HCP1200 July 2017 release of high-level rfMRI connectivity analyses

Summary

Resting-state fMRI (rfMRI) data from the 2017 HCP data release was processed, yielding the following outputs:

1. **Group-PCA** – group-average PCA spatial maps (weighted eigenmaps)
2. **Group Average Dense connectomes** - grayordinate-wise functional connectome derived via the group-PCA
3. **PTN (Parcellation, node Timeseries and Netmats)** data:
    a. **Parcellations** - group-average "parcellations", obtained by means of group-ICA.
    b. **Timeseries** - subject-specific sets of "node timeseries" - for each subject, one representative time series per ICA component.
    c. **Netmats** - subject-specific "parcellated connectomes" – for each subject, a nodes x nodes network matrix or "netmat".
4. **Subject-specific parcellations** (spatial maps), derived from all of the above.

All matlab code used to run the analyses described below is provided as part of the PTN download.

Figure 1 (below) shows a useful schematic to understand the relationship between the different rfMRI datasets in this analysis.

In addition to the descriptive text and citations below, which may be useful when writing papers that take advantage of these "higher-level" HCP outputs, please remember to include the generic HCP acknowledgements (and core HCP citations) when using HCP data in your research: see http://www.humanconnectome.org/study/hcp-young-adult/document/hcp-citations.

Data used

From the 1200-subjects 2017 HCP data release (March 2017), the following analyses used the data from all 1003 subjects having four complete rfMRI runs (4800 total timepoints).

Data from the earliest 184 subjects was reconstructed (from the raw, complex, multi-coil, multiband, k-space data) using an initial version of the data reconstruction software (referred to here as "recon1", and identified in the "unrestricted variables" file available from HCP by variable *fMRI_3T_ReconVrs* having value *r177*). Data from the latest 812 subjects used a later, slightly improved version ("recon2", with variable *fMRI_3T_ReconVrs* having value *r227*). Data from 7 subjects was processed using a mixture of the two methods. A full explanation of this change is given on the HCP website: https://wiki.humanconnectome.org/display/PublicData/Ramifications+of+Image+Reconstruction+Version+Differences

The appended slides report on an investigation into the effect (on resting-connectivity) of this change in reconstruction software. Although in many respects the changes in connectivity measures are not large, they may be significant enough to warrant the inclusion of a "recon version" confound regressor in cross-subject analyses of functional connectivity. Additionally, because some researchers may want to take the most conservative approach of only working with recon2 subjects, the group-average PCA outputs, dense connectome, group-ICA parcellations and PTN data release are all available in two versions – one using all subjects from both recon versions (**HCP1200_Parcellation_Timeseries_Netmats.zip**), and a slightly smaller version using only the 812 recon2 subjects (**HCP1200_Parcellation_Timeseries_Netmats_recon2.zip**).

The various connectivity analysis outputs described below are available for download from ConnectomeDB at https://db.humanconnectome.org. One can view the groups of subjects included in these analyses using the "Explore Subjects" or "Open group" functions in ConnectomeDB.

Data pre-processing and group-PCA

Each 15-minute run of each subject's rfMRI data was preprocessed according to [Smith 2013]; it was minimally-preprocessed [Glasser 2013], and had artefacts removed using ICA+FIX [Salimi-Khorshidi 2014, Griffanti 2014]. Inter-subject registration of cerebral cortex was carried out using areal-feature-based alignment and the Multimodal Surface Matching algorithm ('MSMAll') [Robinson 2014, Glasser 2016]. For feeding into group-PCA, each dataset was then temporally demeaned and had variance normalisation applied according to [Beckmann 2004].

Group-PCA output was generated by MIGP (MELODIC's Incremental Group-PCA) from 1003 (recon 1 + recon 2) and 812 (recon 2) subjects. This output comprises the top 4500 weighted spatial eigenvectors (eigenmaps) from a group-averaged PCA (a very close approximation to concatenating all subjects' timeseries and then applying PCA) [Smith NeuroImage 2014]. Each eigenmaps file (2GB each) were used as input to group-ICA and also to generate the group-average dense connectome for each group:
HCP_S1200_1003_rfMRI_MSMAll_groupPCA_d4500_Eigenmaps.dtseries.nii (group average PCA eigenmaps from all recon1 + recon2 subjects)
HCP_S1200_812_rfMRI_MSMAll_groupPCA_d4500_Eigenmaps_recon2.dtseries.nii (group average PCA eigenmaps from recon2 subjects).

Dense connectome

The group-PCA outputs (eigenmaps referred to above) are renormalized, eigenvalue-reweighted and correlated to form the dense connectome (grayordinatesXgrayordinates functional connectivity), a low-noise regularized equivalent to concatenating every subject's grayordinate-timeseries rfMRI data and correlating each grayordinate's (huge) timeseries with each other grayordinate's timeseries. To carry this out optimally is not trivial, but the code for completing these steps is included in each of the main PTN release archives; for more technical details see the mound-and-moat document at https://www.humanconnectome.org/documentation/mound-and-moat-effect.html. The dense connectome files (33GB each) available from ConnectomeDB are:
HCP_S1200_1003_rfMRI_MSMAll_groupPCA_d4500ROW_zcorr.dconn.nii (group average dense connectome from all recon1+ recon2 subjects)
HCP_S1200_812_rfMRI_MSMAll_groupPCA_d4500ROW_zcorr_recon2.dconn.nii (group average dense connectome from recon2 subjects)

If you prefer to view the dense connectome as seed maps in the wb_view visualization software distributed as part of Connectome Workbench (recommended to avoid the large file size downloads), *you do not need to download it*; instead, there are 2 other options (both require internet connection and a ConnectomeDB login):

1. Download the 1200 Subjects Group Average Group Average Workbench dataset available in ConnectomeDB and open the included

HCP_S1200_GroupAvg_v1.scene file to view in wb_view. The connectivity data for an indicated seed is automatically loaded when you open the scene (scene 3) that includes the full correlation data. Further instructions are in the Group Average Tutorial pdf included in the download.

2. In wb_view with appropriate surfaces and/or volumes loaded (e.g. the S1200 group average very inflated surfaces and the S1200_AverageT1w_restore.nii.gz volume), load the following URL(s) in using the File>Open Location>Custom option:
https://db.humanconnectome.org/spring/cifti-average?resource=HCP_Resources:GroupAvg:HCP_S1200_1003_rfMRI_MSMAll_groupPCA_d4500ROW_zcorr.dconn.nii
https://db.humanconnectome.org/spring/cifti-average?resource=HCP_Resources:GroupAvg:HCP_S1200_812_rfMRI_MSMAll_groupPCA_d4500ROW_zcorr_recon2.dconn.nii
Set the top layer in the wb_view Overlay Toolbox to the cifti-average?resource….dconn.nii file you want to view and click a seed on the surface or subcortical gray matter to view the full correlation connectivity data for that seed. The map at the identified seed is dynamically downloaded from ConnectomeDB as you click.

Parcellations: group-ICA

The MIGP group-PCA output was fed into group-ICA using FSL's MELODIC tool [Hyvärinen 1999, Beckmann 2004], applying spatial-ICA at several different dimensionalities (15, 25, 50, 100, 200, 300). The dimensionality determines the number of distinct ICA components; a higher number typically means that the significant areas within the spatial component maps will be smaller. Spatial-ICA was applied in grayordinate space (surface vertices plus subcortical grey matter voxels, [Glasser 2013]). The spatial-ICA maps are released as a separate 'dscalar' grayordinates file (CIFTI format) for each distinct ICA decomposition (thus each file contains between 15 and 300 spatial maps). Volumetric MNI152 3D-space versions of these maps were also generated, primarily for display purposes. The sets of ICA maps can be considered as "parcellations", though they lack some properties often assumed for parcellations - for example, ICA maps are not binary masks but contain a continuous range of "weight" values; they can overlap each other; and a given map can include multiple spatially separated peaks/regions. Group parcellations utilizing other parcellation methods will be released in the future.

Node timeseries (individual subjects)

For a given "parcellation" (group-ICA decomposition), the set of ICA spatial maps was mapped onto each subject's rfMRI timeseries data to derive one representative timeseries per ICA component (for these purposes we consider each ICA component as a network "node"). For each subject, these 15 (or 25, 50, 100, 200 or 300) timeseries can then be used in network analyses, as described below. One method was used to estimate the node-timeseries - the standard "dual-regression stage-1" approach, in which the full set of ICA maps was used as spatial regressors against the full data, estimating one timeseries for each ICA map [Filippini 2009].

Node timeseries – saved as ASCII text files - are estimated for the 1003 subjects having complete rfMRI data (4800 total timepoints).

Network matrices (individual subjects and group-averaged)

Network-matrices (also referred to as "netmats" or "parcellated connectomes") were derived from the node-timeseries. For each subject, the *N* (15-300) node-timeseries were fed into network modelling, creating an *N*x*N* matrix of connectivity estimates. Network modelling was carried out using the FSLNets toolbox (fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLNets). Netmats were estimated for the 1003 subjects (complete timeseries) described above. We applied network modelling in two ways:

1. netmats1: Using "full" normalized temporal correlation between every node timeseries and every other. This is a common approach and is very simple, but it has various practical and interpretational disadvantages [Smith 2012].
2. netmats2: Using partial temporal correlation between nodes' timeseries. This aims to estimate direct connection strengths better than achieved by full correlation. To slightly improve the estimates of partial correlation coefficients, a small amount of L2 regularization is applied (setting rho=0.01 in the Ridge Regression netmats option in FSLNets) [Smith OHBM 2014, FSLNets].

Netmat values were Gaussianised from Pearson correlation scores (r-values) into Z-stats, and are released for individual subjects, as well as group-averaged over all subjects. The average netmats are provided as "pconn" files, readable by HCP Connectome Workbench software (and directly viewable in the wb_view workbench display tool). The individual subjects' netmats are saved as raw text files, with one row per subject; each row contains the *N*x*N* matrix of connectivity estimates, unwrapped to a long single row of N^2 values. The row ordering matches the list of subject IDs saved in the file subjectIDs.txt

Subject-specific parcellations (node maps)

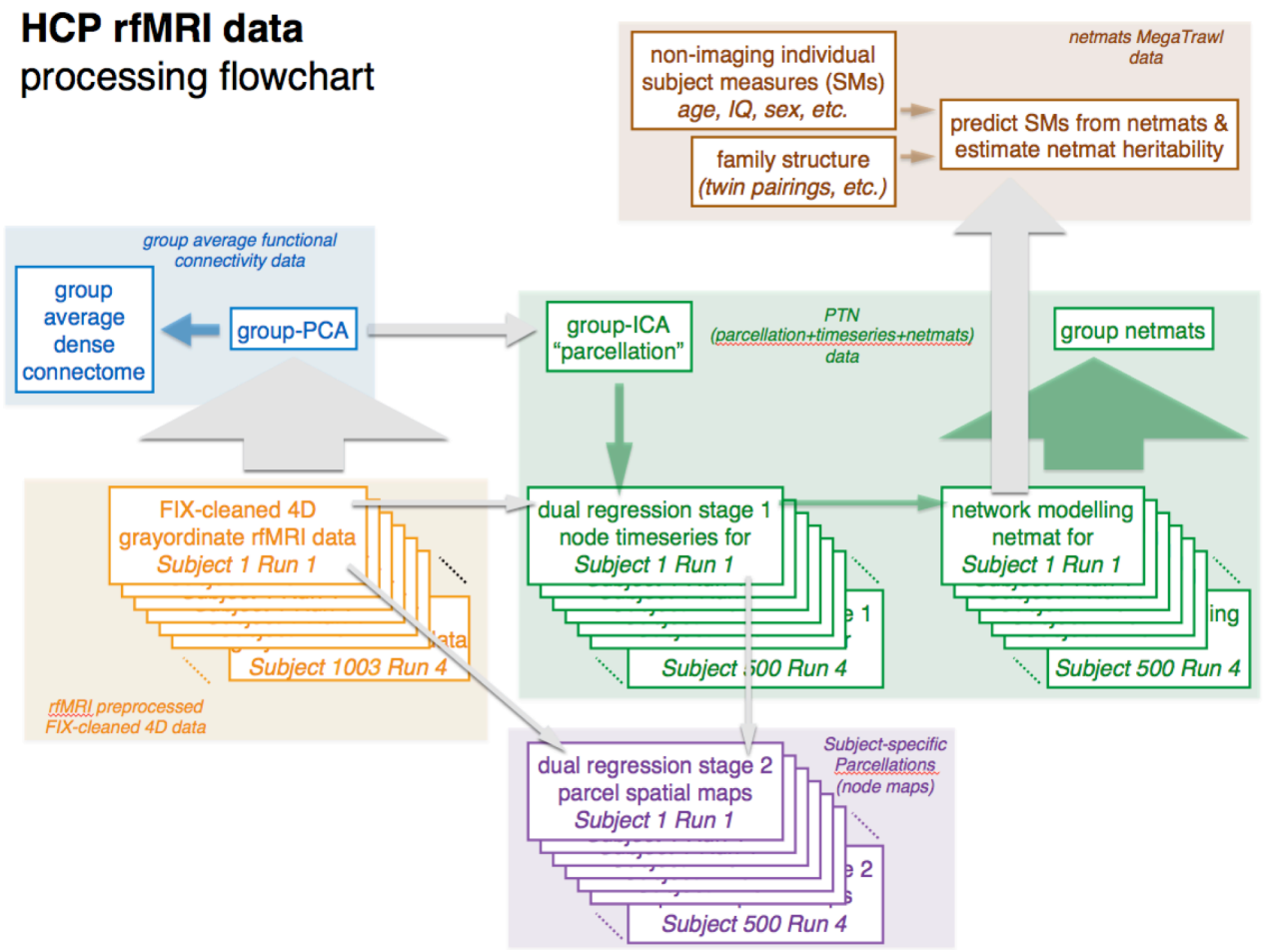
For a given "parcellation" (group-ICA decomposition), the set of group-ICA spatial maps was mapped onto each subject's rfMRI data to derive one representative subject-specific spatial map per ICA component. This is therefore a subject-specific version of the group-ICA parcellation. This is achieved by regressing a subject's node timeseries into the subject's rfMRI grayordinate-wise timeseries data - the standard "dual-regression stage-2" approach [Filippini 2009]. For each subject, these 15 (or 25, 50, 100, 200 or 300) spatial maps can then be used in cross-subject analyses of the group-ICA maps.

These node maps – saved as CIFTI image files with one map ("timepoint") per group-ICA component - are estimated for the 1003 subjects having complete rfMRI data (4800 total timepoints, recon1 + recon2).

Subject-measure predictions/correlations and heritability estimation

The "HCP MegaTrawl" is a large set of analyses based on partial correlation netmats, derived from all subjects with complete rfMRI (see [Smith 2013] for an example). Multivariate analyses seek to model (across subjects) a given "subject measure" (e.g., one of the behavioural variables such as fluid intelligence), finding a set of edge weights in the HCP data that can partially explain that subject measure. Results of this analysis for all the HCP subject measures are presented as web pages with multivariate prediction and univariate regression results for each subject measure, and thumbnail volume images showing the edges (node-pairs) whose connection most strongly correlates with the variable. We also calculate and present heritability estimates for the netmats. The 820-subject MegaTrawl is currently available at:
https://db.humanconnectome.org/megatrawl/index.html – this will be updated soon with 1003 subjects and using family relationships determined by genotyping rather than solely from self reports.

Figure1.



Description of released files

Group-PCA output Eigenmaps and Group Average Dense Connectomes are downloaded as separate files:

```
HCP_S1200_1003_rfMRI_MSMAll_groupPCA_d4500_Eigenmaps.dtseries.nii              group average PCA eigenmaps from all recon1 + recon2 subjects
HCP_S1200_812_rfMRI_MSMAll_groupPCA_d4500_Eigenmaps_recon2.dtseries.nii        group average PCA eigenmaps from recon2 subjects
HCP_S1200_1003_rfMRI_MSMAll_groupPCA_d4500ROW_zcorr.dconn.nii                  group average dense connectome from all recon1+ recon2 subjects
HCP_S1200_812_rfMRI_MSMAll_groupPCA_d4500ROW_zcorr_recon2.dconn.nii            group average dense connectome from recon2 subjects
```

Files within the main PTN downloads, HCP1200_Parcellation_Timeseries_Netmats.zip and HCP1200_Parcellation_Timeseries_Netmats_recon2.zip

```
subjectIDs.txt               Ordered list of all subjects with complete rfMRI data (recon 1 + recon2) included in this PTN release
subjectIDs_recon*.txt        Ordered list of the subjects by recon version

scripts.tar.gz               Scripts for all steps of the connectivity analyses

groupICA_3T_HCP1200_MSMAll.tar.gz
Group-ICA "parcellations" at several dimensionalities (levels of detail)
   melodic_IC.dscalar.nii     ICA spatial maps (unthresholded Zstats); one "timepoint" per map.                    Grayordinates.
   melodic_IC_ftb.dlabel.nii  Summary "find the biggest" labels image for all ICA spatial maps.                    Grayordinates.
   melodic_IC_sum.nii.gz      ICA maps dual-regressed into subjects' 3D data and then averaged across subjects.  MNI152 space.
   melodic_IC_sum.sum         Summary "thumbnail" PNG images created at the most relevant axial slices(s).      Slices of MNI152 space.

NodeTimeseries_3T_HCP1200_MSMAll_d*_ts2.tar.gz
Node-timeseries, with one tarfile for each choice of group-ICA dimensionality
   Inside each tarfile there is one timeseries text file per subject (concatenated across all 4 runs)
   Within each text file there is one column per "node" (ICA component)
   d25 (etc) describes the original group-ICA dimensionality
   ts2: multiple regression (against the set of ICA spatial maps) is used to estimate node timeseries (same as first stage of "dual regression")

netmats_3T_HCP1200_MSMAll_d*_ts2.tar.gz
Netmats (parcellated connectomes), with one tarfile for each choice of group-ICA dimensionality
   netmats1.txt     All subjects' netmats: one subject's unwrapped netmat per row, computed using full correlation, Z-transformed.
   netmats2.txt     As above, but using partial correlation with modest Tikhonov regularisation ("ridgep" in FSLNets with parameter 0.01)
   Mnet1.pconn.nii  Group-average full correlation netmat
   Mnet2.pconn.nii  As above, but for partial correlation
```

Files containing the individual subject parcellation spatial maps for the analysis using the data from all 1003 subjects having four complete rfMRI runs (both recon versions):

```
1200PTNmaps_d15+25+50+100.tar.gz     subject-specific maps; CIFTI files for all group-ICA dimensionalities from 15 up to 100
1200PTNmaps_d200.tar.gz              subject-specific maps; CIFTI files for group-ICA dimensionality 200
1200PTNmaps_d300.tar.gz              subject-specific maps; CIFTI files for group-ICA dimensionality 300
```

References

C.F. Beckmann and S.M. Smith. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. on Medical Imaging*, 23(2):137-152, 2004.

N. Filippini, B.J. MacIntosh, M.G. Hough, G.M. Goodwin, G.B. Frisoni, S.M. Smith, P.M. Matthews, C.F. Beckmann, and C.E. Mackay. Distinct patterns of brain activity in young carriers of the APOE-e4 allele. *Proc Natl Acad Sci USA (PNAS)*, 106:7209-7214, 2009.

M.F. Glasser, S.N. Sotiropoulos, J.A. Wilson, T.S. Coalson, B. Fischl, J.L. Andersson, J. Xu, S. Jbabdi, M. Webster, J.R. Polimeni, D.C. Van Essen, and Jenkinson, M. - for the WU-Minn HCP Consortium. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80:105-124, 2013.

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

L. Griffanti, G. Salimi-Khorshidi, C.F. Beckmann, E.J. Auerbach, G. Douaud, C.E. Sexton, E. Zsoldos, K. Ebmeier, N. Filippini, C.E. Mackay, S. Moeller, J.G. Xu, E. Yacoub, G. Baselli, K. Ugurbil, K.L. Miller, and S.M. Smith. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage*, 95:232-247, 2014.

A. Hyvärinen. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Trans. on Neural Networks* 10:626-634, 1999.

J.X. O'Reilly, C.F. Beckmann, V. Tomassini, N. Ramnani, and H. Johansen-Berg. Distinct and overlapping functional zones in the cerebellum defined by resting state functional connectivity. *Cerebral Cortex*, 20:953-965, 2009.

E.C. Robinson, S. Jbabdi, M.F. Glasser, J. Andersson, G.C. Burgess, M.P. Harms, S.M. Smith, D.C. Van Essen, M. Jenkinson. MSM: A new flexible framework for multimodal surface matching. *NeuroImage* 100:414-426, 2014.

G. Salimi-Khorshidi, G. Douaud, C.F. Beckmann, M.F. Glasser, L. Griffanti, and S.M. Smith. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90:449-468, 2014.

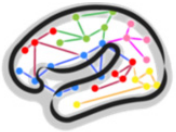
S.M. Smith. The future of FMRI connectivity. *NeuroImage*, 62:1257-1266, 2012.

S.M. Smith, J. Andersson, E.J. Auerbach, C.F. Beckmann, J. Bijsterbosch, G. Douaud, E. Duff, D.A. Feinberg, L. Griffanti, M.P. Harms, M. Kelly, T. Laumann, K.L. Miller, S. Moeller, S. Petersen, J. Power, G. Salimi-Khorshidi, A.Z. Snyder, A. Vu, M.W. Woolrich, J. Xu, E. Yacoub, K. Ugurbil, D.C. Van Essen, and Glasser, M. - for the WU-Minn HCP Consortium. Resting-state fMRI in the Human Connectome Project. *NeuroImage*, 80:144-168, 2013.

S.M. Smith, A. Hyvärinen, G. Varoquaux, K.L. Miller, and C.F. Beckmann. Group-PCA for very large fMRI datasets. *NeuroImage*, 101:738-49, 2014.

S. Smith, M. Glasser, E. Robinson, G. Salimi-Khorshidi, E. Duff, D. Van Essen, M. Woolrich, M. Jenkinson, C. Beckmann. Methods for network modelling from high quality rfMRI data. *Meeting of the Organization for Human Brain Mapping.* 2014.

HCP rfMRI recon version comparisons



Steve Smith, FMRIB Oxford

- **4 subject-grouping datasets**
  - **all820** (from HCP900, mixing recon versions)
  - **all** (1003 subjects from HCP1200, mixing recon versions)
  - **recon1** (184 subjects from HCP1200: *fMRI_3T_ReconVrs=r177*)
  - **recon2** (812 subjects from HCP1200: *fMRI_3T_ReconVrs=r227*)
    (7 subjects have a mix of recon versions)

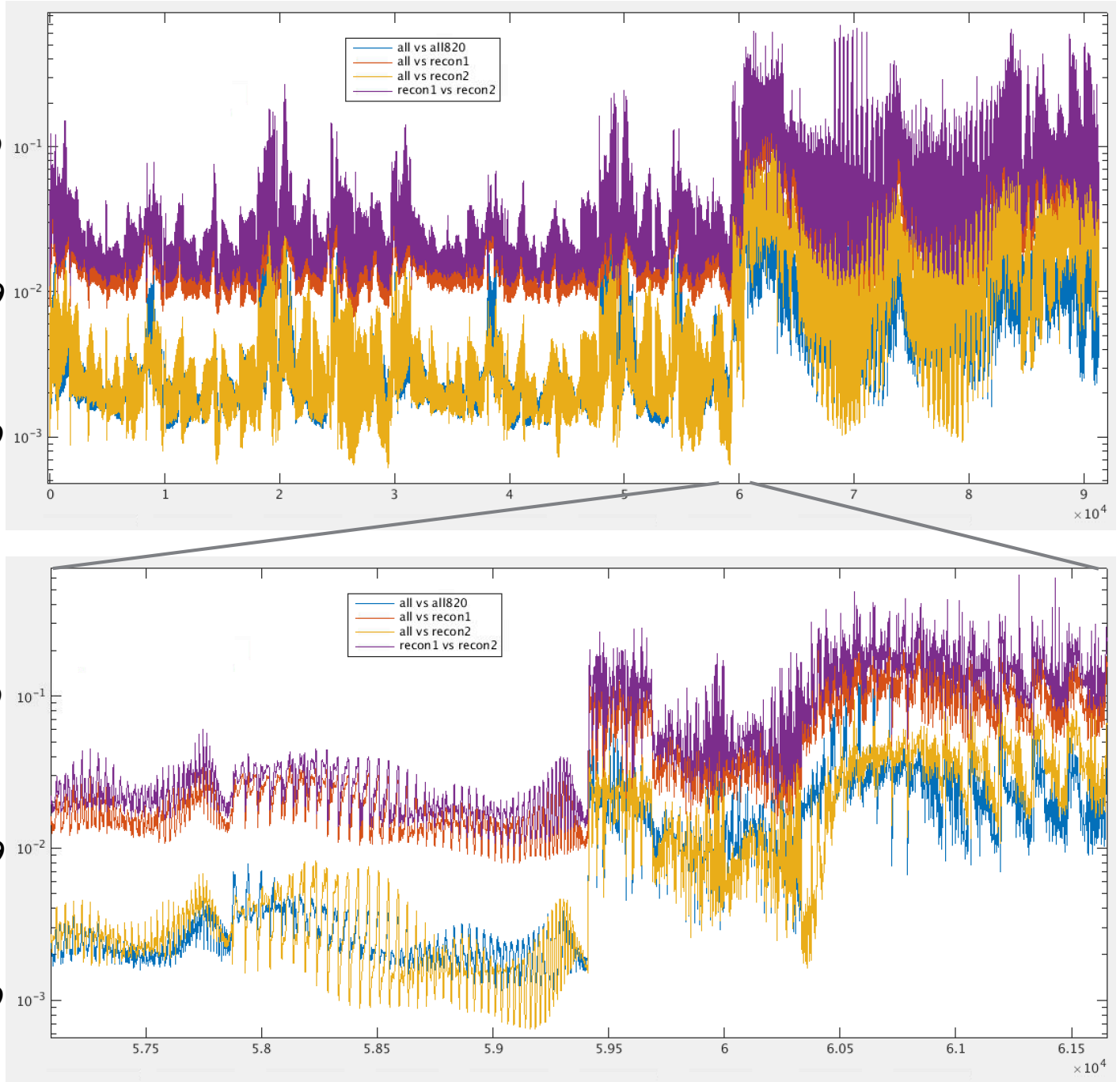
- **Dense connectomes from each dataset** (no-MGTR; MIGP; Wishart eigenadjustment)

- **Correlate dense connectivity maps** (between any two connectomes) **for each grayordinate seed**

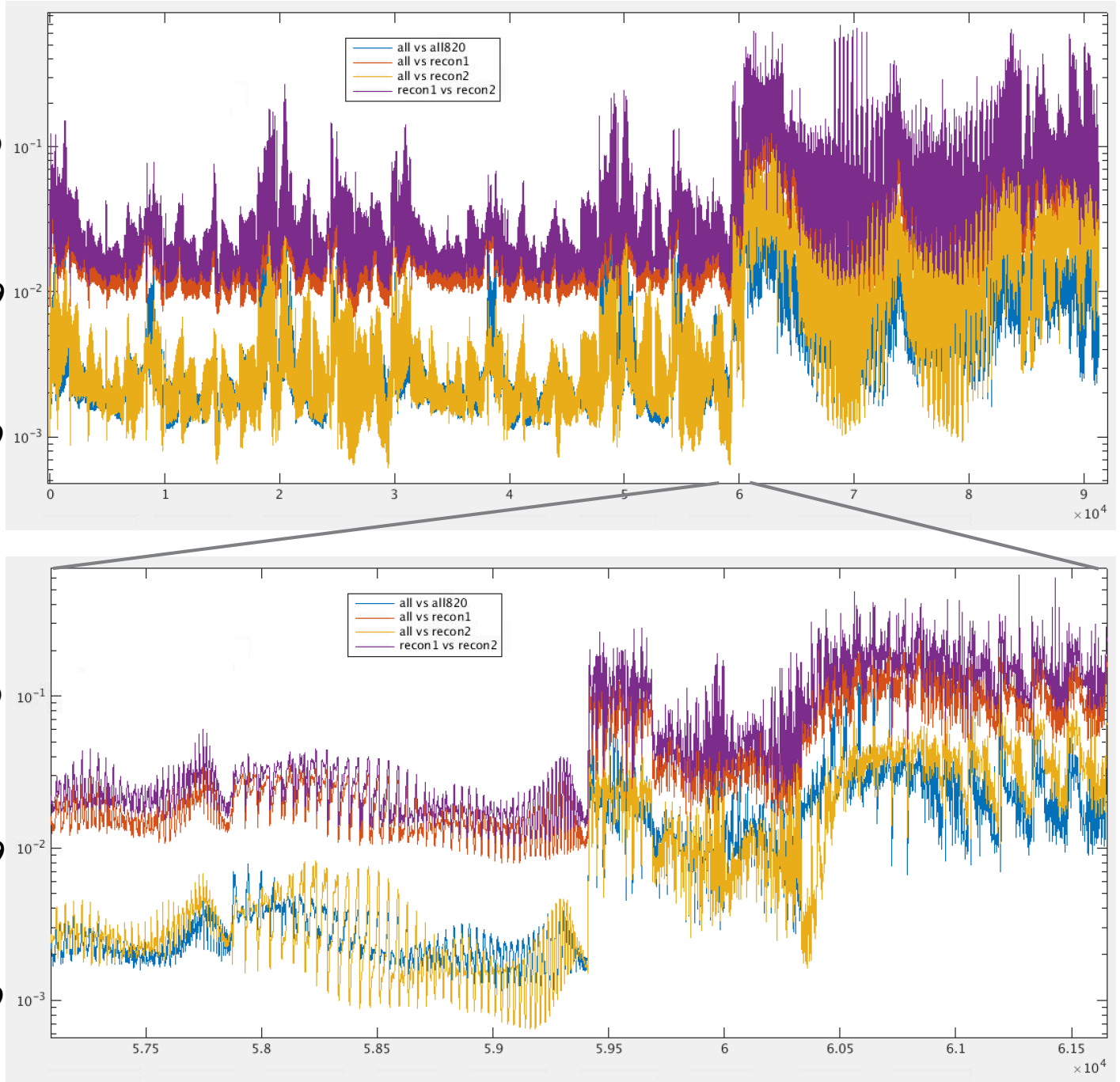
HCP rfMRI recon version comparisons

Each point on the *X* axis is a different grayordinate. A given grayordinate is used as the seed, to estimate the seed-correlation connectivity map. This is done for any two of the 4 group-average datasets. These two (hopefully similar) maps are then correlated with each other to give the correlation value plotted on the *Y* axis.

- *X* axis: grayordinate index
- *Y* axis: *1-r* (lower *Y* = better)

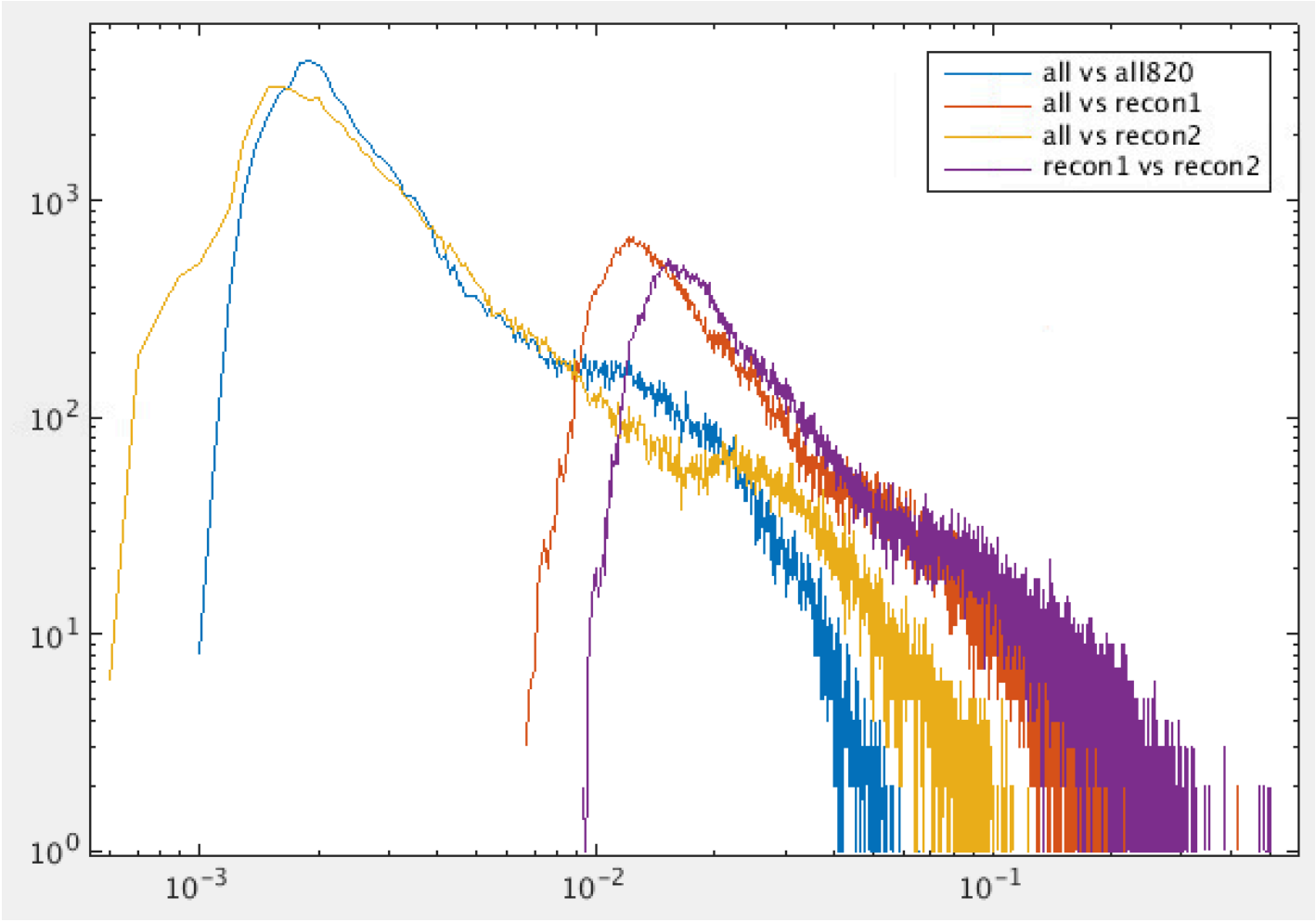


← cortical → sub-cortical



- *X* axis: grayordinate index
- *Y* axis: *1-r* (lower *Y* = better)

HCP rfMRI recon version comparisons



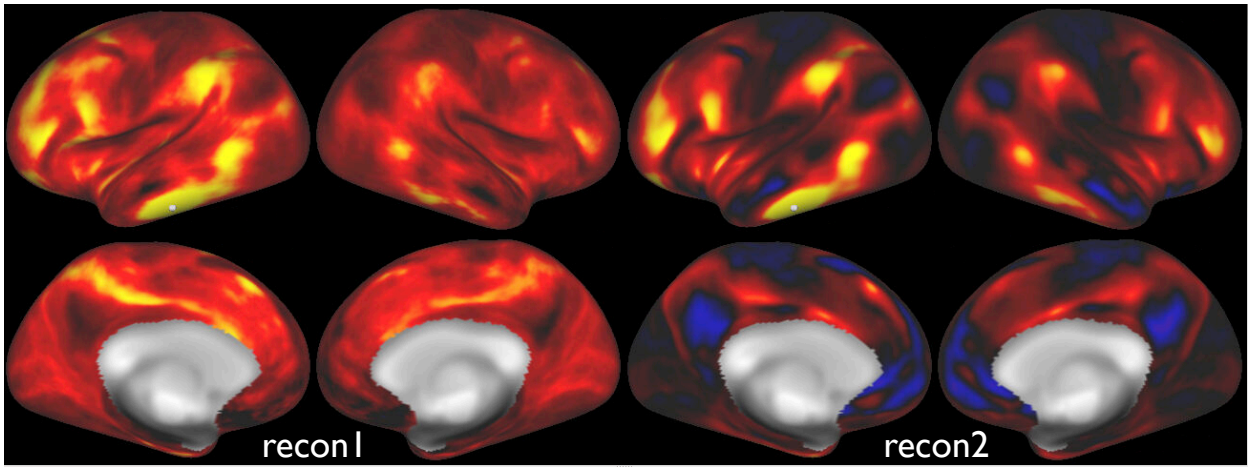
r	mean	median
all vs all820	**0.994**	**0.997**
all vs recon1	0.967	0.981
all vs recon2	**0.991**	**0.997**
recon1 vs recon2	0.954	0.975

- **Histograms over grayordinates**
- *Y* axis is count of grayordinates with a given map-pair-correlation-value
- *X* axis is *1-r* (lower *X* = better)
- "all vs recon2" and "all vs all820" both highly similar

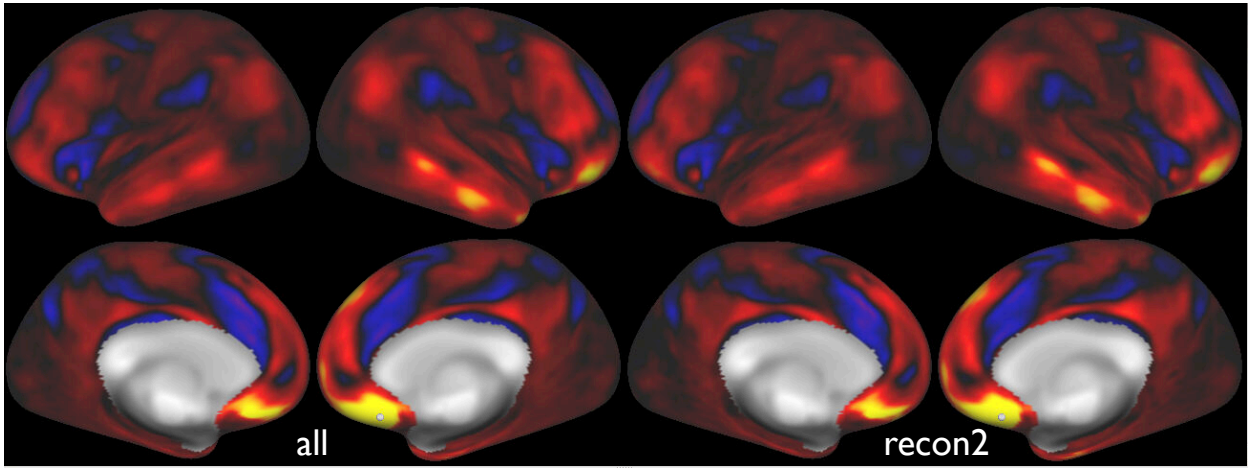
HCP rfMRI recon version comparisons

find cortical seeds with worst map similarity

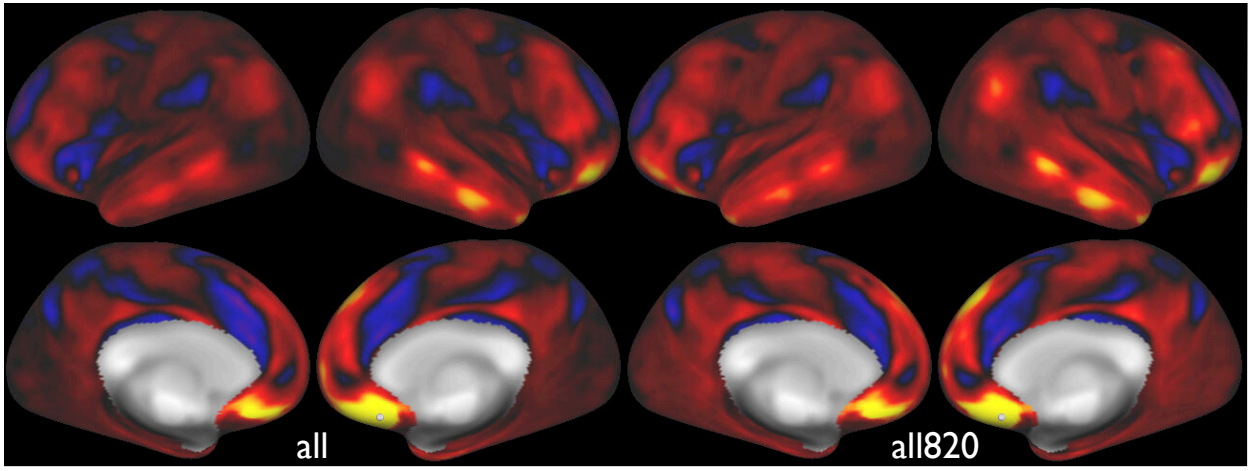
recon1 vs recon2
r = 0.731



all vs recon2
r = 0.969



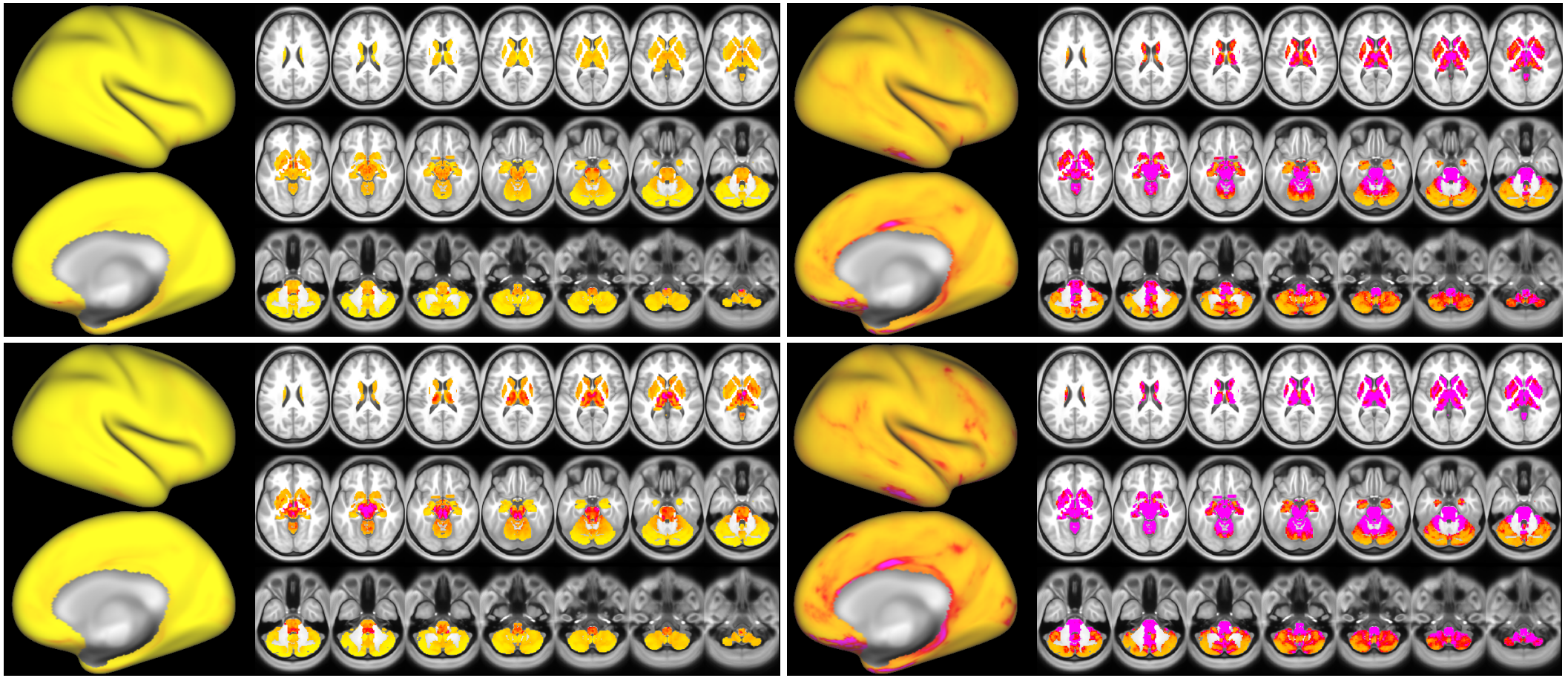
all vs all820
r = 0.948



HCP rfMRI recon version comparisons

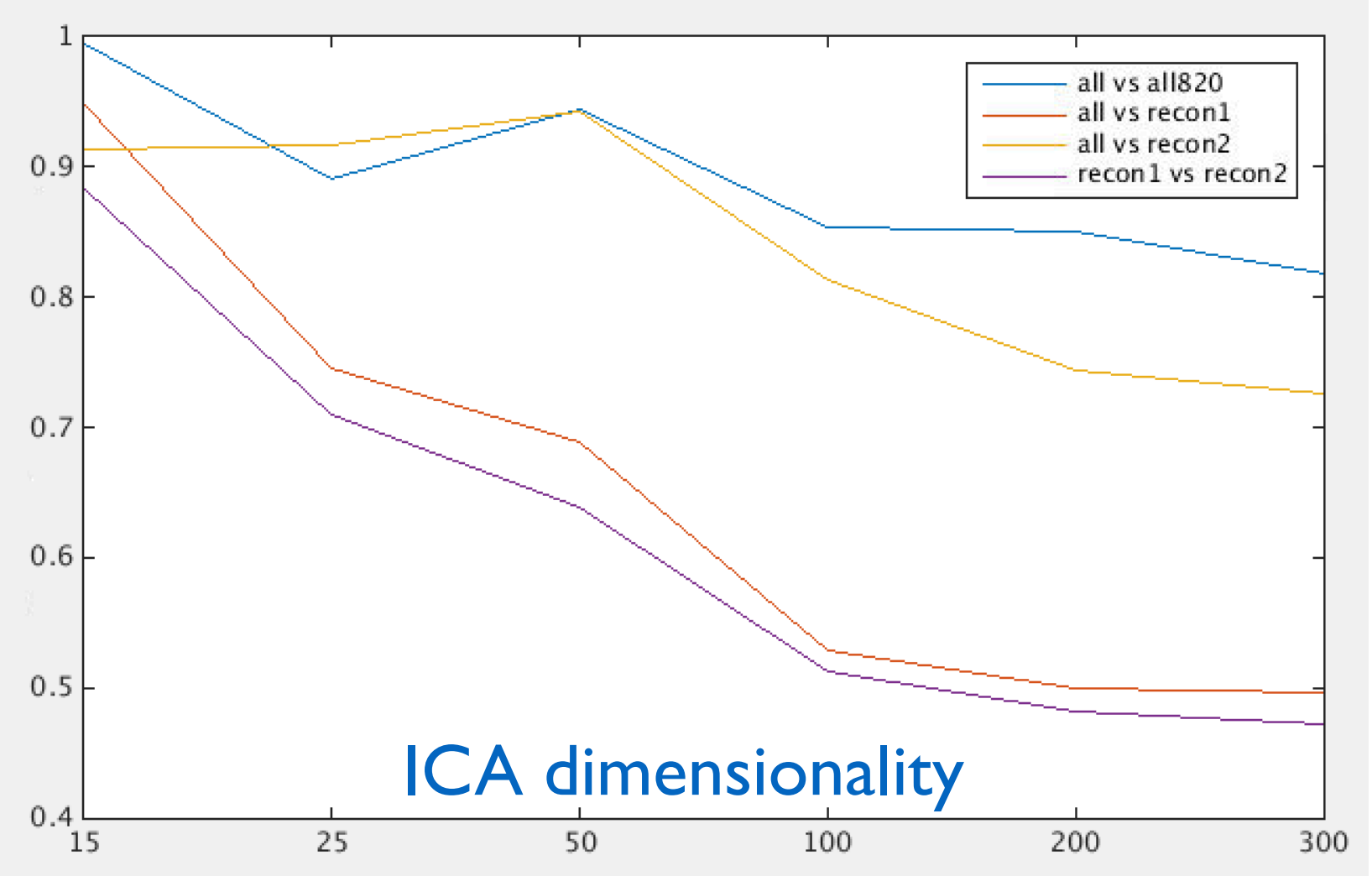
correlation maps

each grayordinate shows the correlation between the dense connectomes (for that seed) from two datasets



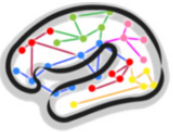
HCP rfMRI recon version comparisons

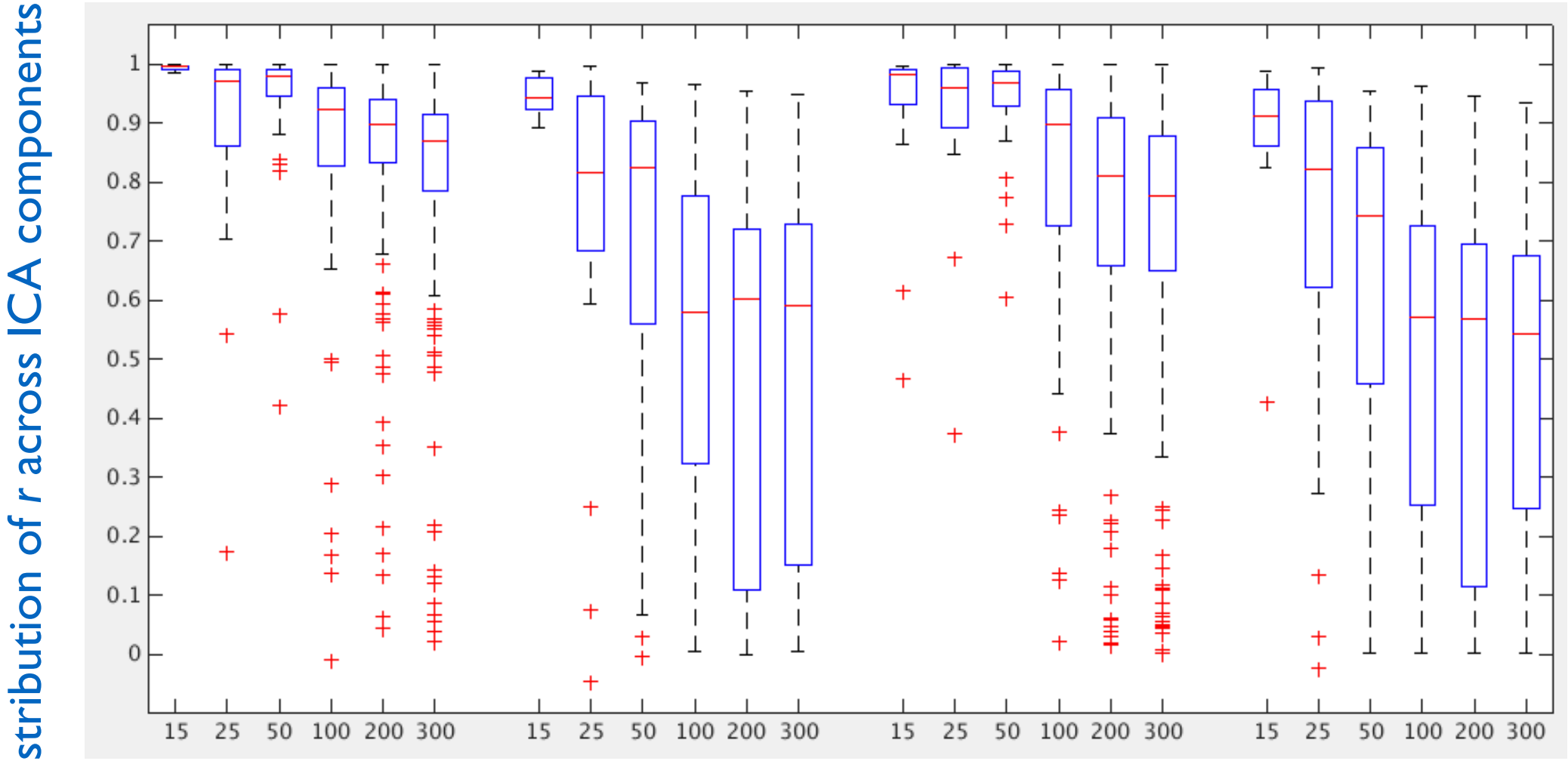
mean *r* (higher is better)



- **4 sets of group-ICA** (4 input datasets, as above)
- **For any two datasets, for a given ICA dimensionality, pair up spatial maps and estimate average pairwise correlation across all maps**

HCP rfMRI recon version comparisons

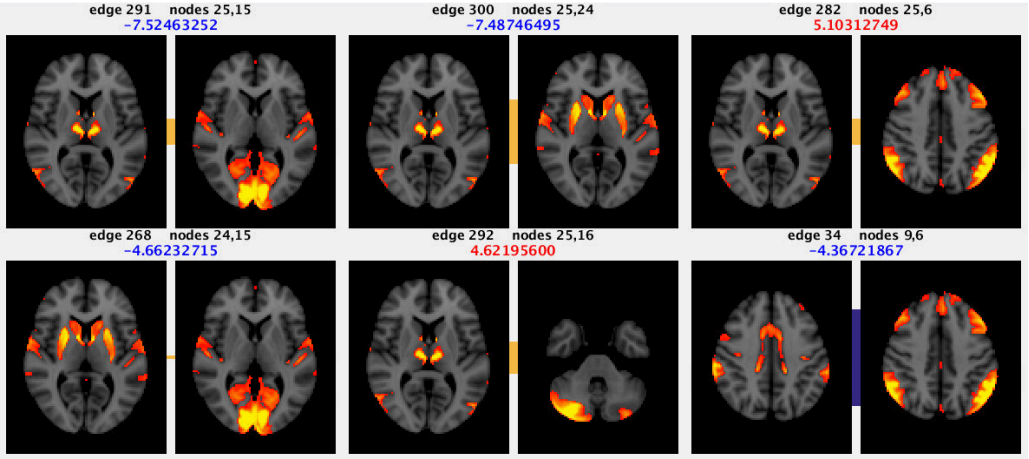




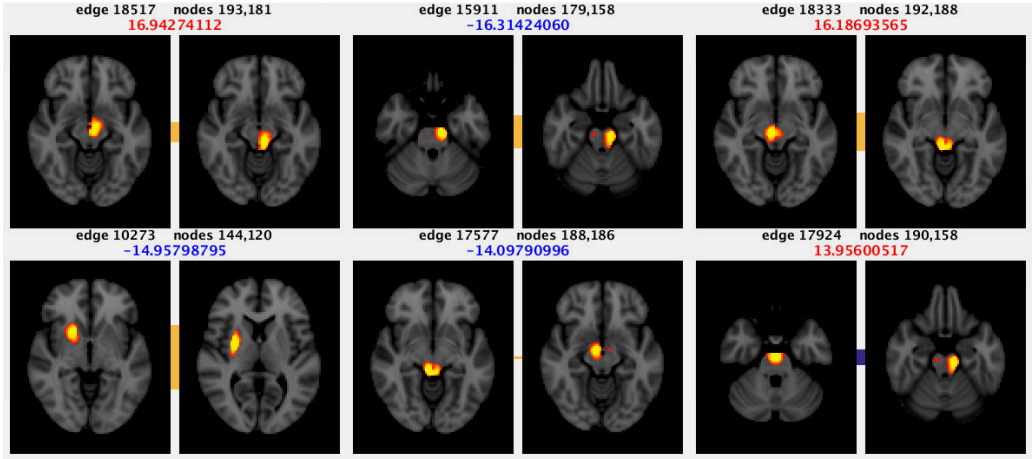
- **4 sets of group-ICA** (4 input datasets, as above)
- **For any two datasets, for a given ICA dimensionality, pair up spatial maps and estimate pairwise correlations for all map-pairs**

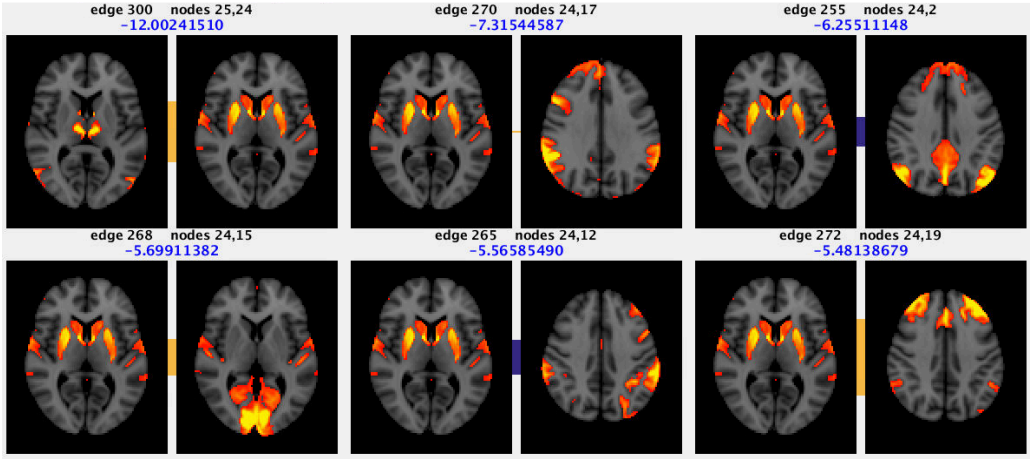
HCP rfMRI recon version comparisons

d=25 full correlation



d=200 full correlation

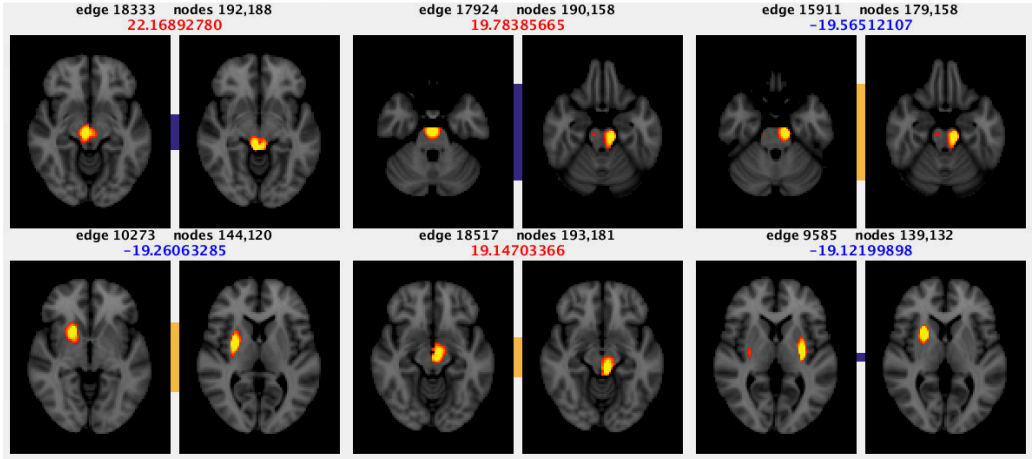




d=25 partial correlation

13 significantly different edges

subjects can be classified (LOO LDA) with **82%** accuracy



d=200 partial correlation

982 significantly different edges

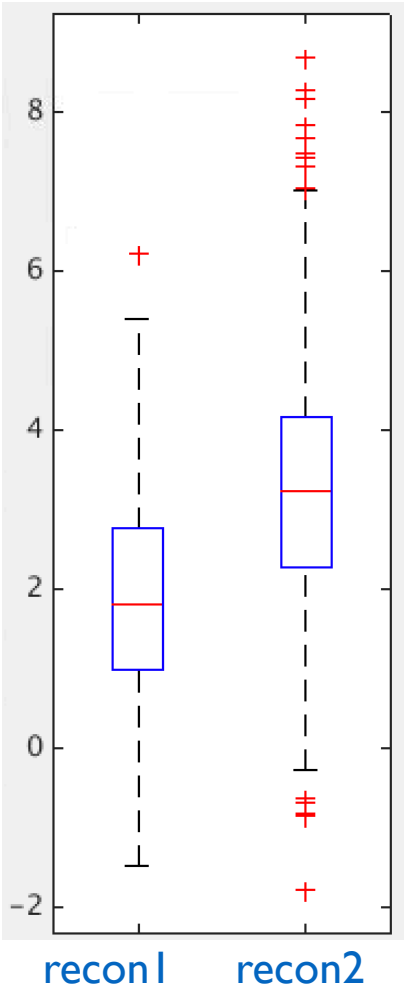
subjects can be classified (LOO LDA) with **99%** accuracy

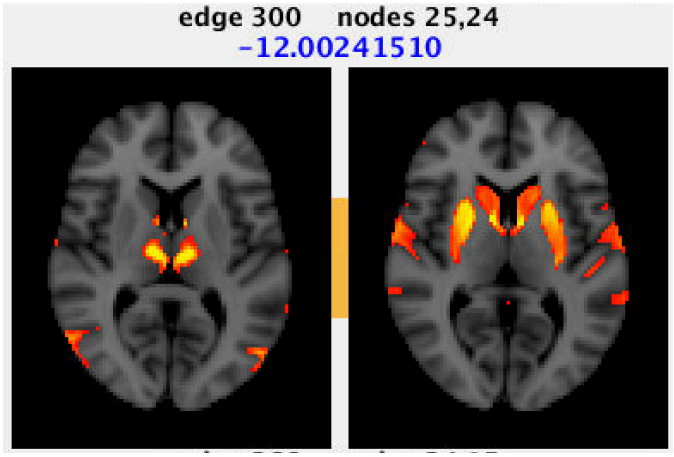
netmats: t-test between recon1 and recon2

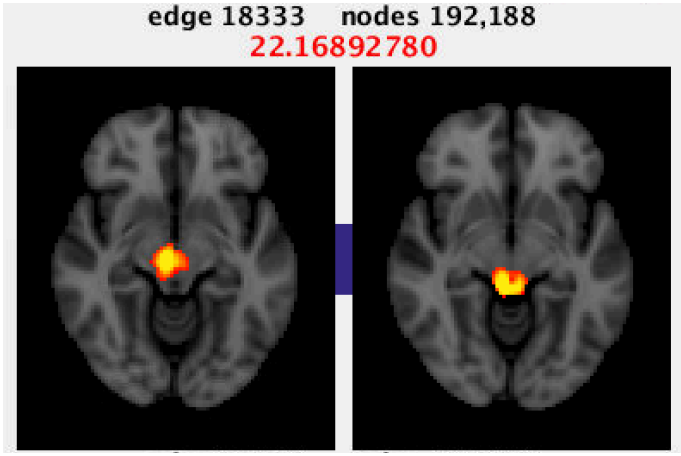
all edges shown p<0.05 (Bonferroni corrected)

HCP rfMRI recon version comparisons

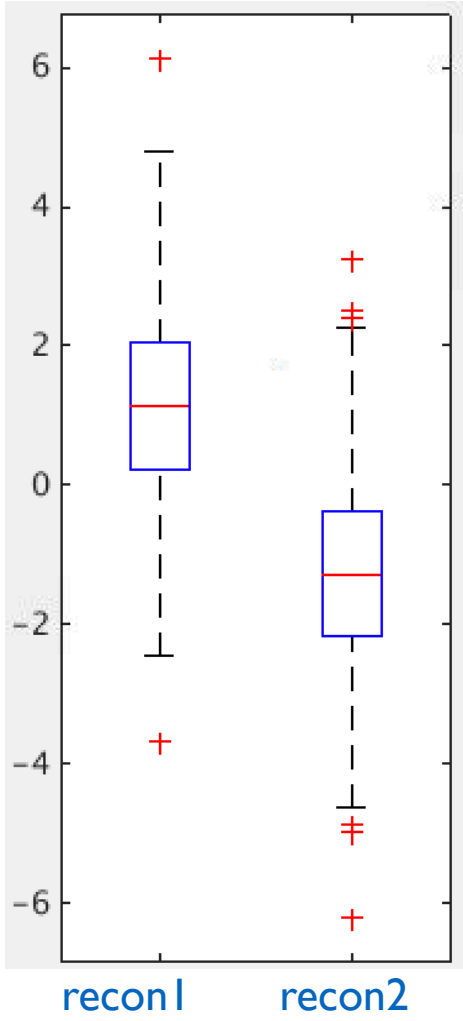
d=25 partial correlation
edge with largest recon change





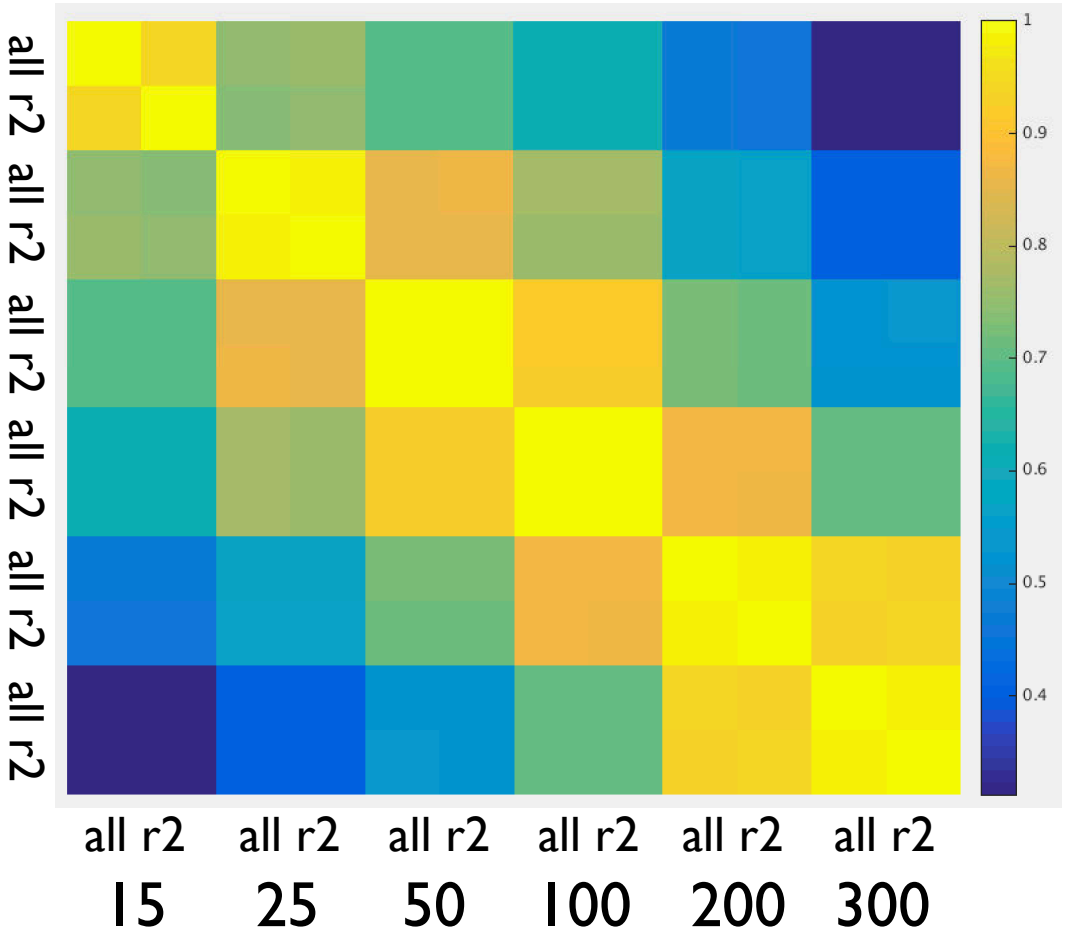


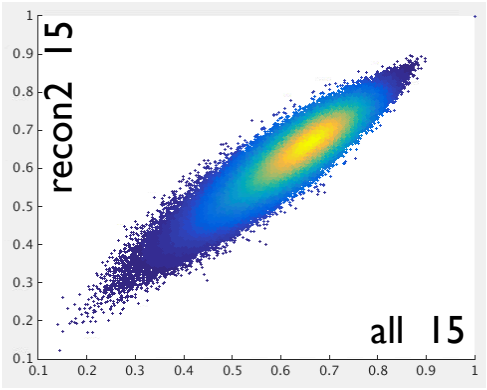
d=200 partial correlation
edge with largest recon change



HCP rfMRI recon version comparisons

- For each dataset, take the *SUBJECTSxEDGES* population-netmat matrix and form the *SUBJECTSxSUBJECTS* correlation matrix ***S***; unwrap to a vector
- Even though the nodes (and hence edges) are distinct for **all** vs **recon2** (and between dimensionalities), ***S*** *can* be compared
- Correlate ***S*** between **all** vs **recon2** and between group-ICA dimensionalities
- Within-dimensionality, correlations between **all** and **recon2** high: >0.95





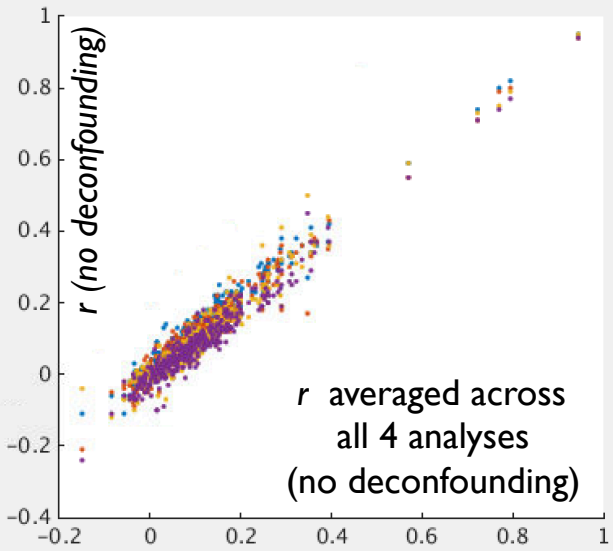
Each point is the correlation between netmats from a given pair of subjects

x axis is correlation estimated from all d=15.
y axis is correlation estimated from recon2 d=15.

HCP rfMRI recon version comparisons

- MegaTrawl prediction accuracies (one point per non-imaging subject-measure)
- Prediction accuracies quantified as *r* (correlation between actual and predicted SM) & *CoD* (1-error/variance)
- Done for:
  - group-ICA d=25,200
  - **all** and **recon2**
- All highly similar

r



CoD