Differential Adaptation and Remodeling of the Tumor Microenvironment in Breast Cancer Cells During the Stages of Brain Metastasis

Xinquan Wu[1], Yan Zhang[2], Zhenyong Huang[3], Nan Li[4§]

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

Correspondence:
Nan Li
Address: Department of Anesthesiology, The Second Affiliated Hospital of Fujian Medical University; The Second Clinical Medical College of Fujian Medical University, 34 Zhongshan North Road Licheng District, Quanzhou, Fujian Province, 361003, China. Phone: +86 059522791140; Email: lir78256@gmail.com

Abstract
Objective To integrate macro-transcriptomics and single-cell sequencing to analyze the differential gene expression between primary and brain metastatic breast cancer cells, as well as between active and dormant cancer cells within brain metastases, exploring their adaptation and remodeling abilities at different stages of brain metastasis. **Methods** Four public datasets were used: three mRNA microarray datasets from breast cancer and brain metastasis tissues, and one single-cell RNA sequencing (scRNA-seq) dataset from active and dormant brain metastatic cells. Gene differential expression, pathway enrichment, and cell clustering analyses were performed to compare primary and metastatic breast cancer cells, as well as active and dormant cells in brain metastases, focusing on gene expression, metabolic pathways, and functional pathways. **Results** Metastatic breast cancer cells showed weakened pathways related to the extracellular matrix and protein digestion. Active cells exhibited enhanced cell cycle regulation, tumor proliferation, and hypoxia resistance pathways compared to dormant cells. Clustering analysis revealed that cluster 6, unique to dormant cells, had enhanced functions in epithelial-mesenchymal transition (EMT), extracellular matrix (ECM), collagen formation, tumor inflammation siganure

(TIS), and IL-10 signaling. Cluster 3 and 4 in active cells had enhanced DNA replication and tumor proliferation pathways, respectively. **Conclusion** This study highlights the role of cancer cell characteristics and heterogeneity in the invasiveness of brain metastasis. Understanding these mechanisms can guide the development of more effective treatment strategies for breast cancer brain metastasis (BCBM) patients.
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 a leading cause 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). A 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 must depart from the primary site, though 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 breast cancer brain metastasis (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 these, GSE209998, GSE125989, and GSE43837 are mRNA-seq datasets containing both breast cancer (BC) tissues and breast cancer brain metastasis (BCBM) tissues. The analysis of these datasets focused on identifying the differences between metastatic and primary lesions. The GSE152818 dataset comprises scRNA-seq data from human BC cell lines metastasized to the mouse brain, serving as an animal model for BCBM. Specifically, dataset GSM4626851 represents proliferating cancer cells from brain metastases, while GSM4626852 corresponds to dormant cancer cells from brain metastases. This portion of the analysis examines the differences between proliferating and dormant cancer cells in BCBM.

By conducting bioinformatics analyses on these two datasets, we compared the differences in metabolic and functional pathways between primary BC cells and metastatic BC cells, as well as between proliferating and dormant cancer cells in BCBM. Additionally, we examined the differences in cell populations between proliferating and dormant cancer cells in brain metastases. By integrating macro-scale transcriptomics with scRNA-seq data, this study explored the potential biological behaviors of BC cells during brain metastasis at two different levels, with the aim of providing valuable 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.

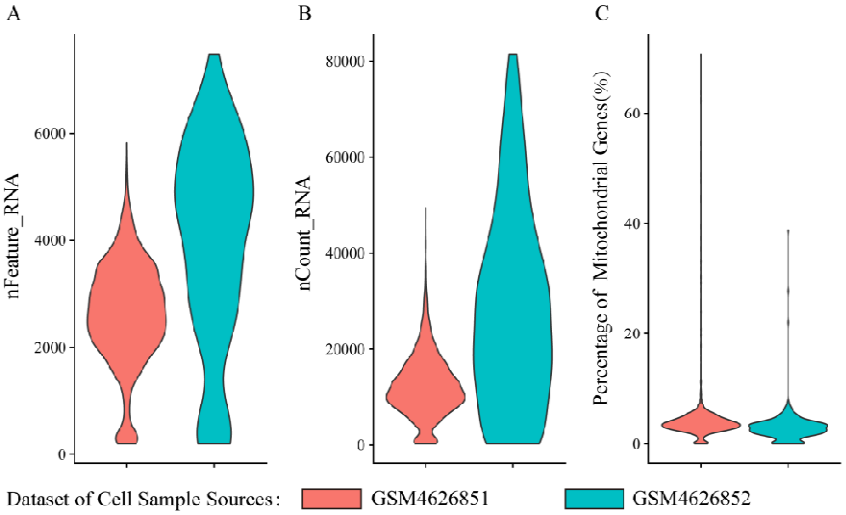


Figure 1. Plots depicting the number of genes detected in each cell (nFeature_RNA) (A), the number of unique RNA molecules detected in each cells (nCount-RNA) (B), and the percentage of reads mapping to the mitochondrial genome (percent.mt) (C).

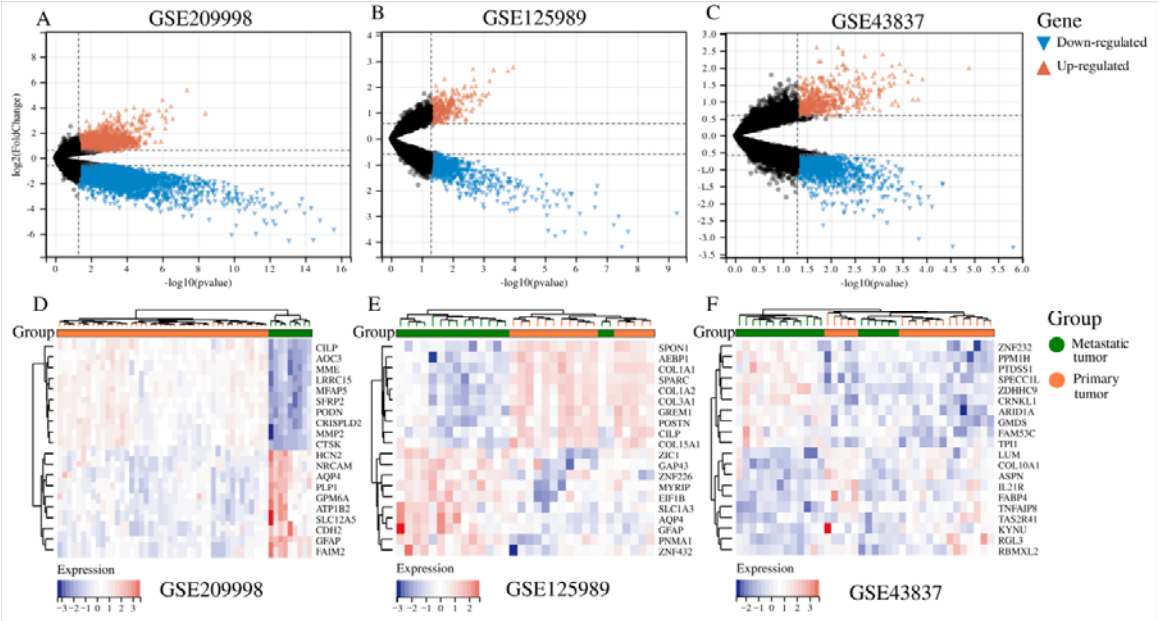


Figure 2. Identification of differentially expressed genes (DEGs) in three gene expression omnibus (GEO) datasets. (A-C) The volcano plots of differentially expressed genes in three datasets. Red represents up-regulated differential genes, black represents no significant difference genes, and blue represents 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.

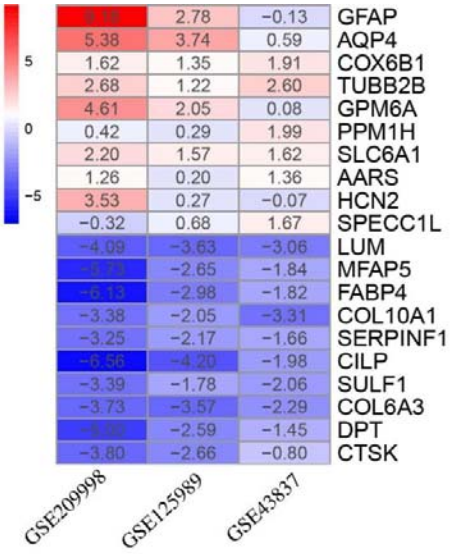


Figure 3. Heatmap shows the top 10 commonly upregulated and top 10 commonly downregulated genes in the RRA analysis of 3 bulk tissue mRNA-seq datasets. Red represents up-regulated genes, while blue represents down-regulated genes. The depth of the color represents the size of the log2 FC.

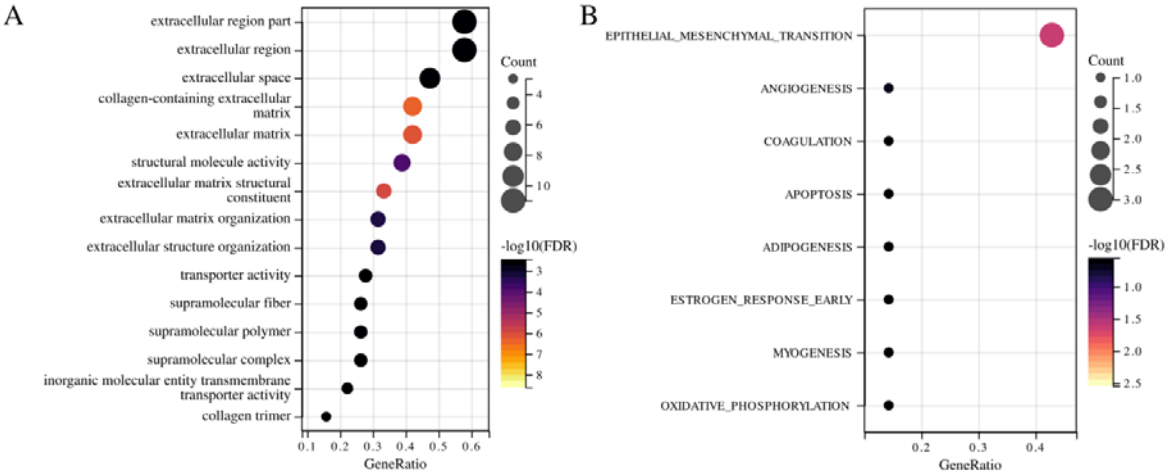


Figure4. GO and Hallmark pathway analysis of commonly upregulated/downregulated genes. (A) GO pathway analysis. (B) Hallmark pathway analysis.

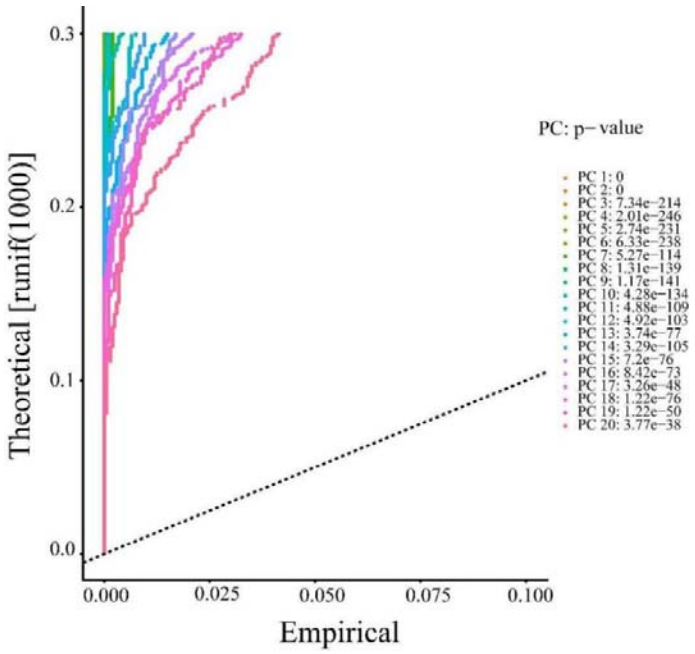


Figure5. Jackstraw plot shows the p-value distribution of each principal component (PC). Significant PCs were defined as those above the dashed line with statistically significant p-values (P < 0.05).

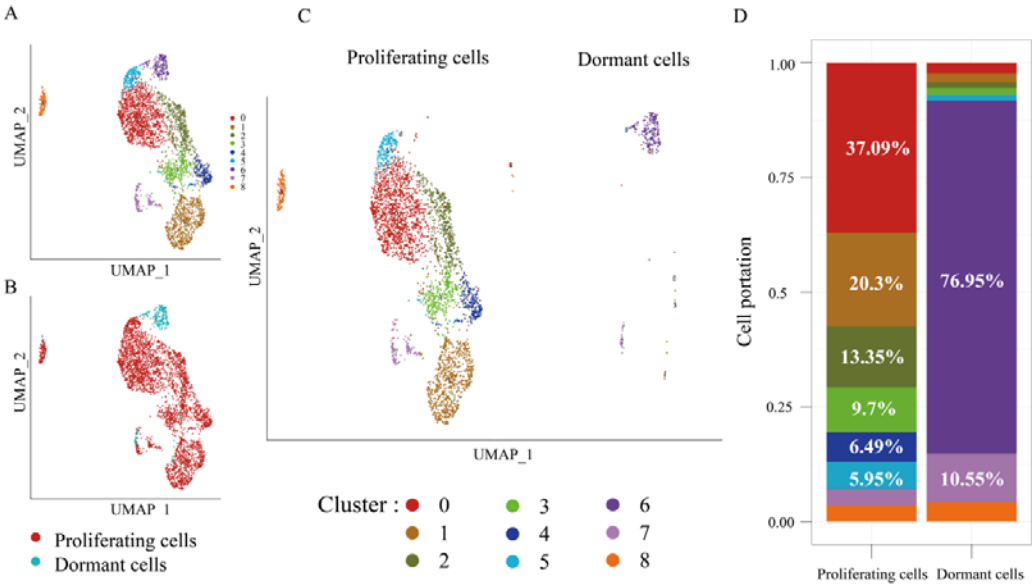


Figure6. scRNA-seq analysis reveals cell cluster heterogeneity in animal models of human BCBM. (A) Cells were clustered into 9 types by the UMAP dimensionality reduction algorithm, with each color representing a cluster. (B) Cells were labeled with two colors according to their source. Red represents 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.

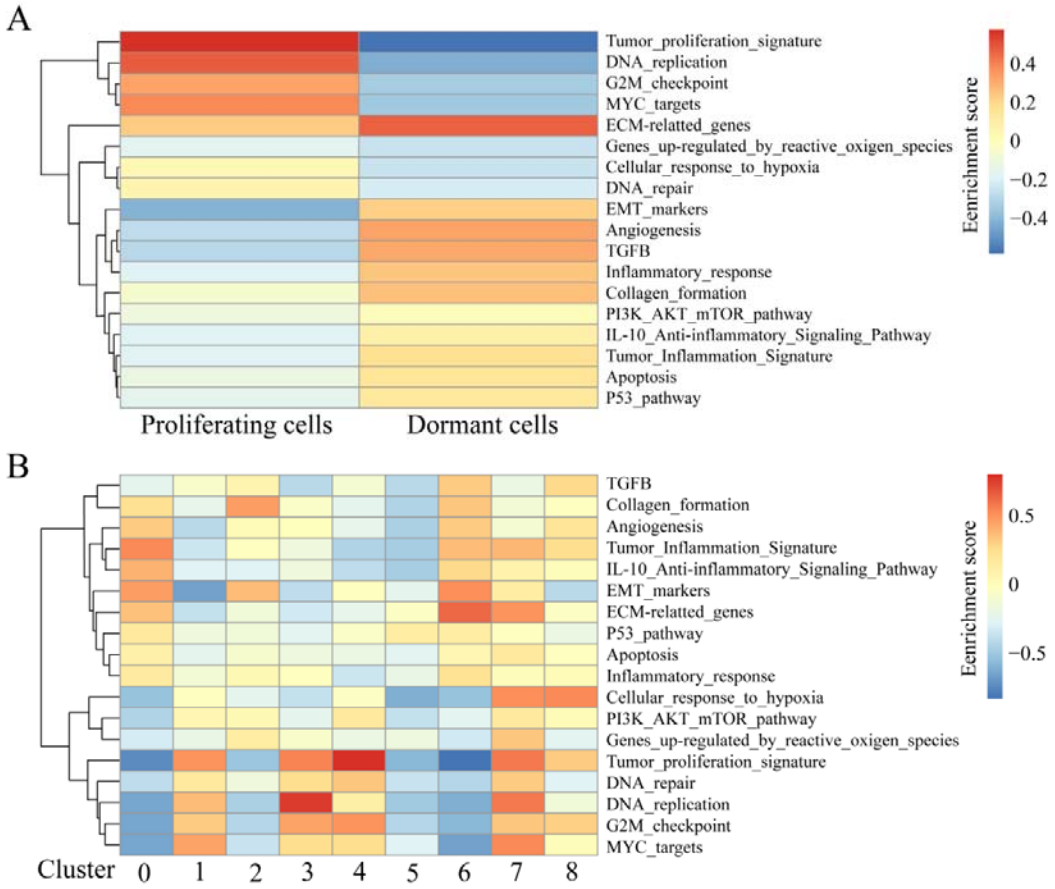


Figure7. The heatmap displays the results of GSVA enrichment analysis performed on the scRNA-seq data. Red color represents positive enrichment scores (pathway activation), while blue color represents negative enrichment scores (pathway inhibition). The depth of color represents the magnitude of the 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.

Supplementary Materials

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

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

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

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).

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

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

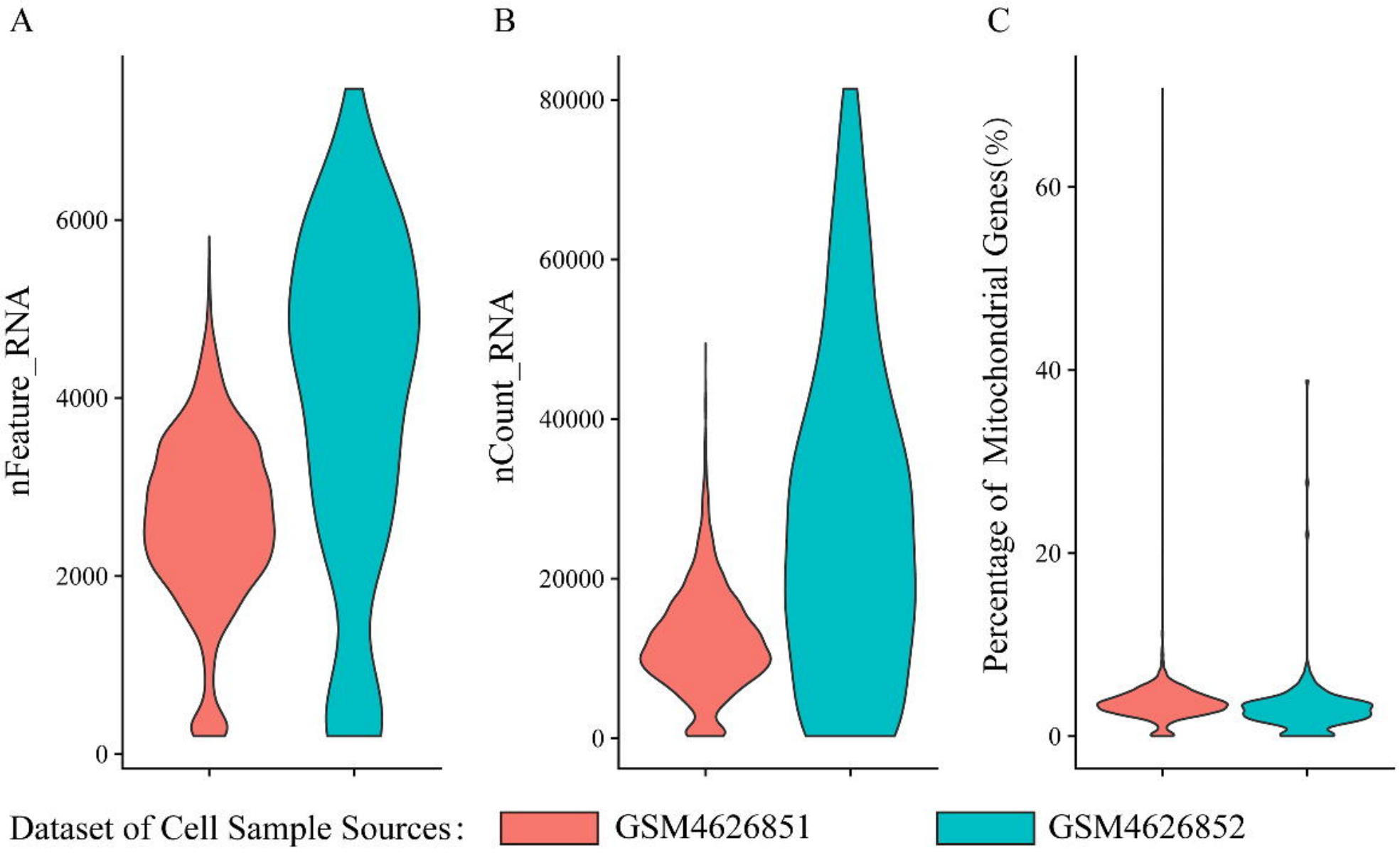
Author Contributions
ZY and ZYH wrote the manuscript, designed the study and interpreted the data. XW performed the statistical analysis and interpreted the results. XW and NL was involved in designing the study and critically revising the manuscript. All authors read and approved the final manuscript.

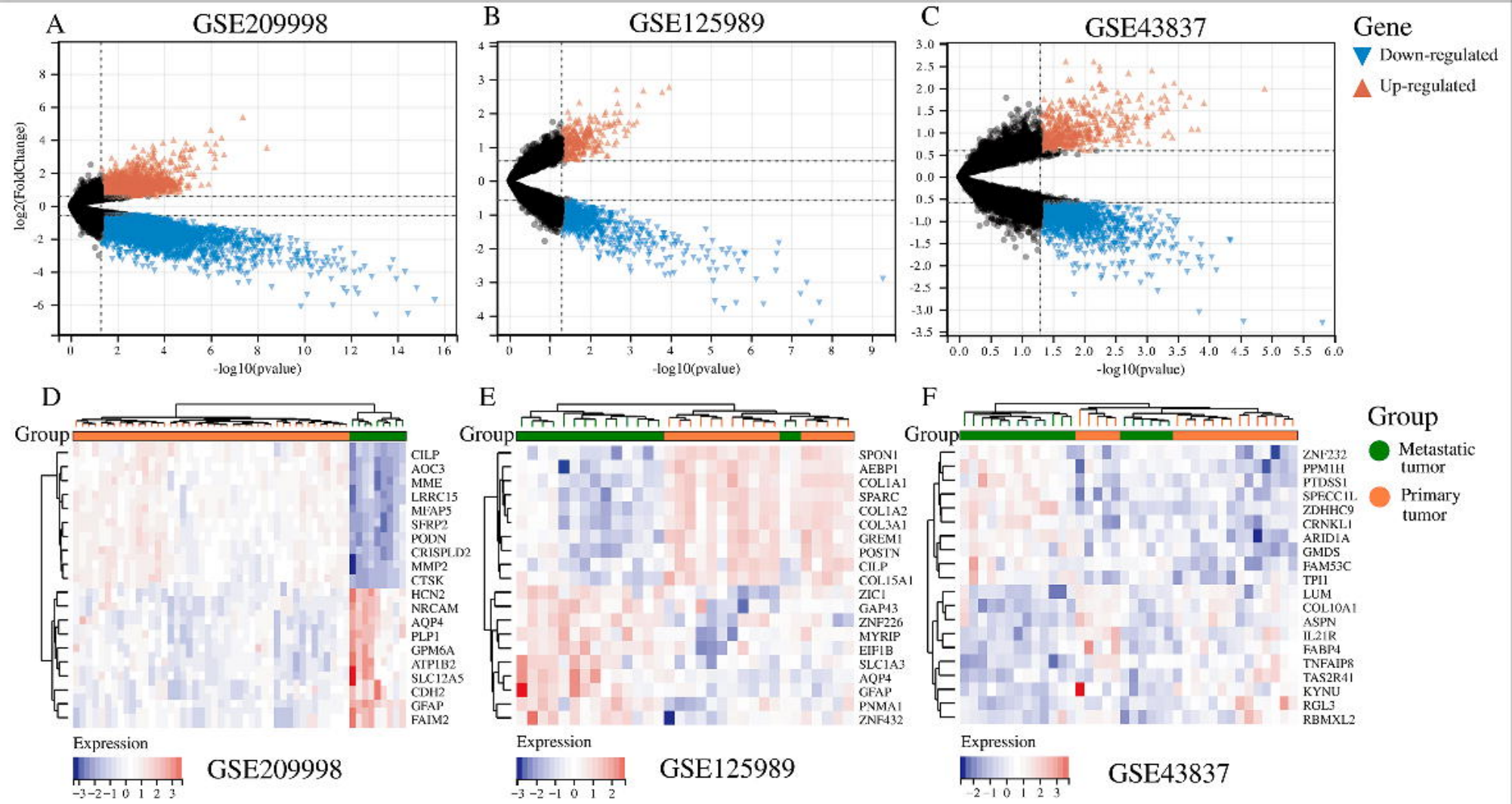
Funding
The study was supported by the Startup Fund for Scientific Research, Fujian Medical University (Grant number: 2020QH1115), the Natural Science Foundation of Fujian Province (No.2024J011341), and Xiamen Medical and Health Guidance Project in 2024 (3502Z20244ZD1215).

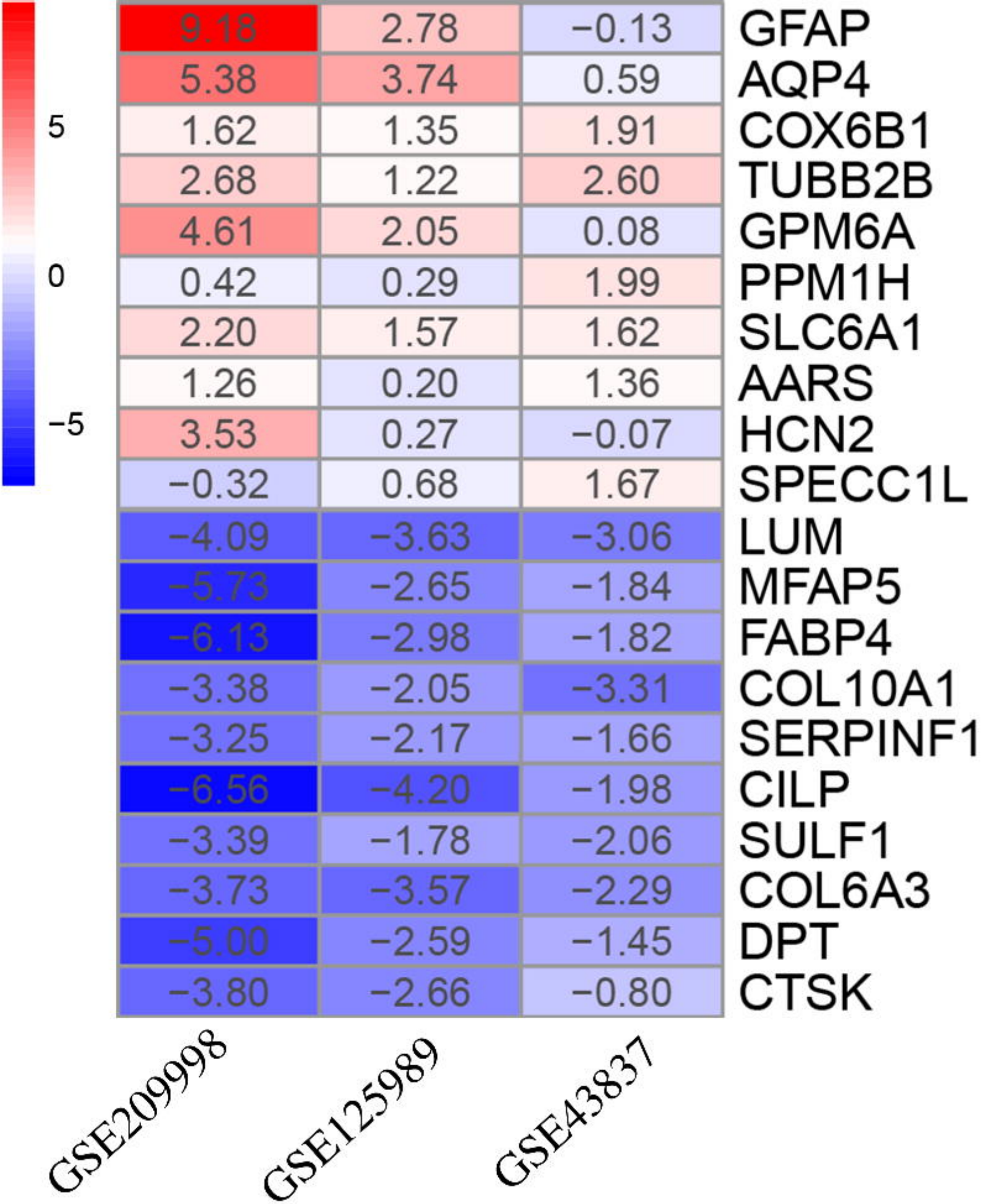
Conflicts of Interest
The authors declare no conflict of interest.

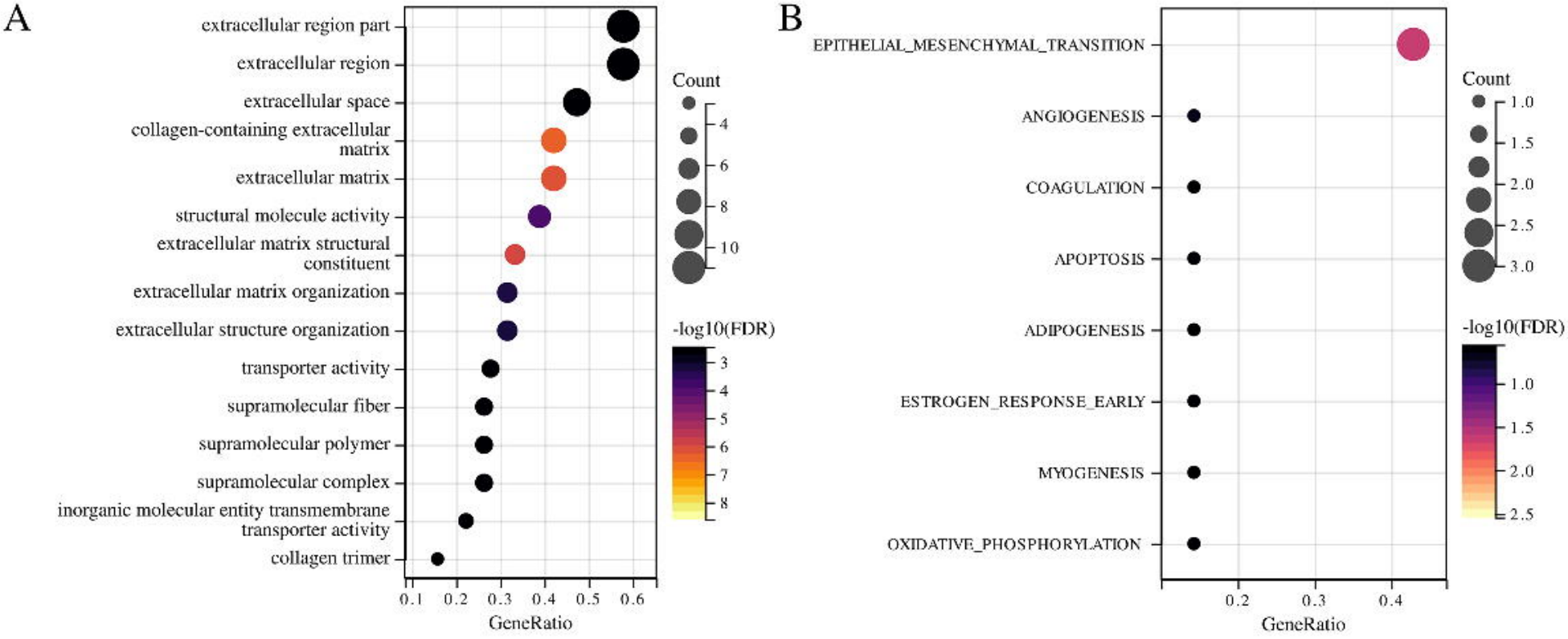
Ethics approval and consent to participate
Not applicable.

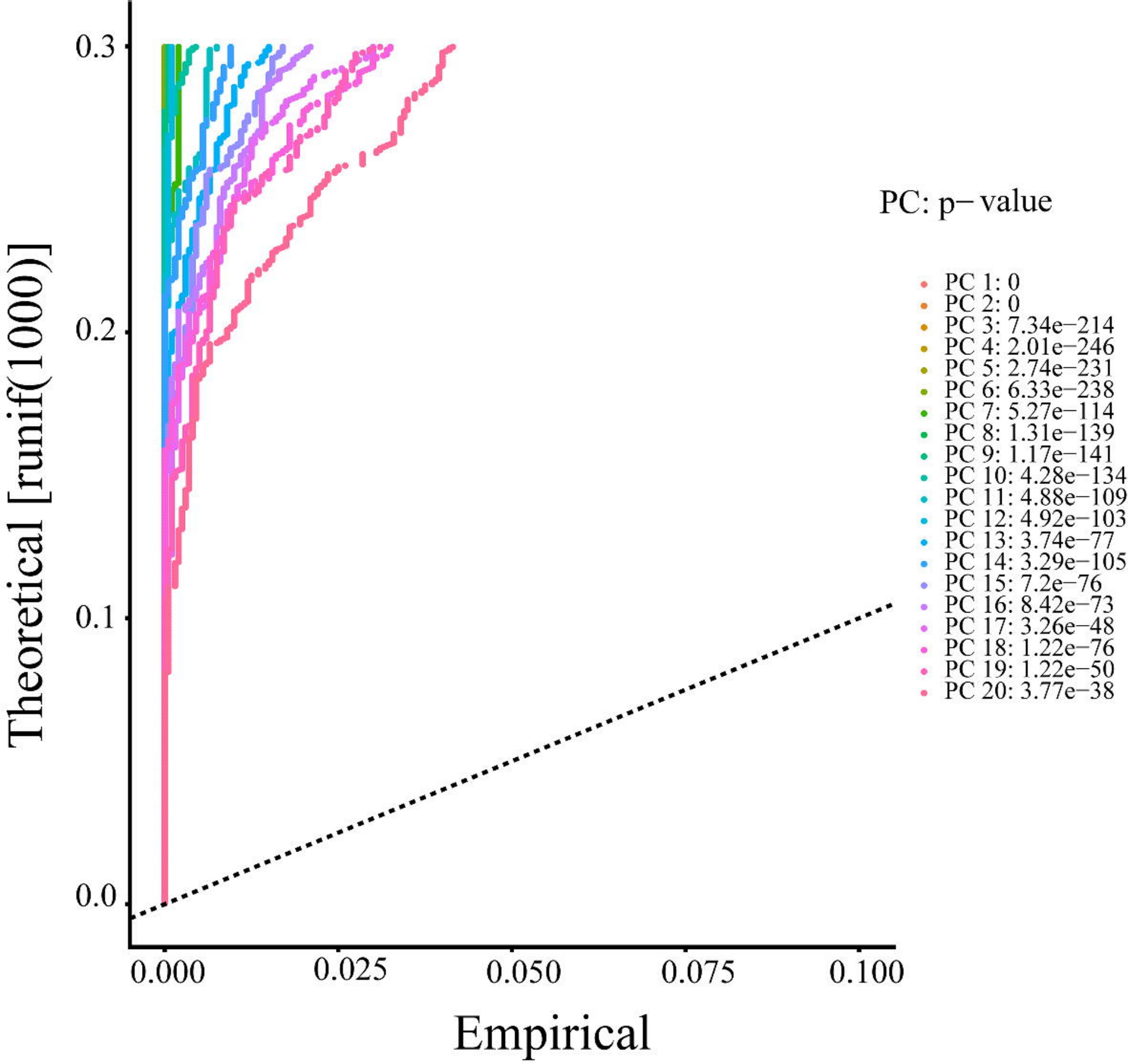
Availability of data and materials
The datasets analyzed during the current study are available in Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/).

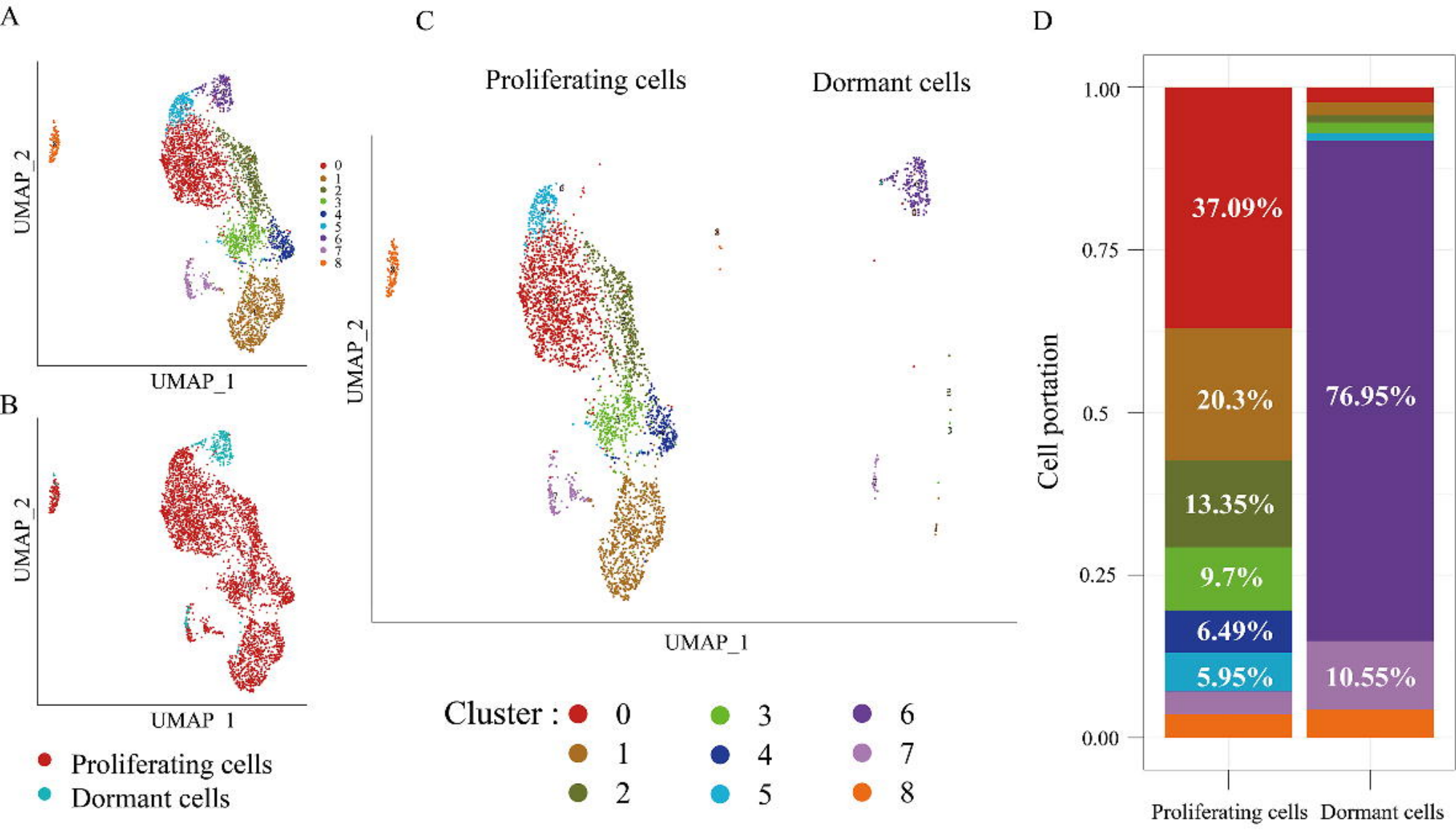

A
B
C
nFeature_RNA
6000
4000
2000
0
nCount_RNA
80000
60000
40000
20000
0
Percentage of Mitochondrial Genes(%)
60
40
20
0
Dataset of Cell Sample Sources:
GSM4626851
GSM4626852

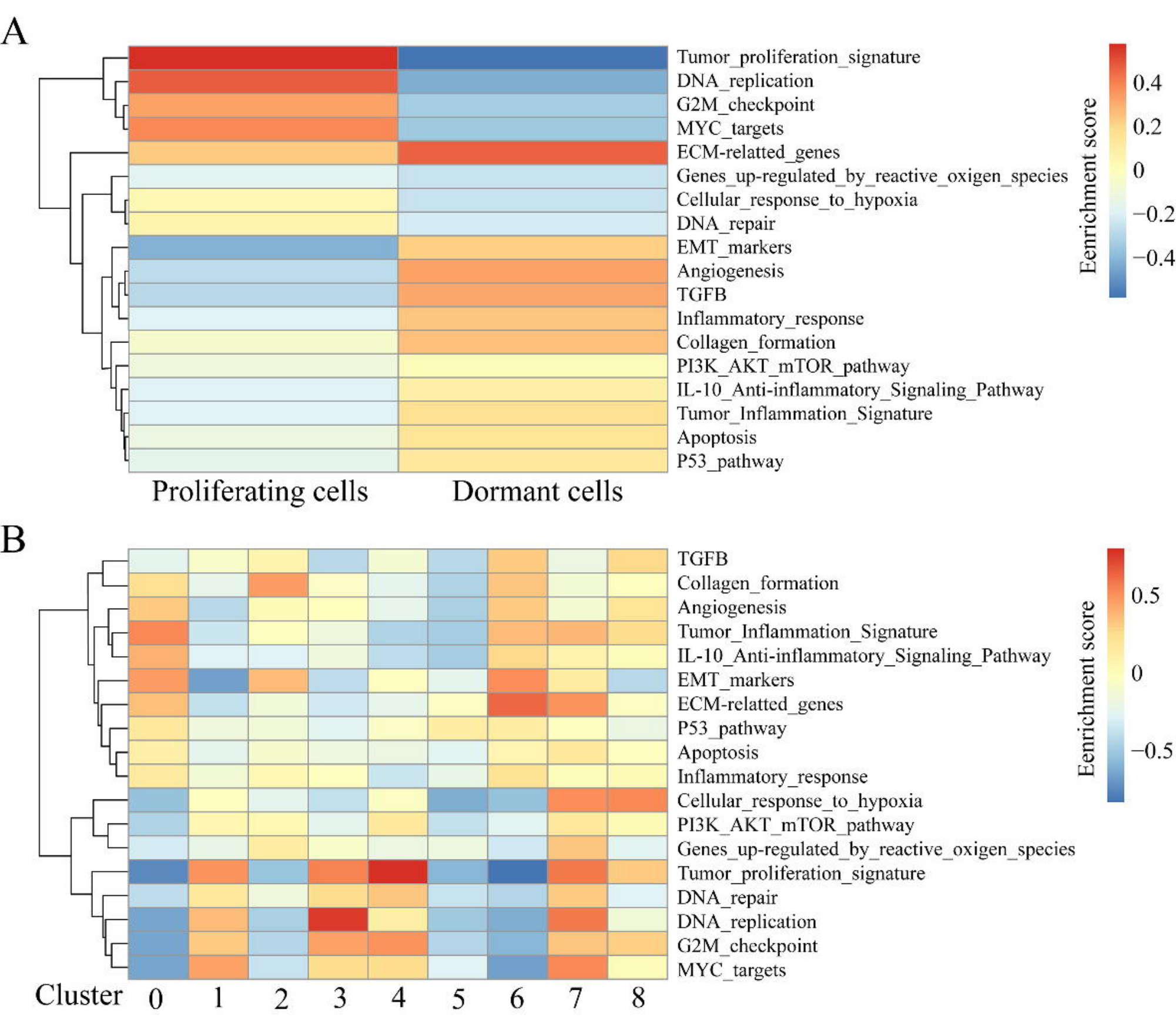

A
GSE209998
B
GSE125989
C
GSE43837
Gene
Down-regulated
Up-regulated
log2(FoldChange)
-log10(pvalue)
D
E
F
Group
Metastatic tumor
Primary tumor
CILP
AOC3
MME
LRRC15
MFAP5
SFRP2
PODN
CRISPLD2
MMP2
CTSK
HCN2
NRCAM
AQP4
PLP1
GPM6A
ATP1B2
SLC12A5
CDH2
GFAP
FAIM2
SPON1
AEBP1
COL1A1
SPARC
COL1A2
COL3A1
GREM1
POSTN
CILP
COL15A1
ZIC1
GAP43
ZNF226
MYRIP
EIF1B
SLC1A3
AQP4
GFAP
PNMA1
ZNF432
ZNF232
PPM1H
PTDSS1
SPECC1L
ZDHHC9
CRNKL1
ARID1A
GMDS
FAM53C
TPI1
LUM
COL10A1
ASPN
IL21R
FABP4
TNFAIP8
TAS2R41
KYNU
RGL3
RBMXL2
Expression
GSE209998
GSE125989
GSE43837


5
0
−5
9.18 2.78 −0.13 GFAP
5.38 3.74 0.59 AQP4
1.62 1.35 1.91 COX6B1
2.68 1.22 2.60 TUBB2B
4.61 2.05 0.08 GPM6A
0.42 0.29 1.99 PPM1H
2.20 1.57 1.62 SLC6A1
1.26 0.20 1.36 AARS
3.53 0.27 −0.07 HCN2
−0.32 0.68 1.67 SPECC1L
−4.09 −3.63 −3.06 LUM
−5.73 −2.65 −1.84 MFAP5
−6.13 −2.98 −1.82 FABP4
−3.38 −2.05 −3.31 COL10A1
−3.25 −2.17 −1.66 SERPINF1
−6.56 −4.20 −1.98 CILP
−3.39 −1.78 −2.06 SULF1
−3.73 −3.57 −2.29 COL6A3
−5.00 −2.59 −1.45 DPT
−3.80 −2.66 −0.80 CTSK
GSE209998
GSE125989
GSE43837


A
extracellular region part
extracellular region
extracellular space
collagen-containing extracellular matrix
extracellular matrix
structural molecule activity
extracellular matrix structural constituent
extracellular matrix organization
extracellular structure organization
transporter activity
supramolecular fiber
supramolecular polymer
supramolecular complex
inorganic molecular entity transmembrane transporter activity
collagen trimer
0.1 0.2 0.3 0.4 0.5 0.6
GeneRatio
Count
4 6 8 10
-log10(FDR)
3 4 5 6 7 8
B
EPITHELIAL_MESENCHYMAL_TRANSITION
ANGIOGENESIS
COAGULATION
APOPTOSIS
ADIPOGENESIS
ESTROGEN_RESPONSE_EARLY
MYOGENESIS
OXIDATIVE_PHOSPHORYLATION
0.2 0.3 0.4
GeneRatio
Count
1.0 1.5 2.0 2.5 3.0
-log10(FDR)
1.0 1.5 2.0 2.5


Theoretical [runif(1000)]
0.3
0.2
0.1
0.0
0.000
0.025
0.050
0.075
0.100
Empirical
PC: p− value
PC 1: 0
PC 2: 0
PC 3: 7.34e−214
PC 4: 2.01e−246
PC 5: 2.74e−231
PC 6: 6.33e−238
PC 7: 5.27e−114
PC 8: 1.31e−139
PC 9: 1.17e−141
PC 10: 4.28e−134
PC 11: 4.88e−109
PC 12: 4.92e−103
PC 13: 3.74e−77
PC 14: 3.29e−105
PC 15: 7.2e−76
PC 16: 8.42e−73
PC 17: 3.26e−48
PC 18: 1.22e−76
PC 19: 1.22e−50
PC 20: 3.77e−38


A
UMAP_2
UMAP_1
0
1
2
3
4
5
6
7
8
B
UMAP_2
UMAP_1
Proliferating cells
Dormant cells
C
Proliferating cells
Dormant cells
UMAP_2
UMAP_1
Cluster :
0
1
2
3
4
5
6
7
8
D
Cell portation
1.00
0.75
0.5
0.25
0.00
37.09%
20.3%
13.35%
9.7%
6.49%
5.95%
76.95%
10.55%
Proliferating cells
Dormant cells


A
Tumor_proliferation_signature
DNA_replication
G2M_checkpoint
MYC_targets
ECM-relatted_genes
Genes_up-regulated_by_reactive_oxigen_species
Cellular_response_to_hypoxia
DNA_repair
EMT_markers
Angiogenesis
TGFB
Inflammatory_response
Collagen_formation
PI3K_AKT_mTOR_pathway
IL-10_Anti-inflammatory_Signaling_Pathway
Tumor_Inflammation_Signature
Apoptosis
P53_pathway
Proliferating cells
Dormant cells
Eenrichment score
0.4
0.2
0
−0.2
−0.4
B
TGFB
Collagen_formation
Angiogenesis
Tumor_Inflammation_Signature
IL-10_Anti-inflammatory_Signaling_Pathway
EMT_markers
ECM-relatted_genes
P53_pathway
Apoptosis
Inflammatory_response
Cellular_response_to_hypoxia
PI3K_AKT_mTOR_pathway
Genes_up-regulated_by_reactive_oxigen_species
Tumor_proliferation_signature
DNA_repair
DNA_replication
G2M_checkpoint
MYC_targets
Cluster
0
1
2
3
4
5
6
7
8
Eenrichment score
0.5
0
−0.5