ALLEN INSTITUTE *for* BRAIN SCIENCE

Visualization of changes in gene expression using Brain Explorer

Generation and use of difference grid files, version 2.0

1. PURPOSE OF DIFFERENCE GRIDS

Analyzing image data from the sleep deprivation study presents multiple challenges, including:

1. The need to compare across conditions and brains.
2. Comparison of replicate datasets.
3. Biological specimens that are sectioned at slightly different plane angles.

To address these challenges, methods were developed to perform statistical analysis of ISH expression data resulting from comparing a pair or multiple pairs of gene expression mapped to 3D reference atlas space at the 3D grid level. An overview of image processing and quantification can be found at http://www.brain-map.org/pdf/InformaticsDataProcessing.pdf. The end result is an interactive file in Brain Explorer that allows visualization of up- or down-regulation of gene expression across two conditions for n=3 replicates per condition.

2. DEFINITIONS

Imageseries. An imageseries is a collection of ISH images (typically 20 images) for a hybridization experiment that consists of one probe on tissue sections from one brain.

Expression Energy. ISH signal is quantified from images using a metric for signal intensity (Expression Level) and for expression area (Expression Density). "Expression Energy" combines features of expression intensity and density into a single measurement. (See http://www.brain-map.org/pdf/NeuroBlast.pdf).

Difference Grid. The difference grid shows the expression differences calculated as fold change across two conditions mapped in 3D.

3. DATA PROCESSING MODULES

There are four modules for the generation of the difference grid files version 2.0.

3.1. Normalization

Each imageseries is normalized such that for all expressing grids, the random variable (X) (representing expression energy) is adjusted to zero mean or $X - E(X)$ where $E(X)$ represents the estimated mean value in the random variable.

Other normalization schemes may be developed for future difference grid versions, including normalization across replicate sets.

3.2. Averaging

For each of the imageseries belonging to a given experiment and a given group, an average estimate is derived from the normalized imageseries (**Figure 1A**). The version 2.0 data release includes only data derived from averaging 3 replicate datasets per condition.

3.3. Change Detection

The version 2.0 data release provides fold change detection. The expression file for one averaged condition dataset is divided by the expression file for a second averaged condition dataset, generating a fold difference grid (**Figure 1B**). Only differences > 1.25-fold or < -1.25-fold are shown.
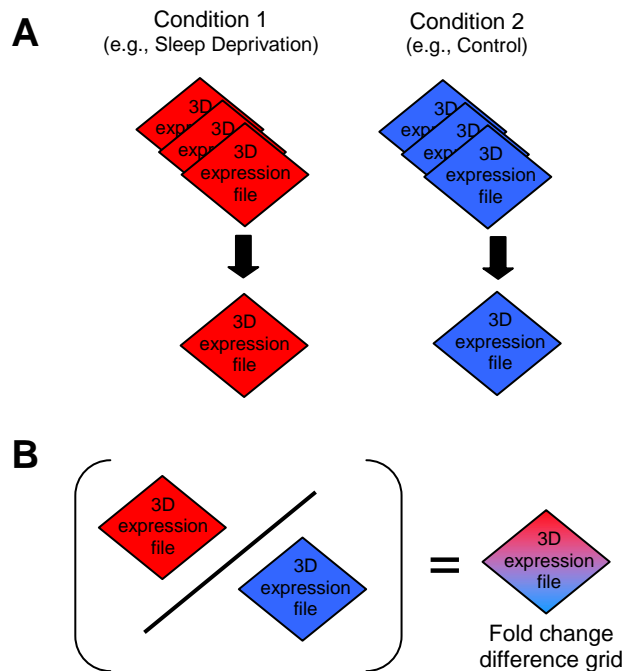
Figure 1. Generation of Difference Grids. After the normalization step (not shown), the three-dimensional expression files of replicate datasets (n=3) are averaged to form one three-dimensional file containing spatial information for average expression energy (A). The fold difference is then calculated from the averaged files of two different conditions for a given gene (B), such that the final difference grid contains information about both positive and negative fold changes for expression energy.
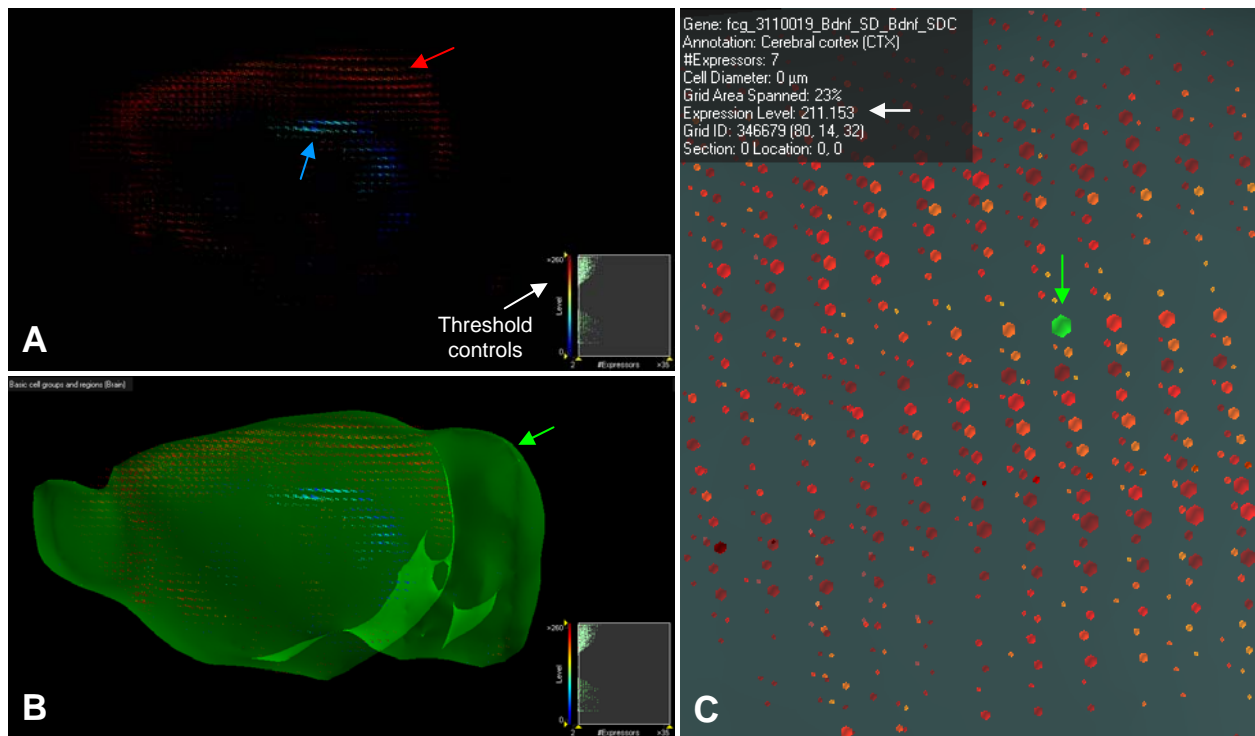
3.4. Visualization

The basic visualization of the difference data is made feasible by loading the XPR files in Brain Explorer (supported by Brain Explorer v. 1.4.1; **Figure 2**).

(1) There are two metrics indicated by the difference grid in Brain Explorer. The color indicates magnitude and directionality of change in "expression energy", indicated as "level" on the vertical axis of the threshold controls. The size of the gridpoints indicates relative magnitude of change in number of expressed cells ("# expressors" on the horizontal axis of threshold controls), but due to the limitations of visualization, this metric is considered an absolute value of fold change, and thus no directionality is indicated. In other words, change in the

number of expressors is a relative value only, and is not equated directly to a fold change.

(2) Customized colormaps allow optimal viewing of the fold change map. In Brain Explorer v. 1.4.1, use the "Jet" colormap for a continuous scale of expression which can be directly related to fold change (**Figure 3**). Generally, red/orange color indicates upregulation, and blue color indicates downregulation in the first condition relative to the second condition.

(3) Threshold the results to filter decreased fold differences (fold change <1.25) or increased fold differences (fold change >1.25).

Figure 2. Visualization of fold change difference grid in Brain Explorer. **A.** The file opened in Brain Explorer shows orange/red expression datapoints for upregulation in condition 1 and blue datapoints for downregulated expression. In this case, the difference between sleep deprivation and its time of day control is shown for *Bdnf*, with upregulated expression in the neocortex (red arrow) and downregulation in the hippocampus (blue arrow). Threshold controls (white arrow) allow filtering of expression based upon fold change in expression energy (vertical axis) or by magnitude of change in density (horizontal axis). **B.** For easier interpretation, surface rendered transparent mouse brain reference atlas structures can be turned on or off, such as a representation of the entire brain (green arrow). **C.** Individual datapoints can be examined by right-clicking on the desired datapoint (green arrow; datapoint turns green when selected), and relevant information will be returned in the upper left of the window, including the anatomical region (cerebral cortex) and change in expression energy (white arrow; listed as "expression level" here). According to the table in Figure 3, an expression level change of 211.153 corresponds to a 3-fold upregulation.

Level	Fold change
255	5
233	4
211	3
189	2
172	1.23
167	1
145	NA
123	-1
118	-1.23
101	-2
79	-3
57	-4
35	-5
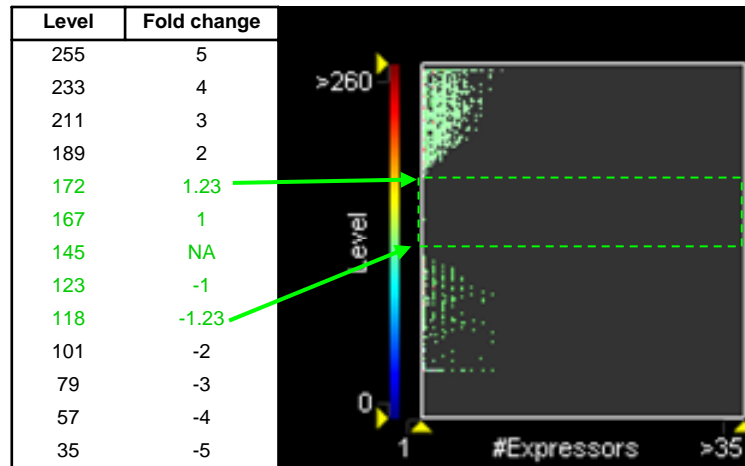
Figure 3. Relationship of "Expression Level" in Brain Explorer difference grid to fold change. The relationship between level and fold change is shown (left). Only fold changes greater than 1.25 or less than -1.25 fold are shown in the difference grid. If the magnitude of change is less than +/- 1.25 fold, the datapoints are removed from the visualization (green arrows and green-dashed box indicate thresholds for fold change visualization).

4. DIRECTIONS FOR OPENING FILES

Directions to set up Brain Explorer for Difference Grid viewing
1) Download Brain Explorer v. 1.4.1.   http://www.brain-map.org/downloadExplorer.do
2) Under "Gene", go to "Colors" and select "Jet". This colormap is calibrated for difference grid files.
3) There are two ways to find files:
    - Browse for files at
    http://download.alleninstitute.org/sleep/DifferenceGrids/
    - Use the "Index to Difference Grid Files" (in the same folder) and click on comparisons for each gene to find the relevant file.
4) The first time you open a Brain Explorer XPR file, you will need to associate it with the Brain Explorer program; alternatively, browse in the difference grid folder and drag and drop file into Brain Explorer.
5) To identify a file once it is open in Brain Explorer, look at the lower right corner. The file name (e.g., fcg_3110010_Arc_SD_Arc_SDC) can be interpreted as follows: "fcg" = fold change grid; 7-digit number is an experiment ID; the first gene symbol and condition (e.g., "Arc_SD") was divided by the second gene and condition (e.g., "Arc_SDC").

5. SUGGESTIONS FOR VIEWING

1) Refer to the Brain Explorer tutorial for a basic overview of Brain Explorer features. To access the tutorial, open Brain Explorer and under "Help", click on "Tutorial".
2) Use threshold controls and expression colormap. Under "Gene", check "Threshold Controls". Under "Gene", go to "Colors", and choose "Jet". Threshold controls will appear in the lower right corner.
3) Downregulation of expression (blue colors) is more visible using the following setting: Under "View", go to "Options". Uncheck the setting "Draw faster but at lower quality". However, using this setting, artifacts can appear as saturated blue or red lines throughout the file; to eliminate these, adjust threshold controls to remove low number of expressors or to remove saturated signal.
4) To view difference grid in the context of brain structures, click on Atlas structures to the right under the "A" column to show the structure, or under the "G" column to show expression in that structure.
5) To identify the atlas structure associated with a grid point, either right click on a grid point to identify the associated atlas structure (**Figure 2C**; upper left text box), or use cross-sections of the atlas: Under "Atlas", choose either "Sagittal sections", coronal, or horizontal. The cross-sections can be moved by right-clicking and dragging.

6. FAQ

Why do difference grids only exist for a subset of genes?
> Difference grids are most accurate when replicate datasets are available, and thus only genes with at least 3 replicate datasets per condition have been used for generation of difference grids.

Why is there signal outside of the brain model?
> There can be multiple reasons for expression that is seemingly outside of the brain model. First, the ISH data was collected from one sagittal hemisphere, but tissue sections were collected slightly past the midline of the brain. The brain model only covers one hemisphere, so any sections collected past the midline appear "outside" of the model. Second, occasionally there is distortion of tissue during dissection, resulting in misplaced tissue that appears outside of the brain model. This occurs most frequently with the olfactory bulb.

How can I tell if the difference is real?
> The difference grid is only a tool for viewing that can help identify regions for further analysis. Any differences observed with these files should be confirmed by examining the ISH data directly. The "Index to Difference Grid Files" in the folder http://download.alleninstitute.org/sleep/DifferenceGrids/ also contains links to the ISH image data. In the difference grids, however, there are some obvious artifacts that appear. For instance, large tissue bubbles may have such a dark staining that even after averaging of replicate files, they may show up as gridpoints in the data files that do not appear to correspond to an anatomical structure (**Figure 4**).
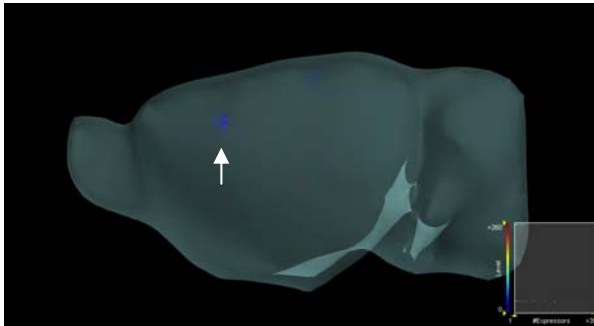
Figure 4. Artifact in difference grid. The white arrow indicates a likely tissue bubble that shows up as a collection of large gridpoints that does not appear to associate with any anatomical structure in the vicinity.

What features of Brain Explorer are incompatible with difference grids?
1) Normal Brain Explorer files allow the user to click on a grid point to retrieve a close-up image of the corresponding ISH experiment. Because difference grids incorporate information from 6 experiments, they cannot be used to link directly to one ISH image.
2) Difference grids are not searchable from within Brain Explorer.