Data-driven dynamic mapping of the human brain

Nick Allgaier, Rob Whelan, Chris Danforth, Josh Bongard, Peter Dodds, Hugh Garavan, & The IMAGEN Consortium

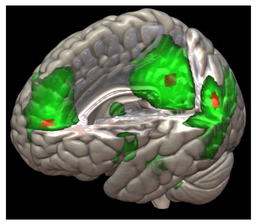
University of Vermont Computational Story Lab, Vermont Complex Systems Center, & Vermont Advanced Computing Core

Abstract

The field of neuroimaging has truly become data rich, and as such, novel analytical methods capable of gleaning meaningful information from large stores of imaging data are in high demand. Those methods which might also be applicable on the level of individual subjects, and thus potentially useful clinically, are of special interest. In the present study we introduce just such a method, called *data-driven dynamic mapping* (3DM), and demonstrate its application in the analysis of resting state fMRI (functional Magnetic Resonance Imaging) from a 242-subject subset of the IMAGEN project, a European study of risk-taking behavior in adolescents that includes longitudinal phenotypic, behavioral, genetic, and neuroimaging data. Dynamic mapping employs a computational technique inspired by biological evolution to discover and mathematically characterize interactions among ROI (regions of interest), without making linear or univariate assumptions. Statistics of the resulting interaction relationships comport with recent independent work, constituting a preliminary cross-validation. Furthermore, nonlinear terms are ubiquitous in the models generated by 3DM, suggesting that some of the interactions characterized here are not discoverable by standard linear methods of analysis. We uncover one such interaction that shows potential for distinguishing between drinking and non-drinking adolescents.

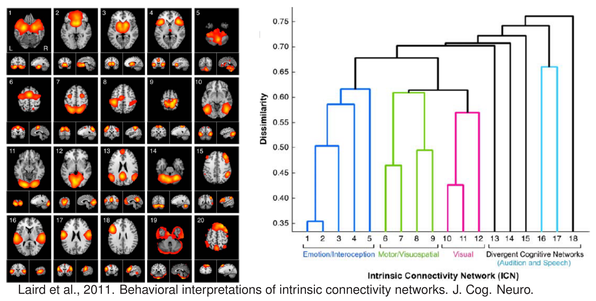
Data from regions of interest (ROI)

Each of the 52 ROI chosen for this study consists of a sphere of 100 voxels, (radius 3 voxels ≈ 1cm). The data come from 6-minute fMRI scans of 242 adolescents who were asked to to keep their eyes open while in the scanner, but were presented with no task or stimulus. Resting-state scans pose a challenge to standard fMRI analysis techniques, which typically involve regression of the BOLD signal (fMRI contrast is Blood Oxygen Level Dependent) in terms of a stimulus or task signal. Here we introduce a methodology that bypasses this challenge by regressing the signal from a particular ROI against the signal from *all other ROI*, and additionally avoids the linear and univariate assumptions typically made in standard approaches.



Locations of the ROI were chosen based on work by Laird et al. in 2011, in which statistical analysis across thousands of imaging studies identified networks of ROI that activate together. (Left) *z*-statistic plot (green) derived from ICA (independent component analysis) and corresponding ROI selected for this study (red) for one network called the default mode network (DMN), which shows damped activity (with respect to resting state) when a subject is presented a stimulus or performing a task.

Recent work from Laird et al.

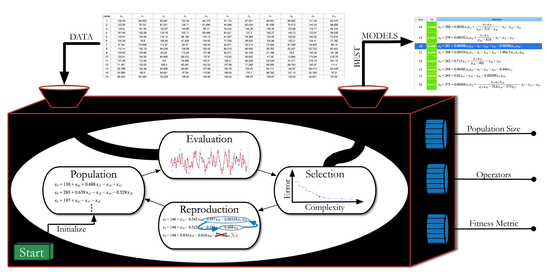


Laird et al., 2011. Behavioral interpretations of intrinsic connectivity networks. J. Cog. Neuro.

(Left) In Laird et al. (2011), 20 ICN (intrinsic connectivity networks) were defined by ICA across the BrainMap database. (Right) HCA (hierarchical cluster analysis) of associated metadata allowed a functional behavioral characterization of the networks. The methodology we introduce reproduces most of the non-artifactual ICN, may extend the above hierarchy, and shows potential for identifying phenotypic variation.

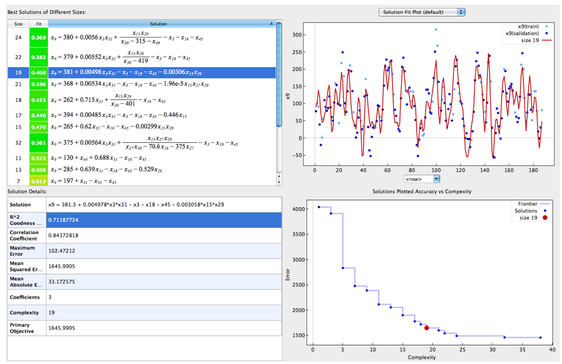
Genetic Programming (GP)

GP is a population based optimization algorithm that searches for equations explaining observations of a particular variable (BOLD signal from an ROI in the brain) as a function of some other observed variable(s) (BOLD signals from other ROI). Symbolic regression is performed, whereby the only assumption made on the form of the models is the set of mathematical building blocks from which they can be constructed, (arithmetic operations in the present work). RMS error was the quantity subject to optimization in this study. An important aspect of GP illustrated in the figure is that the result of a single search is a whole set of potential models along the Pareto front of accuracy vs. parsimony. This is useful for a number of reasons, and in particular, provides a trove of information for statistical analysis that would be lost by the choice of a single model. (Below) A schematic of the GP algorithm.



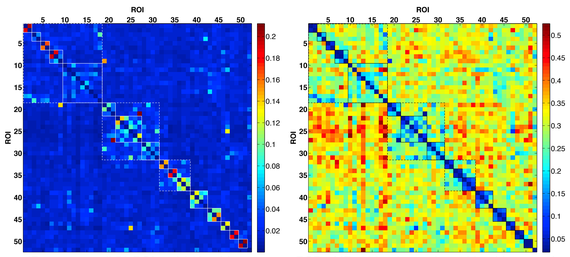
Applying GP to fMRI

To apply GP to the fMRI data, for each of the 242 subjects we extract BOLD signal time series from the 52 selected ROI, and regress each ROI in terms of the others. Note that the algorithm has no knowledge of the networks themselves, but acts on data extracted from regions that come from within these networks. Over 17 total core-years of computation performed on the Vermont Advanced Computing Core (10 days real-time) provide roughly 12 thousand Pareto fronts comprised of a total of about a quarter million models for statistical analysis.



(Above) Screen shot of the GP package Eureqa during a search for models of the activity in ROI 9 as a function of activity in the other 51 regions. (Top left) the current set of potential models along the Pareto front of accuracy vs. parsimony (bottom right), each dot a model, red dot the highlighted model. (Top right) data from ROI 9 (dots) over the 6-minute scan of one of the subjects, with the highlighted model (red line), and (bottom left) statistics for the model fit.

Interaction map



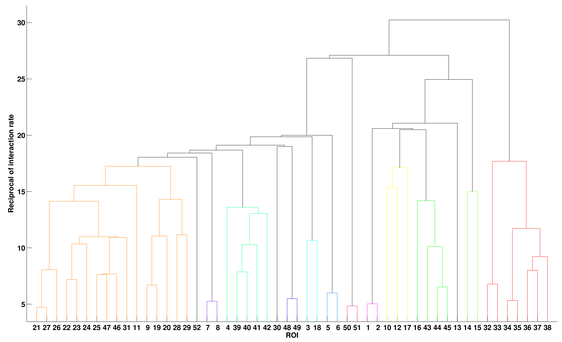
IR map over all 52 ROI and 242 subjects

RSD of IR, 100 subsamples, 100 subjects each

For each ROI we count the number of models, across all 242 subjects, that have terms with a particular (other) region, and then do this for each other region. After normalizing, the result is a vector for each ROI that summarizes its dependence on other regions. We interpret this vector as a distribution of likely interaction and define the computed values to be *interaction rates* (IR). (Left) An interaction map is formed by stacking these IR row vectors. Subsampling is then used to test robustness. For each sample we perform the same counting procedure to produce a corresponding interaction map. (Right) A heat map of RSD (relative standard deviation) of IR over 100 subsamples of 100 subjects each (with replacement).

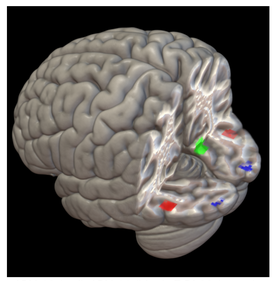
Block-diagonal structure in the IR map corresponds to grouping of ROI into ICN, emphasized by solid outlines. For example, ROI 39-42 make up the DMN. Intra-network IR robustness is supported by matching block-diagonal structure of low subsampling RSD ($< 15\%$). A secondary block structure groups ICN, emphasized by dashed outlines. Consider regions 32-38, composing ICN 10, 11 and 12. A lighter block structure suggests interaction among these ICN, which *together* execute visual processing. Secondary structure for regions 19-31 comprises ICN 6-9, which perform motor and visuospatial tasks. Matching structure of moderate subsampling RSD ($15\% < RSD < 25\%$) indicates robust inter-network interaction.

Hierarchical Cluster Analysis (HCA)



(Above) Dendrogram illustrating HCA of interaction among ROI. The distance between regions is the reciprocal of IR between them, e.g. regions with an IR of 0.2 between them (ROI 1,2) have an interaction distance of 5. The organization of ROI into ICN, and ICN into functional groups is apparent. Also note the following:

- The orange group to the far left includes all but one of the ROI from ICN 6-9, the motor and visuospatial complex. Indicated are interactions with the anterior cingulate gyrus (ROI 9) and ICN 15 (ROI 46,47), thought to be responsible for multiple cognitive processes such as attention and inhibition.

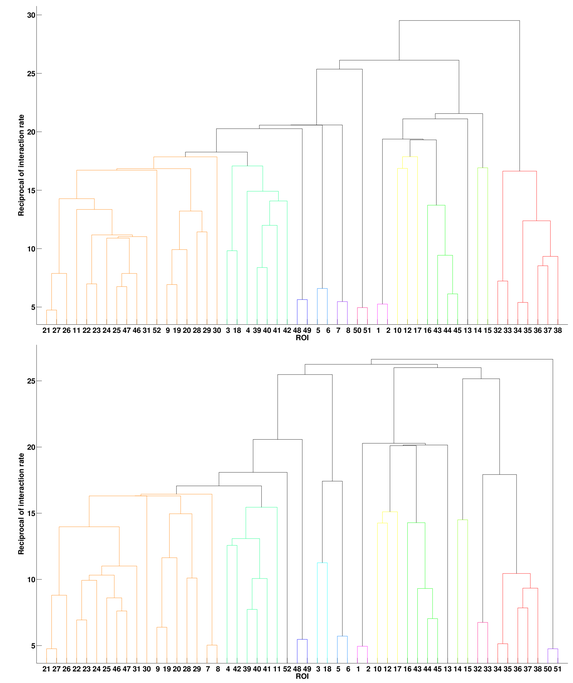


ICN 10 (red); ICN 11 (blue); ROI 36 (green).

- Regions 39-42, the left green group, form ICN 13, the DMN, and show interaction with ROI 4, the orbitofrontal cortex, from ICN 2.
- The red group to the far right forms the visual cluster composed of ICN 10, 11, and 12. (Left) ICN 10 (ROI 32,33) consists of the middle and inferior temporal gyri, while ICN 11 (ROI 34,35) and ICN 12 (ROI 36-38) are the lateral and medial posterior occipital cortices, including V1, V2, and V3. Note the apparently stronger interaction between ICN 11,12.

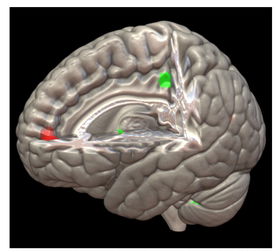
Phenotypic variation

Statistics of IR among ROI may differ between phenotypic groups. The hierarchical organization of ROI induced by IR might illuminate, in such cases, variation in functional dynamics associated with demographic, behavioral, or genetic characteristics. An example illustrating this potential is provided by the contrast between drinking and non-drinking adolescents from the IMAGEN dataset.



(Above) Interaction hierarchies for (top) 100 non-drinking adolescents (NDA), 1 or fewer lifetime drinks, and (bottom) 100 drinking adolescents (DA), 2 or more lifetime drinks. The NDA hierarchy appears strikingly similar to the population-level hierarchy, but there are subtle differences, a notable one being the weaker interaction within the DMN. In contrast, the DA hierarchy is more clearly different at large and small scales. Observe the following:

- In addition to the decreased intra-DMN IR in the NDA hierarchy, there is increased IR between the DMN and the ROI pair 3-18, the subgenual ACC (reward and thirst tasks) and midbrain (interoceptive stimulation), for the NDA.



Ventromedial prefrontal cortex from the DMN (red), and posterior cingulate (larger green)

- For the DA group, the ROI pair 3-18 interacts with ICN 3, the bilateral BG and thalamus, regions linked most strongly to reward tasks, and interoceptive processes such as hunger and thirst.
- At the larger scale, the top level of the DA hierarchy is missing, due to the visual centers becoming more coupled with other ICN.
- (Left) The largest single difference between the DA and NDA groups is the IR between the ventromedial prefrontal cortex (in the DMN) and ROI 11, the posterior cingulate.

Acknowledgments

This work was supported by NASA through the Vermont Space Grant Consortium, DARPA grant FA8650-11-1-7155, and the Vermont Advanced Computing Core, NASA (NNX-08AO96G) at the University of Vermont. The authors also acknowledge Mike Schmidt, creator of the GP package Eureqa used in this study, and Ilknur Icke, Kevin Murphy, Trevor Andrews and Richard Watts for crucial creative and technical support. Finally, the authors wish to acknowledge the following members of the IMAGEN Consortium: Banaschewski, T., Barker, G.J., Bokde, A.L.W., Bromberg, U, Büchel, C., Conrod, P., Flor, H., Frouin, V., Gowland, P., Heinz, A., Ittermann, B., Mann, K., Martinot, J-L., Nees, F., Paus, T., Pausova, Z., Rietschel, M., Robbins, T., Smolka, M.N., Ströhle, A., Schumann, G.