

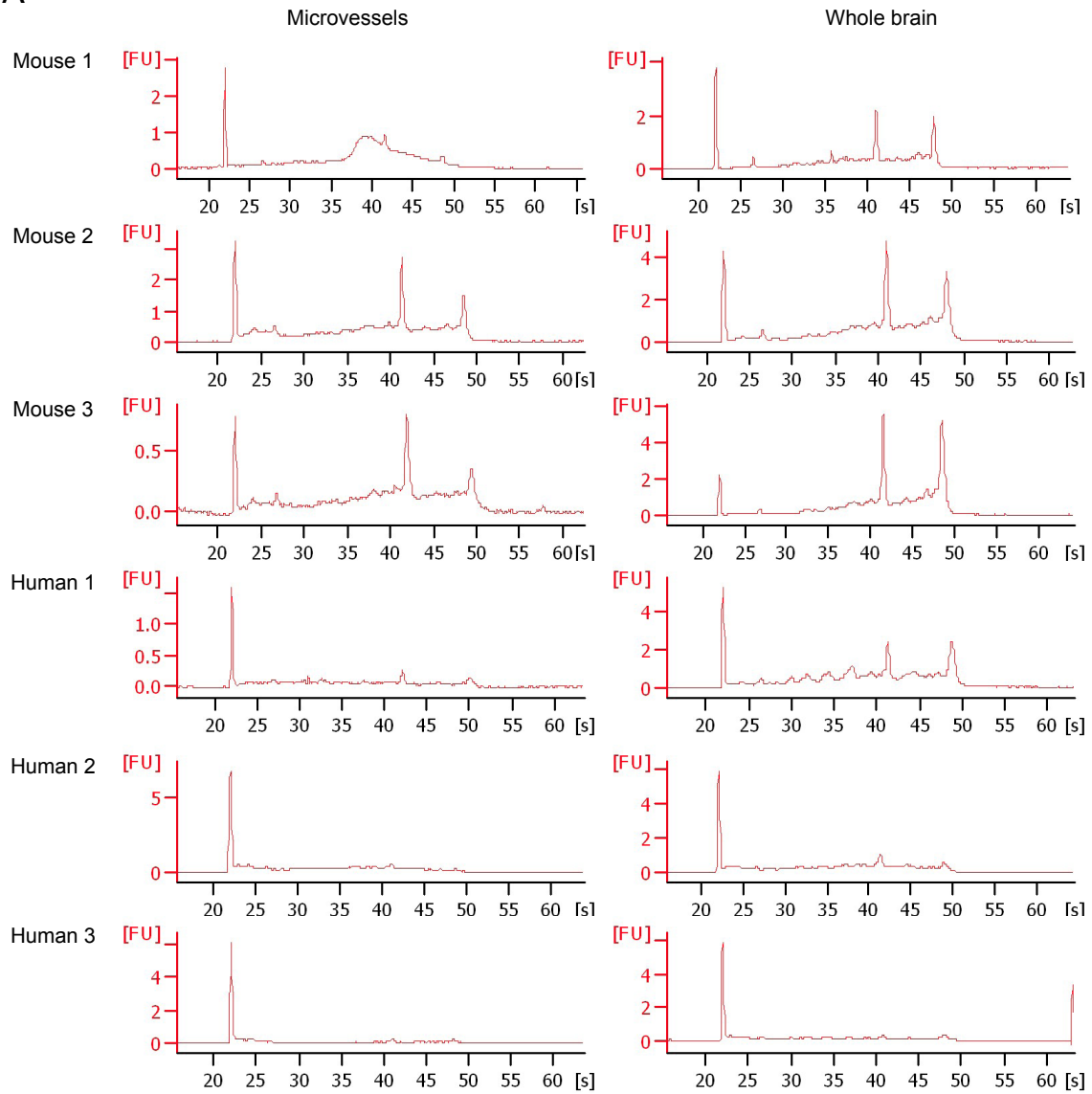
Transcriptomic Comparison of Human and Mouse Brain Microvessels

Hannah W. Song, Koji L. Foreman, Benjamin D. Gastfriend, John S. Kuo, Sean P. Palecek
and Eric V. Shusta

Supplementary information.

- **Figure S1.** RNA electropherograms.
- **Figure S2.** Gene Set Enrichment Analysis (GSEA) of human LCM microvessel datasets.
- **Figure S3.** Lack of *VTN* expression in human brain pericytes.
- **Figure S4.** Validation of putative human-enriched vascular transcripts.
- **Figure S5.** Alternative filtering strategies for identifying species differences in vascular gene expression.
- **Figure S6.** Mouse-human species differences in *SLC* and *ABC* transporter expression.
- **Figure S7.** Validation of putative human brain pericyte-enriched transcripts.
- **Table S1.** Characteristics of RNA samples and associated RNA-seq datasets.

A



B

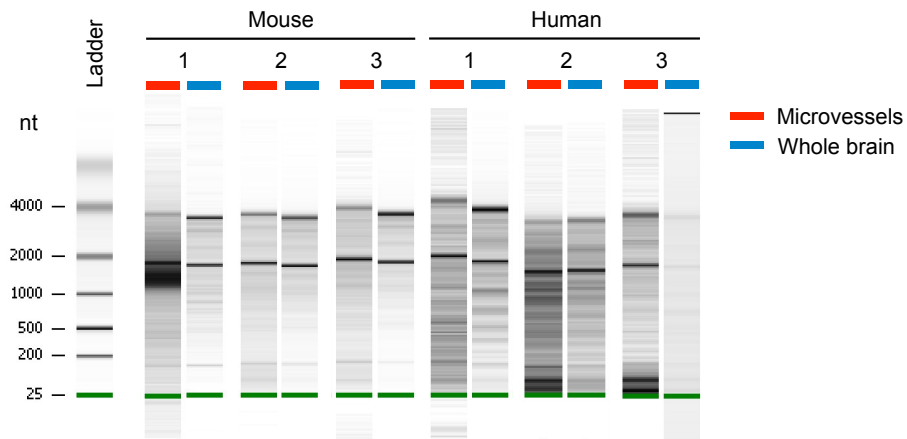


Figure S1. RNA electropherograms. (A,B) Agilent Bioanalyzer analysis of RNA samples from mouse and human LCM microvessel and whole brain samples. Electropherograms (A) and gel-like images (B) are shown.

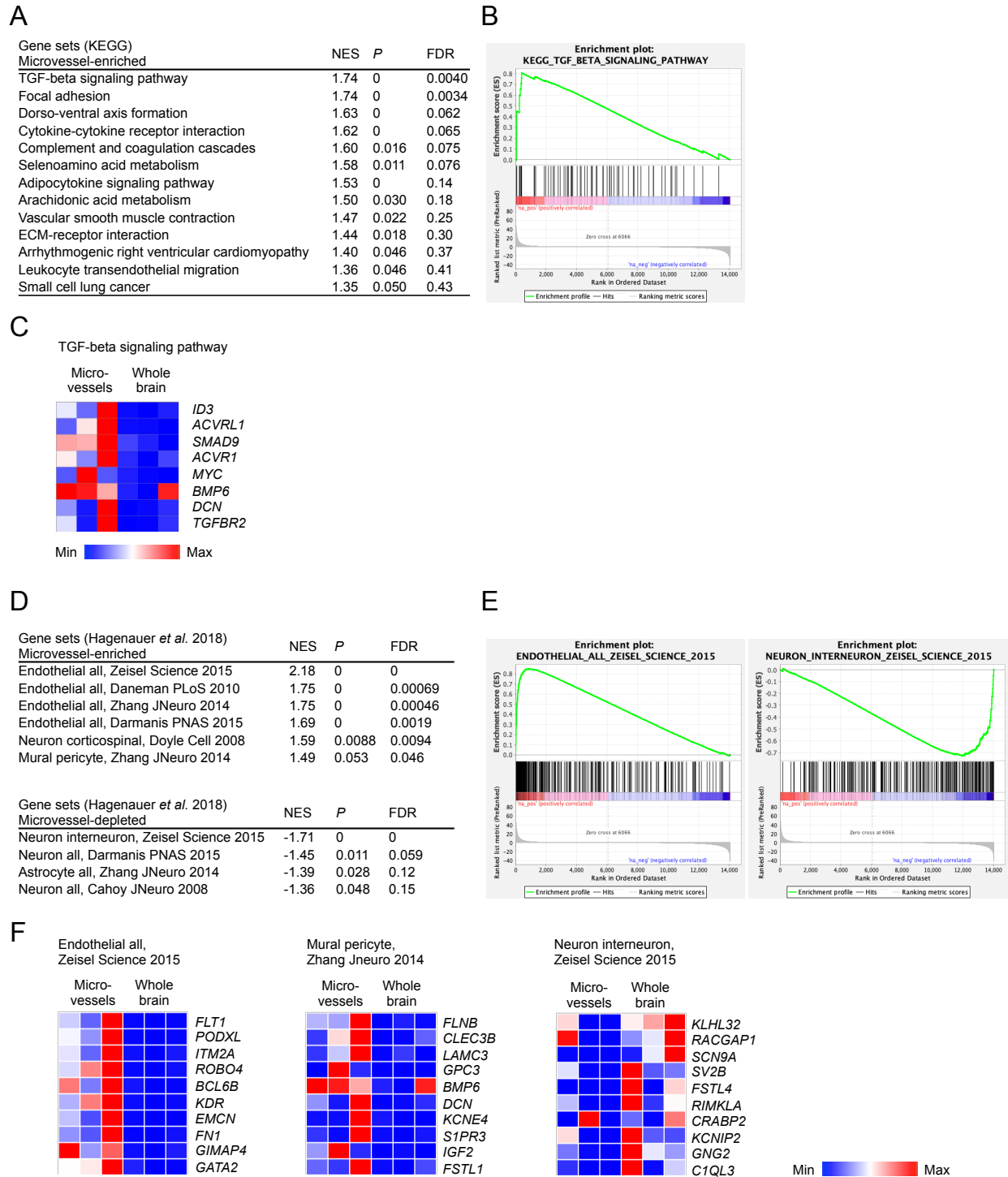


Figure S2. Gene Set Enrichment Analysis (GSEA) of human LCM microvessel datasets. (A)

Gene sets from the KEGG database enriched (P or $FDR < 0.05$) in human LCM microvessels compared to whole brain. NES: normalized enrichment score. The GSEA input was a list of

genes ranked from the highest-confidence LCM microvessel-enriched to the highest-confidence LCM microvessel-depleted gene, using the ranking metric $-\log_{10}(P) \times \text{sign}[\log_2(\text{fold change})]$ (Supplementary Table S4). **(B)** GSEA enrichment plots for the KEGG TGF-beta signaling pathway gene sets. **(C)** Heat maps illustrating transcript abundance in biological triplicates of LCM microvessels and whole brain for the highest ranked LCM microvessel-enriched genes in the indicated gene sets. Color indicates expression that has been normalized within each gene (row). **(D)** Gene sets from Hagenauer *et al.* [30] enriched (P or FDR < 0.05) in human LCM microvessels compared to whole brain (top) and enriched (P or FDR < 0.05) in whole brain compared to LCM microvessels (bottom). GSEA input was as described in (A). **(E)** GSEA enrichment plots for endothelial and interneuron gene sets from Hagenauer *et al.* [30]. **(F)** Heat maps illustrating transcript abundance in biological triplicates of LCM microvessels and whole brain for the highest ranked LCM microvessel-enriched genes in the indicated gene sets. Color indicates expression that has been normalized within each gene (row).

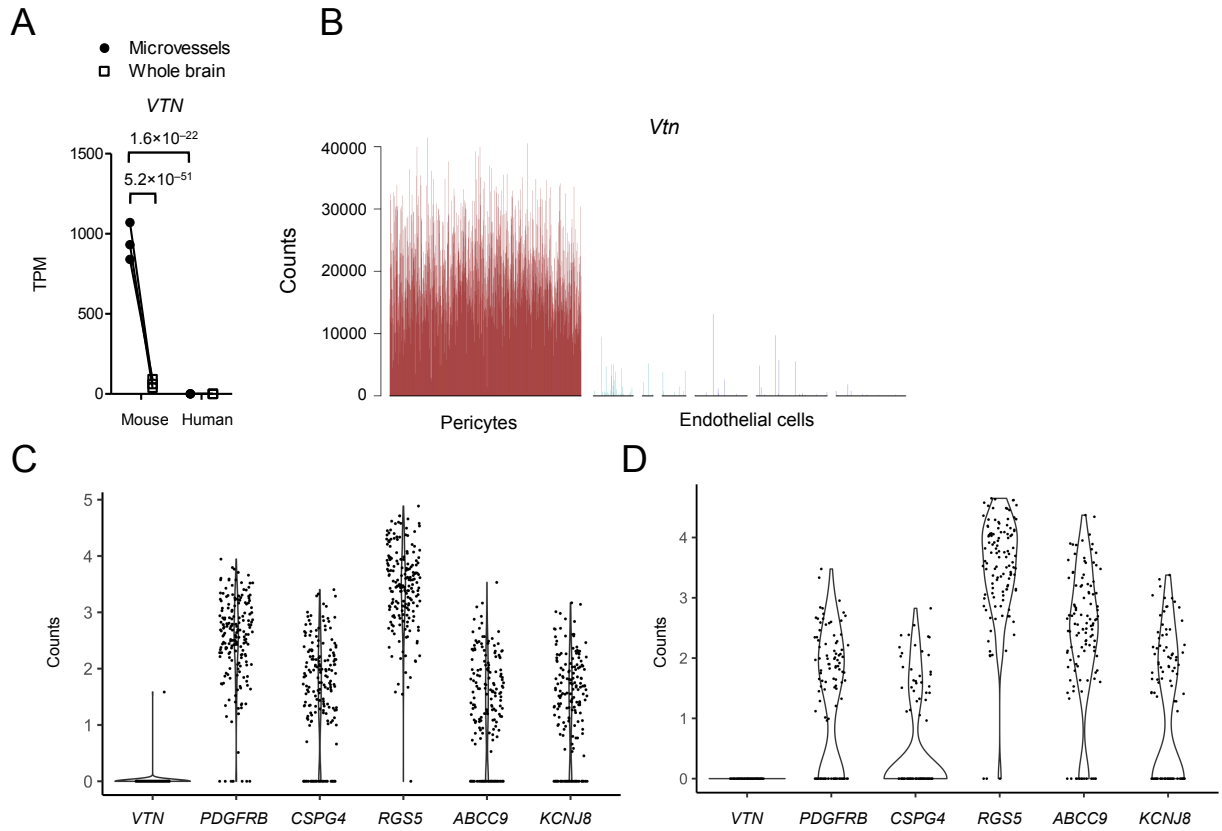
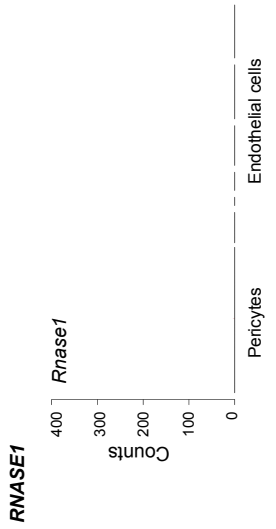
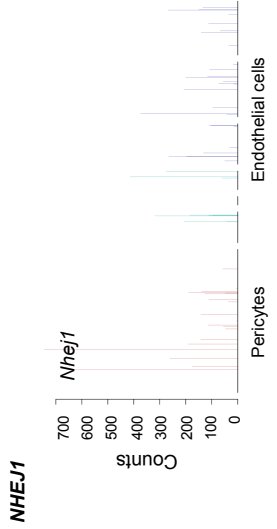
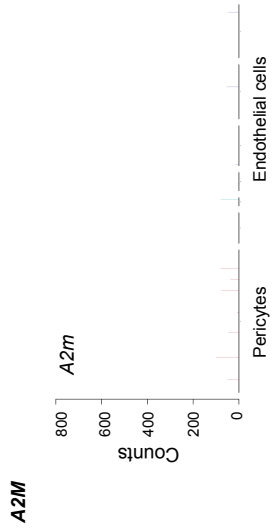
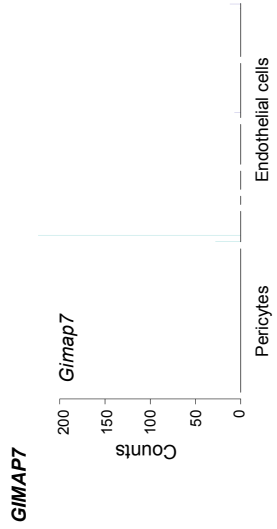
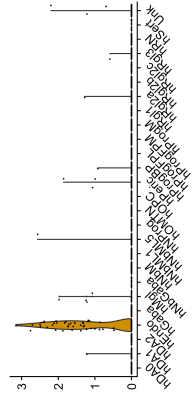
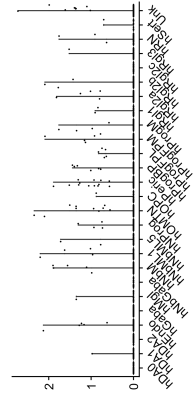
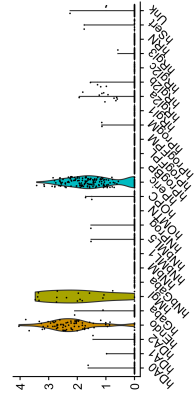
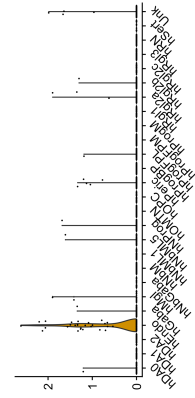


Figure S3. Lack of *VTN* expression in human brain pericytes. **(A)** Transcript abundance of *VTN/Vtn* in biological triplicates of LCM microvessels and whole brain from human and mouse. Lines connect datapoints from matched LCM microvessel and whole brain samples. Adjusted P-values (from DESeq2) are shown; Wald test with Benjamini-Hochberg correction. **(B)** *Vtn* expression in single cell RNA-seq of mouse brain pericytes and endothelial cells [9,10]. Single cell RNA-seq plot in this panel adapted from <http://betsholtzlab.org/VascularSingleCells/database.html>. **(C,D)** *VTN* expression in human brain pericytes. Other pericyte genes (*PDGFRB*, *CSPG4*, *RGS5*, *ABCC9*, and *KCNJ8*) are included for comparison. Single cell RNA-seq data from (C) La Manno *et al.* [18] and (D) Polioudakis *et al.* [19].

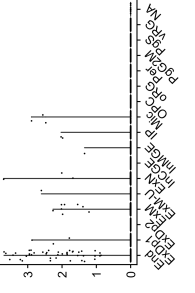
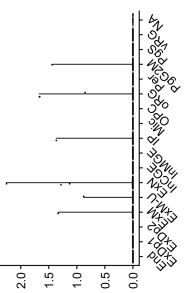
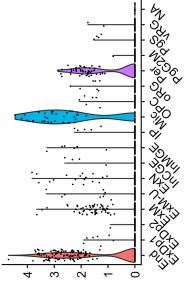
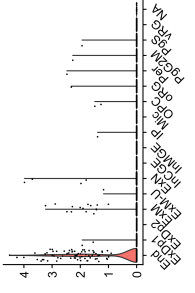
i. Vanlandewijck *et al.* Mouse
brain vascular scRNA-seq



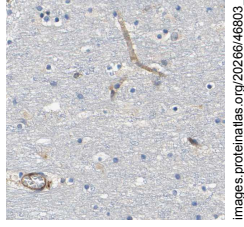
ii. La Manno *et al.* Human
embryonic midbrain scRNA-seq



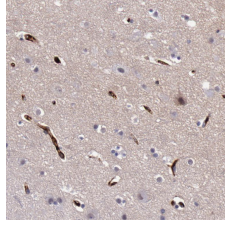
iii. Polioudakis *et al.* Human
embryonic neocortex scRNA-seq



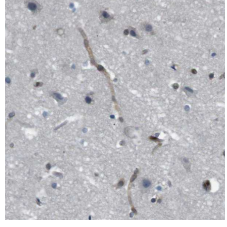
iv. Human Protein Atlas cortex
immunohistochemistry



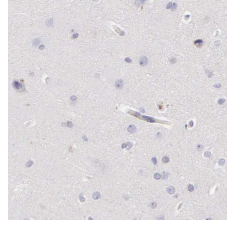
images.proteinatlas.org/20266/46803_B_7_5.jpg



images.proteinatlas.org/22657/607_B_8_5.jpg



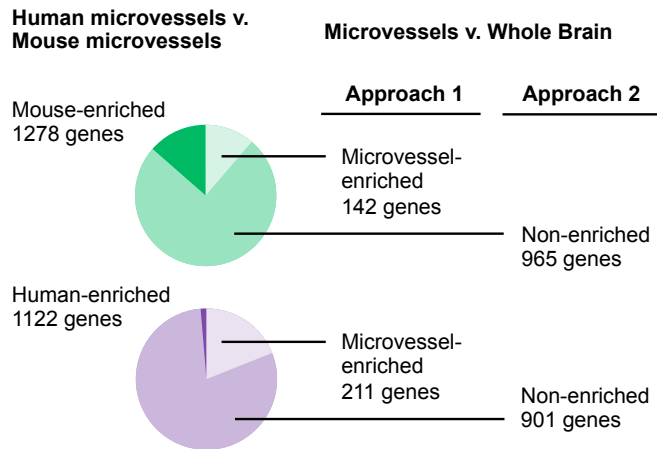
images.proteinatlas.org/12334/32934_B_7_5.jpg



images.proteinatlas.org/11402/634_B_8_5.jpg

Figure S4. Validation of putative human-enriched vascular transcripts. Each row shows data from several sources for one transcript/protein of interest. (i) Transcript expression in single cell RNA-seq of mouse brain pericytes and endothelial cells [9,10]. Single cell RNA-seq plots in these panels adapted from <http://betsholtzlab.org/VascularSingleCells/database.html>. (ii) Transcript abundance (counts) in human embryonic midbrain single cell RNA-seq data [18]. Cell assignment to clusters and cluster names are as reported by La Manno *et al.* [18]. hPeric: human pericytes. hEndo: human endothelial cells. (iii) Transcript abundance (counts) in human embryonic neocortex single cell RNA-seq data [19]. Cell assignment to clusters and cluster names are as reported by Polioudakis *et al.* [19]. Per: pericytes. End: endothelial cells. (iv) Immunohistochemistry on human cortex samples from the Human Protein Atlas (v19.proteinatlas.org) [32]. URLs for each image are provided in the figure.

A



B

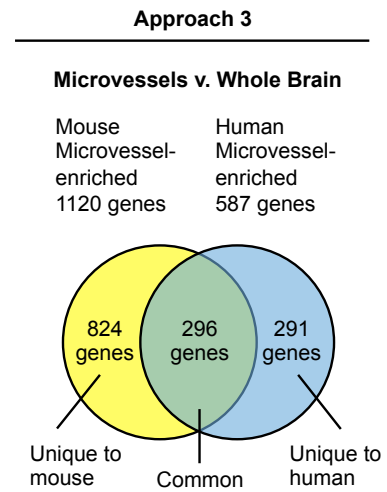


Figure S5. Alternative filtering strategies for identifying species differences in vascular gene expression. (A) Summary of Approaches 1 and 2. Approach 1 is as shown in Figure 4. Of the 1278 mouse-enriched transcripts identified in (A), Approach 2 selects the 965 genes not enriched in mouse microvessels or whole brain (as determined in Fig. 3E). Of the 1122 human-enriched transcripts identified in (A), Approach 2 selects the 901 genes not enriched in human microvessels or whole brain (as determined in Fig. 3F). The small number of genes not selected by Approach 1 or Approach 2 (shaded dark) are whole brain-enriched. Complete filtered gene lists are in Supplementary Table S5. **(B)** Summary of Approach 3, a filtering strategy that identifies species-specific microvessel-enriched genes by comparing the lists of mouse genes with microvessel enrichment versus whole brain (as determined in Fig. 3E) and human genes with microvessel enrichment versus whole brain (as determined in Fig. 3F). 824 genes appear only on the mouse list and 291 genes appear only on the human list. Complete filtered gene lists are in Supplementary Table S5.

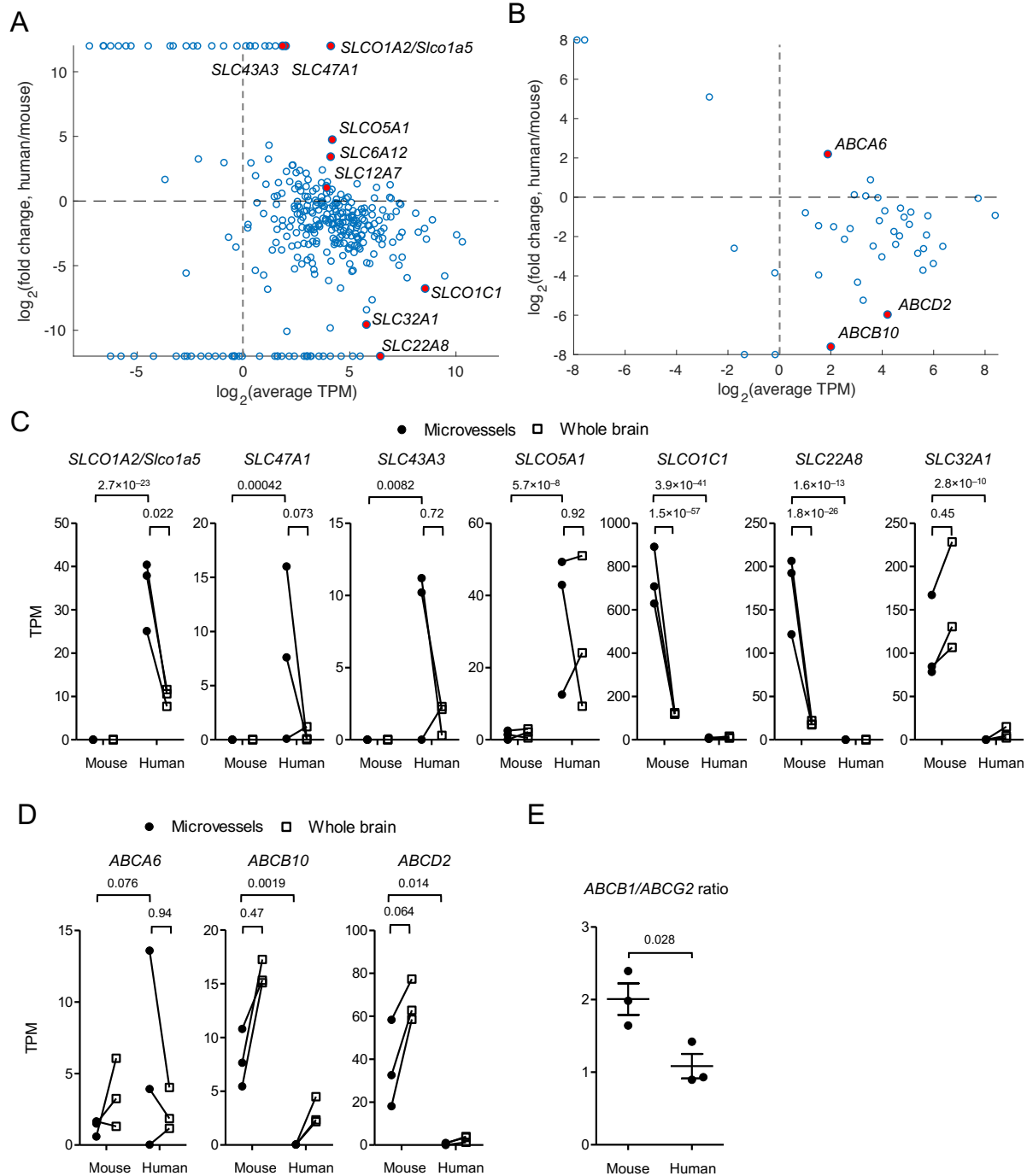
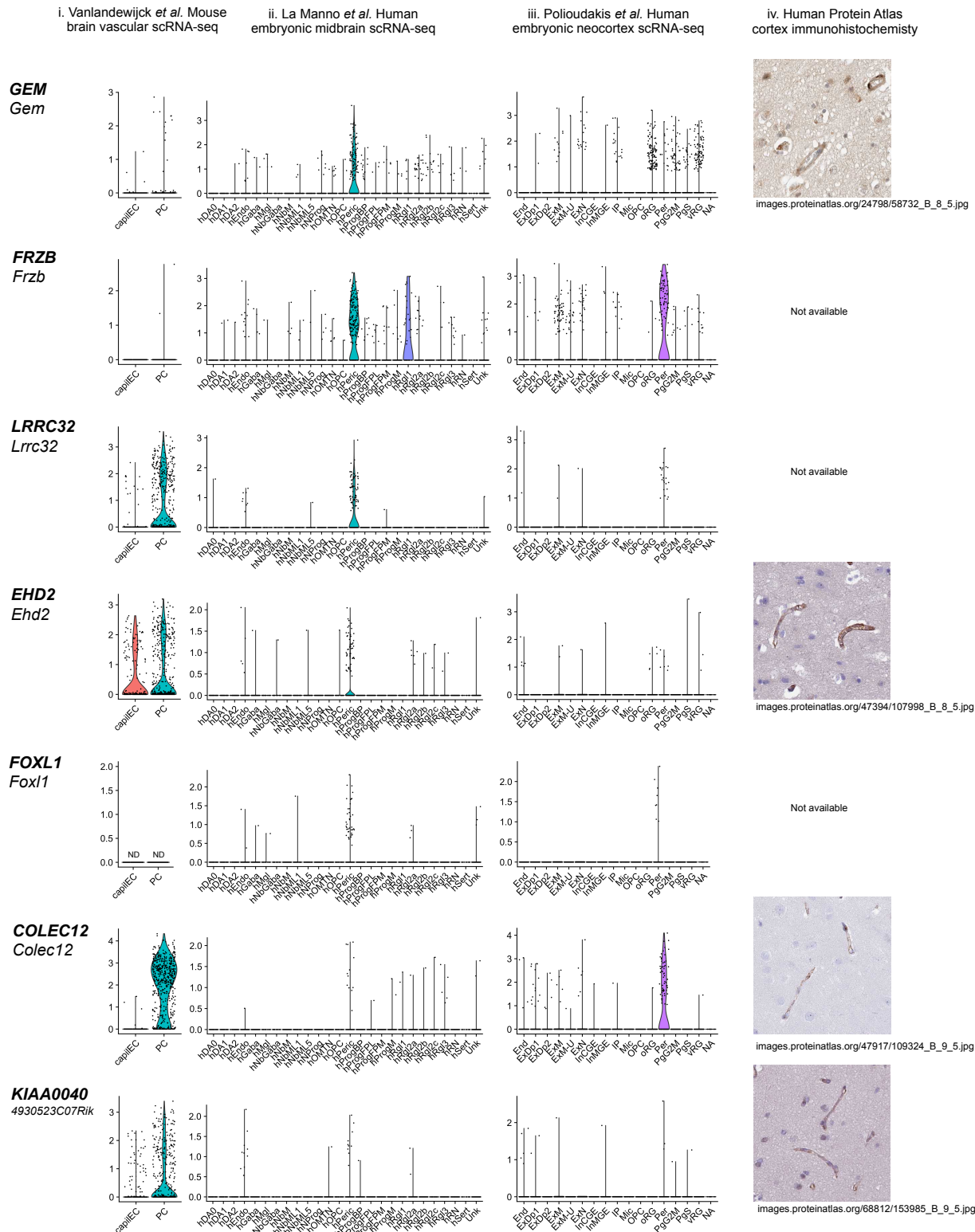
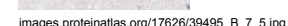
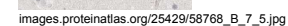
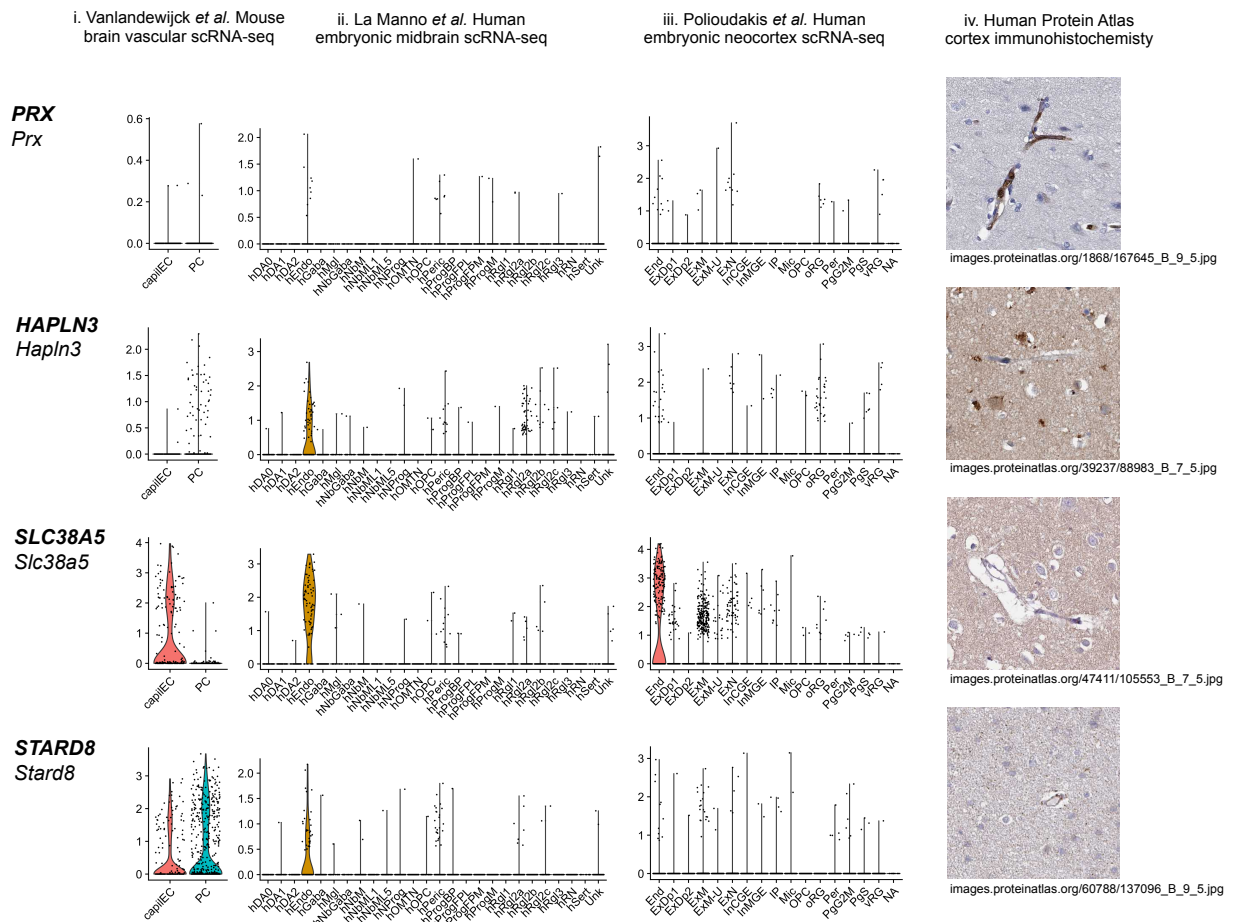


Figure S6. Mouse-human species differences in *SLC* and *ABC* transporter expression. (A) Comparison of solute carrier (*SLC*) transcript abundance in human and mouse LCM microvessels. The $\log_2(\text{fold change})$ of human transcript abundance versus the homologous mouse transcript abundance is plotted against the average abundance in human and mouse. Each

point represents one gene with known mouse-human homology. Genes with zero expression in all samples were excluded. Genes to the right of the vertical line at $\log_2(\text{average TPM}) = 0$ are moderately or highly expressed in at least one species. A subset of human-enriched *SLCs* (*SLCO1A2*, *SLC47A1*, *SLC43A3*, *SLCO5A1*, *SLC6A12*, and *SLC12A7*) and mouse-enriched *SLCs* (*SLCO1C1*, *SLC22A8*, and *SLC32A1*) are highlighted. Genes with zero expression in one species were visualized with an average $\log_2(\text{fold change})$ of ± 12 . Complete results of this analysis are in Supplementary Table S5. **(B)** Comparison of ATP binding cassette (*ABC*) transcript abundance in human and mouse LCM microvessels. The $\log_2(\text{fold change})$ of human transcript abundance versus the homologous mouse transcript abundance is plotted against the average abundance in human and mouse. Each point represents one gene with known mouse-human homology. Genes with zero expression in all samples were excluded. Genes to the right of the vertical line at $\log_2(\text{average TPM}) = 0$ are moderately or highly expressed in at least one species. A subset of human-enriched *ABCs* (*ABCA6*) and mouse-enriched *ABCs* (*ABCB10* and *ABCD2*) are highlighted. Genes with zero expression in one species were visualized with an average $\log_2(\text{fold change})$ of ± 8 . Complete results of this analysis are in Supplementary Table S5. **(C)** Transcript abundance of selected genes in biological triplicates of LCM microvessels and whole brain from human and mouse. *Slco1a5* is the mouse homolog of human *SLCO1A2*. Lines connect datapoints from matched LCM microvessel and whole brain samples. Adjusted P-values (from DESeq2) are shown; Wald test with Benjamini-Hochberg correction. **(D)** Transcript abundance of genes highlighted in (B) in biological triplicates of LCM microvessels and whole brain from human and mouse. Lines connect datapoints from matched LCM microvessel and whole brain samples. Adjusted P-values (from DESeq2) are shown; Wald test with Benjamini-Hochberg correction. **(E)** Ratio of *ABCB1* (mouse *Abcb1a*) to *ABCG2* transcript abundance (TPM) in human and mouse LCM microvessels. P-value shown from Student's *t* test.



iv. Human Protein Atlas
cortex immunohistochemistry



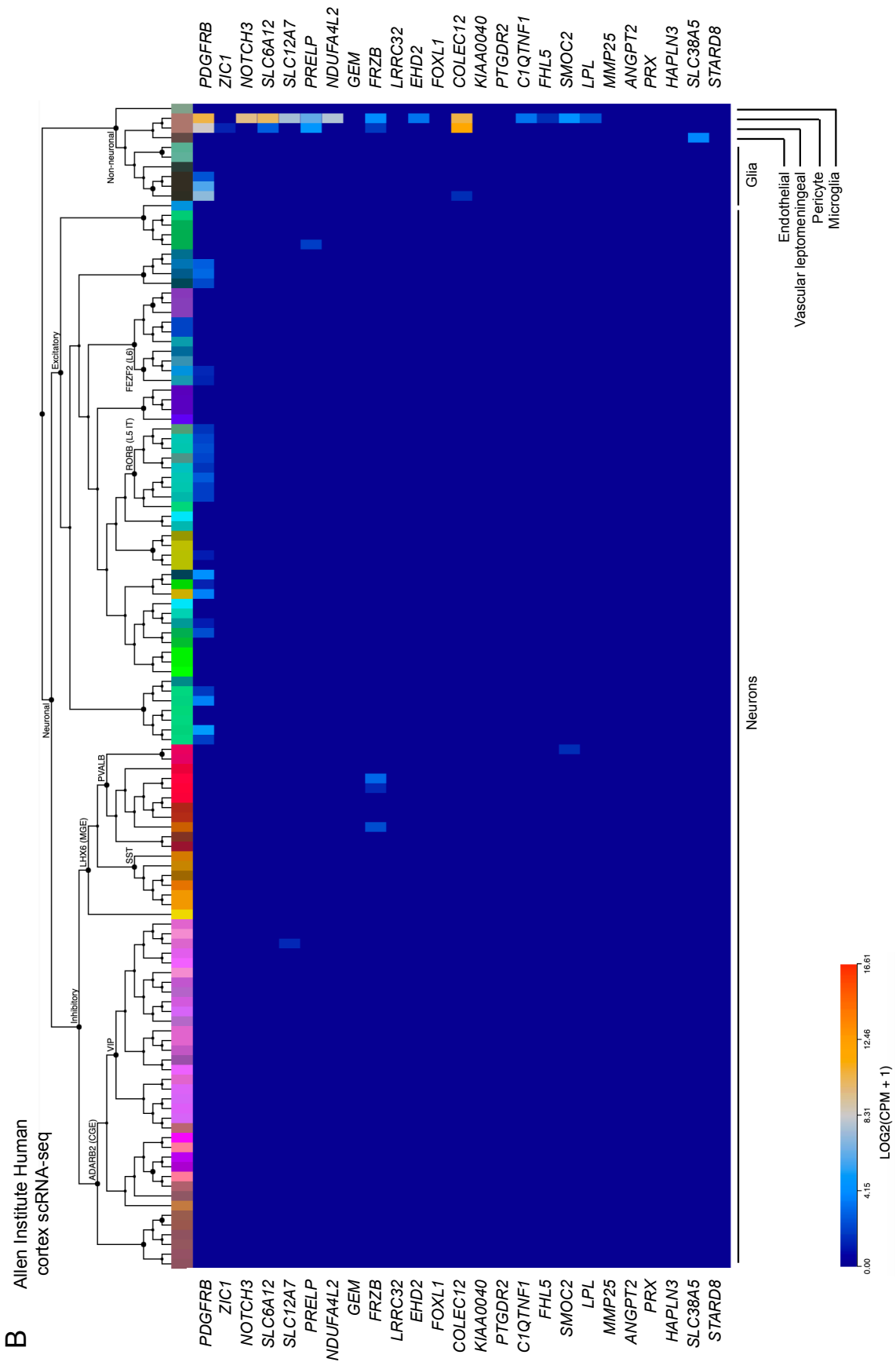


Figure S7. Validation of putative human brain pericyte-enriched transcripts. (A) Each row shows data from several sources for one transcript/protein of interest. (i) Transcript abundance (counts) in mouse brain pericytes and capillary endothelial cells [9,10]. Cell assignment to clusters and cluster names are as reported by Vanlandewijck *et al.* [9,10]. (ii) Transcript abundance (counts) in human embryonic midbrain single cell RNA-seq data [18]. Cell assignment to clusters and cluster names are as reported by La Manno *et al.* [18]. hPeric: human pericytes. hEndo: human endothelial cells. (iii) Transcript abundance (counts) in human embryonic neocortex single cell RNA-seq data [19]. Cell assignment to clusters and cluster names are as reported by Polioudakis *et al.* [19]. Per: pericytes. End: endothelial cells. (iv) Immunohistochemistry on human cortex samples from the Human Protein Atlas (v19.proteinatlas.org) [32]. URLs for each image are provided in the figure. (B) Heat map of transcript abundance in human cortex single cell and single nucleus RNA-seq [39]. Figure adapted from <http://celltypes.brain-map.org/rnaseq/human/cortex>.

Table S1. Characteristics of RNA samples and associated RNA-seq datasets

Sample	Number collected	Total surface area (mm²)	Total RNA (ng)	Total reads	Uniquely mapped reads (%)	Multi-mapping reads (%)
Mouse microvessels 1	1839	2.2	1.0	19,703,651	82.95	9.75
Mouse microvessels 2	3058	3.0	3.1	22,370,744	83.48	11.67
Mouse microvessels 3	2841	2.5	1.0	21,081,514	85.02	10.73
Mouse whole brain 1	28	2.1	2.2	19,617,678	88.88	8.69
Mouse whole brain 2	22	3.0	4.4	19,500,744	80.91	9.23
Mouse whole brain 3	21	2.4	4.4	19,206,425	88.96	8.56
Human microvessels 1	2004	2.3	0.8	27,757,265	58.47	21.82
Human microvessels 2	2577	3.0	4.6	27,762,525	43.84	33.75
Human microvessels 3	2200	2.3	0.9	27,161,926	56.96	24.01
Human whole brain 1	27	2.3	5.5	24,842,174	59.55	22.43
Human whole brain 2	34	3.0	4.7	33,073,030	51.88	28.52
Human whole brain 3	19	2.3	1.5	30,941,286	65.76	15.41