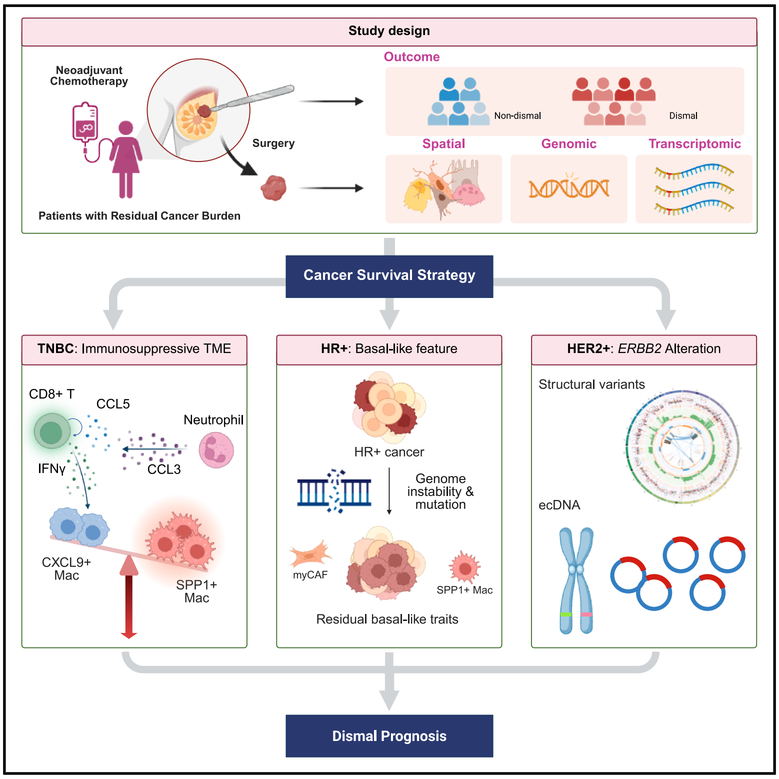
Cell Reports
Medicine

Article

Spatial and genomic profiling of residual breast cancer after neoadjuvant chemotherapy unveil divergent fates for each breast cancer subtype

Graphical abstract



Highlights

- CXCL9+ macrophage-CD8$^+$ T cell interactions improve survival outcome of TNBC
- In non-TNBC, residual basal-like cancer cells are strongly linked to poor prognosis
- Homologous recombination deficiency predicts worse outcomes in HR+HER2− BC
- *ERBB2* DNA alterations correlate with poor outcomes in HER2+ BC

Authors

Eun Seop Seo, Sabin Park,
Eun Yoon Cho, ..., Semin Lee,
Woong-Yang Park, Ji-Yeon Kim

Correspondence

wisekh@skku.edu (H.K.),
seminlee@unist.ac.kr (S.L.),
woongyang.park@samsung.com (W.-Y. P.),
jyeon25@skku.edu (J.-Y.K.)

In brief

Seo et al. demonstrate that CXCL9+ macrophages interacting with CD8$^+$ T cells improve survival outcome in patients with residual TNBC after neoadjuvant chemotherapy, while SPP1+ macrophages and basal-like features combined with HRD predict worse survival in HR+HER2− BCs, and *ERBB2* structural variants predict worse survival outcomes in HER2+ BCs.

Seo et al., 2025, Cell Reports Medicine *6*, 102164
June 17, 2025
https://doi.org/10.1016/j.xcrm.2025.102164

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

Spatial and genomic profiling of residual breast cancer after neoadjuvant chemotherapy unveil divergent fates for each breast cancer subtype

Eun Seop Seo,[1,2,14] Sabin Park,[3,14] Eun Yoon Cho,[4,14] Jeong Eon Lee,[5,6,14] Hae Hyun Jung,[7] Jiyeon Hyeon,[8] Sepil An,[9] Seok Won Kim,[5] Junghoon Shin,[8] Jin Seok Ahn,[6,8] Yeon Hee Park,[6,8] Young-Hyuck Im,[6,7,8] Hoon Kim,[9,10,11,*] Semin Lee,[3,*] Woong-Yang Park,[2,6,12,13,*] and Ji-Yeon Kim[1,6,7,8,15,*]

[1]Department of Digital Health, Samsung Advanced Institute for Health Sciences and Technology, Seoul 06355, Republic of Korea
[2]Geninus Inc., Seoul 05836, Republic of Korea
[3]Department of Biomedical Engineering, College of Information-Bio Convergence Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea
[4]Department of Pathology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Republic of Korea
[5]Division of Breast Surgery, Department of Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Republic of Korea
[6]Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences and Technology, Seoul 06355, Republic of Korea
[7]Biomedical Research Institute, Samsung Medical Center, Seoul 06351, Republic of Korea
[8]Division of Hematology-Oncology, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Republic of Korea
[9]Department of Biopharmaceutical Convergence, School of Pharmacy, Sungkyunkwan University, Suwon-si, Gyeonggi-do 16419, Republic of Korea
[10]Department of Biohealth Regulatory Science, School of Pharmacy, Sungkyunkwan University, Suwon-si, Gyeonggi-do 16419, Republic of Korea
[11]Epigenome Dynamics Control Research Center, Sungkyunkwan University, Suwon-si, Gyeonggi-do 16419, Republic of Korea
[12]Translational Genomics Center, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Republic of Korea
[13]Department of Molecular Cell Biology, Sungkyunkwan University School of Medicine, Suwon, Gyeonggi-do 16419, Republic of Korea
[14]These authors contributed equally
[15]Lead contact
*Correspondence: wisekh@skku.edu (H.K.), seminlee@unist.ac.kr (S.L.), woongyang.park@samsung.com (W.-Y.P.), jyeon25@skku.edu (J.-Y.K.)

https://doi.org/10.1016/j.xcrm.2025.102164

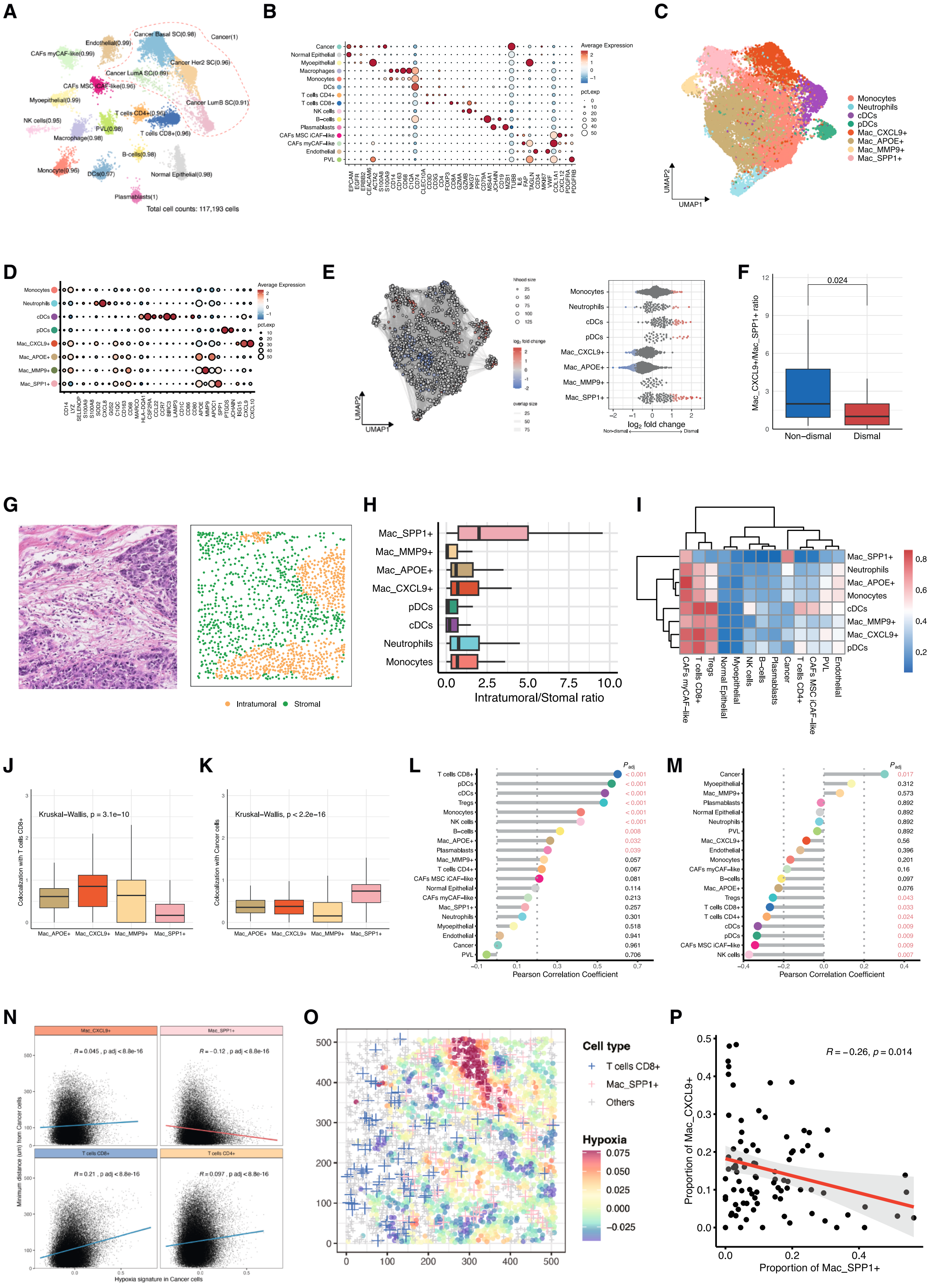
SUMMARY

Residual cancer burden (RCB) is a strong prognostic marker after neoadjuvant chemotherapy (NAC) in breast cancer (BC), yet some BCs defy their predicted outcomes. Using single-cell spatial transcriptomics and genomic profiling, we investigate mechanisms underlying divergent fates of BCs with high RCB across subtypes. In triple-negative BC (TNBC), CXCL9+ macrophage-CD8$^+$ T cell interactions via chemokines and interferon-gamma signaling promote favorable outcomes, while SPP1+ macrophage-cancer cell interactions driven by hypoxia signaling correlate with poor prognosis. In non-TNBC, the extent of basal-like cancer cells and their proximity to scarce immune cells are linked to prognosis. Additionally, tumor-intrinsic features—such as homologous recombination deficiency in hormone receptor (HR)-positive cancers and structural variations, including extrachromosomal *ERBB2* DNA in human epidermal growth factor receptor 2 (HER2)-positive cancers—predict worse outcomes. This study highlights distinct genomic and microenvironmental strategies governing BC subtype-specific fates after NAC.

INTRODUCTION

Neoadjuvant chemotherapy (NAC) followed by curative surgery is the standard treatment strategy in early breast cancer (EBC) according to the treatment guideline.[1] NAC response strongly predicts outcomes, with pathologic complete response (pCR) indicating better disease-free survival (DFS) and overall survival (OS) than non-pCR.[2] In addition, residual cancer burden (RCB), calculated based on residual tumor bed and lymph node metastases after NAC, is the most powerful prognostic factor in EBC.[3,4] Higher RCB class correlates with increased recurrence risk and poorer survival across all subtypes,[5] guiding adjuvant treatment decisions.[6–9]

However, breast cancer (BC) with a high RCB class does not always result in recurrence.[3,5] According to previous studies, up to 40% of RCB class III BC has not experienced recurrence within a 10-year follow-up period. Furthermore, even in RCB class III triple-negative BC (TNBC) after NAC,



(legend on next page)

30% of TNBCs achieve long-term survival without disease recurrence.

The reason for the survival disparity in BC with the same RCB class remains unclear. Researchers hypothesize that tumor genomic characteristics play a role. For example, BC transcriptomic assays including MammaPrint and Oncotype DX precisely predict the benefit of adjuvant chemotherapy for hormone receptor (HR)-positive human epidermal growth factor receptor 2 (HER2)-negative BC.[10,11] The recent HER2DX predicts HER2+ BC prognosis.[12] However, no genomic test currently exists that can precisely predict recurrence or death in BC with RCB class III after NAC.

Here, we explore the spatial and genomic characteristics of residual BC receiving NAC followed by curative surgery. Using spatial transcriptomics combined with bulk RNA sequencing (RNA-seq), we illustrate intricate interactions between tumor and tumor microenvironment (TME). In addition, we demonstrate how genomic features, such as mutational profiles and structural variants, shape the fate of residual BC. Overall, we reveal the spatial and genomic characteristics that significantly influence survival outcomes of BC patients who did not achieve pCR after NAC.

RESULTS

Clinical characteristics

We analyzed 91 BC cases, with 60 classified as dismal prognosis (D group) and 31 as non-dismal (ND group). All 91 BC specimens were residual cancers at curative surgery following NAC. D group patients all experienced distant metastasis within 2 years of surgery, with 50 (83.3%) suffering BC-specific death within 5 years of metastasis. No ND group patients experienced disease progression or BC-specific death during 7 years of follow-up.

Detailed BC characteristics appear in Table S1. The D group comprised 12 (20.0%) HR+HER2−, 4 (6.7%) HR+HER2+, 34 (56.7%) TNBC, and 10 (16.7%) HR−HER2+ BCs. Fourteen cases (23.3%) had RCB class II, while 46 (76.7%) had RCB class III. The ND group included 11 (35.5%) HR+HER2−, 2 (6.5%) HR+HER2+, 14 (45.2%) TNBC, and 4 (12.9%) HR-HER2+ BCs. Sixteen cases (51.6%) had RCB class II, and 15 (48.4%) had RCB class III. Of all patients, 70 completed standard NAC, while 21 did not complete treatment due to non-response or intolerability.

Single-cell spatial analysis indicates spatial separation of CXCL9+ and SPP1+ Macs in residual TNBC after neoadjuvant treatment

We employed the NanoString CosMx Spatial Molecular Imaging (SMI) system to spatially examine gene expression at single-cell resolution across 88 fields of view (FOVs) in fifteen TNBC patients using the 1,000-plex RNA panel (Table S2). To understand the spatial organization of various cell types within TNBC, we first conducted cell type annotation using the InSituType[13] supervised clustering method, with single-cell RNA sequencing (scRNA-seq) data from Wu et al.[14] serving as a reference (Figures 1A and 1B). This analysis identified 15 distinct cell types across 116,193 cells.

In addition, intrinsic subtyping was possible using the supervised clustering method, revealing a strong correlation among the single-cell intrinsic subtypes, those identified in bulk RNA-seq data and reference data (Figures S1A–S1C).

Focusing on myeloid cells, further sub-clustering using Louvain clustering identified monocytes, four distinct macrophages (Macs), two distinct dendritic cell (DC) subtypes, and neutrophils. Specifically, four Mac clusters were distinguished by their unique gene signatures: APOE+ Mac, CXCL9+ Mac, MMP9+ Mac, and SPP1+ Mac (Figures 1C and 1D).

Milo[15] analysis revealed that these myeloid cells were distributed differently according to TNBC prognosis (Figure 1E). APOE + Macs and CXCL9+ Macs were enriched in the ND group, while SPP1+ Macs were more prevalent in the D group. Notably, the ratio of CXCL9+ Mac to SPP1+ Mac was significantly higher in the ND group compared to the D group (Figure 1F).

Next, we investigated spatial distribution of these cells by dividing the FOVs into two areas, intratumoral and stromal regions, using the SPIAT package[16] (Figure 1G). Between these two areas, SPP1+ Mac was more abundant in the intratumoral region compared to the stroma, unlike other myeloid cells (Figure 1H). Further analysis of cell-cell co-localization through normalized mixing scores (NMSs) within a 50 μm radius revealed distinct co-localization patterns for each Mac subtype (Figure 1I). SPP1+ Macs showed a significantly higher co-localization score with cancer cells compared to other Macs ($p < 2.2 \times 10^{-16}$), whereas CXCL9+ Macs had a significantly higher co-localization score with CD8$^+$ T cells ($p = 3.1 \times 10^{-10}$) (Figures 1J and 1K). Further broader analysis of cell co-localization, considering the

Figure 1. Spatial profiling identifies distinct myeloid cell subtypes and spatial features associated with dismal prognosis in post-NAC residual TNBC

(A) Flightpath plot showing InSituType cell identification with posterior probability.
(B) Dotplot of signature marker expression by cell type.
(C) Uniform manifold approximation and projection (UMAP) of reclustered myeloid cells.
(D) Dotplot of signature markers for myeloid cells.
(E) Milo cellular neighborhood differential abundance plots (false discovery rate [FDR] < 0.1) showing myeloid composition changes (left); beeswarm plot of log-fold changes in neighborhoods by subcluster (right).
(F) Boxplots of CXCL9+/SPP1+ Mac ratios.
(G) FOV example showing spatial cell classification relative to margin.
(H) Comparison of spatial distributions of myeloid cells.
(I) Heatmap of myeloid-other cell type co-localization within 50 μm.
(J and K) Boxplots showing the co-localization of macrophages with CD8$^+$ T cells (J) and cancer cells (K).
(L and M) Pearson correlation analyses between the proportions of CXCL9+ Macs (L) or SPP1+ Macs (M) and the counts of other cell types in FOV.
(N) Correlation between cancer cell hypoxia signatures and minimal distance to specified cell types.
(O) Visualization of the cancer hypoxia signature and the spatial distribution of CD8$^+$ T cells and SPP1+ Macs.
(P) Correlation between the proportions of CXCL9+ and SPP1+ in each FOV. Boxplots show medians and interquartile ranges; scatterplots show Pearson's *R* and regression lines with 95% confidence intervals.

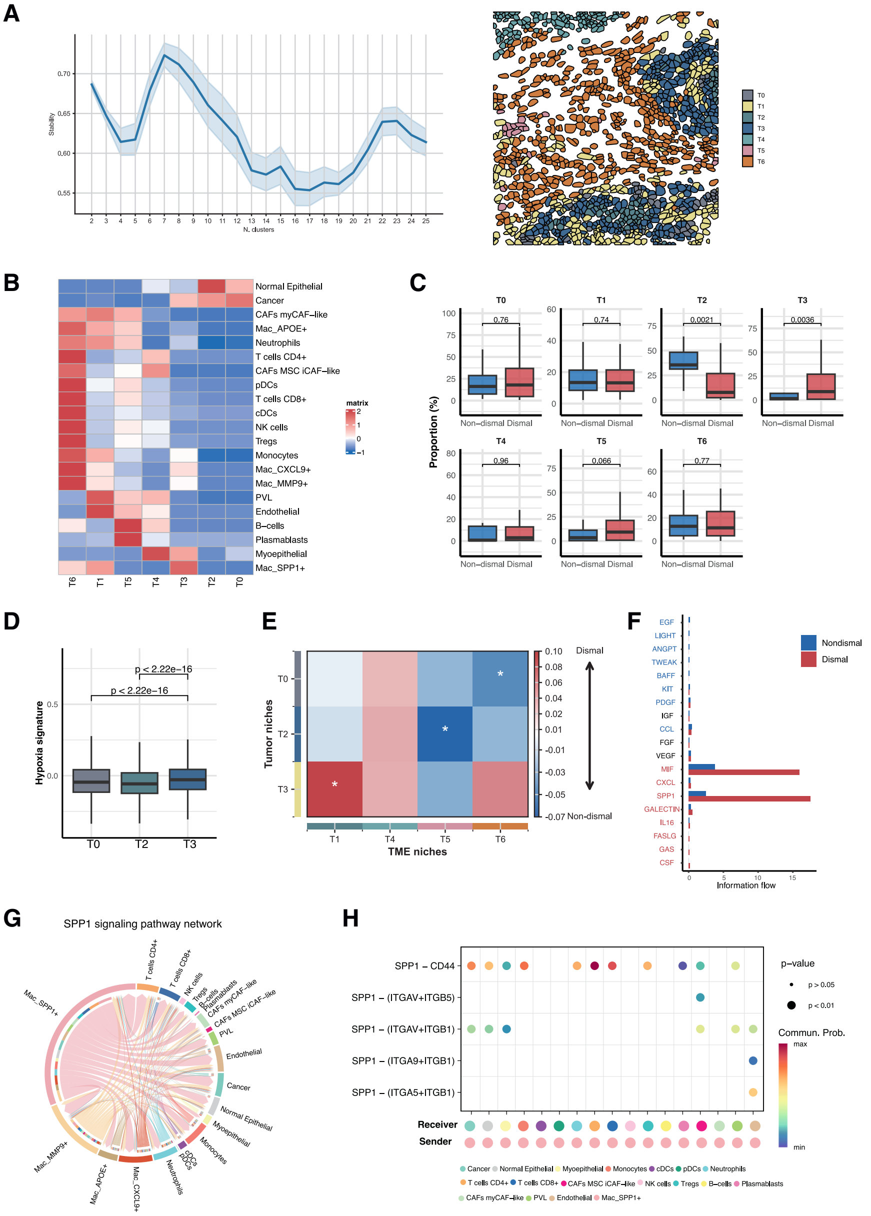


Figure 2. Enriched interaction between hypoxia-driven SPP1+ Macs and tumor niches associated with dismal prognosis in residual TNBC post-NAC

(A) CellCharter cluster stability across varying cluster numbers (left) and a representative spatial distribution of niches derived from clusters (right).

(B) Heatmap of cell type enrichment across identified niches.

(C and D) Boxplots comparing niche proportions (C) and cancer cell hypoxia signatures across tumor-enriched niches (D) between ND and D groups.

(E) Differential neighborhood enrichment analysis between tumor-enriched (source) and TME-enriched (target) niches in ND vs. D. Asterisks: $p < 0.05$; two-sided tests.

(F) Bar plot showing information flow differences in inferred networks between ND and D, with significant signaling pathways ranked by variation.

(G) Chord diagram of SPP1 signaling interactions between cell types.

(H) Bubble plot of significant interactions from SPP1+ Macs to other cells. For all the boxplots, data are presented as median with interquartile range.

entire FOV, revealed that the proportion of CXCL9+ Macs within the total Mac population was strongly associated with the abundance of $CD8^+$ T cells, followed by conventional DCs (cDCs), regulatory T cells (Tregs), monocytes, and natural killer (NK) cells (all adjusted $p < 0.001$) (Figure 1L). In contrast, the proportion of SPP1+ Macs was primarily associated with the abundance of cancer cells (adjusted $p = 0.017$) (Figure 1M).

Hypoxia signal from cancer cells recruits SPP1+ Macs, excludes other immune cells, and predicts dismal prognosis

We evaluated catalytic factors that promote the co-existence of SPP1+ Macs and cancer cells. We focused on the result of enriched gene set signatures[17] in dismal cancer cells (Figures S1D and S1E) and found that the hypoxia signature score, a known contributing factor for SPP1+ Macs, was higher in the D group compared to the ND group in TNBC. Among the key pathways analyzed, the hypoxia signature showed the strongest correlation with the presence of SPP1+ Macs (Figure S1F). We then calculated the distance between cancer cells and four types of immune cells and evaluated the correlation between this distance and the hypoxia signature of the cancer cells. This analysis revealed a negative correlation between hypoxia signature strength and the distance to SPP1+ Macs (R = -0.12, adjusted $p = 8.8 \times 10^{-16}$; Figure 1N) whereas CXCL9+ Macs, $CD8^+$ T cells, and $CD4^+$ T cells were found to be further from hypoxic cancer cells (figures 1n and 1O). Additionally, a reciprocal change in polarity between CXCL9+ and SPP1+ Macs was observed, indicating a negative correlation between their spatial distributions (R = -0.26, $p = 0.014$) (Figure 1P).

We used CellCharter[18] to investigate how cellular niches composed of SPP1+ Macs and cancer cells influence BC prognosis. Through cluster stability analysis, we identified seven stable clusters, which guided us in partitioning our spatial transcriptomic data accordingly (Figure 2A). Among the seven clusters, T0, T2, and T3 primarily consisted of cancer cells, with T3 showing additional enrichment of SPP1+ Macs as well as cancer cells. In addition, T6 exhibited a high enrichment of a diverse immune cell population (Figure 2B). T2, characterized by cancer cells and normal

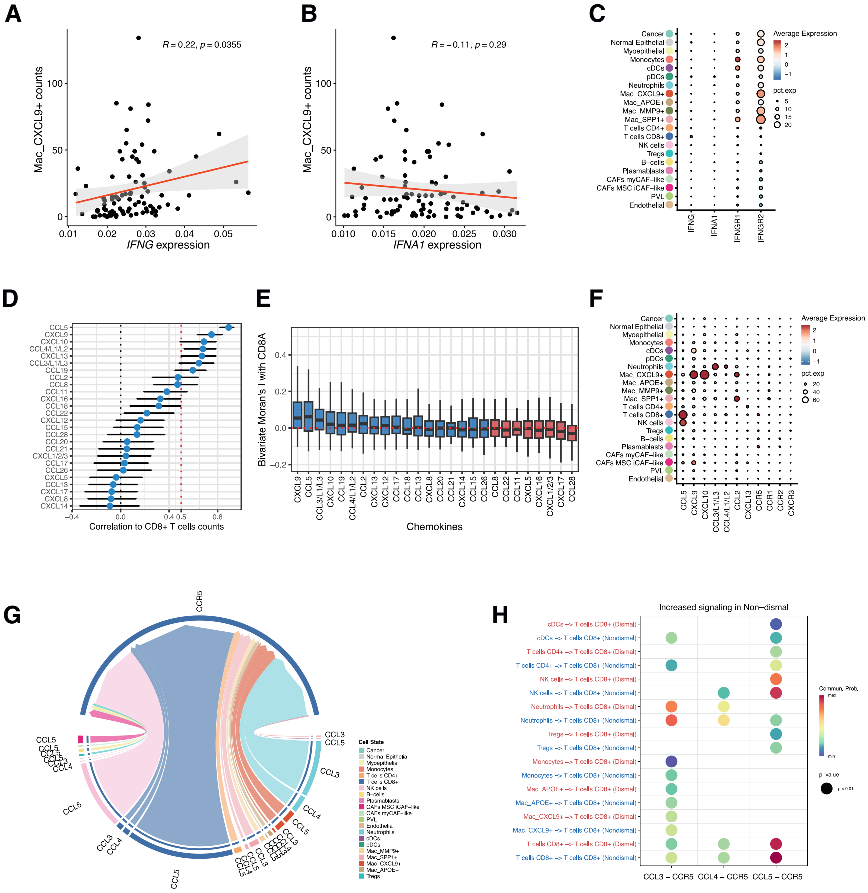


Figure 3. Spatial and molecular interactions between CXCL9+ Macs and CD8$^+$ T cells

(A and B) Scatterplots of correlations between average *IFNG* (A) or *IFNA1* (B) expression levels and CXCL9+ Mac abundance per FOV.

(C) Dotplot showing *IFNG*, *IFNA1*, and receptor expression across cell types.

(D) Correlation analysis of CD8$^+$ T cell abundance with average *CCL/CXCL* chemokine gene expression per FOV. Spearman's rank correlation estimates (circles) are shown with 95% confidence intervals.

(E) Boxplot of spatial autocorrelation using bivariate Moran's I between *CD8A* and chemokine genes. Data are presented as median with interquartile range.

(F) Dotplot showing *CCL/CXCL* chemokine gene expression across cell types.

(G) Chord plot of ligand-receptor pairs in CCL pathways mediating interactions received by CD8$^+$ T cells.

(H) CellChat communication probability analysis of upregulated ligand-receptor interactions in *CCL* pathways received by CD8$^+$ T cells in ND vs. D FOVs ($p < 0.01$).

epithelial cells, was predominantly enriched in ND TNBC, whereas T3, containing cancer cells, myoepithelial cells, and SPP1+ Macs was associated with D TNBC ($p = 0.0021$ and $p = 0.0036$, respectively) (Figure 2C). We also found that the hypoxia score was significantly higher in the T3 niche compared to other cancer-enriched niches ($p < 2.2 \times 10^{-16}$), which may explain why cancer cells and SPP1+ Macs are co-clustered in the T3 niche (Figure 2D). Notably, interaction analyses between cancer niches and other TME niches suggested that T3 preferentially interacted with T1, which is an SPP1+ Mac-enriched TME niche, in D TNBCs ($p = 0.030$), while T2 and T0 more prominently interacted with T5 (B cell- and plasmablast-enriched TME niche) and T6 niche (T cell-enriched niche), respectively, in ND TNBCs ($p = 0.022$, $p = 0.010$, respectively) (Figure 2E).

To address how SPP1+ Macs contribute to dismal prognosis beyond their association with hypoxic tumors, intercellular communication analysis using CellChat[19] was performed. This analysis highlighted that *SPP1* signaling was highly activated in D TNBC (Figure 2F), with SPP1+ Macs emerging as the primary sender cells (Figure 2G). Notably, the *SPP1-CD44* ligand-receptor pair was the most prevalent interaction (Figure 2H). By engaging in *SPP1-CD44* signaling, SPP1+ Macs may enhance intercellular signal boosting therapy resistance and suppress T cell activation.[20,21] These combined effects may create an aggressive TME that drives dismal outcomes.

CXCL9+ Mac polarization reflects a coordinated chemokine network that promotes CD8$^+$ T cell infiltration and predicts improved survival in TNBC

Having identified CXCL9+ Macs as favorable prognostic factors, we sought to uncover the drivers behind their polarization. Given the well-established role of interferons (IFNs) in regulating *CXCL9* expression, we first investigated their impact on these cells. As expected, average IFN-γ expression per FOV was positively correlated with the number of CXCL9+ Macs (R = 0.22, $p = 0.0355$) (Figure 3A), while IFN-α expression did not show any significant correlation (R = −0.11, $p = 0.29$) (Figure 3B). Notably, IFN-γ was predominantly expressed in CD8$^+$ T cells, even though its receptors were not specific to CXCL9+ Macs (Figure 3C).

As we attributed the favorable impact of CXCL9+ Macs to their role in recruiting CD8$^+$ T cells via *CXCL9* and *CXCL10* and recognized that their expression is regulated by IFN-γ from CD8$^+$ T cells, we hypothesized that additional factors might facilitate the early recruitment of CD8$^+$ T cells—potentially setting the stage for subsequent *CXCL9* induction. To explore this further, we identified chemokines associated with CD8$^+$ T cell recruitment, finding that, besides *CXCL9* and *CXCL10*, *CCL5*, *CCL4*, and *CCL3* were also strongly correlated with CD8$^+$ T cell infiltration (Figure 3D). Spatial interactions, analyzed using the bivariate Moran's index[22] further revealed that *CXCL9*, *CCL5*, and *CCL3* were spatially correlated with *CD8A* expression, highlighting the co-existence and mutual interactions between CD8$^+$ T cells and CXCL9+ Macs, as well as the involvement of additional chemokines in supporting this process. Conversely, chemokines such as *CCL28*, *CXCL17*, and

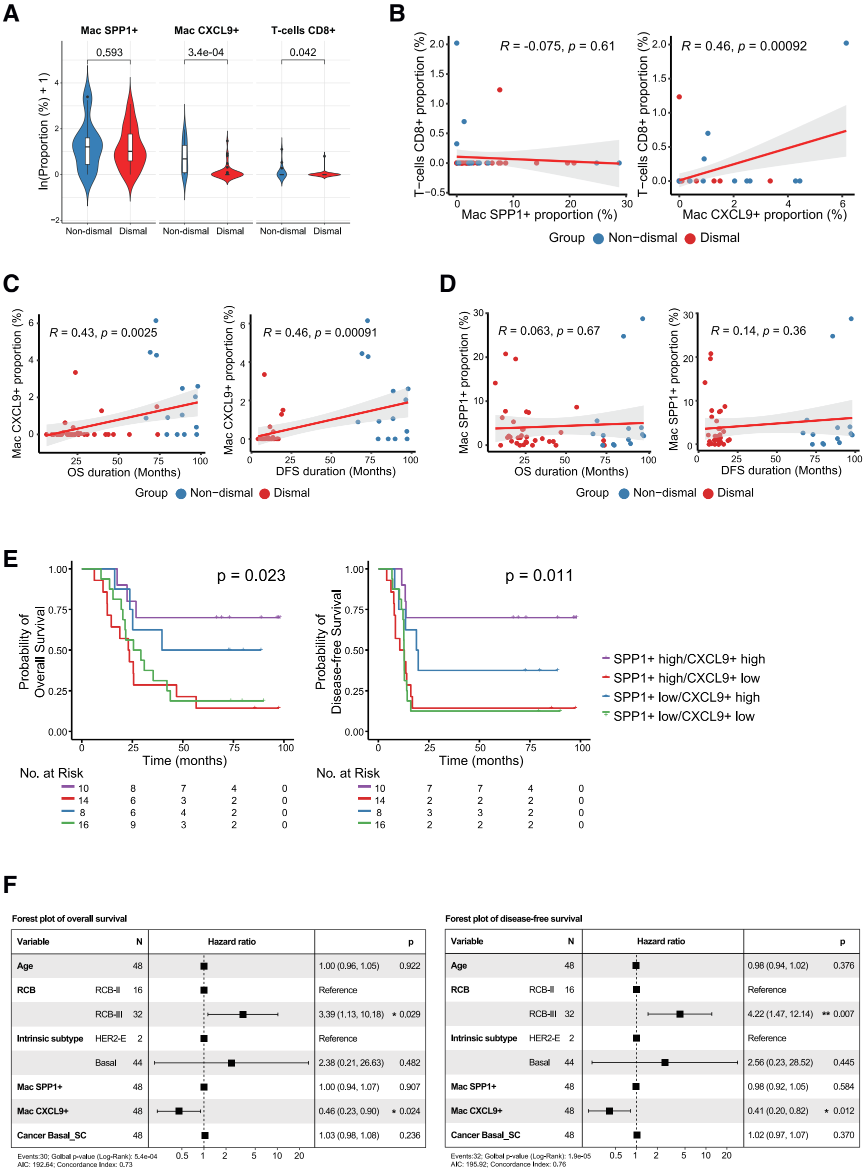


Figure 4. Validation of the prognostic significance of CXCL9+ Macs and CD8$^+$ T cells using bulk RNA-seq

(A) Violin plots comparing CXCL9+ Macs, SPP1+ Macs, and CD8$^+$ T cell proportions by prognosis (Wilcoxon rank-sum test). Data are presented as median and interquartile range.

(B–D) Scatterplots correlating SPP1+ and CXCL9 + Macs with CD8$^+$ T cell proportions (B), CXCL9 + Mac proportion with survival duration (C), and SPP1+ Mac proportion with survival duration (D).

(E) Kaplan-Meier survival curves grouped by SPP1 + Mac and CXCL9+ Mac proportions (groups determined by median values).

(F) Forest plots of Cox multivariable regression for overall survival (left) and disease-free survival (right). Hazard ratios shown as rectangles with 95% confidence intervals. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

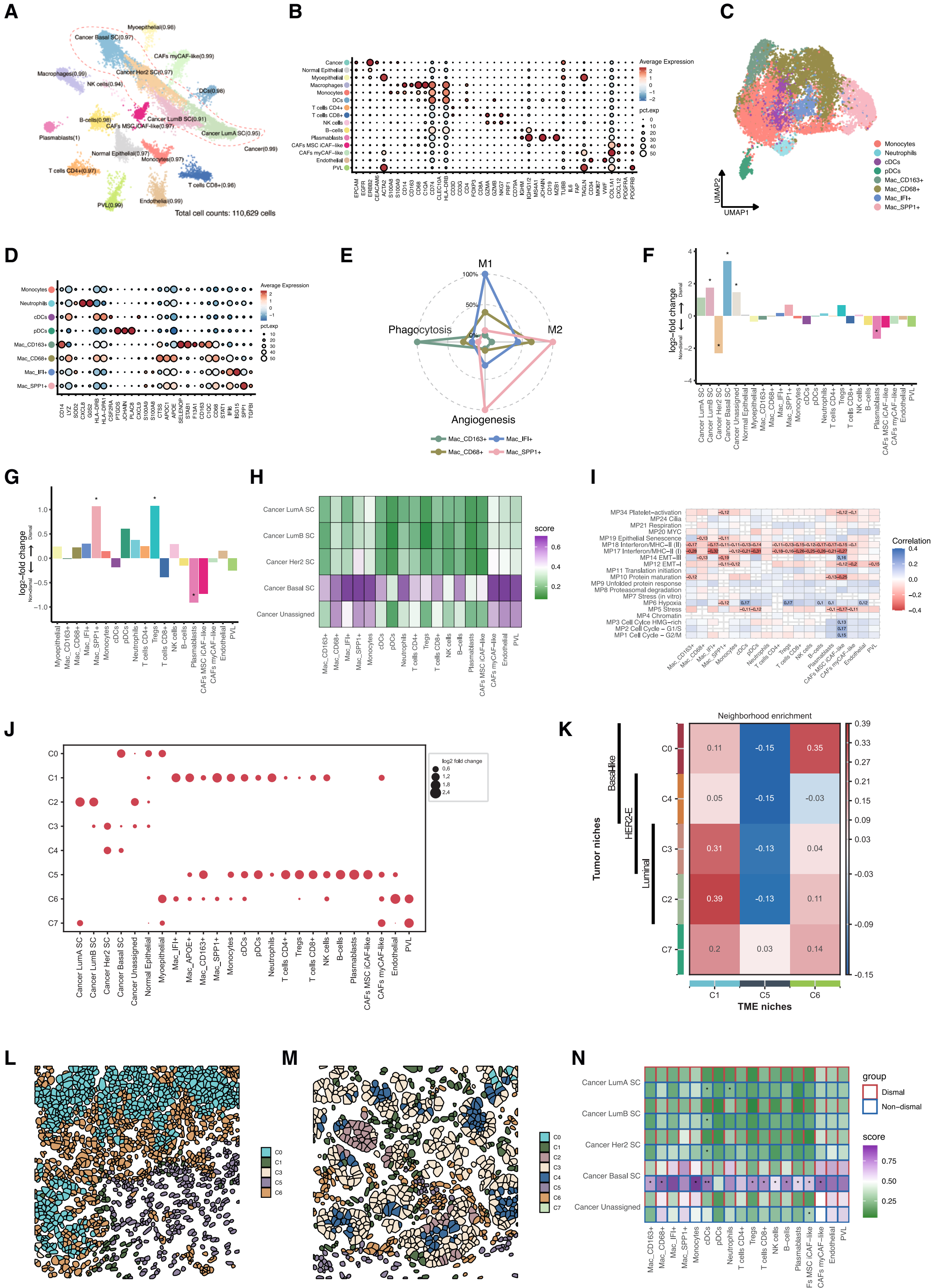
For scatterplots, *R* represents Pearson's correlation coefficient; solid lines show fitted linear regression values with 95% confidence intervals.

CXCL1/2/3 had weaker or negative correlations (Figures 3E, S1G, and S1H).

We further investigated the sources of these chemokines and found that *CCL5* was primarily produced by CD8$^+$ T cells and NK cells, while *CCL3* and *CCL4* were mainly expressed by neutrophils (Figure 3F). A spatial cell-cell communication analysis demonstrated that *CCL3*, predominantly secreted by neutrophils, interacts with *CCR5* expressed on CD8$^+$ T cells, in addition to the *CCL5-CCR5* interaction between CD8$^+$ T cells (Figure 3G). Notably, these interactions (*CCL5-CCR5* and *CCL3/CCL4-CCR5*) were particularly prominent in ND TNBC (Figure 3H), suggesting that robust CD8$^+$ T cell recruitment in ND TNBC occurs via these signals, rather than being solely dependent on *CXCL9/CXCL10*. Ultimately, these findings propose that neutrophils, as early responders, might initiate immune interactions through *CCL3* that prime the subsequent *CXCL9*-driven and *CCL5*-mediated autoreactive T cell infiltration. Therefore, the presence of CXCL9+ Macs appears to mirror a broader, intricately linked network of chemokine-mediated immune responses that foster a more evolved and effective adaptive immune environment.

To validate the prognostic significance of CXCL9+ Mac, we performed CIBERSORTx[23] analysis on bulk RNA-seq data from 34 D TNBCs and 14 ND TNBCs, using SMI data as a reference. This analysis revealed that CXCL9+ Macs and CD8$^+$ T cells were significantly enriched in the ND group whereas SPP1 + Macs did not show an association with BC prognosis ($p = 0.000335$, $p = 0.0419$, and $p = 0.593$, respectively) (Figures 4A and S1I). We also evaluated the association between proportion of CD8$^+$ T cells and both CXCL9+ and SPP1+ Macs. A strong positive correlation was observed between the proportions of CD8$^+$ T cells and CXCL9+ Macs (R = 0.46, $p = 0.00092$), while there was no significant correlation between CD8$^+$ T cells and SPP1+ Macs (R = −0.075, $p = 0.610$) (Figure 4B).

Survival analyses based on the proportions of CXCL9+ and SPP1+ Macs suggested that an increased proportion of CXCL9+ Macs was positively correlated with longer DFS (R = 0.46, $p = 0.00091$) and OS (R = 0.43, $p = 0.0025$) (Figure 4C). However, the proportion of SPP1+ Macs did not have a significant effect on DFS or OS in TNBC (R = 0.14, $p = 0.360$ for DFS, and R = 0.063, $p = 0.670$ for OS) (Figure 4D). Kaplan-Meier analysis based on the median proportions of CXCL9+ and SPP1+ Macs showed that a higher proportion of CXCL9+ Macs was significantly associated with improved outcomes in residual TNBC after NAC ($p = 0.011$ for DFS and $p = 0.023$ for


A
Myoepithelial(0.98)
Cancer Basal SC(0.97)
CAFs myCAF-like(0.99)
Macrophages(0.99)
Cancer Her2 SC(0.97)
NK cells(0.94)
DCs(0.98)
B-cells(0.98)
Cancer LumB SC(0.91)
CAFs MSC iCAF-like(0.97)
Cancer LumA SC(0.95)
Plasmablasts(1)
Cancer(0.99)
Normal Epithelial(0.97)
Monocytes(0.97)
T cells CD4+(0.97)
T cells CD8+(0.96)
PVL(0.99)
Endothelial(0.99)
Total cell counts: 110,629 cells
B
Average Expression
pct.exp
C
UMAP1
UMAP2
Monocytes
Neutrophils
cDCs
pDCs
Mac_CD163+
Mac_CD68+
Mac_IFI+
Mac_SPP1+
D
Average Expression
pct.exp
E
M1
100%
50%
Phagocytosis
M2
Angiogenesis
Mac_CD163+
Mac_IFI+
Mac_CD68+
Mac_SPP1+
F
log2-fold change
Dismal
Non-dismal
G
log2-fold change
Dismal
Non-dismal
H
Cancer LumA SC
Cancer LumB SC
Cancer Her2 SC
Cancer Basal SC
Cancer Unassigned
score
I
Correlation
J
log2 fold change
C0
C1
C2
C3
C4
C5
C6
C7
K
Neighborhood enrichment
Tumor niches
Basal-like
HER2-E
Luminal
TME niches
C1
C5
C6
L
M
N
group
Dismal
Non-dismal
score

OS) (Figure 4E). Multivariable analyses for DFS and OS further confirmed that CXCL9+ Mac was one of the most important prognostic factors, along with RCB class. CXCL9+ Mac showed a hazard ratio of 0.41 (95% confidence interval [CI]: 0.20–0.82, $p = 0.012$) for DFS and a hazard ratio of 0.46 (95% CI: 0.23–0.90, $p = 0.024$) for OS (Figure 4F).

Our initial analysis, which identified the prognostic significance of CXCL9+ Macs, was conducted primarily in post-treatment samples. To verify if these Mac signatures also hold prognostic relevance prior to treatment, we analyzed baseline data from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) project.[24] Among the total of 2,509 BCD patients, we specifically examined 309 patients with TNBC. The baseline analysis confirmed that higher CXCL9+ Mac signatures were significantly associated with better OS ($p < 0.0001$, Figure S2A) and relapse-free survival (RFS) ($p = 0.00036$, Figure S2B) consistent with our findings in post-treatment samples. Using a Cox regression model with CXCL9+ Mac signatures as a continuous variable, the hazard ratio for OS was 0.76 (95% CI: 0.59–0.98, $p = 0.0364$, Figure S2C), and, for RFS, the hazard ratio was 0.74 (95% CI: 0.55–0.99, $p = 0.0409$, Figure S2D). Conversely, higher SPP1+ Mac signatures were not associated with poor outcomes.

No genomic variations distinguish dismal and non-dismal TNBCs in somatic mutational profiles

In further somatic mutational analysis, there was no significant difference between D and ND TNBCs (Figure S2E). All TNBCs harbored *TP53* somatic alterations, regardless of their survival outcomes. Additionally, further analyses, including single base substitution (SBS) signatures,[25] tumor mutation burden, homologous recombination deficiency (HRD) score, and structural variants, did not reveal any significant differences between D and ND TNBCs (Figures S2F–S2I).

Interactions between basal-like cancer cells and limited immune cells drive prognosis in non-TNBC

We further conducted spatial transcriptomic analysis on residual non-TNBC samples using the 6,000-plex RNA panel (Table S3). We evaluated 121 FOVs from 16 BC samples, including six in the ND group and ten in the D group. The ND group included two HER2+ and four HR+HER2− BCs, while the D group consisted of eight HER2+ BCs and two HR+HER2− BCs.

Insitutype[17] with supervision categorized spatial single-cell data into 18 cell types including four intrinsic subtypes across 109,629 cells. (Figures 5A, 5B, and S3A). Louvain clustering further divided the myeloid cells into eight subtypes, including four types of Macs (Figures 5C and 5D). Among the four Mac subtypes, CXCL9+ Mac was not identified. Instead, the Macs observed included $CD68^+$ Mac, CD163+ Mac, IFI+ Mac, and SPP1+ Mac, each with distinct functional signatures. CD163+ Mac was associated with phagocytosis, and IFI+ Mac exhibited an M1 phenotype without *CXCL9/CXCL10* expression, while SPP1+ Mac showed an M2 phenotype and pro-angiogenic features (Figure 5E). To confirm the absence of CXCL9+ Mac clusters in non-TNBC, we analyzed the number of cells expressing at least one count of *CXCL9* or *CXC10*. The results indicated that these genes were rarely detected in this subset (Figure S3B). Similarly, $CD8^+$ T cells—which are typically linked to CXCL9+ Mac cells in TNBC—were scarcely observed in residual non-TNBC samples following NAC (Figure S3C).

After cell type annotation, we examined which cell types were associated with the outcomes of residual BCs after NAC. Among all identified cell types, basal-like cells were significantly more prevalent in residual BCs with a dismal prognosis (adjusted $p < 0.00001$) (Figures 5F and S3D) Additionally, within the TME cell population, we observed an enrichment of SPP1+ Macs and Tregs in the D group (Figure 5G). Similar to findings in TNBC, hypoxia signature in cancer cells was crucial to recruit SPP1+ Mac (Figure S3E), while the other three Mac subtypes were related to IFN-α expression rather than IFN-γ (Figure S3F).

Subsequent analysis of immune cell-cancer cell spatial interactions revealed significantly higher co-localization between immune cells, including $CD8^+$ T cells, and basal-like cells compared to other subtypes (Figures 5H and S3G). To determine the reason for this enhanced co-localization, we employed the SCENIC[26] analysis, which showed that basal-like cancer cells were uniquely driven by inflammation-related transcription factors such as nuclear factor κB1 and STAT1 (Figure S3H). Moreover, we explored the molecular pathways in cancer cells that are spatially associated with immune cell recruitment and discovered that IFN signatures correlate with this recruitment (Figure 5I). These signatures were elevated in basal-like cancer cells relative to other cancer cell types (Figure S3I).

To further dissect TME heterogeneity by cancer subtype, the unsupervised CellCharter algorithm identified eight distinct

Figure 5. Spatial profiling reveals distinct cell type distributions and immune-tumor interactions in residual non-TNBC post-NAC based on patient outcomes

(A) Flightpath plot of InSituType cell identification showing posterior probability.

(B) Dotplot of signature marker expression by cell type.

(C) UMAP of reclustered myeloid cells.

(D) Dotplot of signature markers for myeloid cells.

(E) Radar plot of normalized M1/M2 polarization, angiogenesis, and phagocytosis scores across macrophage subtypes.

(F and G) Bar plots of $\log_2$ fold changes in cell abundance between ND and D groups: all cell types (F) and TME (G), adjusted for intrinsic subtype. $^*p < 0.05$.

(H) Heatmap of average normalized co-localization scores between cancer subtypes (rows) and TME cells (columns) using 50 μm radius.

(I) Heatmap of Pearson correlations between cancer gene signatures (rows) and TME cells (columns) based on minimum distance to cancer cells. Only adjusted $p < 0.05$ shown; $|R| > 0.1$ annotated.

(J) Dotplots of cell type enrichment by niche.

(K) Differential neighborhood enrichment analysis between tumor-enriched (source) and TME-enriched (target) clusters in ND vs. D non-TNBC.

(L and M) Representative FOVs showing a C0-dominant region in D non-TNBC (L) and a region with minimal C0 in ND cases (M).

(N) Heatmaps comparing co-localization scores between cancer subtypes and TME cells in ND vs. D groups (t tests). $^*p < 0.05$; $^{**}p < 0.01$.

spatial niches (Figures 5J and S4A). Among these, five niches were predominantly cancer cell enriched: a basal-like dominant niche (C0), two HER2-dominant niches (C3 and C4), a Luminal A/B niche (C2), and a niche enriched in Luminal A cells (C7). Interestingly, the basal-like dominant niche (C0) was less often adjacent to immune-rich regions (C1 and C5) containing diverse myeloid and adaptive immune cells (Figures 5K and 5L) compared to other cancer niches with lower basal-like dominance (Figure 5M). Instead, C0 tended to neighbor areas that lack adaptive immune cells, composed of myoepithelium, Macs, myCAFs, and endothelial cells.

Supporting these spatial findings, pathway analysis revealed that cancer cells in the C0 niche exhibit significant activation of EMT, hypoxia, proliferation, transforming growth factor β, and Wnt/β-catenin signaling pathways, all of which could contribute to an immunosuppressive TME (Figure S4B), even though IFN signatures remained high in these clusters. In addition, intracellular communication analysis demonstrated that cancer cells in both the C0 niche and its nearby C6 niche communicate with Macs and myofibroblastic cancer-associated fibroblasts (myCAFs) through collagen, amyloid-beta precursor protein, macrophage migration inhibitory factor, and fibronectin 1 signaling, which are known to promote profibrotic or anti-inflammatory environments that suppress adaptive immunity (Figure S4C).

Consistently, the C0 niche was predominantly enriched in the D group (Figure S4D), with most basal-like cancer cells residing in this region (C0) rather than in niches with close immune cell contact (Figure S4E). In contrast, cancer cells in the ND group, although classified as basal like, did not form spatial clusters resembling the C0 niche. Thus, despite their inherent immunogenicity, basal-like cancer cells in dismal cases tend to form densely clustered regions that are less surrounded by immune cells. This is further supported by prognosis-based co-localization analysis within 50 μm, which demonstrated that basal-like cancer cells in the ND group were more closely associated with immune cells (e.g., $CD8^+$ T cells and cDCs) (Figure 5N). A similar pattern was observed when assessing immune contacts considering immunohistochemical tumor subtypes, with ND cases showing closer proximity to immune cells compared to dismal cases (Figures S4F and S4G). Overall, these findings indicate that the intrinsic subtype of BC is a key prognostic factor even at the cellular level, and, within the same subtype, the TME varies significantly depending on prognosis—particularly in basal-like cancer cells, where such variations further influence patient outcomes.

For validation, we performed another CIBERSORTx analysis on bulk RNA-seq data for non-TNBCs, using non-TNBC SMI data as a reference. The results were consistent with the SMI data, demonstrating basal-like cancer cells and SPP1+ Mac enrichment in the D group with limited presence of $CD8^+$ T cells (Figure S5A). However, subgroup analysis revealed that these findings were more pronounced in HR+HER2− patients (Figures S5B and S5C).

Furthermore, survival analysis in the METABRIC cohort (non-TNBC, N = 1,671) confirmed these findings. Specifically, SPP1 + Mac enrichment was significantly associated with worse OS (hazard ratio: 1.20, 95% CI: 1.09–1.32, $p < 0.0001$, Figure S5D) and RFS (hazard ratio: 1.12, 95% CI: 1.12–1.26, $p < 0.0001$, Figure S5E). In addition, basal-like cancer cell enrichment correlated significantly with poorer RFS (hazard ratio: 1.12, 95% CI: 1.12–1.26, $p < 0.0001$, Figure S5E) but did not reach statistical significance regarding OS (hazard ratio: 0.88, 95% CI: 0.77–1.01, $p = 0.0724$, Figure S5D).

Basal-like subtype, associated with HRD, is the foremost factor in determining dismal prognosis in residual HR+HER2− BC post-NAC

Intrinsic subtyping with bulk RNA-seq of HR+HER2− BCs showed that the basal-like subtype was observed in ten of twelve BCs in the D group whereas the ND group consisted of luminal A and B types, not basal-like subtype (Figure 6A). According to genome and exome sequencing, the frequency of *TP53* pathogenic variants (PVs) differed based on prognosis. In the D group, *TP53* PVs were detected in all BCs except one, which harbored *AKT1* PV; furthermore, all samples were classified as basal-like harbored *TP53* mutations (eleven of twelve, $p < 0.01$) (Figure 6B). However, there was no difference in the *TP53* PV loci between the ND and D groups.

Mutational signature analysis[26] found that SBS3 was associated with the D group, whereas SBS1 was associated with the ND group (Figure 6C). Additional analyses for HRD suggested that the HRD score was significantly higher in the D group compared to the ND group ($p = 0.023$) (Figure 6D). Among the components of HRD scores, the loss of heterozygosity (LOH) score was strongly associated with the D group rather than the ND group ($p = 0.0018$). Both the SBS3 and HRD scores were strongly associated with the basal-like intrinsic subtype ($p = 0.00524$ and $p = 0.00172$, respectively) (Figure 6E). In addition, SBS3 and HRD scores were positively correlated in the basal-like type (R = 0.74, $p = 0.014$) but not in the non-basal type (R = 0.29, $p = 0.33$) (Figure 6F).

In survival analysis, the SBS3 signature, high HRD score, and basal-like subtype were associated with shorter DFS and OS compared to other mutational signatures, low HRD scores, and non-basal subtypes ($p < 0.05$) (Figures 6G, 6H, and 6I). In multivariable analysis, only basal-like subtype was strongly associated with poor prognosis (hazard ratio: 105.24, 95% CI: 4.86–2277; $p = 0.003$) (Figure 6J).

Extreme structural variations and genetic alterations of HER2 as key survival strategy in residual HER2+ BC after NAC

All dismal HER2+ BCs harbored *TP53* PVs, and 50% of them also carried *PIK3CA* PVs. Interestingly, *ERBB2* single-nucleotide variants (SNVs) and insertions or deletions were only observed in the D group (Figure 7A). Although only one SNV of *ERBB2* was located in tyrosine kinase domain, SIFT predicted it to be a deleterious alteration (Figures 7B and S6A). Other findings, such as SBS and HRD scores, did not differ between the D and ND groups in HER2+ BCs (Figure 7C).

When analyzing focal oncogene amplification, we found that oncogenic interchromosomal amplicons were significantly more frequent in the D group (ten of twelve, 83.3%) compared to the ND group (none of six) ($p = 0.0015$) (Figures 7D and S6B). Moreover, interchromosomal amplicons co-carrying

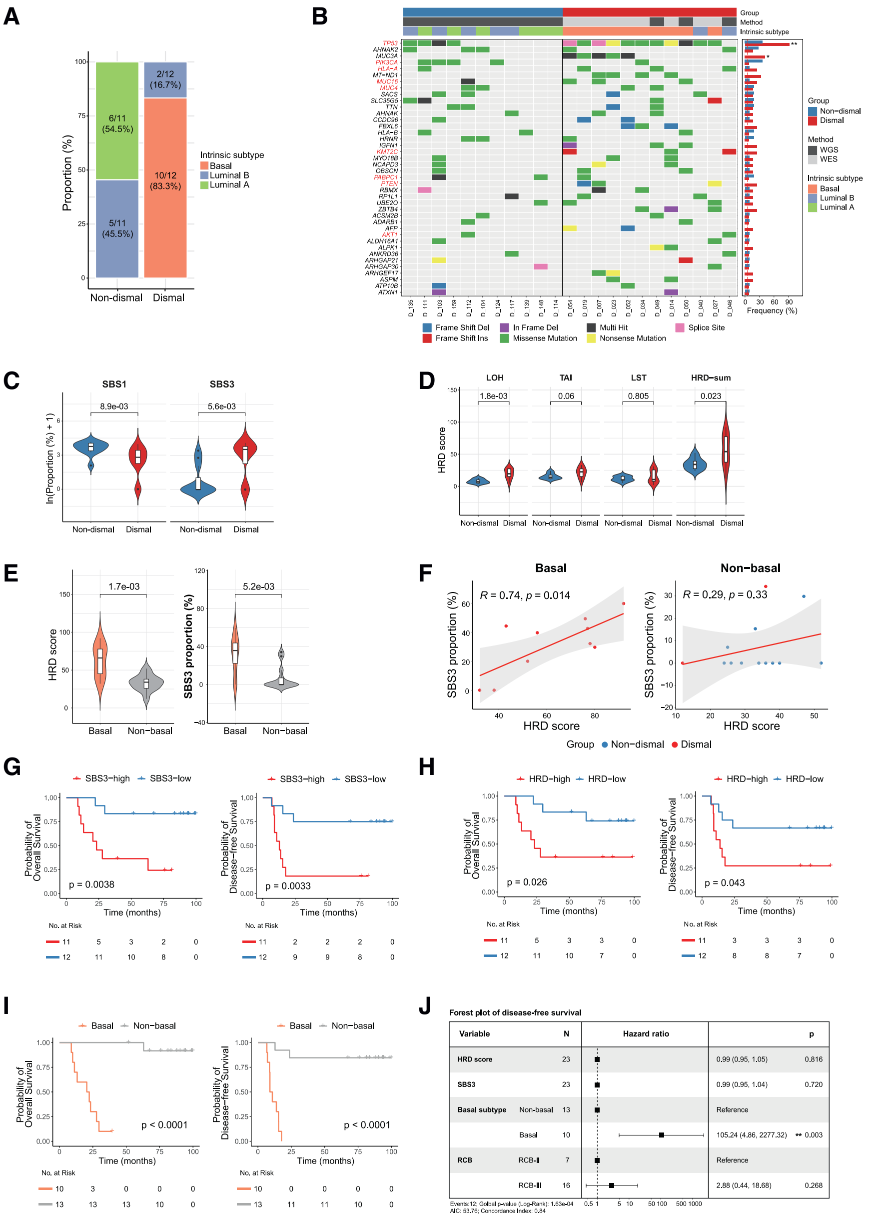


Figure 6. Basal-like subtype and homologous recombination deficiency as determinants of dismal prognosis in residual HR+HER2− BC post-NAC

(A) Intrinsic subtype prediction in HR+HER2− bulk RNA-seq dataset.

(B) Somatic mutation profiles of HR+HER2− BCs by prognosis (Fisher's test). COSMIC Cancer Gene Census v.90 genes are highlighted in red.

(C and D) Violin plots of SBS1 and SBS3 proportions (C) and HRD scores (D) by prognosis.

(E) HRD score and SBS3 proportion comparison between basal and non-basal intrinsic subtypes.

(F) Correlation between SBS3 proportion and HRD score in basal and non-basal subtypes. *R*: Pearson's correlation coefficient; solid line: fitted linear regression with 95% confidence intervals.

(G and H) Kaplan-Meier survival curves stratified by SBS3 proportion (G) and HRD score (H) using median cutoff.

(I) Kaplan-Meier survival curves for basal vs. non-basal intrinsic subtypes.

(J) Forest plot of Cox multivariable regression for disease-free survival. Hazard ratios are shown as rectangles with 95% confidence intervals. $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

For all the violin plots, central lines represent the median; boxes indicate the interquartile range.

ERBB2, *CDK12*, and *C17orf37* were observed in eight of twelve D HER2+ BCs (75.0%) ($p = 0.013$; Fisher's exact test). These amplicon characteristics were summarized in Table S4. These events were closely associated with chromothripsis regardless of their fate, amplicon classification, or other amplicon characteristics (Figure S6C).

Although we did not observe significant differences in extrachromosomal DNA (ecDNA) frequency (Figures 7E and S6D) and median number of chromosomes in an amplicon, as well as the number of ecDNA instances (Figure 7F), we found that the size of all the amplicons or ecDNA only was significantly larger in the D group compared to ND BCs ($p = 0.024$ and $p = 0.034$, respectively) (Figure 7G). Additionally, the number of oncogenes contained within an amplicon was significantly higher in the D BCs than in ND BCs ($p = 0.016$) (Figure 7H). There were no differences in the average copy number of amplicons or the number of intervals within an amplicon between the two groups, however (Figures 7I and 7J).

DISCUSSION

Our study demonstrates that, in residual TNBC after NAC, an inflamed TME driven by CXCL9+ Macs working with $CD8^+$ T cells near cancer cells critically influences outcomes. This interaction was mediated by IFN-γ and other chemokines. In non-TNBC, while immune cells were scarce and less significant than in TNBC, they still contributed to improved survival. Additionally, tumor-intrinsic factors like basal-like features, *TP53* mutations, and *ERBB2* alterations proved critical in non-TNBC outcomes.

Tumor-infiltrating lymphocytes (TILs) as well as $CD8^+$ T cells are known predictive biomarkers for BC prognosis.[27–29] However, TILs' prognostic impact varies by subtype, with stronger influence in TNBC and HER2+ BC than in HR+HER2− disease.[29] Another research indicates that $CD8^+$ T cells specifically affect basal-like BC prognosis rather than luminal or HER2+ subtypes.[30]

In our spatial analysis of residual BC after NAC in TNBC, we evaluated $CD8^+$ T cells and other immune cells. Our study focused on residual BC that did not achieve pCR after NAC, where the

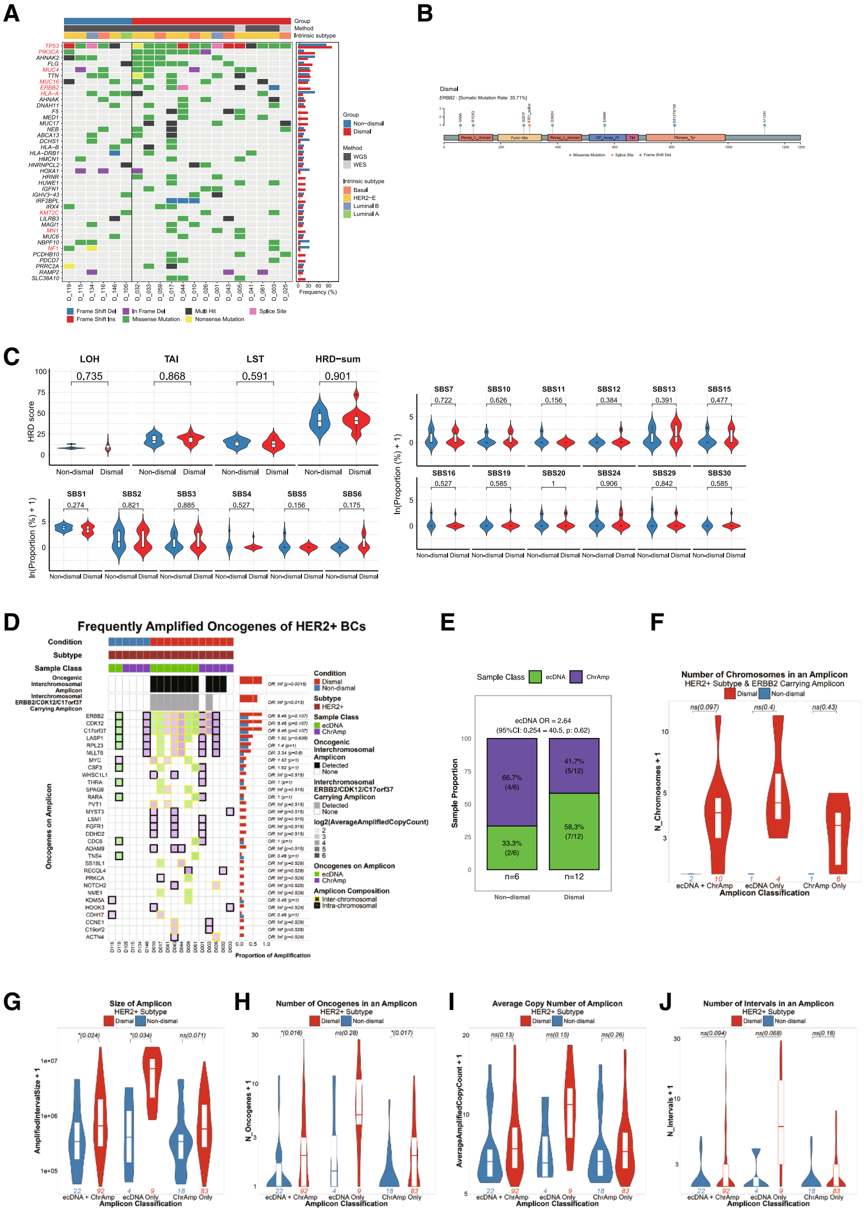


Figure 7. Comparative analysis of genomic and amplicon characteristics in HER2+ BC based on patient outcomes

(A) Somatic mutation profiles of HER2+ BCs by prognosis. COSMIC Cancer Gene Census v.90 genes are highlighted in red.

(B) Lollipop plot showing *ERBB2* somatic mutation loci in D HER2+ group.

(C) Violin plots of HRD score and mutational signature proportion differences between D and ND HER2+ groups.

(D) Distribution of oncogene-carrying focal amplifications with characteristics (interchromosomal, intrachromosomal, and copy counts) in HER2+ patients ($N = 18$) grouped by prognosis and ecDNA presence.

(E) Distribution of ecDNA-bearing patient classification in HER2+ subtype between D and ND conditions.

(F–J) Comparison of amplicon features between D and ND conditions: (F) chromosome counts, (G) amplicon size, (H) oncogene counts, (I) amplicon copy number, and (J) interval counts. ecDNA + ChrAmp, all amplicons; ecDNA Only, ecDNA amplicons; ChrAmp Only, ChrAmp amplicons.

p value is from two-sided Wilcoxon rank-sum test. For all the violin plots, central lines represent the median; boxes indicate the interquartile range.

tumor niche contained only a few CD8$^+$ T cells. Despite this limited presence, these CD8$^+$ T cells actively interacted with CXCL9+ Macs through chemokines and IFN-γ signaling pathway. This positive feedback loop ultimately contributed to improved survival outcomes particularly in TNBC (Figure S7).

CXCL9+ Macs, closely associated with IFN-γ pathway, had a major role in recruiting CD8$^+$ T cells in TNBC data. While previous research suggested that cooperation between tumor-derived *CCL5* and CXCL9+ antigen-presenting cells enables T cell infiltration and immune checkpoint inhibitor (ICI) responsiveness,[31] our findings suggest a more complex picture. Instead of tumor-derived *CCL5*, we identified a coordinated chemokine network where neutrophil-derived *CCL3* and T cell-derived *CCL5* facilitate initial recruitment, followed by *CXCL9*-driven amplification and activation of CD8$^+$ T cells. This highlights the complex, cooperative nature of chemokine signaling pathways in shaping the immune microenvironment in TNBC.

We also demonstrated that hypoxia signals from tumor cells recruited SPP1+Macs, which were more prevalent in the D group compared to the ND group. Moreover, SPP1+ Macs were negatively correlated with CXCL9+ Macs as previous study.[32] Additionally, SPP1+ Macs have been associated with cancer progression in colon cancer, hepatocellular carcinoma, and glioma.[33–35] Our findings are in line with these studies but suggest that SPP1+ Mac alone does not determine poor prognosis across all BC subtypes. Instead, they contribute by reducing the proportion of CXCL9+ Macs, which leads to worse outcomes.

In non-TNBC, CXCL9+ Macs were absent with minimal CD8$^+$ T cells, indicating underdeveloped immune responses. Basal-like cancer cells recruited more immune cells but were associated with poorer prognosis. However, greater immune infiltration in basal-like cells correlated with favorable outcomes. Previous studies have shown that luminal BC has fewer CD8$^+$ T cells and lower PD-L1 expression.[36],[37] In addition, TILs do not have a prognostic role in HR+HER2− BC.[29] However, clinical trials of ICIs have suggested that ICIs improve pCR rates, particularly in HR+HER2− BC with a low estrogen receptor (ER) score and basal-like intrinsic subtype.[38,39] Our study suggests that basal-like cancer cells interact closely with immune cells, potentially facilitating ICI efficacy with NAC.

In HR+HER2− BC, genomic profiling revealed that *TP53* PVs, SBS3, and high HRD scores were associated with the basal-like subtype and a dismal prognosis. These findings were consistent with those from previous studies, which have already demonstrated an association between genomic features and the immune landscape, as well as their impact on BC prognosis.[40] Our findings suggest that, in HR+HER2− tumors with basal-like features after NAC, PARP inhibitors and/or immunotherapy may warrant further investigation to reduce relapse risk.

Previous studies showed TILs predict pCR after NAC with HER2 blockade and DFS in adjuvant settings for HER2+ BC.[29,41,42] However, in the neoadjuvant setting, the impact of TILs on DFS diminished after adjusting for pCR,[41] suggesting that treatment response matters more than TIL presence. In fact, we rarely observed CD8$^+$ T cells in HER2+BC and instead revealed that structural alterations of HER2 itself were pivotal in shaping BC prognosis.

ERBB2 amplification is a hallmark of HER2+ BC. In this study, *ERBB2* mutations and ecDNA containing *ERBB2* were frequently observed in dismal HER2+ BC. *ERBB2* mutations, found in 3%–4% of HER2+ BCs, are considered a mechanism of resistance to anti-HER2 antibodies.[43,44] EcDNA represents the ultimate structural variants of oncogenes, and, in our data, ecDNA containing *ERBB2* was found in up to 80% of dismal HER2+ BC in our cohort. EcDNA has been known to be associated with oncogene amplification and worse prognosis.[45,46] Notably, large interchromosomal ecDNA including *CDK12*, *ERBB2*, and *C17orf37* was frequently observed in dismal BC. A recent study on esophageal cancer also suggested that tumors with ecDNA had inferior survival outcomes when treated with anti-HER2 therapy.[47]

RCB classification has been the strongest surrogate marker for predicting post-NAC survival. However, our study goes beyond this traditional approach by elucidating critical cellular and genomic determinants influencing survival outcomes, especially in residual tumors classified as high RCB. In TNBC, we uncovered a pivotal role for CXCL9+ Macs collaborating with CD8$^+$ T cells through a dynamic, chemokine-driven and IFN-γ-mediated interaction, highlighting previously unexplored therapeutic avenues to enhance immune activation. Conversely, in non-TNBC, intrinsic tumor factors—including basal-like phenotypes, *TP53* mutations, and *ERBB2* genomic alterations—emerged as primary prognostic determinants, surpassing the influence of immune cell infiltration alone. Integrating comprehensive genomic profiling and immune landscape analyses with traditional RCB classification may offer a more precise, personalized prognostic framework for guiding future BC treatment decisions.

Limitations of the study

Our study has several limitations. We only analyzed post-NAC residual tumors, preventing distinction between pre-existing and therapy-induced alterations. The small sample size, particularly in the D non-TNBC subgroup, also reduced statistical power. The study employed SMI in focused regions of interest; thus, intratumoral heterogeneity beyond those sampled regions may be underrepresented. Moreover, the single time point of sampling does not capture temporal shifts in tumor-immune interactions, making it challenging to assess evolving clonal dynamics or immune responses over the full course of treatment. Finally, while we provide key insights into Mac subtypes and spatial interactions, functional validation is needed to strengthen the findings.

RESOURCE AVAILABILITY

Lead contact

Requests for further information, resources, and reagents should be directed to and will be fulfilled by the lead contact, Ji-Yeon Kim (jyeon25@skku.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The CosMx data generated in this study have been deposited in the Zenodo database (https://doi.org/10.5281/zenodo.14039854). The raw sequencing data generated in this study have been deposited in NCBI SRA at PRJNA1256162.
- All scripts used for spatial transcriptomic analysis are archived in a GitHub repository at https://github.com/SGIlabes/CosMx-BRCA-, and a persistent copy of this repository is available via Zenodo (https://doi.org/10.5281/zenodo.15301355).[48] Other analyses were conducted using standard publicly available software.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank NanoString Technologies for their resources and expertise. J.-Y.K. is supported by the 2023 AZ-KHIDI Oncology Research Program and the National Research Foundation of Korea (NRF-2020R1F1A1072616) funded by the Ministry of Science and ICT (MSIT). She also receives support from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI, grant HR20C0025) and Samsung Medical Center (SMO1220551). H.K. is funded by NRF grants NRF-2022M3C1A3092022 and NRF-2023R1A2C2005023 supported by the Korean government. W.-Y. P. receives funding from the Bio & Medical Technology Development Program (NRF-2017M3A9A7050803) and the Korea Health Technology R&D Project (grant RS-2020-KH088686) through KHIDI. S.L. is supported by NRF grants funded by the Ministry of Education (NRF-2018R1A6A1A03025810) and MSIT (NRF-2017M3A9A7050612, NRF-2020M3A9B6038849, NRF-2020M3A9D8038192, NRF-2021R1A2C1094009, RS-2023-00225255, and RS-2023-00261820).

AUTHOR CONTRIBUTIONS

J.-Y.K. conceived the project and directed the study with input from all authors. J.-Y.K., S.L., H.K., and W.-Y.P. designed the study. J.E.L., S.W.K., Y. H.P., and Y.-H.I. constructed the patient cohort, provided samples, and offered intellectual input and guidance. E.Y.C., J.E.L., H.H.J., and J.H. contributed to the experimental design, procured the patient tissues, and assisted with data interpretation. E.S.S. performed the spatial data analysis. S.P. analyzed the RNA-seq and whole-genome sequencing (WGS) data. S.A. and H.K. conducted the amplicon analysis. S.W.K., J.S., and J.S.A. analyzed the clinical information. J.-Y.K., E.S.S., S.P., E.Y.C., J.E.L., H.K., S.L., and W.-Y.P. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

W.-Y.P. is the founder and chief executive officer of Geninus Inc., and E.S.S. is an employee of Geninus Inc.

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
    - Human specimens
- METHOD DETAILS
    - Sequencing experiment
- WHOLE GENOME/EXOME SEQUENCING
- WHOLE TRANSCRIPTOME SEQUENCING
- WGS/WES ANALYSIS
    - Sequence alignment
    - Somatic point mutation
    - Mutational signature
    - Homologous recombination deficiency (HRD)
    - Structural variation
- WTS ANALYSIS
    - Sequence alignment and quantification
    - Intrinsic subtype prediction
    - Cell type Estimation
    - Survival analysis
- CosMX SPATIAL TRANSCRIPTOMICS
    - Tissue preparation
    - Spatial gene expression profiling
    - Cell type annotations
    - Milo and differential abundance test
    - Spatial distribution analysis of myeloid cells
    - Cell-cell Co-localization analysis
    - Cell-cell interaction analysis
    - Transcription factor (TF) activity analysis
    - Scoring of gene signatures
    - Comparing cellular niches
    - Spatial autocorrelation
    - Differential Cell type abundances
- METABRIC COHORT
- WGS COPY NUMBER ANALYSIS
- AMPLICON ANALYSIS
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2025.102164.

Received: January 3, 2025
Revised: April 2, 2025
Accepted: May 14, 2025
Published: June 6, 2025

REFERENCES

1. Loibl, S., André, F., Bachelot, T., Barrios, C.H., Bergh, J., Burstein, H.J., Cardoso, M.J., Carey, L.A., Dawood, S., Del Mastro, L., et al. (2024). Early breast cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann. Oncol. *35*, 159–182. https://doi.org/10.1016/j.annonc.2023.11.016.
2. Cortazar, P., Zhang, L., Untch, M., Mehta, K., Costantino, J.P., Wolmark, N., Bonnefoi, H., Cameron, D., Gianni, L., Valagussa, P., et al. (2014). Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet *384*, 164–172. https://doi.org/10.1016/S0140-6736(13)62422-8.
3. Symmans, W.F., Wei, C., Gould, R., Yu, X., Zhang, Y., Liu, M., Walls, A., Bousamra, A., Ramineni, M., Sinn, B., et al. (2017). Long-Term Prognostic Risk After Neoadjuvant Chemotherapy Associated With Residual Cancer Burden and Breast Cancer Subtype. J. Clin. Oncol. *35*, 1049–1060. https://doi.org/10.1200/JCO.2015.63.1010.
4. Kim, J.Y., Oh, J.M., Lee, S.K., Yu, J., Lee, J.E., Kim, S.W., Nam, S.J., Park, Y.H., Ahn, J.S., Kim, K., and Im, Y.H. (2022). Improved Prediction of Survival Outcomes Using Residual Cancer Burden in Combination With Ki-67 in Breast Cancer Patients Underwent Neoadjuvant Chemotherapy. Front. Oncol. *12*, 903372. https://doi.org/10.3389/fonc.2022.903372.
5. Yau, C., Osdoit, M., van der Noordaa, M., Shad, S., Wei, J., de Croze, D., Hamy, A.S., Laé, M., Reyal, F., Sonke, G.S., et al. (2022). Residual cancer burden after neoadjuvant chemotherapy and long-term survival outcomes in breast cancer: a multicentre pooled analysis of 5161 patients. Lancet Oncol. *23*, 149–160. https://doi.org/10.1016/S1470-2045(21)00589-1.
6. Tutt, A.N.J., Garber, J.E., Kaufman, B., Viale, G., Fumagalli, D., Rastogi, P., Gelber, R.D., de Azambuja, E., Fielding, A., Balmaña, J., et al. (2021). Adjuvant Olaparib for Patients with BRCA1- or BRCA2-Mutated Breast Cancer. N. Engl. J. Med. *384*, 2394–2405. https://doi.org/10.1056/NEJMoa2105215.
7. Masuda, N., Lee, S.J., Ohtani, S., Im, Y.H., Lee, E.S., Yokota, I., Kuroi, K., Im, S.A., Park, B.W., Kim, S.B., et al. (2017). Adjuvant Capecitabine for Breast Cancer after Preoperative Chemotherapy. N. Engl. J. Med. *376*, 2147–2159. https://doi.org/10.1056/NEJMoa1612645.
8. von Minckwitz, G., Huang, C.S., Mano, M.S., Loibl, S., Mamounas, E.P., Untch, M., Wolmark, N., Rastogi, P., Schneeweiss, A., Redondo, A., et al. (2019). Trastuzumab Emtansine for Residual Invasive HER2-Positive Breast Cancer. N. Engl. J. Med. *380*, 617–628. https://doi.org/10.1056/NEJMoa1814017.
9. Johnston, S.R.D., Harbeck, N., Hegg, R., Toi, M., Martin, M., Shao, Z.M., Zhang, Q.Y., Martinez Rodriguez, J.L., Campone, M., Hamilton, E., et al. (2020). Abemaciclib Combined With Endocrine Therapy for the Adjuvant Treatment of HR+, HER2-Node-Positive, High-Risk, Early Breast Cancer (monarchE). J. Clin. Oncol. *38*, 3987–3998. https://doi.org/10.1200/JCO.20.02514.
10. Sparano, J.A., Gray, R.J., Makower, D.F., Pritchard, K.I., Albain, K.S., Hayes, D.F., Geyer, C.E., Jr., Dees, E.C., Goetz, M.P., Olson, J.A., Jr., et al. (2018). Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. N. Engl. J. Med. *379*, 111–121. https://doi.org/10.1056/NEJMoa1804710.
11. Sparano, J.A., Gray, R.J., Ravdin, P.M., Makower, D.F., Pritchard, K.I., Albain, K.S., Hayes, D.F., Geyer, C.E., Jr., Dees, E.C., Goetz, M.P., et al. (2019). Clinical and Genomic Risk to Guide the Use of Adjuvant Therapy for Breast Cancer. N. Engl. J. Med. *380*, 2395–2405. https://doi.org/10.1056/NEJMoa1904819.
12. Prat, A., Guarneri, V., Pascual, T., Brasó-Maristany, F., Sanfeliu, E., Paré, L., Schettini, F., Martínez, D., Jares, P., Griguolo, G., et al. (2022). Development and validation of the new HER2DX assay for predicting pathological response and survival outcome in early-stage HER2-positive breast cancer. EBioMedicine *75*, 103801. https://doi.org/10.1016/j.ebiom.2021.103801.
13. Danaher, P., Zhao, E., Yang, Z., Ross, D., Gregory, M., Reitz, Z., Kim, T.K., Baxter, S., Jackson, S., He, S., et al. (2022). Insitutype: likelihood-based cell typing for single cell spatial transcriptomics. Preprint at bioRxiv. https://doi.org/10.1101/2022.10.19.512902.
14. 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. https://doi.org/10.1038/s41588-021-00911-1.
15. Dann, E., Henderson, N.C., Teichmann, S.A., Morgan, M.D., and Marioni, J.C. (2022). Differential abundance testing on single-cell data using k-nearest neighbor graphs. Nat. Biotechnol. *40*, 245–253. https://doi.org/10.1038/s41587-021-01033-z.
16. Feng, Y., Yang, T., Zhu, J., Li, M., Doyle, M., Ozcoban, V., Bass, G.T., Pizzolla, A., Cain, L., Weng, S., et al. (2023). Spatial analysis with SPIAT and spaSim to characterize and simulate tissue microenvironments. Nat. Commun. *14*, 2697. https://doi.org/10.1038/s41467-023-37822-0.

17. Gavish, A., Tyler, M., Greenwald, A.C., Hoefflin, R., Simkin, D., Tschernichovsky, R., Galili Darnell, N., Somech, E., Barbolin, C., Antman, T., et al. (2023). Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. Nature *618*, 598–606. https://doi.org/10.1038/s41586-023-06130-4.
18. Varrone, M., Tavernari, D., Santamaria-Martínez, A., Walsh, L.A., and Ciriello, G. (2024). CellCharter reveals spatial cell niches associated with tissue remodeling and cell plasticity. Nat. Genet. *56*, 74–84. https://doi.org/10.1038/s41588-023-01588-4.
19. Jin, S., Plikus, M.V., and Nie, Q. (2025). CellChat for systematic analysis of cell–cell communication from single-cell transcriptomics. Nat. Protoc. *20*, 180–219. https://doi.org/10.1038/s41596-024-01045-4.
20. Cheng, M., Liang, G., Yin, Z., Lin, X., Sun, Q., and Liu, Y. (2023). Immunosuppressive role of SPP1-CD44 in the tumor microenvironment of intrahepatic cholangiocarcinoma assessed by single-cell RNA sequencing. J. Cancer Res. Clin. Oncol. *149*, 5497–5512. https://doi.org/10.1007/s00432-022-04498-w.
21. Liu, F., Zhang, J., Gu, X., Guo, Q., and Guo, W. (2024). Single-cell transcriptome sequencing reveals SPP1-CD44-mediated macrophage–tumor cell interactions drive chemoresistance in TNBC. J. Cell Mol. Med. *28*, e18525. https://doi.org/10.1111/jcmm.18525.
22. Soupir, A.C., Hayes, M.T., Peak, T.C., Ospina, O., Chakiryan, N.H., Berglund, A.E., Stewart, P.A., Nguyen, J., Segura, C.M., Francis, N.L., et al. (2023). Increased spatial coupling of integrin and collagen IV in the immunoresistant clear cell renal cell carcinoma tumor microenvironment. Preprint at bioRxiv. https://doi.org/10.1101/2023.11.16.567457.
23. 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. https://doi.org/10.1038/s41587-019-0114-2.
24. 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. https://doi.org/10.1038/nature10983.
25. Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. Nature *578*, 94–101. https://doi.org/10.1038/s41586-020-1943-3.
26. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.
27. Mahmoud, S.M.A., Paish, E.C., Powe, D.G., Macmillan, R.D., Grainge, M.J., Lee, A.H.S., Ellis, I.O., and Green, A.R. (2011). Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. J. Clin. Oncol. *29*, 1949–1955. https://doi.org/10.1200/JCO.2010.30.5037.
28. Denkert, C., Loibl, S., Noske, A., Roller, M., Müller, B.M., Komor, M., Budczies, J., Darb-Esfahani, S., Kronenwett, R., Hanusch, C., et al. (2010). Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. J. Clin. Oncol. *28*, 105–113. https://doi.org/10.1200/JCO.2009.23.7370.
29. Denkert, C., von Minckwitz, G., Darb-Esfahani, S., Lederer, B., Heppner, B.I., Weber, K.E., Budczies, J., Huober, J., Klauschen, F., Furlanetto, J., et al. (2018). Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. Lancet Oncol. *19*, 40–50. https://doi.org/10.1016/S1470-2045(17)30904-X.
30. Liu, S., Lachapelle, J., Leung, S., Gao, D., Foulkes, W.D., and Nielsen, T.O. (2012). CD8+ lymphocyte infiltration is an independent favorable prognostic indicator in basal-like breast cancer. Breast Cancer Res. *14*, R48. https://doi.org/10.1186/bcr3148.
31. Dangaj, D., Bruand, M., Grimm, A.J., Ronet, C., Barras, D., Duttagupta, P.A., Lanitis, E., Duraiswamy, J., Tanyi, J.L., Benencia, F., et al. (2019). Cooperation between Constitutive and Inducible Chemokines Enables T Cell Engraftment and Immune Attack in Solid Tumors. Cancer Cell *35*, 885–900. https://doi.org/10.1016/j.ccell.2019.05.004.
32. Bill, R., Wirapati, P., Messemaker, M., Roh, W., Zitti, B., Duval, F., Kiss, M., Park, J.C., Saal, T.M., Hoelzl, J., et al. (2023). CXCL9:SPP1 macrophage polarity identifies a network of cellular programs that control human cancers. Science *381*, 515–524. https://doi.org/10.1126/science.ade2292.
33. Qi, J., Sun, H., Zhang, Y., Wang, Z., Xun, Z., Li, Z., Ding, X., Bao, R., Hong, L., Jia, W., et al. (2022). Single-cell and spatial analysis reveal interaction of FAP(+) fibroblasts and SPP1(+) macrophages in colorectal cancer. Nat. Commun. *13*, 1742. https://doi.org/10.1038/s41467-022-29366-6.
34. Fan, G., Xie, T., Li, L., Tang, L., Han, X., and Shi, Y. (2024). Single-cell and spatial analyses revealed the co-location of cancer stem cells and SPP1+ macrophage in hypoxic region that determines the poor prognosis in hepatocellular carcinoma. npj Precis. Oncol. *8*, 75. https://doi.org/10.1038/s41698-024-00564-3.
35. Tang, W., Lo, C.W.S., Ma, W., Chu, A.T.W., Tong, A.H.Y., and Chung, B.H.Y. (2024). Revealing the role of SPP1(+) macrophages in glioma prognosis and therapeutic targeting by investigating tumor-associated macrophage landscape in grade 2 and 3 gliomas. Cell Biosci. *14*, 37. https://doi.org/10.1186/s13578-024-01218-4.
36. Stanton, S.E., Adams, S., and Disis, M.L. (2016). Variation in the Incidence and Magnitude of Tumor-Infiltrating Lymphocytes in Breast Cancer Subtypes: A Systematic Review. JAMA Oncol. *2*, 1354–1360. https://doi.org/10.1001/jamaoncol.2016.1061.
37. Sobral-Leite, M., Van de Vijver, K., Michaut, M., van der Linden, R., Hooijer, G.K.J., Horlings, H.M., Severson, T.M., Mulligan, A.M., Weerasooriya, N., Sanders, J., et al. (2018). Assessment of PD-L1 expression across breast cancer molecular subtypes, in relation to mutation rate, BRCA1-like status, tumor-infiltrating immune cells and survival. OncoImmunology *7*, e1509820. https://doi.org/10.1080/2162402X.2018.1509820.
38. Sherene Loi, G.G., Salgado, R., Díaz, R.I.R., Delaloge, S., Rojas, C., and Kok, M. (2023). LBA20 A randomized, double-blind trial of nivolumab (NIVO) vs placebo (PBO) with neoadjuvant chemotherapy (NACT) followed by adjuvant endocrine therapy (ET) ± NIVO in patients (pts) with high-risk, ER+ HER2− primary breast cancer (BC). 2023 ESMO. Ann. Oncol. *34*, S1259–S1260.
39. Fatima Cardoso, H.M., Schmid, P., Cortes, J., Harbeck, N., Telli, M.L., and Cescon, D. (2023). LBA21 KEYNOTE-756: Phase III study of neoadjuvant pembrolizumab (pembro) or placebo (pbo) + chemotherapy (chemo), followed by adjuvant pembro or pbo + endocrine therapy (ET) for early-stage high-risk ER+/HER2– breast cancer. 2023 ESMO. Ann. Oncol. *34*, S1260–S1261.
40. Ballot, E., Galland, L., Mananet, H., Boidot, R., Arnould, L., Desmoulins, I., Mayeur, D., Kaderbhai, C., Ilie, S., Hennequin, A., et al. (2022). Molecular intrinsic subtypes, genomic, and immune landscapes of BRCA-proficient but HRD-high ER-positive/HER2-negative early breast cancers. Breast Cancer Res. *24*, 80. https://doi.org/10.1186/s13058-022-01572-6.
41. Liefaard, M.C., van der Voort, A., van Seijen, M., Thijssen, B., Sanders, J., Vonk, S., Mittempergher, L., Bhaskaran, R., de Munck, L., van Leeuwen-Stok, A.E., et al. (2024). Tumor-infiltrating lymphocytes in HER2-positive breast cancer treated with neoadjuvant chemotherapy and dual HER2-blockade. NPJ Breast Cancer *10*, 29. https://doi.org/10.1038/s41523-024-00636-4.
42. Dieci, M.V., Conte, P., Bisagni, G., Brandes, A.A., Frassoldati, A., Cavanna, L., Musolino, A., Giotta, F., Rimanti, A., Garrone, O., et al. (2019). Association of tumor-infiltrating lymphocytes with distant disease-free survival in the ShortHER randomized adjuvant trial for patients with early HER2+ breast cancer. Ann. Oncol. *30*, 418–423. https://doi.org/10.1093/annonc/mdz007.

43. Subramanian, J., Katta, A., Masood, A., Vudem, D.R., and Kancha, R.K. (2019). Emergence of ERBB2 Mutation as a Biomarker and an Actionable Target in Solid Cancers. Oncologist *24*, e1303–e1314. https://doi.org/10.1634/theoncologist.2018-0845.

44. Sun, Z., Shi, Y., Shen, Y., Cao, L., Zhang, W., and Guan, X. (2015). Analysis of different HER-2 mutations in breast cancer progression and drug resistance. J. Cell Mol. Med. *19*, 2691–2701. https://doi.org/10.1111/jcmm.12662.

45. Kim, H., Nguyen, N.P., Turner, K., Wu, S., Gujar, A.D., Luebeck, J., Liu, J., Deshpande, V., Rajkumar, U., Namburi, S., et al. (2020). Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. Nat. Genet. *52*, 891–897. https://doi.org/10.1038/s41588-020-0678-2.

46. Morton, A.R., Dogan-Artun, N., Faber, Z.J., MacLeod, G., Bartels, C.F., Piazza, M.S., Allan, K.C., Mack, S.C., Wang, X., Gimple, R.C., et al. (2019). Functional Enhancers Shape Extrachromosomal Oncogene Amplifications. Cell *179*, 1330–1341. https://doi.org/10.1016/j.cell.2019.10.039.

47. Tsai, C., Suman, S., Chu, C., Chou, J.F., Krein, P.M., Capanu, M., Ilson, D. H., Ku, G.Y., Janjigian, Y.Y., Drilon, A.E., et al. (2024). Association of ERBB2 extrachromosomal DNA (ecDNA) with first-line HER2-targeted treatment outcomes in advanced esophagogastric adenocarcinoma (EGC). J. Clin. Oncol. *42*, 4041. https://doi.org/10.1200/JCO.2024.42.16_suppl.4041.

48. Seo, E.S. (2025). SGIlabes/CosMx-BRCA-: CosMx Analysis (Zenodo).

49. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

50. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

51. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B.S., and Swanton, C. (2016). DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. *17*, 31. https://doi.org/10.1186/s13059-016-0893-4.

52. Sztupinszki, Z., Diossy, M., Krzystanek, M., Reiniger, L., Csabai, I., Favero, F., Birkbak, N.J., Eklund, A.C., Syed, A., and Szallasi, Z. (2018). Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. NPJ Breast Cancer *4*, 16. https://doi.org/10.1038/s41523-018-0066-6.

53. Cameron, D.L., Schröder, J., Penington, J.S., Do, H., Molania, R., Dobrovic, A., Speed, T.P., and Papenfuss, A.T. (2017). GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. Genome Res. *27*, 2050–2060. https://doi.org/10.1101/gr.222109.117.

54. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

55. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. *12*, 323. https://doi.org/10.1186/1471-2105-12-323.

56. Gendoo, D.M.A., 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://doi.org/10.1093/bioinformatics/btv693.

57. He, S., Bhatt, R., Brown, C., Brown, E.A., Buhr, D.L., Chantranuvatana, K., Danaher, P., Dunaway, D., Garrison, R.G., Geiss, G., et al. (2022). High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. Nat. Biotechnol. *40*, 1794–1806. https://doi.org/10.1038/s41587-022-01483-z.

58. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587. https://doi.org/10.1016/j.cell.2021.04.048.

59. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.

60. Asaeedi, S., Didehvar, F., and Mohades, A. (2017). α-Concave hull, a generalization of convex hull. Theor. Comput. Sci. *702*, 48–59. https://doi.org/10.1016/j.tcs.2017.08.014.

61. Green, P.J., and Silverman, B.W. (1979). Constructing the convex hull of a set of points in the plane. Comput. J. *22*, 262–266. https://doi.org/10.1093/comjnl/22.3.262.

62. Cheng, S., Li, Z., Gao, R., Xing, B., Gao, Y., Yang, Y., Qin, S., Zhang, L., Ouyang, H., Du, P., et al. (2021). A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. Cell *184*, 792–809. https://doi.org/10.1016/j.cell.2021.01.010.

63. Bivand, R. (2022). R Packages for Analyzing Spatial Data: A Comparative Case Study with Areal Data. Geogr. Anal. *54*, 488–518. https://doi.org/10.1111/gean.12319.

64. Chen, Y., Chen, L., Lun, A.T.L., Baldoni, P.L., and Smyth, G.K. (2024). edgeR 4.0: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. Preprint at bioRxiv. https://doi.org/10.1101/2024.01.21.576131.

65. Priestley, P., Baber, J., Lolkema, M.P., Steeghs, N., de Bruijn, E., Shale, C., Duyvesteyn, K., Haidari, S., van Hoeck, A., Onstenk, W., et al. (2019). Pan-cancer whole-genome analyses of metastatic solid tumours. Nature *575*, 210–216. https://doi.org/10.1038/s41586-019-1689-y.

66. Cameron, D.L., Baber, J., Shale, C., Valle-Inclan, J.E., Besselink, N., van Hoeck, A., Janssen, R., Cuppen, E., Priestley, P., and Papenfuss, A.T. (2021). GRIDSS2: comprehensive characterisation of somatic structural variation using single breakend variants and structural variant phasing. Genome Biol. *22*, 202. https://doi.org/10.1186/s13059-021-02423-x.

67. Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput. Biol. *12*, e1004873. https://doi.org/10.1371/journal.pcbi.1004873.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
ER Antibody	Immunotech, Marseille, France	Cat# IM1344; RRID: AB_131617
PgR Antibody	Novocastra Laboratories Ltd., Newcastle upon Tyne, UK	Cat# NCL-L-PGR-312; RRID:AB_563967
HER2 Antibody	Dako, Glostrop, Denmark	Cat# SK001; RRID:AB_2935822
Biological samples		
Human breast cancer samples	Samsung Medical Center	N/A
Critical commercial assays		
QIAamp DNA Mini Kit	Qiagen	Cat# 51304
QIAamp DNA Blood Mini Kit	Qiagen	Cat# 51106
RNeasy Mini Kit	Qiagen	Cat# 74104
NovaSeq 6000 platform	Illumina	Cat# 20012866
CosMx SMI	NanoString	N/A
Deposited data		
Raw and processed CosMx SMI data of human breast cancer patients	This paper	https://doi.org/10.5281/zenodo.14039854
METABRIC	Curtis et al., (2012)	https://www.cbioportal.org
Software and algorithms		
BioRender	BioRender	https://biorender.com/
BWA	Li et al., (2009)	https://github.com/lh3/bwa
Genome Analysis Toolkit (GATK)	McKenna et al., (2009).	https://gatk.broadinstitute.org/hc/en-us
Mutect2	Cibulskis et al., (2013)	https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2
scarHRD	Sztupinszki et al., (2018)	https://github.com/sztup/scarHRD
deconstructSigs	Rosenthal et al., (2016)	https://github.com/raerose01/deconstructSigs
GRIDSS	Cameron et al., (2017)	https://github.com/PapenfussLab/gridss
STAR	Dobin et al., (2013)	https://github.com/alexdobin/STAR
CIBERSORTx	Newman et al., (2019)	https://cibersortx.stanford.edu/
genefu	Gendoo et al., (2015)	https://www.bioconductor.org/packages/release/bioc/html/genefu.html
forestmodel	Kennedy et al., (2020)	https://github.com/NikNakk/forestmodel
Seurat	Stuart et al., (2019)	https://satijalab.org/seurat/
InSituType	Danaher et al., (2022)	https://github.com/Nanostring-Biostats/InSituType
Harmony	Korsunsky et el., (2019)	https://github.com/immunogenomics/harmony
Milo	Dann et al., (2022)	https://github.com/MarioniLab/miloR
SPIAT	Feng el al., (2023)	https://trigosteam.github.io/SPIAT/
SCENIC	Aibar et al., (2017)	https://github.com/aertslab/SCENIC
CellCharter	Varrone et al., (2024)	https://github.com/CSOgroup/cellcharter
spdep	Binvant et al., (2018)	https://github.com/r-spatial/spdep
edgeR	Robinson et al., (2010)	https://bioconductor.org/packages/release/bioc/html/edgeR.html
CellChat	Jin et al., (2024)	https://github.com/jinworks/CellChat
hmftools	Hartwig Medical Foundation	https://github.com/hartwigmedical/hmftools
AmpliconSuite	Luebeck et al., (2024)	https://github.com/AmpliconSuite/AmpliconSuite-pipeline

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human specimens

Using the breast cancer registry in the Cancer Clinical Data Warehouse (CDW) of Samsung Medical Center, we identified data from EBCs that had received NAC followed by curative surgery between 2009 and 2020. Patient characteristics including age, sex, and other clinically relevant factors are summarized in Table S1. From this EBC dataset, we collected BCs that did not achieve a pCR and experienced distant recurrence within two years after curative surgery. This group was labeled the D group. In addition, we collected data from BCs that did not achieve a pCR but did not experience any recurrence, forming the control group (ND group). For these BCs, fresh frozen (FF) tumor specimens and blood samples at curative surgery were provided by the BioBank at Samsung Medical Center. This study was approved by institutional review board (IRB) (IRB number: 2021-02-101) and was conducted in accordance with the tenets of the Declaration of Helsinki and the Guidelines for Good Clinical Practice. All participants gave written informed consent. All available hematoxylin and eosin-stained slides for FF tissues were collected. Pathology specimens were reviewed by independent pathologists to determine tumor histology and immunohistochemical (IHC) findings (estrogen receptor [ER] and progesterone receptor [PgR] expression and HER2 overexpression). ER and PgR positivity, where the presence of either was defined as HR positivity, was determined by Allred scores ranging from 3 to 8, based on IHC with antibodies to ER (Immunotech, Marseille, France) and PgR (Novocastra Laboratories Ltd., Newcastle upon Tyne, UK). HER2 status was evaluated using a specific antibody (Dako, Glostrop, Denmark), fluorescent *in situ* hybridization (FISH), or silver *in situ* hybridization (SISH). A HER2 IHC score of 0 or 1 was defined as a negative result, while a score of 3 was defined as positive. HER2 amplification was confirmed by FISH or SISH if the HER2 score was 2. TNBC was defined as negative for ER, PgR and HER2. Baseline clinical characteristics, histologic characteristics, and treatment response and survival outcomes were collected. Breast magnetic resonance imaging (MRI), chest computed tomography (CT), abdomino-pelvic CT, and bone scans were performed for staging. Staging was conducted using the anatomical staging system of the American Joint Committee of Cancer (AJCC) 8th edition.

METHOD DETAILS

Sequencing experiment

Extraction of DNA and RNA

Nucleic acid extraction was performed on samples with sufficient tumor tissue. FF samples were split for RNA and DNA extraction, and cryo-sectioned tissues were stained with hematoxylin and eosin (H&E) to assess tumor content. For quality control, we excluded samples with less than 30% tumor content. Genomic DNA was extracted from FF tissues using the QIAamp DNA Mini Kit (Qiagen) from buffy coats using the QIAamp DNA Blood Mini Kit (Qiagen). Total RNA extraction from FF tissues was carried out using the RNeasy Mini Kit (Qiagen), following the manufacturer's protocol. Nucleic acid yield and purity were assessed using the NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Thermo-Fisher Scientific, MA, USA)

WHOLE GENOME/EXOME SEQUENCING

The integrity of the genomic DNA was checked by running an agarose gel electrophoresis and gDNA was quantified using Quant-IT PicoGreen (Invitrogen). The sequencing libraries were prepared according to the manufacturer's instructions of the TruSeq DNA Nano Library Prep Kit (Illumina, Inc., San Diego, CA, USA). Briefly, Fragmentation of 100 ng of genomic DNA was performed using adaptive focused acoustic technology (AFA; Covaris) and the fragmented DNA underwent end-repair to create 5′-phosphorylated, blunt-ended dsDNA molecules. Following end-repair, DNA was size selected with a bead-based method. These DNA fragments went through the addition of a single 'A' base, and ligation of the truseq indexing adapters. The purified libraries were quantified using qPCR according to the qPCR Quantification Protocol Guide (KAPA Library Quantification kits for Illumina Sequencing platforms) and qualified using the Agilent Technologies 2200 TapeStation (Agilent Technologies). We then sequenced using the Novaseq6000 (Illumina)

WHOLE TRANSCRIPTOME SEQUENCING

Total RNA concentration was calculated by Quant-IT RiboGreen (Invitrogen). To determine the DV200 (% of RNA fragments >200 bp) value, samples were run on the TapeStation RNA Screentape (Agilent). A total of 100 ng of total RNA was subjected to sequencing library construction using a TruSeq RNA Access library prep kit (Illumina, San Diego, CA, USA) according to the manufacturer's protocols. Briefly, the total RNA was first fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first strand cDNA using SuperScript II reverse transcriptase (Invitrogen, #18064014) and random primers. This was followed by second strand cDNA synthesis using DNA Polymerase I, RNase H, and dUTP. These cDNA fragments then went through an end repair process, the addition of a single 'A' base, and then ligation of the adapters. The products were then purified and enriched with PCR to create the cDNA library. All libraries were normalized and six libraries were pooled into a single hybridization/capture reaction. Pooled libraries were incubated with a cocktail of biotinylated oligos corresponding to coding regions of the genome. Targeted library molecules were captured via hybridized biotinylated oligo probes using streptavidin-conjugated beads. After two rounds of hybridization/capture reactions, the enriched library molecules were subjected to a second round of

PCR amplification. The captured libraries were quantified using KAPA Library Quantification kits for Illumina Sequencing platforms according to the qPCR Quantification Protocol Guide (KAPA BIOSYSTEMS, #KK4854) and qualified using the TapeStation D1000 ScreenTape (Agilent Technologies, #5067–5582). Indexed libraries were then sequenced using the NovaSeq 6000 platform (Illumina, San Diego, USA) by Macrogen Incorporated.

WGS/WES ANALYSIS

Sequence alignment

Using the Burrows Wheeler aligner (BWA v0.7.17),[49] paired-end reads were aligned to the GRCh38 Human reference genome. Genome Analysis Toolkit (GATK) v4.2[50] was used for the local realignment and quality score recalibration procedures.

Somatic point mutation

To facilitate more comprehensive comparative analysis between BC subtypes within the expanded sample set, we incorporated WGS as well as WES samples into our analysis. For the WGS dataset, we focused on mutations within exome regions corresponding to those targeted by WES. Somatic mutations were identified using MuTect2 with paired tumor and matched normal samples. Variants were filtered for false positives using ‘FilterMutectCalls’ tool from GATK, and annotated using VEP v.92.1.

Mutational signature

The deconstructSigs R package[51] was used to analyze single base substitution signatures. Consistent with our approach for somatic point mutation analysis, both WGS and WES datasets were utilized. For the input variant call format (VCF) files, variants present in exome regions were considered. The significance between D and ND groups was determined by Wilcoxon rank-sum test.

Homologous recombination deficiency (HRD)

To calculate HRD score, scarHRD R package[52] was used. HRD score was determined by the sum of the number of loss of heterozygosity (LOH), telomeric allelic imbalances (TAI), and large scale transitions (LST) events. The segmentation files derived from Sequenza results served as input for scarHRD, with ’grch38′ set as the reference genome. Both WGS and WES datasets were used for the analysis, focusing on events within exome regions. Differences between groups were assessed by Wilcoxon rank-sum test.

Structural variation

The genomic rearrangements were detected in WGS dataset using GRIDSS v.2.13.2.[53] For somatic filtering, ‘gridss_somatic_filter’ tool was additionally performed. Structural variants classified as high quality, defined as QUAL score $\geq$1000 and assemblies from both breakpoint sides (AS>0 & RAS>0), were included in the analysis.

WTS ANALYSIS

Sequence alignment and quantification

Sequencing reads were mapped to the GRCh38 human reference genome of GENCODE v36 using STAR v2.7.9a.[54] The expression values for Fragments Per Kilobase of exon model per Million Mapping Fragments (FPKM) were determined by RSEM v1.3.3.[55]

Intrinsic subtype prediction

The ‘molecular.subtyping’ function with ‘pam50’ subtyping classification model in the genefu[56] bioconductor package (v2.26.0) was used to identify the intrinsic subtypes of each sample.

Cell type Estimation

CIBERSORTx[23] was performed to estimate cell types of each sample. Reference cell types from TNBC and non-TNBC SMI datasets were used for deconvolution in both TNBC and non-TNBC samples. Statistical significance between groups was assessed by Wilcoxon rank-sum test.

Survival analysis

DFS and OS were evaluated using the Kaplan-Meier method to estimate survival curves. To assess the impact of covariates on survival, Cox proportional-hazards regression was applied, and hazard ratios along with 95% CIs were calculated. The ‘forest_model’ function of forestmodel R package was used to generate forest plots.

CosMx SPATIAL TRANSCRIPTOMICS

Tissue preparation

Tissue microarray (TMA) slides containing tumor tissue from our study cohort were used for CosMx. To address the distribution of immune cells in the tumor microenvironment and their association with clinical outcomes, a large core (5 mm core size) TMA was generated. Five-micron tissue sections were cut from TMA blocks using a microtome, placed in a heated water bath, and mounted on Superfrost Plus microscopy slides (VWR, Cat #: 48311-703) according to NanoString's sample preparation guide specifications. Details of spatial transcriptomic experiment were described in well-established documents.[57]

Spatial gene expression profiling

CosMx SMI profiling was conducted on 15 TNBC samples, 6 HR + HER2-samples, and 10 HER2-positive (HR±/HER2+) samples. In total, 31 formalin-fixed paraffin-embedded (FFPE) samples from 31 patients were distributed across four slides (two for TNBC and two for non-TNBC) for tissue microarray (TMA) creation with a 5 mm core diameter. These slides were sent to NanoString for profiling as part of their early Technical Access Program (TAP). We employed a protocol that allowed an average of 6.7 fields of view (FOVs) per sample, with each FOV measuring 0.51 mm × 0.51 mm. The CosMx Human Universal Cell Characterization RNA panel (1000-plex) was applied to TNBC samples across 88 FOVs, while the CosMx Human 6K Discovery Panel (6000-plex) was applied to non-TNBC samples across 121 FOVs. Sample processing, staining, imaging, and cell segmentation followed previously established methods.[57] In brief, 5 μm sections of an FFPE TMA block were cut using a microtome and placed onto VWR Superfrost Plus Micro Slides (Cat# 48311-703), adhered to slides after being placed in a heated water bath. These slides were then dried overnight at 37°C, vacuum-sealed, and stored at 4°C until they were analyzed. Manual FFPE tissue preparation, *in situ* hybridization (ISH), coverslip application, and cyclic RNA readout on the SMI followed established protocols. After all cycles were completed, additional markers for morphology and cell segmentation, including DAPI, pancytokeratin (PanCK), CD45, CD298/B2M, and/or CD3, were added. A 3D multichannel image stack was obtained at each FOV location. Registration, feature detection, localization, identification of individual transcripts, and cell segmentation were performed using established methods. The final segmentation assigned each transcript location in the registered image to the corresponding cell, including the specific cell compartment (nucleus, cytoplasm, membrane) where the transcript was found. Additional features generated included shape metrics (area, aspect ratio) and fluorescence intensity statistics (minimum, maximum, average) per cell.

Cell type annotations

We annotated cell types independently for TNBC and non-TNBC samples due to differences in their panels. For each tissue sample, we collected and processed probe counts, spatial coordinates, and cell segmentation data using Seurat v5.0.[58] We excluded cells with fewer than 20 transcript counts or unusually large areas (defined as exceeding five times the geometric mean, indicative of segmentation errors) from further analysis. For cell type identification, we employed the InSituType[13] supervised clustering method for initial phenotyping with scRNA reference data from Wu et el.[14] Cells were assigned a type if their cell type probability exceeded 80%. For cancer cell clusters, we used InSituType again to identify subclusters as defined in the reference dataset. Cancer cells that met the 80% threshold for the cancer cluster but did not meet the threshold for any subcluster were labeled as 'Cancer Unassigned'. After initial InSituType phenotyping, we focused on lymphocyte and myeloid cell clusters. We performed subcluster identification using principal component analysis (PCA) and UMAP projections after batch correction. Batch correction was achieved using the "RunHarmony()" function from the Harmony package (v1.2.0),[59] followed by UMAP using the Seurat package to generate cell embeddings representing the integrated space. Differentially expressed genes (DEGs) among these subclusters were identified using the "FindAllMarkers" function in Seurat. These DEGs were instrumental in further refining cell phenotypes, such as identifying regulatory T cells.

Milo and differential abundance test

We employed Milo[15] (v2.0) to examine the differential abundance of myeloid cells within specific neighborhoods between dismal and non-dismal prognosis groups in TNBC. Initially, we used the "buildGraph" function to create a KNN graph with k = 20, utilizing 20 principal components. Subsequently, the "makeNhoods" function was employed with parameters set to prop = 0.2. For differential abundance testing, we used the "testNhoods" function with TMM normalization and included disease outcome as a covariate in the model. To manage multiple testing, we utilized the spatial FDR approach in Milo, employing the graph-overlap weighting, with the SpatialFDR cutoff set at the 0.1 level. The spatial FDR and $\log_2$ fold change in cell numbers between conditions for each neighborhood were visualized.

Spatial distribution analysis of myeloid cells

To understand the spatial distribution of myeloid cell populations, we used SPIAT[16] (v1.6.0) to identify the tumor margin. SPIAT uses an alpha hull[60] algorithm, which generalizes the convex hull[61] by using curves instead of straight lines to better define tissue margins. The radius (α) of these curves determines their curvature, with larger α values producing less curved hulls for complex borders and smaller α values giving finer resolution for clearer borders.

In our analysis, we used the "identify_bordering_cells" function from SPIAT with the reference_cell parameter set to Cancer, Normal Epithelial, and Myoepithelial cells using a default α value. This function designates the cells on the alpha hull as the margin cells of the structure. We then calculated the distance of myeloid cells from the margin using the "calculate_distance_to_margin" function and classified cells based on their location into intratumor and stromal areas, using a distance of 1 cell from the tumor margin as the cutoff with the "define_structure" function.

Cell-cell Co-localization analysis

To evaluate the spatial relationships between two or more cell types, we utilized the NMS metric from SPIAT.[16] The mixing score measures the interactions between two different cell types by calculating the ratio of interactions between reference cells and target cells to the interactions among reference cells themselves. An interaction is defined when two cells are within a specified radius of each other. Since the mixing score can be influenced by the total number of reference and target cells, SPIAT provides the NMS to account for all possible interactions within an image, regardless of radius limitations. For our analysis, we set the interaction radius to 50 μm and calculated NMS between the cells of interest, assigning a value of 0 when no target cells were within the specified radius of the reference cells.

Cell-cell interaction analysis

We employed CellChat to infer cell-cell interactions and compared CCL pathways between non-dismal and dismal samples. The analysis was conducted separately for each prognostic group. A CellChat object was created for each group using the normalized gene expression matrix and cell type annotations. We then executed the recommended CellChat workflow, which included calculating communication probabilities while applying a distance constraint that limited ligand-receptor interactions to a maximum range of 100 μm. The trim fraction was set to 0.1, and a minimum threshold of 5 cells per group was applied to qualify for communication analysis. After completing these steps, we compared the CCL pathway networks between the two groups. To further analyze differences, the non-dismal and dismal CellChat objects were merged into a single object using the "mergeCellChat" function.

Transcription factor (TF) activity analysis

To identify TF activities characterizing specific cell types, we used SCENIC[26] (v1.3.1), which reconstructs gene regulatory networks and identifies cell states based on co-expression and DNA motifs. This analysis involved identifying the regulatory relationships between TFs and their target genes and predicting an activity score for each TF. Initially, we filtered out genes expressed in fewer than 1% of cells and selected those available in the RcisTarget database. Using the "runCorrelation" function, we created a correlation matrix and identified potential TF targets with GENIE3. Potential direct-binding targets were then identified with RcisTarget based on DNA motif analysis. The specificity of TF signatures in each cell type was determined by scoring intracellular network activity using AUCell.

Scoring of gene signatures

To identify signatures in cancer cells, genes from previously defined metaprograms (MPs)[17] were utilized. Out of a total of 41 consensus MPs, those enriched in breast cancer ($N = 18$) and interferon signatures ($N = 2$) were selected for further analysis. Signature scores were computed using the "AddModuleScore" function in Seurat. To analyze the spatial association between specific signatures in individual cancer cells and immune cells, we calculated the Euclidean distance between cancer cells and other cells. For macrophage-related signatures, gene sets were obtained from the study by Cheng et al.[62] and signature scores were also calculated using the "AddModuleScore" function.

Comparing cellular niches

To identify and compare spatial cellular niches, we utilized the CellCharter[18] algorithm, which facilitated the unsupervised identification of spatial clusters. Through stability analysis, we determined the optimal number of clusters and selected 7 clusters for further investigation. We then compared the proportion of each niche between the D and ND groups of TNBC using the Wilcox test. Additionally, we conducted a differential neighborhood enrichment analysis between the D and ND groups in TNBC, performing 1,000 permutations to ensure robustness.

Spatial autocorrelation

To evaluate the spatial proximity of CD8A and chemokines, we used bivariate Moran's I[22]. We identified the five nearest neighbors for each cell using the spdep R package (version 1.3–4).[63] For each FOV, we converted gene expression data to spatial objects, calculated the k-nearest neighbors, and created a neighbor list. The bivariate Moran's I for every gene pair was then calculated by inputting gene expression data and weights into the "moran_bv" function in spdep. This metric ranges from −1 (indicating a strong inverse spatial autocorrelation) to +1 (indicating a strong positive spatial autocorrelation), with values around 0 suggesting no spatial relationship between the gene expressions.

Differential Cell type abundances

In non-TNBC data, cell type abundances were calculated per FOV. Differential abundance analysis was performed using a generalized linear model with a quasi-likelihood approach. The design matrix for the model included factors for cancer subtype and disease outcome group. Dispersion was estimated with the "estimateDisp" function, and statistical testing was conducted using "glmQLFTest" in edgeR[64] (v4.2.0). Significant changes in cell type abundances were identified and visualized.

METABRIC COHORT

To validate and further investigate the prognostic significance of specific cell types, we used breast cancer profiling data from the METABRIC dataset available through cBioPortal. Gene expression and clinical data were downloaded. Expression data were extracted from the Illumina microarray platform, and clinical data, including OS and RFS, were matched with patient samples. Signature scores were calculated by Z score normalization of marker gene expression levels, followed by averaging across samples. Patients were stratified into high and low signature expression groups based on the optimal cut-off point, which was determined by maximizing the log rank test statistic. Kaplan–Meier survival curves were generated to compare survival outcomes, and differences between groups were assessed using log rank tests. Cox proportional hazards models were used to estimate hazard ratio and 95% CI.

WGS COPY NUMBER ANALYSIS

We used hmftools[65] to perform CNV analysis of the WGS data. (https://github.com/hartwigmedical/hmftools). PURPLE (v3.7.2) was used to determine total copy number, tumor ploidy and purity. PURPLE combines B-allele frequency (AMBER v3.9), read depth ratios (COBALT v1.13), somatic variants (SAGE v3.2.2, PAVE v1.4.1) and structural variants (GRIDSS v2.13.2,[66] GRIPSS v2.3.2). Unsupported copy number segment, which is copy number segment without structural variant support on either side, can be identified using PURPLE.

AMPLICON ANALYSIS

AmpliconArchitect (part of AmpliconSuite-pipeline, version 0.1344.2) was run using default settings. This includes BAM file downsampling to 10X coverage prior to detection of seed regions, to normalize sequencing depth between samples. Candidate seed regions for inputs to AmpliconArchitect were identified with AmpliconSuite-pipeline.py which uses CNVkit[67] for detecting DNA copy number alterations, and were defined as at least 50 kb in length and at minimum 4.5 copy number. AmpliconClassifier (v0.4.16) was invoked from AmpliconSuite-pipeline to predict the class of focal amplification and refine gene coordinate involved in the specific focal amplifications. Amplicons were classified into two categories: ecDNA (circular) and ChrAmp (linear or complex). Samples with at least one circular amplicon (ecDNA) were classified as ecDNA. Samples with amplicons but without ecDNA were classified as ChrAmp, and samples without any amplicons were classified as NoAmp. Amplicons overlapping with unsupported copy number segments were excluded.

QUANTIFICATION AND STATISTICAL ANALYSIS

The quantification methods and statistical analyses for CosMx, DNA, and RNA sequencing are detailed in the corresponding sections of the STAR Methods. Relevant statistical parameters are provided in the legend of each figure unless specified otherwise.