Supplementary Information

Spatial transcriptomics reveals substantial heterogeneity in triple negative breast cancer with potential clinical implications

Authors:

Xiaoxiao Wang[1,2], David Venet[1], Frédéric Lifrange[3], Denis Larsimont[4], Mattia Rediti[1], Linnea Stenbeck[5], Floriane Dupont[1], Ghizlane Rouas[1], Andrea Joaquin Garcia[1], Ligia Craciun[4], Laurence Buisseret[1,2], Michail Ignatiadis[1,2], Marcela Carausu[1], Nayanika Bhalla[5], Yuvarani Masarapu[5], Eva Gracia Villacampa[5], Lovisa Franzén[5], Sami Saarenpää[5], Linda Kvastad[5], Kim Thrane[5], Joakim Lundeberg[5], Françoise Rothé[1] and Christos Sotiriou[1,2]

Affiliations:

[1]Breast Cancer Translational Research Laboratory J-C Heuson, Institut Jules Bordet, Université Libre de Bruxelles, Hôpital Universitaire de Bruxelles, Brussels, Belgium

[2]Medical Oncology Department, Institut Jules Bordet, Université Libre de Bruxelles, Hôpital Universitaire de Bruxelles, Brussels, Belgium

[3]Department of Pathology, University Hospital Center of Liège, Liège, Belgium

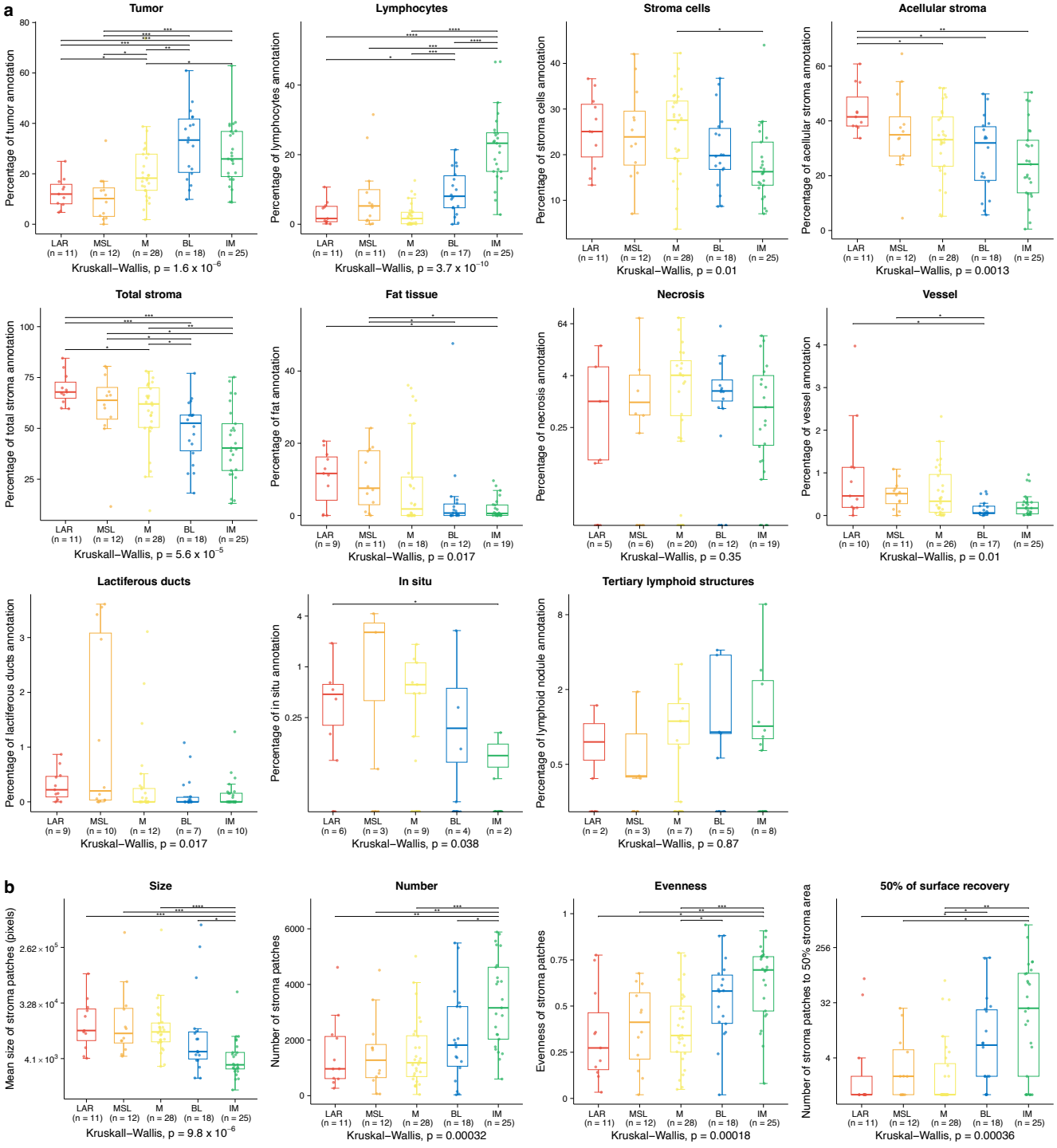
[4]Department of Pathology, Institut Jules Bordet, Université Libre de Bruxelles, Hôpital Universitaire de Bruxelles, Brussels, Belgium

[5]Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden

This file contains:
-All Supplementary Figures (1 to 33)
-All Supplementary Tables (1 to 16)

SUPPLEMENTARY FIGURES

Supplementary Figure 1



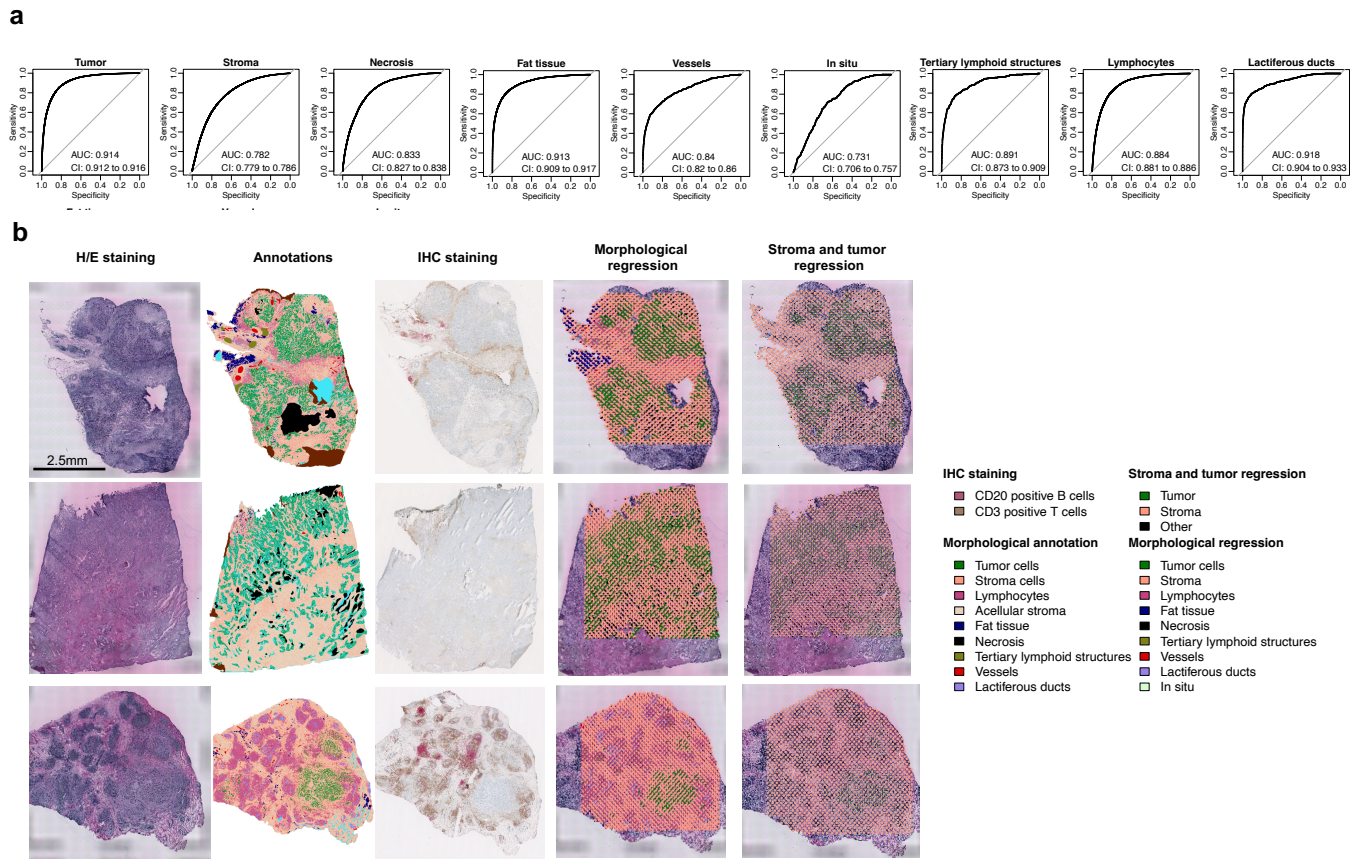
Distribution of morphological annotations (a) and characteristics of stroma patches (b) across the five TNBC molecular subtypes in the ST TNBC cohort (*N* = 94).

Statistical differences across subtypes were assessed using Kruskal-Wallis tests and Wilcoxon rank sum tests (when comparing each subtype against each one of the others). For Wilcoxon tests, FDRs were obtained by adjusting two-sided P values using Benjamini & Hochberg method. FDRs < 0.05 are shown. In the panels: *FDR < 0.05 and ≥ 0.01; **FDR < 0.01 and ≥ 0.001; ***FDR < 0.001 and ≥ 0.0001; ****FDR < 0.0001. In boxplots, the boxes are defined by the upper and lower quartile; the median is shown as a bold colored horizontal line; whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. TNBC molecular subtypes are computed on the ST global PB. The number under each TNBC molecular subtype refers to the

number of samples with the presence of a morphological annotation or stroma patches-related metrics.
Source data and exact P values are provided as a Source Data file.

BL: basal-like; IM: immunomodulatory; LAR: luminal androgen receptor; M: mesenchymal; MSL: mesenchymal stem-like.

Supplementary Figure 2



Morphological regression and deconvolution of gene expression of the nine histomorphological categories.

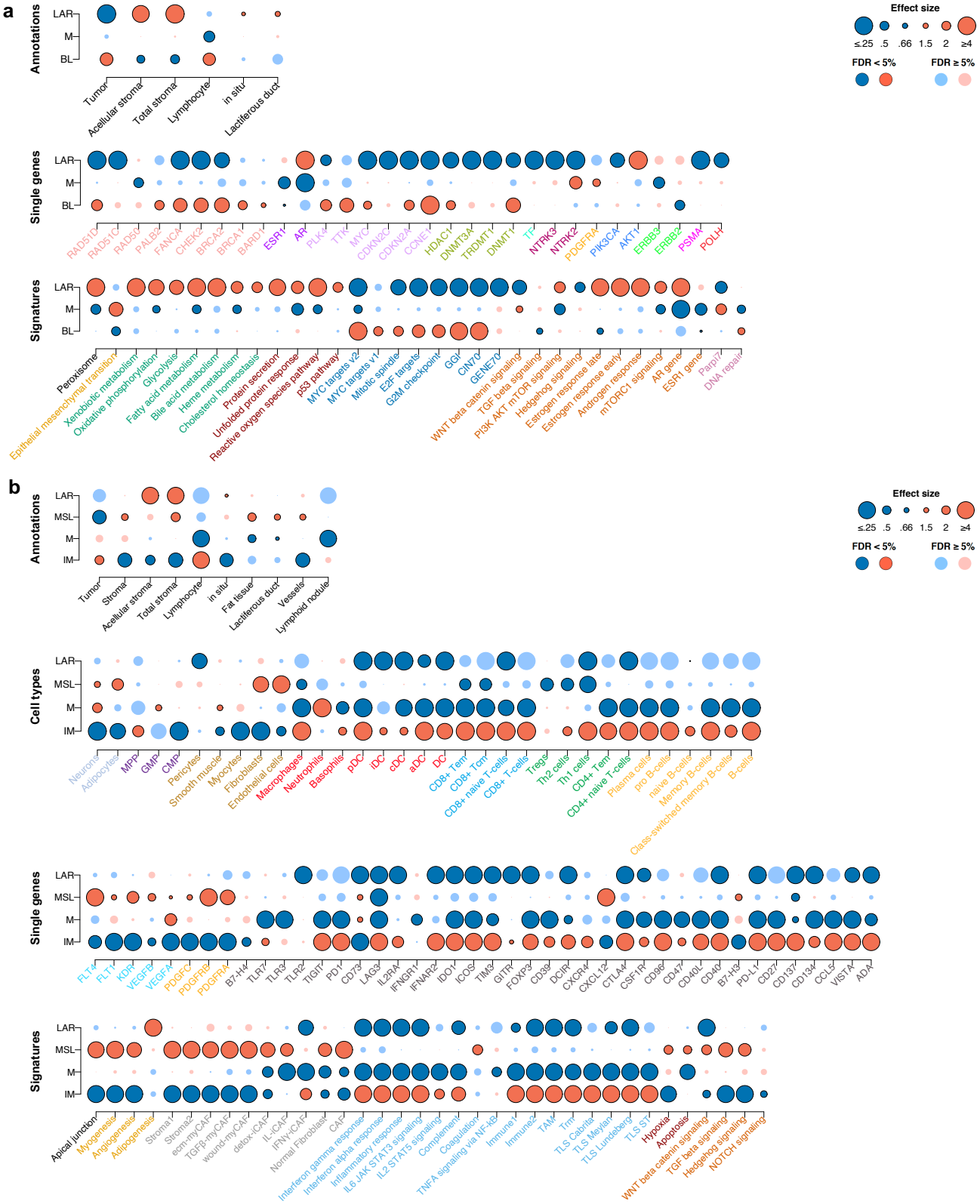
a Areas Under the Curve (AUCs) for the prediction of each histomorphological category (binarized as present in a spot if proportion > 25%) by the regressor. Confidence intervals (CI) for the AUC of each histomorphological category are reported.

b Visualization of three illustrative TNBC samples by different methods: ST-associated H/E image, morphological annotation of the ST-associated H/E image, double CD3/CD20 IHC staining, and projections of all histomorphological or stroma/tumor compartments derived from ST expression-based per-spot regression.

Source data are provided as a Source Data file.

AUC: area under the curve; H/E: hematoxylin-eosin; IHC: immunohistochemistry.

Supplementary Figure 3



Molecular and cellular characterization of stroma and tumor compartments computed on the deconvoluted pseudobulk in the ST TNBC cohort (*N* = 94).

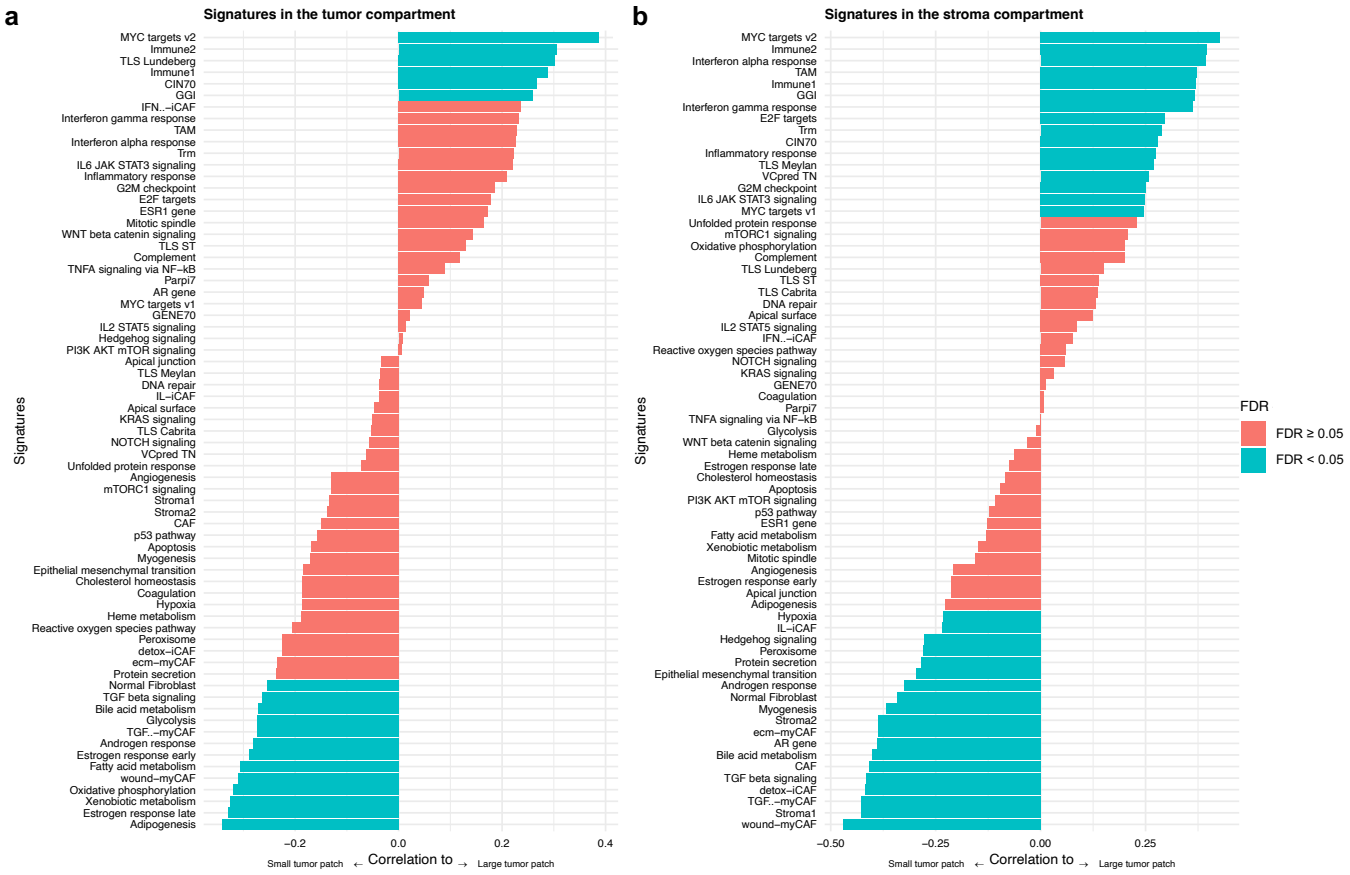
a Associations between TNBC molecular subtypes computed on tumor pseudobulk and different tumor specific features: morphological annotations, expression of single genes of interest and expression-based signatures.

b Associations between TNBC molecular subtypes computed on stroma pseudobulk and different stroma specific features: morphological annotations, cell type enrichment scores by xCell, expression of single genes of interest and expression-based signatures. A logistic regression model

was used to evaluate associations between each feature and each subtype. Two-sided P values were obtained from Wilcoxon rank sum tests and corrected for multi testing. Dots are bordered and dark-colored when FDRs < 0.05, compared to lighter-colored dots when FDRs ≥ 0.05. Negative and positive associations are represented in blue and red, respectively.
Source data are provided as a Source Data file.

aDC: activated dendritic cells; AR: androgen receptor; AUC: Areas under the curve; BL: basal-like; cDC: conventional dendritic cells; DC: dendritic cells; GMP: granulocyte-macrophage progenitor; HE: hematoxylin and eosin; iDC: immature dendritic cells; IHC: immunohistochemistry; IM: immunomodulatory; LAR: luminal androgen receptor; M: mesenchymal; MPP: multipotent progenitor; MSC: mesenchymal stem cell; MSL: mesenchymal stem-like; pDC: plasmacytoid dendritic cells; sTILs: stromal tumor infiltrating lymphocytes; Tcm: central memory T cells; Tem: effector memory T cells; Th1: T helper 1; Th2: T helper 2; Tregs: regulatory T cells.

Supplementary Figure 4



Molecular characterization of tumor and stroma compartments in correlation with the size of tumor patches in the ST TNBC cohort (*N* = 94).

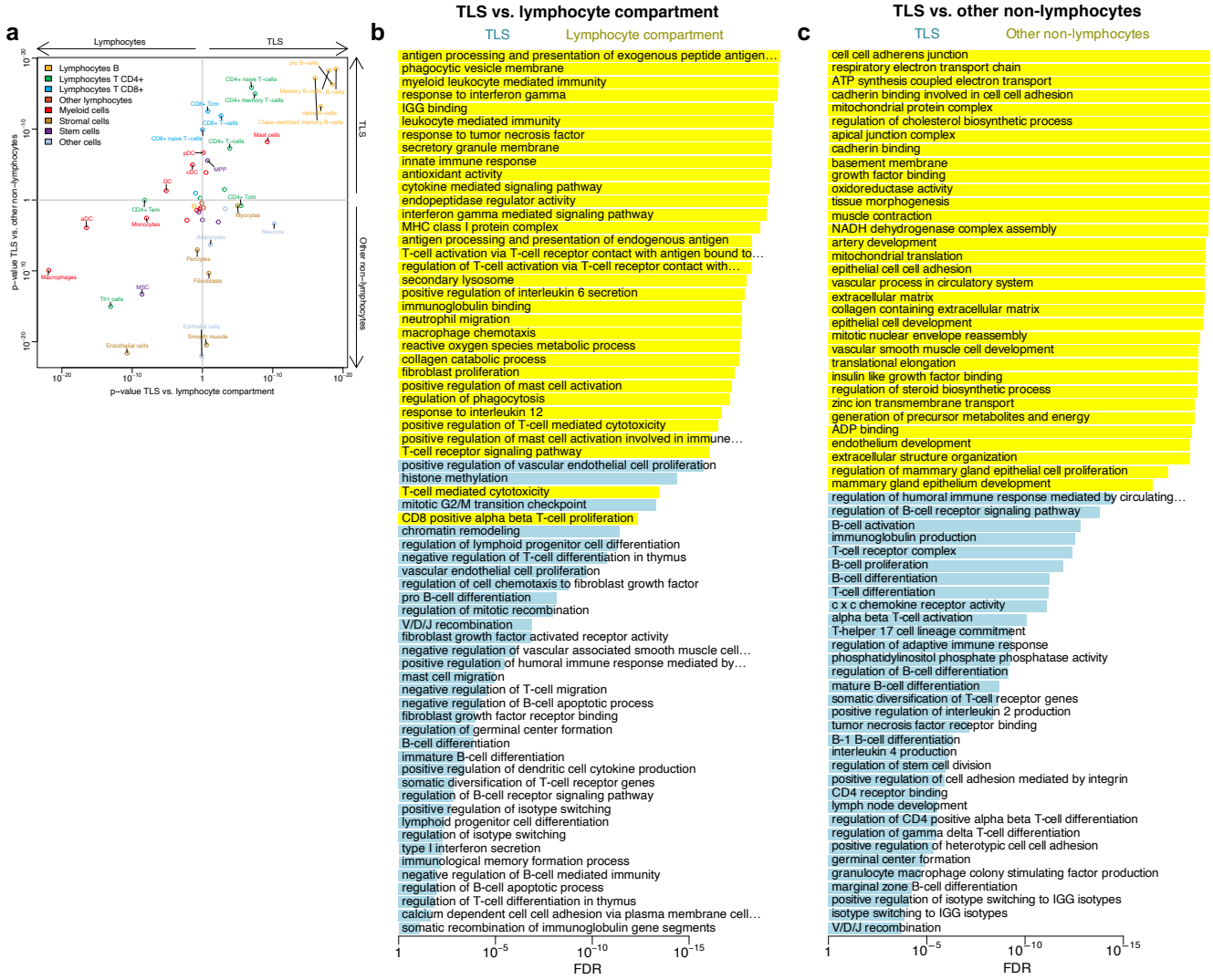
a Association of the mean size of tumor patches with different expression-based signatures computed in the tumor compartment pseudobulk.

b Association of the mean size of tumor patches with different expression-based signatures computed in the stroma compartment pseudobulk. Barplots are ranked by Spearman's Rho coefficient values and colored by the corresponding FDRs (blue: < 0.05; red: ≥ 0.05). Positive correlation indicates association to larger patches and negative correlation indicates association to smaller patches.

Source data are provided as a Source Data file.

CAF: cancer associated fibroblast S1; FDR: false-discovery rate; GGI: genomic grade index; IL2: interleukin 2; IL6: interleukin 6; Parpi7: parp inhibitor 7; TAM: tumor associated macrophages; TLS: tertiary lymphoid structure; TNFA: tumor necrosis factor alpha; Trm: tissue-resident memory T cell; VCpredTN: veliparib carboplatin prediction triple negative.

Supplementary Figure 5



Cellular and molecular characterization of tertiary lymphoid structures in the ST TNBC cohort (*N* = 94).

a Distribution of cell type enrichment scores by xCell specific to TLS compartment compared to lymphocytes (x-axis) and compared to other non-lymphocytes compartments (y-axis).

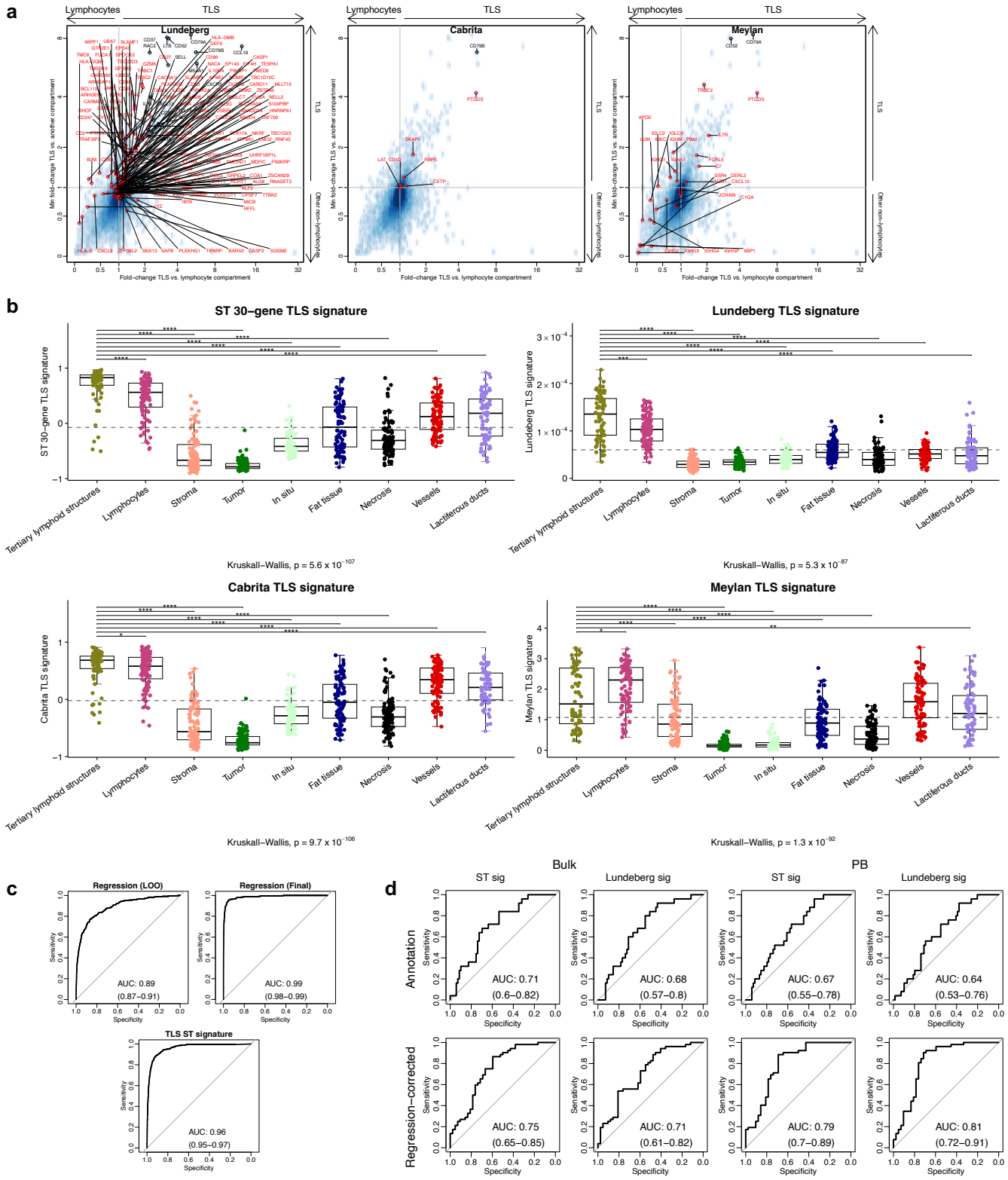
b Selected list of the most differentially expressed biological pathways assessed using GO: BP between TLS and lymphocyte compartments. Blue and yellow bars referred to processes enriched in the TLS and lymphocytes compartments, respectively.

c Selected list of the most differentially expressed biological pathways assessed using GO: BP between TLS and other non-lymphocytes compartments. Blue and yellow bars refer to processes enriched in the TLS and other non-lymphocytes compartments, respectively. Statistical differences were assessed using one-sided Wilcoxon rank sum test. All comparisons shown are significant with FDRs < 0.05.

Source data are provided as a Source Data file.

aDC: activated dendritic cells; cDC: conventional dendritic cells; DC: dendritic cells; GO:BP: Gene Ontology—Biological Processes; iDC: immature dendritic cells; MDSC: myeloid-derived suppressor cells; MPP: multipotent progenitor; MSC: mesenchymal stem cell; pDC: plasmacytoid dendritic cells; Tcm: central memory T cells; Tem: effector memory T cells; Th1: type 1 helper; Th2: T helper 2; Tregs: regulatory T cells; TLS: tertiary lymphoid structure.

Supplementary Figure 6



Morphological and molecular significance of 30-genes TLS ST signature and comparison with other TLS signatures.

a Relative expression of each gene in the TLS compartment compared to the lymphocyte compartment (x-axis) or other compartments (y-axis) in the ST TNBC cohort (*N* = 94). Each panel highlights genes from a published TLS signature[1–3] with shared genes with our signature in black and those exclusive to the referenced signature in red.

b Distribution of the expression of various TLS signatures across the nine histomorphological compartments in the ST TNBC cohort. Statistical differences across compartments were assessed using Kruskal-Wallis test and Wilcoxon rank sum test (when comparing each compartment to each

of the others). For Wilcoxon tests, FDRs were obtained by adjusting two-sided P values using Benjamini & Hochberg method (*FDR < 0.05 and ≥ 0.01; **FDR < 0.01 and ≥ 0.001; ***FDR < 0.001 and ≥ 0.0001; ****FDR < 0.0001). Only the significant FDRs for paired comparisons to TLS are shown. In boxplots, the boxes are defined by the upper and lower quartile; the median is shown as a bold black horizontal line; whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. The dashed horizontal line represents the mean TLS signature across the entire cohort.

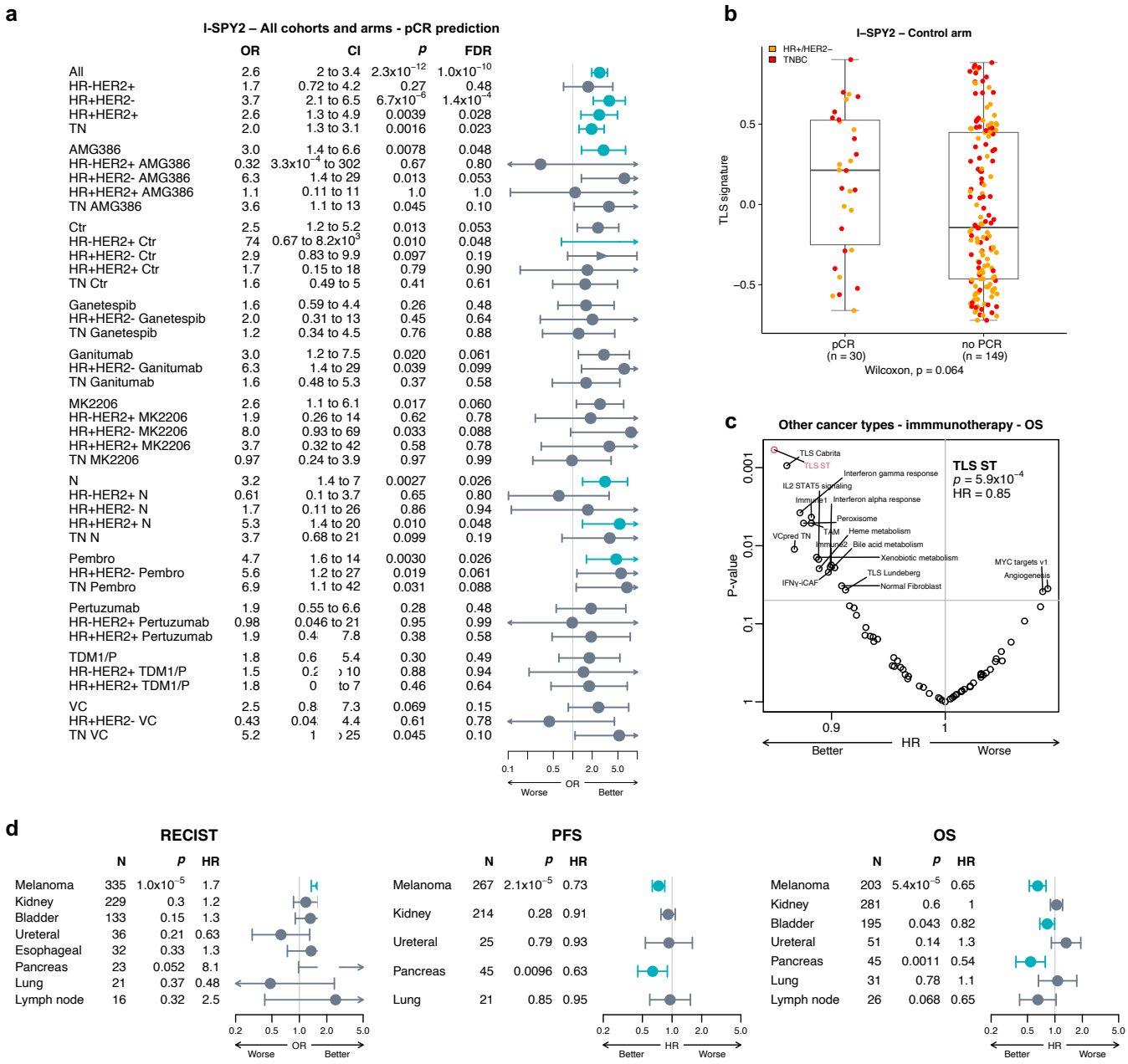
c AUCs for the prediction of the TLS, using the regressor in a leave-a-patient-out cross-validation (top left), the final regressor computed by using all data (top right) and the 30-gene TLS ST signature (bottom). Spots with TLS as at least 25% of their morphological annotation were binarized as presence of TLS.

d Comparison of AUCs of our TLS ST (ST sig) and Lundeberg's TLS (Lundeberg sig) signatures computed on the RNA bulk-seq (Bulk) and on the ST global pseudobulk (PB) of the ST TNBC cohort to predict the presence of TLS identified by morphological annotation and by the morphological regressor.

Source data and exact P values are provided as a Source Data file.

AUC: area under the curve; LOO: leave-one-out; PB: pseudobulk; ST: spatial transcriptomics; TLS: tertiary lymphoid structure

Supplementary Figure 7



Clinical values of 30-gene TLS ST signature in the I-SPY2 trial and in the metastatic non-breast cancers undergoing treatment with immune checkpoint blockades.

a Prediction of pCR by the TLS ST signature in the I-SPY2 trial[4] (*N* = 987), by BC subtypes and/or treatment arms, univariate analysis.

b TLS ST signature levels computed on the bulk RNA-seq data between no pCR and pCR patients in the control arm (treated by paclitaxel only) of the I-SPY2 trial (*N* = 179). Two-sided P value was computed using Wilcoxon rank sum test. TNBC samples are represented by red dots, while HR+/HER2- samples are represented by orange dots.

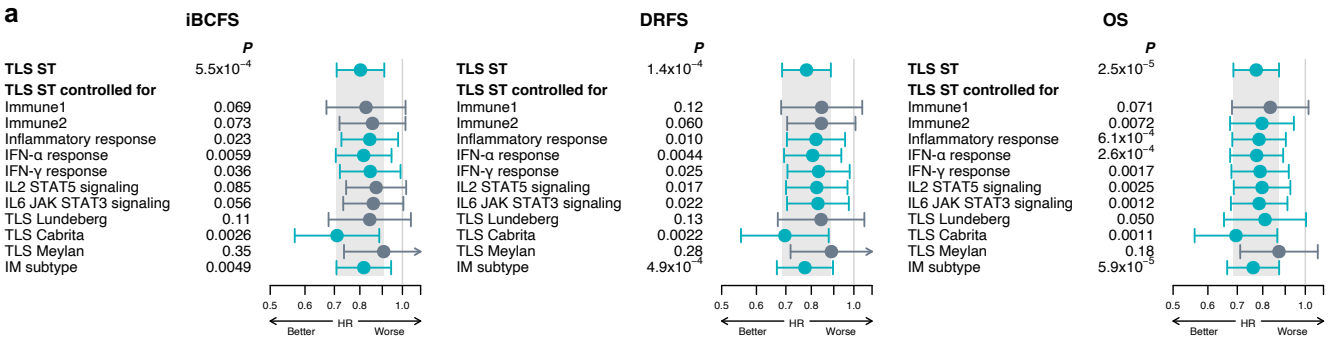
c Volcano plot showing the predictive value of TLS ST signature compared to other existing signatures to predict OS in metastatic non-breast cancers undergoing treatment with immune checkpoint blockades[5] (*N* = 555).

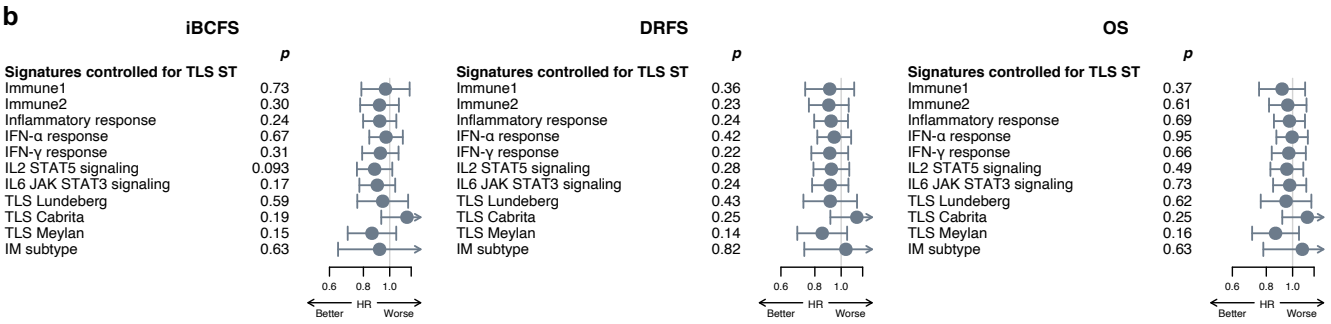
d Prediction of radiological response (RECIST; *N* = 842) and survival outcomes (PFS; *N* = 572 and OS; *N* = 856) by the TLS ST signature computed on bulk RNA-seq data in metastatic non-breast cancers undergoing treatment with immune checkpoint blockades, univariate analysis.

Forest plots: P values were obtained by Wilcoxon rank sum tests, OR or HR by logistic regression. Non-significant FDRs ≥ 0.05 are shown in dark grey, significant FDRs are shown in blue (< 0.05). Circles indicate OR or HR, and error bars the 95% confidence interval (95% CI).
Source data are provided as a Source Data file.

CI: confidence interval; Ctr: control arm (paclitaxel); FDR: false-discovery rate; HR: hazard ratio; HR+/-: hormone receptor positive/negative; IFN**γ**-iCAF: interferon gamma signaling pathway cancer associated fibroblast S1; N: neratinib; OR: odds ratio; pCR: pathological complete response; PFS: progression-free survival; OS: overall survival; RECIST: Response Evaluation Criteria in Solid Tumours; ST: spatial transcriptomics; TAM: tumor associated macrophages; TDM1/P: trastuzumab emtansine/ pertuzumab; TLS: tertiary lymphoid structures; TN or TNBC: triple negative breast cancer; VC: veliparib; VCpredTN: veliparib carboplatin prediction triple negative.

Supplementary Figure 8





Interaction between 30-gene TLS ST signature and other immune signatures for the prediction of survival outcomes in TNBC patients.

a Association between the TLS ST signature (before and after adjusting for various immune signatures) and the prediction for various survival outcomes in the merged TNBC cohorts (ST TNBC, METABRIC and SCAN-B)[6,7] (*N* = 1101).

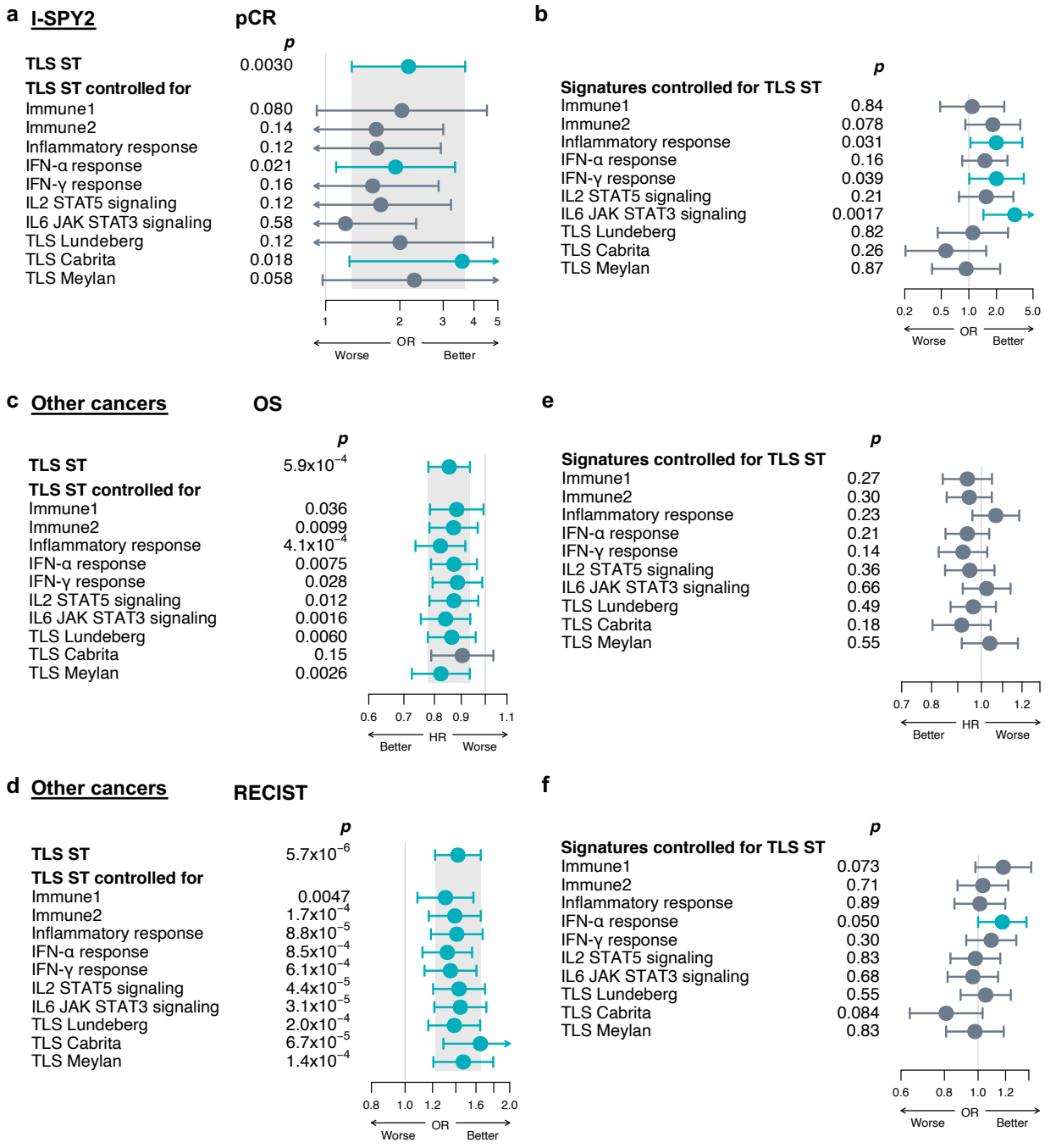
b Association between the various immune signatures and IM subtype after adjusting for TLS ST signature, and the prediction for various survival outcomes in the merged TNBC cohorts (ST TNBC, METABRIC and SCAN-B).

P values were obtained with a likelihood ratio test on nested models using the study ID as a strata, significant P values are shown in blue (< 0.05). Circles indicate HR, and error bars the 95% confidence interval (95% CI).

Source data are provided as a Source Data file.

DRFS: distant relapse–free survival; HR: hazard ratio; iBCFS: invasive breast cancer–free survival; IM: immunomodulatory; OS: overall survival; ST: spatial transcriptomics; TLS: tertiary lymphoid structures.

Supplementary Figure 9



Interaction between 30-gene TLS ST signature and other immune signatures for the prediction of therapeutic response or clinical outcome in other BC and non-BC datasets.

a Association between the TLS ST signature (before and after adjusting for various immune signatures) and the prediction for pCR in all BC subtypes patients treated with paclitaxel plus pembrolizumab in I-SPY2 trial (*N* = 69).

b Association between the various immune signatures after adjusting for TLS ST signature, and the prediction for pCR in all BC subtypes patients treated with paclitaxel plus pembrolizumab in I-SPY2 trial.

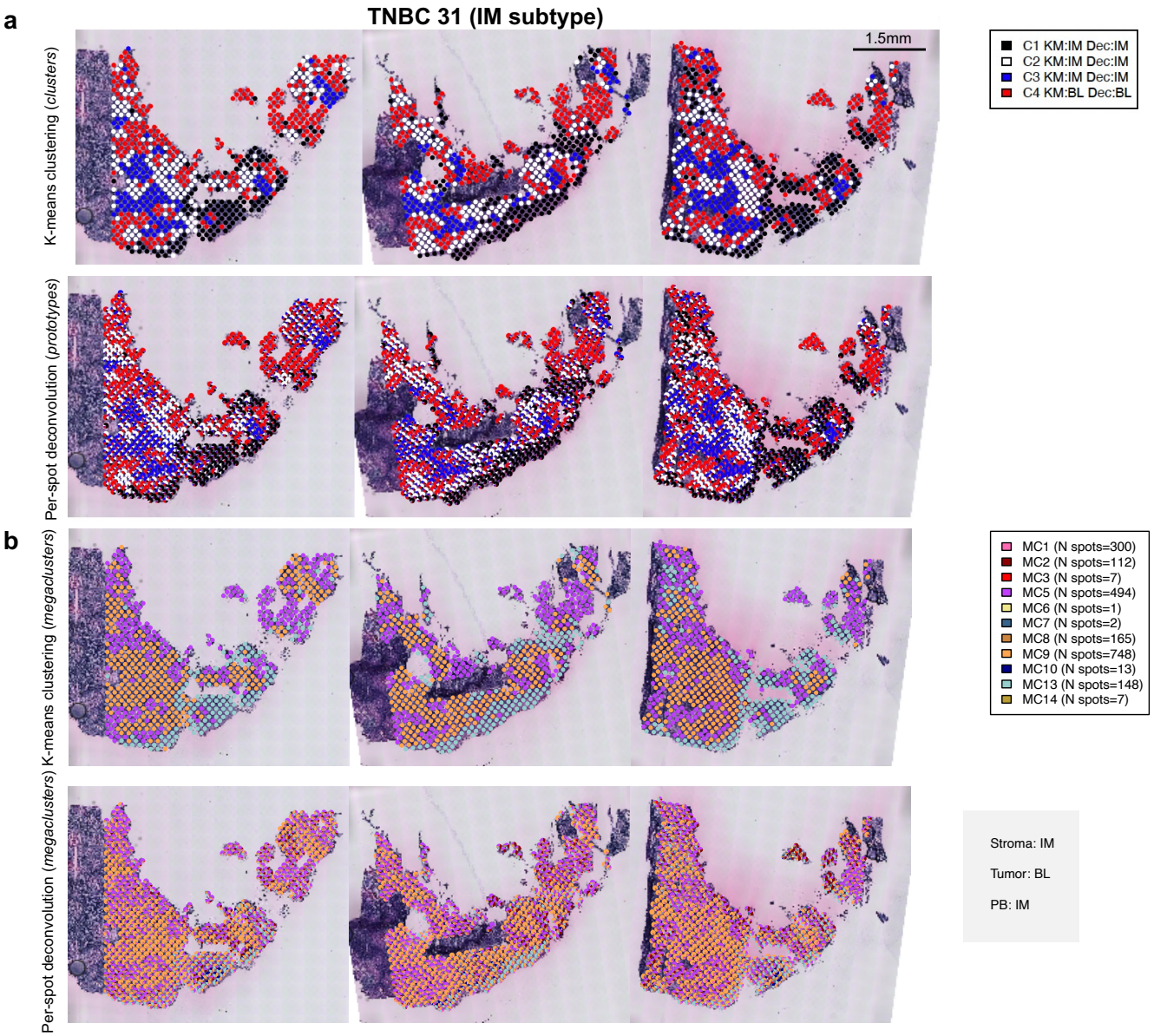
c,d Association between the TLS ST signature (before and after adjusting for various immune signatures) and the prediction for OS (*N* = 856) **(c)** and radiological response (RECIST) (*N* = 842) **(d)** in metastatic non-breast cancers undergoing treatment with immune checkpoint inhibitors.

e,f Association between the various immune signatures after adjusting for TLS ST signature, and the prediction for OS **(e)** and radiological response (RECIST) **(f)** in metastatic non-breast cancers undergoing treatment with immune checkpoint inhibitors.

P values were obtained with a likelihood ratio test on nested models, significant P values are shown in blue (< 0.05). Circles indicate HR, and error bars the 95% confidence interval (95% CI).
Source data are provided as a Source Data file.

HR: hazard ratio; OR: odds ratio; OS: overall survival; pCR: pathological complete response; RECIST: Response Evaluation Criteria in Solid Tumours; ST: spatial transcriptomics; TLS: tertiary lymphoid structures.

Supplementary Figure 10



Projections of intra-patient and inter-patient clusters in a TNBC sample classified as IM subtype on the ST global pseudobulk (ST_TNBC_ID 31).

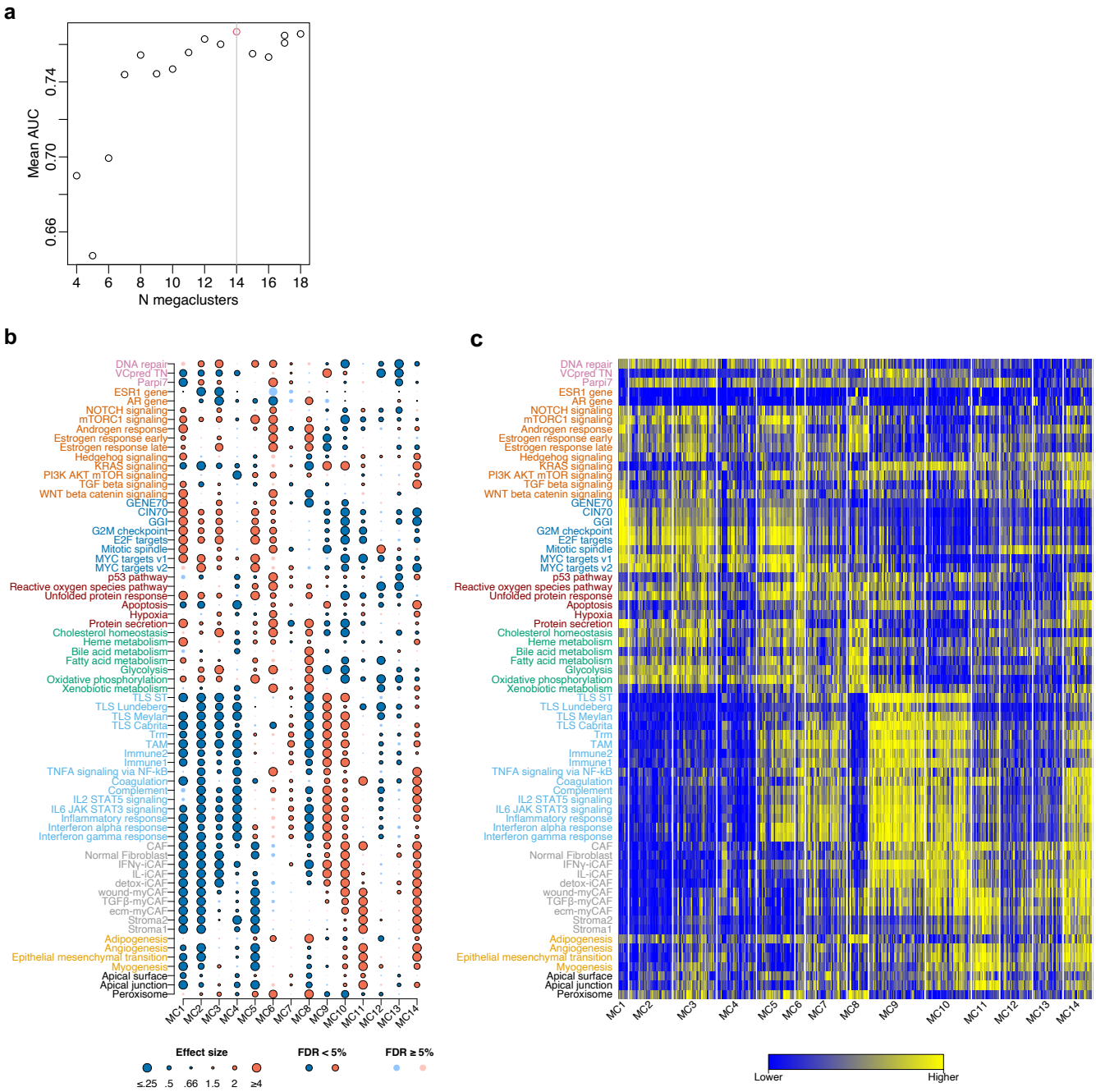
a Projections of intra-patient clusters defined by K-means (*top*) and by per-spot deconvolution of the intra-patient clusters (*bottom*). Clusters C1, C2, C3 are classified as IM and cluster C4 as BL.

b Projections of inter-patient megaclusters based on K-means (*top*) and independent projection of per-spot deconvoluted megaclusters (*bottom*). K-means-derived megaclusters matched with the spatial distribution of K-means-derived intra-patient clusters. The prevalence of per-spot deconvoluted megaclusters is represented by the number of cumulated spots assigned to the given megacluster (*N spot*). K-means identifies 3 megaclusters (MC 5, MC 9 and MC 13). Per-spot deconvolution identifies different contribution of 11 megaclusters.

At the sample level, the TNBC molecular subtypes are computed on the ST global pseudobulk (PB), stroma and tumor PBs (stroma, tumor).

Dec: deconvolution; IM: immunomodulatory; KM: K-means; PB: global pseudobulk; ST: spatial transcriptomics; TNBC: triple negative breast cancer.

Supplementary Figure 11



Optimization of megacluster numbers and their molecular characterization based on gene signatures in the ST TNBC cohort (*N* = 94).

a Optimal number of megaclusters based on the highest mean AUC from the recovery of each MC in the bulk RNA-seq of the ST TNBC cohort.

b Associations between gene signatures and megaclusters. Logistic regressions are used to evaluate associations between each specific gene signature and megacluster (one vs. rest). Two-sided P values were obtained from Wilcoxon rank sum tests and corrected for multi testing. Dots are bordered and dark-colored when FDRs < 0.05, compared to lighter-colored dots when FDRs ≥ 0.05. Negative and positive associations are represented in blue and red, respectively.

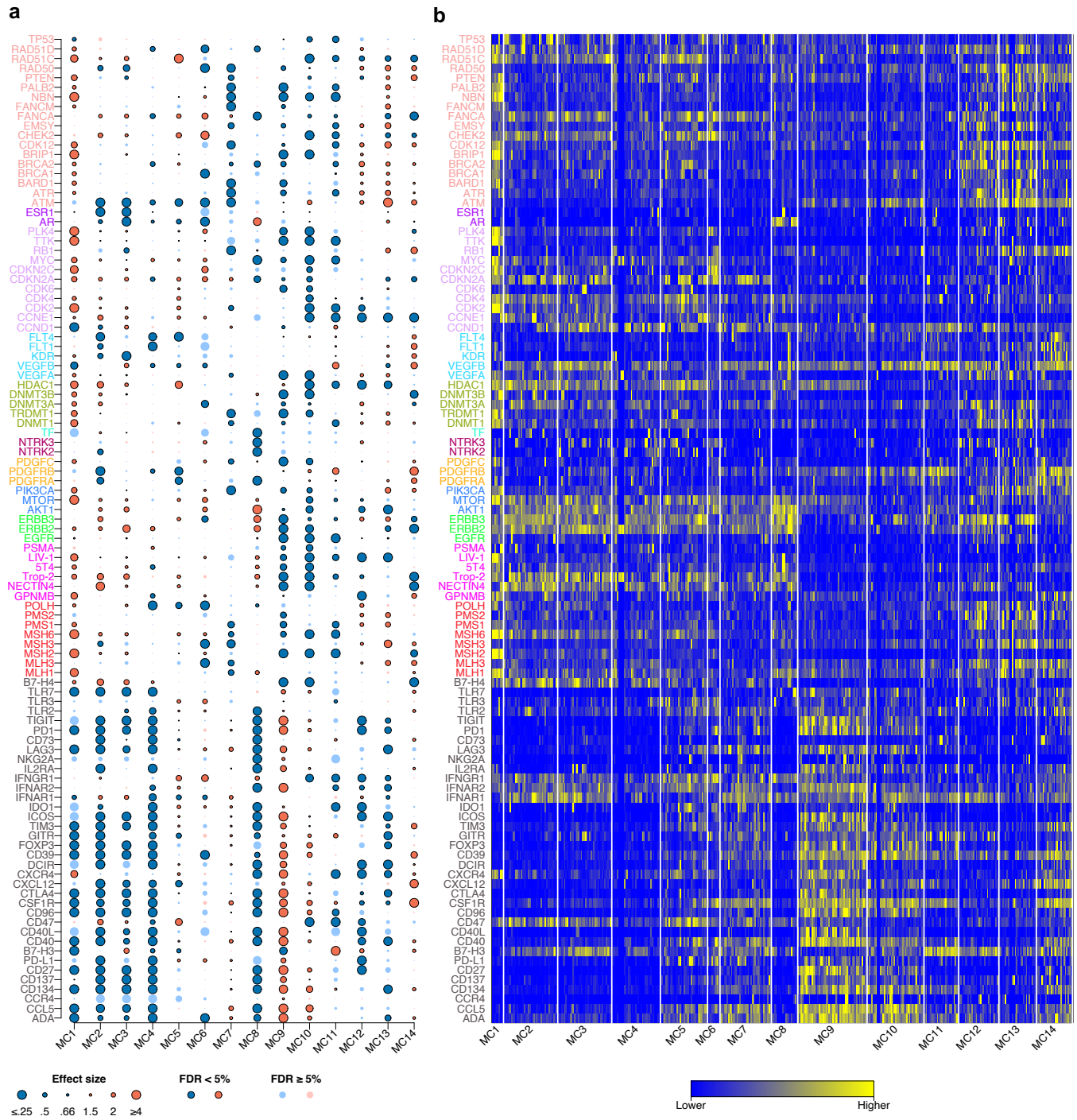
c Heatmap showing gene signatures on individual samples across the 14 megaclusters. The signatures are selected and ordered as in the adjacent dot plot.

Source data are provided as a Source Data file.

AUC: area under the curve; CAF: cancer associated fibroblast S1; FDR: false-discovery rate; GGI: genomic grade index; IL2: interleukin 2; IL6: interleukin 6; MC: megacluster; Parpi7: parp inhibitor

7; ST: spatial transcriptomics; TAM: tumor associated macrophages; TLS: tertiary lymphoid structure; TNF A: tumor necrosis factor alpha; Trm: tissue-resident memory T cell; VCpredTN: veliparib carboplatin prediction triple negative.

Supplementary Figure 12



Characterization of the fourteen megaclusters using the expression of single genes of interest in the ST TNBC cohort (*N* = 94).

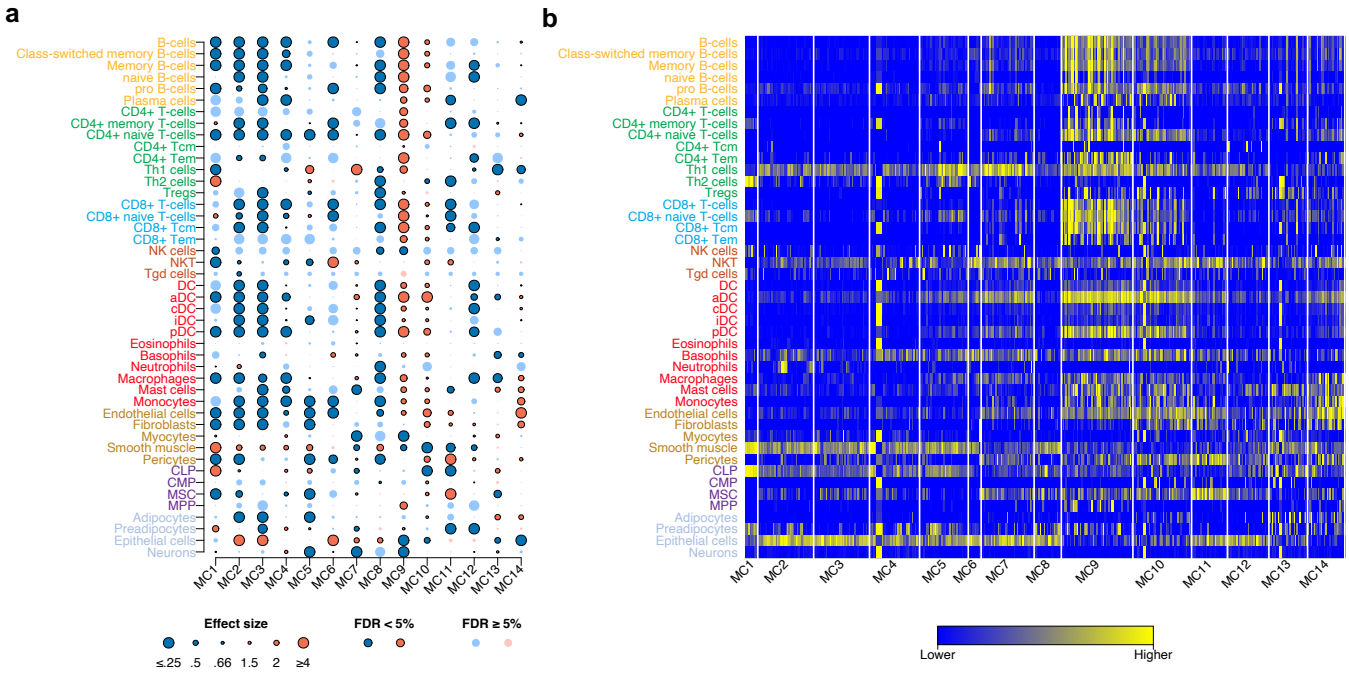
a Associations between single genes of interest and megaclusters. Logistic regressions are used to evaluate associations between each single gene expression and megacluster (one vs. rest). Two-sided P values were obtained from Wilcoxon rank sum tests and corrected for multi testing. Dots are bordered and dark-colored when FDRs < 0.05, compared to lighter-colored dots when FDRs ≥ 0.05. Negative and positive associations are represented in blue and red, respectively.

b Heatmap showing single genes of interest by individual samples across the 14 megaclusters. The single genes are selected and ordered as in the adjacent plot.

Source data are provided as a Source Data file.

FDR: false-discovery rate; MC: megacluster.

Supplementary Figure 13



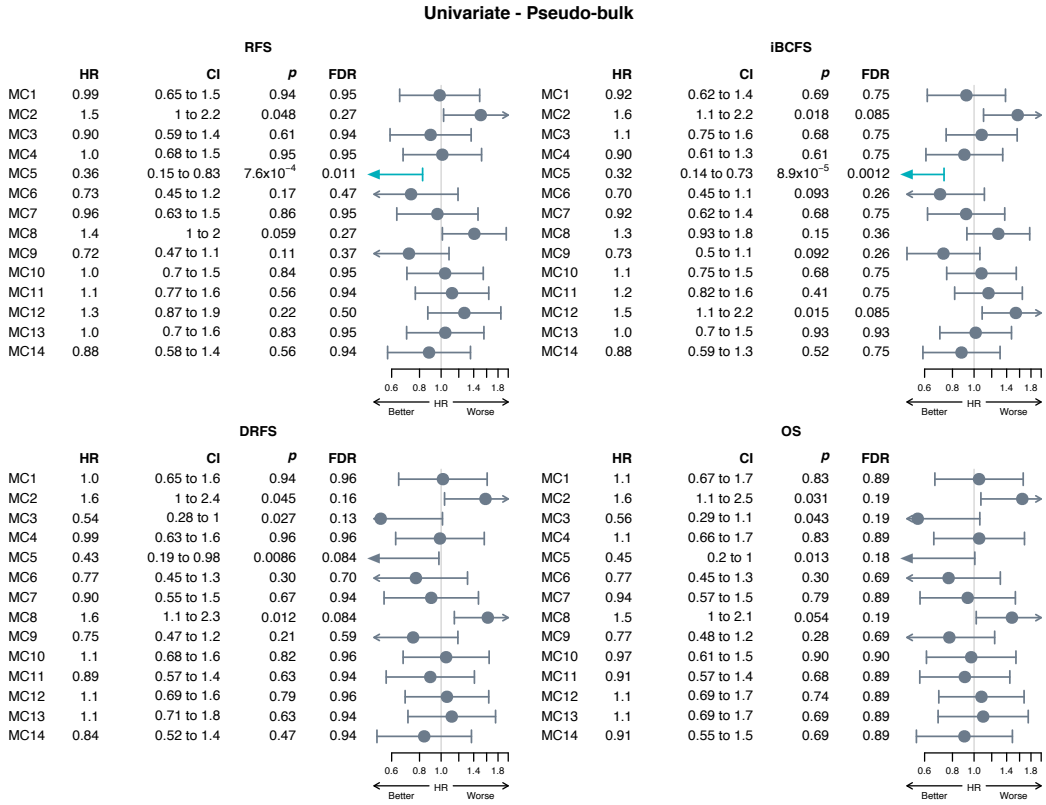
Characterization of the fourteen megaclusters using cell type enrichment analysis in the ST TNBC cohort (*N* = 94).

a Associations between cell type enrichment scores by xCell and each megacluster. Logistic regressions are used to evaluate associations between each cell type and each megacluster (one vs. rest). Two-sided P values were obtained from Wilcoxon rank sum tests and corrected for multi testing. Dots are bordered and dark-colored when FDRs < 0.05, compared to lighter-colored dots when FDRs ≥ 0.05. Negative and positive associations are represented in blue and red, respectively. **b** Heatmap showing cell type enrichment scores by individual samples across the 14 megaclusters. The cell types are selected and ordered as in the adjacent plot.
Source data are provided as a Source Data file.

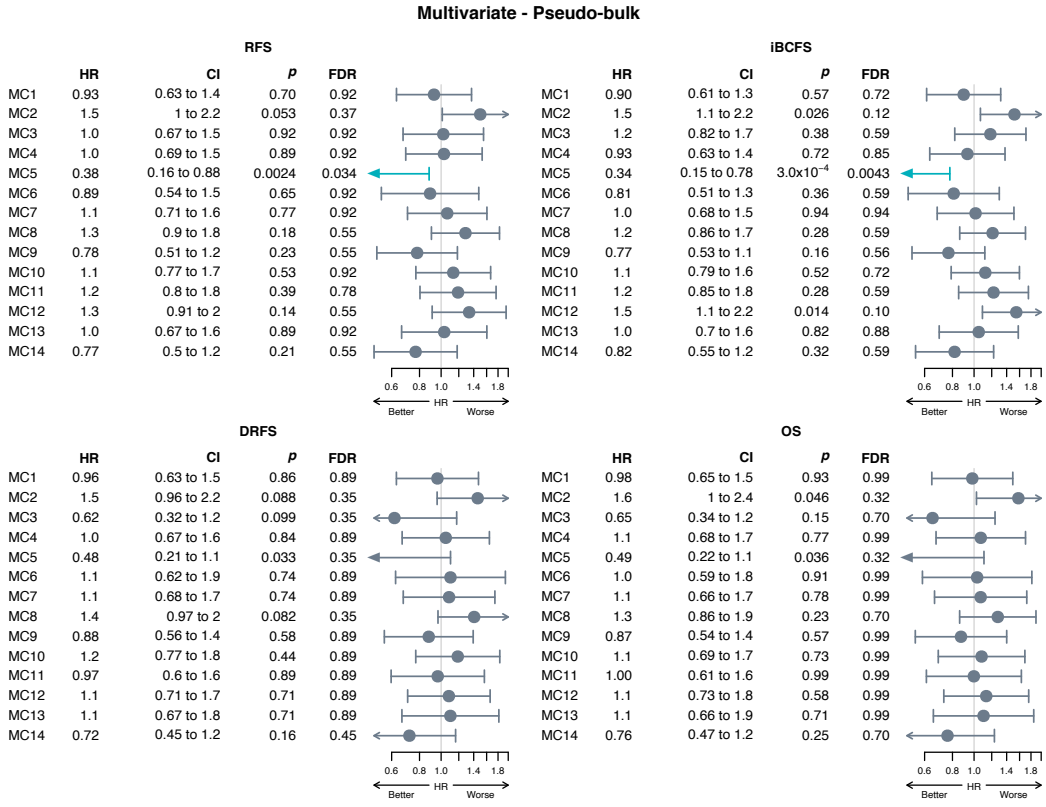
aDC: activated dendritic cells; cDC: conventional dendritic cells; CLP: common lymphoid progenitor; CMP: common myeloid progenitor; DC: dendritic cells; GMP: granulocyte-macrophage progenitor; iDC: immature dendritic cells; MC: megacluster; MPP: multipotent progenitor; MSC: mesenchymal stem cell; NK: natural killer; pDC: plasmacytoid dendritic cells; Tcm: central memory T cells; Tem: effector memory T cells; Th1: type 1 helper; Th2: T helper 2; Tregs: regulatory T cells.

Supplementary Figure 14

a



b



Survival analysis of the fourteen megaclusters using the ST global pseudobulk in the ST TNBC cohort (*N* = 94).

Association between survival outcomes and proportion of each megacluster within individual spots, estimated by the mean of the proportions in all ST spots, calculated by per-spot deconvolution. Proportions below 1% were put to 1%, then proportions were log transformed and scaled to a standard deviation of 1.

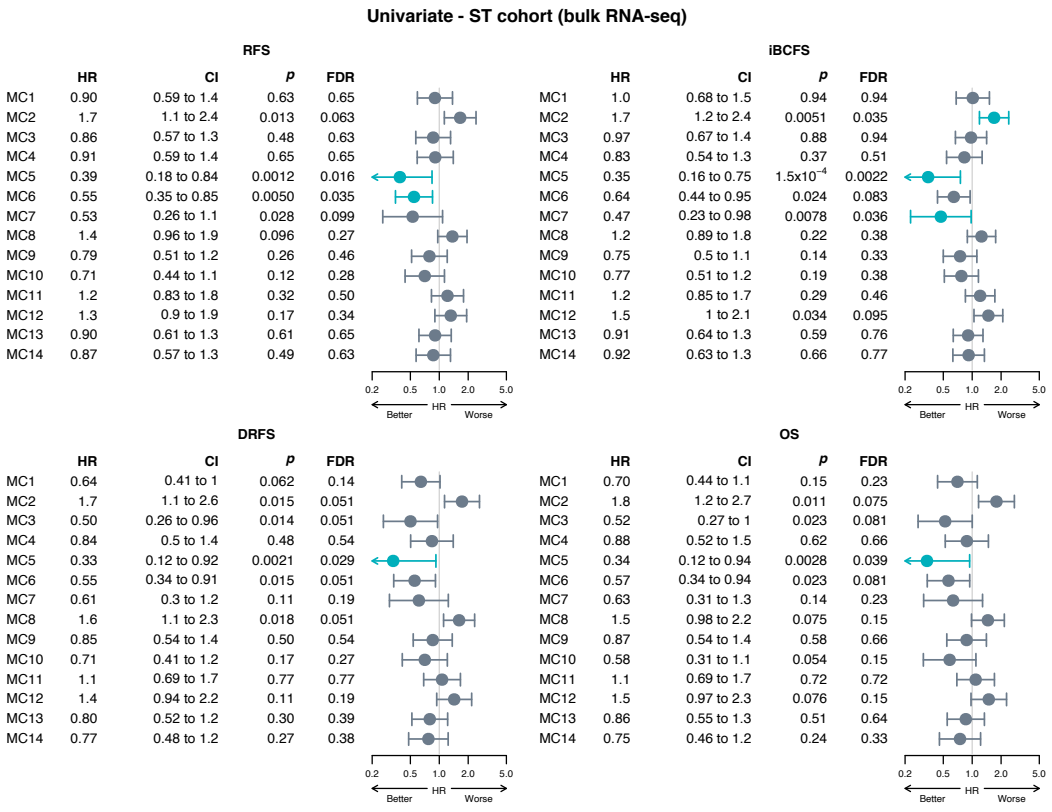
a Univariate analysis. P values were obtained using the likelihood ratio test.

b Multivariate analysis, correcting for clinic-pathological parameters (age, tumor size, nodal status). P values were obtained with a likelihood ratio test on nested Cox models. Significant FDRs are shown in blue (< 0.05). Circles indicate HR, and error bars the 95% confidence interval (95% CI). Source data are provided as a Source Data file.

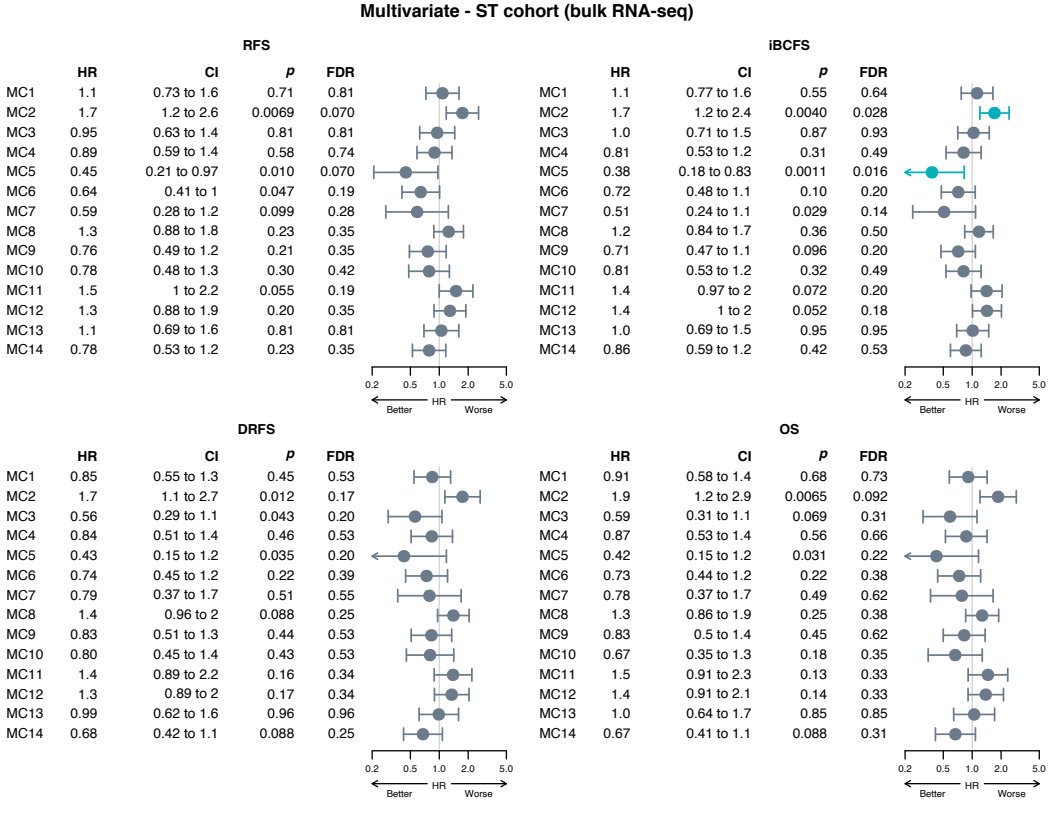
CI: confidence interval; DRFS: distant relapse-free survival; FDR: false-discovery rate; HR: hazard ratio; iBCFS: invasive breast cancer-free survival; MC: megacluster; OS: overall survival; PB: pseudobulk; RFS: recurrence-free survival; ST: spatial transcriptomics; TNBC: triple negative breast cancer.

Supplementary Figure 15

a



b



Survival analysis of the fourteen megaclusters using the bulk RNA-seq data in the ST TNBC cohort (*N* = 94).

Association between survival outcomes and proportion of each megacluster, as estimated by deconvolution from the bulk expression data. Proportions below 1% were put to 1%, then proportions were log transformed and scaled to a standard deviation of 1.

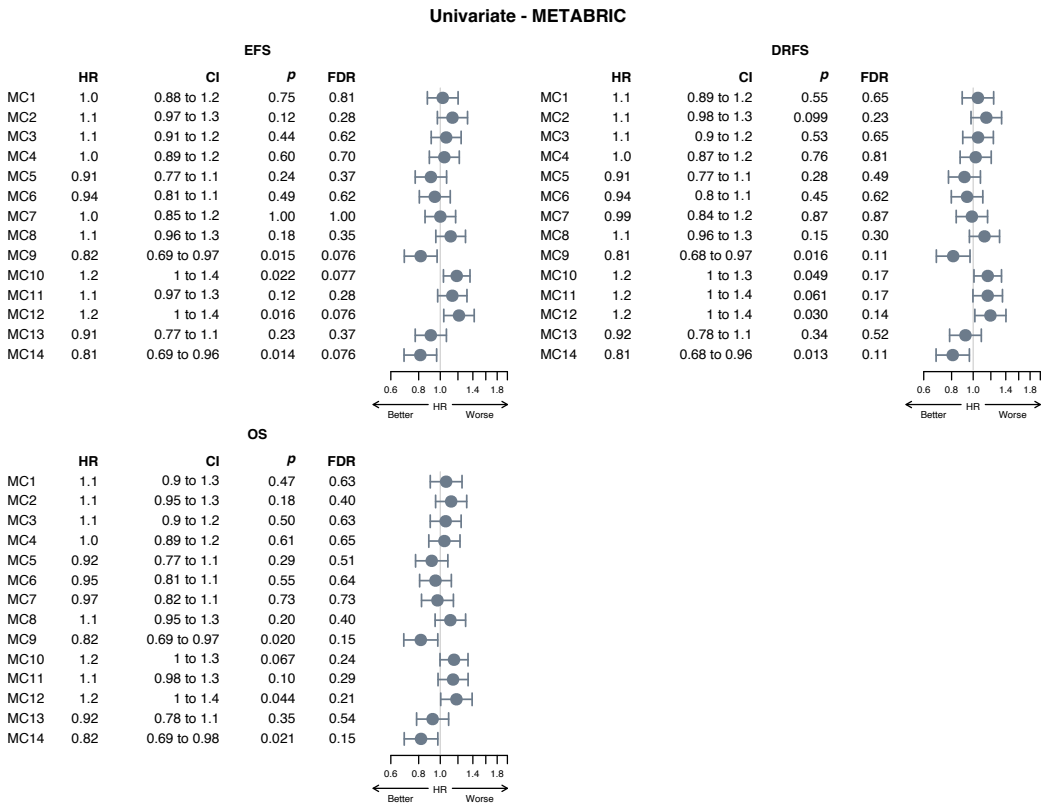
a Univariate analysis. P values were obtained using the likelihood ratio test.

b Multivariate analysis, correcting for clinic-pathological parameters (age, tumor size, nodal status). P values were obtained with a likelihood ratio test on nested Cox models. Significant FDRs are shown in blue (< 0.05). Circles indicate HR, and error bars the 95% confidence interval (95% CI). Source data are provided as a Source Data file.

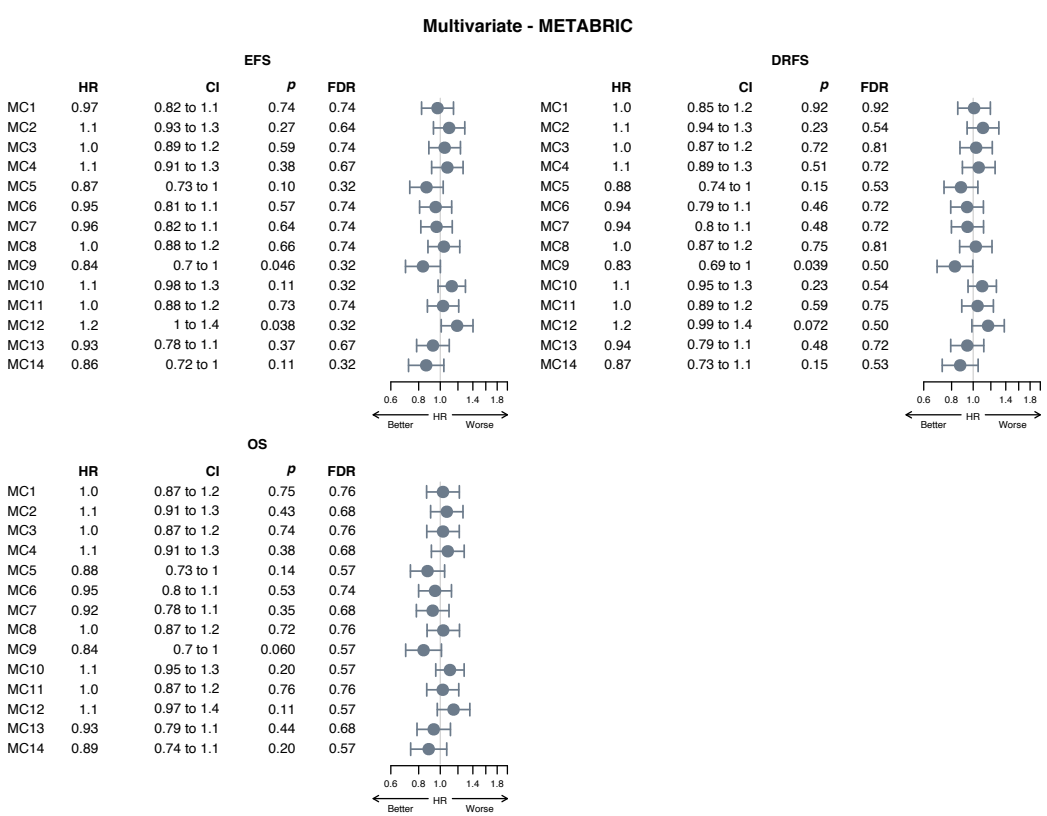
CI: confidence interval; DRFS: distant relapse-free survival; FDR: false-discovery rate; HR: hazard ratio; iBCFS: invasive breast cancer free survival; MC: megacluster; OS: overall survival; RFS: recurrence-free survival; ST: spatial transcriptomics; TNBC: triple negative breast cancer.

Supplementary Figure 16

a



b



Survival analysis of the fourteen megaclusters using the bulk RNA-seq data in the METABRIC TNBC cohort (*N* = 335).

Association between survival outcomes and proportion of each megacluster, as estimated by deconvolution from the bulk expression data. Proportions below 1% were put to 1%, then proportions were log transformed and scaled to a standard deviation of 1.

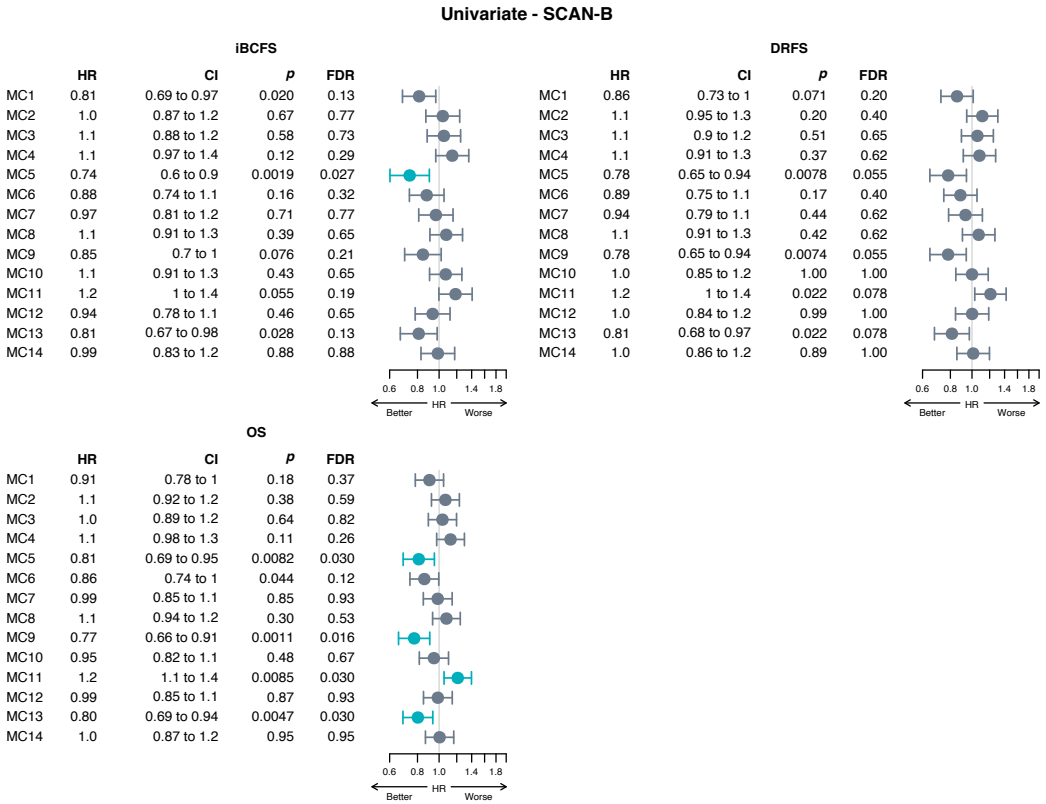
a Univariate analysis. P values were obtained using the likelihood ratio test.

b Multivariate analysis, correcting for clinic-pathological parameters (age, tumor size, nodal status). P values were obtained with a likelihood ratio test on nested Cox models. Significant FDRs are shown in blue (< 0.05). Circles indicate HR, and error bars the 95% confidence interval (95% CI). Source data are provided as a Source Data file.

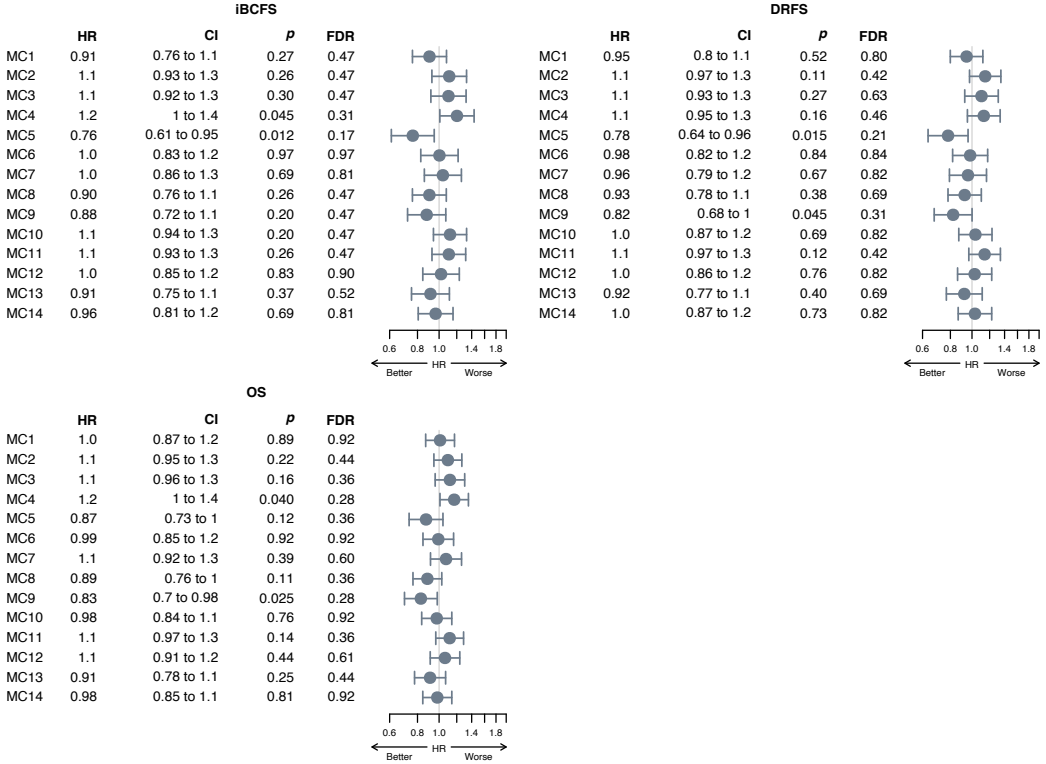
CI: confidence interval; DRFS: distant relapse-free survival; EFS: event-free survival; FDR: false-discovery rate; HR: hazard ratio; MC: megacluster; OS: overall survival; ST: spatial transcriptomics.

Supplementary Figure 17

a



b



Survival analysis of the fourteen megaclusters using the bulk RNA-seq data in the SCAN-B TNBC cohort (*N* = 672).

Association between survival outcomes and proportion of each megacluster, as estimated by deconvolution from the bulk expression data. Proportions below 1% were put to 1%, then proportions were log transformed and scaled to a standard deviation of 1.

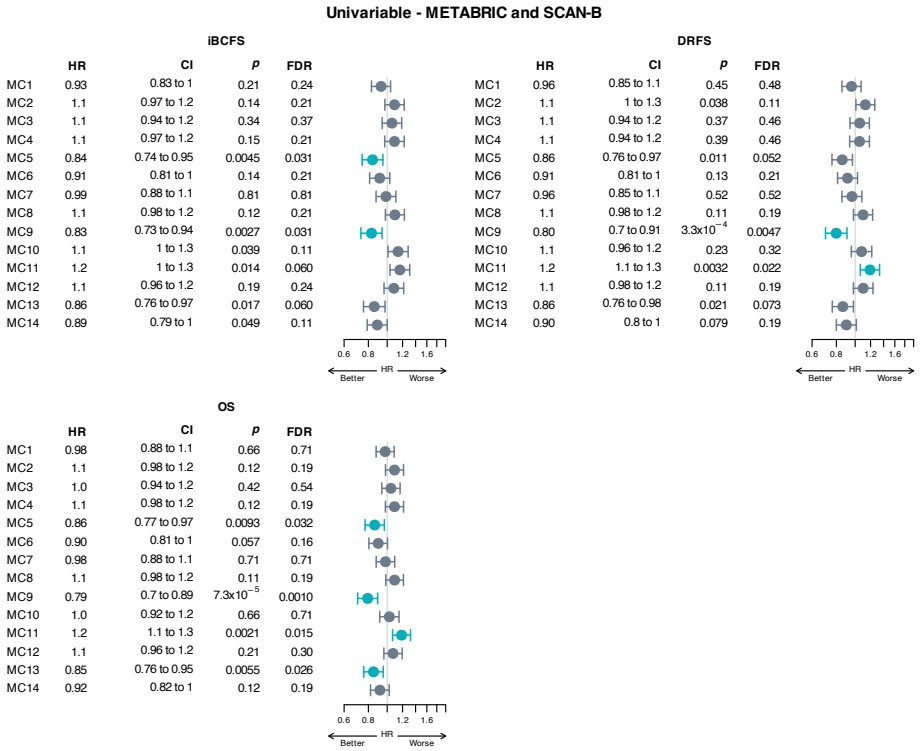
a Univariate analysis. P values were obtained using the likelihood ratio test.

b Multivariate analysis, correcting for clinic-pathological parameters (age, tumor size, nodal status). P values were obtained with a likelihood ratio test on nested Cox models. Significant FDRs are shown in blue (< 0.05). Circles indicate HR, and error bars the 95% confidence interval (95% CI). Source data are provided as a Source Data file.

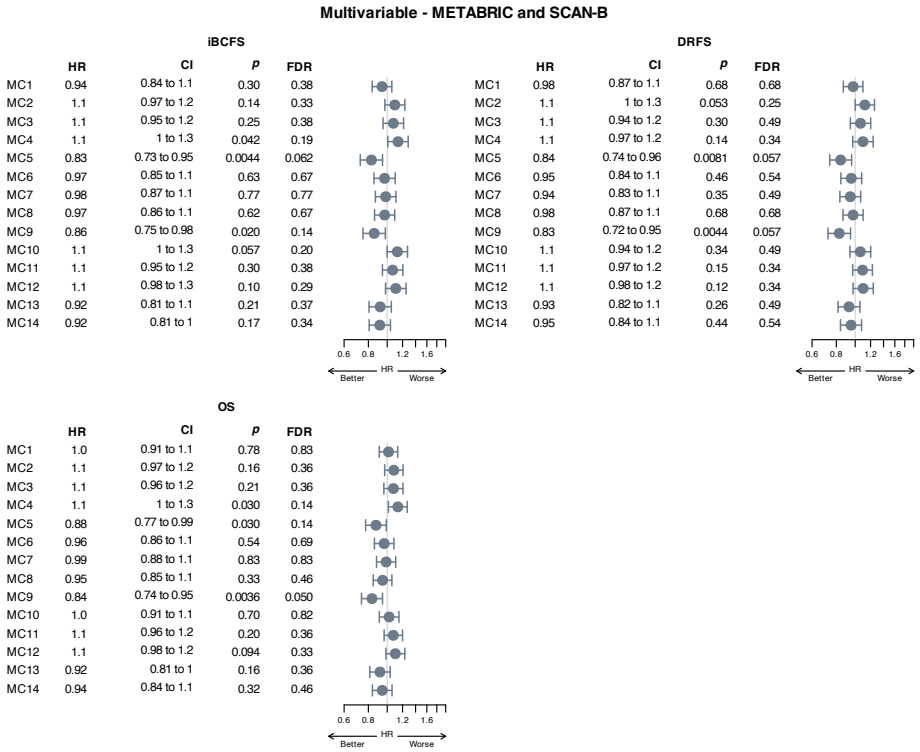
CI: confidence interval; DRFS: distant relapse-free survival; FDR: false-discovery rate; HR: hazard ratio; iBCFS: invasive breast cancer-free survival; MC: megacluster; OS: overall survival; ST: spatial transcriptomics.

Supplementary Figure 18

a



b



Survival analysis of the fourteen megaclusters using the bulk RNA-seq data in the merged METABRIC and SCAN-B TNBC cohorts (*N* = 1007).

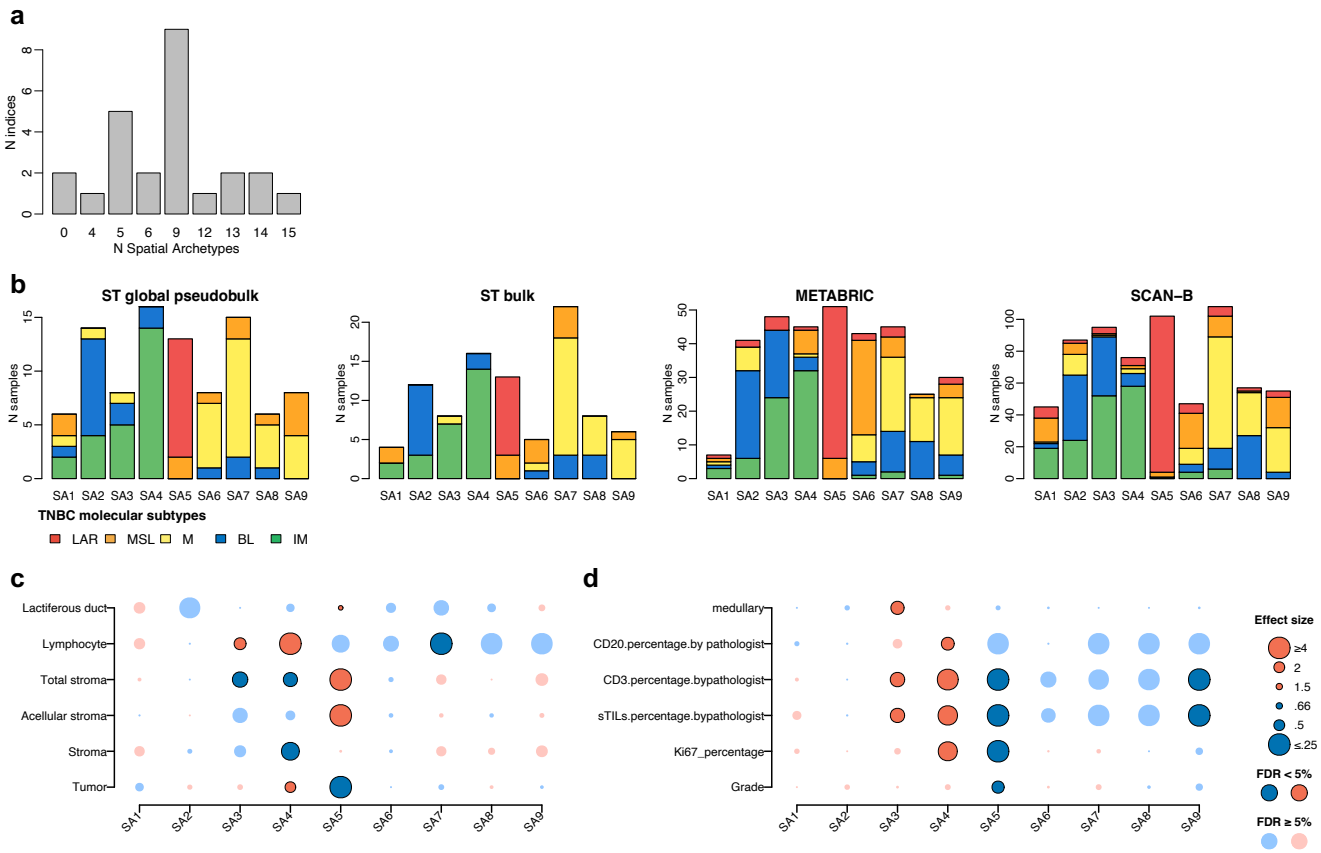
Association between survival outcomes and proportion of each megacluster, as estimated by deconvolution from the bulk expression data. Proportions below 1% were put to 1%, then proportions were log transformed and scaled to a standard deviation of 1. Survival analyses were stratified by study. The endpoint of iBCFS included also EFS from the METABRIC cohort.

a Univariate analysis. P values were obtained using the likelihood ratio test.

b Multivariate analysis, correcting for clinic-pathological parameters (age, tumor size, nodal status). P values were obtained with a likelihood ratio test on nested Cox models. Significant FDRs are shown in blue (< 0.05). Circles indicate HR, and error bars the 95% confidence interval (95% CI). Source data are provided as a Source Data file.

CI: confidence interval; DRFS: distant relapse-free survival; FDR: false-discovery rate; HR: hazard ratio; iBCFS: invasive breast cancer-free survival; MC: megacluster; OS: overall survival; ST: spatial transcriptomics.

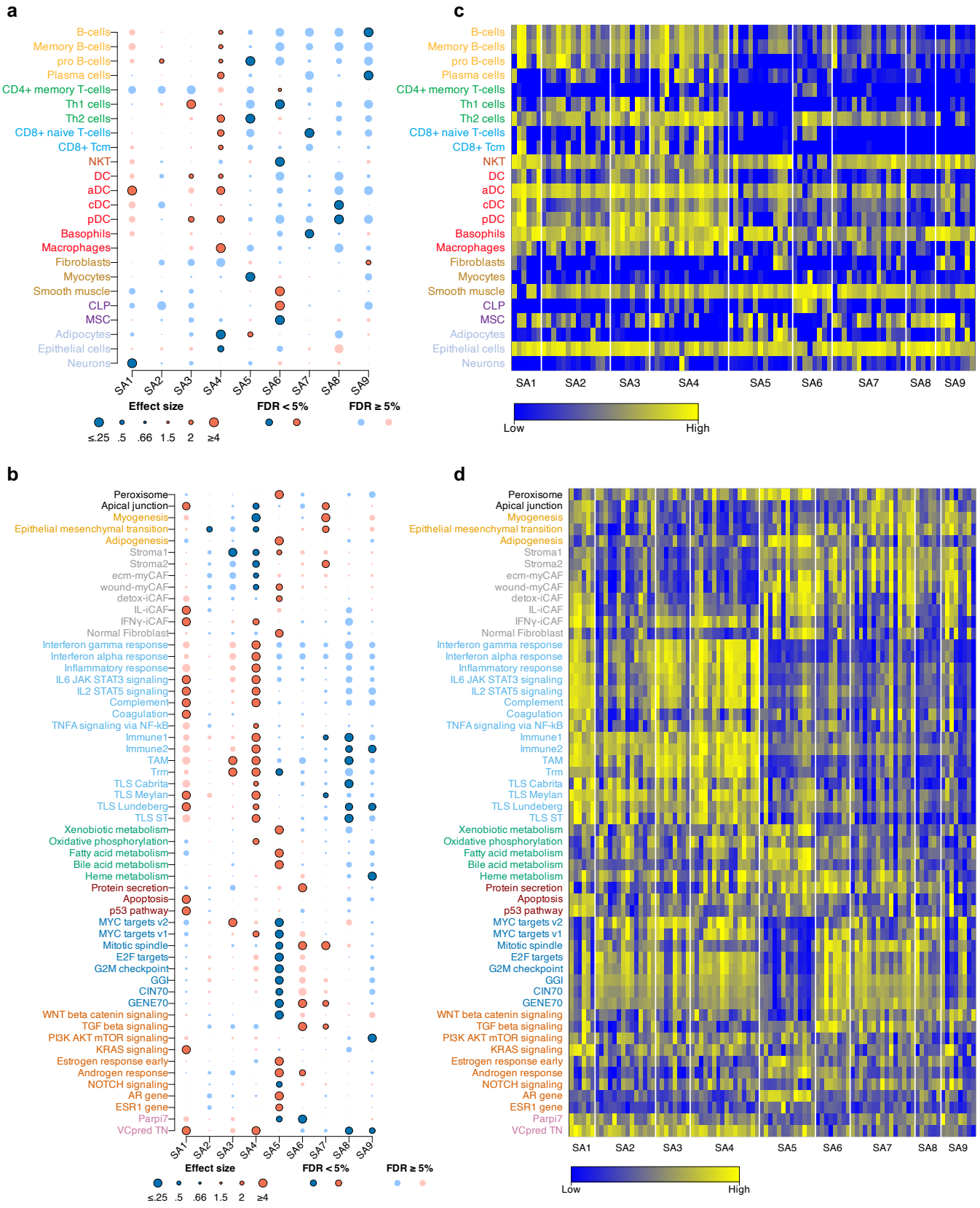
Supplementary Figure 19



Clinic-morphological and molecular characterization of the nine spatial archetypes.
a Optimal number of clusters (denoted as N spatial archetypes) determined using 23 different indices from hierarchical clustering of megacluster proportions with the ward.D2 method. The highest value (N Clusters) represents the optimal number of clusters to be considered.
b Distributions of the TNBC molecular subtypes within each of the 9 spatial archetypes in the different cohorts: ST TNBC global pseudobulk (*N* = 94), ST TNBC bulk RNA-seq (*N* = 94), METABRIC TNBC (*N* = 335) and SCAN-B TNBC cohorts (*N* = 672).
c Associations between contribution of morphological annotations and spatial archetypes in the ST TNBC cohort (global pseudobulk).
d Associations between clinic-pathological data and spatial archetypes in the ST TNBC cohort (global pseudobulk). A logistic regression was used to evaluate associations between each feature and spatial archetype. P values were obtained from Wilcoxon rank sum tests and corrected for multi testing. Dots are bordered and dark-colored when FDRs < 0.05, compared to lighter-colored dots when FDRs ≥ 0.05. Negative and positive associations are represented in blue and red, respectively. Source data are provided as a Source Data file.

BL: basal-like; IM: immunomodulatory; LAR: luminal androgen receptor; M: mesenchymal; MSC: mesenchymal stem cell; MSL: mesenchymal stem-like; SA: spatial archetype; ST: spatial transcriptomics; sTILs: stromal tumor infiltrating lymphocytes;

Supplementary Figure 20

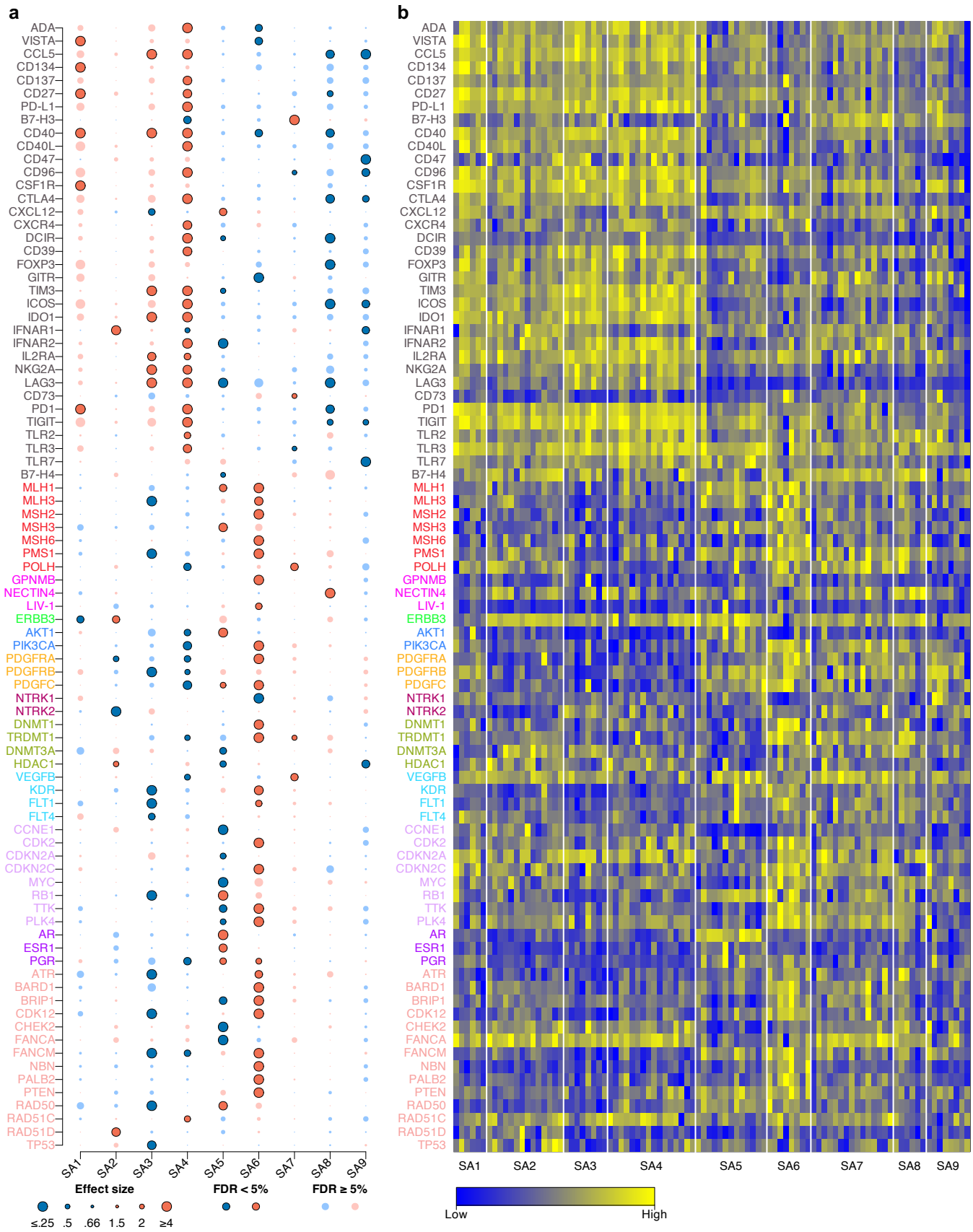


Molecular characterization of the nine spatial archetypes in the ST TNBC cohort (*N* = 94). **a,b** Associations between cell type enrichment scores by xCell **(a)** or gene signatures **(b)** and spatial archetypes in the ST TNBC cohort (global pseudobulk). A logistic regression was used to evaluate associations between each feature and spatial archetype. P values were obtained from Wilcoxon rank sum tests and corrected for multi testing. Dots are bordered and dark-colored when FDRs < 0.05, compared to lighter-colored dots when FDRs ≥ 0.05. Negative and positive associations are represented in blue and red, respectively.

c,d Heatmap showing cell type enrichment scores **(c)** or gene signatures **(d)** by individual samples across the 9 spatial archetypes. Cell types and gene signatures are selected and ordered as in the adjacent dot plot.
Source data are provided as a Source Data file.

aDC: activated dendritic cells; AR: androgen receptor; BL: basal-like; CAF: cancer associated fibroblast S1; cDC: conventional dendritic cells; CLP: common lymphoid progenitor; DC: dendritic cells; FDR: false-discovery rate; GGI: genomic grade index; IL2: interleukin 2; IL6: interleukin 6; IM: immunomodulatory; LAR: luminal androgen receptor; M: mesenchymal; MSC: mesenchymal stem cell; MSL: mesenchymal stem-like; NKT: natural killer T cells; Parpi7: parp inhibitor 7; pDC: plasmacytoid dendritic cells; SA: spatial archetype; ST: spatial transcriptomics; sTILs: stromal tumor infiltrating lymphocytes; Tcm: central memory T cells; Th1: T helper 1; Th2: T helper 2; TAM: tumor associated macrophages; TLS: tertiary lymphoid structure; TNF A: tumor necrosis factor alpha; Trm: tissue-resident memory T cell; VCpredTN: veliparib carboplatin prediction triple negative.

Supplementary Figure 21



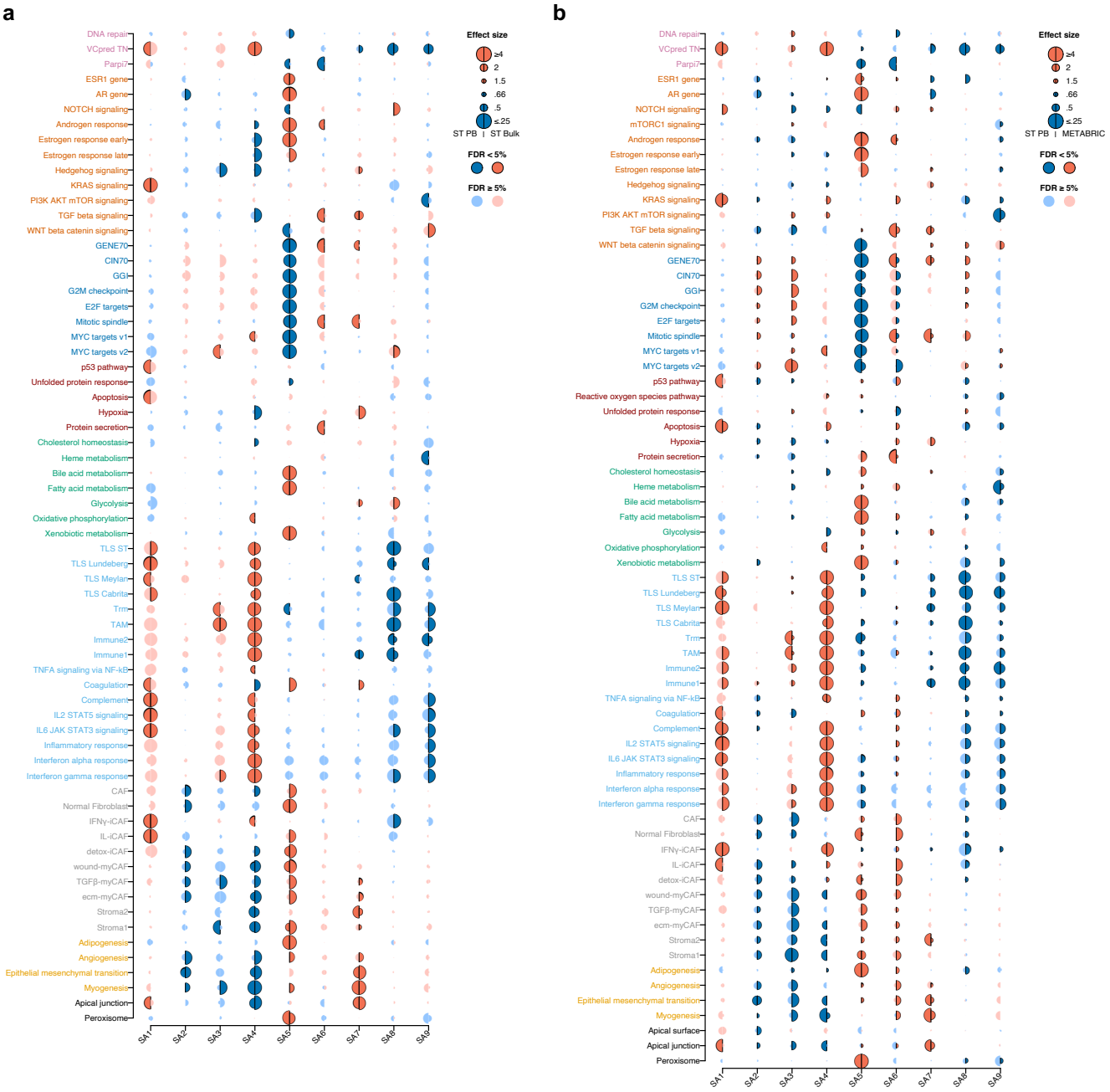
Molecular characterization of the nine spatial archetypes by the expression of single genes of interest in the ST TNBC cohort (*N* = 94).

a Associations between single genes of interest and spatial archetypes. A logistic regression model was used to evaluate associations between each single gene expression and spatial archetype. P values were obtained from Wilcoxon rank sum tests and corrected for multi testing. Dots are

bordered and dark-colored when FDRs < 0.05, compared to lighter-colored dots when FDRs ≥ 0.05. Negative and positive associations are represented in blue and red, respectively.
b Heatmap showing single genes of interest by individual samples across the 9 spatial archetypes. The single genes are selected and ordered as in the adjacent plot.
Source data are provided as a Source Data file.

FDR: false-discovery rate; SA: spatial archetypes.

Supplementary Figure 22



Comparison of gene signatures across the nine spatial archetypes in the ST TNBC (*N* = 94) and METABRIC TNBC cohorts (*N* = 335).

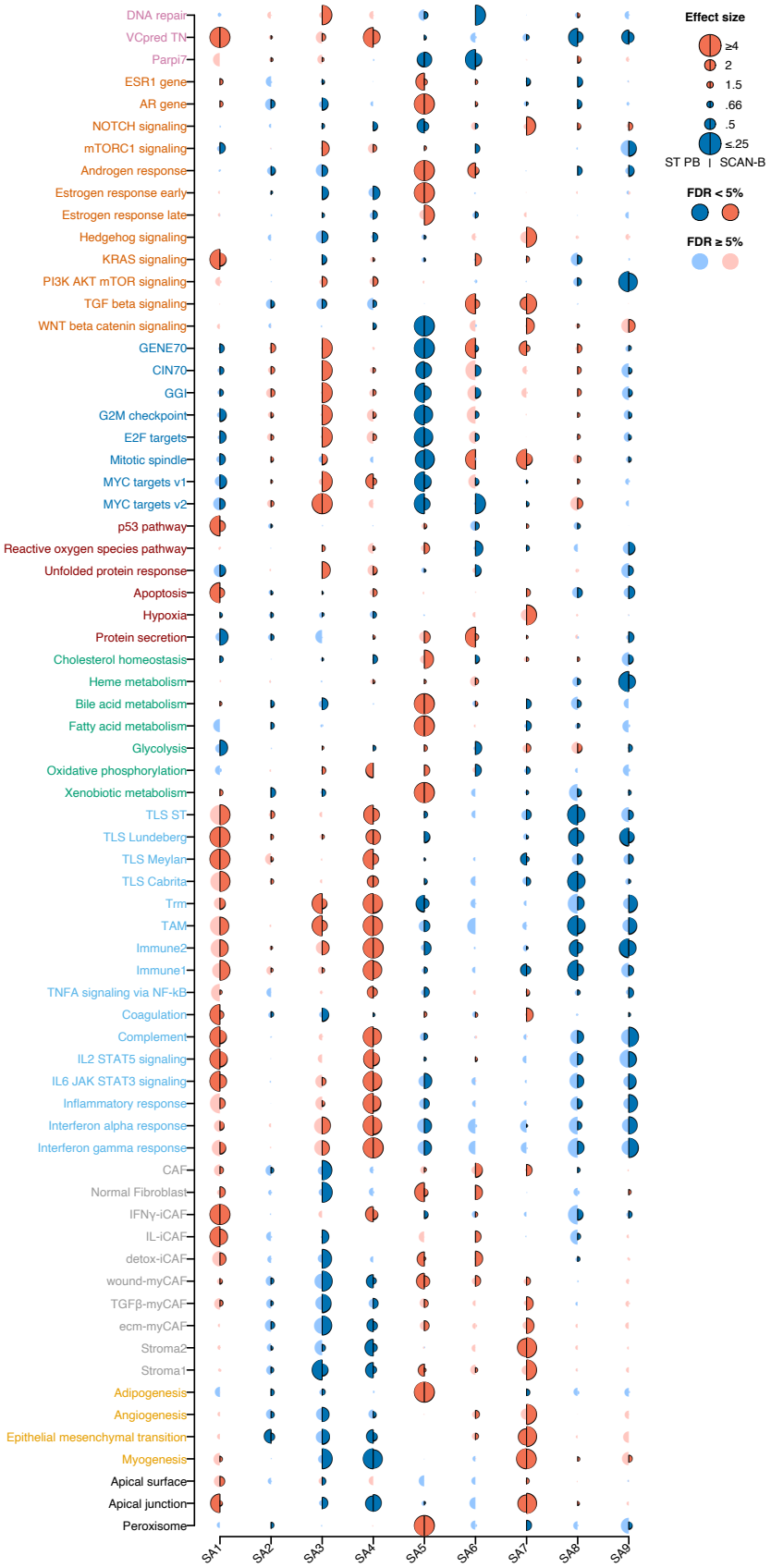
a,b Associations between gene signatures and spatial archetypes in the ST TNBC and METABRIC cohorts. The left half-circle represents ST TNBC global pseudobulk and the right half-circle represents **(a)** ST TNBC bulk RNA-seq, **(b)** METABRIC TNBC cohorts.

A logistic regression model was used to evaluate associations between each specific gene signature and each spatial archetype. P values were obtained from Wilcoxon rank sum tests and corrected for multi testing. Dots are bordered and dark-colored when FDRs < 0.05, compared to lighter-colored dots when FDRs ≥ 0.05. Negative and positive associations are represented in blue and red, respectively.

Source data are provided as a Source Data file.

AR: androgen receptor; CAF: cancer associated fibroblast S1; GGI: genomic grade index; Parpi7: parp inhibitor 7; PB: pseudobulk; SA: spatial archetype; ST: spatial transcriptomics; TAM: tumor associated macrophages; TLS: tertiary lymphoid structure; Trm: tissue-resident memory T cell; VCpredTN: veliparib carboplatin prediction triple negative

Supplementary Figure 23



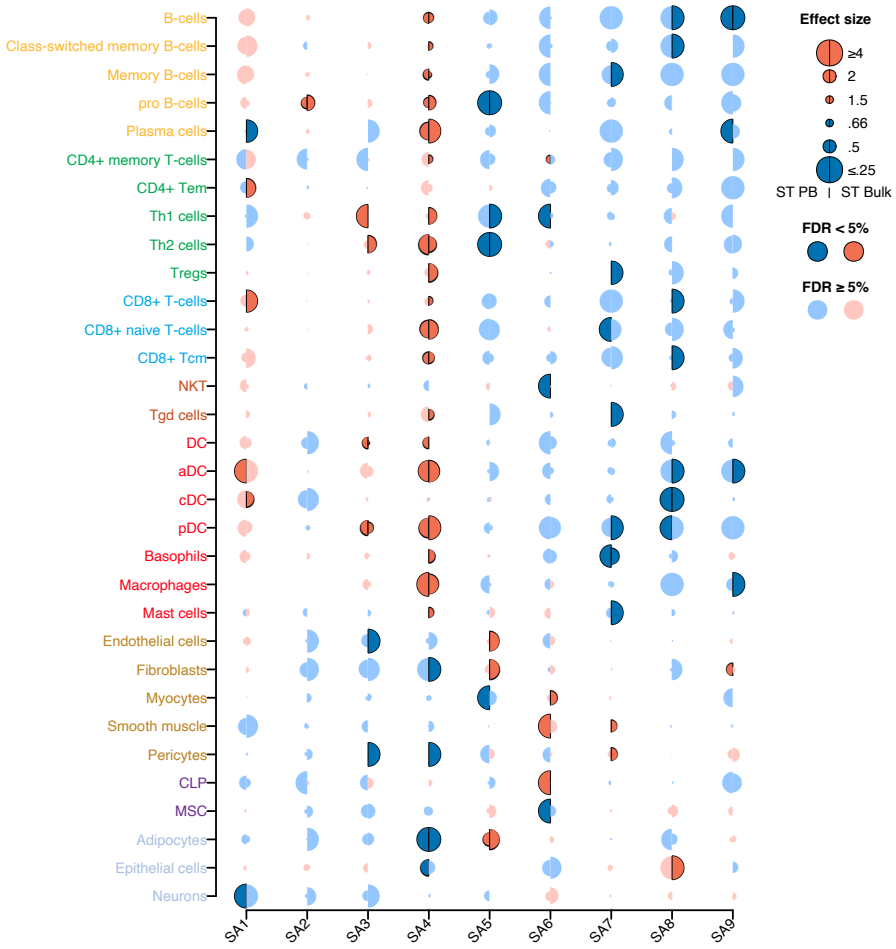
Comparison of gene signatures across the nine spatial archetypes in the ST TNBC (*N* = 94) and SCAN-B TNBC cohorts (*N* = 672).

Associations between expression-based signatures and spatial archetypes in the ST TNBC and SCAN-B TNBC cohorts. The left and right half-circles represent ST TNBC global pseudobulk and SCAN-B TNBC cohorts respectively. A logistic regression model was used to evaluate associations between each specific gene signature and each spatial archetype. P values were obtained from

Wilcoxon rank sum tests and corrected for multi testing. Dots are bordered and dark-colored when FDRs < 0.05, compared to lighter-colored dots when FDRs ≥ 0.05. Negative and positive associations are represented in blue and red, respectively.
Source data are provided as a Source Data file.

AR: androgen receptor; CAF: cancer associated fibroblast S1; GGI: genomic grade index; Parpi7: parp inhibitor 7; PB: pseudobulk; SA: spatial archetypes; ST: spatial transcriptomics; TAM: tumor associated macrophages; TLS: tertiary lymphoid structure; Trm: tissue-resident memory T cell; VCpredTN: veliparib carboplatin prediction triple negative.

Supplementary Figure 24

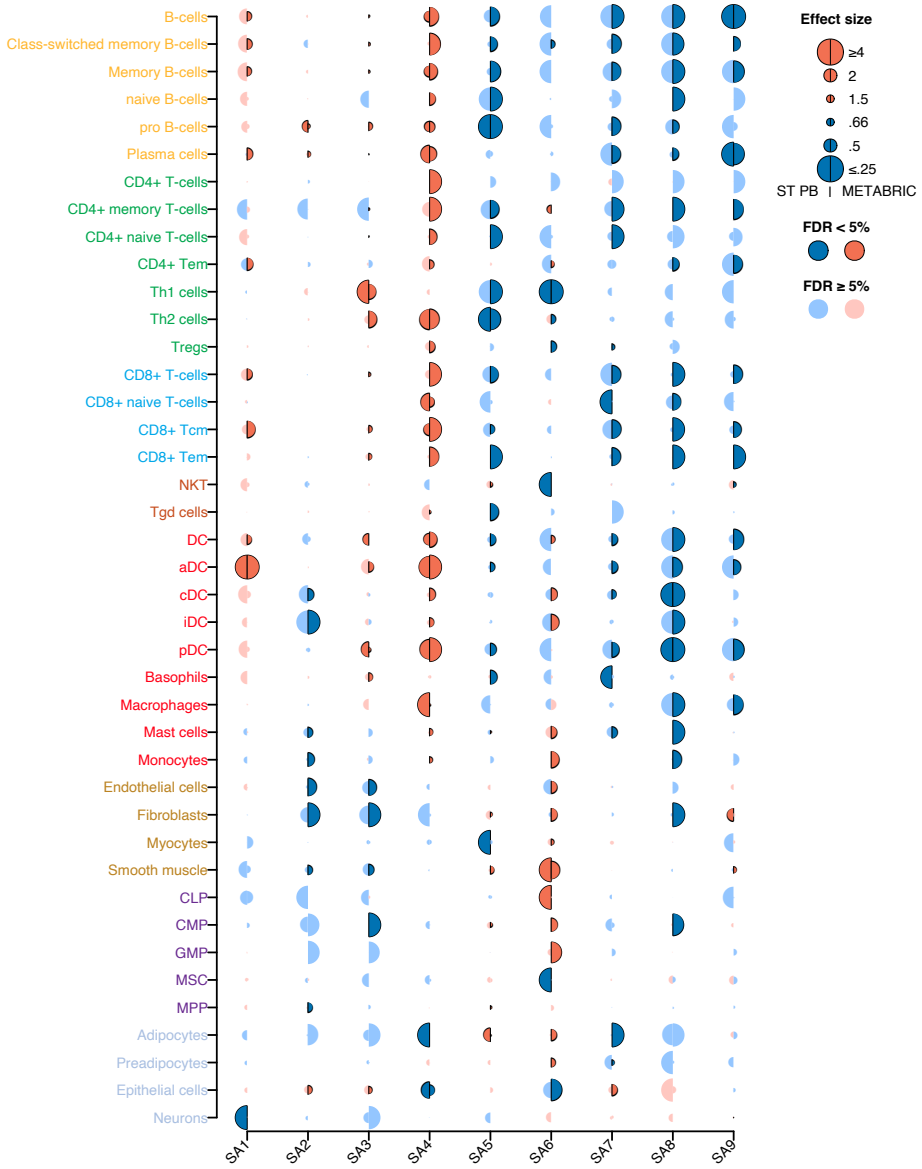


Comparison of cell type enrichment analysis across the nine spatial archetypes in the ST TNBC cohort (*N* = 94).
Associations between cell type enrichment scored by xCell and spatial archetypes. The left half-circle represents ST TNBC global pseudobulk and the right half-circle represents ST TNBC bulk RNA-seq.

A logistic regression model was used to evaluate associations between each cell type enrichment score and each spatial archetypes. P values were obtained from Wilcoxon rank sum tests and corrected for multi testing. Dots are bordered and dark-colored when FDRs < 0.05, compared to lighter-colored dots when FDRs ≥ 0.05. Negative and positive associations are represented in blue and red, respectively.
Source data are provided as a Source Data file.

aDC: activated dendritic cells; cDC: conventional dendritic cells; CLP: common lymphoid progenitor; CMP: common myeloid progenitor; DC: dendritic cells; GMP: granulocyte-macrophage progenitor; iDC: immature dendritic cells; MPP: multipotent progenitor; MSC: mesenchymal stem cell; NK: natural killer; PB: pseudobulk; pDC: plasmacytoid dendritic cells; SA: spatial archetype; ST: spatial transcriptomics; Tcm: central memory T cells; Tem: effector memory T cells; Th1: type 1 helper; Th2: T helper 2; Tregs: regulatory T cells.

Supplementary Figure 25

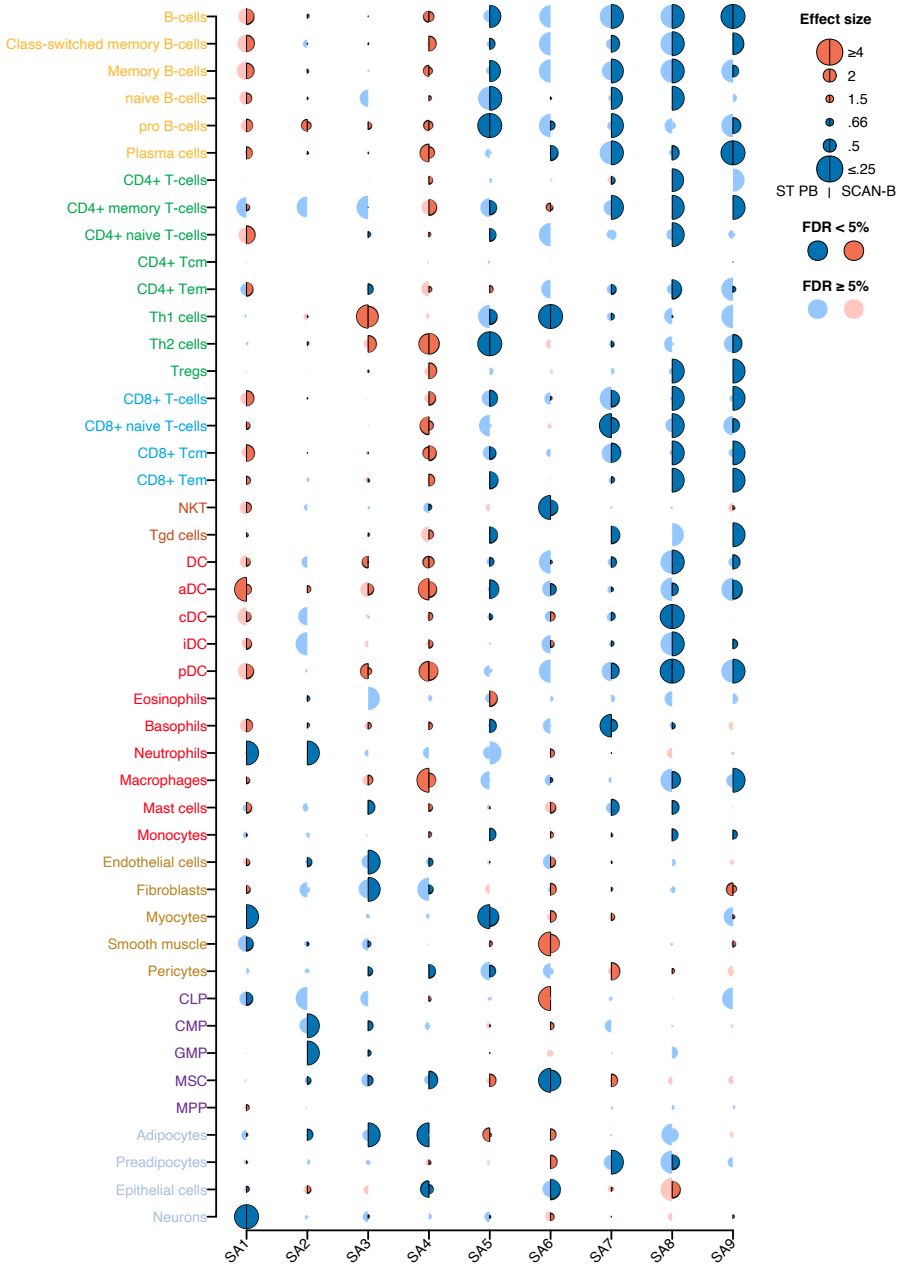


Comparison of cell type enrichment analysis across the nine spatial archetypes in the ST TNBC (*N* = 94) and METABRIC TNBC cohorts (*N* = 335).
Associations between cell type enrichment scored by xCell and spatial archetypes. The left half-circle represents ST TNBC global pseudobulk and the right half-circle represents METABRIC TNBC cohort.

A logistic regression model was used to evaluate associations between each cell type enrichment score and each spatial archetype. P values were obtained from Wilcoxon rank sum tests and corrected for multi testing. Dots are bordered and dark-colored when FDRs < 0.05, compared to lighter-colored dots when FDRs ≥ 0.05. Negative and positive associations are represented in blue and red, respectively.
Source data are provided as a Source Data file.

aDC: activated dendritic cells; cDC: conventional dendritic cells; CLP: common lymphoid progenitor; CMP: common myeloid progenitor; DC: dendritic cells; GMP: granulocyte-macrophage progenitor; iDC: immature dendritic cells; MPP: multipotent progenitor; MSC: mesenchymal stem cell; NK: natural killer; PB: global pseudobulk; pDC: plasmacytoid dendritic cells; SA: spatial archetype; ST: spatial transcriptomics; Tcm: central memory T cells; Tem: effector memory T cells; Th1: type 1 helper; Th2: T helper 2; Tregs: regulatory T cells.

Supplementary Figure 26



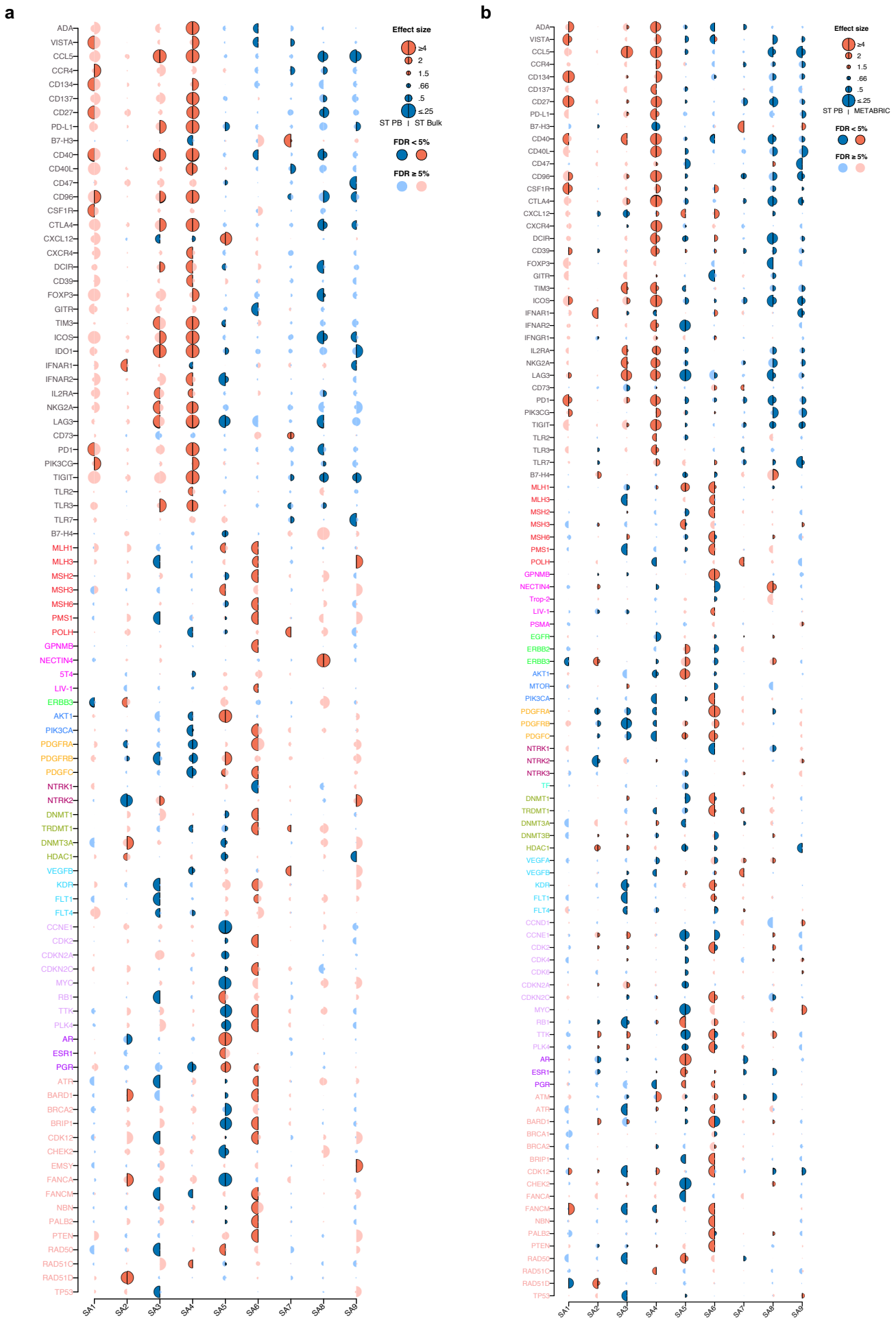
Comparison of cell type enrichment analysis across the nine spatial archetypes in the ST TNBC (*N* = 94) and SCAN-B TNBC cohorts (*N* = 672).
Associations between cell type enrichment scored by xCell and spatial archetypes. The left half-circle represents ST TNBC global pseudobulk and the right half-circle represents SCAN-B TNBC cohort.

A logistic regression model was used to evaluate associations between each cell type enrichment score and each spatial archetype. P values were obtained from Wilcoxon rank sum tests and corrected for multi testing. Dots are bordered and dark-colored when FDRs < 0.05, compared to lighter-colored dots when FDRs ≥ 0.05. Negative and positive associations are represented in blue and red, respectively.
Source data are provided as a Source Data file.

aDC: activated dendritic cells; cDC: conventional dendritic cells; CLP: common lymphoid progenitor; CMP: common myeloid progenitor; DC: dendritic cells; GMP: granulocyte-macrophage progenitor; iDC: immature dendritic cells; MPP: multipotent progenitor; MSC: mesenchymal stem cell; NK:

natural killer; PB: global pseudobulk; pDC: plasmacytoid dendritic cells; SA: spatial archetype; ST: spatial transcriptomics; Tcm: central memory T cells; Tem: effector memory T cells; Th1: type 1 helper; Th2: T helper 2; Tregs: regulatory T cells.

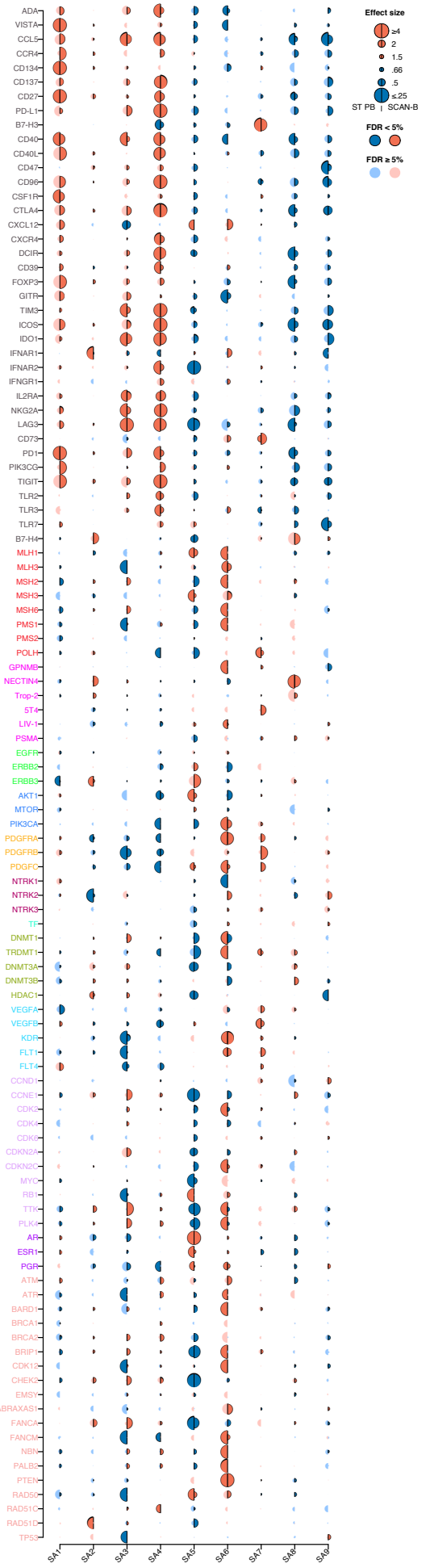
Supplementary Figure 27



Comparison of expression levels of single genes of interest across the nine spatial archetypes in the ST TNBC (*N* = 94) and METABRIC TNBC cohorts (*N* = 335).
Association between the expression levels of genes of interest and spatial archetypes. The left half-circle represents ST TNBC global pseudobulk and the right half-circle represents **(a)** ST TNBC bulk RNA-seq, **(b)** METABRIC TNBC cohorts.

A logistic regression model was used to evaluate associations between each cell type enrichment score and each spatial archetype. P values were obtained from Wilcoxon rank sum tests and corrected for multi testing. Dots are bordered and dark-colored when FDRs < 0.05, compared to lighter-colored dots when FDRs ≥ 0.05. Negative and positive associations are represented in blue and red, respectively.
Source data are provided as a Source Data file.

FDR: false-discovery rate; PB: global pseudobulk; SA: spatial archetype; ST: spatial transcriptomics.



Supplementary Figure 28

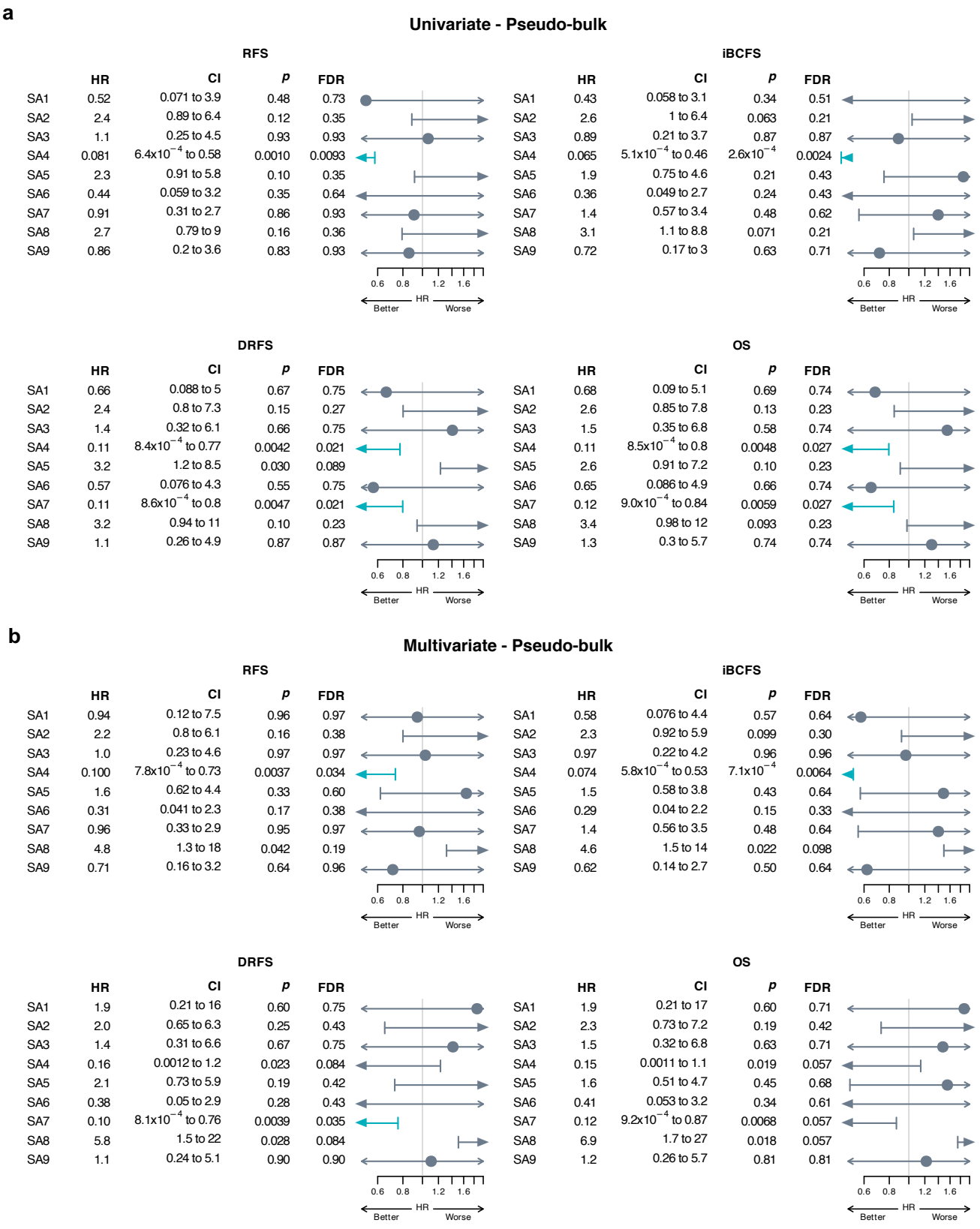
Comparison of expression levels of genes of interest across the nine spatial archetypes in the ST TNBC (*N* = 94) and SCAN-B TNBC cohorts (*N* = 672).

Association between the expression levels of single genes of interest and spatial archetypes in the ST TNBC and SCAN-B TNBC cohorts. The left half-circle and the right half-circle represent ST TNBC global pseudobulk and SCAN-B TNBC cohorts, respectively.

A logistic regression model was used to evaluate associations between each cell type enrichment score and each spatial archetype. P values were obtained from Wilcoxon rank sum tests and corrected for multi testing. Dots are bordered and dark-colored when FDRs < 0.05, compared to lighter-colored dots when FDRs ≥ 0.05. Negative and positive associations are represented in blue and red, respectively.
Source data are provided as a Source Data file.

FDR: false-discovery rate; PB: global pseudobulk; SA: spatial archetype; ST: spatial transcriptomics.

Supplementary Figure 29



Prognostic value of the nine spatial archetypes in the ST TNBC cohort (*N* = 94) using global ST pseudobulk data.

a Forest plots of the 9 spatial archetypes for RFS, iBCFS, DRFS and OS, univariate analysis.

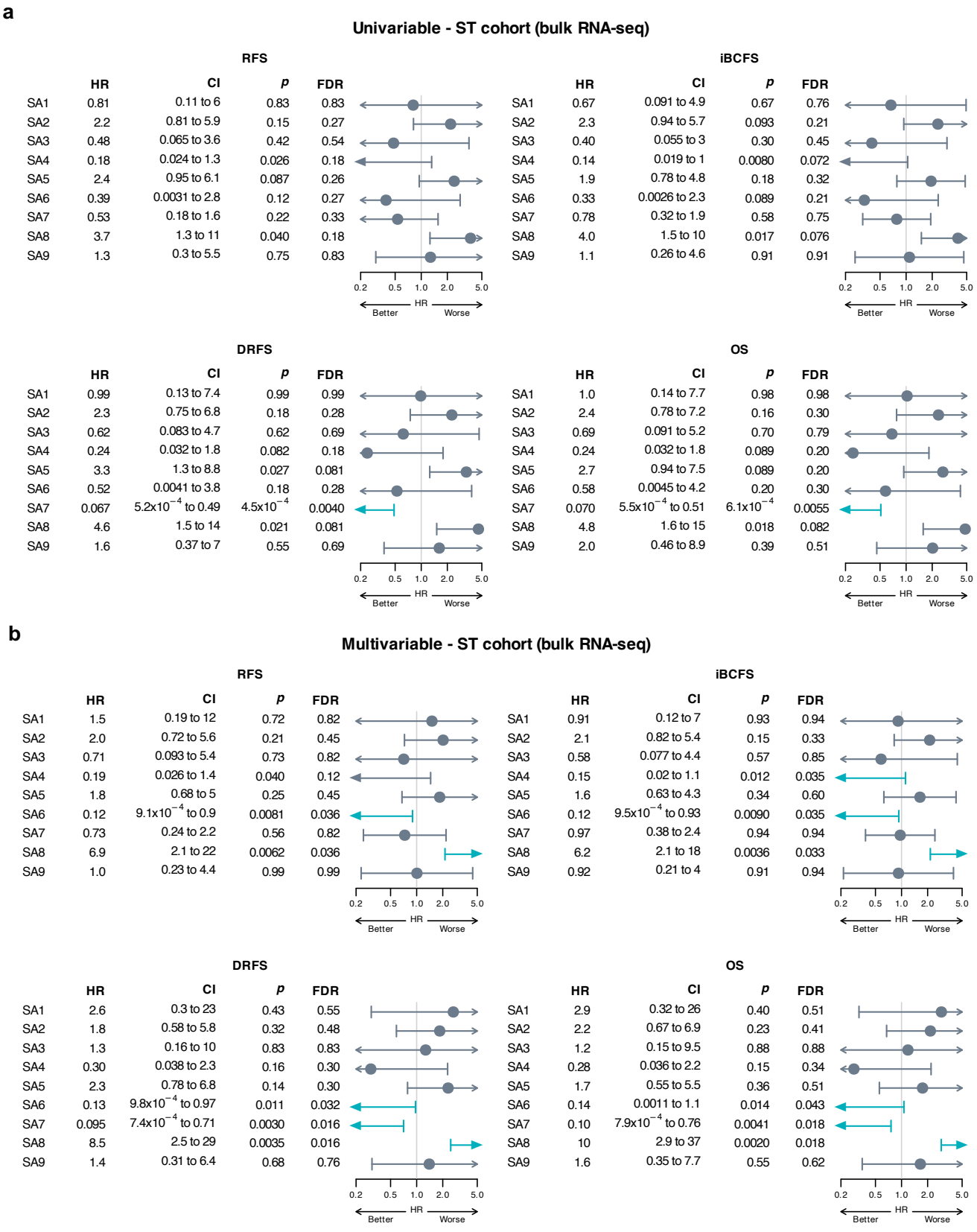
b Forest plots of the 9 spatial archetypes for RFS, iBCFS, DRFS and OS, multivariate analysis correcting for clinic-pathological parameters (age, tumor size, nodal status).

For univariate analysis, P values are from likelihood ratio test. When correcting for clinic-pathological characteristics, P values were obtained with a likelihood ratio test on nested Cox models. Significant

FDRs are shown in *blue* (< 0.05). Circles indicate HR, and error bars the 95% confidence interval (95% CI).
Source data are provided as a Source Data file.

CI: confidence interval; DRFS: distant relapse-free survival; FDR: false-discovery rate; HR: hazard ratio; iBCFS: invasive breast cancer free survival; OS: overall survival; RFS; recurrence-free survival; SA: spatial archetype; ST: spatial transcriptomics; TNBC: triple negative breast cancer.

Supplementary Figure 30



Prognostic value of the nine spatial archetypes in the ST TNBC cohort (*N* = 94) using bulk RNA-seq data.

a Forest plots of the 9 spatial archetypes for RFS, iBCFS, DRFS and OS, univariate analysis.

b Forest plots of the 9 spatial archetypes for RFS, iBCFS, DRFS and OS, multivariate analysis correcting for clinic-pathological parameters (age, tumor size, nodal status).

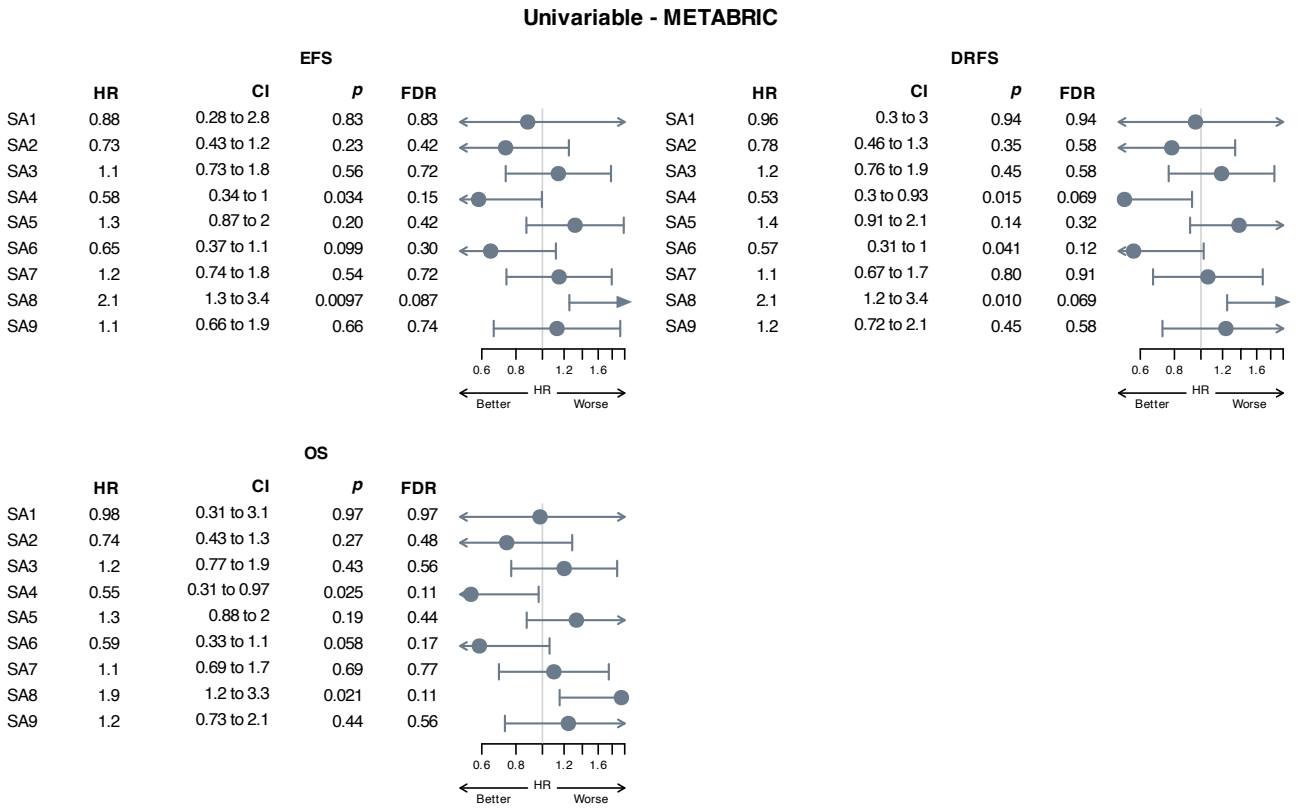
For univariate analysis, P values are from likelihood ratio test. When correcting for clinic-pathological characteristics, P values were obtained with a likelihood ratio test on nested Cox models. Significant

FDRs are shown in blue (< 0.05). Circles indicate HR, and error bars the 95% confidence interval (95% CI).
Source data are provided as a Source Data file.

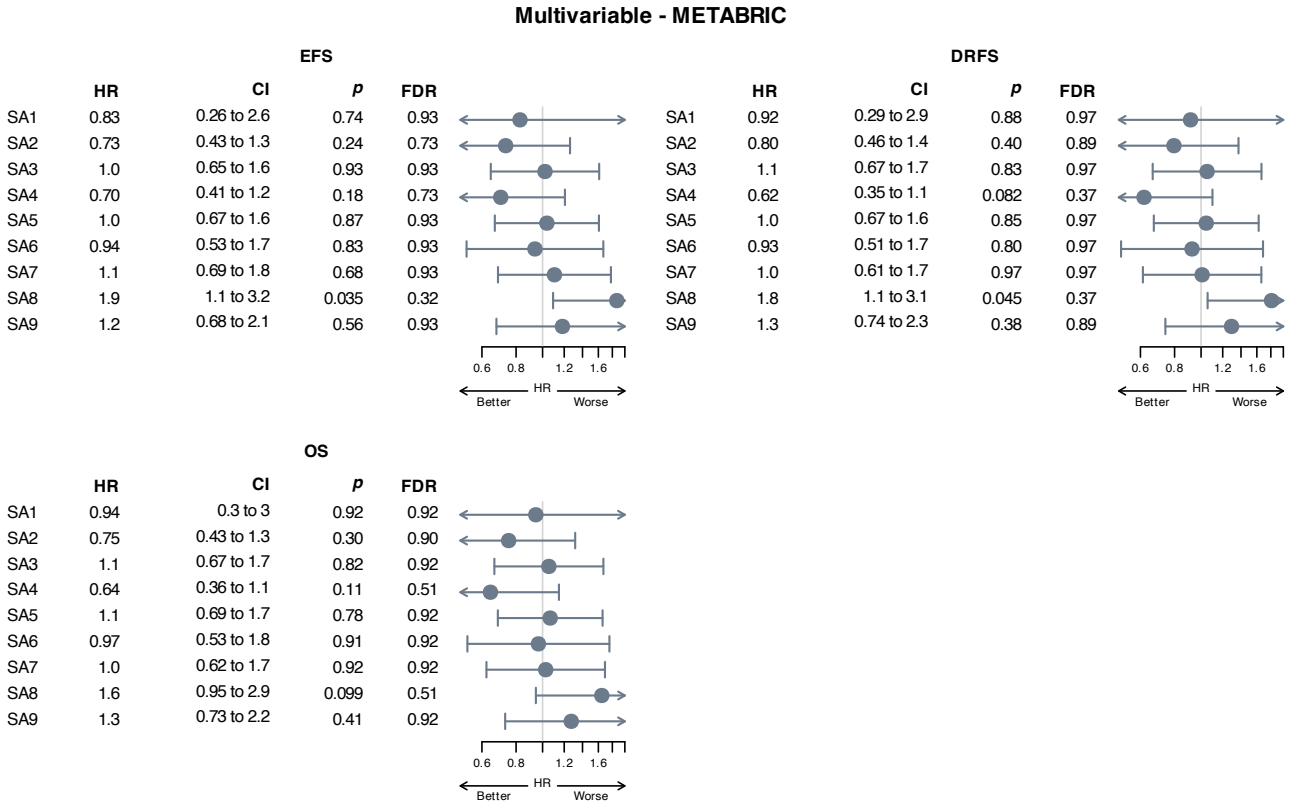
CI: confidence interval; DRFS: distant relapse-free survival; FDR: false-discovery rate; HR: hazard ratio; iBCFS: invasive breast cancer free survival; OS: overall survival; RFS: recurrence-free survival; SA: spatial archetype; ST: spatial transcriptomics; TNBC: triple negative breast cancer.

Supplementary Figure 31

a



b



Prognostic value of the nine spatial archetypes in the METABRIC TNBC cohort (*N* = 335) using bulk RNA-seq data.

a Forest plots of the 9 spatial archetypes for EFS, DRFS and OS, univariate analysis.

b Forest plots of the 9 spatial archetypes for EFS, DRFS and OS, multivariate analysis correcting for clinic-pathological parameters (age, tumor size, nodal status).

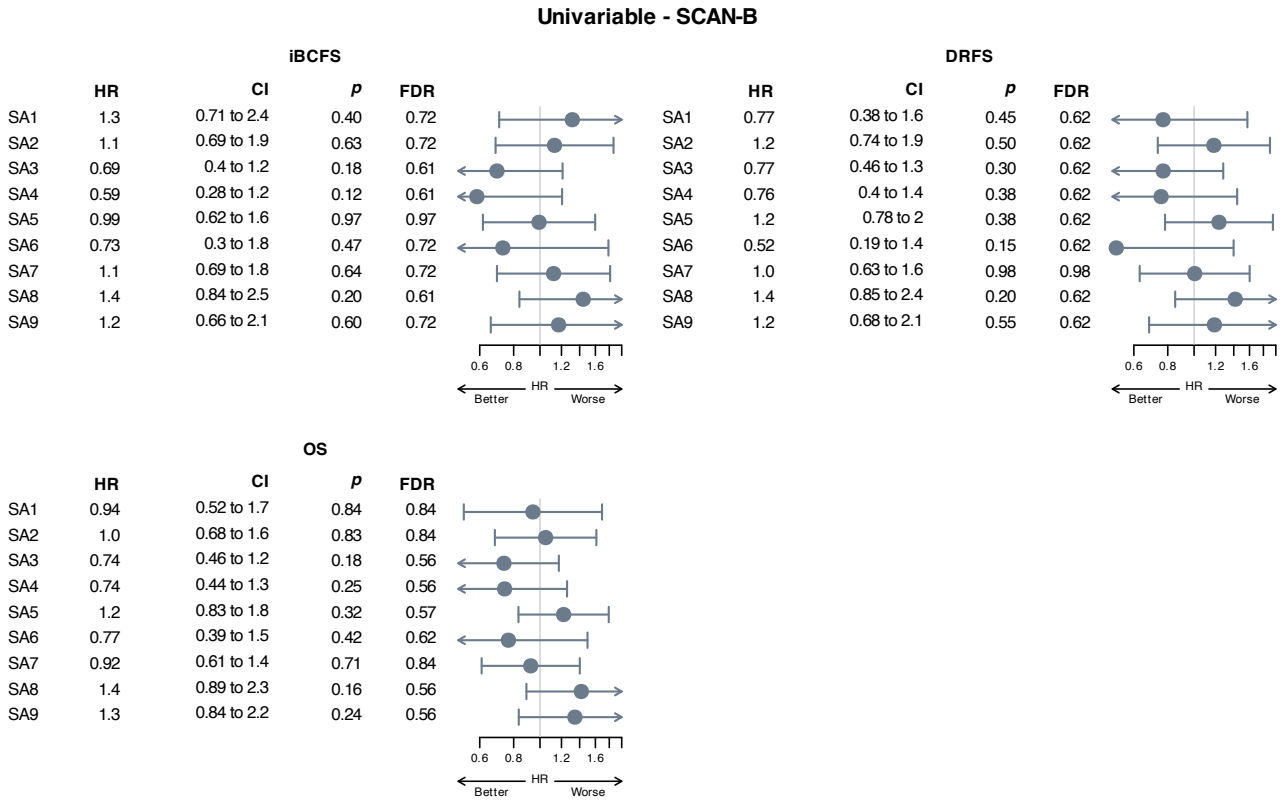
For univariate analysis, P values are from likelihood ratio test. When correcting for clinic-pathological characteristics, P values were obtained with a likelihood ratio test on nested Cox models. Significant

FDRs are shown in blue (< 0.05). Circles indicate HR, and error bars the 95% confidence interval (95% CI).
Source data are provided as a Source Data file.

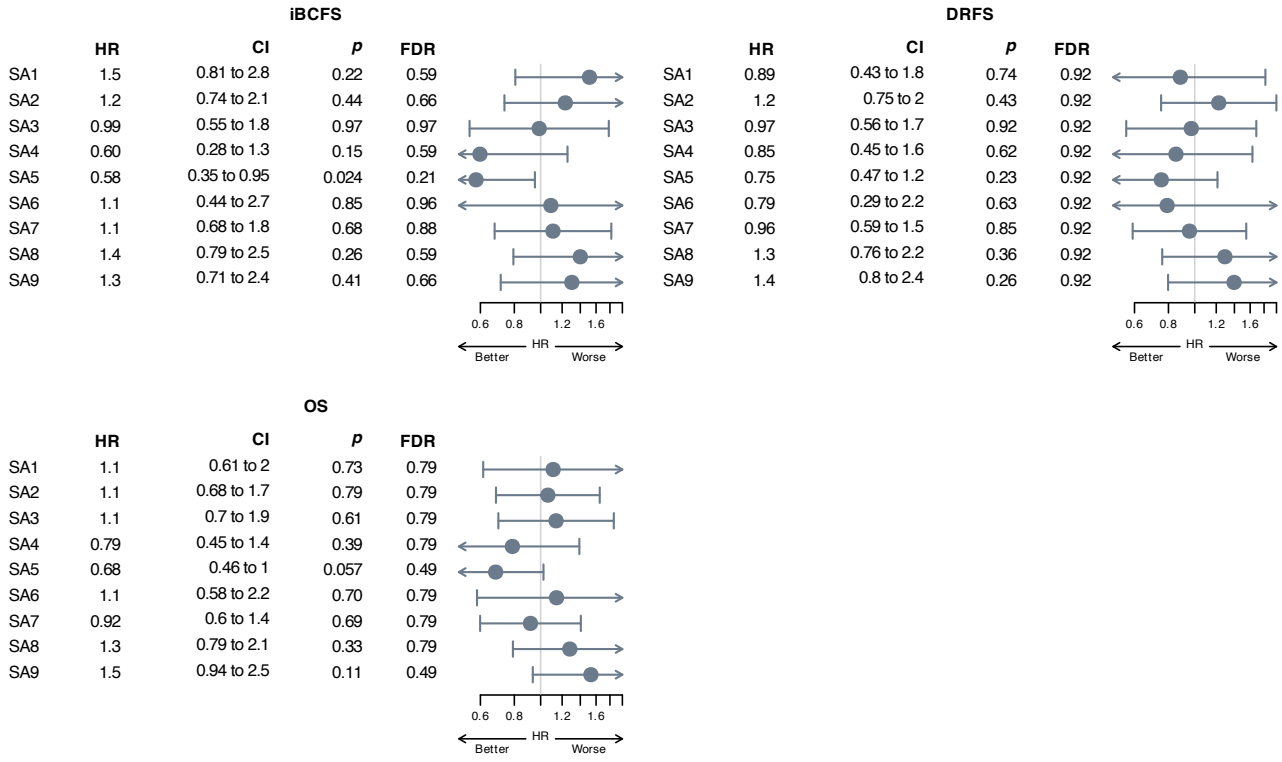
CI: confidence interval; DRFS: distant relapse-free survival; EFS: event-free survival; FDR: false-discovery rate; HR: hazard ratio; OS: overall survival; SA: spatial archetype.

Supplementary Figure 32

a



b



Prognostic value of the nine spatial archetypes in the SCAN-B TNBC cohort (*N* = 672) using bulk RNA-seq data.

a Forest plots of the 9 spatial archetypes for iBCFS, DRFS and OS, univariate analysis.

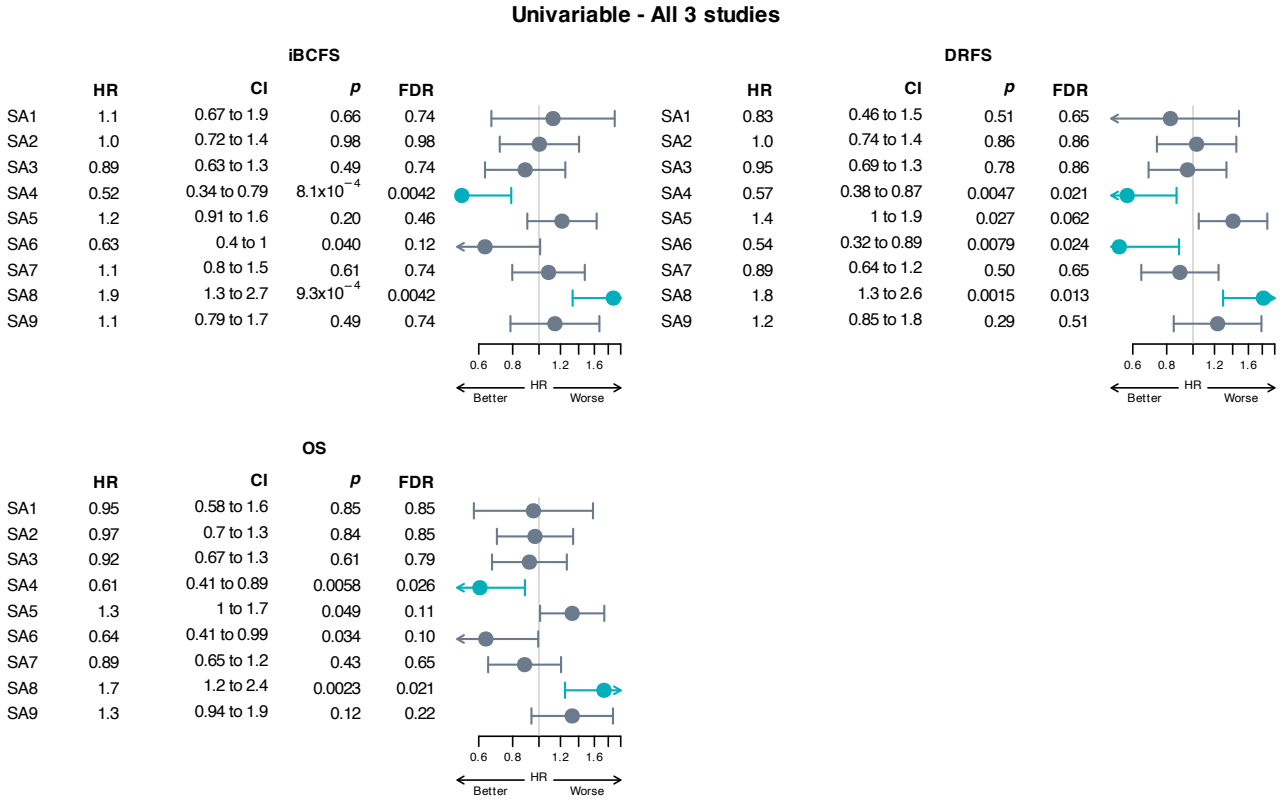
b Forest plots of the 9 spatial archetypes for iBCFS, DRFS and OS, multivariate analysis correcting for clinic-pathological parameters (age, tumor size, nodal status).

For univariate analysis, P values are from likelihood ratio test. When correcting for clinic-pathological characteristics, P values were obtained with a likelihood ratio test on nested Cox models. Significant FDRs are shown in *blue* (< 0.05). Circles indicate HR, and error bars the 95% confidence interval (95% CI).
Source data are provided as a Source Data file.

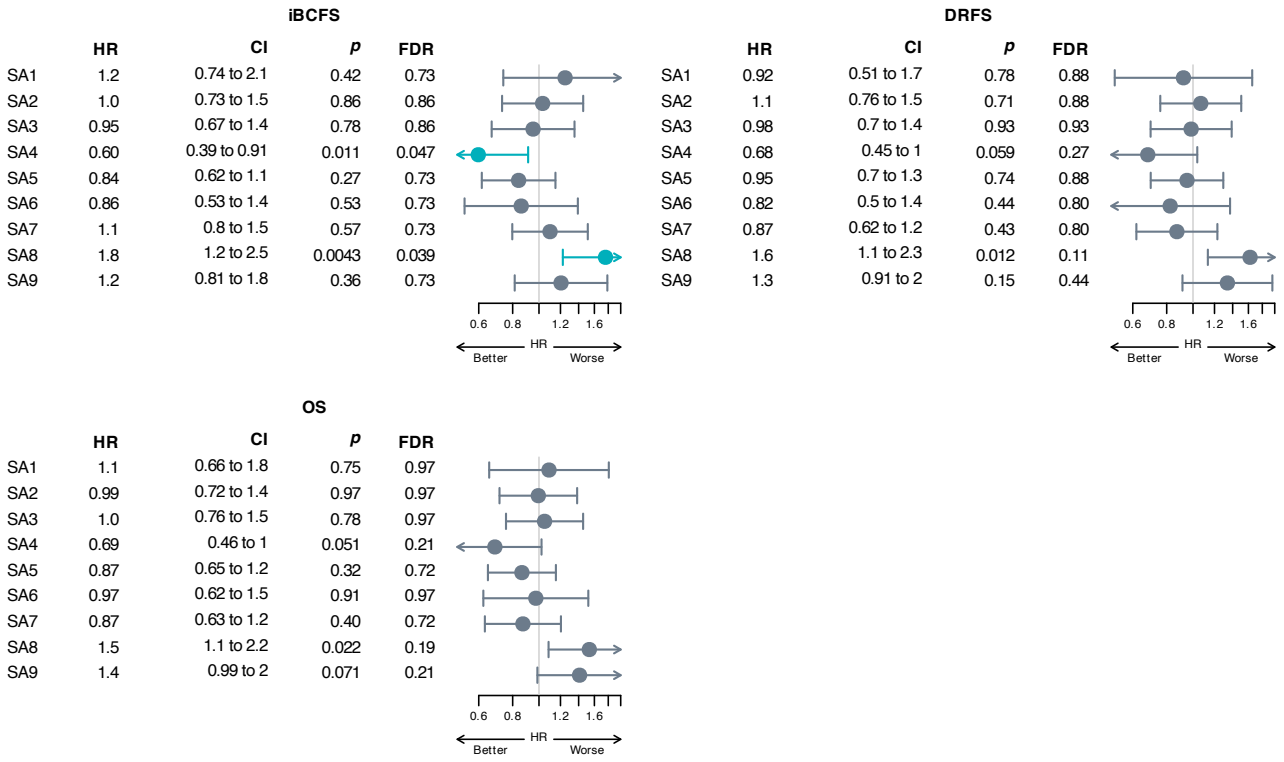
CI: confidence interval; DRFS: distant relapse-free survival; FDR: false-discovery rate; HR: hazard ratio; iBCFS: invasive breast cancer free survival; OS: overall survival; SA: spatial archetype.

Supplementary Figure 33

a



b



Prognostic value of the nine spatial archetypes in the merged ST TNBC, METABRIC and SCAN-B TNBC cohorts (*N* = 1101) using bulk RNA-seq data.

a Forest plots of the 9 spatial archetypes for iBCFS, DRFS and OS, univariate analysis.

b Forest plots of the 9 spatial archetypes for iBCFS, DRFS and OS, multivariate analysis correcting for clinic-pathological parameters (age, tumor size, nodal status).

The endpoint of iBCFS defined on ST TNBC and SCAN-B cohorts included also EFS from the METABRIC cohort. For univariate analysis, P values are from likelihood ratio test. When correcting

for clinic-pathological characteristics, P values were obtained with a likelihood ratio test on nested Cox models. All analysis were stratified by study. Significant FDRs are shown in *blue* (< 0.05). Circles indicate HR, and error bars the 95% confidence interval (95% CI).
Source data are provided as a Source Data file.
CI: confidence interval; DRFS: distant relapse-free survival; FDR: false-discovery rate; HR: hazard ratio; iBCFS: invasive breast cancer free survival; OS: overall survival; SA: spatial archetype.

SUPPLEMENTARY TABLES

Supplementary Table 1: Patients' clinic-pathological characteristics of the global ST TNBC cohort and across each TNBC molecular subtype.

	Total (*N* = 92)	BL (*N* = 17)	IM (*N* = 25)	LAR (*N* = 11)	M (*N* = 27)	MSL (*N* = 12)	P value
Age (years)							
<45	24 (26.1%)	3 (17.6%)	7 (28.0%)	2 (18.2%)	10 (37.0%)	2 (16.7%)	**0.596**[1]
≥45	68 (73.9%)	14 (82.4%)	18 (72.0%)	9 (81.8%)	17 (63.0%)	10 (83.3%)	
T size (cm)							
≤2	57 (62.0%)	8 (47.1%)	18 (72.0%)	6 (54.5%)	18 (66.7%)	7 (58.3%)	**0.503**[1]
>2	35 (38.0%)	9 (52.9%)	7 (28.0%)	5 (45.5%)	9 (33.3%)	5 (41.7%)	
Node status							
Negative	72 (78.3%)	13 (76.5%)	21 (84.0%)	6 (54.5%)	21 (77.8%)	11 (91.7%)	**0.293**[1]
Positive	20 (21.7%)	4 (23.5%)	4 (16.0%)	5 (45.5%)	6 (22.2%)	1 (8.3%)	
HER2+ status (IHC)							
0	86 (93.5%)	14 (82.4%)	25 (100.0%)	10 (90.9%)	26 (96.3%)	11 (91.7%)	**0.146**[1]
1+	6 (6.5%)	3 (17.6%)	0 (0.0%)	1 (9.1%)	1 (3.7%)	1 (8.3%)	
Grade							
I/II	8 (8.9%)	1 (5.9%)	0 (0.0%)	4 (36.4%)	0 (0.0%)	3 (25.0%)	**< 0.001**[1]
III	82 (91.1%)	16 (94.1%)	24 (100.0%)	7 (63.6%)	26 (100.0%)	9 (75.0%)	
Unknown	2	0	1	0	1	0	
Ki67 (%)							
Median	60.000	75.000	80.000	20.000	70.000	45.000	**< 0.001**[2]
Q1, Q3	40.000, 80.000	50.000, 90.000	60.000, 90.000	17.500, 30.000	40.000, 80.000	23.750, 62.500	
Histological type							
IDC	76 (82.6%)	15 (88.2%)	17 (68.0%)	9 (81.8%)	24 (88.9%)	11 (91.7%)	**0.013**[1]
ILC	2 (2.2%)	0 (0.0%)	0 (0.0%)	1 (9.1%)	1 (3.7%)	0 (0.0%)	
MED	8 (8.7%)	1 (5.9%)	7 (28.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
MET	2 (2.2%)	1 (5.9%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (8.3%)	
Other	4 (4.3%)	0 (0.0%)	1 (4.0%)	1 (9.1%)	2 (7.4%)	0 (0.0%)	
sTILs (%)							
Median	15.000	15.000	40.000	5.000	5.000	8.000	**< 0.001**[2]
Q1, Q3	5.000, 30.000	15.000, 20.000	30.000, 80.000	3.000, 10.000	0.000, 10.000	3.000, 20.000	
TIME by pathologist							
FI	17 (18.7%)	1 (5.9%)	15 (60.0%)	0 (0.0%)	0 (0.0%)	1 (9.1%)	**< 0.001**[1]
ID	23 (25.3%)	0 (0.0%)	0 (0.0%)	5 (45.5%)	13 (48.1%)	5 (45.5%)	
MR	18 (19.8%)	3 (17.6%)	2 (8.0%)	2 (18.2%)	8 (29.6%)	3 (27.3%)	

SR	33 (36.3%)	13 (76.5%)	8 (32.0%)	4 (36.4%)	6 (22.2%)	2 (18.2%)	
Unknown	1	0	0	0	0	1	
Germline mutation							
No	21 (58.3%)	1 (33.3%)	10 (71.4%)	3 (100.0%)	5 (41.7%)	2 (50.0%)	**0.280**[1]
BRCA1	13 (36.1%)	1 (33.3%)	4 (28.6%)	0 (0.0%)	6 (50.0%)	2 (50.0%)	
BRCA2	1 (2.8%)	1 (33.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
CHEK2	1 (2.8%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (8.3%)	0 (0.0%)	
Unknown	56	14	11	8	15	8	
Adjuvant chemotherapy							
No	11 (12.1%)	4 (23.5%)	2 (8.0%)	1 (10.0%)	3 (11.1%)	1 (8.3%)	**0.666**[1]
Yes	80 (87.9%)	13 (76.5%)	23 (92.0%)	9 (90.0%)	24 (88.9%)	11 (91.7%)	
Unknown	1	0	0	1	0	0	
Adjuvant radiotherapy							
No	3 (3.3%)	0 (0.0%)	2 (8.0%)	1 (10.0%)	0 (0.0%)	0 (0.0%)	**0.216**[1]
Yes	88 (96.7%)	17 (100.0%)	23 (92.0%)	9 (90.0%)	27 (100.0%)	12 (100.0%)	
Unknown	1	0	0	1	0	0	
Logoregional relapse							
No	81 (88.0%)	16 (94.1%)	23 (92.0%)	9 (81.8%)	23 (85.2%)	10 (83.3%)	**0.747**[1]
Yes	11 (12.0%)	1 (5.9%)	2 (8.0%)	2 (18.2%)	4 (14.8%)	2 (16.7%)	
Contralateral relapse							
No	85 (92.4%)	15 (88.2%)	24 (96.0%)	11 (100.0%)	24 (88.9%)	11 (91.7%)	**0.791**[1]
Yes	7 (7.6%)	2 (11.8%)	1 (4.0%)	0 (0.0%)	3 (11.1%)	1 (8.3%)	
Distant relapse							
No	79 (85.9%)	13 (76.5%)	24 (96.0%)	6 (54.5%)	25 (92.6%)	11 (91.7%)	**0.014**[1]
Yes	13 (14.1%)	4 (23.5%)	1 (4.0%)	5 (45.5%)	2 (7.4%)	1 (8.3%)	
Death							
No	71 (77.2%)	10 (58.8%)	23 (92.0%)	6 (54.5%)	22 (81.5%)	10 (83.3%)	**0.047**[1]
Yes	21 (22.8%)	7 (41.2%)	2 (8.0%)	5 (45.5%)	5 (18.5%)	2 (16.7%)	

[1]Fisher's Exact Test for Count Data with simulated P value (based on 2000 replicates).

[2]Kruskal-Wallis rank sum test.

BL: basal-like; FI: full inflamed; ID: immune desert; IDC: invasive ductal carcinoma; ILC: invasive lobular carcinoma; IM: immunomodulatory; LAR: luminal androgen receptor; M: mesenchymal; MED: medullary carcinoma; MET: metaplastic carcinoma; MSL: mesenchymal stem-like; MR: margin restricted; SR: stroma restricted; TIME: Tumor Immune Micro-Environment; ST: spatial transcriptomics; sTILs: stromal tumor infiltrating lymphocytes; TNBC: triple negative breast cancer.

Supplementary Table 2: Definition of the morphological annotations
(referred to **Figures 3a-c, Supplementary Figure 1)**

Fifteen Histomorphological Categories	Definition	Methods	Merged Histomorphological Categories
Tumor cells (1)	Tumor cells	Machine learning based on training on different tumor cells across different areas of the section	Tumor = (1) + (2)
Stroma cells (3)	Main non-immune cellular component of the stroma = fibroblasts	Machine learning based on training on different fibroblasts across different areas of the section	Stroma cells [A] = (3) + ½ x (4) + ½ x (5)
Lymphocytes (6)	Lymphocytes infiltrating the tumor (=TILs) but also lymphocyte distant from the tumor edge	Machine learning based on training on different lymphocytes across different areas of the section	Lymphocytes = (6) + ½ x (5)
Low TILs stroma (4)	Stroma including acellular stroma, stroma cells with estimated <30% lymphocytes on the area	Manual delineation by the pathologist when application by machine learning tool is not feasible	/
High TILs stroma (5)	Stroma including acellular stroma, stroma cells with estimated >=30% lymphocytes on the area	Manual delineation by the pathologist when application by machine learning tool is not feasible	/
Acellular stroma(7)	Stroma without cells enriched by collagen fibers	Manual delineation	Acellular stroma [B] = (7) + ½ x (4)
			Stroma or Total stroma = [A] + [B]
Fat tissue (8)	Fat tissue	Manual delineation	Fat tissue = (8)
Necrosis (9)	Necrosis which could be infiltrated by immune cells	Manual delineation	Necrosis = (9)
In situ (10)	Carcinoma in situ	Manual delineation	In situ = (10)
Tertiary lymphoid structures (11)	Tertiary lymphoid structure including a dense cellular aggregate (GC) or lymphoid aggregates and lymphoid follicles without GC and consisted of CD20-positive B zones with CD3-positive T zone aggregates	Manual delineation	TLS = (11)
Vessels (12)	Vessel	Manual delineation	Vessels = (12)
Lactiferous ducts (13)	Normal ductal glands including those with hyperplasia	Manual delineation	Lactiferous ducts = (13)
Nerves (14)	Nerve	Manual delineation	Nerves = (14)
Heterologous element (15)	Ectopic structures as bone, teeth…	Manual delineation	Heterologous element = (15)
Tumor region (2)	Area with majority of tumor cells	Manual delineation by the pathologist when application by machine learning tool is not feasible	/
2 Technical categories	**Definition**	**Methods**	
Whitespace	Whitespace without any histological structures or false intercellular spaces due to frozen artefact	Manual delineation	/
Artefact	Technical issues as tissue folds which should be excluded from the spot selection for the analysis	Manual delineation	/

Supplementary Table 3: Definition of the nine grouped histomorphological categories for the regression
(referred to **Supplementary Figure 2a)**

9 grouped categories	Definition
Tumor	Sum of Tumor cells and Tumor region annotations
Stroma	Total stroma = sum of Stroma cell, Low TIL stroma, acellular stroma and half of High TIL stroma
Necrosis	Necrosis
Fat tissue	Fat tissue
Vessels	Vessels
In situ	Carcinoma in situ
Tertiary lymphoid structures	Tertiary lymphoid structure including a dense cellular aggregate (GC) or lymphoid aggregates and lymphoid follicles without GC and consisted of CD20-positive B zones with CD3-positive T zone aggregates
Lymphocytes	Sum of Lymphocytes and half of High TIL stroma annotations
Lactiferous ducts	Normal ductal glands including those with hyperplasia

Supplementary Table 4: List of the expressed-based signatures

Class	Signature	Abbreviation	Source
Cellular component	Peroxisome	/	MSigDB Hallmarks v7.0
	Apical junction	/	MSigDB Hallmarks v7.0
	Apical surface	/	MSigDB Hallmarks v7.0
Development	Myogenesis	/	MSigDB Hallmarks v7.0
	Epithelial mesenchymal transition	/	MSigDB Hallmarks v7.0
	Angiogenesis	/	MSigDB Hallmarks v7.0
	Adipogenesis	/	MSigDB Hallmarks v7.0
Stroma	Stroma1	/	PMID_19122658
	Stroma2	/	PMID_18698033
	ecm-myCAF	Extracellular matrix myofibroblastic cancer associated fibroblast S1	PMID_32434947
	TGFbeta-myCAF	TGFbeta signaling pathway myofibroblastic cancer associated fibroblast S1	PMID_32434947
	wound-myCAF	Wound healing myofibroblastic cancer associated fibroblast S1	PMID_32434947
	detox-iCAF	Detoxification pathway inflammatory cancer associated fibroblast S1	PMID_32434947
	IL-iCAF	IL pathway inflammatory cancer associated fibroblast S1	PMID_32434947
	IFNgamma-iCAF	Interferon gamma signaling pathway cancer associated fibroblast S1	PMID_32434947
	Normal fibroblast	Normal fibroblast	PMID_32434947
	CAF	Cancer associated fibroblast S1	PMID_29455927
Immune	Interferon gamma response	/	MSigDB Hallmarks v7.0
	Interferon alpha response	/	MSigDB Hallmarks v7.0
	Inflammatory response	/	MSigDB Hallmarks v7.0
	IL6 JAK STAT3 signaling	/	MSigDB Hallmarks v7.0
	IL2 STAT5 signaling	/	MSigDB Hallmarks v7.0
	Complement	/	MSigDB Hallmarks v7.0
	Coagulation	/	MSigDB Hallmarks v7.0
	TNFA signaling via NF−kB	/	MSigDB Hallmarks v7.0
	Immune1	/	PMID 17683518
	Immune2	/	PMID_18698033
	TAM	Tumor associated macrophage	PMID_30930117
	Trm	Tissue-resident memory CD8 T cells	PMID_36026440
	TLS Cabrita	Tertiary lymphoid structures by Cabrita et al	PMID_31942071
	TLS Meylan	Tertiary lymphoid structures by Meylan et al	PMID_35231421
	TLS Lundeberg	Tertiary lymphoid structures by Lundeberg et al	PMID_34650042
	TLS ST	Tertiary lymphoid structures by spatial transcriptomics	In-house developed 30-gene TLS ST signature
Metabolic	Xenobiotic metabolism	/	MSigDB Hallmarks v7.0
	Oxidative phosphorylation	/	MSigDB Hallmarks v7.0

	Glycolysis	/	MSigDB Hallmarks v7.0
	Fatty acid metabolism	/	MSigDB Hallmarks v7.0
	Bile acid metabolism	/	MSigDB Hallmarks v7.0
	Heme metabolism	/	MSigDB Hallmarks v7.0
	Cholesterol homeostasis	/	MSigDB Hallmarks v7.0
Pathway	Protein secretion	/	MSigDB Hallmarks v7.0
	Hypoxia	/	MSigDB Hallmarks v7.0
	Apoptosis	/	MSigDB Hallmarks v7.0
	Unfolded protein response	/	MSigDB Hallmarks v7.0
	Reactive oxygen species pathway	/	MSigDB Hallmarks v7.0
	p53 pathway	/	MSigDB Hallmarks v7.0
Proliferation	MYC targets v2	/	MSigDB Hallmarks v7.0
	MYC targets v1	/	MSigDB Hallmarks v7.0
	Mitotic spindle	/	MSigDB Hallmarks v7.0
	E2F targets	/	MSigDB Hallmarks v7.0
	G2M checkpoint	/	MSigDB Hallmarks v7.0
	GGI	Genomic grade index	PMID_16478745
	CIN70	/	PMID_16921376
	GENE70	/	PMID_11823860
Signaling	WNT beta catenin signaling	/	MSigDB Hallmarks v7.0
	TGF beta signaling	/	MSigDB Hallmarks v7.0
	PI3K AKT mTOR signaling	/	MSigDB Hallmarks v7.0
	KRAS signaling	/	MSigDB Hallmarks v7.0
	Hedgehog signaling	/	MSigDB Hallmarks v7.0
	Estrogen response late	/	MSigDB Hallmarks v7.0
	Estrogen response early	/	MSigDB Hallmarks v7.0
	Androgen response	/	MSigDB Hallmarks v7.0
	mTORC1 signaling	/	MSigDB Hallmarks v7.0
	NOTCH signaling	/	MSigDB Hallmarks v7.0
	AR gene	/	single gene
	ESR1 gene	/	single gene
DNA damage	Parpi7	Parp inhibitor 7	PMID_35623341
	VCpred_TN	Veliparib carboplatin prediction triple negative	PMID_35623341
	DNA repair	/	MSigDB Hallmarks v7.0

Supplementary Table 5: List of the single genes of interest

Class	Gene names	Full Name	Role	Function in immune response
Immune	*ADA*	Adenosine deaminase	Catalyzation of hydrolysis of adenosine to inosine	positive immune regulation - adenosine pathway
	ADORA2A	Adenosine A2A receptor	adenosine receptor	immunosuppression - adenosine pathway
	VISTA	V-domain containing Ig Suppressor of T cell activation	Immunomudulatory ligand	T cell co-signaling pathway -negative immune regulation
	CCL5	CC-chemokine ligand 5	chemokine-ligand	mixed leukocyte recruitment
	CCR4	CC-chemokine receptor 4	chemokine-receptor	NK cell, T cell recruitment
	CCR8	CC-chemokine receptor 8	chemokine-receptor	lymphocyte and DC homing
	CD134	TNF-receptor superfamily member 4	immune checkpoint stimulatory receptor	positive immune regulation
	CD137	Cluster of differentiation 137	immune checkpoint stimulatory receptor	positive immune regulation
	CD27	TNF-receptor superfamily member 7	immune checkpoint stimulatory receptor	positive immune regulation
	PD-L1	Programmed Death-1 ligand	Immunomudulatory ligand	T cell co-signaling pathway -negative immune regulation
	B7-H3	B7 superfamily molecule	Immunomudulatory ligand	T cell co-signaling pathway -negative immune regulation
	CD40	TNF-receptor superfamily member 5	immune checkpoint stimulatory receptor	positive immune regulation
	CD40L	Cluster of differentiation 40-ligand	Immunomudulatory ligand	T and B cell co-signaling pathway - positive immune regulation
	CD47	integrin-associated protein	Immunomudulatory ligand	negative immune regulation of myeloid cells
	CD96	tactile	immune checkpoint inhibitory receptor	negative immune regulation of NK cells and T cells
	CSF1R	Colony stimulating factor 1 receptor	cytokine receptor	production, differenciation and function of macrophages
	CTLA4	Cytotoxic T-lymphocyte-associated molecule 4	immune checkpoint inhibitory receptor	T cell co-signaling pathway -negative immune regulation
	CXCL12	C-X-C chemokine ligand-12 (stromal cell-derived factor 1)	chemokine-ligand	mixed leukocyte recruitment
	CXCR4	C-X-C chemokine receptor type 4	chemokine-receptor	mixed leukocyte recruitment
	DCIR	Dendritic Cell Immuno Receptor	immune checkpoint inhibitory receptor	innate immune system - negative regulation
	CD39	Cluster of differentiation 39	ectoenzyme generating AMP	immunosuppression - adenosine pathway
	FOXP3	Forkhead box P3	Transcription factor	Treg marker
	GITR	Glucocorticoid-Induced TNF Receptor	TNF receptor superfamily	Negative immune regulation of Treg
	TIM3	T-cell Ig and Mucin-domain-Molecule 3	immune checkpoint inhibitory receptor	T cell co-signaling pathway -negative immune regulation
	B7-H7	B7 superfamily molecule	Immunomudulatory ligand	T cell co-signaling pathway -negative immune regulation
	ICOS	Inducible T-cell Co-stimulator	Immunomudulatory ligand	T cell co-signaling pathway - positive immune regulation
	IDO1	Indoleamine 2,3-dioxygenase	enzyme degrading tryptophan	inflammation and immunosuppresion

	IFNAR1	Interferon alpha and beta receptor 1	cytokine receptor	inflammation and immunosuppresion
	IFNAR2	Interferon alpha and beta receptor 2	cytokine receptor	inflammation and immunosuppresion
	IFNGR1	Interferon gamma receptor 1	cytokine receptor	inflammation and immunosuppresion
	IL2RA	Interleukin-2 receptor	cytokine receptor	positive immune regulation
	NKG2A	Natural killer G2 A receptor	NK cell inhibiotry receptor	innate immune system - negative regulation of NK cells
	NKG2D	Natural killer G2 D receptor	NK cell activating receptor	innate immune system -positive regulation of NK cells
	LAG3	Lymphocyte activating gene 3	immune checkpoint inhibitory receptor	negative immune regulation
	CD73	Cluster of differentiation 73	ectoenzyme generating adenosine	immunosuppression - adenosine pathway
	PD1	Programmed Death-1	immune checkpoint inhibitory receptor	T cell co-signaling pathway -negative immune regulation
	PIK3CG	Phosphoinositide 3-kinase -gamma	enzyme involved in cellular function	macrophages differenciation
	TIGIT	T cell Immunoglobulin and ITIM domain	immune checkpoint inhibitory receptor	T cell co-signaling pathway -negative immune regulation
	TLR2	Toll-Like Receptor 2	Toll-like receptor	Toll-like receptor signaling pathway
	TLR3	Toll-Like Receptor 3	Toll-like receptor	Toll-like receptor signaling pathway
	TLR7	Toll-Like Receptor 7	Toll-like receptor	Toll-like receptor signaling pathway
	TLR9	Toll-Like Receptor 9	Toll-like receptor	Toll-like receptor signaling Pathway
	B7-H4	B7 superfamily molecule	Immunomudulatory ligand	T cell co-signaling pathway -negative immune regulation
	KIR3DL1	Killer Immunoglobulin-like receptor	NK cell inhibiotry receptor	innate immune system- negative regulation of Nk cells
	GARP	Glycoprotein A repitions predominant	transmembrane protein	Treg marker
DNA mismatch repair	*MLH1*	MutL Homolog 1	DNA mismatch repair	
	MLH3	MutL Homolog 3	DNA mismatch repair	
	MSH2	MutS Homolog 2	DNA mismatch repair	
	MSH3	MutS Homolog 3	DNA mismatch repair	
	MSH6	MutS Homolog 6	DNA mismatch repair	
	PMS1	PMS1 Homolog 1	DNA mismatch repair	
	PMS2	PMS1 Homolog 2	DNA mismatch repair	
	POLH	DNA Polymerase Eta	DNA polymerase specifically involved in the DNA repair by translesion synthesis	
Targetable antigens	*GPNMB*	Transmembrane glycoprotein NMB	promotion of cell migration, invasion, and metastasis	
	NECTIN4	Nectin Cell Adhesion Molecule 4	Cell adhesion	
	Trop-2	Tumor-associated calcium signal transducer 2	Cell surface receptor that transduces calcium signals	
	5T4	Trophoblast Glycoprotein	Cell adhesion and antagonist of Wnt/β-catenin signalling pathway	

	LIV-1	Solute Carrier Family 39 Member 6	Control of gene transcription, growth, development, and differentiation	
	PSMA	Zinc metalloenzyme		
Signaling pathway - EGF	*EGFR*	Epidermal Growth Factor Receptor	Epidermal growth factor signalling pathway including cellular growth	
	ERBB2	Erb-B2 Receptor Tyrosine Kinase 2	Epidermal growth factor signalling pathway including cellular growth	
	ERBB3	Erb-B2 Receptor Tyrosine Kinase 3	Epidermal growth factor signalling pathway including cellular growth	
Signaling pathway - AKT/PI3K	*AKT1*	AKT Serine/Threonine Kinase 1	AKT/PI3K related signalling pathways: cell proliferation, survival, metabolism, and angiogenesis	
	mTOR	Mechanistic Target Of Rapamycin Kinase	AKT/PI3K pathway: cellular responses to stresses (DNA damage) and promotion of cell survival and cell cycle progression	
	PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha	AKT/PI3K related signalling pathways: cell proliferation, survival, metabolism, and angiogenesis	
Signaling pathway - PDGF/VEGF	*PDGFRA*	Platelet derived growth factor receptor alpha	PDGF/VEGF signaling pathways including cellular growth and differentiation	
	PDGFRB	Platelet derived growth factor receptor beta	PDGF/VEGF signaling pathways including cellular growth and differentiation	
	PDGFC	Platelet-derived growth factor C	PDGF/VEGF signaling pathways including cellular growth and differentiation	
Signaling pathway - MAPK	*NTRK1*	Neurotrophic Receptor Tyrosine Kinase 1	MAPK pathway including cell differentiation	
	NTRK2	Neurotrophic Receptor Tyrosine Kinase 2	MAPK pathway including cell differentiation	
	NTRK3	Neurotrophic Receptor Tyrosine Kinase 3	MAPK pathway including cell differentiation	
Signaling pathway - Coagulation	*TF*	Coagulation Factor III (Tissue factor)	Coagulation pathway: normal hemostasis by initiating the cell-surface assembly and propagation of the coagulation protease cascade	
Methylation	*DNMT1*	DNA methyltransferase 1	DNA methylation	
	TRDMT1	TRNA Aspartic Acid Methyltransferase 1	DNA (cytosine-5-)-methyltransferase activity	
	DNMT3A	DNA Methyltransferase 3 Alpha	DNA de novo methylation	
	DNMT3B	DNA Methyltransferase 3 Beta	DNA de novo methylation	
	HDAC1	Histone Deacetylase 1	Catalyzation of acetyl group removal from lysine residues in histones and non-histone proteins, causing transcriptional repression	

Angiogenesis	*VEGFA*	Vascular endothelial growth factor A	PDGF/VEGF signaling pathways including angiogenesis	
	VEGFB	Vascular Endothelial Growth Factor B	PDGF/VEGF signaling pathways including angiogenesis	
	KDR	Kinase Insert Domain Receptor (VEGFR2)	VEGF family receptor for angiogenesis	
	FLT1	Fms Related Receptor Tyrosine Kinase 1 (VEGFR1)	VEGF family receptor for angiogenesis	
	FLT4	Fms Related Receptor Tyrosine Kinase 4 (VEGFR3)	VEGF family receptor for angiogenesis	
Cycling	*CCND1*	Cyclin D1	Cell cycle progression by forming a complex with CDK4 or CDK6 (G1/S transition)	
	CCNE1	Cyclin E1	Cell cycle progression by forming a complex with CDK2 (G1/S transition)	
	CDK2	Cyclin Dependent Kinase 2	Cell cycle progression by forming a complex with cyclin A or E, CDK inhibitor p21Cip1 (CDKN1A), and p27Kip1 (CDKN1B)	
	CDK4	Cyclin Dependent Kinase 4	Cell cycle progression (G1 phase) and phosphorylation of retinoblastoma gene product (Rb)	
	CDK6	Cyclin Dependent Kinase 6	Cell cycle progression (G1 phase) and phosphorylation of retinoblastoma gene product (Rb)	
	CDKN2A	Cyclin Dependent Kinase Inhibitor 2A	Cell cycle inhibition as inhibitors of CDK4 kinase and degradation of p53	
	CDKN2C	Cyclin Dependent Kinase Inhibitor 2C	Cell cycle inhibition as inhibitors of CDK4 and CDK6 kinases	
	MYC	MYC Proto-Oncogene	Cell cycle progression by acting as transcription factor and angiogenesis	
	RB1	RB Transcriptional Corepressor 1	Cell cycle inhibition by binding to E2F (inactive)	
	TTK	Dual Specificity Protein Kinase TTK	Cell cycle progression as mitotic checkpoint protein for accurate segregation of chromosomes during mitosis	
	PLK4	Polo Like Kinase 4	Cell cycle progression by regulating centriole duplication	
Hormonal	*AR*	Androgen Receptor	Androgen-induced cell proliferation	
	ESR1	Estrogen Receptor 1	Estrogen-induced cell proliferation	
	PGR	Progesterone Receptor	Progesterone-induced cell proliferation	
HRD	*ATM*	Ataxia Telangiectasia Mutated	Activation of checkpoint signaling upon double strand breaks	
	ATR	Ataxia Telangiectasia And Rad3 Related	Activation of checkpoint signaling upon double strand breaks	

	BARD1	BRCA1 Associated RING Domain 1	Interaction with the N-terminal region of BRCA1	
	BRCA1	BRCA1 DNA Repair Associated	Homologous recombination pathway for double-strand DNA repair	
	BRCA2	BRCA2 DNA Repair Associated	Homologous recombination pathway for double-strand DNA repair	
	BRIP1	BRCA1 Interacting Helicase 1	Interaction with the BRCT repeats of BRCA1	
	CDK12	Cyclin Dependent Kinase 12	Regulator of transcription elongation by phosphorylating RNA polymerase II and genes involving in DNA repair	
	CHEK2	Checkpoint Kinase 2	Stabilization of p53 leading to cell cycle arrest in G1 and interaction with BRCA1 to restore survival after DNA damage	
	EMSY	EMSY Transcriptional Repressor	Interaction with BRCA2	
	ABRAXAS1	Abraxas 1	Interaction with C-terminal repeats of BRCA1	
	FANCA	FA Complementation Group A	DNA postreplication repair, cell cycle checkpoint function	
	FANCM	FA Complementation Group M	Fanconi anemia complementation group	
	NBN	Nibrin	Member of the MRE11/RAD50 double-strand break repair complex	
	PALB2	Partner And Localizer Of BRCA2	Interaction with BRCA2	
	PTEN	Phosphatase And Tensin Homolog	Tumor suppression by negatively regulating AKT/PKB signaling pathway	
	RAD50	RAD50 Double Strand Break Repair Protein	Component of the MRN complex for double-strand break (DSB) repair, DNA recombination, maintenance of telomere	
	RAD51C	RAD51 Paralog C	Member of the RAD51 family involved in Holliday junction resolution	
	RAD51D	RAD51 Paralog D	Member of the RAD51 family	
	TP53	Tumor Protein P53	Cell cycle inhibition, apoptosis, senescence, DNA repair, or changes in metabolism	

Supplementary Table 6: List of cell types scored by xCell deconvolution tool

Class	Cell types	Full name
Lymphocytes B	B-cells	/
	Class-switched memory B-cells	/
	Memory B-cells	/
	naive B-cells	/
	pro B-cells	/
	Plasma cells	/
Lymphocytes T CD4+	CD4+ T-cells	/
	CD4+ memory T-cells	/
	CD4+ naive T-cells	/
	CD4+ Tcm	CD4+ central memory T cells
	CD4+ Tem	CD4+ effector memory T cells
	Th1 cells	Lymphocytes T helper 1
	Th2 cells	Lymphocytes T helper 2
	Tregs	Regulatory T cells
Lymphocytes T CD8+	CD8+ T-cells	/
	CD8+ naive T-cells	/
	CD8+ Tcm	CD8+ central memory T cells
	CD8+ Tem	CD8+ effector memory T cells
Other lymphocytes	NK cells	Natural killer cells
	NKT	Natural killer T cells
	Tgd cells	Lymphocytes T gamma delta
Myeloid cells	DC	Dendritic cells
	aDC	Activated dendritic cells
	cDC	Conventional dendritic cells
	iDC	Immature dendritic cells
	pDC	Plasmacytoid dendritic cells
	Eosinophils	/
	Basophils	/
	Neutrophils	/
	Macrophages	/
	Mast cells	/
	Monocytes	/
Stromal cells	Endothelial cells	/
	Fibroblasts	/
	Myocytes	/
	Smooth muscle	/
	Pericytes	/
Stem cells	CLP	Common Lymphoid Progenitor
	CMP	Common Myeloid Progenitor
	GMP	Granulocyte-macrophage progenitor
	MSC	Mesenchymal stem cell
	MPP	Multipotent progenitor
Other cells	Adipocytes	/
	Preadipocytes	/
	Epithelial cells	/
	Neurons	/

Supplementary Table 7: Selection of different molecular features based on the stroma (S) and tumor (T) compartments specificity
(referred to **Figure 4c, Supplementary Figures 3a and 3b**)

Signatures			Cell type enrichment scores by xCell			Single genes of interest		
Class	**Signatures**	**Compartment specificity [1]**	**Class**	**Cell types**	**Compartment specificity [1]**	**Class**	**Genes**	**Compartment specificity [1]**
Cellular component	Peroxisome	T	**Lympho B**	B-cells	S	**Immune**	*ADA*	S
	Apical junction	S		Class-switched memory B-cells	S		*ADORA2A*	S
	Apical surface	S		Memory B-cells	S		*VISTA*	S
Development	Myogenesis	S		naive B-cells	S		*CCL5*	S
	Epithelial mesenchymal transition	T		pro B-cells	S		*CCR4*	S
	Angiogenesis	S		Plasma cells	S		*CCR8*	S
	Adipogenesis	S	**Lympho T CD4+**	CD4+ T-cells	S		*CD134*	S
Stroma	Stroma1	S		CD4+ memory T-cells	S		*CD137*	S
	Stroma2	S		CD4+ naive T-cells	S		*CD27*	S
	ecm-myCAF	S		CD4+ Tcm	S		*PD-L1*	S
	TGFbeta-myCAF	S		CD4+ Tem	S		*B7-H3*	S
	wound-myCAF	S		Th1 cells	S		*CD40*	S
	detox-iCAF	S		Th2 cells	S		*CD40L*	S
	IL-iCAF	S		Tregs	S		*CD47*	S
	IFNgamma-iCAF	S	**Lympho T CD8+**	CD8+ T-cells	S		*CD96*	S
	Normal fibroblast	S		CD8+ naive T-cells	S		*CSF1R*	S
	CAF	S		CD8+ Tcm	S		*CTLA4*	S
Immune	Interferon gamma response	S		CD8+ Tem	S		*CXCL12*	S
	Interferon alpha response	S	**Other lympho**	NK cells	S		*CXCR4*	S
	Inflammatory response	S		NKT	S		*DCIR*	S
	IL6 JAK STAT3 signaling	S		Tgd cells	S		*CD39*	S
	IL2 STAT5 signaling	S	**Myeloid cells**	DC	S		*FOXP3*	S
	Complement	S		aDC	S		*GITR*	S
	Coagulation	S		cDC	S		*TIM3*	S
	TNFA signaling via NF−kB	S		iDC	S		*B7-H7*	S
	Immune1	S		pDC	S		*ICOS*	S
	Immune2	S		Eosinophils	S		*IDO1*	S
	TAM	S		Basophils	S		*IFNAR1*	S
	Trm	S		Neutrophils	S		*IFNAR2*	S
	TLS Cabrita	S		Macrophages	S		*IFNGR1*	S
	TLS Meylan	S		Mast cells	S		*IL2RA*	S
	TLS Lundeberg	S		Monocytes	S		*NKG2A*	S

	TLS ST	S
Metabolic	Xenobiotic metabolism	T
	Oxidative phosphorylation	T
	Glycolysis	T
	Fatty acid metabolism	T
	Bile acid metabolism	T
	Heme metabolism	T
	Cholesterol homeostasis	T
Pathway	Protein secretion	T
	Hypoxia	S
	Apoptosis	S
	Unfolded protein response	T
	Reactive oxygen species pathway	T
	p53 pathway	T
Proliferation	MYC targets v2	T
	MYC targets v1	T
	Mitotic spindle	T
	E2F targets	T
	G2M checkpoint	T
	GGI	T
	CIN70	T
	GENE70	T
Signaling	WNT beta catenin signaling	T and S
	TGF beta signaling	T and S
	PI3K AKT mTOR signaling	T
	KRAS signaling	T
	Hedgehog signaling	T and S
	Estrogen response late	T
	Estrogen response early	T
	Androgen response	T
	mTORC1 signaling	T
	NOTCH signaling	T and S
	AR gene	T
	ESR1 gene	T
DNA damage	Parpi7	T
	VCpred_TN	T
	DNA repair	T

Stromal cells	Endothelial cells	S
	Fibroblasts	S
	Myocytes	S
	Smooth muscle	S
	Pericytes	S
Stem cells	CLP	S
	CMP	S
	GMP	S
	MSC	S
	MPP	S
Other cells	Adipocytes	S
	Preadipocytes	S
	Epithelial cells	T
	Neurons	S

	NKG2D	S
	LAG3	S
	CD73	S
	PD1	S
	PIK3CG	S
	TIGIT	S
	TLR2	S
	TLR3	S
	TLR7	S
	TLR9	S
	B7-H4	S
	KIR3DL1	S
	GARP	S
DNA mismatch repair	*MLH1*	T
	MLH3	T
	MSH2	T
	MSH3	T
	MSH6	T
	PMS1	T
	PMS2	T
	POLH	T
Targetable antigens	*GPNMB*	T
	NECTIN 4	T
	Trop-2	T
	5T4	T
	LIV-1	T
	PSMA	T
Signaling pathway - EGF	*EGFR*	T
	ERBB2	T
	ERBB3	T
Signaling pathway - AKT/PI3K	*AKT1*	T
	mTOR	T
	PIK3CA	T
Signaling pathway - PDGF	*PDGFRA*	T and S
	PDGFRB	T and S
	PDGFC	T and S
Signaling pathway - MAPK	*NTRK1*	T
	NTRK2	T
	NTRK3	T
Signaling pathway – Coagulation	*TF*	T

Methylation	*DNMT1*	T
	TRDMT1	T
	DNMT3A	T
	DNMT3B	T
	HDAC1	T
Angiogenesis	*VEGFA*	S
	VEGFB	S
	KDR	S
	FLT1	S
	FLT4	S
Cycling	*CCND1*	T
	CCNE1	T
	CDK2	T
	CDK4	T
	CDK6	T
	CDKN2A	T
	CDKN2C	T
	MYC	T
	RB1	T
	TTK	T
	PLK4	T
Hormonal	*AR*	T
	ESR1	T
	PGR	T
HRD	*ATM*	T
	ATR	T
	BARD1	T
	BRCA1	T
	BRCA2	T
	BRIP1	T
	CDK12	T
	CHEK2	T
	EMSY	T
	ABRAXAS1	T
	FANCA	T
	FANCM	T
	NBN	T
	PALB2	T
	PTEN	T
	RAD50	T
	RAD51C	T
	RAD51D	T
	TP53	T

[1] T = tumor compartment; S = stroma compartment as defined by the regressor (see METHODS: *Stroma compartment was defined as the sum of all categories, except tumor, in situ and necrosis categories*).

Supplementary Table 8: Morphological and molecular characterizations of different molecular subtypes defined on the tumor compartment (or tumor pseudobulk, PB)
(Referred to **Figure 4c**, tumor compartment; **Supplementary Figure 3a**)

Tumor compartment/ pseudobulk	Annotations [1] [2]		Signature [1]		Single genes [1]		Summary
	↑	↓	↑	↓	↑	↓	
LAR	Lactiferous duct, in situ, Total stroma, acellular stroma	Tumor	mTORC1 sig, AR gene, androgen response, estrogen response early/late, PI3K/AKT/mTOR signaling, protein secretion, unfolded protein response, reactive oxygen species path, p53 pathway, Heme/Bile acid/fatty acid/xenobiotic metabolisms, oxydative phosphorylation, cholesterol homeostasis, glycolysis, peroxisome	MYC targets v2, mitotic spindle, E2F target, GENE70, GGI, CIN70, G2M checkpoint, hedgehog signaling, Parpi7	AR, AKT1	POLH, PSMA, PIK3CA, NTRK2/3, TF, HDAC1, DNMT3A, TRDMT1, DNMT1, CCNE1, CDKN2A/C, MYC, PLK4, BRCA2, CHEK2, FANCA, RAD51D/C	Metabolism PI3K/AKT/ mTOR sig
M	/	Lymphocytes	EMT, WNT beta catenin signaling, Parpi7	PI3K/AKT/mTOR sig, mTORC1 sig, DNA repair, ESR1/AR gene MYC targets v2, unfolded protein response, reactive oxygen species path, fatty acid/heme metabolism, oxydative phsophorylation, peroxisome	PDGFRA, NTRK2	ERBB3, AR, ESR1, RAD50	EMT *PDGFRA* gene
BL	Tumor, lymphocytes	Acellular stroma, total stroma	G2M, CIN70, GGI, E2F, Mitotic spindle, MYC targets v1/2, DNA repair	EMT, TGF beta signaling, estrogen response late, ESR1 gene	DNMT1, HDAC1, CCNE1, CDKN2A, MYC, TTK, PLK4 BARD1, BRCA1/2, CHEK2, FANCA, PALB2, RAD51D	ERBB2, ESR1	Proliferation DNA repair

[1] Only significant clinical and molecular features with FDRs <0.05 were reported
[2] Annotations -> morphological annotations represented the annotations for the whole section including stroma and tumor compartments

Supplementary Table 9: List of selected molecular features relevant to characterize TNBC molecular subtypes in the tumor compartment (referred to **Figure 4c** – tumor compartment)

Main function	Type of molecular feature	List of selected molecular features
Pathway	Signatures	Epithelial mesenchymal transition PI3K AKT mTOR signaling WNT beta catenin signaling Estrogen response late Androgen response mTORC1 signaling
	Single genes	*ESR1* *AR* *NTRK2* *PDGFRA* *AKT1* *DNMT1*
Metabolism	Signatures	Glycolysis Fatty acid metabolism Oxidative phosphorylation
Proliferation	Signatures	MYC targets v2 GGI G2M checkpoint Mitotic spindle
	Single genes	*MYC* *CCNE1* *CDKN2A* *PLK4*
Stress response	Signatures	DNA repair
	Single genes	*BRCA1* *BRCA2* *CHEK2* *PALB2*

Supplementary Table 10: Morphological and molecular characterizations of different molecular subtypes defined on the stroma compartment (or stroma pseudobulk, PB)
(Referred to **Figure 4c**, stroma compartment; **Supplementary Figure 3b**)

Stroma compart-ment/ pseudo-bulk	Annotations [1] [2]		Signature [1]		Cell type enrichment scores [1]		Single genes [1]		Summary
	↑	↓	↑	↓	↑	↓	↑	↓	
LAR	Total stroma, acelllular stroma, in situ	/	Adipogenesis	WNT beta catenin sig, immune 1/2, TLS Lundeberg, complement, IL6 sig, inflammatory response, IFNa/g response, Trm, TAM, IFNg-iCAF	/	CD4 naive, Th1, CD8 naive, all DCs, pericytes	/	Immune genes: ADA, VISTA, CD134, CD137, PD-L1, CD40, CSF1R, CTLA4, DCIR, FOXP3, TIM3, GITR, ICOS, IDO1, IFNAR2, IL2RA, LAG3, CD73, TLR2	Adipoge-nesis Immune-depletion
MSL	Vessels, lactiferou s duct, fat tissue, total stroma, stroma	Tumor	Stroma1/2, CAF, ecm-myCAF, TGFb-myCAF, wound-myCAF, IL-iCAF, detox-iCAF, normal fibroblast, TGFbeta sig, WNT beta catenin sig, Hedgehog, apoptosis, hypoxia, coagulation, angiogenesis, myogenesis, apical junction	/	Endothelial cells, fibroblasts, adipocytes, neurons	Th1/Th2, Treg, CD8 Tcm/Tem, macrophag es	B7-H3, CXCL12, CD73, PDGFRA/B, PDGFC, VEGFA/B, KDR, FLT1/4	LAG3, CD137	Angioge-nesis CAF
M	/	Lymphoid nodule (TLS), lactiferous duct, fat tissue, lymphocyt es CD20, CD3, sTILs	/	CAF, detox-iCAF, IL-CAF, IFNg-iCAF, normal fibroblast, apoptosis, TNF alpha sig, Immune 1/2, TLS lundeberg/ST, complement, IL2/6 sig, inflammatory response, IFNa/g res, Trm, TAM	Neutrophils , smooth muscle, GMP, neurons	All B cells (except naive B cells), CD4 naive/Tem, CD8 naive/Tcm/ wsqTem, DC/aDC/c DC/pDC, basophils, macrophag es	CD73, VEGFA	ADA, VISTA, CCL5, CD134, CD137, CD27, PD-L1, CD40, CD40L, CD47, CD96, CSF1R, CTLA4, DCIR, CD39, FOXP3, TIM3, ICOS, IDO1, LAG3, PD1, TIGIT, TLR3/7, IFNGR1	Immune-depletion Neutrophils
IM	Lymphoc ytes, tumor Medullar y, CD3/CD 20/sTILs, Ki67	In situ, total stroma, acellular stroma, stroma, vessels	Immune 1/2, TLS lundeberg/ST, complement, IL2/6 sig, inflammatory response, IFNa/g res, Trm, TAM, IFNg-iCAF	Stroma1/2, CAF, ecm-myCAF, TGFb-myCAF, wound-myCAF, detox-iCAF, normal fibroblast, TGFbeta sig, WNT beta catenin sig, Hedgehog sig, NOTCH sig, hypoxia, angiogenesis, myogenesis, apical junction	All B cells, plasma cells, Th1/2, CD4 naive/Tem, CD8 naive/Tcm/ Tem, all DCs, macrophag es, MPP, basophils	Endothelial cells, fibroblast, myocytes, smooth muscle, CMP, adipocytes, neurons	Immune genes: ADA, VISTA, CCL5, CD134, CD137, CD27, PD-L1, CD40, CD40L, CD47, CD96, CSF1R, CTLA4, CXCR4, DCIR, FOXP3, CD39, GITR, TIM3, ICOS, IDO1, IFNAR2, IL2RA, LAG3, PD1, TIGIT, TLR7	CXCL12, B7-H3, CD73, B7-H4 PDGFRA/B, PDGFC, VEGFA/B, KDR, FLT1/4	Immune-activation

[1] Only significant clinical and molecular features with FDRs <0.05 were reported
[2] Annotations -> morphological annotations represented the annotations for the whole section including stroma and tumor compartments

Supplementary Table 11: List of selected molecular features relevant to characterize TNBC molecular subtypes in the stroma compartment
(referred to **Figure 4c** – stroma compartment)

Main function	Type of molecular feature	List of selected molecular features
Immune	Signatures	TLS Lundeberg
		Interferon gamma response
		Interferon alpha response
	Single genes	*CD73*
		PD-L1
		PD1
		CTLA4
		LAG3
		B7-H3
		IDO1
		ICOS
		CXCL12
		CCL5
	Cell types	Neutrophils [1]
		Tregs [1]
		CD8+ T-cells [1]
		CD4+ naive T-cells [1]
		Plasma cells [1]
		Memory B-cells [1]
		Trm [2]
		TAM [2]
Stroma	Signatures	Angiogenesis
		Adipogenesis
	Single genes	*VEGFA*
		VEGFB
	Cell types	CAF [2]
		IFNgamma-iCAF [2]
		Smooth muscle [1]
		Neurons [1]
Pathway	Signatures	TGF beta signaling
		WNT beta catenin signaling
		Hedgehog signaling
	Single genes	*PDGFRA*
Stress response	Signatures	Hypoxia
		Apoptosis

[1] Computed by deconvolution Xcell
[2] Computed by signature

Supplementary Table 12: Definition of the clinical outcomes

(Referred to **Supplementary Data 1**: Clinic-pathological data of the ST TNBC cohort**)**

Clinical Outcome (1)	Name	Definition (2)							
			Homolateral breast relapse	Local-Regional invasive recurrence	Contralateral invasive breast cancer	Other primary cancer (non-breast cancer)	Distant recurrence	Death from breast cancer	Death from any cause
RFS	Recurrence-free survival	Defined as the time interval from the date of diagnosis to: (3)	X	X			X	X	X
iBCFS	Invasive breast cancer-free survival	Defined as the time interval from the date of diagnosis to:	X	X	X		X	X	X
iDFS	Invasive disease-free survival	Defined as the time interval from the date of diagnosis to:	X	X	X	X	X	X	X
DRFS	Distant relapse–free survival	Defined as the time interval from the date of diagnosis to:					X	X	X
OS	Overall survival	Defined as the time interval from the date of diagnosis to:						X	X

(1) Censoring at 10 years of follow up

(2) Definition based on STEEP version2.0 criteria[8]

(3) No censoring when contralateral relapse occurred

Supplementary Table 13: List of the 30 genes related to TLS ST signature (referred to **Figure 5d)**

Genes (1)	p_value	Fold-Change	Name	Category	Subcategory	Role
RIPOR2	3,2644E-12	2,04916914	RHO Family Interacting Cell Polarization Regulator 2	Immune	T cell	Lymphocyte T proliferation (inhibition of chemokine-induced T lymphocyte responses, such as cell adhesion, polarization and migration) and neutrophil polarization
IL16	1,2317E-11	2,06325624	Interleukin 16	Immune	Immune	Migratory response in CD4+ lymphocytes, monocytes, and eosinophils. Primes CD4+ T-cells for IL-2 and IL-15 responsiveness. Also induces T-lymphocyte expression of interleukin 2 receptor.
POU2AF1	5,5622E-11	2,11378752	POU Class 2 Homeobox Associating Factor 1	Nuclear/immune	Lymphoid nodule initation	Transcription coactivator activity essential for the response of B-cells to antigens and required for the formation of germinal centers, Regulates IL6 expression in B cells
SELL	1,1176E-14	2,16217582	Selectin L	Immune	Lymphoid nodule structure	Selectin L, adherence of lymphocytes to endothelial cells of high endothelial venules in peripheral lymph nodes
ATP2A3	5,0607E-10	2,22866006	ATPase Sarcoplasmic/Endoplasmic Reticulum Ca2+ Transporting 3	Cellular	Contraction	One of the SERCA Ca(2+)-ATPases, which are intracellular pumps located in the sarcoplasmic or endoplasmic reticula of muscle cells.
CD19	3,9592E-15	2,22866673	CD19 Molecule	Immune	B cell	Coreceptor for the B-cell antigen receptor complex (BCR) on B-lymphocytes
RAC2	2,4557E-10	2,24319162	Rac Family Small GTPase 2	Immune	Immune	Regulation of cellular responses, such as secretory processes, phagocytose of apoptotic cells and epithelial cell polarization
CXCL13	1,0856E-10	2,28695664	C-X-C Motif Chemokine Ligand 13	Immune	Lymphoid nodule initation	Chemotactic for B-lymphocytes but not for T-lymphocytes, monocytes and neutrophils; function in the homing of B lymphocytes to follicles
TCL1A	2,3557E-16	2,36279363	TCL1 Family AKT Coactivator A	Cellular	Cell survival	Coactivator of the cell survival kinase AKT
AL928768.3	4,4635E-11	2,39659405	Immunoglobulin Heavy Constant Alpha 1	Immune	Immunoglobulin	IGHA1: Constant region of immunoglobulin heavy chains
CCR7	3,8737E-12	2,44743355	C-C Motif Chemokine Receptor 7	Immune	Immune	Expressed in various lymphoid tissues and activates B and T lymphocyte; control the migration of memory T cells to inflamed tissues, as well as stimulate dendritic cell maturation
LINC00926	7,4513E-12	2,52887166	Long Intergenic Non-Protein Coding RNA 926	Immune	B cell	Potential B cell-specific long non-coding RNA
CXCR5	1,1068E-16	2,54287549	C-X-C Motif Chemokine Receptor 5	Immune	Lymphoid nodule initation	Cytokine receptor leading to B-cell migration into B-cell follicles
RASGRP2	3,7363E-13	2,55882536	RAS Guanyl Releasing Protein 2	Immune	T cell	Calcium- and DAG-regulated nucleotide exchange factor specifically activating Rap through the exchange of bound GDP for GTP, aggregation of platelets and adhesion of T-lymphocytes and neutrophils probably through inside-out integrin activation
NIBAN3	2,7961E-15	2,58174917	Protein Niban 3	Immune	B cell	B-Cell Novel Protein 1
IKZF3	2,7797E-13	2,58204749	IKAROS Family Zinc Finger 3	Immune	B cell	Transcription factor that plays an important role in the regulation of B-cell differentiation
FCRLA	3,9354E-17	2,59059435	Fc Receptor Like A	Immune	Immunoglobulin	Development of B cells, mediation of the destruction of IgG-coated antigens and of cells induced by antibodies
BLK	1,5114E-17	2,7086826	BLK Proto-Oncogene, Src Family Tyrosine Kinase	Immune	B cell	B-cell receptor signaling and B-cell development
TNFRSF13C	4,2609E-18	2,75803266	TNF Receptor Superfamily Member 13C	Immune	B cell	Survival of mature B-cells and B-cell response

TCF7	4,7982E-13	2,85385418	Transcription Factor 7	Immune	T cell	T-cell lymphocyte differentiation
FCMR	1,036E-14	2,87286465	Fc Mu Receptor	Immune	Immunoglobulin	Binding to the Fc region of immunoglobulins (Igs)
VPREB3	9,7905E-17	3,17337177	V-Set Pre-B Cell Surrogate Light Chain 3	Immune	Immunoglobulin	B-cell maturation and role in assembly of the pre-B cell receptor (pre-BCR)
LTB	1,1633E-16	3,24759913	Lymphotoxin beta	Immune	Lymphoid nodule initation	Type II membrane protein of the TNF family; inducer of the inflammatory response system and involved in normal development of lymphoid tissue; May play a specific role in immune response regulation.
CD37	3,5418E-16	3,33691287	CD37 Molecule	Immune	B cell	Role in T-cell-B-cell interactions; member of the transmembrane 4 superfamily = tetraspanin family, cell-surface proteins
CD52	5,0071E-12	3,36687226	CD52 Molecule	Immune	Immune	Present on the surface of mature lymphocytes, role of anti-adhesion allowing cells to freely move around
CD22	1,2008E-16	3,39315468	CD22 Molecule	Immune	B cell	B cell activation, negative regulation of B cell receptor signaling pathway and localization of B-cells in lymphoid tissues
CD79A	2,0142E-18	5,17935038	CD79a Molecule	Immune	B cell	Initiation of the signal transduction cascade activated by binding of antigen to the B-cell antigen receptor complex
CD79B	9,5002E-20	5,67599876	CD79b Molecule	Immune	B cell	Initiation of the signal transduction cascade activated by binding of antigen to the B-cell antigen receptor complex
MS4A1	4,5456E-16	5,78090014	Membrane Spanning 4-Domains A1	Immune	B cell	Development, differentiation, and activation of B-lymphocytes
CCL19	1,47E-12	7,30617032	C-C Motif Chemokine Ligand 19	Immune	Lymphoid nodule initation	T-cell and B-cell migration to secondary lymphoid organs

(1) Genes selected based on the cut off of p<10^-9 and Fold-Change > 2 which are the highest p value and the lowest Fold-Change from all the pairs of comparison (ie. TLS vs lymphocytes, TLS vs tumor, TLS vs vessel…)

Supplementary Table 14: Morphological and molecular characteristics of the fourteen megaclusters

(referred to **Figure 7c, Supplementary Figures 11b-18**)

Megaclu-sters	Main TNBC mole-cular sub-types [1]	Annota-tions [2]/clinic-pathologi cal data	Signatures [1]		xCells [1]		Single genes [1]	
			↑	↓	↑	↓	↑	↓
MC 1	BL	Tumor>stro ma	NOTCH sig, mTORC1 sig, Androgen response, Estrogen late/early, hedgehog, TGFbeta sig, wnt sig, heme/fatty acid metabolism, oxydative phospho, Proliferation sig (G2M, CIN70, GENE70, GGI, E2F, mitotic spindle, Myc v1)	Apoptosis, immune sig (Immune 1/2, allograft rej, coag, IL6 sig, IFNa/g sig, Trm, TLS, TAM) EMT, angiogenesis, myogenesis, Parpi7, VCPredTN, angiogenesis, Stroma1/2, ecm-myCAF, TGFb-myCAF, wound-myCAF, detox-iCAF, IL-iCAF, normal fibroblast, IFNg-iCAF	CD4 memory T cell, Th2, CD8 naive T cells, smooth muscle, CLP, preadipocyt es, neutrophils	B cells, Th1, NK/NKT, DC, macrophag es, endothelial cells, fibroblasts, pericytes	HRD genes (RAD51D/C, PTEN, PALB2, NBN, FANCM, CDK12 BRIP1, BRCA2/1, BARD1), cell cycling genes (PLK4, TTK, MYC, CDKN2A/C, CDK2/4, CCNE1), VEGFA, KDR, methylation genes (HDAC1, DNMT3B/A, TRDMT1, DNMT1), MMR genes (PMS1, MSH6, MSH2, MLH1, POLH), PDGFC, PIK3CA, mTOR, ERBB2, LIV1, 5T4, Trop2, GPNMB, B7-H4	Immune genes (TLR7, PD1, LAG3, TIM3, GITR, FOXP3, CTLA4, ADA, CCL5, CD134) VEGFB
MC 2	BL	Tumor>stro ma	PARPi7, DNA repair, mTORC1 sig, Proliferation sig (G2M, CIN70, GGI, E2F, Myc v1/2) Glycolysis	TGFbeta sig, apoptosis, all immune sig, angiogenesis, EMT, ecm-myCAF, TGFb-myCAF, wound-myCAF, detox-iCAF, IL-iCAF, normal fibroblast, IFNg-iCAF, VCPredTN	Neutrophils, smooth muscle, CLP, epithelial cells	B cells, CD4/CD8 T cells, DC, macrophag es, endothelial cells, fibroblasts, pericytes, adipocytes	Cell cycling genes (PLK4, TTK, MYC, CDKN2A, CDK4, CDK2, CCNE1), VEGFA, TF, ERBB2/3, AKT1/mTOR, Trop2, NECTIN4, B7-H4, CD47, methyl genes (HDAC1, DNMT3B/A, TRDMT1)	ATM, ESR1, AR, immune genes, NTRK2, PDGFRA/B, KDR, FLT1/4
MC 3	M, BL	Stroma > tumor	NOTCH sig, PARPi7, Proliferation sig (G2M, CIN70, GENE70, GGI, E2F, mitotic spindle, Myc v1/2), hypoxia, glycolysis, myogenesis, apical junction, DNA repair	Immune sig(Immune 1/2, allograft rej, IL2/6 sig, IFNa/g sig, Trm, TLS, TAM),VCPredTN, ecm-myCAF, wound-myCAF, detox-iCAF, IL-iCAF, normal fibroblast, IFNg-iCAF	Smooth muscle, epithelial cells	B cells, CD4/CD8 T cells, DC, macorphag es, endothelial cells, fibroblasts, adipocytes, basophils	Cell cycling genes (PLK4, TTK, MYC, CDKN2A/C, CDK2, CCNE1, CCND1), VEGFA, VEGFB, mTOR/AKT1, ERBB2, LIV1, 5T4, Trop2, NECTIN4, B7-H4, B7-H3, CD47, CD73	ATM, ESR1, AR, RB1, immune genes
MC 4	M	Stroma > tumor, lactiferous ducts	Myc v1	PI3K/AKT/mTOR and mTORC1 sig, apoptosis/hypoxia, metabolism (heme, bile acid, xenobiotic), immune sig, angiogenesis, VCPredTN, EMT, ecm-myCAF, TGFb-myCAF, wound-myCAF, IL-iCAF, IFNg-iCAF	Myocytes, smooth muscle, CLP, preadipocyt e, neurons	B cells, CD4/CD8 T cells, macrophag es, pDC/aDC, NKT, endothelial cells, MSC	MYC, ERBB2, PSMA, B7-H4	ESR1, AR, ATM, CDKN2A, FLT1/4, VEGFB, PDGFRB, 5T4, GPNMB, immune genes

MC 5	BL, IM	Tumor >stroma > lymphocytes	PI3K/AKT/mTOR and mTORC1 sig, proliferation sig (G2M, CIN70, GENE70, GGI, E2F, Myc v1/2), reactive oxygen sp path, metabolism (heme/fatty acid), glycolysis, oxydative phosphorylation, immune sig (Immune2, TLS Lund, IFNa/g res, Trm), adipogenesis, DNA repair	TGFbeta sig, Hedgehog sig, hypoxia, angiogenesis, EMT, ecm-myCAF, TGFb-myCAF, wound-myCAF, detox-iCAF, IL-iCAF, normal fibroblast	Th1/Th2, CD8 naive T cells, smooth muscle, CLP, preadipocytes	CD4 naive T cells, Tregs, NKT, iDC, monocytes, endothelial cells, fibroblasts, pericytes, MSC, adipocytes, neurons	Cell cycling genes (PLK4, TTK, MYC, CDKN2A, CDK2/4/6, CCNE1), HDAC1, mTOR, ERBB3, LIV-1, Trop2, NECTIN4, immune genes (LAG3, TIGIT, PD-1, PD-L1, IDO1, TIM3, DCIR, CTLA4, CD47, TLR2, TLR3, ICOS, CXCR4, CD40, IFNAR1/IFNGR1)	AR, FLT4, VEGFB, PDGFRA/B, immune genes (CD73, IFNAR1, GITR, CXCL12, B7-H3)
MC 6	M, BL	Tumor >stroma > lymphocytes	NOTCH sig, Parpi7, DNA repair, PI3K/AKT/mTOR and mTORC1 sig, hormonal path (androgen res, estrogen early/late), wnt sig, prolif sig (G2M, CIN70, GENE70, GGI, E2F, mitotic spindle), apoptosis/hypoxia/glycolysis, metabolism (heme/xenobiotic/cholesterol), complement, adipogenesis, peroxisome	AR gene, TLS cabrita	Class switched memory B cells, NKT, basophils, epithelial cells	B cells/pro B cells, memory/naive CD4 T cells, CD8 naive T cells, monocytes, endothelial cells, pericytes	Cell cycling genes(PLK4, TTK, MYC, CDKN2A/C, CCNE1), NTRK3, PDGFC, PIK3CA/AKT/mTOR, LIV1, GPNMB, immune genes (LAG3, IFNGR1, IDO1, PDL1, TLR2/3)	BRCA1, RAD50, RAD51D, AR, VEGFB, DNMT3A, ERBB3, immune genes (IFNAR1, CD39, ADA)
MC 7	IM	Stroma > tumor > lymphocytes	DNA repair, VCpredTN, Parpi7, KRAS sig, immune sig (Immune1/2, TLS, allograft rej, complement, IL2/6 sig, IFNa/g sig, Trm, TAM), apoptosis, IFNg-iCAF	Androgen response, TGFbeta sig, protein secretion, GENE70	Th1, NKT, DCs, basophils, macrophages, memory B cells, pericytes, epithelial cells	Mast cells, myocytes, smooth muscle, CLP, neurons, preadipocytes	Cell cycling genes (CDKN2A, CCND1), immune genes (TIGIT, PD1, LAG3, IDO1, ICOS, TIM3, GITR, DCIR, CTLA4, CD40, CCL5)	HRD genes (RAD50, PTEN, PALB2, NBN, FANCM, EMSY, CDK12, BRCA2, BARD1, ATR, ATM), RB1, VEGFA, PIK3CA, PDGFC, MMR genes (PMS1, MSH6, MSH3, MLH3, MLH1), methylation genes (TRDMT1, DNMT1)
MC 8	LAR	Stroma > tumor, in situ	PI3K/AKT/mTOR and mTORC1 sig, hormonal path (AR gene, androgen res, estrogen late/early), metabolism (heme/fatty acid/xenobiotic), glycolysis, adipogenesis	Immune sig, detox-iCAF, IL-iCAF, normal fibroblast, IFNg-iCAFWnt sig, hedgehog sig, angiogenesis, EMT, GENE70	Smooth muscle, epithelial cells	B cells, CD4/CD8 T cells, DC, macrophages, endothelial cells, pericytes	AR, RB1, PDGFC, AKT1/mTOR, ERBB2/3, EGFR, LIV1, 5T4, Trop2, NECTIN4	Immune genes, NTRK2/3, PDGFRA, TF, cell cycling genes (MYC, CDKN2A)
MC 9	IM	Stroma > lymphocytes, TLS	CAF, KRAS sig, apoptosis, all immune sig, TGFb-myCAF, wound-myCAF, detox-iCAF, IL-iCAF, normal fibroblast, IFNg-iCAF	NOTCH sig, hormonal path, proliferation sig, hypoxia, glycolysis, adipogenesis	All B cells, plasma cells, CD4/CD8 T cells, all DCs, basophils, macrophages, mast cells, monocytes, endothelial cells	NK cells, myocytes, smooth muscle, epithelial cells, neurons	Immune genes	HRD genes, cell cycling genes, VEGFA, PIK3CA, PDGFC, ERBB2/3, EGFR, B7-H4, B7-H3

MC 10	MSL, IM	Stroma>> lymphocytes, fat tissue, vessels	Stroma1/2, CAF, TGFb-myCAF, wound-myCAF, detox-iCAF, IL-iCAF, normal fibroblast, IFNg-iCAF, KRAS sig, all immune sig, apoptosis, angiogenesis, EMT, myogenesis, apical junction	mTORC1 sig, proliferation sig, hormonal path, hypoxia, glycolysis, oxydative phospho, adipogenesis/fatty acid metab	B cells, CD4 naive T cells, plasma cells, CD8 T cells, all DCs, macrophages, endothelial cells, fibroblasts, pericytes, adipocytes, MSC, CMP	Th2, smooth muscle, CLP, epithelial cells	ATM, ESR1/AR, PDGFRA/B, immune genes	HRD genes (TP53, RAD51C, PTEN, NBN, FANCA, CHEK2, BRIP1), cell cycling genes, methylation genes, PDGFC, PIK3CA/AKT1/mTOR, ERBB2/3, EGFR
MC 11	M, MSL	Stroma	Stroma1/2, CAF, NOTCH sig, hedgehog sig, KRAS sig, TGFbeta sig, apoptosis, hypoxia, coagulation, angiogenesis, EMT, apical junction	mTORC1 sig, proliferation sig, TLS Lund, fatty acid/heme metabolism	NKT, macrophages, endothelial cells, fibroblasts, pericytes, MSC	CD4 memory T Cells, Th2, CD8 T cells, mast cells, smooth muscle, CLP, preadipocytes	VEGFB/KDR/FLT 1-4, PDGFRA/B, immune genes (CD73, GITR, B7-H3, CD134)	HRD genes, cell cycling genes, immune genes (IFNGR1, IFNAR2, IDO1, CXCR4, CD47, CD96, TLR3), MMR genes
MC 12	M	Stroma > necrosis	Mitotic spindle, hedgehog sig, myogenesis, apical junction	NOTCH sig, reactive oxygen sp path, apoptosis, metabolism (fatty acid, heme, xenobiotic), immune sig, adipogenesis, VCPredTN	Fibroblasts, myocytes, pericytes	Memory/naive B cells, CD4 Tem, Th1, CD8 Tcm, DCs, macrophages, smooth muscle, preadipocytes	HRD genes (RAD51D, RAD50, EMSY, CDK12, BRIP1, BRCA2/1, ATR, ATM), VEGFA, NTRK3, PDGFC, ERBB3, MMR genes (PMS1-2, MSH3, MLH3), B7-H3, IFNAR1, methylation genes (TRDMT1, DNMT1)	Cycling genes (CDK2/4, CCNE1), mTOR/AKT1, immune genes
MC 13	M, MSL	Stroma> fat tissue, necrosis	CAF, AR gene, androgen res, hedgehog sig, mitotic spindle, wound-myCAF, detox-iCAF, normal fibroblast	Proliferation sig, NOTCH sig, mTORC1 sig, glycolysis, oxydative phospho, TLS lund, Immune1/2, Trm, VCPredTN, Parpi7	CD4 memory T cells, T regs, mast cells, myocytes, CLP, adipocytes	Th1, NK cells, CD8 Tem, basophils, macrophages, MSC, epithelial cells	HRD genes (RAD50, EMSY, CDK12, BRCA2, BRCA1, BARD1, ATR, ATM), ESR1, AR, RB1, VEGFA/KDR/FLT 1, PIK3CA, ERBB3, EGFR, MMR genes (PMS1-2, MSH3, MLH3), methylation genes (TRDMT1, DNMT1)	Immune genes, CCNE1, CDKN2A, VEGFB, AKT1,
MC 14	MSL	Stroma>fat tissue, vessels	Stroma1/2, CAF, TGFb-myCAF, wound-myCAF, detox-iCAF, IL-iCAF, normal fibroblast, IFNg-iCAF, Hedgehog sig, KRAS sig, TGFbeta sig, apoptosis, hypoxia, immune sig, angiogenesis, EMT, adipogenesis	mTORC1 sig, estrogen late, proliferation sig, glycolysis, oxydative phospho	B cells, memory B cells, naive/memory CD4 T cells, DCs, monocytes, macrophages, endothelial cells, fibroblasts, pericytes, adipocytes	Plasma cells, Th1, basophils, epithelial cells	HRD genes, AR, RB1, VEGFB/KDR/FLT 1-4, PDGFC/PDGFRA/B, PIK3CA, GPNMB, immune genes (CD73, IL2RA, TIM3, CD39, DCIR, CXCL12, CSF1R, B7-H3, CD134, ADA)	CDKN2A, CCNE1, PLK4, ERBB2/3, B7-H4

Megaclusters	Survival - ST TNBC pseudobulk [3]	Survival - ST TNBC bulk [4]	Survival - METABRIC [5]	Survival - SCAN-B [6]	3 cohorts (ST, METABRIC, SCAN-B) [7]	Main characteristics	Summary
MC 1	/	/	/	/	/	High proliferation, NOTCH, Wnt, TGFbeta, immune depletion	Proliferation, Wnt, TGF beta
MC 2	/	iBCFS↓	/	/	iBCFS, DRFS↓	High proliferation, glycolysis, immune depletion, DNA repair Neutrophils, epithelial cells	Proliferation, glycolysis, DNA repair

MC 3	/	/	/	/	/	High proliferation, hypoxia, glycolysis, NOTCH, immune depletion, DNA repair Epithelial cells CD73 gene	Proliferation, hypoxia, glycolysis, DNA repair
MC 4	/	/	/	/	/	Low PI3K/AKT/mTOR and low metabolism, immune depletion, MYC target v1 Neurons, smooth muscle	MYC target v1
MC 5	RFS, iBCFS↑	RFS, iBCFS, DRFS, OS ↑	/	OS↑	iBCFS, DRFS, OS↑	High PI3K/AKT/mTOR, proliferation and immune activation CD8 naive T cells	PI3K/AKT/mTOR, proliferation, immune activation
MC 6	/	RFS ↑	/	/	/	High PI3K/AKT/mTOR, NOTCH, hormone related pathway, proliferation, hypoxia, apoptosis	PI3K/AKT/mTOR, proliferation, hormonal pathway, hypoxia, apoptosis
MC 7	/	iBCFS ↑	/	/	/	Immune activation, response to PARPi, apoptosis	Immune activation, response to PARPi
MC 8	/	/	/	/	/	High PI3K/AKT/mTOR, hormone related pathway, metabolism, adipogenesis, immune depletion	PI3K/AKT/mTOR, hormonal and metabolic pathways, adipogenesis
MC 9	/	/	/	OS↑	iBCFS, DRFS, OS↑	Immune activation, apoptosis, low angiogenesis, CAF	Immune activation, CAF, apoptosis
MC 10	/	/	/	/	/	Stroma and immune activation, CAF, angiogenesis, EMT	CAF, immune activation, angiogenesis
MC 11	/	/	/	OS↓	iBCFS, DRFS, OS↓	Stroma activation, CAF, TGF beta, Hedgehog, high angiogenesis, EMT CD73	CAF, angiogenesis, TGF beta, Hedgehog
MC 12	/	/	/	/	iBCFS↓	High myogenesis, low metabolism, Hedgehog, immune depletion Fibroblasts, myocytes	Myogenesis, Hedgehog
MC 13	/	/	/	OS↑	iBCFS↑	Stroma activation, CAF, Hedgehog, androgen response	CAF, Hedgehog, androgen response
MC 14	/	/	/	/	iBCFS↑	Stroma activation, CAF, TGFbeta signaling, angiogenesis, adipogenesis, EMT, immune activation	CAF, immune activation, angiogenesis, adipogenesis

[1] Assessment of TNBC molecular subtypes, signatures, xCells enrichment score and specific gene expressions based on pseudobulk of all per-spot deconvoluted prototypes/individual clusters belonging to the same megacluster
[2] Means of contributions (%) of morphological annotations based on prototypes/individual clusters (not deconvoluted) belonging to the same megacluster
[3] Survival analysis based on the contribution (%) of per-spot deconvoluted megaclusters (as continous variables): forest plot with univariate analysis where only FDR <0.05 are reported
[4] Survival analysis based on the contribution (%) of deconvoluted megaclusters (as continous variables) from RNA bulk: forest plot with univariate analysis where only FDR <0.05 are reported
[5] Survival analysis based on the contribution (%) of deconvoluted megaclusters (as continous variables) from RNA bulk: forest plot with univariate analysis where only FDR <0.05 are reported
[6] Survival analysis based on the contribution (%) of deconvoluted megaclusters (as continous variables) from RNA bulk: forest plot with univariate analysis where only FDR <0.05 are reported
[7] Survival analysis based on the contribution (%) of deconvoluted megaclusters (as continous variables) from RNA bulk from 3 cohorts corrected for study: forest plot with univariate analysis where only FDR <0.05 are reported
NB. Multivariate analysis controlled for (age (as continous variable), lympho node (>0) and tumor size (>20 mm for METABRIC and SCAN-B, T>1 vs T=1 pour ST cohort))

Supplementary Table 15: Morphological and molecular characteristics of the nine spatial archetypes

(referred to **Figure 8e-f, Figure 9; Supplementary Figures 19b-33**)

Spatial archetypes	Main TNBC molecular subtypes [1]	Main MC [2]	Annotation/ Clinicopathological [2]	Signatures [1]	xCells [1]	Single genes [1]	Survival - ST TNBC pseudo bulk [3]	Survival - ST TNBC bulk [4]	Survival - METABRIC [5]	Survival - SCAN-B [6]	3 cohorts (ST, METABRIC, SCAN-B) [7]	Summary
SA1	IM, MSL	MCs 6, 10	/	High KRAS signaling, high immune signatures including TLS signature, high HRD signature, high IFNg-iCAF	Enriched by DC	High immune genes including CD27, PD1, VISTA, CD134, CSF1R	/	/	/	/		High immune with TLS sig, DNA repair deficiency, KRAS-dependent
SA2	BL	MC2	/	Low stroma signatures with low EMT, higher proliferation in METABRIC and SCAN-B	Enriched by pro B cells	High ERBB3, low NTRK2 and PDGFRA/B	/	/	/	/		High proliferation and high ERBB3 expression
SA3	IM	MC7	Medullary, TILs,	High TAM and Trm signatures, high Myc v2 target, low Stroma 1 signature	Enriched by pDC, depleted by endothelial cells	High immune genes including CCL5, CD40, TIM3, IDO1 and LAG3	/	/	/	/		High proliferation, high immune with PD-L1 expression
SA4	IM	MC5	High Ki67, TILs, CD20	High VCpredTN signature, high immune signatures including TLS, Trm and TAM, high IFNg-iCAF Low Stroma, EMT and myogenesis	Enriched by B cells including memory, pro B cells, plasma cells, Th2, CD8 naive and Tcm, aDC and pDC, macrophages	High immune genes including PD-L1, CTLA4, ICOS, LAG3, PD1, TIGIT, TIM3, ADA	OS, DRFS, iBCFS, RFS ↑	/	/	/	OS, DRFS, iBCFS ↑	High immune with TLS sig and PD-L1 expression, DNA repair deficiency
SA5	LAR	MC8	Low Ki67, grade and TILs	High stroma, high adipogenesis, high AR/ESR1/ER pathway, high metabolisms, high wound-myCAF, detox-iCAF and normal fibroblast Low proliferation	Enriched by adipocytes Depleted by pro B cells and Th2	High AR, ESR1 and AKT1 Low immune genes including IFNAR2, LAG3, B7-H4; low cell-cycling genes including CDKN2A, CCNE1, MYC, TTK, PLK4	/	/	/	/		LAR-enriched, high metabolic pathways, high ERBB2 expression
SA6	M, MSL	MC1	/	High TGFb signature, protein secretion	Enriched by smooth muscle, depleted by Th1	High PDGFRA, PDGFC, KDR, FLT1, CDKN2C	/	/	/	/	DRFS ↑	High stromal, high proliferation and high PDGF expression
SA7	M	MC3	LowTILs and CD20	High stroma with TGFb sig, EMT, myogenesis and apical junction, high hypoxia, glycolysis and angiogenesis in ST bulk, METABRIC and SCAN-B Low immune 1 signature	Depleted by basophils (+ almost immune cells in METABRIC and SCAN-B)	High CD73 in ST bulk and SCAN-B, high VEGFA in METABRIC and SCAN-B, high VEGFB in ST PB and SCAN-B	OS, DRFS ↑	OS, DRFS↑	/	/		High stromal, high EMT, high angiogenesis and adenosine pathway-enriched, high proliferation

SA8	M, BL	MC2	Low TILs	Low HRD signature, low immune including TLS and TAM	Depleted by cDC, B cells and CD8 T cells	High NECTIN4 Low immune genes including CCL5, CD27, CD40, TIGIT, ICOS, CTLA4	/	/	/	/	OS, DRFS, iBCFS ↓	Low immune, high NECTIN4 expression
SA9	M	MC4	Low TILs and CD20	Low HRD signature, low PI3K/AKT/mTOR signaling, low immune 2 signature	Depleted by B cells, plasma cells, macrophages and aDC	High NTRK2 (in ST bulk, METABRIC and SCAN-B) Low immune genes including CCL5 and TIGIT	/	/	/	/		High stromal and low immune

[1] Summary of characterization of spatial archetypes across 4 datasets: pseudobulk ST, bulk ST, METABRIC and SCAN-B

[2] Summary of characterization of spatial archetypes on pseudobulk ST only

[3] Survival data based on spatial archetypes from ST TNBC pseudobulk based on the contribution (%) of per-spot deconvoluted megaclusters (as continous variables): forest plot with univariate analysis where only FDR <0.05 are reported

[4] Survival data based on spatial archetypes deconvoluted from ST TNBC bulk: forest plot with univariate analysis where only FDR <0.05 are reported

[5] Survival data based on spatial archetypes deconvoluted from METABRIC data (TNBC patients): forest plot with univariate analysis where only FDR <0.05 are reported

[6] Survival data based on spatial archetypes deconvoluted from SCAN-B data (TNBC patients): forest plot with univariate analysis where only FDR <0.05 are reported

[7] Survival data based on spatial archetypes deconvoluted from RNA bulk from 3 cohorts corrected for study: forest plot with univariate analysis where only FDR <0.05 are reported

Supplementary Table 16: List of selection of molecular features relevant to characterize the spatial archetypes
(referred to **Figure 8e** and **Supplementary Figures 20-21)**

Main function	Type of molecular features	List of selected molecular features
Immune	Signatures	Complement IL6 JAK STAT3 signaling IL2 STAT5 signaling Interferon gamma response Interferon alpha response TLS ST
	Single genes	*CD73*
	Cell types	TAM [1] Trm [1]
Stroma	Signatures	Adipogenesis Epithelial mesenchymal transition Stroma1 Stroma2
	Single genes	*VEGFB* *PDGFRA* *PDGFC*
	Cell types	Adipocytes [2] Fibroblasts [2]
Oncogenic	Signatures	Myc targets v1 Myc targets v2 GGI Mitotic spindle KRAS signaling Androgen response
	Single genes	*AR* *ESR1* *ERBB3* *NECTIN4*
Stress response	Signatures	VCpredTN Apoptosis
Metabolism	Signatures	Fatty acid metabolism Bile acid metabolism Xenobiotic metabolism Peroxisome

[1] Computed as signatures by GSVA
[2] Computed by xCell deconvolution

SUPPLEMENTARY REFERENCES

1. Andersson, A. *et al.* Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nat. Commun.* **12**, 6012 (2021).
2. Cabrita, R. *et al.* Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* **577**, 561–565 (2020).
3. Meylan, M. Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer. *Immunity* **55**, 527-541 5 (2022).
4. Wolf, D. M. Redefining breast cancer subtypes to guide treatment prioritization and maximize response: Predictive biomarkers across 10 cancer therapies. *Cancer Cell* **40**, 609-623 6 (2022).
5. Bareche, Y. Leveraging big data of immune checkpoint blockade response identifies novel potential targets. *Ann Oncol* **33**, 1304–1317 (2022).
6. METABRIC Group *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
7. Staaf, J. RNA sequencing-based single sample predictors of molecular subtype and risk of recurrence for clinical assessment of early-stage breast cancer. *Npj Breast Cancer* **8**, 94 (2022).
8. Tolaney, S. M. Updated Standardized Definitions for Efficacy End Points (STEEP) in Adjuvant Breast Cancer Clinical Trials: STEEP Version 2.0. *J Clin Oncol* **39**, 2720–2731 (2021).