Translational Oncology 70 (2026) 102840

Contents lists available at ScienceDirect

Translational Oncology

journal homepage: www.elsevier.com/locate/tranon

Original Research

Multi-regional transcriptomics reveals robust consensus subtypes beyond tumor heterogeneity of breast cancer

Jian Ji [a], Qing Zheng [b], Xin Li [b], Can Luo [a], Jingzhi Wang [c,*]

[a] Department of Breast Surgery, The Fifth Affiliated Hospital, Southern Medical University, 566 Congcheng Avenue, Conghua District, Guangzhou, Guangdong, China
[b] Department of Gastrointestinal Surgery, The First Affiliated Hospital of Shantou University Medical College, Shantou, China
[c] Department of Radiotherapy Oncology, The Affiliated Yancheng First Hospital of Nanjing University Medical School; The First People's Hospital of Yancheng, Yancheng, Jiangsu, China

ARTICLE INFO

Keywords:
Breast cancer
Intratumor heterogeneity
Molecular subtype
PAM50
Oncotype DX
MammaPrint

ABSTRACT

Background: Commercial molecular classifiers such as Prediction Analysis of Microarray 50 (PAM50), Oncotype DX, and MammaPrint have revolutionized breast cancer management. However, their clinical reliability is undermined by tumor heterogeneity.
Methods: We quantified the discordance in clinical molecular typing across multi-regional samples. Gene set enrichment analysis (GSEA) was employed to explore potential sources of intratumoral heterogeneity (ITH). Gene-level heterogeneity was evaluated by quantifying expression variation within versus between tumors. A robust subtyping framework was constructed via Non-negative Matrix Factorization (NMF) based on a screened low-ITH gene set.
Results: ITH was observed in over 50% of tumors, triggering significant classification discordance across commercial tests. GSEA revealed that transcriptomic heterogeneity might be driven by stromal and matrix-related pathways. The low-ITH-based NMF framework established an optimal three-subtype classification that exhibited superior prognostic stratification and maintained high stability across metastatic sites.
Conclusions: We established a low-ITH subtyping framework that overcomes the limitations of single-biopsy-based molecular typing. Our findings offer a novel strategy to enhance the precision of breast cancer management.

Introduction

Tumor heterogeneity drives malignant progression and therapeutic resistance [1]. Solid tumors comprise subclones with distinct genetic backgrounds, invasive capacities, and drug sensitivities. Breast cancer frequently exhibits polyclonal invasion, where spatial heterogeneity correlates with lymph node metastasis and poor prognosis [2,3]. However, not all tumor subclones drive disease progression [4]. Clonal selection acts on cell phenotypes rather than genotypes [5], and the heterogeneity of breast cancer also shows dissociation characteristics at the genomic and transcriptomic levels [6,7]. In fact, studies have shown that a large number of intratumoral genomic variations in cancer do not

Abbreviations: AUC, Area Under the Curve; AUCell, Area Under the Curve Cell; CAF, Cancer-associated Fibroblast; CDF, Cumulative Distribution Function; CIBERSORT, Cell-type Identification by Estimating Relative Subsets of RNA Transcripts; CNV, Copy Number Variation; ECM, Extracellular Matrix; EMT, Epithelial-Mesenchymal Transition; ER, Estrogen Receptor; ESTIMATE, Estimation of STromal and Immune cells in Malignant Tumors using Expression data; FDA, Food and Drug Administration; FDR, False Discovery Rate; GSE, Gene Expression Omnibus Series; GSEA, Gene Set Enrichment Analysis; H&E, Hematoxylin and Eosin; HER2, Human Epidermal Growth Factor Receptor 2; HR, Hormone Receptor; ICB, Immune Checkpoint Blockade; IHC, Immunohistochemistry; IJMS, International Journal of Molecular Sciences; IMS, Intrinsic Molecular Subtypes; ITH, Intratumoral Heterogeneity; KM, Kaplan-Meier; MCP-counter, Microenvironment Cell Populations-counter; METABRIC, Molecular Taxonomy of Breast Cancer International Consortium; NF-κB, Nuclear Factor Kappa-light-chain-enhancer of Activated B cells; NMF, Non-negative Matrix Factorization; OS, Overall Survival; PAM50, Prediction Analysis of Microarray 50; PC1–PC3, Principal Component 1 to 3; PCA, Principal Component Analysis; PR, Progesterone Receptor; PVL, Perivascular (cells); SC, Stem Cell; scRNA-seq, Single-cell RNA sequencing; ST, Spatial Transcriptomics; TGF-β, Transforming Growth Factor-beta; TIDE, Tumor Immune Dysfunction and Exclusion; TME, Tumor Microenvironment; TNBC, Triple-negative Breast Cancer; TNF, Tumor Necrosis Factor; UMAP, Uniform Manifold Approximation and Projection; WNT, Wingless-related integration site.

* Corresponding author.
E-mail address: wangjingzhi@ycsyy.com (J. Wang).

https://doi.org/10.1016/j.tranon.2026.102840
Received 26 February 2026; Received in revised form 21 April 2026; Accepted 26 May 2026

necessarily translate into significant phenotypic differences [5]. This phenomenon the crucial role of phenotypic plasticity [8], but the clinical value of intratumoral heterogeneity at the transcriptomic level in breast cancer still needs to be further explored.

The transcriptomic classification system for breast cancer is dominated by a series of commercial gene expression profiling assays detection tools, such as the commonly used Prediction Analysis of Microarray 50 (PAM50) [9], Oncotype DX (21-gene recurrence score) [10], and MammaPrint (70-gene prognostic signature) [11]. PAM50 categorizes tumors into five molecular subtypes (Luminal A, Luminal B, HER2-enriched, basal-like, and normal-like), which reflect distinct biological behaviors and correlate with prognosis and treatment response [9]. Oncotype DX generates a continuous recurrence score based on 16 tumor-related genes and 5 reference genes, and is applicable for chemotherapy decision-making in early-stage breast cancer patients with positive hormone receptors and negative HER2 [10,12]. As the first FDA-approved gene expression profiling test, MammaPrint identifies prognostic groups and shows unique value in the screening of early-stage invasive breast cancer patients [11,13]. However, these commercial molecular classification systems are constructed based on local tissue biopsies of a single lesion. This technical approach makes it difficult to avoid the bias caused by intratumoral heterogeneity, and may even lead to the mixing of all subtypes in tumors at different proportions [14,15]. This limitation significantly weakens the reliability of commercial classification as a predictive biomarker. Single-cell transcriptomic studies have demonstrated that tumor microenvironmental components substantially influence the overall transcriptome, intratumoral heterogeneity, and subtype classification [16]. Indeed, the microenvironmental background substantially interferes with the classification accuracy and prognostic performance efficacy of existing classification systems [17–19]. In contrast, although the intrinsic endogenous expression characteristics of cancer cells are also regulated by the microenvironment [20], they are relatively less sensitive to tumor spatial heterogeneity [21]. Based on this principle, alternative classification frameworks such as the Intrinsic subtyping have emerged [16]. However, the clinical translational value of these methodological strategies focusing on the intrinsic properties of cancer cells remains to be further verified by large-scale studies.

Single-cell transcriptome sequencing enables deciphering tumor heterogeneity but is limited by high costs, technical noise, and biological variations in large cohort studies [22]. As an alternative strategy, multi-regional bulk transcriptome sampling offers a cost-effective alternative to capture both intratumoral heterogeneity and inter-individual differences. While few studies have applied this strategy to reveal transcriptome heterogeneity and sampling bias in breast cancer [23–25], none explored mitigating its impact on molecular subtyping. Here, we analyzed 2–5 multi-region samples from 36 primary breast cancer patients (123 samples total). Integrating single-cell and spatial omics data, we confirmed intratumoral heterogeneity within existing commercial subtypes. We further developed a classification system based on inherent cancer cell expression characteristics (stable across tumor regions), demonstrating robustness to spatial heterogeneity. This method recapitulated intrinsic molecular subtypes at the single-cell level and enhanced prognostic prediction under intratumoral heterogeneity. In summary, our study establishes a novel classification framework, providing important theoretical support for precision diagnosis and treatment of breast cancer.

Methods

Data collection

The study included three independent multi-regional breast cancer transcriptome datasets: 50 samples from 18 patients provided by the Duke Breast Cancer Program [24], covering different receptor statuses; 39 samples from 8 patients with ER+/PR+/HER2- primary breast tumors provided by the National Cancer Centre Singapore [23]; and 34 tissue samples from 10 patients with triple-negative breast cancer (TNBC) provided by the Nottingham City Hospital [25]. These datasets included a total of 36 patients, with a total of 123 samples.

The data of 1980 patients with primary breast cancer from the METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) cohort [26] were used for computational modeling and survival analysis validation. In addition, two microarray datasets (GSE211729 and GSE23593) were included in the study for the construction of the intratumoral heterogeneity (ITH) scoring system, and a single-cell RNA sequencing dataset (GSE176078, including 26 primary tumors) was used to analyze the composition of the tumor microenvironment and the heterogeneity characteristics of malignant epithelial cells. GSE180286 and GSE32489 were used to evaluate the applicability of the ITH score.

Molecular typing and heterogeneity assessment of multi-regional samples

This study used three mainstream transcriptome detection methods to analyze samples from different regions: PAM50 molecular typing, MammaPrint recurrence risk score, and Oncotype DX recurrence score. PAM50 analysis was implemented using the genefu R package [27], and the samples were divided into five molecular subtypes: Luminal A, Luminal B, HER2-enriched, Basal-like, and Normal-like. The MammaPrint and Oncotype DX scores were calculated based on the 70-gene prognostic signature and the 21-gene recurrence score algorithm, respectively.

The criteria for determining intratumoral heterogeneity were as follows: For PAM50 typing, if multiple regional samples from the same patient were assigned to different PAM50 subtypes, the tumor was defined as exhibiting PAM50 heterogeneity; for MammaPrint and Oncotype DX, if different regional samples from the same tumor were classified into different risk categories. To quantify the overall heterogeneity degree of the transcriptome, this study calculated the maximum Euclidean distance of each sample in the space of the first to third principal components (PC1-PC3) of the principal component analysis (PCA) within the same patient group.

Construction of the ITH scoring system

First, two microarray datasets, GSE211729 and GSE23593, were integrated. After batch-removal processing, they were merged to construct a gene expression matrix of multi-regional samples. For each protein-coding gene, the degree of expression variation within the tumor (within-tumor) relative to that between tumors (between-tumor) was calculated to establish the ITH score. Specifically, the ITH score reflects the ratio of the expression fluctuation of a gene between different regions of the same patient to the variation between different patients. A higher score indicates greater intratumoral spatial heterogeneity of the gene.

Single-cell sequencing and spatial transcriptomics data analysis

The Seurat R package was used for data quality control, normalization, and dimensionality reduction analysis. Unsupervised clustering analysis was used to identify the main cell types. Malignant epithelial cells were distinguished from non-malignant cells based on copy number variation (CNV) inference and the expression pattern of epithelial marker genes. The AUCell package was used to quantify the expression enrichment of the ITH-low gene set in tissue micro-regions to verify its specific expression in malignant regions.

Development of a robust classification framework based on low-ITH genes

The Non-negative Matrix Factorization (NMF) algorithm was used, with the ITH-low gene set as the feature input, to discover subtypes in

1980 primary tumor samples from the METABRIC cohort. By systematically evaluating the clustering stability of different k values (number of clusters), the optimal number of clusters was determined. To verify whether low-ITH genes could capture the intrinsic molecular subtypes of breast cancer, the NMF clustering results were compared with the classic intrinsic molecular subtypes (IMS) classification [16].

Heterogeneity assessment model for single samples

First, template gene sets for each PAM50 subtype were constructed through differential gene expression analysis. Then, the singscore package was used to calculate the enrichment score of each subtype template gene set in each sample. Tumors that were significantly enriched in multiple subtypes ($P < 0.05$) were defined as heterogeneous tumors, and the subtype with the highest enrichment score was the main subtype.

Pathway enrichment analysis

The limma R package was used for differential expression analysis, and the screening criteria were set as adjusted *P*-value (FDR) < 0.05 and $|\log2 \text{ fold change}| > 0.05$. Gene set enrichment analysis (GSEA) was used to evaluate the activation status of specific biological pathways.

Immune infiltration analysis

Estimation of STromal and Immune cells in Malignant Tumors using Expression data (ESTIMATE) [28] was used to calculate the purity of the tumor. For specific cell population quantification, Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) [29] was applied with 1000 permutations to estimate the relative proportions of 22 distinct immune cell types. Meanwhile, the Microenvironment Cell Populations-counter (MCP-counter) algorithm [30] was utilized to determine the absolute abundance of immune cell types. Furthermore, the Tumor Immune Dysfunction and Exclusion (TIDE) score [31] was calculated to predict the potential response to immune checkpoint blockade (ICB) therapy and to assess mechanisms of tumor immune evasion.

Statistical analysis

All statistical analyses were performed using R software (version 4.4.2). For continuous variables, the *t*-test or Wilcoxon rank-sum test was used to compare differences between groups, and for categorical variables, the chi-square test or Fisher's exact test was used. Pearson or Spearman correlation tests were used for correlation analysis. The log-rank analysis was used for Kaplan-Meier (KM) curves. A *P*-value < 0.05 was considered statistically significant.

Results

Intratumoral transcriptomic heterogeneity affects commercial breast cancer typing

We systematically searched public multi-regional breast cancer transcriptome data and finally included 3 datasets for analysis. The Duke Breast Cancer Program provided 50 samples from 18 patients, covering different receptor statuses; the National Cancer Centre Singapore contributed 39 samples from 8 patients with ER+/PR+/HER2- primary breast tumors; and the Nottingham City Hospital provided 34 tissue samples from 10 patients with TNBC. These datasets included a total of 36 patients with primary breast cancer. For each patient, 2–5 representative tumor region samples were collected through a standardized surgical procedure to capture the spatial heterogeneity characteristics within the tumor. After strict quality control and RNA integrity assessment, a total of 123 samples were finally included in the transcriptome analysis.

To initially explore the intratumoral transcriptomic heterogeneity of breast cancer, we used three mainstream transcriptome detection methods: PAM50 molecular typing, MammaPrint recurrence risk score, and Oncotype DX recurrence score, to analyze samples from different regions (Fig. 1A-1I). The study found that 53% of the tumors showed intratumoral PAM50 heterogeneity, 36% of the tumors showed intratumoral Oncotype DX heterogeneity, and 25% of the tumors showed intratumoral MammaPrint heterogeneity. Additionally, we averaged the gene expression values across all regions from the same patient to simulate the whole-tumor global level, and evaluated the concordance between multi-regional sample classification and patient-level averaged PAM50 subtyping. The results revealed substantial discordance in subtype classification between the two approaches (Fig. S1).

PAM50 heterogeneity analysis showed that Luminal A and Luminal B often mixed with other subtypes. The Luminal type, especially Luminal B, is not a homogeneous molecular entity pure breed in terms of gene expression profiles but a highly heterogeneous collection [32]. It is essentially in an intermediate state between hormone-driven and proliferation-driven phenotypes, so it frequently harbors transcriptional features is easy to contain the characteristics of other subtypes. In addition, even for clinically pathological TNBC, its PAM50 is not all Basal-like (Fig. 1G). This inconsistency between receptor protein expression and gene expression is exactly the challenge that needs to be overcome in the precision treatment of breast cancer. Previous study showed that non-Basal TNBC in PAM50 has a better response to capecitabine treatment [33]. IHC-based subtyping alone is insufficient for uniform treatment decisions. Deeper molecular characterization is needed to guide individualized therapy.

Although the incidence of gene expression heterogeneity in Oncotype DX and MammaPrint is low (Fig. 1B-1I), their clinical significance cannot be ignored. These two tests have been used for chemotherapy decision-making in early-stage breast cancer, and their accuracy depends on the representativeness of the biopsy sample for the overall tumor risk. If the puncture happens to sample from a heterogeneous region, it will lead to a deviation between the test and the actual biological behavior, resulting in overtreatment or chemotherapy exemption. Both MammaPrint and Oncotype DX were developed based on the gene expression characteristics of the specific subtype of (HR+)/HER2-negative. We tried to extend the application of these molecular detection tools to other pathological subtypes. In the TNBC dataset, all samples showed a high MammaPrint recurrence risk score (Fig. 1H). The Oncotype DX recurrence score also classified most TNBC samples as high-risk (Fig. 1I). The Oncotype DX and MammaPrint gene signatures mainly evaluate the expression levels of genes related to cell proliferation, invasion, and metastasis. The design principle of these tests just captures the malignant biological behavior of TNBC. Merely confirming of known malignant characteristics was insufficient for TNBC, demanding the development of novel biomarkers capable of capturing its profound internal heterogeneity to achieve the goals of precision medicine.

To further quantitatively evaluate the overall heterogeneity degree of the transcriptome, we calculated the maximum Euclidean distance between samples of each sample in the PC1-PC3 space within the same patient group. This index effectively reflects the transcriptome dispersion degree within the tumor by capturing the farthest expression profile difference between samples. The results showed that in the three molecular typing systems of PAM50, MammaPrint, and Oncotype DX, the maximum Euclidean distance of heterogeneous tumors was significantly greater than that of homogeneous tumors (Fig. 1J-1L), suggesting that the heterogeneous tumors identified by these commercial transcriptomic assays gene expression profile detection methods have higher transcriptome diversity.

Further analysis of the impact of sample size on heterogeneity assessment (Fig. 1M) found that as the number of single-tumor samples increased from 2 to 4, the heterogeneity level showed an upward trend;

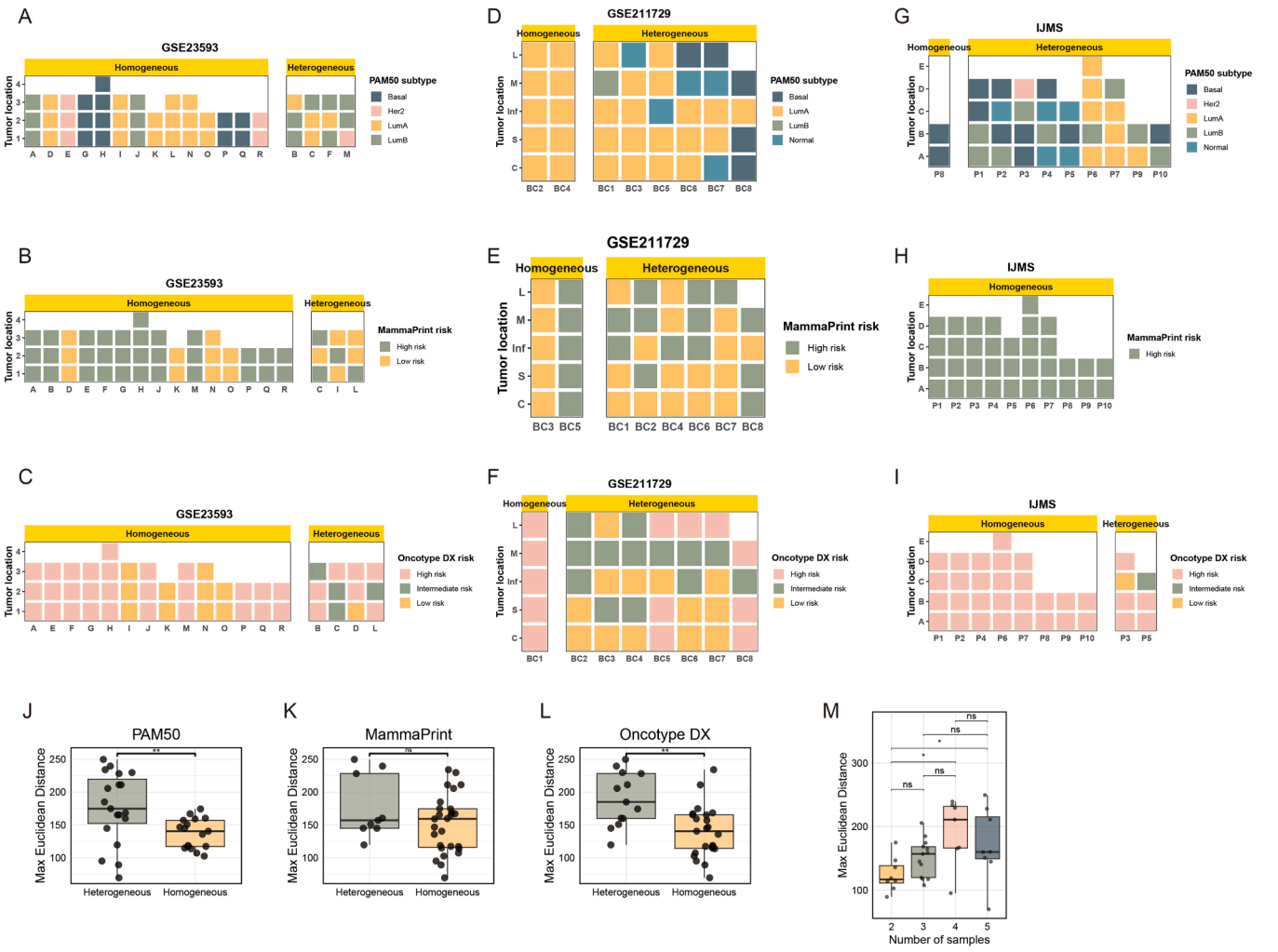

Fig. 1. Intratumoral transcriptomic heterogeneity impacts clinical molecular subtyping in breast cancer. (A–I) Intratumoral subtyping consistency and heterogeneity across different datasets. Heatmaps display results from three datasets (GSE23593, GSE211729, and IJMS) using mainstream transcriptomic platforms: PAM50 (A, D, G), MammaPrint (B, E, H), and Oncotype DX (C, F, I). Each column represents an individual patient (labeled with patient IDs), and each row represents a distinct sampling location within the same tumor. Patients are categorized into "Homogeneous" (consistent results across all sites) and "Heterogeneous" (conflicting subtypes or risk scores within the same tumor). (J–L) Quantitative assessment of transcriptomic divergence. Boxplots compare the Maximum Euclidean Distance between heterogeneous and homogeneous tumors based on PAM50 (J), MammaPrint (K), and Oncotype DX (L) scores. The Euclidean distance quantifies the transcriptomic dissimilarity between multiple samples from the same patient; higher values indicate greater intratumoral divergence. (M) Influence of sampling density on heterogeneity detection. Distribution of Maximum Euclidean Distances as a function of the number of samples collected per tumor (ranging from 2 to 5). $*P < 0.05$; $**P < 0.01$; ns, not significant., not significant. PAM50: Prediction Analysis of Microarray 50; LumA: Luminal A; LumB: Luminal B; Her2: HER2-enriched.

however, when the sample size reached 5, the heterogeneity level decreased instead. This non-linear change suggests that a small sample size may be insufficient to fully capture tumor heterogeneity, while increasing the sample size (≥5) may lead to the homogenization of heterogeneity due to covering the core area of the tumor. All these results indicate that for high-heterogeneity breast cancer, transcriptome detection based on a single biopsy sample may be insufficient to guide systemic treatment decisions.

Transcriptomic heterogeneity is associated with stromal infiltration

Oncotype DX and MammaPrint are only limited to chemotherapy guidance for HR+/HER2- early-stage breast cancer, while PAM50 has a wider clinical positioning and is suitable for comprehensive typing of all molecular subtypes. Given that this study included a multi-pathological subtype cohort, PAM50 was selected for further analysis. We performed GSEA on a selected set of breast cancer-related gene sets (n = 54) to explore the biological characteristics of tumors with heterogeneous PAM50 classification. In two breast cancer datasets containing multiple pathological subtypes, tumors with heterogeneous PAM50 classification showed significant enrichment of stromal-like and matrix characteristics (Fig. 2A, 2B). Specifically, these included the activation of pathways such as stromal infiltration, epithelial-mesenchymal transition (EMT), extracellular matrix components, and fibroblast activation markers. Interestingly, in TNBC, tumors with heterogeneous PAM50 classification showed a completely different enrichment pattern, including significant activation of metabolic pathways, such as arachidonic acid metabolism, linoleic acid metabolism, and tyrosine metabolism (Fig. 2C). This indicates that the heterogeneity of TNBC may mainly originate from metabolic reprogramming rather than changes in the stromal microenvironment. It is worth noting that in all three datasets, tumors with homogeneous PAM50 classification showed consistent characteristics of enhanced immune infiltration and activation of MYC targets (Fig. 2A-2C). These results suggest that immune activation and MYC signaling may be the key driving factors for maintaining tumor molecular homogeneity.

We then used MCP-counter [30] to quantify fibroblast abundance (Fig. S2). Patients with heterogeneous PAM50 subtypes showed greater

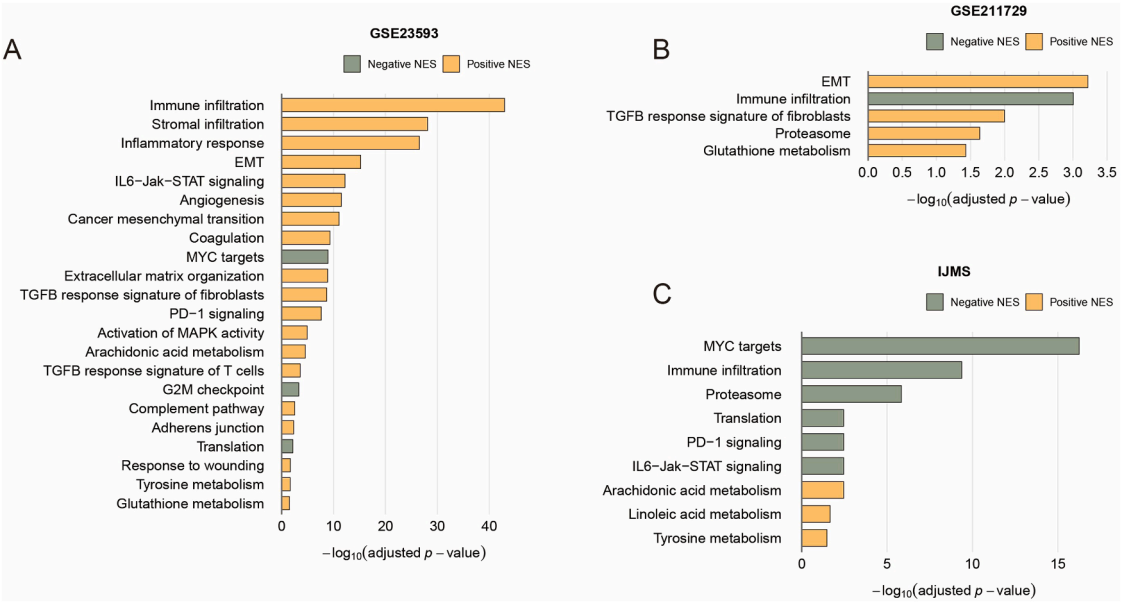

Fig. 2. Biological landscapes and signaling pathways associated with intratumoral PAM50 heterogeneity. (A, B) Enrichment of stromal and mesenchymal signatures in heterogeneous breast tumors. Bar plots illustrate significantly enriched pathways in tumors with heterogeneous PAM50 classification within two independent datasets: GSE23593 (A) and GSE211729 (B). (C) Heterogeneous TNBC tumors show distinct activation of metabolic pathways, including arachidonic acid metabolism, linoleic acid metabolism, and tyrosine metabolism. PAM50: Prediction Analysis of Microarray 50; GSE: Gene Expression Omnibus Series; TNBC: Triple-Negative Breast Cancer.

fibroblast variability than homogeneous cases. Following correction for tumor purity based on breast carcinoma-specific epithelial gene expression (EPCAM, KRT8, KRT18, KRT19, CDH1), fibroblast abundance persisted at higher levels in the heterogeneous group. These findings suggest a potential role for tumor stroma in modulating transcriptomic heterogeneity.

Prognostic impact of intratumoral PAM50 heterogeneity

To verify the clinical significance of intratumoral PAM50 heterogeneity in a larger sample, we performed computational analysis on 1980 patients with primary breast cancer in the METABRIC cohort. First, the genefu package was used to perform PAM50 molecular typing on all patients, and it was confirmed that different subtypes had significant prognostic differences (Fig. 3A). Given the lack of survival information in the multi-regional dataset, we developed a heterogeneity assessment model based on single samples: template gene sets for each PAM50 subtype were constructed through differential gene expression analysis, the singscore package was used to calculate the enrichment score of each sample, and tumors that were significantly enriched in multiple subtypes ($P < 0.05$) were defined as heterogeneous tumors, and the subtype with the highest enrichment score was the main subtype. The overall consistency rate of this model with the original PAM50 centroid predictor was 70.99% (Fig. 3B). The inconsistent cases mainly originated from a higher proportion of determination of Luminal A and Luminal B subtypes in the enrichment analysis, suggesting that single-region sampling may underestimate the internal heterogeneity of luminal-type tumors. In the METABRIC cohort, 56.9% of the tumors showed PAM50 heterogeneity, which was higher than the observed value in the multi-regional sequencing data, possibly reflecting the improvement of statistical power brought by a larger sample size.

Survival analysis showed that there was no significant difference in overall survival between most heterogeneous tumors and homogeneous tumors (Fig. 3C). A reasonable speculation is that the prognosis of breast cancer is mainly determined by the dominant clone. During tumor evolution, the more adaptable clone will becomes the dominant component in the population through selective expansion [34]. The biological characteristics of this clone, including proliferation rate, invasion ability, and treatment sensitivity, directly determine the clinical behavior of the tumor [34]. In other words, this dominant subtype already carries the strongest prognostic information. Therefore, although ITH objectively exists, its prognostic value may be masked have been diluted by the dominant subtype. We speculate that when the dominant clone itself has a good prognosis, the genomic instability represented by ITH may become an additional risk modifier; while when the dominant clone is already highly invasive, the additional impact of ITH is difficult to appear. The results showed that only in the Luminal A subtype (Fig. 3C), the overall survival of heterogeneous tumors was significantly lower than that of homogeneous tumors, while no significant difference was observed in the Basal-like, HER2-enriched, and Luminal B subtypes (Fig. 3C). This finding supports our inference: ITH is not an independent prognostic driving factor but produces a context-dependent effect through interaction with the dominant clone. In a low-invasive background, the genomic instability reflected by ITH constitutes an additive risk dimension and has a synergistic effect with the inert characteristics of the dominant clone; while in a high-invasive background, the dominant clone already occupies the dominant weight in prognosis determination, and the incremental information value of ITH is significantly compressed.

Identification of low-ITH genes with inherent expression signals

To analyze the transcriptomic basis of ITH, we first integrated two microarray datasets (GSE211729 and GSE23593). After batch-removal processing (Fig. S3), they were merged to construct a gene expression matrix of multi-regional samples, and the ITH scoring system was established by quantifying the degree of expression variation within the tumor relative to that between tumors. The distribution of the ITH score showed a significant right-skew feature (Fig. 4A), and the long tail on the right indicated that there was a small number of genes with high intratumoral heterogeneity. Based on this distribution feature, we divided protein-coding genes into three groups (Fig. 4B): ITH-high genes (the top 8.3% in score), ITH-low genes (the bottom 36.5% in score), and an intermediate group. This grouping strategy allowed us to distinguish

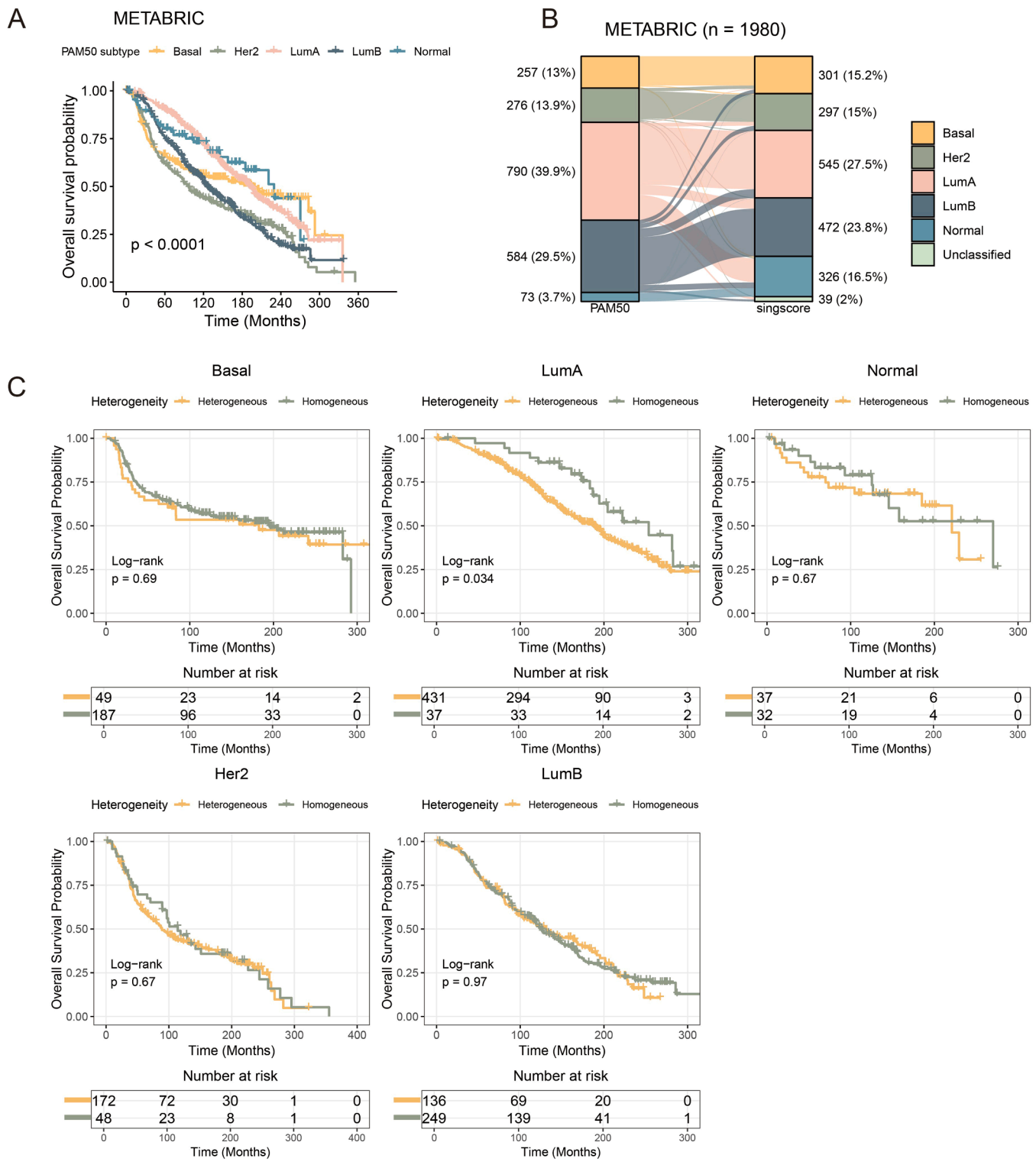

Fig. 3. Prognostic impact of intratumoral PAM50 heterogeneity in the METABRIC cohort. (A) Overall survival across PAM50 molecular subtypes. Kaplan-Meier curves show the overall survival probability for the five intrinsic molecular subtypes in the METABRIC dataset. (B) Concordance between subtyping methods. The alluvial plot illustrates the consistency between the original PAM50 centroid-based predictor and the singscore-based heterogeneity assessment model. (C) Context-dependent prognostic impact of ITH. Survival analyses were performed by stratifying patients into "Heterogeneous" and "Homogeneous" groups within each PAM50 subtype (Basal, LumA, Normal, Her2, and LumB). PAM50: Prediction Analysis of Microarray 50; LumA: Luminal A; LumB: Luminal B; Her2: HER2-enriched; ITH: intratumoral heterogeneity.

between two gene sets: those with large spatial expression variation and those with uniform spatial expression. To explore the biological significance of different ITH gene sets, we performed PCA based on ITH-high and ITH-low genes, respectively. The results showed that the PC1 of tumor samples based on ITH-high genes had the strongest correlation with stromal characteristics (Fig. 4C). This finding indicates that the main source of variation in ITH-high genes is the stromal components of the tumor microenvironment rather than the cancer cells themselves. In

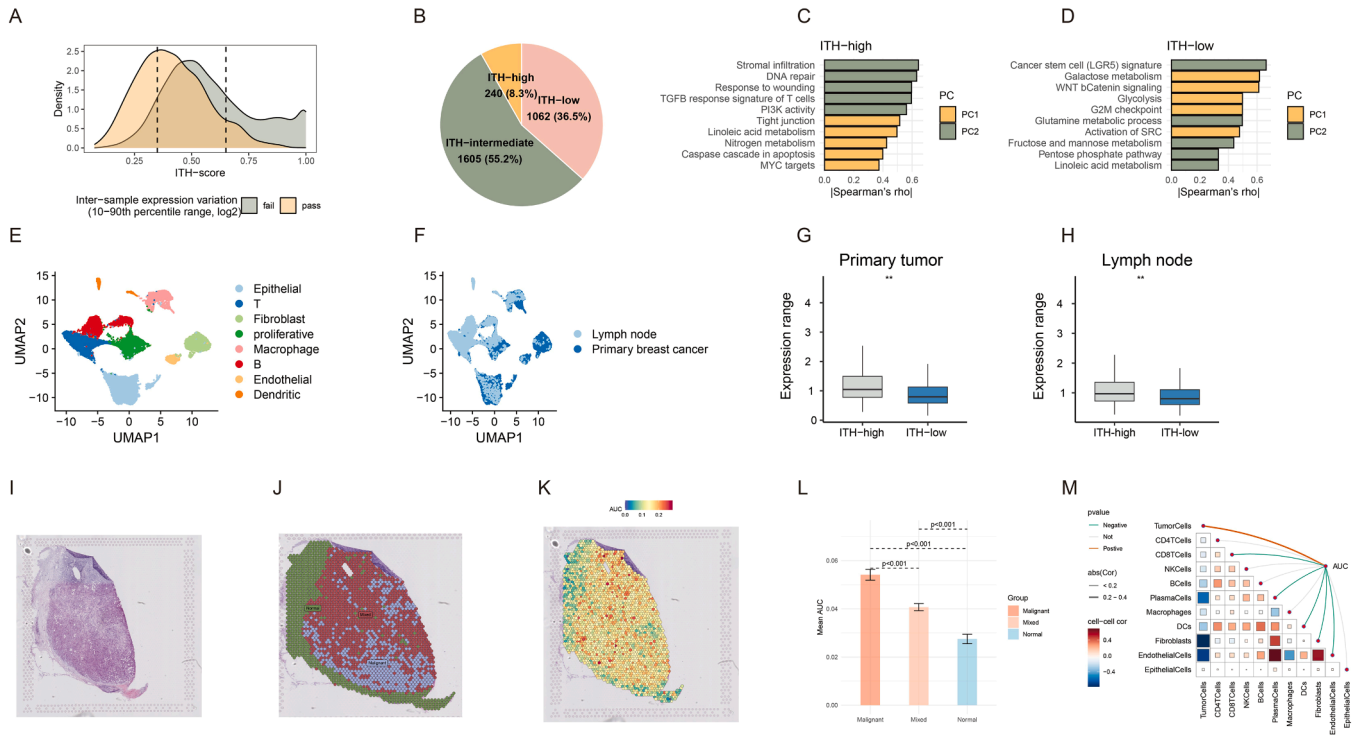

Fig. 4. Identification and multi-omics validation of gene sets with intratumoral expression stability. (A, B) Stratification of genes based on ITH scores. (A) Density plot illustrating the distribution of ITH scores. (B) Classification of protein-coding genes into three distinct groups: ITH-high, ITH-low, and ITH-intermediate. (C, D) Biological signatures of ITH-high and ITH-low genes. (C) ITH-high genes are primarily associated with stromal and microenvironmental characteristics, including stromal infiltration and TGF-β signaling. (D) ITH-low genes are intrinsically linked to tumor hallmarks, with PC1 reflecting cell cycle (G2M checkpoint) and WNT/β-catenin signaling, and PC2 correlating with cancer stem cell signatures and metabolic reprogramming. (E, F) UMAP plots showing the landscape of major cell types (E) and sample origins (F) in GSE180286. (G, H) Boxplots comparing the expression range of ITH-high and ITH-low gene sets in primary tumors (G) and lymph node metastases (H). (I–K) Representative ST slides showing H&E staining (I), unsupervised clustering (J), and the spatial activity (AUC score) of the ITH-low gene set (K). (L) Comparison of mean AUC scores across different spatial domains. (M) Correlation matrix showing the relationship between the ITH-low signature and various cell type proportions. ITH: Intratumoral Heterogeneity; TGF-β: Transforming Growth Factor-Beta; PC1/PC2: Principal Component 1/2; WNT: Wingless-related Integration Site; UMAP: Uniform Manifold Approximation and Projection; GSE: Gene Expression Omnibus Series; ST: Spatial Transcriptomics; H&E: Hematoxylin and Eosin; AUC: Area Under the Curve.

other words, the expression fluctuations observed in multi-regional sampling largely reflect the spatial differences in the stromal content within the tumor. In contrast to ITH-high genes, ITH-low genes are uniformly expressed within the tumor region (Fig. 4D), suggesting that they are less affected by the microenvironment and can better reflect the inherent biological characteristics of cancer cells. PCA based on ITH-low genes revealed the structural variation of the intrinsic programs of cancer cells: PC1 was closely related to the characteristics of the cell cycle and Wingless-related integration site (WNT) signaling pathway, while PC2 was related to the characteristics of cell stemness (Fig. 4D). This shows that after excluding the noise of stromal heterogeneity, ITH-low genes provide a more refined characterization of tumor subtypes based on the intrinsic characteristics of cancer cells.

Using single-cell RNA sequencing data from paired tumor core and lymph node samples of 15 cases of primary breast cancer, we confirmed that ITH-high genes exhibited higher expression variability among cells in paired samples, while ITH-low genes showed relatively stable expression (Fig. 4E-4H), supporting the applicability of the ITH score at the single-cell resolution. By quantifying the expression enrichment of the ITH-low gene set in tissue microregions using the AUCell package, we found that ITH-low signature it was specifically highly expressed in malignant regions, and its expression level was only positively correlated with the proportion of tumor cells (Fig. 4I-4M), further demonstrating that ITH-low genes are reliable markers of the intrinsic characteristics of cancer cells.

In summary, the ITH scoring system effectively distinguishes two gene sets with different biological essences: ITH-high genes mainly capture the heterogeneity of the tumor microenvironment, while ITH-low genes reflect the intrinsic molecular programs of cancer cells. By focusing on ITH-low genes, the interference of stromal components can be filtered, and the essential characteristics of cancer cells and their spatial evolution patterns can be more accurately characterized.

Transcriptomic classification less affected by ITH

The Intrinsic Molecular Subtypes (IMS) of cancer are considered to have low ITH because they mainly reflect the autonomous molecular characteristics of cancer cells and are less affected by stromal components in the tumor microenvironment [16]. To elucidate the cellular structure of breast cancer, we analyzed 26 cases of primary untreated tumors using single-cell RNA sequencing (GSE176078), including 11 cases of ER+, 5 cases of HER2+, and 10 cases of TNBC. Through unsupervised clustering analysis, 9 major cell clusters and 29 minor cell clusters were identified in the original annotation (Fig. 5A, 5B). The major populations included B cells, cancer-associated fibroblasts (CAFs), malignant epithelial cells, endothelial cells, myeloid cells, normal epithelial cells, plasmablasts, perivascular (PVL) cells, and T cells. Differential expression analysis showed a significant separation pattern in the average expression of epithelial marker genes between malignant and non-malignant cells (Fig. 5C, 5D), confirming that epithelial characteristics can effectively distinguish cancer cells from stromal components in the tumor microenvironment. Notably, major cell types were distributed across different clinical subtypes (Fig. 5E). Further analysis revealed significant heterogeneity among malignant epithelial cells,

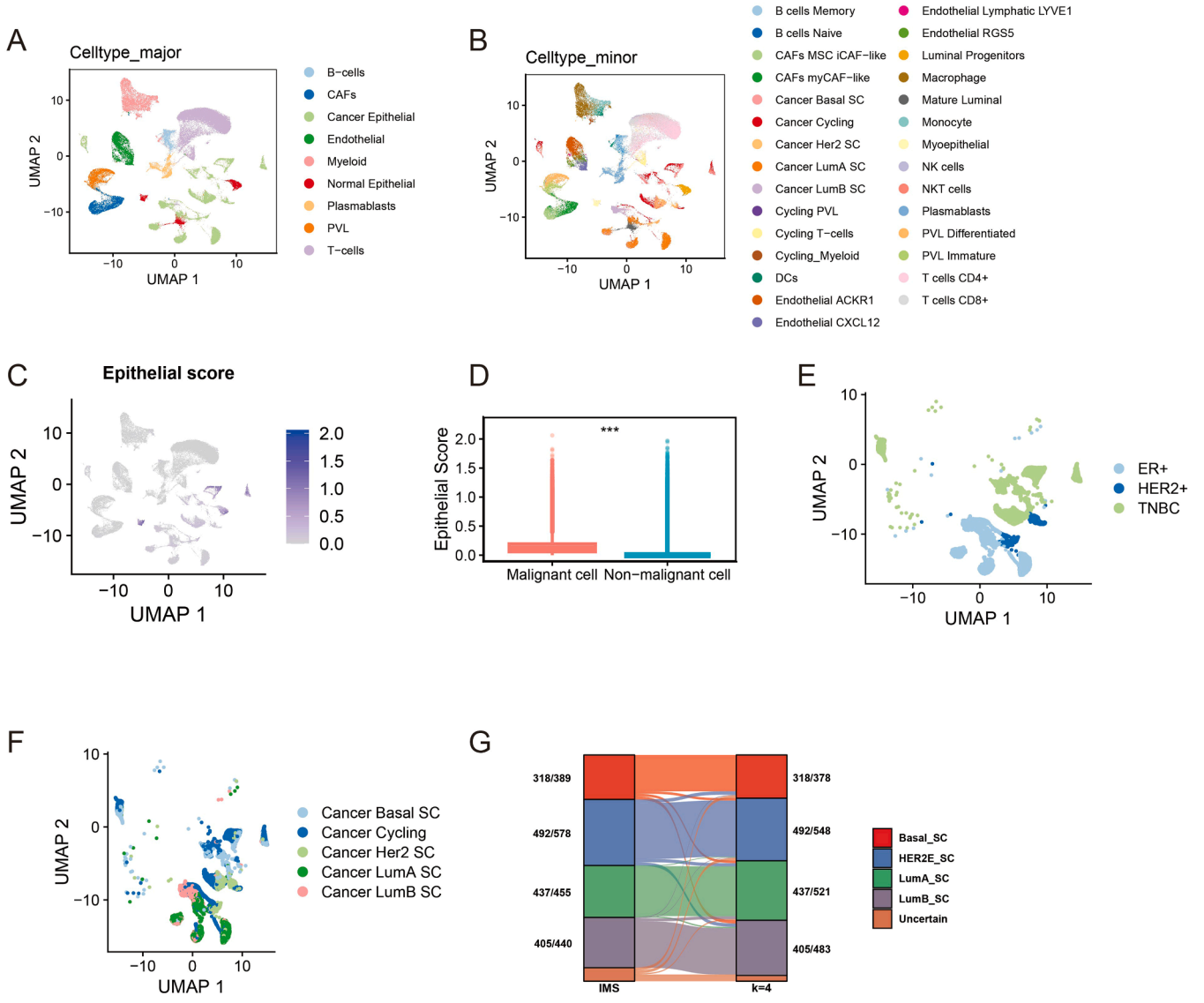

Fig. 5. Transcriptomic classification less affected by intratumoral heterogeneity. (A, B) Unsupervised clustering analysis of 26 primary untreated tumors identifying 9 major cell clusters (A) and 29 minor cell clusters (B) in GSE176078. (C) UMAP projection illustrating the transcriptomic activity of epithelial-specific marker genes. (D) Quantitative boxplot showing significantly higher epithelial scores in malignant cells compared to non-malignant counterparts. (E) Distribution of cancer epithelial cells color-coded by clinical classification (ER+, HER2+, and TNBC). (F) Identification of five distinct states within cancer epithelial cells, including Stem-like, Cycling, and Luminal-like populations. (G) Alluvial plot demonstrating the robust alignment between NMF clustering (based on stable low-ITH genes) and the Intrinsic Molecular Subtypes. GSE: Gene Expression Omnibus Series; UMAP: Uniform Manifold Approximation and Projection; ER+: Estrogen receptor positive; HER2+: Human epidermal growth factor receptor 2 positive; TNBC: Triple-negative breast cancer; NMF: Non-negative matrix factorization. ITH: Intratumoral heterogeneity.

which were further subdivided into five functional subtypes (Fig. 5F): Basal Cancer SC (stem cells), Cancer Cycling, Cancer Her2 SC, Cancer LumA SC, and LumB SC.

In the multi-region sample set, low-ITH genes retained expression variation between tumors, and their inter-tumor expression range (10th to 90th percentiles) was higher than that of high-ITH or medium-ITH genes (Fig. 4A). Based on this finding, we developed a robust transcriptomic classification strategy based on low-ITH genes. Specifically, we used the Non-negative Matrix Factorization (NMF) algorithm, with the low-ITH gene set as the feature input, to perform subtype discovery on a cohort of 1980 primary tumor samples (preset number of clusters k = 4). The clustering results (denoted as k4 subtypes) showed a high degree of consistency (Fig. 5G) with the classic four IMS classifications (Luminal A, Luminal B, HER2-enriched, Basal-like). This result verified that low-ITH genes can capture the intrinsic molecular subtypes of breast cancer; secondly, through the multi-region sequencing strategy, the intrinsic subtypes of cancer cells revealed by single-cell sequencing can be approximately reproduced at the bulk RNA-seq level.

Consistent classification based on low-ITH genes

By evaluating the clustering stability at different k values of NMF, we determined that k = 3 was the optimal number of clusters (Fig. 6A-6C), dividing the patients into three molecular subtypes (A, B, C). Survival analysis (Fig. 6D) showed that this classification method was significantly correlated with the overall survival of breast cancer patients: patients with subtype A had the best prognosis, while those with subtype C had the worst prognosis. This finding suggests that the low-ITH gene characteristics have potential prognostic stratification value.

In view of the clinical translational potential of this classification, we further verified its applicability in metastatic tumors. Using referring to previous studies on the expression profiles of primary breast cancer and

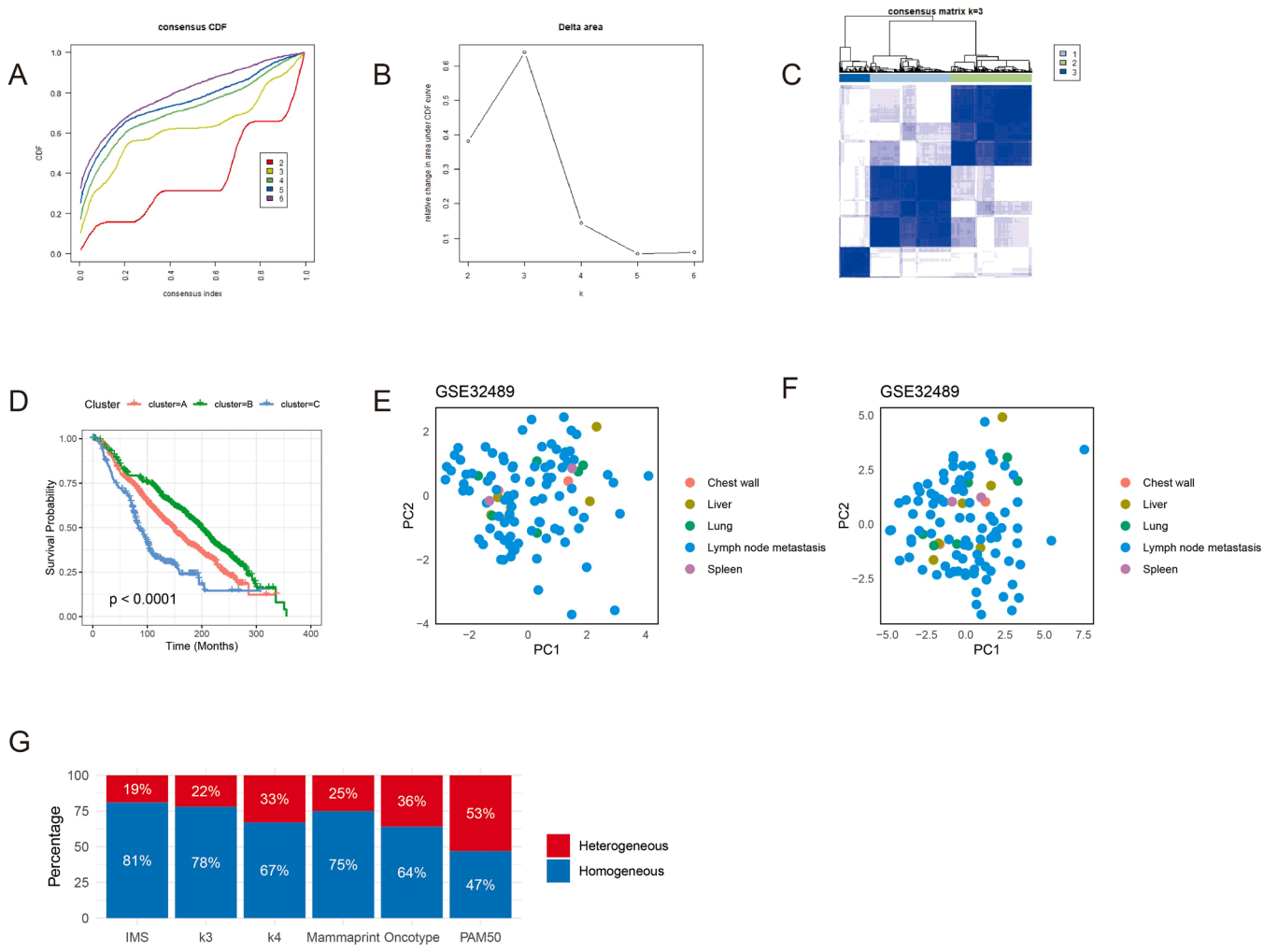

Fig. 6. Consistent classification based on low-ITH genes. (A–C) Evaluation of clustering stability using Cumulative Distribution Function curves (A), the Delta area plot (showing k = 3 as the elbow point) (B), and the consensus matrix heat map (C), which collectively defined k = 3 as the optimal number of molecular subtypes.. (D) Kaplan-Meier survival analysis demonstrating a significant difference in survival probability among the three identified molecular clusters. (E) PCA showing that gene expression profiles based on low-ITH genes do not cluster by anatomical site. (F) Comparative PCA using classic Intrinsic Molecular Subtype genes. (G) A stacked bar chart comparing the proportions of heterogeneous versus homogeneous classifications across different models. ITH: Intratumoral Heterogeneity; PCA: Principal Component Analysis.

multi-site metastases (liver, chest wall, lymph nodes, lungs, spleen), we found that PCA based on either low-ITH genes or IMS template genes failed to distinguish tumor samples from different anatomical sites (Fig. 6E, 6F). Therefore, this classification system may be applicable to metastatic lesions.

Furthermore, we employed the ESTIMATE algorithm [28] to assess tumor purity across multi-region data and applied NMF clustering following purity correction (Fig. S4A-S4D). Although this approach improved subtype stability (Fig. S4E), the low-ITH strategy achieved higher cross-regional concordance without relying on tumor purity correction (Fig. 6G), indicating that it captures stable intrinsic molecular programs underlying tumor evolution rather than merely eliminating stromal noise. In addition, the consistent classification based on low-ITH genes showed low tumor heterogeneity bias, further supporting its reliability as a cross-lesion molecular typing method.

To elucidate the biological basis for the prognostic differences among different subtypes, we performed differential gene expression analysis (Fig. 7A) and pathway enrichment studies. The enrichment analysis results were highly consistent with the survival data: compared with subtype A and B, subtype C was significantly enriched in cancer-promoting pathways such as TNF signaling, NF-κB pathway, and cell cycle regulation (Fig. 7B, 7C); compared with subtype B, subtype A also showed relative inhibition of these pathways (Fig. 7D). Quantitative analysis of immune infiltration based on the CIBERSORT algorithm [29] revealed significant heterogeneity in immune microenvironment composition across the three subtypes (Fig. S5). Notably, subtype C exhibited significantly higher M2 macrophage infiltration than the other two groups, displaying a pro-tumorigenic immune phenotype; whereas subtype B was characterized by T cell infiltration, suggesting more active immune responses in this subtype. Further analysis demonstrated that subtype C had the highest TIDE score and Exclusion Score among the three subtypes (Fig. S6A), indicating the poorest potential for immunotherapy response. However, drug sensitivity analysis confirmed that subtype C tumors were more likely to benefit from standard chemotherapeutic agents (Fig. S6B).

Discussion

The pervasive impact of transcriptomic heterogeneity on clinical subtyping

This study demonstrates that established commercial transcriptomic tools for breast cancer, such as PAM50, Oncotype DX, and MammaPrint, are vulnerable to intratumoral transcriptomic heterogeneity. Our findings reveal that 53% of tumors exhibit discordance in PAM50 subtyping

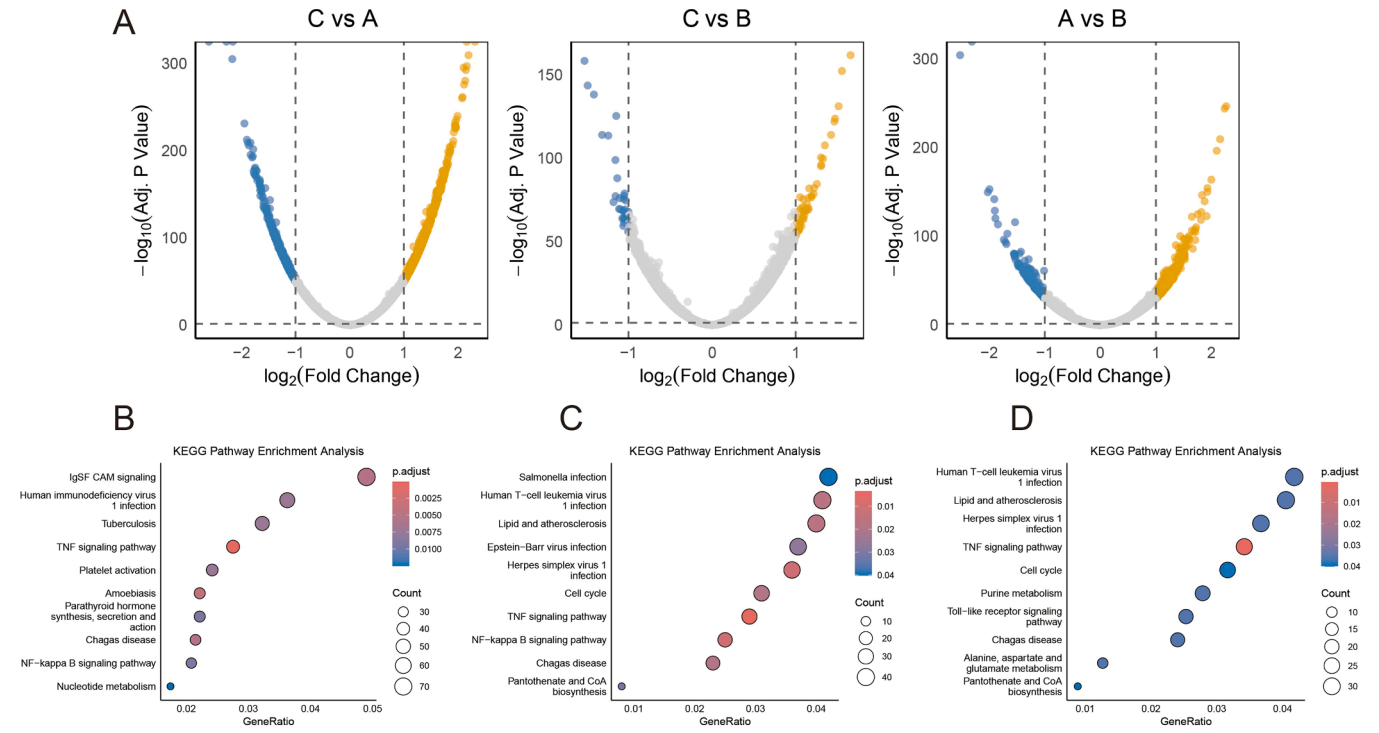

Fig. 7. Biological basis for prognostic differences among subtypes. (A) Differential gene expression analysis among subtypes A, B, and C. (B, C) Subtype C enriched in cancer-promoting pathways including TNF signaling, NF-κB pathway, and cell cycle regulation compared with subtypes A and B. (D) Subtype A enriched in cancer-promoting pathways compared with subtype B. TNF, Tumor Necrosis Factor; NF-κB, Nuclear Factor Kappa-Light-Chain-Enhancer of Activated B Cells.

across different regions. This proportion differed from the inconsistency rate reported by Hurson et al. recently [35], possibly due to the rigorous multi-regional sampling strategy (2–5 regions per patient). Luminal tumors, particularly the Luminal B subtype, showed the highest degree of heterogeneity. This supports the biological premise that Luminal B represents an intermediate state between hormone-driven and proliferation-driven phenotypes, rather than a stable category. Even in clinically diagnosed TNBC, significant transcriptomic variation exists. Existing tests primarily confirm known malignant features but fail to resolve the internal heterogeneity that drives differential treatment responses.

The effect of ITH on prognosis may depend on the tumor's underlying clinical subtype

Our analysis of the METABRIC cohort challenges the traditional assumption that high heterogeneity always equates to poor prognosis, revealing instead a context-dependent effect. In high-risk subtypes like Basal-like or HER2-enriched, the prognosis is primarily dictated by the highly invasive dominant clone, making the additional impact of ITH less detectable. In contrast, for the low-invasive Luminal A subtype, the presence of high ITH is associated with significantly worse survival. This suggests that ITH assessment is a critical risk modifier for traditionally low-risk patients, identifying individuals who may be underestimated by standard subtyping.

Stromal components contribute to transcriptomic variability

Identifying the sources of ITH is paramount for accurate molecular subtyping. While previous research has established that cancer cell evolution contributes to transcriptomic variability, our analysis reveals that stromal components within the TME are also important drivers of transcriptomic fluctuations. This is highly consistent with the tumor-stroma interface heterogeneity revealed by recent spatial transcriptomic studies [36]. High-ITH genes are markedly enriched in pathways related to stromal infiltration and EMT, which partially explains why bulk RNA-seq is susceptible to sampling bias. Conversely, low-ITH genes reflect the stable, intrinsic programs of cancer cells, such as cell cycle regulation and WNT signaling. By distinguishing between these two signals, this study demonstrates the feasibility of utilizing low-ITH genes to filter microenvironmental noise and extract the core biological features of the tumor.

A low-ITH gene-based framework enables robust molecular subtyping of breast cancer

The primary contribution of this study is the development of a robust transcriptomic classification framework based on genes with low-ITH. The advantages of this system are demonstrated across three dimensions. First, it exhibits stability and resistance to interference. Unlike traditional subtyping tools that are often compromised by microenvironmental variations across different anatomical sites, the subtypes identified in this study remain highly consistent across both spatial regions within the primary tumor and metastatic lesions, achieving robust subtyping independent of anatomical location. Second, it demonstrates cross-methods biological reproducibility. This approach successfully recapitulates the four functional malignant epithelial subtypes previously defined by scRNA-seq at the bulk RNA-seq, validating its precision in capturing core biological features. Third, it offers significant potential for clinical application. Addressing the bottlenecks of single-cell technologies—such as prohibitive costs and the loss of spatial context—our strategy provides a cost-effective alternative. It not only retains the crucial resolution, but also provides a potential tool for the large-scale implementation of precision medicine.

The low-ITH classification framework developed in this study preserves classical biological characteristics. First, the observed consistency aligns with the hypothesis that the tumor microenvironment is, at least partially, shaped by malignant epithelial cells, which themselves are capable of recapitulating PAM50 and single-cell intrinsic subtypes. This coherence is a prerequisite for successful classification across diverse

preclinical model systems. Our findings support the view that features associated with PAM50 or single-cell intrinsic subtypes provide a genuine phenotypic stratification of breast cancer; however, the exact cellular interactions defining these microenvironment-rich subtypes remain to be fully elucidated. Furthermore, pathway enrichment analysis reveals distinct biological essences underlying each subtype. Subtype C is significantly enriched in oncogenic pathways, such as TNF signaling, NF-κB signaling, and cell cycle regulation. The aberrant activity of these pathways provides a mechanistic explanation for its aggressive clinical phenotype. In contrast, Subtype A exhibits an inactive state across these pro-inflammatory and proliferative pathways. This divergence in molecular profiles precisely maps the heterogeneity of intrinsic programs within cancer cells. Finally, this biology-based classification system translates into robust clinical prognostic stratification. The three-subtype classification demonstrates significant prognostic power: Subtype C, characterized by the most aggressive molecular features, shows the poorest outcomes, whereas Subtype A, with suppressed oncogenic pathways, correlates with the highest overall survival.

Limitations

While this study establishes a robust framework for overcoming tumor heterogeneity, several limitations warrant acknowledgment. First, the sample size of our multi-region cohort remains relatively modest (36 patients, 123 samples). Although the findings have been validated using the extensive METABRIC dataset, the clinical utility of our classification system requires further verification in large-scale prospective clinical trials. Second, the ITH scoring system was originally developed based on microarray data. Despite validation via RNA-seq, potential systematic biases between different platforms may still affect gene comparability. Finally, our investigation focused primarily on the transcriptomic level. Future studies integrating genomic, epigenetic, and proteomic data will be essential to fully dissect the multi-dimensional architecture of ITH.

Conclusion

In summary, this study provides a new perspective for overcoming the interference of tumor heterogeneity in precision medicine and is expected to promote the transformation of breast cancer molecular typing towards a higher level of precision medicine.

Ethics approval and consent to participate

All studies were approved by their corresponding ethics review boards, and all subjects provided informed consent. This study used publicly available data without individual information, and thus no ethical approval was required.

Consent for publication

Not applicable.

Availability of data and material

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author/s.

Funding

This work was supported by the Yancheng Science and Technology Bureau (Grant No. YCBK2025033) and the Open Research Project of Key Laboratories of Jiangsu Province Universities (Grant No. XZSYSKF2025025).

CRediT authorship contribution statement

Jian Ji: Writing – original draft. **Qing Zheng:** Writing – original draft, Methodology, Conceptualization. **Xin Li:** Writing – original draft. **Can Luo:** Writing – original draft. **Jingzhi Wang:** Writing – review & editing.

Declaration of competing interest

All authors declare that there are no potential conflicts of interest that may affect the research results, and there are no commercial, financial, or other associated relationships involved in the research. This study did not receive any funding from institutions or individuals that may affect the research design.

Acknowledgements

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2026.102840.

References

[1] N. Andor, T.A. Graham, M. Jansen, L.C. Xia, C.A. Aktipis, C. Petritsch, H.P. Ji, C. C. Maley, Pan-cancer analysis of the extent and consequences of intratumor heterogeneity, Nat. Med. 22 (1) (2016) 105–113, https://doi.org/10.1038/nm.3984.

[2] N. Barzgar Barough, F. Sajjadian, N. Jalilzadeh, H. Shafaei, K. Velaei, Understanding breast cancer heterogeneity through non-genetic heterogeneity, Breast. Cancer 28 (4) (2021) 777–791, https://doi.org/10.1007/s12282-021-01237-w.

[3] Y. Liang, H. Zhang, X. Song, Q. Yang, Metastatic heterogeneity of breast cancer: molecular mechanism and potential therapeutic targets, Semin. Cancer Biol. 60 (2020) 14–27, https://doi.org/10.1016/j.semcancer.2019.08.012.

[4] J.R.M. Black, N. McGranahan, Genetic and non-genetic clonal diversity in cancer evolution, Nat. Rev. Cancer 21 (6) (2021) 379–392, https://doi.org/10.1038/s41568-021-00336-2.

[5] J. Househam, T. Heide, G.D. Cresswell, I. Spiteri, C. Kimberley, L. Zapata, C. Lynn, C. James, M. Mossner, J. Fernandez-Mateos, et al., Phenotypic plasticity and genetic control in colorectal cancer evolution, Nature 611 (7937) (2022) 744–753, https://doi.org/10.1038/s41586-022-05311-x.

[6] H. Shen, Q. Zheng, Z. Wang, D. Zheng, Z. Gong, H. Wang, T. Sun, J. Pan, Y. Jin, X. Zheng, et al., Breaking the heterogeneity barrier: a robust prognostic signature for survival stratification and immune profiling in triple-negative breast cancer, Front. Immunol. 16 (2025) 1611917, https://doi.org/10.3389/fimmu.2025.1611917.

[7] H. Shen, Q. Zheng, J. Pan, Y. Jin, X. Zheng, Q. Yuan, D. Tan, Q. Zhou, J. Wang, T. Sun, Intra-tumor heterogeneity-resistant gene signature predicts prognosis and immune infiltration in breast cancer, Front. Immunol. 16 (2025) 1598858, https://doi.org/10.3389/fimmu.2025.1598858.

[8] K. Ganesh, H. Basnet, Y. Kaygusuz, A.M. Laughney, L. He, R. Sharma, K. P. O'Rourke, V.P. Reuter, Y.H. Huang, M. Turkekul, et al., L1CAM defines the regenerative origin of metastasis-initiating cells in colorectal cancer, Nat. Cancer 1 (1) (2020) 28–45, https://doi.org/10.1038/s43018-019-0006-x.

[9] J.S. Parker, M. Mullins, M.C. Cheang, S. Leung, D. Voduc, T. Vickery, S. Davies, C. Fauron, X. He, Z. Hu, et al., Supervised risk predictor of breast cancer based on intrinsic subtypes, J. Clin. Oncol.: Off J. Am. Soc. Clin. Oncol. 27 (8) (2009) 1160–1167, https://doi.org/10.1200/jco.2008.18.1370.

[10] S. Paik, S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F.L. Baehner, M.G. Walker, D. Watson, T. Park, et al., A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer, N. Engl. J. Med. 351 (27) (2004) 2817–2826, https://doi.org/10.1056/NEJMoa041588.

[11] L.J. van 't Veer, H. Dai, M.J. van de Vijver, Y.D. He, A.A. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M.J. Marton, A.T. Witteveen, et al., Gene expression profiling predicts clinical outcome of breast cancer, Nature 415 (6871) (2002) 530–536, https://doi.org/10.1038/415530a.

[12] P.P.M. Berton Giachetti, A. Carnevale Schianca, D. Trapani, A. Marra, A. Toss, C. Marchiò, M.V. Dieci, O.D. Gentilini, C. Criscitiello, K. Kalinsky, et al., Current controversies in the use of oncotype DX in early breast cancer, Cancer Treat. Rev. 135 (2025) 102887, https://doi.org/10.1016/j.ctrv.2025.102887.

[13] G. Kunz, Use of a genomic test (MammaPrint™) in daily clinical practice to assist in risk stratification of young breast cancer patients, Arch. Gynecol. Obs. 283 (3) (2011) 597–602, https://doi.org/10.1007/s00404-010-1454-9.

[14] S. Chowdhury, M. Hofree, K. Lin, D. Maru, S. Kopetz, J.P. Shen, Implications of intratumor heterogeneity on consensus molecular subtype (CMS) in colorectal cancer, Cancers 13 (19) (2021) 4923, https://doi.org/10.3390/cancers13194923.
[15] K. Sirinukunwattana, E. Domingo, S.D. Richman, K.L. Redmond, A. Blake, C. Verrill, S.J. Leedham, A. Chatzipli, C. Hardy, C.M. Whalley, et al., Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning, Gut 70 (3) (2021) 544–554, https://doi.org/10.1136/gutjnl-2019-319866.
[16] S.Z. Wu, G. Al-Eryani, D.L. Roden, S. Junankar, K. Harvey, A. Andersson, A. Thennavan, C. Wang, J.R. Torpy, N. Bartonicek, et al., A single-cell and spatially resolved atlas of human breast cancers, Nat. Genet. 53 (9) (2021) 1334–1347, https://doi.org/10.1038/s41588-021-00911-1.
[17] R. Dienstmann, G. Villacampa, A. Sveen, M.J. Mason, D. Niedzwiecki, A. Nesbakken, V. Moreno, R.S. Warren, R.A. Lothe, J. Guinney, Relative contribution of clinicopathological variables, genomic markers, transcriptomic subtyping and microenvironment features for outcome prediction in stage II/III colorectal cancer, Ann. Oncol.: Off J. Eur. Soc. Med. Oncol. 30 (10) (2019) 1622–1629, https://doi.org/10.1093/annonc/mdz287.
[18] Q. Zheng, Z. Gong, B. Li, H. Wang, S. Lin, Development and validation of the antibody-dependent cellular phagocytosis-based signature: a prognostic risk model of gastric cancer, Adv. Clin. Exp. Med.: off Organ Wroc Med. Univ. 34 (3) (2025) 433–446, https://doi.org/10.17219/acem/189914.
[19] Q. Zheng, Z. Gong, S. Lin, D. Ou, W. Lin, P. Shen, Integrated analysis of a competing endogenous RNA network reveals a ferroptosis-related 6-lncRNA prognostic signature in clear cell renal cell carcinoma, Adv. Clin. Exp. Med: Off Organ Wroc Med. Univ. 33 (12) (2024) 1391–1407, https://doi.org/10.17219/acem/176050.
[20] H.O. Lee, Y. Hong, H.E. Etlioglu, Y.B. Cho, V. Pomella, B. Van den Bosch, J. Vanhecke, S. Verbandt, H. Hong, J.W. Min, et al., Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer, Nat. Genet. 52 (6) (2020) 594–603, https://doi.org/10.1038/s41588-020-0636-z.
[21] P.D. Dunne, M. Alderdice, P.G. O'Reilly, A.C. Roddy, A.M.B. McCorry, S. Richman, T. Maughan, S.S. McDade, P.G. Johnston, D.B. Longley, et al., Cancer-cell intrinsic gene expression signatures overcome intratumoural heterogeneity bias in colorectal cancer patient classification, Nat. Commun. 8 (2017) 15657, https://doi.org/10.1038/ncomms15657.
[22] D. Lähnemann, J. Köster, E. Szczurek, D.J. McCarthy, S.C. Hicks, M.D. Robinson, C. A. Vallejos, K.R. Campbell, N. Beerenwinkel, A. Mahfouz, et al., Eleven grand challenges in single-cell data science, Genome Biol. 21 (1) (2020) 31, https://doi.org/10.1186/s13059-020-1926-6.
[23] H.S.H. Lau, V.K.M. Tan, B.K.T. Tan, Y. Sim, J. Quist, A.A. Thike, P.H. Tan, S. Pervaiz, A. Grigoriadis, K. Sabapathy, Adipose-enriched peri-tumoral stroma, in contrast to myofibroblast-enriched stroma, prognosticates poorer survival in breast cancers, NPJ Breast. Cancer 9 (1) (2023) 84, https://doi.org/10.1038/s41523-023-00590-7.
[24] W.T. Barry, D.N. Kernagis, H.K. Dressman, R.J. Griffis, J.D. Hunter, J.A. Olson, J. R. Marks, G.S. Ginsburg, P.K. Marcom, J.R. Nevins, et al., Intratumor heterogeneity and precision of microarray-based predictors of breast cancer biology and clinical outcome, J. Clin. Oncol.: Off J. Am. Soc. Clin. Oncol. 28 (13) (2010) 2198–2206, https://doi.org/10.1200/jco.2009.26.7245.
[25] Q. Xu, J. Kaur, D. Wylie, K. Mittal, H. Li, R. Kolachina, M. Aleskandarany, M. S. Toss, A.R. Green, J. Yang, et al., A case series exploration of multi-regional expression heterogeneity in triple-negative breast cancer patients, Int. J. Mol. Sci. 23 (21) (2022) 13322, https://doi.org/10.3390/ijms232113322.
[26] C. Curtis, S.P. Shah, S.F. Chin, G. Turashvili, O.M. Rueda, M.J. Dunning, D. Speed, A.G. Lynch, S. Samarajiwa, Y. Yuan, et al., The genomic and transcriptomic architecture of 2000 breast tumours reveals novel subgroups, Nature 486 (7403) (2012) 346–352, https://doi.org/10.1038/nature10983.
[27] D.M. Gendoo, N. Ratanasirigulchai, M.S. Schröder, L. Paré, J.S. Parker, A. Prat, B. Haibe-Kains, Genefu: an R/bioconductor package for computation of gene expression-based signatures in breast cancer, Bioinformatics 32 (7) (2016) 1097–1099, https://doi.org/10.1093/bioinformatics/btv693.
[28] K. Yoshihara, M. Shahmoradgoli, E. Martínez, R. Vegesna, H. Kim, W. Torres-Garcia, V. Treviño, H. Shen, P.W. Laird, D.A. Levine, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, Nat. Commun. 4 (2013) 2612, https://doi.org/10.1038/ncomms3612.
[29] A.M. Newman, C.L. Liu, M.R. Green, A.J. Gentles, W. Feng, Y. Xu, C.D. Hoang, M. Diehn, A.A. Alizadeh, Robust enumeration of cell subsets from tissue expression profiles, Nat. Methods 12 (5) (2015) 453–457, https://doi.org/10.1038/nmeth.3337.
[30] E. Becht, N.A. Giraldo, L. Lacroix, B. Buttard, N. Elarouci, F. Petitprez, J. Selves, P. Laurent-Puig, C. Sautès-Fridman, W.H. Fridman, et al., Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression, Genome Biol. 17 (1) (2016) 218, https://doi.org/10.1186/s13059-016-1070-5.
[31] P. Jiang, S. Gu, D. Pan, J. Fu, A. Sahu, X. Hu, Z. Li, N. Traugh, X. Bu, B. Li, et al., Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response, Nat. Med. 24 (10) (2018) 1550–1558, https://doi.org/10.1038/s41591-018-0136-1.
[32] P. Das, G.M. Siegers, L.M. Postovit, Illuminating luminal B: QSOX1 as a subtype-specific biomarker, Breast Cancer Res.: BCR 15 (3) (2013) 104, https://doi.org/10.1186/bcr3417.
[33] K. Asleh, A. Lluch, A. Goytain, C. Barrios, X.Q. Wang, L. Torrecillas, D. Gao, M. Ruiz-Borrego, S. Leung, J. Bines, et al., Triple-negative PAM50 non-basal breast cancer subtype predicts benefit from extended adjuvant capecitabine, Clin. Cancer Res.: Off J. Am. Assoc. Cancer Res. 29 (2) (2023) 389–400, https://doi.org/10.1158/1078-0432.Ccr-22-2191.
[34] M. Greaves, C.C. Maley, Clonal evolution in cancer, Nature 481 (7381) (2012) 306–313, https://doi.org/10.1038/nature10762.
[35] A.N. Hurson, A.M. Hamilton, L.T. Olsson, E.L. Kirk, M.E. Sherman, B.C. Calhoun, J. Geradts, M.A. Troester, Reproducibility and intratumoral heterogeneity of the PAM50 breast cancer assay, Breast. Cancer Res. Treat. 199 (1) (2023) 147–154, https://doi.org/10.1007/s10549-023-06888-1.
[36] N. Sharma, J. Rájová, G. Mermelekas, K. Thrane, J. Lundeberg, A. Shamikh, S. Vikström, H. Babačić, M. Jensdottir, J. Lehtiö, et al., In-depth patient-specific analysis of tumor heterogeneity in melanoma brain metastasis: insights from spatial transcriptomics and multi-region bulk sequencing, Transl. Oncol. 59 (2025) 102468, https://doi.org/10.1016/j.tranon.2025.102468.