BMC Medical Genomics

RESEARCH

Open Access

Multi-omics single-cell dissection of malignant epithelial heterogeneity identifies GJB2 as an EMT-driving biomarker in triple-negative breast cancer

Hongjie Xu[1,2,3†], Chengzhi Zhang[1,2,3†], Xinyi Zhang[1,2], Shuyi Zhang[1,2] and Guangjun Zhou[1,2*]

Abstract

Background Tumor heterogeneity poses a major challenge to elucidating the mechanisms underlying tumor progression and metastasis. However, the cellular heterogeneity of triple-negative breast cancer (TNBC) remains incompletely understood.

Methods We performed an integrated analysis of publicly available single-cell RNA sequencing (scRNA-seq) datasets from TNBC to identify distinct malignant cell subtypes. Multiple analytical approaches were then applied to comprehensively characterize the metastasis-related subtype. Subsequently, we validated GJB2 as a biomarker of this subtype and confirmed its association with epithelial–mesenchymal transition (EMT) in TNBC cell lines.

Results Based on six scRNA-seq datasets from GSE180286 and GSE161529, we constructed a comprehensive landscape of malignant-cell heterogeneity in TNBC and identified three malignant cell subtypes: an EMT-Subtype, a secretory-like subtype (Sec-like Subtype), and a metabolic subtype (Metab-Subtype). We further identified GJB2 as a marker of the EMT-Subtype. Enrichment analyses indicated that GJB2 was closely associated with the TGF-β signaling pathway, and this finding was further supported by in vitro functional assays.

Conclusion By classifying malignant TNBC cells into three subtypes, we identified GJB2 as a marker of the EMT-Subtype. Mechanistically, GJB2 may promote EMT through the TGF-β signaling pathway and therefore represent a potential therapeutic target in TNBC. Targeting the GJB2–TGF-β axis may provide a promising strategy for suppressing the highly metastatic behavior of TNBC.

Keywords Triple-negative breast cancer, GJB2, Epithelial–mesenchymal transition, Single-cell RNA sequencing, Biomarker

†Hongjie Xu and Chengzhi Zhang contributed equally to this work.

*Correspondence:
Guangjun Zhou
710020220011@xzhmu.edu.cn

[1]Department of General Surgery, The Yancheng Clinical College of Xuzhou Medical University, Yancheng, China
[2]Department of General Surgery, The First People's Hospital of Yancheng, Yancheng, China
[3]Yancheng Key Laboratory of Molecular Epigenetics, Yancheng Medical Research Center of Nanjing University Medical School, Yancheng, China

Introduction

TNBC, defined by the absence of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), represents a biologically and clinically aggressive subtype of breast cancer. TNBC accounts for approximately 10–20% of all breast cancer cases and is closely associated with a markedly increased risk of distant metastasis, recurrence, and mortality [1, 2].

Tumor metastasis is a dynamic and complex biological process that involves not only the progressive acquisition of invasive and migratory capabilities by tumor cells but also extensive interactions with components of the tumor microenvironment. Increasing evidence indicates that immune cells and stromal cells within the tumor microenvironment play essential roles in shaping tumor progression, metastatic dissemination, and therapeutic responses. In particular, immune components such as tumor-associated macrophages, T lymphocytes, and other immune regulators can actively influence tumor evolution by modulating inflammation, immune evasion, and metastatic niche formation. Recent studies have highlighted the importance of the tumor immune microenvironment in determining tumor heterogeneity and clinical outcomes in breast cancer and other malignancies [3–5]. In parallel, a variety of signaling pathways, including EMT-related transcriptional programs, extracellular matrix remodeling, and pro-metastatic signaling pathways such as TGF-β, PI3K/AKT, and NF-κB, have been identified as key drivers of tumor invasion and metastasis [6, 7]. However, most of these insights are derived from bulk transcriptomic analyses, which average signals across heterogeneous cell populations and may obscure the contributions of highly invasive malignant cell subsets [8].

GJB2, also known as gap junction beta-2, encodes connexin 26 (Cx26), a member of the β-connexin family. It forms intercellular gap junction channels that mediate the direct exchange of ions, metabolites, and second messengers between adjacent cells. GJB2 is located on chromosome 13q11–q12 and encodes a transmembrane protein of approximately 26 kDa with four transmembrane domains. This protein oligomerizes to form connexons, which assemble at the plasma membrane into gap junction plaques [9]. GJB2 is one of the most well-established causative genes in hereditary hearing loss, particularly DFNB1-associated deafness [10]. Recent studies have further suggested that dysregulation of GJB2 expression or function is closely associated with tumor initiation and progression, and that GJB2 may exert either tumor-suppressive or oncogenic effects in a context-dependent manner [11].

ScRNA-seq is a rapidly evolving high-throughput omics technology that enables the profiling of genomic, transcriptomic, and even epigenomic information at single-cell resolution [12]. Compared with conventional bulk sequencing, scRNA-seq does not average signals across large cell populations; instead, it labels and sequences thousands to tens of thousands of individual cells, allowing a refined characterization of transcriptional heterogeneity between cells [13, 14].

Previous studies have proposed several transcriptome-based classification frameworks for TNBC, including claudin-low-like [15], immunomodulatory [16], and immune-suppressed [17] categories. However, these subtype classification systems were largely derived from bulk transcriptomic profiles and may not fully capture malignant-cell-state heterogeneity at single-cell resolution.

In this study, we integrated multiple scRNA-seq datasets of TNBC to systematically characterize the heterogeneity of malignant epithelial cells at single-cell resolution. Moreover, our findings identify a potential therapeutic target for limiting TNBC metastasis.

Results

Identification of three subtypes within the TNBC malignant-cell heterogeneity landscape

To dissect the heterogeneity of TNBC tumor cells from a single-cell perspective, we analyzed six TNBC samples from the GSE161529 and GSE180286 datasets (Fig. 1A). After stringent quality control, batch effect correction, selection of highly variable genes, and principal component analysis (Supplementary Fig. 1A–F), the cells were classified into seven major lineages, including epithelial cells, T cells, macrophages, B cells, fibroblasts, endothelial cells, and mast cells (Fig. 1B).

Because breast carcinoma originates from epithelial cells, malignant cells were presumed to be of epithelial origin [18]. We employed inferCNV to infer chromosomal copy number alterations and thereby distinguish malignant from non-malignant cells, ultimately identifying 16,512 malignant cells (Fig. 1C, Table S1). These malignant cells were then subjected to a second round of PCA and UMAP dimensionality reduction, followed by unsupervised clustering at a resolution of 0.3, which yielded nine tumor cell clusters. To assess the relationships among these clusters, we first calculated the average expression profile of each cluster, selected the 50 genes with the highest standard deviation within each cluster, and took the union of these genes as a feature set. Based on this feature set, we computed pairwise Spearman correlation coefficients [19] between clusters and subsequently merged highly correlated clusters into three subtypes: Subtype 1 (clusters 0, 1, 2, 3, and 7), Subtype 2 (clusters 4 and 8), and Subtype 3 (clusters 5 and 6) (Fig. 1D-F).

We further performed non-negative matrix factorization (NMF) [20] on the malignant cells. Specifically,

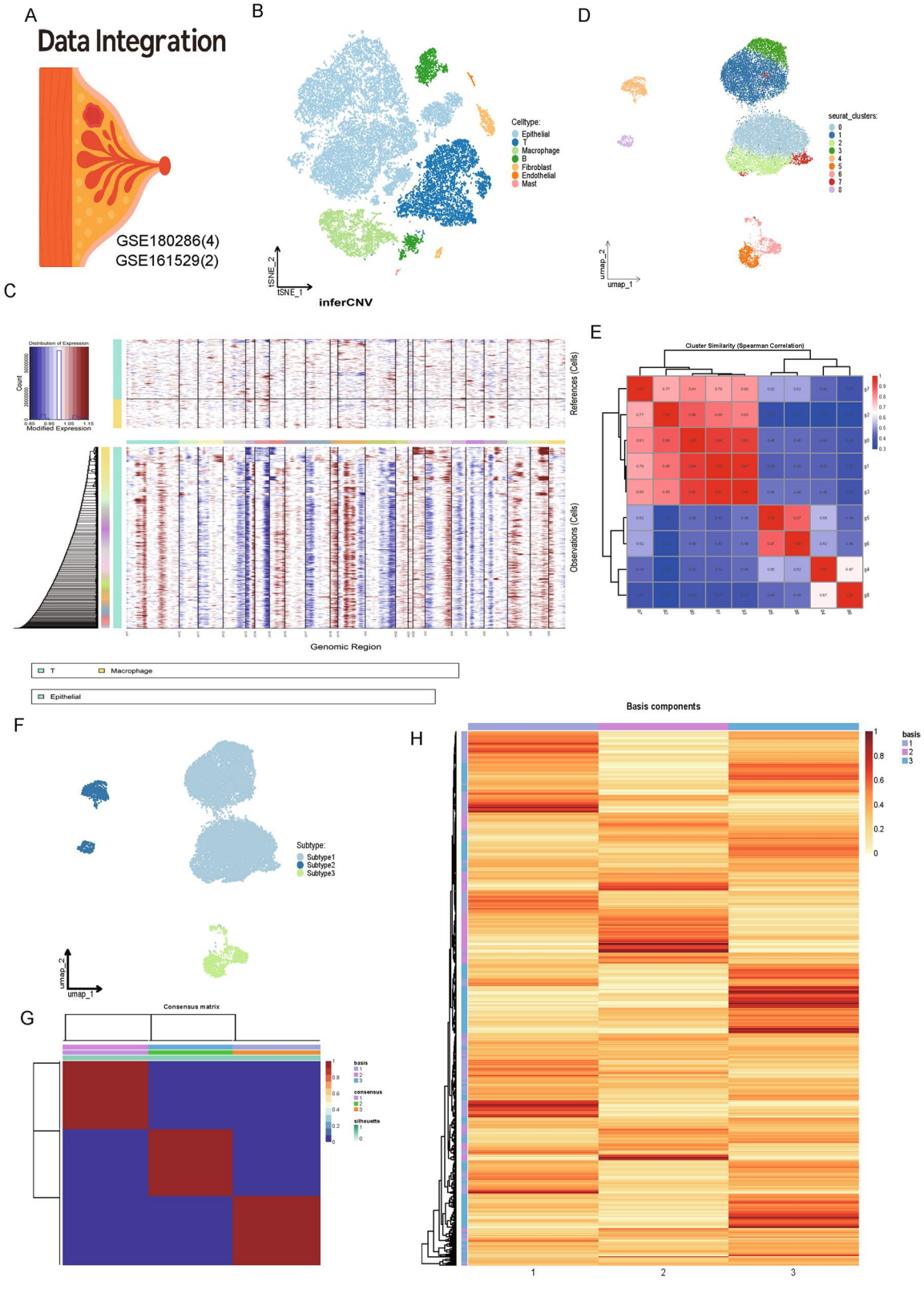

Fig. 1 Single-cell transcriptome analysis of TNBC reveals three malignant cell subtypes. **A** Overview of the six TNBC scRNA-seq datasets included in this study. **B** tSNE plot showing major cell types after data integration. **C** CNV analysis of epithelial cells distinguishing malignant from non-malignant cells. **D** UMAP plot of malignant epithelial cells. **E** Spearman correlation matrix of the average expression profiles of the nine malignant cell clusters.(F) UMAP plot of malignant cells colored by the three merged subtypes. **G** Evaluation of NMF factorization rank showing that k = 3 was selected as the optimal number of programs. **H** Heatmap of NMF-derived gene expression modules across malignant cells

50 malignant cells were randomly sampled from each cluster to construct an expression matrix, and candidate factorization ranks (k = 2–10) were evaluated based on cophenetic correlation coefficients and consensus matrices, from which k = 3 was selected as the optimal rank (Fig. 1G). Consistently, NMF also partitioned the malignant cells into three groups. Notably, these three NMF-derived groups showed a high concordance (79.2%) with the three subtypes defined by correlation-based clustering, and the heatmap (Fig. 1H) illustrated that NMF decomposed the malignant cell expression profiles into three stable gene modules. We next assessed the robustness of the NMF-based clustering by repeated resampling. Across 10 independent runs, the median concordance was 76.3%, further supporting the validity of the three-subtype classification.

Marker genes and functional characterization of the three TNBC malignant cell subtypes

The molecular characteristics of the three TNBC tumor cell subtypes were next examined. Using the Seurat *FindAllMarkers* function, we performed differential expression analysis by comparing each subtype with the remaining malignant cells and thereby identified subtype-specific differentially expressed genes (DEGs) (Fig. 2A). The top DEGs in Subtype 1 were predominantly associated with invasion and metastasis, including S100A2 [21] and MMP7 [22]. In contrast, the leading DEGs in Subtype 2 were mainly associated with glandular differentiation and secretory functions, such as SUSD3 [23] and STC2 [24]. Subtype 3 was enriched for metabolism-related genes, with top DEGs including MFSD2A [25] and ACOX2 [26].

To further delineate the biological characteristics of the three malignant subtypes, we performed GO enrichment analysis [27] based on the top 50 DEGs (Table S2) of each subtype and observed markedly distinct enrichment patterns in biological processes. Subtype 1 was significantly enriched in invasion- and metastasis-related pathways. Consistently, GSVA [28] revealed higher pathway activity scores for EMT and TGF-β signaling in this subtype; therefore, we designated it as the "EMT-Subtype". DEGs in Subtype 2 were mainly enriched in glandular development and secretory functions, and GSVA further revealed increased activity of secretion-related pathways, leading us to define this cluster as the Sec-like Subtype. In contrast, Subtype 3 displayed strong enrichment in metabolic and biosynthetic processes, with GSVA indicating higher scores in energy metabolism and mTORC1-related pathways; accordingly, we termed it the "Metab-Subtype" (Fig. 2B–C).

DE-ERGs and molecular characterization in TNBC patients

To investigate the dysregulation of EMT-related genes in TNBC, we first retrieved 200 EMT-related genes (ERGs) from the MSigDB database [19] (Table S3). Using expression data from 117 TNBC samples and 113 normal breast tissue samples in the TCGA cohort, we performed differential expression analysis and identified 6,056 DEGs ($|\log2FC| > 1$, $p < 0.05$). Among the 200 EMT genes, 99 were differentially expressed in the TCGA cohort (Fig. 3A), of which 58 were upregulated and 41 were downregulated in TNBC (Table S4), indicating that EMT-related programs are profoundly dysregulated in TNBC. To further expand the EMT-related gene set, we used these 99 differentially expressed EMT genes as seed genes and performed Pearson correlation analysis, identifying 7,389 significantly associated EMT-related genes (DE-ERGs) ($|R| > 0.5$, $p < 0.001$; Table S5). Among these, 133 genes were themselves significantly differentially expressed, including 75 upregulated and 58 downregulated genes in tumor tissues ($|\log2FC| > 4$, $p < 0.05$; Table S6).

Identification of GJB2 as an EMT-associated molecular marker in TNBC

To identify potential molecular targets that could effectively inhibit TNBC metastasis, we focused on the top-ranked DEGs within the EMT subtype. Integrating our previous analyses, we found that GJB2 emerged as a particularly noteworthy gene: it ranked among the leading DEGs in the EMT subtype and was also one of the significantly differentially expressed EMT-related genes (DE-ERGs) ($|\log2FC| > 4$, $p < 0.05$). This finding was further supported by analysis of the GEPIA2 database (Fig. 3B). ROC curve analysis based on the TCGA cohort demonstrated that GJB2 exhibited strong diagnostic performance for distinguishing TNBC tissues from normal breast tissues, with an AUC of 0.943 (Fig. 3C). In addition, Kaplan–Meier Plotter analysis (https://kmplot.com/analysis/) showed that elevated GJB2 expression was significantly associated with unfavorable overall survival (OS) and distant metastasis-free survival (DMFS) in patients with TNBC (Fig. 3D–E). These results indicate that GJB2 may have potential diagnostic and prognostic value in TNBC.

The relationship between GJB2 and EMT-related signaling in TNBC was further clarified by evaluating the association between GJB2 expression and Hallmark pathway activity in TNBC samples from the TCGA-BRCA cohort. Correlation analysis revealed that GJB2 was positively associated with the EMT Hallmark and multiple EMT-related signaling pathways, with several pathways exhibiting a significant positive Pearson correlation with GJB2 (Fig. 3F). Moreover, TNBC samples were stratified into GJB2-high and GJB2-low groups based on its

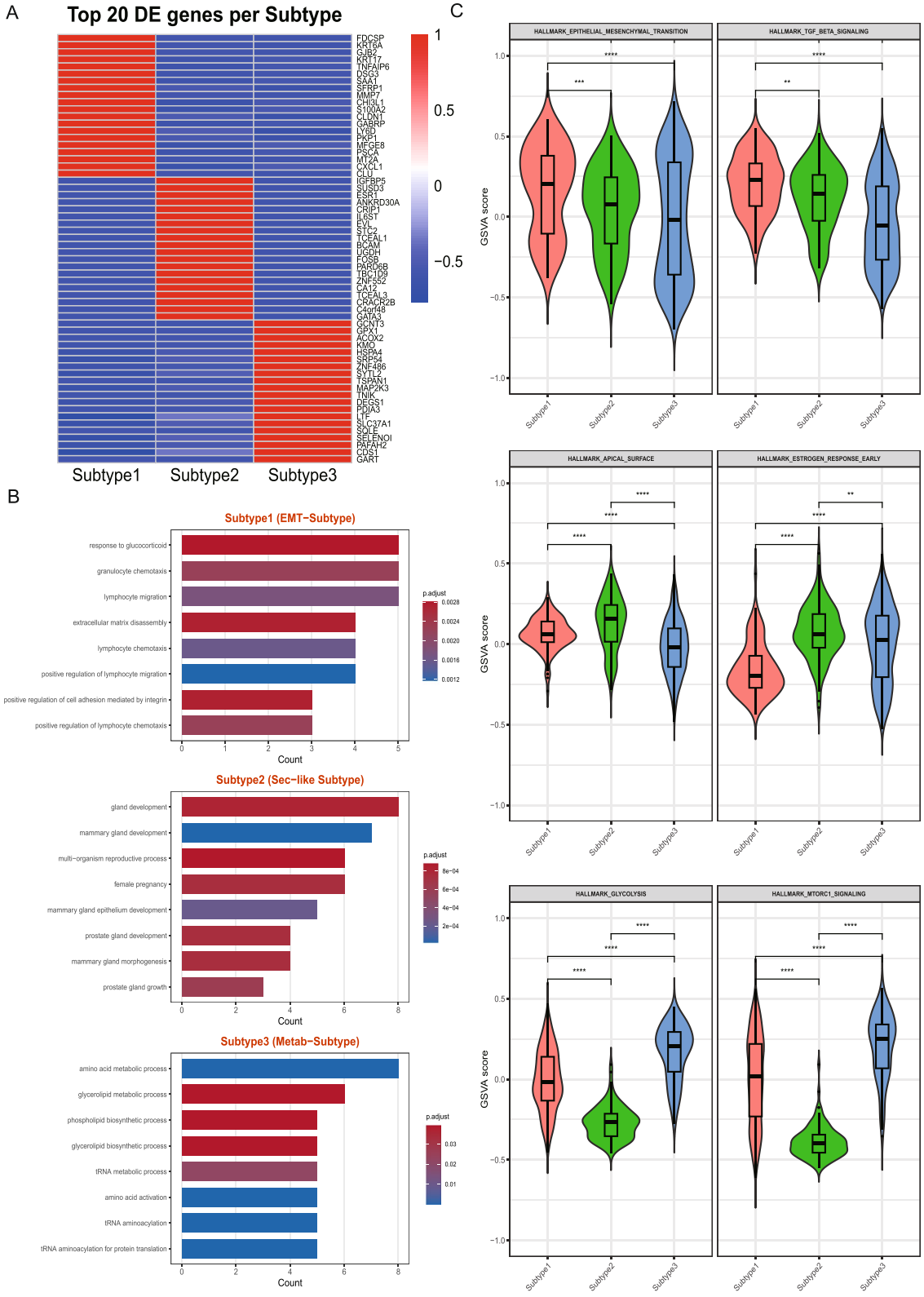

Fig. 2 Molecular signatures and pathway enrichment of TNBC tumor cell subtypes. **A** Heatmap of subtype-specific marker genes identified by differential expression analysis. **B** GO biological process enrichment analysis of the top 50 marker genes for each subtype. **C** Heatmap of GSVA scores for representative pathways across the three malignant cell subtypes

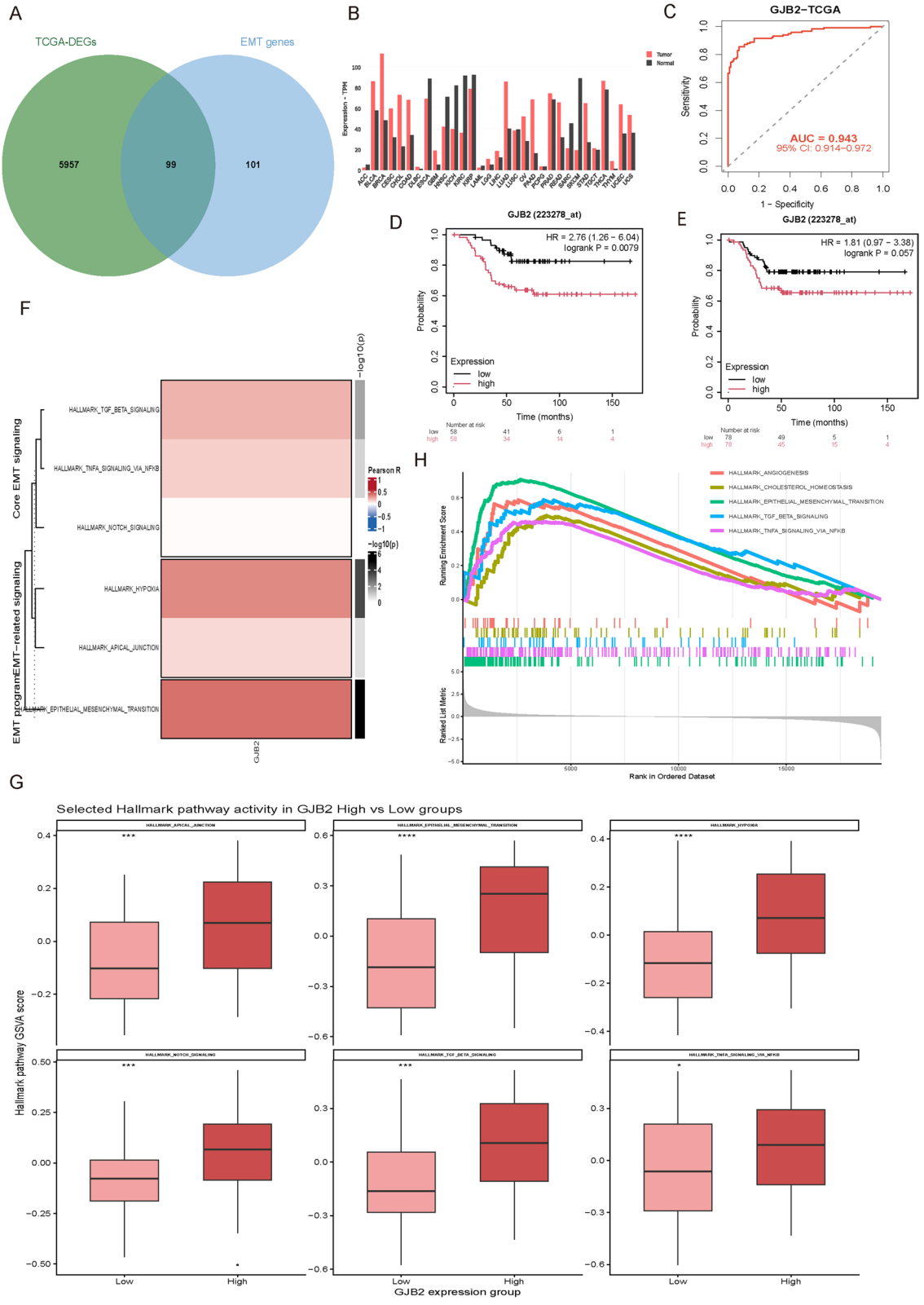

Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 GJB2 is associated with diagnostic/prognostic value and EMT-related signaling in TNBC. **A** Differential expression analysis of Hallmark EMT-related genes between TCGA TNBC tumors and normal breast tissues, identifying significantly dysregulated EMT-related genes. **B** Boxplot showing GJB2 mRNA expression in TNBC and normal breast tissues based on the GEPIA2 database. **C** Receiver operating characteristic (ROC) curve evaluating the diagnostic performance of GJB2 for distinguishing TNBC tissues from normal breast tissues in the TCGA cohort. **D** Kaplan–Meier analysis of overall survival (OS) according to GJB2 expression in patients with TNBC. **E** Kaplan–Meier analysis of distant metastasis-free survival (DMFS) according to GJB2 expression in patients with TNBC. **F** Correlation analysis between GJB2 expression and Hallmark pathway activity scores in TCGA TNBC samples. **G** Comparison of GSVA scores for selected Hallmark pathways between GJB2-high and GJB2-low TNBC groups. **H** Single-gene GSEA of GJB2 in TCGA TNBC samples, showing enrichment of EMT- and TGF-β-related Hallmark pathways

expression, and GSVA scores of selected Hallmark pathways were compared between these two groups (Fig. 3G). Collectively, these findings support the notion that GJB2 is an EMT-associated gene in TNBC.

GJB2 is best known for its pivotal role in autosomal recessive hereditary deafness and has been extensively studied in the field of genetic hearing loss. However, systematic investigations of GJB2 in breast cancer are still lacking, and reports in other tumor types remain relatively limited.

Single-gene GSEA was performed on bulk transcriptomic data from 117 TNBC samples in the TCGA cohort, and the top five Hallmark gene sets with the most significant positive enrichment (NES > 0) were displayed (Fig. 3H). In this analysis, high GJB2 expression was significantly enriched in the EMT and TGF-β Hallmark pathways. Together with the strong positive correlation between GJB2 expression and TGF-β pathway activity in TCGA-BRCA TNBC samples, these findings suggest that GJB2 may mediate EMT in TNBC through the TGF-β pathway and serve as a potential EMT-related biomarker.

GJB2 promotes EMT through the TGF-β/Smad pathway

We next experimentally validated the role of GJB2 in TGF-β–driven EMT in breast cancer cells. To evaluate the subtype specificity of GJB2 expression, we compared GJB2 levels between TNBC and non-TNBC breast cancer cell lines. Western blot analysis showed that GJB2 expression was consistently higher in TNBC cell lines, whereas non-TNBC breast cancer cell lines exhibited relatively low GJB2 expression (Fig. 4A). Based on these findings, MDA-MB-231 and MDA-MB-436 cells, which displayed relatively high endogenous GJB2 expression, were selected for subsequent loss-of-function experiments. GJB2 knockdown increased E-cadherin expression while decreasing N-cadherin, vimentin, Snail, and Slug, suggesting a mesenchymal-to-epithelial phenotypic reversal (Fig. 4B). In addition, GJB2 knockdown markedly reduced the ratio of phosphorylated Smad2/3 to total Smad2/3, indicating attenuation of TGF-β/Smad signaling. Treatment with exogenous TGF-β1 partially restored Smad2/3 phosphorylation, suggesting that GJB2 promotes EMT through activation of the TGF-β pathway (Fig. 4C). Consistent with these findings, Transwell migration assays (Fig. 4D and F) and wound-healing assays (Fig. 4E and G) further demonstrated that GJB2 knockdown significantly impaired migratory capacity, whereas exogenous TGF-β1 treatment partially rescued this effect.

Discussion

In this study, we used a multi-omics strategy to systematically dissect the lineage heterogeneity of malignant epithelial cells in TNBC and to identify potentially actionable molecular targets. By integrating six publicly available scRNA-seq datasets of TNBC, inferring CNV profiles in epithelial cells, and applying NMF-based unsupervised clustering, we delineated three malignant cell subtypes with distinct transcriptional and functional programs: an EMT-Subtype, a Sec-like Subtype, and a Metab-Subtype. Focusing on the EMT-Subtype, we identified GJB2 as a subtype-specific marker gene that was markedly upregulated in TNBC tissues relative to normal breast tissues, strongly correlated with multiple EMT-related gene programs, and among the most significantly dysregulated EMT-related genes (ERGs). Concordant evidence from single-gene GSEA, GSVA, and in vitro functional assays in TNBC cell lines demonstrated that high GJB2 expression is closely associated with activation of EMT and the TGF-β/Smad signaling pathway, as well as enhanced migratory capacity, supporting the role of GJB2 as an EMT-related biomarker and a potential therapeutic target in TNBC.

Compared with conventional bulk transcriptomic profiling, scRNA-seq enables the resolution of cellular subpopulations with distinct lineage origins, differentiation states, and functional phenotypes at the single-cell level. In multiple solid tumors, scRNA-seq studies have revealed EMT-like cells, stem/progenitor-like populations, and immune-evasive cell subsets, and have proposed numerous novel molecular markers and therapeutic targets on this basis. For example, Sun et al. used single-cell RNA sequencing to compare endothelial cells from breast cancer and normal breast tissue, and identified a gene expression signature specific to tumor-associated endothelial cells, which may serve as a potential vascular-targeting biomarker [29]. In HR + /HER2 – metastatic breast cancer, Luo et al. compared scRNA-seq profiles of tumor cells before and after CDK4/6 inhibitor treatment and identified a set of molecular biomarkers capable of predicting late progression, thereby providing

potential targets for individualized therapy and resistance monitoring [30].

GJB2 encodes connexin 26, a member of the gap junction protein family, and is one of the best-established causative genes in autosomal recessive non-syndromic sensorineural hearing loss. Beyond non-syndromic deafness, GJB2 abnormalities have also been implicated in several syndromes characterized by skin hyperkeratosis combined with hearing loss [31]. Increasing evidence indicates that GJB2 and GJB2-mediated gap junctions exert context-dependent dual roles in tumorigenesis and tumor progression: in some tumor types, GJB2 appears to have tumor-suppressive properties [32], whereas in others, such as lung adenocarcinoma [33] and colorectal cancer [34], high GJB2 expression is associated with poor prognosis. In breast tissue, GJB2 is markedly upregulated during pregnancy and lactation [35]. Early studies reported hypermethylation of the GJB2 promoter region, suggesting a potential tumor-suppressive role [36]; however, subsequent work showed that in a subset of breast cancers, particularly those with aggressive phenotypes, GJB2 mRNA and protein levels are significantly elevated and correlate with lymph node metastasis and advanced clinical stage [37].

It should be noted that this study has several limitations. First, although we demonstrated at the cellular level that GJB2 promotes EMT and migration through the TGF-β/Smad pathway, in vivo evidence from animal models is still lacking to evaluate the efficacy and safety of targeting GJB2 to inhibit distant TNBC metastasis. Second, it will be necessary to validate the association between GJB2 expression, metastatic risk, and clinical outcomes in larger, independent TNBC cohorts with detailed metastasis and follow-up information, and to systematically assess its sensitivity and specificity as a diagnostic and prognostic biomarker.

Conclusion

Overall, we systematically delineated the heterogeneity of malignant epithelial cells in TNBC and identified GJB2 as a marker of the EMT-Subtype. Elevated GJB2 expression was closely associated with EMT and may represent a potential therapeutic target in TNBC.

Materials and methods

Single-cell data collection and basic analysis

Six TNBC scRNA-seq samples (GSM4909281, GSM4909282, GSM4909283, GSM4909284, GSM5457199, and GSM5457208) were downloaded from the GEO database, including four samples from GSE180286 and two from GSE161529 for integrated analysis. Raw count matrices were imported into the Seurat R package for processing and underwent standard quality control. Low-quality cells were filtered using the following thresholds: each gene expressed in ≥ 5 cells; ≥ 300 detected genes per cell; mitochondrial gene percentage < 20%; ribosomal gene percentage > 3%; hemoglobin gene percentage < 1%; and total detected genes ≤ 6000 per cell. Expression values were then normalized, log-transformed, centered, and scaled to unit variance.

Highly variable genes (n = 3000) were selected using the variance-stabilizing transformation (VST) method and used as input for principal component analysis (PCA). Selected principal components were further corrected for batch effects using *RunHarmony*, and Harmony embeddings were visualized by UMAP. Shared nearest-neighbor graphs were constructed using *FindNeighbors*, followed by graph-based clustering with *FindClusters*. Clusters were manually annotated into seven major lineages based on canonical marker genes reported in the literature.

Identification of malignant cells

Copy number variation (CNV) profiles were inferred from scRNA-seq data using the inferCNV R package. T cells and macrophages were designated as reference populations, whereas epithelial cells were defined as the observation group. After denoising, epithelial cells exhibiting large-scale chromosomal gains and losses were classified as malignant, while those with relatively flat CNV patterns were considered non-malignant.

Transcriptional similarity analysis among malignant cell clusters

To evaluate transcriptional similarity across malignant clusters, the average gene expression profile of each cluster was calculated. For each cluster, the top 50 genes with the highest standard deviation were selected, and their union was used as a feature gene set. Spearman correlation coefficients were computed based on the expression of these genes across clusters to quantify pairwise similarity. The resulting correlation matrix was subjected to hierarchical clustering to group clusters into transcriptionally related subpopulations.

NMF clustering

NMF analysis was performed on downsampled malignant cells, with 50 randomly selected cells from each cluster included for each run; all cells were retained when fewer than 50 were available. A log-normalized expression matrix was constructed from scRNA-seq data. The optimal factorization rank was determined by inspecting rank plots and consensus heatmaps across candidate ranks, and $k = 3$ was selected. Final NMF modeling was performed with rank = 3 and repeated 10 times to obtain stable solutions.

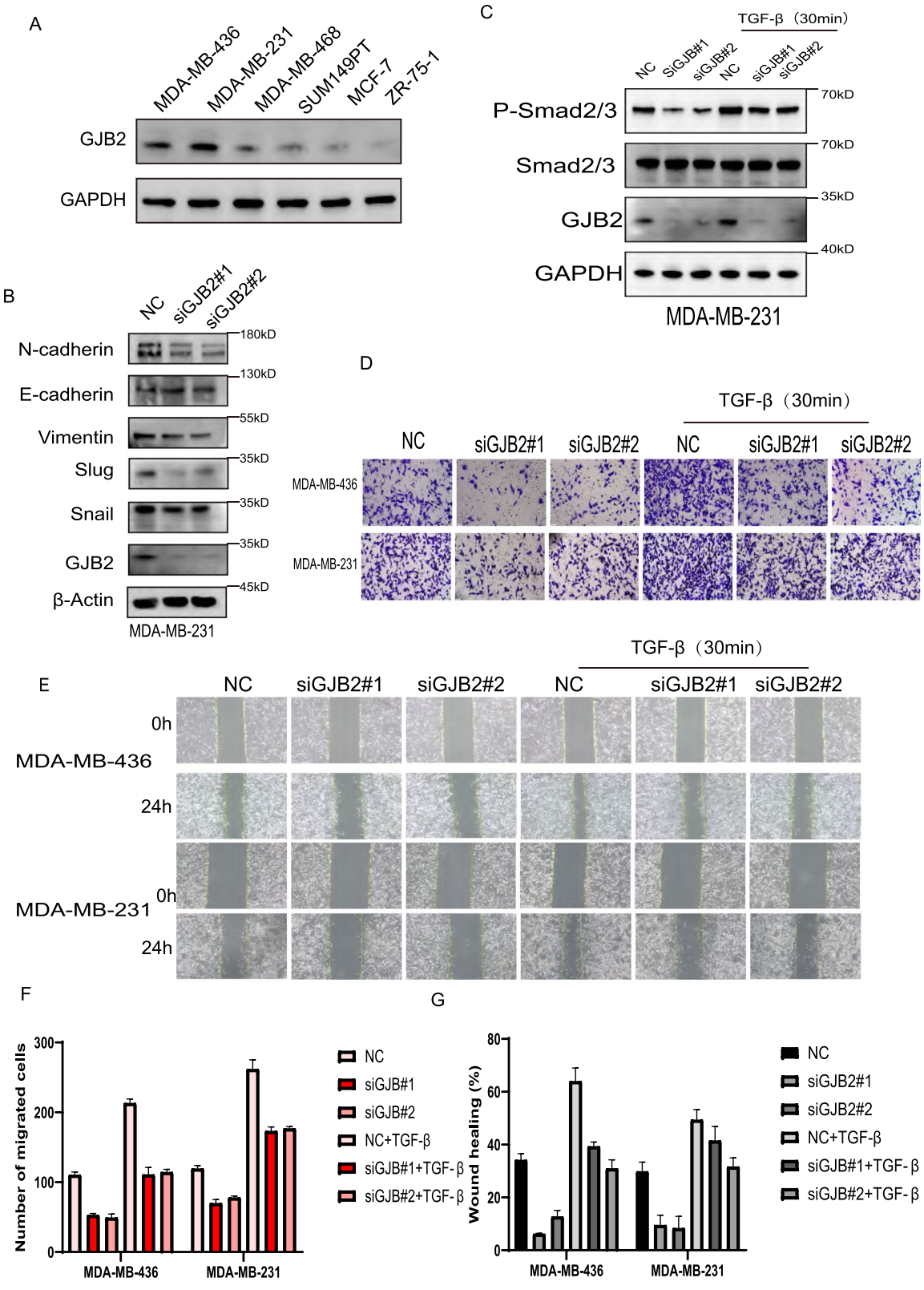

Fig. 4 Experimental validation of the role of GJB2 in EMT and migration in TNBC cells. **A** Western blot analysis of GJB2 protein expression in TNBC and non-TNBC breast cancer cell lines. **B** Western blot analysis of EMT-related proteins (E-cadherin, N-cadherin, vimentin, Snail, and Slug) in MDA-MB-231 cells transfected with siGJB2#1, or siGJB2#2. **C** Western blot analysis of total Smad2/3 and phosphorylated Smad2/3 in control and GJB2-knockdown MDA-MB-231 cells with or without exogenous TGF-β1 treatment. **D** Representative images of Transwell migration assays. **E** Representative images of wound-healing assays. **F** Quantification of migrated cells in the Transwell assay.). **G** Quantification of wound closure rates

Identification of subtype-specific marker genes

Subtype-specific markers were identified using the *FindAllMarkers* function, with subtype identity used as the grouping variable. Genes significantly upregulated in a given subtype compared with the others were designated as positive marker genes.

GO enrichment analysis and GSVA scoring

GO functional enrichment of subtype-specific DEGs was performed using the *clusterProfiler* R package. To quantitatively compare functional differences across subtypes, GSVA analysis was conducted on the log-normalized matrix using the MSigDB Hallmark gene sets to generate pathway activity scores.

Correlation and differential expression analysis of EMT-related genes

Differential expression analysis between TNBC and normal breast tissues from the TCGA-BRCA cohort was conducted using the *limma* package ($|\log2FC| > 1$, $p < 0.05$) to identify TNBC-associated DEGs. These DEGs were intersected with MSigDB-defined EMT gene sets to obtain differentially expressed EMT-related genes (DE-ERGs). Using DE-ERGs as seed genes, Pearson correlation coefficients were computed between each seed gene and all protein-coding genes in the normalized TCGA matrix. Genes significantly correlated with at least one DE-ERG ($|R| > 0.5$, $p < 0.001$) were defined as EMT-related genes (ERGs).

Single-gene GSEA

For each dataset, correlation coefficients between the target gene and all other genes were calculated and used to generate a preranked gene list. GSEA was then performed on the preranked list using MSigDB gene sets. Enrichment significance was evaluated using the normalized enrichment score (NES) and FDR q-value; pathways with FDR $q < 0.25$ were considered significantly enriched.

siRNA, cell transfection, and antibodies

Two siRNAs targeting human GJB2 were synthesized by GenePharma (Shanghai, China) and used for transient knockdown. The sequences were as follows: siGJB2-1 (GJB2-Homo-391), sense: 5′-GCUGCAAGAACGUGUGCUATT-3′, antisense: 5′-UAGCACACGUUCUUGCAGCTT-3′; siGJB2-2 (GJB2-Homo-634), sense: 5′-UCUUCUUCCGGGUCAUCUUTT-3′, antisense: 5′-AAGAUGACCCGGAAGAAGATT-3′. A scrambled siRNA was used as a negative control (siNC). Transfection was performed using a commercial transfection reagent according to the manufacturer's instructions. Antibodies used in this study included:GJB2 (16,960–1-AP, Proteintech, China), E-cadherin (20,874–1-AP, Proteintech, China), N-cadherin (22,018–1-AP, Proteintech, China), Vimentin (10,366–1-AP, Proteintech, China), Snail (13,099–1-AP, Proteintech, China), Slug (12,129–1-AP, Proteintech, China), β-Actin (66,009–1-Ig, Proteintech, China), GAPDH (GB15004, Servicebio, China), Smad2/3 (HY-P80893, MCE, China), P-Smad2/3 (HY-P86139, MCE, China).

Western blot

Cells were transfected with negative control siRNA (NC) or two independent siRNAs targeting GJB2 (siGJB2#1 and siGJB2#2) and harvested after 72 h for protein extraction. For EMT marker detection, cells were lysed in RIPA buffer containing 1% protease and phosphatase inhibitor cocktail. To assess Smad pathway activation, NC- and GJB2-knockdown cells were serum-starved for 24 h and then stimulated with recombinant human TGF-β1 (2 ng/mL, 30 min) or left untreated before lysis. Protein lysates were denatured and separated by SDS-PAGE on 4–20% gradient gels.

Cell migration assay

Transwell chambers were used to assess cell migration. Briefly, digested cells were resuspended in serum-free DMEM and seeded into the upper chamber of 24-well Transwell inserts. The lower chamber was filled with DMEM containing 10% FBS to establish a serum gradient as a chemoattractant. After 24 h, non-migrated cells on the upper surface were removed with a cotton swab, whereas migrated cells on the lower surface were fixed with 4% paraformaldehyde and stained with 0.5% crystal violet. Excess dye was washed away with PBS, and migrated cells were imaged and counted under a microscope.

Wound-healing assay

Cells ($5 \times 10\text{^}5$ per well) were seeded into 6-well plates and grown in complete medium until reaching 90–100% confluence. A straight wound was created using a sterile 200-μL pipette tip. Detached cells were removed by gentle washing with PBS, after which serum-free DMEM was added to minimize proliferation-driven effects during wound closure.

Statistical analysis

Data are presented as mean ± SD from at least three independent experiments. Statistical analyses were performed using GraphPad Prism 10.1.2 and R (version 4.5.2). Comparisons among groups were analyzed using two-way ANOVA followed by Šídák's multiple comparisons test, with comparisons made between selected columns within each row. Comparisons between two groups were performed using an unpaired two-tailed Student's t-test. Statistical significance was defined as follows: ns, not significant; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12920-026-02378-7.

Supplementary Material 1.

Authors' contributions

Hongjie Xu: Data curation; Formal analysis; Investigation; Project administration; Writing – original draft. **Chengzhi Zhang:** Data curation; Formal analysis; Project administration; Writing – original draft. **Xinyi Zhang:** Data curation; Project administration. **Shuyi Zhang:** Data curation. **Guangjun Zhou:** Data curation; Formal analysis; Investigation; Project administration; Writing – review.

Funding

This work was supported by the Research Project of Yancheng Municipal Health Commission (grant numbers YK2024074), and the Basic Research Program of the Yancheng Municipal Bureau of Science and Technology (grant number YCBK2024074).

Data availability

Publicly available single-cell RNA sequencing datasets were obtained from the Gene Expression Omnibus (GEO) under accession numbers GSE161529 and GSE180286. TCGA-BRCA data were accessed via the NCI Genomic Data Commons portal. GEPIA2 was used for independent expression validation. All other data supporting the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 17 December 2025 / Accepted: 17 April 2026
Published online: 27 April 2026

References

1. O'Reilly D, Sendi MA, Kelly CM. Overview of recent advances in metastatic triple negative breast cancer. World J Clin Oncol. 2021;12(3):164–82.
2. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. Nat Rev Clin Oncol. 2016;13(11):674–90.
3. Ke X, Li K, Jiang A, Zhang Y, Wang Q, Li Z, et al. Cloud-based GWAS platform: an innovative solution for efficient acquisition and analysis of genomic data. Med Research. 2025;1(3):397–411.
4. Zhang P, Liang X, Ye B, Wang X, Wang Y, Gong Z, et al. Metabolic reprogramming signature predicts immunotherapy efficacy in lung adenocarcinoma: targeting SLC25A1 to overcome immune resistance. Chin J Cancer Res. 2025;37(6):1000–19.
5. Gong Z, Du M, Li Y, Ye B, Huang Y, Gong H, et al. Machine learning identifies TIME subtypes linking EGFR mutations and immune states in lung adenocarcinoma. NPJ Digit Med. 2025;8(1):796.
6. Haque M, Shyanti RK, Mishra MK. Targeted therapy approaches for epithelial-mesenchymal transition in triple negative breast cancer. Front Oncol. 2024;14:1431418.
7. Shi X, Wang X, Yao W, Shi D, Shao X, Lu Z, et al. Mechanism insights and therapeutic intervention of tumor metastasis: latest developments and perspectives. Signal Transduct Target Ther. 2024;9(1):192.
8. Han X, Li X, Bai L, Zhang G. Single-cell transcriptomics in metastatic breast cancer: mapping tumor evolution and therapeutic resistance. Front Genet. 2025;16:1669741.
9. Chen P, Wu W, Zhang J, Chen J, Li Y, Sun L, et al. Pathological mechanisms of connexin26-related hearing loss: Potassium recycling, ATP-calcium signaling, or energy supply? Front Mol Neurosci. 2022;15:976388.
10. Del Castillo FJ, Del Castillo I. DFNB1 non-syndromic hearing impairment: diversity of mutations and associated phenotypes. Front Mol Neurosci. 2017;10:428.
11 Chen CX, Luo KJ, Yang JP, Huang YC, Cardenas ER, Nicholson BJ, et al. Connexins and cAMP cross-talk in cancer progression and metastasis. Cancers (Basel). 2020;13(1):58.
12. Chen G, Ning B, Shi T. Single-cell RNA-Seq technologies and related computational data analysis. Front Genet. 2019;10:317.
13. Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun. 2017;8:15081.
14 Wang S, Sun ST, Zhang XY, Ding HR, Yuan Y, He JJ, et al. The evolution of single-cell RNA sequencing technology and application: progress and perspectives. Int J Mol Sci. 2023;24(3):2943.
15. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 2010;12(5):R68.
16. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest. 2011;121(7):2750–67.
17. Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua SA, et al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. Clin Cancer Res. 2015;21(7):1688–98.
18. Keller PJ, Arendt LM, Skibinski A, Logvinenko T, Klebba I, Dong S, et al. Defining the cellular precursors to human breast cancer. Proc Natl Acad Sci U S A. 2012;109(8):2772–7.
19. Luthria K, Shah P, Caldwell B, Melms JC, Abuzaid S, Jakubikova V, et al. Single-cell profiling of sarcomas from archival tissue reveals programs associated with resistance to immune checkpoint blockade. Clin Cancer Res. 2024;30(19):4530–41.
20 Kotliar D, Veres A, Nagy MA, Tabrizi S, Hodis E, Melton DA, et al. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. Elife. 2019;8:e43803.
21. Wang H, Hu X, Yang F, Xiao H. miR-325-3p promotes the proliferation, invasion, and EMT of breast cancer cells by directly targeting S100A2. Oncol Res. 2021;28(7):731–44.
22. Liao HY, Da CM, Liao B, Zhang HH. Roles of matrix metalloproteinase-7 (MMP-7) in cancer. Clin Biochem. 2021;92:9–18.
23. Moy I, Todorović V, Dubash AD, Coon JS, Parker JB, Buranapramest M, et al. Estrogen-dependent sushi domain containing 3 regulates cytoskeleton organization and migration in breast cancer cells. Oncogene. 2015;34(3):323–33.
24. Jellinek DA, Chang AC, Larsen MR, Wang X, Robinson PJ, Reddel RR. Stanniocalcin 1 and 2 are secreted as phosphoproteins from human fibrosarcoma cells. Biochem J. 2000;350 Pt 2(Pt 2):453–61.
25. Tiwary S, Morales JE, Kwiatkowski SC, Lang FF, Rao G, McCarty JH. Metastatic brain tumors disrupt the blood-brain barrier and alter lipid metabolism by inhibiting expression of the endothelial cell fatty acid transporter Mfsd2a. Sci Rep. 2018;8(1):8267.
26. Zhang Y, Chen Y, Zhang Z, Tao X, Xu S, Zhang X, et al. Acox2 is a regulator of lysine crotonylation that mediates hepatic metabolic homeostasis in mice. Cell Death Dis. 2022;13(3):279.
27. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25–9.
28. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.
29. Sun Z, Wang CY, Lawson DA, Kwek S, Velozo HG, Owyong M, et al. Single-cell RNA sequencing reveals gene expression signatures of breast cancer-associated endothelial cells. Oncotarget. 2018;9(13):10945–61.
30. Luo L, Yang P, Mastoraki S, Rao X, Wang Y, Kettner NM, et al. Single-cell RNA sequencing identifies molecular biomarkers predicting late progression to CDK4/6 inhibition in patients with HR+/HER2- metastatic breast cancer. Mol Cancer. 2025;24(1):48.
31. Terrinoni A, Codispoti A, Serra V, Bruno E, Didona B, Paradisi M, et al. Connexin 26 (GJB2) mutations as a cause of the KID syndrome with hearing loss. Biochem Biophys Res Commun. 2010;395(1):25–30.
32. Tanaka M, Grossman HB. Connexin 26 induces growth suppression, apoptosis and increased efficacy of doxorubicin in prostate cancer cells. Oncol Rep. 2004;11(2):537–41.

33. Lu A, Shi Y, Liu Y, Lin J, Zhang H, Guo Y, et al. Integrative analyses identified ion channel genes GJB2 and SCNN1B as prognostic biomarkers and therapeutic targets for lung adenocarcinoma. Lung Cancer. 2021;158:29–39.
34. Nomura S, Maeda K, Noda E, Inoue T, Fukunaga S, Nagahara H, et al. Clinical significance of the expression of connexin26 in colorectal cancer. J Exp Clin Cancer Res. 2010;29(1):79.
35. Stewart MK, Simek J, Laird DW. Insights into the role of connexins in mammary gland morphogenesis and function. Reproduction. 2015;149(6):R279-290.
36. Tan LW, Bianco T, Dobrovic A. Variable promoter region CpG island methylation of the putative tumor suppressor gene Connexin 26 in breast cancer. Carcinogenesis. 2002;23(2):231–6.
37. Shettar A, Damineni S, Mukherjee G, Kondaiah P. Gap junction β-2 expression is negatively associated with the estrogen receptor status in breast cancer tissues and is a regulator of breast tumorigenesis. Oncol Rep. 2018;40(6):3645–53.

Publisher's note