CNN-based learning of single-cell transcriptomes reveals a blood-detectable multi-cancer signature of brain metastasis

Ryan Lusby[1], Debojyoti Chowdhury[2], Sarah Carl[3], Vijay K. Tiwari[1,2,4,5]*

1. Wellcome-Wolfson Institute for Experimental Medicine, School of Medicine, Dentistry & Biomedical Science, Queens University, Belfast, BT9 7BL, United Kingdom.
2. Institute for Molecular Medicine, University of Southern Denmark, Odense M, Denmark
3. Scailyte AG, Lichtstrasse 35, 4056 Basel, Switzerland
4. Danish Institute for Advanced Study (DIAS), Odense M, Denmark
5. Department of Clinical Genetics, Odense University Hospital, Odense C, Denmark

Correspondence should be addressed to: tiwari@health.sdu.dk

Running title: Decoding pan-cancer signatures of brain metastasis

Keywords: Deep learning, single-cell transcriptome, Tumour heterogeneity, Brain Metastasis, Cancer

Abstract

Brain metastasis (BrM) is a serious complication of advanced cancers and remains difficult to predict before clinical symptoms appear. To investigate shared transcriptional features of BrM across tumour types, we integrated single-cell RNA sequencing (scRNA-seq) data from malignant epithelial cells derived from six carcinoma types, including lung, breast, colorectal, renal, prostate, and melanoma. We applied ScaiVision, a supervised representation learning method, to classify tumour samples based on BrM status. The models achieved high predictive accuracy (area under the ROC curve > 0.90) across all six cancer types. This analysis identified a consistent multi-cancer gene expression signature associated with BrM, defined at single-cell resolution. To evaluate the clinical relevance of this signature, we assessed its presence in tumour-educated platelets (TEPs) from blood samples of patients with and without BrM. The signature was detectable in platelet RNA and distinguished patients with BrM from those without, indicating that features of the BrM-associated expression program are reflected in blood-derived material. These findings demonstrate that a transcriptional signature of brain metastasis can be identified across multiple tumour types using scRNA-seq and neural network-based analysis. The detectability of this signature in TEPs supports its relevance in a non-invasive context and provides a basis for further investigation into its utility for BrM risk assessment.

Introduction

Brain metastases (BrM) are the most common intracranial tumours, occurring in approximately 20% of adult cancer patients[1,2] . They represent a devastating complication of systemic malignancy, with limited treatment options and poor survival outcomes[1,3]. Most BrMs arise from primary cancers of the lung, breast, melanoma, and other solid tumours, and their development often portends neurologic decline and dismal prognosis. Despite improvements in cancer care, effective means to predict, detect, or prevent BrM remain an unmet clinical need[3]. In particular, the absence of robust biomarkers for early identification of patients at risk for BrM hampers timely intervention and surveillance strategies. This clinical burden and diagnostic gap underscore the urgency for deeper biological insights into BrM pathogenesis and markers of metastatic potential.

Recent advances in single-cell RNA sequencing (scRNA-seq) have begun to characterise the cellular and molecular architecture of BrMs, although significant knowledge gaps persist[4,5]. Single-cell approaches allow high-resolution dissection of the tumour microenvironment, revealing diverse cell types and states that contribute to brain metastatic colonisation. Notably, emerging scRNA-seq studies have identified certain conserved stromal and immune features across BrMs from different cancers. For example, cancer-associated fibroblasts within brain metastatic lesions display a remarkably conserved activation profile characterised by abundant type I collagen and wound-healing signatures, mirroring fibrotic programs seen in primary tumours[6]. Similarly, metastasis-associated myeloid cells (including resident microglia and infiltrating macrophages) in BrM often converge on shared phenotypic states; for instance, coexisting pro-inflammatory (M1-like) and immunosuppressive (M2-like) macrophage subsets have been observed across brain metastases of various origins[7]. These conserved fibroblast and macrophage responses suggest that disparate tumour types co-opt parallel microenvironmental programs when colonising the brain. However, translating such biological insights into clinical diagnostics or therapies is not straightforward. In particular, the question remains how to leverage single-cell discoveries to identify predictive markers of BrM before metastases become clinically evident.

A multi-cancer approach to this question is especially valuable given that BrMs arise from multiple tumour lineages. To date, many studies of BrM have been siloed by cancer type, focusing, for example, on BrM in lung cancer or in breast cancer separately. While such studies have yielded important tumour-specific insights (for instance, single-cell analyses in lung adenocarcinoma have identified a distinct subpopulation of brain metastasis-associated tumour cells in the primary lung tumour, marked by genes such as S100A9, which correlates

with subsequent BrM development[2]), they may overlook metastasis mechanisms that are shared across diseases. A comparative, cross-tumour analysis can reveal universal drivers and biomarkers of brain metastatic spread that single-tumour studies might miss. Indeed, evidence is mounting that certain BrM-associated gene expression changes in the tumour microenvironment (e.g. in stromal or immune compartments) are conserved irrespective of the cancer's tissue of origin[6,7]. Identifying such common signatures could inform broadly applicable diagnostic tools.

In light of the above, the present study leverages scRNA-seq across multiple cancer types to identify BrM-associated markers in primary tumours. We used an interpretable deep learning framework, Scailyte AG's in-house platform, ScaiVision to extract biologically meaningful features from scRNAseq data while maintaining model transparency. ScaiVision employs supervised representation learning to generate low-dimensional, yet biologically relevant, embeddings that distinguish BrM cells from their primary tumour counterparts. Through integrated gradients-based feature attribution analysis, our model identifies the individual genes that drive classification decisions, thus enabling direct biological interpretation of its output. This architecture allows for cross-tumour generalisation while preserving cell-type specific signals critical for the identification of metastasis-associated features. We further employed these genes to create a scoring system that uncovers cells responsible for brain tropism and analysed the plausible mechanism of action of these cells intrinsically as well as in coordination with the microenvironment. The resultant BrM signature was also detectable in liquid biopsies, specifically tumour-educated platelets, highlighting its potential as a non-invasive biomarker for early detection and risk stratification. More broadly, this strategy demonstrates how interpretable machine learning applied to single-cell data can uncover clinically actionable biomarkers and mechanistic insights into metastatic progression across cancer types, offering a blueprint for similar applications in other metastatic or treatment-resistant disease contexts.

Results

Modelling Brain Metastasis and Primary Cancers with Interpretable Neural Networks

Despite advancements in diagnostic techniques and therapeutic interventions, brain metastases (BrMs) remain associated with disproportionately high rates of morbidity and mortality. This clinical challenge underscores the urgent need to uncover robust molecular biomarkers that can facilitate early diagnosis, patient stratification, and development of targeted therapies across primary malignancies predisposed to cerebral dissemination. To address this unmet need, we designed a computational workflow centred on elucidating the transcriptional determinants of brain metastasis using single-cell transcriptomic profiles. Recognizing that conventional deep learning frameworks often function as "black boxes" and lack interpretability with respect to the biological relevance of features, we implemented a fully interpretable representation learning architecture tailored for single-cell classification. This framework enabled us not only to distinguish BrM samples from their primary tumour counterparts with high accuracy, but also to extract highly discriminative gene signatures that may define the metastatic phenotype at the single-cell level.

Our approach is grounded in the biological hypothesis that cells occupying transcriptionally similar states are likely to share functional and developmental programs, allowing us to infer the metastatic potential of individual cells based on gene expression similarities. Using this principle, we aimed to identify BrM-specific molecular programs and construct a brain metastasis signature enrichment score to evaluate their presence across individual cells (Figure 1A). We curated a dataset comprising single-cell RNA-sequencing (scRNA-seq) data from 115 patient samples, including 21 brain metastases and 94 primary tumours (Figure 1B, Supplementary Figure 1A). This dataset was divided into a training cohort (n=70; 13 BrM, 57 primary) and a validation cohort (n=45; 8 BrM, 37 primary) to support robust model development and independent performance assessment. To avoid bias from differences in cellular composition between primary and metastatic tumours, we restricted the model training to epithelial cells only (Figure 1C). Previous studies support the improved accuracy of epithelial cell-specific scRNA-seq signatures, emphasizing the importance of cell-type specificity in biomarker development [8].

To initiate model training (Figure 1D), we first selected highly variable genes (HVGs) from the gene expression matrix, reducing dimensionality while retaining key variance associated with biological signal. From this representation, representative cells were randomly sampled to ensure diversity and avoid sampling bias. Subsequently, the data were processed through a

series of convolutional filters, which scanned across features to learn localized transcriptional patterns indicative of BrM or primary origin. These filter activations were passed through a Rectified Linear Unit (ReLU) activation function to introduce non-linear transformations, followed by Top-K pooling, which retained the most informative features for classification optimizing model performance through extensive hyperparameter tuning[9]. The pooled features were then fed into a fully connected output layer with a SoftMax function, yielding class probabilities that distinguish BrM from primary samples. We trained a total of 300 deep learning models across six independent Monte Carlo cross-validation splits, each for up to 50 epochs, and retained the model weights corresponding to the epoch with the lowest validation loss to prevent overfitting. Evaluation metrics such as classification accuracy and cross-entropy log-loss demonstrated consistent performance and training stability across splits (Supplementary Figures 1B-D), attesting to the robustness and generalizability of our learning architecture.

For downstream biological interpretation, we selected models exhibiting high discriminatory performance, defined as Area Under the Receiver Operating Characteristic Curve (AUC) > 0.9 in the training set and AUC > 0.8 in the validation set (Figure 1E). Further evaluation was made including accuracy and cross-entropy log-loss. This stringent filtering yielded a final ensemble of 223 predictive models (Figure 1C; Supplementary Figures 1B-D), each capable of reliably identifying transcriptional features associated with metastatic potential. The high predictive accuracy of these models, coupled with their interpretability, lays a strong foundation for dissecting the molecular programs underpinning BrM.

Identification of Gene Signature with High Accuracy in Prediction of BrM from Epithelial cells

To uncover the molecular drivers underpinning model predictions, we performed a systematic feature attribution analysis using Integrated Gradients (IG), implemented via the Captum.ai interpretability library[10]. This approach enables a principled estimation of each input gene's contribution to the model's output by integrating gradients of the prediction probability with respect to the input features. We applied IG across our ensemble of high-performing deep learning models to generate an aggregated attribution matrix that prioritizes genes based on their relevance in distinguishing brain metastasis (BrM) from primary tumour cells. Through this aggregation, we derived a consensus ranking of genes based on their mean attribution scores, ultimately identifying 173 highly variable genes (HVGs) with consistently high explanatory power across models. The top 20 most influential genes from this list are visualized in Figure 2A, while the complete ranked gene set is provided in Supplementary

Table 2. We collectively refer to these genes as the BrM signature, representing a core set of transcriptional features that may be mechanistically or functionally linked to brain metastatic progression and therefore warrant further experimental validation and therapeutic exploration.

To assess the biological relevance and predictive utility of the BrM signature across individual cells, we computed a relative metastatic potential score for each epithelial cell, spanning both BrM and primary tumor origins, using the UCell scoring framework, a rank-based enrichment scoring method optimized for single-cell datasets[11] (Figure 2B). As anticipated, epithelial cells derived from histologically confirmed brain metastases exhibited consistently high metastatic potential scores, validating the sensitivity of our signature to known metastatic phenotypes (Figure 2C). Importantly, a subset of epithelial cells within primary tumour samples also received high metastatic potential scores, suggesting the presence of transcriptionally primed subpopulations with latent metastatic capabilities. This finding reinforces the notion that metastatic potential can be heterogeneous within primary tumours and highlights the potential of our signature to prospectively identify high-risk cell populations prior to dissemination. Moreover, the BrM signature demonstrated robust predictive capacity across multiple cancer types (Figure 2D), further supporting its cross-tumour applicability and biological generalizability. To further stratify by metastatic potential, cells were binned into high-scoring (top 20%), mid-scoring (20-80%), and low-scoring (bottom 20%) groups. Differential pseudobulk gene expression analysis comparing high- and low-scoring cells identified significant transcriptional differences associated with elevated metastatic potential (Figure 2E), including upregulation of pathways related to extracellular matrix remodelling, extracellular structural organization, synaptic signalling, and cell adhesion in cells with high metastatic scores (Figure 2E). These pathways are well-known contributors to metastatic processes, underscoring the biological validity of our signature.

To capture the dynamic progression from primary to metastatic states, we analysed paired scRNA-seq data from primary lung tumours and matched brain metastases from two patients[12], enabling direct assessment of our BrM signature across the metastatic continuum. Following standard preprocessing, BrM scores were computed for each cell using our established gene signature. As expected, scores were elevated in brain metastases, particularly within epithelial cells (Supplementary Figures 2A-B). To map the trajectory of metastatic progression, we employed pseudotime analysis using Monocle2[13]. Cells were embedded using DDRTree, revealing a pseudotemporal axis that strongly correlated with BrM scores, further validating the signature's ability to capture transcriptional progression from primary to metastatic states (Figure 2F). To pinpoint genes with dynamic regulation across this trajectory, we applied GeneSwitches R package[14], identifying key genes activated or

repressed over pseudotime. Genes such as ZBTB16, NFIA, NKD1, and SOX6 were switched on, while DOCK2, IGHA1, and MUC1 were switched off (Supplementary Figure 2C). These genes include transcription factors and surface proteins with potential roles in metastatic transition and niche adaptation.

Gene ontology analysis of these temporally regulated genes revealed enrichment of biological processes including cell proliferation, adhesion, lymphocyte activation, and immune development. Notably, immune-related pathways, such as B-cell receptor signalling and phagocytic engulfment, were selectively upregulated during later stages of pseudotime, implicating them in the final stages of metastatic establishment within the brain (Supplementary Figure 2D).

VEGF Signalling Correlates with BrM Score and Distinguishes Cell-Cell Networks in Primary Tumours and Brain Metastases

To broaden the scope of our analysis, we applied the BrM signature across the entire single-cell dataset, which includes a range of tumour and stromal cell types. This confirmed high BrM scores in cells from established brain metastases but also identified high-scoring cells within primary tumours, supporting the signature's potential to flag cells at risk of metastasis even before dissemination. High-scoring cells were found across multiple cancer types (Supplementary Figures 3A-C), and across diverse cell populations including monocytes, endothelial cells, neuroepithelial cells, and neutrophils. Following our established stratification (top 20%, middle 60%, bottom 20%), differential pseudobulk gene expression analysis revealed distinct transcriptional programs in high-scoring cells. These were enriched in pathways related to extracellular matrix remodelling, immune modulation, chemotaxis, and cell adhesion (Supplementary Figure 3D), processes closely tied to metastatic progression.

Recognizing the BrM signature's ability to identify metastatic potential across both tumour and stromal compartments, we next investigated cell-cell communication networks that might drive this phenotype. Tumour-stroma communication is well-documented as a key factor in cancer metastasis and treatment response, particularly within the complex microenvironment of the brain[15–17]. Using CellChat[18] to identify the differential number of interactions between high- vs low-scoring cells, we found that several signalling pathways were significantly upregulated in high-scoring cells, with ANGPT emerging as the most enriched. VEGF signalling was also markedly more active in high-scoring cells compared to low-scoring ones (Figure 3A). Endothelial cells were key players in this communication, particularly through the VEGFA-VEGFR1 ligand-receptor axis, known to be associated with metastatic progression in various cancers[19–21] (Supplementary Figures 3E). To validate these findings, we used LIANA (LIgand-

receptor ANAlysis tool[22]), a consensus-based tool aggregating multiple cell-cell interaction algorithms. LIANA analysis confirmed the dominant role of epithelial cells as senders and endothelial cells as receivers in high-scoring populations and again highlighted VEGFA-VEGFR1 as a central interaction (Figures 3B-C).

To determine whether communication patterns varied by tumour site, we separately analysed high-scoring cells from primary tumours and BrM samples. While VEGF signalling remained a consistent feature across both contexts, the structure of the interaction networks shifted. In primary tumours, epithelial cells played a more prominent signalling role, whereas dendritic cells were more influential within BrM sites (Figure 3B). These findings point to a dynamic reconfiguration of VEGF-related communication networks as metastases establish in the brain.

Finally, we explored the transcriptional dynamics of VEGF target genes during metastatic progression. Using literature-derived VEGF targets[23–25] and UCell scoring, we observed a gradual increase in VEGF target expression across cells binned by BrM score (in 10% increments), with expression peaking in the highest-scoring group (Figure 3D). This trend was supported by a strong correlation between VEGF signalling scores and BrM scores ($R = 0.7$, $p < 2.2e-16$) (Supplementary Figure 3F). These results underscore the importance of VEGF signalling not just as a marker, but potentially as a functional driver of metastatic transition. Together, our findings suggest that VEGF-mediated tumour-stroma interactions, particularly via VEGFA-VEGFR1, are critical in supporting the emergence and survival of metastatic cells in the brain.

Gene Regulatory Network Analysis Identifies ETS1 as a Putative Regulator of VEGF-driven Brain Metastasis

VEGF signalling has been consistently implicated in angiogenesis and metastatic progression, and our previous analyses suggest its involvement across diverse tumour lineages in brain metastasis. To investigate the regulatory control underpinning this VEGF-driven BrM phenotype, we examined transcriptional regulatory networks that may govern VEGF pathway activity during metastatic progression.

We analysed single-cell RNA-seq data from two patients with matched primary lung tumours and brain metastases, previously used in our pseudotime analyses. Cells were stratified into nine equal-sized bins according to their BrM signature score distribution, allowing us to trace regulatory changes across a continuum of BrM potential. Gene regulatory networks (GRNs) were independently inferred for each bin using CellOracle[26], enabling the identification of

transcription factors (TFs) with differential connectivity as a function of metastatic progression. In the highest BrM score group, where VEGF pathway gene expression was also most pronounced, TFs such as MYC, STAT1, MEF2C, and notably ETS1 demonstrated high degree centrality (Figure 4A), indicating their potential importance in coordinating pro-metastatic transcriptional programs. Among these, ETS1 emerged as a particularly compelling candidate due to its well-established role in regulating VEGF-mediated angiogenesis and tissue invasion. By comparing TF connectivity across bins, we observed a progressive increase in ETS1 regulatory influence, paralleling VEGF pathway activation and BrM score enrichment (Figure 4B).

We next focused on characterising ETS1 activity in the broader multi-cancer dataset to assess its relevance in a VEGF-driven BrM context. Although ETS1 transcript abundance was modest, consistent with known limitations of scRNA-seq in capturing TFs, its expression levels increased with higher BrM scores (Figure 4C). To better quantify its regulatory activity, we extracted ETS1 target genes from the inferred CellOracle networks and calculated a UCell enrichment score for each cell based on the expression of these downstream targets. This GRN-derived ETS1 activity score exhibited a clear upward trajectory across BrM score bins (Figure 4D), suggesting that ETS1 not only correlates with VEGF-driven metastatic potential, but may also act as a functional regulator of the transcriptional programs underlying this phenotype.

Together, these findings highlight ETS1 as a putative transcriptional regulator of VEGF-associated gene networks in brain metastasis. By integrating regulatory network inference with activity scoring and expression profiling, we reveal a layered regulatory architecture that underlies the VEGF-driven BrM phenotype. This approach allows us to sensitively detect key transcriptional drivers, even when expression alone is insufficient, and underscores the value of network-based strategies for dissecting the regulatory basis of metastasis to the brain.

Spatial Transcriptomics Reveal VEGF-Driven Microenvironments in BrM Regions

To expand our transcriptomic insights, we analysed spatial transcriptomics data derived from four patients diagnosed with colorectal brain metastases. This approach enabled us not only to validate the BrM signature directly within the tissue architecture[27] but also to investigate how metastatic potential varies across distinct microanatomical regions of the tumour and improve our understanding of metastasis biology[28].

We applied UCell scoring to map BrM signature activity across the entirety of each spatial transcriptomic slide. Remarkably, cells exhibiting high BrM signature scores were not confined solely to the tumour core; instead, they extended into peripheral tumour regions, including areas proximal to blood vessels and sites of tumour-associated inflammation (Figure 5A, Supplementary Figure 4A). This spatial distribution revealed pronounced heterogeneity in metastatic potential, underscoring the complex microenvironmental landscape that shapes tumour progression.

To further dissect molecular differences underpinning these variations, we performed differential gene expression analysis comparing high-scoring regions (top 20% BrM signature activity) to all other tumour areas. This analysis identified enrichment of pathways involved in mitochondrial function, endosomal transport, vesicle localization, and Ras signalling (Figure 5B). These findings highlight active cellular processes potentially driving or sustaining metastatic behaviour within these spatially distinct niches.

Given the well-established role of VEGF signalling in promoting brain metastasis, we next utilised CellChat to characterize spatial patterns of intercellular communication. VEGF-mediated interactions were consistently prominent across all patient samples, particularly between tumour cells and blood vessel-associated stromal regions (Figure 5C-D, Supplementary Figure 4B). Notably, these VEGF-enriched zones strongly overlapped with regions exhibiting elevated BrM signature scores (Figure 5A), reinforcing the link between VEGF signalling and enhanced metastatic capacity. To visualize these dynamics in situ, we projected VEGF pathway activity directly onto spatial maps of tumour sections. This projection revealed pronounced enrichment of VEGF signalling in tumour-adjacent vascular niches and inflammation-associated areas, microenvironments critical for tumour-stromal crosstalk and likely contributors to metastatic progression (Figure 5D, Supplementary Figure 4C).

Additionally, we examined the spatial distribution of the key VEGF signalling ligand-receptor pair VEGFA-VEGFR1, which we previously identified as in the pan-cancer scRNA-seq data. Spatial mapping revealed enrichment of VEGFA-VEGFR1 interactions specifically within high-scoring metastatic regions. This ligand-receptor interaction was predominantly localized in regions associated with active tumour growth and inflammation, indicating its likely role in promoting metastatic dissemination and tumour cell survival within the metastatic microenvironment (Figure 5E). This suggests that VEGFA-VEGFR1 signalling plays a pivotal role in supporting tumour cell survival, angiogenesis, and dissemination within the brain metastatic niche, as a central axis that facilitate metastatic progression[29].

Drug repurposing analysis identifies Pazopanib as a potential therapeutic candidate targeting VEGF-driven brain metastasis

To identify potential therapeutic compounds capable of reversing the transcriptional programs associated with high metastatic risk, specifically those linked to VEGF-driven brain metastasis, we employed the ASGARD R package[30] on our integrated multi-cancer single-cell RNA-seq dataset. Cells were stratified according to their BrM scores, with the top 20% defined as high-BrM and the bottom 20% as low-BrM. Differential gene expression analysis was then conducted using limma, comparing high- versus low-scoring cells to identify genes consistently up- or downregulated in the metastatic state.

The resulting differentially expressed genes served as input to ASGARD's drug repurposing pipeline, which mines the extensive L1000 drug-response dataset, comprising over 590,000 compound profiles, to identify small molecules capable of reversing the metastatic gene expression signature. Compounds were ranked based on their predicted efficacy in modulating BrM-associated gene expression, with candidates filtered by a false discovery rate (FDR) threshold of <0.05. Promising drug candidates with a predicted drug score >0.5 and FDR <0.05 were identified across a range of malignancies, including breast, colorectal, lung, nasopharyngeal carcinoma (NPC), ovarian, and pancreatic ductal adenocarcinoma (PDAC) (Figure 6A).

Among the top-ranking candidates was Pazopanib, a multi-kinase inhibitor with known clinical activity in metastatic cancer, and a well-established inhibitor of VEGF receptors[31]. Given our earlier findings implicating VEGF signalling as a key driver of the BrM phenotype, we further examined Pazopanib's mechanism of action across cancer types in our dataset. Using ASGARD, we extracted the predicted Pazopanib-responsive genes in high BrM-scoring cells and mapped these targets onto cell types and signalling pathways (Figure 6B). Pazopanib's regulatory footprint included modulation of VEGF-associated gene expression in epithelial tumour cells, endothelial compartments, and stromal fibroblasts, indicating potential to disrupt both tumour-intrinsic and microenvironmental drivers of metastasis. Further experimental validation is required to assess its efficacy in BrM-specific models, and to explore potential synergy with existing therapies. By directly targeting the transcriptional underpinnings of VEGF-associated metastatic programs, Pazopanib could serve as a precision therapeutic agent to mitigate the risk or progression of brain metastasis across multiple tumour types.

BrM Scoring Is Detectable in Tumour-Educated Platelets of BrM patients

Given the clinical need for non-invasive, real-time biomarkers to monitor metastatic disease progression, we sought to determine whether our brain metastasis (BrM) gene signature could

be detected in peripheral blood samples from cancer patients. Liquid biopsy approaches, such as blood-based transcriptomic profiling, offer several advantages over tissue biopsies, including minimal patient discomfort, feasibility for repeated sampling, and the potential for dynamic disease monitoring over time [32,33].

To investigate this, we leveraged a publicly available bulk RNA-sequencing dataset derived from tumour-educated platelets (TEPs) collected from patients representing seven distinct cancer types. This cohort included individuals with and without metastatic disease, allowing us to compare transcriptional profiles between these two clinical states. As a preliminary step, we focused on the vascular endothelial growth factor (VEGF) signalling pathway, owing to its well-established role in promoting angiogenesis and metastasis[34]. Analysis revealed that the mean mRNA expression of VEGF pathway genes was significantly elevated in patients with metastatic cancer compared to healthy controls ($p < 0.01$). However, when comparing patients with primary tumours to those with brain metastases, VEGF gene expression did not differ significantly (p = NS) (Figure 7A), indicating that VEGF, although elevated in general metastatic disease, lacks the specificity required to distinguish brain metastases from primary tumours.

We next assessed the performance of our 20-gene BrM signature in the same dataset. Remarkably, all 20 BrM genes were significantly upregulated in TEPs from patients with brain metastases compared to both healthy donors ($p < 0.01$ for each gene) and patients harbouring only primary tumours ($p < 0.05$ for each gene) (Figure 7B). Importantly, none of the BrM signature genes showed differential expression between the primary tumour and healthy groups (p = NS for all), underscoring the specificity of the signature for metastatic brain involvement.

These findings demonstrate that, unlike VEGF, the BrM signature yields a robust, metastasis-specific transcriptional signal in circulating platelets. This distinct molecular profile enables accurate discrimination between brain-metastatic and non-metastatic cases, thereby positioning the BrM signature as a promising candidate for a blood-based biomarker of metastatic progression in a clinically relevant liquid biopsy context. They underscore its translational potential for early detection, prognosis, and monitoring of metastatic brain disease in cancer patients, offering a powerful, non-invasive tool for guiding therapeutic decision-making and surveillance.

Discussion

Brain metastases (BrMs) represent a difficult clinical challenge in oncology. They are the most common intracranial tumours in adults and occur in an estimated 9-17% of all cancer patients[35] . Notably, BrMs arise most frequently from lung cancers, breast cancers, and melanomas, these three primaries collectively accounting for roughly 70% of all BrM cases[35]. The prognosis of patients with BrM is poor, and early detection is notoriously difficult because there are no routine brain metastasis screening protocols and current imaging modalities have limited sensitivity[36]. Consequently, many BrMs go undiagnosed until they become symptomatic, at which point therapeutic options are largely palliative. This clinical reality underlines the urgent need for improved strategies to predict and detect BrMs at a stage when intervention might substantially improve patient outcomes[37].

A central insight from our study is the power of a multi-cancer analytical framework to uncover shared mechanisms underlying brain metastasis (BrM), beyond tumour-specific constraints. Prior efforts to define metastasis-associated transcriptional programs have largely been restricted to individual tumour types, limiting their generalisability across malignancies[38,39]. In contrast, recent evidence suggests that metastatic lesions arising from distinct primaries can converge on common gene expression states[40,41]. Guided by this principle, we integrated scRNA-seq data from multiple tumour types to identify a BrM-associated gene signature that transcends tissue of origin[42]. We applied ScaiVision's supervised representation learning framework to train neural networks that classify BrM status at single-cell resolution. This approach learns tumour-agnostic low-dimensional representations that maximise separation between BrM and non-BrM cells, while preserving biological granularity.

Unlike conventional cluster-based or pseudobulk analyses, this method avoids averaging artefacts and enables interpretable models. Feature attribution using integrated gradients yielded a robust set of BrM-associated genes, ranked by their contribution to classification. Importantly, the resulting signature was reproducible across independent cohorts and computational strategies, supporting its generalisability. Collectively, these findings establish that a robust, cross-cancer BrM transcriptional programme can be derived from single-cell data, offering a framework for mechanistic dissection and translational development.

Scoring cells across our integrated dataset confirmed that the BrM signature is broadly conserved across tumour types and cellular compartments. Brain metastases from diverse primaries, lung, breast, renal, converged on a shared transcriptional programme, underscoring a common set of molecular adaptations required for brain colonisation. Despite distinct genetic

lineages, metastatic cells consistently upregulated genes linked to angiogenesis, extracellular matrix (ECM) remodelling, and intercellular signalling. Importantly, BrM signature genes extended beyond tumour-intrinsic expression. Many encode secreted ligands or receptors mediating tumour-microenvironment communication. Ligand-receptor interaction analysis revealed robust crosstalk between malignant and host brain cells, with pro-angiogenic signalling, particularly via the VEGF pathway, emerging as a recurrent axis. This interactional footprint was conserved across metastases, suggesting a generalised mechanism of stromal engagement. Notably, VEGF family members such as VEGFA and PlGF were consistently upregulated, implicating VEGF, VEGFR1 signalling in endothelial disruption[43] and blood-brain barrier (BBB) permeability[44,45], consistent with prior studies. These signals are known to trigger MMP-mediated ECM degradation and junctional disassembly, facilitating tumour extravasation into brain parenchyma. Our signature's enrichment for angiogenesis and ECM remodelling aligns closely with established mechanisms of metastasis[46], reinforcing its biological relevance and potential utility in identifying pan-cancer vulnerabilities in BrM.

To further validate and contextualise the BrM signature, we integrated spatial transcriptomics and external datasets. Spatial profiling of brain metastasis tissue confirmed that key signature genes exhibit precise localisation patterns consistent with tumour-host interaction niches. Angiogenic ligands such as VEGFA were enriched at the tumour-stroma interface, while their cognate receptors (e.g., VEGFR1) were expressed by adjacent endothelial and immune cells, providing spatial evidence for the ligand-receptor crosstalk inferred from single-cell data. This co-localisation underscores the in-situ relevance of the BrM programme and its role in shaping the metastatic niche. Cross-cohort validation using independent scRNAseq datasets from matched primary and brain metastatic samples demonstrated consistent upregulation of BrM signature genes in brain lesions across cancer types. Signature scores robustly distinguished metastases from their primary counterparts, confirming the signature's generalisability and specificity for the metastatic state.

To explore the BrM signature's clinical utility, we assessed its detectability in peripheral blood. Transcriptomic analysis of tumour-educated platelets (TEPs) revealed clear expression of signature genes in patients with brain metastases, but not in non-metastatic controls, supporting the notion that circulating platelets capture tumour-derived signals[47]. While VEGF-related transcripts were consistently enriched, they lacked sufficient specificity. By contrast, the broader BrM signature reliably distinguished brain metastases from primary tumours across cancer types, reflecting a distinct transcriptional programme associated with brain tropism. These observations highlight the potential of the BrM signature as a non-invasive biomarker. Its expression across diverse tumour origins and enrichment in matched metastatic

samples position it as a candidate for early detection and risk stratification. In particular, the ability to detect signature transcripts in TEPs offers a feasible path toward liquid biopsy-based diagnostics. Such an assay could complement imaging by identifying metastasis prior to radiographic detection and enabling real-time monitoring of disease activity behind the blood-brain barrier. Given the frequency of brain metastases in malignancies such as lung, breast, and melanoma, a platelet-based test could have broad translational relevance in guiding surveillance and treatment response.

Among high BrM-scoring cells, MYC and ETS1 emerged as dominant transcriptional regulators, highlighting their potential role in driving brain metastasis. Both factors are established oncogenic drivers, with c-Myc coordinating programmes spanning proliferation, metabolic reprogramming, immune modulation, and angiogenesis, key processes in metastatic progression[48]. ETS1, a master regulator of invasion and neo-angiogenesis, was especially prominent due to its direct activation of matrix-remodelling enzymes and VEGF signalling pathways[49]. Notably, ETS1 activity increased in parallel with BrM scores, implicating it in the gradual acquisition of metastatic traits. Elevated ETS1 activity suggests a central role in preparing the metastatic niche, particularly through ECM remodelling and vascular co-option—mechanisms critical for colonisation of the brain. ETS1 has also been linked to the induction of L1CAM, a neural adhesion molecule that facilitates tumour cell adaptation to the brain microenvironment[50], further reinforcing its contribution to brain-tropic behaviour.

The functional relevance of this axis is underscored by our drug repurposing analysis, which identified pazopanib, a VEGF receptor inhibitor with brain-penetrant properties, as a candidate therapeutic. Pazopanib has shown intracranial activity in clinical settings, including regression of brain metastases in patients resistant to prior therapies[51]. Similar agents, such as sunitinib, have demonstrated efficacy in lung cancer with brain metastasis[52], supporting the premise that BrM-high cells depend on angiogenic signalling for survival and expansion. These findings point to the feasibility of anti-angiogenic strategies in curbing metastatic outgrowth within the brain.

In conclusion, our study delineates a transcriptional programme underpinning brain metastasis, anchored by regulators such as MYC and ETS1, and captured by a robust BrM gene signature. This work provides mechanistic insight into metastatic adaptation at the brain interface and presents actionable avenues for diagnosis and treatment. Further functional studies and prospective validation will be essential to translate these insights into clinical tools for early detection and targeted intervention, ultimately improving outcomes for patients at risk of or living with brain metastases.

Methods

Data Acquisition and Processing

Publicly available single-cell RNA sequencing (scRNA-seq), spatial transcriptomics (ST), and bulk RNA-sequencing datasets were utilized. scRNA-seq data, primarily generated using 10x Genomics technology, were aggregated from sources detailed in Supplementary Table 1, including paired primary lung cancer and brain metastasis (BrM) samples from GSE223503. ST data were obtained from GEO accession GSE179572, and bulk RNA-seq data from GSE68086. Raw UMI count matrices and associated metadata from public scRNA-seq sources were aggregated. All data are publicly available from the cited sources.

scRNA-seq Analysis

Analyses were performed in R (v4.1.1) using Seurat. Cells were retained if they expressed >200 genes and <20% mitochondrial reads (PercentageFeatureSet, prefix='^MT-') Potential doublets were identified and removed using scDblFinder (v1.6.0, default parameters) Data were normalized using log-transformation (NormalizeData, method='LogNormalize', scale.factor=10,000). The top 2,000 highly variable genes (HVGs) were identified (FindVariableFeatures, method='vst') and scaled (ScaleData). Principal component analysis (PCA) was performed on HVGs (RunPCA), and the first 30 PCs were used for downstream analyses, including K-nearest neighbour graph construction (FindNeighbors) and Louvain clustering (FindClusters, resolution=0.2). Uniform Manifold Approximation and Projection (UMAP) was used for visualization (RunUMAP, input=30 PCs)

Cell Type Annotation

Cell types were annotated using SingleR, with the Human Primary Cell Atlas reference from celldex.

ScaiVision

In the first step, ScaiVision uses data augmentation through subsampling, which increases the number of usable training samples. This data is structured into multi-cell inputs, with genes on the horizontal axis and cells on the vertical axis. Next, ScaiVision initialises filter weights randomly and computes response values for each cell-filter pair via dot products. These filters serve as connections linking gene expression values to outcome predictions (Primary or BrM), enabling automatic feature selection without the need for upfront cell clustering, ensuring the approach remains unbiased and agnostic to cell types and features. In the third step, ScaiVision aggregates cell responses for whole-sample predictions, leveraging top-k pooling, a flexible method that calculates mean values over a defined top percentage of cells, allowing sensitive detection of BrM or primary cells. Subsequently, filter values are combined in the output layer through matrix multiplication, generating probabilities for each output class.

Lastly, in network training, the network undergoes training across multiple epochs using a loss function that compares predicted and actual labels, with backpropagation and parameter updates reducing loss in each batch until a specified end condition is met, yielding a trained network with high accuracy in predicting BrM or primary.

Gene Set Analysis

Feature importance scores were calculated using Integrated Gradients within the Captum framework, defining a "BrM signature". This signature, along with other gene sets, was scored per cell using UCell, which employs a rank-based Mann-Whitney U statistic[11]. For certain analyses, cells were categorized based on their metastatic score relative to other cells within the BrM samples: 'high-scoring' (top 20%) and 'low-scoring' (bottom 20%). Additionally, scores were binned into deciles (10% intervals) for visualization and analysis of score distributions.

Trajectory and Gene Switch Analysis

Pseudotime trajectory analysis was performed on paired primary lung and BrM scRNA-seq data (GSE223503) using Monocle 2 to model metastatic progression. A CellDataSet object was created (expressionFamily=negbinomial.size()). Feature selection utilized HVGs identified in Seurat. Dimensionality was reduced using DDRTree (reduceDimension, max_components=10), and cells were ordered in pseudotime (orderCells), defining primary lung tumour cells as the root state. Branch-dependent gene expression analysis was performed using GeneSwitches[14], identifying genes significantly associated with the high BrM branch.

Spatial Transcriptomics Analysis

ST data generated using the Visium platform from 10X Genomics (GSE179572) were processed using Seurat. Data were normalized using SCTransform. Spatially differential expression between pre-annotated regions was assessed using FindMarkers (Wilcoxon test, Benjamini-Hochberg FDR < 0.05). The metastatic score signature was mapped onto ST spots using UCell.

Cell-Cell Interaction Analysis

Potential ligand-receptor interactions were inferred from normalized scRNA-seq data using CellChat[18] with the CellChatDB.human database. Cells were grouped based on cell type annotation and metastatic potential score (top 20% defined as 'high', bottom 20% defined as 'low' within BrM samples). CellChat analysis was performed using default parameters to identify significant ligand-receptor pairs and communication pathways. Interaction patterns, including the number and strength of interactions, were compared between high-scoring and low-scoring metastatic cells interacting with other cell types (e.g., stromal, immune) to identify

differential communication networks potentially driving metastasis. Results were visualized using netVisual_circle and other relevant functions.

Gene Regulatory Network (GRN) Analysis

GRNs were inferred using CellOracle[26] following standard documentation (https://morris-lab.github.io/CellOracle.documentation/). The workflow involved GRN construction based on TF binding motifs (JASPAR database) within putative regulatory regions (hg38), integration with normalized scRNA-seq data, and network analysis/simulation.

Bulk RNA-seq Analysis

Bulk RNA-seq reads (GSE68086) were aligned to hg38 using STAR. Gene-level counts were quantified using featureCounts. Normalisation was performed using DESeq2[53] (Wald test, Benjamini-Hochberg FDR < 0.05).

Drug Repurposing Screen

In silico drug repurposing was performed using ASGARD[30] with default parameters. Normalized scRNA-seq counts from cells grouped by metastatic score (low score = control) served as input, comparing against the LINCS L1000 drug perturbation database. Drug scores were calculated, identifying candidates with significant scores across relevant comparisons.

Statistical Analysis

Analyses were performed in R (v4.3.1). Comparisons between groups were made using the Wilcoxon rank-sum test. P-values from multiple tests were adjusted using the Benjamini-Hochberg method (FDR < 0.05) A p-value < 0.05 was considered significant for other tests.

Figure Legends

Figure 1 | Identification and validation of a brain metastasis (BrM) signature in epithelial cells.

(A) Workflow overview illustrating the strategy for identifying the BrM signature from single-cell RNA sequencing (scRNA-seq) data across multiple cancer types. **(B)** UMAP plot illustrating full multi-cancer dataset prior to subsetting and model training. Cells are coloured by label, cancer type and cell type. **(C)** Representative UMAP plots demonstrating epithelial cell selection for downstream analysis. Cells are coloured according to their cancer type origin from primary tumour or BrM samples. **(D)** Detailed schematic of the ScaiVision network architecture used for model training. **(E)** Scatter plots summarising Area Under the Curve (AUC) scores for the predictive models across training and validation datasets. The threshold (AUC > 0.9 and Validation AUC > 0.8) used for selecting high-performing models is indicated by a dashed line.

Figure 2 | Prediction of differential features of BrM using Machine Learning. (A) Chart showing the top 20 genes ranked by Integrated Gradients attribution scores, highlighting their contribution to model predictions. These genes represent the core components of the identified BrM signature (complete list provided in Supplementary Table 2). Points show the mean scores and error bars indicate the standard deviation of scores across models (n =223) **(B)** UMAP projection of epithelial cells, coloured according to BrM signature scores calculated using UCell. Higher scores indicate greater inferred metastatic potential, highlighting cell populations from confirmed BrM samples and subsets within primary tumours. **(C)** Violin plots comparing BrM signature scores (UCell) for epithelial cells between primary and BrM samples. **(D)** Violin plots comparing BrM signature scores (UCell) for epithelial cells among cancer types. **(E)** Gene Ontology (GO) enrichment analysis illustrating biological processes significantly enriched in high-scoring epithelial cells (top 20%) compared to low-scoring epithelial cells (bottom 20%), derived from differential pseudobulk gene expression analysis. Dot size indicates gene count, and colour gradient represents adjusted p-values. Notably enriched pathways include extracellular matrix organisation, cell-cell adhesion, and synapse-related processes linked to metastatic capability. **(F)** Monocle 2 trajectory plot showing inferred pseudotime progression for cells from paired primary lung and BrM samples. Cells are coloured by sample origin, BrM signature score (UCell) and inferred pseudotime.

Figure 3 | BrM signature scoring and cell-cell communication network analysis across the multi-cancer dataset.

(A) CellChat pathway analysis comparing signalling pathway activity between high-scoring and low-scoring cells. Bar size represents relative pathway strength between the two conditions; colour indicates activity in high or low BrM scored cells (red=high, blue=low). Significant pathways are coloured on the y-axis **(B)** CellChat network heatmaps comparing the architecture of VEGF signalling pathways specifically within high-scoring cells derived from primary tumour sites and BrM sites. **(C)** Plot showing significantly enhanced activity scores for VEGF ligand-receptor pairs, particularly VEGFA-VEGFR1, in high-scoring cells compared to low-scoring cells, based on LIANA analysis. **(D)** Box plot showing UCell scores for a literature-derived VEGF signalling target gene set across cells binned into deciles based on their BrM signature score.

Figure 4 | Gene regulatory network analysis implicates VEGF signalling drivers

(A) Dot plot illustrating transcription factor (TF) degree centrality within the gene regulatory network (GRN) inferred by CellOracle for the highest BrM score bin from paired primary/BrM samples. **(B)** Line plots showing the change in network connectivity (degree centrality) for key TFs (MYC, STAT1, ETS1) across nine BrM score bins. **(C)** Violin plot showing ETS1 gene expression distribution across BrM score bins in the multi-cancer scRNA-seq dataset. **(D)** Box plots showing UCell scores for ETS1 target genes (derived from CellOracle GRN) across BrM score bins in the multi-cancer scRNA-seq dataset.

Figure 5 | Spatial transcriptomics analysis reveals spatial organization of BrM potential and VEGF signalling.

(A) Spatial feature plots showing BrM signature scores (UCell) mapped onto tissue section of a colorectal cancer BrM patient. Colour scale indicates BrM score. Key anatomical regions are annotated. **(B)** Gene Ontology (GO) enrichment analysis results for biological processes upregulated in high-scoring regions (top 20% BrM score) compared to other regions, based on spatially-resolved differential gene expression analysis. Dot size represents gene count; colour indicates significance level. **(C)** CellChat VEGF network plots projecting inferred VEGF pathway activity scores onto tissue sections. Arrow direction indicates signalling flow; line width indicates interaction strength. **(D)** CellChat network heatmap plot summarizing inferred VEGF signalling interactions between annotated tissue regions. **(E)** Spatial feature plots showing the enrichment score for the VEGFA-VEGFR1 ligand-receptor interaction pair mapped onto tissue sections. Colour scale indicates expression of each ligand.

Figure 6 | Pazopanib as a repurposing candidate for targeting VEGF signaling in brain metastasis.

(A) Bar plots summarizing ASGARD drug repurposing results. Y-axis indicates the predicted drug score (reversal potential). **(B)** Gene ontology enrichment of predicted Pazopanib target genes (identified by ASGARD) within high BrM-scoring cells across different cell types.

Figure 7 | BrM signature aligns with metastatic progression trajectory and is detectable in tumour-educated platelets (TEPs).

(A) Box plots comparing the expression levels of VEGF pathway genes in bulk RNA-seq data from TEPs of metastatic patients, primary tumour-only patients, and healthy controls across seven cancer types. Significance determined by t-test (* $p < 0.05$, ** $p < 0.01$, ns = not significant). **(B)** Box plots comparing the expression levels of the 20 BrM signature genes in bulk RNA-seq data from TEPs across the same patient groups as in (D). Significance determined by t-test (* $p < 0.05$, ** $p < 0.01$, ns = not significant).

Legends for supplementary figures

Supplementary Figure 1 | Model training and BrM signature performance in epithelial cells.

(C) Area under the ROC curve for training and validation samples across all folds, indicating the performance of models on training and validation data. Each point represents a separate model with its own set of hyperparameters. **(D)** Model accuracy for training and validation samples across all folds at a classification threshold of 0.5. Each point represents a separate model with its own set of hyperparameters. **(E)** Cross-entropy log-loss for validation samples across all folds. Each point represents a separate model with its own set of hyperparameters. **(F)** Violin plots comparing BrM signature scores (UCell) for epithelial cells between cancer types.

Supplementary Figure 2 | Extended BrM signature and pseudotime-based enrichment analysis.

(A) UMAP plots of scRNA-seq data from paired primary lung and BrM samples, coloured by endpoint label, original patient ID and BrM signature score (UCell). **(B)** Violin plot highlighting higher BrM scores in BrM samples **(C)** GeneSwitches plot identifying genes significantly activated ('on', positive R2) or inactivated ('off', negative R2) along the pseudotime trajectory. Key genes are labelled. Colour indicates annotation of gene type **(D)** Gene Ontology (GO) enrichment analysis results for genes dynamically regulated across pseudotime.

Supplementary Figure 3 | Extended cell-cell communication analysis.

(A) UMAP plots showing BrM signature scores (UCell) projected onto all cells across multiple cancer

types. Colour scale indicates BrM score. **(B)** Violin plots comparing BrM signature scores (UCell) between cancer types and endpoint labels in the multi-cancer full scRNA-seq dataset. **(C)** Violin plots comparing BrM signature scores (UCell) between cell types in the multi-cancer full scRNA-seq dataset. (D) Gene Ontology (GO) enrichment analysis results for biological processes upregulated in high-scoring (top 20%) compared to low-scoring (bottom 20%) cells across all cell types, based on differential pseudobulk gene expression analysis. Dot size length represents gene count; colour indicates significance level. **(E)** CellChat VEGF network visualization of the consensus cell types between low and high scored cells and ligand interactions in the consensus cell types **(F)** Scatter plot illustrating the cell-wise correlation between BrM signature scores (UCell) and VEGF target gene set scores (UCell). Correlation coefficient (R) and p-value are indicated.

Supplementary Figure 4 | Additional spatial transcriptomics results.
(A) Spatial feature plots showing BrM signature scores (UCell) mapped onto additional tissue sections or regions from the colorectal cancer BrM patients. **(B)** Additional CellChat network heatmap plots summarizing inferred VEGF signalling interactions between annotated tissue regions. **(C)** Additional CellChat spatial interaction plots detailing VEGF signalling interactions between annotated tissue regions in specific samples.

DECLARATIONS

Ethical Approval
Not applicable

Data Access
All the datasets used for this study are publicly available.

Authors' contributions
R. L. and V.K.T. designed the study, analyzed data, and wrote the manuscript. D.C. provided critical edits to the manuscript. S.C. co-supervised analysis and wrote the manuscript.

Funding
This study was supported by the Department for the Economy (DfE), BBSRC Innovation to Commercialisation of University Research (ICURe)(Project Reference: Mid-H-13), Novo

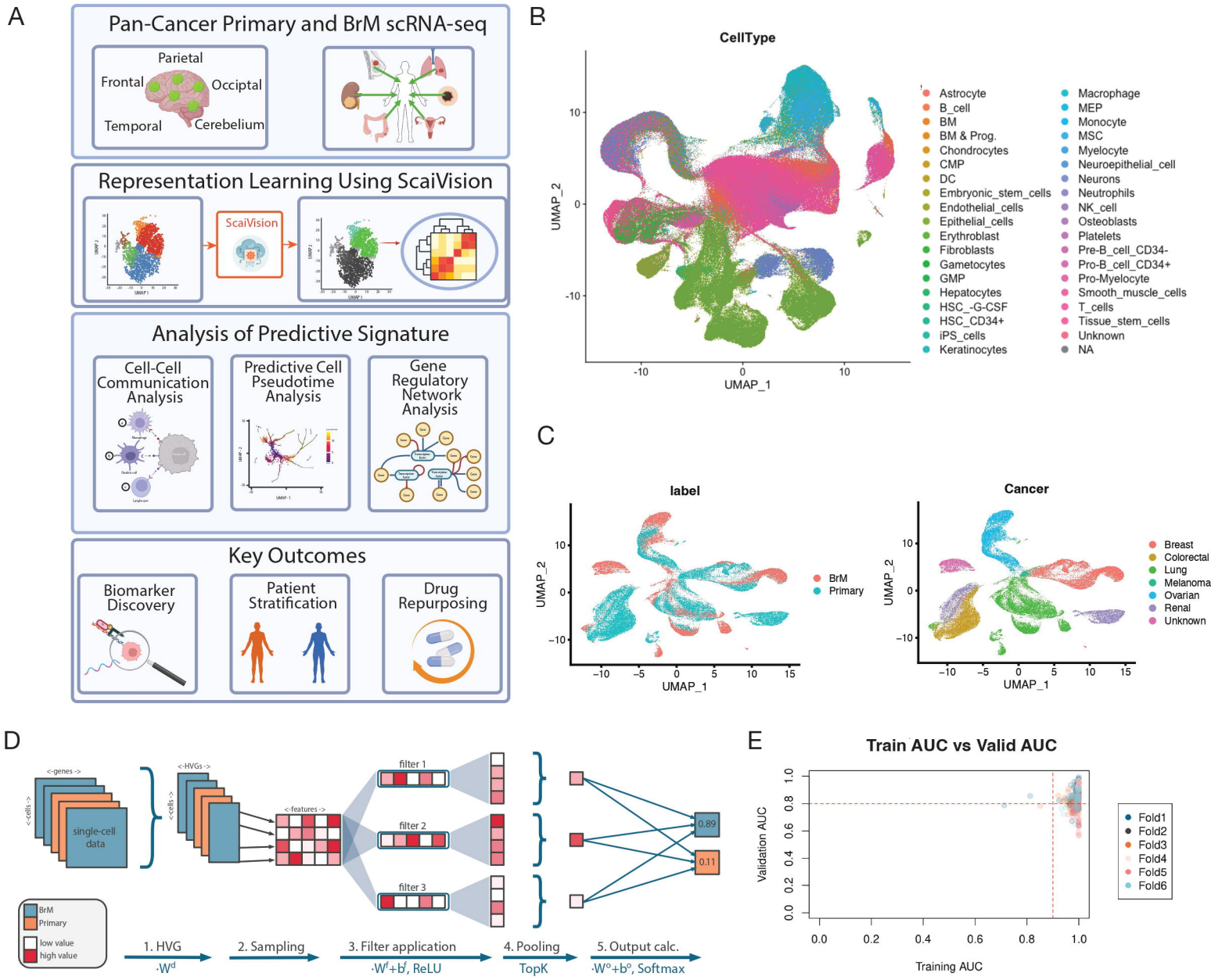
Nordisk Foundation 3110103, Danish National Research Foundation DNRF177 and Danish Cancer Society R374-A22449 grants to VKT.

References

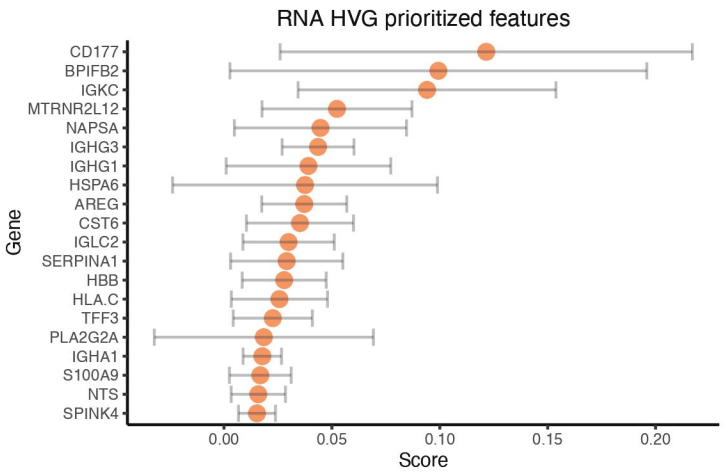
1. Achrol, A. S. *et al.* Brain metastases. *Nat. Rev. Dis. Primer* **5**, 5 (2019).
2. Wang, Z. *et al.* Single-cell transcriptomic analyses provide insights into the cellular origins and drivers of brain metastasis from lung adenocarcinoma. *Neuro-Oncol.* **25**, 1262–1274 (2023).
3. Tabouret, E. *et al.* Recent trends in epidemiology of brain metastases: an overview. *Anticancer Res.* **32**, 4655–62 (2012).
4. Steindl, A., Brastianos, P. K., Preusser, M. & Berghoff, A. S. Precision medicine biomarkers in brain metastases: applications, discordances, and obstacles. *Neuro-Oncol. Adv.* **3**, v35–v42 (2021).
5. Song, Q. *et al.* Single-cell sequencing reveals the landscape of the human brain metastatic microenvironment. *Commun. Biol.* **6**, 760 (2023).
6. Bach, K. *et al.* Time-resolved single-cell analysis of Brca1 associated mammary tumourigenesis reveals aberrant differentiation of luminal progenitors. *Nat. Commun.* **12**, 1502 (2021).
7. Zou, Y. *et al.* Single-cell atlas profiling revealed cellular characteristics and dynamic changes after PD-1 blockade therapy of brain metastases from laryngeal squamous cell carcinoma. *Mol. Cell. Biochem.* **480**, 2377–2400 (2025).
8. Fisher, N. C. *et al.* Biological Misinterpretation of Transcriptional Signatures in Tumor Samples Can Unknowingly Undermine Mechanistic Understanding and Faithful Alignment with Preclinical Data. *Clin. Cancer Res.* **28**, 4056–4069 (2022).
9. Arvaniti, E. & Claassen, M. Sensitive detection of rare disease-associated cell subsets via representation learning. *Nat. Commun.* **8**, 14825 (2017).
10. Kokhlikyan, N. *et al.* Captum: A unified and generic model interpretability library for PyTorch. (2020) doi:https://doi.org/10.48550/arXiv.2009.07896.
11. Andreatta, M. & Carmona, S. J. UCell: Robust and scalable single-cell gene signature scoring. *Comput. Struct. Biotechnol. J.* **19**, 3796–3798 (2021).
12. Tagore, S. *et al.* Single-cell and spatial genomic landscape of non-small cell lung cancer brain metastases. *Nat. Med.* **31**, 1351–1363 (2025).
13. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
14. Cao, E. Y., Ouyang, J. F. & Rackham, O. J. L. GeneSwitches: ordering gene expression and functional events in single-cell experiments. *Bioinformatics* **36**, 3273–3275 (2020).
15. Bergfeld, S. A. & DeClerck, Y. A. Bone marrow-derived mesenchymal stem cells and the tumor microenvironment. *Cancer Metastasis Rev.* **29**, 249–261 (2010).
16. Bussard, K. M., Mutkus, L., Stumpf, K., Gomez-Manzano, C. & Marini, F. C. Tumor-associated stromal cells as key contributors to the tumor microenvironment. *Breast Cancer Res.* **18**, 84 (2016).
17. Khan, I. M. *et al.* TME-targeted approaches of brain metastases and its clinical therapeutic evidence. *Front. Immunol.* **14**, (2023).
18. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
19. Kim, L. S., Huang, S., Lu, W., Chelouche Lev, D. & Price, J. E. Vascular endothelial growth factor expression promotes the growth of breast cancer brain metastases in nude mice. *Clin. Exp. Metastasis* **21**, 107–118 (2004).
20. Yano, S. *et al.* Expression of vascular endothelial growth factor is necessary but not sufficient for production and growth of brain metastasis. *Cancer Res.* **60**, 4959–67 (2000).

21. Claffey, K. P. *et al.* Expression of vascular permeability factor/vascular endothelial growth factor by melanoma cells increases tumor growth, angiogenesis, and experimental metastasis. *Cancer Res.* **56**, 172–81 (1996).
22. Dimitrov, D. *et al.* Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat. Commun.* **13**, 3224 (2022).
23. Matsumoto, T. & Claesson-Welsh, L. VEGF Receptor Signal Transduction. *Sci. STKE* **2001**, (2001).
24. Cross, M. J., Dixelius, J., Matsumoto, T. & Claesson-Welsh, L. VEGF-receptor signal transduction. *Trends Biochem. Sci.* **28**, 488–494 (2003).
25. Ferrara, N., Gerber, H.-P. & LeCouter, J. The biology of VEGF and its receptors. *Nat. Med.* **9**, 669–676 (2003).
26. Kamimoto, K. *et al.* Dissecting cell identity via network inference and in silico gene perturbation. *Nature* **614**, 742–751 (2023).
27. Cheng, M. *et al.* Spatially resolved transcriptomics: a comprehensive review of their technological advances, applications, and challenges. *J. Genet. Genomics* **50**, 625–640 (2023).
28. Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).
29. Ceci, C., Atzori, M. G., Lacal, P. M. & Graziani, G. Role of VEGFs/VEGFR-1 Signaling and Its Inhibition in Modulating Tumor Invasion: Experimental Evidence in Different Metastatic Cancer Models. *Int. J. Mol. Sci.* **21**, 1388 (2020).
30. He, B. *et al.* ASGARD is A Single-cell Guided Pipeline to Aid Repurposing of Drugs. *Nat. Commun.* **14**, 993 (2023).
31. Jacobs, C., Kim, D. W. N., Straka, C., Timmerman, R. D. & Brugarolas, J. Prolonged Survival of a Patient With Papillary Renal Cell Carcinoma and Brain Metastases Using Pazopanib. *J. Clin. Oncol.* **31**, e114–e117 (2013).
32. Marrugo-Ramírez, J., Mir, M. & Samitier, J. Blood-Based Cancer Biomarkers in Liquid Biopsy: A Promising Non-Invasive Alternative to Tissue Biopsy. *Int. J. Mol. Sci.* **19**, 2877 (2018).
33. Uttley, L. *et al.* Building the Evidence Base of Blood-Based Biomarkers for Early Detection of Cancer: A Rapid Systematic Mapping Review. *eBioMedicine* **10**, 164–173 (2016).
34. Brown, A. M., Nock, S., Musgrave, K. & Unsworth, A. J. Vascular Endothelial Growth Factor (VEGF) as a Biomarker for Cancer-Associated Venous Thrombosis: A Meta-analysis. *TH Open* **09**, (2025).
35. Tabor, J. K. *et al.* Brain metastasis screening in the molecular age. *Neuro-Oncol. Adv.* **5**, (2023).
36. Brenner, A. W. & Patel, A. J. Review of Current Principles of the Diagnosis and Management of Brain Metastases. *Front. Oncol.* **12**, (2022).
37. Nayak, L., Lee, E. Q. & Wen, P. Y. Epidemiology of Brain Metastases. *Curr. Oncol. Rep.* **14**, 48–54 (2012).
38. Schell, M. J. *et al.* A Composite Gene Expression Signature Optimizes Prediction of Colorectal Cancer Metastasis and Outcome. *Clin. Cancer Res.* **22**, 734–745 (2016).
39. Fan, C. *et al.* Concordance among Gene-Expression–Based Predictors for Breast Cancer. *N. Engl. J. Med.* **355**, 560–569 (2006).
40. Hakobyan, A. *et al.* Pan-cancer analysis of the interplay between mutational signatures and cellular signaling. *iScience* **27**, 109873 (2024).
41. Xu, Q. *et al.* Pan-cancer transcriptome analysis reveals a gene expression signature for the identification of tumor tissue origin. *Mod. Pathol.* **29**, 546–556 (2016).
42. Xing, X. *et al.* Pan-cancer human brain metastases atlas at single-cell resolution. *Cancer Cell* (2025) doi:10.1016/j.ccell.2025.03.025.
43. Wu, Y. *et al.* Single-cell RNA sequencing reveals epithelial cells driving brain metastasis in lung adenocarcinoma. *iScience* **27**, 109258 (2024).
44. Li, B. *et al.* Elevated PLGF contributes to small-cell lung cancer brain metastasis. *Oncogene* **32**, 2952–2962 (2013).

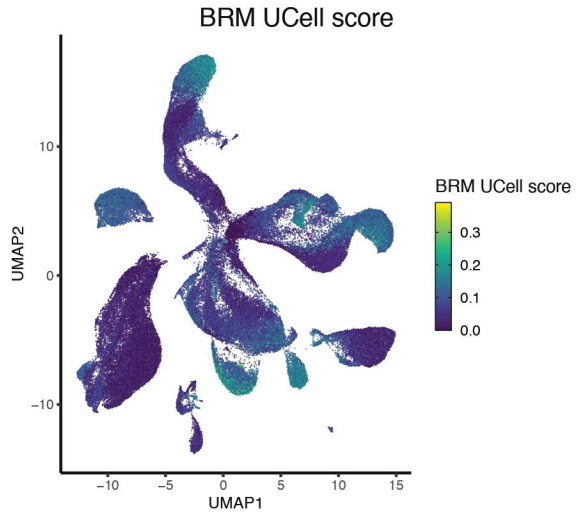
45. Lyle, L. T. *et al.* Alterations in Pericyte Subpopulations Are Associated with Elevated Blood–Tumor Barrier Permeability in Experimental Brain Metastasis of Breast Cancer. *Clin. Cancer Res.* **22**, 5287–5299 (2016).
46. Yuzhalin, A. E. & Yu, D. Critical functions of extracellular matrix in brain metastasis seeding. *Cell. Mol. Life Sci.* **80**, 297 (2023).
47. Best, M. G. *et al.* RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics. *Cancer Cell* **28**, 666–676 (2015).
48. Meškytė, E. M., Keskas, S. & Ciribilli, Y. MYC as a Multifaceted Regulator of Tumor Microenvironment Leading to Metastasis. *Int. J. Mol. Sci.* **21**, 7710 (2020).
49. Dittmer, J. The Biology of the Ets1 Proto-Oncogene. *Mol. Cancer* **2**, 29 (2003).
50. Feng, X., Guan, N., Xu, E., Miao, Y. & Li, C. Transcription Factors Leading to High Expression of Neuropeptide L1CAM in Brain Metastases from Lung Adenocarcinoma and Clinical Prognostic Analysis. *Dis. Markers* **2021**, 1–15 (2021).
51. Hingorani, M., Dixit, S. & Maraveyas, A. Pazopanib-Induced Regression of Brain Metastasis After Whole Brain Palliative Radiotherapy in Metastatic Renal Cell Cancer Progressing on First-Line Sunitinib: A Case Report. *World J. Oncol.* **5**, 223–227 (2014).
52. Novello, S. *et al.* Phase II Study of Sunitinib in Patients with Non-small Cell Lung Cancer and Irradiated Brain Metastases. *J. Thorac. Oncol.* **6**, 1260–1266 (2011).
53. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).


A
Pan-Cancer Primary and BrM scRNA-seq
Parietal
Frontal
Occipital
Temporal
Cerebelium
Representation Learning Using ScaiVision
ScaiVision
Analysis of Predictive Signature
Cell-Cell Communication Analysis
Predictive Cell Pseudotime Analysis
Gene Regulatory Network Analysis
Key Outcomes
Biomarker Discovery
Patient Stratification
Drug Repurposing
B
CellType
UMAP_2
UMAP_1
Astrocyte
B_cell
BM
BM & Prog.
Chondrocytes
CMP
DC
Embryonic_stem_cells
Endothelial_cells
Epithelial_cells
Erythroblast
Fibroblasts
Gametocytes
GMP
Hepatocytes
HSC_-G-CSF
HSC_CD34+
iPS_cells
Keratinocytes
Macrophage
MEP
Monocyte
MSC
Myelocyte
Neuroepithelial_cell
Neurons
Neutrophils
NK_cell
Osteoblasts
Platelets
Pre-B_cell_CD34-
Pro-B_cell_CD34+
Pro-Myelocyte
Smooth_muscle_cells
T_cells
Tissue_stem_cells
Unknown
NA
C
label
BrM
Primary
Cancer
Breast
Colorectal
Lung
Melanoma
Ovarian
Renal
Unknown
D
<-genes->
<-cells->
single-cell data
<-HVGs->
<-features->
filter 1
filter 2
filter 3
0.89
0.11
BrM
Primary
low value
high value
1. HVG
2. Sampling
3. Filter application
4. Pooling
TopK
5. Output calc.
Softmax
E
Train AUC vs Valid AUC
Validation AUC
Training AUC
Fold1
Fold2
Fold3
Fold4
Fold5
Fold6

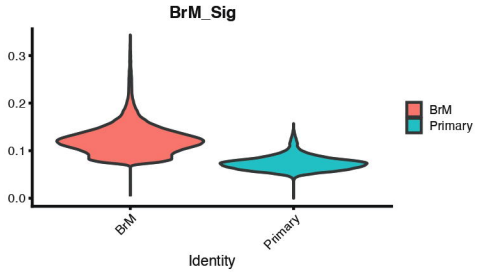
A


RNA HVG prioritized features
CD177
BPIFB2
IGKC
MTRNR2L12
NAPSA
IGHG3
IGHG1
HSPA6
AREG
CST6
IGLC2
SERPINA1
HBB
HLA.C
TFF3
PLA2G2A
IGHA1
S100A9
NTS
SPINK4
Gene
0.00
0.05
0.10
0.15
0.20
Score

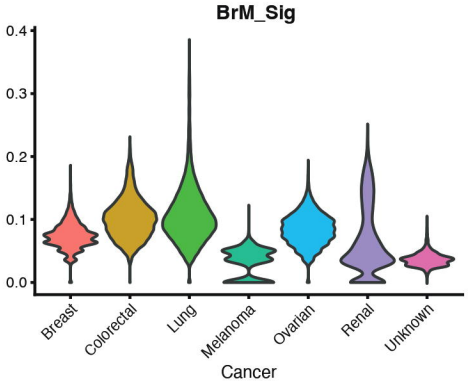
B


BRM UCell score
UMAP2
UMAP1
BRM UCell score
0.3
0.2
0.1
0.0

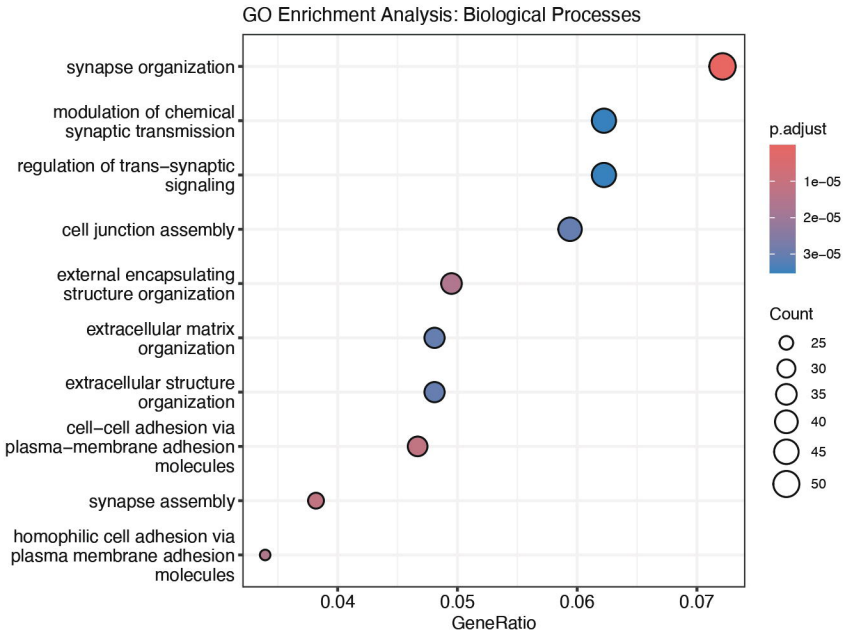
C


BrM_Sig
BrM
Primary
Identity

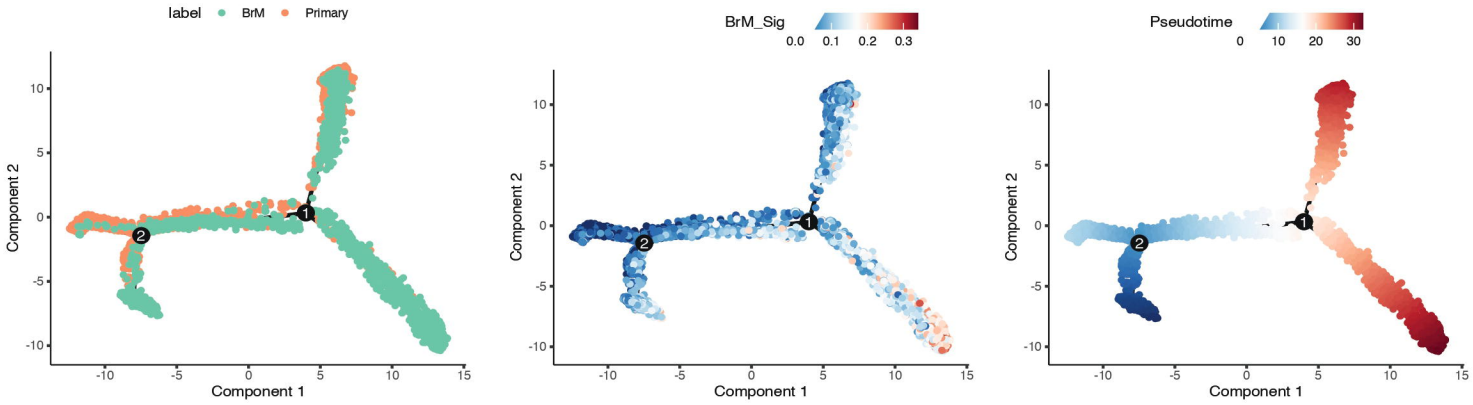
D

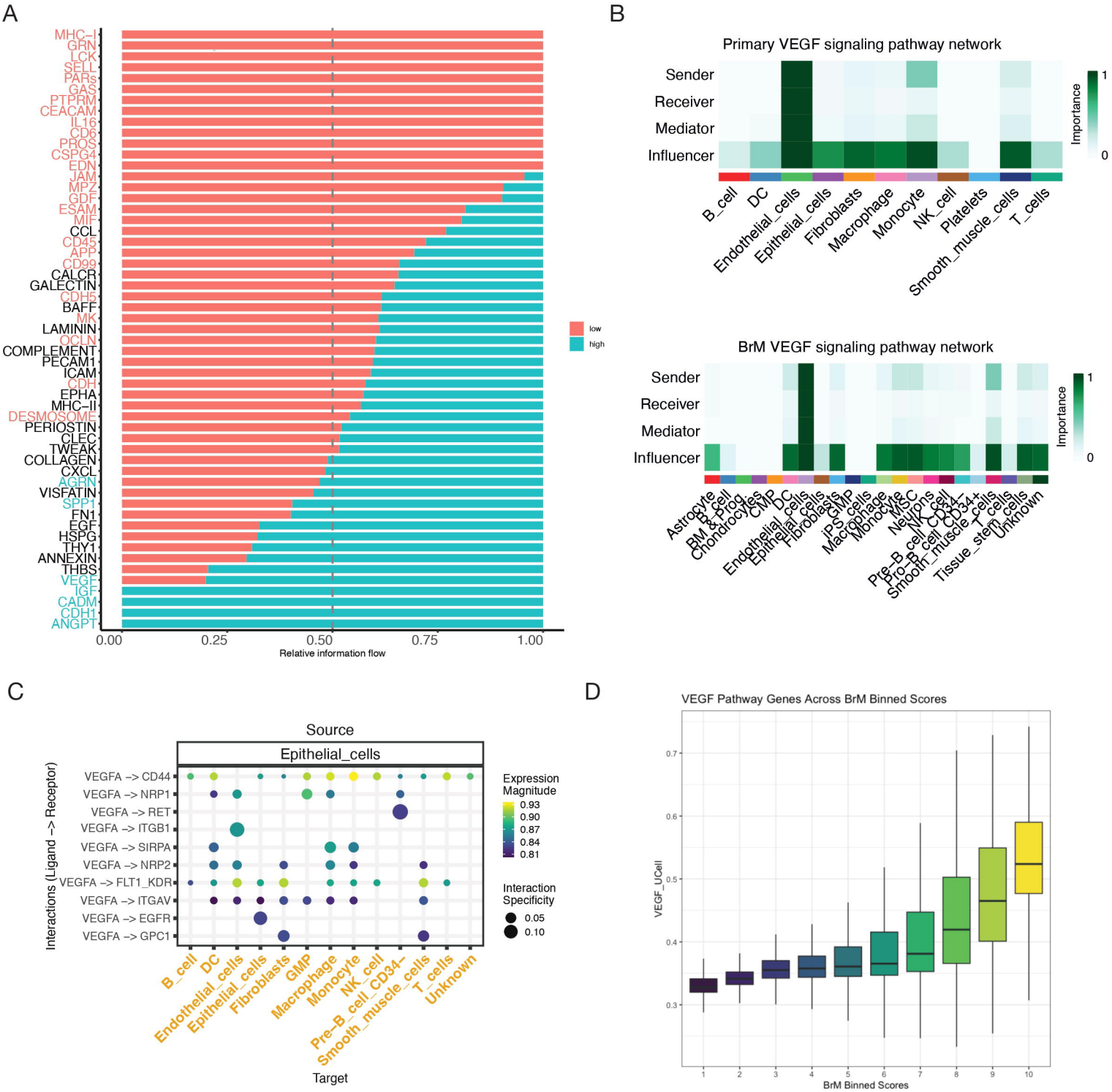

BrM_Sig
Breast
Colorectal
Lung
Melanoma
Ovarian
Renal
Unknown
Cancer

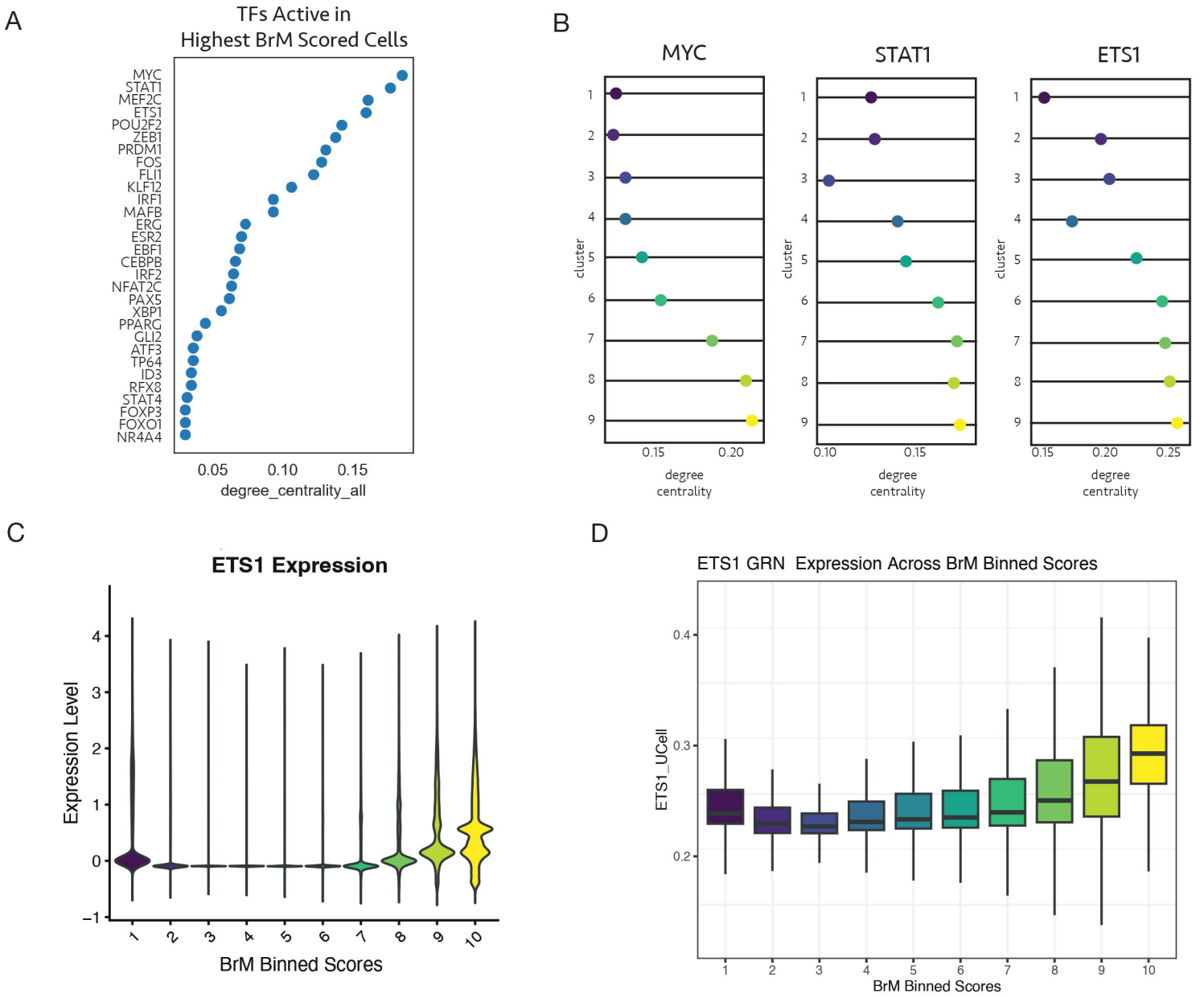
E


GO Enrichment Analysis: Biological Processes
synapse organization
modulation of chemical synaptic transmission
regulation of trans−synaptic signaling
cell junction assembly
external encapsulating structure organization
extracellular matrix organization
extracellular structure organization
cell−cell adhesion via plasma−membrane adhesion molecules
synapse assembly
homophilic cell adhesion via plasma membrane adhesion molecules
GeneRatio
p.adjust
Count

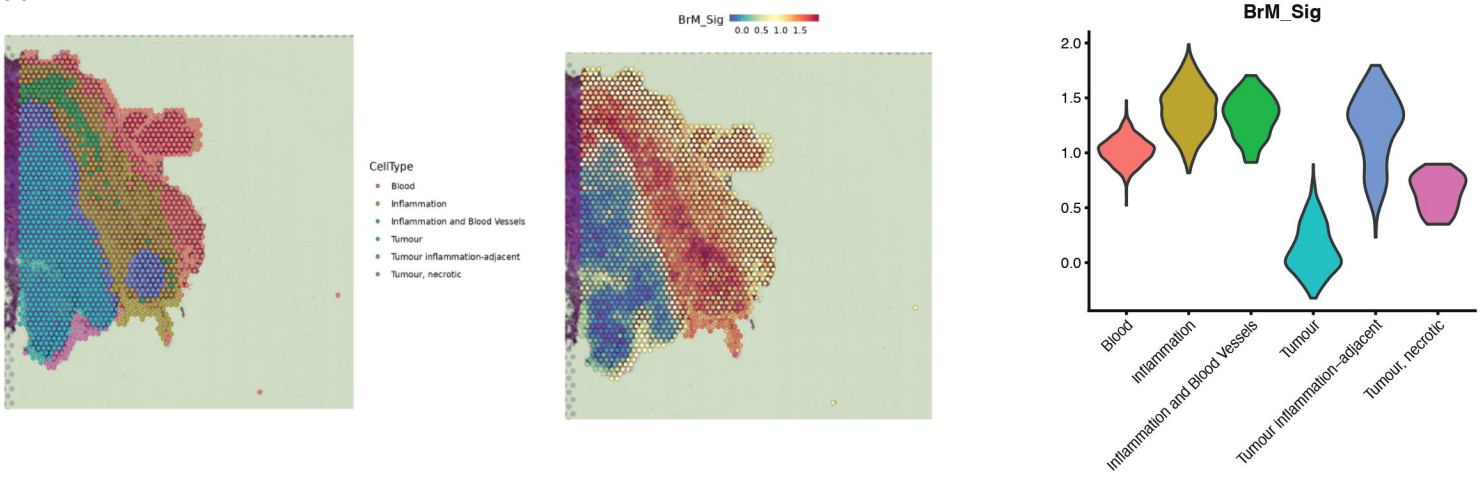
F


label
BrM
Primary
BrM_Sig
Pseudotime
Component 1
Component 2

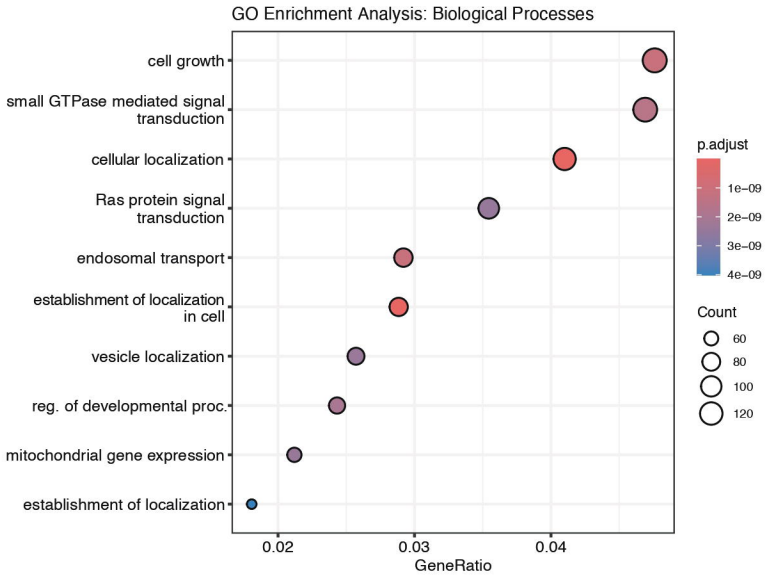

A
MHC-I
GRN
LCK
SELL
PARs
GAS
PTPRM
CEACAM
IL16
CD6
PROS
CSPG4
EDN
JAM
MPZ
GDF
ESAM
MIF
CCL
CD45
APP
CD99
CALCR
GALECTIN
CDH5
BAFF
MK
LAMININ
OCLN
COMPLEMENT
PECAM1
ICAM
CDH
EPHA
MHC-II
DESMOSOME
PERIOSTIN
CLEC
TWEAK
COLLAGEN
CXCL
AGRN
VISFATIN
SPP1
FN1
EGF
HSPG
THY1
ANNEXIN
THBS
VEGF
IGF
CADM
CDH1
ANGPT
low
high
0.00
0.25
0.50
0.75
1.00
Relative information flow
B
Primary VEGF signaling pathway network
Sender
Receiver
Mediator
Influencer
Importance
1
0
B_cell
DC
Endothelial_cells
Epithelial_cells
Fibroblasts
Macrophage
Monocyte
NK_cell
Platelets
Smooth_muscle_cells
T_cells
BrM VEGF signaling pathway network
Sender
Receiver
Mediator
Influencer
Importance
1
0
Astrocyte
B_cell
BM & Prog
Chondrocytes
CMP
DC
Endothelial_cells
Epithelial_cells
Fibroblasts
GMP
iPS_cells
Macrophage
Monocyte
MSC
Neurons
NK_cell
Pre-B_cell_CD34-
Pro-B_cell_CD34+
Smooth_muscle_cells
T_cells
Tissue_stem_cells
Unknown
C
Source
Epithelial_cells
Interactions (Ligand -> Receptor)
VEGFA -> CD44
VEGFA -> NRP1
VEGFA -> RET
VEGFA -> ITGB1
VEGFA -> SIRPA
VEGFA -> NRP2
VEGFA -> FLT1_KDR
VEGFA -> ITGAV
VEGFA -> EGFR
VEGFA -> GPC1
Expression Magnitude
0.93
0.90
0.87
0.84
0.81
Interaction Specificity
0.05
0.10
B_cell
DC
Endothelial_cells
Epithelial_cells
Fibroblasts
GMP
Macrophage
Monocyte
NK_cell
Pre-B_cell_CD34-
Smooth_muscle_cells
T_cells
Unknown
Target
D
VEGF Pathway Genes Across BrM Binned Scores
VEGF_UCell
0.7
0.6
0.5
0.4
0.3
1
2
3
4
5
6
7
8
9
10
BrM Binned Scores


A
TFs Active in
Highest BrM Scored Cells
MYC
STAT1
MEF2C
ETS1
POU2F2
ZEB1
PRDM1
FOS
FLI1
KLF12
IRF1
MAFB
ERG
ESR2
EBF1
CEBPB
IRF2
NFAT2C
PAX5
XBP1
PPARG
GLI2
ATF3
TP64
ID3
RFX8
STAT4
FOXP3
FOXO1
NR4A4
0.05
0.10
0.15
degree_centrality_all
B
MYC
STAT1
ETS1
cluster
degree
centrality
C
ETS1 Expression
Expression Level
BrM Binned Scores
D
ETS1 GRN Expression Across BrM Binned Scores
ETS1_UCell
BrM Binned Scores

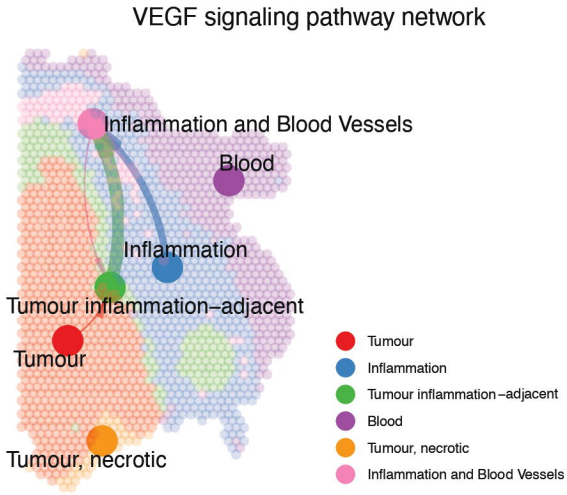
A


CellType
Blood
Inflammation
Inflammation and Blood Vessels
Tumour
Tumour inflammation-adjacent
Tumour, necrotic
BrM_Sig
0.0 0.5 1.0 1.5
BrM_Sig
2.0
1.5
1.0
0.5
0.0
Blood
Inflammation
Inflammation and Blood Vessels
Tumour
Tumour inflammation–adjacent
Tumour, necrotic

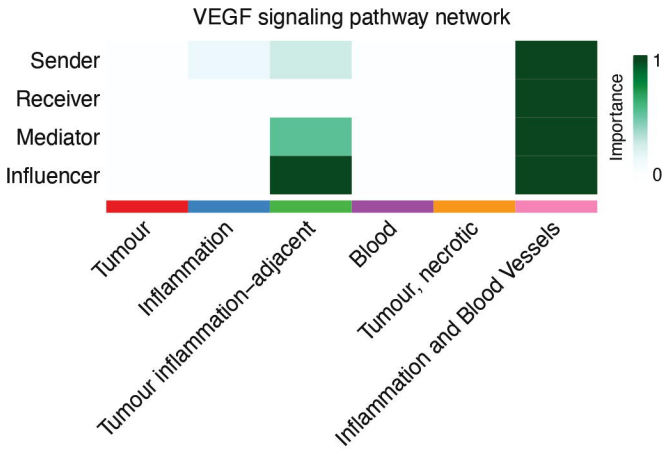
B


GO Enrichment Analysis: Biological Processes
cell growth
small GTPase mediated signal transduction
cellular localization
Ras protein signal transduction
endosomal transport
establishment of localization in cell
vesicle localization
reg. of developmental proc.
mitochondrial gene expression
establishment of localization
0.02
0.03
0.04
GeneRatio
p.adjust
1e−09
2e−09
3e−09
4e−09
Count
60
80
100
120

C


VEGF signaling pathway network
Inflammation and Blood Vessels
Blood
Inflammation
Tumour inflammation–adjacent
Tumour
Tumour, necrotic
Tumour
Inflammation
Tumour inflammation–adjacent
Blood
Tumour, necrotic
Inflammation and Blood Vessels

D


VEGF signaling pathway network
Sender
Receiver
Mediator
Influencer
Importance
1
0
Tumour
Inflammation
Tumour inflammation–adjacent
Blood
Tumour, necrotic
Inflammation and Blood Vessels

E

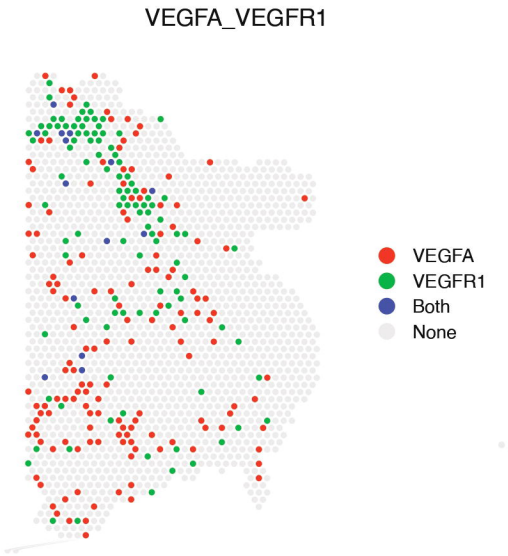
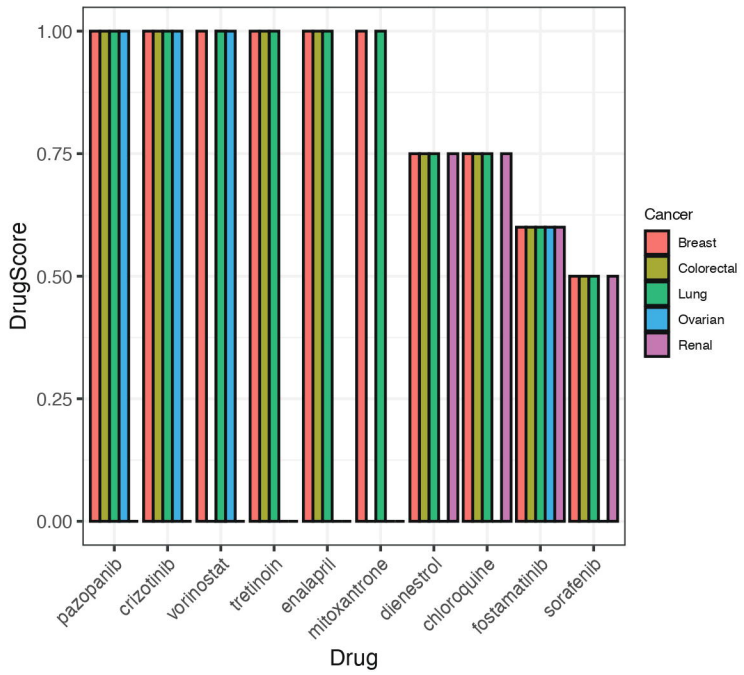

VEGFA_VEGFR1
VEGFA
VEGFR1
Both
None

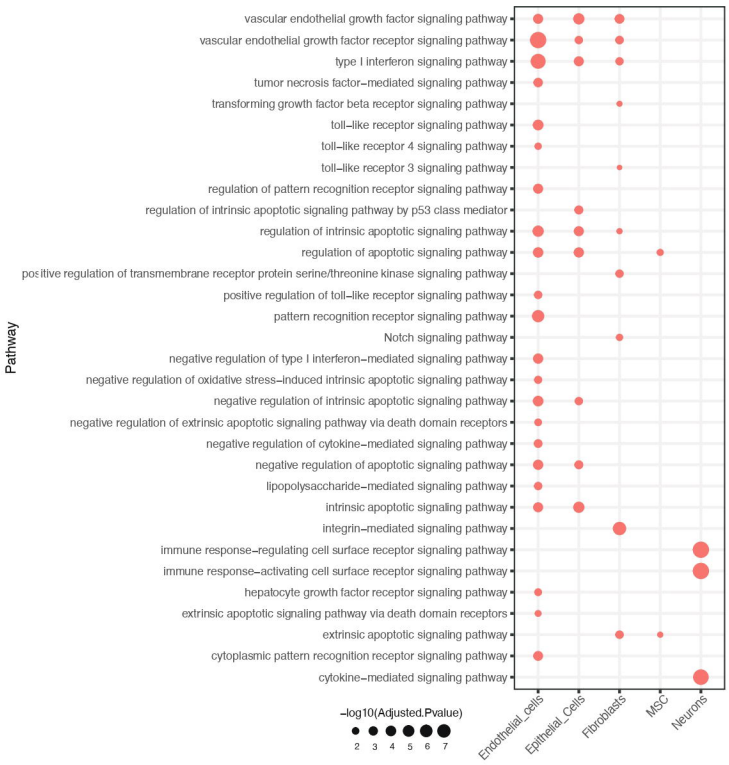
Figure 6

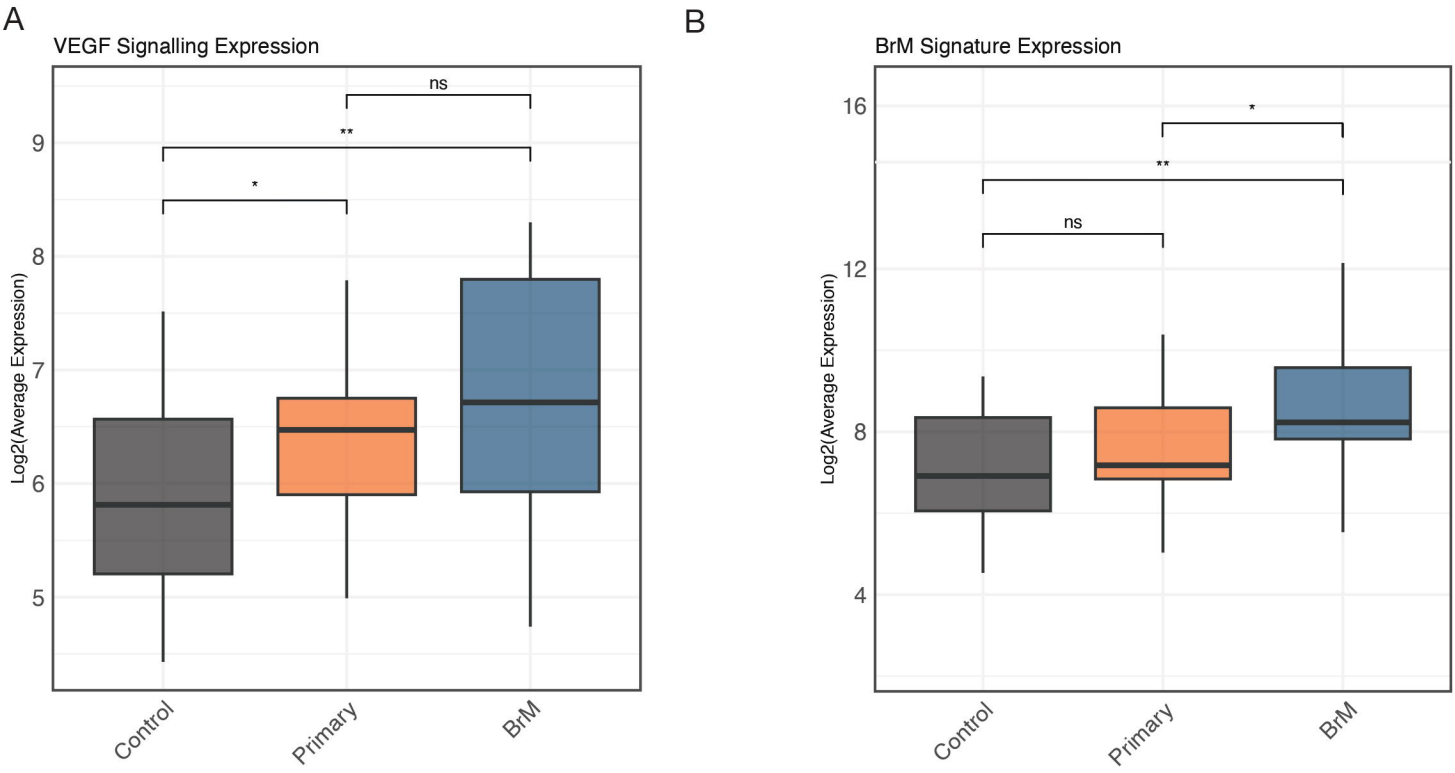
Lusby et al

A



B




A
VEGF Signalling Expression
Log2(Average Expression)
9
8
7
6
5
ns
**
*
Control
Primary
BrM
B
BrM Signature Expression
Log2(Average Expression)
16
12
8
4
*
**
ns
Control
Primary
BrM