

HCP beta-release of group-ICA maps, node timeseries and network matrices

August, 2014 "HCP500-PTN" (Parcellation+Timeseries+Netmats) data release

Summary

Resting-state fMRI (rfMRI) data from the 500 subjects 2014 HCP data release was processed, using data from all 468 subjects having four rfMRI runs, yielding the following outputs:

1. A group-average "parcellation", obtained by means of group-ICA.
2. Subject-specific sets of "node timeseries" – for each subject, a representative time series per ICA component ("parcel").
3. A subject-specific "parcellated connectome" – for each subject, a nodes x nodes matrix – the functional connectivity between node timeseries.

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/documentation/citations.html>

Data pre-processing and Group-ICA

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]. 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 468 subjects. This comprises the top 4500 weighted spatial eigenvectors from a group-averaged PCA (a very close approximation to concatenating all subjects' timeseries and then applying PCA) [Smith NeuroImage 2014].

The MIGP output was fed into group-ICA using FSL's MELODIC tool [Hyvärinen 1999, Beckmann 2004], applying spatial-ICA at several different dimensionalities (25, 50, 100, 200, 300). The dimensionality determines the number of distinct ICA components; a higher number typically means that the significantly modulated 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 25 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 25 (or 50, 100, 200 or 300) timeseries can then be used in network analyses, as described below. Two distinct methods were used to estimate the node-timeseries:

1. The more traditional "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].
2. A newer approach based on estimating the principal eigen-timeseries within each ICA component; this approach aims to be more robust against component misalignment (between the group-ICA maps and individual subjects' data) and artifacts [O'Reilly 2009, Smith OHBM 2014].

Node timeseries – represented as ASCII text files - are estimated for the 461 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 (25-300) node-timeseries were fed into network modelling, creating an $N \times N$ 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 461 subjects (complete timeseries) described above. We applied network modelling in two ways:

1. 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. 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 averaged over all 461 subjects. This "average" is the Z-transformed result of a one-group t-test across all subjects' individual Z-stat netmats. Netmats are provided as "pconn" files, readable by HCP workbench software (and directly viewable in the `wb_view` workbench display tool).

Subject-measure predictions/correlations and heritability estimation

We plan to release the results of a large set of analyses based on partial correlation netmats, derived from the 462 subjects from the 500 Subjects release 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 will be presented as web pages with multivariate prediction and univariate regression results for each subject measures, and thumbnail volume images showing the edges (node-pairs) whose connection most strongly correlates with the variable. We will also calculate and present heritability estimates for the netmats.

Description of released files

```
groupICA_3T_Q1-Q6related468_MSMSulc.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 spatial maps projected from the surface into the volume, for reporting purposes. MNI152 space.
  melodic_IC_sum.sum          Summary "thumbnail" PNG images created at the most relevant axial slices(s).    Slices of MNI152 space.

NodeTimeseries_3T_Q1-Q6related468_MSMSulc_ICAd* ts*.tar.gz
Node-timeseries, with one tarfile for each choice of group-ICA dimensionality and timeseries estimation approach
  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)
  ICAd25 (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")
  ts3: eigentimeseries is used to estimate node timeseries; first PCA component within a spatial map is used

netmats_3T_Q1-Q6related468_MSMSulc_ICAd* ts*.tar.gz
Netmats (parcellated connectomes), with one tarfile for each choice of group-ICA dimensionality and node timeseries estimation approach
  *_netmat1/*      One netmat file per subject, computed using full correlation, Z-transformed.
  *_netmat2/*      As above, but using partial correlation with modest Tikhonov regularisation ("ridgep" in FSLNets with parameter 0.01)
  *_Znet1.pconn.nii Group-level netmat; a one-group t-test across all subjects' full correlation netmats, Z-transformed
  *_Znet2.pconn.nii As above, but for partial correlations
```

Example simple MATLAB/FSLNets code

```
BATCH='3T_Q1-Q6related468_MSMSulc'; % main name of this data release
d=25; % ICA dimensionality
M='ts3'; % node-timeseries method

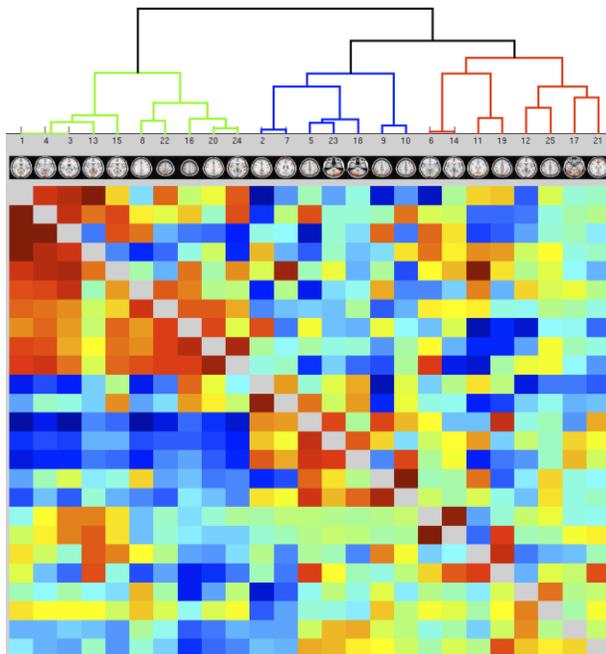
% load in node timeseries
ts_dir=sprintf('node_timeseries/%s_d%d_%s',BATCH,d,M);
ts=nets_load(ts_dir,0.72,1,1);

% view mean spectra
ts_spectra=nets_spectra(ts);

% estimate subjectwise netmats
netmat1=nets_netmats(ts,1,'corr'); % full correlation
netmat2=nets_netmats(ts,1,'ridgep',0.01); % regularised partial correlation

% estimate groupwise netmats
[Znet1,Mnet1]=nets_groupmean(netmat1,0,1);
[Znet2,Mnet2]=nets_groupmean(netmat2,0,1);

% view hierarchical clustering of nodes
SUMPICS=sprintf('groupICA/groupICA_%s_d%d.ica/melodic_IC_sum',BATCH,d);
[hier,linkages]=nets_hierarchy(Znet1,Znet2,ts.DD,SUMPICS);
set(gcf,'PaperPositionMode','auto','Position',[1 1 46*(ts.Nnodes+1) 1574 ]);
```



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.
- 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(3):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*, 2009.
- 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*, 2014. In press.
- 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.