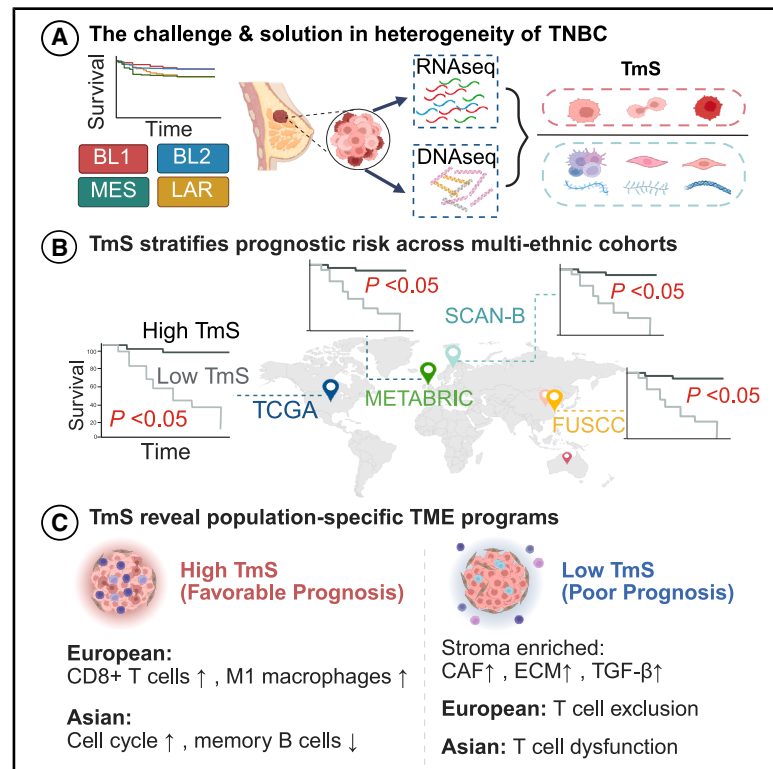


Tumor microenvironment transcriptional activity enables robust stratification of chemotherapy response in triple-negative breast cancer

Graphical abstract



Authors

Yaoyi Dai, Xiaoxi Pan, Shuai Guo, ..., Bora Lim, Yinyin Yuan, Wenyi Wang

Correspondence

wwang7@mdanderson.org

In brief

Dai et al. demonstrate that tumor-specific total mRNA expression (TmS) outperforms existing classifications in stratifying triple-negative breast cancer chemotherapy responses across diverse populations. This deconvolution approach reveals population-specific tumor microenvironment programs and identifies extracellular matrix targeting as a therapeutic strategy for patients with low-TmS.

Highlights

- TmS subtypes outperform existing TNBC subtypes in predicting chemotherapy response
- Stromal barriers and immune exclusion define treatment-resistant low-TmS tumors
- Population-specific immune patterns distinguish Asian and European low-TmS tumors
- ECM signatures represent conserved therapeutic targets across all populations



Article

Tumor microenvironment transcriptional activity enables robust stratification of chemotherapy response in triple-negative breast cancer

Yaoyi Dai,^{1,2} Xiaoxi Pan,^{3,4} Shuai Guo,¹ Shuangxi Ji,¹ Shaolong Cao,^{1,16} Matthew D. Montierth,^{1,2} Yujie Jiang,^{1,17} Jeffrey T. Chang,⁵ Leming Shi,⁶ Shabnam Shalapour,⁷ Gloria V. Echeverria,^{8,9,10,11} Lucy Yates,¹² Johan Staaf,¹³ Bora Lim,^{14,15} Yinyin Yuan,^{3,4} and Wenyi Wang^{1,18,*}

¹Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

²Graduate Program of Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, TX 77030, USA

³Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

⁴Institute for Data Science in Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

⁵McGovern Medical School, Houston, TX 77030, USA

⁶State Key Laboratory of Genetic Engineering, School of Life Sciences and Human Phenome Institute, Fudan University, Shanghai 200093, China

⁷Department of Cancer Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

⁸Department of Medicine, Baylor College of Medicine, Houston, TX 77030, USA

⁹Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA

¹⁰Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA

¹¹Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, TX 77030, USA

¹²Wellcome Sanger Institute, Hinxton CB10 1SA, UK

¹³Division of Translational Cancer Research, Department of Laboratory Medicine, Lund University, Medicon Village, 223 81 Lund, Sweden

¹⁴Department of Breast Medical Oncology, The Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

¹⁵The Morgan Welch IBC Research Program and Clinic, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

¹⁶Present address: Biogen Inc., Cambridge, MA 02142, USA

¹⁷Present address: AbbVie Inc., North Chicago, IL 60064, USA

¹⁸Lead contact

*Correspondence: wwang7@mdanderson.org

<https://doi.org/10.1016/j.xcrm.2026.102610>

SUMMARY

Triple-negative breast cancer (TNBC) exhibits heterogeneous treatment responses, yet molecular subtypes based on predefined biological pathways show limited prognostic value. We introduce tumor-specific total mRNA expression (TmS), a pathway-agnostic deconvolution metric derived from matched RNA/DNA sequencing, as a robust stratification tool. Analyzing 575 TNBC patients across Western and East Asian populations, TmS outperforms established subtypes in predicting chemotherapy outcomes, stratifying patients into high TmS with favorable prognosis and low TmS with poor prognosis. Stromal enrichment with immune exclusion emerges as a universal feature of chemotherapy-resistant low-TmS tumors across all cohorts. Population-specific features distinguish Asian cohorts: high-TmS tumors exhibit cell cycle-driven proliferation programs, and low-TmS tumors display immune dysfunction with memory B cell enrichment and divergent RAS/mitogen-activated protein kinase (MAPK) activation, compared to Western populations. Despite these differences, extracellular matrix organization represents a conserved therapeutic vulnerability in treatment-resistant low-TmS patients. TmS provides a unifying framework for dissecting TNBC heterogeneity and enabling precision therapy across diverse populations.

INTRODUCTION

Triple-negative breast cancer (TNBC), which lacks expression of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2),^{1–3} accounts for 15%–20% of all breast cancers.⁴ TNBC has the worst prognosis across diverse ethnic groups.^{4–6} While immunotherapy and

PARP inhibitors have been integrated into standard-of-care practices, their effectiveness varies considerably across patient subgroups due to TNBC's remarkable molecular and cellular heterogeneity.^{7–9} Further efforts are needed to unravel the molecular underpinnings of TNBC heterogeneity, identify actionable molecular targets, and develop effective patient stratification strategies to guide personalized treatment approaches.



The recently developed classification strategies for TNBC include transcriptome-based classifications (e.g., Lehmann's TNBCtype-6,¹⁰ TNBCtype-4,¹¹ Burstein's four TNBC groups,¹² and gene expression modules¹³), DNA repair deficiency metrics (homologous recombination deficiency [HRD]; HRD scores),¹⁴ and cellular activity assessments (stromal-derived TGF- β signatures¹⁵ and immune cell prevalence¹⁶). Each approach has provided valuable insights into specific aspects of TNBC biology but none provides a unified framework for clinical decision-making. One potential limitation is that these methods examine single molecular characteristics or predefined gene set programs in isolation, failing to capture the integrated dynamics between tumor cells and their surrounding microenvironment. The concerted efforts across many cell types within the tumor microenvironment (TME) play a major role on the tumor cell transcriptional activity,^{17–20} and all cell-type-specific activities collectively determine therapeutic response.^{16,21–24} While single-cell technologies can offer high-resolution insights into these co-occurring programs, their clinical utility remains constrained by small sample sizes and technical variability.^{25,26} Therefore, we propose that an integrative bulk RNA/DNA sequencing (DNA-seq)-based deconvolution metric for transcriptomic plasticity, tumor-specific total mRNA expression (TmS),²⁷ which has been experimentally benchmarked and shown to have pan-cancer biological utility, may provide a powerful solution for patient stratification at scale by comprehensively characterizing TNBC heterogeneity (Figure 1A).

For each tumor sample, TmS calculates the ratio of tumor cell to non-tumor cell (stromal plus immune cells) total mRNA expression levels per cell per haploid genome, thereby simultaneously quantifying both tumor cell transcriptional output and the collective transcriptional activity of the surrounding microenvironment. Our previous pan-cancer study²⁷ demonstrated that TmS measures tumor transcriptome ploidy, analogous to tumor genome ploidy. Transcriptome ploidy represents an important feature of tumor plasticity, which was previously overlooked due to a lack of computational technique to measure it using high-throughput sequencing data from matched DNA/RNA. Therefore, calculating and interpreting TmS for patient tumor samples, both retrospectively and prospectively, represents a promising approach to identify cancer-specific biological mechanisms and inform therapeutic strategies.³⁰

In this study, we performed large-scale multi-omics profiling of 599 TNBC patients across four ethnically diverse cohorts to improve TNBC stratification and therapeutic insights. Specifically, we aim to (1) validate TmS's prognostic value in modern clinical cohorts from European (the Sweden Cancerome Analysis Network – Breast [SCAN-B]³¹) and East Asian (the Fudan University Shanghai Cancer Center [FUSCC]³²) populations, comparing its performance against TNBCtypes and evaluating concordance with discovery cohorts²⁷ (the Cancer Genome Atlas [TCGA] projects³³ and the Molecular Taxonomy of Breast Cancer International Consortium [METABRIC]³⁴); (2) validate TmS-associated stromal and immune activities using digital pathology; (3) delineate co-occurring molecular and cellular features of TmS-defined subtypes across ethnic groups (Figure 1B), with validation using single-cell RNA sequencing (RNA-seq) from two East Asian cohorts (Figure 1C); and (4) identify candidate

therapeutic targets for chemotherapy-resistant low-TmS tumors. This multi-cohort approach demonstrates that TmS effectively stratifies patients and reveals population-specific TME programs without requiring *a priori* specification of individual gene sets or pathways.

RESULTS

TmS estimation in multi-ethnic TNBC patients before chemotherapy

We calculated TmS from four TNBC cohorts with matched DNA-seq/RNA-seq and >5-year follow-up (STAR Methods): TCGA ($n = 83$), METABRIC ($n = 118$), SCAN-B ($n = 144$), and FUSCC ($n = 230$) (Figure 1C). These cohorts represent diverse ethnic populations: TCGA comprises mixed ancestry (55% European, 25.3% African, and 19.3% other), METABRIC consists of European patients, SCAN-B is a population-based study of mainly Swedish patients, and FUSCC represents East Asian (Chinese) patients (Table S1). These cohorts span different time periods with varying survival outcomes: the TCGA cohort (2006–2016) showed a median disease-free survival (DFS) of 1.79 years, the historical METABRIC (1999–2010) showed a median relapse-free survival (RFS) of 1.57 years, the SCAN-B cohort (2010–2015) reported a median relapse-free interval (RFI) of 3.61 years, and the FUSCC cohort (2007–2014) showed a median RFS of 3.24 years. Chemotherapy treatment protocols differed across cohorts (Table S1): SCAN-B patients predominantly received anthracycline-taxane combination regimens, while FUSCC patients received primarily taxane-based regimens, reflecting Asian treatment preferences. The improved survival outcomes in more recent cohorts likely reflect advancements in treatment efficacy and supportive care over time.^{32,35}

TmS values vary significantly across and within the four cohorts (Figure 1D, one-way ANOVA test, $p < 0.001$). TCGA and METABRIC cohorts showed comparable distributions (Figure 1D, mean $\pm$ SD: 3.87 ± 3.25 and 3.68 ± 2.38 , respectively). In contrast, the more recent cohorts SCAN-B and FUSCC showed lower mean values, with FUSCC exhibiting higher variation (Figure 1D, mean $\pm$ SD: 2.14 ± 1.17 and 2.74 ± 4.35 , respectively). The cohort-level TmS differences likely reflect complex interactions between technical factors (sequencing platforms and library preparation protocols), population genetics, and clinical practice variations (Table S1). Mathematically, TmS values may still be affected by a study-specific scalar that represents an interplay between technical and biological variations of each study (STAR Methods). In the following analyses, we performed within-cohort patient stratification to evaluate whether TmS predicts clinical outcomes independent of these technical confounders. Therefore, the numerical threshold defining high versus low TmS was determined separately for each cohort based on the range of the TmS values per study (Table S1).

TmS stratifies prognosis after chemotherapy across multi-ethnic cohorts (1999–2016)

To establish the clinical relevance of TmS for TNBC, we first examined whether TmS associates with molecular subtypes of TNBC in a similar trend across all cohorts (Figure 2A). Indeed, tumors with prognostically favorable subtypes, such as basal-like

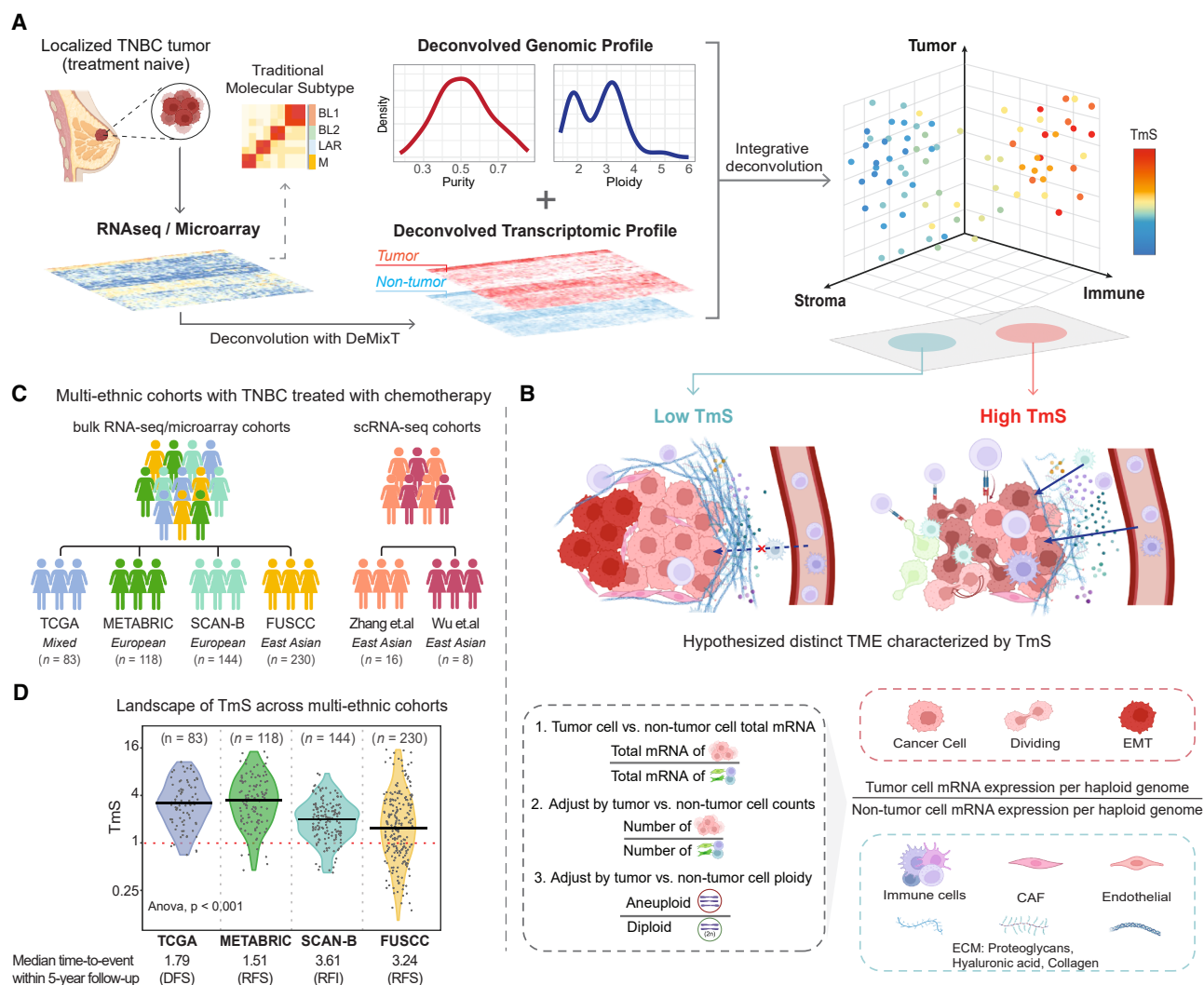


Figure 1. Study design to uncover the tumor microenvironment of TNBC through TmS deconvolution

(A) Study overview: in contrast to traditional molecular subtyping method, tumor-specific total mRNA expression (TmS) is derived from integrative deconvolution, reflecting the dynamic RNA/DNA interplay between tumor cells, stroma, and immune components.

(B) Hypothesized TME dynamics: low TmS (e.g., stroma-dominant, potentially immunosuppressive) vs. high TmS (e.g., enhanced immune activation). TmS derivation integrates tumor cellularity and ploidy in a joint RNA-seq/DNA-seq deconvolution framework.

(C) Cohort composition: four bulk RNA-seq/microarray cohorts with triple-negative breast cancer (TNBC) treated with chemotherapy: The Cancer Genome Atlas (TCGA), Molecular Taxonomy of Breast Cancer International Consortium (METABRIC), Sweden Cancerome Analysis Network – Breast (SCAN-B), and Fudan University Shanghai Cancer Center (FUSCC). Two single-cell RNA-seq cohorts provide independent validation of East Asian-specific TmS mechanisms: Zhang et al.²⁸ (Chinese patients, $n = 16$) and Wu et al.²⁹ (East Asian patients, $n = 8$).

(D) TmS landscape in 575 TNBC patients with bulk RNA-seq: violin plots show TmS distributions for each cohort (n noted above each plot). Black horizontal lines mark median TmS per cohort. Median time-to-event within 5-year follow-up shown below each cohort represents the median time to relapse/progression among patients who experienced an event within 5 years: disease-free survival (DFS) for TCGA, relapse-free survival (RFS) for METABRIC and FUSCC, and relapse-free interval (RFI) for SCAN-B. ANOVA ($p < 0.001$) indicates significant variation in TmS across cohorts.

1 (BL1) and immunomodulatory (IM), consistently exhibited higher TmS values, whereas unfavorable subtypes, such as mesenchymal (MES), displayed lower TmS values (TCGA: adjusted $p < 0.05$; METABRIC: adjusted $p = 0.08$; SCAN-B: adjusted $p < 0.001$; FUSCC: adjusted $p < 0.001$). Our previous pan-cancer study identified TmS to uniquely stratify TNBC patients in TCGA and METABRIC,²⁷ with high TmS significantly associated with favorable outcomes after chemotherapy, an

opposite trend to 12 other cancer types. Therefore, the observed consistency in molecular subtype associations across all four cohorts provides support that the prognostic effect of TmS may be replicated in the more modern cohorts of SCAN-B and FUSCC and that its unique direction may shed light on TNBC biology.

We evaluated the predictive performance of TmS in disease-free or relapse-free survival, in comparison to established

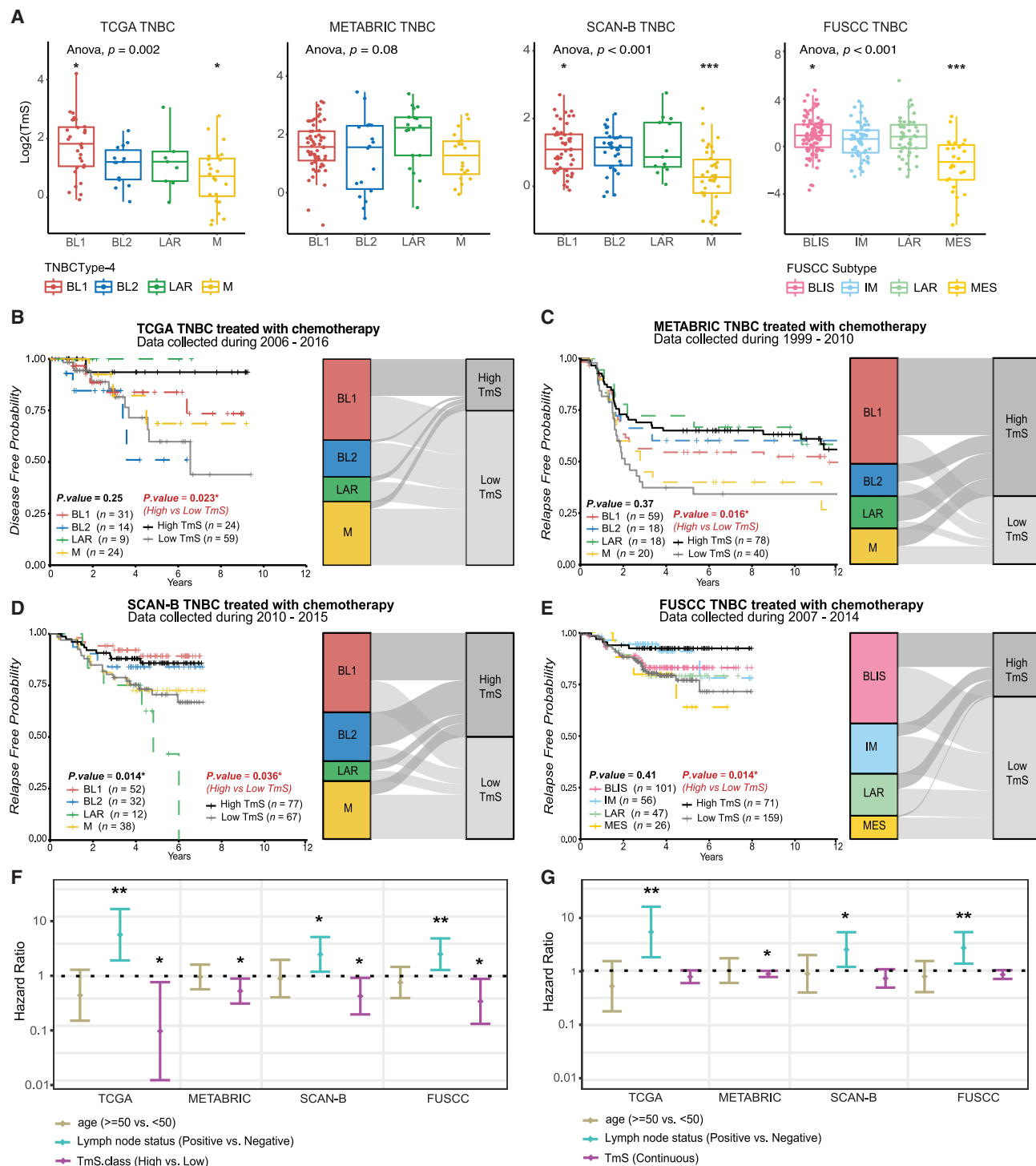


Figure 2. TmS is associated with known TNBC subtypes and stratifies patients' prognostic outcomes with chemotherapy treatment

(A) Distributions of TmS across TNBCtype-4 in TCGA, METABRIC, SCAN-B, and across FUSCC subtypes in FUSCC. Center line represents median; box represents interquartile range (IQR); whiskers extend to $1.5 \times$ IQR. * indicates statistically significant differences between each subtype and other subtypes combined using Tukey's HSD test.

(B–D) Kaplan-Meier curves stratified by TmS (high versus low) and TNBCtype-4 subtypes across TCGA (B), METABRIC (C), SCAN-B (D).

(legend continued on next page)

TNBC classification systems: Lehmann's TNBCtype-4 classification in TCGA, METABRIC, and SCAN-B cohorts, and FUSCC TNBC subtypes within the FUSCC cohort (STAR Methods). The survival benefit associated with high TmS is consistent across all cohorts (Figures 2B–2E, all log rank $p < 0.05$). In contrast, TNBCtype-4 subgroups of the same patients showed no significant differences in survival outcomes across all cohorts. This cross-cohort consistency in TmS stratification is notable given the heterogeneity across cohorts discussed earlier (Tables S1), including population-specific treatment preferences: East Asian patients (FUSCC) predominantly received taxane-based regimens (79.1%), while European patients (SCAN-B) received anthracycline-taxane combinations (97.2% anthracyclines, 87.5% taxanes). TmS stratification remained significant in treatment-specific subsets (anthracycline-taxane in SCAN-B: $n = 91$, log rank $p = 0.04$; taxane-based in FUSCC: $n = 182$, log rank $p = 0.023$; Figures S1A and S1B).

To estimate the relative contribution of TmS compared to the current clinical and pathological features, we performed multivariable Cox proportional hazards regression using the cohort-specific survival outcomes, adjusting for age, lymph node status, tumor size, and tumor grade. When analyzed as a binary categorical variable, high TmS consistently correlated with a lower risk of adverse events, indicating a 43%–90% risk reduction (Figure 2F; TCGA: hazard ratio [HR] = 0.1, 95% confidence interval [CI]: 0.01–0.77; METABRIC: HR = 0.53, 95% CI: 0.3–0.89; SCAN-B: HR = 0.43, 95% CI: 0.2–0.93; FUSCC: HR = 0.34, 95% CI: 0.13–0.89). Similar trends were observed when TmS was analyzed as a continuous variable (Figure 2G; TCGA: HR = 0.77, 95% CI: 0.59–1.01; METABRIC: HR = 0.87, 95% CI: 0.76–0.99; SCAN-B: HR = 0.72, 95% CI: 0.49–1.06; FUSCC: HR = 0.85, 95% CI: 0.70–1.02).

The distinction between TmS and established TNBC classification systems is evident from their patient-level assignment patterns (Figures 2B–2E, alluvial plots). The high- and low-TmS patient groups do not correspond to any aggregation of TNBCtype-4 subgroups or FUSCC subtypes into two groups (Table S2). In TCGA and FUSCC, high-TmS patients primarily comprised a subset of BL1 or basal-like immune-suppressed (BLIS) subtypes plus a small group of patients from each of the other subtypes. The discordance with the FUSCC subtypes is further increased as we divide the non-high-TmS patients further into two categories: medium and low TmS, with each group consisting of patients from each of the FUSCC subtypes. This three-category TmS stratification also demonstrates increased statistical significance in their prognosis differences (Figures S1C and S1D, log rank $p = 0.011$, STAR Methods). We therefore use three-category TmS stratification for FUSCC in downstream analyses.

In summary, our findings suggest that TmS is an independent and robust prognostic biomarker that outperforms existing TNBC classifications, including population-specific systems like FUSCC subtypes, in stratifying patients' responses to chemotherapy. The distinctions between TmS and previous TNBC classification systems suggest that a potential set of biological activities may be uncovered by contrasting TmS subtypes, as TmS captures transcriptome-wide activities across different cell types in the TME.

Contrasting TmS subtypes reveal distinct spatial tumor-stromal contact

Using available histology imaging data from SCAN-B and TCGA, we explored the TmS subtype-associated biological activities at the cellular level. Single-cell studies have documented notable differences in transcriptome size between immune cells (~1,000–1,300 expressed genes per cell) and stromal cells (~2,000–3,000 expressed genes per cell).^{27,36–38} Therefore, we hypothesize that TNBC tumors with high TmS may reflect immune-rich microenvironments (i.e., a small denominator driven by immune cells with compact transcriptomes), whereas tumors with low TmS may indicate stromal-dominated niches (i.e., a large denominator driven by stromal cells with expansive transcriptional programs, Figure 1B).

In SCAN-B, high-TmS tumors showed significantly higher tumor-infiltrating lymphocytes (TILs estimated by H&E images³⁹) as compared to low-TmS tumors (Figure 3A, two-sided Wilcoxon rank-sum test, $p < 0.001$), consistent with their improved relapse-free probabilities (Figure 2E). In TCGA, where raw high-resolution H&E image data are available, we performed digital segmentation of 78 whole H&E slides from TNBC using a deep learning model optimized for breast tumor morphology⁴⁰ (Figures 3B and 3C, STAR Methods). Low-TmS tumors exhibited significantly higher stromal content (median ± median absolute deviation [MAD] pixel counts: 38.9% ± 16.5% vs. 28.1% ± 13.5%, two-sided Wilcoxon rank-sum test $p < 0.05$) and lower immune cell infiltration (median ± MAD pixel counts: 1.2% ± 1.4% vs. 3.1% ± 3.2%, two-sided Wilcoxon rank-sum test $p < 0.01$), while the percentage of tumor tissue remained comparable between groups (Figure 3D).

The H&E slides in TCGA also allow us to study the spatial organization of cellular components, which is expected to critically influence tumor progression and therapeutic response.^{21,22,24,41,42} Using a joint-count analysis to quantify pixel-pixel contacts (Figure 3C, STAR Methods), we found that high-TmS tumors displayed significantly increased tumor-immune cell contacts (median ± MAD contact frequency percentage: 0.83% ± 0.86% vs. 0.27% ± 0.27%, two-sided Wilcoxon rank-sum test $p < 0.01$, Figure 3E), while low-TmS tumors exhibited significantly higher tumor-stromal contacts (median ± MAD contact frequency

(E) Kaplan-Meier curves stratified by TmS (high versus low) and FUSCC TNBC subtypes in the FUSCC cohort. (B–E) Alluvial diagrams of the same patients who are stratified by either TmS or the TNBCtype-4 subtypes within each cohort.

(F and G) Forest plots of hazard ratios (HRs) (center points) and 95% CIs (error bars) of multivariate Cox proportional hazard models within each cohort. Age (≥ 50 versus < 50), lymph node status (Positive versus Negative) and TmS (High versus Low) (F) or continuous TmS (G) as predictors for TCGA, SCAN-B, and FUSCC cohorts. METABRIC cohort only includes age and TmS (High versus Low) (F) or continuous TmS (G) as predictors. For (F) and (G), p values of two-sided Wald tests for the covariates are indicated by asterisks. For all p values, significance levels are denoted as follows: * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Only patients who were treated with chemotherapy are shown in all panels (see Table S1 for treatment details).

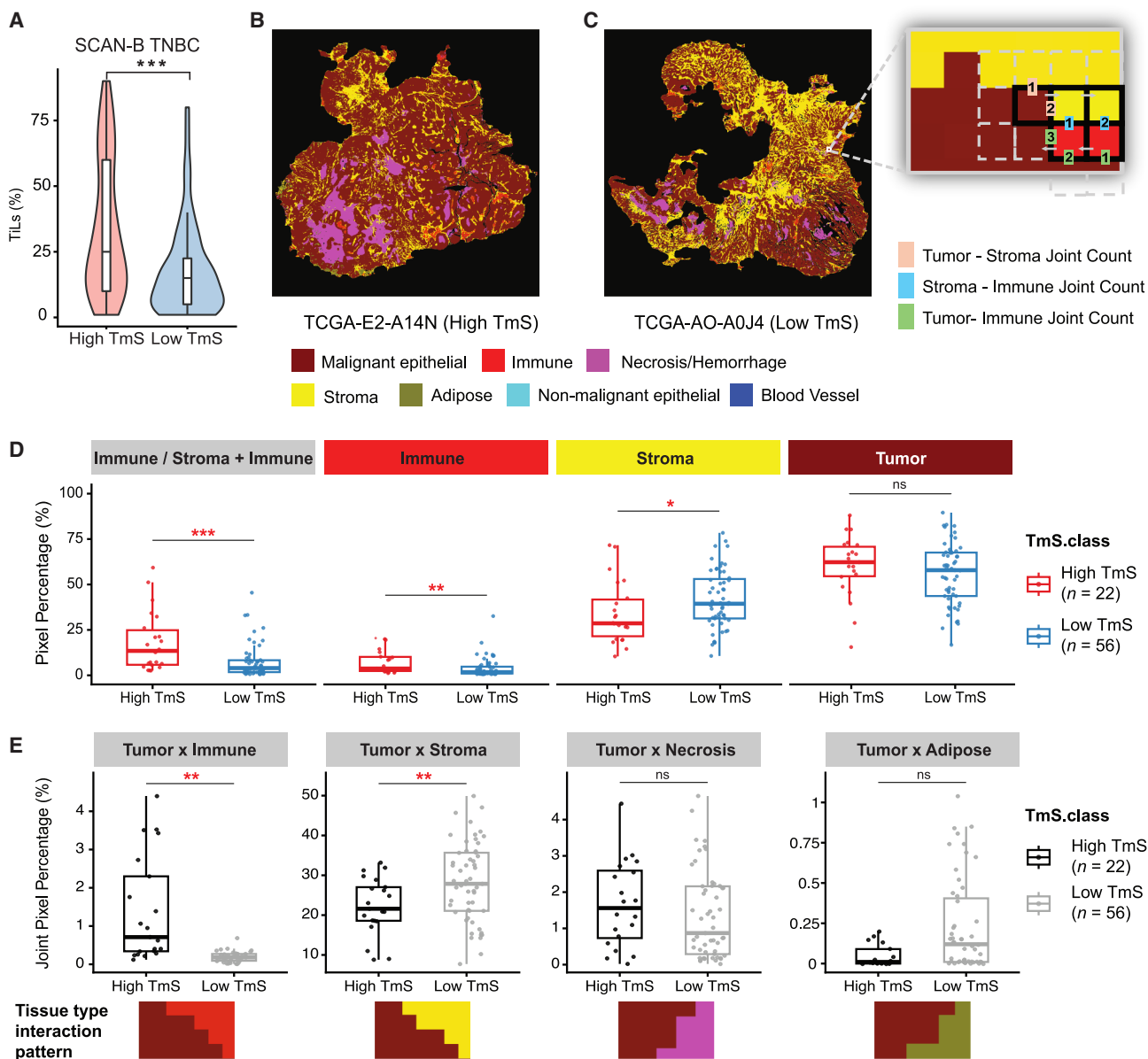


Figure 3. Digital pathology identifies a dynamic microenvironment associated with TmS

(A) Distribution of tumor-infiltrating lymphocyte (TIL) percentage across TmS category in SCAN-B. Center line represents median; box represents IQR. (B) Representation of tissue types in a tumor sample with high TmS, delineating malignant epithelial cells (dark red), immune cells (red), necrosis/hemorrhage (magenta), stromal components (yellow), adipose tissue (olive), non-malignant epithelial (cyan), and blood/blood vessel (blue). (C) Cell type distribution in a tumor sample with low TmS, employing the same color key as in (A); with a zoomed-in screenshot of a segmented H&E image, illustrating the methodology for quantifying joint frequency between different tissue types at the pixel level. (D) Boxplots of the percent pixels that are annotated as tumor, stroma, or inflammatory cells, and stroma TILs (Immune/Stroma+Immune) between high- and low-TmS groups, including corresponding *p* values for statistical relevance. (E) Boxplots detailing the joint pixel percentages for the tumor's interaction with inflammatory cells, stroma, necrosis/hemorrhage and adipose tissue. For (D) and (E), center line represents median; box represents IQR; whiskers extend to $1.5 \times$ IQR; individual samples shown as dots. The Benjamini-Hochberg (BH)-adjusted *p* values for two-sided Wilcoxon rank-sum tests are indicated by asterisks. For all *p* values, significance levels are denoted as follows: **p* < 0.05, ***p* < 0.01, and ****p* < 0.001.

percentage: $27.86\% \pm 10.54\%$ vs. $22.31\% \pm 6.71\%$, two-sided Wilcoxon rank-sum test *p* < 0.05, Figure 3E). The predominant tumor-stromal interface in low-TmS tumors likely promotes im-

mune cell exclusion, as described previously,⁴³ thereby potentially limiting the efficacy of immunotherapy in these chemotherapy-resistant cases.

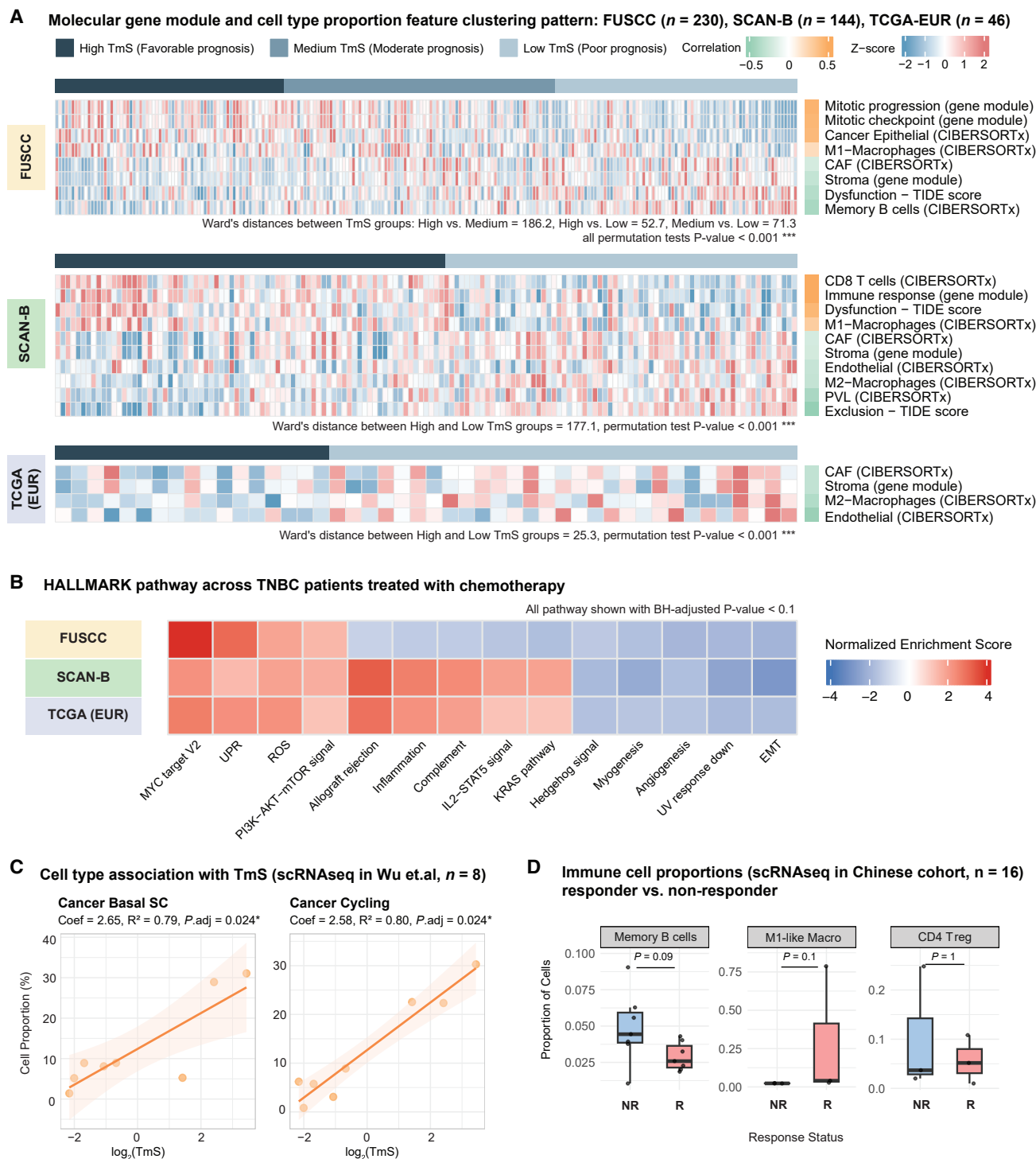


Figure 4. TmS unveils tumor microenvironment dynamics across multi-ethnic TNBCs

(A) Molecular and cellular features clustering pattern comparing Asian (FUSCC) and European (SCAN-B and TCGA subset) cohorts showing features significantly correlated with TmS (Pearson correlation ≥ 0.2 , $p < 0.05$). TmS categories are annotated as high (favorable prognosis), medium (moderate prognosis, FUSCC only), and low (poor prognosis). Right annotation bars indicate correlation with TmS (green to orange) while feature expression is shown as Z scores (blue to red). Statistical analysis confirms significant clustering with Ward's distance tests (shown below each heatmap).

(B) Heatmap of normalized enrichment scores (NES) from gene set enrichment analysis (GSEA) of HALLMARK pathways across FUSCC, SCAN-B, and TCGA-EUR. Genes were ranked by their expression correlation with TmS within each cohort. Positive NES (red) indicates higher pathway expression in high-TmS

(legend continued on next page)

In TCGA, TmS predicted DFS in Cox regression, while individual histological features (TILs and stromal-TILs [sTILs]) did not (Figure S1E). In SCAN-B, while TILs also showed excellent prognostic values (log rank $p = 0.012$, Figure S1F), TmS subgroups are not replaceable by TIL subgroups (Figure S1G). These findings indicate that TmS captures tumor-microenvironment interactions beyond major cellular proportions measured from H&E images, motivating comprehensive genomic and transcriptomic analysis to elucidate the biological basis of TmS in TNBC.

TmS reveals population-specific co-occurrence of TME features

Noting the differences in TmS distributions and in prognosis between Western and Asian cohorts, we systematically evaluated the shared and distinct patterns between the two ethnic groups (STAR Methods). For gene-expression-based analysis, we focused on the three RNA-seq cohorts (TCGA, SCAN-B, and FUSCC), as METABRIC's microarray platform has technical limitations that make it incomparable to others (Figure S2A). Since TCGA comprises patients from multiple ethnic groups (Table S1), we present comparisons for both the full mixed population and the European descent subgroup (55%, as TCGA-EUR) for concordance with SCAN-B. Our comprehensive molecular profiling evaluated (STAR Methods) (1) genetic/genomic features including canonical driver mutations and copy number alterations; (2) breast cancer-specific gene expression modules¹³ including mitotic checkpoint, mitotic progression, stroma, immune response, steroid response, lipid, early response, and basal; (3) deconvolved cell-type proportions including epithelial cells and major immune and stromal cell types; (4) T cell dysfunction and exclusion scores measured by TIDE⁴⁴; (5) canonical marker gene expression analysis for identified key cellular activities; and (6) HALLMARK/KEGG/REACTOME pathway enrichment analysis.^{45–47} We also used two single-cell RNA-seq datasets^{28,29} from East Asian TNBC patients without clinical follow-up information to replicate population-specific findings in FUSCC (Figure 1C, STAR Methods).

Intrinsic tumor cell genomic features show no association with TmS across cohorts

Consistent with our hypothesis that the nontumor-cell transcriptional activity in the denominator drives TmS in TNBC, we found none of the established genetic and genomic features for breast cancer cells, i.e., factors influencing the numerator of the TmS calculation (Figure 1B), are associated with TmS across cohorts. They include the canonical driver mutations (*TP53*, *BRCA1*, *PTEN*, *PIK3CA*, *RAD51D*, and *ATM*), copy number alterations (*MYC* and *PTEN*), *BRCA1/2* germline mutation status, tumor mutation burden (TMB), and HRD (Figure S2B; Table S3).

Low TmS suggests stromal activation across cohorts

Consistent with the digital pathology analysis in TCGA and SCAN-B, expression-level stromal activation emerged as a defining feature of low-TmS tumors across populations, including the East Asian. We observed negative correlations of stromal-associated gene module¹³ expression with continuous TmS in FUSCC, SCAN-B, and TCGA-EUR (Pearson $r = -0.36$, -0.21 , and -0.3 , respectively; Figures 4A and S3A). Cell-type deconvolution⁴⁸ showed negative correlations between TmS and proportions of stromal populations, including cancer-associated fibroblasts (CAFs) across all three cohorts and perivascular-like (PVL) cells and endothelial cells in the European cohorts (Pearson r ranges from -0.48 to -0.1 ; Figures 4A and S3A). Pathway analysis presented statistically significant enrichment of epithelial-mesenchymal transition (EMT), angiogenesis, myogenesis, and TGF- β signaling programs in low-TmS tumors (Figures 4B and S2C). The coordinated activation of EMT, stromal remodeling, and angiogenic pathways supports a desmoplastic microenvironment that promotes tumor invasion, metastasis, and immune evasion.^{17–20}

Contrasting tumor cell cycling and RAS/MAPK activation in East Asian TNBC

FUSCC exhibited a positive correlation between continuous TmS values and cell cycle control modules¹³ for breast cancer (mitotic checkpoint: Pearson $r = 0.46$; mitotic progression: Pearson $r = 0.48$, Figures 4A and S4A; cell cycle marker genes: Pearson $r = 0.37$; Figure S2D), which is corroborated by cell-type deconvolution⁴⁸ showing an increased proportion of cancer epithelial cells (Pearson $r = 0.41$; Figures 4A and S4A). Single-cell RNA-seq data from eight East Asian TNBC patients²⁹ replicated observations in FUSCC, with cancer basal stem cells and cancer cycling cells showing positive correlations with single-cell-derived TmS (linear regression $R^2 = 0.79$, 0.80 , adjusted $p < 0.05$; Figure 4C, STAR Methods). Clinical studies support the observed difference, as Asian breast cancer patients showed significantly greater benefit from CDK4/6 inhibitor therapy than non-Asian counterparts.⁴⁹ We posit that patients with high TmS are more likely to benefit from such treatment.

Pathway analysis showed consistent activation of core proliferation pathways (MYC target V2 and PI3K-AKT-mTOR), stress response programs (reactive oxygen species [ROS] and unfolded protein response [UPR]), and protein synthesis activities in high-TmS tumors across cohorts (Figures 4B and S2C, STAR Methods). However, in contrast to SCAN-B and TCGA showing enrichment of KRAS signaling in high-TmS tumors, FUSCC showed opposite enrichment in low-TmS tumors, specifically in KRAS signaling and mitogen-activated protein kinase (MAPK) family cascades (Figure S2E). Given that low-TmS tumors are associated with less favorable outcomes after chemotherapy

tumors; negative NES (blue) indicates higher expression in low-TmS tumors. Pathways shown met significance thresholds of BH-adjusted $p < 0.1$ in all three cohorts and $|NES| \geq 1.5$ in at least one cohort.

(C) Correlations between single-cell-based TmS (defined as the normalized UMI counts in cancer epithelial cells versus TME cells) and proportions of cell subtypes derived from single-cell RNA-sequencing (scRNA-seq) (Wu et al.,²⁹ $n = 8$). Scatterplots show significant positive correlations (adjusted $p < 0.05$) between TmS and cancer basal stem cells (left) and cancer cycling cells (right). Regression lines are shown, with coefficient R^2 , and adjusted p values displayed for both cell types. (D) Comparison of immune cell proportions in scRNA-seq (Zhang et al.,²⁸ $n = 16$) between chemotherapy responders (R, blue) and non-responders (NR, pink). Boxplots show distributions of memory B cells, M1-like macrophages, and CD4⁺ regulatory T cells. Center line represents median; box represents IQR; whiskers extend to $1.5 \times$ IQR; individual samples shown as dots. p values from two-sided Wilcoxon rank-sum tests are shown.

across populations, this population-specific KRAS activation pattern may indicate distinct therapeutic vulnerabilities and warrants further investigation.

Population-specific immune mechanisms distinguish low-TmS tumors

Low-TmS tumors across all cohorts are associated with worse prognosis after chemotherapy, yet exhibit distinct immunological mechanisms.

Macrophage. Macrophage polarization is consistently observed across cohorts: M1-like macrophage is enriched in high-TmS tumors, and M2-like macrophage is enriched in low-TmS tumors (Figures 4A and S3B).

T and B cells. T and B cell phenotypes differ substantially between populations. In SCAN-B, high-TmS tumors that are associated with improved outcomes after chemotherapy displayed an “inflamed-exhausted” phenotype: strong CD8⁺ T cell infiltration, enriched immune response gene module, and elevated cytotoxic gene expression (Pearson $r = 0.50$, 0.50 , and 0.53 , respectively; Figures 4A, S2D, and S3B), alongside activated interferon response and inflammatory signaling pathways (Figures 4B and S2C), yet with elevated exhaustion gene expression and dysfunction scores (Pearson $r = 0.51$; Figures S2D, S4B, and S4C), indicating T cells are present but functionally impaired. Pathway analysis indicated the presence of inflammation and anti-tumor immunity associated with TmS, with enrichment of T cell and B cell receptor signaling pathways in high-TmS tumors across both the SCAN-B and TCGA-EUR cohorts (Figures 4B and S2C). Conversely, low-TmS tumors showed an “excluded-cold” phenotype with significantly elevated exclusion scores (Pearson $r = -0.53$; Figure 4A) and significantly decreased dysfunction scores (Wilcoxon rank-sum test, Benjamini-Hochberg [BH]-adjusted $p < 0.001$; Figure S4C), suggesting that stromal cells prevent T cell infiltration and limit immunotherapy potential within the European population.

In FUSCC, low-TmS tumors that are associated with worse outcomes exhibited an opposite pattern: significantly increased dysfunction scores (Wilcoxon rank-sum test, BH-adjusted $p < 0.001$, Figures 4A and S4C) alongside activated immune pathways, including T cell and B cell receptor (T/BCR) pathways (Figures 4B and S2C), with little change in T cell exclusion, when compared to high-TmS tumors. Notably, memory B cell proportions showed a negative correlation with TmS (Pearson $r = -0.41$; Figures 4A and S3B), a pattern that is absent in European patients. The concurrent activation of BCR signaling pathways in these tumors suggests ongoing B cell engagement with tumor-associated antigens (Figure S2C). In the TME, chronic antigen stimulation can drive B cells toward functional exhaustion or terminal differentiation rather than stable memory pools.^{50,51} Memory B cells in certain tumor contexts, particularly those with IgA⁺ humoral responses, have been associated with immunosuppression and impaired T cell function.^{52,53} The association between higher memory B cell abundance and poor prognosis in Asian TNBC warrants further investigation, though functional validation through protein or functional assays would be required to establish mechanistic roles. Using an immune-enriched single-cell dataset from 16 Chinese TNBC patients²⁸ and the treat-

ment response as a proxy for TmS subgroups (response represents high TmS, nonresponse represents low TmS, STAR Methods), we replicated that the memory B cells are elevated in the non-response group at a marginal statistical significance (Wilcoxon rank-sum test $p = 0.09$, Figure 4D). The original study also reported an enriched B cell receptor signaling pathway in non-responders, consistent with our bulk RNA-seq findings in low-TmS FUSCC tumors (Figure S2C).

These findings reveal population-specific immunosuppressive mechanisms in low-TmS tumors: European cohorts show physical exclusion of immune cells by stromal barriers, while Asian cohorts exhibit functional immune dysfunction despite immune cell presence, with a distinctive role for memory B cells.

Population-specific TME programs identified by TmS subtypes

Integrating all TmS-associated features into TME programs, we identified population-specific patterns that collectively differentiate TmS-stratified patients in each cohort (Ward distance⁵⁴ permutation-based p values < 0.001 ; Figure 4A, STAR Methods). In the European descent SCAN-B and TCGA cohorts, 10 features define the TME programs: gene modules related to immune response and stroma, cellular proportions of CD8⁺ T cells, M1/M2-like macrophages, CAFs, endothelial cells and PVL cells, plus T cell dysfunction and exclusion scores (Figures 4A and 4B, STAR Methods). In the Asian descent FUSCC cohort, 8 features characterize the TME programs: gene modules for mitotic progression/checkpoint and stroma, cellular proportions of cancer epithelial cells, M1 macrophages, CAFs and memory B cells, and T cell dysfunction scores (Figures 4A and 4B, STAR Methods).

Western women with low-TmS TNBC tumors present an ECM-enriched signature

As TNBC patients with low TmS show resistance to chemotherapy and are unlikely to respond to immunotherapy due to the strong consistent signals from activated stroma, we performed differential expression analyses between high- and low-TmS tumors, aiming to identify gene signatures that may inform therapeutic strategies for these patients. We prioritized shared gene signatures from both TCGA and SCAN-B cohorts to ensure robustness in the Western populations.

A total of 164 differentially expressed genes (DEGs) were consistently altered across both cohorts (Figures 5A and 5B; ≥ 1.5 -fold change in expression and BH-adjusted $p < 0.05$), with 95 genes upregulated and 69 downregulated in high- versus low-TmS tumors (Table S4). Gene Ontology⁵⁵ (GO) enrichment analysis revealed distinct biological programs associated with TmS subtypes. The 69 genes upregulated in low-TmS tumors were predominantly associated with extracellular matrix (ECM) organization, cell-cell adhesion, and structural components of the TME (Figure 5C). Conversely, the 95 genes upregulated in high-TmS tumors demonstrated significant enrichment for immune-related functions, including innate and adaptive immune responses, chemokine signaling, and lymphocyte activation pathways (Figure 5D).

GO enrichment analysis of DEGs in FUSCC also revealed similar biological patterns, with low-TmS tumors showing enrichment in ECM organization, collagen-containing ECM, and

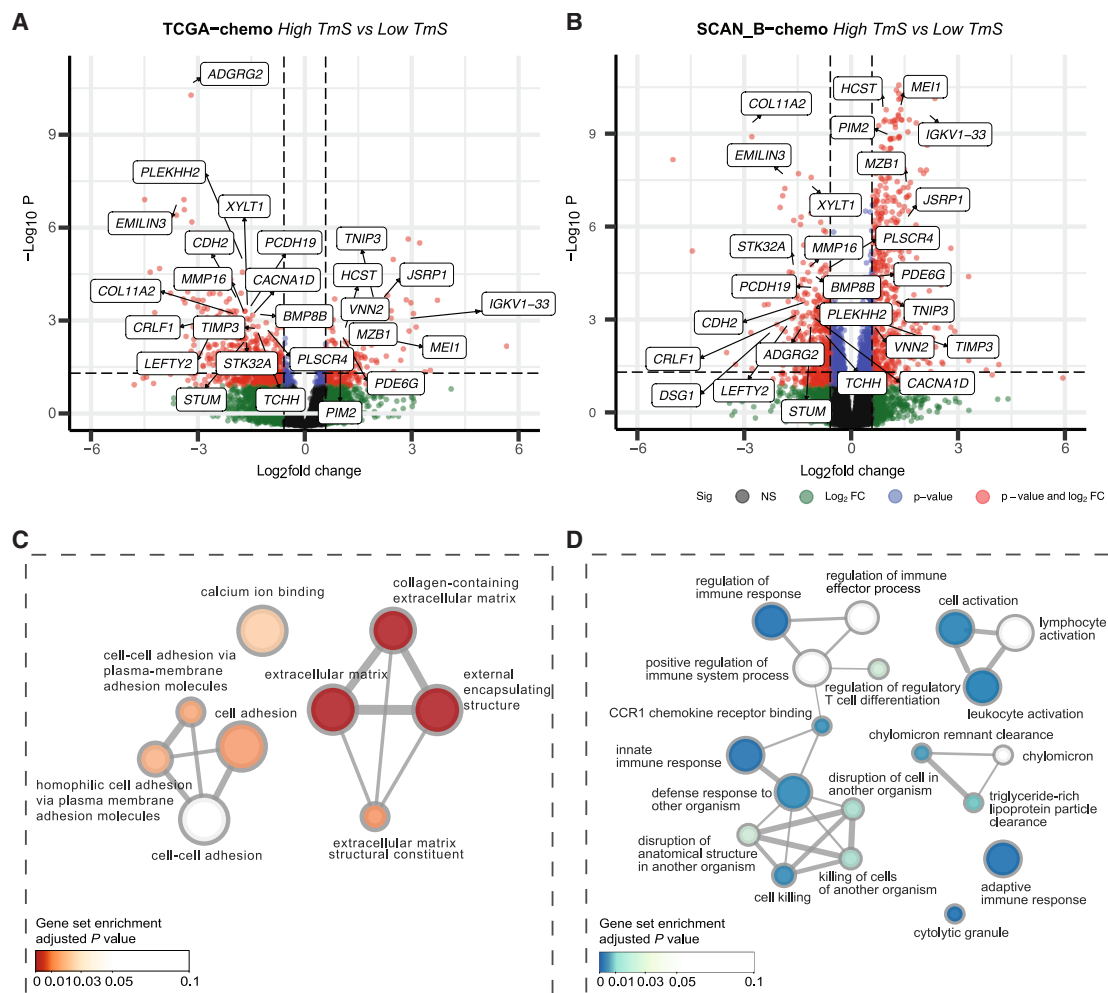


Figure 5. Integrative transcriptomic analysis reveals extracellular matrix signatures in treatment-resistant low-TmS tumors

(A and B) Volcano plot of overlapping differentially expressed genes (DEGs) in high versus low TmS comparison in TCGA (A) and SCAN-B (B) chemo-treated patients.

(C and D) Enrichment map shows pathways enriched in consensus low-TmS DEGs (C) and high-TmS DEGs (D). Nodes in the network represent pathways, and edge width represents the number of genes that overlap between enriched pathways. Nodes are colored by the hypergeometric p values from the pathway enrichment tests.

muscle system processes (Figure S4D), while high-TmS tumors again showing enrichment in nuclear division and cell cycle regulatory pathways (Figure S4E).

The conserved ECM signature in low-TmS tumors across all populations suggests that ECM-targeting approaches, such as ECM remodeling agents or TGF- β pathway inhibitors, represent potential therapeutic avenues that warrant further investigation in this high-risk TNBC subgroup,⁵⁶ regardless of ethnicity.

DISCUSSION

Using a deconvolution-derived transcriptomic plasticity measure, TmS, we provide a framework for addressing TNBC heterogeneity across ethnic groups. Patients can be first designated as high or low TmS, a distinct subtyping system from existing TNBC subtypes. Treatments to disrupt stromal barriers may be priori-

tized for low-TmS patients who would otherwise not perform well with chemotherapy across groups. European and East Asian populations showed contrasting cellular activities in both the epithelial and immune cell compartments for low-TmS tumors, supporting the transcriptomic plasticity measure, i.e., tumor cell versus nontumor cell total mRNA expression per cell per haploid genome, to represent a universal downstream feature of critical cellular and molecular events and, hence, an ideal biomarker. Further investigations guided by TmS subtypes will be useful for understanding these differences in cellular interactions. Such insights can inform personalized therapeutic strategies for TNBC patients.

TmS is calculated through an integrative deconvolution of matched bulk RNA/DNA-seq data from four multi-ethnic patient cohorts ($n = 575$), among which SCAN-B and FUSCC represent more modern and clinically relevant cohorts. These two cohorts

were not used in the original TmS development, thereby serving as independent validation cohorts for prognostic performance. All TNBC patient-specific TmS values are available for further biological exploration from the TmS shiny app (<https://www.wylyab.github.io/TmS/articles/shinyapp.html>).

The clinical utility of TmS is highlighted by current therapeutic challenges in TNBC. Current clinical trials in unselected TNBC show variable efficacy (Table S5): checkpoint inhibitors (~50%–60% ORR, objective response rate, in PD-L1+ patients), anti-angiogenic combinations (~37%–41% ORR with progression-free survival [PFS] benefit but no overall survival [OS] gain), PARP inhibitors (36%–44% CBR, clinical benefit rate, in maintenance, linked to prior platinum), and TGF- β /ECM agents (~9% ORR in early cohorts). Based on our TmS distributions, biomarker-guided selection could improve these outcomes by matching ~40%–50% of stromal-enriched, low-TmS patients to ECM-targeting therapies and ~50%–60% of immune-enriched, high-TmS patients of European descent to checkpoint inhibitors. Future trials incorporating TmS stratification as a primary biomarker could optimize therapeutic selection across diverse populations.

The ethnicity-specific TME patterns revealed by TmS have important implications for cross-population application of emerging combination therapies. The emerging practice of developing anti-angiogenesis plus anti-PD-L1 combinations predominantly in Chinese patient populations²³ (e.g., VEGFR inhibitors like apatinib or famitinib combined with PD-L1 blockers) strongly supports our findings that Asian low-TmS patients exhibit both vascular abnormalities and T cell dysfunction. However, this strategy may not work the same for the Western population. The Western low-TmS patients, where T cell exclusion dominates without significant dysfunction, might require strategies that first disrupt stromal barriers (e.g., TGF- β inhibitors) before applying checkpoint inhibitors. Conversely, immunotherapy alone might be beneficial to the Western high-TmS patients, while Asian high-TmS patients may additionally benefit from cell cycle-targeting agents to address cancer cell's proliferation-driven phenotype. These insights emphasize the importance of biomarker-guided, population-aware clinical trial designs to optimize therapeutic regimens across diverse patient groups.

Beyond TNBC, our findings pave the way for future research to reveal the mechanistic underpinnings of clinically relevant TME dynamics across diverse patient populations that are beyond European/Asians and across cancer types, including ER+ breast cancers and many others. Our study advocates the utility of cancer “transcriptome ploidy,” i.e., the total mRNA expression for tumor versus nontumor cells, to quantify tumor plasticity in cancer patients at scale.

Limitations of the study

TmS estimation faces several practical challenges that limit immediate clinical adoption. The requirement for accurate deconvolution of both bulk RNA and DNA sequencing data restricts implementation in resource-constrained settings. Our comprehensive cross-platform analysis reveals heterogeneity across cohorts that impact TmS clinical translation. The substantial variation in TmS distributions across cohorts likely reflects complex interac-

tions between technical factors (Table S1, sequencing platforms, RNA extraction protocols, and sequencing depths), population-specific biology, and temporal changes in treatment protocols. The confounding of ancestry, clinical practice, and technical platform across our multi-cohort study prevents definitive separation of biological from technical effects.

Our retrospective design enabled large-scale molecular analyses across diverse populations but limits causal inference between TmS-associated features and treatment outcomes. Translating these associations into therapeutic strategies requires distinguishing causal drivers from downstream correlates of resistance. Because our analyses are based on pre-treatment samples, TmS-predictive features are likely causally relevant, though functional validation remains necessary. Computational pathway analyses reveal robust transcriptional differences, yet proteomic or functional studies would strengthen mechanistic insight into population-specific patterns. Target prioritization should consider regulatory positioning in signaling cascades, functional centrality in pathway networks, cross-cohort consistency, and established mechanistic roles in drug resistance. Among ECM-related alterations in low-TmS tumors, TGF- β pathway components may represent more promising targets than individual matrix proteins, given their regulatory role in ECM remodeling and therapy resistance. In contrast, while high-TmS tumors show molecular hallmarks of immunotherapy responsiveness (enhanced CD8⁺ T cell infiltration and reduced cytotoxic T cell exclusion), direct clinical validation of immunotherapy benefit in TmS-stratified TNBC patients remains to be established.

Several barriers must be overcome before clinical implementation: (1) standardized TmS measurement protocols across institutions, (2) prospective validation of treatment-predictive cutoffs through clinical trials, (3) cost-effectiveness analysis for routine RNA/DNA-seq, and (4) integration with existing clinical decision-making frameworks. Definitive clinical cutoffs for treatment assignment require prospective validation through clinical trials, as retrospective thresholds may not translate directly across different clinical contexts and treatment protocols. We recommend an evidence-based implementation approach: research applications may use platform- and population-matched reference cutoffs from our study, while clinical trials should implement cohort-specific TmS cutoffs derived from platform-matched reference populations, with prospective validation of these thresholds as primary or secondary endpoints.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Wenyi Wang (wwang7@mdanderson.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- TCGA: RNA-seq raw read counts, copy number variations, mutation data, and clinical information for 134 TNBC samples are available from the Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov/>); METABRIC: Affymetrix SNP 6.0 CEL files and Illumina HT-12 gene expression arrays are accessible via the

European Genome-phenome Archive (<https://ega-archive.org/studies/EGAS00000000083>). Data access requires formal request to METABRIC Data Access Committee. Clinical metadata are available at https://www.cbioportal.org/study/clinicalData?id=brca_metabric.

Accession: EGAS00000000083; FUSCC: omics data for 465 TNBC samples are accessible via National Omics Data Encyclopedia (NODE: OEP000155, <http://www.biosino.org/node>). Sequence data deposited in NCBI Gene Expression Omnibus (OncoScan array: GSE118527) and Sequence Read Archive (WES and RNA-seq: SRP157974); SCAN-B: somatic mutation data for 237 TNBC samples are available at <https://data.mendeley.com/datasets/2mn4ctdpxp/1>. RNA-seq raw counts for 235 samples are available at <https://data.mendeley.com/datasets/yztxn4nmd/3> from Staaf et al.⁵⁷; Wu et al.²⁹: 10 TNBC samples (8 with cancer epithelial cells) are accessible via Broad Institute Single Cell Portal (https://singlecell.broadinstitute.org/single_cell/study/SCP1039) and Gene Expression Omnibus (GSE176078); Zhang et al.²⁸: 16 pre-chemotherapy Chinese TNBC samples are available via Gene Expression Omnibus (GSE169246). Analysis is focused on pre-treatment tumor samples, with responders (R: partial response) vs. non-responders (NR: stable/progressive disease); H&E whole-slide images for the Cancer Genome Atlas projects are available from the Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov/>).

- TmS analysis pipeline: <https://github.com/wwylab/TmS>; TmS values for all patients: <https://wwylab.github.io/TmS/articles/data.html>; all custom code used in this study is available from the [lead contact](#) upon request.
- Any additional information required to reanalyze the data reported in this work paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

The authors would like to acknowledge the SCAN-B study, all patients and clinicians participating in the SCAN-B study, personnel at the central SCAN-B laboratory at the Division of Oncology, Department of Clinical Sciences Lund, Lund University, as well as the Swedish national breast cancer quality registry (NKBC), Regional Cancer Center South (RCC Syd), Regional Biobank Center South (RBC Syd), and the South Sweden Breast Cancer Group (SSBCG). The authors also thank Dr. Jessica C. Lal for her critical reading of the manuscript and editorial assistance. This work was supported by Spatial Ecology & QUantitative pathOlogy Image Analytical platform (SEQUOIA) through the MD Anderson STRategic Research Initiative Development Program (STRIDE). This work was supported by National Cancer Institute R01CA268380 (to W.W.), Department of Defense PC210079 (to W.W.), Cancer Prevention and Research Institute of Texas RR200009 (to G.V.E.), National Cancer Institute 1R37CA269783-01A1 (to G.V.E.), American Cancer Society Research Scholar Grant RSG-22-093-01-CCB (to G.V.E.), and Lyda Hill Philanthropies FP00020550 (to Y.Y.).

AUTHOR CONTRIBUTIONS

Y.D. and W.W. conceived and designed the project. Y.D. performed data analysis and interpretation, planned figure design, and wrote the manuscript. X.P. developed the imaging segmentation tool and conducted imaging analysis. S.G., Y.J., M.D.M., and S.J. assisted with data analysis and interpretation. S.C. provided statistical expertise and computational analysis. S.S. contributed to data interpretation and manuscript editing. L.S., L.Y., J.S., G.V.E., and J.T.C. provided access to critical data resources and supervision for key components of the project. B.L. provided clinical perspective and validated findings. Y.Y. supervised imaging analysis and provided methodological expertise. W.W. supervised the overall project, secured funding, contributed to data interpretation, and critically revised the manuscript. All authors reviewed and approved the final manuscript.

DECLARATION OF INTERESTS

This work is associated with patent filings by W.W., Y.D., X.P., Y.Y., and S.C. at MD Anderson Cancer Center. G.V.E. receives research funding sponsorship

from Chimerix, Inc. and experimental compounds from Chimerix, Inc., and the Lead Discovery Center of Germany. B.L. receives research funding from Merck, Genentech, Puma Biotechnology, Takeda Oncology, Amgen, Celcuity, and Novartis. L.Y. is funded by a Wellcome Trust Clinical Research Career Development Fellowship (ref. 214584/Z/18/Z). These activities are unrelated to this research.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Patient cohorts and clinical characteristics
 - Single-cell validation cohorts
 - Ethics statement
- **METHOD DETAILS**
 - TmS mathematical framework
 - Cohort-specific TmS estimation
 - TNBC molecular subtyping
 - Treatment characterization
 - Digital pathology analysis
 - Pathway and gene module analyses
 - Cell type deconvolution
 - T cell dysfunction and exclusion analysis
 - Differential gene expression and integrative enrichment analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Sample size and data characteristics
 - Normalization of raw counts from microarray or RNAseq data
 - Association with clinical variables
 - Survival analyses
 - Cox regression analyses
 - Association between categorical variables
 - Ward distance analyses
 - Multiple testing correction
 - Statistical software and significance

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xcrm.2026.102610>.

Received: May 27, 2025

Revised: October 22, 2025

Accepted: January 12, 2026

Published: February 17, 2026

REFERENCES

1. Valastyan, S., and Weinberg, R.A. (2011). Tumor metastasis: molecular insights and evolving paradigms. *Cell* 147, 275–292.
2. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424.
3. Hammond, M.E.H., Hayes, D.F., Dowsett, M., Allred, D.C., Hagerty, K.L., Badve, S., Fitzgibbons, P.L., Francis, G., Goldstein, N.S., Hayes, M., et al. (2010). American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J. Clin. Oncol.* 28, 2784–2795.
4. Bianchini, G., Balko, J.M., Mayer, I.A., Sanders, M.E., and Gianni, L. (2016). Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat. Rev. Clin. Oncol.* 13, 674–690.

5. Bianchini, G., De Angelis, C., Licata, L., and Gianni, L. (2022). Treatment landscape of triple-negative breast cancer — expanded options, evolving needs. *Nat. Rev. Clin. Oncol.* **19**, 91–113.
6. Wang, F., Zheng, W., Bailey, C.E., Mayer, I.A., Pietenpol, J.A., and Shu, X.O. (2021). Racial/Ethnic Disparities in All-Cause Mortality among Patients Diagnosed with Triple-Negative Breast Cancer. *Cancer Res.* **81**, 1163–1170.
7. Abraham, J.E., Pinilla, K., Dayimu, A., Grybowicz, L., Demiris, N., Harvey, C., Drewett, L.M., Lucey, R., Fulton, A., Roberts, A.N., et al. (2024). The PARTNER trial of neoadjuvant olaparib with chemotherapy in triple-negative breast cancer. *Nature* **629**, 1142–1148.
8. Cortes, J., Rugo, H.S., Cescon, D.W., Im, S.A., Yusof, M.M., Gallardo, C., Lipatov, O., Barrios, C.H., Perez-Garcia, J., Iwata, H., et al. (2022). Pembrolizumab plus Chemotherapy in Advanced Triple-Negative Breast Cancer. *N. Engl. J. Med.* **387**, 217–226.
9. Schmid, P., Cortes, J., Dent, R., McArthur, H., Pusztai, L., Kümmel, S., Denkert, C., Park, Y.H., Hui, R., Harbeck, N., et al. (2024). Overall Survival with Pembrolizumab in Early-Stage Triple-Negative Breast Cancer. *N. Engl. J. Med.* **391**, 1981–1991.
10. Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., and Pietenpol, J.A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* **121**, 2750–2767.
11. Lehmann, B.D., Jovanović, B., Chen, X., Estrada, M.V., Johnson, K.N., Shyr, Y., Moses, H.L., Sanders, M.E., and Pietenpol, J.A. (2016). Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. *PLoS One* **11**, e0157368.
12. Burstein, M.D., Tsimelzon, A., Poage, G.M., Covington, K.R., Contreras, A., Fuqua, S.A.W., Savage, M.I., Osborne, C.K., Hilsenbeck, S.G., Chang, J.C., et al. (2015). Comprehensive Genomic Analysis Identifies Novel Subtypes and Targets of Triple-negative Breast Cancer. *Clin. Cancer Res.* **21**, 1688–1698.
13. Fredlund, E., Staaf, J., Rantala, J.K., Kallioniemi, O., Borg, A., and Ringnér, M. (2012). The gene expression landscape of breast cancer is shaped by tumor protein p53 status and epithelial-mesenchymal transition. *Breast Cancer Res.* **14**, R113.
14. Telli, M.L., Timms, K.M., Reid, J., Hennessy, B., Mills, G.B., Jensen, K.C., Szallasi, Z., Barry, W.T., Winer, E.P., Tung, N.M., et al. (2016). Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. *Clin. Cancer Res.* **22**, 3764–3773.
15. Bholra, N.E., Balko, J.M., Dugger, T.C., Kuba, M.G., Sánchez, V., Sanders, M., Stanford, J., Cook, R.S., and Arteaga, C.L. (2013). TGF- β inhibition enhances chemotherapy action against triple-negative breast cancer. *J. Clin. Invest.* **123**, 1348–1358.
16. Saltz, J., Gupta, R., Hou, L., Kurc, T., Singh, P., Nguyen, V., Samaras, D., Shroyer, K.R., Zhao, T., Batiste, R., et al. (2018). Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. *Cell Rep.* **23**, 181–193.e7.
17. Joyce, J.A., and Fearon, D.T. (2015). Cell exclusion, immune privilege, and the tumor microenvironment. *Science* **348**, 74–80.
18. Kalluri, R. (2016). The biology and function of fibroblasts in cancer. *Nat. Rev. Cancer* **16**, 582–598.
19. Hinshaw, D.C., and Shevde, L.A. (2019). The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res.* **79**, 4557–4566.
20. Chen, X., and Song, E. (2019). Turning foes to friends: targeting cancer-associated fibroblasts. *Nat. Rev. Drug Discov.* **18**, 99–115.
21. Keren, L., Bosse, M., Marquez, D., Angoshtari, R., Jain, S., Varma, S., Yang, S.R., Kurian, A., Van Valen, D., West, R., et al. (2018). A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell* **174**, 1373–1387.e19.
22. Jackson, H.W., Fischer, J.R., Zanotelli, V.R.T., Ali, H.R., Mechera, R., Soysal, S.D., Moch, H., Muenst, S., Varga, Z., Weber, W.P., and Bodenmiller, B. (2020). The single-cell pathology landscape of breast cancer. *Nature* **578**, 615–620.
23. Li, Q., Wang, Y., Jia, W., Deng, H., Li, G., Deng, W., Chen, J., Kim, B.Y.S., Jiang, W., Liu, Q., and Liu, J. (2020). Low-Dose Anti-Angiogenic Therapy Sensitizes Breast Cancer to PD-1 Blockade. *Clin. Cancer Res.* **26**, 1712–1724.
24. Donati, B., Reggiani, F., Torricelli, F., Santandrea, G., Rossi, T., Bisagni, A., Gasparini, E., Neri, A., Cortesi, L., Ferrari, G., et al. (2024). Spatial Distribution of Immune Cells Drives Resistance to Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer. *Cancer Immunol. Res.* **12**, 120–134.
25. Wills, Q.F., and Mead, A.J. (2015). Application of single-cell genomics in cancer: promise and challenges. *Hum. Mol. Genet.* **24**, R74–R84.
26. Stuart, T., and Satija, R. (2019). Integrative single-cell analysis. *Nat. Rev. Genet.* **20**, 257–272.
27. Cao, S., Wang, J.R., Ji, S., Yang, P., Dai, Y., Guo, S., Montierth, M.D., Shen, J.P., Zhao, X., Chen, J., et al. (2022). Estimation of tumor cell total mRNA expression in 15 cancer types predicts disease progression. *Nat. Biotechnol.* **40**, 1624–1633.
28. Zhang, Y., Chen, H., Mo, H., Hu, X., Gao, R., Zhao, Y., Liu, B., Niu, L., Sun, X., Yu, X., et al. (2021). Single-cell analyses reveal key immune cell subsets associated with response to PD-L1 blockade in triple-negative breast cancer. *Cancer Cell* **39**, 1578–1593.e8.
29. Wu, S.Z., Al-Eryani, G., Roden, D.L., Junankar, S., Harvey, K., Andersson, A., Thennavan, A., Wang, C., Torpy, J.R., Bartonicek, N., et al. (2021). A single-cell and spatially resolved atlas of human breast cancers. *Nat. Genet.* **53**, 1334–1347.
30. Dai, Y., Guo, S., Pan, Y., Castignani, C., Montierth, M.D., Van Loo, P., and Wang, W. (2026). A guide to transcriptomic deconvolution in cancer. *Nat. Rev. Cancer* **26**, 84–103. <https://doi.org/10.1038/s41568-025-00886-9>.
31. Saal, L.H., Vallon-Christersson, J., Häkkinen, J., Hegardt, C., Grabau, D., Winter, C., Brueffer, C., Tang, M.H.E., Reuterswärd, C., Schulz, R., et al. (2015). The Sweden Cancerome Analysis Network - Breast (SCAN-B) Initiative: a large-scale multicenter infrastructure towards implementation of breast cancer genomic analyses in the clinical routine. *Genome Med.* **7**, 20.
32. Jiang, Y.-Z., Ma, D., Suo, C., Shi, J., Xue, M., Hu, X., Xiao, Y., Yu, K.D., Liu, Y.R., Yu, Y., et al. (2019). Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies. *Cancer Cell* **35**, 428–440.e5.
33. Cancer Genome Atlas Research Network; Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120.
34. Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352.
35. Pan, J.-W., Zabidi, M.M.A., Ng, P.S., Meng, M.Y., Hasan, S.N., Sandey, B., Sammut, S.J., Yip, C.H., Rajadurai, P., Rueda, O.M., et al. (2020). The molecular landscape of Asian breast cancers reveals clinically relevant population-specific differences. *Nat. Commun.* **11**, 6433.
36. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196.
37. Kumar, M.P., Du, J., Lagoudas, G., Jiao, Y., Sawyer, A., Drummond, D.C., Lauffenburger, D.A., and Raue, A. (2018). Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics. *Cell Rep.* **25**, 1458–1468.e4.
38. Kürten, C.H.L., Kulkarni, A., Cillo, A.R., Santos, P.M., Roble, A.K., Onkar, S., Reeder, C., Lang, S., Chen, X., Duvvuri, U., et al. (2021). Investigating immune and non-immune cell interactions in head and neck tumors by single-cell RNA sequencing. *Nat. Commun.* **12**, 7338.

39. Aine, M., Boyaci, C., Hartman, J., Häkkinen, J., Mitra, S., Campos, A.B., Nimeus, E., Ehinger, A., Vallon-Christersson, J., Borg, Å., and Staaf, J. (2021). Molecular analyses of triple-negative breast cancer in the young and elderly. *Breast Cancer Res.* 23, 20.
40. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., and Luo, P. (2021). SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *Adv. Neural Inf. Process. Syst.* 34, 12077–12090.
41. Li, J.J., Tsang, J.Y., and Tse, G.M. (2021). Tumor Microenvironment in Breast Cancer—Updates on Therapeutic Implications and Pathologic Assessment. *Cancers* 13, 4233.
42. Bassiouni, R., Idowu, M.O., Gibbs, L.D., Robila, V., Grizzard, P.J., Webb, M.G., Song, J., Noriega, A., Craig, D.W., and Carpten, J.D. (2023). Spatial Transcriptomic Analysis of a Diverse Patient Cohort Reveals a Conserved Architecture in Triple-Negative Breast Cancer. *Cancer Res.* 83, 34–48.
43. Tiwari, A., Oravec, T., Dillon, L.A., Italiano, A., Audoly, L., Fridman, W.H., and Clifton, G.T. (2023). Towards a consensus definition of immune exclusion in cancer. *Front. Immunol.* 14, 1084887.
44. Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., Li, Z., Traugh, N., Bu, X., Li, B., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24, 1550–1558.
45. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* 1, 417–425.
46. Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30.
47. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550.
48. Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* 37, 773–782.
49. Lu, Y.-S., Yeo, W., Yap, Y.S., Park, Y.H., Tamura, K., Li, H., and Cheng, R. (2021). An Overview of the Treatment Efficacy and Side Effect Profile of Pharmacological Therapies in Asian Patients with Breast Cancer. *Target. Oncol.* 16, 701–741.
50. Laumont, C.M., Banville, A.C., Gilardi, M., Hollern, D.P., and Nelson, B.H. (2022). Tumour-infiltrating B cells: immunological mechanisms, clinical impact and therapeutic opportunities. *Nat. Rev. Cancer* 22, 414–430.
51. Bod, L., and Shalapour, S. (2024). B cells spatial organization defines their phenotype and function in cancer ‘Tell me with whom you consort, and I will tell you who you are’ - Goethe. *Curr. Opin. Immunol.* 91, 102504.
52. Hao, D., Han, G., Sinjab, A., Gomez-Bolanos, L.I., Lazcano, R., Serrano, A., Hernandez, S.D., Dai, E., Cao, X., Hu, J., et al. (2022). The Single-Cell Immunogenomic Landscape of B and Plasma Cells in Early-Stage Lung Adenocarcinoma. *Cancer Discov.* 12, 2626–2645.
53. Ma, J., Wu, Y., Ma, L., Yang, X., Zhang, T., Song, G., Li, T., Gao, K., Shen, X., Lin, J., et al. (2024). A blueprint for tumor-infiltrating B cells across human cancers. *Science* 384, ead4857.
54. Ward, J.H., Jr. (1963). Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* 58, 236–244.
55. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.
56. Yi, M., Wu, Y., Niu, M., Zhu, S., Zhang, J., Yan, Y., Zhou, P., Dai, Z., and Wu, K. (2022). Anti-TGF- β /PD-L1 bispecific antibody promotes T cell infiltration and exhibits enhanced antitumor activity in triple-negative breast cancer. *J. Immunother.* 45, e005543.
57. Staaf, J., Häkkinen, J., Hegardt, C., Saal, L.H., Kimbung, S., Hedenfalk, I., Lien, T., Sörle, T., Naume, B., Russnes, H., et al. (2022). RNA sequencing-based single sample predictors of molecular subtype and risk of recurrence for clinical assessment of early-stage breast cancer. *NPJ Breast Cancer* 8, 94.
58. Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70.
59. Carrot-Zhang, J., Chambwe, N., Damrauer, J.S., Knijnenburg, T.A., Robertson, A.G., Yau, C., Zhou, W., Berger, A.C., Huang, K.L., Newberg, J.Y., et al. (2020). Comprehensive Analysis of Genetic Ancestry and Its Molecular Correlates in Cancer. *Cancer Cell* 37, 639–654.e6.
60. Lin, C.Y., Lovén, J., Rahl, P.B., Paranal, R.M., Burge, C.B., Bradner, J.E., Lee, T.I., and Young, R.A. (2012). Transcriptional Amplification in Tumor Cells with Elevated c-Myc. *Cell* 151, 56–67.
61. Nie, Z., Hu, G., Wei, G., Cui, K., Yamane, A., Resch, W., Wang, R., Green, D.R., Tessarollo, L., Casellas, R., et al. (2012). c-Myc Is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells. *Cell* 151, 68–79.
62. Gulati, G.S., Sikandar, S.S., Wesche, D.J., Manjunath, A., Bharadwaj, A., Berger, M.J., Ilagan, F., Kuo, A.H., Hsieh, R.W., Cai, S., et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* 367, 405–411.
63. Athanasiadis, E.I., Botthof, J.G., Andres, H., Ferreira, L., Lio, P., and Cvejic, A. (2017). Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nat. Commun.* 8, 2045.
64. Li, C., and Wong, W.H. (2001). Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *Proc. Natl. Acad. Sci.* 98, 31–36.
65. Bolstad, B.M., Irizarry, R.A., Åstrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185–193.
66. Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4, 249–264.
67. Lovén, J., Orlando, D.A., Sigova, A.A., Lin, C.Y., Rahl, P.B., Burge, C.B., Levens, D.L., Lee, T.I., and Young, R.A. (2012). Revisiting Global Gene Expression Analysis. *Cell* 151, 476–482.
68. Upender, M.B., Habermann, J.K., McShane, L.M., Korn, E.L., Barrett, J.C., Difilippantonio, M.J., and Ried, T. (2004). Chromosome Transfer Induced Aneuploidy Results in Complex Dysregulation of the Cellular Transcriptome in Immortalized and Cancer Cells. *Cancer Res.* 64, 6941–6949.
69. Van Loo, P., Nordgard, S.H., Lingjærde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., et al. (2010). Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci.* 107, 16910–16915.
70. Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118–127.
71. Wang, Z., Cao, S., Morris, J.S., Ahn, J., Liu, R., Tyekucheva, S., Gao, F., Li, B., Lu, W., Tang, X., et al. (2018). Transcriptome Deconvolution of Heterogeneous Tumor Samples with Immune Infiltration. *iScience* 9, 451–460.
72. Chen, X., Li, J., Gray, W.H., Lehmann, B.D., Bauer, J.A., Shyr, Y., and Pietersen, J.A. (2012). TNBCtype: A Subtyping Tool for Triple-Negative Breast Cancer. *Cancer Inf.* 11, CIN-S9983.
73. Gendoo, D.M., Ratanasirigulchai, N., Schröder, M.S., Paré, L., Parker, J.S., Prat, A., and Haibe-Kains, B. (2016). Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics* 32, 1097–1099. <https://academic.oup.com/bioinformatics/article/32/7/1097/1743901>.
74. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2010.11929>.

75. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (IEEE). <https://ieeexplore.ieee.org/document/5206848>.
76. Amgad, M., Elfandy, H., Hussein, H., Atteya, L.A., Elsebaie, M.A.T., Abo Elnasr, L.S., Sakr, R.A., Salem, H.S.E., Ismail, A.F., Saad, A.M., et al. (2019). Structured crowdsourcing enables convolutional segmentation of histology images. *Bioinformatics* 35, 3461–3467.
77. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles | PNAS. <https://www.pnas.org/doi/full/10.1073/pnas.0506580102>.
78. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740.
79. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
80. Paczkowska, M., Barenboim, J., Sintupisut, N., Fox, N.S., Zhu, H., Abd-Rabbo, D., Mee, M.W., Boutros, P.C., and PCAWG Drivers and Functional Interpretation Working Group; and Reimand, J.; PCAWG Consortium (2020). Integrative pathway enrichment analysis of multivariate omics data. *Nat. Commun.* 11, 735.
81. Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* 46, D649–D655.
82. Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G.D. (2010). Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. *PLoS One* 5, e13984.
83. Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., et al. (2007). Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* 2, 2366–2382.
84. Wu, H., Wang, C., and Wu, Z. (2013). A new shrinkage estimator for dispersion improves differential expression detection in RNA-seq data. *Biostatistics* 14, 232–243.
85. Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., and Omberg, L. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* 173, 400–416.e11.
86. Therneau, T. & Atkinson, E. An Introduction to Recursive Partitioning Using the RPART Routines. in (2015).
87. Cherasia, K.E., Cabrera, J., Fernholz, L.T., and Fernholz, R. (2023). Data Nuggets in Supervised Learning. In *Robust and Multivariate Statistical Methods: Festschrift in Honor of David E. Tyler, M. Yi and K. Nordhausen*, eds. (Springer International Publishing), pp. 429–449. https://doi.org/10.1007/978-3-031-22687-8_20.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
TCGA BRCA	Genomic Data Commons	https://portal.gdc.cancer.gov/
METABRIC	European Genome-phenome Archive	EGAS00000000083
FUSCC	National Omics Data Encyclopedia; GEO; SRA	OEP000155; GSE118527; SRP157974
SCAN-B	Mendeley Data	https://data.mendeley.com/datasets/2mn4ctdpxp/1
Wu et al. scRNA-seq	GEO	GSE176078
Zhang et al. scRNA-seq	GEO	GSE169246
Software and algorithms		
TmS pipeline	Github	https://github.com/wwylab/TmS
ASCAT v2.4.3	Github	https://github.com/VanLoo-lab/ascats
DeMixT	Bioconductor R package	https://bioconductor.org/packages/DeMixT/
CIBERSORTx	Web tool	https://cibersortx.stanford.edu/
TIDE	Github	https://github.com/jingxinfu/TIDEpy
Segformer	Github	https://github.com/NVlabs/SegFormer
DESeq2	Bioconductor R package	https://bioconductor.org/packages/DESeq2/
ClusterProfiler	Bioconductor R package	https://bioconductor.org/packages/clusterProfiler/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Patient cohorts and clinical characteristics

This retrospective study analyzed 575 chemotherapy-treated TNBC patients with matched genomic and transcriptomic profiling and >5-year clinical follow-up.

- TCGA (n = 83): Mixed ancestry (55% European, 25.3% African, 19.3% other). Among patients with events, median disease-free survival: 1.79 years. Variable chemotherapy regimens: 36.1% anthracyclines, 26.5% taxanes.
- METABRIC (n = 118): European patients. Among patients with events, median relapse-free survival: 1.57 years. Diverse treatment protocols from historical cohort (1999–2010).
- SCAN-B (n = 144, excluded one sample without survival information): Swedish population-based cohort. Among patients with events, median relapse-free interval: 3.61 years. Predominantly anthracycline-taxane combinations (97.2% anthracyclines, 87.5% taxanes).
- FUSCC (n = 230): East Asian (Chinese) patients. Among patients with events, median relapse-free survival: 3.24 years. Primarily taxane-based regimens (79.1%).

Triple-negative status was defined by absence of ER, PR, and HER2 expression via immunohistochemistry or molecular profiling criteria. For TCGA samples with missing values in IHC status, we followed the criteria established in a TNBC study by Koboldt et al.⁵⁸ All patients received chemotherapy as part of their treatment regimen. Complete treatment details are provided in Table S1. Ethnic composition data obtained from Carrot-Zhang et al.⁵⁹(Table S1). Cohort-specific technical specifications including sequencing platforms, library preparation methods, and TmS estimation parameters are detailed in Table S1.

Single-cell validation cohorts

- Wu et al.²⁹ (n = 8): East Asian TNBC samples with cancer epithelial cells from treatment-naïve patients. Validated cell-type-specific TmS associations.
- Zhang et al.²⁸ (n = 16): Chinese TNBC patients with pre-treatment tumor samples classified as responders (partial response) vs. non-responders (stable/progressive disease). Replicated immune microenvironment findings in Asian populations.

Ethics statement

All patient data were obtained from publicly available databases with appropriate institutional review board approvals and patient consent as described in the original studies. This study involved analysis of de-identified data and did not require additional ethical approval.

METHOD DETAILS

TmS mathematical framework

Total mRNA transcript amount in bulk tissues has been recognized as a fundamental feature in cancer biology. From early studies on MYC activation,^{60,61} which drives global transcriptional amplification, to recent single cell studies where researchers have found that changes in overall expression output per cell being the best indicator of tumorigenesis than any specific pathways.^{62,63} Measuring such a feature in human tissues at-scale, however, poses several analytical challenges, as total tumor-cell mRNA expression information is masked during standard bulk data analysis, thus requiring deconvolution. Variation in the total expression level is removed by routine normalization, together with technical biases, including read depth and library preparation.^{64–67} DNA and RNA sequencing data generated from cancer studies contain reads from both tumor and other surrounding cells. Furthermore, copy number aberrations such as gain or loss of chromosomal copies (i.e., ploidy) in tumor cells affect gene expression through dosage effects.⁶⁸ Our previous work²⁷ addressed these challenges by introducing *TmS* as the ratio of total mRNA transcript level per cell per haploid genome of tumor cells to their surrounding non-tumor cells. Its mathematical formulation is as follows:

The average global mRNA transcript level per cell per haploid genome for a cell population is:

$$S = \sum_{c=1}^C \left(\sum_{g=1}^G u_{gc} / p_c \right) / C$$

Where u_{gc} denotes the number of mRNA transcripts of gene g in cell c , G is the total gene count, C is the cell count, and p_c is the ploidy of cell c . In practice, we use average ploidy Ψ of the corresponding cell group to approximate the cell level ploidy p_c , as the latter is unobservable:

$$S \approx \sum_{c=1}^C \sum_{g=1}^G u_{gc} / (C\Psi)$$

For each bulk tumor sample containing both tumor (T) and non-tumor (N) cells, *TmS* is calculated as:

$$TmS = \frac{S_T}{S_N} = \frac{T_+ / C_T \Psi_T}{N_+ / C_N \Psi_N}$$

Where T_+ and N_+ are total mRNA transcript level per cell per haploid genome in tumor and non-tumor cells respectively.

This can be reformulated using tumor-specific mRNA proportion (π) and tumor purity (ρ):

$$TmS = \frac{\pi(1 - \rho)\Psi_N}{\rho(1 - \pi)\Psi_T}$$

Where π denotes for the proportion of total bulk mRNA transcript level derived from tumor cells, $\pi = \frac{T_+}{T_+ + N_+}$; and ρ denotes for the tumor cell proportion, $\rho = \frac{C_T}{C_T + C_N}$. In practice, these parameters are estimated from bulk tumor samples: $\hat{\pi}$ is estimated through RNA/Microarray deconvolution; $\hat{\rho}$ and $\hat{\Psi}_T$ are estimated through DNA/SNP array deconvolution; $\hat{\Psi}_N$ is typically assumed to be 2 (diploid).

The final estimator for *TmS* is therefore:

$$\widehat{TmS} = \frac{\hat{\pi}(1 - \hat{\rho})\Psi_N}{\hat{\rho}(1 - \hat{\pi})\Psi_T}$$

This metric quantifies the relative total mRNA transcript level of tumor cells compared to their surrounding non-tumor cells, providing a non-biased measurement that accounts for cellular composition and genome dosage effect.

Cohort-specific TmS estimation

To acquire the input for the *TmS* model, we performed DNAseq deconvolution to estimate tumor purity and ploidy, and RNAseq deconvolution to derive tumor-specific mRNA proportion. Detail of the analysis for each cohort is described below.

METABRIC TNBC

TmS values retrieved from previously published study²⁷ (<https://wwwylab.github.io/TmS>). Triple-negative status identified by negative entries for 'ER.status', 'PR.status', and 'HER2.status' in clinical metadata. Only chemotherapy-treated patients included ($n = 118$). METABRIC excluded from pathway and cell-type deconvolution analyses due to technical limitations of microarray platforms (Figure S2A).

TCGA TNBC

TmS values retrieved from previously published study²⁷ (<https://wwwylab.github.io/TmS>). From the 127 TNBC samples with *TmS* estimates, we applied sequential exclusion criteria: (1) 11 samples with positive values for ER, PR, or HER2 IHC status were excluded based on clinical data from cBioPortal (TCGA BRCA, Firehose Legacy); for samples with missing IHC values, classification followed

Koboldt et al.⁵⁸ criteria; (2) 31 samples without documented chemotherapy treatment were excluded; (3) 1 sample with metastasis at diagnosis and 1 sample lacking clinical follow-up were excluded. The final analytic cohort comprised 83 TNBC patients with valid IHC status, chemotherapy treatment, localized primary tumors, and disease-free survival data.

SCAN-B TNBC

Tumor purity and ploidy were estimated from WGS data using ASCAT (v2.4.3).⁶⁹ To compute tumor-specific mRNA proportion (π) for SCAN-B TNBC tumors that lacked corresponding adjacent normal samples, we applied the DeMixT deconvolution pipeline to raw RNA-seq counts ($n = 235$) using two different sets of normal references: (1) GTEx breast tissue samples without significant pathology, and (2) TCGA breast cancer adjacent normal samples. We employed the ComBat⁷⁰ algorithm to remove batch effects when merging SCAN-B tumor samples with GTEx and TCGA normal samples. The optimal parameters for DeMixT are 1,500 selected genes and 50 spike-ins. Only tumor samples ($n = 208$) with a tumor purity range of ≥ 0.2 and ≤ 0.85 were chosen for TmS calculation (Figure S5A). To affirm the robustness of consensus TmS estimation, we subsequently fitted a linear regression model using $\log_2$ -transformed TmS calculated by GTEx and $\log_2$ -transformed TmS calculated by TCGA adjacent normal. The two sets of TmS estimates were highly correlated (Spearman $r = 0.86$, Figure S5B). Samples with a Cook's distance $\geq 4/n$ ($n = 10$) were discarded, and for the remaining samples ($n = 198$), the final TmS was calculated as: $TmS = \sqrt{TmS_{GTEx-N} \times TmS_{TCGA-N}}$.

FUSCC TNBC

For DNA-based deconvolution, Affymetrix OncoScan CNV SNP assays ($n = 401$) were processed with OncoScan Console (v1.3) software (Affymetrix, Inc.). Tumor purity and ploidy are estimated by ASCAT (v2.4.3)⁶⁹ using probe-level output from the OncoScan Console.

For RNA sequencing deconvolution, we downloaded fastq files of 360 primary tumors and 88 matched normal tissue from SRA with accession number SRP157974. Raw read counts were acquired via the standard RNA alignment pipeline from NCI GDC Documentation (https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline), where the fastq files were mapped to a human reference genome (Hg19, GRCh37_snp_tran). We applied the DeMixT⁷¹ deconvolution pipeline to estimate tumor-specific mRNA proportions using the 88 adjacent normal samples as the reference. Based on the simulation study in Cao et al. and observed distributions of gene selection scores in real data, we set the DeMixT parameters as top 1,500 genes and 50 spike-ins to ensure accuracy in proportion estimation.

We then kept samples ($n = 245$) that contained tumor purity, tumor ploidy, and tumor-specific mRNA proportions to derive the TmS value (Figure S5C).

Normalized total UMI counts (single-cell derived TmS) in Wu, S.Z. et al. study

With processed count matrix, we performed scale normalization on the scRNA-seq data to account for technical differences in UMI counts between patient samples. For each patient sample i , we calculated the average UMI count ($\overline{UMI}_i$) across all cells (C_i) in that sample. A population baseline was established as the median of these averages across all n patients:

$$baseline = median\{\overline{UMI}_1, \overline{UMI}_2, \dots, \overline{UMI}_n\}$$

For each patient sample i , a scale factor (r_i) was calculated as: $r_i = \frac{\overline{UMI}_i}{baseline}$. This scale factor was then applied to normalize UMI counts for all cells within each sample i : $UMI_i^{norm} = \frac{UMI_i}{r_i}$.

After normalization, we summarized metrics of each cell type t for each patient i , including average normalized UMI ($\overline{UMI}_{norm,i,t} = \frac{1}{n_{i,t}} \sum_{j=1}^{n_{i,t}} UMI_{norm,j}$). The single-cell derived TmS is defined as: $TmS_i = \frac{\overline{UMI}_{total,i,cancer}}{\overline{UMI}_{total,i,non-cancer}}$.

Where $\overline{UMI}_{total,i,cancer}$ represents the average normalized UMI counts in cancer epithelial cells, and $\overline{UMI}_{total,i,non-cancer}$ represents the average normalized UMI counts in all non-cancer cells within each patient sample.

To assess associations between single-cell derived TmS and cell type proportions, we first calculated cell type proportions as the percentage of cells of each type relative to total cells for each patient. Cell types with mean proportion $\geq 1\%$ and present in at least 6 patients were retained for further analysis. We then performed linear regression analysis with cell proportion as the dependent variable and $\log_2$ -transformed TmS as the independent variable for each cell type. Statistical significance was assessed using p -values from the regression models, with Benjamini-Hochberg adjustment for multiple testing.

This methodology enabled quantification of relative transcriptional output between cancer epithelial cells and surrounding non-malignant cells while accounting for technical variations between patient samples. Importantly, the single-cell derived TmS does not adjust for tumor cell ploidy, which is not available from this data type, so its biological implication may not completely align with the ploidy adjusted TmS calculated using bulk data.

TNBC molecular subtyping

TNBCtype-4: All samples classified using the TNBCtype online subtyping tool⁷² (<http://cbc.mc.vanderbilt.edu/tNBC/>). Each sample assigned to subtype (Basal-like—BL1, BL2, luminal androgen receptor [LAR], mesenchymal [M]) based on highest Pearson correlation to expression centroid and lowest p -value. Following Lehmann et al.,¹¹ immunomodulatory (IM) and mesenchymal stem-like (MSL) subtypes reassigned to second-highest correlated centroid due to stromal contamination.

FUSCC subtypes: Applied classification from Jiang et al.³²—basal-like immune-suppressed (BLIS), immunomodulatory (IM), mesenchymal-like (MES), and LAR.

PAM50 intrinsic subtypes: Predicted using R 'genefu' package⁷³ based on PAM50 signature.

Treatment characterization

Systematically analyzed chemotherapy regimens across cohorts to document treatment heterogeneity. Drug classes included anthracyclines (doxorubicin, epirubicin), taxanes (paclitaxel, docetaxel), and platinum agents. Treatment patterns detailed in Table S1.

Digital pathology analysis

TIL quantification

SCAN-B tumor-infiltrating lymphocyte (TIL) percentages were obtained from H&E-based quantification by Aine et al.³⁹

H&E tissue segmentation

We trained a deep learning model, Segformer,⁴⁰ on H&E images to segment the tissue into tumor, stroma, inflammatory (tissue with infiltrated lymphocytic, plasma cells, other infiltrated immune cells), necrosis, adipose tissue, benign epithelium, and blood vessel. The Segformer model was initially developed for semantic segmentation on natural scene images, unifying Transformers⁷⁴ and multi-layer perceptron decoders. Segformer has been demonstrated with efficiency, accuracy, and robustness. We adapted Segformer-MiT-B3 architecture, pertained on ImageNet,⁷⁵ for the use with segmentation on histopathology images. The annotations were collected from TCGA BRCA datasets,⁷⁶ consisting of 151 pairs of cropped H&E images and corresponding masks with varied sizes at 40x magnification ($\sim 0.25 \mu\text{m}/\text{pixel}$). Images with artifacts, including staining ink, pen marks, etc., were added as negative samples. The H&E images and corresponding masks were consistently divided into patches with a size of 896×896 , downsampled to 512×512 ($\sim 0.44 \mu\text{m}/\text{pixel}$), totaling 3110 patches. To verify the effectiveness of the model, we used patches annotated from slides of OL, LL, E2, EW, GM, and S3 as testing set, suggested by Amgad et al.,⁷⁶ achieving DICE coefficients of 0.832, 0.831, 0.735, 0.593, and 0.664 for tumor, stroma, inflammatory, necrosis, and adipose tissue, respectively. Note that DICE coefficients were calculated only for images containing the relevant types. For necrosis segmentation performance, 4 out of 13 images have a DICE coefficient of less than 0.2, while the remaining images have a DICE coefficient of 0.75 or higher. The low DICE coefficients (< 0.2) are attributed to images containing smaller areas of necrosis ($p = 0.0028$, Wilcoxon test). These data suggested that the fine-tuned Segformer model achieved a satisfactory performance for delineating tumor microenvironmental components. We then used all the patches to train the model and apply the model to whole slide images.

At the inference stage, the whole slide images were first divided into tiles of 2000×2000 pixels with a magnification of 20x ($\sim 0.44 \mu\text{m}/\text{pixel}$). Each tile was then fed to the trained deep learning model, which, in turn, generated corresponding masks for the seven tissue types detected at the pixel level. Unrecognizable tissue regions were assigned as background. The tile masks were then stitched and further downsampled to $1.25 \times$ ($\sim 7.04 \mu\text{m}/\text{pixel}$) for efficiency. We then used image processing methods, including filling holes, morphology closing operations with a disk shape of structuring element and a radius of 21, and removing individual components smaller than 4.46 mm^2 , to remove the peripheral tissue, while preserving the main tumor bed.

Spatial interaction quantification

We quantified the spatial interactions between different tissue components in the tumor microenvironment through a joint count analysis of the segmented H&E whole-slide images. For each slide, we analyzed direct pixel-to-pixel contacts between pairs of the seven tissue types (tumor, immune, stroma, necrosis, adipose tissue, non-malignant epithelial, and blood vessel).

Our algorithm evaluates each foreground pixel in the segmented image and examines its eight adjacent neighbors. For each pixel of a specific tissue type, it records all instances where neighboring pixels belonged to a different tissue type, excluding double counting, i.e., each pixel can only be designated once. For 7 tissue types, it constructs a 7×7 interaction matrix representing all possible tissue-tissue interfaces in the image.

To enable standardized comparison across samples of varying sizes, we normalized these counts by dividing the raw interaction counts by the total number of foreground pixels in each slide. We focused our analysis on four key interactions: tumor-immune cell, tumor-stromal, tumor-necrosis, and tumor-adipose interfaces. We compared these normalized interaction percentages between high-TmS and low-TmS groups using Kruskal-Wallis tests to identify significant differences in spatial organization.

Pathway and gene module analyses

We examined the association of TmS with transcriptomic alteration. For each TNBC cohort, we conducted GSEA⁷⁷ on the HALLMARK⁴⁵ and KEGG⁴⁶ pathways. All genes were ranked by the Spearman correlation coefficient between their expression levels and TmS across samples within each cohort; they were then put through GSEA in the 'pre-ranked' mode. For GSEA, we adopted permutation tests (1,000 iterations) to generate a normalized enrichment score (NES) for each candidate pathway. To ensure robust cross-population signals with meaningful biological effect sizes, pathways were retained for visualization if they met significance thresholds of BH-adjusted $p < 0.1$ across all cohorts and achieved $|\text{NES}| \geq 1.5$ in at least one cohort.

For population-specific mechanistic analysis, we performed targeted GSEA focusing on refined RAS/MAPK and proliferation pathway sets from MSigDB collections.^{45,47,78} The analysis included 4 RAS/MAPK pathways (KRAS signaling up, MAPK family signaling cascades, MAPK1/ERK2 activation, MAPK3/ERK1 activation) and 7 core proliferation pathways (E2F targets, G2M checkpoint, MYC targets V1/V2, DNA replication, mitotic prometaphase, mitotic metaphase and anaphase).

Gene expression modules were obtained from Fredlund, E. et al.¹³ The Association of TmS and gene modules was evaluated by calculating the Spearman correlation coefficient between the aggregated gene expression for each gene module and the TmS across samples within each cohort.

Spearman correlations between continuous $\log_2(\text{TmS})$ values and mean expression of predefined gene sets representing specific cell types and functional states. Gene sets included cytotoxic T cell markers (*CD8A*, *CD8B*, *GZMA*, *GZMB*, *PRF1*), exhaustion markers (*PDCD1*, *CTLA4*, *LAG3*, *TIGIT*, *HAVCR2*), CAF markers (*COL1A1*, *COL3A1*, *FN1*, *FAP*, *ACTA2*), and other immune/stromal signatures. Only correlations with $p < 0.05$ were considered significant.

Cell type deconvolution

The cellular composition was estimated using the impute cell fractions module of CIBERSORTx (<https://cibersortx.stanford.edu/>). The reference matrix of cell-type-specific expression was derived from a single-cell atlas for breast cancers,²⁹ which included 10 primary untreated TNBC samples. In the reference matrix, nine distinct cell types were annotated: myeloid cells, plasmablast cells, T cells, B cells, cancer epithelial cells, normal epithelial cells, perivascular-like cells (PVLs), cancer-associated fibroblasts (CAFs), and endothelial cells. A maximum of 500 cells per cell type was selected, and all available cells were utilized for cell types with fewer than 500 cells. Subsequently, we applied CIBERSORTx's batch correction method (S-mode batch correction) for the deconvolution of bulk samples. The deconvolution mode was set to 'absolute' with 1000 permutations.

Additionally, for a focused estimation of immune cell populations, we performed deconvolution using the LM22 reference matrix. To account for technical variation, we applied the recommended built-in batch correction method (batch mode = "B"). Deconvolution was conducted in absolute mode with 1,000 permutations.

T cell dysfunction and exclusion analysis

To evaluate the immunomodulatory roles in the high/low TmS group, we calculated T cell Dysfunction and Exclusion (TIDE) scores for all four cohorts using the TIDE⁴⁴ python tool. TIDE first used the average expression level of *CD8A*, *CD8B*, *GZMA*, *GZMB*, and *PRF1* to estimate the cytotoxic T-lymphocyte (CTL) level in each sample within each cohort. Patients with a higher and lower CTL level compared to the mean CTL level within the cohort were stratified into high and low CTL groups, respectively. For T cell dysfunction analysis, TIDE evaluates whether gene signatures from each cohort influence the beneficial effect of CTL levels on patient prognosis. This is performed using the interaction coefficient d from the Cox proportional hazard (Cox-PH) model to evaluate how the interaction between a candidate gene and the CTL affects the death hazard. For T cell exclusion analysis, TIDE evaluates the correlation between the CTL level and the expression profiles of three cell types that have been reported to restrict T cell infiltration in tumors – cancer-associated fibroblasts (CAFs), myeloid-derived suppressor cells (MDSCs) and tumor-associated macrophages (TAMs). In each cohort, a Wilcoxon rank-sum test was used to compare the distributions of dysfunction and exclusion scores between the two TmS groups. The exclusion score reflects potential barriers to T cell infiltration through gene signatures associated with immunosuppressive microenvironments. These signatures include factors secreted by myeloid cells (i.e., *VEGFA*, *CSF1*, *IL10*) and cancer-associated fibroblasts (i.e., *TGFB1*, *FAP*, *CXCL12*) that can affect T cell recruitment and migration.⁴⁴ Meanwhile, the dysfunction score captures T cell exhaustion based on expression of inhibitory receptors and related factors (i.e., *PDCD1*, *CTLA4*, *LAG3*, *TIGIT*, *HAVCR2*).⁴⁴

Differential gene expression and integrative enrichment analysis

Differential gene expression analysis was conducted using the DESeq2⁷⁹ package in R, following its recommended workflow for differential gene expression analysis. Adjusted p -values were computed using the Benjamini-Hochberg procedure to control the false discovery rate (FDR), with an alpha level set at 0.05 for statistical significance. The fold-change thresholds were established at $\geq \log_2(1.5)$ for upregulated genes and $\leq -\log_2(1.5)$ for downregulated genes. Given the intrinsic heterogeneity and potential batch effects between the TCGA and SCAN-B cohorts, each dataset was analyzed independently before the results were integrated. Inter-section analyses were performed to identify consistently differentially expressed genes (DEGs) across both cohorts.

To gain biological insights into the genes found to be differentially expressed across the two cohorts, we performed integrative pathways analysis using the ActivePathways⁸⁰ package in R. We kept consistent DEGs and ranked the genes by adjusted p -values from both TCGA and SCAN-B. The input for ActivePathways includes a matrix of adjusted p -values of genes as rows and datasets as columns and a list of gene sets in the form of a GMT (Gene Matrix Transposed) file, the gene sets we acquired are from Gene Ontology⁵⁵ and Reactome⁸¹ databases. Enrichment map visualization was done by the EnrichmentMap app⁸² of Cytoscape⁸³ for network visualization of similar pathways and their coloring according to supporting evidence.

QUANTIFICATION AND STATISTICAL ANALYSIS

Sample size and data characteristics

Total $n = 575$ chemotherapy-treated TNBC patients across four cohorts. No statistical methods were used to predetermine sample size. All available patients meeting inclusion criteria were analyzed. Cohort-specific technical specifications in Table S1.

Normalization of raw counts from microarray or RNAseq data

Scale normalization at the seventy-fifth percentile based on the DSS package⁸⁴ was then applied to the post quality-control tumor and normal samples. Two criteria applied to filter spurious genes: (1) filtered genes with zero count in either mixed tumor or normal samples, (2) filtered genes with large variance in normal reference samples.

Association with clinical variables

Kruskal-Wallis tests compared TmS distributions between subgroups defined by each clinical variable. *p*-values adjusted using Benjamini-Hochberg correction across all available clinical variables within each TNBC cohort.

Survival analyses

Association between TmS and survival outcomes examined in context of DFS, RFS, and RFI, contingent upon available survival metrics within each cohort. Employment of DFS, RFS, and RFI as primary clinical endpoints follows methodological guidelines established by Liu et al.,⁸⁵ as they are comparable for primary TNBC.

For all association analyses with clinical outcomes across all cohorts, we used recursive partitioning survival tree model (rpart⁸⁶) to find optimal TmS cutoff (high vs. low; high vs. medium vs. low in FUSCC, Table S1) separating different survival outcomes. Splits assessed using Gini index, maximum tree depth set to 2. Survival disparities between high- and low-TmS groups statistically assessed using log rank tests.

Cox regression analyses

Two multivariate Cox proportional hazard models utilized in this study. In each model, we considered age (categorized as ≥ 50 years vs. < 50 years), LN status (positive vs. negative), tumor size (categorized as > 20 mm vs. ≤ 20 mm), and tumor grade (graded as 3 vs. 2) as covariates, along with either TmS as categorical (high vs. low) or continuous variable. Both models employed to interpret RFS (or DFS when RFS not available) for TNBC.

Association between categorical variables

Contingency table analyses were performed to assess associations between TmS classifications (high/low or high/medium/low) and TNBC molecular subtypes (TNBCType-4, FUSCC subtypes) using Chi-square tests. Effect sizes were quantified using Cramer's *V* to evaluate the strength of associations. Statistical significance was set at $p < 0.05$.

Ward distance analyses

To assess how well the identified feature set, where each feature presented moderate correlations with TmS, collectively separates TmS-defined groups, we calculated Ward's distance with permutation testing for uncertainty. Features significantly correlated with TmS scores ($|r| \geq 0.2$, adjusted $p < 0.05$) were selected for each cohort and Z score normalized before calculating Ward's minimum variance distances between groups.

SCAN-B and TCGA European cohorts were dichotomized as high versus low TmS groups using their respective optimal cutoffs (Table S1), while the FUSCC cohort was stratified into high (TmS ≥ 2.5), medium ($0.88 \leq \text{TmS} < 2.5$), and low (TmS < 0.88) groups. Statistical significance was determined through 1,000 permutation iterations, with observed distances compared against the 0.5th and 99.5th percentile thresholds of the empirical distribution ($p < 0.001$, two-tailed). For SCAN-B, the Ward distance between high and low TmS groups was 177.1 (permutation-based 99% CI: 0.1–18.2, $p < 0.001$). For TCGA European sub-population, the distance was 25.3 (99% CI: 0.2–16.4, $p < 0.001$). For FUSCC's three-group stratification, Ward distances were 186.2 (high vs. medium), 52.7 (high vs. low), and 71.3 (medium vs. low), with permutation-based 99% CIs of 0.1–18.6, 0.1–19.6, and 0.1–17.2 respectively (all $p < 0.001$). All analyses were conducted using the WCluster package⁸⁷ in R.

Multiple testing correction

Benjamini-Hochberg procedure applied for multiple testing correction within analytical categories as appropriate throughout the study.

Statistical software and significance

All analyses performed in R (version ≥ 4.0). Significance levels: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. Exact sample sizes, statistical tests, and *p*-values specified in figure legends.