Automated ROI to ROI functional connectivity analysis and visualization with ANNA

Automated Functional Connectivity (FC) analysis is fun and easy with ANNA!

Two Simple Steps to using ANNA:
After preprocessing your data and ensuring that the directory structure requirements are met (described below):
1. Copy "anna.sh", "anna.r", and your favorite ROI list (created as described in the "ROI List Creation" section below) to the appropriate group folder (as described in the "Directory Structure" section).
2. Run the appropriate command (as described below).

That's it!

ANNA will retrieve and analyze the appropriate data for each ROI in each network, providing progress related output to keep you updated. ANNA will create a FC dataset containing "traditional/static" connectivity estimates for each connection within and between each network, as well as summary measures and a variety of dynamic connectivity estimates. Each FC estimate (i.e. a single column in the dataset) can be analyzed like any other behavioral measure; think of the correlated neural activity as a behavior in and of itself as it requires effort and individuals differ in performance. FC estimates can be predictor or outcome variables, as well as mediators/moderators of known relationships.

Here are the details:

Directory Structure
The "top level" directory should be a "StudyFolder" that contains a "GroupFolder" which holds all of the subject data. ANNA will create two additional directories in the StudyFolder – an "ROIs" folder that contains a basic text file for each ROI in your list, and an "OutputFiles" folder that contains all of the group level output files. For the GroupFolder, ANNA expects to find one or more subject folders to be included in the analysis. No other folders should be in the GroupFolder directory, but other files are ok (an "ls" command should show only subject folders and other files like anna.sh). Each of the subject folders should contain data for one subject (a preprocessed brain ready for FC analysis); if the study N=20, then 20 subject folders should exist in the GroupFolder. No other folders. anna.sh, anna.r, and your ROI list.txt file will also be placed in the GroupFolder.

EXAMPLE: The basic directory structure should be:

StudyFolder/
StudyFolder/ROIs/
StudyFolder/OutputFiles/
StudyFolder/GroupFolder/
StudyFolder/GroupFolder/S001/PREPROCESSED BRAIN
StudyFolder/GroupFolder/S002/PREPROCESSED BRAIN
.
.
StudyFolder/GroupFolder/S00N/PREPROCESSED BRAIN
StudyFolder/GroupFolder/anna.sh
StudyFolder/GroupFolder/anna.r
StudyFolder/GroupFolder/ROIList.txt

ROI List Creation

Using a basic text editor, create a list of (a-priori!!) ROIs comprised of **two** neural networks – a Network1 and a Network2. All of the ROIs for Network1 should be listed first, followed by those for Network2. The file should be tab delimited and follow the convention in the example below (it is probably best to edit an existing list until you are comfortable with the format). ANNA was designed to test hypotheses regarding the functioning and interactions of two neural networks. If hypotheses include only one network, a second "dummy" network can be defined and ignored (a future version of ANNA will explicitly allow a single network definition). Save the file in the GroupFolder directory. The name of the ROIList.txt file should be specific to your analysis as the resulting datasets will be named, in part, by this file (see Output File Naming Convention below). This is the only file you need to edit to use ANNA (it will be passed to ANNA via the command line).

EXAMPLE: (ROIListAnderson.txt)

-04	-52	32	PCC-04-5232
04	-53	35	PCC04-5335
-06	36	-15	vmPFC-0636-15
06	27	-15	vmPFC0627-15
-49	-62	34	TPJ-49-6234
50	-57	36	TPJ50-5736
-43	-38	46	IPS-43-3846
40	-39	51	IPS40-3951
-21	-04	59	EYE-21-0459
27	-06	54	EYE27-0654
-38	14	08	aINS-381408
40	15	08	aINS401508

- In this example, the DMN is defined as Network1 (the first 6 ROIs) and the task active network is defined as Network2 (the following 6 ROIs) (as defined by Anderson, 2011). Notice that leading 0s should be included; try to follow this format exactly. A future version of ANNA may require only a label for the fourth column (no coords).

Generating a functional connectivity dataset using ANNA

Once you have created the appropriate directory structure containing the appropriate data and your ROIList.txt file is saved in the GroupFolder, copy your most recent version of anna.sh and anna.r to the GroupFolder directory. Then, from the command line (i.e. using a terminal window), navigate to the GroupFolder and type "./anna.sh ROIList.txt X Y" where X is the desired radius of the ROIs in mm and Y is the number of ROIs in Network1.

EXAMPLE: If ROIs with a 5mm radius are desired and Network1 contains 6 ROIs, type:

./anna.sh ROIList.txt 5 6 (then press enter)

ANNA will run, pick up speed after the first ROI, and let you know when it is done. Results will be in a set of .csv files in the "OutputFiles" directory. The primary files to be used for analyses contain the Z converted connectivity estimates, but the correlations are also output for reporting purposes; all filenames will be in accordance with the naming convention below. Two group level files are most important. The first is a "NetEstimatesZROIList...csv" file that contains the FC estimates, both traditional and dynamic. This is the primary file to be used for FC analyses. Also generated is a "SampleNetStatsROIList...csv" file that contains a sample level summary. As noted a "NetEstimatesROIList...csv" file is also created that contains the raw connectivity estimates.

IMPORTANT!!: A good amount of info is also output to the screen, including a report of unusable data for each subject. Unusable data generally occurs when a TR (i.e. a single timepoint) was censored due to motion during preprocessing, but may also occur for other reasons. An entire list of causes for unusable data is not currently known, but results in a 0 for that TR in the data timeseries (NAs are also sometimes present and are dealt with similarly); this interferes with creation of the connectivity datasets and ANNA bypasses the issue by changing the 0s to a very small random number near 0 (.0001 to .0009). Typically, this adjustment is only needed for a small amount of data from a few subjects, but large amounts of unusable data should be a signal to check the subject to ensure that their data should be included in the analysis (they may need to be excluded due to excessive motion).

Output File Naming Convention

All files output from a single run of ANNA will have the same time stamp and follow a naming convention that allows you to easily identify the files you need.

An example is: "NetEstimatesZROIListAnderson-5mm-08-30-16-17-07-43.csv".
The convention is: "NetEstimatesZROIFILE-Xmm-month-day-year-hour-minute-seconds"

ROIFILE is the name of your ROIList.txt file and should named to help you remember what network definitions you were testing. X is the radius of the ROIs in mm. The time stamp is the same for all files to aid identification of which files were created during a single run and to ensure that no data will ever be overwritten.

Visualizing dynamic functional connectivity using ANNA

Visualizing network functioning using ANNA is easy if you can generate a dataset using the steps above. If one subject folder exists in the GroupFolder, then ANNA will automatically visualize the functioning. If more than one subject folder exists, then ANNA will automatically generate sample level stats and will not visualize anything. The visualization will loop through all TRs in the dynamic connectivity dataset and end when the scan is finished. ANNA will plot the ROIs in a 3D, MNI sized, interactive space and show the networks as they function over time and interact. Colors are Bonferroni corrected, so bright red or blue (for Network1 and Network2, respectively) represents a statistically significant connection (green is for negative connectivity).

How is the visualization accomplished?

ANNA begins like many resting state scripts by accepting a list of ROIs that is used to generate a correlation matrix of all possible pairwise comparisons. These values are vectorized (i.e. made into a row) for each subject and placed into a spreadsheet. The primary difference for the visualization-based calculations is that instead of correlating all time points, ANNA starts at the third time point and calculates a "window correlation" including the previous two time points and the next two time points, for a total of five timepoints per correlation estimate. Thus, the first estimates used by ANNA to depict network functioning include timepoints 1,2,3,4, and 5 (note that TRs have already been removed to allow for scanner stabilization). The next time point is generated by "sliding" down one TR and repeating the process; thus, the second set of estimates used by ANNA includes timepoints 2,3,4,5, and 6. In this way, only one new piece of information is added to the connectivity estimate per frame, adding stability and reducing the effect of short-term fluctuations in the signal. This process is repeated for all remaining TRs in the scan. The result of this process is a "dynamic connectivity" dataset that contains connectivity estimates that vary across time. Very simply, these values are passed to ANNA for visualization.

Want more info?

Feel free to contact Mike McCormick (mjm0102@auburn.edu) for all your ANNA related questions.