ALLEN INSTITUTE *for* BRAIN SCIENCE

ALLEN Mouse Brain Connectivity Atlas

TECHNICAL WHITE PAPER:
INFORMATICS DATA PROCESSING

OVERVIEW

The Allen Mouse Brain Connectivity Atlas is a compendium of high-resolution images of axonal projections from targeted neuronal populations throughout the brain. These projection data are generated by two-photon serial imaging, and the resultant high resolution 2D images are processed by the informatics data processing (IDP) pipeline developed by the Allen Institute (Ng *et al.*, 2007; Dang *et al.*, 2007). The IDP is designed to enable the navigation, analysis and visualization of this large and complex dataset to identify projectional relationships and provide novel insights into how neurons are connected in the brain.

In addition to the two-photon serial imaging data, more recent experiments also include intrinsic signal imaging (ISI) data to enable functional identification of V1 and higher visual areas for more precise targeting of viral injections. The ISI data undergoes processing to enable mapping to the Allen Mouse Common Coordinate Framework (CCF) and analysis by functional area. This white paper describes the informatics data processing for the integration of the two-photon serial tomography and the intrinsic signal imaging into the CCF to support analysis and navigation tools.

PROCESSING OF SERIAL TWO-PHOTON IMAGING

The informatics data processing pipeline consists of the following components: a Preprocessing module, a 3D reference model, an Alignment module, a Detection module, a Gridding module, and a Structure Unionizer module. The output of the pipeline is quantified signal values at a grid voxel level and a structure level according to the Allen Reference Atlas ontology. The grid level data are used downstream to provide on-the-fly data and correlative search services and to support visualization of projectional relationships.

In particular, the informatics data processing supports the following features in the Web application (**Figure 1**):

1. A **Segmentation Mask** for each projection image shows areas of detected signal with color-coding of each segmented class type.
2. A **Projection Summary** for each experiment and interactive bar plot representing the summation of projection signal over annotation structures.
3. A cross-plane, point-based **Synchronize** feature in the Zoom and Pan (Zap) Image Viewer allows multiple experiments to be synced to the same approximate position in the brain based on a linear alignment of the images to a 3D reference model.
4. Visualization of projection profiles in a 3D format as pre-rendered views and as interactive 3D computational paths in the **Brain Explorer® 2** desktop software application.
5. A **Data Search** feature allows users to find experiments by user-specified source and target structures or voxels.

6. A **Correlative Search** feature allows users to search for projection profiles highly correlated to a seed experiment.
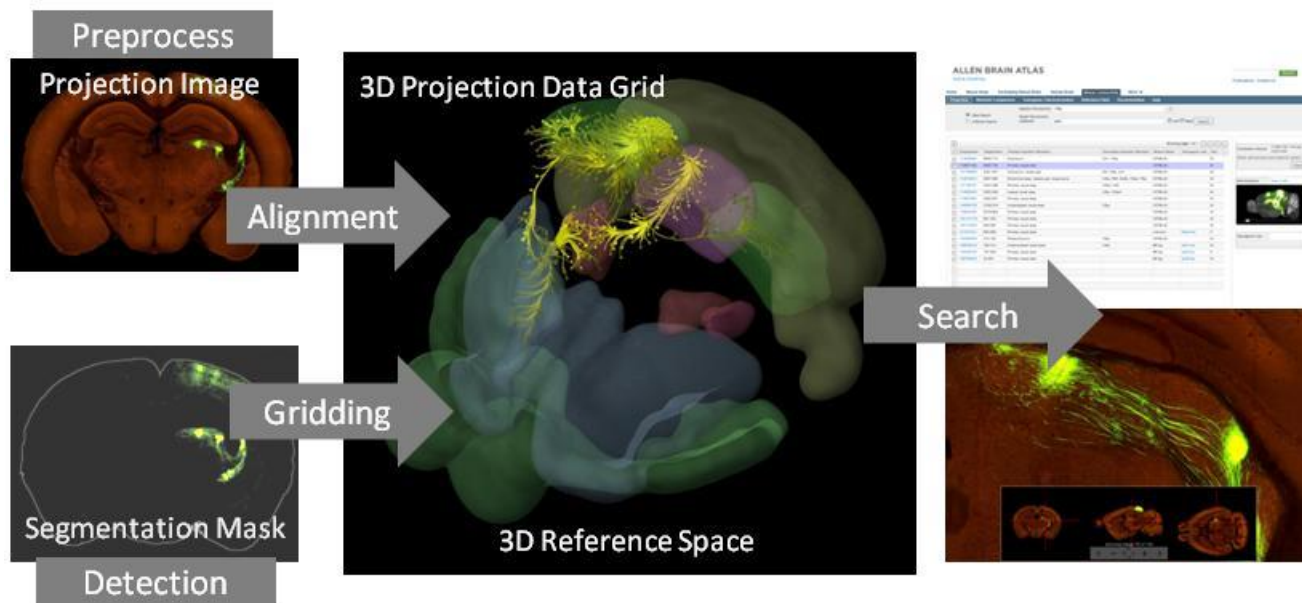
**Figure 1. The informatics data processing pipeline.**
The Alignment module registers each projection image to the common coordinates of a 3D reference model. The Detection module separates signal from background. The Gridding module produces a 3D summary of the detected signal for each data set. Further downstream, the gridded data is used for data and correlation searches. The projection image above shows terminal and passing fiber signal in the thalamus arising from viral infection in the primary visual area.

PREPROCESSING AND QUALITY CONTROL

Scanned image tiles are stitched to form a single high-resolution image in a three step process that consists of tile positioning using a calibration matrix, intensity correction, and tile transition blending.

Tile Positioning and Calibration
For each imaging unit, the spatially dependent stage positioning offsets were modeled as a 2x2 calibration matrix. Each unit was calibrated using a pattern of fluorescent microspheres (**Figure 2**). Each tile pair was registered by maximizing image intensity correlation in the tile overlap region to obtain a relative x-y position. A 2x2 calibration matrix for the system was obtained by solving for the global minimum error solution using over 200 tile pairs. Once calibration was complete, the calibration matrix was used to spatially position scanned projection image tiles without further registration.
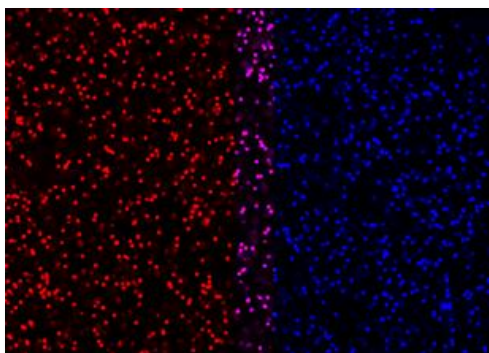
**Figure 2. A pair of registered calibration tiles.**
A microsphere pattern is used to compute a 2x2 calibration matrix. For visualization, the left tile is colorized in red, the right tile in blue, and the overlapping region in pink.

Intensity Correction

Systematic tile intensity distortion was corrected before stitching tiles into an image. It was observed that the intensity distortion varied between units and over time, and hence correction parameters had to be computed on a per-specimen basis. For each specimen, a simple model of the intensity distortion was obtained by averaging all tiles over a whole brain specimen for each channel (Ragan *et al.*, 2012). The intensity profile of each tile was then modified by a pixel by pixel division by the average tile for each channel (**Figure 3**).

Tile Transition Blending

The intensity-corrected tiles were then stitched together. In the overlapping region between adjacent tiles, linear blending was used to create a smooth transition. Note that intensity correction and blending were not applied to older experiments.
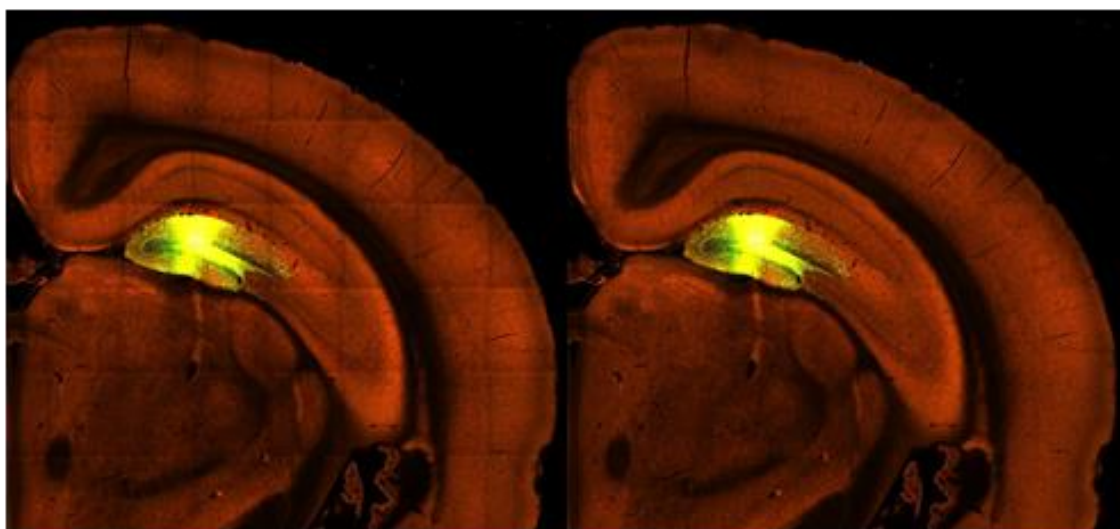
Figure 3. Effect of intensity correction.
Systematic tile intensity distortion is reduced by dividing by a brain-wide average tile. A partial image is shown with (right) and without (left) intensity correction.

Image Display

For multiresolution display on the Web, stitched images were downsampled to form a pyramid with each level of the pyramid divided into 256x256 pixel tiles. To handle 16-bit per channel image content, each image tile was encoded in JPEG2000 format. When a tile is requested, a user-specified window and level for each channel is applied to generate 8-bit JPEG images for display by standard Web browsers. To provide a uniform look over all experiments, default window and level values were computed using intensity histograms. For each experiment, the upper threshold defaults to (2.33 x 95th percentile value) for the red channel and (6.33 x 95th percentile value) for the green channel.

A rigorous manual Quality Control (QC) protocol was also established that included decisions on failing an experiment due to production issues such as specimen and imaging quality. In some cases, the QC process extended to identification of areas of high intensity/high frequency artifacts and areas of signal dropout. This information was used in downstream search and analysis to reduce false positive and false negative returns. Additionally, identification of the injection site location(s) was manually done by delineation with closed polygons on each image. This information was used to computationally association injection structures according to the Allen Reference Atlas ontology. For retinal projectome data, the optic chiasm (identified by registration to the reference space – see methods below) was used informatically as a surrogate injection site location.

3D REFERENCE MODELS

The cornerstone of the automated pipeline is an annotated 3D reference space. For this purpose, a next generation Allen Mouse Common Coordinate Framework (CCF) was created using an average population of 1675 specimens. See the Allen Mouse Common Coordinate Framework whitepaper in the Documentation tab for detailed information (**Figure 4**).
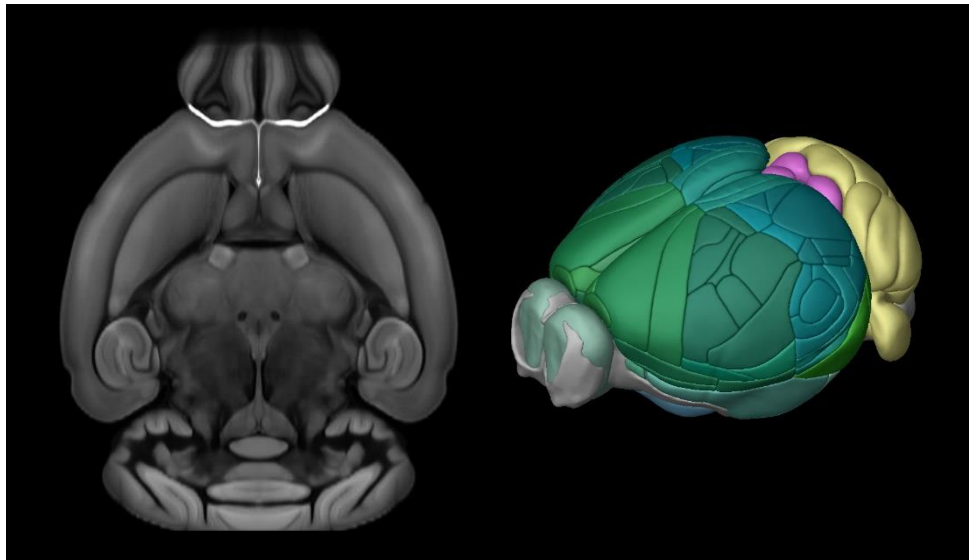
Figure 4. Version 3 of the Allen Mouse Common Coordinate Framework.
The Allen Mouse Common Coordinate Framework is based on the shape and intensity average of data from 1,675 specimens in the Allen Mouse Brain Connectivity Atlas. 662 structures have been delineated on the anatomical template.

ALIGNMENT MODULE

The serial imaging method ensured alignment of section images throughout a single specimen. Thus, section images can be simply stacked together to form a coherent 3D volume and subsequently generate virtual (low-resolution) sagittal and horizontal slices for Web presentation.

Next, each 3D image stack was globally aligned to the 3D reference model. The global alignment process consists of three steps. First, a coarse registration was initialized by matching the image moments of the image stack and the template. Second, a rigid (rotation plus translation) registration was performed, followed by a 12-parameter affine registration. Each step was based on maximizing normalized mutual information between the red channel of the image stack and template using a multi-resolution gradient-descent algorithm.

To further increase the alignment accuracy a secondary local/deformable registration was introduced to increase fidelity by modeling local deformations. Similar to the global alignment stage, the B-spline based deformable registration was also conducted sequentially from-coarse-to-fine in 4 resolution levels with decreasing smoothness constraints (see Kuan *et al.*, 2015 for algorithmic details).

Once registration was achieved, information from the 3D reference model could be transferred to the experimental specimen and vice versa. The resulting transform information (a 2D affine transform per image and 3D affine transform per specimen) was stored to support the image synchronization feature (**Figure 5**) in the Zap viewer and generation of grid-level projection signal summaries.
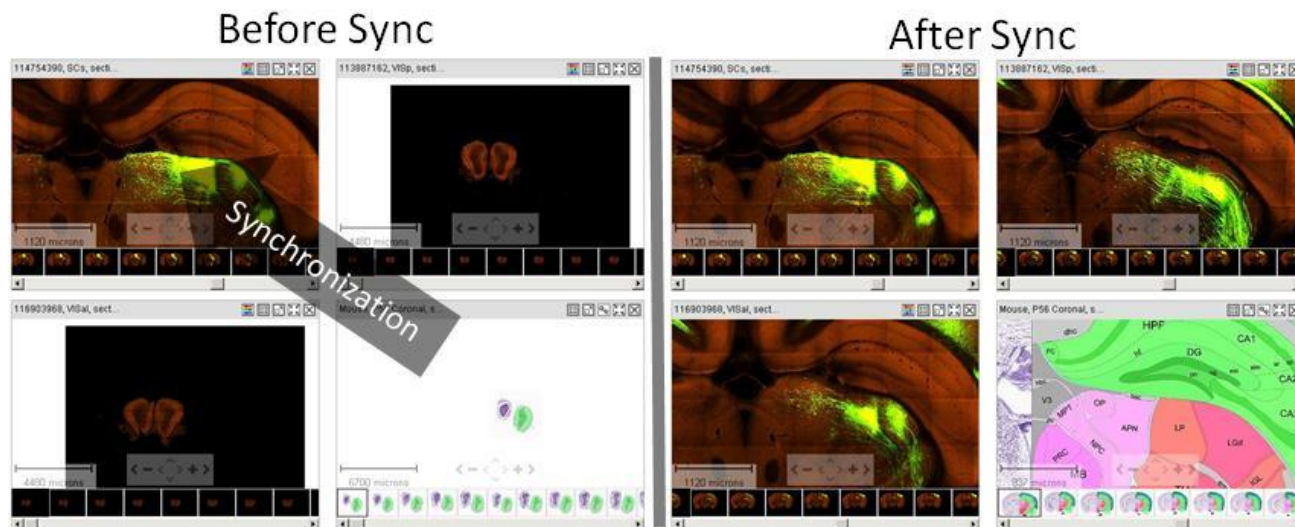
Figure 5. Point-based image synchronization.
Multiple image-series in the Zoom-and-Pan (Zap) viewer can be synchronized to the same approximate location. Before and after synchronization screenshots show projection data with injection in the superior colliculus (SCs), primary visual area (VISp) anteolateral visual area (VISal), and the relevant coronal plates of the Allen Reference Atlas. All experiments show strong signal in the thalamus.

PROJECTION DETECTION MODULE

A detection algorithm was applied to each projection image to separate signal from background (**Figure 6**). There are three stages in the algorithm: 1) intensity rescaling and noise reduction, 2) tissue region segmentation, 3) projection signal segmentation.

Intensity Rescaling and Noise Reduction
Due to scanner variability and biological specimen variations, image intensity can vary significantly across experiments. In the first step of the detection algorithm, images were rescaled in an attempt to fit all experiments to a similar intensity profile using the middle sections of each brain. The image was transformed using the square root of pixel intensity to remove second-order effects. The resulting histogram was matched at midpoint to a template profile. Median filtering and large kernel low pass filtering were applied to remove noise.

Tissue Region Segmentation
Based on the red channel, the tissue area was identified using adaptive thresholding and morphology operations in combination with connected component analysis. Small holes within the segmented area were filled before the exterior boundary was produced.

Projection Signal Segmentation
The signal detection process was based on a combination of adaptive edge/line detection and morphological processing (**Figure 6**). High-thresholded edge information was combined with spatial distance-conditioned low-thresholded edge results to form candidate signal object sets. The candidate objects were then filtered based on their morphological attributes such as length and area using connected component labeling. In addition, high intensity pixels near the detected objects were included into the signal pixel set. In a post-segmentation step, detected objects near hyper-intense artifacts occurring in multiple channels were removed (see Kuan *et al.*, 2015 for algorithmic details). Two variations of the algorithm were employed depending upon the type of virus used for the experiment; one algorithm was tuned for EGFP fiber detection, and one for SypEGFP terminal detection. Note the detection algorithm operates on a per-image basis and not in 3D. At this stage, passing fibers and terminals were not distinguished. When detecting EGFP-expressing fibers in the target-defined anterograde projection data, the autofluorescence background signal around the tissue and blood vessel edges can be enhanced, mimicking low-intensity projection signal, which can result in false positive signal detection

along tissue surfaces. In these circumstances, a supervised decision tree classifier was used to filter out these segmentation artifacts based on morphological measurements, location context and the normalized intensities of all three channels.
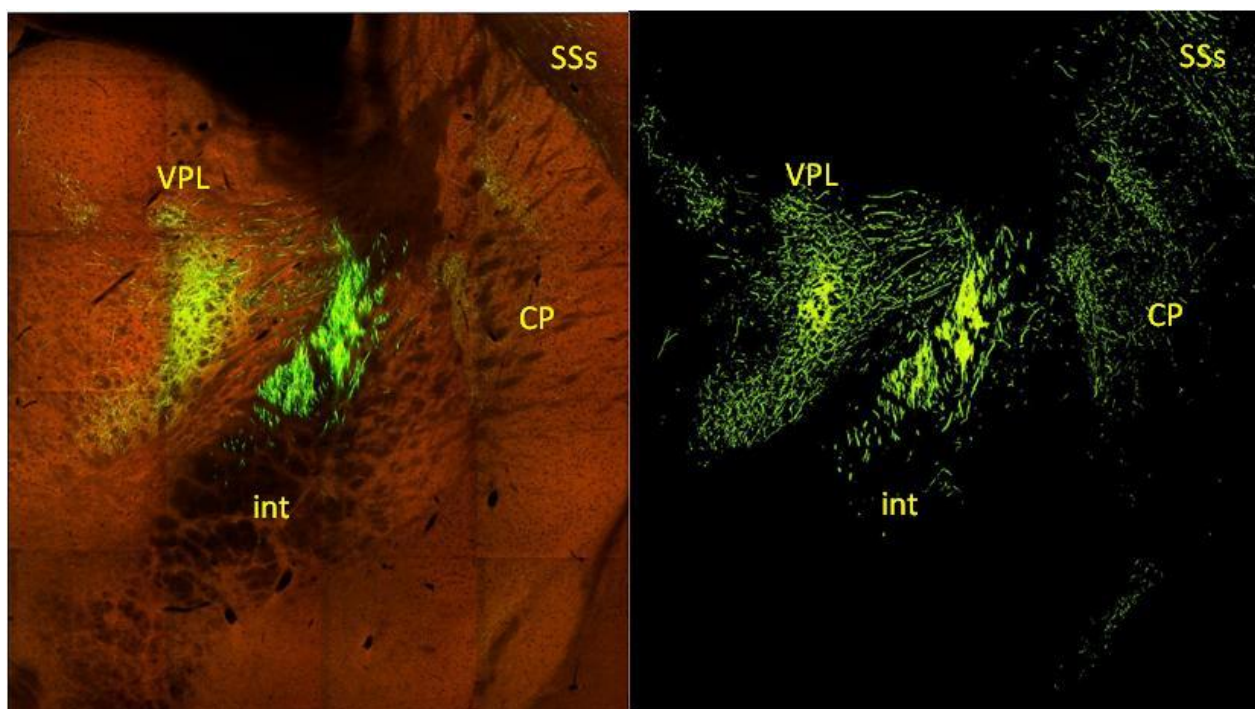
Figure 6. Signal detection for projection data with injection in the primary motor area.
Screenshot of a segmentation mask showing detected signal in the ventral posterolateral nucleus of the thalamus (VPL), internal capsule (int), caudoputamen (CP), and supplemental somatosensory area (SSs).

PROJECTION GRIDDING MODULE

The aim of the Gridding module was to create an isotropic 3D summary of projection data and resample the data to a common coordinate space to enable spatial comparison between data from different specimens. The data grids were used for downstream search and analysis, and were further processed as computational paths for viewing in the Brain Explorer 2 3D viewer (similar to the Brain Explorer viewer; Lau *et al.*, 2008), alongside the Allen Mouse Common Coordinate Framework.

The Gridding module operates on a per specimen basis. Images were divided into 10µm x 10µm grids. In each division, the sum of the number of detected pixels and the sum of detected pixel intensity were collected. A second set of summations was computed for regions identified as belonging to the injection site for injection site quantification. The 2D per-image grids were combined to form 3D grids of resolution 10µm x 10µm x 100µm. Using the transform parameters computed by the Alignment module, each 3D grid was transformed to the standard 10µm isotropic reference space using linear interpolation to generate sub-grid values.

From these summations measurements were obtained for:

- *Signal density* = sum of detected pixels / sum of all pixels in a grid
- *Signal intensity* = sum of detected pixel intensity / sum of detected pixels
- *Signal energy* = sum of detected pixel intensity / sum of all pixels in a grid

A preview of the projection data grid is shown on the Web application as a maximum density projection thumbnail that can be interactively rotated about the dorsal-ventral (y) axis and anterior-posterior (x) axis. Grids

at 10, 25, 50 and 100 μm isotropic resolution are also generated and are available for download through the Allen Brain Atlas Application Programming Interface (API).

INTRINSIC SIGNAL IMAGING DATA PROCESSING

Intrinsic signal imaging (ISI) data was acquired and processed as described in the Overview whitepaper located in the Documentation tab to create auto-segmented and annotated sign maps for experiments targeting visual areas. Along with each ISI experiment, a brain surface vasculature image was collected for each experiment.

Mapping ISI to Serial Two Photon Imaging Data

Major vessels from the brain surface vasculature were used as fiducials to align the ISI data to the two photon imaging data. After mapping the two photon image data to the CCF, a virtual surface vasculature image was generated by obtaining the minimum intensity projection in a narrow band along the pial surface of the brain. The resulting image shows the major blood vessels as dark pixels along the surface. The reconstructed brain was also rotated to match the imaging angle of the ISI experiment. Corresponding vasculature segments were then manually annotated on the vasculature image obtained with the ISI experiment and the virtual vasculature image (**Figure 7**). The annotations were used to compute an affine transform to bring the two images into registration. The transform was then used to project the ISI sign maps into the CCF.
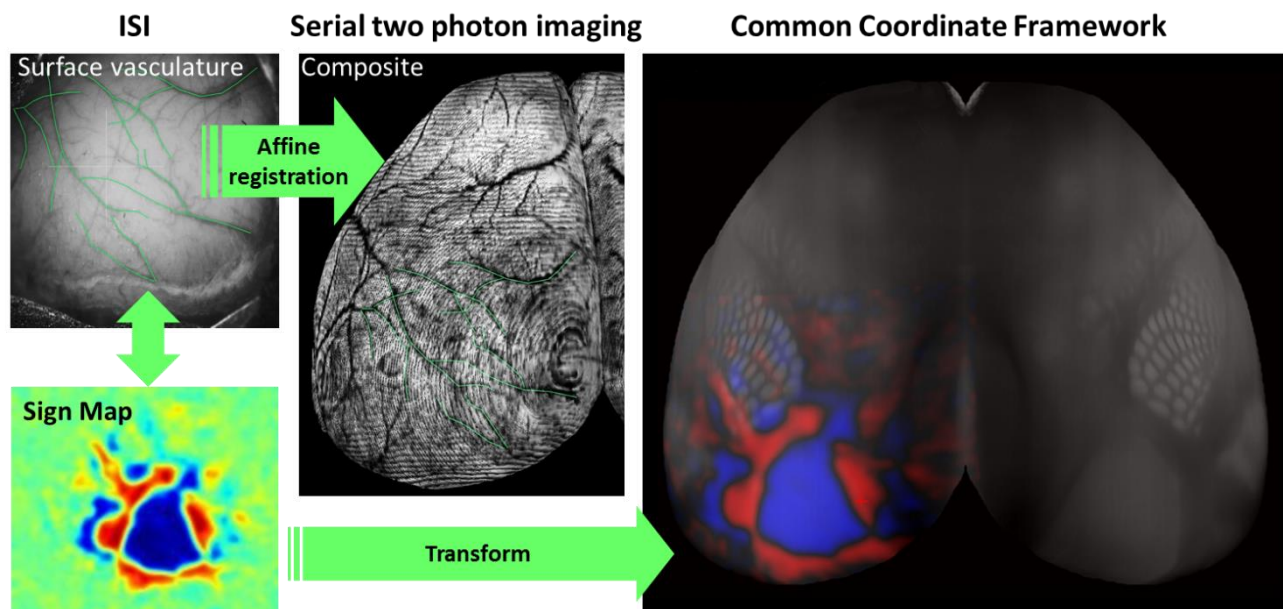

Figure 7. Schematic showing the process of mapping an ISI sign map to the CCF.

GRID VISUALIZATION

The aim of interactive grid visualization in the Brain Explorer 2 3D viewer is to provide a spatial relationship between projection fields and pathways and allow users to view the original 2-photon images at points of interest (**Figure 8**). First, the fast marching algorithm (Sethian, 1999) was run on the signal density grid with seed points at the injection site. This step calculates a time of arrival, which can be thought of as a weighted distance, at each point in the 3D segmentation mask from the injection site. The fast marching computation also yields vectors at each voxel pointing back towards the injection site. Computational paths are fit through this vector field starting from a sampling of points of low density in the grid image. Cubes are rendered at the origin points. These cubes can be clicked to view their location (coordinates and structure), density, and image intensity. The injection site grids are shown as spheres where the injection site has been manually delineated.
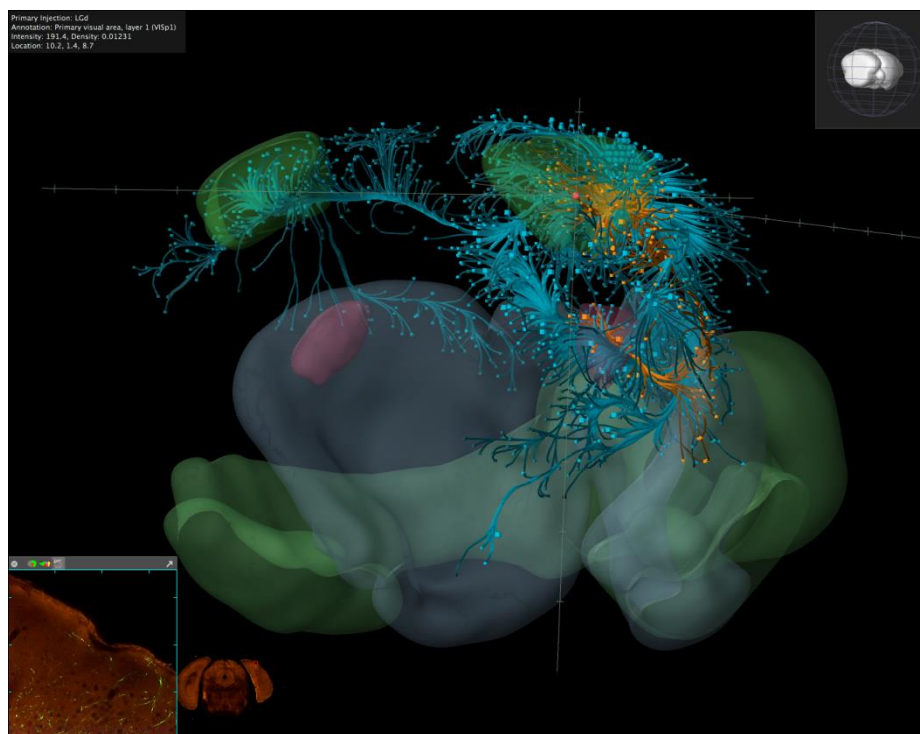
Figure 8. Projection signal visualization in the Brain Explorer 2 3D viewer.
Screenshot showing an injection from the primary visual area (VISp) in blue and the dorsal part of the lateral geniculate complex (LGd) in orange. The outlines of LGd, VISp, striatum, and olfactory areas are also shown for context. A grid from the LGd injection in VISp, layer 1 is highlighted as a red sphere and its corresponding 2D image is shown in the bottom left hand corner.

GRID SEARCH SERVICE

An on-the-fly projection grid search service has been implemented to allow users to instantly search over all experiments to find specific projection profiles:

- The **Source Search** function retrieves experiments by the anatomical location of the injection site.
- The **Target Search** function returns a list of experiments ranked by signal volume in the user-specified target structure(s).
- The **Spatial Search** function in target mode returns a list of experiments ranked by density of signal in the user-specified target voxel location. In injection mode this function returns a list of experiments ordered by user-specified distance from their injection site to a seed location.
- The **Correlative Search** feature allows the user to search for projection profiles highly correlated to a seed experiment, with an option to search over a specified domain.

To perform these computations quickly over the entire data set, all 489,393 non-background voxels from every experiment are loaded into the server memory as a matrix. Each user search request is computed on demand. To activate a **Source Search**, a user can specify one or more *source* brain regions. The search returns all experiments where the injection site location is within the specified region. This search takes into account all primary and secondary injection site structures with an option to search just by the primary injection site.
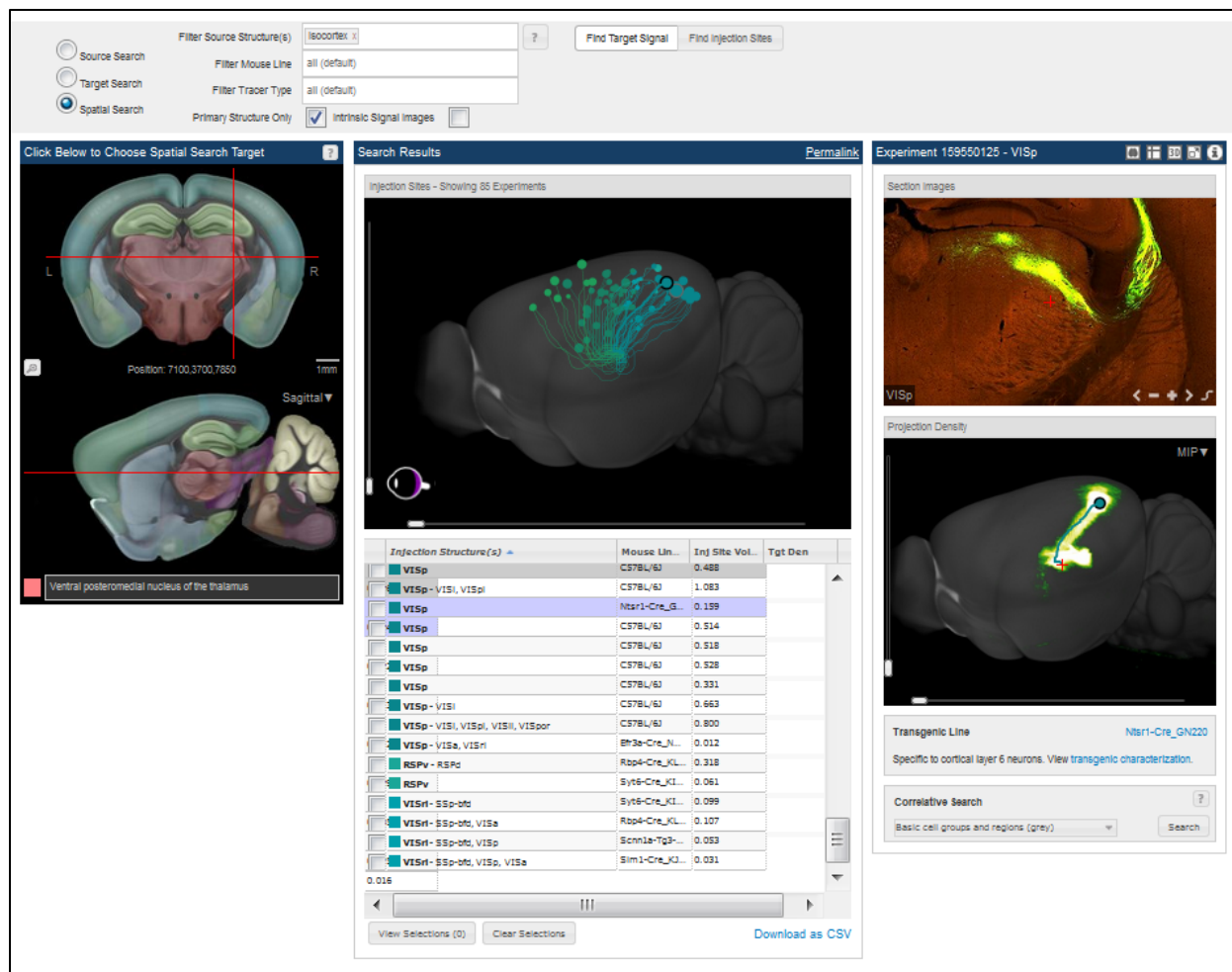
Figure 9. Search results for experiments from the isocortex (source region) to the target location in the dorsal lateral geniculate complex.
Each of the 85 returned experiments are depicted as a circle at the injection location and a computational path from the injection site to the target location. When a user selects an experiment (by clicking on a row), an interactive 3D thumbnail is displayed on the panel on the right. The thumbnail is generated by maximum density projection of the gridded density data. From this panel a user can launch the high-resolution image viewer, view the experiment details page or view the data in full 3D in the Brain Explorer application.

In **Target Search**, a target brain region is also specified along-side the source region. A target brain region is defined by selecting one or more structures from the ontology. The service first computes the signal density in the target brain region by summing all voxels belonging to that area. The experiments are then ranked in decreasing order of signal density. A user can optionally restrict the target density computation to one hemisphere. All injection-hemisphere voxels within structures identified as primary or secondary injection structures are excluded in the computation.

In **Spatial Search** (target mode) a user interactively selects a seed location by manipulating a set of crosshairs on cross-sections of the reference atlas (**Figure 9**). The search returns experiments ranked by density of signal in the selected location. This feature allows the user to interactively explore connectional topography between two regions of the brain.

In **Spatial Search** (injection mode) a user interactively selects a seed location by manipulating a set of crosshairs on cross-sections of the reference atlas. The search returns experiments by injection site within a user defined distance ranked by increasing distance (in μm) from the selected voxel location.

To start a **Correlative Search**, a user selects a seed experiment by selecting a row in any search result and a domain over which the similarity comparison is to be made. All voxels belonging to any of the domain structures will form the domain voxel set. Pearson's correlation coefficient is computed between the domain voxel set from the seed experiment and every other experiment in the dataset. The results are then sorted by descending correlation coefficient and displayed (**Figure 10**).
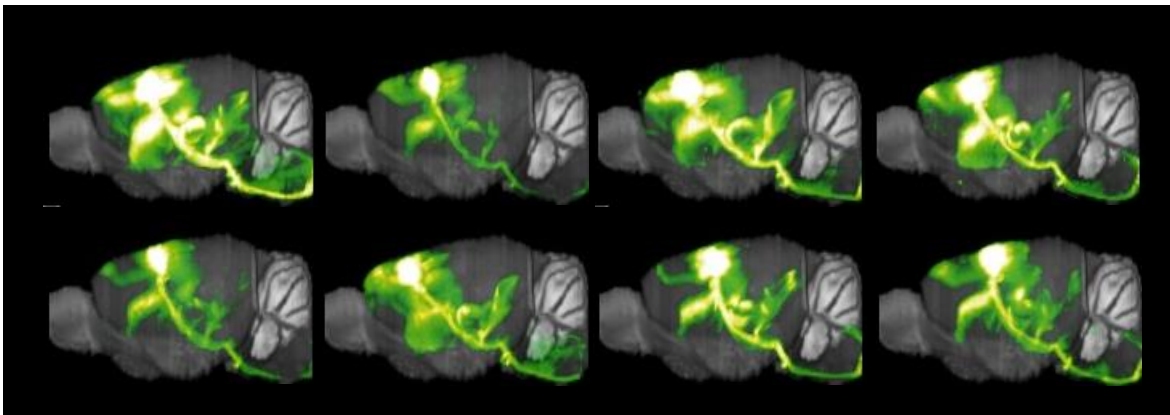
Figure 10. Correlative search for other experiments similar to an injection in the primary motor area (MOp).
Screenshot of the top seven results with the seed experiment in the top-left corner. Correlative search yields other experiments also with injection in the primary motor area.

It should be noted that these on-the-fly search services are derived from a fully automated processing pipeline. False positive and false negative results can occur due to artifacts on the tissue section or slide and/or algorithmic inaccuracies. Users should confirm results with visual inspection of the projection images.

STRUCTURE UNIONIZER MODULE

Projection signal statistics can be computed for each structure delineated in the reference atlas by combining or unionizing grid voxels with the same 3D structural label. While the CCF is typically annotated at the lowest level of the ontology tree, statistics at upper level structures can be obtained by combining measurements of the hierarchical children to obtain statistics for the parent structure. Further, the unionization process separates out left versus right hemisphere contributions as well as injection versus non-injection components. The end result is per structure *projection volume* (sum of detected signal in mm^3) and *projection density* (detected signal volume / structure volume) values for each experiment which are stored in the database and used in the web application to display projection summary bar graphs.

Selecting the experiment opens up a set of preview section images and projection density; at the upper right corner there is an "i" icon that takes you to an experiment detail page that quantifies the projection data. The Injection Summary (**Figure 11A**) gives you the calculated primary and secondary injection sources, the stereotaxic coordinates (or the ISI coordinates for injections that utilized ISI) and the calculated injection site (by percentage). The histogram (**Figure 11B**) quantifies the amount of signal in each structure either by projection volume (mm^3) or by projection density (fraction of area occupied by signal compared to the whole structure). The two image viewers show the 3D rotating thumbnail (**Figure 11C**) or the 2D Zap viewer (**Figure 11D**). The image viewers and the histogram are synchronized such that clicking on a point in one will highlight the same location in the other two.
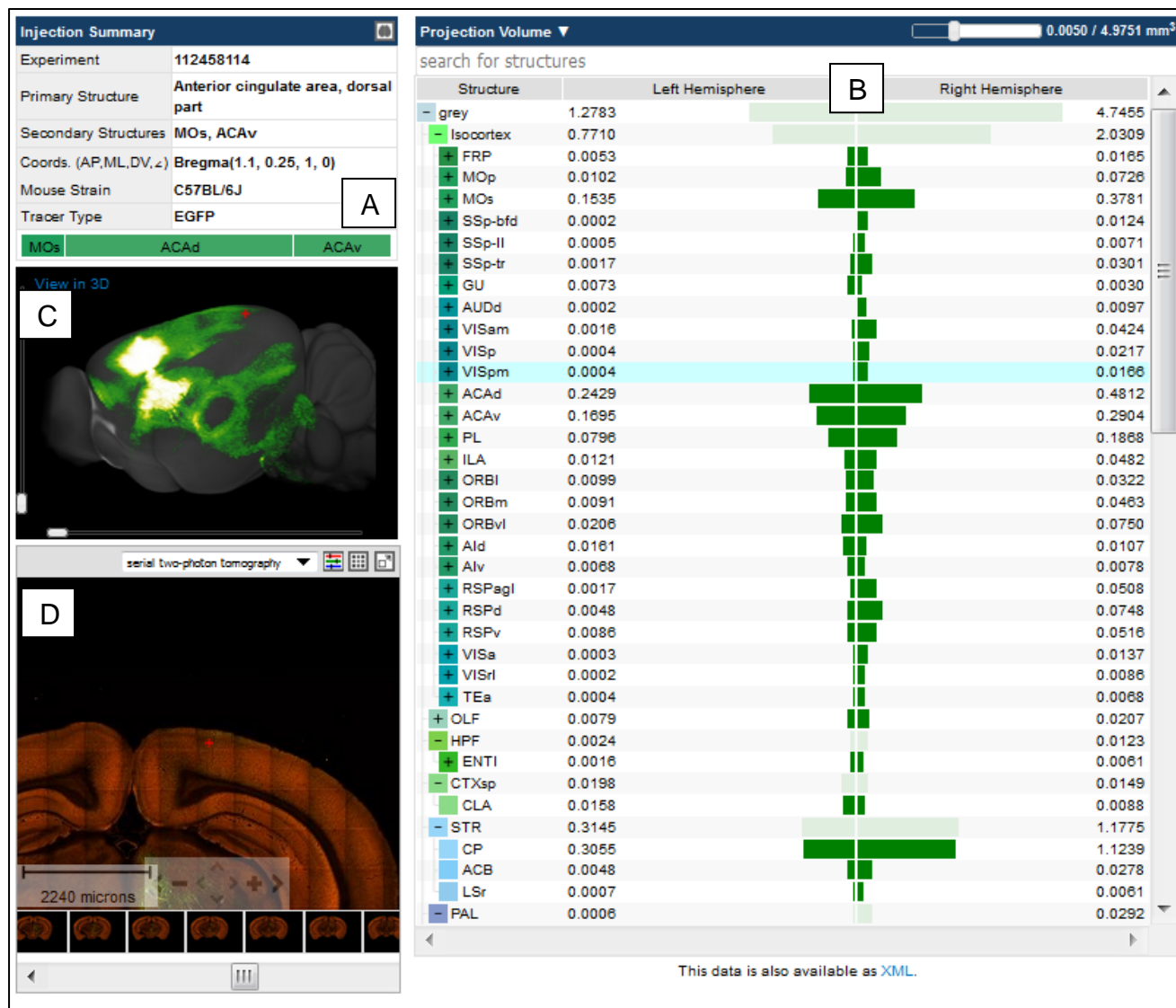
Figure 11. Screenshot of experiment details page for an anterior cingulate area (ACAd) injection experiment.
Quantification of the projection data including **(A)** Manual Injection Summary and a stacked bar graph showing relative structural composition of the injection site as quantified by the informatics processing pipeline. **(B)** Interactive projection summary bar graph display in the context of a simplified Allen Reference Atlas ontology. Length of bar represents Projection Volume. A user can manually open/close each ontology node or alternatively use the threshold slider to control the view. Only structure nodes with values greater than the selected threshold are visible. A dark green bar represent a leaf node in the current view and a light green bar represents a parent node. **(C)** Interactive **3D thumbnail** of projection signal density rendered by maximum density projection. **(D)** 2D Zap Viewer. **Image preview** syncs with both the bar graph and the 3D thumbnail.

INJECTION SITE QUANTIFICATION

Injection site location(s) were manually defined on the high resolution images by delineating with enclosed polygons to create an injection mask. Similar to expression gridding, the injection masks were quantized into 10µm x 10µm grids. In each division, the sum and fraction of pixels and (if applicable) the sum of detected pixels within the injection mask were collected. The 2D per-image grids were combined to form 3D grids of resolution 10µm x 10µm x 100µm. Using the transform parameters computed by the Alignment module, each 3D grid was transformed to the standard 10µm isotropic reference space. Injection statistics can be computed for each

structure delineated in the reference atlas by combining or unionizing grid voxels with the same 3D structural label.

REFERENCES

Dang C, Sodt A, Lau C, Youngstrom B, Ng L, Kuan L, Pathak S, Jones A, Hawrylycz M (2007) The Allen Brain Atlas: Delivering neuroscience to the Web on a genome-wide scale. *Data Integration in the Life Sciences* in *Lecture Notes in Computer Science* 4544:17-26.

Kuan L, Li Y, Lau C, Feng D, Bernard A, Sunkin SM, Zeng H, Dang C, Hawrylycz M, Ng L (2015) Neuroinformatics of the Allen Mouse Brain Connectivity Atlas. *Methods* 73:4-17.

Lau C, Ng L, Thompson C, Pathak S, Kuan L, Jones A, Hawrylycz M (2008) Exploration and visualization of gene expression with neuroanatomy in the adult mouse brain. *BMC Bioinformatics* 9:153.

Ng L, Pathak SD, Kuan C, Lau C, Dong H, Sodt A, Dang C, Avants B, Yushkevich P, Gee JC, Haynor D, Lein E, Jones A, Hawrylycz M (2007) Neuroinformatics for genome-wide 3D gene expression mapping in the mouse brain. *IEEE/ACM Trans Comput Biol Bioinform* 4:382-393.

Ng L, Bernard A, Lau C, Overly CC, Dong HW, Kuan C, Pathak S, Sunkin SM, Dang C, Bohland JW, Bokil H, Mitra PP, Puelles L, Hohmann J, Anderson DJ, Lein ES, Jones AR, Hawrylycz M (2010) An anatomic gene expression atlas of the adult mouse brain. *Nature Neuroscience* 12:356-362.

Ragan T, Kadiri LR, Venkataraju KU, Bahlmann K, Sutin J, Taranda J, Arganda-Carreras I, Kim Y, Seung HS, Osten P (2012) Serial two-photon tomography for automated ex vivo mouse brain imaging. *Nature Methods* 9:255-258.

Sethian JA (1999) *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics*. Cambridge University Press.

Yushkevich P, Avants B, Ng L, Hawrylycz M, Burstein P, Zhang H, Gee J (2006) 3D mouse brain reconstruction from histology using a coarse-to-fine approach. *Third International Workshop on Biomedical Image Registration (WBIR)* 230-237.