Integrating single-cell transcriptomics with Artificial Intelligence reveals pan-cancer biomarkers of brain metastasis

Ryan Lusby[1], Sarah Carl[2], Vijay K. Tiwari[1, 3,4,5,6]*

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

* Correspondence should be addressed to tiwari@health.sdu.dk or v.tiwari@qub.ac.uk

Keywords: Genomics, Metastasis, Artificial Intelligence, Biomarkers. Prognosis, Therapy

Running title: Pan-Cancer Biomarkers for Brain Metastasis

ABSTRACT

Brain metastasis (BrM) represents a devastating complication across various cancer types, posing as a significant contributor to global morbidity and mortality. Hence, identifying robust biomarkers for early detection across various cancer types with a propensity for BrMs and their therapeutic targeting is highly timely and critical. In this study, we leveraged single-cell RNA sequencing (scRNA-seq) data from six cancer types and combined with convolutional neural network (CNN)-based ScaiVision algorithm to identify a pan-cancer BrM signature that achieved remarkable accuracy in distinguishing BrM from primary tumour cells. Further analysis revealed that the BrM signature was not only prognostic but also detectable in bulk RNA-seq data, providing a stratification tool for patients with high or low metastatic potential. Strikingly, this signature was detected at high levels in the tumour educated platelets, showcasing its potential as a minimally invasive tool for metastasis detection. High BrM signature scores were associated with reduced patient survival, particularly in cancers prone to brain metastasis, such as renal and colorectal cancers. Further analysis uncovered VEGF signalling as a central driver of communication networks in high BrM-scored cells. Accordingly, drug repurposing analysis identified Pazopanib as a candidate for targeting highly metastatic cells that disrupts VEGF signalling networks, and potentially impedes brain metastatic progression in multiple cancer types. This study presents a comprehensive pan-cancer BrM signature with clinical implications for early detection and therapeutic intervention in brain metastasis.

INTRODUCTION

The occurrence of brain metastasis (BrM) poses a formidable challenge in the clinical care of cancer patients, significantly impacting prognosis and treatment approaches(1). This complex condition involves the migration of cancer cells from primary tumours to the brain, leading to substantial morbidity and restricted therapeutic choices(2). Importantly, brain metastasis is not exclusive to a particular type of cancer, as the majority of cancers demonstrate the potential to metastasize to the brain(3). This non-specificity underscores the universal impact and clinical significance of brain metastasis, emphasising the need for comprehensive strategies to address its challenges across diverse cancer types.

Despite numerous advancements facilitated by technologies like scRNA-seq, a notable gap in understanding persists regarding the fundamental biological processes driving BrMs(4). This knowledge deficit significantly hampers the identification of effective interventions and the enhancement of patient outcomes. Recent scRNA-seq studies have provided insights into a shared transcriptomic landscape in brain metastasis (BrM) tumours. In one such investigation focusing on BrMs originating from breast and lung tumours, researchers identified conserved fibroblast signatures(5). These signatures suggested the presence of a proinflammatory tumour microenvironment (TME). These findings not only contribute to our understanding of the commonalities in the molecular landscape of BrMs but also highlight potential biomarkers that could serve as therapeutic targets for intervention strategies. Another such study focusing on BrMs from lung cancer identified a signature for BrM risk prediction(2). Further investigation is required to decode conserved predictive biomarkers across cancer types and identify novel therapeutic interventions to ultimately improve the outlook for patients facing the challenges of brain metastasis.

BrM presents a key challenge across diverse cancer types and therefore a pan-cancer investigation offers the most logical approach with significant advantages. Firstly, it can reveal conserved molecular pathways crucial for BrM, facilitating the identification of therapeutic targets applicable across a spectrum of cancers(6). Moreover, the incorporation of multiple cancer types enhances statistical power, enabling the detection of subtle yet biologically significant differences in gene expression and cell behaviour linked to BrM(7). The insights garnered are inherently generalizable, ensuring relevance across diverse patient populations and, consequently, elevating clinical significance(8). Ultimately, pan-cancer studies play a pivotal role in biomarker detection, shaping therapies not only for BrM but also for addressing the intricacies associated with different primary cancers(9). The advantages of a pan-cancer study on BrM scRNA-seq have been highlighted in a recent study that showed that BrMs

shared characteristics across diverse cancers(10). Notably, metastasis-associated macrophages exhibit two distinct states, and the analysis identifies eight metaprograms defining two coexisting archetypes of metastatic tumour cells (MTCs). These archetypes are distinguished by one exhibiting proliferation and the other displaying an inflammatory profile. The study also identifies an immune-evasive programme that supports the proliferative state of MTCs. These findings contribute valuable insights into the heterogeneity and immune dynamics within BrMs, shedding light on common features across different cancer types.

In this investigation, we conducted the first comprehensive pan-cancer study utilising scRNA-seq data to pinpoint BrM markers within the primary tumour across diverse cancer types. Leveraging the capabilities of single-cell transcriptomics and deep learning, here we reveal both shared and distinct molecular signatures associated with BrM and explore the underlying therapeutic opportunities. Through this study, we contribute to a deeper comprehension of the underlying biology of BrM progression, providing valuable insights with the potential to improve the clinical management of patients of several major cancer types.

RESULTS

Identification of 48-gene signature with high accuracy in Predicting Brain metastasis

Given the high mortality rate and incidence of BrMs, there is a critical need to discover biomarkers for diagnosis, stratification, and novel therapeutic targets that apply to multiple cancer types with BrMs. To address this, we combined single-cell RNA sequencing (scRNA-seq) datasets from multiple cancers with Scailyte AG's convolutional neural network (CNN)-driven algorithm, ScaiVision, to enable high accuracy identification of cells in the primary tumour with a high potential for metastasising to the brain (Fig 1A).

Initially, droplet-based scRNA-seq data were collected from 17 studies encompassing six different cancer types, including breast, colorectal, lung, melanoma, ovarian, and renal (Supplementary Table 1). After excluding samples with fewer than 100 cells, conducting quality control, and mitigating batch effects, we retained 1,000,000 cells from 200 patients for proceeding with predictive analyses (Fig 1B). To ensure that we would not detect differences which arose from different cell type populations in the primary and metastatic site, we subset the dataset to include only epithelial cells (Supplementary Fig 1A). Additionally, previous studies have demonstrated that signatures derived from scRNA-seq perform more effectively when derived from epithelial cells exclusively(11).

Our analytical methodology employed Scailyte AG's proprietary software platform, ScaiVision. Within this platform, we leveraged a specialised neural network designed for identifying features relevant to endpoint prediction, distinguishing between BrM or Primary tumour using scRNA-seq data. To determine the optimal model for addressing each scientific inquiry, we executed a hyperparameter tuning process. In this approach, we created a dedicated validation dataset by partitioning 30% of the samples, leaving 70% for model training across 6 cross-validation splits. Subsequently, we trained a total of 50 models with randomised hyperparameters across 50 epochs using the training dataset (Supplementary Fig 1B. The best-performing models, characterised by their superior predictive capabilities on the validation dataset, were chosen for further analysis (training AUC =1 validation AUC = 0.98) (Fig 1C). The outcomes of our ScaiVision analysis comprised two crucial components. Firstly, we established a trained model capable of predicting whether the sample corresponds to BrM or Primary. Secondly, our analysis facilitated the identification of gene features, to extract a signature, utilising captum(12), defining BrM that were relevant to the prediction, setting the stage for subsequent in-depth analysis (Supplementary Fig 1C).

We next sought to characterise cells within our consolidated scRNA-seq data using the BrM signature, assigning them to a score based on their metastatic potential, ranging from low to high. This analytical approach aimed to unveil the underlying biological processes crucial for driving BrM progression. By employing gene set scores derived from the refined BrM signature, serving as a relative measure of cells' metastatic potential, we utilised UCell scoring(13) to position each cell along the spectrum of low to high metastatic potential within the epithelial cells (Supplementary Fig 1D). We next expanded this scoring to all cells (Fig 1D), Interestingly, while the BrM signature distinctly scored BrM samples as having high metastatic potential, it also demonstrated the ability to identify high-scoring cells within the primary tumour. This intriguing finding suggested that the signature might be potent to pinpoint cells in the primary tumour with an elevated likelihood of metastasising to the brain. Furthermore, the signature displayed a lack of specificity to a particular cancer type, exhibiting high scores across various cancer types (Supplementary Fig 1E-F).

In order to stratify cells and patients, we binarized each cell’s score, creating a categorisation into three distinct groups: high (top 20% score), low (bottom 20% score), and mid (20% to 80% score) within the dataset. Subsequently, we performed a pseudobulk differential gene expression analysis, to comprehensively explore the transcriptional landscape in cells exhibiting a high BrM score compared to a low BrM score (Fig 1E). This analysis showed that cells characterised by a high metastatic potential display significantly elevated gene expression in biological processes associated with metabolism, including Adenosine

triphosphate (ATP) metabolism (Fig 1F). Increased expression of ATP metabolism in cancer cells has been correlated with an aggressive phenotype, featuring enhanced stem-like properties and heightened capabilities for cell migration, invasion, and metastasis.(14).

BrM signature can spatially resolve metastatic cells in colorectal BrM samples

While single-cell RNA sequencing (scRNA-seq) provides comprehensive transcriptomic information for all cells in a sample, its limitation lies in the absence of an spatial context(15). To overcome this, we next assessed the effectiveness of BrM signature in detecting metastatic cells within colorectal BRM spatial transcriptomics data, aiming to reveal region-specific gene expression profiles and understand the spatial dynamics of metastatic progression within the secondary site and the surrounding tumour microenvironment (TME). The elucidation of such spatially resolved molecular landscapes deepens our comprehension of metastasis and has potential to offer novel therapeutic targets and strategies (16).

Through the implementation of UCell scoring on our BrM signature, we successfully identified clusters of cells with heightened metastatic potential on individual patient slides of BrM originating from primary colorecta cancer. Importantly, our BrM signature could identify high-scoring cells beyond the central regions of the BrM tumour to include cells at the periphery and those associated with blood regions (Fig 2A). This intriguing observation suggests the potential utility of our BrM signature as a blood-based biomarker panel, introducing the possibility of a minimally invasive approach for metastasis detection.

These findings suggest that our BrM signature gene is not only accurate in identifying cells with a high potential for BrM progression but can also be utilised in both scRNA-seq and spatial transcriptomics data, enabling characterisation of the underlying biology of brain metastasis.

Elevated expression of BrM signature genes correlates with reduced patient survival

We next aimed to evaluate the effectiveness of our Brain Metastasis (BrM) signature in stratifying patients using bulk RNA data. We had earlier ensured its specificity to cancer cells by employing ScaiVision analysis and constructing the signature solely from epithelial cells. To test whether this specificity is retained in the bulk RNA-seq data, we first measured the average expression of our refined metastatic gene list across diverse cancer types within the pan-cancer TCGA dataset(17). Interestingly. this analysis revealed statistically significantly higher expression of our signature genes across most cancer types compared to normal tissue, particularly in cancers originating from epithelial cells (Fig 3A).

Having observed the presence of our metastatic gene signature across diverse cancer types, we next investigated whether it holds any prognostic potential for the patients. Utilising the TCGA pan-cancer data, we computed a relative BrM score for all patients, facilitating their stratification into high or low metastatic potential using the median score within each cancer type as the threshold. The BrM signature exhibited elevated expression across various cancer types, with the highest levels observed in renal and colorectal cancers (Fig 3B). Notably, both of these cancers are known to be prone to metastasize to the brain(18, 19).

Employing a model integrating patients' relapse-free survival (RFS) with the refined signature's activity, accounting for tumour type, stage, age at diagnosis, and tumour purity, we identified a compelling association. When taking median expression of our BrM signature patients in the high cohort were significantly linked to reduced RFS (Fig 3C) (Cox hazard ratio: 2.65; $p = 7.18 \times 10^{-5}$). Whilst the metastatic status of these patients is unknown, these results reveal a novel prognostic role for our metastatic gene-set. Irrespective of the cancer type, the BrM signature demonstrated proficiency at accurately classifying patients into high and low metastatic potential, with direct correlations to survival outcomes.

BrM signature exhibits an elevated expression in the blood of metastatic patients

Blood-based biomarkers represent the gold standard in cancer research, offering a non-invasive and accessible avenue for monitoring cancer progression and diagnosis. The appeal lies in their ease of sampling, which minimises patient discomfort and allows for repeated assessments over time (20). In the context of cancer, blood-based biomarkers hold significant promise for early detection, prognosis, and monitoring treatment responses(21). As the BrM signature was shown to have a high specificity in blood regions on the colorectal spatial transcriptomics slide we next sought to assess the detectability of our BrM signature in metastatic patients' blood. To achieve this, we leveraged RNA-seq data from tumour educated platelets across seven cancer types consisting of both metastatic and non-metastatic patients. Remarkably, all 48 genes comprising the BrM signature exhibited significantly higher expression in metastatic patients (Fig 4A). Recognising the practical utility of biomarker panels with fewer than 20 genes, we refined our focus to the top 11 genes based on ScaiVision scoring. Upon re-evaluation, the differences between metastatic and non-metastatic patients were even more pronounced, suggesting that the BrM signature possesses robust stratification capabilities and is discernible in the bloodstream (Fig 4B). Furthermore, by subsetting the metastatic patients into patients with brain metastasis and other sources, we again measured the expression of the 11 genes and found that, the average expression was higher in patients presenting with brain metastasis (Fig 4C). Interestingly, the Gene Ontology

analysis of these genes reveals their association with biological processes related to neutrophil activation (Fig 4D). In recent years, accumulating evidence suggests that such activation, facilitated through cell-cell communication with tumour cells, plays a crucial role in the colonisation and metastasis of cancer cells(22, 23).

Notably, the absence of specific details regarding the metastatic site in patients introduces an element of uncertainty regarding the BrM signature's specificity to brain metastases. Consequently, its potential applicability for detecting metastatic disease comprehensively offers an intriguing avenue for further investigation.

VEGF Signalling is a central driver of cell-cell communications underlying brain metastasis

Given the capacity of our BrM signature to discern cells with heightened metastatic potential within both the primary tumour and the surrounding tumour microenvironment, our subsequent analysis was to elucidate the pan-cancer cell-to-cell communication networks steering BrM progression. Existing evidence underscores the pivotal role of communication between tumours and adjacent stromal cells in processes such as tumour proliferation, metastasis, and treatment response(24, 25). The intricate microenvironment of the brain, where metastases often occur, adds an additional layer of complexity to this investigation(26).

To unravel the molecular connections among low and high-scoring BrM cells, we conducted an analysis of potential signalling interactions using CellChat(27). The cell-cell communication network was calculated for all cell types in high-scoring BrM cells, where the edge width indicates the strength of communication. Subsequently, we compared all signalling pathways between low and high BrM-scored cells(Supplementary Fig 2A-B). This analysis unveiled the activation of VEGF signalling exclusively in high BrM cells and its absence in low BrM cells (Fig 5A).

VEGF signalling has previously been established as a pivotal factor in the progression of BrMs(28–30). Despite its recognised significance, the fundamental cell-cell communication networks underlying BrM development at both the primary and metastatic sites remain elusive. To address this key knowledge gap, we focused our analysis on VEGF signalling in cells with high and low BrM scores. Notably, this analysis uncovered that VEGF signalling was exclusive to high metastatic scored cells, with no detectable activity in low scored cells. Within high-scored cells, endothelial cells emerged as the dominant cell type regulating VEGF signalling, engaging in interactions with multiple cell types in both the tumour and the surrounding tumour microenvironment in both the primary and secondary site (Fig 5B). Moreover, our analysis

unveiled that the principal ligand interaction governing this communication was VEGFA-VEGFR1 (Fig 5D). VEGFR1 has been previously demonstrated to mediate various responses to VEGF signalling, encompassing cell migration, invasion, and dissemination(31, 32). Importantly, it has been shown to significantly correlate with tumour metastasis in diverse cancers, including lung, liver, and glioblastoma(33).

To gain insight into the distinct cell-cell communications occurring between primary tumors and brain metastasis (BrM) sites, we implemented a subsetting approach focusing on the high-scored patients based on the source of the tumours (primary or BrM). Notably, within this analysis, VEGF signalling emerged as a prominent signalling pathway, consistent with previous observations. However, a deeper examination revealed notable differences in the cell-cell communication networks between primary tumours and BrM sites (Fig 5D-E).

In primary tumours, endothelial cells emerged as the dominant cell type involved in cell-cell communication networks. This suggests a significant role of endothelial cells in orchestrating intercellular communication within the primary tumour microenvironment. Conversely, in BrM tumours, epithelial cells appeared to play a more substantial role compared to primary tumours. This shift suggests a potential alteration in the cellular landscape and communication dynamics upon metastatic colonization in the brain.

Furthermore, the analysis highlighted dendritic cells (DCs) as exhibiting higher importance scores in BrM patients compared to primary tumours. This observation suggests an enhanced involvement of DCs in the communication networks within the BrM microenvironment. Such findings may reflect the unique immune microenvironment of the brain and the potential implications for immune surveillance and response mechanisms in the context of brain metastasis.

These findings underscore the acknowledged significance of VEGF signalling in the progression of brain metastasis (BrM), simultaneously unveiling the precise cell-cell communication networks that underlie BrM development at both primary and metastatic sites. This highlights the crucial role of VEGF signalling and identifies potential targets for therapeutic intervention.

Gene Regulatory Network analysis idenfied MAFB as a potential driver of brain betastasis progression

While unified transcriptional signatures of BrM hold clinical relevance, understanding the core regulatory mechanisms driving pan-cancer brain metastasis (BrM) progression is essential.

Uncovering these mechanisms can reveal common features influencing shared disease progression and identify regulatory dependencies that may warrant pan-cancer therapeutic approaches. Notably, our exploration into the intricate interactions within the tumour microenvironment (TME) has suggested the involvement of shared VEGF pathways in driving BrM progression at the pan-cancer level.

To investigate the network dynamics during the progression of brain metastasis (BrM), we undertook the task of inferring gene regulatory network (GRN) configurations spanning from low to high metastatic scores. In the first instance, we binarized the metastatic scores into 10 distinct groups to enable quantification of GRNs across metastatic timepoints. To dissect the roles of individual TFs within these networks, we employed CellOracle, resulting in the identification of key TFs across each pseudo-timepoint. A fundamental metric, degree centrality, quantifies the number of edges connected to a node, offering valuable insights into the connectivity of a given TF within the network.

In the high BrM cluster GRN configuration, degree centrality scoring successfully identified key TFs associated with the BrM stage, including MAFB and RXRG (Fig 6A). Conversely, the low and medium BrM score clusters revealed TFs linked to the reduction of BrM progression and the prevention of disease advancement (Fig 6A). With detailed information on each regulon across every metastatic stage, we analysed how network connectivity changes during metastatic progression, highlighting alterations in TF activity across BrM timepoints. Our specific focus on MAFB and RXRG, due their previous involvement in in VEGF signalling (34, 35), within the high BrM score clusters showed that its network scores mirrored the BrM progression scores, exhibiting reduced activity in low to medium scored cells and a sharp increase in high BrM scored cells (Fig 6B & Supplementary Fig 3B). Although increased expression of MAFB has been significantly correlated with local lymph node metastasis, high recurrence rates, and poor pathological stage in lung cancer, its precise role in metastasis remains unclear. Furthermore, when examining the GRNs of each TF we could see that many genes, including several from our BrM signature, were targets of both MAFB and RXRG (Supplementary Fig 3B), hinting that both of these TFs could be directly regulating the BrM signature and leading to increased BrM progression at the pan-cancer level.

Our subsequent investigation delved into the expression patterns of MAFB across normal, primary, and metastatic cancer samples. Notably, MAFB exhibited elevated expression in metastatic patients in colorectal, breast and ovarian samples (Fig 6C). To underscore the significance of MAFB in metastatic progression, we assessed its impact on overall survival using the TCGA pan-cancer data. Strikingly, high expression of MAFB was correlated with reduced survival (Fig 6D).

Through this analysis we have uncovered the putative roles of each hub TF and how they change across the metastatic timepoints. Collectively, these findings affirm the recognised role of MAFB as a pivotal driver of metastasis, while simultaneously shedding light on its potential influence at the pan-cancer level for the first time. These insights may provide an opportunity to prevent metastatic progression by directly targeting TFs identified in high metastatic score cells or promoting the overexpression of TFs found in cells with low metastatic scored cells. This could provide novel therapeutic avenues which require further exploration.

In silico MAFB knockout reverses brain metastasis progression

The CellOracle GRN inference and network analysis successfully pinpointed central transcription factor (TF) regulators of high-scored BrM cells. To further elucidate the roles of these TFs and prioritise them for novel drug target identification, we subset our pan-cancer scRNA-seq dataset to contain only breast cancer samples and performed *in silico* perturbation for MAFB. Employing our refined signature on a force-directed graph (FDG) layout of the breast cancer subset, we observed a distinct clustering of metastatic cells (Fig 7A-B), reaffirming our BrM signatures capability in defining subsets of cells with a high potential for metastasis to the brain. Pseudotime and developmental flow analysis using CellOracle mirrored the sample progression as indicated by our metastatic score.

Subsequently, through CellOracle-based simulation of MAFB perturbation, we sought to recapitulate the progression from low to high BrM cell fate. Leveraging the 10 metastatic GRN configurations inferred by CellOracle, we simulated MAFB knockout signal propagation (setting MAFB GRN expression to 0 across all cells), allowing the prediction of future gene expression and, consequently, the direction of cell identity transitions at single-cell resolution. Notably, this simulation revealed a distinct transition, wherein high metastatic cell identities deviated from their predicted trajectory towards a low metastatic signature. Notably, this deviation was represented by a drastic change in direction of the arrows depicting the anticipated cell fate paths prior to the knockout event. (Fig 7B-C). Collectively these findings suggest that, if metastatic patients are treated with a MAFB inhibitor it could potentially prevent or reverse metastatic progression.

Drug repurposing analysis identifies Pazopanib as a key candidate for reversing brain metastasis progression

Drug repurposing, a strategy aimed at discovering new applications for drugs beyond their

original medical approval or investigation(36), presents a promising avenue for identifying novel therapeutic strategies with lower risks of failure in comparison to standard drug discovery pipelines(37). Recently, ASGARD, among various pipelines developed for the discovery of Food and Drug Administration (FDA) approved drugs for new applications, has demonstrated superior performance(38).

In pursuit of potential drug candidates targeting genes in cells with a high metastatic score, we utilised the R package ASGARD to subset our merged pan-cancer scRNA-seq dataset into low BrM cells (bottom 20% score) and high BrM cells (top 20% score). Employing low BrM cells as a control, we conducted differential gene expression analysis using limma to identify genes consistently differentially expressed between metastatic low and high cells, accounting for the inherent heterogeneity. Subsequently, these consistently differentially expressed genes were employed as inputs to identify drugs that could significantly (FDR < 0.05) reverse their expression levels in the L1000 drug response dataset, encompassing 591,697 drug/compound treatments. Our initial application of ASGARD for single drug repurposing prediction yielded drugs (FDR < 0.05 and overall drug score > 0.99) across breast, colorectal, lung, NPC, ovarian, and PDAC cancers (Fig 8A). To explore the potential molecular mechanisms of these drugs, we next investigated the target genes and pathways of the top 5 drugs across each cancer type (Fig 8B). Of key interest, these drugs appear to target genes involved in VEGF signalling in highly metastatic cells and where in expressed in many of the cell types identified in the cell chat analysis. Of particular interest to this study was the VEGF inhibitor Pazopanib. While its efficacy has been firmly established in clinical trials for metastatic cancers(39), our study sheds light on its potential applicability at the pan-cancer level in the context of BrMs. Future studies will be required to test Pazopanib's precise mechanisms of action in controlling BrMs and to explore its synergistic effects with other treatment modalities. Altogether, these findings suggest that by disrupting VEGF signalling communication networks in highly metastatic cells, treatment with Pazopanib holds promise for targeting brain metastatic progression across multiple cancer types and improve the clinical outcomes from a wide range of patients.

DISCUSSION

Brain metastases, found in approximately 40% of cancer cases(40), pose a formidable challenge due to limited treatment options and a grim prognosis. Despite extensive efforts to identify biomarkers, progress in improving the prognosis of metastatic brain conditions has been minimal in recent decades(4). Patients diagnosed with brain metastasis often present

with advanced systemic disease, leading to an alarming survival rate of less than 6 months (41).

To confront the persistent challenges posed by elevated mortality rates and the incidence of brain metastases (BrMs) across diverse cancer types(3), our investigation harnessed the power of single-cell data and neural networks. Through this we successfully constructed a predictive model and provided a molecular characterisation of cells associated with BrMs. Remarkably, the BrM signature not only adeptly scored BrM samples but also exhibited the ability to identify high-BrM-potential cells within primary tumours, showcasing its versatility and wider applicability. The signature's lack of specificity to a particular cancer type highlight its potential for pan-cancer scRNA-seq and RNA-seq applications. Moreover, the spatial efficacy of the BrM signature was demonstrated by its capability to reveal clusters of high-potential metastatic cells in spatial transcriptomics data. Crucially, the BrM signature was also shown to possess prognostic significance, as evidenced by its correlation with reduced relapse-free survival across various cancer types. This multifaceted effectiveness positions the BrM signature as a promising tool with implications for both diagnostic precision and prognostic insights across the spectrum of cancer-associated brain metastases.

A distinctive characteristic of the genes within the BrM signature lies in their engagement in ATP synthesis, a feature underscored by recent studies illustrating the pivotal role of ATP in fostering metastatic dissemination and conferring intrinsic resistance to a spectrum of chemotherapy drugs(42–44). Within the context of brain metastasis, ATP assumes a critical role in the Tumour Microenvironment (TME), facilitating communication among astrocytes, neurons, microglia cells, and notably, due to VEGF signalling, blood vessels(45). Our investigation has unveiled the VEGF signalling pathway as a central driver of BrM progression, not only within the confines of the primary tumour site but also extending its influence on the brain metastatic site. The analysis of cell-to-cell communication networks in high-scoring BrM cells suggests that VEGF signalling emanating from the stroma plays a pivotal role in priming tumour cells for dissemination from the primary site while concurrently influencing the priming of the metastatic site in the brain. While the involvement of VEGF signalling in metastatic progression is documented(46), our findings, for the first time, delineate the specific cell-to-cell communication networks at both the single-cell and pan-cancer levels associated with VEGF signalling.

In our exploration of gene regulatory networks (GRNs), a noteworthy observation emerged regarding the transcription factor MAFB, revealing diminished activation in low BrM cells and heightened activation in high BrM cells. This observation gains significance due to the intricate

and context-dependent interplay between MAFB and VEGF signalling in the realm of cancer biology. MAFB, has previously been shown to critically contribute to various cellular processes, encompassing differentiation, proliferation, and metastasis(47–49). This study reveals an intriguing potential for crosstalk between MAFB and VEGF signalling pathways which is particularly interesting in light of recent studies demonstrating the activation of MAFB by VEGFR-3. Moreover, MAFB has been implicated in exerting regulatory control over the expression of genes pivotal to metastasis, including those integral to the VEGF signalling cascade(50). This potential interplay, revealed here for the first time at the pan-cancer level, within the intricate landscape of brain metastasis, underscores the need for further in-depth investigations. Subsequent research is imperative to comprehensively unravel the underlying biological model governing this critical crosstalk, shedding light on the complexities of these interconnected pathways and their implications in the broader context of cancer BrM.

In concordance with the elevated ATP processes highlighted by our BrM signature, our investigation reveals a potential interplay between ATP and VEGF signalling. This bears crucial clinical significance, given that drugs targeting VEGF have demonstrated the ability to compete with ATP binding, thereby disrupting signalling pathways and mitigating metastatic and cancer progression(51). Through *in silico* drug repurposing screens, we have pinpointed the small molecule VEGF inhibitor Pazopanib as a leading candidate capable of targeting and reversing the BrM potential, steering cell fate away from high BrM states across diverse cancer types. Pazopanib exerts inhibitory effects by blocking the activation of VEGF signalling pathways through the inhibition of receptor tyrosine kinases. This is achieved by competing with ATP for binding sites on the kinases, resulting in the inhibition of angiogenesis and, subsequently, hindering metastatic progression(52). This mechanism highlights the potential of Pazopanib as a therapeutic agent targeting the VEGF pathways crucial in the progression of BrMs. Pazopanib has been the focus of several clinical trials, particularly in targeting metastatic renal cancer, where advanced metastatic renal carcinoma patients exhibited significantly longer progression-free survival (PFS) compared to a placebo group(39). However, our analysis extends the understanding of its potential utility beyond specific cancer types, highlighting its specific efficacy in preventing the progression of BrM at the pan-cancer level. This seminal discovery not only underscores the critical role of VEGF signalling in BrM but also paves the way for prospective research and potential therapeutic interventions targeting the intricate interplay between ATP and VEGF signalling in the context of BrM progression.

In a significant finding, our investigation into blood-based biomarkers has unveiled an 11-gene panel that holds the potential for improving the landscape of brain metastasis detection. This

distinctive panel could not only pioneer early detection strategies but also introduces to possibility for the development of a minimally invasive approach to predict metastatic potential, addressing the critical need for less intrusive diagnostics in the context of metastatic brain conditions. The genes within this refined BrM signature are notably associated with neutrophil biological processes. This association gains significance due to the observed high levels of these genes in the blood of metastatic patients. Neutrophils play a well-established role in interacting with circulating tumour cells, facilitating their binding to the endothelium(22, 23). Furthermore, neutrophils can actively promote tumour progression by inducing tumour cell proliferation, stimulating angiogenesis and matrix remodelling, and impairing T cell-dependent antitumor immunity(53). This link enhances the importance of the refined 11-gene panel, connecting it not only to detection but also to a hinting at their key role in driving the metastatic process.

Anticipating the transformative impact of the Brain Metastasis (BrM) biomarker panel, it is imperative to underscore the importance of careful consideration regarding specificity, sensitivity, and validation through comprehensive clinical trials. The potential integration of this panel with existing diagnostic modalities offers the promise of synergistic augmentation, ushering in a new era in our understanding and management of brain metastasis. However, further study is warranted to unlock the full potential of this biomarker panel in reshaping brain metastasis diagnostics and prognostics.

Acknowledgements

We would also like to thank the members of the Tiwari lab and Scailyte team for their cooperation and critical feedback throughout this study. The support from the Core Facilities of the Queen's University Belfast and University of Southern Denmark is gratefully acknowledged. This study was supported by Innovation to Commercialisation of University Research (ICURe) and Department for the Economy (DfE).

AUTHOR CONTRIBUTIONS

R. L. and V.K.T. designed the study, analyzed data, and wrote the manuscript. S.C. supervised analysis on the ScaiVision platform and related data analysis workflows.

COMPETING INTERESTS

ScaiVision is Scailyte's proprietary data analysis platform.

METHODS

Processing and Annotation of scRNA-seq

Single-cell RNA sequencing data obtained through droplet-based 10x Genomics technology was selectively utilised for meaningful comparisons in the analysis (Supplementary Table 1). Raw count matrices consisting of Unique Molecular Identifiers (UMIs) and corresponding cell metadata were aggregated from diverse sources.

Quality Control

Preliminary quality control procedures were conducted for each individual dataset using the Seurat R package (v4.0). Cells selected for subsequent analysis were those with more than 200 detected genes and genes identified in a minimum of 3 cells. To identify and eliminate potential cell doublets, scDblFinder v1.6.0 was applied with default parameters. Cells manifesting a mitochondrial transcript content exceeding 20% were also excluded to mitigate confounding effects.

Normalisation and Transformation

Normalisation was achieved by scaling UMI counts with respect to library size, followed by a log transformation to stabilise variance. Subsequently, the dataset underwent Principal Components Analysis (PCA) to reduce dimensionality.

Clustering and Visualisation

Utilising the first 30 principal components, a nearest neighbour graph was constructed, facilitating subsequent clustering via the Louvain algorithm at a lower resolution (FindClusters, resolution = 0.2). For illustrative purposes, uniform manifold approximation and projection (UMAP) embeddings were produced based on 30 principal components, enabling effective visualisation.

Cell Annotation

SingleR v1.6.1 in conjunction with the Human Primary Cell Atlas from the celldex R package (v1.2.0) was employed to accurately annotate individual cell identities within each sample. Importantly, the references for annotation encompassed expected cell types within the tumour microenvironment and BrM

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.

GeneSet Extraction

Following selection of the top learner (AUC Train =1 and AUC Validation =0.98) The gene set extraction was performed using the Captum framework(12).

GeneSet Scoring

To score gene sets, we employed the UCell R package (v1.0.0). UCell scores are computed using the Mann-Whitney U statistic. This rank-based approach ensures that scores remain independent of the dataset's cell-type composition. These scores provide interpretable information about the gene set's relative significance within each cell's transcriptome.

Pan-cancer TCGA Pre-processing

The bulk RNA-seq profiles and clinical outcome data were downloaded from the online repository at https://gdc.cancer.gov/node/905. For each sample, RNA-seq count data underwent normalisation to counts per million, followed by a logarithmic transformation to achieve stability in the data.

Pan-cancer TCGA BrM Signature Calculation

To quantify metastatic signature scores, the average gene-level Z score of cancer cell-specific genes was computed for each sample. This calculation provided a measure of metastatic signature activity within each sample.

BrM Signature Association with Clinical Outcomes

We assessed the relationship between metastatic signature activity and progression-free interval (PFI) using a Cox proportional hazards model. This model incorporated tumour type, purity, stage, and age as covariates, along with continuous metastatic scores. These analyses aimed to uncover potential associations between metastatic signature activity and clinical outcomes. To address the issue of multiple comparisons, P values were adjusted using the Benjamini-Hochberg method. This correction ensured robustness in our statistical inferences while considering the possibility of chance findings.

Spatial Transcriptomics Analysis

Spatial data was downloaded from GSE179572. The UMI count matrix underwent processing using the R package Seurat (v4.0) with default parameters. Initially, the data were normalised using sctransform, and principal component analysis was conducted to reduce the dimensionality of the log-transformed gene-barcode matrices of top variable genes. To pinpoint molecular features correlating with spatial location within a tissue, Seurat was employed for differential expression based on pre-annotated anatomical regions within the tissue. Subsequently, the metastatic score for each ST spot was calculated using UCell.

CellChat

To evaluate cellular interactions between different high and low metastatic cells, we applied CellChat (v.1.1.3)(27) to infer ligand-receptor interactions from the scRNA-seq data. We used the normalised count data as an input with default parameters and CellChatDB.human as the interaction database. Cellular interactions were visualised using the *netVisual_circle* function.

RNA-seq Analysis

Raw RNA-seq data from GSE68086 were downloaded and processed using STAR. The reads were mapped to the hg38 reference genome and. Subsequent normalisation was performed using DESeq2.

CellOracle

We built gene regulatory networks (GRNs) with CellOracle (v.10.10) following tutorials from https://morris-lab.github.io/CellOracle.documentation/.

MAFB *in silico* Perturbation

In silico perturbation was performed using CellOracle(54) as previously defined using default parameters with MAFB expression set to 0.

ASGARD Drug Repurposing Screen

We used the normalised count data as an input, with metastatic low-scoring cells set as control in ASGARD(38) at default parameters. We calculated individual drug scores for each cancer type to identify drugs with a significant score across each.

Statistical Analysis

All the statistical analyses were performed using R (version 4.1.1). Student's *t*-test, Wilcoxon rank-sum test and Kaplan–Meier were utilised in this study. *p*-values of less than 0.05 were considered statistically significant.

FIGURE LEGENDS

Figure 1. Identification of Gene Signature Predictive of BrM

A) Graphical representation of the study B) UMAP projection of pan-cancer BrM scRNA-seq data annotated for cancer types and label used for ScaiVision Analysis C) ScaiVision Results highlighting model accuracy and predictive capabilities of the top network D) UMAP projection with cells scored from low to high BrM potential using UCell E) DEG analysis of low vs high BrM scored cells F) GO terms of genes with higher expression in BrM high cells.

Figure 2. Spatial transcriptomics scoring of colorectal BrM samples

A-B) Spatial transcriptomics data of colorectal BrM tissue sections, annotated with cell-type classifications adapted from the original study. C-D) UCell scoring of BrM signature within the spatially resolved tissue unveils. Notable amplification in select regions, particularly observed in the blood vessels and periphery of tumour cells.

Figure 3. BrM signature is Expressed Across Multiple Cancer Types and Associated with Reduced RFS

A) Expression of BrM signature across each cancer type and matched normal tissue in the TCGA pan-cancer cohort. Red indicates significantly higher expression in cancer B) Distribution of BrM signature score of each TCGA pan-cancer patient and stratification into high and low BrM potential C) Survival plot based-on stratification of patients based on expression of BrM signature in TCGA pan-cancer cohort **(**Kaplan–Meier)

Figure 4. BrM signature is Expressed in Metastatic Patients Blood Samples

A) Expression of BrM signature in pan-cancer metastatic and non-metastatic patients' tumour educated platelets RNA-seq. B) Expression of refined BrM signature in pan-cancer metastatic and non-metastatic patients' tumour educated platelets RNA-seq C) Expression of refined signature in patients with brain metastasis vs other sites of metastasis D) GO terms of genes of refined BrM signature

Figure 5. Identification of VEGF signalling as a key driver of communication networks in BrM cells

A) High vs low signalling pathways, driving cell-cell communications in both BrM and primary samples. Each bar represents the strength of signalling pathway in either high (red) or low (blue). Pathways are coloured on y-axis based on their significance for either cohort B) Communication networks via VEGF signalling in metastatic high scored cells in both BrM and primary samples C) Top ligand interactions driving VEGF signalling in metastatic high cells D) Communication networks via VEGF signalling in metastatic high scored cells in primary

samples only E) Communication networks via VEGF signalling in metastatic high scored cells in BrM samples only.

Figure 6. Identification of MAFB as a key driver of BrM progression in high-scored cells
A) BrM Stage GRN TFs B) Network dynamics of MAFB across metastatic stages C) Expression of MAFB in normal, tumour and metastatic RNA-seq D) Survival plots of MAFB in TCGA pan-cancer

Figure 7. *In silico* Perturbation of MAFB in Breast Cancer Reverse BrM Progression
A) BrM Stage GRN TFs B) transition matrix showing cellular dynamics from low to high BrM potential C) in silico perturbation of MAFB alters BrM progression, as denoted by arrows length and change in direction from predicted progression shown in 7.B

Figure 8. *In silico* drug repurposing analysis reveals FDA approved drugs to target BrM scored cells
A) Results of our *in silico* drug repurposing analysis conducted using ASGARD. Each dot represents an FDA approved drug with a score in targeting BrM high scored cells, with the size of the dot reflecting the drug score, and colour indicating the corresponding cancer type. B) Signalling Pathways Targeted by Identified Drugs: A notable presence of VEGF signalling-associated pathways is observed, suggesting the potential relevance of angiogenesis and vascular remodelling processes in the context of brain metastasis and the potential for targeting these pathways

Supplementary Fig 1 Identification of BrM Signature
A) UMAP showing epithelial sub-set of pan-cancer dataset B) Top models derived from ScaiVision C) Contribution of each gene in prediction BrM or Primary in learner 12 (Top Learner) D) UCell scoring of pan-cancer dataset containing only epithelial cells E) expression of BrM signature across all cell types F) expression of BrM signature across all cell types.

Supplementary Fig 2 Cell Chat Analysis
A) Total signeting interactions and strengths between all cells in BrM High vs Low B) Overall Signalling pathways activity in High and Low scored cells

Supplementary Fig 3 GRN Analysis
A) RXRG activity across BrM Timepoints B) GRN Networks of MAFB and RXRG showing common target genes

REFERENCES

1. Achrol AS, Rennert RC, Anders C, Soffietti R, Ahluwalia MS, Nayak L, et al. Brain metastases. Nat Rev Dis Prim [Internet]. 2019; 5(1):5. Available from: https://www.nature.com/articles/s41572-018-0055-y.
2. Wang Z, Wang Y, Chang M, Wang Y, Liu P, Wu J, et al. Single-cell transcriptomic analyses provide insights into the cellular origins and drivers of brain metastasis from lung adenocarcinoma. Neuro Oncol [Internet]. 2023; 25(7):1262–74. Available from: https://academic.oup.com/neuro-oncology/article/25/7/1262/6992983.
3. Tabouret E, Chinot O, Metellus P, Tallet A, Viens P, Gonçalves A. Recent trends in epidemiology of brain metastases: an overview. Anticancer Res [Internet]. 2012; 32(11):4655–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23155227.
4. Steindl A, Brastianos PK, Preusser M, Berghoff AS. Precision medicine biomarkers in brain metastases: applications, discordances, and obstacles. Neuro-Oncology Adv [Internet]. 2021; 3(Supplement_5):v35–42. Available from: https://academic.oup.com/noa/article/3/Supplement_5/v35/6444218.
5. Song Q, Ruiz J, Xing F, Lo H-W, Craddock L, Pullikuth AK, et al. Single-cell sequencing reveals the landscape of the human brain metastatic microenvironment. Commun Biol [Internet]. 2023; 6(1):760. Available from: https://www.nature.com/articles/s42003-023-05124-2.
6. Aaltonen LA, Abascal F, Abeshouse A, Aburatani H, Adams DJ, Agrawal N, et al. Pan-cancer analysis of whole genomes. Nature [Internet]. 2020; 578(7793):82–93. Available from: https://www.nature.com/articles/s41586-020-1969-6.
7. Cooper LA, Demicco EG, Saltz JH, Powell RT, Rao A, Lazar AJ. PanCancer insights from The Cancer Genome Atlas: the pathologist's perspective. J Pathol [Internet]. 2018; 244(5):512–24. Available from: https://pathsocjournals.onlinelibrary.wiley.com/doi/10.1002/path.5028.
8. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet [Internet]. 2013; 45(10):1127–33. Available from: https://www.nature.com/articles/ng.2762.
9. Cindy Yang SY, Lien SC, Wang BX, Clouthier DL, Hanna Y, Cirlan I, et al. Pan-cancer analysis of longitudinal metastatic tumors reveals genomic alterations and immune landscape dynamics associated with pembrolizumab sensitivity. Nat Commun [Internet]. 2021; 12(1):5137. Available from: https://www.nature.com/articles/s41467-021-25432-7.
10. Gonzalez H, Mei W, Robles I, Hagerling C, Allen BM, Hauge Okholm TL, et al. Cellular architecture of human brain metastases. Cell [Internet]. 2022; 185(4):729-745.e20.

Available from: https://linkinghub.elsevier.com/retrieve/pii/S0092867421015658.

11. Fisher NC, Byrne RM, Leslie H, Wood C, Legrini A, Cameron AJ, et al. Biological Misinterpretation of Transcriptional Signatures in Tumor Samples Can Unknowingly Undermine Mechanistic Understanding and Faithful Alignment with Preclinical Data. Clin Cancer Res [Internet]. 2022; 28(18):4056–69. Available from: https://aacrjournals.org/clincancerres/article/28/18/4056/708987/Biological-Misinterpretation-of-Transcriptional.
12. Kokhlikyan N, Miglani V, Martin M, Wang E, Alsallakh B, Reynolds J, et al. Captum: A unified and generic model interpretability library for PyTorch [Internet]. 2020. Available from: http://arxiv.org/abs/2009.07896.
13. Andreatta M, Carmona SJ. UCell: Robust and scalable single-cell gene signature scoring. Comput Struct Biotechnol J [Internet]. 2021; 19:3796–8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2001037021002816.
14. Fiorillo M, Ózsvári B, Sotgia F, Lisanti MP. High ATP Production Fuels Cancer Drug Resistance and Metastasis: Implications for Mitochondrial ATP Depletion Therapy. Front Oncol [Internet]. 2021; 11. Available from: https://www.frontiersin.org/articles/10.3389/fonc.2021.740720/full.
15. Cheng M, Jiang Y, Xu J, Mentis A-FA, Wang S, Zheng H, et al. Spatially resolved transcriptomics: a comprehensive review of their technological advances, applications, and challenges. J Genet Genomics [Internet]. 2023; 50(9):625–40. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1673852723000759.
16. Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. Nature [Internet]. 2021; 596(7871):211–20. Available from: https://www.nature.com/articles/s41586-021-03634-9.
17. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet [Internet]. 2013; 45(10):1113–20. Available from: http://www.nature.com/articles/ng.2764.
18. Morotti A, Gentile F, Lopez G, Passignani G, Valenti L, Locatelli M, et al. Epigenetic Rewiring of Metastatic Cancer to the Brain: Focus on Lung and Colon Cancers. Cancers (Basel) [Internet]. 2023; 15(7):2145. Available from: https://www.mdpi.com/2072-6694/15/7/2145.
19. Hasanov E, Yeboa DN, Tucker MD, Swanson TA, Beckham TH, Rini B, et al. An interdisciplinary consensus on the management of brain metastases in patients with renal cell carcinoma. CA Cancer J Clin [Internet]. 2022; 72(5):454–89. Available from: https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.21729.
20. 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 [Internet].

2018; 19(10):2877. Available from: http://www.mdpi.com/1422-0067/19/10/2877.

21. Uttley L, Whiteman BL, Woods HB, Harnan S, Philips ST, Cree IA. Building the Evidence Base of Blood-Based Biomarkers for Early Detection of Cancer: A Rapid Systematic Mapping Review. eBioMedicine [Internet]. 2016; 10:164–73. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2352396416303097.
22. Huh SJ, Liang S, Sharma A, Dong C, Robertson GP. Transiently Entrapped Circulating Tumor Cells Interact with Neutrophils to Facilitate Lung Metastasis Development. Cancer Res [Internet]. 2010; 70(14):6071–82. Available from: https://aacrjournals.org/cancerres/article/70/14/6071/559651/Transiently-Entrapped-Circulating-Tumor-Cells.
23. Zhang W, Gu J, Chen J, Zhang P, Ji R, Qian H, et al. Interaction with neutrophils promotes gastric cancer cell migration and invasion by inducing epithelial-mesenchymal transition. Oncol Rep [Internet]. 2017; 38(5):2959–66. Available from: https://www.spandidos-publications.com/10.3892/or.2017.5942.
24. Bergfeld SA, DeClerck YA. Bone marrow-derived mesenchymal stem cells and the tumor microenvironment. Cancer Metastasis Rev [Internet]. 2010; 29(2):249–61. Available from: http://link.springer.com/10.1007/s10555-010-9222-7.
25. Bussard KM, Mutkus L, Stumpf K, Gomez-Manzano C, Marini FC. Tumor-associated stromal cells as key contributors to the tumor microenvironment. Breast Cancer Res [Internet]. 2016; 18(1):84. Available from: https://breast-cancer-research.biomedcentral.com/articles/10.1186/s13058-016-0740-2.
26. Khan IM, Khan SU, Sala HSS, Khan MU, Ud Din MA, Khan S, et al. TME-targeted approaches of brain metastases and its clinical therapeutic evidence. Front Immunol [Internet]. 2023; 14. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1131874/full.
27. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C-H, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun [Internet]. 2021; 12(1):1088. Available from: https://www.nature.com/articles/s41467-021-21246-9.
28. Kim LS, Huang S, Lu W, Chelouche Lev D, Price JE. Vascular endothelial growth factor expression promotes the growth of breast cancer brain metastases in nude mice. Clin Exp Metastasis [Internet]. 2004; 21(2):107–18. Available from: http://link.springer.com/10.1023/B:CLIN.0000024761.00373.55.
29. Yano S, Shinohara H, Herbst RS, Kuniyasu H, Bucana CD, Ellis LM, et al. Expression of vascular endothelial growth factor is necessary but not sufficient for production and growth of brain metastasis. Cancer Res [Internet]. 2000; 60(17):4959–67. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10987313.
30. Claffey KP, Brown LF, Aguila LF del, Tognazzi K, Yeo KT, Manseau EJ, et al.

Expression of vascular permeability factor/vascular endothelial growth factor by melanoma cells increases tumor growth, angiogenesis, and experimental metastasis. Cancer Res [Internet]. 1996; 56(1):172–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8548760.

31. Mohammed RAA, Green A, El-Shikh S, Paish EC, Ellis IO, Martin SG. Prognostic significance of vascular endothelial cell growth factors -A, -C and -D in breast cancer and their relationship with angio- and lymphangiogenesis. Br J Cancer [Internet]. 2007; 96(7):1092–100. Available from: https://www.nature.com/articles/6603678.
32. Perrot-Applanat M, Benedetto M Di. Autocrine functions of VEGF in breast tumor cells. Cell Adh Migr [Internet]. 2012; 6(6):547–53. Available from: http://www.tandfonline.com/doi/abs/10.4161/cam.23332.
33. Ceci C, Atzori MG, Lacal PM, 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 [Internet]. 2020; 21(4):1388. Available from: https://www.mdpi.com/1422-0067/21/4/1388.
34. Jeong H-W, Hernández-Rodríguez B, Kim J, Kim K-P, Enriquez-Gasca R, Yoon J, et al. Transcriptional regulation of endothelial cell behavior during sprouting angiogenesis. Nat Commun [Internet]. 2017; 8(1):726. Available from: https://www.nature.com/articles/s41467-017-00738-7.
35. Pawlikowski B, Wragge J, Siegenthaler JA. Retinoic acid signaling in vascular development. genesis [Internet]. 2019; 57(7–8). Available from: https://onlinelibrary.wiley.com/doi/10.1002/dvg.23287.
36. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov [Internet]. 2004; 3(8):673–83. Available from: http://www.nature.com/articles/nrd1468.
37. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov [Internet]. 2019; 18(1):41–58. Available from: http://www.nature.com/articles/nrd.2018.168.
38. He B, Xiao Y, Liang H, Huang Q, Du Y, Li Y, et al. ASGARD is A Single-cell Guided Pipeline to Aid Repurposing of Drugs. Nat Commun [Internet]. 2023; 14(1):993. Available from: https://www.nature.com/articles/s41467-023-36637-3.
39. Sternberg CN, Davis ID, Mardiak J, Szczylik C, Lee E, Wagstaff J, et al. Pazopanib in Locally Advanced or Metastatic Renal Cell Carcinoma: Results of a Randomized Phase III Trial. J Clin Oncol [Internet]. 2023; 41(11):1957–64. Available from: https://ascopubs.org/doi/10.1200/JCO.22.02622.
40. Venur VA, Ahluwalia MS. Targeted Therapy in Brain Metastases: Ready for Primetime? Am Soc Clin Oncol Educ B [Internet]. 2016; (36):e123–30. Available from:

https://ascopubs.org/doi/10.1200/EDBK_100006.

41. AlTamimi JO, AlJohani HA, Naaman N, Johar RA, Allam TA, Lary DAI. Brain Metastases in Adults: A Five-Year Observational Study From King Abdulaziz Medical City. Cureus [Internet]. 2022. Available from: https://www.cureus.com/articles/120084-brain-metastases-in-adults-a-five-year-observational-study-from-king-abdulaziz-medical-city.
42. Fiorillo M, Scatena C, Naccarato AG, Sotgia F, Lisanti MP. Bedaquiline, an FDA-approved drug, inhibits mitochondrial ATP production and metastasis in vivo, by targeting the gamma subunit (ATP5F1C) of the ATP synthase. Cell Death Differ [Internet]. 2021; 28(9):2797–817. Available from: https://www.nature.com/articles/s41418-021-00788-x.
43. Zhou Y, Tozzi F, Chen J, Fan F, Xia L, Wang J, et al. Intracellular ATP Levels Are a Pivotal Determinant of Chemoresistance in Colon Cancer Cells. Cancer Res [Internet]. 2012; 72(1):304–14. Available from: https://aacrjournals.org/cancerres/article/72/1/304/575796/Intracellular-ATP-Levels-Are-a-Pivotal-Determinant.
44. Schneider V, Krieger ML, Bendas G, Jaehde U, Kalayda G V. Contribution of intracellular ATP to cisplatin resistance of tumor cells. JBIC J Biol Inorg Chem [Internet]. 2013; 18(2):165–74. Available from: http://link.springer.com/10.1007/s00775-012-0960-6.
45. Termini J, Neman J, Jandial R. Role of the Neural Niche in Brain Metastatic Cancer. Cancer Res [Internet]. 2014; 74(15):4011–5. Available from: https://aacrjournals.org/cancerres/article/74/15/4011/592685/Role-of-the-Neural-Niche-in-Brain-Metastatic.
46. Yang Y, Cao Y. The impact of VEGF on cancer metastasis and systemic disease. Semin Cancer Biol [Internet]. 2022; 86:251–61. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1044579X22000670.
47. Moon EJ, Mello SS, Li CG, Chi J-T, Thakkar K, Kirkland JG, et al. The HIF target MAFF promotes tumor invasion and metastasis through IL11 and STAT3 signaling. Nat Commun [Internet]. 2021; 12(1):4308. Available from: https://www.nature.com/articles/s41467-021-24631-6.
48. Dieterich LC, Tacconi C, Menzi F, Proulx ST, Kapaklikaya K, Hamada M, et al. Lymphatic MAFB regulates vascular patterning during developmental and pathological lymphangiogenesis. Angiogenesis [Internet]. 2020; 23(3):411–23. Available from: https://link.springer.com/10.1007/s10456-020-09721-1.
49. Deng Y, Lu L, Zhang H, Fu Y, Liu T, Chen Y. The role and regulation of Maf proteins in cancer. Biomark Res [Internet]. 2023; 11(1):17. Available from:

https://biomarkerres.biomedcentral.com/articles/10.1186/s40364-023-00457-w.

50. Dieterich LC, Klein S, Mathelier A, Sliwa-Primorac A, Ma Q, Hong Y-K, et al. DeepCAGE Transcriptomics Reveal an Important Role of the Transcription Factor MAFB in the Lymphatic Endothelium. Cell Rep [Internet]. 2015; 13(7):1493–504. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2211124715011419.

51. Liu Y, Li Y, Wang Y, Lin C, Zhang D, Chen J, et al. Recent progress on vascular endothelial growth factor receptor inhibitors with dual targeting capabilities for tumor therapy. J Hematol Oncol [Internet]. 2022; 15(1):89. Available from: https://jhoonline.biomedcentral.com/articles/10.1186/s13045-022-01310-7.

52. Hamberg P, Verweij J, Sleijfer S. (Pre-)Clinical Pharmacology and Activity of Pazopanib, a Novel Multikinase Angiogenesis Inhibitor. Oncologist [Internet]. 2010; 15(6):539–47. Available from: https://academic.oup.com/oncolo/article/15/6/539/6399187.

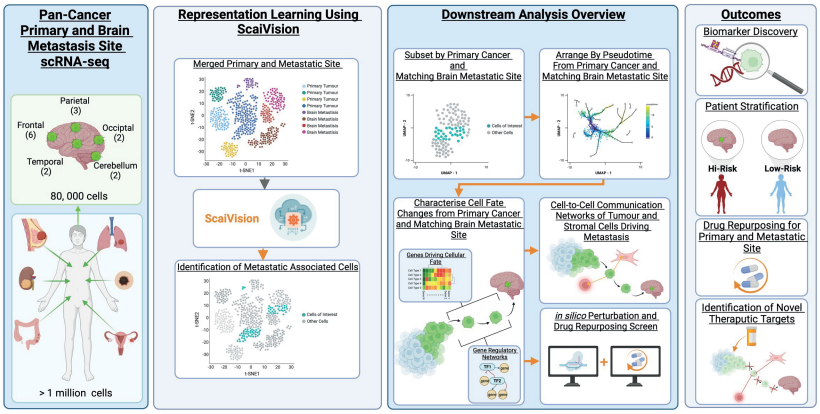
53. Mollinedo F. Neutrophil Degranulation, Plasticity, and Cancer Metastasis. Trends Immunol [Internet]. 2019; 40(3):228–42. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1471490619300067.

54. Kamimoto K, Stringa B, Hoffmann CM, Jindal K, Solnica-Krezel L, Morris SA. Dissecting cell identity via network inference and in silico gene perturbation. Nature [Internet]. 2023; 614(7949):742–51. Available from: https://www.nature.com/articles/s41586-022-05688-9.

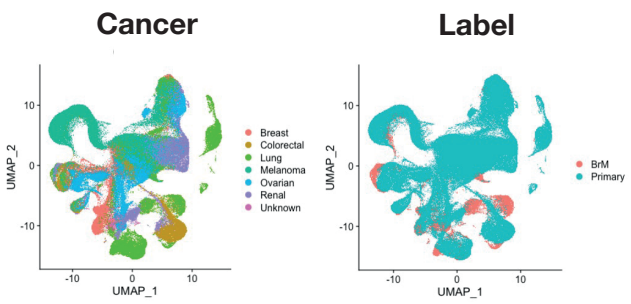
Figure 1

Lusby et al

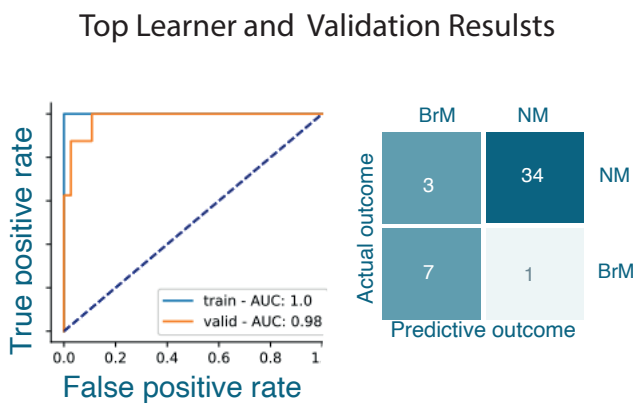
A



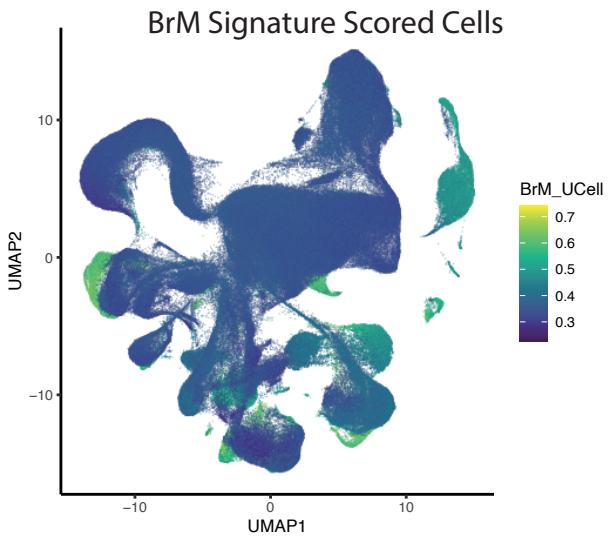
B



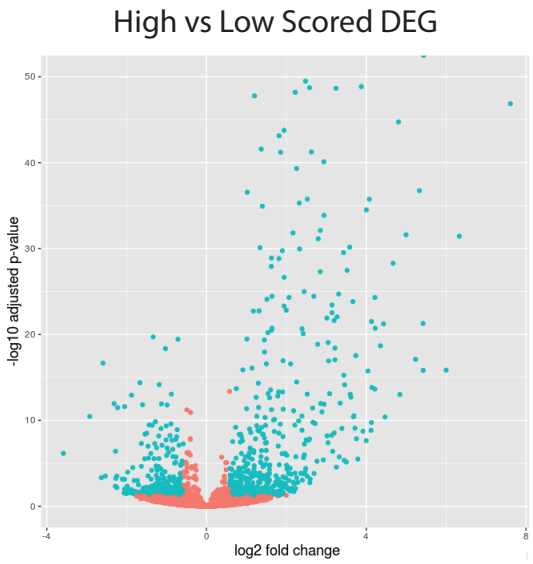
C



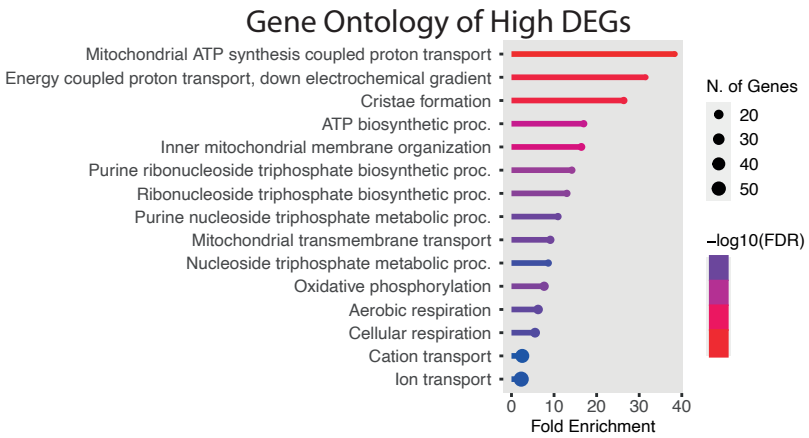
D



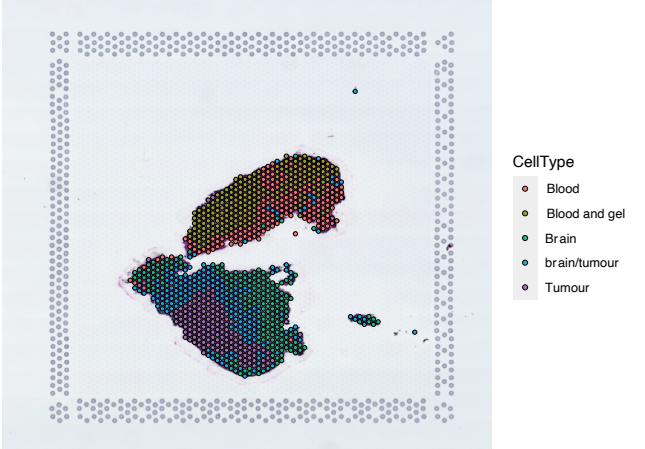
E



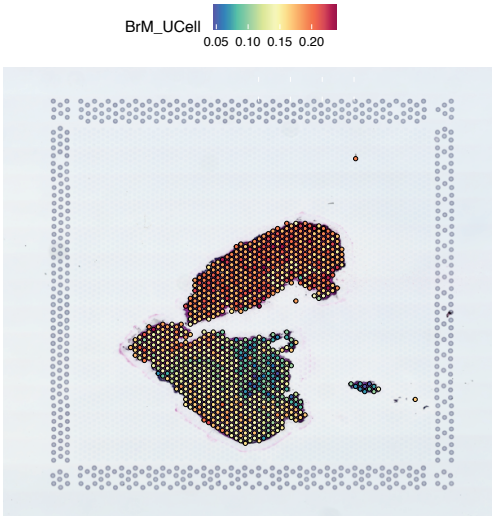
F



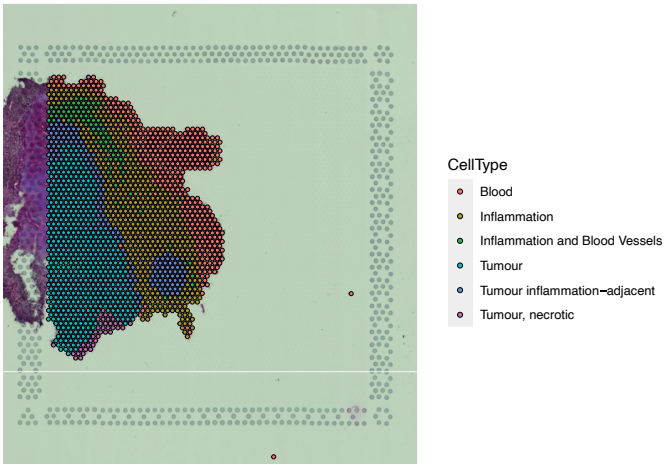
A


CellType
Blood
Blood and gel
Brain
brain/tumour
Tumour

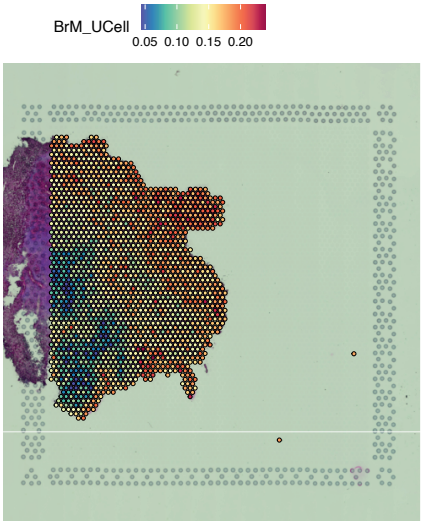
C


BrM_UCell
0.05 0.10 0.15 0.20

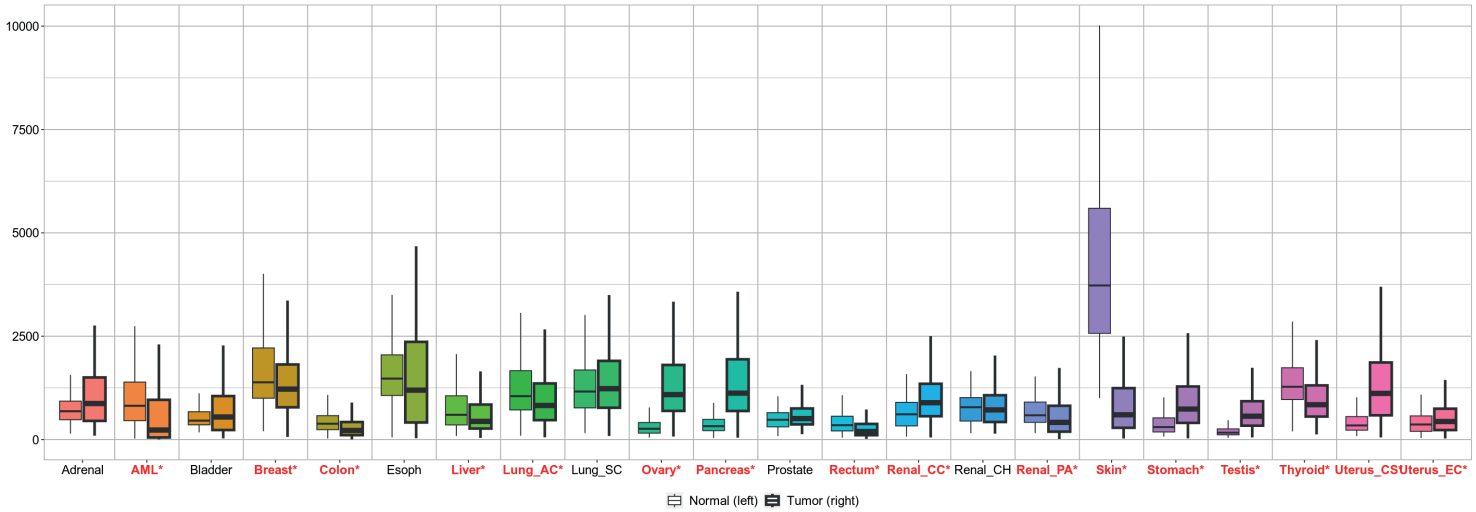
B


CellType
Blood
Inflammation
Inflammation and Blood Vessels
Tumour
Tumour inflammation−adjacent
Tumour, necrotic

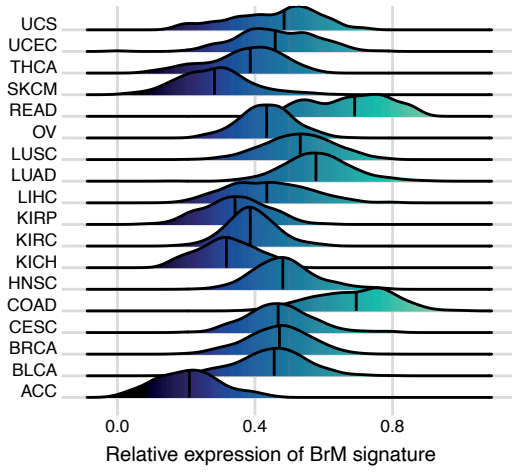
D


BrM_UCell
0.05 0.10 0.15 0.20

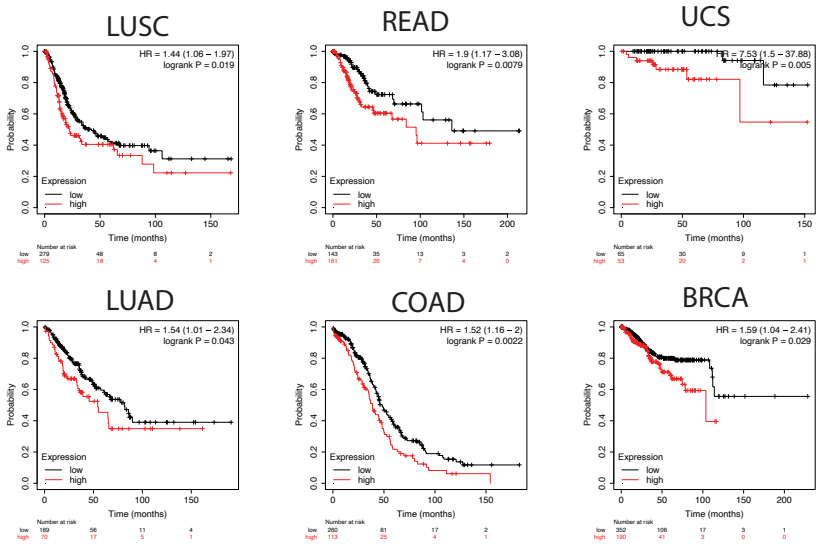
A


10000
7500
5000
2500
0
Adrenal
AML*
Bladder
Breast*
Colon*
Esoph
Liver*
Lung_AC*
Lung_SC
Ovary*
Pancreas*
Prostate
Rectum*
Renal_CC*
Renal_CH
Renal_PA*
Skin*
Stomach*
Testis*
Thyroid*
Uterus_CS
Uterus_EC*
Normal (left)
Tumor (right)

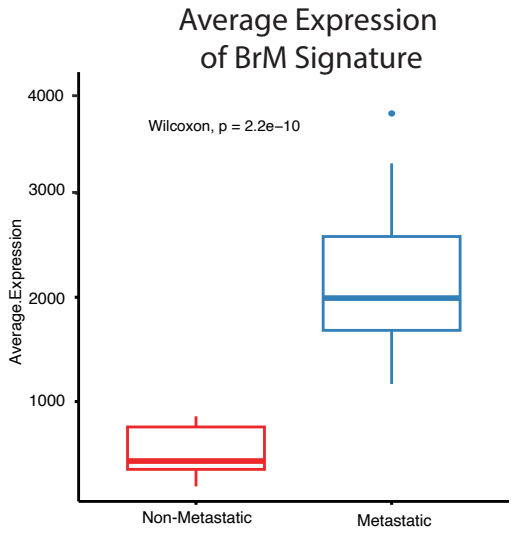
B


UCS
UCEC
THCA
SKCM
READ
OV
LUSC
LUAD
LIHC
KIRP
KIRC
KICH
HNSC
COAD
CESC
BRCA
BLCA
ACC
0.0
0.4
0.8
Relative expression of BrM signature

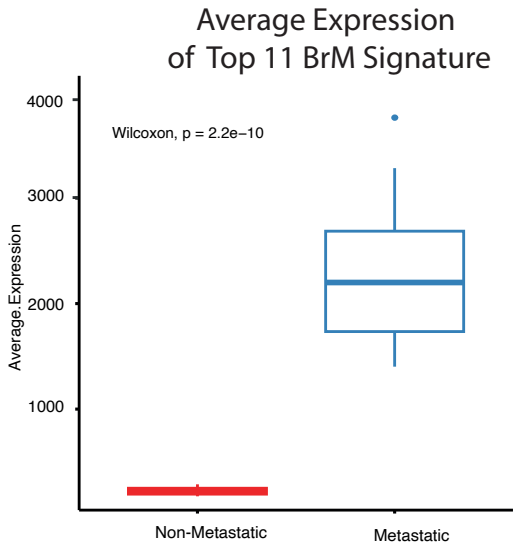
C


LUSC
HR = 1.44 (1.06 – 1.97)
logrank P = 0.019
READ
HR = 1.9 (1.17 – 3.08)
logrank P = 0.0079
UCS
HR = 7.53 (1.5 – 37.88)
logrank P = 0.005
LUAD
HR = 1.54 (1.01 – 2.34)
logrank P = 0.043
COAD
HR = 1.52 (1.16 – 2)
logrank P = 0.0022
BRCA
HR = 1.59 (1.04 – 2.41)
logrank P = 0.029
Probability
Time (months)
Expression
low
high
Number at risk

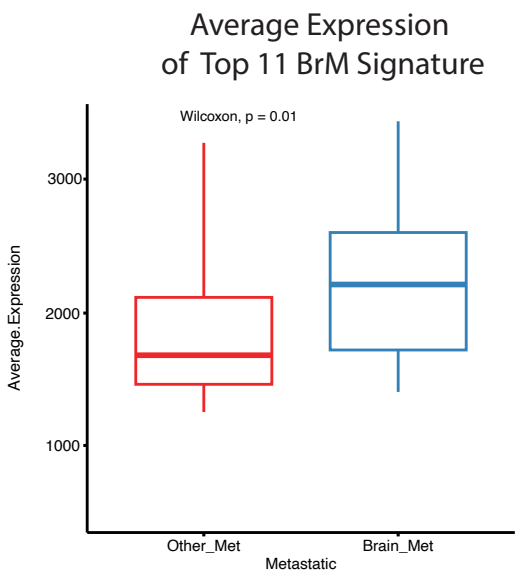
A


Average Expression
of BrM Signature
Wilcoxon, p = 2.2e−10
Average.Expression
4000
3000
2000
1000
Non-Metastatic
Metastatic

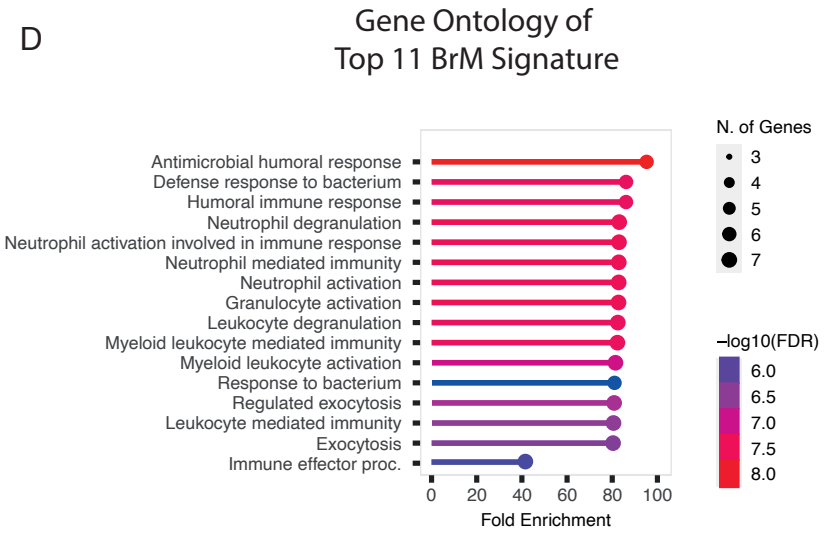
B


Average Expression
of Top 11 BrM Signature
Wilcoxon, p = 2.2e−10
Average.Expression
4000
3000
2000
1000
Non-Metastatic
Metastatic

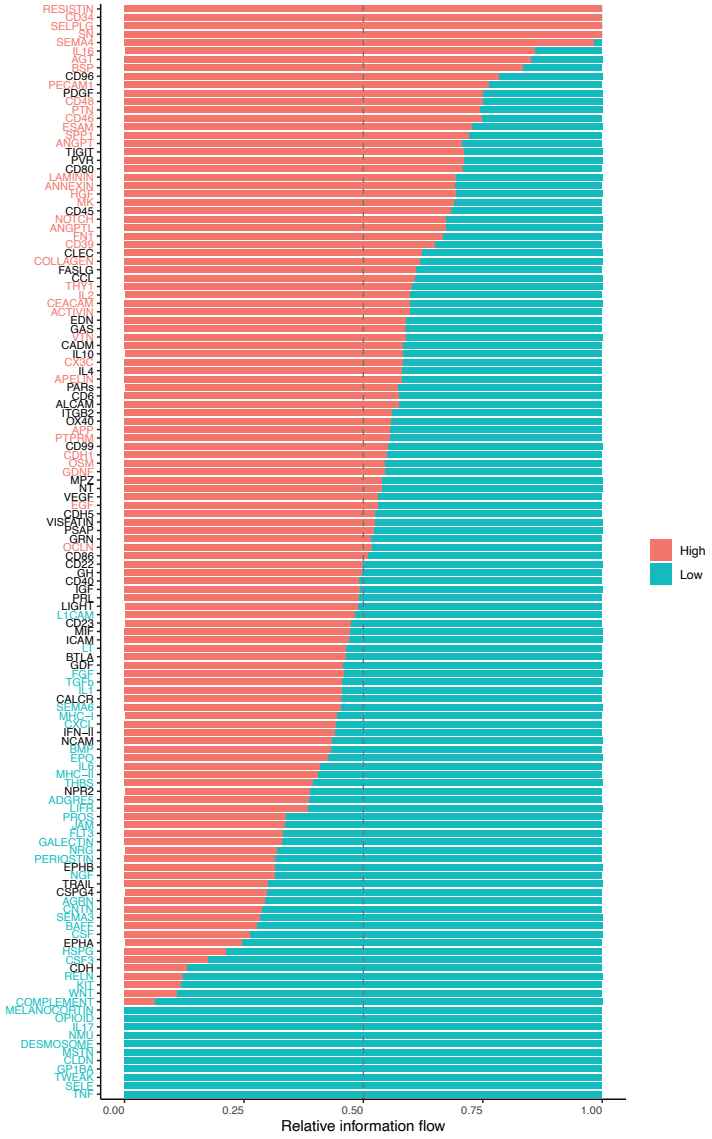
C


Average Expression
of Top 11 BrM Signature
Wilcoxon, p = 0.01
Average.Expression
3000
2000
1000
Other_Met
Brain_Met
Metastatic

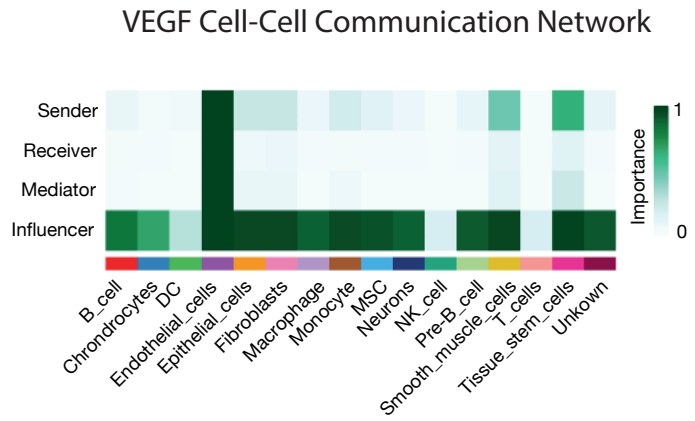
D


Gene Ontology of
Top 11 BrM Signature
Antimicrobial humoral response
Defense response to bacterium
Humoral immune response
Neutrophil degranulation
Neutrophil activation involved in immune response
Neutrophil mediated immunity
Neutrophil activation
Granulocyte activation
Leukocyte degranulation
Myeloid leukocyte mediated immunity
Myeloid leukocyte activation
Response to bacterium
Regulated exocytosis
Leukocyte mediated immunity
Exocytosis
Immune effector proc.
0
20
40
60
80
100
Fold Enrichment
N. of Genes
3
4
5
6
7
−log10(FDR)
6.0
6.5
7.0
7.5
8.0

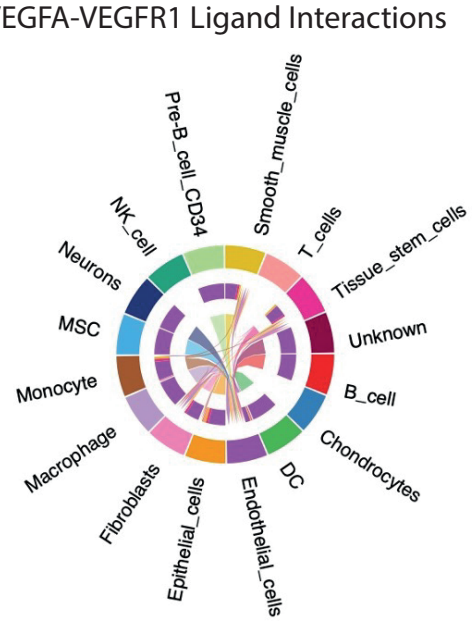
A


High
Low
0.00
0.25
0.50
0.75
1.00
Relative information flow

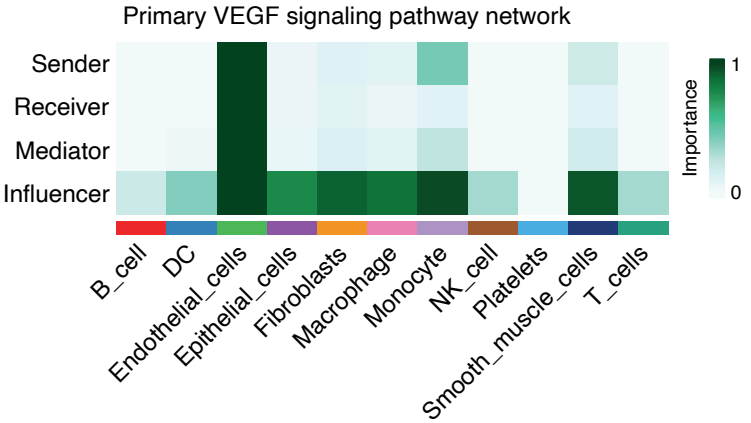
B


VEGF Cell-Cell Communication Network
Sender
Receiver
Mediator
Influencer
Importance
1
0
B_cell
Chrondrocytes
DC
Endothelial_cells
Epithelial_cells
Fibroblasts
Macrophage
Monocyte
MSC
Neurons
NK_cell
Pre-B_cell
Smooth_muscle_cells
T_cells
Tissue_stem_cells
Unkown

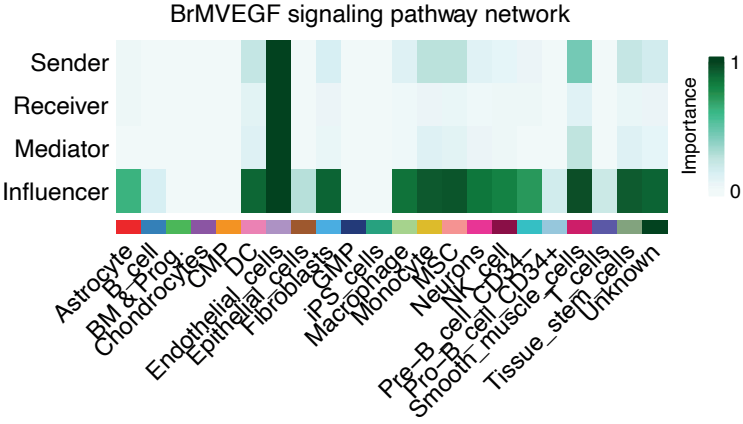
C

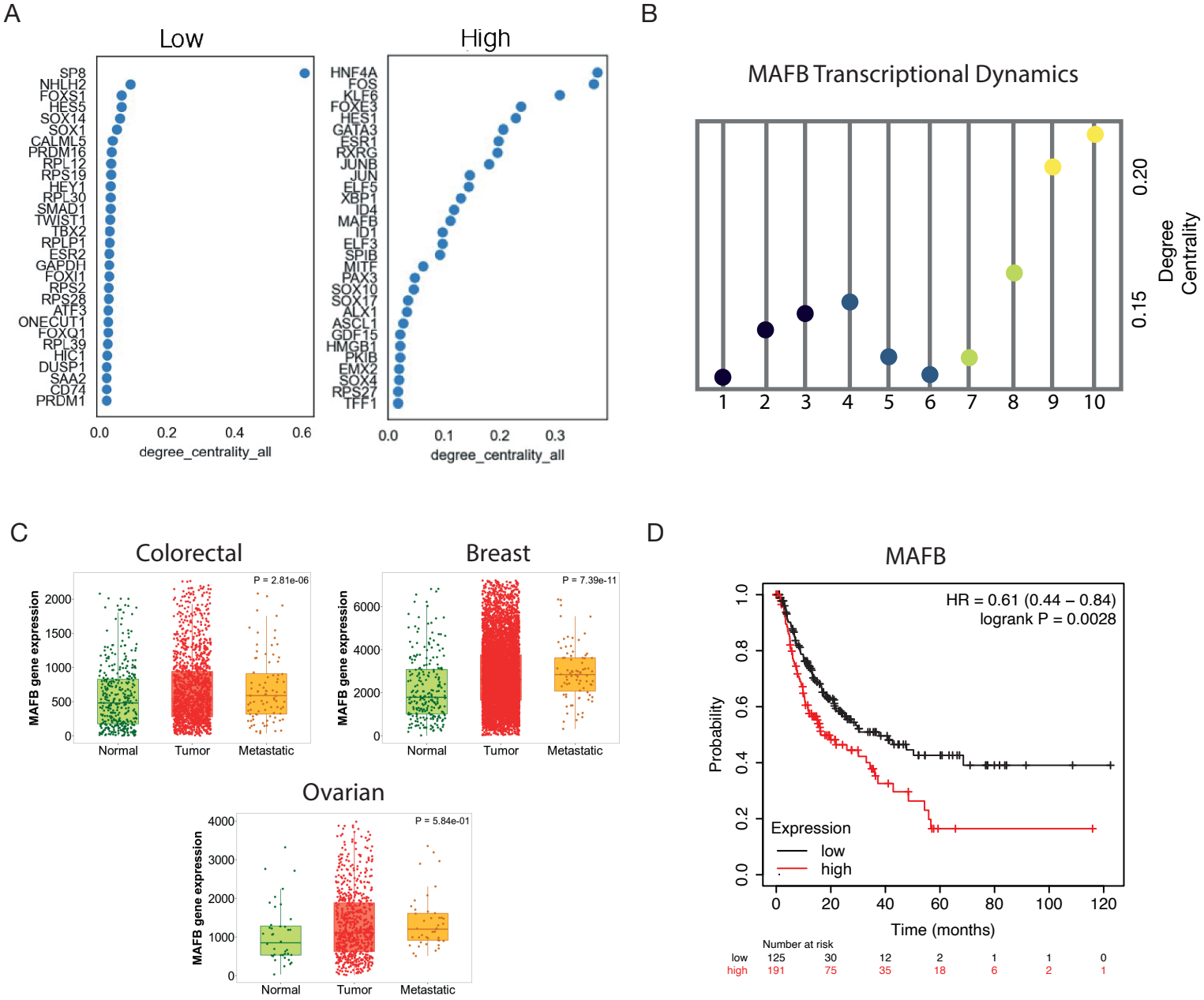

VEGFA-VEGFR1 Ligand Interactions
Pre-B_cell_CD34
Smooth_muscle_cells
NK_cell
T_cells
Neurons
Tissue_stem_cells
MSC
Unknown
Monocyte
B_cell
Macrophage
Chondrocytes
Fibroblasts
DC
Epithelial_cells
Endothelial_cells

D

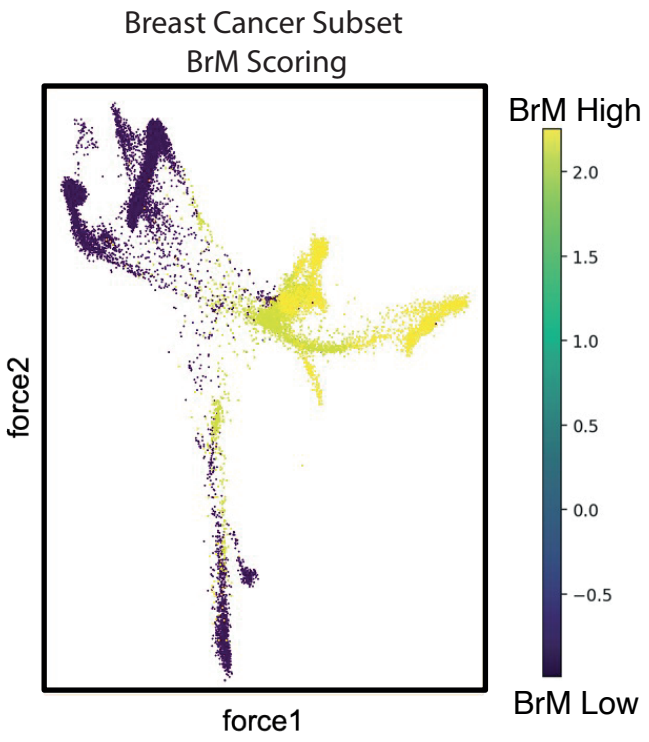

Primary VEGF signaling pathway network
Sender
Receiver
Mediator
Influencer
Importance
1
0
B_cell
DC
Endothelial_cells
Epithelial_cells
Fibroblasts
Macrophage
Monocyte
NK_cell
Platelets
Smooth_muscle_cells
T_cells

E

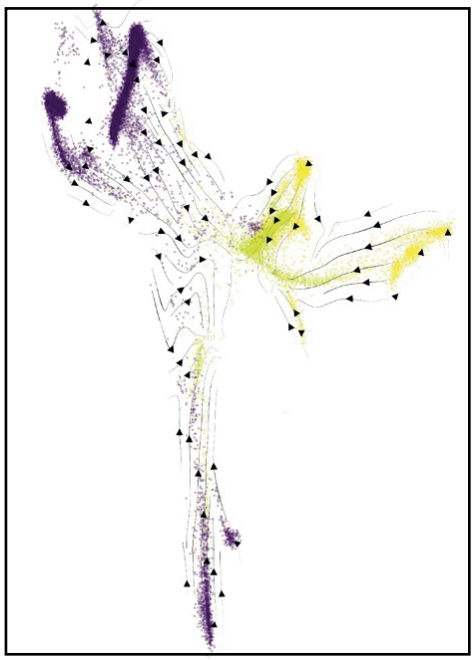

BrMVEGF signaling pathway network
Sender
Receiver
Mediator
Influencer
Importance
1
0
Astrocyte
B_cell
BM & Prog.
Chondrocytes
CMP
DC
Endothelial_cells
Epithelial_cells
Fibroblasts
GMP
iPS_cells
Macrophage
Monocyte
MSC
Neurons
NK_cell
Pre-B_cell_CD34-
Pro-B_cell_CD34+
Smooth_muscle_cells
T_cells
Tissue_stem_cells
Unknown


A
Low
High
SP8
NHLH2
FOXS1
HES5
SOX14
SOX1
CALML5
PRDM16
RPL12
RPS19
HEY1
RPL30
SMAD1
TWIST1
TBX2
RPLP1
ESR2
GAPDH
FOXI1
RPS2
RPS28
ATF3
ONECUT1
FOXQ1
RPL39
HIC1
DUSP1
SAA2
CD74
PRDM1
0.0 0.2 0.4 0.6
degree_centrality_all
HNF4A
FOS
KLF6
FOXE3
HES1
GATA3
ESR1
RXRG
JUNB
JUN
ELF5
XBP1
ID4
MAFB
ID1
ELF3
SPIB
MITF
PAX3
SOX10
SOX17
ALX1
ASCL1
GDF15
HMGB1
PKIB
EMX2
SOX4
RPS27
TFF1
0.0 0.1 0.2 0.3
degree_centrality_all
B
MAFB Transcriptional Dynamics
0.20
0.15
Degree Centrality
1 2 3 4 5 6 7 8 9 10
C
Colorectal
P = 2.81e-06
MAFB gene expression
0 500 1000 1500 2000
Normal Tumor Metastatic
Breast
P = 7.39e-11
MAFB gene expression
0 2000 4000 6000
Normal Tumor Metastatic
Ovarian
P = 5.84e-01
MAFB gene expression
0 1000 2000 3000 4000
Normal Tumor Metastatic
D
MAFB
HR = 0.61 (0.44 – 0.84)
logrank P = 0.0028
Probability
0.0 0.2 0.4 0.6 0.8 1.0
Expression
low
high
0 20 40 60 80 100 120
Time (months)
Number at risk
low 125 30 12 2 1 1 0
high 191 75 35 18 6 2 1

A


Breast Cancer Subset
BrM Scoring
BrM High
2.0
1.5
1.0
0.5
0.0
−0.5
BrM Low
force2
force1

B


Cell Fate Aligns with BrM scoring

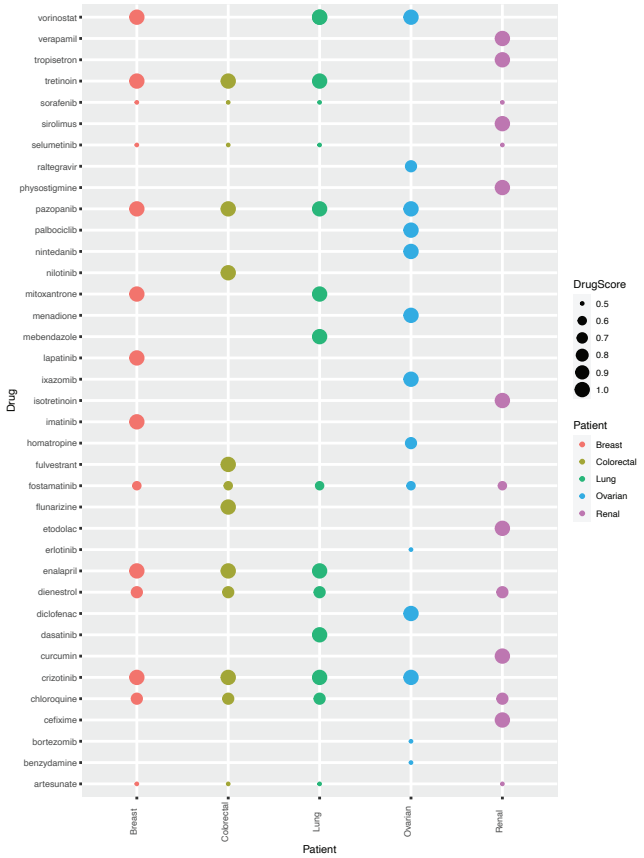
C

MAFB GRN KO

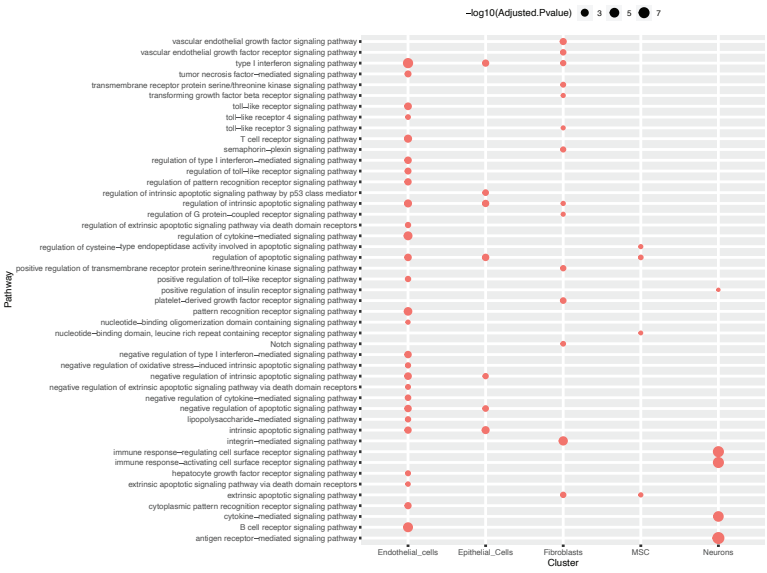
Figure 8

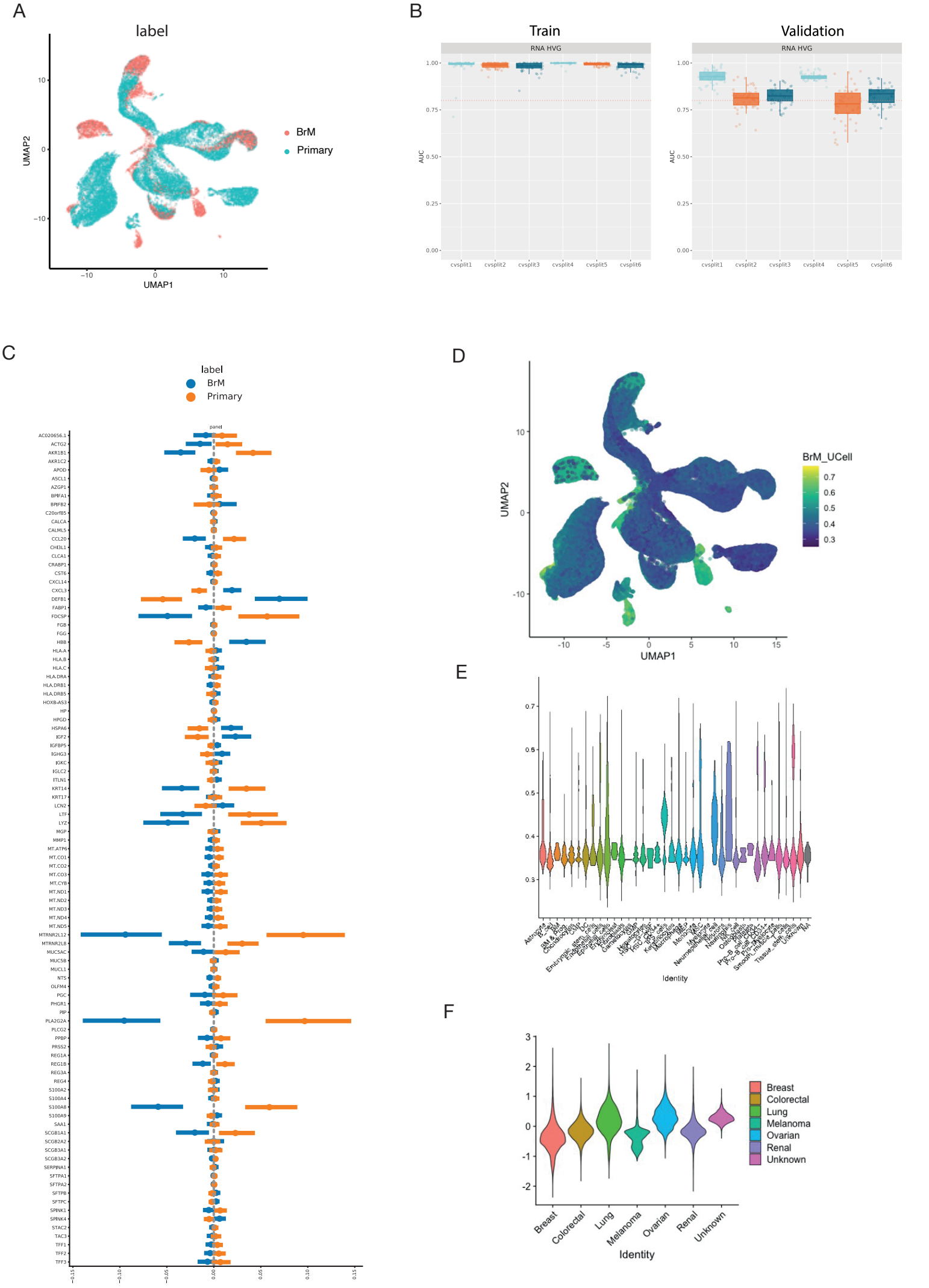
Lusby et al

A



B

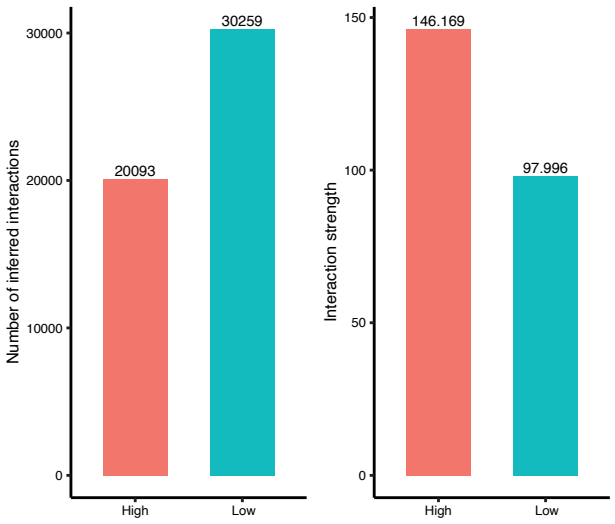



A
label
BrM
Primary
UMAP2
UMAP1
B
Train
Validation
RNA HVG
AUC
cvsplit1
cvsplit2
cvsplit3
cvsplit4
cvsplit5
cvsplit6
C
label
BrM
Primary
panel
D
BrM_UCell
UMAP2
UMAP1
E
Identity
F
Breast
Colorectal
Lung
Melanoma
Ovarian
Renal
Unknown
Identity

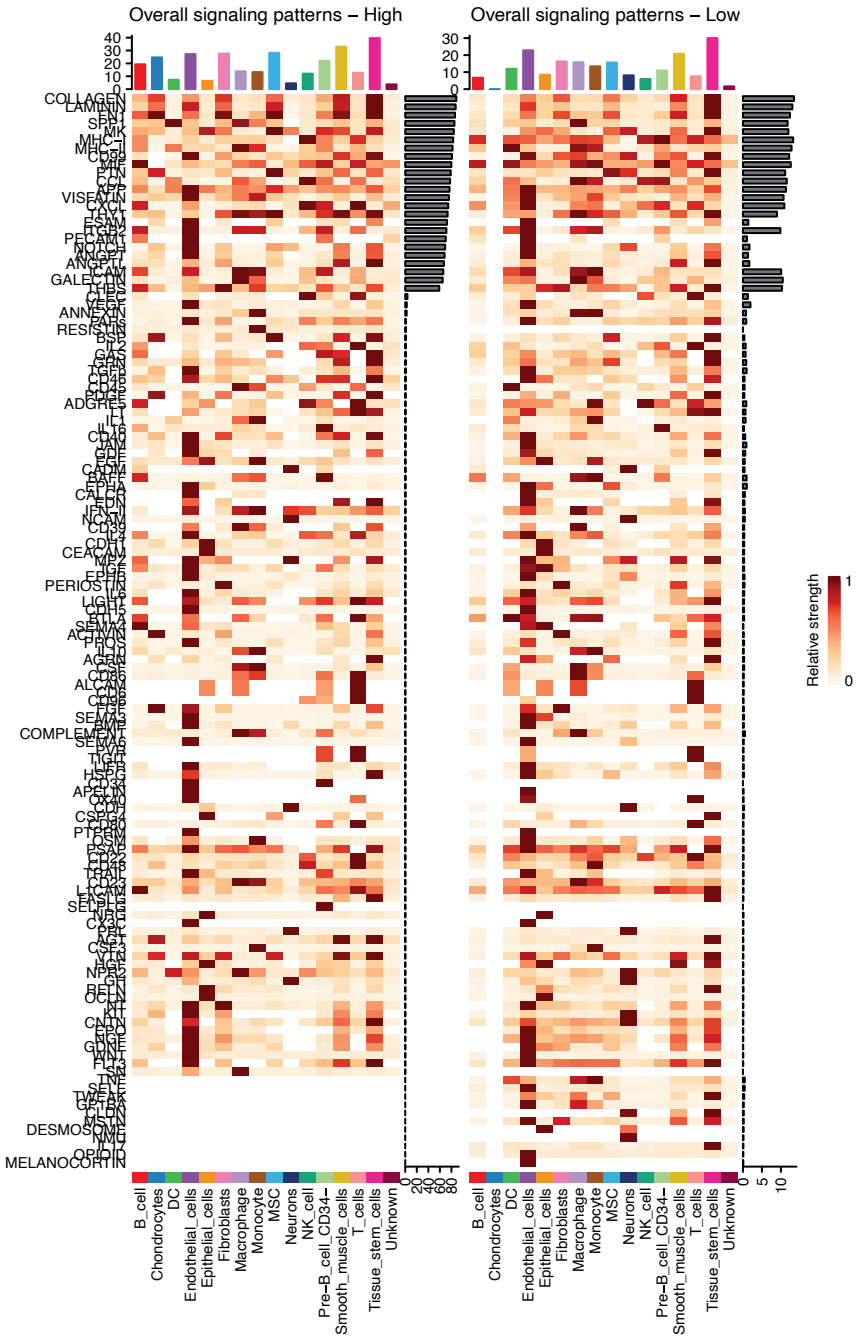
Supplementary Figure 2

Lusby et al

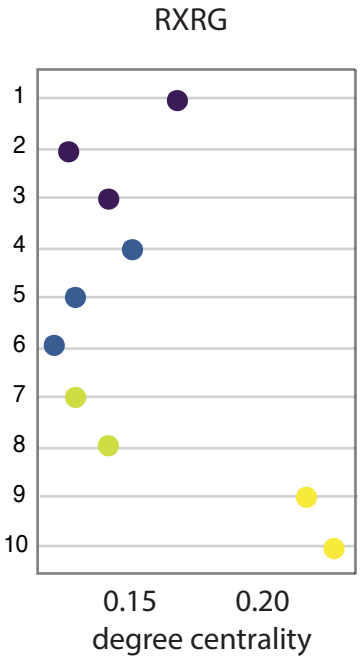
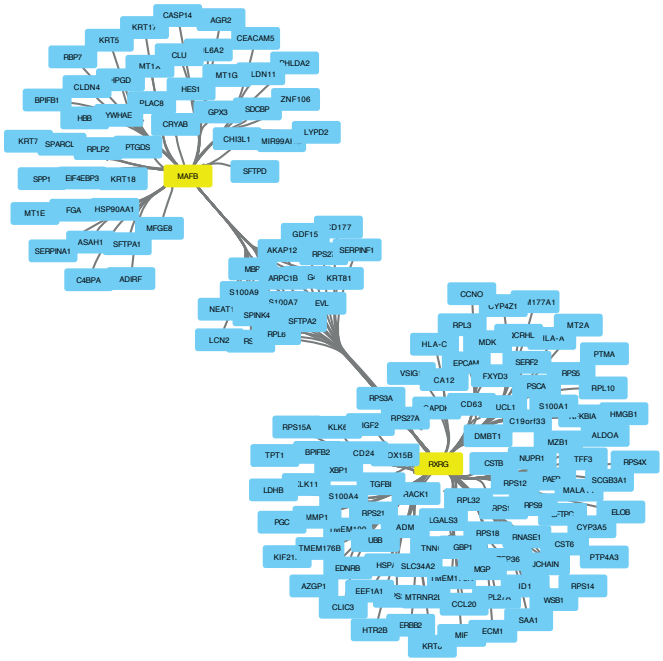
A



B



A



B