions in the network (Friston et al., 1993; Biswal et al., 1995; Xiong et al., 1999; Arfanakis et al., 2000; Huettel et al., 2004; Poldrack, 2007; Cohen et al., 2008; Dawson et al., 2013). Secondly, they can lengthen time-courses: this is often achieved by concatenating

data from several fMRI sessions (Smith et al., 2011; Dawson et al., 2013). Thirdly, alternative methods that do not have this constraint are another way of dealing with the issue. Researchers who were interested in retaining a multivariate approach, capturing the essence of partial correlation, though with more flexibility with respect to the number of time-points versus the number of ROIs, have used the approach of imposing sparsity in the covariance matrix, effectively limiting the number of regressions that would be required. This approach is called regularization. We, too, adopted this approach for dealing with balancing network complexity vs. time-course length (manuscript 3 in this thesis). We chose it as it was one of the most successful models in two prior studies evaluating network modeling methods (Smith et al., 2011; our own study, Dawson et al., 2013) and also due to its simplicity and intuitive interpretability. We discuss further below.

Partial correlation can indeed be computed via multiple regressions and linear correlation, but it can also be equivalently computed via the inverse of the covariance matrix (Cov).

$$\text{Partial correlation } (x, y \mid Z) = Cov^{-1}(i,j) / (Cov^{-1}(i,i) * Cov^{-1}(j,j))$$

where x and y are the network ROIs between which we are establishing the relationship, Z are the full set of time-courses from the other ROIs in the network, and i & j are the indices of x and y in the Covariance matrix, respectively (Marrelec et al., 2006). Spin-off methods from Partial correlation, meant to induce sparsity in the predicted network include the Graphical Lasso (which we will refer to as ICOV, for Inverse COVariance; Friedman et al. 2008, Banerjee et al. 2006) and L2 ridge regularized Inverse Covariance (Pervaiz et al., 2020), among others (Meinshausen and Buhlmann, 2006; Yuan and Lin, 2007; Li and Gui, 2006; and Schafer and Strimmer, 2005; Schmittmann et al., 2015; Pervaiz et al., 2020). In ICOV, our chosen approach for regularization, the covariance matrix is made sparse by the iterative Lasso method, the level of sparsity controlled by regularization parameter lambda, and then once the Lasso procedure converges, the regularized covariance matrix is inverted and partial correlation is computed. The algorithm minimizes matrix entries that are very small down to zero, thereby introducing sparsity. The result is a sparser set of direct connectivity measures, with all the pros of partial correlation (i.e. indirect connectivity removed), and fewer cons (less limited by session duration in terms of possible network complexities). One possible additional con, however, is that the sparsity of the connectivity in the network is artificially produced by an arbitrarily selected

regularization parameter; thus it is possible that one could be inducing inappropriate sparsity into the network (Wang et al., 2016; Liu and Luo, 2015). In the various studies presented in this thesis, we will look at how well these models perform under varying experimental conditions, evaluating and applying them, as well as elucidating further their applicability.

Researching Methods in the Context of Resting-State Functional Connectivity

Methods research in the context of RS-FC has been a very popular area of study in recent years, with mathematical models being evaluated, refined, and developed for optimal accuracy of the connectivity estimated by the model (ex. Smith, 2012; Haak et al., 2018; Allen et al., 2014; Beckmann et al., 2005; Griffanti et al., 2015; Power et al., 2014; Park et al., 2017; Pervaiz et al., 2020). Methods are optimized differently depending on the scale and goal of the study. For example, many studies look at whole-brain FC using atlas-based ROIs or data driven parcellations (ex. Jarrahi et al., 2015, Allen et al., 2011; Cohen et al., 2008; Finn et al., 2015; Smith et al. 2013; Yeo at al., 2011). Given the widespread nature of the connectivity to be observed in these studies, more sensitive models or models better able to work with a large number of ROIs, such as correlation, tend to be most appropriate (ex. van Wijk et al., 2010; Smith et al., 2011; Ryali et al., 2012). In finer-scaled FC analyses, the opposite is true: more discerning methods are most appropriate as greater discrimination is needed in order to observe local connectivity amidst the larger-scaled connectivity which is superimposed (ex. Raemaekers et al. 2014; Dawson et al., 2013; Buckner et al., 2011; Rausch et al., 2016; Huang and Ding, 2016; Hinne et al., 2015), hence our interest in partial correlation and ICOV. Some researchers may be interested in capturing a global measure of the extent of connectivity in a given network (such as characteristic path length, mean clustering coefficient, etc.) (ex. Garrison et al., 2015; van den Heuval et al., 2017; Fornito et al., 2010; Rubinov and Sporns, 2010; Hayasaka and Laurienti, 2010), whereas others may be interested in contrasting sets of connectivity measures characterizing within network relationships (ex. van den Heuval and Hulshoff, 2010; Yeo et al., 2011; Smith et al., 2011; Dawson et al., 2013). All of these possibilities leave us still in search of ideal methods and measures for use in studying RS-FC. One particular focus of this thesis will be the evaluation of the performance of FC measures in terms of identifying direct connectivity on a fine-scale.

Identifying Direct Connectivity from Resting-State Functional MRI

One might wonder why one would apply functional connectivity to capture direct-anatomical connectivity. Why not use another modality which aims more directly to capture anatomical information? One could consider looking at structural MRI, for example. In this modality, the spatial resolution of scans is typically on the order of 1 millimeter. However, neuronal axon diameters are on the scale of microns, and so consequently, axons cannot be captured directly via structural MRI. One could also consider looking at Diffusion MRI. There are many approaches to describing white matter tracts using Diffusion MRI, but the most common is that of Streamlines. Streamlines between gray matter regions are tracts formed by successively stepping in the direction of the principal eigenvector representing the direction of fastest diffusion of water along tissues in the brain (O'Donnell and Westin, 2011). The issues with these measures are that long pathways are not well captured due to the accumulation of small errors during tractography computation (Morris et al., 2008; Clayden et al., 2008) and that there is difficulty in deciding how one should deal with voxels on the periphery of the tract. Diffusion is well suited for large white matter tracts, but if small axon bundles or single axons are running in the gray matter, diffusion can't capture these. Also, there isn't yet strong validation of Diffusion MRI-based studies as this imaging modality is immature, having only started to see a ramp-up of research in the last nine years or so (Eguchi et al. 2017). We hypothesize that at a fine-scale, diffusion-based tractography will fail to accurately and reliably capture anatomical connectivity. However, as Diffusion MRI methods and technology improve, we hope to use RS-FC as a compliment to diffusion tractography, making our current pursuit beneficial regardless of the state of Diffusion advancement.

We have already established that RS-FC is thought to reflect anatomical linkages (De Luca et al., 2006; Fox and Raichle, 2007; Koch et al., 2002; Quigley et al., 2003; van den Heuvel et al., 2009), though this reflection is not trivially identical (Honey et al. 2010). Of note, the inverse approach, studying anatomical connectivity, allows one to capture RS networks across the whole brain (Van Den Heuvel et al., 2009; Greicius et al., 2009; Haimovici et al., 2013). In other words, networks of anatomical connections include all connections in known RS functional networks. Importantly, upon pursuing our fine-scaled RS-FC analysis within the visual cortex in Manuscript 1, we observed that Partial correlation was predicting connections that were in line

with what we would expect from anatomical connectivity, based on previous tracer injection studies in primates. This ignited our curiosity further with respect to the possibility of using RS-FC methods to extract/approximate anatomical connectivity in the context of a fine-scaled network of which no prior information exists. Though there is likely more information contained within these functional signals beyond anatomy, we expect that anatomical linkages contribute significantly to driving the resting-state signal.

Testing Network Models in the Context of No Prior Knowledge of the System of Study

The hypothesis that FC can capture direct-anatomical connectivity inspires the study presented in Manuscript 2. Our hope is that measures of direct connectivity can be used as approximations to the underlying anatomy in the system, something not yet measurable on a fine-scale non-invasively. In this study, we test and evaluate a number of network modeling methods, including correlation, partial correlation, and ICOV (mentioned above), in terms of their abilities to predict direct connectivity in the context of an unknown network. Marellec et al. (2006) outline nicely in their paper why it is important to develop good methods for the situation where one has no prior knowledge of the system with considerations to cross-modality generalization, incomplete understanding of networks, assumptions underlying computational methods, etc. Previously, we evaluated network models in the context of *having prior knowledge* of the system, in Dawson et al., 2013 (see Appendix B for clear details on how Dawson et al., 2013, and manuscript 2 in this thesis differ). This knowledge made it possible to set thresholds on FC measures in order to assure no more than 5% false positive identification of direct connections. However, one is not always fortunate as to have prior knowledge of the system. Even in the case where one does have prior knowledge, perhaps that information is incomplete, and it is unclear what effect the incompleteness might have on the results of connectivity analyses. Thus, in order to make the decision of whether a connection can be retained as a direct connection in the network, we apply standard statistical thresholding to the measures of FC obtained with a variety of network modeling methods. We also look to identify which models can accurately reflect anatomy with stability across different thresholding levels, such that for future use of these models, a specific threshold level is not critical, making the methods more generalizable. Another angle we consider is different types of thresholds: we compare connection density-based thresholding to thresholding based on standard statistical

comparison to zero. We further evaluate the impact that the sample size can have on the results of statistical testing. For the latter, we undertake an analysis with the Human Connectome Project data, which allows us new insight into the performance of the models under the condition of a larger sample size, and also confirms the behaviors of the models we had observed with our own data set.

In this work we also present a novel method for accounting for network effects: Local Principal Component (LPC) regression. "Network effects" are the widespread effects caused either by broad propagation of information from within the network or from sources outside the network. The LPC method utilizes a model of the global signal, which we estimate by computing the first principal component from a principal component decomposition of the data. Carbonell et al. (2011) show that the first principal component is reliably analogous to the global signal, or the common fluctuations to all voxels provided to the model. They also show that this global signal estimation is expected to be robust against generating spurious anticorrelations when regressed from ROI time-course data, making it a model of the global signal not prone to confounding effects. We utilize this concept within a local network, not over the whole brain, thus capturing the set of important network effects within the local network. We demonstrate that regression of the LPC is effective for models such as correlation in terms of allowing them to capture direct connectivity with statistical thresholding, whereas without this regression, they fail to differentiate between direct and indirect connectivity completely. We also show that LPC regression can be beneficial even for multivariate models such as partial correlation and ICOV. This is an important finding as Partial correlation and ICOV, by design, also remove "network effects" via their multiple regressions (Marrelec et al., 2006), including from sources outside the network if the effect that they exert is widespread. Of note, the LPC is computed over the entirety of V1, V2 and V3, and therefore includes information from more voxels than our final set of ROIs after quality control.

Other important study findings from Manuscript 2 include that Partial correlation and ICOV were most reliable at predicting a network of connections with a high true positive rate and a low false-positive rate relative to our defined inferred ground truth network across tests (often with no false positives at all). Bivariate models such as correlation were not capable of producing useable network predictions even when high threshold levels were applied if the LPC

was not regressed. However, correlation and another bivariate method, Kappa, were top performers in tests when the LPC was regressed. In the extension of our study with a large sample size using Human Connectome Project data, network predictions were found to converge toward the inferred ground truth if statistical thresholding was used, but they did not if density-based thresholding was used. Thus, we conclude that multivariate models Partial correlation and ICOV, or correlation with the LPC regressed, are optimal approaches to estimating direct anatomical connectivity in a fine-scale local brain network using common statistical significance thresholding.

FC in Pathological Brain States including Alzheimer's Disease and Dementia

With the promising findings utilizing direct FC measures on a fine-scale in the healthy brain from Manuscripts 1 and 2, we sought next to challenge these measures, and investigate their utility in the context of a pathological brain state. As one might imagine, given that we still have room to expand on our understanding of functional connectivity in primary sensory cortical areas in healthy brains, we most certainly have a need for a greater understanding of the brain in diseased states. With methods in fMRI RS-FC analyses advancing every day and showing promise in terms of new insights to be drawn, researchers are now applying RS-FC analyses to pathological states hoping to observe unique network structure or altered connectivity compared to healthy controls. If differences between the FC networks in pathological states and healthy states are found, there is potential to gain new insights into the mechanisms behind the studied diseases and/or to contribute to diagnostic/screening strategies. A small sample of the many recent studies showing altered FC in pathological states include Hart et al. (2017), who studied patients with right temporoparietooccipital glioblastoma; Sourty et al. (2016) who studied patients with Lewy Body Dementia; Zhang et al. (2012) who studied patients with hepatic encephalopathy; Haak et al. (2016) who studied individuals with macular degeneration; Rausch et al (2016) who studied individuals with autism spectrum disorder; and Onu et al. (2015) who studied patients with Parkinson's disease. In the same spirit as these studies, a number of RS-FC studies have looked at the network alterations observed in patients with Alzheimer's disease (AD)/dementia, or mild cognitive impairment (MCI), a state of cognitive decline which is often a precursor state to dementia. These studies consistently showed altered connectivity involving the default mode network (ex. Balachandar et al., 2015; Binnewijzend et al., 2012; Grothe and

Teipel, 2016; Li et al., 2013; Weiler et al., 2014; Zhou et al., 2015; Tam et al., 2015; Greicus and Kimmel, 2012), as well as (less consistently) other brain regions (ex. Bai et al., 2012, Liu et al., 2015; Cai et al., 2015ab; Sohn et al., 2014; Yao et al., 2013). AD dementia is a great burden to society afflicting approximately one in nine people over age 65 and one in three people over 85, resulting in over 36 million people worldwide who have AD dementia (Hsu and Marshall, 2017). In Manuscript 3 of the thesis, we seek to get a more detailed understanding of the brain changes occurring in those who develop AD/dementia through relatively fine-scaled FC analyses in the context of MCI. We assess if our findings are consistent with the larger-scaled prior studies as well as if there is finer-scaled connectivity information that may allow us a more precise understanding of the changes that are occurring in the MCI brain. It is our hope that, with the addition of more details to our body of knowledge around MCI, we may contribute to the development of early prediction of the expected prognosis and viable screening protocols or potential targets for treatment.

In this MCI study, we chose to do a whole-brain ROI-based data-driven analysis rather than focusing on one region of the brain, as we did in the first two studies where we focused on the visual cortex. We made this choice for three reasons. 1) Our focus is on methods, so the region in the brain studied isn't critical, but what is critical is that the regions we are studying are appropriate for evaluating the methods in the given context. 2) We hope that by moving out of the visual cortex we may be able to comment on the generalizability of the methods we developed and/or evaluated for identifying direct FC to other brain regions. 3) Given the importance of AD and dementia and in society, we have a special interest in studying MCI, and the visual cortex is not a particularly relevant model of connectivity or anatomical changes for this pathological state, according to the current literature. Thus, given the broad set of brain regions implicated in MCI, we have chosen a whole-brain analysis, expecting that it would be most appropriate.

AD is a disease that develops over a long period of time, on the scale of 20 years, before symptoms begin to surface (Fleisher et al., 2012). Interventions that begin after symptoms have appeared have not yet proven effective in disease reversal. The field is in agreement that therapies that will have the most likelihood for success will be those which can be started early in the disease, halting its progression (Hsu and Marshall, 2017). In order to be able to deliver

potential therapies at an earlier stage in the disease, it is critical that we have reliable strategies to identify individuals who will develop dementia. One of the most promising approaches to making these predictions is using machine learning techniques.

Machine Learning Approach to Predicting the Prognosis of Pathology

Machine learning is a computational approach in neuroimaging studies that has increased immensely in its use in recent years (Glaser et al., 2019). There is supervised machine learning which utilizes trends in the data associated with labels in order to predict what label should be associated with a new dataset. Alternatively, in unsupervised machine learning, the algorithm determines both the labels and the trends before predicting a label for new unseen data. Many different machine learning paradigms and pipelines exist and have been tested in the context of predicting the development of pathology, including looking at the question of whether patients with MCI will progress to dementia. Some studies which have looked at this include Cui et al. (2011), Liu et al. (2013), Li et al. (2014), Yu et al. (2014), Cheng et al. (2015), Tong et al. (2017), Dansereau et al. (2017), Mathotaarachchi et al. (2017), and Gómez-Sanchoa et al. (2018). Machine Learning methods can vary in many ways. One way they may vary is in terms of the type of classifier used. Common choices include Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR). Validation methods may also vary; one may do a leave-one out cross-validation, K-fold cross-validation with varying K's, nested cross-validation, etc. The data type and the number of features utilized may also vary; some studies focus on one type of data, while others are multi-modal. How one selects the features to be used is also a choice to be made; there are many approaches to this, including statistical selection such as the Chi-squared test, correlation-based selection, or tree-based selection. Finally, the hyperparameters associated with the classification model need to be selected or optimized in each case, some of which may include regularization parameters, kernels, or weights. It can be very difficult to know which set of parameters and choices in designing a machine learning classifier will lead to optimal results. One strategy for overcoming this is by doing a grid search over many parameters. In this approach, one does cross-validation (CV) within the training set of data in order to optimize parameters and then tests the winning classifier on test/validation data. One can either do this with one left-out validation dataset or over several in so-called nested cross-validation. Nested CV has an inner loop of CV to optimize parameter selection,

done on training data, and an outer loop CV (several splits of training and test data) where the optimized classifiers are tested on validation data, allowing one to see if the optimization pipeline within the inner-loop is generalizable across multiple data splits (Krstajic et al., 2014). This is the approach we will use in Manuscript 3 while testing many different choices in a grid search to build our classifiers.

RS-FC and Structural MRI-based Prediction of Progression from Mild Cognitive Impairment to Dementia

As in Manuscript 2, we pursue the development of novel methodological approaches in our third study, looking to find a more clinically relevant optimization of machine learning results on the difficult to classify, heterogeneous population of individuals with MCI. Specifically, the classification we pursue is that of determining which patients will remain stable in their MCI state, and which will progress further to dementia. This earlier phase of identification is critical as up to 50% of individuals with MCI progress to dementia (most significantly of the AD type) within 5 years (Gauthier et al., 2006; Arbabshirani et al. 2017). As previously mentioned, many studies have already looked at this question; however, the overall accuracy is typically under 80%, and rarely are both sensitivity and specificity high: best results with robust validation are on the scale of 85% sensitivity and 73% specificity (Mathotaarachchi et al., 2017). Inspired by our own observations in preliminary classifications and by a study by Dansereau et al. (2017), we strive to make results more clinically relevant by optimizing precision, or positive predictive value (PPV), for a subset of MCI individuals who can be reliably classified. While this approach means we do not make predictions regarding the progression to dementia for all subjects tested, the goal is to be able to make confident predictions, with PPV>95% about a subset of the subjects. To this end, we have developed a multi-classifier approach optimizing PPV for data-driven subtypes of MCI patients.

In Manuscript 3 we found that it is stable MCI subtypes that we can reliably classify with high PPV. In these subtypes, we found an overall prevalence corrected PPV of 96%, achieving our goal of a PPV>95%. We also found that while both the finer scaled brain parcellation and the larger scaled brain parcellation are successful in the stable-subtype classification task with our multi-classifier approach, the finer-scaled parcellation led to more subjects being correctly classified. FC features were found to be more important than structural measures in contributing

to our overall classification results; however, both were relevant predictors. We found that 3 progressor and 2 stable subtypes were optimal given our data set. Features involving visual, medial parietal and sensoritmotor cortical areas were most often implicated in tests seeking to distinguish the subtypes. We conclude that a multi-classifier, PPV-optimized approach is promising for distinguishing subtypes within a heterogenous population such as that of individuals with MCI.

Summary

To summarize, we began our first study with a curiosity for looking at functional connectivity (FC) on a fine-scale in the visual cortex, and in this pursuit we sought to evaluate the relationship between the distance between regions and the functional connectivity between those regions. We also looked at two different functional connectivity measures: correlation and partial correlation as these both have strengths and weaknesses. One of the main observations from this study was that partial correlation was far superior to correlation in identifying direct connections in the visual cortex. We wondered if there were any other models better suited to the task of identifying direct connectivity in a fine-scaled network. Furthermore, we were especially curious to see which of these would be successful in a situation where we have no prior knowledge of the system of study. We asked these questions in our second study and observed that there were a handful of approaches that were quite reliable in the task of identifying direct connectivity in a fine-scaled brain network. These approaches included regression-based multivariate models partial correlation and ICOV, or correlation with regression of a local model of the global signal, the Local Principal Component. Given the success of these models in our network of evaluation, the visual cortex, we sought to expand our application of these models outside of the visual cortex, into a whole-brain study, but still using relatively fine-scaled regions of interest. Having already shown the success of partial correlation in the visual cortex of healthy individuals, we also wished to test these models now in a pathological state, assessing how useful they might be clinically. To this end, we chose to study individuals with Mild Cognitive Impairment (MCI) in our third study, looking to see if fine-scaled FC could be helpful in identifying which subjects are likely to progress to dementia, and those who will remain stable in an MCI state. We found that FC was indeed helpful in this

classification task and that a finer scaled parcellation of the brain was beneficial to improve classification results.

Overall Hypotheses of the Thesis

(1) Resting-state fMRI carries information that can be used to identify direct connections in a fine-scaled, densely connected network of which no prior knowledge exists. This information can be extracted with computational methods.

(2) Relatively fine-scaled RS fMRI-based FC can improve the accuracy of classification of patients with mild cognitive impairment (MCI) into subgroups of those who will progress to dementia and those who will not.

This thesis will tackle each of these hypotheses, each step utilizing the findings from the previous steps in the hopes of improving and pushing further the methods behind and findings from resting-state functional connectivity analyses using fMRI data.

Introduction to Manuscript 1

We begin the body of the thesis with studying the fine-scale functional connectivity in the early visual cortex of humans, i.e. areas V1, V2, and V3. Within this work we consider within visual area connectivity as well as between visual area connectivity, exploring these both in terms of the difference in retinotopic eccentricity of the regions and the physical distance along the cortex between regions.

The physical distance between regions is known to be inversely correlated with FC measures, i.e., the closer in proximity the regions are, the greater the probability of a high FC (e.g., Katsuki et al., 2013; Liao et al., 2013; Liang et al., 2013), but it has also been observed that strong FC exists between distant brain areas in large scale networks (e.g. Liao et al., 2013). In our study, we seek to observe the trends in FC with both correlation and partial correlation in the context of this *fine-scale* visual cortex network as well as the relationships that these measures have with distance along the cortex. Our hope in examining these relationships is to extend our understanding of them beyond only large-scale networks to also finer-scale, local networks.

In the following published manuscript, for a detailed summary of the process of retinotopic mapping the reader is referred to a prior publication from our group, Dawson et al. (2013). Given the importance of retinotopic mapping to the definition of the regions of interest in both parts 1 and 2 of the thesis, we include such a summary here, followed by Manuscript 1.

Retinotopic mapping and definition of visual areas

Retinotopy is the spatial correspondence of the visual cortex's functional organization to the inverse of the visual field representation (Inouye, 1900, Holmes, 1945). Retinotopic mapping allows the delineation of cortical regions which process different features of the visual field and at different levels of detail. Here we focus on the delineation of visual areas V1, V2, and V3 only. These cortical regions can be defined according to conventional phase encoding retinotopy (Engel et al., 1994, Larsson and Heeger, 2006, Sereno et al., 1995). In order to pursue phase encoding retinotopy, subjects must undergo fMRI scanning with stimuli of expanding or contracting rings and rotating wedges which traverse the visual field in a periodic manner.

Eccentricity and polar angle positions are obtained from the ring and wedge stimuli respectively. The temporal phases of the evoked fMRI responses are matched to the phases of the two presented stimuli by shifting voxel time-series in accordance with hemodynamic delay (~6 seconds). The Fast Fourier Transform of the resultant time-series allows for a phase-amplitude decomposition of the signal (Larsson and Heeger, 2006). Retinotopic visual areas were identified according to the criteria documented in Larsson and Heeger (2006). Briefly, visual areas V1, V2 and V3 can be defined by identifying the cortical boundary where the reversal of direction of increasing polar angle phase maps occurs. These boundaries should be perpendicular to the eccentricity phase maps, and typically, they are manually drawn (as is the case in our studies). See Figure 2 below for a representation of the phase maps generated using both polar angle and eccentricity-based stimuli and the corresponding visual area delineations.

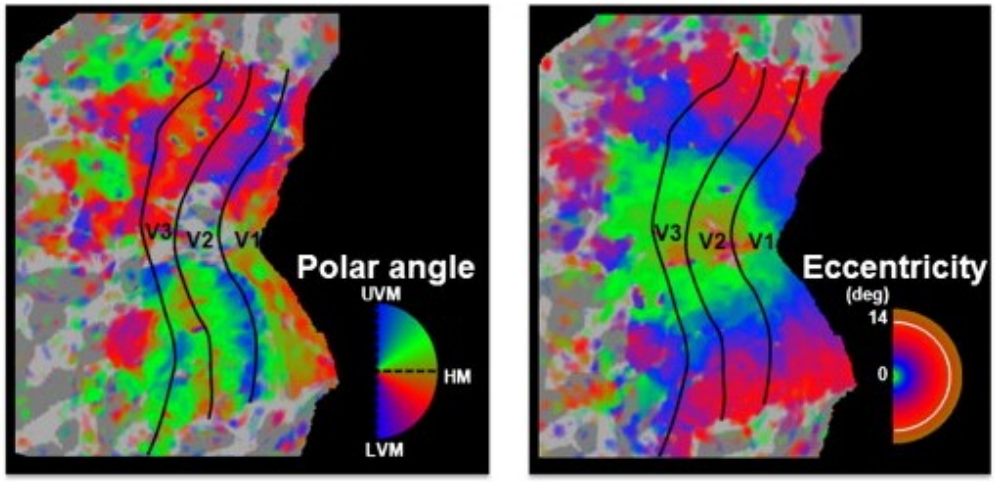


Figure 2. Retinotopy and the delineation of visual areas on the cortex. Visual areas V1, V2 and V3 were defined by examining the reversal of direction of increasing polar angle phases perpendicular to the iso-eccentricity delineation. On the left we see the functional response to a stimulus changing in terms of polar angle and on the right we see the response to a stimulus changing in terms of eccentricity. Each hemisphere was divided for analysis according to its dorsal and ventral parts, based on the mapping of the horizontal meridian, resulting in brain regions representing the four quadrants of visual space. UVM = upper vertical meridian, HM = horizontal meridian, LVM = lower vertical meridian. Figure originally published in Dawson et al. (2013).

Manuscript 1

Partial-correlation based retinotopically organized resting-state functional connectivity within and between areas of the visual cortex reflects more than cortical distance

Debra Ann Dawson[1,2], Jack Lam[1,2], Lindsay B. Lewis[1,3], Felix Carbonell[1,2], Janine D. Mendola[1,3] and Amir Shmuel[1,2,4]

[1]McConnell Brain Imaging Centre, Montreal Neurological Institute

[2]Department of Neurology and Neurosurgery

[3]McGill Vision Research, Department of Ophthalmology

[4]Departments of Physiology and Biomedical Engineering

McGill University, Montréal, Québec, Canada

Correspondence:

Amir Shmuel, Ph.D.

3801 University Street, room NW109
Montreal, Quebec H3A 2B4, Canada

amir.shmuel@mcgill.ca

Acknowledgements

We thank Sepide Movaghati for her comments on an earlier version of the manuscript, and Ilana Leppert and Bruce Pike for support with MRI physics. This work was supported by NIH grant R01 EY015219 awarded to JM, and grants from the Human Frontier Science Program (RGY0080/2008) and the Natural Sciences and Engineering Research Council of Canada (RGPIN 375457-09 and RGPIN 2015-05103) awarded to AS.

Abstract

Numerous studies have demonstrated fMRI-based resting-state-functional-connectivity (RSFC) between cortical areas. Recent evidence suggests that synchronous fluctuations in BOLD-fMRI reflect functional organization at a sub-visual area scale. Here we investigated whether RSFCs within and between lower visual areas are retinotopically organized, and whether retinotopically organized RSFC merely reflects cortical distance. Subjects underwent retinotopic mapping and separately, resting-state fMRI. Visual areas V1, V2 and V3 were subdivided into regions of interest (ROIs) according to quadrants and visual-field eccentricity. FC was computed based on Pearson's linear-correlation (correlation), and Pearson's linear partial-correlation (correlation between two time-courses after the time-courses from all other regions in the network are regressed-out). Within a quadrant, within visual areas, all correlation and partial-correlation FC measures showed statistical significance. Consistently in V1, V2, and to a lesser extent in V3, correlation decreased with increasing eccentricity separation. Consistent with previously reported monkey anatomical connectivity, correlation/partial-correlation values between regions from adjacent areas (V1-V2 and V2-V3) were higher than those between non-adjacent areas (V1-V3). Within a quadrant, partial-correlation showed consistent significance between regions from 2 different areas with the same or adjacent eccentricities. Pairs of ROIs with similar eccentricity showed higher correlation/partial-correlation than pairs of ROIs distant in eccentricity. Between dorsal and ventral quadrants, partial-correlation between common and adjacent eccentricity regions within a visual-area showed statistical significance; this extended to more distant eccentricity regions in V1. Within and between quadrants, correlation decreased approximately linearly with increasing distances separating the tested ROIs. Partial-correlation showed a more complex dependence on cortical distance: it decreased exponentially with increasing distance within quadrant, but fitted best a quadratic function between quadrants. We conclude that RSFCs within and between lower visual areas are retinotopically organized. Correlation-based FC is non-selectively high across lower visual areas, even between regions that do not share direct anatomical connections. The mechanisms likely involve network effects caused by the dense anatomical connectivity and projections from higher visual areas. FC based on partial-correlation, which minimizes network effects, follows expectations based on direct anatomical connections in the monkey visual cortex better than Corr does. Lastly, partial-correlation based retinotopically organized RSFC reflects more than cortical distance effects.

Key words

Resting-state, Resting-state networks, Spontaneous activity, Functional connectivity, Correlation, Partial correlation, Visual cortex, V1, V2, V3, Retinotopy, Retinotopic organization.

Introduction

Since Hubel and Wiesel's seminal work (1962), the functional architecture of the visual cortex has been one of the most rigorously studied themes in the field of neuroscience. This theme covers a wide range of subjects, including the functional properties of visual brain areas such as retinotopic organization and functional specificity. The majority of fine-scaled findings about the functional architecture of the cortex have been achieved using traditional invasive techniques such as electrophysiological recordings (e.g., Hubel and Wiesel, 1962) and optical imaging (e.g., Bonhoeffer and Grinvald, 1991; Shmuel and Grinvald, 2000). Another subject of interest is the thalamo-cortical and cortico-cortical connections. Traditionally, this subject has been studied by anatomical tract tracing (e.g., Felleman and Van Essen, 1991). More recently, functional connectivity (FC) analysis based on functional MRI (fMRI) allows studying the connectivity of the visual cortex non-invasively, thus it is readily usable on human subjects.

FC refers to the temporal correlation between spatially remote neurophysiological events (Friston et al., 1993). FC analyses can be pursued during subject stimulation, task performance, or rest. It can be applied to data obtained with any modality of measurement over time, such as positron emission tomography (PET), blood oxygenation level dependent functional magnetic resonance imaging (BOLD fMRI), electroencephalogram (EEG), and magnetoencephalogram (MEG). FC analyses based on BOLD signal are particularly promising because they can offer high spatial resolution and high spatial specificity relative to where the corresponding changes in neurophysiological signals take place compared to all other non-invasive imaging modalities (Shmuel et al., 2007). FC based on spontaneous signal fluctuations in general introduce no confounds to consider due to co-activation of areas in response to a particular stimulus. In addition, this approach offers convenience in paradigm. Another advantage of this method is that it can elucidate brain connectivity at a large-scale system level, for example, revealing coherent systems of somatomotor, visual, default mode, episodic memory, and attention (Fox & Raichle, 2007). This allows to identify the locations of the systems within the brain readily in order to pursue analyses within a system or between systems. This is complementary to FC analyses at the meso scale, i.e. at the scale of local circuits measured by neurophysiology. Resting state FC analyses can potentially further understanding of neuro-anatomical models (Fox & Raichle, 2007; Wang et al., 2013).

To date, only a few studies have applied fMRI-based resting-state FC analysis at a sub-area spatial scale to the visual cortex. Vincent et al. (2007) demonstrated preliminary results on retinotopically organized resting-state FC between areas V1 and MT of the anesthetized monkey. Heinzle et al. (2011) have demonstrated that the FC between spontaneous fMRI fluctuations recorded in two visual areas depends on their retinotopic coordinates. Raemaekers et al. (2014) have recently considered FC within human visual areas V1, V2 and V3 and indicated that the small scale connectivity information was superimposed on the FC of larger scale networks reflected in the same data.

One criticism of resting state FC analysis using BOLD fMRI data is that the connectivity observed is largely a reflection of distance effects between regions. There have been several studies that focussed on large scale networks and have demonstrated a tendency of FC to be greatest with close proximity of the regions involved (e.g., Katsuki et al., 2013; Liao et al., 2013; Liang et al., 2013). This phenomenon is usually attributed to the abundant presence of anatomical connections at a local scale, as postulated by the small world theory of connectivity in the brain (Kaiser and Hilgetag, 2004). This theory of connectivity has been widely supported with evidence, such as in the macaque brain (Sepulcre et al., 2010). However, in these large scale networks, long range highly functionally connected regions have been identified, indicating that FC analyses can indeed capture information which is more meaningful than a reflection of distance (Liao et al., 2013). We strive to elucidate that within smaller, local networks that are densely connected via anatomical connections such as the visual cortex, there is also meaningful information contained in the FC networks beyond distance.

Using BOLD fMRI resting state FC, we assess the functional connections within and between visual areas of the human occipital cortex, putting an emphasis on how FC depends on retinotopic eccentricity and cortical distance. We evaluate FC using two measures: correlation (Corr) and partial correlation (Pcorr). We have chosen to pursue analyses via these two measures due to their simplicity, the fact that they are widely used in FC analyses, and given that they have been shown to be relatively successful in this type of analysis compared to other common models (Smith et al., 2011; Dawson et al., 2013). Corr is indeed very widely used. However in small, local networks we expect the observed correlations to be high for nearly all connections, due to contribution to synchronization from nodes that are not directly connected. Because of this we include Pcorr which is potentially better equipped to deal with local networks, given that

time-courses from all ROIs in the network are regressed out before assessing the Corr between two regions.

We hypothesize that: 1) resting-state FC within and between lower visual areas is retinotopically organized; 2) the relatively dense anatomical connectivity and projections from higher visual areas cause network effects that increase Corr-based FC across lower visual areas, even between regions that do not share a direct anatomical connection; 3) FC measured by Pcorr, that reduces network effects, roughly follows expectations based on the direct anatomical connectivity in the monkey visual cortex; 4) If network effects are minimized, retinotopically organized resting-state FC reflects more than cortical distance effects.

To test our hypotheses, we look separately at the connectivity within the dorsal and ventral quadrants of the visual cortex in a hemisphere, and between these quadrants, due to the largely symmetric nature of their visual field representation. This separation of within quadrant versus between quadrants also presents a logical separation of generally closer range connections along the cortex (within a quadrant) and longer range connections (between quadrants). In addition, the symmetric arrangement of the visual areas between quadrants creates a distribution of distances within an area that are short (e.g., V1), long (e.g., V3) or intermediate (e.g. V2). We expect higher FC (FC) measures within a visual area, even when computing FC between quadrants. For example, we expect relatively high FC measures between regions representing similar eccentricities in V2 and residing in the dorsal and ventral quadrants respectively, or similarly, regions representing similar eccentricities in V3 and residing in these different quadrants. This creates a scenario in which distant regions residing in different quadrants show relatively high FC compared to ROI pairs that are closer (such as a pair of ROIs respectively in V1 and V2 one residing in the dorsal quadrant and the other in the ventral quadrant, and representing distant eccentricities). This arrangement of visual areas allowed us to test our fourth hypothesis; namely that within the network of retinotopic visual areas, FC measures do no solely reflect distance.

Materials and Methods

Six normally sighted subjects participated in this study after giving written consent in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Each subject was scanned in two sessions: one to obtain high-resolution anatomical images and

fMRI for retinotopy and the other to obtain high-resolution anatomical images and fMRI in the resting-state. Subjects were healthy young adults, both male (4) and female (2).

Resting-state fMRI

Data were acquired on a 3T Magnetom TIM Trio scanner (Siemens, Erlangen, Germany). A phased array head coil was used, with 32 channels employed for resting-state scans. Echo-planar imaging was used to measure blood oxygenation level-dependent (BOLD) changes in image intensity. Resting-state functional MRI was obtained in 8 runs of 8.5 minutes (68 minutes of data in total) from each of the subjects. Note the very long data acquisition for each subject, compared to more common resting state run lengths. This long duration ensures high signal to noise ratio and convergence of resting-state FC (Birn et al., 2013; Dawson et al., 2013). In order to check for sleep, we communicated with the subject at the end of each scan and they were all immediately responsive. Each run consisted of 256 contiguous EPI whole-brain functional volumes [repetition time (TR) = 2000 ms; echo time (TE) = 30 ms; flip angle = 90°; 30 slices; matrix = 112 × 112; field of view (FOV) = 224 mm; acquisition voxel size = 2 × 2 × 2.2 mm].

Resting state runs were preprocessed using FSL's FEAT software package (Jenkinson et al., 2012). This deletes the first 3 volumes, performs motion correction, high pass filtering (0.01 Hz), slice time correction, and registration to the T1-weighted (MPRAGE) anatomical volume from the retinotopy session (see below). In terms of denoising, a cerebrospinal fluid (CSF) mask and white matter (WM) mask were created by segmenting the resting state anatomical image using FSL's FAST software package (Zhang et al., 2001). The masks were then eroded by one voxel, thresholded (80% tissue type probability), and registered to the preprocessed functional image. The average time courses of the CSF and WM masks were obtained and the two time courses were regressed out of the functional image time-series.

Retinotopic mapping and regions of interest for functional connectivity analysis

This study applied new analyses to data from one of our recently published articles (Dawson et al., 2013). Detailed descriptions of methods pertaining to the retinotopy data acquisition and preprocessing can be found therein.

Retinotopic mapping was pursued with axial-oblique slices (repetition time (TR)=2000 ms; echo time (TE)=30 ms; flip angle=90°; 28 slices; matrix=128×128; field of view (FOV)=263

mm; acquisition voxel size=2.05×2.05×3 mm). Retinotopy runs were motion corrected and slice time corrected using Freesurfer (http://surfer.nmr.mgh.harvard.edu/).

Figure 1 presents methodology related to the definition of ROIs and the network under investigation. We evaluated a local visual network defined within a hemisphere, which is subdivided into ventral and dorsal sections. Data were obtained from the two hemispheres of each subject, thus from 4 quadrants of the visual cortex (cortical regions which respond to stimuli in the 4 quadrants of the visual field). To achieve this, we measured the subject specific retinotopic response evoked by rotating wedge and expanding/contracting ring stimuli. Phase encoding analysis was pursued in order to define visual areas V1, V2 and V3. The area-specific ROIs were defined with a gap between visual areas V1, V2, and V3, by means of selecting voxels with polar angle within ±35° of the oblique meridian in each quadrant of the visual field. This left a 20° gap in polar angle between the midpoint polar angle of each included voxel and that of the closest possible voxel in another visual area across the inter-areal border. This gap minimized the possibility of mixing signals with origin in two different visual areas, which can be detrimental to estimation of FC (Smith et al., 2011).

The visual areas in each quadrant were further subdivided into five eccentricity regions of approximately equal size based on the numbers of their voxels. The eccentricity boundaries for the quadrant based regions of interest were 0.5° to 1.7°, 1.7° to 4.1°, 4.1° to 7.1°, 7.1° to 10.4°, and 10.4° to 14°. This definition of ROIs allows us to observe fine spatial specificity of functional connectivity measures within our network.

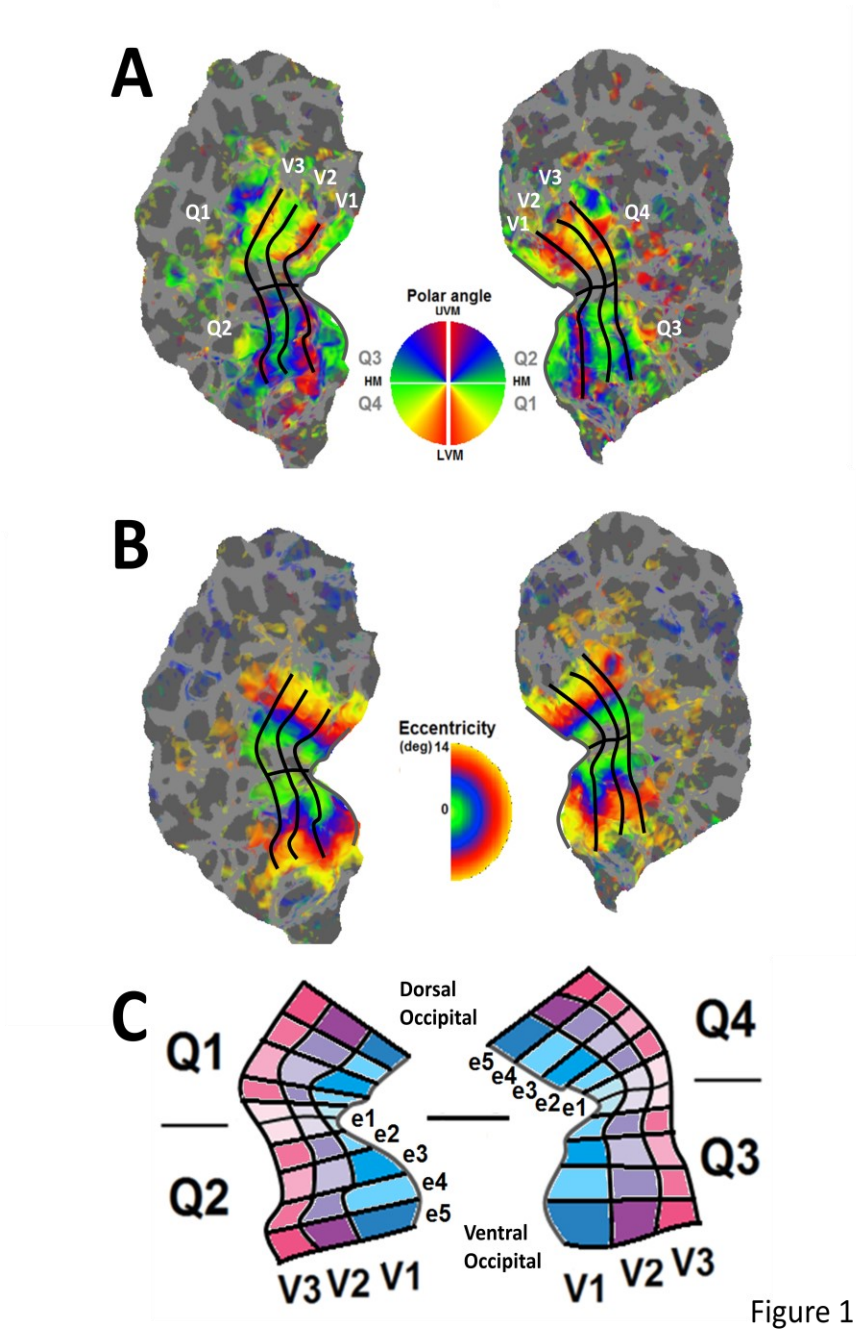


Figure 1

Figure 1. Retinotopy and ROI definition. Data presented are from one subject. Visual areas V1, V2 and V3 were defined by examining the reversal of direction of increasing polar angle phases perpendicular to the iso-eccentricity delineation. Next, each hemisphere was divided into dorsal and ventral sections corresponding to quadrants of visual cortex. In A, we see the functional response to a stimulus changing in terms of polar angle, and in B, we see the response to a stimulus changing in terms of eccentricity. UVM=upper vertical meridian, HM=horizontal meridian, LVM=lower vertical meridian. C) Voxels from each visual area in a quadrant were further classified into five eccentricity bins: e1 (central), e2, e3, e4, and e5 (peripheral), with the boundaries 0.5° to 1.7°, 4.1° to 7.1°, and 10.4° to 14°, respectively.

Functional connectivity analysis

The resting-state fMRI time-series were averaged over all voxels in each ROI to pursue ROI-based FC. Resting-state FC measures were obtained for each run by computing the Corr and Pcorr coefficients between the time-courses of each pair of ROIs. In this case, Corr refers to Pearson's linear correlation and Pcorr refers to Pearson's linear correlation between time-course residuals from two ROIs which have the time-courses from all other ROIs in the network regressed out. Since in the current study we were interested in the FC within the small network of retinotopic visual areas, we did not regress out the global signal average. Analyses were based on a network configuration of 60 ROIs (2 hemispheres × 2 quadrants within a hemisphere × 3 visual areas × 5 eccentricity-defined ROIs). Using the FC measures obtained from this network, we analyzed (1) within-quadrant FC, with the results from four quadrants (dorsal and ventral quadrants from each hemisphere) analysed separately and their respective results combined statistically, and (2) between-quadrant, within-hemisphere, FC with the results from the two hemispheres analysed separately and their respective results combined statistically.

Statistics

FC was quantified within our network of 60 nodes by computing matrices (each with 60×60 entries) of Corr and Pcorr coefficients for each of the 8 data sets available from each subject (resulting in 8 coefficients per subject for each ROI pair). To examine statistical significance, coefficients were tested against zero, and separately against a time-shuffled surrogate distribution (Supplementary Figures 1-4), and a Fourier surrogate distribution (Supplementary Figures 5-8). Time shuffled surrogate data were generated by computing the Corr/Pcorr between ROI time-courses shuffled in time. Shuffling was pursued based on the indices of the 8 scans, while keeping the time-course from each scan unchanged; we computed functional connectivity measures for the maximal possible number of permutations: $([8 \times 8] - 8)/2 = 28$. Thus for each subject, 28 surrogate coefficients were computed for each ROI pair. To create the Fourier surrogate, we extracted the amplitude and phase values from the Fourier transform of the original time-series, shuffled the original phases, assigned these newly shuffled phase values to the amplitudes of the original Fourier transforms, and then computed the inverse Fourier transform.

The results obtained when comparing FC measures to zero (Figures 4-8) and to the surrogate data (Suppl. Figures) were consistent.

Fisher's z transform was applied to individual Corr/Pcorr coefficients (r) prior to use in statistical tests by calculating:

$$z = \frac{1}{2} \ln \left(\frac{1+r}{1-r}\right)$$

The Fisher transformed coefficients were then submitted to a mixed effects, repeated measures General Linear Model (GLM) obtained from the SurfStat MATLAB toolbox (http://www.math.mcgill.ca/keith/surfstat/#mixed). Within a quadrant, 32 ([8 functional runs of data] x [4 quadrants]) Fisher-transformed Corr and Pcorr coefficients were computed for all pairs of ROIs for each subject, and were treated as separate data acquired for the given subject. For the within-hemisphere, between-quadrant statistical tests, we had 16 ([8 functional runs of data] x [2 pairs of quadrants]) sets of Fisher-transformed Corr and Pcorr coefficients for all pairs of ROIs for each subject.

A *t* statistic, degrees of freedom and p-value for the contrast of interest, namely the contrast between the Corr/Pcorr coefficients and zero, were generated by the GLM analysis for each pair of ROIs. Using an alpha of 0.05, p-values were assessed for significance after applying the False Discovery Rate (FDR) correction (Bejamini and Hochberg, 1995). In all FC figures, we present statistical results for both Corr and Pcorr results. Considering the FDR correction, the adjusted alpha for the 2 tests (within quadrant and between quadrants) were 10^-16 and 10^-16 for Corr and 0.0077 and 0.0163 for Pcorr, respectively.

In both the within and between quadrant analyses, we statistically compared Corr/Pcorr values as a function of eccentricity bin difference (Figures 6 and 8). This analysis was performed according to the number of eccentricity regions separating the two ROIs being correlated (see Figure 2 for a schematic showing these connections). With 5 eccentricity regions, we could have 0 (if the ROIs are in different visual areas), 1, 2, 3, or 4 eccentricity bin separations between ROIs (we denote these eccentricity bin separations as 0es, 1es, 2es, 3es, and 4es respectively, "es" signifying "Eccentricity Separation"). A 0es separation indicates that the two ROIs whose connectivity is being evaluated share the same eccentricity bin, i.e. they are both e1, both e2, both e3, etc. A 1es separation indicates that the ROIs are adjacent in eccentricity, i.e. e1 and e2, e2 and e3, etc. A 2es separation indicates that the ROIs are separated by a gap the width of one

eccentricity region, i.e. e1 and e3, e2 and e4, or e3 and e5. A 3es separation indicates that the ROIs are separated by a gap the width of two eccentricity regions, i.e. e1 and e4, or e2 and e5. Finally, a 4e separation means a connection between the most central and most peripheral regions, i.e. e1 and e5. Within a visual area there are zero 0es connections within a quadrant and five 0es connections between quadrants, four 1es connections, three 2es connections, two 3es connections, and one 4es connection. Between visual areas there are five 0es connections, eight 1es connections, six 2es connections, four 3es connections, and two 4es connections. These analyses were also done using the repeated measures GLM, though p-values were not adjusted for multiple comparisons here as only a few comparisons were made.

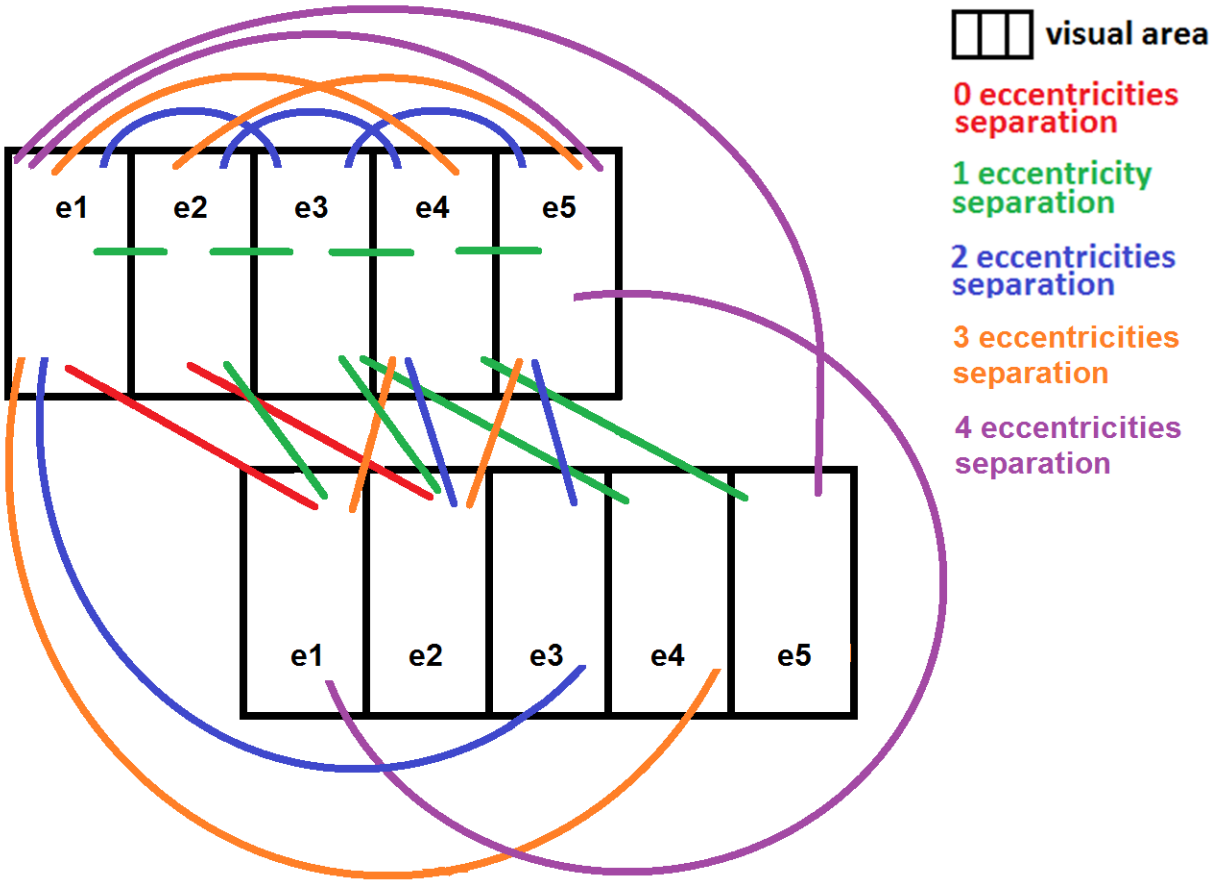


Figure 2. Schematic of separations between ROIs within and between visual areas. For statistical testing, ROIs were grouped according to the number of eccentricity regions separating them. Note that these are a few representative connections. For the total number of connections of each type, see the Methods section.

Cortical distance calculations

Cortical distance was estimated using Freesurfer's "mris_pmake" function along the smoothened surface of the cortex. For each subject, central voxels within each ROI were estimated and then cortical distances between these centroids were computed. Distances were computed for all ROI combinations within the hemisphere, these being within and between quadrants. Figure 3 presents the cortical distances averaged across hemispheres and subjects.

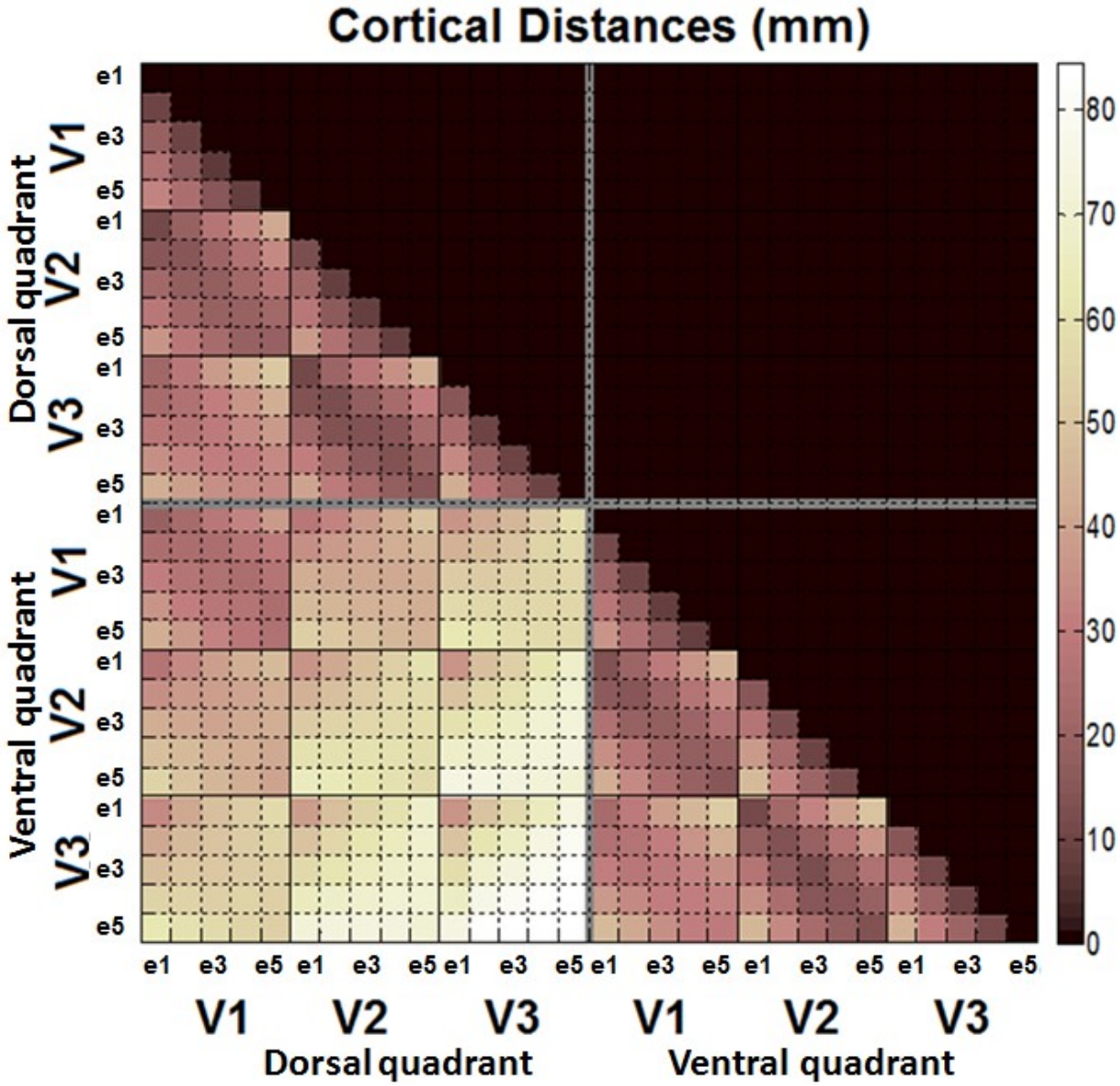


Figure 3. Distances along the surface of the cortex between pairs of ROIs. Distances were computed within a hemisphere using the quadrant division of ROIs, such that there are two full quadrants considered per hemisphere, one dorsal and one ventral. The matrix presents mean distances computed across all five subjects and their two hemispheres. The defined eccentricity regions are labelled as e1, e2, e3, e4, and e5.

The relationship between the Corr/Pcorr coefficients and the cortical distance between ROIs was examined with first and second order regressions, as well as an exponential fit. These functions were fit to the mean of the Corr/Pcorr values and cortical distances across subjects and hemispheres. The linear, quadratic, and exponential trend lines were computed using Matlab's *fit* function. Goodness of fit of the models was evaluated using a Bayesian Information Criterion (BIC) measure:

$$BIC = n * \ln(error\ variance) + v * \ln(n)$$

Where n is the number of data points, v is the degrees of freedom of the model, and the error variance is:

$$error\ variance = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

Where y_i is the mean Corr/Pcorr for a given cortical distance, x_i; and $f(x_i)$ is the model prediction at that cortical distance.

This definition of the BIC assumes that the model errors are independent and normally distributed and that the derivative of the log likelihood with respect to the true variance is zero. When comparing several models, the model with the lowest BIC is expected to best represent the data (Priestley, 1981).

In order to quantify the confidence one can have in the model selected based on the BIC values, we computed the Bayes Factor (Sotero et al., 2009):

$$BayesFactor_i = e^{-0.5\,(BIC_i - BIC_0)}$$

Where the subscript "i" represents the different models, and the subscript "0" represents the reference model, here selected as the model with the lowest BIC. The Bayes Factor allows us to compute the percentage confidence we can have in the fact that the model with the lowest BIC fits better than the other models:

$$\%\ confidence_i = 100\% * (1 - BayesFactor_i)$$

Results

Below we present the results from grouped quadrant analyses. First, for within quadrant analyses, we consider all four quadrants as separate data sets for a given subject. Next, for

between quadrant analyses, we consider pairs of quadrants from the same hemisphere as separate data sets for a given subject. For the results of the full visual cortex analysis see Figure 4.

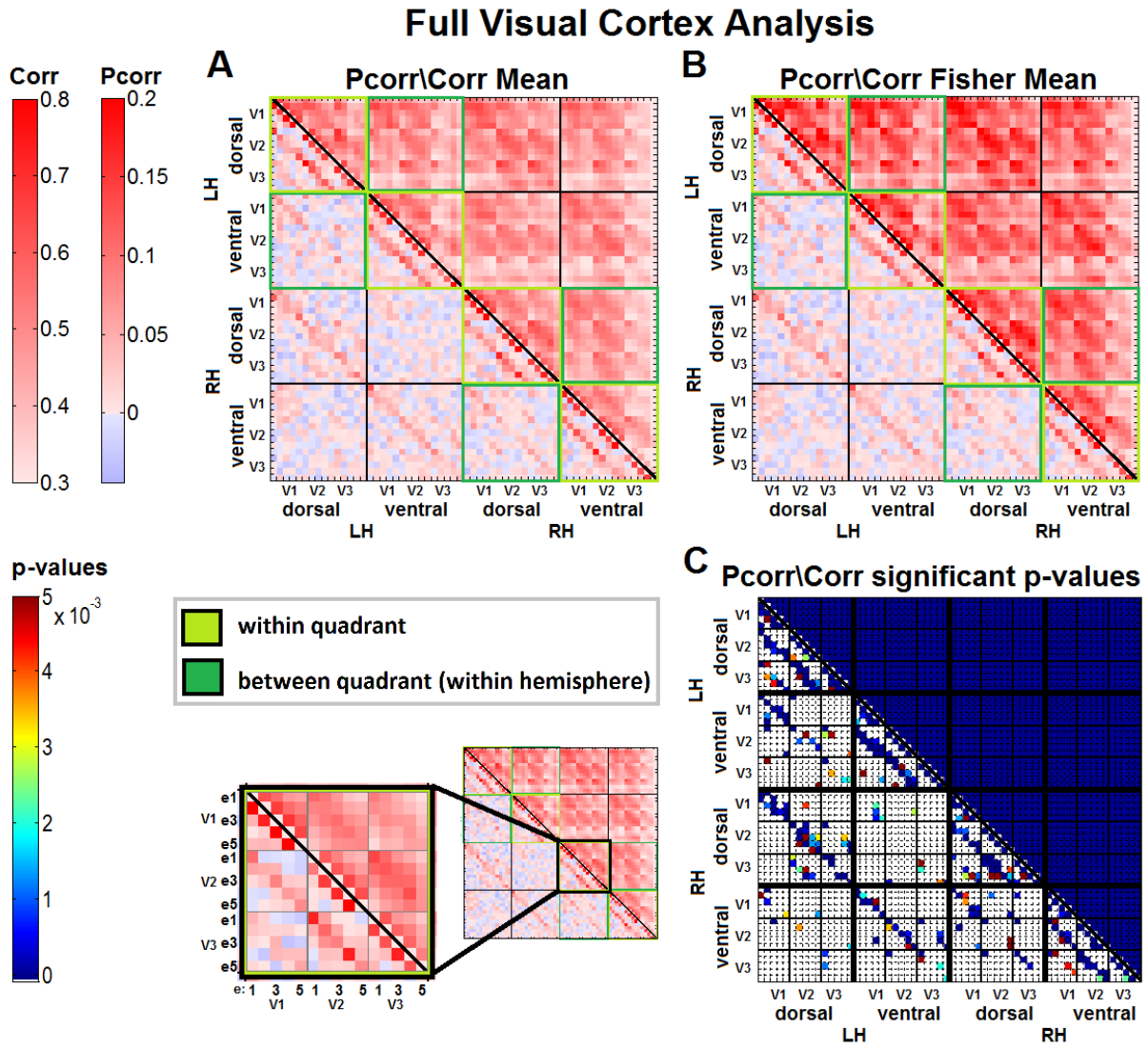


Figure 4. Mean Pcorr/Corr, Fisher coefficients, and p-values associated with functional connections of the full visual cortex network (all four quadrants analyzed together). In each matrix, Pcorr coefficients / p values are presented in the lower left triangle and Corr coefficients / p values are in the upper right triangle. A) Mean Pcorr / Corr coefficients across all subjects (6) and runs per subjects (8) (6*8). B) Mean Fisher coefficients across all subjects and runs. C) P values of statistically significant Pcorr / Corr associated with the functional connections in the full visual cortex network. The five defined eccentricity regions are labelled as e1, e2, e3, e4, and e5. Significance is thresholded at 0.05, with false discovery rate (FDR) correction. Connections that did not show statistical significance are presented in white. LH=left hemisphere; RH=right hemisphere.

Retinotopic organization of functional connectivity: within quadrant analysis

FC within and between areas in the human visual cortex including V1, V2, and V3 was measured by calculating the Corr and, separately, Pcorr coefficients between BOLD fMRI time courses acquired in the resting state. Each area was divided into four quadrants (regions responding to stimuli in the quadrants of the visual field). To observe the retinotopic organization of FC in these early visual areas, voxels were classified into 5 retinotopic ROIs based on the retinotopic phase encoding in eccentricity (see Methods for details).

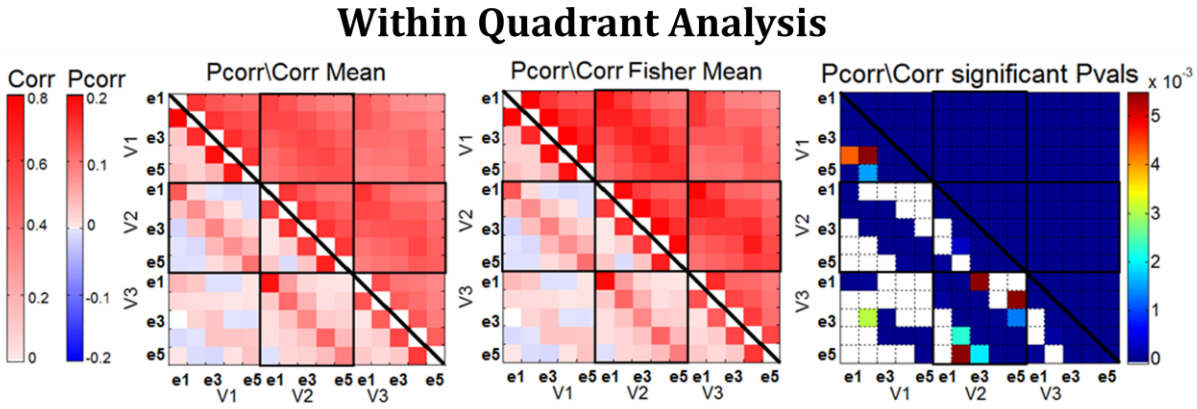


Figure 5. Mean functional connectivity measures within a quadrant, computed across all subjects, runs per subject, and the four quadrants. Mean Pcorr and Corr connectivity measures are reflected below and above the diagonal, respectively. White connections (far right matrix) did not show statistical significance. Presentation is as in Figure 4.

Functional connectivity within an area

In the within quadrant results (Fig. 5), the Corr/Pcorr between two nearby regions in eccentricity were typically larger than those between more distant regions. This pattern of retinotopic organization of FC based on eccentricity is consistently observed throughout the three visual areas, and in both Corr and Pcorr analyses. Pcorr between regions adjacent within a visual area and those along the diagonal between adjacent visual areas stand out as being among the strongest.

In order to determine which pairs of regions take part in the resting-state functional network, Fisher transformed Corr and Pcorr coefficients (in separate tests) were input to a repeated measures GLM, these values tested against zero. Figure 5 (right) shows the results of statistical

testing of the within quadrant connections, with an alpha of 0.05, corrected for multiple comparisons using the FDR correction to 10^-16 for Corr and 0.0086 for Pcorr. All Corr connections within a quadrant were significantly different from zero. Most within-area Pcorr connections showed statistical significance, and all connections between same eccentricity regions between visual areas showed statistical significance. Between adjacent visual areas, all connections reaching to regions with adjacent eccentricities were also significant. Thus, Pcorr based FC between visual areas reflected retinotopic organization.

To quantify the effect of retinotopic distance on the Corr/Pcorr measures, we pursued repeated measures GLM statistical testing, while grouping ROI pairs which were separated by the same number of eccentricity regions. Within visual areas (Figure 6A), all Corr-based connection strengths were different than zero (all $p<10^{-16}$). There was a statistically significant effect of decreasing Corr values with increasing eccentricity separation in V1, V2 and V3 as demonstrated in Figure 6B via differences in eccentricity separations (Figure 6B above diagonal; Table 1 presents the corresponding p-values). In Pcorr-based functional connections, nearly all eccentricity separations were found to have a strength that was significantly different than zero as well (with the exception of V2 3es; see p-values in Table 1). There was a statistically significant effect of decreasing Pcorr measures with increasing eccentricity separation though the drop off is more abrupt than in correlation: all adjacent eccentricity connections were found to be significantly different from the connections between non-adjacent ROIs (Figure 6B below diagonal; Table 1).

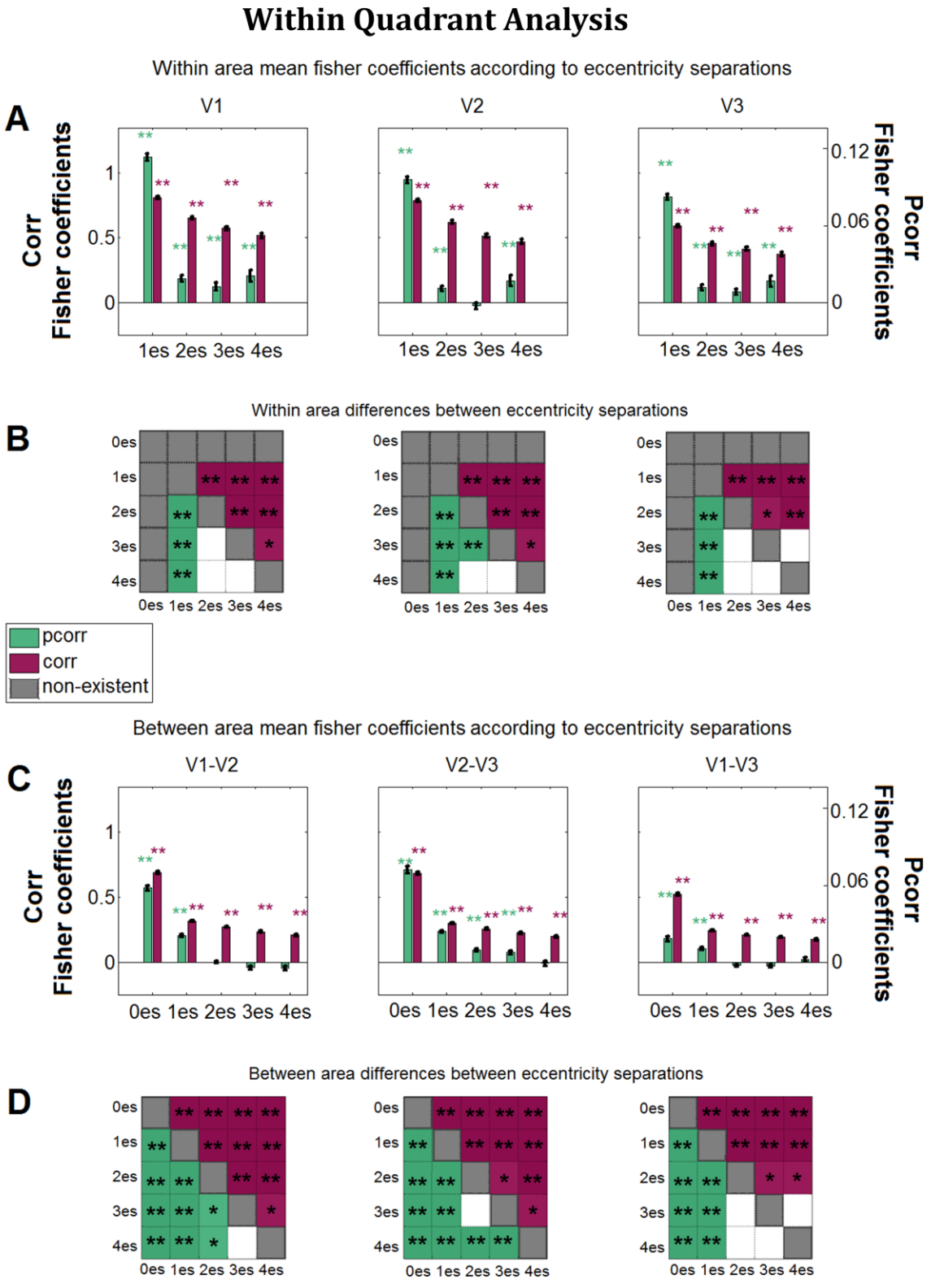

Within Quadrant Analysis
Within area mean fisher coefficients according to eccentricity separations
A
V1
V2
V3
Corr Fisher coefficients
Pcorr Fisher coefficients
1es 2es 3es 4es
Within area differences between eccentricity separations
B
0es 1es 2es 3es 4es
pcorr
corr
non-existent
Between area mean fisher coefficients according to eccentricity separations
C
V1-V2
V2-V3
V1-V3
Between area differences between eccentricity separations
D

Figure 6. Mean functional connectivity measures for connections grouped according to eccentricity separations within a quadrant. A & C) Across all plots, green and purple bars represent the mean Fisher transformed Pcorr and Corr, respectively, computed across all subjects, runs and quadrants. Error bars reflect the standard error of the mean across measurements. A shows within-visual area Fisher coefficients for different eccentricity separations. Plots in C show the between-area Fisher coefficients of different eccentricity separations. A zero eccentricity separation (0es) indicates regions with the same eccentricity; a separation of one eccentricity (1es) means the regions' eccentricities were adjacent, and a separation of two eccentricities (2es) means there was a gap between the two regions' eccentricities, etc. (See methods for further details). Individual connections are shown to be significant with an asterisk directly above the bar. B & D) Differences between eccentricity separation bins are shown to be significant with an asterisk within the colored squares of the matrices. Green and purple squares represent Pcorr and Corr, respectively. White squares represent differences which were not significant. Grey squares represent differences not testable (non-existent) given the analysis. For A, B, C and D statistical significance was assessed using alpha values of 0.05 (*) and 0.001 (**), following repeated measures statistical analyses.

Table 1: Within quadrant significant p-values in eccentricity bin analyses (associated with Figure 6).

WITHIN QUADRANT SIGNIFICANT *p* VALUES					
	Partial Correlation			Correlation	
Visual areas	Eccentricity bin coeffs > zero	Eccentricity bin differences		Eccentricity bin coeffs > zero	Eccentricity bin differences
V1-V1	1es p<10^-16 2es p=1.0x10^-14 3es p=1.7x10^-5 4es p=4.1x10^-7	1es>2es p<10^-16 1es>3es p<10^-16 1es>4es p<10^-16		All p<10^-16	1es>2es p<10^-16 1es>3es p<10^-16 1es>4es p<10^-16 2es>3es p=1.4x10^-6 2es>4es p=1.8x10^-10 3es>4es p=0.0059
V2-V2	1es p<10^-16 2es p=2.9x10^-7 4es p=1.9x10^-5	1es>2es p<10^-16 1es>3es p<10^-16 1es>4es p<10^-16	2es>3es p=3.4x10^-5	All p<10^-16	1es>2es p<10^-16 1es>3es p<10^-16 1es>4es p<10^-16 2es>3es p=1.3x10^-10 2es>4es p=5.5x10^-13 3es>4es p=0.0273
V3-V3	1es p<10^-16 2es p=1.4x10^-8 3es p=5.8x10^-4 4es p=3.1x10^-5	1es>2es p<10^-16 1es>3es p<10^-16 1es>4es p<10^-16		All p<10^-16	1es>2es p<10^-16 1es>3es p<10^-16 1es>4es p<10^-16 2es>3es p=0.0135 2es>4es p=1.6x10^-4
V1-V2	0es p<10^-16 1es p<10^-16	0es>1es p<10^-16 0es>2es p<10^-16 0es>3es p<10^-16 0es>4es p<10^-16	1es>2es p<10^-16 1es>3es p<10^-16 1es>4es p<10^-16 2es>3es p=0.0081 2es>4es p=0.0242	All p<10^-16	0es>1es p<10^-16 0es>2es p<10^-16 0es>3es p<10^-16 0es>4es p<10^-16 1es>2es p=6.0x10^-6 1es>3es p=2.2x10^-13 1es>4es p=1.3x10^-13 2es>3es p=4.4x10^-4 2es>4es p=1.3x10^-6 3es>4es p=0.0310
V2-V3	0es p<10^-16 1es p<10^-16 2es p<10^-16 3es p=4.4x10^-9	0es>1es p<10^-16 0es>2es p<10^-16 0es>3es p<10^-16 0es>4es p<10^-16	1es>2es p<10^-16 1es>3es p<10^-16 1es>4es p<10^-16 2es>4es p=9.9x10^-7 3es>4es p=1.2x10^-4	All p<10^-16	0es>1es p<10^-16 0es>2es p<10^-16 0es>3es p<10^-16 0es>4es p<10^-16 1es>2es p=8.0x10^-6 1es>3es p=4.9x10^-12 1es>4es p=5.3x10^-13 2es>3es p=0.0018 2es>4es p=3.0x10^-6 3es>4es p=0.0221
V1-V3	0es p<10^-16 1es p<10^-16	0es>1es p=9.3x10^-5 0es>2es p<10^-16 0es>3es p<10^-16 0es>4es p=4.1x10^-10	1es>2es p<10^-16 1es>3es p=2.2x10^-16 1es>4es p=1.2x10^-4	All p<10^-16	0es>1es p<10^-16 0es>2es p<10^-16 0es>3es p<10^-16 0es>4es p<10^-16 1es>2es p=9.0x10^-5 1es>3es p=5.3x10^-8 1es>4es p=1.8x10^-8 2es>3es p=0.0377 2es>4es p=7.4x10^-4

The first column indicates the visual areas involved, including within area measures (first 3 rows) and between areas measures (next 3 rows). The second column presents the results of testing for statistical significance of Pcorr measures as a function of eccentricity bin difference. The third column presents the results of testing for differences in PCorr values with increasing eccentricity separation. The fourth and fifth columns present results for Corr measures.

Functional connectivity between areas

Retinotopic organization of FC was also observed between visual areas upon examination of the Corr and Pcorr connectivity matrices (Figure 5, first two matrices). Retinotopically close pairs of ROIs from different visual areas show higher Corr/Pcor values than distant pairs of ROIs. For example, common eccentricity regions tend to be connected with the highest Corr/Pcorr values, particularly between V1 and V2, and V2 and V3.

While assessing the strength of connection between two ROIs, we found that Pcorr between adjacent visual areas (V1-V2, V2-V3) significantly differed from zero for all common eccentricity connections (all $p<10^{-16}$) and all but one adjacent eccentricity connections (all $p<4x10^{-5}$ (Figure 5, right). V1 and V3 had 4 out of 5 same eccentricity region significant Pcorr as well as a few adjacent eccentricity connections, though not all. These findings suggested that the functional connections between adjacent visual areas (V1-V2 and V2-V3) are stronger than those between non-adjacent areas (V1-V3). Corr/Pcorr values between regions from adjacent visual areas were indeed significantly higher than Corr/Pcorr from V1-V3 connections ($p<10^{-16}$ for Corr and Pcorr; two-tailed paired t-test).

In evaluating the range of Corr/Pcor along the eccentricity dimension between visual areas we turn our attention to Figure 6C. The between area eccentricity bin separation plots show consistent significance of correlations for all eccentricity separations (significantly different from zero; p-values, all $p<10^{-16}$, are presented in Table 1). There were significant differences between the magnitudes of the coefficients for all the different eccentricity bin separations (except for 3es compared to 4es between V1 and V3): the Corr coefficients decreased with increasing eccentricity separation (Figure 6D above diagonal, Table 1).

The between area plots for the Pcorr analysis show consistent significance with 0es and 1es (significantly different from zero; Figure 6C; p-values presented in Table 1). Connections with 2es and 3es between adjacent visual areas V2 and V3 showed statistical significance too. In all cases, connections between ROIs separated by 4 eccentricity bins did not show statistically significant Pcorr. We observed a statistically significant effect of decreasing Pcorr coefficients as the eccentricity separation increased, particularly in adjacent visual areas (Figure 6D below diagonal, Table 1). These results lead us to conclude that Pcorr based FC between visual areas is largely limited to identical and adjacent eccentricity regions, whereas Corr measures are not.

See Figure 8 for a summary of the connections identified within a quadrant based on the Pcorr eccentricity bin separation analyses.

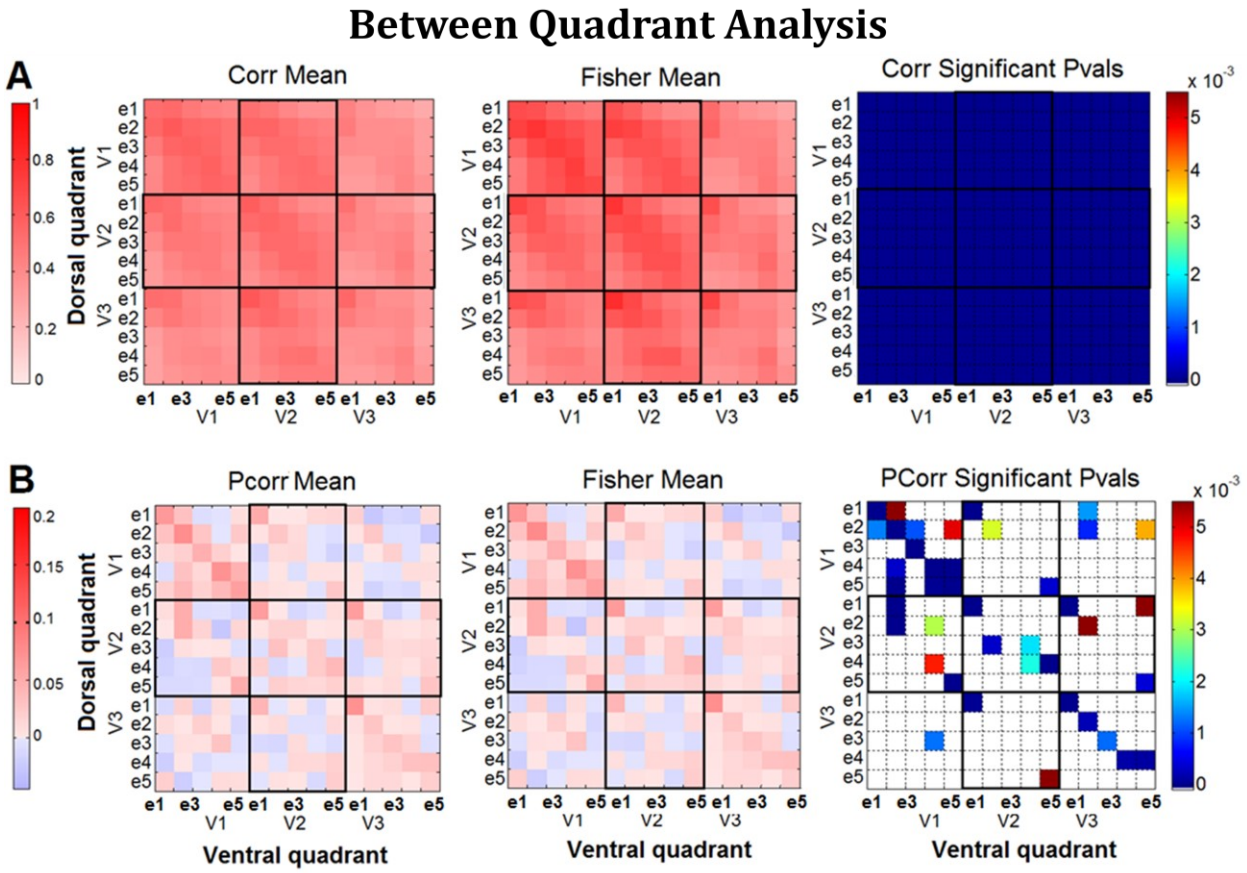


Figure 7. Mean functional connectivity measures associated with functional connections between ROIs residing in different quadrants within a hemisphere. The mean connectivity measures were computed across all subjects, runs, and the two quadrant pairs. A & B reflect Corr and Pcorr connectivity measures, respectively. White entries (far right matrices) represent connections that did not show statistical significance. The format of presentation is as in Figure 4.

Retinotopic organization of functional connectivity: between-quadrant within-hemisphere analysis

Functional connectivity within an area

Figure 7 (A and B) show the Corr and Pcorr coefficients for within-hemisphere, between-quadrant connectivity. The Corr analysis (Figure 7A) shows that e1-e1 connections, and to a lesser extent connections between other retinotopically close regions within a visual area (close

to the diagonal) stand out with higher correlations. Pcorr matrices (Figure 7B) also reflect stronger e1-e1 connections, though less noticeably so than Corr. They also show relatively high Pcorr between other regions of the same eccentricity (e2-e2, e3-e3) within a visual area (along the diagonal).

Figures 7A and 7B (far right matrices) show the results of between quadrant statistical tests against zero, with an alpha of 0.05, FDR corrected to 10^-16 for Corr and 0.0086 for Pcorr. All Corr-based functional connectivities (Figure 7A, far right) were significantly different from zero. The Pcorr statistical analysis (Figure 7B, far right) shows that 11 out of 15 connections along the diagonal, representing connections between regions of same eccentricity belonging to the same visual area have partial-correlations significantly greater than zero. In addition, several within-area adjacent-eccentricity region connections show statistically significant Pcorr, particularly in V1. V1 also shows three other significant connections: dorsal e4 – ventral e2, dorsal e5 – ventral e2, and dorsal e2 – ventral e5. From this we concluded that the range of Pcorr based connectivity within a visual area between dorsal and ventral quadrants of a hemisphere is at most to the adjacent eccentricity region in V2 and V3, unlike within the quadrant, where some non-adjacent eccentricity regions were significantly different than zero as well.

Within area correlations were statistically significant for all eccentricity separations for V1, V2 and V3 (Figure 8A; all p-values for between quadrant eccentricity bin analyses are presented in Table 2). In order to test the effect of retinotopic distance within a visual area between quadrants of a hemisphere, we analyzed the difference between the various eccentricity bin fisher transformed correlations. There were significant differences between nearly all eccentricity bin separations in V1, V2, and V3; the only separations that did not show a statistically significant difference were 2es and 3es, and 3es and 4es in V3 (Figure 8B, Table 2). We concluded that between quadrants but within a visual area, Corr coefficients differ in a similar manner to within a quadrant, such that there is a notable decreasing effect in coefficients with increasing retinotopic distance.

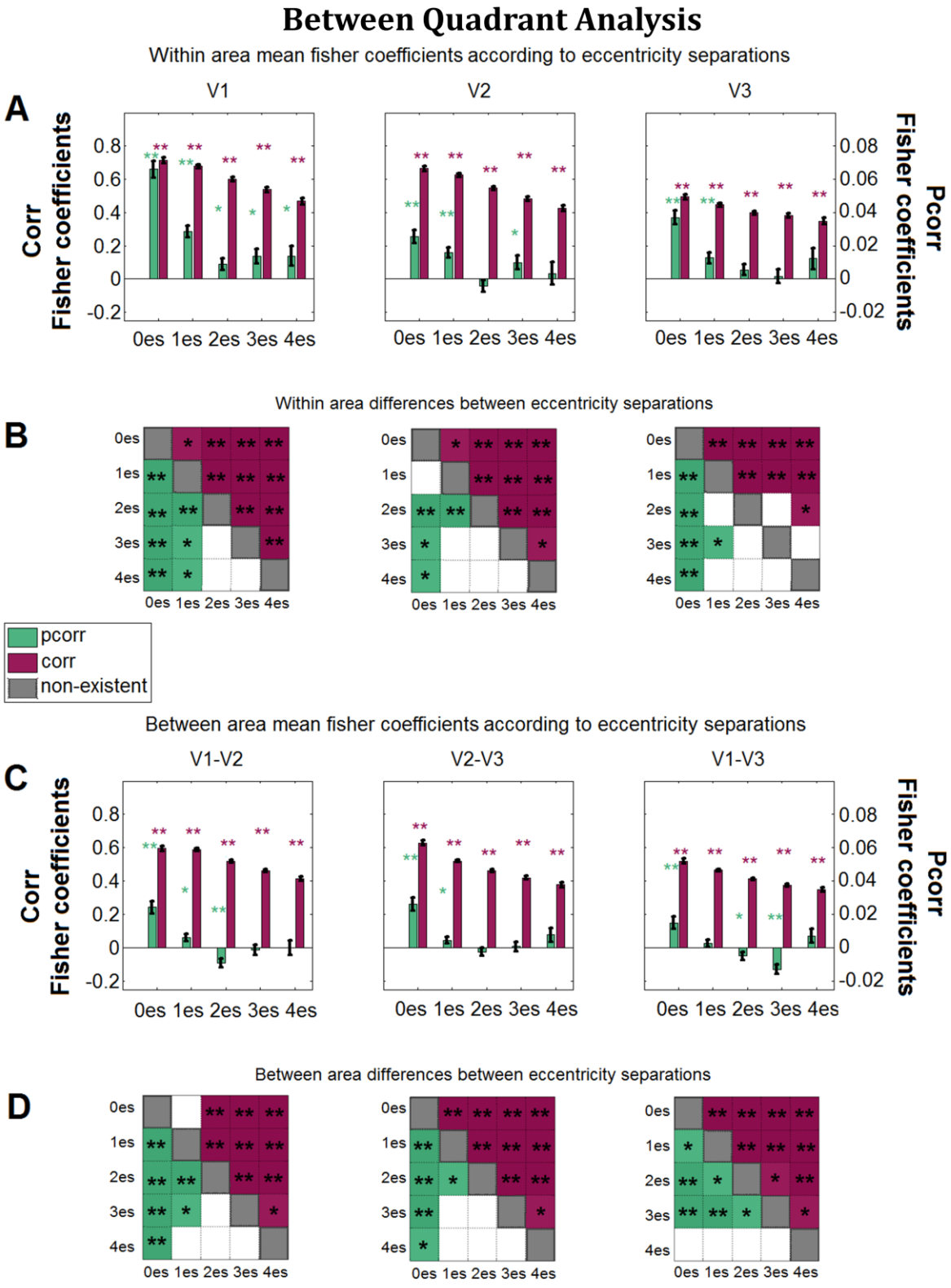

Between Quadrant Analysis
Within area mean fisher coefficients according to eccentricity separations
A
V1
V2
V3
Corr Fisher coefficients
Pcorr Fisher coefficients
0es 1es 2es 3es 4es
Within area differences between eccentricity separations
B
pcorr
corr
non-existent
Between area mean fisher coefficients according to eccentricity separations
C
V1-V2
V2-V3
V1-V3
Between area differences between eccentricity separations
D

Figure 8. Mean functional connectivity measures for connections grouped according to eccentricity separations between quadrants within a hemisphere. Across all plots, green and purple bars represent the mean Fisher transformed Pcorr and Corr, respectively, computed across all subjects, runs and hemispheres. Error bars reflect the standard error of the mean across measurements. Green and purple entries in B and D represent significant differences in Pcorr and Corr eccentricity bin separations (es), respectively. For A, B, C and D statistical significance was assessed using alpha values of 0.05 (*) and 0.001 (**), following repeated measures statistical analyses. Presentation is as in Figure 6.

Table 2: Between quadrant significant p-values in eccentricity bin analyses (associated with Figure 8).

BETWEEN QUADRANT SIGNIFICANT *p* VALUES						
	Partial Correlation			Correlation		
Visual areas	Eccentricity bin coeffs > zero	Eccentricity bin differences		Eccentricity bin coeffs > zero	Eccentricity bin differences	
V1-V1	0es p<10^-16 1es p<10^-16 2es p=0.0141 3es p=0.0026 4es p=0.0175	0es>1es p=2.0x10^-10 0es>2es p<10^-16 0es>3es p=8.6x10^-14 0es>4es p=5.0x10^-9	1es>2es p=9.1x10^-5 1es>3es p=0.0096 1es>4es p=0.0451	All p<10^-16	0es>1es p=0.0297 0es>2es p=2.8x10^-11 0es>3es p<10^-16 0es>4es p<10^-16	1es>2es p=3.0x10^-7 1es>3es p<10^-16 1es>4es p<10^-16 2es>3es p=1.7x10^-4 2es>4es p=2.2x10^-10 3es>4es p=8.3x10^-4
V2-V2	0es p=4.0x10^-10 1es p=7.4x10^-8 3es p=0.0145	0es>2es p=1.6x10^-8 0es>3es p=0.0073 0es>4es p=0.0041	1es>2es p=7.0x10^-6	All p<10^-16	0es>1es p=0.0067 0es>2es p=3.4x10^-14 0es>3es p<10^-16 0es>4es p<10^-16	1es>2es p=2.7x10^-9 1es>3es p<10^-16 1es>4es p<10^-16 2es>3es p=2.6x10^-5 2es>4es p=7.3x10^-10 3es>4es p=0.0052
V3-V3	0es p<10^-16 1es p=4.6x10^-5	0es>1es p=1.1x10^-6 0es>2es p=6.9x10^-10 0es>3es p=9.7x10^-10 0es>4es p=7.6x10^-4	1es>3es p=0.0371	All p<10^-16	0es>1es p=9.8x10^-4 0es>2es p=3.3x10^-10 0es>3es p=8.5x10^-11 0es>4es p=9.9x10^-10	1es>2es p=8.7x10^-5 1es>3es p=4.1x10^-6 1es>4es p=3.5x10^-7 2es>4es p=0.0059
V1-V2	0es p=1.7x10^-10 1es p=0.0047 2es p=2.2x10^-4 (for 2es, negative Pcorr)	0es>1es p=4.0x10^-5 0es>2es p=1.4x10^-13 0es>3es p=1.9x10^-7 0es>4es p=2.6x10^-5	1es>2es p=3.0x10^-6 1es>3es p=0.0476	All p<10^-16	0es>2es p=3.6x10^-8 0es>3es p<10^-16 0es>4es p<10^-16	1es>2es p=8.8x10^-12 1es>3es p<10^-16 1es>4es p<10^-16 2es>3es p=4.2x10^-7 2es>4es p=7.3x10^-13 3es>4es p=0.0012
V2-V3	0es p=9.3x10^-11 1es p=0.0234	0es>1es p=4.9x10^-7 0es>2es p=2.0x10^-10 0es>3es p=1.8x10^-7 0es>4es p=0.0016	1es>2es p=0.0241	All p<10^-16	0es>1es p=4.4x10^-16 0es>2es p<10^-16 0es>3es p<10^-16 0es>4es p<10^-16	1es>2es p=9.9x10^-10 1es>3es p<10^-16 1es>4es p<10^-16 2es>3es p=1.2x10^-4 2es>4es p=1.9x10^-9 3es>4es p=0.0023
V1-V3	0es p=5.3x10^-5 2es p=0.0378 3es p=5.7x10^-6 (for 2es and 3es, negative Pcorr)	0es>1es p=0.0042 0es>2es p=5.8x10^-6 0es>3es p=3.7x10^-9	1es>2es p=0.0178 1es>3es p=1.7x10^-5 2es>3es p=0.0325	All p<10^-16	0es>1es p=8.5x10^-5 0es>2es p=1.9x10^-14 0es>3es p<10^-16 0es>4es p<10^-16	1es>2es p=5.3x10^-8 1es>3es p=3.3x10^-15 1es>4es p=1.3x10^-15 2es>3es p=0.0015 2es>4es p=4.1x10^-6 3es>4es p=0.0344

The format of presentation is similar to that used for Table 1.

Within area Pcorr values were statistically significant between common eccentricity regions (0es separation) and regions adjacent in eccentricity (1es) for V1, V2 and V3 (Figure 8A). In addition, Pcorr between more distant eccentricity regions within V1 (2es, 3es and 4es) showed statistical significance. In order to test the effect of retinotopic distance within a visual area between quadrants of a hemisphere, we analyzed the difference between the various eccentricity bin fisher transformed partial correlations. In all visual areas, same eccentricity partial

correlations were significantly higher than those with eccentricity bin separations (Figure 8B, Table 2). A similar effect was observed for differences between adjacent eccentricity and higher eccentricity bin separations consistently in V1, as well as for 1es and 2es in V2, and 1es and 3es in V3. We concluded that as retinotopic distance increases within visual areas V1, V2, and V3 across dorsal and ventral quadrants, there is an effect of decreasing resting-state Pcorr values, and most notably so in V1.

Functional connectivity between areas

Between visual areas and between quadrants of a hemisphere, the e1-e1 connections stand out with higher coefficients, particularly in the Corr coefficients (Figure 7A). As with the within quadrant analyses, connections between V1 and V3 have somewhat lower coefficients than between V1 and V2, and between V2 and V3. In a two-tailed paired t-test of correlations, the connections between adjacent visual areas, i.e. V1 and V2 together with V2 and V3, were found to have significantly higher correlations than V1-V3 connections ($p<10^{-16}$). Partial correlations were also significantly different in this between quadrants test ($p=0.0120$).

Between visual areas, all Corr values showed statistically significant differences relative to zero in both the analysis between individual ROIs (Figure 7A, far right matrix), and in the eccentricity bin analyses (Figure 8C). This being said, Corr was able to differentiate between all eccentricity bin separations in all pairs of visual areas, except V1-V2 0es and 1es (Figure 8D above diagonal, Table 2).

The findings for Pcorr are quite different than those of Corr between visual areas, between quadrants. Partial correlations that showed statistical significance relative to zero were mainly those of the same eccentricity connections between adjacent visual areas (Figure 7B, far right). V1-V2, V2-V3, and V1-V3 0es connections showed significant Pcorr (Figure 8C; Table 2). In adjacent visual areas V1-V2 and V2-V3, the 1es connections were also found significantly greater than zero. Some negative partial correlations were found to be significant between V1-V2 (2es), and V1-V3 (2es, 3es). As was observed within visual areas, between visual areas the Pcorr values in the 0es bin were found significantly different than those in the other eccentricity bins (except V1-V3 0es and 4es) (Figure 8D). There are also several differences between 1es connections and those with greater separations (all visual area combinations). This means that connections between regions with the same, or similar eccentricity are particularly strong in all

between quadrant connections, with those differing in eccentricity showing significantly smaller Pcorr values.

Cortical distances analyses

Given the observed relationship between retinotopic distance and Corr/Pcorr, and the relationship between retinotopy and distance along the cortex, we undertook an analysis comparing distances between ROI centroids along the cortex and Corr/Pcorr connectivity measures between those ROIs. We have continued with the separation of "within quadrant" and "between quadrants" connections given that we saw different connectivity profiles in these subsets of connections. With differing trends in correlations/partial correlations, it stands to reason that the relationship of cortical distance to the correlations/partial correlations may also differ.

Figure 3 shows the mean computed cortical distances within a hemisphere. The mean is across all subjects and hemispheres. The analysis showed shorter distances for connections between regions from the same quadrant, with the same eccentricity, and from adjacent visual areas (for example, within quadrant adjacent visual area distances between e1-e1: ~11-13mm, mean=11.9mm, standard deviation (STD) =0.96mm; e3-e3: ~13-18mm, mean=15.7mm, STD=2.6mm; e5-e5: ~15-19mm, mean=16.2mm, STD=1.9mm). Between quadrants, the shortest connections are between e1 and e1, with e2-e2, e3-e3, e4-e4, and e5-e5 being progressively more distant. In addition, distances between V1 ROIs belonging to the dorsal and ventral quadrants (between quadrant connections) were relatively short (<42mm). These observations follow expectations from the cortical organization of V1, V2 and V3 (see Figure 1).

Figure 9 presents plots of the mean Corr (A) and Pcorr (B) coefficients versus mean cortical distance between ROI pairs within a quadrant, the figure reflecting dorsal and ventral results together in the same plot. There is a clear trend in the within quadrant scatter plots of data, with decreasing coefficients for greater distances. The linear and quadratic regressions, as well as the exponential fit were computed for each plot and were included in the figure for reference. The linear function seems to fit best the Corr values vs. distance (Bayes factor confidence of 84.7% and 85.7% for linear fit being better than quadratic and exponential, respectively, Table 3). The exponential function fits best to the Pcorr values vs. distance (~100% confidence for exponential function better than both linear and quadratic, Table 3).

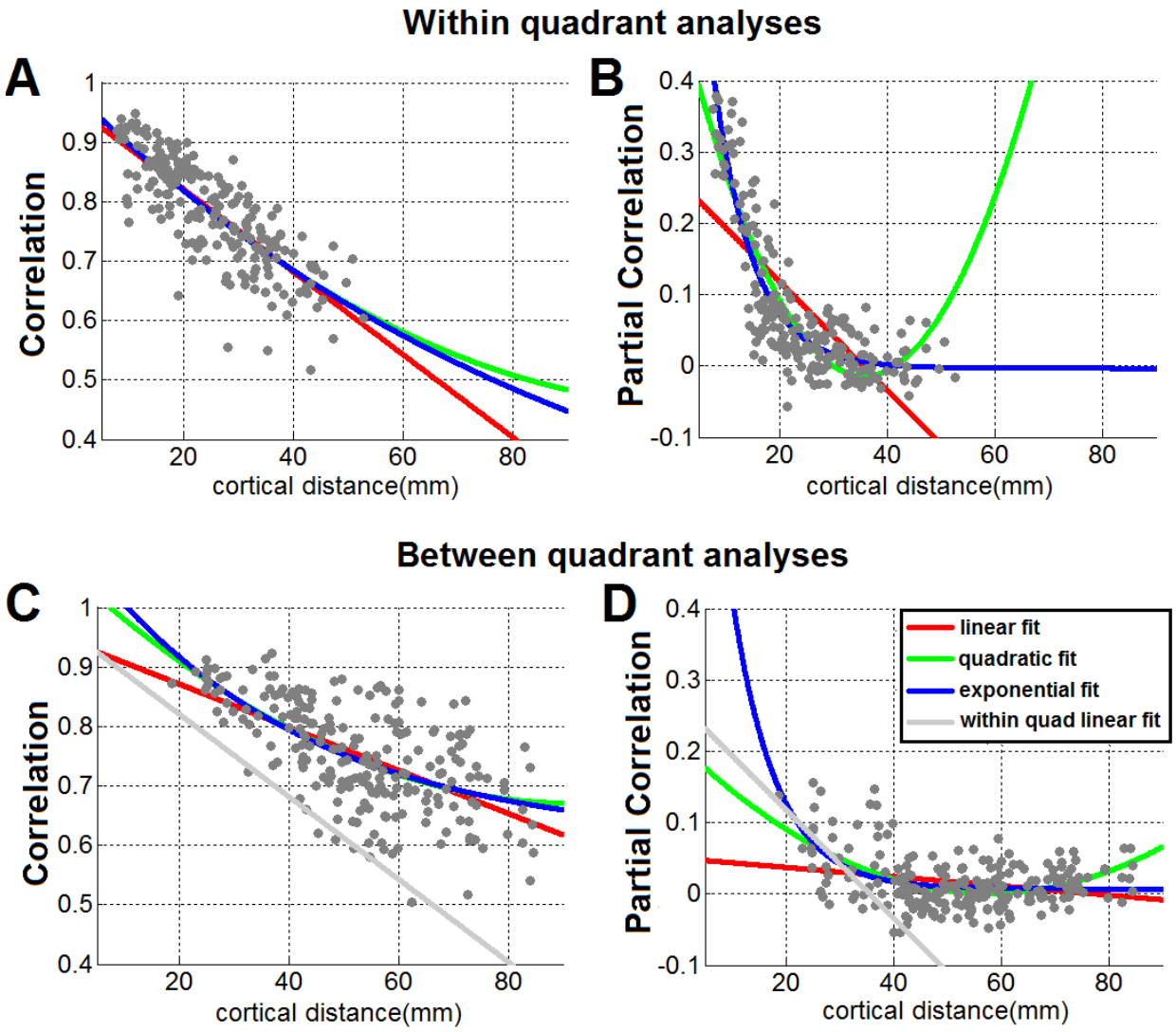


Figure 9. Mean Corr/Pcorr coefficients versus cortical distance plots for within and between quadrant connections. Within quadrant coefficients are presented on the top row (A and B) and between quadrants coefficients are on the bottom row (C and D). The mean coefficients and distances were computed across subjects and hemispheres.

If resting-state FC within a network with dense connections depends on cortical distance alone, we should expect a similar approximately linear trend between Corr and cortical distance between quadrants too. We therefore pursued similar analysis of the mean Corr (Figure 9C) and Pcorr (Figure 9D) coefficients versus mean cortical distance between ROI pairs between quadrants of a hemisphere. The exponential fit, linear, and quadratic regression results are plotted as well as the mean linear trend (in gray) from the within quadrant analysis for easy comparison. While the Corr plot once again looks mostly linear, where Pcorr is concerned,

following the initial drop in connectivity with increased distance, there seems to be an increase in connectivity measures when distances are greater than approximately 55 mm.

In order to investigate if the increase in Pcorr for distances longer than ~55 mm was statistically significant, we compared the Pcorr values of the most distant connections and mid-range connections between quadrants. To this end, we pursued a two tailed t-test between the Fisher transformed mean partial correlations for the 10 most distant connections (largest average distances across all subjects and hemispheres: distances of 79mm-85mm) and 10 connections with distances starting at 54mm (54mm-55mm). The mean Pcorr for the distant connections was found to be significantly greater than that for the mid-range distance connections ($p < 10^{-3}$, two-tail ttest).

In order to evaluate whether the trends in the distributions of Corr/Pcorr versus cortical distance were different for within and between quadrants, we quantified the goodness of fit of the linear, quadratic and exponential regressions (Figure 9). The first data column of Table 3 shows that both the within quadrant and between quadrants Corr coefficients were significantly correlated with distance along cortex (p~E-49 and p~E-24, respectively). Pcorr showed statistically significant correlation with distance too, for both within (p~E-36) and between (p~0.0001) quadrants. The slopes of the linear fits within and between quadrants were found significantly different for Pcorr (p=10^-11) with steeper slopes within quadrant than between quadrants.

To compare the goodness of fits of the 3 regressions in the 4 conditions (Corr/Pcorr, within/between quadrant), we computed the Bayesian Information Criterion (BIC) of each regression, and then we computed the Bayes Factor, which allows us to determine the confidence which we can have that the model with the lowest BIC is better than the other models (see Methods; values in Table 3). We found that for Corr, the linear fit is the best within quadrant (~85% confidence). Although the exponential fit was best between quadrants, the associated level of confidence was low (under 42% for exponential fit better than linear; under 6% for exponential better than quadratic). For Pcorr, the confidences observed were very strong, and we saw an interesting difference between the within quadrant and between quadrant cases. Within a quadrant, the exponential fit was the best fit (~100% confidence it is better than both linear and quadratic, Table 3). Between quadrants, the quadratic was the best fit (~100% and 97.8% it is better than linear and exponential, respectively, Table 3). These results suggest that the

relationship between Pcorr and distance along the cortex was not identical within and between quadrants.

Table 3: Goodness of fit of the linear, quadratic, and exponential fits to correlation/partial correlation vs. cortical distance plot.

		P-value of correlation between coefficients & distance	BIC of linear fit to mean	BIC of quadratic fit to mean	BIC of exponential fit to mean	% confidence that the model with the best BIC is better than the other models: 100% * (1-Bayesfactor)	
Corr	**Within quadrant**	4.28E-49	**-1228.52**	-1224.77	-1224.63	84.68 Lin>quad	85.69 Lin>exp
	Between quadrants	5.05E-24	-1187.35	-1188.31	**-1188.43**	41.76 Exp>lin	5.85 Exp>quad
Pcorr	**Within quadrant**	9.33E-36	-1086.76	-1218.09	**-1261.53**	100 Exp>lin	99.99999996 Exp>quad
	Between quadrants	9.37E-05	-1487.55	**-1534.07**	-1526.42	99.999999992 Quad>lin	97.81 Quad>exp

All columns quantify the data plotted in Fig. 9. For each connection we computed the mean cortical distance across hemispheres and mean correlation/partial correlation coefficient across runs, hemispheres and subjects. Thus, for within quadrant we considered 210 connections: two sets (dorsal and ventral) of 105 coefficients each (number of within quadrant connections), and between quadrants there are 225 coefficients (number of between quadrant connections). The first column presents the p-value of testing the statistical significance of correlation between functional connectivity measures and cortical distances. Goodness of fit (columns 2-4) was evaluated using the Bayesian Information Criterion (BIC). A smaller BIC (more negative) implies a better fit relative to the other models. Bold values indicate the best BIC. In order to test differences between measures of goodness of fit, a Bayes Factor was computed. We present the percent confidence that one model is better than the others tested (columns 5-6). Shaded squares indicate a strong confidence that one model is a better choice than the others.

Discussion

There are two types of connections which we investigated: 1) within visual area connections and 2) between visual area connections. Both of these connections were evaluated within the quadrant and between quadrants of the same hemisphere. We look separately at the connectivity within the dorsal and ventral quadrants of the visual cortex in a hemisphere, and between these quadrants, due to the largely symmetric nature of their visual field representation. This separation of within quadrant versus between quadrants analyses also presents a logical separation of generally closer range connections along the cortex (within a quadrant) and longer range connections (between quadrants). In addition, the symmetric arrangement of the visual areas between quadrants creates a distribution of distances within an area that are short (e.g., V1) long (e.g., V3) or intermediate (e.g. V2). This wide distribution of distances has allowed us to demonstrate a complex relationship between FC measures and distance.

Within visual area connections

Within a quadrant, within visual areas, all Corr and Pcorr FC measures showed statistical significance. Across visual areas, Corr (all visual areas) and Pcorr (all visual areas, 1es > all other es) values decreased with increasing eccentricity separation, demonstrating the dependence of resting-state FC on retinotopic eccentricity.

We went into these analyses with some expectations based on the well-studied organization and anatomical connectivity of the primate visual cortex. Horizontal connections, namely direct neuronal linkages within a visual area, have been shown to be generally quite short in the macaque monkey visual cortex, the longest being of less than 2.5mm in V1 and 5mm in V2 in the macaque brain (Amir et al., 1993; Markov et al., 2011). In humans, these connections may be up to twice as long (up to ~5 mm in V1 and ~10 mm in V2, respectively). This estimation is based on the fact that the cycle of ocular dominance columns in humans is approximately double that in the monkey (Adams et al., 2007). Assuming that this scaling factor applies to other features of the organization, the length of horizontal connections in humans should be approximately double their length in monkeys. Now, to put this into the context of our study across different eccentricity regions within a quadrant of the visual cortex, visual areas in a quadrant have a length of around 4.5-6 cm on the cortex (Tootell et al., 1998). Therefore, our 5 eccentricity regions within the visual quadrant would have an average length of around 9-12mm.

This scale of distances along the cortex was confirmed in our distance analysis between central voxels of ROIs, where pairs of adjacent ROIs from within a visual area of a single quadrant were separated by average distance of 10.6mm (STD=2.3mm). Pairs of ROIs separated by a gap (one of the 5 eccentricity regions) (i.e. 2es as depicted in figure 2), also from within a visual area of a single quadrant, were separated by an average distance of 20.4mm (STD=3.9mm). Between the most central and the most peripheral regions (4es), within visual area ROIs the distance is of mean=40.1mm (STD=5.7mm). Notice that both the mean distances for 2es and 4es connections are much longer than the maximum of 5 and 10 mm expected from monosynaptic horizontal connections in V1 and V2 respectively (see above). In spite of the long distances associated with 2es and 4es connections in V1 and V2, these connections are found to be statistically significant.

Anatomical projections in V1 with origin in the LGN and to a lesser extent from monosynaptic horizontal connections in V1 can drive action potential responses in the neurons they project to. However, these two types of connections show high retinotopic specificity. For example, the connections from the LGN to V1 (Salin and Bullier, 1995) and monosynaptic horizontal connections within V1 (Angelucci et al., 2002) are by far more visuotopic than the connections from V3 to V1 (Angelucci et al., 2002). Therefore, given that within areas V1 and V2, within a quadrant, we observe significant correlations/partial correlations between regions separated by ~20mm (2es) and ~40mm (4es), we expect that the underlying mechanisms of fMRI-based functional connectivity measures are not limited to input from the LGN or to direct, monosynaptic neuronal linkages through horizontal connections. Alternative or additional mechanisms may include feedback from higher visual areas (Salin and Bullier, 1995; Angelucci et al., 2002; Shmuel et al., 2005) or from the pulvinar (Liu et al., 2012; de Zwart et al., 2013). We therefore hypothesize that the long range correlation-based functional connectivity within visual areas V1, V2, and V3 is the result of polysynaptic connections, involving local connectivity, higher visual areas and the pulvinar. Indeed, feedback from higher visual areas reaches regions beyond those targeted by monosynaptic horizontal connections (Salin and Bullier, 1995; Angelucci et al, 2002). Similarly, the input from the pulvinar to retinotopic visual areas can drive responses, and neuronal responses in the pulvinar show wide tuning to stimulus position in the visual space. For correlation-based functional connectivity in this very local scale network, these two types of inputs will vary gradually with distance. We can therefore expect the functional connectivity correlation-coefficients to reflect nearly all nodes being functionally

connected, with a gradual decrease in correlation values with increasing distance. Indeed, correlation-based functional connectivity decreases with increasing distance in an approximate linear manner (Figure 9).

Given the regression of time-courses from all nodes in the network when using Pcorr, we expect that these polysynaptic functional connections should be largely removed. The reason for this expectation is that there will be additional regions between or near the two regions whose functional connectivity is computed, which will receive similar polysynaptically-transmitted signals. Signals based on propagation along polysynaptic horizontal connections will have to traverse the distance between the two ROIs in question, thus leaving traces that can be regressed out by Pcorr. Retinotopically broad feedback signals from higher visual areas or the pulvinar will make their impression on additional ROIs near the two ROIs in question, thus making it possible to regress them out. Indeed, Pcorr values were significantly lower than Corr values. Moreover, Pcorr-based FC is more specific to regions with similar eccentricities than Corr-based FC (Figures 5-8), and it does not change linearly with increasing distance (Figure 9).

Note that the longest horizontal connections expected in V1 (2.5mm*2=5mm) are much shorter than in V2 (5mm*2=10mm). Comparatively, projections from V1 to V2 and V1 to V3 are known to be more spread out than V1 to V1 projections (Amir et al., 1993). Given the relatively narrow horizontal connections expected in V1, one could expect that the range of FC within V1 would be relatively small. Remarkably, we see the opposite when comparing within area connections in V1 to those in V2 and V3: across all tests, V1 is among the most abundantly functionally connected visual areas, with high correlations in particular, but also partial correlations. These findings are consistent with previous studies that combined fMRI simultaneously with intra-cortical neurophysiological recordings (Shmuel and Leopold, 2008; Schoelvinck et al., 2010). These studies demonstrated wide regions in V1 whose spontaneous fMRI signals correlated with time-series of gamma band-limited-amplitude recorded by one electrode (one cortical site) in V1.

Within and between quadrants, V1 is found to have significantly higher correlations and partial correlations than V2 (Within: Corr p=4.2x10^-11, Pcorr p=1.4x10^-10; between: Corr p=2.9x10^-4, Pcorr p=2.2x10^-7) and V3 (Within: Corr p=10^-16, Pcorr p=1.3x10^-7; between: Corr p=10^-16, Pcorr p=7.7x10^-11). This evidence further emphasizes that Corr/Pcorr-based

FC involves mechanisms beyond monosynaptic horizontal connections, probably involving feedback from higher visual areas or the thalamus.

Between visual area connections

Within a quadrant, between visual areas, Corr values showed statistical significance for all eccentricity separations. Pcorr showed consistent significance between regions with the same eccentricity and adjacent eccentricities, and, in addition, between V2 and V3 with 2es and 3es. Pairs of ROIs with similar eccentricity showed higher Corr/Pcorr than pairs of ROIs distant in eccentricity, demonstrating that resting-state FC between lower visual areas is retinotopically organized.

The main expectation we had of between-visual area, within-quadrant connections was that there would be a bias toward stronger connectivity between ROIs of similar eccentricities (Heinzle et al., 2011). In fact, Heinzle et al. (2011) identified that the maximum eccentricity range of significant connections between visual areas V1 and V3 was around pi/3 rad in phase values, or 2.1º ([(pi/3 rad) / (2 pi rad)] = [2.1º /12.5º]; 12.5º was the maximal eccentricity they mapped). In terms of eccentricity degrees, the boundaries around the ROIs that we used were 0.5° to 1.7°, 4.1°to 7.1°, and 10.4° to 14°. With a maximum potential range of around 2.1º, we should expect statistically significant FC limited to ROIs of same or possibly adjacent eccentricity. Indeed, between V1 and V3, the only statistically significant Pcorr were obtained for ROIs of same eccentricity and adjacent eccentricities. Thus, our within quadrant Pcorr analysis, which is comparable to the Corr-based weights and permutation test analysis pursued by Heinzle et al., agrees with that expectation. Here we expand on Heinzle et al.'s results by quantifying the functional connections between V1 and V2 and V2 and V3. V1-V2, V2-V3, and V1-V3 show common eccentricity region as well as adjacent eccentricity region Pcorr-based connectivity. V2-V3 shows, in addition, more distant connectivity (up to 3es).

A recent study by Haak et al. (2013) shows that connective fields maps based on visual stimulation are retinotopically organized, similarly to the retinotopic organization we show here in the resting-state. They accomplished this by computing the stimulus-based connective field maps between regions in area V1 and sites in areas V2-hV4. They then demonstrate that the eccentricity and polar angle maps for visual areas derived with conventional receptive field mapping are spatially correlated with connectivity maps computed using the time courses of

voxels in these areas. These results are in line with our findings of retinotopically organized FC of lower visual areas in the resting state.

When considering the effect of retinotopic distance along the eccentricity dimension on functional connectivity in the human visual cortex, one factor to consider is the population receptive field sizes (pRFs) in each of the visual areas. Dumoulin and Wandell (2008) and Harvey and Dumoulin (2011) demonstrated that pRFs become progressively larger as we move from V1 to V2, and further to V3. In addition, within each visual area the pRFs become larger as we move from central to more peripheral eccentricities. This trend of increasing pRFs with eccentricity is strongest in V3 and weakest in V1, and has been supported by evidence in human and monkey studies (Hubel and Wiesel, 1974; Van Essen et al., 1984; Smith et al., 2001; Kay et al., 2008; Amano et al., 2009; Winawer et al., 2010; Haak et al., 2013). V2 and V3 are expected to have broader pRFs relative to V1, thus leading us to expect significant functional connectivity across a wider range of eccentricity regions between these areas. This is confirmed in Figure 5 (right) and Figure 6C, where we see significant partial correlations between regions in V2 and V3 separated by 2 or 3 of the 5 eccentricity regions. In contrast, between V1 and V2 or V1 and V3 the most we see is significance with a separation of 1 of 5 eccentricity regions.

Within a quadrant, between visual areas, Corr/Pcorr values between regions from adjacent visual areas (V1-V2 and V2-V3) were higher than those between non-adjacent areas (V1-V3). These results are in agreement with the anatomical connectivity in the monkey visual cortex. Markov et al. (2011) show that in the macaque visual cortex, the output from V1 to V2 is at least an order of magnitude higher than the communication between any other areas. More generally, the direct, monosynaptic connectivity between cortical areas decreases as the distance between the areas increases (Ercsey-Ravasz et al., 2013). Thus, within quadrant V1-V3 Pcorr statistically significant functional connections are limited to same eccentricity regions because A) they are weaker, and/or B) the regression of the time courses of V2 nodes regresses out a large part of the contributions between V1 and V3, because they are expected to be similar.

Considerations of cortical distance

Within quadrant, Corr and Pcorr showed a strong effect of decreasing coefficients for greater distances (an approximately linear effect for Corr; a strong exponential effect for Pcorr). Between quadrants, the linear effect was still no worse than quadratic or exponential trends for

modeling Corr as a function of cortical distance. However for Pcorr the optimal fit was no longer exponential; in particular, coefficients first decreased and then increased with increasing distance, with the minimum of the quadratic trend-line appearing at approximately 55 mm. The BIC Bayes factor showed high confidence that the quadratic fit was better than the linear and exponential fits for Pcorr as a function of cortical distance (Fig. 9, Table 3). These results demonstrate that the relationship between Pcorr-based FC measures and distance along the cortex is different within and between quadrants.

One could suspect that the retinotopic effect in correlations/partial correlations observed between within-quadrant ROIs is determined largely by cortical distance and not necessarily by retinotopic differentiation. There are several reasons for this expectation. First, the cortical distance between ROIs within a quadrant and within an area is a continuous function of retinotopic distance. This is the essence of retinotopic mapping: cortical distance depends on retinotopic distance within a quadrant in a continuous manner, as shown by Sereno et al. (1995), DeYoe et al. (1996), Engel et al. (1997), Duncan and Boynton (2003), Larsson and Heeger (2006), Wandell et al. (2007), and others. In addition, the majority of anatomical connections impinging on any cortical neuron are of local origin with distances at the mm scale (Braitenberg and Shuez, 1998). This creates an effect of locally dense connectivity that drops with distance. Lastly, since lower visual areas are densely intra- and inter-connected, one may expect 'network effects' to play a significant role, creating both network-scale but also region-scale synchronized fluctuating signals. Therefore, dependence of FC measures on distance in a small, densely connected network could reflect a combination of the underlying anatomical connections and network effects. Consistent with these expectations, the scatter plots in Figure 9 demonstrate that within a quadrant there is an inverse relationship between Corr and cortical distance (see also statistical results from Table 3).

Note that for Corr within a quadrant, not only the quadratic fit's BIC value was not smaller than that of the linear fit, (Table 3), qualitatively we can observe in the plots that the minimum of the quadratic curve is outside of the range of the data. Thus, the fact that for Corr the quadratic fit Bayes factor does not differ with high percentage confidence from that of the linear fit should not be interpreted as evidence against linearity. A similar phenomenon can be observed for Corr between quadrants: although the BIC for the exponential fit is smallest in this case, it is not significantly better than the linear fit. For both within a quadrant and between quadrants, the only

effect that is really observed is the approximately linear decreasing of coefficients with increasing cortical distance. We conclude that the FC Corr measures found within a quadrant and between quadrants vary approximately linearly with cortical distance.

These observations, which show a strong relationship between Corr and distance, raise concerns that resting-state FC within a small, densely connected network does not carry any information on the anatomical connectivity structure of the network. Given these concerns, we sought evidence indicating that FC measures reflect more than cortical distance alone. We verified that there exist regions within the small network we analyzed, which are quite separated in terms of cortical distance, and yet maintain a Pcorr significantly different from zero, demonstrating that the effect which we extract in such a Pcorr analysis is not simply a reflection of the effect of proximity of regions. In our between-quadrant analysis where we have measured the mean cortical distance between ROIs, some examples of partial correlations that are contrary to the effects of cortical distance include V3-e4 to V3-e5 (~84mm) and V3-e4 to V3-e4 (~83mm), which are relatively distant and are statistically significant when compared to the zero ($p= 10^{-5}$ and $p=1.2x10^{-4}$, respectively), while closer connections between V1 and V3 between quadrants (~35-60mm) do not show statistically significant Pcorr.

Several additional findings show differences between the relationship of Pcorr FC and cortical distance within and between quadrants. Firstly, between quadrants, there is an initial decrease in coefficients with increasing cortical distance, as observed within a quadrant, followed by a minimum coefficient (occurring around ~55 mm cortical distance between ROIs), and then finally an increase in coefficients with greater distance. Indeed, the 10 most distant ROI pairs have significantly greater between quadrant Pcorr strengths than the 10 mid-range distanced ROI pairs starting at 54 mm. Quantitatively, a true (with minimum within the range of the data points) quadratic fit is better than the simpler linear fit (~100% confidence) and exponential fit (97.8% confidence, Table 3) for between quadrant Pcorr versus distance analyses (yet this is not the case within a quadrant where exponential is the clear best fit: 99+% confidence). Lastly, there is a significant difference between within and between quadrant slopes of linear fits and quadratic curvatures for Pcorr. Taken together, these findings demonstrate that within the small, densely connected network of lower visual areas, there is no simple mapping between FC based on Pcorr and cortical distance.

The reason for the difference in trends in the plots for within quadrant and between quadrant FC versus cortical distance can be analyzed by tracking which connections populate the critical parts of the scatter plots of Pcorr against cortical distance. Since Pcorr-based FC within a visual area is high even when computing FC between quadrants, regions in V3 residing in the dorsal and ventral quadrants can be expected to show relatively high Pcorr at the tail end of the Pcorr vs. distance curve. This creates a scenario in which relatively distant regions residing in different quadrants show relatively high FC, compared to ROIs that are closer (such as regions in V1 and V2 residing in the dorsal and ventral quadrants, respectively, and representing distant eccentricities. Indeed, upon examining the cortical distances observed for within and between quadrant connections, the majority (51.6%) of the between quadrant connections are longer than the longest within quadrant connection (52.7mm). Within a quadrant, it is only connections with 3 or 4 eccentricity regions separation which are longer than 39mm, with most (18/21) of these connections connecting ROIs in different visual areas and showing no statistically significant Pcorr. In the Pcorr versus distance plots between quadrants, we see a minimum at around 55mm. The most distant connections either connect regions within V3, or regions in V2 and V3, thus either connecting regions within an area or between areas adjacent in the hierarchy of visual areas, with the majority of these connections show or approach statistically significant Pcorr

Correlation versus partial correlation

Pcorr is becoming increasingly used in connectivity analyses due to its greater ability to weed out false positives in the search for direct connections over simple Corr (Smith et al., 2011; Calabro and Vaina, 2012; Dawson et al., 2013; Pandit et al., 2013). Pcorr uses a linear regression to regress out fluctuations from other network nodes during the analysis of connectivity between two nodes, and in this way is able to ensure that predicted connections are more representative of direct linkages. In our analyses, this was vital to obtaining utilizable results, given that Corr predicted all connections as significantly connected in all quadrant-based analyses (within and between quadrants). These kinds of results can be expected when the network of study is densely connected, with network effects of indirect connections potentially accounting for significant correlations. What we wish to portray is that Pcorr identifies direct connections better than Corr, as evidenced by the fact that it can identify some direct connections we know to exist in the monkey (same eccentricity regions between visual areas), while not

identifying some we know to not be directly connected from the monkey (regions with a large eccentricity difference that reside in two different visual areas).

Raemaekers et al. (2014) found that in their Corr analysis, small scale connectivity was nearly completely obscured by large scale network effects unless the effects of larger networks (the first several Independent Component Analysis components) were removed, which could be interpreted as a global-type signal affecting the network of interest. With the use of Pcorr instead of Corr, we are able to reflect synchrony between nodes with reduced contributions from other network nodes, even with the global signal intact in the data. Pcorr regresses out time-courses of all other nodes in the network from the two nodes between which connectivity is being assessed, in order to eliminate contributions from indirect connections from its network prediction. This in effect also removes any global signal which is observable within the network's nodes. Therefore, statistically significant Pcorr suggests the potential existence of a direct, monosynaptic anatomical connection within the network. This being said, Pcorr coefficients can reflect common input from outside of the network to those two ROIs only.

Conclusions

We report our findings on FC at a sub-areal level within and between visual areas V1, V2, and V3. Within-quadrant Corr and Pcorr are found to be significant for all connections within a visual area. Corr, and to a lesser extent Pcorr, spanning beyond nearby eccentricities within a visual area demonstrates that FC reaches further than direct horizontal connections. Consistently in V1 and V2, and somewhat less so in V3, Corr and Pcorr values decrease with increasing eccentricity separation. Between-visual area within quadrant Pcorr spans to same eccentricity regions and to adjacent eccentricity regions, with connections between V2 and V3 reaching even further. Pairs of ROIs with similar eccentricity in two different areas show higher Corr/Pcorr than pairs of ROIs distant in eccentricity (See Figure 8 for a summary of within quadrant Pcorr connectivity). Significant differences were found between the Corr/Pcorr for adjacent visual areas V1-V2 and V2-V3 and between more distant areas V1-V3. The relationship between Corr FC and cortical distance separating the tested ROIs was approximately linear. Pcorr showed a more complex relationship to distance, with exponential and quadratic curves fitting best for within and between quadrants, respectively.

We conclude that resting-state functional connectivities within and between lower visual areas are retinotopically organized. Furthermore, Pcorr FC follows expectations based on the anatomical connectivity in the monkey visual cortex better than Corr does. Lastly, partial-correlation based retinotopically organized resting-state FC reflects more than cortical distance effects.

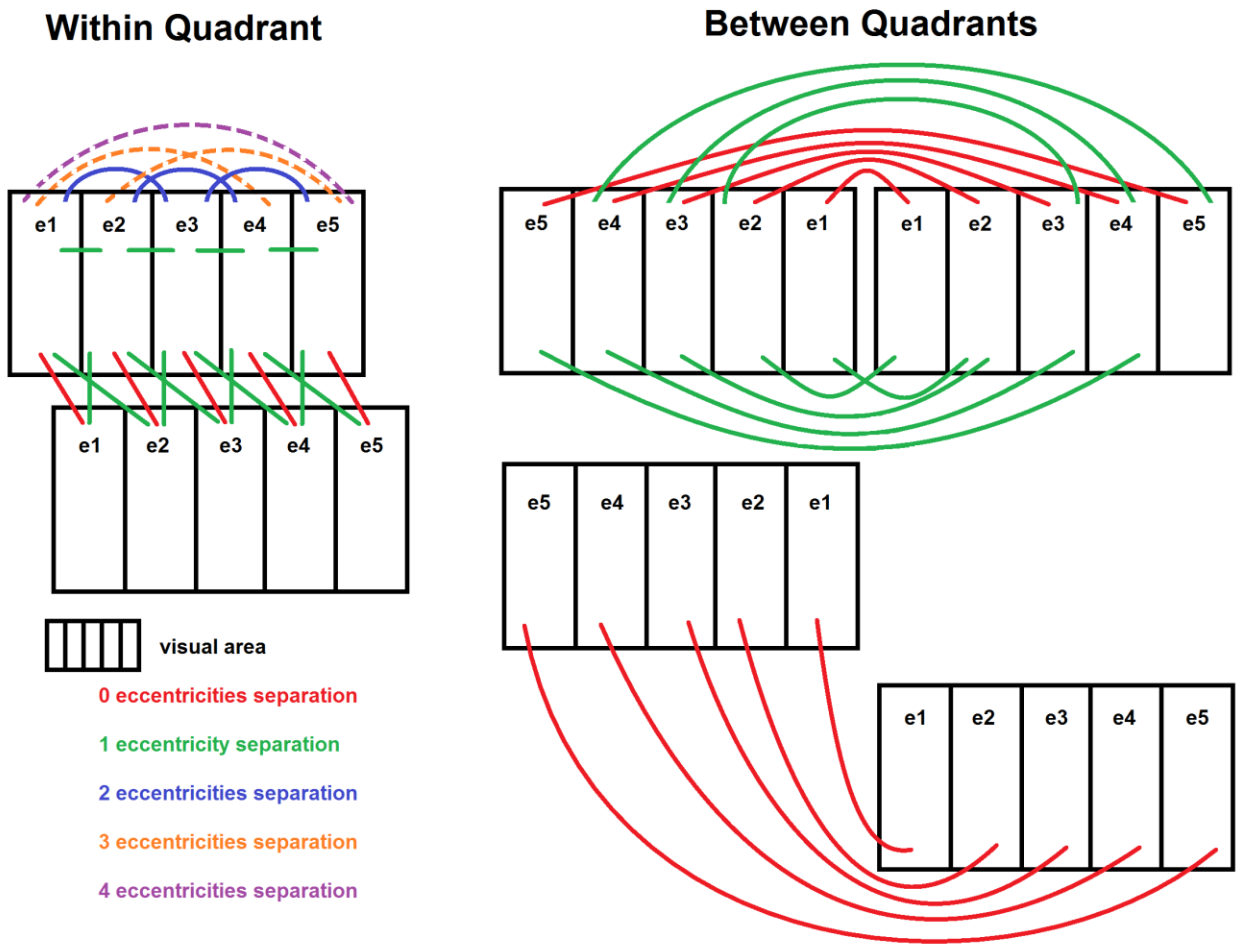


Figure 10. Summary figure of the statistically significant connections from the eccentricity bin analyses for partial correlation. Connections which seem like they may be anomalous were omitted. Dashed lines represent weaker connections.

References

Adams DL, Sincich LC, Horton JC (2007). Complete Pattern of Ocular Dominance Columns in Human Primary Visual Cortex. *J Neurosci, 27*, 10391–10403.

Amano K, Wandell BA, Dumoulin SO (2009). Visual field maps, population receptive field sizes, and visual field coverage in the human MT complex. *J Neurophysiol 102*, 2704–2718.

Amir Y, Harel M, Malach R (1993). Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. *J Comp Neurol., 334(1)*,19-46.

Angelucci A, Levitt JB, Walton EJ, Hupe JM, Bullier J, Lund JS (2002) Circuits for local and global signal integration in primary visual cortex. *J Neurosci., 22(19),* 8633-8646.

Benjamini, Yoav, and Hochberg, Yosef (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B,* 57(1), 289–300.

Birn RM, Molloy EK, Patriat R, Parker T, Meier TB, Kirk GR, Nair VA, Meyerand ME, Prabhakaran V (2013) The effect of scan length on the reliability of resting-state fMRI connectivity estimates. Neuroimage. 2013 Dec;83:550-8. doi: 10.1016/j.neuroimage.2013.05.099. Epub 2013 Jun 6.

Bonhoeffer T and Grinvald A (1991). Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature, 353(6343),* 429-431.

Braitenberg V, Shuez A (1998). Cortex: Statistics and Geometry of Neuronal Connectivity. *Springer,* Berlin.

Calabro FJ, and Vaina LM (2012). Interaction of cortical networks mediating object motion detection by moving observers. *Exp Brain Res., 221(2),*177-89.

Dawson DA, Cha K, Lewis LB, Mendola JD, Shmuel A (2013). Evaluation and calibration of functional network modeling methods based on known anatomical connections. *NeuroImage*, *67*, 331-343.

DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser J, Cox R, Miller D, Neitz J (1996) Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc Natl Acad Sci U S A. 93(6)*, 2382-2386.

de Zwart JA, van Gelderen P, Liu Z, Duyn JH (2013) Independent sources of spontaneous BOLD fluctuation along the visual pathway. *Brain Topogr., 26*, 525-537.

Dumoulin, Serge O., and Brian A. Wandell (2008). Population receptive field estimates in human visual cortex. *Neuroimage, 39(2)*: 647–660.

Duncan, RO, Boynton GM (2003). Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron, 38(4),* 659-671.

Engel, SA, Glover GH, Wandell BA (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex. 7(2),* 181-192.

Ercsey-Ravasz, M, Markov NT, Lamy C, Van Essen DC, Knoblauch K, Toroczkai Z, Kennedy H (2013). A predictive network model of cerebral cortical connectivity based on a distance rule. *Neuron, 80(1),* 184-197

Felleman, Daniel J., & Van Essen, David C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex, 1,* 1-47.

Fox, Michael D, & Raichle, Marcus E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature reviews. Neuroscience*, *8*(9), 700-11.

Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. (1993). Functional connectivity: the principal-component analysis of large (PET) data sets. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, *13*(1), 5-14.

Haak, Koen V., Jonathan Winawer, Ben M. Harvey, Remco Renken, Serge O. Dumoulin, Brian A. Wandell, and Frans W. Cornelissen (2013). Connective field modeling. *Neuroimage, 66,* 376–384.

Harvey BM, and Dumoulin SO. (2011). The relationship between cortical magnification factor and population receptive field size in human visual cortex: constancies in cortical architecture. *J Neurosci. 31(38)*, 13604-12.

Heinzle, J., Thorsten Kahnt, John-Dylan Haynes (2011). Topographically specific functional connectivity between visual field maps in the human brain. *Neuroimage, 56(3),* 1426-1436.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, *160*(3), 106-54.

Hubel, D. H., & Wiesel, T. N. (1974). Uniformity of monkey striate cortex: a parallel relationship between field size, scatter, and magnification factor. *J Comp Neurol., 158(3)*, 295-305.

Jenkinson M., C.F. Beckmann, T.E. Behrens, M.W. Woolrich, S.M. Smith (2012). FSL. *NeuroImage, 62,* 782-90.

Kaiser M and Hilgetag CC (2004). Spatial growth of real-world networks. *Phys Rev E Stat Nonlin Soft Matter Phys, 69(3 Pt 2)*, 036103.

Katsuki, Fumi, Xue-Lian Qi, Travis Meyer, Phillip M. Kostelic, Emilio Salinas and Christos Constantinidis (2013). Differences in Intrinsic Functional Organization Between Dorsolateral Prefrontal and Posterior Parietal Cortex. *Cereb. Cortex. doi: 10.1093/cercor/bht087.*

Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008). Identifying natural images from human brain activity. *Nature 452*, 352–355.

Larsson J and Heeger DJ (2006). *Two retinotopic visual areas in human lateral occipital cortex. J Neurosci., 26(51),* 13128-13142.

Liang, Xia, Qihong Zou, Yong He, and Yihong Yang (2013). Coupling of functional connectivity and regional cerebral blood flow reveals a physiological basis for network hubs of the human brain. *PNAS, 110(5),* 1929-1934

Liao XH, Xia MR, Xu T, Dai ZJ, Cao XY, Niu HJ, Zuo XN, Zang YF, and He Y (2013). Functional brain hubs and their test-retest reliability: A multiband resting-state functional MRI study. *Neuroimage.* doi: 10.1016/j.neuroimage.2013.07.058.

Liu Z, de Zwart JA, Yao B, van Gelderen P, Kuo LW, Duyn JH (2012). Finding thalamic BOLD correlates to posterior alpha EEG. *Neuroimage, 63,* 1060-1069.

Markov NT, Misery P, Falchier A, Lamy C, Vezoli J, Quilodran R, Gariel MA, Giroud P, Ercsey-Ravasz M, Pilaz LJ, Huissoud C, Barone P, Dehay C, Toroczkai Z, Van Essen DC, Kennedy H, Knoblauch K (2011). Weight Consistency Specifies Regularities of Macaque Cortical Networks. *Cerebral Cortex, 21*, 1254-1272.

Pandit AS, Expert P, Lambiotte R, Bonnelle V, Leech R, Turkheimer FE, Sharp DJ (2013). Traumatic brain injury impairs small-world topology. *Neurology. 80(20),* 1826-33.

Priestley, M.B. (1981). Spectral Analysis and Time Series. *Academic Press. ISBN 0-12-564922-3*, 375

Raemaekers Mathijs, Wouter Schellekens, Richard J.A. van Wezel, Natalia Petridou, Gert Kristo, Nick F. Ramsey (2014). Patterns of resting state connectivity in human primary visual cortical areas: A 7 T fMRI study. *NeuroImage, 84,* 911-921.

Salin, P., and Bullier J (1995). Corticocortical connections in the visual system: structure and function. *Physiol Rev, 75*, 107-154.

Schölvinck, M. L., Maier, A., Yec, F. Q., Duynd, J. H., and Leopold, D. A. (2010). Neural basis of global resting-state fMRI activity. *PNAS, 107,* 10238–10243.

Sepulcre J, Liu H, Talukdar T, Martincorena I, Yeo BTT, Buckner RL (2010). The Organization of Local and Distant Functional Connectivity in the Human Brain. *PLoS Comput Biol 6(6),* e1000808.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science (New York, N.Y.)*, *268*(5212), 889-893.

Shmuel A, Grinvald A (2000). Coexistence of linear zones and pinwheels within orientation maps in cat visual cortex. *Proceedings of the National Academy of Sciences US,A 97(10),* 5568- 5573.

Shmuel A, Leopold DA (2008). Neuronal correlates of spontaneous fluctuations in fMRI signals in monkey visual cortex: Implications for functional connectivity at rest. *Hum Brain Mapp 29,* 751-761.

Shmuel A, Korman M, Sterkin A, Harel M, Ullman S, Malach R, Grinvald A (2005) Retinotopic Axis Specificity and Selective Clustering of Feedback Projections from V2 to V1 in the Owl Monkey. *J Neurosci, 25(8),* 2117-2131.

Shmuel A, Yacoub E, Chaimow D, Logothetis NK, Ugurbil K (2007). Spatio-temporal point-spread function of fMRI signal in human gray matter at 7 Tesla. *NeuroImage, 35(2),* 539-552.

Smith A.T., K.D. Singh1, A.L. Williams, and M.W. Greenlee (2001). Estimating Receptive Field Size from fMRI Data in Human Striate and Extrastriate Visual Cortex. *Cereb. Cortex, 11 (12):* 1182-1190.

Smith, S. M., Karla L. Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F. Beckmann, Thomas E. Nichols, Joseph D. Ramsey, Mark W. Woolrich (2011). Network modelling methods for FMRI. *Neuroimage, 54,* 875-891.

Sotero RC, Trujillo-Barreto NJ, Jiménez JC, Carbonell F, Rodríguez-Rojas R (2009). Identification and comparison of stochastic metabolic/hemodynamic models (sMHM) for the generation of the BOLD signal. *J Comput Neurosci., 26(2)*, 251-69.

Tootell, Roger B. H., Nouchine K. Hadjikhani, Wim Vanduffel, Arthur K. Liu, Janine D. Mendola, Martin I. Sereno, and Anders M. Dale (1998). Functional analysis of primary visual cortex (V1) in humans. *PNAS, 95(3),* 811-817.

Van Essen, D.C., Newsome, W.T., Maunsell, J.H., 1984. The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. Vision Res. 24, 429–448.

Vincent, J. L., Patel, G. H., Fox, M D, Snyder, A. Z., Baker, J. T., Van Essen, D. C., et al. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. *Nature*, *447*(7140), 83-86.

Wandell BA, Dumoulin SO, Brewer AA (2007). Visual field maps in human cortex. *Neuron, 56(2),* 366-383.

Wang, Zheng, Li Min Chen, Laszlo Negyessy, Robert M. Friedman, Arabinda Mishra, John C. Gore, Anna W. Roe (2013). The Relationship of Anatomical and Functional Connectivity to Resting-State Connectivity in Primate Somatosensory Cortex. *Neuron, 78,* 1116-1126.

Winawer J, Horiguchi H, Sayres RA, Amano K, Wandell BA (2010). Mapping hV4 and ventral occipital cortex: the venous eclipse. *J Vis 10 (5),* 1.

Zhang, Y. and Brady, M. and Smith, S (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imag, 20 (1)*, 45-57.

Introduction to Manuscript 2 and the Corresponding Data Manuscript

In the introduction to this thesis, we explained how RS-FC is thought to reflect direct-anatomical linkages. However, from previous work (Smith et al., 2011; Raemaekers et al., 2014), it is clear that not all FC measures behave in the same way: they can be more or less sensitive to indirect connections or network effects. Thus, direct connectivity is often lost amidst the broader connectivity of a network. In Manuscript 1, we noticed that the functional connectivity pattern produced by partial correlation in V1, V2, and V3 had some similarities to the connectivity we expect from direct, anatomical connections. This observation is in line with what we would expect theoretically, given that partial correlation is indeed expected to account for network effects through its multiple regressions. However, it has not been concretely established that partial correlation can capture functional connections that reflect anatomical connectivity, and this ability may be context-dependent. Features of the specific network and analyses can have a large impact on the computed FC with any model. Thus, based on the observation in Manuscript 1 that partial correlation seemed to reflect anatomical connectivity, we decided in Manuscript 2 to test how well other network modeling methods/FC measures as well as partial correlation would perform at this specific task in a similar context.

The context in question is that of a small scale, densely connected network, as modeled by the early visual cortex. In fact, this network of study was chosen for three reasons. The first reason is that, as mentioned above, it was in this context that we made our observation that partial correlation was seemingly performing well at reflecting direct connectivity; thus, it was a reasonable extension to pursue this context further. Second, we made this choice as we have access to a relatively complete ground truth comparator in the form of the homologous network in the monkey visual cortex (which is known to compare well with the human visual cortex, see below). Third, if the models can differentiate between true and false positives in a challenging, fine-scale, densely connected network, they will likely also limit the presence of false positives in less challenging, longer-range connectivity tests.

A point that was brought to attention during the journal review process of Manuscript 2 was the fact that we chose to pursue this study with human RS imaging rather than monkey

imaging; a relevant point given that our inferred ground truth is based largely on monkey tracer-injection studies. Certainly, a more direct comparison of monkey imaging to monkey tracer-injection results would remove some uncertainty from the interpretability of the results. Despite this fact, we chose to study humans as our interests lie in discovering the applicability of these methods to human imaging for eventual relevance and use in a clinical setting. We lack reliable studies of anatomical connections of the human brain using tracer injections due to ethical considerations. However, from the functional MRI studies that exist, the human visual cortex compares very well to the monkey visual cortex, particularly in the functional properties of retinotopic visual areas V1, V2, and V3 (e.g., Sereno et al., 1995; Engel et al., 1997; Dumoulin and Wandell, 2008). Finally, subjecting "non-consenting" animals to study presents ethical questions, such as would be the case if we were to image monkeys for this study. Thus it was not an avenue we pursued. Instead, we relied on the known, well studied anatomical connections from previous tracer injection studies in macaque monkeys and on the similar functional organization of the visual cortex in these monkeys and humans, as reflected by previous neurophysiology studies and recent functional MRI studies, respectively.

To more precisely outline the approach in the upcoming study, we tested 16 network models used in the field in terms of their abilities to estimate/approximate the connectivity we know to exist based on the ground truth, thoroughly studied monkey anatomy in the context of a fine-scale, densely connected network such as the visual cortex. Importantly, we test these models under the assumption of no prior knowledge of the system of study. This important distinction inspires Manuscript 2 as the “next step” study to our previous work (Dawson et al. 2013), which undertook similar analyses using a thresholding approach that *must rely* on prior knowledge of the system of study. Key differences between these two works are explained in detail in Appendix B. Briefly, the pursuits explored in Manuscript 2 not considered in the 2013 paper include: (1) comparing interaction measures to zero and performing standard statistical testing with an α of 0.05 (corrected for multiple comparisons); this is a necessity that emerged from the assumption that the user has no information on the studied network, (2) introducing a novel method - *Local* Principal Component-based estimation of the local-global signal, for the removal of network effects, (3) evaluating the performance of statistical and density-based thresholds, (4) testing the effect of increasing sample size on the performance of the different methods, and (5) evaluating the retention of triangle motifs in network predictions.

| Introduction to Manuscript 2 and the Corresponding Data Manuscript

In what follows, we will pursue these specific analyses and more, elaborating on the motivation behind our questions and the literature surrounding the concepts brought forward, as well as presenting a detailed set of results given a number of analytical approaches. In the end, we conclude that estimating direct connectivity *is possible* using RS-FC in an unknown, fine-scaled, densely connected network, and we summarize the methodological approaches best suited to the task.

There are three appendices with content relating to Manuscript 2, the first of which we would like to mention now. **Appendix D presents a data article** (submitted to Data in Brief) elaborating the methodology we used in terms of preparing the Human Connectome Project (HCP) resting-state fMRI data for use in our study and making the data we processed available to the scientific community. The data that we offer for download and study by the larger neuroscience community consists of resting-state fMRI time-series sampled from 60 fine-scaled retinotopically organized regions in cortical areas V1, V2, and V3 (as used in Manuscript 2), obtained from 220 subjects. These data are vital to the study presented in Manuscript 2 as, in using them, we were able to observe the convergence of the statistical trends we observed with smaller sample size. Small sample sizes are commonplace in brain imaging studies, though this issue is actively being tackled in the community by the initiation of open-access data repositories such as the HCP. The prepared data from our data article with defined ROIs in V1, V2, and V3 will allow the brain imaging research community to study the visual cortex with high-quality data from a large cohort of subjects.

Manuscript 2

Local Principal Component regression, partial correlation and inverse covariance identify direct connections in an unknown, fine-scaled, densely connected functional network

Debra Ann Dawson[1,2], Jack Lam[1,2] and Amir Shmuel[1,2,3]

[1]McConnell Brain Imaging Centre, Montreal Neurological Institute

[2]Department of Neurology and Neurosurgery

[3]Departments of Physiology and Biomedical Engineering

McGill University, Montréal, Québec, Canada

Correspondence:

Amir Shmuel, Ph.D.

3801 University Street, room NW109
Montreal, Quebec H3A 2B4, Canada

amir.shmuel@mcgill.ca

Acknowledgements

We thank Lindsay Lewis and Janine Mendola for sharing part of the data we used for this manuscript, and Ilana Leppert and Bruce Pike for support with MRI physics. This work was supported by grants from the Human Frontier Science Program (RGY0080/2008) and the Natural Sciences and Engineering Research Council of Canada (RGPIN 375457-09 and RGPIN 2015-05103) awarded to AS. DD was supported by a Ph.D. student fellowship from the Fonds de Recherche du Québec - Santé.

Abstract

Various models are being implemented in order to study functional brain networks of the human brain. These models often utilize Blood Oxygenation Level Dependent functional Magnetic Resonance Imaging (BOLD fMRI) data because it can offer high spatial specificity and resolution. Here we aim to evaluate the ability of 16 functional network modeling methods to identify *direct connectivity in a densely connected network of which the user has no prior knowledge*. In other words, we strive to approximate anatomical connectivity with functional data, a valuable pursuit given that there are not yet good non-invasive methods for obtaining anatomical measurements in the brain on a fine scale.

The choice of a statistical threshold is always difficult and arbitrary. In our current study, we evaluate which of the network modeling methods performs well using a 'conventional' p-value based significance level of 0.05, corrected for multiple comparisons. This assures the user that there is no need to 'guesstimate' at a threshold. Our findings indicate that the conventional threshold level is indeed appropriate. We include an evaluation of the stability of the network predictions across a range of absolute statistical thresholds as well as a range of density-based thresholds applied to the network connectivity measures.

We also introduce the novel concept of LOCAL Principal Component regression (LPC), and have quantified its effect on all models given an FDR p-value threshold with a maximal FPR of 0.05. This approach results in some of the best performances from network models. While typical global signal approximations are made using information from all voxels in the brain, or all grey matter voxels, we have chosen to approximate a LOCAL average signal using only the voxels in V1, V2, and V3, our network of study. Regions outside of the visual cortex are less retinotopic than those within, therefore, their influence will be widespread in V1, V2, and V3. This leads to the expectation that global contributions can be adequately captured across the voxels contained in V1, V2 and V3.

Multivariate models, such as Partial correlation and Inverse covariance (ICOV), were the only option in order to differentiate between direct and indirect connectivity in this local network without LPC regression. On the other hand, correlation and Kappa, both bivariate models, were top performers in tests when the LPC was regressed. *LPC regression was beneficial to all successful models in terms of bringing network predictions closer to our expectations.* While density-based thresholding shows promise of being able to perform similarly or better than

absolute thresholding, the problem remains that there is no way to determine the appropriate density of network in an unknown brain network.

In addition to analyzing our own data, we support the results of our study with a large number (220) of Human Connectome Project data sets, following power and sample size considerations. This analysis allows us to observe which of the modeling methods will converge (given a large number of data sets) to results very similar to the inferred ground truth direct connectivity. Network predictions with any of fullcorr, Partialcorr, Kappa, or ICOV are excellent in both the smaller and larger networks when the LPC is regressed. Importantly, results from this analysis show that density-based network predictions are not improved when a large number of data-sets is used, whereas network predictions based on absolute statistical thresholding are indeed improved and trend toward our inferred ground truth expectations.

We recommend regression of the Local Principal Component from data before computing network measures and using a lenient significance level with appropriate multiple comparisons correction in future analyses of small, densely connected networks. Our results aid in the selection of models and thresholding strategies for identification of direct connectivity using functional imaging.

Introduction

Though we are continually expanding our understanding about how the brain works, there is still much to be discovered. The human brain is particularly mysterious due to its complexity and the difficulty in obtaining measurements of neuronal activity with wide coverage and high resolution both temporally and spatially. Blood Oxygenated Level Dependent functional Magnetic Resonance Imaging (BOLD fMRI) is one of the better and more widely used data acquisition techniques for studying brain activity in humans. This method offers coverage of the entire brain, high spatial resolution and relatively high spatial specificity, for a non-invasive imaging technique (Shmuel et al., 2007).

A current theme of interest is the analysis of brain networks, examining how regions, distant or close, interact. The term "functional connectivity" (FC) is commonly used to characterize such interactions. It refers to the correlations in time between spatially remote neurophysiological events (Friston et al., 1993). Computing functional connectivity measures associated with resting-state fMRI data, i.e. data obtained in the absence of a task (Biswal et al., 1995), is a popular choice for analyzing brain networks. This is a particularly interesting paradigm for connectivity analyses as resting state functional connectivity (RS-FC) is thought to reflect anatomical linkages (De Luca et al., 2006; Fox and Raichle, 2007; Koch et al., 2002; Quigley et al., 2003; van den Heuvel et al., 2009), although this reflection is not trivially identical (Honey et al. 2010). In this study, we aim to explore the association between FC and anatomy more closely. There is likely more information contained within these functional signals beyond anatomy, however we expect that anatomical linkages contribute significantly to driving the resting state signal and we seek to identify computational models which may be able to extract or approximate this information.

While the working definition of functional connectivity mentions "correlation in time", measures other than correlation per se can quantify a relationship in the temporal domain between neurophysiological events. Many network models are currently being implemented in order to study functional brain networks in humans, with varying degrees of success. A study by Smith et al. (2011) quantified "C-sensitivity" for 38 models using simulated neural network data in order to evaluate each model's ability to predict the true simulated network. In this context, "C-sensitivity" is the proportion of a model's true positive direct connections having interaction measures greater than the 95th percentile of its false positive distribution. Dawson et al. (2013)

further analyzed and tested 30 of these models using resting state fMRI data from the human visual cortex and compared the functional connectivity results to known anatomical connections. Importantly, both Smith et al. (2011) and Dawson et al. (2013) looked to determine which models could identify direct connections by means of a threshold generated using previous knowledge of the network of study. In the current study we evaluate the performance of functional connectivity modeling methods with no prior information available about the networks to which they are applied.

One major concern in functional connectivity network analyses is eliminating false positive results while maximizing true positive results (Khalili-Mahani et al., 2013; Zalesky et al., 2012; Lindquist et al., 2012; Smith et al., 2011; Abou Elseoud et al., 2011; Drakesmith et al., 2015). This is true regardless of how one is interpreting the FC measures. The most common approach to addressing this concern is thresholding the connectivity measures. This, however, requires a careful choice of severity of the threshold such that the false positive results are minimized and the true positive results are maximized (Langer et al., 2013; Dawson et al., 2013; Gerhard et al., 2013; Baria et al., 2013). Such choices cannot be insightfully made with no knowledge of the analyzed network. Therefore, in most cases the choice of the threshold is arbitrary. Often these thresholds are driven by the data themselves, selecting a percentage of the highest interaction measures (connection density-based thresholding), but once again, thresholds determined this way are largely arbitrary (for example, Fornito et al., 2010; van den Heuvel et al., 2017). As of yet, there are no good standards for setting this threshold without prior knowledge of the system of study (Smith et al., 2011; Dawson et al., 2013; Simpson et al., 2013; Garrison et al., 2015; Drakesmith et al., 2015). Here we aim to evaluate various promising models in terms of their ability to minimize false positives and maximize true positives, using common statistical significance thresholding. In essence, we are looking for which models show strong separability of true and false positives, and we are testing in conditions where the ground truth is not known. We further evaluate the sensitivity of these models to the choice of threshold by varying the statistical threshold and in considering the full range of connection density-based thresholds from 100% to 0%.

In more detail, in the current study we aim to evaluate the performance of functional network modeling methods in identifying direct connections in a small, densely connected network by computing fMRI-based resting state functional connectivity. We ask the question: Which of the

tested models are able to identify correctly an expected network of fine-scaled anatomical connectivity with common statistical thresholding and with minimal sensitivity to chosen threshold? To address this question, we analyze fluctuations in spontaneous BOLD signal in the visual cortex of human subjects. We chose to evaluate the models in the visual cortex because of the wealth of knowledge we have in this system from monkey studies. Importantly, from what has been seen thus far, the human visual cortex compares very well to the monkey visual cortex, particularly in the functional properties of retinotopic visual areas V1, V2, and V3, imaged in this study (e.g. Sereno et al., 1995; Engel et al., 1997; Dumoulin and Wandell, 2008). We compare the connectivity results within our visual cortex network to the ground-truth thoroughly investigated connectivity of the monkey model, basing our expectations on the results from anatomical tracer injections in primate studies. The network we study is similar to that outlined in Dawson et al. (2013) within a quadrant of the visual cortex, i.e. a region in the visual cortex responding to a quadrant of the visual field (Figure 1). Within this network, we define connections we expect to exist, and connections we expect to not exist directly.

One might wonder why we focus our attention on such a small portion of the brain, when most recent FC studies look at the whole brain. We feel that evaluation of a fine-scaled network such as the visual cortex in isolation from the rest of the brain is a valuable strategy to observe the local connectivity without unnecessary noise introduced by outside nodes which show far less specificity in their connections with the network of interest (visual cortex) (Salin and Bullier, 1995). Given the broad impact of other brain regions on the network of interest, the information from these outside regions is not expected to be important in the task of identifying short-range, fine-scaled connectivity. Some network models suffer more from additional data entering the model than others (i.e. multivariate models); the large-scaled connectivity patterns and noise can obscure the fine-scaled connectivity (Raemaekers et al. 2014). Thus, utilizing information from the whole brain could potentially be detrimental to these analyses. For these reasons, we choose to focus our analyses only on the information contained in the local network of interest: the early visual cortex. We push this focus further by limiting our choice of "global signal" to the local network we are studying. So rather than referring to "global signal", more accurately, we will refer to the local average signal. We present a novel method: regression of the Local Principal Component (LPC – our model of the local average signal). The LPC is

regressed from our ROI time-courses and we evaluate the effect on the output of the methods we test.

In summary, we aim to 1) Evaluate the abilities of functional network modeling methods to identify direct connectivity in a densely connected network of which the user has no prior knowledge; 2) Introduce and evaluate the "Local Principal Component" regression method and its impact on fine-scaled, densely connected network estimations; and 3) Investigate whether statistical thresholds or connection density-based thresholds are appropriate for use in functional connectivity analyses, including under the conditions of a large sample size.

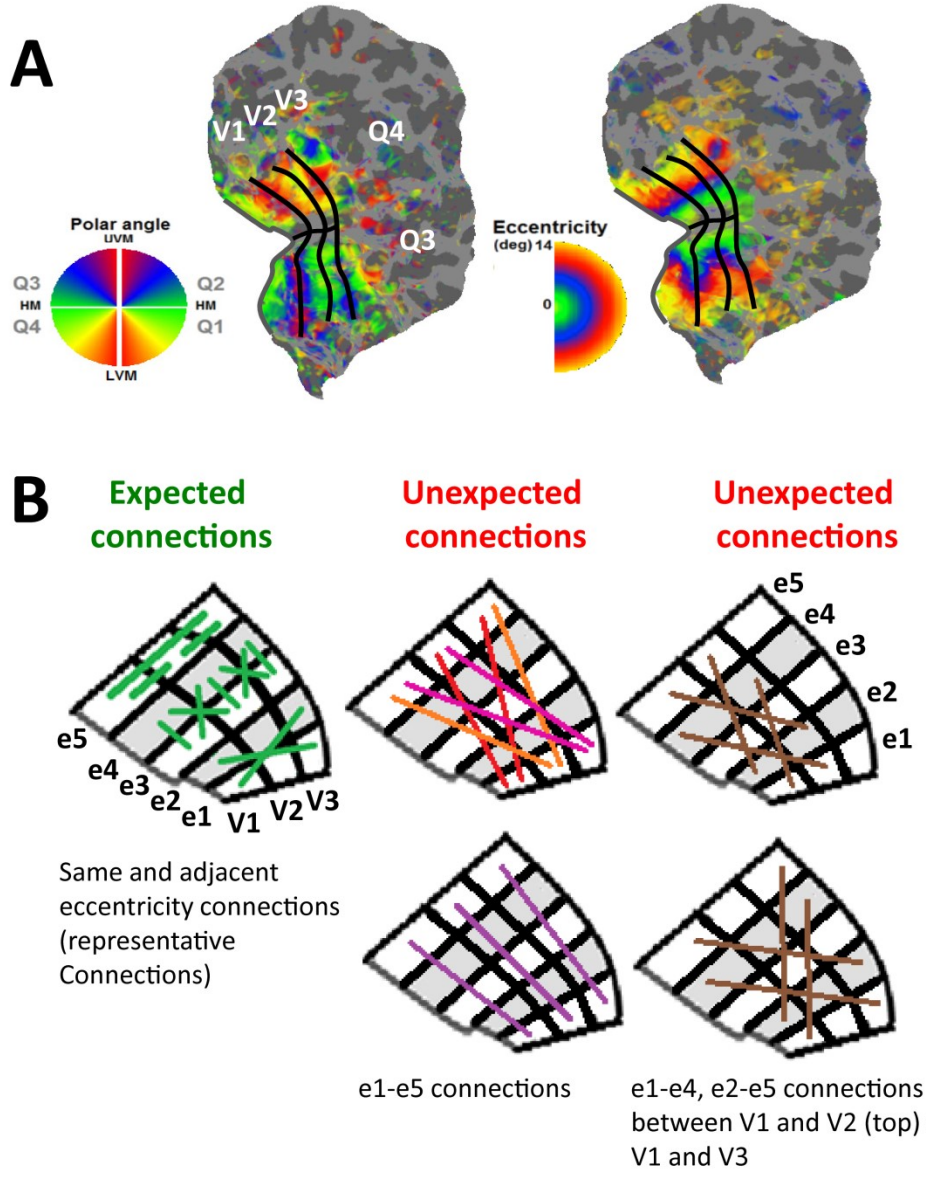


Fig. 1. Retinotopy and connections tested for the network evaluations. A) On the left we see the phase mapping of the functional response to a stimulus changing in terms of polar angle. On the right we see the response to a stimulus changing in terms of eccentricity. Visual areas V1, V2 and V3 were defined by examining the reversal of direction of polar angle phase progression along the surface maps perpendicular to the iso-eccentricity delineation. Voxels from each area were further classified into five eccentricity bins with the boundaries 1.25° to 2.44°, 2.44° to 4.44°, 4.44° to 7.06°, 7.06° to 10.04°, and 10.04° to 14.00°. B) Connections between ROIs tested for evaluations of network predictions. Each hemisphere was divided according to its dorsal and ventral parts, based on the mapping of the horizontal meridian in V1, resulting in brain regions representing the four quadrants of visual space (Q1-Q4). All expected and unexpected connections were set within a quadrant. The left most panel shows representatives of the expected connections as green lines (all eccentricities were tested for; the green lines show only representative expected connection, for clarity of presentation). Connections determined as expected included connections between ROIs of the same eccentricity in 2 different visual areas, and connections between adjacent eccentricity regions either within the same visual area or between visual areas. Connections determined as unexpected included connections between regions with distant eccentricities within an area or in 2 different areas. Unexpected connections were between eccentricity bins e1-e5 (middle panels; represented in red, orange, pink, and purple lines), as well as between e1-e4 and e2-e5 between V1 and V2 or V1 and V3 (right side panels; represented in brown lines). UVM=upper vertical meridian, HM=horizontal meridian, LVM=lower vertical meridian.

Methods

This study applied new analyses to data from one of our recent articles (Dawson et al., 2016), and to data from the Human Connectome Project (HCP; see Appendix D). The preprocessing and ROI definition methods we applied to the two data sets are described below.

Data acquisition and pre-processing

A detailed description of the methods pertaining to our data acquisition, preprocessing, and ROI definition can be found in (Dawson et al., 2016). Six healthy young adults (4 males and 2 females), normally sighted subjects participated in this study after giving written consent in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Each subject was scanned in two sessions: one to obtain high-resolution anatomical images and fMRI for retinotopy and the other to obtain high-resolution anatomical images and fMRI in the resting-state.

All data were acquired on a 3 Tesla Magnetom TIM Trio scanner (Siemens, Erlangen, Germany) using a phased array head coil with 32 channels. For resting-state functional MRI, we used Echo-planar imaging (EPI) to measure changes in oxygenation level-dependent (BOLD).

Resting-state functional MRI was obtained in 8 runs of 8.5 minutes each (68 minutes of data in total) from each of the subjects. Each run consisted of 256 contiguous whole-brain functional volumes [repetition time (TR) = 2000 ms; echo time (TE) = 30 ms; flip angle = 90°; 30 axial-oblique slices; matrix = 112 × 112; field of view (FOV) = 224 mm; acquisition voxel size = 2 × 2 × 2.2 mm].

Retinotopic mapping was pursued using part (the 20 posterior channels) of the same phased array head coil [repetition time (TR) = 2000 ms; echo time (TE) = 30 ms; flip angle = 90°; 28 axial-oblique slices; matrix = 128×128; field of view (FOV) = 263 mm; acquisition voxel size = 2.05×2.05×3 mm].

Resting-state data were preprocessed using the FSL's FEAT software package (Jenkinson et al., 2012). We deleted the first three volumes, performed motion correction, high-pass filtering (0.01 Hz), slice-time correction, and registration to the T1-weighted (MPRAGE) anatomical volume from the retinotopy session. A white matter (WM) mask and a cerebrospinal fluid (CSF) mask were outlined from the resting-state session's anatomical volume using FSL's FAST software package (Zhang et al., 2001). The masks were first eroded by one voxel, thresholded (80% tissue type probability), and registered to the preprocessed functional image. These masks were utilized in a denoising procedure where the average time course from each of the WM and CSF regions was regressed out of the functional image voxel time series.

Regions of interest and networks of the visual cortex

We evaluated a local visual network defined within a hemisphere, which is subdivided into ventral and dorsal sections. Data were obtained from the two hemispheres of each subject, thus from 4 quadrants of the visual cortex (cortical regions which respond to stimuli in the 4 quadrants of the visual field).

Figure 1 presents methodology related to the definition of ROIs and the network under investigation. To define the ROIs, we measured the subject specific retinotopic response evoked by rotating wedge and expanding/contracting ring stimuli. Phase encoding analysis was pursued in order to define visual areas V1, V2 and V3 (Engel et al., 1994; Larsson and Heeger, 2006; Sereno et al., 1995). ROIs were subsections (5) of visual areas V1, V2, and V3 (3) within each quadrant (4) of the visual cortex (total number of ROIs = 5 subsections × 3 areas × 4 quadrants = 60). Quadrants were defined by separating the visual cortex into hemispheres, and then along

the boundary between the dorsal and ventral regions, defined as the retinotopic horizontal meridian. These quadrants as defined on the cortex also respond to a quadrant of the visual field. Misclassifying voxels belonging to one functional region into a neighboring region causes detrimental errors in FC analysis (Smith et al., 2011). In order to avoid this detrimental effect, we excluded from the analysis a 10° polar angle section from either side of the computed boundary between adjacent visual areas. We also excluded eccentricity values below 1.25° due to measurement bias near the edge of the stimulus range and noisy eccentricity mapping near the fovea (Wandell et al., 2007; Benson et al., 2014). Boundaries along the eccentricity dimension within each visual area delineated our fine-scaled ROIs as follows: 1.25° to 2.44°, 2.44° to 4.44°, 4.44° to 7.06°, 7.06° to 10.04°, and 10.04° to 14.00°. These boundaries were selected by computing the distribution of retinotopic eccentricities measured in all voxels, and approximately equalizing the number of voxels included in each ROI across subjects. The eccentricity ranges of the regions are expected to increase as we move more peripherally due to the lower cortical magnification in peripheral representations of the visual field relative to foveal representations (Duncan and Boynton, 2003). We denote these five regions as e1, e2, e3, e4, and e5, respectively. ‘e1’ is an abbreviation for ‘eccentricity 1’, representing the cortical region which responds to stimuli most central in the visual field (with eccentricities in the range of 1.25° to 2.44°). ‘e5’ represents the region which responds to stimuli most peripheral in the visual field (10.04° to 14.00°).

The performance of functional network modeling methods may vary with varying complexity of the network they are applied to. In order to account for this possibility, our study evaluates the performance of the tested network modeling methods in analyzing two different networks, a smaller network with 15 nodes (ROIs) and a larger network with 60 nodes. The networks consist of five eccentricity regions within each of V1, V2 and V3 in a single quadrant of the visual cortex (15 node network, one quadrant of the visual cortex), and from the four quadrants analyzed together (60 node network, entire early visual cortex). For the quadrant network, for each functional run (8) from each subject (6), a 15×15 connectivity matrix is computed for each of the 4 quadrants (8×6×4=192 data sets total). For the entire visual cortex network, for each functional run (8) from each subject (6), a 60×60 connectivity matrix is computed (8×6=48 data sets total).

Network modeling methods

In the current study, we do not evaluate methods that require as input a model of the anatomical connections, such as structural equation modeling (McIntosh and Gonzalez-Lima, 1991), and dynamical causal modeling (Friston et al., 2003). Here we aim to evaluate performance of functional network models in the task of identifying direct connections between regions with no prior knowledge or hypotheses on the connectivity of the analyzed network.

The network models we study are a selection from those previously studied in Dawson et al. (2013) and Smith et al. (2011). We have chosen to carry on with only the 16 models which stood out as being more promising at the task of identifying direct connections in our earlier fine-scaled network analysis (Dawson et al., 2013). The models we have selected are those which consistently achieved higher than 0.4 C-sensitivity (proportion of true positives having interaction measures of magnitude greater than the 95th percentile of the false positive distribution). We tested the following models: Pearson's linear correlation (Fullcorr), Partial correlation (Partialcorr), Regularized Inverse covariance (ICOV, with parameter of 5×10^-3; note for simplicity in notation, all ICOV parameters will be assumed to be ×10^-3, thus this method will be denoted as "ICOV_5"), Mutual Information (MI), Coherence (Wavelet Coherence, CohA; normalized cross-spectral density, CohB), Generalized Synchronization (GenSynch) measures (three measures, including H, N, and S, each with lags of 1 and 2 fMRI volumes tested), Patel's Kappa (also with the data binarized with a threshold of 0.75), and the Bayesian networks (BayesNet) "Peter and Clark" (PC) and "Greedy Equivalence Search" (GES) models. These models are briefly explained in Dawson et al. (2013). Of the models with 0.4 C-sensitivity or greater in Dawson et al. (2013), we have chosen to exclude fullcorr1/2 (correlation performed on the ROI mean times-series after bandpass filtering such that only the bottom half of the frequency range is utilized) as well as ICOV_100. We excluded fullcorr1/2 because it showed no meaningful difference from the results of correlation analyses using the unfiltered time-series. We excluded ICOV_100 because it consistently performed worse than ICOV_5 for the task at hand. Note that Appendix F shows the results of our analyses at different lambdas - pursued independently and subsequently to the analyses in this manuscript - and the main finding is that ICOV with lambda>5x10^-3 in the 15 node network, and lambda>10x10^-3 in the 60 node network found more connections to be statistically significant than was optimal at an alpha of 0.05 (many FPs).

Statistical approach

The set of interaction measures from each model (in separate analyses) for all pairs of ROIs were calculated and were submitted to a mixed effects (repeated measures) General Linear Model (GLM) obtained from the SurfStat matlab toolbox (see website for the model details: http://www.math.mcgill.ca/keith/surfstat/#mixed). Fisher's z transform was applied to measures where appropriate (Correlation, Partialcorr, ICOV, Kappa, Kappa075) prior to statistical testing. To determine whether 2 ROIs are functionally connected, the interaction measures were compared to zero. The mixed effects GLM does not assume that all observations are independent with equal variances. The *SurfStatLinMod* function was used to compute the model, and then the *SurfStatT* function computes a t-statistics for the contrast of interest (in our case, the contrast between real coefficients and zero, 2-tailed t-test). Finally, the GLM's t-statistic was used to obtain a p-value. Using a maximum acceptable false positive rate (FPR) of 0.05, p-values were assessed for significance after identifying the False Discovery Rate (FDR) p-value threshold (Bejamini and Hochberg, 1995). The FDR p-value threshold is based on the set of p-values resulting from the computed contrasts and is thus different for each of the models and in the two networks. These FDR thresholds (p1) are presented in the sub-titles of the individual plots of Figures 4, 5, 8, 9, 13, 14, 15 and 16.

In setting the criteria for which connections were included in our network predictions, we only considered the connections which had a positive mean interaction value. The negative functional connectivity output by the models may contain insightful information, however, in a local, fine-scaled network we expect the overall direct connectivity – which we are interested in – to be positive. There is also concern in the field around possible spurious anti-correlations being induced with a GS regression, a topic currently under investigation. Matrices including both positive and negative statistically significant connections are presented in the supplementary material. As an example, in the 15 node network analysis without LPC regression (Suppl. Fig. 3-B) we found 6 statistically significant negative connections for Pcorr, and 7 for ICOV, whereas with the LPC regressed (Suppl. Fig. 9-B) we found 17 and 18 negative connections, respectively.

Specifically for the visual cortex, we expect that the positive (excitatory) effect of the feedforward connections outweighs negative (inhibitory) feedback connectivity such that the

total cumulative connectivity observed would be positive (Power et al., 2013). This is further discussed in the Discussion.

Statistics-based thresholds and connection density-based thresholds

FC studies use a variety of thesholding methods and levels, however there are two main categories of commonly used thresholds. The first category comprises thresholds based on statistical hypothesis testing, and the second category includes proportional – or density-based – thresholds. For both types of thresholds the cutoff is usually defined arbitrarily. Density-based thresholds may lead to more stable results when comparing brain network global and node-level measures across various thresholds. For example, van den Heuvel et al. (2017) characterized the stability of the functional connectivity network measures of global efficiency and clustering under various proportional thresholds. In our study, to evaluate the performance of the different models we applied both statistical thresholds and connection density-based thresholds. We present a range of statistical thresholds, with p-values beginning at the FDR p-value associated with a maximum FPR of 0.05, i.e. p1, and then p1×10^-2, p1×10^-4… p1×10^-50 (Figures 4, 5, 8, 9, 13, 14, 15 and 16). We retain all connections with interaction measures with p-value smaller than the defined threshold. We also present a range of density-based thresholds, starting with connection density of 100% (i.e., all connections are included), then repeatedly decreasing the threshold by 4% until we reach a 0% network density (Figures 6, 7, 10, and 11). For applying density-based thresholds, we rank all connections according to their interaction measures, and retain all connections with interaction measure greater than the defined threshold.

Evaluating the performance of network modeling methods - expected and unexpected connections

Once the complete of set connections within a quadrant are assessed for significance, we then focus on the set of connections on which we have expectations based on our knowledge of the anatomical connectivity in the monkey visual cortex (Figure 1B). The early visual cortex is densely connected, with forward projections from V1 to V2 and V3, and likewise from V2 to V3, but also feedback projections between each of these areas (Felleman and Van Essen, 1991; Salin and Bullier, 1995). Salin and Bullier (1995) provide an excellent review of tracer injections studies of the visual cortex. Inter areal connections between lower visual areas are specific to similar eccentricities (Salin and Bullier, 1995; Shmuel et al., 2005; Heinzle et al.,

2011). This finding is utilized in our set of expected connections (Figure 1). We also expect intrinsic, or horizontal connections within a visual area. The length of such connections is expected to be on the order of 1 mm to 8 mm. More precise length estimates have been identified as 2.5 mm or less in V1 and approx. 5 mm in V2 in the macaque brain (Amir et al., 1993; Markov et al., 2011). Amir et al. (1993) also found that the length of horizontal connections increases along the hierarchy of visual areas, such that the length observed in V1 would be shortest, followed by V2, then V3. In the human brain, we expect the length of connections to be approximately double those found in macaque brain. Adams et al. (2007) found that the cycle of ocular dominance columns in humans in approximately twice the length in the monkey. We assume that this approximate scaling factor applies to other features of visual organization. This means that we expect to see horizontal connections of approximately 5 mm in V1, 10 mm in V2 and up to 16 mm in V3. The length of visual areas measured on the cortical manifold in humans is within the range of 45-60mm (Tootell et al., 1998). Therefore, the average length of our 5 ROIs should be approximately 10 mm long. Given the scale of our ROIs and the scale of the intrinsic connections, we included connections between adjacent ROIs within a visual area in our set of expected connections. Salin and Bullier also comment on the divergence and convergence of connections between visual areas. For example, the data from paired injections in area V1 of macaque monkeys (Perkel et al., 1986) shows that the divergence of the feedback connections from V3 to V1 is on the order of 5-7 mm (comparable to 10-14mm in humans). Therefore, we can expect that connections between visual areas exist not only for ROIs with identical eccentricities but also for ROIs with similar eccentricities.

These observations in monkey anatomical studies form the basis of our expectations for the connectivity we anticipate in a successful approximation from our tested models of direct connectivity in humans using RS-fMRI. We call these our “expected” connections. From the same observations we can also form a group of connections which we expect do not exist as direct neuronal linkages. These are connections between ROIs with distant eccentricities, which do not correspond retinotopically (i.e. regions which respond to stimuli in a different eccentricity region of the visual field), and we call them our “unexpected connections”. The unexpected connections are those between regions that are separated by at least 2 eccentricity regions, representing a minimum of 20 mm along the cortex, and in most cases more (Figure 1). The set of connections we selected represent conservative choices, with high confidence regarding their

classification as expected or unexpected. We make no assumptions on any other connections which may be present within our defined network. These choices of expected and unexpected connections represent a reliable set of connections we use as reference. For evaluating the performance of the models, we consider only the subset of network connections defined as our expected and unexpected connections. All predictions of other possible connections in the network are ignored. However, for the data analysis, we use all time-courses sampled from all ROIs. Therefore, any influence connections that do not belong to the set of expected or unexpected connections may have in multi-variate analysis on the set of these connections should be considered by good multi-variate models. Note that while our ground truth connections for evaluating the performance of the models only includes the set of expected and unexpected connections, we do present the complete matrices of functional connectivity measures of all potential connections, not only those included in the sets of expected and unexpected connections (Figures 3 and 12).

Based on this set of expectations on the anatomical connectivity we should observe in the visual cortex, we assess the performances of the various models at identifying direct connections. There are 51 expected direct anatomical connections between ROIs within a single quadrant: between 2 ROIs from the same eccentricity bins but different visual areas (e.g., V1e1 and V3e1, or V1e4 and V2e4); and between ROIs from adjacent eccentricity bins, both from the same and different visual areas. We considered 17 connections expected not to exist as direct anatomical linkages: those between ROIs from the same or different visual areas that would connect the most central eccentricity (e1) and the most peripheral eccentricity (e5), as well as those from e1-e4 and e2-e5 between V1 and V2 or V1 and V3. This definition of "expected" and "unexpected" connections is based on the ground-truth, thoroughly studied anatomy in the monkey visual cortex, as outlined above. While it is possible that the e1-e4 and e2-e5 connections between V2 and V3 ROIs are also direct, based on considerations of receptive field sizes, we preferred to remain conservative in our expectations; thus, they were not included in the set of unexpected connections. It has been shown that population receptive fields (pRFs) become progressively larger as we move from V1 to V2, and V2 to V3 (Harvey and Dumoulin, 2011). This trend has been noted in several studies (Smith et al., 2001; Dumoulin and Wandell, 2008; Kay et al., 2008; Amano et al., 2009; Winawer et al., 2010; Haak et al., 2013). It is likely that the mechanisms involves a larger sampling extent of V2 by V3 than the sampling extent of V1 by V2 (Harvey

and Dumoulin, 2011; Haak et al., 2013). Therefore, we considered the connections e1-e4 and e2-e5 as unexpected only between pairs of areas that include V1 with its small receptive fields, i.e. V1-V2 and V1-V3.

For the complete visual cortex analysis (60 nodes), the defined network remains the same. The difference lies in the number of connections considered simultaneously by multivariate modeling methods, and in the amount of data available for statistical tests. Note that no additional expectations on between quadrant connections are made. In order to focus the comparison of performances of the models using small and large networks, the set of expected and unexpected connections remained the same in both the quadrant network and the complete visual cortex network. Thus, the results obtained for the small and large networks can elucidate the impact of increasing the complexity of a network on the relative performances of the various network models.

In evaluating the performance of the models, we define the proportion of true positive connections (TP) as the number of identified expected connections divided by the overall number of expected connections. We also define the proportion of false positive connections (FP) as the number of unexpected connections that passed the significance threshold divided by the overall number of unexpected connections. The main measures we used to evaluate performance are the proportions of TPs and FPs. These proportions are assessed at a range of thresholds. The proportion of TPs is similar to the "C-sensitivity" introduced by Smith et al. (2011) and used in our previous paper (Dawson et al., 2013), however there are some important differences. The key difference is that "C-sensitivity" relied on prior knowledge of the network, being defined as the proportion of true positive connections with greater magnitude than the 95th percentile of the false positive distribution. Thus, C-sensitivity evaluated the separation or overlap between the distributions of true positive and false positive values. In contrast, in our current study we do not use any prior knowledge of the network: we compare the interaction measures by standard statistical testing to 0. Another difference is that C-sensitivity considered only one threshold. In contrast, our analysis here aims to identify network modeling methods that perform well across a wide range of thresholds and remain stable across thresholds, again emphasizing performance with no prior knowledge of the network.

For presenting a comparison between distributions of interaction measures having different scales across models in Figures 2, 3, and 12, mean interaction coefficients were normalized by dividing by the model's maximum mean interaction value. This was done only for presentation purposes.

Local Principal Component regression

A model's performance may also be affected by the influence of the global signal (GS) on the data. The GS may completely or partially underlie observed relationships between regions of interest: the similarity in signal could potentially originate from large scale brain dynamics including fluctuations driven by respiratory and cardiac cycles (Fox et al., 2009) and neuronal activity (David Leopold, PNAS, 2010). The influence of the GS can also distort a present relationship between two nodes given that input from other brain regions may boost the existing interaction/co-fluctuations between two nodes.

It is for these reasons that regression of a GS is a step often taken in whole brain functional connectivity analyses (Carbonell et al., 2011; He et al., 2012; Carbonell et al., 2014). With GS regression, one strives to remove any fluctuations from the data that would be common to all brain regions and thus not contain any unique information about the connectivity between two specific regions of interest. Therefore, in addition to the analysis without GS regression, we also pursue the analysis following GS regression.

Importantly, when analyzing a fine-scaled and densely connected network without some kind of global normalization, the distributions of interaction values by some of the models – such as correlation – are not necessarily centered around zero. Comparing the connectivity measures to zero for this type of model without a GS regression might miss the potential of the model. Given that the result of this regression is to reduce connectivity measures closer to zero, and the concern for spurious anticorrelations during such a regression, we have chosen to proceed with a two-tailed t-test, and then only consider the significant connections with a positive mean value (as in the case without GS regression). By regressing out a GS, we remove a large portion of co-fluctuations with amplitudes that may be higher than those of the network's direct connections, and increase the likelihood that the observed relationships originate from local interactions of the pairs of considered nodes.

While typical GS approximations are made in the context of whole brain analysis using information from all voxels in the brain or all grey matter voxels, we have chosen to regress out a local GS, obtained from only the voxels in V1, V2, and V3. The main reason for this choice is that the fMRI data we acquired consisted of axial-oblique slices covering the visual cortex but not the entire brain. However, we do believe that our approximation of the GS is appropriate given the focus of our analyses. Areas outside of the lower visual areas are less retinotopic than those within. In addition, areas outside of the lower visual areas project to widespread regions in V1, V2, and V3 (Salin and Bullier, 1995). This leads to the expectation that global contributions of neuronal origin (and non- neuronal origin) can be adequately captured across the voxels contained in V1, V2 and V3. Additionally, Yang et al. (2016) showed that contributions to the global signal are non-uniform, finding a spatial distribution of areas variable contributions to the global signal. This supports the idea that a more appropriate "global signal" estimate in local networks should be computed with local information only.

The model of GS that we use (on the local scale) is that presented by Carbonell et al. (2011) for regressing out the first PC obtained from whole brain RS-fMRI. Specifically, we approximate the local GS fluctuations in our small network using the first principal component in a principal component decomposition of the spontaneous activity patterns in V1, V2 and V3. We refer to this approximation as a "Local Principal Component" signal (LPC).

As in our analysis with no global signal regression, following the GS regression we proceeded with a two-tailed t-test, and retained the significant connections with a positive mean value.

Since partial correlation and other multivariate models already consider the common fluctuations across all ROIs in the network, a "global effect" is implicitly considered in these analyses regardless of whether a LPC regression is performed. Despite this, we expect that there may be slight differences in the results from the multivariate models between analyses with and without LPC regression. The reason for this expected difference is that we compute our LPC across all voxels in V1, V2 and V3 from all quadrants, and in our ROI definition, we exclude voxels which lie close to the boundaries between visual areas, or very near to the foveal representation on the cortex. This means that there is a small amount of additional information captured in the LPC beyond that which is found in the ROIs. The LPC is computed across voxels in all quadrants so in the 15 node network we expect to see a more pronounced effect of

removing the Local Principal Component than in the 60 node network since a fair amount of additional information is introduced.

Testing with a large sample size: Human Connectome Project data

Following this study with our own data, we have redone identical analyses using data from the Human Connectome Project (HCP) in order to observe the effect on the results when a much larger sample size is used. In this part of the study, we sought to achieve a power of approximately 95%; however, the sample size required to achieve this power is different for each model and analysis. As a result, we have chosen a sample size which provides approximately 95% or more power for most models and analyses. Exact power values will be identified for each model and analysis in the HCP relevant figures (Supplementary Figures 23, 27, 31, and 35). We have based our sample size calculations on a conservative algorithm for repeated measures studies published by Morgan and Case (2013). This method makes several assumptions, but always such that the estimated sample size required is potentially over-estimated, and not underestimated. They provide an estimate of a multiplier, V, by which one can decrease the sample size given a certain number of repeated measures, k:

$$V=(k+1)^2/(4k^2);$$

The required sample size for a power of 95% was computed via Matlab's "*sampsizepwr*" function. We used the 5th percentile of a model's distribution of mean interaction magnitudes from the connections found to be statistically significant in our 6 subject study as the threshold at which we would like to see a difference from zero in the HCP study. We note that we used the complete set of statistically significant connections, and not only those from our set of expected connections, thus we did not utilize "prior knowledge" of the system in order to set our desired effect size; we simply used results from a pilot study. We maintain our 6 subject "pilot study" as the main set of presented results here due to the fact that for these subjects the ROIs were defined via *measured* retinotopy, an approach that was not possible for the HCP data set. See below for our strategy for identifying boundaries between visual areas in the HCP analyses. Taking into consideration the sample size requirements from each model and each analysis adjusted for repeated measures, we have analyzed data from 220 subjects as a very conservative estimate of the number of subjects required. This is given the 4 runs of resting state data available from subjects within the HCP database as repeated measures for each subject. Subjects were

randomly selected from those available in the 900 subject data release. Preprocessed resting-state data provided in the HCP900 dataset were used.

Since we did not have retinotopy scans for the HCP subjects, we have delineated V1, V2, and V3 using the template (version 2.5) described by Benson et al. (2012, 2014). The template was produced by aligning aggregate retinotopy from a group of subjects to create a model of V1, V2, and V3 retinotopy based on surface topology in the visual cortex. This template was applied to the HCP subjects' *orig.mgz* (output file from Freesurfer's surface extraction) after computing the registration of the template to the subject. The registration is accomplished via alignment of the *sphere.reg* subject file to *fsaverage_sym.sphere.reg*, an average surface atlas available in Freesurfer to which the template is already aligned.

There are two differences in the HCP analyses compared to what we have done with our data. First, we only present results from the 7 most promising models based on the analysis of our 6 subjects: fullcorr, Partialcorr, ICOV, Kappa, Kappa075, GES, and PC. Second, the decision-making process for defining the boundaries for the eccentricity regions differs. Given that we have computed the retinotopic data for the HCP subjects via a template, the range of eccentricity degrees present in the data differed from our experimentally designed range of eccentricity values obtained from out 6 subjects (where the maximum presented was 14°). With the HCP data, the eccentricity values given by the model range from 0° to 90°. We use the same eccentricity boundaries for the HCP data as in our six subjects study, therefore, the extent of the studied cortical surface should be similar. In order to check that this is indeed the case, we computed the average visual area cortical length in our six subjects across all visual areas and quadrants (57 mm, standard deviation 13mm). We then identified the HCP eccentricity values associated with that mean cortical length. The mean eccentricity value at the outer edge of the range was found to be 15.86° – a value close to the boundary of 14° we expected based on the eccentricity maps from our subjects. Therefore, for the HCP data analysis, we used the same eccentricity boundaries (1.25° to 2.44°, 2.44° to 4.44°, 4.44° to 7.06°, 7.06° to 10.04°, and 10.04° to 14.00°) we applied in our six subjects analysis The remainder of the data preparation and analyses were done as in the study with our six subjects' data.

The HCP 900 subjects data release was comprised of 45% males, 55% females, and the average age of participants was approximately 29 (age ranges were provided). The subject pool is of healthy young adults. We selected the 220 subjects of this pool randomly.

Given that we perform many analyses, we present a full list of the analyses undertaken in this study in Table 1.

Table 1: The set of analyses we will present from this study either in the main manuscript or in supplementary material

	15 nodes (1 quadrant)		**60 nodes (4 quadrants)**	
Our 6 subjects	*LPC not regressed*	All data sets	*LPC not regressed*	All data sets
		Reduced data to match 60 nodes analysis		
	LPC regressed	All data sets	*LPC regressed*	All data sets
		Reduced data to match 60 nodes analysis		
220 HCP subjects	*LPC not regressed*	All data sets	*LPC not regressed*	All data sets
		6 random subjects		6 random subjects
	LPC regressed	All data sets	*LPC regressed*	All data sets
		6 random subjects		6 random subjects

Results

One of the primary goals of this study is to determine if there are network models among those tested which are capable of accurately predicting direct connections in a fine-scaled, densely connected network of which no prior information is available, using common statistical hypothesis testing. This is a valuable feature a model can have given that we do not yet have methods to determine appropriate thresholds without prior knowledge of the system. The early visual cortex is a densely connected network, with indirect connections existing between all ROIs in the network. Thus, this challenging network will test the models' abilities to differentiate direct from indirect connections. We consider direct connections obtained from non-invasively obtained functional MRI data as predictive of anatomical connections. It is important to note that when we judge a method as having good or poor performance, it is specifically with respect to this particular task: identifying direct connectivity in the context of fine-scaled, densely connected, unknown networks. In network or FC analyses, there is certainly other information

that may be of interest to the researcher (e.g., broader functional relationships). Here, we do not draw conclusions on the usefulness of these models for the extraction of other information.

Throughout the results section, we examine how the models tested responded under the different conditions of each test. Items which vary across tests include network complexity/number of nodes, abundance of data/sample size, "global signal" regression via the LPC regression, and thresholding approach (statistical hypothesis testing vs. density-based thrsholding). It is important to test the models under different conditions in order to understand their strengths and limitations, and also for considerations of generalizability of findings. We note that MI, cohA, cohB, and the GenSynch models H, N, and S perform poorly at the task at hand in nearly all tests (except for density-based analyses without LPC regression - see Figure 6, Supplemental Figure 17 and Tables 2 and 4, column 8) thus we do not report on them in detail. Their results are presented in the figures for interest.

15-node network: Main results

Our benchmark analysis is that of the 15-node (single quadrant) network, without removal of our global signal, the LPC. We will present this analysis in detail, and from there present further analyses in less detail, highlighting the items that agree or differ from this benchmark analysis. We begin with Figure 2 which shows the distribution of normalized mean interaction measures for each model. The complete distribution of interaction measures is represented in the background of each panel by the gray curve. In the foreground of each panel we have the interaction measures for the expected (green) and unexpected (red) connections. Ideally, there should be little or no overlap between the expected and unexpected connection distributions. At first glance, most models do show relatively good separation between these two distributions, which is consistent with the results of Smith et al. (2011) and Dawson et al. (2013). The next item to consider is if this separability can be captured with common statistical hypothesis techniques, with no prior knowledge that can inform the choice of statistical threshold.

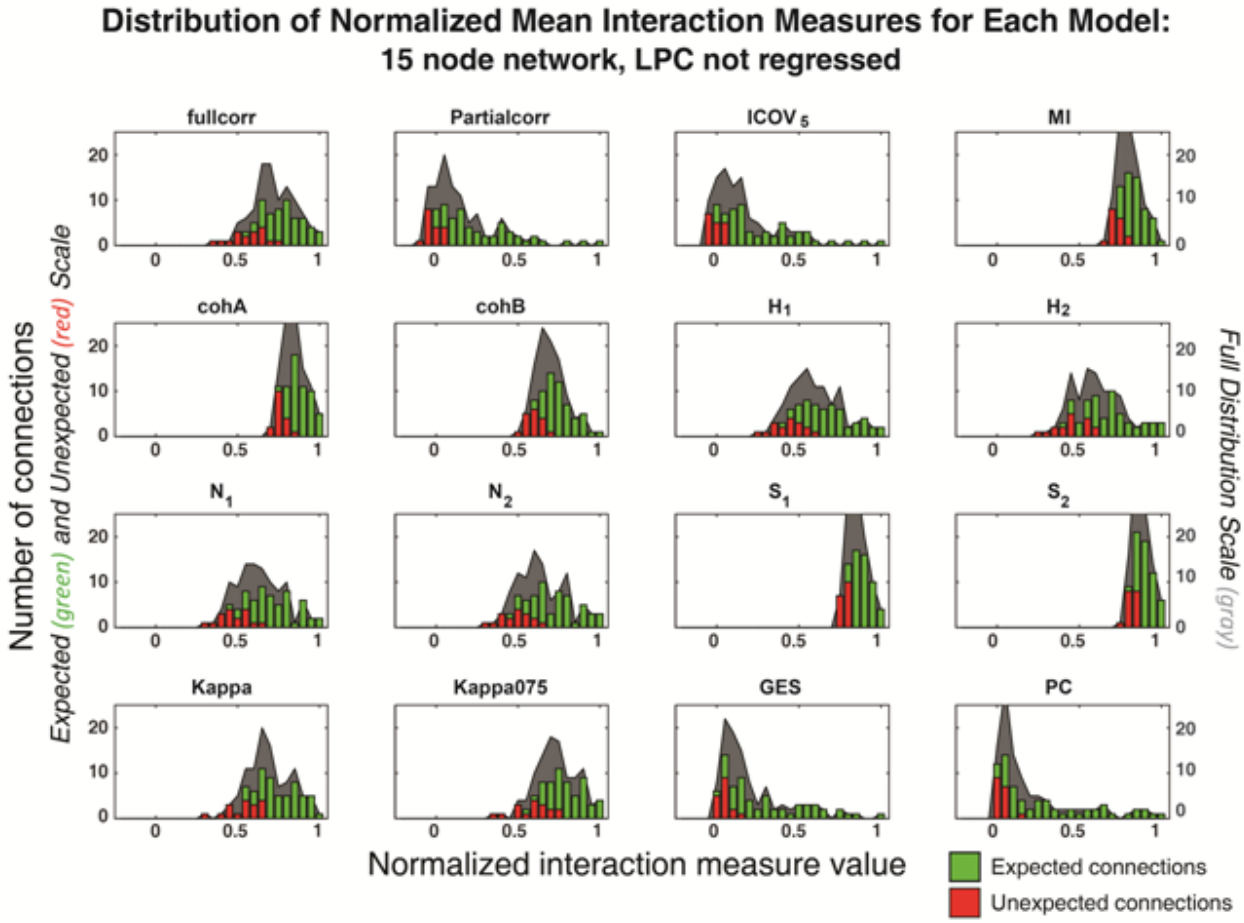


Fig. 2. Normalized mean interaction measures obtained by each of the models for each of the connections tested within the quadrant, with no LPC regression. Normalization is done by dividing all measures by the maximal value, such that for each method the maximal value presented is 1. The gray plot reflects the complete distribution of mean interaction measures (over subjects, runs, and quadrants) of all computed connections ([15×15-15]/2=105). The vertical axis scale for gray plots is on the right-hand side of the panels. Green bars reflect the mean interaction measures associated with the 51 expected connections. Red bars reflect the mean interaction measures associated with the 17 unexpected connections. The vertical axis scale for the green and red bars is on the left-hand side of the panels. Note that all connections are presented, not only those showing statistical significance.

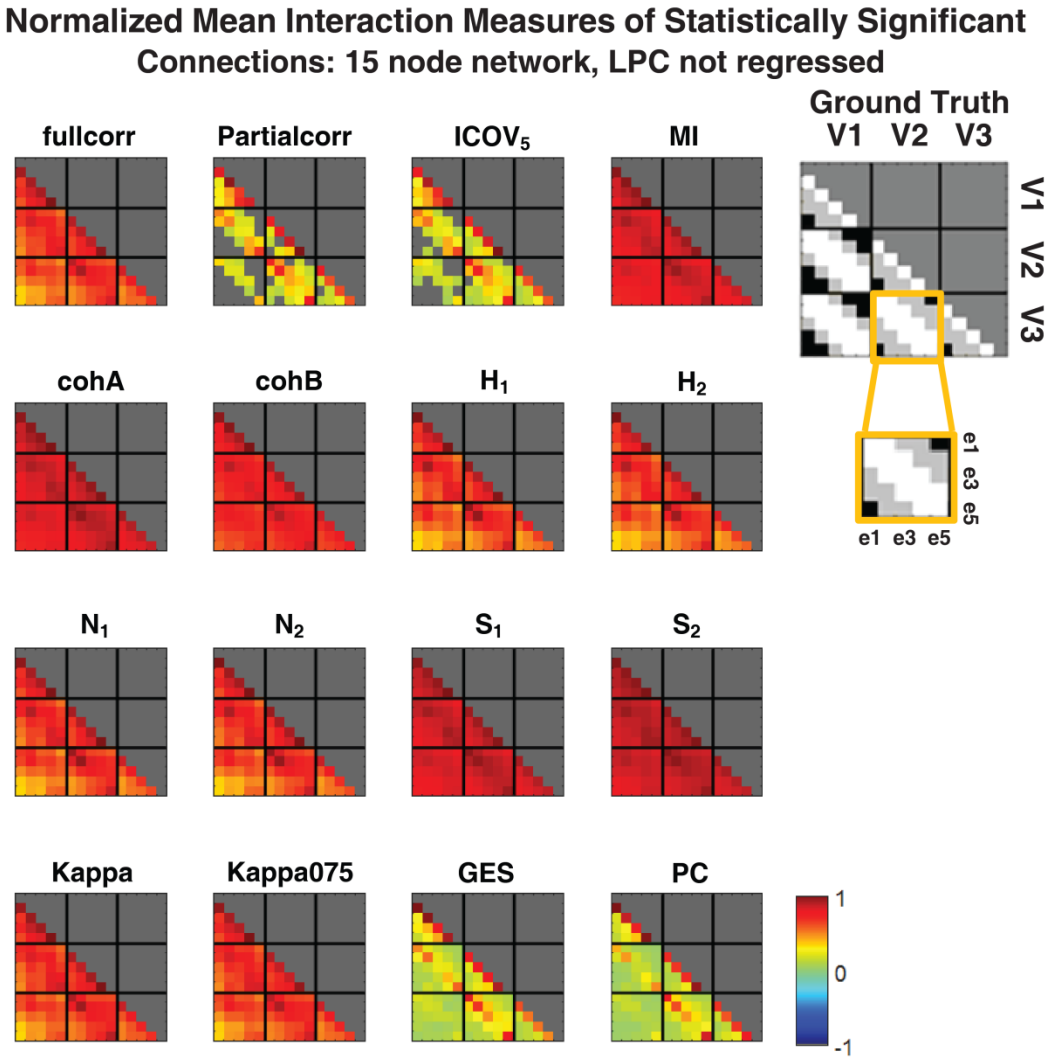


Fig. 3. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 15 node, quadrant network with no LPC regression. Normalization was achieved by dividing all mean measures by the maximal value. The "Ground Truth" connectivity matrix in the upper right-hand corner has been constructed based on knowledge of the monkey visual cortex, and includes the expected connections (white) and the unexpected connections (black) presented in Figure 1. Light gray squares in the ideal matrix are connections on which we placed no expectations. Dark grey squares above the diagonal are connections that were not tested. In the model matrices, colored squares represent connections that showed statistically significant mean interaction measure different than 0, whereas dark gray squares represent connections that did not show statistical significance. Statistical significance was tested against the FDR p-value threshold given a maximum FPR of 0.05.

Figure 3 shows the 15 node networks predicted by each model without regressing the LPC, given an FDR p-value threshold associated with a maximum FPR of 0.05. The majority of models, including FullCorr, Kappa, GES, PC, MI, coherence, and GenSynch models H, N, and S find that all ROIs within the network are connected. The only models that begin to successfully weed out some of the indirect connectivity are PartialCorr and ICOV. Across these two models and two additional models – Bayesian networks GES and PC – we see a strong trend for higher connectivity measures between ROIs with similar eccentricities, along the diagonal in each subsection (smaller black square) of the matrices of normalized interaction measures. This observed trend is consistent with our expectations (see "Ground Truth" in the upper right panel of Figure 3). Within this figure we can see which connections *are* identified as significant although we *did not* expect them to be significant (i.e. FPs), and which connections *are not* found to be significant that we *did* expect to be significant (i.e. missing TPs, or false negatives). In this 15 node network analysis with the LPC not regressed, for ICOV and Partialcorr (the only models to not predict all connections as significant) the FPs are mostly within visual area e1-e5 connections, as well as one e1-e5 connections between V2 and V3. Missing TPs are V1-V3 adjacent eccentricity connections. Ideas regarding why these connections in particular may not agree with our inferred ground truth expectations can be found in the Discussion.

Part of the interaction measures we estimated showed significant negative values, that we retained in Supplemental Figure 3-B. The negative connections we see are almost exclusively those we have classified as unexpected connections. In all analyses, the bulk of the negative connections comprises between area connections e1-e5, e1-e4, e2-e5, and some e2-e4 connections. The interpretability of negative connections is addressed in the Discussion.

For further examination of the performance of each model, we draw the reader's attention to Figure 4 where we present the proportions of TPs (green; TPR or true positive rate) and proportions of FPs (red; FPR or false positive rate) associated with each model. In this figure, we also examine these proportions across a range of statistical thresholds. In presenting the proportion of TP and FP across a range of thresholds we can assess the impact of the threshold on the network prediction, and also the stability of the network prediction across a range of thresholds. In general, PartialCorr and ICOV show good optimal network predictions around lenient p-value thresholds (near the FDR p-value threshold associated with a maximum FPR of 0.05, p1).

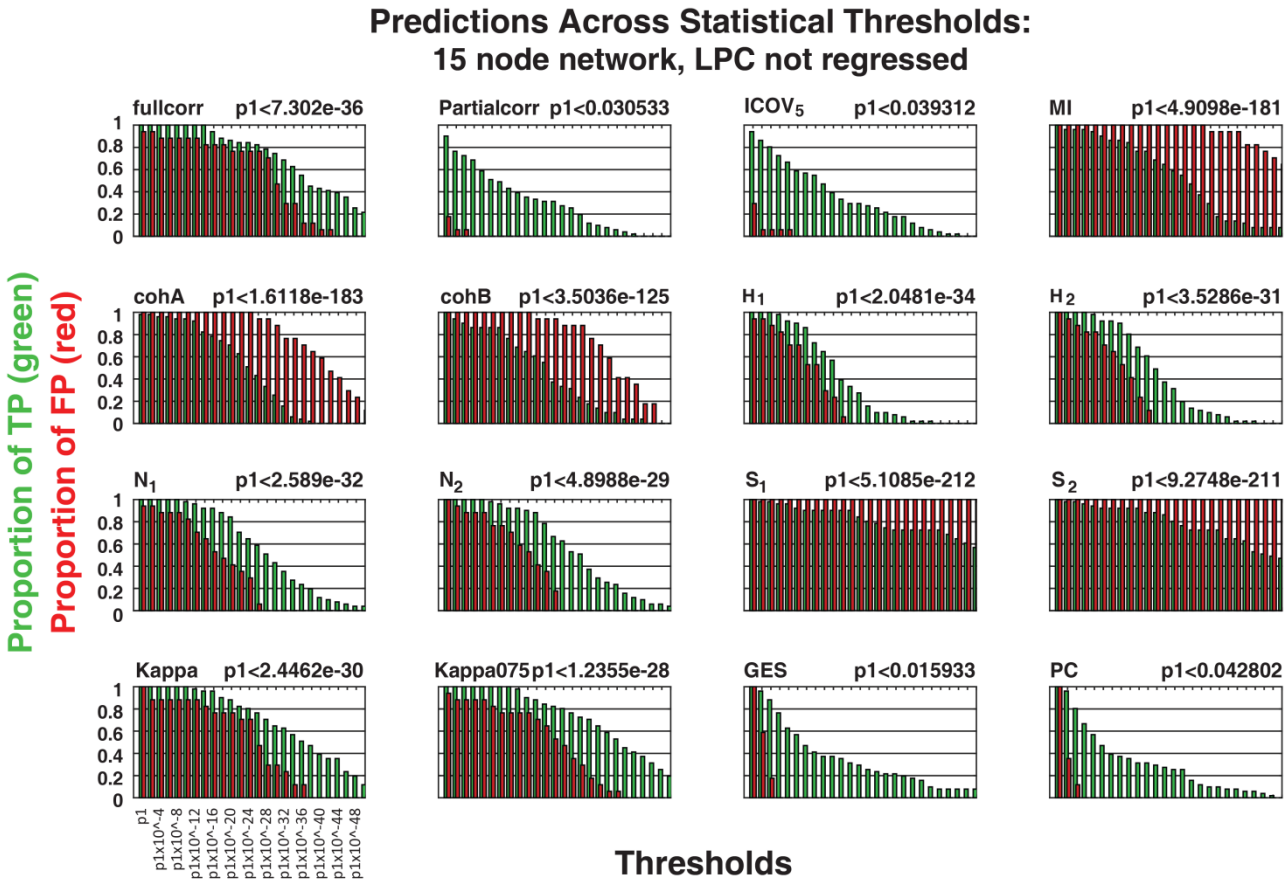


Fig. 4. Proportion of true and false positive connections found for each of the models tested at various statistical significance thresholds. The results shown here were obtained for the 15 node network within the quadrant with no LPC regression. There are 51 expected connections tested for and 17 unexpected connections. Green and red bars represent the true positive (TP) and the false positive (FP) predictions, respectively. The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The p-value thresholds become progressively smaller by a multiplication factor of 10^-2 at each step (up to a minimum of 10^-50). For example, Partial correlation's labels along the horizontal axis would read 0.03, 0.03x10^-2, 0.03x10^-4, 0.03x10^-6, 0.03x10^-8, … , 0.03x10^-50.

These plots (Figure 4) allow us to observe the stability of the TPR and FPR as the absolute p-value threshold is made more and more stringent. In these bar plots, we wish to see high TPR, low FPR and only a slow change in TPR and FPR as we increase the strictness of the threshold. The stability of the TPR and FPR depends on the range of thresholds. The absolute thresholds we selected decrease exponentially. With decreasing threshold (i.e., increasing

strictness), there is a trend of decreasing TPR and FPR across all models. We emphasize that one cannot evaluate the performance of a single model based on stability across thresholds because of the arbitrary nature of the incrementation between thresholds; however, we can compare the relative stabilities of the different models with the given incrementation.

Table 2: Threshold level of significant p-values that is necessary to maximize the proportion of True Positives minus the proportion of False Positives (TPR-FPR) for the 15 node network, ***LPC not regressed***. Second column: the FDR corrected alpha = 0.05 threshold. Third column: the optimal threshold. Fourth column: the optimal TPR-FPR, obtained with the optimal threshold. Fifth and sixth columns: The stability of the network prediction is assessed by evaluating the mean TPR-FPR across thresholds. TPR-FPR at the optimal density (given in column 7) is presented in column 8. The 10th column presents the proportion of triangles predicted by each model at alpha=0.05, FDR corrected. This is presented next to the proportion of all expected connections (i.e. TPR, 9th column). We also present the probability of a triangle being predicted given the probability of existence of each individual connection in the triangle (column 11). A value close to 1 should reflect minimal or no bias against triangle connectivity. Keep in mind that these values do not reflect the FP rate. Models which do not have a ~100% FP rate are identified in bold.

Good performance

Poor performance

Model name	p1: FDR corrected α=0.05	Threshold <=0.05 such that TPR-FPR maximized	TPR-FPR with optimal absolute threshold	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-20}]	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-50}]	Density such that TPR-FPR maximized	TPR-FPR with optimal density	α=0.05 TPR: All conn-ections	α=0.05 TPR: Triangles	P(Triangle | TPR)= [triangles ÷ (all conns)^3]
'Fullcorr'	7.30E-36	p1 E-36	0.43	0.11 ±0.04	0.20 ±0.13	52	0.76	1.00	1.00	1.00
'Partialcorr'	3.05E-02	p1	0.73	0.53 ±0.15	0.29 ±0.25	64	0.80	**0.90**	**0.71**	**0.97**
'ICOV_5'	3.93E-02	p1 E-2	0.80	0.58 ±0.14	0.32 ±0.26	64	0.84	**0.94**	**0.83**	**1.00**
'MI'	4.91E-181	p1	0.00	-0.11 ±0.08	-0.39 ±0.28	52	0.82	1.00	1.00	1.00
'CohA'	1.61E-183	p1	-0.02	-0.11 ±0.10	-0.33 ±0.24	60	0.82	0.98	1.00	1.00
'CohB'	3.50E-125	p1	0.00	-0.18 ±0.12	-0.29 ±0.21	56	0.84	1.00	1.00	1.00
'H_1'	2.05E-34	p1 E-16	0.35	0.21 ±0.11	0.13 ±0.13	48	0.71	1.00	1.00	1.00
'H_2'	3.53E-31	p1 E-22	0.49	0.25 ±0.16	0.18 ±0.16	48	0.69	1.00	1.00	1.00
'N_1'	2.59E-32	p1 E-26	0.53	0.22 ±0.14	0.23 ±0.16	52	0.75	1.00	1.00	1.00
'N_2'	4.90E-29	p1 E-26	0.63	0.20 ±0.14	0.25 ±0.18	48	0.71	1.00	1.00	1.00
'S_1'	5.11E-212	p1	0.00	-0.06 ±0.04	-0.18 ±0.13	60	0.86	1.00	1.00	1.00
'S_2'	9.27E-211	p1	0.00	-0.06 ±0.04	-0.21 ±0.17	56	0.80	1.00	1.00	1.00
'Kappa'	2.45E-30	p1 E-38	0.47	0.12 ±0.05	0.23 ±0.14	48	0.71	1.00	1.00	1.00
'Kappa075'	1.24E-28	p1 E-32	0.53	0.16 ±0.06	0.29 ±0.15	48	0.71	1.00	1.00	1.00
'GES'	1.59E-02	p1 E-6	0.76	0.46 ±0.21	0.29 ±0.21	52	0.78	1.00	1.00	1.00
'PC'	4.28E-02	p1 E-4	0.69	0.43 ±0.20	0.25 ±0.21	48	0.75	1.00	1.00	1.00

The stability of network predictions across statistical thresholds was also evaluated in Table 2 (columns 5 and 6) where we present the mean of TPR-FPR across a range of thresholds. These numbers are not meaningful in an absolute sense as discussed above, however they are informative in a relative sense, allowing us to compare between models. We present results from two threshold ranges, first from p1 to p1×10^-20, and then the entire range of thresholds tested, p1 to p1×10^-50. This mean TPR-FPR value can be viewed as an Area Under the Curve (AUC) type of measure, which represents the stability of the TPR-FPR measure over the range of

thresholds. The larger the value, the higher the TPR-FPR value over a longer range, and thus a better and more stable network prediction. A perfect network prediction would have a TPR-FPR of 1 (100% TPs, 0 FPs). Good models are highlighted in green in the table. For the 15 node network with LPC not regressed, ICOV performed best with mean TPR-FPR 0.58 (from p1 to p1×10^-20). PartialCorr trailed slightly behind with mean TPR-FPR of 0.53.

15-node network: Triangle Connectivity

We have also given consideration to the feature of triangle connectivity within the network. This is an important consideration due to concerns that regression models such as partial correlation and ICOV will eliminate triangles (Schmittmann et al., 2014). The elimination of triangle-connectivity is in some cases a positive feature of these models in densely connected networks, resulting in the elimination of indirect connections from the network prediction. That being said, if there is indeed true direct connectivity in a triangle configuration, we wish to retain these connections in our network prediction. There are 48 triangles in the set of expected connections we have defined (for example: V1e1-V1e2-V2e1-V1e1). We have examined the proportion of the total number of triangles that each model has predicted at the FDR p-value threshold associated with a maximum FPR of 0.05 (results presented in Table 2; column 10). Alongside the proportion of triangles we present the total proportion of predicted expected connections (TPR) for perspective (Table 2, column 9). Finally, we present the probability of obtaining a completed triangle (P(T)) given the probability of each of the individual connections (P(C)) within the triangle being present, P(T|C). This probability can be found in Table 2, column 11:

$$P(T|C) = \frac{P(T)}{P(C)^3}$$

The probability of a complete triangle is dependent on the TPR given that with fewer predicted expected connections, there will necessarily be fewer triangles. The closer this conditional probability is to 1, the less biased the model is toward eliminating triangles.

Our findings with regards to triangles show that there is no large bias observed in any of the models in terms of the elimination of triangles. The scaled probability is always (across all tests) over 88% for Partialcorr and ICOV, and is always over 94% for fullcorr, Kappa, and Kappa075. For example, in table 2 for the 15 node network without the LPC regressed, it is 97%

for Partialcorr and 100% for ICOV. Given these findings, in conjunction with the observations from our other analyses, we conclude that in a small-scaled, densely-connected network such as in the early visual cortex, regression models do not necessarily show a strong bias against triangle connectivity.

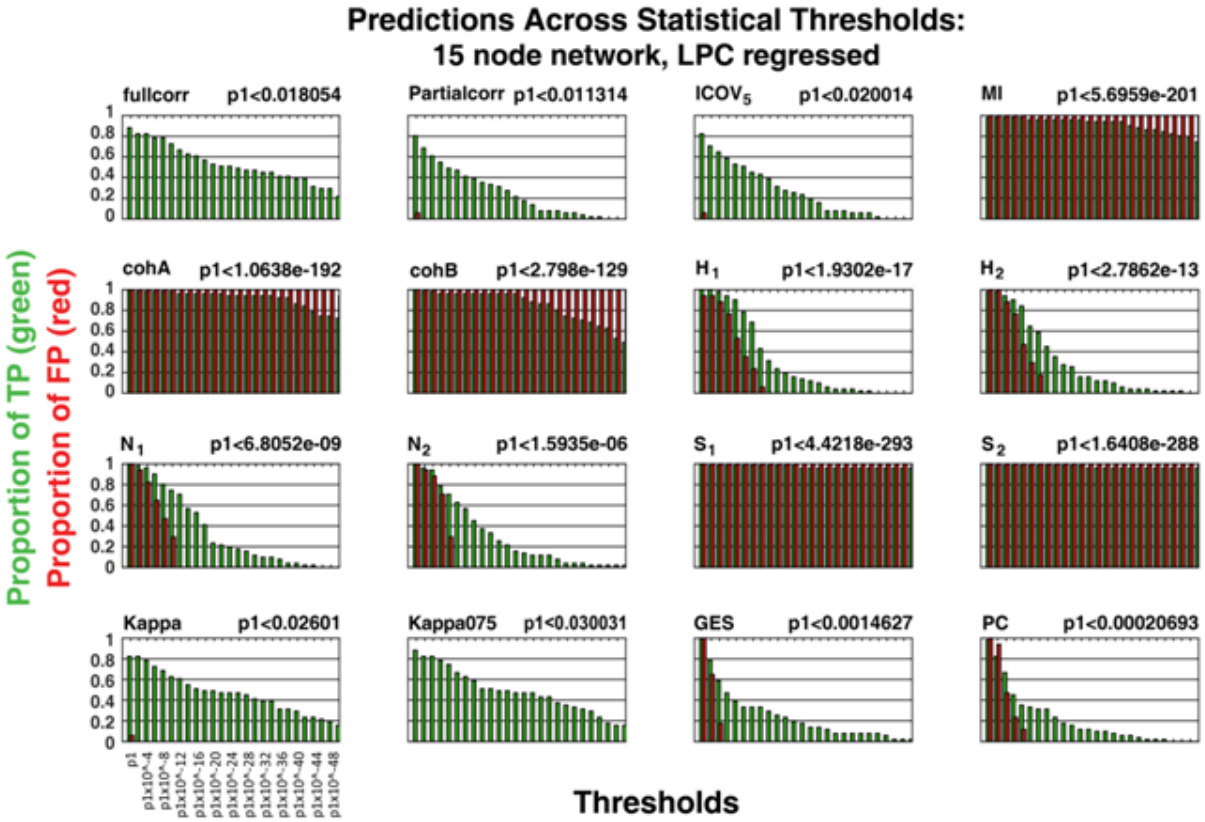


Fig. 5. Proportion of true and false positive connections found for each of the models tested at various statistical significance thresholds. The results shown here were obtained for the 15 node network within the quadrant. *The processing included regression of the LPC.* The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The threshold p-values become progressively smaller by a multiplication factor of 10^-2 at each step, as described for Figure 4. The format of presentation is as in Figure 4.

15-node network: LPC regression

Given that some of the models we test are not expected to be centered around zero, finding positive connectivity for nearly all pairs of nodes, we examined if these models would fare better with the regression of the LPC. We expect that this signal would reflect common fluctuations originating from network effects within and spatially wide common input from outside of the network of study. Therefore, regressing the LPC will leave only connectivity that

is more direct between specific nodes. Figure 5 shows that this is indeed the case in our 15-node network (compare with figure 4). Fullcorr, Kappa, and Kappa075 in particular show impressive improvements when the LPC is regressed as compared to the case where it is not, with few or no FPs and high proportions of TPs. Partialcorr and ICOV still perform well and have fewer FPs in the case where the LPC is regressed out, but there are more missed TPs as well (though this is not a drastic effect).

Since there are either nearly no FPs (using fullcorr, Partialcorr, ICOV, Kappa, Kappa075, GES and PC) or nearly all FPs found (using cohA, cohB, MI, S1, and S2), no trends are observed in which FPs are found. The TPs that are missing are V1-V3 adjacent eccentricity connections, as in the analysis without LPC regression, some V1-V3 same eccentricity connections, as well as V1-V2 adjacent eccentricity connections for ICOV and Partialcorr (see Supplemental Figure 9). Where triangles are concerned, once again we observe no strong bias against triangle connectivity in our multivariate methods (Partialcorr, ICOV, GES and PC; scaled probability of triangle predictions 88-100%), although good bivariate models (fullcorr and Kappa/Kappa075) have an even better triangle retention rate if the Local Principal Component is regressed from the ROI time-courses (scaled probability of triangle predictions 94-100%; see Table 3, column 11).

With LPC regression, GES and PC behave somewhat differently than the other top models. They in fact see *more* FPs, and not much change in the proportion of TPs. This may be explained by the method of computing GES and PC interaction measures. The output of the GES and PC models are directed graphs. The models determine whether a connection is present in the graph and then decide if the connection is forward, backward, bidirectional, or undirected. For our purposes, we are only interested in whether a connection is present, not the direction of the connectivity. Given this, we ignore the directionality predicted by GES and PC and simply note the presence or absence of a connection. We obtain a connection strength via the method used in Smith et al. (2011) and Dawson et al. (2013) using the sensitivity parameters input into the models. For the other models, a shift to decrease connectivity measures would result in negative connectivity values which we have chosen to omit given concerns around spurious anti-correlations. For GES and PC, these “negative connections” would still be reflected by positive magnitudes. Thus, the increased proportion of FP for GES and PC following LPC regression is likely a reflection of spurious “anti-correlations”.

Table 3: Threshold level of significant p-values that is necessary to maximize the proportion of True Positives minus the proportion of False Positives (TPR-FPR) for the 15 node network, ***LPC regressed***. Second column: the FDR corrected alpha = 0.05 threshold. Third column: the optimal threshold. Fourth column: the optimal TPR-FPR, obtained with the optimal threshold. Fifth and sixth columns: The stability of the network prediction is assessed by evaluating the mean TPR-FPR across thresholds. TPR-FPR at the optimal density (given in column 7) is presented in column 8. The 10th column presents the proportion of triangles predicted by each model at alpha=0.05, FDR corrected. This is presented next to the proportion of all expected connections (i.e. TPR, 9th column). We also present the probability of a triangle being predicted given the probability of existence of each individual connection in the triangle (column 11). A value close to 1 should reflect minimal or no bias against triangle connectivity. Keep in mind that these values do not reflect the FP rate. Models which do not have a ~100% FP rate are identified in bold.

Good performance

Poor performance

Model name	p1: FDR corrected α=0.05	Threshold <=0.05 such that TPR-FPR maximized	TPR-FPR with optimal absolute threshold	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-20}]	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-50}]	Density such that TPR-FPR maximized	TPR-FPR with optimal density	α=0.05 TPR: All conn-ections	α=0.05 TPR: Triangles	P(Triangle | TPR)= [triangles ÷ (all conns)^3]
'Fullcorr'	1.81E-02	p1	0.88	0.71 ±0.12	0.53 ±0.18	76	0.90	**0.88**	**0.65**	**0.94**
'Partialcorr'	1.13E-02	p1	0.75	0.49 ±0.15	0.25 ±0.23	72	0.86	**0.80**	**0.46**	**0.88**
'ICOV_5'	2.00E-02	p1	0.76	0.51 ±0.16	0.26 ±0.24	68	0.84	**0.82**	**0.50**	**0.90**
'MI'	5.70E-201	p1	-0.02	-0.03 ±0.01	-0.08 ±0.07	32	0.61	0.98	1.00	1.00
'CohA'	1.06E-192	p1	0.00	-0.03 ±0.01	-0.08 ±0.08	28	0.51	1.00	1.00	1.00
'CohB'	2.80E-129	p1	0.00	-0.03 ±0.01	-0.15 ±0.15	32	0.61	1.00	1.00	1.00
'H_1'	1.93E-17	p1 E-12	0.45	0.25 ±0.15	0.13 ±0.14	32	0.51	1.00	1.00	1.00
'H_2'	2.79E-13	p1 E-14	0.45	0.24 ±0.17	0.14 ±0.15	32	0.59	0.98	1.00	1.00
'N_1'	6.81E-09	p1 E-12	0.71	0.33 ±0.22	0.19 ±0.20	40	0.59	1.00	1.00	1.00
'N_2'	1.59E-06	p1 E-10	0.63	0.29 ±0.22	0.17 ±0.18	32	0.59	1.00	1.00	1.00
'S_1'	4.42E-293	p1	0.00	-0.02 ±0.01	-0.01 ±0.01	40	0.73	1.00	1.00	1.00
'S_2'	1.64E-288	p1	0.00	-0.02 ±0.01	-0.02 ±0.02	48	0.67	1.00	1.00	1.00
'Kappa'	2.60E-02	p1 E-2	0.82	0.64 ±0.12	0.46 ±0.19	72	0.86	**0.82**	**0.56**	**1.02**
'Kappa075'	3.00E-02	p1	0.88	0.68 ±0.14	0.48 ±0.21	76	0.90	**0.88**	**0.65**	**0.94**
'GES'	1.46E-03	p1 E-6	0.47	0.29 ±0.13	0.17 ±0.14	36	0.51	1.00	1.00	1.00
'PC'	2.07E-04	p1 E-10	0.33	0.19 ±0.14	0.11 ±0.12	36	0.61	0.98	1.00	1.00

Following LPC regression, the majority of the negative connectivity measures are between area e1-e5, e1-e4, e2-e5, and some e2-e4 connections; this trend is similar to the results with no LPC regression (Suppl. Fig. 9-B). In addition, two of V1-V3 adjacent eccentricity connections show statistical significance with a negative mean following LPC regression. In general, more negative connections are present following LPC regression for Partialcorr, ICOV, fullcorr, and Kappa. Note that fullcorr and Kappa have only positive connections without LPC regression (Suppl. Fig. 3-B). See the Discussion on the interpretability of negative interaction measures.

The stability of the model predictions across statistical thresholds in the 15 node network with the LPC regressed is presented in Table 3, using same format presented in Table 2 for the previous analysis. Fullcorr showed the highest TPR-FPR values: 0.71 in the smaller range of thresholds (column 5) and 0.53 in the full range of thresholds (column 6). In fact, these are the highest measures of stability achieved in models with good performance across all tests using our data (including with and without LPC regression).

Overall, Fullcorr, Kappa, Kappa075, Partialcorr, and ICOV perform very well with the LPC regressed, nearly always showing their best performance at the FDR p-value threshold. In fact, these results are better than without the LPC regressed. The best performances among all tests we undertook are from Fullcorr and Kappa075 in the 15 node network with LPC regressed.

15-node network: Density-based thresholding

The approach to selecting a threshold in FC analyses is not straightforward. This is particularly true in studies where one has no prior knowledge of the system. It is for this reason that we have chosen to focus our analyses on statistical testing at the FDR p-value threshold associated with a maximum FPR of 0.05, a typical standard in significance testing. We were interested, however, in how much of an impact the choice of threshold would have on the network prediction. In order to examine this, we have covered the 2 main strategies for thresholding connectivity measures: 1) statistical hypothesis testing and absolute thresholds, and 2) density-based thresholds, relative to the number of all possible connections in the network. Density-based results are presented in Figures 6 and 7, and in the supplemental material. Similar to the way we evaluated the performance of the models with statistical hypothesis testing, we can identify good performance of the models via inspection of where the TPR is highest for no or few FPs. Across the thresholds tested, the density-based network predictions appear relatively stable, the green bars often showing a stable plateau before dropping off as the threshold becomes increasingly strict. The top models in our statistical threshold analyses (Fullcorr, Partialcorr, ICOV, Kappa, and Kappa075) also show steep reduction in FP with decreasing density, resulting in network predictions with high TPs and low FPs at a certain connection densities, especially when the LPC is regressed (Figure 7).

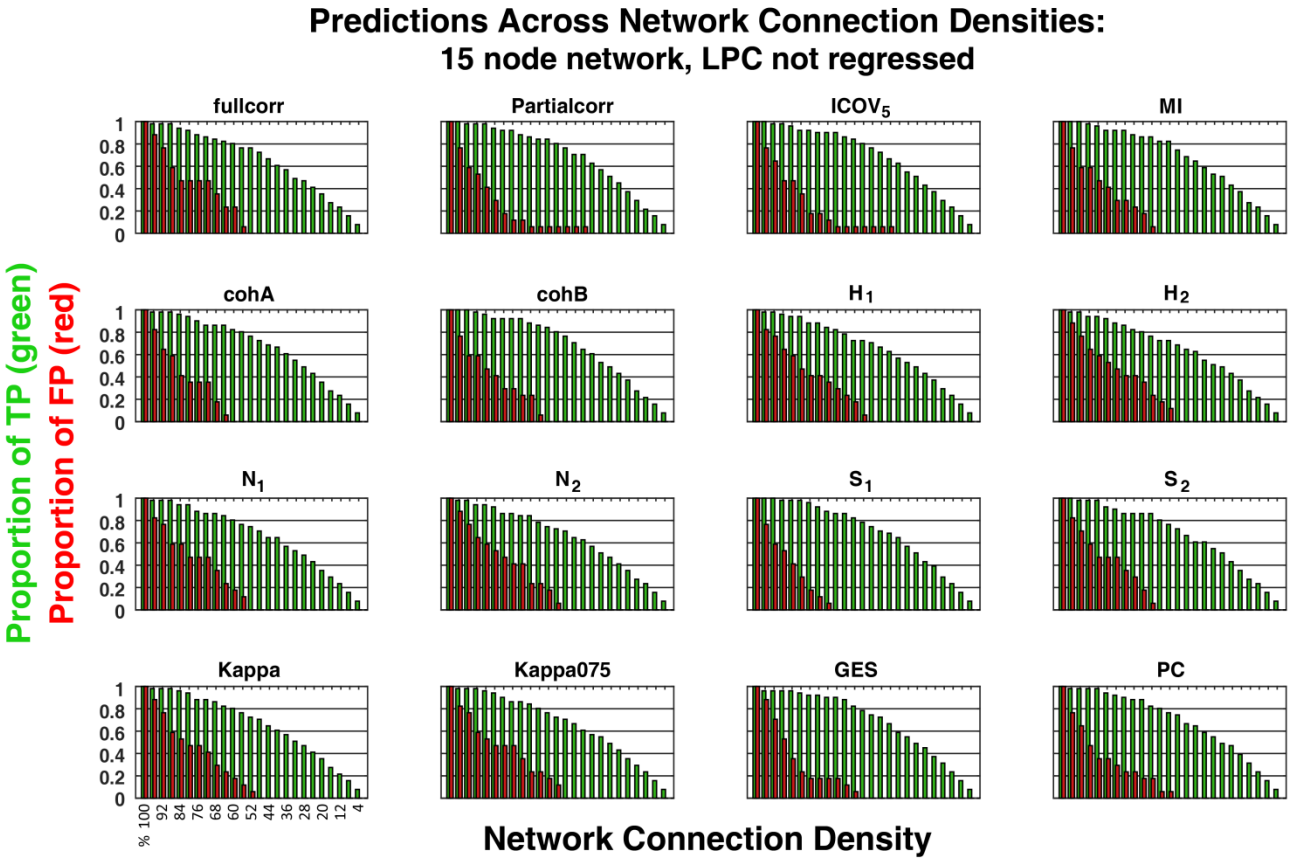


Fig. 6. Proportion of true and false positive connections found for each of the models tested at various *network connection densities*. The results shown here were obtained for the 15 node network within the quadrant with no LPC regression. There are 51 expected connections tested for and 17 unexpected connections. Green and red bars represent the true positive (TP) and the false positive (FP) predictions, respectively. The network connection densities along the horizontal axis begin at 100% and go down by 4% until reaching 0%.

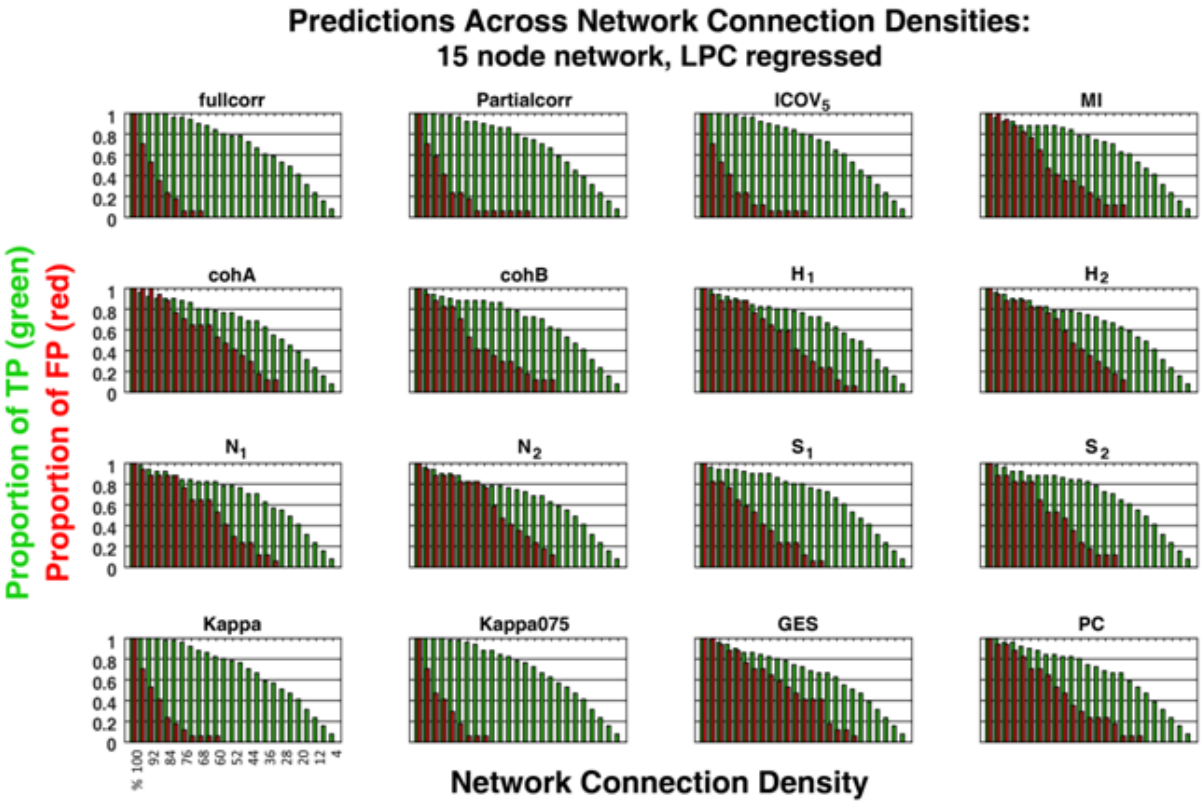


Fig. 7. Proportion of true and false positive connections found for each of the models tested at various *network connection densities*. The results shown here were obtained for the 15 node network within the quadrant. *The processing included regression of the LPC*. There are 51 expected connections tested for and 17 unexpected connections. The network connection densities along the horizontal axis begin at 100% and go down by 4% until reaching 0%. The format of presentation is as in Figure 6.

When density-based thresholding is used, we see on inspection of the figures that the optimal performance achieved by each model is either much better (LPC not regressed - Figure 6), or slightly better (LPC regressed - Figure 7) than the optimal performance for absolute statistical thresholding (Figures 4 and 5 for LPC not regressed and LPC regressed, respectively). We quantify this performance using the mean TPR minus the mean FPR as we did with the statistical thresholding (see Tables 2 and 3, column 4: statistical thresholding, and column 8: density-based thresholding). For example, from Table 2, ICOV has a maximal TPR-FPR of 0.80 in the 15 node analysis without LPC regression when statistical thresholding, and the same analysis has a maximal TPR-FPR of 0.84 when density-based thresholding is applied. When the LPC is regressed (Table 3), ICOV's TPR-FPR goes from a maximum of 0.51 with statistical

thresholding to a maximum of 0.84 with density-based thresholding; Fullcorr's TPR-FPR (LPC regressed) changes from 0.71 to 0.90, an excellent result.

Looking more closely at the bivariate models, including Fullcorr, Kappa, MI, cohA, cohB, and the GenSynch models (H, N, and S), the majority of these models are unusable with no LPC regression and statistical thresholding, however, they perform far better with density based thresholding (compare Figures 4 and 5 to Figures 6 and 7). In fact, S1 is a top performer with density-based thresholding and without LPC regression achieving 0.86 TPR-FPR (Table 2, column 8) at its optimal density. Except for Fullcorr and Kappa, these models perform well with density-based thresholding without LPC regression. Fullcorr and Kappa achieve better results with LPC regression. Where multivariate models are concerned, Partialcorr and ICOV perform well in either case (LPC regressed or not) with density-based thresholding, but do show some subtle improvements in their performance when the LPC is regressed.

While the density-based thresholds can generate more accurate network predictions when compared to networks generated using absolute thresholds (at optimal density) the difficulty still remains that there is no way to decide on the "best", or even a "good" threshold without prior knowledge of the system. The best density-based threshold is also not consistent across different tests including different network complexities or whether the LPC is regressed (see supplemental figures). However, with absolute statistical thresholding, several models are often at or near their optimal performance at the FDR p-value threshold associated with a maximum FPR of 0.05. This is very useful as we do not need to try to improve upon this threshold with arbitrarily decided strictness, especially in cases in which we do not have any information on the network. For example, in the 15 node network case with LPC regression and statistical hypothesis-based thresholding, Fullcorr and Kappa both achieve 0.88 TPR-FPR (Table 3, column 4) with zero FPs. This result is comparable to the best density-based outcome (TPR-FPR=0.90), but at the convenient and conventional statistical significance level of 0.05. In contrast, density-based thresholding in an unknown network requires information on the optimal density specific to the network at hand, which in the majority of the cases is unknown.

Analysis with a large sample size – Human Connectome Project data: Main results

Next, we test the effect of increasing the sample size on the results of our study. We use 220 subjects from the Human Connectome Project with the goal of achieving 95% or more power in each analysis undertaken (the power of the analysis is presented in Supplemental Figures 23, 27). Figures 8-11 (and Supplemental Figures 22-29) show the results of the HCP analyses in the 15 node network. We present only Fullcorr, Kappa, Kappa075, Partialcorr, ICOV, GES and PC as these models were most successful with our own small sample size data.

Figure 8 presents the results obtained without LPC regression, in a format similar to that used for presenting Figure 4. Compared to the results obtained with the small sample size (Fig. 4), using the large number of data sets increased the proportions of both TP and FP obtained by all models at any threshold level. The high proportion of FPs is likely the result of increased statistical significance due to more abundant data. The increase in FPR renders these models practically unusable. This includes both bivariate models (e.g., Fullcorr and Kappa) and multivariate models (e.g., Partialcorr, ICOV and Bayes nets GES and PC). In contrast, with the LPC regressed (Figure 9), Fullcorr, Kappa, Kappa075, Partialcorr and ICOV show very impressive results with nearly perfect network predictions at their FDR p-value threshold associated with a maximum FPR of 0.05. No FPs are found (except for a small number in Kappa075) and the proportion of TPs found is very high in all cases. The connectivity matrices for these 5 models resemble the inferred ground truth matrix remarkably well (see Supplemental Figure 27).

Analysis with a large sample size: Triangle connectivity

We next tested the effect of increasing the number of data sets on triangle connectivity. Do the models show improved retention of triangles with the greater sample size in our HCP analyses, as they have shown improved prediction rates of expected connections? In all analyses, given that there are more significant connections, we indeed see more entire triangles predicted (if they were not already all predicted). See Tables 6-7, columns 10 and 11, in the supplemental material for details. The adjusted probability of triangle retention of the models is 91% at its lowest, and is most often in the range of 97-100%.

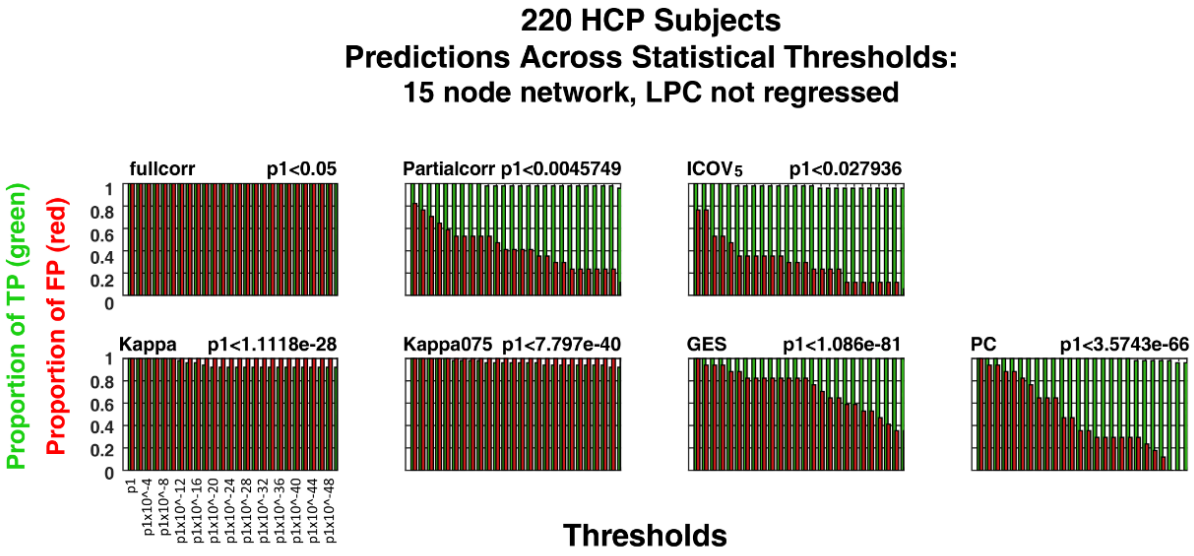


Fig. 8. Proportion of true and false positive connections found for each of the models tested at various statistical significance thresholds. The results shown here were obtained for the 15 node network within the quadrant with no LPC regression *using data from 220 HCP subjects*. The format of presentation is as in Figure 4.

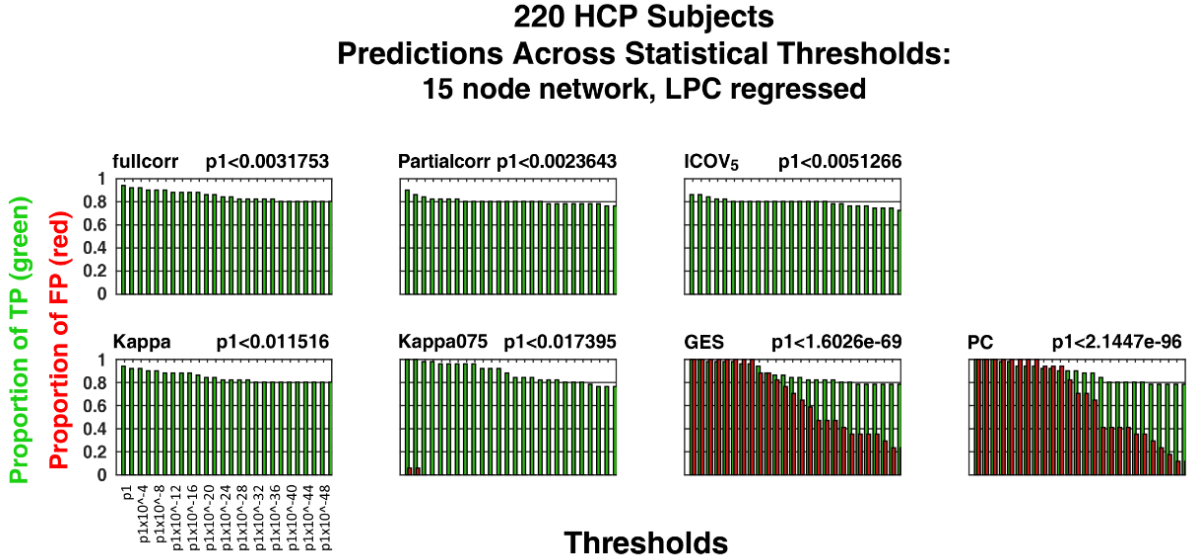


Fig. 9. Proportion of true and false positive connections found for each of the models tested at various statistical significance thresholds. The results shown here were obtained for the 15 node network within the quadrant *using data from 220 HCP subjects*. *The processing included regression of the LPC.* The format of presentation is as in Figure 4.

Analysis with a large sample size: Density-based thresholding

We next present the results of our HCP 15 node analyses with density-based thresholding (Figures 10 and 11). Although the sample size is much larger, the results are largely the same as those found in the analogous analysis with our small-sample data. There are somewhat steeper drops in the FP distribution across thresholds and somewhat more stable TP distributions near the optimal network density. These slightly improved results are reflected in Supplemental Tables 6 and 7, column 8, where the maximal TPR-FPR is as high at 0.96 in these analyses. Recall that the highest TPR-FPR was 0.90 in the analyses with our data (fullcorr, LPC regressed). The results obtained with density-based thresholding following LPC regression (Fig. 11) are not qualitatively different than those obtained without LPC regression (Fig. 10). In summary, unlike the results obtained with statistical hypothesis testing following LPC regression (Fig. 9), density-based thresholding does not converge to reflecting the direct connectivity, even when the data sample size increases substantially.

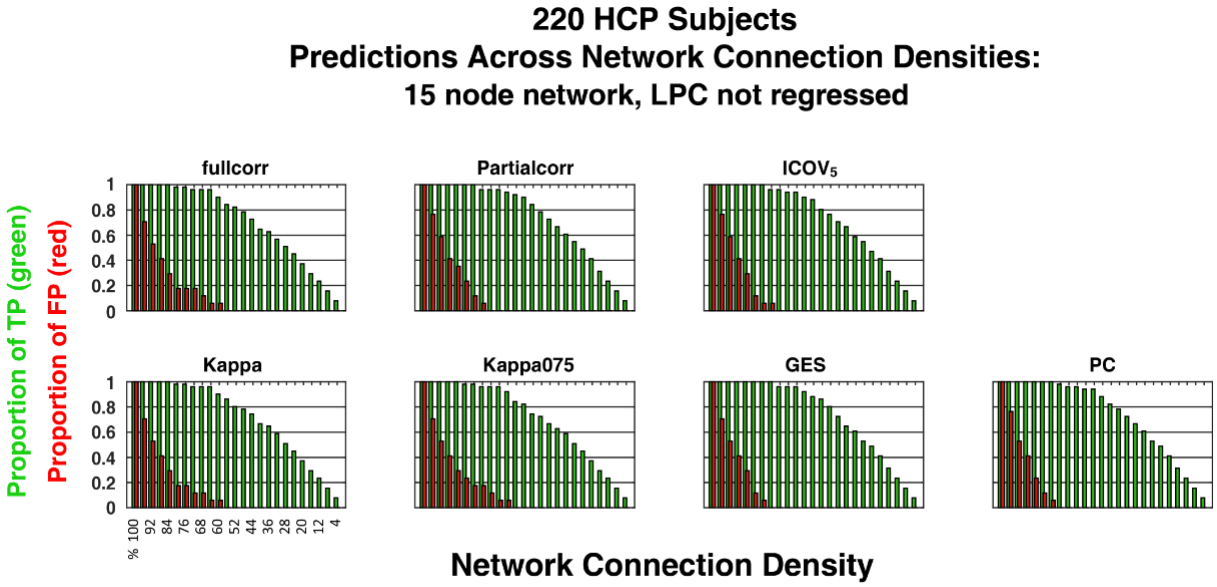


Fig. 10. Proportion of true and false positive connections found for each of the models tested at various *network connection densities*. The results shown here were obtained for the 15 node network within the quadrant with no LPC regression *using data from 220 HCP subjects*. The network connection densities along the horizontal axis begin at 100% and go down by 4% until reaching 0%. The format of presentation is as in Figure 6.

Analysis with a large sample size: a random sampling of Human Connectome Project data for comparisons with our small-size data set

The results we obtained with the HCP data were qualitatively different than those obtained by our small sample data set. We next sought to test whether this difference was caused by the difference in the size of data-sets, or possibly due to some other differences such as data quality or the use of the template to define the ROIs rather than retinotopy. To this end, we have conducted the same analyses using only 6 randomly selected HPC subjects (4 resting-state runs each), an equl number of subjects to that used in our small size sample. These results are presented in the supplemental material, Supplemental Figures 38-53. Overall, the results are comparable between the 6 subjects tests with our data and the tests with the 6 random HCP subjects. In some of the tests, for example Fullcorr in the 15 node analysis with LPC regressed, the HCP data networks are somewhat less stable (sharper decreases in the barplots; compare Supplemental Figures 10 - our data, and 44 - HCP). This could possibly be due to the fact that there are more runs per subject with our data, allowing for slightly more convergence toward the "true" network. Given the similar results obtained in these two sets of analyses with 6 subjects, we concluded that the improvements we see in the results with 220 HCP subjects can indeed be attributed to the large sample size, and that more data leads to models such as Fullcorr, Kappa, Partialcorr and ICOV to converge to the "true" network when statistical thresholding is used and the LPC is regressed. The same cannot be said on connection density-based thresholding.

.

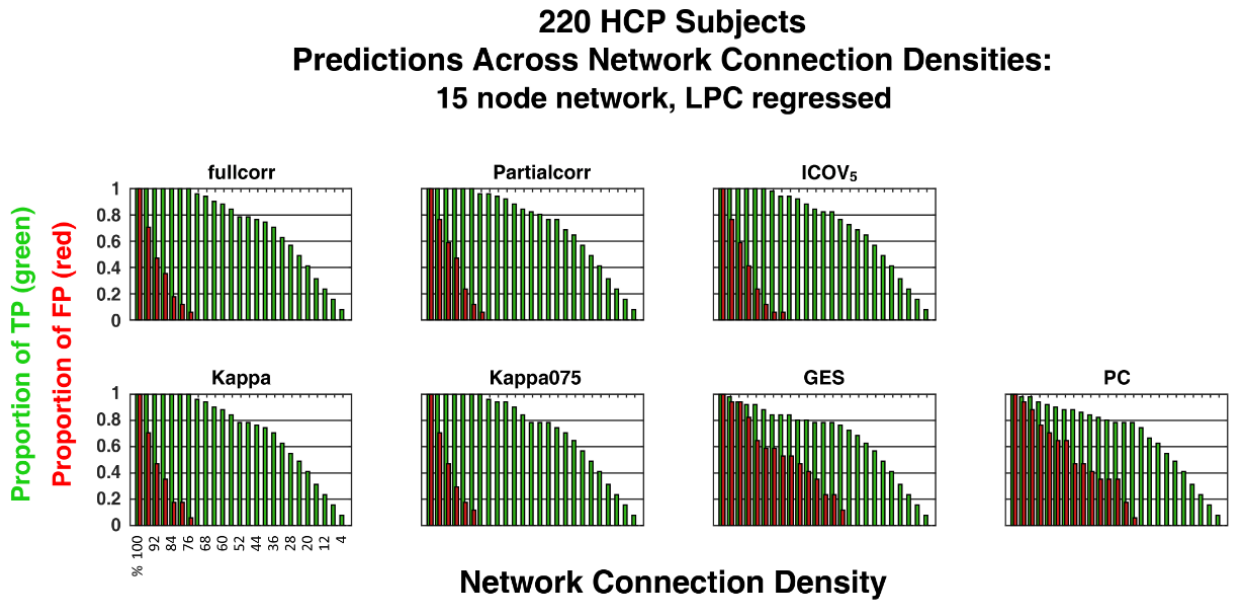


Fig. 11. Proportion of true and false positive connections found for each of the models tested at various *network connection densities*. The results shown here were obtained for the 15 node network within the quadrant *using data from 220 HCP subjects*. *The processing included regression of the LPC*. The network connection densities along the horizontal axis begin at 100% and go down by 4% until reaching 0%. The format of presentation is as in Figure 6.

60-node network: Main results

The next set of analyses we will present are those involving the whole visual cortex as a network, combining all four quadrants, resulting in 60 nodes total. These analyses will allow us to determine the impact of network complexity on the model outputs, and to determine which models are most useful when studying networks with more nodes. Figure 12 shows the networks predicted by each model within the visual cortex given an FDR p-value threshold associated with a maximum FPR of 0.05 in the 60 node network without regressing the LPC. Most models find that all ROIs within the network are connected, as in the 15-node network case (Fig. 3). Once again, Partialcorr and ICOV are among the exceptions showing some discrimination between expected and unexpected connections, this time along with PC and GES. Across these more successful models, connectivity is more abundant within a quadrant than between quadrants. In addition, we see a strong trend for connectivity between ROIs with similar eccentricities, as we did in the 15-node network (along the diagonal in each small black square). Only a few FPs were observed, with no specific trend. Expected connections that were missed were largely those

between V1 and V3 and between adjacent-eccentricity ROIs. The models retained triangles to a comparable degree in both the 15 and 60 node network analyses (See Tables 4 and 5, column 11). As in the 15 node network, negative interaction measures were observed between area e1-e5, e1-e4, e2-e5, and some e2-e4 connections (see Supplemental Figure 15B).

Figure 13 shows the proportion of TPs (green) and of FPs (red) predicted by each model across varying statistical thresholds. Overall, the level of statistical significance in the 60 node network is lower compared to the 15 node network. This reflects the smaller size of data sets: here, one time series comprises data obtained simultaneously from all 4 quadrants. In the 15 node analysis, the same data is split to 4 separate time series. In addition, the thresholding is more stringent due to correction for a larger number of comparisons. The decrease in statistical significance is even greater for the multivariate methods (ICOV, Partialcorr, GES, and PC). This is likely due to the fact that they account for more information from more nodes via regression or other approaches, removing the influence of additional common inputs (direct or indirect) from these nodes' time-courses as well as introducing more noise into the system.

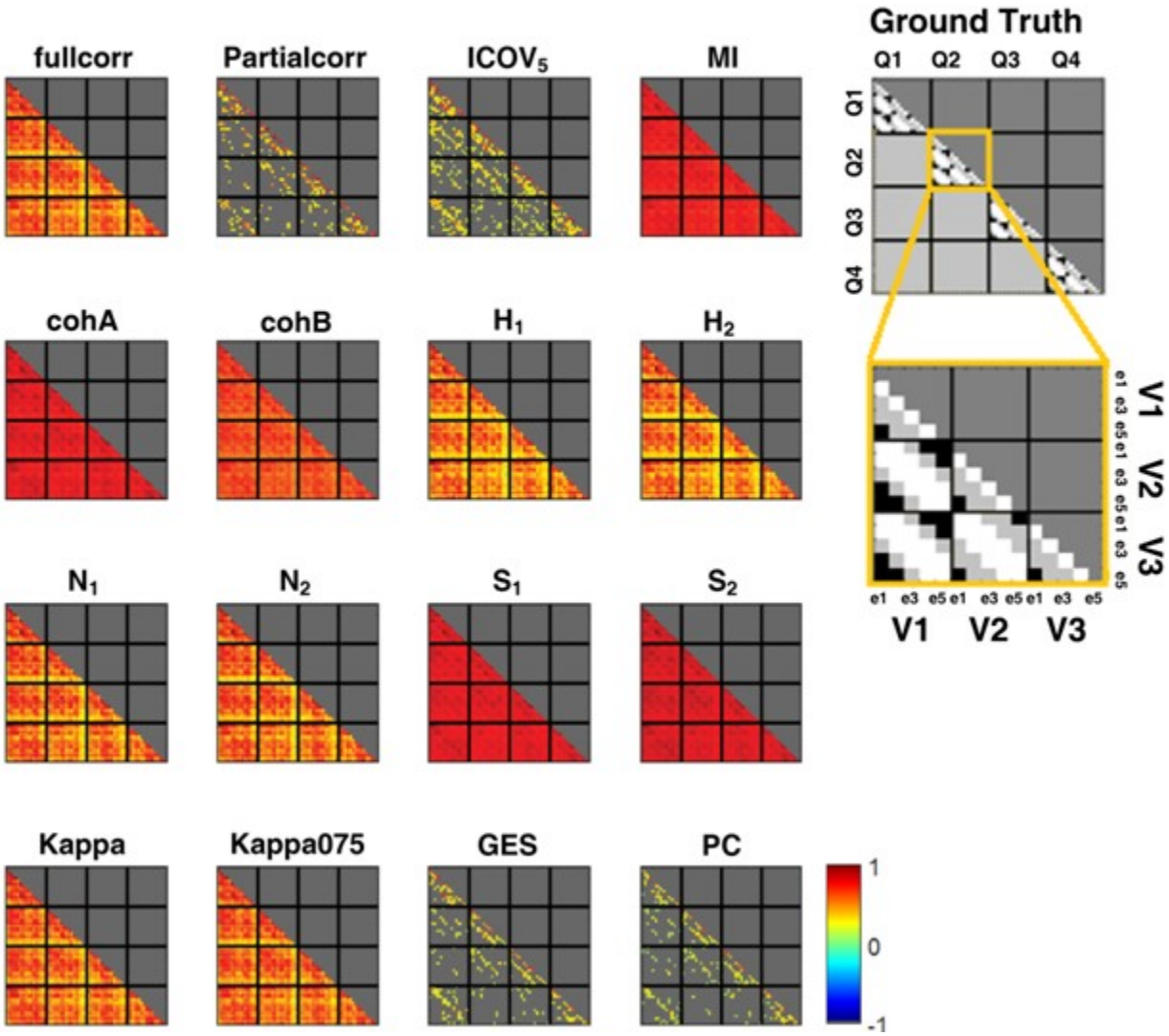


Fig. 12. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 60 nodes from all 4 quadrants of the complete visual cortex network with no LPC regression. Normalization was achieved by dividing all mean measures by the maximal value. The "Ground Truth" connectivity matrix in the upper right-hand corner has been constructed based on knowledge of the monkey visual cortex, and includes the expected connections (white) and the unexpected connections (black) presented in Figure 1. Light gray squares in the ideal matrix are connections on which we placed no expectations. Statistical significance was tested against the FDR p-value threshold given a maximum FPR of 0.05. Q=Quadrant. The format of presentation is as in Figure 3.

In the 60 node network, the TPR-FPR mean values across a range of thresholds are lower than in the 15 node network, in part reflecting lower TPR to begin with, but also reduced stability (Table 4, columns 5 and 6). The models show less stability with the 60 nodes in the network as compared to only 15 nodes. ICOV performs best in the case with LPC not regressed at 0.20 TPR-FPR (from p1 to p1×10^-20).

Table 4: Threshold level of significant p-values that is necessary to maximize the proportion of True Positives minus the proportion of False Positives (TPR-FPR) for the 60 node network, ***LPC not regressed***. Second column: the FDR corrected alpha = 0.05 threshold. Third column: the optimal threshold. Fourth column: the optimal TPR-FPR, obtained with the optimal threshold. Fifth and sixth columns: The stability of the network prediction is assessed by evaluating the mean TPR-FPR across thresholds. TPR-FPR at the optimal density (given in column 7) is presented in column 8. The 10th column presents the proportion of triangles predicted by each model at alpha=0.05, FDR corrected. This is presented next to the proportion of all expected connections (i.e. TPR, 9th column). We also present the probability of a triangle being predicted given the probability of existence of each individual connection in the triangle (column 11). A value close to 1 should reflect minimal or no bias against triangle connectivity. Keep in mind that these values do not reflect the FP rate. Models which do not have a ~100% FP rate are identified in bold.

Good performance

Poor performance

Model name	p1: FDR corrected α=0.05	Threshold <=0.05 such that TPR-FPR maximized	TPR-FPR with optimal absolute threshold	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-20}]	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-50}]	Density such that TPR-FPR maximized	TPR-FPR with optimal density	α=0.05 TPR: All conn-ections	α=0.05 TPR: Triangles	P(Triangle | TPR)= [triangles ÷ (all conns)^3]
'Fullcorr'	3.70E-02	p1 E-20	0.27	0.10 ±0.10	0.06 ±0.09	32	0.50	1.00	1.00	1.00
'Partialcorr'	7.74E-03	p1	0.62	0.16 ±0.21	0.07 ±0.15	40	0.68	**0.62**	**0.25**	**1.05**
'ICOV_5'	1.33E-02	p1	0.70	0.20 ±0.24	0.08 ±0.18	32	0.72	**0.77**	**0.45**	**0.98**
'MI'	1.44E-44	p1	0.00	-0.20 ±0.16	-0.10 ±0.14	40	0.60	1.00	1.00	1.00
'CohA'	6.91E-45	p1	0.00	-0.14 ±0.12	-0.06 ±0.10	44	0.63	1.00	1.00	1.00
'CohB'	1.09E-29	p1 E-2	0.00	-0.12 ±0.10	-0.06 ±0.09	36	0.59	1.00	1.00	1.00
'H_1'	3.96E-07	p1 E-10	0.29	0.10 ±0.11	0.04 ±0.08	44	0.51	1.00	1.00	1.00
'H_2'	5.03E-07	p1 E-10	0.34	0.13 ±0.14	0.06 ±0.11	44	0.49	1.00	1.00	1.00
'N_1'	4.16E-07	p1 E-10	0.41	0.17 ±0.14	0.07 ±0.12	36	0.52	1.00	1.00	1.00
'N_2'	3.17E-06	p1 E-10	0.48	0.19 ±0.16	0.08 ±0.14	48	0.49	1.00	1.00	1.00
'S_1'	8.16E-50	p1 E-48	0.01	-0.11 ±0.11	-0.11 ±0.11	40	0.72	1.00	1.00	1.00
'S_2'	5.85E-52	p1	0.00	-0.20 ±0.19	-0.19 ±0.17	32	0.55	1.00	1.00	1.00
'Kappa'	3.35E-03	p1 E-16	0.29	0.13 ±0.11	0.07 ±0.10	36	0.49	1.00	1.00	1.00
'Kappa075'	3.43E-02	p1 E-16	0.36	0.17 ±0.14	0.09 ±0.12	44	0.47	1.00	1.00	1.00
'GES'	7.07E-03	p1	0.53	0.13 ±0.17	0.06 ±0.13	28	0.62	**0.53**	**0.14**	**0.89**
'PC'	6.84E-03	p1	0.52	0.11 ±0.16	0.05 ±0.12	36	0.64	**0.54**	**0.13**	**0.83**

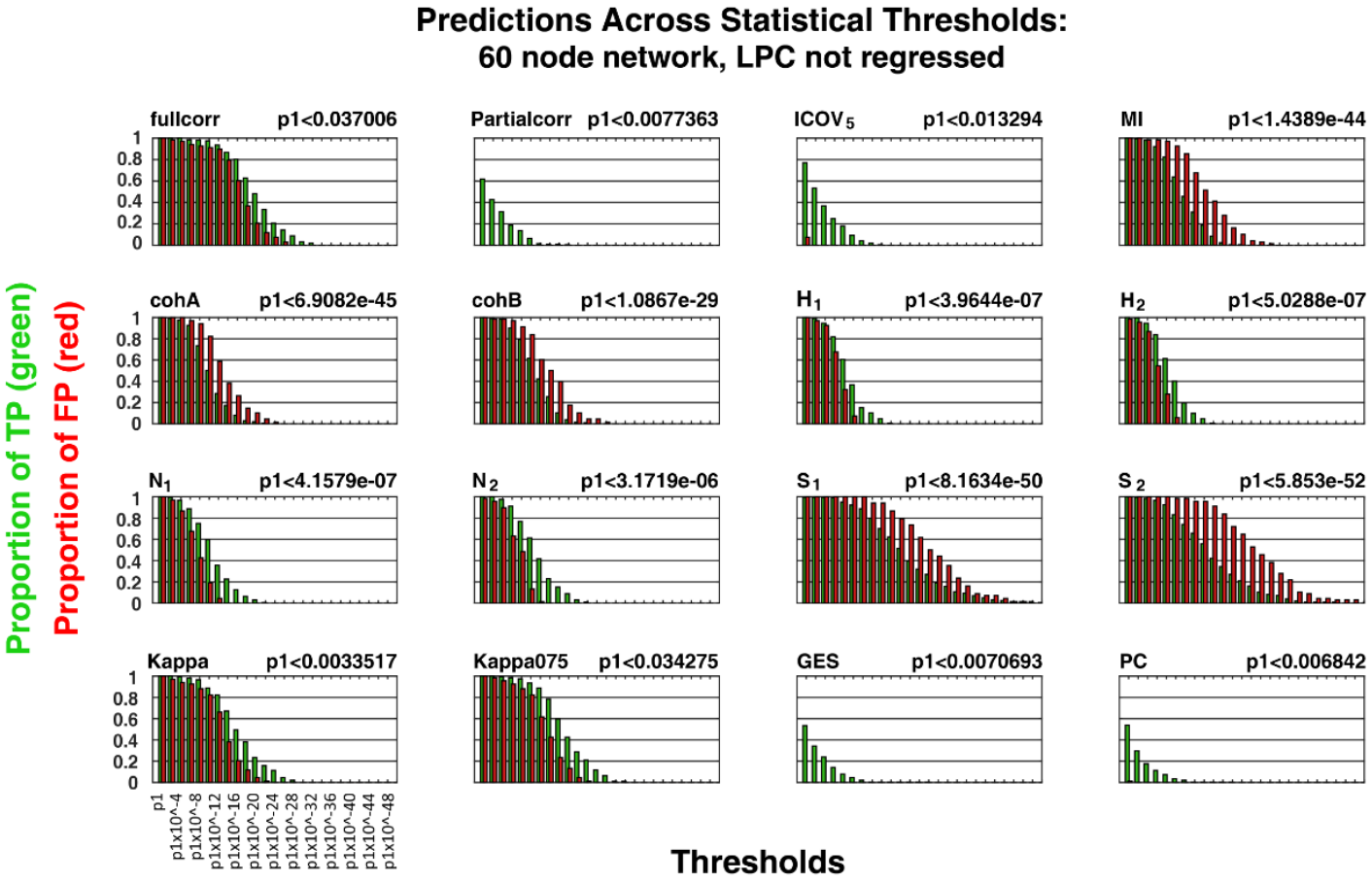


Fig. 13. Proportion of true and false positive connections found for each of the models tested at various statistical significance thresholds. The results shown here were obtained for the 60 nodes from all 4 quadrants of the complete visual cortex network with no LPC regression. There are 51×4=204 expected connections tested for and 17×4=68 unexpected connections. Green and red bars represent the true positive (TP) and the false positive (FP) predictions, respectively. The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The p-value thresholds become progressively smaller by a multiplication factor of 10^-2 at each step (up to a minimum of 10^-50). The format of presentation is as in Figure 4.

60-node network: Comparison to a 15-node network with the same sample size

Did the increased complexity of the network or the smaller sample size underlie the difference in stability across thresholds? Recall that our 15 node network is representative of a quadrant of the visual cortex, or ¼ of the data from the entirety of V1, V2, and V3 in the visual cortex. This means that when we take the individual quadrants as separate data sets, we have 4 times the number of data sets in the 15 node analysis compared to the 60 node, whole visual cortex analysis.

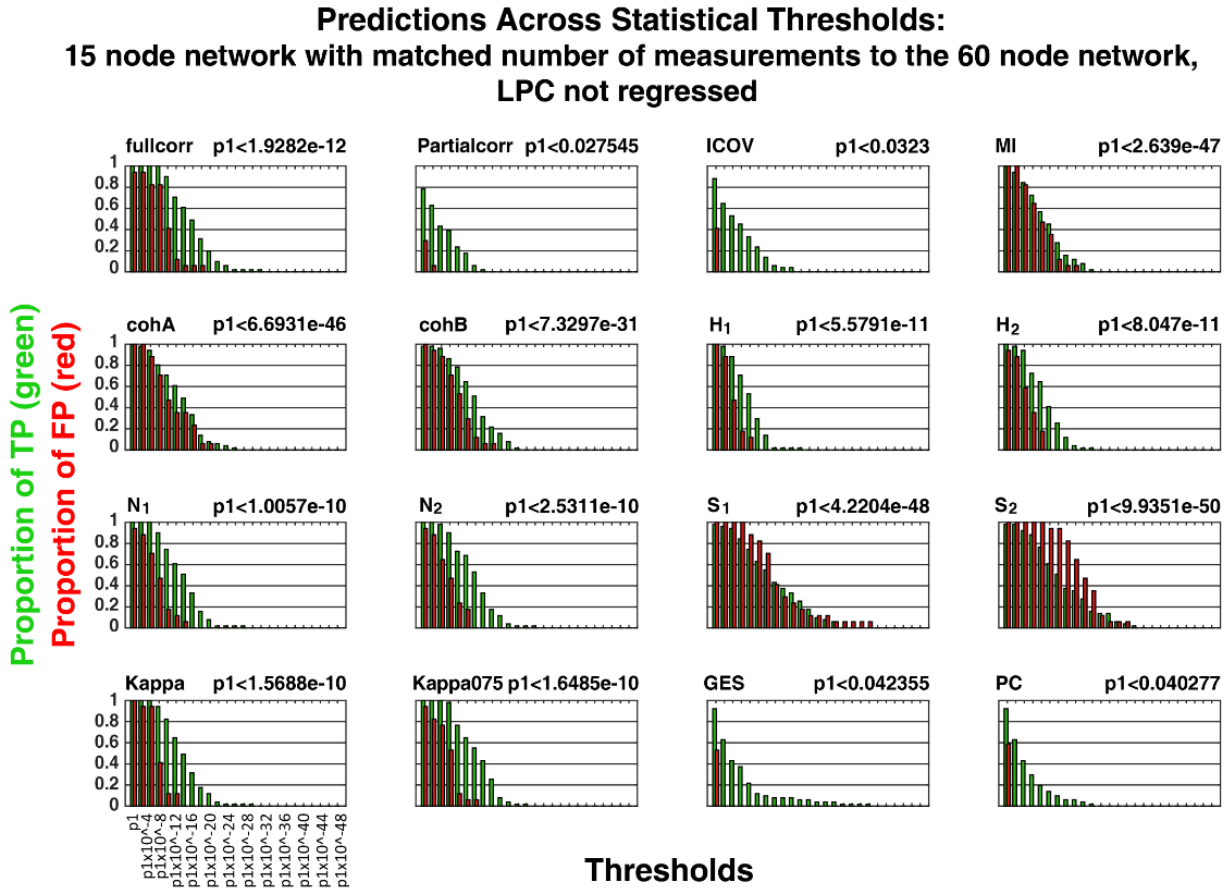


Fig.14. Proportion of true and false positive connections found for each of the models tested at various statistical significance thresholds. The results shown here were obtained for the 15 node network within the quadrant with no LPC regression, *this time using data obtained from only a single quadrant from each of the 8 runs per subject*. These results are presented for comparison to the 60 node network, which also has 8 runs of data per subject. The format of presentation is as in Figure 4.

In order to compare results from the 15 node and 60 node networks fairly, we have done a 15 node analysis with the number of data points matched to the number available in the 60 node network. We used only 8 data sets for each subject, when we actually had 8×4 data sets available (8 runs, 4 quadrants). The runs and quadrants we used were selected in the following way: each quadrant was randomly selected from two different runs, with each run getting equal representation. With this new analysis presented in figure 14, we can now make a more fair comparison between the 15 and 60 node networks.

The results obtained from the 60 node network (Figure 13) and the matched-sample-size 15 node network (Figure 14) are comparable. The main observed difference is a slight decrease in the number of significant connections for the multivariate models, especially GES and PC, in the 60 node network compared to the matched-sample-size 15 node network. The additional information from the many nodes in this larger network has likely contributed to removing more network effects from the measured multivariate connectivity thus resulting in fewer false positive results (i.e. Partialcorr, ICOV, GES and PC). Another difference between the tests with different numbers of nodes is the FDR p-value threshold. In the case of the 15 node network, the connections we evaluate are those below the diagonal in a 15×15 connectivity matrix (i.e. 105 connections) (see Figure 3). In the 60 node network, the comparisons computed are those below the diagonal in a 60×60 connectivity matrix (i.e. 1770 connections) (see Figure 12). This increase in number of statistical comparisons for the 60 node network would result in a lower FDR corrected p-value threshold, or p1. Multivariate models possibly see a drop in the proportion of significant connections due to this more stringent multiple comparisons threshold. While fullcorr, Kappa, Kappa075 and the other bivariate models are observed to be relatively stable to the change in network complexity tested here, multivariate models Partialcorr, ICOV, GES and PC are seemingly more sensitive. That being said, the multivariate models fare better with greater network complexity (and as a result a more strict statistical threshold).

Next, we take advantage of this 15 node study with reduced data to consider the impact of the abundance of data on the models, without any other variation to the study. For this we compare the results of the 15 node network analyses in the cases with 32 data sets (4 quadrants × 8 runs; Figure 4) and 8 data sets (as described above; Figure 14) per subject. As is typical in statistical testing, with more data, all models show a trend of more significant connections at all p-value thresholds. When the LPC is not regressed, the differences between the 4×8 datasets tests and the 8 datasets tests are minor (at the FDR p-value threshold associated with a maximum FPR of 0.05) for nearly all models. Amongst these minor differences, notably, for Partialcorr and ICOV, more data allows the connectivity estimates to converge closer to the ground truth, thus we see fewer false positive predictions.

60-node network: LPC regression

In the 60 node network analysis with or without LPC regression, the trends remain similar for Partialcorr, ICOV, GES and PC. Fullcorr and Kappa/Kappa075 show dramatic improvement with the LPC regressed (Figure 15), as they did in the 15 node network. These models go from all FPs to zero FPs, becoming the best performing methods. Overall there is a small drop in statistical significance across models when the LPC is regressed, reflecting the baseline shift of positive interactions measures closer to zero. The maximum mean TPR-FPR observed in the 60 node analyses is of 0.41 in the smaller range and 0.18 in the full range (Fullcorr achieves these, see Table 5, columns 5 and 6).

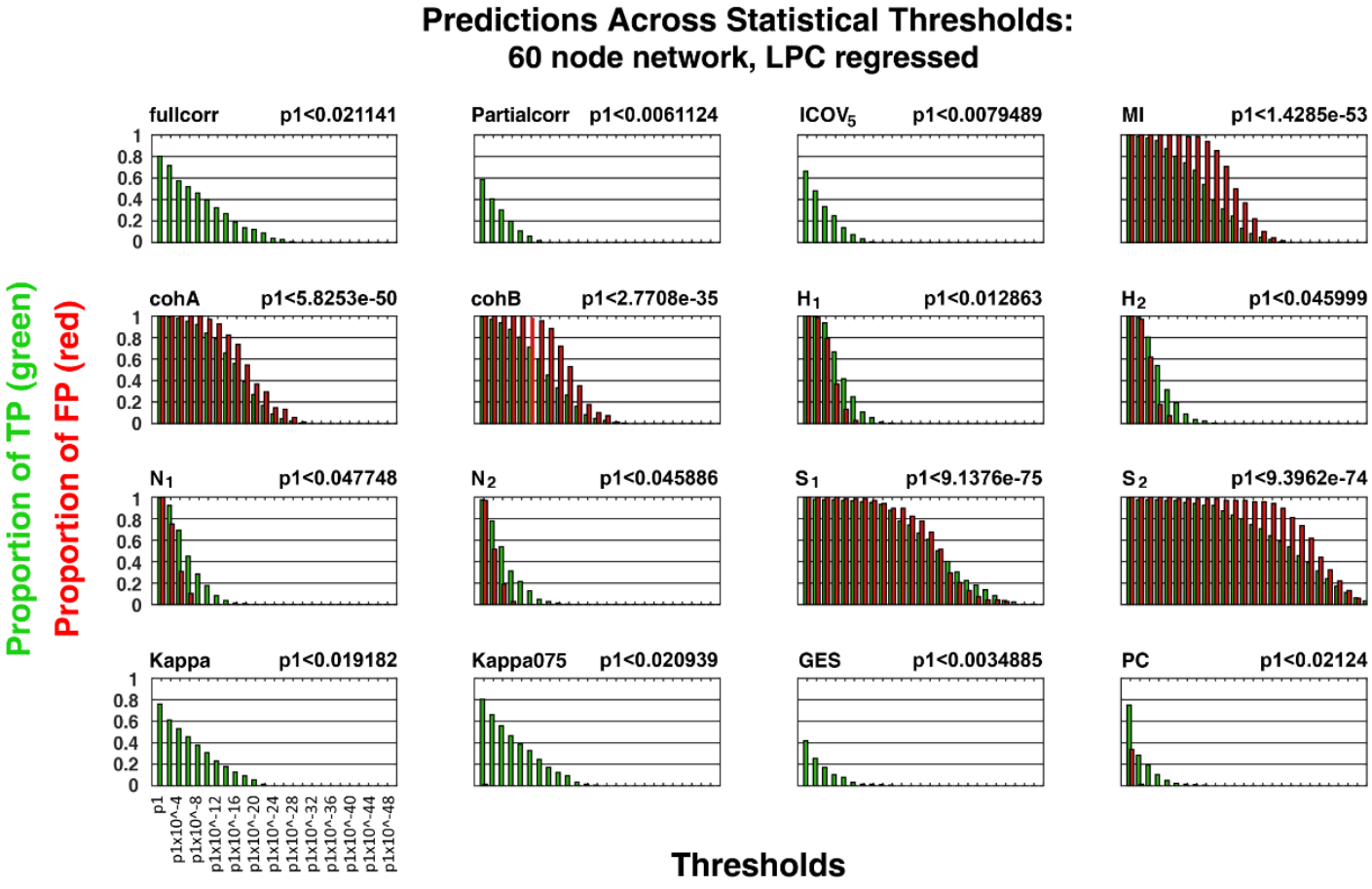


Fig. 15. Proportion of true and false positive connections found for each of the models tested at various statistical significance thresholds. The results shown here were obtained for the 60 nodes from all 4 quadrants of the complete visual cortex network. *The processing included regression of the LPC*. There are 51×4=204 expected connections tested for and 17×4=68 unexpected connections. Green and red bars represent the true positive (TP) and the false positive (FP)

predictions, respectively. The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The p-value thresholds become progressively smaller by a multiplication factor of 10^-2 at each step (up to a minimum of 10^-50). The format of presentation is as in Figure 4.

60-node network: Analysis with a large sample size – Human Connectome Project data

Finally, to examine whether the models can improve performance with abundant data as they did in the 15 node network, we present the results of a 60 node analysis with 220 HCP subjects. We remind the reader that we use 220 subjects from the Human Connectome Project with the goal of achieving 95% or more power in each analysis undertaken (power of the analysis for the 60 node network is presented in Supplemental Figures 31, and 35). With the LPC regressed (Figure 16), Fullcorr, Kappa, Kappa075, Partialcorr and ICOV show very substantially improved results with nearly perfect results at their FDR p-value threshold associated with a maximum FPR of 0.05. No FPs were found and the connectivity matrices resemble the inferred ground truth matrix remarkably well (see Supplemental Figure 47). The proportion of TPs found is also very high in all cases.

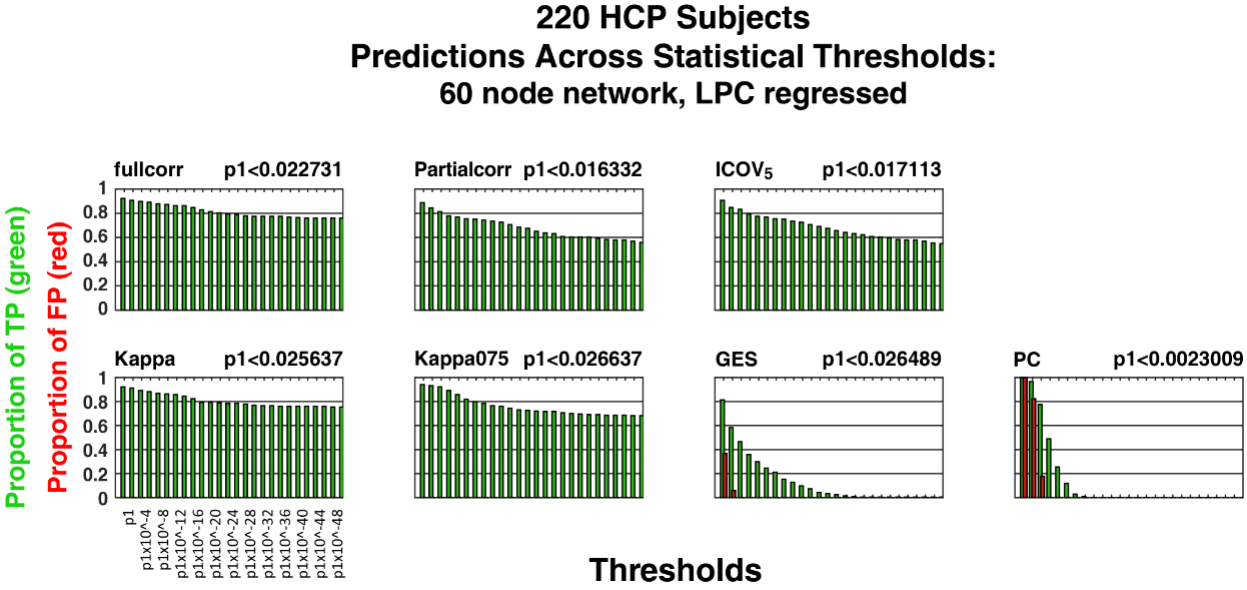


Fig. 16. Proportion of true and false positive connections found for each of the models tested at various statistical significance thresholds. The results shown here were obtained for the 60 nodes from all 4 quadrants of the complete visual cortex network *using data from 220 HCP subjects*. *The processing included regression of the LPC*. The format of presentation is as in Figure 4.

Without LPC regression, Partialcorr and ICOV, as well as PC and GES improved the limiting of the FPs in the network prediction in the 60 node network than in the 15 node network (compare Supplemental Figure 32 – 60 nodes to Figure 8/Supplemental Figure 24 – 15 nodes), benefitting from the additional regressions and likely also from the more stringent correction for multiple comparisons. They show excellent network predictions with p-value thresholds at p1x10^-6 or smaller (Supplemental Figure 32).

The stability of the network prediction across statistical thresholds is very good in this analysis with large sample size. That being said, the 60 node analysis is somewhat less stable than the 15 node analysis. Supplemental Table 9 shows the mean TPR-FPR across a threshold range of [p1 to p1×10-20] and of [p1 to p1×10-50] in the 60 node analysis. These values drop slightly from where they were in the 15 node network (Supplemental Table 7). For example, Fullcorr has mean TPR-FPR of 0.87 (short threshold range) and 0.81 (complete threshold range) in the 60 node analysis, whereas in the 15 node analysis it had means of 0.90 and 0.85, respectively. In spite of this drop, the models still show high level of stability and the network predictions are highly reliable across any statistical threshold for Fullcorr, Partialcorr, ICOV and Kappa. The same cannot be said for the density-based thresholds (Supplemental Figures 33 and 37), where the model performance is very threshold dependent.

Overall, the best performing models observed in the HCP analysis with 15 nodes remain the top performers in the 60 node analysis. With LPC regression, Fullcorr and Kappa075 are the top models, achieving 0.92 and 0.94 maximal TPR-FPR, respectively, at their FDR p-value threshold associated with a maximum FPR of 0.05. These results are highly stable and include no FPs. Therefore, we conclude that with large sample size, Fullcorr, Kappa/Kappa075, Partialcorr, and ICOV are excellent and highly reliable in predicting direct connections within a fine-scaled, densely-connected network at any statistical threshold, including an alpha of 0.05 where they perform optimally.

Discussion

What makes a good model? A good model should minimize false positive results, maximize true positive results, show invariance under variable parameters, be usable with reasonable amounts of data, and not require the user to provide abundant a priori details about the network of study. In the discussion sections below we address these considerations and

attempt to narrow in on successful models of direct/anatomical connectivity from functional MRI data.

Regions of interest and networks of the visual cortex

a. Small, densely connected networks

We chose a small scale, densely connected network for the following reasons. First, we have extensive knowledge on the homologous network in the monkey. Second, we believe that extending the methods of resting-state functional connectivity to fine-scale networks is the next challenge for utilization of this imaging modality. Such extension will substantially increase the information we can gain, both on functional/anatomical connectivity and on the pathophysiology of neurological diseases. Small scale networks can be analyzed in a framework of hierarchical, multi-scale analysis of functional connectivity (Smith et al., 2013; Shams et al., 2015). While many groups prefer to analyze FC in a whole brain analysis, either in a voxel-wise or ROI-based approach, in the interest of obtaining small scale functional connectivity that is meaningful on the local scale, rather than a global scale, we have chosen to look at small ROIs within only a small section of the cortex (early visual cortex). More, we pursue this without including extraneous information from outside of the local network which could overpower the fine effects we observe locally. To provide a sense of the scale of our ROIs: a whole brain parcellation with comparably sized ROIs would be composed of about 660 regions. This number is based on an estimate of approximately 100,000 surface voxels on cortex (given a voxel size of 2mm x 2mm x 2.2mm), and the approximate 9,000 voxels in our composite mask of V1, V2, and V3 (or 9,000/60=150 voxels per ROI). Lastly, we motivate our choice of network in the following way: If the models can differentiate between true and false positives in a challenging, small, densely connected network, they will likely also do so in less challenging, longer range connectivity tests.

A big assumption that we make in defining our inferred ground truth network is that the principles governing visuotopic anatomical connections within and between V1, V2, and V3 are similar in humans and macaque monkeys. There are a number of human fMRI studies (e.g. Sereno et al., 1995; Engel et al., 1997; Dumoulin and Wandell, 2008) that have shown a close resemblance between the functional retinotopy of these early visual areas in the human brain and the monkey brain. Topographical specificity has also been shown between these areas with resting-state functional connectivity (Heinzle et al., 2011). These findings strongly indicate that

retinotopically specific anatomical connections exist between V1, V2, and V3 in the human brain, as they do in the monkey brain.

As previously explained, the early visual cortex is densely connected, with forward projections from V1 to V2 and V3, and likewise from V2 to V3, but also feedback projections between each of these areas (Felleman and Van Essen, 1991; Salin and Bullier, 1995). Inter areal connections between lower visual areas are specific to similar eccentricities (Salin and Bullier, 1995; Shmuel et al., 2005; Heinzle et al., 2011). This finding is the basis for our choice of expected and unexpected connections. In addition we expect horizontal connections (i.e. within a visual area), on a scale of approximately 5-16mm, or between adjacent eccentricity regions in our network (see methods for explanation). Moreover, the set of connections we selected represent conservative choices. Since we preferred to be conservative, we did not include in our set of expected and unexpected connections any connections other than those outlined above within or between lower visual areas. These "other connections" include some inter-areal connections between regions with less similar eccentricities. While connections between certain ROIs from different quadrants or different hemispheres may exist, they are substantially weaker (Salin and Bullier, 1995; Saenz and Fine, 2010; Restani and Caleo, 2016), their spatial extent is variable and not fully clear, and they are not validated to the same level as within quadrant connections (Van Essen et al., 1982). Therefore, we excluded any between-quadrant or between-hemisphere connections from our set of expected and unexpected connections.

Given the existence of inter-areal connections, horizontal connections, and weak inter-hemispheric connections, one can easily see that the entirety of the early visual cortex would be in communication either through direct or indirect connections. This can certainly be said of connections between ROIs from the same quadrant. Therefore, this visual cortex-based network we used for our evaluations is a small, densely connected network. We chose this network in order to challenge the models to differentiate between direct and indirect connectivity, with the goal of identifying only direct connectivity from RS-FC, which may be a good approximation of the underlying anatomy (De Luca et al., 2006; Fox and Raichle, 2007; Koch et al., 2002; Quigley et al., 2003; van den Heuvel et al., 2009).

b. Network complexity

Network complexity is an important consideration because with more nodes, there is more information that can be accounted for, but also more noise and fewer degrees of freedom, and thus the quality of the network prediction may be altered. Additional information can be helpful in predicting a network with greater accuracy, but the additional noise and reduction in degrees of freedom can be quite detrimental. Our parallel analyses in networks with different numbers of nodes, each run under the assumption of no prior knowledge of the network of study, seek to illuminate the effect of network complexity on the models' performances at the task of approximating anatomical connectivity from RS-FC.

Comparing the 15 node network analysis using only 8 runs to the 60 node analysis (8 runs also), there are somewhat lower proportions of significant connections in the larger network for the multivariate models. The reduced proportion of significant connections in the larger network analyses (for example, Partialcorr, ICOV, GES, PC) is something we observed previously (Dawson et al., 2013), where we used prior knowledge of the network to set a false positive based threshold for significance. We note that the difference in network size is unsurprisingly most impactful for the models which are multivariate, since for those which are bivariate, only the two nodes being compared are considered, and no "network effects" are accounted for; the comparison between time courses remains identical. That being said, the bivariate network predictions do differ somewhat between the 15 node and 60 node analyses because not all quadrant data is used in the 15 node study with only 8 runs, and the correction for multiple comparisons is more stringent given more comparisons in the 60 node network.

In the case of multivariate models, there are several explanations for why we see a reduced proportion of significant connections predicted in the larger network. Regression of time courses is a noisy process. When the number of regressions (e.g., in Partialcorr and ICOV) increases, some of the fluctuations that contribute to measures of connectivity are potentially obscured or lost (Bright and Murphy, 2015). In addition, true for all multivariate models, with increasing number of nodes, the models need to lock on to an optimal solution among increased number of possible solutions, due to the increase in number of nodes and connections. In other words, the number of degrees of freedom of the problem decreases. Indeed, all multivariate models see a drop in proportion of connections with statistically significant connectivity when we compare the 60 node network to the 15 node network (Figures 13 and 4, respectively). The

drop in connectivity observed for the multivariate models from the 15 node analysis to the 60 node analysis ended up having a beneficial effect, eliminating most FP, but at the expense of a few TP as well. The network predictions in the larger network can be seen as more conservative than those in the smaller network.

Network modeling methods: multivariate and bivariate models

We hypothesized that multivariate models, which take into account the data from all regions within the network when assessing RS-FC, would be more successful at approximating direct, anatomical connections than bivariate models (Smith at al., 2011; Dawson et al., 2013). This hypothesis was especially convincing in a small scale, densely connected network such as the one comprising V1, V2 and V3 in a quadrant of the visual cortex, where all regions are expected to be indirectly connected through a relatively small number of synapses. Bivariate models that, by definition, do not account for signal contributions from other regions were expected to show non-zero interaction measures between all pairs of ROIs. That being said, as mentioned above, taking into consideration the data from other regions in the network increases the complexity of the computation (i.e. increased number of degrees of freedom) and the level of noise which enters into the computation of the functional connectivity measure. Accounting for a large number of regions present in the network can lead to the relationship between two nodes being obscured by blending with the noise contained within the signals (Bright and Murphy, 2015). In our study, we find that the inclusion of additional nodes in the multivariate models does, to a certain extent, obscure true connectivity (missed TPs), but still allows for proper classification of many direct and indirect connections (i.e. relatively high proportion of TP and very low FP). Given our results, it is clear that bivariate models are incapable of differentiating between direct and indirect connections in a small scale network such as within a single quadrant of the visual cortex unless the LPC is regressed out. On the other hand, some multivariate models are indeed capable.

The multivariate models are the better choices within the models we have tested for predicting the defined expected network with common statistical methods, if the LPC is not regressed. Partialcorr, ICOV, and the BayesNet methods are the only multivariate models among those considered. In the smaller network, Partialcorr is a simple and straightforward regression model which is successful and reliable (Marrelec et al., 2006). For the larger network,

Partialcorr is still a good option, though ICOV can offer some improvement, this being a regularized version of Partialcorr. The regularization reduces small interaction measures between minimally correlated regions, bringing the measure closer to zero, with the goal of increasing the sparsity of the estimated network (Friedman et al., 2008). This is achieved by way of an L1 graphical lasso regression (Tibshirani, 1996; Schmidt, 2006). The increased sparsity (or increased number of regions having connection strength ~0) reflects greater conditional independence between the nodes in the network. Regularization should improve upon methods which predict a network with an inappropriately high density of connections, such as correlation. This occurrence of densely connected predicted networks is unsurprising in small scaled networks where the regions within are expected to be indirectly connected. Thus, the effect of ICOV to decrease this density of connections is very well suited to the scale of network worked with here and given that we wish to differentiate between direct and indirect connectivity.

Both ICOV and Partialcorr can only work under the condition that the number of nodes in the network is smaller than the number of time points available in a dataset. However, even if this condition is met, abundant time points allows for easier convergence of the data to the true result. Part of the problem with shorter time courses is the introduction of relatively large amounts of noise into the FC estimation through multiple regressions (more nodes in the network means more regressions). With longer time courses, the common trends across time between nodes are more robustly observable and thus can withstand more noise before they are obscured. ICOV was introduced in part as an improvement to Partialcorr, the sparsity condition making it such that the noise introduced through regression is less impactful. This means that the need for long time courses is less pronounced for ICOV than it is for Partialcorr. Another strategy to overcome the need for long time courses in network analyses with many nodes is to concatenate time courses obtained in multiple sessions from the same subject (for example as done in Dawson et al., 2013), or even time courses from different subjects (as done in Tian et al., 2013).

Choosing a method which favors sparsity is one option for limiting the number of connections in densely connected networks. Another option for modifying the density of the network prediction would be to do so via density-based thresholding. Model choice, as well as number of nodes in the network or the number of data points, can indeed impact the density of the network. We would like to emphasize, however, that in this case we are not looking for a particular density of network, since we assume that we have no knowledge of the network in

general and, more specifically, on the true density of anatomical connections in the network. Instead, we are interested in finding accurate predictions of direct anatomical connections, and thus we prefer to limit density via appropriate modeling and not by arbitrary selection.

The BayesNet models are quite successful in eliminating false positives in the larger network (LPC not regressed), though this is at the expense of identifying the full set of true positives. They suffer when the LPC is regressed with regards to the presence of FPs, likely due to the appearance of spurious anti-correlations not accounted for by sign, and thus not readily distinguishable from positive connectivity. They also show optimal results with slightly more stringent thresholding than the FDR p-value threshold associated with a maximum FPR of 0.05 (i.e. p1) (e.g. $p1x10^{-6}$ or less, see Figure 4 and Table 2, column 3) when the network is small and less information is taken into account (particularly when the LPC is not regressed). These use a Bayesian approach which seeks to identify a parent subset of nodes within the system which are activated before the nodes which are being assessed for connectivity. More specifically, PC ("Peter and Clark"; Meek, 1995) is a pattern search which assumes that the network of study is acyclic (though not strictly), and does not consider the impact of input by nodes outside of the network (Scheines et al., 1996). Beginning with a complete undirected graph, the algorithm tests for adjacencies (ie. "parent" or "child" relationships, or connectivity) between nodes which may be through another node, and then removes any such adjacencies (Spirtes et al., 1993). GES (Chickering, 2002) is a score-based search algorithm which makes the same assumptions as PC. The network connections are added or removed based on the existence of parent or descendant nodes in the undirected graph, and then the new graph is scored using the Bayesian Information Criterion (BIC), and the configuration which optimizes this score is the network predicted by the model. The BIC score is such that it increases when an edge is added to the graph if this edge eliminates an independence constraint between nodes in the network (Chickering, 2002). Alternatively, the BIC will decrease if an edge is added and it does not eliminate such a constraint. Thus, PC and GES are well equipped, by design, to seek out only direct connections, independent of influence from any other network nodes, though they seem to require as much information from the network of interest as possible in order to minimize FP results.

Fullcorr and Kappa/Kappa075 are the bivariate models which are successful in our analyses when the LPC is regressed. Kappa aims to identify the joint activation of a pair of brain voxels using a multinomial likelihood with a Dirichlet prior distribution, within the context of a

Bernouilli Bayesian model. It is calculated using the probabilities of activation of two nodes as well as the probabilities of each node given the other (Patel et al., 2006). These probabilities are calculated based on binarized data. The data is binarized such that any value in the normalized time-course which is less than 0.1 is set to 0, and anything at or above is set to 1 (Kappa). We also binarize the data at a 0.75 threshold (Kappa075). Fullcorr is simply Pearson's linear correlation, which utilizes the covariance between two time-courses to assess a relationship. Given that these models are bivariate, only considering two ROI time-courses at a time, the values tend to be positive, reflecting relationships that are both direct and indirect. With the regression of the LPC, we are able to account for much of the covariance contained with the time-courses that are due to indirect connectivity, and we then end up with a distribution of scaled down "direct" connectivity measures which are centered around zero. Fullcorr and Kappa manage to perform better than Partialcorr and ICOV under the right conditions (i.e. with LPC regression). This may be due to the fact that a single regression of the LPC introduces less noise into the model than multiple regressions from all other network nodes, as done in Partialcorr and ICOV.

Statistical approach

Some of the models tested are better able to avoid unexpected connections with additional thresholding (beyond the FDR p-value threshold associated with a maximum FPR of 0.05). The difficulty with such an approach is defining what the threshold should be. The thresholds that we can identify as optimal in the networks tested are identifiable only because we know which connections are expected and which are not. In order to avoid arbitrarily setting the threshold, we have previously proposed using data from a known network to establish a threshold based on the false positive distribution in that known network, and to use this threshold in order to make inferences about related unknown networks (Dawson et al., 2013). This type of approach would be useful only in analyses where it is possible to establish a related "ground-truth" network on a similar scale, which is not always feasible. Because of this limitation, it is important to have reliable models which can be trusted to accurately predict the network without prior information or assumptions. These models are what we have sought to identify in this study. In the 15 node network, PC and GES benefit from additional thresholding (~p1×10^-6,

LPC not regressed) in the task of eliminating false positive results. ICOV_5 and Partialcorr have optimal network predictions at or near (~p1x10^-2) the FDR p-value threshold in all analyses, and the same goes for GES and PC in the larger network. Fullcorr and Kappa are always optimal at their FDR p-value threshold if the LPC is regressed. If the LPC is not regressed, these methods fail fully.

A measurement we would like to draw attention to is the actual value of p1, the FDR p-value threshold associated with a maximum FPR of 0.05. The FDR threshold is determined according to the distribution of the p-values from all comparisons made. Some of the adjusted thresholds are very small, for example p1 for cohA in Figure 4 is 1.6x10^-183. This scale of p-value threshold is not especially useful. The t-statistic corresponding to these very small p-values will change dramatically even for small changes in the p-value. This size of threshold reflects the fact that the associated models are not useful in the task we have presented them with (approximating anatomical connectivity with functional data). Models with such low FDR thresholds are always among the worst models with essentially unusable network predictions at any threshold (see Figures 4, 5, and 14). Thus, one can use this as a way to gauge the likelihood of a successful network prediction: if the FDR threshold is very small, it is likely that the choice of model is poor for the analysis at hand.

An important decision in our statistical interpretations was to only consider positive connections in our network and omit any significant connections with a negative mean interaction measure. The connections that we observe as significant with a negative mean, most often in the case of LPC regression, but also scantly for Partialcorr and ICOV without LPC regression, are almost always among the connections we have included in our set of unexpected connections. This finding leads us to believe that our decision to only consider positive connections is justified and important. At least, it is important to differentiate between positive and negative mean connectivity measures as these likely compose different groups of connections, representing different effects.

As mentioned, the negative interaction measure we see, occur most often when the LPC is regressed, and in Partialcorr and ICOV. The methodology of regression models like partial correlation and ICOV bears some similarities to global signal regression (here accomplished via LPC regression), in that the multiple regressions that occur should account for the common fluctuations to all nodes considered. These methods are in fact alternatives to global signal

regression, but importantly they do not produce an identical effect. This is especially true if there is additional information contained within the global signal that will not be part of the regressions in Partialcorr, as is the case in our study. Indeed, Partialcorr and ICOV without LPC regression achieve results which are by far the most comparable to those with the LPC regressed, demonstrating the similar effect that can be achieved using these models. But importantly, regressing the LPC improves results further by regressing out additional information that was not accounted for in the network's nodes. This is because the LPC is based on all V1, V2, and V3 voxels, whereas the regressions done by Partialcorr and ICOV only contain information from the refined selection of voxels which are contained in the defined ROIs for that network. As for why these regressions lead to negative values, by regressing the LPC we are removing some level of common signal from the time courses. This should reduce the correlation (and partial correlation if the information regressed in the global signal is not accounted for in its regressions). Thus lower correlation, Partialcorr, and ICOV to some degree is expected. This is what we observe. For Partialcorr and ICOV who have many values already nearly of zero magnitude, this results in negative interaction measures.

Looking deeper at what these negative partial correlation/ICOV values might represent, is not straightforward to interpret. If we take two nodes, A and B, which have low or zero correlation for which we are computing the partial correlation, when we regress out information from another node, C, for which the relationships with A and B are not zero and share a common effect structure, we are removing the variance A and B share through the common effect they generate within C. This leads to a decrease in the correlation between A and B, and for sufficiently small correlations to begin with, this results in negative relationships (Zhang et al., 2008; Epskamp and Fried, 2017). This effect will be stronger with more regressions, which explains the presence of negative partial correlations in our network, especially at the unexpected connections, which we a priori expect to have relatively low correlation. When we regress out the LPC, we eliminate even more shared variance, resulting in more negative partial correlations/ICOVs. Thus, we hypothesize that the negative interaction measures we find in our analyses represents largely shared variance through common effects with other network nodes between two nodes which are not strongly connected; certainly this includes our unexpected connections.

A final item to note with regards to negative connections: the models which do not have the distinguishing feature of sign within the set of connectivity measures do not perform as well as those which do distinguish between positive and negative connectivity. A prime example of this are the BayesNet models, GES and PC, which show poorer results with LPC regression, a process which tends to capture more of the indirect connections for these models. These multivariate models do however associate directionality to each connection and perhaps in evaluating the direction of the connections within the network, we might observe a similar grouping.

Another factor we considered with regards to statistical approach is the degree and type of thresholding applied to determine the presence of a connection in the network. The reason our initial p-value threshold is based on a FPR of 0.05 is because there is no clear way to determine an alternate threshold without prior knowledge of the system. Despite this, we were curious to see if network predictions might improve with stricter thresholding and also to observe the stability of the network prediction across thresholds. Garrison et al. (2015) also found this to be an important question. They had evaluated the stability of a range of absolute and proportional thresholds applied to various network measures computed over atlas-based ROIs from the whole brain. The network measures they considered included characteristic path length, global efficiency, transitivity, and modularity. This was along with some node-level measures including clustering coefficient, local efficiency, degree, betweenness centrality, and participation coefficient. They found that proportional thresholds were in general more stable than absolute thresholds on these measures.

In our study, we look at a progression through the full range of network connection densities (100-0%, step size=4%). Upon visual inspection, the TP bars decrease with a concave down trend while the FP bars decrease with a concave up trend as the density is decreased toward 0% (Figures 6, 7, 10, 11, and 16; plus others included in supplementary material). The FPs decrease more rapidly than the TPs. This effect is strongest in our successful models when the LPC is regressed. When the LPC is not regressed, MI, cohA, cohB, the GenSynch models, GES, and PC tend to show more of a linear decrease in FPs, the rate of decrease slowing compared to when the LCP is regressed. On inspection of the barplots, the stability of the density-based network predictions appears greater than the absolute threshold plots (i.e. less steep decline of TPs as thresholds become increasingly strict or density decreases). This,

however, is largely a reflection of the critical range of the p-values for that network; in other words, the range of p-value thresholds selected would have needed to be more narrow to observe similar stability. Thus we conclude that in analyses evaluating FC measures, it is more important to find the right scale of threshold than to use a density-based vs. absolute threshold, since in the correct range, these can show equivalent stability.

An issue with the density-based thresholding in our study is that there is no consistent density range for optimal network prediction and network stability for any of the models under the various test such that we could be confident in a different study that the threshold we use is appropriate. For the absolute thresholds, at or near the FDR threshold with a maximum FPR of 0.05 was reliably best for models like fullcorr, Kappa, ICOV, and Partialcorr under relatively broad conditions (i.e. if the LPC is regressed, if the network size is large etc.). We conclude that the stability of the network predictions across density-based thresholds is good given the appropriate density range, and that a significance level of 0.05 is a good standard to use for absolute thresholding of the p-values in the absence of prior knowledge of the system.

Evaluating the performance of network modeling methods - expected and unexpected connections

In this section we will discuss how well the predictions by the network models lined up with our set expected and unexpected connections. Of note in the observed FPs, i.e. unexpected connections that were found to be statistically significant, are the connections between e1 and e5, mostly within a visual area, but occasionally between visual areas. The span from e1-e5 in humans is expected to be somewhere on the scale of around 45-60mm on the cortex (Tootell et al., 1998), as mentioned while motivating our expectations. Horizontal connections, however, are not expected to be more than 16 mm at their longest, and likely much shorter in V1 and V2 (Salin and Bullier, 1995). Between visual areas, the divergence of information is greatest in the feedback direction (Salin and Bullier, 1995). These connections are expected to extend their broadest reach of up to around 14mm between V3 and V1. If we use this scale of ~14-16mm as the maximum horizontal reach either within or between visual areas for any direct connection, we should not see connectivity between 2 regions separated by more than one eccentricity region (i.e. not beyond e1-e3, e2-e4, e3-e5). Because of this, the e1-e5 connections we are observing

we suspect are artifactual. We note that these connections are largely eliminated when the LPC is regressed, a sign that this regression is leading the models closer to the inferred ground truth in direct connectivity. One possible explanation for this observed connectivity when the LPC is not regressed is that e5 is located relatively near to the lateral occipital complex (LOC) which processes shape and receives abundant input from the central visual field (Wang et al., 2016). It may be that feedback from LOC to nearby (i.e. e5) visual areas is strong and thus the influence of V1 from LOC is resulting in this heightened measured connectivity between e1 and e5.

The next inconsistencies we observed between our expectations and the network predictions are the false negative results for expected connections between V1 and V3. One possible explanation for the reduced number of V1-V3 connections found to be significant is that given the strong influence of V1 on both V2 and V3 (Markov et al., 2014), perhaps in accounting for the influence of V2 by the multivariate models, some of the similarity between V1 and V3 might also be removed. Since V1 and V3 are less in communication compared to V1 and V2 (for example as seen in Markov et al., 2014, V1-V3 adjacent eccentricity connections should be weaker than V1-V2 connections), these connections would be more likely to drop off in significance than the connections between V1 and V2 (due to considerations of influence from V1 to V3). That being said, it may in fact be due to similarities in connectivity between V1-V3 and V1-V2 that some V1-V2 adjacent eccentricity connections are not found to be significant. These false negative connections are fewer than between V1 and V3.

These observations of the potential shortcomings in the network predictions show us how important accurate ROI definition is as well as removal of the influence of potential confounds. Perhaps information from outside of V1, V2, and V3, from the rest of the visual cortex, would make a good contribution to our LPC estimation, and would help minimize the e1-e5 FP presence. These issues also remind us of how connected the early visual cortex is, and just how sensitive a process it is for a model to diligently account for information, without accounting for too much such that true relationships are obscured. And finally, while we are utilizing the RS-FC to approximate anatomical connectivity here, there is additional information contained within RS-FC that we may need to account for in order to better make our approximation.

False negatives are also a concern with regression methods such as Partialcorr and ICOV. In particular, these models may struggle to retain triangle connectivity when it truly exists, as the regression procedure may remove some of a triangle's edges, interpreting the connections as

indirect. We test for many triangles in our expected connections, for example a horizontal connection between adjacent eccentricity regions from the same visual area, a vertical connection between same eccentricity regions in different visual areas, and a diagonal connection between adjacent eccentricity regions in different visual areas. Let us consider specifically such triangles between V1-V2, or V2-V3 since we just outlined above that there are missing connections between V1-V3. If we examine the connectivity matrices in Figures 3 and 12 (plus others included in supplementary material), we find that most of these such triangles are preserved in the 15 node network analyses, and they are not incomplete in alarming number in the 60 node networks (where there are more missing connections in general; see Tables 2-5, column 10). There are also likely triangles contained within each between-area, same-eccentricity connection. These connections could be viewed as capturing the entire triangular connectivity that exists on a smaller scale than our ROIs. Thus, in local networks with strong connections between areas, triangles do not seem to be a major issue for regression models.

Local Principal Component regression

While the major concern with the removal of the global signal is the introduction of spurious anti-correlations in the computed distribution (Murphy et al., 2009), our focus here is not the sign or scale of the individual correlations, but the identification of the presence of direct connections. Removal of a global signal results in decreased functional connectivity between regions, which could in fact be helpful in our study, potentially limiting artificially high connectivity to below significance thresholding. This is especially true for bivariate models like correlation and Kappa which do not have the benefit of information from other nodes in orders to differentiate between direct and indirect connectivity. Because of this, we chose to pursue our analyses without the removal of the LPC in order to present a more conservative set of results, as well as to present the results with LPC regression in order to give bivariate models a chance. Although a carefully chosen global signal model which should not introduce spurious anti correlations is possible (e.g., Carbonell et al., 2011, 2014; He et al., 2012), we have decided to eliminate the possibility of spurious anticorrelations altogether by examining only positive connections in our analyses, with and without LPC regression. We support this decision by emphasizing the strength of the feedforward, excitatory connections between these early visual areas, as they compare to the inhibitory feedback connectivity that may be present (Power et al.,

2013). Examining the results of our analyses with LPC regression, our approach provides very good network predictions compared to expectations (Figures 5, 9, 14, and 15), further supporting our choice.

Another approach we could have taken with respect to regression of a GS, would be to compute the GS over the whole brain: in removing a truly "global" signal from the data we would have had the opportunity to take into account input from regions outside of V1, V2 and V3. This choice of GS could potentially contribute to eliminating connections being predicted due to common inputs specifically to two ROIs, from outside of the considered network, leading to fewer FP. We decided against this approach, however, because we feel that if any such effects occur, they would be very minor. Input from higher visual areas or regions outside of the visual cortex are far less retinotopic than input from within the early visual cortex and are likely to have a similar effect on all regions in our network (Salin and Bullier, 1995). Any information to be gathered from outside of the network should already be captured in the LPC from within.

Another reason we choose to approximate the GS with the voxels contained only within the early visual cortex (i.e. the LPC) is that we are striving to approximate anatomical connections, not capture the full extent and content of the functional relationships of the regions in our networks. Thus, if the model is able to successfully approximate anatomical connectivity without the use of time-courses from outside of the defined network, then we do not need to consider a larger network. In the end, we chose to limit our "global signal" analysis to a local network in order to evaluate whether the models can see success in spite of potential, though not expected, limitations in extent of information contributing to the model.

An interesting future direction of study would be to directly compare the performances of the LPC and a whole-brain GS in the context of estimating the fine-scaled direct FC of local networks, however, this was beyond the scope of this study.

Testing with a large sample size: Human Connectome Project data

The results of our analyses with data from 220 HCP subjects are very impressive. When the LPC is regressed, the results from both the 15 and 60 node network analyses are nearly perfect and show zero FPs for fullcorr, Kappa, Kappa075, Partialcorr, and ICOV. It is important for us to have pursued these tests with a greater sample size and power considerations in order to validate our current study and to observe any shortcoming that we may have suffered given the

smaller data set. The results with our data compare quite well with the results with HCP data, with the strongest models in the tests with our data remaining strong in the HCP analyses. The trends observed remain across performed tests. With the larger sample size, however, we see the successful models converging to the inferred ground truth network with the LPC regressed and when statistical thresholding is used. Importantly, we do not see this convergence happening with density-based thresholding. This is a somewhat trivial finding given the design of a density-based analysis looking through the full range of network connection densities. At 100% connection density, all TPs and FPs will be found, and at 0% no connectivity will be present in the network, and thus the optimal network prediction will always lie somewhere in the middle ranges of network density. The improvement possible for density-based thresholding with greater sample size is minor (compare tables 2 and 6, 3 and 7, 4 and 8, 5 and 9, and density-based figures found in the supplementary material), and presents as a greater stability of TP connections (more gradual decline of green bars) and decreased stability of FP connections (steeper decline of red bars). Nonetheless, with large sample size the problem remains of determining the most appropriate connection density for the network.

GES and PC are found to be not as stable in their network predictions in spite of having a large number of subjects, and they are also found to have many FPs even when the LPC is regressed, both of these observations being counter trend compared to the other tested models. The preservation of FPs when the LPC is regressed is, as previously mentioned, largely due to the fact that these measures are not signed and thus "negative" connections remain in the network prediction, whereas for the other models we are able to remove these from the network. The relative instability could likely be overcome by testing a finer range of sensitivity parameters, which are used in order to associate a connection strength with a given connection, since these models output a binary indicating presence or non-presence of connection. Through varying the sensitivity parameter, more or less connections are found to be present in the network. A finer range of sensitivity parameters was not tested here due to the computational demands that these models require. The memory and time requirements are much higher given the volume of data being tested in these "large-sample size" analyses.

In order to be sure that the improvement we saw in fullcorr, Partialcorr, ICOV, Kappa and Kappa075 was due to the greater sample size, and not due to differences in the data acquisition/preprocessing/quality, we ran the HCP analyses with only 6 randomly selected

subjects (Supplementary Figures 38-53). These analyses made it clear that the improvements observed did indeed come from greater sample size.

It is particularly encouraging to observe how closely the results with abundant, high quality data match our set inferred ground truth (based on anatomical connectivity identified in tracer injection studies in the monkey - see Supplementary Figures 27 and 35). This is strong support for the methods we have used here as a possible means of approximating anatomical connectivity from RS-FC on a fine scale, something not yet achievable with other methods such as diffusion MRI tractography. Diffusion studies are still quite limited in terms of the spatial scale they can achieve (van den Heuvel et al., 2016). Based on the results of our study with 6 subjects and the results using the HCP data, our recommended choices for approximating anatomical connectivity in a network study without a priori knowledge of the connectivity contained within from RS fMRI data are the following: 1) LPC regression is beneficial, and care to avoid spurious anti-correlations by working with only the positive connectivity is a step worth considering (if it is appropriate in your network of study); 2) fullcorr, Kappa, ICOV, and Partialcorr are all capable of producing meaningful results in this context, though if LPC regression is not undertaken, Partialcorr or ICOV are better choices; 3) statistical significance can be evaluated using p-value thresholding at an significance level of 0.05, with correction for multiple comparisons.

Conclusion

A variety of network models have been implemented with the goal of identifying direct connections based on resting state fMRI data, with no prior information of the networks. This is achieved by using common statistical significance thresholding at an FDR p-value threshold associated with a maximum FPR of 0.05. The network we chose for this study within the early visual cortex of humans is expected to be densely connected. One major issue with many models is the presence of false positives in the predicted networks when one is seeking only direct connections. False positives are especially problematic in fine-scale, densely connected networks. We show that when using common statistical practices alone, in the absence of a "global signal" regression (in our case the Local Principal Component, LPC), multivariate models are the only option for differentiating between direct and indirect connectivity in such a network. Partial correlation and Regularized Inverse Covariance (ICOV) are most reliable at

predicting expected connections while not predicting unexpected connections. Bivariate models such as correlation are not capable of producing useable network predictions even when high threshold levels are applied when the LPC is not regressed. That being said, correlation and Kappa, both bivariate models, are top performers in tests when the LPC is regressed (top performances overall are in these tests). Even for the multivariate models, LPC regression results in network predictions closer to our expectations. If density-based thresholding is used instead of absolute statistical thresholding, network predictions can closely resemble expectations for some models. The trouble with using density-based thresholding in a system where we do not have any prior knowledge, is that there is no way to determine which network density is appropriate. In an extension of our study using a large sample size from Human Connectome Project data, we observe similar trends and are able to achieve excellent results with any of fullcorr, Partialcorr, Kappa, or ICOV in both the smaller and larger networks when the LPC is regressed. Importantly, with a large sample size, network predictions converge toward the ground truth if statistical thresholding is used, but they do not if density-based thresholding is used. For future study of small scale, densely connected, unknown brain networks, we recommend regression of a Local Principal Component signal, and to consider the effects of such a regression on the connection strengths of your model (for example, here it was important to recognize the shift toward smaller interaction values given the LPC regression, and only consider positive connections to avoid spurious anticorrelations). If this step is not done, it is advisable to use a multivariate model such as Partialcorr or ICOV. These recommendations will be helpful in predicting only direct connections between network nodes.

Note: There are three appendices with content relating to Manuscript 2. The reader can refer to these appendices as needed.

Appendix D presents a data article (submitted to Data in Brief; see BioRxiv doi: https://doi.org/10.1101/340174) elaborating the methodology we used in terms of preparing the Human Connectome Project (HCP) resting-state fMRI data for use in our study. The data that we have prepared and offer for download and study by the larger neuroscience community consists of 60 fine-scaled retinotopically organized regions in cortical areas V1, V2, and V3 (as used in Manuscript 2).

Appendix E contains a comment on computer precision, relevant given that we quote some computed values at very low powers (>e-100) in Manuscript 2. Appendix F presents results of analyses as conducted in Manuscript 2 but only for ICOV pursued at different regularization parameters.

In Appendix F we motivate our choice of regularization parameter further and give the reader a sense of potential variation in the performance of this model should a different regularization parameter be used.

References

Abou Elseoud, A., Littow H, Remes J, Starck T, Nikkinen J, Nissilä J, Timonen M, Tervonen O, Kiviniemi V (2011). "Group-ICA Model Order Highlights Patterns of Functional Brain Connectivity." Front Syst Neurosci **eCollection**: 5-37.

Adams DL, Sincich LC, Horton JC (2007). "Complete Pattern of Ocular Dominance Columns in Human Primary Visual Cortex". J Neurosci **27:** 10391–10403.

Amano, K., B. A. Wandell, et al. (2009). "Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex." J Neurophysiol **102**(5): 2704-2718.

Amir, Y., Harel M, Malach R (1993). "Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex." J Comp Neurol. **334**(1): 19-49.

Baria, A., Mansour A, Huang L, Baliki MN, Cecchi GA, Mesulam MM, Apkarian AV (2013). "Linking human brain local activity fluctuations to structural and functional network architectures." Neuroimage **73**: 144-155.

Benjamini, Y., & Hochberg, Y. (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing." Journal of the Royal Statistical Society, Series B ***57****(1): 289–300.*

Benson, NC, OH Butt, R Datta, PD Radoeva, DH Brainard, GK Aguirre (2012). "The retinotopic organization of striate cortex is well predicted by surface topology." Curr. Biol. **22**(21):2081-2085.

Benson, NC, OH Butt, DH Brainard, GK Aguirre (2014). "Correction of distortion in flattened representations of the cortical surface allows prediction of V1-V3 functional organization from anatomy." PLoS Comput. Biol. **10**(3)

Biswal, B., F. Zerrin Yetkin, Victor M. Haughton, James S. Hyde (1995). "Functional connectivity in the motor cortex of resting human brain using echo-planar mri." Magnetic Resonance in Medicine **34**(4): 537-541.

Bright, MG, Murphy K (2015). "Is fMRI "noise" really noise? Resting state nuisance regressors remove variance with network structure." Neuroimage **114**: 158-69.

Carbonell, F., Bellec P, Shmuel A (2011). "Global and System-Specific Resting-State fMRI Fluctuations Are Uncorrelated: Principal Component Analysis Reveals Anti-Correlated Networks." Brain Connectivity **1**(6): 496-510.

Carbonell, F., Bellec, P., Shmuel, A., 2014. Quantification of the impact of a confounding variable on functional connectivity confirms anti-correlated networks in the resting-state. NeuroImage 86, 343-353.

Chickering, DM (2002). "Optimal Structure Identification With Greedy Search." Journal of Machine Learning Research **3**; 507-554.

Dawson, DA, Cha K, Lewis LB, Mendola JD, Shmuel A (2013). "Evaluation and calibration of functional network modeling methods based on known anatomical connections." Neuroimage **67**: 331-343.

Dawson, DA, Lam J, Lewis LB, Carbonell F, Mendola JD, Shmuel A (2016). "Partial Correlation-Based Retinotopically Organized Resting-State Functional Connectivity Within and Between Areas of the Visual Cortex Reflects More Than Cortical Distance." Brain Connectivity **6**(1) :57-75.

De Luca, M., Beckmann, C. F., De Stefano, N., Matthews, P. M., Smith, S. M (2006). "fMRI resting state networks define distinct modes of long-distance interactions in the human brain." Neuroimage **29**: 1359-1367.

Drakesmith, M, Caeyenberghs K, Dutt A, Lewis G, David AS, Jones DK (2015). "Overcoming the effects of false positives and threshold bias in graph theoretical analyses of neuroimaging data." Neuroimage **118**: 313-333.

Dumoulin, S. O., Brian A. Wandell (2008). "Population receptive field estimates in human visual cortex." NeuroImage **39**: 647-660.

Duncan, R., Boynton GM. (2003). "Cortical magnification within human primary visual cortex correlates with acuity thresholds." Neuron 38(4): 659-671.

Engel, S., Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN. (1994). "fMRI of human visual cortex." Nature 369(6481).

Engel, S., Glover GH, Wandell BA (1997). "Retinotopic organization in human visual cortex and the spatial precision of functional MRI." Cereb Cortex **7**(2): 181-192.

Felleman**,** DJ, Van Essen DC (1991). **"**Distributed hierarchical processing in the primate cerebral cortex.**"** Cereb Cortex 1(1):1-47.

Fox, M. D., Marcus E. Raichle (2007). "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging." Nature **8**: 700-711.

Fox, M.D., D. Zhang, A.Z. Snyder, M.E. Raichle (2009). "The global signal and observed anticorrelated resting state brain networks." J. Neurophysiol. **101**: 3270–3283

Friedman, J, Hastie T, Tibshirani R (2008). "Sparse inverse covariance estimation with the graphical lasso." Biostatistics **9**(3): 432–441.

Friston, K., Harrison L, Penny W (2003). Dynamic causal modelling. NeuroImage **19**:1273‐302.

Garrison, KA, Scheinost D, Finn ES, Shen X, Constable RT (2015). "The (in)stability of functional brain network measures across thresholds." Neuroimage **118**: 651-661.

Gerhard, F., Kispersky T, Gutierrez GJ, Marder E, Kramer M, Eden U (2013). "Successful reconstruction of a physiological circuit with known connectivity from spiking activity alone." PLoS Comput Biol **9**(7).

Gotts, S., Saad ZS, Jo HJ, Wallace GL, Cox RW, Martin A (2013). "The perils of global signal regression for group comparisons: a case study of Autism Spectrum Disorders." Front Hum Neurosci **eCollection**: 7:356.

Haak, K. V., J. Winawer, et al. (2013). "Connective field modeling." Neuroimage **66**: 376-384.

Harvey, B. M. and S. O. Dumoulin (2011). "The relationship between cortical magnification factor and population receptive field size in human visual cortex: constancies in cortical architecture." J Neurosci **31**(38): 13604-13612.

He, H., Liu TT (2012). "A geometric view of global signal confounds in resting-state functional MRI." Neuroimage **59**(3): 2339-2348.

Heinzle, J., Kahnt T, Haynes J-D (2011). "Topographically specific functional connectivity between visual field maps in the human brain." NeuroImage **56** (3), 1426–1436

Honey, C., Thivierge JP, Sporns O (2010). "Can structure predict function in the human brain?" Neuroimage **52**(3): 766-776.

Kay, K. N., T. Naselaris, et al. (2008). "Identifying natural images from human brain activity." Nature **452**(7185): 352-355.

Khalili-Mahani, N., Chang C, van Osch MJ, Veer IM, van Buchem MA, Dahan A, Beckmann CF, van Gerven JM, Rombouts SA (2013). "The impact of "physiological correction" on functional connectivity analysis of pharmacological resting state fMRI." Neuroimage **65**: 499-510.

Koch, M. A., Norris, D. G., Hund-Georgiadis, M. (2002). "An investigation of functional and anatomical connectivity using magnetic resonance imaging." Neuroimage **16**(241-250).

Langer, N., Pedroni A, Jäncke L (2013). "The problem of thresholding in small-world network analysis." PLoS One **8**(1).

Larsson, J., Heeger DJ (2006). "Two retinotopic visual areas in human lateral occipital cortex." J Neurosci. 26(51): 13128-13142.

Lindquist, M., Spicer J, Asllani I, Wager TD (2012). "Estimating and testing variance components in a multi-level GLM." Neuroimage **59**(1): 490-501.

Markov NT, Misery P, Falchier A, Lamy C, Vezoli J, Quilodran R, Gariel MA, Giroud P, Ercsey-Ravasz M, Pilaz LJ, Huissoud C, Barone P, Dehay C, Toroczkai Z, Van Essen DC, Kennedy H, Knoblauch K (2011). Weight Consistency Specifies Regularities of Macaque Cortical Networks. Cerebral Cortex **21**: 1254-1272.

Markov, NT, Vezoli J, Chameau P, Falchier A, Quilodran R, Huissoud C, Lamy C, Misery P, Giroud P, Ullman S, Barone P, Dehay C, Knoblauch K, Kennedy H (2014). "Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex." J Comp Neurol. **522**(1):225-259.

Marrelec, G., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehéricy, S., Doyon, J., Benali, H (2006). "Partial correlation for functional brain interactivity investigation in functional MRI." Neuroimage **32**, 228–237.

McIntosh, A., Gonzalez-Lima F (1991). "Structural modeling of functional neural pathways mapped with 2-deoxyglucose: effects of acoustic startle habituation on the auditory system." Brain Res. **547**:295-302.

Meek, C (1995). "Causal inference and causal explanation with background knowledge." Proceedings of the 11th Annual Conference on Uncertainty in Artificial Intelligence; 403–410.

Morgan, Timothy M. and L. Douglas Case (2013). "Conservative Sample Size Determination for Repeated Measures Analysis of Covariance". Ann Biom Biostat. **1**(1).

Murphy, K., Birn, RM, Handwerker DA, Jones TB, Bandettini PA (2009). "The impact of global signal regression on resting state correlations: are anti-correlated networks introduced?" Neuroimage **44**(3): 893-905.

Patel, R., Bowman, F., Rilling, J (2006). "A Bayesian approach to determining connectivity of the human brain." Hum. Brain Mapp **27**: 267-276

Perkel, D. J., J. Bullier, And H. Kennedy (1986). "Topography of the afferent connectiivty of area 17 of the macaque monkey: a doublelabelling study". J. Comp. Neurol. **253:** 374-402.

Power, JD, Schlaggar BL, Lessov-Schlaggar CN, Petersen SE (2013). "Evidence for hubs in human functional brain networks." Neuron. **79**(4):798-813.

Quigley, M. A. e. a. (2003). "Role of the corpus callosum in functional connectivity." Am. J. Neuroradiol **24**(208-212).

Restani, L, Caleo M (2016). "Reorganization of Visual Callosal Connections Following Alterations of Retinal Input and Brain Damage." Front Syst Neurosci. **10**: 86.

Saenz, M, Fine I (2010). "Topographic organization of V1 projections through the corpus callosum in humans." Neuroimage **52**(4):1224-9

Salin, P., Bullier, J., 1995. "Corticocortical connections in the visual system: structure and

function." Physiol. Rev. **75**, 107–154.
Scheines, R., Spirtes, P., Glymour, C., Meek, C., Richardson, T (1996). "TETRAD 3: Tools for Causal Modeling User's Manual." http://www.phil.cmu.edu/projects/tetrad/old/tet3/master.htm
Schmidt, M (2006). "L1precision: Matlab code for MAP estimation of Gaussian Graphical Model Precision with L1-regularizer." http://www.cs.ubc.ca/~schmidtm/Software/L1precision.html.
Schmittmann, V. D., S. Jahfari, et al. (2015). "Making Large-Scale Networks from fMRI Data." PLoS One **10**(9): e0129074
Schölvinck, M.L., Maier, A., Ye, F.Q., Duyn, J.H., Leopold, D.A., 2010. Neural basis of global resting-state fMRI activity. Proceedings of the National Academy of Sciences 107, 10238.
Sereno, M., Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB. (1995). "Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging." Science **269**(5212): 889-893.
Shams, S. M., Afshin-Pour B, Soltanian-Zadeh H, Hossein-Zadeh GA, Strother SC (2015). "Automated iterative reclustering framework for determining hierarchical functional networks in resting state fMRI." Hum Brain Mapp. **36**(9): 3303-3022.
Shmuel A, Korman M, Sterkin A, Harel M, Ullman S, Malach R, Grinvald A (2005). "Retinotopic axis specificity and selective clustering of feedback projections from V2 to V1 in the owl monkey." J Neurosci. **25**(8):2117-31.
Smith, A. T., K. D. Singh, et al. (2001). "Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex." Cereb Cortex **11**(12): 1182-1190.
Smith, S. M., Karla L. Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F. Beckmann, Thomas E. Nichols, Joseph D. Ramsey, Mark W. Woolrich (2011). "Network modelling methods for FMRI." Neuroimage **54**: 875-891.
Tian, L., Kong, Y., Ren, J., Varoquaux, G., Zang, Y., Smith, S. M. (2013). "Spatial vs. Temporal Features in ICA of Resting-State fMRI – A Quantitative and Qualitative Investigation in the Context of Response Inhibition." *PLoS ONE*, **8**(6), e66572.
Tibshirani, R (1996). "Regression shrinkage and selection via the Lasso." Journal of Royal Statistical Society: Series B. **58**(1): 267–288.
Tootell, Roger B. H., Nouchine K. Hadjikhani, Wim Vanduffel, Arthur K. Liu, Janine D. Mendola, Martin I. Sereno, and Anders M. Dale (1998). "Functional analysis of primary visual cortex (V1) in humans." PNAS, **95**(3), 811-817.
Spirtes, P., Glymour, C., Scheines, R., (1993). "Causation, Prediction, and Search." Lecture Notes in Statistics, **81**. Springer-Verlag.
van den Heuvel, M., Mandl RC, Kahn RS, Hulshoff Pol HE (2009). "Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain." Hum Brain Mapp **30**: 3127-3141.
van den Heuvel, Martijn P., Edward T. Bullmore, Olaf Sporns (2016). "Comparative Connectomics" Trends in Cognitive Sciences **20** (5), 345-361.
van den Heuvel, MP, Siemon C. de Langea, Andrew Zaleskyb, Caio Seguinb, B.T. Thomas Yeoc, Ruben Schmidt (2017). "Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations." Neuroimage. **152:** 437-449.
Van Essen, D. C., Newsome WT, Bixby JL (1982). "The pattern of interhemispheric connections and its relationship to extrastriate visual areas in the macaque monkey". J Neurosci. 2(3):265-283.
Wandell BA, Dumoulin SO, Brewer AA. Visual Field Maps in Human Cortex. *Neuron.* 2007/10/25/ 2007;56(2):366-383.
Wang, B., Guo, J., Yan, T., Ohno, S., Kanazawa, S., Huang, Q., Wu, J., 2016. Neural Responses to Central and Peripheral Objects in the Lateral Occipital Cortex. Front Hum Neurosci 10, 54.

Winawer, J., H. Horiguchi, et al. (2010). "Mapping hV4 and ventral occipital cortex: the venous eclipse." J Vis **10**(5): 1.

Yang, G. J., J. D. Murray, et al. (2016). "Altered Global Signal Topography in Schizophrenia." Cereb Cortex.

Zalesky, A., Cocchi L, Fornito A, Murray MM, Bullmore E (2012). "Connectivity differences in brain networks." Neuroimage **60**(2): 1055-1062.

Introduction to Manuscript 3

In Manuscript 2, we have successfully implemented network models in a system within the brains of healthy individuals with relatively comprehensive prior studies to which we can compare our results. Next, we seek to utilize our promising methods from Manuscript 2 in less explored conditions: relatively fine-scale whole-brain network connectivity of patients with mild cognitive impairment (MCI). Extending our methods to a pathological state and into other brain areas beyond the visual cortex are important next steps in order to test the generalizability of the methods, to push the boundaries of our knowledge of MCI, and to determine whether the methods that best predict direct connections in fine-scaled networks can add information for fMRI-based prediction of MCI progression. The ultimate goal is to improve the clinical applicability of computational prediction approaches endeavoring to identify MCI individuals who will progress to dementia vs. those who will not. This improvement would potentially lead to more informed prognoses and earlier diagnosis. We may also learn about the pathophysiology of MCI/dementia by seeing which MRI measures are most informative in the classification.

Many imaging studies have been conducted in the last decade which look at the changes in brain structure that occur in AD as well as studies looking at differences in functional connectivity observed in patients with AD or MCI as compared to healthy age-matched controls. It has been frequently confirmed that there are clear patterns of cortical atrophy associated with AD, determined by computing cortical thickness or gray matter volume measures, among others, using T1-weighted anatomical MRI scans. The most common pattern of cortical atrophy is in the medial-temporal lobe and the associative cortex, but other patterns including limbic-predominant and hippocampal-sparing have also been shown (e.g. Noh et al., 2014; Persson et al., 2017; Ferriera et al., 2017; Ferriera et al., 2019). Strong evidence has also shown altered FC within the default-mode network (DMN), and in the connectivity of other regions with the DMN (Balachandar et al., 2015; Binnewijzend et al., 2012; Grothe and Teipel, 2016; Li et al., 2013; Weiler et al., 2014; Zhou et al., 2015; Tam et al., 2015; Greicus and Kimmel, 2012).

Aside from the DMN, the salience network (He et al., 2014), executive control network (Balachandar et al., 2015; Weiler et al., 2014), and frontoparietal-control network (Grothe and Teipel, 2016) have also been shown to have altered FC in MCI or mild AD compared to healthy controls (HC). Additionally, the fusiform gyrus (Cai et al., 2015a), thalamus (Cai et al., 2015b;

Zhou et al. 2013), parahippocampal gyrus (Liu et al., 2015; Cai et al., 2015a) , hippocampus (Sohn et al., 2014), amygdala (Yao et al., 2013), middle frontal gyrus (Cai et al., 2015a; Cai et al., 2015b; Zhou et al., 2013) , and temporal lobe (Zhou et al., 2013; Zhou et al., 2015) have been shown to vary from healthy controls to MCI or mild AD patients in their FC with regions within the DMN or other brain areas. In a meta-analysis combining four datasets with healthy controls and amnesic MCI patients (Tam et al., 2015), the largest differences in resting-state FC (RS-FC) involved the superior medial frontal cortex (including the anterior cingulate), dorsomedial prefrontal cortex, striatum, and middle temporal lobe. Other studies have considered whole-brain FC (Wang et al., 2013; Zhou et al., 2015) in the study of MCI/AD. A good qualitative summary of the past studies showing altered fMRI in the context of MCI or AD is presented in Zhou et al. (2017).

The broad involvement of cortical and subcortical regions leads us to pursue a whole-brain, data-driven approach to studying MCI. In support of this choice, Badhwar et al., 2017, recommend pursuing whole-brain analyses to avoid biases. Based on Manuscript 2 of this thesis, we have chosen to look at a relatively fine-scaled brain parcellation (the HCP cortical Parcellation), as compared to other frequently used atlases such as AAL or the Freesurfer cortical segmentation. We opted to use a fine-scaled brain parcellation with the goal of refining our understanding of what underlies the larger-scaled findings that have previously been observed.

Our computational approach in our pursuit to better understand MCI is that of machine learning classifiers, allowing us to predict within the population which patients will progress to dementia. However, it is important to note that individuals with dementia and MCI are quite heterogeneous in terms of their presentation. Because of this heterogeneity, the overall accuracy of prediction of progression from a healthy state or MCI to AD has been too low and/or not adequately reliable for widespread clinical implementation. From this challenge have sprung studies looking at combining data across modalities and studies utilizing more complex computational models in order to try to capture the complex relationship between AD precursor states and AD.

There are many studies that have pursued this classification question in recent years. Therefore, in the literature review in Manuscript 3, we have focused on only the most relevant articles which met the following criteria: 1) MRI and/or fMRI markers were the only imaging

features utilized in the model (clinical or socio-demographic features were allowed as well) (i.e., the studies used relatively comparable methods to those of our own study); 2) both sensitivity and specificity must be over 70%, overall accuracy over 70%, or either sensitivity or specificity over 90% with the other over 60% (i.e., they were relatively successful in their classification task in terms of their chosen metric); 3) validation using an inner and outer loop, or in other words, a cross-validation procedure utilized on one dataset in order to optimize the model, and then a second validation of the optimized model on unseen data (i.e., the studies followed good practice regarding validation in their machine learning approaches). Inspired by the findings in our literature review and keeping population heterogeneity in mind, we, too, seek to evolve our computational approach beyond the standard by proposing a multi-classifier method looking to optimize precision for a subgroup of the MCI population rather than overall accuracy. Our hope is that this approach may contribute to developing a clinically useful classification procedure that considers both computationally optimal conditions and real-world clinical feasibility.

Manuscript 3

High positive predictive value in classification of non-progression from mild cognitive impairment to dementia

Debra Dawson[1], Ziqi Hao[1,2], Wenzong Xia[1], Tianzhu Fu[1], Kelvin Mok[1], Li Lin[4], Ying Han[4,5,6,7], Pierre Bellec[4], Amir Shmuel[1], for the Alzheimer's Disease Neuroimaging Initiative*

1. McGill University, Montreal, Québec,

2. Electronic Information Engineering, Sichuan University, Chengdu, Sichuan,

3. Centre de recherche de l'institut de gériatrie de Montréal, Montréal, Québec

4. Department of Neurology, XuanWu Hospital of Capital Medical University, Beijing, China, 100053

5. Center of Alzheimer's Disease, Beijing Institute for Brain Disorders, Beijing, China, 100053

6. Beijing Institute of Geriatrics, Beijing, China, 100053

7. National Clinical Research Center for Geriatric Disorders, Beijing, China, 100053

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Correspondence:

Amir Shmuel, Ph.D.
3801 University Street, room NW109
Montreal, Quebec H3A 2B4, Canada
amir.shmuel@mcgill.ca
Ying Han, Ph.D.
Xuanwu Hospital, Capital Medical University, Beijing, China
hanying@xwh.ccmu.edu.cn

Acknowledgements:

Supported by The National Key Research and Development Program of China (2016YFC1306300), National Natural Science Foundation of China (Grant 616 33018), Beijing Municipal Commission of Health and Family Planning (PXM2019_ 026283_000002), awarded to Dr. Ying Han, and the Natural Sciences and Engin-eering Research Council of Canada (RGPIN 2015-05103) awarded to Amir Shmuel.

Data collection and sharing for this project was in part funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Data were provided in part by OASIS-3: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P50AG00561, P30NS09857781, P01AG026276, P01AG003991, R01AG043434, UL1TR000448, R01EB009352. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly.

Abstract:

Currently, there are no clear methods for determining the prognosis of individuals presenting with Mild Cognitive Impairment (MCI). Early identification of those MCI who will progress to dementia and those who will not may be possible with a machine learning computational analysis of MRI images. We hypothesize that accurate classification of groups of individuals who progress from MCI to dementia (Progressor) and those who do not (Stable) can be achieved using fine-scaled whole-brain resting-state functional connectivity (RSFC), gray matter volume, and cortical thickness. Given the heterogeneity of the MCI population and limited accuracy achieved in prior studies that considered all MCI together as one group, we apply a subtyping algorithm to the MRI-based measures. The data analyzed (RSFC and T1-weighted-MRI obtained at 3 Tesla) comes from the Alzheimer's Disease Neuroimaging Initiative, Dr. Ying Han's lab at the XuanWu Hospital - Capital Medical University, and the Open Access Series of Imaging Studies (OASIS) 3 database. Analyses were done on the whole brain with the Human Connectome Project (HCP) cortical parcellation combined with the Freesurfer subcortical segmentation. A grid-search with nested cross-validation was performed and classifiers with high class-wise positive predictive value (PPV) were identified. While each individual classifier only allows for confident classification in a small number of subjects, combined results from a selected set of high-PPV analyses allow for confident classification in a higher proportion of subjects. The set of classifiers identified with high PPV in a *stable class* performs very well with a mean prevalence-corrected PPV of 96%. We also examined the impact of multivariate (L1 regularized Inverse Covariance) vs. bivariate (Pearson's Correlation) functional connectivity measures. These analyses reveal that correlation is more informative than L1 regularized Inverse Covariance. In addition, we tested the impact of the size of ROIs by repeating the analysis with the relatively coarse Desikan-Killiany-Tourville (DKT) delineation. While the results on training data with the coarser scaled DKT ROIs were fairly successful, fewer subjects were correctly classified overall when compared to the analyses with the HCP ROIs. Lastly, we examined the differences in functional connectivity, gray matter volume, and cortical thickness between the identified MCI subtypes. The most important brain regions implicated in these differences were found in the visual/occipital cortex, sensorimotor cortices, and medial parietal regions. Other regions, such as those in the default mode network, were also implicated. We conclude that high-PPV classification of a subset of individuals using a set of

complementary classifiers is a promising method for predicting non-progression from mild cognitive impairment.

Introduction/Background:

Alzheimer's Disease (AD) is a debilitating neurodegenerative disease affecting millions around the world. The causes of this disease are not well understood, nor are the risk factors for individuals who may develop AD in the future or the protective factors for those who will not. Mild Cognitive impairment (MCI) is defined as a cognitive impairment that is not sufficient to meet the criteria for dementia, and it is often a precursor state for AD and other forms of dementia. Studies suggest that as much as 50% of MCI patients will progress to dementia within 5 years (Gauthier et al., 2006; Arbabshirani et al., 2017). Investigating the characteristics of those with MCI may prove useful for identifying potential screening procedures for individuals who will develop AD.

Early identification of at-risk individuals via screening will allow us to study the progression of MCI from an earlier stage and will eventually allow for the study and implementation of treatments before the cognitive decline becomes severe. Similarly, if we can identify which individuals are not likely to progress to dementia from a state of MCI, the health care system and groups conducting clinical trials can save their resources for those who have a higher risk of progression. Additionally, there is the benefit of peace of mind that families can have knowing that their loved one is not expected to deteriorate cognitively. Although clinical criteria – such as insidious and worsening cognitive impairment - have been developed to identify individuals who have AD (McKann et al., 2011), as of yet no reliable screening has been established for asymptomatic individuals. There have been some genetic linkages made (Dyukov et al., 2015), but these genetic links (ex. the APOE gene) are not exclusive to individuals with AD, and alternatively, not all who have AD will have these markers. Imaging-based studies are providing results that paint MRI as a promising screening tool, reliably showing detectable differences between healthy individuals and those with AD. Identifying MRI-based differences within the group of individuals with MCI may be a viable option for predicting those with a higher likelihood of progression and those with a higher likelihood of remaining stable in a state of MCI.

The question of how to predict whether an individual with MCI will progress to AD through imaging or other biomarkers is actively pursued by many groups. Early studies with small sample sizes showed that MRI brain imaging could be useful in the differentiation of AD subjects from non-AD subjects (e.g. Kloppel et al., 2008; Fan et al., 2008; Magnin et al., 2009; Plant et al., 2010; Liu et al., 2013). Other studies showed that fMRI, PET, as well as clinical measures could be predictors with varying degrees of success (e.g. Chen et al., 2011; Gray et al., 2012; Vandenberghe et al., 2013; Challis et al., 2015; Battista et al., 2017). Larger cohorts of subjects are now becoming possible with the emergence of shared databases. One of the largest existing MCI/AD imaging databases is the Alzheimer's Disease Neuroimaging Initiative (ADNI, *adni.loni.usc.edu*). In 2003 the ADNI project began as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD, with the goal of gathering abundant MRI and other imaging data from individuals with AD, as well as healthy controls (HC). Since then, the project has expanded to include data from individuals with MCI and other states with some levels of cognitive decline (ex. the "Significant Memory Concern" group). This data is available for use by researchers around the world and has encouraged many MRI/brain imaging-based studies of MCI and AD. Other imaging databases have also arisen since ADNI began, including the Open Access Series of Imaging Studies (OASIS), which began in 2007.

The use of machine learning methods in imaging neuroscience has taken off along a similar trajectory as the increase in AD MRI population studies. In recent years there have been numerous studies that utilized the promising methodological approaches possible with machine learning to the problem of predicting whether an individual with MCI will progress to AD. These studies are not only numerous but also highly diverse in their approaches to the problem. While the main threads of utilizing multiple biomarkers for predicting AD and using machine learning computational approaches remain consistent among them, the choice of biomarkers, the details of how those biomarkers are extracted/quantified, and the choice of machine learning algorithm and hyperparameters differ in meaningful and sometimes dramatic ways.

There have been a handful of review articles published in recent years which aimed to capture the current evidence and level of progress on this research question. Each of them can only begin to address a part of the evidence given the abundance of past and new articles on the topic. Nevertheless, these reviews are a great starting point to capture a general sense of recent trends in the field and to familiarize oneself with the research groups actively looking into this

question. Some more recent review articles include Leandrou et al. (2018), Liu et al. (2018), Sarica et al. (2017), and Rathore et al. (2017).

Leandrou et al. (2018) reviewed MRI-based studies looking at classification of MCI, AD, and HC using quantitative structural MRI techniques such as voxel-based morphometry, volumetric measurements in ROIs, cortical thickness, shape analysis, and texture analysis. Included in their results are classification accuracies for MCI Progressors vs. HC. Overall, most of the reviewed studies find 80% or less accuracy of classification. Some studies find up to 90% accuracy, though their validation procedures are questionable. They also present MCI Progressors vs. Stables studies where the accuracies of results are approximately 75%. Of relevance, the insights gained from this article include that ROI-based volume measures were most informative in the MRI-based MCI classification task. In fact, studies had found that structural MRI could better predict conversion than PET measures (Desikan et al., 2010) and CSF measures (Desikan et al., 2010, and Vermuri et al., 2009).

Sarica et al. (2017) provide a review of random forest classifiers applied to AD which covers a handful of Stable vs. Progressors studies. The most relevant studies - Moradi et al. (2015) and Ardekani et al. (2017) - use 1.5T MRI data and neuropsychological scores as features and achieve an overall accuracy of 82%.

Rathore et al. (2017) reviewed 81 studies published between 1985 and June 2016 utilizing various feature modalities including MRI. They focussed on high-quality studies that utilized some form of cross-validation in their methods in order to better ensure generalizability of the results as compared to studies that did not validate. Study accuracies of the most successful models in classifying Progressors vs. Stable were ~80%. They also found that FC measures specifically showed promise, achieving high classification results in two studies: 97.00% for AD vs. MCI (Challis et al., 2015) and 91.90% for MCI vs. HC (Jie et al., 2014). This review also emphasizes the importance of appropriate validation methods in machine learning. The most robust test of a model is with cross-validation during model optimization and further testing on completely unseen data. The results currently achievable with this high standard for demonstrating generalizability is illustrated in the CADDementia 2015 challenge (Bron et al., 2015), where the best performing study yielded an area under the receiver-operating-characteristic curve of 78.8%.

Liu et al. (2018) review papers with multimodal models utilizing MRI, PET and other markers. Perhaps the most promising study utilizes amyloid PET and achieves 86% accuracy, 81% sensitivity, and 90% specificity, though it is important to note that amyloid PET is not widely available and thus has limited current clinical and research use. The other studies with top results utilize data from many modalities and achieve accuracies in the high 70's or up to 80 percent. Liu et al. provide a set of recommendations for future directions in MCI machine learning studies. One of their strong recommendations is subtyping the MCI subjects in order to deal with the heterogeneity of the disease and disease progression.

This leads to another element that we would like to introduce further: the heterogeneity in the population of AD/MCI. One can see from the results that the studies mentioned above struggled to achieve an overall accuracy that would be sufficient to warrant classifier implementation in a clinical setting in order to aid with prognosis. Theories as to why we are not able to achieve better accuracy include the fact that we have not yet landed upon the right predictor features within the MCI population, or that the data is too heterogeneous to be classified into 2 groups (progressors and non-progressors). In the latter case, it is possible that either or both of these groups have subtypes and combining these subtypes into one big group may be muddying the feature representation of progression. Possible etiologies of MCI include several forms of dementia including AD, fronto-temporal lobe degeneration, dementia with Lewy bodies, vascular disorders, and psychiatric disorders (Petersen and Negash, 2008; Arbabshirani et al., 2017). There is also evidence that subtyping based on the stage of disease progression can improve the accuracy of classification (Eskildsen et al., 2013, Wee et al., 2013). The need for subtyping and an acknowledgement of the heterogeneity in AD/MCI was emphasized in the conclusion from the review by Rathore et al. (2017). Given these findings and recommendations, our manuscript seeks to explore whether subtyping can lead to the successful classification of stable or progressor MCI subjects. Past studies that implemented subtyping in AD/MCI using MRI-based measures include Noh et al. (2014), Hwang et al. (2015), Ferreira et al. (2017), Badhwar et al. (2017), and Ferreira et al. (2019). These studies use predominantly structural MRI measures such as cortical thickness and gray matter volume, though Orban et al. (2017) applied subtyping in AD utilizing resting-state fMRI. Most prior subtyping studies have found between 3 or 4 subtypes for MCI/AD using these measures. Therfore, in our study we also split our MCI progressor and stable groups into up to 4 groups each.

Another approach we can take to make studies with limited accuracy results more clinically relevant is by shifting our focus to other metrics which can still be meaningful and clinically relevant despite the limited overall accuracy. These measures include sensitivity, specificity, positive predictive value, and negative predictive value - the typical measures one obtains from a confusion matrix. The ideal test would have either both high sensitivity (Sens) and high specificity (Spec), or high positive predictive value (PPV) and high negative predictive value (NPV). This implies a balance in the usefulness of both positive and negative test results (PPV and NPV), or, on the other hand, the predictability of both the positive and negative diagnostic states (Sens and Spec). Sensitivity is defined as the proportion of individuals correctly identified as having a given diagnosis relative to the total number of individuals with that diagnosis. There are instances where a diagnostic method with high sensitivity can be useful, despite low specificity. For example, high sensitivity can be very useful despite low specificity if it is important to be sure to identify all individuals who may indeed have a diagnosis, even at the expense of identifying a number of FP. One example of such a situation is in the case of brain hemorrhage. It is better to catch all those who have a brain hemorrhage and pursue active monitoring, along with some who are falsely identified as having a hemorrhage and who undergo unnecessary monitoring, as the monitoring is harmless; in contrast, missing a hemorrhage can very often be deadly. The converse can be true when it is critical to be sure to identify negative results (high Spec, low Sens): examples include situations involving quarantine– one must be sure that those released from quarantine are indeed not carriers of the pathogen.

The metric we will focus on here is positive predictive value; i.e., the proportion of individuals correctly identified as having a given diagnosis relative to the total number of individuals testing positive for that diagnosis. This is useful in the case where a test provides a positive result. If the PPV is high, one can be confident in any positive test result obtained, but this is often at the expense of low negative predictive value, where a negative test result is not very reliable. We are aware of two studies that have attempted to predict MCI converting to AD with imaging measures and utilized subtyping to identify a subgroup within which we can have greater confidence in the results. These studies are Dansereau et al. (2017) and Tam et al. (2019). Dansereau et al. (2017) implemented subtyping in a population of AD patients based on differences in gray matter volume and functional connectivity as compared to HC. They then hypothesized that MCI patients who will progress to AD show MRI features that are similar –

although of a lesser magnitude – to those seen in AD patients. Therefore, they used a classifier to assign MCI individuals to the AD subtypes. One of the subtypes identified resulted in a 90% precision in predicting progression to AD. Tam et al. used structural MRI and clinical measures and achieved, for one of the subgroups of MCI progressors, 87.8% PPV, adjusted for the prevalence of MCI progression in the population. These studies show once again the promise of subtyping in attempting to classify individuals with MCI, and also that overall accuracy may not be the ideal metric on which to focus. Perhaps a metric associated with individual subtypes would allow for more meaningful insights. In our current study, we examine subgroups of MCI individuals and identify which subgroups have high PPV. If we can develop a method using classifiers that predict high PPV in specific subtypes, we can have high confidence in the case of a positive test result (a clinically relevant finding), despite the fact that the overall accuracy across all subjects is not sufficient for clinical usefulness.

Given all of the above, we are inspired to test the boundaries of what can be achieved in classification with MRI alone, as the evidence shows that it is a meaningful predictor of progression from MCI to AD, and that there are still promising avenues to explore with this modality, such as subtyping and different ways of quantifying the MRI trends. Thus, here we will classify MCI subjects into groups of those who will progress and those who will not progress, following subtyping within these groups. We aim to capture the different phenotypes or stages of disease present in a diverse grouping of individuals with MCI. We will also use a brain parcellation with relatively small ROIs, allowing for finer-scaled regional implications while still limiting the volume of data relative to a voxel-wise analysis. Based on our own recent studies (Manuscript 2 in this thesis), we test whether L1-norm Regularized Inverse Covariance (ICOV) would be more useful in applying fine-scaled functional connectivity within the context of MCI, as it has been shown to be a strong method for fine-scaled connectivity in the model network of the visual cortex (Dawson et al., 2013; Dawson et al., under revision for Neuroimage, Manuscript 2 in this thesis). Inverse covariance-based analyses regress out "network effects" and thus have the potential of demonstrating differences that are masked out by broadly reaching large-scaled network effects.

Our study hypotheses are the following:

1. MRI-based measures such as cortical thickness, gray matter volume, and functional connectivity are different between MCI individuals who will progress to dementia and those who will not. The differences are sufficient to be informative in a machine learning classification aiming to separate these groups.

2. The MCI cohort is heterogeneous within those who will progress to dementia and possibly among those who will not. Thus, subtyping will be required to find meaningful feature sets representing the cohort.

3. Given the heterogeneity of the MCI cohort, a single classifier is not likely to be well-tuned to properly classify all MCI individuals. Thus, a combination of separately optimized classifiers may allow for precise classification of a greater number of subjects.

4. Selecting model parameters to successfully optimize PPV may be more feasible than optimizing overall accuracy.

5. Along with the traditional choice of correlation-based FC, ICOV-based FC may be informative in distinguishing progressors from non-progressors.

6. A finer-scaled parcellation may be more successful in classifying subjects than a larger-scaled parcellation due to the greater spatial specificity of measures utilized in classification.

And our Specific Aims are:

1. To apply a novel data-driven approach to identify the minimum necessary number of successful and non-redundant classifiers in order to correctly classify the largest possible number of subjects.

2. To determine the overall PPV of a set of classifiers selected for optimizing PPV under their own parameters, and to determine the reliability of these results by using nested cross-validation.

3. To observe how the choice of parcellation scale affects classification results by runing all analyses on both a finer-scaled parcellation and a coarser-scaled parcellation.

4. To generate classifiers utilizing various numbers of subtypes within the Progressor and Stable MCI groups and to observe which classifiers are most often selected among the set of classifiers used in the overall classification of subjects. This will allow us to examine more closely the feature profiles of the subtypes in the selected classifiers and to identify the number of meaningful subtypes likely to exist.

5. To generate classifiers using 1) the structural measures of cortical thickness and gray matter volume of regions and 2) the FC measures of ICOV or correlation, and to observe which classifiers are most often selected to be among the set of classifiers used in the overall classification of subjects. This will allow us to compare how informative structural vs. functional measures as well as ICOV-based vs. correlation-based FC are in the context of predicting progression from MCI to dementia.

Methods:

Data

The data we analyzed comes from three sources: the *Alzheimer's Disease Neuroimaging Initiative* (*ADNI*) database, the *Open Access Series of Imaging Studies (OASIS)* 3 database, and from our collaborators at Dr. Ying Han's lab from XuanWu Hospital – Capital Medical University, China. We have used data from MCI subjects who have undergone T1-weighted structural MRI and resting-state fMRI scanning using a 3T scanner. We then analyzed the fMRI data of individuals with MCI who have been followed over a period of at least 6 months showing either a stable diagnosis of MCI or MCI progressing to dementia. Progressor subjects were followed for, on average, 27 months before progression. Many subjects have undergone multiple scans. We have used all available scans, accounting for repeated measures in statistical testing, and treating individual scans as independent datasets when testing our classifiers.

The ADNI database contains freely available data utilizable in assessing the rate of change of cognition, function, brain structure, and biomarkers in elderly controls as well as subjects with MCI or AD. Detailed data acquisition parameters vary slightly given the site of acquisition or scanner type and can be found at http://adni.loni.usc.edu/methods/documents/mri-protocols. In the ADNI protocol, a participant was classified as having MCI if they had memory complaints and objective cognitive loss (based on neuropsychological testing via the Wechsler

Memory Scale Logical Memory II test), but had intact functional abilities. The ADNI database contains 47 stable MCI subjects and 19 subjects with MCI who progress to dementia within the time frame of the study.

OASIS-3 is a longitudinal dataset comprised of neuroimaging data as well as clinical and cognitive biomarkers for normal aging and Alzheimer's Disease. Details of the MRI data acquisition can be found in Lamontagne et al. (2019). Cognitive status was assessed with the Clinical Dementia Rating (CDR) Scale, with CDR 0 indicating normal cognitive function, CDR 0.5 very mild impairment, CDR 1 mild impairment, and CDR 2 or greater indicating a demented state. The OASIS-3 database contains 76 stable MCI subjects and 107 subjects with MCI who progressed to dementia within the time frame of the study.

Finally, we thank our collaborators in Dr. Ying Han's lab from XuanWu Hospital – Capital Medical University, China, who generously provided us with MRI and fMRI data from 21 stable MCI subjects and 32 progressors. Inclusion criteria for MCI subjects included: (1) memory complaint; (2) objective memory impairment as measured by the Mini-Mental State Examination (MMSE), the Montreal Cognitive Assessment (MoCA), or the auditory verbal learning test (AVLT); (3) no or minimal impairment of daily life activities; (4) a CDR score of 0.5; (5) not being diagnosed with dementia; (6) hippocampal atrophy confirmed by structural MRI; and (7) right-handed individuals of the Han nationality. See Li et al. (2016) for details of the data acquisition.

	Progressors	**Stables**	**Total**
Subjects	158 (52.3%)	144 (47.7%)	302
Sex	55.7% M	54.9% M	55.3% M
Age: mean ± std	73.7 ±7.7	73.9 ±8.9	73.8±8.3
Sites represented	12/12	12/12	12/12
Scans per subject: mean ± std	1.56 ± 1.26	2.53 ± 1.81	2.03± 1.62
Total scans	247 (40.4%)	365 (59.6%)	612

Table1. Group Demographics. In our full data set, there are 8 sites represented across the ADNI subjects, 3 sites represented across the OASIS3 subjects, and 1 site across subjects from Dr. Han's group.

Preprocessing

Resting-state runs were preprocessed in the following way: FSL's FEAT software package was used to delete the first three volumes, as well as perform motion correction, high-pass filtering (0.01 Hz), slice-time correction, and registration to the baseline T1-weighted (MPRAGE) anatomical volume. Next, we used ICA-AROMA to remove remaining motion artifacts from the data, a step especially important to consider in patient populations (Pruim et al., 2015a; Pruim et al., 2015b). This strategy first computes ICA on a subject's functional scan, then classifies the components as motion based on the maximum realignment parameter correlation, the edge fraction, the CSF fraction, and the high-frequency content. Then, the motion components are linearly regressed from the subject's scan. Following ICA-AROMA, a white matter (WM) mask and a cerebrospinal fluid (CSF) mask were created from the subject's anatomical volume using FSL's FAST software package. The masks were then eroded by one voxel, thresholded (80% tissue type probability), and registered to the preprocessed functional image. Finally, in order to denoise the signal, the average time course from within each of the two masks was regressed out of the functional image voxel time courses. No spatial smoothing was performed in order to retain fine-scaled spatial variations in the signal.

Regions of Interest (ROIs)

Many imaging studies have been conducted over the last decade that looked at the changes in brain structure that occur in AD as well as studies looking at differences in functional connectivity observed in patients with AD or MCI as compared to healthy age-matched controls. It has been frequently confirmed that there are some clear patterns of cortical atrophy associated with AD, and synaptic loss is known to be an early finding in AD, contributing to altered FC (Tampellini, 2015). Strong evidence has shown differences in FC within the default-mode network (DMN; a group of brain regions which activate when the brain is at rest), and in the connectivity of other regions with the DMN (Balachandar et al., 2015; Binnewijzend et al., 2012; Grothe and Teipel, 2016; Li et al., 2013; Weiler et al., 2014; Zhou et al., 2015; Tam et al., 2015; Greicus and Kimmel, 2012). The DMN includes the posterior cingulate cortex and precuneus, medial prefrontal cortex, medial temporal lobes, and inferolateral parietal regions (Greicus and Kimmel, 2012).

Aside from the DMN, a number of brain regions have been implicated in MCI/dementia. Table 2 shows a summary of a number of networks and regions outside of the DMN which have shown altered structure or function.

Network implicated:	**Study**
the salience network	(He et al., 2014; Badhwar et al., 2017)
executive control network	(Balachandar et al., 2015; Weiler et al., 2014; Wu et al., 2013)
frontoparietal-control network	(Grothe and Teipel, 2016)
limbic network	(Badhwar et al., 2017)
Other implicated regions:	**Study**
the fusiform gyrus	(Cai et al., 2015a)
thalamus	(Cai et al., 2015b; Zhou et al. 2013)
parahippocampal gyrus	(Liu et al., 2016; Cai et al., 2015a)
hippocampus	(Sohn et al., 2014;)
amygdala	(Yao et al., 2013)
middle frontal gyrus	(Cai et al., 2015a; Cai et al., 2015b; Zhou et al., 2013)
temporal lobe	(Zhou et al., 2013; Zhou et al., 2015)

Table 2: Regions outside of the DMN having been shown to be altered in the context of AD/MCI and studies which have shown the effect (a non-comprehensive list).

Based on the wide range of regions implicated in the AD/MCI literature, we pursue a whole-brain data-driven study. We include gray matter brain regions from both the cortex and the subcortex. Thus, we add the Freesurfer subcortical segments (16; 8 per hemisphere) to the ROIs obtained from the cortical parcellations that we apply.

One of the aims of our study is to test the effect of the scale of the brain parcellation used. To do this, we have run the full set of analyses on two different cortical parcellations. First, we have used the multimodal group cortical parcellation generated by Glasser et al. (2016), which has 360 ROIs (180 per hemisphere). We apply the group parcellation as an atlas registered to the anatomical scans of the study's subjects. As far as we know, this is the first time this parcellation - which is finer-scaled than most in use for clinical research - is being used in the context of MCI classification. Second, we use the Desikan–Killiany–Tourville (DKT; Klein and Tourville, 2012) Atlas available through Freesurfer, which has 62 ROIs (31 per hemisphere). We selected this larger-scaled atlas after reviewing the choices made by other groups in similar classification studies. While a wide range of different atlases and approaches to building ROIs have been used, the DKT atlas, as well as other parcellations on a similar spatial scale, have been commonly applied.

In the analyses with the HCP parcellation there are 376 ROIs total, with the addition of the Freesurfer subcortical regions. This full set of 376 regions is used when we compute the structural measure of gray matter volume (GMV) for our classification problem (see "Structural Features" below). The 360 cortical regions are used when we compute the cortical thickness (CT) (see "Structural Features" below). However, while assessing the quality of the functional scans and parcel alignment, we noticed that some ROIs overlap areas of consistent signal drop out in the functional MRI data. We removed these ROIs from the functional analyses (see "Functional Connectivity" below). These regions with signal drop out are located in the orbital frontal cortex and inferior temporal cortex: areas known to be susceptible to signal drop out caused by air in the sinuses and the Eustachian tubes. 38 ROIs were removed resulting in 338 ROIs total for the FC analyses. Similarly for the DKT analyses, we compute the GMV in78 regions, and the CT in 62 regions. For the FC analyses, we use 72 ROIs (6 were removed).

Structural Features

As mentioned in the introduction, there are a number of ROI-based MRI measures that have been shown to be useful in the classification of MCI individuals. In these studies, structural features, in particular, have been found very informative as there is often a distinct pattern of atrophy observed in AD (Dansereau et al., 2017; Grothe and Teipel, 2016). Therefore, we include two structural MRI-based measures - cortical thickness (ex. Querbes et al., 2009;

Eskildsen et al., 2013) and gray matter volume (ex. Karas et al., 2004) – as part of our set of features used to classify the subjects. Given the distribution of values obtained across all regions for a given subject, the CT and GMV are normalized to a range of 0 to 1.

Functional Connectivity

There are a number of model choices that can be made in order to quantify FC. We have looked at Pearson's correlation, the most commonly used measure in FC analyses, and a measure known to capture meaningful connectivity in larger-scaled whole-brain analyses. Recent studies have shown that partial correlation-based methods are an appropriate model for studying a small, dense network (Dawson et al., 2013; Kim et al., 2015). We use an L1-norm regularized Inverse Covariance measure (ICOV) in order to compute a sparse partial correlation matrix to reflect the FC between ROIs. The implementation we have used is called L1precision (Schmidt, 2006) and is available at www.cs.ubc.ca/~schmidtm/Software/L1precision.html. We have chosen to include ICOV as a FC measure in our study in large part because it was found to be among the most successful models in Dawson et al. (under revision for Neuroimage, Manuscript 2 in this thesis) where the performance of various models was evaluated in terms of their abilities to approximate direct anatomical connectivity in the fine-scaled, local network of visual areas V1, V2, and V3. Here we are looking to assess whether this promising model can be of use in a broader context, that of the whole-brain, and in the context of brain pathology, but still in the context of relatively fine-scaled connectivity given that we define relatively fine-scaled ROIs. ICOV and correlation FC analyses are done on the entire brain with each of the HCP and DKT cortical parcellations separately combined with the FreeSurfer subcortical regions.

Multi-Classifier Grid-Search Pipeline

Past studies have concluded that a simple, single classifier is not sufficient to achieve predictions that will be clinically useful in the MCI population with MRI-based measures alone. Given this, here we present a concept which we think may lead to improved reliability of results. The flow of analyses of our proposed pipeline is presented in Figure 1.

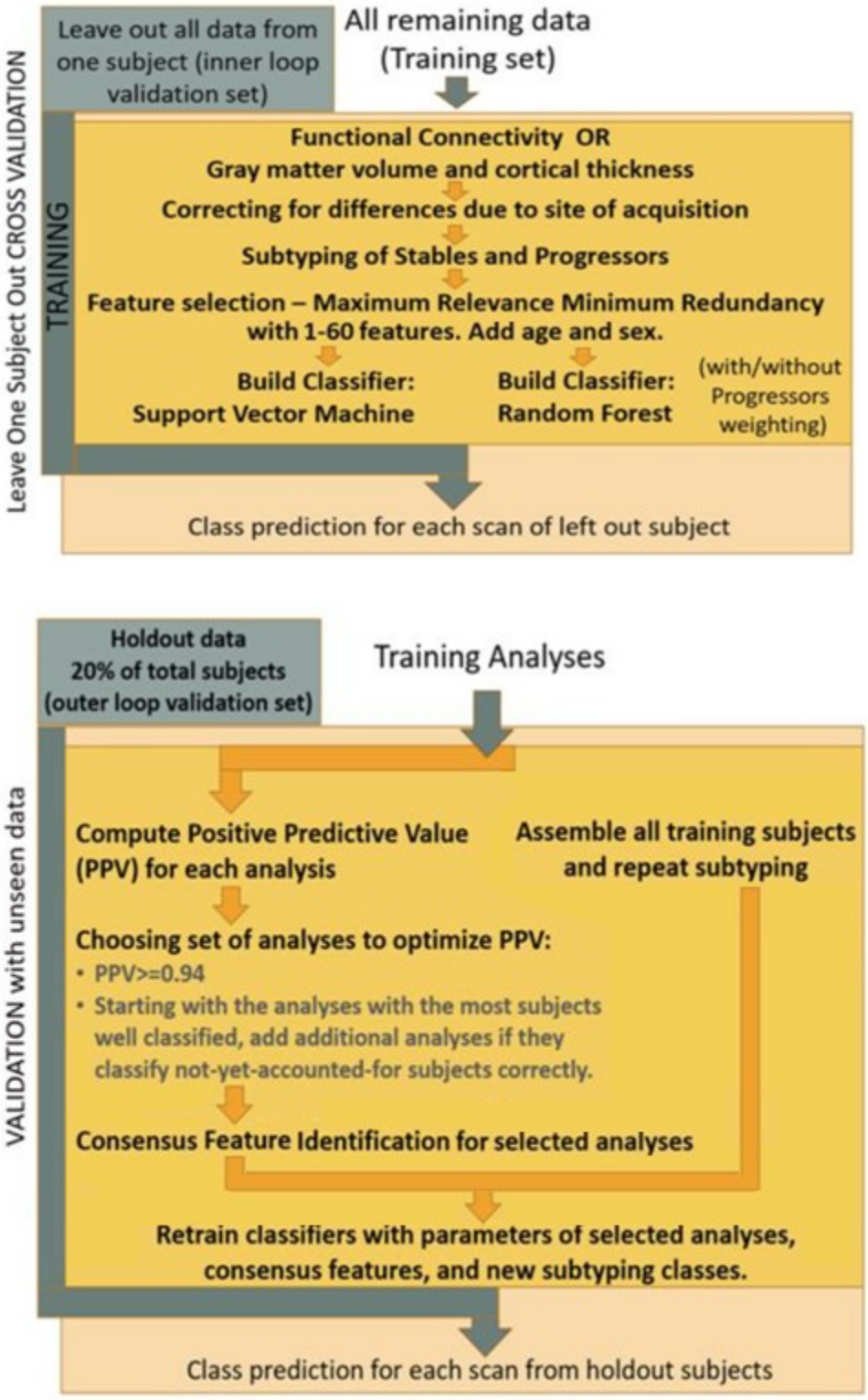


Figure 1. Pipeline workflow for inner loop and outer loop classification. In the inner loop (top half of figure), we use a "leave one subject out" cross-validation procedure in order to classify subjects as Stable (in 1, 2, 3, or 4 subtypes) or Progressor (in 1, 2, 3 or 4 subtypes). We assessed

the functional connectivity (FC) differences between groups via a general linear model. The 10% FC values with the lowest the p-values or the gray matter volume and cortical thickness of all areas are used as possible features for use in Support Vector Machine or Random Forest classifiers. Given that the data is acquired at many different sites, we correct for site-specific effects in the validation data via the ComBat harmonization method. FC / gray matter volume and cortical thickness measures are submitted to the Maximum Relevance Minimum Redundancy feature selection algorithm. A weighting may also be submitted to the classifier to prioritize correct classification of progressor subjects. In the outer loop (bottom half of figure), we identify which of the inner loop analyses are successful by computing the class-wise PPV. Among analyses with >94% PPV we first select the analysis that classifies the most subjects correctly and then add additional analyses if they capture new subjects. Once we have identified this set of high-PPV analyses, subtypes are recomputed using all outer-loop training data, and consensus features are identified.

We use a machine learning (ML) approach, in which we have an outer loop and an inner loop (nested cross-validation). The inner loop is presented in the top half of Figure 1, and the outer loop is in the lower half. Overall, we do 5-fold cross validation (CV) in our outer loop. The first step is to leave out 20% of the data to be our testing data for an outer loop fold. In the inner loop we do leave-one-subject-out (LOSO) CV in order to tune the hyperparameters of classification. We have chosen this cross-validation method as we have multiple scans per subject, and the number of scans per subject is not always consistent, thus for the sake of keeping the number of subjects and scans relatively consistent across inner loop iterations, LOSO was the most logical choice. In each iteration of the inner loop, all of the scans for a given subject from the outer loop training data are left out and become the inner loop test data set. All other subjects from the outer loop training data become the inner loop training data. This way we ensure that no scans from the inner loop test subject remain in the inner loop training data, removing any subject-specific bias. We have pursued a grid search over a large number of different parameters and selections in the design of each individual classifier. These choices will be elaborated below.

1. Choose feature type

The first step is to identify the type of feature we are going to use. This is either FC-based features or structural features, i.e., GMV and CT. If we are using GMV&CT, the full set of all of the measures is carried on to the next step of the pipeline. If we are using FC features, we perform a statistical contrast in order to do a first phase of data reduction/ feature selection. This

step is motivated by the huge quantity of FC-based features we have relative to the number of features we expect to use in our classifier. Recall that with the HCP parcellation, we have 338 ROIs defined in the brain (72 ROIs for the DKT parcellation). This would result in [338x338-338]/2=56,953 connections (or 2,556 for the DKT parcellation). Given that, typically, the optimal number of features in related classification problems is on the order of 10 features, we clearly have far too many features at over 50,000. Thus, the FC measures are assessed for statistically significant differences between the progressor and stable groups using a repeated measures general linear model (GLM). The GLM accounts for the site of origin of the data, the scanner type, as well as the age and sex of the subjects. The FC measures are Fisher-transformed before being submitted to the mixed-effects, repeated-measures GLM obtained from the SurfStat MATLAB toolbox (www.math.mcgill.ca/keith/surfstat/#mixed). In a preliminary analysis, the p-values resulting from statistical testing were not significant after correction for multiple comparisons via the false discovery rate correction. Thus, we take a density-based thresholding approach to identify the connections most relevant to this contrast between groups. To this end, we identify the 10% of connections with lowest p-values in the between-group contrast (10%=5,695 for HCP ROIs and 256 for DKT ROIs). It is with these connections that we proceed to the next step in the pipeline.

2. Correction for site of acquisition

Given that we have 3 different datasets - each one containing data obtained from multiple sites – in the next step we correct for the site of acquisition. The method we used to accomplish this is called ComBat Harmonization (Fortin et al., 2017; Fortin at al. 2018; Yu et al., 2018). ComBat is a regression based method which has been tested and validated on fMRI and MRI measures as an effective means of removing effects related to site of data acquisition. The method allows the user to provide a vector of site labels associated with each data point (this vector's trends are to be removed from the data), and vectors representing variables who's effects should not be removed from the data. The variables we wished to retain were group label (i.e., progressor or stable), age, and sex. The ComBat step is computed on the feature type selected in step 1 only (i.e., FC or GMV&CT).

3. Subtyping

After correcting for site of acquisition, we pursue subtyping of both the stables and the progressors separately: we considered 1-4 subtypes in both of these groups, as previous studies indicated this was the meaningful range for number of subtypes. The method we use for subtyping is called Similarity Network Fusion (SNF; Wang et al., 2014). SNF computes the similarity matrix (squared Euclidean distance) across all data sets for each data type and then integrates of these matrices into a single similarity matrix using a nonlinear combination method based on message-passing theory. This method utilizes spectral clustering techniques that make use of the eigenvalues of the similarity matrix to perform dimensionality reduction. Once we obtain the subtypes in each group, the original labels of "progressor" and "stable" are replaced with the subtype identifiers, and these become the labels associated with the training data in building the classifiers. Once again, the subtyping is computed on the feature type selected in step 1 only (i.e., GMV&CT *or* FC). We chose this method because in our GMV&CT analyses we wished to combine these two types of data into one overall structural-feature-based subtyping schema. We also wished to observe the overarching subtyping considering all 3 data types (GMV, CT, and FC). However, we discovered that the subtyping and subsequent feature selection given the new labels associated with the subtypes followed almost exclusively the trends associated with the structural features. Given this observation, we decided to keep our analyses with GMV&CT vs. FC completely separate, allowing FC features a chance to have a voice in the classification problem. See the results section "Subtypes Characterization" for a demonstration of this finding.

4. Feature selection

In order to maintain adequate predictive power in the classifier, and to minimize computational time, we next proceed to implement a feature selection algorithm, which allows us to further rein in the number of features. We retain 1-60 features in our grid search, and always add age and sex to the final set of features, as these have been shown to be correlated with the development of dementia (Riedel et al., 2016). The feature selection algorithm is called Multivariate Minimal Redundancy Maximal Relevance (MRMR, Yang et al., 2013). This particular algorithm was selected following work presented in Hojjati et al. (2017) where they compared the success of several feature selection algorithms in their classification of the same

populations: MCI stables and progressors. The MRMR algorithm was the most successful across their tests seeking to classify MCI subjects using various network measures derived from FC matrices. The MRMR algorithm uses mutual information to determine the redundancy of the information contained in any given feature, and seeks to maintain only highly informative, non-redundant features.

5. Classifier selection and training

Once the features are selected, we can now build our classifiers with the selected features and subtype labels. We use one of two possible ML approaches: Support Vector Machine (SVM) or Random Forest (RF). The implementations we use are Matlab's *fitcecoc* with *templateSVM*, and *fitcensemble* with the *'Bag'* method respectively. Since in our first tests of the classifiers, we noticed that only a few subjects were predicted into the progressor classes, we consider a weighting value that biases the classifier to put more importance on correctly classifying the progressor data over the stables. It has also been observed that, in the MCI population, which has an imbalance in the prevalence of progression vs. non-progression, classifiers tend to be biased to predict the most prevalent class, i.e., stables (Mathotaarachchi et al., 2017; Yen and Lee, 2009). Our use of progressor weighting is an approach to overcome this potential bias. Of course, we do also look at no weighting (weight = 1). Finally, after all of these selections and algorithms, we end up with a class prediction for the scans of the left out subject.

To recap, our grid search includes all combinations of the following selections and parameters:

Variable parameters:

{ Gray matter volume & Cortical thickness or FC (ICOV or correlation),
1, 2, 3, or 4 progressor subtypes,
1, 2, 3, or 4 stable subtypes,
Number of features [1-60],
Random Forest or Support Vector Machine,
Weight associated with progressor data [1, 5, 10, 15, 25, 35, 45] }

6. Outer loop

Now that we have done this large number of analyses in our grid search, we need to identify which of the analyses are successful and with which we want to move forward in the pipeline. As motivated in the introduction, our metric for a "successful" analysis is that of PPV. We compute the PPV for each class in each analysis. To arrive at the set of analyses we move forward with, we identify the analyses that have PPV of at least 94% for at least one class. Next we verify that the high PPV analysis remains with high PPV across a minimum of 3 consecutive "number of features" selections, to reduce the likelihood that the success of this analysis is due to chance. Consecutive analyses with high PPV are narrowed down to one single analysis with the highest PPV, or, if the PPV is equivalent for several within the range, the analysis with the median number of features. Of the high PPV analyses that remain after enforcement of this criterion, we first choose the analysis that classifies the most subjects correctly, and then we add additional analyses if they correctly classify additional subjects not yet accounted for. We also recompute the subtypes on the whole outer loop training set. This is necessary because the subtyping done in each of the inner loop iterations was done with the data from one subject left out (i.e., LOSO cross-validation). Next, we identify consensus features across the inner loop iterations. This means that we produce a histogram of the features selected in each of the LOSO iterations (recall that features are selected independently for each LOSO iteration and may differ somewhat between iterations) given the parameters of each selected analysis and identify the top X most commonly selected features, X being the number of features associated with that analysis. If a selected analysis has a poor consensus, meaning fewer than 60% of the consensus features are selected in over 60% of the LOSO iterations, that analysis is excluded. We can then retrain the classifier with the parameters of the selected analyses, the full set of subtype labels, and the consensus features and we end up getting a class prediction for each of the left out scans from the outer loop (the 20% outer loop test set).

Given that we are doing a 5-fold outer loop CV, the entire process from steps 1-6 is repeated 5 times with unique test sets comprised of 20% of the data. This means that the subtypes, features, and high-PPV analyses are optimized separately for each outer-fold, and thus we present them separately and comment on the trends observed between them, as well as

computing an "overall result" across the 5 folds, as explained in the next sub-section (Methods sub-section 7).

7. Final interpretation and summarization

Recall that our high PPV-classifiers optimize one class only: the high prediction confidence is only associated with test set classifications that fall within the optimized class of interest. Because of this design, our interpretation of outer-loop test subject class predictions is as foloows: if the class of interest is a progressor (stable) class, and the test subject is a progressor (stable), this is a true positive result; if the class of interest is a progressor (stable) class, and the test subject is a stable (progressor), this is a false positive result; all predictions of test subjects into classes other than the class of interest are indeterminate.

In order to summarize the results of our 5 outer-loop folds, each consisting of multiple high-PPV classifiers, we first compute the "majority voting" subject predictions across the high PPV analyses within a given outer-fold. This results in one single set of true positive (TP) and false positive (FP) values per outer fold. Finally, to characterize the overall success of the pipeline and to quantify the generalizability of the results, we compute the mean and standard deviation of the sensitivity, specificity, PPV, indeterminate rate, and FP rate across folds. See Results section "Pipeline Results".

Subtypes Characterization

In order to be able to better understand the subtypes we classify, we pursue an additional step outside of our ML pipeline, which strives to characterize the subtypes generated with the methodology employed within the pipeline. The first step to summarizing the subtypes in our sample of MCI subjects is to identify the optimal number of subtypes for stables and progressors. The optimal number of subtypes represented in this data was evaluated in several ways: 1) the frequency of the subtype numbers in *all selected high-PPV analyses* across all 5 outer loop folds; 2) the frequency of the subtype numbers in *only the selected stable-optimized high-PPV analyses* across all 5 outer loop folds (this choice was made given the finding that the stable-optimized analyses achieved high PPV in the outer loop: see Results sections below); 3) using eigen-gaps computed separately for FC, GMV&CT, and for all data types together; and 4) using rotation cost computed separately for FC, GMV&CT, and for all data types together. The eigen-gap is

defined as the magnitude of the difference between eigenvalues of the similarity matrix. The larger the eigen-gap is, the closer the eigenvectors of the ideal case and the perturbed case are, and hence the better the spectral clustering fits the data (von Luxburg, 2006). Thus, the number of subtypes that maximizes the eigen-gaps is optimal. The rotation cost is a cost function defined using a rotation of the eigenvector matrix, and the optimal number of subtypes is the value that minimizes this cost (Zelnik-Manor and Perona, 2004).

Once the number of subtypes was identified, we gathered all features selected across all selected high-PPV analyses from all 5 outer loop folds and computed the subtype labels once again with the SNF method. The subtypes are computed separately for the structural data and the functional data, as in our pipeline. With the subtype labels identified, we then proceeded to perform a statistical contrast between subtypes generated using each data type separately. We used a two tailed t-test and a significance level of 0.05 corrected for multiple comparison using the Bonferroni correction. In this t-test, each of the features is contrasted between pairs of subtypes.

Results

We will first present the overall results of the pipeline and then look at the results on a fold-by-fold basis for the 5 outer-loop folds. In delving deeper into the outer-loop folds, we will discuss the trends observed within our pipeline's high-PPV analyses, including the trends in each of the variable parameters. The parameters we will discuss include the selected number of features, feature type, classifier type, and the number of progressor and stable subtypes. Finally, we will proceed to summarizing the findings with regards to the subtypes identified. Throughout, we will present our findings for both the finer scaled parcellation (HCP) and the larger scaled parcellation (DKT) which will allow us to observe the differences in their performances.

Pipeline Results

The final subject classification results from each of our 5 outer-loop pipeline iterations are presented in Figure 2. We present both the results with the HCP ROIs (plots in column 1) and the DKT ROIs (plots in column 2). Each row of a scatterplot shows the classification results of a single high-PPV analysis. Each point on the scatterplot identifies a subject who was

classified into that specific analysis' class of interest. Areas on the plots with a white background contain true results, and areas with a gray background contain false results. Above the horizontal line, we present the analyses which were optimized for a progressor class, and below the horizontal line, are those that were optimized for a stable class. One can see at a glance that the stable-optimized analyses all have very few false results and often have a fairly large number of true results (i.e., have high PPV), whereas the progressor-optimized analyses often have very many false results, despite a large number of true results (i.e., low PPV). This result is true for both the HCP and DKT ROI-based analyses.

To quantify the classification results in Figure 2, we present various metrics associated with the stable-optimized HCP and DKT ROI-based analyses in Table 3. We focus much of the upcoming results summary and discussion on the stable-optimized analyses only, as the pipeline was only successful in that context. When considering only the stable-optimized analyses, if a stable subject is correctly classified into a stable class of interest, this is considered a TP result. A FP result is when a progressor subject is classified into a stable class of interest. We consider any subject not classified into the class of interest as an indeterminate result, and therefore by this construct, there are no "negative" results per se. As a consequence of this, we only present TP and FP results (and no TN and FN). We compute the Sensitivity as the TP rate (TPR) and the Specificity as 1-[FP rate] (FPR). We present both the sample PPV as well as a PPV corrected (PPVc) given the prevalence of non-progression (i.e., stability) from MCI to dementia (Tam et al., 2019).

$$\text{PPVc} = \frac{\text{Sensitivity} * \text{Prevalence}}{(\text{Sensitivity} * \text{Prevalence}) + (1 - \text{Specificity}) * (1 - \text{Prevalence})} \quad (1)$$

We estimate the prevalence of progression of MCI subjects in the population to be 33.6% (Mitchell et al., 2009; Tam et al., 2019) given our average time of follow up of approximately 3.5 years.

We find that with the HCP ROIs, there are on average (across the 5 outer-loop folds) 18.6 TP subject classifications (±standard deviation, STD, of 6.6), and only 1.8±1.5 FP subject classifications. This results in a sample PPV of 92.5% (±5.1%), and a PPVc of 96.1% (±2.6%). Mean specificity is high at 94.8% (±4.2%), though this is at the cost of the sensitivity which is only at 60.4% (±14%). The indeterminate rate is high at 68.7% (±9.7%). However, we believe

that the successful identification of even a subset of subjects with high PPV is worthwhile given the high confidence we can have in our prognostic prediction for those individuals. Clinical decisions around treatment, qualification for clinical trials, and future planning can be made accordingly if the subject and clinician know with a high likelihood that the subject will not develop dementia. For the DKT ROIs, there are on average 9±5.8 TPs, 1.6±1.8 FPs, a sample PPV of 88.2% (±10.9%), and a PPVc of 93.5% (±6.1%). Mean specificity is still high at 95.3% (±5.3%), sensitivity is very low at 27.8% (±16%), and the indeterminate rate is 84.1% (±10.4%).

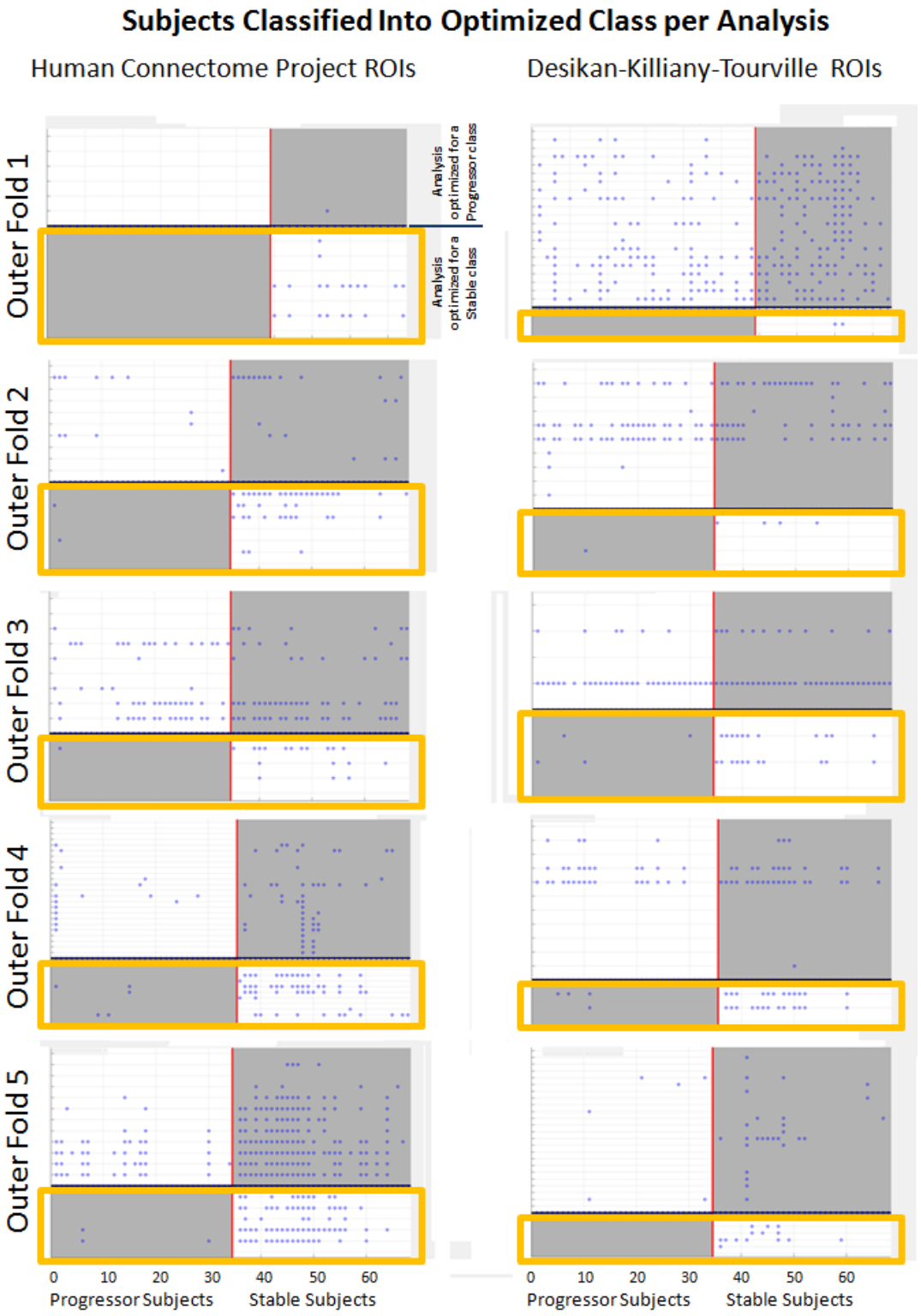


Figure 2. Test subjects who were classified into the class which was optimized-for in each selected high-PPV analysis in the 5 outer-loop pipeline iterations. Subjects are presented on the x-axis with progressors appearing to the left of the red line and stable subjects appearing to the

right. In each graph, analyses above the horizontal black line are those which optimized a progressor class, and analyses below the black line optimized a stable class. Given this, sections with white background within each graph represent "True" predictions (i.e., Progressor/Stable subject classified into a Progressor/Stable class), and sections with a gray background represent "False" predictions. Yellow boxes are highlighting that analyses optimized for stable subtypes show high positive predictive value. Rows of graphs represent the 5 outer-loop folds and the two columns of graphs represent the two different parcellations tested: the Human Connectome Project Cortical Parcellation (HCP), and the Desikan-Killiany-Tourville Atlas (DKT). ROIs: regions of interest.

Outer Fold	Test Group Stables	Test Group Prog.	TP	FP	SENS-ITIVITY	SPEC-IFICITY	IND. RATE	FP RATE	PPV	PPV corrected
HCP										
1	18	29	10	0	0.556	1.000	0.787	0.000	1.000	1.000
2	34	34	22	2	0.647	0.941	0.647	0.059	0.917	0.956
3	34	34	13	1	0.382	0.971	0.794	0.029	0.929	0.963
4	33	35	24	4	0.727	0.886	0.588	0.114	0.857	0.926
5	34	34	24	2	0.706	0.941	0.618	0.059	0.923	0.960
Mean	30.6	33.2	18.600	1.800	0.604	0.948	0.687	0.052	0.925	0.961
Std			6.618	1.483	0.140	0.042	0.097	0.042	0.051	0.026
DKT										
1	18	29	2	0	0.111	1.000	0.957	0.000	1.000	1.000
2	34	34	4	1	0.118	0.971	0.926	0.029	0.800	0.888
3	34	34	14	4	0.412	0.882	0.735	0.118	0.778	0.874
4	33	35	15	3	0.455	0.914	0.735	0.086	0.833	0.913
5	34	34	10	0	0.294	1.000	0.853	0.000	1.000	1.000
Mean	30.6	33.2	9.000	1.600	0.278	0.953	0.841	0.047	0.882	0.935
Std			5.831	1.817	0.160	0.053	0.104	0.053	0.109	0.061

Table 3. Metrics summarizing the classification results associated with the stable-optimized HCP and DKT ROI-based analyses. These metrics reflect the results presented in Figure 1. A stable subject that is correctly classified into a stable class of interest is considered a true positive (TP) result. A false positive (FP) result is a progressor subject classified into a stable class of interest. We consider any subject not classified into the class of interest as an indeterminate (IND) result, and therefore by this construct, there are no "negative" results, (i.e., true or false negative predictions). We compute the Sensitivity as the TP rate, and the Specificity as 1-[FP rate]. We present both the sample PPV and a PPV corrected given the prevalence of non-progression (i.e., stability) from MCI to dementia. The prevalence of progression is estimated at 33.6%, thus the prevalence of non-progression would be at 66.4%. The metrics associated with each of the outer-loop folds are presented in the rows, as well as the mean and standard deviation across these values, summarizing the overall metrics for the pipeline. Metrics for the pipeline iterations using HCP ROIs are presented in the top half of the table, and those for the iterations using the DKT ROIs are presented in the bottom half.

Despite the fact that the progressor-optimized analyses had poor results, we present a similar table to Table 3 in the supplementary material for the interest of the reader, this time presenting both the progressor-optimized results and the stable-optimized results together (Supplementary Table 1). In Supplementary Table 1, a progressor result is considered a "positive" result, and a stable result is considered a "negative" result. Thus we present TP and FP, as well as TN and FN. The PPV reflects the confidence of a progressor prediction, and the NPV reflects the confidence of a stable prediction. The NPV in Supplementary Table 1 is different from the PPV presented above in Table 3 because the overall majority voting class-prediction result across only stable-optimized analyses may be different than the class-prediction across both progressor-and-stable-optimized analyses. The NPVc for the HCP and DKT ROI-based analyses are 95.8% (±8.4%) and 89.6% (±14.7%), respectively. The PPVc is only 48.8% (±16.4%) for the HCP ROI-based analyses and 38.3% (±6.1%) for the DKT ROI-based analyses.

A Closer Look at the Pipeline-Selected High-PPV Analyses

Now that we have seen the overall classification results from the pipeline, to understand better how these results were achieved, we will examine the parameters associated with each of the high-PPV analyses selected by the pipeline. The parameters from the high-PPV analyses from each outer-loop fold are presented in Supplementary Tables 2 (HCP ROIs) and 3 (DKT ROIs). Within these tables, we present each of the parameters associated with the selected analysis (i.e., number of features, classifier type, type of feature used, number of progressor and stable subtypes, and progressor weighting). In addition, we present some of the key results associated with that classifier (i.e., the subtype that was classified with a high PPV, the value of that high PPV, the number of subjects correctly classified into that class, and the extent over which that classifier was stable in its high PPV). We also present the mean of the PPV, number of features, subjects correctly classified, and extent of high PPV, as well as the mode of classifier type, feature type, number of subtypes, and weight. To quantify the confidence of these means/modes, we compute the standard deviation and the occurrence rate, respectively. Finally, we also present these mean/mode results both across *all of the high-PPV selected analyses* and over only the *stable-optimized high-PPV analyses* (since only stable-optimized analyses were successful in testing on outer loop validation data sets). These fold-by-fold results are presented for the interest of the reader. Here, we will focus our summary on the mean/mode results across

pooled high-PPV analyses from all 5 outer-loop folds. These pooled results are presented in Table 4. In the summary of the results presented below, we will focus only on the stable-optimized high-PPV analyses.

In both the HCP and DKT ROI-based high-PPV analyses, the mean of the PPV values resulting from the inner-loop of the pipeline is approximately 98% for the class of interest. It is unsurprising that we have a PPV of such high magnitude as we select only analyses with a PPV of 94% or greater. Because of this imposed elevated PPV, we do not present the standard deviation on this value, as its range is artificially created to be from 94-100%. The mean number of features selected for the stable-optimized high-PPV analyses is 31 for both HCP (±14) and DKT (±12) ROIs. For both parcellations, there were, on average, 11 subjects (±9, ±14 respectively) classified correctly by a single high-PPV analysis. The high PPV for these analyses was stable for on average 26±16 successive numbers of features for the HCP ROI-based analyses and for 15.5 ±1.5 successive numbers of features for the DKT ROI-based analyses. In all cases, RF classifiers were most numerous among the selected high-PPV classifiers. RF was used in 58.1% and 81.3% of selected analyses when HCP ROIs and DKT ROIs were used, respectively. FC feature-based classifiers (as opposed to GMV&CT feature-based classifiers) were most numerous among the high-PPV analyses, being utilized in 74.2% and 87.5% of classifiers for HCP and DKT ROI-based analyses, respectively. In all cases across all selected high-PPV analyses from all outer-loop folds using both HCP and DKT ROIs, if FC-based features were used, the FC measure was correlation. Classifiers using ICOV-based FC features were never among the high PPV analyses. Some comments on why we think this is the case will be presented in the discussion. The mode progressor weightings selected were low for both the HCP and DKT ROI-based analyses, being 1 (a weighting value of 1 means no weight, 39% occurrence rate) and 5 (lowest weight tested, 25% occurrence rate), respectively.

The final characteristic of the high-PPV analyses from Table 4 that we will discuss is the number of progressor and stable subtypes used in the analyses. The most commonly used number of stable subtypes was 3 for both HCP and DKT ROI-based analyses, being selected 54.8% and 50.0% of the time, respectively. The second most common number of stables subtypes was two being selected 35.5% and 43.8% of the time in the HCP and DKT ROI-based analyses, respectively. For the progressors, the most commonly used number of subtypes was in

fact one, being selected 41.9% of the time for the HCP ROI-based analyses and 43.8% of the time for DKT ROI-based analyses. The second most common number of progressor subtypes in the HCP ROI-based analyses was three (29.0% occurrence). The second most common number of progressor subtypes in the DKT ROI-based analyses was four (31.3% occurrence).

Subtypes Characterization

Number of subtypes

The premise of our study is that a subset of MCI individuals can be classified with high confidence, and our results show that indeed, a subset of stable-MCI subjects can be classified with high PPV. Since we have now identified which branch of the pipeline is successful, i.e., stable-subtype-optimized analyses, we now seek to characterize those stable subtypes and the ways in which they differ from the represented progressor subtypes. In order to have an "overall characterization", we need to identify the optimal numbers of subtypes for both progressors and stables. There are several approaches we considered for identifying these optimal numbers, as presented in the methods section. Below, we present the results of these approaches.

The first approach we considered for establishing the optimal number of subtypes was identifying the most frequently occurring number of subtypes across *all high-PPV analyses*. The results of this optimization were 4 progressor subtypes and 3 stable subtypes (Table 4) for both HCP and DKT ROIs. The second most frequently occurring number of subtypes was 3 for progressors and 2 for stables, also for both parcellations. Despite the agreement in the results between HCP and DKT ROI-types, the poor performance of the progressor-optimized high-PPV analyses utilized to generate these optimal values led us to question the reliability of the results.

Our second approach to establishing the optimal number of subtypes was by identifying the most frequently occurring number of subtypes in only the *stable-optimized high-PPV analyses*. The results of this optimization were one progressor subtype and three stable subtypes (Table 4). However, we know from the literature and the well-established heterogeneity of MCI and dementia that there are more than one progressor subtypes. Because of this, we rejected this approach of identifying the optimal number of subtypes.

	PPV	Num Features	Subjects Right	Extent of High PPV	Classifier Type	Feature Type	FC Measure	Prog Classes	Prog Classes	Stable Classes	Stable Classes	Weight
All HCP Outer Loops												
mean/mode all analyses	mean: 0.984	25.276	12.724	15.678	mode: RF	FC	corr	4	2nd: 3	3	2nd: 2	1
std/occurrence all analyses		std: 15.126	14.437	15.235	occurrence 0.678	0.598	1.000	0.483	0.345	0.471	0.299	0.276
mean/mode stable analyses	mean: 0.981	31.484	11.161	26.065	mode: RF	FC	corr	1	2nd: 3	3	2nd: 2	1
std/occurrence stable analyses		std: 14.069	9.321	15.966	occurrence 0.581	0.742	1.000	0.419	0.290	0.548	0.355	0.387
All DKT Outer Loops												
mean/mode all analyses	mean: 0.986	36.258	12.015	9.606	mode: RF	GMCT	corr	4	2nd: 3	3	2nd: 2	35
std/occurrence all analyses		std: 15.062	12.020	10.275	occurrence 0.803	0.667	1.000	0.470	0.424	0.379	0.364	0.197
mean/mode stable analyses	mean: 0.981	30.813	10.875	15.500	mode: RF	FC	corr	1	2nd: 4	3	2nd: 2	5
std/occurrence stable analyses		std: 11.968	14.226	14.528	occurrence 0.813	0.875	1.000	0.438	0.313	0.500	0.438	0.250

Table 4. Mean/Mode of parameters selected across all high-PPV analyses from the 5 outer-loop folds and computed results. The mean is presented for results: PPV, subjects correctly classified (Subjects Right), and the extent of the stability of the high-PPV score across successive numbers of features (Extent of High PPV). In addition, we present the mean of the number of features (Num Features). We report the standard deviation (std) to quantify the confidence on the mean measures. The std is not presented for the PPV as these values are selected to be within the range of 0.94-1. The mode is presented for parameters: classifier type (either Random Forest – RF, or Support Vector Machine), feature type (FC or GMCT), FC measure (corr or ICOV), number of progressor subtypes ("Prog classes": 1, 2, 3, or 4), number of stable subtypes ("Stable classes": 1, 2, 3, or 4), and weight associated with the progressor training data (1, 5, 10, 15, 25, 35, or 45). We also present the second most common numbers of progressor and stable subtypes ("2nd"). The confidence in these mode values is quantified via its occurrence rate ("occurrence"). All values are presented across all high-PPV analyses (rows with white background), and across only stable-optimized high-PPV analyses (shaded rows). The values obtained with the HCP ROIs are presented in the top half of the table, and those associated with the DKT ROI-based analyses are in the bottom half.

Next, we pursued a computational approach to optimizing the number of subtypes. We looked at two separate measures that seek to optimize the number of clusters in a cluster analysis such as the SNF method we have used. The SNF package is equipped to compute both the eigen-gaps and the rotation cost metrics, which we can use to optimize the number of subtypes. In all iterations, both the eigen-gaps and rotation cost were optimal for the same number of subtypes; therefore, we only quote one set of results. We computed these metrics separately for subtyping generated using FC features, GMV&CT features, and then all feature types together on the training data from each of the outer-loop folds. The results of the optimizations are

presented in Supplementary Table 4. Overall, for both HCP and DKT ROIs, the optimal number of subtypes was 3 for progressors and 3 for stables with GMV&CT features, three for progressors and two for stables with FC (i.e., correlation) features, and 3 for progressors and 2 for stables with the combined feature set. We also computed a weighted sum over the most often identified best fitting (1st ranked) number of subtypes and the most often identified 2nd best fitting (2nd ranked) number of subtypes across all five outer-loop training data sets (Supplementary Table 4, last 4 columns; sum computed as [2*rank1+1*rank2]). The weighted sum identified three progressor subtypes and two stable subtypes as optimal. Three Stable subtypes also had a relatively high weighted sum, with only a slightly lower score than two subtypes.

Finally, we computed the eigen-gaps and rotation cost once again using the full set of data and found again that three progressor and two stable subtypes were optimal (Figures 3 and 4). Considering all of these results, the most commonly identified optimal number of subtypes were three progressor and two stable subtypes. Thus, as we proceed to further characterize the MCI subtypes, these are the numbers we will use.

Subtypes Associated With Different Feature-Types

The affinity matrices representing the SNF generated subtypes computed using the HCP ROIs on each of the data types (GMV&CT, FC, and all combined) are presented in Figure 3. Likewise, the subtypes generated using the DKT ROIs are presented in Figure 4. At a glance, despite the order of subjects not being consistent between matrices (by design, affinity matrices rearrange the subjects into clusters), it is clear that the GMV&CT features generate very different subtypes than the FC features for both the HCP and DKT ROIs. This is evident by the size of the subtypes. What is also evident is that the subtypes generated using all data types highly resembles those generated using GMV&CT. The algorithm that generates these subtypes works by combining the individual cluster patterns found using each feature type, putting a greater weight on the pattern that is more clearly represented in the data. Thus, the fact that the combined feature set results in a subtyping pattern more similar to that of the structural features implies that the structural features allow for more clearly defined subtypes. This is also reflected by the smaller values of affinity presented in the FC affinity matrices as compared to the GMV&CT affinity matrices. Because the GMV&CT subtype pattern seems to be overpowering

the FC subtype pattern when all features are used together, and because in preliminary analyses with all feature types together the MRMR feature selection algorithm selected only GMV&CT features (and no FC features), we present the characteristics of the subtypes separately for structural measures and for functional measures. This approach allows the functional measures to retain a "voice" in the characterization of the subtypes.

Subtype Demographics

The allocation of the subjects into the 5 subtypes (3 progressor and 2 stable) is presented in Figure 5. Recall that the progressor and stable subtypes are generated separately so there is never a progressor allocated to a stable subtype or a stable allocated to a progressor subtype. What we can see from Figure 5 is that subjects from all 3 data sets (ADNI, Han Lab, and OASIS3) are allocated across the different subtypes: i.e., *we do not see* subtyping simply based on dataset. This is encouraging with respect to the success of our correction for the site of data acquisition. This figure also illustrates well how very different the subtyping is between structural and functional features, with the distributions being highly different from one another.

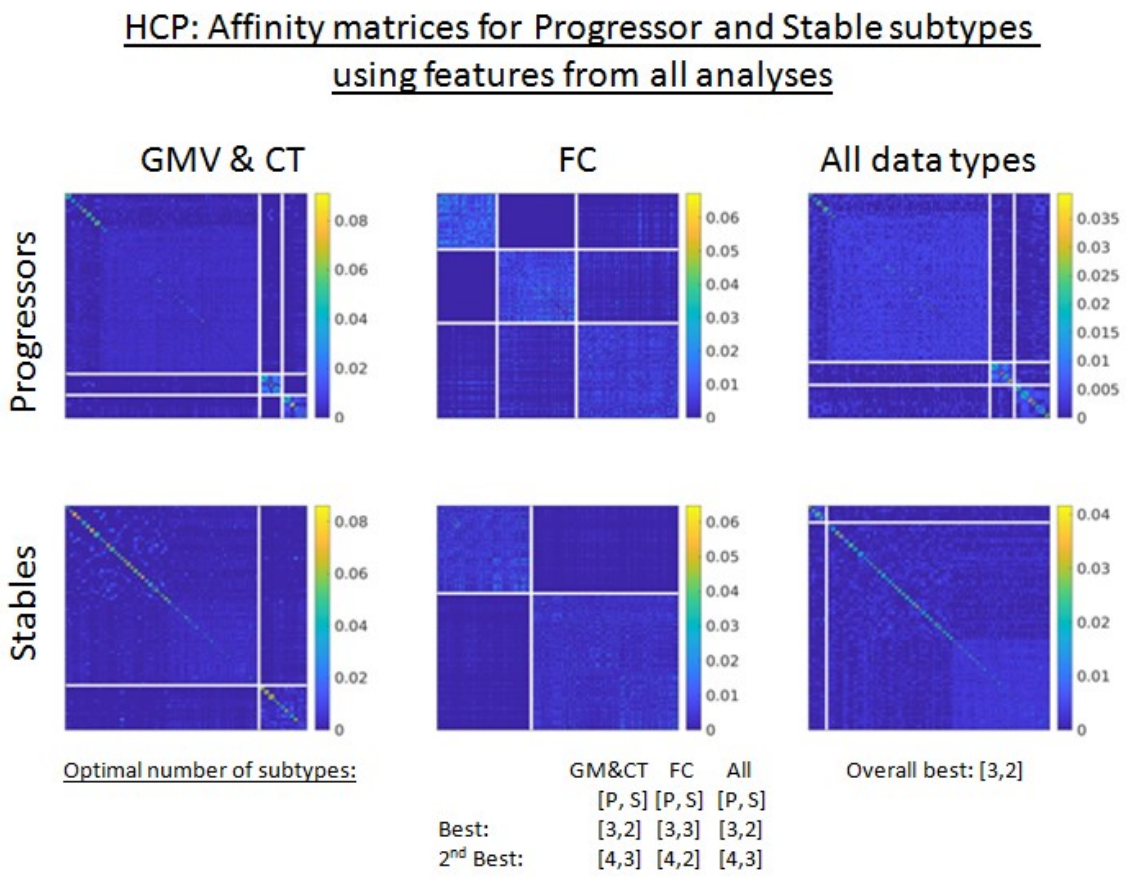


Figure 3. Affinity matrices showing the separation of the progressor and stable subjects into subtypes using different sets of HCP ROI-based features. The optimal number of progressor subtypes (3) are shown on the top row and optimal number of stable subtypes (2) are shown on the bottom row. The first column of matrices shows the subtypes generated using GMV&CT features, the second column shows the subtypes generated using FC features, and the third column shows the subtypes generated using all of the feature types together. Subtypes are generated using the Similarity Network Fusion (SNF) method. The optimal number of subtypes was computed using the eigen-gaps and rotation cost metrics. Both methods always found the same optimal values. They are presented below the matrices.

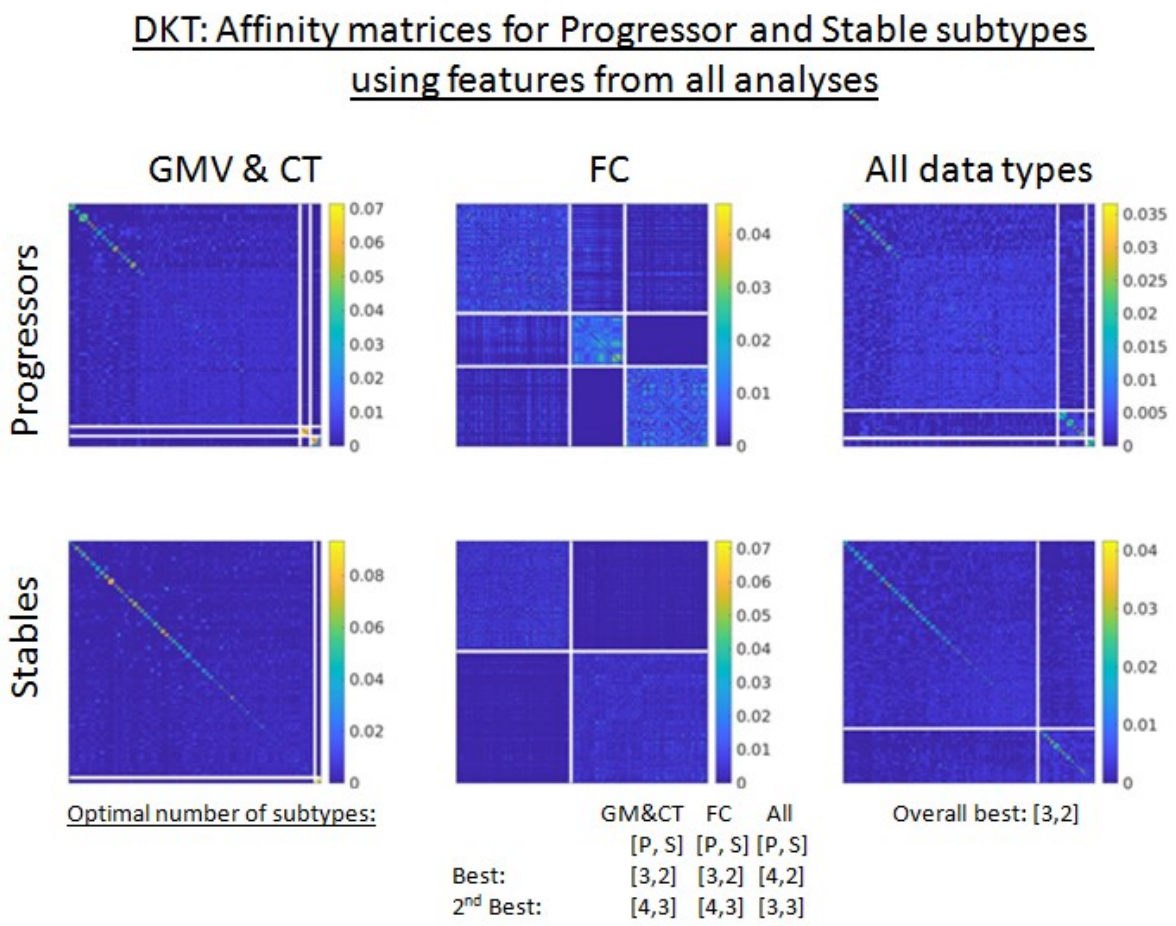


Figure 4. Affinity matrices showing the separation of the progressor and stable subjects into subtypes using different sets of DKT ROI-based features. The presentation is as in Figure 3.

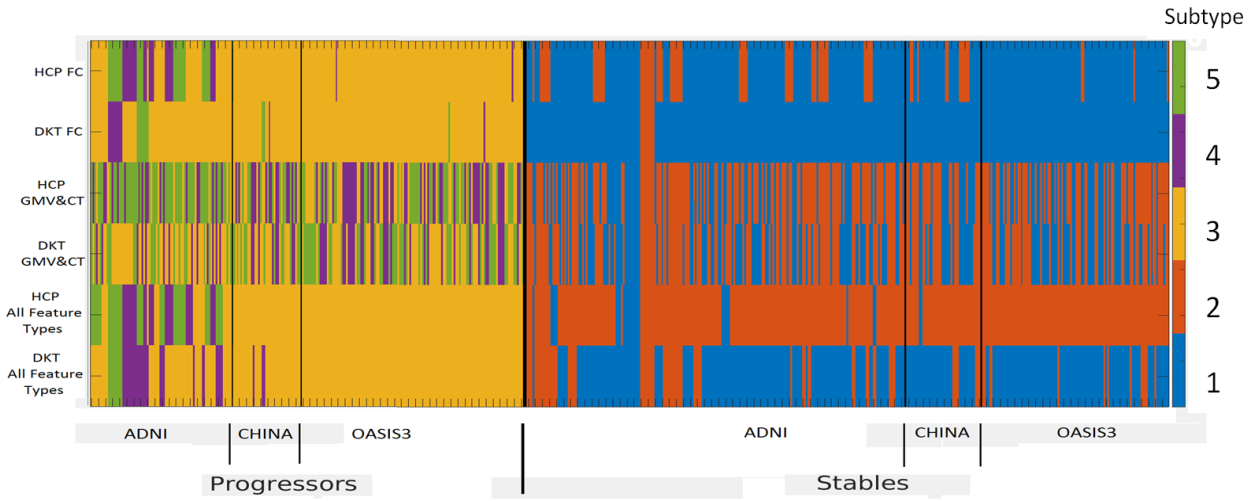


Figure 5. Allocation of subjects to 3 progressor and 2 stable subtypes. Progressor subjects appear to the left along the horizontal axis followed by the stable subjects to the right. Within the sets of progressor and stable subjects, the ADNI data appears first, followed by the Han Lab data, then the OASIS3 data. These datasets are all separated by a vertical black line. Rows within the graph present the allocation of subjects into subtypes generated by the different feature types: row 1- HCP ROI-based FC features, row 2- DKT ROI-based FC features, etc. (as labeled along the vertical axis).

Stable Subtype	**Data Type**	**Number Of Subjects**	**Sex** (# Males)	(%)	**Mean Age** (yrs ± std)	**Median Age** (yrs)	**Follow Up Time (months ± std)**	**Sites Represented**
HCP								
1	GMCT	124	66	53.23%	74.9±8.8	75.7	59.1±48	12/12
	FC	88	49	55.68%	73.1±8.5	74.0	50.7±36	11/12
2	GMTC	21	13	61.9%	67.1±6.6	67.0	35.8±18	9/12
	FC	107	57	53.27%	73.1±9.0	73.7	55.6±48	12/12
DKT								
1	GMCT	143	78	54.55%	73.9±8.9	74.1	56.0±46	12/12
	FC	96	56	58.33%	73.0±8.6	73.9	50.6±38	12/12
2	GMCT	1	1	100%	70.6±0	70.6	26	1/12
	FC	99	50	50.51%	73.3±8.9	74.0	54.9±48	12/12

Table 5. Demographics of subjects within the stable subtypes. yrs= years. std= standard deviation.

Beyond looking at the dataset of origin of the subjects, we also looked at a number of other demographics across the stable subtypes to evaluate whether these could account for the subtypes defined. These demographics are presented in Table 5. Largely, there are no evident demographic differences observable between groups for both the subtypes generated using GMV&CT features and FC features for both HCP and DKT ROIs. Most stable subtypes have from 50-58% males, a mean and median age from 73-76, 50-59 months of follow up establishing they have remained stable, and 11-12 of 12 sites of data acquisition accounted for across their subjects. There are 2 subtypes that are somewhat exceptional, both of them for GMV&CT-subtype 2: the first using HCP ROIs, and the second using DKT ROIs. For the HCP ROI GMV&CT-subtype 2, the percentage of males is slightly higher at 62%, and the age is slightly younger at a mean/median of 67. There are also only 9 of 12 sites accounted for and the mean follow-up is shorter at only 35.8 months. Of note, this subtype has only 21 subjects: it is smaller than the other subtypes. This smaller subject pool could account for the slightly deviated demographic trends. It is also plausible that this younger group with less time of follow up is at an earlier stage of their cognitive decline (whether their end-points are stable MCI or dementia). For the DKT ROI GMV&CT-subtype 2, there is only one subject that has been allocated to that subtype. Recall that each subject has multiple scans, and thus this subtype is in fact capturing all scans from this one subject. We can conclude from this subtype allocation that there are relatively minimal between-subject similarities in GMV&CT features when DKT ROIs are used. To test if DKT ROI GMV&CT feature-based subtypes could capture more than one subject, we also looked at generating 3 stable subtypes (the second most optimal of subtypes identified), and found once again that individual subjects were being separated out, rather than groups of subjects (i.e., there were two subtypes with scans from only one subject each, and all other subjects were in the third subtype). Overall, there are no clear trends identifiable in the subject demographics in terms of the definition of the subtypes.

Statistical Differences Between Subtype Features

Now that we have identified the optimal number of subtypes and have not been able to explain the definition of these subtypes by demographics alone, we look to see the differences in mean structural and functional measures between groups. Figures 6 and 7 show the features which are significantly different between subtypes among the set of features selected across all

high-PPV analyses for the HCP ROI-based analyses and the DKT ROI-based analyses, respectively. There were 252 FC features and 163 structural features tested in the contrasts with HCP ROIs (tested separately, see Figure 6). Likewise, there were 196 FC features and 111 structural features in the contrasts with DKT ROIs. Recall that there are 56,953 possible HCP ROI-based FC features that could be selected in the pipeline by considering all pairing combinations of 338 brain regions. Similarly, there are 736 structural features possible including the GMV of all 360 HCP cortical regions plus the 16 subcortical regions, as well as the CT from the 360 HCP cortical regions. This reminder is to illustrate that our set of features used for testing group differences is quite pared-down, taking advantage of numerous iterations of feature selection to identify only the most relevant features for this testing. Testing only these pared-down feature sets will minimize the effects of multiple comparisons.

We can see in the upper portion of Figure 6 that HCP ROI FC-based stable subtypes 1 and 2 differ from each other in nearly all considered features, whereas subtypes 1 and 3 (subtypes 3, 4 and 5 are the progressor subtypes), and subtypes 2 and 4 are quite similar to one another with some minor differences. Subtype 5 appears to be an intermediate group of sorts: it is less different than 1 and 2 are to one another but certainly has a high number of statistically significant differences in feature magnitudes. The trends in HCP ROI structural feature-based subtype differences apparent in the lower portion of Figure 6 differ from those observed in the FC-based subtypes. This is not surprising based on our observation that the subtypes defined using these different feature types are quite different. The subtypes identified using structural features in general show many differences between one another. That being said, subtype 2 overall shows more differences from the other subtypes, and subtype 1 and 4 are quite similar.

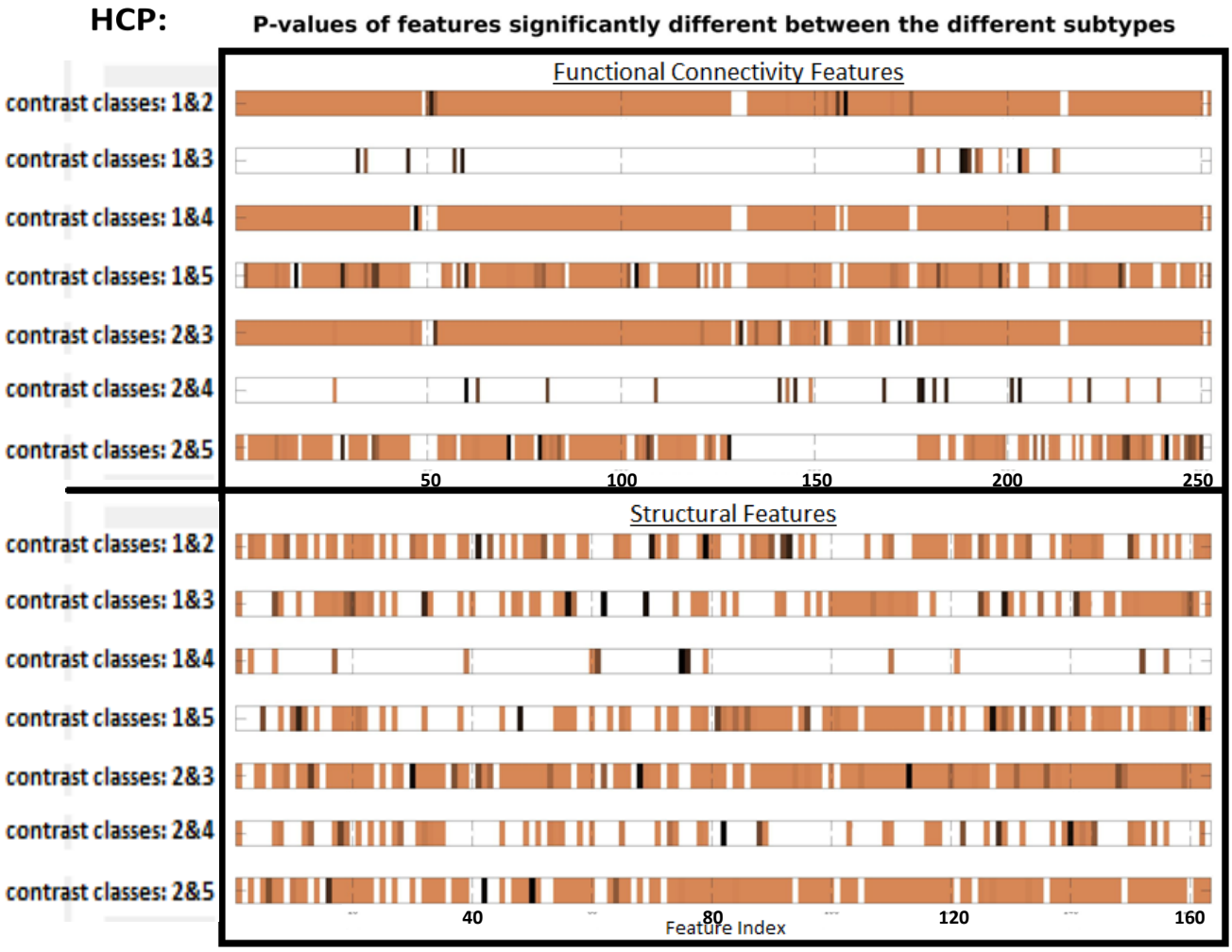


Figure 6. HCP ROI-based features showing statistical differences between subtypes of MCI individuals. Significantly different features between subtypes are presented as colored vertical bars, with smaller p-values being darker in color. White color represents a feature that is not significantly different between groups. The p-value threshold for significance is set such that the false discovery rate is 0.05. Class (i.e., subtypes) combinations are presented in separate rows. We show only the contrasts between stable subtypes (classes 1 and 2) and between stable and progressor subtypes (classes 3, 4 and 5), as only the stable subtypes were classified with high PPV. The results of the statistical contrasts for subtypes generated using functional connectivity features are presented in the top half of the figure and the results for subtypes generated using structural features are on the bottom half. The detailed results of these statistical tests can be found in Supplementary Tables 5-7.

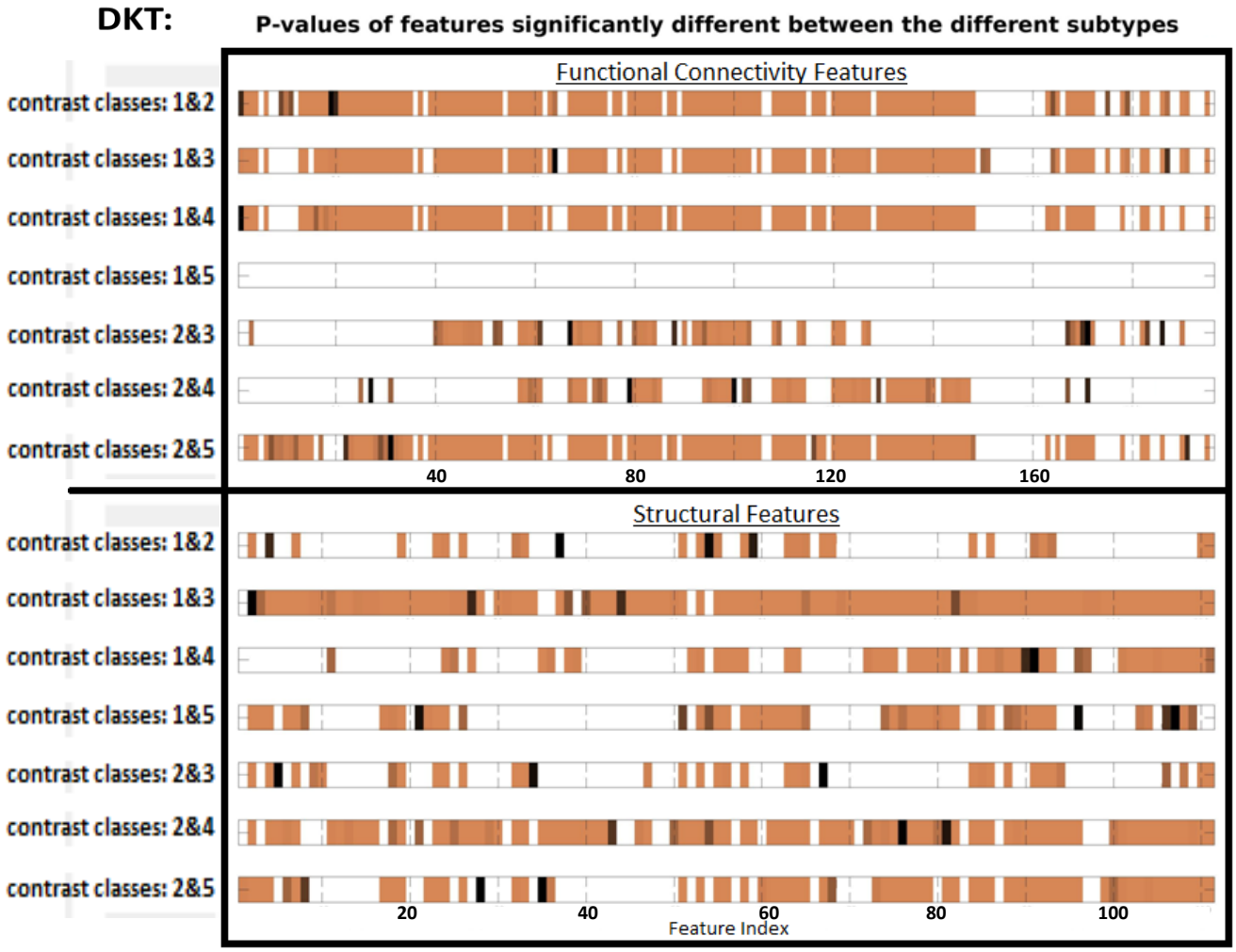


Figure 7. DKT ROI-based features showing statistical differences between subtypes of MCI individuals. The presentation is as in Figure 6. The detailed results of these statistical tests can be found in Supplementary Tables 8-10.

Examining next the upper portion of Figure 7 for the statistically significant differences between FC-based features using the DKT ROIs, we see that the features which differ between subtypes 1 and 2, 1 and 3, and 1 and 4 are very similar. This would imply that stable subtype 2 is more similar to progressor subtypes 3 and 4 than is stable subtype 1, which is consistent with the reduced number of significantly different features between subtypes 2 and 3, and 2 and 4. Stable subtype 1 and progressor subtype 5 show no statistically significant differences, whereas stable subtype 2 and progressor subtype 5 show a very similar distribution of differences to that between subtypes 1 and 2, which is consistent with minimal differences between subtypes 1 and 5. When we look at the differences in the subtypes generated using DKT ROI-based structural

features in the lower portion of Figure 7, we can see that stable subtype 1 and progressor subtype 3 are very different, with nearly all tested features showing statistically significant differences, and similarly, stable subtype 2 and progressor subtype 4 are very different. Conversely, subtypes 1 and 4 show fewer differences from one another, as do subtypes 2 and 3. Both of the stable subtypes, 1 and 2, also show relatively few differences. This set of distributions of differences leads to the interpretation that progressor subtypes 3 and 4 are likely more "extreme" versions of subtypes 2 and 1 respectively (greater changes in feature magnitudes of a similar distribution), resulting in the observed similarity with one of the stable subtypes and increased level of difference with the other. It is important to note that the subtypes generated with structural features using the DKT ROIs were poorly generalizable. As seen in Table 5, stable subtype 2 only contains 1 subject, and similarly, as illustrated in Figure 4, progressor subtypes 2 and 3 are small, each containing only 3 subjects. Thus, any results presented with regards to differences between DKT ROI-based structural subtypes should be interpreted with caution.

The full list of significantly different features between pairs of subtypes is presented in Supplementary Tables 5-7 for the HCP ROIs and Supplementary Tables 8-10 for the DKT ROIs.

Brain Regions Implicated in Statistically Significant Differences Between Subtypes

Thus far the only comments we have made on the differences between subtypes is the proportion or distribution of features different between them. Figure 6 is helpful to observe these trends, but it does not facilitate understanding which specific brain regions are implicated in the differences given that there are often up to 200+ significantly different features, which is too many to legibly identify on the graph. The detailed results presented in Supplementary Figures 5-10 are helpful, but the volume of information is too large to easily grasp the big picture. To facilitate the illustration of which brain regions have been implicated in the statistically significant differences, we present them on the surface of the brain in Figure 8. We have also plotted the frequency with which a given brain region was implicated in a statistical difference between groups in Figures 9 and 10.

Recall that our FC measures quantify the correlation between pairs of brain regions. Thus, when we count the number of occasions a brain region is implicated in a set of FC-based group differences, each of the two brain regions in each pair being correlated is counted. It

follows that despite the fact that there were 252 FC features in the HCP ROI contrasts and 196 in the DKT ROI contrasts, there are in fact only 82 brain regions (of 376) implicated in the HCP FC feature differences and 45 (of 78) implicated in the DKT FC feature differences (due to the same brain regions being implicated in several FC features). The most frequently implicated HCP and DKT ROIs in FC subtype differences are shown in the upper portions of Figures 9 and 10, respectively. Those regions which were implicated in more than ten statistically significant FC feature differences between subtypes are labeled along the horizontal axis (with reduced labeling for legibility). For the HCP ROIs (ROI naming convention is as in Glasser et al., 2016), these are areas 1 (primary somatosensory complex anterior), 2 (primary somatosensory complex posterior), 3a (primary somatosensory cortex - fundus of central sulcus), 3b (primary somatosensory cortex - posterior bank of central sulcus), 4 (primary motor cortex), IP0 (intra-parietal region 0), PH (basal parietal region most likely belonging to the visual cortex), and V2 (2^{nd} visual area) from the left hemisphere (LH), and from the right hemisphere (RH), areas 1, 2, 3a, 3b, 4, 5m (Brodmann's area 5 medial), 5mv (area 5m ventral), V2, V3, V6 (visual areas), and VMV1 (ventromedial visual area 1). For the DKT ROIs (ROI naming convention is as in Klein and Tourville, 2012), these are areas postcentral (primary somatosensory cortex), precentral (primary motor cortex), paracentral (medial sensorimotor area), posterior cingulate, precuneus (medial parietal), pericalcarine (early visual cortex), inferior parietal, parahippocampal, and pars triangularis from the LH, and from the RH, areas postcentral, precentral, posterior cingulate, pericalcarine, parahippocampal, pars triangularis, and superior temporal. The regions implicated agree in general between these two parcellations with a strong presence of differences in the primary somatosensory and motor cortices, the visual cortex, and medial regions of the parietal lobe. There is also a strong agreement between hemispheres. The DKT parcellation also shows a fair presence of frontal regions that are significantly different between groups (Figure 8), though these were implicated fewer than ten times (Figure 10). As the subcortical regions cannot be plotted on the surface in Figure 8, we will emphasize here that for the HCP parcellation's FC measures, there were no significant differences between groups implicating subcortical regions. For the DKT parcellation, the subcortical regions implicated were the caudate from both LH and RH, hippocampus from the RH, and the accumbens form the LH (though these were all implicated fewer than ten times).

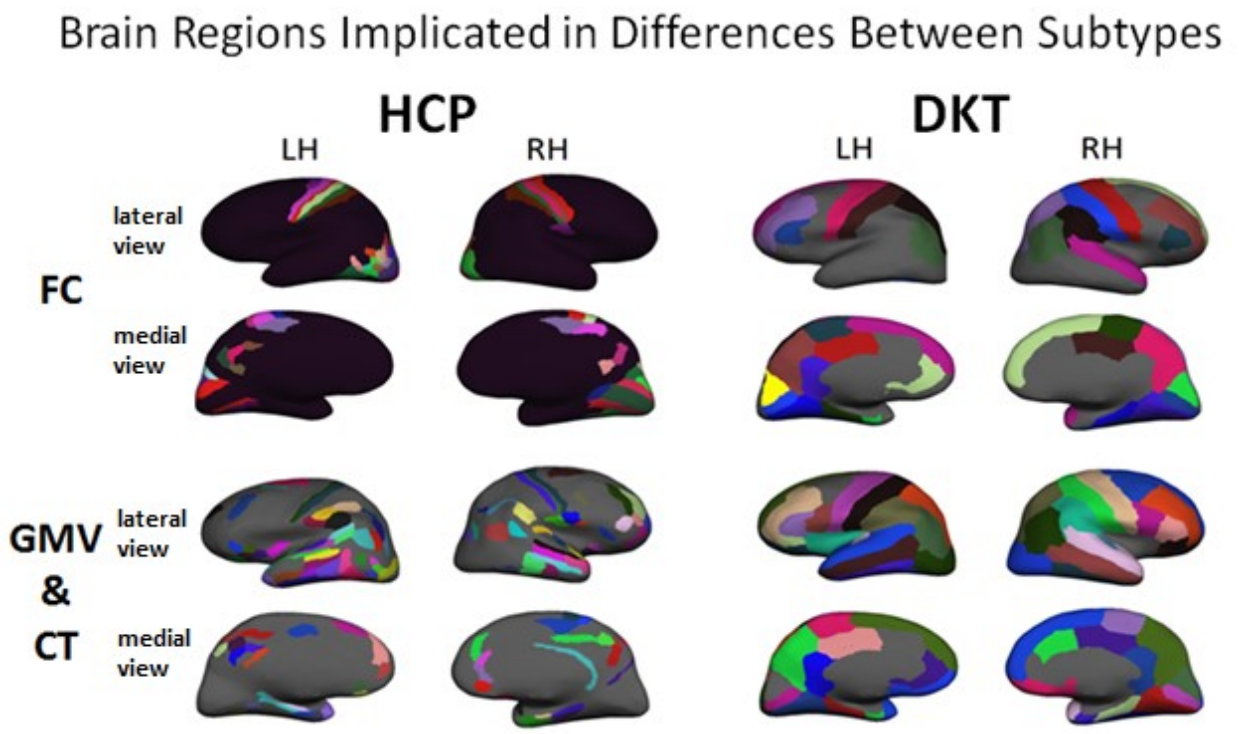


Figure 8. Brain regions implicated in the statistical differences between MCI subtypes. The regions implicated in the differences between subtypes generated using FC features are shown in the first two rows of brain images (row 1: lateral view of the brain's surface; row 2: medial view of the brain's surface), whereas the regions implicated in the differences between subtypes generated using GMV&CT features are shown in the bottom two rows. The results using HCP ROIs are shown in the first two columns (column 1: left hemisphere – LH; column 2: right hemisphere – RH), and the results using DKT ROIs are shown in the last two columns.

The brain regions implicated in the subtype differences between GMV&CT features are presented in Figures 8-10. Figures 9 and 10 quantify again the frequency with which a given region was implicated. When we consider our structural measures of GMV and CT, each of the brain regions is quantified only once for each of these measures, thus, the maximum number of times a region could be implicated would be twice. There are 119 (of 376) brain regions implicated in the HCP ROI-based structural feature contrasts with differences between subtypes, and these are presented along the horizontal axis of Figure 9. Those regions which were implicated for both GMV and CT are labeled (reduced labeling for legibility). They include areas A4 (Auditory 4 Complex), EC (entorhinal cortex), OFC (orbital frontal complex), p32 (the most anterior portion of the paralimbic cortex), PCV (Precuneus Visual Area), PeEc (Perirhinal Ectorhinal Cortex), PHT (middle temporal gyrus), PreS (PreSubiculum), PSL (Perisylvian

Language Area), STGa (auditory association area on the anterior superior temporal gyrus), STSdp (posterior of superior banks of Superior Temporal Sulcus), STSvp (posterior of inferior banks of STS), STV (Superior Temporal Visual area), TE1p (middle temporal gyrus posterior), TE2a (inferior temporal gyrus anterior) from the LH, and from the RH, areas 52 (Brodmann area 52), a9-46v (anterior region bridging the border between Brodmann areas 9 and 46 ventral), EC, PHA2 (Parahippocampal Area 2), PH, TE1p, TE2p (inferior temporal sulcus posterior), and V3B (visual area 3B). Of course, there were also many other regions implicated only once either for GMV or CT; all of these regions are presented on the cortex in Figure 8. Broadly, the implicated HCP regions in the GMV&CT-based subtype differences lie in the temporal lobe, parietal lobe, inferior frontal lobe, as well as other regions in the frontal lobe, and some of the lateral occipital lobe (including higher visual areas), especially in the LH. Other regions, including the posterior cingulate, primary somatosensory cortex, and primary motor cortex, were also implicated but only once for either GMV or CT. For the subcortical regions, we found that the caudate and thalamus were implicated in both hemispheres, as well as the globus pallidus and accumbens in the LH and the amygdala in the RH.

There were 66 (of 78, so in this case a high proportion) DKT ROI-based brain regions that were implicated in the contrasts between GMV&CT features between subtypes. Of these, those that were implicated for both GMV and CT include entorhinal, fusiform, interior temporal, lateral orbitofrontal, medial orbitofrontal, middle temporal, parahippocampal, paracentral, pars opercularis, postcentral, posterior cingulate, precentral, precuneus, rostral anterior cingulate, rostral middle frontal, superior frontal, superior parietal, superior temporal, supramarginal, and transverse temporal from the LH, and from the RH, entorhinal, fusiform, inferior parietal, isthmus cingulate, lateral orbitofrontal, lingual, medial orbitofrontal, lingual, medial orbitofrontal, middle temporal, parahippocampal, paracentral, pars opercularis, pars orbitalis, pars triangularis, pericalcarine, postcentral, posterior cingulate, precentral, and precuneus. In the case of the DKT ROIs, nearly all of the ROIs have been implicated in the differences between subtypes so it is more informative to mention those that were *not* implicated. Looking at Figure 8, the only areas which are notably absent from both hemispheres include anterior portions of the cingulate gyrus. Additionally, the subcortical regions implicated include the amygdala, caudate, globus pallidus, thalamus, and accumbens from both hemispheres and the hippocampus from the LH.

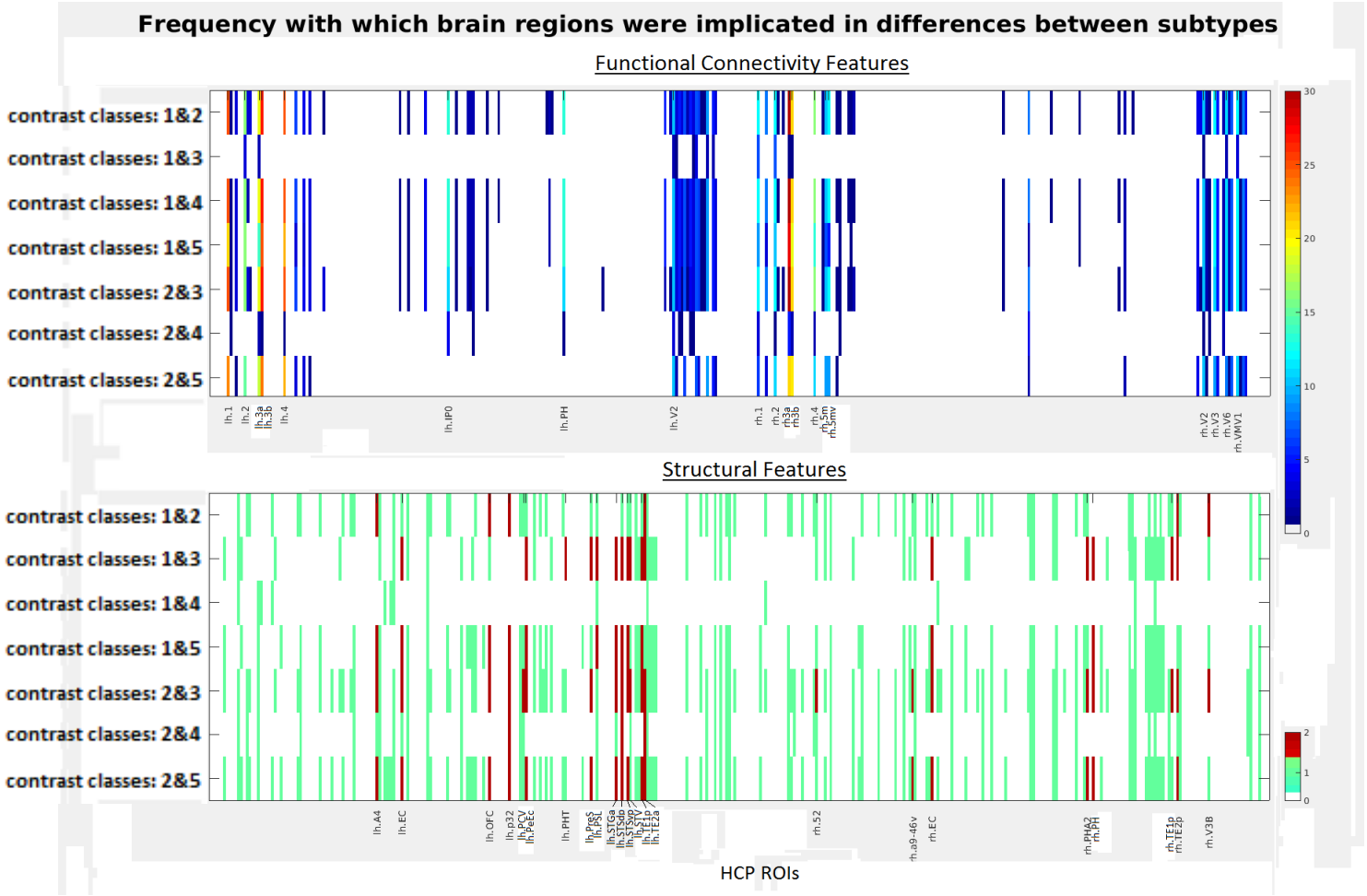


Figure 9. The frequency with which HCP parcellation-defined brain regions were implicated in the statistically significant differences between MCI subtypes. FC measures reflect the interaction between two brain regions, thus, for any one FC measure, two brain regions will be implicated. The FC features contrasted between subtypes involved 82 brain regions and these are shown along the horizontal axis of the top figure. For adequate readability, only those which were implicated more than 10 times are labelled. The structural features we considered were GMV and CT, each being measured only once per brain region, thus, the maximum frequency a region could be implicated in the structural feature differences between groups is twice. The structural features that were contrasted between subtypes involved 119 brain regions whose frequencies of implication are presented in the bottom figure. For adequate readability, only those which were implicated twice are labelled. The combinations of subtypes being contrasted are presented in the rows of each figure. Recall that subtypes 1 and 2 are stable subtypes, and subtypes 3, 4 and 5 are progressor subtypes.

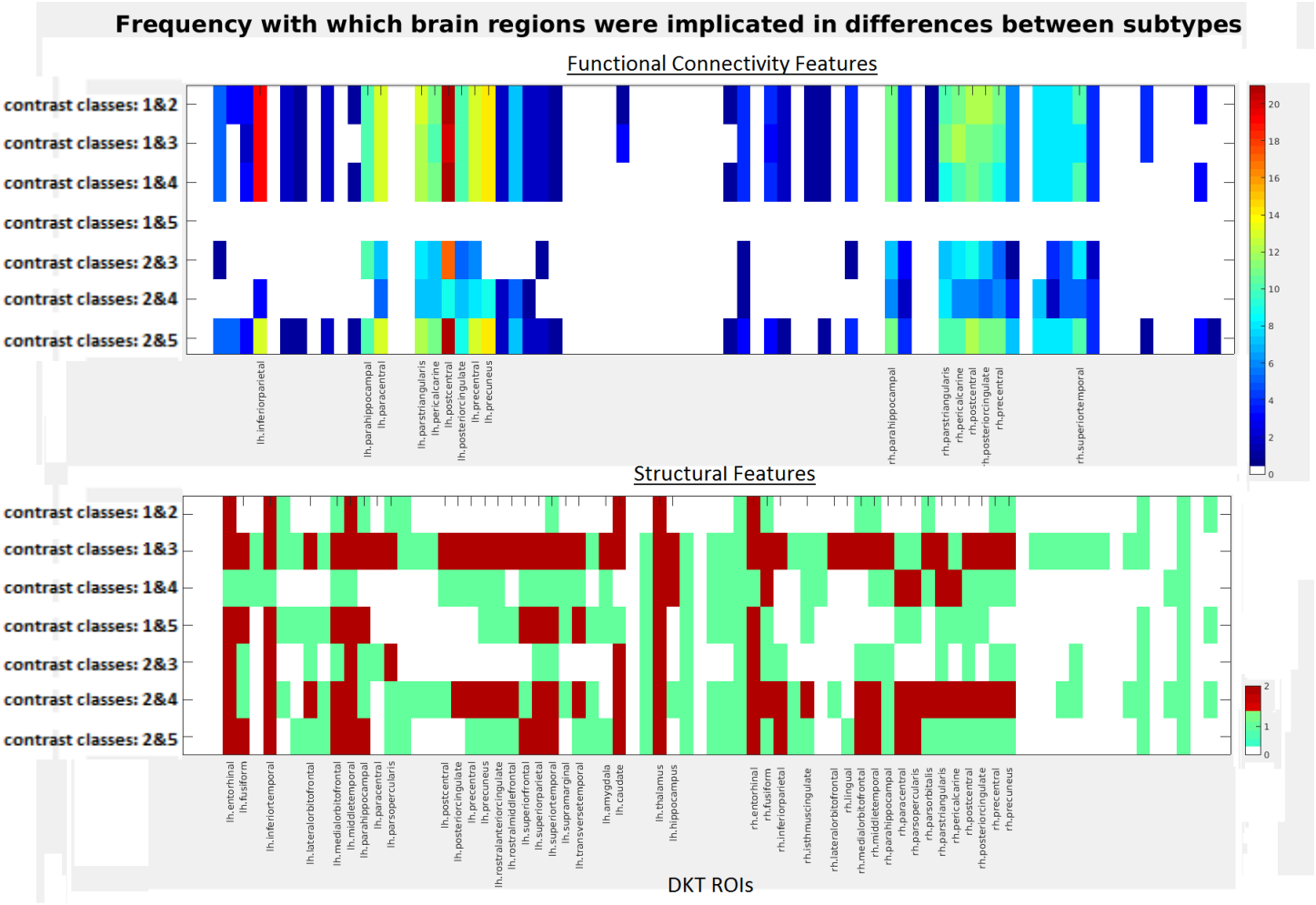


Figure 10. The frequency with which DKT parcellation-defined brain regions were implicated in the statistically significant differences between MCI subtypes. The FC features contrasted between subtypes involved 45 brain regions, shown along the horizontal axis of the top panel. The structural features contrasted between subtypes involved 66 brain regions whose frequencies of implication are presented in the bottom panel. The combinations of subtypes being contrasted are presented in the rows of each panel. The format of the presentation is as in Figure 9.

Discussion

A Stable-MCI-Subtype-Optimized Multi-Classifier Machine-Learning Pipeline

We have proposed a multi-classifier machine learning pipeline that is optimized for high PPV in a subtype of MCI individuals. We found that the set of classifiers that are optimized for stable subtypes are very reliable with high PPV across a 5-fold cross-validation, with the mean PPVc (i.e., corrected PPV considering the prevalence of non-progression) being as high as

96.1% (std 2.6%). The highest mean PPVc is achieved when we use the relatively fine-scaled HCP cortical parcellation (with the addition of subcortical regions). When we use our coarser-scaled parcellation, the DKT atlas, the mean PPVc is still quite good at 93.5% (std 6.1%). The finding that a subset of MCI individuals can be identified as non-progressors with high confidence is important and clinically relevant. The prevalence of non-progression to dementia in MCI individuals is high, estimated to be around 66.4%. In clinical trials that aim to target individuals with MCI for early treatment with the intent to delay or halt the progression to dementia, if one could identify individuals in that 66.4% of non-progressors (in the absence of treatment), a lot of time and money could be saved. By excluding stable subjects from these trials, the effects of the tested treatments would also be more easily observed, with fewer subjects having a positive response regardless of treatment. Our HCP ROI-based classifiers have an average sensitivity of 60.4% for identifying stable subjects, with a specificity of 94.8%. However, given the clinical scenario, it is more important to include all individuals who will progress to dementia in such a clinical trial, potentially at the expense of retaining some stable subjects. Our pipeline can facilitate the reduction of this unnecessary inclusion of stable subjects.

PPV and the Prevalence of Non-Progression in MCI

As of yet, the only other studies we are aware of which seek to optimize PPV in the MCI/AD population are Dansereau et al. (2017) and Tam et al. (2019). Dansereau et al. were able to achieve 90% PPV in a subset of MCI progressors using MRI-based features, including GMV, CT, and seed-based FC maps for 20 intrinsic connectivity brain networks. Tam et al. were able to achieve 87.8% PPVc in a subset of MCI progressors using GMV as well as several cognitive scores, including measures from the Boston Naming Test and the Alzheimer's Disease Assessment Scale−Cognitive (ADAS13). By shifting the focus of our study to classifying a subset of *stable MCI* subjects rather than progressors, we are able to improve upon the high PPVs these researchers have achieved, with our PPVc at 96.1%, as mentioned above, in our case using only MRI-based features (both structural and functional). Important differences between our study and these two prior studies include a high PPV in a subset of *stable* subjects, the finding that the use of the HCP ROIs allowed for precise classification of more stable subjects than the use of the DKT ROIs, and a larger sample size: our fMRI data set (612 RS-fMRI scans)

was significantly larger than that in Dansereau et al. (129 RS-fMRI scans; Tam et al. did not use fMRI scans).

We present our PPVc under the assumption that the population prevalence of MCI non-progression is approximately 66.4%. However, the sample's mean PPV was slightly different. In the HCP ROI-based and the DKT ROI-based analyses the mean PPVs were 92.5% (std 5.1%), and 88.2% (std 10.9%), respectively. If we reverse equation (1) to solve for prevalence using the sample PPV, we can find the pipeline's predicted sample prevalence of non-progression, which is 61.1% for the HCP ROIs, and 77.4% for the DKT ROIs. These are compared to the actual sample prevalence for our test data sets, which is on average 47.4% (kept approximately stable across the outer loop data partitions, reflecting the total sample prevalence of 47.7%). These values show that the use of the HCP ROIs as opposed to the DKT ROIs lends to a) a higher PPV/PPVc, and b) a predicted sample prevalence closer to the estimated population prevalence and the actual sample prevalence. There are a number of other ways that the analyses with the HCP ROIs have outperformed the analyses with the DKT ROIs, and these are discussed in section "*Finer-Scaled vs. Larger-Scaled Brain Parcellation*".

Novel/Unique Elements in our Pipeline Design

As illustrated above, our pipeline results in the novel finding that *stable* MCI subjects can be identified with high PPV. Our pipeline itself is also novel, or unique, in its multi-classifier design. Most classification studies will apply different classifier designs and parameters, and then identify the single best option from within the set. We argue that the separate classifiers may very well be tuned to optimally capture different subsets of subjects. In 4 of our 5 outer-loop iterations using the HCP ROIs, the set of high-PPV classifiers captured more subjects correctly than any one single "best" classifier in the set. For the analyses with the DKT ROIs, for 3 of 5 iterations, this is the case. In the 2 other iterations, only a few stable-optimized classifiers were among the set of high-PPV analyses. With the low representation of stable-optimized analyses, there was little opportunity for the DKT iterations to demonstrate the potential for multiple complimentary stable-optimized classifiers.

Another way in which our classifier is unique is that we have trained it exclusively using MRI measures from *MCI subjects*. Most classifiers that classify MCI subjects make use of data

from subjects with AD and HC in addition to data from MCI in order to give the classifier clearer examples of what progressor and stable features would resemble in their extreme states (assuming that AD is the 'extreme' state of progressors and HC is the 'extreme' state of stables). These additional clinical cohorts may also be utilized in subtyping studies, once again to draw out clearer differences in features between groups. Our success here in identifying a subset of stable MCI subjects with high PPV using only MRI-based MCI features is encouraging for those who would like to build and test their classifiers with MCI data alone, allowing for a more data driven approach, without providing pre-determined end points representative of the classes. For example, if the only set of patients with dementia represented at the extreme case for progressors are those with AD, other forms of dementia would be unrepresented and likely more poorly classified. That being said, researchers have made this choice to include AD and HC subjects in order to improve the accuracy of their classifications. It would be an interesting future direction to test our pipeline with additional cohorts of patients (with dementia and various clinical presentations of stable MCI) to reinforce the trends of the progressor and stable subtypes in the data.

Choice of Feature-Types to Test

In addition to structural and functional MRI measures, previous studies have also used other measures in their classification problems relating to MCI/AD. Other imaging modalities such as Positron Emission Tomography (ex. Zhang and Shen, 2012) have been shown to be informative, as have clinical scores (ex. Tam et al., 2019), and a number of other biomarkers such as genetic markers (ex. Ye et al., 2012), CSF amyloid content (ex. Hansson et al., 2006), blood tau levels (ex. Karikari et al., 2020), etc. Some of the best overall accuracies in MCI classifiers are achieved in with classification models that make use of features from multiple modalities. For example, Mathotaarachchi et al. (2017) achieved an overall accuracy of 84%, with sensitivity of 71% and specificity 87%, using amyloid PET and genetic markers (recall that we achieved 60.4% sensitivity and 94.8% specificity with the HCP ROIs). One of the aims of our study was to test the limits of what is possible with MRI-based measures alone. The reasoning behind this is simple: if one wishes to optimize a classification problem using multiple feature-types, the optimal design of each set of features should be used. The most straightforward way to identify the optimal design of each feature set is to study them

independently. With this goal in mind, we have pursued our classification study on MRI-based measures alone and tested the FC and structural features separately, allowing the observation of which choices optimize each individual feature type. For example, here, the HCP ROIs led to higher PPV in stable subjects than the DKT ROIs; therefore, future studies would likely be better served using the HCP ROIs. Upon identifying successful classifier designs on MRI-based features alone, a future direction for study would be to add in other optimized feature-types to determine whether the use of multiple feature types can improve the overall classification result. We have not pursued such additional tests here because of limitations in the compatibility of additional features available for our subjects across the 3 data sets used. The clinical measures and biomarkers available across the ADNI, OASIS3, and Han Lab data sets are not identical, and even *within* the larger data sets, the same measures or scan-types were not available for all subjects. If we were to try to add PET data, for example, only a small subset of our subjects would have these scans available and so our sample size would decrease significantly. Each of the different data sets does indeed have a clinical score assessing the degree of cognitive impairment of each subject, however, the scoring systems differ between data sets, and so to avoid confounds due to different metrics, we omit these feature types. We expect that the design and optimization of classifiers with multiple feature-types will be more promising once larger pools of subject data become available with consistent and diverse types of imaging and non-imaging biomarkers.

Another dimension to consider when building a classifier with multiple data types is the feasibility of gathering the data from new clinical subjects. If the goal is to eventually use the classifier in the clinical setting, the data required to apply the classifier must be affordable and accessible. The larger the set of measures one must gather on any given subject, the more costly and burdensome the pursuit becomes. With this in mind, a smaller set of highly informative features is ideal. There are also features being tested - such as amyloid PET-based features that show great promise for the classification of MCI/AD subjects - that are not yet widely accessible in the clinical setting (Guo et al., 2012). While T1 anatomical images – similar to those we have used in this study to compute our structural features - are regularly obtained, fMRI scans are not yet widely used in the clinical setting. Blood Oxygenation Level-Dependent (BOLD) fMRI is relatively new compared to MRI, though it has been used extensively in research for nearly 30 years. Some limitations to its use in the clinical setting include within and between-subject

variability in signal and the lack of standardized acquisition and analysis methods (Specht, 2020). That being said, there are also numerous studies that have shown that BOLD fMRI can be helpful in many clinical contexts, such as in neurosurgical planning (Leuthardt et al., 2018). Resting-state fMRI, in particular, is quite well suited to the clinical setting as the scans can be short (5-15 minutes typically) and they do not require the patient to cooperate in performing any task. fMRI scanning can also be accomplished using existing MRI machines, which may simply need some retrofitting with fMRI optimized hardware. At centres such as Washington University in St. Louis, resting-state fMRI is considered an integral part of standard imaging for neurosurgical planning (Leuthardt et al., 2018). All considered, and with examples of successfully implemented clinical fMRI in existence, we believe that resting-state fMRI-based FC is feasible in the clinical setting and therefore worth testing and exploring for use in clinical applications, as we have done here.

Mean/Mode Classifier Parameters and How They Compare to the Literature

Mode Feature Type: We tested our pipeline with a variety of parameters. The mean or mode of those parameters across the high-PPV analyses is presented in Table 4. In general, we saw that FC feature-based classifiers were selected more often than GMV&CT based classifiers. This is in contrast to what is typically found in MCI classification problems (ex. Schouten et al., 2016; Dansereau et al., 2017; Leandrou et al., 2018), where structural measures alone perform better than functional measures alone. The relative importance of FC vs. GMV&CT features in classification may be influenced by the fact that we have done our classification using subtypes of both stable and progressor MCI subjects, and have only used MCI subject data (no AD or HC data). There are, in fact, many studies showing that functional measures are valuable in differentiating AD vs. HC, and some studies do find that the functional features they use are more informative than the structural features (ex. Wee et al., 2013). Note that our finding that FC-based analyses were more frequent among the high-PPV analyses is true for the case with HCP ROIs and for the DKT ROI stable-optimized analyses. In the case where we have used DKT ROIs and considered *all high-PPV analyses*, we see that GMV&CT features are more frequently used among the high PPV analyses than FC features. Larger scaled ROIs, and a focus on the full set of subjects (i.e., both progressors and stables) are more in line with the approach that most other studies employ, so perhaps it is because of our novel choices (using the HCP

parcellation and optimizing separately for stable and progressor subtypes) that we find FC most useful. However, we do still find that GMV&CT are positive contributors to our classifications since the FC-based high-PPV analyses make up only 60-88% of the full set of high-PPV analyses (74-88% of stable-optimized high-PPV analyses), the remainder being GMV&CT-based. We know that in FC analyses, if voxels from outside of the functional region of interest are included in the mean time-course for that region, the observed FC with that region will be severely impacted (Smith et al., 2011). It is possible that with larger-scaled ROIs, meaningful FC differences that exist between finer-scaled ROIs are being lost due to this effect of signal cross over from neighboring regions. Possibly finer-scaled ROIs are better suited to detecting meaningful FC differences between MCI subtypes. Further discussion on structural vs. functional features as it pertains to the differences between subtypes observed in this study will be presented below in discussion section "*Subtypes Based on Different Feature-Types*".

Mode FC Measure: An additional item to discuss regarding functional features is the measure used to quantify the FC. In this study we evaluated two measures: correlation and ICOV. Correlation is the most widely used FC measure and the classic choice for use in whole-brain studies. ICOV, a regularized version of partial correlation which seeks to induce sparsity in the connectivity matrix and results in an estimated network of direct connections (as opposed to direct *and indirect* connections), is a method we have studied extensively and found to be very promising in the context of fine-scaled networks (Smith et al., 2011; Dawson et al., 2013, Dawson et al., under revision for Neuroimage, Manuscript 2 of this thesis). With our interest in pursuing this classification on relatively fine-scaled brain regions, we decided to test whether ICOV would continue to provide promising results. While the overall network we are evaluating in this whole-brain study is not small, the ROIs are relatively small, with relatively dense ROIs in any part of the brain. Therefore, we hoped that ICOV could contribute to refining the final set of FC-measures used in classification, avoiding what would be arguably redundant information in the form of indirect connections. However, none of the FC-based high-PPV analyses using either HCP ROIs (finer scaled) or DKT ROIs (larger scaled) made use of ICOV as a measure. All of the high-PPV analyses used correlation-based FC features. It is possible that ICOV failed in our network with finer-scaled ROIs (HCP ROIs) because of the complexity of the network. In this case, there were 338 ROIs between which we computed the FC. The ICOV method iteratively regresses out the information contained in all other ROIs from the ROIs being

correlated before computing the relationship. This high number of regressions results in a large amount of noise being introduced into the residual ROI time courses. While ICOV was shown to perform well in a local network with 60 ROIs (Dawson et al., under revision for Neuroimage, Manuscript 2 of this thesis), the larger and more complex network of the whole brain tested here is perhaps outside of its range of good performance given the current experimental context. We conclude that correlation is better suited than ICOV to this whole-brain classification, as supported by correlation being the sole FC measure utilized in our high-PPV analyses.

Mean Number of Features: Another parameter we can look at across high-PPV analyses is the number of features selected. Whether all high-PPV analyses were considered or only the stable-optimized high-PPV analyses, the mean number of features used was from 25-36 features (with both HCP and DKT ROIs). The typical number of features found to be optimal in other similar classification problems is ~5-20 (Khondocker et al., 2016; Ramirez et al., 2018; Eskildsen et al., 2013, 2015; Hojjati et al., 2017). The number of features we found to be optimal is slightly higher than the norm. However, in Hojjati et al. (2017), when they tested the optimal number of features using the MRMR algorithm (to predict MCI progression using resting-state fMRI), they found multiple "peak" numbers of features at approximately 10, 23, and 34 features. This scale of the optimal number of ROIs is indeed in line with what we have found. Other studies that selected a significantly higher number of features have found optimal numbers in the hundreds or thousands (Beheshti et al., 2016, 2017). We tested only up to 60 features in our pipeline in order to limit computational time. This is justified because most groups have found smaller numbers of features to be optimal/adequate. A potential future direction of study for our pipeline could be to examine higher numbers of features utilized in classification.

Mode Classifier Type: The final item from Table 4 we will discuss is the mode classifier type across high-PPV analyses. In all cases, RF was selected more frequently than SVM, being utilized in 58-83% of high-PPV classifiers. In the analyses with the HCP ROIs, the occurrence of RF was 58% and 68% for stable-optimized and all high-PPV analyses, respectively, whereas with the DKT ROIs the occurrence was on the order of 80%. In the literature, many different machine learning approaches have been used to build classifiers for the MCI/AD population. SVM and RF are among the more commonly used, however other methods such as logistic

regression, linear discriminant analysis, naïve Bayes, multilayer perceptrons, deep learning neural networks, etc. are also used (ex. Dai et al., 2012; Moradi et al., 2015; Clark et al., 2016; Amoroso et al., 2018). There is little agreement in the field as to which approach is the best as this is an active research question and one of the aims of this study. Khondocker et al. (2016) found that SVM performs better than RF on a classification using ADNI 1.5T MRI based estimation of regional cortical thickness and volumetric measures, in contrast to our findings. On the other hand, Sarica et al. (2017 review) found that RF produced one of the best accuracies of prediction of MCI progression at the time of study when used with multi-modal data, in line with our findings. Sarica et al.'s review encourages further studies comparing commonly used classification approaches in order to get a clearer sense of which approach would be best suited to the MCI progression classification problem. We are in full agreement with this sentiment and have attempted to contribute to the literature comparing RF and SVM. Our results indicate that while both RF and SVM can contribute to the correct classification of stable MCI subjects, RF classifiers contribute more to this successful classification. A more direct comparison will be needed in the future, especially with the development of more complex and modified ML approaches.

Finer-Scaled vs. Coarser-Scaled Brain Parcellation

In the Results section, we have discussed the outcomes of both the HCP and DKT ROI-based analyses and have described the trends in the results of each one. In this section, we will contrast the two parcellations emphasizing any differences between them. Recall that, overall, the HCP stable optimized high-PPV analyses achieved a PPVc of 96.1%, whereas the DKT analyses achieved a PPVc of 93.5% (Table 3). In addition to the higher PPV/PPVc, the analyses with the HCP ROIs also have a lower indeterminate rate (68.7% vs. 84.1%), higher sensitivity (60.4% vs. 27.8%), relatively equivalent specificity (94.8% vs. 95.3%), and more clearly defined structural feature-based subtypes. Given all of this, it is clear that the HCP ROI-based analyses perform better than the DKT ROI-based analyses.

The poorly defined subtypes using the GMV&CT features for the DKT ROI analyses are an important finding. This poor subtyping limits the ability of the pipeline to classify subjects into subtypes, with the majority of all stable subjects being clustered into a single subtype (all but one). This is effectively the same as doing no subtyping at all. This same trend of very small

subtypes (i.e., containing very few subjects) using DKT ROIs and GMV&CT features holds true for both the stable and progressor subtypes. One might consider that the poor subtyping observed with the DKT ROI-based GMV&CT feature-set is a result of forced subtypes on a data-set that does not lend itself to subtyping - i.e., could the GMV&CT of MCI subjects be uniformly or randomly distributed? The fact that clear subtypes can be identified using HCP ROI-based GMV&CT shows us that this is not strictly the case: subtypes *do exist* within the set of GMV&CT measures of MCI subjects when appropriate ROIs are used. With poor subtyping in all of the GMV&CT analyses in the DKT ROI-based grid search, it is not surprising that the DKT analyses capture fewer correctly classified subjects since there are fewer "viable" classifiers with meaningful subtypes for it to choose from. Importantly, the classifiers that do have meaningful subtypes are all built using the same feature type (FC). Given that the goal of the pipeline is to identify a subset of subjects with classifiers optimized for different information, classifiers built with more varied feature-types are valuable. Thus, having only FC-based classifiers to meaningfully contribute to the set of high-PPV classifiers is a shortcoming of the DKT ROI-based analyses.

We can also compare the HCP and DKT ROI-based analyses in terms of the brain regions implicated in the statistical differences between subtypes. Group differences between the FC-feature-based subtypes using the HCP vs. DKT ROIs were comparable, with the exceptions of more frontal and temporal regions being implicated when DKT ROIs were used. The brain regions implicated in the DKT GMV&CT differences between subtypes are broad: nearly all regions are implicated. This unspecific set of implicated regions is possibly related to the poor subtyping using these features; thus, we do not compare these results to those found using the HCP ROIs. A more detailed investigation of which regions were implicated in the subtype differences for each of the parcellations can be found below in the section "*Subtypes Based on Different Feature-Types*".

MCI Subtyping

Brain Regions Implicated in the Differences Between Subtypes

In our pipeline, we divide our groups of progressors and stables into subtypes and proceed to attempt to classify the subjects into those defined subtypes. A natural question to ask

at the end of this procedure is: What are the characteristics of the subtypes? As presented in the results section, the known demographics (age, sex, site, etc.) of the subjects do not explain the observed subtypes. Therefore, we are left to examine the differences in the features between subtypes to understand what underlies them. A visualization of the full set of implicated regions is presented in Figure 8 and the full list of differences between subtypes is presented in Supplementary Tables 5-10. Here we will discuss the subset of regions implicated in differences between specific subtypes and whether these regions have previously been shown to be implicated in the literature.

Recall that we have examined the subtypes of both progressors and stables separately and have also considered separate subtypes for structural and functional features. Given that our pipeline was only successful for the stable subtypes, it is largely those that we will discuss here: progressor subtypes will only be discussed in terms of how they differ from the stable subtypes. We start by discussing the differences in the subtypes defined using the FC (i.e., correlation) features. Recall that the set of features used in testing for differences between subtypes has already been pared down to those which were retained after two rounds of feature reduction in the pipeline. The first feature reduction procedure identified the connections with the lowest p-values after contrasting the entire progressor and stable groups, and the second procedure was the MRMR algorithm, a mutual information-based algorithm that prioritizes informative and non-redundant features predictive of the subtypes. For the HCP parcellation, the main sets of functional connectivity differences we see between these subtypes involve connections within and between both the primary somatosensory/sensorimotor and visual cortices (including visual association areas, and usually excluding V1). The altered connectivity is between regions are within-hemisphere and between-hemisphere relative to one another. There is also some involvement of medial parietal areas responsible for processing visual and somatosensory information, though these are less important (i.e., fewer altered connections with these areas) in terms of group differences.

Previous studies, like our study, have implicated sensorimotor regions in AD/MCI. Stephen et al. (2010) showed that the somatosensory cortex is affected early in the progression of AD and expected it should indeed affect functional measures. Tudela et al. (2019) showed differences in somatosensory and sensorimotor resting-state networks between groups at

different stages of progression in an AD rat model. In addition, Mattia et al. (2003) showed a bilateral reduction in cerebral blood volume response (measured via dynamic susceptibility contrast-enhanced MRI) which correlates with changes in the EEG power spectra over the sensoritmotor cortex in AD. All of these studies are consistent with our finding that sensory and motor areas of the brain show functional changes within the MCI cohort, especially among those who are likely to progress to AD. Therefore, in contrasting the stable subtypes with the progressor subtypes, statistically significant differences between FC measures in sensorimotor areas may represent a relative sparing in stable subjects of the changes seen in AD progressors. In fact, what we see is that for HCP and DKT ROIs, stable subtype 1 has broadly *greater FC* than all but one of the three progressor subtypes in sensorimotor areas (HCP subtype 3/DKT subtype 5 and subtype 1 show few statistically significant differences in FC in these areas, though the bulk of those that are significant show that the progressor subtype has greater FC). The opposite is true for stable subtype 2 which shows broadly *decreased FC* relative to all but one of the three progressor subtypes in sensorimotor areas (HCP subtype 4/DKT subtype 4 and subtype 2 show fewer statistically significant differences in FC in these areas, though the bulk of those that are significant show that the progressor subtype has decreased FC). These differing findings between the two stable subtypes are an important finding and suggest that future MCI studies should include subtyping in order to capture the heterogeneous presentation of this population.

As for studies implicating the occipital/visual cortex, as our study has, Wee et al. (2013) found that the functional connection between paracentral (LH) to lateral occipital (RH) is among the most discriminative features used in their MCI progressors vs. stable classification using the DKT parcellation. Zhang et al. (2009) showed altered FC between the posterior cingulate and bilateral visual cortex in mild AD compared to HC. Yamasaki et al. (2019), too, studied the visual system in AD and explained how visual motion deficits are common and that these deficits are largely expected to be due to altered connectivity in the posterior parietal cortex. Yamaski et al. found that V1 is also involved, despite most prior studies showing that V1 is relatively spared in early AD. Our study does implicate cortical areas which include V1 when the DKT ROIs are used, though V1 is less commonly implicated when the HCP ROIs are used. In another interesting study, Wang et al. (2019) showed that the visual cortex is hyper-connected in MCI relative to healthy controls, perhaps reflecting disease-related adaptations of brain

networks. While we did not test MCI vs. HC, we do observe relative hyper- and hypo-connectivity between stable and progressor MCI subtypes. Our findings with respect to the connectivity of the visual cortex in MCI subtypes is that stable subtype 1 is largely hyper-connected relative to the progressor subtypes, and stable subtype 2 is largely relatively hypo-connected (for both DKT and HCP ROIs, See Supplementary Tables 7 and 10). Finally, Deng et al. (2016) showed that higher-level visual networks are altered in AD patients, but not in MCI patients. We do find that some of the higher-level visual processing cortical areas are implicated in the differences between our MCI subtypes. The difference in findings relative to Deng et al. may be due to the fact that we are testing within MCI, not between MCI and HC. With the set of findings listed above, it is unclear if the functional changes observed in AD are adaptive or destructive or a combination of both. It is also not yet clear if these changes are consistent within the MCI population. However, given the strength of the finding in our study, we suggest that altered FC in the visual system is indeed present in MCI and is a major contributor to differentiating between subtypes.

In the DKT ROI FC-based subtype differences, the visual and sensorimotor areas are still identified, in addition to medial and inferior temporal and frontal areas, as well as more extensive inclusion of parietal regions. Broadly, the differing FC observed is within the visual cortex (intra- and inter-hemispheric connections), as well as between the visual cortex and parietal areas, sensorimotor areas, the posterior/isthmus cingulate, and some more frontal regions, most notably the pars triangularis. The pericalcarine area (likely including parts of V1 and V2) also showed altered connectivity with the parahippocampus and hippocampus. Likewise, sensorimotor areas show altered interconnectivity as well as altered connectivity with bilateral visual cortex, parietal areas, posterior/isthmus cingulate, pars triangularis, and the parahippocampal area. In addition, the postcentral gyrus (i.e., primary somatosensory area) show altered FC with the ipsilateral hippocampus. The inferior and medial portions of the parietal regions are mostly showing altered FC with the posterior/isthmus cingulate, visual cortex, and sensorimotor areas bilaterally, as well as some frontal areas. In addition to the connections already mentioned, the posterior cingulate shows altered FC with the inferior temporal region and the parahippocampal region as well as with its contralateral equivalent. The pars triangularis is the frontal region most commonly implicated in differing FC, with differences in connections with other frontal, parietal, sensorimotor, visual, parahippocampal, and hippocampal regions, as

well as with its contralateral equivalent. The only subcortical areas implicated were the hippocampus, and caudate regions, mostly from the right hemisphere.

Most of the additional regions showing altered FC in the DKT analyses relative to the HCP analyses are regions of the DMN. All DKT-regions of the DMN show altered FC with the exception of the lateral temporal lobe, which does not show altered FC in our study. As explained in the introduction, the DMN is the most frequently identified set of regions showing altered connectivity in the context of AD. Additionally, our findings show that the hippocampus, parahippocampal gyrus, and caudate are implicated, consistent with the literature (Sohn et al., 2014; Liu et al., 2016; Cai et al., 2015a; Badhwar et al., 2017) and with the altered memory profile in MCI.

There are a number of previous studies implicating the DKT regions commonly implicated in our study, mentioned above. Let us begin by discussing the pars triangularis. In the dominant hemisphere, the pars triangularis is one of two regions that make up Broca's area (expressive language), and in the non-dominant hemisphere, it is involved in non-verbal communication (Foundas et al., 1996). Serra et al. (2017) showed that the white matter tract connecting Broca's area to the supplementary motor area and pre-supplementary motor areas is damaged bilaterally in patients with AD. They also comment that this damage corresponds well to a known deficit in constructional praxis in AD patients. Our findings show decreased FC between the pars triangularis and the paracentral region (which includes the supplementary motor area) for progressor subtypes 3 and 4 relative to stable subtype 1, within the LH and also between the left paracentral region and the right pars triangularis. We did not, however, find FC differences between subtypes between these regions within the RH. Metzger et al. (2016) showed that a behavioral-subtype of frontotemporal dementia has a pattern of increased cortical activation in frontopolar regions including Broca's area. This finding is analogous to our finding that stable subtype 2 shows broadly decreased FC involving the pars triangularis relative to progressor subtypes 3 and 5. We also found that stable subtype 1 showed broadly *increased* FC relative to progressor subtypes 3 and 4 involving the pars triangularis. This finding is possibly analagous to increased functional activation in Broca's area in MCI found in Farina et al. (2017). Farina et al. also showed decreased functional activation in AD: This mixed picture of decreased vs. increased functional activity in the frontal cortex depending on type and stage of disease is

consistent with our finding that stable subtypes 1 and 2 have opposite relative FC to the progressor subtypes (i.e., subtype 1 is hyperconnected and subtype 2 is hypoconnected relative to the progressor subtypes), accounting for both trends. Farina et al. (2017) contribute to the body of literature implicating the parietal cortex, in addition to the pars triangularis/frontal cortex. They found a gradually declining activation of the fronto-parietal network as cognitive impairment worsens. Additionally implicating both of parietal and frontal regions, Grothe and Teipel (2016) reported that the frontoparietal-control network demonstrates altered FC in MCI or mild AD compared to healthy controls. Our results show that the FC between parietal and frontal regions are largely *increased* in stable subtype 1 relative to progressor subtypes 3 and 4, and *decreased* in stable subtype 2 relative to progressor subtype 5, as well as in some cases progressor subtype 3. Stable subtype 2 also sometimes shows an *increased* FC relative to progressor subtype 4 between parietal and frontal regions. Clearly, there is a mixed picture in our results, possibly reflecting different stages of disease captured in our subtypes, which would be consistent with the Farina et al. and Grothe and Teipel findings. The body of evidence showing functional, behavioral, and white matter tract alterations in AD and MCI indicate that is it likely that frontal and parietal regions of the brain are relevant in differentiating MCI subtypes, as our findings show.

Where the implicated subcortical regions are concerned, the hippocampus is well known to be an area implicated in MCI/AD, with atrophy and altered functional connections (Chen et al., 2015). Our findings are consistent with this with the left hippocampus showing greater CT in stable subtype 1 relative to progressor subtypes 3 and 4, and right hippocampus showing altered FC in the DKT ROI-based analyses. Sohn et al. (2014) found that the left hippocampus has decreased FC with the posterior cingulate cortex and lateral parietal areas. While our study did not find the left hippocampus to be implicated, the left parahippocampal region was indeed implicated and showed greater FC with contralateral posterior cingulate and superior parietal areas in stable subtype 1 relative to progressor subtypes 3 and 4. In addition to finding that the hippocampus is implicated in MCI, we also found that the caudate is implicated, as Madsen et al. (2010) did. Madsen et al. showed that normal caudate asymmetry is lost in AD with a higher level of right caudate atrophy than left (consistent with our results).

Having discussed thoroughly the regions implicated in differences between the subtypes generated using FC features, we next discuss those implicated in differences between the subtype generated using GMV&CT features. We only cover the subtypes created using the HCP ROIs because the DKT ROI-based subtypes identified single subjects as subtypes, indicating a lack of between-subject similarity, and thus poor clustering into subtypes. The HCP GMV&CT subtypes showed differences between groups involving the dorsal and ventral visual streams, the MT+ complex and its neighbors, as well as several sensorimotor areas (early somatosensory cortex, motor cortex, sensori-motor associated paracentral lobular, and mid cingulate cortex), compatible with findings from the FC-based subtypes. However, the predominant regions showing differences in GMV&CT between MCI subtypes are frontal (medial prefrontal cortex, orbital and polar frontal cortex, the dorsolateral prefrontal cortex, insular and frontal opercular cortex) and temporal regions (medial and lateral temporal cortex, auditory association cortex) as well as some parietal regions and posterior portions of the cingulate (temporal-parietal-occipital junction, the inferior parietal cortex, the posterior cingulate cortex). In addition, there are also a number of subcortical regions that were implicated, namely the caudate and thalamus from both hemispheres, the amygdala from the RH, and the globus pallidus and accumbens from the LH. Most of these regions, including regions in the default mode network, are well known to be implicated in AD, however, there is a broader representation of frontal areas, temporal areas, and the visual and sensorimotor cortices.

There also exists evidence supporting the findings that sensorimotor and visual cortex regions are implicated in terms of structural measures, as we have found in our study. For example, Yuan et al. (2008) found that the motor cortex has lower GMV in AD compared to HC and Wagner et al. (2020) showed that a somatosensory structural covariance network is informative in classifying AD patients and healthy controls. These studies are consistent with our findings that stable subtype 1 has greater GMV than progressor subtype 3, bilaterally, and that structural measures involving the somatosensory regions of the brain were selected as informative features by our pipeline. Regarding the visual/occipital cortex, Delli Pizzi et al. (2014) found that the pericalcarine and lingual gyri, cuneus, precuneus, and superior parietal gyrus have reduced cortical thickness bilaterally in Lewy Body Dementia when compared to controls, consistent with our findings (with the exception of the pericalcarine and cuneus regions

from the LH, and the superior parietal region from the RH which were not clearly found to have decreased CT in progressor subtypes relative to stable subtypes).

We have not yet discussed many subcortical regions (aside from hippocampus and caudate above). However, many subcortical regions have indeed also been implicated in dementia, including here in our study. For example, Moller et al. (2015) found that the GMV of the accumbens and caudate differ between individuals with frontotemporal dementia (FTD), AD, and subjective memory complaints, with the most severe atrophy in FTD. The globus pallidus, hippocampus, amygdala, thalamus, and putamen GMVs are smaller for both dementia types relative to the volume in those with only subjective complaints. Supplementary Tables 5-10 include the full list of regions implicated in the subtype differences for our study; indeed, the accumbens, amygdala, thalamus, and the globus pallidus are each implicated to some capacity (we do not find the putamen to be implicated). Consistent with the literature, the subcortical regions most strongly implicated in our study are the hippocampus and the caudate. However, as demonstrated above in our discussion of the literature, there is likely a broader influence of the subcortex in MCI/dementia (note that additional subcortical regions are implicated sparsely in our study, as listed in Supplementary Tables 5-10).

Subtypes Based on Different Feature-Types

It is clear that the regions we have found to be implicated have all been previously implicated in MCI, AD, or another form of dementia in some capacity, though rarely with such a fine-grained definition of regions. The literature is vast on these topics, and a wide variety of brain regions have indeed been implicated. Perhaps what is more interesting than what *has* been implicated is what *has not* been. Notably, the DMN is implicated in the DKT ROI FC-based subtype differences, but not when the HCP ROIs are used. The same can be said for the fronto-parietal regions. A meta-analysis performed by Tam et al. (2015) showed that the superior medial frontal cortex (including the anterior cingulate), dorsomedial prefrontal cortex, striatum, and middle temporal lobe were the regions most commonly showing altered FC in the context of amnestic MCI: none of which were flagged in the HCP ROI FC subtype differences. Similarly, in a meta-analysis by Gu and Zhang (2019), MCI related functional activity was most commonly altered in the superior temporal gyrus, posterior cingulate, precuneus, uncus (i.e., anterior parahippocampal gyrus, and hippocampus), insula, lateral parietal lobe, and middle and inferior

frontal gyri. In comparison to GU and Zhnag (2019), the only area that was implicated in our HCP ROI FC-based subtype differences is the precuneus. This finding could be a reflection of the fact that the bulk of the literature studies ROIs of a comparable scale to the DKT atlas and very few are done with smaller ROIs such as those in the HCP parcellation. It is possible that upon parcellation of the brain into finer ROIs, different features would be extracted, and thus different trends are elucidated. This seems to be the case here, given that with the same data in larger scaled ROIs, the results are much more in line with the bulk of the literature. That being said, as illustrated above, the regions which our study implicated in the HCP FC subtype differences (visual/occipital cortex, sensorimotor cortices, and medial parietal regions) have been implicated by the literature, just less frequently. As of yet, the changes in these primary sensory and motor areas are not well understood in the context of dementia. Despite this, there is accumulating evidence that disturbances to sensory and motor function are indeed linked to AD and can, in fact, precede cognitive symptoms (Albers et al., 2015). For example, it has been shown that a subset of individuals with AD will have impairments of electrophysiological and psychophysical measures of central visual function and impairment of certain visual functions (Albers et al., 2015). Reduced walking speed is also thought to precede the onset of cognitive impairment by 5–15 years, and is a strong risk factor for developing AD dementia (Verghese et al., 2007). Finally, it has been suggested that individuals with AD have altered pain sensitivity compared to healthy controls (Stubbs et al., 2016), suggesting that changes in the cortical processing of sensation may be impacted. With this in mind, possibly our MCI subtypes indeed identify these subsets of individuals with sensory and motor changes. Given that we successfully classify the stable subjects, it could well be that these individuals have fewer sensorimotor disturbances.

The HCP ROI-based GMV&CT subtypes included statistically significant differences in some sensoritmotor and visual areas, as well as a number of other areas expected given the literature, as discussed above. Interestingly, however, analyses using subtypes generated using FC-based features were found almost always more informative in our classification task. One possible reason for this could be the quality of the subtyping using the different feature types. Recall that when the DKT ROIs were used, the GMV&CT subtypes were comprised of a single subject each, and then the remainder of subjects were assigned into a single group. This is a phenomenon that occurs when the inter-subject similarity is low for all subjects (particularly in

relation to intra-subject similarity) and therefore meaningful subtypes are difficult to identify. It then follows that with weaker trends within a subtype, a classifier would have a harder time identifying to which subtype a new subject should be allocated. With stronger between-subject trends and similarity represented in the subtypes defined using FC data, it is unsurprising that FC feature-based classifiers were more often successful and thus selected to be among the set of high-PPV analyses.

Age and Sex Features

Age and chromosomal sex are well-established risk factors for developing AD. Age is the strongest risk factor for AD and over 60% of individuals who develop AD are female (Riedel et al., 2016). It stands to reason then that these characteristics would be important features in a classification problem involving individuals who may develop AD. Interestingly, none of the high-PPV analyses selected used age or sex as a feature for classification after optimizing the features via the MRMR algorithm. Based on the design of the algorithm, this means that either age and sex were less informative than the other features, or the trends they described across subjects were redundant in terms of the characterization of the subtypes. As demonstrated in Table 5, the mean age of the individuals across the stable subtypes is quite consistent at around 73-75 years (std=8.5-9.0 years). This is comparable to the sample mean of approximately 74 years (std=7.7 years) of age. The exception to this is stable subtype 2 as defined using HCP ROI-based GMV&CT where the mean age was younger, at 67 9 (std=6.6 years). As discussed above, however, we suspect that the subtypes generated using GMV&CT features are less reliable than those generated using the FC measures given weaker between-subject similarities and poorer representation of GMV&CT-based classifiers in the set of high-PPV analyses. This subtype also contains far fewer subjects (21 vs. 88-143) and perhaps the slightly deviated mean age is a byproduct of the smaller sample size. If these factors are causative of the lower age observed for this subtype, the exclusion of age as an important predictive feature would be because of low relevance. It is indeed possible, however, that this group represents a sample of younger stable individuals and the age vector is simply redundant, and the trend of age would already have been captured in the patterns of the GMV or CT of the subjects. The brain is known to atrophy even with normal aging (Fjell et al., 2014), so this is entirely plausible. Considering sex, the distribution of subjects into the stable subtypes was once again stable relative to the entire

sample with 53-58% males in all subtypes except for, once again, stable subtype 2 defined using HCP ROI-based GMV&CT where there were 62% males. As stated above, there is a slightly higher likelihood of developing dementia for females, so stable subtype 2 could plausibly be representing the relative stability in males. As discussed for age, it is also possible that this is an artifact of the smaller sample size in this subtype.

Predicting Stability vs. Progression

While we initially set out with the goal of identifying individuals who will progress to dementia *as well as* those who will not, our pipeline was only successful in predicting stable MCI. We wondered what might be the reason for the relative ease of high-PPV classification of stables compared to progressors. Progression from MCI to dementia is known to be a difficult classification problem, hence the continued and abundant work on it. Most prior studies have focused their optimizations on correctly identifying progressors rather than stables. Given this, it is possible that an easier classification of stables has always been possible, but it was not emphasized due to different study priorities. The reason behind the ease of classification of stables is not clear, however, we do have some hypotheses as to elements that may be contributing to this. It is known that some individuals with MCI not only fail to progress to dementia, but they actually return to what is deemed as "normal function". Within the Kungsholmen Project, it was found that a quarter of the individuals with MCI reverted to normal functioning or showed improved cognition within three years, and only one-third of the individuals with MCI progressed to dementia (Fratiglioni et al., 2007). The subset of individuals who do indeed regain cognitive function may be much closer in terms of feature trends to healthy controls than to MCI itself.

In a similar vein, there is currently no clear consensus on the criteria to differentiate normal from mildly pathological cognition in the elderly (Díaz-Mardomingo et al., 2017). For example, in a direct comparison, the Petersen/Winblad criteria classified 34% of participants as having MCI, while the Jak/Bondi criteria classified only 24% as having MCI (Jak et al., 2016). Therefore, it is possible that one of the stable subtypes identified in our study is indeed a variant of normal decline in aging, once again making that subtype distinct from the other subtypes of MCI patients.

Another possible contributor to our successful classification of stables only is that the data sources we used emphasized AD dementia specifically (as opposed to other forms of dementia) and thus targeted their data gathering to those who had clinical or genetic characteristics compatible with AD, and often excluded those with risk factors for other forms of dementia. In doing so, the heterogeneity of our MCI cohort would be decreased, particularly for the progressor groups, as most of the individuals would progress to dementia of the AD type. This type of data distribution may have led to: 1) less clear subtypes for progressors, and 2) a more clear separation of stables from the set of relatively homogeneous MCI progressors.

Finally, there are many factors such as motivation, functional status, comorbidities, etc., that influence the likelihood of progression. It is possible that the functional or structural MRI differences we observed between groups captured the brain's representations of these risk factors lending the data to make more clearly separable stable subtypes.

Conclusion

In this study we built a multi-classifier pipeline where each classifier optimized the PPV of an individual MCI subtype, the set of which allowed for correct classification of a greater number of subjects. We found that we were able to classify stable MCI subjects with a high PPV. We tested a finer scaled (HCP) and coarser scaled (DKT) brain parcellation and both of these were successful with mean prevalence-corrected PPVs of 96.1% and 93.5%, respectively. However, it is important to note that the HCP parcellation resulted in more stable (as opposed to progressor) subjects being correctly classified and a lower indeterminate rate than the DKT parcellation. We analyzed both FC-based features and GMV & CT features in building our classifiers. For the FC features, we tested two measures - ICOV and correlation - and we found that only correlation-based FC classifiers were selected to be among our set of high-PPV classifiers. FC-based classifiers were selected more often than GMV&CT based classifiers, possibly due to more clearly defined subtypes using FC features when compared to GMV&CT features. Overall, 3 progressor and 2 stable subtypes were optimal within the data we have studied here. The brain regions implicated in the differences between subtypes were largely in line with those implicated in the literature, including regions of the default mode network,

prefrontal cortex, temporal lobe, parietal regions, and the hippocampus and caudate nucleus of the subcortex. In addition, we found a strong trend of visual cortex and sensorimotor areas being implicated in group differences. Our pipeline shows promise for identifying a subset of MCI individuals with high PPV. Notably, the stable subtypes appear more easily classified. For now, we present this pipeline as a concept and recognize more testing is required before any strong claims on its results can be made. Future directions include validating the pipeline on simulated data and testing the pipeline on data from a site not accounted for in the training data set.

References

Albers, M.W., Gilmore, G.C., Kaye, J., Murphy, C., Wingfield, A., Bennett, D.A., Boxer, A.L., Buchman, A.S., Cruickshanks, K.J., Devanand, D.P., Duffy, C.J., Gall, C.M., Gates, G.A., Granholm, A.-C., Hensch, T., Holtzer, R., Hyman, B.T., Lin, F.R., McKee, A.C., Morris, J.C., Petersen, R.C., Silbert, L.C., Struble, R.G., Trojanowski, J.Q., Verghese, J., Wilson, D.A., Xu, S., Zhang, L.I., 2015. At the interface of sensory and motor dysfunctions and Alzheimer's disease. Alzheimer's & Dementia 11, 70-98.

Amoroso, N., Diacono, D., Fanizzi, A., La Rocca, M., Monaco, A., Lombardi, A., Guaragnella, C., Bellotti, R., Tangaro, S., 2018. Deep learning reveals Alzheimer's disease onset in MCI subjects: Results from an international challenge. Journal of Neuroscience Methods 302, 3-9.

Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D., 2017. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. NeuroImage 145, 137-165.

Ardekani, B.A., Bermudez, E., Mubeen, A.M., Bachman, A.H., 2017. Prediction of Incipient Alzheimer’s Disease Dementia in Patients with Mild Cognitive Impairment. Journal of Alzheimer's Disease 55, 269-281.

Badhwar, A., Tam, A., Dansereau, C., Orban, P., Hoffstaedter, F., Bellec, P., 2017. Resting-state network dysfunction in Alzheimer's disease: A systematic review and meta-analysis. Alzheimers Dement (Amst) 8, 73-85.

Balachandar, R., John, J.P., Saini, J., Kumar, K.J., Joshi, H., Sadanand, S., Aiyappan, S., Sivakumar, P.T., Loganathan, S., Varghese, M., Bharath, S., 2015. A study of structural and functional connectivity in early Alzheimer's disease using rest fMRI and diffusion tensor imaging. Int J Geriatr Psychiatry 30, 497-504.

Battista, P., Salvatore, C., Castiglioni, I., 2017. Optimizing Neuropsychological Assessments for Cognitive, Behavioral, and Functional Impairment Classification: A Machine Learning Study. Behavioural neurology 2017, 1850909-1850909.

Beheshti, I., Demirel, H., Farokhian, F., Yang, C., Matsuda, H., 2016. Structural MRI-based detection of Alzheimer's disease using feature ranking and classification error. Computer Methods and Programs in Biomedicine 137, 177-193.

Beheshti, I., Demirel, H., Matsuda, H., 2017. Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. Computers in Biology and Medicine 83, 109-119.

Binnewijzend, M.A., Schoonheim, M.M., Sanz-Arigita, E., Wink, A.M., van der Flier, W.M., Tolboom, N., Adriaanse, S.M., Damoiseaux, J.S., Scheltens, P., van Berckel, B.N., Barkhof, F., 2012. Resting-state fMRI changes in Alzheimer's disease and mild cognitive impairment. Neurobiol Aging 33, 2018-2028.

Cai, S., Chong, T., Zhang, Y., Li, J., von Deneen, K.M., Ren, J., Dong, M., Huang, L., 2015a. Altered Functional Connectivity of Fusiform Gyrus in Subjects with Amnestic Mild Cognitive Impairment: A Resting-State fMRI Study. Frontiers in Human Neuroscience 9, 471.

Cai, S., Huang, L., Zou, J., Jing, L., Zhai, B., Ji, G., von Deneen, K.M., Ren, J., Ren, A., 2015b. Changes in thalamic connectivity in the early and late stages of amnestic mild cognitive impairment: a resting-state functional magnetic resonance study from ADNI. PLoS ONE [Electronic Resource] 10, e0115573.

Challis, E., Hurley, P., Serra, L., Bozzali, M., Oliver, S., Cercignani, M., 2015. Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI. NeuroImage 112, 232-243.

Chen, G., Ward, B.D., Xie, C., Li, W., Wu, Z., Jones, J.L., Franczak, M., Antuono, P., Li, S.-J., 2011. Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging. Radiology 259, 213-221.

Clark, D.G., McLaughlin, P.M., Woo, E., Hwang, K., Hurtz, S., Ramirez, L., Eastman, J., Dukes, R.-M., Kapur, P., DeRamus, T.P., Apostolova, L.G., 2016. Novel verbal fluency scores and structural brain imaging for prediction of cognitive outcome in mild cognitive impairment. Alzheimer's & dementia (Amsterdam, Netherlands) 2, 113-122.

Dai, Z., Yan, C., Wang, Z., Wang, J., Xia, M., Li, K., He, Y., 2012. Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3). NeuroImage 59, 2187-2195.

Dansereau, C., Tam, A., Badhwar, A., Urchs, S., Orban, P., Rosa-Neto, P., Bellec, P., 2017. A brain signature highly predictive of future progression to Alzheimer's dementia. arXiv: Quantitative Methods.

Dawson, D.A., Cha, K., Lewis, L.B., Mendola, J.D., Shmuel, A., 2013. Evaluation and calibration of functional network modeling methods based on known anatomical connections. NeuroImage 67, 331-343.

Dawson, D.A., Jack Lam, Amir Shmuel, 2019. Local Principal Component regression, partial correlation and inverse covariance identify direct connections in an unknown, fine-scale, densely connected functional network. submitted.

Delli Pizzi, S., Franciotti, R., Tartaro, A., Caulo, M., Thomas, A., Onofrj, M., Bonanni, L., 2014. Structural alteration of the dorsal visual network in DLB patients with visual hallucinations: a cortical thickness MRI study. PLoS One 9, e86624.

Deng, Y., Shi, L., Lei, Y., Wang, D., 2016. Altered topological organization of high-level visual networks in Alzheimer's disease and mild cognitive impairment patients. Neurosci Lett 630, 147-153.

Desikan, R.S., Cabral, H.J., Settecase, F., Hess, C.P., Dillon, W.P., Glastonbury, C.M., Weiner, M.W., Schmansky, N.J., Salat, D.H., Fischl, B., 2010. Automated MRI measures predict progression to Alzheimer's disease. Neurobiology of Aging 31, 1364-1374.

Díaz-Mardomingo, M.D.C., García-Herranz, S., Rodríguez-Fernández, R., Venero, C., Peraita, H., 2017. Problems in Classifying Mild Cognitive Impairment (MCI): One or Multiple Syndromes? Brain sciences 7, 111.

Dyukov, Y.V., Bachinskaya, N.Y., Cholin, V.A., Kolyada, A.K., Vaiserman, A.M., 2015. [GENETIC AND EPIGENETIC DETERMINANTS OF ALZHEIMER'S DISEASE]. Adv Gerontol 28, 299-306.

Eskildsen, S.F., Coupé, P., Fonov, V.S., Pruessner, J.C., Collins, D.L., 2015. Structural imaging biomarkers of Alzheimer's disease: predicting disease progression. Neurobiology of Aging 36, S23-S31.

Eskildsen, S.F., Coupe, P., Garcia-Lorenzo, D., Fonov, V., Pruessner, J.C., Collins, D.L., 2013. Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. NeuroImage 65, 511-521.

Fan, Y., Batmanghelich, N., Clark, C.M., Davatzikos, C., Alzheimer's Disease Neuroimaging, I., 2008. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. NeuroImage 39, 1731-1743.

Farina, E., Baglio, F., Pomati, S., D'Amico, A., Campini, I.C., Di Tella, S., Belloni, G., Pozzo, T., 2017. The Mirror Neurons Network in Aging, Mild Cognitive Impairment, and Alzheimer Disease: A functional MRI Study. Front Aging Neurosci 9, 371.

Ferreira, D., Pereira, J.B., Volpe, G., Westman, E., 2019. Subtypes of Alzheimer's Disease Display Distinct Network Abnormalities Extending Beyond Their Pattern of Brain Atrophy. Frontiers in Neurology 10.

Ferreira, D., Verhagen, C., Hernández-Cabrera, J.A., Cavallin, L., Guo, C.-J., Ekman, U., Muehlboeck, J.S., Simmons, A., Barroso, J., Wahlund, L.-O., Westman, E., 2017. Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: longitudinal trajectories and clinical applications. Scientific reports 7, 46263-46263.

Fjell, A.M., McEvoy, L., Holland, D., Dale, A.M., Walhovd, K.B., Alzheimer's Disease Neuroimaging, I., 2014. What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus. Progress in neurobiology 117, 20-40.

Fortin, J.P., Cullen, N., Sheline, Y.I., Taylor, W.D., Aselcioglu, I., Cook, P.A., Adams, P., Cooper, C., Fava, M., McGrath, P.J., McInnis, M., Phillips, M.L., Trivedi, M.H., Weissman, M.M., Shinohara, R.T., 2018. Harmonization of cortical thickness measurements across scanners and sites. NeuroImage 167, 104-120.

Fortin, J.P., Parker, D., Tunc, B., Watanabe, T., Elliott, M.A., Ruparel, K., Roalf, D.R., Satterthwaite, T.D., Gur, R.C., Gur, R.E., Schultz, R.T., Verma, R., Shinohara, R.T., 2017. Harmonization of multi-site diffusion tensor imaging data. NeuroImage 161, 149-170.

Foundas, A.L., Leonard, C.M., Gilmore, R.L., Fennell, E.B., Heilman, K.M., 1996. Pars triangularis asymmetry and language dominance. Proc Natl Acad Sci U S A 93, 719-722.

Fratiglioni, L., Winblad, B., von Strauss, E., 2007. Prevention of Alzheimer's disease and dementia. Major findings from the Kungsholmen Project. Physiol Behav 92, 98-104.

Gauthier, S., Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, Belleville S, Brodaty FL, Bennett D, Chertkow H, 2006. Mild cognitive impairment. Lancet 367, 1262–1270.

Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., Smith, S.M., Van Essen, D.C., 2016. A multi-modal parcellation of human cerebral cortex. Nature 536, 171-178.

Gray, K.R., Wolz, R., Heckemann, R.A., Aljabar, P., Hammers, A., Rueckert, D., 2012. Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease. NeuroImage 60, 221-229.

Greicius, M.D., Kimmel, D.L., 2012. Neuroimaging insights into network-based neurodegeneration. Curr Opin Neurol 25, 727-734.

Grothe, M.J., Teipel, S.J., 2016. Spatial patterns of atrophy, hypometabolism, and amyloid deposition in Alzheimer's disease correspond to dissociable functional brain networks. Hum Brain Mapp 37, 35-53.

Gu, L., Zhang, Z., 2019. Exploring Structural and Functional Brain Changes in Mild Cognitive Impairment: A Whole Brain ALE Meta-Analysis for Multimodal MRI. ACS Chemical Neuroscience 10, 2823-2829.

Guo, S., Getsios, D., Hernandez, L., Cho, K., Lawler, E., Altincatal, A., Lanes, S., Blankenburg, M., 2012. Florbetaben PET in the Early Diagnosis of Alzheimer's Disease: A Discrete Event Simulation to Explore Its Potential Value and Key Data Gaps. International Journal of Alzheimer’s Disease 2012, 548157.

Hansson, O., Zetterberg, H., Buchhave, P., Londos, E., Blennow, K., Minthon, L., 2006. Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study. The Lancet Neurology 5, 228-234.

Hojjati, S.H., Ebrahimzadeh, A., Khazaee, A., Babajani-Feremi, A., 2017. Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM. Journal of Neuroscience Methods 282, 69-80.

Huang, H., Ding, M., 2016. Linking Functional Connectivity and Structural Connectivity Quantitatively: A Comparison of Methods. Brain Connect 6, 99-108.

Hwang, J., Kim, C.M., Jeon, S., Lee, J.M., Hong, Y.J., Roh, J.H., Lee, J.H., Koh, J.Y., Na, D.L., 2016. Prediction of Alzheimer's disease pathophysiology based on cortical thickness patterns. Alzheimers Dement (Amst) 2, 58-67.

Jak, A.J., Preis, S.R., Beiser, A.S., Seshadri, S., Wolf, P.A., Bondi, M.W., Au, R., 2016. Neuropsychological Criteria for Mild Cognitive Impairment and Dementia Risk in the Framingham Heart Study. J Int Neuropsychol Soc 22, 937-943.

Jie, B., Zhang, D., Gao, W., Wang, Q., Wee, C., Shen, D., 2014. Integration of Network Topological and Connectivity Properties for Neuroimaging Classification. IEEE Transactions on Biomedical Engineering 61, 576-589.

Karas, G.B., Scheltens, P., Rombouts, S.A.R.B., Visser, P.J., van Schijndel, R.A., Fox, N.C., Barkhof, F., 2004. Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease. NeuroImage 23, 708-716.

Karikari, T.K., Pascoal, T.A., Ashton, N.J., Janelidze, S., Benedet, A.L., Rodriguez, J.L., Chamoun, M., Savard, M., Kang, M.S., Therriault, J., Schöll, M., Massarweh, G., Soucy, J.-P.,

Höglund, K., Brinkmalm, G., Mattsson, N., Palmqvist, S., Gauthier, S., Stomrud, E., Zetterberg, H., Hansson, O., Rosa-Neto, P., Blennow, K., 2020. Blood phosphorylated tau 181 as a biomarker for Alzheimer's disease: a diagnostic performance and prediction modelling study using data from four prospective cohorts. The Lancet Neurology 19, 422-433.

Khondoker, M., Dobson, R., Skirrow, C., Simmons, A., Stahl, D., 2016. A comparison of machine learning methods for classification using simulation with multiple real data examples from mental health studies. Statistical methods in medical research 25, 1804-1823.

Kim, J., Wozniak, J.R., Mueller, B.A., Pan, W., 2015. Testing group differences in brain functional connectivity: using correlations or partial correlations? Brain Connect 5, 214-231.

Klein, A., Tourville, J., 2012. 101 Labeled Brain Images and a Consistent Human Cortical Labeling Protocol. Frontiers in Neuroscience 6.

Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack, C.R., Jr., Ashburner, J., Frackowiak, R.S.J., 2008. Automatic classification of MR scans in Alzheimer's disease. Brain : a journal of neurology 131, 681-689.

LaMontagne, P.J., Benzinger, T.L.S., Morris, J.C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A., Raichle, M.E., Cruchaga, C., Marcus, D., 2019. OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease. medRxiv, 2019.2012.2013.19014902.

Leandrou, S., Petroudi, S., Kyriacou, P.A., Reyes-Aldasoro, C.C., Pattichis, C.S., 2018. Quantitative MRI Brain Studies in Mild Cognitive Impairment and Alzheimer's Disease: A Methodological Review. IEEE Rev Biomed Eng 11, 97-111.

Leuthardt, E.C., Guzman, G., Bandt, S.K., Hacker, C., Vellimana, A.K., Limbrick, D., Milchenko, M., Lamontagne, P., Speidel, B., Roland, J., Miller-Thomas, M., Snyder, A.Z., Marcus, D., Shimony, J., Benzinger, T.L.S., 2018. Integration of resting state functional MRI into clinical practice - A large single institution experience. PLoS One 13, e0198349.

Li, R., Yu, J., Zhang, S., Bao, F., Wang, P., Huang, X., Li, J., 2013. Bayesian network analysis reveals alterations to default mode network connectivity in individuals at risk for Alzheimer's disease. PLoS One 8, e82104.

Li, Y., Xiaoni Wang,Yongqiu Li, Yu Sun,Can Sheng,Hongyan Li,Xuanyu Li,Yang Yu,Guanqun Chen,Xiaochen Hu,Bin Jing,Defeng Wang,Kuncheng Li,Frank Jessen,Mingrui Xia, Ying Han, 2016. Abnormal Resting-State Functional Connectivity Strength in Mild Cognitive Impairment and Its Conversion to Alzheimer’s Disease. Neural Plasticity 2016.

Liu, J., Zhang, X., Yu, C., Duan, Y., Zhuo, J., Cui, Y., Liu, B., Li, K., Jiang, T., Liu, Y., 2016. Impaired Parahippocampus Connectivity in Mild Cognitive Impairment and Alzheimer's Disease. J Alzheimers Dis 49, 1051-1064.

Liu, X., Chen, K., Wu, T., Weidman, D., Lure, F., Li, J., 2018. Use of multimodality imaging and artificial intelligence for diagnosis and prognosis of early stages of Alzheimer's disease. Transl Res 194, 56-67.

Liu, X., Tosun, D., Weiner, M.W., Schuff, N., Alzheimer's Disease Neuroimaging, I., 2013. Locally linear embedding (LLE) for MRI based Alzheimer's disease classification. NeuroImage 83, 148-157.

Madsen, S.K., Ho, A.J., Hua, X., Saharan, P.S., Toga, A.W., Jack, C.R., Weiner, M.W., Thompson, P.M., 2010. 3D maps localize caudate nucleus atrophy in 400 Alzheimer's disease, mild cognitive impairment, and healthy elderly subjects. Neurobiology of Aging 31, 1312-1325.

Magnin, B., Mesrob, L., Kinkingnehun, S., Pelegrini-Issac, M., Colliot, O., Sarazin, M., Dubois, B., Lehericy, S., Benali, H., 2009. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology 51, 73-83.

Mathotaarachchi, S., Pascoal, T.A., Shin, M., Benedet, A.L., Kang, M.S., Beaudry, T., Fonov, V.S., Gauthier, S., Rosa-Neto, P., 2017. Identifying incipient dementia individuals using machine learning and amyloid imaging. Neurobiol Aging 59, 80-90.

Mattia, D., Babiloni, F., Romigi, A., Cincotti, F., Bianchi, L., Sperli, F., Placidi, F., Bozzao, A., Giacomini, P., Floris, R., Grazia Marciani, M., 2003. Quantitative EEG and dynamic susceptibility contrast MRI in Alzheimer's disease: a correlative study. Clin Neurophysiol 114, 1210-1216.

McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Jr., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., Mohs, R.C., Morris, J.C., Rossor, M.N., Scheltens, P., Carrillo, M.C., Thies, B., Weintraub, S., Phelps, C.H., 2011. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 7, 263-269.

Metzger, F.G., Schopp, B., Haeussinger, F.B., Dehnen, K., Synofzik, M., Fallgatter, A.J., Ehlis, A.C., 2016. Brain activation in frontotemporal and Alzheimer's dementia: a functional near-infrared spectroscopy study. Alzheimers Res Ther 8, 56.

Mitchell, A.J., Shiri-Feshki, M., 2009. Rate of progression of mild cognitive impairment to dementia--meta-analysis of 41 robust inception cohort studies. Acta Psychiatr Scand 119, 252-265.

Möller, C., Dieleman, N., van der Flier, W.M., Versteeg, A., Pijnenburg, Y., Scheltens, P., Barkhof, F., Vrenken, H., 2015. More atrophy of deep gray matter structures in frontotemporal dementia compared to Alzheimer's disease. J Alzheimers Dis 44, 635-647.

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., 2015. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. NeuroImage 104, 398-412.

Noh, Y., Jeon, S., Lee, J.M., Seo, S.W., Kim, G.H., Cho, H., Ye, B.S., Yoon, C.W., Kim, H.J., Chin, J., Park, K.H., Heilman, K.M., Na, D.L., 2014. Anatomical heterogeneity of Alzheimer disease: based on cortical thickness on MRIs. Neurology 83, 1936-1944.

Orban, P., Madjar, C., Savard, M., Dansereau, C., Tam, A., Das, S., Evans, A.C., Rosa-Neto, P., Breitner, J.C., Bellec, P., 2015. Test-retest resting-state fMRI in healthy elderly persons with a family history of Alzheimer's disease. Sci Data 2, 150043.

Petersen, R.C., Negash, S., 2008. Mild Cognitive Impairment: An Overview. CNS Spectrums 13, 45-53.

Plant, C., Teipel, S.J., Oswald, A., Böhm, C., Meindl, T., Mourao-Miranda, J., Bokde, A.W., Hampel, H., Ewers, M., 2010. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. NeuroImage 50, 162-174.

Pruim, R.H., Mennes, M., Buitelaar, J.K., Beckmann, C.F., 2015a. Evaluation of ICA-AROMA and alternative strategies for motion artifact removal in resting state fMRI. NeuroImage 112, 278-287.

Pruim, R.H., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J.K., Beckmann, C.F., 2015b. ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. NeuroImage 112, 267-277.

Querbes, O., Aubry, F., Pariente, J., Lotterie, J.-A., Démonet, J.-F., Duret, V., Puel, M., Berry, I., Fort, J.-C., Celsis, P., The Alzheimer's Disease Neuroimaging, I., 2009. Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. Brain : a journal of neurology 132, 2036-2047.

Ramírez, J., Górriz, J.M., Ortiz, A., Martínez-Murcia, F.J., Segovia, F., Salas-Gonzalez, D., Castillo-Barnes, D., Illán, I.A., Puntonet, C.G., 2018. Ensemble of random forests One vs. Rest classifiers for MCI and AD prediction using ANOVA cortical and subcortical feature selection and partial least squares. Journal of Neuroscience Methods 302, 47-57.

Rathore, S., Habes, M., Iftikhar, M.A., Shacklett, A., Davatzikos, C., 2017. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. NeuroImage 155, 530-548.

Riedel, B.C., Thompson, P.M., Brinton, R.D., 2016. Age, APOE and sex: Triad of risk of Alzheimer's disease. The Journal of steroid biochemistry and molecular biology 160, 134-147.

Sarica, A., Cerasa, A., Quattrone, A., 2017. Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. Front Aging Neurosci 9, 329.

Schmidt, M., 2006. L1precision: Matlab code for MAP estimation of Gaussian Graphical Model Precision with L1-regularizer.

Schouten, T.M., Koini, M., de Vos, F., Seiler, S., van der Grond, J., Lechner, A., Hafkemeijer, A., Moller, C., Schmidt, R., de Rooij, M., Rombouts, S., 2016. Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate Alzheimer's disease. Neuroimage Clin 11, 46-51.

Serra, L., Gabrielli, G.B., Tuzzi, E., Spanò, B., Giulietti, G., Failoni, V., Marra, C., Caltagirone, C., Koch, G., Cercignani, M., Bozzali, M., 2017. Damage to the Frontal Aslant Tract Accounts for Visuo-Constructive Deficits in Alzheimer's Disease. J Alzheimers Dis 60, 1015-1024.

Smith, S.M., Karla L. Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F. Beckmann, Thomas E. Nichols, Joseph D. Ramsey, Mark W. Woolrich, 2011. Network modelling methods for FMRI. NeuroImage 54, 875-891.

Sohn, W.S., Yoo, K., Na, D.L., Jeong, Y., 2014. Progressive changes in hippocampal resting-state connectivity across cognitive impairment: a cross-sectional study from normal to Alzheimer disease. Alzheimer Dis Assoc Disord 28, 239-246.

Specht, K., 2020. Current Challenges in Translational and Clinical fMRI and Future Directions. Frontiers in psychiatry 10, 924-924.

Stephen, J.M., Montaño, R., Donahue, C.H., Adair, J.C., Knoefel, J., Qualls, C., Hart, B., Ranken, D., Aine, C.J., 2010. Somatosensory responses in normal aging, mild cognitive impairment, and Alzheimer's disease. J Neural Transm (Vienna) 117, 217-225.

Stubbs, B., Thompson, T., Solmi, M., Vancampfort, D., Sergi, G., Luchini, C., Veronese, N., 2016. Is pain sensitivity altered in people with Alzheimer's disease? A systematic review and meta-analysis of experimental pain research. Experimental Gerontology 82, 30-38.

Tam, A., Dansereau, C., Badhwar, A., Orban, P., Belleville, S., Chertkow, H., Dagher, A., Hanganu, A., Monchi, O., Rosa-Neto, P., Shmuel, A., Wang, S., Breitner, J., Bellec, P., 2015. Common Effects of Amnestic Mild Cognitive Impairment on Resting-State Connectivity Across Four Independent Studies. Front Aging Neurosci 7, 242.

Tam, A., Dansereau, C., Iturria-Medina, Y., Urchs, S., Orban, P., Sharmarke, H., Breitner, J., Bellec, P., 2019. A highly predictive signature of cognition and brain atrophy for progression to Alzheimer's dementia. Gigascience 8.

Tampellini, D., 2015. Synaptic activity and Alzheimer's disease: a critical update. Frontiers in Neuroscience 9.

Tudela, R., Muñoz-Moreno, E., Sala-Llonch, R., López-Gil, X., Soria, G., 2019. Resting State Networks in the TgF344-AD Rat Model of Alzheimer's Disease Are Altered From Early Stages. Front Aging Neurosci 11, 213.

Vandenberghe, R., Nelissen, N., Salmon, E., Ivanoiu, A., Hasselbalch, S., Andersen, A., Korner, A., Minthon, L., Brooks, D.J., Van Laere, K., Dupont, P., 2013. Binary classification of (1)(8)F-flutemetamol PET using machine learning: comparison with visual reads and structural MRI. NeuroImage 64, 517-525.

Vemuri, P., Wiste, H.J., Weigand, S.D., Shaw, L.M., Trojanowski, J.Q., Weiner, M.W., Knopman, D.S., Petersen, R.C., Jack, C.R., 2009. MRI and CSF biomarkers in normal, MCI, and AD subjects. Neurology 73, 294.

Verghese, J., Wang, C., Lipton, R.B., Holtzer, R., Xue, X., 2007. Quantitative gait dysfunction and risk of cognitive decline and dementia. J Neurol Neurosurg Psychiatry 78, 929-935.

von Luxburg, U., 2006. A Tutorial on Spectral Clustering. Max Planck Institute for Biological Cybernetics Technical Report No. TR-149.

Wagner, F., Duering, M., Gesierich, B.G., Enzinger, C., Ropele, S., Dal-Bianco, P., Mayer, F., Schmidt, R., Koini, M., 2020. Gray Matter Covariance Networks as Classifiers and Predictors of Cognitive Function in Alzheimer's Disease. Frontiers in psychiatry 11, 360.

Wang, B., A Mezlini, F Demir, M Fiume, T Zu, M Brudno, B Haibe-Kains, A Goldenberg, 2014. Similarity Network Fusion: a fast and effective method to aggregate multiple data types on a genome wide scale. . Nature Methods Online.

Wang, J., Liu, J., Wang, Z., Sun, P., Li, K., Liang, P., 2019. Dysfunctional interactions between the default mode network and the dorsal attention network in subtypes of amnestic mild cognitive impairment. Aging (Albany NY) 11, 9147-9166.

Wee, C.Y., Yap, P.T., Shen, D., 2013. Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. Hum Brain Mapp 34, 3411-3425.

Weiler, M., Fukuda, A., Massabki, L.H., Lopes, T.M., Franco, A.R., Damasceno, B.P., Cendes, F., Balthazar, M.L., 2014. Default mode, executive function, and language functional connectivity networks are compromised in mild Alzheimer's disease. Curr Alzheimer Res 11, 274-282.

Yamasaki, T., Aso, T., Kaseda, Y., Mimori, Y., Doi, H., Matsuoka, N., Takamiya, N., Torii, T., Takahashi, T., Ohshita, T., Yamashita, H., Doi, H., Inamizu, S., Chatani, H., Tobimatsu, S., 2019. Decreased stimulus-driven connectivity of the primary visual cortex during visual motion

stimulation in amnestic mild cognitive impairment: An fMRI study. Neuroscience Letters 711, 134402.

Yang, J., Zhu, Z., He, S., Ji, Z., 2013. Minimal-redundancy-maximal-relevance feature selection using different relevance measures for omics data classification. 2013 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), pp. 246-251.

Ye, J., Farnum, M., Yang, E., Verbeeck, R., Lobanov, V., Raghavan, N., Novak, G., DiBernardo, A., Narayan, V.A., for the Alzheimer's Disease Neuroimaging, I., 2012. Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data. BMC Neurology 12, 46.

Yen, S.-J., Lee, Y.-S., 2009. Cluster-based under-sampling approaches for imbalanced data distributions. Expert Systems with Applications 36, 5718-5727.

Yu, M., Linn, K.A., Cook, P.A., Phillips, M.L., McInnis, M., Fava, M., Trivedi, M.H., Weissman, M.M., Shinohara, R.T., Sheline, Y.I., 2018. Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. Hum Brain Mapp 39, 4213-4227.

Yuan, Q., Zou, L., Chen, Q., 2008. [Voxel-based morphometric study of brain structure in Alzheimer's disease]. Sichuan Da Xue Xue Bao Yi Xue Ban 39, 496-499.

Zelnik-Manor, L., P Perona, 2004. Self-Tuning Spectral Clustering. Neural Information Processing Systems Conference paper.

Zhang, D., Shen, D., Alzheimer's Disease Neuroimaging, I., 2012. Predicting Future Clinical Changes of MCI Patients Using Longitudinal and Multimodal Biomarkers. PLoS One 7, e33182.

Zhang, H.Y., Wang, S.J., Xing, J., Liu, B., Ma, Z.L., Yang, M., Zhang, Z.J., Teng, G.J., 2009. Detection of PCC functional connectivity characteristics in resting-state fMRI in mild Alzheimer's disease. Behav Brain Res 197, 103-108.

Zhou, B., Yao, H., Wang, P., Zhang, Z., Zhan, Y., Ma, J., Xu, K., Wang, L., An, N., Liu, Y., Zhang, X., 2015. Aberrant Functional Connectivity Architecture in Alzheimer's Disease and Mild Cognitive Impairment: A Whole-Brain, Data-Driven Analysis. Biomed Res Int 2015, 495375.

Thesis Discussion

In this section, we present some additional points of discussion as they relate to one or more of the three manuscripts in the thesis. We divide the discussion into three themes that recur in the three manuscripts of the thesis: the definition of network ROIs, the importance of computational methods, and considerations around sample size. At the end of this section, we will summarize the main findings of the overall thesis and the significance of these findings.

The Definition of Network ROIs

Of utmost importance when studying FC networks is the selection of the network of interest as well as the definition of ROIs within the network. There are pros and cons to choosing larger versus smaller networks, and likewise for choosing larger or smaller ROIs. What is well known, however, is that careful definition of the boundaries between functionally distinct brain regions is critical (Smith et al., 2011). If there is signal mixing from functionally distinct regions due to poor ROI boundary definition, meaningful connectivity information can be lost. While in each of the 3 presented studies we have already motivated our choice of network, here we will discuss additional considerations in terms of the definition of the network ROIs which may have an impact on computed results.

The effect of external inputs

It is possible that communication from nodes outside of the network of interest – such as the visual cortex in manuscripts 1 and 2 – may play a role in the functional connectivity we observe with the network in our analyses. This is a phenomenon that was briefly introduced in the thesis introduction. While methods like partial correlation or ICOV remove network effects with their multiple regressions, effects that are not widespread throughout the network and originate from outside the network are not likely to be accounted for. This means that FC between two nodes may be spuriously high or reflect indirect connectivity through external nodes rather than reflect the direct within-network connectivity we are interested in identifying.

If we strive for completeness, we should be sure to include any and all brain regions which could potentially have this type of limited communication with the network through only a small proportion of the nodes. That being said, our goal in manuscript 2 is to approximate anatomical connections, not capture the full extent and content of the functional relationships of the regions in our networks. Thus, if the model is capable of successfully approximating anatomical connectivity without the use of time-courses from outside of the defined network, then we do not need to consider a larger network. Given the complementary nature of the information in areas V1, V2, and V3 and the subregions we define based on eccentricity, our assumption is that external interaction with the early visual cortex will be quite widespread (Felleman and Van Essen, 1991; Angelucci et al., 2002; Salin and Bullier, 1995). If this is indeed the case, then these more "global" effects from external inputs should be picked up in the network effects and/or the LPC. Given these considerations, we feel that not including regions outside of the visual cortex in our network is unlikely to be problematic with respect to external inputs. In the more general case, analyzing a local network as we do in manuscripts 1 and 2 assumes that external input to the network is not confined to only 2 of the defined ROIs.

While discussing the effect of external inputs, it is also relevant to discuss Partial correlation's network effects as a model of the global signal. Another way of viewing the regression procedure carried out in Partial correlation or ICOV is that it removes so-called "network effects" from the computed correlation (Chen et al., 2018). One commonly studied and utilized model of network effects is that of the global signal. The global signal is typically computed by averaging the signal across all gray matter voxels in the brain, capturing several global components including the common input to all voxels. Thus, upon regressing the global average signal from the time-courses of the voxels in the network, the remaining time-courses do not reflect common responses that influence the entire brain/network. If the set of nodes considered in a partial correlation analysis captures the entire brain, then computing the partial correlation is one possible model for how to remove the global signal, as the signal from all brain regions will be considered and removed before computing the connectivity (Liang et al., 2012; Chen at al., 2013). If the network considered in the partial correlation analysis is not complete, i.e. there are other sources of information flow into the network that are not accounted for, then the removal of the effect of all nodes in the network is more appropriately called "network effects", rather than "global signal". If the information flow between network nodes and external

nodes is relatively widespread in the network, it is likely the impact will be accounted for in partial correlation's multiple regressions, as discussed above (Marrelec et al., 2006). However, if there are more focused exchanges between network nodes and the external nodes, connectivity relationships within the network may be skewed and inaccurate in terms of reflecting direct connectivity, the connectivity analysis suffering in the same way correlation does with respect to indirect relationships (Marrelec et al., 2006). Thus, network completeness is an important consideration when planning for a partial correlation analysis. In our studies of the visual cortex, as mentioned in Manuscript 2, we expect that the influence from external nodes should indeed be widespread and we are therefore confident in our choice of network.

It is, however, an interesting question to consider: Does including more regions – external to the network of interest – help or hinder the network prediction (in terms of capturing anatomical connectivity)? Some of the methods we use are limited in terms of how many regions we can define in the network relative to the length of the time-courses of our data and may suffer with more regions being included (for example, partial correlation). There is also the issue of added noise to the system when one includes a greater number of nodes which would likely negatively impact the computed results of the multivariate models, as we demonstrate in manuscript 2. This phenomenon of increased noise may have been at play when we consider the results of our 60-node analyses as compared to the 15-node analyses in Manuscript 2: There are fewer statistically significant network connections and a lower proportion of TPs identified for multivariate models when more nodes are considered. Because of the potentially detrimental effects on some of our more promising models of study (e.g., partial correlation and ICOV) with a high number of network ROIs, we chose to limit our analyses to a local network, thereby limiting the total number of ROIs. In this way, we were able to evaluate whether the models could make successful network predictions, despite limitations in the extent of information contributing to the model in the form of ROIs.

<u>An incomplete ground truth network</u>

Next, rather than discussing the completeness of the set of ROIs in the network, we will discuss the completeness of the *a priori* information we have about the set of ROIs in the network. In Manuscript 2, we define our inferred ground truth set of connections within the network as a subset of the full extent of the connections that likely exist within the network. The

connections we have omitted are not as well understood as those we have included, and it is unclear if these connections are expected to be direct or indirect. The choice to study a network on which we can only evaluate a subset of the connectivity we compute may have consequences in terms of the overall TPR or FPR of a given method in that network. Perhaps the models that were successful in our study here, with few to no FPs, would have far more FPs if we could evaluate the full set of connections. That is to say; with the entire ground truth known, the results may indeed be different. This is certainly a limitation of our study; however, we do not see a more appropriate choice of network and ground truth in the context of real human imaging data. There are few systems in the human brain that are very well understood in terms of fine-scaled direct anatomical connectivity; among these systems, the visual cortex is the best understood, based not only on studies in humans but also in primates. Other regions outside of the visual cortex, which are well understood include the primary motor or auditory cortices. However, these brain regions still require exploration to understand the full extent of the direct connections within them and to understand how well animal models compare to human cortex (Häkkinen and Rinne, 2018; Card and Gharbawie 2020). There is in fact no network on a scale comparable to that of the early visual cortex, which is fully characterized in terms of anatomical linkages between neurons and clusters of neurons, based on primate studies. The Non-invasive functional MRI studies, and investigations conducted with human patients who suffer from lesions in the visual cortex show that the functional anatomy of the human and non-human primates are similar. Given this, and despite the limitations of our *a priori* knowledge of the early visual cortex, it is our hope that even with just a subset of the real anatomical connectivity in the network accounted for in our evaluations, we do identify models that succeed in approximating the underlying anatomy.

The HCP cortical parcellation

The HCP cortical parcellation (Glasser et al., 2016) is a relatively new approach to defining cortical areas using multiple modalities of imaging. The benefit of using a parcellation built using multiple modalities of imaging data is that the edges between the regions it defines are more likely to converge to a ground-truth set of highly functionally and structurally similar voxels representing meaningful cortical regions. This parcellation is also interesting to us in

particular because of the finer scale of the defined regions relative to other whole-brain parcellations. With 180 cortical regions defined in each hemisphere, a more fine-grained analysis of brain regions is possible relative to traditional brain atlases involving ~45 regions per hemisphere (e.g., Desikan-Killiany-Tourville, Automated Anatomical Labeling). The HCP parcellation was also built and validated using a relatively large number of subjects, contributing to the confidence one can have in the brain regions defined.

While we believe the HCP parcellation is an excellent tool, it is important to also mention some of its limitations given the context of our choice to apply it to MCI subjects. Firstly, the parcellation is built using brain imaging data from young healthy subjects and not from elderly subjects, which would have been a more appropriate standard given our clinical cohort. No multi-modal parcellation of a similar standard and scale built on elderly subjects exists, thus, we have chosen to continue with the HCP parcellation despite this imperfect fit. Another limitation of our implementation of the HCP cortical parcellation is that we have applied it as an atlas, registering the group average to each subject's T1 anatomical image. The HCP parcellation was conceived as a pipeline for computing a cohort-specific group parcellation, or ideally, a group-parcellation-informed subject-specific cortical parcellation. It is well known that inter-subject variability in the exact locations of cortical gyri and sulci exists. It is also well known that functionally defined regions do not necessarily overlie the same brain coordinates between individuals: There are typically millimetric variations in the position of boundaries of different subjects at a minimum. Given this, the ideal situation is indeed to have a subject-specific parcellation which seeks to retain the same number and general location of regions, but allows for flexibility in terms of exact location based on subject-specific data. It would have been ideal to apply the HCP subject-specific parcellation pipeline to our data; however, the code for this procedure was not available at the time of study. In order to minimize error from poorly aligned atlas regions to a specific subject's anatomy, we retained the voxels from the centre of the ROIs and avoided any that were in contact with the edges between regions. An interesting future direction for Manuscript 3 would be to run the study with the HCP pipeline-derived subject-and-cohort-specific parcellations and observe the differences relative to what we have computed using the HCP cortical parcellation as an atlas.

Additionally, it is important to mention that the HCP cortical parcellation is built using a surface-based analysis; thus, it is comprised of brain regions defined on the surface. In our case,

we have proceeded to project these surface-based regions into volume space so that we can use the regions to compute volume-based metrics such as gray matter volume, cortical thickness, and voxel-based mean ROI fMRI time-courses. While this parcellation is not optimized for volume based analysis, we remain confident in our use of the ROIs in volume space through quality control measures. Specifically, we accomplish this by visualizing the regions as they are defined in volume space, ensuring good alignment with the full extent of the cortical gray matter depth.

Finally, the relevance of the RS-fMRI information depends on good alignment between a subject's functional and anatomical scans as well as the alignment between the anatomical scan and the atlas. Given this, one might question the appropriateness of applying such a fine-scaled, relatively high-resolution parcellation to relatively low-quality fMRI scans. This is a limitation in all fMRI-based studies. While a perfect alignment between scans is likely not achieved, we strive for the best alignment possible, and expect that by eroding the edges of the defined ROIs, we can compensate for some of the possible misalignment. The field of imaging neuroscience proceeds with fMRI-based study broadly, despite the limitations in resolution and consequently alignment, though it is important to keep these factors in mind when interpreting results.

A whole-brain study of MCI

In Manuscripts 1 and 2 we focused our study on a local network of the brain: the early visual cortex. A natural extension of the first two studies into Manuscript 3 would have been to study local networks within the MCI population. One could also argue that a targeted approach to studying brain regions implicated in MCI rather than mining the whole-brain would be more insightful. Following a different school of thought, Badhwar et al. (2017), for example, recommend whole-brain analyses of MCI/AD to avoid the biases inherent in pre-selected regions. For our study, evidently, we decided against the study of local brain networks in MCI in favor of a whole-brain approach. There are a number of reasons for this choice which we will now discuss. It should be clear from Manuscript 3 that the scope of brain regions implicated in MCI and/or AD is vast. There are indeed a number of brain regions that are more frequently implicated in the literature than others (DMN, temporal lobe, etc.); however, it is clear that focusing only on these brain regions would not have been sufficient to capture the full extent of the regions implicated in MCI. In fact, had we taken this approach, we would not have observed the importance of the visual and sensorimotor cortices in differentiating our MCI subtypes, as

these are not among the most commonly implicated brain regions. It is also important to keep in mind that we study not-yet-well-defined MCI subtypes; thus, a data driven approach with no a priori assumptions or constraints seemed most appropriate. A future direction in the study of MCI subtypes could indeed involve a targeted study of local networks. However, we believe that a more thorough definition of the subtypes and their characteristic local networks as identified in Manuscript 3 will be important to achieve before such a pursuit.

Importance of Computational Methods

One of the main themes of the thesis is that of evaluating/optimizing computational methods. Through this pursuit, we have developed two novel approaches we recommend for further investigation/development: LPC Regression and our PPV-optimized multi-classifier machine learning approach. In addition to this, we have elucidated FC methods that are good choices for use in fine-scaled, densely connected networks of which no information exists. Another equally important methodological component to the work we have pursued is our statistical approach. Next, we discuss further some of our statistical choices, as well as other relevant considerations for our methodological approaches.

The repeated measures General Linear Model and statistical contrasting of functional connectivity measures to zero

In both Manuscripts 1 and 2 we make use of a repeated measures General Linear Model (GLM) for our statistical tests, and the contrast in question is that comparing our FC measures to zero. We have chosen to apply a repeated-measures GLM in order to avoid averaging our data, which is the common alternate approach when multiple measures are taken for individual subjects. A benefit of considering a repeated measures design is that we have an increase in statistical power relative to the case where we average within-subject measures. The overall goal in choosing a statistical approach is to make optimal and appropriate use of the power that is inherent in our data. Given the high number of repeated resting-state measurements per subject in our data (8.5 mins × 8 runs), our GLM approach also allows the opportunity to evaluate within-subject and between-subjects variability, something not inherent in a 2 sample t-test.

As mentioned above, we have chosen to contrast our FC measures with zero, as opposed to another common choice of comparator: a surrogate null distribution. In Manuscript 1, we had originally chosen to include a surrogate comparator in order to generate a customized estimate of the baseline of the correlation and partial correlation coefficients for our analyses. Following comments from reviewers during the publication process, we were convinced the surrogate was not necessary in this context. Based on their comments, we checked whether the mean correlation and partial correlation between "time-courses" of our two types of surrogates (time-shuffled surrogate and Fourier surrogate) were different from zero. The mean and standard deviation of the distribution of the time-shuffled surrogate data were: correlation mean=0.0074, std=0.0054, and partial correlation mean=5.9x10^-6, std=0.0049; and the mean and standard deviation of the distribution of the Fourier surrogate data were: correlation mean=-0.0076, std=0.0187, and partial correlation mean=0.0021, std=0.0157. The surrogate mean was approximately zero for both correlation and partial correlation and for both the time-shuffled and Fourier surrogates. Thus, the chosen FC measures showed no bias toward a positive or negative mean. These findings support the comments by the reviewers that the use of surrogate data is not strictly needed here. Therefore, we now include in the main manuscript functional connectivity measures compared to zero rather than surrogate data. However, in the supplementary material (Appendix A) we include the comparisons to surrogate data.

As mentioned, in Appendix A we include a version of our Manuscript 1 results where the FC measures were contrasted with each of the two tested surrogate distributions. To create the time-shuffled surrogate data, we shuffled the original data in time and computed the correlations and partial correlations for a given ROI pair between all combinations of non-simultaneous runs. Given the data we had from 8 runs, the maximal number of surrogate correlations/partial correlations we could produce per subject was 28 ([8runs x 8runs – 8 simultaneous runs]/2 = 28 non-simultaneous surrogate data sets). These sets of time-shuffled FC measures made up the time-shuffled surrogate distribution. For the Fourier surrogate, the ROI time-courses were first transformed to Fourier space, then the imaginary portions of the Fourier transformed data were randomized before computing the inverse of the Fourier transform. The correlations and partial correlations were then computed between the phase-randomized surrogate ROI time-courses, resulting in the Fourier surrogate distribution. Contrasting our real correlation and partial-correlation measures to the surrogate distributions contributes to the minimization of false-

positive connections in our predicted connectivity network. The comparison against both types of surrogate data showed very similar results to those we presented in the main manuscript, where we compared FC measures to zero.

Given the above analyses, we believe that it is appropriate to continue this statistical approach using the repeated measures GLM and a contrast of FC measures to zero, as we did in Manuscript 2.

Correction for multiple comparisons: False Discovery Rate

Another important decision related to statistical testing is the choice to identify the FDR corrected p-value threshold associated with a maximal false positive rate of 5% and using this as our threshold for statistical significance given multiple comparisons. There are a number of approaches for accounting for multiple comparisons in statistical testing, such as the Bonferroni correction, the Sidak correction, the Holm-Bonferroni correction, Holm's step-down procedure, etc. (Ludbrook, 1998). Most of these traditional methods are known to be quite strict, resulting in a large number of false negative results. Other methods – such as Gaussian random field corrections or Monte Carlo simulation based corrections (Chen et al., 2018) – have been developed more recently. These methods aim to apply less stringent correction procedures, though often at the cost of higher family-wise error rates (FWER). Other methods will utilize these less stringent methods, but only after the application of an omnibus test, showing that a global effect does exist before going into the detailed testing for effects (Keselman et al., 1999; Bellec et al., 2015). In any study, the choice of method for correction for multiple comparisons results in balancing a trade-off between discovery rate and FWER. In our studies, particularly in Manuscript 2, we prioritize the minimization of false positives, thus imposing a relatively low FWER, though of course we still strive to retain as many true positives as possible under this criterion. For this reason, we decided to use the FDR correction, which is shown in Chen at al. (2018) to have a low FWER in the context of RS-FMRI analysis, and is also designed to be more lenient in terms of significance criterion than methods such as the Bonferroni correction. Our results indicate that, for a number of the models we studied, this is a well-balanced choice of correction for multiple comparisons, with high TP rates and low FP rates.

An important consideration for the FDR correction is that the generated p-value threshold is dependent on the data. Bellec et al. (2015) examined this effect in detail in the context of

whole-brain RS-fMRI-based FC. They found that the discovery rate depends on the number of regions in the parcellation of the brain (i.e. network complexity). They found that sensitivity was greater with lower network complexity, and that sensitivity was generally at an acceptably high level in networks with fewer than approximately 50 regions. Given the variation observed at different network complexities when the FDR correction is used, they recommend selecting a single appropriate parcellation prior to statistical testing, and to avoid direct comparisons between tests with different network complexities. This is the approach we took in Manuscript 2 where we studied only 2 pre-selected network complexities, and we compare them only qualitatively, not directly and quantitatively.

In another study, Efron (2008) tested the hypothesis that the results of an analysis using FDR would be relatively invariant across network complexities if the connectivity within the network was strong. The study results supported this, indicating that under strong connectivity conditions, the FDR correction should be equally appropriate at any network complexity. Given the dense connectivity expected in the visual cortex network we studied in Manuscripts 1 and 2, we once again believe that the FDR is an appropriate method, likely to be minimally affected by network complexity.

Another encouraging finding in Bellec at al. (2015) was that, across testing on 3 real datasets, statistical maps at resolutions above 30 ROIs covering all of the brain's gray matter, the matched closely the maps generated with hundreds of parcels (findings also in line with Shehzad et al., 2014). This finding reinforced our expectation that FDR correction would be relatively unaffected by network complexity in our chosen network with real fMRI data from the visual cortex, despite the fact that we considered relatively low network complexities at 15 and 60. Bellec et al. suggested that there was a compromise to be made in implementing the FDR correction and the resolution of the connectivity information that can be studied. Their tests, however, only considered whole-brain parcellations, which would indeed require a low resolution to keep the number of regions at ~50 or less (to retain reasonable sensitivity) while still covering the whole brain. Our approach to this issue was to not compromise on resolution, but instead to compromise on the extent of the network: looking at a relatively low complexity network (60 regions or less) at a fine-scale (or, in other words, a relatively high resolution) in a localized, sub-whole-brain region of cortex. Given our strong and consistent findings between the 15 node and 60 node networks in manuscript 2, in particular in the HCP analysis with large

sample size, we are confident that our approach is appropriate, though we recognize its limitations.

Selecting the ICOV regularization parameter

The purpose of pursuing ICOV FC analyses as opposed to Partial correlation analyses is to take advantage of the regularization possible with ICOV, which results in a sparser set of connections within the predicted network. The level of sparseness achieved, however, is determined by the regularization parameter, Lambda. The Lambda that one should use, or in other words, the level of sparseness appropriate for the network of study, is analogous to determining the density of the network: one needs prior information in order to make this choice in an informed way. Our choice of Lambda for Manuscript 2 was inspired by the results of the Smith et al. (2011) study in which they tested a handful of Lambdas (5, 10, 20, 50, 100, and 200, all ×10^-3) and found that 5×10^-3 and 100×10^-3 were best across simulations. Upon reviewing their results, we saw that in most simulations, ICOV_5 and ICOV_100 performed very similarly, or ICOV_5 performed slightly better. The exception to this was in the simulation with a higher number of nodes (50), where ICOV_100 performed somewhat better. Given the overall trend with these values, we were inclined to proceed with ICOV_5 only. We also expected that in the context of analyses with real data, as opposed to simulated data as in Smith et al. (2011), the proportion of significant connections in our larger network may already be relatively low given the abundant noise that would enter into the computation via the multiple regressions inherent in the methodology. This expectation grew in part with the results from Manuscript 1, where Partial correlation saw a reduced proportion of significant connections in the 60 node network relative to the 15 node network.

Despite our relative confidence in our choice of the regularization parameter, we were curious to see whether a better choice in Lambda was possible, and also to observe the trends in ICOV true positive rate (TPR) and false-positive rate (FPR) at different levels of regularization. We performed several analyses in the same format as done in Manuscript 2 at various Lambdas and the results of these analyses are presented in *Appendix F.* The results confirm that ICOV_5, namely ICOV with a Lambda of 5x10^-3 is a good default choice, being a good performer across analytical conditions such as a different number of network nodes, LPC regressed, sample size, etc. Even in the analysis with 60 nodes, the best performance was achieved at Lambdas of 5 and

10 without additional p-value thresholding (slight improvement was achieved at Lambda=300 but with stricter p-value thresholding). This confirms that our choice of pursuing only ICOV_5 and not ICOV_100 as well (as might have been suggested by the results of the Smith et al. study) was appropriate. In Manuscript 2, we utilized the findings from Smith et al. (2011) to make our choice of Lambda. Making use of the Smith et al. study to inform our choice for Lambda uses prior knowledge of the ICOV method and its behavior under certain conditions, though not prior knowledge of our particular system of study. For future studies of unknown networks on a fine-scale, the analyses in *Appendix F* could inform the appropriate choice of Lambda.

There are other methods that have been developed to select the regularization parameter in a data-driven way, looking to optimize statistical parameters associated with computations across a range of Lambdas. In essence, these methods perform a similar analysis across Lambdas as that done in *Appendix F*, however, the deciding factor for which Lambda is best to move forward with is not based on a comparison to the ground truth but rather through optimization of a pre-selected statistical score. Examples of these include penalized likelihood-function based methods (Fan and Li, 2001; Huang et al., 2006), Mean Squared Error based criteria (Schafer and Strimmer, 2005), the Aikeke Information Criterion (Buhlmann, 2006; Zou et al., 2007), the Bayesian Information Criterion, covariance penalty methods, and cross-validation methods (Zou et al., 2007; Peng at al., 2009). In our case, we chose not to proceed with such Lambda-selecting methods in order to minimize computational time associated with ICOV, as increased computational time is one of the costs associated with these parameter optimization procedures. Given that we were interested in a variety of methods, we chose to dedicate our computational time to more methods rather than optimization within one method. Given the success of ICOV_5 in the study presented in Manuscript 2, an interesting next-step study could indeed explore different regularization parameter optimization strategies, evaluating their performances in terms of predicting a ground truth network, as done in Manuscript 2.

Testing the machine learning pipeline on data from sites not represented in training data

A possible shortcoming that one might identify in the testing of our multi-classifier pipeline is that the sites of data acquisition of testing data are all represented in the training data. This choice is potentially inflating the perceived generalizability of the results, and, given that the condition is not tested, there is certainly the possibility that the pipeline will fail on data from

new sites. We made this design choice out of necessity, given that there is a well-known issue with combining MRI-based data from multiple sites without performing a correction to take this into account. The method of correction for the site of acquisition that we used is called ComBat Harmonization and the correction that is computed can only be applied to data from sites represented in the set used to compute the correction. This method adjusts the magnitude of features from subjects from the same site in a similar way and differently for subjects from a different site. Without the representation of a test subject's site in the training data, it would not be possible to know which adjustment to apply to its data. So, the question still remains: Is it possible to apply our multi-classifier approach to testing data from a site not accounted for in the training data? We expect that our findings are likely to be relatively generalizable given that we make use of data from many sites (20), whereas many studies use data from only one or two sites of acquisition, thereby dramatically limiting the variability they are accounting for due to site.

Our pipeline should still be readily applicable to data from other sites of acquisition, so long as a handful of datasets is available to be used to compute the ComBat correction. Once computed, this correction can be applied to the testing data from the new site. Alternatively, this handfuls of datasets from the new site can be used as training data. In this case, to test data from a new site, one would simply retrain the pipeline, including the data of the new training subjects from the new site, and then we would be ready to classify test subjects from that new site (as the ComBat step is built into the pipeline). The addition of the extra training subjects is likely to make the pipeline more robust so the addition of data from more sites is a valuable future pursuit.

Sample Size

The final category of topics we will discuss is that of topics relating to sample size or abundance of data. At a variety of levels in a computational analysis, the factor of data abundance can come into play. Whether this is as it relates to study design and number of subjects to include in order to achieve the desired level of statistical power, the quantity, and the coarseness of the parcellation of the imaging data, or the number of iterations of analyses needed to summarize the distribution of the results, the sample size is a critical point to consider and reflect upon when interpreting experimental results.

Small sample size in Manuscripts 1 and 2

In Manuscripts 1 and 2, the data analysed comes from 6 healthy, young, adult subjects. Ideally, when striving to make generalizable study conclusions, a large sample size with balanced relevant demographics should be used. With a sample comprised of data from only 6 subjects, one might question the generalizability of the presented results. In the case of our analyses of the fine-scaled connectivity in the visual cortex, however, we feel confident that our results are indeed generalizable. The connectivity of the early visual cortex has been widely studied in animal models such as in the macaque monkey and in cats (Hubel and Weisel 1962, 1968). In these animal models, the the anatomical and functional connectivity in early visual areas has been shown to be consistent across indivduals and studies. As previously introduced, the literature shows that the lower visual areas in humans are highly analogous to those in the monkey, thus the connectivy therein should also be highly consistent across individuals. With this in mind, and given the particular information we are seeking to extract from our data set, the risk of biases due to individual subject differences is minimal. In Manuscript 2, we also validate the results generated using the data from our 6 subjects by repeating the analyses on a data set with a larger sample size and similar demographics (young healthy adults). The analyses with large sample size confirmed the validity of the results with a smaller sample size, while also emphasizing the benefits of a larger sample size in the network of study: stronger effect sizes more easily captured via statistical testing (i.e., greater statistical power).

Also important to note is that while we only had data from 6 subjects in Manuscripts 1 and 2, recall that each subject underwent 8 runs (each 8.5minutes long) of resting-state scanning. These repeated measures contributed to reinforcing the observable effects in the data, resulting in relatively abundant data (48 resting-state fMRI runs), despite few subjects.

Recall that these studies are done in young, healthy adults. Given this, we only expect the generalizability of our results to extend to individuals who fall within this demographic. Any individuals with visual or cortical pathology may show differences in connectivity relative to our findings and study of those specific populations would be required to characterize them.

Improved network predictions with high-quality HCP data

In Manuscript 2, we observed that with a larger data set (the HCP data set) statistical testing seeking direct connections with the LPC regressed converged to the inferred ground truth.

We have largely attributed this highly successful performance to the increased sample size: 220 subjects relative to 6. It is possible, however, that there are other features of the HCP data which contributed to the relative success we see. The main differences between our data with 6 subjects and the HCP data we use are the length of the RS-fMRI time courses and the repetition time of the scans (TR). The TR represents the time between consecutive scans or between each time point in the time courses. For our data, the TR was 2s, and for the HCP data it was 0.7060s. One might expect that this shorter TR contributed to the time-courses better reflecting the fast neuronal exchanges between local, fine-scaled ROIs. Given the low-pass filtering feature of the BOLD signal relative to neuronal activity, we expect that this is not the case. The shorter TR most certainly contributed to the possible length of the time-courses while still keeping the duration of data acquisition relatively short for the subject. Our RS-fMRI scans were 8.5 minutes long, and given a TR of 2s, this provided us with 256 time points per scan. For the HCP data, there are 1200 time-points within a time course, and with TR of 0.7060s, this leads to about 14 minutes of scanning. We previously showed that the performance of most network models improves with longer time-courses (Dawson et al., 2013). Besides this intuitive increase in statistical strength with more data, some of the methods are designed in such a way that longer time courses are needed in order to make the data robust against the various manipulations of it. For example, partial correlation and ICOV see improved performance when the time-courses are longer relative to the number of ROIs as this allows more degrees of freedom despite multiple regressions.

Recall that we also tested a random sampling of 6 HCP subjects to see how these compared to our 6 subjects and the results were comparable. This strongly indicates that the main effect at play is that of increased sample size. Additionally, models like correlation and Kappa also converged toward the inferred ground truth with the HCP data and LPC regression, and these methods do not depend on long time-courses to retain a high number of degrees of freedom. Given all of these findings, we believe the highly successful performances of the models tested with the larger sample of subjects from the HCP data repository are predominantly due to the number of data sets, rather than the length of the time-courses or the smaller TR.

Power and sample size in the identified MCI subtypes

When designing an experiment in which statistical testing will be pursued, it is important to consider what sample size is appropriate in order to pick up the effect size that is expected to

be present in your data sample. Ideally, this sample size is decided upon before the experiment is begun. However, it is often the case that the effect size is not well known and data is limited such that pursuing an ideal sample size is not possible. In such cases, one can proceed with the experiment and observe the effect size that emerges and then compute the power associated with the available sample size. In neuroimaging studies, this is often the approach taken as the sample sizes tend to be quite small. In our MCI study, we made every effort to increase our sample size for the benefit of the machine learning algorithm and, of course, to optimize the statistical power. We proceeded with as much data as we could accumulate and sought to identify the effect sizes between previously undefined subtypes of MCI individuals. We found that the mean effect size (between subtype difference in mean feature magnitudes) for the HCP ROI-based GMV&CT analyses (GMV and CT measures normalized to a range from 0 to 1 within a subject) was 0.3788 (mean standard deviation, std, of 0.2255), which resulted in 100% power for even our smallest subtype with 21 subjects. In reality, we have multiple scans per subject, so the amount of data within a subtype is in fact, much larger than the number of subjects quoted, and thus we have more than enough data to pick up the effect sizes present between subtypes. The mean effect size for HCP ROI-based FC analyses (between subtype difference in mean correlation magnitudes) was 0.3635 (mean std 0.3027), once again with 100% power. For the DKT ROI-based FC analyses, the mean effect size was 0.3599 (mean std 0.2748), and the power was 100%. For the DKT ROI-based GMV&CT analyses, recall that we have one stable subtype with only one subject, and the other subtype contains all remaining subjects. Because of this poorly separated distribution, we will not comment further on the power or sample size of the subtypes associated with DKT ROI-based GMV&CT analyses.

Based on our strong findings with a power of 100% in all subtypes, we conclude that for the effect sizes observed between the subtypes defined in our MCI study, our sample size is more than adequate. That being said, for the machine learning algorithm to perform well, more data is better, so we would not limit the amount of data used to the minimum required. Our machine learning pipeline still has room for improvement with respect to correctly classifying subjects (improving progressor classification and limiting the indeterminate rate), something that could potentially be helped by even more data.

Receiver Operating Characteristic curves for MCI pipeline outcomes

One common way of presenting the results of machine learning pipelines, in particular when seeking to demonstrate parameter optimization, is through a receiver operating characteristic (ROC) curve. This method plots the false positive rate (FPR) on the horizontal axis and the true positive rate (TPR) on the vertical axis, which allows the reader to clearly see the trade-off between these two results across iterations of classification. From this plot, one can compute the corresponding area under the curve (AUC) which quantifies the overall success of the model and the generalizability of the results across iterations. An AUC close to 1 implies very high TPR with very low FPR in all cases, whereas an AUC close to 0 implies low TPR and high FPR. The most common use of these curves is when evaluating a parameter selection: the value or range which gives a desirable balance between TPs and FPs is identified and selected. In our case, parameters are selected and optimized within the pipeline with class-wise PPV; thus the ROC curve is not required for that use. Also, in our case, the ROC curve can be useful in visualizing the generalizability of the pipeline across different partitions of the data into training and testing sets, i.e. across our outer loop folds (see Figure 3 below). As noted above, a well-balanced model with high TPR and low FPR should have an AUC closer to 1. In Figure 3, we see that when the HCP parcellation is used (results across all selected high-PPV analyses for the 5 outer-loop iterations), the AUC is 0.88, whereas when the DKT parcellation is used it is 0.77. This reflects the finding (presented in Manuscript 3) that the HCP-based classification was more successful than the DKT-based classification overall, achieving on average a larger number of TPs across iterations of the outer loop.

We have decided to present these results here rather than in the manuscript as our sampling of 5 outer loop iterations is fairly small for such a plot. With 5 partitions of the data only, we are likely not capturing the full distribution of TPRs and FPRs of our pipeline, however, the consistency of the results suggests that these results may be representative. Despite that, for a fully informed ROC curve, the more data, the better, and so on the order of 50+ iterations would likely be more fitting. A high number of outer loop iterations was not an option for this particular study due to computational time considerations. Also, typically an ROC curve is obtained by varying a parameter such as a threshold. This makes it possible to obtain data points along all sections of the horizontal axis, and an optimally fitted function. In our case, the fitting is based on essentially one data point, with variability due to cross validation. In any case, given

that we present our pipeline as a concept to be further explored, these 5 representative data partitions are adequate to encourage more exploration and testing. A future direction of the study of our pipeline would include testing on a larger sample size and including a greater number of iterations in the outer loop cross-validation.

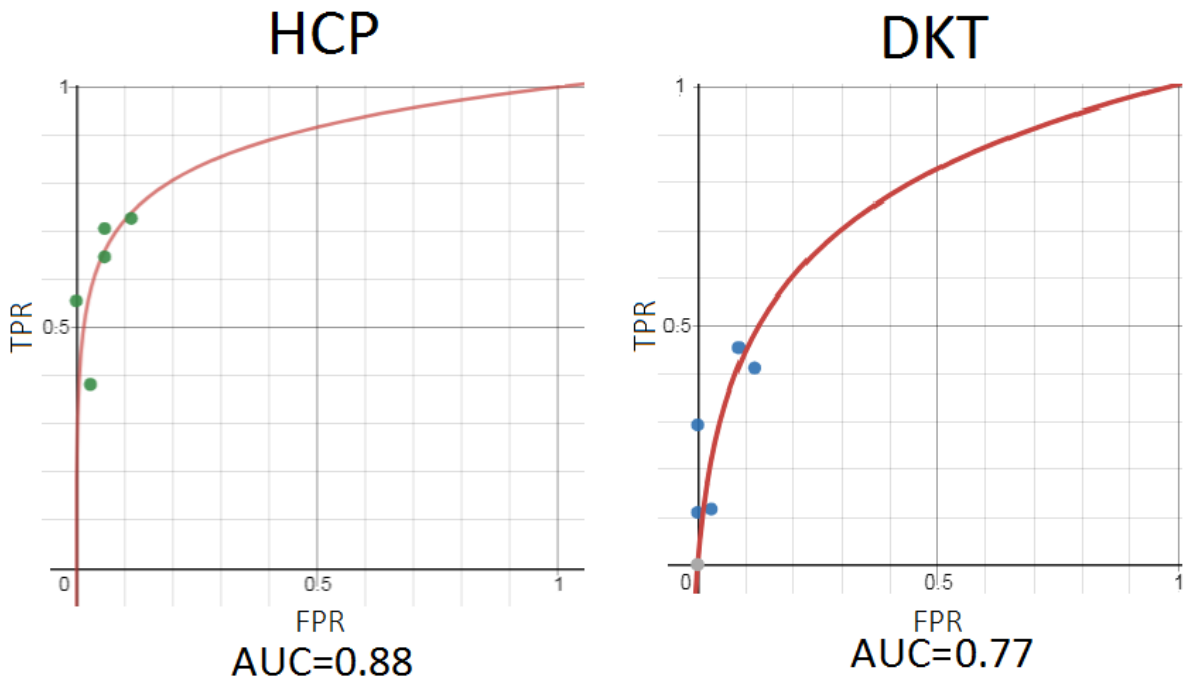


Figure 3. Receiver Operating Characteristic curves for each of the HCP (left plot) and DKT (right plot) cortical parcellation outer-loop iterations. The plots reflect the results from the set of stable-optimized classifiers (best-fit logarithmic curve is plotted). In this case, a TP (FP) result is a stable (progressor) test subject being classified into the stable-class of interest of one or more of the high-PPV classifiers. TPR=true positive rate; FPR=false positive rate; AUC=area under the curve.

Conclusions/Summary

Work in this thesis has been done with the goal of advancing knowledge regarding the information we can gather from the resting-state fMRI signal. We focus on the study of fine-scaled functional connectivity, first analyzing the visual system in healthy individuals, second evaluating methods in terms of their ability to extract direct neuronal connectivity between brain regions, and third applying the more promising methods in a whole-brain study of patients with mild cognitive impairment. The results of this work clarify the potential of various computational models in the contexts of healthy and pathological states and in multiple regions of the brain. The results also contribute to the growing knowledge of the functional connectivity in the studied brain regions, leading to a better understanding of the functioning of the brain as well as pushing the boundaries of the ways we utilize resting-state fMRI data.

Specific results from this thesis include the following:

- RS fMRI carries information that can be extracted with computational methods and which identifies direct connections in a fine-scaled, densely connected network of which the user has no prior information.
- Partial correlation can be used to estimate the direct connectivity in a fine-scaled study of the human visual cortex, and this measure captures information beyond simply cortical distance.
- Regression of the LPC from network time-courses is beneficial to many methods in terms of their ability to estimate direct connectivity in a fine-scale network.
- Using partial correlation or correlation combined with LPC regression in a fine-scaled, densely connected network and lenient *statistical thresholding* (α=0.05) allows one to converge toward the ground truth direct connectivity of the network.
- ICOV-based whole-brain direct connectivity is not an important predictor of progression from MCI to dementia.
- Finer-scaled parcellations may allow for correct classification of a greater number of subjects than larger scaled parcellations in the context of machine learning-based classification of the prognosis of individuals with MCI.

- Subtypes in the MCI population appear to differ depending on the type of feature used to define the subtypes. The brain regions most important for differentiation between these subtypes include areas in the occipital, medial parietal, and sensorimotor cortices.
- A PPV-optimized multi-classifier approach to classifying a heterogeneous population into subtypes shows promise for confident predictions on a subset of the population.

Throughout these studies, a focus is on evaluating and proposing methodological approaches which we optimize for how closely they can reflect reality. The methods we have developed and propose include the LPC, and a PPV-optimized multi-classifier machine learning classification pipeline. Better methods are an important tool to help in the meaningful interpretation of experimental measurements. We, therefore, believe that our work is important and a positive contribution to the field of imaging neuroscience. Specifically, our work contributes to the identification of an accurate non-invasive measure of fine-scaled anatomical connectivity, something as of yet not achievable. We also contribute to a better understanding of the important features in distinguishing individuals with MCI. This may allow for better prognostication and eventually may lead to the development of better treatments through clarification of the mechanism underlying this pathological state. Finally, our set of machine learning analyses optimizing PPV can contribute to future identification of individuals for whom we can make confident prognosis, allowing to optimize care from an earlier time frame and optimize patient selection for clinical trials. There is still much to learn in terms of the scope of application of fMRI measures in the study of local and whole-brain networks, as well as in the application of multi-classifier and precision-optimized machine learning approaches to identify subtypes in a heterogeneous population. We look forward to contributing to the body of knowledge yet to come.

References

Allen, E.A., Damaraju, E., Plis, S.M., Erhardt, E.B., Eichele, T., Calhoun, V.D., 2014. Tracking whole-brain connectivity dynamics in the resting state. Cereb Cortex 24, 663-676.

Allen, E.A., Erhardt, E.B., Damaraju, E., Gruner, W., Segall, J.M., Silva, R.F., Havlicek, M., Rachakonda, S., Fries, J., Kalyanam, R., Michael, A.M., Caprihan, A., Turner, J.A., Eichele, T., Adelsheim, S., Bryan, A.D., Bustillo, J., Clark, V.P., Feldstein Ewing, S.W., Filbey, F., Ford, C.C., Hutchison, K., Jung, R.E., Kiehl, K.A., Kodituwakku, P., Komesu, Y.M., Mayer, A.R., Pearlson, G.D., Phillips, J.P., Sadek, J.R., Stevens, M., Teuscher, U., Thoma, R.J., Calhoun, V.D., 2011. A baseline for the multivariate comparison of resting-state networks. Front Syst Neurosci 5, 2.

Angelucci, A., Levitt, J.B., Walton, E.J., Hupe, J.M., Bullier, J., Lund, J.S., 2002. Circuits for local and global signal integration in primary visual cortex. J Neurosci 22, 8633-8646.

Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D., 2017. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. NeuroImage 145, 137-165.

Arfanakis, K., Cordes D., Haughton V.M., Moritz C.H., Quigley M.A., Meyerand M.E., 2000. Combining independent component analysis and correlation analysis to probe interregional connectivity in fMRI task activation datasets. Magn. Reson. Imaging 18, 921 – 930.

Badhwar, A., Tam, A., Dansereau, C., Orban, P., Hoffstaedter, F., Bellec, P., 2017. Resting-state network dysfunction in Alzheimer's disease: A systematic review and meta-analysis. Alzheimers Dement (Amst) 8, 73-85.

Bai, F., Shi, Y., Yuan, Y., Wang, Y., Yue, C., Teng, Y., Wu, D., Zhang, Z., Jia, J., 2012. Altered self-referential network in resting-state amnestic type mild cognitive impairment. Cortex 48, 604-613.

Balachandar, R., John, J.P., Saini, J., Kumar, K.J., Joshi, H., Sadanand, S., Aiyappan, S., Sivakumar, P.T., Loganathan, S., Varghese, M., Bharath, S., 2015. A study of structural and functional connectivity in early Alzheimer's disease using rest fMRI and diffusion tensor imaging. Int J Geriatr Psychiatry 30, 497-504.

Banerjee, O., El Ghaoui, L., d'Aspremont, A., Natsoulis, G., 2006. Convex optimization techniques for fitting sparse Gaussian graphical models. Proceedings of the 23rd International Conference on Machine Learning, 96.

Beckmann, C.F., DeLuca, M., Devlin, J.T., Smith, S.M., 2005. Investigations into resting-state connectivity using independent component analysis. Philos Trans R Soc Lond B Biol Sci 360, 1001-1013.

Bellec, P., Benhajali, Y., Carbonell, F., Dansereau, C., Albouy, G., Pelland, M., Craddock, C., Collignon, O., Doyon, J., Stip, E., Orban, P., 2015. Impact of the resolution of brain parcels on connectome-wide association studies in fMRI. NeuroImage 123, 212-228.

Binnewijzend, M.A., Schoonheim, M.M., Sanz-Arigita, E., Wink, A.M., van der Flier, W.M., Tolboom, N., Adriaanse, S.M., Damoiseaux, J.S., Scheltens, P., van Berckel, B.N., Barkhof, F., 2012. Resting-state fMRI changes in Alzheimer's disease and mild cognitive impairment. Neurobiol Aging 33, 2018-2028.

Biswal, B., F. Zerrin Yetkin, Victor M. Haughton, James S. Hyde, 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. Magnetic Resonance in Medicine 34, 537-541.

Boly, M., Phillips, C., Tshibanda, L., Vanhaudenhuyse, A., Schabus, M., Dang-Vu, T.T., Moonen, G., Hustinx, R., Maquet, P., Laureys, S., 2008. Intrinsic brain activity in altered states of consciousness: how conscious is the default mode of brain function? Ann N Y Acad Sci 1129, 119-129.

Buckner, R.L., Krienen, F.M., Castellanos, A., Diaz, J.C., Yeo, B.T., 2011. The organization of the human cerebellum estimated by intrinsic functional connectivity. J Neurophysiol 106, 2322-2345.

Cai, S., Chong, T., Zhang, Y., Li, J., von Deneen, K.M., Ren, J., Dong, M., Huang, L., 2015a. Altered Functional Connectivity of Fusiform Gyrus in Subjects with Amnestic Mild Cognitive Impairment: A Resting-State fMRI Study. Frontiers in Human Neuroscience 9, 471.

Cai, S., Huang, L., Zou, J., Jing, L., Zhai, B., Ji, G., von Deneen, K.M., Ren, J., Ren, A., 2015b. Changes in thalamic connectivity in the early and late stages of amnestic mild cognitive impairment: a resting-state functional magnetic resonance study from ADNI. PLoS ONE [Electronic Resource] 10, e0115573.

Carbonell, F., Bellec P, Shmuel A, 2011. Global and System-Specific Resting-State fMRI Fluctuations Are Uncorrelated: Principal Component Analysis Reveals Anti-Correlated Networks. Brain Connectivity in press.

Card, N.S., Gharbawie, O.A., 2020. Principles of Intrinsic Motor Cortex Connectivity in Primates. The Journal of Neuroscience 40, 4348.

Chen, T., Ryali, S., Qin, S., Menon, V., 2013. Estimation of resting-state functional connectivity using random subspace based partial correlation: A novel method for reducing global artifacts. NeuroImage 82, 87-100.

Chen, X., Lu, B., Yan, C.-G., 2018. Reproducibility of R-fMRI metrics on the impact of different strategies for multiple comparison correction and sample sizes. Human Brain Mapping 39, 300-318.

Cheng, B., Liu, M., Zhang, D., Munsell, B.C., Shen, D., 2015. Domain Transfer Learning for MCI Conversion Prediction. IEEE Transactions on Biomedical Engineering 62, 1805-1817.

Clayden, J.D., A. J. Storkey, M. E. Bastin, 2008. A comparison of seeding strategies for group tractography Proc. Intl. Soc. Mag. Reson. Med. 16.

Cohen, A., D. Fair, N. Dosenbach, F. Miezin, D. Dierker, D. Van Essen, B. Schlaggar, S. Petersen, 2008. Defining functional areas in individual human brains using resting functional connectivity MRI. NeuroImage 41, 45-57.

Cui, Y., Liu, B., Luo, S., Zhen, X., Fan, M., Liu, T., Zhu, W., Park, M., Jiang, T., Jin, J.S., the Alzheimer's Disease Neuroimaging, I., 2011. Identification of Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using Multivariate Predictors. PLoS One 6, e21896.

Dansereau, C., Benhajali, Y., Risterucci, C., Pich, E.M., Orban, P., Arnold, D., Bellec, P., 2017. Statistical power and prediction accuracy in multisite resting-state fMRI connectivity. NeuroImage 149, 220-232.

Dawson, D.A., Cha, K., Lewis, L.B., Mendola, J.D., Shmuel, A., 2013. Evaluation and calibration of functional network modeling methods based on known anatomical connections. NeuroImage 67, 331-343.

Dawson, D.A., Jack Lam, Amir Shmuel. Local Principal Component regression, partial correlation and inverse covariance identify direct connections in an unknown, fine-scale, densely connected functional network. under revision for Neuroimage.

De Luca, M., Beckmann, C. F., De Stefano, N., Matthews, P. M., Smith, S. M, 2006. fMRI resting state networks define distinct modes of long-distance interactions in the human brain. NeuroImage 29, 1359-1367.

Dumoulin, S.O., Brian A. Wandell, 2008. Population receptive field estimates in human visual cortex. NeuroImage 39, 647-660.

Edward, D., 2000. Introduction to Graphical Modelling. New York: Springer 2.

Efron, B., 2008. Simultaneous inference: When should hypothesis testing problems be combined? Ann. Appl. Stat. 2, 197-223.

Eguchi, Y., Kanamoto, H., Oikawa, Y., Suzuki, M., Yamanaka, H., Tamai, H., Kobayashi, T., Orita, S., Yamauchi, K., Suzuki, M., Inage, K., Aoki, Y., Watanabe, A., Furuya, T., Koda, M., Takahashi, K., Ohtori, S., 2017. Recent advances in magnetic resonance neuroimaging of lumbar nerve to clinical applications: A review of clinical studies utilizing Diffusion Tensor Imaging and Diffusion-weighted magnetic resonance neurography. Spine Surg Relat Res 1, 61-71.

Engel, S., Rumelhart, D.E., Wandell, B.A., Lee, A.T., Glover, G.H., Chichilnisky, E.J., Shadlen, M.N., 1994. fMRI of human visual cortex. Nature 369 (6481).

Engel, S.A., G H Glover and B A Wandell, 1997. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb Cortex. 7, 181-192.

Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. Cerebral cortex (New York, N.Y. : 1991) 1, 1-47.

Ferreira, D., Pereira, J.B., Volpe, G., Westman, E., 2019. Subtypes of Alzheimer's Disease Display Distinct Network Abnormalities Extending Beyond Their Pattern of Brain Atrophy. Frontiers in Neurology 10.

Ferreira, D., Verhagen, C., Hernández-Cabrera, J.A., Cavallin, L., Guo, C.-J., Ekman, U., Muehlboeck, J.S., Simmons, A., Barroso, J., Wahlund, L.-O., Westman, E., 2017. Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: longitudinal trajectories and clinical applications. Scientific reports 7, 46263-46263.

Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., Constable, R.T., 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat Neurosci 18, 1664-1671.

Fleisher, A.S., Chen, K., Quiroz, Y.T., Jakimovich, L.J., Gomez, M.G., Langois, C.M., Langbaum, J.B.S., Ayutyanont, N., Roontiva, A., Thiyyagura, P., Lee, W., Mo, H., Lopez, L., Moreno, S., Acosta-Baena, N., Giraldo, M., Garcia, G., Reiman, R.A., Huentelman, M.J., Kosik, K.S., Tariot, P.N., Lopera, F., Reiman, E.M., 2012. Florbetapir PET analysis of amyloid-β deposition in the presenilin 1 E280A autosomal dominant Alzheimer's disease kindred: a cross-sectional study. The Lancet Neurology 11, 1057-1065.

Fornito, A., Zalesky, A., Bullmore, E.T., 2010. Network scaling effects in graph analytic studies of human resting-state FMRI data. Front Syst Neurosci 4, 22.

Fox, M.D., Raichle, M.E., 2007. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci 8, 700-711.

Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9, 432-441.

Friston, K., 2011. Functional and effective connectivity: a review. Brain Connect 1, 13-36.

Friston, K.J., 1994. Functional and effective connectivity in neuroimaging: A synthesis. Human Brain Mapping 2, 56-78.

Friston, K.J., Frith C.D., Liddle P.F., Frackowiak R.S.J., 1993. Functional connectivity: the principal component analysis of large (PET) data sets. J. Cereb. Blood Flow Metab. 13, 5 – 14.

Garrison, K.A., Scheinost, D., Finn, E.S., Shen, X., Constable, R.T., 2015. The (in)stability of functional brain network measures across thresholds. NeuroImage 118, 651-661.

Gauthier, S., Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, Belleville S, Brodaty FL, Bennett D, Chertkow H, 2006. Mild cognitive impairment. Lancet 367, 1262–1270.

Glaser, J.I., Benjamin, A.S., Farhoodi, R., Kording, K.P., 2019. The roles of supervised machine learning in systems neuroscience. Prog Neurobiol 175, 126-137.

Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., Smith, S.M., Van Essen, D.C., 2016. A multi-modal parcellation of human cerebral cortex. Nature 536, 171-178.

Glover, G.H., 2011. Overview of functional magnetic resonance imaging. Neurosurgery clinics of North America 22, 133-vii.

Gómez-Sancho, M., Tohka, J., Gómez-Verdejo, V., 2018. Comparison of feature representations in MRI-based MCI-to-AD conversion prediction. Magnetic Resonance Imaging 50, 84-95.

Gravel, N., Harvey, B., Nordhjem, B., Haak, K.V., Dumoulin, S.O., Renken, R., Curcic-Blake, B., Cornelissen, F.W., 2014. Cortical connective field estimates from resting state fMRI activity. Front Neurosci 8, 339.

Greicius, M.D., Kimmel, D.L., 2012. Neuroimaging insights into network-based neurodegeneration. Curr Opin Neurol 25, 727-734.

Greicius, M.D., Supekar, K., Menon, V., Dougherty, R.F., 2009. Resting-state functional connectivity reflects structural connectivity in the default mode network. Cereb Cortex 19, 72-78.

Griffanti, L., Baglio, F., Lagana, M.M., Preti, M.G., Cecconi, P., Clerici, M., Nemni, R., Baselli, G., 2015. Individual Thresholding of Voxel-based Functional Connectivity Maps. Estimation of Random Errors by Means of Surrogate Time Series. Methods Inf Med 54, 227-231.

Grothe, M.J., Teipel, S.J., 2016. Spatial patterns of atrophy, hypometabolism, and amyloid deposition in Alzheimer's disease correspond to dissociable functional brain networks. Hum Brain Mapp 37, 35-53.

Haak, K.V., Marquand, A.F., Beckmann, C.F., 2018. Connectopic mapping with resting-state fMRI. NeuroImage 170, 83-94.

Haak, K.V., Morland, A.B., Rubin, G.S., Cornelissen, F.W., 2016. Preserved retinotopic brain connectivity in macular degeneration. Ophthalmic Physiol Opt 36, 335-343.

Haimovici, A., Tagliazucchi, E., Balenzuela, P., Chialvo, D.R., 2013. Brain organization into resting state networks emerges at criticality on a model of the human connectome. Phys Rev Lett 110, 178101.

Häkkinen, S., Rinne, T., 2018. Intrinsic, stimulus-driven and task-dependent connectivity in human auditory cortex. Brain Structure and Function 223, 2113-2127.

Hart, M.G., Price, S.J., Suckling, J., 2017. Functional connectivity networks for preoperative brain mapping in neurosurgery. J Neurosurg 126, 1941-1950.

Hayasaka, S., Laurienti, P.J., 2010. Comparison of characteristics between region-and voxel-based network analyses in resting-state fMRI data. NeuroImage 50, 499-508.

He, H., Sui J, Yu Q, Turner JA, Ho BC, Sponheim SR, Manoach DS, Clark VP, Calhoun VD., 2012. Altered Small-World Brain Networks in Schizophrenia Patients during Working Memory Performance. PLoS One. 7, e38195.

Heinzle, J., Thorsten Kahnt, John-Dylan Haynes, 2011. Topographically specific functional connectivity between visual field maps in the human brain. NeuroImage 56, 1426-1436.

Hinne, M., Janssen, R.J., Heskes, T., van Gerven, M.A., 2015. Bayesian Estimation of Conditional Independence Graphs Improves Functional Connectivity Estimates. PLoS Comput Biol 11, e1004534.

Holmes, G., 1945. The organisation of the visual cortex in man. Ferrier lecture. Proc. R. Soc. Ser. B 132, 348–361.

Honey, C., Thivierge JP, Sporns O, 2010. Can structure predict function in the human brain? NeuroImage 52, 766-776.

Hsu, D., Marshall, G.A., 2017. Primary and Secondary Prevention Trials in Alzheimer Disease: Looking Back, Moving Forward. Current Alzheimer research 14, 426-440.

Huang, H., Ding, M., 2016. Linking Functional Connectivity and Structural Connectivity Quantitatively: A Comparison of Methods. Brain Connect 6, 99-108.

Hubel DH & Wiesel TN (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160, 106–154.

Hubel DH & Wiesel TN (1968). Receptive fields and functional architecture of monkey striate cortex. J Physiol 195, 215–243.

Huettel, S.A., Song, A.W., McCarthy, G., 2004. Functional Magnetic Resonance Imaging. Sinauer, Sunderland.

Inouye, T. (1900, 2000). "Visual disturbances following gunshot wounds of the cortical visual area based on observations of the wounded in the recent Japanese wars." Brain -, 1(−, 101): 1904–1905, 1123.

Jarrahi, B., Mantini, D., Mehnert, U., Kollias, S., 2015. Exploring influence of subliminal interoception on whole-brain functional network connectivity dynamics. Conf Proc IEEE Eng Med Biol Soc 2015, 670-674.

Katsuki, F., Constantinidis, C., 2013. Time Course of Functional Connectivity in Primate Dorsolateral Prefrontal and Posterior Parietal Cortex during Working Memory. PLoS One 8, e81601.

Keselman, H.J., Cribbie, R., Holland, B., 1999. The pairwise multiple comparison multiplicity problem: An alternative approach to familywise and comparison wise Type I error control. Psychological Methods 4, 58-69.

Khazaee, A., Ebrahimzadeh, A., Babajani-Feremi, A., 2015. Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory. Clin Neurophysiol 126, 2132-2141.

Koch, M.A., Norris, D. G., Hund-Georgiadis, M., 2002. An investigation of functional and anatomical connectivity using magnetic resonance imaging. NeuroImage 16.

Krstajic, D., Buturovic, L.J., Leahy, D.E., Thomas, S., 2014. Cross-validation pitfalls when selecting and assessing regression and classification models. Journal of cheminformatics 6, 10-10.

Larsson, J., Heeger, D.J., 2006. Two retinotopic visual areas in human lateral occipital cortex. J. Neurosci. 26 (51), 13128–13142.

Larson-Prior, L.J., Zempel, J.M., Nolan, T.S., Prior, F.W., Snyder, A.Z., Raichle, M.E., 2009. Cortical network functional connectivity in the descent to sleep. Proc Natl Acad Sci U S A 106, 4489-4494.

Li, F., Tran, L., Thung, K.-H., Ji, S., Shen, D., Li, J., 2014. Robust Deep Learning for Improved Classification of AD/MCI Patients. In: Wu, G., Zhang, D., Zhou, L. (Eds.), Machine Learning in Medical Imaging. Springer International Publishing, Cham, pp. 240-247.

Li, H., Gui, J., 2006. Gradient directed regularization for sparse Gaussian concentration graphs, with applications to inference of genetic networks. Biostatistics 7, 302-317.

Li, R., Yu, J., Zhang, S., Bao, F., Wang, P., Huang, X., Li, J., 2013. Bayesian network analysis reveals alterations to default mode network connectivity in individuals at risk for Alzheimer's disease. PLoS One 8, e82104.

Liang, X., Wang, J., Yan, C., Shu, N., Xu, K., Gong, G., He, Y., 2012. Effects of different correlation metrics and preprocessing factors on small-world brain functional networks: a resting-state functional MRI study. PLoS One 7, e32766-e32766.

Liang, X., Zou, Q., He, Y., Yang, Y., 2013. Coupling of functional connectivity and regional cerebral blood flow reveals a physiological basis for network hubs of the human brain. Proceedings of the National Academy of Sciences 110, 1929.

Liao, X.-H., Xia, M.-R., Xu, T., Dai, Z.-J., Cao, X.-Y., Niu, H.-J., Zuo, X.-N., Zang, Y.-F., He, Y., 2013. Functional brain hubs and their test–retest reliability: A multiband resting-state functional MRI study. NeuroImage 83, 969-982.

Liu, W., & Luo, X. , 2015. Fast and Adaptive Sparse Precision Matrix Estimation in High Dimensions. Journal of multivariate analysis 135, 153–162.

Liu, X., Lauer, K.K., Ward, B.D., Roberts, C.J., Liu, S., Gollapudy, S., Rohloff, R., Gross, W., Xu, Z., Chen, G., Binder, J.R., Li, S.J., Hudetz, A.G., 2017. Fine-Grained Parcellation of Brain Connectivity Improves Differentiation of States of Consciousness During Graded Propofol Sedation. Brain Connect 7, 373-381.

Liu, X., Tosun, D., Weiner, M.W., Schuff, N., Alzheimer's Disease Neuroimaging, I., 2013. Locally linear embedding (LLE) for MRI based Alzheimer's disease classification. NeuroImage 83, 148-157.

Ludbrook, J., 1998. MULTIPLE COMPARISON PROCEDURES UPDATED. Clinical and Experimental Pharmacology and Physiology 25, 1032-1037.

Marrelec, G., J. Daunizeau, M. Pelegrini-Issac, J. Doyon, H. Benali, 2005. Conditional correlation as a measure of mediated interactivity in fMRI and MEG/EEG. IEEE Transactions on Signal Processing 53, 3503-3516.

Marrelec, G., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehéricy, S., Doyon, J., Benali, H., 2006. Partial correlation for functional brain interactivity investigation in functional MRI. NeuroImage 32.

Mathotaarachchi, S., Pascoal, T.A., Shin, M., Benedet, A.L., Kang, M.S., Beaudry, T., Fonov, V.S., Gauthier, S., Rosa-Neto, P., 2017. Identifying incipient dementia individuals using machine learning and amyloid imaging. Neurobiol Aging 59, 80-90.

McDowell, A.R., Carmichael, D.W., 2019. Optimal repetition time reduction for single subject event-related functional magnetic resonance imaging. Magnetic Resonance in Medicine 81, 1890-1897.

Meinshausen, N., Buhlmann, P., 2006. High-dimensional graphs and variable selection with the Lasso. Ann. Statist. 34, 1436-1462.

Morris, D.M., Embleton, K.V., Parker, G.J.M., 2008. Probabilistic fibre tracking: Differentiation of connections from chance events. NeuroImage 42, 1329-1339.

Noh, Y., Jeon, S., Lee, J.M., Seo, S.W., Kim, G.H., Cho, H., Ye, B.S., Yoon, C.W., Kim, H.J., Chin, J., Park, K.H., Heilman, K.M., Na, D.L., 2014. Anatomical heterogeneity of Alzheimer disease: based on cortical thickness on MRIs. Neurology 83, 1936-1944.

O'Donnell, L.J., Westin, C.F., 2011. An introduction to diffusion tensor image analysis. Neurosurg Clin N Am 22, 185-196, viii.

Onu, M., Badea, L., Roceanu, A., Tivarus, M., Bajenaru, O., 2015. Increased connectivity between sensorimotor and attentional areas in Parkinson's disease. Neuroradiology 57, 957-968.

Park, B., Eo, J., Park, H.J., 2017. Structural Brain Connectivity Constrains within-a-Day Variability of Direct Functional Connectivity. Front Hum Neurosci 11, 408.

Peng, J., Pei Wang, Nengfeng Zhou, Ji Zhu 2009. Partial Correlation Estimation by Joint Sparse Regression Models. Journal of the American Statistical Association 104, 735-746.

Persson, K., Eldholm, R.S., Barca, M.L., Cavallin, L., Ferreira, D., Knapskog, A.-B., Selbæk, G., Brækhus, A., Saltvedt, I., Westman, E., Engedal, K., 2017. MRI-assessed atrophy subtypes in Alzheimer’s disease and the cognitive reserve hypothesis. PLoS One 12, e0186595.

Pervaiz, U., Vidaurre, D., Woolrich, M.W., Smith, S.M., 2020. Optimising network modelling methods for fMRI. NeuroImage 211, 116604.

Plant, C., Teipel, S.J., Oswald, A., Böhm, C., Meindl, T., Mourao-Miranda, J., Bokde, A.W., Hampel, H., Ewers, M., 2010. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. NeuroImage 50, 162-174.

Poldrack, R.A., 2007. Region of interest analysis for fMRI. Social cognitive and affective neuroscience 2, 67-70.

Power, Jonathan D., Schlaggar, Bradley L., Petersen, Steven E., 2014. Studying Brain Organization via Spontaneous fMRI Signal. Neuron 84, 681-696.

Quigley, M.A.e.a., 2003. Role of the corpus callosum in functional connectivity. Am. J. Neuroradiol 24.

Raemaekers, M., Schellekens, W., van Wezel, R.J., Petridou, N., Kristo, G., Ramsey, N.F., 2014. Patterns of resting state connectivity in human primary visual cortical areas: a 7T fMRI study. NeuroImage 84, 911-921.

Rausch, A., Zhang, W., Haak, K.V., Mennes, M., Hermans, E.J., van Oort, E., van Wingen, G., Beckmann, C.F., Buitelaar, J.K., Groen, W.B., 2016. Altered functional connectivity of the amygdaloid input nuclei in adolescents and young adults with autism spectrum disorder: a resting state fMRI study. Mol Autism 7, 13.

Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. NeuroImage 52, 1059-1069.

Ryali, S., Chen, T., Supekar, K., Menon, V., 2012. Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. NeuroImage 59, 3852-3861.

Schafer, J., Strimmer, K., 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat Appl Genet Mol Biol 4, Article32.

Schmittmann, V.D., Jahfari, S., Borsboom, D., Savi, A.O., Waldorp, L.J., 2015. Making Large-Scale Networks from fMRI Data. PLoS One 10, e0129074.

Sereno, M., Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 269, 889-893.

Shehzad, Z., Kelly, C., Reiss, P.T., Cameron Craddock, R., Emerson, J.W., McMahon, K., Copland, D.A., Xavier Castellanos, F., Milham, M.P., 2014. A multivariate distance-based analytic framework for connectome-wide association studies. NeuroImage 93, 74-94.

Smith, S.M., 2012. The future of FMRI connectivity. NeuroImage 62, 1257-1266.

Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., Kelly, M., Laumann, T., Miller, K.L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A.Z., Vu, A.T., Woolrich, M.W., Xu, J., Yacoub, E., Uğurbil, K., Van Essen, D.C., Glasser, M.F., 2013. Resting-state fMRI in the Human Connectome Project. NeuroImage 80, 144-168.

Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F., 2009. Correspondence of the brain's functional architecture during activation and rest. Proc Natl Acad Sci U S A 106, 13040-13045.

Smith, S.M., Karla L. Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F. Beckmann, Thomas E. Nichols, Joseph D. Ramsey, Mark W. Woolrich, 2011. Network modelling methods for FMRI. NeuroImage 54, 875-891.

Sohn, W.S., Yoo, K., Na, D.L., Jeong, Y., 2014. Progressive changes in hippocampal resting-state connectivity across cognitive impairment: a cross-sectional study from normal to Alzheimer disease. Alzheimer Dis Assoc Disord 28, 239-246.

Sourty, M., Thoraval, L., Roquet, D., Armspach, J.P., Foucher, J., Blanc, F., 2016. Identifying Dynamic Functional Connectivity Changes in Dementia with Lewy Bodies Based on Product Hidden Markov Models. Front Comput Neurosci 10, 60.

Tagliazucchi, E., von Wegner, F., Morzelewski, A., Brodbeck, V., Jahnke, K., Laufs, H., 2013. Breakdown of long-range temporal dependence in default mode and attention networks during deep sleep. Proceedings of the National Academy of Sciences of the United States of America 110, 15419-15424.

Tam, A., Dansereau, C., Badhwar, A., Orban, P., Belleville, S., Chertkow, H., Dagher, A., Hanganu, A., Monchi, O., Rosa-Neto, P., Shmuel, A., Wang, S., Breitner, J., Bellec, P., 2015. Common Effects of Amnestic Mild Cognitive Impairment on Resting-State Connectivity Across Four Independent Studies. Front Aging Neurosci 7, 242.

Tong, T., Gao, Q., Guerrero, R., Ledig, C., Chen, L., Rueckert, D., Initiative, A.D.N., 2017. A Novel Grading Biomarker for the Prediction of Conversion From Mild Cognitive Impairment to Alzheimer's Disease. IEEE Transactions on Biomedical Engineering 64, 155-165.

Tong, Y., Chen, Q., Nichols, T.E., Rasetti, R., Callicott, J.H., Berman, K.F., Weinberger, D.R., Mattay, V.S., 2016. Seeking Optimal Region-Of-Interest (ROI) Single-Value Summary Measures for fMRI Studies in Imaging Genetics. PLoS One 11, e0151391-e0151391.

van den Heuvel, M., Mandl RC, Kahn RS, Hulshoff Pol HE, 2009. Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. Hum Brain Mapp 30, 3127-3141.

van den Heuvel, M.P., de Lange, S.C., Zalesky, A., Seguin, C., Yeo, B.T.T., Schmidt, R., 2017. Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations. NeuroImage 152, 437-449.

van den Heuvel, M.P., Hulshoff Pol, H.E., 2010. Specific somatotopic organization of functional connections of the primary motor network during resting state. Hum Brain Mapp 31, 631-644.

Van Dijk, K.R., Hedden, T., Venkataraman, A., Evans, K.C., Lazar, S.W., Buckner, R.L., 2010. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. J Neurophysiol 103, 297-321.

Wang, B., Guo, J., Yan, T., Ohno, S., Kanazawa, S., Huang, Q., Wu, J., 2016. Neural Responses to Central and Peripheral Objects in the Lateral Occipital Cortex. Front Hum Neurosci 10, 54.

Wang, J., Zuo, X., Dai, Z., Xia, M., Zhao, Z., Zhao, X., Jia, J., Han, Y., He, Y., 2013. Disrupted functional brain connectome in individuals at risk for Alzheimer's disease. Biological Psychiatry 73, 472-481.

Wang, L., Mruczek, R.E., Arcaro, M.J., Kastner, S., 2015. Probabilistic Maps of Visual Topography in Human Cortex. Cereb Cortex 25, 3911-3931.

Weiler, M., Fukuda, A., Massabki, L.H., Lopes, T.M., Franco, A.R., Damasceno, B.P., Cendes, F., Balthazar, M.L., 2014. Default mode, executive function, and language functional connectivity networks are compromised in mild Alzheimer's disease. Curr Alzheimer Res 11, 274-282.

Whittaker, J., 1990. Graphical Models in Applied Mathematical Multivariate Statistics. Chichester: John Wiley & Sons Ltd.

Xiong, J., Parsons L.M., Gao J.-H., Fox P.T., 1999. Interregional connectivity to primary motor cortex revealed using MRI resting state images. Hum. Brain Mapp 8, 151 – 156.

Yao, H., Liu, Y., Zhou, B., Zhang, Z., An, N., Wang, P., Wang, L., Zhang, X., Jiang, T., 2013. Decreased functional connectivity of the amygdala in Alzheimer's disease revealed by resting-state fMRI. Eur J Radiol 82, 1531-1538.

Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zollei, L., Polimeni, J.R., Fischl, B., Liu, H., Buckner, R.L., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol 106, 1125-1165.

Yu, G., Liu, Y., Thung, K.-H., Shen, D., 2014. Multi-Task Linear Programming Discriminant Analysis for the Identification of Progressive MCI Individuals. PLoS One 9, e96458.

Yuan, M., Lin, Y., 2007. Model Selection and Estimation in the Gaussian Graphical Model. Biometrika 94, 19-35.

Zhang, L., Qi, R., Wu, S., Zhong, J., Zhong, Y., Zhang, Z., Lu, G., 2012. Brain default-mode network abnormalities in hepatic encephalopathy: a resting-state functional MRI study. Hum Brain Mapp 33, 1384-1392.

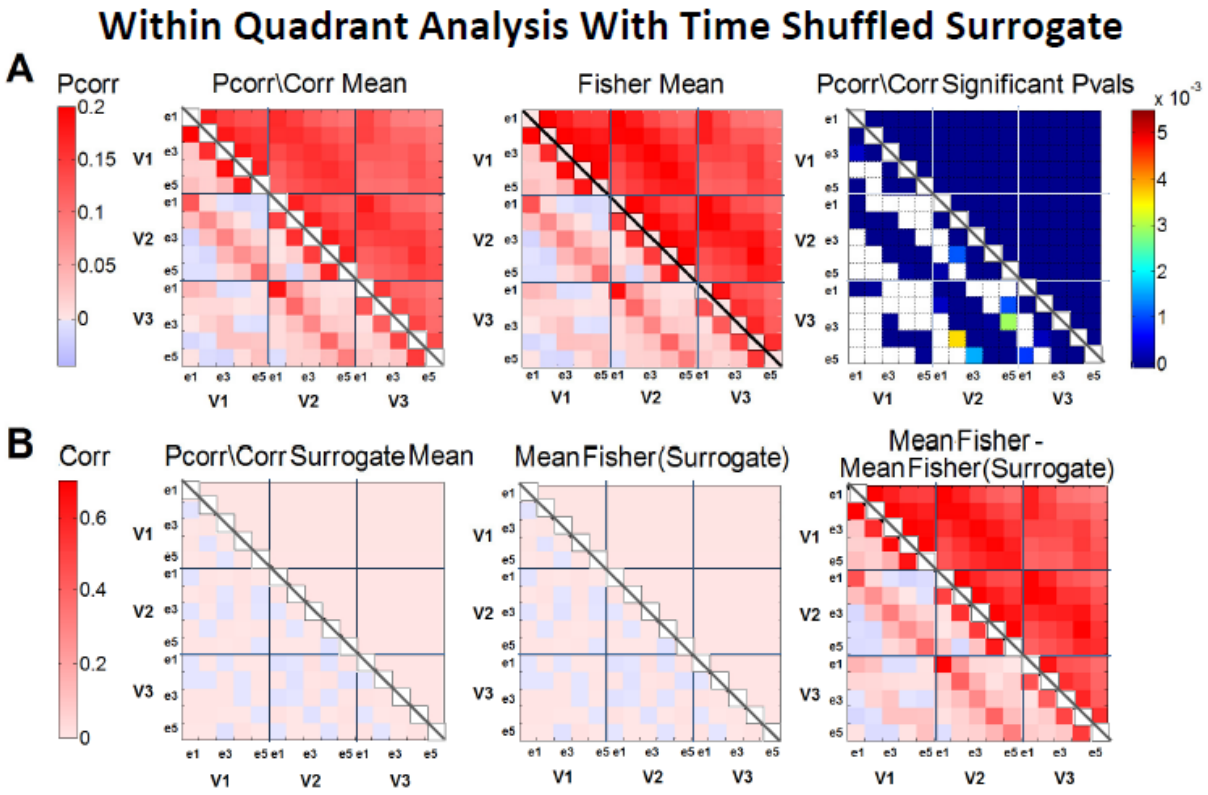
Zhou, B., Liu, Y., Zhang, Z., An, N., Yao, H., Wang, P., Wang, L., Zhang, X., Jiang, T., 2013. Impaired functional connectivity of the thalamus in Alzheimer's disease and mild cognitive impairment: a resting-state fMRI study. Curr Alzheimer Res 10, 754-766.

Zhou, B., Yao, H., Wang, P., Zhang, Z., Zhan, Y., Ma, J., Xu, K., Wang, L., An, N., Liu, Y., Zhang, X., 2015. Aberrant Functional Connectivity Architecture in Alzheimer's Disease and Mild Cognitive Impairment: A Whole-Brain, Data-Driven Analysis. Biomed Res Int 2015, 495375.

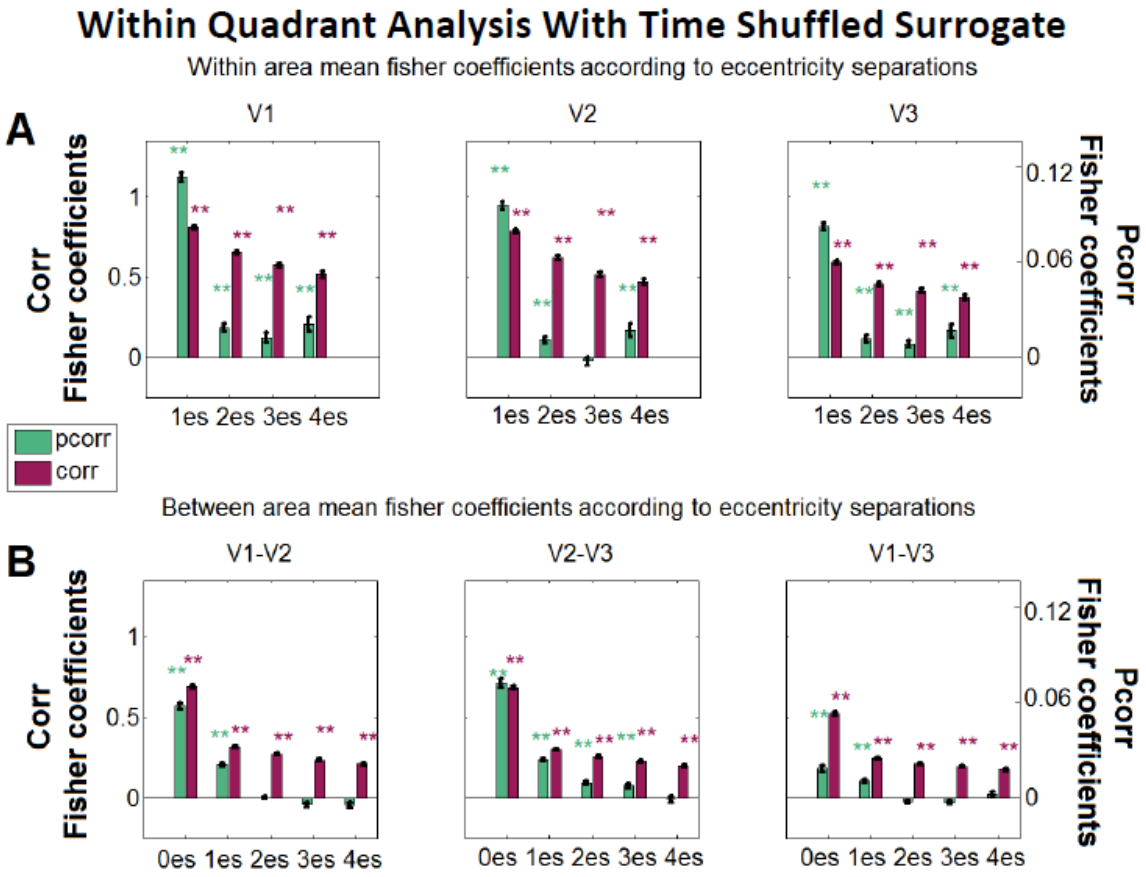
Zhou, J., Liu, S., Ng, K.K., Wang, J., 2017. Applications of Resting-State Functional Connectivity to Neurodegenerative Disease. Neuroimaging Clin N Am 27, 663-683.

Appendix A

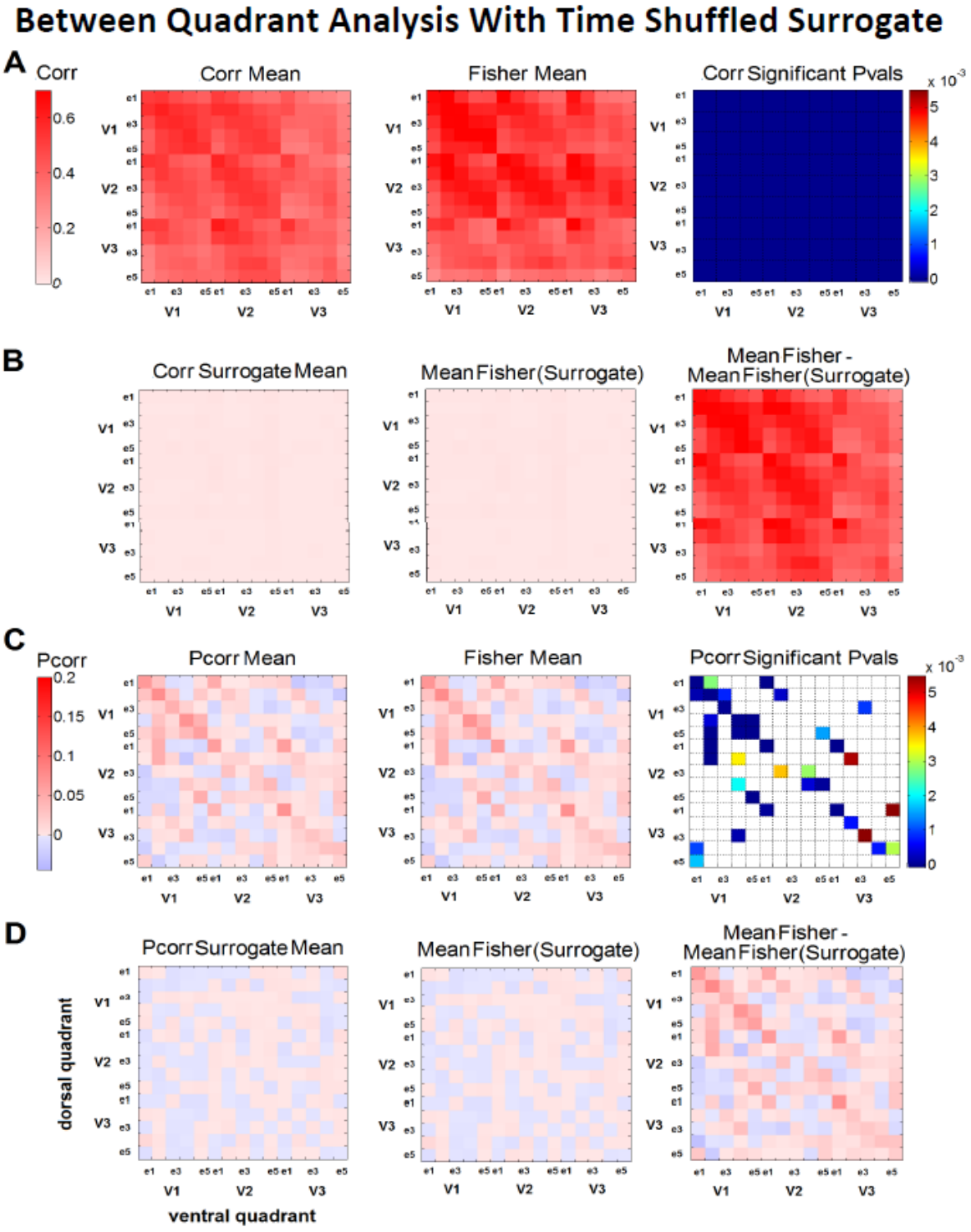
Supplementary Material from Manuscript 1



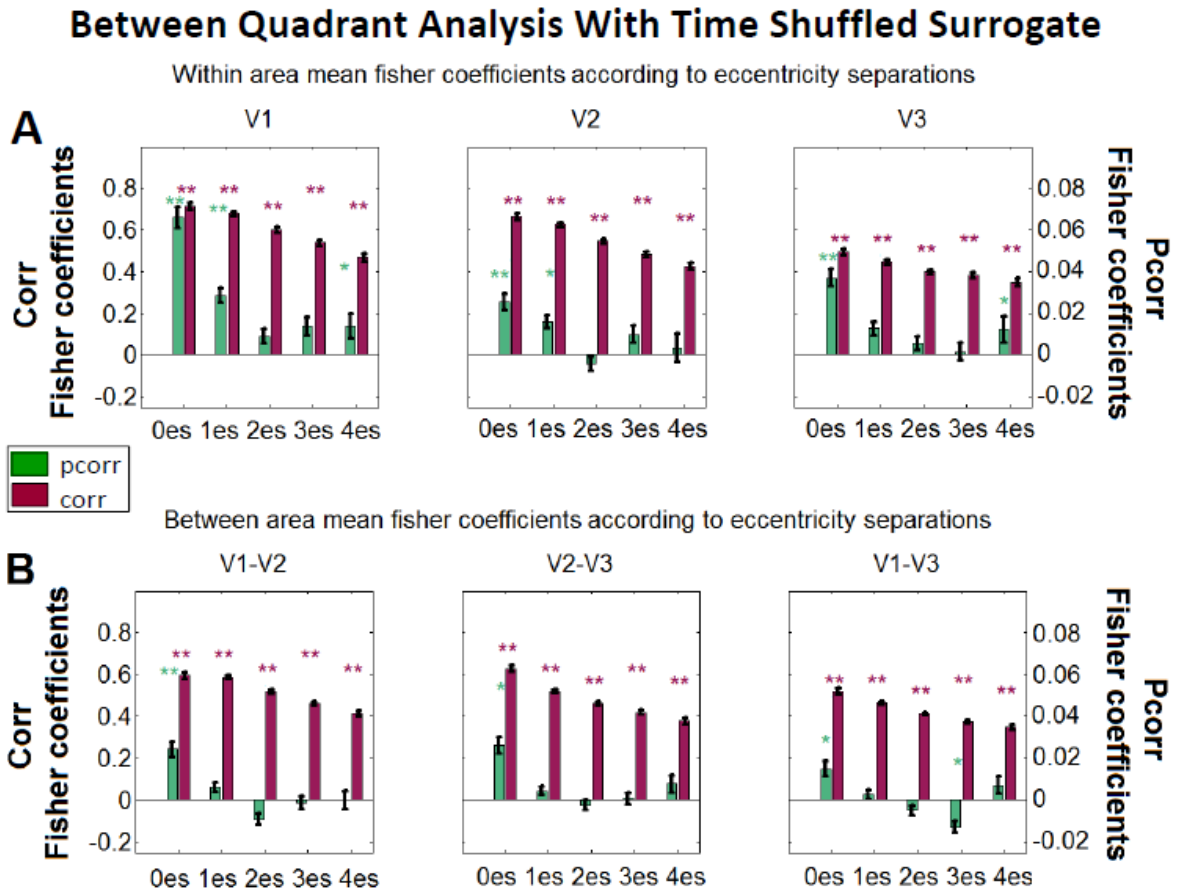
Supplementary Figure 1. Mean functional connectivity measures within a quadrant, computed across all subjects, runs per subjects, and the four quadrants; statistics done with use of a Time-Shuffled Surrogate. Shuffling was pursued based on the indices of the 8 scans, while keeping the time-course from each scan unchanged; we computed functional connectivity measures for the maximal possible number of permutations: ([8 × 8] – 8)/2 =28. Thus for each subject, 28 surrogate coefficients were computed for each ROI pair. A) Left: Mean Pcorr / Corr coefficients across all subjects (6), quadrants (4). And runs per subjects (8) (6*4*8). Center: Mean Fisher coefficients across all subjects and runs. Right: p values of statistically significant Pcorr / Corr associated with the functional connections in this 5 eccentricity region network. Significance is thresholded at 0.05, with false discovery rate (FDR) correction. White connections did not show statistical significance. B) Left: Mean Pcorr / Corr Surrogate coefficients across all subjects (6), quadrants (4) and shuffled runs per subjects (28) (6*4*28). Center: Mean Fisher coefficients of Surrogate across all subjects and runs. Right: Mean Fisher Pcorr/Corr minus the Mean Fisher of the Surrogate distributions.



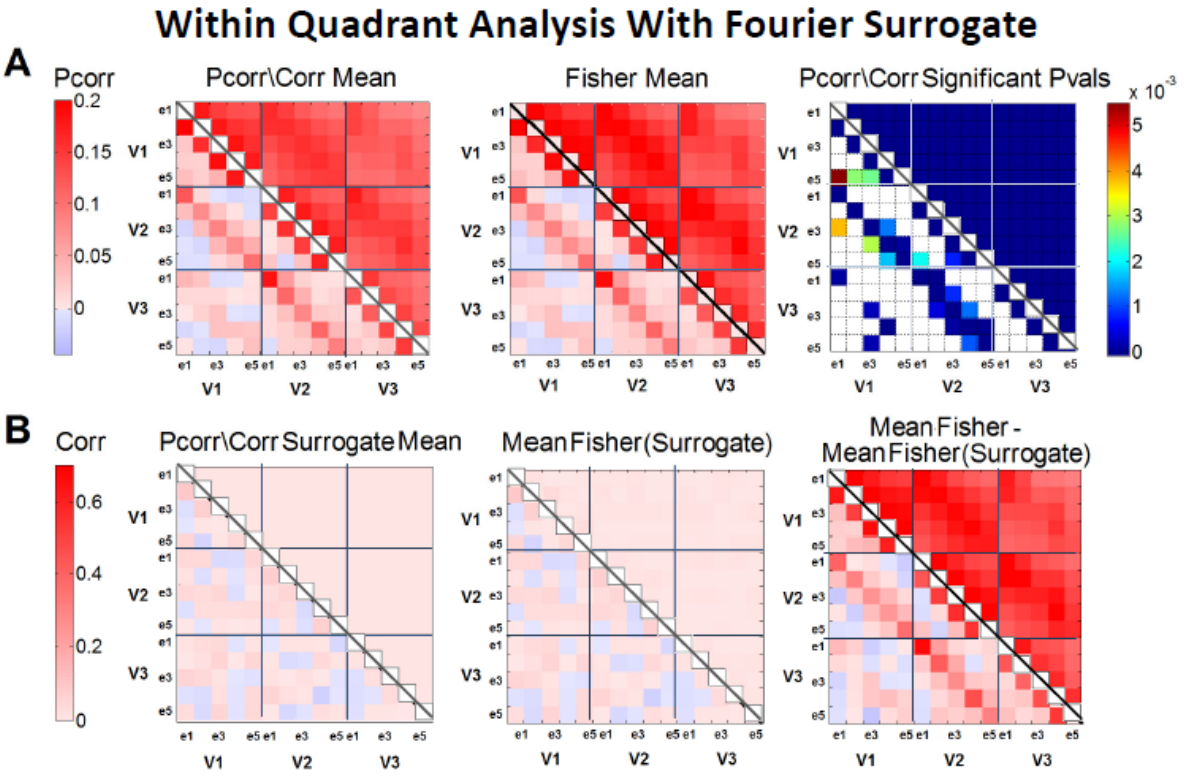
Supplementary Figure 2. Mean functional connectivity measures according to eccentricity bin groupings associated with ROI pairs composed of regions residing within the quadrant; statistics done with use of a Time-Shuffled Surrogate. Across all plots, green and purple bars represent the mean Fisher transformed Pcorr and Corr, respectively, computed across all subjects, runs and quadrants. Error bars reflect the standard error of the mean across measurements. A shows within-visual area Fisher coefficients for different eccentricity separations. Plots in B show the between-area Fisher coefficients of different eccentricity separations. A zero eccentricity separation (0es) indicates regions with the same eccentricity; a separation of one eccentricity (1es) means the regions' eccentricities were adjacent, a separation of two eccentricities (2es) means there was an eccentricity region separating the two regions' eccentricities. A separation of three eccentricities (3es) means there were two eccentricity regions separating the two regions' eccentricities. Finally, a separation of four eccentricities (4es) means there were three eccentricity regions separating the two regions' eccentricities. Statistical significance was assessed using alpha values of 0.05 (*) and 0.001 (**), following repeated measures statistical analyses. Individual connections are shown to be significant with an asterisk directly above the bar.



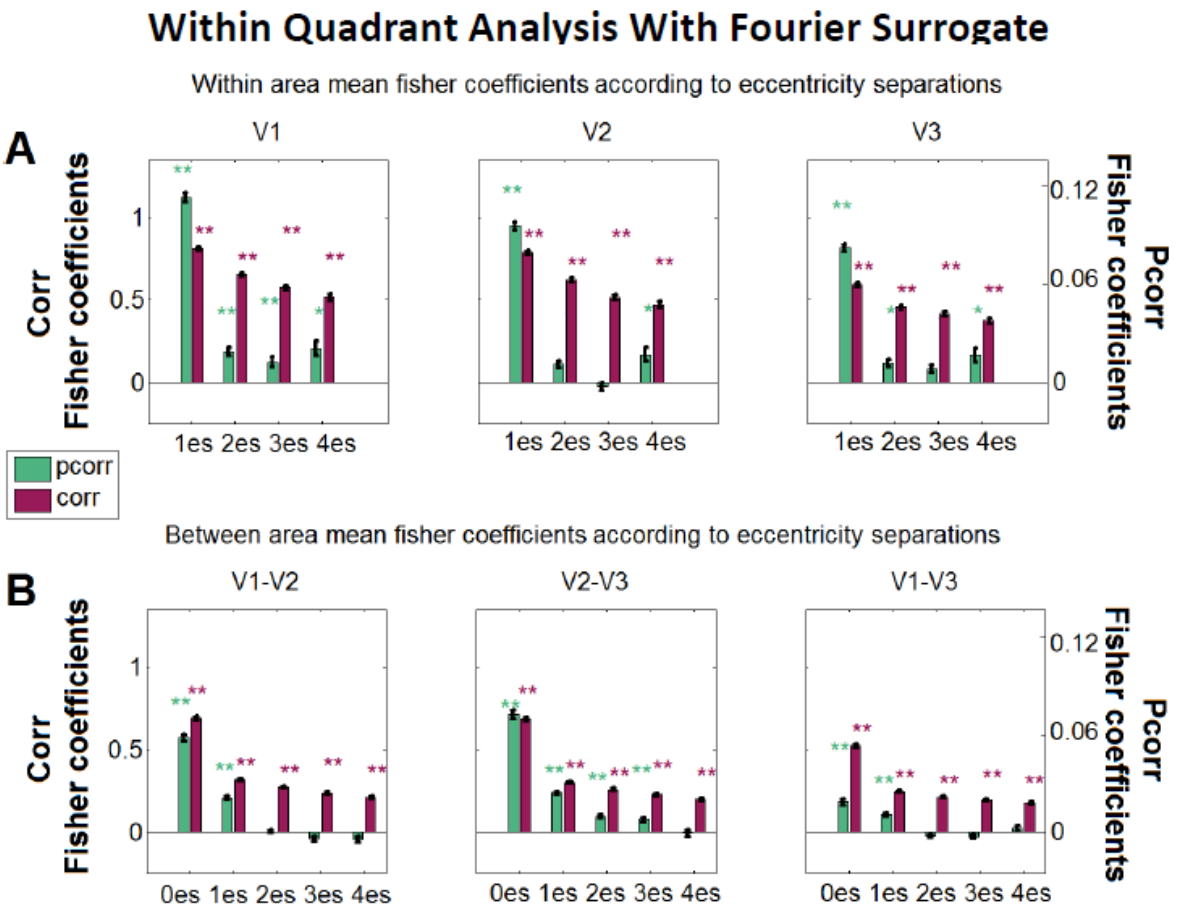
Supplementary Figure 3. Mean functional connectivity measures associated with functional connections between ROIs residing in different quadrants within a hemisphere; statistics done with use of a Time-Shuffled Surrogate. The mean connectivity measures were computed across all subjects, runs, and the two quadrant pairs. A & B reflect Corr connectivity measures and C & D reflect Pcorr connectivity measures. Presentation is as in Figure 7 and Supplementary Figures 1A and 1B.



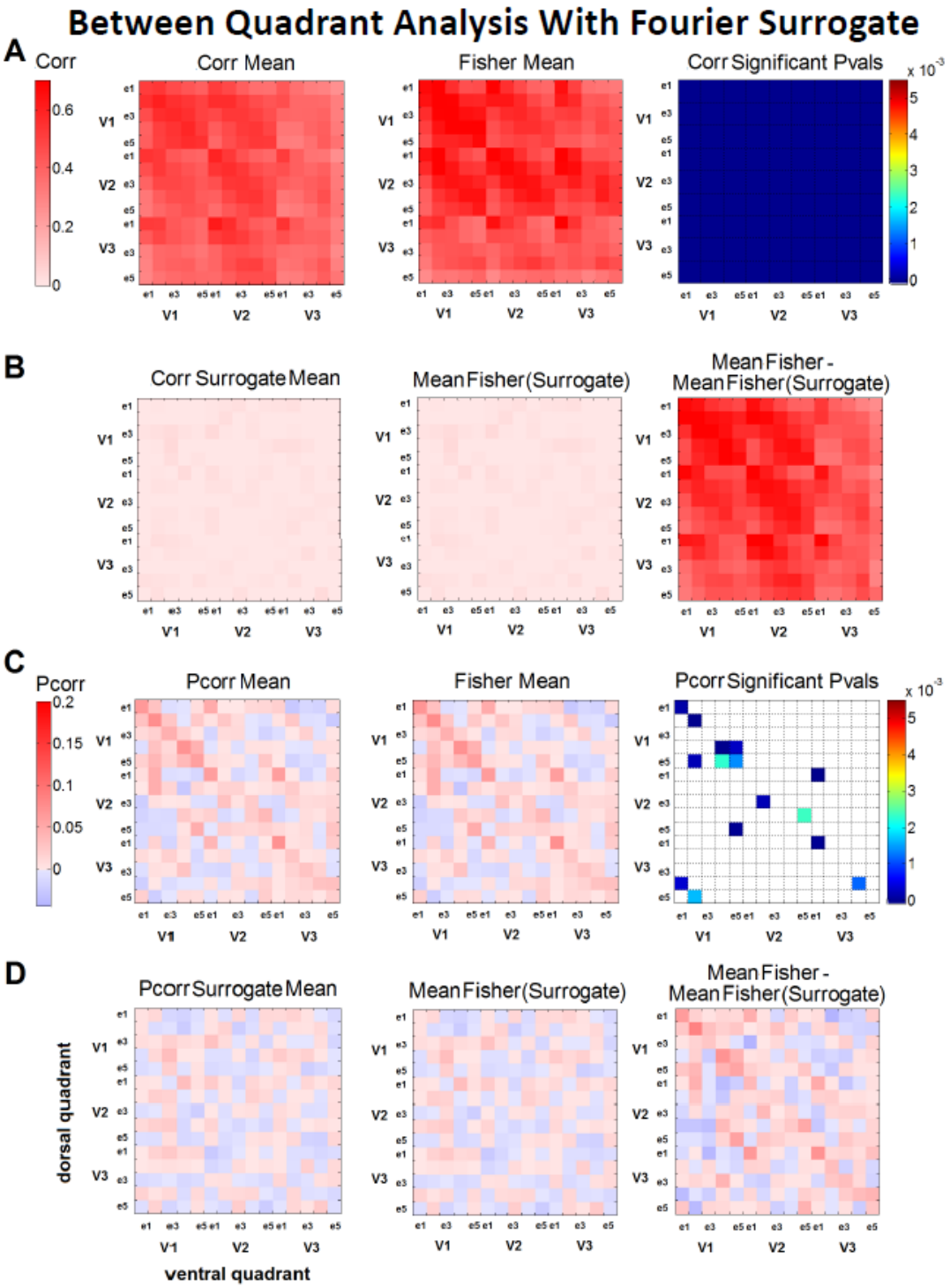
Supplementary Figure 4. Mean functional connectivity measures according to eccentricity bin groupings associated with ROI pairs composed of regions residing in different quadrants of a hemisphere; statistics done with use of a Time-Shuffled Surrogate. Across all plots, green and purple bars represent the mean Fisher transformed Pcorr and Corr, respectively, computed across all subjects, runs and pairs of quadrants. Error bars reflect the standard error of the mean across measurements. The format of presentation is as in Figures 8 A and C in the main manuscript and Supplementary Figure 2. * $p<0.05$; ** $p<0.001$.



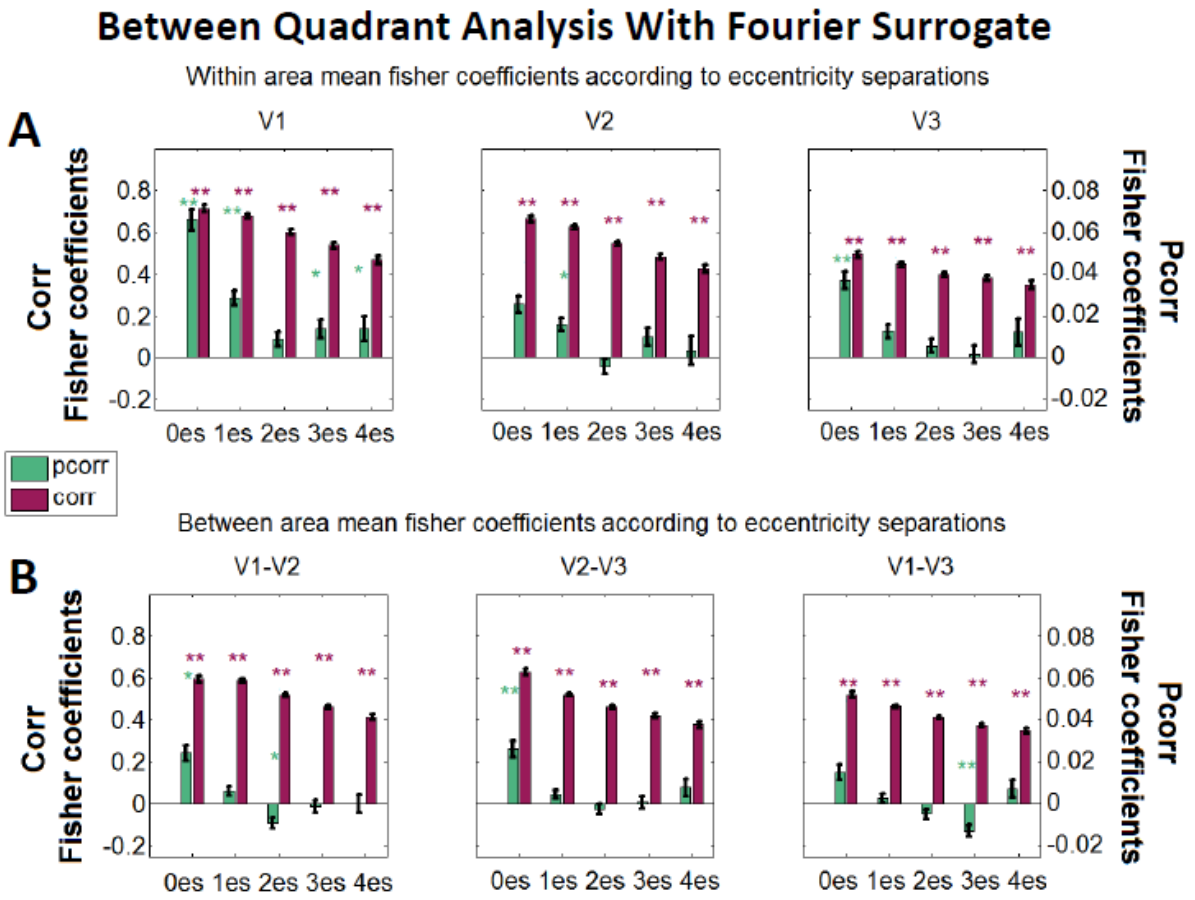
Supplementary Figure 5. Mean functional connectivity measures within a quadrant, computed across all subjects, four quadrants, and runs per subjects; statistics done with use of a Fourier Surrogate. To create these surrogate data, we shuffled the phase values from the Fourier transform of the original time-courses, then computed the inverse Fourier transform. Presentation is as in Supplementary Figure 1.



Supplementary Figure 6. Mean functional connectivity measures according to eccentricity bin groupings associated with ROI pairs composed of regions residing within the quadrant; statistics done with use of a Fourier Surrogate. Across all plots, green and purple bars represent the mean Fisher transformed Pcorr and Corr, respectively, computed across all subjects, runs and pairs of quadrants. Error bars reflect the standard error of the mean across measurements. The format of presentation is as in Supplementary Figure 2. * $p<0.05$; ** $p<0.001$.



Supplementary Figure 7. Mean functional connectivity measures associated with functional connections between ROIs residing in different quadrants in a hemisphere; statistics done with use of a Fourier Surrogate. Presentation is as in Suppl. Fig. 3. All entries along the diagonal that show statistical significance here also do in Fig. 7. Not all entries along the diagonal that show statistical significance in Fig. 7 and Suppl. Fig. 3 show it here too, most likely due to the lesser statistical power here (comparing 8 measured values to only 8 surrogate values; Suppl. Fig. 3 has 28 surrogate values; Fig. 7 compares 8 measured values to non-variable 0).



Supplementary Figure 8. Mean functional connectivity measures according to eccentricity bin groupings associated with ROI pairs composed of regions residing in different quadrants of a hemisphere; statistics done with use of a Fourier Surrogate. Presentation is as in Supplementary Figure 4.

Appendix B

Details regarding the differences between Part 2 of the thesis and Dawson et al., NeuroImage 2013

We acknowledge the similarity in questions posed by Dawson et al. (2013) and Part 2 of the thesis. That being said, there are important differences in the studies and much added valuable information. As a reminder, Dawson et al. (2013) examined the proportion of expected connections with interaction magnitudes greater than the 95th percentile of the unexpected connections distribution. These expected and unexpected connections were based in the visual cortex with three eccentricity regions defined in each of V1, V2, and V3. The following is a list of the differences between Dawson et al. (2013) and Part 2 of the PhD proposal.

(1) The most important difference is that the thresholding of connectivity measures employed by Dawson et al. (2013) was *based on prior knowledge of the system of study*. In other words, one cannot determine the 95th percentile of the unexpected connections distribution, unless one has complete knowledge of which connections are unexpected. In contrast, in Part 2 of the thesis our aim is to evaluate functional network modeling methods *given no prior knowledge of the system of study*. In the current study we emphasize evaluating the performance of models under statistical hypothesis testing with the null hypothesis that "no interaction" is represented by a zero-magnitude interaction measure. Based on Dawson et al. 2013, we could qualitatively predict that partial correlation and ICOV would be good candidates for identifying direct connections with no prior knowledge on the examined network. In the current study we demonstrate this directly and quantitatively.

(2) The choice of a statistical threshold is always difficult and arbitrary. In our current study, we evaluate which of the network modeling methods performs well using a 'conventional' p-value based significance level of 0.05, corrected for multiple comparisons. This assures the user that there is no need to 'guesstimate' a threshold. We find that the conventional threshold level is appropriate.

(3) Once again we are motivated by the fact that the choice of a statistical threshold is difficult and arbitrary. In addition to evaluating model performance at the FDR p-value

threshold based on a maximal FPR of 0.05, part 2 of the thesis includes an evaluation of the stability of the network predictions across a range of absolute statistical thresholds.

(4) We also consider a range of density-based thresholds applied to the network connectivity measures. Through these analyses we are able to evaluate whether the network models tested are more successful and/or more stable in the task of identifying direct connectivity through considerations of network density instead of via FPR-based statistical thresholding.

(5) We evaluate the network models against an inferred ground truth network with additional expected and unexpected connections compared to the 2013 study. With this new set of connections, we are now able to consider how well the model predictions fare with respect to triangle connectivity (see *Triangle Connectivity* below).

(6) We have now introduced the novel concept of LOCAL Principal Component regression (LPC), and have quantified its effect on all models (including models whose interaction measures are not centered around zero) given an FDR p-value threshold with a maximal FPR of 0.05. In preliminary analyses, this approach results in some of the best performances from network models. More details on this method are included below (see *Local Principal Component Regression*).

(7) Finally, in addition to analyzing our own data, we now support the results of our study with a large number (224) of Human Connectome Project data sets, following power and sample size considerations. This analysis allows us to observe which of the modeling methods will converge (given a large number of data sets) to results very similar to the inferred ground truth direct connectivity. Importantly, results from this analysis show that density-based network predictions are not improved when a large number of data-sets is used, whereas network predictions based on absolute statistical thresholding are indeed improved.

Local Principal Component Regression

While typical global signal approximations are made using information from all voxels in the brain, or all grey matter voxels, we have chosen to approximate the LOCAL average signal using only the voxels in V1, V2, and V3. Regions outside of the visual cortex are less retinotopic than

those within, therefore, their influence will be widespread in V1, V2, and V3 (Salin and Bullier, 1995). This leads to the expectation that global contributions can be adequately captured across the voxels contained in V1, V2 and V3. The LOCAL principal component method we will be applying is equivalent to a regression of the first principal component in a principal component decomposition (Carbonell et al., 2011) of the data from within a brain mask outlining the local network. Given our choice, we refer to our approximation as a "Local Principal Component" signal (LPC).

Triangle Connectivity

We have proceeded to include considerations of triangle connectivity within the network. With Partialcorr and ICOV there is concern that any true triangle connectivity in the network will be lost due to the regressions undertaken by these models. In the expected connections originally defined in Dawson et al. (2013) we did not look explicitly at how well the models do with triangle connectivity. In order to expand on our earlier results and address this concern in the current study (Part 2 of the thesis) we have enlarged our inferred ground truth network to include more connections, including some triangle motifs. For example, the three nodes including v1e1, v1e2, and v2e1 form a triangle connected by the following connections: v1e1-v1e2, v1e2-v2e1, v2e1-v1e1.

Some of the new content which will be included in coming drafts of Part 2 of the thesis includes the following regarding triangle connectivity:
"We have also given consideration to the idea of triangle connectivity within the network. This is an important consideration due to concerns that regression models such as partial correlation and ICOV will destroy triangles (Schmittmann et al., 2014). Partial correlation and ICOV do well in general in our densely connected network, because they eliminate 'network effects', i.e. indirect connections, from the network prediction. That being said, if there is indeed true direct connectivity in a triangle configuration, we wish to retain these connections in our network prediction. In the set of expected connections we have defined there are 48 triangles (for example: V1e1-V1e2-V2e1-V1e1). We have examined the proportion of the total number of triangles that each model has predicted at the FDR p-value threshold based on a maximal FPR of

0.05. Alongside the proportion of preserved triangles, we also present, for perspective, the total proportion of expected connections predicted (TPR) in the groupings of 1) connections not involved in triangles, and 2) connections involved in triangles. Finally, we present the difference in these two TPRs. This difference should be small if there is no bias toward eliminating triangles."

Our findings with regards to triangles reflect that there is no large bias observed in any of the models in terms of the elimination of triangles. Moreover, unlike partial correlation and ICOV, the LPC regression we introduce in conjunction with Pearson correlation as an interaction measure is not prone to eliminating triangles.

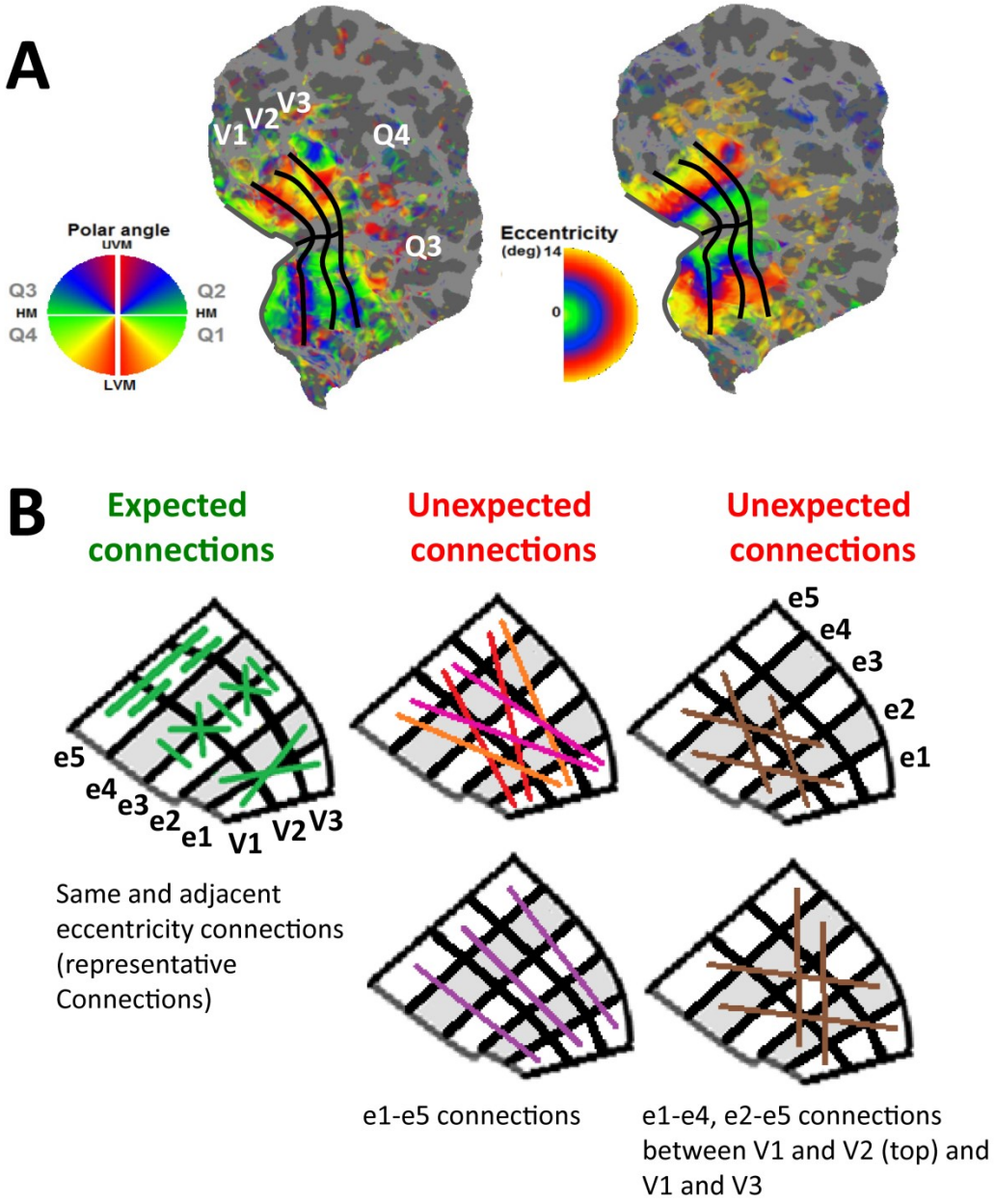
Additional References

Salin, P., Bullier J, 1995. Corticocortical connections in the visual system: structure and function. Physiol Rev 75, 107-154.

Schmittmann, V.D., Jahfari, S., Borsboom, D., Savi, A.O., Waldorp, L.J., 2015. Making Large-Scale Networks from fMRI Data. PLoS One 10, e0129074.

Appendix C

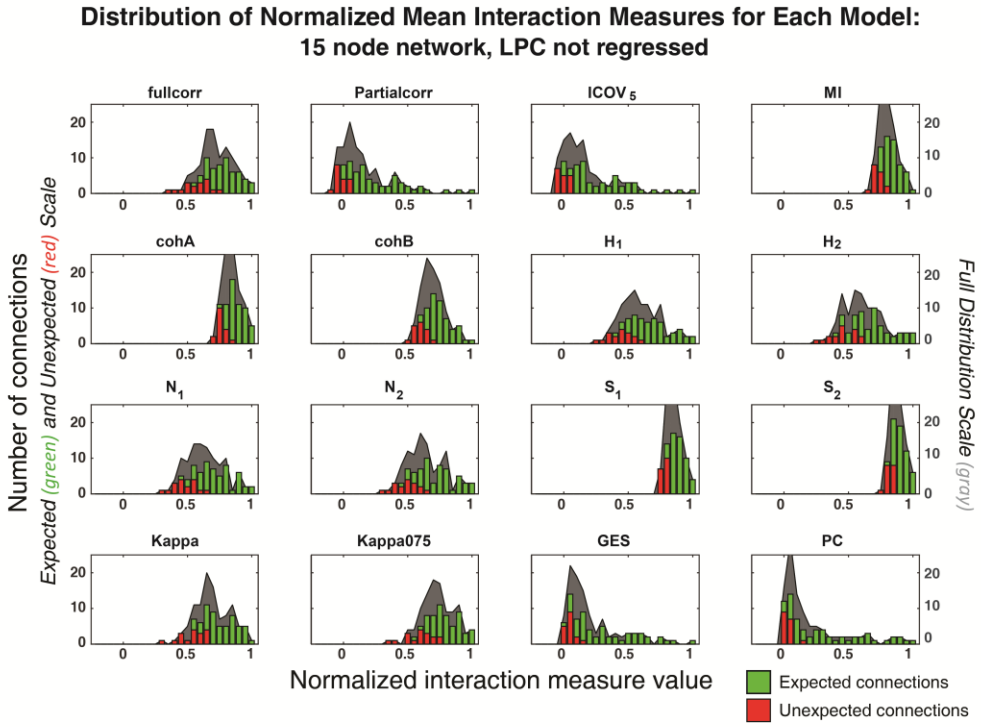
Supplementary Material from Manuscript 2.



Supplementary Fig. 1. Retinotopy and connections tested for the network evaluations. A) Visual areas V1, V2 and V3 were defined by examining the reversal of direction of polar angle phase progression along the surface maps perpendicular to the iso-eccentricity delineation. Voxels from each area were further classified into five eccentricity bins with the boundaries 1.25° to 2.44°, 2.44° to 4.44°, 4.44° to 7.06°, 7.06° to 10.04°, and 10.04° to 14.00°. On the left we see the phase mapping of the functional response to a stimulus changing in terms of polar angle. On the right we see the response to a stimulus changing in terms of eccentricity. B) Connections tested

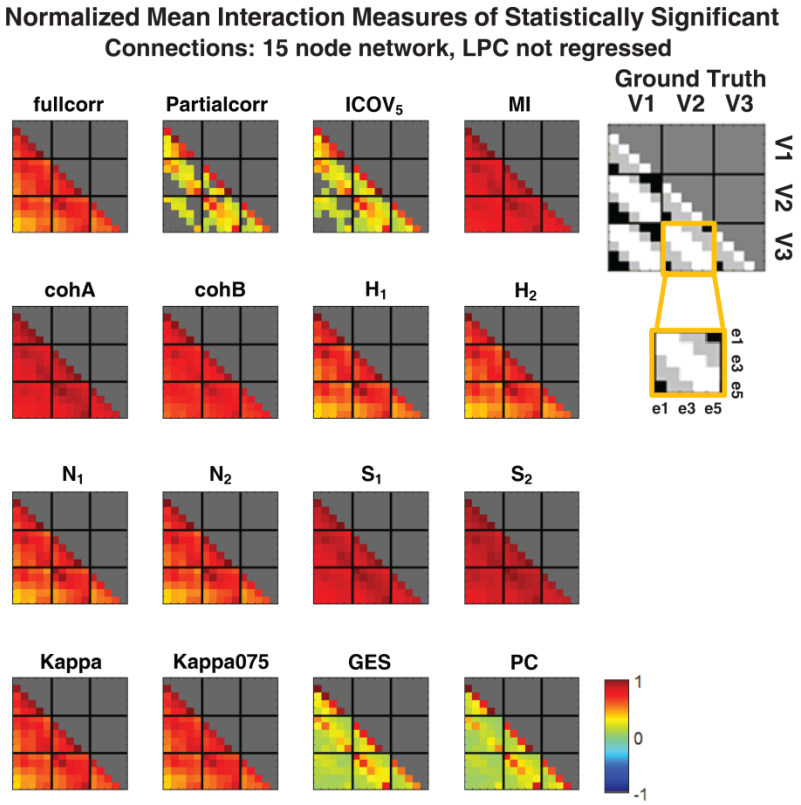
between ROIs for evaluations of accuracy of network predictions. Each hemisphere was divided for analysis according to its dorsal and ventral parts, based on the mapping of the horizontal meridian in V1, resulting in brain regions representing the four quadrants of visual space. Within a quadrant, connections between ROIs of the same eccentricity in 2 different visual areas were expected to be connected, as well as those from adjacent eccentricity regions either within the same visual area or between visual areas (left; connections presented as green lines; all eccentricities are tested for, those shown are just representative). Distant eccentricity regions in 2 different areas were expected to not be directly connected: between e1-e5 (middle; represented in red, orange, pink, and purple lines), as well as between e1-e4 and e2-e5 between V1 and V2 or V1 and V3 (bottom right; represented in brown lines). UVM=upper vertical meridian, HM=horizontal meridian, LVM=lower vertical meridian.

Our 6 subjects:

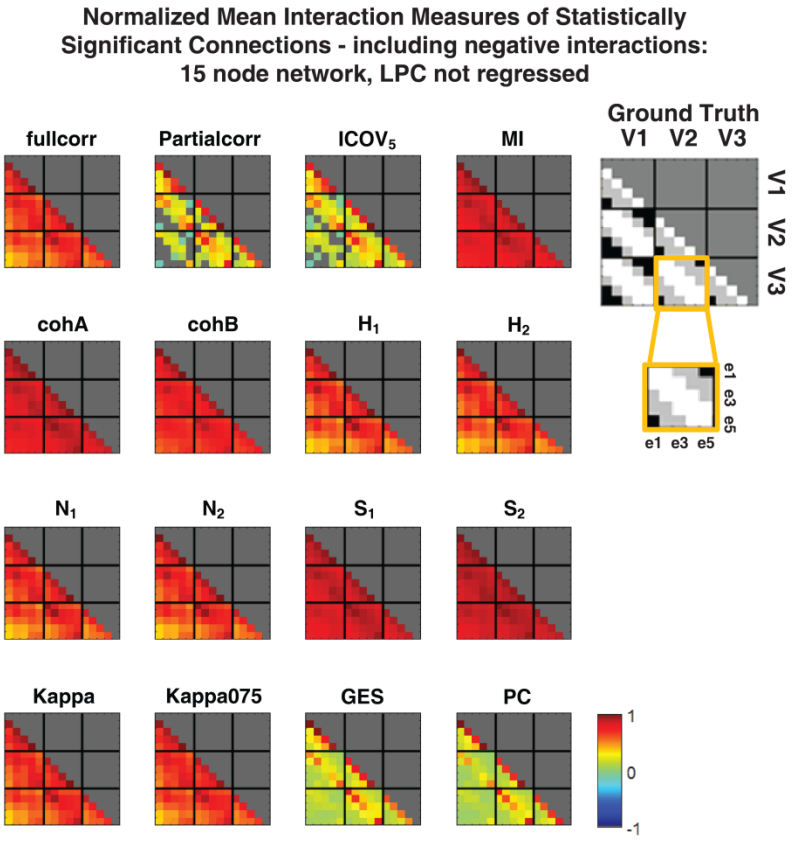


Supplementary Fig. 2. Normalized mean interaction measures for each of the connections tested within the quadrant, for each model (no LPC regression). Normalization is done by dividing all measures by the maximal value, such that for each method the maximal value presented is 1. The gray plot reflects the full distribution of mean interaction measures (over subjects, runs, and quadrants) of all computed connections ([15x15-15]/2=105). The vertical axis scale for gray plots is on the right-hand side of the Supplementary Figures. Green bars reflect the mean

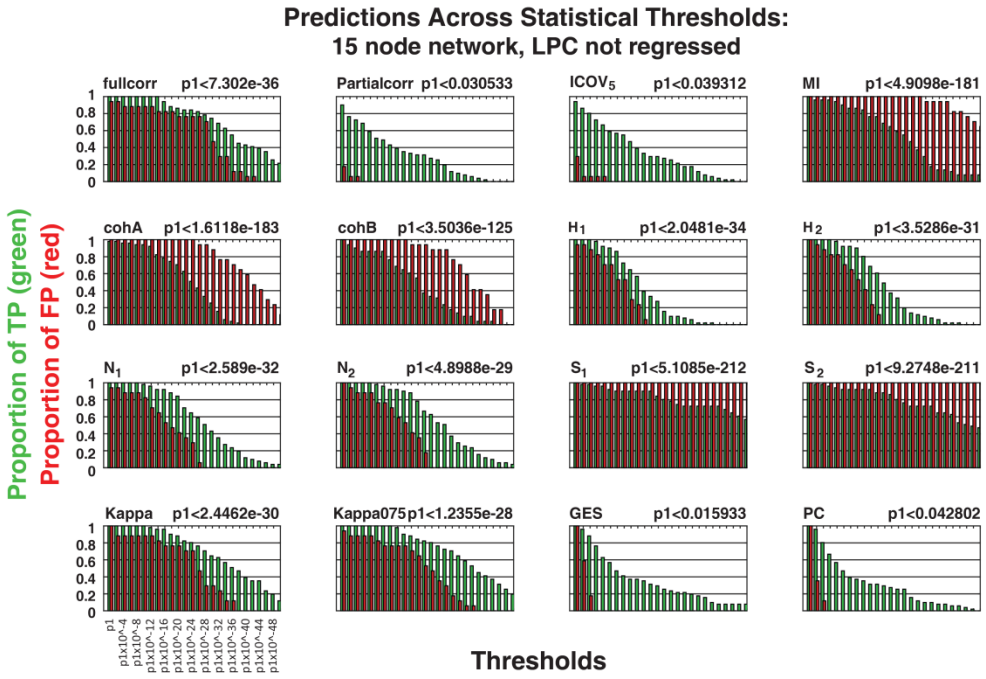
interaction measures associated with the 51 expected connections within the quadrant. Red bars reflect the 17 unexpected connections within the quadrant. Note that all connections are presented, not only those showing statistical significance. The vertical axis scale for the red and green bars is on the left-hand side of the plots.



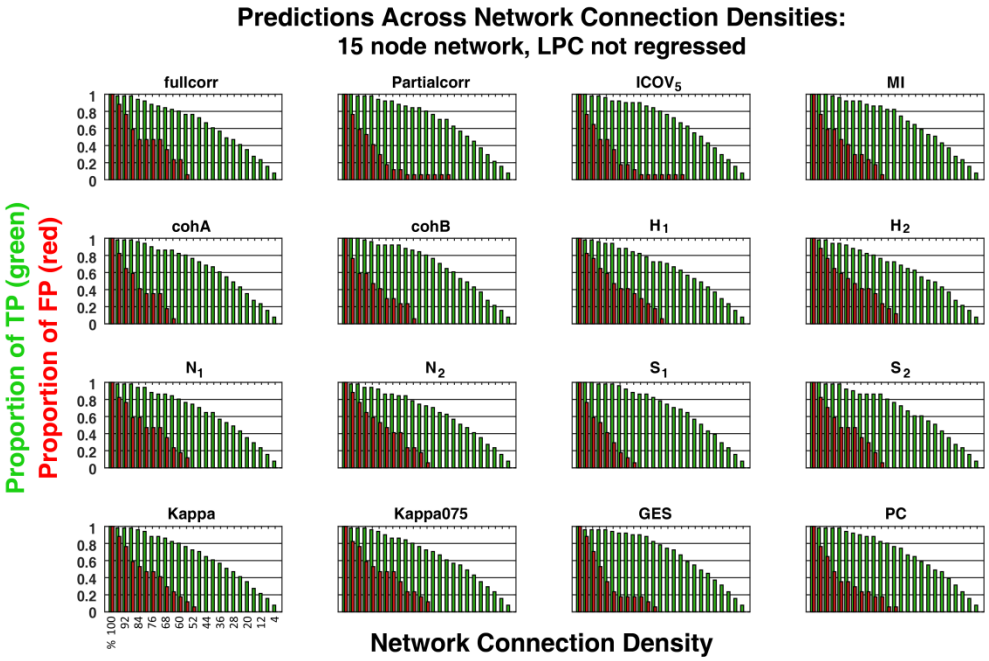
Supplementary Fig. 3. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 15 node, quadrant network without the LPC regressed. Normalization is achieved by dividing all mean measures by the maximal value. The "Ground Truth" matrix in upper right-hand corner has been constructed based on knowledge of the monkey visual cortex, and includes the expected connections (white) and the unexpected connections (black) presented in Supplementary Figure 1. Light gray squares in the ideal matrix are connections on which we place no expectations. Dark grey squares above the diagonal are connections that are not tested. In the model matrices, dark gray represents a connection that is not statistically significant, while a colored square represents a significant connection. Significance is based the FDR p-value threshold given a maximum FPR of 0.05.



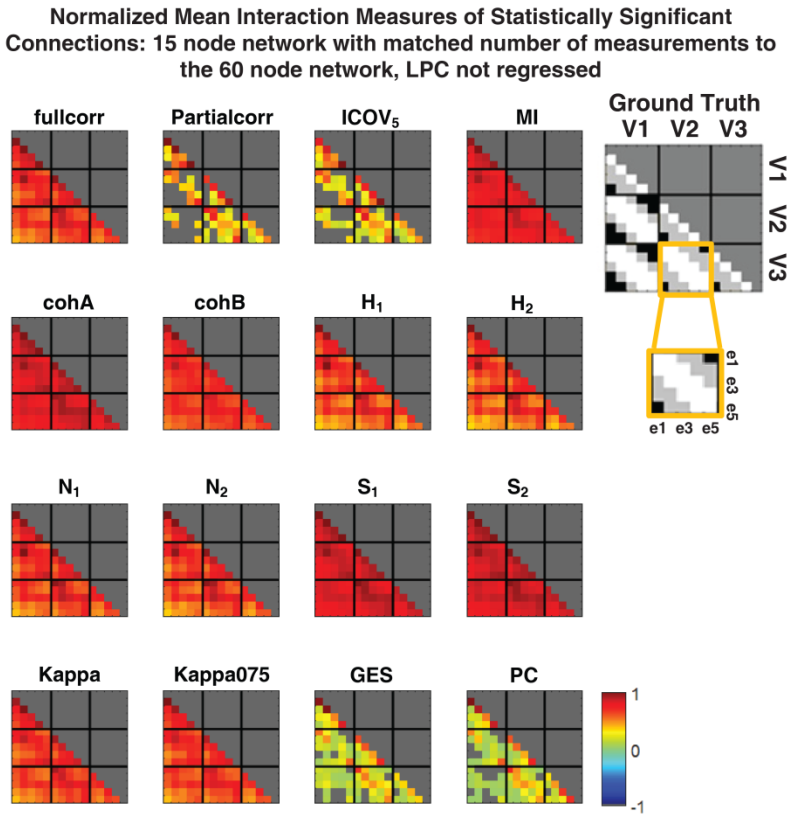
Supplementary Fig. 3-B. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 15 node, quadrant network without the LPC regressed, **including negative interactions**. Presentation is as in Supplementary Figure 3. See "Ground Truth" matrix on top right. Significance is based on the FDR p-value threshold.



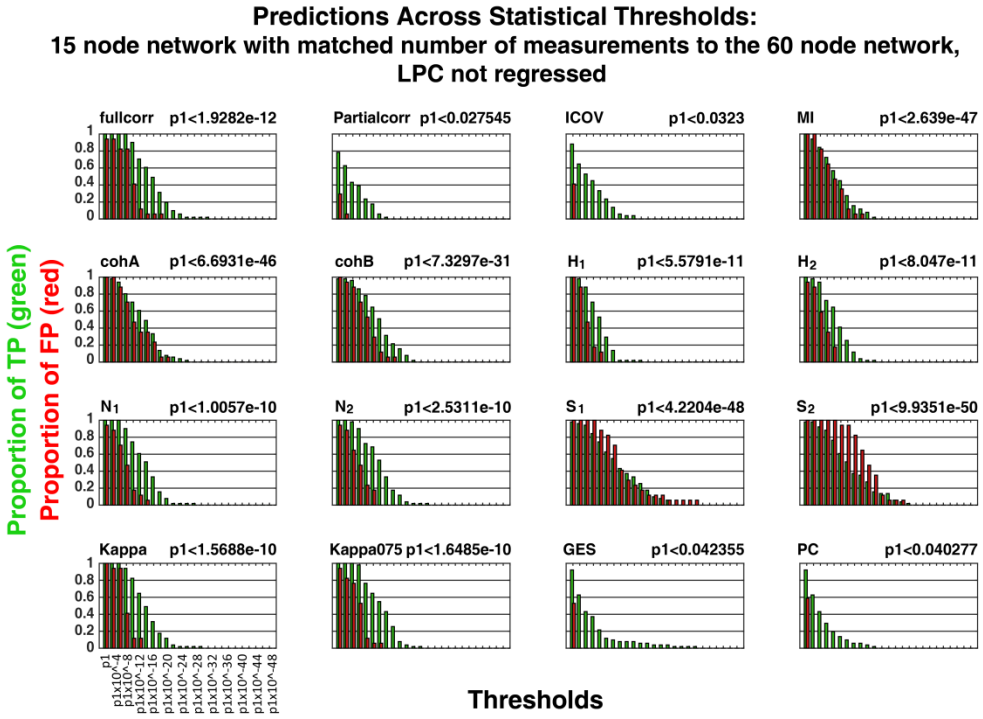
Supplementary Fig. 4. Proportion of true and false positive connections found for each of the models tested at various significance thresholds. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 15 node network within the quadrant with no LPC regression, there are 51 expected connections tested for and 17 unexpected connections. The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The p-value threshold becomes progressively smaller by a multiplication factor of 10^-2 at each step (up to a minimum of 10^-50). For example, Partial correlation's labels along the horizontal axis would read 0.03, 0.03x10^-2, 0.03x10^-4, 0.03x10^-6, 0.03x10^-8,…, 0.03x10^-50.



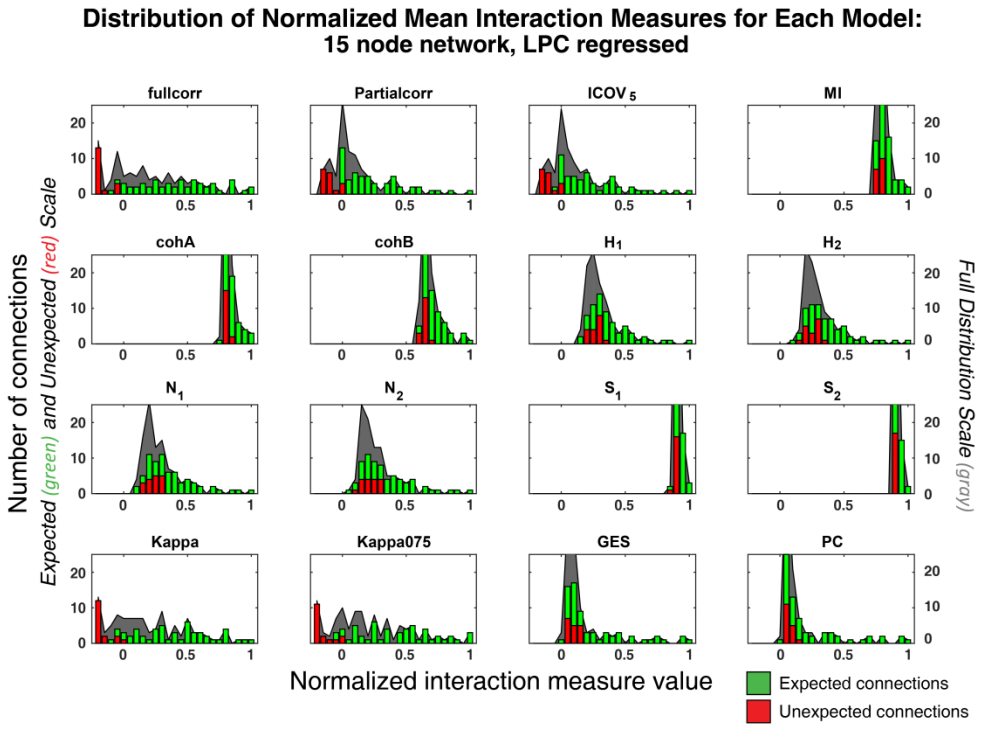
Supplementary Fig. 5. Proportion of true and false positive connections found for each of the models tested at various *network connection densities*. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 15 node network within the quadrant with no LPC regression, there are 51 expected connections tested for and 17 unexpected connections. The network connection densities along the horizontal axis begin at 100% and go down by 4% until reaching 0%.



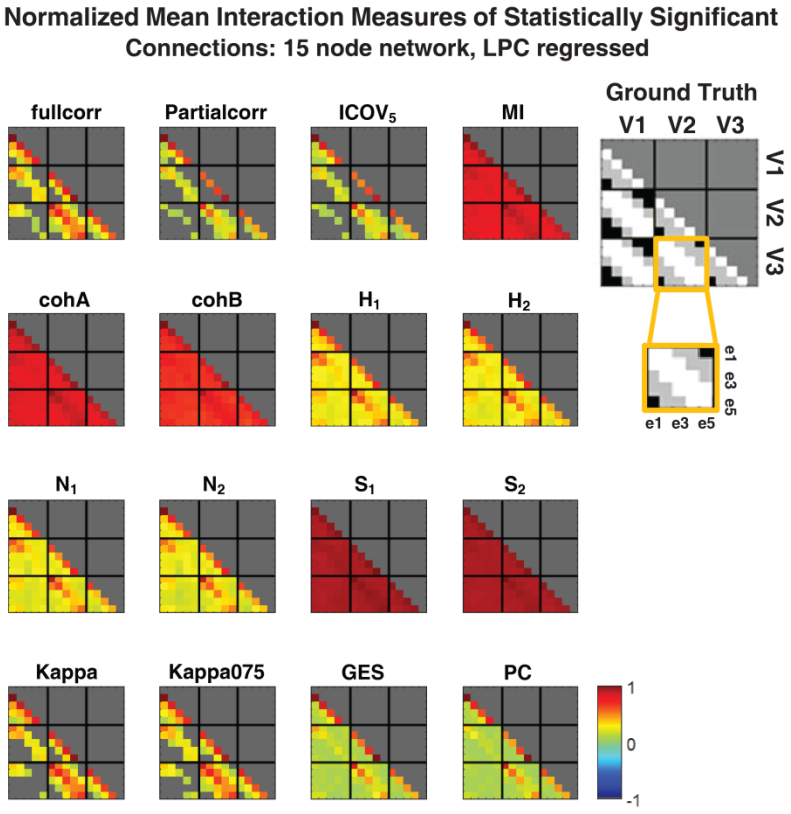
Supplementary Fig. 6. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model. Presented is the 15 node network without LPC regression, *this time using only 8 runs of data per subject* for comparison to the 60 node network, which also has 8 runs of data per subject. Normalization is achieved by dividing all mean measures by the maximal value. In the model matrices, dark gray represents a connection that is not statistically significant, while a colored square represents a significant connection. Presentation is as in Supplementary Figure 3.



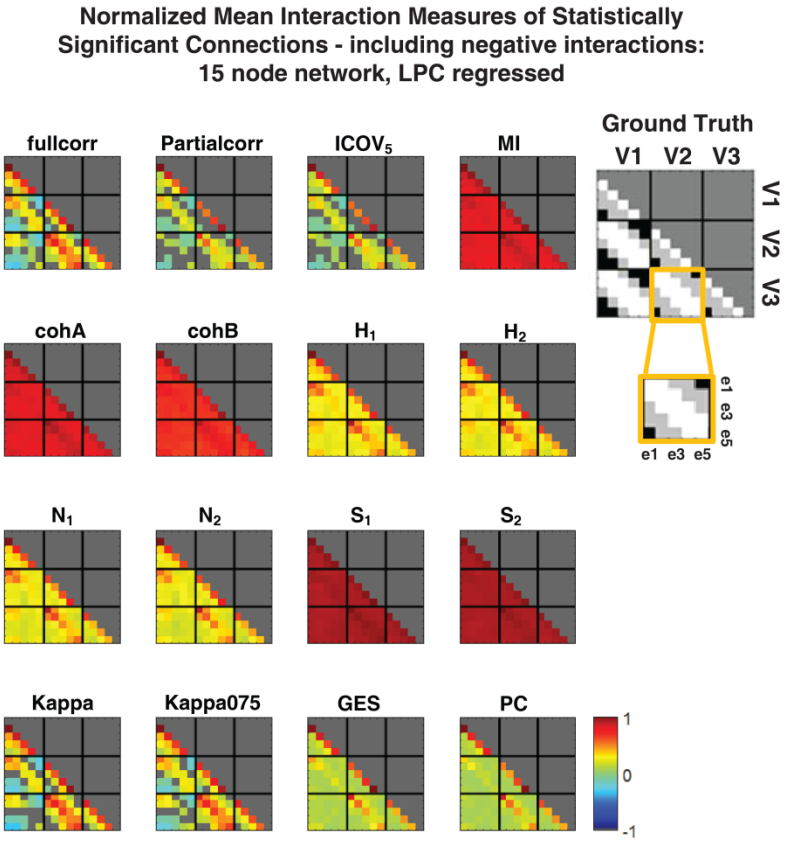
Supplementary Fig.7. Proportion of true and false positive connections found for each of the models tested at various significance thresholds. Green bars are the true positives (TP), and red bars are the false positives (FP). Presented is the 15 node network without LPC regression, *this time using only 8 runs of data per subject* for comparison to the 60 node network, which also has 8 runs of data per subject. The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The threshold p-values become progressively smaller by a multiplication factor of 10^-2 at each step, as described for Supplementary Figure 4. Note that this Supplementary Figure is only referred to when comparing between 15 and 60 node analyses.



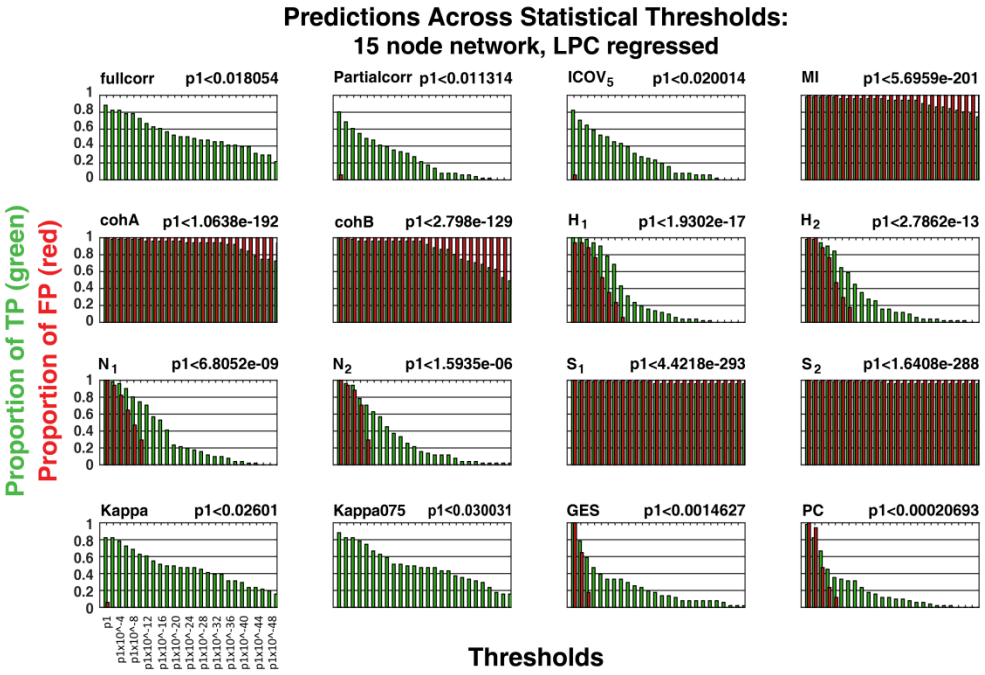
Supplementary Fig. 8. Normalized mean interaction measures for each of the connections tested within the quadrant, for each model (*with LPC regression*). The gray plot reflects the full distribution of mean interaction measures (over subjects, runs, and quadrants) of all computed connections ([15x15-15]/2=105). The vertical axis scale for gray plots is on the right-hand side of the Supplementary Figures. Green bars reflect the mean interaction measures associated with the 51 expected connections within the quadrant. Red bars reflect the 17 unexpected connections within the quadrant. The vertical axis scale for the red and green bars is on the left-hand side of the plots.



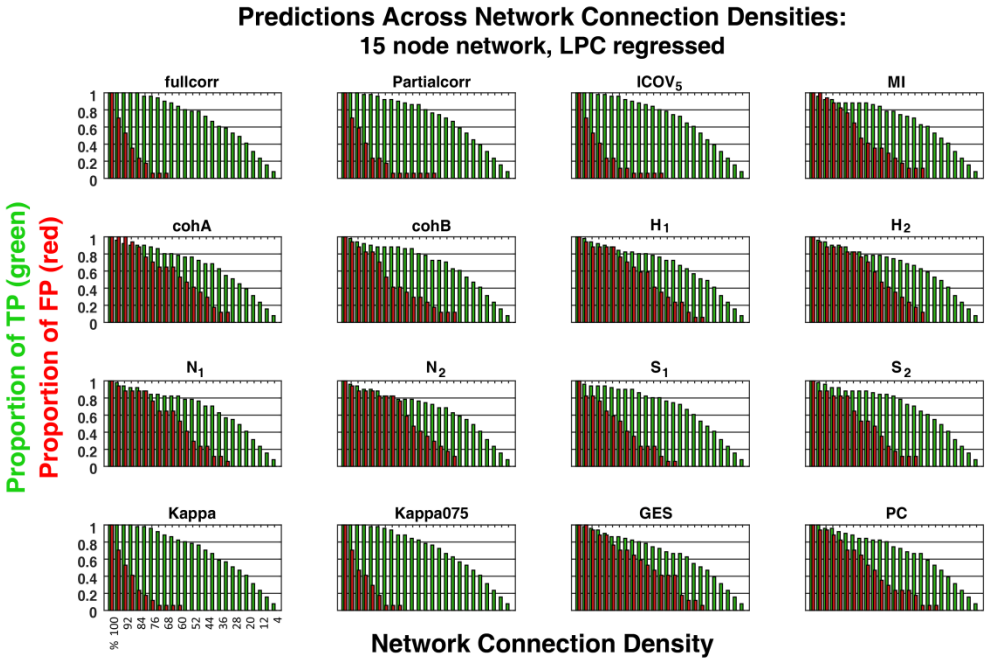
Supplementary Fig. 9. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 15 node, quadrant network *with LPC regressed*. Presentation is as in Supplementary Figure 3. See "Ground Truth" matrix on top right. Significance is based on the FDR p-value threshold.



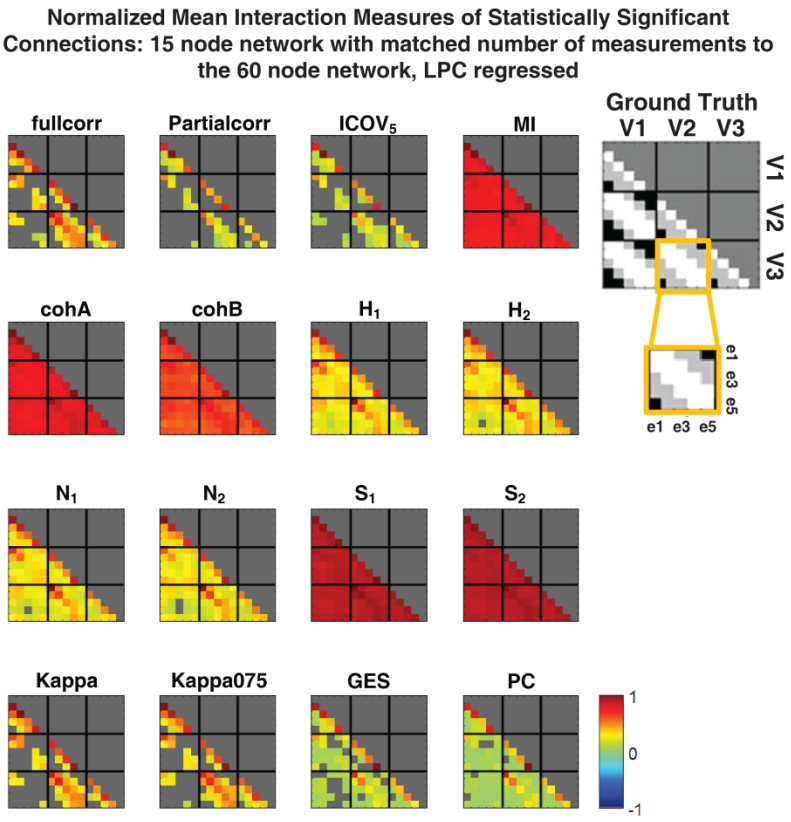
Supplementary Fig. 9-B. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 15 node, quadrant network *with LPC regressed,* **including negative interactions**. Presentation is as in Supplementary Figure 3. See "Ground Truth" matrix on top right. Significance is based on the FDR p-value threshold.



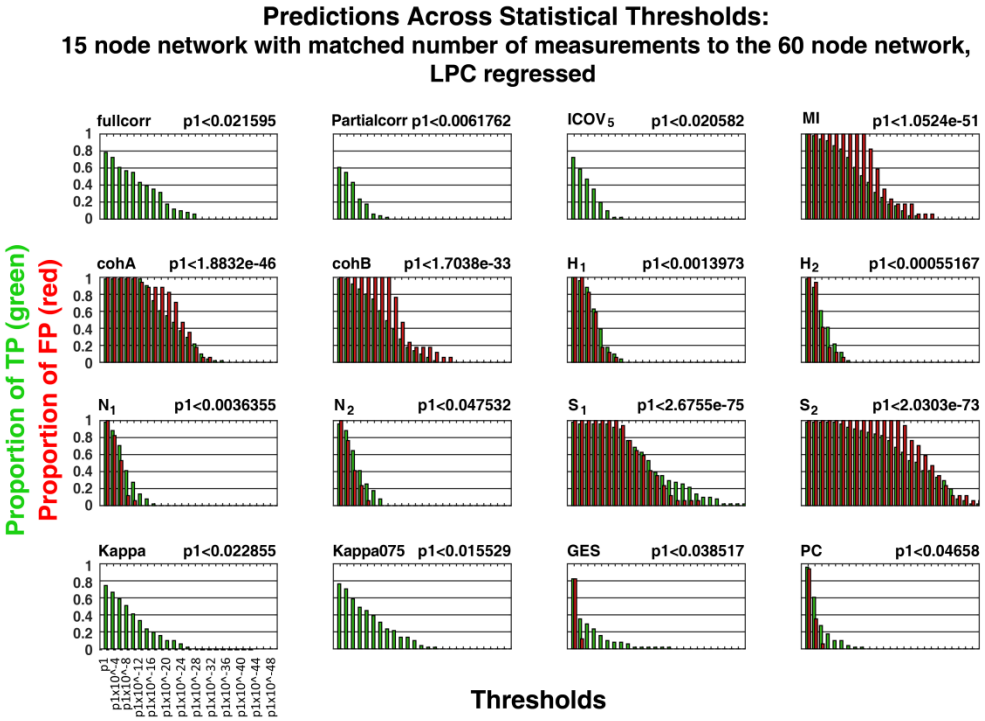
Supplementary Fig. 10. Proportion of true and false positive connections found for each of the models tested at various significance thresholds. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 15 node network within the quadrant *with LPC regressed*, there are 51 expected connections tested for and 17 unexpected connections. The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The threshold p-values become progressively smaller by a multiplication factor of 10^-2 at each step, as described for Supplementary Figure 4.



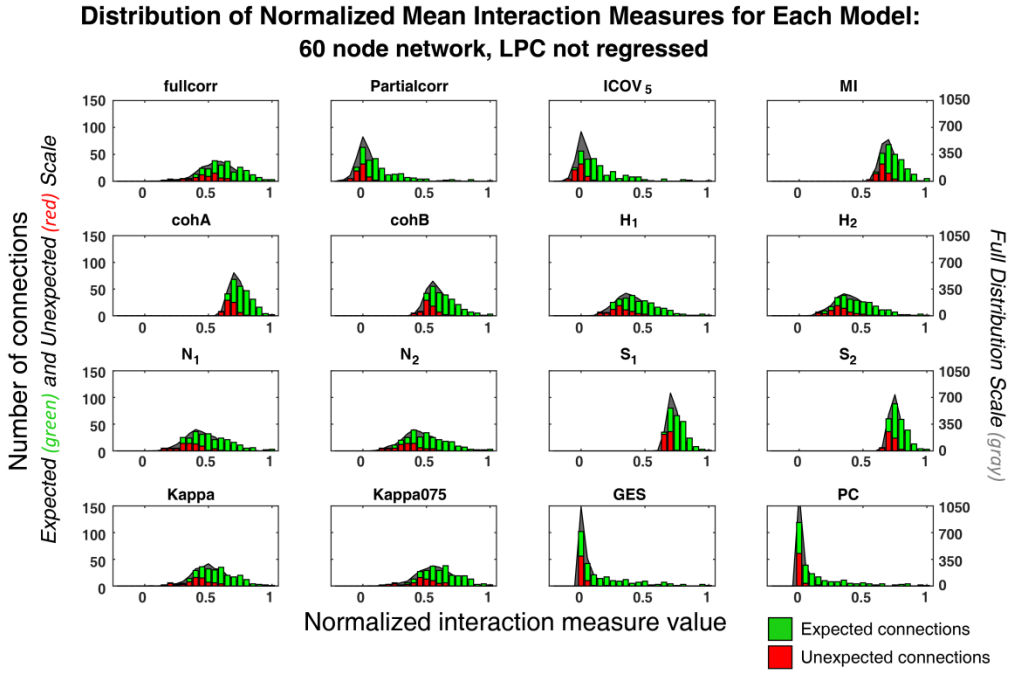
Supplementary Fig. 11. Proportion of true and false positive connections found for each of the models tested at various *network connection densities*. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 15 node network within the quadrant *with LPC regression*, there are 51 expected connections tested for and 17 unexpected connections. The network connection densities along the horizontal axis begin at 100% and go down by 4% until reaching 0%.



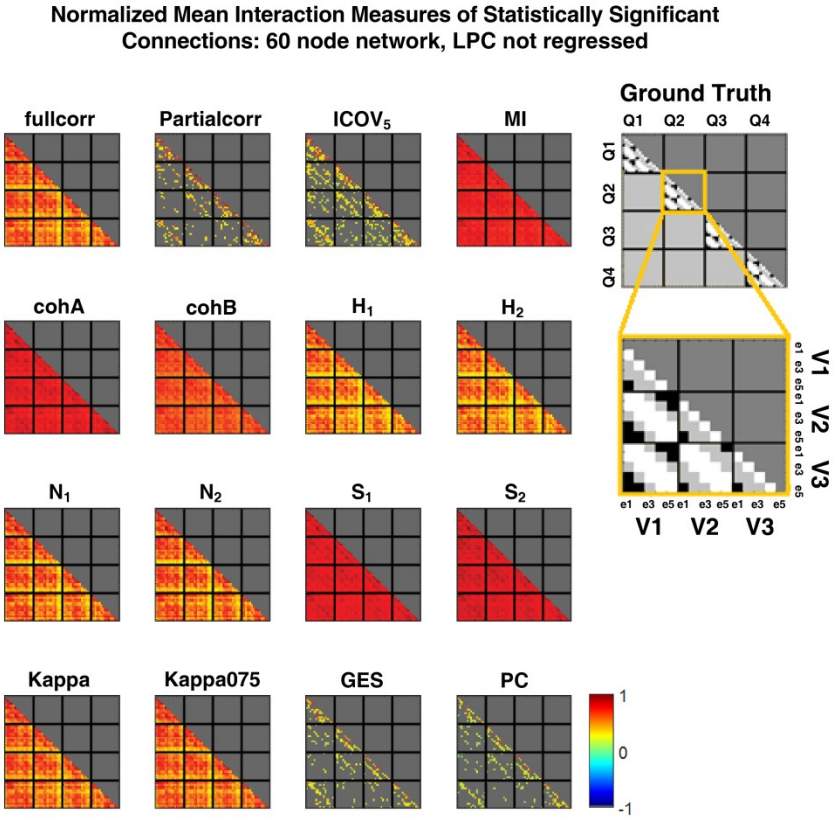
Supplementary Fig. 12. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model. Presented is the 15 node network *with LPC regression*, *this time using only 8 runs of data per subject* for comparison to the 60 node network, which also has 8 runs of data per subject. Normalization is achieved by dividing all mean measures by the maximal value. In the model matrices, dark gray represents a connection that is not statistically significant, while a colored square represents a significant connection. Presentation is as in Supplementary Figure 3.



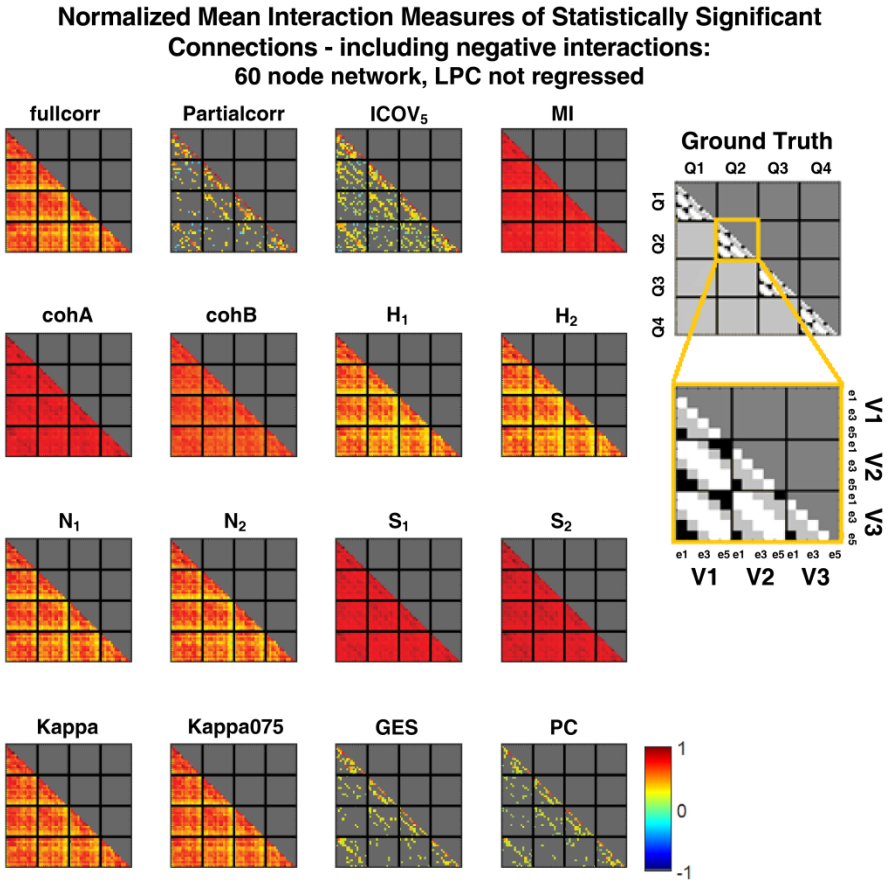
Supplementary Fig.13. Proportion of true and false positive connections found for each of the models tested at various significance thresholds . Green bars are the true positives (TP), and red bars are the false positives (FP). Presented is the 15 node network *with LPC regression*, *this time using only 8 runs of data per subject* for comparison to the 60 node network, which also has 8 runs of data per subject. The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The threshold p-values become progressively smaller by a multiplication factor of 10^-2 at each step, as described for Supplementary Figure 4. Note that this Supplementary Figure is only referred to when comparing between 15 and 60 node analyses.



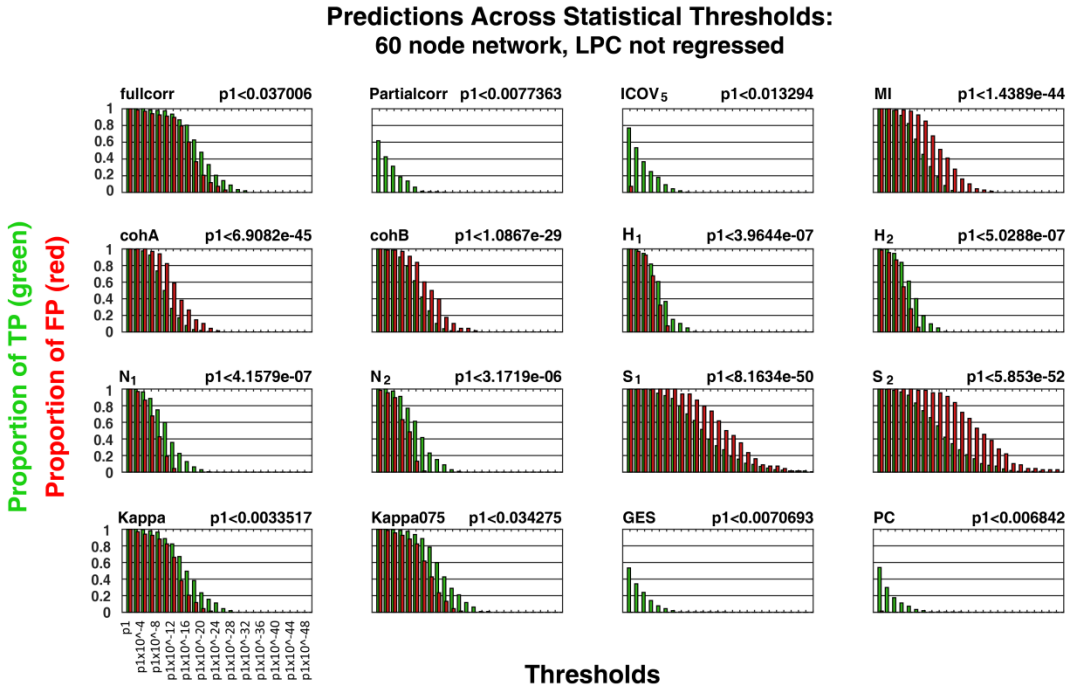
Supplementary Fig. 14. Normalized mean interaction measures for each of the connections tested within all four quadrants, for each model (LPC not regressed). Normalization is done by dividing all measures by the maximal value, such that for each method the maximal value presented is 1. The gray plot reflects the full distribution of mean interaction measures (over subjects and runs) of all computed connections ([60x60-60]/2=1770). The vertical axis scale for gray plots is on the right-hand side of the Supplementary Figures. Green bars reflect the mean interaction measures associated with the 51x4=204 expected connections within the quadrants. Red bars reflect the 17x4=68 unexpected connections within the quadrants. The vertical axis scale for the red and green bars is on the left-hand side of the plots.



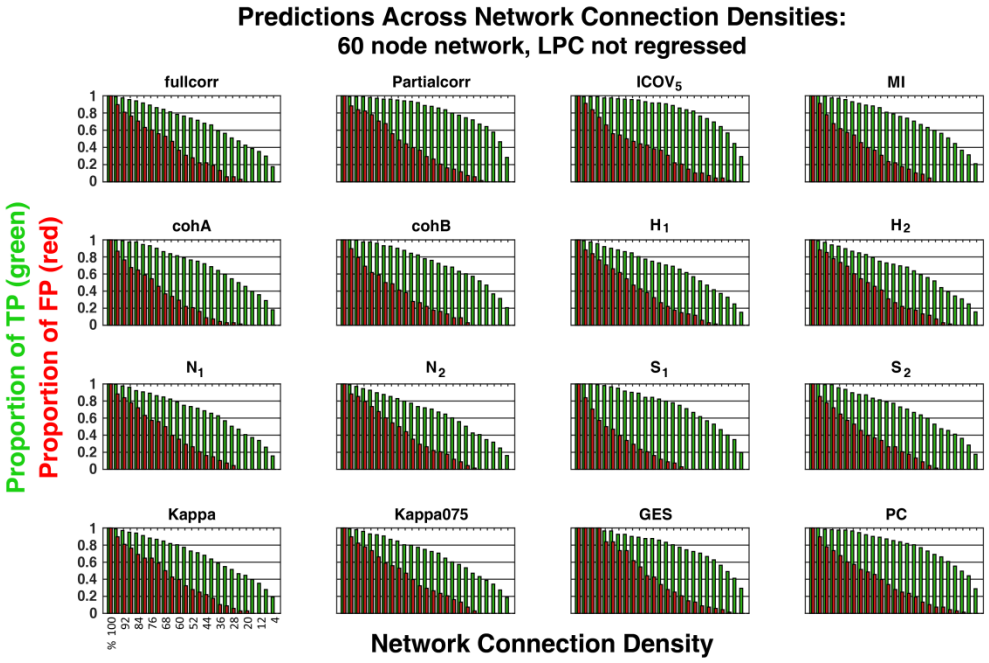
Supplementary Fig. 15. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 60 node, full visual cortex network without the LPC regressed. Presentation is as in Supplementary Figure 3. See "Ground Truth" matrix on top right. Significance is based on the FDR p-value threshold. Q=Quadrant.



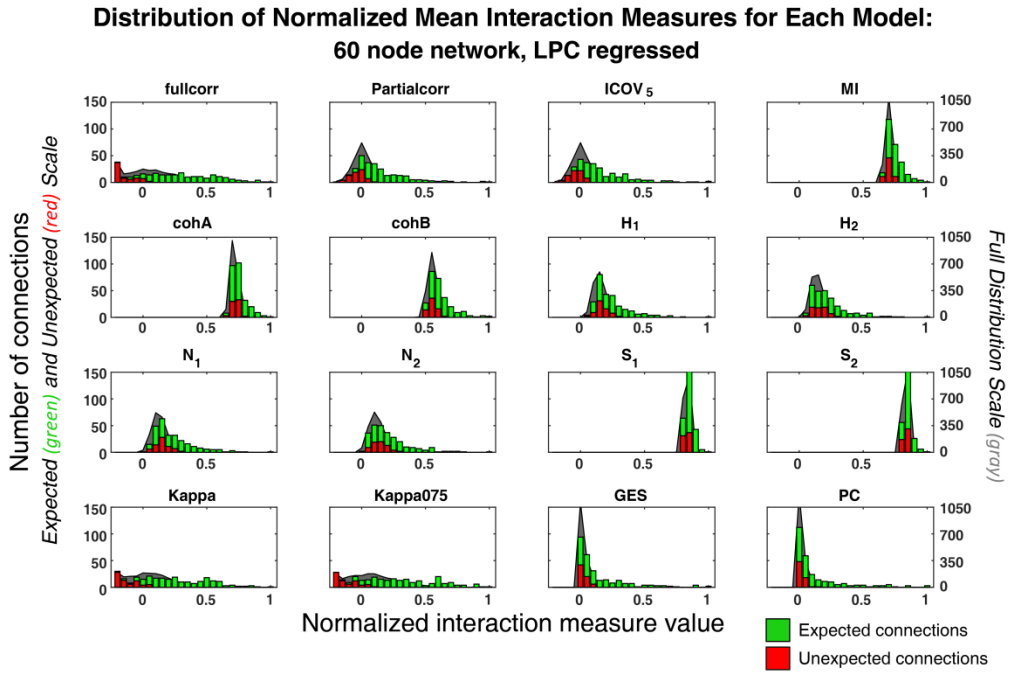
Supplementary Fig. 15-B. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 60 node, full visual cortex network without the LPC regressed, **including negative interactions**. Presentation is as in Supplementary Figure 3. See "Ground Truth" matrix on top right. Significance is based on the FDR p-value threshold.



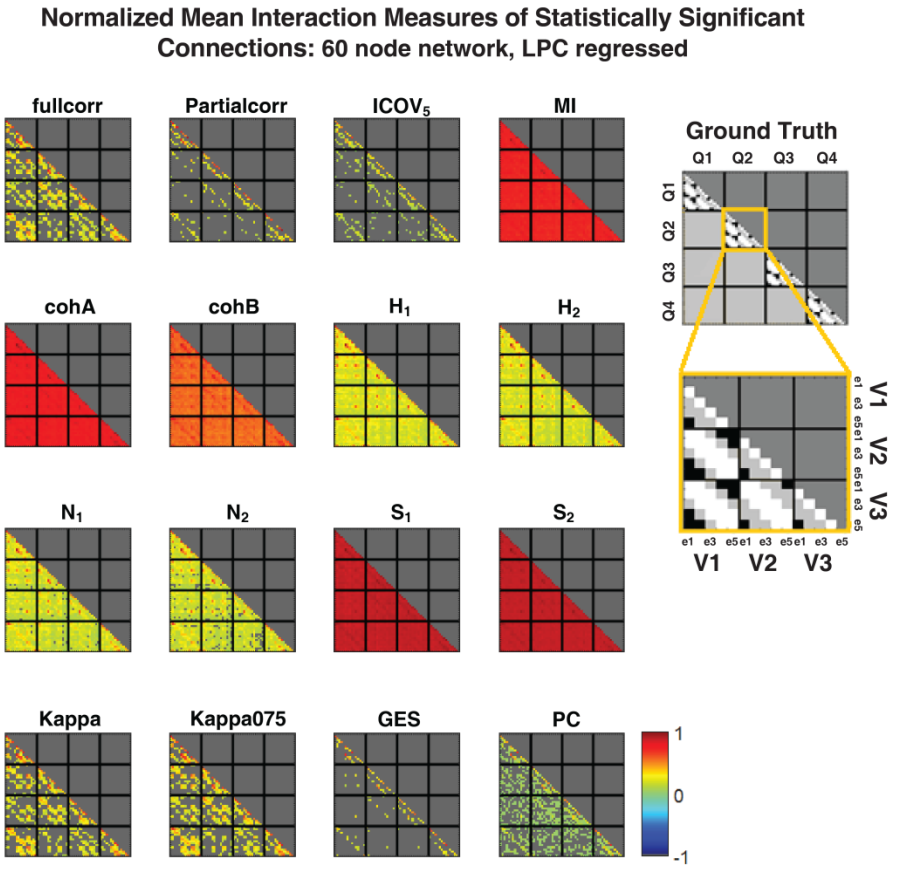
Supplementary Fig. 16. Proportion of true and false positive connections found for each of the models tested at various significance thresholds. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 60 node network without the LPC regressed, there are 51×4=204 expected connections tested for and 17×4=68 unexpected connections. The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The threshold p-values become progressively smaller by a multiplication factor of 10^-2 at each step, as described for Supplementary Figure 4.



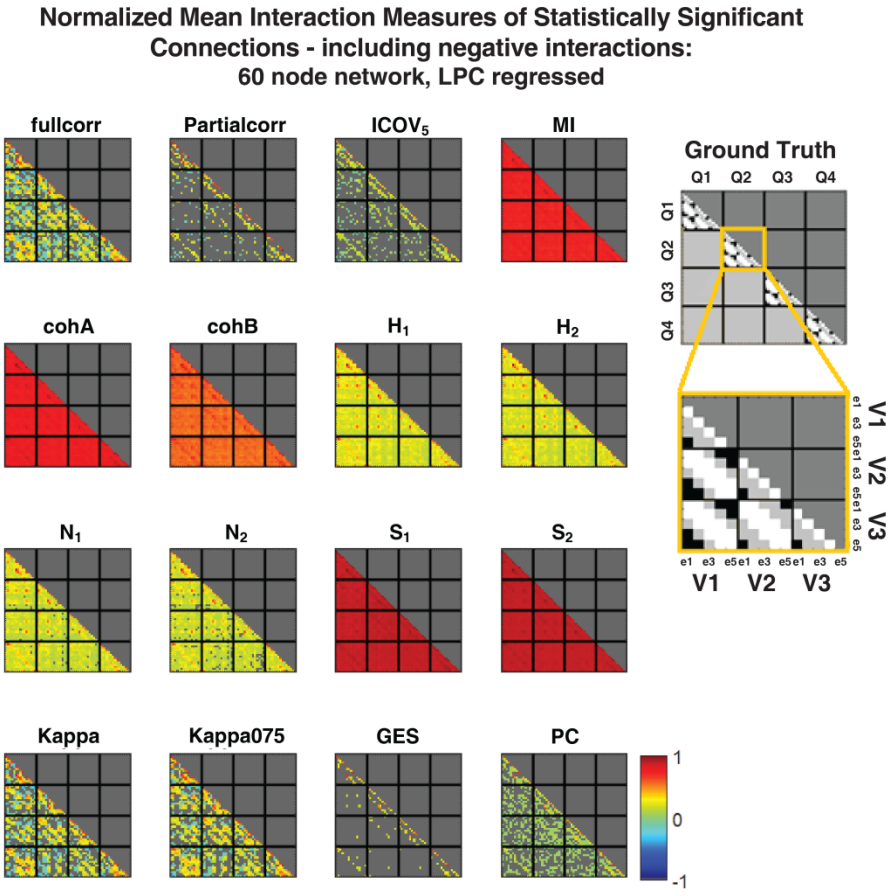
Supplementary Fig. 17. Proportion of true and false positive connections found for each of the models tested at various *network connection densities*. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 60 node network without LPC regression, there are 51×4=204 expected connections tested for and 17×4=68 unexpected connections. The network connection densities along the horizontal axis begin at 100% and go down by 4% until reaching 0%.



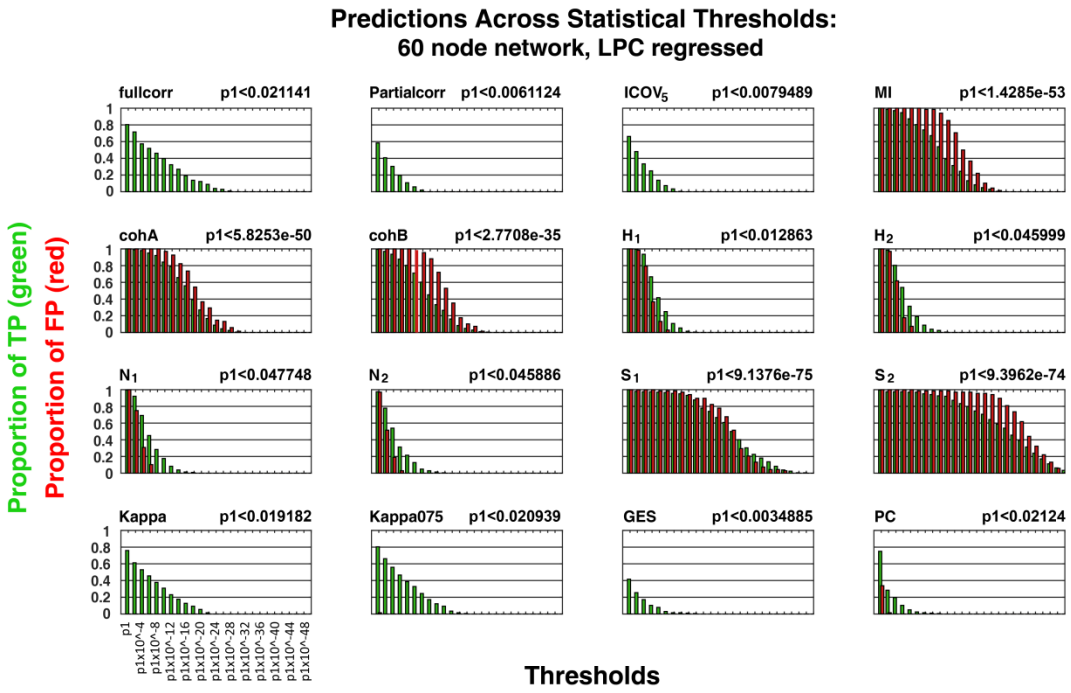
Supplementary Fig. 18. Normalized mean interaction measures for each of the connections tested within all four quadrants, for each model (*LPC regressed*). Normalization is done by dividing all measures by the maximal value, such that for each method the maximal value presented is 1. The gray plot reflects the full distribution of mean interaction measures (over subjects and runs) of all computed connections ([60x60-60]/2=1770). The vertical axis scale for gray plots is on the right-hand side of the Supplementary Figures. Green bars reflect the mean interaction measures associated with the 51x4=204 expected connections within the quadrants. Red bars reflect the 17x4=68 unexpected connections within the quadrants. The vertical axis scale for the red and green bars is on the left-hand side of the plots.



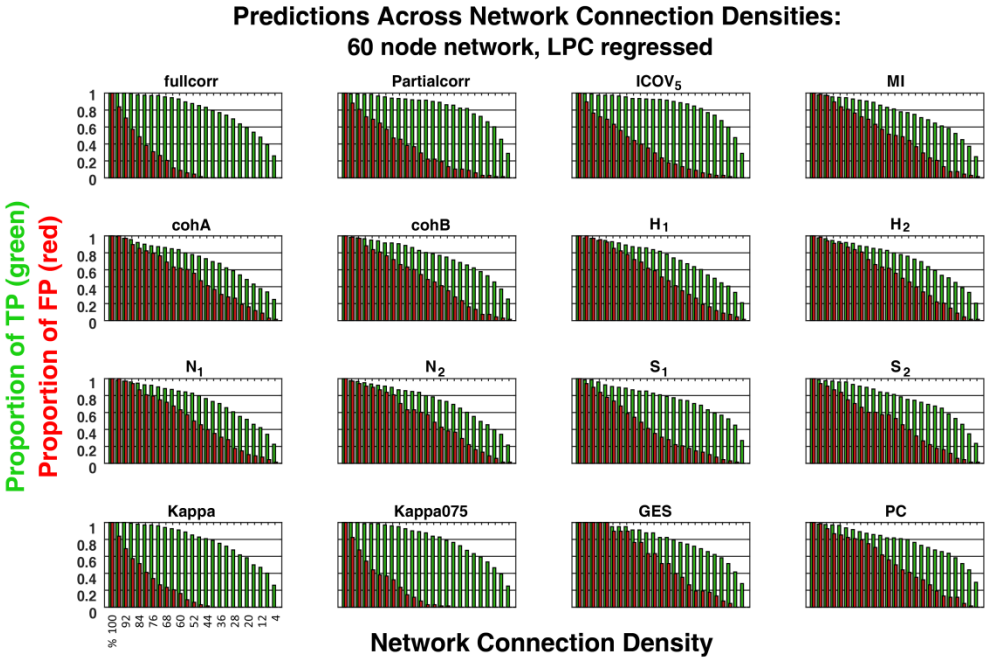
Supplementary Fig. 19. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 60 node, full visual cortex network *with LPC regressed*. Presentation is as in Supplementary Figure 3. See "Ground Truth" matrix on top right. Significance is based on the FDR p-value threshold. Q=Quadrant.



Supplementary Fig. 19-B. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 60 node, full visual cortex network *with LPC regressed,* **including negative interactions**. Presentation is as in Supplementary Figure 3. See "Ground Truth" matrix on top right. Significance is based on the FDR p-value threshold.

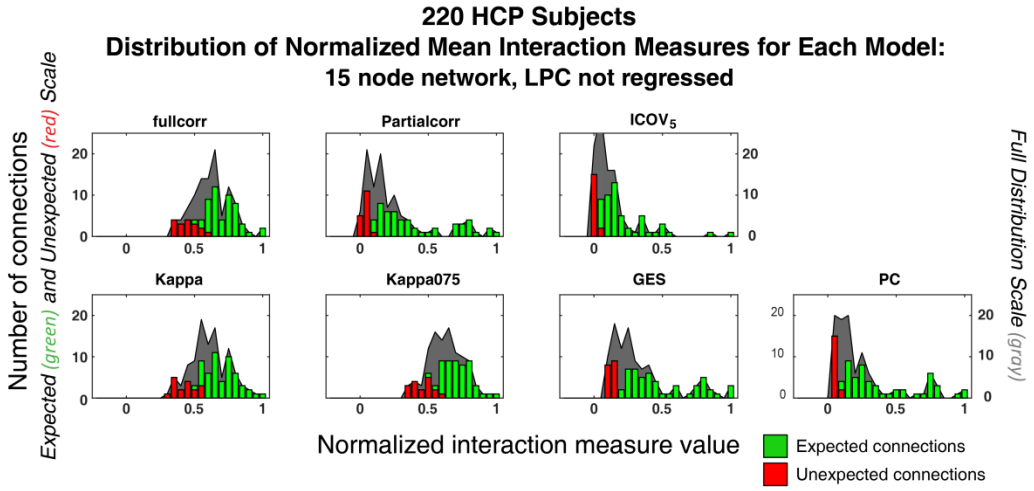


Supplementary Fig. 20. Proportion of true and false positive connections found for each of the models tested at various significance thresholds. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 60 node network *with LPC regressed*, there are 51×4=204 expected connections tested for and 17×4=68 unexpected connections. The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The threshold p-values become progressively smaller by a multiplication factor of 10^-2 at each step, as described for Supplementary Figure 4.

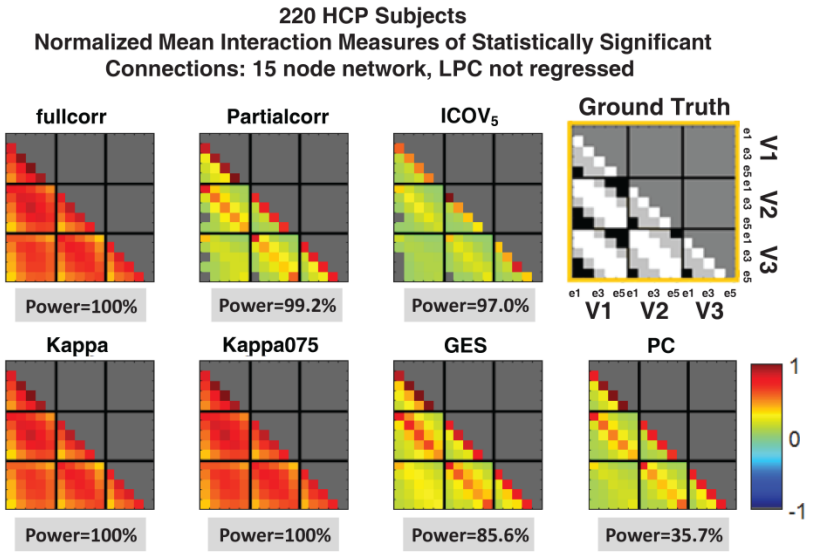


Supplementary Fig. 21. Proportion of true and false positive connections found for each of the models tested at various *network connection densities*. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 60 node network *with LPC regression*, there are 51×4=204 expected connections tested for and 17×4=68 unexpected connections. The network connection densities along the horizontal axis begin at 100% and go down by 4% until reaching 0%.

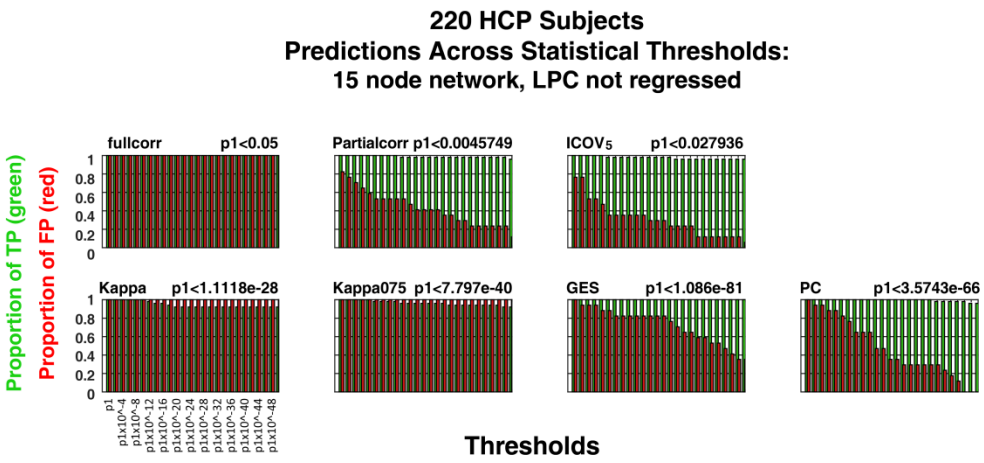
HCP 220 subjects:



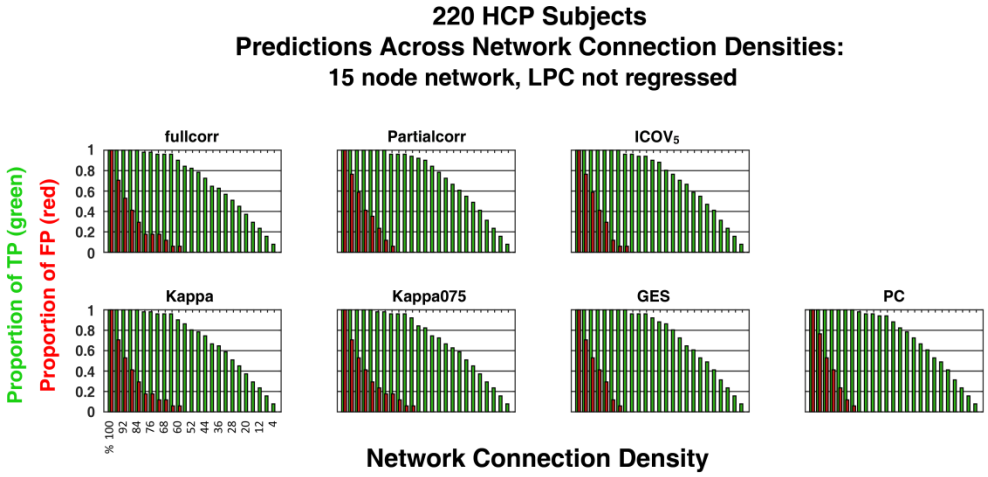
Supplementary Fig. 22. Normalized mean interaction measures for each of the connections tested within the quadrant **using 220 HCP subjects**, for each model (no LPC regression). The gray plot reflects the full distribution of mean interaction measures (over subjects, runs, and quadrants) of all computed connections ([15x15-15]/2=105). The vertical axis scale for gray plots is on the right-hand side of the Supplementary Figures. Green bars reflect the mean interaction measures associated with the 51 expected connections within the quadrant. Red bars reflect the 17 unexpected connections within the quadrant. The vertical axis scale for the red and green bars is on the left-hand side of the plots.



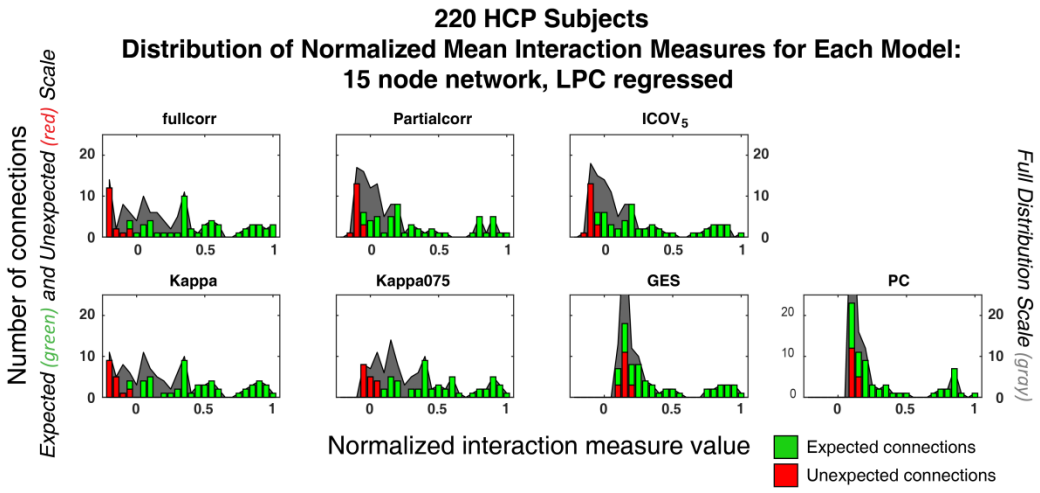
Supplementary Fig. 23. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 15 node, quadrant network **using 220 HCP subjects** without the LPC regressed. Presentation is as in Supplementary Figure 3. See "Ground Truth" matrix on top right. Significance is based on the FDR p-value threshold.



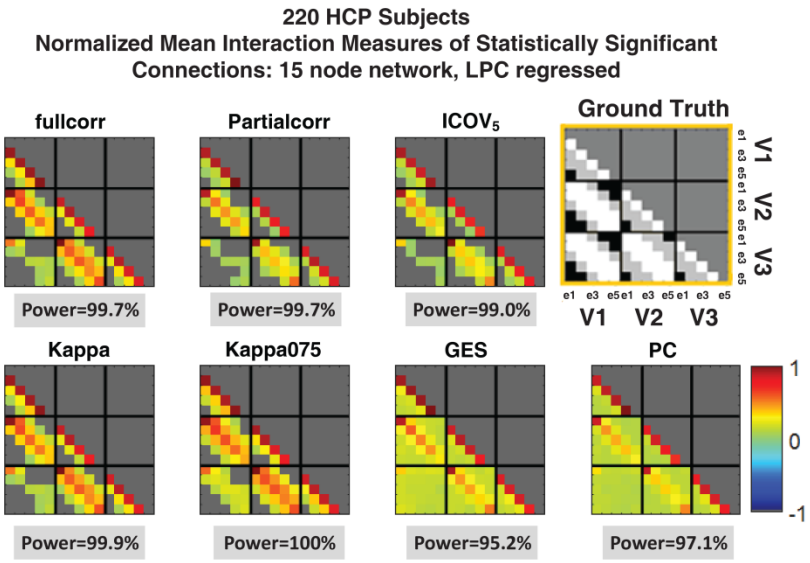
Supplementary Fig. 24. Proportion of true and false positive connections found for each of the models tested at various significance thresholds **using 220 HCP subjects**. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 15 node network within the quadrant without the LPC regressed, there are 51 expected connections tested for and 17 unexpected connections. The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The threshold p-values become progressively smaller by a multiplication factor of 10^-2 at each step, as described for Supplementary Figure 4.



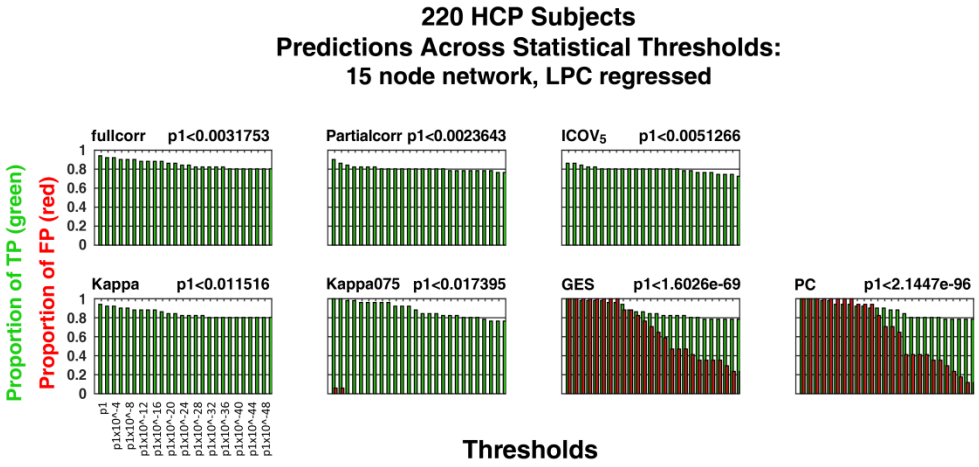
Supplementary Fig. 25. Proportion of true and false positive connections found for each of the models tested at various *network connection densities* **using 220 HCP subjects**. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 15 node network within the quadrant with no LPC regression, there are 51 expected connections tested for and 17 unexpected connections. The network connection densities along the horizontal axis begin at 100% and go down by 4% until reaching 0%.



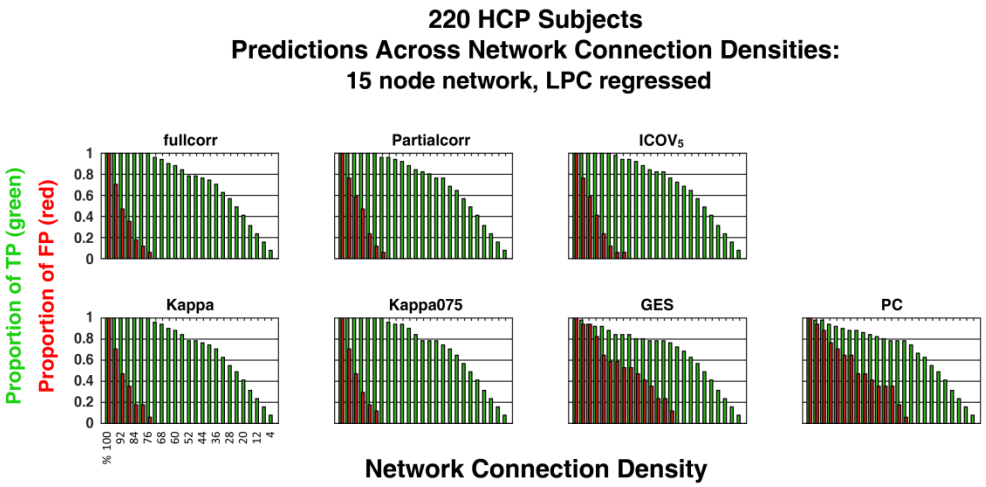
Supplementary Fig. 26. Normalized mean interaction measures for each of the connections tested within the quadrant **using 220 HCP subjects**, for each model (*with LPC regression*). The gray plot reflects the full distribution of mean interaction measures (over subjects, runs, and quadrants) of all computed connections ([15x15-15]/2=105). The vertical axis scale for gray plots is on the right-hand side of the Supplementary Figures. Green bars reflect the mean interaction measures associated with the 51 expected connections within the quadrant. Red bars reflect the 17 unexpected connections within the quadrant. The vertical axis scale for the red and green bars is on the left-hand side of the plots.



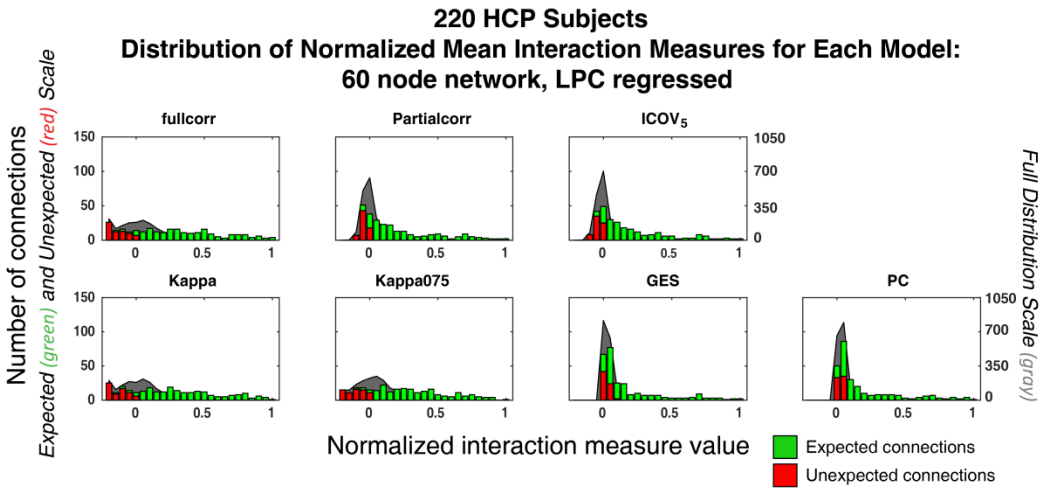
Supplementary Fig. 27. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 15 node, quadrant network **using 220 HCP subjects** *with LPC regressed*. Presentation is as in Supplementary Figure 3. See "Ground Truth" matrix on top right. Significance is based on the FDR p-value threshold.



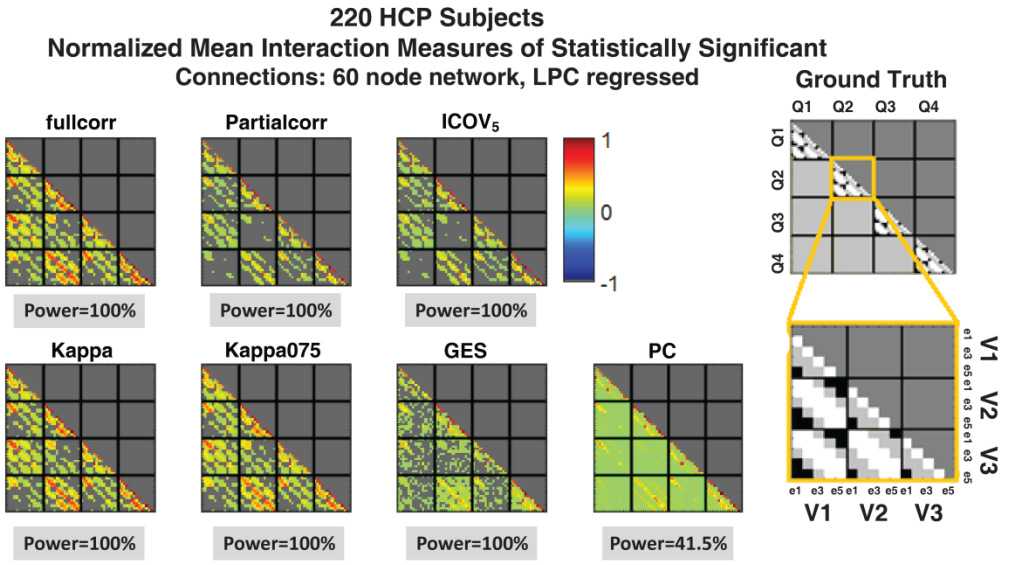
Supplementary Fig. 28. Proportion of true and false positive connections found for each of the models tested at various significance thresholds **using 220 HCP subjects**. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 15 node network within the quadrant *with LPC regressed*, there are 51 expected connections tested for and 17 unexpected connections. The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The threshold p-values become progressively smaller by a multiplication factor of 10^-2 at each step, as described for Supplementary Figure 4.



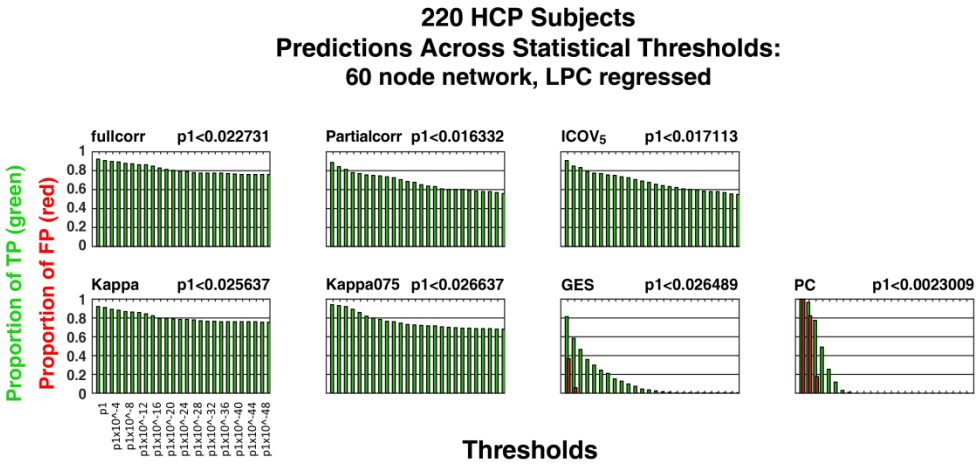
Supplementary Fig. 29. Proportion of true and false positive connections found for each of the models tested at various *network connection densities* **using 220 HCP subjects**. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 15 node network within the quadrant *with LPC regression*, there are 51 expected connections tested for and 17 unexpected connections. The network connection densities along the horizontal axis begin at 100% and go down by 4% until reaching 0%.



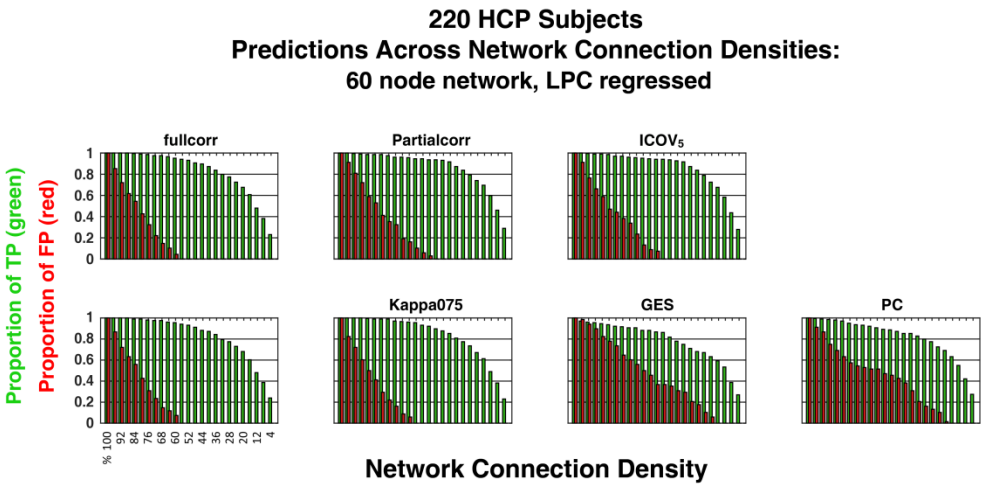
Supplementary Fig. 30. Normalized mean interaction measures for each of the connections tested within all four quadrants **using 220 HCP subjects**, for each model (*LPC regressed*). Normalization is done by dividing all measures by the maximal value, such that for each method the maximal value presented is 1. The gray plot reflects the full distribution of mean interaction measures (over subjects and runs) of all computed connections ([60x60-60]/2=1770). The vertical axis scale for gray plots is on the right-hand side of the Supplementary Figures. Green bars reflect the mean interaction measures associated with the 51x4=204 expected connections within the quadrants. Red bars reflect the 17x4=68 unexpected connections within the quadrants. The vertical axis scale for the red and green bars is on the left-hand side of the plots.



Supplementary Fig. 31. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 60 node, full visual cortex network **using 220 HCP subjects** *with LPC regressed*. Presentation is as in Supplementary Figure 3. See "Ground Truth" matrix on top right. Significance is based on the FDR p-value threshold. Q=Quadrant.

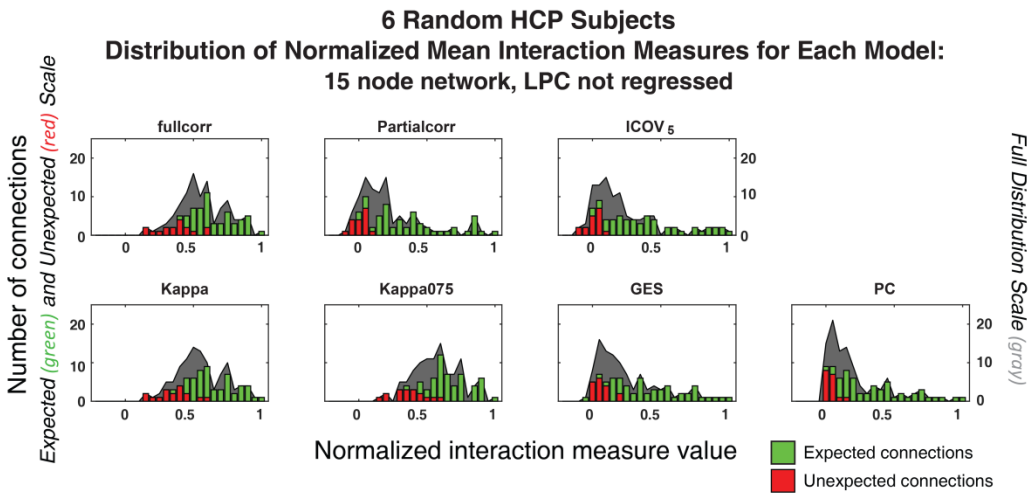


Supplementary Fig. 32. Proportion of true and false positive connections found for each of the models tested at various significance thresholds **using 220 HCP subjects**. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 60 node network *with LPC regressed*, there are 51×4=204 expected connections tested for and 17×4=68 unexpected connections. The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The threshold p-values become progressively smaller by a multiplication factor of 10^-2 at each step, as described for Supplementary Figure 4.

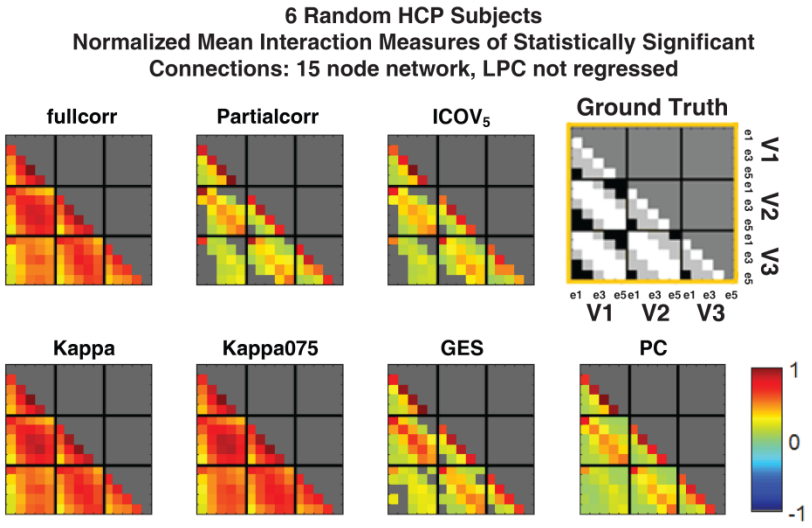


Supplementary Fig. 33. Proportion of true and false positive connections found for each of the models tested at various *network connection densities* **using 220 HCP subjects**. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 60 node network *with LPC regression*, there are 51×4=204 expected connections tested for and 17×4=68 unexpected connections. The network connection densities along the horizontal axis begin at 100% and go down by 4% until reaching 0%.

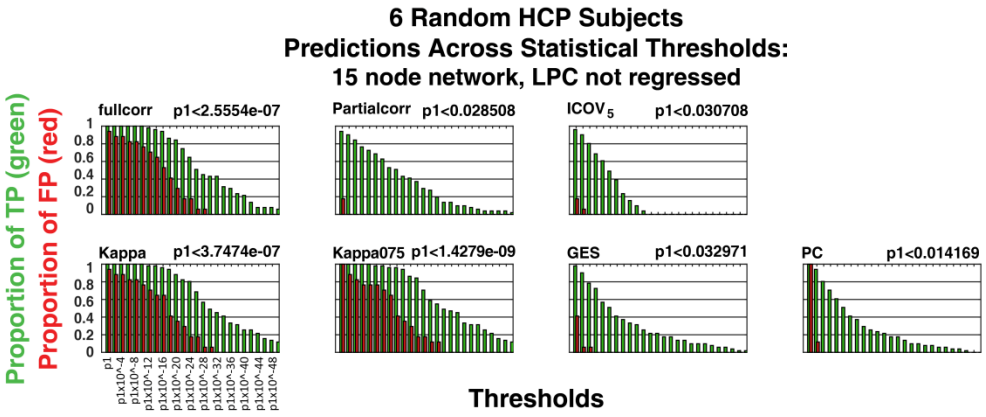
HCP 6 subjects:



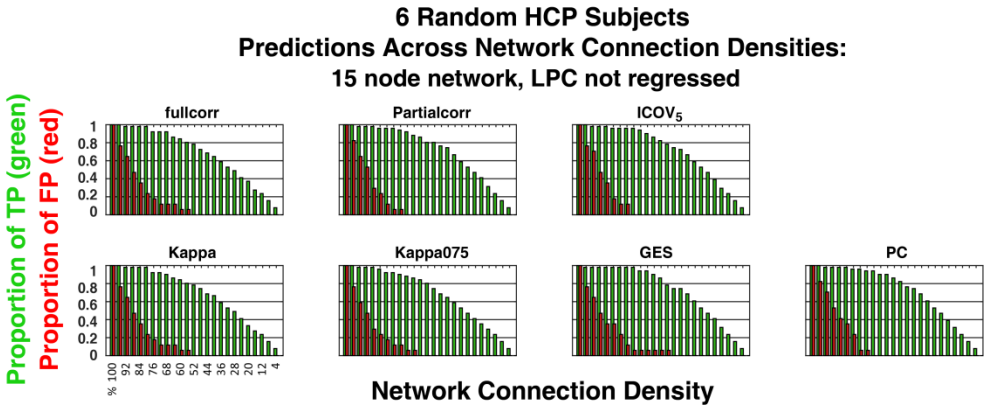
Supplementary Fig. 34. Normalized mean interaction measures for each of the connections tested within the quadrant **using 6 random HCP subjects**, for each model (no LPC regression). The gray plot reflects the full distribution of mean interaction measures (over subjects, runs, and quadrants) of all computed connections ([15x15-15]/2=105). The vertical axis scale for gray plots is on the right-hand side of the Supplementary Figures. Green bars reflect the mean interaction measures associated with the 51 expected connections within the quadrant. Red bars reflect the 17 unexpected connections within the quadrant. The vertical axis scale for the red and green bars is on the left-hand side of the plots.



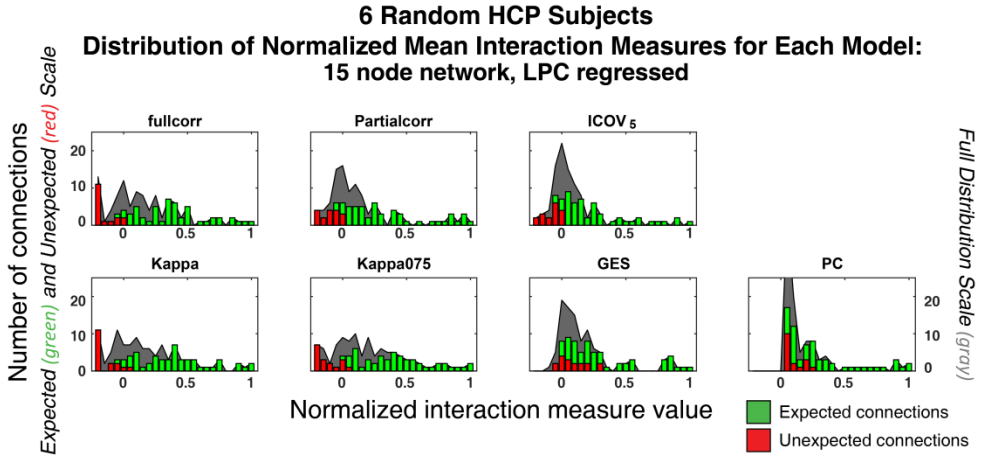
Supplementary Fig. 35. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 15 node, quadrant network **using 6 random HCP subjects** without the LPC regressed. Presentation is as in Supplementary Figure 3. See "Ground Truth" matrix on top right. Significance is based on the FDR p-value threshold.



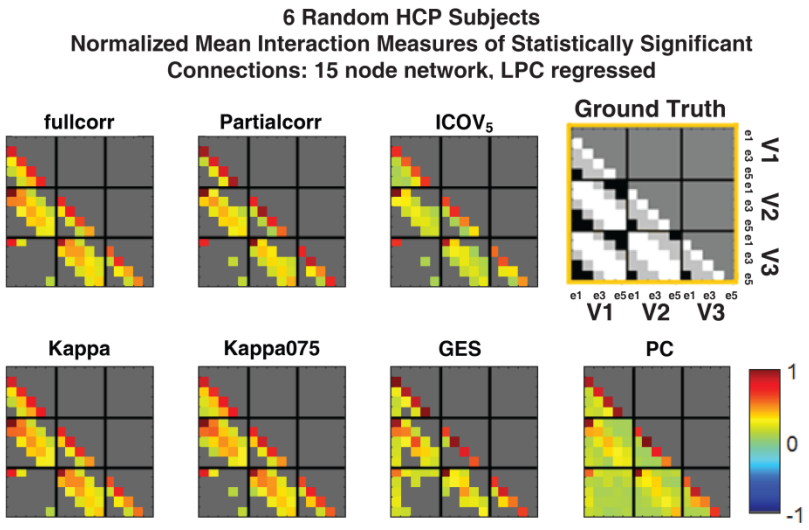
Supplementary Fig. 36. Proportion of true and false positive connections found for each of the models tested at various significance thresholds **using 6 random HCP subjects**. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 15 node network within the quadrant without the LPC regressed, there are 51 expected connections tested for and 17 unexpected connections. The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The threshold p-values become progressively smaller by a multiplication factor of 10^-2 at each step, as described for Supplementary Figure 4.



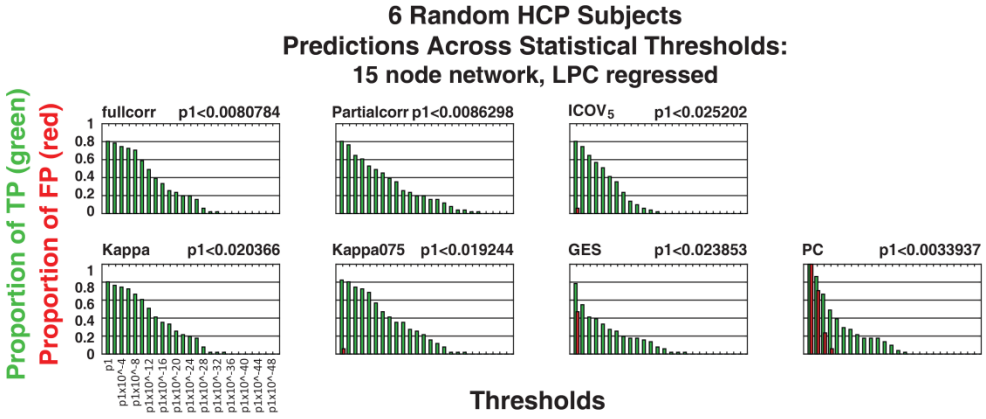
Supplementary Fig. 37. Proportion of true and false positive connections found for each of the models tested at various *network connection densities* **using 6 random HCP subjects**. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 15 node network within the quadrant with no LPC regression, there are 51 expected connections tested for and 17 unexpected connections. The network connection densities along the horizontal axis begin at 100% and go down by 4% until reaching 0%.



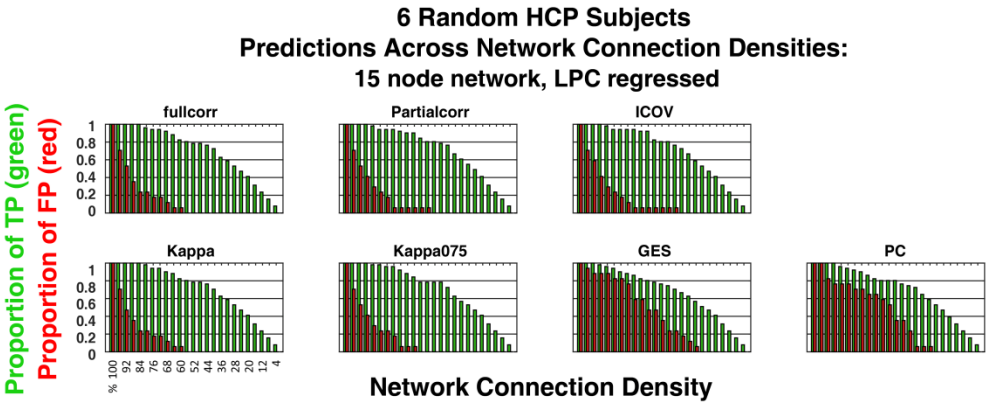
Supplementary Fig. 38. Normalized mean interaction measures for each of the connections tested within the quadrant **using 6 random HCP subjects**, for each model (*with LPC regression*). The gray plot reflects the full distribution of mean interaction measures (over subjects, runs, and quadrants) of all computed connections ([15x15-15]/2=105). The vertical axis scale for gray plots is on the right-hand side of the Supplementary Figures. Green bars reflect the mean interaction measures associated with the 51 expected connections within the quadrant. Red bars reflect the 17 unexpected connections within the quadrant. The vertical axis scale for the red and green bars is on the left-hand side of the plots.



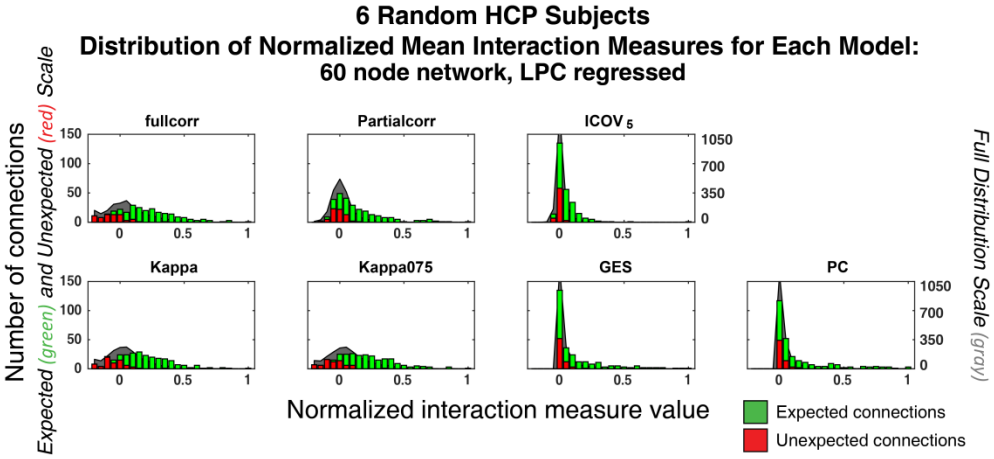
Supplementary Fig. 39. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 15 node, quadrant network **using 6 random HCP subjects** *with LPC regressed*. Presentation is as in Supplementary Figure 3. See "Ground Truth" matrix on top right. Significance is based on the FDR p-value threshold.



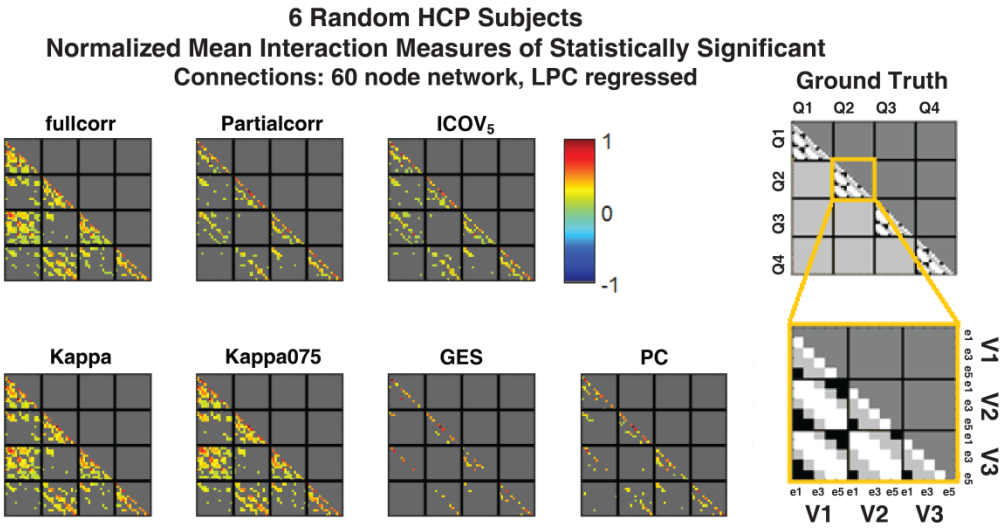
Supplementary Fig. 40. Proportion of true and false positive connections found for each of the models tested at various significance thresholds **using 6 random HCP subjects**. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 15 node network within the quadrant *with LPC regressed*, there are 51 expected connections tested for and 17 unexpected connections. The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The threshold p-values become progressively smaller by a multiplication factor of 10^-2 at each step, as described for Supplementary Figure 4.



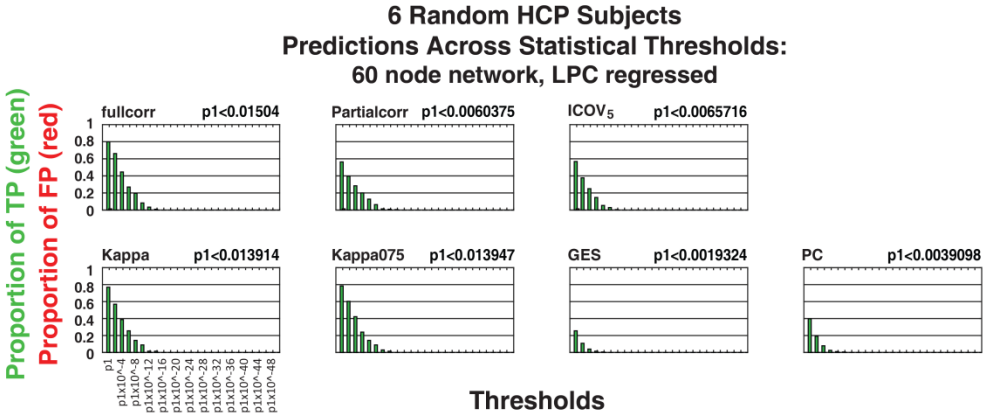
Supplementary Fig. 41. Proportion of true and false positive connections found for each of the models tested at various *network connection densities* **using 6 random HCP subjects**. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 15 node network within the quadrant *with LPC regression*, there are 51 expected connections tested for and 17 unexpected connections. The network connection densities along the horizontal axis begin at 100% and go down by 4% until reaching 0%.



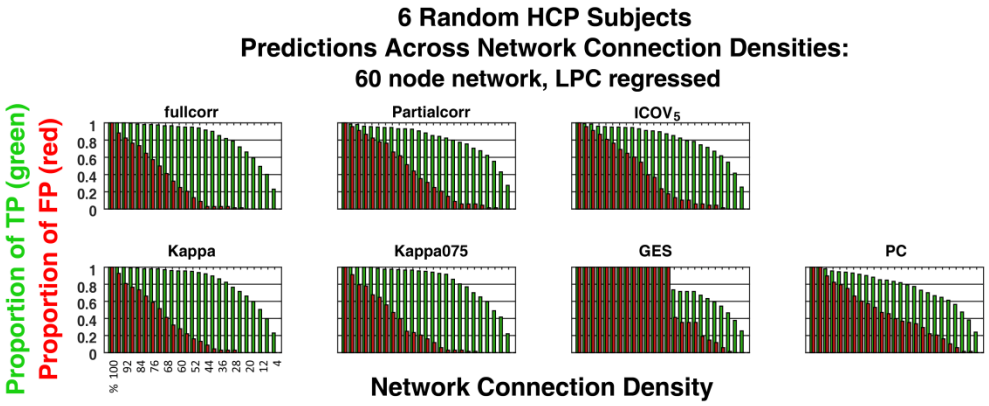
Supplementary Fig. 42. Normalized mean interaction measures for each of the connections tested within all four quadrants **using 6 random HCP subjects**, for each model (*LPC regressed*). Normalization is done by dividing all measures by the maximal value, such that for each method the maximal value presented is 1. The gray plot reflects the full distribution of mean interaction measures (over subjects and runs) of all computed connections ([60x60-60]/2=1770). The vertical axis scale for gray plots is on the right-hand side of the Supplementary Figures. Green bars reflect the mean interaction measures associated with the 51x4=204 expected connections within the quadrants. Red bars reflect the 17x4=68 unexpected connections within the quadrants. The vertical axis scale for the red and green bars is on the left-hand side of the plots.



Supplementary Fig. 43. Normalized mean interaction measures of ROI pairs with statistically significant connectivity predicted by each model in the 60 node, full visual cortex network **using 6 random HCP subjects** *with LPC regressed*. Presentation is as in Supplementary Figure 3. See "Ground Truth" matrix on top right. Significance is based on the FDR p-value threshold. Q=Quadrant.



Supplementary Fig. 44. Proportion of true and false positive connections found for each of the models tested at various significance thresholds **using 6 random HCP subjects**. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 60 node network *with LPC regressed*, there are 51×4=204 expected connections tested for and 17×4=68 unexpected connections. The p-value thresholds along the horizontal axis begin at the FDR p-value threshold (max FPR of 0.05) indicated in the title above each plot ('p1'). The threshold p-values become progressively smaller by a multiplication factor of 10^-2 at each step, as described for Supplementary Figure 4.



Supplementary Fig. 45. Proportion of true and false positive connections found for each of the models tested at various *network connection densities* **using 6 random HCP subjects**. Green bars are the true positives (TP), and red bars are the false positives (FP). In this 60 node network *with LPC regression*, there are 51×4=204 expected connections tested for and 17×4=68 unexpected connections. The network connection densities along the horizontal axis begin at 100% and go down by 4% until reaching 0%.

Table 1: The set of analyses we will present from this study either in the main manuscript or in supplemental material

	15 nodes (1 quadrant)		60 nodes (4 quadrants)	
Our 6 subjects	*LPC not regressed*	All data sets	*LPC not regressed*	All data sets
		Reduced data to match 60 nodes analysis		
	LPC regressed	All data sets	*LPC regressed*	All data sets
		Reduced data to match 60 nodes analysis		
220 HCP subjects	*LPC not regressed*	All data sets	*LPC not regressed*	All data sets
		6 random subjects		6 random subjects
	LPC regressed	All data sets	*LPC regressed*	All data sets
		6 random subjects		6 random subjects

Table 2: Threshold level of significant p-values that is necessary to maximize the proportion of True Positives minus the proportion of False Positives (TPR-FPR) for the 15 node network, ***LPC not regressed***. Second column: the FDR corrected alpha = 0.05 threshold. Third column: the optimal threshold. Fourth column: the optimal TPR-FPR, obtained with the optimal threshold. Fifth and sixth columns: The stability of the network prediction is assessed by evaluating the mean TPR-FPR across thresholds. TPR-FPR at the optimal density (given in column 7) is presented in column 8. The 10th column presents the proportion of triangles predicted by each model at alpha=0.05, FDR corrected. This is presented next to the proportion of all expected connections (i.e. TPR, 9th column). We also present the probability of a triangle being predicted given the probability of existence of each individual connection in the triangle (column 11). A value close to 1 should reflect minimal or no bias against triangle connectivity. Keep in mind that these values do not reflect the FP rate. Models which do not have a ~100% FP rate are identified in bold.

Good performance

Poor performance

Model name	p1: FDR corrected α=0.05	Threshold <=0.05 such that TPR-FPR maximized	TPR-FPR with optimal absolute threshold	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-20}]	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-50}]	Density such that TPR-FPR maximized	TPR-FPR with optimal density	α=0.05 TPR: All conn-ections	α=0.05 TPR: Triangles	P(Triangle| TPR)= [triangles ÷ (all conns)^3]
'Fullcorr'	7.30E-36	p1 E-36	0.43	0.11±0.04	0.20±0.13	52	0.76	1.00	1.00	1.00
'Partialcorr'	3.05E-02	p1	0.73	0.53±0.15	0.29±0.25	64	0.80	**0.90**	**0.71**	**0.97**
'ICOV_5'	3.93E-02	p1 E-2	0.80	0.58±0.14	0.32±0.26	64	0.84	**0.94**	**0.83**	**1.00**
'MI'	4.91E-181	p1	0.00	-0.11±0.08	-0.39±0.28	52	0.82	1.00	1.00	1.00
'CohA'	1.61E-183	p1	-0.02	-0.11±0.10	-0.33±0.24	60	0.82	0.98	1.00	1.00
'CohB'	3.50E-125	p1	0.00	-0.18±0.12	-0.29±0.21	56	0.84	1.00	1.00	1.00
'H_1'	2.05E-34	p1 E-16	0.35	0.21±0.11	0.13±0.13	48	0.71	1.00	1.00	1.00
'H_2'	3.53E-31	p1 E-22	0.49	0.25±0.16	0.18±0.16	48	0.69	1.00	1.00	1.00
'N_1'	2.59E-32	p1 E-26	0.53	0.22±0.14	0.23±0.16	52	0.75	1.00	1.00	1.00
'N_2'	4.90E-29	p1 E-26	0.63	0.20±0.14	0.25±0.18	48	0.71	1.00	1.00	1.00
'S_1'	5.11E-212	p1	0.00	-0.06±0.04	-0.18±0.13	60	0.86	1.00	1.00	1.00
'S_2'	9.27E-211	p1	0.00	-0.06±0.04	-0.21±0.17	56	0.80	1.00	1.00	1.00
'Kappa'	2.45E-30	p1 E-38	0.47	0.12±0.05	0.23±0.14	48	0.71	1.00	1.00	1.00
'Kappa075'	1.24E-28	p1 E-32	0.53	0.16±0.06	0.29±0.15	48	0.71	1.00	1.00	1.00
'GES'	1.59E-02	p1 E-6	0.76	0.46±0.21	0.29±0.21	52	0.78	1.00	1.00	1.00
'PC'	4.28E-02	p1 E-4	0.69	0.43±0.20	0.25±0.21	48	0.75	1.00	1.00	1.00

Table 3: Threshold level of significant p-values that is necessary to maximize the proportion of True Positives minus the proportion of False Positives (TPR-FPR) for the 15 node network, ***LPC regressed***. Second column: the FDR corrected alpha = 0.05 threshold. Third column: the optimal threshold. Fourth column: the optimal TPR-FPR, obtained with the optimal threshold. Fifth and sixth columns: The stability of the network prediction is assessed by evaluating the mean TPR-FPR across thresholds. TPR-FPR at the optimal density (given in column 7) is presented in column 8. The 10th column presents the proportion of triangles predicted by each model at alpha=0.05, FDR corrected. This is presented next to the proportion of all expected connections (i.e. TPR, 9th column). We also present the probability of a triangle being predicted given the probability of existence of each individual connection in the triangle (column 11). A value close to 1 should reflect minimal or no bias against triangle connectivity. Keep in mind that these values do not reflect the FP rate. Models which do not have a ~100% FP rate are identified in bold.

Good performance

Poor performance

Model name	p1: FDR corrected α=0.05	Threshold <=0.05 such that TPR-FPR maximized	TPR-FPR with optimal absolute threshold	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-20}]	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-50}]	Density such that TPR-FPR maximized	TPR-FPR with optimal density	α=0.05 TPR: All conn-ections	α=0.05 TPR: Triangles	P(Triangle| TPR)= [triangles ÷ (all conns)^3]
'Fullcorr'	1.81E-02	p1	0.88	0.71±0.12	0.53±0.18	76	0.90	**0.88**	**0.65**	**0.94**
'Partialcorr'	1.13E-02	p1	0.75	0.49±0.15	0.25±0.23	72	0.86	**0.80**	**0.46**	**0.88**
'ICOV_5'	2.00E-02	p1	0.76	0.51±0.16	0.26±0.24	68	0.84	**0.82**	**0.50**	**0.90**
'MI'	5.70E-201	p1	-0.02	-0.03±0.01	-0.08±0.07	32	0.61	0.98	1.00	1.00
'CohA'	1.06E-192	p1	0.00	-0.03±0.01	-0.08±0.08	28	0.51	1.00	1.00	1.00
'CohB'	2.80E-129	p1	0.00	-0.03±0.01	-0.15±0.15	32	0.61	1.00	1.00	1.00
'H_1'	1.93E-17	p1 E-12	0.45	0.25±0.15	0.13±0.14	32	0.51	1.00	1.00	1.00
'H_2'	2.79E-13	p1 E-14	0.45	0.24±0.17	0.14±0.15	32	0.59	0.98	1.00	1.00
'N_1'	6.81E-09	p1 E-12	0.71	0.33±0.22	0.19±0.20	40	0.59	1.00	1.00	1.00
'N_2'	1.59E-06	p1 E-10	0.63	0.29±0.22	0.17±0.18	32	0.59	1.00	1.00	1.00
'S_1'	4.42E-293	p1	0.00	-0.02±0.01	-0.01±0.01	40	0.73	1.00	1.00	1.00
'S_2'	1.64E-288	p1	0.00	-0.02±0.01	-0.02±0.02	48	0.67	1.00	1.00	1.00
'Kappa'	2.60E-02	p1 E-2	0.82	0.64±0.12	0.46±0.19	72	0.86	**0.82**	**0.56**	**1.00**
'Kappa075'	3.00E-02	p1	0.88	0.68±0.14	0.48±0.21	76	0.90	**0.88**	**0.65**	**0.94**
'GES'	1.46E-03	p1 E-6	0.47	0.29±0.13	0.17±0.14	36	0.51	1.00	1.00	1.00
'PC'	2.07E-04	p1 E-10	0.33	0.19±0.14	0.11±0.12	36	0.61	0.98	1.00	1.00

Table 4: Threshold level of significant p-values that is necessary to maximize the proportion of True Positives minus the proportion of False Positives (TPR-FPR) for the 60 node network, ***LPC not regressed***. Second column: the FDR corrected alpha = 0.05 threshold. Third column: the optimal threshold. Fourth column: the optimal TPR-FPR, obtained with the optimal threshold. Fifth and sixth columns: The stability of the network prediction is assessed by evaluating the mean TPR-FPR across thresholds. TPR-FPR at the optimal density (given in column 7) is presented in column 8. The 10th column presents the proportion of triangles predicted by each model at alpha=0.05, FDR corrected. This is presented next to the proportion of all expected connections (i.e. TPR, 9th column). We also present the probability of a triangle being predicted given the probability of existence of each individual connection in the triangle (column 11). A value close to 1 should reflect minimal or no bias against triangle connectivity. Keep in mind that these values do not reflect the FP rate. Models which do not have a ~100% FP rate are identified in bold.

Good performance

Poor performance

Model name	p1: FDR corrected α=0.05	Threshold <=0.05 such that TPR-FPR maximized	TPR-FPR with optimal absolute threshold	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-20}]	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-50}]	Density such that TPR-FPR maximized	TPR-FPR with optimal density	α=0.05 TPR: All conn-ections	α=0.05 TPR: Triangles	P(Triangle| TPR)= [triangles ÷ (all conns)^3]
'Fullcorr'	3.70E-02	p1 E-20	0.27	0.10±0.10	0.06±0.09	32	0.50	1.00	1.00	1.00
'Partialcorr'	7.74E-03	p1	0.62	0.16±0.21	0.07±0.15	40	0.68	**0.62**	**0.25**	**1.00**
'ICOV_5'	1.33E-02	p1	0.70	0.20±0.24	0.08±0.18	32	0.72	**0.77**	**0.45**	**0.98**
'MI'	1.44E-44	p1	0.00	-0.20±0.16	-0.10±0.14	40	0.60	1.00	1.00	1.00
'CohA'	6.91E-45	p1	0.00	-0.14±0.12	-0.06±0.10	44	0.63	1.00	1.00	1.00
'CohB'	1.09E-29	p1 E-2	0.00	-0.12±0.10	-0.06±0.09	36	0.59	1.00	1.00	1.00
'H_1'	3.96E-07	p1 E-10	0.29	0.10±0.11	0.04±0.08	44	0.51	1.00	1.00	1.00
'H_2'	5.03E-07	p1 E-10	0.34	0.13±0.14	0.06±0.11	44	0.49	1.00	1.00	1.00
'N_1'	4.16E-07	p1 E-10	0.41	0.17±0.14	0.07±0.12	36	0.52	1.00	1.00	1.00
'N_2'	3.17E-06	p1 E-10	0.48	0.19±0.16	0.08±0.14	48	0.49	1.00	1.00	1.00
'S_1'	8.16E-50	p1 E-48	0.01	-0.11±0.11	-0.11±0.11	40	0.72	1.00	1.00	1.00
'S_2'	5.85E-52	p1	0.00	-0.20±0.19	-0.19±0.17	32	0.55	1.00	1.00	1.00
'Kappa'	3.35E-03	p1 E-16	0.29	0.13±0.11	0.07±0.10	36	0.49	1.00	1.00	1.00
'Kappa075'	3.43E-02	p1 E-16	0.36	0.17±0.14	0.09±0.12	44	0.47	1.00	1.00	1.00
'GES'	7.07E-03	p1	0.53	0.13±0.17	0.06±0.13	28	0.62	**0.53**	**0.14**	**0.89**
'PC'	6.84E-03	p1	0.52	0.11±0.16	0.05±0.12	36	0.64	**0.54**	**0.13**	**0.83**

Table 5: Threshold level of significant p-values that is necessary to maximize the proportion of True Positives minus the proportion of False Positives (TPR-FPR) for the 60 node network, ***LPC regressed***. Second column: the FDR corrected alpha = 0.05 threshold. Third column: the optimal threshold. Fourth column: the optimal TPR-FPR, obtained with the optimal threshold. Fifth and sixth columns: The stability of the network prediction is assessed by evaluating the mean TPR-FPR across thresholds. TPR-FPR at the optimal density (given in column 7) is presented in column 8. The 10th column presents the proportion of triangles predicted by each model at alpha=0.05, FDR corrected. This is presented next to the proportion of all expected connections (i.e. TPR, 9th column). We also present the probability of a triangle being predicted given the probability of existence of each individual connection in the triangle (column 11). A value close to 1 should reflect minimal or no bias against triangle connectivity. Keep in mind that these values do not reflect the FP rate. Models which do not have a ~100% FP rate are identified in bold.

Good performance

Poor performance

Model name	p1: FDR corrected α=0.05	Threshold <=0.05 such that TPR-FPR maximized	TPR-FPR with optimal absolute threshold	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-20}]	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-50}]	Density such that TPR-FPR maximized	TPR-FPR with optimal density	α=0.05 TPR: All conn-ections	α=0.05 TPR: Triangles	P(Triangle| TPR)= [triangles ÷ (all conns)^3]
'Fullcorr'	2.11E-02	p1	0.80	0.41±0.23	0.18±0.25	60	0.84	**0.80**	**0.53**	**1.00**
'Partialcorr'	6.11E-03	p1	0.58	0.15±0.20	0.06±0.15	36	0.75	**0.58**	**0.18**	**0.92**
'ICOV_5'	7.95E-03	p1	0.66	0.18±0.23	0.08±0.17	36	0.77	**0.66**	**0.28**	**0.97**
'MI'	1.43E-53	p1	0.00	-0.20±0.17	-0.11±0.15	20	0.51	1.00	1.00	1.00
'CohA'	5.83E-50	p1	0.00	-0.09±0.06	-0.05±0.06	36	0.37	1.00	1.00	1.00
'CohB'	2.77E-35	p1 E-30	0.00	-0.21±0.15	-0.10±0.14	20	0.50	1.00	1.00	1.00
'H_1'	1.29E-02	p1 E-6	0.30	0.10±0.12	0.04±0.09	28	0.41	1.00	1.00	1.00
'H_2'	4.60E-02	p1 E-6	0.36	0.10±0.12	0.04±0.09	28	0.37	1.00	1.00	1.00
'N_1'	4.77E-02	p1 E-4	0.38	0.14±0.15	0.06±0.11	28	0.43	1.00	0.98	1.00
'N_2'	4.59E-02	p1 E-4	0.35	0.12±0.13	0.05±0.10	24	0.39	0.98	0.92	0.99
'S_1'	9.14E-75	p1 E-30	0.11	-0.03±0.03	0.00±0.06	36	0.56	1.00	1.00	1.00
'S_2'	9.40E-74	p1 E-50	0.02	-0.04±0.03	-0.11±0.11	28	0.50	1.00	1.00	1.00
'Kappa'	1.92E-02	p1	0.76	0.34±0.23	0.14±0.22	48	0.80	**0.76**	**0.44**	**1.00**
'Kappa075'	2.09E-02	p1	0.79	0.35±0.25	0.15±0.24	52	0.85	**0.80**	**0.53**	**1.00**
'GES'	3.49E-03	p1	0.42	0.10±0.13	0.04±0.10	32	0.53	**0.42**	**0.05**	**0.72**
'PC'	2.12E-02	p1	0.41	0.10±0.14	0.04±0.10	24	0.52	**0.75**	**0.43**	**1.00**

Table 6: Threshold level of significant p-values that is necessary to maximize the proportion of True Positives minus the proportion of False Positives (TPR-FPR) for the 15 node network, ***LPC not regressed,* with 220 HCP subjects**. Second column: the FDR corrected alpha = 0.05 threshold. Third column: the optimal threshold. Fourth column: the optimal TPR-FPR, obtained with the optimal threshold. Fifth and sixth columns: The stability of the network prediction is assessed by evaluating the mean TPR-FPR across thresholds. TPR-FPR at the optimal density (given in column 7) is presented in column 8. The 10th column presents the proportion of triangles predicted by each model at alpha=0.05, FDR corrected. This is presented next to the proportion of all expected connections (i.e. TPR, 9th column). We also present the probability of a triangle being predicted given the probability of existence of each individual connection in the triangle (column 11). A value close to 1 should reflect minimal or no bias against triangle connectivity. Keep in mind that these values do not reflect the FP rate. Models which do not have a ~100% FP rate are identified in bold. Fullcorr connections all had p-values ~0 and thus FDR thresholding was not possible.

Good performance

Poor performance

Model name	p1: FDR corrected α=0.05	Threshold <=0.05 such that TPR-FPR maximized	TPR-FPR with optimal absolute threshold	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-20}]	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-50}]	Density such that TPR-FPR maximized	TPR-FPR with optimal density	α=0.05 TPR: All conn-ections	α=0.05 TPR: Triangles	P(Triangle| TPR)= [triangles ÷ (all conns)^3]
'Fullcorr'	N/A	N/A	0.00	0.00±0.00	0.00±0.00	64	0.90	1	1	1
'Partialcorr'	4.57E-03	p1 E-50	0.84	0.39±0.11	0.56±0.17	64	0.96	**1**	**1**	**1**
'ICOV_5'	2.79E-02	p1 E-50	0.90	0.52±0.15	0.67±0.18	68	0.96	**1**	**1**	**1**
'Kappa'	1.11E-28	p1	0.00	-0.02±0.03	-0.05±0.03	64	0.90	1	1	1
'Kappa075'	7.80E-40	p1	0.00	-0.01±0.02	-0.04±0.02	64	0.90	1	1	1
'GES'	1.09E-81	p1 E-48	0.65	0.12±0.06	0.28±0.19	72	1.00	1	1	1
'PC'	3.57E-66	p1 E-46	0.98	0.21±0.16	0.53±0.31	72	0.98	1	1	1

Table 7: Threshold level of significant p-values that is necessary to maximize the proportion of True Positives minus the proportion of False Positives (TPR-FPR) for the 15 node network, ***LPC regressed,* with 220 HCP subjects**. Second column: the FDR corrected alpha = 0.05 threshold. Third column: the optimal threshold. Fourth column: the optimal TPR-FPR, obtained with the optimal threshold. Fifth and sixth columns: The stability of the network prediction is assessed by evaluating the mean TPR-FPR across thresholds. TPR-FPR at the optimal density (given in column 7) is presented in column 8. The 10th column presents the proportion of triangles predicted by each model at alpha=0.05, FDR corrected. This is presented next to the proportion of all expected connections (i.e. TPR, 9th column). We also present the probability of a triangle being predicted given the probability of existence of each individual connection in the triangle (column 11). A value close to 1 should reflect minimal or no bias against triangle connectivity. Keep in mind that these values do not reflect the FP rate. Models which do not have a ~100% FP rate are identified in bold.

Good performance

Poor performance

Model name	p1: FDR corrected α=0.05	Threshold <=0.05 such that TPR-FPR maximized	TPR-FPR with optimal absolute threshold	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-20}]	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-50}]	Density such that TPR-FPR maximized	TPR-FPR with optimal density	α=0.05 TPR: All connections	α=0.05 TPR: Triangles	P(Triangle|TPR)= [triangles ÷ (all conns)^3]
'Fullcorr'	3.18E-03	p1	0.94	0.90±0.02	0.85±0.04	72	0.96	**0.94**	**0.83**	**1.00**
'Partialcorr'	2.36E-03	p1	0.90	0.83±0.03	0.81±0.03	72	0.96	**0.90**	**0.67**	**0.91**
'ICOV_5'	5.13E-03	p1	0.86	0.82±0.02	0.80±0.03	68	0.94	**0.86**	**0.63**	**0.97**
'Kappa'	1.15E-02	p1	0.94	0.89±0.03	0.85±0.05	72	0.96	**0.94**	**0.83**	**1.00**
'Kappa075'	1.74E-02	p1 E-4	0.98	0.95±0.02	0.88±0.08	76	1.00	**1.00**	**1.00**	**1.00**
'GES'	1.60E-69	p1 E-48	0.55	-0.01±0.03	0.21±0.21	40	0.73	1.00	1.00	1.00
'PC'	2.14E-96	p1 E-48	0.67	-0.03±0.02	0.23±0.26	40	0.75	1.00	1.00	1.00

Table 8: Threshold level of significant p-values that is necessary to maximize the proportion of True Positives minus the proportion of False Positives (TPR-FPR) for the 60 node network, ***LPC not regressed,* with 220 HCP subjects**. Second column: the FDR corrected alpha = 0.05 threshold. Third column: the optimal threshold. Fourth column: the optimal TPR-FPR, obtained with the optimal threshold. Fifth and sixth columns: The stability of the network prediction is assessed by evaluating the mean TPR-FPR across thresholds. TPR-FPR at the optimal density (given in column 7) is presented in column 8. The 10th column presents the proportion of triangles predicted by each model at alpha=0.05, FDR corrected. This is presented next to the proportion of all expected connections (i.e. TPR, 9th column). We also present the probability of a triangle being predicted given the probability of existence of each individual connection in the triangle (column 11). A value close to 1 should reflect minimal or no bias against triangle connectivity. Keep in mind that these values do not reflect the FP rate. Models which do not have a ~100% FP rate are identified in bold.

Good performance

Poor performance

Model name	p1: FDR corrected α=0.05	Threshold <=0.05 such that TPR-FPR maximized	TPR-FPR with optimal absolute threshold	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-20}]	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-50}]	Density such that TPR-FPR maximized	TPR-FPR with optimal density	α=0.05 TPR: All connections	α=0.05 TPR: Triangles	P(Triangle|TPR)= [triangles ÷ (all conns)^3]
'Fullcorr'	1.09E-134	p1 E-48	0.21	0.03±0.02	0.09±0.06	60	0.61	1.00	1.00	1.00
'Partialcorr'	3.39E-02	p1 E-6	0.95	0.86±0.07	0.78±0.09	48	0.94	**0.98**	**0.94**	**0.99**
'ICOV_5'	3.53E-02	p1 E-10	0.94	0.86±0.08	0.76±0.11	48	0.97	**0.99**	**0.95**	**1.00**
'Kappa'	3.86E-122	p1 E-46	0.24	0.03±0.02	0.11±0.08	60	0.60	1.00	1.00	1.00
'Kappa075'	4.19E-10	p1	0.00	-0.02±0.02	-0.05±0.03	60	0.59	1.00	1.00	1.00
'GES'	3.63E-2	p1 E-2	0.70	0.35±0.22	0.22±0.15	48	0.84	**0.98**	**0.92**	**0.99**
'PC'	1.30E-02	p1 E-8	0.94	0.63±0.33	0.60±0.22	48	0.89	1.00	1.00	1.00

Table 9: Threshold level of significant p-values that is necessary to maximize the proportion of True Positives minus the proportion of False Positives (TPR-FPR) for the 60 node network, ***LPC regressed,* with 220 HCP subjects**. Second column: the FDR corrected alpha = 0.05 threshold. Third column: the optimal threshold. Fourth column: the optimal TPR-FPR, obtained with the optimal threshold. Fifth and sixth columns: The stability of the network prediction is assessed by evaluating the mean TPR-FPR across thresholds. TPR-FPR at the optimal density (given in column 7) is presented in column 8. The 10th column presents the proportion of triangles predicted by each model at alpha=0.05, FDR corrected. This is presented next to the proportion of all expected connections (i.e. TPR, 9th column). We also present the probability of a triangle being predicted given the probability of existence of each individual connection in the triangle (column 11). A value close to 1 should reflect minimal or no bias against triangle connectivity. Keep in mind that these values do not reflect the FP rate. Models which do not have a ~100% FP rate are identified in bold.

Good performance

Poor performance

Model name	p1: FDR corrected α=0.05	Threshold <=0.05 such that TPR-FPR maximized	TPR-FPR with optimal absolute threshold	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-20}]	Mean TPR-FPR ± standard deviation over the range of [p1 to p1x10^{-50}]	Density such that TPR-FPR maximized	TPR-FPR with optimal density	α=0.05 TPR: All conn-ections	α=0.05 TPR: Triangles	P(Triangle| TPR)= [triangles ÷ (all conns)^3]
'Fullcorr'	2.27E-02	p1	0.92	0.87±0.03	0.81±0.05	56	0.94	**0.92**	**0.77**	**0.98**
'Partialcorr'	1.63E-02	p1	0.89	0.77±0.05	0.68±0.09	44	0.94	**0.89**	**0.70**	**1.01**
'ICOV_5'	1.71E-02	p1	0.91	0.78±0.06	0.68±0.10	48	0.94	**0.91**	**0.74**	**0.99**
'Kappa'	2.56E-02	p1	0.92	0.86±0.04	0.81±0.05	56	0.94	**0.92**	**0.77**	**0.98**
'Kappa075'	2.66E-02	p1	0.94	0.84±0.07	0.76±0.08	56	0.96	**0.94**	**0.81**	**0.97**
'GES'	2.65E-02	P1 E-2	0.52	0.27±0.16	0.12±0.17	16	0.59	**0.81**	**0.58**	**1.07**
'PC'	2.30E-03	p1 E-4	0.60	0.15±0.21	0.06±0.15	20	0.68	1.00	1.00	1.00

Appendix D:
A Manuscript

(2018). Human Connectome Project resting state fMRI data organized into 60 fine-scaled retinotopically organized regions in cortical areas V1, V2 and V3. bioRxiv. Epub: June 6, 2018. https://doi.org/10.1101/340174

Submitted to Data In Brief

Data Article

Title: Human Connectome Project resting state fMRI data organized into 60 fine-scaled retinotopically organized regions in cortical areas V1, V2 and V3

Authors: Debra Ann Dawson, Zixuan Yin, Jack Lam, Amir Shmuel

Affiliations: McConnell Brain Imaging Centre, Montreal Neurological Institute; Department of Neurology and Neurosurgery, McGill University

Contact email: amir.shmuel@mcgill.ca

Abstract

The data comprises 60 regions of interest (ROIs) from V1, V2, and V3 of the human visual cortex. Preprocessed data from the Human Connectome Project (HCP) 900 subjects public data release were utilized: 220 subjects were randomly selected, each with 4 scans of resting state fMRI data. Given that these subjects did not have retinotopy scans performed, the visual areas were defined using an anatomical template from Benson et al. (2014). Visual areas from each hemisphere were further divided along dorsal-ventral lines into quadrants, resulting in 4 quadrants per subject. Within each quadrant, fine-scaled ROIs were defined by subdividing each visual area into 5 regions according to eccentricity. These data may be useful for studying retinotopically organized functional connectivity in the visual cortex using the HCP 3 Tesla dataset.

Specifications Table

Subject area	*Neuroscience*
More specific subject area	*NeuroImaging*
Type of data	*Functional MRI (fMRI) region of interest average time courses.*
How data was acquired	*fMRI at 3 Tesla, Human Connectome Project public release data.*
Data format	*Voxel time courses averaged into regions of interest; Matlab array.*
Experimental factors	*Template of early visual areas applied to preprocessed data*
Experimental features	
Data source location	*Human Connectome Project public data release*
Data accessibility	*Data is with this article (with Data In Brief)*
Related research article	Dawson DA, Lam J, Lewis LB, Carbonell F, Mendola JD, Shmuel A. Partial Correlation-Based Retinotopically Organized Resting-State Functional Connectivity Within and Between Areas of the Visual Cortex Reflects More Than Cortical Distance. Brain Connect. Feb 2016;6(1):57-75.

Value of the Data

- Most datasets of subjects having undergone retinotopy include data from only a small number of subjects, limiting the power possible from the study. Here we present high quality data obtained from a large number (220) of subjects.
- The data can be used for studying the functional connectivity within and between areas V1, V2 and V3 of the visual cortex.
- Most of the previously acquired data from study subjects having undergone retinotopy scanning has been done with a 3 Tesla scanner. We provide a large, high quality sample

of V1, V2, and V3 resting state data acquired from a 3 Tesla scanner, allowing for the opportunity to undertake comparative studies using other datasets, including datasets from patient populations (e.g. Mendola et al., 2018).

Data

The data comprises 60 regions of interest (ROIs) from V1, V2, and V3 of the human visual cortex. 3D volume preprocessed resting state fMRI data from the Human Connectome Project (HCP) 900 subjects public data release were utilized: 220 random subjects were selected, each with 4 scans of 3 Tesla resting state fMRI data. Given that these subjects did not have retinotopy scans performed, the visual areas were defined using an anatomical template from Benson et al. (2014). Visual areas from each hemisphere were further subdivided along dorsal-ventral lines into quadrants, resulting in 4 quadrants per subject. Within each quadrant, fine-scaled ROIs were defined by subdividing each visual area into 5 regions according to eccentricity.

Experimental Design, Materials, and Methods

Original data and preprocessing

Subjects were randomly selected from those available in the HCP900 data release. 3D volume preprocessed resting state fMRI data were used, along with files resulting from a FreeSurfer-based surface extraction pipeline. Important details regarding scanning follows (see also Table 1): the resting state fMRI scans were done with 3 Tesla scanners, the TR was 0.72s, there are 1200 time points per scan. Full details on the data acquisition can be found in a special Connectome issue of NeuroImage, the most important contributions to the data used here being the following: Van Essen et al., 2013, Ŭgurbil et al., 2013, Glasser et al.2013, Smith et al., 2013, Sotiropoulos et al., 2013.

Parameter	Value
Sequence	Gradient-echo EPI
TR	720 ms
TE	33.1 ms
flip angle	52 deg
FOV	208x180 mm (RO x PE)
Matrix	104x90 (RO x PE)
Slice thickness	2.0 mm; 72 slices; 2.0 mm isotropic voxels
Multiband factor	8
Echo spacing	0.58 ms
BW	2290 Hz/Px

Table D.1. Parameters relevant for the resting state fMRI images from the HCP dataset. This table is taken directly from the HCP 900 Subjects Data Release Reference Manual (page 32). https://www.humanconnectome.org/storage/app/media/documentation/s900/HCP_S900_Release_Reference_Manual.pdf

The preprocessed HCP data have undergone "the Minimal Preprocessing Pipelines". The specific pipelines the data utilized here underwent are the so called PreFreesurfer, Freesurfer, PostFreeSurfer, and fMRIVolume pipelines. Details on these and all of the Minimal Preprocessing Pipelines can be found in Glasser et al. (2013), with additional details in Jenkinson et al. (2012), Fischl (2012), and Jenskinson et al. (2002). These pipelines utilize tools From FSL and FreeSurfer.

The preprocessed data downloaded for this data release came from the following HCP repository locations:

/HCP_900/{subjectID}/T1w/{subjectID}/*

/HCP_900/{subjectID}/MNINonLinear/Results/rfMRI_REST{sessionID}_{runID}/

…

rfMRI_REST{sessionID}_{runID}_hp2000_clean.nii.gz

The first contains the post-FreeSurfer brain extracted volumes and surfaces which are needed for ROI definition. The second contains the four resting state scans per subject, with {sessionID}

holding the value of either 1, or 2, and {runID} holding the value of either LR or RL (this is meant to represent the phase encoding direction of the oblique axial acquisitions, alternated in consecutive runs: LR=left-right, RL=right-left). {subjectID} is the six digit subject number assigned to all participants in the HCP.

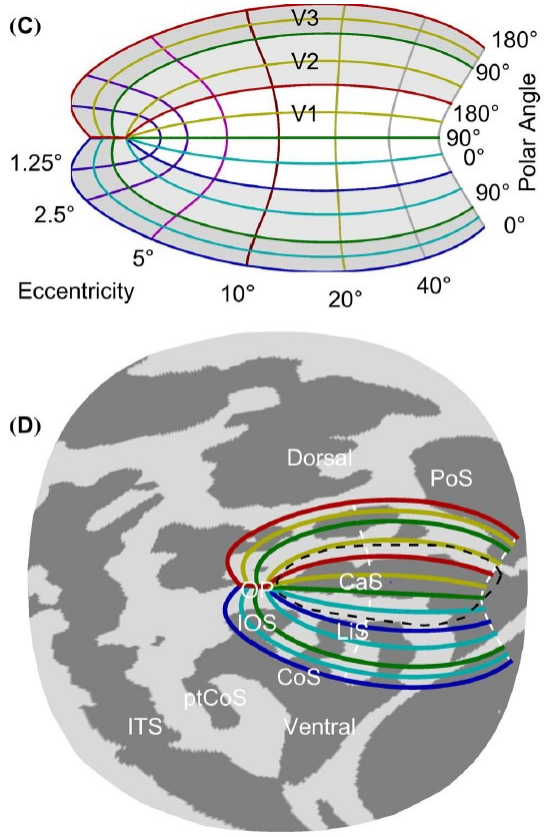


Figure D.1. Figure (panels C and D) from Benson et al. (2014). "(C) The algebraic model of retinotopic organization. V1, V2, and V3 are colored white, light gray, and dark gray, respectively. (D) The cortical surface atlas space (fsaverage_sym) from the occipital pole after flattening to the 2D surface. The Hinds V1 border (Hinds et al., 2008) is indicated by the dashed black line. The algebraic model of retinotopic organization used in registration is plotted with all 0°, 90°, and 180° polar angle lines colored according to their color in (C), and the 10° and 90° eccentricity lines dashed and colored white. Shown are the Calcarine Sulcus (CaS), the Parietal-occipital Sulcus (PoS), the Lingual sulcus (LiS), the Inferior Occipital Sulcus (IOS), the Collateral Sulcus (CoS), the posterior Collateral Sulcus (ptCoS), the Inferior Temporal Sulcus (ITS), and the Occipital Pole (OP)." Reprinted from Benson et al. (2014) with permission.

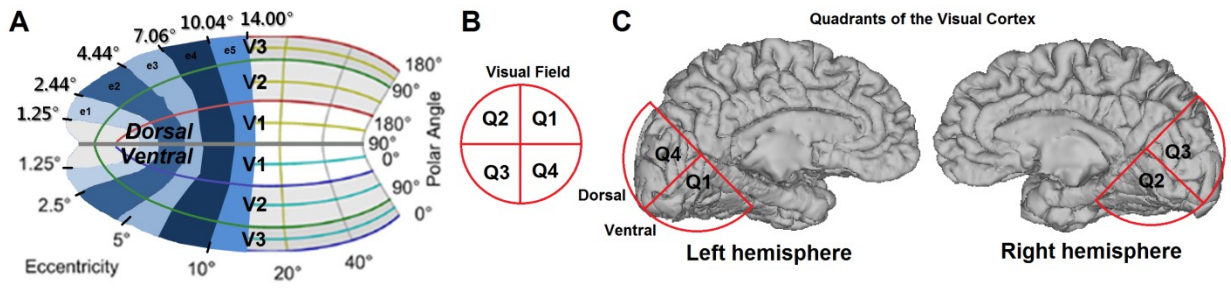


Figure D.2. ROIs defined on a template representation of one hemisphere of the visual cortex. (A) We define the regions using the template from Benson et al. (2014). We first split each hemisphere along dorsal-ventral lines, forming two quadrants per hemisphere. Voxels from each visual area and quadrant were classified into five eccentricity bins with the boundaries 1.25° to 2.44°, 2.44° to 4.44°, 4.44° to 7.06°, 7.06° to 10.04°, and 10.04° to 14.00°. Panel (C) presents an approximate scheme of the four quadrants. These quadrants reflect the visual field representations as identified in (B). For preparing Figure 2A, we modified Figure 1C from Benson et al (2014) with permission.

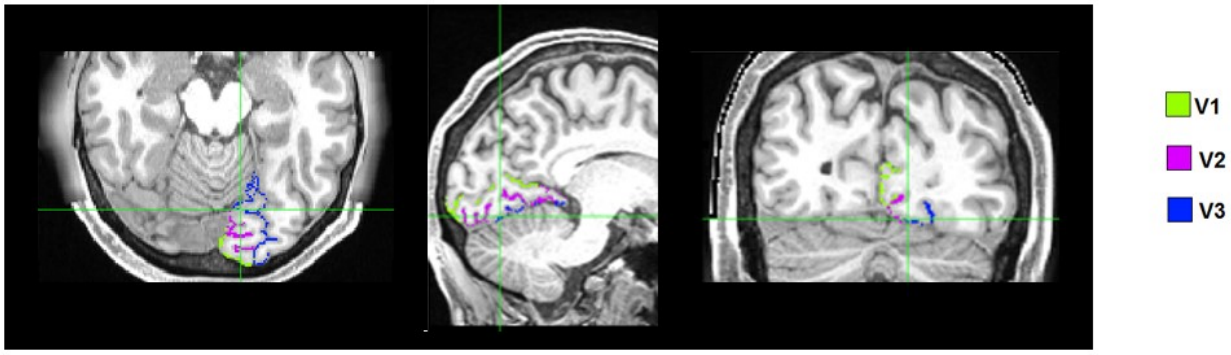


Figure D.3. Results of template alignment on an example subject, ID 1004008. The green outline indicates the spanned stretch of the cortical surface assigned to V1, purple is V2, and blue is V3.

Regions of interest

We define a local visual network within a hemisphere, which is subdivided into ventral and dorsal sections along the horizontal meridian of the polar angle phase map (>90° = dorsal

quadrant, <90° = ventral quadrant, see figure 2A). Anatomically, this horizontal meridian is expected to lie at the deepest curvature of the calcarine sulcus (figure 2C). These data were obtained from the two hemispheres of each subject, thus forming 4 quadrants of the visual cortex (cortical regions which respond to stimuli in the 4 quadrants of the visual field) (Figure 2C).

Typically, visual areas are defined on the cortex via retinotopic scanning and mapping. Rotating wedge and expanding/contracting ring stimuli are presented to the subjects during a retinotopic scanning session in order to elicit polar angle and eccentricity phase maps respectively. Visual area delineation is achieved by examining the reversal of direction of increasing polar angle phases perpendicular to the iso-eccentricity lines represented on the cortex (Sereno et al., 1995; Wandell et al., 2007).

We were interested in having access to the broad range of subjects who were only scanned in a 3 Tesla scanner and thus sought a method which would allow us to define the visual areas without having retinotopic scans. As a result, we have delineated V1, V2, and V3 using a template (version 2.5) described by Benson et al. (2012, 2014). The template was produced by aligning aggregate retinotopy from a group of subjects to a model of V1, V2, and V3 retinotopy based on surface topology in the visual cortex (Figure 2). We applied this template to the HCP subjects' *orig.mgz* (output file from Freesurfer's surface extraction) after computing the registration of the template to the subject (Figure 3). The registration is accomplished via alignment of the *sphere.reg* subject file to *fsaverage_sym.sphere.reg*, an average surface atlas available in Freesurfer to which the template is already aligned.

ROIs are subsections (5) of visual areas V1, V2, and V3 (3) within each quadrant (4) of the visual cortex (total number of ROIs=5×3×4 = 60) (Figure 2). Quadrants are defined by separating the visual cortex into hemispheres, and then along dorsal/ventral lines according to polar phase. These quadrants as defined on the cortex respond to a quadrant of the visual field. It can be very problematic in FC analyses if voxels belonging to one functional region are misclassified into a neighboring region (Smith et al., 2011). In order to avoid this detrimental effect, we removed a 10 polar angle degree range from either side of the border between visual areas. We also excluded eccentricity values below 1.25° due to measurement bias near the edge of the stimulus range and difficulty in obtaining eccentricity mapping near the fovea (Wandell et al., 2007; Schira et al., 2009; Benson et al., 2014). The retinotopy template assigns values to voxels using the full extent of possible phases in eccentricity and polar angle dimensions, i.e. 0°

to 90°. This range of eccentricity values goes beyond what is typical in experimental retinotopic mapping (where the maximum eccentricity of a stimulus presented is usually under 15°). We therefore have selected a more "realistic" maximal eccentricity boundary for use in defining our ROIs. We chose our maximal eccentricity to be 14°. This value is the same as we had used in a previous study by our group with measured retinotopy (Dawson et al., 2016) and is within the range of values tested in the Benson et al. (2014) study which defined the template.

In Benson et al., they look predominantly at an eccentricity range of 0° to10° (avoiding 1.25° at each edge), but also considered a maximum of 20°. They found a somewhat decreased match between template predicted visual area boundaries and measured retinotopic boundaries with a maximum eccentricity of 20° relative to 10°, however the results were still relevant and the errors similar in scale. Importantly, the authors point out that the error in visual area boundary definition between two retinotopic scans is greater than the prediction error of their anatomical template (as compared to the measured retinotopy). Thus, for maximal eccentricity values under 20°, the discordance of the template with measured retinotopic visual area delineation is expected to be within the range of the uncertainty of the measured retinotopy. Our choice of 14° is a compromise in order to capture a broader field while minimizing prediction error.

Given the above choices regarding upper and lower limits of eccentricity, we selected boundaries along the eccentricity dimension within each visual area, delineating our fine-scaled ROIs as follows: 1.25° to 2.44°, 2.44° to 4.44°, 4.44° to 7.06°, 7.06° to 10.04°, and 10.04° to 14.00°. The middle boundaries were chosen via an optimization algorithm in order to approximately equalize the number of voxels included in each ROI across subjects. The eccentricity ranges of the regions are expected to increase as we move more peripherally due to the smaller cortical magnification in peripheral representations of the visual field relative to foveal representations (Duncan and Boynton, 2003). We denote these five regions as e1, e2, e3, e4, and e5, respectively (Figure 2A). 'e1' is an abbreviation for 'eccentricity 1', representing the cortical region which responds to stimuli most central in the visual field. 'e5' represents the region which responds to stimuli most peripheral in the visual field. ROI time courses are generated by computing the mean across all voxels within the ROI. These average time courses have been normalized by subtracting the mean and dividing by the standard deviation across time points.

Details of data organization

DATAHCP200 is a 1x880 cell Matlab array. Each cell represents an independent run. The order of the runs is as follows:

Subject1_run1_LR, Subject1_run1_RL, Subject1_run2_LR, Subject1_run2_RL, Subject2_run1_LR, Subject2_run1_RL, Subject2_run2_LR,…, Subject220_run2_RL

Within each cell is a 1200x60 double matrix. Column vectors contain the time-courses (1200 time points) for each of the 60 ROIs. The TR of these resting state data sets is 0.72s. The order of the ROIs is as follows:

Q1_V1_e1, Q1_V1_e2, Q1_V1_e3, Q1_V1_e4, Q1_V1_e5, Q1_V2_e1, Q1_V2_e2, …, Q1_V3_e5, Q2_V1_e1,…, Q4_V3_e5.

Where "Qi" represents the quadrant (i=1,2,3,4), "Vj" represents the visual area (j=1,2,3), and "ek" represents the eccentricity region (k=1, 2, 3, 4, 5) (e1 is most central, e5 is most peripheral; see figure 2A). Quadrant numbering is as presented in Figure 2B and C.

Acknowledgments

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

References

1. Benson NC, Butt OH, Brainard DH, Aguirre GK. Correction of distortion in flattened representations of the cortical surface allows prediction of V1-V3 functional organization from anatomy. PLoS Comput Biol. Mar 2014;10(3):e1003538.
2. Benson NC, Butt OH, Datta R, Radoeva PD, Brainard DH, Aguirre GK. The retinotopic organization of striate cortex is well predicted by surface topology. Curr Biol. Nov 6 2012;22(21):2081-2085.
3. Dawson DA, Cha K, Lewis LB, Mendola JD, Shmuel A. Evaluation and calibration of functional network modeling methods based on known anatomical connections. *NeuroImage.* Feb 15 2013;67:331-343.
4. Dawson DA, Lam J, Lewis LB, Carbonell F, Mendola JD, Shmuel A. Partial Correlation-Based Retinotopically Organized Resting-State Functional Connectivity Within and Between Areas of the Visual Cortex Reflects More Than Cortical Distance. Brain Connect. Feb 2016;6(1):57-75.
5. Fischl B. FreeSurfer. NeuroImage. Aug 15 2012;62(2):774-781.
6. Glasser MF, Sotiropoulos SN, Wilson JA, et al. The minimal preprocessing pipelines for the Human Connectome Project. NeuroImage. 2013/10/15/ 2013;80:105-124.
7. Hinds OP, Rajendran N, Polimeni JR, et al. Accurate prediction of V1 location from cortical folds in a surface coordinate system. *NeuroImage.* Feb 15 2008;39(4):1585-1599.
8. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage. Oct 2002;17(2):825-841.
9. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. NeuroImage. Aug 15 2012;62(2):782-790.
10. Mendola JD, Lam J, Rosenstein M, Lewis LB, Shmuel A. Partial Correlation Analysis Reveals Abnormal Retinotopically Organized Functional Connectivity of Visual Areas in Amblyopia. NeuroImage Clinical, 2018. 18:192-201. https://doi.org/10.1016/j.nicl.2018.01.022.
11. Schira MM, Tyler CW, Breakspear M, Spehar B. The Foveal Confluence in Human Visual Cortex. Journal of Neuroscience. 2009; 29: 9050-9058.

12. Sereno M, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science.* 1995;269(5212):889-893.
13. Smith SM, Karla L. Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F. Beckmann, Thomas E. Nichols, Joseph D. Ramsey, Mark W. Woolrich. Network modelling methods for FMRI. Neuroimage. 2011;54:875-891.
14. Smith SM, Beckmann CF, Andersson J, et al. Resting-state fMRI in the Human Connectome Project. NeuroImage. 2013/10/15/ 2013;80:144-168.
15. Sotiropoulos SN, Jbabdi S, Xu J, et al. Advances in diffusion MRI acquisition and processing in the Human Connectome Project. NeuroImage. 2013/10/15/ 2013;80:125-143.
16. Uğurbil K, Xu J, Auerbach EJ, et al. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. NeuroImage. 2013/10/15/ 2013;80:80-104.
17. Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K. The WU-Minn Human Connectome Project: An overview. NeuroImage. 2013/10/15/ 2013;80:62-79.
18. Wandell BA, Dumoulin SO, Brewer AA. Visual Field Maps in Human Cortex. Neuron. 2007/10/25/ 2007;56(2):366-383.

Appendix E

A Comment On Computer Precision

In Manuscript 2, we are often quoting values of magnitude <10-16 ("p1" values in figures, the False Discovery Rate threshold associated with an alpha of 0.5 in General Linear Model statistical testing). The reader may question these values as computers can't really keep number smaller than this that are very meaningful. Matlab, which we use for our computations, has a default precision of 16 digits.

We acknowledge this problem, but for the sake of consistent presentation of the results we chose to keep the p1 values as they are. Additionally, we do find that there is some meaning behind these values, and this is explained here. Firstly, we are using floating point representation, which allows for us to retain 16 digits at any given power (not just from 0 to 10^-16). The loss in numerical precision at any power only becomes an issue when adding or subtracting. In a subtraction which involves digits beyond the 16 digit range captured in the largest number, any resultant digits beyond the original 16 digit power range will be randomly allocated by the computer when they are accessed. Subtractions occur at most 4 times in the computation of the p-value thresholds. In these subtractions involving very small p-values, the power of the result remains meaningful, it is the number that loses its meaning and is random. Since the number is random, it is possible that this will impact the power of the p-value, but given the few subtractions we do, the power shouldn't be inaccurate by more than a 10^+/-4. Thus, the approximate power of the p-value threshold should still be relevant. In any case, since the models which have very low p1 values show poor performance at all thresholds, getting a highly accurate threshold is not practically relevant.

Additional Reference:

Greaves, D.J., 2010. Floating Point Computation, A six-lecture course. Computer Laboratory, University of Cambridge.
https://www.cl.cam.ac.uk/teaching/1011/FPComp/fpcomp10slides.pdf

Appendix F

ICOV Pursued At Different Regularization Parameters

ICOV, or Regularized Inverse Covariance, is the model which has shown the most promise for outperforming all other methods tested including partial correlation, GES, and PC. ICOV can be viewed as a regularized version of partial correlation. The regularization used here is the Lasso method (Banerjee et al., 2006; Friedman et al., 2008), which has a parameter (lambda), that regularizes the covariance matrix to below a certain threshold. The goal of such regularization is to minimize false positive results by inducing a smaller spread in the interaction coefficients and making stronger connectivity measures less pronounced. This type of model seems perfectly suited for the task at hand where we seek models which minimize false positive results in a network analysis on a fine scale.

ICOV_5, ICOV with lambda=5x10^-3, is the computation that has been previously presented in both Smith et al. (2011) and Dawson et al. (2013), hence we have used this computation in our main presentation of ICOV. ICOV_5 was able to achieve the highest TPR-FPR among the models tested in both the quadrant analysis and the full visual cortex analysis with no additional thresholding. It was also the most stable model across thresholds in general. Given these findings, we tested ICOV at 8 different lambdas increasing from 0 (Partialcorr), with the other lambdas selected such that variation in the results was observable (Figure F.1, Table F.1). For the quadrant network the lambdas we pursued were 0, 1, 2.5, 5, 7.5, 10, 25, and 50 (all x10^-3). For the full visual cortex network the lambdas pursued were 0, 1, 5, 10, 50, 75, 100, and 300 (all x10^-3). With a higher regularization parameter, more significant connections are found because a smaller spread in interaction coefficients results in smaller standard deviation across measurements for a given interaction. ICOV with lambda>5 in the quadrant network, and lambda>10 in the full visual cortex network found more connections to be statistically significant than was optimal at an alpha of 0.05, finding many FP. Figure F.1 shows the proportion of TP and of FP for the 15 node network (top two rows), and the 60 node network (bottom two rows) for the different regularization parameters. Figure F.1 shows that at an appropriate lambda, ICOV is capable of identifying any TP that partial correlation has missed, but that capturing these additional TP often comes along with additional FP. The optimal TPR-FPR thresholds of the presented computations are given in Table F.1.

In the 15 node quadrant network, ICOV_5 is able to achieve a perfect network prediction (TPR-FPR=1) which is quite stable with 11 thresholds with TPR-FPR>0.65 within the range of [p1 to p1x10^{-20}], and 15 within the range of [p1 to p1x10^{-50}]. The perfect prediction is also found at higher lambdas (25, and 50) with even more stability in the range of [p1 to p1x10^{-50}] (17, and 18 thresholds with TPR-FPR>0.65 respectively), however additional thresholding is required. (Figure F.1; Table F.1)

In the 60 node full visual cortex network, ICOV is able to notably improve upon Partialcorr's performance both at high and low lambdas and with and without additional thresholding. ICOV_300 is able to achieve a TPR-FPR that is higher than that of ICOV_5, 0.88 compared to 0.83, but this requires thresholding beyond an alpha of 0.05. The ICOV networks are more stable than Partialcorr at any lambda in this larger network analysis, however there are never more than 3 thresholds with TPR-FPR>0.65 within the range of thresholds tested.

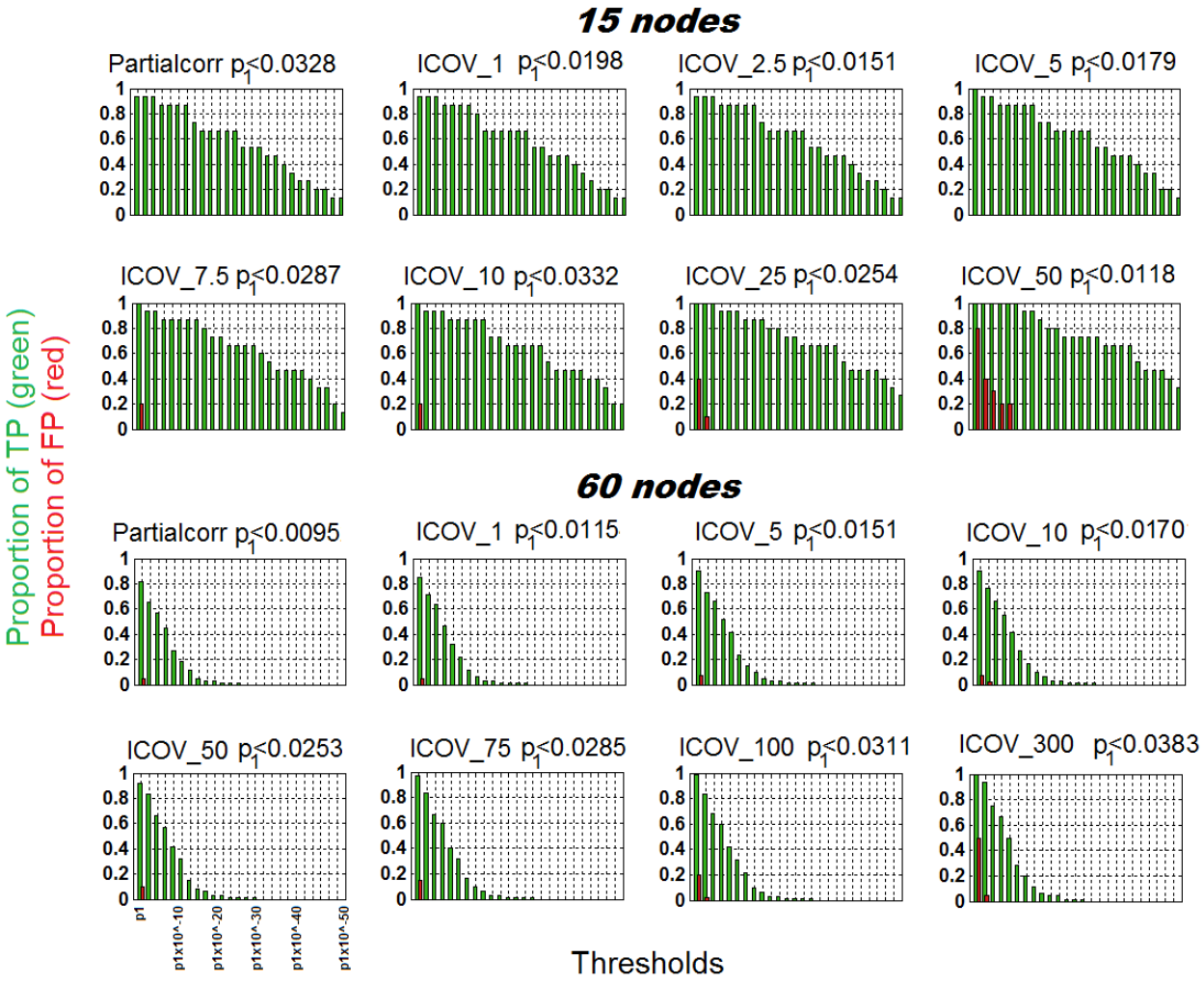


Fig. F.1. Proportion of true and false positive connections found for selected ICOV computations with differing regularization parameters (lambda) tested at various significance thresholds (alpha). ICOV is a regularized version of partial correlation which acts to increase high correlations, while decreasing low correlations, resulting in a greater spread of values and thus more statistical significance in the high correlations. Given the relative success of partial correlation, a range of regularization parameters were tested in order to observe if an improvement on the performance of Partialcorr could be seen. A selection of representative plots are shown at a range of lambdas. The top row presents the results given the 15 node, quadrant network; the bottom row presents the results given the 60 node, full visual cortex network. Green bars are the true positives (TP), and red bars are the false positives (FP). The p-value thresholds along the horizontal axis begin at the FDR corrected alpha (corrected from 0.05) indicated in the title above each plot ('p1'). The threshold p-values become progressively smaller by a multiplication factor of 10^-2 at each step (up to a minimum of 10^-50).

Table F.1: Threshold level of significant p-values that is necessary to maximize the proportion of True Positives minus the proportion of False Positives (TP-FP) for **ICOV** computed with various regularization parameters (lambda).

lambda	Common significance threshold; α=0.05, FDR corrected (p1):	Threshold <=0.05 such that TP-FP maximized; orders of magnitude of p1	TP-FP with optimal threshold	Number of thresholds with TP-FP within 15% of where TP-FP is optimized	Average step size (%)	Number of thresholds with TP-FP>=0.65 within range of [p1 to p1x10^{-20}]	Number of thresholds with TP-FP>=0.65 within range of [p1 to p1x10^{-50}]
15 node network							
0 (Partialcorr)	0.0328	p1	0.93	7	3.20	11	13
1	0.0199	p1	0.93	8	3.20	11	14
2.5	0.0151	p1	0.93	8	3.20	11	14
5	0.0180	p1	1.00	8	3.47	11	15
7.5	0.0287	-2	0.93	8	3.73	11	15
10	0.0332	-2	0.93	8	3.47	11	16
25	0.0254	-4	1.00	7	4.53	10	17
50	0.0118	-10	1.00	4	5.87	9	18
60 node network							
0 (Partialcorr)	0.0095	p1	0.77	1	6.25	1	1
1	0.0115	p1	0.80	2	6.03	2	2
5	0.0151	p1	0.83	2	5.77	3	3
10	0.0170	p1	0.83	2	5.77	3	3
50	0.0253	-2	0.83	1	5.95	3	3
75	0.0285	-2	0.83	1	5.95	3	3
100	0.0311	-2	0.81	1	5.83	3	3
300	0.0383	-2	0.88	1	9.62	3	3

Appendix G

Supplementary Material from Manuscript 3.

Outer Fold	TP	FP	TN	FN	SENS-ITIVITY	SPEC-IFICITY	IND. RATE	FPR	PPV	PPV corrected	NPV	NPV corrected	Accuracy
HCP													
1	0	0	10	0	-	1.000	0.787	0.000	-	-	1.000	-	60.9%
2	7	5	18	0	1.000	0.783	0.559	0.074	0.583	0.699	1.000	1.000	52.8%
3	28	26	2	0	1.000	0.071	0.176	0.382	0.519	0.353	1.000	1.000	61.0%
4	8	7	15	3	0.727	0.682	0.515	0.103	0.533	0.536	0.833	0.832	61.3%
5	14	23	3	0	1.000	0.115	0.412	0.338	0.378	0.364	1.000	1.000	57.9%
Mean	11.40	12.20	9.60	0.60	0.932	0.530	0.490	0.179	0.503	0.488	0.967	0.958	58.7%
Std	10.53	11.56	7.09	1.34	0.136	0.415	0.223	0.170	0.088	0.164	0.075	0.084	
DKT													
1	28	18	0	0	1.000	0.000	0.021	0.383	0.609	0.336	-	-	100.0%
2	27	24	1	1	0.964	0.040	0.221	0.353	0.529	0.337	0.500	0.689	83.3%
3	30	23	6	0	1.000	0.207	0.132	0.338	0.566	0.390	1.000	1.000	53.6%
4	15	11	4	1	0.938	0.267	0.544	0.162	0.577	0.393	0.800	0.894	69.7%
5	4	8	7	0	1.000	0.467	0.721	0.118	0.333	0.487	1.000	1.000	42.5%
Mean	20.80	16.80	3.60	0.40	0.980	0.196	0.328	0.271	0.523	0.388	0.825	0.896	62.1%
Std	11.08	7.12	3.05	0.55	0.029	0.188	0.294	0.122	0.110	0.061	0.236	0.147	

Supplementary Table 1. Metrics summarizing the classification results across all high-PPV HCP and DKT analyses. These metrics reflect the results presented in Figure 1. For the purpose of this table, a progressor subject that is correctly classified into a progressor class of interest is considered a true positive (TP) result; a stable subject that is correctly classified into a stable class of interest is considered a true negative (TN) result; a false positive (FP) result is a stable subject classified into a progressor class of interest; and a false negative (FP) result is a progressor subject classified into a stable class of interest. We consider any subject not classified into a class of interest as an indeterminate (IND) result. We present both the sample PPV (and NPV) and a PPV corrected (and NPV corrected) given the prevalence of progression from MCI to dementia. The prevalence of progression is estimated at 33.6%. Accuracy is computed as (TP+TN)/(TP+TN+FP+FN).The metrics associated with each of the outer-loop folds are presented in the rows, as well as the mean and standard deviation across these values, summarizing the overall metrics for the pipeline. Metrics for the pipeline iterations using HCP ROIs are presented in the top half of the table, and those for the iterations using the DKT ROIs are presented in t he bottom half.

Analysis Number	PPV	Num Features	Subjects Right	Duration Of High PPV	Classifier Type	Feature Type	FC Measure	Prog Classes	Stable Classes	Class Of Interest	Weight
HCP Outer Loop 1											
1	0.940	38	36	22	RF	FC	corr	3	3	2	1
2	0.969	44	19	28	RF	FC	corr	1	3	2	1
3	1.000	40	17	9	RF	FC	corr	1	3	2	5
4	0.952	56	13	3	RF	FC	corr	4	1	3	5
5	0.976	28	8	10	RF	GMCT		1	2	1	15
6	1.000	32	6	3	RF	FC	corr	3	3	2	10
7	0.955	27	5	8	RF	GMCT		1	3	1	25
8	1.000	22	4	4	RF	GMCT		3	1	4	35
9	1.000	13	3	3	SVM	GMCT		3	1	4	1
10	1.000	52	3	10	RF	GMCT		3	3	5	25
11	1.000	40	3	6	RF	GMCT		3	3	6	35
mean/mode all analyses	mean: 0.981	35.636	10.636	9.636	mode: RF	GMCT	corr	3	3		1
std/occurrence all analyses		std: 12.808	10.191	8.189	occurrence 0.909	0.545	1.000	0.545	0.636		0.273
mean/mode stable analyses	mean: 0.981	34.000	15.000	12.667	mode: RF	FC	corr	1	3		1
std/occurrence stable analyses		std: 6.882	11.754	9.543	occurrence 1.000	0.667	1.000	0.667	0.833		0.333
HCP Outer Loop 2											
1	0.955	60	23	45	RF	FC	corr	1	2	1	1
2	0.951	6	20	10	RF	FC	corr	3	2	1	5
3	0.980	34	19	45	RF	FC	corr	1	2	1	5
4	1.000	13	5	4	RF	GMCT		4	3	6	45
5	0.947	32	4	5	SVM	GMCT		1	2	2	1
6	1.000	21	4	4	RF	GMCT		3	1	4	15
7	1.000	11	4	4	RF	GMCT		4	1	3	15
8	1.000	28	4	3	RF	FC	corr	3	3	4	1
9	1.000	31	3	3	SVM	GMCT		3	3	3	1
10	1.000	9	3	4	RF	GMCT		3	1	3	10
11	1.000	10	3	4	RF	GMCT		3	1	3	15
12	1.000	10	3	4	RF	GMCT		4	2	5	10
13	1.000	12	3	5	RF	GMCT		4	2	4	15
14	1.000	28	3	4	SVM	FC	corr	3	3	4	1
15	0.947	36	3	7	RF	FC	corr	3	3	1	15
mean/mode all analyses	mean: 0.985	22.733	6.933	10.067	mode: RF	GMCT	corr	3	2		1
std/occurrence all analyses		std: 14.748	7.176	14.290	occurrence 0.800	0.600	1.000	0.533	0.400		0.333
mean/mode stable analyses	mean: 0.963	33.167	12.000	19.167	mode: RF	FC	corr	1	2		1
std/occurrence stable analyses		std: 17.163	9.592	20.144	occurrence 0.667	0.667	1.000	0.500	0.667		0.500

Supplementary Table 2. Parameters and options selected across all HCP ROI-based high-PPV analyses from the 5 outer-loop folds and computed results. Mean and mode results for individual folds are also presented. The mean is presented for results: PPV, subjects correctly classified (Subjects Right), and the duration of the stability of the high-PPV score across successive numbers of features (Duration of High PPV). The mean is also presented for parameter choice: number of features (Num Features). Standard deviation (std) is presented also to quantify the confidence on the mean measures presented. The std is not presented for the PPV as these values are selected to be within the range of 0.94-1. The mode is presented for parameters and options: classifier type (either Random

Forest –RF, or Support Vector Machine) , feature type (FC or GMCT), FC measure (corr or ICOV), number of progressor subtypes ("Prog classes": 1, 2, 3, or 4), number of stable subtypes ("Stable classes": 1, 2, 3, or 4), and weight associated with the progressor training data (1, 5, 10, 15, 25, 35, or 45). The second most common numbers of progressor and stable subtypes ("2nd") are also presented. The confidence in these mode values is quantified via the occurrence rate ("occurrence") of the mode value. All values are presented across all high-PPV analyses (all rows), and across stable-optimized only high-PPV analyses (shaded rows).

Analysis Number	PPV	Num Features	Subjects Right	Duration Of High PPV	Classifier Type	Feature Type	FC Measure	Prog Classes	Stable Classes	Class Of Interest	Weight
HCP Outer Loop 3											
1	1.000	16	38	39	SVM	FC	corr	1	3	3	1
2	0.968	56	31	8	RF	FC	corr	4	3	4	15
3	0.956	30	26	39	SVM	FC	corr	4	3	4	35
4	0.959	27	20	34	RF	FC	corr	4	3	2	1
5	1.000	28	18	36	SVM	FC	corr	4	3	2	1
6	0.944	26	12	4	RF	FC	corr	3	3	4	1
7	1.000	5	4	3	SVM	GMCT		4	3	5	5
8	1.000	11	3	3	SVM	GMCT		3	2	5	25
9	1.000	4	3	3	SVM	GMCT		4	3	4	1
10	1.000	18	3	7	SVM	GMCT		4	1	5	1
11	1.000	36	3	3	RF	GMCT		4	1	5	45
mean/mode all analyses	mean: 0.984	23.364	14.636	16.273	mode: SVM	FC	corr	4	3		1
std/occurrence all analyses		std: 15.028	12.824	16.572	occurrence 0.636	0.545	1.000	0.727	0.727		0.545
mean/mode stable analyses	mean: 0.986	23.667	25.333	36.333	mode: SVM	FC	corr	4	3		1
std/occurrence stable analyses		std: 6.658	11.015	2.517	occurrence 0.667	1.000	1.000	0.667	1.000		1.000
HCP Outer Loop 5											
1	0.982	6	66	44	SVM	FC	corr	3	2	5	15
2	0.946	6	65	9	SVM	FC	corr	3	1	4	15
3	0.946	4	63	8	SVM	FC	corr	3	1	4	10
4	0.941	7	58	44	SVM	FC	corr	3	3	6	10
5	0.962	20	23	3	RF	FC	corr	4	3	4	15
6	0.955	26	21	4	RF	FC	corr	4	2	3	45
7	0.955	9	20	3	RF	FC	corr	4	2	3	15
8	0.947	12	17	3	RF	FC	corr	4	1	2	10
9	0.987	21	6	45	RF	FC	corr	4	3	3	5
10	0.979	60	6	36	RF	FC	corr	1	3	3	1
11	1.000	15	5	3	RF	GMCT		4	1	3	25
12	1.000	38	5	40	RF	FC	corr	1	3	3	5
13	0.953	22	5	35	RF	FC	corr	1	2	2	1
14	1.000	32	3	3	RF	GMCT		4	1	4	10
15	0.983	46	3	41	SVM	FC	corr	4	2	2	1
16	1.000	17	3	7	RF	FC	corr	3	2	3	1
mean/mode all analyses	mean: 0.971	21.313	23.063	20.500	mode: RF	FC	corr	4	2		15
std/occurrence all analyses		std: 15.903	24.815	18.673	occurrence 0.688	0.875	1.000	0.500	0.375		0.250
mean/mode stable analyses	mean: 0.980	37.400	5.000	39.400	mode: RF	FC	corr	1	3		1
std/occurrence stable analyses		std: 16.517	1.225	4.037	occurrence 0.800	1.000	1.000	0.600	0.600		0.600

Supplementary Table 2 (continued).

Analysis Number	PPV	Num Features	Subjects Right	Duration Of High PPV	Classifier Type	Feature Type	FC Measure	Prog Classes	Stable Classes	Class Of Interest	Weight
HCP Outer Loop 4											
1	0.975	54	32	12	RF	FC	corr	4	2	6	35
2	1.000	21	31	45	SVM	FC	corr	4	1	5	15
3	0.972	60	26	18	RF	FC	corr	4	3	7	45
4	1.000	4	25	39	SVM	FC	corr	4	3	7	5
5	0.971	48	25	17	RF	FC	corr	4	2	6	45
6	0.971	44	24	15	RF	FC	corr	4	2	6	25
7	0.974	24	23	15	RF	FC	corr	4	3	7	35
8	0.952	20	17	7	RF	FC	corr	4	2	6	10
9	0.941	20	15	5	RF	FC	corr	4	2	6	5
10	1.000	29	14	33	SVM	FC	corr	1	3	3	5
11	0.964	15	14	33	RF	FC	corr	1	3	3	15
12	0.991	58	9	45	SVM	FC	corr	4	2	2	1
13	1.000	30	9	19	RF	FC	corr	4	3	3	15
14	0.992	16	9	45	RF	FC	corr	1	2	2	1
15	0.992	32	8	45	SVM	FC	corr	3	2	2	1
16	1.000	19	5	3	SVM	GMCT		4	3	6	10
17	1.000	12	5	3	RF	GMCT		3	3	5	5
18	1.000	27	5	20	RF	GMCT		4	3	5	45
19	1.000	13	5	19	SVM	FC	corr	3	2	4	1
20	1.000	31	5	10	RF	FC	corr	3	3	3	15
21	1.000	28	5	3	RF	FC	corr	3	1	4	1
22	1.000	12	4	3	SVM	GMCT		4	2	4	25
23	1.000	26	4	4	RF	GMCT		3	2	5	10
24	1.000	4	4	3	RF	GMCT		4	3	5	1
25	1.000	23	4	18	RF	GMCT		4	1	3	10
26	1.000	13	4	43	RF	FC	corr	3	3	2	5
27	1.000	7	4	23	RF	FC	corr	1	3	3	25
28	1.000	52	3	6	SVM	GMCT		3	3	1	5
29	1.000	10	3	3	SVM	GMCT		3	1	4	35
30	1.000	16	3	7	SVM	GMCT		4	3	6	35
31	1.000	25	3	6	RF	GMCT		4	3	5	10
32	1.000	20	3	9	RF	GMCT		4	1	3	5
33	1.000	28	3	19	RF	GMCT		4	1	3	35
34	1.000	27	3	5	SVM	FC	corr	4	2	1	25
mean/mode all analyses	mean: 0.991	25.529	10.471	17.647	mode: RF	FC	corr	4	3		5
std/occurrence all analyses		std: 14.943	9.229	14.641	occurrence 0.647	0.618	1.000	0.618	0.471		0.206
mean/mode stable analyses	mean: 0.994	28.182	7.455	27.909	mode: RF=SVM	FC	corr	1	3		5
std/occurrence stable analyses		std: 15.766	4.034	16.078	occurrence 0.500	0.909	1.000	0.364	0.636		0.273

Supplementary Table 2 (continued).

Analysis Number	PPV	Num Features	Subjects Right	Duration Of High PPV	Classifier Type	Feature Type	FC Measure	Prog Classes	Stable Classes	Class Of Interest	Weight
DKT Outer Loop 1											
1	1.000	40	17	9	RF	FC	corr	1	3	2	5
2	0.952	56	13	3	RF	FC	corr	4	1	3	5
3	0.976	28	8	10	RF	GMCT		1	2	1	15
4	1.000	32	6	3	RF	FC	corr	3	3	2	10
5	0.955	27	5	8	RF	GMCT		1	3	1	25
6	1.000	22	4	4	RF	GMCT		3	1	4	35
7	1.000	13	3	3	SVM	GMCT		3	1	4	1
8	1.000	52	3	10	RF	GMCT		3	3	5	25
9	1.000	40	3	6	RF	GMCT		3	3	6	35
mean/mode all analyses	0.987	34.444	6.889	6.222	RF	GMCT	corr	3	3		5
std/occurrence all analyses		13.929	4.986	3.073	0.889	0.667	1.000	0.556	0.556		0.222
mean/mode stable analyses	0.983	31.750	9.000	7.500	RF	FC= GMCT	corr	1	3		#N/A
std/occurrence stable analyses		5.909	5.477	3.109	1.000	0.500	1.000	0.667	0.667		0.000
DKT Outer Loop 2											
1	0.950	60	49	6	RF	GMCT		3	3	4	5
2	1.000	16	46	36	SVM	GMCT		3	3	4	10
3	0.968	48	44	14	RF	GMCT		3	2	3	5
4	1.000	44	40	9	RF	GMCT		3	1	2	5
5	1.000	48	18	7	SVM	GMCT		4	1	5	35
6	0.946	30	16	29	RF	FC	corr	4	2	2	10
7	0.964	60	12	5	SVM	GMCT		4	2	6	25
8	1.000	27	9	6	RF	GMCT		4	2	5	35
9	0.964	20	8	13	RF	FC	corr	4	2	2	15
10	1.000	38	4	14	RF	GMCT		4	2	5	15
11	0.944	58	4	8	SVM	FC	corr	3	1	2	10
12	1.000	13	3	7	RF	FC	corr	3	1	1	35
mean/mode all analyses	mean: 0.978	38.500	21.083	12.833	mode: RF	GMCT	corr	3	2		5
std/occurrence all analyses		std: 17.107	18.163	9.824	occurrence 0.667	0.667	1.000	0.500	0.500		0.250
mean/mode stable analyses	mean: 0.978	40.250	19.250	12.250	mode: RF	FC	corr	4	2		#N/A
std/occurrence stable analyses		std: 8.544	6.557	11.372	occurrence 1.000	1.000	1.000	0.667	0.667		0.000

Supplementary Table 3. Parameters and options selected across all DKT ROI-based high-PPV analyses from the 5 outer-loop folds and computed results. Presentation is as in Supplementary Table 2.

Analysis Number	PPV	Num Features	Subjects Right	Duration Of High PPV	Classifier Type	Feature Type	FC Measure	Prog Classes	Stable Classes	Class Of Interest	Weight
DKT Outer Loop 3											
1	0.992	40	57	38	SVM	FC	corr	1	2	2	1
2	0.948	58	34	3	SVM	GMCT		4	2	3	10
3	0.957	13	29	6	RF	FC	corr	4	3	2	1
4	1.000	56	4	7	SVM	GMCT		4	3	7	35
5	1.000	60	3	3	SVM	FC	corr	3	1	2	35
mean/mode all analyses	mean: 0.980	45.400	25.400	11.400	mode: SVM	FC	corr	4	2		1
std/occurrence all analyses		std: 19.769	22.612	14.977	occurrence 0.800	0.600	1.000	0.600	0.400		0.400
mean/mode stable analyses	mean: 0.975	26.500	43.000	22.000	mode: RF=SVM	FC	corr	#N/A	#N/A		1
std/occurrence stable analyses		std: 19.092	19.799	22.627	occurrence 0.500	1.000	1.000	0.000	0.000		1.000
DKT Outer Loop 4											
1	0.941	13	12	6	SVM	FC	corr	4	3	7	1
2	1.000	38	8	3	RF	GMCT		4	2	5	25
3	1.000	46	5	7	RF	GMCT		3	3	6	45
4	1.000	20	5	45	SVM	FC	corr	3	2	2	1
5	1.000	26	5	43	RF	FC	corr	1	2	2	1
6	1.000	29	4	15	RF	GMCT		3	2	5	10
7	1.000	44	4	10	RF	GMCT		3	1	4	15
8	1.000	36	4	4	RF	GMCT		4	2	5	10
9	1.000	19	4	5	SVM	FC	corr	3	3	4	25
10	1.000	31	4	3	RF	FC	corr	3	2	3	5
11	1.000	13	3	5	RF	GMCT		4	1	4	45
12	1.000	58	3	3	RF	FC	corr	4	2	3	1
mean/mode all analyses	mean: 0.995	31.083	5.083	12.417	mode: RF	FC= GMCT	corr	3	2		1
std/occurrence all analyses		std: 13.944	2.539	15.157	occurrence 0.750	0.500	1.000	0.500	0.583		0.333
mean/mode stable analyses	mean: 1.000	23.000	5.000	44	mode: RF=SVM	FC	corr	#N/A	2		1
std/occurrence stable analyses		std: 4.243	0.000	1.414	occurrence 0.500	1.000	1.000	0.000	1.000		1.000

Supplementary Table 3 (continued).

Analysis Number	PPV	Num Features	Subjects Right	Duration Of High PPV	Classifier Type	Feature Type	FC Measure	Prog Classes	Stable Classes	Class Of Interest	Weight
DKT Outer Loop 5											
1	0.969	58	19	5	RF	GMCT		3	1	4	35
2	1.000	21	18	4	RF	GMCT		4	3	5	15
3	0.964	58	17	4	RF	GMCT		3	1	4	45
4	0.967	34	17	7	RF	GMCT		4	2	4	45
5	1.000	28	16	9	RF	GMCT		4	3	5	10
6	0.969	34	16	6	RF	GMCT		4	2	4	35
7	1.000	58	15	3	RF	GMCT		3	1	4	15
8	1.000	48	13	3	RF	GMCT		3	2	5	35
9	1.000	32	13	5	RF	GMCT		4	3	5	25
10	0.963	18	13	18	RF	GMCT		4	2	4	10
11	1.000	26	13	43	RF	FC	corr	3	3	4	15
12	0.950	24	12	5	RF	GMCT		3	1	4	25
13	1.000	48	10	7	RF	GMCT		4	3	5	35
14	1.000	29	10	3	RF	GMCT		4	3	5	45
15	0.958	21	10	4	RF	GMCT		4	2	4	15
16	0.960	26	10	6	RF	GMCT		4	2	4	25
17	1.000	56	9	3	RF	GMCT		3	3	6	10
18	1.000	15	9	6	RF	GMCT		4	1	3	35
19	1.000	56	8	5	RF	GMCT		3	3	6	35
20	1.000	24	8	4	RF	GMCT		4	1	3	10
21	1.000	50	7	3	RF	GMCT		3	2	5	25
22	1.000	24	7	5	RF	GMCT		4	3	5	5
23	1.000	29	5	23	RF	GMCT		4	1	3	5
24	1.000	50	3	3	SVM	FC	corr	4	2	1	25
25	0.961	36	3	5	RF	FC	corr	3	3	2	5
26	0.970	24	3	7	RF	FC	corr	4	3	2	5
27	0.972	54	3	8	RF	FC	corr	1	3	3	5
28	1.000	40	3	14	RF	FC	corr	1	3	3	10
mean/mode all analyses	mean: 0.986	36.464	10.357	7.786	mode: RF	GMCT	corr	4	3		35
std/%occurrence all analyses		std: 14.198	4.998	8.319	occurrence 0.964	0.786	1.000	0.571	0.464		0.214
mean/mode stable analyses	mean: 0.981	40.800	3.000	7.400	mode: RF	FC	corr	4	3		5
std/%occurrence stable analyses		std: 11.883	0.000	4.159	occurrence 0.667	0.500	1.000	0.667	0.667		0.667

Supplementary Table 3 (continued).

		GMV & CT	FC	all feature types	number of instances of number of classes being optimal			
HCP								
outer loop	rank	[prog classes, stable classes]	[prog classes, stable classes]	[prog classes, stable classes]	Prog 3	Prog 4	Stable 2	Stable 3
1	1	[4,3]	[3,2]	[3,2]	2	1	2	1
	2	[3,2]	[4,3]	[4,3]	1	2	1	2
2	1	[3,2]	[3,2]	[3,2]	3	0	3	0
	2	[4,3]	[4,3]	[4,3]	0	3	0	3
3	1	[3,3]	[3,3]	[4,3]	2	1	0	3
	2	[4,2]	[4,2]	[3,2]	1	2	3	0
4	1	[3,3]	[3,2]	[3,3]	3	0	1	2
	2	[4,2]	[4,3]	[4,2]	0	3	2	1
5	1	[3,3]	[3,2]	[4,2]	2	1	2	1
	2	[4,2]	[4,3]	[3,3]	1	2	1	2
Highest ranked by feature type		[3,3]	[3,2]	[3,2]	weighted sum of instances across folds (2*rank1+1*rank2):			
					27	18	23	22
DKT								
1	1	[3,3]	[4,2]	[3,2]	2	1	2	1
	2	[4,2]	[3,3]	[4,3]	1	2	1	2
2	1	[3,2]	[3,2]	[4,3]	2	1	2	1
	2	[4,3]	[4,3]	[3,2]	1	2	1	2
3	1	[4,2]	[3,2]	[3,3]	2	1	2	1
	2	[3,3]	[4,3]	[4,2]	1	2	1	2
4	1	[3,3]	[3,2]	[3,2]	3	0	2	1
	2	[4,2]	[4,3]	[4,3]	0	3	1	2
5	1	[3,3]	[3,2]	[3,2]	3	0	2	1
	2	[4,2]	[4,3]	[4,3]	0	3	1	2
Highest ranked by feature type		[3,3]	[3,2]	[3,2]	weighted sum of instances across folds (2*rank1+1*rank2):			
					27	18	25	20

Supplementary Table 4. Optimal number of subtypes computed on the training data from each of the 5 outer-loop folds. The optimal number of subtypes is computed both via the eigen-gaps and rotation costs methods, both yielding the same results. The best (rank1) and second best (rank2) number of subtypes are presented. Subtypes are computed separated for the structural features (i.e. GMVT&CT), and functional features (I.e. FC), as well as all features types together using the Similarity Network Fusion method. Within each set of square brackets are the optimal [number of progressor (prog) classes, and number of stable classes]. The "highest ranked" value in each column is selected as the number of subtypes most often identified as optimal. The number of instances a given number of subtypes was identified as optimal across feature type used is counted in the last 4 columns of the table. A weighted sum across these values is computed to identify the overall optimal number of subtypes. The weighted sum is computed in each column as [2*rank1+1*rank2]. Shaded table entries identify the optimal numbers of subtypes.

HCP ROI-based features	Subtypes contrasted									
GMV	12	13	14	15	23	24	25	34	35	45
GMV. lh.24dv	-1	1	1		1	1	1			
GMV. lh.31pd										
GMV. lh.46	-1		-1				1	-1		1
GMV. lh.9a	-1				1					
GMV. lh.A4	-1			1	1		1			
GMV. lh.AIP							1			1
GMV. lh.d23ab	-1	1	1		1	1	1			
GMV. lh.EC	-1	1		1	1	1	1		1	1
GMV. lh.IP1	-1				1					
GMV. lh.LO2				-1			-1			
GMV. lh.OFC	-1	1		1	1		1	-1		1
GMV. lh.p32	-1			1	1	1	1			
GMV. lh.PCV					1	1				
GMV. lh.PeEc	-1	1		1	1		1		1	1
GMV. lh.PHT		1								
GMV. lh.PreS	-1	1			1		1			
GMV. lh.PSL	-1	1	1	1	1	1	1			
GMV. lh.STGa		1		1	1	1	1			
GMV. lh.STSdp	-1	1		1	1	1	1		1	1
GMV. lh.STSvp	-1	1		1	1		1		1	1
GMV. lh.STV	-1	1		1	1	1	1			
GMV. lh.TE1p	-1	1		1	1		1			
GMV. lh.TE2a	-1	1			1	1	1			
GMV. lh.V3A										
GMV. lh.caudate	-1	1		1	1	1	1		1	1
GMV. lh.globuspallidus										
GMV. lh.thalamus	-1	1		1	1	1	1		1	1
GMV.rh.43					1	1				
GMV.rh.45										
GMV.rh.46	-1				1		1			
GMV.rh.4	-1				1	1				
GMV.rh.52	-1	1		1	1	1	1			
GMV.rh.7m		1			1	1	1			
GMV.rh.9-46d	-1				1	1	1			
GMV.rh.a24	-1				1	1	1			
GMV.rh.a32pr										
GMV.rh.a9-46v					1		1			1
GMV.rh.EC	-1	1		1	1		1	-1		1
GMV.rh.FFC	-1		-1		1		1	-1		1
GMV.rh.IFSa		1								
GMV.rh.IP0	-1				1					
GMV.rh.MBelt					1		1			
GMV.rh.OP4	-1				1					
GMV.rh.PBelt										
GMV.rh.PeEc	-1	1		1	1	1	1			
GMV.rh.PF					1					
GMV.rh.PHA2	-1	1			1		1			
GMV.rh.PH		1		1	1		1	-1		1
GMV.rh.TE1m	-1				1	1	1			
GMV.rh.TE1p	-1	1			1		1			
GMV.rh.TE2p	-1	1			1	1	1			
GMV.rh.V3B	-1				1					
GMV.rh.amygdala					1	1				
GMV.rh.caudate	-1	1		1	1	1	1			1
GMV.rh.thalamus	-1	1		1	1	1	1		1	1

Supplementary Table 5. HCP ROI-based GMV features which were found to have statistically significant differences between MCI subtypes. Subtypes 1 and 2 are stable subtypes and subtypes 3, 4, and 5 are progressor subtypes. A "1" indicates a statistically significant difference where the subtype named with the smaller number has *greater GMV* than that named with the larger number. A "-1" indicates a statistically significant difference where the

subtype named with the smaller number has *smaller GMV* than that named with the larger number. lh=left hemisphere, rh=right hemisphere. ROI naming convention is as defined in Glasser et al. (2016).

HCP ROI-based features	Subtypes contrasted										HCP ROI-based features	Subtypes contrasted									
CT	12	13	14	15	23	24	25	34	35	45	CT	12	13	14	15	23	24	25	34	35	45
CT. lh.13l		1		1	1		1		1	1	CT. lh.TF		1		1	1		1			1
CT. lh.25		1		1			1		1	1	CT. lh.TGd		1		1	1		1		1	1
CT. lh.31a	-1				1	1	1				CT. lh.TGv		1		1	1		1			1
CT. lh.31pd	-1				1		1	-1		1	CT. lh.V4	-1	1		1	1		1	-1		1
CT. lh.3a			-1	1	1	-1	1	-1	1	1	CT. lh.V8	-1			1	1		1		1	1
CT. lh.3b			1					1			CT. lh.VMV1	-1				1	1	1			
CT. lh.47l		1			1						CT. lh.VVC	-1	1		1	1	1	1			
CT. lh.47s				1			1			1	CT.rh.23c	-1			1	1	1	1		1	1
CT. lh.5mv	-1				1		1				CT.rh.24dd	-1				1		1			
CT. lh.6a	-1			1	1	1	1				CT.rh.3a				1	1		1			1
CT. lh.7Am	-1			1	1		1			1	CT.rh.3b	1		1		-1			1		-1
CT. lh.7PL											CT.rh.52	-1			1	1	1	1		1	1
CT. lh.7Pm					1		1				CT.rh.5m	-1				1		1	-1		1
CT. lh.8BL		1			1						CT.rh.6a					1					
CT. lh.8BM	-1				1		1				CT.rh.7PL	-1	1			1		1			
CT. lh.9m	-1			1	1	1	1		1	1	CT.rh.7Pm	-1	1		1	1	1	1		1	
CT. lh.a24pr											CT.rh.a47r				1			1			1
CT. lh.A4	-1			1	1	1	1		1	1	CT.rh.A5	-1			1	1	1	1		1	1
CT. lh.A5	-1	1		1	1	1	1		1	1	CT.rh.a9-46v	-1	1		1	1	1	1			
CT. lh.AAIC			-1				1	-1		1	CT.rh.AAIC		1		1	1		1	-1		1
CT. lh.AVI			-1				1	-1		1	CT.rh.d32	-1				1		1			
CT. lh.EC		1		1	1		1		1	1	CT.rh.DVT	-1	1		1	1	1	1			
CT. lh.FFC	-1	1		1	1	1	1				CT.rh.EC	-1				1		1			
CT. lh.H	-1	1	1	1	1	1	1				CT.rh.FOP1				1	1		1			1
CT. lh.i6-8	-1				1		1				CT.rh.FOP5		1		1	1		1			1
CT. lh.IP0	-1			1	1		1				CT.rh.IFJp					1			-1		1
CT. lh.LIPd		1		1	1	1	1				CT.rh.IP2	-1			1	1	1	1			
CT. lh.LO1				1			1	-1	1	1	CT.rh.LIPd		1		1	1		1			
CT. lh.LO3		1		1	1		1			1	CT.rh.p10p	-1				1	1	1			
CT. lh.MBelt	-1			1	1		1		1	1	CT.rh.p24	-1			1	1	1	1			
CT. lh.MST				1			1			1	CT.rh.PCV	-1	1		1	1	1	1		1	1
CT. lh.OFC	-1			1	1		1		1	1	CT.rh.PGi	-1	1		1	1	1	1		1	1
CT. lh.p32	-1			1	1	1	1		1	1	CT.rh.PHA1	-1	1			1	1	1			
CT. lh.PBelt	-1			1	1	1	1		1	1	CT.rh.PHA2	-1			1	1	1	1		1	
CT. lh.PCV	-1			1	1		1				CT.rh.PH	-1			1	1		1		1	1
CT. lh.PeEc		1		1	1		1		1	1	CT.rh.Pir		1		1	1		1			1
CT. lh.PF	-1	1		1	1		1				CT.rh.pOFC		1		1	1		1	-1		1
CT. lh.PFop	-1			1	1		1		1	1	CT.rh.RI				1	1		1		1	1
CT. lh.PFt					1						CT.rh.RSC		1			1			-1		1
CT. lh.PGi	-1			1	1		1			1	CT.rh.s32	-1	1		1	1	1	1			1
CT. lh.PGs		1		1	1		1				CT.rh.STGa	-1	1			1	1	1			
CT. lh.PH	-1				1		1				CT.rh.STSda		1	1	1	1	1	1			
CT. lh.PHT		1			1		1				CT.rh.STSdp		1		1	1		1			
CT. lh.PoI2				1			1		1	1	CT.rh.STSva	-1	1		1	1	1	1			
CT. lh.PreS		1		1	1		1				CT.rh.STSvp		1		1	1		1			1
CT. lh.PSL		1		1							CT.rh.STV	-1	1	1	1	1	1	1			
CT. lh.STGa		1		1	1		1			1	CT.rh.TA2		1		1	1		1			
CT. lh.STSdp		1		1	1	1	1		1	1	CT.rh.TE1p	-1	1		1	1		1			1
CT. lh.STSvp		1		1	1		1		1	1	CT.rh.TE2p		1		1	1		1		1	1
CT. lh.STV		1			1						CT.rh.TF		1			1					
CT. lh.TE1a	-1	1		1	1		1				CT.rh.V3B	-1	1		1	1		1			1
CT. lh.TE1m		1		1	1		1		1	1											
CT. lh.TE1p		1		1	1		1														
CT. lh.TE2a	-1	1		1	1	1	1			1											
CT. lh.TE2p	-1	1	1	1	1	1	1														

Supplementary Table 6. HCP ROI-based CT features which were found to have statistically significant differences between MCI subtypes. Subtypes 1 and 2 are stable subtypes and subtypes 3, 4, and 5 are progressor subtypes. A "1" indicates a statistically significant difference where the subtype named with the smaller number has *greater CT* than that named with the larger number. A "-1" indicates a statistically significant difference where the subtype named with the smaller number has *smaller CT* than that named with the larger number.. lh=left hemisphere, rh=right hemisphere. ROI naming convention is as defined in Glasser et al. (2016).

HCP ROI-based features		Subtypes contrasted										HCP ROI features		Subtypes contrasted									
FC		12	13	14	15	23	24	25	34	35	45	FC		12	13	14	15	23	24	25	34	35	45
lh.1	lh.3a	1		1		-1		-1	1	1	-1	lh.3a	rh.V2	1		1	1	-1		-1	1	1	-1
lh.1	lh.5m	1		1		-1		-1	1	1	-1	lh.3a	rh.V3	1		1	1	-1		-1	1	1	-1
lh.1	lh.6mp	1		1	1	-1			1	1	-1	lh.3a	rh.V6	1		1	1	-1		-1	1	1	-1
lh.1	lh.V2	1		1	1	-1		-1	1	1	-1	lh.3a	rh.V8	1		1	1	-1		-1	1	1	-1
lh.1	lh.V3A	1		1	1	-1		-1	1	1	-1	lh.3a	rh.VMV1	1		1	1	-1		-1	1	1	-1
lh.1	lh.V3	1		1	1	-1		-1	1	1	-1	lh.3a	rh.VMV3	1		1	1	-1		-1	1	1	-1
lh.1	lh.V6	1		1	1	-1		-1	1	1	-1	lh.3a	rh.VVC	1		1	1	-1		-1	1	1	-1
lh.1	lh.V7	1		1	1	-1		-1	1	1	-1	lh.3b	lh.4	1		1	1	-1		-1	1	1	-1
lh.1	h.VMV3	1		1	1	-1		-1	1	1	-1	lh.3b	lh.6mp	1		1	1	-1			1	1	-1
lh.1	lh.VVC	1		1	1	-1		-1	1	1	-1	lh.3b	lh.V2	1		1	1	-1		-1	1	1	-1
lh.1	rh.2	1		1	1	-1		-1	1	1	-1	lh.3b	lh.V3	1		1	1	-1		-1	1	1	-1
lh.1	rh.3a	1		1	1	-1		-1	1	1	-1	lh.3b	lh.V6	1		1	1	-1		-1	1	1	-1
lh.1	rh.3b	1		1	1	-1		-1	1	1	-1	lh.3b	lh.VMV1	1		1	1	-1		-1	1	1	-1
lh.1	rh.4	1		1	1	-1		-1	1	1	-1	lh.3b	lh.VMV3	1		1	1	-1		-1	1	1	-1
lh.1	rh.5m	1		1		-1		-1	1	1	-1	lh.3b	rh.1	1		1	1	-1			1	1	-1
lh.1	rh.5mv	1		1	1	-1		-1	1	1	-1	lh.3b	rh.24dd	1		1	1	-1		-1	1	1	-1
lh.1	rh.OP4	1		1		-1			1	1	-1	lh.3b	rh.2	1		1	1	-1		-1	1	1	-1
lh.1	rh.V1	1		1	1	-1		-1	1	1	-1	lh.3b	rh.3a	1		1	1	-1	1	-1	1	1	-1
lh.1	rh.V2	1		1	1	-1		-1	1	1	-1	lh.3b	rh.3b	1		1	1	-1		-1	1	1	-1
lh.1	rh.V3	1		1	1	-1		-1	1	1	-1	lh.3b	rh.4	1		1	1	-1		-1	1	1	-1
lh.1	rh.V4	1		1	1	-1		-1	1	1	-1	lh.3b	rh.5m	1		1	1	-1		-1	1	1	-1
lh.1	rh.V6	1		1	1	-1		-1	1	1	-1	lh.3b	rh.5mv	1		1	1	-1		-1	1	1	-1
lh.1	rh.V8	1		1	1	-1		-1	1	1	-1	lh.3b	rh.OP4	1		1		-1			1	1	-1
lh.1	rh.VMV1	1		1	1	-1		-1	1	1	-1	lh.3b	rh.V1	1		1	1	-1		-1	1	1	-1
lh.1	rh.VMV3	1		1	1	-1		-1	1	1	-1	lh.3b	rh.V2	1		1	1	-1		-1	1	1	-1
lh.23c	rh.V3B	1		1	1	-1	1		1	1	-1	lh.3b	rh.V3A	1		1	1	-1		-1	1	1	-1
h.24dd	lh.5m	1		1	1	-1			1	1		lh.3b	rh.V3	1		1	1	-1		-1	1	1	-1
h.24dd	rh.3b	1		1	1	-1		-1	1	1	-1	lh.3b	rh.V4	1		1	1	-1		-1	1	1	-1
h.24dd	rh.4	1		1	1	-1			1	1		lh.3b	rh.V6	1		1	1	-1		-1	1	1	-1
lh.2	lh.V2	1		1	1	-1		-1	1	1	-1	lh.3b	rh.V8	1		1	1	-1		-1	1	1	-1
lh.2	lh.V3	1		1	1	-1		-1	1	1	-1	lh.3b	rh.VMV1	1		1	1	-1		-1	1	1	-1
lh.2	lh.V6	1	-1	1	1	-1		-1	1	1	-1	lh.3b	rh.VMV3	1		1	1	-1		-1	1	1	-1
lh.2	lh.V7	1		1	1	-1		-1	1	1	-1	lh.3b	rh.VVC	1		1	1	-1		-1	1	1	-1
lh.2	h.VMV3	1	-1	1	1	-1		-1	1	1	-1	lh.4	lh.V2	1		1	1	-1		-1	1	1	-1
lh.2	rh.1	1		1	1	-1			1	1		lh.4	lh.V3	1		1	1	-1		-1	1	1	-1
lh.2	rh.3a	1		1	1	-1		-1	1	1	-1	lh.4	lh.V6	1		1	1	-1		-1	1	1	-1
lh.2	rh.3b	1		1	1	-1		-1	1	1	-1	lh.4	lh.VMV1	1		1	1	-1		-1	1	1	-1
lh.2	rh.ProS	1		1	1	-1		-1	1	1	-1	lh.4	lh.VVC	1		1	1	-1		-1	1	1	-1
lh.2	rh.V2	1		1	1	-1		-1	1	1	-1	lh.4	rh.1	1		1	1	-1			1	1	-1
lh.2	rh.V3	1		1	1	-1		-1	1	1	-1	lh.4	rh.24dd	1		1		-1			1	1	-1
lh.2	rh.V4	1		1	1	-1		-1	1	1	-1	lh.4	rh.2	1		1	1	-1		-1	1	1	-1
lh.2	rh.V6	1		1	1	-1		-1	1	1	-1	lh.4	rh.3a	1		1	1	-1		-1	1	1	-1
lh.2	rh.V7	1		1	1	-1		-1	1	1	-1	lh.4	rh.3b	1		1	1	-1		-1	1	1	-1
lh.2	rh.V8	1		1	1	-1		-1	1	1	-1	lh.4	rh.4	1		1	1	-1		-1	1	1	-1
lh.2	rh.VMV1	1	-1	1	1	-1		-1	1	1	-1	lh.4	rh.5mv	1		1		-1		-1	1	1	-1
lh.31a	lh.7m	1				-1						lh.4	rh.OP4	1		1		-1	1		1	1	-1
lh.31a	rh.7m	1		1		-1						lh.4	rh.ProS	1		1	1	-1		-1	1	1	-1
lh.31a	rh.POS2	1		1		-1			1			lh.4	rh.V2	1		1	1	-1		-1	1	1	-1
lh.31pd	lh.d23ab											lh.4	rh.V3	1		1	1	-1		-1	1	1	-1
lh.31pd	rh.31pv	1										lh.4	rh.V4	1		1	1	-1		-1	1	1	-1
lh.31pd	rh.RSC	1										lh.4	rh.V6	1		1	1	-1		-1	1	1	-1
lh.31pd	rh.v23ab	1				-1						lh.4	rh.V8	1		1	1	-1		-1	1	1	-1
lh.3a	lh.3b	1		1		-1		-1	1	1	-1	lh.4	rh.VMV1	1		1	1	-1		-1	1	1	-1
lh.3a	lh.4	1		1	1	-1		-1	1	1	-1	lh.4	rh.VMV2	1		1	1	-1		-1	1	1	-1
lh.3a	lh.V2	1		1	1	-1		-1	1	1	-1	lh.4	rh.VMV3	1		1	1	-1		-1	1	1	-1
lh.3a	lh.V6	1		1	1	-1		-1	1	1	-1	lh.4	rh.VVC	1		1	1	-1		-1	1	1	-1
lh.3a	rh.1	1	-1	1		-1		-1	1	1	-1	lh.5m	rh.24dd	1		1	1	-1			1	1	
lh.3a	rh.24dd	1		1	1	-1			1	1	-1	lh.5m	rh.31a	1		1		-1					
lh.3a	rh.2	1	-1	1		-1		-1	1	1	-1	lh.5m	rh.3a	1		1	1	-1		-1	1	1	-1
lh.3a	rh.3a	1		1	1	-1	1	-1	1	1	-1	lh.5m	rh.5mv	1		1		-1		-1	1	1	-1

HCP ROI features		Subtypes contrasted										HCP ROI-based features		Subtypes contrasted									
FC		12	13	14	15	23	24	25	34	35	45	FC		12	13	14	15	23	24	25	34	35	45
lh.3a	rh.3b	1		1	1	-1		-1	1	1	-1	lh.6d	h.VMV1	1		1	1	-1		-1	1	1	-1
lh.3a	rh.4	1		1	1	-1		-1	1	1	-1	lh.6d	rh.3a	1		1	1	-1			1	1	-1
lh.3a	rh.OP4	1		1		-1	1	-1	1		-1	lh.6d	rh.5mv	1		1		-1		-1	1	1	-1
lh.6d	rh.VMV1	1		1	1	-1		-1	1	1	-1	lh.V6	rh.3b	1	-1	1	1	-1		-1	1	1	-1
lh.6mp	rh.3a	1		1	1	-1		-1	1	1	-1	lh.VMV1	rh.1	1	-1	1	1	-1		-1	1	1	-1
lh.7m	lh.d23ab											lh.VMV1	rh.2	1	-1	1	1	-1		-1	1	1	-1
lh.7m	lh.RSC					-1						lh.VMV1	rh.3a	1		1	1	-1		-1	1	1	-1
lh.7m	rh.31pv					-1						lh.VMV1	rh.3b	1		1	1	-1		-1	1	1	-1
lh.7m	rh.v23ab											lh.VMV1	rh.4	1		1	1	-1		-1	1	1	-1
lh.DVT	rh.3a	1		1	1	-1			1	1	-1	lh.VMV1	rh.5m	1		1	1	-1		-1	1	1	-1
lh.FFC	lh.V7	1		1	1	-1			1	1		lh.VMV3	rh.2	1	-1	1	1	-1		-1	1	1	-1
lh.FST	lh.IP0	1		1	1	-1			1	1		lh.VMV3	rh.3a	1		1	1	-1		-1	1	1	-1
lh.FST	lh.IPS1	1		1	1	-1			1	1		lh.VVC	rh.7Am	1		1	1	-1			1	1	
lh.FST	lh.V3B	1		1	1	-1			1	1		rh.1	rh.3a	1		1		-1	1		1	1	-1
lh.FST	rh.V6	1		1	1	-1			1	1		rh.1	rh.5mv	1		1		-1			1	1	-1
lh.IP0	lh.LO1	1		1	1	-1			1	1		rh.1	rh.V2	1	-1	1	1	-1	1	-1	1	1	-1
lh.IP0	lh.LO2	1		1	1	-1			1	1		rh.1	rh.V6	1	-1	1	1	-1		-1	1	1	-1
lh.IP0	lh.LO3	1		1	1	-1	1		1	1		rh.1	rh.VMV1	1	-1	1	1	-1		-1	1	1	-1
lh.IP0	lh.PH	1		1	1				1	1		rh.24dd	rh.3a	1		1		-1			1	1	-1
lh.IP0	lh.V3CD	1		1	1		1		1			rh.24dd	rh.3b	1		1		-1		-1	1	1	-1
lh.IP0	lh.V3	1		1	1	-1			1	1		rh.24dd	rh.4	1		1		-1			1	1	
lh.IP0	lh.V4t	1		1	1	-1	1		1	1		rh.24dd	rh.5m	1		1		-1		-1	1	1	
lh.IP0	lh.V8	1		1	1	-1			1	1		rh.2	rh.7AL	1		1		-1			1	1	
lh.IP0	rh.LO3	1		1	1	-1			1	1		rh.2	rh.V2	1		1	1	-1		-1	1	1	-1
lh.IP0	rh.V3	1		1	1	-1			1	1		rh.2	rh.V6	1	-1	1	1	-1		-1	1	1	-1
lh.IP0	rh.V6A	1		1	1	-1	1		1	1		rh.2	rh.VMV1	1	-1	1	1	-1		-1	1	1	-1
lh.IP0	rh.VMV3	1		1	1	-1			1	1		rh.31pv	rh.PGs										
lh.LO1	lh.PH	1		1	1	-1			1	1		rh.31pv	rh.POS1										
lh.LO2	lh.PGp	1		1	1				1			rh.3a	rh.3b	1		1	1	-1	1		1	1	-1
lh.LO3	lh.V3B	1		1	1	-1			1	1		rh.3a	rh.4	1		1	1	-1		-1	1	1	-1
lh.MT	lh.VIP	1		1	1	-1			1	1		rh.3a	rh.5L	1		1	1	-1			1	1	-1
lh.OP4	rh.PBelt	1		1					1		-1	rh.3a	rh.5m	1		1	1	-1		-1	1	1	-1
lh.PGi	rh.v23ab	1										rh.3a	rh.5mv	1		1	1	-1			1	1	-1
lh.PGp	lh.V4t	1		1	1				1			rh.3a	rh.6mp	1		1	1	-1	1		1	1	-1
lh.PGs	rh.v23ab	1										rh.3a	rh.V1	1		1	1	-1		-1	1	1	-1
lh.PH	lh.V3A	1		1	1	-1			1	1		rh.3a	rh.V2	1		1	1	-1		-1	1	1	-1
lh.PH	lh.V3B	1		1	1	-1			1	1		rh.3a	rh.V3	1		1	1	-1		-1	1	1	-1
lh.PH	lh.V3CD	1		1	1	-1			1	1		rh.3a	rh.V6	1		1	1	-1			1	1	-1
lh.PH	lh.V4	1		1	1	-1			1	1		rh.3a	rh.V8	1		1	1	-1		-1	1	1	-1
lh.PH	lh.V6A	1		1	1	-1			1	1		rh.3a	rh.VMV1	1		1	1	-1		-1	1	1	-1
lh.PH	lh.V7	1		1	1	-1			1	1		rh.3a	rh.VMV3	1		1	1	-1		-1	1	1	-1
lh.PH	rh.PGp	1		1	1				1			rh.3b	rh.4	1		1	1	-1		-1	1	1	-1
lh.PH	rh.V3A	1		1	1	-1			1	1		rh.3b	rh.5mv	1		1	1	-1		-1	1	1	-1
lh.PH	rh.V3	1		1	1	-1			1	1		rh.3b	rh.OP4	1		1		-1	1	-1	1	1	-1
lh.PH	rh.V6A	1		1	1	-1	1		1	1		rh.3b	rh.V1	1		1	1	-1		-1	1	1	-1
lh.PH	rh.V7	1		1	1	-1			1	1		rh.3b	rh.V2	1		1	1	-1		-1	1	1	-1
lh.TPOJ3	lh.V3CD	1		1	1				1			rh.3b	rh.V3	1		1	1	-1		-1	1	1	-1
lh.TPOJ3	lh.V4	1		1	1				1			rh.3b	rh.V6	1		1	1	-1		-1	1	1	-1
lh.TPOJ3	lh.V4t	1		1	1	-1			1			rh.3b	rh.VMV1	1		1	1	-1		-1	1	1	-1
lh.TPOJ3	rh.V3	1		1	1				1			rh.3b	rh.VMV3	1		1	1	-1		-1	1	1	-1
lh.TPOJ3	rh.V4	1		1	1	-1			1	1		rh.4	rh.5mv	1		1		-1			1	1	-1
lh.v23ab	rh.7m	1				-1			1			rh.4	rh.OP4	1		1		-1	1		1	1	-1
lh.v23ab	rh.v23ab	1										rh.4	rh.V2	1		1	1	-1		-1	1	1	-1
lh.V2	rh.1	1	-1	1	1	-1	1	-1	1	1	-1	rh.4	rh.V3	1		1	1	-1		-1	1	1	-1
lh.V2	rh.2	1	-1	1	1	-1	1	-1	1	1	-1	rh.4	rh.V6	1		1	1	-1			1	1	-1
lh.V2	rh.3a	1		1	1	-1		-1	1	1	-1	rh.4	rh.VMV1	1		1	1	-1		-1	1	1	-1
lh.V2	rh.3b	1		1	1	-1		-1	1	1	-1	rh.5m	rh.5mv	1		1		-1		-1	1	1	
lh.V2	rh.4	1		1	1	-1	1	-1	1	1	-1	rh.5m	rh.OP4	1		1	1	-1			1	1	-1
lh.V3A	rh.2	1	-1	1	1	-1		-1	1	1	-1	rh.5m	rh.V8	1		1	1	-1		-1	1	1	-1
lh.V3B	lh.V4t	1		1	1	-1			1	1		rh.5m	rh.VMV3	1		1	1	-1		-1	1	1	-1

FC		12	13	14	15	23	24	25	34	35	45	FC		12	13	14	15	23	24	25	34	35	45
lh.V3B	lh.V6A	1		1	1	-1	1		1	1		rh.5m	rh.VVC	1		1	1	-1		-1	1	1	-1
lh.V3	rh.3a	1		1	1	-1		-1	1	1	-1	rh.5mv	rh.6ma	1		1		-1		-1	1		-1
lh.V3	rh.3b	1		1	1	-1		-1	1	1	-1	rh.5mv	rh.V3	1		1	1	-1		-1	1	1	-1
lh.V4t	lh.V7	1		1	1	-1			1	1		rh.d23ab	rh.PGs										
lh.V6A	rh.1	1	-1	1	1	-1			1	1	-1	rh.OP4	rh.VMV1	1		1	1	-1			1	1	-1
lh.V6	rh.3a	1	-1	1	1	-1		-1	1	1	-1	lh.V7	rh.3a	1		1	1	-1		-1	1	1	-1

Supplementary Table 7. HCP ROI-based FC features which were found to have statistically significant differences between MCI subtypes. Subtypes 1 and 2 are stable subtypes and subtypes 3, 4, and 5 are progressor subtypes. A "1" indicates a statistically significant difference where the subtype named with the smaller number has *greater FC* than that named with the larger number. A "-1" indicates a statistically significant difference where the subtype named with the smaller number has *smaller FC* than that named with the larger number. lh=left hemisphere, rh=right hemisphere. ROI naming convention is as defined in Glasser et al. (2016).

DKT ROI-based features	Subtypes contrasted									
GMV	12	13	14	15	23	24	25	34	35	45
lh.caudalanteriorcingulate							1			1
lh.entorhinal	-1	1		1	1	1	1		1	1
lh.fusiform		1		1			1		1	1
lh.inferiortemporal	-1	1		1	1	1	1		1	1
lh.lateralorbitofrontal		1			1	1				
lh.medialorbitofrontal		1		1		1	1			
lh.middletemporal	-1	1		1	1	1	1		1	1
lh.parahippocampal		1		1			1	-1		1
lh.paracentral		1			1			-1		
lh.parsopercularis		1			1					
lh.postcentral		1	-1			-1		-1		
lh.posteriorcingulate		1				-1		-1		
lh.precentral		1				-1				
lh.precuneus		1				-1				
lh.rostralanteriorcingulate		1				-1				
lh.rostralmiddlefrontal		1				-1				
lh.superiorfrontal		1		1			1			1
lh.superiorparietal		1		1	1	1	1			1
lh.superiortemporal	-1	1		1	1	1	1			1
lh.supramarginal		1								
lh.transversetemporal		1		1		-1				
lh.amygdala		1		1			1		1	1
lh.caudate	-1	1		1	1	1	1		1	
lh.thalamus	-1	1	1	1	1	1	1		1	1
lh.hippocampus		1	-1			-1		-1		
rh.entorhinal	-1	1		1	1	1	1			
rh.fusiform		1	1			1		1		
rh.inferiorparietal		1				-1	-1	-1	-1	
rh.isthmuscingulate						-1				
rh.lateralorbitofrontal		1				-1				
rh.lingual		1								
rh.medialorbitofrontal	-1	1			1	1	1			
rh.middletemporal	-1	1			1	1	1			
rh.parahippocampal		1			1					
rh.paracentral			-1			-1	-1	-1		1
rh.parsopercularis			-1			-1	-1	-1		1
rh.parsorbitalis	1	1				-1		-1		1
rh.parstriangularis		1	-1			-1		-1		1
rh.pericalcarine			-1			-1		-1		
rh.postcentral		1				-1		-1		
rh.posteriorcingulate		1				-1		-1		
rh.precentral		1				-1				
rh.precuneus		1				1				
rh.rostralmiddlefrontal		1								
rh.superiorfrontal		1						-1		
rh.superiorparietal		1				1				
rh.superiortemporal		1			1	1				
rh.supramarginal		1							-1	
rh.transversetemporal		1								
rh.amygdala		1				-1				
rh.caudate	-1	1		1	1	1	1			
rh.globuspallidus			-1			-1		-1		
rh.thalamus	-1	1	1	1	1	1	1			
rh.accumbens	-1			-1		1				

Supplementary Table 8. DKT ROI-based GMV features which were found to have statistically significant differences between MCI subtypes. Subtypes 1 and 2 are stable subtypes and subtypes 3, 4, and 5 are progressor subtypes. A "1" indicates a statistically significant difference where the subtype named with the smaller number has *greater GMV* than that named with the larger number. A "-1" indicates a statistically significant difference where the

subtype named with the smaller number has *smaller GMV* than that named with the larger number. lh=left hemisphere, rh=right hemisphere. ROI naming convention is as defined in Klein and Tourville (2012).

DKT ROI-based features	Subtypes contrasted									
CT	12	13	14	15	23	24	25	34	35	45
lh.entorhinal	-1	1	1	1	1	1	1		1	1
lh.fusiform		1	1	1	1	1	1		1	
lh.inferiorparietal		1	1							-1
lh.inferiortemporal	-1	1	1	1	1	1	1		1	1
lh.isthmuscingulate	1	1		1		-1				1
lh.lateraloccipital		1		1			1			1
lh.lateralorbitofrontal		1		1		1	1			1
lh.lingual		1		1		1	1			1
lh.medialorbitofrontal	-1	1	1	1	1	1	1		1	1
lh.middletemporal	-1	1	1	1	1	1	1		1	1
lh.parahippocampal	-1	1		1	1	1	1			
lh.paracentral		1								
lh.parsopercularis	1	1			-1	-1	-1			
lh.parsorbitalis	1	1				-1	-1			
lh.parstriangularis		1				-1				1
lh.pericalcarine		1				-1				1
lh.postcentral		1								
lh.posteriorcingulate		1	1			1				-1
lh.precentral		1	1			1	1			-1
lh.precuneus		1	1	1		1	1			
lh.rostralanteriorcingulate		1	1	1		1	1			
lh.rostralmiddlefrontal		1		1		1	1			
lh.superiorfrontal		1	1	1		1	1			
lh.superiorparietal		1	1	1		1	1			
lh.superiortemporal		1	1	1		1	1		1	
lh.supramarginal		1	1	1		1				
lh.transversetemporal		1	1	1		1	1			
lh.insula		1		1		1	1		1	1
lh.amygdala		1	1							
lh.caudate	-1	1			1	1	1			
lh.globuspallidus		1	1	1	1	1	1			
lh.thalamus	-1	1	1	1	1	1	1			
lh.hippocampus		1	1							-1
lh.accumbens		1	1	1	1	1	1	1		-1
rh.caudalanteriorcingulate		1	1	1		1	1			
rh.caudalmiddlefrontal		1	1	1		1	1		1	
rh.cuneus	-1	1	1	1	1	1	1		1	1
rh.entorhinal	-1	1	1	1	1	1	1			
rh.fusiform	-1	1	1	1	1	1	1			
rh.inferiorparietal		1			1	1	1			
rh.inferiortemporal		1				1	1			
rh.isthmuscingulate		1	1	1		1	1			
rh.lateraloccipital		1	1							
rh.lateralorbitofrontal		1								1
rh.lingual		1					1			1
rh.medialorbitofrontal		1				1	1			
rh.middletemporal		1	1			1	1			
rh.parahippocampal		1	1			1	1			
rh.paracentral		1	1	1		1	1			
rh.parsopercularis		1	1	1		1	1			
rh.parsorbitalis		1	1			1	1			
rh.parstriangularis		1	1	1	1	1	1			
rh.pericalcarine		1	1	1		1	1			
rh.postcentral		1	1	1	1	1	1			
rh.posteriorcingulate		1	1	1		1	1			
rh.precentral	-1	1	1		1	1	1			
rh.precuneus	-1	1	1		1	1	1			

Supplementary Table 9. DKT ROI-based CT features which were found to have statistically significant differences between MCI subtypes. Subtypes 1 and 2 are stable subtypes and subtypes 3, 4, and 5 are progressor subtypes. A "1" indicates a statistically significant difference where the subtype named with the smaller number has *greater CT* than that named with the larger number. A "-1" indicates a statistically significant difference where the subtype named with the smaller number has *smaller CT* than that named with the larger number. lh=left hemisphere, rh=right hemisphere. ROI naming convention is as defined in Klein and Tourville (2012).

DKT ROI-based features		Subtypes contrasted									
FC		12	13	14	15	23	24	25	34	35	45
lh.cuneus	lh.inferiorparietal	1	1	1							
lh.cuneus	lh.paracentral	1	1	1				-1	1	-1	-1
lh.cuneus	lh.postcentral	1	1	1		-1		-1	1	-1	-1
lh.cuneus	lh.precuneus	1	1	1				-1		-1	-1
lh.cuneus	lh.thalamus										
lh.cuneus	rh.precuneus	1	1	1				-1	1	-1	-1
lh.cuneus	rh.accumbens							-1			
lh.entorhinal	rh.parstriangularis							-1			
lh.entorhinal	rh.postcentral	1						-1			
lh.entorhinal	rh.posteriorcingulate	1						-1			
lh.entorhinal	rh.precentral	1						-1			
lh.entorhinal	rh.precuneus							-1			-1
lh.fusiform	lh.postcentral	1	1	1				-1	1	-1	-1
lh.fusiform	rh.paracentral	1	1	1				-1		-1	-1
lh.fusiform	rh.superiortemporal	1		1				-1	1	-1	-1
lh.inferiorparietal	lh.isthmuscingulate	1	1	1						-1	
lh.inferiorparietal	lh.lingual	1	1	1				-1		-1	-1
lh.inferiorparietal	lh.posteriorcingulate	1	1	1						-1	
lh.inferiorparietal	rh.cuneus	1	1	1							-1
lh.inferiorparietal	rh.inferiorparietal	1	1	1						-1	-1
lh.inferiorparietal	rh.isthmuscingulate	1	1	1						-1	
lh.inferiorparietal	rh.lateraloccipital	1	1	1				-1		-1	-1
lh.inferiorparietal	rh.lingual	1	1	1				-1		-1	-1
lh.inferiorparietal	rh.parahippocampal	1	1	1				-1		-1	-1
lh.inferiorparietal	rh.parstriangularis	1	1	1			1	-1		-1	-1
lh.inferiorparietal	rh.pericalcarine	1	1	1				-1		-1	-1
lh.inferiorparietal	rh.postcentral	1	1	1			1	-1	1	-1	-1
lh.inferiorparietal	rh.posteriorcingulate	1	1	1				-1		-1	-1
lh.inferiorparietal	rh.precentral	1	1	1				-1		-1	-1
lh.inferiorparietal	rh.precuneus	1	1	1				-1		-1	-1
lh.inferiorparietal	rh.rostralmiddlefrontal	1	1	1			1	-1		-1	-1
lh.inferiorparietal	rh.superiorfrontal	1	1	1				-1		-1	-1
lh.inferiorparietal	rh.superiorparietal	1	1	1				-1		-1	-1
lh.isthmuscingulate	lh.precentral	1	1	1				-1		-1	-1
lh.lateraloccipital	lh.precuneus	1	1	1				-1	1	-1	-1
lh.lateraloccipital	rh.accumbens										
lh.lingual	lh.paracentral	1	1	1				-1		-1	-1
lh.lingual	lh.accumbens										
lh.middletemporal	rh.superiortemporal	1		1				-1	1		-1
lh.parahippocampal	lh.parstriangularis	1	1	1		-1		-1	1	-1	-1
lh.parahippocampal	lh.pericalcarine	1	1	1		-1		-1	1	-1	-1
lh.parahippocampal	lh.postcentral	1	1	1		-1		-1	1	-1	-1
lh.parahippocampal	rh.parstriangularis	1	1	1		-1		-1	1	-1	-1
lh.parahippocampal	rh.pericalcarine	1	1	1		-1		-1	1	-1	-1
lh.parahippocampal	rh.postcentral	1	1	1		-1		-1	1	-1	-1
lh.parahippocampal	rh.posteriorcingulate	1	1	1		-1		-1	1	-1	-1
lh.parahippocampal	rh.precentral	1	1	1		-1		-1	1	-1	-1
lh.parahippocampal	rh.superiorparietal	1	1	1		-1		-1	1	-1	-1
lh.parahippocampal	rh.superiortemporal	1	1	1		-1		-1	1	-1	-1

Supplementary Table 10. DKT ROI-based FC features which were found to have statistically significant differences between MCI subtypes. Subtypes 1 and 2 are stable subtypes and subtypes 3, 4, and 5 are progressor subtypes. A "1" indicates a statistically significant difference where the subtype named with the smaller number has *greater FC*

than that named with the larger number. A "-1" indicates a statistically significant difference where the subtype named with the smaller number has *smaller FC* than that named with the larger number. lh=left hemisphere, rh=right hemisphere. ROI naming convention is as defined in Klein and Tourville (2012).

DKT ROI-based features		Subtypes contrasted									
FC		12	13	14	15	23	24	25	34	35	45
lh.paracentral	lh.parstriangularis	1	1	1				-1	1	-1	-1
lh.paracentral	lh.pericalcarine	1	1	1				-1	1	-1	-1
lh.paracentral	lh.posteriorcingulate	1	1	1		-1		-1	1	-1	-1
lh.paracentral	lh.precentral	1	1	1		-1		-1	1	-1	-1
lh.paracentral	lh.thalamus										
lh.paracentral	rh.cuneus	1	1	1				-1	1	-1	-1
lh.paracentral	rh.lingual	1	1	1				-1	1	-1	-1
lh.paracentral	rh.parstriangularis	1	1	1		-1	1	-1	1	-1	-1
lh.paracentral	rh.pericalcarine	1	1	1		-1	1	-1	1	-1	-1
lh.paracentral	rh.posteriorcingulate	1	1	1		-1	1	-1	1	-1	-1
lh.paracentral	rh.precentral	1	1	1		-1	1	-1	1	-1	-1
lh.paracentral	rh.superiorparietal	1	1	1		-1	1	-1	1	-1	-1
lh.paracentral	rh.thalamus										
lh.parstriangularis	lh.postcentral	1	1	1				-1	1	-1	-1
lh.parstriangularis	lh.caudate	-1	-1								
lh.parstriangularis	lh.thalamus										
lh.parstriangularis	lh.accumbens										
lh.parstriangularis	rh.parahippocampal	1	1	1		-1	1	-1	1	-1	-1
lh.parstriangularis	rh.parstriangularis	1	1	1		-1	1	-1	1	-1	-1
lh.parstriangularis	rh.pericalcarine	1	1	1		-1	1	-1	1	-1	-1
lh.parstriangularis	rh.postcentral	1	1	1		-1	1	-1	1	-1	-1
lh.parstriangularis	rh.superiorfrontal	1	1	1		-1		-1	1	-1	-1
lh.parstriangularis	rh.superiorparietal	1	1	1		-1	1	-1	1	-1	-1
lh.parstriangularis	rh.superiortemporal	1	1	1		-1	1	-1	1	-1	-1
lh.parstriangularis	rh.supramarginal	1	1	1			1	-1	1	-1	-1
lh.parstriangularis	rh.thalamus										
lh.parstriangularis	rh.hippocampus	1		1				-1			-1
lh.pericalcarine	lh.postcentral	1	1	1		-1		-1	1	-1	-1
lh.pericalcarine	lh.accumbens										
lh.pericalcarine	rh.parahippocampal	1	1	1			1	-1	1	-1	-1
lh.pericalcarine	rh.paracentral	1	1	1		-1	1	-1	1	-1	-1
lh.pericalcarine	rh.parstriangularis	1	1	1		-1	1	-1	1	-1	-1
lh.pericalcarine	rh.pericalcarine	1	1	1		-1	1	-1	1	-1	-1
lh.pericalcarine	rh.postcentral	1	1	1		-1	1	-1	1	-1	-1
lh.pericalcarine	rh.superiortemporal	1	1	1		-1	1	-1	1	-1	-1
lh.pericalcarine	rh.supramarginal	1	1	1			1	-1	1	-1	-1
lh.pericalcarine	rh.thalamus										
lh.pericalcarine	rh.hippocampus	1		1				-1			-1
lh.postcentral	lh.superiorparietal	1	1	1		-1		-1	1	-1	-1
lh.postcentral	lh.accumbens										
lh.postcentral	rh.cuneus	1	1	1		-1		-1	1	-1	-1
lh.postcentral	rh.fusiform	1	1	1				-1	1	-1	-1
lh.postcentral	rh.lingual	1	1	1		-1		-1	1	-1	-1
lh.postcentral	rh.parahippocampal	1	1	1		-1		-1	1	-1	-1
lh.postcentral	rh.paracentral	1	1	1		-1	1	-1	1	-1	-1
lh.postcentral	rh.parstriangularis	1	1	1		-1	1	-1	1	-1	-1
lh.postcentral	rh.pericalcarine	1	1	1		-1	1	-1	1	-1	-1
lh.postcentral	rh.postcentral	1	1	1		-1	1	-1	1	-1	-1
lh.postcentral	rh.posteriorcingulate	1	1	1		-1	1	-1	1	-1	-1

Supplementary Table 10 (continued).

DKT ROI-based features		Subtypes contrasted									
FC		12	13	14	15	23	24	25	34	35	45
lh.postcentral	rh.precentral	1	1	1		-1	1	-1	1	-1	-1
lh.postcentral	rh.precuneus	1	1	1		-1	1	-1	1	-1	-1
lh.postcentral	rh.superiorfrontal	1	1	1		-1		-1	1	-1	-1
lh.postcentral	rh.superiorparietal	1	1	1		-1	1	-1	1	-1	-1
lh.postcentral	rh.superiortemporal	1	1	1		-1	1	-1	1	-1	-1
lh.postcentral	rh.hippocampus	1		1				-1			-1
lh.posteriorcingulate	lh.precuneus	1	1	1				-1		-1	-1
lh.posteriorcingulate	lh.thalamus										
lh.posteriorcingulate	lh.accumbens										
lh.posteriorcingulate	rh.postcentral	1	1	1		-1	1	-1	1	-1	-1
lh.posteriorcingulate	rh.posteriorcingulate	1	1	1		-1	1	-1	1	-1	-1
lh.posteriorcingulate	rh.precentral	1	1	1			1	-1	1	-1	-1
lh.posteriorcingulate	rh.precuneus	1	1	1			1	-1	1	-1	-1
lh.posteriorcingulate	rh.rostralmiddlefrontal	1	1	1			1	-1	1	-1	-1
lh.posteriorcingulate	rh.superiortemporal	1	1	1		-1	1	-1	1	-1	-1
lh.posteriorcingulate	rh.supramarginal	1	1	1		-1	1	-1	1	-1	-1
lh.posteriorcingulate	rh.thalamus										
lh.precentral	lh.precuneus	1	1	1				-1		-1	-1
lh.precentral	lh.rostralmiddlefrontal	1	1	1				-1	1	-1	-1
lh.precentral	lh.superiortemporal	1	1	1				-1	1	-1	-1
lh.precentral	lh.accumbens										
lh.precentral	rh.parstriangularis	1	1	1		-1	1	-1	1	-1	-1
lh.precentral	rh.pericalcarine	1	1	1		-1	1	-1	1	-1	-1
lh.precentral	rh.postcentral	1	1	1		-1	1	-1	1	-1	-1
lh.precentral	rh.posteriorcingulate	1	1	1			1	-1	1	-1	-1
lh.precentral	rh.precuneus	1	1	1			1	-1	1	-1	-1
lh.precentral	rh.rostralmiddlefrontal	1	1	1			1	-1	1	-1	-1
lh.precentral	rh.superiortemporal	1	1	1		-1	1	-1	1	-1	-1
lh.precentral	rh.supramarginal	1	1	1		-1	1	-1	1	-1	-1
lh.precentral	rh.thalamus										
lh.precuneus	rh.cuneus	1	1	1			1	-1		-1	-1
lh.precuneus	rh.lingual	1	1	1				-1		-1	-1
lh.precuneus	rh.parstriangularis	1	1	1			1	-1	1	-1	-1
lh.precuneus	rh.pericalcarine	1	1	1			1	-1	1	-1	-1
lh.precuneus	rh.posteriorcingulate	1	1	1			1	-1	1	-1	-1
lh.precuneus	rh.precentral	1	1	1			1	-1	1	-1	-1
lh.precuneus	rh.precuneus	1	1	1			1	-1	1	-1	-1
lh.precuneus	rh.rostralmiddlefrontal	1	1	1			1	-1	1	-1	-1
lh.precuneus	rh.superiorfrontal	1	1	1			1	-1	1	-1	-1
lh.precuneus	rh.superiorparietal	1	1	1			1	-1	1	-1	-1
lh.rostralanteriorcingulate	rh.parahippocampal	1	1	1			1	-1	1	-1	-1
lh.rostralanteriorcingulate	rh.rostralmiddlefrontal	1	1	1			1	-1	1	-1	-1
lh.rostralmiddlefrontal	lh.superiorfrontal	1	1	1				-1		-1	-1
lh.rostralmiddlefrontal	rh.parahippocampal	1	1	1			1	-1	1	-1	-1
lh.rostralmiddlefrontal	rh.precentral	1	1	1			1	-1	1	-1	-1
lh.rostralmiddlefrontal	rh.rostralmiddlefrontal	1	1	1			1	-1	1	-1	-1
lh.rostralmiddlefrontal	rh.superiorfrontal	1	1	1			1	-1	1	-1	-1
lh.rostralmiddlefrontal	rh.superiorparietal	1	1	1			1	-1	1	-1	-1
lh.superiorfrontal	rh.rostralmiddlefrontal	1	1	1			1	-1	1	-1	-1

Supplementary Table 10 (continued).

DKT ROI-based features		Subtypes contrasted									
FC		12	13	14	15	23	24	25	34	35	45
lh.superiorparietal	rh.inferiorparietal	1	1	1				-1		-1	-1
lh.transversetemporal	lh.insula										
lh.caudate	rh.parstriangularis		-1								
lh.caudate	rh.pericalcarine		-1								
lh.thalamus	rh.parstriangularis										
lh.thalamus	rh.pericalcarine										
lh.thalamus	rh.postcentral										
lh.thalamus	rh.posteriorcingulate										
lh.thalamus	rh.precentral										
lh.accumbens	rh.parstriangularis										-1
lh.accumbens	rh.pericalcarine								1		-1
lh.accumbens	rh.postcentral										
lh.accumbens	rh.posteriorcingulate										
lh.accumbens	rh.precentral										
lh.accumbens	rh.rostralmiddlefrontal										
rh.caudalmiddlefrontal	rh.fusiform	1		1				-1		-1	-1
rh.fusiform	rh.parsorbitalis	1	1	1							-1
rh.fusiform	rh.postcentral	1	1	1				-1	1	-1	-1
rh.lingual	rh.accumbens										
rh.parahippocampal	rh.parstriangularis	1	1	1		-1	1	-1	1	-1	-1
rh.parahippocampal	rh.pericalcarine	1	1	1		-1		-1	1	-1	-1
rh.parahippocampal	rh.postcentral	1	1	1		-1		-1	1	-1	-1
rh.parahippocampal	rh.posteriorcingulate	1	1	1		-1		-1	1	-1	-1
rh.parahippocampal	rh.precentral	1	1	1		-1	1	-1	1	-1	-1
rh.paracentral	rh.posteriorcingulate	1	1	1		-1		-1	1	-1	-1
rh.paracentral	rh.thalamus										
rh.paracentral	rh.accumbens										
rh.parstriangularis	rh.caudate	-1	-1							1	
rh.parstriangularis	rh.thalamus										
rh.parstriangularis	rh.accumbens										
rh.pericalcarine	rh.superiorfrontal	1	1	1		-1		-1	1	-1	-1
rh.pericalcarine	rh.caudate	-1	-1								
rh.pericalcarine	rh.thalamus										
rh.pericalcarine	rh.accumbens										-1
rh.postcentral	rh.superiorfrontal	1	1	1		-1		-1	1	-1	-1
rh.postcentral	rh.superiorparietal	1	1	1		-1		-1	1	-1	-1
rh.postcentral	rh.thalamus										
rh.postcentral	rh.accumbens										
rh.posteriorcingulate	rh.superiortemporal	1	1	1		-1		-1	1	-1	-1
rh.posteriorcingulate	rh.caudate	-1	-1							1	
rh.posteriorcingulate	rh.thalamus										
rh.posteriorcingulate	rh.accumbens										
rh.precentral	rh.superiortemporal	1	1	1		-1		-1	1	-1	-1
rh.precentral	rh.caudate	-1	-1					1		1	
rh.precentral	rh.thalamus										
rh.precentral	rh.accumbens										
rh.precuneus	rh.accumbens										
rh.rostralmiddlefrontal	rh.superiorfrontal	1	1	1				-1		-1	-1
rh.rostralmiddlefrontal	rh.accumbens										

Supplementary Table 10 (continued).

Appendix H

Manuscript Specific Acknowledgements

We thank Lindsay Lewis and Janine Mendola for sharing part of the data we used for manuscripts 1 and 2. We thank Sepide Movaghati for her comments on an earlier version of manuscript 1, and Ilana Leppert and Bruce Pike for support with MRI physics. We thank Robert Brown and Jacob Vogel for their advice with respect to machine learning methods for Manuscript 3. We also thank Samuel Gregoire for his early contributions to quality checks in Manuscript 3.

This work was supported by NIH grant R01 EY015219 awarded to Janine Mendola.

The work is also supported by grants from the Human Frontier Science Program (RGY0080/2008) and the Natural Sciences and Engineering Research Council of Canada (RGPIN 375457-09 and RGPIN 2015-05103) awarded to Amir Shmuel.

Debra Dawson was supported by a Ph.D. student fellowship from the Fonds de Recherche du Québec – Santé as well as a Canadian Institutes of Health Research MD-PhD fellowship.

Manuscript 3 was in part supported by The National Key Research and Development Program of China (2016YFC1306300), National Natural Science Foundation of China (Grant 61633018), Beijing Municipal Commission of Health and Family Planning (PXM2019_026283_000002) all awarded to Dr Ying Han.

Part of the data utilized in Manuscript 3 comes from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The data collection and sharing for the project was funded by the ADNI (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug

Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Data for Manuscript 3 were provided in part by OASIS-3: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P50AG00561, P30NS09857781, P01AG026276, P01AG003991, R01AG043434, UL1TR000448, R01EB009352. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly