



Cerebellar Spatiotemporal Gene Expression Mapper Technical Documentation

Shesan Govindasamy, Queenie Tsang, Gayathri Ravindra
April, 11th, 2019

Cerebellar Spatiotemporal Gene Expression Mapper Technical Documentation

Summary

The cerebellum is a region of the brain which is important for controlling voluntary motor movement, motor learning, processing sensory information and relaying signals to motor neurons to regulate motor movements. Cerebellar Spatiotemporal Gene Expression Mapper is a R Shiny web application which enables users to understand and visualize the gene expression in the cerebellum with quickly generated figures. The product uses datasets on the cerebellum from Allen Human Brain Atlas (<http://portal.brain-map.org/>), and BrainSpan (<http://www.brainspan.org/>). The Allen Human Brain dataset is comprised of normalized microarray data from 6 adult human brains, while the BrainSpan dataset uses RNA-Seq Gencode v10 summarized to genes and Exon microarray summarized to genes datasets from BrainSpan. The R Shiny webapp uses these datasets as input to produce images of gene expression within the cerebellum in Scalable Vector Graphics (SVG) format. Additional functions of the webapp enables users to input a gene of interest to display a line graph of the log of gene expression as a function of age.

The R Shiny app uses the Shiny package in an R environment and calls upon R scripts which first processes the Human Brain Atlas and BrainSpan datasets to produce long form .csv files. This process utilizes various R libraries including magrittr, here, dplyr, ggplot2, reshape2, and readr to filter the datasets for data related to the cerebellum and save information such as gene symbol, expression level and age into a .csv file.

Product Description

Our R Shiny app produces a total of two products that we can discuss in further detail. These products are indicated by the tabs of our app with each in reference to a dataset we have used to plot gene expression. The first image generated using our program is a SVG sagittal view image using Allen Human Brain Atlas. The Allen HBA dataset consists of temporal gene expression data from six adult brain samples.

This image highlights the cerebrum and cerebellum regions of the brain with only the cerebellum regions highlighted after running the app. The program takes in input of genes present in the dataset which include TTR, BRCA1, or TP53 for example. The image generated includes a legend which corresponds to levels of expression for the gene of interest. Regions highlighted in yellow would have higher expression for the gene while regions highlighted in purple would have lower expression. The legend also indicates expression values in terms of a numeric scale.

In terms of interpreting this data, there are a few methods that this SVG image could be further used for. Our initial design kept neuroscientists and researchers in mind because they were the target audience for our web application. Thus, we predict one will be able to quickly characterize expression patterns using the figure to identify genes with coordinated behavior or it can serve as a reference to cell activity of genes. For instance, generating the SVG for multiple genes, a similar pattern of expression may be observed which could suggest a link between genes. With regard to cell activity, we know that cells are the direct result of gene

expression. Therefore, we can compare expression levels of genes in the cerebellum to deviant cell activity to check for possible correlation in the values.

The second product developed using R Shiny was the line plot generated through the use of the BrainSpan dataset. BrainSpan consists of spatial gene expression data for developing human brains (fetal stage to early adulthood). This graph maps the log of the expression level for a particular gene (y-axis) as a function of time (x-axis). Looking at the overall pattern of the line plotted, one can determine points in development with high expression of the gene (regions with peaks) and areas with low expression (regions with troughs). This plot essentially provides a method for identifying expression patterns for genes on a time basis which can provide insight to the relationship that exists between these genes and age-related diseases.

Operating Instructions

Cerebellar Spatiotemporal Gene Expression Mapper requires R, a programming language for statistical computing and graphics which can be downloaded and installed from <https://www.r-project.org/>. Additionally, RStudio, a free and open-source integrated development environment for R, should also be downloaded and installed for editing and running R scripts. RStudio can be downloaded from <https://www.rstudio.com>.

Initial Setups/Downloading Files

- 1) Clone/download the project file from <https://github.com/leonfrench/CerebellumSeneca>. It is currently a private repository, but will be public once the full application is ready to be deployed.
- 2) Set up Allen Human Brain Atlas.
 - Download files containing all normalized microarray expression values as well as probe and sample metadata necessary for analysis: “H0351.2001”, “H0351.2002”, “H0351.1009”, “H0351.1012”, “H0351.1015”, “H0351.1016”. From <http://human.brain-map.org/static/download>.
 - Uncompress the files.
 - Move the files into a new directory inside “CerebellumSeneca/R/” subfolder.
 - Set up file path location in Config.R script, “ALLEN_HBA_LOCATION”, to the directory created to house the Human Brain Atlas dataset.
- 3) Set up BrainSpan.
 - Download archive files containing normalized expression values and meta-data (as displayed in heatmap) for analysis: “Exon microarray summarized to probe sets”, and “Exon microarray summarized to genes”. From <http://www.brainspan.org/static/download.html>.
 - Uncompress the files.
 - Move the files into a new directory inside “CerebellumSeneca/R/” subfolder.
 - Set up file path location in Config.R script to the directory created to house the BrainSpan dataset. “BRAINSPAN_LOCATION_RNASEQ” shows the file path to

the gene_matrix_csv. "BRAINSPAN_LOCATION_EXONARRAY" shows the file path to gene_array_matrix_csv.

- Run config.R in Rstudio IDE or R application.

4) Load PackageList.R script to install package library.

*(Other datasets setup's will be updated as project progresses in the future).

Pre-processing Allen Human Brain Atlas data

- 1) Run AllenHBA_data_processing.R script. This is used to extract information from all the normalized microarray expression values and metadata to create a *LONG_FORM.csv file.

Pre-processing BrainSpan data

- 1) Run BrainSpan_data_processing.R script. This is used to extract information from *RNASEQ data or *EXONARRAY data to create a *LONG_FORM.csv file.

App.R setup (Shiny R)

- 1) Formatting location of the source input files in the app.R may be required to identify location of testAllenHBA_SVG.R, testLegend.r, testBrainSpan.R. Use "here()" to find current directory and work from there.
- 2) The app must be run on an external viewer, such as Google Chrome to display the SVG files properly. This can be changed by altering the Run App value, found at the top of the R input region. Example below.



App.R Usage (Shiny R)

The initial loading of the shiny application is presented below.

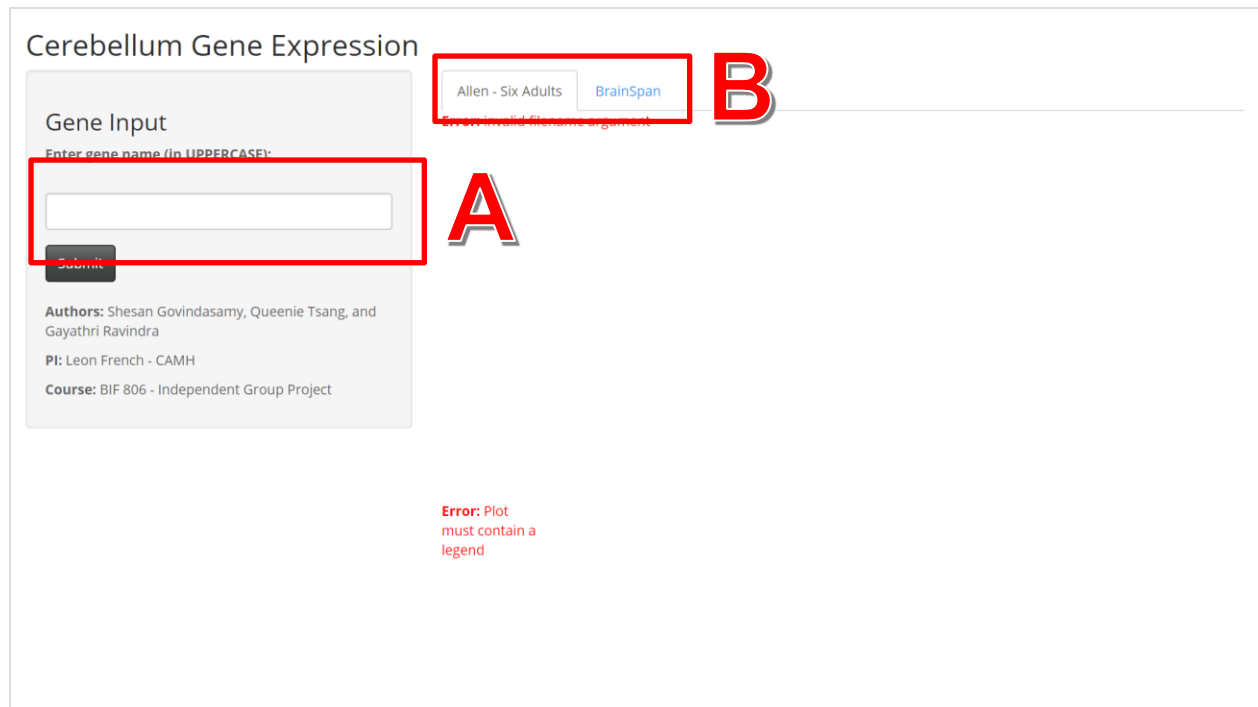


Figure 1: Startup of the app view. The A-value would accept the Gene Input (in uppercase) to be used for the expression mapping. The submit button must be pressed for plots and images to compile correctly. The B-value would allow the user to switch tabs between Allen – Six Adults (Allen Human Brain Atlas dataset) and BrainSpan (BrainSpan dataset).

*(2 error results occur upon initial loading of shiny app, however, this does not affect the program once a gene is input.)

1) Allen – Six Adults (with TTR gene input)

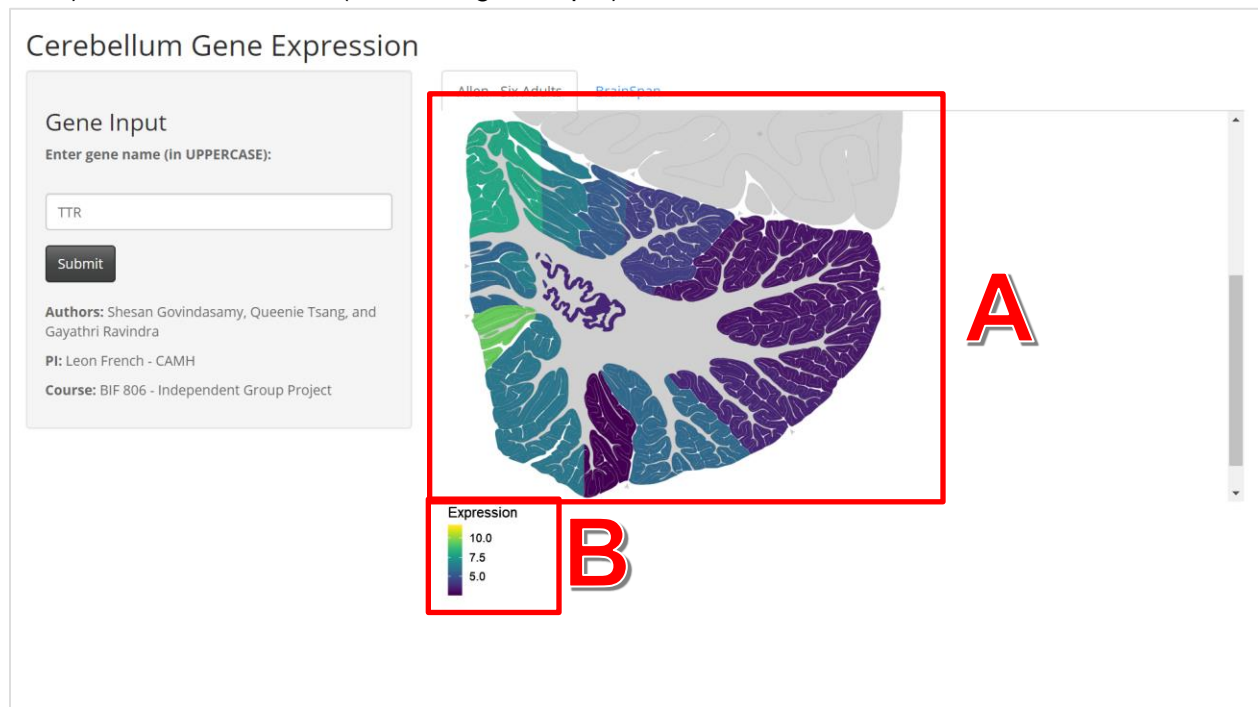


Figure 2: Allen Human Brain Atlas tab results. The A-value displays the SVG file of the mapped expression values from the pre-processed Allen Human Brain Atlas data results. The B-value displays the expression level gradient as a legend, the colors refers to the SVG map to show the levels of expressions.

2) BrainSpan



Figure 3: BrainSpan tab results. The A-value shows the generated ggplot of the age correlation to log₂Expression levels from the pre-processed BrainSpan results. Generation of the plot takes time and is dependent on your computational power, allow some time to let it plot.

Troubleshooting Techniques

Program file pre-processing troubleshooting

ERROR (dataset)	CAUSE	FIX
Files not found (AllenHBA_data_processing.R and BrainSpan_data_processing.R)	File not found due to multiple reasons	Ensure your files are found at the specified folders. Update the config.R script. Use here() to find the base location from where all scripts are run.

App troubleshooting

ERROR (tab)	CAUSE	FIX
"Error: invalid filename argument" (Allen – Six Adults)	SVG plot could not be generated due to no gene input	Input gene ID and hit submit
"Error: Plot must contain a legend" (Allen – Six Adults)	Legend image could not be generated due to no gene input	Input gene ID and hit submit
"Error: x must be length 1 or 2" (BrainSpan)	Plot could not be generated due to no gene input	Input gene ID and hit submit
Legend values do not load properly (Allen – Six Adults)	Rendering issue	Hit submit
SVG image does not display any color (Allen – Six Adults)	Invalid gene input	Select "Run External" from the "Run App" drop down menu in Rstudio. Then click "Run App". Input valid gene ID and hit submit.
Input gene entered correctly, but no display is shown (Allen – Six Adults and BrainSpan)	Gene input not in uppercase lettering or gene id not found in database.	Input uppercase gene ID and/or valid gene ID and hit submit

References

- Allen Institute for Brain Science. (2019). Complete normalized microarray datasets: H0351.2001, H0351.2002, H0351.1009, H0351.1012, H0351.1015, H0351.1016. *Allen Brain Atlas*. Retrieved from <http://human.brain-map.org/static/download>
- Allen Institute for Brain Science. (2019). Developmental Transcriptome Dataset: Exon microarray summarized to genes, RNA-Seq Gencode v10 summarized to genes. BrainSpan Atlas of the Developing Human Brain. Retrieved from <http://www.brainspan.org/static/download.html>
- R Core Team (2018). R version 3.5.2: A Language and Environment for Statistical Computing [software program]. *R Foundation for Statistical Computing*, Vienna, Austria. Accessed from <https://www.R-project.org>
- RStudio, Inc. (2018). RStudio Version 1.1.463 [software program], *the Foundation for Open Access Statistics*. Accessed from <https://www.rstudio.com/products/rstudio/>