PubAnatomy: Cross Domain Neurobiology Literature and Data Exploration

Weijian Xuan[1,4], Manhong Dai[1], Buckner Josh[1], Barbara Mirel[2,4], Jean Song[3,4], Brian Athey[1,4], Huda Akil[1], Stanley J. Watson[1], Fan Meng[1,4]

[1]Psychiatry Department and Molecular and Behavioral Neuroscience Institute, [2]School of Education, [3] Health Science Libraries,

[4]National Center for Integrative Biomedical Informatics, University of Michigan

Abstract

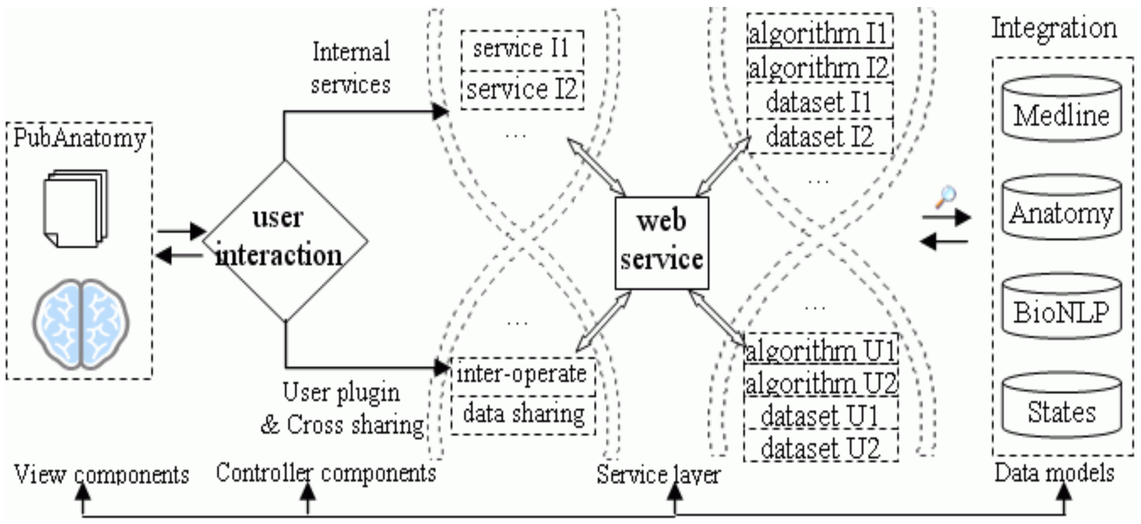
Understanding the biomedical implications of data from high throughput experiments requires solutions for effective cross-scale and cross-domain data exploration. Our work integrates molecular level data with high level biological functions using anatomical structure as a scaffold. The Flex-based PubAnatomy web application we developed enables highly interactive visual exploration of literature and experimental data for understanding the relationships between molecular level changes, pathways, brain circuits and pathophysiological processes. PubAnatomy also allows the sharing of intermediate data exploration results with other web applications, greatly increasing the power of cross-domain data exploration and mining. The prototype of PubAnatomy is freely accessible at:

http://brainarray.mbni.med.umich.edu/Brainarray/prototype/PubAnatomy.

Overall Architecture

The high level architecture of PubAnatomy follows the Model-View-Controller-Service design pattern, as shown in Fig. 1. The view component (UI) is a Flex-based Rich Internet application compatible with virtually all major browsers. The UI is further divided into a series of views that each is a reusable component. The controller component handles user interactions (e.g. keyword search, gene expression heatmap drawing) and associated events, and call services to analyze and exchange data. The service layer provides the bridge between UI and the PubAnatomy backend or external databases. There are already over 30 web services developed for PubAnatomy. Except services dedicated to providing UI information, the web services can also serve other programs. The data models maintain datasets and stores the state of the user exploration in PubAnatomy. Overall, the isolation of interaction, UI design, data models and analysis enables us to handle nonlinear and complex exploration process.

Fig. 1. PubAnatomy architecture



Data Exploration in PubAnatomy

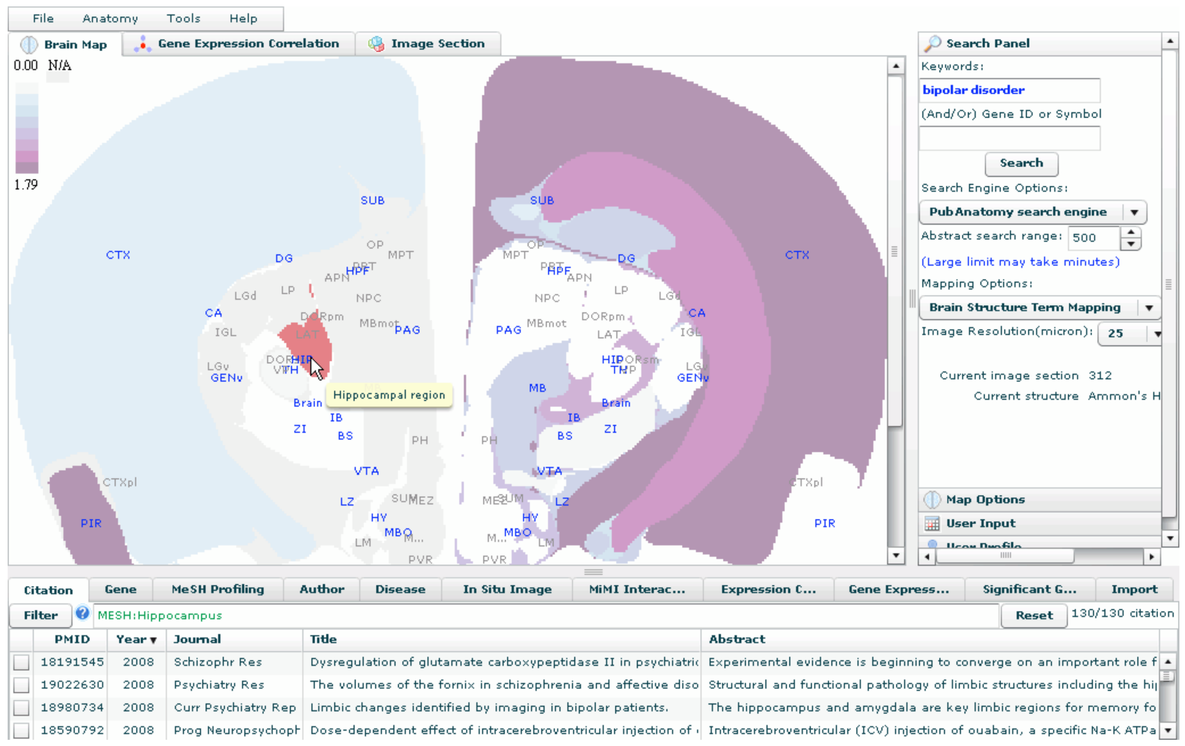
As a literature and data exploration tool, PubAnatomy allows users to search literature and provides different view and data tables for filtering and sense making. The UI has three major components: 1) graphic views on the main window provide data overview and starting points for data exploration 2) tabulated data tabs in the bottom contain information relevant to the current view and selection, such as current brain structure, citation set, selected genes, etc.; 3) Tabs and menus on top of the main window are for selecting parameters and initiating analysis. The right panel contains search functions, user input and user history management (Fig.2). Here we highlight some unique functions in PubAnatomy.

1. BrainMap view for presenting related structural and functional information.

When a user issue a keyword search, PubAnatomy will search literatures by using either a local search engines that we implemented for brain-related literature, or by calling the NCBI E-utilities web service on PubMed. The related citations will be mapped to brain structures using our pre-indexed tables. A coronal brain section with the largest number of PMID hit is than selected for the main window. Each colored region is a brain structure. Users can click on a structure to retrieve related Medline records. The color of labels of each structure indicates if there are citations mapped to the structure (Fig.2).

When a different structure is selected, contents in the data tabs under the main window, such as Mesh profiling, related disease, In situ image mappings, and protein-protein interactions associated with genes in the related Medline records will be automatically updated.

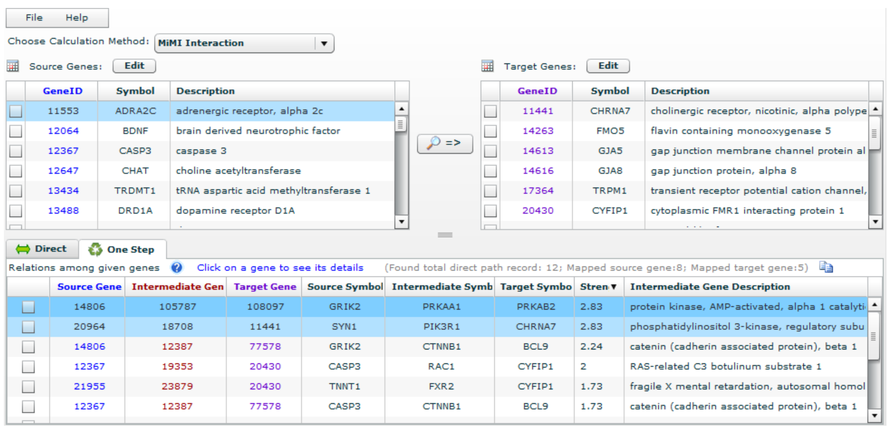
Fig. 2. PubAnatomy Interface Overview



2. Find potential gene-gene relationships through ArrowSmith like path finding

When researchers obtained a gene list from high throughput experiment, a common question is how the list of gene can potentially be related to the targeted biological process. In PubAnatomy, this question can be answered in a straightforward manner by first performing a keyword search for the biological process and then choosing "Find path for all genes" in the right click menu of the Gene tab, which lists all genes associated with related Medline abstract as annotated by NCBI. In the resulting PubPath window, the user can paste in their gene list and click the path search button to identify all direct gene-gene relationships and indirect relationships with one intermediate gene based on either protein-protein interaction relationship in the Michigan Molecular Interaction Database (MiMI) or the gene-gene co-expression relationship derived from the Allen Brain Atlas mouse brain data. Fig. 3 Illustrates researchers can quickly identify how genes in three deleted regions associated with schizophrenia patients may interact with genes related to schizophrenia based on Medline annotation.

Fig. 3. Link query gene list to schizophrenia related Medline genes through MiMI



3. Interoperability with other applications

PubAnatomy emphasize the interoperability by adopting the aforementioned data sharing and user management schema. When a user login from PubAnatomy, he/she can export datasets, e.g. citation PMIDs, to a central database. He can name the dataset, record its parameters, write description and choose whether to share the dataset. Other applications or PubAnatomy itself can retrieve the dataset from the central database for additional analysis.

Summary

We developed PubAnatomy to facilitate the association of molecular level data with higher order pathophysiological processes in the context of anatomical structures and their functional annotations. Our implementation can be easily extended to any macro- and micro-anatomical structure representations for more effective cross-domain data exploration and analysis in the context of relevant biology.

Acknowledgement: We want to thank Glenn Tarcea's help in mirroring the MiMI database. W. Xuan, M. Dai, J. Buckner, F. Meng, H. Akil, and S. Watson are members of the Pritzker Neuropsychiatric Disorders Research Consortium, which is supported by the Pritzker Neuropsychiatric Disorders Research Fund L.L.C. This work is also partly supported by the National Center for Integrated Biomedical Informatics through NIH grant 1U54DA021519-01A1.