Identification of gene expression patterns from 7.5k sorted cells from a human invasive ductal carcinoma dataset

Author: Victoria Palecek

Abstract:

Invasive ductal carcinoma, or IDC, is an aggressive type of breast cancer. It begins as ductal carcinoma in-situ, DCIS, within the milk ducts and then progresses to IDC by invading surrounding breast tissue. The exact mechanism of invasion is not yet known. The aim of this study was to identify possible drivers for disease progression from DCIS to IDC and possible routes to further understand subclone formation. This was accomplished by performing scRNA sequencing on sorted IDC cells, performing cluster analysis and identifying differentially expressed genes within the identified clusters. In total, 11 clusters were identified and subsequently compared using p-values and log2 fold changes. Clusters 5-10 showed well known poor prognostic indicators for breast cancer. Cluster 1 showed a decrease in expression of several key mitochondrial genes. Cluster 3 showed a decrease in expression in several ribosomal genes, including one that has been identified as a poor prognostic indicator. Clusters 0, 2 and 4 contained the most cells and had high differential expression of many genes that have yet to be functionally annotated. All these distinct patterns are indicative of possible sub-clones. Any of the genetic abnormalities identified in each of the clusters could contribute to disease progression. It is worth looking into the genes that have not yet been functionally annotated to identify if they may be contributing to IDC cell invasion. Further studies should focus on functional characterization of the unannotated genes and on exploring therapeutic approaches for mitochondrial and ribosomal gene dysregulation.

Introduction:

Invasive ductal carcinoma is a type of breast cancer that begins in the milk ducts. Unlike ductal carcinoma in-situ, invasive ductal carcinoma does not stay confined to the milk ducts. It spreads to the surrounding breast tissue and can metastasize to other parts of the body if left untreated ("Invasive Ductal Carcinoma (IDC)", n.d.). This type of cancer typically metastasizes to the lung, bone, liver or brain (Borst and Ingold, 1993). It is the most common type of breast cancer. IDC typically presents in a dense tumor, usually resulting in quick diagnosis. Before invasion into the surrounding breast tissue from the ducts IDC is considered ductal carcinoma in-situ (DCIS).
The progression of IDC is explained by a grading system that includes three grades. At grade 1, the IDC cells maintain the same morphology of healthy breast cells and are growing slowly. At grade 2 they are beginning to take on morphological differences and begin to grow a slightly faster rate than grade one. Once the IDC cells reach grade 3, they are very morphologically distinct in comparison to grade 1cells and are growing/spreading much faster than grade 1. The use of cytological grading allows for a quick, visual way to continuously monitor the effectiveness of treatment and the relative progression of the disease (Bansal et al.,2014).
IDC tumor cells are prognostically characterized by the quantity of HER2 (HER2 positive), the presence of estrogen receptors (ER positive), or the presence of progesterone receptors (PR positive). Tumors that have all three are considered triple positive and tumors without any of the

three abnormalities are considered triple negative (Adedayo et al,2009). These abnormalities are well characterized and studied. However, these three abnormalities cannot always be blamed for the invasion of IDC cells into surrounding breast tissue. There are many cases where DCIS tumors contain any variation of the HER2+, ER+ or PR+ abnormalities but never progress to IDC. In fact, it is not well understood why DCIS becomes IDC. Although factors such as tumor size, nuclear grade and Ki67 index are closely linked with IDC diagnosis, the why is still yet to be determined. Single cell RNA (scRNA) studies have been profoundly valuable in identifying answers to these questions. This study will use scRNA data from sorted IDC cells collected from a 65-year-old female patient. In this study we are trying to determine if there are gene expression patterns that may contribute to the invasive or metastatic nature of this type of tumor in this specific case.

Methods:

Invasive Ductal Carcinoma (IDC) cells were obtained from a 65-year-old female donor through 10x Genomics from Discovery Life Sciences. Cells were sequenced on the Illumina NovaSeq 6000 platform to a depth of 54,504 mean reads per cell. Sequencing was performed as paired-end, dual indexing with a 16bp barcode. Data was then processed using Cell Ranger which generated an H5 output file.

The H5 file was imported into R, and a Seurat object was created (Hao et al.,2024). Across the dataset, 22,914 features were identified across 5,621 cells. The mitochondrial RNA percentage per cell was calculated using PercentageFeatureSet. Initial quality control involved generating violin plots for the number of features, total RNA counts, and mitochondrial percentage in each cell. Scatter plots of RNA count vs. mitochondrial percentage and RNA count vs. feature count were also generated to help define filtering thresholds.

Because higher mitochondrial RNA percentages are often observed in aggressive cancers, the dataset was filtered to retain cells with ≤20% mitochondrial RNA and between 200 and 8,500 detected features (figure 1). Post-filtering visualizations confirmed that the data quality was sufficient for further analysis. Normalization was performed using the LogNormalize method with a scale factor of 10,000. Variable feature selection was used to identify the top 2,000 features using the vst method, from which the top 10 were labeled for reference. This step helped reduce noise in downstream analysis by removing low-variance genes.

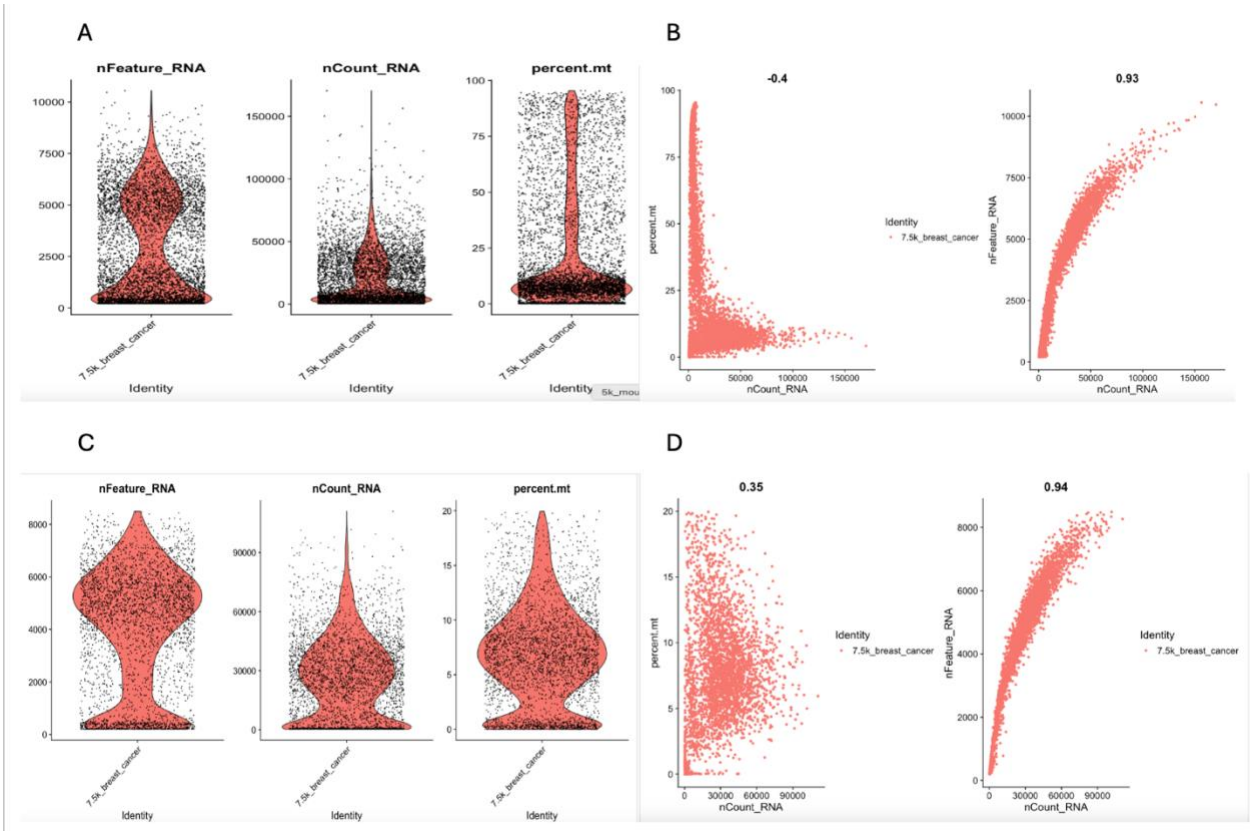

Figure 1: Quality control and filtering of invasive ductal carcinoma single-cell RNA-seq data. Violin plots display the distribution of detected features, total RNA counts, and mitochondrial RNA percentage across cells. Scatter plots of RNA count versus mitochondrial percentage and RNA count versus feature count illustrate thresholds used for filtering.

Data scaling was performed using ScaleData, followed by principal component analysis (RunPCA) on the top 2,000 variable features. The top 30 loadings from the first 20 principal components (PCs) were inspected to ensure that none were strongly driven by mitochondrial genes (figure 2). Heatmaps for PCs 1–20 were generated to identify patterns for downstream analysis, and based on these results, PCs 1–15 were selected (figure 3). This decision was supported by an elbow plot (figure 4).

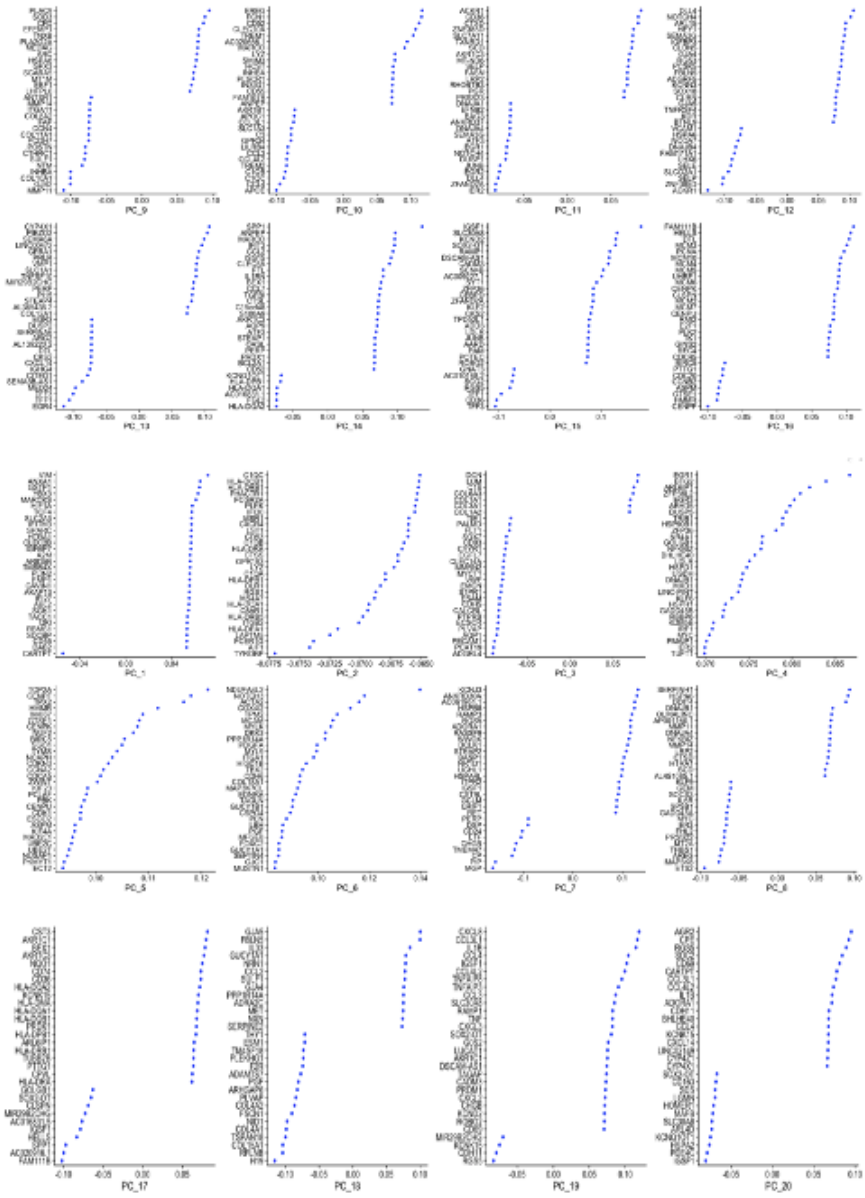

Figure 2: Loadings of the top 30 genes for the first 20 principal components. This figure visualizes the contribution of individual genes to each principal component, providing an overview of the major sources of variation in the single-cell RNA-seq dataset. The analysis confirms that no single group of genes, such as mitochondrial genes, disproportionately influences the principal components.

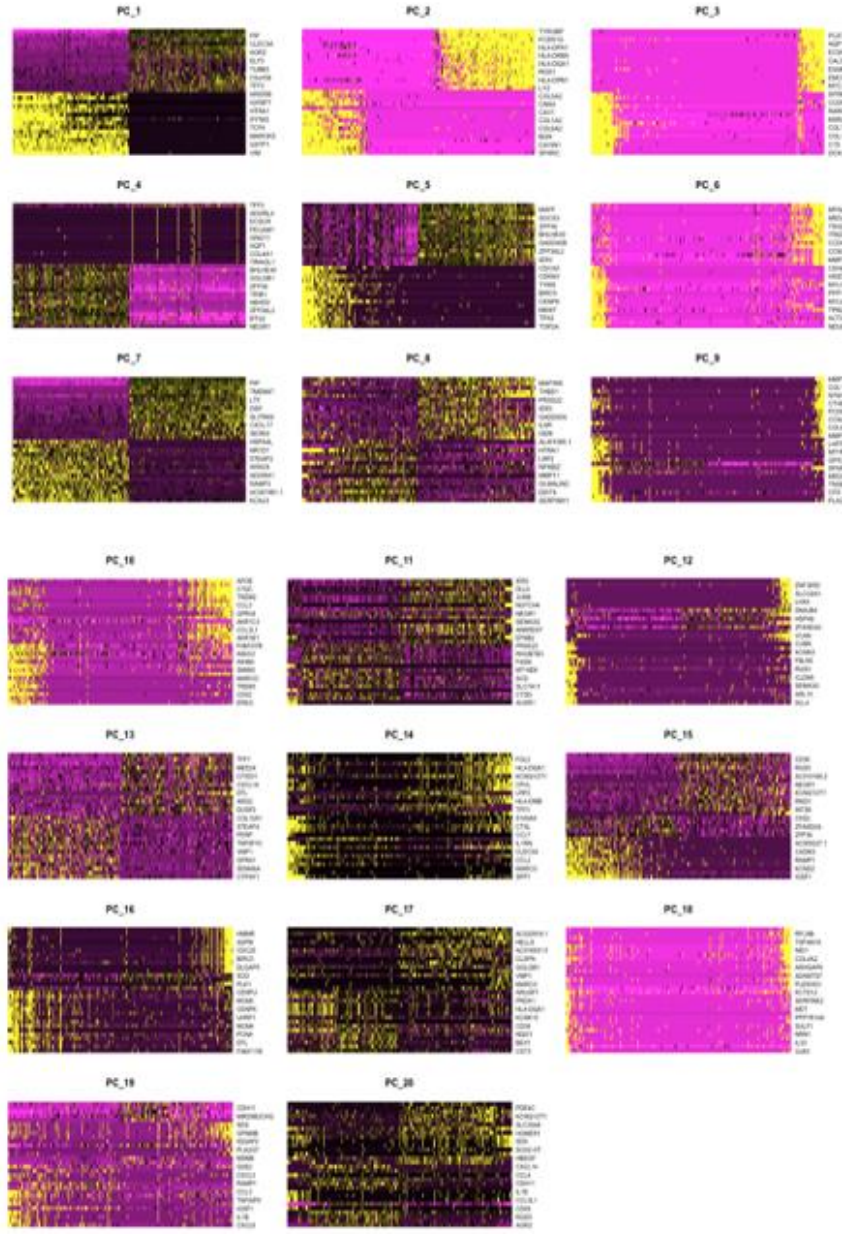

Figure 3: Heatmaps of principal components 1–20. These heatmaps display patterns of gene expression contributing to each principal component, allowing visualization of major sources of variation across the single-cell RNA-seq dataset. The observed patterns helped guide the selection of the first 15 PCs for downstream clustering and dimensionality reduction analyses.

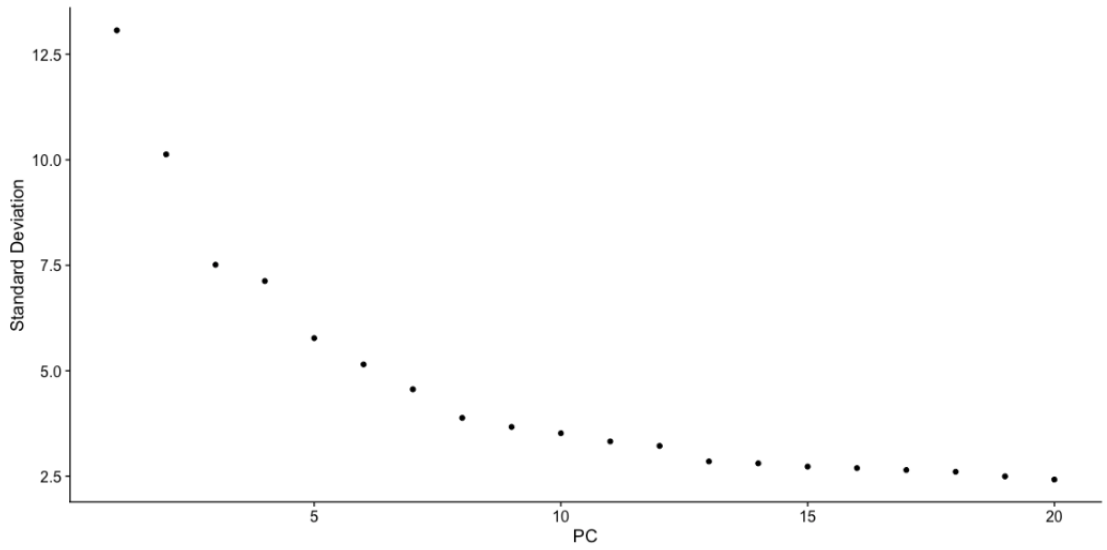

Figure 4: Elbow plot used to determine the optimal number of principal components for downstream analysis. The plot displays the standard deviation explained by each principal component, with a noticeable bend around PC 15, supporting the selection of the first 15 PCs for clustering and dimensionality reduction.

Cells were clustered using a Jaccard similarity distance matrix to identify nearest neighbors based on the first 15 PCs. The Louvain algorithm (FindClusters) was applied with a resolution of 0.7. Both UMAP (figure 5) and t-SNE (figure 6) were run for non-linear dimensionality reduction, but UMAP was chosen for further visualization of the 11 identified clusters.

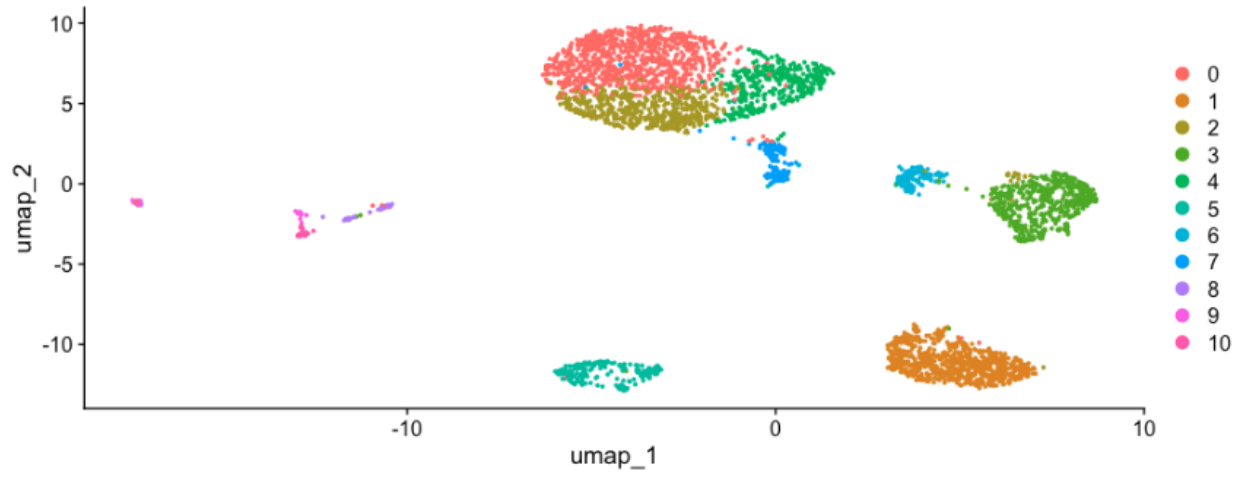

Figure 5: UMAP visualization of 7,500 single invasive ductal carcinoma cells clustered into 11 distinct groups. Each point represents a single cell, and colors correspond to identified clusters based on shared gene expression profiles.

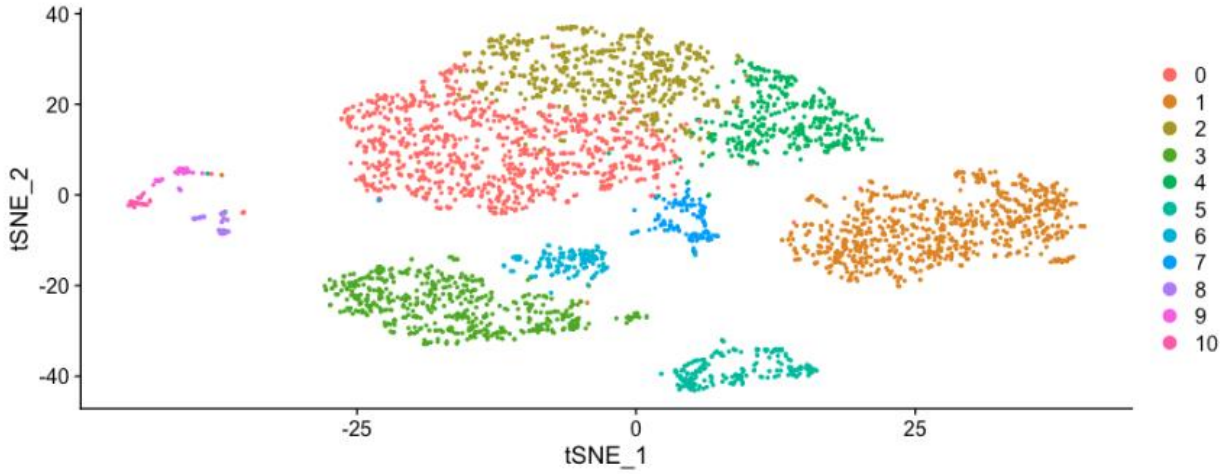

Figure 6: t-SNE visualization of 7,500 single invasive ductal carcinoma cells clustered into 11 groups. Each point represents an individual cell, with colors indicating cluster membership. The plot provides an alternative non-linear dimensionality reduction to UMAP.

Marker genes for each cluster were identified using FindAllMarkers with both positive and negative markers reported. The top 30 most significant markers per cluster were annotated using UniProt, BioMart, and published studies. Heatmaps, FeaturePlots, and violin plots were then generated to visualize gene expression patterns and validate marker gene relevance.

Results:

A total of 11 clusters were identified during analysis. Clusters 0, 2 and 4 were the largest in size and were close enough in proximity to make up one large, main cluster with similar differential gene expressions. The smallest of the clusters, 8,9 and 10, were also very close in proximity and had very similar expressions of known breast cancer prognostic indicators. 8, 9 and 10 were analyzed together due to their size. Cluster 1 was the largest of the well-defined clusters, while clusters 5, 6 and 7 were approximately the same size as 8, 9 and 10 combined.

Clusters 8, 9 and 10 showed the highest expression of many of the top 10 variably expressed genes. These genes consisted of SPARC ($p < 2.2e-308$, Log2FC = 10.3), DCN ($p = 9.9e-230$, Log2FC = 11.4), LUM ($p = 7.2e-259$, Log2FC = 11.0), COL1A1 ($p = 1.4e-171$, Log2FC =9.0), COL1A2 ($p < 2.2e-308$, Log2FC = 9.8) and COL3A1($p < 2.2e-308$, Log2FC = 10.0) and all play a large role in remodeling of the extracellular matrix or ECM. IGFBP7($p < 2.2e-308$, Log2FC = 10.0) was also highly expressed across clusters 8, 9 and 10 and is linked to disease progression in breast cancer (figure 11). In addition, LUM and DCN, when increased expression of both is observed, is typically associated with cancer associated fibroblast (CAFs).

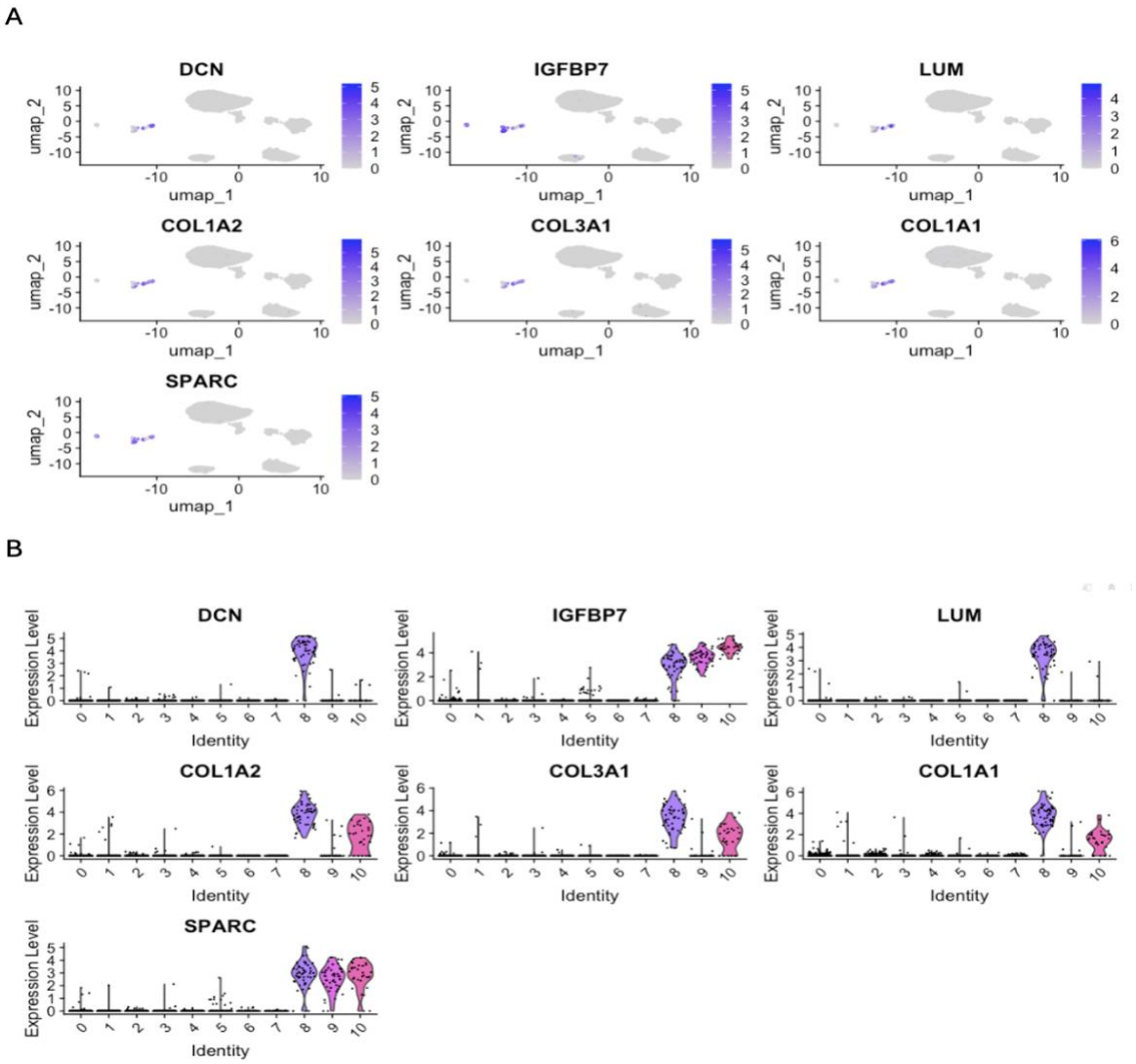

Figure 11: (A) Feature plots showing the spatial expression of DCN, IGFBP7, LUM, COL1A2, COL3A1, COL1A1, and SPARC across the UMAP embedding of cells. High expression is indicated by deeper shades of blue. (B) Violin plots depicting the distribution of expression levels for the same genes across all clusters.

In cluster 3 decreased expression of several ribosomal protein genes, including RPS9(p = 3.6e-218, Log2FC =-1.3), was observed (figure 12). The down regulation of this specific ribosomal

protein has been linked to poor prognostic outcomes in patients with breast cancer.

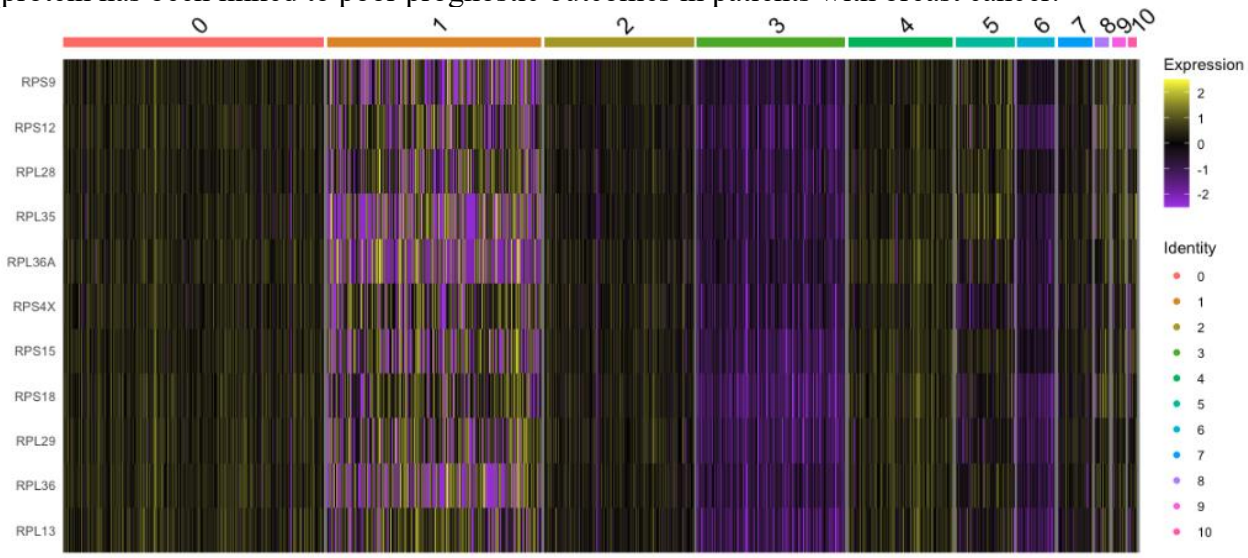

Figure 12: Heatmap of ribosomal protein gene expression across all clusters, highlighting cluster 3. Genes included are RPS9, RPS12, RPL28, RPL35, RPL36A, RPS4X, RPS15, RPS18, RPL29, RPL36, and RPL13. Each row represents a gene and each column represents a single cell grouped by cluster. The color gradient from purple to yellow indicates low to high expression levels, illustrating the decreased ribosomal gene expression characteristic of cluster 3.

Cluster 5 observed significantly increased expression of the remaining top 10 genes as well as additional poor prognostic biomarkers. Increased expression of APOE ($p = 1.3e-144$, Log2FC = 7.6), CCL3($p < 2.2e-308$, Log2FC = 9.4 and TYROBP ($p < 2.2e-308$, Log2FC = 9.6) (figure 13) was observed. TYROBP is a protein tyrosine binding protein and when highly expressed is associated with a poor prognosis in breast cancer and other malignancies. APOE is a lipoprotein that mediates lipid transport between organs and is has been linked to breast cancer metastasis when highly expressed. CCL3 is an inflammatory chemokine that can inhibit apoptosis and aid in invasion to other tissues when expression is increased.

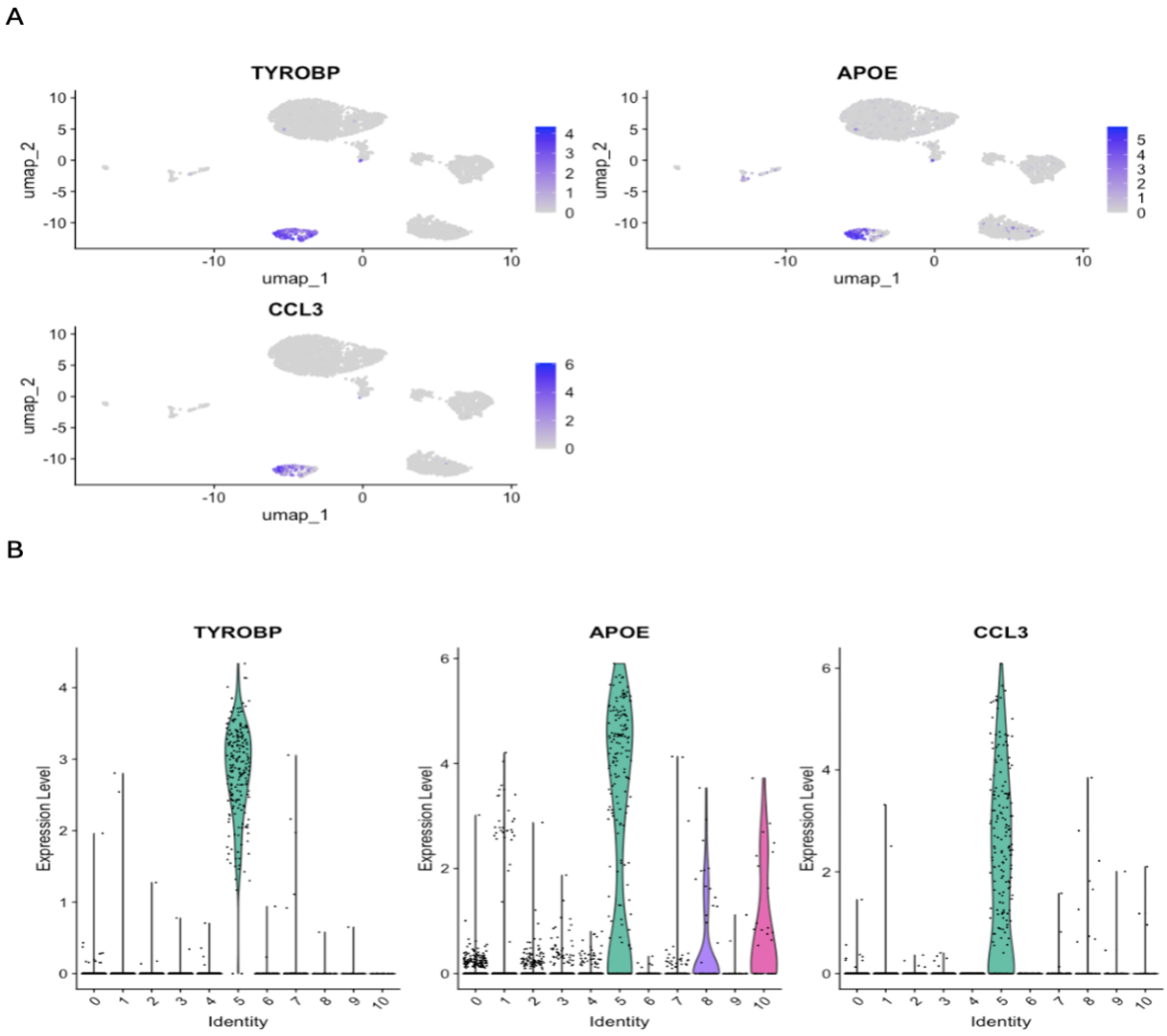

Figure 13: (A) Feature plots showing the spatial distribution of TYROBP, APOE, and CCL3 across UMAP clustering. High expression levels are indicated by deeper shades of blue, highlighting cluster 5-specific upregulation. (B) Violin plots displaying the distribution of TYROBP, APOE, and CCL3 expression within cluster 5 relative to other clusters.

Cluster 6 is a small cluster relative to 0, 1, and 2. This cluster of cells had an increased expression of DSCAM-AS1($p = 2.3e-308$, $Log2FC = 6.8$), IGSF1($p = 1.3e-129$, $Log2FC = 6.1$) and SOX2-OT($p = 7.6e-205$, $Log2FC = 5.3$) (figure 14). These genes are linked to chemotherapy resistance and disease relapse when increased expression is observed.

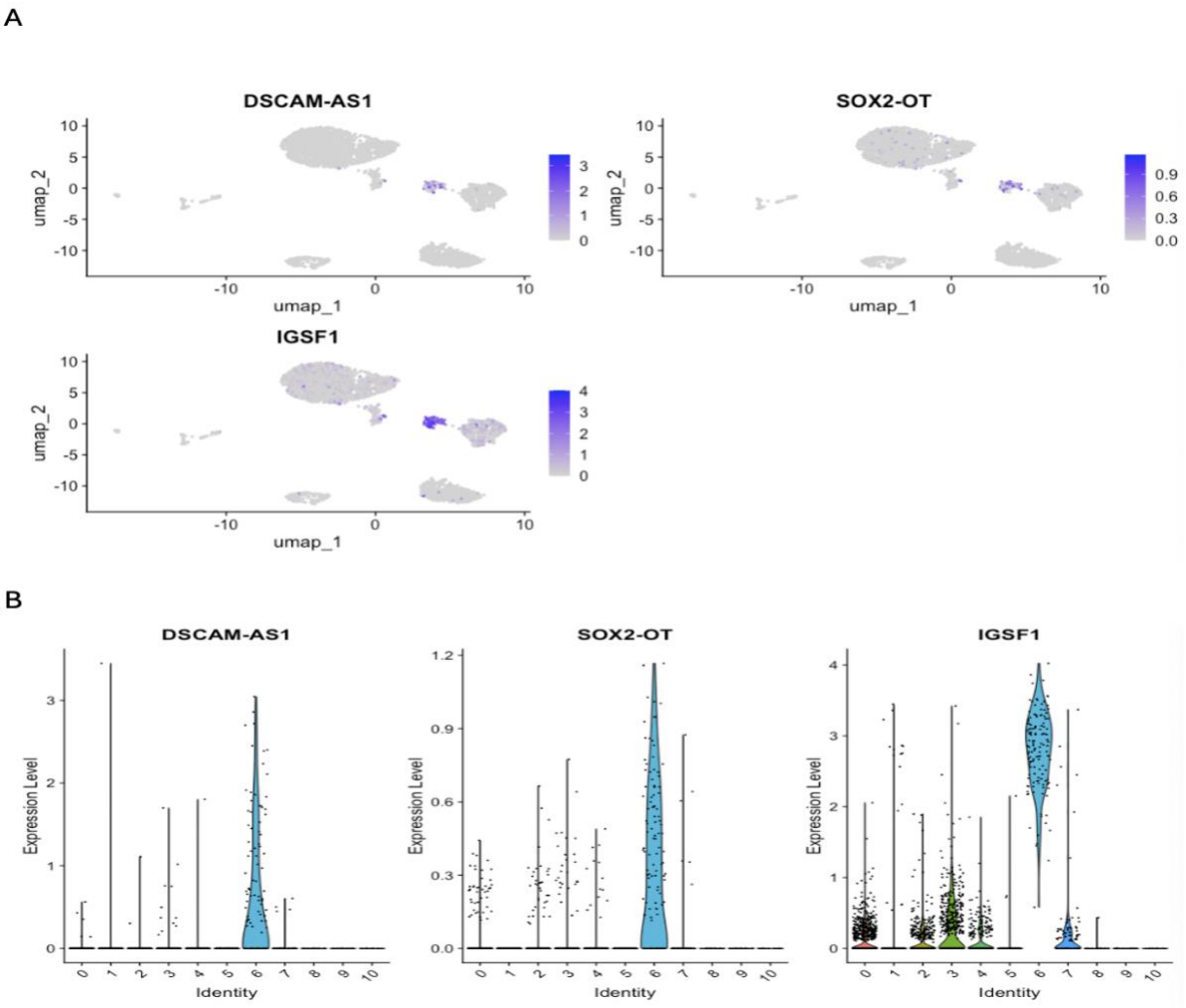

Figure 14: (A) Feature plot illustrating the expression patterns of DSCAM-AS1, SOX2-OT, and IGSF1 across the UMAP clustering. Elevated expression in cluster 6 is indicated by deeper shades of blue. (B) Violin plots showing the distribution of DSCAM-AS1, SOX2-OT, and IGSF1 expression within cluster 6 relative to other clusters.

Cluster 7 contains approximately the same quantity of cells as cluster 6 and saw a very significant increase in the expression of several poor prognostic markers including EXO1($p = 1.2e-192$, $Log2FC = 6.4$), NCAPH($p < 2.2e-308$, $Log2FC = 6.7$), PKMYT1($p = 7.4e-291$, $Log2FC = 5.8$) and CENPF($p = 7.6e-267$, $Log2FC = 6.0$) (figure 15). All three of these genes are linked to cancer cell proliferation.

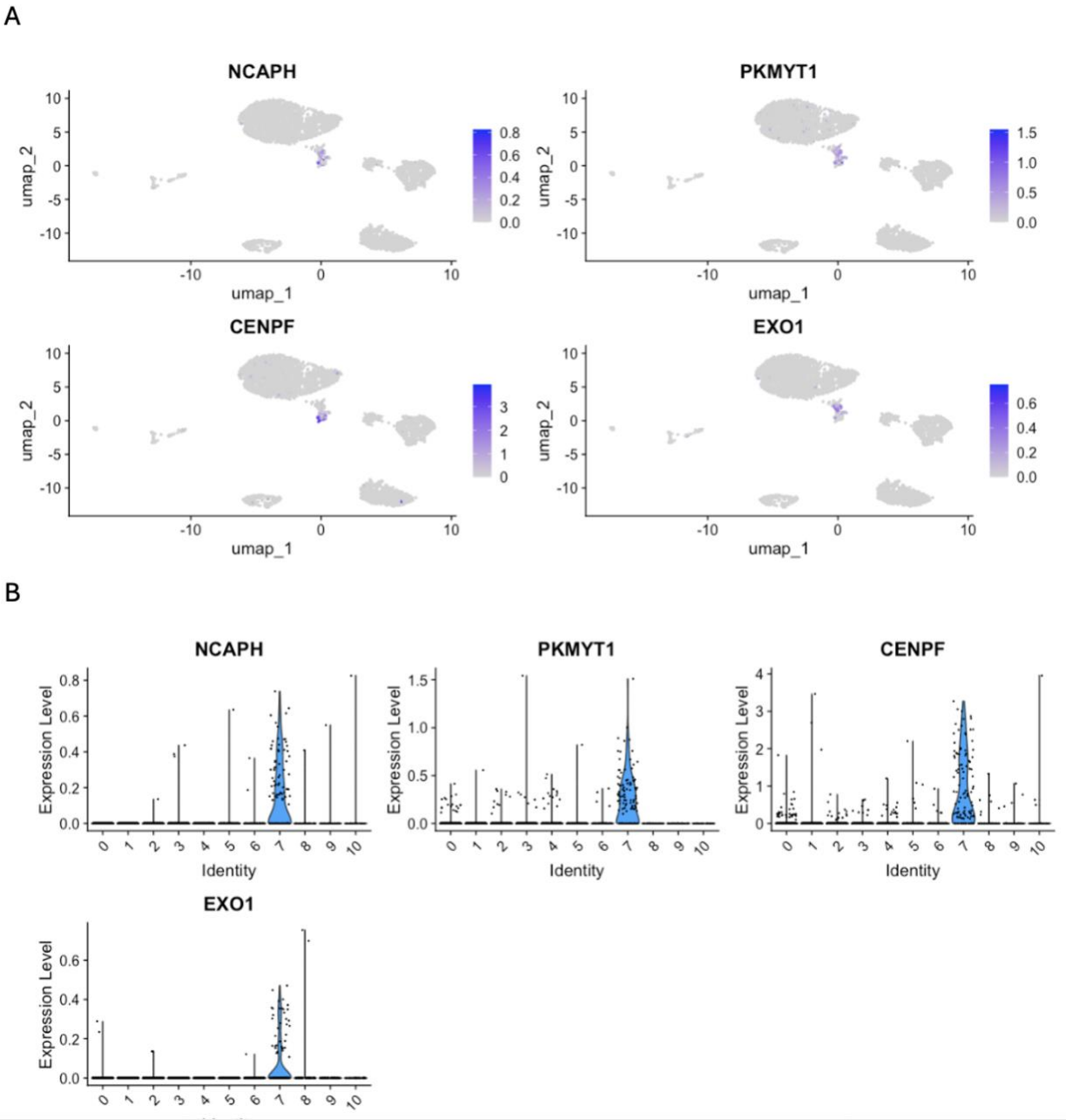

Figure 15: (A) Feature plots showing the UMAP distribution of NCAPH, PKMYT1, CENPF, and EXO1. Elevated expression in cluster 7 is indicated by deeper shades of blue. (B) Violin plots displaying the expression levels of NCAPH, PKMYT1, CENPF, and EXO1 within cluster 7 relative to other clusters.

Decreased expression of several critical genes was observed in cluster 1. Cluster 1 had a much lower expression of EGR1(p = 5.5e-147, Log2FC = -1.4), TNFSF10(p = 3.5e-264, Log2FC = -1.6) and several key mitochondrial genes (MT-CO2(p = 5.1e-295, Log2FC = -2.9), MT-CYB(p = 7.1e-293, Log2FC = -2.8), MT-ND4(p = 7.5e-289, Log2FC = -2.9)) (figure 16). EGR1 and TNFSF10 are known tumor suppressors. Their expression was also reduced in cluster 5 but remained mostly normal in all other clusters. The reduced expression or mutation of mitochondrial genes has been linked to poor prognosis.

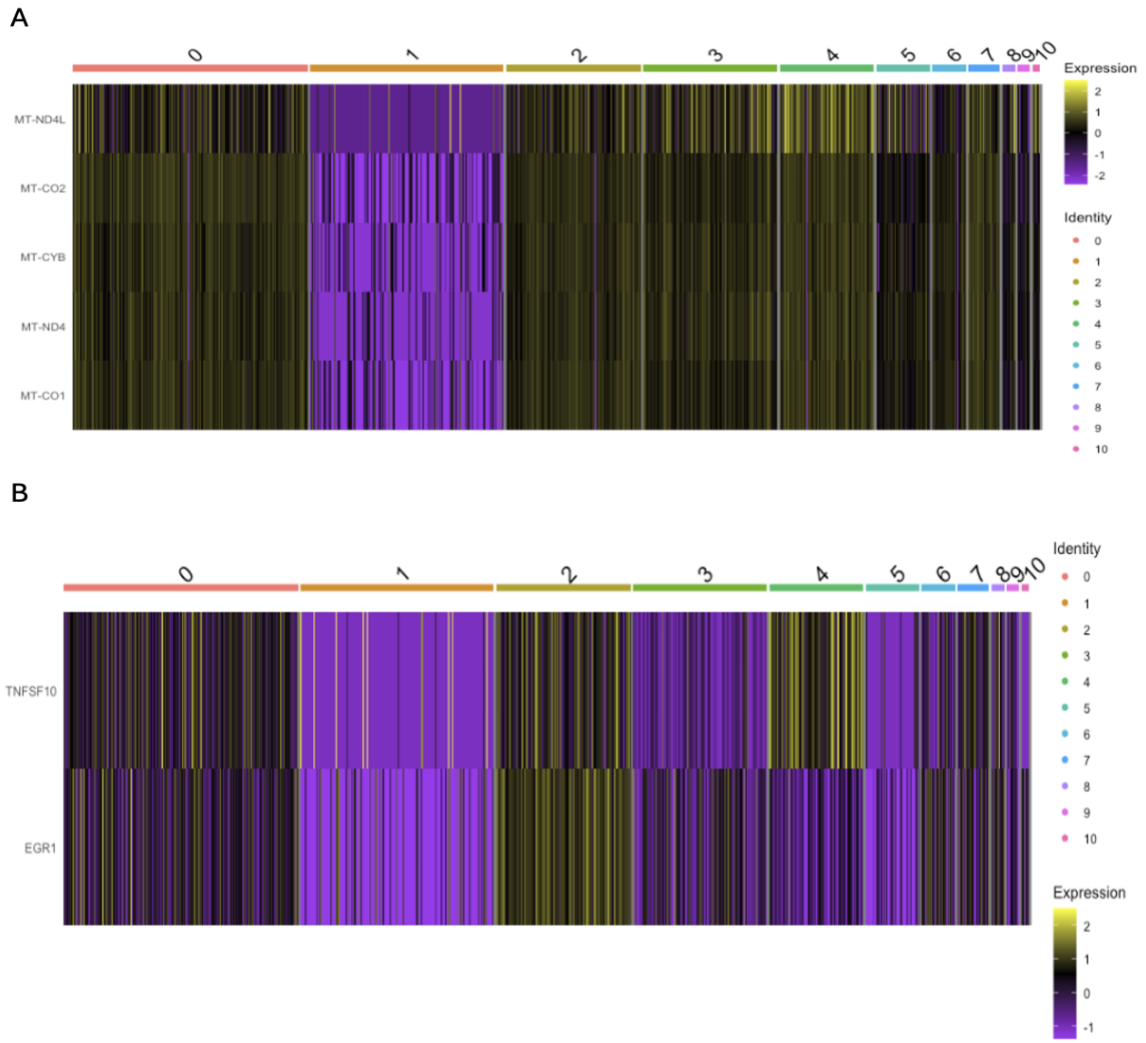

Figure 16: Heatmap analysis of cluster 1. (A) Expression of mitochondrial genes (MT-ND4L, MT-CO2, MT-CYB, MT-ND4, MT-CO1) across all clusters, showing elevated mitochondrial activity in cluster 1. (B) Expression of selected marker genes ,TNFSF10 and EGR1, across all clusters, highlighting their relative decreased expression in cluster 1. Colors indicate relative expression levels, with yellow representing high expression and purple representing low expression.

Clusters 0, 2 and 4 all observed higher expression of a substantial amount of un-named and unclassified genes, some of which are AC091173.1(p = 2.8e-15, Log2FC = 2.8), AC005841.1(p = 4.5e-6, Log2FC = 2.9) and AC011471.2(p = 1.47e-1, Log2FC = 3.5). Because these genes are unannotated/unclassified we do not know what the impact of their overexpression is yet. Clusters 0, 2 and 4 also did not contain the same differential expression of poor prognostic markers seen in all of the other clusters.

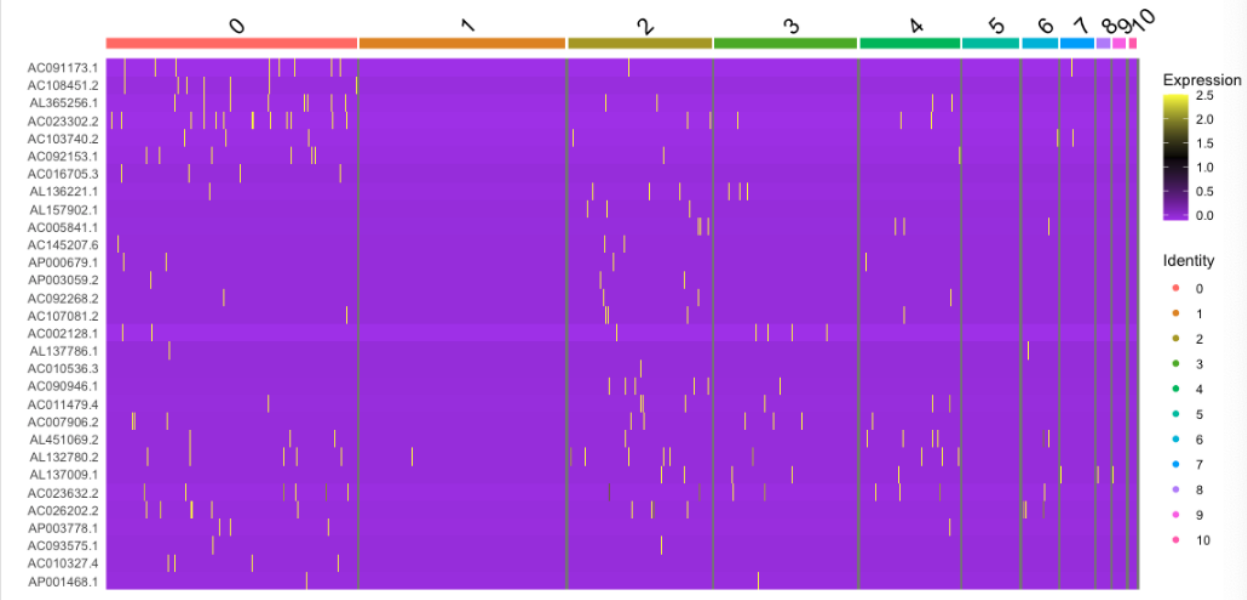

Figure 17: Heatmap of unclassified genes across clusters 0, 2, and 4. Expression patterns of 30 uncharacterized genes are shown across all clusters. Colors indicate relative expression levels, with yellow representing high expression and purple representing low expression.

Discussion:

There were several significant findings for each of the clusters identified in this study. The types of genes with significant variable gene expression could indicate several biological processes that could be contributing to the progression of this patient's IDC. Some of these processes are well characterized and have been linked to poor prognosis, therapy resistance, metastasis, and most importantly, invasion.

The most highly variably expressed genes were almost all found to be solely highly expressed in clusters 8, 9 and 10. These genes were DCN, LUM, COL1A1, COL1A2, COL3A1, and SPARC and are all implicated in ECM remodeling. Remodeling of the ECM can impact the presence/performance of cell surface receptors and the ability of chemotherapies to enter or impact the cell. The decreased effectiveness of chemotherapy caused by EMC remodeling can lead to cancer proliferation and metastasis (Yuan et al, 2023). IGFBP7, insulin-like growth factor binding protein-7, was also found to be highly expressed al clusters 8, 9 and 10 and is heavily link to breast cancer progression. The genes with the most significant increase in expression in mainly cluster 8 compared to all other cells were DCN and LUM. Typically, an increase in either of these genes' expression on their own could result in tumor suppression. However, when both genes are down regulated it can be permissive to excessive tumor growth. In this case we are seeing increased expression of both genes. An increase in both DCN and LUM in conjunction is typically associated with cancer associated fibroblasts (CAFs). This type of cell morphology is linked to progression to invasive types of cancer and resistance to chemotherapy. (Cai et al.,2024)

The remaining most highly variably expressed genes, CCL3 and APOE, were significantly more highly expressed in cluster 5, along with TYROBP, than any other cluster. CCL3 and

TYROBP are both known to inhibit apoptosis and support invasion in breast cancer, which is a key component of IDC cells (Luo et al,2020). APOE, when not expressed as an inherited biomarker, is closely linked to breast cancer metastasis (Ganggang et al,2023). All three of these biomarkers could be implicated in further metastasis or invasion of this patient's IDC if proliferation continues.

Similarly, both clusters 6 and 7 contain their own poor prognostic indicators that are completely different from cluster 5 or each other. In cluster 6 IGSF1 or Immunoglobulin superfamily member 1, was found to be highly expressed. It is a glycoprotein in the cell membrane and it is almost always only expressed in tumors and not in normal tissues. When inhibited it has proven helpful in insuring tumor cells cannot avoid immune cell evasion (Koh et al., 2024). In this case, the cells of cluster 6 may be evading immune cell destruction and could benefit from an immune-checkpoint inhibitor. SOX2-OT was also found to be significantly highly expressed in cluster 6. Overexpression of SOX2-OT, a long non-coding RNA, allows for therapy resistance, proliferation and invasion into surrounding tissues. It is being looked at currently as a potential biomarker for breast cancer progression and a target for personalized therapies (Yousefi et al, 2025). The increased expression of DSCAM-AS1 is particularly of note for cluster 6 because it is a long non-coding RNA that is involved in regulation of 908 genes. These genes are mostly involved in inflammatory responses and the cell cycle. It has a correlation to breast cancer relapse, but not to progression of the disease (Elhasnaoui et al.,2020).

In cluster 7 EXO1, NCAPH, CENPF and PKMYT1 were found to all be more highly expressed than in any other cluster. EXO1 is an exonuclease encoding gene that aides in DNA repair and maintenance after double stranded breaks. In breast cancer it is a poor prognostic marker when highly expressed (Liu and Zhang, 2021). NCAPH is associated with a poor prognosis in breast cancer when it is highly expressed. The level of its expression is positively correlated with the aggressiveness of the tumor (Mendiburu-Elicabe et al., 2023). PKMYT1 is a protein kinase associated with the cell membrane. It has been indicated as a poor prognostic marker in breast cancer. When it is knocked out, invasion, migration and proliferation is inhibited (Li eta l., 2023). Finally, Centromere protein F (CENPF) is an antigen associated with the cell cycle. As the cell progresses from G0 to G2/M the expression of CENPF increases. Excessive expression of CENPF has been identified as a poor prognostic indicator in many types of cancers, including breast cancers. In breast cancer it has been linked to poor prognosis, cell proliferation and metastasis.

It is worth noting that even if combined, clusters 5, 6 and 7 do not contain as many cells as cluster 0. When interpreting the meaning of these clusters all having their own poor prognostic indicators, it may be worth considering that these cells may be progressing into sub-clones along with clusters 8, 9 and 10. Although these populations are small at the time of sampling, they could progress into dominant tumor types if left to proliferate. Based on the genetic profiles identified by this study, it may be worth considering a personalized cancer therapy to target the highly studied poor prognostic indicators for this specific case and cases like it.

When looking at cluster 1 we identified a decrease in the expression of several key mitochondrial genes, such as MT-COS and MT-CYB. The low expression of several mitochondrial genes, including

essential genes MT-CYB and MT-ND, in comparison to their expression in the main clusters, 0, 2 and 4 , may indicate an increase in reactive oxygen species (ROS) linked to an increased cancer cell survival. It is suggested that disruption of oxidative phosphorylation (OXPHOS) caused by mtDNA mutations could be linked to suppression of apoptosis (Yadav and Chandra, 2014). At present, mitochondrial transplantation is being looked at as a possibility in treating mitochondrial dysfunction in cancers like IDC and other diseases (Zong et al.,2024). A significant reduction in the expression of EGR1 and TNFSF10 was also observed in the cells of cluster 1. EGR1 in breast cancer has demonstrated itself to be a tumor suppressor gene. When it is normally expressed it can arrest breast cancer tumor cells in G0 or G1, inhibiting further growth and proliferation. When EGR1 is down regulated the opposite is true and can be seen as a common poor prognostic marker for breast cancers (Wei et al.,2017). TNFSF10 both initiates apoptosis and regulates metastasis of cancers when normally regulated. It can slow or stop the metastasis or invasion of cancer cells. A reduced expression of TNFSF10 removes one of the cells normal protections against proliferation and invasion (Oh and Sun, 2021).

Cluster 3, curiously, mainly showed reduced ribosomal gene expression. Of note, RPS9 was expressed significantly lower in the cells of cluster 3 than in any other cluster. This is related to poor overall survival in breast cancer. In general, reduced ribosomal gene expression in breast cancer is associated with poor survival. The down regulation of RPS9 and other ribosomal genes can allow for increased cell proliferation (Khoury and Nasr, 2021).

During the analysis of the most variably expressed genes in clusters 0, 2 and 4 compared to the rest of the sorted cells it was discovered that many of the most significantly, variably expressed genes were currently un-named and were without functional annotation at the time of this study. These genes may warrant functional studies due to the quantity and their presence in the largest of cell clusters in this study.

Conclusion:

The aim of this study was to identify novel biomarkers or drivers in the progression of IDC from DCIS. This study identified several possible sub-clones and a route for continued studies. Clusters 5-10 variably expressed genes that are known poor prognostic indicators with known targeted therapies. If these clusters progress into sub-clones, there are personalized therapy options available. In clusters 1 and 3, variable expression of important mitochondrial and ribosomal genes was identified respectively. Both abnormalities are currently being research, without there being a distinct way to treat them to date. Finally, clusters 0, 2 and 4, the largest groups, contained a substantial number of unannotated genes with significantly higher expression than any other clusters. Functional studies on these genes may help uncover novel mechanisms of tumor progression. Future work should investigate possible treatments for the ribosomal and mitochondrial gene abnormalities identified in clusters 1 and 3 respectively, as there is not currently a well-documented treatment option for either abnormality. It may also be worth looking into a possible correlation between the low expression of EGR1 and TNFSF10 and the low expression of mitochondrial genes from cluster 1 in future studies. These findings highlight potential therapeutic targets and paths towards identifying new drivers for IDC progression.

References:

(1) *What Is a Breast Cancer's Grade? | Grading Breast Cancer*. https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-grades.html (accessed 2025-08-03).

(2) Li, X.; Jin, Y.; Xue, J. Unveiling Collagen's Role in Breast Cancer: Insights into Expression Patterns, Functions and Clinical Implications. *Int J Gen Med* **2024**, *17*, 1773–1787. https://doi.org/10.2147/IJGM.S463649.

(3) Xu, H.-R.; Chen, J.-J.; Shen, J.-M.; Ding, W.-H.; Chen, J. TYRO Protein Tyrosine Kinase-Binding Protein Predicts Favorable Overall Survival in Osteosarcoma and Correlates with Antitumor Immunity. *Medicine* **2022**, *101* (39), e30878. https://doi.org/10.1097/MD.0000000000030878.

(4) Wu, X.; Srinivasan, P.; Basu, M.; Zhang, P.; Saruwatari, M.; Thommandru, B.; Jacobi, A.; Behlke, M.; Sandler, A. Tumor Apolipoprotein E Is a Key Checkpoint Blocking Anti-Tumor Immunity in Mouse Melanoma. *Front. Immunol.* **2022**, *13*. https://doi.org/10.3389/fimmu.2022.991790.

(5) Koh, D.-I.; Lee, M.; Park, Y. S.; Shin, J.-S.; Kim, J.; Ryu, Y. S.; Lee, J. H.; Bae, S.; Lee, M. S.; Hong, J. K.; Jeong, H.-R.; Choi, M.; Hong, S.-W.; Kim, D. K.; Lee, H.-K.; Kim, B.; Yoon, Y. S.; Jin, D.-H. The Immune Suppressor IGSF1 as a Potential Target for Cancer Immunotherapy. *Cancer Immunol Res* **2024**, *12* (4), 491–507. https://doi.org/10.1158/2326-6066.CIR-23-0817.

(6) Cai, L.; Kolonin, M. G.; Anastassiou, D. The Fibro-Adipogenic Progenitor APOD+DCN+LUM+ Cell Population in Aggressive Carcinomas. *Cancer Metastasis Rev* **2024**, *43* (3), 977–980. https://doi.org/10.1007/s10555-024-10181-y.

(7) Oh, Y.-T.; Sun, S.-Y. Regulation of Cancer Metastasis by TRAIL/Death Receptor Signaling. *Biomolecules* **2021**, *11* (4), 499. https://doi.org/10.3390/biom11040499.

(8) Shi, S.; Ma, H.-Y.; Han, X.-Y.; Sang, Y.-Z.; Yang, M.-Y.; Zhang, Z.-G. Prognostic Significance of SPARC Expression in Breast Cancer: A Meta-Analysis and Bioinformatics Analysis. *Biomed Res Int* **2022**, *2022*, 8600419. https://doi.org/10.1155/2022/8600419.

(9) Troup, S.; Njue, C.; Kliewer, E. V.; Parisien, M.; Roskelley, C.; Chakravarti, S.; Roughley, P. J.; Murphy, L. C.; Watson, P. H. Reduced Expression of the Small Leucine-Rich Proteoglycans, Lumican, and Decorin Is Associated with Poor Outcome in Node-Negative Invasive Breast Cancer. *Clin Cancer Res* **2003**, *9* (1), 207–214.

(10) Li, H.; Wang, L.; Zhang, W.; Dong, Y.; Cai, Y.; Huang, X.; Dong, X. Overexpression of PKMYT1 Associated with Poor Prognosis and Immune Infiltration May Serve as a Target in Triple-Negative Breast Cancer. *Front Oncol* **2023**, *12*, 1002186. https://doi.org/10.3389/fonc.2022.1002186.

(11)
Sun, J.; Huang, J.; Lan, J.; Zhou, K.; Gao, Y.; Song, Z.; Deng, Y.; Liu, L.; Dong, Y.; Liu, X. Overexpression of CENPF Correlates with Poor Prognosis and Tumor Bone Metastasis in Breast Cancer. *Cancer Cell Int* **2019**, *19*, 264. https://doi.org/10.1186/s12935-019-0986-8.

(12)
Mendiburu-Eliçabe, M.; García-Sancha, N.; Corchado-Cobos, R.; Martínez-López, A.; Chang, H.; Mao, J. H.; Blanco-Gómez, A.; García-Casas, A.; Castellanos-Martín, A.; Salvador, N.; Jiménez-Navas, A.; Pérez-Baena, M. J.; Sánchez-Martín, M. A.; Abad-Hernández, M. D. M.; Del Carmen, S.; Claros-Ampuero, J.; Cruz-Hernández, J. J.; Rodríguez-Sánchez, C. A.; García-Cenador, M. B.; García-Criado, F. J.; Vicente, R. S.; Castillo-Lluva, S.; Pérez-Losada, J. NCAPH Drives Breast Cancer Progression and Identifies a Gene Signature That Predicts Luminal A Tumor Recurrence. *Res Sq* **2023**, rs.3.rs-3231230. https://doi.org/10.21203/rs.3.rs-3231230/v2.

(13)
Luo, A.; Meng, M.; Wang, G.; Han, R.; Zhang, Y.; Jing, X.; Zhao, L.; Gu, S.; Zhao, X. Myeloid-Derived Suppressor Cells Recruited by Chemokine (C-C Motif) Ligand 3 Promote the Progression of Breast Cancer via Phosphoinositide 3-Kinase-Protein Kinase B-Mammalian Target of Rapamycin Signaling. *J Breast Cancer* **2020**, *23* (2), 141–161. https://doi.org/10.4048/jbc.2020.23.e26.

(14)
Zong, Y.; Li, H.; Liao, P.; Chen, L.; Pan, Y.; Zheng, Y.; Zhang, C.; Liu, D.; Zheng, M.; Gao, J. Mitochondrial Dysfunction: Mechanisms and Advances in Therapy. *Sig Transduct Target Ther* **2024**, *9* (1), 124. https://doi.org/10.1038/s41392-024-01839-8.

(15)
Yadav, N.; Chandra, D. Mitochondrial DNA Mutations and Breast Tumorigenesis. *Biochim Biophys Acta* **2013**, *1836* (2), 336–344. https://doi.org/10.1016/j.bbcan.2013.10.002.

(16)
Huang, X.; Taeb, S.; Jahangiri, S.; Emmenegger, U.; Tran, E.; Bruce, J.; Mesci, A.; Korpela, E.; Vesprini, D.; Wong, C. S.; Bristow, R. G.; Liu, F.-F.; Liu, S. K. miRNA-95 Mediates Radioresistance in Tumors by Targeting the Sphingolipid Phosphatase SGPP1. *Cancer Res* **2013**, *73* (23), 6972–6986. https://doi.org/10.1158/0008-5472.CAN-13-1657.

(17)
Borst, M. J.; Ingold, J. A. Metastatic Patterns of Invasive Lobular versus Invasive Ductal Carcinoma of the Breast. *Surgery* **1993**, *114* (4), 637–642. https://doi.org/10.5555/uri:pii:003960609390249D.

(18)
Godina, C.; Pollak, M. N.; Jernström, H. Low IGFBP7 Expression Identifies a Subset of Breast Cancers with Favorable Prognosis and Sensitivity to IGF-1 Receptor Targeting with Ganitumab: Data from I-SPY2 and SCAN-B. medRxiv December 19, 2023, p 2023.12.18.23300129. https://doi.org/10.1101/2023.12.18.23300129.

(19)
Invasive Ductal Carcinoma (IDC): Symptoms, Treatments, and More. https://www.breastcancer.org/types/invasive-ductal-carcinoma (accessed 2025-06-26).

(20)

Invasive Ductal Carcinoma (IDC): Overview, Treatment & Prognosis. National Breast Cancer Foundation. https://www.nationalbreastcancer.org/invasive-ductal-carcinoma/ (accessed 2025-08-01).

(21)

Invasive breast cancer. https://www.cancerresearchuk.org/about-cancer/breast-cancer/types/invasive-breast-cancer (accessed 2025-06-26).

(22)

Lebok, P.; Schütt, K.; Kluth, M.; Witzel, I.; Wölber, L.; Paluchowski, P.; Terracciano, L.; Wilke, C.; Heilenkötter, U.; Müller, V.; Schmalfeldt, B.; Simon, R.; Sauter, G.; Von Leffern, I.; Krech, T.; Krech, R. H.; Jacobsen, F.; Burandt, E. High Mitochondrial Content Is Associated with Breast Cancer Aggressiveness. *Mol Clin Oncol* **2021**, *15* (4), 203. https://doi.org/10.3892/mco.2021.2365.

(23)

Bansal, C.; Pujani, M.; Sharma, K. L.; Srivastava, A. N.; Singh, U. S. Grading Systems in the Cytological Diagnosis of Breast Cancer: A Review. *Journal of Cancer Research and Therapeutics* **2014**, *10* (4), 839. https://doi.org/10.4103/0973-1482.140979.

(24)

Hernandez, L.; Wilkerson, P. M.; Lambros, M. B.; Campion-Flora, A.; Rodrigues, D. N.; Gauthier, A.; Cabral, C.; Pawar, V.; Mackay, A.; A'Hern, R.; Marchiò, C.; Palacios, J.; Natrajan, R.; Weigelt, B.; Reis-Filho, J. S. Genomic and Mutational Profiling of Ductal Carcinomas in Situ and Matched Adjacent Invasive Breast Cancers Reveals Intra-Tumour Genetic Heterogeneity and Clonal Selection. *J Pathol* **2012**, *227* (1), 42–52. https://doi.org/10.1002/path.3990.

(25)

Tai, I. T.; Dai, M.; Owen, D. A.; Chen, L. B. Genome-Wide Expression Analysis of Therapy-Resistant Tumors Reveals SPARC as a Novel Target for Cancer Therapy. *J Clin Invest* **2005**, *115* (6), 1492–1502. https://doi.org/10.1172/JCI23002.

(26)

Miao, G.; Zhuo, D.; Han, X.; Yao, W.; Liu, C.; Liu, H.; Cao, H.; Sun, Y.; Chen, Z.; Feng, T. From Degenerative Disease to Malignant Tumors: Insight to the Function of ApoE. *Biomedicine & Pharmacotherapy* **2023**, *158*, 114127. https://doi.org/10.1016/j.biopha.2022.114127.

(27)

Fluorescence-activated cell sorting (FACS) of live cells | Abcam. https://www.abcam.com/en-us/technical-resources/guides/flow-cytometry-guide/fluorescence-activated-cell-sorting-live-cells (accessed 2025-06-26).

(28)

Yuan, Z.; Li, Y.; Zhang, S.; Wang, X.; Dou, H.; Yu, X.; Zhang, Z.; Yang, S.; Xiao, M. Extracellular Matrix Remodeling in Tumor Progression and Immune Escape: From Mechanisms to Treatments. *Molecular Cancer* **2023**, *22* (1), 48. https://doi.org/10.1186/s12943-023-01744-8.

(29)

Yousefi, M.; Karimi, M.; Rezaei, A.; Mortezazadeh, M.; Akhshabi, S.; Yousefi Ladmakhi, A.; Shirsalimi, N.; Kamali Yazdi, E.; Seyyed Mahmoudi, S. T.; Kashani, M.; Mofidi, A. Evaluation of lncRNA SOX2OT Gene Expression in Breast Cancer Patients and Its Association with Prognosis of Disease: A Paired Sample Study. *Egyptian Journal of Medical Human Genetics* **2025**, *26* (1), 123. https://doi.org/10.1186/s43042-025-00748-x.

(30)

Liu, J.; Zhang, J. Elevated EXO1 Expression Is Associated with Breast Carcinogenesis and Poor Prognosis. *Ann Transl Med* **2021**, *9* (2), 135. https://doi.org/10.21037/atm-20-7922.

(31) Elhasnaoui, J.; Miano, V.; Ferrero, G.; Doria, E.; Leon, A. E.; Fabricio, A. S. C.; Annaratone, L.; Castellano, I.; Sapino, A.; De Bortoli, M. DSCAM-AS1-Driven Proliferation of Breast Cancer Cells Involves Regulation of Alternative Exon Splicing and 3'-End Usage. *Cancers (Basel)* **2020**, *12* (6), 1453. https://doi.org/10.3390/cancers12061453.

(32) Wei, L.-L.; Wu, X.-J.; Gong, C.-C.; Pei, D.-S. Egr-1 Suppresses Breast Cancer Cells Proliferation by Arresting Cell Cycle Progression via down-Regulating CyclinDs. *Int J Clin Exp Pathol* **2017**, *10* (10), 10212–10222.

(33) Hao, Y.; Stuart, T.; Kowalski, M. H.; Choudhary, S.; Hoffman, P.; Hartman, A.; Srivastava, A.; Molla, G.; Madad, S.; Fernandez-Granda, C.; Satija, R. Dictionary Learning for Integrative, Multimodal and Scalable Single-Cell Analysis. *Nat Biotechnol* **2024**, *42* (2), 293–304. https://doi.org/10.1038/s41587-023-01767-y.

(34) El Khoury, W.; Nasr, Z. Deregulation of Ribosomal Proteins in Human Cancers. *Biosci Rep* **2021**, *41* (12), BSR20211577. https://doi.org/10.1042/BSR20211577.

(35) Järvinen, T. A. H.; Prince, S. Decorin: A Growth Factor Antagonist for Tumor Growth Inhibition. *Biomed Res Int* **2015**, *2015*, 654765. https://doi.org/10.1155/2015/654765.

(36) Saadat, M. Apolipoprotein E (APOE) Polymorphisms and Susceptibility to Breast Cancer: A Meta-Analysis. *Cancer Res Treat* **2012**, *44* (2), 121–126. https://doi.org/10.4143/crt.2012.44.2.121.