**1   LaminaRGeneVis: a tool to visualize gene expression across the laminar**
**2   architecture of the human neocortex**

3   Authors: Ethan Kim[1,2], Derek Howard[1], Yuxiao Chen[1], Shreejoy J. Tripathy[1,2,4,*], Leon

4   French[1,3,4,*]

5        1.  Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health

6            (CAMH), Toronto, Canada

7        2.  Institute for Medical Science, Temerty Faculty of Medicine, University of Toronto,

8            Toronto, Canada

9        3.  Campbell Family Mental Health Research Institute, Centre for Addiction and

10           Mental Health, Toronto, Canada

11       4.  Department of Psychiatry, University of Toronto, Toronto, Canada

12   * Denotes equal contribution

13   Corresponding Author: Leon French (leonfrench@gmail.com)

14   **Abstract**

15   Application of RNA sequencing has enabled the characterization of genome-wide gene

16   expression in the human brain, including distinct layers of the neocortex.

17   Neuroanatomically, the molecular patterns that underlie the laminar organization of the

18   neocortex can help link structure to circuitry and function.  To advance our

19   understanding of cortical architecture, we created *LaminaRGeneVis*, a web application

20   that displays across-layer cortical gene expression from multiple datasets. These

21   datasets were collected using bulk, single-nucleus, and spatial RNA sequencing

22   methodologies and these data were harmonized to facilitate comparisons between

23    datasets. The online resource facilitates single- and multi-gene analyses by providing

24    figures and statistics for user-friendly assessment of laminar gene expression patterns

25    in the adult human neocortex.

26    **Availability and implementation**

27    *LaminaRGeneVis* is available at https://ethanhkim.shinyapps.io/laminargenevis. The

28    source code and data is accessible at https://github.com/ethanhkim/laminargenevis.

29    **1. Introduction**

30    RNA sequencing has provided molecular markers of human brain anatomy by revealing

31    spatial gene expression patterns. The application of these techniques has provided

32    genome-wide profiles of expression, but the brain's complexity has limited our

33    understanding of its cellular, molecular, and laminar architecture. Specifically, while the

34    cytoarchitecture of the recently evolved neocortex has been characterized, we lack a

35    strong understanding of its layer-specific gene expression.

36    Currently, few web applications visualize gene expression in the adult human

37    neocortex. The Allen Brain Atlases from the Allen Institute of Brain Science (AIBS) and

38    other tools allow viewing spatial expression patterns (Guo et al., 2019; Maynard et al.,

39    2021; Shen et al., 2012; Zeng et al., 2012). However, there are no visualization tools to

40    analyze expression across human neocortical layers. Several datasets provide this

41    laminar data but due to differences in the scopes and methods used, accessing and

42    comparing this data is difficult and time-consuming.

43    Here, we present LaminaRGeneVis, a web application for analyses of gene expression

44    across human neocortical layers. LaminaRGeneVis enables visualization and analysis

45  of data from layer-specific bulk-tissue, single-nucleus, and spatial transcriptomic RNA

46  sequencing studies.

47  **2. Data and Methods**

48  2.1 Datasets

49  We used data from three studies that assayed genome-wide expression across the

50  layers of the human neocortex in neurotypical donors. First, He and colleagues

51  transversely sliced dorsolateral prefrontal cortex samples (DLPFC) from postmortem

52  brains (He et al., 2017). Guided by their analyses, we focused on their first dataset

53  (DS1) which contains expression data from four brains and was obtained from SRA

54  using project code SRP065273. A second study of the DLPFC from 3 adult donors

55  employing the spatial transcriptomics 10X Genomics Visium platform was obtained from

56  the spatialLIBD R package (Maynard et al., 2021). The third dataset is from AIBS and

57  assayed expression with single-nucleus RNA sequencing (snRNA-seq) in 3 brains

58  (https://portal.brain-map.org/atlases-and-data/rnaseq/human-multiple-cortical-areas-

59  smart-seq). We chose to use data only from the middle temporal gyrus, as this region

60  had the most samples. The data was also split into three cell types as labelled by AIBS:

61  GABAergic, glutamatergic and non-neuronal. While methods for spatial dissection vary,

62  all three of these studies employed RNA sequencing and profiled the adult human

63  neocortex. The characteristics of these datasets are described in Supplementary Table

64  1. Similarity analyses were performed on the datasets to validate subsequent

65  expression correlation and layer-specific enrichment analyses. The results of those

66  analyses are available in Supplementary Results, Tables S2 and S3.

67    2.1.1 Dataset processing

68    Each dataset was processed using a similar pipeline as described in Supplementary

69    Methods. This processing resulted in each dataset being represented in a gene

70    expression matrix in the shape of genes (rows) by layer (columns), with gene

71    expression represented as normalized and log-transformed read counts (log2(counts

72    per million)).

73    2.2 Statistical analysis

74    2.2.1 Gene expression correlation

75    Pearson correlation was used to assess agreement for single genes in the bulk-tissue

76    data. This is calculated using a given gene's seven layer-specific expression values

77    from the He and Maynard datasets. For multiple genes, average correlation is used. To

78    add a genome-wide perspective, the percentile of these correlations in reference to all

79    other gene-to-gene values are also calculated.

80    2.2.2 Layer-specific gene set enrichment

81    To test for the enrichment of a set of genes in a given layer, we first removed genes

82    with CPM < 0.1 across all layers, and normalized through log2-transformation then z-

83    score normalized. We sorted the normalized expression matrices per layer by ranking

84    the remaining genes in each dataset from the most to least normalized expression.

85    Within this ranking, the area under the receiver operating characteristic curve (AUC)

86    was used to test whether the inputted set of genes were enriched or depleted

87    (enrichment: AUC > 0.5, depletion: AUC < 0.5). The Mann–Whitney U test was used to

88    determine statistical significance and multiple-test corrected with Bonferroni correction.

89    2.2.3 Availability of data and code

90    Data used in the application and the code to process the data are available online at

91    https://github.com/ethanhkim/laminargenevis. Scripts to process the raw data from He

92    et al. are available at https://github.com/derekhoward/he_seq.

93    **The LaminaRGeneVis application**

94    *LaminaRGeneVis* is a web application implemented in shiny (Chang et al., 2021) that

95    can be accessed at https://ethanhkim/shinyapps.io/laminargenevis. Users can choose

96    either Single or Multi-gene mode to visualize gene expression for one or a set of genes,

97    respectively. The application will also report statistical analyses that assay agreement in

98    the bulk-tissue data and test layer-specific enrichment for a set of genes.

99    Single gene mode

100   The user can input their gene of interest by typing in its gene symbol and selecting it

101   from the drop-down bar. Once submitted, the gene's normalized expression across the

102   cortical layers in each dataset is displayed as a bar plot. A text box below notes which

103   datasets assayed the queried gene and the agreement statistics in bulk-tissue datasets.

104   Multi-gene mode

105   In the Multi-gene mode, the application generates visualizations for the queried gene

106   set's layer-specific enrichment and normalized expression of each gene across

107   datasets.

108   Layer-specific enrichment is visualized as a heatmap that displays the AUC values

109   across layers and datasets. Normalized gene expression is displayed as a heatmap for

110   30 genes or less and a scatterplot otherwise. The scatterplots also show the queried

111   genes' median expression in each layer. A summary textbox at the bottom reports

112     information such as the number of inputted genes assayed in each bulk-tissue dataset

113     and dataset agreement statistics across the bulk-tissue data.
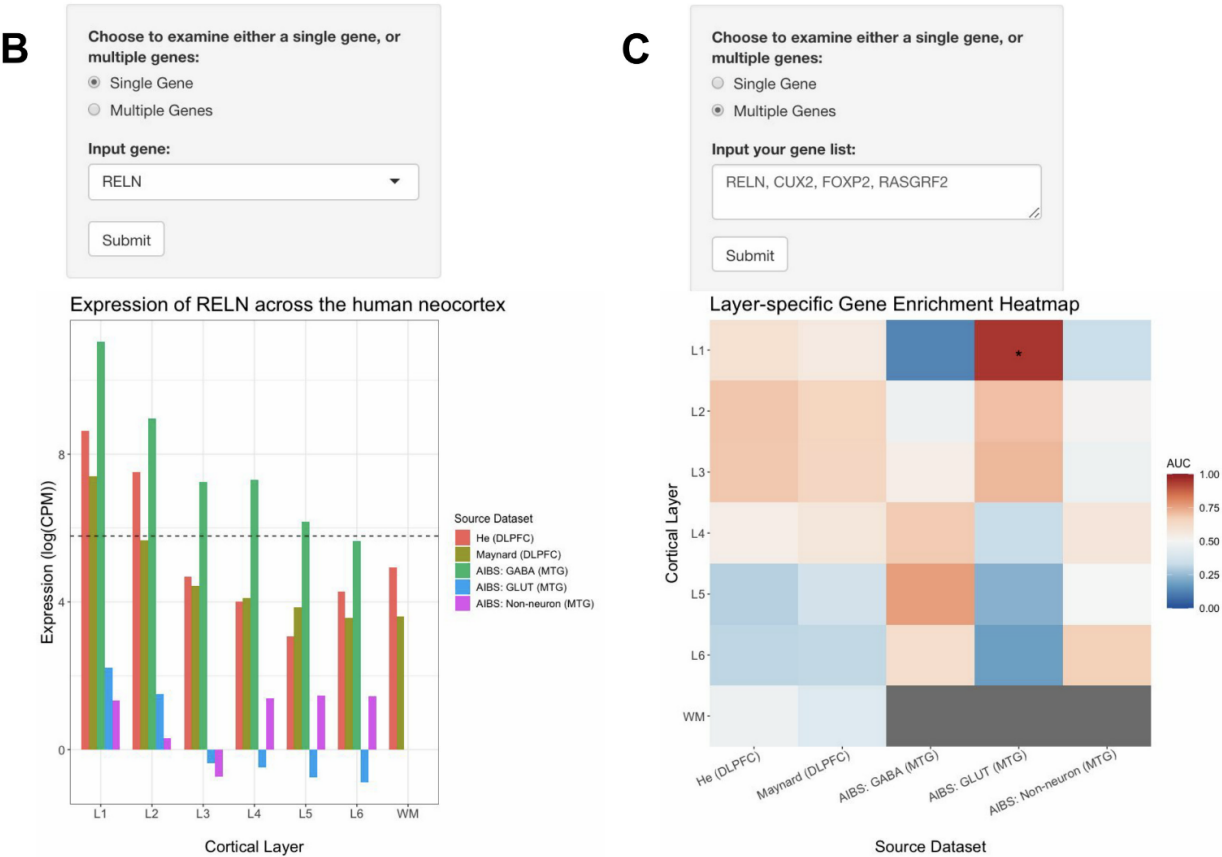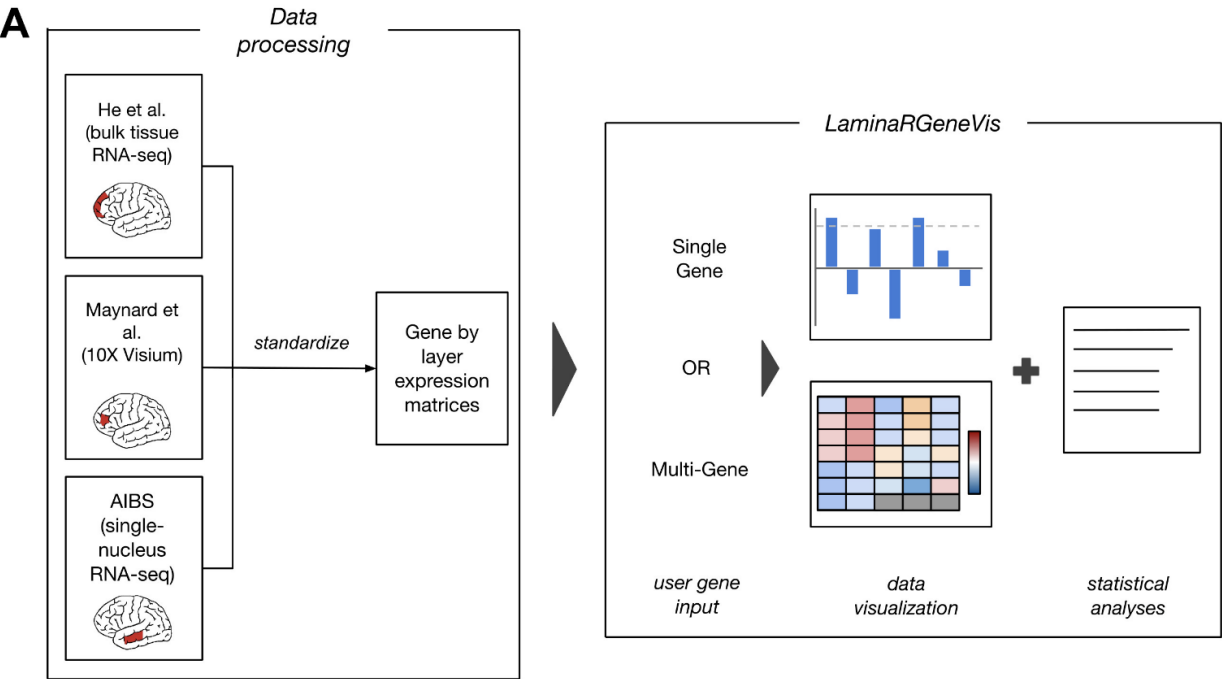
114     **Conclusion**

115     We have developed a web application for visualizing gene expression across the

116     laminar architecture of the adult human neocortex. It reports cross-dataset correlation

117     and the enrichment of layer-specific expression. These functionalities provide easily

118     accessible figures and statistics for quick assessment of expression across the layers of

119     the human cortex.

120     **Acknowledgements**

124    **Fig. 1**



125

126    **Fig. 1.** Workflow diagram and visualizations of gene expression across the cortical

127    layers. (**A**) Diagram of data flow and application usage. Each dataset was uniformly

128    processed and standardized to create normalized gene expression matrices. These

129    matrices are used in the web application for visualizations and statistical analyses.

130    Dataset characteristics are summarized in Supplementary Table S1. **(B)** When the user

131    chooses to examine a single gene with the selection options, the application will display

132    a bar plot showing the normalized expression of the gene across the cortical layers and

133    white matter. **(C)** Upon choosing to examine multiple genes, the application will

134    visualize the genes' layer-specific enrichment (as shown) as a heatmap with asterisks

135    marking statistically significant AUC scores.

136

137    **References**

138    Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J.,

139        McPherson, J., Dipert, A., & Borges, B. (2021). *shiny: Web Application*

140        *Framework for R.* (1.6.0) [R]. https://CRAN.R-project.org/package=shiny

141    Guo, L., Lin, W., Zhang, Y., Li, W., & Wang, J. (2019). BEST: A web server for brain

142        expression Spatio-temporal pattern analysis. *BMC Bioinformatics*, *20*(1), 632.

143        https://doi.org/10.1186/s12859-019-3222-6

144    He, Z., Han, D., Efimova, O., Guijarro, P., Yu, Q., Oleksiak, A., Jiang, S., Anokhin, K.,

145        Velichkovsky, B., Grünewald, S., & Khaitovich, P. (2017). Comprehensive

146        transcriptome analysis of neocortical layers in humans, chimpanzees and

147        macaques. *Nature Neuroscience*, *20*(6), 886–895.

148        https://doi.org/10.1038/nn.4548

149 Maynard, K. R., Collado-Torres, L., Weber, L. M., Uytingco, C., Barry, B. K., Williams, S.

150   R., Catallini, J. L., Tran, M. N., Besich, Z., Tippani, M., Chew, J., Yin, Y.,

151   Kleinman, J. E., Hyde, T. M., Rao, N., Hicks, S. C., Martinowich, K., & Jaffe, A. E.

152   (2021). Transcriptome-scale spatial gene expression in the human dorsolateral

153   prefrontal cortex. *Nature Neuroscience*, 1–12. https://doi.org/10.1038/s41593-

154   020-00787-0

155 Shen, E. H., Overly, C. C., & Jones, A. R. (2012). The Allen Human Brain Atlas:

156   Comprehensive gene expression mapping of the human brain. *Trends in*

157   *Neurosciences*, *35*(12), 711–714. https://doi.org/10.1016/j.tins.2012.09.005

158 Zeng, H., Shen, E. H., Hohmann, J. G., Oh, S. W., Bernard, A., Royall, J. J., Glattfelder,

159   K. J., Sunkin, S. M., Morris, J. A., Guillozet-Bongaarts, A. L., Smith, K. A., Ebbert,

160   A. J., Swanson, B., Kuan, L., Page, D. T., Overly, C. C., Lein, E. S., Hawrylycz,

161   M. J., Hof, P. R., … Jones, A. R. (2012). Large-scale cellular-resolution gene

162   profiling in human neocortex reveals species-specific molecular signatures. *Cell*,

163   *149*(2), 483–496. https://doi.org/10.1016/j.cell.2012.02.052