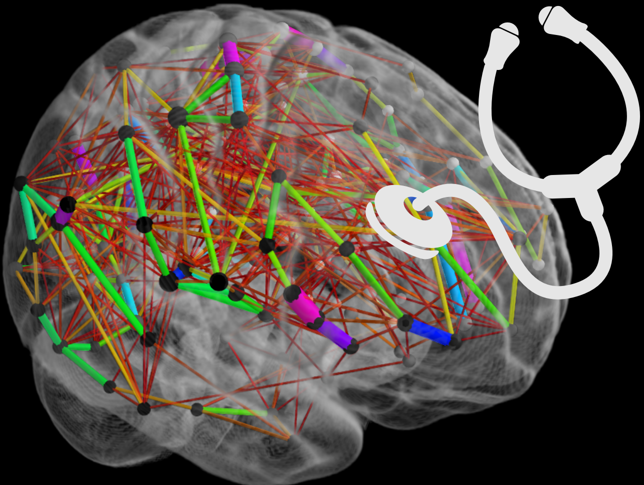
Extracting neuro-phenotypes from the brain at rest

Gaël Varoquaux

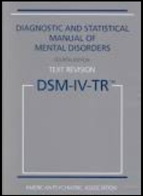
PARIETAL Inria NeuroSpin



Probing variations of the mind

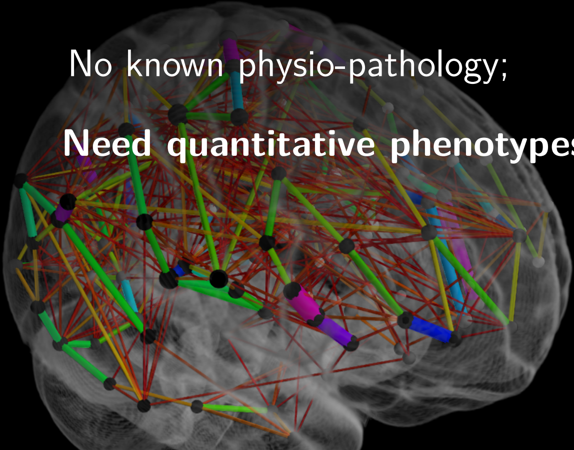
Psychiatry is defined by symptoms

Diagnostic and Statistical
Manual of Mental Disorders



No known physio-pathology; Autism $\stackrel{?}{=}$ Asperger

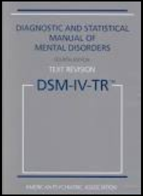
Need quantitative phenotypes of brain function



Probing variations of the mind

Psychiatry is defined by symptoms

Diagnostic and Statistical Manual of Mental Disorders



No known physio-pathology; Autism $\stackrel{?}{=}$ Asperger

Need quantitative phenotypes of brain function

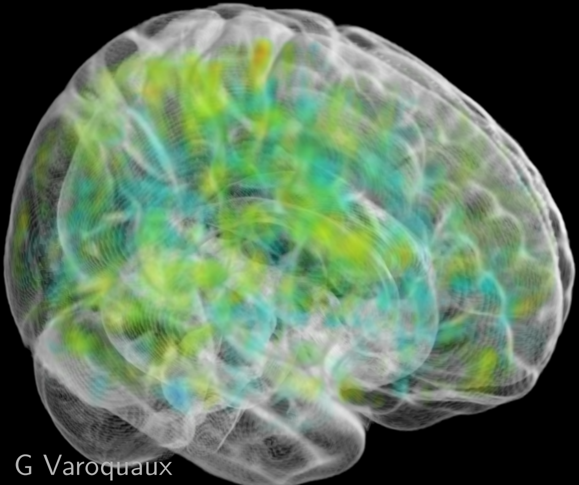
Population imaging with rest fMRI

UK Biobank [Miller... 2016]

- Easy to set up reproducibly
- Suitable for diminished patients
- Connectivity captures traits

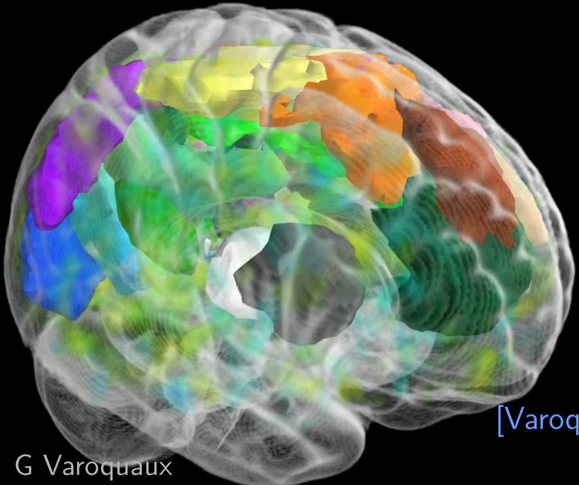
Functional connectomes

No salient features in rest fMRI



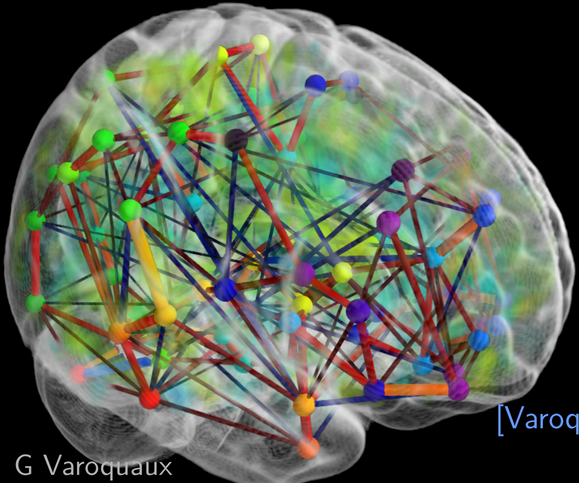
Functional connectomes

- Define functional regions



[Varoquaux and Craddock 2013]

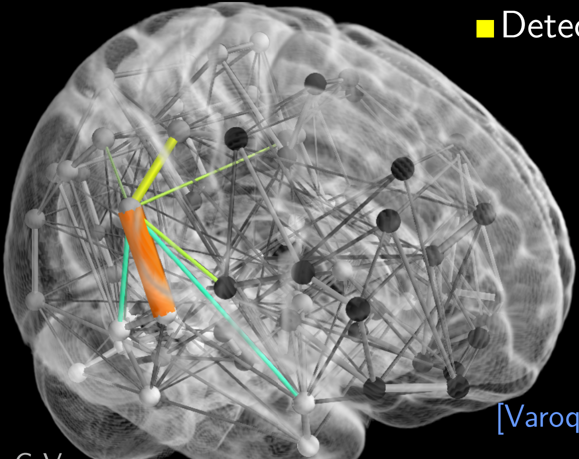
Functional connectomes



- Define functional regions
- Learn interactions

[Varoquaux and Craddock 2013]

Functional connectomes

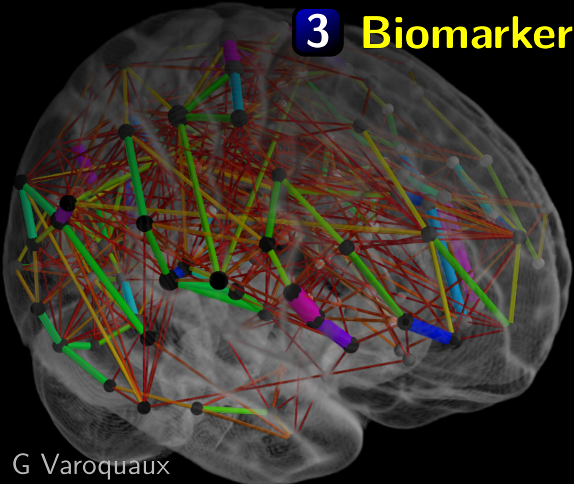


- Define functional regions
- Learn interactions
- Detect differences

[Varoquaux and Craddock 2013]

Outline

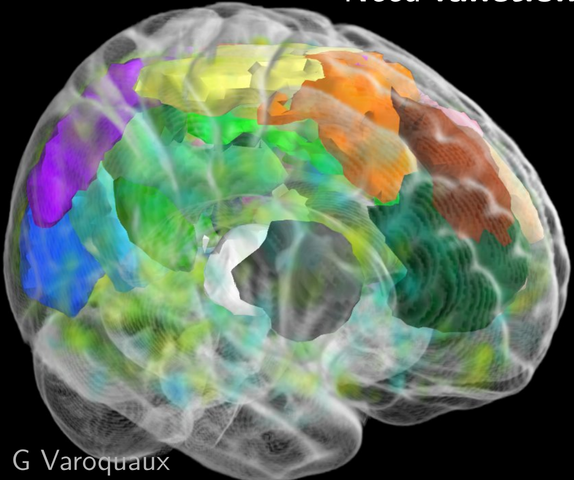
1. Functional regions
2. The connectome matrix
3. Biomarkers of autism



1 Functional regions

Need **functional** regions for nodes

⇒ Spatial analysis



1 Functional regions

Available “on the market”

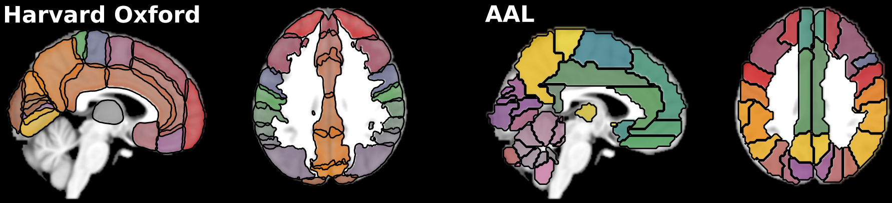
anatomical atlases, functional atlases, region extraction methods

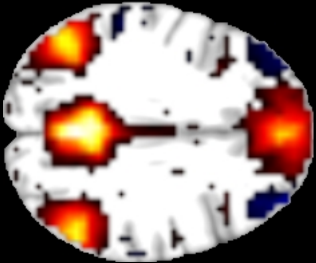
1 Functional regions

- Atlases based on anatomy
- Clustering tools
- Linear decomposition

1 Anatomical

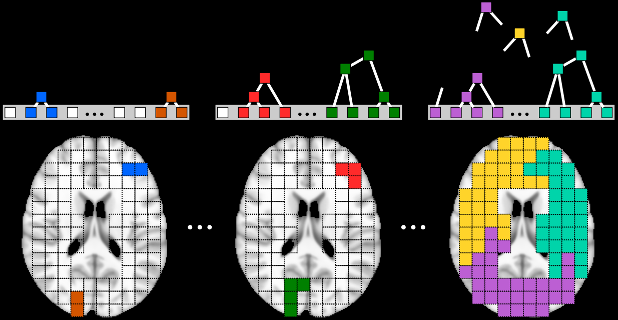
- Anatomical atlases do not resolve functional structures





1 Clustering approaches

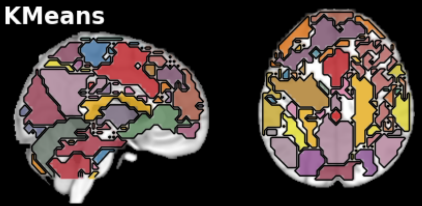
- Group together voxels with similar time courses



1 Clustering approaches

K-Means

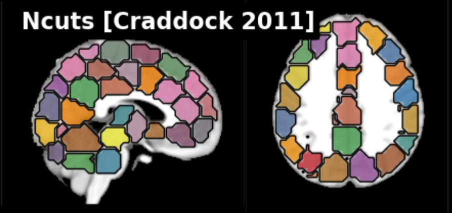
- Fast
- No spatial constraint
  (smooth the data)
- Related to [Yeo... 2011]



Normalized cuts

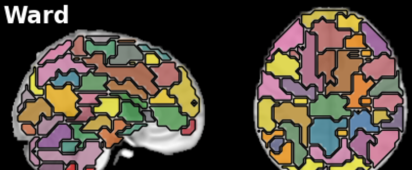
[Craddock... 2012]

- Slow
- Spatial constraints
- Very geometrical



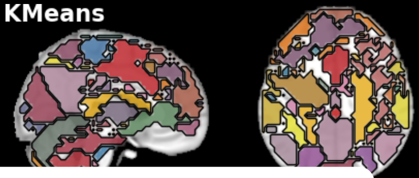
Ward clustering

- Very fast
  (even with many clusters)
- Spatial constraints



K-Means

- Fast
- No spatial constraint
  (smooth the data)



Empirical choice

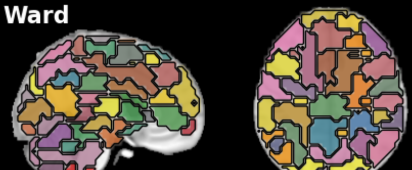
Based on cluster stability and fit to data

- Large number of clusters: Ward
- Small number of clusters: Kmeans

[Thirion... 2014]

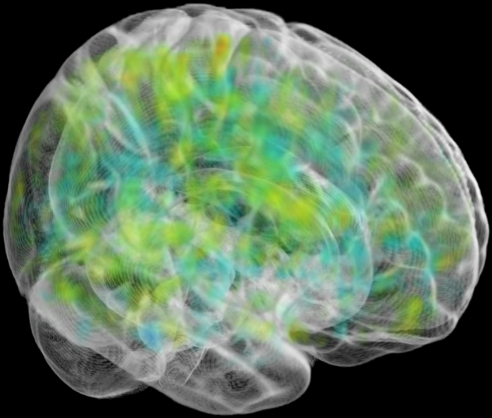
Ward clustering

- Very fast
  (even with many clusters)
- Spatial constraints



Working hypothesis / model:
Observing linear mixtures of networks at rest

Time courses

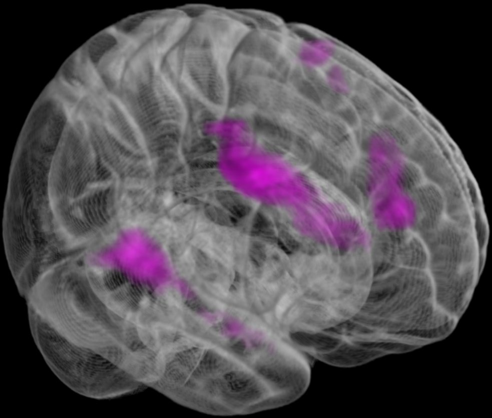


Working hypothesis / model:

Observing linear mixtures of networks at rest

Time courses

Language

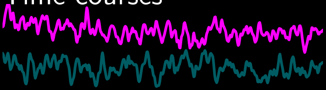


1 Mixture models: linear decompositions

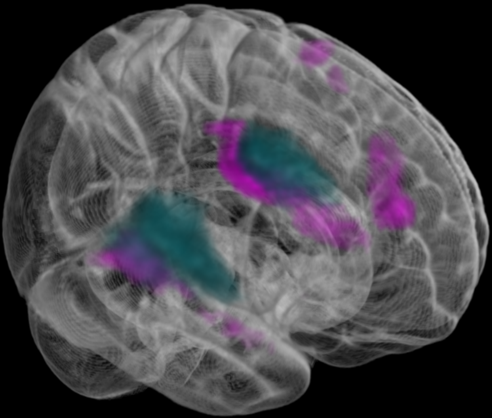
Working hypothesis / model:

Observing linear mixtures of networks at rest

Time courses

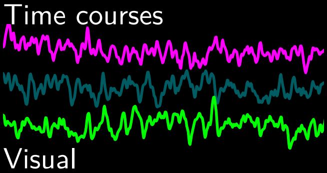


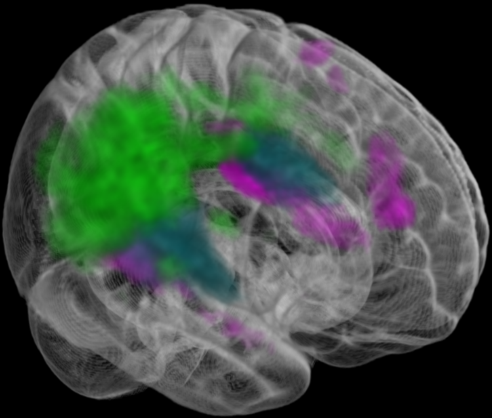
Audio



Working hypothesis / model:

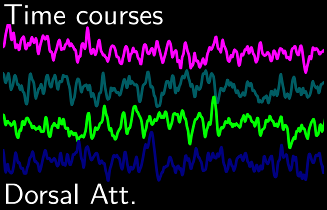
Observing linear mixtures of networks at rest

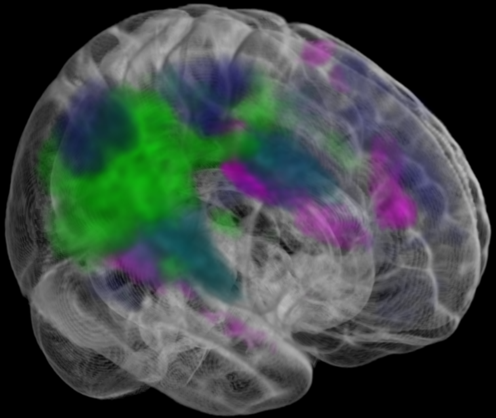




Working hypothesis / model:

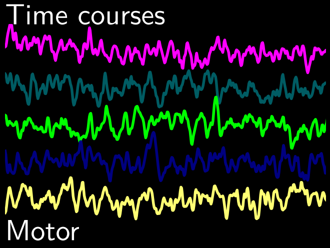
Observing linear mixtures of networks at rest

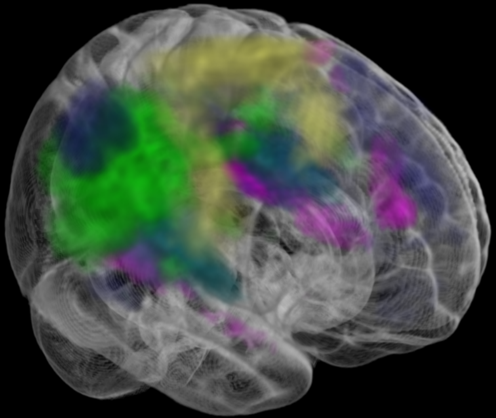




Working hypothesis / model:

Observing linear mixtures of networks at rest



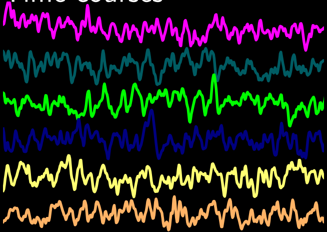


1 Mixture models: linear decompositions

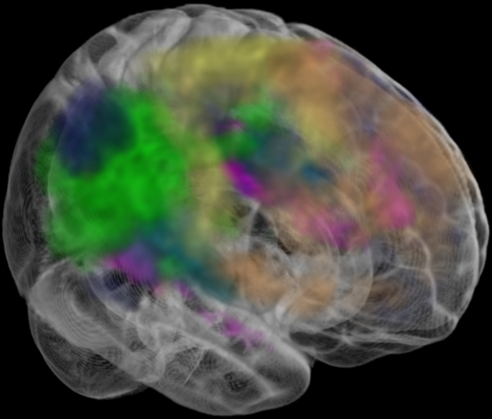
Working hypothesis / model:

Observing linear mixtures of networks at rest

Time courses



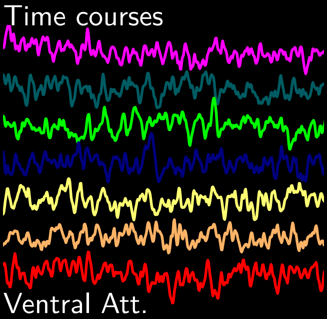
Salience

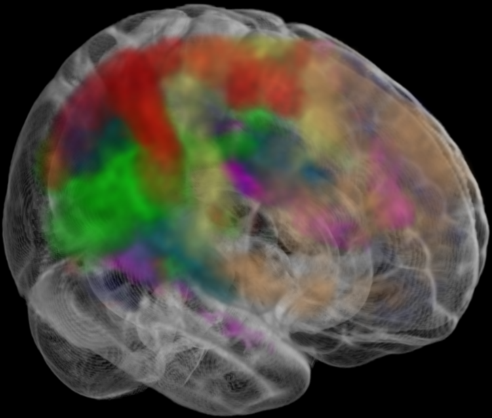


1 Mixture models: linear decompositions

Working hypothesis / model:

Observing linear mixtures of networks at rest

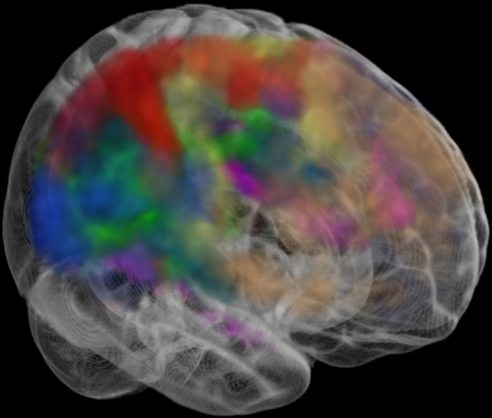




1 Mixture models: linear decompositions

Working hypothesis / model:

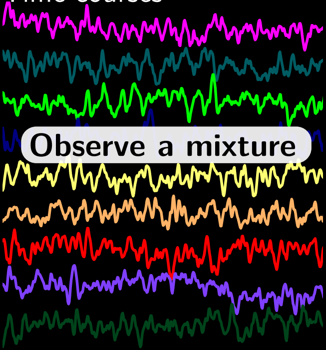
Observing linear mixtures of networks at rest

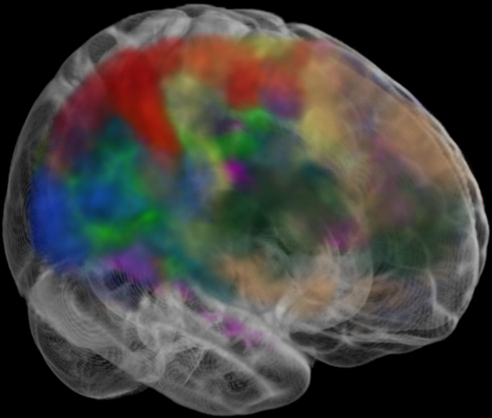


Working hypothesis / model:

Observing linear mixtures of networks at rest

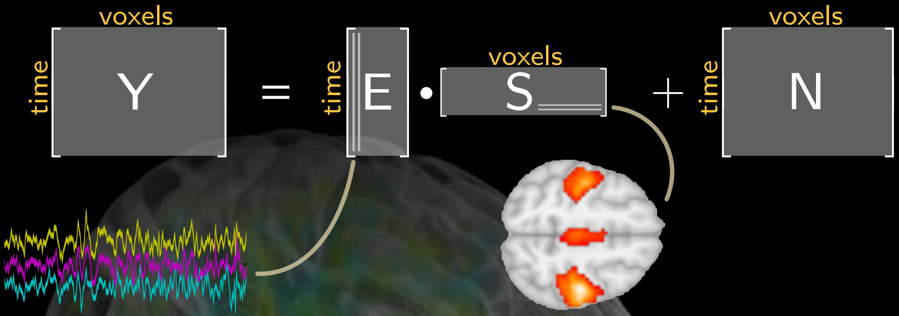
Time courses





How to unmix networks?

1 Spatial modes: ICA decomposition

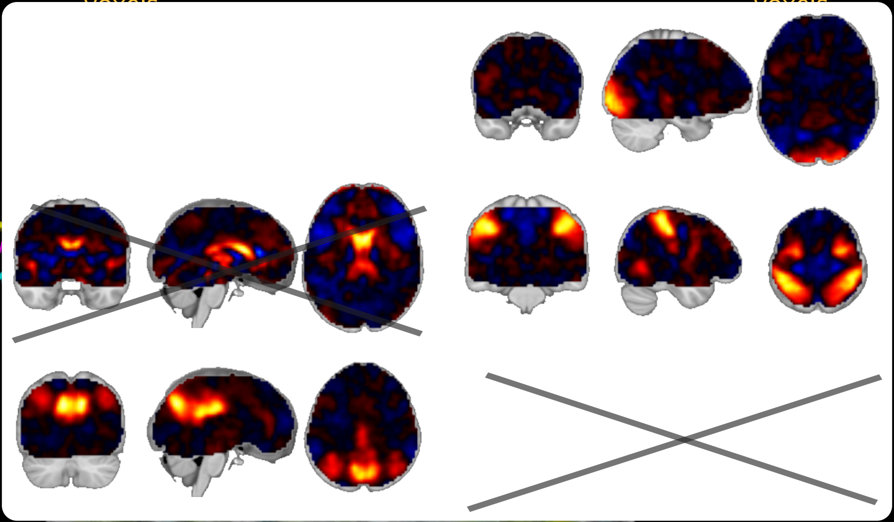


Decomposing time series into:

- covarying spatial maps, **S**
- uncorrelated residuals, **N**

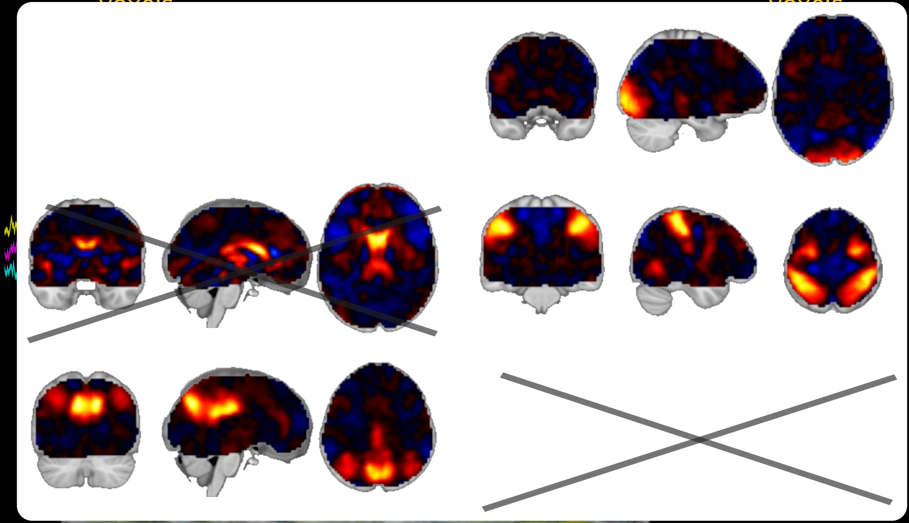
ICA: minimize mutual information across S

1 Spatial modes: ICA decomposition



ICA: minimize mutual information across S

1 Spatial modes: ICA decomposition

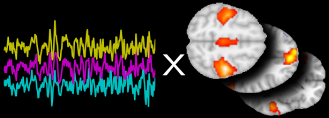


Sparse decompositions: sparse penalty on maps

ICA

1. Select signal of interest
2. Select “maximaly independent” ICs

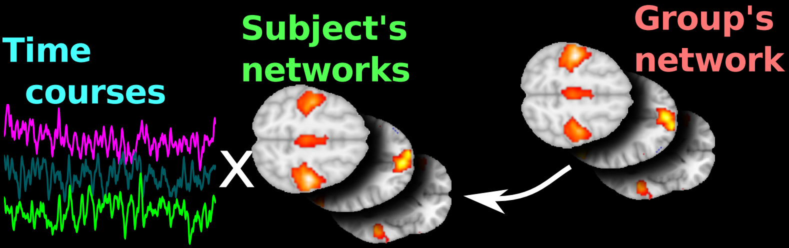
Sparse decomposition



$$\hat{\mathbf{E}}, \hat{\mathbf{S}} = \underset{\mathbf{S},\,\mathbf{E}}{\operatorname{argmin}} \left\| \mathbf{Y} - \mathbf{E}\,\mathbf{S} \right\|_2^2 + \lambda \left\| \mathbf{S} \right\|_1$$

Data fit — Penalization: sparse maps

Joint estimation of signal space + components



Multi-Subject Dictionary Learning

$$\underset{\mathbf{E}^s,\mathbf{S}^s,\mathbf{S}}{\operatorname{argmin}} \sum_{subjects} \left(\|\mathbf{Y}^s - \mathbf{E}^s {\mathbf{S}^s}^T\|^2_{\text{Fro}} + \mu \|\mathbf{S}^s - \mathbf{S}\|^2_{\text{Fro}} \right) + \lambda\, \Omega(\mathbf{S})$$

Data fit

Subject variability

Penalization: inject structure

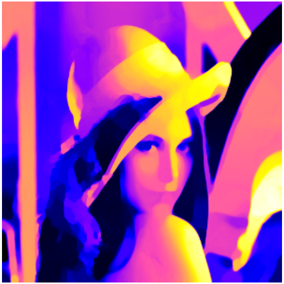
[Varoquaux... 2011, Abraham... 2013]

Create a region-forming penalty:

Original

Clustering

Total-variation



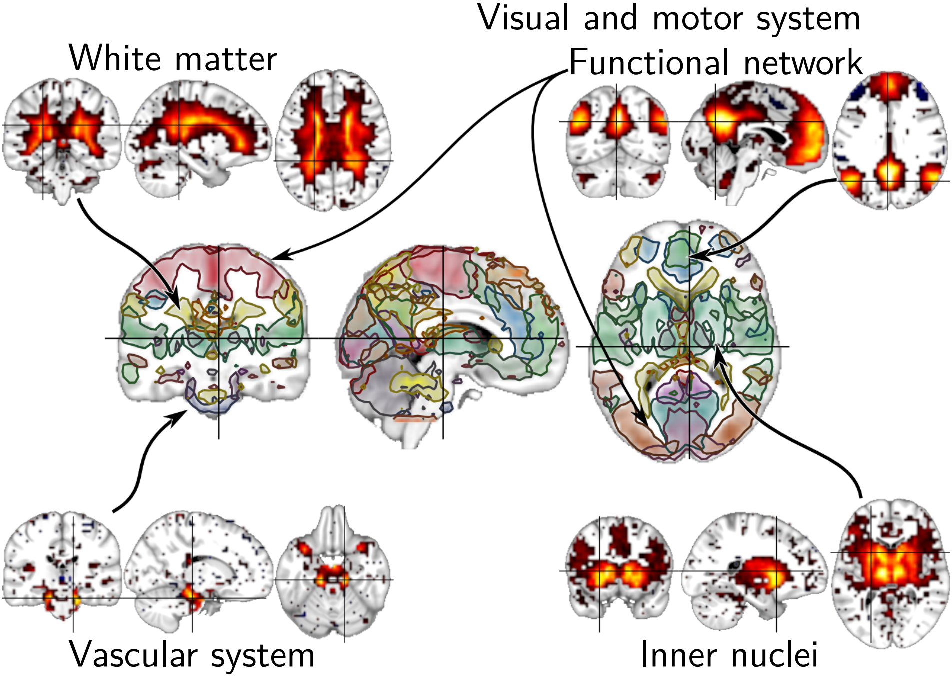
$$\underset{\mathbf{E}^s,\mathbf{S}^s,\mathbf{S}}{\operatorname{argmin}} \sum_{subjects} \left(\|\mathbf{Y}^s - \mathbf{E}^s {\mathbf{S}^s}^T\|_{\text{Fro}}^2 + \mu \|\mathbf{S}^s - \mathbf{S}\|_{\text{Fro}}^2 \right) + \lambda\, \Omega(\mathbf{S})$$

Data fit

Subject variability

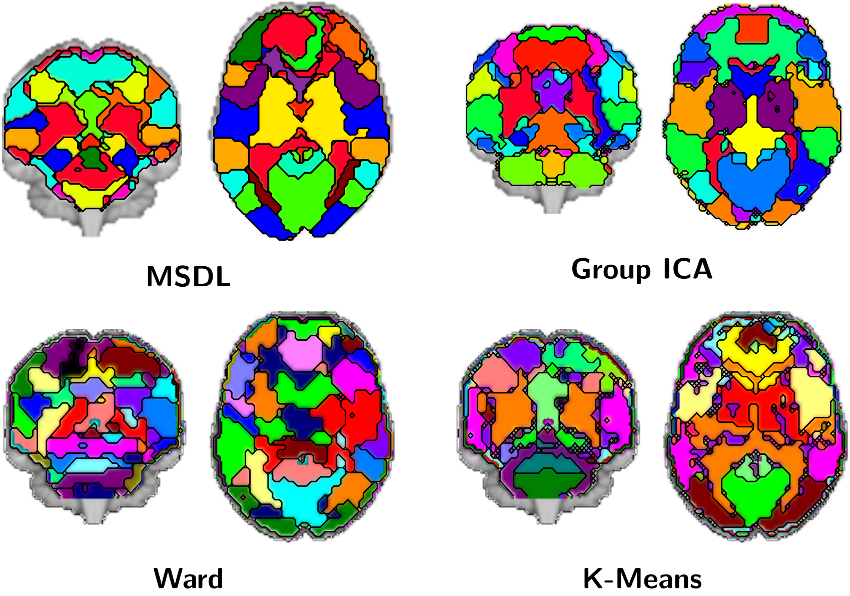
Penalization: inject structure

[Varoquaux... 2011, Abraham... 2013]



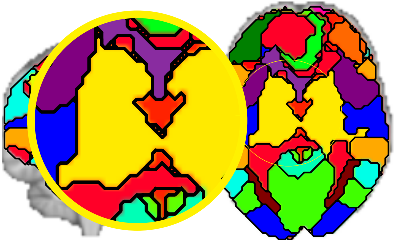
Downloadable from Parietal webpage `http://team.inria.fr/parietal`

Brain parcellations

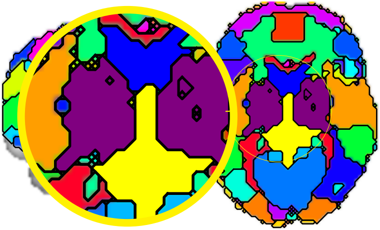


[Abraham... 2013]

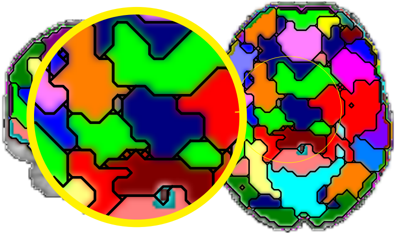
Brain parcellations



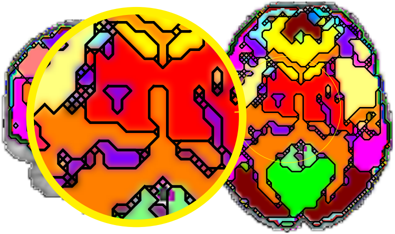
MSDL



Group ICA



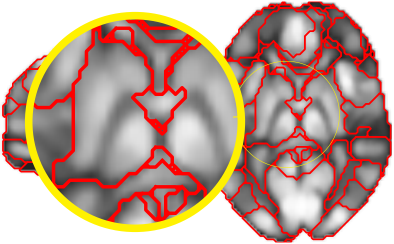
Ward



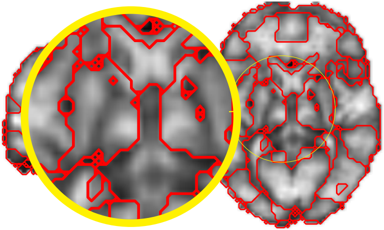
K-Means

[Abraham... 2013]

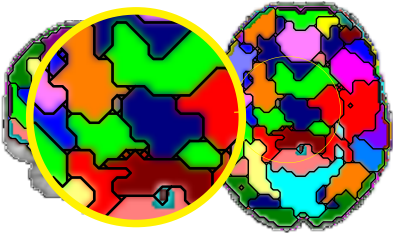
Brain parcellations



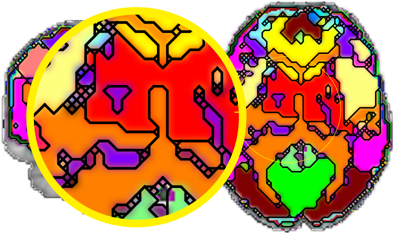
MSDL



Group ICA



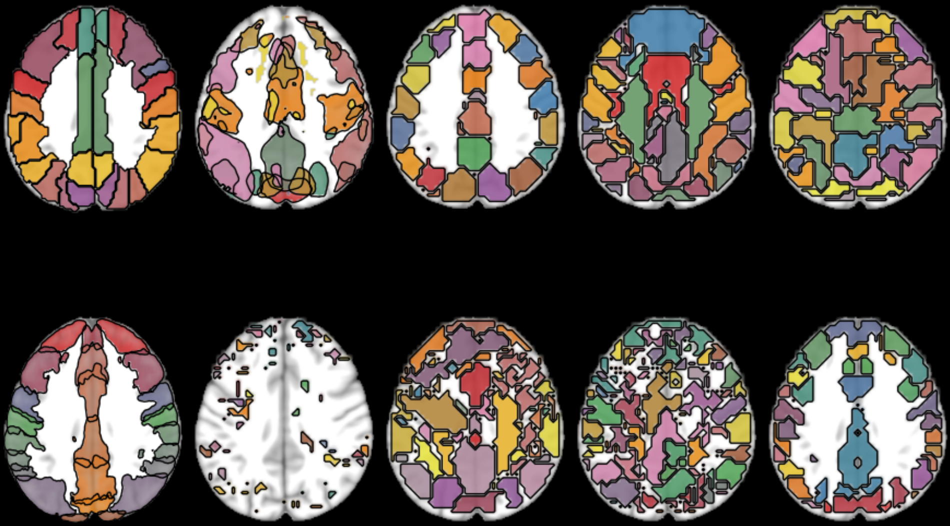
Ward



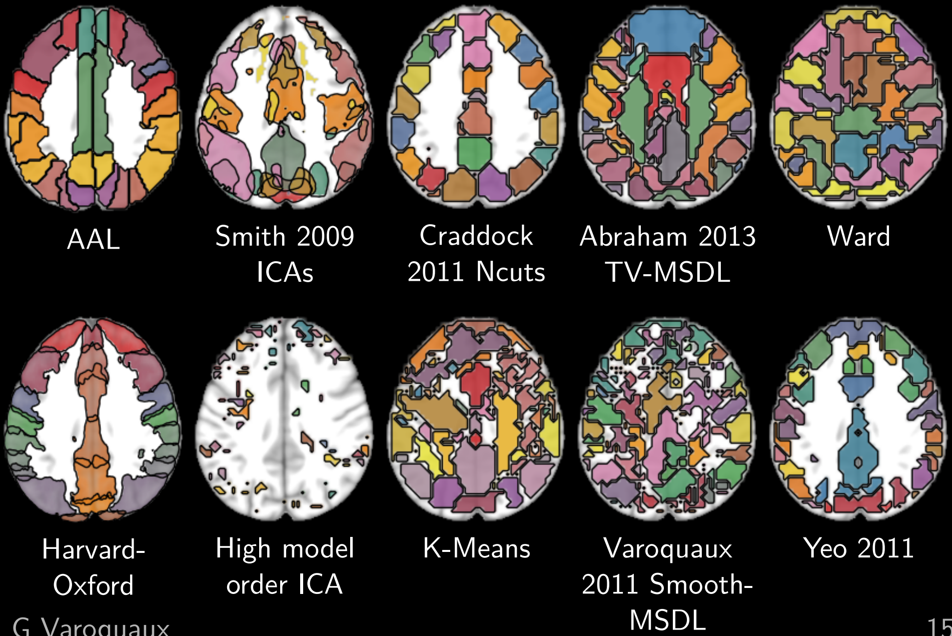
K-Means

[Abraham... 2013]

Functional regions

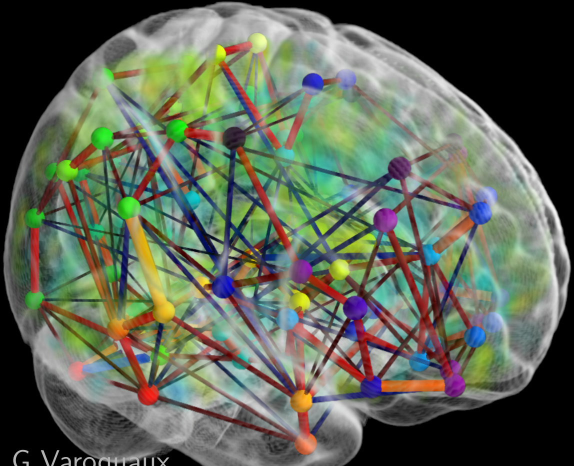


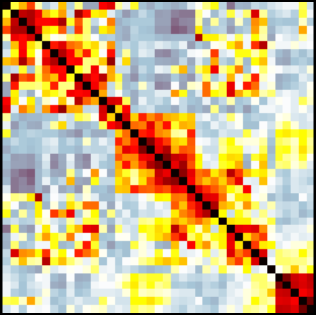
Functional regions



2 The connectome matrix

How to capture and represent interactions?

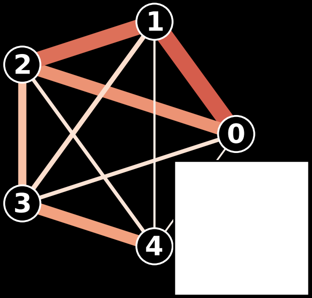




2 Correlations: observations and indirect effects

Observations

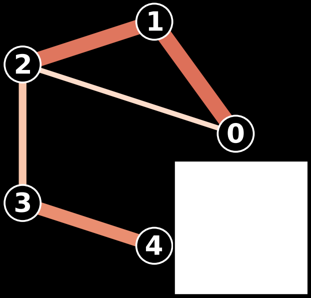
Correlation



Covariance:
scaled by variance

Direct connections

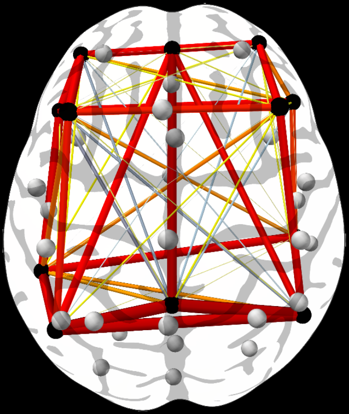
Partial correlation



Inverse covariance:
scaled by partial variance

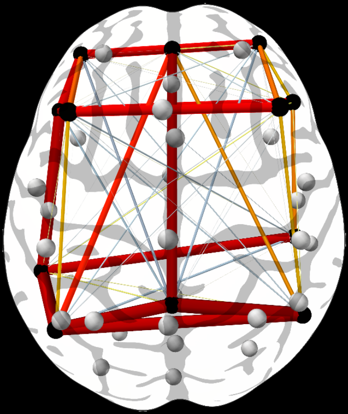
Observations

Correlation



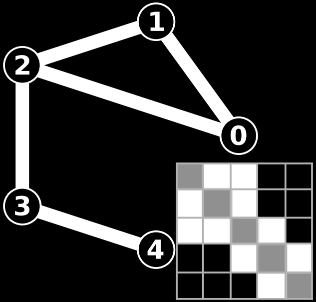
Direct connections

Partial correlation



Gaussian graphical models

Zeros in inverse covariance give **conditional independence**



$$\boldsymbol{\Sigma}^{-1}_{i,j} = 0 \quad \Leftrightarrow \quad \text{$\mathbf{x}_i$, $\mathbf{x}_j$ independent conditionally on } \{\mathbf{x}_k, k \neq i, j\}$$

Sparse inverse covariance

Estimator imposes zeros

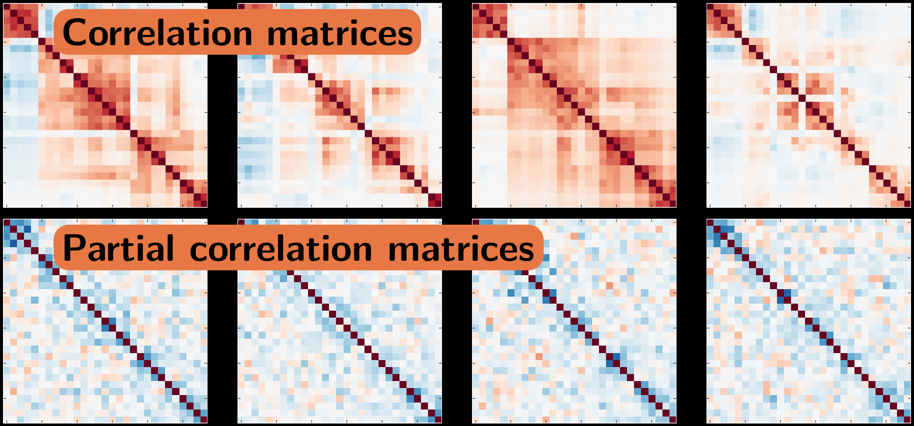
[Smith... 2011, Varoquaux... 2010b]

Shrunk estimator

Estimates closer to 0

[Varoquaux and Craddock 2013]

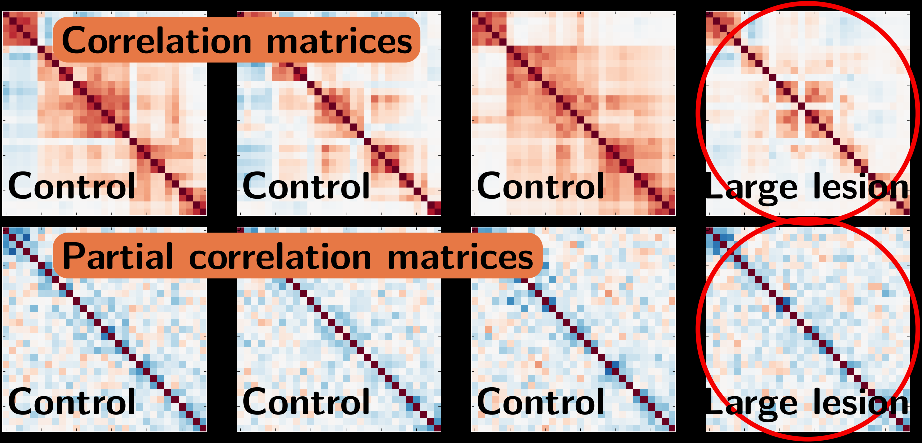
2 Differences in correlations across subjects



3 controls, 1 severe stroke patient

Which is which?

2 Differences in correlations across subjects



- Spread-out variability in correlation matrices
- Noise in partial-correlations

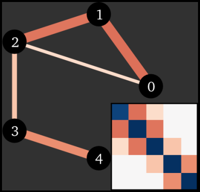
Strong dependence between coefficients

[Varoquaux... 2010a]

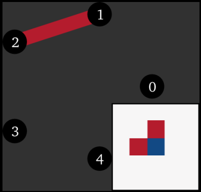
2 A toy model of differences in connectivity

- Two processes with different partial correlations

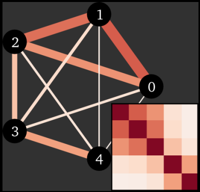
$\mathbf{K}_1$:



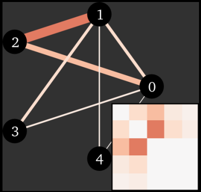
$\mathbf{K}_1 - \mathbf{K}_2$:



$\mathbf{\Sigma}_1$:

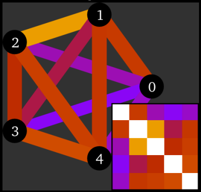


$\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$:

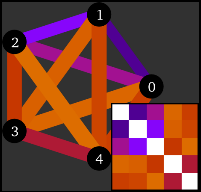


- $+$ jitter in observed covariance

$\text{MSE}(\mathbf{K}_1 - \mathbf{K}_2)$:



$\text{MSE}(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2)$:

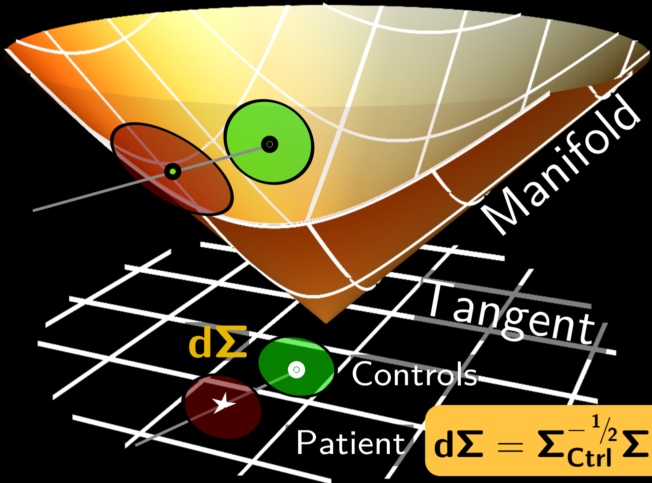


Non-local effects and non homogeneous noise

2 Reparametrization for uniform error geometry

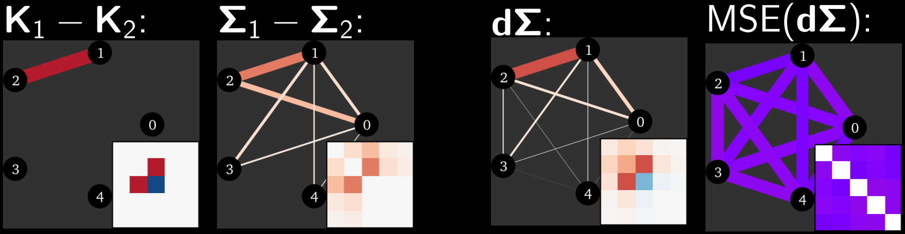
- Disentangle parameters (edge-level connectivities)
- Connectivity matrices form a manifold

$\Rightarrow$ project to tangent space



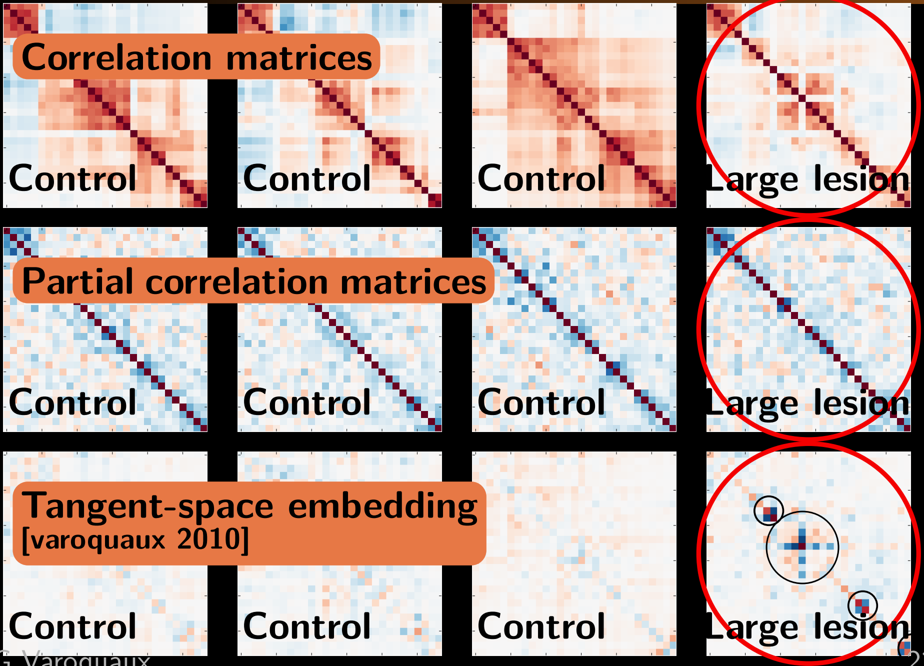
$$\mathbf{d\Sigma} = \mathbf{\Sigma}_{\mathbf{Ctrl}}^{-1/2}\mathbf{\Sigma}_{\mathbf{Patient}}\mathbf{\Sigma}_{\mathbf{Ctrl}}^{-1/2}$$

The simulations



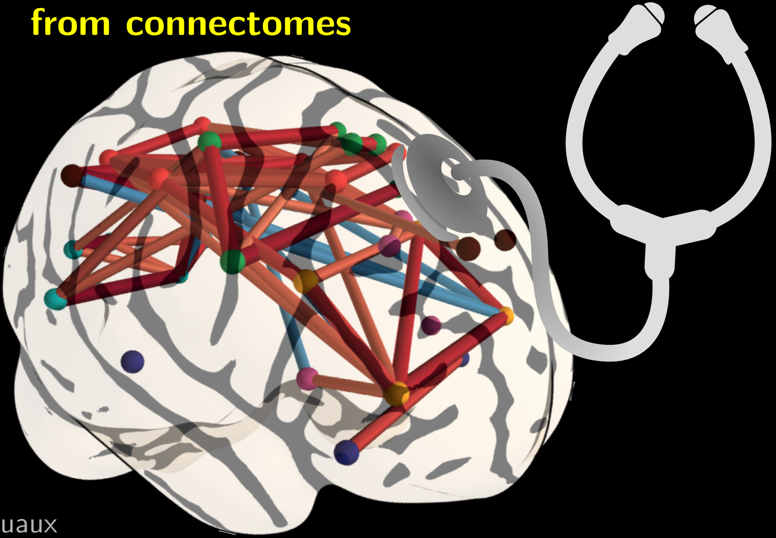
Semi-local effects and homogeneous noise

2 Which parametrization capture differences



3 Biomarkers of autism

from connectomes

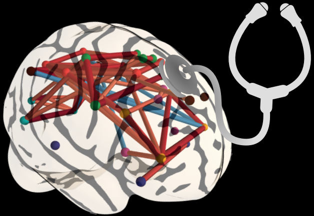


Predicting diagnostic status a good success metric

Multi-site large autism dataset: ABIDE

[Di Martino... 2014]

- Autism Spectrum Disorder
  $\Rightarrow$ Patient/Control classification
- 16 sites
- $\sim$ 1000 subjects

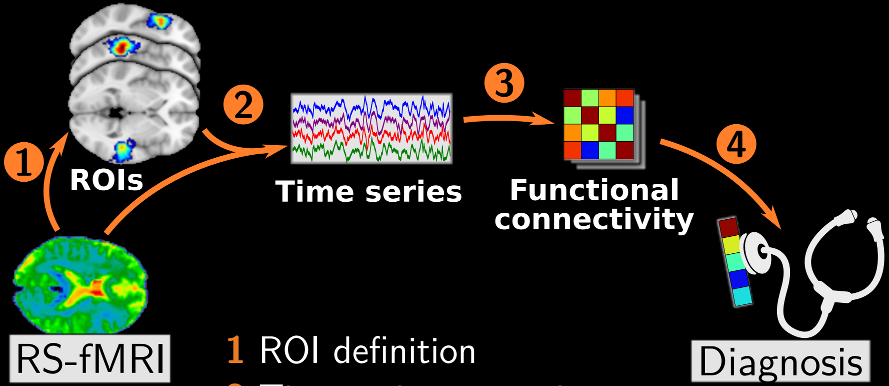


Biomarkers robust to inter-site variations

- Cross-validation predicting to new sites

α	β	γ	δ		ϵ
Training set					Testing set

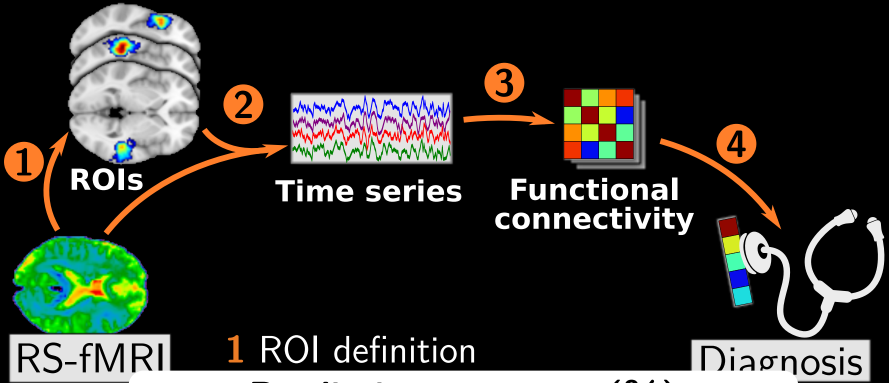
A connectome classification pipeline



1 ROI definition
2 Time-series extraction
3 Connectivity matrices
4 Supervised learning

[Abraham... 2016]

A connectome classification pipeline



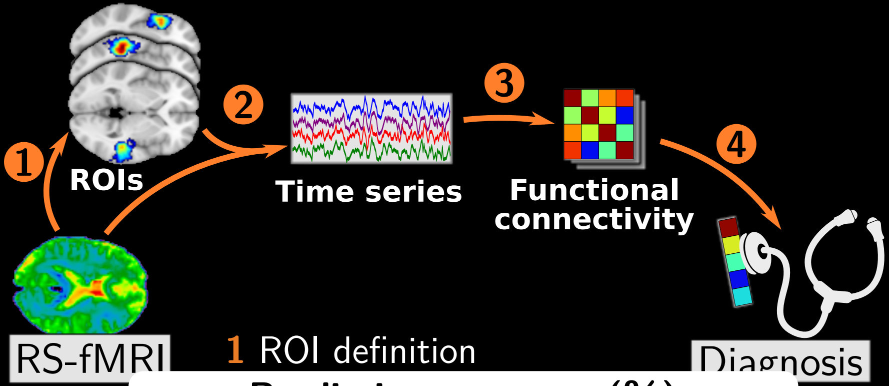
1 ROI definition

Prediction accuracy (%)

Seen sites	**67±3**
Unseen sites	**67±5**

016]

A connectome classification pipeline



1 ROI definition

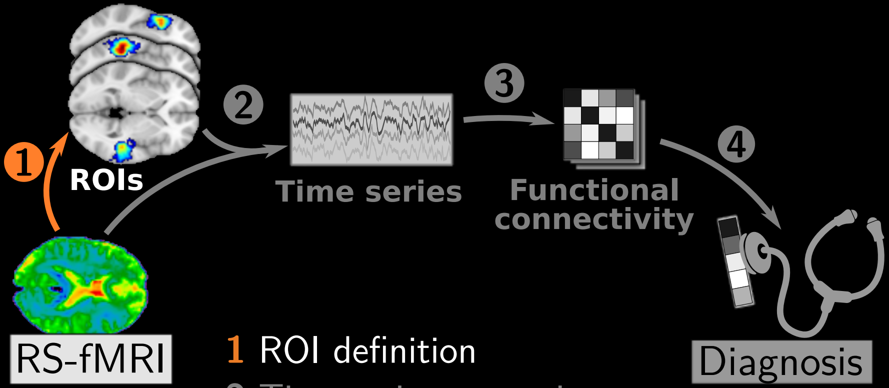
Prediction accuracy (%)

Seen sites	**67 ± 3**
Unseen sites	**67 ± 5**

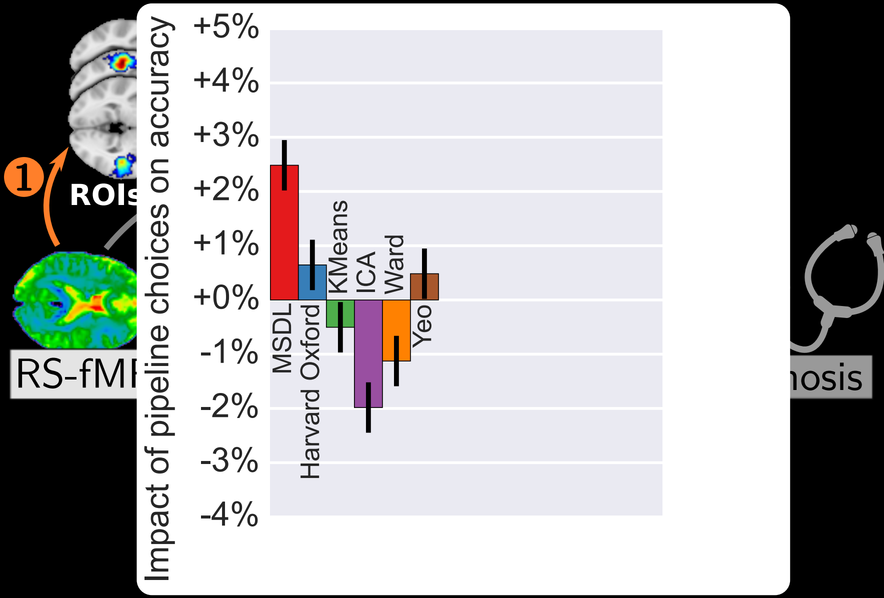
What is important to predict?

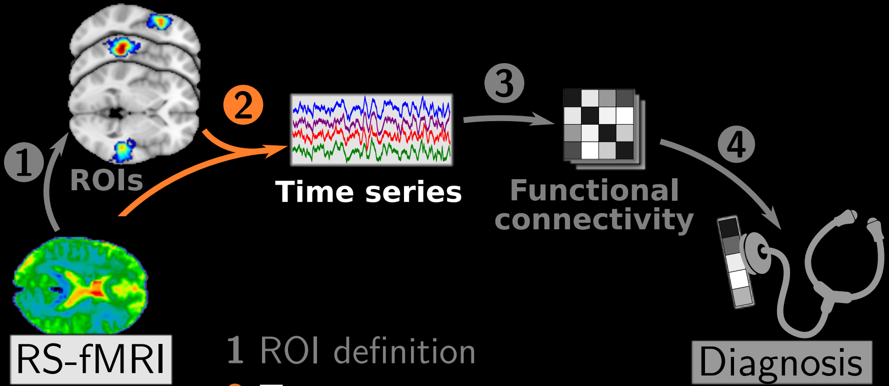
016]

3 ROI definition: impact of choice



1 ROI definition
2 Time-series extraction
3 Connectivity matrices
4 Supervised learning


ROIs
RS-fMRI
nosis
Impact of pipeline choices on accuracy
+5%
+4%
+3%
+2%
+1%
+0%
-1%
-2%
-3%
-4%
MSDL
Harvard Oxford
KMeans
ICA
Ward
Yeo



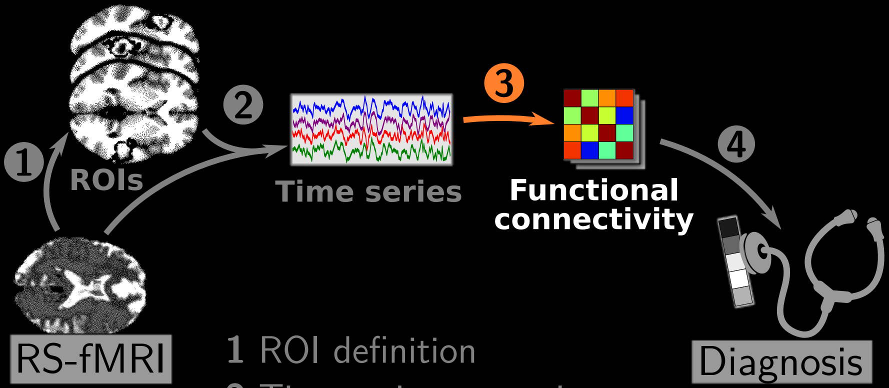
1 ROI definition

2 Time-series extraction

- Remove motion regressors
- Compcorr
- Global mean regression

Empirically: different ways work

3 Functional-connectivity matrix

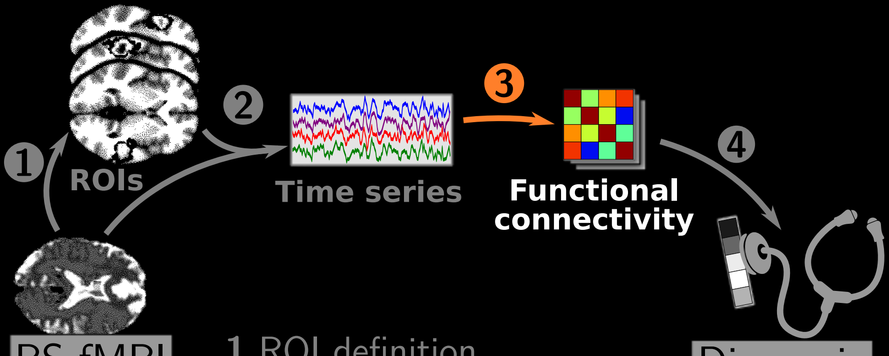


1 ROI definition

2 Time-series extraction

3 Connectivity matrices

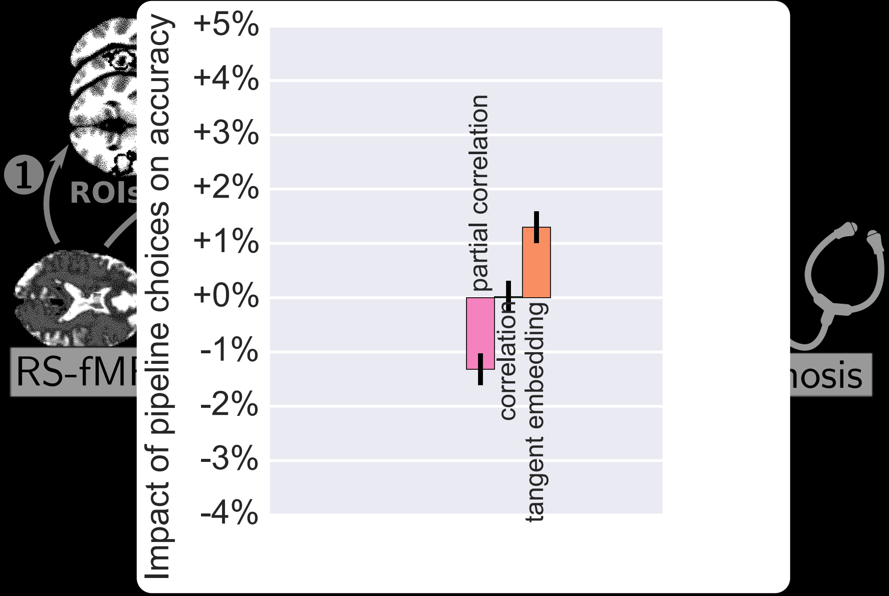
4 Supervised learning

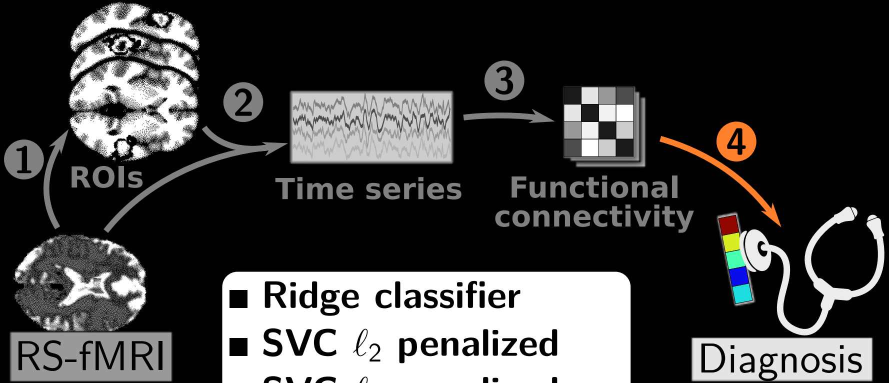


- Correlation matrix?
- Partial correlation matrix?
- Tangent-space embedding?

[Varoquaux... 2010a]

3 Functional-connectivity matrix

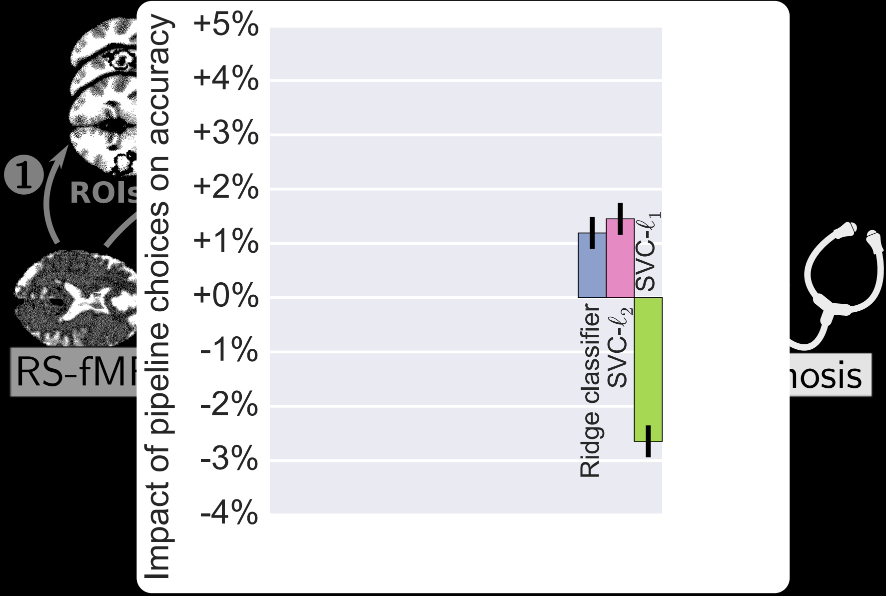




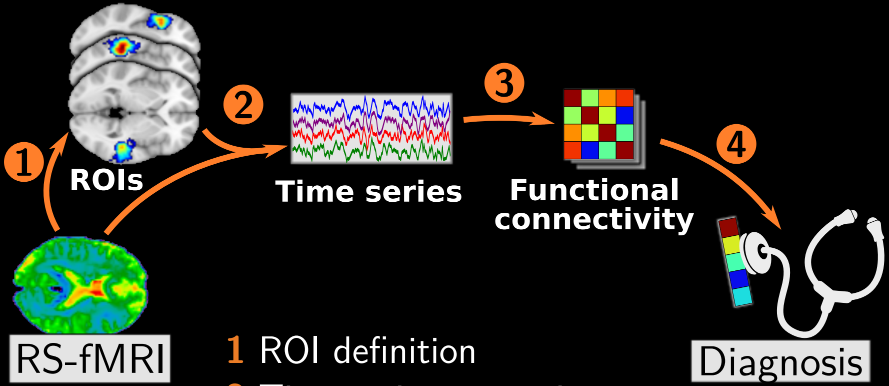
- **Ridge classifier**
- **SVC ℓ_2 penalized**
- **SVC ℓ_1 penalized**

3 Connectivity matrices

4 Supervised learning

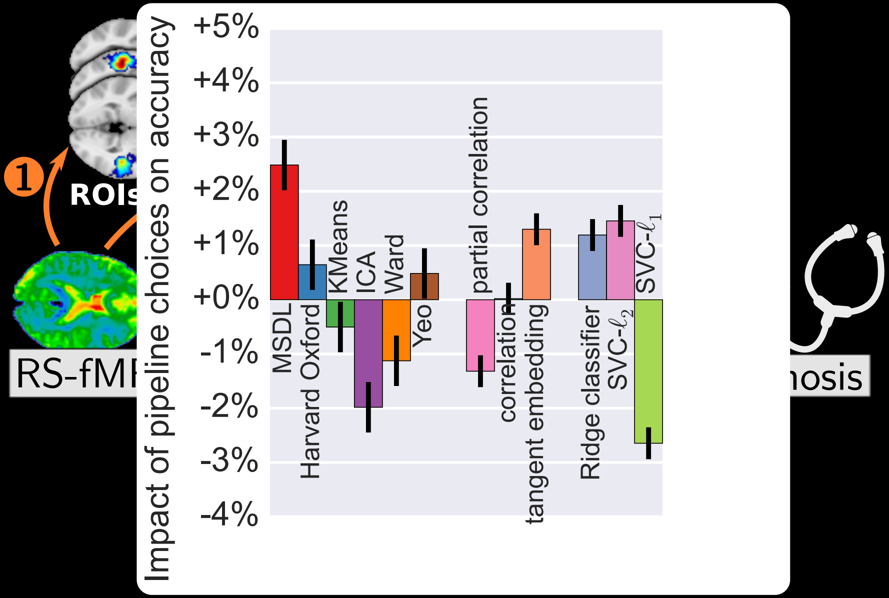

ROIs
RS-fMRI
nosis
Impact of pipeline choices on accuracy
+5%
+4%
+3%
+2%
+1%
+0%
-1%
-2%
-3%
-4%
Ridge classifier
SVC-ℓ_2
SVC-ℓ_1

Importance of pipeline steps

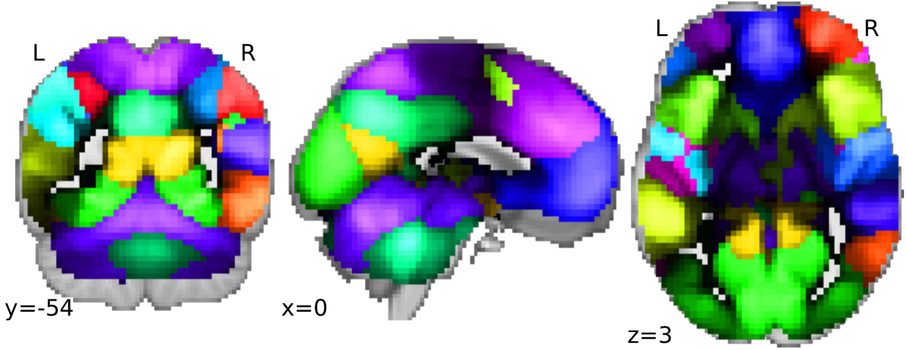


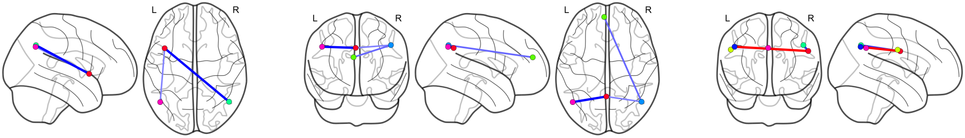
1 ROI definition
2 Time-series extraction
3 Connectivity matrices
4 Supervised learning

Importance of pipeline steps

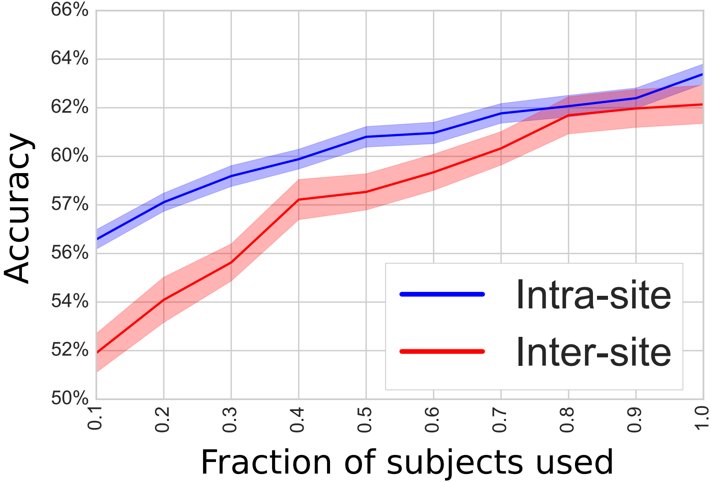


MSDL atlas





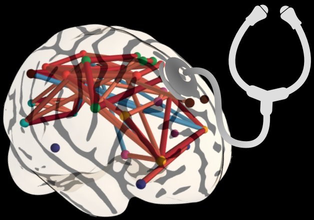
More data is better



Multivariate processing of a 1Tb of heterogeneous data is worth the trouble

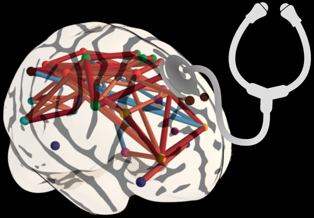
3 Pyschiatric neurophenotypes from rest-fMRI

Viable from data accumulation



- ABIDE is a post-hoc aggregate
- Prediction across sites

Viable from data accumulation



- ABIDE is a post-hoc aggregate
- Prediction across sites

- Not (yet) for clinical diagnostic
- Capture neural signatures of disorders

$\Rightarrow$ **Towards a redefinition of disorders**

Requires huge data accumulation

nilearn: machine learning for neuroimaging

Make it easy for

- Neuroscientists to use machine learning
- Machine learning research to do neuroimaging

Design goal: runs out of the box

Strong points

- Fast and versatile
- High-quality brain plotting
- Simple syntax

Meaningful neuroimaging analysis in examples.

Try it – `http://nilearn.github.io`

[Abraham... 2014]

Neurophenotypes from rest

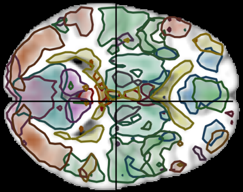
Recipe for good neurophenotypes

- Choice of regions critical (learn them)
- Tangent-space embedding
- Standard SVM

Neurophenotypes from rest

Recipe for good neurophenotypes

- Choice of regions critical (learn them)
- Tangent-space embedding
- Standard SVM



Dictionary learning MSDL

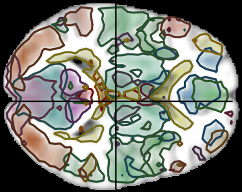
Good definitions of regions

Validation is very hard

Neurophenotypes from rest

Recipe for good neurophenotypes

- Choice of regions critical (learn them)
- Tangent-space embedding
- Standard SVM



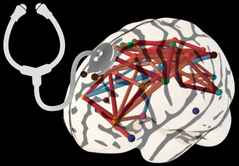
Dictionary learning MSDL

Good definitions of regions

Validation is very hard

Prediction of autism across sites

[Abraham... 2016]



@GaelVaroquaux

References I

A. Abraham, E. Dohmatob, B. Thirion, D. Samaras, and G. Varoquaux. Extracting brain regions from rest fMRI with total-variation constrained dictionary learning. In *MICCAI*, page 607. 2013.

A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Mueller, J. Kossaifi, A. Gramfort, B. Thirion, and G. Varoquaux. Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics*, 8, 2014.

A. Abraham, M. Milham, A. Di Martino, R. C. Craddock, D. Samaras, B. Thirion, and G. Varoquaux. Deriving robust biomarkers from multi-site resting-state data: An autism-based example. *bioRxiv*, page 075853, 2016.

R. C. Craddock, G. A. James, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33(8): 1914–1928, 2012.

References II

A. Di Martino, C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto, ... The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19:659, 2014.

K. L. Miller, F. Alfaro-Almagro, N. K. Bangerter, D. L. Thomas, E. Yacoub, J. Xu, A. J. Bartsch, S. Jbabdi, S. N. Sotiropoulos, J. L. Andersson, ... Multimodal population brain imaging in the uk biobank prospective epidemiological study. *Nature Neuroscience*, 2016.

S. Smith, K. Miller, G. Salimi-Khorshidi, M. Webster, C. Beckmann, T. Nichols, J. Ramsey, and M. Woolrich. Network modelling methods for fMRI. *Neuroimage*, 54:875, 2011.

B. Thirion, G. Varoquaux, E. Dohmatob, and J. Poline. Which fMRI clustering gives good brain parcellations? *Name: Frontiers in Neuroscience*, 8:167, 2014.

References III

G. Varoquaux and R. C. Craddock. Learning and comparing functional connectomes across subjects. *NeuroImage*, 80:405, 2013.

G. Varoquaux, F. Baronnet, A. Kleinschmidt, P. Fillard, and B. Thirion. Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling. In *MICCAI*, pages 200–208. 2010a.

G. Varoquaux, A. Gramfort, J. B. Poline, and B. Thirion. Brain covariance selection: better individual functional connectivity models using population prior. In *NIPS*. 2010b.

G. Varoquaux, A. Gramfort, F. Pedregosa, V. Michel, and B. Thirion. Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. In *Inf Proc Med Imag*, pages 562–573, 2011.

References IV

B. Yeo, F. Krienen, J. Sepulcre, M. Sabuncu, ... The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysio*, 106:1125, 2011.