



LUND
UNIVERSITY

Discovery of Novel Fibroblast Subtypes in Glioblastoma Using Single-Cell RNA Sequencing Analysis

Ida L. Neinhardt

Bachelor's Programme in Biomedicine, Lund University, Lund, Sweden

Biomedical Research Project, Bachelor's Degree (BIMK60)

DATE OF EXAMINATION

Monday, 22nd May 2023 at 10:00

SUPERVISOR

Jonas Sjölund

Division of Translational Cancer Research, Department of Laboratory Medicine, Lund University Cancer Centre, Lund, Sweden

EXPERT OPPONENT

Dimitra Manou

Division of Translational Cancer Research, Department of Laboratory Medicine, Lund University Cancer Centre, Lund, Sweden

STUDENT OPPONENT

Marina Peltonen

Bachelor's Programme in Biomedicine, Lund University, Lund, Sweden

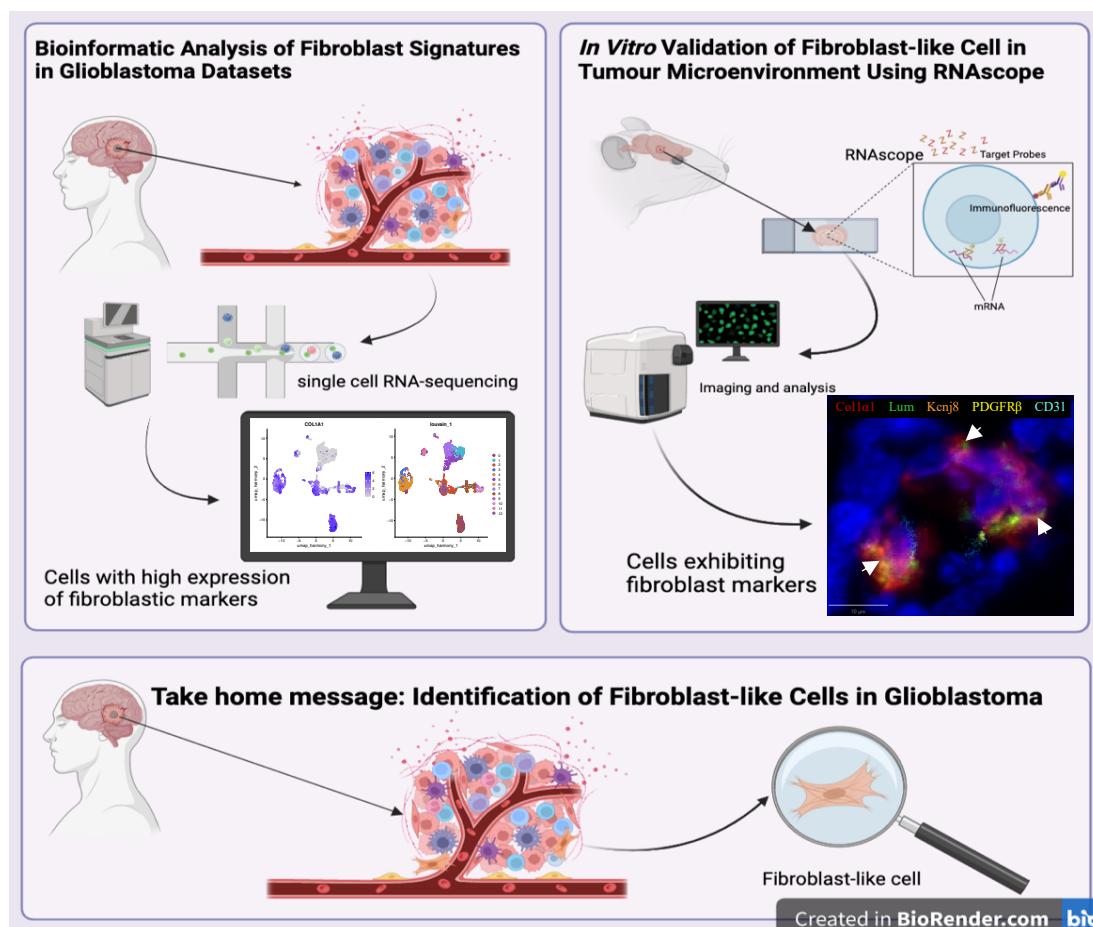
Table of Contents

Abstract	3
Graphical Abstract.....	3
Introduction	4
Materials and Methods	6
<i>Animal Model.....</i>	<i>6</i>
<i>Data Availability.....</i>	<i>6</i>
<i>Re-clustering of Endothelial Cells and Pericytes</i>	<i>6</i>
<i>Applying Gene Signatures.....</i>	<i>7</i>
<i>Kaplan-Meier Survival Curves and Biological Analysis.....</i>	<i>7</i>
<i>RNAscope</i>	<i>8</i>
<i>Scanning and Image Analysis.....</i>	<i>8</i>
<i>Statistical Analysis.....</i>	<i>8</i>
<i>Ethical Aspects</i>	<i>8</i>
Results	8
<i>Fibroblast Genes Expressed Higher Than Pericyte Genes in Abdelfattah's Pericyte Cluster</i>	<i>8</i>
<i>Gene Signatures Suggest a Fibroblast Genotype in Pericyte Clusters</i>	<i>10</i>
<i>Annotation of Clusters Using ScType</i>	<i>13</i>
<i>CAF Signatures of Potential Fibroblast Subtypes Plotted on Abdelfattah's Dataset</i>	<i>14</i>
<i>RNAscope Staining Showing Expression of Fibroblast Genes in mouse model of GBM</i>	<i>16</i>
<i>Validation of Findings in Additional Datasets</i>	<i>18</i>
<i>Population Survival and Biological Analysis Based on New Cluster Analysis</i>	<i>19</i>
Discussion	21
<i>Limitations of the Study.....</i>	<i>22</i>
Acknowledgements.....	23
References	24
Comments to Key References	31
Supplementary Materials	32

Abstract

Glioblastoma is the most common form of glioma and is also one of the most lethal cancers. Despite extensive research, much remains unknown about glioblastoma due to its highly invasive nature and the intricate microenvironment that it creates within the brain. The glioblastoma microenvironment is complex, involving various types of cells that interact with one another. These interactions contribute to a multitude of functions including tumour growth and invasion. Though fibroblasts are known to be present in the microenvironment of many other cancers, it has remained unknown if they exist within glioblastoma. The current study aims to examine this, and the results proposed that they do exist, showing on a previously unidentified fibroblast-like cell-type as being present in the tumour microenvironment of glioblastoma. This was seen in single cell RNA-sequencing analyses and RNAscope experiments where fibroblast-like cells were found in glioblastoma datasets and mouse models. A subtype of these cells with a myofibroblastic phenotype was additionally shown to correlate with a worse prognosis for patients with the disease. Though further studies are needed to confirm their role within disease progression, a new cell type has been suggested which could become a target for potential future treatments.

Graphical Abstract



Introduction

Glioblastoma (GBM) is the most common form of glioma, accounting for 69% of documented cases (1). It is also one of the most lethal cancers found in the central nervous system (CNS), being a grade IV malignant glioma according to the WHO (World Health Organization) classification of tumours, with an approximate 5% survival rate after 5 years and a median survival of 15 months (2,3). There are many symptoms of GBM, including headaches and nausea, cognition changes, personality changes, visual field defects, and seizures, however they often remain unspecific until the cancer has progressed beyond what is currently treatable, which is one of the reasons for the poor survival rate (4).

An increased risk for the development of GBM has been seen in males as well as in patients with certain rare hereditary syndromes such as Li-Fraumeni (4). Risk factors have been studied for their correlation to gliomas and the only well-established environmental risk factor is ionizing radiation, especially when exposure occurs during childhood.

Cellularly, GBM is characterised by proliferation of the vasculature, mitotic activity and a high level of necrosis, with its cells being very irregular both in their shape and size (5). GBM is additionally a highly invasive cancer which often leads to much of the surrounding brain parenchyma being infiltrated by the tumour. Intriguingly, GBM has not so far been seen to venture beyond the CNS and metastasis is improbable.

The GBM tumour microenvironment (TME) has in recent years been highlighted as an important part of the GBM pathology and, as such, it becomes important to get an understanding of its components and functions to get a comprehensive understanding of this disease (6). It has been shown that the TME has influence over many of the characteristic mechanisms contributing to the lethality of GBM, including invasion and tumour growth. Many questions remain of its structure and function, however, one major question is if there are fibroblasts associated with the GBM TME (7)*.

Fibroblasts are stromal cells which play an essential role in the synthesis of the extracellular matrix (ECM) that provides structural integrity in many tissues (8). There exists a large transcriptomic heterogeneity within this overarching cell type suggesting specialised functions of different cells, however, much remains to be elucidated of what the biological consequences of these differences actually are.

Cancer associated fibroblasts (CAFs) are major constituents of the TME in many cancers (9)*. Current knowledge suggests that there is a large diversity and heterogeneity among CAFs and there is not just one but many subpopulations of this cell type. The different subtypes each have specific gene expression patterns and diverse functionalities. Additionally, it has also been shown that their spatial location within the TME differs. Different names have been given to CAFs identified by different research groups, though there is a remarkable overlap of the documented features, suggesting that there might be more universal CAF populations than first postulated (9–12). CAFs have been identified in breast, pancreatic, lung, and renal cell cancer, however, if they are also present within the brain, and in cancers associated with it, remains to be revealed.

Fibroblast-like cells have been identified in the perivascular space in the brain where they are implicated in the barrier between the cerebrospinal fluid and the brain (13). Additionally, two subtypes of these fibroblasts have been described, showing distinct clustering and differing in their expressed markers. Fibroblasts have also been shown to be present in the meninges where they protect the brain and spinal cord by generating components of basement membrane supporting the brain parenchyma, creating fibrotic scars that induce wound healing, and inducing a CNS immune response (14,15). Furthermore, fibroblasts have been observed within the choroid plexus where they are thought to trigger immune cell recruitment and activation, and contribute to oedema in response to head trauma (16).

Pericytes are cells that are known to lay close to blood-vessels and support the structure of these (17). They have been shown to promote resistance to the current most effective drug for GBM, temozolomide, and further contribute to progression of the disease. One major issue with studying pericytes, however, is that there are few known pericyte-specific markers, making what one might think of as a pericyte actually being a smooth muscle cell (SMC) or even a fibroblast (18).

In recent years, single-cell RNAseq (scRNA-seq) has become a highly useful tool to analyse individual cells to determine their transcriptome and thereby classify and characterise them into previously unknown and rare, but important, cell populations (19). In one single experiment, a researcher is able to analyse hundreds of thousands of cells with one of the many described scRNA-seq technologies. This high resolution of analysis has proven incredibly useful for e.g., the determination of the previously discussed CAF subtypes (9–11).

As it remains unknown if fibroblasts exist within the brain tissue or if CAFs are present in the GBM TME like they do in many other cancer types, it is impossible to develop treatments that would target these in GBM. This study aims to look for CAFs in the GBM microenvironment using described and evaluated markers and bioinformatic approaches. The hypothesis of this thesis work is that cells that resemble CAFs will be found within datasets of GBM and its microenvironment. Due to the small portion of fibroblasts that is expected to be found, it could be difficult to elucidate if subtypes of CAFs are present or if differences between cells are due to other variable factors; regardless, this will also be examined. Additionally, multiplex stainings will be performed that possibly could reveal cells with fibroblast markers within a GBM mouse model.

Whether fibroblasts exist within the brain parenchyma and/or in the brain TME is something that remains controversial (20). One of the main reasons for this uncertainty is the fact that there are no described fibroblast markers that are specific enough to differentiate fibroblasts from other alike cell types like pericytes, especially within the brain. Recently, an article published in The Journal of Clinical Investigation claimed to have found CAFs within patient samples of GBM (21)*. The researchers behind the paper stated that through CAF purification from GBM tumour samples and bioinformatic analysis to have found fibroblasts in the GBM TME. However, it should be noted that, as previously mentioned, there are no known specific markers for fibroblasts within the brain parenchyma and this study used markers that signify fibroblasts and CAFs outside the current area of interest. Additionally, the markers used by the scientists do not fully line up with the markers suggested by literature of

best differentiating CAFs from other cell types. The finding of CAFs in GBM patients, if more evidence is provided, could revolutionise the way the GBM TME is researched. They could provide a potential target for future treatments, and as there today are few available therapies for GBM, none of which leaving patients with a high chance of survival, new treatments are irrefutably needed. This thesis project focuses on finding further evidence for fibroblasts within the GBM TME and could show additional confirmation for the hypothesis laid out by the previously mentioned article, i.e., that also in cancers within the CNS, CAF-like cells exist.

Materials and Methods

Animal Model

Tissue used for the RNAscope part of this project was obtained from adult mice with induced high-grade glioma. The induction took place through *in vitro* *Pdgfβ* overexpression and *Tp53* knockdown in mammalian cells using short hairpin RNAs before transplantation of these into the subventricular zone of mouse brains. For tissue retrieval, the mice were anaesthetised with isoflurane prior to being euthanised by decapitation. The brain tissues were quick-frozen over dry ice in optimal cutting temperature compound. The fresh-frozen tissues were sectioned using a microtome into 10 μ m sections and placed on SuperFrost Plus glass slides. The samples were stored at -80°C until further use.

Data Availability

The raw scRNA-seq data used for the majority of this project was acquired as a publicly available human GBM dataset published by Abdelfattah et al. (SRA accession code PRJNA787981) (22)*. The analyses in this study mainly used a subset of the dataset from Abdelfattah et al. which was comprised of cells classified as endothelial cells or pericytes by the authors. The pericytes were of predominant interest, however, the endothelial cells were also included due to their known interactions with pericytes and fibroblasts, as well as to provide a reliable control when examining gene expression in the cells.

For validation, a dataset by Pietras et al. was used. The dataset contained induced GBM single-cell data from mice. The induction of GBM took place similarly as described in the animal model part of this methods section with the difference being that one group of mice had normal pericyte expression and the other was pericyte-poor. As the presence of pericytes was not of main interest in this project, the dataset was not divided based on pericyte expression. The dataset will be available upon publication.

Re-clustering of Endothelial Cells and Pericytes

For the re-clustering of the subset of the human GBM dataset, the preparatory steps were completed as outlined by Abdelfattah et al. on the raw data of the full dataset in R (22). For this, cells that were identified to be of low-quality were removed. These were cells that contained <500 expressed genes, >50% ribosomal transcripts, and/or >20% mitochondrial transcripts. Probable doublets, i.e., two cells that have been sequenced as one, were removed

using the DoubleFinder v2 package following the developer's directives of using a 7.5% expected doublet rate assuming Poisson statistics (23). The Seurat v4 package was used to normalise the dataset before batch-correcting it via integration using Harmony v1 (24,25). Following this, the gene expression of the single-cells was normalised and log₂-transformed. To reduce the dimensionality, the top 2000 most variable genes were used for principal component (PC) analyses. Batch correction for patient and sex in the computed PCs was completed using Harmony. The integration of the data was verified with the do_DimPlot function of the SCpubr v1 package with the cells grouped by patient (26). The PCs were used for clustering using the Louvain algorithm and the resolutions parameters tested were between 0.1 and 2.0. Resolutions of interest were identified as 0.2 and 1.0, and the dataset was clustered using the Seurat v4 package and its FindClusters function (24)*. The colour palette that was used for the clustering graphs was created using the software *i want hue* available on GitHub (27).

To visualise the movement of cells between clusters in increased resolutions to decide the resolutions to go forth with for further investigations, the Clustree package v0.5 was used (28).

For validation purposes, a package used for automatically naming clusters based on gene expression, ScType v1, was run on the human GBM data (29). This method is based on reference datasets and uses known genetic signatures to make its predictions. The basic pipeline outlined by the creator of the package was followed.

Applying Gene Signatures

The gene signatures used to indicate what cell types were present in the analysed datasets were from a brain scRNA-seq article published by Winkler et al. (30). The CAF signature lists used for comparison of this dataset with other cancers where the concept of CAFs is well established, were acquired from articles published by Elyada et al., and Wu et al. (10,31).

To add each signature to the dataset, the function AddModuleScore from the Seurat v4 package was implemented and feature plots were subsequently plotted using the FeaturePlot_scCustom function from the scCustomize v1 package in R (32). Cellular state plots for the signatures were made using the SCpubr v1 package and its function do_CellularStatesPlot (26). The heatmaps also utilised the SCpubr package, implementing the do_EnrichmentHeatmap function.

Kaplan-Meier Survival Curves and Biological Analysis

The Cancer Genome Atlas' (TCGA's) GBM mRNA expression and clinical data (Firehose Legacy) were downloaded from the cBioPortal database (33,34). Overall survival (OS) and disease-specific survival (DSS) data were used to analyse clinical outcomes of the human GBM dataset. The R function ggsurvplot from the survminer v0.4 package was used to plot Kaplan-Meier survival curves (35).

clusterProfiler v4 was used for the biological analysis in this project, utilising hallmarks from msigdbr v7 and visualised in a DotPlot to characterise the involvement of each cluster's cells in various biological processes (36–38).

RNAscope

The RNAscope *in situ* hybridisation combined with immunofluorescence staining was performed according to the RNAscope Fluorescent Multiplex Assay protocol (Advanced Cell Diagnostics, USA) (39,40).

Scanning and Image Analysis

Slides were scanned in the Vectra Polaris machine from Akoya Biosciences. For analysis the softwares Phenochart v1, inForm v2 and QuPath v0.4 were used. Phenochart was used to define regions of interest while QuPath and inForm were used for image processing.

Statistical Analysis

For the analysis of TCGA data, the Cox proportional hazards model was used for statistical analysis and was corrected for age and sex. By using the Cox model, we can examine how various variables affect survival while taking into account potential confounders. The mean expression of the top 20 differentially expressed genes (adjusted p-value) of each cluster was used to group patient samples into low and high groups based on the 25th and 75th percentile, respectively. This analysis provided significance regarding the clinical relevance of the findings in this project.

Other packages used, such as ScType, based their analysis on statistics and gave results in probabilities. The use of these methods has been outlined in their relevant method sections.

Ethical Aspects

There were no new ethical approvals needed to be in place for this study. This is because this project's main focus was on analysis of data which had already been gathered. However, for the studies acquiring the datasets that have been analysed, along with the wet-lab portion of this project, ethical permits were needed.

For the publicly available datasets that were used in this project, the ethical approval has come from the institutional review boards at those institutions in which the original experiment was conducted. The animal experiments conducted prior to the wet-lab portion of this project were approved by the local Ethical Committees for Animal Experimentation in Lund (14122/2020). Additionally, the mouse data was acquired in a way which complied with the Australian Code for the care and use of animals for scientific purposes.

Results

Fibroblast Genes Expressed Higher Than Pericyte Genes in Abdelfattah's Pericyte Cluster

The dataset from Abdelfattah et al., was created using patient samples from diagnoses of astrocytoma, oligodendrogloma, GBM and recurrent GBM. Most cells were from Grade IV tumours, however, there was also a minority that were from Grade II tumours. The workflow performed by Abdelfattah et al., prior to this study included collection of tissue samples, droplet-based scRNA-seq and analysis of the generated dataset (**Figure 1A**).

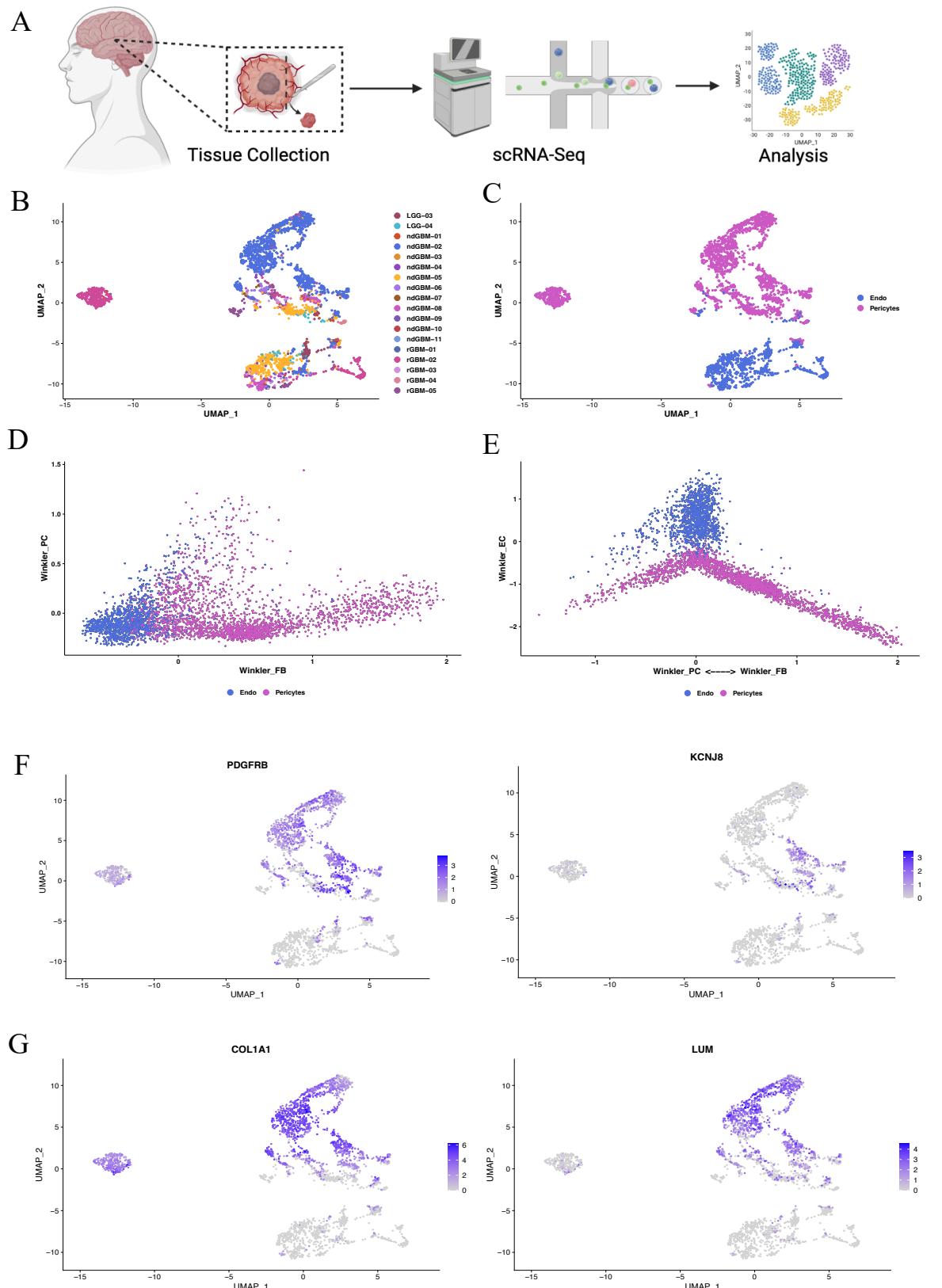


Figure 1: Pericyte assignment does not encapsulate the full nature of cluster in dataset by Abdelfattah et al.

(A) Outline of experiment workflow performed by Abdelfattah et al. prior to the start of this study. Tissue samples from glioma patients were harvested and droplet-based scRNA-seq was performed preceding analysis. Created in BioRender.com.
 (B-C) Visualisation of non-integrated dataset generated by Abdelfattah et al. in a UMAP grouped by (B) patient and (C) assigned cell type, Endo (endothelial cells) in purple and Pericytes in pink.

Legend continued on next side...

(D-E) Cellular state plot showing each cell's features associated with (D) pericyte (PC) and fibroblast (FB), and (E) pericyte (PC), fibroblast (FB), and endothelial cell (EC) signatures proposed by Winkler et al. The cells are grouped by the cell type assignment completed by Abdelfattah et al.

(F) Dimplots of traditional pericyte markers PDGFR β and KCNJ8 on non-integrated dataset.

(G) Dimplots of traditional fibroblast markers COL1 α 1 and LUM on non-integrated dataset.

Abbreviations: GBM: glioblastoma, TME: tumour microenvironment, scRNA-seq: single cell RNA sequencing, UMAP: Uniform Manifold Approximation and Projection

The subset of interest in the dataset, including clusters identified by Abdelfattah et al. as either pericytes or endothelial cells, contained 3580 cells from 18 patients, as seen in the metadata of the dataset, and we visualised these in a Uniform Manifold Approximation and Projection (UMAP) to look at the heterogeneity among the cells (**Figure 1B**). Both male and female samples were used, however the majority of the analysed cells were derived from males with 3320 cells to the females' 260 cells. The non-integrated dataset had clear batch-effects with certain clusters only containing cells from few patients, with in particular patient rGBM-02 and patient ndGBM-02 being very distinct from the others, however, the separation between what had been assigned previously by Abdelfattah et al. as endothelial cells remained separate from the cells assigned as pericytes (**Figure 1C**).

We plotted the cells in a cellular states plot against gene signatures described by Winkler et al. (30). The endothelial cells matched the established gene signature, however, the so-called pericytes did not, and resembled fibroblasts to a greater extent than pericytes, indicating that the assigned pericytes are, at least partly, not pericytes (**Figures 1D and 1E**). There was also a subset of the pericyte cluster which had cells with smooth muscle cell features (**Supplemental Figure 1**). To further confirm the findings in figure 1D and 1E, markers often used to characterise pericytes and fibroblasts were plotted on the non-integrated data (**Figures 1F and 1G**). PDGFR β is one of the most common markers for pericytes but has also been shown to be expressed by fibroblasts, while KCNJ8 is specifically enriched in pericytes in the brain (41). COL1 α 1 and LUM are both fibroblast markers and are not expressed in pericytes. The expression of COL1 α 1 and LUM was higher than both PDGFR β and KCNJ8 and more widespread than KCNJ8 in the assumed pericyte cluster. Altogether, these results suggest a so far unknown complexity of the GBM TME in which what is often referred to as pericytes may in truth be another many cell types.

Gene Signatures Suggest a Fibroblast Genotype in Pericyte Clusters

To get a better view of the, as defined by Abdelfattah et al., pericyte cluster, we integrated the dataset. This manipulation of the cells homogenised them and allowed cells from one patient to overlap more with the cells from another, thus removing some of the batch effects. However, this was a balancing act as, if integrated too much, the biological significance could be removed. This led the dataset to not fully be integrated and some batch effects could still be observed. The two integrations that were performed were done in R with reciprocal principal component analysis (RPCA) from the Seurat package and with Harmony from the Harmony package before clustering according to the Louvain algorithm took place (**Figures 2A and 2B**). Though similar, the Harmony integration was chosen due to ease of use.

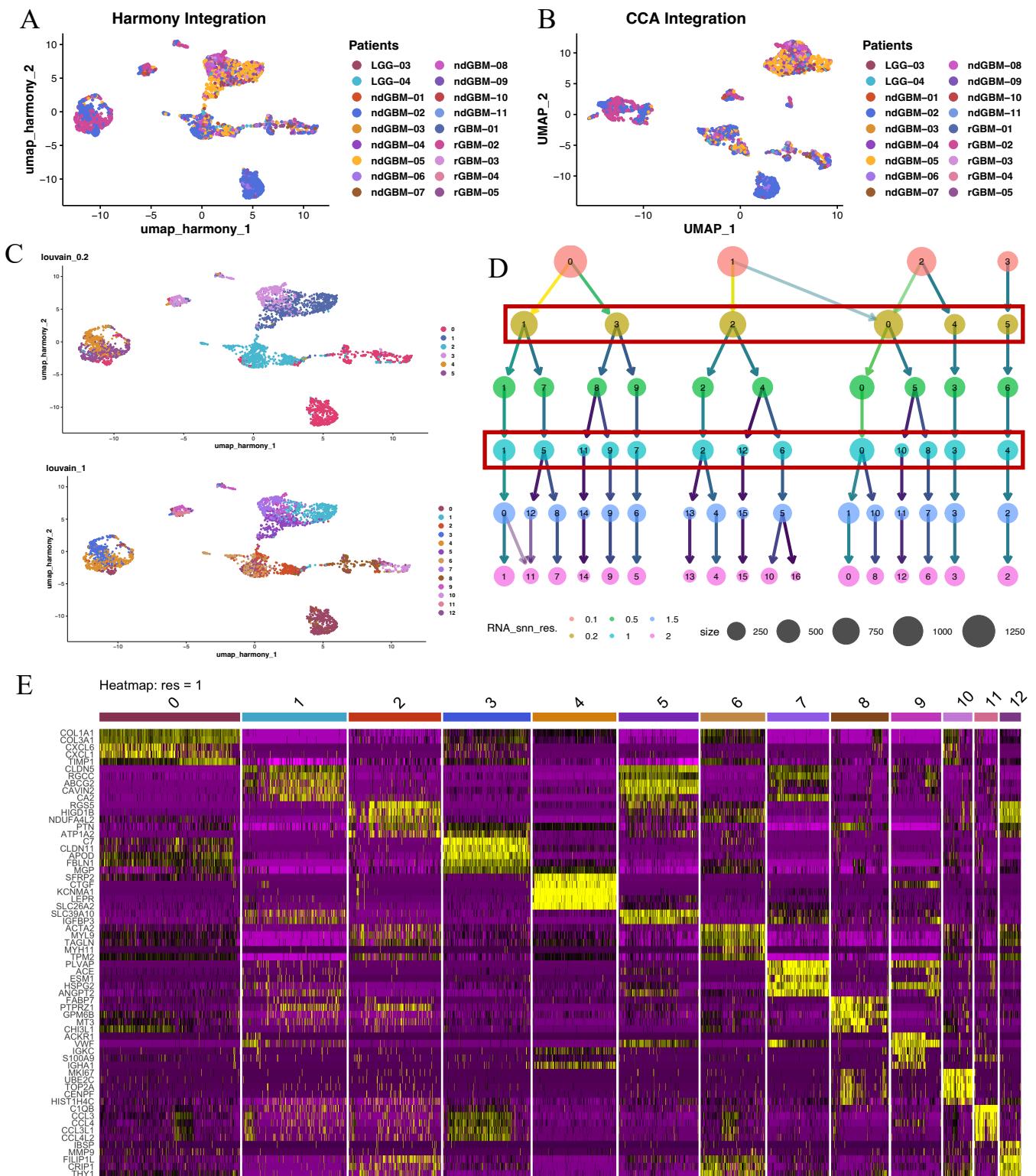


Figure 2: Re-clustering of dataset and gene enrichment analysis in new clusters.

(A-B) Integration of raw data, coloured based on patients, using (A) Harmony and (B) RPCA integration.

(C) Visualisation with Dimplot of chosen resolutions, 0.2 and 1.0, after Harmony integration and clustering, coloured based on clusters.

(D) Clustree visualisation of how cells move between clusters with increasing resolution. Red rectangles show the resolutions that were chosen for further investigations.

(E) Top differentially expressed genes in each cluster with resolution 1.0 visualised in a heatmap.

Abbreviations: RPCA: reciprocal principal component analysis

The resolutions were decided based on what seemed reasonable for the objective of this study, where one higher resolution, 1.0, and one lower, 0.2, was chosen to be used in later analyses (**Figure 2C** and **Supplemental Figure 2A and 2B**). Alternative resolutions, integrated both with RPCA and Harmony, were examined but ultimately decided against (**Supplemental Figure 3A and 3B**). In the Clustree, the movement of the cells between clusters is visualised and interestingly, the clusters 4 and 5 in the 0.2 resolution overlapped fully with the clusters 3 and 4 in the 1.0 resolution whereas the other clusters split at least once (**Figure 2D**).

We identified the top 5 differentially expressed genes, based on log₂-fold change, for each cluster in both the 1.0 and 0.2 resolutions (**Figures 2E** and **Supplemental Figure 4**). Cluster 0 in the 0.2 resolution had a clear division where some of the cells expressed CHI3L1, a protein known for its effect on angiogenesis and cell proliferation, expressed by peritumoral macrophages, cancer cells, including GBM cells, as well as by CAFs (42–44). Other genes expressed by Cluster 0 in the 0.2 resolution included *POSTN*, *COL1α1*, *COL3α1*, and *TIMP1*, all of which are known to be expressed by CAFs (45–47). In resolution 1.0, the Cluster 0 contained similar CAF genes as in the 0.2 resolution, however, *CXCL6* and *CXCL1* were also highly expressed which they are known to be in fibroblasts (48). This added up to the conclusion that Cluster 0 in resolution 1.0 was made up of mainly fibroblast-like cells. Cluster 1 had enriched gene expression in genes correlated to endothelial and tumour endothelial cells, e.g., *CLDN5* and *CA2* (49,50). These genes were also upregulated in Cluster 5 and part of Cluster 7. Additionally, in Cluster 5, *SLC39A10*, serving to promote resistance to apoptosis in cancer, and *IGFBP3*, a marker for endothelial cells, were enriched (51,52). Cluster 7 expressed genes such as *PLVAP*, a known endothelial-specific protein, *ACE*, which has long been known for its functions within the endothelium, and *ESMI* which contributes to angiogenesis and vascular permeability in endothelial cells (53–55). Part of Cluster 9 had high expression of the same genes as Cluster 7. These cells also had high expression of *VWF*, an endothelial marker, which, unsurprisingly, also was increased in Cluster 5 and Cluster 7 (56). Overall, these findings overlapped with the assignment done by Abdelfattah et al. in which Cluster 1, 5, 7, and part of Cluster 9 in the high resolution, 1.0, were endothelial cell clusters. Cluster 2 had high expression of *HIGD1B* and *NDUFA4L2*, both common and well-established pericyte genes, as well as *RGS5* which has been described to be expressed by pericytes within the tumour microenvironment (57,58). In other cancers, *CLDN11* and *FBLNI*, both highly expressed in Cluster 3, have been shown to be enriched in CAFs, suggesting that also this cluster contain fibroblast-like cells (59,60). Cluster 4 appeared to be yet another fibroblast cluster as it had an enrichment of *SFRP2* and *CTGF*, both of which are mainly expressed by fibroblasts (61,62). *ACTA2* and *MYL9* are both genes known to be expressed highly in smooth muscle cells (63,64). These genes and *MYH11*, another gene highly expressed in smooth muscle cells, were enriched in part of Cluster 6, which suggested that the cells in Cluster 6 were smooth muscle-like and were the ones that exhibited smooth muscle features mentioned previously (**Supplemental Figure 1**) (65). Cluster 8 was enriched for CHI3L1 which, as previously described, is a protein involved in proliferation. Additionally, *MT3*, also expressed highly in Cluster 8, has been shown to be involved with increased cell proliferation in cancer,

leading to the assumption that this cluster contained proliferating cells (66). Though not expressing *CHI3L1* to the same degree as Cluster 8, Cluster 10 contained other genes known to be heavily upregulated in cell proliferation, such as *MKI67*, *UBE2C*, *TOP2A*, and *CENPF*, indicating that this cluster also contained cells undergoing proliferation (67–70). Part of Cluster 9 was previously said to be endothelial cells, however, the cluster also had enriched genes related to immune responses and inflammation in cancer and other diseases, e.g., *S100A9* and *IGHA1*, suggesting that this cluster also had an involvement in these functions (71,72). *CCL3* and *CCL4* are both inflammatory cytokines suggesting that Cluster 11, in which *CCL3* and *CCL4* were enriched was, like Cluster 9, made up of cells that have roles related to immune responses and inflammation (73). Lastly, Cluster 12 contained enriched genes which overlapped with Cluster 2 and Cluster 6. Additionally, this Cluster had genes such as *THY1*, generally expressed by fibroblasts, *FILIP1L*, a marker for epithelial-to-mesenchymal transition, and *MMP9*, secreted by a variety of cells, including fibroblasts (74–76). Despite these gene enrichments indicating that Cluster 12 might also be a fibroblast cluster, it was difficult to make any substantial conclusions about Cluster 12 as it was the smallest of the clusters and had enriched genes with a multitude of functions.

Cluster 0 of the low, 0.2, resolution has already been assumed to be CAFs and proliferating cells. Overlapping the other cluster labels from the high to the low resolution showed that Cluster 1 and 3 should be endothelial cells. Cluster 2 was likely pericytes where the left-most cells had more of a SMC genotype, indicated by the pericyte genes, e.g., *HIGD1B* and *NDUFA4L2* that were still present in the cells which had high SMC associated genes (**Supplemental figure 4**). Cluster 4 and 5 in the low resolution overlapped well with Cluster 3 and 4 respectively in the high resolution indicating that they were both fibroblast clusters.

Annotation of Clusters Using ScType

With the aim of being less subjective in the naming of clusters, the dataset from Abdelfattah et al. in the low resolution was run through the ScType package in R.

A reference dataset from brain was put to predict the cell types and identified the endothelial cell clusters the same as previous manual prediction (**Figure 3A**). The remaining clusters, however, were classified differently and as no fibroblasts were part of the reference dataset, the clusters could not get this classification. The same results were found using a dataset from the pancreas where Cluster 4 and Cluster 5 were predicted to be ductal and beta cells respectively, i.e., pancreas specific cell types (**Figure 3B**).

Using a lung dataset as a reference revealed a prediction that Cluster 0 and Cluster 4 were fibroblasts, however, it had a great uncertainty regarding Cluster 2 which in manual prediction in the low resolution correlated to pericytes (**Figure 3C**). The liver dataset identified Cluster 0 and Cluster 4 as hepatic stellate cells which is a fibroblast-like cell type and shows these clusters' similitude to fibroblast-like cells, however also Cluster 2 was categorised as hepatic stellate cells pointing to the fact that pericytes is known to closely resemble fibroblasts (**Figure 3D**) (77).

An eye dataset seemingly contained all the cell types of interest and identified the clusters in a similar manner as the manual prediction (**Figure 3E**). Cluster 1 and Cluster 3 were

predicted to be endothelial cells while Cluster 0 and Cluster 4 were said to be fibroblasts. Cluster 2 was identified as pericytes and the only cluster that did not match up with previous predictions was Cluster 5 which was named as retinal pigment epithelial cells. As Cluster 5 was predicted as an epithelial cell type rather than that of a stromal cell with a reference dataset where the latter was present, an uncertainty of the identity of Cluster 5 became prevalent.

As the results from running through this pipeline were very different depending on the reference dataset, the main outcome of this experiment was a conclusion that depending on the data one inputs into the programme, the results will be very different. As the brain reference dataset had not annotated any cells as fibroblasts, using this specific dataset as a reference understandably does not give the fibroblast label to any of the clusters. Because of this, the further analyses and discussions will mainly take the manual classification of clusters into consideration.

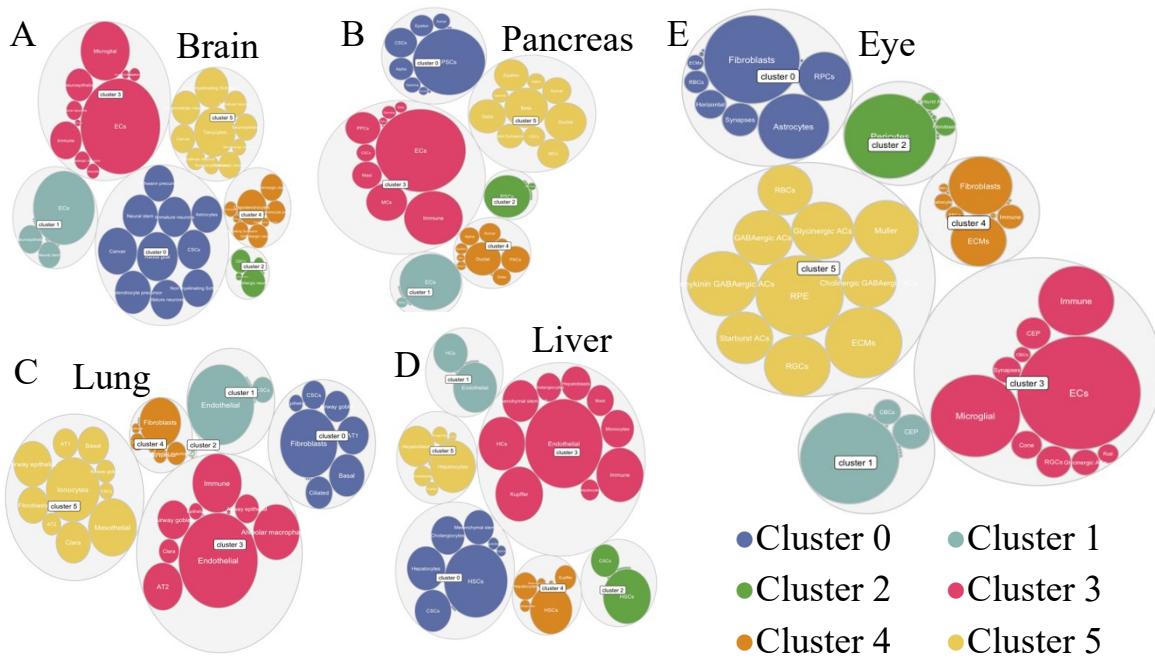


Figure 3: ScType analyses on dataset by Abdelfattah et al. in low resolution, 0.2.

(A-E) Using the reference datasets of (A) brain, (B) pancreas, (C) lung, (D) liver and (E) eye, the cell types of the clusters in their low resolutions were predicted.

CAF Signatures of Potential Fibroblast Subtypes Plotted on Abdelfattah's Dataset

Next, we wanted to look at established CAF signatures from other cancers to see if these also overlap with our suggested fibroblast signatures in the Abdelfattah et al. dataset. The signatures used were from Elyada et al., and Wu et al. (10,12,31).

From Elyada et al., signatures for myofibroblastic CAFs (myCAFs) and antigen presenting CAFs (apCAFs) were studied together with an inflammatory CAF (iCAF) signature which were removed from further investigation due to its low expression in the dataset (**Supplemental Figure 5**). All signatures from Wu et al., differentiated perivascular-like cells (dPVLs), iCAFs, immature perivascular-like cells (imPVLs) and myCAFs, were also included.

Functions of the subtypes of CAFs and PVLs remain elusive, however, studies suggest that they have a multitude of functions related to affecting the immune system and inflammation as well contributing to the ECM in the TME (**Figure 4A**) (10,12,31).

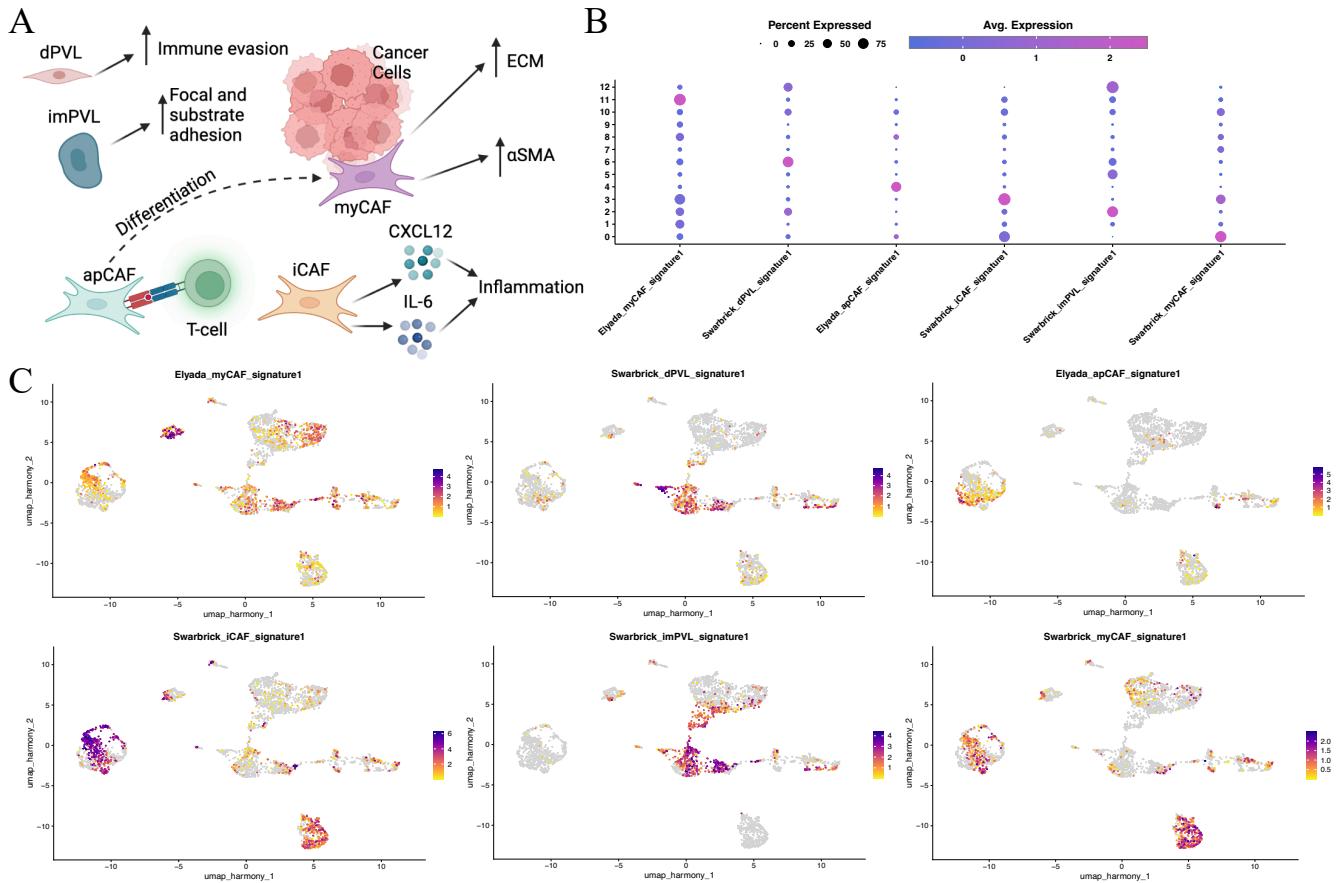


Figure 4: Plotting of established CAF signatures on dataset by Abdelfattah et al.

(A) Depiction of subtypes of CAFs and PVLs and their actions in the tumour TME. dPVL cells have properties contributing to immune evasion. imPVLs have functions related to focal and substrate adhesion. myCAFs are located in close proximity to cancer cells, have high expression of α SMA and produce ECM. iCAFs express cytokines, including CXCL12 and IL-6, which cascade to produce inflammation. apCAFs interact and activate T-cells through MHC class II and may, under certain conditions, differentiate into myCAFs. Created using BioRender.com.

(B) Dot plot of expression of CAF and PVL subtype signatures by Elyada et al. and Wu et al. (denoted as Swarbrick) per cluster from clustering with resolution 1.0.

(C) Signatures by Elyada et al. and Wu et al. (denoted as Swarbrick) of CAF and PVL subtypes plotted in a feature plot on Harmony integrated dataset.

Abbreviations: CAF: Cancer Associated Fibroblast, PVL: perivascular-like cell, TME: tumour microenvironment, dPVL: differentiated PVL, imPVL: immature PVL, myCAF: myofibroblastic CAF, α SMA: alpha-smooth muscle actin, ECM: extracellular matrix, iCAF: inflammatory CAF, CXCL12: stromal cell-derived factor 1, IL-6: interleukin 6, apCAF: antigen-presenting CAF, MHC: major histocompatibility complex

We showed that the CAF and PVL signatures were present in the Abdelfattah et al. dataset and primarily expressed by different clusters (**Figure 4B and 4C**). Interestingly, the myCAF signatures from Elyada et al. and Wu et al. did not match up where the former had an enriched expression in Cluster 11 of the high resolution and the latter in Cluster 0. The dPVL signature had an increased expression in Cluster 6, the cluster previously identified to be SMC-like. An explanation to this was found in looking at the gene list for dPVLs in which common smooth muscle cell genes like *ACTA2* and *MYH11* were found, suggesting that Cluster 6

contained cells which not necessarily were SMCs but might be perivascular-like cells, explaining the labelling in the low resolution of these as pericytes. Cluster 4 had an enrichment for the apCAF signature. The iCAF signature was enriched in Cluster 3. Many cells expressing an iCAF signature were also located in Cluster 0, however, their expression tended to be less than among the cells in Cluster 3. This was reversely true for the myCAF signature where Cluster 0 had the higher expression of the two according to the Wu et al. signatures. The imPVL signature was highest in Cluster 2, however, an enrichment was also seen in especially Cluster 5 and Cluster 12. We previously supposed that Cluster 2 was a pericyte cluster, Cluster 5 was an endothelial cluster and Cluster 12 was inconclusive. These new results created an ambiguity of what types of cells these clusters contained and what imPVLs actually are and what their function in the tumour TME is.

In the lower resolution, Cluster 0 had an increased expression of the Wu et al. myCAF and iCAF signatures. The same was true for Cluster 4 though, in this cluster, the iCAF signature was higher than the myCAF one. As they are endothelial clusters, no major CAF signature was found in Cluster 1 and 3, however, the imPVL signature reached partly into Cluster 1, indicating a division between the cells in this cluster. Cluster 2, which was classified as pericytes in this resolution showed the highest correlation with the dPVL and imPVL signatures. Lastly, Cluster 5 showed an increased expression of the apCAF signature.

RNAscope Staining Showing Expression of Fibroblast Genes in mouse model of GBM

Targets for the RNAscope multiplex staining were identified bioinformatically in previous experiments and further validated in the integrated dataset (**Figure 5A**). 3 probes and 2 antibodies were used. *Kcnj8* has been shown to be a specific marker for pericytes in a tumour environment, and *PDGFR β* is known to be expressed by mesenchymal cells, including both pericytes and fibroblasts, however, it is often used as a pericyte marker in stainings (78). Both these genes were seen in the datasets of Abdelfattah et al. and Pietras et al. (**Figure 1F and 6C**). *Coll α 1* and *Lum* are both suggested specific markers for fibroblasts, however, it remains difficult to definitively say so, and such they were chosen due to them looking promising in the Abdelfattah et al. and Pietras et al. datasets (**Figure 1G and 6C**). Additionally, CD31, encoded by Pecam-1 and highly expressed in endothelial cells, was used in order to visualise the blood vessels (79). *PDGFR β* and CD31 were the antibodies used and coloured as yellow and turquoise respectively, while *Coll α 1*, *Lum* and *Kcnj8* were probes with the colours red, green and orange (**Figure 5B**).

This RNAscope staining in a GBM mouse model showed cells co-expressing *Coll α 1* and *Lum*, indicating that fibroblast-like cells are part of the GBM TME (**Figure 5C**). Furthermore, many of the *Coll α 1*-positive cells, and some of the *Lum*-positive cells, expressed *PDGFR β* and/or *Kcnj8*. Interestingly, a few cells which expressed *PDGFR β* , *Coll α 1*, *Lum* as well as *Kcnj8* were observed which may indicate a stronger relationship between pericyte-like and fibroblast-like cells than previously thought.

Most of the pericyte-like cells, PDGFR β and *Kcnj8* positive, were seen close to the endothelial cells marked by CD31, however, though many of the fibroblast-like cells were also close to the endothelium, a number of them were located further away from the blood vessels (**Figure 5D and 5E**). This migration from endothelial cells was especially seen in, however not exclusive to, the *Col1 α 1*-positive cells which were negative for all the other examined markers.

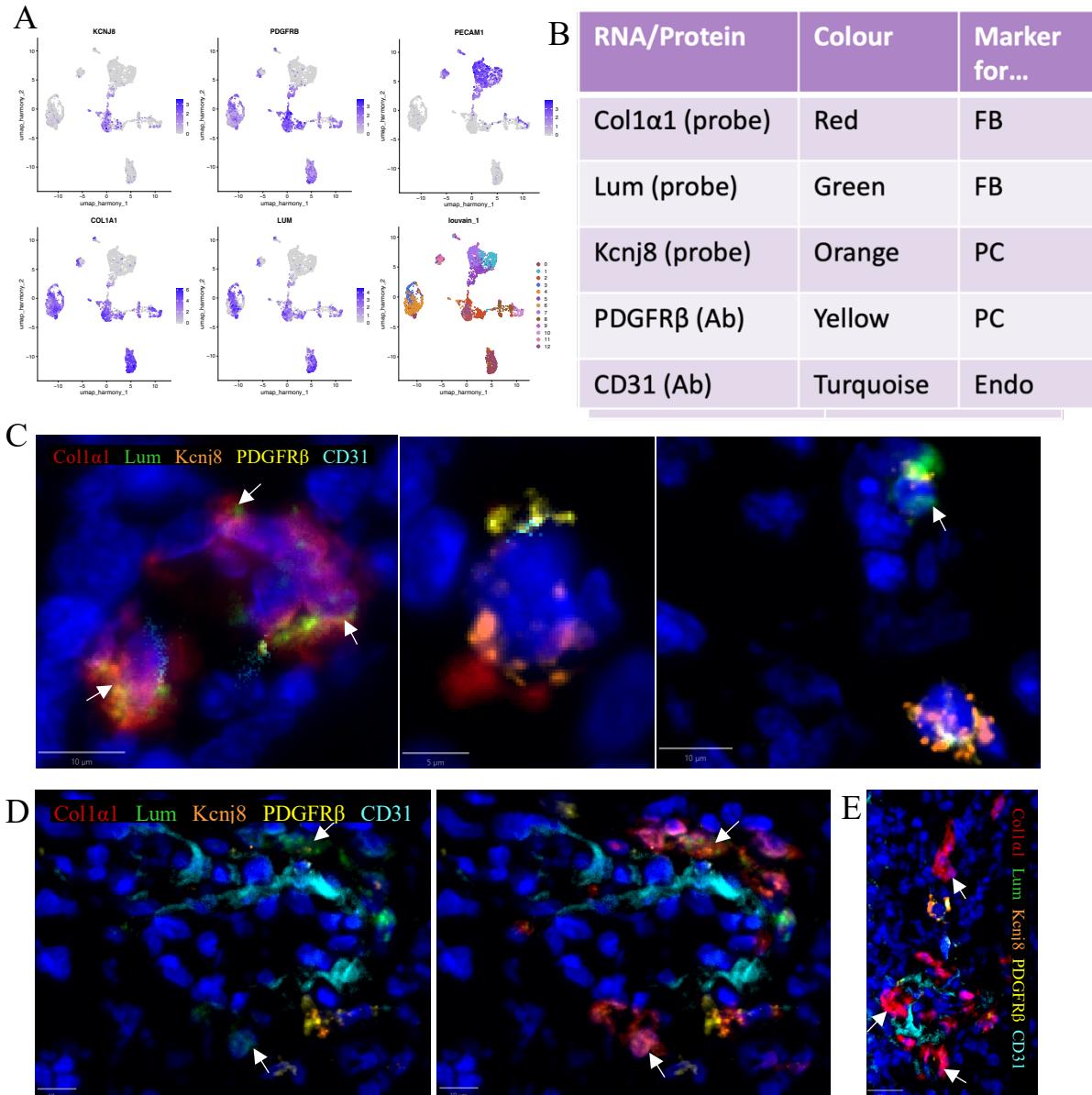


Figure 5: Multiplex staining of fibroblast and pericyte markers on GBM mouse model

- (A) Feature plots on Harmony integrated dataset by Abdelfattah et al. of target genes and surface markers for RNAscope as well as Dimplot with resolution 1.0.
- (B) Table outlining the targets used in the multiplex staining, their corresponding colours as seen in the image analysis as well as if they are a marker for fibroblasts (FB), pericytes (PC) or endothelial cells (Endo).
- (C) Multiplex staining of fibroblast markers *Col1 α 1* (red) and *Lum* (green) as well as co-localisation with pericyte markers PDGFR β (yellow) and *Kcnj8* (orange).
- (D) Multiplex staining of cells with double-positive fibroblast marker expression localised adjacent to endothelial cells marked by CD31 (turquoise).
- (E) Multiplex staining of *Col1 α 1*-positive (red) cells localised adjacent to endothelium as well as further away.

Validation of Findings in Additional Datasets

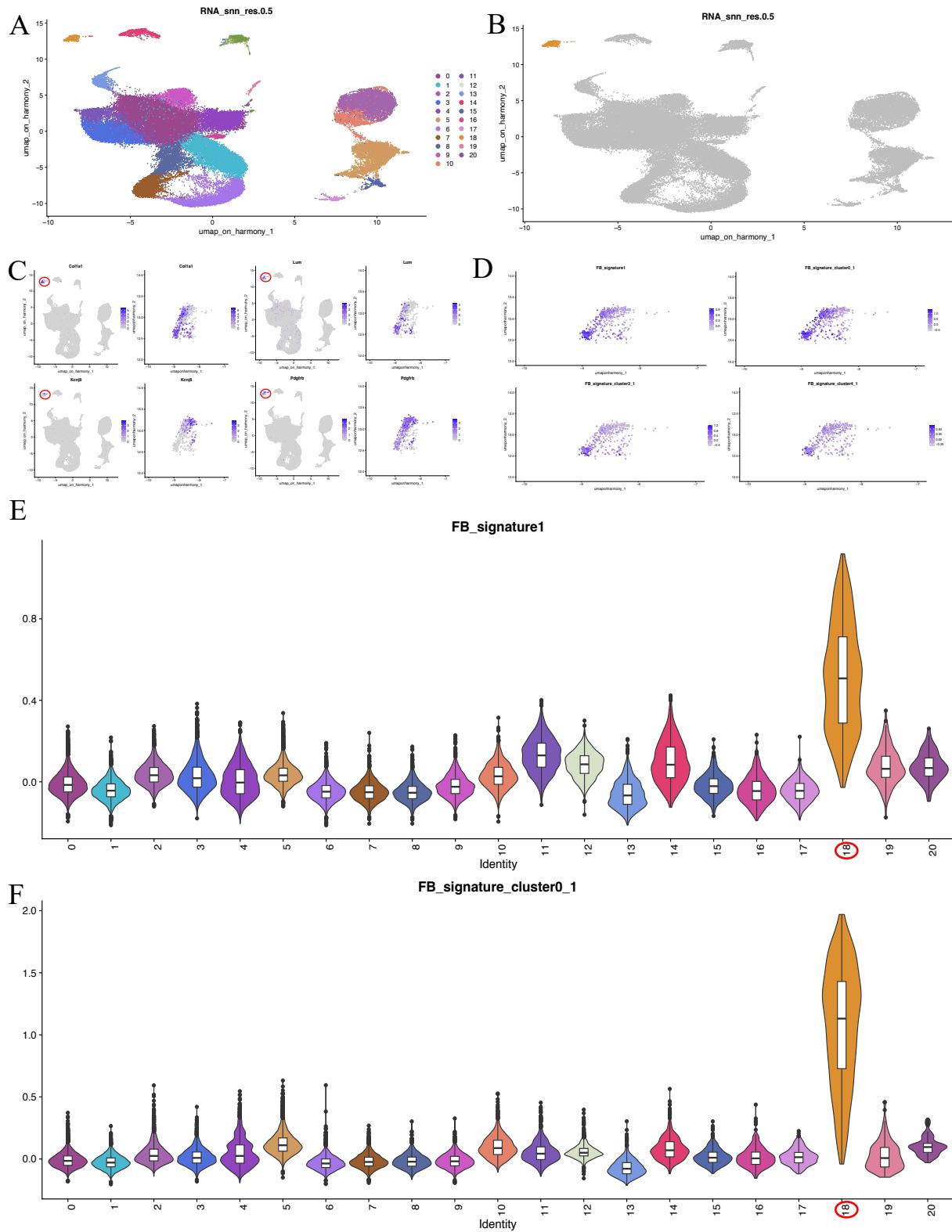


Figure 6: Plotting of fibroblast genes and signatures on the dataset from Pietras et al. focusing on pericyte cluster.

(A) Harmony integrated dataset from Pietras et al. visualised in a Dimplot with resolution 0.5.

(B) Harmony integrated dataset from Pietras et al. with pericyte assigned cluster, Cluster 18, highlighted in orange.

Legend continued on next side...

(C) Fibroblast genes, Col1 α 1 and Lum, and pericyte genes, Kcnj8 and Pdgfr β mapped as Feature plot on full dataset as well as in Cluster 18 alone. Red rings in full dataset encompass Cluster 18.

(D) Feature plots with Cluster 18 subsetted from dataset by Pietras et al. showing complete fibroblast signature established from dataset by Abdelfattah et al., as well as specific signatures from each suggested fibroblast cluster.

(E-F) Violin plots showing expression of (E) overall fibroblast signature and (F) signature from Cluster 0 identified from dataset from Abdelfattah et al. on each cluster in Pietras et al.'s dataset. Red rings highlight Cluster 18, assigned pericyte cluster from Pietras et al. The y-axis denotes the signature score from the signature.

In order to validate the findings that fibroblasts may be present in the GBM TME, a dataset from Pietras et al. which had been created for prior experiments was used. The UMAP from the dataset, integrated with Harmony, was generated and the cluster assigned as pericytes (Cluster 18) was taken out for further analysis (**Figure 6A and 6B**). The previously identified genes for identification of fibroblasts and pericytes were applied to the complete dataset as well as to Cluster 18 alone and revealed that seemingly there was an enrichment in the fibroblast-associated genes, especially in the lower left of the cluster, whereas the pericyte gene expression was more located towards the top right (**Figure 6C**). Additionally, the complete fibroblast signature, along with the specific signatures of the high resolution, 1.0, from Cluster 0, 3 and 4 in the Abdelfattah dataset and the pericyte-like Cluster 2, were clustered on the Cluster 18 from Pietras et al., highlighting the same as what was found in the previous gene expression analysis (**Figure 6D and Supplemental Figure 6**). Cluster 3 had the strongest specific signature as can be seen by the expression scale legend as well with most of the cells with a high expression of the signature clustering to the lower left of the cluster. The signature from Cluster 0 was overall strong while Cluster 4's signature was rather weak altogether. The total fibroblast signature confirmed that the fibroblast signature was higher in the cluster of interest, i.e., Cluster 18 in the dataset by Pietras et al., compared to the other clusters (**Figure 6E**). Additionally, the signature from Cluster 0, and to a lesser extent Cluster 3, had a clear difference in expression between Cluster 18 and the remaining clusters whereas the Cluster 4 signature was weaker and more evenly distributed between the clusters (**Figure 6F and Supplementary Figure 7A-B**). With these results we saw further support for the outcomes from previous experiments in that we have what looks like fibroblast-like cells in GBM.

Population Survival and Biological Analysis Based on New Cluster Analysis

From the analyses on the dataset from Abdelfattah et al., 3 main clusters were suggested to contain fibroblast-like cells, cluster 0, 4 and 5 in the low 0.2 resolution. To put this into a clinical perspective, an analysis on the population survival was conducted. Patients with gliomas, the majority having GBM or recurrent GBM, were examined based on their expression levels of the various signatures described previously.

Interestingly, Cluster 0, the cluster with the highest myofibroblastic signature, in its low resolution was the only cluster with a statistically significant difference in survival based on OS and DSS where patients with high expression of genes associated with Cluster 0 had a higher mortality rate than those with low expression (**Figure 7A and Supplementary Figure 8A-E**). The difference in survival was deemed statistically significant for p-values < 0.05. To get a first preliminary indication on the differences between the clusters, we performed a biological enrichment analysis (**Figure 7B**). This showed that Cluster 0, and no other cluster, had enriched gene-sets associated with hypoxia and glycolysis.

In order to summarise the results of this project, the newly identified fibroblast-like cells within the GBM microenvironment were plotted as part of the full dataset by Abdelfattah et al. (**Figure 7C**). Here Cluster 0 and Cluster 4 were named CAF_1 and CAF_2 respectively. Cluster 5 was named unknown because although the gene expression indicating it as a fibroblast cluster, the ScType analysis revealed that it had a relatively strong epithelial phenotype rather than a stromal one.

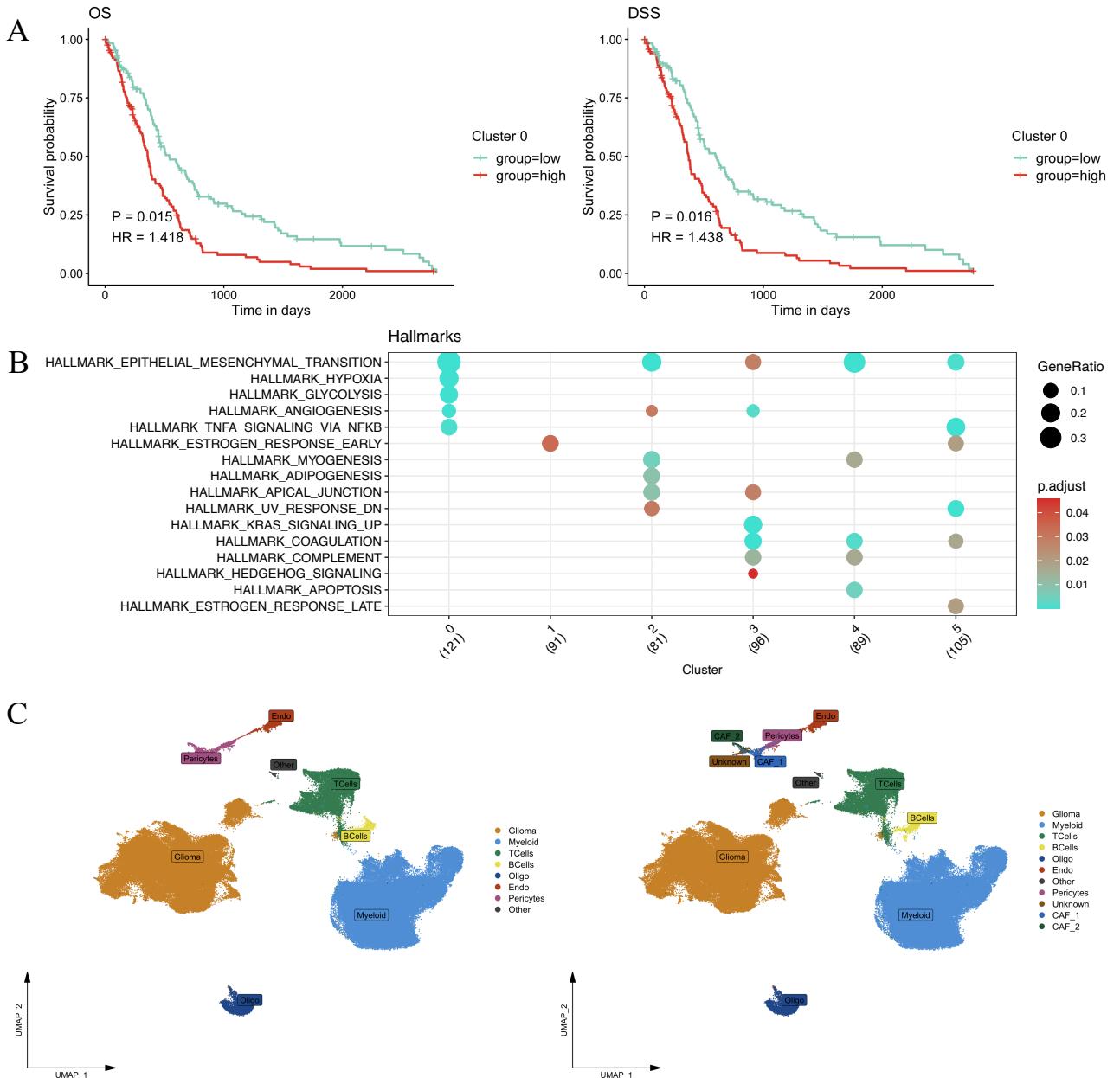


Figure 7: Survival and biological analysis of fibroblast signatures.

(A) Kaplan-Meier survival curves of OS and DSS on GBM population based on signature from low, 0.2, resolution Cluster 0. The difference in survival between a high and low expression of the signature was deemed statistically significant for p-values < 0.05.

(B) Biological functions of each cluster visualised in a DotPlot with larger dots indicating a greater percentage of cells involved in the function and turquoise dots indicating a higher certainty of the results conveyed through an adjusted p-value.

(C) Dimplot of cell types in the full dataset by Abdelfattah et al. where the right figure contains the updated inclusion of the cell types identified in this project.

Abbreviations: OS: overall survival, DSS: disease-specific survival, GBM: glioblastoma

Discussion

It has been suggested that fibroblasts and fibroblast-like cells do not exist within the brain parenchyma, however, this project aims to challenge this belief (13). A hypothesis discussed in an article by Schaffenrath et al., is that upon injury in the brain, fibroblasts migrate to the damaged area (7). This would explain why they have not been observed in the healthy brain, however, how they would pass the blood brain barrier remains unspecified.

Our study provides a bioinformatic analysis showing a strong fibroblast signature in certain cell populations in the GBM brain. A staining performed additionally revealed a great number of *Collα1*-positive and *Lum*-positive cells which clearly indicate that there are fibroblast-like cells in the brain in GBM. Whether this applies also to the healthy brain, however, remains unknown.

In the analysis with fibroblast signatures, it was clear that the cells with the strongest fibroblast phenotype were the ones which had previously been labelled as pericytes. We could show a relationship between these cell types in the staining where the co-expression of PDGFR β and especially *Collα1*, but also *Lum*, was evident. This could suggest an intermediate cell state between pericytes and fibroblasts where the expression of certain genes such as PDGFR β vary in expression depending on the cell state, e.g., having an increased expression in pericyte-like cells. In an article by Hosaka et al., pericyte-fibroblast transition is a suggested mechanism in tumour metastasis and invasion in solid tumours (80). If this transition also take place in GBM it could perhaps explain how some cells in the staining had expression of both pericyte and fibroblast genes. Alternatively, these results may simply indicate that PDGFR β might not be a specific marker for pericytes but also marks fibroblast-like cells. What remains unknown in this case would be why *Kcnj8*, a distinct pericyte marker, is at times co-expressed with the fibroblast markers.

In this work, we also suggest that a high expression of a special type of fibroblast-like cell, likely with a myofibroblastic CAF phenotype, is an indicator for a poor prognosis among patients with GBM. Thus, if further validated, these cells could potentially be an interesting target for drug development. The fact that Cluster 0 is involved in hypoxia and glycolysis could be a possible explanation as to why it is associated with a worse prognosis. A hypoxic microenvironment is known to cause hypervascularisation and necrosis, both characteristics of GBM (81). Tying in with hypoxia, glycolysis, which was also elevated in Cluster 0, is a natural adaptation to hypoxia and is mediated by HIF-1 α which has been the mainly targeted factor in GBM when attempting to inhibit the hypoxia pathway for treatment (81,82). As these fibroblast-like cells of Cluster 0 might be involved in both hypoxia and glycolysis they could provide an interesting alternative target when trying to dampen the detrimental effects of hypoxia in GBM.

Though the incidence of GBM is relatively low, the incidence is on the rise while survival frequency remains low, highlighting a need for further research and interventions to be developed (83). Additionally, breast cancer, lung cancer and melanoma are among the cancers that may metastasise to the brain leaving patients with an incredibly low chance of survival (84). If a better understanding of the tumour microenvironment in the brain can be

established, these cancer patients, as well as those suffering from other cancers originating in the brain, might benefit along with GBM patients.

The next steps following this study are to characterise the cells of especially Cluster 0 further biologically. It would be interesting to examine the communication pathways between Cluster 0 and other cells in the GBM TME and the tumour itself. Additionally, a knockout of these cells could further reveal their role in GBM progression and development. Lastly, a scRNA-seq experiment focusing on these cells along with the other examined fibroblast-like cells and pericytes in this project would be a great tool for further analysis as it could provide a larger dataset which would make the results more reliable.

All in all, GBM is an extraordinarily complex disease and we have yet to understand all of its intricacies. This project has aimed to provide an additional puzzle piece which might help to better understand the GBM TME and could provide an interesting perspective in the development of novel treatments.

Limitations of the Study

As the study by Abdelfattah et al., was not focused on pericytes/fibroblasts, and these cell populations being generally small in comparison to other cell types in the GBM TME, there were quite few cells that could be used for analysis. This decreases the likelihood that all findings in this study are universally applicable. The clusters that were identified may not represent the full picture of what fibroblasts look like in the GBM TME and there may be gene signatures that were overlooked due to too few cells in the subset expressing them. In order to get more a reliable hypothesis of the fibroblast population in the brain, a scRNA-seq experiment could be done, focusing on these cell populations. This was not possible to do during this project due to the high cost in running a scRNA-seq experiment as well as the long timeline it requires.

Additionally, with the small number of cells that were included in the datasets analysed in this study, it proved difficult to make an integration where the batch effects of each patient's sample were minimised. An integration completely removing the batch effects was tried, but as this almost completely homogenised the cells, the biological aspects were not apparent and further experiments could not be conducted. Due to this, there may be reason to think that some of the results of this thesis work may be related to the differences in acquisition of the samples for prior scRNA-seq experiments where some surgeries may e.g., have taken out more of the surrounding brain tissue or different structures depending on the spatial location of the tumour.

There were no cancer-free mice in the wet-lab experiments for this project due to a lack of time. As it has been suggested that fibroblasts might migrate into the brain parenchyma in cancer, looking for fibroblasts in healthy mice would provide an interesting contrast to the results in the current study (7).

Acknowledgements

I would first and foremost like to thank my supervisor Dr Jonas Sjölund who has guided me through this project and who contributed to me developing a deep fascination for the field of bioinformatics.

I also want to thank Dr Kristian Pietras for allowing me a place in his lab group and for providing a calm and encouraging working atmosphere in the lab. Furthermore, the lab group as a whole has been very welcoming and has allowed me to feel at home in the lab.

My fellow student Michelle Lü has provided invaluable help when it came to the wet-lab portion of this project. Additionally, she has been a great office companion, listening to proposed ideas and coming with her own insights. I lastly want to thank fellow students Alexander, Florian, Teia, Ruben, Anna, Gustav, and Naftali for amusing debates over lunch and for very entertaining after-works.

References

1. Ohgaki H, Kleihues P. Population-Based Studies on Incidence, Survival Rates, and Genetic Alterations in Astrocytic and Oligodendroglial Gliomas. *J Neuropathol Exp Neurol.* 2005 Jun 1;64(6):479–89.
2. Tykocki T, Eltayeb M. Ten-year survival in glioblastoma. A systematic review. *J Clin Neurosci.* 2018 Aug 1;54:7–13.
3. Organisation mondiale de la santé, Centre international de recherche sur le cancer, editors. Central nervous system tumours. 5th ed. Lyon: International agency for research on cancer; 2021. (World health organization classification of tumours).
4. Omuro A. Glioblastoma and Other Malignant Gliomas: A Clinical Review. *JAMA.* 2013 Nov 6;310(17):1842.
5. Claes A, Idema AJ, Wesseling P. Diffuse glioma growth: a guerilla war. *Acta Neuropathol (Berl).* 2007 Nov 1;114(5):443–58.
6. Bikfalvi A, da Costa CA, Avril T, Barnier JV, Bauchet L, Brisson L, et al. Challenges in glioblastoma research: focus on the tumor microenvironment. *Trends Cancer.* 2023 Jan 1;9(1):9–27.
7. Schaffernath J, Wyss T, He L, Rushing EJ, Delorenzi M, Vasella F, et al. Blood-brain barrier alterations in human brain tumors revealed by genome-wide transcriptomic profiling. *Neuro-Oncol.* 2021 Dec 1;23(12):2095–106.
8. Lynch MD, Watt FM. Fibroblast heterogeneity: implications for human disease. *J Clin Invest.* 128(1):26–35.
9. Bartoschek M, Oskolkov N, Bocci M, Lövrot J, Larsson C, Sommarin M, et al. Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. *Nat Commun.* 2018 Dec 4;9(1):5150.
10. Wu SZ, Roden DL, Wang C, Holliday H, Harvey K, Cazet AS, et al. Stromal cell diversity associated with immune evasion in human triple-negative breast cancer. *EMBO J.* 2020 Oct 1;39(19):e104063.
11. Khalil AM, Erdogan C, Kurt Z, Turgut SS, Grunwald MW, Rand T, et al. Refining colorectal cancer classification and clinical stratification through a single-cell atlas. *Genome Biol.* 2022 May 11;23:113.
12. Öhlund D, Handly-Santana A, Biffi G, Elyada E, Almeida AS, Ponz-Sarvise M, et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J Exp Med.* 2017 Mar 6;214(3):579–96.

13. Vanlandewijck M, He L, M  e MA, Andrae J, Ando K, Del Gaudio F, et al. A molecular atlas of cell types and zonation in the brain vasculature. *Nature*. 2018 Feb 22;554(7693):475–80.
14. DeSisto J, O’Rourke R, Jones HE, Pawlikowski B, Malek AD, Bonney S, et al. Single-Cell Transcriptomic Analyses of the Developing Meninges Reveal Meningeal Fibroblast Diversity and Function. *Dev Cell*. 2020 Jul;54(1):43-59.e4.
15. Derk J, Jones HE, Como C, Pawlikowski B, Siegenthaler JA. Living on the Edge of the CNS: Meninges Cell Diversity in Health and Disease. *Front Cell Neurosci*. 2021 Jul 1;15:703944.
16. Dorrier CE, Jones HE, Pintari  L, Siegenthaler JA, Daneman R. Emerging roles for CNS fibroblasts in health, injury and disease. *Nat Rev Neurosci*. 2022 Jan;23(1):23–34.
17. Zhang XN, Yang KD, Chen C, He ZC, Wang QH, Feng H, et al. Pericytes augment glioblastoma cell resistance to temozolomide through CCL5-CCR5 paracrine signaling. *Cell Res*. 2021 Oct;31(10):1072–87.
18. Yao Y. Challenges in Pericyte Research: Pericyte-Specific and Subtype-Specific Markers. *Transl Stroke Res*. 2022 Dec 1;13(6):863–5.
19. Jovic D, Liang X, Zeng H, Lin L, Xu F, Luo Y. Single-cell RNA sequencing technologies and applications: A brief overview. *Clin Transl Med*. 2022 Mar 29;12(3):e694.
20. Xu L, Yao Y. Central Nervous System Fibroblast-Like Cells in Stroke and Other Neurological Disorders. *Stroke*. 2021 Jul;52(7):2456–64.
21. Jain S, Rick JW, Joshi RS, Beniwal A, Spatz J, Gill S, et al. Single-cell RNA sequencing and spatial transcriptomics reveal cancer-associated fibroblasts in glioblastoma with protumoral effects. *J Clin Invest*. 2023 Mar 9;133(5).
22. Abdelfattah N, Kumar P, Wang C, Leu JS, Flynn WF, Gao R, et al. Single-cell analysis of human glioma and immune cells identifies S100A4 as an immunotherapy target. *Nat Commun*. 2022 Feb 9;13(1):767.
23. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst*. 2019 Apr 24;8(4):329-337.e4.
24. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021 Jun;184(13):3573-3587.e29.

25. Korsunsky I, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, et al. Fast, sensitive, and accurate integration of single cell data with Harmony. *Bioinformatics*; 2018 Nov.
26. Blanco-Carmona E. Generating publication ready visualizations for Single Cell transcriptomics using SCpubr. *Bioinformatics*; 2022 Mar.
27. iWantHue [Internet]. médialab Sciences Po; 2023 [cited 2023 Mar 21]. Available from: <https://github.com/medialab/iwanthue>
28. Zappia L, Oshlack A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *GigaScience*. 2018 Jul 1;7(7).
29. Ianevski A, Giri AK, Aittokallio T. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nat Commun*. 2022 Mar 10;13(1):1246.
30. Winkler EA, Kim CN, Ross JM, Garcia JH, Gil E, Oh I, et al. Single cell atlas of the normal and malformed human brain vasculature. *Science*. 2022 Mar 4;375(6584):eabi7377.
31. Elyada E, Bolisetty M, Laise P, Flynn WF, Courtois ET, Burkhardt RA, et al. Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. *Cancer Discov*. 2019 Aug;9(8):1102–23.
32. Marsh S, Salmon M, Hoffman P. samuel-marsh/scCustomize: Version 1.1.1 [Internet]. Zenodo; 2023 [cited 2023 Mar 22]. Available from: <https://zenodo.org/record/5706430>
33. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012 May;2(5):401–4.
34. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013 Apr 2;6(269):pl1.
35. Kassambara A. survminer [Internet]. 2021. Available from: <https://cloud.r-project.org/web/packages/survminer/index.html>
36. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov Camb Mass*. 2021 Aug 28;2(3):100141.
37. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J Integr Biol*. 2012 May;16(5):284–7.

38. Dolgalev I. msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format [Internet]. 2022. Available from: <https://CRAN.R-project.org/package=msigdbr>
39. ACD. RNAscope® Multiplex Fluorescent v2 Assay combined with Immunofluorescence - Integrated Co-Detection Workflow (ICW). 2021.
40. ACD. RNAscope® 4-plex Ancillary Kit for Multiplex Flurescent Reagent Kit v2. 2019.
41. Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, Van Der Zwan J, et al. Molecular Architecture of the Mouse Nervous System. *Cell*. 2018 Aug;174(4):999-1014.e22.
42. Junker N, Johansen JS, Andersen CB, Kristjansen PEG. Expression of YKL-40 by peritumoral macrophages in human small cell lung cancer. *Lung Cancer Amst Neth*. 2005 May;48(2):223–31.
43. Junker N, Johansen JS, Hansen LT, Lund EL, Kristjansen PEG. Regulation of YKL-40 expression during genotoxic or microenvironmental stress in human glioblastoma cells. *Cancer Sci*. 2005 Mar;96(3):183–90.
44. Watanabe K, Shiga K, Maeda A, Harata S, Yanagita T, Suzuki T, et al. Chitinase 3-like 1 secreted from cancer-associated fibroblasts promotes tumor angiogenesis via interleukin-8 secretion in colorectal cancer. *Int J Oncol*. 2022 Jan;60(1):3.
45. Yue H, Li W, Chen R, Wang J, Lu X, Li J. Stromal POSTN induced by TGF- β 1 facilitates the migration and invasion of ovarian cancer. *Gynecol Oncol*. 2021 Feb 1;160(2):530–8.
46. Ucaryilmaz Metin C, Ozcan G. Comprehensive bioinformatic analysis reveals a cancer-associated fibroblast gene signature as a poor prognostic factor and potential therapeutic target in gastric cancer. *BMC Cancer*. 2022 Jun 23;22(1):692.
47. Pape J, Magdeldin T, Stamati K, Nyga A, Loizidou M, Emberton M, et al. Cancer-associated fibroblasts mediate cancer progression and remodel the tumouroid stroma. *Br J Cancer*. 2020 Sep 29;123(7):1178–90.
48. Chen Y, Wang H, Yang Q, Zhao W, Chen Y, Ni Q, et al. Single-cell RNA landscape of the osteoimmunology microenvironment in periodontitis. *Theranostics*. 2022;12(3):1074–96.
49. Annan DA, Maishi N, Soga T, Dawood R, Li C, Kikuchi H, et al. Carbonic anhydrase 2 (CAII) supports tumor blood endothelial cell survival under lactic acidosis in the tumor microenvironment. *Cell Commun Signal*. 2019 Dec 17;17(1):169.

50. Hashimoto Y, Greene C, Munnich A, Campbell M. The CLDN5 gene at the blood-brain barrier in health and disease. *Fluids Barriers CNS*. 2023 Mar 28;20(1):22.
51. Ma Z, Li Z, Wang S, Zhou Q, Ma Z, Liu C, et al. SLC39A10 Upregulation Predicts Poor Prognosis, Promotes Proliferation and Migration, and Correlates with Immune Infiltration in Hepatocellular Carcinoma. *J Hepatocell Carcinoma*. 2021 Aug 7;8:899–912.
52. Rhee S, Paik DT, Yang JY, Nagelberg D, Williams I, Tian L, et al. Endocardial/endothelial angiocrines regulate cardiomyocyte development and maturation and induce features of ventricular non-compaction. *Eur Heart J*. 2021 Jul 19;42(41):4264–76.
53. Denzer L, Muranyi W, Schroten H, Schwerk C. The role of PLVAP in endothelial cells. *Cell Tissue Res*. 2023 Feb 13;
54. Miyazaki M, Okunishi H, Nishimura K, Toda N. Vascular Angiotensin-Converting Enzyme Activity in Man and other Species. *Clin Sci*. 1984 Jan 1;66(1):39–45.
55. Rocha SF, Schiller M, Jing D, Li H, Butz S, Vestweber D, et al. Esm1 Modulates Endothelial Tip Cell Behavior and Vascular Permeability by Enhancing VEGF Bioavailability. *Circ Res*. 2014 Aug 29;115(6):581–90.
56. Pusztaszeri MP, Seelentag W, Bosman FT. Immunohistochemical Expression of Endothelial Markers CD31, CD34, von Willebrand Factor, and Fli-1 in Normal Human Tissues. *J Histochem Cytochem*. 2006 Apr;54(4):385–95.
57. Baek SH, Maiorino E, Kim H, Glass K, Raby BA, Yuan K. Single Cell Transcriptomic Analysis Reveals Organ Specific Pericyte Markers and Identities. *Front Cardiovasc Med*. 2022 Jun 1;9:876591.
58. Shen J, Shrestha S, Yen YH, Scott MA, Soo C, Ting K, et al. The pericyte antigen RGS5 in perivascular soft tissue tumors,. *Hum Pathol*. 2016 Jan;47(1):121–31.
59. Karagiannis GS, Schaeffer DF, Cho CKJ, Musrap N, Saraon P, Batruch I, et al. Collective migration of cancer-associated fibroblasts is enhanced by overexpression of tight junction-associated proteins claudin-11 and occludin. *Mol Oncol*. 2014;8(2):178–95.
60. Orr B, Riddick ACP, Stewart GD, Anderson RA, Franco OE, Hayward SW, et al. Identification of stromally expressed molecules in the prostate by tag-profiling of cancer-associated fibroblasts, normal fibroblasts and fetal prostate. *Oncogene*. 2012 Mar 1;31(9):1130–42.
61. Tabib T, Morse C, Wang T, Chen W, Lafyatis R. SFRP2/DPP4 and FMO1/LSP1 define major fibroblast populations in human skin. *J Invest Dermatol*. 2018 Apr;138(4):802–10.

62. Igarashi A, Okochi H, Bradham DM, Grotendorst GR. Regulation of connective tissue growth factor gene expression in human skin fibroblasts and during wound repair. *Mol Biol Cell*. 1993 Jun;4(6):637–45.
63. Rockey DC, Weymouth N, Shi Z. Smooth Muscle α Actin (Acta2) and Myofibroblast Function during Hepatic Wound Healing. *PLoS ONE*. 2013 Oct 29;8(10):e77166.
64. Kruthika BS, Sugur H, Nandaki K, Arimappamagan A, Paturu K, Santosh V. Expression pattern and prognostic significance of myosin light chain 9 (MYL9): a novel biomarker in glioblastoma. *J Clin Pathol*. 2019 Oct 1;72(10):677–81.
65. Gilbert MA, Schultz-Rogers L, Rajagopalan R, Grochowski CM, Wilkins BJ, Biswas S, et al. Protein-elongating mutations in MYH11 are implicated in a dominantly inherited smooth muscle dysmotility syndrome with severe esophageal, gastric, and intestinal disease. *Hum Mutat*. 2020;41(5):973–82.
66. Juang HH, Chung LC, Sung HC, Feng TH, Lee YH, Chang PL, et al. Metallothionein 3: An androgen-upregulated gene enhances cell invasion and tumorigenesis of prostate carcinoma cells. *The Prostate*. 2013;73(14):1495–506.
67. Uxa S, Castillo-Binder P, Kohler R, Stangner K, Müller GA, Engeland K. Ki-67 gene expression. *Cell Death Differ*. 2021 Dec;28(12):3357–70.
68. Lu ZN, Song J, Sun TH, Sun G. UBE2C affects breast cancer proliferation through the AKT/mTOR signaling pathway. *Chin Med J (Engl)*. 2021 Oct 20;134(20):2465–74.
69. Wang T, Lu J, Wang R, Cao W, Xu J. TOP2A promotes proliferation and metastasis of hepatocellular carcinoma regulated by miR-144-3p. *J Cancer*. 2022 Jan 1;13(2):589–601.
70. Zhang J, Wang Z, Liu Z, Chen Z, Jiang J, Ji Y, et al. CENPF promotes the proliferation of renal cell carcinoma in vitro. *Transl Androl Urol*. 2023 Feb 28;12(2):320–9.
71. Marinković G, Koenis DS, de Camp L, Jablonowski R, Graber N, de Waard V, et al. S100A9 Links Inflammation and Repair in Myocardial Infarction. *Circ Res*. 2020 Aug 14;127(5):664–76.
72. Hong WF, Gu YJ, Wang N, Xia J, Zhou HY, Zhan K, et al. Integrative Characterization of Immune-relevant Genes in Hepatocellular Carcinoma. *J Clin Transl Hepatol*. 2021 Jun 28;9(3):301–14.
73. Maurer M, von Stebut E. Macrophage inflammatory protein-1. *Int J Biochem Cell Biol*. 2004 Oct 1;36(10):1882–6.
74. Croft AP, Campos J, Jansen K, Turner JD, Marshall J, Attar M, et al. Distinct fibroblast subsets drive inflammation and damage in arthritis. *Nature*. 2019 Jun;570(7760):246–51.

75. Jing R, Hu C, Qi T, Yue J, Wang G, Zhang M, et al. FILIP1L-mediated cell apoptosis, epithelial-mesenchymal transition and extracellular matrix synthesis aggravate posterior capsular opacification. *Life Sci.* 2021 Dec 1;286:120061.
76. Yabluchanskiy A, Ma Y, Iyer RP, Hall ME, Lindsey ML. Matrix Metalloproteinase-9: Many Shades of Function in Cardiovascular Disease. *Physiology*. 2013 Nov;28(6):391–403.
77. Higashi T, Friedman SL, Hoshida Y. Hepatic stellate cells as key target in liver fibrosis. *Adv Drug Deliv Rev.* 2017 Nov 1;121:27–42.
78. Östman A. PDGF receptors in tumor stroma: Biological effects and associations with prognosis and response to treatment. *Adv Drug Deliv Rev.* 2017 Nov 1;121:117–23.
79. Fox SB, Gatter KC, Bicknell R, Going JJ, Stanton P, Cooke TG, et al. Relationship of Endothelial Cell Proliferation to Tumor Vascularity in Human Breast Cancer. 1993 Sep 15;(CANCER RESEARCH 53.4161-4163).
80. Hosaka K, Yang Y, Seki T, Fischer C, Dubey O, Fredlund E, et al. Pericyte–fibroblast transition promotes tumor growth and metastasis. *Proc Natl Acad Sci.* 2016 Sep 20;113(38).
81. Domènech M, Hernández A, Plaja A, Martínez-Balibrea E, Balañà C. Hypoxia: The Cornerstone of Glioblastoma. *Int J Mol Sci.* 2021 Nov 22;22(22):12608.
82. Park JH, Lee HK. Current Understanding of Hypoxia in Glioblastoma Multiforme and Its Response to Immunotherapy. *Cancers.* 2022 Feb 24;14(5):1176.
83. Grech N, Dalli T, Mizzi S, Meilak L, Calleja N, Zrinzo A. Rising Incidence of Glioblastoma Multiforme in a Well-Defined Population. *Cureus.* 12(5):e8195.
84. Yuzhalin AE, Yu D. Brain Metastasis Organotropism. *Cold Spring Harb Perspect Med.* 2020 May;10(5):a037242.

Comments to Key References

Schaffernath et al. (7): This paper discusses the brain vasculature in GBM and shows the differences between it and healthy brain vasculature and vasculature in lung cancer. It discusses fibroblasts in the brain and lays forth the hypothesis that they activate and start migrating in disease.

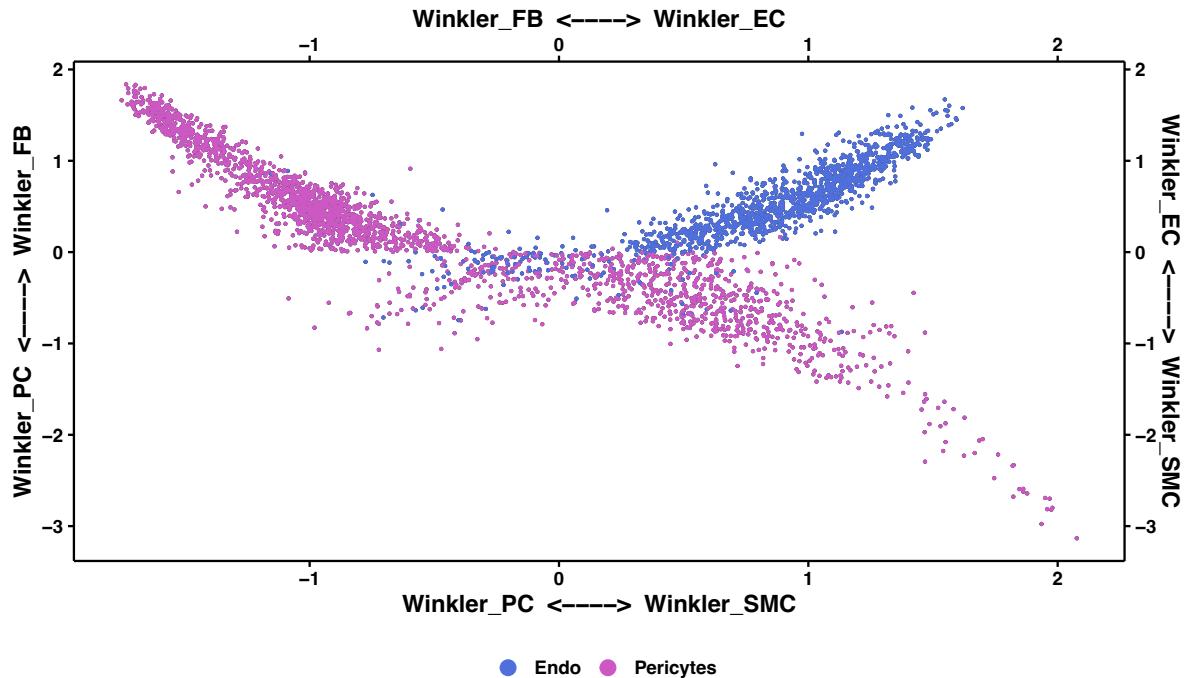
Bartoschek et al. (9): This article discusses CAFs within breast cancer and proposes one of the most known classifications of them. It describes the large heterogeneity among CAFs which has evolved into a big field of study and also inspired part of this project.

Jain et al. (21): This is one of the first articles that have been published which suggests evidence of fibroblast-like cells not only in the perivascular space, meninges, or choroid plexus, but in the actual brain parenchyma which has previously been said to be devoid of fibroblasts. These CAFs, as Jain et al. describe them, open up the possibility of new treatments being developed with this target in mind. As similar results to the article were found in this thesis project, further confirmation of the results has been shown.

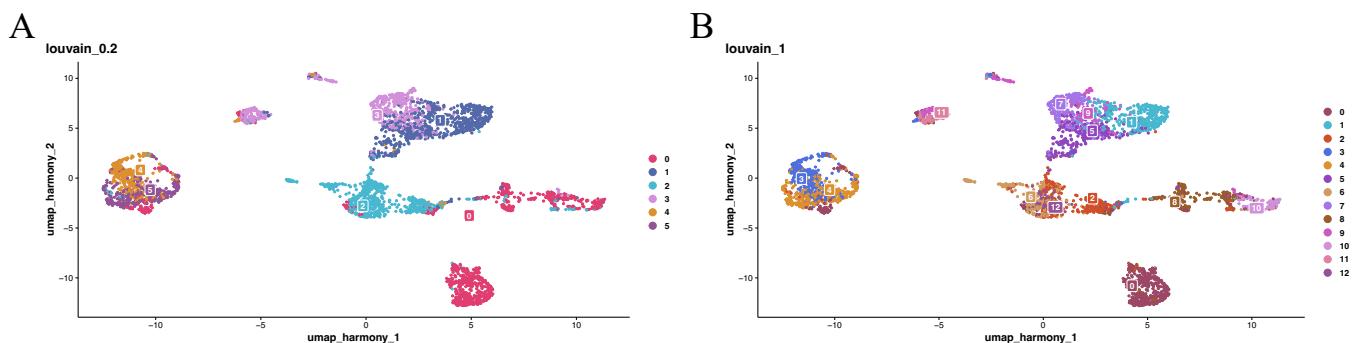
Abdelfattah et al. (22): This paper was especially important for this research project as it provided the dataset on which most of the analysis was done. It, together with many other datasets which are starting to be published with open access, opens the door for research to be made more publicly available and reproducible.

Hao et al. (24): This reference is for the Seurat package in R. This package has been incredibly influential in the single cell analysis field as it provides many functions and analyses that helps when studying this type of data. Though numerous other packages have also been used in this project, Seurat has by far been the most utilized and many of these other packages also have their code based in code first made in Seurat.

Supplementary Materials

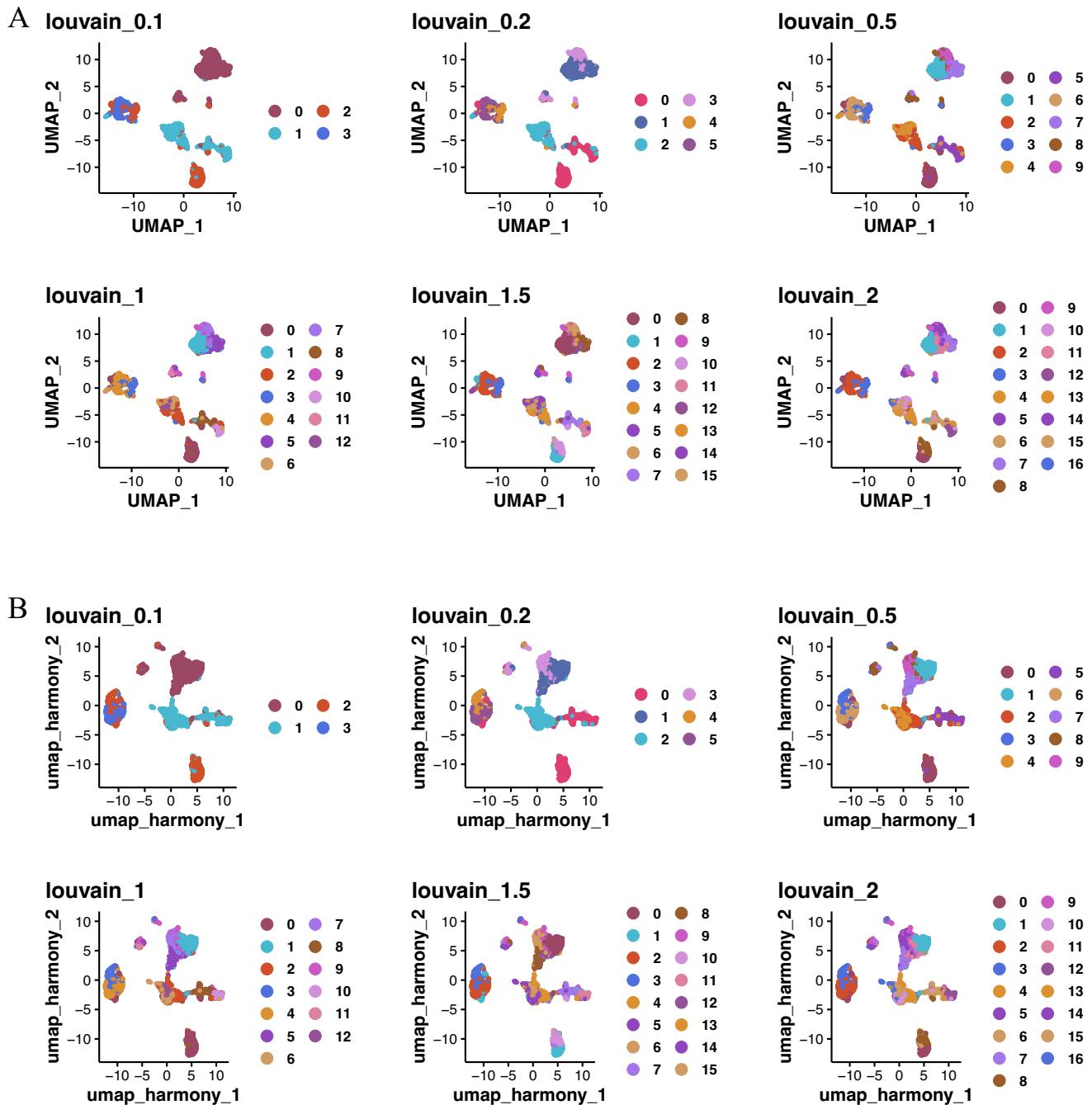


Supplemental Figure 1: Cellular States Plot showing the similarity to different cell types of the cells in the Abdelfattah et al. dataset compared to signature lists by Winkler et al. The blue dots represent cells characterised by Abdelfattah et al. as endothelial cells and lie primarily in the Winkler endothelial (EC) signature, whereas the pink, pericyte identified cells, look more fibroblast-like than pericyte-like and also has more cells that are smooth muscle cell-like than pericyte-like.



Supplemental Figure 2: Re-clustering of dataset with labels.

- (A) Visualisation with Dimplot using resolution 0.2 after Harmony integration and clustering, coloured and labelled based on clusters.
- (B) Visualisation with Dimplot using resolution 1.0 after Harmony integration and clustering, coloured and labelled based on clusters.

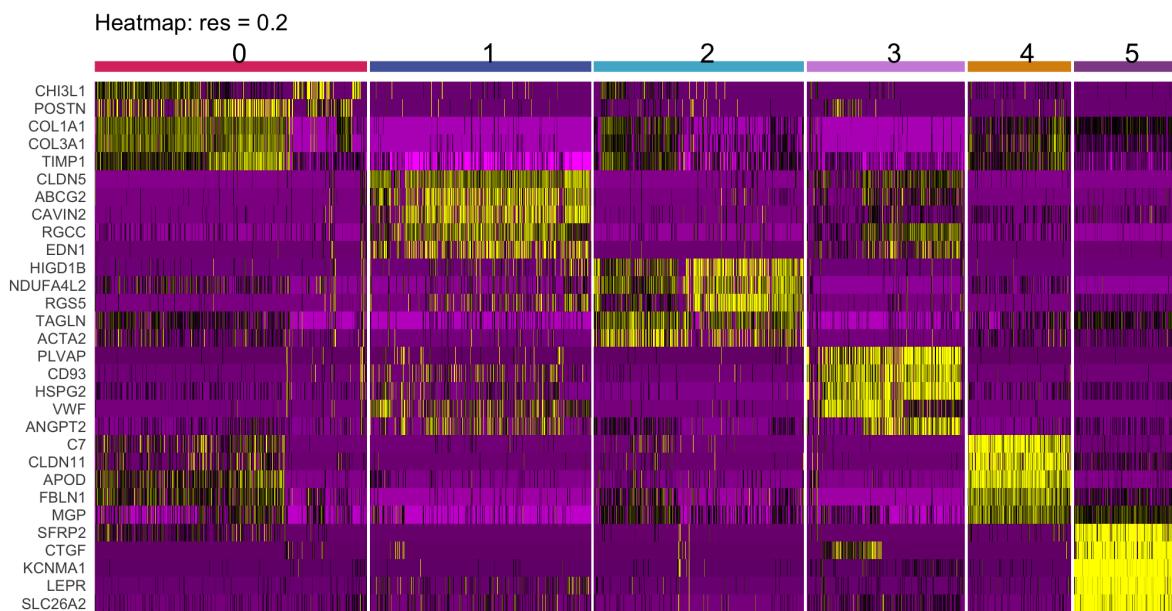


Supplemental Figure 3: Clustering of RPCA and Harmony integrated datasets in different resolutions.

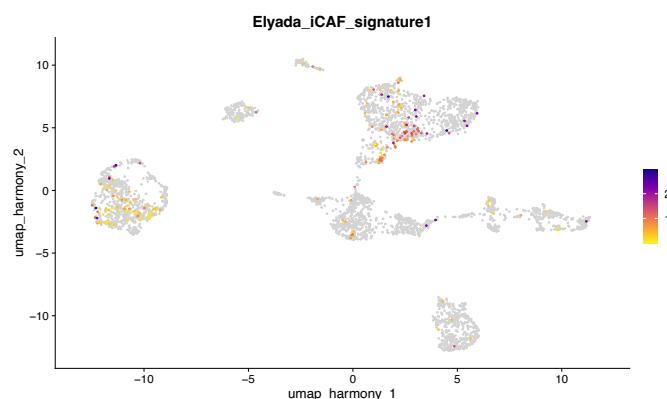
(A) RPCA integration of dataset clustered in resolutions 0.1, 0.2, 0.5, 1.0, 1.5 and 2.0.

(B) Harmony integration of dataset clustered in resolutions 0.1, 0.2, 0.5, 1.0, 1.5 and 2.

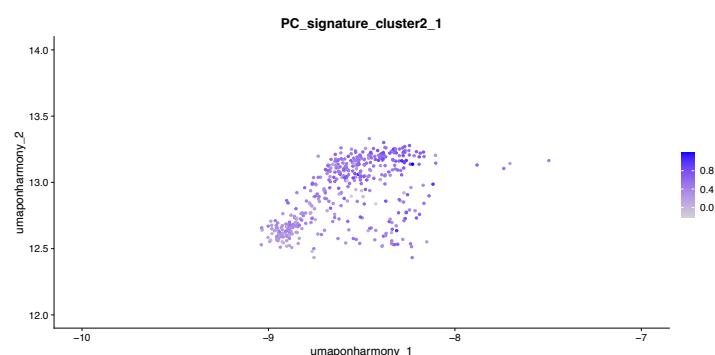
Abbreviations: RPCA: reciprocal principal component analysis



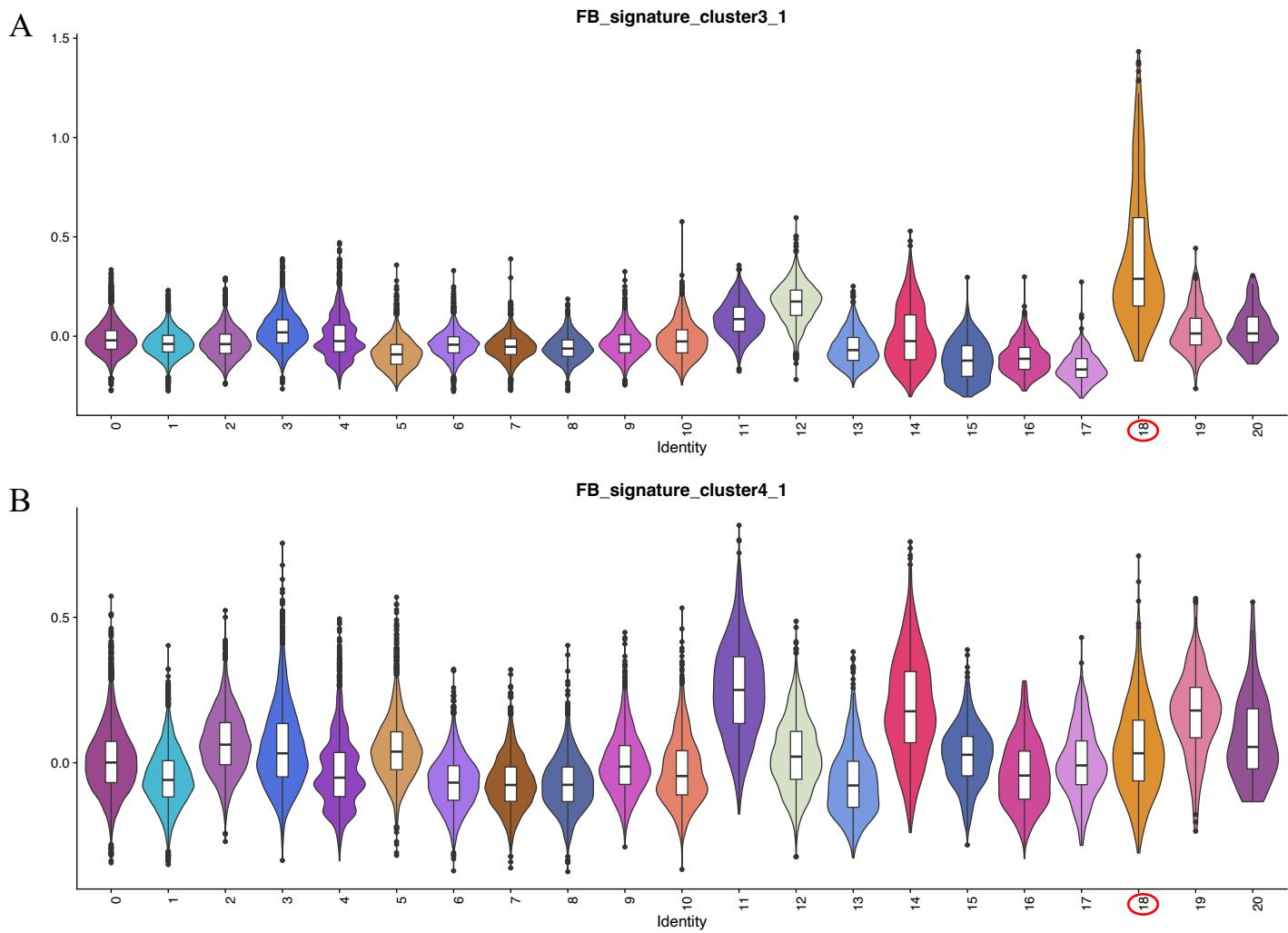
Supplemental Figure 4: Top differentially expressed genes in each cluster with resolution 0.2 visualised in a heatmap.



Supplemental Figure 5: Feature plot with iCAF signature from Elyada et al. on Harmony integrated dataset from Abdelfattah et al.



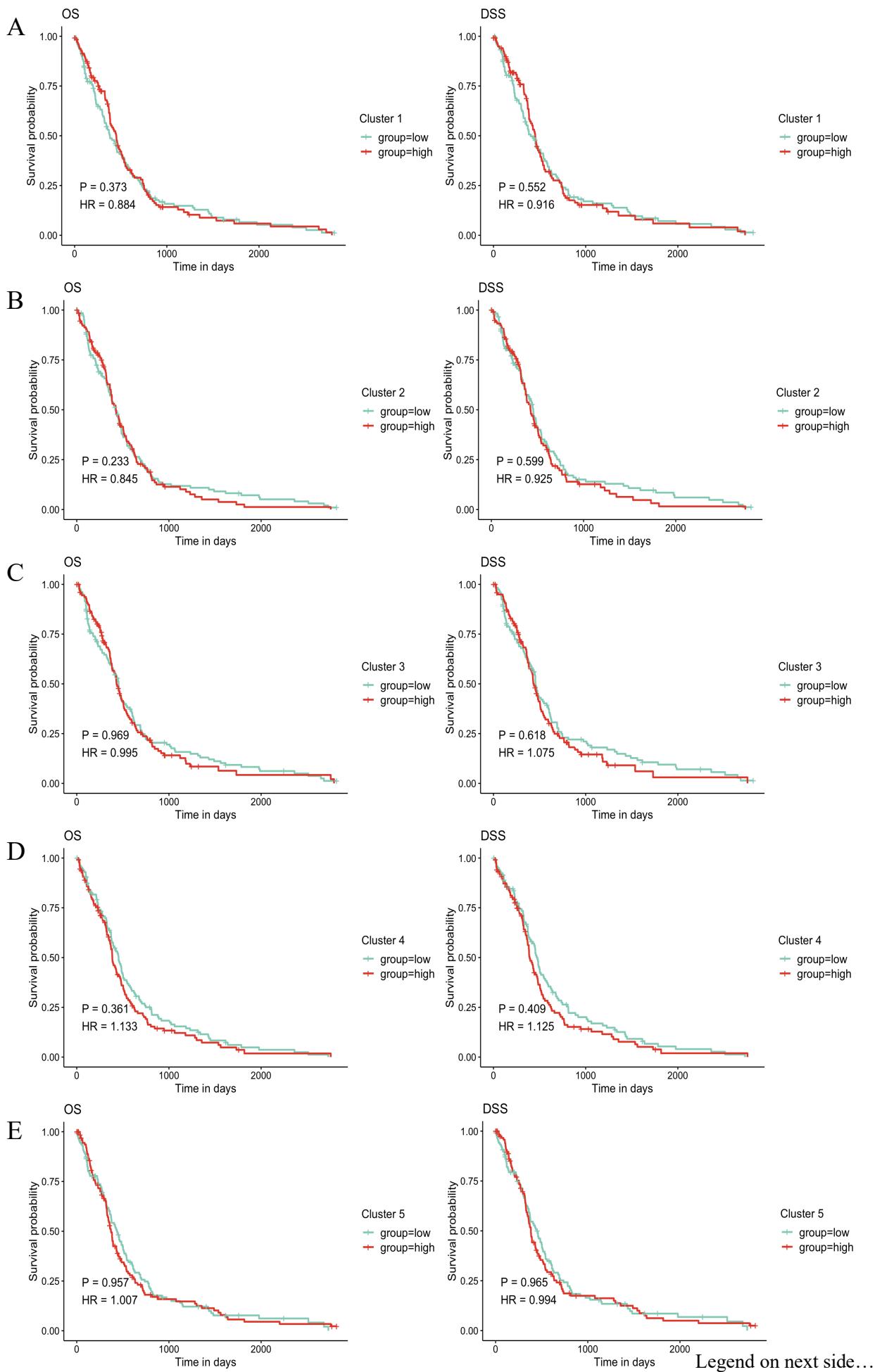
Supplemental Figure 6: Feature plots with Cluster 18 subsetted from dataset from Pietras et al. showing pericyte signature from Cluster 2 of the high resolution from dataset by Abdelfattah et al.



Supplemental Figure 7: Fibroblast signatures in validation dataset

(A) Violin plot showing expression of fibroblast signature from high resolution, 1.0, Cluster 3 identified from dataset from Abdelfattah et al. on each cluster in Pietras et al.'s dataset. Red ring highlights Cluster 18, assigned pericyte cluster from Pietras et al. The y-axis denotes the signature score from the signature.

(B) Violin plot showing expression of fibroblast signature from Cluster 4 identified from dataset from Abdelfattah et al. on each cluster in Pietras et al.'s dataset. Red ring highlights Cluster 18, assigned pericyte cluster from Pietras et al. The y-axis denotes the signature score from the signature.



Supplemental Figure 8: Survival plots based on expression of low resolution cluster signatures.

- (A) Kaplan-Meier survival curves of OS and DSS on GBM population based on signature from low resolution, 0.2, Cluster 1.
(B) Kaplan-Meier survival curves of OS and DSS on GBM population based on signature from low resolution, 0.2, Cluster 2.
(C) Kaplan-Meier survival curves of OS and DSS on GBM population based on signature from low resolution, 0.2, Cluster 3.
(D) Kaplan-Meier survival curves of OS and DSS on GBM population based on signature from low resolution, 0.2, Cluster 4.
(E) Kaplan-Meier survival curves of OS and DSS on GBM population based on signature from low resolution, 0.2, Cluster 5.

Abbreviations: OS: overall survival, DDS: disease-specific survival, GBM: glioblastoma