Joint analysis of single-cell and bulk tissue sequencing data to reveals the adaptation and remodeling of the tumor microenvironment by breast cancer cells at different stages of brain metastasis process

Yan Zhang[1], *Zhenyong Huang*[2], *Xinquan Wu*[3§]

[1]*Department of Integrated Traditional Chinese and Western Medicine Oncology, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, Fujian Province, 362000, China*
[2]*Department of Neurosurgery, Hui'an County Hospital, Quanzhou, Fujian Province, 362000, China*
[3]*Department of Breast Surgery, Women and Children's Hospital, School of Medicine, Xiamen university, Xiamen, Fujian Province, 361003, China*

Correspondence:
Xinquan Wu
Address: Department of Breast Surgery, Women and Children's Hospital, NO.10, Zhenhai Road, Xiamen, Fujian Province, 361003, China. Phone: +86 13506007103; Email:dr.xinquanwu@gmail.com

Abstract
Brain metastasis is a major cause of death in breast cancer (BC) patients.In this study, we integrated scRNA-seq data and bulk tissue mRNA data to explore the adaptation and remodeling of the tumor microenvironment by breast cancer cells during different stages of brain metastasis. Four datasets were utilized in our study, including three with mRNA-seq data from bulk BC tissue and BCBM tissue, and one with scRNA-seq data of BCBM cells. Through a series of bioinformatics analyses, we compared gene expression disparities, metabolic pathways, and functional pathways between BC cells in primary and metastatic sites, as well as between proliferating and dormant cancer cells in BCBM. We also examined the heterogeneity of cancer cells within the metastatic site and compared subgroups between proliferating and dormant cells.BCBM cells exhibited reduced functional pathways related to extracellular matrix (ECM), protein digestion and absorption, compared to primary site cells. Proliferating metastatic cells showed decreased pathways related to ECM, apoptosis, and inflammation, while pathways associated with cell cycle regulation, tumor cell proliferation, and anti-hypoxia were enhanced. Cluster 6 of dormant metastatic cells showed significant enhancement in pathways related to epithelial-mesenchymal transition, ECM, tumor inflammation, and anti-inflammatory response signaling.

Cluster 3 of proliferating metastatic cells exhibited enhanced DNA replication pathways, while cluster 4 showed enhanced tumor proliferation signalin. Our study suggests that the characteristics of BC cells at different stages and the heterogeneity of subgroups play crucial roles in brain metastasis, involving adaptation to the brain microenvironment and remodeling of the tumor microenvironment.
Keywords Breast cancer; brain metastasis; tumor heterogeneity; single-cell sequencing; tumor microenvironment.

Background

Breast cancer (BC) is one of the most common types of cancer among women worldwide and is also one of the leading causes of cancer-related deaths in women (1). One of the major challenges in treating BC is managing distant metastasis, which often leads to disease progression and even death. Brain metastasis is a particularly devastating form of metastasis with poor treatment outcomes, resulting in a higher mortality rate (2). Thorough and comprehensive understanding of the underlying biological behavior of brain metastasis is crucial due to the limited treatment options available for patients with brain metastases. When metastasis begins, BC cells need to depart from the primary site, and the molecular mechanisms involved in this process are not fully understood. Some studies suggest a potential association with the tumor microenvironment (TME) (3). After BC cells reach distant organs through the bloodstream or lymphatic system, some cells remain in an active state and form metastases, while others enter a dormant state. The mechanisms underlying this phenomenon are complex and not yet fully elucidated. One hypothesis is that the tumor microenvironment plays a critical role in determining whether cancer cells remain dormant or proliferate. For instance, the recruitment of immune cells to the microenvironment may contribute to tumor cell proliferation and survival (4-6). Recent studies have emphasized the importance of the microenvironment in BC progression, particularly in the context of brain metastasis. For example, one study found significant differences in the tumor microenvironment between brain metastases and primary breast tumors, suggesting that targeted therapies may need to be adjusted to specific microenvironments in the metastatic sites (7, 8). Another study identified key molecular pathways involved in brain metastasis development, including immune signaling and ECM remodeling (8). These findings may serve as potential targets for novel therapies.

Moreover, multiple studies have confirmed the presence of cellular heterogeneity within the same lesion of BC , which can result in different functions and behaviors of tumor cells. Some cancer cells within the same tumor may exhibit more aggressiveness and higher invasive and metastatic capabilities (9, 10). After BC cells metastasize to the brain, they need to adapt to the different brain microenvironment compared to the primary site. However, there is limited research on whether cancer cells from the same primary site undergo differentiation in the brain and what characteristics are associated with these differentiated populations. Studies have found that triple-negative BC can be classified into different subtypes based on gene expression patterns, each with distinct clinical outcomes and responses to treatment (11). Therefore, gaining a deeper understanding of the heterogeneity of

BCBM holds positive implications for developing more precise and personalized treatment approaches.

In this study, we utilized four datasets obtained from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo). Among them, GSE209998, GSE125989, and GSE43837 are mRNA-seq datasets containing BC tissues and BCBM tissues. The analysis of this portion represents the differences between metastatic and primary lesions. The GSE152818 dataset consists of scRNA-seq data derived from human BC cell lines metastasized to the mouse brain, serving as an animal model for BCBM. Dataset GSM4626851 represents proliferating cancer cells from brain metastases, while dataset GSM4626852 represents dormant cancer cells from brain metastases. This section of the analysis represents the differences between proliferating and dormant cancer cells in BCBM. By conducting bioinformatics analysis on the two parts of data mentioned above, our study compared the differences in metabolic pathways and functional pathways between primary BC cells and metastatic BC cells, as well as between proliferating and dormant cancer cells in BCBM. Additionally, the study compared the differences in cell populations between proliferating and dormant cancer cells in brain metastases. By combining macro-scale transcriptomics and scRNA-seq, this study explored the potential biological behaviors of BC cells during brain metastasis at two different levels, aiming to provide insights for the diagnosis and treatment of BC patients.

Materials and Methods

Bulk tissue and single-cell sequencing datasets

GEO database is a freely accessible public database. In this study, we utilized four datasets obtained from this database. Among them, GSE209998, GSE125989, and GSE43837 are mRNA-seq datasets containing bulk BC tissues and BCBM tissues, providing a total of 44 brain metastasis samples and 79 primary BC samples. Specifically, GSE209998 consists of 9 brain metastasis samples and 44 primary BC samples; GSE125989 consists of 16 brain metastasis samples and 16 primary BC samples; GSE43837 consists of 19 brain metastasis samples and 19 primary BC samples. The raw data have undergone batch correction during the upload process.

The GSE152818 dataset is a scRNA-seq dataset from a mouse model of human BCBM. In this model, the BC cell line HMT-3522-T4-2 was injected into the left ventricle of four 6-8-week-old female mice under ultrasound guidance. The mice were euthanized one month after injection, and brain dissection was performed to obtain samples. For detailed processing methods, please refer to the sample descriptions (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM4626851;https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM4626852). Dataset GSM4626851 represents proliferating BC cells from brain metastases, while dataset GSM4626852 represents dormant BC cells from brain metastases.

Data Preprocessing

First, the GSE209998, GSE125989, and GSE43837 datasets were subjected to data cleaning. After obtaining the expression matrix, we removed rows (genes) with an NA value proportion greater than 50% and columns (samples) with an NA value proportion greater than 50%. Then, the impute.knn function from the impute package

(version 3.17) in R software (http://www.r-project.org/, version 4.3) was used for missing value imputation, with the Number of neighbors set to 10 for data completion.

After downloading the raw data of GSM4626851 and GSM4626852, the Seurat package (version 4.3) in R software (https://satijalab.org/seurat/) was used to convert the scRNA-seq data into seurat objects and filter out low-quality cells. Following the recommended criteria in the Seurat tutorial based on the number of unique genes per cell (**Figure 1**), cells meeting any of the following conditions were defined as low-quality cells in this study: (1) expressing fewer than 200 genes, (2) expressing more than 8000 genes, or (3) having 5% or more unique molecular identifiers (UMIs) mapped to mitochondrial genes. Finally, the ScaleData function from the Seurat package was applied for batch normalization.

Screening of Differentially Expressed Genes

We used the limma package in R software (http://bioconductor.org/packages/release/bioc/html/limma.html, version 3.56) to identify differentially expressed genes (DEGs) between tumor tissues and normal tissues. Fold change (FC) represents the magnitude of gene expression differences. A log |FC| > 1.5 and a p-value < 0.05 were set as the cutoff criteria for determining DEGs. After performing the calculations, differential gene sets were obtained for three datasets: GSE209998, GSE125989, and GSE43837. To visualize the differential results, we used the ggplot2 package in R software (https://ggplot2.tidyverse.org/, version 3.4.1) to create a volcano plot and a heat map, where specific genes are marked.

Perform Robust Rank Aggregation (RRA) analysis on the datasets of DEGs

The three datasets we used were generated under different experimental conditions and utilized different microarray platforms, making it unreliable to merge and analyze the three gene sets together (12). The principle of Robust Rank Aggregation (RRA) (13) was to compare the actual rankings of each dataset with random order null hypotheses and filter DEGs by assigning significance scores to each gene. In this study, RRA calculated the significance scores for all genes by combining the p-values and logFC values from the differential gene lists of the three datasets. The RRA package (https://cran.r-project.org/web/packages/RobustRankAggreg/index.html, version 1.2) was used for meta-analysis of the ranked gene lists. After adjusting for false discovery rate (FDR), a log |FC| > 1 and a p-value < 0.05 were chosen as the threshold for selecting commonly DEGs. We selected the top 10 upregulated and top 10 downregulated genes (a total of 20 genes) with the highest significance to generate a heatmap and proceeded with further analysis.

Perform pathway analysis on the commonly upregulated/downregulated genes

The purpose of gene functional enrichment analysis was to assist researchers in understanding the biological roles of a set of genes. In this study, pathway enrichment analysis was performed using background gene sets from the Hallmark, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Ontology (GO) gene sets in the MSIGDB database (14). The R package clusterProfiler (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html, version 4.8)

was used to conduct GO, KEGG, and Hallmark pathway analyses on the top 10 significantly upregulated and downregulated genes obtained from RRA. The cutoff criteria for pathway analysis were adjusted false discovery rate (FDR) with a p-value < 0.05 and a pathway gene count ≥ 3.

Perform principal component analysis (PCA) and dimensionality reduction on the preprocessed single-cell data

For batch-corrected scRNA-seq data, the FindVariableFeatures function from the Seurat package was used to select highly variable genes. The top 2000 genes with the highest standard deviation were defined as highly variable genes. The RunPCA function from the Seurat package was then utilized to perform PCA on the highly variable genes, calculating the top 20 principal components. Subsequently, the JackStraw plot function from the Seurat package was used to visualize the distribution of p-values for each principal component. The p-value distribution was compared to a uniform distribution (dashed line), and significant principal components were defined as those above the dashed line with statistically significant p-values ($P < 0.05$). Important principal components are retained for visualization and clustering. Next, the FindNeighbors and FindClusters functions from the Seurat package were employed for cell clustering (resolution 0.5). Finally, the RunUMAP function from the Seurat package was used to perform UMAP (15), a nonlinear dimensionality reduction technique, for visualization purposes.

Marker genes identification and cell types clustering

The FindAllMarkers function from the Seurat package was used to identify marker genes in each principal component. A cutoff of $|\log2FC| > 1$ and adjusted p-value < 0.05 was applied, and the top 10 marker genes with the highest rankings in each subcluster were selected for cell type clustering.

Cell type and metabolic pathway analysis of proliferating tumor cell population and dormant tumor cell population

Based on the grouping and cell clustering data, a stacked bar plot of the proportions of different cell groups in each group was generated using the ggplot2 package in R software (https://cran.r-project.org/web/packages/ggplot2/index.html, version 3.4.2). The AverageExpression function was used to calculate the average expression values of genes in the cells included in different groups. Then, the Gene Set Variation Analysis (GSVA) algorithm was applied to load the MSigDB database for differential analysis at the gene set level between different groups. We performed one calculation based on the grouping of cells according to their sources (proliferating cells and dormant cells), and then we performed another calculation by re-grouping the cells based on their clusters. A heatmap was created using the pheatmap package (CRAN.R-project.org/package=pheatmap, version 1.0.12) based on the score of each group for the corresponding pathways.

Results

Identification of DEGs between BCBM and primary tumor and RRA analysis of DEGs

Based on the cutoff criteria, we identified 7,974 DEGs in GSE209998, including 2,038 upregulated genes and 5,936 downregulated genes. In GSE125989, we

identified 979 DEGs, with 307 upregulated genes and 672 downregulated genes. Additionally, in GSE43837, we identified 1,832 DEGs, including 519 upregulated genes and 1,313 downregulated genes. The volcano plots (**Figure 2A-C**) display these DEGs, while the heatmaps (**Figure 2D-F**) show the top 10 upregulated and top 10 downregulated genes. After RRA analysis on the three DEGs sets, 32 commonly upregulated genes and 158 commonly downregulated genes were identified. Heatmaps were generated for the top 10 upregulated genes (GFAP, AQP4, COX6B1, TUBB2B, GPM6A, PPM1H,SLC6A1, AARS, HCN2, SPECC1L) and top 10 downregulated genes (LUM, MFAP5, FABP4, COL10A1, SERPINF1, CILP, SULF1, COL6A3, DPT, CTSK) (**Figure 3**).

Pathway analysis of commonly upregulated/downregulated genes

According to the cutoff criteria, pathway analysis was performed on the top 10 commonly DEGs in terms of upregulation and downregulation. The GO analysis results showed that the commonly differentially expressed genes were significantly enriched in pathways related to the extracellular region, extracellular matrix, and extracellular structural organization. The genes enriched in these pathways were downregulated in the metastatic foci compared to the primary foci (**Figure 4A, Supplementary Table 1**). Hallmark pathway analysis revealed that LUM, MFAP5, and COL6A3 were enriched in the epithelial-mesenchymal transition pathway (adj $P < 0.05$) (**Figure 4B, Supplementary Table 2**). These three genes were downregulated in the metastatic foci compared to the primary foci. No significant enrichment of meaningful pathways was found in the KEGG analysis (**Supplementary Table 3**).

scRNA-seq clustering analysis revealed heterogeneous cell populations in animal models of human BCBM

According to the visualization of P-value distributions for the 20 principal components using the JackStraw function, a total of 14 important principal components were determined and retained for visualization and clustering (**Figure 5**). All cells were projected onto a two-dimensional plane using the UMAP method, and further identified into 9 cell clusters based on cell types (**Figure 6A**). By considering the tissue origin, it can be observed that dormant BC cells in the tumor microenvironment are mainly concentrated in cluster 6, while the majority of proliferating BC cells are composed of the remaining clusters (**Figure 6B, C**). The stacked bar plot of cell proportions in different groups reveals that both the proliferating metastatic tumor cell population and dormant metastatic tumor cell population consist of distinct heterogeneous BC cell clusters, and there are significant differences in the composition proportions of cell clusters. Subcluster 6 dominates in the dormant tumor cell population, while clusters 0, 1, 2, 3, 4, and 5 dominate in the proliferating tumor cell population. Clusters 7 and 8 are present in both the proliferating tumor cell population and dormant tumor cell population, occupying a certain proportion in each (**Figure 6D**).

GSVA analysis based on scRNA-seq data revealed metabolic pathway differences between clusters of proliferating tumor cells and dormant tumor cells

Pathway analysis revealed that compared to the primary tumor, the metastatic lesion exhibited significant weakening of pathways related to ECM and protein digestion

and absorption. When comparing proliferating tumor cells to dormant tumor cells, the pathways associated with ECM and structure, apoptosis, and inflammatory response were significantly attenuated. However, pathways involved in cell cycle regulation, tumor cell proliferation, and hypoxia response were significantly enhanced in proliferating tumor cells compared to dormant metastatic cells (**Figure 7A, Supplementary Table 4**). Further analysis was performed on different heterogeneous subpopulations of metastatic BC cells. The results showed that Subcluster 6 was a unique subpopulation of dormant tumor cells, exhibiting enhanced functions in EMT-related genes, ECM-related genes, collagen formation, tumor inflammation index (TIS), and IL-10 anti-inflammatory signaling pathway. Among the heterogeneous cancer cell subpopulations constituting the proliferating tumor metastatic group, cluster 3 showed significant enhancement in DNA replication pathway, while cluster 4 exhibited significant enhancement in tumor proliferation signaling pathway, and cluster 4 was uniquely associated with proliferating tumor cells (**Figure 7B, Supplementary Table 5**).

Discussion

BC is the most common cancer among women worldwide. According to previous research statistics, approximately 10-30% of patients with metastatic BC have brain metastases (16, 17). The process of metastasis involves a complex series of steps, including invasion, survival, and proliferation of cancer cells in distant organs. Recent studies have emphasized the importance of the tumor microenvironment (TME) in this process. The TME of brain metastases is a complex network composed of cells, signaling molecules, and ECM components that support the growth and survival of cancer cells (18). The ECM is a major component of the tumor microenvironment, providing structural support and regulating cell behavior (19). Changes in the structure and composition of the TME, particularly ECM remodeling, are closely associated with cancer growth, progression, and metastasis (20, 21). In BC, ECM plays a crucial role in invasion and metastasis. Some studies have found that ECM degradation is a common feature of BC invasion and metastasis (22), while others suggest that ECM stiffness and enhancement promote cancer cell migration and invasion (23, 24). The reason for this contradiction may be the different remodeling of the tumor microenvironment by BC cells at different stages of metastasis, exhibiting different biological behaviors.

Our research found that compared to the primary tumor, the metastatic lesion exhibited significant functional pathway attenuation in ECM, extracellular structure, and protein digestion and absorption. We believe that the ECM stiffening caused by BC cells before metastasis may promote cancer cell metastasis. Previous studies have shown that ECM stiffening, characterized by increased collagen deposition and thickening of interstitial collagens, can enhance the invasiveness of breast cells (23, 25). However, during the process of cancer cell colonization and progression in the brain, excessive ECM may hinder cancer cell dissemination. Our research discovered that brain metastatic cancer cells downregulated the extracellular region of ECM, extracellular structure-related pathways, and EMT pathway. Therefore, ECM degradation may be one of the adaptive behaviors of BCBM to the brain

microenvironment, favoring the development of metastatic lesions. Single-cell analysis also confirmed the downregulation of ECM and EMT in the dormant cancer cell population of metastatic lesions, while dormant cancer cells did not proliferate upon brain colonization and retained the biological characteristics of the primary tumor, namely ECM stiffening capability. These facts support the notion that BC cells exhibit different remodeling behaviors of the tumor microenvironment at different stages of metastasis.

Once BC cells establish themselves in the brain, they must adapt to the unique characteristics of the brain TME. In a study by Valiente et al., brain metastatic cells exhibited specific phenotypic changes, including increased proliferation, decreased apoptosis, and enhanced hypoxia response (26). These changes may be related to the brain TME, which is characterized by a low-oxygen environment. Enhancing anti-hypoxia pathways may also be a means by which BC cells adapt to the environment. Another study by Dekker et al. found that the YAP/TAZ mechanotransduction pathway serves as an important link between hypoxia and ECM in the tumor microenvironment, collectively driving BC progression (27). This study indirectly supports our findings.

The heterogeneity of cancer cells within tumors can lead to different functions and behaviors of tumor cells. As early as 1984, Heppne's research indicated that cancer cells within tumors can have different phenotypes and functions, including the ability to invade, grow, and metastasize (28). We analyzed cell subgroups and the results showed that cluster 3 exhibited significantly enhanced DNA replication pathways, while cluster 4 showed significantly enhanced tumor proliferation signaling pathways. Cluster 4 was identified as a unique subgroup of proliferating metastatic cells, similar to the highly proliferative characteristics of proliferating metastatic cancer cells. cluster 1, 3, 4, and 5 exhibited downregulation of ECM construction abilities such as key genes involved in EMT, ECM-related genes, and collagen formation pathways compared to dormant metastatic lesions. Furthermore, their tumor inflammation pathway functions were significantly weakened compared to other subgroups. These behaviors may facilitate tumor cell migration and immune escape. Cluster 6 was identified as a unique subgroup of dormant metastatic lesion cancer cells, which exhibited upregulation of ECM construction abilities and significantly enhanced tumor inflammation pathway function compared to other subgroups. The enhanced matrix and inflammation may cause difficulties in tumor cell migration and aggregation of surrounding immune cells, which could be the reason for these cancer cells entering a dormant state.

Zou et al.'s research suggests that the metastatic ecosystem formed during BCBM exhibits significant immune suppressive cell reprogramming to aid immune escape of brain metastatic cells (29). Another study found the absence of cellular immunity in the microenvironment of BCBM, which may be an important condition for cancer cells to establish themselves in the brain (30). We believe that primary tumor cells may consist of different heterogeneous cancer cell populations, and upon entering the brain environment, populations that adapt to the environment can survive and develop, while those that are not adaptable may gradually become dormant or

even undergo apoptosis. proliferating BC metastatic cells, by reducing inflammatory responses for immune escape, facilitate the growth and colonization of metastatic cancer cells in the brain TME .

In summary, the process of BC metastasis to the brain is complex. This study found that the characteristics of cancer cells at different stages and the heterogeneity of different subgroups of cancer cells play important roles in the process of BC cell colonization and growth from the primary site to brain metastatic lesions. This effect can be achieved through adaptation to the brain environment and reshaping of the tumor microenvironment. By gaining a deeper understanding of these underlying mechanisms, researchers and clinicians can develop more effective treatment strategies, ultimately improving the prognosis of BCBM patients.

Reference

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F: **Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries**. *CA Cancer J Clin* 2021, **71**(3):209-249.
2. Mills MN, Figura NB, Arrington JA, Yu HM, Etame AB, Vogelbaum MA, Soliman H, Czerniecki BJ, Forsyth PA, Han HS *et al*: **Management of brain metastases in breast cancer: a review of current practices and emerging treatments**. *Breast Cancer Res Treat* 2020, **180**(2):279-300.
3. Pickup MW, Mouw JK, Weaver VM: **The extracellular matrix modulates the hallmarks of cancer**. *EMBO Rep* 2014, **15**(12):1243-1253.
4. Malladi S, Macalinao DG, Jin X, He L, Basnet H, Zou Y, de Stanchina E, Massague J: **Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT**. *Cell* 2016, **165**(1):45-60.
5. Yamauchi M, Barker TH, Gibbons DL, Kurie JM: **The fibrotic tumor stroma**. *J Clin Invest* 2018, **128**(1):16-25.
6. Oskarsson T, Batlle E, Massague J: **Metastatic stem cells: sources, niches, and vital pathways**. *Cell Stem Cell* 2014, **14**(3):306-321.
7. Ogiya R, Niikura N, Kumaki N, Yasojima H, Iwasa T, Kanbayashi C, Oshitanai R, Tsuneizumi M, Watanabe KI, Matsui A *et al*: **Comparison of immune microenvironments between primary tumors and brain metastases in patients with breast cancer**. *Oncotarget* 2017, **8**(61):103671-103681.
8. Liang Y, Zhang H, Song X, Yang Q: **Metastatic heterogeneity of breast cancer: Molecular mechanism and potential therapeutic targets**. *Semin Cancer Biol* 2020, **60**:14-27.
9. Asleh K, Riaz N, Nielsen TO: **Heterogeneity of triple negative breast cancer: Current advances in subtyping and treatment implications**. *J Exp Clin Cancer Res* 2022, **41**(1):265.
10. Marusyk A, Tabassum DP, Altrock PM, Almendro V, Michor F, Polyak K: **Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity**. *Nature* 2014, **514**(7520):54-58.

11. Jovanovic B, Mayer IA, Mayer EL, Abramson VG, Bardia A, Sanders ME, Kuba MG, Estrada MV, Beeler JS, Shaver TM *et al*: **A Randomized Phase II Neoadjuvant Study of Cisplatin, Paclitaxel With or Without Everolimus in Patients with Stage II/III Triple-Negative Breast Cancer (TNBC): Responses and Long-term Outcome Correlated with Increased Frequency of DNA Damage Response Gene Mutations, TNBC Subtype, AR Status, and Ki67**. *Clin Cancer Res* 2017, **23**(15):4035-4045.
12. Marshall E: **Getting the noise out of gene arrays**. *Science* 2004, **306**(5696):630-631.
13. Kolde R, Laur S, Adler P, Vilo J: **Robust rank aggregation for gene list integration and meta-analysis**. *Bioinformatics* 2012, **28**(4):573-580.
14. Liberzon A: **A description of the Molecular Signatures Database (MSigDB) Web site**. *Methods Mol Biol* 2014, **1150**:153-160.
15. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F, Newell EW: **Dimensionality reduction for visualizing single-cell data using UMAP**. *Nat Biotechnol* 2018.
16. Tabouret E, Chinot O, Metellus P, Tallet A, Viens P, Goncalves A: **Recent trends in epidemiology of brain metastases: an overview**. *Anticancer Res* 2012, **32**(11):4655-4662.
17. Saha A, Ghosh SK, Roy C, Choudhury KB, Chakrabarty B, Sarkar R: **Demographic and clinical profile of patients with brain metastases: A retrospective study**. *Asian J Neurosurg* 2013, **8**(3):157-161.
18. Schulz M, Salamero-Boix A, Niesel K, Alekseeva T, Sevenich L: **Microenvironmental Regulation of Tumor Progression and Therapeutic Response in Brain Metastasis**. *Front Immunol* 2019, **10**:1713.
19. Mongiat M, Buraschi S, Andreuzzi E, Neill T, Iozzo RV: **Extracellular matrix: the gatekeeper of tumor angiogenesis**. *Biochem Soc Trans* 2019, **47**(5):1543-1555.
20. Quail DF, Joyce JA: **Microenvironmental regulation of tumor progression and metastasis**. *Nat Med* 2013, **19**(11):1423-1437.
21. Lu P, Weaver VM, Werb Z: **The extracellular matrix: a dynamic niche in cancer progression**. *J Cell Biol* 2012, **196**(4):395-406.
22. Ronnov-Jessen L, Petersen OW, Bissell MJ: **Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction**. *Physiol Rev* 1996, **76**(1):69-125.
23. Acerbi I, Cassereau L, Dean I, Shi Q, Au A, Park C, Chen YY, Liphardt J, Hwang ES, Weaver VM: **Human breast cancer invasion and aggression correlates with ECM stiffening and immune cell infiltration**. *Integr Biol (Camb)* 2015, **7**(10):1120-1134.
24. Boghaert E, Gleghorn JP, Lee K, Gjorevski N, Radisky DC, Nelson CM: **Host epithelial geometry regulates breast cancer cell invasiveness**. *Proc Natl Acad Sci U S A* 2012, **109**(48):19632-19637.
25. Panzetta V, Verde G, Pugliese M, Artiola V, Arrichiello C, Muto P, Commara M, Netti PA, Fusco S: **Adhesion and Migration Response to Radiation**

Therapy of Mammary Epithelial and Adenocarcinoma Cells Interacting with Different Stiffness Substrates. *Cancers (Basel)* 2020, **12**(5).

26. Valiente M, Obenauf AC, Jin X, Chen Q, Zhang XH, Lee DJ, Chaft JE, Kris MG, Huse JT, Brogi E *et al*: **Serpins promote cancer cell survival and vascular co-option in brain metastasis**. *Cell* 2014, **156**(5):1002-1016.
27. Dekker Y, Le Devedec SE, Danen EHJ, Liu Q: **Crosstalk between Hypoxia and Extracellular Matrix in the Tumor Microenvironment in Breast Cancer**. *Genes (Basel)* 2022, **13**(9).
28. Heppner GH: **Tumor heterogeneity**. *Cancer Res* 1984, **44**(6):2259-2265.
29. Zou Y, Ye F, Kong Y, Hu X, Deng X, Xie J, Song C, Ou X, Wu S, Wu L *et al*: **The Single-Cell Landscape of Intratumoral Heterogeneity and The Immunosuppressive Microenvironment in Liver and Brain Metastases of Breast Cancer**. *Adv Sci (Weinh)* 2023, **10**(5):e2203699.
30. Jiang J, Wu L, Yuan F, Ji J, Lin X, Yang W, Wu J, Shi M, Yang H, Ma Y *et al*: **Characterization of the immune microenvironment in brain metastases from different solid tumors**. *Cancer Med* 2020, **9**(7):2299-2308.

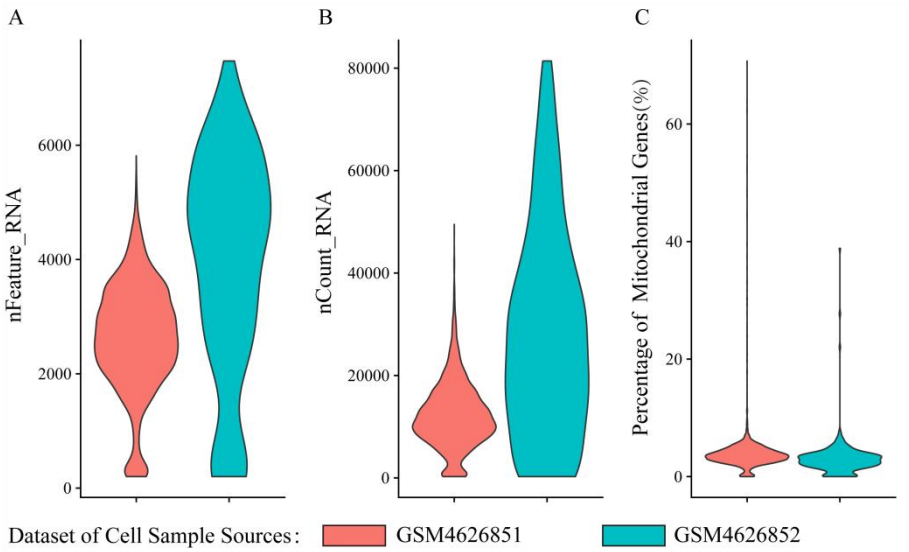

Figure 1. Plots depicting the number of genes detected in each cell (nFeature_RNA) (A), the number of unique RNA molecules detected in each cells (nCount-RNA) (B), and the percentage of reads mapping to the mitochondrial genome (percent.mt) (C).

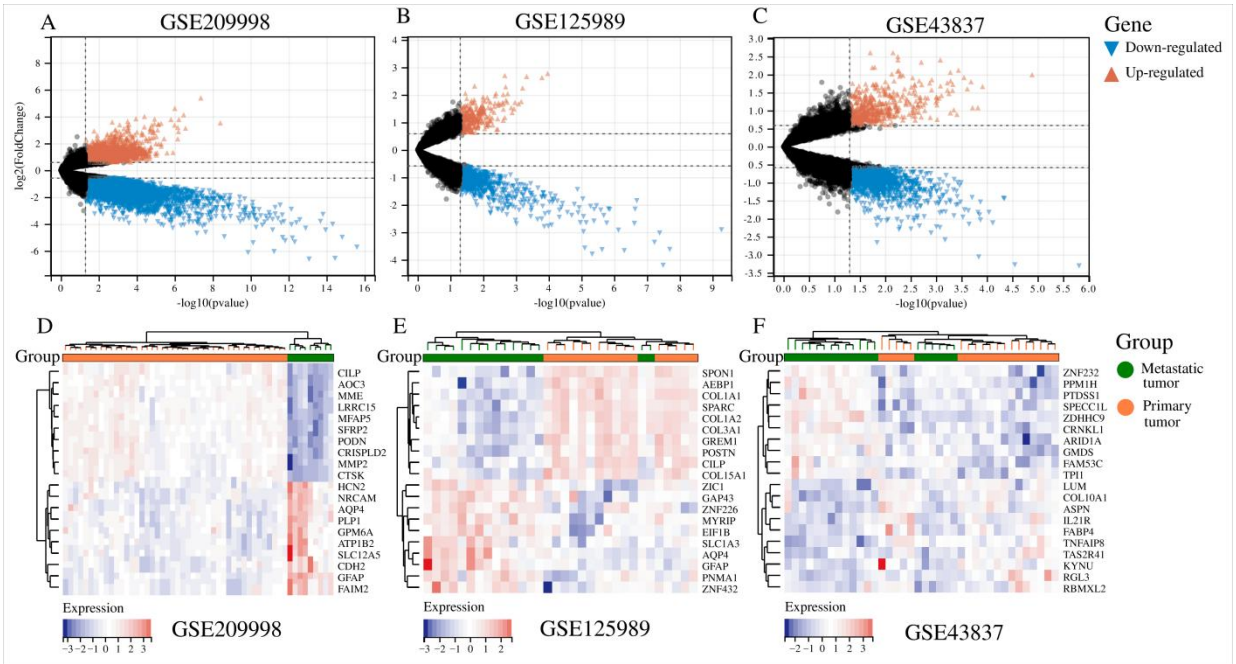

Figure 2. Identification of differentially expressed genes (DEGs) in three gene expression omnibus (GEO) datasets. (A-C) The volcano plots of differentially expressed genes in three datasets. Red represents up-regulated differential genes, black represents no significant difference genes, and blue represents down-regulated differential genes.（D-F）Heatmap of the top 10 upregulated and downregulated differentially expressed genes in three datasets. Blue represents down-regulated expression and red up-regulated expression. The green color of the top bars represents the metastatic tumor group, and the orange color of the top bars represents the primary tumor group.

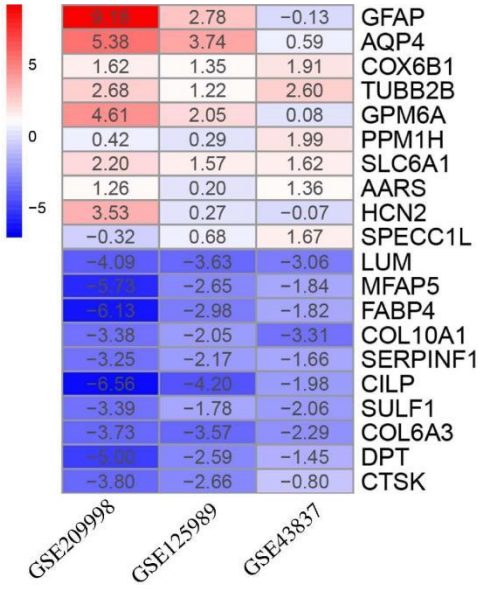

Figure 3. Heatmap shows the top 10 commonly upregulated and top 10 commonly downregulated genes in the RRA analysis of 3 bulk tissue mRNA-seq datasets. Red represents up-regulated genes, while blue represents down-regulated genes. The depth of the color represents the size of the log2 FC.

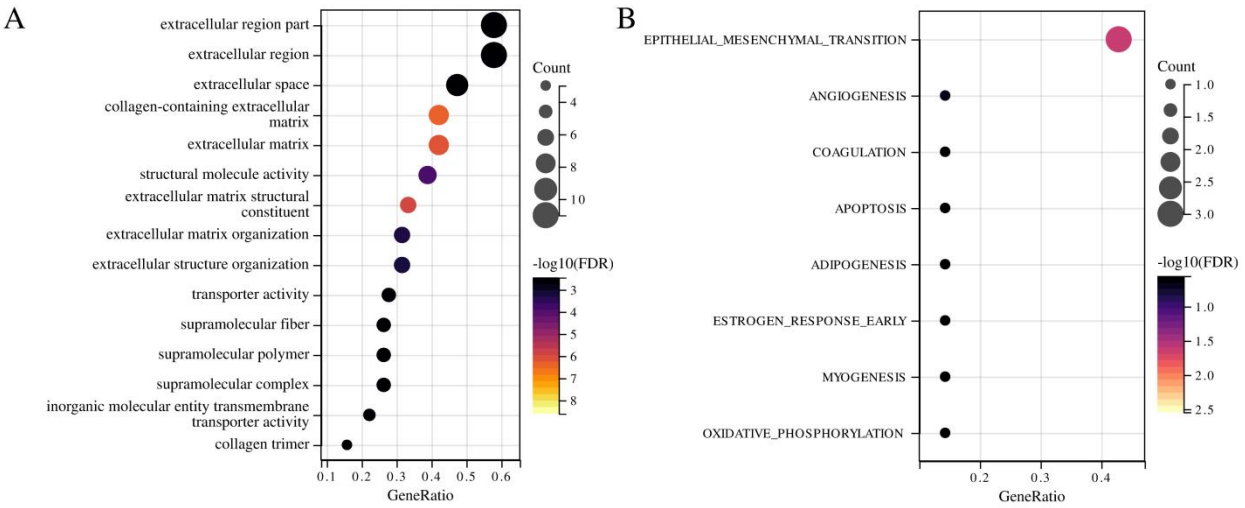

Figure4. GO and Hallmark pathway analysis of commonly upregulated/downregulated genes. (A) GO pathway analysis. (B) Hallmark pathway analysis.

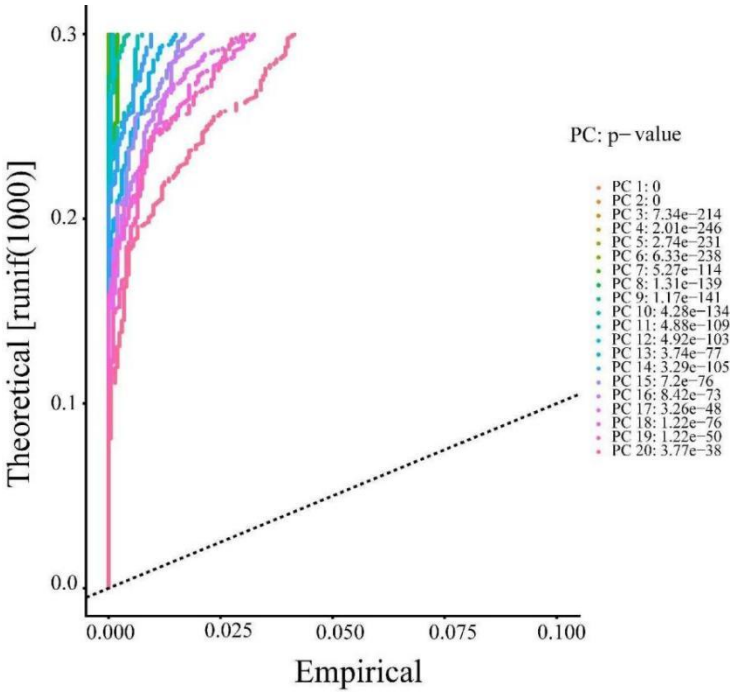

Figure5. Jackstraw plot shows the p-value distribution of each principal component (PC). Significant PCs were defined as those above the dashed line with statistically significant p-values (P < 0.05).

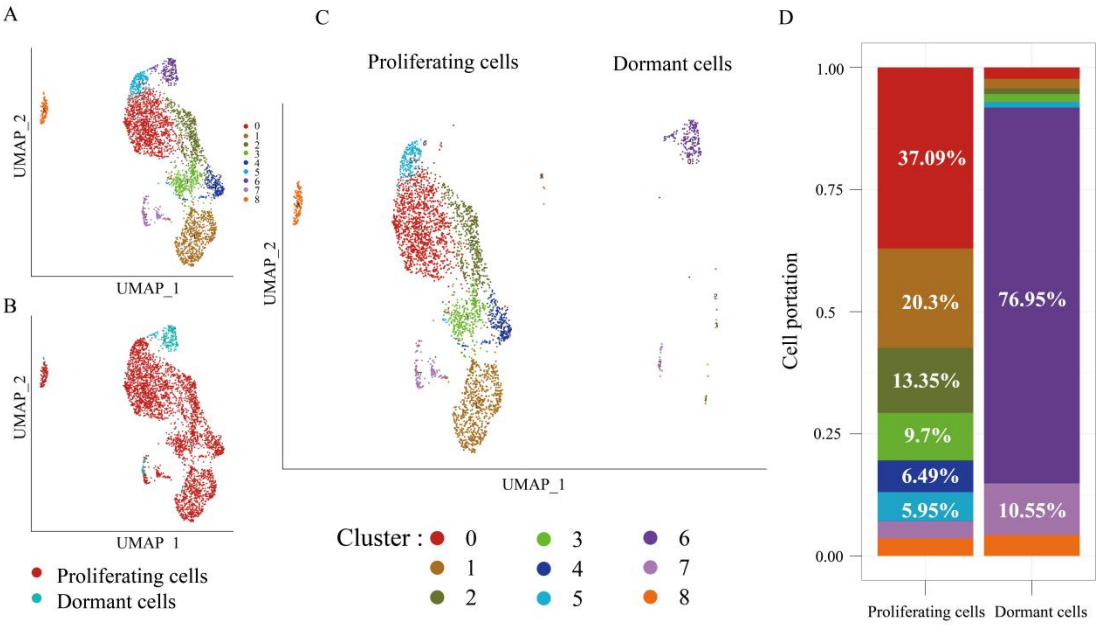

Figure6. scRNA-seq analysis reveals cell cluster heterogeneity in animal models of human BCBM. (A) Cells were clustered into 9 types by the UMAP dimensionality reduction algorithm, with each color representing a cluster. (B) Cells were labeled with two colors according to their source. Red represents proliferating cells, and blue represents dormant cells. (C-D) Cells were labeled with nine colors according to their clusters. Bar charts show the proportion of cell clusters in proliferating cells and dormant cells.

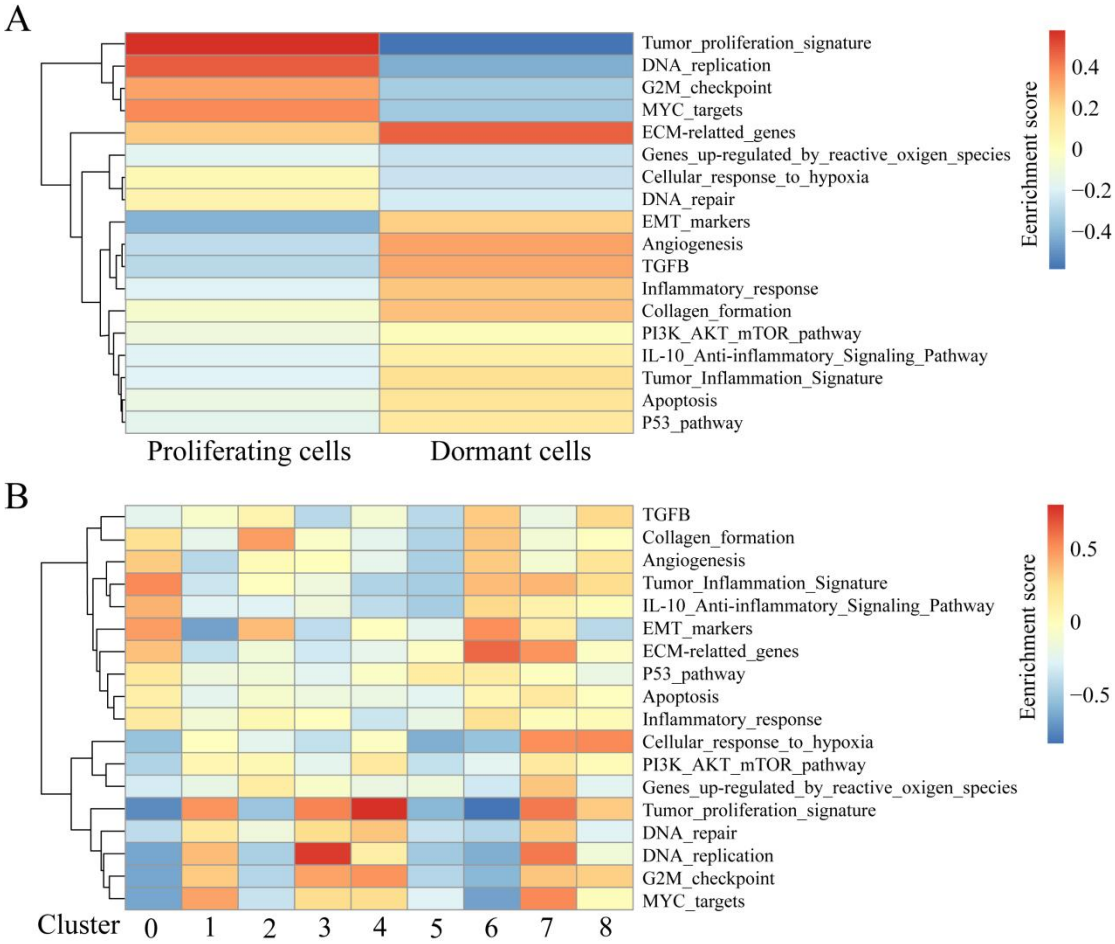

Figure7. The heatmap displays the results of GSVA enrichment analysis performed on the scRNA-seq data. Red color represents positive enrichment scores (pathway activation), while blue color represents negative enrichment scores (pathway inhibition). The depth of color represents the magnitude of the enrichment score, with darker colors indicating higher scores and lighter colors indicating lower scores. (A) Divide the heatmap into two columns based on cell sources (proliferating cells and dormant cells). Each row represents a metabolic pathway. (B) Divide the heatmap into nine columns based on cell clusters (0-8). Each row represents a metabolic pathway.

Acknowledgments
We thank GEO for generating the scRNA-seq data and mRNA-seq data.

Author Contributions
ZY and ZYH wrote the manuscript, designed the study and interpreted the data. XW performed the statistical analysis and interpreted the results. XW was involved in designing the study and critically revising the manuscript. All authors read and approved the final manuscript.

Funding
The study was supported by the Startup Fund for Scientific Research, Fujian Medical University (Grant number: 2020QH1115).

Conflicts of Interest
The authors declare no conflict of interest.

Ethics approval and consent to participate
Not applicable.

Availability of data and materials
The datasets analyzed during the current study are available in Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/).

Supplementary Materials

Supplementary Table 1. GO enrichment analysis results of DEGs in bulk tissue sequencing data.

ID	Pathway	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count	ONT
GO:0062023	collagen-containing extracellular matrix	8/19	399/18675	2.49115381567067e-09	3.6869076471926e-07	2.59604450264628e-07	LUM/MFAP5/COL10A1/SERPINF1/CILP/SULF1/COL6A3/DPT	8	CC
GO:0031012	extracellular matrix	8/19	468/18675	8.69257518177506e-09	6.43250563451355e-07	4.52928917366175e-07	LUM/MFAP5/COL10A1/SERPINF1/CILP/SULF1/COL6A3/DPT	8	CC
GO:0005201	extracellular matrix structural constituent	6/18	158/16967	1.00343223324784e-08	1.07367248957518e-06	7.07683785553737e-07	LUM/MFAP5/COL10A1/CILP/COL6A3/DPT	6	MF
GO:0005198	structural molecule activity	7/18	653/16967	2.66023890179152e-06	0.000142322781245847	9.38084244315958e-05	GFAP/LUM/MFAP5/COL10A1/CILP/COL6A3/DPT	7	MF
GO:0030198	extracellular matrix organization	6/19	334/17910	8.89068725288633e-07	0.000671246887592918	0.000550286747862859	LUM/MFAP5/COL10A1/SULF1/COL6A3/CTSK	6	BP
GO:0043062	extracellular structure organization	6/19	387/17910	2.09367473563201e-06	0.000790362212701082	0.000647937233974537	LUM/MFAP5/COL10A1/SULF1/COL6A3/CTSK	6	BP
GO:0005581	collagen trimer	3/19	86/18675	8.66233880908265e-05	0.00368526656518171	0.00259488897548356	LUM/COL10A1/COL6A3	3	CC
GO:0044421	extracellular region part	11/19	3312/18675	9.96017990589651e-05	0.00368526656518171	0.00259488897548356	GPM6A/LUM/MFAP5/FABP4/COL10A1/SERPINF1/CILP/SULF1/COL6A3/DPT/CTSK	11	CC
GO:0044420	extracellular matrix component	2/19	47/18675	0.00103156862863166	0.0218752478838201	0.0154029128058193	LUM/MFAP5	2	CC
GO:0005576	extracellular region	11/19	4295/18675	0.00109244259410598	0.0218752478838201	0.0154029128058193	GPM6A/LUM/MFAP5/FABP4/COL10A1/SERPINF1/CILP/SULF1/COL6A3/DPT/CTSK	11	CC
GO:0099512	supramolecular fiber	5/19	845/18675	0.00128209544771797	0.0218752478838201	0.0154029128058193	GFAP/TUBB2B/LUM/MFAP5/COL6A3	5	CC
GO:0099081	supramolecular polymer	5/19	851/18675	0.00132329063833896	0.0218752478838201	0.0154029128058193	GFAP/TUBB2B/LUM/MFAP5/COL6A3	5	CC
GO:0099080	supramolecular complex	5/19	852/18675	0.00133025156050258	0.0218752478838201	0.0154029128058193	GFAP/TUBB2B/LUM/MFAP5/COL6A3	5	CC
GO:0005615	extracellular space	9/19	3051/18675	0.00157449724211684	0.0233025591833292	0.0164079186283755	GPM6A/LUM/FABP4/SERPINF1/CILP/SULF1/COL6A3/DPT/CTSK	9	CC
GO:0030020	extracellular matrix structural constituent conferring tensile strength	2/18	41/16967	0.000850560226337338	0.0303366480726984	0.0199956263735444	COL10A1/COL6A3	2	MF
GO:0099056	integral component of presynaptic membrane	2/19	74/18675	0.00253567131510704	0.0313497689323773	0.0220741616237934	GPM6A/SLC6A1	2	CC
GO:0005589	collagen type VI trimer	1/19	3/18675	0.0030492676825864	0.0313497689323773	0.0220741616237934	COL6A3	1	CC
GO:0098647	collagen beaded filament	1/19	3/18675	0.0030492676825864	0.0313497689323773	0.0220741616237934	COL6A3	1	CC
GO:0098855	HCN channel complex	1/19	3/18675	0.0030492676825864	0.0313497689323773	0.0220741616237934	HCN2	1	CC
GO:0098889	intrinsic component of presynaptic membrane	2/19	83/18675	0.00317734144584905	0.0313497689323773	0.0220741616237934	GPM6A/SLC6A1	2	CC
GO:0043202	lysosomal lumen	2/19	94/18675	0.00405407259760963	0.0375001715278891	0.0264048149449575	LUM/CTSK	2	CC
GO:0036021	endolysosome lumen	1/19	5/18675	0.00507721683134765	0.0417460050577474	0.029394413234118	CTSK	1	CC

GO:0097450	astrocyte end-foot	1/19	5/18675	0.00507721683134765	0.0417460050577474	0.029394413234118	AQP4	1	CC
GO:0005518	collagen binding	2/18	61/16967	0.00187451383153867	0.044365524930905	0.0292424020695587	LUM/CTSK	2	MF
GO:0005215	transporter activity	5/18	908/16967	0.00207315537060304	0.044365524930905	0.0292424020695587	AQP4/COX6B1/GPM6A/HCN2/FABP4	5	MF
GO:0005222	intracellular cAMP-activated cation channel activity	1/18	3/16967	0.00317946064676877	0.0477294459121358	0.0314596446248214	HCN2	1	MF
GO:0008449	N-acetylglucosamine-6-sulfatase activity	1/18	3/16967	0.00317946064676877	0.0477294459121358	0.0314596446248214	SULF1	1	MF
GO:0015318	inorganic molecular entity transmembrane transporter activity	4/18	619/16967	0.00356855670371109	0.0477294459121358	0.0314596446248214	AQP4/COX6B1/GPM6A/HCN2	4	MF
GO:0042383	sarcolemma	2/19	125/18675	0.00705457585395906	0.0549514329676811	0.0386926875092491	AQP4/COL6A3	2	CC
GO:0022838	substrate-specific channel activity	3/18	338/16967	0.00512032895116282	0.0566435999034509	0.0373351814415269	AQP4/GPM6A/HCN2	3	MF
GO:0005324	long-chain fatty acid transporter activity	1/18	5/16967	0.00529379438350008	0.0566435999034509	0.0373351814415269	FABP4	1	MF
GO:0015267	channel activity	3/18	370/16967	0.00658061291854321	0.0591162399526348	0.0389649589456619	AQP4/GPM6A/HCN2	3	MF
GO:0022803	passive transmembrane transporter activity	3/18	371/16967	0.0066298586862768	0.0591162399526348	0.0389649589456619	AQP4/GPM6A/HCN2	3	MF
GO:0043203	axon hillock	1/19	8/18675	0.00811181441318476	0.0600274266575672	0.0422668224686995	SERPINF1	1	CC
GO:0042734	presynaptic membrane	2/19	141/18675	0.00889791131486098	0.0627090892666393	0.0441550486301372	GPM6A/SLC6A1	2	CC
GO:0005221	intracellular cyclic nucleotide activated cation channel activity	1/18	8/16967	0.00845735495350186	0.0644448496176901	0.0424771758424519	HCN2	1	MF
GO:0043855	cyclic nucleotide-gated ion channel activity	1/18	8/16967	0.00845735495350186	0.0644448496176901	0.0424771758424519	HCN2	1	MF
GO:0015250	water channel activity	1/18	9/16967	0.00950976179416707	0.0644448496176901	0.0424771758424519	AQP4	1	MF
GO:0022857	transmembrane transporter activity	4/18	838/16967	0.0103781437837907	0.0644448496176901	0.0424771758424519	AQP4/COX6B1/GPM6A/HCN2	4	MF
GO:0005372	water transmembrane transporter activity	1/18	10/16967	0.0105611136214876	0.0644448496176901	0.0424771758424519	AQP4	1	MF
GO:0022890	inorganic cation transmembrane transporter activity	3/18	444/16967	0.0108411896553123	0.0644448496176901	0.0424771758424519	COX6B1/GPM6A/HCN2	3	MF
GO:0001527	microfibril	1/19	10/18675	0.0101300057451985	0.0659348403846668	0.0464263812096872	MFAP5	1	CC
GO:0099699	integral component of synaptic membrane	2/19	152/18675	0.0102772435691404	0.0659348403846668	0.0464263812096872	GPM6A/SLC6A1	2	CC
GO:0005583	fibrillar collagen trimer	1/19	11/18675	0.0111376419568694	0.0659348403846668	0.0464263812096872	LUM	1	CC
GO:0098643	banded collagen fibril	1/19	11/18675	0.0111376419568694	0.0659348403846668	0.0464263812096872	LUM	1	CC
GO:0099240	intrinsic component of synaptic membrane	2/19	164/18675	0.0118834327891594	0.066568790530203	0.0468727614686351	GPM6A/SLC6A1	2	CC
GO:0097449	astrocyte projection	1/19	12/18675	0.01214430638051	0.066568790530203	0.0468727614686351	AQP4	1	CC

GO:0005775	vacuolar lumen	2/19	171/18675	0.0128683011067856	0.0680181629930096	0.0478933011117208	LUM/CTSK	2	CC
GO:0004065	arylsulfatase activity	1/18	12/16967	0.012660656217103	0.0712994850121063	0.046995233603651	SULF1	1	MF
GO:0099059	integral component of presynaptic active zone membrane	1/19	14/18675	0.0141547234037187	0.0722378987500127	0.0508645233019293	GPM6A	1	CC
GO:0008324	cation transmembrane transporter activity	3/18	484/16967	0.0136830741741479	0.0732044468316913	0.0482508405088373	COX6B1/GPM6A/HCN2	3	MF
GO:0036041	long-chain fatty acid binding	1/18	14/16967	0.0147559906935575	0.0751852859147929	0.049556459973351	FABP4	1	MF
GO:0070062	extracellular exosome	6/19	2162/18675	0.0170953698184108	0.0768074961127827	0.054082091857507	GPM6A/LUM/FABP4/SERPINF1/CILP/COL6A3	6	CC
GO:1903561	extracellular vesicle	6/19	2185/18675	0.0179509826473051	0.0768074961127827	0.054082091857507	GPM6A/LUM/FABP4/SERPINF1/CILP/COL6A3	6	CC
GO:0043230	extracellular organelle	6/19	2187/18675	0.0180267976198261	0.0768074961127827	0.054082091857507	GPM6A/LUM/FABP4/SERPINF1/CILP/COL6A3	6	CC
GO:0036019	endolysosome	1/19	18/18675	0.0181639348915364	0.0768074961127827	0.054082091857507	CTSK	1	CC
GO:0098644	complex of collagen trimers	1/19	18/18675	0.0181639348915364	0.0768074961127827	0.054082091857507	LUM	1	CC
GO:0098945	intrinsic component of presynaptic active zone membrane	1/19	18/18675	0.0181639348915364	0.0768074961127827	0.054082091857507	GPM6A	1	CC
GO:0008484	sulfuric ester hydrolase activity	1/18	15/16967	0.0158020823686301	0.0768555824292465	0.0506573932391492	SULF1	1	MF
GO:0005248	voltage-gated sodium channel activity	1/18	17/16967	0.0178911195478486	0.0803142945008543	0.0529371149194023	HCN2	1	MF
GO:0004129	cytochrome-c oxidase activity	1/18	19/16967	0.0199759684294397	0.0803142945008543	0.0529371149194023	COX6B1	1	MF
GO:0015002	heme-copper terminal oxidase activity	1/18	19/16967	0.0199759684294397	0.0803142945008543	0.0529371149194023	COX6B1	1	MF
GO:0016676	oxidoreductase activity, acting on a heme group of donors, oxygen as acceptor	1/18	19/16967	0.0199759684294397	0.0803142945008543	0.0529371149194023	COX6B1	1	MF
GO:0016675	oxidoreductase activity, acting on a heme group of donors	1/18	20/16967	0.0210168247291955	0.0803142945008543	0.0529371149194023	COX6B1	1	MF
GO:0030552	cAMP binding	1/18	20/16967	0.0210168247291955	0.0803142945008543	0.0529371149194023	HCN2	1	MF
GO:0030021	extracellular matrix structural constituent conferring compression resistance	1/18	21/16967	0.0220566369174811	0.081381384488637	0.0536404600171046	LUM	1	MF
GO:0001968	fibronectin binding	1/18	24/16967	0.0251698186673962	0.0897723532470465	0.0591711526566858	CTSK	1	MF
GO:0005217	intracellular ligand-gated ion channel activity	1/18	25/16967	0.0262054642593215	0.0904511185724967	0.0596185434762152	HCN2	1	MF
GO:0045202	synapse	4/19	1091/18675	0.0221988776752952	0.0912620526651024	0.064259909060065	TUBB2B/GPM6A/PPM1H/SLC6A1	4	CC
GO:0048787	presynaptic active zone membrane	1/19	23/18675	0.0231537296544618	0.0926149186178473	0.0652124960395938	GPM6A	1	CC

GO:0097386	glial cell projection	1/19	24/18675	0.024148801450332	0.0940532267012932	0.0662252449745948	AQP4	1	CC
GO:0005261	cation channel activity	2/18	254/16967	0.0291703844999183	0.0950230354565336	0.0626320056624471	GPM6A/HCN2	2	MF
GO:0005504	fatty acid binding	1/18	28/16967	0.0293061698136973	0.0950230354565336	0.0626320056624471	FABP4	1	MF
GO:0015077	monovalent inorganic cation transmembrane transporter activity	2/18	266/16967	0.0317591746397695	0.0990126561031262	0.0652616621633985	COX6B1/HCN2	2	MF
GO:0015075	ion transmembrane transporter activity	3/18	672/16967	0.0323873174169104	0.0990126561031262	0.0652616621633985	COX6B1/GPM6A/HCN2	3	MF
GO:0005887	integral component of plasma membrane	4/19	1158/18675	0.0269776067626774	0.0990877407219271	0.0697701730545575	AQP4/GPM6A/SLC6A1/HCN2	4	CC
GO:0000323	lytic vacuole	3/19	659/18675	0.0278159536215677	0.0990877407219271	0.0697701730545575	GFAP/LUM/CTSK	3	CC
GO:0005764	lysosome	3/19	659/18675	0.0278159536215677	0.0990877407219271	0.0697701730545575	GFAP/LUM/CTSK	3	CC
GO:0044295	axonal growth cone	1/19	28/18675	0.028119493988655	0.0990877407219271	0.0697701730545575	GPM6A	1	CC
GO:0030551	cyclic nucleotide binding	1/18	32/16967	0.0334259409543508	0.0993493245032094	0.0654835685363013	HCN2	1	MF

Supplementary Table 2. Hallmark enrichment analysis results of DEGs in bulk tissue sequencing data.

Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
EPITHELIAL_MESENCHYMAL_TRANSITION	3/7	200/4383	0.00285880475689797	0.0228704380551838	0.0210648771560903	LUM/MFAP5/COL6A3	3
ANGIOGENESIS	1/7	36/4383	0.0561348800392154	0.224539520156862	0.206812715933952	LUM	1
COAGULATION	1/7	138/4383	0.20076249612173	0.279033363336343	0.257004413599264	CTSK	1
APOPTOSIS	1/7	161/4383	0.230607622268114	0.279033363336343	0.257004413599264	LUM	1
ADIPOGENESIS	1/7	200/4383	0.279033363336343	0.279033363336343	0.257004413599264	FABP4	1
ESTROGEN_RESPONSE_EARLY	1/7	200/4383	0.279033363336343	0.279033363336343	0.257004413599264	TUBB2B	1
MYOGENESIS	1/7	200/4383	0.279033363336343	0.279033363336343	0.257004413599264	COL6A3	1
OXIDATIVE_PHOSPHORYLATION	1/7	200/4383	0.279033363336343	0.279033363336343	0.257004413599264	COX6B1	1

Supplementary Table 3. KEGG enrichment analysis results of DEGs in bulk tissue sequencing data.

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count	one_type
hsa04974	Protein digestion and absorption	2/13	95/7914	0.0102045899121614	0.336751467101325	0.336751467101325	COL10A1/COL6A3	2	Organismal Systems
hsa04962	Vasopressin-regulated water reabsorption	1/13	44/7914	0.0699656968007636	0.338488972116741	0.338488972116741	AQP4	1	Organismal Systems
hsa04923	Regulation of lipolysis in adipocytes	1/13	55/7914	0.0867363550761457	0.338488972116741	0.338488972116741	FABP4	1	Organismal Systems
hsa04929	GnRH secretion New	1/13	64/7914	0.100249687323194	0.338488972116741	0.338488972116741	HCN2	1	Organismal Systems
hsa00970	Aminoacyl-tRNA biosynthesis	1/13	66/7914	0.1032274740064	0.338488972116741	0.338488972116741	AARS	1	Genetic Information Processing
hsa04976	Bile secretion	1/13	72/7914	0.11210633945809	0.338488972116741	0.338488972116741	AQP4	1	Organismal Systems
hsa03320	PPAR signaling pathway	1/13	76/7914	0.117980425067018	0.338488972116741	0.338488972116741	FABP4	1	Organismal Systems
hsa04721	Synaptic vesicle cycle	1/13	78/7914	0.12090399636588	0.338488972116741	0.338488972116741	SLC6A1	1	Organismal Systems
hsa04260	Cardiac muscle contraction	1/13	86/7914	0.132509058117147	0.338488972116741	0.338488972116741	COX6B1	1	Organismal Systems
hsa04512	ECM-receptor interaction	1/13	88/7914	0.135388142751793	0.338488972116741	0.338488972116741	COL6A3	1	Environmental Information Processing
hsa04540	Gap junction	1/13	88/7914	0.135388142751793	0.338488972116741	0.338488972116741	TUBB2B	1	Cellular Processes
hsa04727	GABAergic synapse	1/13	89/7914	0.136824375072803	0.338488972116741	0.338488972116741	SLC6A1	1	Organismal Systems
hsa05323	Rheumatoid arthritis	1/13	93/7914	0.142547310015091	0.338488972116741	0.338488972116741	CTSK	1	Human Diseases
hsa04620	Toll-like receptor signaling pathway	1/13	104/7914	0.158105283345089	0.338488972116741	0.338488972116741	CTSK	1	Organismal Systems
hsa04142	Lysosome	1/13	123/7914	0.184365830826913	0.338488972116741	0.338488972116741	CTSK	1	Cellular Processes
hsa04380	Osteoclast differentiation	1/13	128/7914	0.191149698787406	0.338488972116741	0.338488972116741	CTSK	1	Organismal Systems
hsa00190	Oxidative phosphorylation	1/13	133/7914	0.19788145022877	0.338488972116741	0.338488972116741	COX6B1	1	Metabolism
hsa04210	Apoptosis	1/13	136/7914	0.201895638038981	0.338488972116741	0.338488972116741	CTSK	1	Cellular Processes
hsa05012	Parkinson disease	1/13	142/7914	0.209868439957483	0.338488972116741	0.338488972116741	COX6B1	1	Human Diseases
hsa04932	Non-alcoholic fatty liver disease (NAFLD)	1/13	149/7914	0.219077094513302	0.338488972116741	0.338488972116741	COX6B1	1	Human Diseases

hsa04145	Phagosome	1/13	152/7914	0.222993249428303	0.338488972116741	0.338488972116741	TUBB2B	1	Cellular Processes
hsa04310	Wnt signaling pathway	1/13	160/7914	0.23334787832927	0.338488972116741	0.338488972116741	SERPINF1	1	Environmental Information Processing
hsa04630	Jak-STAT signaling pathway	1/13	162/7914	0.235916556323789	0.338488972116741	0.338488972116741	GFAP	1	Environmental Information Processing
hsa05010	Alzheimer disease	1/13	171/7914	0.247377595399736	0.3394962968195	0.3394962968195	COX6B1	1	Human Diseases
hsa05016	Huntington disease	1/13	193/7914	0.274728728937408	0.3394962968195	0.3394962968195	COX6B1	1	Human Diseases
hsa04510	Focal adhesion	1/13	199/7914	0.282027231764925	0.3394962968195	0.3394962968195	COL6A3	1	Cellular Processes
hsa05130	Pathogenic Escherichia coli infection	1/13	201/7914	0.284444959883068	0.3394962968195	0.3394962968195	TUBB2B	1	Human Diseases
hsa05205	Proteoglycans in cancer	1/13	204/7914	0.288057463968061	0.3394962968195	0.3394962968195	LUM	1	Human Diseases
hsa04024	cAMP signaling pathway	1/13	216/7914	0.302339784091146	0.344041823276132	0.344041823276132	HCN2	1	Environmental Information Processing
hsa04714	Thermogenesis	1/13	231/7914	0.319820715510916	0.351802787062007	0.351802787062007	COX6B1	1	Organismal Systems
hsa05165	Human papillomavirus infection	1/13	330/7914	0.42542970738126	0.452876785276825	0.452876785276825	COL6A3	1	Human Diseases
hsa04151	PI3K-Akt signaling pathway	1/13	354/7914	0.448641470570437	0.462661516525764	0.462661516525764	COL6A3	1	Environmental Information Processing
hsa01100	Metabolic pathways	1/13	1439/7914	0.92654589625078	0.92654589625078	0.92654589625078	COX6B1	1	Metabolism

Supplementary Table 4. Results of GSVA enrichment analysis performed on the scRNA-seq data, which were grouped based on cell sources (proliferating cells and dormant cells).

Pathway	Proliferating cells	Dormant cells
Tumor_Inflammation_Signature	-0.171262175	0.188187818
Cellular_response_to_hypoxia	0.043667379	-0.238397916
Tumor_proliferation_signature	0.58832394	-0.572656497
EMT_markers	-0.401735153	0.239331599
ECM-relatted_genes	0.241610223	0.479810903
Angiogenesis	-0.267314132	0.3386892
Apoptosis	-0.11297245	0.168013017
DNA_repair	0.071036527	-0.212751478
G2M_checkpoint	0.340764695	-0.319396773
Inflammatory_response	-0.175248928	0.261205576
PI3K_AKT_mTOR_pathway	-0.093371947	0.028171619
P53_pathway	-0.150952285	0.145450729
MYC_targets	0.399708948	-0.332056885
TGFB	-0.273439175	0.324281897
IL-10_Anti-inflammatory_Signaling_Pathway	-0.169261852	0.090334055
Genes_up-regulated_by_reactive_oxigen_species_(ROS)	-0.16167908	-0.246458347
DNA_replication	0.490278057	-0.41557711
Collagen_formation	-0.047701567	0.272898204

Supplementary Table 5. Results of GSVA enrichment analysis performed on the scRNA-seq data, which were grouped based on cell cluster.

Pathway	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Tumor_Inflammation_Signature	0.54676056	-0.338551251	-0.008813567	-0.144436135	-0.43409071	-0.465514732	0.377117111	0.407875666	0.27422269
Cellular_response_to_hypoxia	-0.512339305	-0.01123549	-0.219382877	-0.375427307	-0.033046406	-0.599587314	-0.513501406	0.526527822	0.547741163
Tumor_proliferation_signature	-0.734016994	0.519660123	-0.501843057	0.565156788	0.817159555	-0.559588692	-0.817878307	0.600007042	0.337584868
EMT_markers	0.481454652	-0.653742586	0.389911883	-0.379081051	-0.006979076	-0.222245084	0.534978208	0.16114354	-0.403797409
ECM-relatted_genes	0.374381881	-0.373334688	-0.117820471	-0.319178018	-0.208389733	-0.018021246	0.646471191	0.516757072	-0.030823717
Angiogenesis	0.335493002	-0.397246225	0.037512992	0.007203761	-0.2120972	-0.451278246	0.342773985	-0.093072073	0.227725151
Apoptosis	0.129260705	-0.225603079	-0.071524893	-0.148829091	-0.173389483	-0.236074377	0.070386482	0.193278215	-0.010308832
DNA_repair	-0.377316861	0.183105157	-0.134932932	0.26841695	0.350281245	-0.355544586	-0.42350959	0.342410935	-0.260398269
G2M_checkpoint	-0.628887802	0.338814849	-0.415572489	0.458690478	0.521634824	-0.417061712	-0.561271688	0.347817951	0.310385503
Inflammatory_response	0.1713642	-0.110219075	0.055921183	-0.012711872	-0.330013383	-0.190305768	0.229234602	0.025063233	0.041553451
PI3K_AKT_mTOR_pathway	-0.43743092	0.07715685	0.056810542	-0.213011894	0.18055104	-0.371944256	-0.237991786	0.183421782	0.045860957
P53_pathway	0.18433204	-0.136844399	-0.11890501	-0.233742992	-0.044421871	0.151867126	0.152865893	-0.006581697	-0.168633898
MYC_targets	-0.637180968	0.471487599	-0.348102384	0.272592288	0.275956277	-0.256917665	-0.646002641	0.542697204	0.030525086
TGFB	-0.218805819	-0.058958825	0.09467468	-0.393584042	-0.085350132	-0.406402041	0.328823287	-0.16391675	0.278328264
IL-10_Anti-inflammatory_Signaling_Pathway	0.417137037	-0.257100028	-0.255167114	-0.134664092	-0.386772879	-0.463225586	0.29368834	0.112060086	0.022214087
Genes_up-regulated_by_reactive_oxigen_species_(ROS)	-0.310126804	-0.182159922	0.161244084	-0.05963265	-0.164564716	-0.155397178	-0.325861899	0.352551329	-0.230383964
DNA_replication	-0.628876071	0.388242924	-0.442511135	0.773173753	0.143524844	-0.485786532	-0.589473738	0.600888266	-0.122014971
Collagen_formation	0.24917257	-0.201314683	0.478068128	-0.033292505	-0.216707098	-0.434971964	0.344169793	-0.11055777	-0.010305336