

<https://doi.org/10.1038/s43856-026-01548-z>

Intratumor heterogeneity in human basal like breast cancer is revealed by single cell molecular subtyping

Check for updates

Jing Li^{1,6}, Yuan Gao^{2,6}, Sirui Zhang^{1,6}, Shouhui Guo¹, Qingzhen Hou¹, Shisong Zhang^{3,4}✉, Weixing Yu⁵✉ & Ke Liu¹✉

Abstract

Background Breast cancer is a molecularly heterogeneous disease composed of multiple intrinsic subtypes. Recent studies have highlighted the substantial intratumor heterogeneity of breast cancer, wherein malignant cells of distinct intrinsic subtypes co-exist within the same tumor. However, most existing subtyping methods are designed for bulk transcriptomic data and are therefore limited in their ability to resolve such intratumor heterogeneity at single-cell resolution.

Methods We develop UBS93, a computational framework that enables robust molecular subtyping of both bulk tumor samples and individual breast cancer cells. We rigorously validate UBS93 and demonstrate its superior performance relative to existing approaches, particularly in identifying the highly aggressive Claudin-low subtype.

Results Applying UBS93 to single-cell RNA sequencing data from human Basal-like breast cancers, we identify the co-existence of Basal-like and Claudin-low cancer cell populations within the same tumor—a form of intratumor heterogeneity previously observed in mouse models with genetically engineered RAS pathway alterations. Further analyses suggest that Claudin-low cancer cells originate from Basal-like population, with down-regulation of transcription factor *ELF3* playing a pivotal role in the Basal-like/Claudin-low transition.

Conclusions Our findings establish UBS93 as a powerful tool for breast cancer subtyping and uncover the intratumor heterogeneity in human Basal-like breast cancer.

Plain language summary

Breast cancer is not a single disease and different subtypes of cancer cells can exist within the same tumor. Most current methods used to classify breast cancer look at tumors as a whole and cannot detect differences between individual cells. In this study, we developed a computational tool called UBS93 which can classify breast cancer using data from either whole tumor or single cells. Using UBS93, we found that two aggressive subtypes can exist together within the same tumor. We also discovered that one subtype may arise from the other through changes in gene activity. These findings improve our understanding of breast cancer diversity and may help guide more precise diagnosis and treatment in the future.

Breast cancer is a heterogeneous disease characterized by diverse genomic profiles, transcriptomic alterations, and clinical outcomes. This heterogeneity presents significant challenges for achieving clinically meaningful tumor classification¹. Parker et al. utilized microarray technology to develop an intrinsic breast cancer classifier, PAM50, which stratifies the disease into five subtypes: luminal A, luminal B, HER2-enriched, basal-like, and normal-like². This classification system provides superior prognostic and predictive capabilities compared to traditional methods, such as pathological staging, histological grading, standard clinical biomarkers, and so on ref.³. In addition to PAM50, Herschkowitz et al. identified the highly aggressive claudin-low subtype which is characterized by low expression levels of cell adhesion

molecules (e.g., *CLDN3*, *CLDN4*, *CLDN7*, and *CDH1*) and enriched for high stemness and mesenchymal features^{3–5}. Fougner et al. revealed distinguished characteristics of claudin-low tumors through integrative multi-omics analyses across multiple cohorts, although its cellular origin remains unclear⁶.

The use of single-cell omics technology has profoundly enhanced our understanding of intratumor heterogeneity in breast cancer^{7–15}. Especially, several recent studies have reported the co-existence of cancer cells of multiple molecular subtypes within the same tumor. For instance, Wu et al. identified confident basal-like cancer cells within a human estrogen receptor-positive (ER⁺) breast tumor (which are usually subtyped as luminal

¹Department of Medical Dataology, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, China. ²Precision Medicine Laboratory for Chronic Non-communicable Diseases of Shandong Province, Institute of Precision Medicine, Jining Medical University, Jining, China. ³Department of Thoracic and Oncological Surgery, Children's Hospital Affiliated to Shandong University, Jinan, China. ⁴Department of Thoracic and Oncological Surgery, Jinan Children's Hospital, Jinan, China. ⁵Department of Biochemistry and Molecular Biology, School of Basic Medicine, Jining Medical University, Jining, China. ⁶These authors contributed equally: Jing Li, Yuan Gao, Sirui Zhang. ✉e-mail: zsl323@163.com; weixingy@mail.jnmc.edu.cn; keliu.iluke@email.sdu.edu.cn

A or luminal B)¹⁶; in addition, Rädler et al. demonstrated that modulation of the RAS signaling pathway in mouse models can induce tumors containing both basal-like and claudin-low cancer cell populations⁵. These discoveries underscore the limitations of traditional bulk omics-based subtyping strategy and highlight the importance of achieving single-cell resolution for more precise classification.

Despite advances in computational subtyping, significant challenges remain in delineating the intratumoral heterogeneity described above. First, most existing subtyping tools are designed for bulk omics data and the intrinsic sparsity and dropout of single-cell omics data, therefore, substantially limit their applicability to individual cells. Second, there is a lack of a uniform gene panel capable of robustly classifying breast cancer subtypes across heterogeneous cellular states. To the best of our knowledge, SCSubtype is the only available tool, which enables molecular subtyping of breast cancer at single-cell resolution; however, it internally relies on the PAM50 gene panel and overlooks the highly aggressive claudin-low subtype. Addressing these limitations is critical for refining breast cancer subtyping and guiding precise diagnostic and therapeutic interventions.

In this study, we develop a computational framework (UBS93), which enables molecular subtyping of both bulk tumor samples and single cancer cells. Applying UBS93 to single-cell RNA sequencing (scRNA-seq) data, we identify intratumor heterogeneity in human basal-like breast cancer, characterized by the co-existence of basal-like and claudin-low cancer cell populations within the same tumor. Further analysis suggests that claudin-low cancer cells arise from basal-like populations, with downregulation of the transcription factor *ELF3* can induce a partial claudin-low phenotype.

Methods

Development of the UBS93 gene panel

We obtained a breast cancer scRNA-seq dataset consisting of 8 samples from Wu et al.'s research¹⁶. Given a gene *g*, its *C-M*-ratio value in sample *S* is calculated with the following formula:

$$C-M\text{-ratio} = (C_g + 0.1)/(M_g + 0.1)$$

Here, C_g and M_g represent the average expression of *G* in cancer cells and non-cancer cells (of sample *S*), respectively.

Next, we divided all the samples into two groups: the TNBC group (2 samples) and the non-TNBC group (6 samples). We computed the group-wise *C-M*-ratio of gene *g* by taking the mean *C-M*-ratio within each group. Consequently, we obtained two group-wise *C-M*-ratio values for each gene *g*.

Finally, we ranked all genes according to their average group-wise *C-M*-ratio. We performed a quantile regression analysis between the average group-wise *C-M*-ratio (*Y*) and rank (*X*), and noticed that there appears to be a change point of the slope. We picked out the top 400 genes to re-perform the analysis and found 300 appeared to be a reasonable estimate of the change point. Therefore, we finally kept the top 100 genes (Supplementary Fig. 1). After excluding 7 genes that lack ENSEMBL gene IDs, we established the UBS93 gene panel.

Determination of breast cancer subtypes

The CCLE categorized breast cancer into five subtypes: basal_A, basal_B, luminal, HER2-amp, and luminal-HER2-amp. In the t-SNE plot (Fig. 1c), basal_A and basal_B form their own distinct cluster while the other three subtypes form the third cluster. The basal_A cluster is significantly enriched with cell lines whose PAM50 subtype is basal-like; therefore, it is reasonable to rename it as basal-like; similarly, basal_B was renamed as claudin-low. The luminal-HER2-amp subtype is not used in our classifier since the two cell lines of this subtype do not appear to form a distinct cluster.

Differentiating luminal and HER2-amp subtypes

We conducted differential gene expression analysis between CCLE HER2-amp and luminal cell lines and identified 860 significantly up-regulated

genes. We picked out the 20 genes, which are present in the UBS93 panel, and derived the HER2-amp signature.

In our classifier, ssGSEA analysis (UBS93 genes as background, and HER2-amp signature as a geneset) is utilized to differentiate luminal and HER2-amp subtypes. We validated such strategy on a pooled bulk RNA-seq dataset (GSE212143 and GSE48213) and obtained an AUC (area under the curve) value of 0.964 (Supplementary Fig. 2). In the final classifier, we determined the cutoff value to be 1.662 based on Youden's index calculated in the validation dataset.

Construction of validation dataset using TCGA breast cancer samples

We downloaded clinical data of the Cancer Genome Atlas (TCGA) breast cancer samples from cBioPortal (https://datahub.assets.cbioportal.org/brca_tcga.tar.gz)^{17,18} and utilized the following pipeline to assign subtypes.

First, hierarchical clustering of metaplastic breast cancer samples identified three samples characterized by reduced expression of *CLDN3*, *CLDN4*, *CLDN7*, and *EPCAM*, together with elevated *VIM* expression (Supplementary Fig. 3). These samples were therefore classified as the claudin-low subtype.

Second, non-metaplastic triple-negative breast cancer (TNBC) samples were classified as the basal-like subtype.

Third, samples reported to harbor HER2 amplification were categorized as the HER2-amplified subtype.

Fourth, the remaining samples showing positivity for at least one of ER, PR, or HER2 were classified as the luminal subtype.

Finally, samples that did not meet any of the above criteria were excluded from further analysis. In total, 974 samples were retained for performance evaluation and comparison.

Breast cancer subtyping with Genefu

In the Genefu package, the gene sets used for claudin-low and PAM50 subtyping are different. When analyzing a breast cancer sample, we first employ the function `claudinLow` to determine whether its subtype is claudin-low; if not, we subsequently conduct PAM50 subtyping. In PAM50 subtyping, we removed the normal-like subtype and merged luminal A and luminal B as luminal.

Gene set enrichment analysis

The EMT signature was downloaded from MSigDB¹⁹. To derive the claudin-low-specific depletion (CLSD) gene set, z-score normalization was applied to gene expression values of CCLE breast cancer cell lines. For each gene, the median expression value across claudin-low cell lines was computed. If this value is below -1 , then the gene was considered as a CLSD gene. In total, this approach identified 960 CLSD genes. The clusterProfiler R package was used to compute enrichment score and *p*-value²⁰.

Processing of bulk RNA-seq data

We mapped raw bulk RNA-seq reads to the transcriptome index with STAR²¹ and then utilized RSEM²² to quantify each gene's transcripts per million value. The genome versions used are hg38 and mm10, respectively. The genome annotation files were downloaded from GENCODE²³. The processed CCLE (version: CCLE 2019) and TCGA bulk RNA-seq data were downloaded from DepMap (https://depmap.org/portal/data_page/?tab=allData) and UCSC Xena (<https://xena.ucsc.edu/>).

The t-SNE analysis

Given two single cells *C1* and *C2*, their distance is defined as $1 - \text{cor}(C1, C2)$, where $\text{cor}(C1, C2)$ is the Spearman rank correlation value across used genes. After the computation of the distance matrix, `Rtsne` function is used to generate the final embedding.

Cell lines and cell culture conditions

HCC70 cells and HEK293T cells were obtained from Pricella Biotechnology Co., Ltd. and American Type Culture Collection. HCC70 cells maintained

in RPMI 1640 (catalog no. 22400089, Gibco). HEK293T cells maintained in DMEM (catalog no. SH30285.01, Cytiva Life Sciences). All medium supplemented with 10% FBS (catalog no. 164210-50, Pricella Biotechnology).

Lentivirus packaging and infection

HEK293T cells (2×10^6) were seeded in a 100-mm cell culture dish and transfected with plasmids using PolyJet in vitro DNA transfection reagent PEI (C0537-20 mg, Beyotime Biotechnology). The transfection experiments were operated according to the manufacturer's instructions. For knock-down target proteins, we co-transfected the cells with the plasmids of pMDL, pVSVG, pREV, and pLKO.1 (sh-*ELF3*, scrambled) to generate lentivirus. For lentivirus infection, 500 μ l of virus-containing medium was added to the 5×10^4 cells with 0.1% polybrene.

Quantitative real-time PCR

Total RNA was extracted from cells using an RNAsimple Total RNA Kit (TIANGEN, DP419). Complementary DNA was prepared using the high-capacity cDNA Reverse Transcription Kit purchased from Vazyme Biotech (catalog no. R333-01). Real-time PCR analysis was performed using $2 \times$ Universal SYBR Green Fast qPCR Mix (catalog no. RK21203, Abclonal Technology) with the following conditions: 3 min at 95 °C, and then 40 cycles at 95 °C for 5 s, 60 °C for 25 s, using a real-time system (Roche LightCycler® 480). Data were normalized according to the expression of a control gene (*ACTB*) in each experiment. The data are presented as the mean $\pm$ s.d. from three independent experiments. The *p*-values were computed with an unpaired t-test in GraphPad Prism 10 software.

Single-cell RNA-sequencing of HCC70

The single-cell RNA-sequencing of the HCC70 cell line was performed by a commercial company Novogene. Briefly, RNA from the barcoded cells was subsequently reverse-transcribed, and sequencing libraries were constructed with reagents from a Chromium Single Cell 3' v2 reagent kit (10X Genomics) according to the manufacturer's instructions. Sequencing was performed with Illumina sequencing machines according to the manufacturer's instructions. After filtering low-quality reads with fastp²⁴, raw reads were demultiplexed and mapped to the human reference genome by cellranger 7.2.0²⁵.

Software tools

All the analysis was conducted in R. Seurat was used for processing scRNA-seq data. In DE analysis^{26–29}, genes with adjusted *p* < 0.05 and $\text{abs}(\log_2\text{FC}) > 0.5$ were considered as differentially expressed genes. The ggplot2 package was used for data visualization, pheatmap and ComplexHeatmap packages were used to generate heat maps^{30–32}, GSVA was used for ssGSEA analysis^{33,34}.

Statistics and reproducibility

If not specified, the two-sided Wilcoxon rank-sum test was used to compute the *p*-value in hypothesis testing.

Ethical approval

No ethics approval was sought for this study, and participant consent was not required as all data included was publicly available/deidentified.

Results

Development of the UBS93 classifier for breast cancer subtyping

Solid tumor tissue is a heterogeneous mixture of various cell types¹⁶. In bulk RNA sequencing, the measured expression of a gene reflects a weighted average of its expression across all constituent cells. Specifically, for a given gene *g*, if its average expression in cancer cells and non-cancer cells is denoted by C_g and M_g , respectively, and the tumor purity is α , the observed expression in a bulk sample can be approximated as $\alpha \times C_g + (1 - \alpha) \times M_g$. As a result, gene expression rankings in a bulk tumor sample may deviate substantially from those in pure cancer cells due to the influence of the tumor microenvironment (Fig. 1a, left panel); however, for genes

highly specific to cancer cells (i.e., $M_g \approx 0$), the rankings in a bulk tumor sample and pure cancer cells are expected to remain highly concordant (Fig. 1a, right panel). This observation suggests that the relative expression ranks of cancer cell-specific genes may serve as robust features for molecular subtyping.

To identify such genes, we analyzed the single-cell RNA sequencing (scRNA-seq) dataset from Wu et al.'s study and computed the “C–M ratio” value for each gene, representing its expression specificity to cancer cells. Among the top 100 genes with the highest C–M ratios, seven lacked ENSEMBL IDs and were excluded. The remaining 93 genes formed a uniform gene panel for breast cancer subtyping, termed UBS93 (Supplementary Data 1). As expected, UBS93 genes exhibited significantly higher expression in cancer cells compared to non-cancer cells (Fig. 1b).

To validate this gene panel, we utilized breast cancer cell line data from the cancer cell line encyclopedia (CCLE), which classifies cell lines into five subtypes: basal_A, basal_B, luminal, HER2-amplified (HER2-amp), and luminal-HER2-amp³⁵. A t-distributed stochastic neighbor embedding (t-SNE) analysis of CCLE bulk RNA-seq data revealed clustering patterns consistent with these subtypes (Fig. 1c), supporting the validity of the CCLE classification. For consistency with widely used breast cancer nomenclature, we renamed basal_A and basal_B as basal-like and claudin-low, respectively, and excluded the luminal-HER2-amp subtype from further analysis. This yielded four refined subtypes: basal-like, claudin-low, luminal, and HER2-amp. We then performed t-SNE using only the UBS93 gene set on the same CCLE dataset (Supplementary Fig. 4). The resulting subtype-specific clusters mirrored those from the full transcriptome analysis, confirming the utility of UBS93 in distinguishing breast cancer subtypes.

Building on these observations, we developed our UBS93 subtype classifier (Fig. 1d). We first created a 93×3 subtype centroid matrix, where each column represented the average expression profile of UBS93 genes across the CCLE cell lines corresponding to basal-like, claudin-low, luminal or HER2-amp, respectively. For a given bulk RNA-seq or scRNA-seq gene expression profile, we computed its Spearman rank correlation with each subtype centroid and assigned it to the subtype with the highest correlation. If the sample is subtyped as luminal or HER2-amp, we further calculated an enrichment score based on HER2-amp signature genes to determine the final subtype assignment.

Evaluation of the UBS93 classifier in human and mice

We first evaluated the performance of the UBS93 classifier using an assembled dataset, which contains the bulk RNA-seq profile of 47 breast cancer cell lines extracted from the Gene Expression Omnibus database^{36–38} (Supplementary Data 2). We found 93.6% of the samples were distributed on the diagonal of the confusion matrix, suggesting the high overall accuracy of our method (Fig. 2a). Notably, our classifier also demonstrated optimal subtype-specific performance and the F1 scores for the four subtypes were 0.97 (basal-like), 0.80 (claudin-low), 1.00 (HER2-amp), and 0.92 (luminal), respectively. In addition to cell lines, we also evaluated our classifier on the breast cancer samples from TCGA and derived ideal overall accuracy (81.9%) and subtype-specific performance (basal-like: 0.78; claudin-low: 0.67; Her2-amp: 0.51; luminal: 0.89)³⁹ (Fig. 2b).

Next, we validated our classifier using an assembled scRNA-seq dataset of human breast cancer cell lines (Supplementary Data 3)^{40,41}. For each cell line, we computed the accuracy as the proportion of cells that were correctly subtyped. Our classifier achieved an accuracy higher than 80% in 19 of the 30 cell lines (63.3%), with a median accuracy of 85.7%. The subtype-specific median accuracy values were 99.0% for basal-like, 98.3% for claudin-low, 78.0% for HER2-amp, and 75.7% for luminal (Fig. 2c).

We also validated the classifier on a dataset of 27 bulk RNA-seq samples from murine-derived studies. This dataset included 17 samples from Hollern's study⁴², which were experimentally confirmed as claudin-low, and 10 samples from Rädler's study⁵, with unknown subtype labels.

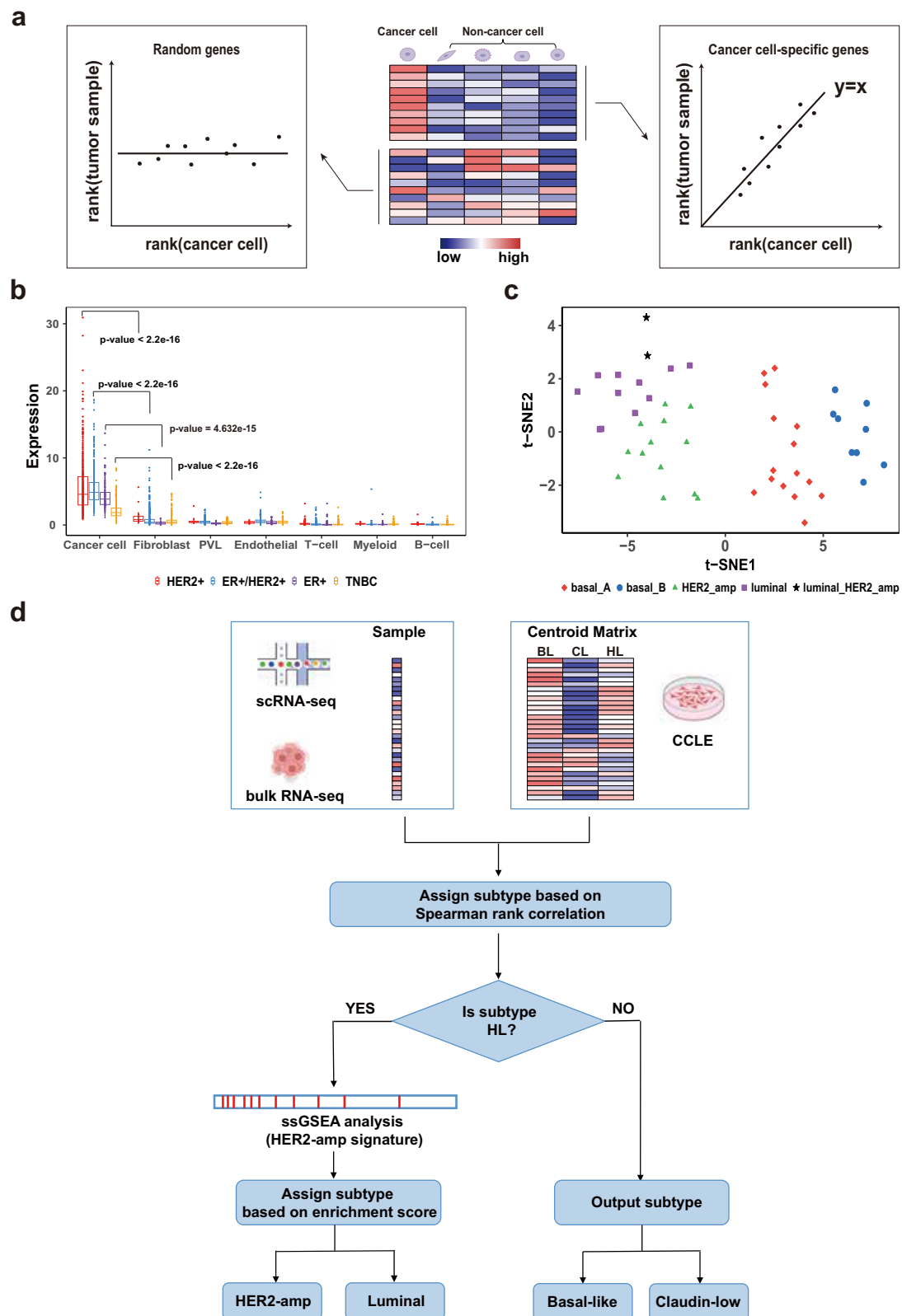


Fig. 1 | Development of the UBS93 classifier. a Schematic representation illustrating the rationale for utilizing cancer cell-specific genes in molecular subtyping. **b** Average expression values of UBS93 genes across various cell types in breast cancer patients. In the boxplots, each dot represents a single cell, the central line indicates the median, and the bounds represent the 25th and 75th percentiles (interquartile

range). Whiskers extend to 1.5 times the interquartile range. *P*-values were calculated using the two-sided Wilcoxon rank-sum test. **c** The t-SNE visualization of CCLE breast cancer cell lines based on the full gene expression matrix, colored by CCLE subtype annotations. **d** Workflow diagram for the UBS93 classifier. BL basal-like, CL claudin-low, HL HER2-amp or luminal.

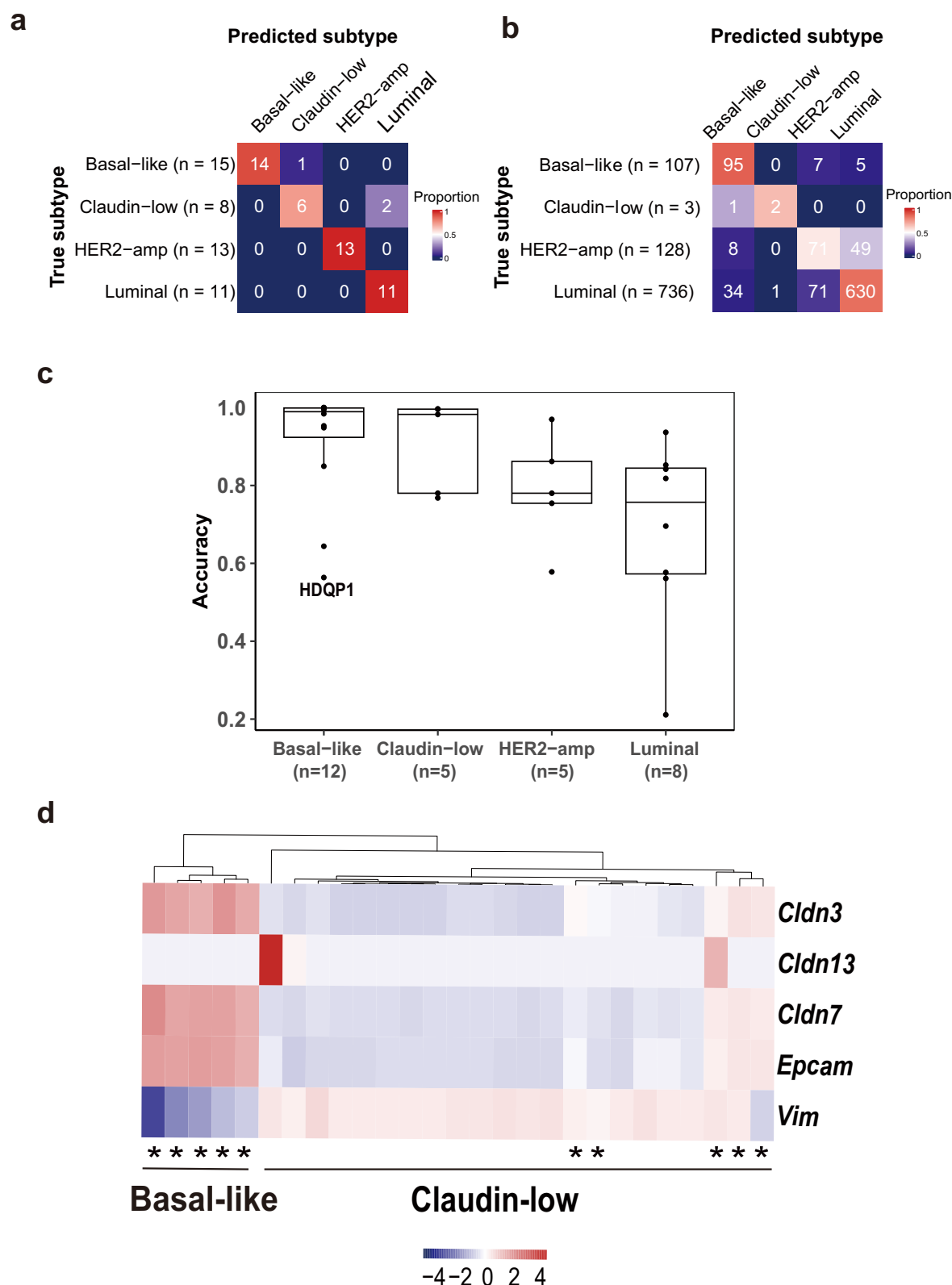


Fig. 2 | Evaluation of the UBS93 classifier in human and mouse. a Evaluation of UBS93 using bulk RNA-seq data of breast cancer cell lines. In the confusion matrix, rows represent the true subtypes and columns represent the predicted subtypes. Numbers in each cell denote sample counts, while the background color indicates the row-wise proportion. Row labels indicate the total number of samples for each true subtype. **b** Evaluation of UBS93 using bulk RNA-seq data from TCGA breast cancer samples. The confusion matrix follows the same format as in (a). **c** Evaluation of

UBS93 using scRNA-seq data of breast cancer cell lines. In the boxplots, each dot represents a cell line, the central line indicates the median, and the bounds represent the 25th and 75th percentiles (interquartile range). Whiskers extend to 1.5 times the interquartile range. **d** Scaled expression values of five claudin-low indicative genes in the assembled murine dataset (27 samples). Asterisks highlight samples whose subtypes were not reported in Rädler's study.

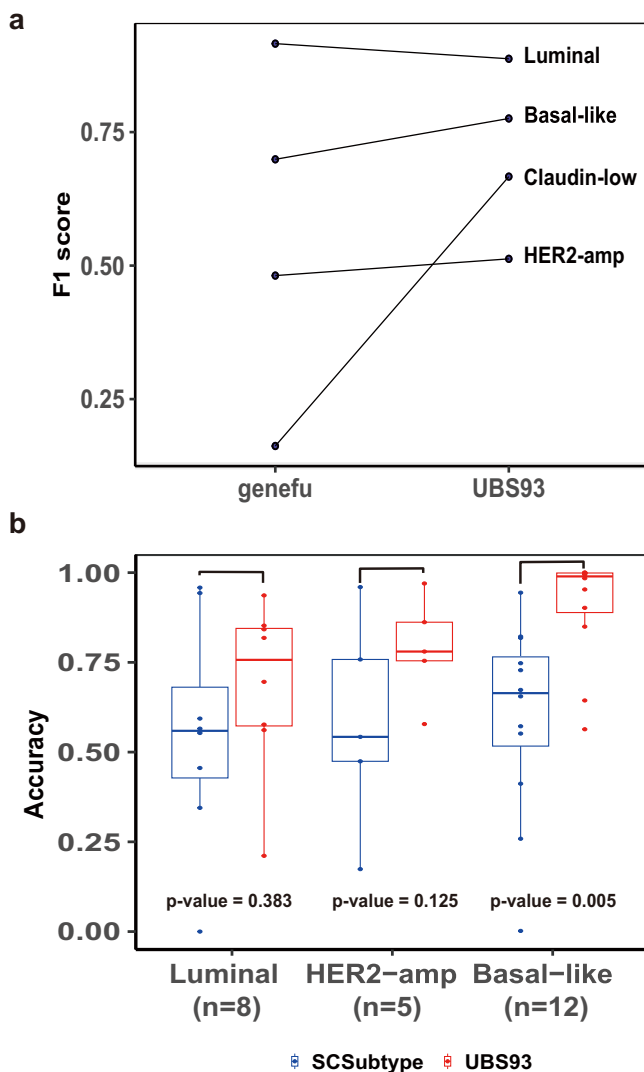


Fig. 3 | Comparison to existing methods. **a** Comparison of UBS93 with Genefu using TCGA breast cancer samples. **b** Comparison of UBS93 with SCSubtype using scRNA-seq data. In the boxplots, each dot represents a cell line, the central line indicates the median, and the bounds represent the 25th and 75th percentiles (interquartile range). Whiskers extend to 1.5 times the interquartile range. *P*-values were calculated using the two-sided Wilcoxon signed-rank test.

Based on hierarchical clustering of five “claudin-low indicative” genes (*Cldn3*, *Cldn13*, *Cldn7*, *Vim*, and *Epcam*), we imputed the missing subtype labels (Fig. 2d). The UBS93 classifier accurately assigned subtypes to all 27 samples, with results aligning with the known subtypes (Supplementary Data 4).

Comparison to existing methods

We systematically compared the UBS93 classifier with Genefu, a widely used R package for molecular subtyping of breast cancer⁴³. Notably, only six genes were shared between UBS93 and PAM50 panels (Supplementary Fig. 5a); moreover, UBS93 genes exhibited significantly higher *C-M*-ratio values and cancer cell specificity compared to PAM50 genes (Supplementary Fig. 5b).

When comparing classifier performance using bulk RNA-seq data from TCGA breast cancer samples, Genefu achieved a slightly higher overall accuracy (83.7%). However, UBS93 outperformed Genefu in classifying basal-like, claudin-low, and HER2-amp subtypes (Fig. 3a). In murine datasets, Genefu misclassified all basal-like samples as claudin-low, whereas UBS93 accurately identified them as basal-like (Supplementary Data 5).

In addition to Genefu, we also compared UBS93 with SCSubtype, a method designed for subtyping individual breast cancer cells¹⁶, using the same scRNA-seq dataset used to evaluate our classifier. Since SCSubtype does not include the claudin-low subtype, the comparison focused on the other three subtypes. UBS93 demonstrated significantly higher accuracy than SCSubtype for the basal-like subtype (Fig. 3b).

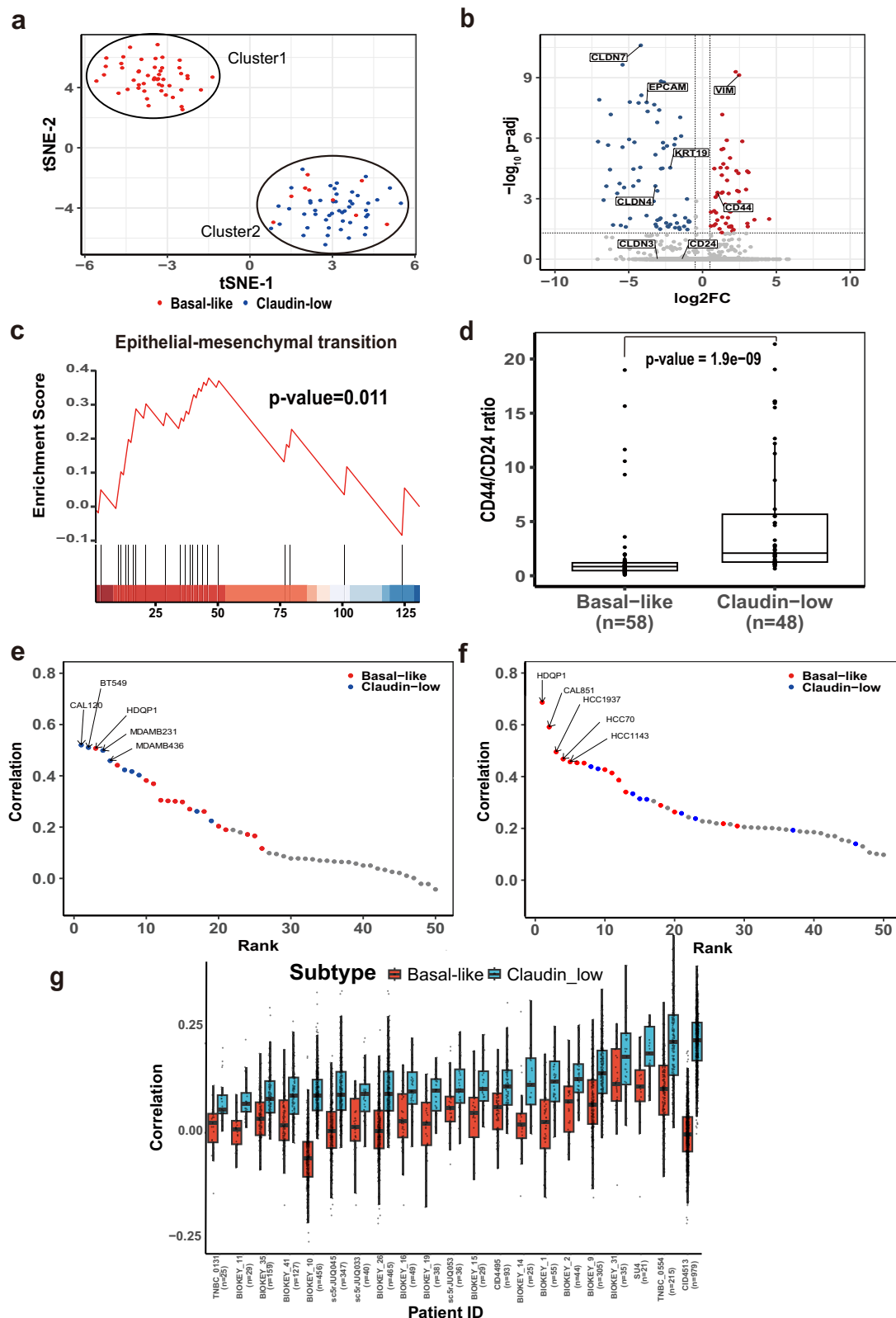
UBS93 uncovers intratumor heterogeneity in basal-like breast cancer

Prior to our study, Wu et al. identified basal-like cancer cells within human ER⁺ breast tumors, highlighting the presence of intratumor heterogeneity in ER⁺ breast cancer¹⁶. However, the intratumoral heterogeneity of TNBC remains largely underexplored. To investigate this, we re-examined UBS93 subtyping results derived from the scRNA-seq datasets of 12 established basal-like breast cancer cell lines. In 10 of these lines, over 80% of single cells were classified as basal-like, consistent with subtyping outcomes derived from bulk RNA-seq data. Strikingly, in the canonical basal-like cell line HDQP1, only 56.4% of cells were classified as basal-like, while the remaining cells were subtyped as claudin-low (Fig. 2c). These findings led us to hypothesize that, despite being classified as basal-like based on bulk transcriptomic profile, HDQP1 in fact consists of a mixture of basal-like and claudin-low cancer cell populations.

To test this hypothesis, we performed t-SNE visualization of the HDQP1 scRNA-seq data and observed two distinct clusters (Fig. 4a). UBS93 analysis revealed that cluster 1 consisted exclusively of basal-like cells, while 84.2% of the cells in cluster 2 were subtyped as claudin-low. Differential gene expression (DE) analysis (claudin-low vs. basal-like) suggested that the identified claudin-low cells exhibited canonical features of claudin-low breast cancer^{4–6,44,45}. These included: (1) reduced expression of cell adhesion genes (e.g., *CLDN4*, *CLDN7*, *KRT19*, and *EPCAM*) (Fig. 4b, Supplementary Fig. 6, and Supplementary Data 6); (2) elevated epithelial-to-mesenchymal transition (EMT) level, evidenced by up-regulation of *VIM* and the enrichment of EMT signature genes among up-regulated DE genes (Fig. 4c); and (3) heightened stemness, as indicated by increased *CD44* (marker of cancer stem cell) expression and higher *CD44/CD24* ratio^{4,45,46} (Fig. 4d).

To further validate these findings, we generated a pseudo-bulk transcriptomic profile using the HDQP1 cells subtyped as claudin-low and computed its transcriptomic correlation with breast cancer cell lines from the CCLE project³⁵. CAL120—a well-established claudin-low cell line—exhibited the highest correlation, with four of the top five most correlated lines being claudin-low (Fig. 4e). Conversely, similar analysis utilizing HDQP1 cells subtyped as basal-like yielded HDQP1 as the top correlated cell line, with all five top correlates being basal-like (Fig. 4f). To double confirm our finding, we replicated our analysis on an independent HDQP1 scRNA-seq data (GSE202771) and derived the same conclusion (Supplementary Figs. 7, 8, and Supplementary Data 7). Therefore, our computational analysis suggested that the basal-like cell line HDQP1 was an admixture of basal-like and claudin-low cell populations.

To our knowledge, such a “basal-like/claudin-low admixture” phenotype has been previously reported in mouse models following genetic modulation of the RAS signaling pathway⁵, with no prior evidence in the human basal-like breast cancer cell line HDQP1. To further investigate this phenotype at single-cell resolution in patients, we next applied the UBS93 classifier to the scRNA-seq data from 50 TNBC samples. These samples were then ranked based on the median transcriptomic correlation of claudin-low cells with the UBS93 claudin-low centroid (Fig. 4g and Supplementary Data 8). As expected, the metaplastic breast cancer patient CID4513 (from dataset GSE176078) ranked highest, which may not be surprising since previous studies have confirmed that metaplastic breast cancer exhibits characteristics of the claudin-low subtype^{47,48}. Intriguingly, the patient 0554 (from dataset GSE161529), who was classified as basal-like in the literature, ranked second⁴⁹. Of the cancer cells in this patient, 91.1% were subtyped as basal-like, while 7.9% were claudin-low. Similar to HDQP1, claudin-low cells tended to form a distinct cluster on the t-SNE plot (Fig. 5a), and meanwhile, exhibited canonical features of claudin-low



breast cancer in DE analysis (claudin-low vs. basal-like): down-regulation of adhesion genes, enrichment of EMT signatures, and elevated stemness markers (Fig. 5b–d, Supplementary Fig. 9, and Supplementary Data 9). In addition, we also performed transcriptomic correlation analysis between pseudo-bulk samples and CCLE breast cancer cell lines and double confirmed our results (Fig. 5e, f).

In summary, our integrative analyses reveal intratumoral heterogeneity in human Basal-like breast cancer, characterized by the co-existence of basal-like and claudin-low cancer cells. These findings highlight the capability of UBS93 to detect subtle but biologically meaningful substructures in breast cancer and underscore its potential utility for studying tumor evolution and plasticity.

Fig. 4 | HDQP1 is a cell line with “basal-like/claudin-low admixture” phenotype. **a** The t-SNE visualization of HDQP1 scRNA-seq data (GSE17364), with cells colored according to their UBS93 subtypes. **b** Differential gene expression analysis comparing claudin-low and basal-like cells of the HDQP1 cell line. Red and blue dots represent genes significantly up-regulated and down-regulated in claudin-low cells, respectively. **c** Gene set enrichment analysis (GSEA) of epithelial-mesenchymal transition (EMT) signature in the differentially expressed genes. The barcode beneath the heatmap represents the ranked position of EMT genes, from red (high log2FC) to blue (low log2FC). **d** Comparison of stemness between basal-like and claudin-low cells. Each dot represents a cell, the central line indicates the median value, and the bounds represent the 25th and 75th percentiles (interquartile range).

ELF3 knockdown induces a partial claudin-low phenotype in basal-like breast cancer

To investigate the underlying cause of the observed heterogeneity, we first examined the cellular origin of claudin-low cancer cells. Pommier et al. conducted large-scale computational analyses of bulk cancer genomics data and proposed three potential origins: direct malignant transformation from normal mammary stem cells⁵⁰, or trans-differentiation from either luminal or basal-like cancer cells⁵¹. However, robust single-cell-level evidence from patient samples has been lacking. The identification of basal-like/claudin-low admixture in our study provides a unique opportunity to revisit this question. Using copy number variation-based hierarchical clustering in patient 0554, we observed that claudin-low cancer cells were most closely related to basal-like cancer cells (Fig. 6a), supporting the hypothesis that claudin-low cancer cells may arise from a basal-like lineage.

To identify genes which may drive the basal-like/claudin-low transition, we intersected DE gene lists from both the HDQP1 cell line and patient 0554, revealing 18 commonly up-regulated and 37 commonly down-regulated genes (Supplementary Data 10). Among these, the transcription factor *ELF3* emerged as a gene of interest for two reasons (Supplementary Fig. 10). First, overexpression of *ELF3* in claudin-low cancer cell lines (MDA-MB-231 and BT549) inhibits the EMT process⁵²; second, *ELF3* promotes *CLDN4* expression in ovarian cancer and non-small cell lung cancer^{53,54}. These findings prompted us to test whether depletion of *ELF3* in basal-like cancer cells might induce a claudin-low phenotype.

We selected HCC70, a cell line with a transcriptomic profile closely resembling basal-like breast cancer⁵⁵, as our experimental model. scRNA-seq followed by UBS93 subtyping confirmed that 98.7% of untreated HCC70 cells were classified as basal-like. We then performed *ELF3* knockdown in HCC70 and repeated scRNA-seq to determine whether this perturbation could promote claudin-low features. UBS93 analysis showed that 98.1% of cells remained basal-like, suggesting that *ELF3* knockdown alone is insufficient to trigger a complete subtype switch (Supplementary Fig. 11a).

Despite failing to induce complete basal-like/claudin-low transition, we asked whether *ELF3* knockdown could induce partial claudin-low features. To account for variable knockdown efficiency, we selected the 100 cells with the lowest *ELF3* expression (sh-*ELF3* group) and the 100 cells with the highest *ELF3* expression (control group) for further analysis (Supplementary Fig. 11b). Interestingly, the sh-*ELF3* group exhibited significantly lower correlation with the basal-like centroid and higher correlation with the claudin-low centroid (Supplementary Fig. 11c, d) and DE analysis between these two groups identified 210 up-regulated and 79 down-regulated genes, respectively (Fig. 6b and Supplementary Data 11). As anticipated, *CLDN4* was significantly down-regulated ($p = 3.57 \times 10^{-14}$), as was the epithelial cell marker *KRT19* ($p = 3.59 \times 10^{-13}$). We performed hierarchical clustering analysis on the 79 down-regulated genes based on their expression pattern across CCLE breast cancer cell lines and derived three clusters. Interestingly, we found genes in cluster 1 exhibited claudin-low specific depletion (Fig. 6c). We curated a set of such genes (which we termed as CLSD genes) and gene set enrichment analysis revealed significant enrichment of CLSD genes among the down-regulated DE genes (Supplementary Fig. 11e).

Whiskers extend to 1.5 times the interquartile range. *P*-values were calculated using the two-sided Wilcoxon rank-sum test. **e, f** Ranking of CCLE breast cancer cell lines based on transcriptomic correlation with claudin-low (or basal-like) pseudo-bulk sample. The basal-like and claudin-low cell lines are colored red and blue, respectively. **g** Ranking of TNBC patients based on median correlation with claudin-low centroid. In the boxes, each dot represents an individual cell classified as claudin-low by the UBS93 classifier. The central line indicates the median value, and the bounds represent the 25th and 75th percentiles (interquartile range). The whiskers extend to 1.5 times the interquartile range. For comparison, correlation values with the basal-like centroid are also shown.

In addition to scRNA-seq, we also used qRT-PCR to assess how *ELF3* knockdown influenced the expression of *CLDN3*, *CLDN4*, *CLDN7*, *EPCAM*, and *VIM* (Fig. 6d). Consistent with scRNA-seq results, *CLDN4* was significantly down-regulated upon *ELF3* knockdown ($p = 0.002254$), and *CLDN7* also showed reduced expression, despite not meeting DE gene thresholds ($p = 0.018557$). No significant changes were observed in *CLDN3* and *EPCAM*. Unexpectedly, *VIM* expression was down-regulated ($p = 0.003883$), which contrasts with the reported role of *ELF3* as a suppressor of EMT and warrants further investigation. Replicated experiments double confirmed these findings (Supplementary Fig. 12).

Together, these results suggest that *ELF3* knockdown induces a partial claudin-low phenotype in basal-like cancer cells, as evidenced by the downregulation of CLSD genes, such as *CLDN4* and *CLDN7*.

Discussion

In this study, we introduced a novel classifier, UBS93, for the molecular subtyping of breast cancer. We validated its performance using both bulk and single-cell RNA sequencing data from human and mouse models. Notably, UBS93 outperformed classical methods, particularly in the identification of claudin-low subtypes.

Using UBS93, we uncovered intratumoral heterogeneity in human basal-like breast cancer at single-cell resolution, characterized by the co-existence of basal-like and claudin-low cancer cell populations within the same tumor. In our opinion, this finding carries three major implications. First, it advances our understanding of basal-like breast cancer biology. Previously, such heterogeneity had only been reported in mouse models following genetic modulation of the RAS signaling pathway⁵. Whether this phenotype also occurs in humans remained unclear, given the substantial molecular differences between human and murine models. Our study provides direct evidence that this heterogeneity exists in human basal-like breast tumors. Second, our findings offer compelling support for the hypothesis that claudin-low cancer cells could originate from a basal-like lineage, an idea previously proposed by Pommier et al. but lacking conclusive evidence. Third, our findings highlight the necessity of incorporating all intrinsic subtypes—including the claudin-low subtype—into molecular subtyping frameworks, particularly in single-cell analyses. Notably, although Wu et al. introduced the SCSubtype classifier, its omission of the claudin-low subtype led to the failure to detect this heterogeneity in human basal-like scRNA-seq datasets.

Beyond the advancements in computational subtyping, our research has raised several intriguing scientific questions that warrant further investigation.

First, besides *ELF3* knockdown, what additional genetic perturbations are needed to fully convert basal-like cancer cells into claudin-low cancer cells? While high EMT activity is a hallmark of claudin-low cells, *ELF3* knockdown alone did not significantly up-regulate mesenchymal markers, such as *VIM* or *FN1*. Interestingly, *ZEB1*—a key EMT regulator—was consistently up-regulated in the DE analyses comparing claudin-low and basal-like populations. This prompts the hypothesis that combined *ELF3* knockdown and *ZEB1* overexpression may be necessary to induce a complete claudin-low phenotype. In addition, since Fougner et al. claimed that the claudin-low phenotype may permeate breast tumors of various

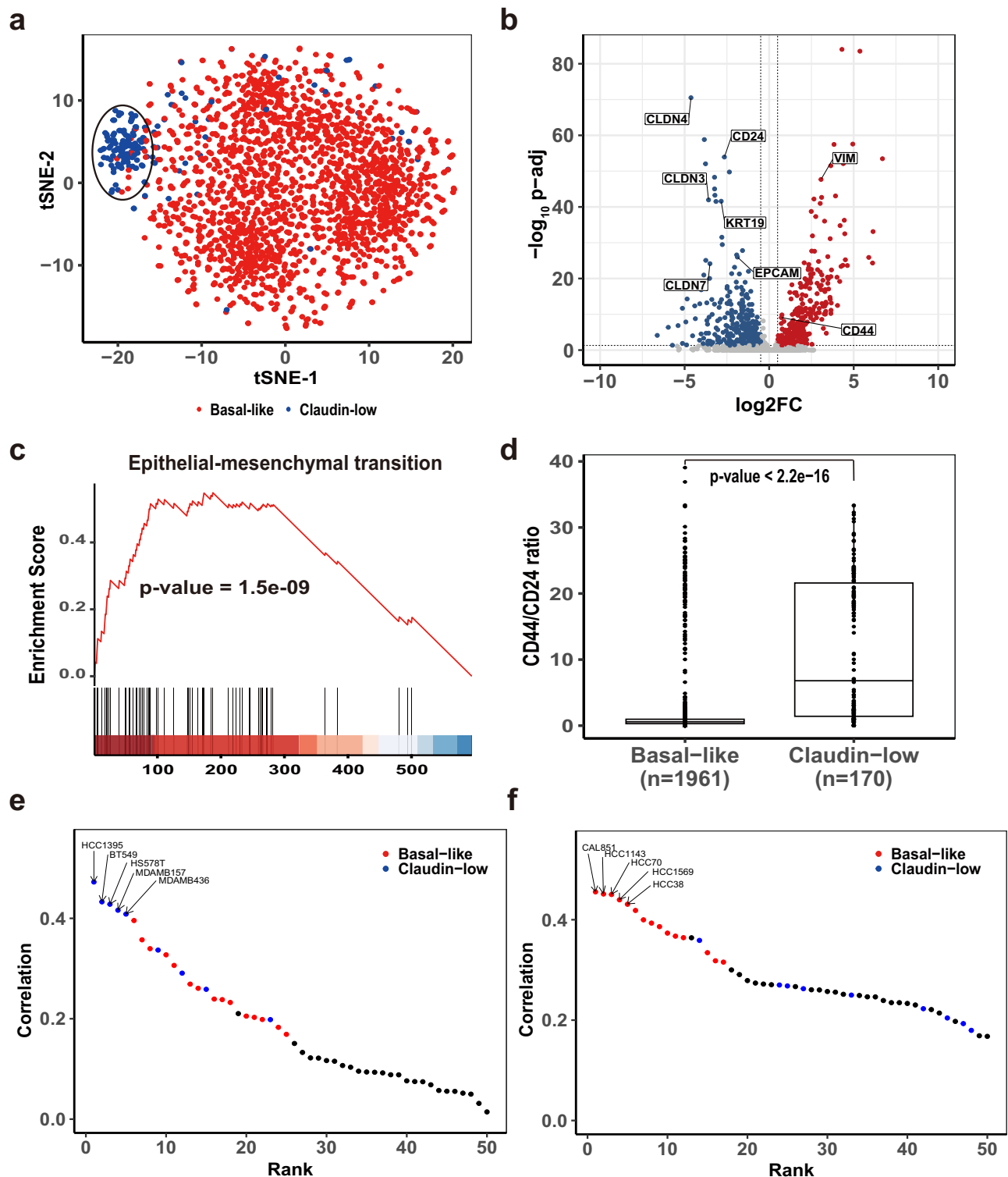


Fig. 5 | Identification of TNBC patient with “basal-like/claudin-low admixture” phenotype. **a** The t-SNE visualization of scRNA-seq data from TNBC patient 0554, with cancer cells colored according to their UBS93 subtypes. **b** Differential gene expression analysis comparing claudin-low and basal-like cells in TNBC patient 0554. Red and blue dots represent genes significantly up-regulated and down-regulated in claudin-low cells, respectively. **c** Gene set enrichment analysis (GSEA) of epithelial-mesenchymal transition (EMT) signature in the differentially expressed genes. The barcode beneath the heatmap represents the ranked position of EMT

genes, from red (high log2FC) to blue (low log2FC). **d** Comparison of stemness between basal-like and claudin-low cells. Each dot represents a cell, the central line indicates the median value, and the bounds represent the 25th and 75th percentiles (interquartile range). The whiskers extend to 1.5 times the interquartile range. P-values were calculated using the two-sided Wilcoxon rank-sum test. **e, f** Ranking of CCLE breast cancer cell lines based on transcriptomic correlation with the claudin-low (or basal-like) pseudo-bulk sample. The basal-like and claudin-low cell lines are colored red and blue, respectively.

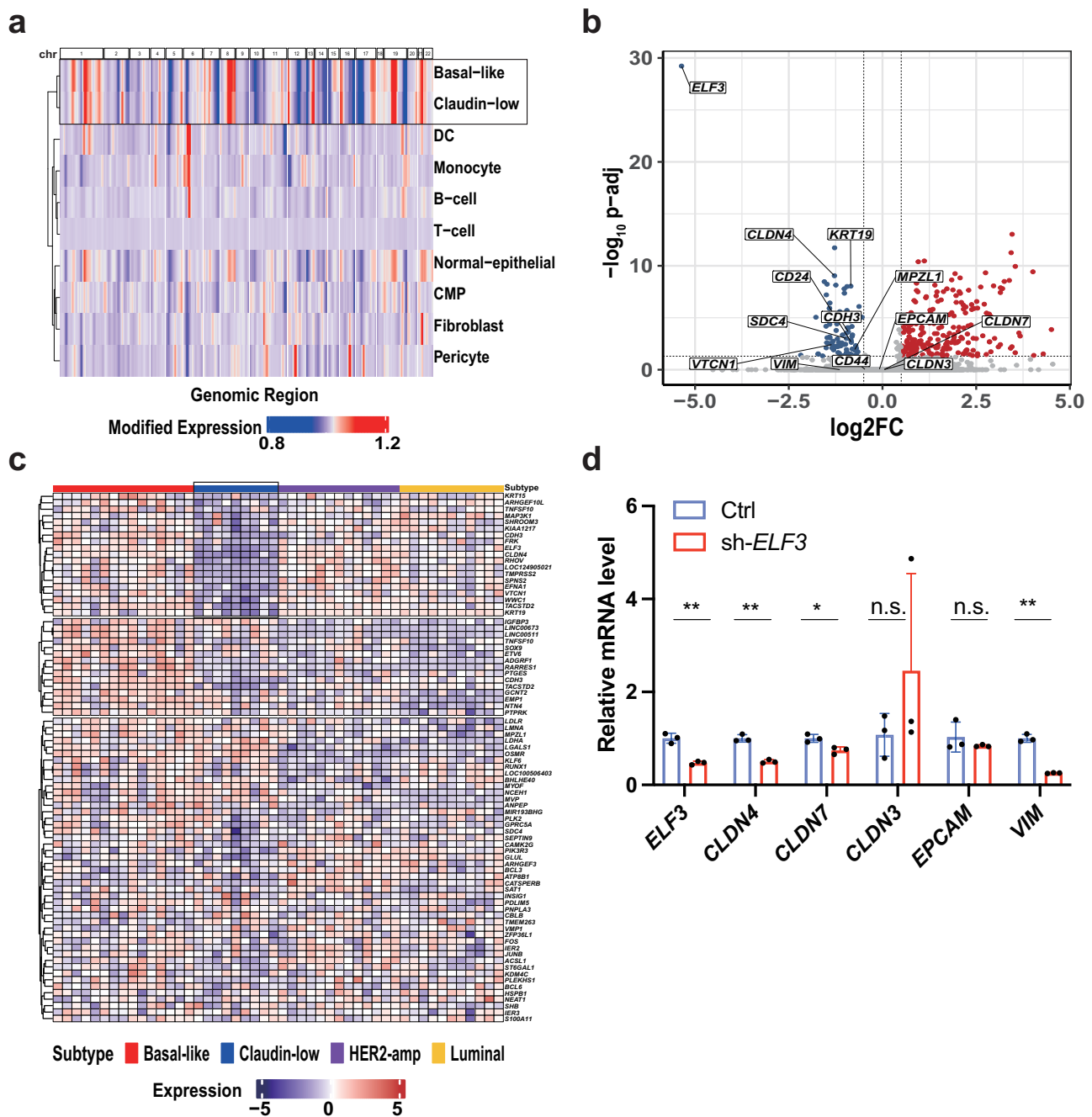


Fig. 6 | ELF3 knock-down induces a partial claudin-low phenotype in basal-like cancer. **a** Hierarchical clustering of different cell types (from TNBC patient 0554) based on average copy number variation profile. **b** Differential expression analysis comparing sh-ELF3 and control groups. Red and blue dots represent genes significantly up-regulated and down-regulated in the sh-ELF3 group, respectively. **c** Scaled expression of the 79 down-regulated genes across CLE breast cancer cell lines. Genes were hierarchically clustered by their expression patterns across cell

lines and displayed as three-row modules. The top module corresponds to the claudin-low-specific depletion (CLSD) genes. **d** Relative mRNA levels (quantified by RT-PCR) of ELF3, CLDN4, CLDN7, CLDN3, EPCAM, and VIM in control (blue) and sh-ELF3 (red) groups. Each dot represents a parallel sample. *P*-values were calculated using an unpaired two-sided *t*-test. n.s. (*p* > 0.05), * (*p* < 0.05), ** (*p* < 0.01), *** (*p* < 0.001).

subtypes⁶, therefore, it is also important to explore the transition mechanisms between other subtypes and claudin-low.

Second, is the *KRAS* mutation necessary for the generation of claudin-low cancer cells? Rädler et al. induced the tumor of “basal-like/claudin-low admixture” through activation of endogenous *KRAS* mutations (or exogenous *KRAS* expression). However, *KRAS* mutation is rare in human breast cancer. Using cBioPortal, we only identified six TCGA breast cancer patients carrying the *KRAS* mutation; in addition, the cell line HDQP1 does not carry the *KRAS* mutation either. These observations suggest that there

might be *KRAS*-independent mechanisms which may also give rise of claudin-low cancer cells.

Third, what are the clinical implications of the “basal-like/claudin-low admixture”? Given that distant metastasis is the primary cause of cancer-related mortality⁵⁶ and that claudin-low tumors are associated with greater aggressiveness, it is plausible that the presence of such a phenotype may correlate with poor prognosis in basal-like breast cancer. In the future, we will integrate single-cell RNA-seq data with clinical outcome information to test this hypothesis.

Finally, does the presence of “basal-like/claudin-low admixture” alter the tumor microenvironment? Fougner et al. reported that claudin-low tumors are characterized by high lymphocyte infiltration and elevated PD-L1 expression⁶. We therefore hypothesize that claudin-low cells within the tumors may shape distinct immune niches. To explore this, we will apply the UBS93 classifier to spatial transcriptomics datasets to identify regions enriched with claudin-low cells and compare their surrounding microenvironments to those of basal-like regions. This analysis may offer insights into immunotherapeutic vulnerabilities in this heterogeneous basal-like breast cancer.

In conclusion, our study introduces a robust and versatile tool for delineating intratumoral heterogeneity in breast cancer. As high-quality single-cell transcriptomic data continue to emerge across diverse tumor types, we anticipate that this strategy will be broadly applicable to other cancers. By enabling precise dissection of tumor subpopulations, our approach holds promise for uncovering novel biological mechanisms that drive tumor progression and therapeutic resistance.

Data availability

The authors declare that all data supporting the findings of this study are available within the article and its Supplementary information files or from the corresponding author upon reasonable request. The generated HCC70 scRNA-seq data are available at Zenodo⁵⁷.

Code availability

The code, which can be used to reproduce the figures, is available at GitHub⁵⁸.

Received: 15 August 2025; Accepted: 9 March 2026;

Published online: 13 April 2026

References

- Nolan, E., Lindeman, G. J. & Visvader, J. E. Deciphering breast cancer: from biology to the clinic. *Cell* **186**, 1708–1728 (2023).
- Parker, J. S. et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
- Herschkowitz, J. I. et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol.* **8**, R76 (2007).
- Prat, A. et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* **12**, R68 (2010).
- Rädler, P. D. et al. Highly metastatic claudin-low mammary cancers can originate from luminal epithelial cells. *Nat. Commun.* **12**, 3742 (2021).
- Fougner, C., Bergholtz, H., Norum, J. H. & Sørli, T. Re-definition of claudin-low as a breast cancer phenotype. *Nat. Commun.* **11**, 1787 (2020).
- Jordan, N. V. et al. HER2 expression identifies dynamic functional states within circulating breast cancer cells. *Nature* **537**, 102–106 (2016).
- Chung, W. et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat. Commun.* **8**, 15081 (2017).
- Azizi, E. et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308 (2018).
- Karaayvaz, M. et al. Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. *Nat. Commun.* **9**, 3588 (2018).
- Kim, C. et al. Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. *Cell* **173**, 879–893 (2018).
- Savas, P. et al. Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. *Nat. Med.* **24**, 986–993 (2018).
- Gulati, G. S. et al. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* **367**, 405–411 (2020).
- Bassez, A. et al. A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. *Nat. Med.* **27**, 820–832 (2021).
- Pal, B. et al. A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. *The EMBO journal* **40**, e107333 (2021).
- Wu, S. Z. et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat. Genet.* **53**, 1334–1347 (2021).
- Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
- Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, p11 (2013).
- Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
- Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J. Integr. Biol.* **16**, 284–287 (2012).
- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinform.* **12**, 323 (2011).
- Frankish, A. et al. GENCODE: reference annotation for the human and mouse genomes in 2023. *Nucleic Acids Res.* **51**, D942–d949 (2023).
- Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
- Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
- Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
- Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
- Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
- The ggplot2 package <https://cran.r-project.org/web/packages/ggplot2/index.html>.
- The pheatmap package <https://github.com/raivokolde/pheatmap>.
- Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
- Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
- Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7 (2013).
- Barretina, J. et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
- Greer, Y. E. et al. Mitochondrial matrix protease ClpP agonists inhibit cancer stem cell function in breast cancer cells by disrupting mitochondrial homeostasis. *Cancer Res. Commun.* **2**, 1144–1161 (2022).
- Daemen, A. et al. Modeling precision treatment of breast cancer. *Genome Biol.* **14**, R110 (2013).
- Marcotte, R. et al. Functional genomic landscape of human breast cancer drivers, vulnerabilities, and resistance. *Cell* **164**, 293–309 (2016).
- Weinstein, J. N. et al. The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat. Genet.* **45**, 1113–1120 (2013).

40. Gambardella, G. et al. A single-cell analysis of breast cancer cell lines to study tumour heterogeneity and drug response. *Nat. Commun.* **13**, 1714 (2022).
41. Jovanović, B. et al. Heterogeneity and transcriptional drivers of triple-negative breast cancer. *Cell Rep.* **42**, 113564 (2023).
42. Hollern, D. P. et al. B cells and T follicular helper cells mediate response to checkpoint inhibitors in high mutation burden mouse models of breast cancer. *Cell* **179**, 1191–1206 (2019).
43. Gendoo, D. M. et al. Genefu: an R/bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics* **32**, 1097–1099 (2016).
44. Pan, C. et al. Research progress of claudin-low breast cancer. *Front. Oncol.* **13**, 1226118 (2023).
45. Dias, K. et al. Claudin-low breast cancer; clinical & pathological characteristics. *PLoS ONE* **12**, e0168669 (2017).
46. Li, W. et al. Unraveling the roles of CD44/CD24 and ALDH1 as cancer stem cell markers in tumorigenesis and metastasis. *Sci. Rep.* **7**, 13856 (2017).
47. Reddy, T. P. et al. A comprehensive overview of metaplastic breast cancer: clinical features and molecular aberrations. *Breast Cancer Res.* **22**, 121 (2020).
48. Weigelt, B. et al. Metaplastic breast carcinomas display genomic and transcriptomic heterogeneity [corrected]. *Mod. Pathol.* **28**, 340–351 (2015).
49. Pal, B. et al. A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. *EMBO J.* **40**, e107333 (2021).
50. Creighton, C. J. et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc. Natl. Acad. Sci. USA* **106**, 13820–13825 (2009).
51. Pommier, R. M. et al. Comprehensive characterization of claudin-low breast tumors reflects the impact of the cell-of-origin on cancer evolution. *Nat. Commun.* **11**, 3431 (2020).
52. Park, S. & Kwon, Y. The role of ELF3 in metastasis of triple negative breast cancer. *Ann. Oncol.* **28**, x34 (2017).
53. Kuang, L. & Li, L. E74-like factor 3 suppresses microRNA-485-5p transcription to trigger growth and metastasis of ovarian cancer cells with the involvement of CLDN4/Wnt/ β -catenin axis. *Saudi J. Biol. Sci.* **28**, 4137–4146 (2021).
54. Zhang, X. et al. Single-cell transcriptomics reveal metastatic CLDN4⁺ cancer cells underlying the recurrence of malignant pleural effusion in patients with advanced non-small-cell lung cancer. *Clin. Transl. Med.* **14**, e1649 (2024).
55. Liu, K. et al. Evaluating cell lines as models for metastatic breast cancer through integrative analysis of genomic data. *Nat. Commun.* **10**, 2138 (2019).
56. Park, M. et al. Breast cancer metastasis: mechanisms and therapeutic implications. *Int. J. Mol. Sci.* **23**, 6806 (2022).
57. Liu, K. HCC70 scRNA-seq data. Zenodo <https://doi.org/10.5281/zenodo.15482443> (2025).
58. bioklab. UBS93. GitHub <https://github.com/bioklab/UBS93> (2025).

Acknowledgements

We thank Dr. Xiaohui Lin for helping us revise the manuscript. The research was supported by the National Natural Science Foundation of China (Fund

32370715, 82330108), the Science Fund for Distinguished Oversea Young Scholars of Shandong Province (2023HWYQ-015), the Taishan Young Scholar Program of Shandong Province (tsqn202312020), and the Cheeloo Young Scholar Program of Shandong University. The content is solely the responsibility of the authors and does not necessarily represent the official views of sponsors.

Author contributions

J.L., K.L. conceived the study. J.L. performed the majority of computational analysis, Y.G., W.X.Y. performed all the experimental assays, K.L., S.Z.Z., and W.X.Y. supervised the study. S.R.Z., K.L. revised the manuscript. All authors contributed to writing, reviewing, and editing the manuscript and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains Supplementary material available at <https://doi.org/10.1038/s43856-026-01548-z>.

Correspondence and requests for materials should be addressed to Shisong Zhang, Weixing Yu or Ke Liu.

Peer review information *Communications Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026