Graph-based learning in medical imaging and imaging genetics

Jonas Richiardi

@TranslationalML W3 https://unil.ch/tml

UNIL | Université de Lausanne

TRANSLATIONAL MACHINE LEARNING LAB

Stats Forum / DBC Comp Bio Seminar 19/05/2022

Agenda

Background and motivation: why graphs ?

Graph estimation: from images to graphs

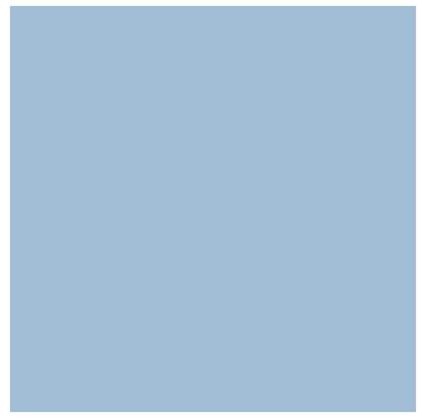
Graph analysis: from graphs to models and inference

Graph imaging genetics: using graphs as endophenotypes

1. Background & motivation

Medical imaging data

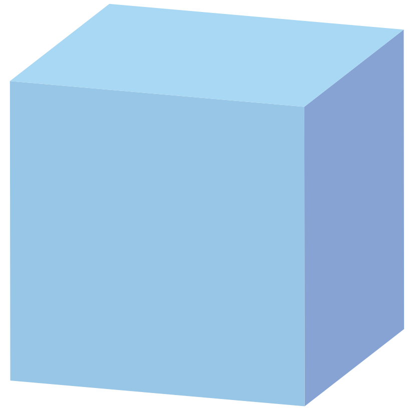
2D



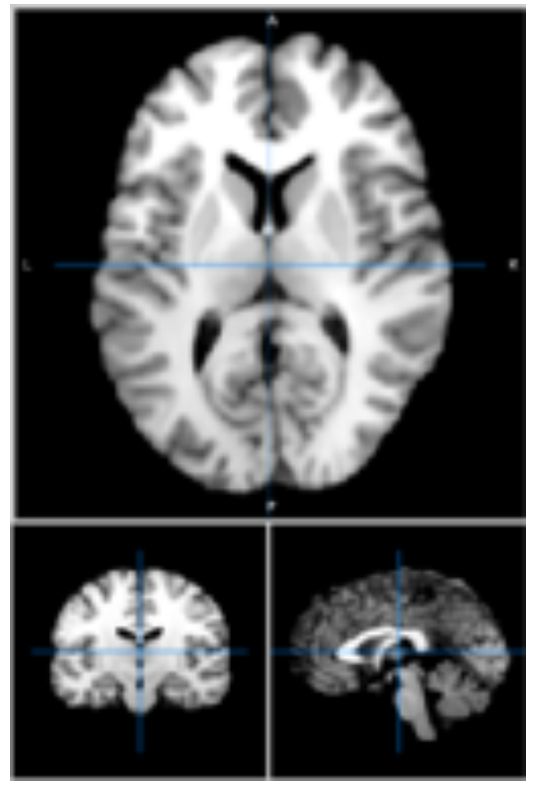
Radiograph

1500x2000

3D

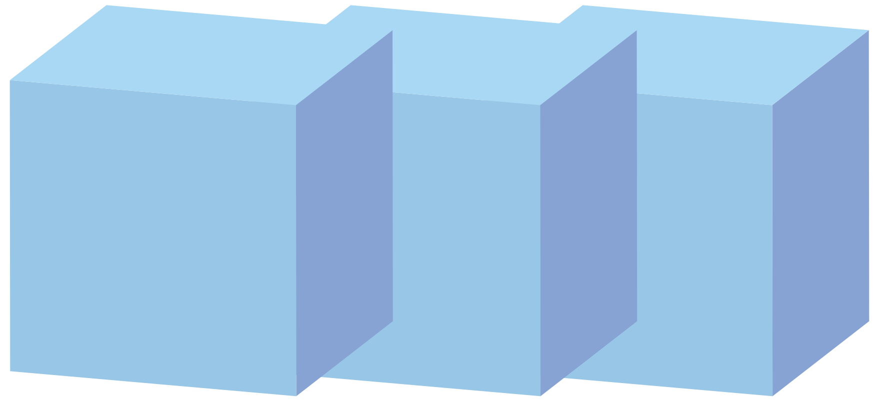


T1w sMRI

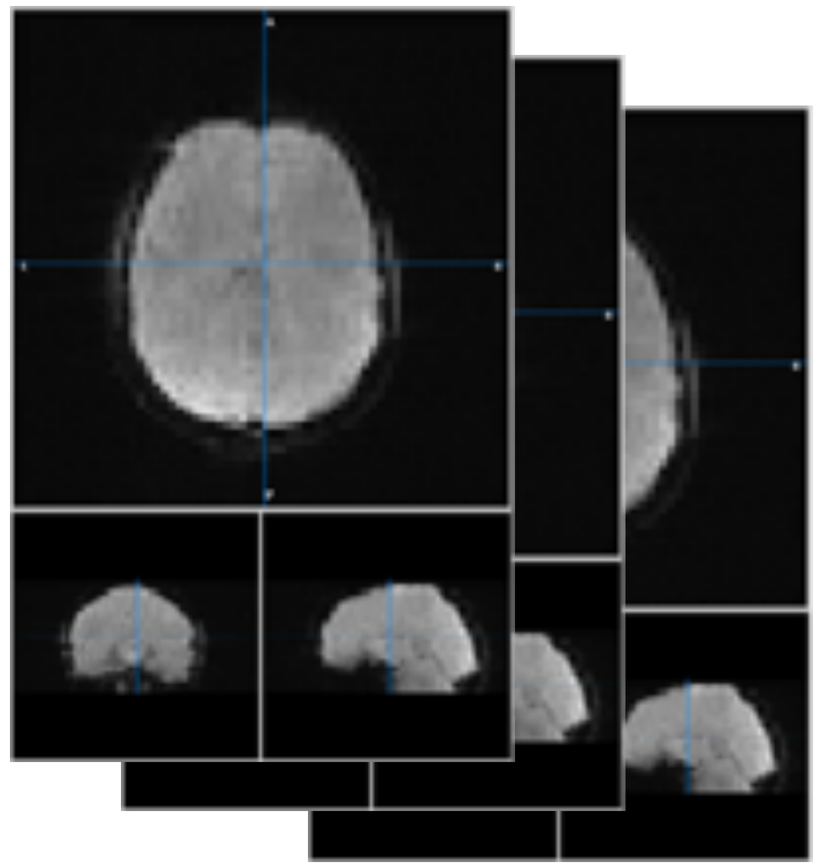


161x191x151

4D



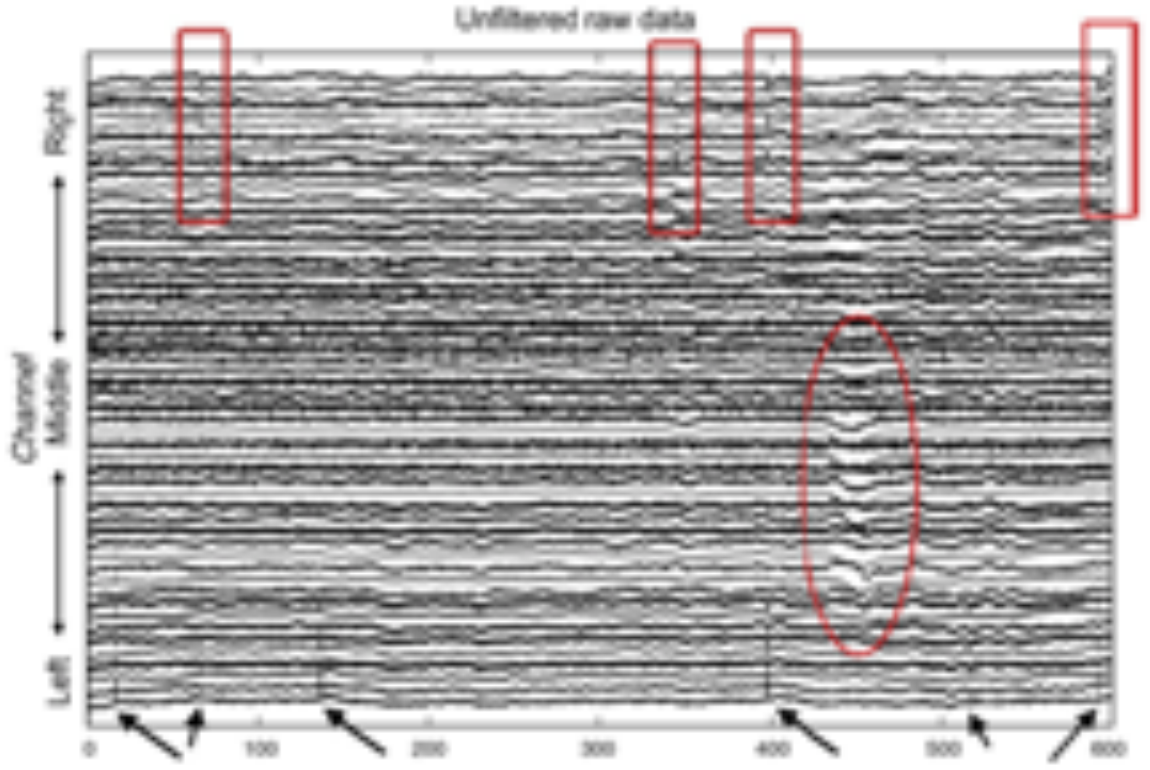
BOLD fMRI



64x61x21@0.9 Hz

MVTS

fNIRS



52x600@10Hz

What types of analyses can we perform?

Image → Image prediction

[Dou et al., IEEE TMI, 2016]

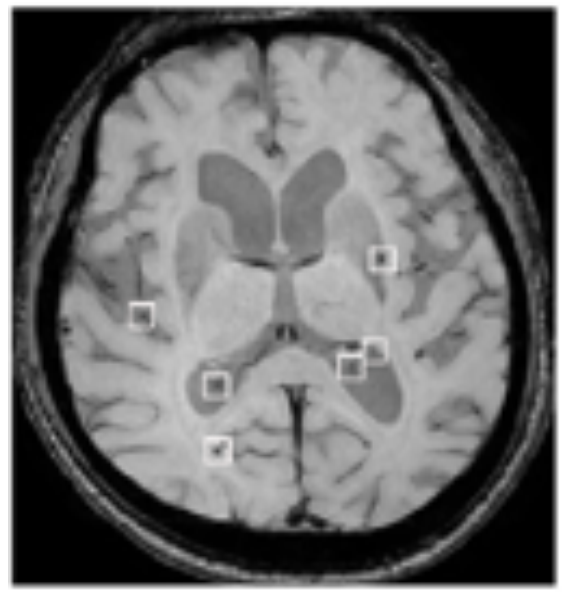
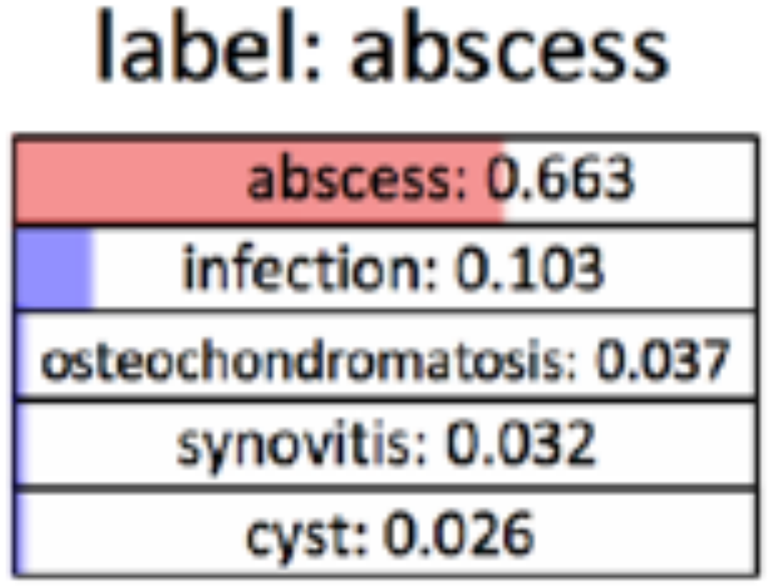


Image (+X) → Clinical prediction

[Shin et al., JMLR, 2016]



Acquisition/Reconstruction/Image processing

- I→I Synthetic contrast generation
- I→I Multi-site data
- S→I Sequence/Reconstruction optimisation (RF pulse design)

Diagnosis

- I→I: lesion segmentation, abnormality detection (e.g. Aneurysms)
- I→C: direct Dx / DDx
- I→C: patient subtyping

Prognosis

- I→C: clinical score change, survival
- I→C: high-risk/low-risk stratification (care / trial enrichment)

Treatment planning

- I→C: drug repurposing, responder/non-responder

How do we perform these analyses?

Imaging data are quite high-dimensional (in terms of voxels) compared to clinical data

Brain structural imaging - 4.6M voxels per image
Brain functional imaging - 82K voxels per time point (image), 37M per session

Deep ML	$\tilde{\mathbf{X}} \in \mathbb{R}^{\tilde{D} \times N}$	$\longrightarrow$		$\arg\min_{\Theta} \mathcal{L}(f(\tilde{\mathbf{X}}; \Theta), Y)$
	Original image			trained model
"Classical" statistical ML	$\tilde{\mathbf{X}} \in \mathbb{R}^{\tilde{D} \times N}$ $\rightarrow$	$\mathbf{X} \in \mathbb{R}^{D \times N}; D \leq \tilde{D}$	$\rightarrow$	$\arg\min_{\Theta} \mathcal{L}(f(\mathbf{X}; \Theta), Y)$
	Original image	handcrafted features		trained model

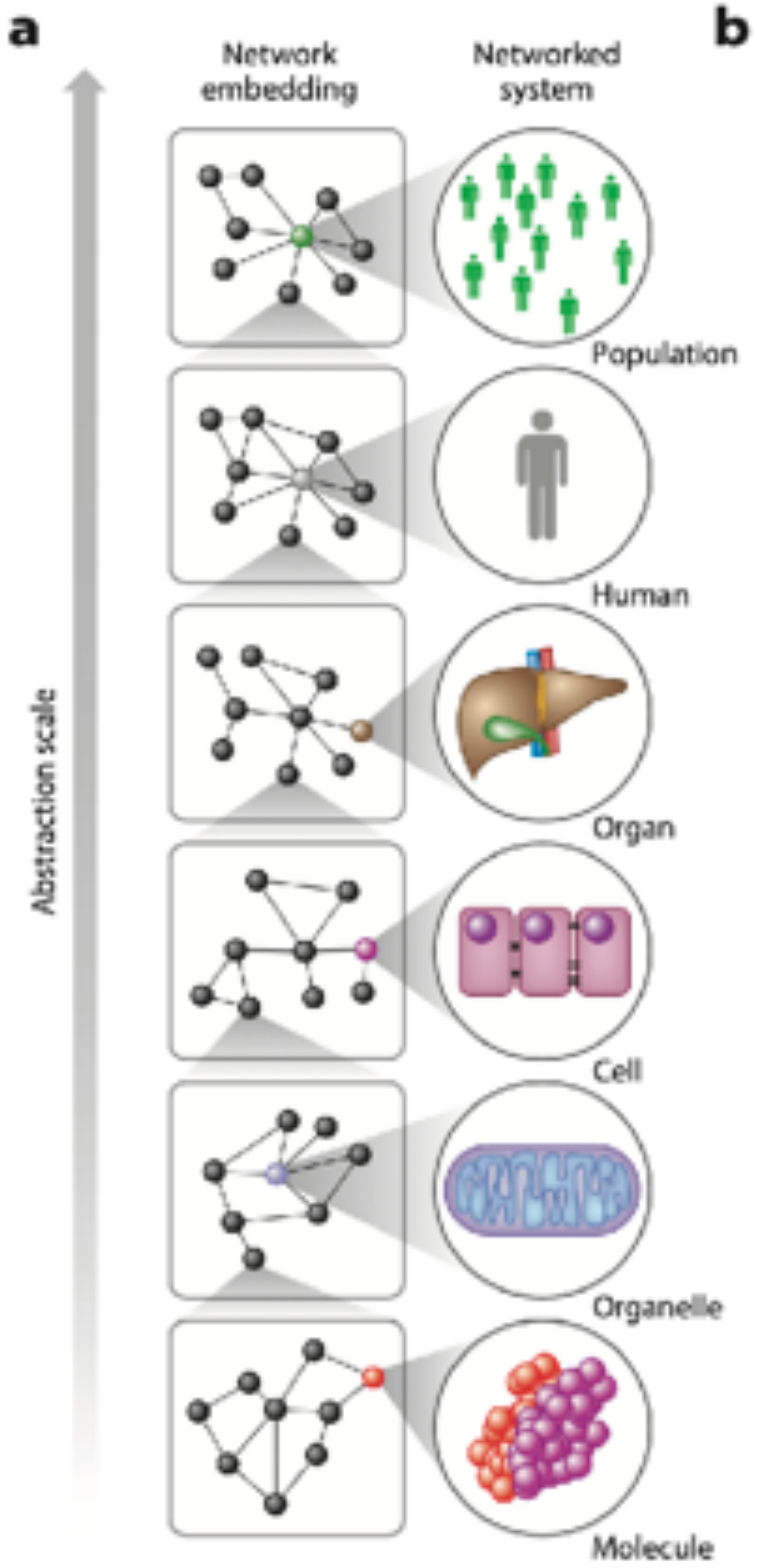
Use prior anatomical knowledge, signal processing, assumptions, and analysis goals to define a lower-dimensional representation

Graphs in biology and biomedical imaging

Annual Review of Biomedical Data Science

Network Analysis as a Grand Unifier in Biomedical Data Science

Patrick McGillivray,[1] Declan Clarke,[1] William Meyerson,[2] Jing Zhang,[1,2] Donghoon Lee,[2] Mengting Gu,[2,3] Sushant Kumar,[1] Holly Zhou,[1] and Mark Gerstein[1,2,3]



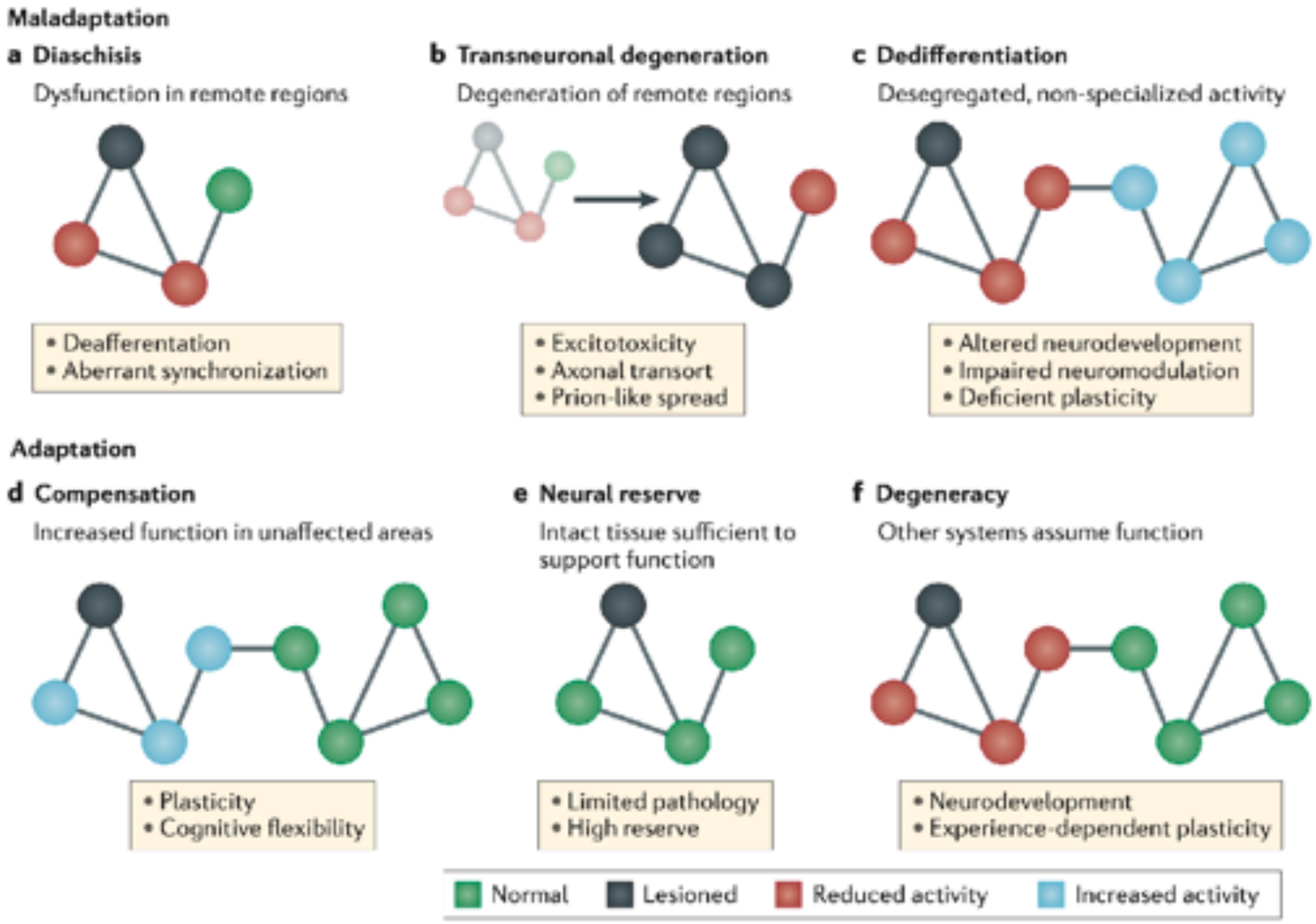
(brain) graphs: a meso-scale representation for (brain) imaging analysis.

node = brain region
edge = “connection” (structural or functional)

GRAIL
Workshop on GRaphs in biomedicAl Image anaLysis

GeoMedIA Workshop 2022
Geometric Deep Learning in Medical Image Analysis

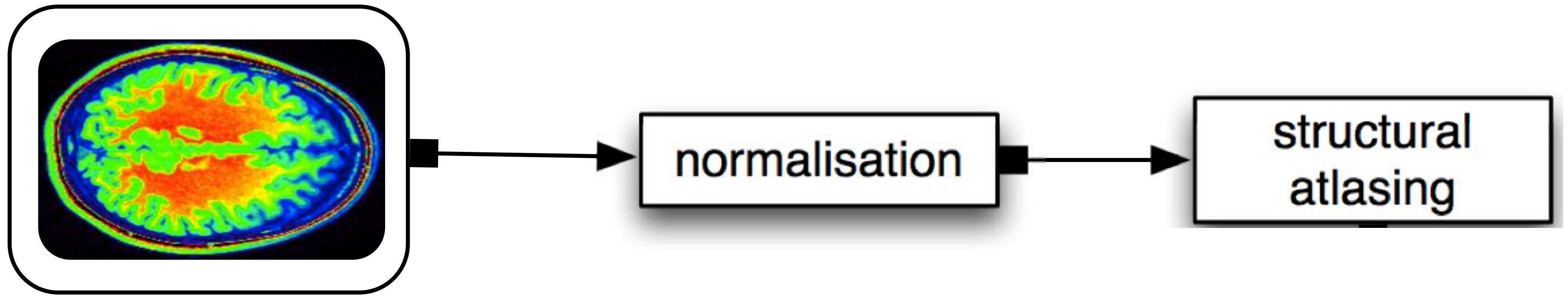
Brain graphs and interpretability



[Fornito et al Nat. Rev. Neuro 2015]

2. Graph estimation

From spatiotemporal brain imaging to graphs



structural image

valid adjacency
matrix **A** (- diagonal)
for a weighted,
complete,
undirected graph

ML with brain graphs overview [Richiardi et al, IEEE Sig. Proc. Mag. 2013] [Richiardi, EMBC 2015]

Inter-regional correlations

A problem with applications in many fields

Miscellanea 179

IV. The Elimination of Spurious Correlation due to position in Time or Space.

By "STUDENT."

[Gosset/Student, Biometrika, 1914]

Appl. Statist. (1977),
26, *No.* 2, p. 179

Estimation of Interclass Correlation from Familial Data

By B. Rosner, A. Donner and C. H. Hennekens

[Rosner et al, J. Royal Stat Soc. C (App. Statist.), 1977]

2454 JOURNAL OF CLIMATE VOLUME 8

Estimating Spatial Correlations from Spatial–Temporal Meteorological Data

Richard F. Gunst

[Gunst, J. Climate, 1995]

Population and Community Ecology for Insect Management and Conservation, Baumgärtner et al. (eds)
© 1998 Balkema, Rotterdam, ISBN 90 5410 930 0

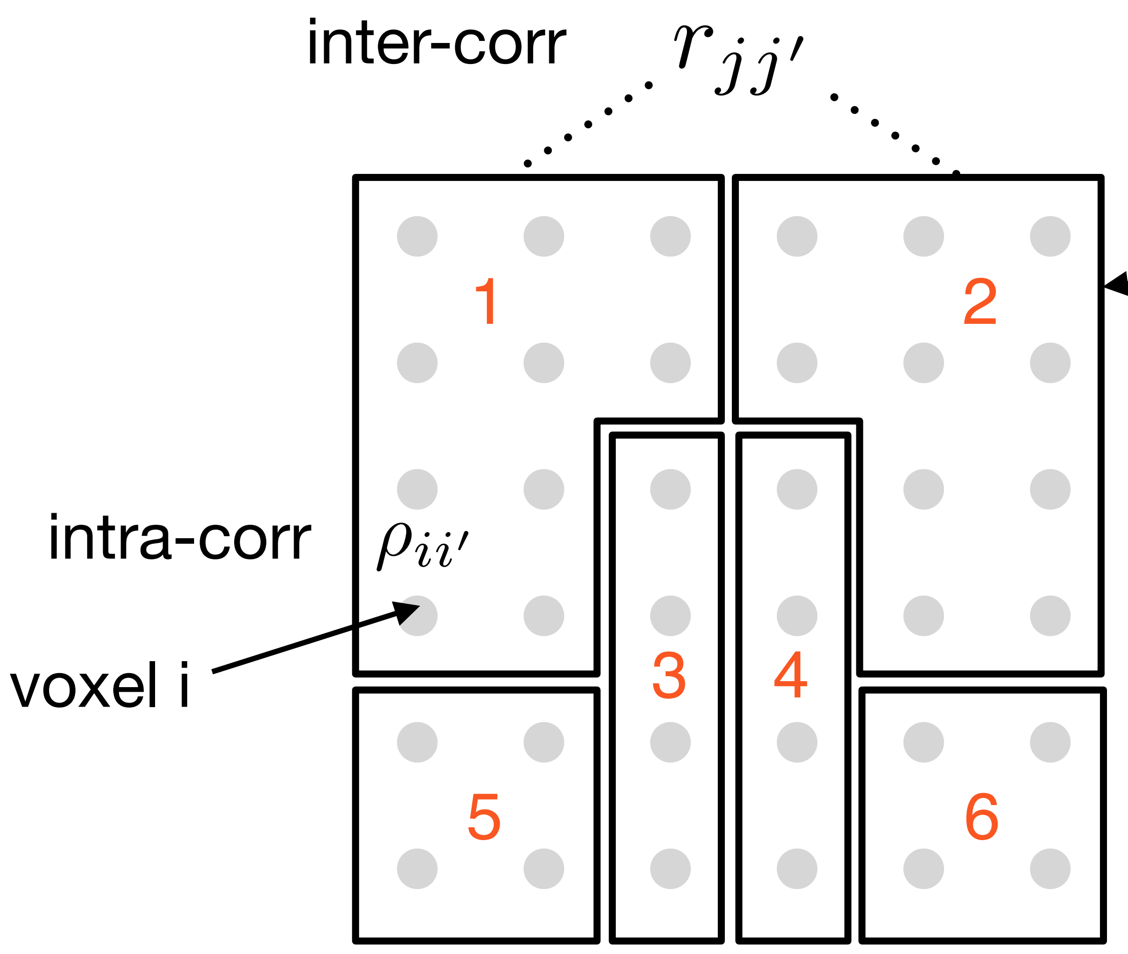
Testing for correlation in the presence of spatial autocorrelation in insect count data

A.M.Liebhold
Northeastern Forest Experiment Station, USDA Forest Service, Morgantown, W.Va., USA
A.A.Sharov
Department of Entomology, Virginia Polytechnic Institute and State University Blacksburg, Va., USA

[Liebhold & Sharov, Proc. Int. Congress on Entomology 1996]

Inter-regional correlations

Notation, model, and assumptions



$\mathcal{C}$ set of indices of all voxels in the brain

$\mathcal{R}_j$ set of voxels in region j

Signal model

$$Y_i(t) = X_i(t) + \varepsilon_i(t) + e(t)$$

Observed voxel value at time t (after DWT) — Neural signal — local noise — global noise

Assumptions

all RVs centered

$X_i(\cdot), \varepsilon_i(\cdot), e(\cdot)$ mutually $\perp$, $\perp$ in time

$E[e(t)^2] = \sigma_e^2$ global noise homoskedastic

$\rho_{ii'} \approx 1$ for small distances

$(X_i(t), i \in \mathcal{R}_j), (\varepsilon_i(t), i \in \mathcal{C})$ stationary random fields

$\forall (i, i', j)\ \rho_{ii'} \in \mathcal{R}_j$

$\forall (i, i', j)\ \eta_{ii'} \in \mathcal{R}_j$

stationary and isotropic with respect to distance

Covariance model under assumptions

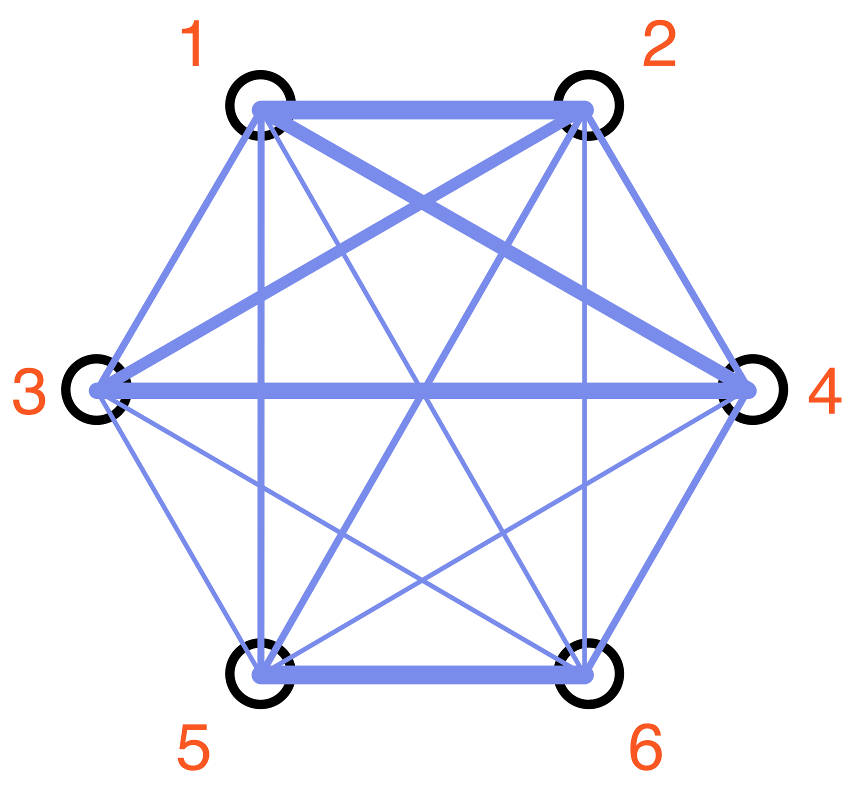
inter-regional correlation = parameter of interest

global noise variance

$$\mathrm{cov}(Y_i(t), Y_{i'}(t)) = \begin{cases} \sigma_j \sigma_{j'} r_{jj'} + \sigma_e^2 & \text{if } j \neq j' \\ \sigma_j^2 \rho_{ii'} + \sigma_\varepsilon^2 \eta_{ii'} + \sigma_e^2 & \text{if } j = j' \end{cases}$$

intra-region correlation

local noise correlation



(OLD PREPRINT) [Achard, Coeurjolly, Lafaye de Micheaux and Richiardi https://arxiv.org/abs/2011.08269, 2020]

[Richiardi et al, IEEE Sig. Proc. Mag. 2013]

[Achard, Coeurjolly, Lafaye de Micheaux, **Lbath** and Richiardi, in prep.]

Inter-regional correlations

Estimators (selected from 9)

Desirable properties: 1) handle regions with different number of voxels , 2) handle small intra-correlations, 3) robust to local noise, (4) robust to global noise)

"correlation of averages"

$$\widehat{r}^{\mathrm{CA}}_{jj'} = \frac{\widehat{\mathrm{cov}}(\bar{\mathbf{Y}}_{\mathcal{R}_j}, \bar{\mathbf{Y}}_{\mathcal{R}_{j'}})}{\widehat{\sigma}(\bar{\mathbf{Y}}_{\mathcal{R}_j})\widehat{\sigma}(\bar{\mathbf{Y}}_{\mathcal{R}_{j'}})}.$$

⊕ local noise: smoothed

⊖ size effect: highly dependent on intra-correlations: estimate for

$$r_{jj'}/\sqrt{\bar{\rho}_{\mathcal{R}_j}\bar{\rho}_{\mathcal{R}_{j'}}}$$

can degenerate if regions are not positively intra-correlated enough

"average of correlations"

$$\widehat{r}^{\mathrm{AC}}_{jj'} = \frac{1}{N_j N_{j'}} \sum_{i \in \mathcal{R}_j, i' \in \mathcal{R}_{j'}} \widehat{\mathrm{cor}}(\mathbf{Y}_i, \mathbf{Y}_{i'}).$$

⊖ local noise: not robust

⊕ size effect: corrected

"local correlation of averages"

$$\widehat{r}^{\ell\mathrm{CA}}_{jj'} = \frac{1}{B}\sum_{b=1}^{B} \widehat{\mathrm{cor}}(\bar{\mathbf{Y}}_{\mathcal{V}_j^{(b)}}, \bar{\mathbf{Y}}_{\mathcal{V}_{j'}^{(b)}}).$$

⊕ local noise: smoothed

⊕ size effect: corrected

fMRI: e.g. [Achard et al, J. Neurosci, 2006]

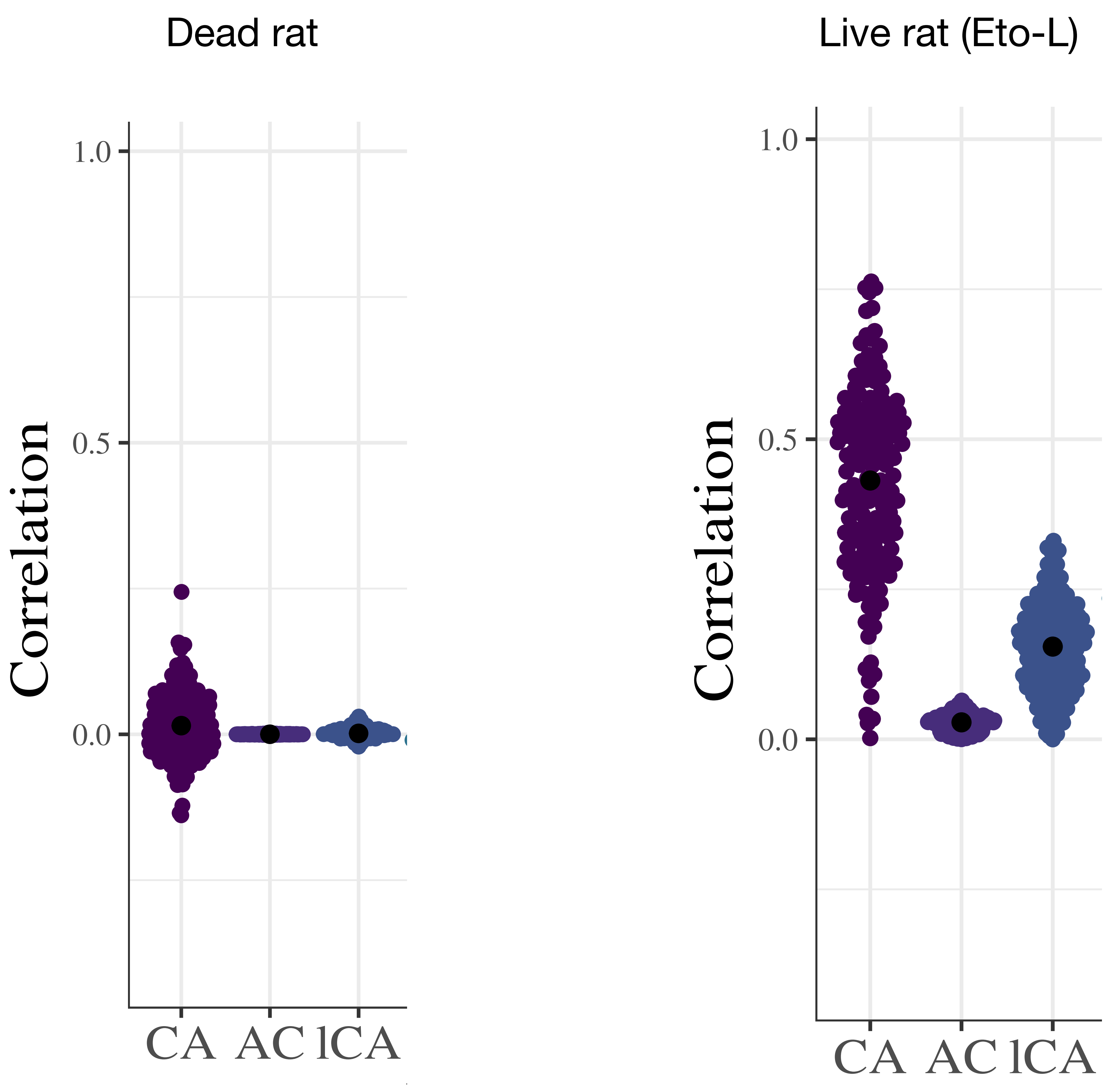
[Achard et al, IEEE SSP, 2011]

[Achard, Coeurjolly, Lafaye de Micheaux, Lbath and Richiardi, in prep.]

Inter-regional correlations

Empirical results - face validity

live vs dead rats
rats, N=2, R=22



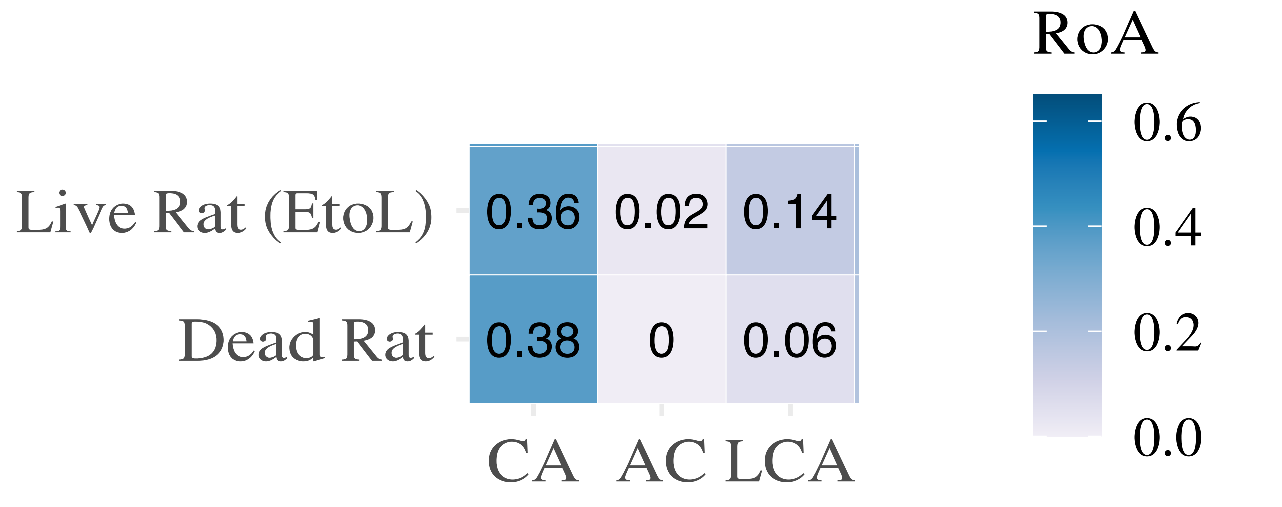
Rat data [Becq et al, J. Neural Eng., 2020]

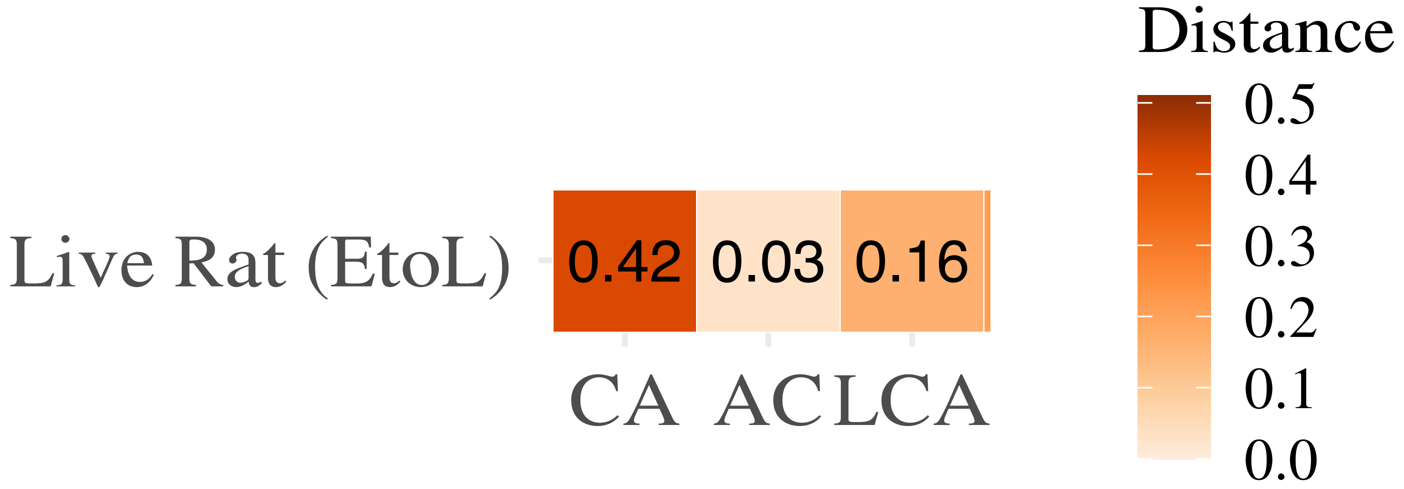
[Achard, Coeurjolly, Lafaye de Micheaux, Lbath and Richiardi, in prep.]

Inter-regional correlations

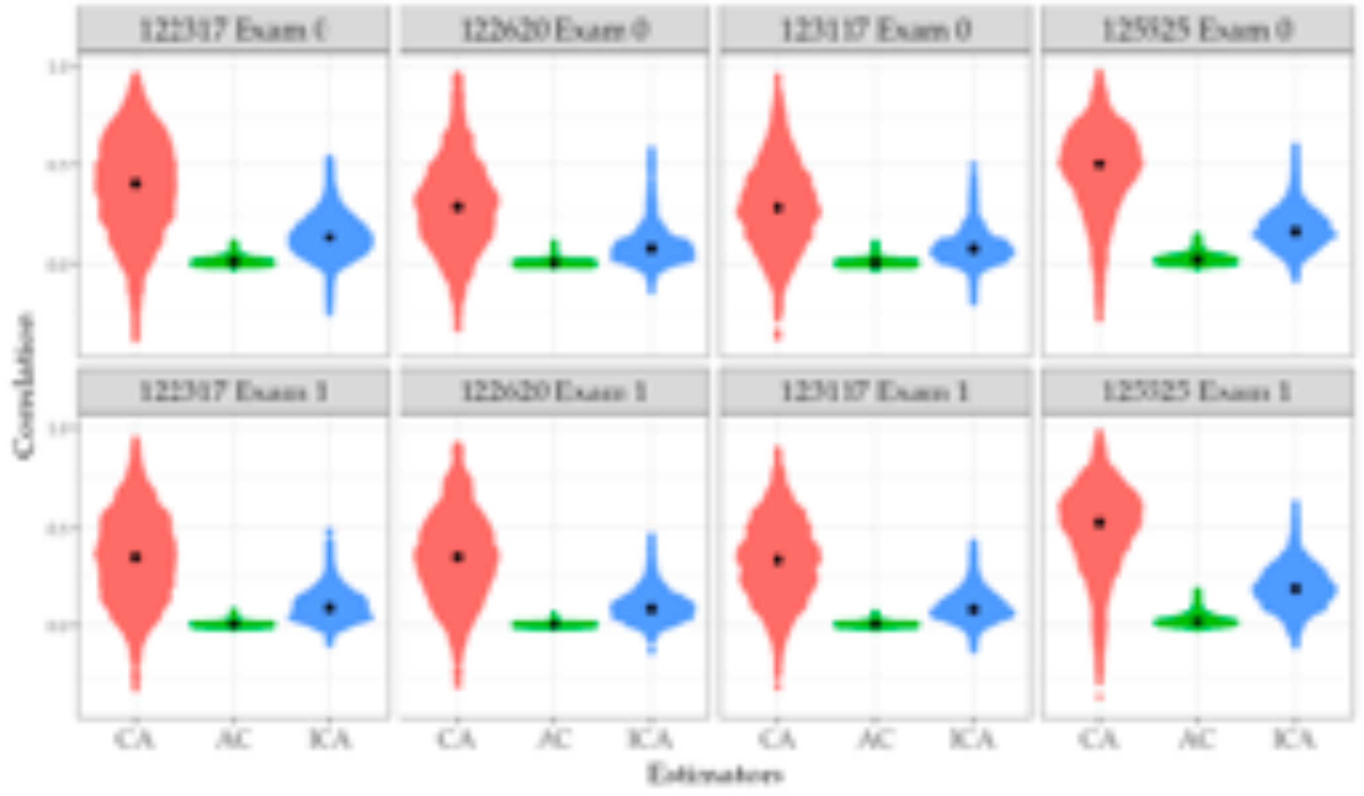
Empirical results - repeatability & discriminative power

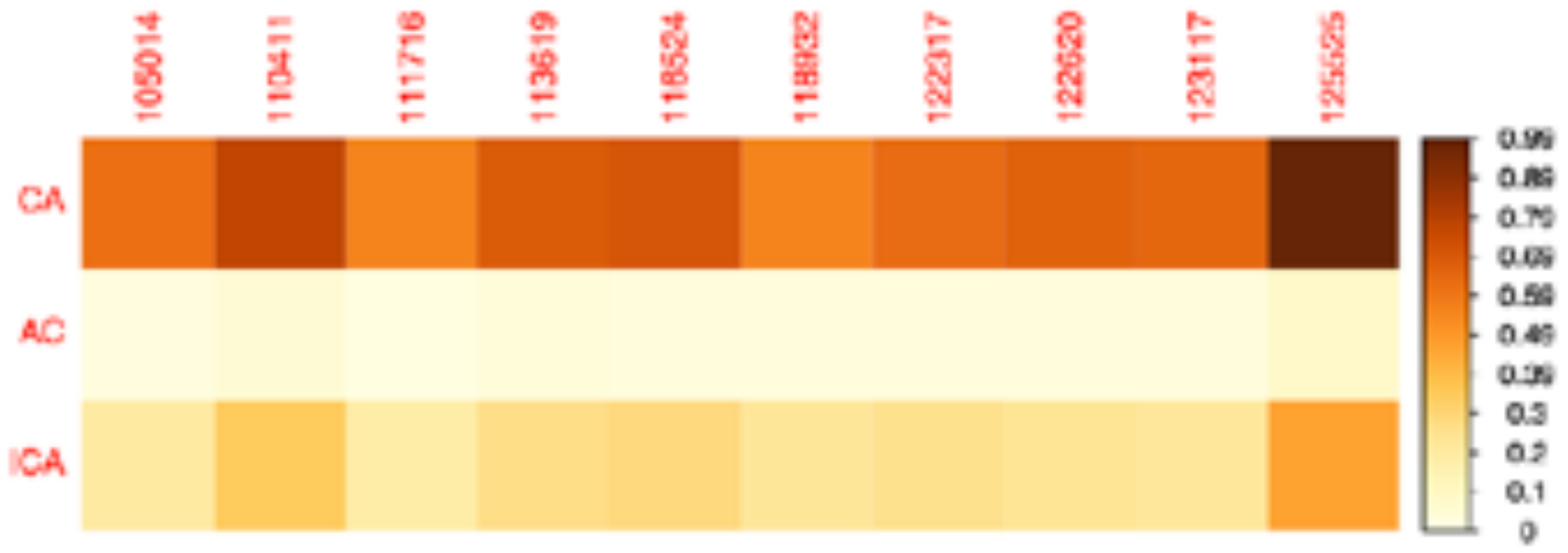
Repeatability
split-half Bland-Altmann & Wasserstein
rats, N=2, R=22





Repeatability
split-half Bland-Altmann
humans[1], N=10, R=90





Discriminative power
fingerprinting[2] & direct embedding distance
humans, N = 10, R=90

estimator	CA	AC	LCA
intra-subject mean (sd)	0.36 (0.15)	0.27 (0.13)	0.35 (0.14)
inter-subject mean (sd)	0.60 (0.18)	0.46 (0.17)	0.57 (0.17)
W	-3.95	-3.36	-3.58
p	3.9×10^{-5}	3.9×10^{-4}	1.7×10^{-4}
identification rate	80%	50%	60%

Conclusion: LCA has face validity, seems to offer increased repeatability, possibly at the expense of discriminative power

[1]Human data https://humanconnectomeproject.org [Achard, Coeurjolly, Lafaye de Micheaux, Lbath and Richiardi, in prep.] [2][Finn et al., Nature Neuroscience, 2015]

Inter-regional correlations
Limitations

Filtering: bandpass filtering

Aggregation: decomposition approaches (ICA, EMD, EEMD…)

Non-linear/causal estimators: Mutual Information, Granger Causality, …

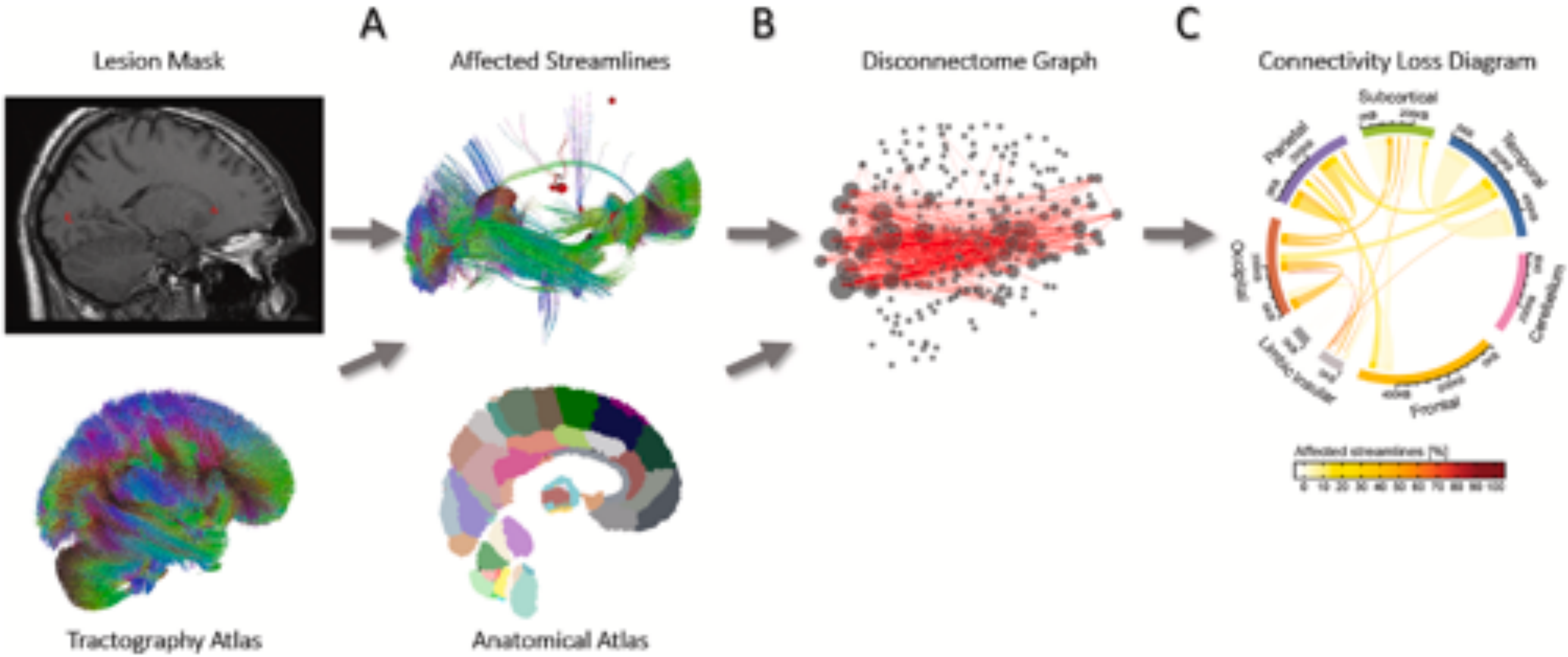
Full structure learning: partial correlation, inverse covariance, PC*…

Regularisation: Ledoit-Wolfe, gLasso…

Noise: multiplicative, convolutional…

Brain disconnectomics

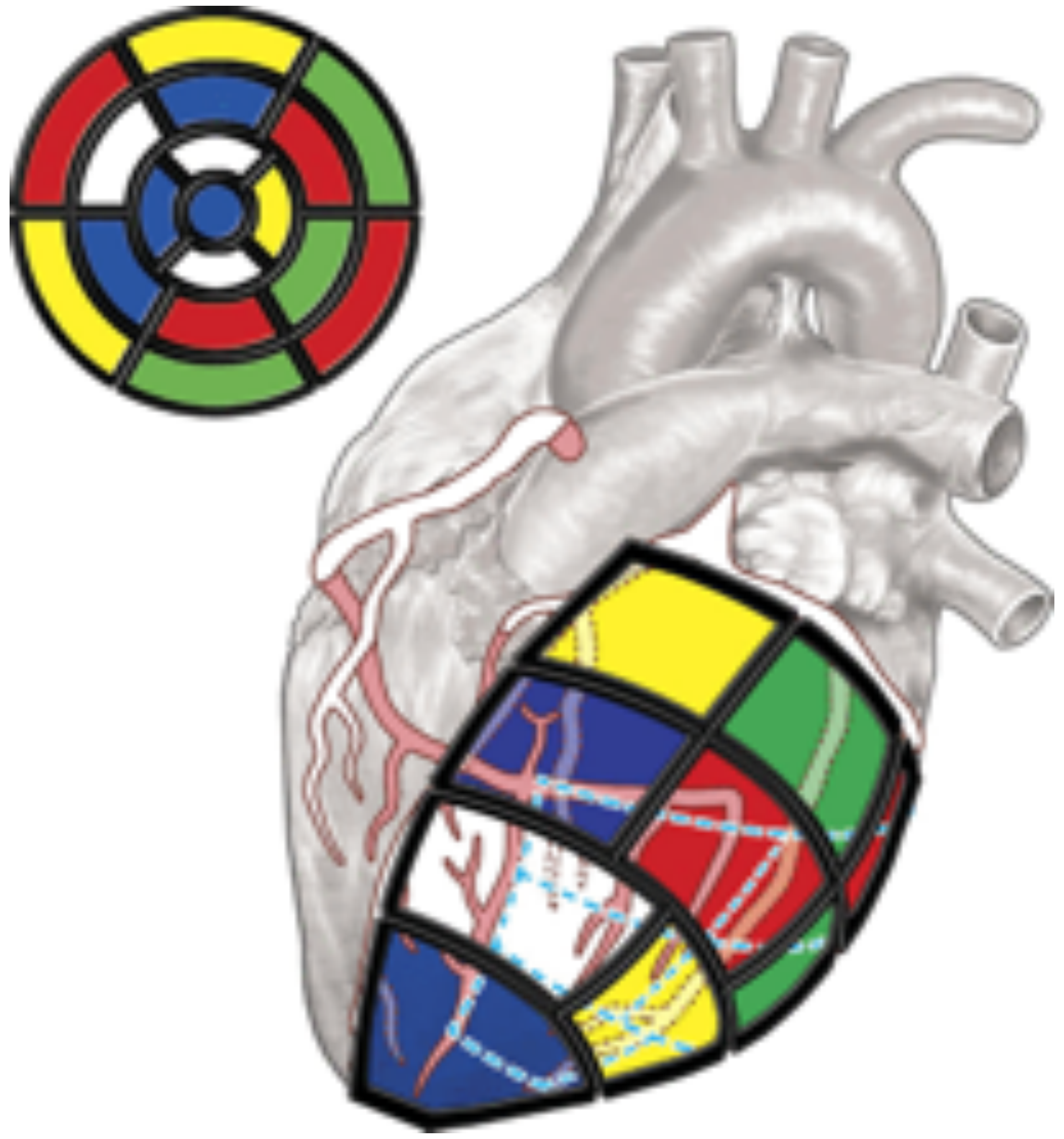
Atlas-based estimation of white matter tract damage in MS



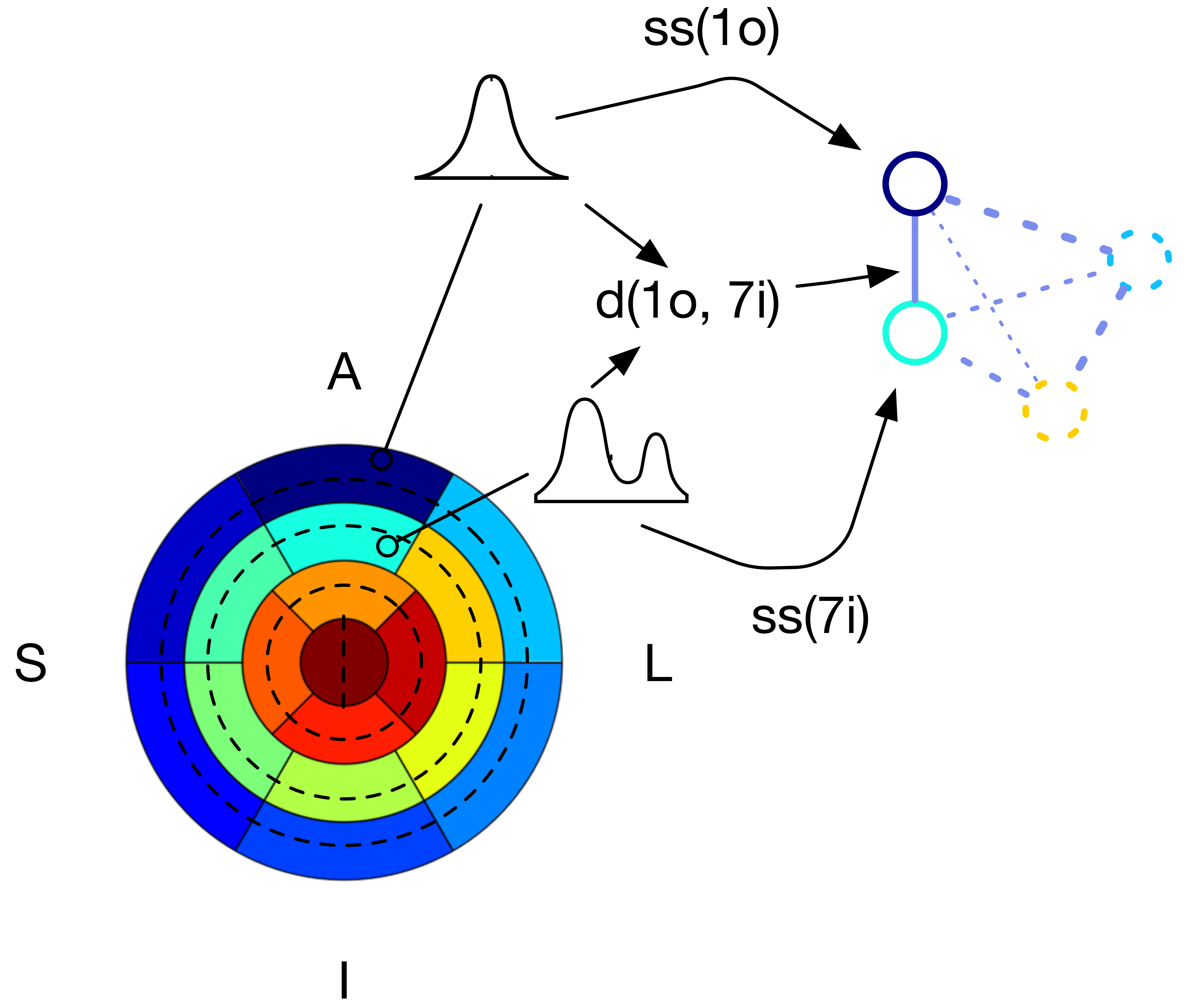
[Ravano et al., Neuroimage: clinical, 2021]

Heart graphs

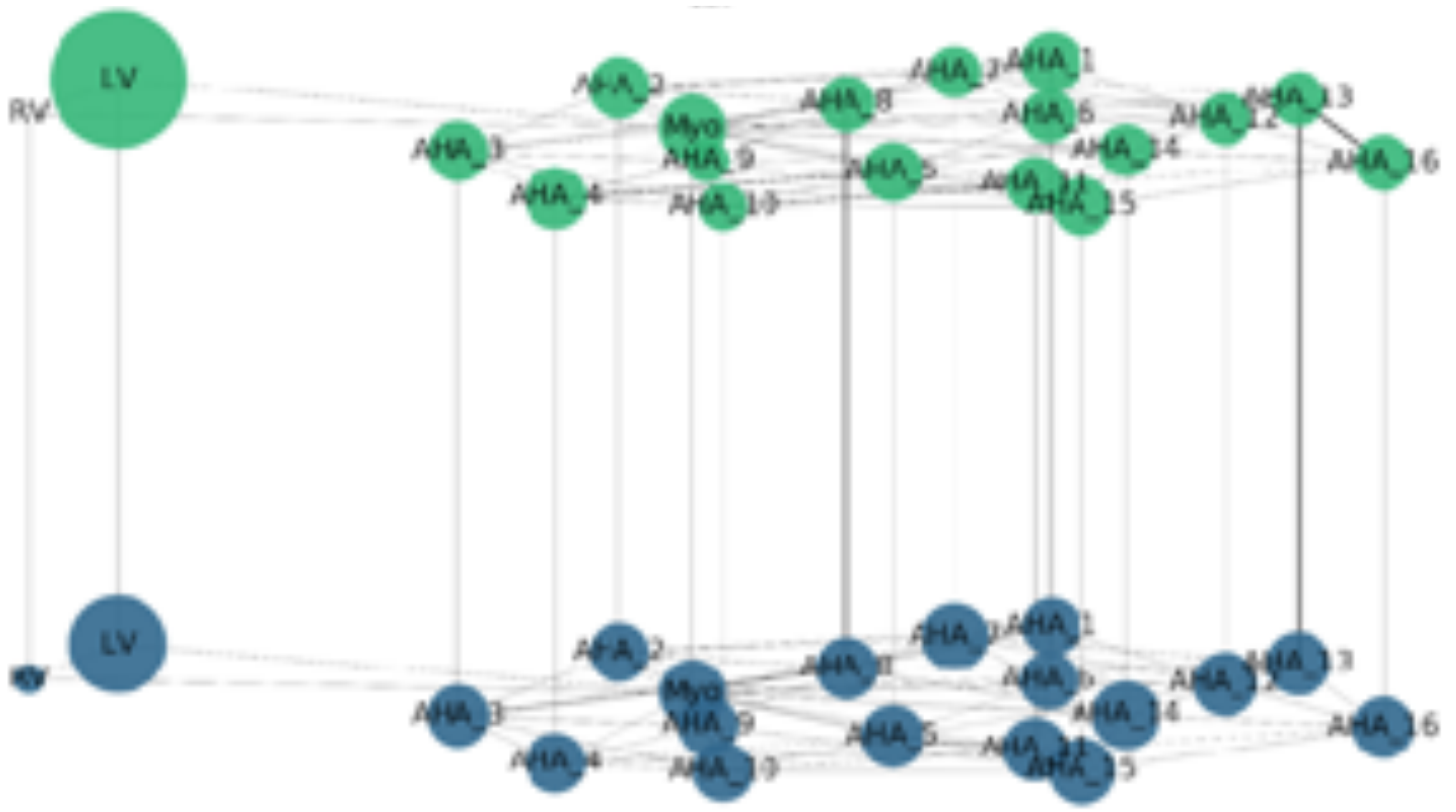
Compact representation of cardiac structure and function



anatomical mapping of
AHA segments



building labelled graphs with
edge labels and vertex labels

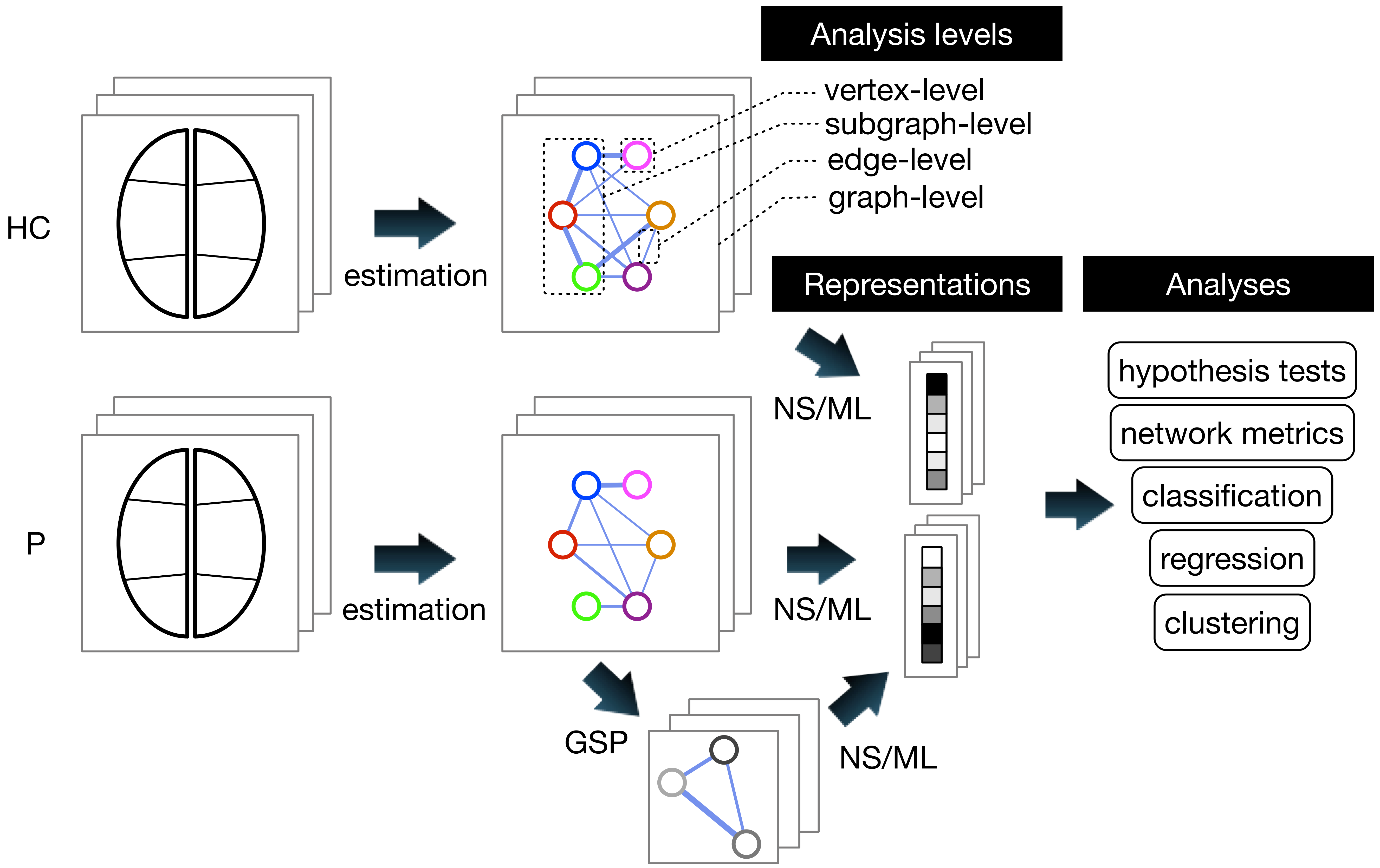


example dilated cardiomyopathy
patient graph, end-diastole (top) and
end-systole (bottom)

[Banus-Cobo et al., in prep.]

3. Graph analysis

Analysis levels, representations, and approaches



e.g. multi-level (subgraph + node/edge) screening-filtering[1]/step-up[2] multiple comparisons correction procedure to control FWER or FDR

[Damoiseaux et al., Advanced Neuro MR Techniques and applications, vol.4, chap. 21, 2021]

Broad overview of brain graph analysis: [De Vico Fallani et al., Phil. Trans. Royal Soc. B, 2014]

[1][Meskaldji et al., NeuroImage, 2015]

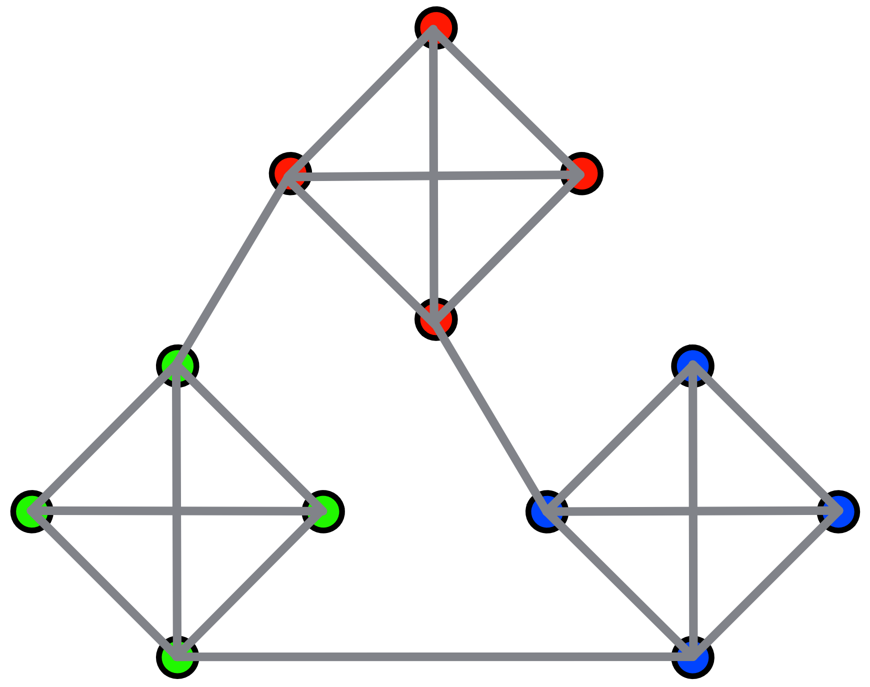
[2][Meskaldji et al., Stat Papers, 2018]

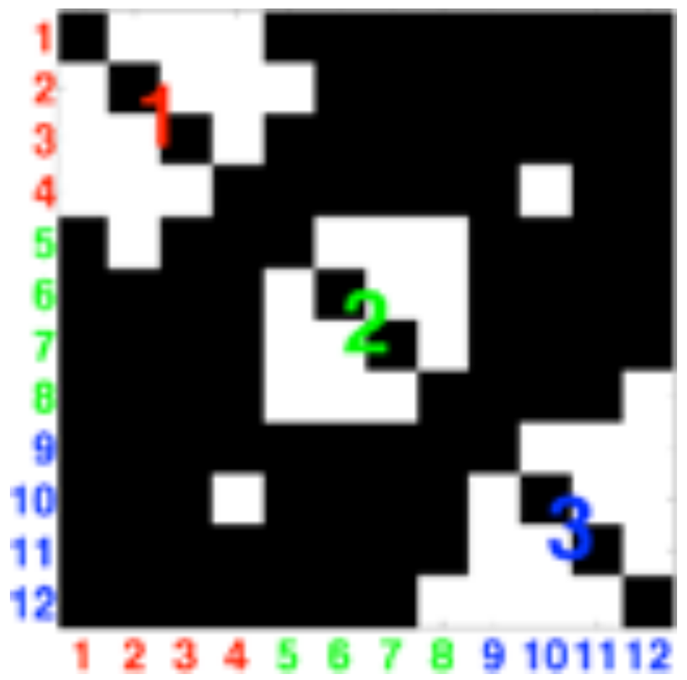
Network science approaches

Focus on subgraphs or topological properties

Principle

Brain graphs have identifiable subgraphs (“modules”, “communities”) in several modalities





The partition into communities can be used to compare brain graphs between subjects at various scales

- Whole-brain: graphwise community structure
- “Subnetwork of regions”: individual communities
- Single region: community membership (not shown)

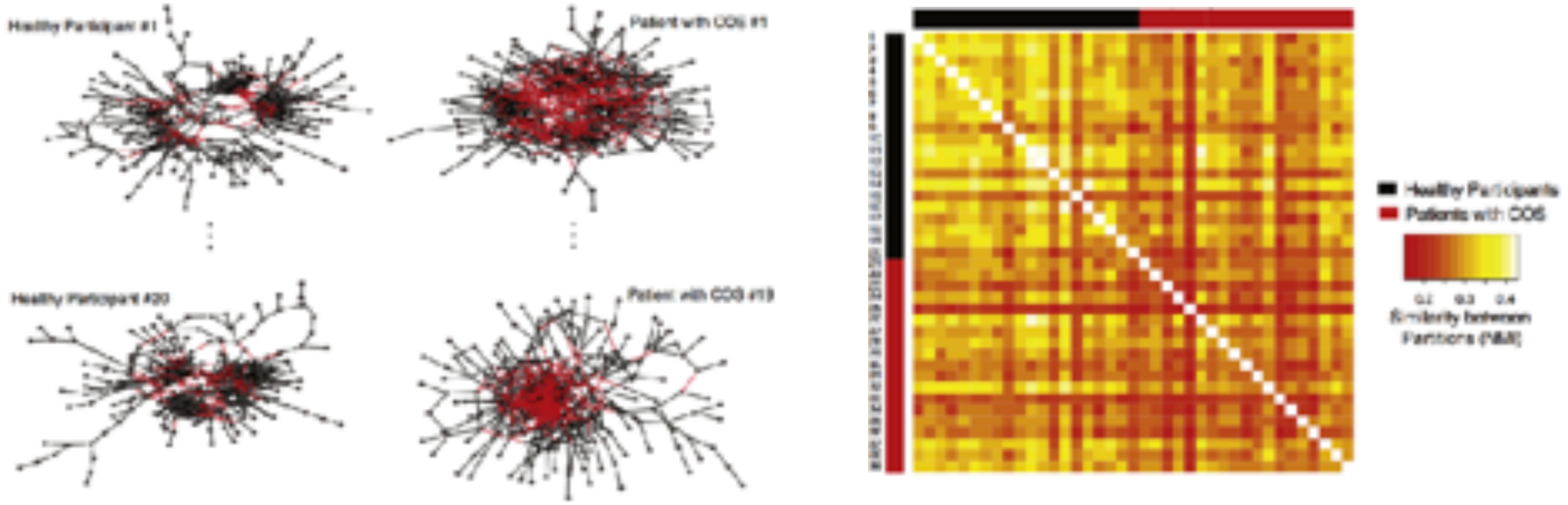
Application

Goal: discriminate patients with schizophrenia

Data: fMRI, 23 HC, 23 SZ, TR=2.3s, rest, 2x3 min (144 points)

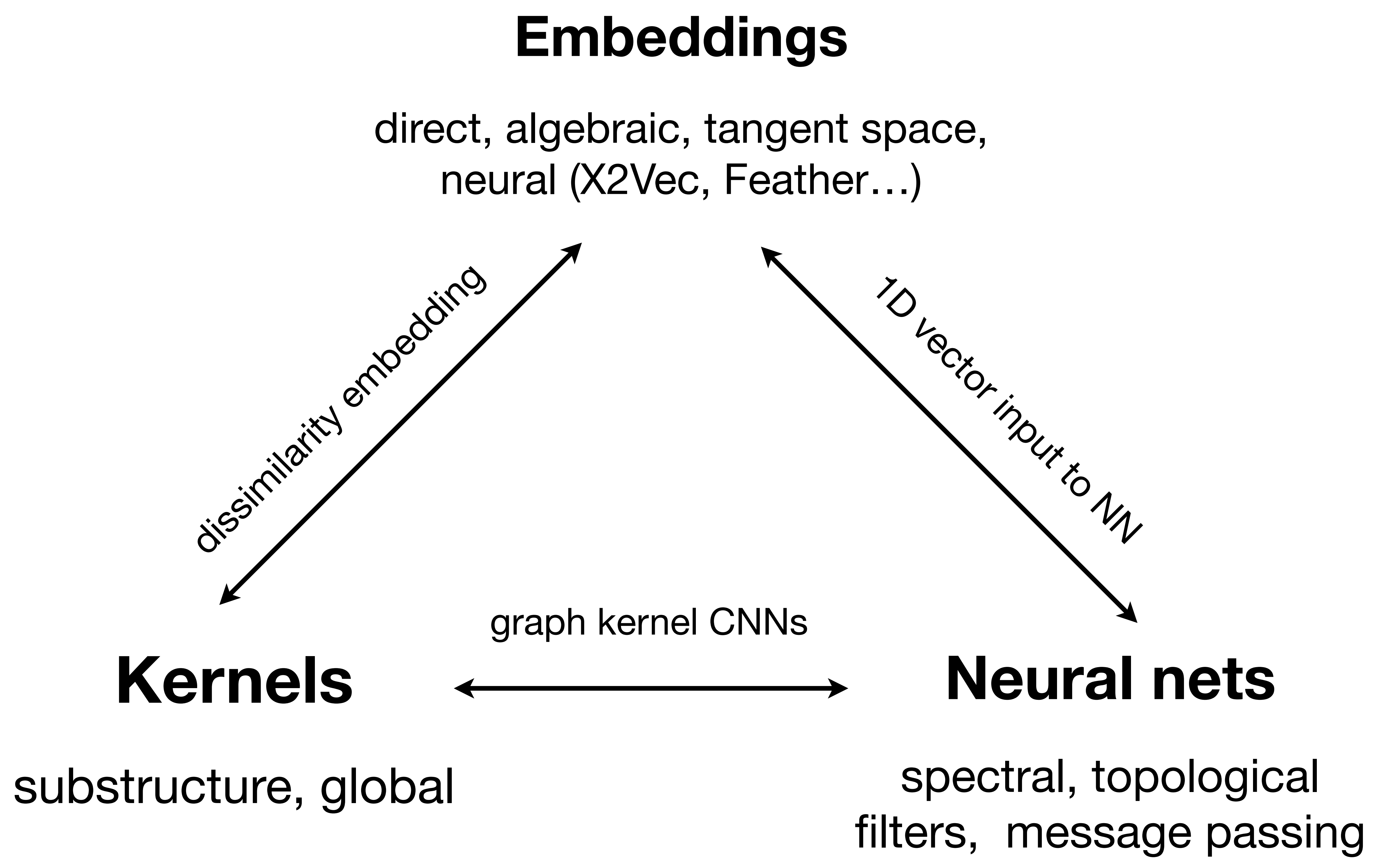
Vertices: Subparcellated Harvard-Oxford, 278 regions

Edge labels: thresholded and binarised absolute wavelet correlation, 0.05-0.1Hz



[Alexander-Bloch et al., NeuroImage, 2012]

Machine learning approaches



Benchmark of estimators, atlases, classifiers: [Dadi et al., Neuroimage, 2019]

Graph embeddings

Graph embedding maps graphs to points in $\mathbb{R}^D$

With G a set of graphs, a graph embedding $\varphi : G \to \mathbb{R}^D$
maps graphs to D-dimensional vectors:

$$\varphi(g) = (x_1 \ \ldots \ x_D)^T$$

For brain graphs, we are generally interested in preserving edge label information

Vertex labels can be dropped because of the correspondence between atlas regions

Once we have vectors we can use any ML algorithm we want

Direct embedding and behaviour/phenotype prediction

Principle

Use the upper-triangular part of **A**[1,2,3]

$$\begin{pmatrix} (1,1) & \dots & (1,|V_i|) \\ & \ddots & \\ & & (|V_i|,|V_i|) \end{pmatrix} \longrightarrow \begin{pmatrix} (1,2) \\ \vdots \\ (|V_i|-1,|V_i|) \end{pmatrix}$$

$$\mathbf{A}_i \in \mathbb{R}^{|V_i| \times |V_i|} \qquad \mathbf{a}_i \in \mathbb{R}^{\binom{|V_i|}{2} \times 1}$$

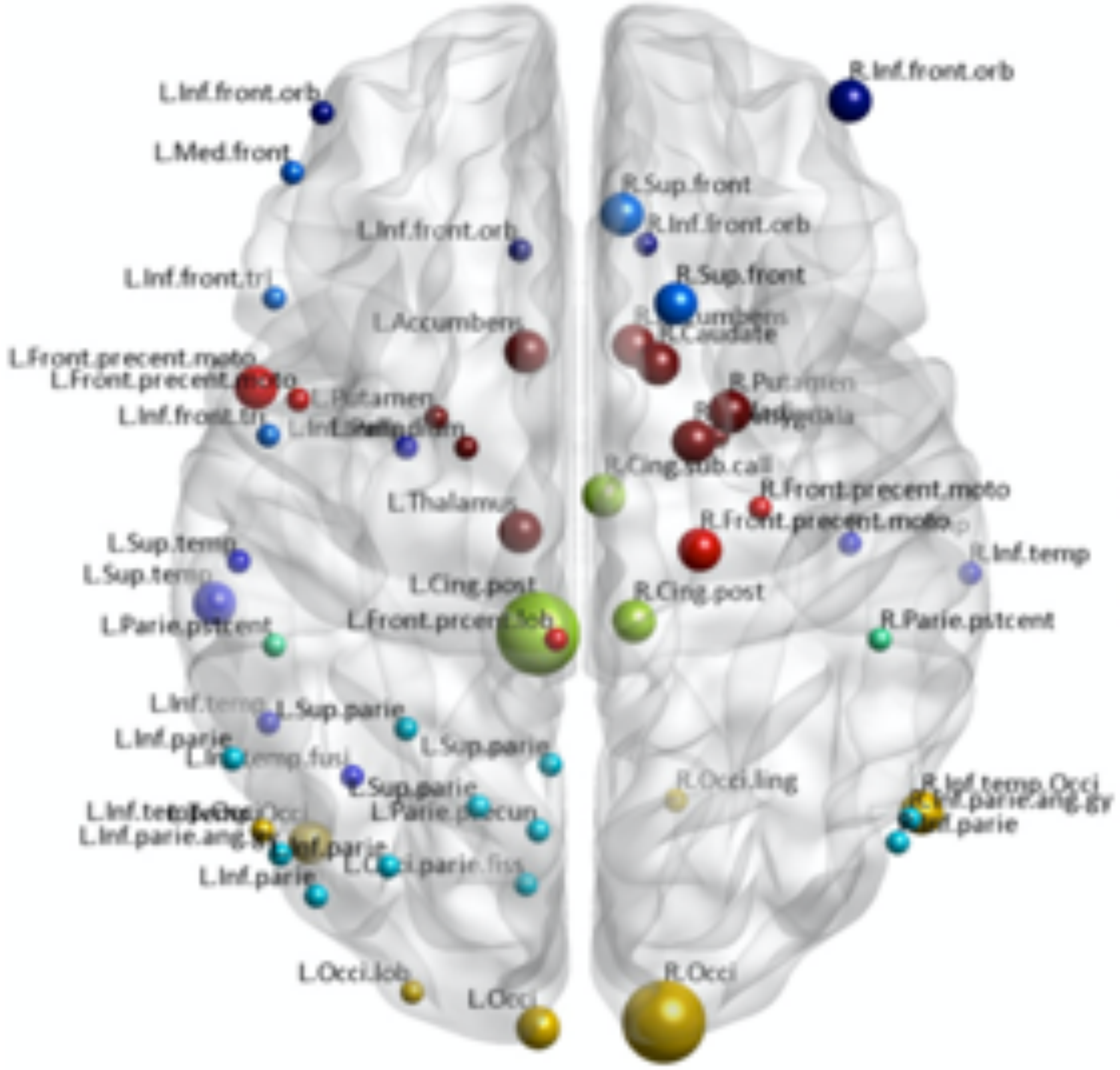
"Cursed" representation and loss of posdefness, but generally a competitive baseline (at least with ~100 vertices, fMRI)

Links Mantel test statistic with Linear kernel SVM - same notion of similarity

1 [Wang et al., MICCAI, 2006]
2 [Craddock et al., MRM, 2009]
3 [Richiardi et al., ISBI 2010]
[Richiardi et al., ICPR 2010]
[Richiardi et al., NeuroImage, 2011+12]

Application

Task: Dx, 30 depressive vs 30 matched HC
Data: task fMRI, 137 regions
Classifier: L1-SVM on direct embedding
Results: 85% acc



[Rosa et al., Neuroimage, 2015]

Graph convolutional networks and phenotype prediction

Principle[1]

layer weight matrix (trainable)

A+I

Approximate spectral filtering on graphs[2]

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)})$$

Degree matrix

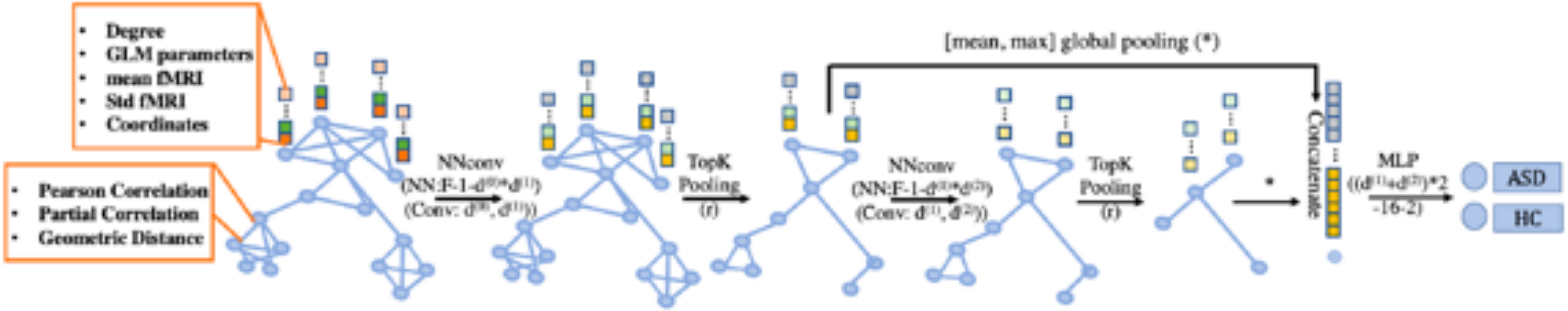
layer activations

[1][Kipf and Welling, ICLR, 2017]

[2][Cvetkovic et al., Spectra of Graphs, AP, 1980] [Luo et al., Pattern Recognition, 2003] [Deferrard et al., NeurIPS, 2016]

Application

Task: Dx, 75 ASD depressive vs 43 matched HC
Data: task fMRI, 148 regions
Classifier: GNN with edge and node attributes
Results: 76% acc

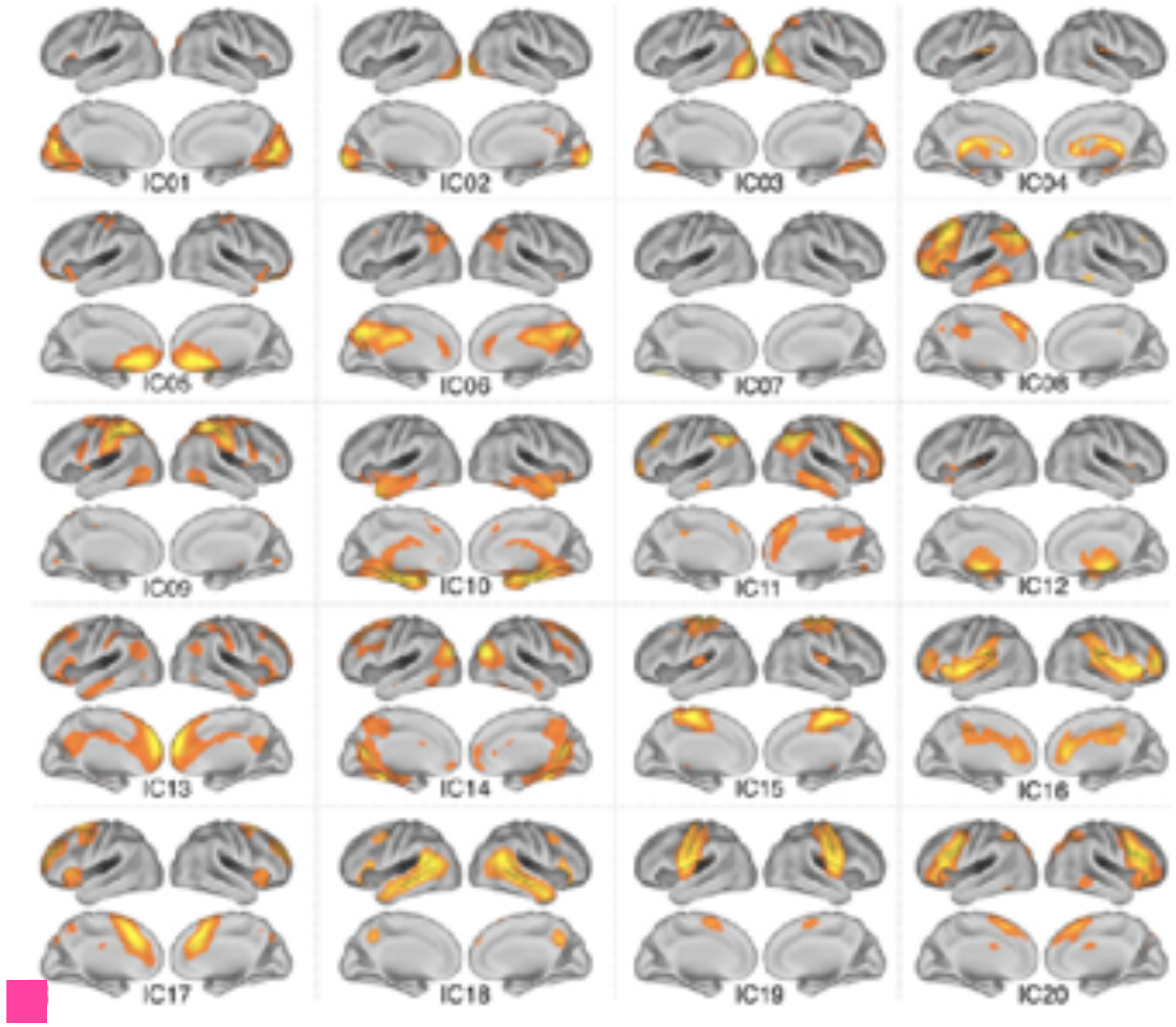


[Li et al., MICCAI, 2019]

4. Graph imaging genetics

Genetics of intrinsic brain activity

In humans and lower mammals, spatially consistent, synchronised, intrinsic activity forming "functional networks" is observed reproducibly across the lifespan



BOLD fMRI, ICA(K=20), N=1093, C=24

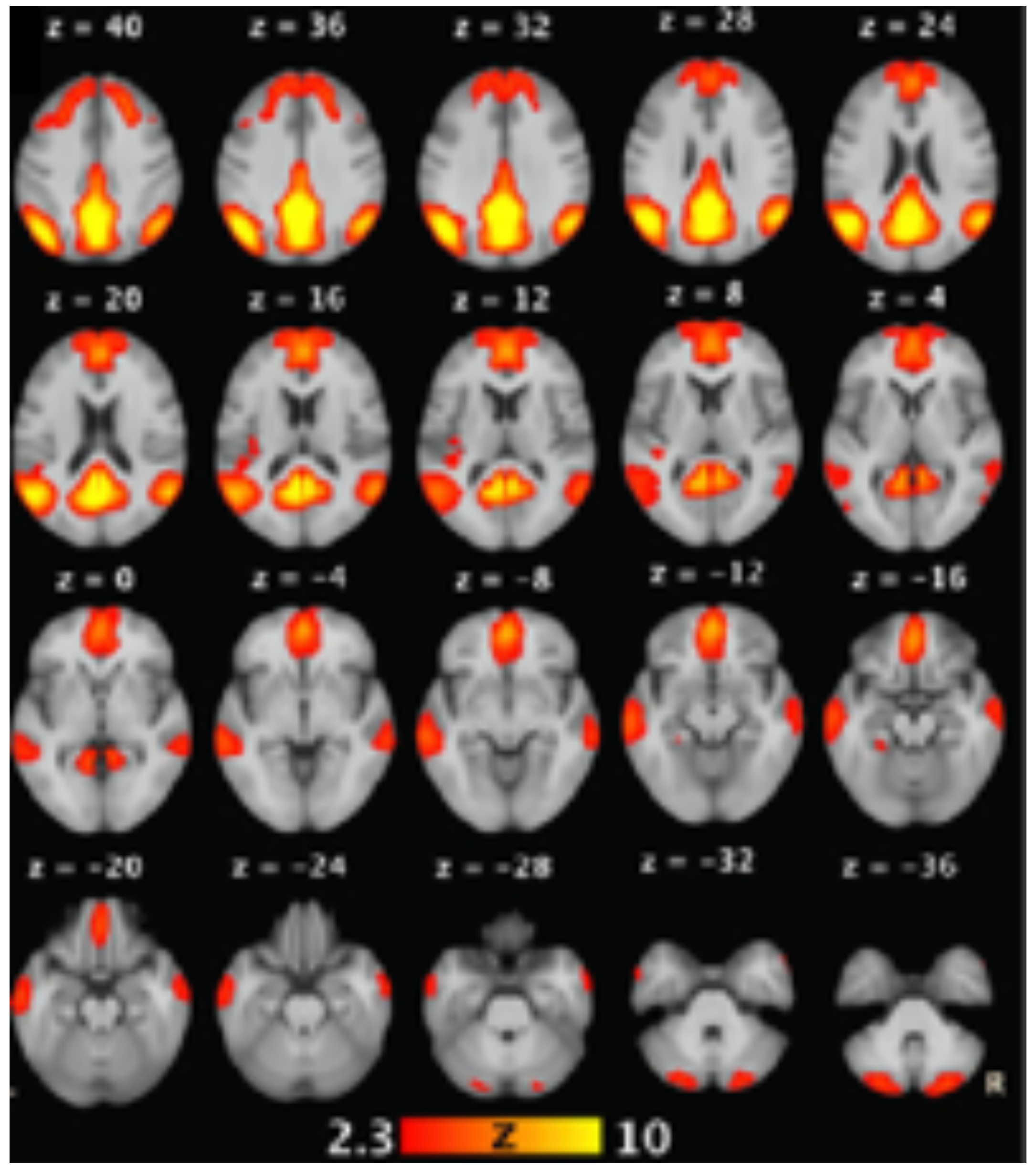
What are the genetic correlates of this consistency?

[Biswal et al., PNAS, 2010]

Functional networks are heritable

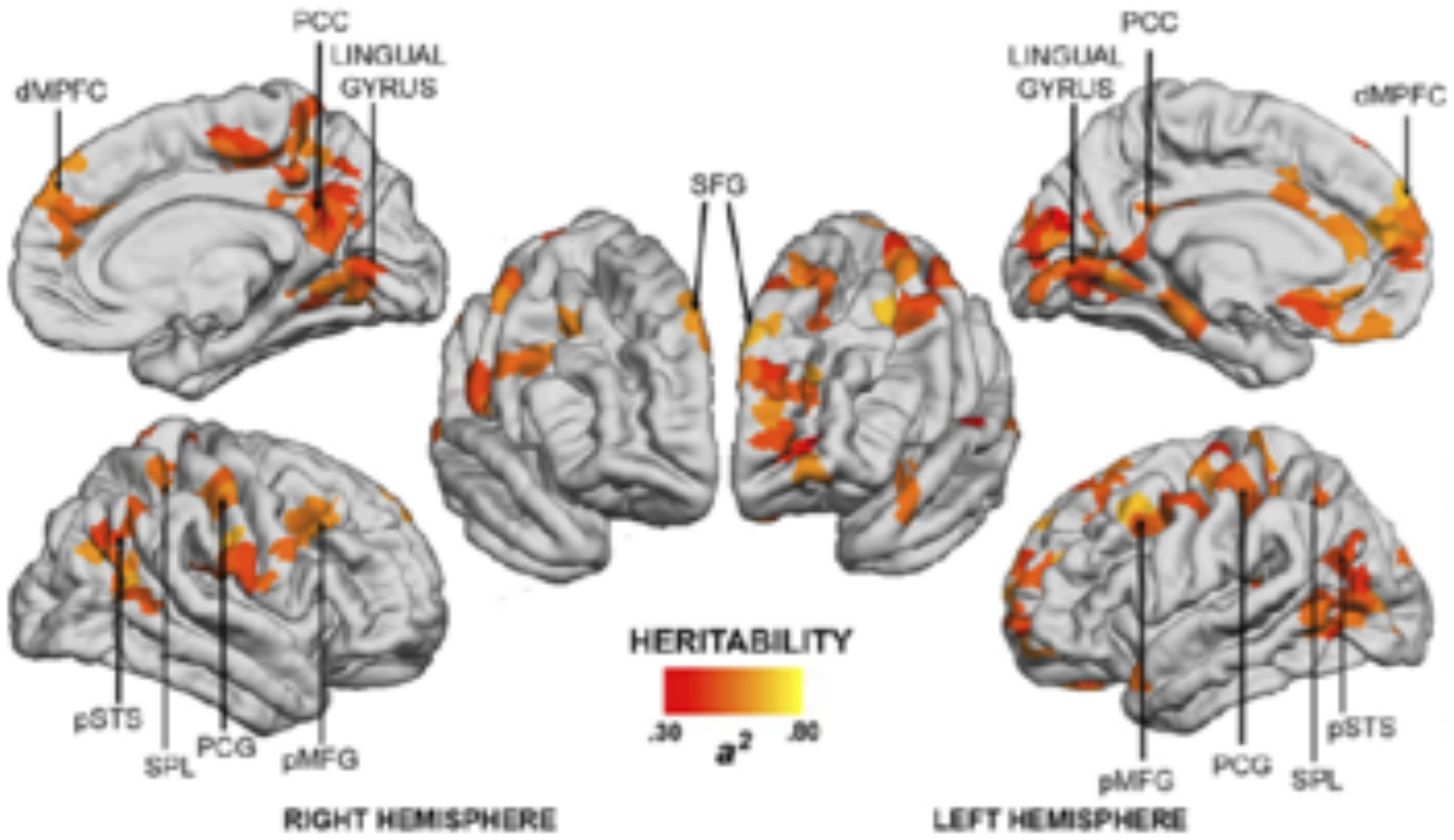
Like other neuroimaging phenotypes[1,2], some aspects of FNs are heritable

Connectivity within the DMN is heritable[3] (0.4, p=0.005)



BOLD fMRI, ICA (K=17), N=333

Some topological metrics of FNs are heritable[4] (0.6 for CE)



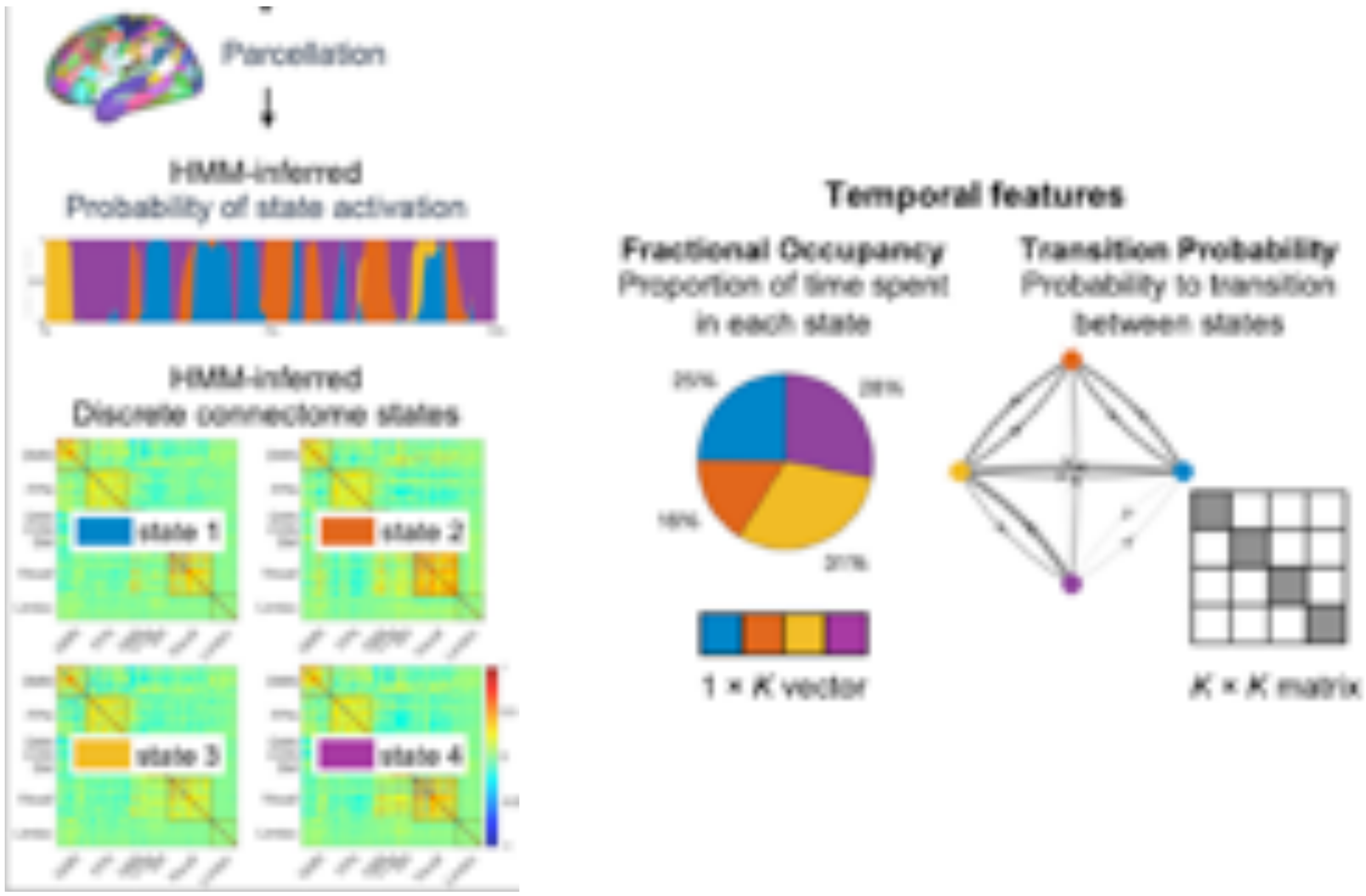
BOLD fMRI, R=1041, N=48

[1][den Braber et al., NeuroImage, 2013]

[2][Chiang et al., J. Neurosci, 2009]

[3][Glahn et al., PNAS 2010][4][Fornito et al., J. Neurosci, 2011]

Dynamics of functional networks are also heritable



N=1003 (aged 22–37, 534 females):

- 120 monozygotic (MZ) twin pairs
- 65 sex-matched dizygotic (DZ) twin pairs
- 96 sex-matched non-twin (NT) sibling pairs
- 62 pairs of sex-matched unrelated individuals.

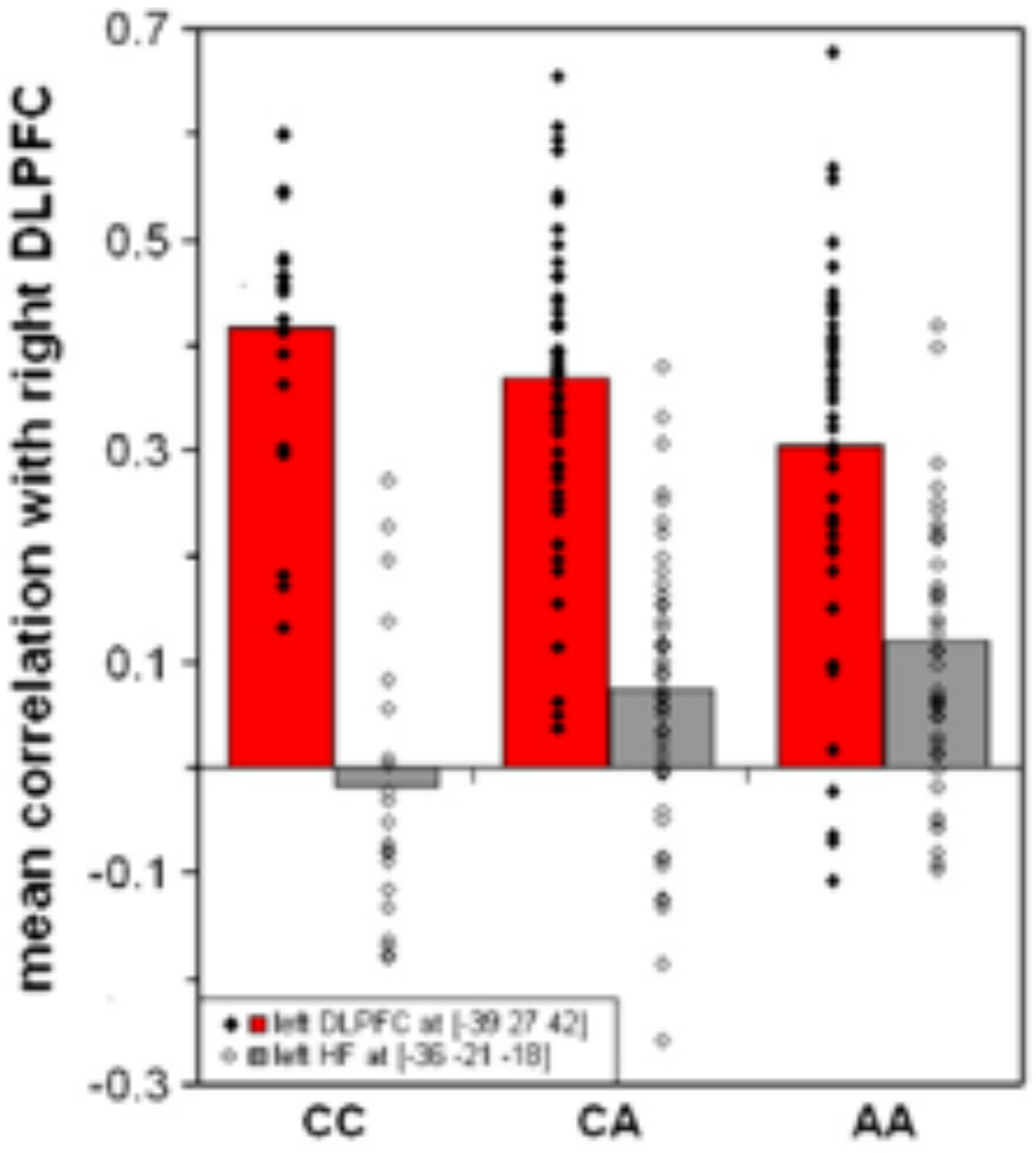
(h^2=0.39, 95% CI= [.24,.54] for FO
h^2=0.43, 95% CI=[.29,.57] for TP)

[Jun et al., NeuroImage, 2022]

SNPs modulate functional networks

Genetic risk factors for many diseases affect FNs[1]

ZNF804 / rs1344706 (SZ risk)[2]

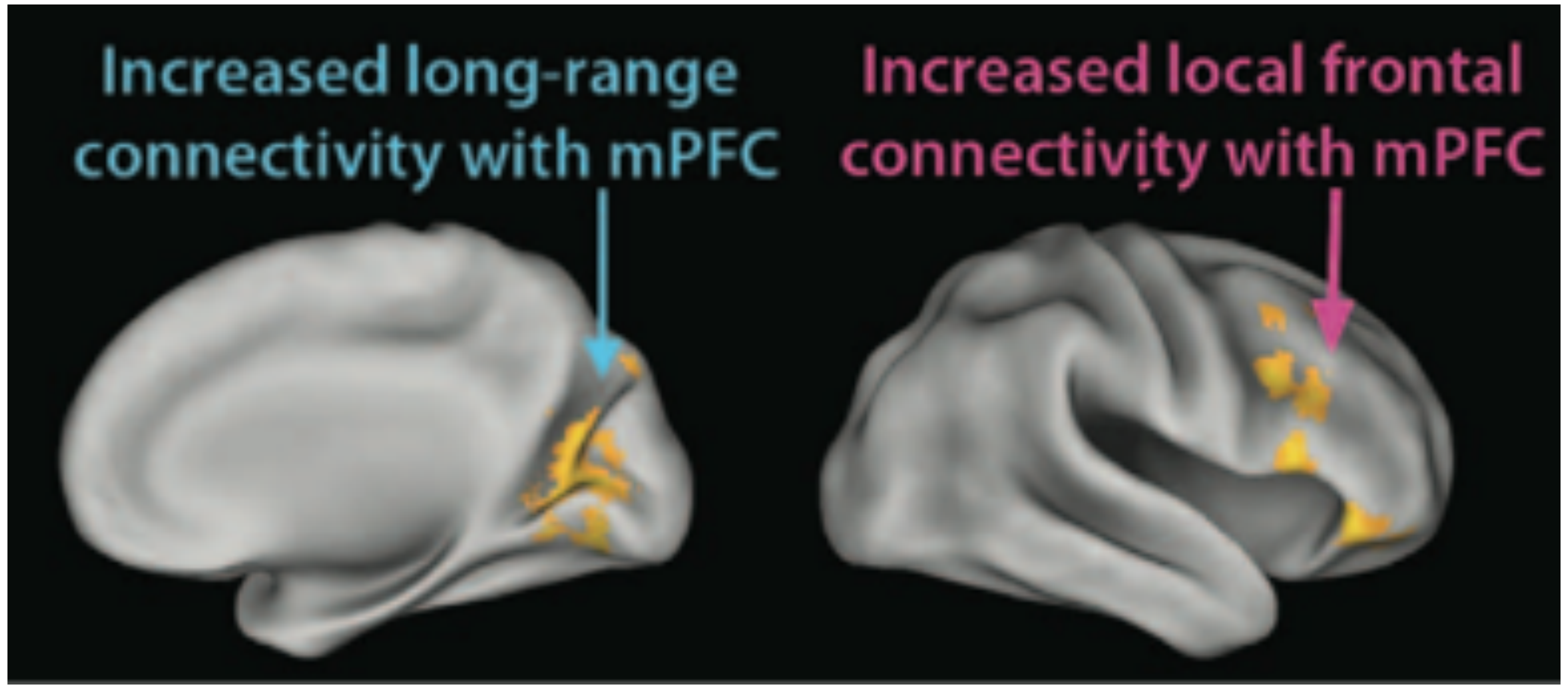


BOLD fMRI, *seed corr*, N=115 HC

CNTNAP2 / rs2710102 (ASD risk)[3]

nonrisk>risk

risk>nonrisk



BOLD fMRI, *seed corr*, N=71 (16+39 HC, 16 ASD)

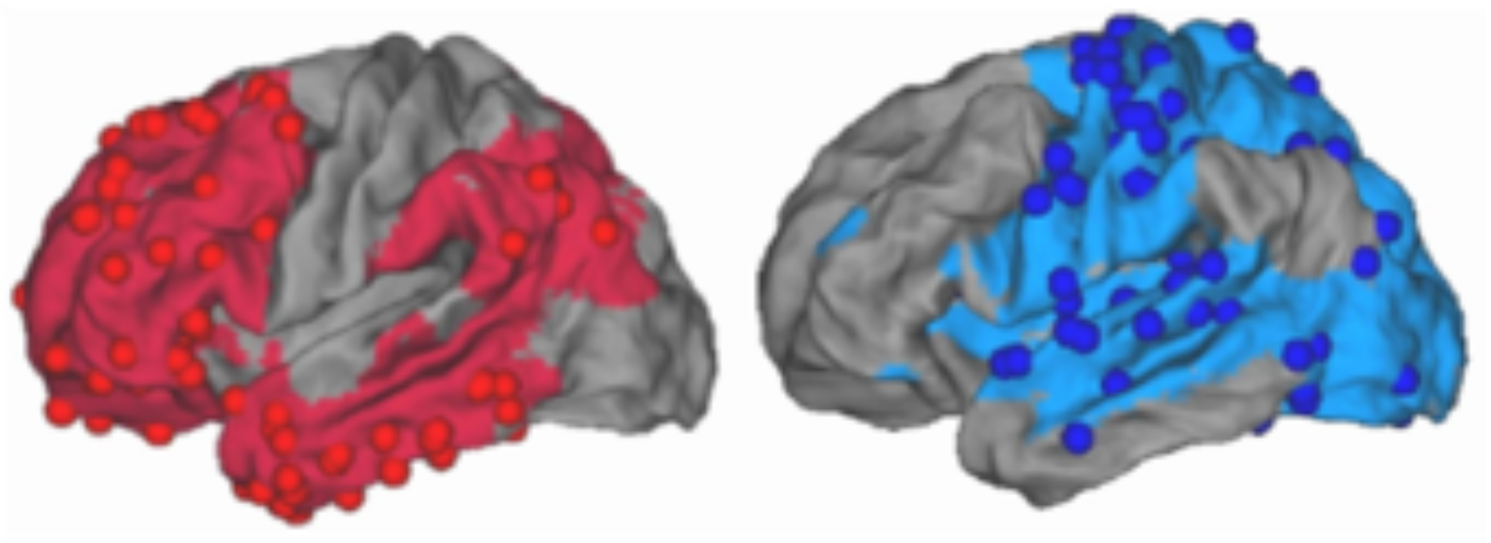
[1][Tost et al., Neuroimage, 2012] [2][Esslinger et al., Science, 2009]

[3][Scott-Van Zeeland et al., Science Transl. Med., 2010]

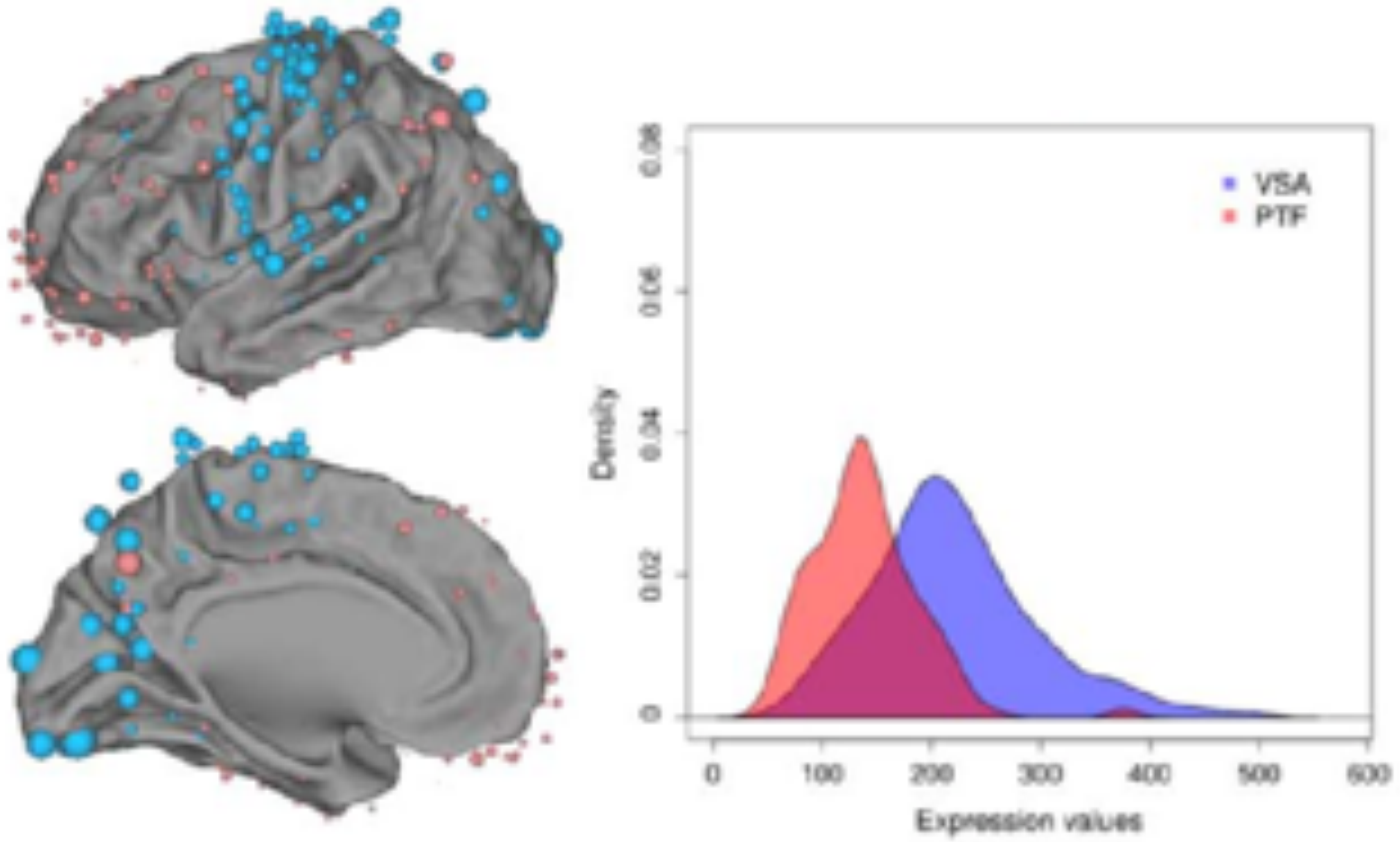
Gene expression relates to functional networks

Different (cortical) FNs[1] have different gene expression[2]

parieto-temporo-frontal   visual-sensorimotor-auditory



SCN1A



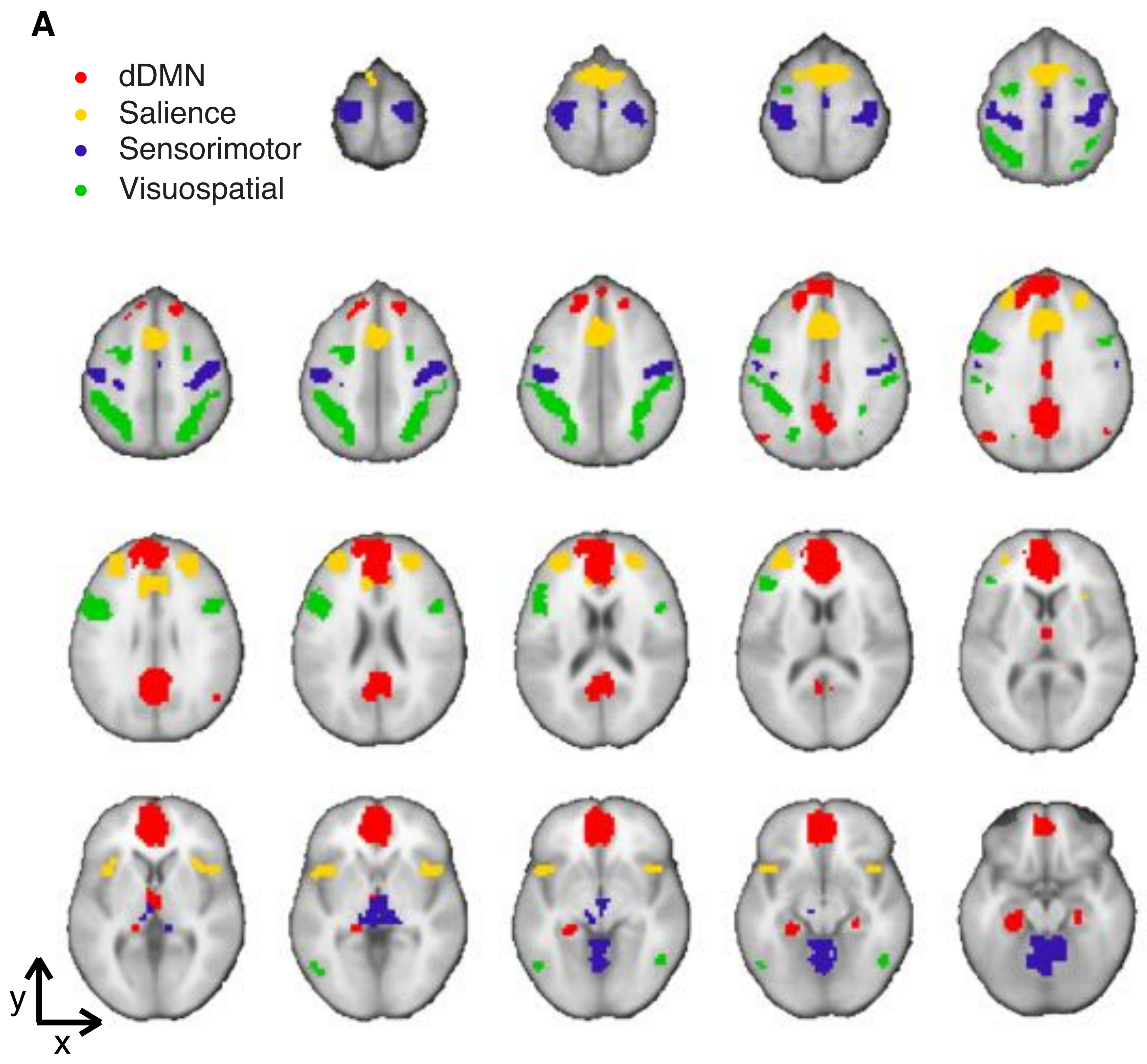
N=1 (394 samples), G=938

1[Mesmoudi et al-, PLoS one, 2013] (dual rings)

2[Cioli et al., PLoS one, 2014]

Functional networks genetics

fMRI data
(ICA-defined networks)

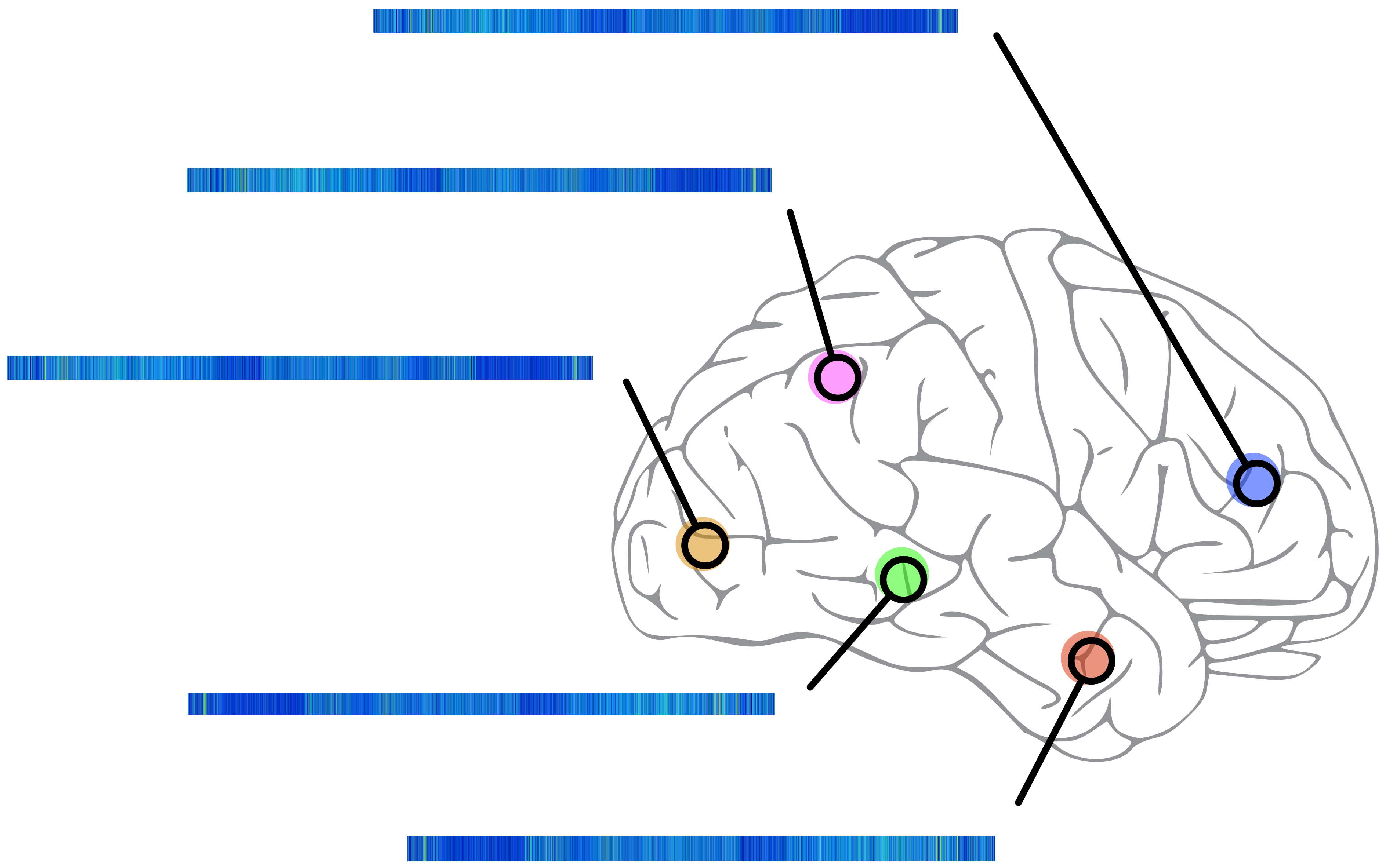


N=15, 18-29 y.o.
in vivo

[Richiardi, Altmann et al., Science, 2015]

Transcriptional similarity within functional networks

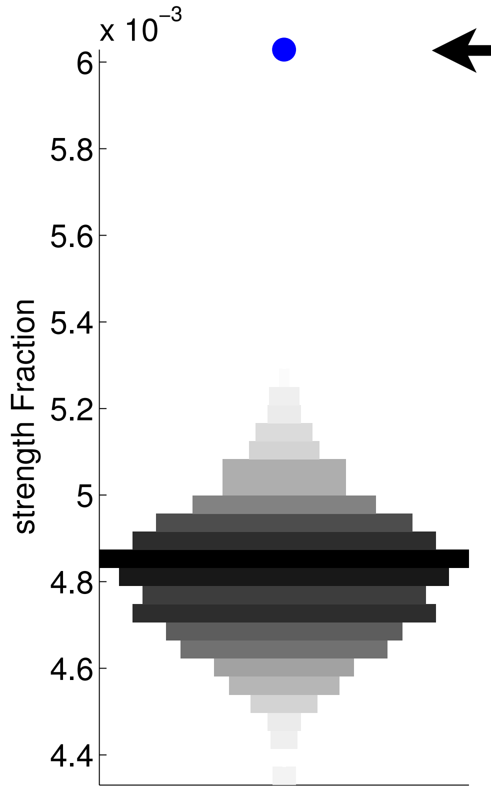
$$S_g = \frac{\sum_1}{\sum_{1+0} - \sum_1}$$



1 1 1 0 0

Gene expression per region

Regions grouped as belonging to FNs (1) or rest of brain (0)



grouped by networks

random groupings

p = 0.001

Holds with distance-corrected data, distance-preserving permutation

Genes driving the test

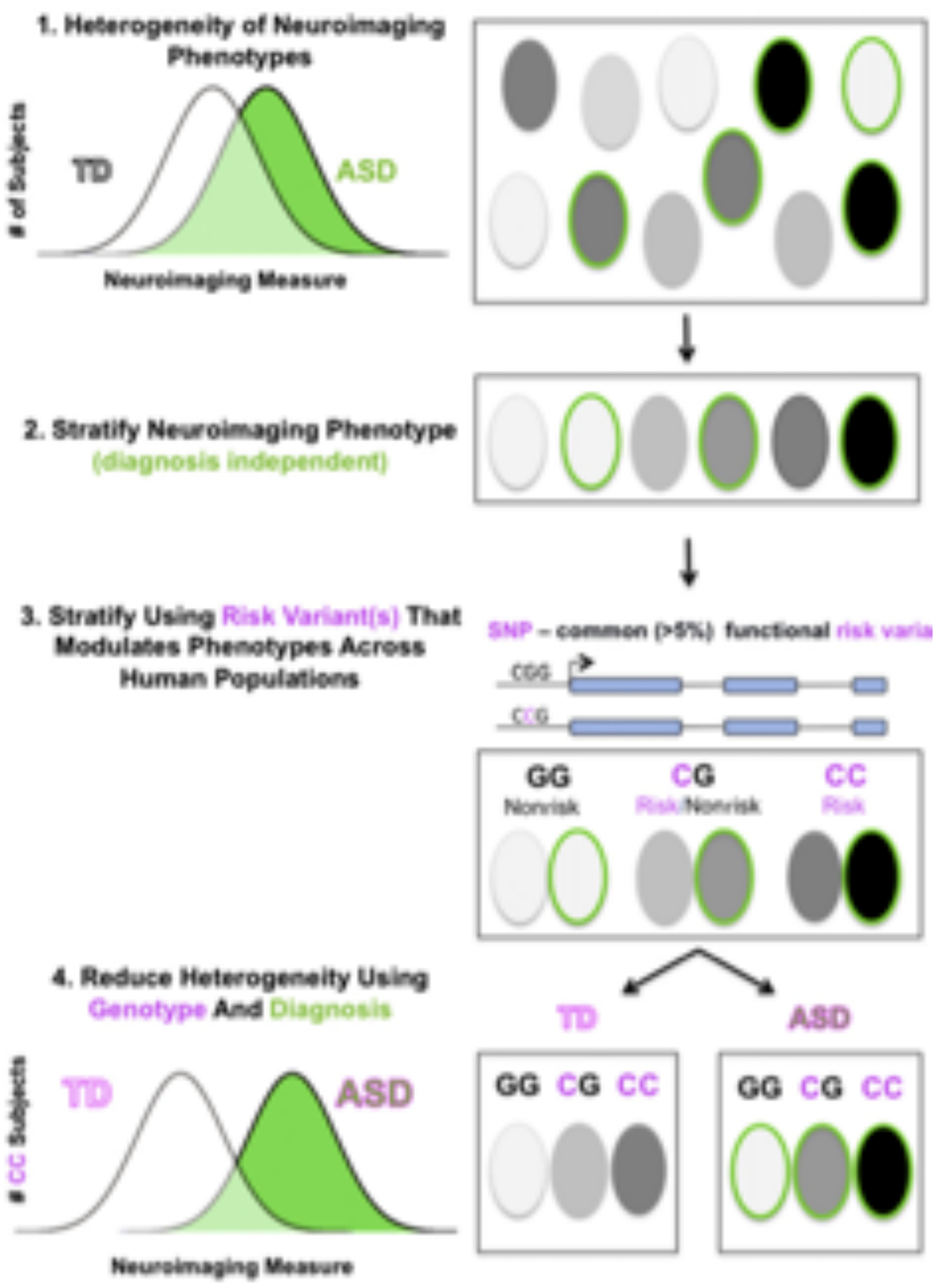
list overlap test + stability selection results

All splits (most stable)	ADAM23 ANKRD6 ATP6V1C2 BAIAP3 C3orf55 CARTPT CCDC39 CD70 CDK1 CNTN6 CRYBA2 CTXN3 CXXC11 DMRT3 EPN3 FEZF1 FZD7 GAL **GLRA3** GNA14 GNGT2 GRP HSD11B1 KANK4 **KCNA1 KCNA3 KCNA5 KCNC1 KCTD15** KRT1 KRT31 LAIR2 LINC00238 LMOD3 LRRC38 LYPLA2 MGP MYH7 MYLK3 NEB **NECAB2** NEFH NEUROD6 NGFR NOL4 NOV NRP1 ONECUT2 PCP4 PIRT PNMT PRR15 PRSS35 PTGS1 RBP4 RBPMS2 RHOBTB2 RSPH9 SCARA5 **SCN1B SCN4B** SEMA7A SHD SHISA9 SIX3-AS1 SLC16A6 SLC22A10 SLN SV2C SYT10 **SYT2** TDO2 TGFBI TINCR TLX2 TNNT2 TRIM29 TSHZ3
5/6 splits	AMDHD1 ASGR2 CD163L1 COL5A2 CYP2C18 FAM163A **GABRA5** GALP GPR26 GPR88 GPX3 HPCAL1 IL13RA2 ISCU NEXN NKAIN4 NPBWR2 ONECUT3 OR51E2 PLCH1 **PVALB** **SNAP25** SPHKAP TMEM52 TSPAN8 ZCCHC18
4/6 splits	ALOX12 CALB1 **CCBE1** CD6 CDR2L CPLX1 ENPP6 GMPR GOLT1A GPR20 HOXD1 HPCA IL33 IQCJ KLK1 KLK8 LGR6 LINC00617 MS4A8 MYBPC1 NUPR1L PYDC1 RTP1 SEMA3C SH3RF2 SLC16A5 SLC20A2 SLC39A12 SOST WISP1 WISP2 WNT4

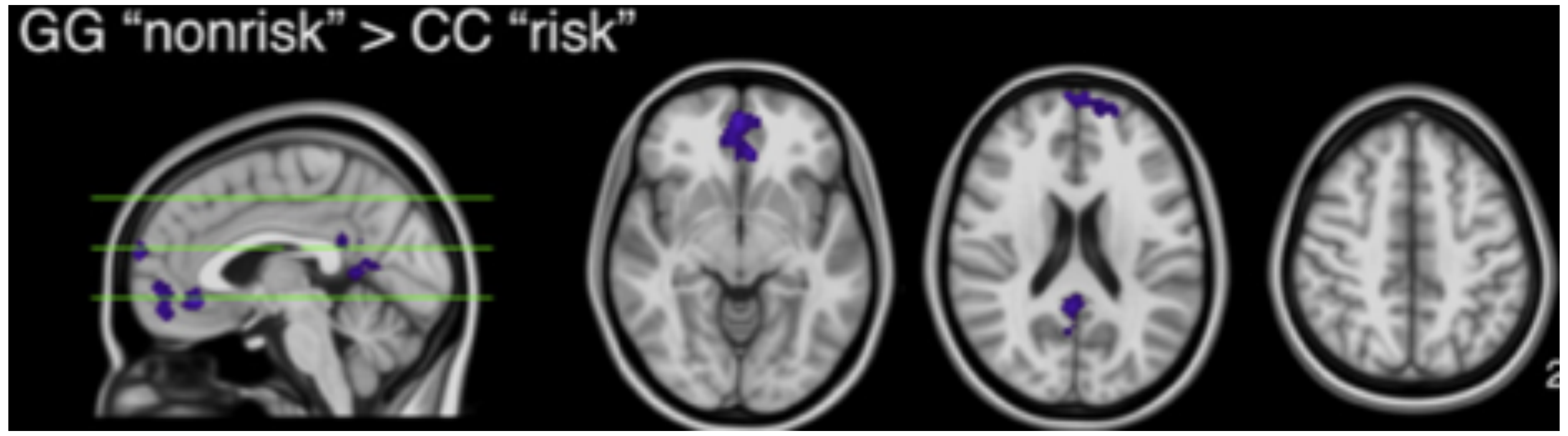
validated (n=259) in association study using SNPs of these genes with in-vivo connectivity as quantitative phenotype.

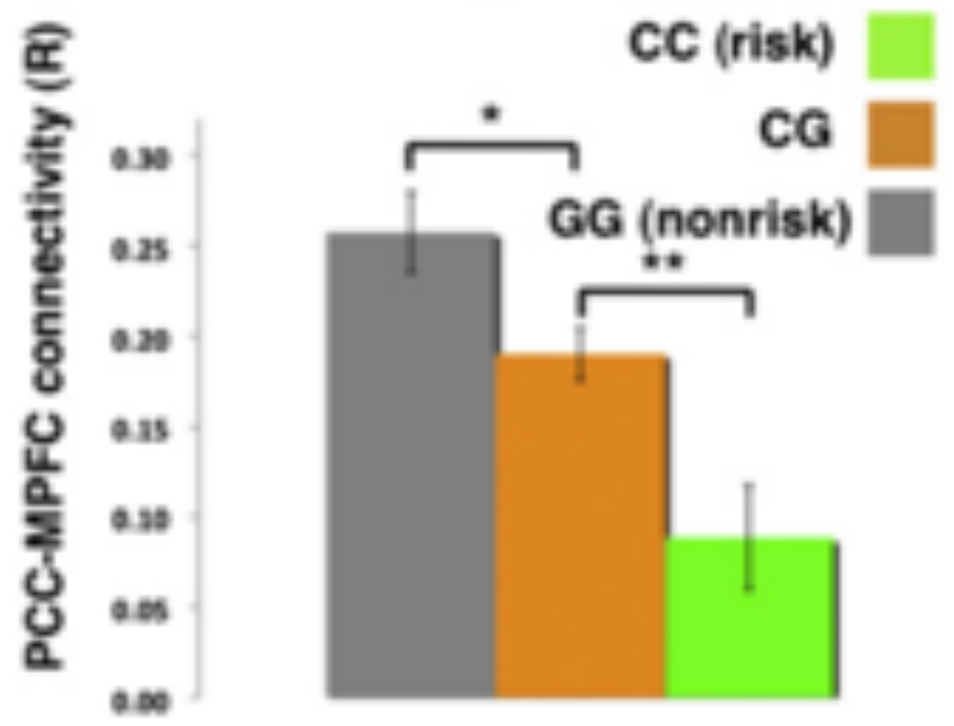
A few well-known genes like SNAP25 or GABRA5
Many potassium channels (KCN*)
Significant enrichment for voltage-gated ion channels

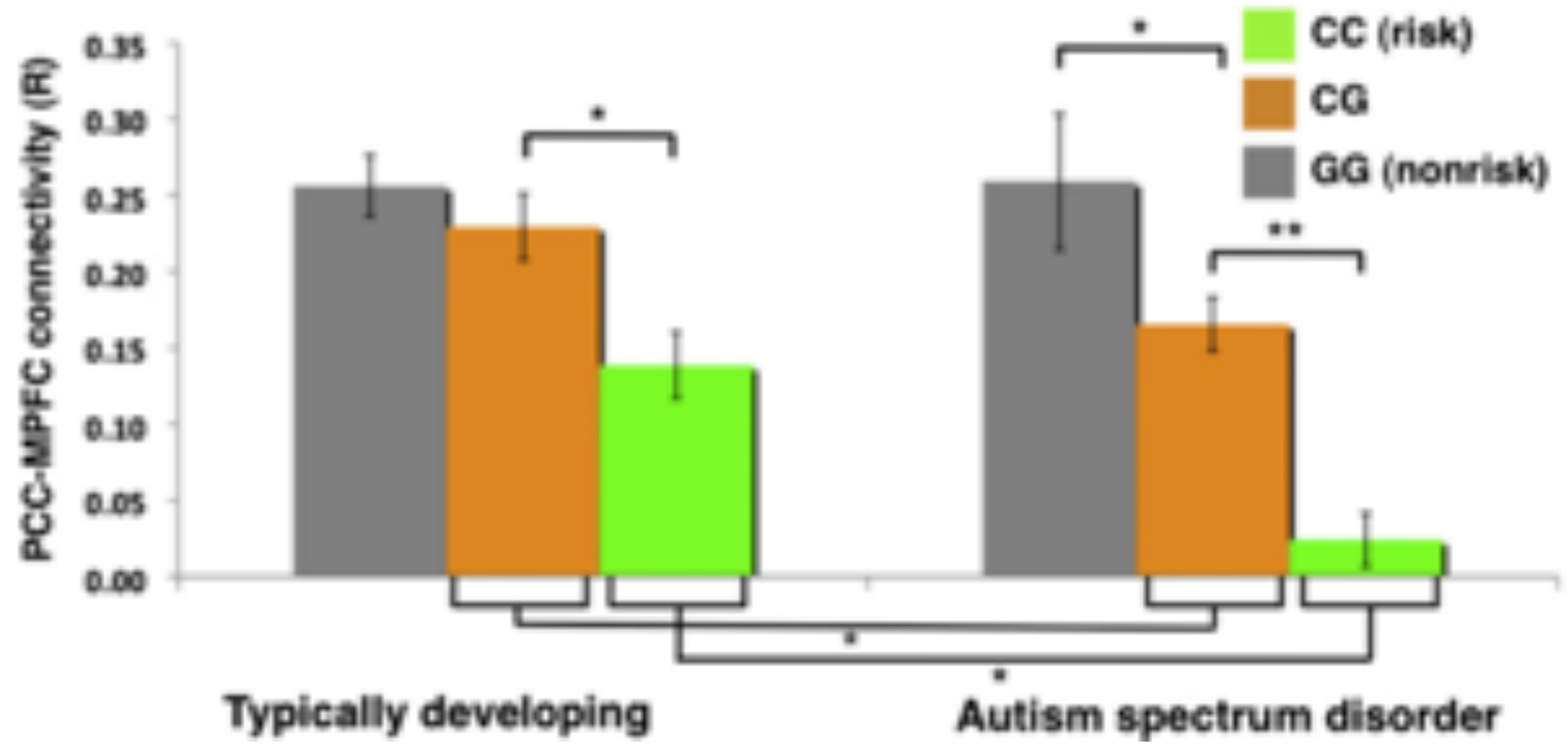
Brain network imaging genetics for stratification



MET genotype: CC lowers PCC↔MPFC conn







	CC	CG	GG	∑
TD	9	15	9	33
ASD	7	24	7	38
∑	16	39	16	**71**

[Rudie et al., Neuron, 2012]

Wrap-up and take-home points

Graphs offer an interpretable ‘meso-scale’ for medical image analysis, and a useful ‘inter-lingua’ between different modalities, organs, and biological scales

Graph estimation from medical imaging data can be improved from the status quo, yielding more desirable estimators

There has been tremendous progress over the past 10 years in graph analysis, particularly in scalable and predictive ML approaches

Graphs obtained from medical imaging, and their properties, can be used as a heritable and sensitive endophenotypes for genetic analyses

Thanks

Translational Machine Learning Lab

back Tommaso Di Noto**, Antoine Madrona, Xavier Sieber***, Dr. Jonathan Patiño-Lopez, Jonas Richiardi
front Veronica Ravano*, *Dr Elda Fischi-Gomez*, Costa Georgantas, Dr Jaume Banus Cobo, (Adrien Jammot)

Collaborators

INRIA/CNRS/UGA
DrSc. Sophie Achard
Prof. JF Coeurjolly
Ms Hana Lbath
UNSW
Prof. P. Lafaye de Micheaux
BYU/UCSB
Prof. A. Petersen
UCL
DrSc Andre Altmann
Stanford
Prof. M. Greicius
CHUV/UNIL
DrSc Ruud van Heeswijk
Prof. Roger Hullin
HUG/UNIGE
Dr Philippe Meyer

Funding

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Innosuisse – Swiss Innovation Agency

Swiss National Science Foundation

CHUV

anr agence nationale de la recherche
AU SERVICE DE LA SCIENCE

NSF

co-supervisions: *Siemens Healthcare, **UNIL/MIAL, **CHUV/Connectomics Lab, ***CHUV/CVMR