AGEA: An Anatomic Gene Expression Atlas

1. What is *AGEA*?

The Allen Institute has built a data-driven three-dimensional atlas of the adult C57Bl/6J mouse brain based on the ISH gene expression images of the Allen Brain Atlas (ABA). The project is called AGEA and stands for Anatomic Gene Expression Atlas. Essentially, AGEA characterizes the multi-scale spatial relationship in the mouse brain as derived from gene expression data without *a prior* knowledge of classical anatomy. In the AGEA online application, you can

- View and navigate 3D spatial relationship maps (***Correlation*** mode);
- Explore a transcriptome based spatial organization of the brain (***Clusters*** mode) and,
- Search for genes with local regionality as defined by AGEA (***Gene Finder*** mode).

Upon entering the AGEA application the screen in Figure 1 should appear.

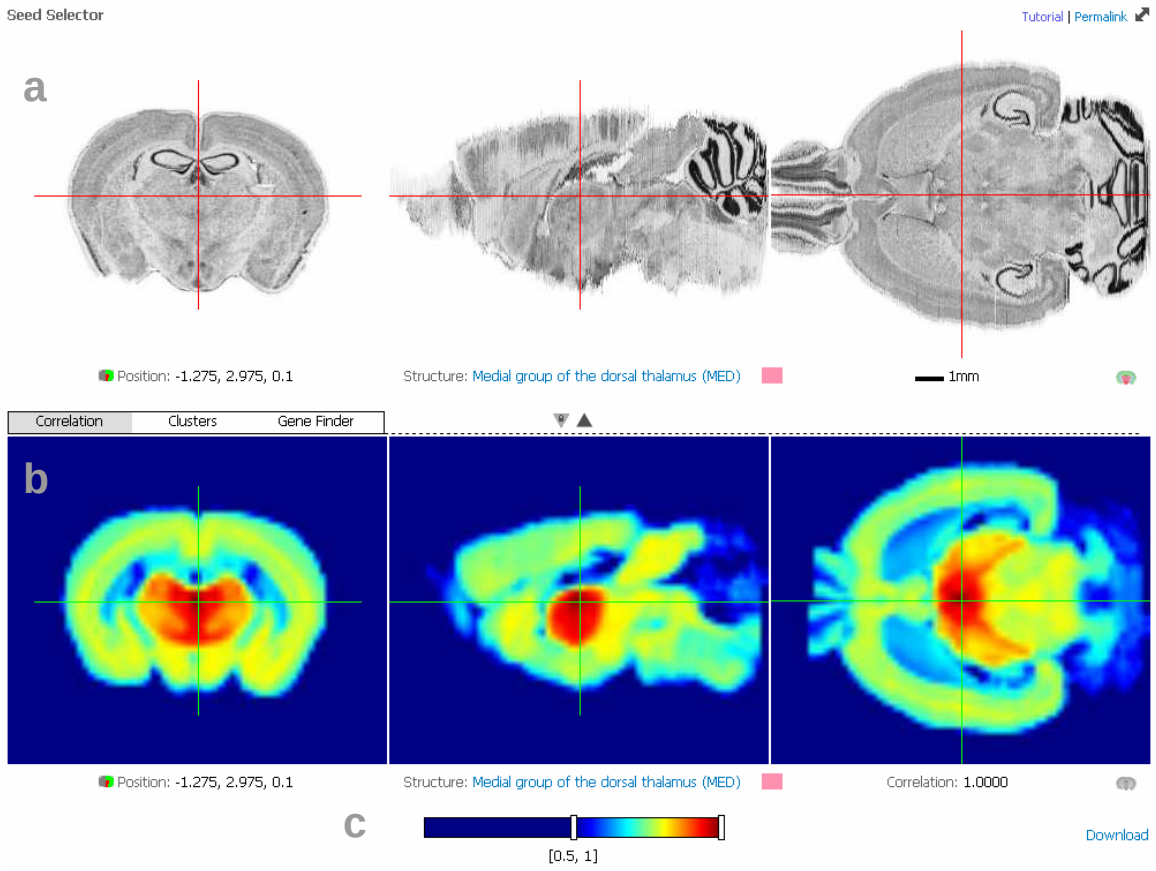

Figure 1. The AGEA interface in *Correlation* mode. Images in the upper row (a) shows three orthogonal views of the 3D Nissl based Allen Reference Atlas used to select the "seed" voxel (3D pixel). The corresponding AGEA correlation map shown in the lower panel (b) illustrates the average gene expression correlation (see Section 1: Spatial Relationship Maps) between the seed voxel and other voxels in the 3D brain. Users can threshold the correlation mapping to change the dynamic range by the slider control shown (c). This particular AGEA map can be accessed using this permalink.

Spatial Relationship Maps

The upper panels of images in Figure 1 are orthogonal views of a 3D Nissl reference atlas volume. In ***Correlation*** mode, the lower panels show orthogonal views of the selected spatial relationship map. The AGEA spatial relationship maps are computed from spatially mapping 4,376 coronal images series of the ABA. This initial gene set was chosen from the ABA because it contains many of the genes of known neuroscientific interest as well as genes exhibiting marked or unique expression patterns.

Each ISH image series is processed through an automated pipeline [1] that detects the location of expressing cells in the images. The ISH image series are then reconstructed into a 3D volume and expression statistics from the images are pooled so that the data can be accessed by a standard coordinate system with (200µm x 200µm x 200µm) voxels. The ontology of the Allen Reference Atlas is used to label individual voxels with their anatomic nomenclature. To quantify the overall expression effect at a given voxel, we compute an *expression energy* measure that represents the product of expression level and density over cells contained in that voxel (see [3] for details).

Once the expression energy map is obtained for each gene (M=4,376) and for each voxel (N=51,533), we may use correlation to explore the relationship of gene expression and anatomic location. For each voxel, the Pearson's correlation coefficient from the seed voxel to every other voxel is computed by comparing expression vectors of length M. Performing this computation over all N voxels results in 51,533 three-dimensional correlation maps. For each correlation map, we denote the reference voxel to which all other voxels are compared as the "seed" voxel. A web viewer was developed to provide easy navigation between the maps and within each 3D map. The correlation values are displayed as a 24-bit false color images using a blue-to-red ("jet") color scale.

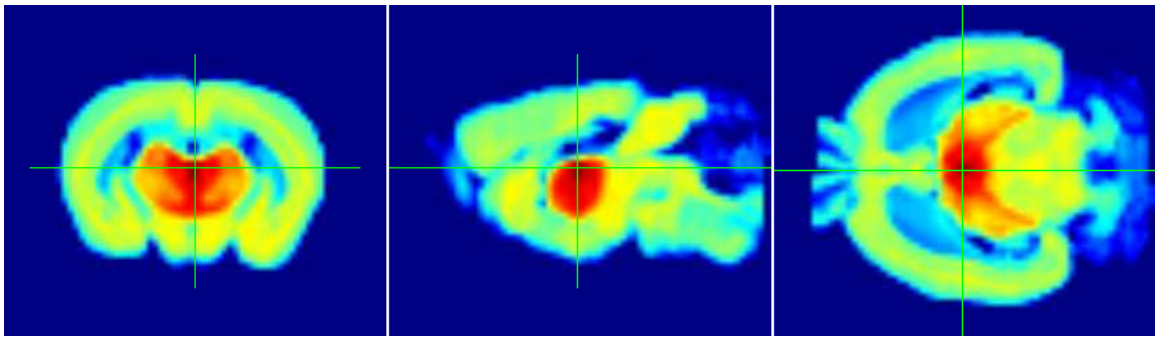

Figure 2. Sample AGEA spatial map in coronal, sagittal and horizontal planes. The dark red voxel corresponds to the seed voxel and is observed to be highly correlated with other regions of the thalamus. Dragging the green cross hairs in one image moves the viewing plane shown in the other two images [permalink].

The correlation values in Figure 2 can be interpreted as a measure of average co-expression between two voxels. The higher the correlation value between voxels, the more common it is for genes from the input set to be co-expressed. Higher correlation between two voxels indicates more spatial correlation of expression and thus potentially higher possibility that the spatial regions spanned by the voxels are anatomically related. This may indicate that the voxels compared share common cell types or represent a coherent functional area. For example, in Figure 2, the dark red cluster in the map approximately overlaps the intermediodorsal nucleus of the thalamus (IMD). As correlation values decrease, a local coherent group may become apparent. For example, the dark orange versus light orange voxels likely corresponds to differential functional groups in the thalamus. The correlation map can also be used to locate co-expressing areas in other brain regions. In Figure 2, the map indicates that there is higher co-expression with layer 5 than other layers of the cerebral cortex.

Hierarchical Spatial Organization

Selecting ***Clusters*** mode switches the lower panels to view a data-driven hierarchical spatial organization of the brain computed from the AGEA correlation maps. The spectrum of gene expression patterns in the brain is complex displaying both intra-structure widespread expression and multifarious regional specificity. We have used a simple strategy that captures the various scales of spatial co-expression as described below.

To construct the decomposition, voxels are spatially organized as a binary tree (bi-tree) [4] where each node represents a collection of voxels. To initialize all 51,533 voxels were assigned to the root node of the tree. As we descend the tree, a node is bifurcated into two nodes to achieve maximal dissimilarity between two groups of voxels based on correlation values. The bifurcation terminates whenever a node contains only a single voxel. This recursive division scheme is effectively multi-scale processing. The final bi-tree consists of 103,065 nodes with a maximum depth of 53 levels and 51,533 leaf nodes (one for each voxel in the brain). At each bifurcation an ordering (albeit arbitrary) is assigned to each child to enable the definition a global "depth first" ordering for all leaf nodes. Effective visualization of this large data structure is a challenging problem. A viewer has been implemented to provide an easy-to-use mechanism to navigate the bi-tree, provide 3D context and visualize the multi-scale partitioning.

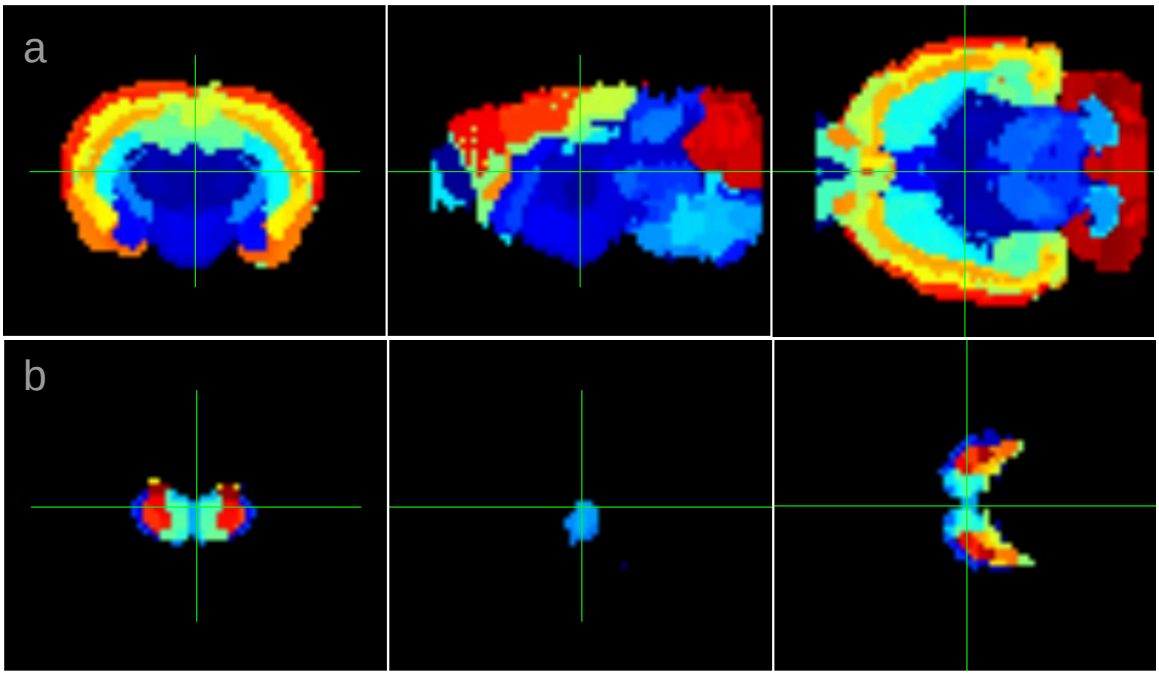

Figure 3: Two nodes in a data-driven hierarchical spatial organization of the mouse brain. Each coronal, sagittal, and horizontal image represents one node of a 3D data driven spatial clustering of the expression modes of the C57Bl/6J mouse brain. Each node in the decomposition tree represents a set of voxels and is visualized by a systematic color coding of the voxels using a "jet" color scheme. The colorization provides a preview to the sub-tree organization. Panel (a) [permalink] is the visualization of the root node spanning all voxels in the brain. Coarse parcellation into gross anatomical structures can be inferred from the colorization. The node in Panel (b) [permalink] spans the thalamus and colorization indicates a medial-lateral as well as a rostral-caudual gradient.

In the AGEA ***Clusters*** viewer navigation works on the principle that from each leaf node (representing a single voxel) there is only one path back to the root. To navigate the bi-tree, a user selects a leaf node by choosing a voxel within the reference volume as in Figure 1(a) and moves up and down the tree (along the root to leaf path) using the *Tree Depth* slider control. At each level of the tree, the voxels corresponding to the node can be browsed within its 3D spatial context as shown in Figure 3(a,b).

The voxels of a node are visualized with a systematic color coding scheme. Each leaf node is assigned a global ordering such that each internal node of the tree represents voxels with a contiguously clustered ordering. All voxels of a node are then assigned a color based on the 'jet' color scheme where the leaf node with low order number is assigned shades of blues. The colors then run through green, yellow and orange. Finally, higher order voxels are assigned shades of reds. Two example node visualizations are shown in Figure 3 for the root node (a) consisting of all voxels of the brain and (b) for a node overlapping the thalamus. Schematic coloring of the voxels by global ordering essentially provides a preview to the sub-tree organization. For the root node Figure 3(a), coarse parcellation into gross anatomical structures can be observed while in Figure 3(b), the arrangement of the colors indicates medial-lateral and rostral-caudal sub-organization in the thalamus.

AGEA Gene Finder

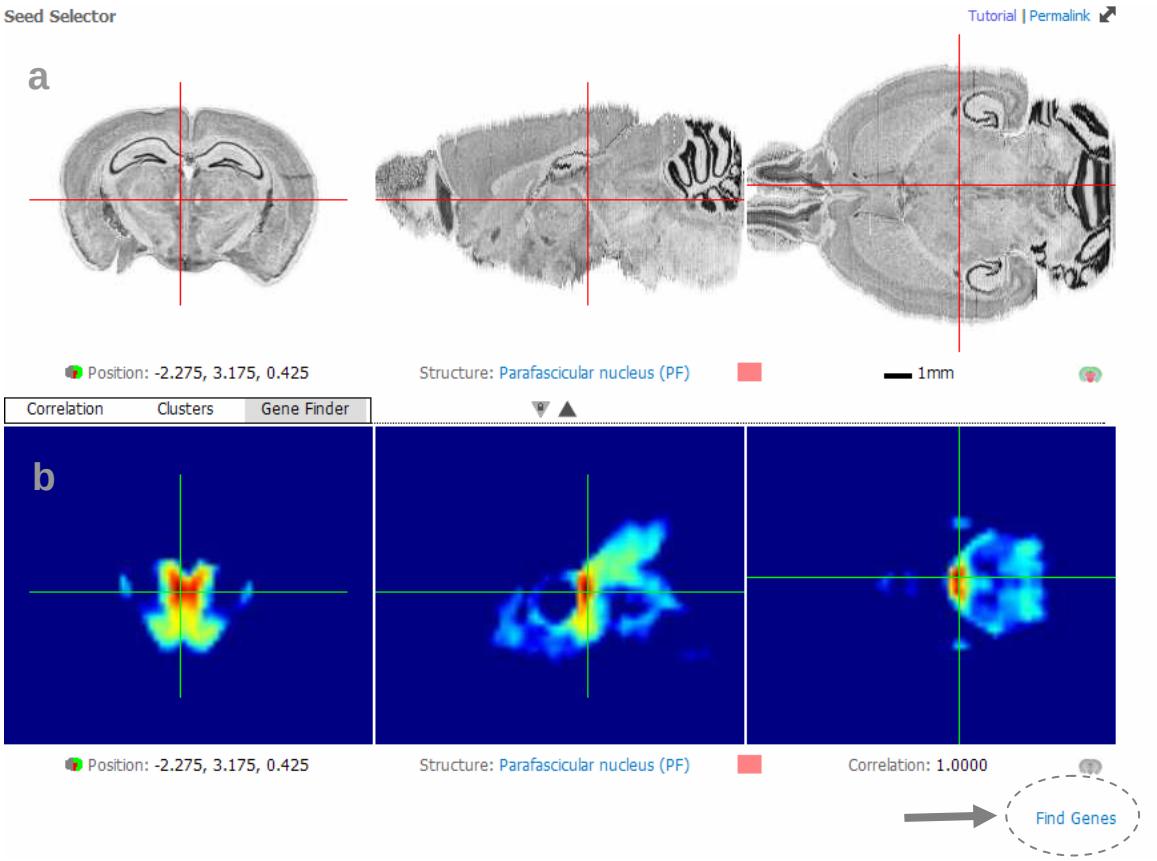

Figure 4: Gene Finder correlation map with seed voxel in the parafascicular nucleus of the thalamus. A fixed AGEA correlation domain is shown in the lower panel. Genes exhibiting the highest specificity to the correlated portion around the seed can be found through the "Find Genes" button shown. The construction of the domain of search and ranking method is described in the text [permalink].

The ***Gene Finder*** search facility is among the most powerful aspects of AGEA's functionality. It enables users to search a local anatomic region of interest for genes within the ABA database that exhibits localized enrichment. Finding genes with highly localized expression patterns is of neuroscientific interest to study structural relationships and/or provide evidence for refinement of

structural boundaries. Using the full complement of tools in AGEA, a user can connect structural neuroanatomy with average gene expression profile and finally to molecular neurogenomics via the original in situ hybridization data.

To use the ***Gene Finder*** feature, the user navigates to a voxel of interest in the reference atlas volume and a *fixed threshold* AGEA correlation map appears (described below in the Gene Finder Algorithm) as in Figure 4b. By clicking on the "Find Genes" button shown highlighted in the lower right a gene list from the ABA is returned. Two of the top ranking genes are illustrated in Figure 5 below. Note that the search facility operates on one fixed threshold. Both computational constraints and the practicalities of finding genes at every threshold level prevent arbitrary control of the threshold in the ***Gene Finder*** search.

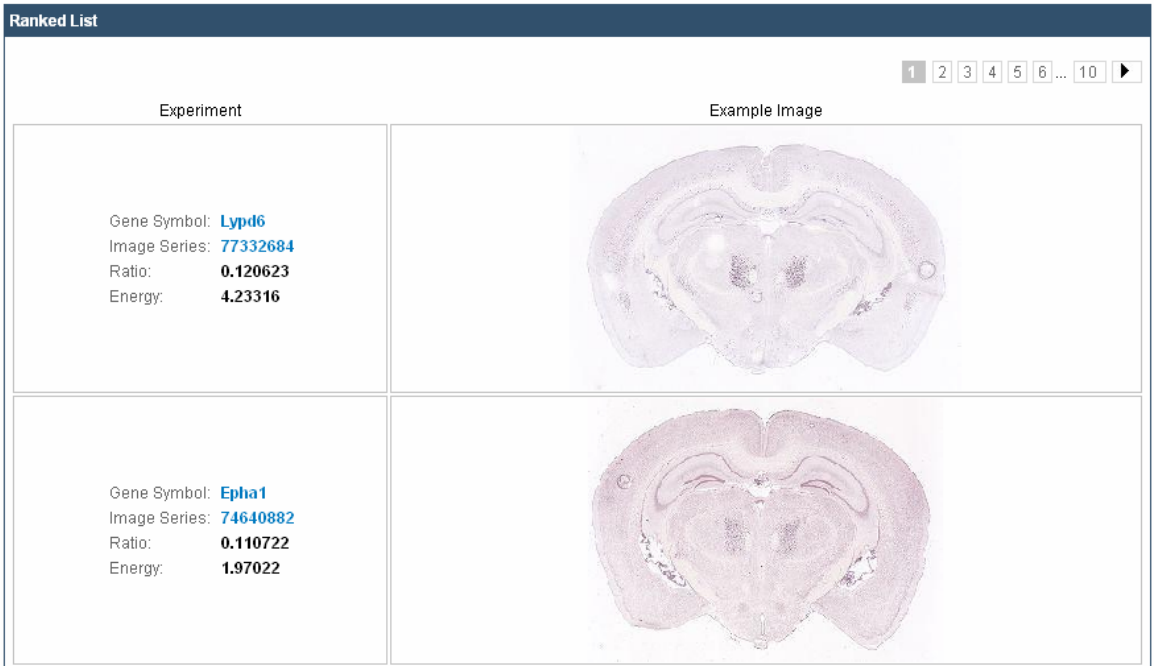

Figure 5: Top two ranking genes derived from correlation map in Figure 4 showing enriched parafascicular nucleus expression. Users can click on the example image to open up the standard multi-resolution image viewer. Click on the image series number to see all the other images in the series. Click on the gene symbol to navigate to other image series related to the gene.

The ***Gene Finder*** Algorithm

For each seed voxel, a fixed correlation threshold is pre-computed and the corresponding correlation map is displayed in the lower panel of Figure 4. Voxels in the top third of the correlation range (orange to red voxels) form the local region of interest *A* and all voxels above threshold make up the domain or larger neighborhood *B* against which local specificity is to be measured. Images in the ABA database are ranked by the ratio of expression energy in *A* over expression energy in *B*. The top 200 ranked image series are returned for manual verification. The AGEA correlation map shown is selected to best represent the local correlation in the spatial neighborhood of the seed and is constructed as follows.

1. The computation proceeds independently for each of 16 regions *R* in the brain defined by unique intra-correlation patterns. These regions are approximately the cortex, hippocampus, striatum, thalamus, olfactory bulb, cerebellar cortex, hypothalamus, midbrain and hindbrain. Further the ventricular areas, medial

habenula, caudoputamen, deep cortical layers, olfactory nerve layer of the olfactory bulb, zona incerta and inferior olive are treated as specialized regions due their unique co-expression profiles. For each of these regions, a second "target domain" region R^+ is also selected for determining the domain region *B* for specificity comparison. For example, for voxels in the cortex, we select the cortex as the specificity comparator. While for highly uniquely expressing areas such as the ventricles and medial habenula we select the whole brain as the specificity comparator. The regions *R* and R^+ are defined using the AGEA bi-tree clustering results. The corresponding AGEA ***Clusters*** permalinks are listed Appendix C.

2. For any initial seed *s* and correlation threshold value *t*, let *N(t,s)* denote the set of AGEA voxels whose correlation values ρ with the seed exceed *t.* In symbols,

$$N(t,s) = \{v_i : \rho(s, v_i) > t\}.$$

For each seed *s* in region *R*, we define an AGEA correlation region *B(s)* that most closely agrees with the target domain R^+ by selecting an optimal threshold *T* chosen so that

$$T = \arg\max_t d\left(R^+, N(t,s)\right)$$

where the distance metric *d* is defined as the Dice similarity metric[1]. The advantage of the Dice measure over simple intersection is that it also penalizes for differences in region size. Now let *B(s) = N(T,s)* represents the set of voxels exceeding this optimal threshold and is represented as the complete set of non-zero (non dark blue) rescaled voxels in Figure 4b.

3. The most local neighborhood *A(s)* of highest correlated voxels with the seed *s,* (corresponding to the deepest orange-red colored voxels in Figure 4b) may be approximated by retrieving the voxels having highest correlation values from *B(s)*. Set

$$A(s) = \{v_j \in B(s) : \rho(v_j, s) > \rho'\}, \text{ where } \rho' = (T+2)/3.$$

4. Together the voxel sets *A(s)* and *B(s)* determine the local and larger anatomic context for AGEA correlation values centered at a fixed seed *s*. We can use the sets *A(s)* and *B(s)* to prioritize genes from the ABA in terms of how specific expression is to the local region as compare to the larger neighborhood*.* For a given gene *g* compute the expression energy E_g in both sets and define the E_{ratio} of the gene with respect to the seed *s* to be

$$E_{ratio}(g,s) = \frac{E_g(A(s))}{E_g(B(s))}$$

Values of the E_{ratio} range from 0.0 (not expressing in *A(s)*) to 1.0 (ideal specificity to *A(s)*) and genes can be ranked by descending values.

Clicking on the "Find Genes" link (highlighted in Figure 4) searches the ABA image data for genes potentially enriched in the local region *A(s)*. Figures 4 and 5 shows an example search for enriched expression in the parafascicular nucleus of the thalamus. In Figure 5 two of the top gene returns are shown for the given search together with their E_{ratio} and $E_g(A(s))$.

[1] Dice similarity between two regions P and Q is given by (2 |P ∩ Q| / (|P| + |Q|)).

In the ***Gene Finder*** search false positives can results from dark artifacts and/or registration inaccuracies causing leakage of expression from neighboring regions of *A(s)*. The latter is particularly problematic when the region *A(s)* is near the boundary of the domain region *B(s)*. On the other hand, false negatives can occur from poor focusing, pale/white bubbles in the region of interest or alternatively dark artifacts and non specific binding occurring in the domain region. Further, to eliminate false positives due to noise, an image series is only included in a list if the average energy in A is greater than 1. It may be possible that a gene with sparse and/or lighter intensity expression in the region of interest might be excluded by this criterion. A good analogy to AGEA ***Gene Finder*** would be an internet search engine where the user must further examine return results to find those most applicable to the query. Additionally, due to nature of gene expression some brain areas have a large number of genes with robust expression while other areas have only a few genes and/or less robust expression.

2. The AGEA Viewer

This section describes the controls of the ***Correlation***, ***Cluster*** and ***Gene Finder*** modes of AGEA.

Correlation Mode

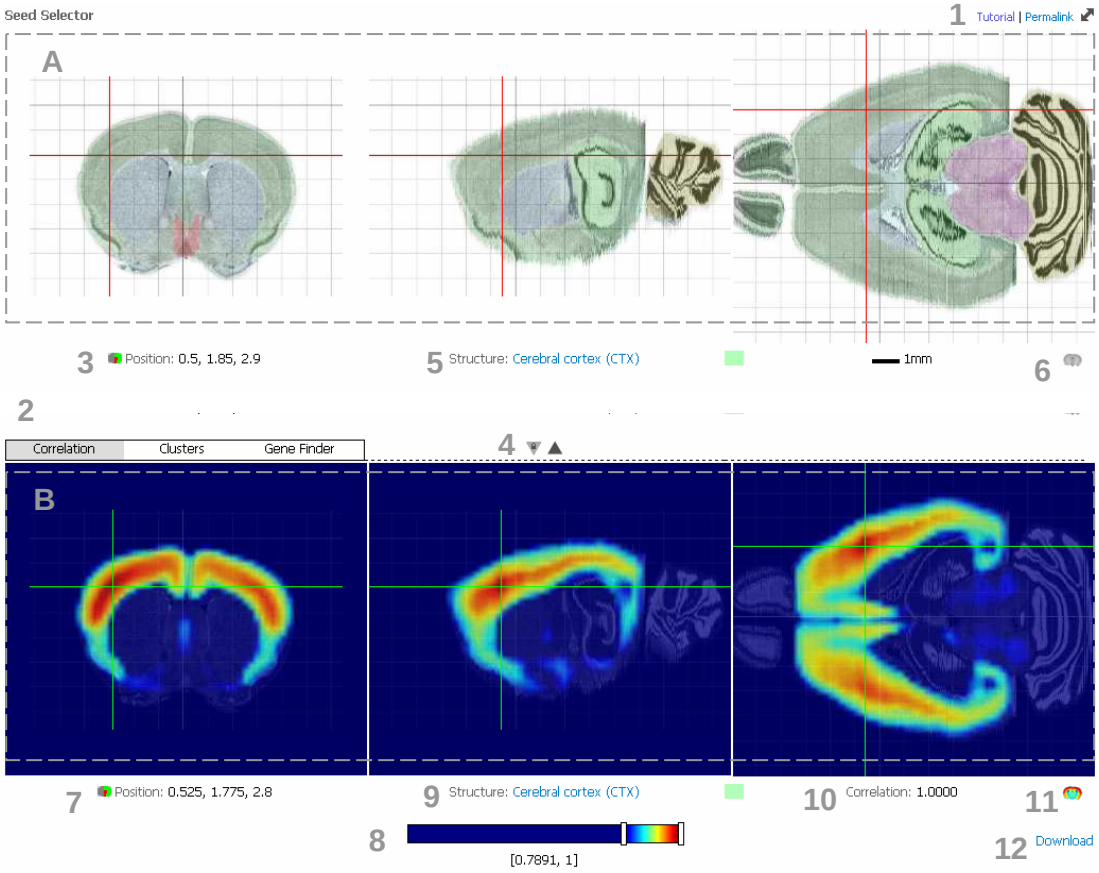

Figure 6. The AGEA interface in *Correlation* mode. The workspace is divided into two major panels: the seed selector panel (top) and selected correlation map panel (bottom) [permalink].

A. **Seed Selector panel:** Use this area to select the seed voxel as marked by the red crosshairs. Either click or click and drag to navigate to a different correlation map.
B. **Map panel:** This area displays the three orthogonal views of the currently selected correlation map in coronal, sagittal and horizontal planes. The green crosshair marks the currently selected voxel. Either click or click and drag to move to a new 3D location. Note that this volume can be navigated in 3D for any selected voxel location from A.

1. **Permalink/Zoom:** Clicking on "Permalink" creates an URL in the address bar that effectively save information about your current viewing state. This URL can be saved for latter access to directly take you back to the current view and color scale settings. Clicking the arrows will toggle between zooming the images to fit in the window and a higher resolution more zoomed mode. At a fixed zoom level, you may need to use the browser scroll bars to view all the images.
2. **Mode Selector:** Click on **Correlation**, **Clusters** or **Gene Finder** to switch the viewer to different modes.
3. **Position/ARA:** The position of the crosshairs in the seed selector in millimeters from bregma. Click on the icon next to the position to view the closest coronal section of the Allen Reference Atlas.
4. **Lock/Sync:** The icon on the left locks the planes shown in the selected expression map B to the same position as the seed map A. Click to toggle the locking behavior. Conversely, click on the right icon to move the seed voxel to the current selected map voxel.
5. **ARA Label:** The structure or structural grouping from the Allen Reference Atlas to which the seed voxel (indicated by red crosshairs) belongs. Click on the name to get information about the structure.
6. **ARA Blend:** This icon toggles blending Allen Reference Atlas structural delineations on the Seed Selector images for direct anatomic comparision. A 1mm grid is also included with the reference atlas colors. The darker lines indicate the origin of the coordinate system, which is at bregma.
7. **Position/ARA:** The position of the crosshairs in the correlation map in millimeters from bregma. Click on the icon next to the position to view the closest coronal section of the reference atlas.
8. **Color scale control:** Use to adjust the false color mapping of the correlation map to threshold the images for regions of higher significance. All voxels with correlation within the select range are rescaled to span the color scale.
9. **ARA Label:** The structure or structural grouping to which the selected voxel (indicated by green crosshairs) belongs. Click on the name to get information about the structure.
10. **Correlation:** Value of the correlation at the currently selected voxel with respect to the seed voxel selected in panel A. This shows numerically how well the target voxel is correlated with the seed.
11. **Nissl Blend:** This icon toggles blending the Nissl reference atlas volume on the correlation map. As with (6), a 1mm grid is shown, and the brighter lines indicate the origin.
12. **Download:** Click to download the currently selected correlation map as raw flat file with numbers saved as floats.

Using the Seed Selector Panel

A user can browse from one map to another by clicking and dragging the red crosshairs in Panel A to the desired location. Panel A is a multi-planar view of the 3D reference atlas showing (left to right) one coronal, one sagittal and one horizontal slice. The red crosshairs mark the same 3D position in all views, hence clicking or dragging in one plane automatically updates the other two planes. The red crosshairs also mark the seed voxel of the correlation map being displayed in Panel B. Any movement of the red crosshairs in A also updates the multi-planar view of the correlation map in B.

Using the Map Panel

A user can explore the 3D correlation map by clicking and dragging the green crosshairs in Panel B to different 3D locations. Panel B is a multi-planar view of the currently selected correlation map displaying (left to right) one coronal, one sagittal and one horizontal slice. The green crosshairs mark the same 3D position in all views, hence clicking or dragging in one plane updates the other two planes.

In the AGEA viewer, correlation is spatially visualized by assigning one of 256 colors to a range of correlation values. The color assignment is determined by the color scale slider bar (8). The vertical white bars define the lower and upper bounds of the correlation values to be mapped. The range of values between the bars is divided into 256 bins and each bin is assigned to a color according to the jet color scheme, where low values are assigned shades of blues. As the correlation values increase the colors move through green, yellow, and then orange. Finally, high values are assigned shades of red.

As a user moves the location of the green crosshairs, the information below the correlation maps changes to reflect structure (or structural grouping) membership and the correlation coefficient (10) of the currently selected voxel.

Clusters Mode

In the ***Clusters*** mode (Figure 7), the user selects a given voxel from the reference atlas (Panel A) and descends the hierarchical bi-tree to the leaf represented by that voxel. Initially the map in Panel B shows the highest level or root of the tree. By using the *Tree Depth* control (13) the nodes along the bifurcation path to the leaf can be explored. By moving this control to the right eventually one obtains a cluster consisting of the single seed voxel. As with other modes in AGEA the cluster nodes can be explored in 3D by moving the crosshairs.

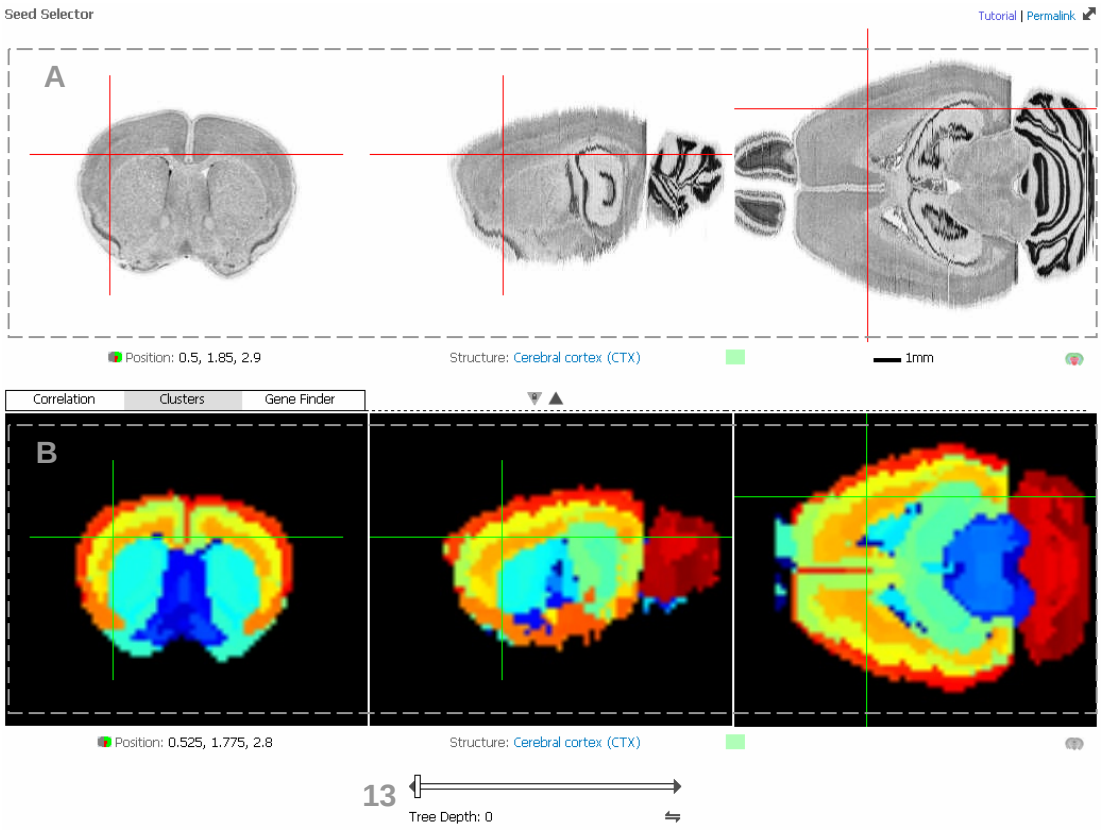

Figure 7. The AGEA interface in *Clusters* mode. The workspace is divided into two major panels: the voxel/path selector panel (top) and visualization of the voxels associated the currently selected node (bottom) [permalink].

A. **Seed Selector panel:** Use this area to select the tree path to traverse. The path is defined by the selected leaf node (corresponding to selected voxel) as marked by the red crosshairs and its unique path to the root node. Either click or click and drag to navigate to a different tree path.
B. **Map panel:** This area display three orthogonal views of the currently selected tree node cluster. Voxels associated with the node are systematically color coded using a blue-to-red ("jet") color scheme. The green crosshair marks the currently selected viewing planes. Either click or click and drag to move to a new 3D location.

13. **Tree Navigation:** Move slider to traverse the path. Moving the slider to the right corresponds to moving towards the leaf node. Moving the slider to the left corresponds to walking towards the root node. At each node, you can take a short-cut to the sibling node via the double arrow icon. This action takes you to the median voxel of the sibling cluster node while retaining the current tree depth.

Gene Finder Mode

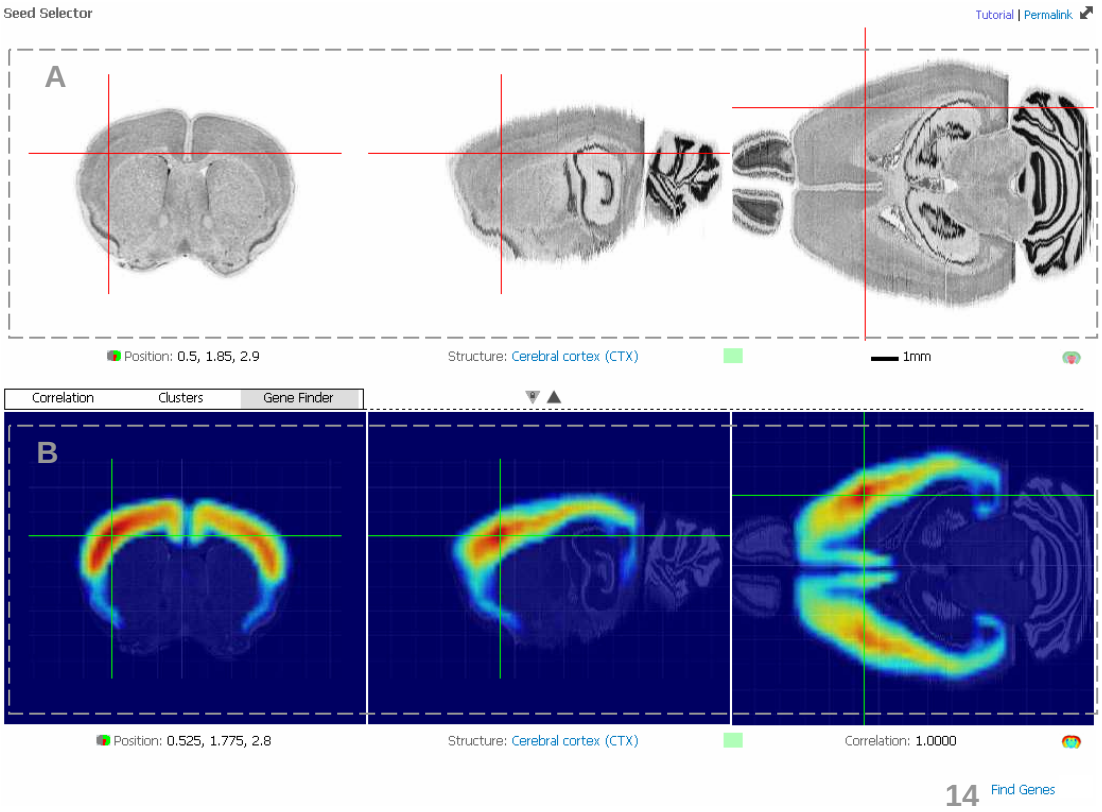

Figure 8. The AGEA interface in *Gene Finder* mode. The workspace is divided into two major panels: the seed selector panel (top) and pre-computed selected correlation map panel (bottom) [permalink].

A. **Seed Selector panel:** Use this area to select the seed voxel as marked by the red crosshairs. Either click or click and drag to navigate to a different correlation map.
B. **Map panel:** This area displays three orthogonal views of the currently selected correlation map at a fixed pre-computed threshold (as explained in Section 1) used for searching the ABA database. The green crosshair marks the currently selected voxel. Either click or click and drag to move to a new 3D location.

14. **Find Genes:** Click to find genes potentially enriched in the local correlated region as defined by the currently selected correlation map. Voxels in the orange to red range (top third of the range) are considered the local region of interest while all voxels above threshold cutoff (all non dark blue voxels) forms the domain region (see Section 1: Gene Finder Algorithm). An energy ratio method is use rank the images in the ABA database. The top 200 image series are returned by rank order in a separate page.

Gene Finder Return Page

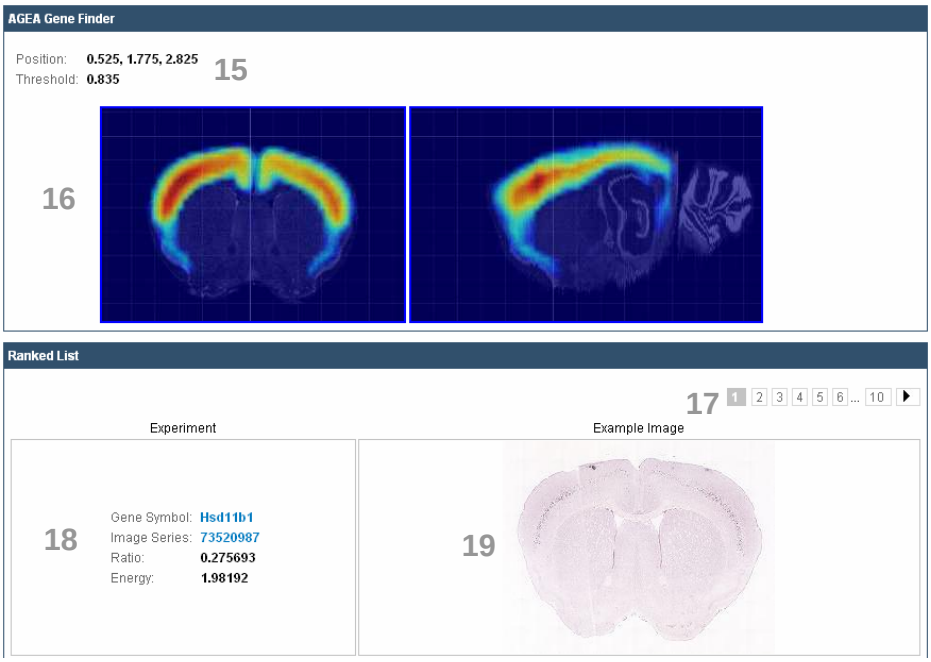

Figure 9. Top portion of an example *Gene Finder* return page. The top 200 ranked image series from an energy ratio search of the ABA database is displayed as a table with level matched thumbnail image.

15. **Position/Threshold:** The position (in millimeters from bregma) and threshold of the correlation map use to define the local region of interest. Note: This correlation map displayed on the ***Gene Finder*** return page can be reproduced in ***Correlation*** mode by setting the threshold to the specific value shown here.
16. **Correlation Map:** Coronal and sagittal plane of the correlation map containing the seed voxel
17. **Page Navigation:** 20 search results are returned per page. Click on page number or arrow to browse more returns.
18. **Image Series:** Click on the image series number to view all images associated with the series. Click on the gene symbol to get information about the gene and navigate to other image series assayed for this gene. The energy expression with the local region of interest and the specificity ratio is also reported for each return.
19. **Example Image:** Thumbnail image of the median section from the coronal series where the seed voxel is located. Click on the image to launch the multi-resolution image viewer. Note that if there is a large discrepancy in cutting angle or specimen bending, it is possible that the selected region of interest may not appear in the “example” image. In those cases, the neighboring sections can be viewed by clicking on the image series number.

3. Examples of using *AGEA*

Laminar organization of the cerebral cortex

Follow this link to navigate to a correlation map where the seed voxel is within Layer 5 of the cerebral cortex as shown in Figure 10a.

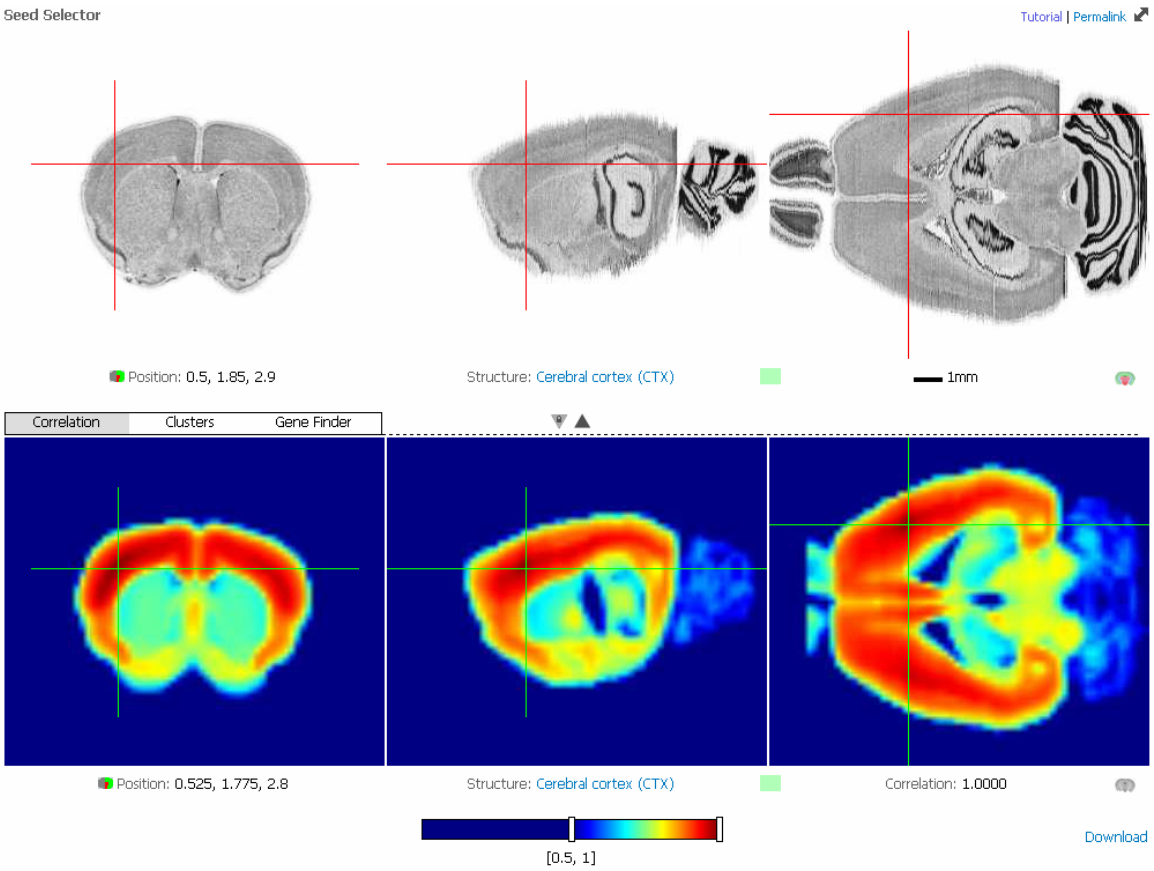

Figure 10a. An AGEA spatial correlation map with seed voxel selected in Layer 5 of the cortex [permalink]

Clicking on the icon to the left of Position opens the Allen Reference Atlas (ARA) coronal atlas plate closest to the position shown (Figure 10b). The coordinates are given in millimeters relative to bregma and are along the following axes: rostral-caudal, dorsal-ventral, lateral-medial. In Figure 10a, clicking on the icon automatically opens the plate at 0.545 mm as shown in the lower right-hand corner of Figure 10b. The arrow at the upper left of Figure 10b indicates the coordinates (1.85 mm, 2.924 mm), which can be compared to coronal images in Figure 10a. Note that the ARA is symmetric about the midline as it is formed by reflecting one hemisphere; whereas the AGEA atlas on the other hand is based on the full Nissl images and hence is not symmetric. As a result the corresponding coordinates indicated are only an approximation.

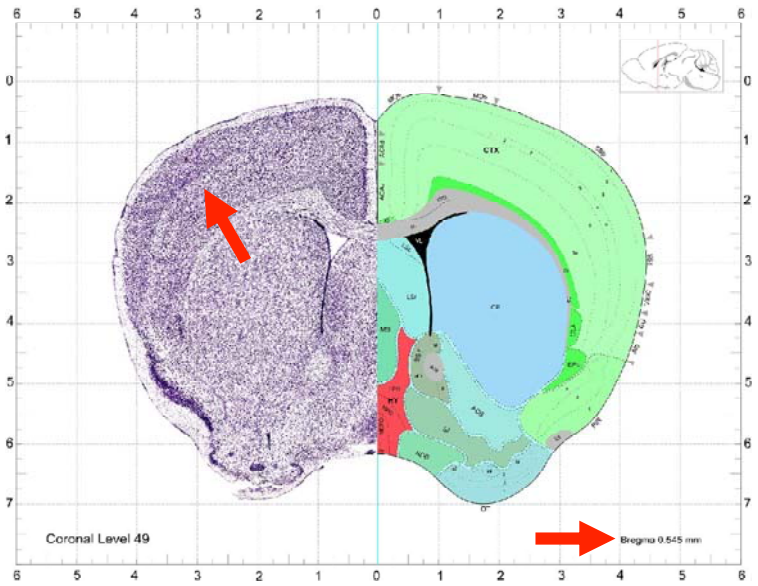

Figure 10b. The Allen Reference Atlas coronal atlas plate [URL] corresponding to the cursor location in Figure 10a.

By default, the correlation display range is set to [0.5,1]. On the threshold control slider, push the white bar representing the lower limit towards the white bar representing the upper limit.

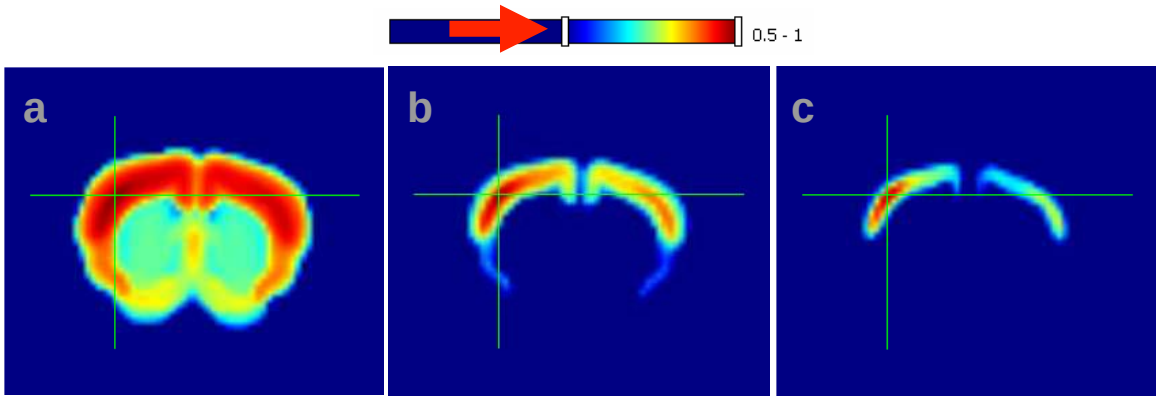

Figure 11. Changing Correlation Threshold. The effect of changing the threshold color scale bar for a voxel seed in Layer 5 of the cortex. Contracting toward a more restricted dynamic range focuses the spatial relationship first on the cortex–wide scale and then subsequently on a more regional intra-layer 5 scale.

Changing the dynamic range of the correlation map provides a multi-scale view of the spatial relationship within the cerebral cortex. In the first panel Figure 11a, the mass of orange/red voxels indicates the cerebral cortex as a co-expressing unit. A subtle change from dark orange and light orange reflects the demarcation between the isocortex (or neocortex) and the olfactory areas. As we decrease the mapping range to [0.85,1] in Figure 11b, the existence of a coherent laminar layer 5 becomes more apparent. By decreasing the range even further to [0.92,1] in Figure 11c, a more local region in the somatosensory cortex emerges. This example illustrates the multiple scales on which gene expression correlation can be analyzed.

Note: As we decrease the mapping range to only focus on the top end of the correlation scale, a noticeable asymmetry between the left and right side can be observed. Rather than being a result of actual left-right expression dimorphism this asymmetry is usually due to either a) correlation as a result of the spatial averaging inherent in the registration process, or b) artifacts in the

production of the ISH imagery. Any artifacts in the ISH images (dark/light bubbles, tears, folds, inhomogeneity gradient, etc.) are likely to have only a local spatial effect and thus typically do not affect the corresponding location in the other hemisphere in exactly the same way. The net result of these effects is typically a lower correlation value in the corresponding location in the other hemisphere and the false appearance of a higher correlated local cluster around the seed. One of the benefits of only using the coronal images from the ABA data is that information from both hemispheres is represented. Correlation patterns from the other hemisphere can be used to confirm if a local cluster is real as the same pattern should be present, albeit at lower correlation values.

To explore the laminar organization of expression in the cortex more fully, follow this link to navigate to a correlation map where the seed voxel is within Layer 2/3 of the cerebral cortex with a correlation mapping range of [0.85, 1]. The result should agree with Figure 12 depicting the bilateral laminar pattern in coronal and horizontal sections.

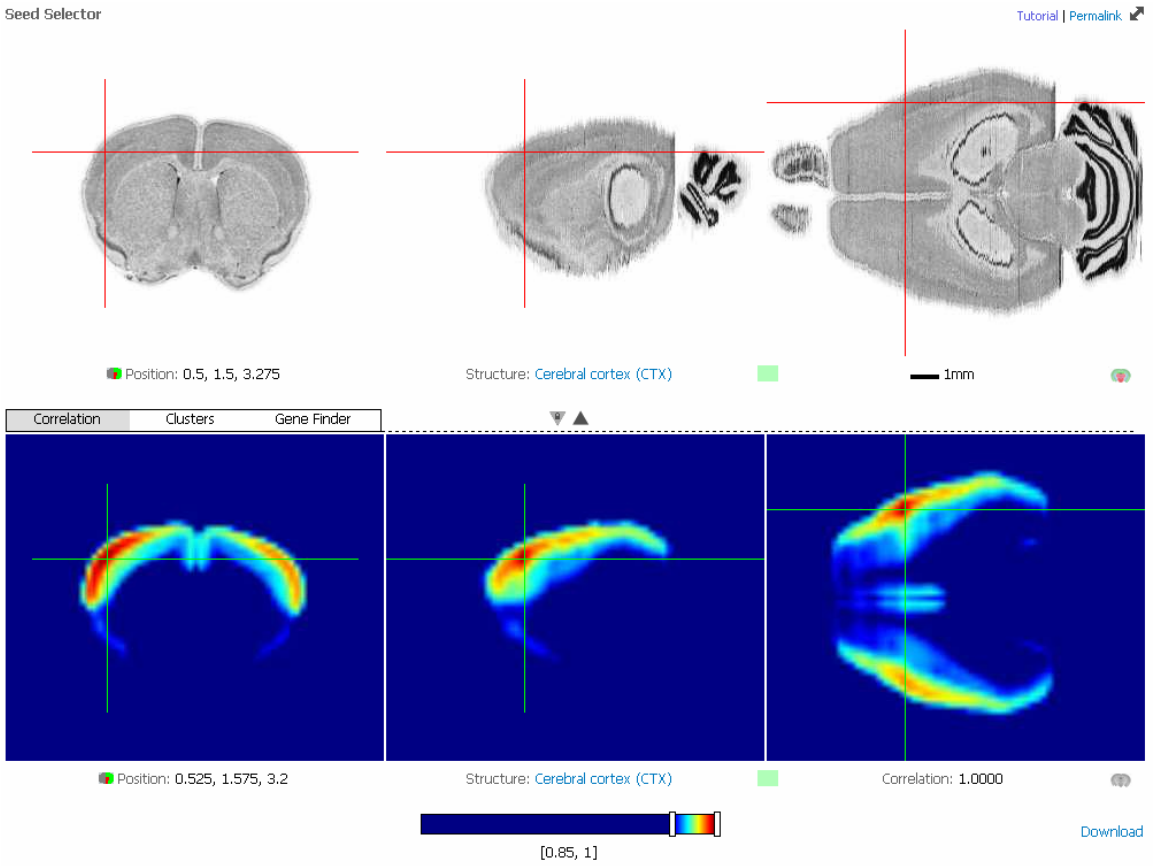

Figure 12. AGEA display with seed voxel in Layer 2/3 of the neocortex and dynamic range of color scale set to [0.85,1] [permalink]

On the coronal view of the 3D reference atlas, click and drag the red crosshairs perpendicular to the cerebral cortex surface (as indicated by the red arrow in Figure 13). The effect of laminar over areal dominance in the isocortex is characteristic of gene expression in the ABA dataset and well illustrated by correlation in the AGEA maps.

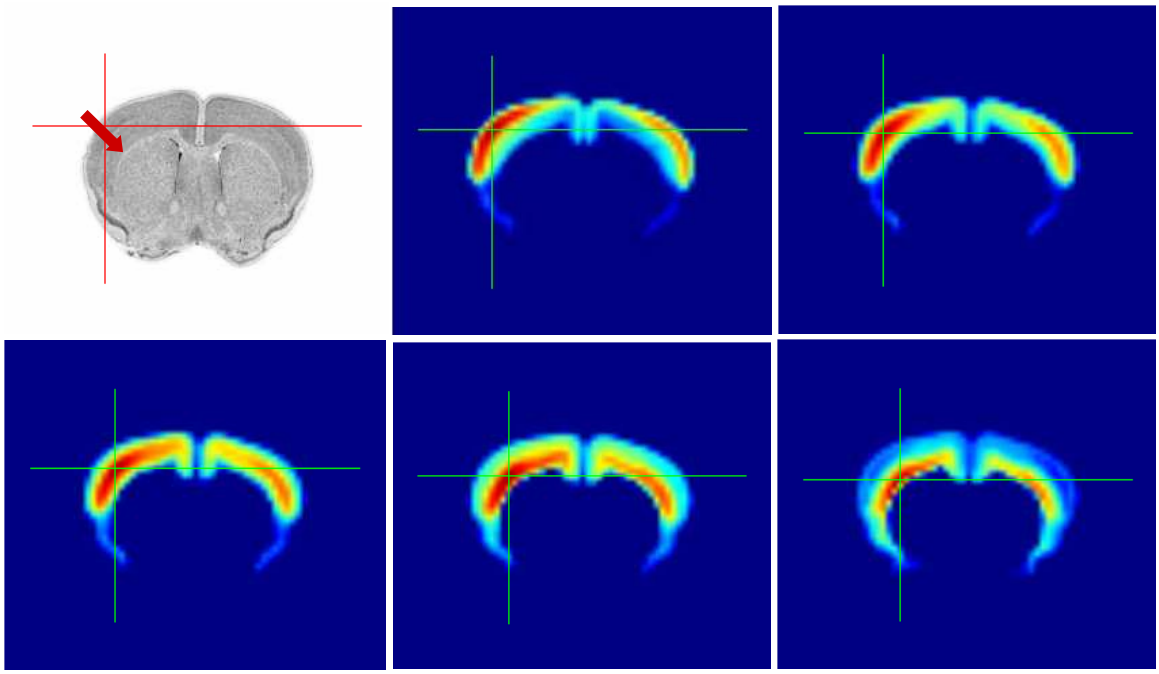

Figure 13. Moving the cursor across the laminar regions of the neocortex and perpendicular to the outer boundary reveals the laminar structure of neocortex as depicted by gene expression.

Co-expression with the bed nuclei of the stria terminalis (BST)

Follow this link to navigate to a correlation map where the seed voxel is located in the principal nucleus of the bed nuclei of the stria terminalis with a correlation mapping range of [0.684, 1].

Browsing through the 3D correlation map shown in Figure 14 reveals strong co-expression between BST and the hypothalamus (HY) and the medial amygdalar nucleus (MEA). The striatum-like amygdala (CEA, MEA) shares close bi-directional neural connections with the BST, and together they provide predominant inputs to the hypothalamus to control somatic and visceral motor activities associated with motivated behavior [2]. Also of note is differential expression in other regions of the brain manifesting as local cool and warm spots warranting further investigation. In particular, other higher correlated regions are the magnocellular nucleus (MA) in the pallidum, the red nucleus (RN) in the midbrain, and the nucleus of the solitary tract (NTS) in the medulla.

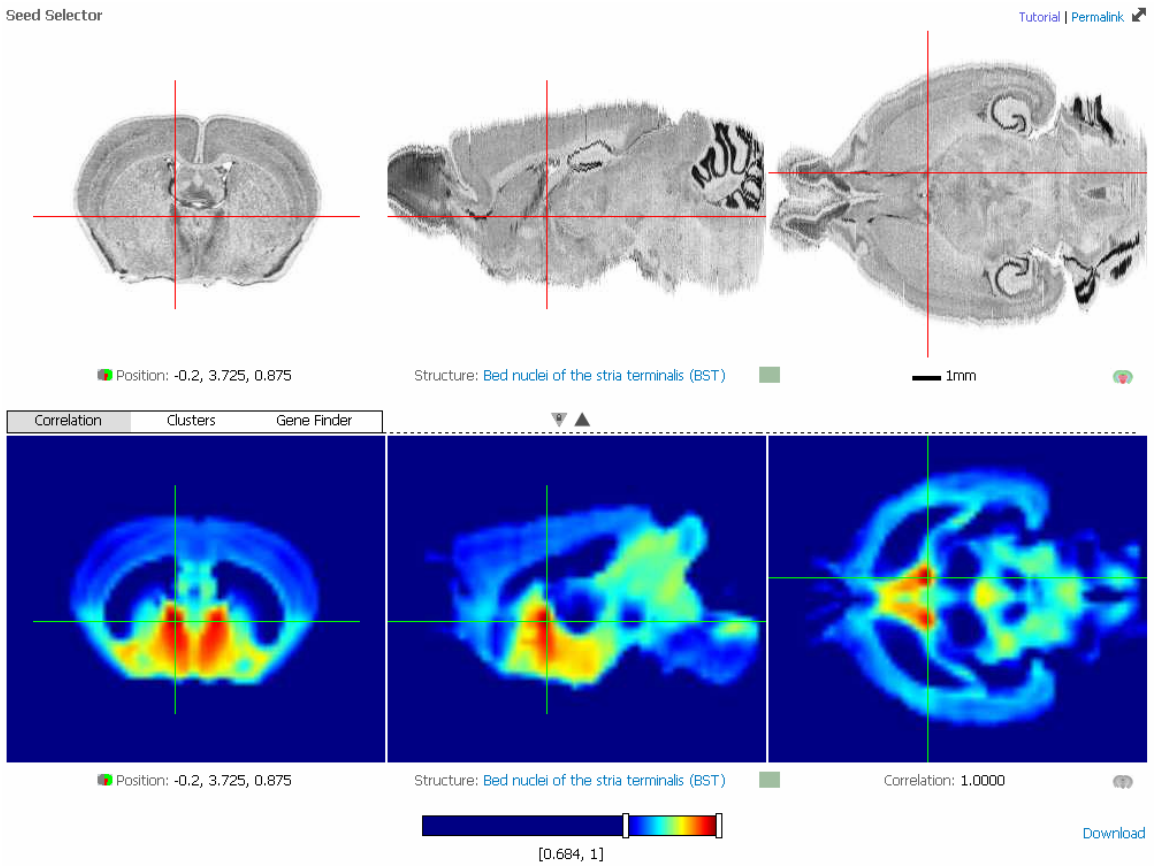

Figure 14. AGEA correlation map seeded at principal nucleus of the bed nuclei of the stria terminalis [permalink].

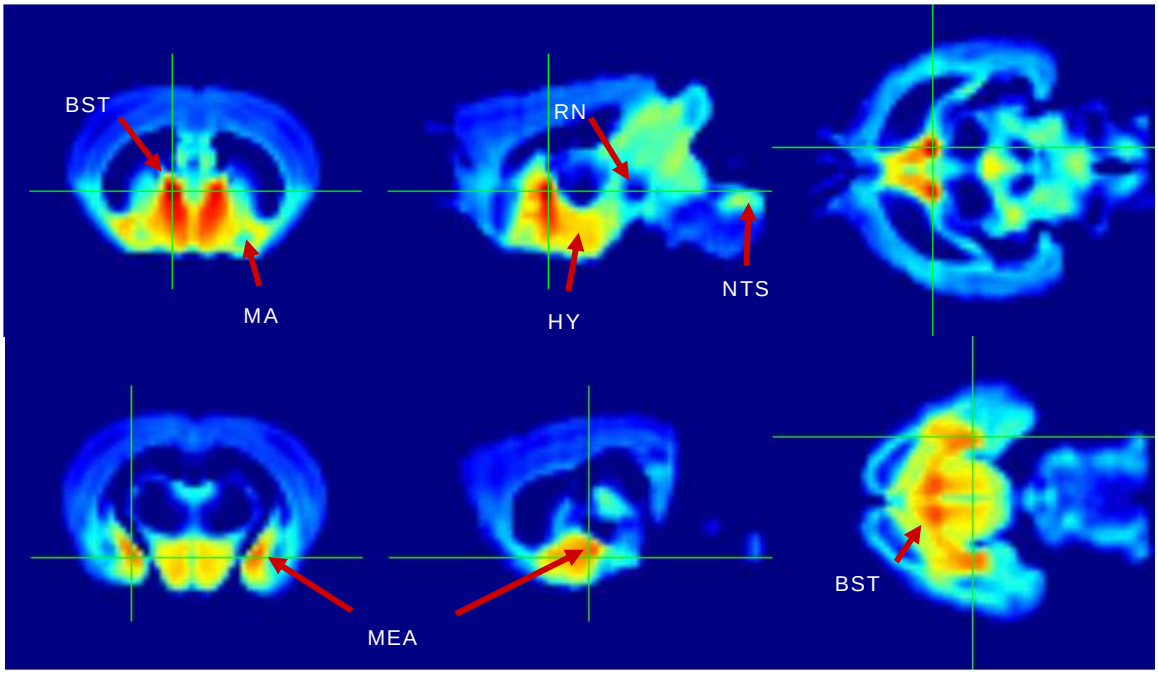

Figure 15. Strong co-expression between BST and the hypothalamus (HY) and the medial amygdalar nucleus (MEA) shown in AGEA correlation plots. (BST, bed nuclei of the stria terminalis, MA, magnocellular nucleus, RN, red nucleus, NTS, nucleus of the solitary tract.

Searching for structural markers

Using the ***Gene Finder*** feature Figure 16 shows example search regions for initial seed voxels in four locations (a) dorsal nucleus raphé (DR), (b) suprachiasmatic nucleus (SCH), (c) medial habenula (MH) and (d) paraventricular nucleus of the thalamus (PVT). Permalinks to the seed voxels and other metadata URL are listed in Appendix A. Example results from these searches are shown in Figure 17.

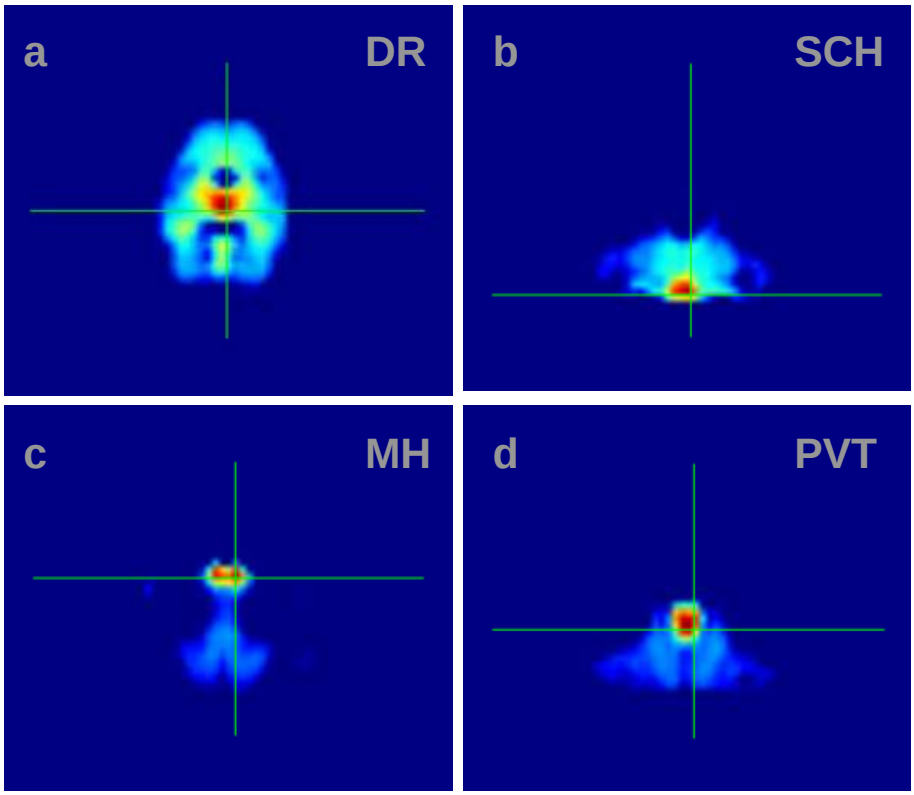

Figure 16. Sample AGEA *Gene Finder* regions for seed voxels located in the (a) dorsal nucleus raphé (DR), (b) suprachiasmatic nucleus (SCH), (c) medial habenula (MH) and (d) paraventricular nucleus of the thalamus (PVT).

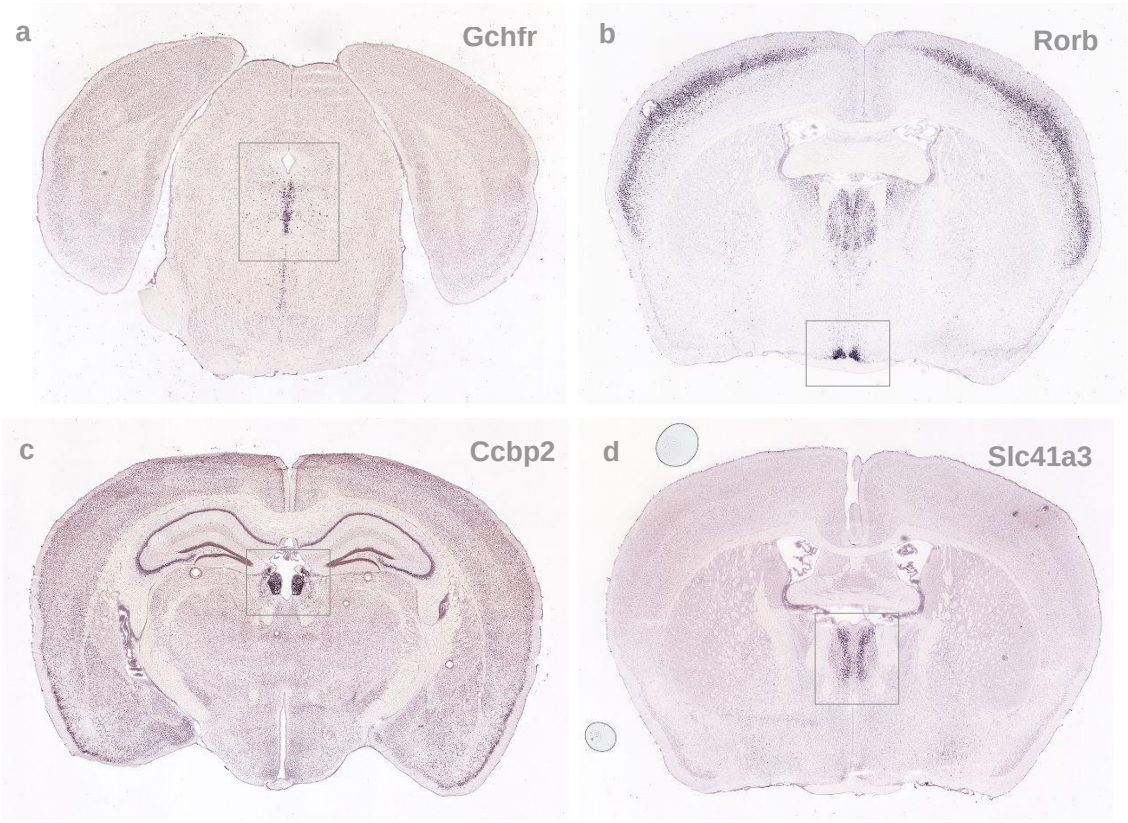

Figure 17. Sample AGEA *Gene Finder* results for seed voxels located in the (a) dorsal nucleus raphé (DR), (b) suprachiasmatic nucleus (SCH), (c) medial habenula (MH) and (d) paraventricular nucleus of the thalamus (PVT).

Transcriptome derived anatomical ontology

In AGEA **Clusters** mode, a user can explore an expression driven spatial organization of the adult mouse brain. The full AGEA set of 51,533 voxels is organized as a hierarchical binary tree (bi-tree) [4] where each node represents an anatomical compartment of correlated gene expression at multiple scale. This bi-tree can also be viewed as a novel transcriptome derived anatomical ontology reflecting grouping of anatomical regions by frequency of co-expression. Thus, the ***Clusters*** viewer is a potentially useful research tool to investigate the difference between classical anatomic boundaries and gene expression patterns. Effective visualization of this large clustered data structure is a challenging problem. Figure 18 shows one possible rendering of the complete bi-tree as a fractal tree to enable a more compact representation.

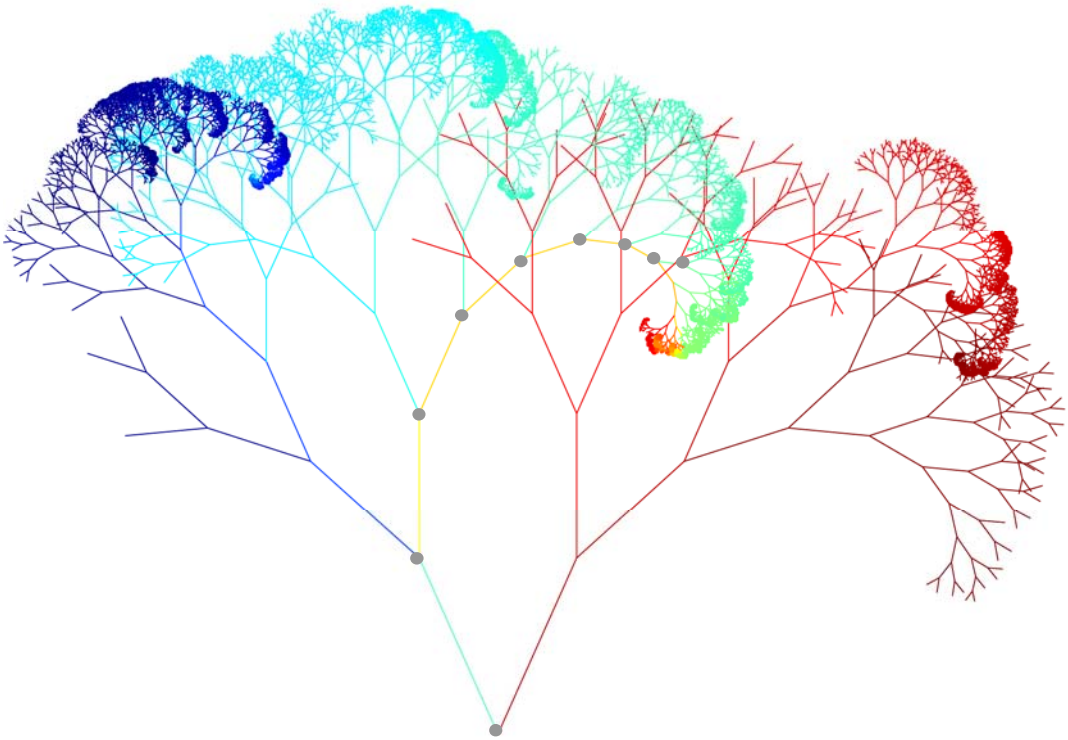

Figure 18. The AGEA bi-tree rendered as fractal tree. At each bifurcation the branch length is reduced by a factor of 0.75 and offset by 24 degrees. Each branch is colored with the same color as the median voxel of the node being represented. Grey circles denotes the path from the root node to a voxel located in the hippocampus [permalink].

In the viewer, the tree navigation paradigm is to select a voxel from the 3D reference atlas (Panel A, Figure 7). As each voxel in the brain is associated with one leaf node, the selected voxel effectively defines one root-to-leaf path in the bi-tree. Once this path is selected, a slider bar control is used to traverse from the root node to the leaf node. Further, the double-arrow icon provides a shortcut to the sibling node. In the bi-tree clustering model, the collection of voxels in each node is divided into two sets with maximum dissimilarity with respect to correlation. Therefore the sibling node *S* of a node *N* represents those voxels whose correlation profile is most unlike *N* given that a decision to split the set into two groups is made. For example, Figure 19 shows the first 9 bifurcations for a path defined by selecting a seed voxel in the hippocampus. Nodes along the path are shown in column A and the corresponding siblings in column B. This path is also graphically illustrated in Figure 18.

Each node is visualized in its 3D context by systematic color coding of the associated voxels in the Map panels. Each leaf node (and therefore each voxel on the brain) is assigned a "depth first" global ordering such that all nodes always represent a set of voxels with contiguous ordering. The voxels of a node are assigned a color based on the 'jet' color scheme with the lowest ordered voxels assigned shades of blues. The colors then blend through green, yellow and orange. Finally, the highest ordered voxels are assigned shades of reds. This global order base coloring scheme basically provides a preview of the sub-tree organization.

The top row of Figure 19 represents the root node (depth 0) spanning all voxels in the brain. The color coding scheme shows a gross decomposition of the brain similar to classical anatomy. At the first bifurcation (depth 1) the cerebellar cortex (1B) splits away from the rest of the brain (1A). Noteworthy is the grouping of the cerebellar nuclei with the hindbrain instead of with the cerebellar cortex and hence in the bi-tree a coherent cerebellum node does not exist. At depth 2,

the cerebrum (2A) divides off from a large hypothalamus, thalamus, midbrain and hindbrain cluster (2B). At the next bifurcation (depth 3), a large striatum cluster (3B) is demarcated noting that this cluster does not include the lateral septal complex in contrast to the ARA ontology.

At depth 4, a small cluster of voxels overlapping the lateral olfactory tract in the olfactory bulb divides off (4B). This is followed by voxels in the stratum lacunosum-moleculare of the hippocampus (5B) and the main and accessory olfactory bulb (6B). At depth 7, a few boundary voxels overlapping the top edge of the cortex split off as a small cluster (7B). A large size hippocampus cluster (8A) breaks from the cortex (8B) at depth 8. The color coding scheme indicates further sub-organization of the hippocampus into the dentate gyrus, CA1 and CA3 fields. Also significant is the absence of the ventral tip of the hippocampus which has been grouped with other cortical regions. This reflects the fact the ventral tip often co-express independently from the dorsal hippocampus and co-express frequently with surrounding cortical areas. AGEA paths and sibling links are given in Appendix B.

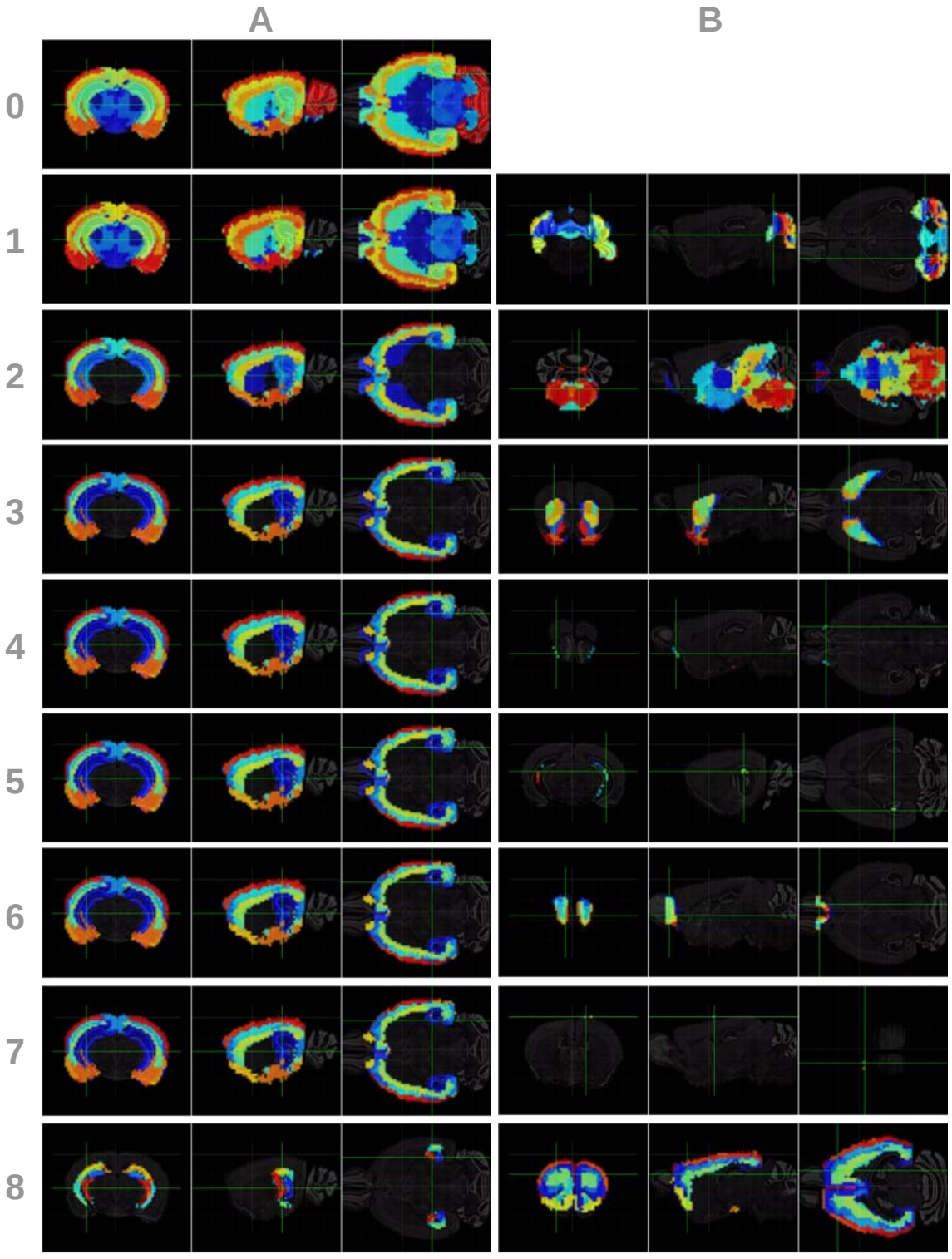

Figure 19. Traversing a bi-tree path. First 9 bifurcations on the path from root to a leaf node located in the hippocampus. Column A shows the node along the path while column B shows the corresponding sibling node. This same path is graphically illustrated in Figure 17. Permalinks to AGEA is provided in Appendix B.

4. References

[1] L Ng, S Pathak, L Kuan, C Lau, HW Dong, A Sodt, C Dang, B Avants, P Yushkevich, J. Gee, D Haynor, E Lein A Jones, M Hawrylycz: **Neuroinformatics for Genome-Wide 3-D Gene Expression Mapping in the Mouse Brain**. IEEE/ACM Transaction on Computational Biology and Bioinformatics, vol 4(3), pp 382-393, Jul-Sept, 2007.

[2] R. Gaykema, C Chen, L Goehler, **Organization of immune-responsive medullary projections to the bed nucleus of the stria terminalis, central amygdala, and paraventricular nucleus of the hypothalamus: Evidence for parallel viscerosensory pathways in the rat brain**. Brain Research 1130, pp 130-145, 2007

[3] C.-K. Lee, S.M. Sunkin, C. Kuan, C.L. Thompson, S. Pathak, L.Ng, C.Lau, S. Fischer, M. Mortrud, C. Slaughterbeck, A. Jones, E. Lein, M. Hawrylycz, **Quantitative methods for genome-scale analysis of *in situ* hybridization and correlation with microarray data**, Genome Biology, 2008, R:23

[4] M. Sultan, D.A. Wigle, C.A. Cumbaa, M. Maziarz, J. Glasgow, M.S. Tsao, I, Jurisica, **Binary tree-structured vector quantization approach to clustering and visualizing microarray data**, Bioinformatics, Vol. 18, Suppl. 1, 2002, S111-S119

Appendix A

Structure URL, AGEA Gene Finder permalink, gene and image series URLs associated with Figures 16 and 17

a) structure DR, gene Gchfr
http://www.brain-map.org/aba/mouse/atlas/coronal/DR.html
http://www.brain-map.org/agea/all_coronal?geneFinder&seedPoint=9659,4272,5725&mapPoint=9659,4272,5725
http://www.brain-map.org/aba/mouse/brain/Gchfr.html
http://www.brain-map.org/aba/mouse/brain/Gchfr/74511806/thumbnails.html

b) structure SCH, gene Rorb
http://www.brain-map.org/aba/mouse/atlas/coronal/SCH.html
http://www.brain-map.org/agea/all_coronal?geneFinder&seedPoint=5853,6741,5603&mapPoint=5853,6741,5603
http://www.brain-map.org/aba/mouse/brain/Rorb.html
http://www.brain-map.org/aba/mouse/brain/Rorb/79556597/thumbnails.html

c) structure MH, gene Ccbp2
http://www.brain-map.org/aba/mouse/atlas/coronal/MH.html
http://www.brain-map.org/agea/all_coronal?geneFinder&seedPoint=7306,3350,5481&mapPoint=7306,3350,5481
http://www.brain-map.org/aba/mouse/brain/Ccbp2.html
http://www.brain-map.org/aba/mouse/brain/Ccbp2/74724637/thumbnails.html

d) structure PVT, gene Slc41a3
http://www.brain-map.org/aba/mouse/atlas/coronal/PVT.html
http://www.brain-map.org/agea/all_coronal?geneFinder&seedPoint=5611,4804,5522&mapPoint=5611,4804,5522
http://www.brain-map.org/aba/mouse/brain/Slc41a3.html
http://www.brain-map.org/aba/mouse/brain/Slc41a3/73512484/thumbnails.html

Appendix B

AGEA Cluster permalinks associated with Figure 19

Nodes on root-to-leaf path

0: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7957,4033,8375&mapPoint=7957,4033,8375&treeLevel=0
1: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7957,4033,8375&mapPoint=7957,4033,8375&treeLevel=1
2: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7957,4033,8375&mapPoint=7957,4033,8375&treeLevel=2
3: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7957,4033,8375&mapPoint=7957,4033,8375&treeLevel=3
4: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7957,4033,8375&mapPoint=7957,4033,8375&treeLevel=4
5: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7957,4033,8375&mapPoint=7957,4033,8375&treeLevel=5
6: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7957,4033,8375&mapPoint=7957,4033,8375&treeLevel=6
7: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7957,4033,8375&mapPoint=7957,4033,8375&treeLevel=7
8: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7957,4033,8375&mapPoint=7957,4033,8375&treeLevel=8

Sibling Nodes

0: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=2600,4600,7400&mapPoint=2600,4600,7400&treeLevel=0
1: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=11000,3600,4000&mapPoint=11000,3600,4000&treeLevel=1
2: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=12200,5200,5200&mapPoint=12200,5200,5200&treeLevel=2
3: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=4400,4000,7400&mapPoint=4400,4000,7400&treeLevel=3
4: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=2400,4800,7200&mapPoint=2400,4800,7200&treeLevel=4
5: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=8400,3400,2800&mapPoint=8400,3400,2800&treeLevel=5
6: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=1800,4200,6400&mapPoint=1800,4200,6400&treeLevel=6
7: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=5800,1000,4600&mapPoint=5800,1000,4600&treeLevel=7
8: http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=3400,2800,7200&mapPoint=3400,2800,7200&treeLevel=8

Appendix C

AGEA ***Clusters*** permalinks used for computing ***Gene Finder*** correlation thresholds as described in Section 1: The Gene Finder Algorithm. For each region, the first URL defines *R* and the second URL defines R^+, the target domain region.

1. CBX
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=12250,2550,7275&mapPoint=12250,2550,7275&treeLevel=1
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=12250,2550,7275&mapPoint=12250,2550,7275&treeLevel=0

2. ventricles
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7400,4475,8525&mapPoint=7400,4475,8525&treeLevel=4
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7400,4475,8525&mapPoint=7400,4475,8525&treeLevel=0

3. onl
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=1800,3600,5800&mapPoint=1800,3600,5800&treeLevel=6
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=1800,3600,5800&mapPoint=1800,3600,5800&treeLevel=0

4. MH
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=6550,3425,5775&mapPoint=6425,3350,5775&treeLevel=8
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=6550,3425,5775&mapPoint=6425,3350,5775&treeLevel=0

5. STR
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=4800,4200,7800&mapPoint=4800,4200,7800&treeLevel=3
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=4800,4200,7800&mapPoint=4800,4200,7800&treeLevel=3

6. MOB
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=1800,3800,4600&mapPoint=1800,3800,4600&treeLevel=6
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=1800,3800,4600&mapPoint=1800,3800,4600&treeLevel=6

7. TH
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7400,3600,7000&mapPoint=7400,3600,7000&treeLevel=7
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7400,3600,7000&mapPoint=7400,3600,7000&treeLevel=7

8. HIP
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7400,2000,6800&mapPoint=7400,2000,6800&treeLevel=8
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7400,2000,6800&mapPoint=7400,2000,6800&treeLevel=7

9. CTX
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7400,1400,8400&mapPoint=7400,1400,8400&treeLevel=8
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7400,1400,8400&mapPoint=7400,1400,8400&treeLevel=8

10. ZI
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7400,5150,7000&mapPoint=7400,4975,7075&treeLevel=31
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7400,5150,7000&mapPoint=7400,4975,7075&treeLevel=9

11. CP
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=4800,4200,7800&mapPoint=4800,4200,7800&treeLevel=10
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=4800,4200,7800&mapPoint=4800,4200,7800&treeLevel=10

12. HY
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7400,5725,5825&mapPoint=7400,5725,5825&treeLevel=11
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7400,5725,5825&mapPoint=7400,5725,5825&treeLevel=11

13. CTX-deep
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7400,2000,8400&mapPoint=7400,2000,8400&treeLevel=12
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=7400,2000,8400&mapPoint=7400,2000,8400&treeLevel=12

14. IO
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=12150,7025,5725&mapPoint=12150,7025,5725&treeLevel=15
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=12150,7025,5725&mapPoint=12150,7025,5725&treeLevel=14

15. HB
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=10825,5675,5850&mapPoint=10825,5675,5850&treeLevel=16
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=10825,5675,5850&mapPoint=10825,5675,5850&treeLevel=16

16. MB
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=8600,4125,5725&mapPoint=8600,4125,5725&treeLevel=16
http://www.brain-map.org/agea/all_coronal?clusters&seedPoint=8600,4125,5725&mapPoint=8600,4125,5725&treeLevel=16