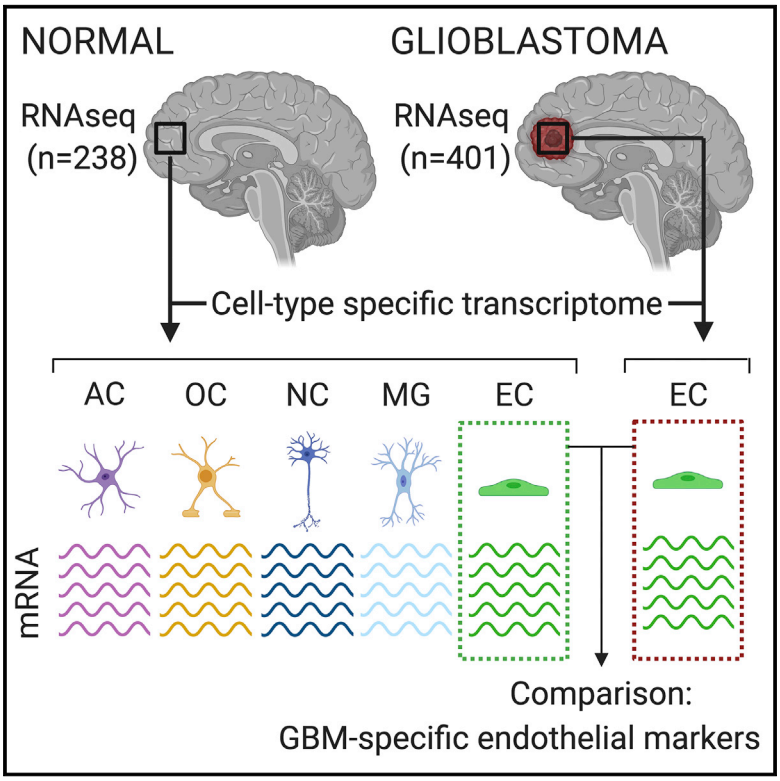
Resource

Cell Reports

A Systems-Based Map of Human Brain Cell-Type Enriched Genes and Malignancy-Associated Endothelial Changes

Graphical Abstract



Authors

Philip Dusart, Björn Mikael Hallström, Thomas Renné, Jacob Odeberg, Mathias Uhlén, Lynn Marie Butler

Correspondence

lynn.butler@ki.se

In Brief

Dusart et al. use a correlation-based bulk RNA-seq analysis method to identify cell-type-enriched transcriptomes in human brain. The endothelial transcriptome in glioblastoma is profiled and compared to normal brain to predict tumor-specific endothelial markers. A web-based portal is provided to allow exploration of cell-type enrichment profiles.

Highlights

- Cell-type reference mRNA used to profile cell-specific transcripts from bulk RNA-seq
- RNA-seq of 238 unfractionated human brain samples from two cohorts was analyzed
- Endothelium-enriched transcriptome was profiled, together with four other cell types
- RNA-seq of 401 unfractionated glioblastomas analyzed for tumor endothelial markers

Dusart et al., 2019, Cell Reports *29*, 1690–1706
November 5, 2019
https://doi.org/10.1016/j.celrep.2019.09.088

CellPress

OPEN ACCESS
CellPress

Cell Reports
Resource

A Systems-Based Map of Human Brain Cell-Type Enriched Genes and Malignancy-Associated Endothelial Changes

Philip Dusart,[1,2] Björn Mikael Hallström,[1] Thomas Renné,[3] Jacob Odeberg,[1,2,4,5] Mathias Uhlén,[1] and Lynn Marie Butler[1,2,3,6,7,*]

[1]Science for Life Laboratory, Department of Protein Science, Royal Institute of Technology (KTH), 171 21 Stockholm, Sweden
[2]K.G. Jebsen Thrombosis Research and Expertise Centre, Department of Clinical Medicine, The Arctic University of Norway, 9019 Tromsø, Norway
[3]Institute for Clinical Chemistry and Laboratory Medicine, University Medical Centre Hamburg-Eppendorf, 20246 Hamburg, Germany
[4]The University Hospital of North Norway (UNN), PB100, 9038 Tromsø, Norway
[5]Department of Hematology, Karolinska University Hospital, 171 77 Stockholm, Sweden
[6]Clinical Chemistry and Blood Coagulation Research, Department of Molecular Medicine and Surgery, Karolinska Institute, 171 76 Stockholm, Sweden
[7]Lead Contact
*Correspondence: lynn.butler@ki.se
https://doi.org/10.1016/j.celrep.2019.09.088

SUMMARY

Changes in the endothelium of the cerebral vasculature can contribute to inflammatory, thrombotic, and malignant disorders. The importance of defining cell-type-specific genes and their modification in disease is increasingly recognized. Here, we develop a bioinformatics-based approach to identify normal brain cell-enriched genes, using bulk RNA sequencing (RNA-seq) data from 238 normal human cortex samples from 2 independent cohorts. We compare endothelial cell-enriched gene profiles with astrocyte, oligodendrocyte, neuron, and microglial cell profiles. Endothelial changes in malignant disease are explored using RNA-seq data from 516 lower-grade gliomas and 401 glioblastomas. Lower-grade gliomas appear to be an "endothelial intermediate" between normal brain and glioblastoma. We apply our method for the prediction of glioblastoma-specific endothelial biomarkers, providing potential diagnostic or therapeutic targets. In summary, we provide a roadmap of endothelial cell identity in normal and malignant brain, using a method developed to resolve bulk RNA-seq into constituent cell-type-enriched profiles.

INTRODUCTION

Comprehensive characterization of human organs, and the constitutive cell types, is required to fully understand biological processes and disease development, concepts underlying large-scale tissue and cell profiling projects, e.g., the Human Protein Atlas (Uhlén et al., 2015) and Human Cell Atlas (Regev et al., 2017). RNA sequencing (RNA-seq) data from unfractionated human normal and diseased tissue is available through online portals, e.g., Genotype-Tissue Expression (GTEx) Project (http://gtexportal.org/home/index.html) (GTEx Consortium, 2015) and The Cancer Genome Atlas (https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga), but using mixed-cell "bulk" data to decipher changes in cell-type-specific transcriptome profiles in disease is challenging. Recent advances have facilitated sequencing of cell populations or individual cells, but practical and technical challenges, such as sourcing of material, occurrence of artifacts due to processing, compromised read depth, and financial constraints, limit the accessibility of such methods (Beliakova-Bethell et al., 2014; Rizzetto et al., 2017; Saliba et al., 2014; Ziegenhain et al., 2017). At the cell-type level, one of the most well-studied organs is the brain, tissue composed primarily of neurons (NCs), astrocytes (ACs), oligodendrocytes (OCs), microglia (MGs), and a vascular network of endothelial cells (ECs) and associated mural cells (Azevedo et al., 2009; von Bartheld et al., 2016; Zhao et al., 2015). Previous studies have isolated and analyzed the transcriptomes of these cell types, primarily from mouse (La Manno et al., 2018; Pandey et al., 2018; Saunders et al., 2018; Vanlandewijck et al., 2018; Zeisel et al., 2015), but also human brain (Darmanis et al., 2015; Reddy et al., 2017; Zhang et al., 2016). The most common primary brain malignancy is glioma, of which 60%–70% are glioblastoma (GBM), an incurable disease with a short median survival (Lim et al., 2018; Osuka and Van Meir, 2017). Bulk RNA-seq can identify GBM molecular signatures (Jovčevska, 2018) but does not resolve disease-associated cell-type-specific changes. Few studies have analyzed GBM on a cell-by-cell basis, and profiling of relatively low abundance cells, such as ECs, remains challenging due to the small number of tumors analyzed (Darmanis et al., 2017; Patel et al., 2014; Yuan et al., 2018).

Previously, we identified the core *body-wide* EC-enriched transcriptome from unfractionated RNA-seq from 32 different organs (Butler et al., 2016), using an analysis based on the relative proportion of ECs across organs. Here, we analyzed

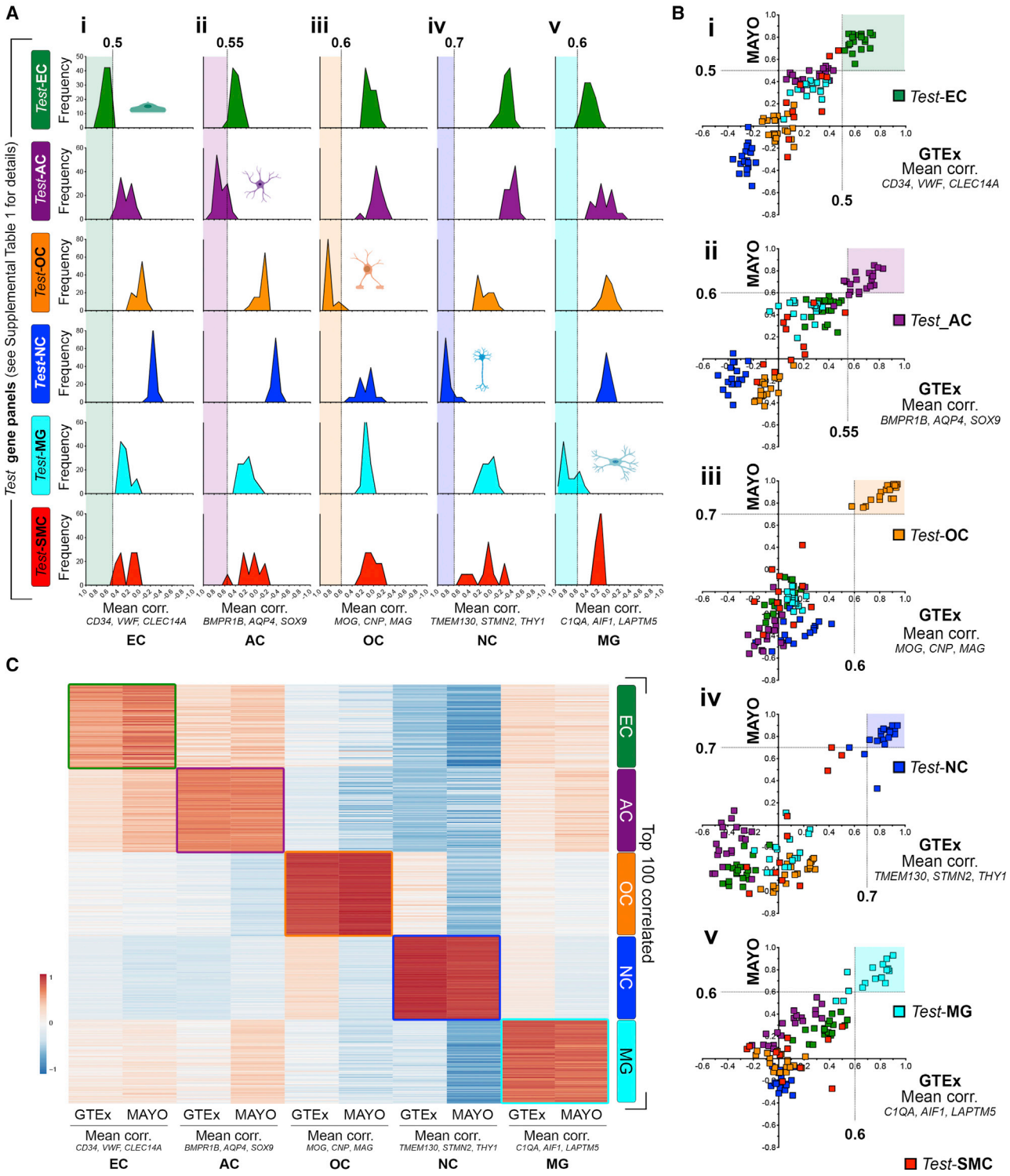


Figure 1. Ref. Transcript-Based Analysis Identifies Cell-Type-Enriched Genes from Unfractionated Human Cortex Data

(A) Cortex RNA-seq data from GTEx (n = 158) was used to calculate corr. coefficients between ref. transcripts for: (i) endothelial cells (ECs) (*CD34*, *CLEC14A*, *VWF*); (ii) astrocytes (ACs) (*BMPR1B*, *AQP4*, *SOX9*); (iii) oligodendrocytes (OCs) (*MOG*, *CNP*, *MAG*); (iv) neurons (NCs) (*TMEM130*, *STMN2*, *THY1*); and (v) microglial cells (MGs) (*C1QA*, *AIF1*, *LAPTM5*) and "*test*-panels"—genes categorized in the literature as enriched in: brain ECs (*test*-ECs), ACs (*test*-ACs), OCs (*test*-OCs), NCs (*test*-NCs), MGs (*test*-MGs), or smooth muscle cells (*test*-SMCs). Frequency distribution of coefficient values are plotted, and shaded areas indicate thresholds selected.

(legend continued on next page)

unfractionated RNA-seq of a *single* organ, the human brain, to identify EC, AC, OC, NC, and MG cell-type-enriched genes. We used RNA-seq from 516 lower-grade gliomas (LGGs) and 401 GBMs to decipher global EC-compartment modifications and predict tumor-specific EC biomarkers. A searchable web-based interface is provided for an exploration of the generated datasets at https://cell-enrichment.shinyapps.io/brain. Our approach allows direct comparison of cell-type-enriched transcriptome profiles between normal and diseased tissue, adding another method to the toolkit of existing RNA-seq deconvolution approaches, such as those that can be used to establish cell-type enriched profiles in normal tissue and changes in the proportion of cell types in diseased tissue (Kelley et al., 2018; Newman et al., 2015; Yoshihara et al., 2013). In addition, our approach, which generates results consistent with isolated cell and single-cell RNA-seq, does not require high-level bioinformatics expertise or complex modeling, making it an accessible tool for the wider research community.

RESULTS

Cell-Type Reference Transcripts Correlate across Unfractionated Cortex RNA-Seq Data

Unfractionated normal human brain RNA-seq was sourced from (1) GTEx V7 (cortex, n = 158) (GTEx Consortium, 2015) and (2) Accelerating Medicines Partnership-Alzheimer's Disease (AMP-AD) knowledge portal (MAYO RNA-seq study) (controls, n = 80) (Allen et al., 2016). We selected 3 "reference" (ref.) gene transcripts, which encode for cell-type-specific markers: (A) ECs (*CD34*, *CLEC14A*, *VWF*); (B) ACs (*BMPR1B*, *AQP4*, *SOX9*); (C) OCs (*MOG*, *CNP*, *MAG*); (D) NCs (*TMEM130*, *STMN2*, *THY1*); and (E) MGs (C1QA, AIF1, LAPTM5) (Butler et al., 2016; Cahoy et al., 2008; Darmanis et al., 2015; Fonseca et al., 2017; Ito et al., 1998; Pfeiffer et al., 1993; Sun et al., 2017; Zhang et al., 2014). We analyzed GTEx cortex RNA-seq data to calculate correlation (corr.) coefficient values between these cell-type ref. transcripts across samples. Consistent with co-expression, ref. transcripts within each cell-type group correlated with each other: ECs (*CD34*, *CLEC14A*, *VWF*) mean corr., 0.66 $p > 0.0001$; ACs (*BMPR1B*, *AQP4*, *SOX9*) mean corr., 0.81 $p > 0.0001$; OCs (*MOG*, *CNP*, *MAG*) mean corr., 0.92 $p > 0.0001$; NCs (*TMEM130*, *STMN2*, *THY1*) mean corr., 0.89 $p > 0.0001$; MGs (C1QA, AIF1, LAPTM5) mean corr., 0.91 $p > 0.0001$; while ref. transcripts between cell-type groups did not (Figure S1).

Ref. Transcript Analysis Can Resolve Cell-Type Genes from Cortex RNA-Seq Data

We analyzed the GTEx cortex RNA-seq data to produce corr. values between each ref. transcript and the other >20,000 mapped protein-coding genes. A high mean corr. with cell-type ref. transcripts should indicate enrichment of the gene(s) in question in that cell type. To test sensitivity and specificity of the method, we compared corr. between the ref. transcripts and 6 "*test*-panels"—genes categorized in the literature as enriched in (1) brain ECs (*test*-ECs), (2) ACs (*test*-ACs), (3) OCs (*test*-OCs), (4) NCs (*test*-NCs), (5) MGs (*test*-MGs), and (6) smooth muscle cells (*test*-SMCs) (Cahoy et al., 2008; Chu and Peters, 2008; Conley, 2001; Darmanis et al., 2015; Dreiza et al., 2010; He et al., 2016; Long et al., 2009; Miwa et al., 1991; Rensen et al., 2007; Wang et al., 2003; Yamawaki et al., 2001; Zhang et al., 2014) (Table S1, tabs 1–5: column A). Each set of ref. transcripts correlated most highly with genes in the corresponding *test*-panel (Figure 1A), with no overlap with genes from any other *test*-panel (Table S1, tabs 1–5: column A). We performed an equivalent *test*-panel analysis using RNA-seq data from the AMP-AD knowledge portal (MAYO RNA-seq study). Corr. between the *test*-panels and the EC, AC, OC, NC and MG ref. transcripts in the MAYO (Figures 1Bi–1Bv, respectively), were comparable to the GTEx (Table S1, tabs 1–5). For *test*-EC, AC, OC, NC, and MG genes, corr. values versus the corresponding ref. transcripts were high in both GTEx and MAYO data; the resultant cluster lying in the upper-right quadrant of the comparative plot (Figure 1B); the shaded box and dashed lines indicate the selected threshold requirement for classification as cell-type enriched. We observed comparable results in the *test*-panel analysis when we used a larger panel of ref. transcripts for each cell type (6 or 12, instead of 3) (see Figure S2). *Test*-SMC genes did not highly correlate with the EC, AC, OC, NC, or MG ref. transcripts (Figures 1Ai–1Av). However, to further verify that SMC-enriched transcripts would not be incorrectly classified as EC-, AC-, OC-, NC-, or MG-enriched, we selected 3 known SMC transcripts (*FHL5*, *ACTA2*, *ACTG2*) (Vanlandewijck et al., 2018) and calculated their corr. with the *test*-panels (GTEx) (Table S1, tab 6). All *test*-transcripts correlated more highly with the corresponding ref. transcripts, than with SMC ref. transcripts (Figures S3Ai–S3Av). Higher corr. values of the SMC ref. transcripts with the *test*-EC panel (versus *test*-AC, OC, NC, or MG) indicated that SMC genes were most likely to be incorrectly classified as EC-enriched (Figure S3Ai), rather than AC-, OC-, NC-, or MG-enriched. Transcripts identified as EC-, AC-, OC-, NC-, or MG-enriched were excluded if the mean corr. with the ref. transcripts < mean corr. versus SMC transcripts (GTEx versus GTEx). As predicted, most exclusions were from the EC-enriched list (Figure S4A).

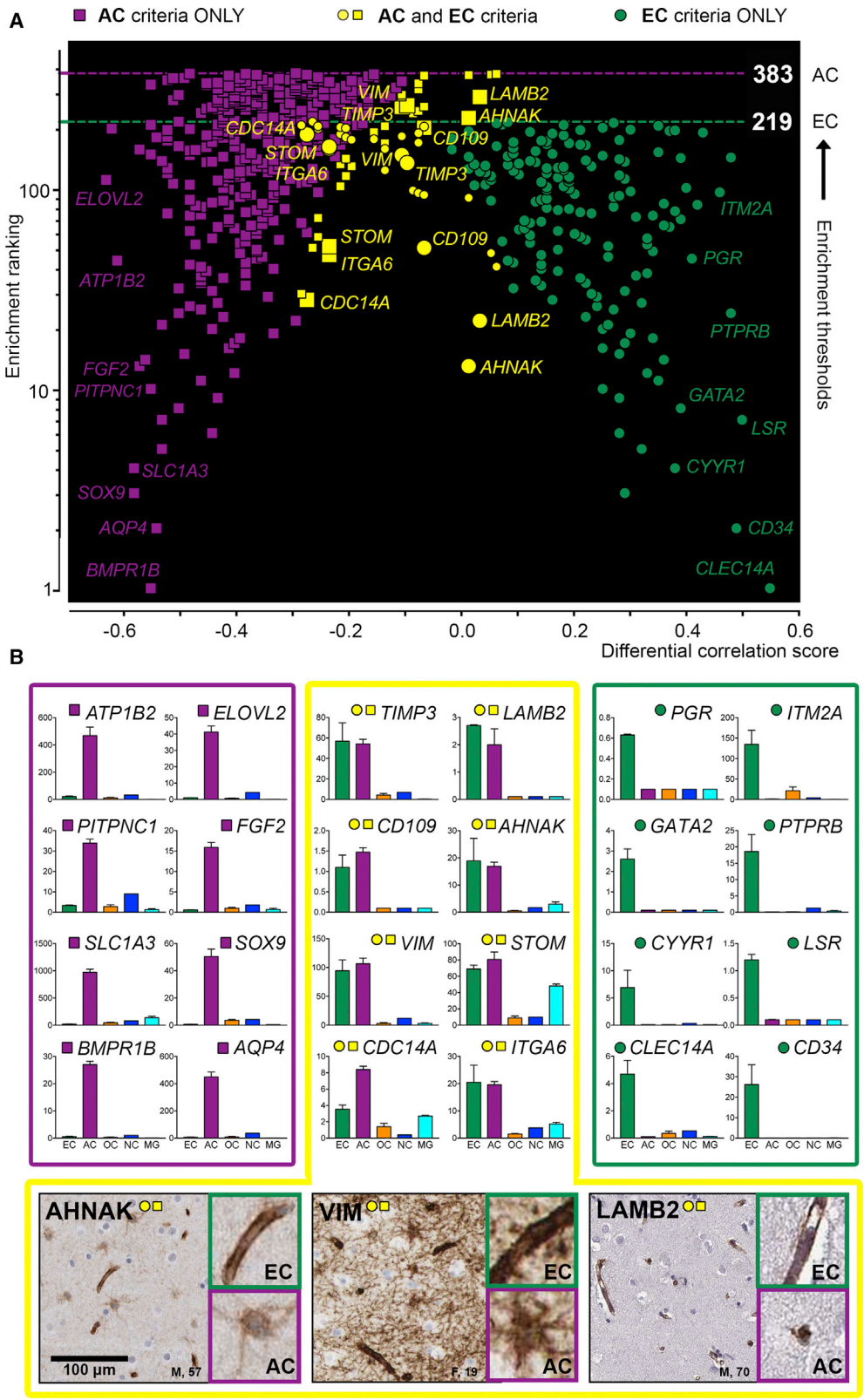
ECs and ACs Have a Panel of Dual-Enriched Transcripts

We compared the relationships between transcripts in the cell-type-enriched lists. To compare groups, e.g., EC- versus AC-enriched, the following values were calculated for each transcript featuring in either list: (1) the difference between the mean corr. for both sets of corresponding ref. transcripts,

(B) Equivalent *test*-panel analysis was performed on cortex RNA-seq data from the AMP-AD knowledge portal MAYO RNA-seq study (controls, n = 80) and means compared to the GTEX results for (i) EC, (ii) AC, (iii) OC, (iv) NC, and (v) MG ref. transcripts. The dotted lines indicate the dual-thresholding criteria applied for classification as cell-type enriched.

(C) Heatmap plot of corr. values of the top 150 genes identified as EC-, AC-, OC-, NC-, or MG-enriched versus all the ref. transcript sets (adapted from ClustVis, https://biit.cs.ut.ee/clustvis/; values subject to row centering and unit variance scaling).

See also Tables S1, S2, S3, S4, S5, and S6.



(legend on next page)

i.e., ECs (*CD34*, *CLEC14A*, *VWF*) and ACs (*BMPR1B*, *AQP4*, *SOX9*) (the "differential corr. score"), and (2) the "enrichment ranking," based on the corr. value with each set of ref. transcripts (highest corr.: rank = 1) (Figures S3Bi–S3Biv). Threshold lines indicate the rank number below which transcripts were classified as cell-type enriched. EC and AC-enriched transcripts and (to a lesser extent) EC- and MG-enriched transcripts "crossed" on the plot (31 and 5 common genes, respectively) (Figures S3Bi and S3Biv, yellow points), while EC transcripts were well separated from the enriched gene profiles of OCs and NCs (Figures S3Bii and S3Biii). We further investigated transcripts that fulfilled the criteria to be classified as EC *and* AC enriched (Figure 2A, yellow points [plot as for Figure S3B, but transcripts ranked below enriched classification threshold excluded]) (Table S3, tab 5). We sourced data from a previously published study where human brain was sorted into cell populations prior to RNA-seq (Zhang et al., 2016); for transcripts we identified as (1) AC-enriched *only*, (2) AC *and* EC-enriched, or (3) EC-enriched *only* (Figure 2A: purple, yellow, and green, respectively); expression profiles were consistent with our classifications (Figure 2B). Based on these analyses, transcripts classified as enriched in more than one cell type were excluded from the final list(s) if they fulfilled enrichment criteria for both, or if they had higher corr. values with the other (non-corresponding) ref. transcripts.

Identification of EC-, AC-, OC-, NC-, or MG-Enriched Genes

Following application of selection criteria (ref. transcript corr. in GTEx, replication in MAYO, false discovery rate [FDR] threshold, SMC- and dual-enriched transcript exclusion) (Figure S4), 166 genes were classified as EC-enriched, 351 were AC-enriched, 397 were OC-enriched, 2,015 were NC-enriched, and 205 were MG-enriched (Tables S2–S6 [enriched: tab 1; all values: tab 2]). Expression of selected uncharacterized or lesser-known transcripts was confirmed by immunohistochemistry [IHC] (Figure S5). A heatmap plot of the corr. values of the top 100 genes identified as EC-, AC-, OC-, NC-, or MG-enriched versus the EC, AC, OC, NC, and MG ref. genes, respectively, revealed the resolution of cell-type expression profiles in the GTEx and MAYO data (Figure 1C). The top 15 most highly enriched genes for each category contained known cell-enriched transcripts (Figure 3, marked in bold). Gene Ontology (GO) and Reactome ("pathway") analysis (Ashburner et al., 2000) was performed on the final list of EC-, AC-, OC-, NC-, and MG-enriched transcripts. The most significant biological process GO groups in the EC-enriched list included "*vasculature development*" and "*angiogenesis*" ($p < 3.8 \times 10^{\wedge}17$) and reactome pathways included "*haemostasis*" and "*NOSTRIN mediated eNOS trafficking*" ($p < 3.9 \times 10^{\wedge}5$) (Table S2, tab 4). The most significant biological process GO groups in the AC-enriched list included "*regulation of signaling*" and "*small molecule catabolic process*" ($p < 1.4 \times 10^{\wedge}10$) and reactome pathways included "*metabolism*" and "*transport of small molecules*" ($p < 1.3 \times 10^{\wedge}5$) (Table S3, tab 4). The most significant biological process GO group in the OC-enriched list was "*myelination*" ($p < 1.6 \times 10^{\wedge}17$) and the single reactome pathway identified was "*transport of small molecules*" ($p < 7.8 \times 10^{\wedge}7$) (Table S4, tab 4). The most significant biological process GO groups in the NC-enriched list included "*nervous system development*" and "*trans-synaptic signaling*" ($p < 2.2 \times 10^{\wedge}32$) and reactome pathways included "*neuronal system*" and "*neurotransmitter receptors and postsynaptic signal transmission*" ($p < 9.5 \times 10^{\wedge}17$) (Table S5, tab 4). The most significant biological process GO groups in the MG-enriched list included "*Immune system process*" and "*defense response*" ($p < 2.5 \times 10^{\wedge}45$) and reactome pathways included "*immune system*" and "*cytokine signaling in immune system*" ($p < 4.4 \times 10^{\wedge}18$) (Table S6, tab 4). Summary plots were generated using Reduce and Visualize Gene Ontology (REViGO) (Supek et al., 2011) (Figure 3).

Comparison of Results with Single-Cell or Fractionated Brain Cell-Type RNA-Seq Data

We compared the top 100 most enriched EC, AC, OC, NC, and MG transcripts from our unfractionated (UF) RNA-seq analysis, to existing single-cell RNA sequencing (scRNA-seq) of human brain (H1) (Darmanis et al., 2015) and RNA-seq of fractionated fluorescence-activated cell sorting (FACS)-sorted or immunopanned isolated cell populations from human (H2) (Zhang et al., 2016) or mouse (M1) brain (Zhang et al., 2014). The top 100 most enriched genes we identified in the EC, AC, OC, NC, and MG categories were significantly enriched in the corresponding human single-cell-type sequencing analysis (H1): total with reproducibility-optimized test statistic (ROTS) score (see STAR Methods for ROTS definition) > 2 versus other cell types ECs 72/100, ACs 90/100, OCs 86/100, NCs 97/100 and MGs 73/100 (Figures 4Ai and 4Aii) and the human fractionated cell type analysis (H2): fold expression > 2 versus other cell types: ECs, 72/100; ACs, 87/100; OCs, 94/100; NCs, 83/100; and MGs, 99/100 (all $p > 0.001$ versus other cell types) (Figures 4Bi and 4Bii). Agreement was less consistent in the mouse

Figure 2. Ref. Transcript-Based Analysis Identifies Dual Cell-Type-Enriched Genes from Unfractionated Human Cortex Data

Cortex RNA-seq data from GTEx (n = 158) was used to calculate corr. coefficients between ref. transcripts for ECs (*CD34*, *CLEC14A*, *VWF*) or ACs (*BMPR1B*, *AQP4*, *SOX9*).

(A) For transcripts fulfilling criteria for classification as EC- (circles) or AC-enriched (squares), the "*differential* corr. *score*" (difference between mean corr. with EC and AC Ref. transcripts) was plotted versus "enrichment ranking" (position in each respective enriched list, highest corr. = ranking 1). Threshold lines denote ranking below which transcripts were classified as EC- or AC-enriched. The yellow symbols represent transcripts classified as *both* EC- and AC-enriched (circular *and* square symbol, on the same x axis dimension).

(B) mRNA expression data from 5 isolated brain cell types: ECs (green, n = 2 donors), ACs (purple, n = 12 donors), OCs (orange, n = 5 donors), NCs (blue, n = 1 donors), and MGs (cyan, n = 3 donors), isolated from human brain prior to RNA-seq (Zhang et al., 2016), for 351 genes we classified as AC-enriched (purple box), 31 dual-enriched in ACs and ECs (yellow box), and 166 EC-enriched (green). The bars represent mean ± SD. Protein profiling of human cortex tissue for expression of *AHNAK*, *VIM*, and *LAMB2* in both ACs and ECs. Scale bar, 100 μm.

See also Table S3, tab 5. Letter and number on bottom right of image denote tissue sex (F, female; M, male) and age (years).

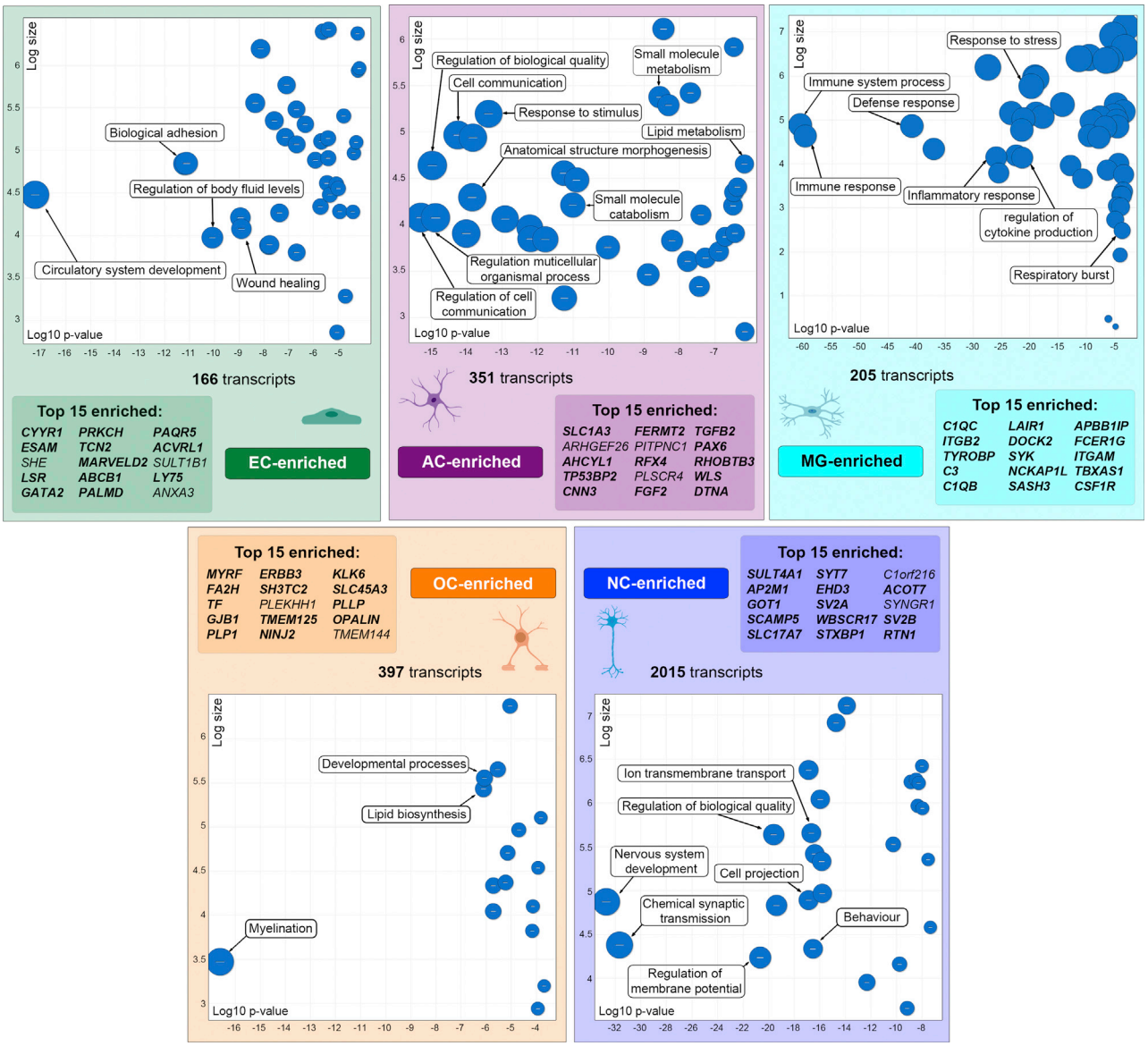


Figure 3. Transcripts Classified as EC-, AC-, OC-, NC-, or MG-Enriched in Human Brain Cortex

Cortex RNA-seq data from GTEx (n = 158), and the AMP-AD knowledge portal MAYO RNA-seq study (n = 80) was used to identify EC-, AC-, OC-, NC-, and MG-enriched transcripts. Total enriched transcript number, top 15 most enriched (bold denotes previously described), and a summary of Gene Ontology (GO) enrichment for all enriched genes in each cell type (generated using REViGO).

See also Tables S2, S3, S4, S5, and S6 and Figure S4.

fractionated cell type analysis (M1): fold expression > 2 versus other cell types: ECs, 68/100; ACs, 53/100; OCs, 53/100; NCs, 64/100; MGs, 87/100 ($p > 0.0001$ versus other cell types) (Figures 4Ci and 4Cii); consistent with differences in gene expression and cell type specificity between human and murine brain. Overall, ~50% of the top 100 EC-, AC-, OC-, NC-, and MG-enriched genes were categorized as enriched in the same cell type in *all* 3 (H1, H2 and M1) datasets (Figure 4D), and ≥98% were classified as enriched in *at least* one other dataset. Thus, our results were consistent with fractionated tissue or scRNA-seq-based cell-type transcript classification.

Comparison of Ref. Transcript-Based and Weighted Corr. Network Analysis

A disadvantage of our analysis is its reliance on a pre-selected panel of cell-type-enriched "ref." transcripts and, thus, its being subject to a possible input bias. To determine the extent of this limitation, we analyzed the same datasets using an alternative "unbiased" approach—weighted corr. network analysis ("WGCNA") (Langfelder and Horvath, 2008); where corrs. were generated between each transcript and all others, and transcripts with similar profiles were clustered together. Analysis of the GTEx and MAYO data produced a total of 37 and 50

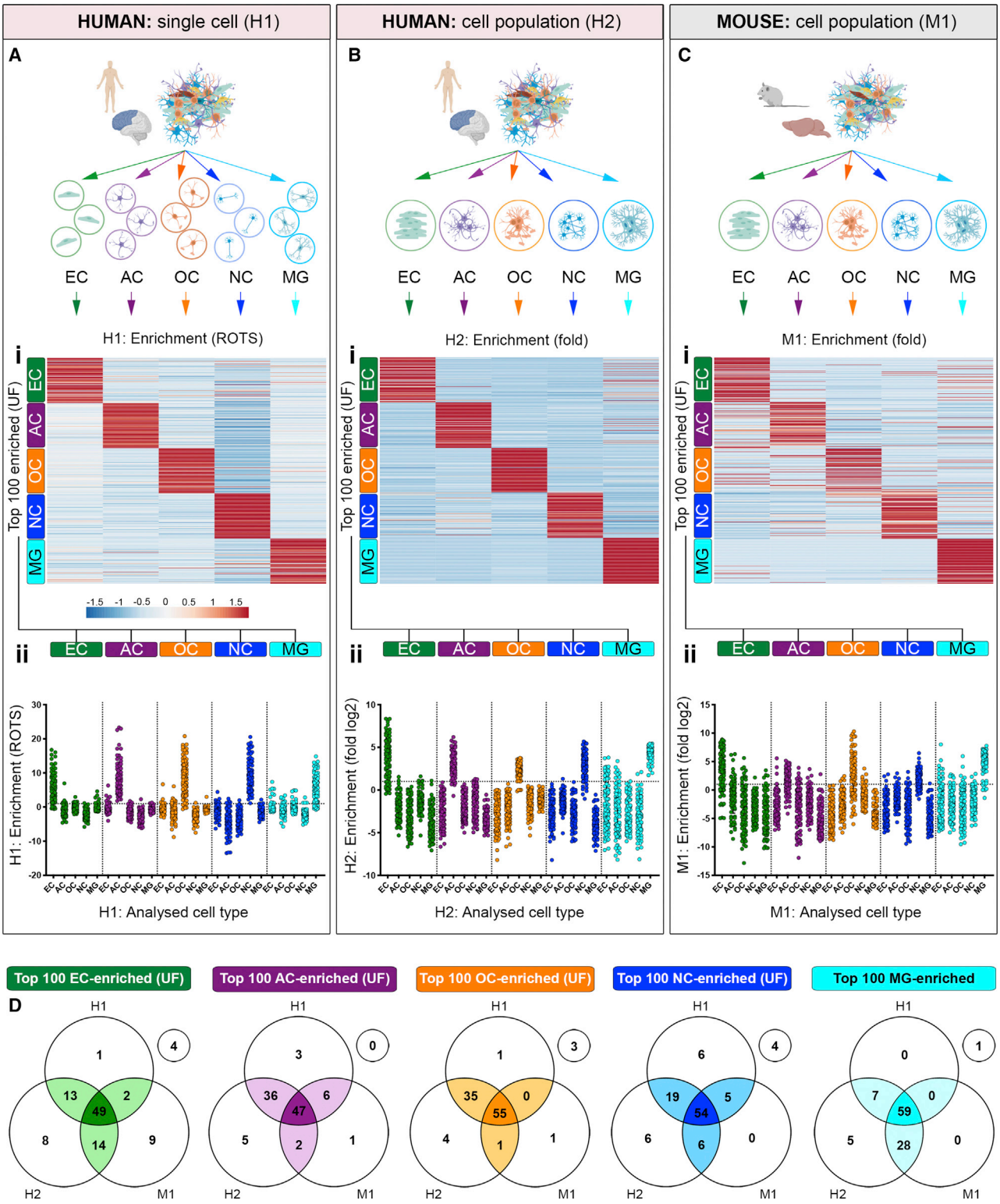


Figure 4. Comparison of Transcripts Classified as EC-, AC-, OC-, NC-, or MG-Enriched with Existing Single-Cell or Fractionated Brain Cell-Type RNA-Seq Data

(A–C) Data from (A) single-cell sequencing of human brain (Darmanis et al., 2015) (H1) or RNA-seq of cell populations from (B) human (H2) (Zhang et al., 2016) or (C) mouse (M1) brain (Zhang et al., 2014) were downloaded. Transcript "enrichment scores" in EC, AC, OC, NC, and MG populations in H1 ("ROTS score"), H2, and

(legend continued on next page)

independent clusters, respectively (Table S7, tab 1, annotated with arbitrary numbers). *Test*-OC and *test*-EC panels were used to identify clusters representing these cell types (Table S7, tab 2) (Figures S6A and S6B, green/orange boxes), and GO enrichment analysis of these clusters were consistent with OCs or ECs identify, respectively (Figures S6C and S6D) (Table S7, tab 4). The majority of genes classified as OC- or EC-enriched by the ref. transcript analysis were also classified as such by WGCNA (Figures S6Ei and S6Fii) (337/397 [85%] and 106/166 [64%], respectively). Of those not replicated, most were classified as OC- or EC-enriched in either the GTEx *or* MAYO WGCNA, but not *both*, and thus were excluded (Figures S6Ei and S6Fi) (Table S7, tab 5, columns A–D and L–O). However, IHC confirmed that many of the genes "excluded" from the EC category were indeed EC-enriched (Figure S6I) and, thus, were likely false negatives in the WGCNA. 337/446 [76%] of genes classified as OC-enriched by the WGCNA were also identified by the ref. transcript analysis (Figure S6Eii). Of those not replicated, 67/109 (61%) were classified as OC-enriched in either the GTEx *or* MAYO ref. transcript analysis, but not in *both* (Figure S6Eii) (Table S7, tab 5, columns F–I). In most cases, values narrowly failed to meet the required threshold in the "negative" dataset (mean corr.: GTEx, 0.57 ± 0.07; MAYO, 0.75 ± 0.08) (threshold required: ≥0.60 in GTEx, ≥0.70 in MAYO). The remaining 42/109 (39%) of those not replicated failed to reach the threshold for OC-enriched classification in either the GTEx or the MAYO ref. transcript analyses, but by a narrow margin (mean corr.: GTEx, 0.53 ± 0.06; MAYO, 0.62 ± 0.07). The WGCNA identified more EC-enriched genes than the ref. transcript (252 versus 166), and 106/252 (42%) of genes classified as EC-enriched by the WGCNA were also identified by the ref. transcript analysis (Figure S6Fii) (Table S7, tab 5, columns Q–T). Of those not replicated, 23/146 (16%) reached the required corr. threshold for classification as EC-enriched in the ref. transcript analysis, but were excluded due to high corr. values with MG-enriched transcripts (n = 9) or dual enrichment with ACs (n = 14), while 29/146 (20%) had corrs. with EC ref. transcripts <0.05 below the designated threshold for classification (mean corr.: GTEx, 0.49 ± 0.04; min., 0.45; mean corr.: MAYO ≥ 0.59 ± 0.08; min, 0.46).

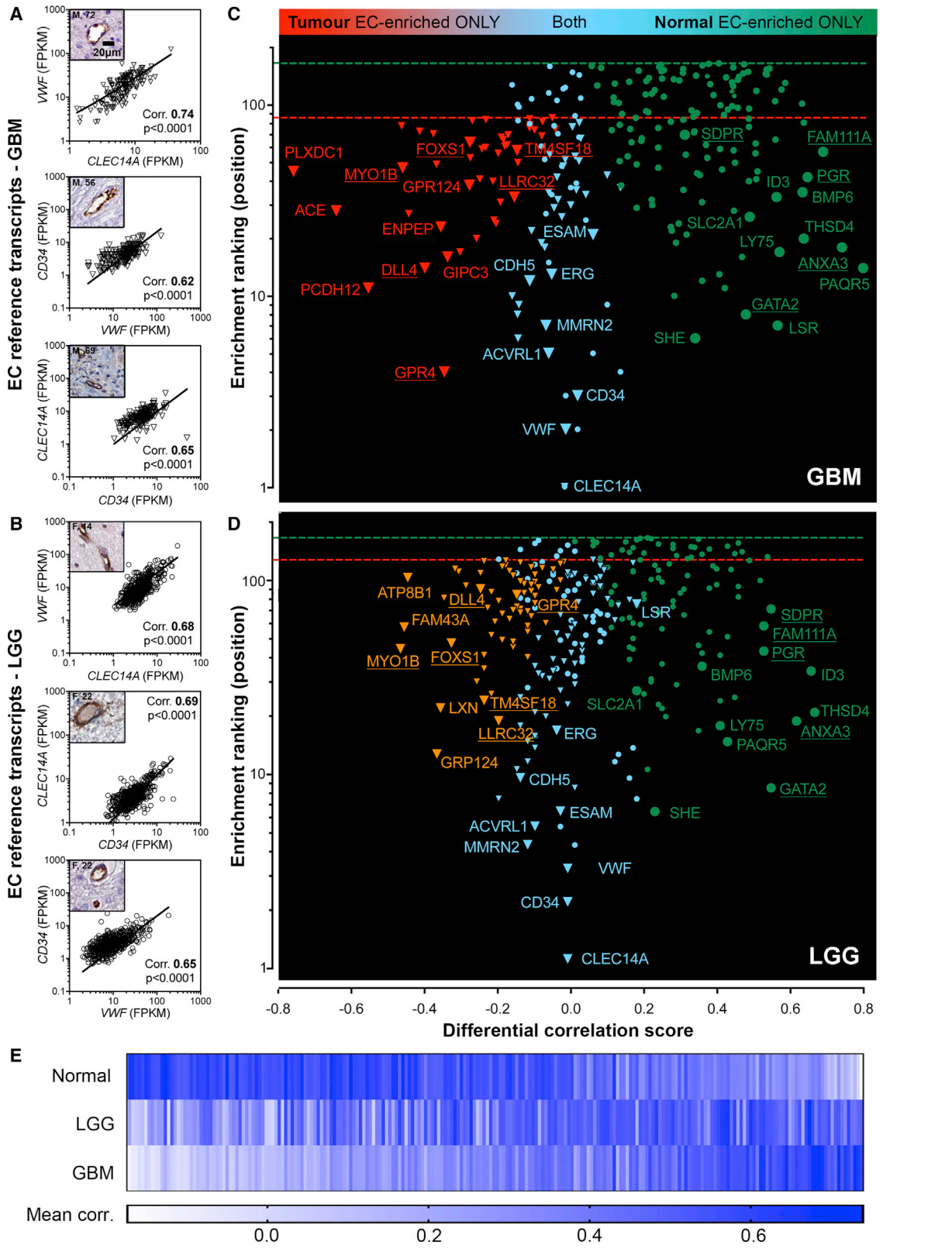
We compared EC- and OC-enriched lists, generated using each method, to brain cortex scRNA-seq (Darmanis et al., 2015) (H1) and isolated brain cell-type RNA-seq (Zhang et al., 2016) (H2) (as in Figure 4). The majority of the OC-enriched genes identified by ref. transcripts or WGCNA, were correspondingly enriched in both the (H1) (ref. transcript, 82%, versus WGCNA, 76%) (Figure S6Gi) and (H2) (ref. transcript, 80%, versus WGCNA, 78%) datasets (Figure S6Gii). Similarly, the majority of EC-enriched genes identified by ref. transcripts or WGCNA, were correspondingly enriched in both the (H1) dataset (ref. transcript, 76%, versus WGCNA, 72%) (Figure S6Hi) and the (H2) dataset (ref. transcript, 83%, versus WGCNA, 67%) datasets (Figure S6Hii). Overall, fewer genes were classified as OC- or EC-enriched by the ref. transcript analysis (OCs, 379 versus 446; and ECs, 166 versus 252), but results were better supported by data from (H1) and (H2), compared to the WGCNA.

Ref. Transcript Analysis Can Be Used to Profile Disease-Associated Modifications

Our data from the analysis of normal cortex were used as a starting point for the prediction of cellular changes in malignant disease. We used ref. transcript analysis to study disease-associated changes in the vasculature of GBM and LGG, where cross talk between ECs and tumor cells promotes proliferation and subsequent disease progression (Yan et al., 2017). Expression tables for GBM (n = 401) and LGG (n = 516) samples from The Cancer Genome Atlas (TCGA) were sourced from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/). As in normal brain, EC genes (*CD34*, *CLEC14A*, *VWF*) correlated highly in GBM (Figure 5A) and LGG (Figure 5B) (corr. mean, GBM, 0.67; LGG, 0.74), with EC restricted expression (Figures 5A and 5B, upper left images), indicating suitability as EC ref. transcripts in these cohorts. In principle, ref. transcripts representing other cell types (e.g., ACs or OCs) could be selected to analyze changes in these cell types in malignancy, but validation of their maintained cell-type-restricted expression in disease, in the absence of the distinctive morphological features of ECs (thin, relatively elongated nucleus, frequently clearly adjacent to a vessel lumen), may be more challenging and require additional validation. GBM and LGG datasets were analyzed as for normal brain (GTEx) (Table S8, tab 1) to identify EC-enriched transcripts. The most significant biological process GO groups in the GBM EC-enriched list (corr. > 0.50) were related to EC function, including "vasculature development" and "blood vessel morphogenesis" (p values < 1.7 × 10^25) (Table S8, tab 5). There are were no GO groups related to immune cell populations, hypoxia, or necrosis, indicating that transcript changes that could be linked to vascular density per se (rather than direct expression by ECs), are not incorrectly identified as EC-enriched. WGCNA of GBM RNA-seq produced 2 related clusters that contained EC markers (n = 104 and n = 35) (Figure S7A). These clusters contained 78/85 (92%) of genes classified as GBM EC-enriched by the ref. transcript analysis (Figure S7B) (Table S8, tab 2). We compared GBM and normal brain, using the differential corr. score and enrichment rank of all transcripts categorized as EC-enriched in GBM *or* normal brain (Figure 5C). 33 transcripts were classified as EC-enriched in normal brain *and* GBM (Figure 5C, blue data points) (data contained in Table S8, tab 3). Of these, we previously identified 29/33 (88%) as body-wide core EC-enriched (Butler et al., 2016), including *ENG*, *TIE1*, *ESAM*, *CDH5*, and *ERG* (which encodes for a key transcription factor in the maintenance of EC identity) (Shah et al., 2016); thus, this group likely represents genes indispensable for vascular function. A panel of genes were classified as

M2 (fold enrichment versus other cell types) were compiled for the top 100 transcripts identified as EC-, AC-, OC-, NC-, or MG-enriched from our unfractionated (UF) RNA-seq analysis and presented as follows: (i) heatmaps and (ii) grouped plots. Horizontal line: threshold for "enrichment."

(D) Venn diagrams showing number of top 100 transcripts we identified as EC-, AC-, OC-, NC-, or MG-enriched that were also identified as enriched in the corresponding cell type in H1, H2, and M1 datasets (free circles, unfractionated only). Heatmaps adapted from images generated by ClustVis (https://biit.cs.ut.ee/clustvis/).



(legend on next page)

EC-enriched in GBM, but not in normal brain, including *PLXDC1*, *ACE*, *PCDH12*, and vice versa, including *PAQR5*, *ANXA3*, and *FAM111A* (Figure 5C, red and green data points, respectively) (Table S8, tab 3). LGG had a higher proportion of EC-enriched genes in common with normal brain, compared to GBM (Figure 5D, more blue data points versus Figure 5C), and correspondingly, fewer genes were modified in LGG versus normal, compared to GBM versus normal. Of those EC-enriched genes that were modified in LGG, parallels were observed with GBM, e.g., *MYO1B*, *DLL4*, *FOXS1*, *GPR4*, *LLRC32*, and *TM4SF18* "gained" EC enrichment, and *SDPR*, *FAM111A*, *PGR*, *ANXA3*, and *GATA2* "lost" EC enrichment, in both tumor grades (Figures 5C and 5D, underlined). A heatmap plot of the raw corr. values, between the EC ref. transcripts and those classified as EC-enriched in normal *or* LGG *or* GBM (Figure 5E), was consistent with the hypothesis that LGG represents an "intermediate state" between normal and GBM.

Further Resolution of GBM-Associated EC-Enriched Transcriptome Modifications Using mRNA Expression Data in Combination with Protein Profiling

We selected a panel of EC genes identified as modified in GBM versus normal brain to further characterize, by combining expression data (normal, LGG, and GBM) from the OASIS portal (Fernandez-Banet et al., 2016) with protein profiling, which we performed as part of the Human Protein Atlas (HPA) Pathology Atlas (Uhlén et al., 2017). ACE (angiotensin converting enzyme) was classified as EC-enriched in GBM, but not normal brain (mean corr., 0.63, GBM, versus −0.02, normal); correspondingly, *ACE* mRNA was higher in GBM versus normal (mean transcripts per million [TPM], 16.0, GBM, versus 4.7, normal; $p < 0.0001$) (Figure 6Ai, exp. [TPM]). The "degree" of *ACE* EC-enrichment and expression was intermediate in LGG, between normal and GBM (Figure 6Ai), indicating changes occurred in LGG, but they were not as pronounced/consistent as those in the higher-grade GBM. ACE protein was not detected in normal brain (Figure 6Ai, IHC 1 and 2) (blue arrows indicate negative vessels), was variable in LGG ECs, and was highly abundant in GBM ECs (green arrows show positive staining). Similar patterns were observed for CD93 (receptor for the C1q complement factor) (Figure 6Aii) and ANGPT2 (angiopoietin 2) (Figure 6Aiii).

When transcripts were classified as EC-enriched in normal brain, but not GBM (Figure 5C, green data points), this could indicate that (1) expression was *lost* from GBM ECs, or (2) expression was *gained* outside the vasculature, i.e., in tumor tissue (illustrated in Figures 7A and 7B). *ANXA3* (Annexin 3A) was EC-enriched in normal brain, but not in GBM (mean corr., 0.66, normal, versus −0.09, GBM) and *ANXA3* mRNA was decreased in GBM versus normal (mean TPM, 1.6, GBM, versus 10.3, normal; $p < 0.0001$) (Figure 6Bi). IHC showed ANXA3 was expressed in normal ECs, was variable in LGG ECs, and was absent from GBM ECs. Similar patterns were observed for *PRX* (Periaxin) (Figure 6Bii) and *ABCB1* (ATP-binding cassette subfamily B member 1) (Figure 6Biii). *CAV2* (Caveolin 2) was also EC-enriched in normal brain, but not GBM (mean corr., 0.52, normal, versus 0.16, GBM) (Figure 6Ci); however, *CAV2* mRNA was *increased* in GBM versus normal (mean TPM, 20.9, GBM, versus 7.1, normal; $p < 0.0001$), and CAV2 was expressed outside the EC compartment (Figure 6Ci). Similar was observed for *A2M* (Figure 6Cii) and *HLA-B* (Figure 6Ciii). Generally, LGG represented an "intermediate state" between normal brain and GBM, the degree of similarity to either condition depending on the transcript.

We used our analysis to predict 10 GBM EC markers (Figure 7C), based on the following criteria: (1) they were *not* core EC-enriched genes across other vascular beds (Butler et al., 2016); (2) they had *low* expression levels in the normal brain (mean TPM < 5); and (3) they had a high differential EC-enriched corr. between GBM and normal (Table S9). The 10 predicted GBM-EC markers correlated with each other in the GBM RNA-seq (Figure 7D), but not the normal RNA-seq (mean corr., GBM, 0.61 ± 0.06, versus normal, 0.18 ± 0.05). Consistent with previous observations, LGG represented an intermediate between GBM and normal (mean corr., LGG, 0.46 ± 0.04) (corr. matrices, Table S8, tab 4). All 10 predicted GBM-EC markers were also in the EC-annotated groups in the WGCNA of the GBM RNA-seq (Figures 7E and S7), consistent with co-expression in GBM ECs. Thus, our method can be applied to perform a systems-based prediction of highly GBM-specific EC-markers that, following experimental validation, could have clinical applications for tumor grading, prognosis prediction, and therapeutic targeting.

DISCUSSION

Here, we have demonstrated how cell-type-enriched transcripts in the brain can be identified using bulk RNA-seq, and performed a systems-level exploration of EC changes associated with GBM. We provide data tables and a web interface (https://cell-enrichment.shinyapps.io/Brain/) for users to look up the specificity profiles of any mapped protein-coding gene in normal cortex and EC enrichment score in GBM or LGG.

Our approach has some advantages over other types of transcriptome analysis, such as single-cell sequencing (scRNA-seq). scRNA-seq typically involves the extraction and sequencing of thousands of individual cells (Darmanis et al., 2015; Villani et al., 2017; Pandey et al., 2018; Papalexi and Satija, 2018) to detect cell type specific genes and identify to rare cell subtypes

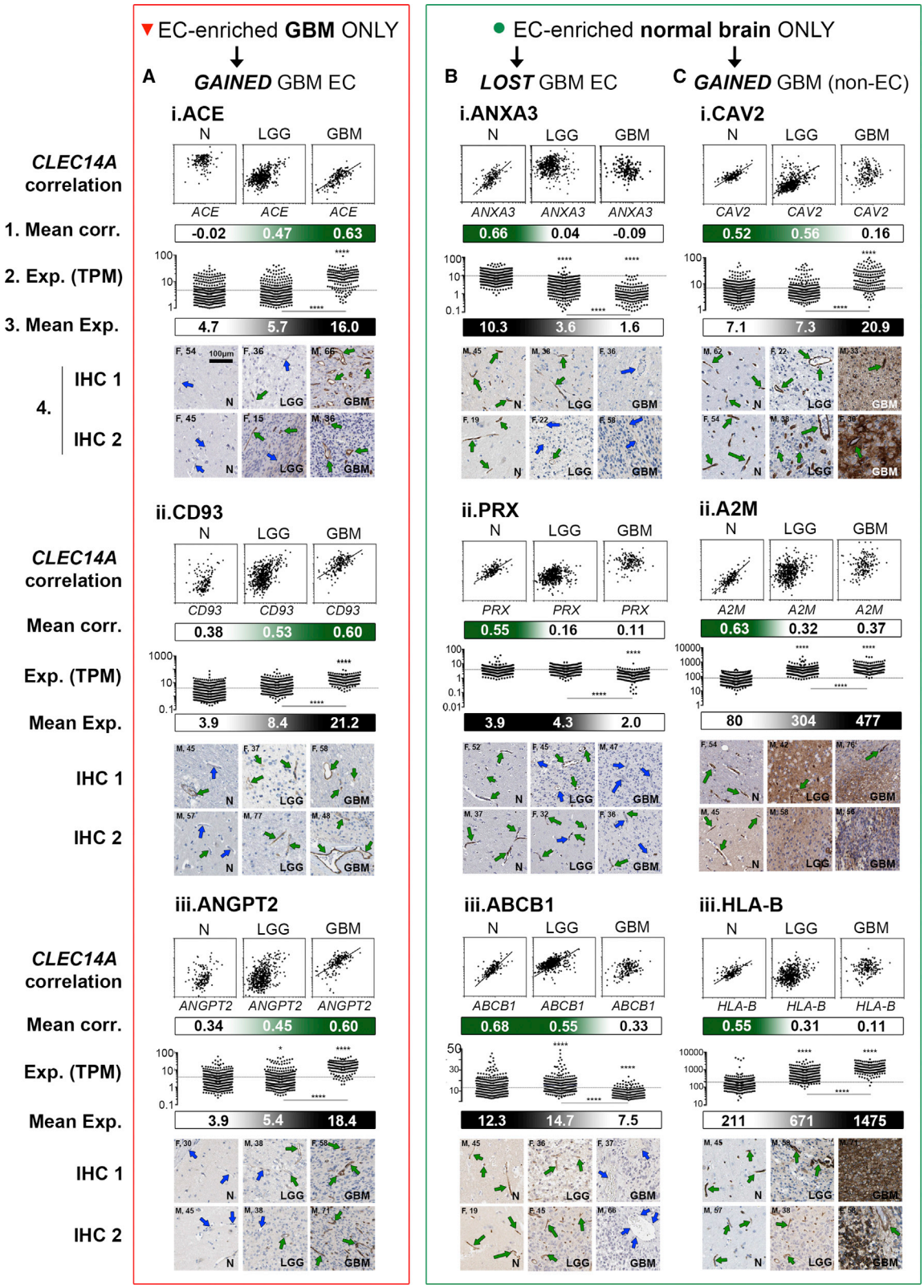
Figure 5. Ref. Transcript-Based Analysis Can Predict Global Modifications of EC-Enriched Transcriptome of Human Glioblastoma and Lower-Grade Glioma

RNA-seq data from unfractionated glioblastoma (GBM) (n = 401) and lower-grade glioma (LGG) (n = 516) was downloaded from The Cancer Genome Atlas (TCGA), and corr. coefficients were calculated between the EC ref. transcripts (*CD34*, *CLEC14A*, *VWF*) and all mapped protein-coding genes.

(A–D) Corrs. between *CD34*, *CLEC14A*, *VWF*, and IHC confirming EC restricted expression (label as for y axis) in (A) GBM and (B) LGG. Plots showing *differential corr. score* and *enrichment rank* of transcripts categorized as EC-enriched in (C) GBM and/or normal brain and (D) LGG and/or normal brain.

(E) Heatmap of corr. values between transcripts classified as EC-enriched in GBM and/or LGG and/or normal brain, and the EC ref. transcripts.

See also Table S8. Letter and number on top left of image denotes tissue sex (F, female; M, male) and age (years).



(legend on next page)

in heterogeneous tissue. In contrast, our analysis does not require removal of cells from the resident niche, circumventing modifications due to loss of environmental cues, or processing-associated activation (Chow and Gu, 2015; Urich et al., 2012; Wilhelm et al., 2016) while avoiding the introduction of technical variability (Beliakova-Bethell et al., 2014; Saliba et al., 2014). As sample collection and processing needs to be well controlled, many brain scRNA-seq studies use non-human tissue (e.g., Pandey et al., 2018; Rosenberg et al., 2018; Vanlandewijck et al., 2018). In contrast, our method can be used to analyze hundreds of human biological replicates concurrently. However, our analysis does have limitations; we applied strict thresholds to maximize confidence in our classifications and to only identify the most *highly* specific genes for each cell type. It was not feasible to identify an extensive panel of transcripts expressed at moderately higher levels in one cell type versus others, which is reflected by the low number of genes we annotated as cell-enriched, compared to isolated brain cell type or scRNA-seq studies (Darmanis et al., 2015; Vanlandewijck et al., 2018; Zhang et al., 2014). A further limitation is possible incorrect classification of transcripts, particularly between closely associated cell types, which could have comparable ratios across samples, such as ECs with pericytes or SMCs. We used test panels to exclude SMCs as a source of false-positive EC-enriched genes, but equivalent analysis for pericytes was challenging, due to a lack of specific markers (Armulik et al., 2011). Genes described as pericyte-enriched, e.g., PDGFRB, ACTA2, DES, MCAM (Smyth et al., 2018), VTN, and IFITM1 (He et al., 2016), were not classified as EC-enriched. ANPEP (CD13), a marker used to identify pericytes in mouse (Crouch and Doetsch, 2018), was classified as brain EC-enriched in our study; however, it was also classified as such by scRNA-seq (Darmanis et al., 2015) and is highly expressed in EC lines (Thul et al., 2017). Thus, while our data indicate that pericyte genes are not incorrectly classified as EC-enriched, we cannot definitively exclude the possibility that the EC-enriched list contains dual pericyte/EC-enriched transcripts.

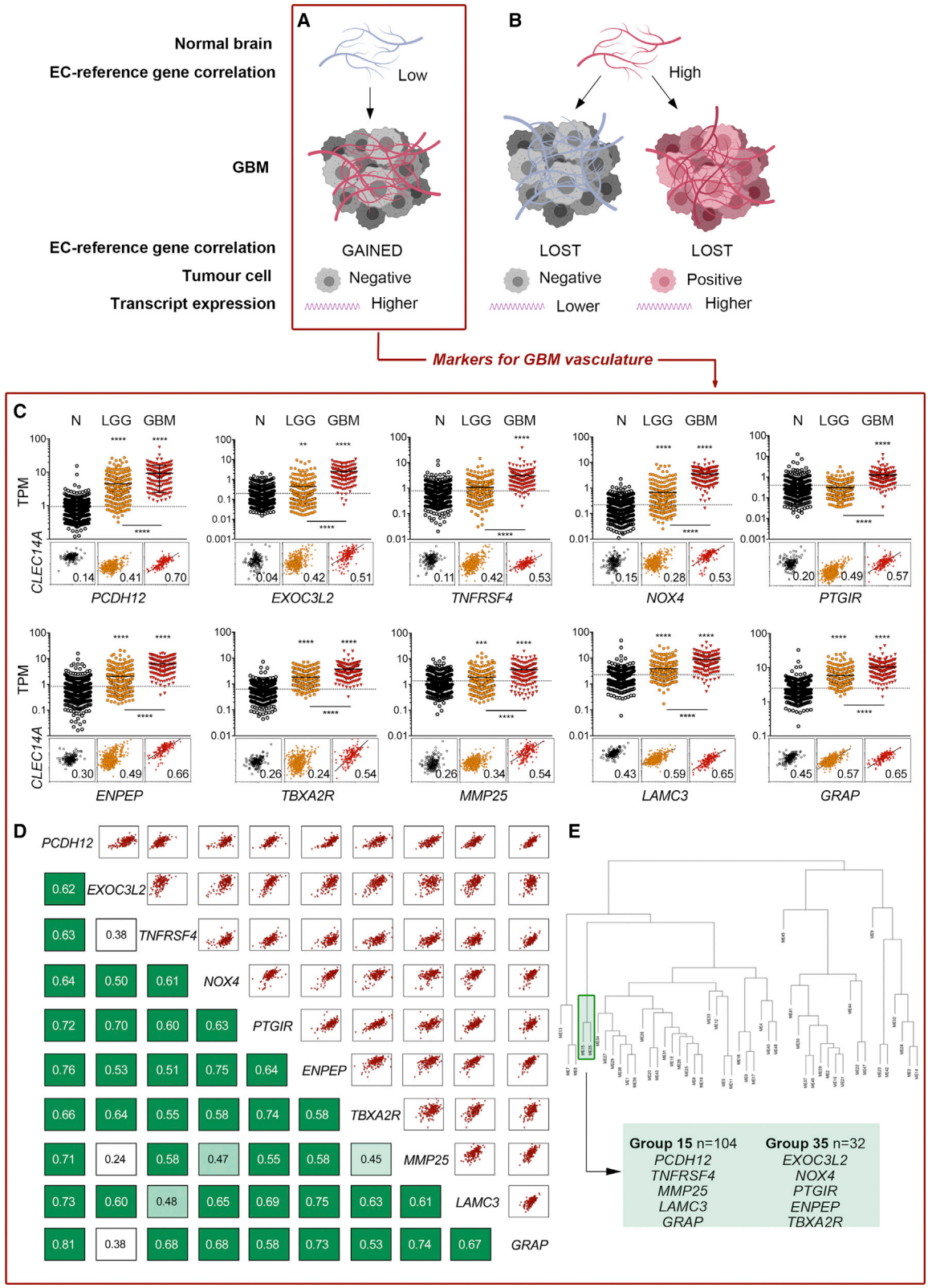
We performed a global analysis to predict changes in the EC-enriched transcriptome in GBM. Bulk RNA-seq has been used to define GBM molecular signatures (Jovčevska, 2018), but not to resolve cell-type-specific changes. A small number of studies have dissociated GBM tissue for transcriptome analysis on a cell-by-cell basis (Darmanis et al., 2017; Dieterich et al., 2012; Patel et al., 2014; Yuan et al., 2018), but it remains challenging to define GBM-specific cell-type signatures, due to the small number of tumors analyzed. Brain ECs are highly specialized with extensive adherens and tight junctions (Tietz and Engelhardt, 2015) and reduced pinocytosis, transcytosis, non-selective fenestration, and adhesion molecule expression (Zhao et al., 2015). We identified transcripts enriched in normal brain ECs, but not in GBM ECs, including those encoding for specialized brain EC proteins, such as the drug efflux pump TP-binding cassette sub-family B member 1 (ABCB1) (Borst and Schinkel, 2013), and the immuno- and permeability regulatory protein Periaxin (PRX) (Wang et al., 2018). As specialized brain EC features are regulated by the local microenvironment (Chow and Gu, 2015) and rapidly lost *ex vivo* (Urich et al., 2012; Wilhelm et al., 2011), one could speculate that GBM ECs fail to express such genes due to the loss of the normal microenvironment. The majority of transcripts we identified as having EC-specific expression in both normal and GBM were "core" EC genes, found across human tissue beds (Butler et al., 2016); suggesting GBM ECs maintain genes critical for function. Studies indicate that a population of ECs in GBM could originate from the transdifferentiation of tumor cells (Soda et al., 2011; Wang et al., 2010), but in the absence of a consensus on this topic and accompanying definitions, we did not attempt to resolve these potential populations in our analysis workflow. We provide a panel of predicted GBM EC markers, some of which have been previously reported as such, including CD93 (Dieterich et al., 2012; Langenkamp et al., 2015) and ANGPT2 (Dieterich et al., 2012; Scholz et al., 2016; Stratmann et al., 1998). The majority, to our knowledge, were previously unknown, e.g., exocyst complex component 3-like 2 (EXOC3L2), TNF receptor superfamily member 4 (TNFRSF4), thromboxane A2 receptor (TBXA2R), and prostaglandin I2 receptor (PTGIR); these are interesting candidates for further investigation, but experimental validation should be performed to confirm the predicted expression profile. EXOC3L2, a component of the exocyst involved in vesicle fusion with the plasma membrane, is poorly characterized; the only published study on the protein describes its upregulation in mouse embryonic sprouting blood vessels and induction in human ECs in response to vascular endothelial growth factor (VEGF) (Barkefors et al., 2011). This is consistent with its expression in GBM ECs; GBM tumor cells produce VEGF, which induces angiogenesis (Das and Marsden, 2013); however, the role of EXOC3L2, or potential as drug target in GBM remains unknown. TNFRSF4 (Ox40) has been a target of experimental tumor immunotherapy, including in glioma, due to its expression on tumor infiltrating T cells (Buchan et al., 2018). In a mouse sarcoma model, TNFRSF4 was also expressed on ECs, which, following TNFRSF4-agonist treatment, responded with (T-cell-independent) VCAM-1 expression (Pardee et al., 2010). Thus, one could speculate that TNFRSF4 on GBM ECs

Figure 6. Using Ref. Transcript-Based Corr. Analysis with mRNA Expression and Tissue Protein Profiling to Investigate EC-Enriched Modifications in LGG and GBM

Normal human cortex RNA-seq data (N) (n = 158), human lower-grade glioma (LGG) (n = 516), and human glioblastoma (GBM) (n = 401) was used to calculate: (1) mean corr. coefficient between EC ref. transcripts (*CD34*, *CLEC14A*, and *VWF*) and selected genes (upper plot shows *CLEC14A* versus gene). mRNA expression (TPM) for selected genes were downloaded from OASIS (Fernandez-Banet et al., 2016) for N, LGG, and GBM and displayed as (2) individual samples and (3) as a mean expression in all. (4) Protein profiling was performed for selected genes on tissue sections of N, LGG, and GBM. Scale bar, 100 μm.

(A–C) Candidate genes were classified as follows: (A) EC-enriched in GBM, but not normal (N) tissue (red box) (i, ACE; ii, CD93; and iii, ANGPT2), EC-enriched in normal (N) tissue, but not GBM (green box) either due to (B) *loss* of expression from GBM EC (i, *ANXA3*; ii, *PRX*; and iii, *ABCB1*) or (C) gain of expression outside the vasculature in GBM (i, CAV2; ii, A2M; and iii, HLA-B); *$p < 0.05$ and ****$p < 0.0001$.

See Table S8. Letter and number on top left of image denote tissue sex (F, female; M, male) and age (years).



(legend on next page)

could indicate a role for EC interaction in the antitumor effect of TNFRSF4-targeted immunotherapy (Curry et al., 2016). TBXA2R and PTGIR encode for prostaglandin receptors, which bind thromboxane A2 and prostacyclin, respectively, products of the short-lived intermediate molecule prostaglandin H_2, generated from arachidonic acid, via cyclooxygenase (COX)-1 or -2. COX-2 is elevated in GBM, and levels positively correlated with tumor grade, reoccurrence, and shorter survival (Qiu et al., 2017); however, the expression or role of prostaglandin receptors on GBM ECs has not been explored. A panel of GBM EC-enriched genes we identified are linked through the renin-angiotensin system (RAS): angiotensin I converting enzyme (ACE), glutamyl aminopeptidase (ENPEP), and NADPH oxidase 4 (NOX4). RAS inhibitors suppress progression and lengthen survival in several cancer types (Ishikane and Takahashi-Yanaga, 2018; Pinter and Jain, 2017), including glioma (Levin et al., 2017), and are currently being considered for inclusion in treatment protocols (Perdomo-Pantoja et al., 2018). ACE and ENPEP are central enzymes in the RAS, catalyzing the conversion of angiotensin I to angiotensin II, and angiotensin II to angiotensin III, respectively. Angiotensin II and III have various effects, mediated mainly through angiotensin receptor I (ATR1) (Jackson et al., 2018), such as the induction of pro-survival signaling, via nuclear factor κB (NF-κB)-mediated anti-apoptotic molecules, or through PI3K-Akt-mediated suppression of caspases (George et al., 2010), the induction of proliferation (Clark et al., 2011; Perdomo-Pantoja et al., 2018), and loss of brain EC barrier properties (Biancardi and Stern, 2016). More high-grade astrocytoma tumor cells express ATR1 than lower grade, and expression is associated with cell proliferation, vascularization, and shorter survival (Arrieta et al., 2008). Angiotensin II induces NOX4, which regulates reactive oxygen species (ROS) production (Nguyen Dinh Cat et al., 2013). ROS play an important role in signal transduction, cell differentiation, tumor cell proliferation, apoptosis, and angiogenesis (Guo and Chen, 2015). NOX4 is more highly expressed in GBM than lower-grade tumors, with a role in tumor proliferation and resistance to chemotherapeutic agent-induced apoptosis (Shono et al., 2008). Despite the acknowledged importance of RAS in cancer, the cell types underlying the increased signaling are not well described. Based on our data, we speculate that GBM ECs drive increased local levels of angiotensin II and II, promoting cancer progression and maintenance.

In summary, we identify cell-enriched genes from unfractionated brain RNA-seq, without the need for complex modeling. We profile system-level EC changes associated with brain malignancy of increasing severity, providing biological insight and identifying potential targets for therapy.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Tissue Profiling: Human Tissue Sections
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Reference Transcript-Based Correlation Analysis
  - Weighted Correlation Network (WGCNA) Analysis
  - Brain Single-Cell and Isolated Cell Fraction RNA-Seq Datasets
  - Gene Ontology (GO) Enrichment Analysis
- DATA AND CODE AVAILABILITY
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.09.088.

ACKNOWLEDGMENTS

Funding was granted to L.M.B. from the Swedish Heart-Lung Foundation (20170759 and 20170537) and the Swedish Research Council (2013-42608-102305-28) and to J.O. from the Stockholm County Council (SLL 2017-0842). P.D. was supported by an SFO grant, KTH, to J.O. and L.M.B. The HPA is funded by The Knut and Alice Wallenberg Foundation. Data usage: we used data from the Genotype-Tissue Expression (GTEx) Project (gtexportal.org) (GTEx Consortium, 2015), the TCGA Research Network (https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga), and the AMP-AD Knowledge Portal (https://www.synapse.org/). The GTEx Project was supported by the Office of the Director of the NIH, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The Mayo RNA-seq study data were provided by The Mayo Clinic Alzheimer's Disease Genetic Studies, led by Dr. Taner and Dr. Younkin, Mayo Clinic, Jacksonville, FL, using samples from the Mayo Clinic Study of Aging, the Mayo Clinic Alzheimer's Disease Research Centre, and the Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, and R01 AG003949; NINDS grant R01 NS080820; the CurePSP Foundation; and support from the Mayo Foundation. Data include samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, AZ. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026; National Brain and Tissue Resource for Parkinson's Disease and Related Disorders); the National Institute on Aging (P30 AG19610; Arizona Alzheimer's Disease Core Center); the Arizona Department of Health

Figure 7. Predicted Glioblastoma (GBM)-Specific EC-Enriched Genes

(A and B) Summary of EC-enriched transcriptome modifications in GBM: (A) EC-transcripts classified as GBM only (*increased* EC-enrichment score) and (B) EC-transcripts absent in GBM EC *or* gained in non-EC (both *decreased* EC-enrichment score).

(C) mRNA expression (TPM) for GBM EC marker candidates in normal brain (N) (n = 158), lower-grade glioma (LGG) (n = 516), and GBM (n = 401), and illustrative corr. plots versus EC Ref. transcript *CLEC14A,* with mean values versus all EC ref. transcripts (*CD34*, *CLEC14A*, and *VWF*) in bottom right of plot.

(D) Corr. matrix of GBM marker candidates in GBM RNA-seq.

(E) Weighted corr. network analysis (WGCNA) dendrogram of unfractionated GBM: EC groups identified by green boxes. $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$.

See also Table S8.

Services (Contract 211002; Arizona Alzheimer's Research Centre); the Arizona Biomedical Research Commission (Contracts 4001, 0011, 05-901, and 1001 to the Arizona Parkinson's Disease Consortium); and the Michael J. Fox Foundation for Parkinson's Research.

AUTHOR CONTRIBUTIONS

Conceptualization, L.M.B.; Methodology, P.D., B.M.H., and L.M.B.; Formal Analysis, P.D., B.M.H., and L.M.B.; Investigation, P.D. and L.M.B.; Resources, T.R. and M.U.; Writing – Original Draft, P.D. and L.M.B.; Writing – Review & Editing, P.D., B.M.H., T.R., J.O., M.U., and L.M.B.; Visualization, P.D. and L.M.B.; Supervision, L.M.B.; Funding Acquisition, J.O. and L.M.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 21, 2019
Revised: April 9, 2019
Accepted: September 27, 2019
Published: November 5, 2019

REFERENCES

Allen, M., Carrasquillo, M.M., Funk, C., Heavner, B.D., Zou, F., Younkin, C.S., Burgess, J.D., Chai, H.-S., Crook, J., Eddy, J.A., et al. (2016). Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. Sci. Data *3*, 160089.

Armulik, A., Genové, G., and Betsholtz, C. (2011). Pericytes: developmental, physiological, and pathological perspectives, problems, and promises. Dev. Cell *21*, 193–215.

Arrieta, O., Pineda-Olvera, B., Guevara-Salazar, P., Hernández-Pedro, N., Morales-Espinosa, D., Cerón-Lizarraga, T.L., González-De la Rosa, C.H., Rembao, D., Segura-Pacheco, B., and Sotelo, J. (2008). Expression of AT1 and AT2 angiotensin receptors in astrocytomas is associated with poor prognosis. Br. J. Cancer *99*, 160–166.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.

Azevedo, F.A.C., Carvalho, L.R.B., Grinberg, L.T., Farfel, J.M., Ferretti, R.E.L., Leite, R.E.P., Jacob Filho, W., Lent, R., and Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. J. Comp. Neurol. *513*, 532–541.

Barkefors, I., Fuchs, P.F., Heldin, J., Bergström, T., Forsberg-Nilsson, K., and Kreuger, J. (2011). Exocyst complex component 3-like 2 (EXOC3L2) associates with the exocyst complex and mediates directional migration of endothelial cells. J. Biol. Chem. *286*, 24189–24199.

Beliakova-Bethell, N., Massanella, M., White, C., Lada, S., Du, P., Vaida, F., Blanco, J., Spina, C.A., and Woelk, C.H. (2014). The effect of cell subset isolation method on gene expression in leukocytes. Cytometry A *85*, 94–104.

Biancardi, V.C., and Stern, J.E. (2016). Compromised blood-brain barrier permeability: novel mechanism by which circulating angiotensin II signals to sympathoexcitatory centres during hypertension. J. Physiol. *594*, 1591–1600.

Borst, P., and Schinkel, A.H. (2013). P-glycoprotein ABCB1: a major player in drug handling by mammals. J. Clin. Invest. *123*, 4131–4133.

Buchan, S.L., Rogel, A., and Al-Shamkhani, A. (2018). The immunobiology of CD27 and OX40 and their potential as targets for cancer immunotherapy. Blood *131*, 39–48.

Butler, L.M., Hallstrom, B.M., Fagerberg, L., Ponten, F., Uhlen, M., Renne, T., and Odeberg, J. (2016). Analysis of body-wide unfractionated tissue data to identify a core human endothelial transcriptome. Cell Syst*3*, pp. 287–301.e3.

Cahoy, J.D., Emery, B., Kaushal, A., Foo, L.C., Zamanian, J.L., Christopherson, K.S., Xing, Y., Lubischer, J.L., Krieg, P.A., Krupenko, S.A., et al. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J. Neurosci. *28*, 264–278.

Chow, B.W., and Gu, C. (2015). The molecular constituents of the blood-brain barrier. Trends Neurosci. *38*, 598–608.

Chu, T.J., and Peters, D.G. (2008). Serial analysis of the vascular endothelial transcriptome under static and shear stress conditions. Physiol. Genomics *34*, 185–192.

Clark, M.A., Tran, H., and Nguyen, C. (2011). Angiotensin III stimulates ERK1/2 mitogen-activated protein kinases and astrocyte growth in cultured rat astrocytes. Neuropeptides *45*, 329–335.

Conley, C.A. (2001). Leiomodin and tropomodulin in smooth muscle. Am. J. Physiol. Cell Physiol. *280*, C1645–C1656.

GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648–660.

Crouch, E.E., and Doetsch, F. (2018). FACS isolation of endothelial cells and pericytes from mouse brain microregions. Nat. Protoc. *13*, 738–751.

Curry, W.T., Jr., Gorrepati, R., Piesche, M., Sasada, T., Agarwalla, P., Jones, P.S., Gerstner, E.R., Golby, A.J., Batchelor, T.T., Wen, P.Y., et al. (2016). Vaccination with irradiated autologous tumor cells mixed with irradiated GM-K562 cells stimulates antitumor immunity and T lymphocyte activation in patients with recurrent malignant glioma. Clin. Cancer Res. *22*, 2885–2896.

Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Hayden Gephart, M.G., Barres, B.A., and Quake, S.R. (2015). A survey of human brain transcriptome diversity at the single cell level. Proc. Natl. Acad. Sci. USA *112*, 7285–7290.

Darmanis, S., Sloan, S.A., Croote, D., Mignardi, M., Chernikova, S., Samghababi, P., Zhang, Y., Neff, N., Kowarsky, M., Caneda, C., et al. (2017). Single-cell RNA-seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. Cell Rep. *21*, 1399–1410.

Das, S., and Marsden, P.A. (2013). Angiogenesis in glioblastoma. N. Engl. J. Med. *369*, 1561–1563.

Dieterich, L.C., Mellberg, S., Langenkamp, E., Zhang, L., Zieba, A., Salomäki, H., Teichert, M., Huang, H., Edqvist, P.H., Kraus, T., et al. (2012). Transcriptional profiling of human glioblastoma vessels indicates a key role of VEGF-A and TGFβ2 in vascular abnormalization. J. Pathol. *228*, 378–390.

Dreiza, C.M., Komalavilas, P., Furnish, E.J., Flynn, C.R., Sheller, M.R., Smoke, C.C., Lopes, L.B., and Brophy, C.M. (2010). The small heat shock protein, HSPB6, in muscle function and disease. Cell Stress Chaperones *15*, 1–11.

Elo, L.L., Filen, S., Lahesmaa, R., and Aittokallio, T. (2008). Reproducibility-optimized test statistic for ranking genes in microarray studies. IEEE/ACM Trans. Comput. Biol. Bioinform *5*, 423–431.

Fernandez-Banet, J., Esposito, A., Coffin, S., Horvath, I.B., Estrella, H., Schefzick, S., Deng, S., Wang, K., AChing, K., Ding, Y., et al. (2016). OASIS: web-based platform for exploring cancer multi-omics data. Nat. Methods *13*, 9–10.

Fonseca, M.I., Chu, S.H., Hernandez, M.X., Fang, M.J., Modarresi, L., Selvan, P., MacGregor, G.R., and Tenner, A.J. (2017). Cell-specific deletion of C1qa identifies microglia as the dominant source of C1q in mouse brain. J. Neuroinflammation *14*, 48.

George, A.J., Thomas, W.G., and Hannan, R.D. (2010). The renin-angiotensin system and cancer: old dog, new tricks. Nat. Rev. Cancer *10*, 745–759.

Guo, S., and Chen, X. (2015). The human Nox4: gene, structure, physiological function and pathological significance. J. Drug Target. *23*, 888–896.

He, L., Vanlandewijck, M., Raschperger, E., Andaloussi Mäe, M., Jung, B., Lebouvier, T., Ando, K., Hofmann, J., Keller, A., and Betsholtz, C. (2016). Analysis of the brain mural cell transcriptome. Sci. Rep. *6*, 35108.

Ishikane, S., and Takahashi-Yanaga, F. (2018). The role of angiotensin II in cancer metastasis: potential of renin-angiotensin system blockade as a treatment for cancer metastasis. Biochem. Pharmacol. *151*, 96–103.

Ito, D., Imai, Y., Ohsawa, K., Nakajima, K., Fukuuchi, Y., and Kohsaka, S. (1998). Microglia-specific localisation of a novel calcium binding protein, Iba1. Brain Res. Mol. Brain Res. *57*, 1–9.

Jaakkola, M.K., Seyednasrollah, F., Mehmood, A., and Elo, L.L. (2017). Comparison of methods to detect differentially expressed genes between single-cell populations. Brief. Bioinform. *18*, 735–743.

Jackson, L., Eldahshan, W., Fagan, S.C., and Ergul, A. (2018). Within the brain: the renin angiotensin system. Int. J. Mol. Sci. *19*, 876.

Jovčevska, I. (2018). Sequencing the next generation of glioblastomas. Crit. Rev. Clin. Lab. Sci. *55*, 264–282.

Kelley, K.W., Nakao-Inoue, H., Molofsky, A.V., and Oldham, M.C. (2018). Variation among intact tissue samples reveals the core transcriptional features of human CNS cell classes. Nat. Neurosci. *21*, 1171–1184.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

Langenkamp, E., Zhang, L., Lugano, R., Huang, H., Elhassan, T.E., Georganaki, M., Bazzar, W., Lööf, J., Trendelenburg, G., Essand, M., et al. (2015). Elevated expression of the C-type lectin CD93 in the glioblastoma vasculature regulates cytoskeletal rearrangements that enhance vessel function and reduce host survival. Cancer Res. *75*, 4504–4516.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Levin, V.A., Chan, J., Datta, M., Yee, J.L., and Jain, R.K. (2017). Effect of angiotensin system inhibitors on survival in newly diagnosed glioma patients and recurrent glioblastoma patients receiving chemotherapy and/or bevacizumab. J. Neurooncol. *134*, 325–330.

Lim, M., Xia, Y., Bettegowda, C., and Weller, M. (2018). Current state of immunotherapy for glioblastoma. Nat. Rev. Clin. Oncol. *15*, 422–442.

Long, X., Tharp, D.L., Georger, M.A., Slivano, O.J., Lee, M.Y., Wamhoff, B.R., Bowles, D.K., and Miano, J.M. (2009). The smooth muscle cell-restricted KCNMB1 ion channel subunit is a direct transcriptional target of serum response factor and myocardin. J. Biol. Chem. *284*, 33671–33682.

Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nat. Protoc. *8*, 1551–1566.

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2016). PANTHER version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Res. *44* (*D1*), D336–D342.

Miwa, T., Manabe, Y., Kurokawa, K., Kamada, S., Kanda, N., Bruns, G., Ueyama, H., and Kakunaga, T. (1991). Structure, chromosome location, and expression of the human smooth muscle (enteric type) gamma-actin gene: evolution of six human actin genes. Mol. Cell. Biol. *11*, 3296–3306.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457.

Nguyen Dinh Cat, A., Montezano, A.C., Burger, D., and Touyz, R.M. (2013). Angiotensin II, NADPH oxidase, and redox signaling in the vasculature. Antioxid. Redox Signal. *19*, 1110–1120.

Osuka, S., and Van Meir, E.G. (2017). Overcoming therapeutic resistance in glioblastoma: the way forward. J. Clin. Invest. *127*, 415–426.

Pandey, S., Shekhar, K., Regev, A., and Schier, A.F. (2018). Comprehensive identification and spatial mapping of habenular neuronal types using single-cell RNA-seq. Curr. Biol. *28*, 1052–1065.e7.

Papalexi, E., and Satija, R. (2018). Single-cell RNA sequencing to explore immune cell heterogeneity. Nat. Rev. Immunol *18*, 35–45.

Pardee, A.D., McCurry, D., Alber, S., Hu, P., Epstein, A.L., and Storkus, W.J. (2010). A therapeutic OX40 agonist dynamically alters dendritic, endothelial, and T cell subsets within the established tumor microenvironment. Cancer Res. *70*, 9041–9052.

Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401.

Perdomo-Pantoja, A., Mejía-Pérez, S.I., Gómez-Flores-Ramos, L., Lara-Velazquez, M., Orillac, C., Gómez-Amador, J.L., and Wegman-Ostrosky, T. (2018). Renin angiotensin system and its role in biomarkers and treatment in gliomas. J. Neurooncol. *138*, 1–15.

Pfeiffer, S.E., Warrington, A.E., and Bansal, R. (1993). The oligodendrocyte and its many cellular processes. Trends Cell Biol. *3*, 191–197.

Pinter, M., and Jain, R.K. (2017). Targeting the renin-angiotensin system to improve cancer treatment: implications for immunotherapy. Sci. Transl. Med. *9*, eaan5616.

Pontén, F., Jirström, K., and Uhlen, M. (2008). The Human Protein Atlas—a tool for pathology. J. Pathol. *216*, 387–393.

Qiu, J., Shi, Z., and Jiang, J. (2017). Cyclooxygenase-2 in glioblastoma multiforme. Drug Discov. Today *22*, 148–156.

Reddy, A.S., O'Brien, D., Pisat, N., Weichselbaum, C.T., Sakers, K., Lisci, M., Dalal, J.S., and Dougherty, J.D. (2017). A comprehensive analysis of cell type-specific nuclear RNA from neurons and glia of the brain. Biol. Psychiatry *81*, 252–264.

Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al.; Human Cell Atlas Meeting Participants (2017). The Human Cell Atlas. eLife *6*, e27041.

Rensen, S.S., Doevendans, P.A., and van Eys, G.J. (2007). Regulation and characteristics of vascular smooth muscle cell phenotypic diversity. Neth. Heart J. *15*, 100–108.

Rizzetto, S., Eltahla, A.A., Lin, P., Bull, R., Lloyd, A.R., Ho, J.W.K., Venturi, V., and Luciani, F. (2017). Impact of sequencing depth and read length on single cell RNA sequencing data of T cells. Sci. Rep. *7*, 12781.

Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L.T., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science *360*, 176–182.

Saliba, A.-E., Westermann, A.J., Gorski, S.A., and Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. Nucleic Acids Res. *42*, 8845–8860.

Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S., et al. (2018). Molecular diversity and specializations among the cells of the adult mouse brain. Cell *174*, 1015–1030.e16.

Scholz, A., Harter, P.N., Cremer, S., Yalcin, B.H., Gurnik, S., Yamaji, M., Di Tacchio, M., Sommer, K., Baumgarten, P., Bähr, O., et al. (2016). Endothelial cell-derived angiopoietin-2 is a therapeutic target in treatment-naive and bevacizumab-resistant glioblastoma. EMBO Mol. Med. *8*, 39–57.

Shah, A.V., Birdsey, G.M., and Randi, A.M. (2016). Regulation of endothelial homeostasis, vascular development and angiogenesis by the transcription factor ERG. Vascul. Pharmacol. *86*, 3–13.

Shono, T., Yokoyama, N., Uesaka, T., Kuroda, J., Takeya, R., Yamasaki, T., Amano, T., Mizoguchi, M., Suzuki, S.O., Niiro, H., et al. (2008). Enhanced expression of NADPH oxidase Nox4 in human gliomas and its roles in cell proliferation and survival. Int. J. Cancer *123*, 787–792.

Smyth, L.C.D., Rustenhoven, J., Scotter, E.L., Schweder, P., Faull, R.L.M., Park, T.I.H., and Dragunow, M. (2018). Markers for human brain pericytes and smooth muscle cells. J. Chem. Neuroanat. *92*, 48–60.

Soda, Y., Marumoto, T., Friedmann-Morvinski, D., Soda, M., Liu, F., Michiue, H., Pastorino, S., Yang, M., Hoffman, R.M., Kesari, S., and Verma, I.M. (2011). Transdifferentiation of glioblastoma cells into vascular endothelial cells. Proc. Natl. Acad. Sci. USA *108*, 4274–4280.

Stratmann, A., Risau, W., and Plate, K.H. (1998). Cell type-specific expression of angiopoietin-1 and angiopoietin-2 suggests a role in glioblastoma angiogenesis. Am. J. Pathol. *153*, 1459–1466.

Sun, W., Cornwell, A., Li, J., Peng, S., Osorio, M.J., Aalling, N., Wang, S., Benraiss, A., Lou, N., Goldman, S.A., and Nedergaard, M. (2017). SOX9 is an astrocyte-specific nuclear marker in the adult brain outside the neurogenic regions. J. Neurosci. *37*, 4493–4507.

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS ONE *6*, e21800.

Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. Science *356*, eaal3321.

Tietz, S., and Engelhardt, B. (2015). Brain barriers: crosstalk between complex tight junctions and adherens junctions. J. Cell Biol. *209*, 493–506.

Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.

Uhlén, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., Benfeitas, R., Arif, M., Liu, Z., Edfors, F., et al. (2017). A pathology atlas of the human cancer transcriptome. Science *357*, eaan2507.

Urich, E., Lazic, S.E., Molnos, J., Wells, I., and Freskgård, P.O. (2012). Transcriptional profiling of human brain endothelial cells reveals key properties crucial for predictive in vitro blood-brain barrier models. PLoS ONE *7*, e38149.

Vanlandewijck, M., He, L., Mäe, M.A., Andrae, J., Ando, K., Del Gaudio, F., Nahar, K., Lebouvier, T., Laviña, B., Gouveia, L., et al. (2018). A molecular atlas of cell types and zonation in the brain vasculature. Nature *554*, 475–480.

Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science *356*, eaah4573.

von Bartheld, C.S., Bahney, J., and Herculano-Houzel, S. (2016). The search for true numbers of neurons and glial cells in the human brain: a review of 150 years of cell counting. J. Comp. Neurol. *524*, 3865–3895.

Wang, Z., Wang, D.Z., Pipes, G.C., and Olson, E.N. (2003). Myocardin is a master regulator of smooth muscle gene expression. Proc. Natl. Acad. Sci. USA *100*, 7129–7134.

Wang, R., Chadalavada, K., Wilshire, J., Kowalik, U., Hovinga, K.E., Geber, A., Fligelman, B., Leversha, M., Brennan, C., and Tabar, V. (2010). Glioblastoma stem-like cells give rise to tumour endothelium. Nature *468*, 829–833.

Wang, M.M., Zhang, X., Lee, S.J., Maripudi, S., Keep, R.F., Johnson, A.M., Stamatovic, S.M., and Andjelkovic, A.V. (2018). Expression of periaxin (PRX) specifically in the human cerebrovascular system: PDZ domain-mediated strengthening of endothelial barrier function. Sci. Rep. *8*, 10042.

Wilhelm, I., Fazakas, C., and Krizbai, I.A. (2011). In vitro models of the blood-brain barrier. Acta Neurobiol. Exp. (Warsz.) *71*, 113–128.

Wilhelm, I., Nyúl-Tóth, Á., Suciu, M., Hermenean, A., and Krizbai, I.A. (2016). Heterogeneity of the blood-brain barrier. Tissue Barriers *4*, e1143544.

Yamawaki, K., Ito, M., Machida, H., Moriki, N., Okamoto, R., Isaka, N., Shimpo, H., Kohda, A., Okumura, K., Hartshorne, D.J., and Nakano, T. (2001). Identification of human CPI-17, an inhibitory phosphoprotein for myosin phosphatase. Biochem. Biophys. Res. Commun. *285*, 1040–1045.

Yan, H., Romero-López, M., Benitez, L.I., Di, K., Frieboes, H.B., Hughes, C.C.W., Bota, D.A., and Lowengrub, J.S. (2017). 3D mathematical modeling of glioblastoma suggests that transdifferentiated vascular endothelial cells mediate resistance to current standard-of-care therapy. Cancer Res. *77*, 4171–4184.

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. *4*, 2612.

Yuan, J., Levitin, H.M., Frattini, V., Bush, E.C., Boyett, D.M., Samanamud, J., Ceccarelli, M., Dovas, A., Zanazzi, G., Canoll, P., et al. (2018). Single-cell transcriptome analysis of lineage diversity in high-grade glioma. Genome Med. *10*, 57.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.

Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N., et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J. Neurosci. *34*, 11929–11947.

Zhang, Y., Sloan, S.A., Clarke, L.E., Caneda, C., Plaza, C.A., Blumenthal, P.D., Vogel, H., Steinberg, G.K., Edwards, M.S.B., Li, G., et al. (2016). Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. Neuron *89*, 37–53.

Zhao, Z., Nelson, A.R., Betsholtz, C., and Zlokovic, B.V. (2015). Establishment and dysfunction of the blood-brain barrier. Cell *163*, 1064–1078.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative analysis of single-cell RNA sequencing methods. Mol. Cell *65*, 631–643.e4.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
AHNAK	Atlas Antibodies	Cat#HPA026643
A2M	Atlas Antibodies	Cat#HPA002265
ANPEP	Atlas Antibodies	Cat#HPA004625
ANXA3	Atlas Antibodies	Cat#HPA013398
C12orf43	Atlas Antibodies	Cat#HPA046148
CALD1	Atlas Antibodies	Cat#HPA008066
CAPN13	Atlas Antibodies	Cat#HPA029497
CARNS1	Atlas Antibodies	Cat#HPA038569
CD34	Atlas Antibodies	Cat#HPA036722
CD93	Atlas Antibodies	Cat#HPA012368
CLEC14A	Atlas Antibodies	Cat#HPA039468
CMTM6	Atlas Antibodies	Cat#HPA026980
ERLIN2	Atlas Antibodies	Cat#HPA002025
FAM163B	Atlas Antibodies	Cat#HPA067336
GGT5	Atlas Antibodies	Cat#HPA008121
HABP4	Atlas Antibodies	Cat#HPA055969
HCLS1	Atlas Antibodies	Cat#HPA019143
HSDL2	Atlas Antibodies	Cat#HPA050453
ITGB1	Atlas Antibodies	Cat#HPA059297
KANK3	Atlas Antibodies	Cat#HPA051153
MCC	Atlas Antibodies	Cat#HPA037391
MID1	Atlas Antibodies	Cat#HPA003715
MOAP1	Atlas Antibodies	Cat#HPA000939
NCKAP1L	Atlas Antibodies	Cat#HPA039490
NUDCD3	Atlas Antibodies	Cat#HPA019529
PHLDB1	Atlas Antibodies	Cat#HPA037959
PLEKHG3	Atlas Antibodies	Cat#HPA074734
PROCR	Atlas Antibodies	Cat#HPA039461
PRSS23	Atlas Antibodies	Cat#HPA030591
PRX	Atlas Antibodies	Cat#HPA001868
PSMB9	Atlas Antibodies	Cat#HPA053280
SASH3	Atlas Antibodies	Cat#HPA001085
SCIN	Atlas Antibodies	Cat#HPA024264
SHROOM4	Atlas Antibodies	Cat#HPA010565
SLFN12	Atlas Antibodies	Cat#HPA022523
SMAP2	Atlas Antibodies	Cat#HPA024424
SNX20	Atlas Antibodies	Cat#HPA043649
SP100	Atlas Antibodies	Cat#HPA016707
TAGLN2	Atlas Antibodies	Cat#HPA001925
TINAGL1	Atlas Antibodies	Cat#HPA048695
TLN1	Atlas Antibodies	Cat#HPA004748
TMEM47	Atlas Antibodies	Cat#HPA046658
VWF	Atlas Antibodies	Cat#HPA001815

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DCN	Abnova	Cat#H00001634-M01
VIM	Agilent	Cat#M7020
ACE	Leica Biosystems	Cat#NCL-CD143-510
ABCB1	Merck	Cat#MAB4120
ANGPT2	Santa Cruz Biotechnology	Cat#sc-74403
HLA-B	Santa Cruz Biotechnology	Cat#sc-55582
LIPE	Santa Cruz Biotechnology	Cat#sc-74489
MYL12A	Santa Cruz Biotechnology	Cat#sc-28329
PGM5	Santa Cruz Biotechnology	Cat#sc-73613
LAMB2	Sigma-Aldrich	Cat#AMAb91097
CAV2	Thermo Fisher Scientific	Cat#41-0700
CLDN5	Thermo Fisher Scientific	Cat#RB-9243
Biological Samples		
Normal human cortex	Uppsala Biobank	http://www.proteinatlas.org
Lower grade glioma	Uppsala Biobank	http://www.proteinatlas.org
Glioblastoma	Uppsala Biobank	http://www.proteinatlas.org
Deposited Data		
All analysis	This paper	Mendeley Data https://doi.org/10.17632/mvdsfk69j6.1
Other		
Human Protein Atlas (HPA)		http://www.proteinatlas.org
Web-based portal for this paper	This paper	https://cell-enrichment.shinyapps.io/Brain/

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact: Dr. Lynn Marie Butler. Email: Lynn.butler@ki.se

This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bulk RNaseq data analyzed in this study was part of the Genotype-Tissue Expression (GTEx) Project (gtexportal.org) (GTEx Consortium, 2015) (dbGaP Accession phs000424.v7.p2) (n = 158, 108 male, 50 female, ages 20-79), the TCGA Research Network (https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga) and *AMP-AD Knowledge Portal* (https://www.synapse.org) (study ID: syn3163039) (n = 80, 41 male, 39 female, ages 53-90). Datasets from normal brain scRNaseq (Darmanis et al., 2015) and RNaseq of isolated human and mouse brain cell types (Zhang et al., 2014, 2016) were downloaded from the Gene Expression Omnibus (GEO) database (Accession IDs: GSE67835, GSE73721 and GSE52564, respectively). Comparative transcript expression between normal, LGG and GBM samples were downloaded from OASIS (oasis-genomics.org) (Fernandez-Banet et al., 2016). Human tissue protein profiling was performed in house as part of the Human Protein Atlas (HPA) project (Pontén et al., 2008; Uhlén et al., 2015, 2017) (http://www.proteinatlas.org). Normal brain (cortex), lower grade glioma (LGG) and glioblastoma (GBM) samples were obtained from the Department of Pathology, Uppsala University Hospital, Uppsala, Sweden; as part of the Uppsala Biobank. Samples were handled in accordance with Swedish laws and regulations, with approval from the Uppsala Ethical Review Board (Uhlén et al., 2015).

METHOD DETAILS

Tissue Profiling: Human Tissue Sections

Tissue sections from human cerebral cortex were generated and stained, as previously described (Pontén et al., 2008; Uhlén et al., 2015). Briefly, formalin fixed and paraffin embedded tissue samples were sectioned, de-paraffinized in xylene, hydrated in graded alcohols and blocked for endogenous peroxidase in 0.3% hydrogen peroxide diluted in 95% ethanol. For antigen retrieval, a Decloaking chamber® (Biocare Medical, CA) was used. Slides were boiled in Citrate buffer®, pH6 (Lab Vision, CA). Primary antibodies and a dextran polymer visualization system (UltraVision LP HRP polymer®, Lab Vision) were incubated for 30 min each at room

temperature and slides were developed for 10 minutes using Diaminobenzidine (Lab Vision) as the chromogen. Slides were counterstained in Mayers hematoxylin (Histolab) and scanned using Scanscope XT (Aperio). Primary antibodies, source, target, and identifier are as follows: Atlas Antibodies: AHNAK (HPA026643), A2M (HPA002265), ANPEP (HPA004625), ANXA3 (HPA013398), C12orf43 (HPA046148), CALD1 (HPA008066), CAPN13 (HPA029497), CARNS1 (HPA038569), CD34 (HPA036722), CD93 (HPA012368), CLEC14A (HPA039468), CMTM6 (HPA026980), ERLIN2 (HPA002025), FAM163B (HPA067336), GGT5 (HPA008121), HABP4 (HPA055969), HCLS1 (HPA019143), HSDL2 (HPA050453), ITGB1 (HPA059297), KANK3 (HPA051153), MCC (HPA037391), MID1 (HPA003715), MOAP1 (HPA000939), NCKAP1L (HPA039490), NUDCD3 (HPA019529), PHLDB1 (HPA037959), PLEKHG3 (HPA074734), PROCR (HPA039461), PRSS23 (HPA030591), PRX (HPA001868), PSMB9 (HPA053280), SASH3 (HPA001085), SCIN (HPA024264), SHROOM4 (HPA010565), SLFN12 (HPA022523), SMAP2 (HPA024424), SNX20 (HPA043649), SP100 (HPA016707), TAGLN2 (HPA001925), TINAGL1 (HPA048695), TLN1 (HPA004748), TMEM47 (HPA046658), VWF (HPA001815). Abnova: DCN (H00001634-M01). Agilent: VIM (M7020). Leica Biosystems: ACE (NCL-CD143-510). Merck: (ABCB1 (MAB4120). Santa Cruz Biotechnology: ANGPT2 (sc-74403), HLA-B (sc-55582), LIPE (sc-74489), MYL12A (sc-28329), PGM5 (sc-73613). Sigma-Aldrich: LAMB2 (AMAb91097). Thermo Fisher Scientific: CAV2 (41-0700), CLDN5 (RB-9243).

QUANTIFICATION AND STATISTICAL ANALYSIS

Reference Transcript-Based Correlation Analysis

This method was adapted from that previously developed to determine the cross-tissue pan-EC-enriched transcriptome (Butler et al., 2016). As different cell types are present in different proportions across individual samples, we used a correlation analysis to identify cell-type specific transcripts. We calculated the pairwise Spearman correlation coefficients between reference transcripts selected as proxy markers for: endothelial cells (EC): [*CD34*, *CLEC14A*, *VWF*] (in GTEx, TCGA [LGG and GBM] and MAYO datasets), astrocytes (AC): [*BMPR1B*, *AQP4*, *SOX9*] (in GTEx and MAYO datasets), oligodendrocytes (OC) [*MOG*, *CNP*, *MAG*] (GTEx and MAYO datasets), neurons (NC) the genes [*TMEM130*, *STMN2*, *THY1*] (GTEx and MAYO datasets), microglial cells (MG) [*C1QA*, *AIF1*, *LAPTM5*] (GTEx and MAYO datasets) and all mapped protein coding genes. To exclude false positives, we also calculated correlations between proxy markers for SMC [*FHL5*, *ACTA2*, *ACTG2*] and selected transcripts identified as cell-type enriched. Reference transcripts used in analysis of **GTEx**: EC, AC, OC, NC, MG, SMC; **LGG**: EC; **GBM**: EC; **MAYO**: EC, AC, OC, NC, MG. Non-coding transcripts and transcripts with mean expression TPM < 0.1 are excluded from final data tables. See results section for full analysis and exclusion criteria required for transcript classification as cell-type enriched. Correlation coefficients were calculated in R (v 3.4.3) using the *corr.test* function from the *psych* package (v 1.8.4). In addition to correlation coefficients False Discovery Rate (FDR) adjusted p values (using Bonferroni correction) and raw p values were calculated.

Weighted Correlation Network (WGCNA) Analysis

The R package WGCNA was used to perform co-expression network analysis for gene clustering, on log2 expression values. The analysis was performed according to recommendations in the WGCNA manual. Genes with too many missing values were excluded using the goodSamplesGenes() function. The remaining genes were used to cluster the samples, and obvious outlier samples were excluded. Using these genes and samples a soft-thresholding power was selected and the networks were constructed using a minimum module size of 15 and merging threshold of 0.05. Eigengenes were calculates from the resulting clusters and eigengene dendrograms were constructed using the plotEigengeneNetworks() function.

Brain Single-Cell and Isolated Cell Fraction RNA-Seq Datasets

Brain cell type expression datasets from scRNaseq (Darmanis et al., 2015) (GSE67835) and isolated cell type RNaseq (Zhang et al., 2014, 2016) (GSE73721 and GSE52564) studies were used to calculate cell type specific enrichment values for each mapped transcript. For scRNaseq, cell-type enrichment was calculated using the ROTS (reproducibility-optimized test statistic) analysis method (Elo et al., 2008), first developed for microarrays, but later applied to RNaseq data (Jaakkola et al., 2017). ROTS analysis uses a modified t-statistic which is adjusted based on inherent properties of the data, and ranks features based on statistical evidence of differential expression. In our analysis enrichment was defined as ROTS score > 2. For the isolated cell type RNaseq fold-enrichment values in each cell type: EC, AC, OC (*myelinating oligodendrocytes* and *newly formed oligodendrocytes* were combined), NC and MG were calculated as described in the original studies; '*FPKM expression in one cell type divided by the average expression level in all other cell types*'. Cell enrichment was defined as a fold-enrichment of > 2.

Gene Ontology (GO) Enrichment Analysis

The Gene Ontology Consortium (Ashburner et al., 2000) and PANTHER classification resource (Mi et al., 2013, 2016) were used to identify over represented terms (biological processes) in the panel of identified cell-type-enriched transcripts from the GO ontology database (release date March 2016).

DATA AND CODE AVAILABILITY

The published article includes all datasets generated during this study (Tables S1, S2, S3, S4, S5, S6, S7, and S8).

The datasets are also available at Mendeley Data, V1 [https://doi.org/10.17632/mvdsfk69j6.1].

ADDITIONAL RESOURCES

The Human Protein Atlas (HPA) website contains details of all antibody-based protein profiling used in this study: http://www.proteinatlas.org. A searchable web-based interface can be used to explore the datasets generated: http://cell-enrichment.shinyapps.io/Brain/.