ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO OPEN SCIENCE FOR OPTIMAL CANCER CARE

ORIGINAL RESEARCH

An immunological signature to predict outcome in patients with triple-negative breast cancer with residual disease after neoadjuvant chemotherapy

C. Blaye[1,2†], É. Darbo[3†], M. Debled[1], V. Brouste[4], V. Vélasco[5], C. Pinard[6], N. Larmonier[2,7], I. Pellegrin[8,9], A. Tarricone[8], M. Arnedos[1], J. Commeny[10], H. Bonnefoi[1,3,7], C. Larmonier[5] & G. MacGrogan[3,5*]

[1]Department of Medical Oncology, Institut Bergonié, Bordeaux; [2]Univ. Bordeaux, CNRS, ImmunoConcEpT, UMR 5164, Bordeaux; [3]Univ. Bordeaux, INSERM U1218, ACTION Laboratory, Bordeaux; Departments of [4]Clinical Research and Medical Information; [5]Biopathology, Institut Bergonié, Bordeaux; [6]Pathology Laboratory, University Hospital of Martinique, Fort de France; [7]Univ. Bordeaux, Bordeaux; [8]Service d'Immunologie et Immunogénétique, University Hospital of Bordeaux, Bordeaux; [9]Centre de Ressources Biologiques Plurithématique, University Hospital of Bordeaux, Bordeaux; [10]Department of Surgery, Institut Bergonié, Bordeaux, France

Available online 24 June 2022

Background: When triple-negative breast cancer (TNBC) patients have residual disease after neoadjuvant chemotherapy (NACT), they have a high risk of metastatic relapse. With immune infiltrate in TNBC being prognostic and predictive of response to treatment, our aim was to develop an immunologic transcriptomic signature using post-NACT samples to predict relapse.

Materials and methods: We identified 115 samples of residual tumors from post-NACT TNBC patients. We profiled the expression of 770 genes related to cancer microenvironment using the NanoString PanCancer IO360 panel to develop a prognostic transcriptomic signature, and we describe the immune microenvironments of the residual tumors.

Results: Thirty-eight (33%) patients experienced metastatic relapse. Hierarchical clustering separated patients into five clusters with distinct prognosis based on pathways linked to immune activation, epithelial-to-mesenchymal transition and cell cycle. The immune microenvironment of the residual disease was significantly different between patients who experienced relapse compared to those who did not, the latter having significantly more effector antitumoral immune cells, with significant differences in lymphoid subpopulations. We selected eight genes linked to immunity (*BLK*, *GZMM*, *CXCR6*, *LILRA1*, *SPIB*, *CCL4*, *CXCR4*, *SLAMF7*) to develop a transcriptomic signature which could predict relapse in our cohort. This signature was validated in two external cohorts (KMplot and METABRIC).

Conclusions: Lack of immune activation after NACT is associated with a high risk of distant relapse. We propose a prognostic signature based on immune infiltrate that could lead to targeted therapeutic strategies to improve patient prognosis.

Key words: triple-negative breast cancer, residual disease, tumor immune microenvironment, transcriptomic signature

INTRODUCTION

Triple-negative breast cancer (TNBC) accounts for 10%-15% of breast cancers and is defined by the absence of estrogen receptor (ER) expression, progesterone receptor (PR) expression and human epidermal growth factor receptor 2 (*HER-2*) amplification.[1] In comparison to other breast cancer subtypes, it affects younger women and has worse prognosis with a higher metastatic relapse rate, mostly in the first years following initial treatment.[1] In TNBC patients treated with neoadjuvant chemotherapy (NACT), pathological response of the primary tumor is a very important prognostic factor as demonstrated in several trials[2-4] and in a large meta-analysis.[5] Patients whose tumors achieved a pathological complete response (pCR) had a risk of invasive relapse of up to 15%, while patients with residual invasive disease had a risk of metastatic relapse of 50%.[2-5] Research approaches have been tested aiming to decrease the TNBC risk of relapse. The first approach focuses on reinforcing adjuvant treatments in patients whose tumors did not achieve a pCR, with capecitabine[6,7] or poly (ADP-ribose) polymerase inhibitors for *BRCA*-mutated patients.[8] A second potential approach could be immune checkpoint inhibitors

*Correspondence to: Dr Gaëtan MacGrogan, Department of Biopathology, Institut Bergonié, 229 Cours de l'Argonne 33000 Bordeaux, France. Tel: +33-(0)-5-56-33-33-36; Fax: +33-(0)-5-56-33-33-38

E-mail: g.macgrogan@bordeaux.unicancer.fr (G. MacGrogan).

†These authors contributed equally to this work.

(ICIs), since TNBC appears to have a higher immunogenicity than other breast cancer subtypes.[9,10] Hence, ICIs were tested in advanced TNBC and two became Food and Drug Administration approved in this setting, atezolizumab[11] and pembrolizumab.[12] Pembrolizumab has been recently approved in the neoadjuvant setting,[13] and durvalumab[14] and atezolizumab[15] showed promising results in the early setting as well.

Although ICIs in the neoadjuvant setting look promising,[13-15] it should be noted that many patients with TNBC treated with standard NACT (without ICI) will not relapse. These patients should not receive these expensive compounds and should be spared their toxicity that can be severe and definitive. Thus, the identification of prognostic factors in TNBC patients remains a research priority. Several trials have successfully identified the prognostic value of tumor-infiltrating lymphocytes (TILs) on the primary tumor before or after medical treatment.[9,10,16-20] In patients treated with NACT, residual cancer burden (RCB), which classifies tumors into four categories, is a strong prognostic factor.[21] However, if RCB 0-I tumors have an excellent prognosis and RCB III tumors a very poor prognosis, RCB II tumor prognosis is very heterogeneous and would need to be refined. In 2019, Luen et al. demonstrated that TILs assessment on residual disease post-NACT adds an independent prognostic value to the RCB.[20] Our group also demonstrated independent prognostic significance of CD4 TILs and RCB.[22] Two limiting issues remain to be overcome: firstly, their quantification is pathologist-specific, and secondly, TILs count does not capture the full immunological microenvironmental picture, which is the result of a very complex interplay between effector and immunosuppressive cells.

The aim of our study was therefore to better describe the immunological microenvironment in residual TNBC after NACT and develop an immunological transcriptomic prognostic signature using patients' residual disease samples.

MATERIALS AND METHODS

Study population and pathological evaluation

One hundred and forty-six consecutive patients with TNBC (diagnosed on the pretreatment core needle biopsies) who underwent surgical resection after NACT from December 2003 to November 2015 at our institute, aged $\geq$18 years and who had residual disease on surgical specimen examination were screened for this study. Formalin-fixed paraffin-embedded (FFPE) surgical specimens were retrieved from the hospital archives. Clinical information was obtained from patients' files. TNBC was defined according to the American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) 2010 guidelines[23] and HER-2 status determined according to the Groupe d'Etude des Facteurs Pronostiques/Prédictifs Immunohistochimiques dans le Cancer du Sein 2014 guidelines.[24] Patients who had a positive hormonal receptor or HER-2 status in the surgical specimen, had received neoadjuvant radiotherapy, had a distant metastasis at the time of diagnosis, had a past history of invasive breast cancer or any other cancer (excluding skin carcinomas) or did not receive at least three cycles of NACT were excluded from the study. Similarly, patients for whom surgical histological material was not available or insufficient for review or RNA extraction were also excluded. Thus, we identified 125 cases matching our selection criteria. After RNA extraction and quality controls, 115 cases were analyzed using the NanoString nCounter® system (NanoString Technologies, Seattle, WA) (flow chart, Supplementary Figure S1, available at https://doi.org/10.1016/j.esmoop.2022.100502).

Pathological examination

Residual tumor hematoxylin—eosin (H&E) slides were reviewed by two pathologists (GM and CP) to determine tumor cellularity, stromal TILs infiltration,[25] RCB according to Breast International Group/North American Breast Cancer Group international working group recommendations[26] as well as area of interest to be sent for RNA extraction.

RNA extraction and analysis

Total tumor RNA [tumor cells and tumor microenvironment (TME)] was extracted from four to six 10-µm tumor sections after identification of the residual tumor by a pathologist on a corresponding H&E section. RNA was extracted using the High Pure FFPET RNA Isolation Kit (Roche®, Meylan France), following the manufacturer's instructions. RNA quantity and quality were assessed by NanoDrop 1000 spectrophotometer (ThermoFisher, Waltham, MA) and 2100 Bioanalyzer with Agilent RNA 6000 Nano Kit (Agilent, Santa Clara, CA). Two hundred nanograms of total RNA was used from each sample for gene expression profiling (acceptation criteria 260/280 ratio: 1.8-2.3, 230/280 ratio: 1.7-2.3).

RNA samples were analyzed based on the NanoString PanCancer Immuno-Oncology panel (IO360), made of 770 genes related to the interplay between tumor, microenvironment and immune response in cancer. Sample runs were carried out on the nCounter® FLEX Analysis System (automated nCounter® Prep station and the nCounter® Digital Analyzer optical scanner, NanoString Technologies) according to the manufacturer's protocol. This panel was designed using biological signatures, including the 18-gene Tumor Inflammation Signature (TIS) from Ayers et al.[27]

Bioinformatics

Housekeeping gene selection. We refined the 20-housekeeping gene list defined in the NanoString IO360 panel using three independent datasets grouping TNBC patients retrieved from xenaBrowser[28] [The Cancer Genome Atlas (TCGA) *BRCA* primary tumors with 123 patients, Chin et al. 2006[29]: 41 patients and Hess et al. 2006[30]: 27 patients]. We computed gene expression variability as the inter-quantile (5%-95%) ratio in each dataset and selected 6 of the 20 genes showing ratios under 1.5 in all datasets: *ERCC3*, *POLR2A*, *SF3A1*, *TBC1D10B*, *TMUB2*, *UBB*.

NanoString data processing. Raw RCC files of 115 patients were loaded into R environment and normalized using NanoStringNorm package.[31] We normalized the expression values with options CodeCount ='geo.mean', Background ='mean.2sd', SampleContent ='housekeeping.geo.mean', round.values = TRUE, take.log = TRUE.

Unsupervised analysis. We first computed gene expression pairwise Spearman's correlations and selected gene pairs with a score of ≥0.7. We then selected modules of correlated genes having at least 10 members using the edge.betweenness.community function from igraph R package.[32] The 573 remaining genes and 115 patients were clustered according to gene expression using PCA and HCPC functions from FactoMineR R package[33] with options ncp = 10 for PCA and default parameters and we extracted four gene clusters and five patient clusters.

Gene set functional enrichment

Gene clusters. We evaluated functional gene cluster (from unsupervised clustering and the extended signature) enrichment using enricher function from ClusterProfiler R package[34] and Reactome, WikiPathways and Hallmarks from the MSigDB[35] and gene sets provided by NanoString. The universe was set as the genes present in the NanoString setup. The most significant term per unsupervised gene cluster was reported. We reported gene sets showing an adjusted *P* value under 0.05 for the extended signature.

Patient comparison. We first computed single-sample gene set enrichment analysis (ssGSEA) scores using GSVA R package[36] with norm.ssgsea option set to FALSE and gene sets provided by NanoString (biological functions and immune cell population signatures). We also computed immune cell population infiltration using Microenvironment Cell Populations (MCP) counter.[37] The scores were then compared between metastatic and non-metastatic patients using Wilcoxon's test.

Supervised analysis for signature gene selection. Due to unbalanced patient groups (77 non-metastatic and 38 metastatic patients), we decided to proceed 1000 times random subsampling of 30 patients in each group and we evaluated the metastatic predictive power of each gene by computing a univariate Cox proportional hazards regression model using the survival R package. The signature was extended by selecting genes showing a Spearman's correlation score above 0.8 with at least one signature member. Signature score was defined as described below.

Signature score calculation. ssGSEA calculates enrichment scores representing how much a set of genes of interest are coordinately up-regulated or down-regulated within a sample, independent of phenotype labeling. We thus computed signature scores by applying this method to each patient and signature gene sets [the eight-gene and extended signature, the Immunological Constant of Rejection (ICR) from Hendrickx et al.[38] and TIS[27]]. A higher score represents a higher activity of the analyzed genes. Cohorts were then separated into two subgroups according to the optimal cut-off defined with surv_cutpoint function from survminer R package.

Table 1. Characteristics of the main cohort population

Clinicopathological characteristics of patients in the study			
		n	%
Age (years)	≤50	60	52.2
	>50	55	47.8
Menopausal status	NR	1	0.9
	0	62	53.9
	1	52	45.2
cT	1	1	0.9
	2	59	51.3
	3	38	33.0
	4	17	14.8
cN	0	59	51.3
	1/2/3	56	48.7
Chemotherapy regimen	Anthracyclines + taxanes	104	90.4
	Anthracyclines	10	8.7
	Taxanes	1	0.9
ypT	1	65	56.5
	2/3	50	43.5
ypN	ypN0	79	68.7
	ypN1/2	36	31.3
Elston and Ellis grade	NR	1	0.9
	2	37	32.2
	3	77	67.0
Ki67 index	<20%	32	27.8
	≥20%	83	72.2
Lymphovascular invasion	NR	1	0.9
	0	87	75.7
	1	27	23.5
TILs	<5%	11	9.6
	≥5%	104	90.4
RCB class	I	14	12.2
	II	73	63.5
	III	28	24.3

NR, not reported; RCB, residual cancer burden; TILs, tumor-infiltrating lymphocytes.

Validation datasets. We selected 442 patients with the filter 'PAM50-basal' from KMplot breast cancer cohort[39] and used online tools to compute the relapse-free survival (RFS) probabilities. We used the median value of our eight-gene and extended signature and separated the cohort according to the optimal score.

We tested our signature on the METABRIC dataset in which we selected 251 ER, PR, HER2-negative status and basal PAM50 patients.[40] These data were accessed through Synapse (synapse.sagebase.org).

Survival analysis. We used survival and survminer R packages to compute Kaplan—Meier curves and log-rank test *P* values to evaluate metastasis-free survival differences.

Statistical analysis

Statistical analysis was carried out using the 9.4 SAS software version (SAS Institute, Inc., Cary, NC). The qualitative data were described by their numbers and percentages, and the quantitative data by their medians and their ranges (min-max). The quantitative data were analyzed as continuous variables.

Distant-relapse-free interval (DRFI) was defined as the time interval between the surgery date and the date of distant recurrence or death from breast cancer, whichever occurred

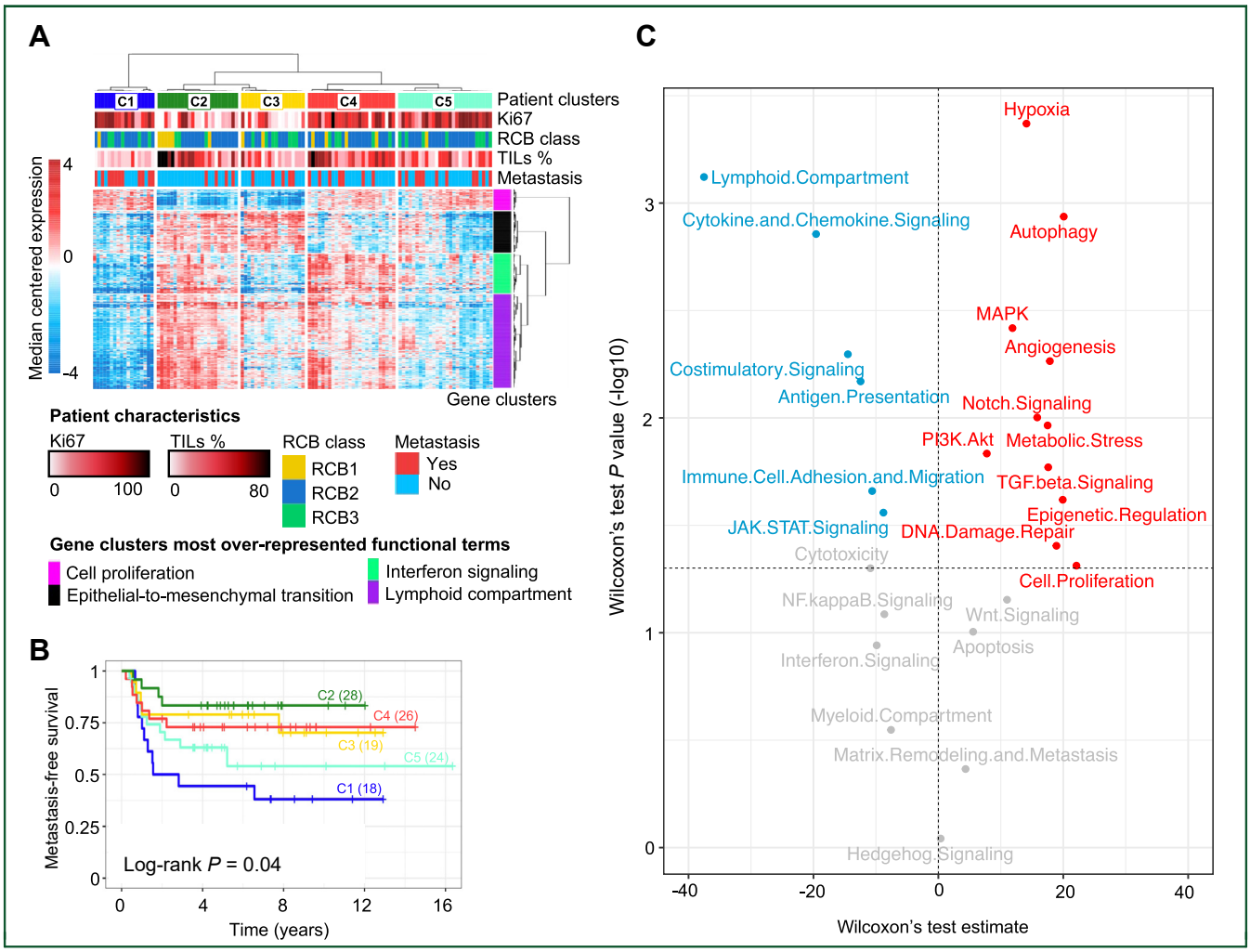


Figure 1. TME of the residual disease.
(A) Unsupervised hierarchical clustering from 115 patients (columns) and 573 genes (rows) highlights five different patient subgroups and four gene clusters for which the most significant enriched functions are presented in the figure legend. (B) Kaplan—Meier analysis of the five patient subgroups defined in A. (C) Volcano plot representing ssGSEA differential analysis. The *x*-axis represents the Wilcoxon's difference estimate and the *y*-axis the corresponding *P* value. Significantly enriched IO360 pathways ($P < 0.05$) in non-metastatic and metastatic patients are highlighted in blue and red, respectively. RCB, residual cancer burden; ssGSEA, single-sample gene set enrichment analysis; TILs, tumor-infiltrating lymphocytes; TME, tumor microenvironment.

first. Follow-up times were censored at last contact if no distant relapse had occurred or if patients had died without distant relapse. The follow-up was calculated using the inverse Kaplan—Meier method, where deaths are censored.

The Kaplan—Meier method was used to estimate DRFI, and the results were compared between groups with log-rank tests. For the univariate analysis of DRFI, hazard ratios (HRs) and 95% confidence interval (95% CI) were calculated with the Cox proportional hazards regression model; *P* Wald is also reported.

Only variables with $P < 0.05$ in the univariate analysis were included in multivariate analyses to investigate their prognostic value. The proportionality of the risks was verified by the Schoenfeld test. A *P* value <0.05 was considered significant.

Ethics

This study was approved by the institutional review board of Institut Bergonié and data collection was in accordance with the French Jardé Law. Patients were informed of the study project and those who were not opposed were included in the study.

DATA AVAILABILITY

The data generated in this study are publicly available in Gene Expression Omnibus (GEO) at GSE185507.

RESULTS

Clinicopathological characteristics of the population

The mean age of the patients was 50.8 years and 46% of them were menopausal. Clinicopathological characteristics of the study cohort are detailed in Table 1. The majority of patients (90.4%) received an anthracycline/taxane-based combination. Fourteen (12%), 73 (63%) and 28 (24%) patients had RCB I, RCB II and RCB III scores, respectively. Thirty-eight (33%) patients experienced a metastatic relapse during follow-up, with a median follow-up of 82.8 months (95% CI 74.2-94.7 months).

Table 2. Univariate and multivariate analysis of factors associated with distant-relapse-free interval (DRFI)

	n = 115	Univariate analysis (DRFI) (Cox)				Multivariate analysis (DRFI) (Cox)			
	n	*P* (Wald)	HR	95% CI HR		*P* (Wald)	HR	95% CI HR	
TILs %									
n	115	0.0036	0.972	0.954	0.991	Not retained $P = 0.86$			
Median value	20								
Signature score									
n	115								
Median value	−63.16	<0.0001	0.985	0.979	0.991	0.0019	0.988	0.980	0.995
RCB score									
n	115	<0.0001	2.712	1.908	3.855	0.0020	1.910	1.266	2.881
Median value	2.1								
LVE									
n	114								
Not seen	87	<0.0001	1			0.0591	1		
Present	27		6.088	3.166	11.71		2.090	0.972	4.493

CI, confidence interval; HR, hazard ratio; LVE, lymphovascular emboli; RCB, residual cancer burden; TILs, tumor-infiltrating lymphocytes.

The tumor microenvironment of residual disease

We carried out unsupervised hierarchical clustering on both samples and genes using the gene expression of 115 residual tumors. We found five clinical clusters (C1-C5), based on the expression of four groups of highly correlated gene modules enriched, respectively, in cell proliferation and epithelial-to-mesenchymal transition, and two immune-related gene clusters (interferon-γ-related response and lymphoid compartment) (Figure 1A). Regarding the survival of the five clusters, three clusters (C2, C3, C4) enriched in immunity genes had a significantly better survival than the two other clusters (C1, C5) (Figure 1B). The subgroups with poorest prognosis, C1 and C5, were markedly enriched in genes implicated in 'cell proliferation' pathways and were rather poor in immune-related gene clusters (Figure 1A). Next, we evaluated biological pathway activity in each patient by carrying out ssGSEA and assessed significant differential activation between patients who later relapsed and those who did not, by applying Wilcoxon's test. For this, we used the functional annotations determined by NanoString on the IO360 panel. In the residual disease of patients who did not experience metastatic relapse, we found a significant up-regulation of pathways implicated in immune response: 'lymphoid compartment', 'cytokine and chemokine signaling', 'costimulatory signaling', 'antigen presentation', 'immune cell adhesion and migration' and 'JAK/STAT' signaling. For patients who later experienced a metastatic relapse, TME of the residual disease was enriched in many processes associated with the tumorigenic and metastatic processes such as 'hypoxia', 'angiogenesis', 'TGF beta signaling', 'Notch signaling' and 'cell proliferation'. Patients who later experienced relapse also displayed other pathways such as 'autophagy', 'MAPK pathway', 'DNA damage repair', 'metabolic stress', 'PI3K.Akt', and 'epigenetic regulation' pathways. Results are shown on a volcano plot (Figure 1C).

Immune profile of post-neoadjuvant chemotherapy triple-negative breast cancer residual disease and metastatic relapse

Next, we evaluated the composition of the immune infiltrate of the residual disease, and explored whether there were differences between patients with and those without metastatic relapse. In our cohort, 30 patients (26%) had a TILs infiltrate of at least 50%. TILs were significantly associated with improved DRFI in univariate analysis (Table 2). The TILs mean percentage in the residual disease of non-metastatic patients was 32.10% versus 17.55% for metastatic patients. Using two different methods (ssGSEA applied to NanoString IO 360 gene sets and MCP counter) to study immune populations, we observed highly significant differences in the immune infiltrate profiles of the post-NACT residual tumors between patients who experienced metastatic recurrence and those who did not. The main population of this infiltrate was cells of monocytic lineage/macrophages (MCP counter method/ssGSEA, respectively) (Figure 2A and B). However, this myeloid infiltrate did not discriminate between metastatic and non-metastatic patients. In lymphoid populations, we found an enrichment in B lymphocytes, in total T cells (with a significant enrichment on the NanoString gene set in Th1 cells, but not in Treg), in CD8 T cells (total and exhausted), on cytotoxic cells and in natural killer (NK) cells for patients who did not experience metastatic relapse (Figure 2A). This enrichment was similar in the MCP counter method (Figure 2A). Both approaches showed very consistent results; indeed, predictions for similar immune populations are highly correlated (Supplementary Figure S2B, available at https://doi.org/10.1016/j.esmoop.2022.100502).

When related to the percentage of TILs on the residual disease as shown in Figure 2B, these differences were less pronounced. The most important difference was the total immune infiltration (Figure 2B). The CD45+ population as measured by the ssGSEA method had a different distribution from the TILs (determined on H&E slides), which was expected since TILs reflect only the lymphocytic population. On these heatmaps, the main immune population was the macrophages/monocytic population; there was also an enrichment in lymphocytic cells (mostly T, CD8, NK and to a lesser extent B cells) that seemed higher in non-metastatic patients (Figure 2B and Supplementary Figure S2C, available at https://doi.org/10.1016/j.esmoop.2022.100502).

Finally, because a high infiltration of B lymphocytes, cytotoxic cells and Th1 cells was detected in the immune

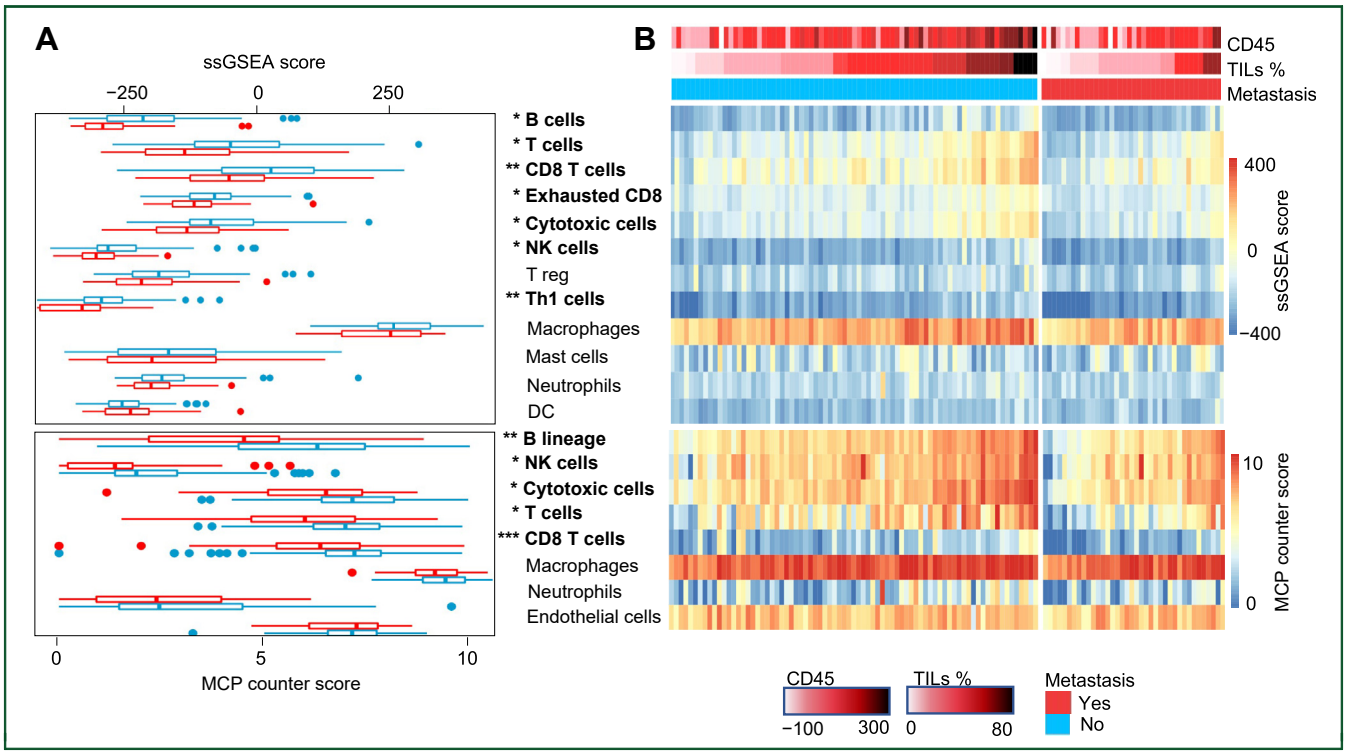


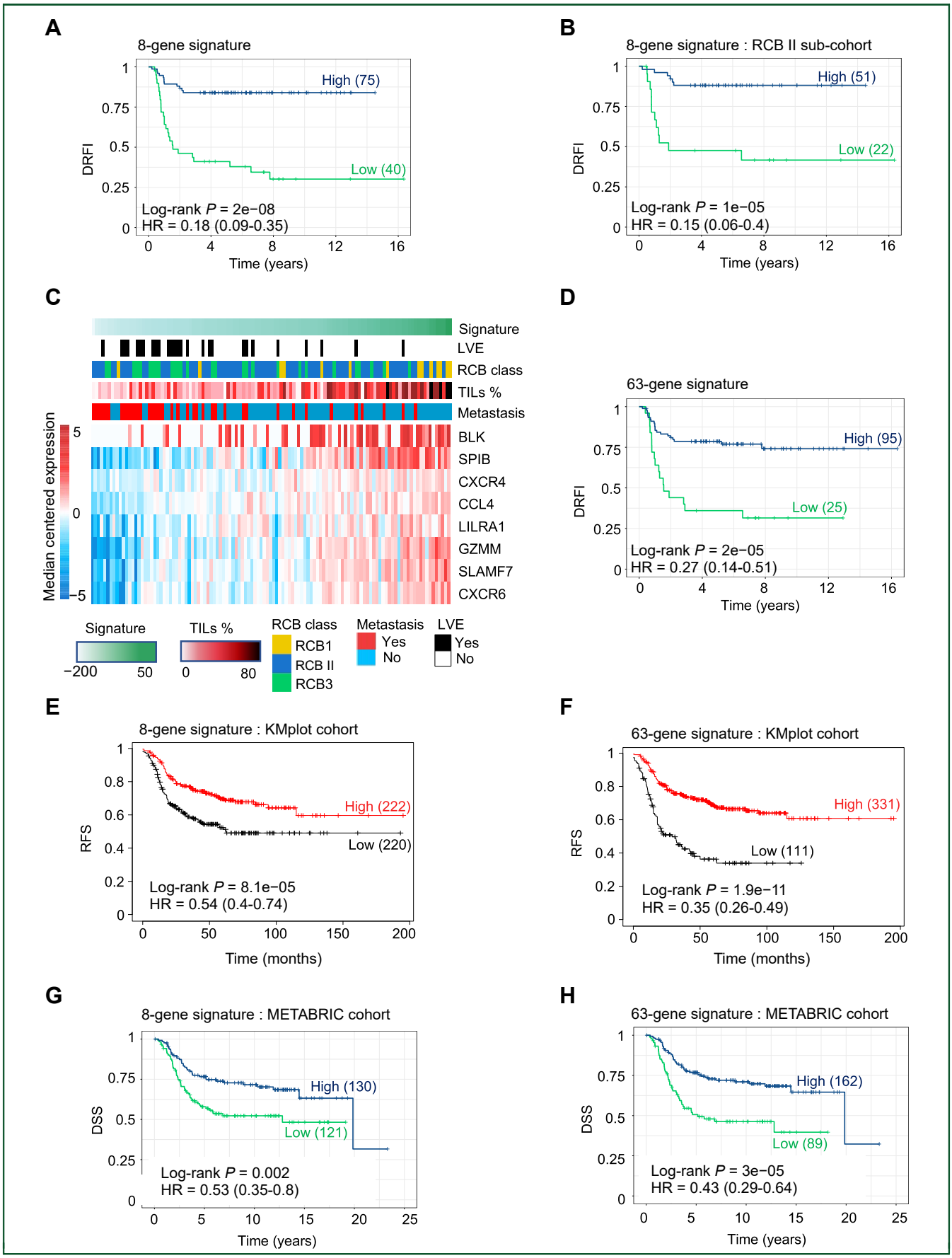
Figure 2. Analysis of the immune microenvironment of the residual disease in non-metastatic versus metastatic relapse patients.
(A) Boxplots summarize score distributions in metastatic (red) and non-metastatic (blue) patients for each corresponding immune cell population in B. (B) Heatmap representing immune infiltration scores computed with ssGSEA applied to NanoString IO 360 gene sets (top panel) and MCP counter (bottom panel). Patients (columns) are ranked according to TILs scores (H&E) in metastatic and non-metastatic patients; scores of CD45 ssGSEA enrichment are reported at the top of the heatmap. The heatmap gives for each immune population its relative enrichment in each sample. (B) Wilcoxon's test significance is noted as follows: *$P < 0.01$, **$P < 0.001$, and ***$P < 0.05$. DC, dendritic cell; H&E, hematoxylin—eosin; MCP, Microenvironment Cell Populations; NK, natural killer; ssGSEA, single-sample gene set enrichment analysis; TILs, tumor-infiltrating lymphocytes.

infiltrate of non-relapsing patients, we applied the 12-chemokine gene expression score previously correlated to the presence of tumor localized ectopic lymph node-like structures[41] and found that it was significantly enriched in patients who did not experience metastatic relapse (Supplementary Figure S2A, available at https://doi.org/10.1016/j.esmoop.2022.100502).

Construction of an eight-gene prognostic signature

We implemented a 1000 times subsampling procedure (detailed in Materials and methods) to determine the robust survival predictive power of each gene and selected eight genes (*BLK*, *GZMM*, *CXCR6*, *LILRA1*, *SPIB*, *CCL4*, *CXCR4*, *SLAMF7*) that were significantly associated with survival (log-rank test P value <0.05 in $>90\%$ of iterations) (Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2022.100502). The score was derived from the eight-gene list using ssGSEA. We tested our signature for the whole cohort and in the RCB II subgroup since it contains patients with very heterogeneous outcomes; we used ssGSEA with the eight selected genes as the gene set. The eight-gene signature could accurately predict DRFI with an HR of 0.18 (95% CI 0.09-0.35), $P < 0.001$ (Figure 3A). Interestingly, our signature could also separate high-risk patients and low-risk patients in the RCB II group [HR 0.15 (95% CI 0.06-0.4), $P < 0.001$] (Figure 3B). The heatmap showing the eight-gene expression according to the signature score is shown in Figure 3C, along with the metastatic status, the RCB class, the presence of lymphovascular emboli and the levels of TILs as seen in H&E.

We next investigated the individual prognostic values of each gene in an independent cohort. We therefore used the KMplot website (https://kmplot.com)[39] and checked the prognostic role of each gene in the basal PAM50 subtype (Supplementary Figure S3A, available at https://doi.org/10.1016/j.esmoop.2022.100502). All genes but one were associated with a significantly improved prognosis on this cohort. The fact that *BLK* had a non-significant trend toward a better survival could be explained by several reasons: lack of sensitivity, inter-cohort variability or because one gene can be associated with several pathways with diverse biological functions. To overcome this problem, we studied the correlations between all genes in our cohort and identified 55 genes that presented a correlation score ≥ 0.8 with at least one gene from our signature. This led to the development of an 'extended signature' composed of 63 genes, which could also predict survival of our patients with good accuracy [HR 0.27 (95% CI 0.14-0.51), $P < 0.001$] (Figure 3D). When applying gene ontology on this set of genes, the main pathways represented were in favor of a strong lymphoid response (lymphoid compartment, costimulatory signaling with *TIGIT* and *PDCD1/PD1* among others, interactions between a lymphoid and a non-lymphoid cell, T-cell receptor-related pathways, etc.)


A
8-gene signature
DRFI
High (75)
Low (40)
Log-rank P = 2e−08
HR = 0.18 (0.09-0.35)
Time (years)
B
8-gene signature : RCB II sub-cohort
DRFI
High (51)
Low (22)
Log-rank P = 1e−05
HR = 0.15 (0.06-0.4)
Time (years)
C
Signature
LVE
RCB class
TILs %
Metastasis
BLK
SPIB
CXCR4
CCL4
LILRA1
GZMM
SLAMF7
CXCR6
Median centered expression
Signature
TILs %
RCB class
RCB1
RCB II
RCB3
Metastasis
Yes
No
LVE
Yes
No
D
63-gene signature
DRFI
High (95)
Low (25)
Log-rank P = 2e−05
HR = 0.27 (0.14-0.51)
Time (years)
E
8-gene signature : KMplot cohort
RFS
High (222)
Low (220)
Log-rank P = 8.1e−05
HR = 0.54 (0.4-0.74)
Time (months)
F
63-gene signature : KMplot cohort
RFS
High (331)
Low (111)
Log-rank P = 1.9e−11
HR = 0.35 (0.26-0.49)
Time (months)
G
8-gene signature : METABRIC cohort
DSS
High (130)
Low (121)
Log-rank P = 0.002
HR = 0.53 (0.35-0.8)
Time (years)
H
63-gene signature : METABRIC cohort
DSS
High (162)
Low (89)
Log-rank P = 3e−05
HR = 0.43 (0.29-0.64)
Time (years)

(Supplementary Figure S3B, available at https://doi.org/10.1016/j.esmoop.2022.100502).

To validate our results, we applied our signatures to the KMplot website selecting only the PAM50 basal subtype,[39] and to the TNBC patients from the METABRIC cohort[40] via the synapse.org website. In the KMplot website, a high eight-gene signature score predicted a significantly better RFS of patients with basal breast cancer: HR 0.54 (95% CI 0.4-0.74), $P < 0.0001$, and the extended signature was even more accurate: HR 0.35 (95% CI 0.26-0.49), $P < 0.0001$ (Figure 3E and F). Similar observations were made after screening the METABRIC cohort. A high signature score predicted a better disease-specific survival, and the extended signature was more precise than the eight-gene signature [HR 0.53 (95% CI 0.35-0.8), $P < 0.01$ and HR 0.43 (95% CI 0.2-0.64), $P < 0.001$, respectively, Figure 3G and H]. Overall, our signatures robustly predicted outcome in two independent cohorts as well as our training set, with the extended signature being more precise.

Univariate and multivariate analyses of factors associated with relapse on triple-negative breast cancer post-neoadjuvant chemotherapy residual disease

In a previous study based on this cohort, TILs, RCB score and lymphovascular emboli were predictive factors for survival in univariate analysis.[22] In the present study, they were also associated with prognosis in univariate analysis, as was our eight-gene signature score (Table 2). We carried out a multivariate analysis to assess the independent prognostic values of these clinicopathological factors and the eight-gene signature score. In this analysis, TILs, RCB score and eight-gene signature score were assessed as continuous factors. The RCB score and the eight-gene signature were two independent predictive factors of metastatic relapse; the lymphovascular emboli were not significant ($P = 0.059$) but improved the predictive value of the model (Table 2).

DISCUSSION

This study is, to the best of our knowledge, the first transcriptomic analysis describing the immune microenvironment of the residual disease in post-NACT TNBC. Residual disease in post-NACT TNBC has been the subject of numerous publications in recent years, due to its particularly poor prognosis. Pathways such as cell-cycle alteration, phosphatidylinositol-3-kinase/mammalian target of rapamycin, DNA repair, Ras/mitogen-activated protein kinase (MAPK) and growth factor receptors were found to be altered in at least 10% of cases in a transcriptomic study from Balko et al. In our cohort, the MAPK pathway was enriched in the residual disease of patients who later experienced relapse. This MEK/MAPK/RAS pathway activation has been inversely correlated with TILs infiltration of residual disease from TNBC patients,[19] and was found overexpressed in the less immunogenic module in the ICR classification.[38] Moreover, MEK inhibition in association with immune agonists was shown to enhance antitumor immune response in preclinical models of TNBC.[19] These findings stress the importance of the immunological context of residual disease.

We identified a combination of eight immune-related genes whose expression accurately predicted RFS in our cohort of 115 TNBC patients with residual disease. Four genes of the signature can be strongly associated with B lymphocytes (*BLK*, *SPIB*, *LILRA1* and *SLAMF7*). Some genes are associated with cytotoxic activation, such as *GZMM* or *SLAMF7* as recently reported.[42] Significance of the chemokine receptors (*CCL4*, *CXCR4*, *CXCR6*) of the signature is more challenging to determine, since they are involved in many different signaling pathways. In our study, these genes seemed linked to lymphocyte activation and recruitment. *CXCR4*, for instance, was correlated to *TNFAIP3* and *CD69* (correlation coefficients 0.69 and 0.62, respectively, data not shown), which are linked to lymphocyte activation. Since most genes are associated with several pathways, we developed an 'extended signature' which includes 63 genes. Pathways involved in this signature were also strongly in favor of a lymphoid activation. This was in line with the analysis of immune cells of the residual disease, with a significant difference between non-metastatic and metastatic patients for all the anti-tumorigenic lymphocytic cells (T lymphocytes, cytotoxic cells, T CD8+ cells, B lymphocytes, NK cells, etc.). Bulk transcriptomic analysis cannot directly show the presence of organized structures such as tertiary lymphoid structures (TLS); however, gene signatures of TLS have been proposed as a reflection of formation of these structures.[41] Since our findings, both on the immune cell infiltrates and of the signature, point toward a coordinated activation of cytotoxic cells, B lymphocytes/plasma cells and Th1 cells, we applied a 12-chemokine gene signature to our cohort and found a significant enrichment in patients who did not experience relapse, suggesting a role of these structures in our TNBC patients.

Finally, this immune activation in non-relapsing patients after chemotherapy may reflect chemotherapy-induced immunogenic cell death, a process that differs from apoptosis and leads to the release of numerous damage-associated molecular patterns by cancer cells, which in turn induces immune response.[43] This process was recently exploited in the TONIC trial, where doxorubicin induction

Figure 3. Design of an immune-related signature to predict survival from patients with residual disease after NACT.
(A and B) KM plot of the eight-gene signature on total cohort and RCB II subgroup. (C) Heatmap representing the median centered gene expression of the eight-gene signature (rows) with patient (columns) ordered by the signature score; column annotations show the presence of metastases, lymphovascular emboli (LVE), the RCB class and TILs levels. (D) Prognostic value of the extended signature on the cohort. For survival analysis, the cohorts are separated by the optimal cut-off. (E and F) 8- and 63- gene signatures on KMplot dataset with basal breast cancer patients. (G and H) 8- and 63- gene signatures on METABRIC dataset selected for TNBC patients ($n = 251$). Cohorts are separated by the optimal score value obtained with ssGSEA on our and METABRIC cohorts and by 'optimal cut-off' in the KMplot cohort. DRFI, disease relapse-free interval; DSS, disease-specific survival; HR, hazard ratio with 95% confidence interval; KM, Kaplan—Meier; NACT, neoadjuvant chemotherapy; RCB, residual cancer burden; RFS, relapse-free survival; TILs, tumor-infiltrating lymphocyte; TNBC, triple-negative breast cancer.

was the most efficient to induce a response to programmed cell death protein 1 blockade therapy.[44]

Another finding of this study is the significant domination of macrophages and/or monocytic cells on the total immune infiltrate. However, we were not able to show any prognostic difference with these cells. This can be explained by the limited number of genes from the panel we used, which did not allow a clear functional analysis of the macrophages.

Our study has some limitations, one being the use of a very limited panel. This does not allow an exhaustive analysis of the metastatic relapse environment. However, the NanoString technology was interesting in this cohort since it can be applied to FFPE samples, which usually have more degraded RNA. Moreover, this technology does not require any amplification, which limits the risk of a bias of poor-quality RNA. The fact that we used older samples with decreased quality of RNA could be seen as another limitation of this study, leading to bias in the results; however, these samples are closer to what can be analyzed in clinical settings. Finally, we validated our signatures on external cohorts. However, RNA expression range values vary between platforms and patient cohorts, thus preventing the definition of a common threshold. Our signatures and their thresholds should be determined in a dedicated validation study to allow the utilization of these signatures in clinical practice.

Overall, we show here that metastatic relapse in TNBC patients with post-NACT residual disease is highly mediated by the immune infiltration, and propose an immunologic prognostic gene signature for TNBC patients with residual disease. This signature can accurately identify high-risk patients from the RCB II subgroup, which could allow the implementation of targeted therapeutic strategies to improve these patients' prognosis.

ACKNOWLEDGEMENTS

We thank the patients for their support in the realization of this study. We thank Pippa McKelvie-Sebileau for medical editing in English and D. Quincy for manuscript submission. The results published here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

FUNDING

This work was supported by the Comité Prévention Dépistage des Cancers (no grant number).

DISCLOSURE

The authors have declared no conflicts of interest.

REFERENCES

1. Dent R, Trudeau M, Pritchard KI, et al. Triple-negative breast cancer: clinical feaStures and patterns of recurrence. *Clin Cancer Res*. 2007;13(15 Pt 1):4429-4434.
2. Liedtke C, Mazouni C, Hess KR, et al. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J Clin Oncol*. 2008;26(8):1275-1281.
3. von Minckwitz G, Untch M, JU Blohmer, et al. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. *J Clin Oncol*. 2012;30(15):1796-1804.
4. Bonnefoi H, Litière S, Piccart M, et al. Pathological complete response after neoadjuvant chemotherapy is an independent predictive factor irrespective of simplified breast cancer intrinsic subtypes: a landmark and two-step approach analyses from the EORTC 10994/BIG 1-00 phase III trial. *Ann Oncol*. 2014;25(6):1128-1136.
5. Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet*. 2014;384(9938):164-172.
6. Masuda N, Lee SJ, Ohtani S, et al. Adjuvant capecitabine for breast cancer after preoperative chemotherapy. *N Engl J Med*. 2017;376(22):2147-2159.
7. Wang X, Wang SS, Huang H, et al. Effect of capecitabine maintenance therapy using lower dosage and higher frequency vs observation on disease-free survival among patients with early-stage triple-negative breast cancer who had received standard treatment: the SYSUCC-001 randomized clinical trial. *JAMA*. 2021;325(1):50-58.
8. Tutt ANJ, Garber JE, Kaufman B, et al. Adjuvant olaparib for patients with BRCA1- or BRCA2-mutated breast cancer. *N Engl J Med*. 2021;384(25):2394-2405.
9. Adams S, Gray RJ, Demaria S, et al. Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. *J Clin Oncol*. 2014;32(27):2959-2966.
10. Denkert C, von Minckwitz G, Brase JC, et al. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers. *J Clin Oncol*. 2015;33(9):983-991.
11. Schmid P, Adams S, Rugo HS, et al. Atezolizumab and nab-paclitaxel in advanced triple-negative breast cancer. *N Engl J Med*. 2018;379(22):2108-2121.
12. Cortes J, Cescon DW, Rugo HS, et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial. *Lancet*. 2020;396(10265):1817-1828.
13. Schmid P, Cortes J, Pusztai L, et al. Pembrolizumab for early triple-negative breast cancer. *N Engl J Med*. 2020;382(9):810-821.
14. Loibl S, Untch M, Burchardi N, et al. A randomised phase II study investigating durvalumab in addition to an anthracycline taxane-based neoadjuvant therapy in early triple-negative breast cancer: clinical results and biomarker analysis of GeparNuevo study. *Ann Oncol*. 2019;30(8):1279-1288.
15. Mittendorf EA, Zhang H, Barrios CH, et al. Neoadjuvant atezolizumab in combination with sequential nab-paclitaxel and anthracycline-based chemotherapy versus placebo and chemotherapy in patients with early-stage triple-negative breast cancer (IMpassion031): a randomised, double-blind, phase 3 trial. *Lancet*. 2020;396(10257):1090-1100.
16. Loi S, Sirtaine N, Piette F, et al. Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. *J Clin Oncol*. 2013;31(7):860-867.
17. Loi S, Drubay D, Adams S, et al. Tumor-infiltrating lymphocytes and prognosis: a pooled individual patient analysis of early-stage triple-negative breast cancers. *J Clin Oncol*. 2019;37(7):559-569.
18. Dieci MV, Criscitiello C, Goubar A, et al. Prognostic value of tumor-infiltrating lymphocytes on residual disease after primary chemotherapy for triple-negative breast cancer: a retrospective multicenter study. *Ann Oncol*. 2014;25(3):611-618.
19. Loi S, Dushyanthen S, Beavis PA, et al. RAS/MAPK activation is associated with reduced tumor-infiltrating lymphocytes in triple-negative breast cancer: therapeutic cooperation between MEK and PD-1/PD-L1 immune checkpoint inhibitors. *Clin Cancer Res*. 2016;22(6):1499-1509.

20. Luen SJ, Salgado R, Dieci MV, et al. Prognostic implications of residual disease tumor-infiltrating lymphocytes and residual cancer burden in triple-negative breast cancer patients after neoadjuvant chemotherapy. *Ann Oncol*. 2019;30(2):236-242.
21. Symmans WF, Peintinger F, Hatzis C, et al. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. *J Clin Oncol*. 2007;25(28):4414-4422.
22. Pinard C, Debled M, Ben Rejeb H, et al. Residual cancer burden index and tumor-infiltrating lymphocyte subtypes in triple-negative breast cancer after neoadjuvant chemotherapy. *Breast Cancer Res Treat*. 2020;179(1):11-23.
23. Hammond ME, Hayes DF, Dowsett M, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Arch Pathol Lab Med*. 2010;134(7):e48-e72.
24. Penault-Llorca F, Vincent-Salomon A, MacGrogan G, et al. [2014 Update of the GEFPICS' recommendations for HER2 status determination in breast cancers in France]. *Ann Pathol*. 2014;34(5):352-365.
25. Dieci MV, Radosevic-Robin N, Fineberg S, et al. Update on tumor-infiltrating lymphocytes (TILs) in breast cancer, including recommendations to assess TILs in residual disease after neoadjuvant therapy and in carcinoma in situ: a report of the International Immuno-Oncology Biomarker Working Group on Breast Cancer. *Semin Cancer Biol*. 2018;52(Pt 2):16-25.
26. Provenzano E, Bossuyt V, Viale G, et al. Standardization of pathologic evaluation and reporting of postneoadjuvant specimens in clinical trials of breast cancer: recommendations from an international working group. *Mod Pathol*. 2015;28(9):1185-1201.
27. Ayers M, Lunceford J, Nebozhyn M, et al. IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest*. 2017;127(8):2930-2940.
28. Goldman MJ, Craft B, Hastie M, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol*. 2020;38(6):675-678.
29. Chin K, DeVries S, Fridlyand J, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell*. 2006;10(6):529-541.
30. Hess KR, Anderson K, Symmans WF, et al. Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol*. 2006;24(26):4236-4244.
31. Waggott D, Chu K, Yin S, Wouters BG, Liu FF, Boutros PC. NanoStringNorm: an extensible R package for the pre-processing of NanoString mRNA and miRNA data. *Bioinformatics*. 2012;28(11):1546-1548.
32. Csardi G, Nepusz T. The igraph software package for complex network research. *Int J Complex Syst*. 2006;1695(5):1-9.
33. Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Softw*. 2008;25(1):18.
34. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-287.
35. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550.
36. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013;14:7.
37. Becht E, Giraldo NA, Lacroix L, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. 2016;17(1):218.
38. Hendrickx W, Simeone I, Anjum S, et al. Identification of genetic determinants of breast cancer immune phenotypes by integrative genome-scale analysis. *Oncoimmunology*. 2017;6(2):e1253654.
39. Gyorffy B. Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. *Comput Struct Biotechnol J*. 2021;19(2021):4101-4109.
40. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403):346-352.
41. Coppola D, Nebozhyn M, Khalil F, et al. Unique ectopic lymph node-like structures present in human primary colorectal carcinoma are identified by immune gene array profiling. *Am J Pathol*. 2011;179(1):37-45.
42. Loyal L, Warth S, Jürchott K, et al. SLAMF7 and IL-6R define distinct cytotoxic versus helper memory CD8(+) T cells. *Nat Commun*. 2020;11(1):6357.
43. Casares N, Pequignot MO, Tesniere A, et al. Caspase-dependent immunogenicity of doxorubicin-induced tumor cell death. *J Exp Med*. 2005;202(12):1691-1701.
44. Voorwerk L, Slagter M, Horlings HM, et al. Immune induction strategies in metastatic triple-negative breast cancer to enhance the sensitivity to PD-1 blockade: the TONIC trial. *Nat Med*. 2019;25(6):920-928.